J Mol Model (2007) 13:951–963
DOI 10.1007/s00894-007-0209-4

ORIGINAL PAPER

QSPR modeling of hyperpolarizabilities

Alan R. Katritzky · Liliana Pacureanu · Dimitar Dobchev · Mati Karelson

Received: 16 November 2006 / Accepted: 23 April 2007 / Published online: 15 June 2007

Abstract The polarizabilities and the first and second hyperpolarizabilities of 219 conjugated organic compounds are modeled by QSPR (quantitative structure activity relationship) based on a large pool of constitutional, topological, electronic and quantum chemical descriptors calculated by CODESSA Pro (comprehensive descriptors for structural and statistical analysis) derived solely from molecular structure. Multilinear models were developed using the BMLR (best multilinear regression) algorithm to relate the experimental (hyper)polarizabilities to their predicted values. The regression equations include AM1 (Austin model 1) calculated (hyper)polarizabilities together with the size, electrostatic and quantum chemical descriptors to compensate for the imprecision of the AM1 computational method. The results emphasize the main factors that influence (hyper)polarizability. All models were validated by the "leave-one-out" method and internal validations that confirmed the stability and good predictive ability.

Electronic supplementary material The online version of this article (doi:10.1007/s00894-007-0209-4) contains supplementary material, which is available to authorized users.

A. R. Katritzky (✉) · L. Pacureanu · D. Dobchev
Center for Heterocyclic Compounds, Department of Chemistry, University of Florida,
Gainesville, FL 32611, USA
e-mail: katritzky@chem.ufl.edu

D. Dobchev · M. Karelson
Department of Chemistry, Tallinn University of Technology,
Ehitajate tee 5,
Tallinn 19086, Estonia

M. Karelson
Department of Chemistry, University of Tartu,
Jakobi Street 2,
Tartu 51014, Estonia

Keywords CODESSA Pro · Multilinear regression · Polarizability · QSPR

Introduction

QSPR (quantitative structure activity relationship) methods enable predictions of the properties of a wide range of chemical compounds based on the correlation between these properties and molecular descriptors. Many descriptors feature molecular electronic properties derived from quantum mechanics [1]. Others are mathematical indices derived from graph theory that characterize the shape and branching of the molecular skeleton [2].

Previous studies have emphasized the importance of molecular polarizability in (1) intermolecular weak interactions between closed shell species, and (2) for diverse molecular properties including boiling points, melting points, vaporization enthalpies, solubilities, and solvent polarity scales [3].

Molecular polarizability measures the response of the outer shell electrons of a molecule toward an external perturbation, whereas chemical binding can be also viewed as a result of reorganization of the valence electrons of atoms due to perturbation effects.

The molecular response that relates the dipole moment μ after the interaction with an external electric field can be written as follows:

$$\mu_i(F) = \mu_i + \sum_j \alpha_{ij} F_{ij} + \frac{1}{2}\sum_{jk} \beta_{ijk} F_j F_k + \frac{1}{6}\sum_{jkl} \gamma_{ijkl} F_j F_k F_l + \ldots\ldots \quad (1)$$

The first order term describe the linear response and the coefficients α_{ij} represent the elements of the linear first-order

polarizability tensor (α polarizability). The coefficients β_{ijk}, γ_{ijkl} are the first and second nonlinear hyperpolarizability terms (β polarizability and γ polarizability). The μ_i is the dipole moment in the absence of the field, and the indeces i, j, k, l denote the Cartesian coordinates x, y and z. Static polarizabilities and hyperpolarizabilities (time independent, or frequency independent) can be expressed by taking the nth order derivatives of the dipole moment with respect to the electric field:

$$\alpha_{ij} = \frac{\delta \mu_i}{\delta F_j}|_{F=0} \tag{2}$$

$$\beta_{ijk} = \frac{\delta^2 \mu_i}{\delta F_j \delta F_k}|_{F=0} \tag{3}$$

$$\gamma_{ijkl} = \frac{\delta^3 \mu_i}{\delta F_j \delta F_k \delta F_l}|_{F=0} \tag{4}$$

Quantum chemistry provides several methodologies to calculate polarizability. The sum over states (SOS) procedure is based on a perturbation treatment where the calculations are performed independently of the field and the response involves the coupling of the excited states of the molecule [4]. The second approach is the finite field (FF) method, which considers the molecule in the presence of an external static electric field. The Hamiltonian is diagonalized in order to calculate the ground state energy (or dipole moment) at various F values and n-order polarizabilities are calculated as numerical derivatives of the energy (dipole moment) with respect to the electric field [4].

The first tentative to calculate first hyperpolarizability was provided by the classical equivalent internal field (EIF) model. The second order response of a substituted conjugated molecule is calculated from the ground state π-electron distribution perturbated by asymmetric distortion due to the substituent [5].

Oudar and Chemla proposed the two states model to describe the large nonlinear responses in donor–acceptor chromofores due to intramolecular charge–transfer interaction between donor and acceptor substituents that can be explained by means of ground and excited states.

$$\beta = \beta_{add} + \beta_{ct} \tag{5}$$

where β_{add} is the additive portion and β_{ct} is provided from the interaction of donor and acceptor groups.

$$\beta_{ct} = \frac{3e^2\hbar^2}{2m} \frac{W}{(W^2 - (2\hbar\omega^2)\left(W^2 - (\hbar\omega)^2\right)} f \Delta\mu_{ex} \tag{6}$$

where $\hbar\omega$ is the energy of a laser photon, $\Delta\mu_{\mathrm{ex}}$ the difference between excited state dipole moment and ground state dipole moment, W the energy of the excited state, and f the oscillator strength [6].

Zyss proposed a three states model that can better explain the nonlinear responses for octopolar systems whose $\Delta\mu_{\mathrm{ex}}$= 0, where β can be calculated from the product of dipolar transition moment coupling three states [7].

Modern ab initio approaches include electron-correlation effects for the ground state calculation in order to overcome the drawbacks of Hartree-Fock approximation. The most popular methods that account for electronic correlations are: (1) Rayleygh-Scrödinger many-body perturbation theory (MBPT) by Møller-Plesset at second (MP2), third (MP3) and fourth (MP4) order; (2) coupled cluster (CC) methods. Correlation effects were proven to be important for small as well as for large molecules. Theoretical investigations demonstrate that MP2 approach is able to describe 90% of the effects of electron correlation on static first hyperpolarizability [8].

Adant et al. [9] investigated the electron-correlation effect on the polarizability and second hyperpolarizabilitiy of benzene and thiophene derivatives at the ab initio MP2 (FF method) using 3–21G + (p,d) basis set and semiempirical INDO-CI/CV (intermediate neglect of diferential overlap-configuration interaction/correction vector). The results showed that the semiempirical approach is able to reproduce the trends observed following the ab initio method.

In most cases the calculation of (hyper)polarizabilities is carried out by using a time-dependent response technique such as the time-dependent Hartree-Fock (TDHF) [4]. Frequency dependent molecular response represents the most complex and reliable method for hyperpolarizability calculation.

$$F(\omega) = F_0 + F_\omega \cos(\omega t) \tag{7}$$

where ω is the frequency of the applied field F.

The frequency-dependent response can be calculated as a higher order derivative of "pseudo-energy" defined as the eigenvalue of the operator H-iħ∂/∂t. Similarly, dynamic hyperpolarizabilities can be calculated using various electronic correlations approaches including MP2 or coupled cluster methods [8].

The response calculated following the time-dependent density-functional theory (TDDFT) was proven to be superior to TDHF calculations with regard to efficiency and accuracy.

An electrostatic model has been developed by Jensen and co-workers [10] to calculate ab initio frequency-dependent molecular polarizability using atom-type parameters able to describe polarizability, the damping of the electric fields and frequency-dependence. The same group [11] developed the electrostatic interaction model based on point dipole interaction (PDI) to assess the frequency-

dependent polarizability for 20 essential amino acids and three small proteins. Comparison of the PDI model with SCF calculation provided an accuracy of 5% for amino acids and 10% for proteins, suggesting that the PDI model is suitable for classical force field calculations or combined QM/MM (quantum mechanics / molecular mechanics) calculations whenever QC calculation is prohibitive.

