Neural Network Correction of PM3-Predicted Infrared Spectra

Christopher J. Barden, Marc K. Boysworth, and Frank A. Palocsay*

Department of Chemistry, James Madison University, Harrisonburg, Virginia 22807

Received December 17, 1997

We describe the application of neural networks to a theoretical problem: the correction of inaccuracies in infrared spectra as predicted by the PM3 semiempirical method. Twenty-eight "peak-correcting" back-propagation neural networks were trained to predict the location of a characteristic infrared peak when given a scaled topological map of one of 1116 literature spectra. The infrared spectra of 200 aliphatics were then calculated using PM3, displayed graphically in Infrared Spectrum Comparison, and submitted to the appropriate "peak-correcting" neural network(s) based on a rule set implemented via an interface to HyperCube's HyperChem software. Results show an average 8-fold decrease in prediction error between PM3-predicted and neural network-corrected peak locations.

INTRODUCTION

Semiempirical Methods. Semiempirical methods have developed alongside the more rigorous ab initio methods as a fast alternative in calculating chemical properties. Ab initio methods are more theoretically sound, having no need for empirical parameters, but this accuracy is bought by costly computation that becomes impractical very quickly as the size of the molecular system increases. Current semiempirical methods (including MNDO, AM1, and PM3) are fast and have been shown to compare favorably with ab initio results using reasonably sized basis sets.[1] The balance of accuracy and speed provided by these methods makes them useful for calculation of heats of formation, geometries, and ultraviolet and infrared spectra, among other properties.

Nevertheless, accuracy remains a significant barrier to practical use of theoretical infrared spectra. According to a review by Seeger and co-workers, PM3 is the most accurate predictor of infrared spectra, but uncorrected errors are as high as 25% when the normal mode in question has a high degree of anharmonicity.[2] Often the specific chemical environment of the vibrational mode has a significant effect on peak location. These inaccuracies manifest themselves even in the case of ab initio studies at the Hartree−Fock limit.[3,4] A popular (albeit untheoretical) solution to such inaccuracies is a scaling process by which particularly bad normal modes may be brought closer to experiment. Early work by Dewar and Ford applied a systematic adjustment to specific normal modes calculated by the MINDO/3 semiempirical method.[5] Similar treatments have been applied to CNDO/2 by Panchenko and co-workers,[6] to 4−21 ab initio calculations using Pulay and co-workers' Scaled Quantum Mechanical (SQM) method,[7] and more recently to MNDO, AM1, and PM3 by Seeger[2] and by Fausto.[4] Alternatively, Schneider and co-workers performed their scaling to the force constants before calculation with similar results.[8] These techniques have been used with more or less complexity to achieve improved results, i.e., different sets of scaling factors for each kind of bond or for each bond in a certain chemical situation.

Clearly these methods have merit in reducing the often systematic-looking errors in theoretical infrared spectra, but for quantitative use these parameters are lacking: when the question is "what is the location of the carbonyl stretch in this unsaturated secondary amide?", a scaling parameter based on the weighted average of carbonyl stretches or even amides may be insufficiently specific. Regardless of how specific the scaling parameters are, they cannot provide quantitative chemical answers on a per compound basis unless the scaling is specific to that compound (and in such a situation, predicting the infrared spectrum is trivial). How, then, does one choose scaling parameters based on a particular compound without having parametrized specifically for that compound? The answer may be a heuristic approach.

Neural Networks. The use of neural networks in chemistry is now widespread and well-characterized.[9] The most frequently used type of neural network in chemistry is the back-propagation neural network. Such networks have been shown to have excellent classification capabilities, with numerous applications including spectrum-structure correlation,[10−12] baseline and analytical curve fitting,[13] and instrumental drift correction.[14] The network approach is particularly useful when the rules for classification or correction are poorly defined or the data is "noisy" (or both, in the case of semiempirical method inaccuracy).

In this study, neural networks have been utilized to apply a "variably scaled" correction to PM3. There is no reason, however, why this technique could not be used on AM1 or an earlier method like CNDO, as it does not depend on PM3 data during training but only on literature data.

