Ab Initio Studies of Amino Acid Conformations. 1. The Conformers of Alanine, Serine, and Cysteine

Scott Gronert*,† and Richard A. J. O'Hair‡

Contribution from the Department of Chemistry and Biochemistry, San Francisco State University, San Francisco, California 94132, and the Department of Chemistry, Kansas State University, Manhattan, Kansas 66506

Received May 9, 1994[⊗]

Abstract: High-level ab initio methods (up to MP2/6-31+G*) have been used to characterize the gas phase conformations of alanine, serine, and cysteine. A wide range of possible structures (36 for alanine and 324 for serine and cysteine) was surveyed at the AM1 level, and then the geometries of the unique conformers were refined at the 3-21G(*) and 6-31G* levels. At the highest theoretical level, 10 conformers were located for alanine, 51 for serine, and 42 for cysteine. The AM1 level provides a poor description of the relative energies. Calculations at the 3-21G(*) level represent a significant improvement, but some bonding schemes are poorly characterized. Better values are obtained at the HF/6-31G* level, but to obtain reasonably accurate relative energies, correlation corrections are required and calculations at the MP2/6-31+G*//HF/6-31G* level give values in good accord with a series of test calculations at the MP2/6-31+G*//MP2/6-31+G* level. Ab initio rotational constants and dipole moments are reported for all the conformers. The results are compared to previous studies of amino acids and are analyzed in terms of intramolecular hydrogen-bonding interactions.

Introduction

Amino acids are an attractive target for computational chemists because they contain a variety of intramolecular interactions, are conformationally labile, and are of a tractable size for high-level ab initio calculations. Moreover, theoretical studies can provide important information to help guide experimentalists in efforts to identify gaseous amino acids by their microwave or IR spectra. Over the years, several groups have reported computational studies of α-amino acids,[1−12] but with a few notable exceptions, most groups have only considered a subset of the possible conformations. Glycine has been the most widely studied of the amino acids, and several groups have presented thorough investigations.[1−3] Most recently, Schaefer[2a] and Császár[2b] have reported high-level calculations with extensive correlation corrections. In the case of alanine, Brown and co-workers[4] have examined the potential energy surface at the HF/6-31G** level and compared the optimized structures with microwave data. For the more complex amino acids, previous studies have been at low levels of theory and have only considered a set of model conformations chosen on the basis of favorable intramolecular interactions (*i.e.*, hydrogen bonding). For example, Schäfer and co-workers have completed a series of studies on amino acids such as serine[6] and cysteine.[8] Because it is impossible to use such criteria to pick, *a priori*, all of the active conformations, this type of approach yields only a subset of the true conformational space. Moreover, this subset is biased by the assumptions made in choosing the initial set of "trial" structures. To truly characterize conformationally labile molecules such as amino acids, the full ensemble of possible conformations must be considered.

As an extension of our preliminary survey of the potential energy surfaces at the 3-21G(*) level,[12] we now report a comprehensive study of the conformations of alanine, serine,

† San Francisco State University.

‡ Kansas State University.

⊗ Abstract published in *Advance ACS Abstracts*, February 1, 1995.

(1) Jensen, J. H.; Gordon, M. S. *J. Am. Chem. Soc.* **1991**, *113*, 7917.

(2) (a) Hu, C.-H.; Shen, M.; Schaefer, H. F., III. *J. Am. Chem. Soc.* **1993**, *115*, 2923. (b) Császár, A. G. *J. Am. Chem. Soc.* **1992**, *114*, 9568. (c) Yu, D.; Armstrong, D. A.; Rauk, A. *Can. J. Chem.* **1992**, *70*, 1762.

(3) (a) Ramek, M.; Cheng, V. K. W.; Frey, R. F.; Newton, S. Q.; Schäfer, L. *J. Mol. Struct. (THEOCHEM)* **1991**, *235*, 1. (b) Schäfer, L.; Sellers, H. L.; Lovas, F. J.; Suenram, R. D. *J. Am. Chem. Soc.* **1980**, *1202*, 6566. (c) Sellers, H. L.; Schäfer, L. *J. Am. Chem. Soc.* **1978**, *100*, 7728.

(4) Godrey, P. D.; Firth, S.; Hatherley, L. D.; Brown, R. D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1993**, *115*, 9687.

(5) (a) Sellers, H. L.; Schäfer, L. *Chem. Phys. Lett.* **1979**, *63*, 609. (b) Siam, K.; Klimkowski, V. J.; Ewbank, J. D.; Van Alsenoy, C.; Schäfer, L. *J. Mol. Struct. (THEOCHEM)* **1984**, *110*, 171.

(6) (a) Van Alsenoy, C.; Scarsdale, J. N.; Sellers, H. L.; Schäfer, L. *Chem. Phys. Lett.* **1981**, *80*, 124. (b) Van Alsenoy, C.; Kulp, S.; Siam, K.; Klimkowski, V. J.; Ewbank, J. D.; Schäfer, L. *J. Mol. Struct. (THEOCHEM)* **1988**, *181*, 169.

(7) Mezey, P. G.; Ladik, J. J.; Suhai, S. *Theor. Chim. Acta* **1979**, *51*, 323.

(8) Schäfer, L.; Siam, K.; Klimkowski, V. J.; Ewbank, J. D.; Van Alsenoy, C. *J. Mol. Struct. (THEOCHEM)* **1990**, *204*, 361.

(9) (a) Wright, L. R.; Borkman, R. F. *J. Am. Chem. Soc.* **1980**, *102*, 6207. (b) Laurence, P. R.; Thomson, C. *Theor. Chim. Acta* **1981**, *58*, 121.

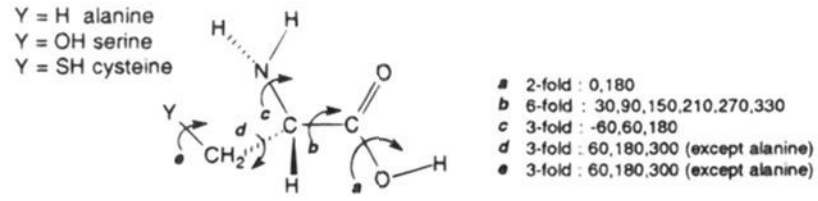
(10) (a) Masamura, M. *THEOCHEM* **1988**, *168*, 227. (b) Ramek, M. *THEOCHEM* **1990**, *67*, 301. (c) Ramek, M. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1990**, *17*, 45. (d) Gatti, C.; Branchi, R.; Destro, R.; Merati, F. *THEOCHEM* **1992**, *87*, 409. (e) Barron, L. D.; Gargaro, A. R.; Hecht, L.; Polavarapu, P. L. *Spectrochim. Acta, Part A* **1991**, *47A*, 1001. (f) Chipot, C.; Maigret, B.; Rivail, J.-L.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 10276. (g) de Dios, A. C.; Pearson, J. G.; Oldfield, E. *J. Am. Chem. Soc.* **1993**, *115*, 9768. (h) Vishveshwara, S.; Pople, J. A. *J. Am. Chem. Soc.* **1977**, *99*, 2422. (i) Ford, G. P.; Wang, B. *J. Am. Chem. Soc.* **1992**, *114*, 10563.

(11) Ab initio calculations have been completed on amino acid derivatives. For example, see the following: (a) Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1991**, *113*, 6256. (b) Perczel, A.; Daudel, R.; Angyan, J. G.; Csizmadia, I. G. *Can. J. Chem.* **1990**, *68*, 1882. (c) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L., III; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989. (d) Price, S. L.; Andrews, J. S.; Murray, C. W.; Amos, R. D. *J. Am. Chem. Soc.* **1992**, *114*, 8268. (e) Böhm, H.-J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129. (f) Ewbank, J. D.; Klimkowski, V. J.; Siam, K.; Schäfer, L. *J. Mol. Struct.* **1987**, *160*, 275. (g) Siam, K.; Klimkowski, V. J.; Van Alsenoy, C.; Ewbank, J. D.; Schäfer, L. *J. Mol. Struct. (THEOCHEM)* **1987**, *152*, 261. (h) Böhm, H.-J. *J. Am. Chem. Soc.* **1993**, *115*, 6152. (i) Bour, P.; Keiderling, T. A. *J. Am. Chem. Soc.* **1993**, *115*, 9602. (j) Schäfer, L.; Newton, S. Q.; Cao, M.; Peeters, A.; Van Alsenoy, C.; Wolinski, K.; Momany, F. A. *J. Am. Chem. Soc.* **1993**, *115*, 272. (k) 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.

(12) O'Hair, R. A. J.; Bowie, J. H.; Gronert, S. *Int. J. Mass Spectrom. Ion Processes* **1992**, *117*, 23.

0002-7863/95/1517-2071$09.00/0

Scheme 1



and cysteine. Since our goal was to identify all of the minima on the conformational potential energy surface, we initially surveyed a wide range of conformers at the semiempirical level (AM1).[13a] The starting sets of conformations were chosen by allowing for all possible combinations of single-bond rotamers (see below). The unique conformations from the AM1 optimizations[13b] were then subjected to further optimization at the HF/3-21G(*) and HF/6-31G* levels. In addition, frequency analysis was completed for the final set of structures, and for selected conformers of each amino acid, ab initio IR intensities are given (supplementary material). Rotational constants and dipole moments are listed for all of the conformers.

Methods

All calculations were completed on HP-720 or HP-735 workstations at the San Francisco State University Computational Chemistry Center using the Gaussian92[14] or Gamess[15] quantum mechanical program. Basis sets were taken from the Gaussian92 library.[16] A series of trial structures was generated for each amino acid by allowing for all combinations of single-bond rotamers (Scheme 1).

For the carboxyl group (***a***), syn or anti conformations were considered. The interaction of the α-carbon with the caboxyl (***b***) is the most complex and leads to six different possibilities. Assuming a staggered conformation, one of the three groups on the α-carbon will be separated from the other two groups by the plane of the carbonyl. Because of asymmetry in the carboxyl, placement of the unique group above or below the carbonyl plane leads to different conformations. The orientation of the NH_2 group (***c***) allows for three rotamers. If the β-carbon is substituted, there are three possible rotamers about the $C_\alpha-C_\beta$ bond (***d***). Finally, for serine and cysteine, the C−O and C−S bonds, respectively, represent 3-fold rotors (***e***). This leads to a total of 36 trial structures for alanine and 324 trial structures for serine and cysteine.

