

NIH Public Access

Author Manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2012 December 10.

Published in final edited form as:

J Chem Theory Comput. 2011 April 12; 7(4): 852–856. doi:10.1021/ct1006373.

Polarized Molecular Orbital Model Chemistry. I. Ab Initio Foundations

Luke Fiedler, **Jiali Gao**, and **Donald G. Truhlar**

Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street S.E., Minneapolis, MN 55455-0431

Jiali Gao: jiali@jialigao.org; Donald G. Truhlar: truhlar@umn.edu

Abstract

The objective of this paper is to examine the minimal requirements for obtaining semiquantitative polarizabilities of molecules, in order to provide a well-founded starting point for a new semiempirical molecular orbital formulation that is more suitable than presently available methods for simulating electronic polarization effects. For this purpose, we present polarizability calculations for 38 molecules with 36 basis sets, including many unconventional ones, and five semiempirical molecular orbital theories based on neglect of diatomic differential overlap. We conclude that two basis sets are particularly promising to serve as bases for semiempirical improvement, namely STO-3G(,P), in which diffuse p functions are added to all hydrogens, and 3- (21,3,21)G in which a minimal basis set is augmented with one extra s function on every atom. We especially recommend the former because all intra-atomic overlap integrals are zero by symmetry, which makes it a better candidate for neglect-of-differential-overlap treatments.

1. Introduction

Semiempirical molecular orbital theory with the neglect of diatomic differential overlap¹ (NDDO, e.g., MNDO,² AM1,³ PM3,⁴ and RM1⁵) has been enormously useful, and it continues to be widely used in applications to macromolecular systems, although it is not without known deficiencies. One of the chief deficiencies is that it does not describe intermolecular interactions very accurately, partially due to underestimation of molecular polarizabilities.6,7 This deficiency is a major limiting factor for quantitative simulation and modeling of systems of biological and materials interest. As an example of trying to remedy the underpolarization of NDDO methods, Schürer et al. 8 reported a parametrized variational method for calculating molecular electronic polarizabilities from an NDDO wave function; that method, however, does not concern the polarization of the electronic wave function itself and hence it is unsuitable for modeling polarization in direct dynamics calculations. Our goal is to include polarization in the model wave function itself to provide a framework for developing a new generation of force fields for large molecules in condensed phases, where the new force fields go beyond fixed-charge molecular mechanics by allowing for self-consistent polarization and charge transfer.⁹

One can make up for "small" qualitative deficiencies in molecular model chemistries by parameterization if one uses an appropriate theoretical framework, but it is hard and dangerous to use parametrization to make up for large quantitative deficiencies or for a framework that does not contain the dominant physical factors. In the latter case, it is

Correspondence to: Jiali Gao, jiali@jialigao.org; Donald G. Truhlar, truhlar@umn.edu.

Supporting Information Available: A complete set of tables of polarizability calculations and geometries optimized at the MP2 level. This material is available free of charge via the Internet at<http://pubs.acs.org>.

desirable to improve the form of the model so the deficiencies that must be overcome are smaller and the model to be parameterized contains the physical features that correctly account for the phenomenon to be modeled.

