Journal of Physical & Theoretical Chemistry
Islamic Azad University of Iran 3 (2)
(2006)
Science and Research Campus
ISSN: 1735-2126

Quantum Chemistry Study & Evaluation of Basis Set Effects on Prediction of Amino Acids Properties:

M.Monajjemi[1*], M. Karimkhani[1, 3], M. R. Gholami[2], A.Ziglari[3], K.Zare[1,4] and S.Afsharnezhad[5]

1. Department of Chemistry, Science and Research Campus, Islamic Azad University, P.O. Box: 14515-775, Tehran, Iran.
2. Department of Chemistry, Sharif University and Technology, Tehran, Iran.
3. Department of Chemistry, Tehran Central Branch, Islamic Azad University, Tehran, Iran.
4. Department of Chemistry, Shahid Beheshti University Evin,Tehran,Iran
5. Department of Biochemistry, Mashhad Azad university of Medical science, Mashhad, Iran.

ABSTRACT

The potential energy surface of gaseous glycine determined years ago in the ab initio B3LYP/6-311++G** calculations is composed of thirteen stable conformers. We performed the ab initio molecular orbital calculations as the starting point to carry out a force field and normal coordinate calculation on the most stable conformer of non-zwitterionic glycine [conformer (I)]. The calculations were carried out at different levels of theory using two methods, namely, the Hartree-Fock (HF) and the Möller-Plesset second order perturbation (MP2) method (including electron correlation), and using the Pople's basis sets, namely, STO-nG (n=2, 3 and 6), 3-21G, 6-21G, 6-31G, 6-311G and also cc-pVnZ to obtain HF limit. This different basis sets accompanied with the different combinations of diffuse and polarization functions were used. Each level of theory, with no symmetry restrictions, did fully optimization of neutral glycine. The atomic charge distributions were obtained using the Mulliken population analysis. The structural characteristics such as the total energies, the complete optimized geometrical parameters including bond lengths, normal and torsion angles, as well as dipole moments, rotational constants, atomic charge distributions, vibrational frequencies and IR intensities of the equilibrium conformation of glycine in gas phase were calculated at a wide range of the levels of theory -as mentioned above- and the results were compared together and with HF limit and the experimental data to examine the reliability of the applied basis sets and to introduce the most efficient ones. We also assayed how the strength of internal H-bonds depended on the variant parameters of basis set via the calculated atomic charges.

* . corresponding author

Keywords: Glycine; Basis sets; Ab initio calculations; Geometry parameters; Atomic charges; IR spectrum; HF Limit .

INTRODUCTION

Amino acids are remarkable subjects for computational chemists because of their diversity of intramolecular interactions and conformationally molecule flexibility. Also they create considerable interest for the understanding of the chemistry of peptides and proteins [1]. The simplest amino acid, Glycine, as one of the most important biological compounds has been the most widely studied of the amino acids experimentally and theoretically. This has been confirmed by experimental studies that in the gas phase glycine exists in non-ionized form, NH_2-CH_2-COOH [2]. Internal rotation about the C-C, C-N, and C-O bonds results in several glycine conformers . During the past three decades, the conformational behavior of glycine especially in gas phase has been the subject of various experimental [2-9] and theoretical [10-22] studies. Crystallized glycine has been explored using X-ray diffraction since 1939 [3] as well as neutron diffraction [4] and spectroscopic techniques [5]. The determination of the spectral and structural characteristics of the conformers of glycine, as well as other natural amino acids, is of great interest because of their relation to the amino acid units in peptides and existence of gas phase glycine in interstellar spaces[23]. The molecular structure of the gaseous glycine were determined by Iijima et al in an electron diffraction study. Conformer **I** (figure 1) with having bifurcated NH2···O=C H-bond was proven to be the most stable form in the gas phase.

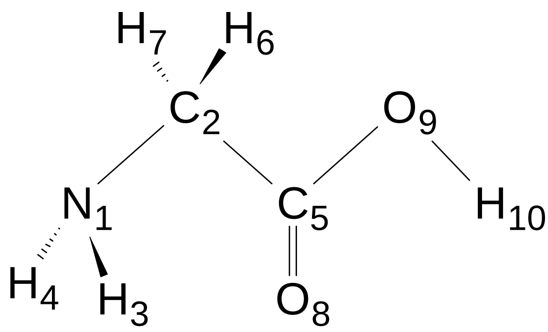


Figure 1. The most stable conformation of glycine (Conformer **I**).

The conformational behavior of glycine has also been the subject of very extensive theoretical studies. Császár predicted the existence of 13 stable conformers using high-level correlated ab initio calculations [15]. This calculations have been consistent in predicting that the conformer **I** is the most stable form of glycine neutral molecule. This subject has been confirmed by other similar work [12-20]. However the stability order of this conformer depends on the level of theory and the basis set used in the calculations. Unfortunately, although it is used a chanceful basis set but the efficacy of basis set on calculations has stayed unknown.

The aim of the present study is try to recognition of the behaviors of structural characteristics of glycine respect to the changes of some variables existing in basis sets for the purpose of specifying the most adequate basis set in ab initio calculations to describe this simplest amino acid. Therefore, we determined various molecular properties including total energies, complete optimized geometrical parameters including bond lengths, normal and torsion angles, as well as dipole moments, rotational constants, atomic charge distributions, vibrational frequencies and IR intensities and of the neutral form of the most stable conformer of glycine in gas phase with quantum chemical calculations. For this purpose we applied a wide range of Pople's basis sets including STO-nG (n=2, 3 and 6), 3-21G, 6-21G, 6-31G, 6-311G augmented by polarization and diffuse functions and two different methods, i.e. the HF and MP2 as the electron uncorrelated and correlated method, respectively. The results were compared together and with the experimental data to understand how these properties depend on the basis sets applied in the calculations and consequently prescribe a suitable basis set for this purpose. In our forthcoming works we continue our studies about other amino acids to survey the dependence of their properties on specific basis sets.

Computational Details

The theoretical results presented in this work were obtained by means of the ab initio molecular orbital calculations as the starting point to carry out a force field and normal coordinate calculation for non-zwitterionic glycine.

The calculations were carried out at the different levels of theory using the methods, namely, the Hartree-Fock (HF) [24] as an electron uncorrelated method, the Möller-Plesset second order many body perturbation method (MP2) [25,26] as a method containing electron correlation and inconsiderably the Becke's three-parameter hybrid functional combined with gradient corrected functional of Lee, Yang and Parr (B3LYP) [27]. The computations also have been performed using the different features of double (DZ) and triple zeta (TZ) qualities of Pople's basis sets , namely, STO-nG (n = 2, 3 and 6) [28], 3-21G [29,30], 6-21G [30], 6-31G [31] and 6-311G [32]. The mentioned basis sets have been chosen based on the difference between the number of primitives in minimal ones, splitting in valence layer and the number of primitives in core and valence layer. The Dunning's correlation consistent basis sets (cc-pVnZ) [33] have also been applied to determine the HF limit of basis sets. This various basis sets were used with different combinations of diffuse [34] and polarization functions [35], as we presented in Table 1. Fully geometry optimization of structure I (the most stable conformer of neutral glycine) was performed using analytical energy gradients by each level of theory, with no symmetry restrictions. RMS of forces and distances for all calculations didn't exceed $9.5*10^{-5}$ Hartree/Bohr and $3*10^{-9}$ Å , respectively.

The atomic charge distributions were obtained using the Mulliken population analysis [36]. The IR spectral characteristics of this structure (I) were calculated by all mentioned above basis sets and two HF and MP2 methods. All calculations were carried out employing the program package GAUSSIAN98 [37].

Results and Discussion

The calculations were done with the HF and MP2 methods - where the former is electron uncorrelated, while the latter one is containing correlation effects - using the various basis sets including the STO-nG series (n=2,3 and 6) and the derivatives of Pople's double and triple zeta basis sets including 3-21G, 6-21G, 6-31G, and 6-311G. They were chosen based on the difference between the number of primitives in minimal ones, splitting in valence layer and the number of primitives in the core and valence layers. They were augmented with the different combinations of diffuse and polarization functions, as we listed in Table 1. The fully geometry optimization of the conformer I of glycine (figure 1) was performed using the analytical energy gradients by each level of theory, with no symmetry restrictions. In the following, first we introduce a simple abbreviation for above basis sets to simplify the diagram presentations, and then the results obtained by aforesaid calculations will be shown and discussed.

Basis sets presentation

Whereas the Pople’s basis sets have massive names for our presentational purposes, we had to use some abbreviations as introduced in table 1 . The nomination has been accomplished based on the form of splitting, the number of polarization and diffuse functions as similar as possible to basic name. We used the letter A for the minimal basis sets followed by a number showing the number of its primitives. The split-valence basis sets were categorized to the double (S’) and triple (S”) zeta.For more illumination, we applied the number of core primitives only for the 6-21G and 3-21G.

The Greek numbers were applied sequentially with the increase of polarization functions, and also □and □ were seated instead of the diffuse function for heavy atoms and hydrogen atoms, respectively.

HF Limit

The solution of the HF equations with an infinite basis set is defined as the HF limit. Actually carrying out such a calculation is almost never a practical possibility. However, it is sometimes the case that one may extrapolate to the HF limit with a fair degree of confidence. Of the basis sets, the cc-pVnZ and cc-pCVnZ were designed expressly for this purpose. As they increase in size in a consistent fashion with each increment of n, one can imagine plotting some particular computed property as a function of n^{-1} and extrapolating the curve fit through those points back to the intercept; the intercept corresponds to n = ∞, i.e. the infinite basis limit[38]. We calculated the HF limit of properties as shown in Table 2 for the geometries and Figure 3 for some other properties by the mentioned method.

Table 1: Classification, presentation and notation of applied basis set.

