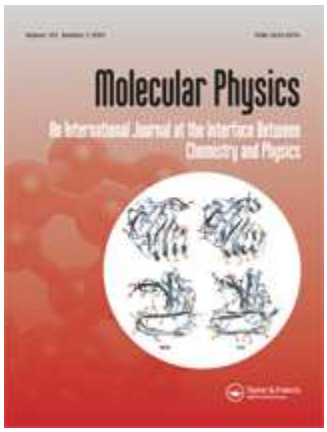
Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6

Zoltan A. Fekete, Eufrozina A. Hoffmann, Tamás Körtvélyesi, Botond Penke



Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6

Journal:	*Molecular Physics*
Manuscript ID:	TMPH-2007-0193.R1
Manuscript Type:	Full Paper
Date Submitted by the Author:	26-Jul-2007
Complete List of Authors:	Fekete, Zoltan; U. Szeged, HPC group HOFFMANN, EUFROZINA; University of Szeged, Department of Physical Chemistry KÖRTVÉLYESI, TAMÁS; University of Szeged, Department of Physical Chemistry PENKE, BOTOND; University of Szeged, Department of Medical Chemistry
Keywords:	Vibrational frequencies, Semiempirical molecular orbital methods, Semiempirical parametrization, Scaled quantum mechanics, Multivariate analysis

Running heads (verso) *Z. A. Fekete* et al.
(recto) *Frequency scaling factors: RM1 and PM6*

Article Type: "Peter Pulay Special Issue" contribution

Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6

ZOLTAN A. FEKETE*†, EUFROZINA A. HOFFMANN‡, TAMÁS KÖRTVÉLYESI†‡ and BOTOND PENKE§

†HPC group, University of Szeged, DNT Szikra u. 2., H-6725 Szeged, Hungary
‡Department of Physical Chemistry, University of Szeged, PO Box 105, H-6720 Szeged, Hungary
§Department of Medical Chemistry, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary

Correspondence *Zoltan A. Fekete. Email: MQM2007.ZAF@inboxclean.com

Abstract

Scaling factors have been derived for obtaining fundamental vibrational frequencies with the recently introduced semiempirical molecular orbital methods RM1 and PM6, implemented in MOPAC2007. A least-squares approach is used with a training set comprised of 90 singlet-state molecules and 922 distinct vibrations, extracted from the NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB). Results are presented both for the conventional Scott-Radom type single-factor fitting, and for a multi-factor linear model: Semiempirical Semiglobal Self-consistently Scaled Quantum Mechanical (S4QM). The new NDDO methods in conjunction with the multi-linear fitting are shown to yield improved prediction of vibrational frequencies. To demonstrate the performance of S4QM//PM6 for calculating vibrational spectra, the examples of indene, indazole and four tetrachlorinated *p*-dibenzodioxins are presented.

Keywords: Vibrational frequencies; Semiempirical molecular orbital methods; Semiempirical parametrization; Scaled quantum mechanics; Multivariate analysis

AMS Subject Classification: 92E99 80A99 81V55 62J05

1. Introduction

Recent years brought on a revival for semiempirical molecular orbital theory [1--7], after a long period of 'consistent reports of its death' [8]. Despite the continuous growth in system sizes available for *ab initio* or DFT treatment, the computational efficiency [9] of semiempirical methods still makes them an attractive alternative for many applications [10--13].

Vibrational frequency calculations by quantum mechanical methods are of major importance in many areas of chemistry. Apart from their most straightforward application, the prediction and interpretation of vibrational spectra, they are crucial in dealing with quantities which depend on the form of vibrations, like infrared and Raman intensities, or the vibrational structure in ultraviolet and photoelectron spectra, as well as vibrational averaging effects on molecular geometries and dipole moments. Another important area is the derivation of thermochemical and kinetic information through statistical thermodynamics.

Computed frequencies typically deviate from experimentally determined ones significantly (with rare exceptions for very high-level calculations, which are only feasible for small molecules due to their extreme computational demand). This has led to the standard practice of scaling the results in order to bring them in line with measured values [14--19]. Two principal types of scaling procedure has emerged in practice. A more convenient, albeit theoretically less justified way is to fit calculated versus experimental data globally, without respect to the structural details involved. A theoretically more sound way is to use the full information content of the quantum mechanical results and scale the fundamental force constants accordingly, as in the Scaled Quantum Mechanical (SQM) procedure by Pulay [16,20,21].

Method development for predicting vibrational frequencies based on semiempirical quantum chemistry [22] has been disfavoured at least since Scott and Radom [18] reported the very poor results of AM1 and PM3 in this respect. This situation may be revised in light of the arrival of the new NDDO methods RM1 [23] and PM6 [24,25]. Both these correct many shortcomings of their predecessors, in particular calculated geometries are much improved. Especially significant is the advancement of PM6 that is parameterized based on a much extended set of data, and incorporates *d*–shell thus extending to the whole periodic table including transition metals. [24]

In this contribution we report on the performance of linearly scaled RM1 and PM6 in predicting vibrational frequencies. Besides showing results with the conventional single–parameter scaling, we are also introducing a multi–parameter protocol that seeks middle ground between the simplicity of global scaling and the detailed mode–specificity of SQM. Using a global fit to frequencies, but incorporating molecular descriptors split according to various types of vibrational modes, the procedure is designated Semiempirical Semiglobal Self–consistently Scaled Quantum Mechanical (S4QM) frequency fitting.

2. Methods

2.1. Data selection procedure

Published experimental vibrational frequencies, as well as geometries, were obtained from the NIST CCCBDB [26]. (Those compounds whose full geometry data is unavailable from the same database were excluded from the current study.) The initial selection has been pruned based on subsequent computations (see section 2.2.). Only species in the singlet electronic state, without spin contamination, have been included in the final analysis. Those polyatomics which are near–linear, *i.e*. have both of their calculated dimensions in the x and y directions (perpendicular to the main molecular axis) smaller than 100 pm, were also excluded. Finally, molecules with excitation energy (computed HOMO–LUMO difference) smaller than 8.00 eV were omitted, too. The dataset so chosen contains 90 molecules and 922 individual frequencies. For an overall

Törölt: exluded

description of this set, we have determined the following arithmetic mean values: there are 10.2 frequencies, 6.8 atoms — of which 3.2 are heavy (non–hydrogen) — and 2.6 elements on the average per species. The constituent elements are (the number of molecules that contain each is listed in parentheses): C(67), H(68), N(20), O(22), F(21), P(8), S(5), Cl(23) and Br(4). It is a characteristics of the CCCBDB that they are mostly small organic molecules with few heteroatoms. Sizes up to a total of 18 atoms (which occurs in cyclo–C_6H_{12}), and up to 8 heavy atoms (in C_2F_6) can be found in this sample.

2.2. Quantum chemical calculations

PM6 and RM1 computations were performed with the MOPAC2007 program package [24]. First, starting from their initial experimental geometry, tightly optimized RHF calculated geometries were determined for all molecules considered. At these geometries, single-point tests were run for spin contamination (with MOPAC keywords '1SCF UHF'), and species with S^2>0.01 were excluded from further consideration. Then bonding parameters were obtained (MOPAC keyword 'BONDS'). We imposed two selection criteria for molecules to be included in the final analysis for this work: no valence of any atom should be larger than 4.25, and no bond order larger than 2.25. For the remaining species, the MOPAC 'FORCE' calculation yielded theoretical harmonic frequencies, as well as the semiempirical vibrational analysis [27] that is utilized to obtain molecular descriptors according the section 2.3.

Törölt: initital

2.3. MOPAC vibrational analysis of Stewart

Normal coordinate calculations in MOPAC provide a supplemental output, with pair–wise atomic partitioning of motions into radial and tangential components. Although details of the scheme were published by its authors [27] long ago, its benefits are rarely recognized. For easy reference, the main points are summarized here:

> The energy absorbed by each atom (E_{AA}, E_{BB}, ...) and the energy absorbed or released by each bond (E_{AB}, E_{BC}, ...) is calculated for each mode. In a given mode, the energy change associated with an atom, E_{AA}, is calculated from its displacement and the force resisting the displacement (the force constants). The energy change associated with the A–B bond, E_{AB}, is calculated from the simultaneous relative displacement of atoms A and B and the net resisting force. E_{AB} may be either positive or negative (unlike the non-negative E_{AA}, E_{BB}, ..., terms). A loose interpretation of this algebraically driven result is that a bond may either absorb part of the energy of the photon stimulating the mode, or it may release energy to the other motions in the mode. The energy for a given pair of atoms is: E(A–B) = E_{AA} + E_{BB} + 2E_{AB}. The total energy for all the pairs of bonded atoms in the molecule in the mode is: E_{tot} =$\Sigma_A\Sigma_B$ E(A–B).

2.4. Regression models

The conventional one–parameter global frequency scaling relation [18,26], taken between theoretical (harmonic) frequencies ${}^{theo}\omega$ and their (anharmonic) experimental counterparts ${}^{obs}\nu$, is given by equation (1):

$$ {}^{obs}\nu = \lambda \cdot {}^{theo}\omega \qquad (1) $$

We introduce an expanded multi-parameter expression (2), based on partitioning equation (1) with a set of molecular descriptors, f_j, utilizing the analysis in section 2.3.

$$ {}^{obs}\nu = \Sigma_j \lambda_j \cdot f_j \cdot {}^{theo}\omega, \quad j=\textbf{Ls, Lbt, Hs, Hbt} \qquad (2) $$

The four descriptors are calculated from the partitioning of energy contributions to the vibrational mode: fractions of stretching (radial motion) and bending+torsional characters (tangential components) are collected from the MOPAC output; respectively **Ls** and **Lbt** are for vibrations involving light atoms (hydrogen isotopes), **Hs** and **Hbt** for those with exclusively heavy atoms. We use the designation S4QM for this model: Semiempirical Semiglobal Self-consistently Scaled Quantum Mechanical frequencies.

