Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 795–805

BIOORGANIC & MEDICINAL CHEMISTRY

The Use of 'Electronic Nose' Sensor Responses to Predict the Inhibition Activity of Alcohols on the Cytochrome P-450 Catalyzed *p*-Hydroxylation of Aniline

Thomas P. Vaid and Nathan S. Lewis*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Received 30 August 1999; accepted 26 October 1999

Abstract—A quantitative structure–activity relationship (QSAR) has been formulated to describe the inhibitory action of a series of alcohols on the cytochrome P-450 catalyzed *p*-hydroxylation of aniline. The descriptors used in the QSAR are the responses of individual sensors in a polymer-based electronic nose, and are all easily generated experimental values. If the various electronic nose sensor response patterns for the family of test alcohols reflect differences in the chemical properties that are involved in the cytochrome P-450 inhibition process, it ought to be possible to correlate the differences in the electronic nose signals of these analytes with the differences in the cytochrome P-450 inhibition by these species. To evaluate this possibility, multiple linear regression was performed on data obtained from exposure of a series of test alcohols to 19 sensors of a conducting polymer composite electronic nose array. A genetic algorithm was then used to select the optimal set of sensors that best described the inhibitory activity of these alcohols within a linear regression model. The regression equation fit the inhibition data of 20 of the alcohols with an R of 0.995. This fit compares favorably with previously published QSARs on this system that have used log P (P≡octanol–water partition coefficient) along with steric parameters of the alcohols, and also compares favorably to QSARs formulated using theoretically calculated parameters.

Introduction

Quantitative structure–activity relationships (QSAR) are frequently used in drug discovery to direct the construction of libraries and to identify target molecules, or classes of molecules, for detailed experimental investigation with respect to a phenomenon of interest.[1,2] Common descriptors include experimental properties, such as log P (P≡octanol–water partition coefficient), which is a measure of hydrophobicity, and Hammett constants, which describe electronic effects of individual substituents, as well as a variety of calculated parameters, including molecular orbital energies, topological indices, and steric parameters.[2] An advantage of calculated descriptors is that test molecules need not be synthesized in order to calculate their predicted activity. However, with large numbers of candidate molecules now being synthesized via combinatorial methods, a straightforward test to predict the biological activity of existing molecules could be quite useful.

The approach evaluated herein for a specific, initial test case uses the experimental data that are generated when an array of polymer/carbon black composite sensors that comprise an 'electronic nose' is exposed to the molecules of interest.[3] The steady-state change in the electrical resistances of such sensors can be related to the equilibrium sorption properties of an analyte vapor into the polymer films of the various sensors.[3] The relative sorption properties of an analyte into a variety of different polymers produces a pattern of sensor signals that allows classification of the analyte based on the resultant electronic nose sensor response data.[3] Although it is clear that individual gas/polymer partition coefficients can provide information on physicochemical properties of an analyte, such as its dipole moment, index of refraction, or water/octanol partition coefficient,[4] the differences in the relative sorption properties of various analytes into an array of polymer-composite based sensors also apparently yields information on the steric and electronic properties of the analyte. This information allows discrimination not only between analytes in distinct functional group classes (alcohols versus alkanes versus aromatics versus chlorinated hydrocarbons)[3,5] but also between members of a class (such as straight chain alkanes, straight chain alcohols,

*Corresponding author. Fax: +1-625-795-7487.

0968-0896/00/$ - see front matter
PII: S0968-0896(99)00304-1

etc.).[6] This relative sorption information probed by a conducting polymer composite sensor array even allows robust discrimination between closely related analytes that primarily differ in their steric properties, such as *o*- and *m*-xylene,[7] as well as between enantiomers (provided that polymers with chiral centers are included in the electronic nose array).[8] If the significant interactions between a molecule and the binding site of an enzyme are reflected either directly or indirectly in the collection of differential binding constants of that molecule to the polymers in an electronic nose, then it might be possible to relate the differences in electronic nose patterns to the relative enzyme binding properties of those same analytes. The purpose of this work was to evaluate this hypothesis on a specific test system.

