Producing ^{13}C NMR, Infrared Absorption, and Electron Ionization Mass Spectrometric Data Models of the Monodechlorination of Chlorobenzenes, Chlorophenols, and Chloroanilines

Richard D. Beger,* James P. Freeman, Jackson O. Lay, Jr., Jon G. Wilkes, and Dwight W. Miller

Division of Chemistry, National Center for Toxicological Research, Food and Drug Administration, Jefferson, Arkansas 72079

Received May 12, 2000

We have developed four spectroscopic data−activity relationship (SDAR) models of monodechlorination of 32 chlorinated benzene compounds in anaerobic estuarine sediment. The SDAR models were based on combinations of ^{13}C nuclear magnetic resonance (NMR), infrared absorption (IR), and electron ionization mass spectrometric (EI MS) data. The SDAR models segregated the 32 compounds into 17 readily monodechlorinated compounds and 15 not readily monodechlorinated compounds. The SDAR model based on ^{13}C NMR, IR, and EI MS data gave a leave-one-out cross-validation of 93.8%. The SDAR model based on a composite of ^{13}C NMR and IR data gave a leave-one-out cross-validation of 90.6%. The SDAR model based on a composite of IR and EI MS data gave a leave-one-out cross-validation of 84.4%. The SDAR model based on a composite of ^{13}C NMR and EI MS data gave a leave-one-out cross-validation of 84.4%. These reliable SDAR models provide a rapid and simple way to predict whether a chlorinated benzene compound will readily go through monodechlorination. The FDA has filed a patent application on methods of using any combination of spectral data (NMR, MS, UV−vis, IR, and fluorescence, phosphorescence) to model a chemical, physical, or biological endpoint.

INTRODUCTION

Producing a quantitative structure−activity relationship (QSAR) model of a molecular process where the site of action is unknown or where multiple sites of action occur simultaneously can be very difficult to implement. The removal of one chlorine atom from 32 different chlorobenzenes, chlorophenols, and chloroanilines in anaerobic estuarine sediment is a reductive microbial transformation process where multiple sites of action can occur simultaneously, since all six positions on a benzene ring are potential sites of action for each chlorinated compound.[1−5] Monodechlorination is part of the reductive biodegradation process of chlorinated benzene compounds. To produce a model of mondechlorination, we looked to the theory of quantum mechanics. The problem is that quantum mechanics calculations are very hard to calculate accurately. To solve the quantum mechanical Schrödinger's equation involves many approximations to all the individual wave functions and to the parts of the Hamiltonian itself. These approximations allow inaccuracies to accumulate in the calculation and the QSAR models that may use them. One way to avoid quantum mechanics calculations in QSAR is to use the results of quantum mechanics experiments that have spectral energy outputs. Spectrometric data experiments such NMR, IR, UV−vis, Raman, and fluorescence all give outputs that directly correspond to different available quantum mechanics energies of a compound. An EI MS experiment is a type of spectrometric data experiment that can be used for obtaining information about the available fragment sizes of a compound.

We decided to test whether the use of any combination of ^{13}C NMR, IR, and EI MS data could be used to produce an accurate spectrometric data−activity relationship (SDAR) of the monodechlorination of 32 chlorobenzenes, chlorophenols, and chloroanilines. We have demonstrated that ^{13}C NMR and EI MS data can be used to produce a reliable SDAR model of a compound's ability to bind to the estrogen receptor.[6] The power of SDAR is that we did not need to solve any quantum mechanic calculations or use the structure of the molecule in the calculation as is done in QSAR techniques.[7−11]

^{13}C nuclear magnetic resonance (NMR) chemical shifts have been used to predict and refine chemical structures.[12,13] Conversely, the chemical structure of a compound has been used to predict its ^{13}C NMR chemical shifts.[14] The ^{13}C NMR spectrum of a compound contains frequencies that correspond directly to the quantum mechanical properties of the molecule. The quantum mechanical description of a molecule depends largely on its electrostatic features and geometry.[15] Ab initio quantum mechanical calculations of ^{13}C chemical shift tensors in proteins reveal that they are dependent on the structural environment.[16] ADC Labs now has software that will take a ^{13}C NMR one-dimensional spectrum and predict a structure.[17]

IR absorption spectra, like ^{13}C NMR chemical shifts, have been used to predict molecular structure.[18] Conversely, like ^{13}C NMR spectra, the chemical structure of a compound has been used to predict its IR spectrum.[19] The IR spectrum, particularly absorptions between 600 and 4000 cm^{-1}, reveals

* Corresponding author. Phone: (870) 543-7080. FAX: (870) 543-7686. E-mail: rbeger@NCTR.FDA.GOV.

10.1021/ci000331v
Published on Web 09/28/2000

how absorbed energy is transformed into the heat energy of molecular vibrations and rotations. Calculating the IR spectra for a molecular structure is computationally demanding and requires an ab initio or a density function approach. QSAR models that have used IR spectra as comprehensive descriptors showed that the information provided by IR absorption is different from those currently used by QSAR and has meaning in relation to biological activity.[20]

Using 3D-QSAR (three-dimensional quantitative structure−activity relationship) modeling results, receptor binding of a compound can be predicted, based in part upon electrostatics and geometrical structure. Therefore, we postulated that we could use ^{13}C NMR, IR, and EI MS data in much the same way that 3D-QSAR uses constitutional, topological, geometrical, electrostatic, and quantum descriptors to model receptor binding of a compound.[21−24] By combining the ^{13}C NMR, IR, and EI MS data into a composite set of descriptors and putting them into pattern recognition programs, it should be possible to produce a predictive SDAR model of the monodechlorination of 32 chlorinated benzene compounds.

