# An Atlas of Forecasted Molecular Data. 2. Vibration Frequencies of Main-Group and Transition-Metal Neutral Gas-Phase Diatomic Molecules in the Ground State

W. Bradford Davis[†] and Ray Hefferlin*[,‡]

Davis Research Group, Rio Linda, California 95673, and Department of Physics, Southern Adventist University, Collegedale, Tennessee 37315

Received October 4, 2005

This atlas of diatomic-molecular vibration frequencies parallels the previously offered Atlas of Internuclear Separations. The Atlas was produced by mining the data from Huber and Herzberg and training neural network software to forecast new data. New protocols were employed with the powerful software, which was originally designed for forecasting the financial markets. The Atlas presents 1920 additional vibration frequencies for use until critical tables are available to fill the needs more precisely. The precision of the predictions is characterized by the average fractional 1% confidence limit, that is, 10.66%. The accuracies of the predictions are determined in two ways. First, 221 of the 224 Huber and Herzberg data values used for training and validation fall within the prediction confidence limits or fall outside by less than 10% of the Huber and Herzberg values, and 181 values agree (within the limits). Second, 87 of 101 comparison data values, consisting of literature data and some additional Huber and Herzberg values, fall within the prediction confidence limits or fall outside by less than half the prediction values, and 44 of the 101 values agree (within the limits).

## 1. INTRODUCTION

This atlas of diatomic-molecular vibration frequencies ($\omega_e$) and the preceding Atlas of Internuclear Separations[1] seek to provide forecasts that should be useful, for instance, in studies of species found in stellar atmospheres. Studies of spectroscopic data, even of moderate accuracy, provide tests of models of stellar interiors and nucleogenesis.

We have formed the outer matrix product of the chart of the elements[2,3] to construct a general periodic system of all diatomic molecules and have used Diophantine algebra to determine all possible diatomic[4] (and larger[5]) main-group molecules whose atoms possess closed valence shells. These procedures are tantamount to assuming that all diatomics, and especially those with closed valence-electron shells, will be found to exist under standard or, more likely, *other* conditions as stable molecules. This assumption entails a responsibility to use whatever means available to estimate the properties of as many of these species as possible. In this paper, we mine the high-quality tabulated data of Huber and Herzberg[6] to find trends that are then used to forecast data for $\omega_e$. Neural networks (NNs) have trained on these data once before.[7,8] We suspect[1] that the input vectors (independent variables plus data) were memorized to the point that no predictions could be made, and in fact, none were presented. Here, we present $\omega_e$ for 2129 molecules with a precision (average fractional 1% confidence limit, CL) of 10.66%. The accuracies are tested by comparison with data from the literature.

Boldyrev and Simons[9] and Krasnov[10] have also assembled data, though not so many. The former used quantum computation and experimentation and the latter depended on experimentation.

## 2. THEORY

Neural networks have been used to forecast properties of carcinogenic,[11,12] toxic,[13] and pharmaceutical[14] species; topological indices were used as independent variables (inputs). Since the atomic topologies of all diatomic molecules are identical, standard topological indices cannot be used as inputs. Even if there were one, then the problem would be that only carbon atoms could be treated. The index would have to be redesigned to take other atoms into account, perhaps as Balaban's $J_{het}$ topological index[15,16] does—including them in a way strongly related to their periodicity—but his parameter that does so is not available for all atoms. For these reasons, we begin by selecting as inputs the coordinates of the molecules in the periodic system of diatomic molecules—the period, or row number ($R$), and group, or column number ($C$), of each atom's location in the chart of the elements. The squares of these four descriptors are also selected as inputs.

As additional inputs, we include the positive differences of the electronegativities, their cubes, and the geometric means of the two electronegativies (en). The en's are included (heuristically) to enhance training of the networks in portions of the vector space where the group numbers are very different and the bonds have a more ionic character; the cubes exaggerate differences to be seen as the bonding moves from covalent to ionic. For these additional inputs, as for the squares of row and column numbers, the primary justification can only be that significant improvements of predictions were achieved; working with these neural networks has proven to be as much an art as a science.

* Corresponding author phone: (423) 238-2869; fax: (423) 238-2349; e-mail: hefferln@southern.edu.
[†] Davis Research Group.
[‡] Southern Adventist University.