In 1973, Cohen and Mannering published the results of a study on the inhibitory effect of alcohols upon the *p*-hydroxylation of aniline by cytochrome P-450.[9] The inhibitory action of the alcohols varied greatly with structure, and most likely occurred through competitive binding to the active site of the enzyme. In general, more hydrophobic and/or less branched alcohols exhibited stronger enzymatic inhibition activities. The authors correlated the activity of the unbranched 1- and 2-alcohols with log *P*, but the activities of the other alcohols could not be adequately modeled. Four subsequent QSARs on the same data set have been published, with varying degrees of success.[10–13] The examined alcohols are volatile organic compounds and the measured biological activity is the simple inhibition of enzymatic activity in a well-understood reaction; consequently, this system was identified as an excellent candidate to evaluate the potential use of electronic nose signals to provide descriptors for such processes. If the polymer composite sensor signals only probe properties such as hydrophobicity, then the correlations between their responses and the test biological activity ought to be similar to the unsatisfactory correlations obtained in the cytochrome P-450 alcohol system using QSAR models that employ only log *P* as a descriptor.[9] Alternatively, if the relative sorption properties into the polymer composite sensors of the electronic nose also contain information on a variety of steric and electronic properties of the analytes, then it might be possible to obtain improved statistical correlations between the relative biological activity of these test analytes and their electronic nose patterns.

Results

Figure 1 presents the relative differential resistance responses for various conducting polymer composite sensors to three representative alcohols, and Table 1 summarizes all of the sensor response data for the various alcohols investigated in this work. Each alcohol produced a distinct, characteristic response pattern with the array of sensors chosen for use in this work.

The responses of the 19 working sensors to 20 of the alcohols (Table 1) were used to build a QSAR model.