EXPERIMENTAL SECTION

Network Architecture. In designing the architecture of a neural network, it is crucial that one take into account the topology of the problem. For this problem we decided to construct our neural networks to accept the infrared spectrum

* To whom correspondence should be addressed.

S0095-2338(97)00293-X CCC: $15.00
Published on Web 04/15/1998

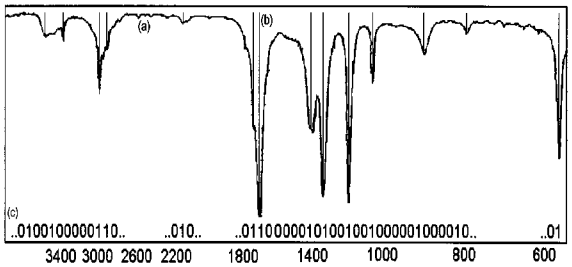

Figure 1. (a) A partial infrared spectrum of acetone. (b) The prominent features of the spectrum listed in the training set, denoted by solid lines. (c) The binary representation of the spectrum. Values between ellipses are all zero.

(approximately 400−4000 cm^{-1}) as input and to receive as output the correction for "theoretical instrumental drift". At first glance (and as implemented by other researchers in some simpler cases),[10,12] the problem suggested a three-layer back-propagation network with 3600 input nodes, an unknown number of hidden nodes, and as many output nodes as there were meaningful peaks in the spectrum. However, such an architecture is unsatisfactory for several reasons:

(1) Having such a large number of input nodes makes training and application of a network computationally impractical (especially important in this case, since speed is PM3's principal advantage over ab initio methods);

(2) Providing extremely high-resolution data to a network can result in erratic, ungeneralized predictions; and perhaps most importantly,

(3) It has been shown that a more specific network, with a more narrowly defined problem, performs significantly better than a more general network.[15,16]

These reasons suggested that building many small, specific "peak-correcting" networks would be better than building one large, general network.

To this end, all networks utilized a fully connected three-layer feed-forward back-propagation architecture with 254 input nodes, 76−650 hidden nodes (depending on the network), and one output node. The input and hidden layers each contained one bias node. The input nodes represented the section of the infrared from approximately 448 cm^{-1} to approximately 4011 cm^{-1}, which was divided into 254 regions in a manner analogous to that used by Munk and co-workers[12]

$$J = \text{int}(6 * (\nu)^{1/2} - 126)$$

where J is the truncated integer value of the region (1, 2, ..., 254) and ν is the wavenumber of vibration. This function gave each region a slightly wider range, starting in the fingerprint region at ±7.1 and ending with ±21.1 wavenumbers around the mid/near-IR boundary. All inputs were binary, with a value of one denoting a peak in that specific region and a value of zero denoting no peak (see Figure 1 for an illustration). These inputs were linearly scaled and transformed logistically by the hidden and output layers. The output node was an unscaled, continuous value representing the network's "best estimate" of the peak location in wavenumbers.

Training Database. A total of 1116 compounds was abstracted from the *Aldrich IR Library*,[17] a collection of 11 000 condensed-phase infrared spectra that is divided into aromatics and nonaromatics and further into functional groups. All compounds in this study were nonaromatics chosen to represent a wide range of compounds: as small as four atoms to as large as 100 atoms, as simple as a straight-chain alkane to as complex as a multifunctional biomolecule. Each compound's spectrum was reduced to a list of visually identifiable peak locations without intensity information. This list was then manually entered into a Microsoft Excel[18] spreadsheet and scaled via an Excel macro, using the nonlinear transformation described above. The data entry process was repeated many times, resulting in a data set of compounds containing a given functional group. Table 1 displays the composition of the carboxylic acid O−H stretch network as an example of a typical data set.