10.1021/ci050439g CCC: $33.50 
Published on Web 02/01/2006

**Table 1.** Characteristics of the Data Sets

| | training set | compensated training set | validation set | prediction set after 16 species removed |
|---|---|---|---|---|
| count[a] | 204 | 1550 | 20 | 2129 |
| low value | 42.02 | 42.02 | 75.5 | 42.18 |
| high value | 2358.60 | 2358.60 | 1173.33 | 2153.92 |
| mean | 523.29 | 517.34 | 431.09 | 310.48 |
| $\sigma$ | 418.16 | 513.42 | 282.04 | 237.49 |
| median | 377.00 | 333.00 | 369.74 | 226.43 |
| average 1% limit[b] | | | | 10.66% |

[a] All entries except those in the first and last rows are in $cm^{-1}$. [b] After all 1% confidence limits less than 9.04% have been replaced by 9.04%.

## 3. DATA FOR TRAINING, VALIDATION, AND COMPARISON

Data for rows 2−6, groups 1−17 diatomics are found in ref 6. After the exclusion of alkaline earth pairs, $Hg_2$, molecules containing inert-gas atoms and lanthanoids, and data flagged with parentheses and square brackets, 224 vectors remained (parentheses indicate uncertainty in the data; square brackets indicate uncertainty in the state to which the data pertain). They were partitioned by us into a learning set and a validation set. The validation set is extracted from the known (tabulated) data and is not included in the NN training, so as to have data that give a measure of how well the NN is doing at prediction. There is some argument in the literature as to how large the data set should be. We choose to keep out approximately 10%, or 20 vectors. A total of 11 (AlO, PS, SiSe, TlF, SrS, SnSe, TlCl, TeAu, BiBr, BiI, and BaI) were taken from the more populated portions of the space, in order to not further deplete the sparsely populated portions and in order to have validation-set exemplars that would be well-predicted. Another nine (SnCl, SbS, SbF, PBr, KRb, $Cu_2$, CdF, AlN, and $Ag_2$) were chosen specifically because they had been forecasted poorly in preliminary trials. Because their data were flagged, 55 molecules from Huber and Herzberg were not used for the training of the neural network model. These molecules form part of the comparison set, which is quite apart from the validation set of 20 molecules, and play their role as described in section 5. The various data sets are described in Table 1.

The training and validation set molecules are entered in AB order, where A is the atom with the smaller Pauling en. If the en's are equal, then A is the atom with the higher period number *R*. In early trials, we had symmetrized the data as in ref 1, but this procedure is here reconsidered, for the following reason: the effect of the symmetrization is that for almost all points in the multidimensional space (i.e., the locations of heteronuclear molecules) there are two different sets of coordinates associated with exactly the same datum. Given that there are relatively few nodes with connections to adjust, this makes the problem of finding a solution in the space much more difficult and time-consuming—a crucial factor for $\omega_e$, which has a far larger range of magnitudes than the internuclear separations.[17]

The training data were sorted in order of increasing magnitude. Their numbers per unit vibration frequency increase (with negative curvature) from low values of $\omega_e$ to a long, rather linear, region and then move up (with positive curvature) to high-value $\omega_e$ values—very much as shown in Figure 2 of ref 1. To eliminate poor performance at the low and high ends due to the deficits, 1346 duplicate values of the vectors were introduced in the following way to form a larger training set of 1550 vectors:

1. Order the data from the smallest magnitude to the largest.

2. Find the magnitude difference between each record and the next. These differences for the small magnitudes are on the order of 10; they drop fairly quickly down to about 1; then, for the large magnitudes, they increase again (rapidly).

3. Form a lagged average (5−10 magnitudes at a time) of differences to see if there are domains where the magnitudes increase smoothly. The reason for the lagged average is that in places it is difficult to see smooth increases on a one-datum to one-datum basis.

4. Estimate how many additional records are needed to smooth the lagged averages. This initial estimate, of about 1300, determined some parameters in the following steps.

5. Split the needed additional records between the lower datum and the adjacent higher datum of the domain. If there are several data in the domain, distribute the needed additional records among the data including the lower border datum and the higher border datum of the domain. Thus, if there are 10 additions, add five more to the smaller datum and five more to the larger one (or distribute the 10 among all the data in the domain). Then, go on to the next domain, which involves as its smaller datum what was the larger datum of the previous domain. If the needed additional records now number seven (for example), add three more of what is now the smaller datum and four more to what is now the larger datum (or distribute the seven among all the data in the domain).

