PARAMETRIZATION OF THE SCALING FACTOR IN THE QUASI-RELATIVISTIC INDO METHOD BASED ON THE EQUILIBRIUM GEOMETRIES AND VIBRATIONAL FREQUENCIES OF ELECTRONEUTRAL CLOSED-SHELL XY₄ MOLECULES POSSESSING THE T_d SYMMETRY

Ladislav TURI NAGY^a, Marek LISKA^a and Daniel TUNEGA^{a,b}

^a Institute of Inorganic Chemistry, Slovak Academy of Sciences, 842 36 Bratislava, The Slovak Republic ^b Department of Physical Chemistry, Slovak Technical University, 812 37 Bratislava, The Slovak Republic

> Received October 21, 1993 Accepted March 25, 1994

The parameter *K* in the quasi-relativistic INDO (QR-INDO) relation $h_{iAjB}^{core} = \frac{1}{2} K S_{i,j} (\beta_i^A + \beta_j^B)$ was optimized by applying the QR-INDO method to the optimization of the geometries and calculation of vibrational wavenumbers of 36 molecules of the XY₄ type with X = C, Si, Ti, Ge, Zr, Sn, Hf and Pb, and Y = H, F, Cl, Br and I. The calculations were performed for 10 values of *K* uniformly spaced across the region of 1.1 to 2.0. The optimum value with which the best fit of the bond lengths and vibrational frequencies was achieved was K = 1.5. This is very close to the value of 1.4 recommended in the literature. The study was performed using recent tabulated gas-phase data. The application of the variable scaling approximation approach was found generally unsuitable.

It is well known that non-relativistic methods, both *ab initio* and semiempirical, fail to successfully describe systems containing heavy atoms. The routine use of relativistic *ab initio* methods, on the other hand, meets with severe problems of technical (and also economic) nature. Despite the enormous success of *ab initio* methods, semiempirical methods still remain a competitive alternative for treating molecules which contain heavy atoms. A quasi-relativistic version of the INDO method (QR-INDO) has been developed by Boca recently¹.

A typical semiempirical method, QR-INDO suffers from all the shortcomings following from the generation of a universal and internally consistent parametrization. One serious problem consists in a proper choice of the target properties whose good fit is the criterion of optimum parameters. There is a certain analogy in *ab initio* calculations where vibrational frequencies serve as the target properties. In this case, the calculated force constants are scaled by empirical factors to achieve a good fit to the observed frequencies. The scaling factor can be the same for all force constants, or different scaling factors are applied to different types of vibration (stretching, bending, etc.), or each different force constant is assigned an individual scaling factor (refs^{2–5} and references therein).

The choice of experimental data for calibration is a next serious problem, particularly with respect to their origin and reliability. Moreover, in the case of systems containing heavy atoms (metals in particular), gas-phase data are frequently unavailable. A successful parametrization requires the use of a large set of data. Often such a large set exists but not all of the data pertain to the gaseous phase. In some cases it is actually more appropriate to use smaller sets of homogeneous data than larger sets of inhomogeneous data. The application of inhomogeneous sets results in statistical noise and bias in the parameters determined; this price for the maximum possible universality is too high.

In the QR-INDO method, the two-center off-diagonal matrix elements of the operator \mathbf{h}^{core} are calculated by the relation

$$h_{\rm iAjB}^{\rm core} = \frac{1}{2} K S_{\rm i,j} \left(\beta_{\rm i}^{\rm A} + \beta_{\rm j}^{\rm B}\right) , \qquad (1)$$

where $S_{i,j}$ is the overlap integral, β_i^A and β_j^B characterize the quality of the *i*-th and *j*-th atomic orbitals of atoms A and B, respectively, and *K* is the scaling factor. This scaling factor is the only adjustable parameter (the β values are not¹). The problems mentioned above give rise to a certain ambiguity in a suitable choice from among the *K* values recommended in ref.¹. Among them is K = 1.4, which was obtained by fitting the equilibrium bond lengths for a wide variety of compounds. Experimental data were taken from refs^{6,7}. The application of the variable scaling factor approach has also been suggested in ref.¹; here the scaling factor is modified according to the atomic electrone-gativities χ_A , χ_B :

$$K_{\rm A,B} = \chi_{\rm A} / \chi_{\rm B} , \quad \chi_{\rm A} > \chi_{\rm B} . \tag{2}$$

This approach is recommended for extremely polar, almost ionic bonds. The treatment in ref.¹ was based on a database of experimental structures, many of which pertained to the condensed phase.