Polarizabilities are usually determined experimentally by refractometric techniques, including studies of relative permitivity, Rayleigh and Raman scattering, and the quadratic Stark effect. Hyperpolarizabilities can be determined by sophisticated techniques including hyper Rayleigh scattering technique (HRS), d.c. electric-field induced optical second-harmonic generation (EFISH), third-harmonic generation experiments (THG), and the Kerr optic effect [12, 13].

The polarizability is closely related to the hardness and electronegativity of a molecule. Hardness and electronegativity are defined as the first and second order derivatives of the total energy with respect to the number of electrons [14]. They provide considerable insight into chemical binding, reactivity and selectivity of molecular systems by describing the response of a molecule to the variation of the number of electrons at constant external potential [15, 16].

An important property used in QSPR analysis is molar refractivity (MR). MR has been related to lipophilicity, molar volume, steric bulk and polarizability [17–19] and can be calculated by the Lorentz-Lorentz equation:

$$MR = \frac{(n^2 - 1)(n^2 + 2) \cdot MW}{\rho} = 1.333\pi N\alpha \qquad (8)$$

In Eq. 8, MW is the molecular weight, $\tilde{n}$ the density, n the refractive index, N is Avogadro's number, and α the polarizability. Thus, refractive index depends on the polarizability, size and polarity of a chemical species.

Hansch and coworkers correlated polarizability, MR and the number of valence electrons (NVE) for organic compounds containing halogen, oxygen, nitrogen, phosphorus, or sulfur [19–22].

Recently much research interest has been devoted to organic materials that display large optical nonlinearities [5, 6, 8, 12, 13]. Conjugated organic compounds with electron donors and electron acceptor groups (push-pull systems) display significant non-linear optical (NLO) properties with respect to harmonic generation (HG). NLO materials have numerous applications in optical devices, storage, information processing, optical switches, etc.; the phenomenon of charge transfer is very important in chemical interactions and also in physical applications, therefore the optimization of charge transfer in a push-pull system is of significant interest. Certain electron donating–accepting group combinations can provide ground state charge asymmetry, while the π-conjugated backbone enables the transfer and redistribution of electric charge under the influence of an external electric field. The first hyperpolarizability depends directly on the strength of the donor–acceptor groups, and on the length of the π-conjugation and symmetry of the molecule [4, 23]. Among the factors that play an important role in the charge transfer process is the optical gap [HOMO–LUMO gap (highest occupied molecular orbital–lowest unoccupied molecular orbital)]. A small optical gap favors the charge transfer process [4]. There is an optimal combination of donor–acceptor groups and an optimal length of the π-conjugation that results in maximum first hyperpolarizability. Recent investigations suggest that a polyenic chain is a better bridge for long path while a polyynic chain is better for short linkers due to the specific mobility of π-electrons [13]. In a recent model, β is correlated with the degree of ground state polarization, which depends on chemical structure. In donor–acceptor polyenes this quantity is related to bond length alternation (BLA) or bond order alternation (BOA), which describe the mixture of the resonance forms in the ground state structure of the chemical system [5, 24].

The design of novel molecules relies on information provided by theoretical and experimental investigations into the relationships between (hyper)polarizability and donor–acceptor strength, nature and length of π-conjugation, etc. [23, 25]. Accessible and cost-effective quantum chemical tools permit the calculation of (hyper)polarizabilities for all types of chemical species. However, the approximations required introduce a degree of uncertainty, so the results have to be validated by the experimental determinations [12].

QSPR methodology has successfully enabled the correlation of a large variety of physico-chemical properties of molecules, therefore it was of interest to investigate (hyper) polarizability by this approach [26–30].

CODESSA Pro was previously employed to correlate and predict many physical properties, including boiling points [27], solvent scales and refractive indices [28], chromatographic properties [29], critical micelle concentration of nonionic and anionic surfactants [30, 31], complexation free energy of β-cyclodextrin and organic compounds [32], and partition coefficients including rat blood:air, saline:air, and olive oil:air [33].

The present paper is devoted to a study of polarizability and hyperpolarizabilities of large sets of organic compounds by means of QSPR in an attempt to construct a statistically significant model that can be used to predict (hyper)polarizabilitiy for new conjugated compounds with NLO properties. In this paper we compare the QSPR models with both the experimentally determined polarizabilities and AM1 FF calculated polarizabilities and hyperpolarizabilities.

Materials and methods

Methodology

The compound 2D structures were drawn using Hyperchem 7.5 software [34] and their geometry were preoptimized using the molecular mechanics force field (MM+) available in the Hyperchem 7.5 package. The final refined equilibrium molecular geometries were obtained using the semiempirical method AM1 (Austin Model-1) without any symmetry constraint [35], applying a gradient norm limit of 0.01 kcal/Å as a stopping criterion for optimized structures.

These geometries were used to calculate up to 700 molecular descriptors using the CODESSA-Pro package, classified as (1) constitutional, (2) topological, (3) geometrical, (4) charge-related, (5) semiempirical, or (6) thermodynamical [36]. The best multilinear regression (BMLR) procedure [29] was used to find the best correlation models from the selected noncolinear descriptors. A correlation is not statistically reliable if the variables are mutually intercorrelated. The BMLR selects the best two-parameter regression equation, the best three-parameter regression equation, etc., based on the highest R^2 and F values in the stepwise regression procedure [37], using only those descriptors with intercorrelation coefficients smaller than a certain threshold value. During the BMLR procedure, the descriptor scales are also normalized, centered automatically and the final result is given in natural scales. The resulting n-parameter regression equations have the best representations of the property in the given descriptors pool. The final correlation equations are selected on the basis of partial and standard tests of significance and the highest regression coefficients.

A major decision in developing successive QSPRs is when to stop adding descriptors to the model during the stepwise regression procedure. The lack of adequate control leads to over-correlated equations, which contain an excess of descriptors and are difficult to analyze in terms of interaction mechanisms. A simple procedure to control the model expansion is the so-called 'break point' resulting in the plot of the number of descriptors involved in the obtained models versus the corresponding squared correlation coefficient. Moreover, augmentation of the number of the descriptors could lead to the inclusion of descriptors that are highly intercorrelated. During the BMLR procedure, the addition of the descriptors to the QSPR equations was monitored. Thus, if no significant improvement of the statistical parameters (s, F and especially R^2) was observed, then the current model with a certain number of descriptors reaching the break point was considered the optimum.

The QSPR models obtained were validated by the "leave-one-out" method and by internal validation, whereby each one-third of the compounds is predicted by a model derived by the remaining two-thirds of the compounds. The corresponding squared cross-validated correlation coefficient (R^2_{cv}) for all selected models is calculated automatically by the validation module implemented in CODESSA Pro package.

Data set

The present study of 219 compounds includes benzene, styrene, stilbene, fluorene, diphenylacetylene, α-phenylpolyene oligomers, diphenylpolyynes, and polyphenyl oligomers derivatives with push-pull systems that consist of electron donor and electron acceptor groups with various strength (see Table 1) [12, 13]. Our dataset is based on literature experimental data measured in solution using various techniques such as refractive index at various wavelengths, dielectric constant, THG, and EFISH amplitudes and coherence lengths on a series of solutions of graded concentrations. Using solution properties, the literature authors [12, 13] calculated significant molecular properties including permanent dipole moments and the tensor projections corresponding to the polarizability and the first and second hyperpolarizabilities in Eq. 1 using the approximations of liquid dielectric behavior. In addition they carried out solvatochromic measurements to assess the charge transfer nature of the low-lying electronic excitations [12, 13].

The experimental determinations of (hyper)polarizabilities of the current data set of NLO molecules were performed in various organic solvents, e.g., acetonitrile, *N*-methylpyrrolidone, $CHCl_3$, *p*-dioxane. It is well known that solvent interactions cause significant changes in the molecular (hyper)polarizabilities as well as excitation energies and transition moments of a solute with respect to the gas phase [38]. In the case of polarized solvents, solute–solvent interactions are appropriately described, including the induced dipole moments [39]. Previous calculations demonstrate that while the hyperpolarizability dispersion and sign of organic compounds with NLO properties change when different solvents were modeled, the first order response is little affected [38, 40, 41]. However, as the AM1 calculated (hyper)polarizabilities refer to the gas phase, they need correction for the condensed phase. Our objective is to do this for the condensed phase using the QSPR approach.

