Direct Scaling of Primitive Valence Force Constants: An Alternative Approach to Scaled Quantum Mechanical Force Fields

Jon Baker,* Andrzej A. Jarzecki,† and Peter Pulay

Department of Chemistry, University of Arkansas, Fayetteville, Arkansas 72701

Received: October 31, 1997

We present an alternative approach to the derivation of scaled quantum mechanical (SQM) force fields involving the direct scaling of *individual* primitive valence force constants from a full set of *redundant* valence coordinates. Our approach is completely general and more flexible than previous SQM schemes. Optimal scaling factors for various primitive stretching, bending, and torsional force constants are derived from a training set of 30 molecules containing C, O, N, H, and Cl and used to scale force constants for a further 30 molecules. Calculated vibrational frequencies are compared with experimental values for over 1500 fundamentals. Using the hybrid three-parameter B3-LYP density functional with the split-valence 6-31G* basis set, our scaling procedure gives an average error of less than 8.5 cm^{-1} in the scaled frequencies. The average percentage error is under 1%.

1. Introduction

Vibrational spectroscopy is an important tool for molecular identification. The presence and intensity of various peaks in an infrared (IR) or Raman spectrum indicate the presence of particular functional groups in the molecule. Vibrational spectroscopy is often used to identify reactive intermediates as well as species in interstellar space.

The direct calculation of vibrational frequencies by ab initio computations can be of considerable help in the interpretation of experimental vibrational spectra. In larger molecules it is virtually impossible to reliably assign vibrational fundamentals without input from theory. Theory can also suggest frequencies that can be used as "fingerprints" for the presence of particular conformers, isomers, or compounds. The computed normal modes can be used to estimate IR and Raman intensities from dipole and polarizability derivatives, as well as vibrational averaging effects on molecular geometries and properties. Comparison of calculated and experimental vibrational spectra has become one of the principal means of identifying unusual molecules, especially in low-temperature matrixes.

In the harmonic approximation, vibrational frequencies are calculated theoretically from computed force constants (second derivatives of the molecular potential energy with respect to atomic displacements), either by direct diagonalization of the (mass-weighted) Cartesian force constant matrix or in some (usually nonredundant) internal coordinate system using the Wilson GF matrix method.[1] Analytical derivative methods are essential for the calculation of force constants. The first systematic calculations of polyatomic force constants[2−4] were performed via numerical differentiation of analytical gradients.[5] The introduction in 1979 of analytical second-derivative techniques at the Hartree−Fock (HF) level further increased the effficiency of such calculations.[6] Over the next decade, the calculation of HF vibrational frequencies for small to medium-sized systems became routine.

When compared with experimental fundamentals, computed HF frequencies were found to be consistently 7−15% too high.[2e,7] This is due partly to limitations in the HF method itself, i.e., the neglect of electron correlation and basis set truncation, and, to a lesser extent, to anharmonicity, which affects the observed fundamentals but is necessarily omitted in the usual harmonic approximation. The overestimation is, however, fairly uniform, and consequently scaling factors are often applied to eliminate the systematic error in the force constants and frequencies and provide better agreement with experiment.

The need for empirical correction diminishes but is not completely eliminated if the quality of the wave function improves by adding electron correlation and increasing the basis set. Unfortunately, most correlated wave functions are too expensive for general use. The exceptions to this statement are the currently popular density functional methods, which include a significant part of the total electron correlation, albeit in a somewhat empirical fashion, and which are only marginally more expensive than basic Hartree−Fock. The best density functionals available at the moment are the so-called hybrid HF−DFT functionals,[8] which mix a percentage of the "exact" Hartree−Fock exchange term in with the density functional. These hybrid functionals are capable of providing molecular structures, energetics, and a wide range of properties—including vibrational frequencies—of a similar quality to those obtained from expensive, high-level ab initio methods (such as coupled cluster methods) at a fraction of the computational cost.[9] For example, Scott and Radom have recently compared the computed harmonic vibrational frequencies for a set of 122 small molecules (a total of 1066 frequencies) using various theoretical methods and basis sets.[10] The methods included HF, second-order Møller−Plesset perturbation theory (MP2), quadratic configuration interaction with singles and doubles substitution (QCISD), and several density functionals, including three hybrid functionals. Global scaling factors for each level of theory were determined by a least-squares fit of the calculated to the experimental vibrational frequencies. Two of the hybrid density functionals—B3-LYP[8,11,12] and B3-PW91[8,11,13]—gave the best overall agreement (after scaling) between theory and experiment.

The first scaling methods applied to ab initio force constants used several different scale factors to correct for systematic

† Current address: Department of Chemistry, Indiana University, Bloomington, IN 47405.

S1089-5639(98)00038-3 CCC: $15.00
Published on Web 02/03/1998

errors in different types of molecular deformations, e.g. stretches, bends, and torsions.[2e,3,7,14] This procedure requires the transformation of the molecular force field to chemically meaningful internal coordinates and cannot be applied directly to the calculated frequencies. It is thus less convenient than global scaling using a single scaling factor. Because of its simplicity, global scaling became popular to correct calculated frequencies at the Hartree−Fock level, particularly if the results need not be very accurate, e.g., for the calculation of zero-point energies. However, scaling with multiple scale factors yields much better results, as demonstrated convincingly by the pioneering work of Blom and Altona.[7] Their method formed the basis of the scaled quantum mechanical (SQM) force field procedure[15] which has been used in this group and elsewhere for over 15 years.

In the original SQM procedure, the molecular geometry is expressed in terms of a full set of nonredundant natural internal coordinates.[16,17] Natural internal coordinates use individual bond displacements as stretching coordinates and linear combinations of bond angles and torsions as deformational coordinates. Suitable linear combinations of bends and torsions (the two are considered separately) are selected using group theoretical arguments based on local pseudosymmetry. On the basis of chemical intuition, the natural internal coordinates of all molecules under consideration are sorted into groups sharing a common scaling factor, and factors for each group are determined by a least-squares fitting procedure to experimental vibrational frequencies. Force constants, originally calculated in Cartesian coordinates, are transformed into an internal coordinate representation, and scaling is applied to the elements of the internal coordinate force constant matrix (*not* to the individual vibrational frequencies) according to

$$\mathbf{F}_{ij}(\text{scaled}) = (s_i s_j)^{1/2}\mathbf{F}_{ij} \qquad (1)$$

where s_i and s_j are scaling factors for natural internal coordinates i and j, respectively.

The accuracy obtained by selective scaling in this way is naturally greater than if just a single overall scaling factor were used. Additionally, scaling the force constant matrix also affects the resultant normal modes, and hence the calculated intensities (which are unaffected if only the frequencies are scaled), leading to better agreement with experimental intensities.

The SQM procedure has been widely used in the interpretation of vibrational spectra. A further important role is the development of *transferable* scale factors which can be used to modify calculated force constants and so predict the vibrational spectrum a priori. For example, Rauhut and Pulay have developed a set of 11 transferable scaling factors for organic molecules containing the atoms C, O, N, and H based on B3-LYP/6-31G* force fields,[18] and Panchenko et al. have successfully transferred scale factors between rotational isomers using HF force fields.[19] A recent review gives several favorable comparisons between experimental IR spectra and spectra predicted using the SQM method.[20]

A prerequisite for implementing the SQM procedure is the generation of a set of natural internal coordinates for the molecules under examination and the classification of these coordinates into "chemically similar" groups to which the same scaling factor will be applied. This is normally straightforward for stretches, as each stretch is a coordinate in its own right, but it can sometimes be awkward for bends and torsions, because these coordinates typically come as linear combinations of several different primitive bends or torsions. Additionally, the manual generation of natural internal coordinates is tedious. The problem has been solved in principle by automatic computer programs;[17] however, the automatic generation may fail occasionally for unusual topologies, e.g., cage compounds. The somewhat delocalized nature of natural internal coordinates, for instance in rings, is also a disadvantage.

It would clearly be advantageous if, instead of classifying and scaling natural internal coordinates, each primitive valence coordinate could be scaled *individually*. This has not been seriously considered in the past, as primitive valence coordinates form a highly redundant coordinate set; that is, there are normally far more individual stretches, bends, and torsions in a molecule than are necessary to describe all possible molecular deformations. By taking appropriate linear combinations of primitives, as in the natural internal coordinates, this redundancy can be eliminated.

Recent work has shown that the concept of a gradient can be meaningfully generalized to redundant internal coordinates,[21] and the latest molecular geometry optimization algorithms[21−23] use this approach. One of the first methods that could use force constants expressed in redundant internal coordinates was the normal coordinate optimization technique of Sellers et al.,[24] although this aspect of the method was not emphasized at the time. Recently, we have shown that second and higher derivatives can also be uniquely expressed in redundant internal coordinates.[25] For our purposes, this means that we can express the force constant matrix in terms of a full set of *redundant* valence coordinates and still extract vibrational frequencies and normal modes from it, exactly as we can from a force constant matrix expressed using a *nonredundant* coordinate set. This opens up the possibility for the direct scaling of *individual* stretches, bends, and torsions.

The purpose of this paper is to present a modified SQM procedure involving the scaling of individual valence coordinates. This has immediate advantages in terms of ease of use, as no natural internals need to be generated (which may fail for complicated molecular topologies) and simplifies the classification and presorting of the coordinates. In section 2 we present our modified SQM procedure including the optimization of scaling factors (which is similar to the original scheme but has not previously been presented in detail). In section 3 we use the new SQM procedure to derive scaling factors for *individual* stretches, bends, and torsions for a training set of 30 molecules containing C, O, N, H, and Cl and subsequently use these scaling factors to predict the fundamental modes of a further 30 molecules, all taken from the recent literature (post-1993). Section 4 comprises a summary and conclusions.