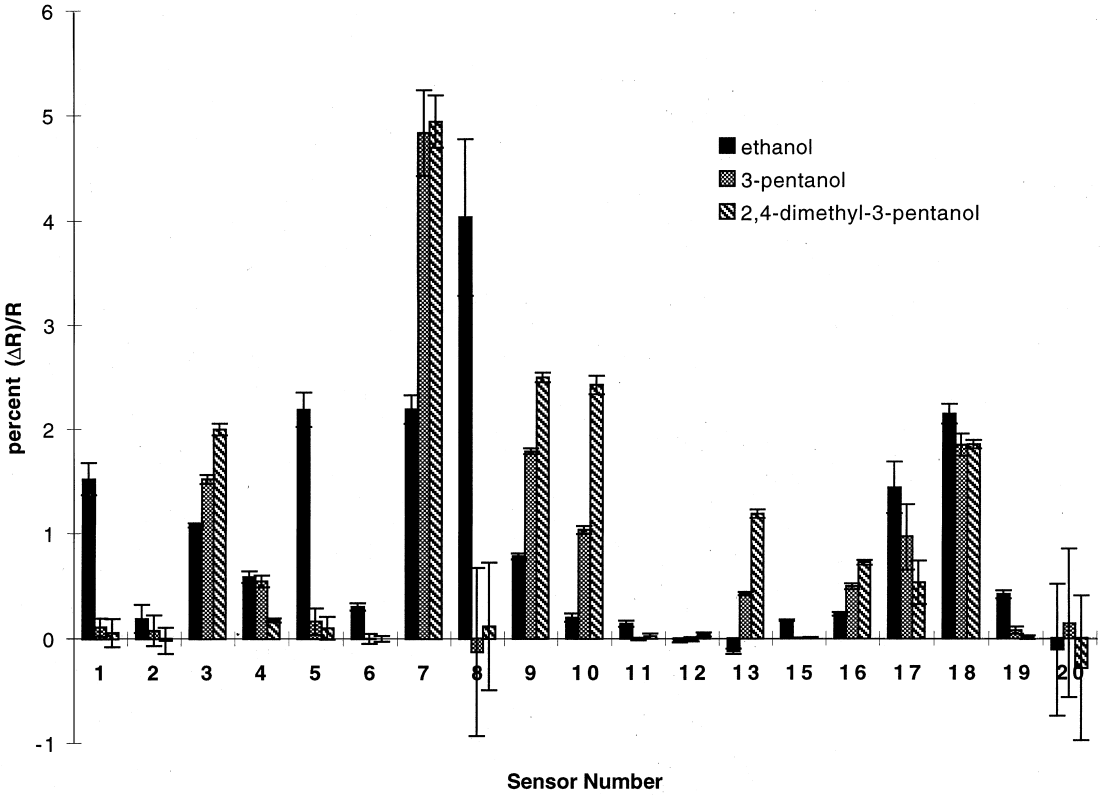

Figure 1. Responses (expressed as percent change in electrical resistance relative to baseline resistance) of 19 different polymer/carbon black sensors upon exposure to ethanol, 3-pentanol, and 2,4-dimethyl-3-pentanol at 5% of their respective vapor pressures. Error bars represent the standard deviation over 10 exposures.

Benzyl alcohol and *tert*-amyl alcohol were excluded from the fit because their biological activities were anomolous.[9] The two diols were also excluded while building the model.

The inhibitory action data of Cohen and Mannering[9] are listed in Table 1 The values are expressed as pI_{50}, where I_{50} is the concentration of the alcohol (in mM) at which the activity of the enzyme is 50% inhibited, and pI_{50} is the negative logarithm of I_{50}. More positive numbers correspond to more strongly inhibiting alcohols.

The QSAR equations consisted of a linear combination of descriptors whose coefficients were obtained by a least-squares fitting of predicted to observed biological activity through multiple linear regression.[14] A simple linear combination was chosen because when the descriptors were given an exponent that was allowed to float, the exponents would generally approach unity in the final fit. A linear description also has the advantage that algorithms such as multiple linear regression can be used to formulate the correlation. Nonlinear data correlation methods such as artificial neural networks[15] could, in principle, also be used to correlate the pI_{50} data with the electronic nose sensor array response patterns, and in some cases these methods can outperform linear methods such as the one employed in this work.[16] However, we preferred to limit the formation of our QSAR to linear methods because of the somewhat limited number of alcohols (≈20) for which pI_{50} data are available and because insight into the predictive power of the various QSAR models with different numbers of sensors or different weightings of sensor responses can then be obtained in a straightforward fashion through reference to conventional statistical metrics.

Due to the large number of combinations of n descriptors ($n \leq 19$) out of 19 possible descriptors, an efficient search procedure was required to explore the various possibilities and to identify the best fit (or, at least, the near-best fit) from the pool of all possible combinations of n descriptors for $n < 20$. The genetic function algorithm[17] of the QSAR module of Cerius[2] was used to select the best sensors for the QSAR. One hundred multiple linear regression models were generated from random combinations of four sensors. These models were ranked according to a lack-of-fit (LOF) parameter,[18] as given by eq (1).

$$\mathrm{LOF} = \frac{\mathrm{LSE}}{\left(1 - \frac{c + dp}{M}\right)^2} \quad (1)$$

LSE is the least-squares error, c and p are both the number of descriptors (sets of relative differential resistance responses of the sensors in the array) for a simple linear model such as the one herein, M is the number of samples (alcohols), and d is the 'smoothing parameter', which is entered by the user (1.0 was used). The LOF value is therefore an inverse measure of how well the model fits the data, with a penalty for the use of a large number of descriptors relative to samples.[10] From the set of 100 models, two 'parents' are chosen, with a probability inversely proportional to their LOF, and 'crossed over'—some of the descriptors from each are used to form a new model. There is then a probability for 'mutation', where a new, randomly chosen, descriptor is added to the 'daughter'. If the daughter is not already present in the population, it replaces the model with the worst LOF from the population. After 5000 rounds of genetic operation, convergence was generally reached, in which the optimal models have been found. The entire procedure was repeated three times, each with a new starting set of 100 randomly chosen four-sensor equations, and in each case the results converged to the same final sensor combination and sensor coefficients. When the 19 sets of responses from the working sensors were given to the GFA, a model that incorporated five of the sensors was found to be optimal. The best fit is described by eq (2).