3. Results and discussion

3.1. RM1 frequency fitting

Results from fitting RM1 frequencies to equations (1–2) are summarized in table 1. To put the overall errors shown in perspective: Scott and Radom [18] in their study involving 1066 fundamentals determined Δ_{RMS} values of 126 cm^{-1} and 159 cm^{-1}, for scaled frequencies from AM1 and PM3 calculations, respectively.

[Insert table 1 about here]

Fitting for the model denoted S4QM//RM1, *i.e.* equation (2) with RM1, is shown figure 1. The inset of this figure (as well as of those following) has a table with summary statistics for both the absolute and relative deviations, as well as a histogram of the latter. Note that the statistics for relative errors are merely displayed for comparison with similar reports in the literature, but all of the calculations carried out here used non-relative values for fitting. (This causes the average of relative deviations to be further from zero than that of the absolute ones.) Both the visual display of the points scattering around the fitted line, and the statistics (i.e. $\Delta_{max} \approx 3\Delta_{RMS}$) confirm that there are no particular outliers.

Since the fingerprint region (500–2000 cm^{-1}) is often of special interest experimentally, the inset of figure 1 displays the histogram of relative errors tabulated from this interval only.

Because the statistics appear poorer than with the PM6 Hamiltonian (section 3.2.), which is a more sophisticated method regarding its quantum chemistry, we will further discuss only the latter below.

Törölt: Since

[Insert figure 1 about here]

3.2. PM6 frequency fitting

Results from fitting PM6 frequencies to equations (1–2) are summarized in table 2. It is noted here that our 4–factor model shows considerable improvement over the single–factor fitting, unlike in the case with RM1. In either case, results are markedly better than with AM1 or PM3.

[Insert table 2 about here]

Figure 2 visualizes fitting to the S4QM//PM6 model, and the histogram of relative errors tabulated from the 500–2000 cm^{-1} interval is shown in the inset.

[Insert figure 2 about here]

From a practical point view, instead of the overall error describing the fitting across the whole training set as presented above, it is more important to consider the molecular error [18] for individual species. In the following sections examples of utilizing this fitted S4QM//PM6 model are presented. They are all for molecules larger than those in the training set.

3.3. Indene and indazole: examples of S4QM//PM6 prediction for individual molecules

Indene is a compound with a well characterized spectrum, which is often used for calibration purposes either in experimental vibrational spectroscopy [28,29] or in theoretical modelling [30].

[Insert figure 3 about here]

Figure 3 plots the experimentally determined fundamentals [31,32] *vs.* those predicted by the S4QM//PM6 model (sections 2.1., 2.4. and 3.2.). Statistics taken over the fingerprint region are summarized on the inset. We emphasize that no fit is made to the experimental data on indene: unmodified λ_j parameters from table 2 are substituted into equation (2) for the prediction. This is a check for the transferability of the four scaling factors determined on the training set, using no fitted parameter determined specifically in connection with the species.

Similarly to the above, figure 4 presents the S4QM//PM6 results for the indazole molecule (experimental fundamentals are taken from [31,33]). As seen from the structure indicated on the figure, this compound is an indene analogoue that contains geminal nitrogens in a heterocyclic ring.

[Insert figure 4 about here]

Even though this structural unit is completely lacking from the training set used to obtain the model parameters, the frequency predictions appear surprisingly good for this species: in the fingerprint region Δ_{RMS}=38 cm^{-1}, or 4% relative error. For comparison, El-Azhary [31] achieved a fit of Δ_{RMS}=9 cm^{-1} with B3LYP/6-31G** computations and a set of three SQM–type [16,20] scaling factors, which had been refined based on six other analogue structures (also reported a similarly refined single–factor scaled fit of Δ_{RMS}=15 cm^{-1}). On the other hand, the SQM fit by Cane [33], based on HF/6-31G** computations (which are now considered inferior to DFT for frequency predictions [31,34--36]), yielded Δ_{RMS}=22 cm^{-1}. One should keep in mind that both these latter methods require orders of magnitude larger computational times than the semiempirical ones.

3.4. Tetrachlorinated p–dibenzodioxins: examples of S4QM//PM6 predictions for an isomer family

In this section results for a set of four isomer tetrachlorinated *p*–dibenzodioxins (**TCDD**) presented, see figure 5 (which also shows the structure and numbering of their skeleton). The comparison made here is with higher–level (SQM//B3LYP/6-31G(d) [37]) theoretical predictions, rather than with experimental data.

[Insert figure 5 about here]

For both methods a total of 146 frequencies are considered, which fall into the fingerprint region. With their 18 heavy atoms, and multi–substituted aromatic system, these molecules substantially exceed the coverage of our training set. In particular, many features of the **TCDD** spectra are affected by the presence of the chlorine, which being a second–row element is expected to scale differently [37]. Therefore the large, and partly systematic, deviations seen on figure 5 are not surprising. Nevertheless, the overall trend is fairly well reproduced. Moreover, the global part (*i.e*. that spread across all atom types rather than characteristic of chlorine) of the systematic difference between our model and that with the higher–level method can be minimized with a simple linear adjustment.

$$^{adj}\nu = m + b \cdot {}^{S4QM}\omega \qquad (3)$$

With the *a posteriori* modification described by equation (3), the ${}^{S4QM}\omega$ original predictions are brought in line with the target dataset, *via* incorporating two extra parameters by fitting ${}^{adj}\nu$ to the target. The errors from this expanded, six–parameter model are summarized on the inset of figure 5. These are indicative of the limits to the predictive power of this simple S4QM//PM6 method, as specified with the λ_{Ls}, λ_{Lbt}, λ_{Hs}, λ_{Hbt} parameter set given in table 2. Clearly, in order to make reliable predictions for molecules very dissimilar to those included in the training set, the diversity of the data as well as of the parameters should be increased. The partial successes of the initial version of S4QM//PM6 show promise for applying the same protocol for expanding the model this way.

4. Conclusion and outlook

Compared to the errors of scaled semiempirical frequency predictions published in the seminal paper by Scott and Radom [18], the new NDDO methods are improved over their predecessors: single–parameter fitting with RM1 yields Δ_{RMS} 96 — instead of Δ_{RMS}=126 cm^{-1} with AM1; with PM6 Δ_{RMS}=108 — instead of Δ_{RMS}=159 cm^{-1} with PM3. Our novel S4QM fitting gives further reduction of error, most notably with the all–element method PM6 (Δ_{RMS}=88 with four scaling constants). Importantly, all four parameters (λ_{Ls}, λ_{Lbt}, λ_{Hs}, λ_{Hbt}) from S4QM//PM6 are determined with high significance from our modestly sized training set currently utilized. This strongly indicates that systematic further improvement of the statistics will be attainable with an enlarged data set and judiciously augmented parameterization. Therefore linearly scaled semiempirical methods can be made a semiquantitative tool for vibrational frequency prediction. With their improved calibration they will yield *a priori* (though not *ab initio*) fundamental frequencies at very small computational expense, even for large systems.

Acknowledgements

Z.A.F. is grateful for the enthusiastic help of J.J.P. Stewart with his program.
The computing facilities of the HPC group at the University of Szeged were utilized.
Partial financial support was provided by the Hungarian National Office for Research and Technology (grant RET 08/2004), and by the Hungarian National Scientific Research Fund (OTKA grant K61577).

Formázott: Angol (egyesült királysági)