6. Difficulties occur in continuing this process for the differences at the largest magnitudes, because the vibration frequency magnitude differences are on the order of several hundred. Limit the increases to 30 or 50 records, and run the neural-network program to see if the NN is underestimating high-magnitude vibration frequencies. If indeed the NN is underestimating them, then increase the numbers of added records by relatively large increments.

7. When successive models indicate that high-magnitude vibration frequencies are being fairly well estimated as a whole, then select an arbitrary percent error (20% to begin with) by which to identify molecules for which the NN failed to predict well. Add records for those molecules that are not being estimated well; the number added is based on how far off the estimates are. The number added is two (adding two more records) per 5% error if they are above the criterion error (20% to begin with). This implies, for example, that if a given record's representation has already been increased by say three in step 6, and the estimate for that molecule is off by 25%, then it is necessary to add 10 more records (making 13 additions altogether).

8. (An adjustment to step 7) If the molecule is in the linear domain, change "two (adding two more records) per 5% error" to "one (adding one more record) per 5% error".

9. (Adjustment to step 7). If the estimate has a very large error, change "two (adding two more records) per 5% error" to "one (adding one more record) per 5% error", knowing that the same molecule will probably come up again as poorly forecasted.

**Table 2.** Excerpt from the Atlas of Predicted Vibration Frequencies for Diatomic Molecules ($cm^{-1}$) and Associated Statistical Measures

| molecule | tabulated data | networks | predicted value | standard deviation | standard error | 1% confidence limit[a] |
|---|---|---|---|---|---|---|
| AgAg | 192.4 | 21 | 169.91 | 13.11 | 2.86 | 9.04 |
| AgAs | | 22 | 325.00 | 64.43 | 13.74 | 10.89 |
| AgAt | | 20 | 149.80 | 28.00 | 6.26 | 10.77 |
| AgAu | | 22 | 185.45 | 19.68 | 4.19 | 9.04 |
| AgB | | 22 | 508.82 | 169.14 | 36.06 | 8.25 |
| AgBi | | 20 | 180.09 | 23.23 | 5.19 | 9.04 |
| AgBr | 247.7 | 21 | 236.74 | 10.19 | 2.22 | 9.04 |

[a] All values less than 9.04% have been replaced by 9.04%.

10. For larger percent errors, limit that record's representation to no more than 30 beyond the increase that it already has after step 6. If, as a hypothetical case, a large vibration frequency has already been duplicated 300 times in step 6 and is off by 100%, then it would be necessary to add 20 more records (one addition for each 5%). Such an increase is too large and is lowered.

11. As the forecasts improve, gradually lower the criterion error in step 7 to 10%.

## 4. CREATION OF NETWORKS AND PRESENTATION OF GLOBAL PREDICTIONS

We used software (Neuralwork's PREDICT) that was originally designed for forecasting financial markets. Many of the procedures used in the operation of PREDICT are described here; additional details can be found in refs 18 (in which criteria are devised for what could become parallel computation for mining molecular properties) and 19. A 3 GHz Pentium IV machine processes a network in 1−3 days. On average, one in three is acceptable, as shown by comparing the predictions made for validation-set molecules against the tabulated data.

A total of 11 inputs were identified in section 2; however, of these, only five or six are truly independent (depending on whether the difference of en is considered independent of the geometric mean of the en) and the rest of them are transformations of the five or six. PREDICT constructs additional transformations of the 11 inputs on the basis of correlational analyses. More than 11 inputs were included in early trials; however, the sensitivity analysis of PREDICT (the numerical partial derivatives with respect to each input) showed that the program systematically marginalized the additional inputs, and so they were eliminated.

We specify that at most six transfer functions, chosen from tanh (represented three times) and sinh, Gaussian, and sigmoid (represented once each), shall be used in building a given node. Quite often, PREDICT chooses tanh six times; otherwise, it tries different sets of six (or fewer) transfer functions, using genetic algorithms, to see which set relates most closely to the original inputs and to the dependent variable.[18] Instead of creating discrete layers, PREDICT adds and trains one, two, three, or four additional nodes, at each node, in each cycle. Each time the program starts a new attempt at training a given node, or moves on to another node, it cycles through the transfer functions listed above.