Once AM1 optimizations were completed for all of the trial structures, a set of unique conformations was identified. Structures were considered identical if their energies differed by less than 10^{-5} hartree and if the root-mean-square difference of their rotational frequencies differed by less than 30 MHz. The unique conformations were then subjected to optimization at the HF/3-21G(*) level, and again the unique structures were identified. Finally, optimizations and analytical frequency analysis[17] were completed at the HF/6-31G* level. For glycine and a derivative of alanine, it has been pointed out that correlation corrections and diffuse functions can have a significant effect on the relative stabilities of the possible conformers.[1-3,11k] To take this into account, MP2/6-31+G* calculations were completed on the HF/6-31G*-optimized structures. To test the validity of this approach,

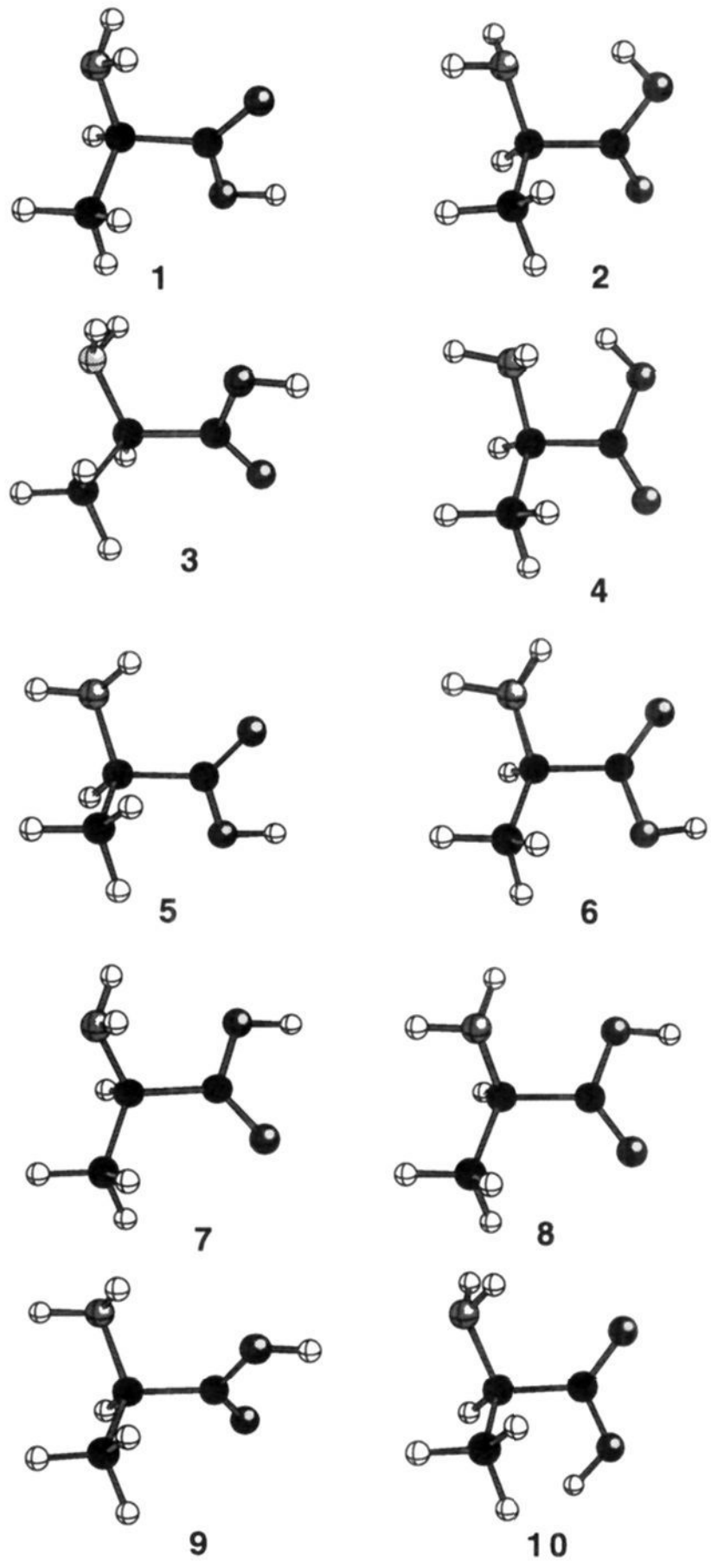


Figure 1. Conformers of alanine optimized at the HF/6-31G* level.

MP2/6-31+G* optimizations were completed for several conformations of each amino acid. The results from these tests suggest that optimizations can be completed without correlation or diffuse functions (see below). In addition, a set of calculations was completed at the MP4 level for alanine and serine (see below).

For the purpose of analyzing the large data sets, conformations were characterized by their near-atom interactions (hydrogen bonding). For the purposes of this study, a distance of 2.75 Å was used as a cutoff for near-atom interactions.[18] Although this approach is somewhat simplistic, it can be more useful than a dihedral angle list for illustrating the complex interactions found in the amino acids (listings of important dihedral angles are given in Appendix A). This is especially true when considering single amino acids as opposed to peptide chains.

(13) (a) AM1 as implemented in Gaussian92-Rev. C (ref 13), see: Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902. (b) PM3 was also investigated, but more unique conformers were identified at the AM1 level. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.

(14) Gaussian92-Rev. C: Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. H.; Foresman, J. B.; Johnson, B. D.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; DeFrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. Gaussian, Inc., Pittsburgh, PA, 1992.

(15) GAMESS: Schmidt, M. W.; Boatz, J. A.; Baldridge, K. K.; Koseki, S.; Gordon, M. S.; Elbert, S. T.; Lam, B. *QCPE Bull.* **1978**, *7*, 115.

(16) For details, see Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Intio Molecular Orbital Theory*; Wiley-Interscience: New York, 1986.

(17) Frequencies are scaled by 0.9. See: Pople, J. A.; Schlegel, H. B.; Krishnan, R.; DeFrees, D. J.; Binkley, J. S.; Frisch, M. J.; Whiteside, R. A.; Hout, R. F.; Hehre, W. J. *Int. J. Quantum Chem. Symp.* **1981**, *15*, 269.

(18) This distance is arbitrary, but it is consistent with the typical intramolecular interaction distances found in these systems. This is long for a typical hydrogen bond, but geometry constraints force them to be long-range interactions in these molecules.

Table 1. Relative Energies, Interactions, and Rotational Data for Alanine Conformers[a]

	relative energies						rotational spectra			
	HF	HF+	MP2+	ZPE	carboxyl interaction[b]	NH_2 interaction[c]	A	B	C	dipole (D)
1	0.0	0.0	0.0	66.02	C=O	C=O	5.146	3.163	2.270	1.50
2	2.9	2.7	0.9	66.36	NH_2		5.074	3.184	2.381	5.69
3	1.5	1.2	1.0	66.02	C=O	C(O)OH	5.137	2.934	2.463	1.79
4	2.7	2.8	1.2	66.31	NH_2		4.969	3.472	2.172	5.64
5	1.8	1.6	1.3	65.95	C=O	C=O	5.158	3.046	2.383	2.48
6	1.9	1.7	1.4	65.95	C=O	C=O	5.142	3.374	2.176	2.42
7	1.6	1.5	1.5	65.98	C=O	C(O)OH	5.137	3.284	2.189	1.52
8	2.0	1.9	1.7	65.99	C=O	C(O)OH	5.184	3.345	2.193	2.74
9	2.5	2.2	2.0	65.94	C=O	C(O)OH	5.123	2.877	2.549	2.98
10	7.0	7.0	6.4	65.84		C=O	5.086	3.155	2.287	3.63

[a] Geometries optimized at the HF/6-31G* level and energies at the HF/6-31G* (HF), HF/6-31+G* (HF+), and MP2/6-31+G* (MP2+) levels (kcal/mol). Zero-Point energies (ZPE) in kcal/mol scaled by 0.9. Rotational constants in GHz. Dipole from HF/6-31+G* wave function. [b] Group that interacts with carboxyl proton. [c] Group that has near-atom interaction with the hydrogen of the NH_2 group. For carboxyl, it is the hydroxy oxygen that interacts.

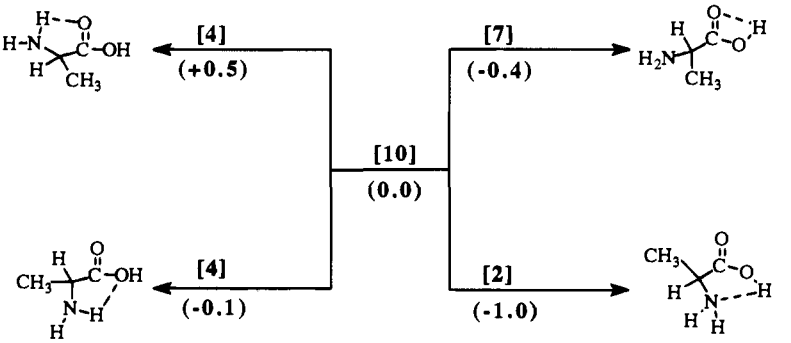


Figure 2. Bonding schemes for alanine. Brackets contain the total number of conformers of each bonding scheme at the 6-31G* level. The average energy of the bonding scheme (relative to the overall average) is given parenthetically for the MP2/6-31+G*//HF/6-31G* level.

Table 2. Linear Regression Analysis of Near-Atom Interactions in Alanine Conformers[a]

interactions[b]			
NH_2 C(O)−OH	NH_2 C=O	COOH NH_2	COOH C=O
−1.3	−2.2	−8.0	−5.7
(0.6)	(0.5)	(0.3)	(0.5)
[4]	[4]	[2]	[7]

[a] MP2/6-31+G*//HF/6-31G*. Energies in kcal/mol. Standard errors given parenthetically. Number of occurrences given in brackets. [b] Hydrogen-bond donor listed above acceptor.

