ELSEVIER

Journal of Molecular Structure: THEOCHEM 808 (2007) 63–70

THEO
CHEM

www.elsevier.com/locate/theochem

On the determination of bond lengths by ab-initio methods II: Simplifications and new considerations

J. Sánchez Márquez, P.N. Arenas Hinojosa, D. Zorrilla Cuenca, M. Fernández Núñez *

Departamento de Química-Física de la, Universidad de Cádiz Facultad de Ciencias, Apartado 40, 11510 Puerto Real, (Cadiz) España, Spain

Received 21 March 2006; received in revised form 28 November 2006; accepted 21 December 2006
Available online 31 December 2006

Abstract

This work analyzes some improvements and simplifications made on procedures to calculate molecular bond lengths, through the combination of different theoretical methods that we have proposed some time ago. The main improvement that we have presently achieved lies in the combined application of the two procedures already presented, one related to the reduction of systematic errors and the other associated to the reduction of random errors. A second part of the work is devoted to explore the possibility to abbreviate the procedure to reduce random errors, either by reducing the basis set size in ab-initio methods or by modifying the parametrization in semiempirical methods. Finally, we present a first attempt of developing a INDO-like method specifically parametrized to obtain bond lengths, suitable to be successfully combined with the well known PM3 semiempirical method.

Keywords: Bond length; ab-initio methods; Semiempirical methods; B3LYP; Moeller-Plesset; INDO; Molecular geometry

1. Introduction

In a previous publication [1] we studied two different procedures to improve ab-initio calculation of bond lengths. The first one minimizes the systematic error by using linear relations:

$$R^{\text{Improved}} = a + b \cdot R^{\text{Calculated}} \tag{1}$$

The *a* and *b* values, being different for each combination (method + basis set), are determined by a least square fitting of experimental and calculated data: $R^{\text{Experimental}} \approx a + b \cdot R^{\text{Calculated}}$. The second procedure reduces the random error by combining outcomes of two *statistically independent* standard methods:

$$R = \alpha_1 R_1 + \alpha_2 R_2 \tag{2}$$

with weights α_1 and α_2, inversely proportional to the estimated quadratic error of each method. As final outcome of this second procedure, we proposed the formula:

$$R^{\text{improved}} = 0.32 \cdot R^{\text{MP2/6-311G(3df,3pd)}} + 0.68 \cdot R^{\text{B3LYP/6-311G(3df,3pd)}} \tag{3}$$

In this work we have improved the results reached in [1] by mixing both procedures, and we have reduced the calculation effort necessary to determinate the bond lengths by using two ways:

a) Through moderate size basis sets.
b) Through semiempirical methodologies.

This is important because if we try to apply the expression (3) to large molecules, we realize that calculations using 6-311G(3df,3pd) basis set become unaffordable.

By using the two ways above indicated, we obtain two equations similar to (3), one by combining B3LYP/6-311G(d,p) with MP2/6-311G(d,p) and another one by combining the PM3 semiempirical method with a modification of the INDO method specifically designed to obtain suitable theoretical bond lengths. This version of INDO is under development in our laboratory, and is briefly described at the end of this paper. We have also analysed

---

* Corresponding author. Tel.: +34 956 016 328; fax: +34 956 016 471.
E-mail address: manuel.fernandez@uca.es (M.F. Núñez).

0166-1280/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.theochem.2006.12.038

the possibility of changing the least square fittings that leads to formula (1) depending on the nature of the elements involved. This has allowed us to improve formula (3) and to complete the classification of errors carried out in [1], by using different values of "*a*" and "*b*" in formula (1) if we take elements from different rows of the periodic table.

2. Reducing total errors (systematic + random)

We have obtained a new improvement by using the results of fittings (1) as data of the formula (2). This way we combine the random error reduction reached by formula (2) with the systematic error reduction reached by the least square fitting (1), obtaining a significant reduction of the total error.

In Table 1 are the coefficients "a_i", "b_i" and "α_i" to apply the formula obtained by mixing (1) and (2):

$$R^{(i+j)} = \alpha_i \cdot (a_i + b_i R^{(i)}) + \alpha_j \cdot (a_j + b_j R^{(j)}) \tag{4}$$

for several representative basis sets. Last two columns of the table quote the values or the Pearson's correlation coefficients R between MP2 and B3LYP results. It's worthwhile to underline that, in order to combine the outcomes of two calculation methods by formula (2) the absolute value $|R|$ of the Pearson's correlation coefficients must be small enough. We can see in Table 1 that the substitution of direct formula (1) by the combined formula (4), significantly reduces the $|R|$ value, except for the 6-311G(3df,3pd) basis set. But the worsening is so slight in this case, that it can be negligible.