Current quanitative structure−activity relationship (QSAR) and structure−activity relationship (SAR) models use computer modeling of the molecule and/or break the molecule into secondary structural motifs. Using this approach, the structure or substructure must be determined before one can either define the QSAR or SAR model or use it for predictive purposes.[21−26] Using a specific molecular structure for computer modeling of each compound dramatically extends the number of calculations required to define the QSAR model. Moreover, the selection of the most appropriate 3D structure for each molecule requires a number of assumptions. The necessary simplifying assumptions in some cases give results that are hard to replicate or are inaccurate. By contrast, an experimentally determined SDAR model requires no knowledge or assumptions regarding the molecular structure of the compounds being modeled. Each NMR chemical shift or IR absorption peak acts like a quantum mechanical identifier of a four to eight atom secondary structural motif. These quantum mechanical identifiers are used in the same manner as in current QSAR and SAR models that break the molecule into secondary structural motifs. This SDAR model is fast and gives a straightforward classification for the fastest rate of monodechlorination compounds. Moreover, it requires only experimental spectral data (^{13}C NMR, IR, and EI MS) that were readily attainable for this SDAR monodechlorination model.

PROCEDURE

The monodechlorination rates for 32 chlorinated benzene compounds were found on the Web site www.aist.go.jp/RIODB/dbefc/nirefate/deg_E.html.[27] We were able to find most of the ^{13}C NMR, IR, and EI MS spectral data from the Integrated Spectral Data Base System for Organic Compounds Web site www.aist.go.jp/RIODB/SDBS/.[28] The *Aldrich Library of ^{13}C and ^{1}H FT NMR Spectra*[29] and the NIST MS database software version 1.6[30] were used to obtain the rest of the needed spectral data. All the ^{13}C NMR spectra were run within $CDCl_3$ solvent. All the IR absorption spectra were done on a condensed-phase KBr disk or liquid film.

We used the ^{13}C NMR, IR, and EI MS data points in the same way that 3D-QSAR uses comprehensive descriptors for structural and statistical analyses (CODESSA).[23] Mass spectroscopic data from *m*/*z* 50−300 were used. Unassigned one-dimensional ^{13}C NMR chemical shifts were segregated into bins over a 0−200 ppm range. The ^{13}C NMR frequencies were shifted to bins 1701−1920, so bin 1701 was the ^{13}C NMR spectrum for frequencies inside 0.5−1.49 ppm and bin 1900 was the ^{13}C NMR spectrum for frequencies inside 199.5−200.49 ppm. Unassigned IR absorption were segregated into bins over a 500−1700 cm^{-1} range. We used only 500−1700 cm^{-1} because the data outside this region were very dependent on the different processes used to obtain the condensed-phase IR experimental data. The IR absorption spectra was put into 10 cm^{-1} wide bins from 500 to 1700^{-1}. The 500 bin was used for IR absorption from 495 to 504 (10 cm^{-1}), the 510 bin was used for IR absorption from 505 to 514 (10 cm^{-1}), and so on to the 1700 cm^{-1} bin.

The ^{13}C NMR spectra were saved as the area under the peak within a certain spectral range and normalized to an integer. A nondegenerate frequency was assigned an area of 25, a doubly degenerate frequency had an area of 50, and so forth. This was done so that all the spectra would have a similar signal-to-noise ratio and to eliminate line width variations due to differences in NMR instrumental field strengths, shimming, temperature, pH, or solvents. The spectral width of the ^{13}C NMR bin used in this study was 1 ppm for each bin.

The only part of the IR spectra that was used was the IR absorption maximum frequencies. This was done because the data obtained from the SDBS Web site[28] only had that data in tabular form, and like ^{13}C NMR the width of a peak is dependent on the rotation states of the atoms in the compound. The IR absorption (A) was determined by $A = 1 - T$, where T is the transmittance. SDAR models where the absorption (A) was determined by $A = -\log(T)$ were not as accurate, and the results are not shown here. One possible reason that SDAR models that used $A = 1 - T$ were more accurate is that all the transmitted data is treated the same way and, when $A = -\log(T)$ is used, the large transmittances (small absorption) are reduced more than the small transmittances (large absorption). This may cause the SDAR to rely too much on the large absorption and statistically diminish the effect of the small absorption that appears to be important.

The monodechlorinations of 32 different chlorobenzenes, chlorophenols, and chloroanilines were classified as readily monodechlorinated when the half-life was less than 30 days to remove one chlorine atom from the chlorinated compound. The rest of the compounds had half-lives greater than 30 days to remove one chlorine atom and were classified as not readily monodechlorinated. We chose 30 days as a cutoff because it left with almost an even distribution of 17 readily monodechlorinated and 15 not readily monodechlorinated in the SDAR models.

The analysis of the SDAR model was done by the leave-one-out (LOO) cross-validation procedure where each compound is systematically excluded from the training set and its monodechlorination is predicted by the model.[31] All 32 compounds were used in eight separate sets of leave-four-out predictions that were performed to determine the predictive accuracy of the four SDAR models.