2. Modified SQM Procedure

In the modified SQM procedure we describe the geometry of a given molecule in terms of a full set of redundant valence coordinates. Following ref 22, we choose all possible stretches, all planar bends, and all proper torsions that can be generated on the basis of the atomic connectivity. Occasionally other primitives may be utilized, such as out-of-plane bends (in cases where there are an insufficient number of proper torsions to span all the degrees of freedom, e.g., formaldehyde) or the special collinear and coplanar bending coordinates for near-linear arrangements of atoms,[26] but usually only stretches, bends, and torsions are used.

The transformation of a force constant matrix from Cartesians into a *redundant* set of valence coordinates, which previously had been problematic, can now be done routinely and in a *consistent* manner using the concept of the generalized inverse. Full details of this transformation, together with a discussion of the interpretation of force constants defined in redundant

internal coordinates, including a possible definition of invariant force constants, are given elsewhere.[25] We emphasize that the only significant difference between the original SQM procedure, which used a nonredundant set of natural internal coordinates, and our modified scheme, which uses a full set of redundant valence coordinates, is the replacement of the regular inverse matrix by the generalized inverse whenever a matrix inverse is needed. Apart from this, the actual steps taken in the two procedures are the same.

With a force constant matrix, $\mathbf{F}^{\text{int}}$, defined in terms of the full redundant set of valence coordinates, we can scale the *individual* primitive stretches, bends, and torsions, using eq 1. If we already have a set of scaling factors, then $\mathbf{F}^{\text{int}}$ can be scaled directly after its formation from the Cartesian force constant matrix, and one solution of the Wilson GF eigenvalue equation will give the predicted (scaled) vibrational frequencies and normal modes. If, on the other hand, we wish to optimize the scaling factors to get the best agreement between our (scaled) frequencies and a given set of experimental frequencies for a particular molecule or set of molecules, then repeated solutions are required, changing the scaling factors, until the best agreement is obtained. This is accomplished using a least-squares minimization (see, for example, ref 27), similar to that used in the SCALE2 program.[28] This program, although used by many groups, has been described only very briefly,[15] and therefore we summarize our optimization method below.

We define a least-squares merit function

$$\chi^2(\mathbf{s}) = \sum\{[\nu_i^{\text{expt}} - \nu_i^{\text{theor}}(\mathbf{s})]w_i\}^2 \qquad (2)$$

minimization of which should lead to the "best" agreement between experimental and theoretical vibrational frequencies. In eq 2, ν_i^{expt} are the experimental frequencies and $\nu_i^{\text{theor}}(\mathbf{s})$ the theoretical frequencies, the nomenclature indicating that the theoretical values (and hence χ^2) depend on the values of the scaling factors, $\mathbf{s}$. The w_i are weighting factors, which can be used to increase or decrease the importance of specific frequencies in the fit. In particular, frequencies that are unobserved experimentally, or that are considered to have large error bars, are assigned low or zero weights. The summation in eq 2 is over the total number of fundamental vibrational frequencies present in all molecules under consideration. For well-established frequencies, it has been suggested[15] that w_i should be inversely proportional to the frequency itself. This is intermediate between assuming that the frequencies have a constant error (e.g., from broadening of the rotational envelope caused by condensed phase effects) and a roughly constant percentage error, as is the case with anharmonic effects.

Differentiating eq 2 with respect to the scaling factors gives

$$\frac{\partial\chi^2(\mathbf{s})}{\partial s_k} = -2\sum\{[\nu_i^{\text{expt}} - \nu_i^{\text{theor}}(\mathbf{s})]w_i\}\frac{\partial\nu_i^{\text{theor}}(\mathbf{s})}{\partial s_k} \qquad k = 1, 2, ..., n_s \quad (3)$$

where n_s is the total number of scaling factors. Differentiating a second time gives

$$\frac{\partial^2\chi^2(\mathbf{s})}{\partial \mathbf{s}_k\partial \mathbf{s}_l} = 2\sum w_i\left\{\frac{\partial\nu_i^{\text{theor}}(\mathbf{s})}{\partial s_k}\frac{\partial\nu_i^{\text{theor}}(\mathbf{s})}{\partial s_l} - [\nu_i^{\text{expt}} - \nu_i^{\text{theor}}(\mathbf{s})]\frac{\partial^2\nu_i^{\text{theor}}(\mathbf{s})}{\partial \mathbf{s}_k\partial \mathbf{s}_l}\right\} \quad (4)$$

Equations 3 and 4 can be considered to define a gradient vector and a Hessian (second-derivative) matrix, respectively, of the merit function with respect to the scaling factors. These quantities can be used to minimize $\chi^2(\mathbf{s})$ using, for example, a standard Newton−Raphson optimization.

To completely define the gradient and Hessian for a given set of scaling factors, we need to evaluate the partial derivative quantities $\partial\nu_i/\partial \mathbf{s}_k$ and $\partial^2\nu_i/\partial \mathbf{s}_k\partial \mathbf{s}_l$. These quantities measure how the theoretical frequencies change with scaling of the force constant matrix elements. The first derivatives are straightforward to determine using vibrational perturbation theory (see, for example, Mills[29]). A change $\delta\mathbf{F}_k$ in the force constant matrix changes the ith vibrational frequency by

$$\delta\nu_i = (4\pi\nu_i)^{-1}(\mathbf{L}\,\delta\mathbf{F}_k\,\mathbf{L})_{ii} \qquad (5)$$

where $\mathbf{L}$ are the original unperturbed modes.

The second derivatives, $\partial^2\nu_i/\partial \mathbf{s}_k\partial \mathbf{s}_l$, are more awkward to evaluate, but they are much less important and can arguably be ignored. Their contribution to the second term in eq 4 vanishes if the calculated frequencies are equal to the observed ones and has a random sign error arising from the factor $[\nu_i^{\text{expt}} - \nu_i^{\text{theor}}(\mathbf{s})]$, causing the partial second derivative contributions to cancel out when summed over i. Consequently, when evaluating eq 4, we take only the first derivative product—the first term on the right-hand side—and ignore any contributions from the partial second-derivative terms. No convergence difficulties have been encountered in the past 20 years using the original SCALE2 program,[28] which also omits the second term in eq 4.

Given $\chi^2(\mathbf{s})$ and its first and second derivatives (eqs 2, 3, and 4, respectively), it is a straightforward matter to minimize χ^2. We use the eigenvector following (EF) algorithm[30] to calculate the optimization step (the change in the scaling factors), recalculating the Hessian matrix using eq 4 on every cycle. Convergence is attained when any *two* of the following three criteria are satisfied: (1) the maximum gradient component is less than 10^{-6}; (2) $\chi^2(\mathbf{s})$ changes between cycles by less than 10^{-6}; (3) the maximum predicted change in the scaling factors, i.e., the next step size, is less than 3×10^{-4}. Convergence is normally rapid, typically requiring less than six optimization cycles.

3. Applications

We have used our modified SQM procedure to first derive a set of scaling factors for individual stretches, bends, and torsions for a training set of 30 molecules containing C, O, N, H, and Cl and then to apply these scaling factors to predict the fundamental vibrational frequencies of a further 30 molecules, all taken from the recent literature (post-1993). Our basic level of theory is B3-LYP/6-31G*, and all molecular geometries were fully optimized at this level using the standard defaults in GAUSSIAN 94,[31] with force constants calculated analytically at the optimized geometries. As mentioned in the Introduction, this was one of the best methods found by Scott and Radom using a single scaling factor[10] and furthermore was the level used in a previous work on transferable scaling factors from this group.[18] We use many of the molecules from this previous study[18] in our training set. For our least-squares fit of the scaling factors, we use a weighting factor (w_i in eq 2) of $1000/\nu_i^{\text{expt}}$, with ν_i^{expt} given in cm^{-1}.

The 30 molecules chosen for our training set are shown in Figure 1. They comprise 17 of the 20 molecules in the training set of ref 18, together with 7 of the additional 11 molecules used in ref 18 to check the reliability of the derived scaling factors, plus 6 extra molecules, 4 of which contain chlorine (specifically C−Cl bonds). Certain of the molecules used in

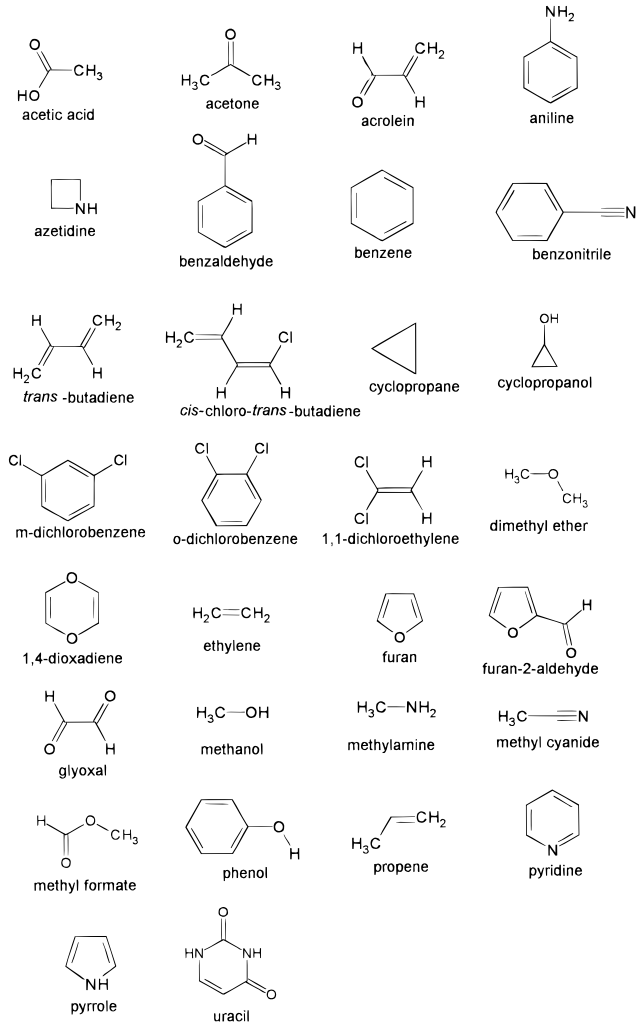


Figure 1. Schematic of the 30 molecules in the training set (see Tables 1 and 2).

ref 18 were deliberately omitted. We left out formaldehyde, as this molecule has no proper torsions and would be the only molecule of the 60 examined with an out-of-plane bend (see the discussion at the beginning of section 2); hydrazine was omitted, as it is the only molecule with an N−N bond (additionally, several of its vibrational frequencies are not well-established); we also left out formic acid and nitrobenzene. Additionally, a review of the experimental data used in ref 18 persuaded us to eliminate ethanol and oxetane; they were replaced by cyclopropanol[32] and 1,4-dioxadiene,[33] molecules with similar structural motifs.