Some may find remarkable that the coefficient $R_2 = -0.019$ obtained for the 6-311G(3df,3pd) basis set and previously unfitted data falls below zero. This fact is however irrelevant, as the value obtained is virtually null: there is no correlation between the MP2 and B3LYP results obtained from this basis set, which makes them very convenient for combination. Should a significant negative value appear, it would mean that data from MP2 and B3LYP, when ordered according to the importance of the error, would end up in opposite positions. Such a thing has not happened with the methods tested here, which implies that their errors grow in the same direction for most molecules; this result seems reasonable enough.

If we examine the columns R_1 and R_2 in Table 1 we notice that this pattern of absence of correlation between the errors corresponding to methods MP2 and B3LYP that we have found for the basis set 6-311G(3df,3pd) is not valid for the rest of the basis sets. As the correlation coefficient approaches the unity, the error values for both methods tend to be proportional, and the procedure for compensating errors becomes less effective. That does not imply that this combination of two methods would yield poorer results than the best of these two methods considered individually, but the error does not approach zero anymore as the data added are no longer the sum of errors of different sign. *Hence it is important to blend methods with small correlation coefficients between their errors.*

Data from Table 1 confirm that for the basis set 6-311G(3df,3pd) we obtain $\alpha_{\text{MP2}} \ll \alpha_{\text{B3LYP}}$ whilst we have $\alpha_{\text{MP2}} = \alpha_{\text{B3LYP}}$ (for 6-311G(2d,p)) or $\alpha_{\text{MP2}} > \alpha_{\text{B3LYP}}$ for all the other basis sets. We believe that the reason for this dissimilarity is the independence between the respective influences of the method and basis set chosen. In the first case ($\alpha_{\text{MP2}} = 0.3361$), we just find that the combination of B3LYP and 6-311G(3df,3pd) is much more effective than the combination of MP2 and 6-311G(3df,3pd). In the second case the "effectiveness" is the same for both procedures, and the two last cases lead to opposite results than the first one.

Table 2 shows bond length calculation standard errors corresponding to the MP2 and B3LYP methods with several representative basis sets: 6-311G(3df,3pd), 6-311G(2d,p), 6-311G(d,p) and 6-311G basis sets, before and after reduction of systematic error by firstly reducing the systematic error by formula (1). Columns σ_{MP2} and σ_{B3LYP} correspond to the direct application of the MP2 and B3LYP methods (without any improvement). By comparing columns $\sigma_{\text{MP2+B3LYP}}$ and $\sigma_{\underline{\text{MP2+B3LYP}}}$ we can see the improvement obtained through formula (4). This is moderate for the most expensive basis set 6-311G(3df,3pd) but very significant for the rest of the basis sets. As an example, we can see that the use of a mix of the MP2 and B3LYP procedures with a basis set 6-311G(2d,p) leads to results as good as the direct use of a basis set 6-311G(3df,3pd).

All data in tables has been obtained from the experimental R_e values of the following 50 well known molecules [2,3]: H_2, C_2, N_2, O_2, LiH, BeH, BH, CH, NH, OH, FH, LiF, BeO, BeF, BO, BF, CO, CF, NO, NF, CF_2, CN_2, CO_2, C_3, H_2O, OF_2, O_3, SiF_2, SO_2, SeO_2, Cl_2, ClO, ClF, NS, PN, PO, P_2, SF, SO, S_2, H_2S, HSi, SiO, CS, HCl,

Table 1

Parameters a_i, b_i and α_i necessary to apply the formula (4) and Pearson's correlation coefficients for the combinations of MP2 and B3LYP methods with several basis sets

Basis	α_{MP2}	a_{MP2}	b_{MP2}	α_{B3LYP}	a_{B3LYP}	b_{B3LYP}	R_1	R_2
6-311G(3df,3pd)	0.3361	0.0185	0.9839	0.6639	0.0155	0.9879	0.075	−0.019
6-311G(2d,p)	0.5000	0.0419	0.9608	0.5000	0.0308	0.9829	0.262	0.515
6-311G(d,p)	0.5735	0.0421	0.9601	0.4265	0.0452	0.9590	0.444	0.644
6-311G	0.5652	0.1750	0.8279	0.4348	0.1569	0.8465	0.725	0.869

R_1 correspond to data refined with formula (1) and R_2 to data obtained directly by MP2 and B3LYP methods in [1]. The values of α_i are proportional to inverse square of standard errors σi after applying the formula (1).

Table 2
Standard errors for the calculation of bond lengths by MP2 and B3LYP methods with several representative basis sets

Basis	σ_{MP2}	σ_{B3LYP}	$\sigma_{MP2+B3LYP}$	σ_{MP2}	σ_{B3LYP}	$\sigma_{MP2+B3LYP}$
6-311G(3df,3pd)	0.0116	0.0081	0.0067	0.0101	0.0071	0.0055
6-311G(2d,p)	0.0187	0.0142	0.0143	0.0112	0.0111	0.0084
6-311 G(d,p)	0.0202	0.0206	0.0191	0.0122	0.0142	0.0108
6-311G	0.0933	0.0839	0.0863	0.0339	0.0378	0.0325

σ_{MP2} and σ_{B3LYP} correspond to MP2 and B3LYP methods directly applied (without any improvement), and $\sigma_{MP2+B3LYP}$ to both methods mixed by formula (2). $\sigma_{\underline{MP2}}$ and $\sigma_{\underline{B3LYP}}$ correspond to calculations improved by formula (1) and $\sigma_{\underline{MP2\ +\ B3LYP}}$ to the complete improvement achieved by formula (4).