The aim of the present paper was to find a K value which provides the best fit to both the observed equilibrium geometries and vibrational frequencies, which reflect the curvature of the energy surface in the near environment of the local minima. Gas-phase data are employed for this; the effect of the condensed phase is better included in some other way.

1902

THEORETICAL

The optimization of the scaling factor K should lead to correct positions of the local extrema on the energy surface. The result of optimization is unaffected by linear transformation of the energy surface, i.e. by a constant shift and/or alternative choice of the energy scale. The optimum K value corresponds to the minimum of the target function

$$U_{\rm g}(K) = \sum_{\rm i,j} \left[P_{\rm i,j}^{\rm exp} - P_{\rm i,j}^{\rm calc}(K) \right]^2 , \qquad (3)$$

where the summation runs over all internal coordinates j of molecules i, and $P_{i,j}^{exp}$ is a target geometrical parameter. Such an approach is consistent with the character of semiempirical methods, which should provide a good forecast of trends of various properties while the absolute accuracy of the values is less strictly required.

Vibrational wavenumbers are proportional to the square roots of the force constants. Force constants include the second derivatives of the energy surface with respect to the internal coordinates. A straightforward inclusion of vibrational wavenumbers in the optimization procedure preserves the invariancy with respect to a constant shift of the energy surface but the invariancy with respect to the energy scaling is lost. Since the "semiempirical attributes" of the parametrization should not change, it is necessary to preserve the invariancy with respect to the energy scale as well, also if the vibrational wavenumbers are to be fitted. This requirement is met by the following target function,

$$U_{\rm f}(K) = \min_{\eta} \left\{ \sum_{i,j} \left[\omega_{i,j}^{\rm exp} - \eta \; \omega_{i,j}^{\rm calc}(K) \right]^2 \right\} = \sum_{i,j} \left[\omega_{i,j}^{\rm exp} - \eta^*(K) \; \omega_{i,j}^{\rm calc}(K) \right]^2 \;, \tag{4}$$

where the subscripts *i* and *j* refer to the molecules and their vibrational modes, respectively. The $\eta^*(K)$ value, for a given *K*, is calculated as

$$\eta^*(K) = \sum_{i,j} \omega_{i,j}^{\exp} \omega_{i,j}^{\text{calc}}(K) / \sum_{i,j} \left[\omega_{i,j}^{\text{calc}}(K) \right]^2 .$$
⁽⁵⁾

A *K* value which is optimum with respect to both target functions $U_g(K)$ and $U_f(K)$ can be obtained by minimization of the combined target function

$$U(K) = w_g U_g(K) + (1 - w_g) U_f(K) , \qquad (6)$$

where w_g is a weighting factor for the geometric criteria ($0 \le w_g \le 1$). By continuously changing this weighting factor from 0 to 1 we pass smoothly from the wavenumber

criterion to the geometric criterion. The choice of a proper w_g is actually a problem, which is also subjectively biassed. On the other hand, the freedom in the choice of w_g can be of assistance when the reliability of experimental data is to be taken into account.