As mentioned in section 2, we describe the molecular geometry in terms of all possible stretches, planar bends, and proper torsions that can be generated on the basis of the atomic connectivity, replacing planar bends by linear coplanar and perpendicular bends for near-linear arrangements of atoms. On the basis of previous experience from ref 18, together with some experimentation, we group the primitive coordinates into the following 11 sets, each with its own scaling factor (note that X is any non-hydrogen atom): (1) all X−X stretches *except* those involving chlorine; (2) C−Cl stretches; (3) C−H stretches; (4) N−H stretches; (5) O−H stretches; (6) all X−X−X bends; (7) all X−X−H bends; (8) all H−C−H bends; (9) all H−N−H bends; (10) all torsions; (11) all linear bends.

Table 1 shows the optimized scaling factors derived from our training set together with the scaling factors derived in ref 18 for comparison (the scaling factor for the C−Cl stretch was obtained in a later study[34]). The two sets of scaling factors are similar, but are not the same. The scaling set from ref 18 has 12 scaling factors, whereas our primitive scaling set has 11, one less. Additionally, the grouping of the coordinate types is somewhat different. In ref 18 *all* X−H stretches are grouped together, whereas we consider C−H, N−H, and O−H stretches separately. On the other hand, we group together all X−X−H bends, while X−C−H bends are considered separately from X−N−H and X−O−H bends in ref 18; we also group *all* torsions together, while ref 18 distinguishes two distinct types. The extra scaling factor in the earlier set comes from the out-of-plane bends, which we do not use. Having said this, there *are* six analogous coordinate groupings that have very similar scaling factors in the old and new schemes. The H−N−H bend is an exception, probably because it is strongly affected by anharmonicity.

TABLE 1: Optimized Scaling Factors Derived from the Training Set of 30 Molecules (see Figure 1) for the Primitive SQM Scheme Together with Scaling Factors Used for the Standard SQM Scheme Taken from Ref 18 for Comparison (X Refers to a Non-Hydrogen Atom; in the Original Work the C-Cl Stretch Was Scaled Separately)

		value		
scaling factor		primitive this work	standard old	(as per ref 18) reoptimized
stretch	X−X	0.9207	0.922	0.9254
stretch	C−Cl	1.0438	1.017[a]	1.0460
stretch	X−H		0.920	0.9182
stretch	C−H	0.9164		
stretch	N−H	0.9242		
stretch	O−H	0.9527		
bend	X−X−X	1.0144	0.990	0.9923
bend	X−X−H	0.9431		
bend	X−C−H		0.950	0.9473
bend	X−N−H; X−O−H		0.876	0.9047
bend	H−C−H	0.9016	0.915	0.9171
bend	H−N−H	0.8753	0.806	0.8358
out-of-plane bends			0.976	0.9711
torsion	all	0.9523		
torsion	conjugated		0.935[b]	0.9389
torsion	single-bonded		0.831[b]	0.8980
linear deformations		0.8847	0.913	0.8905
total no. of scaling factors		11	12	12

[a] Ref 34. [b] Accidentally switched in original reference; corrected in ref 39 (erratum).

The results for our full training set of 30 molecules are given in Table 2, which shows both root-mean-square (rms) and average (mean) differences between calculated and experimental fundamentals for 663 vibrational frequencies. Note that we did not include in our analysis every single frequency for all 30 molecules; the experimental frequencies were carefully screened to eliminate misassignments and uncertain or unobserved fundamentals. This screening had in fact already been done for the previous study,[18] and we took full advantage of the previous analysis for this work. Complete tables of experimental frequencies, including unweighted frequencies and reassignments, were provided as Supporting Information with ref 18.

We made some changes from the assignments used in ref 18. We carried out a thorough review of the previous experimental data, correcting a few misassignments and minor errors; additionally we found more recent experimental data for five of the molecules in the training set which we used in preference to, or together with, the previous data; these were acrolein,[35] furan,[36] pyridine,[37] pyrrole,[36] and uracil.[38] For a full

TABLE 2: Average and Root-Mean-Square (rms) Deviations Between the 663 Calculated and Experimental Vibrational Frequencies for the 30 Molecules in the Training Set Considered: (a) Collectively (Using the Set of Scaling Factors Shown in Table 1) and (b) Individually (with Scaling Factors Optimized Separately for Each Molecule)

	(a) collectively				(b) individually			
	this work		ref 18		this work		ref 18	
molecule	rms	av	rms	av	rms	av	rms	av
acetic acid	11.12	7.76	15.58	9.97	7.17	5.27	13.15	9.36
acetone	12.47	10.01	12.64	10.52	11.51	9.52	12.59	10.46
acrolein	14.08	11.06	15.56	11.96	10.36	7.40	10.38	7.28
aniline	9.60	6.57	11.69	8.43	6.92	5.64	10.54	7.80
azetidine	11.45	8.81	9.04	7.18	6.25	4.89	6.44	5.02
benzaldehyde	11.00	8.40	10.53	8.16	10.13	8.09	9.74	7.72
benzene	8.49	6.07	6.78	5.57	6.33	5.14	5.87	5.01
benzonitrile	10.38	6.67	10.80	6.87	7.37	5.16	7.34	5.51
trans-butadiene	11.58	8.93	11.93	9.39	9.61	6.83	8.60	5.64
cis-chloro-*trans*-butadiene	14.42	10.37	15.69	11.58	8.71	6.82	9.08	7.06
cyclopropane	10.64	9.19	9.27	7.70	8.30	6.85	8.32	6.95
cyclopropanol	9.31	7.61	21.17	11.75	4.83	3.36	20.60	11.38
m-dichlorobenzene	9.69	6.56	8.32	5.18	7.55	5.06	5.97	4.29
o-dichlorobenzene	8.90	6.96	8.72	6.30	7.04	5.44	7.26	5.78
1,1-dichloroethylene	10.40	7.20	10.94	7.88	2.33	1.55	2.33	1.43
dimethyl ether	13.10	8.86	11.75	7.94	10.62	7.33	9.48	6.30
1,4-dioxadiene	16.82	11.41	18.82	12.90	16.48	10.89	16.43	11.15
ethylene	10.95	9.00	9.67	7.87	6.92	5.72	4.12	3.77
furan	9.34	6.91	6.42	4.69	6.38	4.51	4.80	3.70
furan-2-aldehyde	13.25	9.18	13.21	8.98	12.60	9.66	12.78	9.76
glyoxal	9.84	8.34	10.05	8.39	5.55	3.64	3.79	3.05
methanol	19.45	12.76	27.31	17.83	16.19	9.87	26.58	17.97
methylamine	24.69	15.78	18.74	14.89	10.28	8.17	17.01	13.69
methyl cyanide	12.37	9.66	11.98	8.79	9.12	6.22	9.65	6.28
methyl formate	11.38	9.49	11.03	9.48	10.29	8.89	10.75	9.44
phenol	10.27	5.98	15.74	8.18	9.88	5.34	12.34	6.94
propene	12.34	9.25	11.57	8.92	11.33	7.98	10.02	6.76
pyridine	9.60	6.63	9.57	6.72	8.21	6.21	8.25	6.20
pyrrole	13.69	8.72	11.90	8.44	9.52	6.22	8.17	4.91
uracil	12.94	10.49	13.77	10.90	8.59	6.56	8.99	6.60
total (30 molecules)	12.04	8.49	12.60	8.57				
prefingerprint region:	10.94		10.36					
fingerprint region (500−2500):	11.21		10.15					
postfingerprint region:	14.92		19.24					
standard deviation:		8.55		9.24				

listing of references to the experimental fundamentals, see the previous study.[18] A complete listing of the experimental fundamentals used in this work and their weights and symmetry assignments for both the training *and* the test set (60 molecules in total) is available on request from the authors.

Table 2 gives rms and average deviations for the *full* training set, both as a total, and for each molecule in the training set individually, and also with scaling factors optimized for *each* molecule separately. The second set of calculations provides an indication of the quality of the total fit (i.e., considering all 30 molecules in the training set collectively) and acts as a pointer to any "problem" molecules, for which the global fit is poor. Also included in Table 2 for comparison are results using the previous SQM analysis as per ref 18. For a fairer and more direct comparison between the primitive and the "standard" SQM scheme, we reoptimized all the scaling factors derived in ref 18 for our enlarged training set. The new set of scaling factors is shown in Table 1; it is these reoptimized scaling factors that are used for all comparisons between the two schemes in this work. Many of the reoptimized parameters have values similar to those quoted in ref 18; however, some scaling factors have changed more significantly.

The first thing to note is that, despite the different scaling factors and coordinates (primitives compared to natural internals), the performance of the two SQM schemes for the training set collectively is—on the average—very similar. However, this apparent identity *on the average* is rather deceptive, as there are some noticeable differences. If, following ref 18, we divide the vibrational spectrum into three regions, the important fingerprint region (from 500−2500 cm^{-1}) and pre- (below 500 cm^{-1}) and post- (above 2500 cm^{-1}) fingerprint regions, then the previous SQM scheme performs about the same in the prefingerprint region, somewhat better in the fingerprint region, and significantly worse in the postfingerprint region. The better performance of our primitive SQM scheme in the high-frequency region is doubtless due to our use of separate scaling factors for C−H, N−H, and O−H stretches; however, we were unable to improve our results significantly in the fingerprint region by introducing separate scaling factors for some of the bends and torsions, as was done in ref 18 (see Table 1).