LiC, MgF, MgO, PF, SiH_2. Bond length calculations have been carried out by GAUSSIAN 98 [4].

3. Simplification of the basis set

The methods B3LYP/6-311G(3df,3pd) and MP2/6-311G(3df,3pd) used in [1] can be combined because their results are statistically independent. In this work we have analysed the possibility of obtaining adequate results by using simpler basis sets, but maintaining the essential condition of statistical independence between the combined methods. Thus, before mixing a pair of procedures we have required a small Pearson's correlation coefficient between their sets of errors.

In Table 1 we can see the results of MP2 and B3LYP methods carried out with basis sets of decreasing complexity. The clearest advantage of reducing the basis sets – that will be shown in Table 4 – is that they need considerably less calculation effort, therefore allowing to handle much bigger molecules than 6-311G(3df,3pd) basis set.

We can see that the calculation of equilibrium bond lengths with basis sets without polarization leads to results so bad that, if we want to handle large molecules (where it is difficult to introduce polarization in the basis set), it can become preferable to replace ab-initio or DFT methodologies by semiempirical ones. Moreover, semiempirical methods can be very useful in order to obtain a good start point for the ab-initio calculations.

4. Use of semiempirical methods

In Table 3 we can see the Pearson's coefficients between the errors corresponding to several pairs of semiempirical methods. The last row of Table 3 shows the standard errors of the equilibrium bond length obtained by the same semiempirical methods. The procedure used to estimate the error is similar to the one applied in [1] to Hartree-Fock, Moeller-Plesset and DFT methods.

Table 3
Standard errors (σ, last row) of the equilibrium bond lengths calculated by semiempirical methods and Pearson's correlation coefficients (R) between pairs of methods

Methods	CNDO/2	AM1	PM3	INDO(mod.)
CNDO/2	1	0.309	0.236	0.238
AM1		1	0.192	0.489
PM3			1	0.071
INDO(mod.)				1
σ	0.189	0.059	0.045	0.038

The sample of 50 independent molecules is described at the bottom of the second epigraph of this article.

The methods CNDO/2, AM1 and PM3 have been applied by using the GAUSSIAN program [4]. When trying to use the INDO model we have realized that the original INDO (from Pople et al.) is not valid to study molecules containing atoms from the third period and that the MINDO methods lead to convergence problems for many of those molecules. However, as the INDO model doesn't seem to work too bad for molecules with atoms from the first and second period, we have considered useful to build up a parametrization for the INDO method including all the atoms of our study and intended to obtain bond lengths as adequate as possible. This parametrization is described in the appendix of this paper. Its results are shown in Tables 3 and 4 as "INDO(modified)". In Table 4 we can also see that the best combination of the analyzed semiempirical methods is PM3 + INDO(modified), because they are statistically independent ($R = 0.071$, see Table 3), they correspond to the least individual errors, and these errors are similar.

5. Classification attending to bond type

It is worthwhile to remark that the general results quoted in the precedent section can be improved by applying the same ideas, but in a independent way for each type of bond. We have found that a proper classification can be:

Type 1: H atom + Second period atom
Type 2: H atom + Third period atom
Type 3: Second period atom + Second period atom
Type 4: Third period atom + Second or Third period atom

When the formula $R = \alpha_1 \cdot R_1 + \alpha_2 \cdot R_2$ is applied independently for each type of bond, we obtain the results on Tables 5–7. The significant change for the coefficients attached to each type of bond comes in support of choosing different combinations for different types of bond. Table 8 shows the a and b fitting coefficients classified by types of bond, for the ab-initio and semiempirical methods studied in this work.

Table 4
Standard errors and Pearson coefficients (*R*), for the mixing of methods MP2/6-311-G(3df,3pd) + B3LYP/6-311G(3df,3pd), MP2/6-311-G(d,p) + B3LYP/6-311G(d,p) and PM3 + INDO(modified)

Methods	Basis	*R*	$\alpha_{\mathrm{MP2/\ldots}}$	$\alpha_{\mathrm{B3LYP/\ldots}}$	$\alpha_{\mathrm{MP2\,+\,B3LYP}}$
MP2 + B3LYP	6-311G(3df,3pd)	−0.019	0.3278	0.6722	0.0067
MP2 + B3LYP	6-311G(d,p)	0.644	0.5098	0.4902	0.0191
Methods	Basis	*R*	α_{PM3}	$\alpha_{\mathrm{INDO.(mod)}}$	$\sigma_{\mathrm{PM3+INDO(mod.)}}$
PM3 + INDO(mod.)	–	0.071	0.4120	0.5880	0.0298