RESULTS

The parametrization was accomplished using a set of 36 electroneutral XY₄ closed-shell molecules possessing the tetrahedral equilibrium geometry: X = C, Si, Ti, Ge, Zr, Sn, Hf and Pb; Y = H, F, Cl, Br, I. Sufficiently reliable gas-phase geometries and vibrational wavenumbers are available for them^{8,9}. The study was confined to closed-shell molecules to exclude the possibility of occurrence of degenerate ground states and lowering of symmetry of the equilibrium geometry. The X–Y bond length is then the single geometric parameter. Moreover, the vibrational analysis is facilitated by the preservation of the T_d symmetry. In the harmonic approximation, the energy surface is defined by mere five force constants *F*, viz.

$$2 V = F(A_1) S^2(A_1) + F(E) (S_a^2(E) + S_b^2(E)) + F_1(T_2) (S_{1a}^2(T_2) + S_{1b}^2(T_2) + S_{1c}^2(T_2)) + F_2(T_2) (S_{2a}^2(T_2) + S_{2b}^2(T_2) + S_{2c}^2(T_2)) + 2 F_{12}(T_2) (S_{1a}(T_2) S_{2a}(T_2) + S_{1b}(T_2) S_{2b}(T_2) + S_{1c}(T_2) S_{2c}(T_2)) ,$$
(7)

where the symmetry coordinates are defined following Cyvin¹⁰:

$$S(A_{1}) = \frac{1}{2} (r_{1} + r_{2} + r_{3} + r_{4})$$

$$S_{a}(E) = 12^{-1/2} R (2 \alpha_{23} - \alpha_{12} - \alpha_{13} + 2 \alpha_{14} - \alpha_{34} - \alpha_{24})$$

$$S_{b}(E) = 1/2 R (\alpha_{13} - \alpha_{12} + \alpha_{24} - \alpha_{34})$$

$$S_{1a}(T_{2}) = 1/2 (r_{1} - r_{2} + r_{3} - r_{4})$$

$$S_{1b}(T_{2}) = 1/2 (r_{1} + r_{2} - r_{3} - r_{4})$$

$$S_{1c}(T_{2}) = 1/2 (-r_{1} + r_{2} + r_{3} - r_{4})$$

$$S_{2a}(T_{2}) = 2^{-1/2} R (\alpha_{24} - \alpha_{13})$$

$$S_{2b}(T_{2}) = 2^{-1/2} R (\alpha_{34} - \alpha_{12})$$

$$S_{2c}(T_{2}) = 2^{-1/2} R (\alpha_{14} - \alpha_{23}) . \qquad (8)$$

Collect. Czech. Chem. Commun. (Vol. 59) (1994)

Here *R* is the equilibrium bond length and r_i and $\alpha_{i,j}$ are the differences of the X–Y bond lengths and Y_i–X–Y_i bond angles, respectively, from their equilibrium values.

The equilibrium geometries were obtained by direct minimization of the total SCF energy in the T_d geometry using the program MOSEM7 (ref.¹¹). In the symmetry coordinate basis, the harmonic force constants were obtained by mapping the energy surface for the equilibrium geometry. Bond length changes r_i of ±0.008 and ±0.012 . 10^{-10} m and bond angle changes $\alpha_{i,j}$ of ±1° and ±2.5° were used for the mapping. The standard deviations of the force constant estimates were calculated by ordinary linear regression analysis. The kinematic Wilson matrix was set up using relativistic atomic masses corresponding to the natural isotopic composition.