For the present work the values for polarizability and first and second hyperpolarizabilities, expressed in electrostatic units (esu), were collected from references [12, 13] and converted by us into logarithmic units to improve the normal distribution of the experimental data points and to decrease their clusterization.

We built QSPR models for the Full Set of compounds and also for Class I, which consists of 140 conjugated

Table 1 Chemical names of the conjugated π-system compounds considered

No.	Chemical name	No.	Chemical name
1	(E)-3-(4-methoxyphenyl)-2-propenenitrile	111	9-4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl–2,3,6, 7-tetrahydro-1*H*,5*H*-pyrido[3,2,1]quinoline
2	(E)-3-[4-(dimethylamino)phenyl]-2-propenenitrile	112	methyl 4-[(*E*)-2-(4-nitrophenyl)ethenyl]benzoate
3	(E)-3-(4-bromophenyl)-2-propenal	113	4-[(*E*)-2-(4-nitrophenyl)ethenyl]benzaldehyde
4	(E)-3-(4-methoxyphenyl)-2-propenal	114	1-[(*E*)-2-(4-bromophenyl)ethenyl]-4-methoxybenzene
5	(*E*)-3-[4-(dimethylamino)phenyl]-2-propenal	115	methyl 3-[(*E*)-2-(2-nitrophenyl)ethenyl]phenyl ether
6	(E)-4-(4-methoxyphenyl)-3-buten-2-one	116	methyl 2-[(*E*)-2-(3-nitrophenyl)ethenyl]phenyl ether
7	1-[(*E*)-2-nitroethenyl]benzene	117	methyl 3-[(*E*)-2-(3-nitrophenyl)ethenyl]phenyl ether
8	4-[(*E*)-2-nitroethenyl]phenol	118	methyl 4-[(*E*)-2-(3-nitrophenyl)ethenyl]phenyl ether
9	1-methoxy-4-[E)-2-nitroethenyl]benzene	119	methyl 3-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl ether
10	*N*,*N*-dimethyl-4-[(*E*)-2-nitroethenyl]aniline	120	1-bromo-2-[(*E*)-2-(4-nitrophenyl)ethenyl]benzene
11	(*E*)-*N*,*N*-dimethyl-2-(4-nitrophenyl)-1-ethenamine	121	1-bromo-3-[(*E*)-2-(4-nitrophenyl)ethenyl]benzene
12	1,1′-biphenyl	122	1-[(*E*)-2-phenylethenyl]benzene
13	[1,1′-biphenyl]-4-carbonitrile	123	1-[(*Z*)-2-phenylethenyl]benzene
14	1-[1,1′-biphenyl]-4-yl-1-ethanone	124	4-[(*E*)-2-phenylethenyl]aniline
15	4-nitro-1,1′-biphenyl	125	*N*,*N*-dimethyl-4-[(*E*)-2-phenylethenyl]aniline
16	6-[4′-(dimethylamino)[1,1′-biphenyl]-4-yl]sulfonyl-1-hexanol	126	1-nitro-4-[(*E*)-2-phenylethenyl]benzene
17	4′-hydroxy[1,1′-biphenyl]-4-carbonitrile	127	*N*-[(*E*)-(4-nitrophenyl)methylidene]aniline
18	1-(4′-methoxy[1,1′-biphenyl]-4-yl)-1-ethanone	128	4-methyl-*N*-[(*E*)-(4-nitrophenyl)methylidene]aniline
19	4-bromo-4′-nitro-1,1′-biphenyl	129	4-methoxy-*N*-[(*E*)-(4-nitrophenyl)methylidene]aniline
20	4′-nitro[1,1′-biphenyl]-4-ol	130	*N*-[(*E*)-(4-methoxyphenyl)methylidene]-4-nitroaniline
21	4-methoxy-4′-nitro-1,1′-biphenyl	131	4-[(*E*)-2-(4-nitrophenyl)diazenyl]aniline
22	4′-nitro[1,1′-biphenyl]-4-amine	132	2-ethyl-4-[(*E*)-2-(4-nitrophenyl)diazenyl]anilino-1-ethanol
23	*N*,*N*-dimethyl-4′-nitro[1,1′-biphenyl]-4-amine	133	methyl 2-methyl-4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl ether
24	9*H*-fluorene	134	2-methoxy-4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl methyl ether
25	9*H*-fluorene-2-carbonitrile	135	2-fluoro-4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl methyl ether
26	2-nitro-9*H*-fluorene	136	3-methoxy-4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl methyl ether
27	2-bromo-7-nitro-9*H*-fluorene	137	(*Z*)-3-(4-methoxyphenyl)-2-(4-nitrophenyl)-2-propenenitrile
28	2-methoxy-7-nitro-9*H*-fluorene	138	(*Z*)-3-(4-bromophenyl)-2-(4-nitrophenyl)-2-propenenitrile
29	*N*,*N*-dimethyl-*N*-(7-nitro-9*H*-fluoren-2-yl)amine	139	*N*-(4-bromophenyl)-4-nitrobenzenecarboximidoyl cyanide
30	4-nitroaniline	140	4-[(*E*)-2-(2,4-dinitrophenyl)ethenyl] phenyl methyl ether
31	4′-nitro[1,1′-biphenyl]-4-amine	141	benzene
32	4″-nitro[1,1:4,1″-terphenyl]-4-amine	142	toluene
33	4‴-nitro[1,1′:4′,1″:4″,1‴]-quaterphenyl]-4-amine	143	phenol
34	1-methoxy-4-nitrobenzene	144	benzenethiol
35	4-methoxy-4′-nitro-1,1′-biphenyl	145	methyl phenyl ether
36	4-methoxy-4″-nitro[1,1′:4′,1″-terphenyl]	146	1-(methylsulfanyl)benzene
37	4-2-[4-(methylsulfonyl)phenyl]ethynylaniline	147	aniline
38	methyl 4-2-[4-(methylsulfanyl)phenyl]ethynyl benzoate	148	*N*,*N*-dimethylaniline
39	methyl 4-[2-(4-aminophenyl)ethynyl]benzoate	149	9-phenyl-2,3,6,7-tetrahydro-1*H*,5*H*-pyrido[3,2,1-*ij*] quinoline
40	1-(4-2-[4-(methylsulfanyl)phenyl]ethynylphenyl)-1-ethanone	150	1-fluorobenzene
41	1-4-[2-(4-aminophenyl)ethynyl]phenyl-1-ethanone	151	1-chlorobenzene
42	4-[2-(4-aminophenyl)ethynyl]phenyl(phenyl) methanone	152	1-bromobenzene
43	4-2-[4-(methylsulfanyl)phenyl]ethynylbenzonitrile	153	1-iodobenzene
44	4-[2-(4-aminophenyl)ethynyl]benzonitrile	154	methyl phenyl sulfone
45	4-2-[4-(methylamino)phenyl]ethynylbenzonitrile	155	benzenesulfonyl fluoride
46	4-2-[4-(dimethylamino)phenyl]ethynylbenzonitrile	156	benzonitrile
47	1-bromo-4-[2-(4-nitrophenyl)ethynyl]benzene	157	benzaldehyde
48	1-methoxy-4-[2-(4-nitrophenyl)ethynyl]benzene	158	2,2,2-trifluoro-1-phenyl-1-ethanone
49	1-(methylsulfanyl)-4-[2-(4-nitrophenyl)ethynyl] benzene	159	1-nitrosobenzene

Table 1 (continued)