Table 5
Standard error and mixing coefficients for methods: MP2/6-311G(3df,3pd), B3LYP/6-311G(3df,3pd) and the combination $\alpha_1 \cdot R_1 + \alpha_1 \cdot R_2$, separated by bond types and using all the molecules of the sample

Bond	Data	$\sigma_{\mathrm{MP2/6-311\ldots}}$	$\sigma_{\mathrm{B3LYP/6-311\ldots}}$	$\sigma_{\mathrm{MP2/\ldots+B3LYP/\ldots}}$	$\sigma_{\mathrm{MP2/\ldots}}$	$\sigma_{\mathrm{B3LYP/\ldots}}$
Type 1	9	0.005	0.003	0.002	0.2549	0.7451
	_9_	_0.004_*	_0.002_*	_0.001_*	_0.1823_*	_0.8177_*
Type 2	6	0.006	0.009	0.004	0.7209	0.2791
	_6_	_0.004_*	_0.004_*	_0.004_*	_0.4501_*	_0.5499_*
Type 3	17	0.013	0.006	0.006	0.1682	0.8318
	_17_	_0.012_*	_0.005_*	_0.005_*	_0.1322_*	_0.8678_*
Type 4	19	0.016	0.012	0.010	0.3688	0.6312
	_19_	_0.010_*	_0.010_*	_0.006_*	_0.5399_*	_0.4601_*
All	50	0.012	0.008	0.007	0.3263	0.6737
	_50_	_0.009_*	_0.007_*	_0.005_*	_0.3633_*	_0.6367_*

For each pair of rows, the first one correspond to direct combination of data and the ***second one**** to combine data previously refined by the systematic error reduction procedure described in [1].

Table 6
Standard error and mixing coefficients for methods: MP2/6-311G(d,p), B3LYP/6-311G(d,p) and the combination $\alpha_1 \cdot R_1 + \alpha_1 \cdot R_2$, separated by bond types and using all the molecules of the sample

Bond	Data	$\sigma_{\mathrm{MP2/6-311\ldots}}$	$\sigma_{\mathrm{B3LYP/6-311\ldots}}$	$\sigma_{\mathrm{MP2/\ldots+B3LYP/\ldots}}$	$\sigma_{\mathrm{MP2/\ldots}}$	$\sigma_{\mathrm{B3LYP/\ldots}}$
Type 1	9	0.002	0.005	0.002	0.8179	0.1821
	_9_	_0.002_*	_0.003_*	_0.002_*	_0.7631_*	_0.2369_*
Type 2	6	0.004	0.015	0.004	0.9395	0.0605
	_6_	_0.003_*	_0.003_*	_0.003_*	_0.4791_*	_0.5209_*
Type 3	17	0.014	0.006	0.006	0.1570	0.8430
	_17_	_0.013_*	_0.006_*	_0.006_*	_0.1898_*	_0.8102_*
Type 4	17	0.028	0.030	0.028	0.5277	0.4723
	_17_	_0.014_*	_0.020_*	_0.014_*	_0.6891_*	_0.3109_*
All	50	0.018	0.016	0.016	0.4450	0.5550
	_50_	_0.014_*	_0.012_*	_0.011_*	_0.4465_*	_0.5535_*

For each pair of rows, the first one correspond to direct combination of data and the ***second one**** to combine data previously refined by the systematic error reduction procedure.

It is also worth mentioning that the values of α_{MP2} and α_{B3LYP} turn out to be somewhat unpredictable; for some cases, $\alpha_{\mathrm{MP2}} \gg \alpha_{\mathrm{B3LYP}}$ while for others we find that $\alpha_{\mathrm{MP2}} \ll \alpha_{\mathrm{B3LYP}}$ or $\alpha_{\mathrm{MP2}} \approx \alpha_{\mathrm{B3}LYP}$. We have to consider that these values are chosen to be inversely proportional to the square of the standard error of the methods combined. However, these errors are very uneven, as they depend mostly on the bond type for which the calculations are carried out. It could be interesting to compare the results of combining the methods following these criteria to those obtained by assigning equal weight to each of them ($\alpha_{\mathrm{MP2}} = \alpha_{\mathrm{B3LYP}} = 0.5$). The errors $\sigma_{\mathrm{MP2\,+\,B3LYP}}$ (type 1: 0.0016, type 2: 0.0036, type 3: 0.0065 and type 4: 0.0276) become $\sigma'_{\mathrm{MP2\,+\,B3LYP}}$ (type 1: 0.0021, type 2: 0.0075, type 3: 0.0089 and type 4: 0.0277). Therefore, the weights chosen for the combination are essential.

