

The Linear Relationship Between Koopmans' and Hydrogen Bond Energies for some Simple Carbonyl Molecules

Roy E. Bruns^{*,a}, Roberto L. A. Haiduke^a and Antonia T. do Amaral^b

^a Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13083-970 Campinas - SP, Brazil

^b Instituto de Química, Universidade de São Paulo, CP 26077, 05513-970 São Paulo - SP, Brazil

Recentemente Galabov and Bobadova-Parvanova mostraram que a energia de formação da ligação de hidrogênio obtida por cálculo no nível HF/6-31G(d,p) está altamente correlacionada com o potencial eletrostático molecular na região acceptora em alguns compostos carbonílicos simples. Neste trabalho mostramos que o potencial eletrostático pode ser substituído pela energia de Koopmans. A correlação entre esta energia e a energia de formação da ligação de hidrogênio é tão alta quanto aquela observada por Galabov e Bobadova-Parvanova. O potencial de Siegbahn relacionando às energias de Koopmans e cargas GAPT mostra que a energia de ligação de hidrogênio não está simplesmente correlacionada com a carga da região acceptora pois as cargas dos átomos vizinhos são também importantes no processo de ligação de hidrogênio.

Recently Galabov and Bobadova-Parvanova have shown that the energy of hydrogen bond formation calculated at the HF/6-31G(d,p) level is highly correlated with the molecular electrostatic potential at the acceptor site for a number of simple carbonyl compounds. Here it is shown that the electrostatic potential can be replaced by Koopmans' energy. The correlation between this energy and the hydrogen bond formation energy is just as high as the one observed by Galabov and Bobadova-Parvanova. The Siegbahn simple potential relating Koopmans' energies and GAPT charges shows that the hydrogen bond energy is not simply correlated with the charge of the acceptor site because the charges on the neighboring atoms are also important in the hydrogen bonding process.

Keywords: hydrogen bonding, atomic charges, Koopmans' energy, QSAR

Introduction

The success of QSAR studies depends on whether the molecular descriptors chosen are appropriate to explain biological activities. Descriptors are obtained from a number of sources such as experimental physical-chemical data, geometrical structure parameters and theoretical electronic indexes obtained from quantum mechanical calculations.¹ Among the electronic indexes most prominent in QSAR studies are the atomic charges. The charge values used in QSAR studies are dependent on the basis set used, the level of electron correlation treatment and the method used to extract charge values from the molecular wave function.

Considering the models generated by 3D-QSAR, for example using the CoMFA (Comparative Molecular Field

Analysis) approach, they can be used either to predict biological activity – taking into account mainly the statistical aspects of the model – or to get information about the physicochemical molecular surroundings, in order to describe single steps of interactions with receptor binding sites. If the charges are correlated with the activity they can be used for predictions even though the absolute charges are not correct. On the other hand, the electrostatic fields generated from different calculational methods could simulate very different physico-chemical conditions of ligand-protein interactions.

Folkers *et al.*² studied the effect of charge calculation methods in CoMFA models performed for 24 substituted N2-phenylguanines, which inhibit the Herpes Simplex Virus 1 Thymidine Kinase (HSV1 TK inhibitors). Similar CoMFA results were observed with different charge methods. However, their study showed that the observed electrostatic fields were greatly affected by the calculational methods. Kroemer *et al.*³ analyzed 17 different methods at three different levels of theory to

* e-mail: bruns@iqm.unicamp.br

calculate charges and their effects on CoMFA results. Gasteiger-Marsili, semi empirical MNDO, AM1 and PM3) and ab-initio (HF/STO-3G, HF/3-21G* and HF/6-31G*) charges were included. The ESPFIT-derived charges yielded better models than those based on charges calculated from Mülliken population analyses. However the simple Gasteiger-Marsilii charges did not give the worst model.

For 5-HT₃ receptor binding affinity of 39 piperazinopyrrolo-thieno-pyrazines, Bureau *et al.*⁴ compared the CoMFA results obtained with calculated charges computed from electrostatic potentials, quantum mechanically calculated charges using 6-31G* basis and Gasteiger-Hückel charges. The CoMFA model obtained with Gasteiger-Hückel charges was inferior.

