Prediction of Nonlinear Optical Responses of Organic Compounds

Kjell Öberg, Anders Berglund,* Ulf Edlund, and Bertil Eliasson

Department of Chemistry, Organic Chemistry, Umeå University, 901 87 Umeå, Sweden,

Received October 25, 2000

The nonlinear optical quantities, second and third harmonics (β and γ), were predicted using a quantitative structure−property relationship (QSPR) approach. Molecular orbital ab initio calculations were applied to generate easily accessible variables to be used in the partial least-squares analysis. Simplified equations are presented that could be used to predict the experimental β and γ responses, prior to further investigations of potentially interesting molecules for use in optical materials.

INTRODUCTION

The nonlinear optical (NLO) responses induced in various molecules in solution and solids are of great interest in many fields of research.[1,2] These phenomena occur when the optical properties of molecules change in the presence of strong external electrical fields, i.e., high-energy laser beams. Attempts are being made to incorporate organic molecules into various devices with diverse purposes, such as optical communication, computing and data storage, and image processing.[3,4] The design and synthesis of optimized molecules is a key goal in this area of research.[5−8] A method for predicting the macroscopic nonlinear optical characteristics of materials from their molecular electronic properties undoubtedly would be valuable in the search for compounds and materials showing strong NLO effects.

In the QSPR (quantitative structure−property relationship) study reported here we are interested in second and third harmonic generation, denoted SHG and THG, or β and γ, respectively. A general approach is being sought, in which the molecular properties input can be obtained in a straightforward manner. In an earlier study, neural network analysis, standard substituent constants, and ^{13}C NMR chemical shifts were used to model β from a family of substituted nitrobenzenes with varying results.[9]

MODEL BUILDING

In this study we focused on deriving a model in which experimental β and γ responses of molecules with significant structural differences could be predicted. Several electronic properties were calculated by molecular orbital (MO) methods to obtain variables for the model. The compounds are listed in Table 1 together with their experimental β (statvolt^{-1} cm^4) and γ (statvolt^{-2} cm^5) responses obtained from literature. The variables used in the models were as follows: the first seven levels of highest occupied and lowest unoccupied MOs (*HOMOs* and *LUMOs* in eV), ground-state dipole moments (*Dip*, in Debye), mean polarizability (*MeanPol*, in au),[10] and the band gap in eV, also known as "*hardness*", all calculated on the geometry optimized structure.

These properties, or descriptors, comprising an **X** matrix, were then related to the response, **Y**, by building a model using partial least squares (PLS).[11] PLS is an approach to the modeling of the relationship between an x-matrix and a y-matrix, where **X** is summarized in terms of a number of scores, $\mathbf{t}_a$ ($a = 1,2,...,A$). These scores are linear combinations of the original variables, computed with the objective to both well summarize **X** and well model and predict **Y**. There are many examples from various fields in chemistry, where PLS has been used with good results.[12−14] With PLS it is also possible to generate the regression coefficients for each variable. The size of the coefficients gives an estimate of the relative importance of the different x-variables, since the variables are scaled to unit variance before any models are done. All the (MO) calculations were performed using Gaussian 94W software,[15] while Chem3D Pro v. 5.0 was used for drawing and visualizing structures.[16] The structures were optimized on the Hartree−Fock (HF) level with the 3-21G* basis set,[17,18] followed by frequency calculations to verify energy minima. The choice to build a model on data from the relatively low-level HF/3-21G* method, instead from for instance the B3LYP/6-31G* functional method,[19] was taken to facilitate future calculations on larger molecules, where DFT and especially HF calculations using large basis sets may be very time demanding. It is expected that polarizabilities and hyperpolarizabilities are less accurate on the HF/3-21G* level, but the present results do not suggest that this method is inappropriate, vide infra. To validate the results and to decide the number of PLS components in each PLS model, cross-validation was used.[20]

RESULTS AND DISCUSSION

The first PLS model (M1), based on the 25 molecules (Table 1) for which experimental values of β were available, gave three PLS components explaining 87.5% of the response, **Y**, with a corresponding cross-validated value of 82.4%, see Table 2. The scaled coefficients (Figure 1) suggest that the most important variable for modeling β was the dipole moment, closely followed by the *hardness*. The dipole moment is positively correlated with the response, meaning that a high dipole moment for a molecule tends to give a high β value. In contrast, *hardness* is negatively correlated with the response, meaning that a low value is

* Corresponding author phone: +46 (0)90 7865358; fax: +46 (0)90 138885; e-mail: anders.berglund@chem.umu.se.