No.	Chemical name	No.	Chemical name
50	4-[2-(4-nitrophenyl)ethynyl]aniline	160	1-nitrobenzene
51	*N*-methyl-4-[2-(4-nitrophenyl)ethynyl]aniline	161	2-(phenylmethylene)malononitrile
52	*N*,*N*-dimethyl-4-[2-(4-nitrophenyl)ethynyl]aniline	162	4-(methylsulfonyl)phenol
53	4-4-[4-(methylsulfanyl)phenyl]-1,3-butadiynylbenzonitrile	163	4-methylbenzonitrile
54	4-[4-(4-aminophenyl)-1,3-butadiynyl]benzonitrile	164	4-chlorobenzonitrile
55	1-(methylsulfanyl)-4-[4-(4-nitrophenyl)-1,3-butadiynyl]benzene	165	4-bromobenzonitrile
56	4-[4-(4-nitrophenyl)-1,3-butadiynyl]aniline	166	4-phenoxybenzonitrile
57	4-methoxybenzaldehyde	167	4-methoxybenzonitrile
58	(2*E*,4*E*)-5-(4-methoxyphenyl)-2,4-pentadienal	168	4-(methylsulfanyl)benzonitrile
59	(2*E*,4*E*,6*E*)-7-(4-methoxyphenyl)-2,4,6-heptatrienal	169	4-aminobenzonitrile
60	4-(dimethylamino)benzaldehyde	170	4-(dimethylamino)benzonitrile
61	(2*E*,4*E*)-5-[4-(dimethylamino)phenyl]-2,4-pentadienal	171	4-methylbenzaldehyde
62	(2*E*,4*E*,6*E*)-7-[4-(dimethylamino)phenyl]-2,4,6-heptatrienal	172	4-phenoxybenzaldehyde
63	4-(2-ethynyl-1-buten-3-ynyl)phenyl methyl ether	173	4-methoxybenzaldehyde
64	2-[(*E*)-3-(4-methoxyphenyl)-2-propenylidene] malononitrile	174	4-(methylsulfanyl)benzaldehyde
65	*N*-[4-(2-ethynyl-1-buten-3-ynyl)phenyl]-*N*,*N*-dimethylamine	175	4-(dimethylamino)benzaldehyde
66	2-(*E*)-3-[4-(dimethylamino)phenyl]-2-propenylidenemalononitrile	176	4-methoxyphenyl 1,2,2,2-tetrafluoro-1-(trifluoromethyl) ethyl sulfone
67	2-(2*E*,4*E*)-5-[4-(dimethylamino)phenyl]-2,4-pentadienylidenemalononitrile	177	2,2,2-trifluoro-1-(4-methoxyphenyl)-1-ethanone
68	4-[(*E*)-2-(4-methoxyphenyl)ethenyl]benzonitrile	178	2,2,2-trifluoro-1-(4-phenoxyphenyl)-1-ethanone
69	4-[(1*E*,3*E*)-4-(4-methoxyphenyl)-1,3-butadienyl]benzonitrile	179	1-[4-(dimethylamino)phenyl]-2,2,2-trifluoro-1-ethanone
70	4-[(1*E*,3*E*,5*E*)-6-(4-methoxyphenyl)-1,3,5-hexatrienyl]benzonitrile	180	*N*,*N*-dimethyl-4-nitrosoaniline
71	1-[(*E*)-2-(4-bromophenyl)ethenyl]-4-nitrobenzene	181	1-methyl-4-nitrobenzene
72	1-bromo-4-[(1*E*,3*E*)-4-(4-nitrophenyl)-1,3-butadienyl]benzene	182	1-bromo-4-nitrobenzene
73	1-bromo-4-[(1*E*,3*E*,5*E*)-6-(4-nitrophenyl)-1,3,5-hexatrienyl]benzene	183	4-nitrophenol
74	1-methoxy-4-[(*E*)-2-(4-nitrophenyl)ethenyl]benzene	184	1-phenoxy-4-nitrobenzene
75	methyl 4-[(1*E*,3*E*)-4-(4-nitrophenyl)-1,3-butadienyl]phenyl ether	185	1-methoxy-4-nitrobenzene
76	1-methoxy-4-[(1*E*,3*E*,5*E*)-6-(4-nitrophenyl)-1,3,5-hexatrienyl]benzene	186	1-(methylsulfanyl)-4-nitrobenzene
77	(1*E*,3*E*,5*E*,7*E*)-1-(4-methoxyphenyl)-8-(4-nitrophenyl)-1,3,5,7-octatetraene	187	1-(4-nitrophenyl)hydrazine
78	methyl 4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenyl sulfide	188	4-nitroaniline
79	methyl 4-[(1*E*,3*E*)-4-(4-nitrophenyl)-1,3-butadienyl]phenyl sulfide	189	*N*,*N*-dimethyl-4-nitroaniline
80	*N*,*N*-dimethyl-4-[(*E*)-2-(4-nitrophenyl)ethenyl]aniline	190	4-nitrobenzonitrile
81	*N*,*N*-dimethyl-4-[(1*E*,3*E*)-4-(4-nitrophenyl)-1,3-butadienyl]aniline	191	4-nitrobenzaldehyde
82	*N*,*N*-dimethyl-4-[(1*E*,3*E*,5*E*)-6-(4-nitrophenyl)-1,3,5-hexatrienyl]aniline	192	2-[(4-methoxyphenyl)methylene]malononitrile
83	*N*,*N*-dimethyl-4-[(1*E*,3*E*,5*E*,7*E*)-8-(4-nitrophenyl)-1,3,5,7-octatetraenyl] aniline	193	2-[4-(dimethylamino)phenyl]methylenemalononitrile
84	2-[(1*E*,3*E*)-4-(2-methoxyphenyl)-1,3-butadienyl]benzonitrile	194	2-[4-(2,3,6,7-tetrahydro-1*H*,5*H*-pyrido[3,2,1-*ij*]quinolin-9-yl)phenyl]methylenemalononitrile
85	2-[(1*E*,3*E*,5*E*)-6-(2-methoxyphenyl)-1,3,5-hexatrienyl]benzonitrile	195	2-(4-aminophenyl)-1,1,2-ethylenetricarbonitrile
86	2-[(1*E*,3*E*)-4-(4-methoxyphenyl)-1,3-butadienyl]benzonitrile	196	2-[4-(dimethylamino)phenyl]-1,1,2-ethylenetricarbonitrile
87	2-[(1*E*,3*E*,5*E*)-6-(4-methoxyphenyl)-1,3,5-hexatrienyl]benzonitrile	197	2-[4-(2,3,6,7-tetrahydro-1*H*,5*H*-pyrido[3,2,1-*ij*]quinolin-9-yl)phenyl]-1,1,2-ethylenetricarbonitrile
88	4-[(1*E*,3*E*)-4-(2-methoxyphenyl)-1,3-butadienyl]benzonitrile	198	1-methyl-2-nitrobenzene
89	1-methoxy-2-[(*E*)-2-(2-nitrophenyl)ethenyl]benzene	199	1-bromo-2-nitrobenzene
90	methyl 2-[(1*E*,3*E*)-4-(2-nitrophenyl)-1,3-butadienyl]phenyl ether	200	2-nitrophenol
91	1-methoxy-2-[(1*E*,3*E*,5*E*)-6-(2-nitrophenyl)-1,3,5-hexatrienyl]benzene	201	1-methoxy-2-nitrobenzene
92	methyl 4-[(*E*)-2-(2-nitrophenyl)ethenyl]phenyl ether	202	2-nitroaniline
93	1-[(1*E*,3*E*)-4-(4-methoxyphenyl)-1,3-butadienyl]-2-nitrobenzene	203	2-nitrobenzonitrile
94	1-[(1*E*,3*E*,5*E*)-6-(4-methoxyphenyl)-1,3,5-hexatrienyl]-2-nitrobenzene	204	2-nitrobenzaldehyde
95	1-methoxy-2-[(*E*)-2-(4-nitrophenyl)ethenyl]benzene	205	1-methyl-3-nitrobenzene
96	methyl 2-[(1*E*,3*E*)-4-(4-nitrophenyl)-1,3-butadienyl]phenyl ether	206	1-bromo-3-nitrobenzene
97	1-methoxy-2-[(1*E*,3*E*,5*E*)-6-(4-nitrophenyl)-1,3,5-hexatrienyl]benzene	207	3-nitrophenol
98	*N*,*N*-dimethyl-4-[(*E*)-2-(5-nitro-2-furyl)ethenyl]aniline	208	1-methoxy-3-nitrobenzene
99	*N*,*N*-dimethyl-4-[(*E*)-2-(5-nitro-2-thienyl)ethenyl]aniline	209	3-nitroaniline