For angiotensin converting enzyme (ACE) and thermolysin inhibitors, Waller *et al.*⁵ studied the effect of charges calculated by Gasteiger-Hückel and PM3 methods. In the ACE inhibitor series, the two methods gave nearly similar models. For thermolysin inhibitors, a better model (internal predictivity) was observed using Gasteiger-Hückel charges, but the PM3 method gave better external predictivity for 11 test compounds. For non-steroidal aromatase inhibitors related to fadrozole, Recanatini⁶ observed similar models using either the AM1 method for geometry optimization and charge calculations or MAXIMIN2 molecular mechanics for geometry optimization and Gasteiger-Marsili charges. Navajas *et al.*⁷ verified that the mutagenic activity of 16 5H-furan-2-one derivatives was correlated with the LUMO field. The MNDO, PM3 and AM1 Hamiltonians were used to optimize and generate the LUMO field. Only the AM1 and PM3 methods gave satisfactory CoMFA models.

More recently the descriptive power of MS-WHIM, classified as a global 3D-QSAR method to model specific biological interactions, was described and the dependence of MS-WHIM on the type of atomic charge used to compute the electrostatic potential was analyzed. They observed that MS-WHIM descriptors were sensitive to the type of partial atomic charges applied and improved models were obtained using more accurate charges.⁸

CoMFA models derived for artemisinin derivatives, using semiempirical AM1 and HF/3-21G optimized geometries, revealed that the HF/3-21G method was found to be usually but not drastically better than AM1. Additional calculations were performed to investigate the electrostatic field difference using the Gasteiger and Marsili charges, the electrostatic potential fit charges at the AM1 level, and the natural population analysis charges at the HF/3-21G level of theory. For the HF/3-21G optimized structures no difference in predictability was observed,

whereas for the AM1 optimized structures differences were found.⁹

In addition to these problems it is not always clear that the atomic charge is the most appropriate descriptor to be used in QSAR studies. This is particularly true for problems involving hydrogen bonding. Galabov and Bobadova-Parvanova^{10,11} have recently performed theoretical studies at the HF/6-31G(d,p) and HF/6-31+G(d,p) levels showing that the energy of hydrogen bond formation is not highly correlated with the atomic charge of the H-bonding site in the acceptor molecule but rather to the electrostatic potential at this site. For hydrogen bonding of HF to some open chain carbonyl and nitrile compounds the H-bonding energy was not found to be systematically related to the atomic charges of the oxygen or nitrogen acceptor atoms but was linearly related to the electrostatic potentials at these nuclei. An absence of a precise linear relation was found for several types of atomic charges, Mulliken,¹² CHELPG¹³ and MK.¹⁴ This is particularly significant since the charges and potentials were calculated from the same wave functions.

In this paper we investigate the relation between the H-bonding energy and GAPT¹⁵ atomic charges. Even though GAPT charges can be determined experimentally from infrared intensities and also calculated from molecular wave functions¹⁶ their relation to H-bonding has not yet been investigated. Furthermore the GAPT charges have been shown to be related to atomic electrostatic potentials by Siegbahn's simple potential model.^{17,18} These relationships suggest the use of Koopmans' energies for core electrons as QSAR descriptors in activity problems for which the H-bonding phenomenon is important.

Calculations

H-bonding energies for the hydrogen fluoride molecule with the HCOH, HCOOH, HCOSH, HCOOCH₃, HCONH₂, HCONO₂, HCOCN, HCOF, HCOCl, HCOCH₃ and HCOCF₃ molecules were taken from references 10 and 11 and are reproduced in Table 1. These values were calculated at the HF/6-31G(d,p) level and for this reason our GAPT charges and Koopmans' energies were calculated using the same wave functions. All calculations were performed at the theoretical equilibrium geometries of these molecules using the Gaussian 98 program²⁰ on a DEC ALPHA 1000 work station.

The molecular electrostatic potentials at point x in space, where x refers to the Cartesian coordinates of the acceptor nucleus in the molecule are standard output of the Gaussian program and can be calculated by the equation:¹¹

Table 1. HF/6-31G(d,p) *ab initio* calculated energy of hydrogen bond formation, GAPT charge, Koopmans' energy and molecular electrostatic potential of the carbonyl oxygen in the isolated molecule.