Basis set Type	Specification	Core	Valence	Addition Function	Presentation	Symbol[a]	No. of BF	No. of PG
Minimal (A)	2 PG (‘)	-	-	-	STO-2G	A2	30	60
	3 PG (“)	-	-	-	STO-3G	A3	30	90
	6 PG (“)	-	-	-	STO-6G	A6	30	180
Split-Valence (S)	Double Zeta (‘)	3G	21	Simple	3-21G	S’3	55	90
				Polarization	3-21G*	S’3I	55	90
		6G	21	Simple	6-21G	S’62	55	105
				Polarization	6-21G*	S’62I	85	135
			31	Simple	6-31G	S’	55	130
				Polarization	6-31G*	S’I	85	160
					6-31G**	S’II	100	175
				Diffused	6-31+G	S’α	75	150
					6-31++G	S’β	80	155
				Combined	6-31+G*	S’αI	105	180
					6-31++G*	S’βI	110	185
					6-31+G**	S’αII	120	195
					6-31++G**	S’βII	125	200
					6-31++G(2df,pd)	S’βIII	220	310
					6-31++G(3df,3pd)	S’βIV	280	370
	Triple Zeta (“)	6G	311	Simple	6-311G	S”	80	155
				Polarization	6-311G*	S”I	105	185
					6-311G**	S”II	120	200
				Diffused	6-311+G	S”α	100	175
					6-311++G	S”β	105	180
				Combined	6-311+G*	S”αI	125	205
					6-311++G*	S”βI	130	210
					6-311+G**	S”αII	140	220
					6-311++G**	S”βII	145	225
					6-311++G(2df,pd)	S”βIII	230	335
					6-311++G(3df,3pd)	S”βIV	280	370

[a] Due to more simplification, the notation of 63 has omitted from the symbols of all 6-31G and 6-311G basis sets. The notation of α indicates diffuse function on heavy atoms (+), and β moreover shows diffuse function on hydrogen atoms (++).

Table 2: Extrapolated HF limit of geometry parameters.

Bond lengths	HF Limit	Bond Angles	HF Limit	Dihedrals	HF Limit
N1-C2	1.4337	C2-N1-H3	111.370	H3-C2-N1-H4	-
N1-H3	0.9965	C2-N1-H4	111.370	H3-N1-C2-C5	- 59.636
N1-H4	0.9965	N1-C2-C5	115.652	H6-C2-N1-C5	-
C2-C5	1.5179	N1-C2-H6	110.180	H7-C2-N1-C5	-
C2-H6	1.0823	N1-C2-H7	110.180	O8=C5-C2-N1	-
C2-H7	1.0823	C2-C5=O8	125.620	O9-C5-C2-O8	-
C5=O8	1.1787	C2-C5-O9	111.640	H10-O9-C5-C2	-179.965
C5-O9	1.3254	C5-O9-H10	109.335		
O9-H10	0.9439				

As we can observe from Figure 2, the cc-pV6Z calculation for glycine due to great number of its basis functions (about 1200 equals to more than 2200 primitives) take a time more than 220000 minute (about 150 days); then involving cc-pV6Z in extrapolation of the HF limit of properties is impractical.

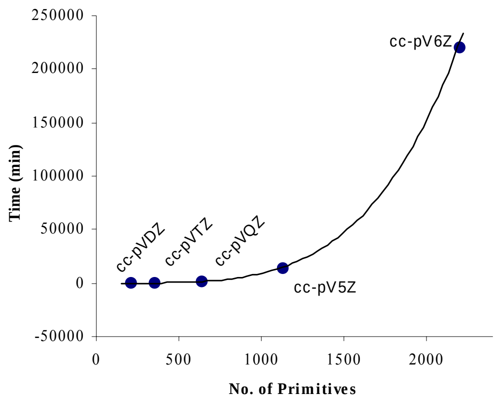
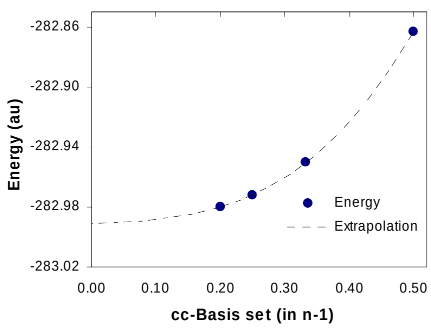
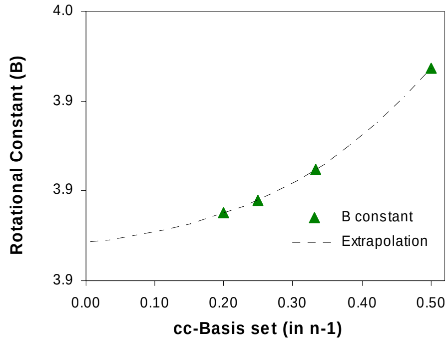


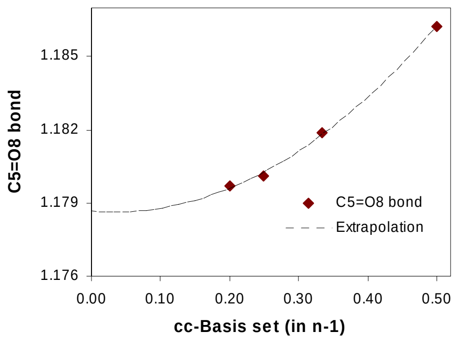
Figure 2. Time of calculation versus number of primitive gaussians in basis set.



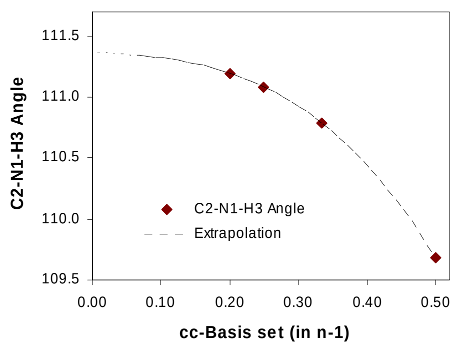
(a)



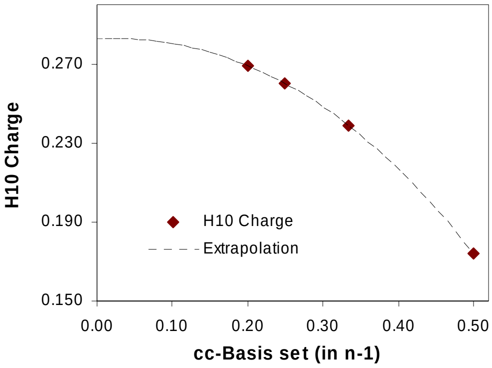
(b)



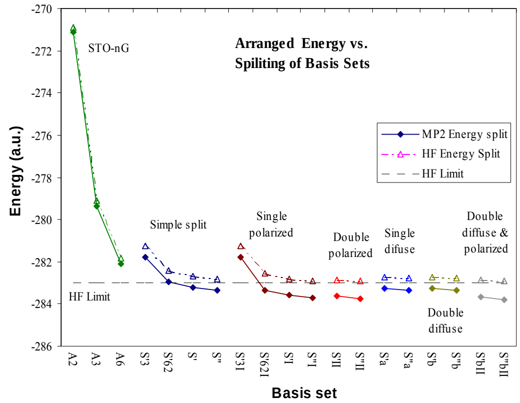
(c)



(d)



(e)

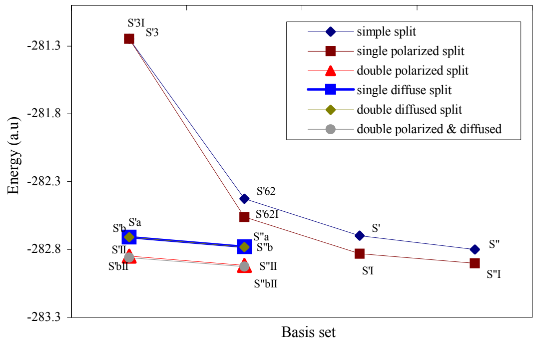


(a)

Energy, ZPE and Dipole moment

The total energy, zero point energy and dipole moment data were calculated by different levels of theory. To better deduction and realize the effects of splitting, the increment of polarization and diffuse functions to basis sets and also the effect of methods, we survey the results in following four categories:

Splitting Effect: Figures 3 to 5 respectively show the changes of total energy, dipole moment and zero point energy calculated by the Hartree-Fock and MP2 methods versus different basis sets and discretion for similar splitting in the basis sets. The results has been compared with HF limit of energy as obtained in 3.2 section. As seen from figure 2, the increment in the number of primitive gaussians From A2 to A6 basis set or the increase of splitting in valence layer in the basis sets cause a continuous decrease in the energy level of system. As one can see from junctions between S'3-S'62 and S'3I-S'62I (clearly in Figure 2), the number of primitives in core layer impress extremely on energy, while the effect of increment in splitting of valence layer and the number of primitives in each splitting valence layer continuously diminish, though generally trails decrease of total energy. In comparison of the lines corresponding with and "diffused splits" lines, we observe that the effect of diffuse functions on system energy is really very teeny and negligible.



(b)

Figure 2. Total energy calculated by HF and MP2 versus basis set compared with HF limit for conformer (I) glycine, ranking on splitting. a) complete comparison. b) comparison between splitted basis sets except STO-nG for HF method.

In Hartree-Fock level for all lines except "single split" line, the falling slope of energy in split increment from double to triple zeta with nearly maximum %6.8 difference show a steady procedure, wherein all lines are truly parallel. The exception is the "single split" line which in absence of each diffuse or polarization functions is more sensible to more splits and have higher exceptional slope. The falling slope of energy in MP2 level is more respect to HF level, which it bodes more sensibility of this method respect to splitting of basis set.

Figure 3 shows dipole moment changes versus different basis sets for similar splitting in them. As we see, increasing in the number of primitives in STO-3G cause an improvement in their predicting of dipole moments. “Simple split” basis sets improve widely the prediction procedure, as there is good agreement to experiment. Also the triple zeta basis sets try to improve the results more than double zeta basis sets. However the increment of polarization functions to basis set for heavy atoms change the results severely, so that the quality of basis sets in predicting dipole moment come down to limit of the smallest STO-nG. Addition of polarization functions for all atoms improves partially the prediction. Addition of diffuse functions on heavy atom basis sets always increases the value of dipole moment. This matter is obvious with comparison of the couples of S’, S” to S’α, S”α and S’II, S”II to S’βII, S”βII. Increase of diffuse function to basis set of hydrogen atoms doesn’t create tangible changes. At last, the comparison between the results obtained by MP2 and HF methods shows the former achieves partly more improvement in dipole moment prediction.

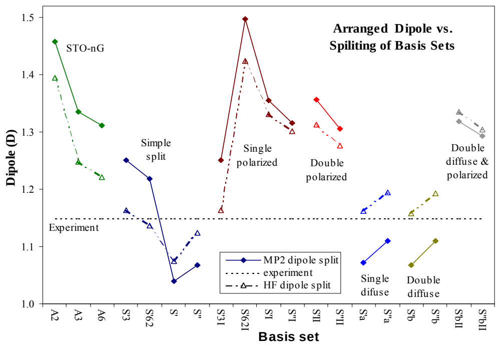


Figure 3. Dipole moment versus basis set for conformer (I) glycine, ranking on splitting.

Figure 4 shows zero point energy (ZPE) changes versus different basis sets for similar splitting in them. As we can see, STO-nG basis sets with a difference more than 4 kcal/mol, which is greatly faraway from splitted basis sets, aren’t able to give correct results for ZPE However a relative improvement in achievements observe when the number of primitives increase.