Current NDDO molecular orbital theory for elements in the 1s, 2s, and 2p blocks of the periodic table usually has the form of Hartree-Fock (HF) theory employing a minimum basis set (MBS) with many integrals systematically neglected and others parameterized rather than evaluated as in *ab initio* calculations.¹ Since the parametrization is carried out against experimental data rather than against a correct HF/MBS calculation, not only does it make up for the neglected integrals, but also it empirically introduces higher-order effects, in particular some of the effects of using a more complete (more polarizable) basis set and some of the effects of electron correlation. Using the STO-3G¹⁰ minimum basis set, *ab initio* Hartree-Fock theory underestimates the mean polarizability of a water molecule by a factor of 3.6, which may be compared to a factor of 3.4 for MNDO and a factor of 2.9 for AM1, PM3, and RM1. Thus, these semiempirical parametrizations only improved the polarizability marginally. When we tried to reparameterize these models to have greater polarizability, many of the other properties became significantly worse, which indicates a fundamental limitation in building the model on the HF/MBS framework. An interesting question is whether there is an *ab initio* formalism, not much more complicated than HF/ MBS, that has a significantly reduced error in computed molecular polarizability, e.g., less than a factor of two (as compared to a factor of 3.4). An NDDO version of such a theory might prove easier to parameterize satisfactorily than the NDDO version of HF/MBS theory. For example, should we explicitly include electron correlation? We already know from accumulated experience with ab initio calculations that this is not necessary, i.e., large-basisset HF theory gives realistic (although of course not quantitatively accurate) polarizabilities. But how much larger a basis set is called for? What is the smallest increase in basis set that yields qualitatively correct polarizabilities? The present article is devoted to this question. Paper II ,¹¹ which immediately follows this paper, builds on the findings of this paper to develop a new NDDO framework, which we call the polarized molecular orbital (PMO) model, that yields reasonable results for polarizabilities.

Section II of the present paper presents a database that will be used to test basis sets for polarizabilities. Section III then presents Hartree-Fock polarizability calculations with 37 different basis sets.

II. Database

To understand the basis set requirements for predicting realistic polarizabilities, we consider only mean dipole polarizabilities α defined as

$$
\alpha = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3} \quad (1)
$$

where a_{ij} is an element of the electric dipole polarizability tensor. The mean polarizability is the trace of the polarizability tensor and is independent of the orientation of the axes. In the rest of this article and the next one we will simply call α the polarizability.

Table 1 gives the 38 polarizabilities that we will use to test basis sets. In most cases these are taken from experiment; 1^{2-21} in a few cases where we did not have experimental values we calculated the polarizability by MP2/aug-cc-pVTZ//MP2/aug-cc-pVDZ, by MP2/aug-ccpVDZ, or by HF/aug-cc-pVDZ, where HF denotes Hartree-Fock, MP2 denotes Møller-Plesset $2nd$ order perturbation theory,²² and we use the usual notation²³ for correlationconsistent basis sets.

III. Basis sets

The basis sets considered are in Table 2. For most of the basis sets, the notation is standard. Nonstandard basis sets are explained in the table; in these explanations, exponential parameters are called exponents and are given in a_0^{-2} . All basis sets are single-zeta for the core orbitals; for valence electrons we consider both regular valence basis functions (indicated by lower case s, p, or d) and diffuse basis functions (indicated by capital S, P, or D). The table shows how many basis functions of each type (excluding core basis functions) are present in a given basis; N_1 is the number of basis functions on each hydrogenic atom, and N_2 is the number of basis functions (excluding the core) on each nonhydrogenic atom. (To obtain the total number of basis functions, add one core basis function for C, N, and O and 5 core basis functions for S.)

IV. Results

Polarizabilities for three typical molecules are given in Table 3. We give results for all 36 basis sets of Table 2 plus for five NDDO methods: MNDO, 2 AM1, 3 PM3, 4 RM1, 5 and PM6.24 A complete table of polarizability calculations for all 38 molecules is presented in supporting information. Table 3 also gives the mean unsigned percentage error (MUPE, that is, the mean of the absolute values of all 38 percentage deviations of the calculated polarizabilities from the reference values of Table 1). The polarizability tensor components of two molecules, water and acetamide, are given in Table 4.

V. Discussion

The final five rows of table 3 confirm that presently available semiempirical methods underestimate molecular polarizabilities, but comparison to the results using STO-3G shows that they have smaller errors than *ab initio* Hartree–Fock calculations with the same-size basis set. The other results in the table explore the effect of expanding the basis set.

The table shows that adding diffuse P functions on hydrogen atoms is particularly effective in increasing the polarizabilities. For example, the STO-3G(,P) basis set has a smaller MUPE than the much larger STO-3G(d), 3-21G, 3-21G(d), or $3\n-21G(p)$ basis sets, and the error is almost as small as for the well-polarized cc-pVDZ basis set or the 3-21+G(d) basis set, which contains more contracted functions than the cc-pVDZ basis set. (We note that extending NDDO methodology to the 4-31G basis set, which is similar to the 3-21G basis set, has already been considered by Thiel;²⁵ however, results for polarizabilities were not presented.)