Table 1 (continued)

No.	Chemical name	No.	Chemical name
100	*N*,*N*-dimethyl-4-[(1*E*,3*E*)-4-(5-nitro-2-furyl)-1,3-butadienyl]aniline	210	3-nitrobenzonitrile
101	2-(4-methoxyphenyl)-5-(4-nitrophenyl)furan	211	3-nitrobenzaldehyde
102	6-(4-(*E*)-2-[4-(pentylsulfonyl)phenyl]ethenyl phenoxy)-1-hexanol	212	5-nitro-2-pyridinamine
103	(E)-1-methoxy-4(4perfluorohexylsulfonyl) phenylethenyl) benzene	213	5-methoxy-2-nitropyridine
104	trifluoromethyl 4-[(*E*)-2-(4-methoxyphenyl)ethenyl] benzoate	214	2-methyl-4-nitroaniline
105	4-[(*E*)-2-(4-hydroxyphenyl)ethenyl]benzonitrile	215	3-chloro-4-nitroaniline
106	4-(*E*)-2-[4-(dimethylamino)phenyl]ethenyl benzonitrile	216	2-methoxy-4-nitroaniline
107	1-methyl-4-[(*E*)-2-(4-nitrophenyl)ethenyl]benzene	217	2-fluoro-4-methoxy-1-nitrobenzene
108	4-[(*E*)-2-(4-nitrophenyl)ethenyl]phenol	218	1,4-difluoro-2-methoxy-5-nitrobenzene
109	1-[(*E*)-2-(4-nitrophenyl)ethenyl]-4-phenoxybenzene	219	1,2,4,5-tetrafluoro-3-methoxy-6-nitrobenzene
110	4-[(*E*)-2-(4-nitrophenyl)ethenyl]aniline		

organic compounds (Table 1 entries 1–140), and Class II, which comprises 79 mono and polysubstituted benzene derivatives (Table 1, entries 141–219) [12, 13].

Results and discussion

The BMLR [29] algorithm was employed to correlate the independent descriptor variables with the logarithm of the polarizability, and the first and second hyperpolarizabilities to obtain the best multilinear QSPR model for the Full Set of organic compounds. The experimental and predicted logarithm of polarizabilty (static values), and first and second hyperpolarizability values are shown in Supplementary Material. Obtaining satisfactory QSPR models required the use of AM1 calculated quantum chemical polarizability, first and second hyperpolarizabilities as descriptors in addition to the usual size, as well as electrostatic and quantum chemical related descriptors. These partially compensate for the drastic approximations needed in the AM1 computational method before they can be applied to calculate polarizabilities [12]. The polarizability of a system is a measure of the response of the electron density distribution to a static electric field. This response can be calculated in two ways, by the application of electric fields (the static method) and by direct analysis of the wavefunction (the sum over states method). For large systems, the application of electric fields method is faster, although the sum over states method is more precise for all systems. Because of the limited precision of the static method, a way has to be provided to allow the user to determine the precision. Static polarizability (and first and second nonlinear optical responses) can be determined from the changing dipole or from the changing heat of formation (see Mopac manual [42]). Since polarizability is strongly related to the dipole moment and the heat of formation it can incorporate similar errors. The error in Debye units for the dipole moment for all elements averages 0.78, 0.52, 0.62, 0.66 for MNDO (modified neglect of differential overlap), AM1, PM3 and PM5 (parametrized method number 3 and 5) methods, respectively [42]. Errors in heats of formation for the above semiempirical methods (e.g., for AM1 the average is 12 kcal mol^{-1}) also contributes to uncertainties in the polarizability. Moreover, ab initio calculations provided by Hinchliffe et al. demonstrate that AM1 calculations provide reliable results for small molecules, e.g., benzene, but not for larger molecules [43]. Therefore, correcting the calculated polarizability by a QSPR equation using molecular descriptors as independent variables could lead to significant improvement.

The experimental and calculated values of polarizability using the best QSPR model for the Full Set of compounds are plotted in Fig. 1. The model indicates that the calculated

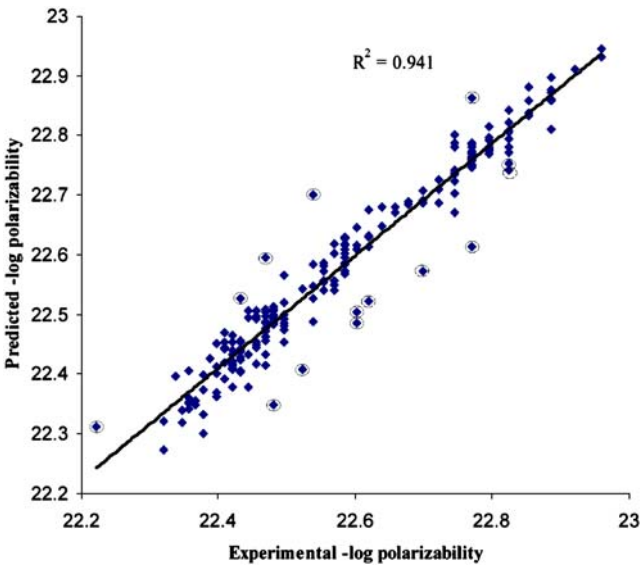

Fig. 1 Plot of 212 experimental values of log polarizability versus quantitative structure activity relationship (QSPR) predicted values (14 outliers are marked; Table 3)

properties correlate well with the experiment data and suggest good prediction ability.

Polarizability

The QSPR analysis, performed for the Full Set of compounds, is not better than the correlation between AM1 calculated and experimental values (Figs. 1, 2). This result is to be expected as the AM1 method is not able to satisfactorily predict this property [44].

First hyperpolarizability

In contrast to polarizability, many conjugated organic compounds display zero experimental values or negative calculated values for first hyperpolarizability. Because the logarithmic scale is used to create a normal distribution of the 208 data points, 15 experimental or calculated values cannot be used.

While we performed calculations for the Full Set and for Class I and II, separately; the QSPR model for the Full Set is better than the models for Class I (134 compounds) and II (59 compounds) separately in terms of statistical characteristics and descriptor significance (see Fig. 3).

Molecular descriptors included in the QSPR model listed in Table 2 are essentially molecular orbital and quantum chemical related. As first hyperpolarizability is a complex physico-chemical characteristic that depends directly on molecular parameters such as optical gap, π-conjugated system, frontier orbital energies, the selected variables of QSPR model (Table 2) are appropriate. The order of importance of the selected variables in the QSPR model

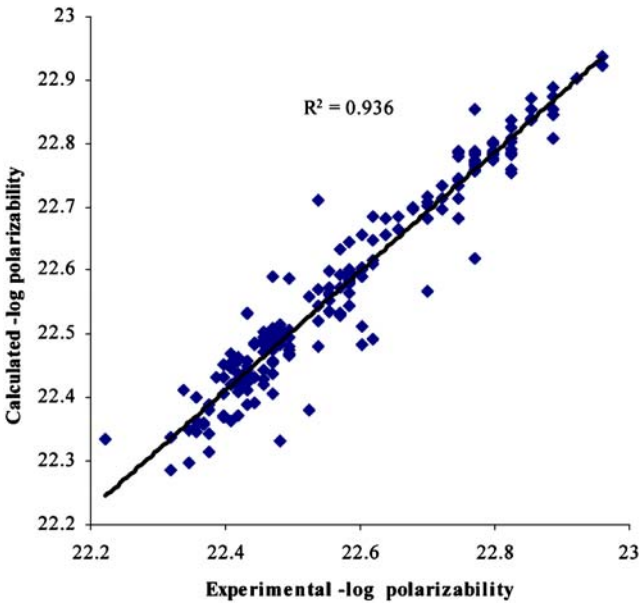

Fig. 2 Scatter plot of 212 AM1-calculated versus experimental values of log polarizability