The effect of adding and training multiple nodes is the more effective handling of the problem of higher dimensionality; the procedure has the same effect as adding multiple layers, but the incremental control is finer. By contrast, another program called Braincell adds multiple layers instead of additional nodes in one layer, as it trains. PREDICT, even when it was prevented from using the genetic algorithms, made much better forecasts than those of Braincell.

**Table 3.** Amount, in Percent of Huber and Herzberg Data, by Which Some Predictions and Their 1% Confidence Intervals Differ from the Tabulated Data[a]

| species | gap $\delta$, percent | species | gap $\delta$, percent | species | gap $\delta$, percent |
|---|---|---|---|---|---|
| SiSe | 0.323 | PbCl | 1.134 | GaBr | 4.675 |
| RbCl | 0.379 | GaI | 1.455 | RbF | 5.175 |
| $N_2$ | 0.462 | RbBr | 1.482 | BiI | 6.106 |
| CuS | 0.539 | SnTe | 1.931 | CsF | 6.244 |
| CsCl | 0.651 | TeS | 2.431 | TeO | 7.948 |
| GeTe | 0.728 | SiO | 2.624 | KSe | 14.062 |
| BrO | 0.778 | $P_2$ | 2.894 | BaI | 15.684 |
| GaO | 0.894 | SbTe | 3.003 | TeSe | 21.056 |
| TeAu | 0.910 | SbO | 3.050 | | |
| GaCl | 1.020 | $Ag_2$ | 4.194 | | |

[a] There are 181 more predictions that do not differ at all.

The noise level is set to "moderate"; Kalman filtering is used in place of back-propagation, and cube-root mean cube is used instead of correlation for error evaluation.

The results obtained from 23 networks are summarized statistically in Table 1. The Atlas itself (next paragraph) originates from 2145 predictions, that is, the nonredundant members of [(7 groups)(periods 2 and 3) + (17 groups)-(periods 3−6)] squared. The averages of the standard errors of the predictions, of their 1% confidence limits, and of their 1% confidence limits expressed as a fraction of the mean are 10.32 $cm^{-1}$, 26.58 $cm^{-1}$, and 9.04%.

A total of 15 nonredundant alkali-earth pairs, having $(C_1,C_2) = (2,2)$, and $Hg_2$ were then culled out, leaving 2129 entries. The results for these molecules are presented as in Table 2, which is an excerpt of the Atlas, Supporting Information. The molecules appear alphabetically in column 1, just as in ref 1 (e.g., BaRh, BAs, BaSe) except that homonuclear molecules contain both atomic names, as in AgAg. Atom A in molecule AB is the one with the smaller Pauling en or with the greater period number if the en's are equal. Data in the training and validation sets are given in column 2. The next columns show how in many of the 23 networks the particular molecular predictions appear, the mean of the predictions and various precision measures. In our estimation, any fractional 1% CL less than 9.04% should considered as equal to 9.04% and is so designated in the Atlas.

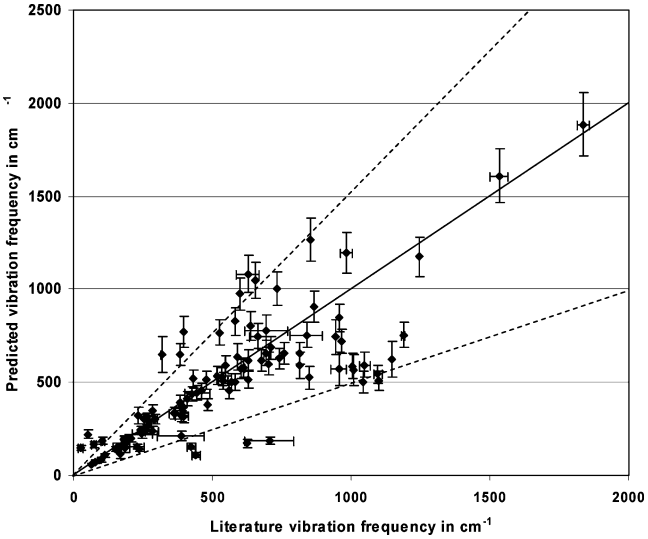


**Figure 1.** Predicted $\omega_e$ values with 1% confidence limits compared with literature values not used in training or validation. The solid line has a unit slope, and the dashed lines have greater and lesser slopes by 50%.