References

[1] M. Kolb; W. Thiel. *J. Comput. Chem.*, **14**, 775 (1993).
[2] W. Thiel. *Adv. Chem. Phys.*, **93**, 703 (1996).
[3] T. Clark. *J. Mol. Struct. THEOCHEM*, **530**, 1 (2000).
[4] J. J. P. Stewart. *J. Mol. Model.*, **10**, 155 (2004).
[5] J. J. P. Stewart. *J. Mol. Model.*, **10**, 6 (2004).
[6] S. Patchkovskii; A. Koslowski; W. Thiel. *Theor. Chem. Acc.*, **114**, 84 (2005).
[7] R. Steiger; C. H. Bischof; B. Lang; W. Thiel. *Future. Gener. Comp. Sy.*, **21**, 1324 (2005).
[8] T. Clark; P. Winget; C. Selcuki; A. Horn; B. Martin. *Abstr. Pap. Am. Chem. Soc.*, **224**, U500 (2002).
[9] J. J. P. Stewart; P. Csaszar; P. Pulay. *J. Comput. Chem.*, **3**, 227 (1982).
[10] P. Murray-Rust; H. S. Rzepa; J. J. P. Stewart; Y. Zhang. *J. Mol. Model.*, **11**, 532 (2005).
[11] A. Monge; A. Arrault; C. Marot; L. Morin-Allory. *Mol. Divers.*, **10**, 389 (2006).
[12] J. Linnanto; J. Korppi-Tommola. *J. Comput. Chem.*, **25**, 123 (2004).
[13] H. M. Senn; W. Thiel. *Top. Curr. Chem.*, **268**, 173 (2007).
[14] C. E. Blom; P. J. Slingerland; C. Altona. *Mol. Phys.*, **31**, 1359 (1976).
[15] C. E. Blom; C. Altona; A. Oskam. *Mol. Phys.*, **34**, 557 (1977).
[16] P. Pulay; G. Fogarasi; G. Pongor; J. E. Boggs; A. Vargha. *J. Am. Chem. Soc.*, **105**, 7037 (1983).
[17] J. A. Pople; A. P. Scott; M. W. Wong; L. Radom. *Israel J. Chem.*, **33**, 345 (1993).
[18] A. P. Scott; L. Radom. *J. Phys. Chem. US*, **100**, 16502 (1996).
[19] V. I. Pupyshev; Y. N. Panchenko; C. W. Bock; G. Pongor. *J. Chem. Phys.*, **94**, 1247 (1991).
[20] G. Fogarasi; P. G. Szalay; P. P. Liescheski; J. E. Boggs; P. Pulay. *J. Mol. Struct. THEOCHEM*, **36**, 341 (1987).
[21] J. Baker; A. A. Jarzecki; P. Pulay. *J. Phys. Chem. A*, **102**, 1412 (1998).
[22] M. B. Coolidge; J. E. Marlin; J. J. P. Stewart. *J. Comput. Chem.*, **12**, 948 (1991).
[23] G. B. Rocha; R. O. Freire; A. M. Simas; J. J. P. Stewart. *J. Comput. Chem.*, **27**, 1101 (2006).
[24] James J. P. Stewart. MOPAC2007, Version 7.0*. Available online at: HTTP://OpenMOPAC.net (accessed 15 Jan. 2007).
[25] J. J. P. Stewart. *J. Mol. Model.*, **accepted**, (2007).
[26] Editor: R.D. Johnson III. NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. Available online at: http://srdata.nist.gov/cccbdb (accessed 15 Dec. 2006).
[27] C.J. Dymek; J.J.P. Stewart. *Inorg. Chem.*, **28**, 1472 (1989).
[28] IUPAC Commission of Molecular Structure and Spectroscopy. *Tables of Wavenumbers for the Calibration of Infrared Spectrometers*, Pergamon Press, New York, (1977).
[29] E. N. Lewis; V. F. Kalasinsky; I. W. Levin. *Anal. Chem.*, **60**, 2658 (1988).
[30] H. Yoshida; K. Takeda; J. Okamura; A. Ehara; H. Matsuura. *J. Phys. Chem. A*, **106**, 3580 (2002).
[31] A. A. El-Azhary. *Spectrochim. Acta A*, **55**, 2437 (1999).
[32] T. D. Klots. *Spectrochim. Acta A*, **51**, 2307 (1995).
[33] E. Cane; P. Palmieri; R. Tarroni; A. Trombetti. *J. Chem. Soc .Faraday. T.*, **89**, 4005 (1993).
[34] G. Rauhut; P. Pulay. *J. Phys. Chem. US*, **99**, 3093 (1995).
[35] J. Baker; P. Pulay. *J. Comput. Chem.*, **19**, 1187 (1998).
[36] A. Navarro; J. J. L. Gonzalez; A. G. Fernandez; I. Laczik; G. Pongor. *Chem. Phys.*, **313**, 279 (2005).

Törölt: submitted

Törölt:

[37] G. Rauhut; P. Pulay. *J. Am. Chem. Soc.*, **117**, 4167 (1995).

Regression results from fitting RM1 frequencies

Table 1. Parameters and summary statistics for RM1 frequency fitting.

Model	Eq. (1)	Equation (2)			
Parameter[a]	λ	λ_{Ls}	λ_{Lbt}	λ_{Hs}	λ_{Hbt}
	0.986(2)	0.984(2)	1.013(5)	0.9(1)	1.05(2)
[b]Δ_{MSE}	-7		-2		
[c]Δ_{MUE}	77		72		
[d]Δ_{max}	302		283		
[e]Δ_{RMS}	96		91		

[a]Standard deviations in the last digit are shown in parentheses
[b]Mean signed error of fitted wavenumbers (cm^{-1})
[c]Mean unsigned error (MUE) of fitted wavenumbers (cm^{-1})
[d]Maximum absolute deviation of fitted wavenumbers (cm^{-1})
[e]RMS deviation of fitted wavenumbers (cm^{-1})

Törölt: abs

Törölt: [b]Mean signed deviation of fitted wavenumbers (cm^{-1})¶
[c]Mean absolute deviation (MAD) of fitted wavenumbers (cm^{-1})¶

Regression results from fitting PM6 frequencies

Table 2. Parameters and summary statistics for PM6 frequency fitting.

Model	Eq. (1)	Equation (2)			
Parameter[a]	λ	λ_{Ls}	λ_{Lbt}	λ_{Hs}	λ_{Hbt}
	1.061(2)	1.099(2)	1.014(5)	0.924(8)	1.06(1)
[b]Δ_{MSE}	0			1	
[c]Δ_{MUE}	87			70	
[d]Δ_{max}	443			279	
[e]Δ_{RMS}	108			88	

[a]Standard deviations in the last digit are shown in parentheses
[b]Mean signed error of fitted wavenumbers (cm^{-1})
[c]Mean unsigned error (MUE) of fitted wavenumbers (cm^{-1})
[d]Maximum absolute deviation of fitted wavenumbers (cm^{-1})
[e]RMS deviation of fitted wavenumbers (cm^{-1})

Formázott: Bal: 0.63"

Törölt: [b]Δ

Törölt: [c]Δ_{abs}

Törölt: [b]Mean signed deviation of fitted wavenumbers (cm^{-1})¶
[c]Mean absolute deviation (MAD) of fitted wavenumbers (cm^{-1})¶

List of figure captions

Figure 1. Fitting of observed *vs.* S4QM//RM1 predicted wavenumbers. Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 2. Fitting of observed *vs.* S4QM//PM6 predicted wavenumbers. Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 3. Experimentally determined fundamentals [31,32] of indene *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2). No fit is made to this species; dotted line indicates the *a priori* y=x line. Circles mark points with deviations exceeding 3σ.
Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 4. Experimentally determined fundamentals [31,33] of indazol *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2). No fit is made to this species; dotted line indicates the *a priori* y=x line.
Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 5. Fundamental frequencies of *p*–dibenzodioxin isomers, as computed by SQM//B3LYP/6-31G(d) [37], *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2).
Legend: filled triangles, 2378-; open triangles, 1469-; squares, 1478-; diamonds, 1378-**TCDD**.
Light dotted line indicates the *a priori* y=x line; heavy dotted line indicates adjusted fit according to equation (3).
Top left inset: histogram of relative deviations in the 500–2000 cm^{-1} region.
Table: summary statistics of absolute (middle column) and relative deviations (right column), for fit to equation (3).
Bottom right inset: data for the 2,3,7,8-**TCDD** isomer plotted separately. The structure of the 2,3,7,8-**TCDD** isomer is shown as well, with the numbering of the *p*–dibenzodioxin skeleton indicated.