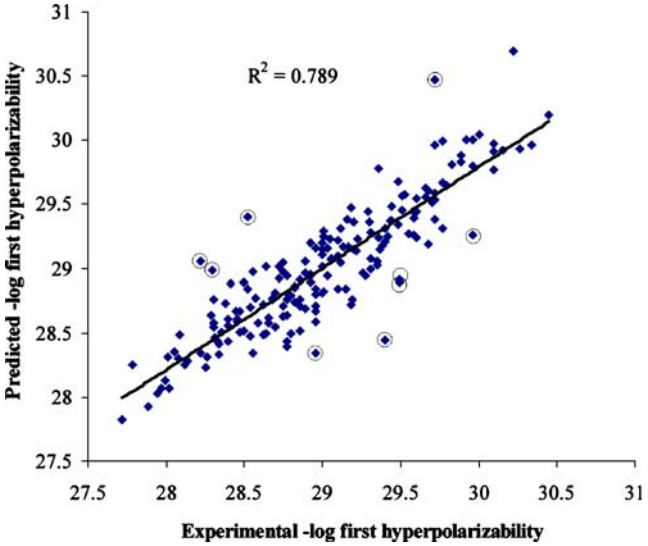

Fig. 3 Plot of 193 QSPR predicted versus experimental values of log first hyperpolarizability (9 outliers are marked)

of Table 2 is D_4>D_5>D_6>D_7 >D_8 according to the *t*-test criteria.

Still, the leading term in the QSPR model given in Table 2 is the logarithm of the AM1 calculated first hyperpolarizability. The expression of first hyperpolarizability in terms of ground and excited state is provided by quantum perturbation theory [45]:

$$\beta_{ijk} = \frac{1}{2}\sum_{P}\sum_{e,e'} \frac{\left\langle 0\left|\mu_i\right|e\right\rangle\left\langle e\left|\mu_j\right|e'\right\rangle\left\langle e'\left|\mu_k\right|0\right\rangle}{(E_e - E_0)(E_{e'} - E_0)} \tag{9}$$

where |e> and E_e refer to an excited state and the energy of the molecule in the excited state, |0> and E_0 to the ground state, and P is any possible permutation of indices.

Frontier orbital energies, in our case HOMO-1 energy (D_5), is important for the description of the low lying electronic excitations of the charge transfer state, which proved to be well correlated with first hyperpolarizability, and can be calculated as follows [12, 37]:

$$E_{HOMO-1} = \left\langle \phi_{HOMO-1}\left|\widehat{F}\right|\phi_{HOMO-1}\right\rangle \tag{10}$$

HOMO−LUMO gap (D_6) represents the optical gap and is also closely related to the first hyperpolarizability; enhanced first hyperpolarizabilities for extended π-conjugated systems are explained by the reduced energy gaps and increased number of hyperpolarizable electrons [13]. Therefore, D_5 has a negative influence in the equation of Table 2.

The total dipole moment of the molecule (D_7) is a charge distribution—related molecular descriptor that is proportional to the first hyperpolarizability as described in Eq. 9. First hyperpolarizability is related to the large difference between the dipole moment in the ground state and excited state; the inclusion of this descriptor in equation from Table 2 is

Table 2 The selected descriptors and statistical parameters for the Full Set of compounds resulting from the best multilinear regression (BMLR) procedure. *AM1* Austin model 1, *HOMO* highest occupied molecular orbital, *LUMO* lowest unoccupied molecular orbital, *SAMNEG* surface area maximum negative charge, *RNCG* relative negative charge

Descriptor	Symbol	b[a]	s[b]	t[c]	$R^{2,d}$	R^2cv[e]	F[f]	s[2,g]
Intercept		−14.904	1.262	−11.810				
AM1 Calculated logarithm of first hyperpolarizability	D_4	0.354	0.053	6.690	0.660	0.639	371.174	0.111
HOMO−1[h] energy	D_5	0.222	0.338	6.566	0.665	0.640	188.921	0.110
HOMO - LUMO energy gap[i]	D_6	−0.263	0.045	−5.847	0.732	0.712	172.128	0.089
Total dipole of the molecule	D_7	0.060	0.014	4.293	0.770	0.731	157.792	0.076
RNCS Relative negative charged SA (SAMNEG*RNCG)	D_8	0.018	0.004	4.044	0.789	0.767	139.812	0.071

[a] Regression coefficients of the linear model
[b] Standard errors for these regression coefficients
[c] *t*-test values for each selected descriptor
[d] Correlation coefficient
[e] Crossvalidated correlation coefficient
[f] Fisher criterion
[g] Errors of the model
[h] The next highest occupied molecular orbital
[i] The difference between the highest occupied molecular orbital and the lowest unoccupied molecular orbital

supported by experimental and theoretical arguments [46]. Dipole moments can be calculated by MOPAC [42] from atomic charges and the lone-pairs as follows:

$$\mu_x = cC\sum_A Q_A x_A + cCa_o 2\sum_A P(s-p_x)_A D_1(A)$$
$$\mu_y = cC\sum_A Q_A y_A + cCa_o 2\sum_A P\left(s-p_y\right)_A D_1(A) \quad (11)$$
$$\mu_z = cC\sum_A Q_A z_A + cCa_o 2\sum_A P(s-p_z)_A D_1(A)$$

$$\mu = \mu_x + \mu_y + \mu_z \quad (12)$$

where c is the speed of light, C the charge of electron, and a_o Bohr radius [37].

The relative negative charge (D_8) is related to the dipole character of the molecule and to the degree of the charge separation, which is directly proportional to the first hyperpolarizability. This model displays nine outliers: entries 25, 26, 27, 29, 33, 64, 65, 110, and 121 (Table 1) that display a large difference between experimental and QSPR predicted values of first hyperpolarizabilily. In order to assess the improvement of the predictions by including additional terms in the QSPR model for the first hyperpolarizability values, we plotted the AM1 calculated versus experimental logarithm of the first hyperpolarizability data (see Fig. 4). The difference between R^2 obtained from the QSPR model (0.789) and R^2 resulting from the dependence between the experimental and AM1 calculated values (0.660) indicate a significant improvement by using the QSPR model.

The QSPR models for the first hyperpolarizability obtained separately for Classes I and II includes five and three descriptors whose origins are quantum chemical- and charge-related, respectively. The model for Class II, which contains benzene derivatives, has lower statistical characteristics and a small number of data than that for Class I, which comprises structurally different compounds. However, the first hyperpolarizability is more complex than polarizability and was more difficult to model. The model for the Full Set was the most stable as the R^2–R^2cv difference is lowest at 0.021 in comparison with Class I (0.059) and Class II (0.107) and the model contains the same number of parameters as the model for Class I (5) (see Table 4).

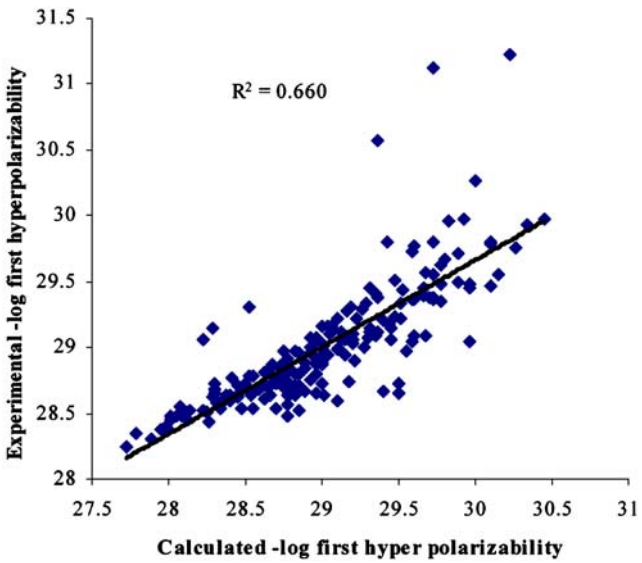

Fig. 4 Plot of 193 AM1 calculated versus experimental values of log first hyperpolarizability