6. Application to thiadiazoles and methylthiocyanate

The α coefficients in Tables 5–7 and the *a* and *b* coefficients in Table 8 were obtained using 50 diatomic and triatomic representative molecules but, of course, our

Table 7
Standard error and mixing coefficients for methods: PM3, INDO(mod.) and the combination $\alpha_1 \cdot R_1 + \alpha_1\ R_2$, separated by bond types and using all the molecules of the sample

Bond	Data	$\sigma_{\text{INDO(mod.)}}$	σ_{PM3}	$\sigma_{\text{PM3INDO(mod.)}}$	α_{PM3}	$\sigma_{\text{INDO(mod.)}}$
Type 1	7	0.012	0.036	0.012	0.1025	0.8975
	7	*0.009**	*0.025**	*0.010**	*0.1274**	*0.8726**
Type 2	7	0.021	0.049	0.013	0.1606	0.8394
	7	*0.012**	*0.028**	*0.010**	*0.1527**	*0.8473**
Type 3	12	0.058	0.045	0.037	0.6279	0.3721
	12	*0.056**	*0.040**	*0.037**	*0.6594**	*0.3406**
Type 4	13	0.030	0.048	0.029	0.2727	0.7273
	13	*0.029**	*0.029**	*0.026**	*0.5086**	*0.4914**
All	39	0.038	0.045	0.030	0.4120	0.5880
	39	*0.036**	*0.032**	*0.027**	*0.5559**	*0.4441**

For each pair of rows, the first one correspond to direct combination of data and the ***second one**** to combine data previously refined by the systematic error reduction procedure.

Table 8
Fitting coefficients of $R^{\text{improved}} = a + b \cdot\ R^{\text{Calculated}}$ for ab-initio and semiempirical methods studied in this work

	Basis	MP2		B3LYP		PM3		INDO (mod.)	
		a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Type1	6-311G(d,p)	0.0073	0.9944	−0.0129	1.0084	0.0304	0.9956	−0.0467	1.0466
	…G(3df,3pd)	0.0137	0.9904	−0.0120	1.0101				
Type 2	6-311G(d,p)	−0.0079	1.0065	−0.0059	0.9939	−0.0352	1.0494	0.0448	0.9609
	…G(3df,3pd)	−0.0031	1.0046	0.0000	0.9945				
Type 3	6-311G(d,p)	0.0046	0.9909	0.0095	0.9917	−0.0399	1.0445	0.0325	0.9620
	…G(3df,3pd)	0.0284	0.9756	0.0142	0.9913				
Type 4	6-311G(d,p)	−0.0388	1.0084	0.0057	0.9837	0.1308	0.9224	0.0674	0.9581
	…G(3df,3pd)	0.0089	0.9877	0.0211	0.9850				
All	6-311G(d,p)	0.0421	0.9601	0.0452	0.9590	−0.0080	1.0247	0.0244	0.9816
	…G(3df,3pd)	0.0185	0.9839	0.0155	0.9879				

purpose is that they be valid for any other molecule. In order to state the usefulness of the proposed combinations, as in our previous work [1], we have applied them to the analysis of the Glossman-Mitnik thiadiazoles geometries [5] and to the methylthiocyanate molecule [6]. Table 9 shows the bond lengths calculated by the different methods and improvements, and Tables 10 and 11 shows the corresponding standard deviations.

In Table 10, we find the error obtained with MP2/6-311G(d,p) for type 4 bonds in thiadiazol and thiocyanate to be remarkably low (this error is $\sigma = 0.0067$, while the one obtained with the same method for the 50 general type reference molecules was $\sigma = 0.0280$). In our opinion, this result happens by mere chance, caused by the similarity of all the AB bonds considered (SC and SN); this means that we could associate a satisfactory error compensation between those from the basis set 6-311G(d,p) and the MP2 method. In support of this hypothesis comes the observation that the replacement of the basis set 6-311G(d,p) by the 6-311G(3df,3pd) set makes the result *worse* ($\sigma = 0.0148$) for MP2 calculations.

7. Conclusions

(a) The formula (1) proposed in [1] can be improved by the combination:

$$R^{\text{improved}} = 0.3361 \cdot \widetilde{R}^{\text{MP2}} + 0.6639 \cdot \widetilde{R}^{\text{B3LYP}} (\sigma = 0.0055\text{Å})$$

by using:

$$\widetilde{R}^{\text{MP2}} = 0.0185 + 0.9839 \cdot R^{\text{MP2}}$$

$$\widetilde{R}^{\text{B3LYP}} = 0.0155 + 0.9879 \cdot R^{\text{B3LYP}}$$

The standard error moves from 0.0067 Å to 0.0055 Å which is a significant improvement even if the quantities seem relatively similar. This kind of combination becomes more efficient when the large basis set 6-311G(3df,3pd) is replaced by smaller ones, such as 6-311G(2d,p), 6-311G(d,p) or 6-311G.