All calculations were performed for 10 values of K within the range of 1.1 to 2.0 with a step of 0.1. The K value calculated by the variable scaling formula (2) was applied as well. The criteria $U_{g}(K)$ and $U_{f}(K)$, calculated for the various K values, were evaluated separately. Three sets of experimental data were employed: data in ref.⁷, data in the JANAF Tables, 2nd edition of 1971 (ref.8) and data in the same tables, 3rd edition of 1985 (ref.⁹). The data set from ref.⁷ includes equilibrium geometries and vibrational wavenumbers for CH₄, CF₄, CCl₄, CBr₄, CI₄, GeH₄, GeF₄, GeCl₄, GeBr₄, Gel₄, HfF₄, HfCl₄, HfBr₄, HfI₄, PbF₄, PbCl₄, PbBr₄, PbI₄, SiH₄, SiF₄, SiCl₄, SiBr₄, SiI₄, SnH₄, SnF₄, SnCl₄, SnBr₄, SnI₄, TiF₄, TiCl₄, TiBr₄, TiI₄, ZrF₄, ZrCl₄, ZrBr₄ and ZrI₄; some pertain to the gaseous phase, other to the crystal phase. The JANAF data pertain all to the gaseous phase. Tables from ref.⁸ provide data for CH₄, CF₄, CCl₄, CBr₄, PbF₄, PbCl₄, PbBr₄, PbI₄, SiH₄, SiCl₄, TiBr₄, TiI₄, ZrBr₄, and ZrI₄, and tables from ref.⁹ provide data for CH₄, CF₄, CCl₄, CBr₄, PbF₄, PbCl₄, PbBr₄, PbI₄, SiH₄, SiF₄, SiCl₄, SiBr₄, SiI₄, TiF₄, TiCl₄, TiBr₄, TiI₄, ZrF₄, ZrCl₄, ZrBr₄, and ZrI₄. The latter set also corrects some values from the former edition. The $U_{\sigma}(K)$ and $U_{f}(K)$ values were normalized by dividing them by the number of experimental data used. For graphical representation, they were normalized further so that the maximum in each plot have the value of 1.

The dependences of U_g on K for the three sets of experimental data are plotted in Fig. 1. It is clear that the "variable scaling" parametrization gives unsatisfactory results. The lowest minimum is obtained for the data set of ref.⁸; the minimum U_g is attained at K = 1.2. Somewhat higher is the minimum value of the function for the set from ref.⁹, which lies at K = 1.5. The data set from ref.⁷ provides the highest minimum U_g value, which is found at K = 1.4. The analogous plots for the U_f criterion are shown in Fig. 2. The "variable scaling" approach gives the poorest results in all cases. All the three $U_f(K)$ plots descend monotonously with the growing K. Thus, higher values of K are preferred for lower $U_f(K)$. Lower U_f values are obtained for the data sets of refs^{7,9}; the two curves approach each other closely. This fact is of importance in view of the different experimental data involved.

The $U_{\rm f}(K)$ plots do not enable an optimum K value to be estimated based on the position of the local extremum of $U_{\rm f}$, but the $U_{\rm f}$ values can help us to decide which set

1906

TABLE I

Force constants (in 10^2 N m⁻¹; upper rows) and their standard deviations (lower rows) for the optimum value of K = 1.5

Molecule	$F(A_1)$	F(E)	$F_{1}(T_{2})$	$F_{2}(T_{2})$	$F_{12}(T_2)$
CBr ₄	17.7668	0.5530	12.4302	0.7916	-0.9633
	0.5287	0.0018	0.1170	0.0058	0.0192
CCl ₄	19.8593	0.6261	14.0218	0.9100	-1.0628
	0.4807	0.0019	0.1073	0.0061	0.0192
CF ₄	32.1578	0.8743	24.8598	1.2929	-1.3314
	0.3525	0.0022	0.0801	0.0070	0.0176
CH4	16.9865	0.8761	15.9860	0.9925	-0.5186
	0.1490	0.0015	0.0336	0.0046	0.0093
CI4	14.7295	0.4623	10.2054	0.6477	-0.8185
	0.6249	0.0018	0.1378	0.0056	0.0208
GeBr ₄	9.1726	0.2116	7.1915	0.2989	-0.3710
	0.2564	0.0006	0.0561	0.0019	0.0075
GeCl ₄	10.1576	0.2278	8.1031	0.3276	-0.3852
	0.2564	0.0007	0.0577	0.0021	0.0082
GeF ₄	12.4164	0.1683	11.2547	0.2485	-0.2080
	0.1762	0.0006	0.0385	0.0018	0.0062
GeH ₄	7.5364	0.3557	7.1578	0.3708	-0.2480
	0.1057	0.0006	0.0240	0.0018	0.0048
GeI ₄	8.1358	0.2009	6.1686	0.2792	-0.3605
	0.2724	0.0006	0.0609	0.0018	0.0079
$HfBr_4$	9.0334	0.2842	7.2283	0.2953	-0.0405
	0.2564	0.0006	0.0561	0.0018	0.0075
HfCl ₄	9.9466	0.3216	8.0301	0.3269	-0.0079
	0.2243	0.0006	0.0497	0.0018	0.0070
HfF_4	12.2294	0.2570	9.9470	0.2323	0.1729
	0.0208	0.0001	0.0045	0.0002	0.0007
HfI_4	7.9347	0.2531	6.3200	0.2800	-0.1054
	0.3204	0.0006	0.0705	0.0021	0.0088
PbBr ₄	7.4693	0.1439	6.1837	0.1984	-0.2525
	0.2083	0.0004	0.0449	0.0013	0.0058
PbCl ₄	8.1522	0.1521	6.8529	0.2129	-0.2567
	0.1923	0.0004	0.0433	0.0014	0.0058
PbF ₄	9.4889	0.0994	8.9473	0.1451	-0.1080
	0.0929	0.0003	0.0208	0.0008	0.0030
PbI ₄	6.6712	0.1442	5.3225	0.1946	-0.2594
	0.2083	0.0004	0.0465	0.0012	0.0056