Deleting the diffuse P functions on any hydrogens, even deleting them only on sp^3 carbon atoms raises the MUPE appreciably, as shown by the results for STO-3G(,P*).

The introduction of p orbitals on hydrogen atoms was previously considered in semiempirical molecular orbital theory most thoroughly by Jug and Geudtner, 26 who added p orbitals to hydrogen atoms in the SINDO1 approximation; their goal was to improve the treatment of hydrogen bonding. In the present work, it was found, through systematic investigation of basis set dependence, that adding P functions (that is, in the notation established in Section III, diffuse p functions) on hydrogen to a minimum basis set in ab initio Hartree-Fock calculations provides a powerful strategy to calculate more accurate polarizabilities. the mean unsigned percentage errors is 32% for adding a P subshell but 56% for adding a p subshell. Adding both further reduces the error, but only to 31%, showing that the diffuse P function is the key to the success of this strategy. One obtains a similar mean error, in particular 33%, with either the $3-21G(p)$ or $3-21G(d)$ basis set. Taking acetylene as an example, Table 2 shows that the number of valence basis functions in the STO-3G(,P)

basis set is 16, whereas the number in the $3-21G(p)$ or $3-21G(d)$ is 26–30. For a water molecule these numbers are reduced to 12 for $STO-3G(P)$ and to $17-18$ for $3-21G(p)$ or 3-21G(d).

Another strategy that may be compared to the successful STO-3G(,P)strategy is to add a d subshell to the STO-3G basis every nonhydrogenic atom, yielding the STO-3G(d) basis set. This might have been anticipated to be a powerful strategy (the natural choice) because most of the electrons in the molecules considered here are in valence p orbitals and these are strongly coupled to d functions by the electric dipole operator. Using this strategy raises the number of valence basis functions to 20 for acetylene and lowers it to 11 for water, but it yields a mean error of 57%. Thus this strategy is less successful.

A second successful strategy revealed by Tables 2 and 3 is to split the valence s subshell of STO-3G on all atoms. If one were to split all the valence subshells, that would give 3-21G, but the splitting only s subshells gives 3-(21,3,21)G. The mean error is 29%, and the number of valence basis functions for the example case of acetylene is only 14. For a water molecule this number is reduced to 9. Of the two successful strategies, i.e., (a) splitting the valence s subshell on all atoms and (b) adding a set of diffuse P functions to hydrogen atoms, the latter is preferred for two reasons. First, the inclusion of a set of P orbitals allows the out-of-plane polarizability as well as the in-plane polarizability to be better represented, which is particularly important for key compounds such as water and benzene or any other planar molecule. Second, when the strategy is applied in the context of the NDDO approximation, choice (b) avoids the treatment of nonorthogonal basis functions on the same center, which is probably very important since neglect of one-center differential overlap has been singled out as a significant shortcoming of the current NDDO formalisms, even with minimal basis sets.27,28 We examine the use of the STO-3G(,P) basis as a starting point for NDDO parametrization in the following article.

The polarizability tensor components for water and acetamide are provided in Table 4, with axes aligned along the principal axes. For comparison, we also give results calculated with the same axis choices with the M06-2X density functional²⁹ and ab initio MP2 wave function theory²² (with standard basis sets^{23,30}). Water is in the yz plane, and acetamide has the yz plane as a plane of C_s symmetry. By comparing the results with the STO-3G basis set to those with the STO-3G(,P) basis and 3-(21,3,21)G bases sets, the table shows that both new basis sets allow out-of-plane polarization, and for the larger molecule the error in outof-plane polarizability is less than 30%.