(b) When the basis set 6-311G(3df,3pd) recommended in [1] to calculate bond lengths becomes too expensive, it is possible to obtain fairly accurate results through the combination:

Table 9
Experimental bond lengths, and calculated by the different methods, for thiadiazoles [5] and methylthiocyanate [6]

Molecule	Enlace	R_e (exp)	MP2/ …(3df,3pd)	B3LYP/ …(3df,3pd)	MP2 + B3LYP/ …(3df,3pd)	MP2/ …(d,p)	B3LYP/ …(d,p)	MP2 + B3LYP/ …(d,p)	PM3	INDO (mod.)	PM3 + INDO (mod.)
1,2,3- Thiadiazole	N1-N2	1.290	1.317	1.276	1.289	1.310	1.266	1.283	1.243	1.202	1.219
	N1-S5	1.692	1.669	1.709	1.689	1.700	1.763	1.701	1.754	1.655	1.696
	N2-C3	1.366	1.348	1.363	1.356	1.358	1.371	1.352	1.421	1.405	1.412
	C3-C4	1.369	1.384	1.367	1.371	1.386	1.367	1.366	1.369	1.347	1.356
	C3-H6	1.078	1.079	1.078	1.080	1.082	1.080	1.081	1.092	1.069	1.078
	C4-S5	1.689	1.674	1.691	1.679	1.684	1.702	1.667	1.716	1.714	1.715
	C4-H7	1.078	1.077	1.076	1.078	1.081	1.079	1.080	1.090	1.068	1.077
1,2,4- Thiadiazole	C1-N2	1.313	1.320	1.305	1.309	1.321	1.315	1.309	1.326	1.325	1.325
	C1-S5	1.707	1.696	1.714	1.702	1.707	1.756	1.702	1.739	1.700	1.716
	C1-H6	1.079	1.079	1.080	1.082	1.083	1.088	1.085	1.091	1.074	1.081
	N2-C3	1.366	1.354	1.365	1.360	1.364	1.383	1.360	1.414	1.373	1.390
	C3-N4	1.317	1.330	1.311	1.316	1.329	1.324	1.317	1.321	1.327	1.324
	C3-H7	1.078	1.079	1.081	1.082	1.083	1.089	1.085	1.096	1.074	1.083
	N4-S5	1.649	1.633	1.653	1.641	1.653	1.702	1.650	1.728	1.637	1.674
1,2,5- Thiadiazole	C1-N2	1.327	1.345	1.319	1.326	1.343	1.330	1.327	1.316	1.340	1.330
	C1-S5	1.630	1.617	1.633	1.622	1.638	1.686	1.635	1.725	1.627	1.667
	C1-H6	1.417	1.399	1.419	1.410	1.410	1.436	1.407	1.455	1.398	1.422
	N2-C3	1.081	1.079	1.081	1.082	1.083	1.089	1.085	1.093	1.069	1.079
1,3,4- Thiadiazole	C1-N2	1.320	1.315	1.295	1.301	1.317	1.308	1.304	1.330	1.348	1.341
	C1-S5	1.721	1.705	1.727	1.713	1.716	1.765	1.710	1.749	1.692	1.716
	C1-H6	1.079	1.078	1.079	1.081	1.082	1.087	1.084	1.091	1.071	1.079
	N2-C3	1.371	1.356	1.365	1.360	1.364	1.387	1.362	1.354	1.239	1.286
MeSCN	S1-C2	1.824	1.809	1.829	1.814	1.815	1.840	1.795	1.811	1.741	1.770
	S1-C6	1.684	1.685	1.689	1.682	1.694	1.699	1.671	1.657	1.732	1.701
	C2-H3	1.073	1.087	1.088	1.090	1.089	1.088	1.088	1.096	1.078	1.085
	C2-H4	1.085	1.085	1.085	1.087	1.091	1.090	1.090	1.095	1.116	1.107
	C6-N7	1.170	1.175	1.155	1.163	1.180	1.159	1.167	1.165	1.226	1.201
Standard deviations			0.013	0.008	0.007	0.009	0.026	0.017	0.035	0.041	0.029

Table 10
Standard deviations for the methods used in this paper, applied to the molecules studied in Section 6, and classified by bond type

	σ_{MP2}	σ_{B3LYP}	σ_{MP2}	σ_{B3LYP}	$\sigma_{INDO(mod.)}$	σ_{PM3}
Basis	...(3df,3pd)	...(3df,3pd)	...(d,p)	...(d,p)	–	–
Type 1	0.0082	0.0055	0.0071	0.0086	0.0148	0.0131
Type 3	0.0144	0.0108	0.0123	0.0126	0.0305	0.0382
Type 4	0.0148	0.0076	0.0067	0.0446	0.0531	0.0389

Abbreviations for basis set are: (3df,3pd) = 6-311G(3df,3pd), and (d,p) = 6-311G(d,p).