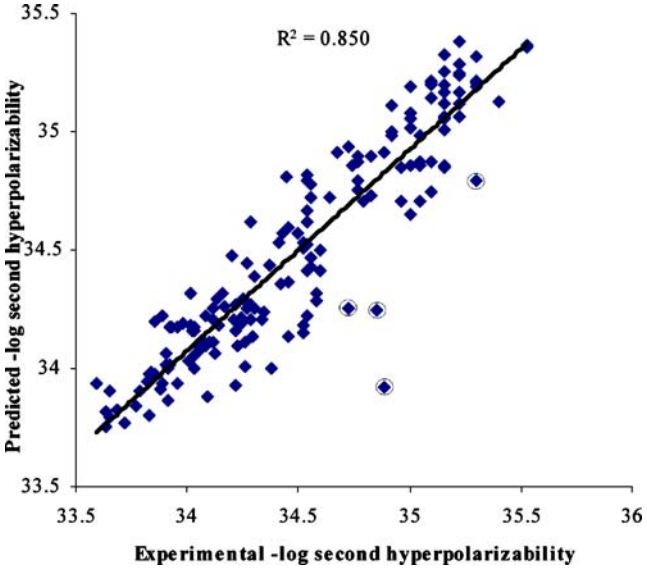

Fig. 5 Scatter plot of 169 QSPR-predicted versus experimental values for log second hyperpolarizability (4 outliers are marked)

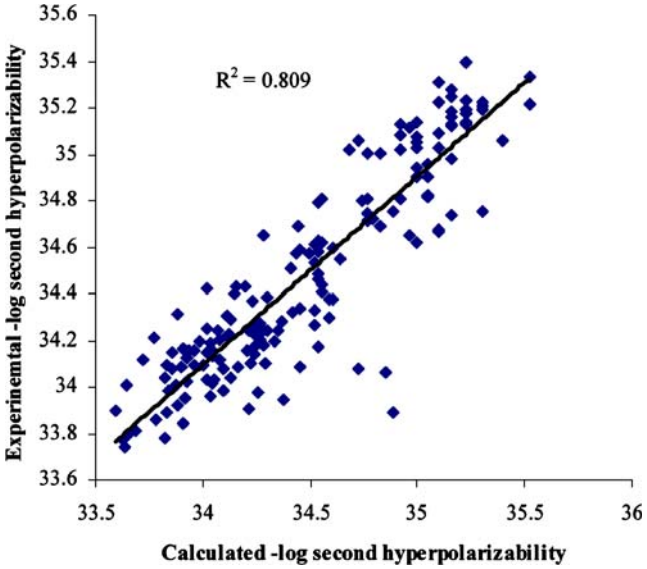

Fig. 6 Plot of 169 AM1-calculated versus experimental values of log second hyperpolarizability

Second hyperpolarizability

In this case, we performed the QSPR modeling for the Full Set and Classes I and II, but the model for the Full Set is better in terms of statistical characteristics, thus we present this Full Set model (Fig. 5) and a short description of the models for Classes I and II.

In the QSPR model of Table 3, three types of molecular descriptors are involved: constitutional (D_{14}), charge-related (D_{15}), and quantum chemical (D_{13}). Their order of importance is: $D_{13}>D_{14}>D_{15}$.

The logarithm of the AM1-calculated second hyperpolarizability (D_{13}) is again the most important descriptor in the QSPR model of Table 3. It is calculated as follows [47]:

$$\gamma_{abcd}(0;0,0) = -Tr\left[H^a D^{abcd}(0,0,0)\right] \tag{13}$$

The number of N atoms (D_{14}) is calculated simply by counting the number of N atoms in the molecule. As most of compounds contain at least one nitrogen atom in the molecule the presence of this descriptor in the model of Table 3 is explainable. FNSA3 fractional atomic charge weighted partial negative surface area (D_{15}) is calculated as [37]:

$$FNSA3 = \frac{PNSA3}{TMSA} \tag{14}$$

$$PNSA3 = \sum_A q_A S_A A \in \{\delta_A < 0\} \tag{15}$$

where q_A is atomic partial charge, S_A is negatively charged solvent-accessible atomic surface area, and TMSA is total molecular surface area. This descriptor is related to the induced charge asymmetry in the molecule and to the TMSA, therefore it manifests a large negative influence in the regression equation of Table 3. The outliers for this model are entries 38, 40, 91, 176 (Table 1).

As in the cases of polarizability and first hyperpolarizabilitiy, we use the plot of experimental versus AM1-calculated second hyperpolarizability (see Fig. 6) to evaluate the effectiveness of the QSPR regression equation in Table 3. The difference between R^2 of the QSPR model and that resulting from the plot in Fig. 6 is 0.041, suggesting a significant improvement.

Table 3 The selected descriptors and statistical parameters resulting from BMLR procedure for the Full Set of compounds. *FNSA-3* Fractional PNSA (PNSA-3/TMSA) (Zefirov PC) Fractional atomic charge weighted partial negative surface area, *PNSA-3* Partial negative surface area, *TMSA* total molecular surface area

Descriptor	Symbol	b	s	t	R^2	R^2cv	F	s^2
Intercept		−8.004	1.560	−5.130				
AM1-calculated logarithm of second hyperpolarizability	D_{13}	0.778	0.054	14.476	0.809	0.805	708.197	0.047
Number of N atoms	D_{14}	0.131	0.021	6.275	0.846	0.840	455.961	0.038
FNSA-3 Fractional PNSA (PNSA-3/TMSA) (Zefirov PC)	D_{15}	−4.308	2.077	−2.074	0.850	0.833	232.649	0.038

Table 4 Leave-one-out validation of the proposed QSPR models

	No compounds	No descriptors	R^2	R^2cv	$\Delta R^2 = R^2 - R^2$cv
Polarizability					
Class I	133	4	0.864	0.845	0.019
Class II	79	4	0.915	0.905	0.010
Full set	212	3	0.941	0.937	0.004
First hyperpolarizability					
Class I	134	5	0.648	0.589	0.059
Class II	59	3	0.544	0.437	0.107
Full set	193	5	0.789	0.768	0.021
Second hyperpolarizability					
Class I	118	5	0.766	0.740	0.026
Class II	51	5	0.802	0.731	0.071
Full set	169	3	0.850	0.833	0.017

The QSPR model obtained for the Full Set display fewer descriptors (3) and better statistics than those for Class I and II and the difference R^2–R^2cv for the Full Set model is lower (0.017) than for Classes I–II (0.026, 0.071) and suggests that stability and predictive power are better (see Table 4).

The QSPR modeling of second hyperpolarizability is less precise than the modeling of polarizability; this is reflected in the statistical characteristics of the models obtained. The number of selected parameters is five in both cases. In the case of Class I their nature is quantum-chemical- or charge-distribution-related, while for Class II a size-related descriptor is included. The correlation coefficient of the model obtained for Class II is better than that for Class I because of the structural complexity of the compounds included in the Class I. However, the difference R^2–R^2cv is larger for the Class II model (0.071) as compared to Class I (0.026), which suggests a better predictive power for the latter model (see Table 4).

Validation of QSPR models

Leave-one-out

The first technique applied for the validation of the proposed QSPR models was cross-validation, namely the leave-one-out algorithm. The corresponding squared cross-validated correlation coefficient (R^2_{cv}) for all selected models, which is calculated automatically by the validation module implemented in the CODESSA Pro package, is shown in Table 4. The squared correlation coefficient of the model is also given in order to emphasize the quality of the models.

A B C internal validation

The A B C internal validation technique predicts the property values for each one-third of the compounds, with the model fitted for the remaining two-thirds of the

Table 5 Internal validation of the models—statistical characteristics.