The pattern recognition software used was Resolve Version 1.2 (Colorado School of Mines, Boulder, CO). The ^{13}C NMR, IR, and EI MS spectroscopic data for all 32 compounds were

Table 1. Monodechlorination Half-life Rates in Days[a]

compound	no.	half-life (days)	input class	output			
				C NMR IR EI MS	C NMR IR	IR EI MS	C NMR EI MS
chlorobenzene	**1**	46.2	N	N	N	N	N
1,2-dichlorobenzene	**2**	36.9	N	N	N	N	N
1,3-dichlorobenzene	**3**	433	N	N	N	N	N
1,4-dichlorobenzene	**4**	385	N	N	N	N	N
1,2,3-trichlorobenzene	**5**	23	R	R	R	N	N
1,2,4-trichlorobenzene	**6**	40.8	N	N	N	N	N
1,3,5-trichlorobenzene	**7**	35	N	N	N	N	N
1,2,3,4-tetrachlorobenzene	**8**	18.2	R	R	R	R	R
1,2,3,5-tetrachlorobenzene	**9**	18.6	R	R	R	R	R
1,2,4,5-tetrachlorobenzene	**10**	28.8	R	R	R	R	R
pentachlorobenzene	**11**	17.8	R	R	R	R	R
hexachlorobenzene	**12**	27.1	R	R	R	N	N
2-chloroaniline	**13**	175	N	N	N	N	N
3-chloroaniline	**14**	672	N	N	N	N	N
4-chloroaniline	**15**	203	N	N	N	N	N
2,3-dichloroaniline	**16**	54	N	N	N	N	N
3,4-dichloroaniline	**17**	55	N	N	N	N	N
3,5-dichloroaniline	**18**	149	N	N	N	N	N
3,4,5-trichloroaniline	**19**	145	N	N	N	N	N
2,3,4,5-tetrachloroaniline	**20**	57	N	N	N	N	N
2-chlorophenol	**21**	6.9	R	R	R	R	R
4-chlorophenol	**22**	11.7	R	R	N	R	R
2,3-dichlorophenol	**23**	13.2	R	R	N	R	R
2,4-dichlorophenol	**24**	15.3	R	R	R	R	R
2,5-dichlorophenol	**25**	46.3	N	R	N	R	R
2,6-dichlorophenol	**26**	5.0	R	R	N	R	R
3,4-dichlorophenol	**27**	69.9	N	N	N	N	N
3,5-dichlororphenol	**28**	17.3	R	N	R	N	N
2,3,4-trichlorophenol	**29**	1.8	R	R	R	R	N
3,4,5-trichlorophenol	**30**	21.7	R	R	R	R	R
2,3,4,5-tetrachlorophenol	**31**	6.5	R	R	R	R	R
pentachlorophenol	**32**	2.1	R	R	R	N	R

[a] R stands for a readily monodechlorinated classification with a half-life less than 30 days, and N stands for a not readily monodechlorinated classification with a half-life greater than 30 days.

input into the Resolve software. The spectroscopic data were then autoscaled and variance-weighted prior to principal component analysis. The discriminant analysis was based on the canonical variate vector, and leave-one-out (LOO) cross-validation was used for each compound to maximize the size of the training set.

Autoscaling scales the quantitative response at each spectral bin to a standard deviation percentage within the bin. An average value with standard deviation is calculated for each bin. Then, for each spectrum, the quantitative response at a bin is expressed as the number of standard deviations above or below the average. This data pretreatment step equalizes the weight of consistent variance of spectrometric data with inherently small magnitudes to spectrometric data whose signals have large magnitudes. Autoscaling, which automatically compensates for gross magnitude variations, is particularly important for this application in which two or more completely different types of analytical spectra are formed into a composite surrogate representative of important molecular characteristics affecting the compound's monodechlorination characteristics.

Variance weighting treats the data in a statistical way that emphasizes those spectral characteristics important in distinguishing defined groups. For every spectral bin, the variance between groups is divided by the variance within groups. The resulting dividend becomes a weighting factor that has a magnitude larger than 1 when a particular spectral bin has an important role in distinguishing groups. Variance weighting all of the raw spectra before pattern recognition increases the power of discriminant function analysis. It is particularly important in this application because it deemphasizes the relative importance of irrelevant spectral information.

RESULTS

Table 1 contains the compound name, the monodechlorination half-life, a training input (N for not readily monodechlorinated and R for readily monodechloroinated), a prediction using ^{13}C NMR, IR, and EI MS data, a prediction using ^{13}C NMR and IR data, a prediction using EI MS and IR data, and a prediction using ^{13}C NMR and EI MS data.

On the basis of only ^{13}C NMR, IR, and EI MS data, the statistical pattern recognition program with 16 principal components (PCs) used 76.0% of the total variance and a cross-validation of 93.8%. Figure 1 shows the discriminant function for the composite ^{13}C NMR, IR, and EI MS SDAR monodechlorination model. There is a large separation in the discriminant function between the readily monodechlorinated (white background) and not readily monodechlorinated compounds (gray background), except for 3,5-dichlorophenol and 2,5-dichlorophenol. 3,5-Dichlorophenol and 2,5-dichlorophenol are part of six dichlorophenols in the SDAR of which four dichlorophenols are readily monodechlorinated and two dichlorophenols are not readily monodechlorinated in 30 days. This inconsistency in dichlorophenol monodechlorinations causes problems in the model classification prediction because the EI MS data are very similar for all six dichlorophenols with a molecular weight of 162.