As shown by “simple split” and “ single polarized” paths in figure 4, Increment in the number of primitives in core and valence layers and also increment of splits on valence layer nearly always make ZPE values greater. The consequence of increment of polarization functions to basis sets is generally in the interest of heavy atoms and culminates in increase of results, but if done for hydrogen atoms will decrease the conclusions. Indeed, the addition of diffuse functions for heavy atoms generally decreases the results.

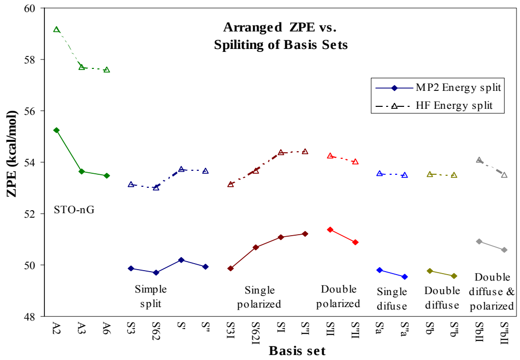


Figure 4. Zero point energy versus basis set for conformer (I) glycine, ranking on splitting.

3.3.2. Polarization Effect: Figures 5 to 7 show the changes of total energy, dipole moment and zero point energy, respectively, calculated by Hartree-Fock method versus different basis sets, but this time division has done based on the addition of polarization functions in the basis sets. As seen from figure 5, in general, the addition of polarization functions in basis sets continuously decrease the energy. It seems that the regular addition of polarization functions in different types of basis set (whether double or triple zeta or diffused basis set) follow a uniform procedure. It seems we can achieve to basis set limit in way of protracting the energy decrement path against the increment of polarization functions.

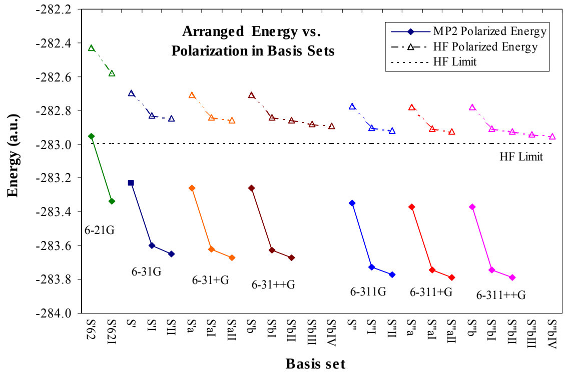


Figure 5. Total energy versus basis set for conformer (I) glycine, ranking on the number of added polarization functions.

As it shown in figure 6, the addition of the first polarization function to basis sets for heavy atoms generally overshoot the predicted value by basis set. This variation in triple zeta basis set is less than double one. However, if we add more polarization function, we encounter the gentle decline run which at last conduce to a definite limit value.

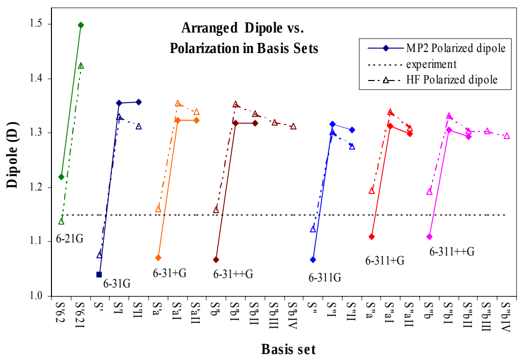


Figure 6. Dipole moment versus basis set, ranking on the number of added polarization functions.

Figure 7 considers zero point energy data. The comparison between figures 6 and 7 reveals that the behavior of ZPE changes against the increment of polarization functions in basis sets is similar to dipole moment. The only two discrepancies consist: first, contradictorily, the ZPE against the inclusion of the first polarization function in precipitate increment of each triple zeta basis sets is less than double zeta ones, and second, the increment of the more number of polarization functions to basis set cause a considerable decrease in ZPE value and at the end on HF limit comes to constant.

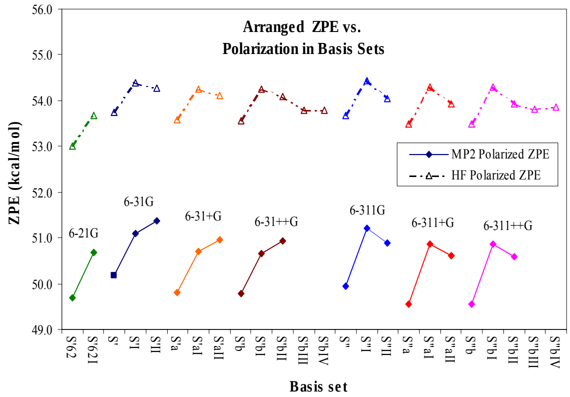


Figure 7. Zero point energy versus basis set, ranking on the number of added polarization functions.

3.3.3. Diffuse Effect: Figures 8 to 10 display the changes of total energy, zero point energy and dipole moment, respectively, calculated by Hartree-Fock method versus different basis sets, but this time division has done based on the addition of diffuse functions in the basis sets. As shown in figure 8, generally the addition of diffuse functions in various splitted basis sets for heavy atoms have the regular forms, which shows a suitable decrement corresponding about –0.01 a.u., while generalizing the increment of diffuse functions for hydrogen atoms only redound to a negligible decrement of energy corresponding about –0.0003 a.u.

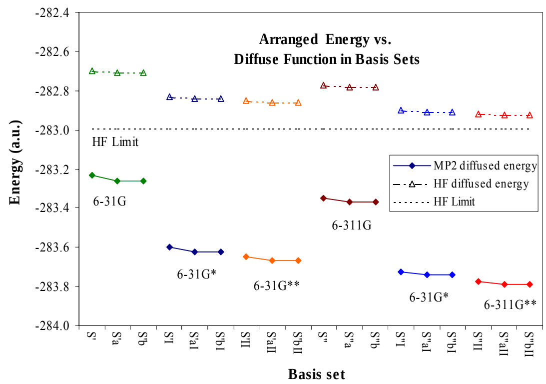


Figure 8. Total energy versus basis set for conformer (I) glycine, ranking on the number of added diffuse functions.

Remarkably this regular procedure exactly is followed for ZPE, as shown in figure 9, but

ZPE is more sensitive respect to the increment of diffuse functions for heavy atoms and shows a decrease between 0.12 – 0.18 kcal/mol, while similar to energy, the inclusion of diffuse function for hydrogen atoms leave out a very small effect only in order of several thousandth.

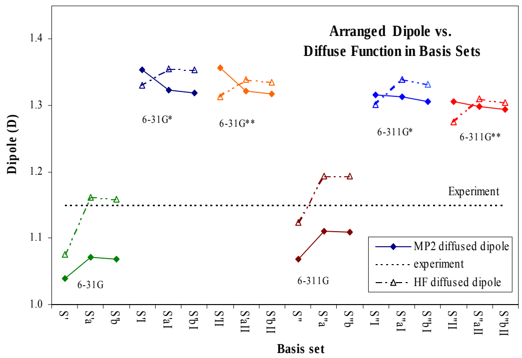


Figure 9. Zero point energy versus basis set, ranking on the number of added diffuse functions.

In contradictory to these two properties, the addition of diffuse functions for heavy atoms, as shown in figure 10, cause the increment of dipole moment in order of several hundredth to 0.1 debyes. Whereas the addition of diffuse function for hydrogen atoms accompanied by a much less effect always decrease the dipole moment about several thousandth. In any case, the addition of polarization and diffuse functions throw away the result far from experimental value, although in triple zeta basis sets this fact occurs gentler than double zeta ones.

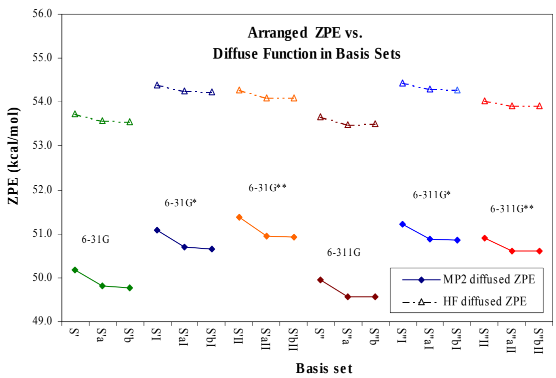


Figure 10. Dipole moment versus basis set, ranking on the number of added diffuse functions.

3.3.4. Method Effect: Notifying to figure 11 reveal that in general, Moller-Pelleset perturbation methods (in any order) due to considering correlation generally predict the lower energy, closer to the fact, although its impression is only about 0.9 a.u. equal to %0.3 of total energy. The B3LYP method offers the lowest energy between these methods.

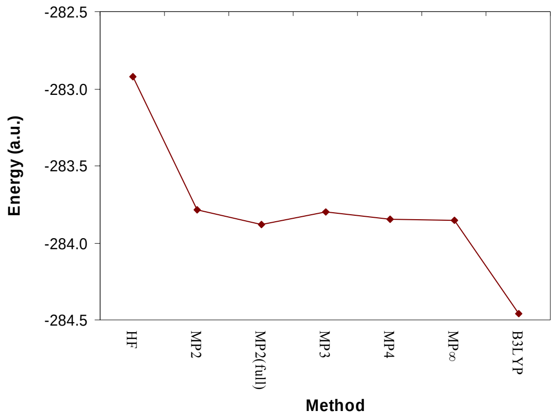


Figure 11. Comparison between total energies calculated by various methods and 6-311++G** basis set for conformer (I) of glycine. The MPn data were taken from ref. [15].

1.1. Geometry

3.4.1. Bond lengths. The theoretically optimized and experimental geometries including bond lengths, bond angles and rotational constants for the equilibrium conformation obtained for the non-zwitterionic glycine (figure 1) by microwave spectroscopy [9] are depicted in Table 3. The calculations have done by the HF method.

Concerning bond distances obtained by STO-nG basis sets, they have calculated the longest values for all of them but since this basis set is known to be less accurate than the other applied basis sets, we can consider that STO-nG basis sets overestimates the bond distances of glycine. As n increase from 2 to 3 and then 6 respectively, this overestimation improves and come nearer to the experiment. Almost in the most cases, the 3-21G series overestimate all the bond lengths except N1-C2 and C2-C5. The same observation can be made for the 6-21G series, but the exceptions spread to the underestimation of the C5=O8 and C5-O9 bonds. The 6-31G, 6-31+G and 6-31++G sets underestimate all the bond lengths except C2-H6 and C5=O8. The underestimation of C2-H6 is just generalized to the other derivatives of the 6-31G basis set. The same expressions can be repeated for all derivatives of the 6-311G basis set, unless 6-311G, 6-311+G and 6-311++G only lead C5=O8 to be shorter.