Another difference is a seemingly more "average" performance on a molecule-by-molecule basis for our primitive SQM scheme; this can be seen by looking at the range of rms values for the collective scaling (8.49−24.69 cm^{-1} for the primitive scheme compared to 6.42−27.31 cm^{-1} for the original) and also at the standard deviation (8.55 cm^{-1} versus 9.24 cm^{-1}).

This trend is even more marked when the scaling factors are optimized for each individual molecule. The primitive SQM scheme has only two molecules with an optimized rms value lower than 5.00 and just one molecule (1,4-dioxadiene) with an average deviation from experiment of more than 10.00 cm^{-1}. This compares to four systems with optimized rms values lower than 5.00 and five with average deviations greater than 10.00 cm^{-1}—including an almost unchanged average error of 17.97 cm^{-1} for methanol—with the original scheme (see columns b in Table 2).

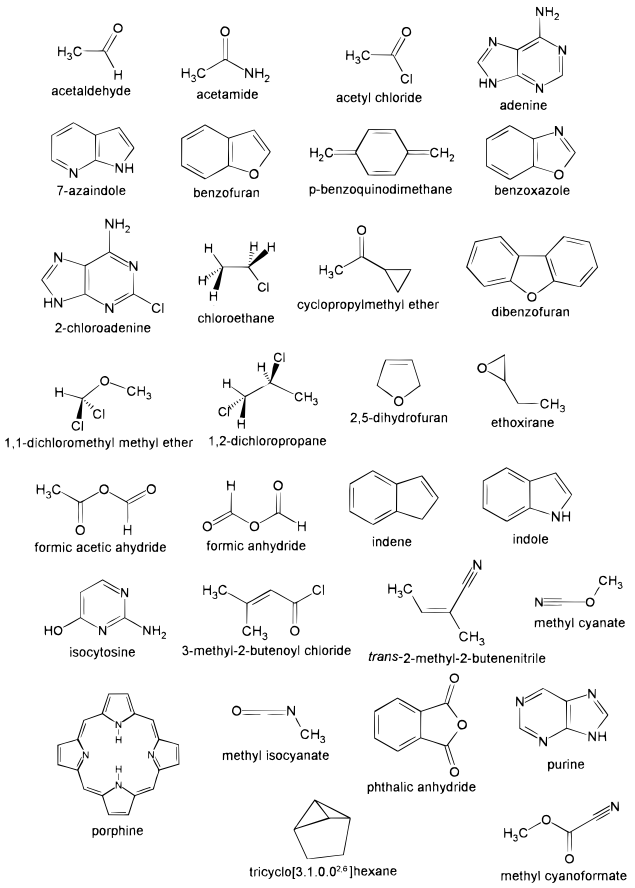


Figure 2. Schematic of the 30 molecules in the test set (see Tables 3 and 4).

We used the scale factors derived from the training set (Table 1) to scale the force constants for the test set of 30 molecules shown in Figure 2. All these molecules were taken from the recent literature; the bulk of them represent new gas-phase IR and Raman vibrational spectra published during or after 1994. A listing of the test set, the experimental references, the type of experiment (gas phase or inert gas matrix), and notes on the fundamentals reassigned or given zero weight is provided in Table 3 (molecules with an asterik in Table 3 are discussed in more detail later in this section).

Table 4 gives rms and average deviations for the *full* test set, both as a total, and for each molecule in the test set individually, and also with scaling factors optimized for *each* molecule separately. We again include in Table 4 results using the previous SQM analysis.[18]

Looking at the results for the test set overall, then for the primitive SQM scheme the total rms and average deviations over all 30 molecules are actually slightly better for the test set than they were for the training set. In other words, the derived scaling factors give an equally good performance for both sets and would thus appear to be highly transferable. The mean error between calculated and experimental fundamentals is just 8.49 cm^{-1} for the training set (Table 2) and 8.23 cm^{-1} for the test set (Table 4). The original SQM scheme does not perform as well; although the mean error for the test set (8.54 cm^{-1}) is virtually the same as for the training set (8.57 cm^{-1}), the total rms deviation has increased, as has the rms deviation in the fingerprint region and the standard deviation. The rms deviations for the test set using the primitive SQM scheme are significantly lower in all regions of the spectrum than are those for the original SQM scheme. The better performance of the primitive scheme is clear and suggests that scaling parameters are not as generally transferable when natural internals are used than they are when individual primitives are scaled. Thus not only is the primitive SQM scheme simpler from a user's point of view but derived scaling factors appear to be more transferable. Scaling individual primitives should allow more flexibility, so this result is perhaps not surprising.

If we look at those molecules that perform well with the original SQM scheme, we find five with rms values less than 8 cm^{-1}: benzofuran, benzoxazole, dibenzofuran, indole, and purine. Every one of these systems is cyclic, contains one or more aromatic rings, and is completely planar. The performance of the original SQM scheme is better than the primitive scheme for these five molecules. For all other molecules apart from indene (which is cyclic, partly aromatic, and almost planar) the rms deviation with the original SQM scheme is similar or higher—often significantly—than with the primitive scheme. The high local symmetry of these planar cyclic systems probably ensures that natural internal coordinates match well the actual deformations (the normal modes) in these molecules; this is probably not the case for the acyclic systems, for which the extra flexibility gained by scaling the individual primitives gives rise to an overall improved performance.

Certain of the molecules in the test set warrant extra attention (those marked with an asterisk in Table 3), and we discuss these further below.

Adenine and 2-Chloroadenine. These two molecules (along with purine) were the subject of a low-temperature IR argon matrix isolation study by Nowak et al.[43] These workers tabulated between 50 and 60 observed frequencies per molecule, but were able to definitely assign only about 25 fundamentals for each species. In calculating the rms and mean deviations for adenine and 2-chloroadenine, we included only those frequencies that had definitely been assigned in ref 43. We of course obtained values for all fundamentals and, on the basis of our results, are able to assign fundamentals to appropriate experimentally observed bands. We give our calculated fundamentals together with all experimentally observed bands in Table 5. Our quoted values were derived using the fixed scaling factors reported in Table 1 (we have not refined the fit by further optimization). As can be seen, the agreement between the calculated and experimental values is good, and we are able to propose reasonable assignments of the observed signals to corresponding fundamentals. Note that our assignments for the fundamentals are simply labels and do not correspond to any particular vibrational motion.

We do have some disagreement with the tentative assignments of Nowak et al.[43] For adenine, peaks at 1639, 1633, and 1612 cm^{-1} are all assigned as fundamentals in ref 43; we calculate only two frequencies in this region (1617 and 1602 cm^{-1}) and consider that the 1639 cm^{-1} signal does not correspond to a fundamental. For 2-chloroadenine there is no band corresponding to our calculated fundamental at 3130 cm^{-1}; we consider this frequency to be unobserved in the experimental spectrum.

7-Azaindole. Experimental fundamentals for 7-azaindole were obtained from gas-phase IR spectra and assigned with the help of (traditional) SQM calculations at the HF/6-31G** level.[44] We omitted two fundamentals from our own SQM procedure: an A″ frequency at 204 cm^{-1}, which was estimated from microwave spectra (and had a quoted uncertainty of 30 cm^{-1}), and an A′ frequency at 1252 cm^{-1}, which came from a polycrystalline IR spectrum of 7-azaindole as a CsI pellet. Overall agreement with the experimental spectrum was good (see Table 4) except for an A″ mode at 925 cm^{-1}. The

TABLE 3: References and Notes to the 30 Molecules in the Test Set (See Figure 2)[a]

molecule	symmetry	type of experiment	notes	no. of modes	ref
acetaldehyde	C_s	gas (IR)	all fundamentals included	15/15	40
acetamide	C_1	Ar matrix	3 fundamentals unassigned NH_2 wag (269 cm^{-1}) given zero weight assignments based on C_s symmetry	17/21	41
acetyl chloride	C_s	gas	2 fundamentals (1417 and 2990 cm^{-1}) only seen in solid given zero weight	13/15	42
adenine*	C_1	Ar matrix	only fundamentals with definite assignments included (see text)	24/39	43
7-azaindole*	C_s	gas (IR)	2 fundamentals (204 and 1252 cm^{-1}) estimated or from solid plus a″ mode at 925 cm^{-1} given zero weight (see text)	36/39	44
benzofuran	C_s	gas	3 fundamentals assigned based on combination bands given zero weight	36/39	45
p-benzoquinodimethane	D_{2h}	Ar matrix	lowest b_{3u} fundamental unobserved	41/42	46
benzoxazole	C_s	gas	fundamental at 971 cm^{-1} assigned from combination bands given zero weight	35/36	45
2-chloroadenine*	C_1	Ar matrix	only fundamentals with definite assignments included (see text)	24/39	43
chloroethane	C_s	gas	2 fundamentals unobserved	16/18	47
cyclopropylmethyl ketone	C_s	gas	a′ fundamental at 2966 cm^{-1} given zero weight; bands at 862 and 900 cm^{-1} reassigned	35/36	48
dibenzofuran	C_{2v}	gas	uncertain a_2 assignment at 890 cm^{-1} (found theoretically at 748 cm^{-1}) and a_1 fundamental at 1170 cm^{-1} given zero weight	55/57	49
1,1-dichloromethyl methyl ether	C_s	gas	2 fundamentals (2855 and 2963 cm^{-1}) given zero weight	19/21	50
1,2-dichloropropane	C_1	gas	all fundamentals included	27/27	51
2,5-dihydrofuran	C_{2v}	gas	2 fundamentals assigned based on combination bands given zero weight	25/27	52
ethyloxirane	C_1	gas	6 fundamentals unobserved	27/33	53
formic acetic anhydride*	C_s	gas (IR)	6 fundamentals unobserved (see text) a″ mode at 1133 cm^{-1} given zero weight	17/24	54
formic anhydride	C_s	gas	2 fundamentals unobserved lowest mode (85 cm^{-1}; from microwave spectrum) given zero weight	12/15	55
indene*	C_s	gas	a′ fundamental at 947.1 cm^{-1} given zero weight (see text)	44/45	56
indole	C_s	gas	3 fundamentals (from liquid or assigned from combinations) given zero weight	39/42	45
isocytosine	C_1	Ar matrix	6 fundamentals unobserved lowest mode given zero weight	26/33	57
3-methyl-2-butenoyl chloride*	C_s	gas	3 fundamentals unobserved (see text) a″ mode at 1071 cm^{-1} given zero weight	32/36	58
trans-2-methyl-2-butenenitrile	C_s	gas	1 fundamental unobserved 2 modes from solid given zero weight	30/33	59
methyl cyanate	C_s	gas (IR)	5 fundamentals unobserved	10/15	60
methyl cyanoformate	C_s	gas	lowest 4 fundamentals from spectra of solid given zero weight	17/21	61
methyl isocyanate	C_s	gas	4 fundamentals only observed in solid spectra given zero weight	11/15	62
phthalic anhydride	C_{2v}	Ar/Ne matrix	6 A_2 fundamentals unobserved	33/39	63
porphyrin*	D_{2h}	Xe matrix	IR data from Radziszewski et al; nonresonance Raman from Kozlowski et al. (primarily inactive a_u, b_{2g} and b_{3g} modes unobserved in experimental spectra)	69/108	64
purine	C_s	Ar matrix	7 fundamentals unobserved/unassigned	26/33	43
tricyclo[3.1.0.0^{2,6}]hexane*	C_{2v}	gas (IR)	2 fundamentals unobserved/unassigned 2 modes (liquid) given zero weight	32/36	65