VI. Conclusions

We found that adding only a few well chosen extra basis functions can considerably reduce the errors in minimum-basis-set calculations of electric dipole polarizabilities. Two particularly efficient strategies are (i) to add a diffuse P subshell to every hydrogen atom and (ii) to split the valence s shell on each atom into a double zeta pair of s functions. It is particularly encouraging that these strategies work better than the seemingly more obvious choice of adding d functions to nonhydrogenic atoms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institutes of Health (grant no. RC1-GM091445) and the National Science Foundation (grant no. CHE09-56776).

References

- 1. Pople JA, Santry DP, Segal GA. J Chem Phys. 1965; 43:S129.
- 2. Dewar MJS, Thiel W. J Am Chem Soc. 1977; 99:4899.
- 3. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP. J Am Chem Soc. 1985; 107:3902.
- 4. Stewart JJP. J Comput Chem. 1989; 10:209.
- 5. Rocha GB, Freire RO, Simos AM, Stewart JJP. J Comp Chem. 2006; 27:1101. [PubMed: 16691568]
- 6. Dewar MJS, Yamaguchi Y, Suck S. Chem Phys Lett. 1974; 59:541.
- 7. Parkinson WA, Zerner MC. J Chem Phys. 1991; 94:478.
- 8. Schürer G, Gedeck P, Gottschalk M, Clark T. Int J Quantum Chem. 1999; 75:17.
- 9. Gao J. J Phys Chem B. 1997; 101:657.Gao J. J Chem Phys. 1998; 109:2346.Xie W, Gao J. J Chem Theory Comput. 2007; 3:1890. [PubMed: 18985172] Xie W, Song L, Truhlar DG, Gao J. J Chem Phys. 2008; 128:234108. [PubMed: 18570492] Xie W, Song L, Truhlar DG, Gao J. J Phys Chem B. 2008; 112:14124. [PubMed: 18937511] Song L, Han J, Lin YL, Xie W, Gao J. J Phys Chem A. 2009; 113:11656. [PubMed: 19618944] Xie W, Orozco M, Truhlar DG, Gao J. J Chem Theory Comput. 2009; 5:459. [PubMed: 20490369] (h) Cembran, A.; Bao, P.; Wang, Y.; Song, L.; Truhlar D., G.; Gao, J., work in progress.
- 10. Hehre WJ, Stewart RF, Pople JA. J Chem Phys. 1969; 51:2657.
- 11. Zheng P, Fiedler L, Leverentz H, Truhlar DG, Gao J. J Chem Theory Comput. following paper.
- 12. (a) Stuart, HA. Landolt-Börnstein "Zahlenwerte und Funktionen". Eucken, A.; Hellwege, KH., editors. Springer-Verlag: berlin; 1951. p. 511(b) Maryott, AA.; Buckley, F. U S National bureau of Standards Circular No 537. U. S. Government; Washington: 1953.
- 13. Hirschfelder, JO.; Curtiss, CF.; Bird, RB. Molecular Theory of Gases and Liquids. Wiley; New York: 1954. p. 950
- 14. Bridge NJ, Buckingham AD. Proc Roy Soc (London). 1966; 295:334.
- 15. Sutter H, Cole RH. J Chem Phys. 1970; 52:132.
- 16. Bose TK, Cole RH. J Chem Phys. 1970; 52:140.
- 17. Bose TK, Cole RH. J Chem Phys. 1971; 54:132.
- 18. Applequist J, Carl JR, Fung KK. J Am Chem Soc. 1972; 94:2952.
- 19. Bose TK, Sochanski JS, Cole RH. J Chem Phys. 1972; 57:3592.
- 20. Marchese FT, Jaffe HH. Theor Chim Acta. 1977; 45:241.
- 21. No KT, Cho KH, Jhon MS, Scheraga HA. J Am Chem Soc. 1993; 115:2005.
- 22. Pople JA, Binkley JS, Seeger R. Int J Quantum Chem Symp. 1976; 10:1.
- 23. Kendall RA, Dunning TH Jr, Harrison RJ. J Chem Phys. 1992; 96:6796.
- 24. Stewart JJP. J Mol Model. 2007; 13:1173. [PubMed: 17828561]
- 25. Thiel W. Theor Chim Acta. 1981; 59:191.
- 26. Jug K, Geudtner G. J Comp Chem. 1993; 14:639.
- 27. Thiel W. Theor Chem Acc. 2000; 103:495.
- 28. Winget P, Selcuki C, Horn AHC, Martin B, Clark T. Theor Chem Acc. 2003; 110:254.
- 29. Zhao Y, Truhlar DG. Theor Chem Acc. 2008; 120:215.
- 30. Clark T, Chandrasekhar J, Spitznagel GW, Schleyer PvR. J Comp Chem. 1983; 4:294.