Table 11
Standard deviations for the improvements proposed in this paper, applied to the molecules studied in Section 6, and classified by bond type

	$\sigma_{MP2+B3LYP}$	$\sigma_{MP2+B3LYP}$ improved	$\sigma_{MP2+B3LYP}$	$\sigma_{MP2+B3LYP}$ improved	$\sigma_{INDO(mod)+PM3}$	$\sigma_{INDO(mod.)+PM3}$ improved
Basis	...(3df,3pd)	...(3df,3pd)	...(d,p)	...(d,p)	–	–
Type 1	0.0054	0.0053	0.0072	0.0073	0.0132	0.0196
Type 3	0.0085	0.0077	0.0100	0.0083	0.0314	0.0357
Type 4	0.0028	0.0132	0.0226	0.0141	0.0374	0.0370

Abbreviations for basis set are: (3df,3pd) = 6-311G(3df,3pd),and (d,p) = 6-311G(d,p).

$$R^{\text{improved}} = 0.5000 \cdot \widetilde{R}'^{\text{MP2}} + 0.5000 \cdot \widetilde{R}'^{\text{B3LYP}} (\sigma = 0.0084\ \text{Å})$$

by using:

$$\widetilde{R}'^{\text{MP2}} = 0.0419 + 0.9608 \cdot R^{\text{MP2/6-311G(2d,p)}}$$
$$\widetilde{R}'^{\text{B3LYP}} = 0.0308 + 0.9729 \cdot R^{\text{B3LYP/6-311G(2d,p)}}$$

or through the combination:

$$R^{\text{improved}} = 0.5735 \cdot \widetilde{R}''^{\text{MP2}} + 0.4265 \cdot \widetilde{R}''^{\text{B3LYP}} (\sigma = 0.0108\text{Å})$$

by using:

$$\widetilde{R}''^{\text{MP2}} = 0.0421 + 0.9601 \cdot R^{\text{MP2/6-311G(2d,p)}}$$
$$\widetilde{R}''^{\text{B3LYP}} = 0.0452 + 0.9590 \cdot R^{\text{B3LYP/6-311G(2d,p)}}$$

The following combination of semiempirical method can be used to obtain starting points for ab-initio geometry optimizations, or to estimate geometries of very large molecules:

$$R^{\text{improved}} = 0.4120 \cdot R^{\text{PM3}} + 0.5880 R^{\text{INDO(mod)}} (\sigma = 0.030\text{Å})$$

here INDO(mod.) represents the results obtained by the INDO method reparametrized by us.

(c) Taking into account if bonds are formed between elements from the first, the second or the third period allows us to make some improvements and to complete the classification of errors began in [1].

(d) Our parametrization of INDO, even tough is still developing, can be used for a quick estimate of bond lengths which could be particularly useful to obtain suitable starting points for more complex methods.

Appendix A. Characteristics of our modification of the INDO method

In order to improve the calculation of bond lengths, we have modified the Pople's original INDO [7] by adding polarization functions to the hydrogen atom basis set and by building up an INDO parametrization for the third period elements.

The description of the hydrogen atom has been improved by adding "1p" polarization functions whose radial part is:

$$R_{1p}(R) = \text{N}e^{-\alpha R} \tag{5}$$

These basis functions were preferred to "2p" ones in order to get the same radial dependence as for "1s" functions, so the spherical average approximation of INDO methods is better fulfilled.

In Pople's original research on INDO methods [7] there were no parameters for third period atoms. Subsequent modifications of INDO method as MINDO [9] lead too frequently to convergence problems during the geometrical optimization, and are not specifically designed to calculate bond lengths. We have built an INDO parametrization for these elements that achieves satisfactory results in calculating bond distances, without the afore mentioned convergence problems. Our parametrization uses as starting point the Pople's CNDO/2 parametrization [8], but using, as PM3 or AM1, a "3s3p" type basis set. The exclusion of Pople's "3d" functions makes the parametrization of the method easier, whilst reducing the convergence problems.

Because our procedure uses the INDO original method as a starting point, most of the parameters we used are the same as those originally used by Pople et al. [7]. So here we will describe only the new parameters necessary for our procedure.

A.1. Parameters for hydrogen

Due to the introduction of polarization functions on hydrogen, we need to set the values of the parameters α_{1p}, β_H χ_{1p}, $G^1$and F^2 for this atom. We choose the α_{1p}

exponent for the Slater's orbital to be equal to the α_{1S} exponent, $\alpha_{1S} = \alpha_{1p} = 1.22$. Our value of the β_H parameter is set to $\beta_H = -8.7$ eV. The "1p" electronegativity is set to $\chi_{1p} = -75.5$ eV. This value seems rather unusual, but it is not unfeasible as it belongs to an orbital without physical meaning, with an extremely high energy.

The G^1 and F^2 parameters for hydrogen, necessary to calculate the one center bielectronic integrals are $G^1 = 0.3326$ u.a. and $F^2 = 0.2001$ u.a. These values have been obtained by the same procedure as the values for third row atoms, whose parameters are described below.