Results and Discussion

Alanine. Optimization at the AM1 level led to 13 unique structures for alanine. After refinement at the highest ab initio level, nine structures (Figure 1) remained with energies varying by ~7 kcal/mol (Table 1). A tenth structure was located by direct optimization at the 3-21G(*)[19] and 6-31G* levels (see below). In our earlier, partial survey of the potential energy surface,[12] we also identified conformer **1** as the global minimum at the 3-21G(*) level. In their studies at the HF/6-31G** level, Brown and co-workers[4] found only six conformations for alanine. Their conformations **I**, **II**, **III**, **IV**, **V**, and **VI** correspond to conformations **1**, **4**, **2**, **3**, **7**, and **10**, respectively, in the present study—they did not locate structures **5**, **6**, **8**, or **9**. The relative energies and rotational constants reported by Brown for the six structures in common are in good accord with those found in the present study. Brown and co-workers[4] do not describe their strategy for locating conformers, but it was probably based on identifying logical interactions (*i.e.*, hydrogen bonding). *The fact that this approach excluded several conformers in a relatively simple amino acid clearly points out the need to undertake a systematic survey of all the reasonable rotamers in amino acid studies.* All of the additional structures are related to those of Brown and co-workers[4] by rotation around the C−N bond, indicating that this bond must be treated as a true 3-fold rotor. The missing structures should not affect the microwave spectrum because there is probably only a very small barrier separating them from more stable conformers (**1** and **3**). Our initial AM1 survey did not locate conformer **IV** in Brown and co-workers' study.[4] It was found by starting with a structure resembling **IV** and then optimizing at the 3-21G(*) and 6-31G* levels. This conformer (**3**) is related to **7** by an 85° rotation around the C_α−COOH bond. The rotation simply switches the identity of the amino hydrogen that interacts with the OH of the carboxyl. The absence of this conformer in our initial search points out an important weakness in the approach. We are dependent on the AM1 method to identify all of the possible conformers before the geometries are refined at higher levels. If any conformers are missed at this level, they cannot be recovered. The potential energy surface separating conformers **3** and **7** is exceptionally flat at the AM1 level, and the method cannot identify them as discrete species. In fact, the AM1 calculation results in a structure where the carboxyl interacts equally with both of the amino hydrogens (a conformer midway between **3** and **7**). Since only one conformer of this type is identified at the AM1 level, only one of the pair is found at the HF/6-31G* level. Although both exist at the HF/6-31G* level, the barrier to interconversion is very small (~0.2 kcal/mol). With the other amino acids in this study (serine and cysteine), we anticipate that some minima may have been missed for the same reason; however, it is likely that they are closely related to identified conformers and separated from those conformers by an exceptionally small barrier. Therefore, their absence should have little effect on the overall conclusions.

(19) The basis set is referred to as 3-21G(*) in the text, but for alanine and serine this is equivalent to a 3-21G basis set. Only in cysteine are polarization functions included (on sulfur).

Table 3. Relative Energies, Interactions, and Rotational Data for Serine Conformers[a]

	relative energies						rotational spectra			
	HF	HF+	MP2+	ZPE	bonding scheme[b]	NH_2 interaction[c]	A	B	C	dipole (D)
1	0.0	0.0	0.0	69.73	**C**	C=O	4.579	1.846	1.446	1.97
2	1.2	1.5	0.1	70.11	**D**	CH_2OH	3.679	2.406	1.734	4.40
3	0.7	0.7	0.2	69.72	**C**	C=O	3.590	2.333	1.816	3.22
4	0.8	1.0	1.2	69.86	**B**	CH_2OH, C(O)OH	3.950	2.292	1.679	2.72
5	2.1	2.4	1.4	70.10	**D**		3.721	2.378	1.531	5.18
6	1.3	1.0	1.4	69.48	**F**	CH_2OH, C=O	3.539	2.328	1.811	2.95
7	1.7	1.7	1.5	69.60	**A**	CH_2OH, C=O	3.505	2.423	1.755	2.93
8	1.7	1.7	1.8	69.66	**C**	C(O)OH*	4.837	1.837	1.389	3.90
9	1.4	1.2	2.0	69.71	**B**	C(O)OH	3.613	2.354	1.573	3.35
10	3.5	3.1	2.0	69.87	**D**	CH_2OH	4.608	1.853	1.485	4.81
11	2.8	2.7	2.1	69.73	**C**	C(O)OH	3.628	2.287	1.829	4.00
12	1.5	1.5	2.2	69.73	**B**	C(O)OH	3.717	2.353	1.553	3.05
13	3.7	3.5	2.2	69.90	**D**	CH_2OH	4.569	1.839	1.480	4.57
14	3.1	2.8	2.3	69.74	**D**	CH_2OH	4.787	1.870	1.401	4.58
15	2.3	2.0	2.4	69.52	**F**	CH_2OH, C(O)OH	3.589	2.292	1.818	0.61
16	1.9	2.2	2.5	69.85	**B**	CH_2OH	4.093	2.172	1.704	2.31
17	2.3	2.3	2.6	69.51	**A**	C=O*	3.447	2.459	1.540	1.94
18	3.6	3.3	2.8	69.71	**A**	CH_2OH	3.932	2.104	1.804	1.79
19	3.7	3.5	2.9	69.36	**D**	CH_2OH	4.747	1.852	1.393	4.24
20	2.7	2.6	3.0	69.49	**F**	C=O	3.044	2.563	1.734	2.67
21	2.6	2.1	3.0	69.46	**F**	CH_2OH, C(O)OH	4.983	1.812	1.413	1.09
22	3.0	2.5	3.1	69.44	**F**	C=O	3.087	2.469	1.773	0.63
23	3.4	2.7	3.1	69.38	**F**	CH_2OH, C=O	4.775	1.875	1.401	2.67
24	4.5	4.3	3.3	69.64	**E**	C=O	3.563	2.498	1.548	5.36
25	3.9	3.7	3.4	69.63	**A**	CH_2OH	3.894	2.090	1.839	1.35
26	3.6	3.3	3.4	69.45	**F**	CH_2OH, C=O	4.518	1.807	1.478	2.94
27	3.3	2.9	3.4	69.44	**F**	C(O)OH	3.080	2.442	1.799	1.74
28	3.4	3.1	3.7	69.51	**C**	CH_2OH, C(O)OH	4.986	1.823	1.397	2.57
29	4.2	3.9	3.7	69.40	**A**	C=O	3.343	2.487	1.590	0.93
30	3.7	3.0	3.7	69.35	**F**	C(O)OH	3.167	2.368	1.808	3.07
31	5.4	5.7	3.7	70.23	**E**	C=O	3.987	2.279	1.748	5.04
32	3.1	2.9	3.8	69.56	**B**		3.455	2.346	1.645	1.73
33	3.5	3.1	3.8	69.43	**F**	CH_2OH, C(O)OH	4.954	1.797	1.405	3.25
34	4.2	3.5	4.2	69.35	**F**	CH_2OH, C(O)OH	4.625	1.730	1.546	3.82
35	4.2	4.0	4.2	69.38	**F**	C=O	3.072	2.499	1.748	2.80
36	4.4	3.9	4.2	69.42	**F**	CH_2OH, C(O)OH	4.573	1.739	1.526	1.40
37	4.2	3.5	4.3	69.32	**F**	C=O	3.158	2.363	1.779	2.69
38	4.4	3.9	4.3	69.36	**F**	CH_2OH, C(O)OH	3.509	2.269	1.892	2.93
39	6.4	5.8	4.7	69.79	**E**	CH_2OH	3.931	2.197	1.775	5.34
40	4.7	4.3	4.7	69.30	**F**	C(O)OH	3.100	2.430	1.787	2.92
41	5.4	4.6	4.9	69.22	**F**	CH_2OH, C(O)OH	4.591	1.721	1.539	3.57
42	5.4	4.8	5.0	69.25	**F**	C(O)OH	3.159	2.353	1.792	3.54
43	5.2	4.5	5.2	69.21	**F**	C(O)OH	3.168	2.357	1.795	4.80
44	7.2	6.8	6.1	69.29	**E**	C=O	3.420	2.543	1.593	6.27
45	6.9	6.9	6.3	69.59	**H**	C=O	4.505	1.842	1.458	2.25
46	6.2	5.8	6.6	69.28	**F**		3.172	2.341	1.795	2.88
47	8.6	8.0	7.3	69.45	**E**	CH_2OH	3.866	2.155	1.837	7.36
48	8.6	8.8	8.2	69.70	**I**	CH_2OH	4.066	2.216	1.687	3.70
49	10.0	9.8	9.6	69.92	**G**	CH_2OH, C=O	3.513	2.314	1.811	5.39
50	12.3	12.0	11.6	68.99	**G**	C=O	3.492	2.243	1.885	6.76
51	12.9	12.5	12.0	69.03	**G**	CH_2OH, C=O	4.629	1.850	1.407	5.45

[a] Geometries optimized at the HF/6-31G* level and energies at the HF/6-31G* (HF), HF/6-31+G* (HF+), and MP2/6-31+G* (MP2+) levels (kcal/mol). Zero-Point energies (ZPE) in kcal/mol scaled by 0.9. Rotational constants in GHZ. Dipole from HF/6-31+G* wave function. [b] Bonding schemes as shown in Figure 3. **F** stands for syn carboxyl, but no interaction for the hydrogen of the CH_2OH group. **G** indicates no interactions for the hydrogens of either the carboxyl or the CH_2OH group. **H** indicates no interaction for carboxyl proton, but CH_2OH hydrogen interacts with NH2 group. **I** indicates no interaction for carboxyl proton, but CH_2OH hydrogen interacts with C=O. [c] Groups that have near-atom interaction with the hydrogen(s) of the NH_2 group. An asterisk indicates that both hydrogens interact with the group.

In Table 1, a comparison of the Hartree–Fock and MP2 relative energies indicates that correlation has its greatest affect on conformers **2** and **4**. To test the validity of the MP2/6-31+G*//HF/6-31G* energies, conformers **1** and **4** were reoptimized at the MP2/6-31+G* level, and an energy difference of 1.1 kcal/mol was found (1.2 kcal/mol at MP2/6-31+G*//HF/6-31G*). As a further test, the MP2-optimized geometries were used for MP4(SDQ)/6-31+G* calculations. The resulting MP4 relative energy is 1.5 kcal/mol. These results and others (see below for serine) suggest that the MP2/6-31+G*//HF/6-31G* calculations provide an acceptable estimate of the correlation energy; therefore, these values will be used in all the discussions.