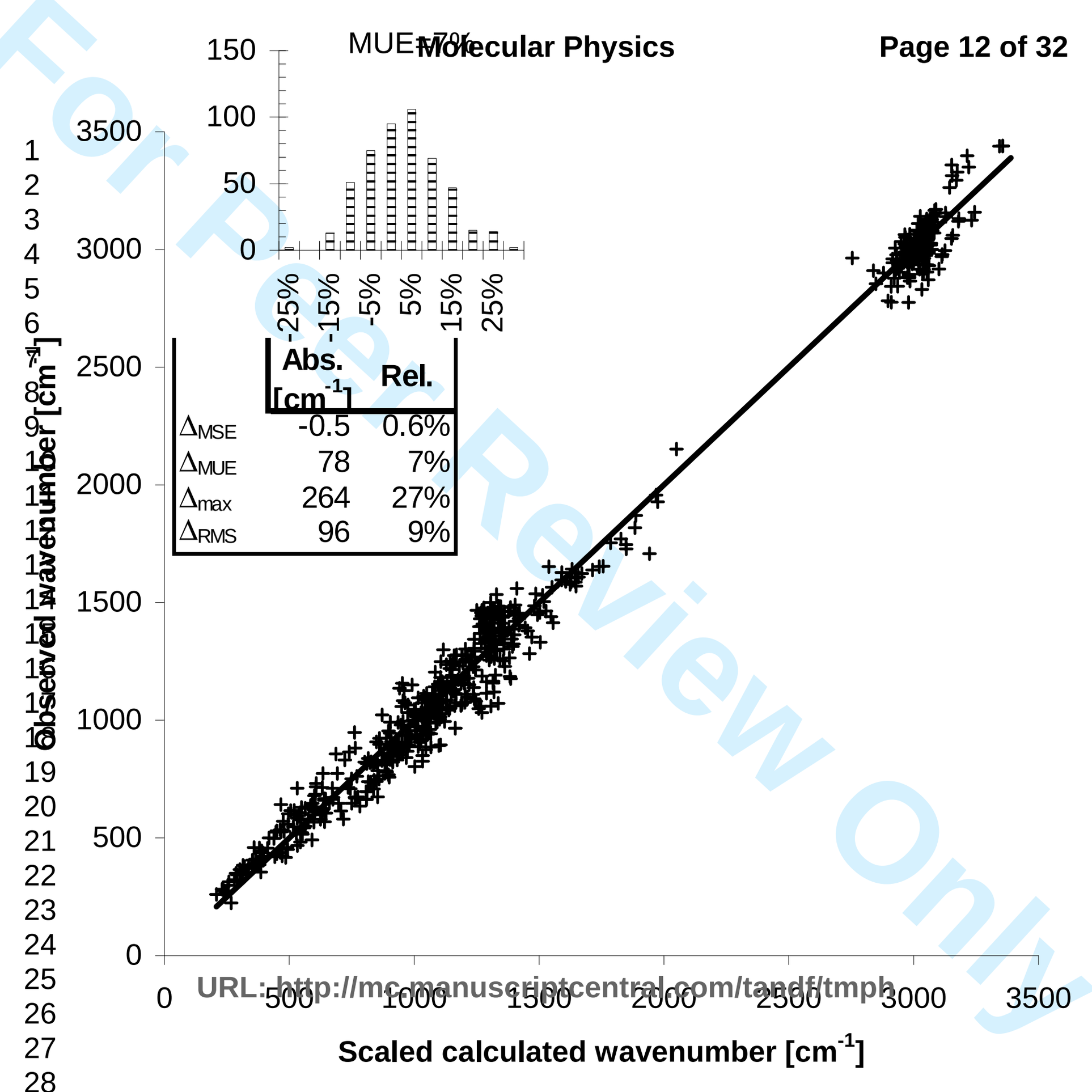
Törölt: ¶

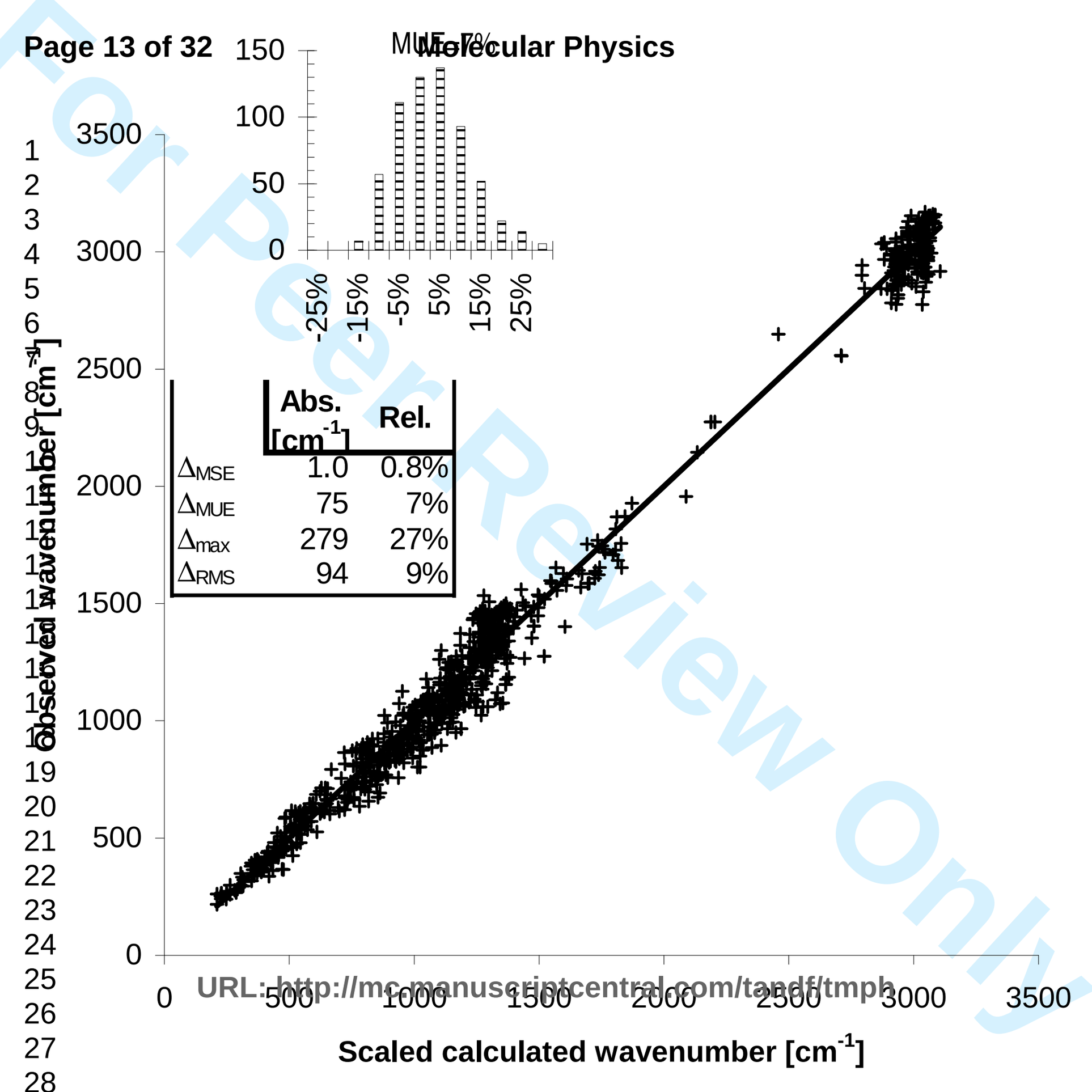
Törölt: ¶

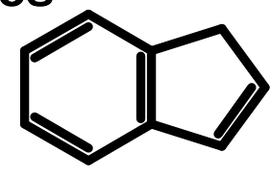
Törölt: . ¶
¶

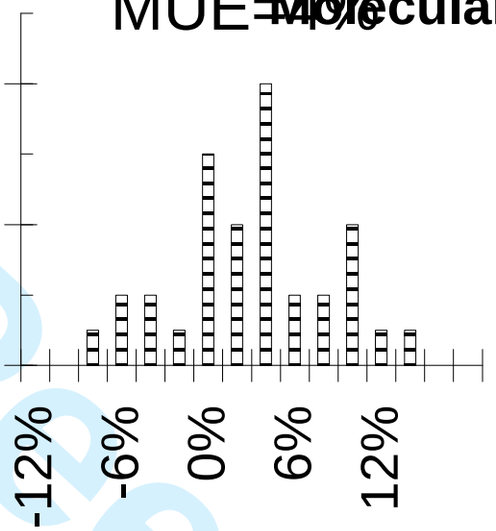
Törölt: Legend: filled triangles, 2378-; open triangles, 1469-; squares, 1478-; diamonds, 1378-**TCDD**. .
Inset