[a] Molecules marked with * are discussed further in the text.

corresponding mode in our theoretical SQM spectrum occurred at around 845 cm^{-1}, below two A′ fundamentals at 898 and 870 cm^{-1}. We consider that an experimental reassignment of this band is perhaps needed. (For the final rms and mean deviations quoted for 7-azaindole in Table 4, we omitted this frequency in addition to the two already mentioned.)

Formic Acetic Anhydride. Experimental fundamentals were obtained from gas-phase IR spectra recorded at room temperature[54] and assigned with the help of (traditional) SQM calculations at the HF level using both 4-21G[16] and 6-31G** basis sets. There were six fundamentals unobserved in the IR spectra; additionally there was an A″ fundamental assigned at 1133 cm^{-1} which was found at 1049 cm^{-1} in the theoretical spectrum. This frequency was given zero weight in our SQM analysis. We note that there is a very intense A′ fundamental also at 1049 cm^{-1} (the most intense band in the entire spectrum), and the A″ fundamental may be "lost" within this intense A′ band. If this is the case, then the band at 1133 cm^{-1} is misassigned.

Indene. Fundamentals for indene were obtained from a high-quality IR/Raman gas- and liquid-phase study by Klots.[56]

TABLE 4: Average and Root-Mean-Square (rms) Deviations between the 843 Calculated and Experimental Vibrational Frequencies for the 30 Molecules in the Test Set Considered: (a) Collectively (Using the Set of Scaling Factors Shown in Table 1) and (b) Individually (with Scaling Factors Optimized Separately for Each Molecule)

	(a) collectively				(b) individually			
	this work		ref 18		this work		ref 18	
molecule	rms	av	rms	av	rms	av	rms	av
acetaldehyde	17.48	10.85	18.76	11.75	14.52	9.62	14.50	9.51
acetamide	15.36	11.97	20.95	14.14	9.66	6.84	10.13	7.48
acetyl chloride	13.99	11.40	15.71	12.57	11.58	8.48	12.84	10.53
adenine	13.28	9.97	18.07	11.57	7.50	5.44	6.46	5.18
7-azaindole	9.05	6.17	10.14	7.06	8.43	5.56	8.43	5.92
benzofuran	5.95	4.36	5.16	4.23	5.23	4.05	4.76	3.73
p-benzoquinodimethane	13.70	10.28	14.46	10.62	11.62	8.43	11.34	8.87
benzoxazole	5.22	4.00	4.83	3.83	4.32	3.46	3.93	2.90
2-chloroadenine	11.82	9.42	14.12	11.39	5.36	4.40	10.35	8.18
chloroethane	11.37	9.57	11.12	8.91	8.98	5.64	8.71	5.74
cyclopropylmethyl ketone	11.30	9.78	10.58	8.52	10.42	8.46	9.25	7.36
dibenzofuran	7.71	5.96	6.55	5.13	5.64	4.58	5.82	4.48
1,1-dichloromethyl methyl ether	14.26	10.61	18.15	14.87	7.59	6.30	9.22	7.59
1,2-dichloropropane	13.98	10.68	15.12	11.13	5.37	4.23	6.02	4.93
2,5-dihydrofuran	11.70	9.34	11.58	8.62	10.65	7.57	11.16	8.16
ethyloxirane	12.78	9.25	12.75	8.14	10.57	8.35	11.33	7.82
formic acetic anhydride	10.63	8.18	12.52	9.13	7.20	5.78	12.07	8.83
formic anhydride	10.23	7.11	11.32	7.98	3.93	3.40	9.08	6.73
indene	12.06	6.64	10.98	5.60	9.87	6.17	9.51	5.70
indole	6.73	4.92	5.43	4.16	5.46	4.48	4.49	3.69
isocytosine	19.43	13.62	22.22	14.95	13.17	9.39	18.01	12.66
3-methyl-2-butenoyl chloride	13.44	8.83	13.48	8.96	12.13	9.25	12.03	8.70
trans-2-methyl-2-butenenitrile	10.15	7.70	10.24	7.90	9.15	6.05	9.66	6.87
methyl cyanate	13.84	12.95	14.02	12.80	11.01	7.66	14.02	12.80
methyl cyanoformate	11.68	10.30	12.10	10.77	9.16	6.81	9.40	6.61
methyl isocyanate	16.11	10.30	19.74	14.13	8.16	5.73	13.34	10.52
phthalic anhydride	14.71	10.25	14.43	9.88	13.36	8.59	13.52	9.16
porphyrin[a]	7.69	5.45	8.39	5.88	6.88	5.06	9.57	5.87
purine	8.15	5.73	7.75	5.55	7.12	4.66	7.00	4.55
tricyclo[3.1.0.0^{2,6}]hexane	16.18	11.38	15.97	11.66	10.90	7.91	10.43	7.83
total (30 molecules)	11.77	8.23	12.85	8.54				
prefingerprint region:	8.81		9.68					
fingerprint region (500−2500):	10.69		11.72					
postfingerprint region:	16.70		18.05					
standard deviation:		8.41		9.71				

[a] Collective scaling for porphyrin excludes N−H stretching frequency; individual scaling includes it (see text).

Overall the agreement was very good, with a mean error of only 6.64 cm^{-1} over 44 fundamentals (Table 4). There was one relatively minor anomaly; the assigned A′ fundamental at 947.1 cm^{-1} occurred in the calculated spectrum below two A″ fundamentals at 926.4 and 943 cm^{-1}. This "misordering" persisted even when the scale factors were optimized specifically for indene. Agreement would noticeably improve in this region if there were a reassignment of this A′ fundamental to A″ and a corresponding reassignment of the A″ fundamental at 926.4 cm^{-1} to A′. However, this is only an observation, and we are not proposing a reassignment at this stage. The "problem" A′ mode was omitted when calculating the rms and mean deviations in Table 4.

We show in Figure 3 a plot of the predicted B3-LYP/6-31G* infrared spectrum of indene in the range 400−1200 cm^{-1}, together with the experimental vapor-phase spectrum taken from ref 56. This gives a better idea as to the overall quality of our SQM fit. We made no attempt to simulate the rotational band envelope; despite this there is good agreement both in the peak positions and relative intensities.

3-Methyl-2-butenoyl Chloride. Reported experimental frequencies are from a Raman and IR study by Durig et al.[58] Three fundamentals were unobserved in the gas-phase spectrum. Additionally an A″ fundamental at 1071 cm^{-1} (assigned as a double methyl rock) was given zero weight in the SQM analysis. This mode occurred at ∼989 cm^{-1} in the calculated spectrum below an A′ fundamental at 1011 cm^{-1}. Perhaps this is a misassignment?

Porphyrin. Free base porphyrin has been extensively studied in this group,[66,67] and a complete vibrational assignment and SQM analysis at the B3-LYP/6-31G* level has been carried out[67,64b] based primarily on the experimental data of Radziszewski et al.[64a] Sixty-nine out of 108 fundamentals were included in the SQM analysis, with primarily unobserved a_u, b_{2g}, and b_{3g} modes omitted. Agreement with experiment is excellent, with the primitive SQM scheme, if anything, performing even better than the original SQM scheme.

Note that, when using the fixed ("transferable") scaling factors, we gave the b_{3u} N−H stretching mode at 3324 cm^{-1} zero weight (the second N−H stretch, with a_g symmetry, is unobserved). The calculated value for this fundamental is ∼80 cm^{-1} in error due to a too large scaling factor. The inappropriateness of "transferable" N−H stretch scaling factors for the labile N−H bonds in porphyrin has already been noted.[66,67] When the scaling factors were optimized specifically for porphyrin, this mode was included with full weighting. The reoptimized N−H stretch scaling factor using the primitive SQM scheme changed from 0.9242 (Table 1) to 0.8787, and the calculated N−H stretch mode was in essentially exact agreement with experiment. With the original SQM scheme, the rms deviation increases with specifically optimized scaling factors (Table 4) because the N−H stretch is not scaled separately from the C−H stretches.