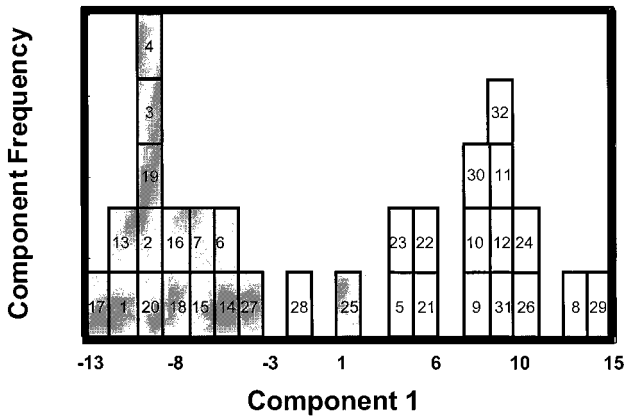

Figure 1. The discriminant function using ^{13}C NMR, IR, and EI MS data in the SDAR model. The *X*-axis is the first principal component, and the *Y*-axis is the second principal component. Compounds with white background are readily monodechlorinated, and compounds with a gray background are not readily monodechlorinated.

The negative peaks in all the panels of Figures 2, 4, 6, and 8 correspond to bins that bias toward not readily monodechlorinated, and positive peaks correspond to bins that bias toward readily monodechlorinated. Figure 2A shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on EI/MS data from *m*/*z* 0−300. Figure 2B shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on ^{13}C NMR data from 100 to 200 ppm. The negative peaks near 145 ppm signify the carbon atom near the amine group in a chloroaniline compound, and positive peaks near 153 ppm signify the carbon near the hydroxy group of a chlorophenol compound. Figure 2C shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on IR data from 700 to 1700 cm^{-1}. The IR data canonical loadings are evenly divided between positive and negative.

Based only on ^{13}C NMR and IR data, the statistical pattern recognition program with 15 principal components (PCs) used 69.7% of the total variance and a cross-validation of 90.6%. Figure 3 shows the discriminant function for the composite ^{13}C NMR and IR SDAR monodechlorination model. There is a very large separation in the discriminant function between the readily monodechlorinated and not readily monodechlorinated compounds when using only ^{13}C NMR and IR data in the SDAR model. 4-Chlorophenol, 2,3-dichlorophenol, and 2,6-dichlorophenol do not cross-validate correctly and all the chloroanilines and chlorobenzenes cross-validate correctly when using only ^{13}C NMR and IR data in the monodechlorination SDAR model.

Figure 4A shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on ^{13}C NMR data from 100 to 200 ppm. Figure 4B shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on IR data from 700 to 1700 cm^{-1}. The ^{13}C NMR factor loading in Figure 4A is very similar to factor loadings seen in Figure 2B, and the IR factor loading of Figure 4B is very similar to factor loadings seen in Figure 2C.

Based only on IR and EI MS data, the statistical pattern recognition program with 17 principal components (PCs) used 81.2% of the total variance and a cross-validation of 84.4%. Figure 5 shows the discriminant function for the composite IR and EI MS SDAR monodechlorination model. There is a very large separation in the discriminant function between the readily monodechlorinated and not readily monodechlorinated compounds, except for 3,5-dichlorophenol, 3,4-dichlorophenol, and 2,5-dichlorophenol which appear to be modeled indecisively. 3,5-Dichlorophenol, 3,4-dichlorophenol, and 2,5-dichlorophenol are part of six dichlorophenols in the SDAR of which four dichlorophenols are readily monodechlorinated and two dichlorophenols are not readily monodechlorinated in 30 days. This inconsistency in dichlorophenol monodechlorination causes problems in the SDAR model classification prediction because the EI MS is very similar for all six dichlorphenols.