## 5. EVALUATIONS OF ACCURACY

The 1% confidence limits of the predictions nicely straddle 196 of the 224 tabulated data used for training and validation. The 28 predictions that do not are shown in Table 3, which shows that only three of these 28 predictions and their CL differ from the tabulated data by a gap, $\delta$, of more than 10% of the tabulated value. Extensive study has shown that the species with the three highest gaps are, compared to other molecules, not more lacking neighbors in the atom−atom space (in terms of Euclidean distances). Transition-metal molecules are under-represented, there being only three, in Table 3. In sum, an explanation is yet to be found.

There were 55 molecules from Huber and Herzberg not used for training (section 2). There were 48 molecules for which over 200 data values were gleaned (with the help of Davis and Eaken[20]) from 25 articles in the Journal of Chemical Physics (JCP), all but one dated from 1999 to 2002. In all, 101 molecules are included in the comparison set. When two or more data values exist for a molecule, they are averaged, and the highest and lowest values (after the elimination of a very apparent outlier in a couple of cases) are subtracted and divided by two to obtain the "half-spread" (HS). For 43 of the 101 molecules, their CL and the HS (if any) of their literature values overlap nicely; for 44 more (making 87 in all), there are gaps between the CL and the HS of their literature values that are less than half the forecasted values (Figure 1); all of these 101 species are shown in Table 4. For two of the molecules, agreement would not be expected: $Mg_2$ is an alkaline-earth dimer, and $Hg_2$ also has a very low $\omega_e$. (This test was done before the alkaline-earth pairs and $Hg_2$ were deleted from the predictions.) Of the 58 data values in Table 4 that are not in total agreement with the NN predictions, over half contain at least one transition-metal atom. All of those with a gap of 50% or more have at least one transition-metal atom. This situation reflects then-current research interests reported in JCP but

**Table 4.** Comparison Molecules and Gap between Predicted and Literature Values' Error Bars

| species | gap $\delta$ | ref | species | gap $\delta$ | ref | species | gap $\delta$ | ref |
|---|---|---|---|---|---|---|---|---|
| $Ag_2$ | *a* | 21 | $Hg_2$ | 66.38% | 6, 24, 30 | RbLi | 4.03% | 6 |
| AgSe | 13.74% | 6 | InO | 6.99% | 6 | RbO | 7.46% | 6 |
| AlAg | 6.33% | 6 | IrC | 97.91% | 6 | RhC | 106.06% | 32 |
| AlB | 29.44% | 22 | KLi | | 6 | SbF | | 6 |
| AlBe | 37.96% | 22 | KO | 31.93% | 6 | SbI | | 6 |
| AlC | 28.15% | 22 | KRb | | 6 | SbN | 14.10% | 6 |
| AlLi | 36.10% | 22 | LaF | 4.93% | 6 | SbS | | 6 |
| AlN | 23.71% | 6, 22 | LaO | 25.84% | 6 | $Sc_2$ | 60.63% | 23 |
| AlO | 17.89% | 22 | LiO | 23.41% | 6 | ScO | 25.34% | 6 |
| $Au_2$ | *a* | 21, 24, 25 | $Mg_2$ | 67.83% | 6 | SeF | 6.27% | 6 |
| AuBr | | 25 | MgCl | | 6 | SeN | 3.98% | 6 |
| AuF | | 25 | MgF | | 6 | SiAg | | 36 |
| AuO | | 25 | $Mn_2$ | 225.31% | 23 | SiAu | | 36 |
| AuGe | | 6 | MnCl | | 6 | SiC | 8.86% | 6 |
| AuSe | | 6 | MnI | | 6 | SiCu | | 36, 37 |
| BeF | | 6 | MnO | | 31 | SnF | 20.90% | 6 |
| $Bi_2$ | | 21 | MoC | 64.25% | 32 | SiPt | 17.00% | 38 |
| BiSe | 4.30% | 6 | NI | | 6 | SnO | 15.44% | 6 |
| $C_2$ | | 27 | NaO | 22.38% | 6 | TaO | 62.28% | 5, 34 |
| CCl | | 6 | $Nb_2$ | 282.38% | 33 | TeCl | | 6 |
| CdF | | 6 | NbN | 59.68% | 6 | TeF | | 39 |
| ClBr | | 26 | NbO | 90.09% | 34 | $Ti_2$ | 149.23% | 23, 40 |
| CaCl | 1.85% | 6 | $Ni_2$ | | 23 | TiS | 13.78% | 6 |
| CaF | 7.29% | 6 | NiF | 8.87% | 6 | $Tl_2$ | 40.88% | 21, 41 |
| $Co_2$ | | 23 | NiO | | 6, 34 | TlAt | | 42 |
| CsK | | 6 | $O_2$ | | 27 | TlI | | 6 |
| CsLi | 19.02% | 6 | OsN | 68.48% | 28 | $V_2$ | 251.70% | 23 |
| CsNa | 2.88% | 6 | PBr | | 6 | VO | 47.85% | 34 |
| $Cu_2$ | | 21, 23 | $Pb_2$ | 31.07% | 21 | WC | 45.89% | 43, 44 |
| CuO | 11.23% | 6 | PbF | | 6 | YF | 13.19% | 6 |
| $Fe_2$ | 29.17% | 23 | PdC | 49.74% | 32 | YS | 18.30% | 6 |
| FeCl | | 6 | PdO | | 34 | ZrCl | 4.99% | 45 |
| FeF | 0.78% | 6 | $Pt_2$ | 21.93% | 21, 24, 35 | ZnF | | 6 |
| $Ge_2$ | 8.60% | 29 | PtO | | 34 | | | |