Molecules	$E_{\text{H bond}}$ (kcal mol ⁻¹) ^a	$q_{\text{O}}^{\text{GAPT}}$ (e) ^b	E_{Koop} (eV) ^b	V_x (eV) ^a
HCOH	-3.7899	-0.673	560.09	-606.77
HCOOH	-4.3727	-0.835	559.67	-607.16
HCOSH	-3.7434	-0.822	560.26	-606.63
HCOOCH ₃	-4.9121	-0.813	559.45	-607.38
HCONH ₂	-6.5051	-0.903	558.65	-608.16
HCONO ₂	-1.4811	-0.608	562.25	-604.67
HCOCN	-2.2123	-0.653	561.48	-605.43
HCOF	-2.7607	-0.745	560.61	-606.26
HCOCI	-2.3317	-0.778	561.12	-605.79
HCOCH ₃	-4.9560	-0.740	559.61	-607.24
HCOCF ₃	-2.5187	-0.626	561.15	-605.74

^aValues taken from Tables 1 and 3 of Reference 1.

^bValues calculated in our laboratory using the Gaussian 98 program.

$$V_x = V(\mathbf{R}_x) = \sum_{A \neq x} \frac{Z_A}{|\mathbf{R}_x - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{R}_x - \mathbf{r}'|} d\mathbf{r}' \quad (1)$$

The GAPT charges are mean dipole moment derivatives given by

$$\bar{p}_\alpha = \frac{1}{3} \left(\frac{\partial p_x}{\partial x_\alpha} + \frac{\partial p_y}{\partial y_\alpha} + \frac{\partial p_z}{\partial z_\alpha} \right) \quad (2)$$

where α represents an atom in a molecule. In other words the GAPT charges are one third of the trace of the atomic polar tensor:^{20,21}

$$\mathbf{P}_X^\alpha = \begin{pmatrix} \frac{\partial p_x}{\partial x_\alpha} & \frac{\partial p_x}{\partial y_\alpha} & \frac{\partial p_x}{\partial z_\alpha} \\ \frac{\partial p_y}{\partial x_\alpha} & \frac{\partial p_y}{\partial y_\alpha} & \frac{\partial p_y}{\partial z_\alpha} \\ \frac{\partial p_z}{\partial x_\alpha} & \frac{\partial p_z}{\partial y_\alpha} & \frac{\partial p_z}{\partial z_\alpha} \end{pmatrix} \quad (3)$$

The molecular polar tensor is a juxtaposition of the atomic polar tensors

$$\mathbf{P}_X = \left\{ \mathbf{P}_X^{(1)} \mathbf{P}_X^{(2)} \dots \mathbf{P}_X^{(N)} \right\} \quad (4)$$

where there are N atoms in the molecule and is calculated from

$$\mathbf{P}_X = \mathbf{P}_Q \mathbf{L}^{-1} \mathbf{U} \mathbf{B} + \mathbf{P}_\rho \boldsymbol{\beta} \quad (5)$$

where \mathbf{L}^{-1} , \mathbf{U} and \mathbf{B} are well known transformation matrixes²² used in molecular vibrational analysis and \mathbf{P}_Q contains the dipole moment derivatives with respect to the normal

coordinates. These latter elements are proportional to the square roots of the experimental infrared intensities.²³ The GAPT charges are automatically calculated with the Gaussian program when the FREQ option is used.

Results

Galabov and Parvanova¹⁰ have shown that the energy of hydrogen bond formation calculated at the HF/6-31G(d,p) level is not highly correlated with the atomic charge on the oxygen atom of the carbonyl group of the isolated acceptor molecule. Regressions of this energy on Mulliken, CHELPG and MK charges for the oxygen atoms of this group of molecules resulted in relatively low coefficients of determination, $R^2 = 0.867$, 0.867 and 0.837 . In Figure 1 the graph of the energy of hydrogen bond formation as a function of the GAPT atomic charges on the oxygen atoms is presented. The energies were taken from reference 10 and the GAPT charges were calculated in our laboratory and are presented in Table 1. The scatter of the points about the regression line, included in the graph, is quite large and reflects the quite low coefficient of determination obtained, 0.602 . The slope of the regression line is positive, as expected, since more negative oxygen acceptor charges are expected to result in more stable hydrogen bonds.