By comparing the values obtained with the double zeta basis sets together, one can deduce that the addition of polarization functions for heavy atoms make a relative improvement in the form of decreasing the overestimations and increasing the underestimations in the bond lengths. Anyway, further addition of polarization for hydrogen atoms not only do not improve the situation but also make it a bit more critical. As one can deduce by comparing 6-31++G(2df,pd) and 6-31++G(3df,3pd), The recuperative effect of more addition of polarization function to basis set is mostly unimportant, then the complexity of the calculations can not help so much to improve the bond lengths. For instance, with going from 6-31G to 6-31++G** and 6-31++G(2df,pd) and then 6-31++G(3df,3pd), the relative errors in N1-C2 decrease corresponding to 2.16, 2.03, 2.02, %1.96, and in O9-H10 increase corresponding to 1.19, 1.82, 2.03, %2.04, respectively.

In 6-311G triple zeta basis set and its derivatives, it is not observed any preferable absolute status respect to corresponding states in the 6-31G double zeta basis set, and even most of the time there is a few tendency in the interest of double zeta basis set. However the difference between their results is negligible and we didn't observe a meaningful discrepancy. Comparing the 6-311G derivatives together, it seems the inclusion of polarization function for heavy atoms affords a relative improvement in the bond lengths except for C5=O8, C5-O9 and O9-H10.

As like 6-31G, the addition of more polarization functions only generates insignificant changes in the bond lengths and most of the time doesn't afford any improvement.

Generally, in corresponding cases, the results of MP2 method have better conformity with the experiment respect to HF method.

3.4.2. Bond angles. The theoretically optimized and experimental bond angles consist of normal and torsion angles for equilibrium conformation glycine are summarized in Table 3. As we can see, the whole derivatives of 6-31G and 6-311G basis sets overestimate N1-C2-C5, C2-C5=O8 and C2-C5-O9, and underestimate C5-O9-H10 (except for 6-31G and 6-311G). The effect of the addition of polarization functions and diffuse functions on angle values is completely inversed together.

If the addition of polarization function increases the angle, the diffuse function has a gentle descending effect on it and at last the addition of more polarization functions enforce the angles to tend to experiment in a gentle run. The impression left on bond angles by the addition of the first polarization function, is very intensive only for C2-N1-H3 and C5-O9-H10 angles equal to about 5.5° for 6-31G derivatives and negligible for others. The more inclusion of polarization function increases these two angles maximum 0.5° or less.

According to just one dihedral angle we have from experiment, whole of basis sets predict successfully.

3.4.3. Rotational constants. Again, the theoretically optimized and experimental rotational constants for equilibrium conformation glycine have been showed in Table 3.

Table 3: Geometry Parameters, Rotational Constants, and Dipole Moments of conformer (I) of Glycine [a], calculated by different levels of theory.

Levels of Theory / Coordinates [a]	HF / Basis Set																	
	A2	A3	A6	S'3	S'3I	S'62	S'62I	S'	S'I	S'II	S'α	S'β	S'αI	S'βI	S'αII	S'βII	S'βIII	S'βIV
R									Bond Lengths									
N1-C2	1.4900	1.4818	1.4818	1.4533	1.4533	1.4545	1.4499	1.4343	1.4390	1.4372	1.4347	1.4347	1.4380	1.4382	1.4362	1.4362	1.4364	1.4373
N1-H3	1.0431	1.0334	1.0334	1.0036	1.0036	1.0035	1.0113	0.9947	1.0017	1.0002	0.9949	0.9949	1.0012	1.0014	0.9996	0.9996	0.9988	0.9991
C2-C5	1.5672	1.5519	1.5519	1.5090	1.5090	1.5108	1.5199	1.5074	1.5156	1.5149	1.5078	1.5078	1.5164	1.5163	1.5158	1.5158	1.5163	1.5161
C2-H6	1.0981	1.0916	1.0916	1.0822	1.0822	1.0821	1.0882	1.0828	1.0846	1.0853	1.0829	1.0829	1.0845	1.0845	1.0850	1.0850	1.0850	1.0847
C5=O8	1.2200	1.2166	1.2166	1.2020	1.2020	1.2046	1.1838	1.2104	1.1877	1.1879	1.2107	1.2107	1.1889	1.1889	1.1890	1.1890	1.1828	1.1834
C5-O9	1.3926	1.3891	1.3891	1.3565	1.3565	1.3608	1.3352	1.3529	1.3299	1.3287	1.3524	1.3524	1.3298	1.3299	1.3290	1.3290	1.3271	1.3275
O9-H10	0.9974	0.9902	0.9902	0.9687	0.9687	0.9693	0.9683	0.9545	0.9525	0.9481	0.9543	0.9543	0.9526	0.9526	0.9484	0.9484	0.9464	0.9463
∠A									Bond angles									
C2-N1-H3	105.99	107.23	107.23	113.33	113.33	113.23	108.32	116.12	110.41	110.55	116.58	116.58	111.20	111.14	111.50	111.50	111.12	110.86
N1-C2-C5	114.21	113.58	113.58	113.42	113.42	113.41	113.78	114.90	115.03	115.08	115.16	115.16	115.43	115.42	115.46	115.46	115.49	115.43
N1-C2-H6	109.13	109.36	109.36	110.24	110.24	110.20	110.10	110.24	109.98	110.15	110.09	110.09	109.88	109.89	110.04	110.04	110.16	110.20
C2-C5=O8	125.32	125.66	125.66	126.67	126.67	126.55	124.81	126.34	125.40	125.38	126.54	126.54	125.62	125.63	125.63	125.63	125.57	125.57
C2-C5-O9	111.31	111.90	111.90	110.54	110.54	110.63	111.53	111.54	111.80	111.85	111.35	111.35	111.67	111.65	111.68	111.68	111.57	111.61
C5-O9-H10	104.28	104.79	104.79	112.26	112.26	111.96	106.49	114.16	108.48	108.68	114.60	114.60	109.08	109.07	109.23	109.23	108.94	108.85
∠τ									Torsion Angles									
H3-C2-N1-H4	107.87	111.42	111.42	127.22	127.22	126.83	112.87	135.88	117.11	117.73	137.10	137.10	118.74	118.57	119.87	119.87	118.94	118.21
H3-N1-C2-C5	-53.94	-55.70	-55.70	-63.61	-63.61	-63.41	-56.43	-67.94	-58.56	-58.87	-68.55	-68.55	-59.37	-59.28	-59.94	-59.94	-59.47	-59.10
H7-C2-N1-C5	121.98	121.23	121.23	121.21	121.21	121.29	121.60	121.73	121.79	121.74	121.74	121.74	121.82	121.80	121.74	121.74	121.71	121.67
O8=C5-C2-N1	0.02	-0.02	-0.02	0.00	0.00	0.00	-0.01	0.00	-0.01	0.03	0.00	0.00	0.01	0.00	0.00	0.00	-0.01	0.01
O9-C5-C2-O8	179.95	180.00	180.00	180.00	180.00	-180.00	-179.99	-180.00	-179.97	179.94	-180.00	-180.00	179.98	180.00	180.00	-180.00	179.98	179.98
								Rotational constants										
A	9.952	10.094	10.094	10.335	10.335	10.303	10.593	10.401	10.645	10.654	10.387	10.387	10.630	10.628	10.636	10.636	10.686	10.682
B	3.713	3.754	3.754	3.911	3.911	3.900	3.955	3.872	3.930	3.933	3.863	3.863	3.914	3.914	3.916	3.916	3.923	3.924
C	2.790	2.825	2.825	2.932	2.932	2.923	2.976	2.916	2.965	2.968	2.909	2.909	2.955	2.955	2.957	2.957	2.964	2.964
								Dipole moment										
μ	1.395	1.248	1.248	1.163	1.163	1.137	1.424	1.075	1.331	1.313	1.162	1.158	1.356	1.354	1.340	1.335	1.320	1.313

Table 3: Continued.

Level of Theory / Coordinates [a]	HF / Basis Set														HF Limit	B3LYP / Basis Set		Experi-mental Value [c]
	S"	S"I	S"II	S"α	S"β	S"αI	S"βI	S"αII	S"βII	S"βIII	S"βIV	CD	CT	CQ		S'βII	CβD [b]	
R								**Bond Lengths**										
N1-C2	1.4379	1.4392	1.4389	1.4378	1.4378	1.4380	1.4380	1.4376	1.4376	1.4347	1.4346	1.4384	1.4363	1.4349	1.4337	1.4484	1.450	1.466
N1-H3	0.9941	0.9983	1.0000	0.9942	0.9942	0.9981	0.9981	0.9996	0.9996	0.9991	0.9975	1.0061	0.9979	0.9969	0.9965	1.0163	1.018	1.001
C2-C5	1.5030	1.5143	1.5139	1.5047	1.5047	1.5154	1.5154	1.5154	1.5154	1.5134	1.5143	1.5120	1.5134	1.5143	1.5179	1.5259	1.524	1.529
C2-H6	1.0803	1.0842	1.0852	1.0805	1.0805	1.0842	1.0842	1.0852	1.0852	1.0843	1.0831	1.0919	1.0835	1.0829	1.0823	1.0966	1.100	1.081
C5=O8	1.2079	1.1820	1.1820	1.2078	1.2078	1.1829	1.1829	1.1827	1.1827	1.1810	1.1805	1.1862	1.1819	1.1801	1.1787	1.2127	1.211	1.204
C5-O9	1.3504	1.3282	1.3285	1.3497	1.3497	1.3280	1.3280	1.3285	1.3285	1.3259	1.3255	1.3292	1.3274	1.3262	1.3254	1.3565	1.359	1.354
O9-H10	0.9497	0.9443	0.9457	0.9505	0.9505	0.9449	0.9449	0.9461	0.9461	0.9468	0.9448	0.9511	0.9455	0.9444	0.9439	0.9728	0.971	0.966
∠A								**Bond angles**										
C2-N1-H3	114.94	110.54	110.47	115.40	115.40	111.20	111.20	111.18	111.18	111.25	111.09	109.68	110.79	111.08	111.37	110.73	110.4	-
N1-C2-C5	115.05	115.25	115.20	115.15	115.15	115.53	115.53	115.54	115.54	115.53	115.52	115.07	115.43	115.54	115.65	115.87	115.9	113.0
N1-C2-H6	110.18	110.04	110.20	110.11	110.11	109.98	109.98	110.11	110.11	110.17	110.19	110.38	110.20	110.19	110.18	109.94	-	-
C2-C5=O8	126.40	125.61	125.48	126.33	126.33	125.75	125.75	125.67	125.67	125.57	125.60	125.20	125.55	125.60	125.62	125.70	125.8	125.1
C2-C5-O9	111.66	111.69	111.66	111.58	111.58	111.50	111.50	111.56	111.56	111.60	111.62	111.96	111.64	111.65	111.64	111.53	111.4	111.5
C5-O9-H10	113.86	109.22	108.66	114.29	114.29	109.74	109.74	109.12	109.12	109.12	108.91	108.13	108.79	109.01	109.33	107.44	107.2	110.5
∠τ								**Torsion angles**										
H3-C2-N1-H4	131.32	117.53	117.39	132.65	132.65	118.90	118.90	119.04	119.04	119.12	118.73	115.05	118.06	-	-	117.87	-	-
H3-N1-C2-C5	-65.66	-58.77	-58.70	-66.32	-66.32	-59.45	-59.45	-59.52	-59.52	-59.56	-59.37	-57.52	-59.03	-59.37	- 59.64	-58.94	-58.3	-
H7-C2-N1-C5	121.78	121.92	121.70	121.74	121.74	121.92	121.92	121.72	121.72	121.73	121.70	121.75	121.73	-	-	122.09	-	-
O8=C5-C2-N1	0.01	0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.02	-	0.02	0.0	0.00
O9-C5-C2-O8	179.98	179.98	- 179.97	- 179.98	- 179.98	- 179.98	- 179.98	179.97	179.97	180.00	180.00	180.00	179.99	-	- 179.97	179.95	-	-
							Rotational constants											
A	10.452	10.701	10.696	10.445	10.445	10.680	10.680	10.684	10.684	10.717	10.727	10.669	10.709	10.732	10.765	10.280	10.283	10.342
B	3.882	3.929	3.934	3.876	3.876	3.918	3.918	3.918	3.918	3.931	3.932	3.944	3.932	3.929	3.924	3.831	3.832	3.876
C	2.925	2.968	2.971	2.921	2.921	2.961	2.961	2.961	2.961	2.971	2.972	2.975	2.971	2.971	2.971	2.882	2.883	2.912
							Dipole moment											
μ	1.125	1.302	1.276	1.194	1.193	1.340	1.333	1.310	1.305	1.303	1.295	1.269	1.262	1.272	1.28	1.212	1.170	1.150