Table 1. Composition of the 126-Compound Training Set for the Carboxylic Acid OCO−H Stretch Network

property	no. of compds
size < 20 atoms	49
size 20−40 atoms	50
size > 40 atoms	27
n-monoacids	35
diacids	33
unsaturated	23
halogenated	20
cyclic	23
multifunctional	33

Training Method and Optimization. All networks were trained using NeuroShell 2 Release 3 neural network software from the Ward Systems Group, running on a PC with at least a Pentium 90 MHz. Approximately 80% of each data set was used as the training set. Ten percent of the data set (denoted "monitoring set") was used by NeuroShell during training to determine how well the network performed. The remaining 10% (denoted "test set" or "validation set" in the literature) was used after training to evaluate the network's generalization ability. Each network was trained using the proprietary TurboProp algorithm which makes uses of a training method similar to cascade correlation,[19] an algorithm discussed in detail by other workers.[20] Since Backprop and Cascade Correlation are not compatible methods, we think that TurboProp is using some variation of Fahlman's QuickProp algorithm. Quickprop is incorporated in Cascade Correlation and has been shown to out-perform other back prop algorithms.[21] Since our initial work suggested that networks trained with different random number seeds were not significantly different, the same random number seed was used to randomize weights at the start of training for each network. Thus, we may not have found the global minimum error for each network. Finding this minimum would not change but only improve our results. The training set was then evaluated by the network and network, weights were adjusted iteratively to improve performance. After an evaluation of a complete training set, overall network performance was assessed using the monitoring set, and the weights were saved to disk when the network performed better than previous networks on the monitoring set. The network training process was terminated after a large number of iterations with no improvement in network performance on the monitoring set. The weights from the best performing network were saved to a permanent Dynamic Linked Library (DLL), generating the permanent network architecture that was used to correct a particular functional group.

One of the difficulties in using neural networks is determining how many hidden nodes a network should have. The discussion about "rules of thumb" has produced several possible guidelines but nothing definitive.[15] During experiments with training these networks, it was noted that the sample size had some qualitative correlation to the number of hidden nodes. Unfortunately, on a quantitative level the situation is much more complex, as two networks differing in architecture by only a small number of hidden nodes can have extremely different performances. Our method for choosing the best number of hidden nodes involved a coarse exploration of the range 50 to 3**N* nodes where *N* is the number of compounds in the training set. Data were collected until a point of minimum average test set error was found.

During the training of these networks, we found, as did Munk and co-workers,[12] that in a training set of 50 compounds, there would always be two or three compounds with a significantly different infrared spectrum than the others. These compounds usually had structural characteristics such as a small size (methanol) or a multifunctional interaction (β-unsaturated ketones) that made them unique in the training set. Since the neural network is a generalizing tool, the best performance is generally from a typical compound in the set, while the worst performance is from an exception compound. By representing these exception compounds more often during training, we found that the accuracy of a prediction became more balanced across the range of peak locations. In our study we have attempted to maximize the network's performance using all compounds; but when two network architectures gave similar overall performances, the network with the best performance on typical compounds was chosen as the best architecture.

Computational Methods. All PM3 calculations were performed using HyperChem 5.01 for Windows from HyperCube.[22] A set of 200 molecules selected from the Aldrich IR library were drawn using HyperChem. The geometry of all molecules was optimized using the Polak-Ribiere conjugate gradient method with a root-mean-square convergence gradient of 0.01 kcal/(Å mol). Molecules that failed to converge satisfactorily were reoptimized after manually adjusting torsional angles to a strained position. All molecules were uncharged closed-shell species, calculated with a singlet multiplicity.

Implementation. Our interface for applying the "peak-correcting" neural networks was Infrared Spectrum Comparison (ISC),[23] shown in Figure 2. ISC is a Visual Basic program written by Robert H. Williams that utilizes Dynamic Data Exchange (DDE) to communicate directly with HyperChem and graphically display the results of a vibrational analysis. Originally designed to compare HyperChem results with experiment through import of MIDAC GRAMS/386 data files,[24] the program's functionality has been extended to interface with NeuroShell-designed networks via a DLL and display the corrected peak locations.