Formázott: Betűszín: Fényeszöld

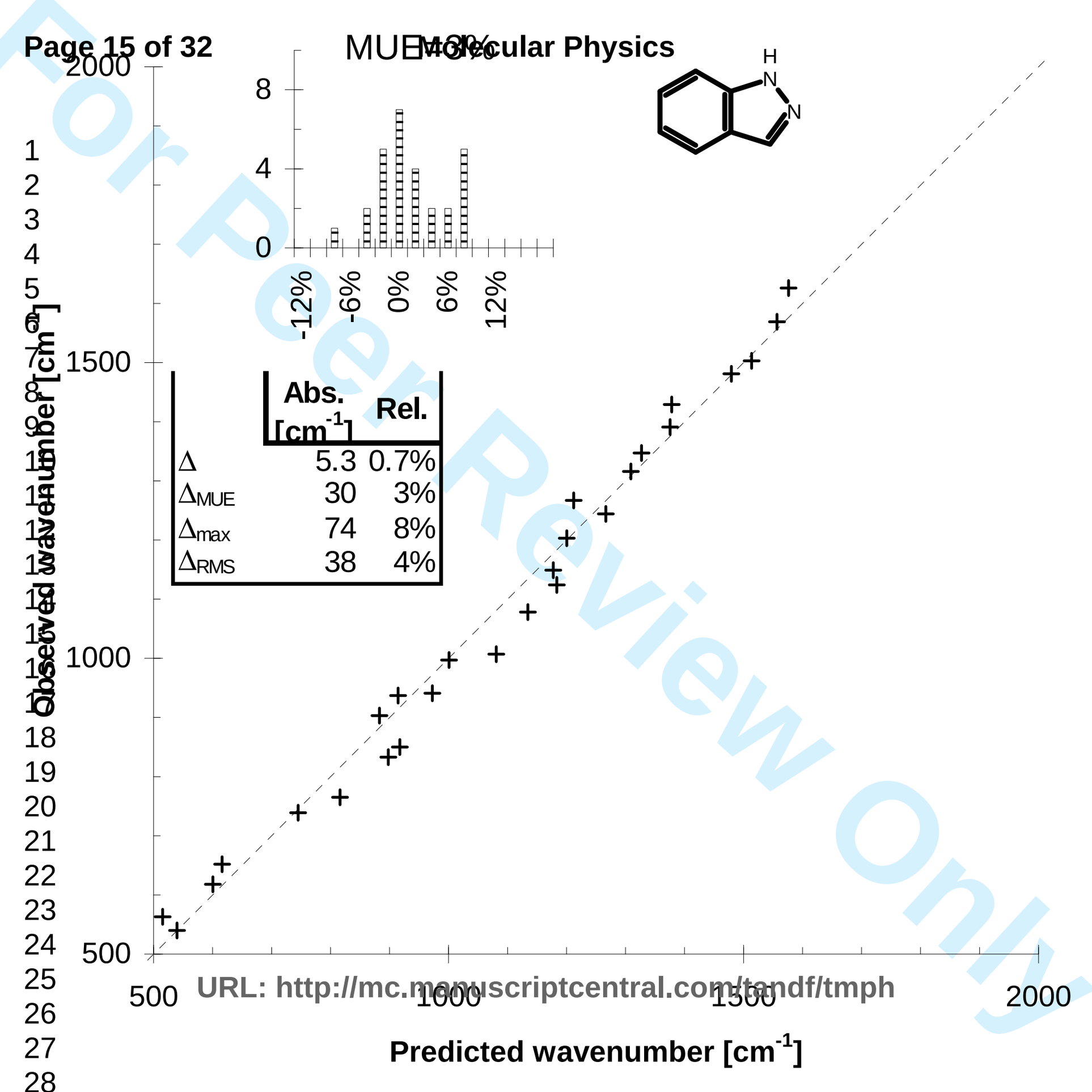


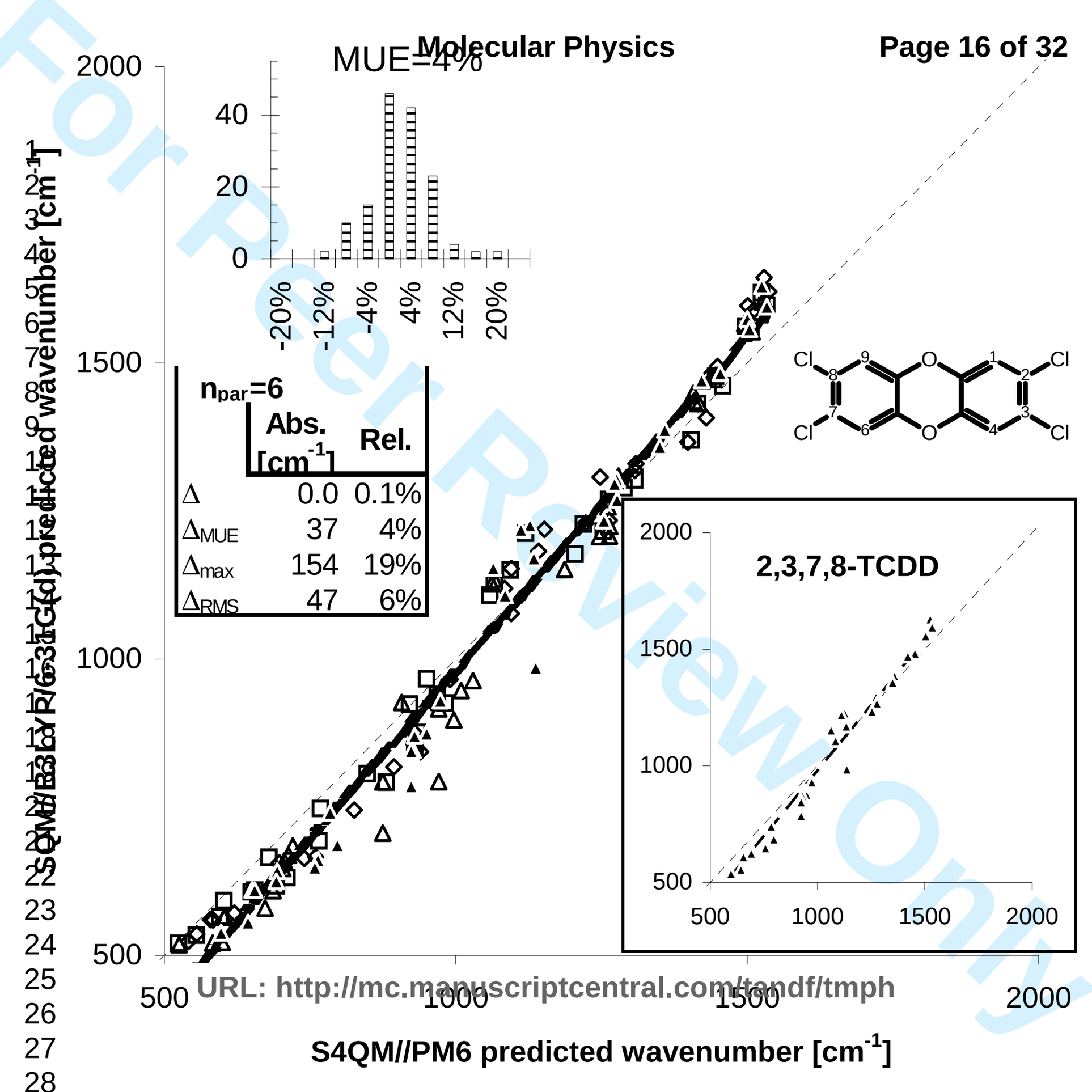






	Abs. [cm^{-1}]	Rel.
Δ	0.0	0.2%
Δ_{MUE}	38	4%
Δ_{max}	134	10%
Δ_{RMS}	49	5%





Running heads (verso) *Z. A. Fekete* et al.
(recto) *Frequency scaling factors: RM1 and PM6*

Article Type: "Peter Pulay Special Issue" contribution

Harmonic vibrational frequency scaling factors for the new NDDO Hamiltonians: RM1 and PM6

ZOLTAN A. FEKETE*†, EUFROZINA A. HOFFMANN‡, TAMÁS KÖRTVÉLYESI†‡ and BOTOND PENKE§

†HPC group, University of Szeged, DNT Szikra u. 2., H-6725 Szeged, Hungary
‡Department of Physical Chemistry, University of Szeged, PO Box 105, H-6720 Szeged, Hungary
§Department of Medical Chemistry, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary

Correspondence *Zoltan A. Fekete. Email: MQM2007.ZAF@inboxclean.com

Abstract

Scaling factors have been derived for obtaining fundamental vibrational frequencies with the recently introduced semiempirical molecular orbital methods RM1 and PM6, implemented in MOPAC2007. A least-squares approach is used with a training set comprised of 90 singlet-state molecules and 922 distinct vibrations, extracted from the NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB). Results are presented both for the conventional Scott-Radom type single-factor fitting, and for a multi-factor linear model: Semiempirical Semiglobal Self-consistently Scaled Quantum Mechanical (S4QM). The new NDDO methods in conjunction with the multi-linear fitting are shown to yield improved prediction of vibrational frequencies. To demonstrate the performance of S4QM//PM6 for calculating vibrational spectra, the examples of indene, indazole and four tetrachlorinated *p*-dibenzodioxins are presented.

Keywords: Vibrational frequencies; Semiempirical molecular orbital methods; Semiempirical parametrization; Scaled quantum mechanics; Multivariate analysis

AMS Subject Classification: 92E99 80A99 81V55 62J05

1. Introduction

Recent years brought on a revival for semiempirical molecular orbital theory [1--7], after a long period of 'consistent reports of its death' [8]. Despite the continuous growth in system sizes available for *ab initio* or DFT treatment, the computational efficiency [9] of semiempirical methods still makes them an attractive alternative for many applications [10--13].

Vibrational frequency calculations by quantum mechanical methods are of major importance in many areas of chemistry. Apart from their most straightforward application, the prediction and interpretation of vibrational spectra, they are crucial in dealing with quantities which depend on the form of vibrations, like infrared and Raman intensities, or the vibrational structure in ultraviolet and photoelectron spectra, as well as vibrational averaging effects on molecular geometries and dipole moments. Another important area is the derivation of thermochemical and kinetic information through statistical thermodynamics.

Computed frequencies typically deviate from experimentally determined ones significantly (with rare exceptions for very high-level calculations, which are only feasible for small molecules due to their extreme computational demand). This has led to the standard practice of scaling the results in order to bring them in line with measured values [14--19]. Two principal types of scaling procedure has emerged in practice. A more convenient, albeit theoretically less justified way is to fit calculated versus experimental data globally, without respect to the structural details involved. A theoretically more sound way is to use the full information content of the quantum mechanical results and scale the fundamental force constants accordingly, as in the Scaled Quantum Mechanical (SQM) procedure by Pulay [16,20,21].

Method development for predicting vibrational frequencies based on semiempirical quantum chemistry [22] has been disfavoured at least since Scott and Radom [18] reported the very poor results of AM1 and PM3 in this respect. This situation may be revised in light of the arrival of the new NDDO methods RM1 [23] and PM6 [24,25]. Both these correct many shortcomings of their predecessors, in particular calculated geometries are much improved. Especially significant is the advancement of PM6 that is parameterized based on a much extended set of data, and incorporates *d*–shell thus extending to the whole periodic table including transition metals. [24]

In this contribution we report on the performance of linearly scaled RM1 and PM6 in predicting vibrational frequencies. Besides showing results with the conventional single–parameter scaling, we are also introducing a multi–parameter protocol that seeks middle ground between the simplicity of global scaling and the detailed mode–specificity of SQM. Using a global fit to frequencies, but incorporating molecular descriptors split according to various types of vibrational modes, the procedure is designated Semiempirical Semiglobal Self–consistently Scaled Quantum Mechanical (S4QM) frequency fitting.

2. Methods

2.1. Data selection procedure

Published experimental vibrational frequencies, as well as geometries, were obtained from the NIST CCCBDB [26]. (Those compounds whose full geometry data is unavailable from the same database were excluded from the current study.) The initial selection has been pruned based on subsequent computations (see section 2.2.). Only species in the singlet electronic state, without spin contamination, have been included in the final analysis. Those polyatomics which are near–linear, *i.e*. have both of their calculated dimensions in the x and y directions (perpendicular to the main molecular axis) smaller than 100 pm, were also excluded. Finally, molecules with excitation energy (computed HOMO–LUMO difference) smaller than 8.00 eV were omitted, too. The dataset so chosen contains 90 molecules and 922 individual frequencies. For an overall

Deleted: exluded

description of this set, we have determined the following arithmetic mean values: there are 10.2 frequencies, 6.8 atoms — of which 3.2 are heavy (non–hydrogen) — and 2.6 elements on the average per species. The constituent elements are (the number of molecules that contain each is listed in parentheses): C(67), H(68), N(20), O(22), F(21), P(8), S(5), Cl(23) and Br(4). It is a characteristics of the CCCBDB that they are mostly small organic molecules with few heteroatoms. Sizes up to a total of 18 atoms (which occurs in cyclo–C_6H_{12}), and up to 8 heavy atoms (in C_2F_6) can be found in this sample.

2.2. Quantum chemical calculations

PM6 and RM1 computations were performed with the MOPAC2007 program package [24]. First, starting from their initial experimental geometry, tightly optimized RHF calculated geometries were determined for all molecules considered. At these geometries, single-point tests were run for spin contamination (with MOPAC keywords '1SCF UHF'), and species with S^2>0.01 were excluded from further consideration. Then bonding parameters were obtained (MOPAC keyword 'BONDS'). We imposed two selection criteria for molecules to be included in the final analysis for this work: no valence of any atom should be larger than 4.25, and no bond order larger than 2.25. For the remaining species, the MOPAC 'FORCE' calculation yielded theoretical harmonic frequencies, as well as the semiempirical vibrational analysis [27] that is utilized to obtain molecular descriptors according the section 2.3.