$$pI_{50} = 0.51\cdot\mathbf{3} + 1.90\cdot\mathbf{9} - 3.58\cdot\mathbf{13} - 2.14\cdot\mathbf{15} - 0.90\cdot\mathbf{18} - 1.29 \quad (2)$$

$$n = 20 \quad R = 0.995 \quad s = 0.092 \quad F = 297$$

The numbers in bold refer to sets of responses from the sensors with those numbers, n is the number of samples, R is the correlation coefficient, and s is the standard error. The correlation coefficient of 0.995 indicates that the fit was quite good. The F statistic of 297 indicates that the overall significance of the fit is very high, in fact is at a level of $1–10^{-13}$. Coefficients for all sensors are significant far beyond the 99.9% level, as attested to by their t statistics (see Table 2). Predicted versus experimental pI_{50} values are plotted in Figure 2.

Methanol has an inhibition activity distinctly different from that of the other alcohols, and this can lead to a misleadingly good fit through a 'point and cluster' effect.[14] A second least-squares fitting of eq (2) was performed with the exclusion of methanol. The coefficient of **15** changed from −2.14 to −2.20, while those of the other sensors remained nearly the same. The overall quality of the fit declined; F decreased from 297 to 109, corresponding to a decrease in the signficance of the fit from the level of $1- (1\times10^{-13})$ to $1- (4\times10^{-10})$. The decreased quality of the fit occurs because methanol is modeled well by the equation, but when methanol is excluded there is much less variation in the data to be fit.

Discussion

Electronic nose-based QSAR

The selection of which molecules to include in a QSAR is crucial. It is desirable to use the broadest set of molecules available to build a QSAR, while not including only one or two molecules from a distinctly different class of compounds. In formulating a QSAR, one can

Table 1.[a]