10.1021/ci000402q CCC: $20.00
Published on Web 05/08/2001

Table 1. Molecules for PLS Studies of β and γ

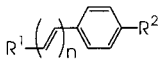
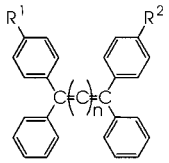
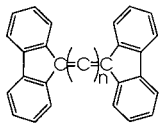
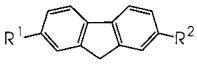
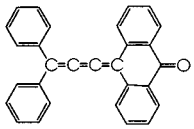
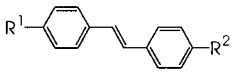
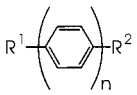
Object no.	Molecules	R^1	R^2	n	$\lg(\beta\times10^{30})$	$\lg(\gamma\times10^{36})$	ref
1a		CN	OMe	1	0.85	1.04	[21]
1b		CN	NMe_2	1	1.36	1.46	[21]
1c		CHO	OMe	1	1.04	1.45	[21]
1d		CHO	OMe	2	1.45	1.63	[21]
1e		CHO	OMe	3	1.62	2.08	[21]
1f		CHO	NMe_2	1	1.48	1.80	[21]
1g		CHO	NMe_2	2	1.72	2.15	[21]
1h		CHO	NMe_2	3	1.94	2.41	[21]
1i		NO_2	OMe	1	1.23	1.54	[21]
1j		NO_2	NMe_2	1	1.70		[21]
1k		CHO	Br	1	0.81	1.56	[21]
1l		NO_2	H	1	0.90	1.46	[21]
2a		COMe	NH_2		1.08	1.46	[21]
2b		CN	NMe_2		1.46	2.00	[21]
2c		NO_2	NMe_2		1.66	2.18	[21]
2d		NO_2	NH_2		1.38	2.08	[21]
2e		CN	SMe		1.18	1.54	[21]
2f		CO_2Me	SMe		0.90	1.15	[21]
3a		H	H	2		3.48	[22]
3b		OCH_3	OCH_3	2		3.85	[22]
3c		H	H	4		4.00	[22]
3d		NO_2	NO_2	4		3.00	[22]
3e		OCH_3	OCH_3	4		3.60	[22]
3f		NO_2	OCH_3	4		4.48	[22]
4				4		4.30	[22]
5a		NO_2	OMe		1.04	1.45	[21]
5b		NO_2	NMe_2		1.60	1.98	[21]
5c		Br	NO_2		0.78	1.48	[21]
6						3.00	[22]
7a		H	NMe_2		0.87	1.79	[23]
7b		NO_2	NMe_2		1.86	2.43	[21]
8b		NO_2	NH_2	3	1.20	2.09	[21]
8b		CN	NH_2	1	0.49	0.78	[23]

Table 2. Results from the PLS Models

model	N[a]	K[b]	A[c]	R^2X[d]	R^2Y[e]	Q^2 [f]
M1 beta	25	17	3	87.5	90.9	82.4
M2 beta	25	5	2	68.4	86.3	78.3
M3 gamma	32	17	3	87.5	90.7	82.1
M4 gamma	32	5	2	86.1	86.4	83.8

[a] Number of compounds in the model. [b] Number of variables. [c] Number of PLS components. [d] By the model explained variance in the descriptor matrix, **X**. [e] By the model explained variance in **Y**. [f] Cross-validated explained variance in **Y**.

associated with a strong response. The energies of the occupied MOs (OMOs) are less important, except for the first *HOMO* energy, which is positively correlated to the β value. For the UMO energies, the *LUMO+3* energy is positively correlated with the β value and somewhat more important than the *LUMO* energy, which is negatively correlated with β.