Molecule α **(A³**

Table 1

Polarizabilities (Å³)

) Ref.

Fiedler et al. Page 7

a present work, calculated by MP2/aug-cc-pVTZ//MP2/aug-cc-pVDZ.

 b present work, calculated by HF/aug-cc-pVDZ.

 c present work, calculated by MP2/aug-cc-pVDZ.

 \$watermark-text \$watermark-text

\$watermark-text

\$watermark-text

Table 2
Basis sets and numbers of regular valence basis functions and diffuse basis functions **Basis sets and numbers of regular valence basis functions and diffuse basis functions**

 α exponent on hydrogen P orbital = 0.141, taken from aug-cc-pVDZ. Note that we use capital S, P, or D to denote diffuse basis functions, thereby distinguishing them from regular valence basis functions exponent on hydrogen P orbital = 0.141, taken from aug-cc-pVDZ. Note that we use capital S, P, or D to denote diffuse basis functions, thereby distinguishing them from regular valence basis functions (indicated by lower case s, p, or d). (indicated by lower case s, p, or d).

b exponent on hydrogen p orbital = 0.727, taken from cc-pVDZ. exponent on hydrogen p orbital = 0.727, taken from cc-pVDZ.

 $\overset{\circ}{c}$ exponents for diffuse functions on non-hydrogenic atoms taken from a
ug-cc-pVDZ exponents for diffuse functions on non-hydrogenic atoms taken from aug-cc-pVDZ

J Chem Theory Comput. Author manuscript; available in PMC 2012 December 10.

 d exponent for non-hydrogenic d functions = 0.8 exponent for non-hydrogenic d functions = 0.8

combination of STO-3G and 3-21G. In 3- (x, y, z) G: x denotes valence basis for nonhydrogenic s orbitals, y denotes basis for nonhydrogenic p orbitals, and z denotes basis for hydrogen. combination of STO-3G and 3-21G. In $3-(x, y, z)$ G: x denotes valence basis for nonhydrogenic s orbitals, y denotes basis for nonhydrogenic p orbitals, and z denotes basis for hydrogen.

 \hat{P} functions only on polar hydrogens, i.e., hydrogens not bonded to carbon. P functions only on polar hydrogens, i.e., hydrogens not bonded to carbon.

 ${}^{\mathcal{E}}\!P$ functions only on polar hydrogens and hydrogens bonded to non-sp ${}^{S}\!P$ functions only on polar hydrogens and hydrogens bonded to non-sp³ carbons.

 h exponent on hydrogen P orbital = 0.123, which was obtained by optimizing it with respect to the polarizabilities of the 38-mlecule test set. exponent on hydrogen P orbital $= 0.123$, which was obtained by optimizing it with respect to the polarizabilities of the 38-mlecule test set.

 i exponent on hydrogen P orbital = 0.082 . exponent on hydrogen P orbital $= 0.082$.

 \$watermark-text \$watermark-text

\$watermark-text

\$watermark-text

Table 3 Polarizabilities and mean unsigned percentage errors in polarizabilities (Å3

)

J Chem Theory Comput. Author manuscript; available in PMC 2012 December 10.

Fiedler et al. Page 11

\$watermark-text

\$watermark-text

 \$watermark-text \$watermark-text **Table 4**

)

\$watermark-text

\$watermark-text

Polarizability tensor components of water and acetamide (Å3