The interactions listed in Table 1 and shown in Figure 2 give insight into the factors that affect the stability of amino acid conformers. In 9 of the 10 structures, the carboxyl proton is involved in interactions with either the carbonyl or the nitrogen. The former is the more common and is incorporated in seven of the structures. The two structures where the carboxyl proton hydrogen bonds to nitrogen are ~1 kcal/mol less stable than the global minimum. It should be noted that zero-point energies vary significantly between the different bonding schemes; therefore, frequency analysis is important in assigning relative energies. The one structure that lacks an interaction with the carboxyl proton (**10**) is by far the least favorable conformation (6.4 kcal/mol less stable than **1**). There appears to be little

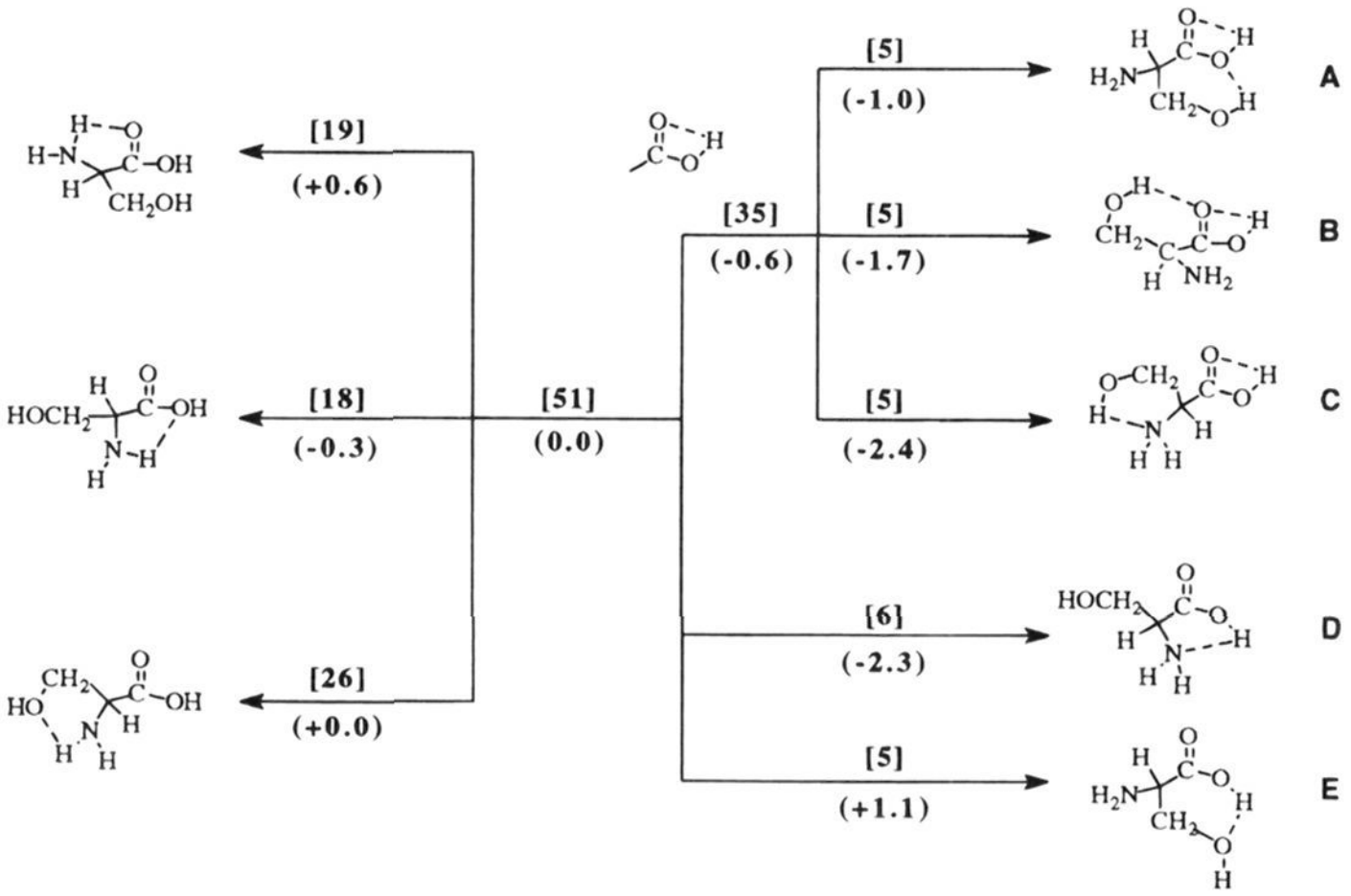


Figure 3. Bonding schemes for serine. Brackets contain the total number of conformers of each bonding scheme at the 6-31G* level. The average energy of the bonding scheme (relative to the overall average) is given parenthetically for the MP2/6-31+G*//HF/6-31G* level.

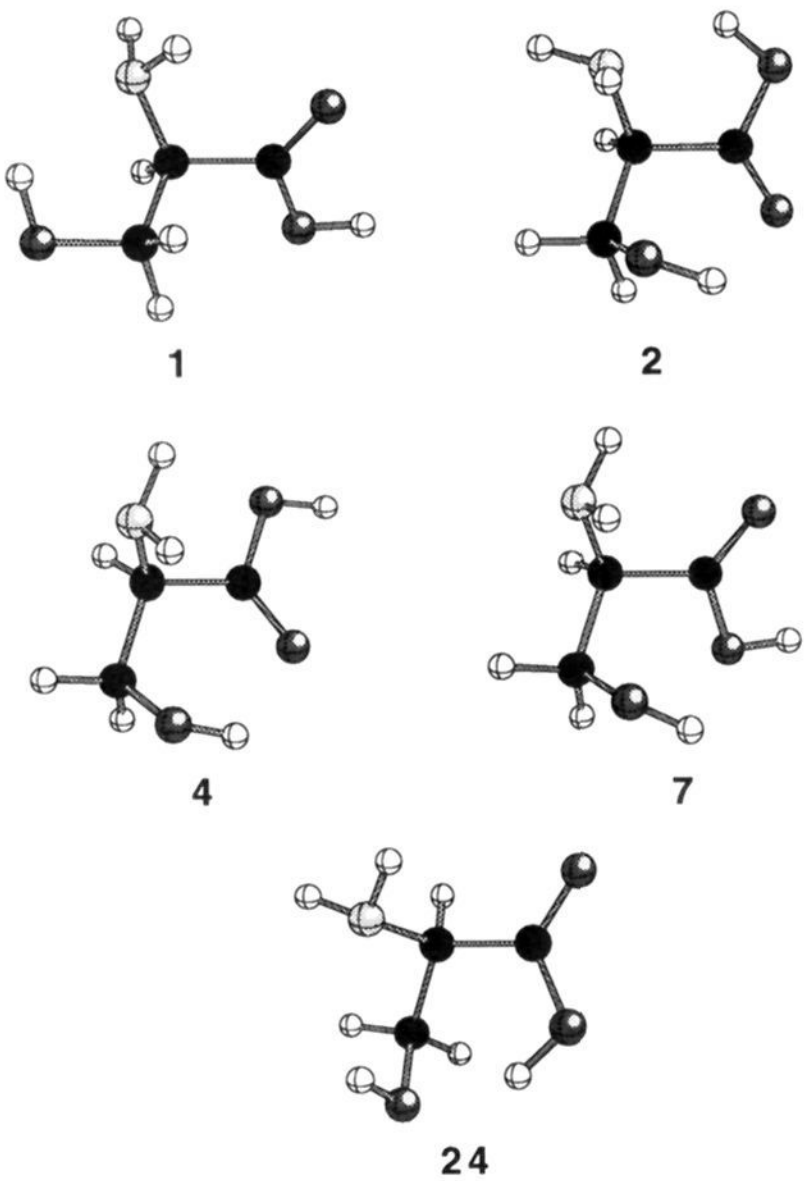


Figure 4. Sample conformers of serine optimized at the MP2/6-31+G* level. Each represents one of the fundamental bonding schemes (**A**−**E**) in Figure 3.

preference in the two interactions available to the amino hydrogens. For example, conformers **6** and **7** have similar energies and are virtually identical except for the fact that the amino hydrogens interact with the carbonyl oxygen in **6** and carboxyl OH in **7**.

The data can also be analyzed by a linear regression approach.[20,21] If one assumes that each interaction provides a constant amount of stabilization and that these stabilizations are additive, then the relative energies can be dissected into contributions from the various hydrogen-bonding interactions. Although this approach has obvious defects, it is useful in *qualitatively* evaluating the importance of the intramolecular interactions, particularly in the more complicated amino acids. Any quantitative interpretation of these data are dangerous, especially when the sample size is small. The derived interaction energies for alanine are listed in Table 2. In accord with the analysis given above, these results indicate the carboxyl proton's interactions are the most important. The regression also suggests that the advantage of a syn carboxyl group (conformers **1**, **3**, **5**−**9**) is that this bonding scheme allows for additional interactions—the amino group is free to act as a hydrogen-bond donor to either the hydroxyl or carbonyl oxygens in this case.

Serine. The presence of a hydrogen-bonding group in the side chain of serine leads to a level of complexity not found in simple amino acids such as glycine or alanine. For one, the substitution of an HO group adds two new 3-fold rotors and therefore increases the number of possible conformations by nearly an order of magnitude (324 trial structures). In addition, the HO group may act as a hydrogen-bond donor or acceptor with the NH_2 or the carboxyl group of the amino acid. It should be noted that the extensive set of trial structures was necessary and that if any of the rotors is eliminated or simplified, some conformations are lost. From the AM1 optimizations, 73 unique conformers were identified and subjected to optimization at the HF/3-21G(*) and HF/6-31G* levels. At the highest level, 51 unique conformations were located with energies varying by ~12 kcal/mol (Table 3). As in alanine, there is a strong

(20) A standard multiple regression approach was used: Excel 4.0, Microsoft Corp., Redmond, WA.

(21) To insure that the intercept was chemically meaningful, a hypothetical conformer with no interactions was included in the data set. For this purpose, constrained structures were used to estimate the energy. The energy used for the interactionless structure will have a modest effect on the absolute interaction energies.

Table 4. Linear Regression Analysis of Near-Atom Interactions in Serine Conformers[a]

interactions[b]								
NH_2 C(O)—OH	NH_2 C=O	NH_2 CH_2OH	COOH NH_2	COOH CH_2OH	COOH C=O	CH_2OH C(O)—OH	CH_2OH C=O	CH_2OH NH_2
−0.9	−2.4	−1.5	−10.9	−7.2	−8.4	−0.8	−3.3	−3.2
(0.5)	(0.4)	(0.3)	(0.6)	(0.6)	(0.5)	(0.7)	(0.5)	(0.5)
[18]	[19]	[26]	[6]	[5]	[35]	[5]	[8]	[7]

[a] MP2/6-31G*//HF/6-31G*. Energies in kcal/mol. Standard errors given parenthetically. Number of occurrences given in brackets. [b] Hydrogen-bond donor listed above acceptor.