It is important to note that ISC must first determine which corrections are appropriate for the molecule. ISC does this by querying HyperChem as to the atomic "types" in a molecule, with the presence or absence of such "types" being the sole criterion on which applicability is based. A special set of type rules was compiled for this purpose using

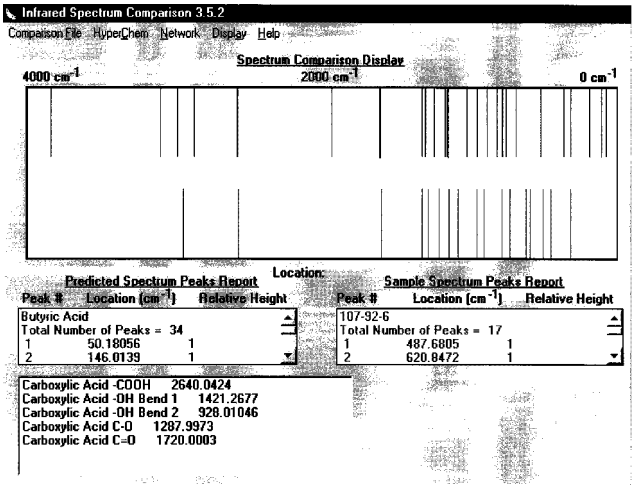

Figure 2. A screen shot of infrared spectrum comparison, showing the experimental spectrum on the bottom and the PM3-predicted, neural network corrected spectrum on the top.

HyperChem's flexible typing language. The following is an example of a typical entry in the file CHEM.RUL:

```
N: ; nitrogen

; primary amide

        connected to (-H)(-H)-C=O?

        =N86.   ; numerical type 86
```

This rule would specify any nitrogen atom connected to two hydrogens and a carbonyl group to be of type N86. A corresponding lookup for the atom type N86 would also be placed in MMPTYP.TXT, and the new rule set would be compiled under HyperChem.[25] Using this method, all functional groups studied were given a unique atomic type based on the atom in question (i.e., nitrogens for amines, oxygens for ethers), and Visual Basic code was written into ISC that would apply "peak-correcting" neural networks for every correctable peak found to exist in the molecule.

RESULTS AND DISCUSSION

Training Results. The results of neural network training are summarized in Table 2. Error is reported as the average deviation in wavenumbers from the experimental peak locations in the test data, excluding the 10% of compounds on range extremes which represent 50% or more of the error in some cases. It can be seen qualitatively from the data that the more characteristic the peak, the better the network performance (that is to say, the smaller the range of peak location, the more accurate the network).

The decision to split amides and amines into separate primary, secondary, and tertiary networks stems from the work done by ourselves and by others that suggests specific networks perform better than general networks.[15,16] Neural networks are excellent categorizing tools, and much work has been done on the design of neural networks that can interpret infrared spectra.[10,11,15,26−28] However, neural network categorization applications use binary outputs almost exclusively, while correction applications use mostly continuous outputs. The networks constructed in this study are indeed capable of recognizing two distinct patterns of

Table 2. Functional Group Neural Network Models

functional group	assigned mode	unique compds	hidden nodes	training set mean	training set SD[a]	av test error[b] in wavenumbers (see *) 90	av test error[b] in wavenumbers (see *) 100
alcohol	C−O−C st	90	650	1149	79	0.6	3.4
	O−H st	91	500	3378	86	32.0	48.7
aldehyde	OC−H st	44	100	2715	11	4.1	5.1
amide	pri C−N st	15	200	1418	17	17.6	17.6
	sec C−N st	28	220	1285	17	7.6	7.6
	ter C−N st	20	400	1237	29	18.1	18.1
	C═O st	68	500	1691	37	16.8	27.8
	pri/sec N−H def	40	600	1574	28	8.4	15.4
	pri/sec N−H st	36	510	3275	63	51.4	68.7
amine	pri N−H st	76	450	3327	37	10.9	20.5
	pri N−H_2 def	93	190	1604	13	5.5	14.9
	pri C−N st	94	130	1064	22	8.8	16.7
	sec N−H st	51	105	3298	49	24.1	37.2
	sec C−N st	56	100	1125	22	13.0	14.1
	ter C−N st	85	115	1158	23	10.0	12.9
bromocarbon	C−Br st	39	130	540	38	12.9	19.8
carboxylic acid	C═O st	136	125	1716	29	11.3	15.6
	OCO−H st	126	130	2662	31	24.9	30.4
	C−O st	135	76	1285	16	11.5	15.3
	C−OH def	136	375	922	18	9.7	10.9
chlorocarbon	C−Cl st	26	142	622	37	15.0	34.4
ester	C═O st	120	524	1737	21	24.0	32.1
	C−O−C asym st	114	475	1203	23	8.0	11.5
ether (dialkyl)	C−O−C st	61	100	1111	26	14.0	14.0
(vinyl)	C−O−C st	33	100	1243	23	23.2	23.2
ketone	C═O st	146	130	1721	24	7.7	11.2
vinyl	C═C st	66	101	1649	15	3.9	8.2
	CH_2 def	77	90	3048	32	10.3	12.5