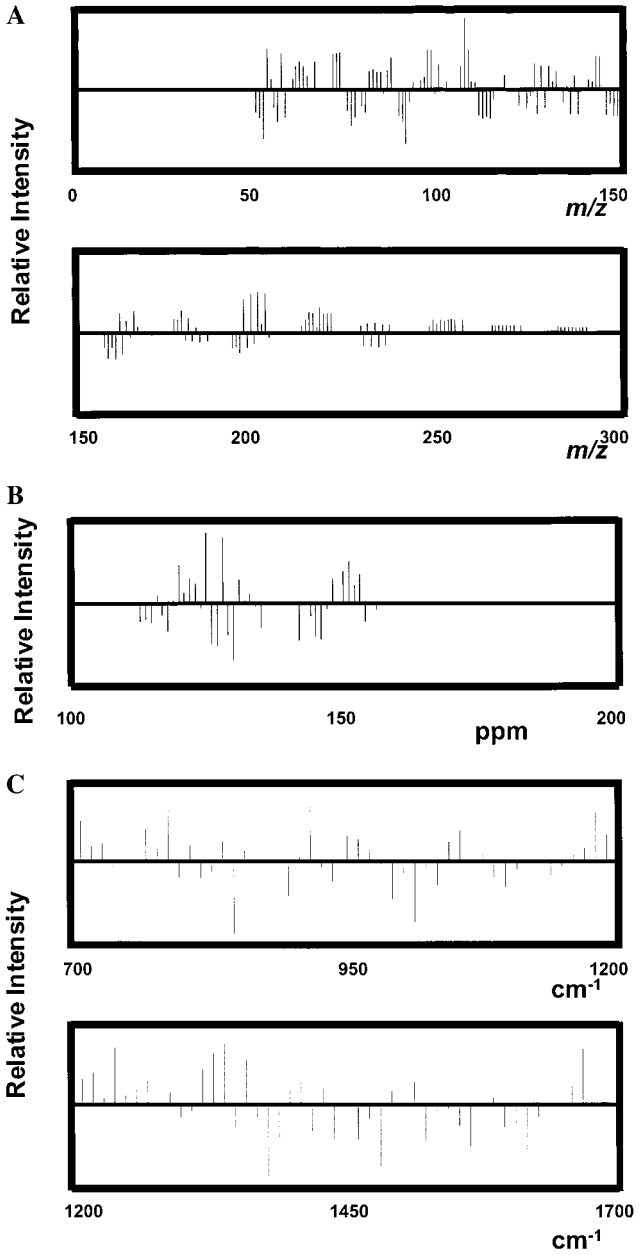

Figure 2. The first canonical variate using ^{13}C NMR, EI MS, and IR data in the SDAR model. The *X*-axis is the bin number and the *Y*-axis is the relative intensity. (A) EI MS data factor loadings from 0 to 300 *m*/*z*. (B) ^{13}C NMR data factor loadings from 100 to 200 ppm. (C) IR data factor loadings from 700 to 1700 cm^{-1}.

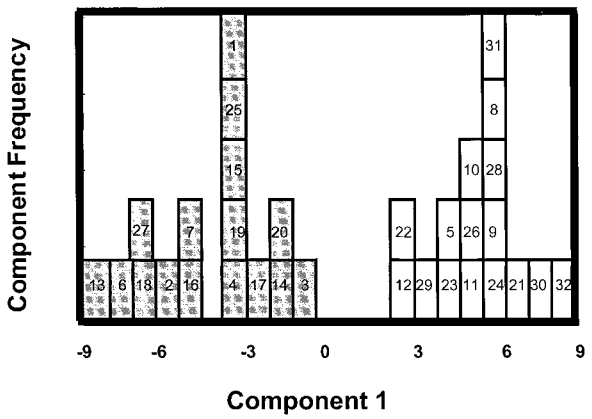

Figure 3. The discriminant function using ^{13}C NMR and IR data in the SDAR model. The *X*-axis is the first principal component, and the *Y*-axis is the second principal component. Compounds with white background are readily monodechlorinated, and compounds with a gray background are not readily monodechlorinated.

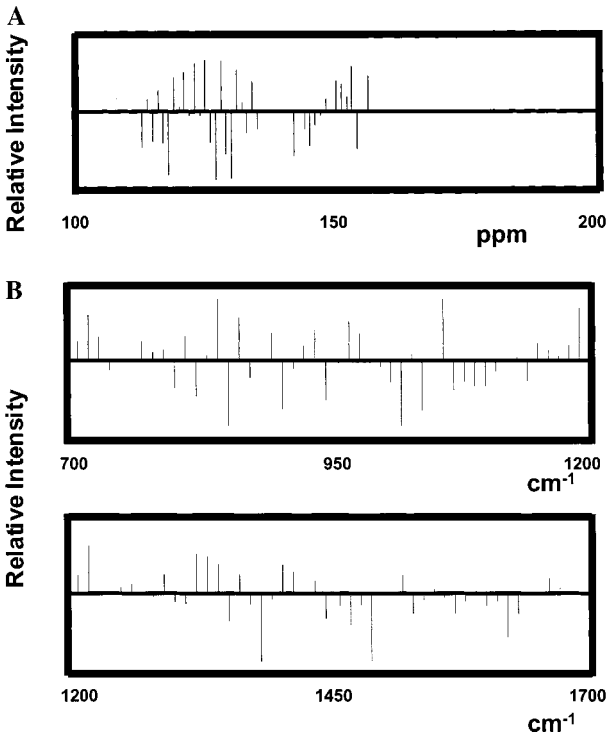

Figure 4. The first canonical variate using ^{13}C NMR and IR data in the SDAR model. The *X*-axis is the bin number, and the *Y*-axis is the relative intensity. (A) ^{13}C NMR data factor loadings from 100 to 200 ppm, (B) IR data factor loadings from 700 to 1700 cm^{-1}.

Figure 6A shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on EI/MS data from *m*/*z* 0−300. Figure 6B shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on IR data from 700 to 1700 cm^{-1}.

Based only on ^{13}C NMR and EI MS data, the statistical pattern recognition program with 20 principal components (PCs) used 93.1% of the total variance and a cross-validation of 84.4%. Figure 7 shows the discriminant function for the composite ^{13}C NMR and EI MS SDAR monodechlorination model. There is a very large separation in the discriminant function between the readily monodechlorinated and not readily monodechlorinated compounds, except for 3,5-dichlorophenol and 2,5-dichlorophenol. Once again, 3,5-dichlorophenol and 2,5-dichlorophenol are part of six dichlorophenols in the SDAR of which four dichlorophenols are readily monodechlorinated and two dichlorophenols are not readily monodechlorinated in 30 days.