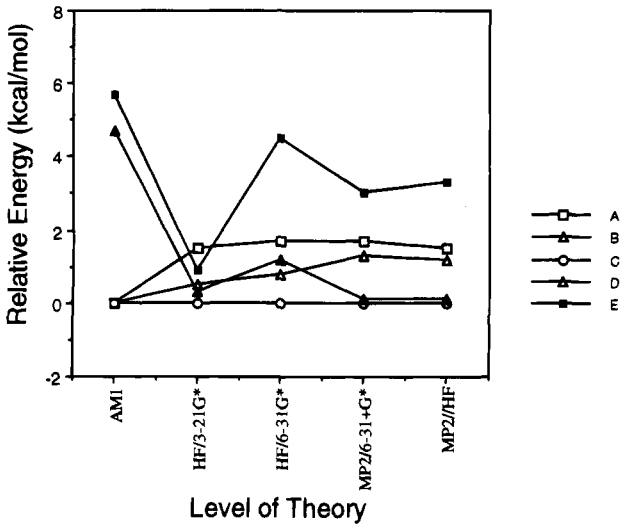


Figure 5. Plot of relative energy vs level of theory for serine conformers in Figure 4. MP2/HF refers to MP2/6-31+G*//HF/6-31G*.

preference for hydrogen bonding to the carboxyl proton. Of the 51 structures (Figure 3), 35 have syn carboxyls (bonding schemes **A**−**C**), 6 have hydrogen-bonding between the carboxyl proton and the nitrogen (**D**), and 5 have hydrogen-bonding between the carboxyl proton and the hydroxyl side chain (**E**). For each of the bonding schemes outlined on the right side of Figure 3 (**A**−**E**), a drawing of the most stable representative is shown in Figure 4. As a way of analyzing the general stability of the bonding schemes, the average energies of the conformers within each scheme (relative to the overall average) are shown in Figure 3. The interaction with the hydroxyl (**E**) is less favorable (average energy +1.1 kcal/mol) than those with the nitrogen (**D**) or the carbonyl (**A**−**C**) group (average energies of −2.3 and −0.6 kcal/mol, respectively). This order of stability is also suggested by the interaction energies in Table 4. These values indicate that the interaction with the amino group is the strongest and is favored by 2−3 kcal/mol over interactions with the carbonyl or the side chain hydroxyl.

As a hydrogen-bond donor, the side-chain hydroxyl can interact with the carboxyl OH (**A**), the carbonyl oxygen (**B**), or the nitrogen (**C**). For syn carboxyls, these three interactions are illustrated in Figure 3. Although each have an equal occurrence rate, interactions with the nitrogen (−2.4 kcal/mol) and the carbonyl (−1.7 kcal/mol) are favored over interactions with the carboxyl OH (−1.0 kcal/mol). The preference for scheme **B** over scheme **A** is obviously a result of the greater basicity of the carbonyl oxygen. Although the nitrogen is more basic than the carbonyl, it appears that geometric factors (**B** involves a 6-membered hydrogen-bonding ring whereas **C** involves a 5-membered ring) counteract this effect to some extent. This is more apparent in the regression values in Table 4.

The NH_2 group also acts as a hydrogen-bond donor and can interact with the carbonyl oxygen, the carboxyl OH, or the side-chain OH. All of these interactions have a high occurrence level and the average energies of conformers containing these interactions differ little from the overall average (Figure 3). The regression analysis provides more insight in this case and indicates that there is a definite preference for hydrogen bonding to the carbonyl oxygen. The interactions with the two OH groups (carboxyl and side-chain) result in smaller stabilization energies.

An important aspect of this project is assessing the abilities of various levels of theory to characterize the conformations of amino acids. This is critical because for practical reasons, ab initio studies of larger systems (*i.e.*, small peptides) must be limited to the lowest level of theory capable of adequately reproducing energies and structures. Serine is a good test case because like many amino acids, it contains the added complexity of a hydrogen-bonding group in the side chain. In our studies, we characterized the full set of serine conformations at the AM1, HF/3-21G(*), and HF/6-31G* levels of theory. In the progression from one level of theory to the next, the total number of conformations dropped dramatically (~20%) between the AM1 and HF/3-21G(*) levels, but only slightly between the HF/3-21G(*) and HF/6-31G* (~4%) levels. This indicates that the AM1 method significantly overestimates the number of conformations presumably because the method's inherent inflexibility leads to conformational barriers that do not really exist on the potential energy surface. For our strategy this is clearly beneficial because we depend on the AM1 method to identify all of the conformations that will be subjected to higher level calculations.

To judge the ability of the methods to characterize energies, a sample set of conformers was chosen, and a plot of their relative energies vs the level of theory is given in Figure 5. For this purpose, the most stable representatives of schemes **A**−**E** were used.[22] From the plot, it is clear that the AM1 method is incapable of characterizing the relative energies of the serine conformers. First, AM1 grossly underestimates the stability of conformers with bonding scheme **D**. In addition, the AM1 calculations do not reflect the subtle preferences found within schemes **A**−**C**. This leads to a root-mean-square (rms) difference of ~2.6 kcal/mol between the relative energies at the AM1 and HF/3-21G(*) levels. Between the HF/3-21G(*) and HF/6-31G* levels, the changes are less significant (rms = 1.3 kcal/mol)—the major difference is that conformer **E** appears to be much less stable at the higher level. Nonetheless, these results suggest that with the exception of bonding scheme **E**, the small, split-valence basis set (3-21G(*)) can roughly approximate the values of the larger, polarized basis set (6-31G*). This is an encouraging result because the 3-21G(*) level is much more tractable for large systems and already has been employed by others for the study of dipeptides and amino acid derivatives.[23] However, neither of these levels incorporates corrections for electron correlation. In earlier work, Schäfer and others have pointed out that correlation corrections are needed to characterize

(22) Energies are shown relative to conformation type C. All comparisons are root-mean-square deviations of the energies relative to the mean.

Table 5. Relative Energies, Interactions, and Rotational Data for Cysteine Conformers[a]

	relative energies						rotational spectra			
	HF	HF+	MP2+	ZPE	bonding scheme[b]	NH_2 interaction[c]	A	B	C	dipole (D)
1	0.0	0.0	0.0	66.74	**D**		3.193	1.566	1.285	4.77
2	−1.4	−1.4	0.3	66.26	**C**	C=O	4.313	1.184	0.999	1.90
3	−0.8	−1.0	0.4	66.29	**F**	C=O	2.948	1.604	1.332	2.61
4	−0.2	−0.4	0.5	66.15	**C**	C=O	2.933	1.641	1.365	3.17
5	−0.5	−0.7	1.0	66.38	**B**	C(O)OH	3.249	1.553	1.261	2.48
6	1.5	1.2	1.3	66.65	**D**	CH_2SH	4.387	1.172	1.015	4.24
7	0.0	−0.4	1.4	66.28	**B**	C(O)OH	3.161	1.486	1.174	2.81
8	0.2	0.0	1.6	66.38	**F**	C(O)OH*	3.101	1.541	1.306	2.64
9	0.6	0.2	1.7	66.27	**F**	C=O	2.795	1.598	1.246	2.39
10	0.4	0.0	1.8	66.22	**F**	C=O	4.289	1.157	1.022	2.22
11	0.9	0.3	1.9	66.20	**F**	C(O)OH	2.845	1.577	1.262	2.01
12	0.1	0.0	1.9	66.19	**C**	C(O)OH	4.649	1.155	0.968	3.47
13	0.8	0.3	2.2	66.16	**F**	C=O	4.555	1.181	0.973	2.80
14	0.4	0.0	2.2	66.21	**F**	C(O)OH	4.793	1.144	0.973	2.92
15	1.4	1.3	2.3	66.15	**F**	C=O	3.044	1.496	1.361	2.54
16	0.7	0.4	2.3	66.28	**B**	C(O)OH	3.393	1.461	1.128	3.01
17	1.1	0.6	2.4	66.12	**F**	C=O	4.291	1.155	1.021	1.07
18	1.2	0.6	2.4	66.19	**F**		2.961	1.522	1.220	1.56
19	1.0	0.5	2.5	66.19	**F**	C(O)OH	4.385	1.125	1.043	1.59
20	1.7	1.7	2.5	66.42	**D**		4.652	1.186	0.975	3.97
21	2.2	2.2	2.5	66.61	**D**		3.553	1.438	1.086	5.16
22	1.7	1.5	2.5	66.18	**C**	C(O)OH	3.047	1.589	1.337	3.92
23	1.4	1.0	2.6	66.20	**F**	C=O	2.844	1.566	1.242	2.47
24	2.3	2.3	2.6	66.55	**D**		3.222	1.538	1.274	6.47
25	0.7	0.7	2.8	66.17	**F**	C=O	4.429	1.185	1.000	2.70
26	1.6	1.0	3.0	66.11	**F**	C(O)OH	4.399	1.124	1.040	2.98
27	2.1	1.6	3.1	66.08	**F**	C(O)OH	2.858	1.571	1.254	3.18
28	2.1	1.7	3.1	66.23	**A**	C=O	3.484	1.455	1.078	0.60
29	1.7	1.5	3.2	66.20	**F**	C=O	2.850	1.627	1.229	0.62
30	2.3	1.8	3.2	66.02	**F**	C(O)OH	2.923	1.528	1.214	3.21
31	1.6	1.3	3.3	66.09	**F**	C=O	4.282	1.166	1.029	2.33
32	2.3	1.9	3.6	66.12	**F**	C(O)OH	2.856	1.576	1.258	3.11
33	3.9	3.7	3.7	66.48	**D**		2.936	1.576	1.410	6.34
34	2.2	1.8	4.1	65.99	**F**	C(O)OH	4.401	1.129	1.052	3.25
35	1.8	1.8	4.1	66.13	**F**	C(O)OH	4.774	1.159	0.972	2.26
36	5.7	5.5	5.5	66.33	**D**		2.979	1.566	1.214	6.14
37	5.5	5.3	5.9	66.26	**E**	C=O	3.274	1.514	1.111	3.72
38	5.6	5.4	6.1	66.22	**E**	C=O	3.023	1.594	1.173	2.30
39	5.8	5.7	6.7	66.12	**H**	C=O	4.229	1.184	1.007	2.24
40	7.2	7.0	7.6	66.07	**G**	C=O*	2.917	1.564	1.365	3.93
41	8.0	8.0	9.4	66.01	**G**	C=O	4.329	1.181	1.012	3.05
42	9.8	9.1	10.0	65.92	**H**		4.385	1.090	1.052	3.56