[a] The absence of an entry indicates agreement with a gap of zero (an overlap of error bars).

also means that, if the molecules selected from the journals had been a random distribution, then the failures in Table 4 would have been fewer and smaller.

## DISCUSSION

The 2145 preliminary results of this work are rooted in 224 tabulated data values. For triatomic molecules, of which there are 22 050 even if we restrict ourselves to those formed only of main-group atoms in periods 2−6 and groups 1, 2, and 13−17, we find 115 enthalpies of atomization in one database, 117 in another, and 93 ionization potentials—and a disproportionate number of these data values relate to $AB_2$ molecules. It seems clear that the method employed here cannot be globally employed to larger molecules. Selected classes of organic molecules have, on the other hand, often been studied by means of neural networks.[11−14]

## ACKNOWLEDGMENT

Darwin Clark, Dunghoon Chun, Emil Dameff, David Gimbel, Chip Hicks IV, Shawn Klein, Ken Parker, Rosalie Parrish, Ken Priddy, Phil Regas, Kevin Shaw, Mike Seaman, and Vincent Tan began this work by plotting many scores of provocative graphs like Figures 5-16 to 5-18 of ref 3 (which culminated in Figure 1 of ref 17) and 6-16 to 6-18. Jonathan Knoll and Jason Ileto played an active role in the analyses of the NN forecasts. All were undergraduate students at the University of Denver or at SAU.

**Supporting Information Available:** Predictions of neutral diatomic molecule ground-state vibration frequencies (Table 2). This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES AND NOTES

(1) Hefferlin, R.; Davis, W. B.; Ileto, J. An Atlas of Forecasted Molecular Data. 1. Internuclear Separations of Main-Group and Transition-Metal Neutral Gas-Phase Diatomic Molecules in the Ground State. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 622−628.

(2) Hefferlin, R. A.; Zhuvikin, G. V.; Caviness, K. E.; Duerksen, P. J. Periodic System of N-atom Molecules. *J. Quant. Spectrosc. Radiat. Transfer* **1984**, *32*, 257−268.

(3) Hefferlin, R. *Periodic Systems and their Relation to the Systematic Analysis of Molecular Data*; Edwin Mellen Press: Lewiston, New York, 1989; pp 257−277.

(4) James, B.; Caviness, K. E.; Geach, J.; Walters, C. J.; Hefferlin, R. Global Molecular Identification from Graphs. Neutral and Ioanized Main-Group Diatomic Molecules. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1−7.

(5) Walters, C. J.; Caviness, K.; Hefferlin, R. Global Identification from Graphs. IV. Molecules with Four Closed p-Shell Atoms and Beyond. *Croat. Chem. Acta* **2004**, *77*, 65−71.

(6) Huber, K. P.; Herzberg, G. *Constants of Diatomic Molecules*; Van Nostrand Reinhold: New York, 1979. Data for individual molecules are available on-line, courtesy of the National Institute of Standards and Technology, by going to http://www.webbook.nist.gov/chemistry/form-ser.html and selecting "Chemical formula," "Constants of Diatomic Molecules," and "submit."