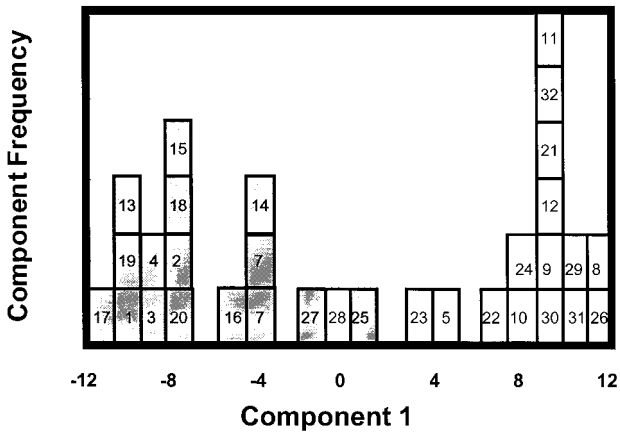

Figure 5. The discriminant function using IR and EI MS data in the SDAR model. The *X*-axis is the first principal component, and the *Y*-axis is the second principal component. Compounds with white background are readily monodechlorinated, and compounds with a gray background are not readily monodechlorinated.

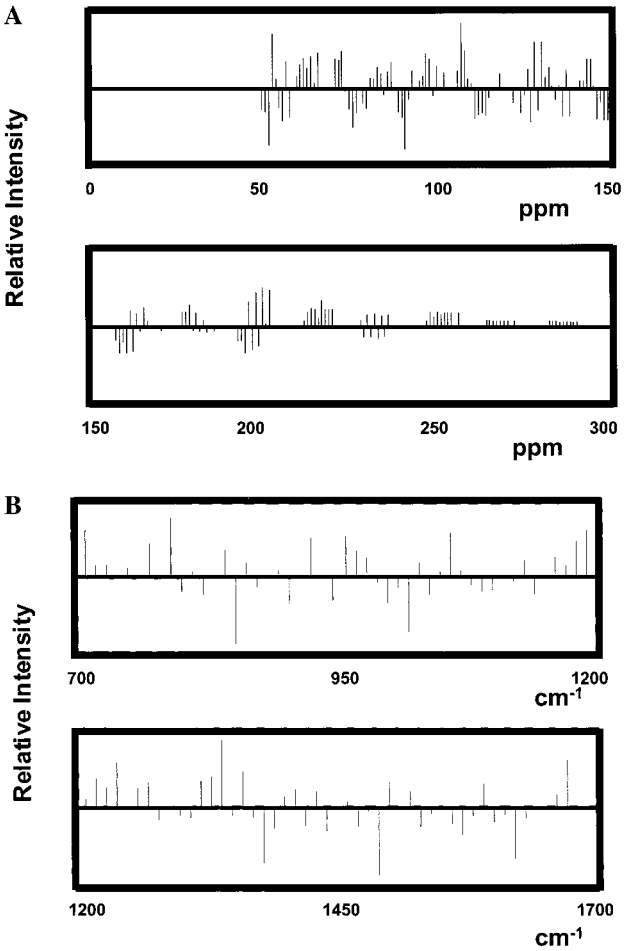

Figure 6. The first canonical variate using EI MS and IR data in the SDAR model. The *X*-axis is the bin number, and the *Y*-axis is the relative intensity. (A) EI MS data factor loadings from 0 to 300 *m*/*z*. (B) IR data factor loadings from 700 to 1700 cm^{-1}.

Figure 8A shows the factor loadings associated with the first canonical variate function used in the pattern recognition

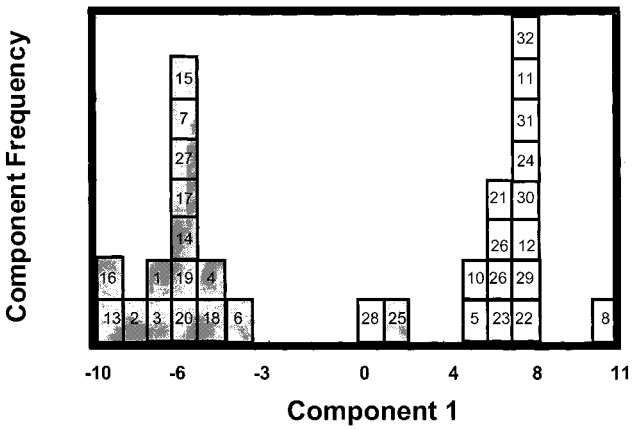

Figure 7. The discriminant function using ^{13}C NMR and EI MS data in the SDAR model. The *X*-axis is the first principal component, and the *Y*-axis is the second principal component. Compounds with white background are readily monodechlorinated, and compounds with a gray background are not readily monodechlorinated.