Deleted: initital

2.3. MOPAC vibrational analysis of Stewart

Normal coordinate calculations in MOPAC provide a supplemental output, with pair–wise atomic partitioning of motions into radial and tangential components. Although details of the scheme were published by its authors [27] long ago, its benefits are rarely recognized. For easy reference, the main points are summarized here:

> The energy absorbed by each atom (E_{AA}, E_{BB}, ...) and the energy absorbed or released by each bond (E_{AB}, E_{BC}, ...) is calculated for each mode. In a given mode, the energy change associated with an atom, E_{AA}, is calculated from its displacement and the force resisting the displacement (the force constants). The energy change associated with the A–B bond, E_{AB}, is calculated from the simultaneous relative displacement of atoms A and B and the net resisting force. E_{AB} may be either positive or negative (unlike the non-negative E_{AA}, E_{BB}, ..., terms). A loose interpretation of this algebraically driven result is that a bond may either absorb part of the energy of the photon stimulating the mode, or it may release energy to the other motions in the mode. The energy for a given pair of atoms is: $E(A–B) = E_{AA} + E_{BB} + 2E_{AB}$. The total energy for all the pairs of bonded atoms in the molecule in the mode is: $E_{tot} = \Sigma_A \Sigma_B E(A–B)$.

2.4. Regression models

The conventional one–parameter global frequency scaling relation [18,26], taken between theoretical (harmonic) frequencies ${}^{theo}\omega$ and their (anharmonic) experimental counterparts ${}^{obs}\nu$, is given by equation (1):

$$ {}^{obs}\nu = \lambda \cdot {}^{theo}\omega \qquad (1) $$

We introduce an expanded multi-parameter expression (2), based on partitioning equation (1) with a set of molecular descriptors, f_j, utilizing the analysis in section 2.3.

$$ {}^{obs}\nu = \Sigma_j \lambda_j \cdot f_j \cdot {}^{theo}\omega, \quad j\textbf{=Ls, Lbt, Hs, Hbt} \qquad (2) $$

The four descriptors are calculated from the partitioning of energy contributions to the vibrational mode: fractions of stretching (radial motion) and bending+torsional characters (tangential components) are collected from the MOPAC output; respectively **Ls** and **Lbt** are for vibrations involving light atoms (hydrogen isotopes), **Hs** and **Hbt** for those with exclusively heavy atoms. We use the designation S4QM for this model: Semiempirical Semiglobal Self-consistently Scaled Quantum Mechanical frequencies.

3. Results and discussion

3.1. RM1 frequency fitting

Results from fitting RM1 frequencies to equations (1–2) are summarized in table 1. To put the overall errors shown in perspective: Scott and Radom [18] in their study involving 1066 fundamentals determined Δ_{RMS} values of 126 cm^{-1} and 159 cm^{-1}, for scaled frequencies from AM1 and PM3 calculations, respectively.

[Insert table 1 about here]

Fitting for the model denoted S4QM//RM1, *i.e.* equation (2) with RM1, is shown figure 1. The inset of this figure (as well as of those following) has a table with summary statistics for both the absolute and relative deviations, as well as a histogram of the latter. Note that the statistics for relative errors are merely displayed for comparison with similar reports in the literature, but all of the calculations carried out here used non-relative values for fitting. (This causes the average of relative deviations to be further from zero than that of the absolute ones.) Both the visual display of the points scattering around the fitted line, and the statistics (i.e. $\Delta_{max} \approx 3\Delta_{RMS}$) confirm that there are no particular outliers.

Since the fingerprint region (500–2000 cm^{-1}) is often of special interest experimentally, the inset of figure 1 displays the histogram of relative errors tabulated from this interval only.

Because the statistics appear poorer than with the PM6 Hamiltonian (section 3.2.), which is a more sophisticated method regarding its quantum chemistry, we will further discuss only the latter below.

Deleted: Since

[Insert figure 1 about here]

3.2. PM6 frequency fitting

Results from fitting PM6 frequencies to equations (1–2) are summarized in table 2. It is noted here that our 4–factor model shows considerable improvement over the single–factor fitting, unlike in the case with RM1. In either case, results are markedly better than with AM1 or PM3.

[Insert table 2 about here]

Figure 2 visualizes fitting to the S4QM//PM6 model, and the histogram of relative errors tabulated from the 500–2000 cm^{-1} interval is shown in the inset.

[Insert figure 2 about here]

From a practical point view, instead of the overall error describing the fitting across the whole training set as presented above, it is more important to consider the molecular error [18] for individual species. In the following sections examples of utilizing this fitted S4QM//PM6 model are presented. They are all for molecules larger than those in the training set.

3.3. Indene and indazole: examples of S4QM//PM6 prediction for individual molecules

Indene is a compound with a well characterized spectrum, which is often used for calibration purposes either in experimental vibrational spectroscopy [28,29] or in theoretical modelling [30].

[Insert figure 3 about here]

Figure 3 plots the experimentally determined fundamentals [31,32] *vs.* those predicted by the S4QM//PM6 model (sections 2.1., 2.4. and 3.2.). Statistics taken over the fingerprint region are summarized on the inset. We emphasize that no fit is made to the experimental data on indene: unmodified λ_j parameters from table 2 are substituted into equation (2) for the prediction. This is a check for the transferability of the four scaling factors determined on the training set, using no fitted parameter determined specifically in connection with the species.

Similarly to the above, figure 4 presents the S4QM//PM6 results for the indazole molecule (experimental fundamentals are taken from [31,33]). As seen from the structure indicated on the figure, this compound is an indene analogoue that contains geminal nitrogens in a heterocyclic ring.

[Insert figure 4 about here]

Even though this structural unit is completely lacking from the training set used to obtain the model parameters, the frequency predictions appear surprisingly good for this species: in the fingerprint region Δ_{RMS}=38 cm^{-1}, or 4% relative error. For comparison, El-Azhary [31] achieved a fit of Δ_{RMS}=9 cm^{-1} with B3LYP/6-31G** computations and a set of three SQM–type [16,20] scaling factors, which had been refined based on six other analogue structures (also reported a similarly refined single–factor scaled fit of Δ_{RMS}=15 cm^{-1}). On the other hand, the SQM fit by Cane [33], based on HF/6-31G** computations (which are now considered inferior to DFT for frequency predictions [31,34--36]), yielded Δ_{RMS}=22 cm^{-1}. One should keep in mind that both these latter methods require orders of magnitude larger computational times than the semiempirical ones.

3.4. Tetrachlorinated p–dibenzodioxins: examples of S4QM//PM6 predictions for an isomer family

In this section results for a set of four isomer tetrachlorinated *p*–dibenzodioxins (**TCDD**) presented, see figure 5 (which also shows the structure and numbering of their skeleton). The comparison made here is with higher–level (SQM//B3LYP/6-31G(d) [37]) theoretical predictions, rather than with experimental data.

[Insert figure 5 about here]

For both methods a total of 146 frequencies are considered, which fall into the fingerprint region. With their 18 heavy atoms, and multi–substituted aromatic system, these molecules substantially exceed the coverage of our training set. In particular, many features of the **TCDD** spectra are affected by the presence of the chlorine, which being a second–row element is expected to scale differently [37]. Therefore the large, and partly systematic, deviations seen on figure 5 are not surprising. Nevertheless, the overall trend is fairly well reproduced. Moreover, the global part (*i.e.* that spread across all atom types rather than characteristic of chlorine) of the systematic difference between our model and that with the higher–level method can be minimized with a simple linear adjustment.

$$^{adj}\nu = m + b \cdot {}^{S4QM}\omega \qquad (3)$$

With the *a posteriori* modification described by equation (3), the ${}^{S4QM}\omega$ original predictions are brought in line with the target dataset, *via* incorporating two extra parameters by fitting ${}^{adj}\nu$ to the target. The errors from this expanded, six–parameter model are summarized on the inset of figure 5. These are indicative of the limits to the predictive power of this simple S4QM//PM6 method, as specified with the λ_{Ls}, λ_{Lbt}, λ_{Hs}, λ_{Hbt} parameter set given in table 2. Clearly, in order to make reliable predictions for molecules very dissimilar to those included in the training set, the diversity of the data as well as of the parameters should be increased. The partial successes of the initial version of S4QM//PM6 show promise for applying the same protocol for expanding the model this way.

4. Conclusion and outlook

Compared to the errors of scaled semiempirical frequency predictions published in the seminal paper by Scott and Radom [18], the new NDDO methods are improved over their predecessors: single–parameter fitting with RM1 yields Δ_{RMS} 96 — instead of Δ_{RMS}=126 cm^{-1} with AM1; with PM6 Δ_{RMS}=108 — instead of Δ_{RMS}=159 cm^{-1} with PM3. Our novel S4QM fitting gives further reduction of error, most notably with the all–element method PM6 (Δ_{RMS}=88 with four scaling constants). Importantly, all four parameters (λ_{Ls}, λ_{Lbt}, λ_{Hs}, λ_{Hbt}) from S4QM//PM6 are determined with high significance from our modestly sized training set currently utilized. This strongly indicates that systematic further improvement of the statistics will be attainable with an enlarged data set and judiciously augmented parameterization. Therefore linearly scaled semiempirical methods can be made a semiquantitative tool for vibrational frequency prediction. With their improved calibration they will yield *a priori* (though not *ab initio*) fundamental frequencies at very small computational expense, even for large systems.

Acknowledgements

Z.A.F. is grateful for the enthusiastic help of J.J.P. Stewart with his program.
The computing facilities of the HPC group at the University of Szeged were utilized.
Partial financial support was provided by the Hungarian National Office for Research and Technology (grant RET 08/2004), and by the Hungarian National Scientific Research Fund (OTKA grant K61577).