(7) Cundari, T. R.; Moody, E. W. A Comparison of Neural Networks versus Quantum Mechanics for Inorganic Systems. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 871−875.

(8) Cundari, T. R.; Moody, E. W. Prediction of Bond Dissociation Energies using Neural Network, Statistical, and Quantum Mechanical Approaches. *THEOCHEM* **1998**, *425*, 43−50.

(9) Boldyrev, A. I.; Simons, J. *Periodic Table of Diatomic Molecules*; John Wiley: New York, 1997; Parts A, B, and C.

(10) Krasnov, K. S. *Molekulyarnye postoynaye neorganicheskikh soedinenij*; Khimia: Leningrad, 1979. An English translation of an earlier edition: Schmorak, J. *Handbook of Molecular Constants of Inorganic Compounds*; Israel Program for Scientific Translations: Jerusalem, 1970.

(11) Vračko, M. A Study of Structure−Carcinogenic Potency Relationship with Artificial Neural Networks. The Using of Descriptors Related to Geometrical and Electronic Structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1037−1043.

(12) Vračko, M. A Study of Structure−Carcinogenic Relationship for 86 Compounds from NTP Database using Topological Indices as Descriptors. *SAR QSAR Environ. Res.* **2000**, *11*, 103−115.

(13) Basak, S. C.; Grunwald, G. D.; Gute, B. D.; Balasubramanian, K.; Opitz, D. Use of Statistical and Neural Net Approaches in Predicting Toxicity of Chemicals. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 885−890.

(14) Xing, L.; Glen, R. Novel Methods for the Prediction of logP, p$K_a$, and logD. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 796−805.

(15) Balaban, A. T. Chemical graphs. XVIII. Graphs of degrees four or less, isomers of annulenes, and nomenclature of bridged polycyclic structures. *Rev. Roum. Chim.* **1973**, *18*, 635−653.

(16) Balaban, A. T. Chemical Graphs, Part 48, Topological Index J for Heterotaom-Containing Molecules Taking into Account the Periodicities of Element Properties. *MATCH* **1986**, *21*, 115−122.

(17) Latysheva, V. A.; Hefferlin, R. Periodic Systems of Molecules as Elements of Shchukarev's "Supermatrix," i.e. the Chemical Element Periodic System. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1202−1209.

(18) Davis, W. B.; Hefferlin, R. Neural Networks in Mining Molecular Properties from Tabulated Data. *WSEAS Trans. Inf. Sci. Appl.* **2005**, *2*, 951−953.

(19) Rumelhart, D. E.; McClelland, J. L.; PDP Research Group. *Parallel Distributed Processing: Exploitation in the Microstructure of Cognition*; MIT Press: Campbrige, Massachussetts, 1986; Vol. 1.

(20) Davis, S. P.; Eakin, D. M. *The Berkeley Newsletter of Molecular Spectra*; Department of Physics, University of California: Berkeley, CA 94720-7300.

(21) Varga, S.; Fricke, B.; Nakamatsu, H.; Mukoyama, T.; Anton, J.; Geschke, D.; Heitmann, A.; Engel, E.; Bastug, T. Four-component relativistic density functional calculations of heavy diatomic molecules. *J. Chem. Phys.* **2000**, *112*, 3499−3506.

(22) Gutsev, G. L.; Jena, P. Structure and stability of the AlX and $AlX_2$ species. *J. Chem. Phys.* **1999**, *110*, 2928−2935.

(23) Barden, C. J.; Rienstra-Kiracofe, J. C.; Schaefer, H. F., III. Homonuclear 3d transition-metal diatomics: A systematic density functional theory study. *J. Chem. Phys.* **2000**, *113*, 690−700.

(24) Motegi, K.; Nakajima, T.; Hirao, K.; Seijo, L. The ab initio model potential method with the spin-free relativistic scheme by eliminating small components Hamiltonian. *J. Chem. Phys.* **2001**, *114*, 6000−6006.

(25) Liu, W.; van Wullen, C. Spectroscopic constants of gold and eka-gold (element 111) diatomic compounds: the importance of spin−orbit coupling. *J. Chem. Phys.* **1999**, *110*, 3730−3735.