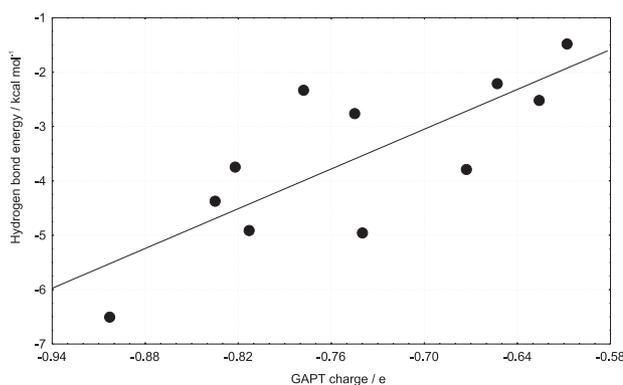


Figure 1. Graph of HF/6-31G(d,p) calculated hydrogen bond formation energy against GAPT charges of the carbonyl oxygen in the isolated acceptor molecules.

Galabov and Bobadova-Parvanova¹⁰ reported that a much higher correlation coefficient is obtained when the energy of hydrogen bond formation is graphed against the molecular electrostatic potential at the carbon atom in the isolated acceptor molecules. The graph can be seen in reference 10 and the corresponding regression has an R^2 value of 0.958 , much larger than the values observed for the various types of atomic charges calculated to pertain to the oxygen atom in the isolated acceptor molecules.

The molecular electrostatic potential at an atom can be accurately estimated by Koopmans' energies of electrons occupying core orbitals. This is clearly illustrated in Figure 2 where a graph of Koopmans' energies for the $1s$ electrons of the oxygen atoms for the group of carbonyls investigated here is graphed against the corresponding molecular electrostatic potentials. The numerical values used have been included in Table 1. This graph shows an almost perfect correlation and its regression has an R^2 value of 1.000.

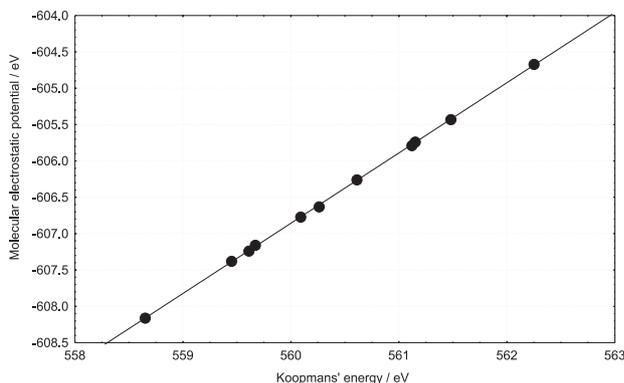


Figure 2. Graph of HF/6-31G(d,p) calculated molecular electrostatic potentials and $1s$ electron Koopmans' energies of the carbonyl oxygen in the isolated acceptor molecules.

As such one can expect that the energy of hydrogen bond formation is highly correlated with Koopmans' energy. This graph is presented in Figure 3 and corresponds to a regression R^2 value of 0.951, and is very similar to the one reported by Galabov and Bobadova-Parvanova¹⁰ for the electrostatic potential. The small difference in the coefficient of determination could come from the fact that besides investigating the HCOR molecules treated in this work Galabov and Bobadova-Parvanova also studied their methyl analogues, CH_3COR .

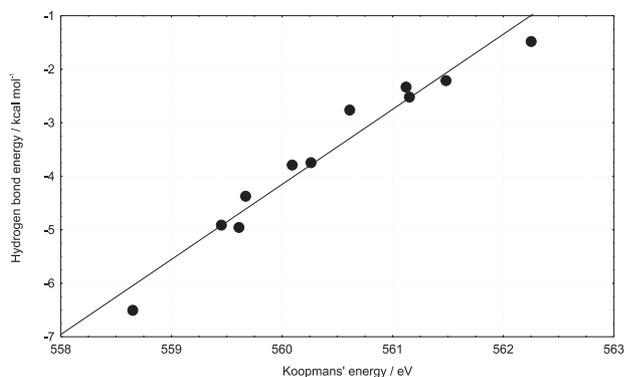


Figure 3. Graph of HF/6-31G(d,p) calculated hydrogen bond formation energy against Koopmans' energies of the $1s$ electron of the carbonyl oxygen in the isolated acceptor molecules.