The scaled coefficients show that the five most important variables are *HOMO*, *LUMO*, *LUMO+3*, *Dip*, and *Hardness* (Figure 1). A new PLS model (M2) was calculated using only these five variables, resulting in a simpler model. The characteristics of this model (M2) are similar to the model with all the variables (M1) as can be seen in Table 2. This indicates that these five variables are sufficient for modeling the response. The plot of observed values versus the calculated values confirms this observation (Figure 2). This model (M2) was used to generate the following equation for predicting β:

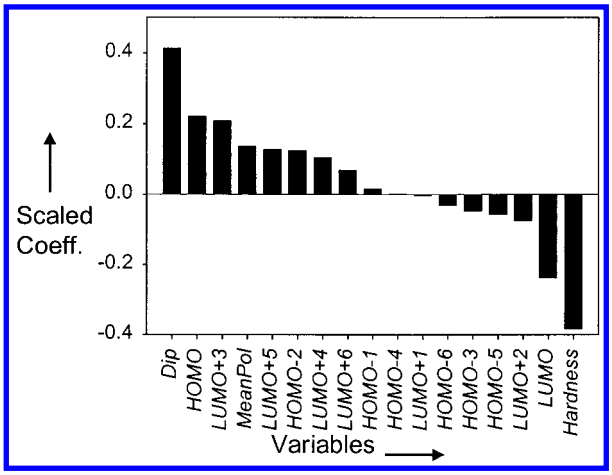


Figure 1. Coefficient plot showing the relative importance of each variable modeling the β value.

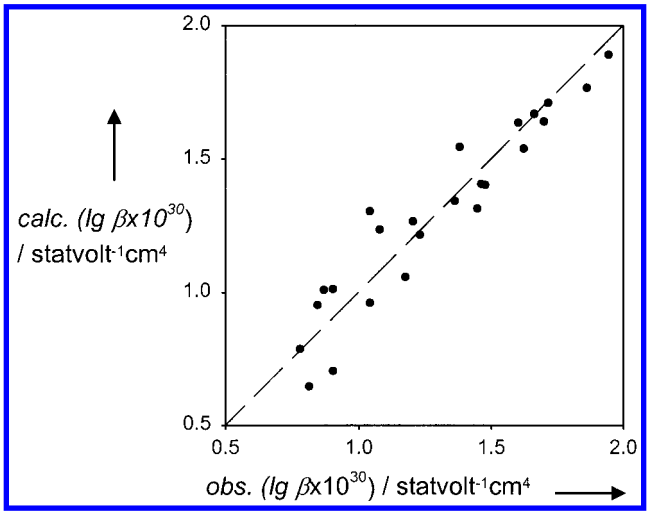


Figure 2. The calculated versus the observed β value.

$$\lg(\beta \times 10^{30}) = 4.204 + 5.781HOMO - 2.563LUMO + 2.852(LUMO{+}3) + 6.098 \times 10^{-2}Dip - 11.927Hardness \quad (1)$$

For the γ models, 32 molecules having experimental γ values were used, see Table 1. The model with all the variables, M3, has an explained variance of 90.7% for the response and 82.1% for the corresponding cross-validated value, which is similar to the β models, see Table 2. The scaled coefficients, see Figure 3, show that the most important variable is *LUMO+3* followed by the mean polarizability and the hardness. One difference for the coefficient plot, compared to the coefficient plot of the β model, is that the dipole moment is not important for the γ model.