[a] Geometries optimized at the HF/6-31G* level and energies at the HF/6-31G* (HF), HF/6-31+G* (HF+), and MP2/6-31+G* (MP2+) levels (kcal/mol). Zero-Point energies (ZPE) in kcal/mol scaled by 0.9. Rotational constants in GHZ. Dipole from HF/6-31G* wave function. [b] Bonding schemes as shown in Figure 6. F stands for syn carboxyl, but no interaction for the hydrogen of the CH_2SH group. G indicates no interactions for the hydrogens of either the carboxyl or the CH_2SH group. H indicates that the hydrogen of CH_2SH interacts with the NH_2 group, but that there is no interaction for the carboxyl proton. [c] Group that has near-atom interaction with a hydrogen of the NH_2 group. * indicates that both hydrogens interact with the group.

the interactions found in amino acids.[1-3,11k] To test for correlation effects, the most stable conformer from each bonding scheme (**A**−**E**) was subjected to optimization at the MP2/6-31+G* level, and the relative energies are shown in Figure 5. Fortunately, the correlation effects are relatively minor (rms = 0.8 kcal/mol)—with correlation, the relative stability of conformers **E** and **D** increase. However, the HF/6-31G* level gives a reasonable representation of the potential energy surface (rms = 0.6 kcal/mol for the entire data set vs that for MP2/6-31+G*//HF/6-31G*). In each of these conformers, correlation has a modest effect on the geometry of the amino acid. For this reason, Schäfer had warned against completing MP2 calculations on geometries optimized at the Hartree−Fock level.[11k] There are some subtle differences in our small data set (Figure 5), but overall this approach is only somewhat less accurate than optimizations at the MP2/6-31+G* level—compared to the fully optimized MP2/6-31+G*//MP2/6-31+G* level, the MP2/6-31+G*//HF/6-31G* level has a rms error of ~0.2 kcal/mol. As a further test of basis set dependence and correlation effects, the relative energy of conformers **1** and **2** was calculated at the MP4SDQ/6-31+G*//MP2/6-31+G* level.[24] The MP4 relative energy (0.3 kcal/mol) differs only slightly from that of the MP2/6-31+G* level (0.1 kcal/mol).

To judge the overall performance of the theoretical methods in this study, expectations must be defined. For systems with hydrogen bonding, the amino acids are surprisingly insensitive to changes in the basis set because the intramolecular nature of the interaction severely limits the basis set superposition error normally encountered in hydrogen-bonding situations. Moreover, the long-range nature of the interactions tends to limit the effects of correlation. This is consistent with Császár's[2b] work on glycine where only small differences (~0.1 kcal/mol) were seen between the MP2/6-311++G** and MP4/6-311++G** levels. Moreover, even at exceptionally high levels

(23) For example, see refs 11a−c,h.

(24) Many other comparisons were made including optimizations at the MP2/6-31G** and HF/6-31+G* levels. In all cases (including MP4SDTQ/6-31G**), polarization functions on hydrogen had almost no effect (<0.1 kcal/mol). Diffuse functions have a modest effect on the relative energies, but optimization at the HF/6-31+G* level has only a minor effect (~0.1−0.2 kcal/mol on the total energy).

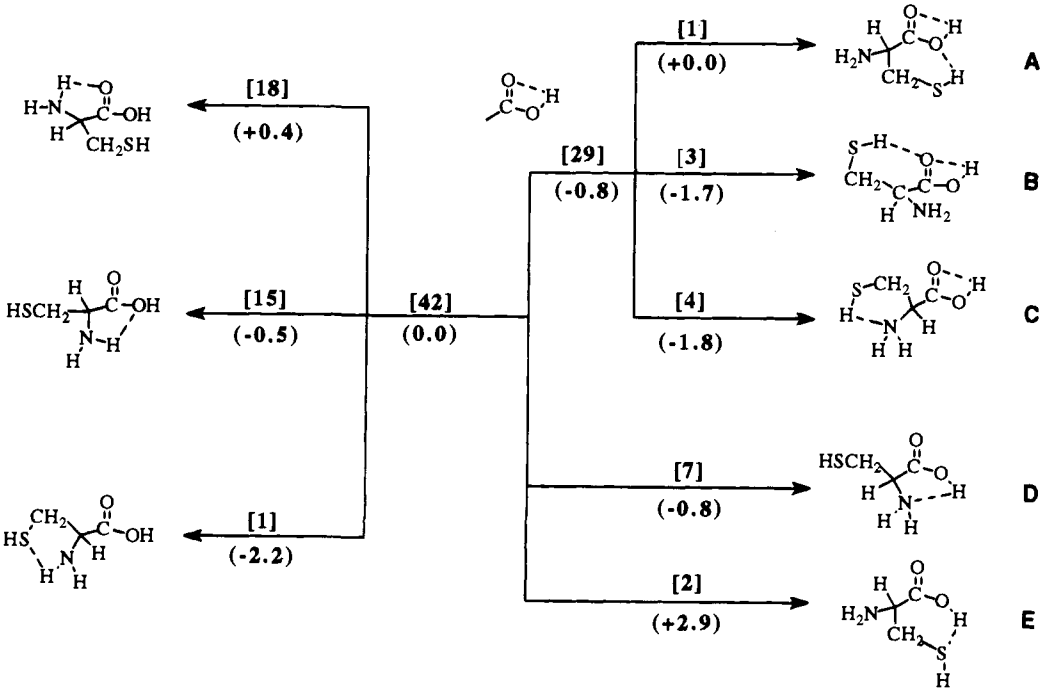


Figure 6. Bonding schemes for cysteine. Brackets contain the total number of conformers of each bonding scheme at the 6-31G* level. The average energy of the bonding scheme (relative to the overall average) is given parenthetically for the MP2/6-31+G*//HF/6-31G* level.

Table 6. Linear Regression Analysis of Near-Atom Interactions in Cysteine Conformers[a]

interactions[b]								
NH_2 C(O)—OH	NH_2 C=O	NH_2 CH_2SH	COOH NH_2	COOH CH_2SH	COOH C=O	CH_2SH C(O)—OH	CH_2SH C=O	CH_2SH NH_2
−1.4	−2.4	−1.8	−9.8	−4.1	−8.0	1.0	−1.5	−1.8
(0.5)	(0.4)	(1.2)	(0.5)	(0.9)	(0.5)	(1.2)	(0.6)	(0.5)
[15]	[18]	[1]	[7]	[2]	[29]	[1]	[4]	[6]

[a] MP2/6-31+G*//HF/6-31G*. Energies in kcal/mol. Standard errors given parenthetically. Number of occurrences given in brackets. [b] Hydrogen-bond donor listed above acceptor.

of theory, Császár's[2b] comparison of the two observed conformers of glycine (error = ~0.4 kcal/mol) was no closer to the experimental value than Schäfer's[11k] work at the MP2/6-311G** level. It appears that the majority of the correlation energy is captured with simple correlation corrections (MP2) and that only minor improvements are made with more extensive correlation treatments. Given the differences between the experimental and theoretical values for glycine, and the extent of convergence in the high-level work of Schaefer[2a] and Császár,[2b] it would be unreasonable to expect an accuracy of better than ±0.5 kcal/mol for our MP2/6-31+G*//MP2/6-31+G* values.[25] Using this level as a standard for comparison, calculations at the AM1, HF/3-21G(*), HF/6-31G*, and MP2/6-31+G*//HF/6-31G* levels give rms errors of 2.4, 0.8, 0.8, and 0.2 kcal/mol, respectively. This suggests that the overall accuracy of MP2/6-31+G*//HF/6-31G* and MP2/6-31+G*//MP2/6-31+G* levels is almost indistinguishable. The HF/6-31G* values are useful in many respects, but can have significant errors (~1.5 kcal/mol) in certain conformations. The HF/3-21G level gives a crude description of the relative energies and is susceptible to large errors in certain conformations. Finally, the AM1 values are very unpredictable in these systems.

From Table 3 it can be seen that the presence of three hydrogen-bond donors and four hydrogen-bond acceptors in serine allows for a wide range of hydrogen-bonding combinations and consequently a large number of stable conformations. It also is apparent that serine can find stabilization via a variety of different hydrogen-bonding schemes. For example, the eight most stable conformers differ in energy by less than 2 kcal/mol and contain representatives of bonding schemes **A**, **B**, **C**, and **D**. The first 30 conformers span less than 4 kcal/mol and contain all of the characteristic bonding schemes. The most stable conformer incorporates bonding scheme **C** as well as an interaction between an amino hydrogen and the carbonyl. The next most stable conformer, **2**, is also stabilized by a combination of strong hydrogen-bonding interactions, (COOH ⇔ NH_2) and (CH_2OH ⇔ O=C). As weaker interactions are incorporated, the stability naturally drops. For example, conformers with bonding scheme **E** are relatively unstable because they are built from the relatively weak (COOH ⇔ $O(H)CH_2$) interaction. Moreover, this scheme severely limits the ability of the CH_2OH group to act as a hydrogen-bond donor. The relative energies in Table 3 suggest that the equilibrium mixture should contain significant contributions from several conformers, with **1**, **2**, and **3** playing the most important role. Microwave spectroscopy should be able to distinguish between these conformers because there are differences in the rotational constants and dipole moments. In addition, there are significant differences in many parts of the IR spectra of **1** and **2**,

(25) Without experimental data, it is impossible to accurately assign expected errors in the computational results. The discussion in this paragraph suggests that our MP2 relative energies are probably close (±0.2−0.5 kcal/mol) to those that would be obtained with higher levels of theory (*i.e.*, MP4). Assuming that the experimental result is correct, Schaefer's and Császár's results with glycine indicate that high-levels of ab initio theory give relative energies with an uncertainty of ~0.4 kcal/mol.

particularly in the O−H stretches of the carboxyl (Table S5, supplementary material).[26]

Our results can be compared to those of Schäfer *et al.*[6] In an exceptionally efficient partial survey of the potential energy surface at the 4-21G level, they located 14 conformations including 8 of the 10 most stable structures found in this study. Labeling Schäfer's conformations in order of stability (roman numerals), conformations **I**−**X** correspond to conformations **1**, **2**, **4**, **3**, **5**, **6**, **17**, **9**, **7**, **24** in this investigation. The average deviation (relative to the mean) in the relative energies (excluding the ZPE corrections of the present work) is 0.6 kcal/mol compared to those of the HF/6-31G* calculations and 0.7 kcal/mol compared to those of the MP2/6-31+G*//HF/6-31G* calculations. The most obvious difference involves conformation **24** where the 4-21G level predicts a significantly greater relative stability. As noted above, small basis sets overestimate the stability of conformers with a hydrogen bond between the acidic proton and the side-chain hydroxyl (scheme **E**). In our own work at the 3-21G(*) level,[12] we identified conformer **2** as the global minimum.