References

[1] M. Kolb; W. Thiel. *J. Comput. Chem.*, **14**, 775 (1993).
[2] W. Thiel. *Adv. Chem. Phys.*, **93**, 703 (1996).
[3] T. Clark. *J. Mol. Struct. THEOCHEM*, **530**, 1 (2000).
[4] J. J. P. Stewart. *J. Mol. Model.*, **10**, 155 (2004).
[5] J. J. P. Stewart. *J. Mol. Model.*, **10**, 6 (2004).
[6] S. Patchkovskii; A. Koslowski; W. Thiel. *Theor. Chem. Acc.*, **114**, 84 (2005).
[7] R. Steiger; C. H. Bischof; B. Lang; W. Thiel. *Future. Gener. Comp. Sy.*, **21**, 1324 (2005).
[8] T. Clark; P. Winget; C. Selcuki; A. Horn; B. Martin. *Abstr. Pap. Am. Chem. Soc.*, **224**, U500 (2002).
[9] J. J. P. Stewart; P. Csaszar; P. Pulay. *J. Comput. Chem.*, **3**, 227 (1982).
[10] P. Murray-Rust; H. S. Rzepa; J. J. P. Stewart; Y. Zhang. *J. Mol. Model.*, **11**, 532 (2005).
[11] A. Monge; A. Arrault; C. Marot; L. Morin-Allory. *Mol. Divers.*, **10**, 389 (2006).
[12] J. Linnanto; J. Korppi-Tommola. *J. Comput. Chem.*, **25**, 123 (2004).
[13] H. M. Senn; W. Thiel. *Top. Curr. Chem.*, **268**, 173 (2007).
[14] C. E. Blom; P. J. Slingerland; C. Altona. *Mol. Phys.*, **31**, 1359 (1976).
[15] C. E. Blom; C. Altona; A. Oskam. *Mol. Phys.*, **34**, 557 (1977).
[16] P. Pulay; G. Fogarasi; G. Pongor; J. E. Boggs; A. Vargha. *J. Am. Chem. Soc.*, **105**, 7037 (1983).
[17] J. A. Pople; A. P. Scott; M. W. Wong; L. Radom. *Israel J. Chem.*, **33**, 345 (1993).
[18] A. P. Scott; L. Radom. *J. Phys. Chem. US*, **100**, 16502 (1996).
[19] V. I. Pupyshev; Y. N. Panchenko; C. W. Bock; G. Pongor. *J. Chem. Phys.*, **94**, 1247 (1991).
[20] G. Fogarasi; P. G. Szalay; P. P. Lieschesk; J. E. Boggs; P. Pulay. *J. Mol. Struct. THEOCHEM*, **36**, 341 (1987).
[21] J. Baker; A. A. Jarzecki; P. Pulay. *J. Phys. Chem. A*, **102**, 1412 (1998).
[22] M. B. Coolidge; J. E. Marlin; J. J. P. Stewart. *J. Comput. Chem.*, **12**, 948 (1991).
[23] G. B. Rocha; R. O. Freire; A. M. Simas; J. J. P. Stewart. *J. Comput. Chem.*, **27**, 1101 (2006).
[24] James J. P. Stewart. MOPAC2007, Version 7.0*. Available online at: HTTP://OpenMOPAC.net (accessed 15 Jan. 2007).
[25] J. J. P. Stewart. *J. Mol. Model.*, **accepted**, (2007).
[26] Editor: R.D. Johnson III. NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. Available online at: http://srdata.nist.gov/cccbdb (accessed 15 Dec. 2006).
[27] C.J. Dymek; J.J.P. Stewart. *Inorg. Chem.*, **28**, 1472 (1989).
[28] IUPAC Commission of Molecular Structure and Spectroscopy. *Tables of Wavenumbers for the Calibration of Infrared Spectrometers*, Pergamon Press, New York, (1977).
[29] E. N. Lewis; V. F. Kalasinsky; I. W. Levin. *Anal. Chem.*, **60**, 2658 (1988).
[30] H. Yoshida; K. Takeda; J. Okamura; A. Ehara; H. Matsuura. *J. Phys. Chem. A*, **106**, 3580 (2002).
[31] A. A. El-Azhary. *Spectrochim. Acta A*, **55**, 2437 (1999).
[32] T. D. Klots. *Spectrochim. Acta A*, **51**, 2307 (1995).
[33] E. Cane; P. Palmieri; R. Tarroni; A. Trombetti. *J. Chem. Soc .Faraday. T.*, **89**, 4005 (1993).
[34] G. Rauhut; P. Pulay. *J. Phys. Chem. US*, **99**, 3093 (1995).
[35] J. Baker; P. Pulay. *J. Comput. Chem.*, **19**, 1187 (1998).
[36] A. Navarro; J. J. L. Gonzalez; A. G. Fernandez; I. Laczik; G. Pongor. *Chem. Phys.*, **313**, 279 (2005).

[37] G. Rauhut; P. Pulay. *J. Am. Chem. Soc.*, **117**, 4167 (1995).

Regression results from fitting RM1 frequencies

Table 1. Parameters and summary statistics for RM1 frequency fitting.

Model	Eq. (1)	Equation (2)			
Parameter[a]	λ	λ_{Ls}	λ_{Lbt}	λ_{Hs}	λ_{Hbt}
	0.986(2)	0.984(2)	1.013(5)	0.9(1)	1.05(2)
[b]Δ_{MSE}	-7			-2	
[c]Δ_{MUE}	77			72	
[d]Δ_{max}	302			283	
[e]Δ_{RMS}	96			91	

[a]Standard deviations in the last digit are shown in parentheses
[b]Mean signed error of fitted wavenumbers (cm^{-1})
[c]Mean unsigned error (MUE) of fitted wavenumbers (cm^{-1})
[d]Maximum absolute deviation of fitted wavenumbers (cm^{-1})
[e]RMS deviation of fitted wavenumbers (cm^{-1})

Deleted: abs

Deleted: [b]Mean signed deviation of fitted wavenumbers (cm^{-1})¶
[c]Mean absolute deviation (MAD) of fitted wavenumbers (cm^{-1})¶

Regression results from fitting PM6 frequencies

Table 2. Parameters and summary statistics for PM6 frequency fitting.

Model	Eq. (1)	Equation (2)			
Parameter[a]	λ	λ_{Ls}	λ_{Lbt}	λ_{Hs}	λ_{Hbt}
	1.061(2)	1.099(2)	1.014(5)	0.924(8)	1.06(1)
[b]Δ_{MSE}	0			1	
[c]Δ_{MUE}	87			70	
[d]Δ_{max}	443			279	
[e]Δ_{RMS}	108			88	

[a]Standard deviations in the last digit are shown in parentheses
[b]Mean signed error of fitted wavenumbers (cm^{-1})
[c]Mean unsigned error (MUE) of fitted wavenumbers (cm^{-1})
[d]Maximum absolute deviation of fitted wavenumbers (cm^{-1})
[e]RMS deviation of fitted wavenumbers (cm^{-1})

Deleted: [b]Δ

Deleted: [c]Δ_{abs}

Deleted: [b]Mean signed deviation of fitted wavenumbers (cm^{-1})¶
[c]Mean absolute deviation (MAD) of fitted wavenumbers (cm^{-1})¶

List of figure captions

Figure 1. Fitting of observed *vs.* S4QM//RM1 predicted wavenumbers. Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 2. Fitting of observed *vs.* S4QM//PM6 predicted wavenumbers. Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 3. Experimentally determined fundamentals [31,32] of indene *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2). No fit is made to this species; dotted line indicates the *a priori* y=x line. Circles mark points with deviations exceeding 3σ.
Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 4. Experimentally determined fundamentals [31,33] of indazol *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2). No fit is made to this species; dotted line indicates the *a priori* y=x line.
Inset: histogram of relative deviations in the 500–2000 cm^{-1} region. Table: summary statistics of absolute (middle column) and relative deviations (right column).

Figure 5. Fundamental frequencies of *p*–dibenzodioxin isomers, as computed by SQM//B3LYP/6-31G(d) [37], *vs.* those predicted by the S4QM//PM6 model (equation (2) with λ_j parameters from table 2).
Legend: filled triangles, 2378-; open triangles, 1469-; squares, 1478-; diamonds, 1378-**TCDD**.
Light dotted line indicates the *a priori* y=x line; heavy dotted line indicates adjusted fit according to equation (3).
Top left inset: histogram of relative deviations in the 500–2000 cm^{-1} region.
Table: summary statistics of absolute (middle column) and relative deviations (right column), for fit to equation (3).
Bottom right inset: data for the 2,3,7,8-**TCDD** isomer plotted separately. The structure of the 2,3,7,8-**TCDD** isomer is shown as well, with the numbering of the *p*–dibenzodioxin skeleton indicated.