[a] The mean and standard deviation for peak locations in a given training set, to the nearest wavenumber. [b] Average error in wavenumbers for the given training set, both with and without the 10% range extremes.

correction in the same network, as in a network containing exception compounds (see above), but this combined categorization and correction comes at the price of lower overall accuracy. For example, a network designed to correct both secondary and tertiary amide C−N stretches had an average test set error of 19 wavenumbers. Comparison to the separate networks in Table 3 shows that most of the error was introduced by the tertiary amide C−N stretch and that the secondary amide network performed significantly better by itself than in the combined secondary/tertiary network. In general, it has been our experience that a network trained to correct for two or three dissimilar types of compounds will invariably perform worse than any specific network derived from that set, even when there is an equal number of compounds from each category.

Application Results. The correction method described above was applied to 200 compounds, approximately half of which had never been evaluated by the networks. The results are summarized in Table 3 and Figure 3. Examination of the data shows a number of important results. As expected, if a given network performed well on test data during training, it tended to perform well when correcting actual PM3 data. The correlation between good network performance and a diagnostic peak appears to hold for actual PM3 application as well as for the test set, as noted in the performance of the ketone C═O stretch network when compared to the secondary amide C−O−C stretch network. On average, this method allows prediction of functional groups to within an error of 22 wavenumbers, which translates into an 8-fold increase in PM3 accuracy (error decreases from approximately 10.0% to approximately 1.25%). Such a performance is similar to the predictions made by quite sophisticated ab initio methods.[2] For comparison purposes, Table 4 displays the results of neural

Table 3. Summary of Neural Network Performance

network name	assigned mode	av error in wavenumbers* PM3 output	av error in wavenumbers* NN corrected output	accuracy increase (fold)
alcco	C−O st	123	24	5.2
alcoh	O−H st	550	31	17.6
ald	OC−H st	208	9	22.1
amd1cn	C−N st	70	18	3.9
amd2cn	C−N st	83	9	9.1
amd3cn	C−N st	52	24	2.2
amdcarb	C═O st	257	10	25.9
amdii	N−H def	108	29	3.7
amdnh	N−H st	70	32	2.2
amn1nh	N−H st	123	27	4.5
amn1nh2	N−H def	71	11	6.5
amn1cn	C−N st	148	21	7.0
amn2nh	N−H st	82	28	3.0
amn2cn	C−N st	58	27	2.2
amn3cn	C−N st	109	8	13.2
bromo	C−Br st	125	18	6.7
cacarb	C═O st	257	19	13.7
cacooh	OCO−H st	1207	22	55.7
caco	C−O st	17	36	0.5
caoh	O−H bend	87	15	5.8
carb	C═O st	265	11	24.0
chloro	C−Cl st	131	18	7.2
estcarb	C═O st	244	33	7.5
estco	C−O st	166	22	7.6
ethco1 (dialkyl)	C−O−C st	37	32	1.2
ethco2 (vinyl)	C−O−C st	136	6	23.9
vinylcc	C═C st	216	23	9.3
vinylch2	C−H def	34	49	0.7
overall averages		180	22	8.2