Cysteine. In cysteine, the side-chain thiol group is highly polarizable but is a relatively poor hydrogen-bond donor and acceptor. As in serine, the combination of all single-bond rotamers leads to 324 trial structures. At the AM1 level, 58 conformers were located on the potential energy surface, but only 42 conformers remained at the highest level (Table 5 and Figure 6). The energies of these conformers vary by ~10 kcal/mol. In 29 of the conformers, the carboxyl hydrogen interacts with the carbonyl, and in 7 cases it interacts with the amino nitrogen (**D**). In only two conformers is there an interaction between the acidic proton and the thiol group (**E**). As noted for serine, the side-chain group (thiol) interacts most strongly as a hydrogen-bond donor with the carbonyl (**B**) and the amino nitrogen (**C**). Given the thiol's poor hydrogen-bonding characteristics, it is not surprising that bonding scheme **E** is much less stable than the others. This effect is also apparent in the interaction with the amino hydrogens. An interaction between the N−H bonds and the thiol sulfur is indicated in only one conformation.[27] This conformer is more stable than the average, but that is probably the result of other bonding interactions. Comparing the interaction energies (Tables 4 and 6) for serine and cysteine, it is clear that the thiol group provides less stabilization than the hydroxy group. For four out of the five interactions involving the side-chain substituent, a significantly larger stabilization energy (~2 kcal/mol) is found for serine. In the one exception (N−H ⟺ X(H)CH_2), the data set is not statistically meaningful for cysteine (one conformer). Because the interactions in cysteine are weaker, the barriers between conformers can be smaller and in some cases disappear. As a result, cysteine exhibits fewer unique conformations than serine.

The most stable representatives of bonding schemes **A**−**E** were further optimized at the MP2/6-31+G* level, and the structures are shown in Figure 7. A plot of relative energy vs level of theory is shown in Figure 8. Again, the AM1 level provides a poor description in comparison to the ab initio levels, and the HF/3-21G(*) level has difficulty with some bonding schemes. As was observed with serine, the HF/6-31G* level clearly underestimates the stability of conformers with bonding scheme **D** but gives reasonable values for the other schemes.

(26) Frequencies and intensities are shown for the most stable representatives of each of the characteristic bonding schemes. Conformers from the same bonding scheme have relatively similar frequencies and intensities.

(27) For cysteine, many of the interactions are weak and are probably not representative of true hydrogen bonds. The interaction simply indicates a near-atom relationship.

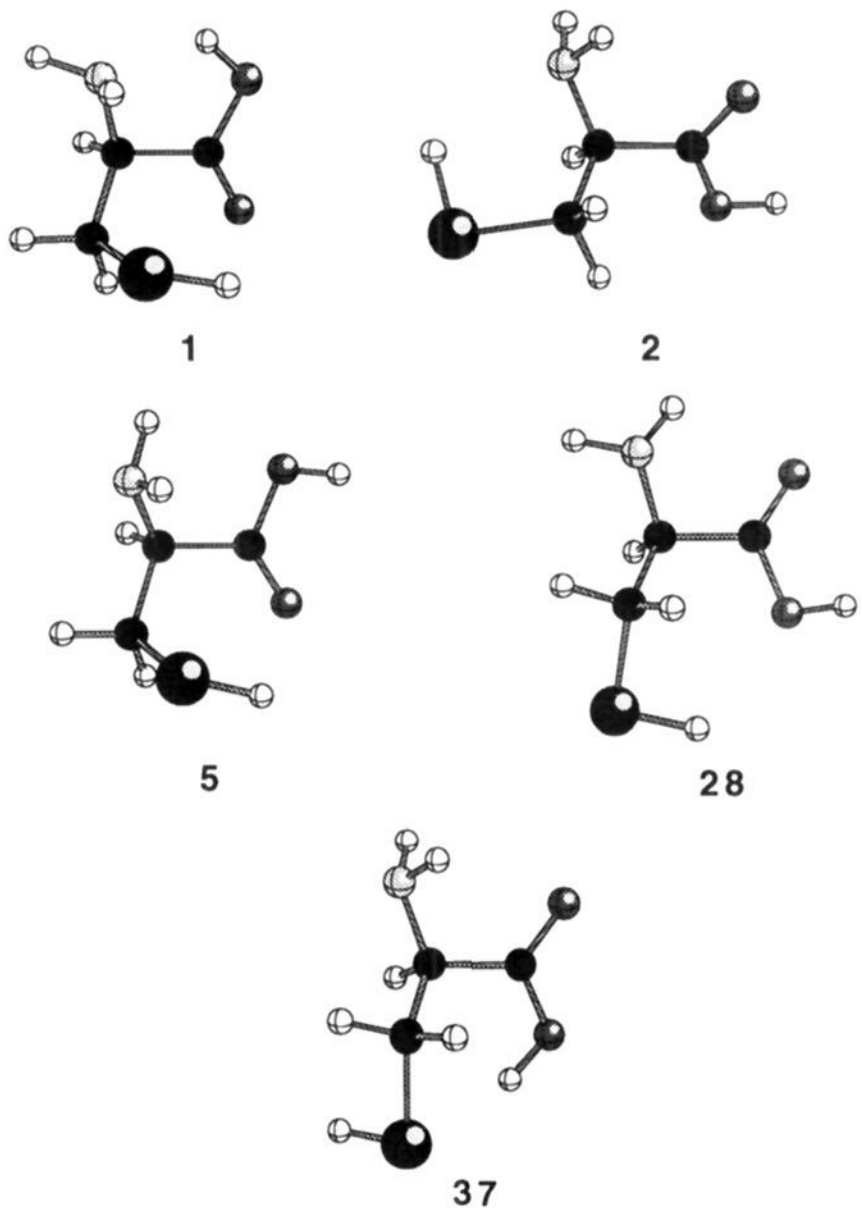


Figure 7. Sample conformers of cysteine optimized at the MP2/6-31+G* level. Each represents one of the fundamental bonding schemes (**A**−**E**) in Figure 6.

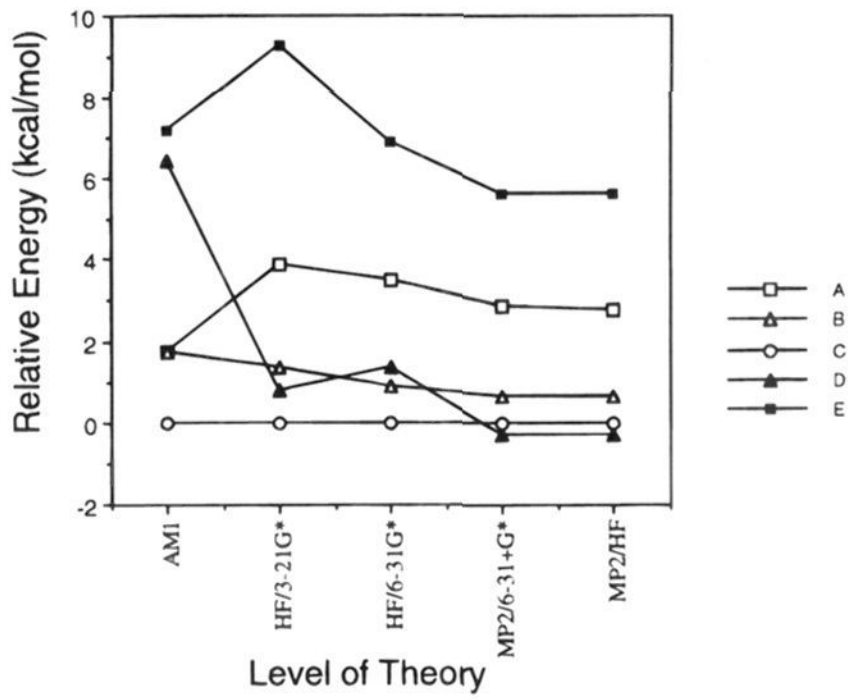


Figure 8. Plot of relative energy vs level of theory for cysteine conformers in Figure 7. MP2/HF refers to MP2/6-31+G*//HF/6-31G*.

Finally, it appears as if the MP2/6-31+G*//HF/6-31G* approach provides a good approximation of the true MP2 energies in cysteine.