Table 3: Continued.

Level of Theory / Coordinate [a]	MP2 / Basis Set															
	A2	A3	A6	S'3	S'3I	S'62	S'62I	S'	S'I	S'II	S'α	S'β	S'αI	S'βI	S'αII	S'βII
R									**Bond Lengths**							
N1-C2	1.5192	1.5152	1.5112	1.4815	1.4815	1.4833	1.4598	1.4642	1.4521	1.4505	1.4628	1.4630	1.4502	1.4502	1.4483	1.4483
N1-H3	1.0669	1.0598	1.0558	1.0232	1.0232	1.0235	1.0271	1.0162	1.0189	1.0144	1.0159	1.0160	1.0183	1.0183	1.0137	1.0137
C2-C5	1.5919	1.5766	1.5759	1.5276	1.5276	1.5300	1.5191	1.5275	1.5169	1.5166	1.5275	1.5272	1.5181	1.5181	1.5182	1.5182
C2-H6	1.1147	1.1090	1.1052	1.0941	1.0941	1.0941	1.0960	1.0984	1.0952	1.0910	1.0988	1.0988	1.0955	1.0955	1.0913	1.0913
C5=O8	1.2557	1.2577	1.2558	1.2374	1.2374	1.2411	1.2134	1.2490	1.2189	1.2183	1.2491	1.2490	1.2214	1.2214	1.2208	1.2208
C5-O9	1.4149	1.4157	1.4116	1.3936	1.3936	1.3991	1.3575	1.4003	1.3593	1.3585	1.4032	1.4031	1.3621	1.3621	1.3615	1.3615
O9-H10	1.0232	1.0189	1.0168	0.9964	0.9964	0.9978	0.9936	0.9851	0.9796	0.9712	0.9867	0.9867	0.9812	0.9812	0.9728	0.9728
∠A									**Bond angles**							
C2-N1-H3	103.14	104.09	104.24	110.27	110.27	110.16	106.50	113.67	108.80	108.57	114.94	114.92	110.36	110.36	110.32	110.32
N1-C2-C5	114.00	113.60	113.64	113.18	113.18	113.16	113.34	114.89	114.88	114.87	115.41	115.39	115.57	115.57	115.60	115.60
N1-C2-H6	108.71	108.82	108.85	109.71	109.71	109.66	109.95	109.61	109.80	109.96	109.47	109.49	109.61	109.61	109.74	109.74
C2-C5=O8	125.82	126.11	126.29	126.59	126.59	126.45	124.64	126.43	125.45	125.38	126.92	126.91	125.89	125.89	125.87	125.87
C2-C5-O9	110.41	111.06	111.17	110.03	110.03	110.16	111.22	110.96	111.36	111.36	110.47	110.47	111.06	111.06	111.00	111.00
C5-O9-H10	101.58	101.76	101.61	108.10	108.10	107.73	103.83	110.04	105.80	105.79	110.81	110.77	106.76	106.76	106.74	106.74
∠τ									**Torsion Angles**							
H3-C2-N1-H4	102.77	105.31	105.64	119.45	119.45	119.14	109.37	128.01	113.86	113.34	132.00	131.90	117.35	117.35	117.50	117.50
H3-N1-C2-C5	-51.39	-52.66	-52.82	-59.77	-59.77	-59.58	-54.69	-64.01	-56.93	-56.67	-66.00	-65.95	-58.68	-58.68	-58.75	-58.75
H7-C2-N1-C5	122.15	121.42	121.36	121.18	121.18	121.25	121.66	121.82	121.95	121.80	121.79	121.71	122.01	122.01	121.83	121.83
O8=C5-C2-N1	-0.01	-0.01	0.00	0.01	0.01	-0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01
O9-C5-C2-O8	180.00	-179.98	180.00	-179.99	-179.99	180.00	-180.00	180.00	-180.00	180.00	-180.00	180.00	-180.00	-180.00	-179.98	-179.98
								Rotational constants								
A	9.440	9.524	9.569	9.808	9.808	9.769	10.217	9.800	10.210	10.228	9.763	9.764	10.164	10.164	10.176	10.176
B	3.625	3.648	3.656	3.813	3.813	3.798	3.940	3.745	3.885	3.894	3.725	3.726	3.853	3.853	3.860	3.860
C	2.701	2.722	2.729	2.836	2.836	2.825	2.938	2.799	2.907	2.912	2.785	2.786	2.886	2.886	2.890	2.890
								Dipole moment								
μ	1.458	1.336	1.311	1.251	1.251	1.219	1.498	1.039	1.355	1.357	1.072	1.068	1.324	1.318	1.323	1.318

Table 3: Continued.

Level of Theory / Coordinate [a]	MP2 / Basis Set											Experimental
	S"	S"I	S"II	S"α	S"β	S"αI	S"βI	S"αII	S" βII	S"βIII	C'I [b]	Value [c]
						Bond Lengths						
R												
N1-C2	1.4630	1.4499	1.4513	1.4618	1.4618	1.4476	1.4476	1.4491	1.4491	1.4444	1.456	1.466
N1-H3	1.0131	1.0128	1.0152	1.0134	1.0134	1.0128	1.0128	1.0148	1.0148	1.0142	1.021	1.001
C2-C5	1.5193	1.5189	1.5195	1.5208	1.5208	1.5196	1.5196	1.5205	1.5205	1.5123	1.523	1.529
C2-H6	1.0943	1.0940	1.0945	1.0947	1.0947	1.0941	1.0941	1.0945	1.0945	1.0919	1.102	1.081
C5=O8	1.2434	1.2096	1.2086	1.2432	1.2432	1.2110	1.2110	1.2100	1.2100	1.2072	1.220	1.204
C5-O9	1.3963	1.3526	1.3540	1.3981	1.3981	1.3549	1.3549	1.3565	1.3565	1.3521	1.367	1.354
O9-H10	0.9788	0.9675	0.9671	0.9811	0.9811	0.9692	0.9692	0.9682	0.9682	0.9705	0.974	0.966
						Bond angles						
∠A												
C2-N1-H3	113.25	109.21	108.56	114.30	114.30	110.48	110.48	109.89	109.89	109.97	109.4	-
N1-C2-C5	115.14	115.18	114.97	115.35	115.35	115.70	115.70	115.53	115.53	115.50	115.4	113.0
N1-C2-H6	109.45	109.81	110.01	109.38	109.38	109.73	109.73	109.87	109.87	110.01	-	-
C2-C5=O8	126.54	125.57	125.35	126.70	126.70	125.92	125.92	125.73	125.73	125.78	125.8	125.1
C2-C5-O9	110.92	111.19	111.25	110.53	110.53	110.80	110.80	110.91	110.91	110.95	111.0	111.5
C5-O9-H10	110.31	106.80	105.40	111.04	111.04	107.65	107.65	106.25	106.25	106.34	106.1	110.5
						Torsion angles						
∠τ												
H3-C2-N1-H4	127.01	115.64	113.46	129.91	129.91	118.29	118.29	116.41	116.41	116.11	-	-
H3-N1-C2-C5	-63.51	-57.82	-56.73	-64.96	-64.96	-59.15	-59.15	-58.21	-58.21	-58.06	-57.3	-
H7-C2-N1-C5	121.72	122.00	121.69	121.73	121.73	122.06	122.06	121.76	121.76	121.88	-	-
O8=C5-C2-N1	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.0	0.00
O9-C5-C2-O8	179.99	180.00	180.00	179.98	179.98	180.00	180.00	180.00	180.00	179.99	-	-
					Rotational constants							
A	9.882	10.300	10.303	9.846	9.846	10.254	10.254	10.262	10.262	10.326	10.130	10.342
B	3.765	3.885	3.894	3.755	3.755	3.866	3.866	3.870	3.870	3.901	3.844	3.876
C	2.816	2.914	2.919	2.808	2.808	2.900	2.900	2.903	2.903	2.925	2.878	2.912
						Dipole moment						
μ	1.068	1.316	1.306	1.110	1.109	1.313	1.306	1.299	1.294	1.298	1.300	1.150

[a] Distances (R) in angstroms, angles (A and τ) in degrees, rotational constants (A, B and C) in GHz, and dipole moments(μ) in debyes. For numbering of atoms see Figure 1. Note that theoretical rotational constants refer to equilibrium values (A_e, B_e and C_e), but available experimental constants do not. [b] From ref. [15]. [c] From ref. [9].