Set to fit	R^2fit	s^2fit	Set to predict	R^2 predict	s^2predicted
Polarizability					
Full Set					
C_{1A}+C_{1B}	0.936	0.002	C	0.950	0.0013
C_{1A}+C_{1C}	0.954	0.001	B	0.916	0.0025
C_{1B}+C_{1C}	0.933	0.002	A	0.952	0.0013
Average	0.941	0.0016		0.939	0.0017
First hyperpolarizability					
Full Set					
C_{1A}+C_{1B}	0.790	0.071	C	0.782	0.075
C_{1A}+C_{1C}	0.790	0.075	B	0.775	0.069
C_{1B}+C_{1C}	0.798	0.065	A	0.769	0.091
Average	0.793	0.070		0.775	0.078
Second hyperpolarizability					
Full set					
C_{1A}+C_{1B}	0.840	0.039	C	0.866	0.035
C_{1A}+C_{1C}	0.819	0.046	B	0.937	0.025
C_{1B}+C_{1C}	0.899	0.025	A	0.764	0.069
Average	0.853	0.037		0.856	0.043

compounds. This procedure was applied to both subsets (Classes I and II) and the Full Set of compounds by using the corresponding QSPR models (Tables 2, 3).

The general algorithm of the internal validation is as follows:

1. Division of the data set to be analyzed into three sets: C_iA (the 1st, 4th, 7th, etc. entries), C_iB (the 2nd, 5th, 8th, etc. entries) and C_iC (the 3rd, 6th, 9th, etc. entries)
2. In each of three combinations, two of the sets are combined into one and the correlation equation with the same descriptors, as in the QSAR model to be validated, is derived
3. The equation developed in step (2) is used to predict the property values for the remaining set
4. A final comparison of the average of squared correlation coefficients for the fitted and predicted sets

The results of the internal validation applied to our data are listed in Table 5.

Conclusions

The relationship of the (hyper)polarizabilities to molecular structure for a series of conjugated molecules has been investigated using molecular descriptors calculated by CODESSA Pro software. Validation by the leave-one-out method and internal validation demonstrated good stability and significant predictive ability for all developed models. The best models obtained were those for the Full set of compounds, which cover the whole range of substitution patterns, geometries and inductive effects. The regression equations provide insight into the structural features that influence (hyper)polarizability. In accordance with experimental determination and quantum chemical studies, polarizability is influenced predominantly by molecular size and charge transfer. The first hyperpolarizability is well correlated with optical gap (HOMO−LUMO) and orbital energies (HOMO−1) showing the influence of orbital energies on nonlinear properties. The magnitude of the HOMO−LUMO gap is important for electronic transitions between the occupied orbitals (HOMO, HOMO−1) of the donor to the vacant orbitals of the acceptor, for the first excited state, and implicitly on nonlinear responses of the molecules. The first hyperpolarizability is well correlated with the dipole moment, which makes a major contribution to the first excited state energy. A significant improvement of the AM1 calculations was achieved for the first and second hyperpolarizability, whereas polarizability values were not improved by additional QSPR descriptors. In addition to their theoretical interest, these predictive models of (hyper)polarizability should be of interest in the development of new nonlinear optical materials and their applications.

References

1. Karelson M, Lobanov VS, Katritzky AR (1996) Chem Rev 96:1027–1043
2. Devillers J, Balaban AT (1999) Topological indices and related descriptors in QSAR/QSPR. Gordon & Breach, New York, pp 59–167
3. Bosque R, Sales J (2002) J Chem Inf Comput Sci 42:1154–1163
4. Davis D, Sreekumar K, Sajeev Y, Pal S (2005) J Phys Chem B 109:14093–14101
5. Marder SR, Kippelen B, Jen AKY, Peyghambarian N (1997) Nature 388:845–851
6. Oudar JL, Chemla DS (1977) J Chem Phys 66:2664–2668
7. Zyss J, Ledoux I (1994) Chem Rev 94:77–105
8. Kanis DR, Ratner MA, Marks TJ (1994) Chem Rev 94:195–242
9. Adant C, Brédas JL, Dupuis M (1997) J Phys Chem A 101: 3025–3031
10. Jensen L, Åstrand PO, Sylvester-Hvid KO, Mikkelsen KV (2000) J Phys Chem A 104:1563–1569
11. Hansen T, Jensen L, Åstrand PO, Mikkelsen KV (2005) J Chem Theory Comput 1:626–633
12. Cheng LK, Tam W, Stevenson SH, Meredith GR, Rikken G, Marder SR (1991) J Phys Chem 95:10631–10643
13. Cheng LK, Tam W, Marder SR, Steigman AE, Rikken G, Spangler CW (1991) J Phys Chem 95:10643–10652
14. Ganthy TK, Gosh SK (1996) J Phys Chem 100:12295–12298
15. Chattaraj PK, Fuentealba P, Jaque P, Toro-Labbé A (1999) J Phys Chem 103:9307–9312
16. Pogliani L (2003) New J Chem 27:919–927
17. Carrasco-Velar R, Padrón JA, Galvez J (2004) J Pharm Pharmaceut Sci 7:19–26
18. Padrón JA, Carasco R, Pellón RF (2002) J Pharm Pharmaceut Sci 5:258–266
19. Verma RP, Hansch C (2005) Bioorg Med Chem 13:2355–2372
20. Hansch C, Kurup A (2003) J Chem Inf Comput Sci 43:1647–1651
21. Hansch C, Steinmetz WE, Leo AJ, Mekapati SB, Kurup A, Hoekman D (2003) J Chem Inf Comput Sci 43:120–125
22. Verma RP, Kurup A, Hansch C (2005) Bioorg Med Chem 13: 237–255
23. Jha PC, Anusooya Pati Y, Ramasesha S (2005) Mol Phys 103:1859–1873
24. Marder SR, Gorman CB, Meyers F, Perry JW, Bourhill G, Brédas JL, Pierce BMA (1994) Science 265:632–635
25. Wu W, Ye C, Wang D (2003) ARKIVOC 59–69
26. Lučić B, Bašic I, Nadramija D, Milicević A, Trinajstić N, Suzuki T, Petrukin R, Karelson M, Katritzky AR (2002) ARKIVOC 45–49
27. Katritzky AR, Lobanov V, Karelson M (1998) J Chem Inf Comput Sci 38:28–41
28. Katritzky AR, Fara DC, Hongfang Y, Karelson M (2004) Chem Rev 104:175–198
29. Katritzky AR, Ignachenko E, Barcock R, Lobanov V, Karelson M (1994) Anal Chem 66:1799–1807
30. Huibers PDT, Lobanov VS, Katritzky AR, Shah OD, Karelson M (1996) Langmuir 12:1462–1470
31. Huibers PDT, Lobanov VS, Katritzky AR, Shah OD (1997) J Colloid Interface Sci 187:113–120
32. Katritzky AR, Fara DC, Yang H, Karelson M, Suzuki T, Solov'ev VP, Varnek A (2004) J Chem Inf Comp Sci 44:529–541
33. Katritzky AR, Kuanar M, Fara DC, Karelson M, Acree WE (2004) Bioorg Med Chem 12:4375–4748
34. http://www.Hyper.com
35. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
36. http://www.codessa-pro.com

37. Karelson M (2000) Molecular descriptors in QSAR/QSPR. Wiley, New York, pp 229–230, 280–282
38. Mikkelsen KV, Luo Y, Ågren H, Jørgensen P (1994) J Chem Phys 100:8240–8250
39. Osted A, Kongsted J, Mikkelsen KV, Christiansen O (2004) J Phys Chem A 108:8646–8658
40. Sylvester-Hvid KO, Mikkelsen KV, Jonsson D, Norman P, Ågren H (1998) J Chem Phys 109:5576–5584
41. Poulsen TD, Ogilby PR, Mikkelsen KV (2002) J Chem Phys 116:3730–3738
42. http://www.cachesoftware.com/mopac/Mopac2002manual/node439.html
43. Hinchliffe A, Nikolaidi B, Machado HJS (2004) Int J Mol Sci 5:224–238
44. Kagawa H, Ichimura A, Kamka NA, Mori K (2001) J Mol Struct (THEOCHEM) 546:127–141
45. Zyss J (1979) J Chem Phys 70:3333–3340
46. Marder SR, Kippelen B, Jen AKY, Peyghambarian N (1997) Nature 388:845–851
47. Karna SP, Dupuis M (1991) Int J Comput Chem 12:487–504