Alcohol			Polymer name (sensor #)									
	Experimental pI50	Run# (bubbler #)	Poly(4-vinylpyridine) **(1)**	Poly(vinyl chloride) **(2)**	Poly(ethylene oxide) **(3)**	Poly(styrene/allyl alcohol) **(4)**	Poly(4-vinylphenol) **(5)**	Poly(vinyl acetate) **(6)**	Ethyl cellulose **(7)**	Poly(*N*-vinylpyrrolidone) **(8)**	Poly(ethylene/acrylic acid) **(9)**	Poly(ethylene/vinyl acetate) **(10)**
1-Butanol	−0.05	1 (8)	0.23 (0.08)	0.01 (0.15)	1.83 (0.22)	0.56 (0.08)	0.41 (0.17)	0.11 (0.08)	4.04 (0.2)	0.31 (0.7)	1.65 (0.14)	0.74 (0.1)
1-Heptanol	0.68	2 (6)	0.05 (0.1)	0.04 (0.16)	1.9 (0.1)	0.23 (0.02)	0.08 (0.11)	−0.07 (0.05)	4.28 (0.21)	0.52 (0.84)	1.66 (0.09)	0.69 (0.04)
1-Hexanol	0.54	3 (6)	0.09 (0.08)	0.03 (0.16)	1.76 (0.04)	0.45 (0.05)	0.11 (0.09)	−0.08 (0.05)	5.32 (0.18)	−0.07 (0.6)	1.84 (0.05)	0.73 (0.02)
1-Pentanol	0.27	3 (7)	0.17 (0.1)	−0.03 (0.13)	1.63 (0.02)	0.58 (0.05)	0.16 (0.1)	0.02 (0.04)	4.97 (0.14)	0.67 (0.58)	1.7 (0.03)	0.67 (0.01)
1-Propanol	−0.48	3 (3)	0.55 (0.14)	−0.03 (0.19)	1.18 (0.01)	0.57 (0.04)	0.7 (0.09)	0.2 (0.03)	3.17 (0.14)	1.08 (0.82)	1.07 (0.02)	0.4 (0.02)
2,4-Dimethyl-3-pentanol	−1.38	2 (1)	0.05 (0.14)	−0.02 (0.13)	2 (0.06)	0.17 (0.02)	0.1 (0.11)	0 (0.03)	4.94 (0.25)	0.11 (0.61)	2.49 (0.05)	2.42 (0.09)
2-Butanol	−0.35	2 (8)	0.2 (0.13)	−0.06 (0.12)	1.35 (0.04)	0.65 (0.1)	0.29 (0.23)	0.14 (0.05)	3.89 (0.29)	0.23 (0.68)	1.62 (0.03)	0.76 (0.04)
2-Heptanol	0.25	1 (2)	0.13 (0.08)	−0.04 (0.09)	2.89 (0.54)	0.28 (0.04)	0.11 (0.13)	−0.09 (0.05)	4.76 (0.18)	0.23 (0.99)	1.91 (0.06)	0.99 (0.04)
2-Hexanol	0.15	2 (2)	0.16 (0.15)	0.01 (0.15)	1.69 (0.09)	0.37 (0.03)	0.17 (0.11)	−0.05 (0.03)	5.1 (0.31)	0.77 (0.42)	1.97 (0.05)	0.92 (0.03)
2-Methyl-1-butanol	−0.15	2 (7)	0.04 (0.11)	0.02 (0.12)	1.74 (0.05)	0.34 (0.04)	0.16 (0.14)	−0.01 (0.04)	4.76 (0.27)	0.09 (0.73)	1.82 (0.04)	0.91 (0.03)
2-Methyl-1-propanol	−0.39	1 (6)	0.12 (0.07)	0.01 (0.12)	1.84 (0.15)	0.48 (0.08)	0.28 (0.14)	0.1 (0.03)	3.98 (0.2)	0.47 (0.59)	1.65 (0.1)	0.78 (0.08)
2-Methyl-3-pentanol	−0.89	1 (1)	0.13 (0.09)	0.06 (0.11)	2.34 (0.33)	0.29 (0.05)	0.19 (0.11)	−0.03 (0.04)	5.24 (0.26)	0.76 (0.61)	2.17 (0.02)	1.59 (0.04)
2-Pentanol	−0.07	3 (8)	0.06 (0.06)	−0.03 (0.13)	1.41 (0.02)	0.57 (0.07)	0.14 (0.09)	0 (0.07)	4.9 (0.19)	0.68 (0.55)	1.77 (0.02)	0.82 (0.02)
2-Propanol	−0.47	1 (7)	0.24 (0.08)	0.13 (0.14)	1.58 (0.23)	0.57 (0.06)	0.62 (0.17)	0.14 (0.06)	3.31 (0.31)	0.4 (0.88)	1.45 (0.07)	0.63 (0.04)
3-Hexanol	−0.47	3 (1)	0.07 (0.08)	0.01 (0.13)	1.57 (0.03)	0.4 (0.04)	0.07 (0.08)	−0.06 (0.05)	5.56 (0.23)	0.02 (1.03)	1.81 (0.04)	1.07 (0.01)
3-Methyl-1-butanol	−0.19	3 (5)	0.08 (0.08)	0.03 (0.08)	1.49 (0.02)	0.39 (0.03)	0.07 (0.07)	0.01 (0.04)	4.82 (0.13)	0.08 (0.83)	1.77 (0.02)	0.75 (0.02)
3-Pentanol	−0.37	2 (4)	0.11 (0.09)	0.08 (0.15)	1.52 (0.04)	0.55 (0.06)	0.16 (0.13)	−0.01 (0.05)	4.83 (0.41)	−0.13 (0.81)	1.79 (0.03)	1.03 (0.04)
Ethanol	−1.1	2 (3)	1.52 (0.15)	0.19 (0.14)	1.08 (0.02)	0.59 (0.05)	2.19 (0.17)	0.31 (0.03)	2.19 (0.14)	4.03 (0.74)	0.78 (0.03)	0.2 (0.04)
Methanol	−3.09	1 (3)	3.71 (0.23)	0.57 (0.12)	1.33 (0.1)	0.55 (0.03)	2.51 (0.21)	0.4 (0.07)	1.82 (0.22)	7.76 (0.78)	0.69 (0.03)	0.15 (0.04)
Neopentanol (solid)	−0.67	3 (2)	0.03 (0.1)	0 (0.18)	1.37 (0.04)	0.14 (0.03)	−0.01 (0.05)	0.02 (0.03)	3.28 (0.2)	−0.13 (0.79)	1.54 (0.05)	0.94 (0.03)
Benzyl alcohol	0.32	1 (4)	0.06 (0.07)	0.04 (0.13)	3.05 (0.91)	0.22 (0.03)	0.1 (0.07)	−0.03 (0.05)	2.07 (1.01)	−0.1 (0.59)	0.58 (0.34)	0.33 (0.17)
tert-Amyl alcohol	−2.56	1 (5)	0.1 (0.1)	−0.07 (0.14)	1.77 (0.24)	0.39 (0.07)	0.26 (0.14)	0.06 (0.05)	3.91 (0.29)	0.35 (0.62)	2.05 (0.12)	1.04 (0.08)
1,3-Propanediol	−1.87	3 (4)	−0.02 (0.1)	0.04 (0.12)	0.17 (0.02)	0.06 (0.02)	−0.01 (0.05)	0.01 (0.03)	0.4 (0.19)	−0.39 (0.8)	0.06 (0.01)	0.02 (0.02)
1,4-Butanediol	−1.41	2 (5)	−0.01 (0.09)	−0.01 (0.15)	0.19 (0.2)	0.06 (0.06)	0.04 (0.06)	−0.02 (0.04)	0.81 (0.68)	−0.09 (0.79)	0.14 (0.15)	0.05 (0.05)