TABLE 5: Experimentally Observed Signals (cm^{-1}) in the IR Spectra of Adenine and 2-chloroadenine isolated in Ar Matrixes Together with Calculated Primitive SQM Frequencies (Fixed Scaling Factors) and Proposed Assignments of Fundamentals (Experimental Signals from Ref 43; Calculated Fundamentals This Work)

adenine						2-chloroadenine					
expt	calc	assign	expt	calc	assign	expt	calc	assign	expt	calc	assign
3565	3575	Q1	1127	1122	Q18	3563	3587	Q1	1203		
3557			1078			3561			1180		
3498	3510	Q2	1061		Q19?	3492	3508	Q2	1170		
3489			1032	1055	Q19?	3489			1137	1130	Q16
3448	3457	Q3	1017			3449	3466	Q3	1057	1054	Q17
3441			1005	997	Q20	3442			1042	1025	Q18
3066	3126	Q4	958	949	Q21		3130	Q4	943	938	Q19
3057	3052	Q5	927	934	Q22	1781			933		Q20?
3041			887	892	Q23	1691			930	933	Q20?
1693			869			1682			847	822	Q21
1659			848	824	Q24	1636			795	780	Q22
1651			802	789	Q25	1633	1617	Q5	781		
1645			717			1609	1601	Q6	776	771	Q23
1639		?	698	707	Q26	1595			736		
1633	1617	Q6	688	670	Q27	1577	1554	Q7	687	675	Q24
1626			655	655	Q28	1466	1484	Q8	674	662	Q25
1619			648			1459	1452	Q9	639	644	Q26
1612	1602	Q7	610	612	Q29	1423	1410	Q10	634	640	Q27
1599	1560	Q8	591			1380			565	570	Q28
1482	1484	Q9	583			1375			530	535	Q29
1474	1471	Q10	566	565	Q30	1365	1372	Q11		529	Q30
1419	1407	Q11		531	Q31	1361			516	516	Q31
1389	1388	Q12		526	Q32	1349			505		
1358			513	507	Q33	1343				395	Q32
1345		Q13?	503	497	Q34	1335	1334	Q12		341	Q33
1334	1335	Q13?		311	Q35	1320			276	273	Q34
1328	1327	Q14	276	293	Q36	1313	1313	Q13	266	247	Q35
1290	1296	Q15	242	273	Q37	1290			239	238	Q36
1246			214	210	Q38	1260				182	Q37
1240	1238	Q16		162	Q39	1251				174	Q38
1229	1221	Q17				1249	1247	Q14		112	Q39
1133						1234	1222	Q15			

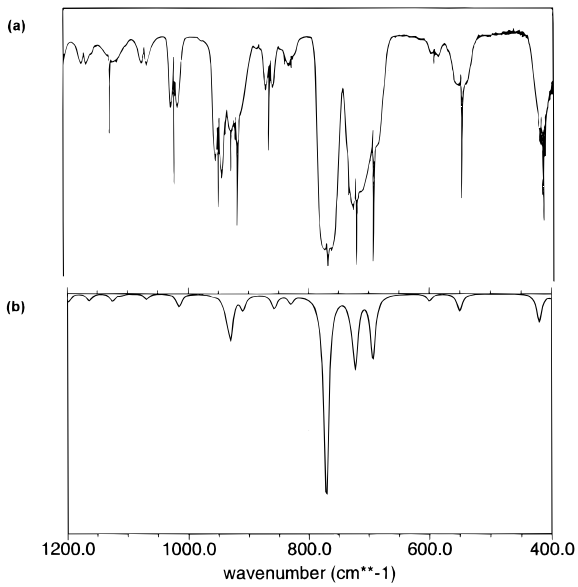


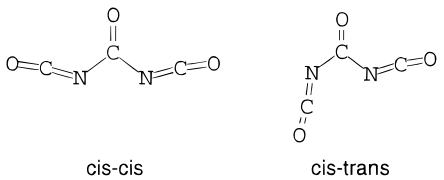
Figure 3. Infrared spectrum of indene from 400 to 1200 cm^{-1}: (a) vapor-phase spectrum from ref 56; (b) predicted B3-LYP/6-31G* spectrum using the scaling factors listed in Table 1 and a Lorentzian band profile with half-width 5 cm^{-1}.

Tricyclo[3.1.0.0^{2,6}]hexane. This molecule is a small "cage compound", highly strained, with two five-membered and two three-membered rings "fused" together (see Figure 2). It was deliberately included as an example of a system with a rather complex topology and one for which standard scaling factors ought not to be particularly appropriate. The experimental data (from Davis and Tan[65]) seem to be of good quality, with IR spectra recorded both in the gas phase and in an argon matrix at 10 K. The Raman spectrum of the liquid was also recorded. The major contaminant in the sample was benzene, and benzene spectra were recorded and subtracted from the observed spectra, to give spectra of "pure" tricyclo[3.1.0.0^{2,6}]hexane.

Two fundamentals were unobserved or unassigned in all recorded spectra, and two fundamentals (an A_2 mode at 260 cm^{-1} and a B_1 mode at 944 cm^{-1}) only observed in the Raman spectrum of the liquid were given zero weight. Apart from these four, all other fundamentals were included in the SQM analysis.

The results, while not among the best, are not too bad, with rms and mean deviations using fixed scaling factors of 16.18 and 11.38 cm^{-1}, respectively. These values of course improve when the scaling factors are optimized, but by no more than for most other molecules. However, as expected, some of the scaling factors change quite dramatically, most notably the CCC bend (from 1.0144 to 0.7428).

Rotational Isomerism in Carbonyl Diisocyanate. Experimentally, carbonyl diisocyanate has been shown to exist in the gas phase as a mixture of two planar conformers cis−cis and cis−trans, with the cis−cis conformer the more stable. (Our calculations, including ZPVE, give an energy difference of ∼1.2 kcal/mol.)



Recently, Balfour et al.[68] have recorded the IR spectrum of carbonyl diisocyanate in the gas phase and both the IR and

TABLE 6: Observed Frequencies (cm^{-1}) in the Gas Phase IR Spectrum of Carbon Diisocyanate and Proposed Assignment of Fundamentals (Taken from Ref 68) Mapped against Calculated SQM Frequencies for the Cis−Cis and Cis−Trans Isomers Using the Standard Scaling Factors Reported in Table 1 (Theoretical IR Intensities Given in km/mol)[a]

experimental spectrum					calculated spectrum					
	intensity	proposed assignment			cis−cis			cis−trans		
2275	vs	ν_{as}(NCO)			2280	0	A_1	2272	498	A′
2241	vs	ν_{as}(NCO)			2239	2488	B_1	2236	2236	A′
1772	vs	ν(C═O)$_{cis-trans}$						1763	524	A′
1741	vs	ν(C═O)$_{cis-cis}$			1731	323	A_1			
1426	vs	ν_s(NCO)			1437	768	B_1	1427	354	A′
1403	s	ν_s(NCO)			1423	2	A_1	1395	114	A′
1075	vs	ν(NCN)$_{cis-cis}$	out-of-phase		1060	582	B_1			
1068	vs	ν(NCN)$_{cis-trans}$	out-of-phase					1061	516	A′
855	mw	ν(NCN)$_{cis-cis}$	in-phase		841	52	A_1			
831	ms	ν(NCN)$_{cis-trans}$	in-phase					805	42	A′
					720	52	B_1			
737	m, sh	γ(CO)$_{cis-cis}$	out-of-plane	?	720	32	B_2			
726	s	γ(CO)$_{cis-trans}$	out-of-plane	?				711	27	A″
687	vw	δ(NCO)	in-plane					682	20	A′
657	vw, sh	δ(NCO)	in-plane		597	13	A_1			
								619	3	A′
615	s	δ(NCO)	out-of-plane		587 ?	52	B_2	586 ?	27	A″
609	s	δ(NCO)	out-of-plane		587 ?	52	B_2	586 ?	27	A″
					585		A_2	576	28	A″
538	w	δ(CO)$_{cis-cis}$	in-plane	?	583	2	B_1			
522	vw	δ(CO)$_{cis-trans}$	in-plane	?				516	6	A′
469[b]	mw	δ(NCN)$_{cis-cis}$		?				468	10	A′
413[b]	m	δ(NCN)$_{cis-trans}$		?	402	1	A_1			
∼155[b]	vw, sh	δ(CNC)			157	1	B_1	149	1	A′
					137		A_2	106	0	A″
					87	0	A_1	80	1	A″
					53	0	B_2	79	0	A′
rms deviation					21.15			11.55		
mean deviation					15.06			8.85		

[a] Abbreviations: s, strong; m, moderate; w, weak; v, very, sh, shoulder. [b] From Raman spectrum of liquid; all other experimental values from gas-phase IR.

Raman spectra of the liquid. Most of the fundamentals were assigned, but in several cases it was unclear which bands came from which isomer. In our final example, we use the primitive SQM scheme to completely assign the experimental spectra.

As a first step, we calculated the fundamentals of both conformers (at the B3-LYP/6-31G* level) using the standard set of scaling factors reported in Table 1. Only three of these scaling factors are applicable to carbon diisocyanate: the X−X stretch, the X−X−X bend, and the torsion (the NCO angle in both species is less than 175°, our cutoff value for near-linearity). We match our calculated frequencies to the experimental values and assignments reported in ref 68 in Table 6. In mapping the calculated fundamentals to experiment, we assume that the classification of each vibration given in ref 68 is essentially correct, at least as far as being in-plane or out-of-plane is concerned.

The cis−cis isomer has C_{2v} symmetry, and its fundamentals span the irreducible representations $7A_1 + 2A_2 + 6B_1 + 3B_2$. The A_2 modes are IR-inactive, and so all observable out-of-plane fundamentals in the IR spectrum must be B_2. This immediately fixes two of the calculated fundamentals for cis−cis carbon diisocyanate: the B_2 mode calculated at 720 cm^{-1} corresponds to the signal at 737 cm^{-1}, and the one at 587 cm^{-1} to either the 615 cm^{-1} or the 609 cm^{-1} signal. Similar considerations apply to the cis−trans isomer (C_s) and the A″ modes. The rest of the calculated fundamentals are mapped around these modes.