Table 12
Parameters G^1 and F^2 (u.a.) for hydrogen and third period atoms

Element	G^1	F^2
H	0.3326	0.2001
Na	0.0802	0.0392
Mg	0.1012	0.0590
Al	0.1249	0.0795
Si	0.1510	0.1005
P	0.1799	0.1222
S	0.2113	0.1446
Cl	0.2451	0.1674

Table 13
Parameters β_{AB} (e.V.) used for third period atoms

(A–B)	β_{AB}	(A–B)	β_{AB}	(A-B)	β_{AB}
Na–H	−12.29	Mg–Cl	−18.84	P–Be	−14.04
Na–Li	−8.36	Al–H	−10.00	P–B	−16.04
Na–Be	−10.36	Al–Li	−10.15	P–C	−18.22
Na–B	−12.36	Al–Be	−12.15	P–N	−25.63
Na–C	−14.36	Al–B	−14.15	P–O	−39.10
Na–N	−16.36	Al–C	−16.15	P–F	−48.85
Na–O	−43.49	Al–N	−18.15	P–P	−15.07
Na–F	−68.88	Al–O	−40.06	P–S	−21.59
Na–Na	−1.81	Al–F	−62.33	P–Cl	−18.05
Na–Mg	−8.58	Al–Al	−6.78	S–H	−13.43
Na–Al	−9.51	Al–Si	−12.18	S–Li	−13.58
Na–Si	−10.39	Al–P	−13.19	S–Be	−15.58
Na–P	−11.40	Al–S	−14.73	S–B	−17.18
Na–S	−12.94	Al–Cl	−20.80	S–C	−20.81
Na–Cl	−19.09	Si–H	−10.88	S–N	−26.19
Mg–H	−12.45	Si–Li	−11.03	S–O	−34.42
Mg-Li	−9.22	Si–Be	−13.03	S–F	−35.53
Mg–Be	−11.22	Si–B	−15.03	S–S	−18.15
Mg–B	−13.22	Si–C	−17.03	S–Cl	−17.62
Mg–C	−15.22	Si–N	−25.27	Cl–H	−15.52
Mg–N	−17.22	Si–O	−38.37	Cl–Li	−15.72
Mg–O	−41.84	Si–F	−55.31	Cl–Be	−17.33
Mg–F	−63.36	Si–Si	−11.24	Cl–B	−18.95
Mg–Mg	−9.45	Si–P	−9.49	Cl–C	−18.72
Mg–Al	−10.37	Si–S	−15.61	Cl–N	−18.48
Mg–Si	−11.26	Si–Cl	−20.67	Cl–O	−27.09
Mg–P	−12.26	P–H	−11.89	Cl–F	−27.81
Mg–S	−13.80	P–Li	−12.04	Cl–Cl	−17.03

A.2. Parameters for third period atoms

We use Slater's exponents a_m, CNDO/2 electronegativities χ_x and χ_p, and standard coulomb integrals γ_{AB}. New G^1 and F^2 parameters, necessary in INDO, have been set starting from analytical values of integrals (sp|sp) and (pp′|pp′), through two steps:

1) An analytical calculation of $G^1 = 3(sp|sp)$ and $F^2 = (pp'|pp')$ for all the atoms (first and second row).
2) A fitting between analytical and experimental values for the second row atoms. In the end we use the parameters obtained by this fitting to calculate the values of G^1 and F^2 for hydrogen and third row atoms, obtaining the results quoted in Table 12.

Displayed in Table 13 are the β_{AB} parameters optimized to reproduce the experimental geometries of 50 representative molecules.

An option to carry out INDO(mod.) calculations is included in our program UCA-MOL, that can be obtained from our web [10] (within the section: "Software/Programas CTPM").

References

[1] J. Sánchez Márquez, M. Fernández Núñez, J. Mol. Struct. (Theochem) 624 (2003) 239–249, **Erratum**: The Table 4 of [1] contains two mistakes: (1) For the 1,2,5-thidiazol molecule the C_1–N_2, C_1–S_5, C_1s–H_6 and N_2–C_3 distances actually belong to C_3–N_2, N_2–S, C_3–C_4 and H–C_3. (2) For the 1,3,4-thidiazol molecule the C_1–N_2, C_1–S_5, C_1–H_6 and N_2–C_3 distances actually belong to C_2–N_3, C_2–S, C_2–H and N_3–N_4.

[2] K.P. Huber, G. Herzberg, IV Constants of Diatomic Molecules Van Nostrand, NY (1979).

[3] CRC Handbook of Chemistry and Physics (76 Ed.). CRC Press (1997).

[4] Gaussian 98, Revision A.9, 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, A.G. Baboul, 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, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, and J.A. Pople, Gaussian, Inc., Pittsburgh PA, (1998).

[5] D. Glossman-Mitnik, A. Márquez-Lucero, J. Mol. Struct. (Theochem.) 548 (2001) 153.

[6] Landolt-Börnstein, New Series, group II volume 7, Springer Verlag, Berlin, 1976.

[7] J.A. Pople, D. Beveridge, P. Dobosh, J. Chem. Phys. 47 (1967) 2026.

[8] G. Segal, J. Pople, J. Chem. Phys. (1966) 3289.

[9] M. Dewar, W. Thiel, J. Am. Chem. Soc. 99 (1977) 4499.

[10] http://www.uca.es/dept/quimica_fisica/.