[a]First three columns give the name of the alcohol, its experimental pI_{50} value,[4] and this run in which it was analyzed (and the bubbler in which it was placed). The remainder of the table lists the responses (expressed as % change in electrical resistance relative to baseline resistance) of the 19 different polymer/carbon black sensors upon exposure to the alcohols at 5% of their respective saturated vapor pressures. The standard deviation of the responses over 10 trials are given in parentheses. Sensor **14** was not functioning. The last four alcohols were not used in building the model.

either try to include molecules that encompass a large range of variation in many properties in an attempt to obtain descriptors that will have predictive value for essentially any molecule of interest, or one can choose a more restricted set of molecules in which many factors are kept constant but only a few properties are varied in an attempt to make predictions that are designed to describe the response to specific changes in a test set of molecules. We have focused this study on the latter approach, and attempted only to probe whether the steric and other chemical differences within a test set of aliphatic alcohols that inhibit cytochrome P-450 hydroxylation could be correlated successfully with the differential sorption signatures of these alcohols on an array of conducting polymer composite electronic nose sensors. This approach was driven not only by the fact that extensive biological inhibition data in this system is currently only available for alcoholic inhibitors, but also by the observation that even in this restricted test set of molecules, log *P*, either alone or in combination with one or two other physicochemical features of the alcohols, is not sufficient to provide satisfactory descriptions of the relative cytochrome P-450 inhibition activity of alcohols in this test set. Further descriptors that involve steric or other chemical attributes of the alcohols are therefore clearly needed in order to capture the experimentally observed variation in pI_{50}, and the goal was to evaluate whether such descriptors were contained in the electronic nose response data.