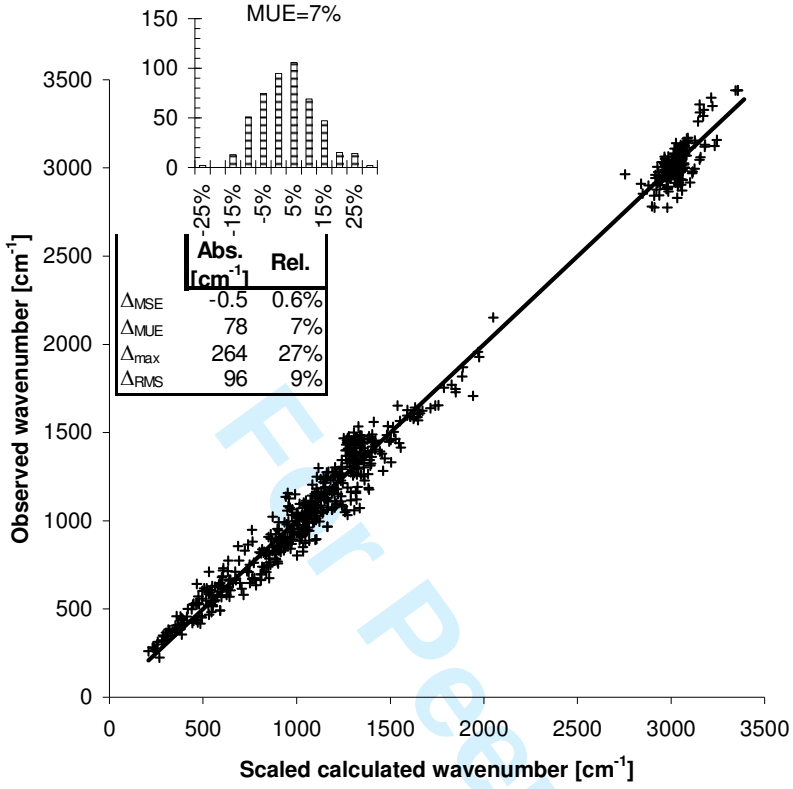
Deleted: ¶

Deleted: ¶

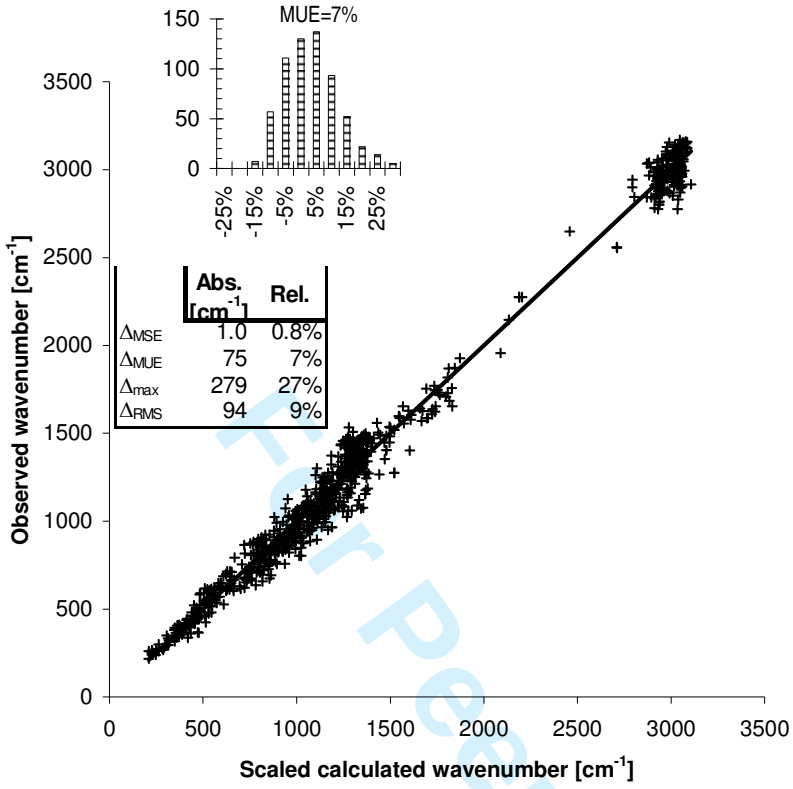
Deleted: . ¶
¶

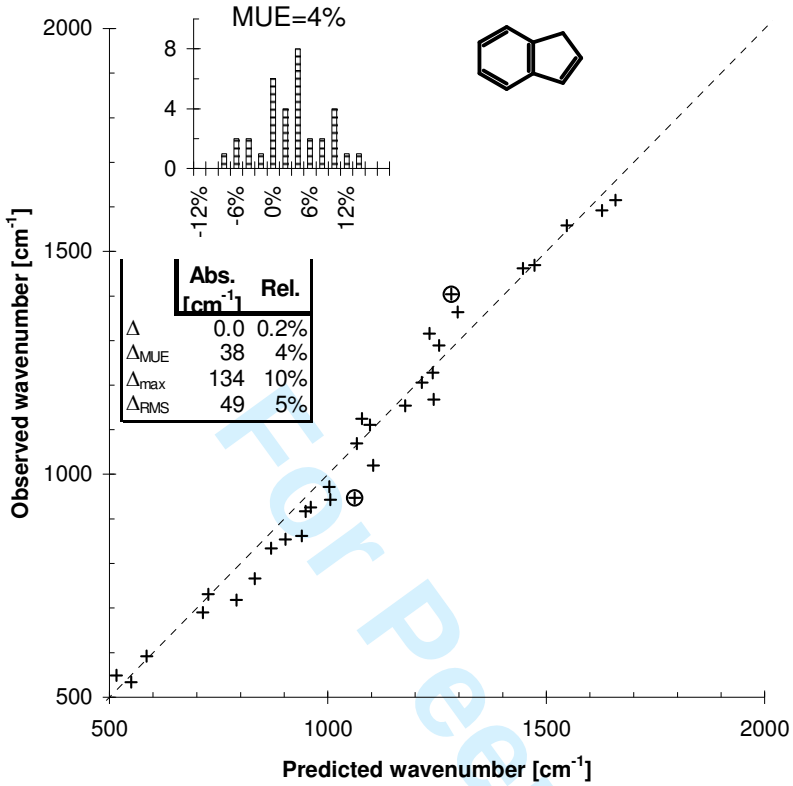
Deleted: Legend: filled triangles, 2378-; open triangles, 1469-; squares, 1478-; diamonds, 1378-**TCDD**. .
Inset

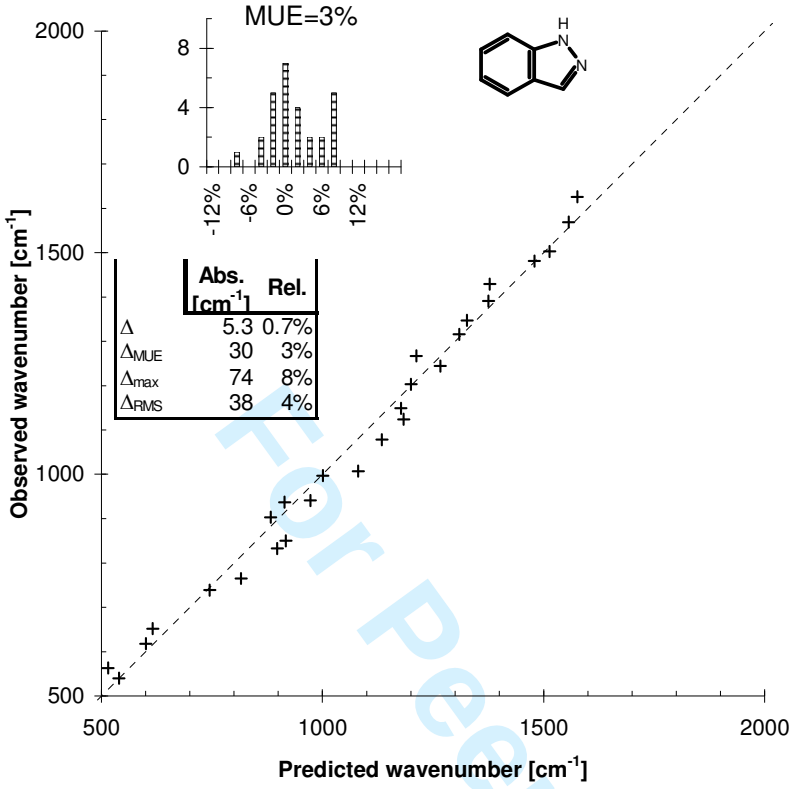
Formatted: Font color: Bright Green

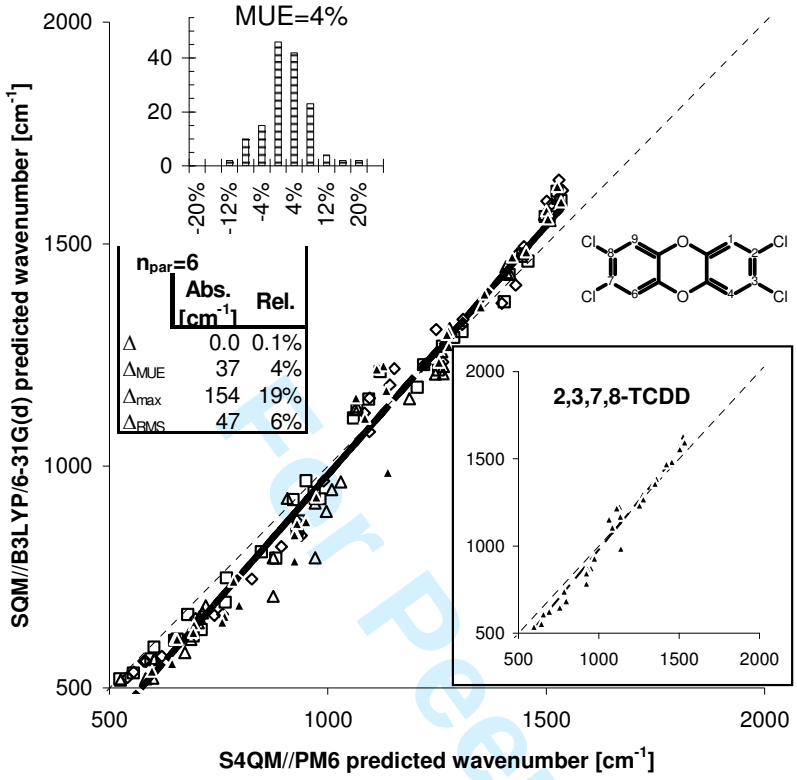


Formatted: Left: 70.85 pt, Right: 70.85 pt, Top: 70.85 pt, Bottom: 70.85 pt, Width: 595.3 pt, Height: 841.9 pt, Header distance from edge: 35.4 pt, Footer distance from edge: 35.4 pt









Formatted: English U.K.