Although not indicated as such in Table 6, Balfour et al.[68] consider that the four vibrations classified as δ(NCO) (NCO bends) are due to rotational isomerism, with two signals coming from each isomer. Given this, then experimentally only 13 fundamentals for each isomer have been assigned. Our calculations show that both isomers have three low-frequency modes, which means that there must be two more fundamentals (there are 18 altogether) for each isomer either not observed or unassigned. In the cis−cis isomer one of these is the inactive A_2 mode (calculated at 585 cm^{-1}), and we have taken the other to be a B_1 mode at 720 cm^{-1}, which may be obscured by the B_2 mode calculated at the same value. In the cis−trans isomer we have taken the additional fundamentals to be an A″ mode at 576 cm^{-1} (calculated) and a weak A′ mode at 619 cm^{-1}.

Note that experimentally it is not certain which of the two bands for the γ(CO), the δ(CO), and the δ(NCN) modes comes from which isomer. We have taken the proposed assignments for the first two of these, but for the δ(NCN) mode our calculations strongly suggest that the given assignment should be reversed, with the band at 469 cm^{-1} coming from the cis−trans isomer and the 413 cm^{-1} band from the cis−cis. Note also that we are not able to assign with certainty the very close δ(NCO) peaks at 609 and 615 cm^{-1}; subsequent calculations (see below) give a somewhat better fit if the signal at 615 cm^{-1} is assigned to the cis−cis isomer.

Given this initial mapping of the calculated and experimentally observed fundamentals, we have refined our calculated fundamentals by optimizing the scaling factors for each isomer separately. Since we are interested in getting the best possible fit (within reason), we have increased the number of scaling factors from 3 to 11, scaling separately the C═O (carbonyl), C−N, N═C, and C═O (cyanate) stretches, the O═C−N, N−C−N, C−N═C, and N═C═O bends, and the O═C═N−C, C═N−C═O, and C═N−C−N torsions. The optimized fit

TABLE 7: Observed Frequencies (cm^{-1}) in the Gas-Phase IR Spectrum of Carbon Diisocyanate and Proposed Assignment of Fundamentals (Modified from Ref 68) Mapped against Calculated SQM Frequencies for the Cis−Cis and Cis−Trans Isomers Using 11 Scaling Factors Optimized Separately for Each Isomer (Theoretical IR Intensities Given in km/mol)

experimental spectrum				calculated spectrum					
	intensity	proposed assignment		cis−cis			cis−trans		
2275	vs	ν_s(NCO)		2277	0	A_1	2275	484	A′
2241	vs	ν_{as}(NCO)		2240	2341	B_1	2240	1640	A′
1772	vs	ν(C═O)$_{cis-trans}$					1772	534	A′
1741	vs	ν(C═O)$_{cis-cis}$		1740	326	A_1			
1426	vs	ν_{as}(CNC)		1431	1077	B_1	1436	459	A′
1403	s	ν_s(CNC)		1410	2	A_1	1396	144	A′
1075	vs	ν(NCN)$_{cis-cis}$	out-of-phase	1052	382	B_1			
1068	vs	ν(NCN)$_{cis-trans}$	out-of-phase				1071	428	A′
855	mw	ν(NCN)$_{cis-cis}$	in-phase	864	53	A_1			
831	ms	ν(NCN)$_{cis-trans}$	in-phase				819	41	A′
				754	77	B_1			
737	m, sh	γ(CO)$_{cis-cis}$	out-of-plane	737	33	B_2			
726	s	γ(CO)$_{cis-trans}$	out-of-plane				726	26	A″
687	vw	δ(NCO)$_{cis-trans}$	in-plane				691	23	A′
657	vw, sh	δ(NCO)$_{cis-cis}$	in-plane	650	9	A_1			
							629	4	A′
615	s	δ(NCO)$_{cis-cis}$	out-of-plane	615	51	B_2			
609	s	δ(NCO)$_{cis-trans}$	out-of-plane				609	27	A″
				612		A_2	598	29	A″
538	w	δ(CO)$_{cis-cis}$	in-plane	549	15	B_1			
522	vw	δ(CO)$_{cis-trans}$	in-plane				520	7	A′
469[b]	mw	δ(NCN)$_{cis-trans}$					471	8	A′
413[b]	m	δ(NCN)$_{cis-cis}$		406	1	A_1			
∼155[b]	vw, sh	δ(CNC)		157	1	B_1	154	1	A′
				140		A_2	108	0	A″
				87	0	A_1	82	1	A″
				54	0	B_2	82	0	A′
rms deviation				8.45			5.08		
mean deviation				5.74			3.30		

[a] Abbreviations: s, strong; m, moderate; w, weak; v, very, sh, shoulder. [b] From Raman spectrum of liquid; all other experimental values from gas-phase IR.

is shown in Table 7. We also give a revised description of the vibrations based on our normal modes. If our fit is reliable, we expect the two unassigned/unobserved fundamentals for each isomer (in addition to the three low-frequency modes) to be A_2 and B_1 modes at ∼612 and ∼754 cm^{-1} for cis−cis diisocyanate and A″ and A′ modes at ∼598 and ∼629 cm^{-1}, respectively, for the cis−trans isomer.

Figure 4 shows various theoretically predicted infrared spectra for carbon diisocyanate (using the data in Table 7) from 400 to 2800 cm^{-1} compared with the observed vapor-phase spectrum from ref 68. All theoretical spectra were plotted using a Lorentzian band profile with a half-width of 8 cm^{-1}. Parts a and b of Figure 4 show the predicted spectrum of each isomer separately, while parts c and d are mixtures; the former has cis−cis and cis−trans isomers in the ratio 1.0:0.26 (the predicted ratio using our calculated Boltzmann factor), and the latter is a 50:50 mixture (equivalent to an energy difference of around 0.4 kcal/mol in favor of the cis−cis isomer). This last spectrum shows perhaps the best agreement with the observed vapor-phase spectrum (Figure 4) in terms of band shape and relative peak intensities, indicating that the two isomers are perhaps closer in energy than our calculations suggest. However, further computations (at the same level of theory, B3-LYP) with a much larger basis set (containing two d and an f polarization function on each atom) fail to significantly reduce the energy difference, and comparison of the predicted HF Raman intensities with the experimental liquid-phase Raman spectrum[68] is inconclusive.

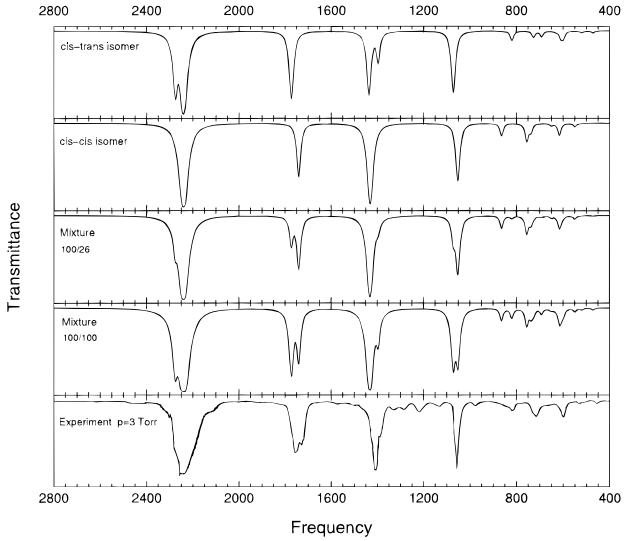


Figure 4. Infrared spectrum of carbonyl diisocyanate from 400 to 2800 cm^{-1}: (a) predicted B3-LYP/6-31G* spectrum of the cis−trans isomer; (b) the same for the cis−cis isomer; (c) predicted spectrum for a mixture of cis−cis and cis−trans isomers in the ratio 1.0:0.26 (this is the predicted ratio at room temperature from the Boltzmann factor, assuming the cis−cis isomer is more stable by ∼1.2 kcal/mol); (d) predicted spectrum for an equal mixture of cis−cis and cis−trans isomers; (e) vapor-phase spectrum at 3 Torr from ref 68.

Summary

We have presented an alternative to the standard scaled quantum mechanical (SQM) force field procedure[15] involving the direct scaling of *individual* (primitive) internal coordinates (stretches, bends, and torsions) from a full set of redundant valence coordinates. Our approach is related to the redundant optimization scheme of Pulay and Fogarasi[21] in that both methods use the concept of a generalized matrix inverse. Our

primitive SQM scheme is simpler and more straightforward than the standard SQM procedure, which requires the generation and sorting of natural internal coordinates for each molecule being examined. Results on a set of 60 molecules with various structural motifs suggest that the additional flexibility involved in the scaling of individual primitive internals, as opposed to the linear combinations of primitives present in natural internals, leads to an increase in accuracy when similar numbers of scaling factors are used.

Using a set of 11 "transferable" scaling factors, the average deviation between calculated and experimental frequencies for 60 molecules involving over 1500 fundamentals is less than 8.5 cm^{-1}. (For the training and test sets combined, all 60 molecules, the rms and mean deviations are 11.89 and 8.34 cm^{-1}, respectively, with a standard deviation of 8.47 cm^{-1}.) To put this in perspective, in their recent study involving 1066 fundamentals using a single (global) scaling factor, Scott and Radom[10] considered a 10% error in the predicted frequencies to be "acceptable", and their best methods had approximately 6% of all calculated fundamentals with an error greater than this target. We have only *eight* fundamentals (out of 1506) with a greater than 10% error, and these are all low frequencies, where the percentage error is always larger. This represents just 0.5% of all calculated fundamentals. Our average percentage error, over all 1506 fundamentals, is 0.9%.

On the basis of scaled B3-LYP/6-31G* force constants, we suggest possible reassignments in the vibrational spectra of 7-azaindole,[44] formic acetic anhydride,[54] and 3-methyl-2-butenoyl chloride,[58] we propose assignments for unassigned fundamentals in the IR spectra of adenine and 2-chloroadenine,[43] and we complete the assignment for the vibrational spectra of the cis−cis and cis−trans isomers of carbonyl diisocyanate.[68]

Acknowledgment. This work was supported by the U.S. Air Force for scientific research under Grant No. F 49620-94-1-0072.

References and Notes

(1) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York, 1955.