In contrast to serine, the global minimum has a hydrogen-bonding interaction between the carboxyl proton and the NH_2 group; however, conformers with syn carboxyls (**2**−**4**) are nearly as stable. The equilibrium mixture should contain important contributions from several conformers (probably **1**−**4**). The rotational constants differ significantly through this series with the exception of **3** and **4** whose constants are separated by only

Table 7. Dihedral Angles of Alanine Conformers (6-31G*)

	I	II	III	IV
1	67.95	−168.52	0.62	1.32
2	−99.75	21.01	177.98	144.58
3	−85.80	38.00	−0.57	1.28
4	−150.53	−22.03	182.79	−146.16
5	83.03	−141.40	0.24	−110.35
6	44.35	166.46	0.24	101.81
7	183.67	−51.38	0.10	−4.06
8	−172.42	−51.05	−2.00	121.27
9	−76.96	49.93	1.96	−127.39
10	70.60	−166.46	−178.41	2.75

Table 8. Dihedral Angles of Serine Conformers (6-31G*)

	I	II	III	IV	V	VI
1	67.08	−170.66	0.43	17.57	−177.77	−47.45
2	−154.82	−28.47	−172.90	−143.74	61.91	−80.09
3	54.72	174.33	2.46	88.27	60.84	50.02
4	179.76	−56.88	−0.62	−9.83	62.94	−74.36
5	−153.51	−25.66	−175.26	−144.19	−65.85	61.45
6	54.40	178.58	1.63	−8.45	60.35	177.26
7	39.65	164.24	−1.95	−12.67	57.33	−87.17
8	−151.78	−26.88	0.75	8.72	−173.42	−46.69
9	−177.66	−53.46	0.69	2.25	−59.46	70.76
10	−104.96	15.49	−179.82	146.68	−179.88	−179.21
11	−124.87	−4.88	−1.81	99.59	64.50	48.64
12	−169.36	−48.59	−1.22	116.46	−65.49	70.30
13	−103.63	16.59	179.59	147.56	179.54	78.40
14	−153.96	−25.65	−176.36	−139.64	−177.85	176.03
15	−127.69	−2.98	−1.42	−6.71	65.46	−176.44
16	153.23	−82.30	−2.98	−144.19	67.87	−65.97
17	54.04	177.95	−0.48	3.51	−72.25	60.65
18	−55.80	68.28	0.82	−9.66	72.84	−65.49
19	−151.64	−23.62	−177.48	−139.52	−177.63	89.66
20	94.62	−144.70	3.71	13.00	−48.88	−64.36
21	−174.02	−53.41	−2.04	125.85	−176.73	−172.65
22	109.10	−130.62	3.06	14.29	−55.07	171.75
23	36.56	158.45	−0.06	110.82	−176.09	−170.07
24	45.78	166.88	176.58	80.26	−66.72	174.57
25	−60.96	65.31	3.37	−143.17	74.96	−63.77
26	79.77	−155.04	−0.12	−103.71	179.28	75.58
27	−69.60	53.38	−2.49	7.11	−46.73	−65.33
28	−166.76	−45.54	−1.19	107.99	176.03	−85.94
29	68.22	−166.30	−1.35	−108.36	−74.30	56.00
30	−57.62	65.21	−2.08	7.77	−52.75	173.76
31	−23.88	95.79	173.67	103.85	61.98	52.57
32	163.35	−71.01	−1.58	−138.28	−57.59	75.45
33	−173.13	−52.61	−2.32	128.34	−178.16	79.95
34	−79.09	47.81	2.56	−123.22	−176.24	−178.82
35	101.26	−134.74	1.86	−114.22	−54.88	−62.86
36	−82.87	43.91	2.23	−118.37	−177.14	77.72
37	119.67	−116.44	0.56	−121.98	−61.42	166.14
38	−113.94	13.84	−0.72	−124.29	65.63	−171.94
39	−44.01	80.92	−176.15	−5.00	64.83	176.61
40	−70.57	55.48	0.01	−127.85	−50.11	−59.70
41	−78.95	48.17	2.36	−128.01	−177.28	−85.34
42	−53.46	72.01	3.50	−135.76	−53.57	84.80
43	−61.70	63.81	0.53	−131.09	−56.76	172.30
44	62.08	−172.11	168.34	−94.97	−68.49	170.73
45	71.75	−166.68	−177.64	18.04	−175.91	−49.47
46	−53.50	67.05	−1.56	121.96	−55.29	−67.18
47	−56.37	70.40	−170.12	−143.02	68.94	−175.04
48	165.53	−72.32	174.81	−13.03	65.49	−69.42
49	57.31	−178.97	179.91	−6.72	57.66	170.49
50	70.31	−163.36	−175.03	−104.32	61.40	−177.76
51	56.28	176.79	−176.15	109.23	−177.03	76.78

~30 MHz. Again, IR spectroscopy (Table S6, supplementary material) could help identify conformers, especially those with bonding scheme **D** (*e.g.*, **1**).

In their work with cysteine, Schäfer *et al.*[8] reported seven fully optimized structures at the 4-21G level. Using their conformer designations, the correspondence is CYS1, **2**; CYS2, **28** CYS3, **3**; CYS1B3, **12**; CYS60, none; CYS180; **29**; CYS300,

Table 9. Dihedral Angles of Cysteine Conformers (6-31G*)

	I	II	III	IV	V	VI
1	−154.20	−24.00	−176.07	−144.31	68.14	−75.09
2	69.82	−167.32	0.52	14.81	−169.98	−55.61
3	47.37	174.37	−0.73	−5.37	62.99	−79.16
4	48.86	172.51	3.06	85.33	58.76	56.88
5	−160.46	−32.58	0.08	−4.88	67.45	−79.68
6	−101.22	19.77	178.54	142.99	−175.84	67.97
7	164.47	−73.40	0.25	3.21	−66.52	77.63
8	−140.92	−13.06	−1.07	5.39	69.45	79.67
9	118.00	−121.79	2.93	9.89	−59.99	−67.10
10	86.04	−148.15	−0.26	−112.06	−177.54	70.07
11	−54.83	67.83	−2.29	4.47	−55.78	−66.13
12	−166.61	−41.41	0.81	3.88	−165.86	−54.58
13	45.07	166.93	0.25	103.59	−174.69	77.62
14	−173.58	−52.45	−1.99	121.48	−174.42	72.85
15	78.01	−153.42	−0.44	−114.67	70.69	−65.24
16	−178.90	−59.54	−1.38	118.43	−72.48	75.11
17	86.84	−147.03	−0.33	−118.06	−177.20	−81.14
18	144.00	−92.14	−1.84	−131.65	−64.53	78.42
19	−77.20	50.05	2.30	−125.16	−174.82	71.13
20	−150.70	−22.17	−176.77	−147.58	−175.71	166.81
21	−159.57	−32.16	−172.63	−139.90	−76.58	68.43
22	−139.84	−15.66	−2.25	102.25	63.77	59.13
23	122.71	−113.86	0.64	−122.46	−62.71	−66.33
24	−153.98	−23.22	−177.43	−141.90	69.50	95.24
25	67.05	−169.56	0.27	8.27	−177.40	152.48
26	−77.15	50.37	2.35	−129.85	−174.67	−80.33
27	−61.03	64.20	0.13	−130.52	−56.52	−61.02
28	30.26	150.86	−0.24	99.83	−83.27	72.25
29	102.70	−137.35	3.65	11.67	−63.37	177.91
30	−67.59	57.45	2.79	−129.56	−67.21	66.07
31	87.33	−147.00	−0.27	−112.57	−172.12	−162.73
32	−58.73	63.65	−2.13	5.26	−61.30	169.14
33	−108.71	15.09	−175.13	136.83	62.15	67.55
34	−75.27	51.93	2.48	−128.63	−171.44	−171.24
35	−169.72	−44.44	0.54	−0.22	−174.86	154.39
36	−80.76	37.48	169.68	132.29	−58.52	−55.72
37	64.85	−172.84	174.26	5.90	−75.08	−73.02
38	81.87	−157.04	174.88	11.82	−65.60	87.85
39	75.44	−162.58	−176.84	15.46	−168.60	−55.23
40	63.15	−171.39	−175.32	−1.84	64.30	−72.29
41	75.65	−162.03	−176.29	10.02	−175.94	153.94
42	−52.60	72.19	−175.35	7.53	−165.88	−54.32

25. The absence of a conformer corresponding to CYS60 probably is a result of the more flexible basis set used in this work. The agreement between the relative energies at the 4-21G and 6-31G* levels is better than in the case of serine (average deviation = 0.3 kcal/mol at the HF/6-31G* level and 0.3 kcal/mol at the MP2/6-31G*//HF/6-31G* level). This is probably the result of a smaller data set limited to relatively similar structures.

Conclusions

Because amino acids are capable of a variety of intramolecular hydrogen-bonding interactions, a large number of stable conformations are possible. To identify them, one must consider all of the possible combinations of single-bond rotamers. In the case of a simple amino acid such as alanine, this results in 36 trial structures, but for serine and cysteine, 324 trial structures are required. AM1 calculations can identify possible conformers but do not adequately characterize the relative energies of these conformers. Improvements are observed at the Hartree−Fock level with 3-21G(*) and 6-31G* basis sets. To accurately assess the relative energies of the conformers, correlation corrections are needed, but calculations at the MP2/6-31+G*//HF/6-31G* level seem adequate for this purpose.

For alanine, 10 stable conformers are identified on the potential energy surface whereas 51 and 42 conformers were located for the more complicated amino acids, serine and

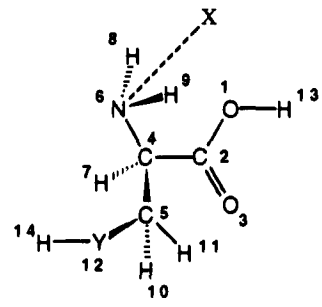
cysteine, respectively. In each system, the conformer vary in energy by ~10 kcal/mol and several conformers are within 2 kcal/mol of the global minimum. For each amino acid, the most stable conformer contains multiple hydrogen-bonding interactions. The majority of the conformers have a syn carboxyl, but structures where the carboxyl proton interacts with the amino nitrogen are nearly as stable and represent the global minimum for cysteine. Interactions between the carboxyl proton and the side-chain substituent (−OH or −SH) are less favorable. As expected, the results suggest that the sulfur in cysteine is a weaker hydrogen-bond donor and acceptor than the oxygen in serine.

It is hoped that the relative energies, rotational constants, and vibrational frequencies reported in this paper will encourage experimental studies of gaseous serine and cysteine and aid in the assignment of their microwave and IR spectra. Work on other amino acids is underway and will be reported in subsequent publications.

Acknowledgment. S.G. wishes to thank the National Science Foundation for their generous financial support (CHE-9208895). In addition, the authors would like to thank Professor Schäfer for providing reprints of his work.

Appendix A

The conformers of alanine, serine, and cysteine can also be described by a series of dihedral angles. The following designations are used in the Tables 7−9: **I** = $\angle C_5-C_4-C_2-O_1$; **II** = $\angle N_6-C_4-C_2-O_1$; **III** = $\angle H_{13}-O_1-C_2-O_3$; **IV** = $\angle X-N_6-C_4-C_2$; **V** = $\angle Y_{12}-C_5-C_4-C_2$; **VI** = $\angle H_{14}-Y_{12}-C_5-C_4$. The X group bisects the angle $\angle H_8-N_6-H_9$.



Supplementary Material Available: A listing of absolute energies and Cartesian coordinates (HF/6-31G*) for all of the alanine, serine, and cysteine conformers as well as frequencies for selected conformers (47 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, can be ordered from the ACS, and can be downloaded from the Internet; see any current masthead page for ordering information and Internet access instructions.

JA941428Z