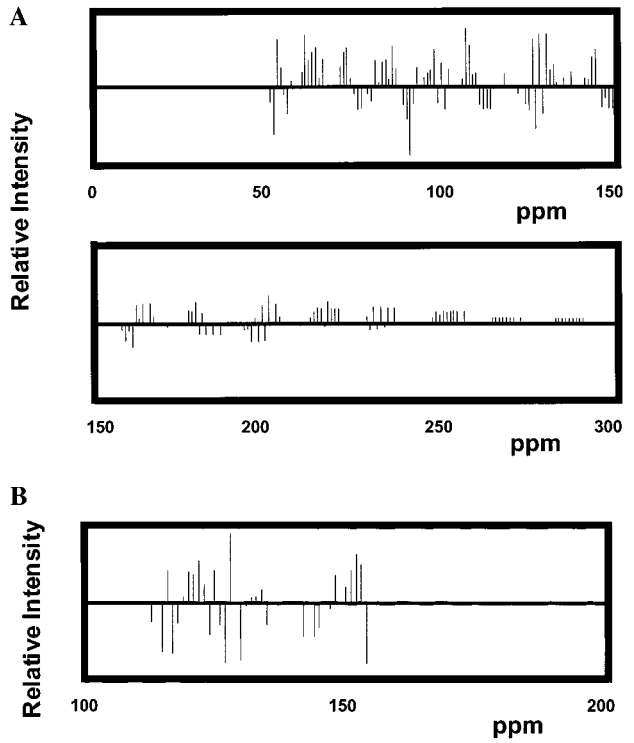

Figure 8. The first canonical variate using EI MS and ^{13}C NMR data in the SDAR model. The *X*-axis is the bin number, and the *Y*-axis is the relative intensity. (A) EI MS data factor loadings from 0 to 300 *m*/*z*. (B) ^{13}C NMR data factor loadings from 100 to 200 ppm

based on EI MS data from *m*/*z* 0−300. Figure 8B shows the factor loadings associated with the first canonical variate function used in the pattern recognition based on ^{13}C NMR data from 100 to 200 ppm.

The eight sets of leave-four-out predictions were performed on all four SDAR models of monodechlorination. The leave-four-out predictions using the SDAR model based on ^{13}C NMR, IR, and EI MS data was 87.5% correct. The leave-four-out predictions using the SDAR model based on ^{13}C NMR and IR data was 75.0% correct. The leave-four-out predictions using the SDAR model based on ^{13}C NMR and EI MS data was 75.0% correct. The leave-four-out predictions using the SDAR model based on EI MS and IR data was 75.0% correct.

CONCLUSIONS

When building a SDAR model, there is no need to break the compound into secondary structural motifs, as is done in a CODESSA[22,23] QSAR. In all four SDAR models of monodechlorination, we were able to make a reliable model of the removal of a chlorine atom at any of the six possible chlorinated positions of 32 different chlorobenznes, chlorophenols, and chloroanilines with accuracy greater than 84.4%. The SDAR model that did use EI MS, ^{13}C NMR, and IR data had a leave-one-out cross-validation of 93.8%. The EI MS and ^{13}C NMR and the EI MS and IR SDAR models appear to have a hard time deciding what to do with the EI MS data of six dichlorophenol cpompounds that are split into four readily monodechlorinated and two monodechlorinated classifications. EI MS data are important, because in these SDAR models the higher the molecular weight of the compound, the more chlorine atoms were in the compound and the more likely the compound was readily monodechlorinated. The same problem occurs with the three trichlorobenzene compounds: they are split into two not readily monodechlorinated compounds and one readily monodechlorinated compound and have a molecular weight of 180. The results of all the SDAR models are very good since there is only an order of magnitude of variation in the lifetimes between compounds that are readily monodechlorinated and compounds that are not readily monodechlorinated.

All four SDAR models had a very strong leave-four-out predictive power. The leave-four-out 87.5% prediction of the model based on ^{13}C NMR, IR, and EI MS data was the best of all four SDAR models. The leave-four-out 75.0% prediction rate for the other three SDAR models was very encouraging.

The factor loadings associated with the first canonical variate function for EI MS, ^{13}C NMR, and IR data in Figures 2, 4, 6, and 8 are very similar. Almost every bin in the factor loadings of EI MS, ^{13}C NMR, and IR data as seen in Figures 2, 4, 6, and 8 does not change its bias toward readily monodechlorinated (pointing upward) and toward not readily monodechlorinated (pointing downward) when comparing one SDAR model to another SDAR model. Since there is a strong consistency between all four SDAR model reliabilities and factor loadings, this supports the hypothesis that there is a strong quantum mechanical connection among structure, spectra, and the microbial anaerobic activity of the monodechlorination of chlorinated benzene compounds.

The strong predictive quality of four SDAR classification model results presented here demonstrates a need to continue this line of research. Therefore, we are planning to use the same monodechlorination data to produce quantitative SDAR models (QSDAR) and compare them to 3D-QSAR and to calculate quantum mechanics properties to produce models as an alternative or complement to producing SDAR models.

REFERENCES AND NOTES

(1) Masunaga S.; Susarla, S.; Gundersen, J. L.; Yonezaw, Y. Pathway and Rate of Chlorophenol Transformation in Anaerobic Estuarine Sediment. *Environ. Sci. Technol.* **1996**, *30* (4), 1253−1260.