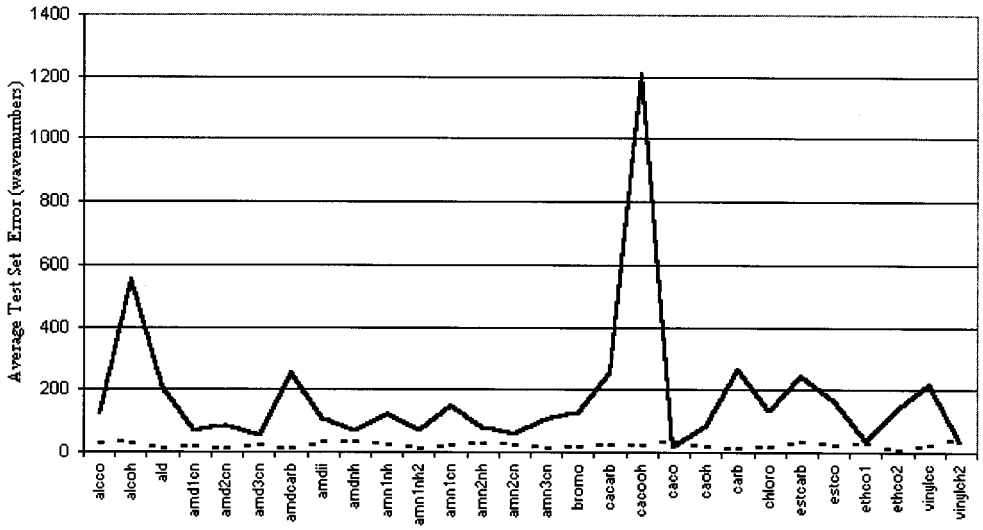

Figure 3. Graphical comparison of error between uncorrected PM3 (solid line) and neural network corrected PM3 (dotted line).

Table 4. Comparison of PM3 Neural Network Correction to the Pulay SQM Method and the 4−21 ab Initio Method[2,7]

compd	mode	SQM	NN	4−21	exp
acrolein	C═O st	1745	1715	1900	1723
	C═C st	1561	1660	1767	1625
butadiene	C═C sym st	1690	1651	1809	1643
	CH_2 asym st	3075	3029	3307	3102
ethylene	C═C st	1628	1644	1783	1630
formaldehyde	C═O st	1753	1727	1921	1746
glyoxal	C═O st	1761	1725	1934	1745

network corrections for the test compounds used in Seeger's application of the Scaled Quantum Mechanical method, along with unscaled 4−21 ab initio results.[2,7] It can be seen that while both scaling methods are significantly more accurate than unscaled ab initio results, the neural network correction method performs as well as or better than SQM with these molecules, with the exception of the vinyl CH_2 stretch in butadiene.

It is networks such as the vinyl CH_2 stretch network or the network trained to find the C−O stretch in carboxylic acids that represent the biggest disadvantage to this method: assignment error. In even the simpler molecules, there are normal modes of vibration that are similar in amplitude and position to the vibrational mode of interest. In the case of the vinyl CH_2 stretch, the general region of 3000−3100 wavenumbers can contain broad absorptions from other functional groups as well as related ═CH stretches. To make matters worse, the stretching manifests itself differently depending on its status as vinyl, vinylene, or vinylidene. The carboxylic acid C−O stretch region of 1200−1300 wavenumbers is riddled with absorptions from other nitrogen- and oxygen-containing functional groups and quickly becomes crowded as the molecule being studied becomes more complex.[28] In most situations, a neural network is robust to such "noise" being introduced to its training set, but in these cases it appears there was too much variation in peak location for the network to generalize acceptably. This result underscores the importance of accurate peak assignment, as the network cannot possibly "learn" how to properly correct a noisy spectrum if it was trained incorrectly.

It should be noted also that our method has limitations in that it always provides one peak for each functional group. A compound of high symmetry, such as methane, might not have a peak where it would customarily be found, while a multifunctional compound, such as an enol, could produce coupled peaks. However, as the data appears to show, generalization is quite good for a variety of functional groups in many different situations. One should be able to parametrize a set looking at specific peaks (syn and anti C−H stretches in vinylenes, for example) provided the peaks can be found in a number of compounds. Further functional group studies are ongoing.