Discussion

It is not surprising that hydrogen bond formation energies are highly correlated with the molecular electrostatic potentials at the acceptor atoms or with Koopmans' energies of the $1s$ electron on the acceptor atoms. The energy of hydrogen bond formation consists of moving a positive hydrogen atom with the rest of the donor molecule from infinity and placing the hydrogen atom *close* to the oxygen nucleus. One can idealize this process by moving a positive test charge from infinity and placing it on the oxygen nucleus. If the test charge does not affect the molecular environment, this is just the molecular electrostatic potential at the acceptor atom. This explains why the molecular electrostatic potential approximates hydrogen bond formation and their values are expected to be correlated. Koopmans' ionization of a $1s$ electron is a process almost exactly opposite to the one used for determining the molecular electrostatic potential. Instead of adding a positive charge a negative charge is removed. As such it is also highly correlated with the energy of hydrogen bond formation.

These relations are only approximate for two reasons. Hydrogen bond formation affects the electron distribution of the acceptor molecule. This, of course, is not the case for molecular electrostatic potentials and Koopmans' energies of the acceptor molecule. Also the latter quantities, as used in this study, are measured at the acceptor atom. The proton involved in hydrogen bonding is close to the acceptor atom but not on it. Close means the hydrogen bond distance. Since these distances are approximately the same for all hydrogen bonding pairs treated here (between 177 and 222 pm according to the theoretical calculations¹⁰) a good correlation between the hydrogen bond formation energy and these two quantities can be expected. The scatter of points about the regression line in Figure 3 could be due, in part, to the different hydrogen bonding distances. However it must be remembered that the molecular electrostatic potentials and Koopmans' energies provide extremely simple models for explaining the much more complicated hydrogen bonding process.

It is also simple to understand why the molecular electrostatic potentials and Koopmans' energies are more highly correlated with the hydrogen bonding energies than the atomic charges. This is more conveniently explained using the GAPT charges. Siegbahn and co-workers¹¹ have shown that Koopmans' energy can be calculated using a simple potential model represented by the equation:

$$E_{O,1s}^{Koop} = kq_O + \sum_{\alpha \neq O} \left(\frac{q_\alpha}{R_{\alpha O}} \right) = kq_O + V \quad (6)$$

Here $E_{O,1s}^{Koop}$ is Koopmans' energy of a 1s electron of the carbonyl oxygen, k is a constant, q_O is the charge on the oxygen atom, q_α the charge on a neighboring atom and $R_{\alpha O}$ the distance between the oxygen and its neighboring α^{th} atom. If this potential model is applied to experimental ESCA energies a relaxation energy term must be added on the right hand side of this equation. This relation has been shown to be valid for both experimental 1s electron ionization energies and Koopmans' energies if GAPT charges obtained from experimentally measured infrared intensities or calculated from molecular wave functions are used.^{17,18} Evidently, the formation energy of the hydrogen bond and GAPT oxygen charges are not highly correlated because the electrostatic potential at the oxygen atom contributed by the neighboring atoms is also important in the hydrogen bonding process. In other words an approaching donor molecule is influenced not only by the charge of the acceptor oxygen atom but also by the charges on the other atoms in the molecule. Analogous arguments could be made for relations between other types of atomic charges and the hydrogen bonding energies. However in those cases the Siegbahn simple potential is not as accurate as for the GAPT charges.

Recommendations

QSAR studies have often been carried out by regressing biological activities on sets of molecular descriptors that often include atomic charges calculated from quantum chemical wave functions. If the main result of this paper, that Koopmans' energies are more highly correlated than atomic charges to hydrogen bond formation energies, can be extended to other groups of molecules besides those investigated here it seems reasonable to use these energies as descriptors for biological processes for which hydrogen bonding is important. Their use instead of electrostatic potentials has several advantages. First, ionization energy processes are more familiar to chemists than the more abstract molecular potential concept. Second, their use instead of atomic charges eliminates two uncertainties in QSAR applications. Which electron correlation level should be used to calculate the molecular wave function? What is the most appropriate way (Mulliken, CHELPG, GAPT, etc.) of determining the atomic charges from this wave function? Since Koopmans' energies are defined at the Hartree-Fock electron correlation level, really significant savings in computer time are also desirable consequences. One could also use experimental Koopmans' energies as QSAR descriptors. However these quantities are only available for molecules in the gas and

solid phases. Furthermore they reflect not only the initial electronic structure of the molecule but also its relaxation properties on core electron ionization that are probably not relevant to biological problems.