TABLE I	
(Continued)	

Molecule	$F(A_1)$	F(E)	$F_{1}(T_{2})$	$F_2(T_2)$	$F_{12}(T_2)$
SiBr ₄	9.0106	0.2190	6.9427	0.3139	-0.3784
	0.2564	0.0006	0.0593	0.0021	0.0082
SiCl ₄	9.8510	0.2350	7.7249	0.3420	-0.3911
	0.2564	0.0007	0.0577	0.0022	0.0083
SiF ₄	11.6305	0.1730	10.3781	0.2582	-0.2112
	0.1602	0.0005	0.0369	0.0018	0.0059
SiH_4	7.5938	0.3791	7.2477	0.4022	-0.2361
	0.1122	0.0006	0.0256	0.0019	0.0053
SiI_4	7.9911	0.2089	5.9583	0.2940	-0.3710
	0.2724	0.0006	0.0625	0.0019	0.0080
SnBr ₄	7.5856	0.1463	6.2874	0.2048	-0.2495
	0.2243	0.0005	0.0497	0.0015	0.0064
SnCl ₄	8.3212	0.1547	7.0035	0.2205	-0.2533
	0.2083	0.0004	0.0449	0.0014	0.0059
SnF_4	9.5180	0.1020	8.9558	0.1519	-0.1093
	0.0721	0.0002	0.0160	0.0006	0.0024
SnH_4	6.0560	0.2659	5.7811	0.2759	-0.1885
	0.0929	0.0004	0.0208	0.0013	0.0038
SnI ₄	6.7643	0.1469	5.4034	0.2012	-0.2587
	0.2243	0.0004	0.0497	0.0013	0.0059
TiBr ₄	5.9932	0.2083	4.3487	0.2275	-0.0226
	0.1586	0.0004	0.0352	0.0012	0.0048
TiCl ₄	6.1525	0.2330	4.3836	0.2539	-0.0009
	0.1490	0.0004	0.0336	0.0012	0.0046
TiF ₄	4.9481	0.1574	2.6366	0.1801	0.0706
	0.0817	0.0003	0.0176	0.0008	0.0029
TiI ₄	5.7131	0.1853	4.2516	0.2127	-0.0709
	0.2083	0.0004	0.0465	0.0013	0.0058
$ZrBr_4$	7.2312	0.2341	5.5696	0.2404	-0.0064
	0.1923	0.0004	0.0449	0.0014	0.0059
ZrCl ₄	7.7619	0.2612	5.9942	0.2648	0.0217
	0.1923	0.0004	0.0433	0.0015	0.0058
ZrF_4	8.1667	0.1797	5.9858	0.1833	0.1366
	0.0881	0.0003	0.0192	0.0008	0.0030
ZrI_4	6.5692	0.2145	5.1061	0.2328	-0.0635
	0.2724	0.0005	0.0593	0.0016	0.0074