A simplified model was also derived for γ, with five selected variables, namely the following: *HOMO*, *LUMO*, *LUMO*+3, *MeanPol*, and *hardness*. *HOMO* was selected instead of *HOMO-1* since they are correlated and *HOMO* is used in eq 1. The model, M4 (Table 2), shows almost the same statistical results as the M3 model. The model (M4) can explain the differences in γ between the molecules, see Figure 4, where the observed values are plotted against the calculated values. The following regression equation was obtained:

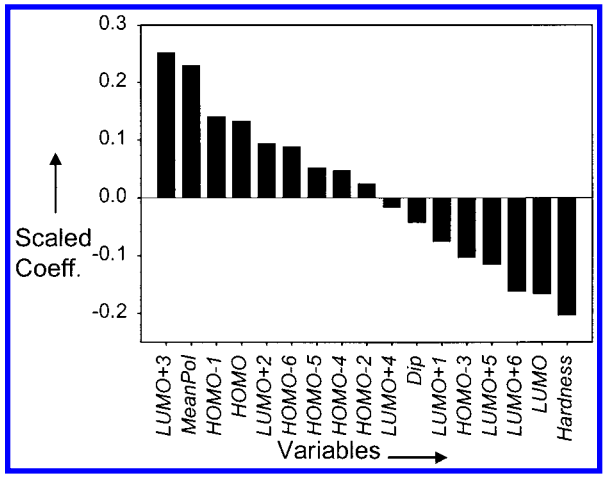


Figure 3. Coefficient plot showing the relative importance of each variable modeling the γ value.

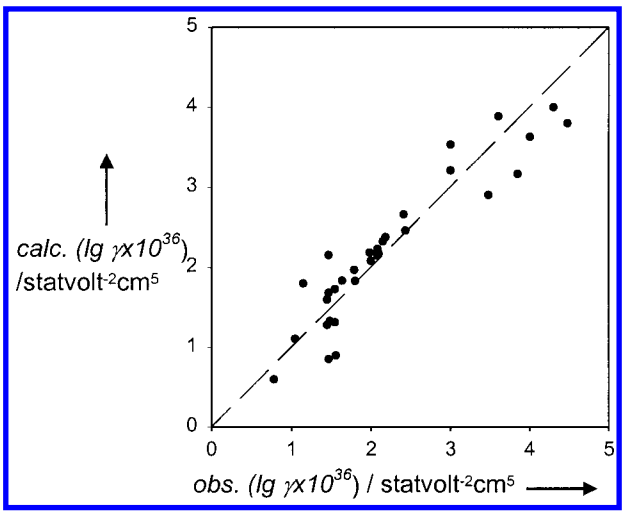


Figure 4. The calculated versus the observed γ value.

$$\lg(\gamma \times 10^{36}) = 7.171 + 12.089HOMO - 4.987LUMO + 2.395(LUMO{+}3) + 3.393 \times 10^{-3}meanPol - 14.875Hardness \quad (2)$$

CONCLUSIONS

Using the above equations experimental β and γ responses can be predicted for new molecules as an alternative to time-consuming synthesis and further studies. These are empirical models, which may be different from a model based on the first principals. A differing importance of the variables in the β and γ models is found. In the β models, the dipole and hardness are the most significant variables. This indicates that a target molecule should have a large dipole moment and a small band gap. For γ responses, a strong mean polarization is an important parameter for a molecule to be of interest. The derived equations for predicting β and γ (see above) give good results, with experimental responses being well predicted with the molecular diversity chosen in this study. An obvious source of error in a study like this is that the responses taken from the literature are measured under different experimental conditions. We believe that these errors are of minor importance in our case since the responses are well modeled. Further investigations are in progress with other types of NLO responses and molecular ranges.

ACKNOWLEDGMENT

We are grateful for support from J. C. Kempe Memorial (K.Ö.), Umeå University (B.E.), Knut and Alice Wallenberg Foundation (A.B.), and Swedish Natural Science Research Council (U.E).