A final word on the design should be mentioned. Data were collected using the method described above, with the training data being the *experimental spectra* along with the literature value of the peak in question. Early work on this problem was focused on designing a carbonyl network using the *theoretical PM3 spectra* as training data along with the literature value of the C═O stretch, which proved to be a considerably more time-consuming method than the one discussed here. Comparison of the early data with the data shown above for the carbonyl stretch shows no statistical difference in performance; in fact, through an accidental creation of a network containing both designs it was discovered that even this "mixed" design corrected acceptably! We believe this phenomenon to be largely due to the especially small variance of the carbonyl stretch (in our data, 24 wavenumbers), which approaches the theoretical limit of resolution of 16 wavenumbers in this region (denoted region 122 by Munk's data reduction equation). It may be possible to improve the performance of these neural networks even further by gradually increasing the resolution of the spectrum being input, keeping the aforementioned caution about complexity in mind.

Conclusions. A new method for correcting inaccuracies of PM3 has been demonstrated. Earlier methods of correcting infrared spectra generated by semiempirical algorithms were scaled based on the weighted average of an arbitrary basis set of test compounds. This new method, neural network correction, is heuristic in nature, which may be more properly suited to an empirical correction process. While there are certainly drawbacks to this method of correction, they are not insurmountable so long as the method is used with a thorough understanding of its strengths and weak-

nesses. The 28 networks described in this paper are a relatively useful tool for the spectra of common organic molecules, providing an average 8-fold increase in accuracy over PM3. Further work is ongoing, with the goal of creating a comprehensive set of networks for use in organic chemistry.

ACKNOWLEDGMENT

We would like to acknowledge Dr. Susan W. Palocsay for her insightful comments and suggestions regarding neural networks. Robert H. Williams made an important contribution to this research through his early work on Infrared Spectrum Comparison. The authors are indebted to Julia Drumheller and Dr. Benjamin A. DeGraff for their help in editing this paper and to SEIT for providing scholarship funds during a portion of this research.

Supporting Information Available: A list of the compounds in each training set, listed by CAS Registry Number, is available (19 pages). See any current masthead page for ordering and Web access instructions.