of data to employ. The data from ref.⁸ do not seem suitable for the parametrization. As the U_g function was calculated only for a discrete set of K values, the U_g minimum at K = 1.5 formed within this set evidently may not be the true one. The closer inspection of the shape of the U_g -curve (Fig. 1) shows that the true minimum of U_g for data from ref.⁹ is somewhere between K = 1.4 and 1.5. However, adding now the monotonously descending function U_f (according to relation (6)) will slightly shift the minimum of the resulting function U back towards higher K values whatever is the weighting factor w_g . Thus, taking the value of K = 1.5 as a minimum of U seems to be acceptable for any w_g . The equilibrium bond lengths obtained with this K value for the 3 sets of data are shown in Figs 3 – 5, respectively, as plots of calculated versus observed values. The



Fig. 1

1908

Dependence of the criterion U_g on the *K* value. Data set: \blacksquare ref.⁷, \bullet ref.⁸, \bigcirc ref.⁹. VS denotes the variable scaling approximation of *K*



FIG. 3

Calculated bond lengths plotted against observed data (ref.⁹)



Fig. 2

Dependence of the criterion $U_{\rm f}$ on the *K* value. Data set: \blacksquare ref.⁷, \bullet ref.⁸, \bigcirc ref.⁹. VS denotes the variable scaling approximation of *K*



Calculated bond lengths plotted against observed data (ref. 8)

linear best fit lines are also shown for illustration. The calculated vibrational wavenumbers (scaled by a factor of $\eta^*(1.5) = 0.512$ obtained for K = 1.5 using the data from ref.⁹) are compared with the observed values in Fig. 6. All the results demonstrate that the requirement placed on semiempirical methods, viz. that they should provide correct trends, is sufficiently well satisfied by the QR-INDO method. The force constants, along with their standard deviations, calculated for K = 1.5 are given in Table I. The magnitude of the standard deviations demonstrates that the harmonic approximation is fully acceptable to the systems studied.



Fig. 5

Calculated bond lengths plotted against observed data (ref.⁷)



Calculated vibrational wavenumbers plotted against observed data (ref.⁹)

CONCLUSIONS

1) The approach of variable scaling approximation of K is generally unacceptable, although it may give satisfactory results in some particular cases.

2) The optimum value of K = 1.5 obtained by the treatment is very close to the value of 1.4 proposed in ref.¹.

3) The effect of condensed phase should be accounted for by some method different from the parametrization of the dominant contribution to the total SCF electron energy.

REFERENCES

- 1. Boca R.: Int. J. Quantum Chem. 34, 385 (1988); 33, 159 (1988); 31, 941 (1987).
- 2. Durig J. R., Phan H. V., Hardin J. A., Berry J. R., Little T. S.: J. Mol. Struct. 198, 365 (1989).
- 3. Durig J. R., Gvirgis G. A., Phan H. V.: J. Raman Spectrosc. 21, 359 (1990).
- 4. Durig J. R., Lin J., van der Veken B. J.: J. Raman Spectrosc. 23, 287 (1992).

1910

- 5. Durig J. R., Wang A. Y., Little T. S., Brletic P. A.: J. Chem. Phys. 93, 905 (1990).
- 6. Huber K. P., Herzberg G.: *Molecular Spectra and Molecular Structure*, Vol. IV. Van Nostrand Reinhold, New York 1979.
- 7. Krasnov K. C. (Ed.): Molecular Constants of Inorganic Compounds. Khimiya, Leningrad 1979.
- 8. Sull D. R., Prophet H.: JANAF Thermochemical Tables, 2nd ed. NSRDS-NBS, Washington 1971.
- 9. JANAF Thermochemical Tables, 3rd ed.: J. Phys. Chem. Ref. Data 14, Suppl. 1 (1985).
- 10. Cyvin S. J.: Molecular Vibrations and Mean Square Amplitudes. Elsevier, Amsterdam 1969.
- 11. Boca R.: Program MOSEM7. Slovak Technical University, Bratislava 1987.

Translation revised by P. Adamek.