1.2. Charge

In this section such as the energy which was discussed at above, we decide to study the point charges obtained with the HF method and different basis sets based on the effects of splitting, polarization and diffuse functions which were used in basis sets. As it is apparent from Figure 1, there are two main internal H-bonds in conformer (I) of glycine molecule, i.e. H3, H4 ··· O8 and H6, H7 ··· O9. Because of the symmetry present in the molecule, we just discuss about H3 ··· O8 and H6 ··· O9 bonds. Undoubtedly, the stability degree of this conformer is directly proportional to the strength of the H-bonds. Unfortunately, due to the lack of comparative data in the literatures to determine the appropriate basis set, we only show how the strength of H-bonds will be changed by the basis sets.

3.5.1. Splitting effect: Figure 12 shows the charge on different atoms versus the basis set for conformer (I) of glycine, mainly ranking on splitting. In the minimal basis sets (STO-nG) as n increase, a tangible increment on the absolute value of charges on the atom centers and so the strength of intramolecular H-bonds observes. The increment in the number of primitive gaussians for the electrons of core (as happened for 3-21G, 6-21G and their derivatives) and valance layer (as regarded to 6-21G and 6-31G) doesn't affect so much on the value of atomic charges, but more splitting (like 6-311G triple zeta against 6-31G double zeta and their derivatives) in basis sets catch a considerable changes on atomic charges towards weaker H-bonding. With comparing the couples of S'II-S"II, S'α- S"α and S'β- S"β, it is obvious that the descending procedure of charge quantities on the atomic centers is totally general, but exceptionally the presence of diffuse function in basis set just on the C2 and H6 atoms contravene severely this generality, so that the atomic charge on these centers increase exceptionally. For instance, the S"β basis set predicts that the C2 atom is even more electronegative than N1 atom. However, the subject of weakening the H-bonds doesn't change.

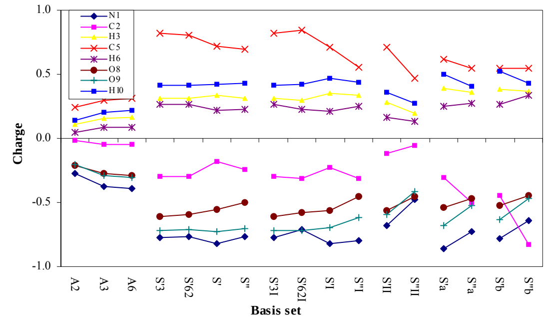


Figure 12. Charge on different atoms versus basis set for conformer (I) glycine, ranking on splitting.

3.5.2. Polarization effect: As a general rule for the addition of polarization functions in different types of double and triple zeta basis sets, we can express: "by the inclusion of more polarization functions to each type of basis set derivatives, there is a propensity to decrease of the absolute charge concentrated on the whole atoms, and consequently to weaken the H-bonds." The decrement intensity is further specifically when the polarization function adds to basis set for the hydrogen atoms. This descending run repeats with going from (d, p) to (2df, pd) except for S"β, while everything is reversed completely in (3df, 3pd) and basis set tends severely to show the charge values on the atoms much intense and make the intermolecular bonds much more polar. This changes are dominant especially for C5 which has become strongly positive, for N1, O8 and O9 which has become strongly negative and also for C2 which in S"α unlike to the negative values predicted from all the other basis sets precipitate strongly positive.

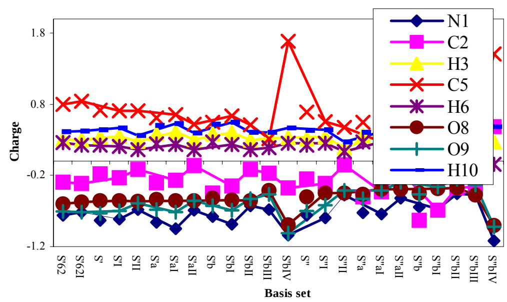


Figure 13. Charge on different atoms versus basis set for conformer (I) glycine,

ranking on the number of added polarization functions.

3.5.3. Diffuse Effect: Involving the diffuse functions in basis sets except for some atomic centers, doesn't afford a considerable change in atomic charges. The centers of C5 and N1 enhance to a relatively severe diminution of charge value and inversely the C2 center to a severe accretion of the negative charge. In the derivatives of 6-311G basis set, these changes is more intensive due to the increment of positive charge on H6, H7 (for C2) and decrement of negative charge on O9 nucleus. Therefore we except to obtain a bit more stable structures through using diffuse functions in the basis sets especially in the 6-311G and their derivatives.

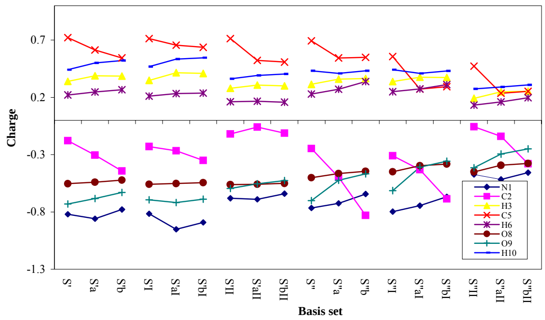


Figure 14. Charge on different atoms versus basis set for conformer (I) glycine, ranking on the number of added diffuse functions.

1.3. IR Spectrum

The frequency calculations were done at the stationary points obtained by optimization separately at each level of theory.

So in Tables 4 and 5, we have listed fifteen normal modes of neutral glycine in gas phase calculated by the HF and MP2 methods, respectively, with 29 different basis sets, in comparison with the corresponding experimental values resulted from the matrix-isolation infrared spectroscopy given by Stepanian and et al [9]. The numbers in the gray cadres show more conformity with the experimental data, but the bold numbers insist on wrong or remote data and also the maximum digression from the experiments. The regular numbers are those we don't have opined about them.

In general, by comparison of the results of calculations by the HF and MP2 methods as shown in Tables 4 and 5 respectively, it is obvious that the HF method due to the lack of electron correlation inclusion, overestimates the frequencies more than the MP2 method. Meanwhile the latter one has less mistakes than the earlier one in predicting the intensities.

The STO-nG basis sets overestimate intensely the frequencies and have completely wrong assessment for the intensities especially in the HF method.

The 3-21G and 6-21G basis sets have an adequate evaluation of the frequencies under 1429 cm^{-1} and over 3410 cm^{-1}. In this range (except for 619 cm^{-1} in HF and 3560 cm^{-1} for MP2), the intensities are assessed accurately, whereas the inmost intensities especially for the HF method are miscued severely. By comparison of the results obtained .

Table 4: IR Frequencies (cm^{-1}) and Intensities of Glycine, calculated by HF method.

HF/ Basis set	IR Normal Modes of Glycine																													
	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int
A2	473	18	547	32	678	18	988	20	1031	0	1219	42	1358	8	1371	2	1373	3	1801	0	2063	4	2240	81	3691	1	4166	0	4377	7
A3	465	24	539	45	659	17	959	25	1026	1	1141	95	1313	11	1360	3	1677	10	1778	2	2023	7	2132	81	3601	1	4088	2	4245	7
A6	467	25	538	48	662	17	959	29	1025	1	1130	100	1313	11	1358	3	1677	11	1774	2	2015	2	2135	80	3601	1	4080	0	4205	6
S'3	493	40	545	66	684	4	795	323	915	52	996	12	1178	82	1222	239	1514	36	1619	24	1861	28	1967	263	3241	10	3781	7	3868	66
S'3I	493	40	545	66	684	4	795	323	915	52	996	12	1178	82	1222	239	1514	36	1619	24	1861	28	1967	263	3241	10	3781	7	3868	66
S'62	491	41	541	66	683	5	800	306	913	59	997	10	1175	97	1218	218	1512	35	1621	22	1859	26	1948	255	3239	11	3773	5	3844	59
S'62I	497	33	543	40	705	110	905	62	1007	5	1062	171	1224	23	1287	298	1556	36	1611	9	1845	19	2054	307	3215	23	3699	0	3924	73
S'	501	49	556	55	688	201	745	380	933	37	1000	14	1220	250	1260	83	1548	27	1618	25	1861	38	1922	328	3230	14	3918	16	3992	93
S'I	504	34	553	39	708	138	909	79	1008	7	1045	199	1232	40	1297	2	1570	36	1610	9	1852	25	2035	386	3237	20	3818	5	4052	110
S'II	505	34	550	40	704	133	903	100	1004	7	1022	179	1228	45	1287	2	1564	44	1593	10	1821	26	2032	385	3215	21	3837	6	4128	111
S'α	501	47	562	65	686	180	729	335	927	31	998	12	1207	290	1250	42	1542	19	1613	27	1863	42	1894	354	3228	13	3919	21	3983	99
S'β	501	47	562	64	685	180	728	322	926	31	997	12	1206	297	1250	38	1542	18	1613	27	1862	42	1894	351	3228	14	3917	19	3983	99
S'αI	502	34	553	42	701	136	903	84	1004	5	1026	179	1228	61	1297	2	1563	35	1605	10	1855	29	2009	430	3237	19	3831	10	4046	117
S'βI	502	34	552	42	700	136	904	80	1004	5	1029	179	1228	61	1297	2	1563	33	1605	11	1855	29	2009	428	3237	19	3827	8	4046	117
S'αII	504	34	548	45	695	11	894	115	997	148	998	6	1225	66	1285	2	1556	42	1588	11	1821	29	2007	428	3217	20	3849	12	4123	123
S'βII	504	34	548	45	695	11	894	113	998	147	998	6	1225	67	1284	2	1556	40	1588	11	1821	29	2007	426	3217	20	3849	11	4123	123
S'βIII	503	31	547	43	695	11	890	100	998	133	999	6	1219	72	1282	3	1546	35	1580	11	1808	24	1996	412	3192	22	3815	11	4096	118
S'βIV	503	31	550	41	695	11	894	88	999	5	1008	136	1219	69	1284	3	1546	35	1578	11	1809	22	1994	403	3188	22	3801	10	4096	119
S"	500	45	555	52	685	190	787	333	939	49	1003	10	1215	244	1261	98	1551	23	1617	25	1873	26	1899	347	3207	17	3874	6	4041	83
S"I	503	34	557	37	719	150	906	76	1009	7	1043	196	1225	44	1298	3	1561	30	1598	11	1869	26	2017	418	3221	24	3827	2	4121	83
S"II	504	31	548	44	700	14	902	86	1004	6	1024	163	1221	55	1285	3	1553	38	1586	11	1816	23	2015	406	3200	22	3808	5	4119	114
S"α	498	45	553	68	679	168	770	324	933	41	997	10	1208	266	1253	73	1544	21	1614	25	1869	22	1883	370	3204	16	3879	11	4031	87
S"β	499	45	555	65	680	169	770	324	933	42	996	10	1207	266	1253	73	1544	21	1613	25	1869	24	1883	367	3204	15	3878	10	4031	88
S"αI	502	34	553	45	703	144	901	79	1004	6	1026	176	1222	60	1297	2	1556	31	1597	12	1868	27	2001	448	3221	22	3835	5	4114	92
S"βI	502	33	554	44	703	145	901	79	1004	6	1026	176	1222	60	1297	2	1556	31	1597	12	1867	27	2001	446	3221	22	3835	5	4114	92
S"αII	504	32	545	51	698	11	895	96	999	6	1004	144	1219	73	1285	3	1548	37	1585	12	1815	25	2000	437	3201	20	3819	10	4115	123
S"βII	504	32	546	50	698	11	895	96	999	6	1004	143	1219	73	1285	2	1548	36	1585	12	1815	25	2000	436	3201	20	3819	10	4115	123
S"βIII	504	30	549	47	698	10	892	100	999	130	1001	6	1218	75	1285	269	1548	35	1583	12	1810	24	1994	416	3192	21	3808	11	4098	120
S"βIV	505	30	549	43	698	11	893	96	1002	6	1002	131	1220	73	1288	3	1550	34	1586	11	1808	21	1997	404	3191	22	3812	10	4109	115
Experiment[a]	463	1	500	1	619	3	801	4	883	2	907	0	1101	5	1136	2	1373	0	1429	0	1630	0	1779	?	2958	1	3410	0	3560	3
Explanation	CC-O bend, CC str		CCOH tor, CCH bend		CCOH tor, NCC-O tor		CC str, C-O str		CNH bend, CC str		CNH bend		CN str, C-O str		CN str, C-O str		CCH bend, C-O str		HCH bend		HNH bend, CNH bend		C=O str		CH6 str, CH7 str		NH str		OH str	