(26) de Jong, W. A.; Styszynski, J.; Visscher, L.; Nieuwpoort, W. C. Relativistic and correlation effects on molecular properties: The interhalogens ClF, BrF, BrCl, IF, ICl, and IBr. *J. Chem. Phys.* **1988**, *108*, 5177−5184.

(27) Sordo, J. A. Performance of CCSDT for first row AB/AB-diatomics: Dissociation energies and electron affinities. *J. Chem. Phys.* **2001**, *114*, 1974−1980.

(28) Ram, R. S.; Lievin, J.; Bernath, P. F. Fourier transform emission spectroscopy and ab initio calculations on OsN. *J. Chem. Phys.* **1999**, *111*, 3449−3456.

(29) Jo, C.; Lee, K. Semiempirical tight binding method study of small Ge and Sn clusters. *J. Chem. Phys.* **2000**, *113*, 7268−7272.

(30) Munro, L. J.; Johnson, J. K.; Jordan, K. D. An interatomic potential for mercury dimer. *J. Chem. Phys.* **2001**, *114*, 5545−5551.

(31) Gutsev, G. L.; Rao, B. K.; Jena, P.; Li, X.; Wang, L.-S. Experimental and theoretical study of the photoelectron spectra of $MnO_x^-$ ($x$=1−3) clusters. *J. Chem. Phys.* **2000**, *113*, 1473−1483.

(32) DaBell, R. S.; Meyer, R. G.; Morse, M. D. Electronic structure of the 4d transition metal carbides: Dispersed fluorescence spectroscopy of MoC, RuC, and PdC. *J. Chem. Phys.* **2001**, *114*, 2938−2954.

(33) Balasubramanian, K.; Xiao, L. Z. Spectroscopic constants and potential energy curves of $Nb_2$ and $Nb_2^+$. *J. Chem. Phys.* **2001**, *114*, 10375−10388.

(34) Rakowitz, F.; Marian, C. M.; Seijo, L.; Wahlgren, U. Spin-free relativistic no-pair ab initio core model potential and valence basis sets for the transition metal elements Sc to Hg. Part I. *J. Chem. Phys.* **1999**, *110*, 3678−3686.

(35) Airola, M. B.; Morse, M. D. Rotationally resolved spectroscopy of $Pt_2$. *J. Chem. Phys.* **2002**, *116*, 1313−1317.

(36) Turski, P.; Barysz, M. Electronic states of the copper, silver, and gold silicides and their ions. *J. Chem. Phys.* **2000**, *113*, 4654−4661.

(37) Turski, P.; Barysz, M. Electronic states of the copper silicide and its ions. *J. Chem. Phys.* **1999**, *111*, 2973−2977.
(38) Shao, L.; Sickafoose, S. M.; Langenberg, J. D.; Brugh, D. J.; Morse, M. D. Resonant two-photon ionization spectroscopy of jet-cooled PtSi. *J. Chem. Phys.* **2000**, *112*, 4118−4123.
(39) Rai, V.; Liebermann, H.-P.; Alekseyev, A. B.; Buenker, R. J. Electronic states and transitions of tellurium fluoride. *J. Chem. Phys.* **2001**, *114*, 8386−8394.
(40) Wei, S. H.; Zeng, Z.; You, J. Q.; Yan, X. H.; Gong, X. G. A density-functional study of small titanium clusters. *J. Chem. Phys.* **2000**, *113*, 11127−11133.
(41) Han, Y.-K.; Hirao, K. On the ground-state spectroscopic constants of $Tl_2$. *J. Chem. Phys.* **2000**, *112*, 9353−9355.
(42) Faegri, K. K., Jr.; Saue, T. Diatomic molecules between very heavy elements of group 13 and group 17: A study of relativistic effects on bonding. *J. Chem. Phys.* **2001**, *115*, 2556−2464.
(43) Balasubramanian, K. Spectroscopic constants and potential energy curves of tungsten carbide. *J. Chem. Phys.* **2000**, *112*, 7425−7436.
(44) Sickafoose, S. M.; Smith, A. W.; Morse, M. D. Optical spectroscopy of tungsten carbide (WC). *J. Chem. Phys.* **2002**, *116*, 993−1002.
(45) Ram, R. S.; Adam, A. G.; Sha, W.; Tsouli, A.; Lievin, J.; Bernath, P. F. The electronic structure of ZrCl. *J. Chem. Phys.* **2001**, *114*, 3977−3987.

CI050439G