(2) (a) Pulay, P. *Mol. Phys.* **1970**, *18*, 473; **1971**, *21*, 329. (b) Pulay, P.; Meyer, W. *J. Mol. Spectrosc.* **1971**, *40*, 59. (c) Meyer, W.; Pulay, P. *J. Chem. Phys.* **1972**, *56*, 2109; **1972**, *57*, 3337. (d) Meyer, W.; Pulay, P. *Theor. Chim. Acta* **1974**, *32*, 253. (e) Pulay, P.; Meyer, W. *Mol. Phys.* **1974**, *27*, 473. (f) Pulay, P.; Török, *F. J. Mol. Struct.* **1975**, *29*, 1123.

(3) (a) Botschwina, P. *Chem. Phys. Lett.* **1974**, *29*, 580. (b) Botschwina, P. *Mol. Phys.* **1974**, *30*, 1029; **1976**, *32*, 729. (c) Botschwina, P.; Meyer, W.; Semkow, A. *Chem. Phys.* **1976**, *15*, 25. (d) Botschwina, P.; Srinivasan, K.; Meyer, W. *Mol. Phys.* **1976**, *35*, 1177.

(4) Schlegel, H. B.; Wolfe, S.; Bernardi, *F. J. Chem. Phys.* **1975**, *63*, 3632; **1977**, *67*, 4194.

(5) Pulay, P. *Mol. Phys.* **1969**, *17*, 197.

(6) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem. Symp.* **1979**, *13*, 225.

(7) (a) Blom, C. E.; Altona, C. *Mol. Phys.* **1976**, *31*, 1377. (b) Blom, C. E.; Otto, L. P.; Altona, C. *Mol. Phys.* **1976**, *32*, 1137. (c) Blom, C. E.; Altona, C. *Mol. Phys.* **1977**, *33*, 875; **1977**, *34*, 177.

(8) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(9) Baker, J.; Muir, M.; Andzelm, J.; Scheiner, A. In *Chemical Applications of Density Functional Theory*; Laird, B. V., Ziegler, T., Ross, R., Eds.; American Chemical Society: Washington, DC, 1996; Chapter 24, pp 342−367 and references therein.

(10) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502.

(11) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(12) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(13) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; AkademieVerlag: Berlin, 1991.

(14) Fogarasi, G.; Pulay, P.; Molt, K.; Sawodny, W. *Mol. Phys.* **1977**, *33*, 1565.

(15) Pulay, P.; Fogarasi, G.; Pongor, G.; Boggs, J. E.; Vargha, A. *J. Am. Chem. Soc.* **1983**, *105*, 7037.

(16) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. *J. Am. Chem. Soc.* **1979**, *101*, 2550.

(17) Fogarasi, G.; Zhou, X.; Taylor, P. W.; Pulay, P. *J. Am. Chem. Soc.* **1992**, *114*, 8191.

(18) Rauhut, G.; Pulay, P. *J. Phys. Chem.* **1995**, *99*, 3093.

(19) Panchenko, Y. N.; De Maré, G. R.; Pupyshev, V. I. *J. Phys. Chem.* **1995**, *99*, 17544.

(20) Pulay, P. *J. Mol. Struct.* **1995**, *347*, 293.

(21) Pulay, P.; Fogarasi, G. *J. Chem. Phys.* **1992**, *96*, 2856.

(22) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.

(23) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **1996**, *17*, 49.

(24) Sellers, H. L.; Klimkowski, V. J.; Schäfer, L. *Chem. Phys. Lett.* **1978**, *58*, 541.

(25) Pulay, P.; Baker, J.; Fleischer, U. To be published.

(26) Califano, S. *Vibrational States*; Wiley: London, 1976.

(27) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes: The Art of Scientifc Computing*, 2nd ed.; Cambridge University Press: New York, 1992.

(28) Pongor, G. *SCALE2*, Eötvös University, Budapest, 1977.

(29) Mills, I. M. *J. Mol. Spectrosc.* **1960**, *5*, 334.

(30) Baker, J. *J. Comput. Chem.* **1986**, *7*, 385.

(31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J. A. *GAUSSIAN 94*, revision C.3; Gaussian, Inc.: Pittsburgh, PA, 1995.

(32) Williams, G. A.; Macdonald, J. N.; Boggs, J. E. *J. Chem. Soc., Faraday Trans.* **1990**, *86*, 2805.

(33) Connett, J. E.; Creighton, J. A.; Green, J. H. S.; Kynaston, W. *Spectrochim. Acta* **1966**, *22*, 1859.

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

(35) Hamada, Y.; Nishimura, Y.; Tsuboi, M. *Chem. Phys.* **1985**, *100*, 365.

(36) Klots, T. D.; Chirico, R. D.; Steele, W. V. *Spectrochim. Acta* **1994**, *50A*, 765.

(37) Chirico, R. D.; Steele, W. V.; Nguyen, A.; Klots, T. D.; Knipmeyer, S. E. *J. Chem. Thermodyn.* **1996**, *28*, 797.

(38) Les, A.; Adamowicz, L.; Nowak, M. J.; Lapinski, L. *Spectrochim. Acta* **1992**, *48A*, 1385.

(39) Rauhut, G.; Pulay, P. *J. Phys. Chem.* **1995**, *99*, 14572 (erratum).

(40) Wiberg, K. B.; Thiel, Y.; Goodman, L.; Leszczynski, J. *J. Phys. Chem.* **1995**, *99*, 13580.

(41) Knudsen, R.; Sala, O.; Hase, Y. *J. Mol. Struct.* **1994**, *321*, 187.

(42) Durig, J. R.; Davis, J. F.; Guirgis, G. A. *J. Raman Spectrosc.* **1994**, *25*, 189.

(43) Nowak, M. J.; Rostkowska, H.; Lapinski, L.; Kwiatkowski, J. S.; Leszczynski, J. *Spectrochim. Acta* **1994**, *50A*, 1081.

(44) Cané, E.; Palmieri, P.; Tarroni, R.; Trombetti, A. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 3213.

(45) Klots, T. D.; Collier, W. B. *Spectrochim. Acta* **1995**, *51A*, 1291.

(46) Yamakita, Y.; Tasumi, M. *J. Phys. Chem.* **1995**, *99*, 8524.

(47) McKean, D. C.; McQuillan, G. P.; Robertson, A. H. J.; Murphy, W. F.; Mastryukov, V. S.; Boggs, J. E. *J. Phys. Chem.* **1995**, *99*, 8994.

(48) Wang, A.; Little, T. S.; Durig, J. R. *Spectrochim. Acta* **1994**, *50A*, 595.

(49) Klots, T. D.; Collier, W. B. *J. Mol. Struct.* **1996**, *380*, 1.

(50) Stidham, H. D.; Vlaservich, A. C.; Hsu, D. Y.; Guirgis, G. A.; Durig, J. R. *J. Raman Spectrosc.* **1994**, *25*, 747.

(51) Guirgis, G. A.; Hsu, D. Y.; Vlaservich, A. C.; Stidham, H. D.; Durig, J. R. *J. Mol. Struct.* **1996**, *378*, 83.

(52) Klots, T. D.; Collier, W. B. *Spectrochim. Acta* **1994**, *50A*, 1725.

(53) Durig, J. R.; Shen, S.; Gounev, T. K.; Wurrey, C. J. *J. Mol. Struct.* **1996**, *379*, 267.

(54) Wu, G.; Shlykov, S.; van Alsenoy, C.; Geise, H. J.; Sluyts, E.; van der Veken, B. J. *J. Phys. Chem.* **1996**, *100*, 11620.

(55) Wu, G.; Shlykov, S.; van Alsenoy, C.; Geise, H. J.; Sluyts, E.; van der Veken, B. J. *J. Phys. Chem.* **1995**, *99*, 8589.

(56) Klots, T. D. *Spectrochim. Acta* **1995**, *51A*, 2307.

(57) Vranken, H.; Smets, J.; Maes, G.; Lapinski, L.; Nowak, M. J.; Adamowicz, L. *Spectrochim. Acta* **1994**, *50A*, 875.

(58) Durig, J. R.; Guirgis, G. A.; Jin, Y. *J. Mol. Struct.* **1996**, *376*, 261.

(59) Durig, J. R.; Drew, A. S.; Guirgis, G. A. *J. Raman Spectrosc.* **1995**, *26*, 933.

(60) Pasinszki, T.; Westwood, N. P. C. *J. Phys. Chem.* **1995**, *99*, 1649.

(61) Durig, J. R.; Groner, P.; Drew, A. S.; Guirgis, G. A.; van der Veken, B. J. *J. Raman Spectrosc.* **1995**, *26*, 43.

(62) Sullivan, J. F.; Heusel, H. L.; Zunic, W. M.; Durig, J. R.; Cradock, S. *Spectrochim. Acta* **1994**, *50A*, 435.

(63) Biernacki, P. R.; Kaszynski, P.; Hess, B. A.; Thulstrup, E. W.; Radziszewski, J. G. *J. Phys. Chem.* **1995**, *99*, 6309.
(64) (a) Radziszewski, J. G.; Nepras, M.; Balaji, V.; Waluk, J.; Vogel, E.; Michl, J. *J. Phys. Chem.* **1995**, *99*, 14254. (b) Kozlowski, P. M.; Jarzecki, A. A.; Pulay, P.; Li, X-Y.; Zgierski, M. Z. *J. Phys. Chem.* **1996**, *100*, 13985.
(65) Davis, S. R.; Tan, P. L. *J. Phys. Chem.* **1994**, *98*, 12236.
(66) Kozlowski, P. M.; Zgierski, M. Z.; Pulay, P. *Chem. Phys. Lett.* **1995**, *247*, 379.
(67) Kozlowski, P. M.; Jarzecki, A. A.; Pulay, P. *J. Phys. Chem.* **1996**, *100*, 7007.
(68) Balfour, W. J.; Fougere, S. G.; Klapstein, D.; Nau, W. M. *Spectrochim. Acta* **1994**, *50A*, 1039.