(2) Masunaga, S.; Susarla, S.; Yonezawa, Y. Dechlorination of Chlorobenzenes in Anaerobic Estuarine Sediment. *Water Sci. Technol.* **1996**, *33* (6), 173−180.
(3) Yonezawa, Y.; Fukui, M.; Masunaga, S.; Urushgawa, Y. Dechlorination of 1,2,4-Trichlorobenzene in the Sediment of Ise Bay. *Chemosophere* **1994**, *28* (12), 2179−2184.
(4) Susarla, S.; Masunaga, S.; Yonezawa, Y. Transformation of Chloronitrobenznes in Anaerobic Sediment. *Chemosphere* **1996**, *32* (5), 967−977.
(5) Susarla, S.; Yonezawa, Y.; Masunaga, S. Reductive Dehalogenation of Chloroanilines in Anaerobic Esturerine Sediment. *Environ. Technol.* **1997**, *18*, 75−83.
(6) Beger, R. D.; Freeman, J. P.; Lay, Jr., J. O.; Wilkes, J. G.; Miller, D. W. ^{13}C NMR and Electron Ionization Mass Spectrometric Data-Activity Relationship Model of Estrogen Receptor Binding. *Toxicol. Appl. Pharmacol.* **2000**, *168*, submitted.
(7) Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S. Prediction of gas chromatographic retention times and response factors using a general quantitative structure−property relationship. *Anal. Chem.* **1994**, *66*, 1799−1807.
(8) Katritzky, A. R.; Mu, L.; Labanov, V. S.; Karelson, M. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.* **1996**, *100*, 10400−10407.
(9) Fujita, T.; Iwasa, J.; Hansch, C. A new substituent constant, π, derived from partition coefficient. *J. Am. Chem. Soc.* **1964**, *86*, 5175−5180.
(10) Branbury, S. P. Quantitative structure−activity relationship and ecological risk assessment: an overview of predictive aquatic toxicology research. *Toxicology* **1995**, *25* (1), 67−89.
(11) Cramer, R. D.; Paterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(12) Beger, R. D.; Bolton, P. H. Protein ϕ and ψ dihedrals restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J. Biomol. NMR* **1997**, *10*, 129−142.
(13) Wishart, D. S.; Sykes B. D. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **1994**, *239*, 363−392.
(14) Kvasnicka, V. An Application of Neural Networks in Chemistry. Prediction of ^{13}C NMR Chemical Shifts *J. Math. Chem.* **1991**, *6*, 63−76.
(15) Emsley, J. W.; Feeney, J.; Sutcliffe, L. H. *High-Resolution Nuclear Magnetic Resonance*; Pergamon Press Ltd.: Oxford, 1965; Vol. I, Chapter 8, p 287.
(16) De Dios, A. C.; Pearson, J. G.; Oldfield, E. Secondary and tertiary structural effects on protein NMR chemical shifts: an *ab initio* approach. *Science* **1993**, *260*, 1491−1496.
(17) *ACD/Labs Structure Elucidator*; Toronto, Canada.
(18) Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. The Prediction of the 3D Structure of Organic Molecules from Their Infrared Spectra. *Vib. Spectrosc*. **1999**, *19*, 151−164.
(19) Gasteiger, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Finding the 3D structure of a molecule in its IR spectrum. *J. Anal. Chem.* **1997**, *359*, 50−55.
(20) Benigni, R.; Passerini, L.; Livingstone, D. J.; Johnson, M. A.; Giuliani, A. Infrared Spectra information and their correlation with QSAR descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 558.
(21) Hansch, C.; Leo, A. *Exploring QSAR—Fundamentals and applications in chemistry and biology*; American Chemical Society: Washington, DC, 1995.
(22) Tong W.; Perkins, R.; Strelitz, R.; Collantes, E. R.; Keenan, S.; Welsh, W. J.; Branham, W. S.; Sheehan, D. M. Quantitative Structure−Activity Relationships (QSARs) for Estrogen Binding to the Estrogen Receptor: Predictions across Species. *Environ. Health Perspect*. **1997**, *105*, 1116−1124.
(23) Tong, W.; Collantes, E.; Chen, Y.; Welsh, W. J. A. A comparative molecular field analysis study of N-benzylpiperidines as acetylcholinesterase inhibitors. *J. Med. Chem.* **1995**, *39* (2), 380−387.
(24) Collantes, E.; Tong, W.; Welsh, W. Use of moment of inertia in comparative molecular field analysis to model chromatographic retention of nonpolar solutes. *J. Anal. Chem.* **1996**, *68*, 2038−2043.
(25) Klopman, G. Artificial intelligence approach to structure−activity studies. Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(26) Klopman, G. MULTICASE1. A hierarchial computer automated structure evaluation program. *Quant. Struct.-Act. Relat.* **1992**, *11*, 176−184.
(27) Database for environmental fate of chemicals Web site www.aist.go.ip/RIODB/dbefc/nirefate/deg_E.html.
(28) Integrated Spectral Data Base System for Organic Compounds Web site www.aist.go.ip/RIODB/SDBS/.
(29) Pouchert, C. J.; Behnke, J. *The Aldrich Library of ^{13}C and ^{1}H FT NMR Spectra*, 1st ed.; Aldrich Chemical Co.: 1993; Vols. 1−3.
(30) *NIST 1998 Mass Spectral Library*, Version 1.6; US Department of Commerce, National Institute of Standards and Technology, Standard Reference Data Program: Gaithersburg, MD.
(31) Cramer, R. D.; Bunce, J. D.; Patterson, D. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.

CI000331V