[a] From ref. [9].

Table 5: IR Frequencies (cm^{-1}) and Intensities of Glycine, calculated by MP2 method.

MP2/ Basis set	IR Normal Modes of Glycine																													
	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int	Freq	IR Int
A2	439	17	523	5	634	48	907	17	968	0	1142	21	1278	85	1293	2	1607	12	1705	0	1961	7	1968	41	3499	1	3879	7	4002	18
A3	428	22	510	9	621	67	880	15	962	0	1081	54	1236	73	1279	2	1556	12	1679	1	1852	35	1917	7	3399	2	3775	12	3850	22
A6	429	22	509	10	621	71	882	17	960	0	1075	57	1237	66	1276	2	1555	13	1675	1	1857	33	1910	8	3395	1	3757	9	3798	20
S'3	458	36	521	50	654	129	794	161	874	121	944	2	1085	211	1140	61	1420	1	1555	19	1744	108	1773	108	3119	9	3498	16	3535	0
S'3I	458	36	521	50	654	129	794	161	874	121	944	2	1085	211	1140	61	1420	1	1555	19	1744	108	1773	27	3119	9	3498	16	3535	0
S'62	454	37	516	49	645	126	794	143	872	130	945	1	1077	204	1141	62	1419	1	1557	18	1724	105	1770	23	3116	9	3466	12	3522	0
S'62I	469	35	521	35	690	93	860	55	956	2	1010	153	1167	135	1206	123	1457	35	1535	10	1742	17	1872	168	3117	13	3505	0	3588	37
S'	461	43	519	41	644	163	763	286	870	67	938	2	1086	281	1167	15	1425	12	1538	21	1713	143	1763	32	3091	12	3595	34	3666	4
S'I	471	36	521	33	682	119	859	64	964	2	993	185	1165	177	1204	109	1452	28	1523	13	1739	22	1849	215	3123	11	3616	4	3691	66
S'II	470	36	518	31	676	114	857	69	941	2	983	165	1160	190	1200	91	1450	27	1516	10	1713	20	1850	210	3147	13	3674	3	3814	62
S'α	460	39	523	64	624	3	736	305	855	45	933	2	1060	316	1155	2	1417	4	1525	27	1684	174	1765	39	3085	11	3563	44	3678	12
S'β	460	39	522	61	624	3	737	297	855	48	934	2	1059	319	1154	2	1417	4	1521	26	1684	173	1763	39	3083	11	3563	43	3673	10
S'αI	468	36	517	42	652	113	848	73	936	1	959	164	1146	253	1193	47	1435	22	1511	17	1739	26	1812	268	3117	12	3631	10	3665	75
S'βI	468	36	515	40	651	114	847	72	936	1	958	160	1146	255	1192	47	1436	21	1507	17	1737	26	1812	266	3117	12	3629	9	3665	74
S'αII	468	35	510	41	644	105	842	88	930	1	941	132	1143	262	1189	38	1433	21	1505	14	1712	25	1813	261	3141	13	3691	10	3791	74
S'βII	468	35	509	39	643	106	841	87	930	1	940	130	1143	264	1189	38	1433	20	1502	14	1711	25	1813	259	3140	13	3691	9	3791	73
S''	461	40	512	40	634	152	781	229	874	100	936	2	1083	288	1170	21	1432	7	1518	21	1700	155	1768	36	3046	14	3625	37	3633	2
S''I	473	34	518	37	683	122	857	62	938	2	994	191	1166	171	1204	142	1439	24	1496	16	1757	25	1842	245	3105	14	3648	3	3772	49
S''II	472	32	507	42	652	92	853	62	937	1	983	151	1155	194	1194	1	1430	28	1481	12	1677	19	1842	228	3102	15	3632	3	3818	66
S''α	460	38	511	70	622	4	752	256	860	61	927	1	1066	316	1159	6	1424	4	1513	24	1679	175	1768	39	3041	12	3598	41	3639	8
S''β	460	38	512	67	622	4	752	255	859	62	927	1	1065	318	1158	6	1423	4	1513	24	1679	174	1766	38	3041	12	3598	41	3638	7
S''αI	472	34	508	52	645	6	849	67	930	2	965	168	1153	240	1194	83	1429	20	1489	17	1758	24	1820	284	3103	13	3655	8	3751	57
S''βI	471	34	509	49	646	112	849	68	929	2	963	166	1153	241	1194	84	1429	20	1488	17	1757	24	1820	282	3102	13	3655	8	3751	57
S''αII	470	32	494	56	641	6	844	72	927	1	951	128	1144	255	1188	64	1419	22	1477	14	1679	22	1821	265	3101	14	3645	9	3805	75
S''βII	470	31	495	54	641	6	844	72	926	1	950	128	1144	255	1188	65	1419	22	1475	14	1679	22	1821	264	3100	14	3644	9	3805	75
Experiment[a]	463	1	500	1	619	3	801	4	883	2	907	0	1101	5	1136	2	1373	0	1429	0	1630	0	1779	?	2958	1	3410	0	3560	3
Explanation	CC-O bend, CC str		CCOH tor, CCH bend		CCOH tor, NCC-O tor		CC str, C-O str		CNH bend, CC str		CNH bend		CN str, C-O str		CN str, C-O str		CCH bend, C-O str		HCH bend		HNH bend, CNH bend		C=O str		CH6 str, CH7 str		NH str		OH str	

[a] From ref. [9].

with the 3-21G and 6-21G sets by both HF and MP2 methods, we can deduce that the increment of primitives for core electrons causes an slight improvement for the frequencies, although most of the time it is negligible.

Comparing the results of the 6-21G and 6-31G simple double zeta split valence basis sets reveals that the basis set containing less primitives in valence layer (i.e. 6-21G) shows a better correspondence with experimental values in both HF and MP2 methods. However, in prediction of intensities, 6-31G are more prosperous.

The comparison of the 6-31G double zeta with 6-311G triple zeta split valence basis sets shows that there is not any substantial difference between their results in the MP2 method, while in HF, it is for the benefit of 6-311G.

We can infer the efficacy of polarization functions via the comparison of basis sets in three groups (6-21G, 6-21G*), (6-31G, 6-31G*, 6-31G**) and (6-311G, 6-311G*, 6-311G**). As one can see, the addition of polarization functions either in the HF or MP2 method especially under frequency 1429 cm^{-1} pervert most of the frequencies from the experiment and misstate the intensities. In the HF method, over the frequency 1429 cm^{-1}, this addition causes an improvement in the intensity predictions. These deductions repeat exactly for the groups (6-31+G, 6-31+G*, 6-31+G**), (6-31++G, 6-31++G*, 6-31++G**, 6-31++G(2d,p), 6-31++G(3d,3p)) and the corresponding groups for 6-311+G and 6-311++G.

On the other hand, with comparing the results obtained by basis sets in two groups (6-31G, 6-31+G, 6-31++G) and (6-311G, 6-311+G, 6-311++G) in both HF and MP2 methods, we realize that the addition of diffuse functions always improves the frequencies without any salient changes. The efficacy of adding diffuse functions for hydrogen atoms is much less than it for heavy atoms. These deduction repeats for the groups (6-31G*, 6-31+G*, 6-31++G*), (6-31G**, 6-31+G**, 6-31++G**) and the corresponding groups for 6-311G* and 6-311G**.

At the end, we can suggest that the simple double zeta 6-21G and 6-31G and triple zeta 6-311G split valence basis sets and also their corresponding diffuse augmented basis sets achieve more success in predicting IR spectrum of glycine respect to the other basis sets either with the HF or MP2 method. It seems that 6-311++G are more successful among them, but if necessary to reduce the calculation time, one can content oneself with those simple double and triple zeta basis sets. Finally, it should be noted that the HF method for determination of the intensities and the MP2 method for predicting more accurate frequencies are more adequate.

2. Conclusions

The calculations were accomplished with two HF and MP2 methods using the various basis sets including the STO-nG series (n= 2,3 and 6) and the derivatives of Pople's double and triple zeta basis sets including 3-21G, 6-21G, 6-31G, and 6-311G which were augmented with the different combinations of diffuse and polarization functions, as we listed in Table 1. Dunning's cc-pVnZ basis sets have also been applied to determine the HF limit of the molecule properties, as we presented in Table 2 and Figure 3. The fully geometry optimization of the conformer I of gaseous glycine (figure 1) was done by each level of theory, without any symmetry restrictions. The results were compared together and with the experiment and also the corresponding HF limits to find how these properties depend on the basis sets.