Acknowledgements

The authors acknowledge partial financial support from FAPESP, the State of São Paulo Research Foundation, and CNPq, the Brazilian National Research Council.

References

1. Karelson, M.; Lobanov, V.S.; Katritzky, A.R.; *Chem. Reviews*, **1996**, *96*, 1027.
2. Folkers, G.; Merz, A.; Rognan, D. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H. ed., ESCOM: Leiden, 1993, 583.
3. Kroemer, R.T.; Hecht, P.; Liedl, K.R.; *J.Comput.Chem.* **1996**, *17*, 1296.
4. Bureau, R.; Rault, S.; Pilo, J.C.; Robba, M.; *Quant. Struct.-Act.Relat.*, **1996**, *15*, 373.
5. Waller, C.L.; Marshall, G.R.; *J.Med.Chem.* **1993**, *36*, 2390.
6. Recanatini, M.; *J.Comput.-Aid.Mol.Design* **1996**, *10*, 74.
7. Navajas, C.; Poso, A.; Tuppurainen, K.; Gynther, J.; *Quant. Struct.-Act.Relat.* **1996**, *15*, 189.
8. Gancia, E.; Bravi, G.; Mascagni, P.; Zaliani, A.; *J.Comput.-Aid.Mol.Design* **2000**, *14*, 293.
9. Tonmunphean, S.; Kokpol, S.; Parasuk, V.; Wolschann, P.; Winger, R.H.; Liedl, K.R.; Rode, B.M.; *J.Comput.-Aid.Mol.Design* **1998**, *12*, 397.
10. Bobadova-Parvanova, P.; Galabov, B.; *J. Phys. Chem. A* **1998**, *102*, 1815.
11. Galabov, B.; Bobadova-Parvanova, P.; *J. Phys. Chem. A* **1999**, *103*, 6793.
12. Mulliken, R. S.; *J. Chem. Phys.* **1955**, *23*, 1833.
13. Breneman, C. M.; Wiberg, K. B.; *J. Comput. Chem.* **1990**, *11*, 361.
14. Besler, B. H.; Merz, K. M.; Kollman, P. A.; *J. Comput. Chem.* **1990**, *11*, 431.
15. Cioslowski, J.; *J. Am. Chem. Soc.* **1989**, *111*, 8333.
16. de Oliveira, A. E.; Haiduke, R. L. A.; Bruns, R. E.; *J. Phys. Chem. A* **2000**, *104*, 5320.
17. Guadagnini, P. H.; de Oliveira, A.E.; Bruns, R. E.; Neto, B. B.; *J. Am. Chem. Soc.* **1997**, *119*, 4224.
18. Haiduke, R. L. A.; de Oliveira, A. E.; Bruns, R. E.; *J. Phys. Chem. A* **2002**, *106*, 1824.
19. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, V. G.; Zakrewski, J. A.; Montgomery Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millasm, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.;

- Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, P. Y.; Ayala, Q.; Cui, K.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; *Gaussian 98 (Revision A.7)*, Gaussian Inc.: Pittsburg, PA, 1998.
20. Biarge, J. F.; Herranz, J.; Morcillo, J.; *J. An. R. Soc. Esp. Fis. Quim.* **1961**, *A57*, 81.
21. Person, W. B.; Newton, J. H.; *J. Chem. Phys.* **1974**, *61*, 1040.
22. Wilson, E. B.; Decius, J. C.; Cross, P. C.; *Molecular Vibrations*; McGraw-Hill: New York, 1955.
23. Overend, J. In *Infrared Spectroscopy and Molecular Structure*; Davies, M., ed., Elsevier: New York, 1963, ch. 10.

Received: April 11, 2002

Published on the web: October 31, 2002

FAPESP helped in meeting the publication costs of this article.