Supporting Information Available: Fifty permutations of the M1, M2, M3, and M4 models. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Chemla, D. S.; Zyss, J. *Nonlinear Optical Properties of Organic Molecules and Crystals*; Academic Press: New York, 1987.
(2) Prasad, P. N.; Williams, D. J. *Introduction to Nonlinear Optical Effects in Molecules and Polymers*; Wiley: New York, 1991.
(3) Shi, S. Molecular-Structures, Macroscopic Nonlinear-Optical Effects, and Photonic Devices. *Contemp. Phys.* **1994**, *35*, 21−36.
(4) Zyss, J. *Molecular nonlinear optics: Materials, Physics and Devices*; Academic Press: Boston, 1994.
(5) Beck, B.; Grummt, U. W. Semiempirical calculations of first-order hyperpolarizabilities: Testing the performance of different methods in comparison to experiment. *J. Phys. Chem. B* **1998**, *102*, 664−670.
(6) Willetts, A.; Rice, J. E.; Burland, D. M.; Shelton, D. P. Problems in the Comparison of Theoretical and Experimental Hyperpolarizabilities. *J. Chem. Phys.* **1992**, *97*, 7590−7599.
(7) Kanis, D. R.; Ratner, M. A.; Marks, T. J. Design and Construction of Molecular Assemblies with Large 2nd- Order Optical Nonlinearities − Quantum-Chemical Aspects. *Chem. Rev.* **1994**, *94*, 195−242.
(8) Marder, S. R.; Beratan, D. N.; Cheng, L. T. Approaches for Optimizing the 1st Electronic Hyperpolarizability of Conjugated Organic-Molecules. *Science* **1991**, *252*, 103−106.
(9) Munn, R. W.; Munn, N. S.; Hardie, K. J. Neural Network Analysis of Molecular Hyperpolarisabilities. *Adv. Mater. Optics Electronics* **1994**, *4*, 19−26.
(10) Maertens, C.; Detrembleur, C.; Dubois, P.; et al. Structure − Second-order polarizability relationship in chromophores incorporating a spacer: A joint experimental and theoretical study. *Chem.-Eur. J.* **1999**, *5*, 369−380.
(11) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III The Collinearity Problem in Linear Regression. The Partial Least Squares Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.
(12) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(13) Kourti, T.; Macgregor, J. F. Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. *Chemometrics Intell. Lab. Syst.* **1995**, *28*, 3−21.
(14) Uppgard, L. L.; Lindgren, A.; Sjostrom, M.; Wold, S. Multivariate quantitative structure−activity relationships for the aquatic toxicity of technical nonionic surfactants. *J. Surfactants Detergents* **2000**, *3*, 33−41.
(15) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; et al. *Gaussian 94 (Revision D.1)*; Gaussian Inc.: Pittsburgh, PA, 1995.
(16) *CS Chem3D Pro*; CambridgeSoft Corporation: Cambridge, MA, 1996.
(17) Pietro, W. J.; Francl, M. M.; Hehre, W. J.; DeFrees, D. J.; Pople, J. A.; Binkley, J. S. ab initio. *J. Am. Chem. Soc.* **1982**, *104*, 5039.
(18) Dobbs, K. D.; Hehre, W. J. *J. Comput. Chem.* **1986**, *7*, 359.
(19) Becke, A. D. Density-Functional Thermochemistry. 3. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(20) Wold, S. Cross Validatory Estimation of the Number of Components in Factor and Principal Components Models. *Technometrics* **1978**, *20*, 397−406.
(21) Cheng, L. T.; Tam, W.; Marder, S. R.; Stiegman, A. E.; Rikken, G.; Spangler, C. W. Experimental Investigations of Organic Molecular Nonlinear Optical Polarizabilities. 2. A Study of Conjugation Dependences. *J. Phys. Chem.* **1991**, *95*, 10643−10652.
(22) Kminek, I.; Klimovic, J.; Prasad, P. N. 3rd-Order Nonlinear Optical-Response of Some Tetrasubstituted Cumulenes. *Chem. Mater.* **1993**, *5*, 357−360.
(23) Cheng, L. T.; Tam, W.; Stevenson, S. H.; Meredith, G. R.; Rikken, G.; Marder, S. R. Experimental Investigations of Organic Molecular Nonlinear Optical Polarizabilities. 1. Methods and Results on Benzene and Stilbene Derivatives. *J. Phys. Chem.* **1991**, *95*, 10631−10643.

CI000402Q