The following mentionable conclusions can be drawn from the present theoretical study:

1. In study of total energy, dipole moment and zero point energy in the 3.3 section, we discussed comprehensively the effects of the increment of splitting, polarization and diffuse functions to basis sets. As seen, the increase of primitives in the minimal basis sets and the splitting in the split valance basis sets cause a continuous decrease

in the energy level of system and coming close to HF limit. The number of primitives in core layer on the contrary to valence layer impress extremely on energy. On the other hand, the triple zeta basis sets tend to improve the results more than double zeta basis sets. In general, the addition of polarization functions in basis sets exponentially decrease the energy, so that the assessment of basis set limit is predictable. The addition of diffuse function for hydrogen atoms doesn't create tangible changes in above properties.

Moreover, The MPn methods (in any order) and the B3LYP, due to considering electron correlation generally predict the energy about %0.3 and %0.5 lower, respectively and also have improvable effect on dipole moment results.

With attention to aforesaid highlights and Figures 2 to 10 in section 3.3, we suggest 6-311G basis set to run a quick calculation with adequate accuracy and the bigger 6-311+G** basis set for more accurate calculations.

2. In geometry studies we find that STO-nG basis sets overestimates the bond distances of glycine. This overestimation improves as n increase. 3-21G overestimate almost all the bond lengths. The same observation can be made for the 6-21G sets.

On the other hand, there is no meaningful difference between the results of 6-31G double zeta and 6-311G triple zeta basis sets and their corresponding derivatives. Anyhow, in both kind of basis sets, the addition of polarization functions for heavy atoms makes a relative improvement in the bond lengths. Anyway, further addition of polarization for hydrogen atoms makes them results a bit critical. Also more addition of polarization function and the complexity of the calculations doesn't improve the results. Also, the basis sets augmented with diffuse functions underestimate the most of bond lengths. At last, the results of MP2 have better agreement with the experiment respect to the HF method.

The effect of the addition of polarization functions and diffuse functions on angle values is completely inversed together. The addition of more polarization functions enforce the angles to tend to experiment in a gentle run. The addition of the first polarization function, impress very intensive only for C2-N1-H3 and C5-O9-H10 angles equal to about 5.5° for 6-31G derivatives. The more inclusion of polarization functions impress them much less.

So with attention to above, it seems that the 6-31G* can be the most adequate to attain the geometry parameters.

3. In the calculation of atomic charges we conclude that:

As n increase in STO-nG, a tangible increment on the absolute value of charges on the atom centers and so the strength of intramolecular H-bonds observes. The increment in the number of primitive gaussians for the electrons of core and valance layer doesn't affect so much on the value of atomic charges. Also more splitting valence basis sets catch a considerable changes on atomic charges towards weaker H-bonding. Furthermore a general rule for the addition of polarization functions in different types of double and triple zeta basis sets consist "with including more polarization functions to each type of basis set derivatives, there is a propensity to decrease the absolute charges on the all atoms, and consequently to weaken the H-bonds". On the other hand, we except to obtain a bit more stable structures through using the diffuse functions especially in the 6-311G basis set and their derivatives.

4. In the study of the IR spectrum, we find that the HF method generally because of the lack of embracing the proportion of electron correlation overestimates the frequency much more than the MP2 method. Moreover it makes more mistake to predict the intensities. Then the HF method for determination of the intensities and the MP2 method for predicting of more accurate frequencies are more adequate.

The STO-nG minimal basis sets overestimate intensely the frequencies and evaluate the intensities completely wrong. By comparing the results obtained by 3-21G and 6-21G basis sets, it is revealed that the effect of the number of primitives in core layer on the results are negligible. Through two simple double zeta split valence 6-21G and 6-31G basis sets, the 6-21G which applies less number of primitives in the valence layer is more valid in the estimation of the frequencies whereas it does inversely in determination of the intensities. Also it is manifested that the increment of splitting in valence layer in HF method improve the results. In deed, the addition of polarization functions to basis sets digress always the results from the experiment. Whereas the addition of diffuse function especially for heavy atoms operates vise versa.

Finally, we suggest that the simple double zeta 6-21G and 6-31G and triple zeta 6-311G split valence basis sets and also their corresponding diffuse augmented basis sets are valid in predicting of the IR spectrum of glycine. It seems that 6-311++G are more successful among them, but to shorten the time of calculations, one can content oneself with those simple double and triple zeta basis sets.

References

1. P. J. Greenstein, M. Winitz, Chemistry of the Amino Acids, **1961**, John Wiley & Sons Ltd, New York, 435- 522.
2. K. Iijima, K. Tanaka, S. Onuma, J. Mol. Struct. (Theochem), **1991**, 246, 257.
3. R. Destro, P. Roversi, M. Barzaghi, R. E. Marsh, J. Phys. Chem. A, **2000**, 104, 1047.
4. J. Almlof, A. Kvick, J.O. Thomas, J. Chem. Phys., **1973**, 59, 3901.
5. M. Kakihana, M. Akiyama, T. Nagumo, M. Okamoto, Naturforsch., **1988**, 43A, 774.
6. R. D. Suenram, F. J. Lovas, J. Am. Chem. Soc., **1980**, 102, 7180.
7. P. D. Godfrey, R. D. Brown, J. Am. Chem. Soc., **1995**, 117, 2019.
8. Y. Zheng, J. J. Neville, C. E. Brion, Science, **1995**, 270, 786.
9. S. G. Stepanian, I. D. Reva, E. D. Radchenko, M. T. S. Rosado, M. L. T. S. Duarte, R. Fausto, and L. Adamowicz, J. Phys. Chem. A, **1998**, 102, 1041.
10. S. Vishveshwara, J. A. Pople, J. Am. Chem. Soc., **1977**, 99, 2422.
11. C. E. Dykstra, R. A. Chiles, M. D. Garrett, J. Comput. Chem., **1991**, 2, 266.
12. J. H. Jensen, M. S. Gordon, J. Am. Chem. Soc., **1991**, 113, 7917.
13. M. Ramek, V. K. W. Cheng, R. F. Frey, S. Q. Newton, L. Schäfer, J. Mol. Struct. (THEOCHEM), **1991**, 235, 1.
14. R. F. Frey, J. Coffin, S. Q. Newton, M. Ramek, V. K. W. Cheng, F. A. Momany, L. Schäfer, J. Am. Chem. Soc., **1992**, 114, 5369.
15. A. Császár, J. Am. Chem. Soc., **1992**, 114, 9568.
16. C. H. Hu, M. Shen, H. F. Schaefer, J. Am. Chem. Soc., **1993**, 115, 2923.
17. V. Barone, C. Adamo, F. Lelj, J. Chem. Phys., **1995**, 102, 364.
18. I. D. Reva, A. M. Plokhotnichenko, S. G. Stepanian, A. Y. Ivanov, E. D. Radchenko, G. G. Sheina, Y. P. Blagoi, Chemical Physics Letters, **1995**, 232, 141.
19. D. Yu, A. Rauk, D. A. Armstrong, J. Am. Chem. Soc., **1995**, 117, 1789.
20. A. Császár, J. Mol. Struct., **1995**, 346, 141.
21. D. T. Neguyen, A. C. Scheiner, J. W. Andzelm, S. Sirois, D. R. Salahub, A. T. Hagler, J. Comput. Chem., **1997**, 18, 1609.
22. S. Hoyau, G. Ohanessian, J. Am. Chem. Soc., **1997**, 119, 2016.
23. L. E. Snyder, J. M. Hollis, R. D. Suenram, F. J. Lovas, L. W. Brown, D. Buhl, Astrophys. J., **1983**, 268, 123.
24. a) C. C. J. Roothan, Rev. Mod. Phys. **1951**, 23, 69. b) J. A. Pople and R. K. Nesbet, J. Chem. Phys. **1954**, 22, 571. c) R. McWeeny, G. Dierksen, J. Chem. Phys. **1968**, 49, 4852.
25. C. Moller, M. S. Plesset, Phys. Rev., **1934**, 46, 618.
26. C. Hampel, K. Peterson, H. J. Werner, Chem. Phys. Lett., **1992**, 190, 1.
27. a) A. D. Becke, J. Chem. Phys., **1993**, 98, 5648. b) C. Lee, W. Yang, R. G. Parr,

Phys. Rev. B, **1988**, 37, 785. c) A. D. Becke, J. Chem. Phys., **1993**, 98, 5648.

28. a) W. J. Hehre, R. F. Stewart, J. A. Pople, J. Chem. Phys., **1969**, 51, 2657. b) J. B. Collins, P. V. R. Schleyer, J. S. Binkley, J. A. Pople, J. Chem. Phys., **1976**, 64, 5142.
29. a) W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. Defrees, J. A. Pople, J. S. Binkley, J. Am. Chem. Soc., **1982**, 104, 5039. b) K. D. Dobbs, W. J. Hehre, J. Comp. Chem., **1986**, 7, 359. c) K. D. Dobbs, W. J. Hehre, J. Comp. Chem., **1987**, 8, 861. d) K. D. Dobbs, W. J. Hehre, J. Comp. Chem., **1987**, 8, 880.
30. a) J. S. Binkley, J. A. Pople, W. J. Hehre, J. Am. Chem. Soc., **1980**, 102, 939. b) M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, W. J. Hehre, J. Am. Chem. Soc., **1982**, 104, 2797.
31. a) R. Ditchfield, W. J. Hehre, J. A. Pople, J. Chem. Phys., **1971**, 54, 724. b) W. J. Hehre, R. Ditchfield, J. A. Pople, J. Chem. Phys., **1972**, 56, 2257. c) P. C. Hariharan, J. A. Pople, Mol. Phys., **1974**, 27, 209. d) M. S. Gordon, Chem. Phys. Lett., **1980**, 76, 163. e) P. C. Hariharan, J. A. Pople, Theo. Chim. Acta., **1973**, 28, 213.
32. a) A. D. McLean, G. S. Chandler, J. Chem. Phys., **1980**, 72, 5639. b) R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, J. Chem. Phys., **1980**, 72, 650. c) A. J. H. Wachters, J. Chem. Phys., **1970**, 52, 1033. d) P. J. Hay, J. Chem. Phys., **1977**, 66, 4377.
33. a) T. H. Dunning Jr., J. Chem. Phys., **1989**, 90, 1007. b) D. E. Woon, T. H. Dunning Jr., J. Chem. Phys., **1993**, 98, 1358.
34. M. J. Frisch, J. A. Pople, J. S. Binkley, J. Chem. Phys., **1984**, 80, 3265.
35. T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, J. Comp. Chem., **1983**, 4, 294.
36. R. S. Mulliken, J. Chem. Phys., 1955, 23, 1833.
37. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, , J. A. Pople, Gaussian 98, Revision A.6, 1998, Gaussian, Inc., Pittsburgh PA.
38. C. J. Cramer, Essentials of Computational Chemistry: Theories and Models, **2002**, John Wiley & Sons Ltd, West Sussex, pp. 164.