REFERENCES AND NOTES

(1) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods. *J. Comput. Chem.* **1989**, *10*, 209−220.
(2) Seeger, D. M.; Korzeniewski, C.; Kowalchyk, W. Evaluation of Vibrational Force Fields Derived by Using Semiempirical and Ab Initio Methods. *J. Phys. Chem.* **1991**, *95*, 6871−6879.
(3) Pulay, P.; Fogarasi, G.; Pang, F.; Boggs, J. E. Systematic ab initio Gradient Calculation of Molecular Geometries, Force Constants, and Dipole-moment Derivatives. *J. Am. Chem. Soc.* **1979**, *101*, 2550−2560.
(4) Fausto, R. Calculation of Vibrational Frequencies of Sulphur Compounds Using Semiempirical Molecular Orbital Methods (AM1, PM3, MNDO). *J. Mol. Struct.* **1994**, *323*, 267−277.
(5) Dewar, M. J. S.; Ford, G. P. MINDO/3 Calculations of Molecular Vibrational Frequencies. *J. Am. Chem. Soc.* **1977**, *99*, 1685−1691.
(6) Panchenko, Y. N.; Pulay, P.; Török, F. Prediction of Vibrational Spectra by the CNDO/2 Force Methodology. 2. Calculation of Vibrational Frequencies of cis and trans Forms of Glyoxal, Acrolein, and 1,3-butadiene. *J. Mol. Struct.* **1976**, *34*, 283−289.
(7) Pulay, P.; Fogarasi, G.; Pongor, G.; Boggs, J. E.; Vargha, A. Combination of Theoretical ab initio and Experimental Information to Obtain Reliable Harmonic Force Constants. Scaled Quantum Mechanical (SQM) Force-fields for Glyoxal, Acrolein, Butadiene, Formaldehyde, and Ethylene. *J. Am. Chem. Soc.* **1983**, *105*, 7037−7047.
(8) Schneider, S.; Freunscht, P.; Brehm, G. Normal Modes of *p*-Aminobenzonitrile and *p*-Dimethylaminobenzonitrile. Vibrational Spectra and PM3 Calculations with a Modified Scaling Procedure. *J. Raman Spectrosc.* **1997**, *28*, 305−314.
(9) An excellent review of neural networks and their applications to chemistry can be found in Jure Zupan and Johann Gasteiger's *Neural Networks for Chemists*; VCH Publishers: New York, 1993.
(10) Robb, E. W.; Madison, M. S.; Munk, M. E.; Neural Network Models for Infrared Spectrum Interpretation. *Mikrochim. Acta* **1991**, *265*, 183−190.
(11) Nović, M.; Zupan, J. Investigation of Infrared Spectra-Structure Correlations Using Kohonen and Counterpropagation Neural Networks. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 454−466.
(12) Munk, M. E.; Madison, M. S.; Robb, E. W. The Neural Network as a Tool for Multispectral Interpretation. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 231−238.
(13) Wythoff, B. J.; Levine, S. P.; Tomellini, S. A. Spectral Peak Verification and Recognition Using a Multilayered Neural Network. *Anal. Chem.* **1990**, *62*, 2702−2709.
(14) Goodacre, R.; Kell, D. B. Correction of Mass Spectral Drift Using Artificial Neural Networks. *Anal. Chem.* **1996**, *68*, 271−280.
(15) Klawun, C.; Wilkins, C. L. Optimization of Functional Group Predictions from Infrared Spectra Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 69−81.
(16) Lu, T.; Lerner, J. Spectroscopy and Hybrid Neural Network Analysis. *Proc. IEEE* **1996**, *84*, 895−904.
(17) *The Aldrich Library of FT-IR Spectra*, 1st ed.; Pouchert, C. J., Ed.; Aldrich Chemical Co.: Milwaukee, 1985.
(18) Excel is a registered trademark of Microsoft Corp., Redmond, Washington.
(19) NeuroShell is a registered trademark and Turboprop is a trademark of Ward Systems Group, Inc., Frederick, Maryland. All details of Turboprop given here are from the NeuroShell manual or from private communication.
(20) Fahlman, S. E.; Lebiere, C. *The Cascade-Correlation Learning Architecture*, Carnegie Mellon University Technical Report CMU-CS-90-100, 1991.
(21) Fahlman, S. E. Faster-Learning Variations on Back-Propogation: An Empirical Study; in Proceedings of the 1988 Connectionist Models Summer School; Morgan Kaufmann.
(22) HyperChem is a registered trademark of Hypercube, Inc., Waterloo, Ontario, Canada.
(23) Williams, R. H. *A Computer Interface to Predict and Compare Infrared Spectra*; Department of Chemistry, James Madison University; 1995.
(24) GRAMS/386 is a registered trademark of MIDAC Corporation, Irvine, CA.
(25) *HyperChem Release 5.0 for Windows: Reference Manual*; Hypercube: Ontario, 1996; pp 463−471.
(26) Fessenden, R. J.; Györgyi, L. Identifying Functional Groups in Infrared Spectra Using Neural Networks. *J. Chem. Soc., Perkin Trans.* **1991**, *2*, 1755−1762.
(27) Meyer, M.; Weigelt, T. Interpretation of Infrared Spectra by Artificial Neural Networks. *Anal. Chim. Acta* **1992**, *265*, 183−190.
(28) Ricard, D.; Cachet, C.; Cabrol-Bass, D.; Forrest, T. P. Neural Network Approach to Structural Feature Recognition from Infrared Spectra. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 202−210.
(29) *CRC Handbook of Chemistry and Physics*; Weast, R. C., Ed.; CRC Press: Cleveland, OH, 1972; Section F, pp 253−256.

CI970293O