Prior to building the model, it was noted that benzyl alcohol, the only aromatic alcohol in the data set, has a higher activity than is predicted by both our QSAR and another QSAR on the cytochrome P-450 system.[13] The anomalous activity of benzyl alcohol could be accounted for with an additional descriptor unique to benzyl alcohol, but the choice of such a parameter is rather arbitrary, so benzyl alcohol was excluded during the building of our QSAR. *tert*-Amyl alcohol was also excluded because there is evidence that tertiary alcohols function through a stimulatory mechanism in addition to the usual inhibitory mechanism.[9] As would be expected if *tert*-amyl alcohol were also acting through this stimulatory mechanism, its inhibitory activity is anomalously low. The two diols were also excluded while building the model. Because of these limitations, our QSAR is expected to be most successful at, and perhaps limited to, predicting the activity of aliphatic

Table 1 contd.

Alcohol	Polymer name (sensor #)								
	Poly(methyl methacrylate) (**11**)	Poly(methylvinylether/maleic anhydride)(**12**)	1,2-Polybutadiene (**13**)	Poly(styrene/acrylonitrile) (**14**)	Poly(methyloctadecylsiloxane) (**16**)	Poly(vinyl butyral) (**17**)	Poly(ethylene glycol) (**18**)	Poly(2,4,6-tribromostyrene) (**19**)	Polystyrene (**20**)
1-Butanol	0 (0.02)	−0.01 (0.01)	0.23 (0.1)	0 (0)	0.42 (0.02)	1.14 (0.27)	2.37 (0.25)	0.12 (0.06)	−0.46 (0.73)
1-Heptanol	0.01 (0.03)	0 (0.01)	0.3 (0.02)	0 (0)	0.41 (0.03)	0.45 (0.22)	1.23 (0.14)	0.01 (0.02)	−0.01 (0.82)
1-Hexanol	0.01 (0.03)	−0.01 (0.02)	0.28 (0.01)	0 (0)	0.49 (0.02)	0.89 (0.14)	1.79 (0.08)	0.04 (0.03)	0.23 (0.94)
1-Pentanol	0.01 (0.02)	0 (0.02)	0.21 (0.01)	0.01 (0)	0.46 (0.02)	1.04 (0.18)	1.95 (0.06)	0.07 (0.06)	−0.21 (0.8)
1-Propanol	0.01 (0.03)	−0.02 (0.02)	−0.03 (0.01)	0.02 (0)	0.28 (0.02)	1.51 (0.2)	2.12 (0.11)	0.25 (0.04)	0.47 (0.59)
2,4-Dimethyl-3-pentanol	0.02 (0.02)	0.03 (0.02)	1.19 (0.04)	0 (0.01)	0.72 (0.02)	0.54 (0.21)	1.85 (0.04)	0.01 (0.02)	−0.29 (0.69)
2-Butanol	0 (0.03)	−0.02 (0.01)	0.26 (0.01)	0 (0)	0.4 (0.02)	1.03 (0.23)	1.95 (0.09)	0.13 (0.02)	0.17 (0.32)
2-Heptanol	0 (0.03)	0.01 (0.03)	0.45 (0.02)	0 (0)	0.49 (0.03)	0.62 (0.19)	2.13 (0.49)	0.03 (0.01)	−0.03 (0.32)
2-Hexanol	−0.01 (0.02)	0.1 (0.02)	0.45 (0.01)	0.01 (0)	0.51 (0.01)	0.77 (0.21)	1.73 (0.11)	0.03 (0.02)	0.08 (0.76)
2-Methyl-1-butanol	−0.01 (0.02)	−0.01 (0.01)	0.41 (0.01)	0 (0.01)	0.45 (0.02)	0.77 (0.24)	2.08 (0.11)	0.03 (0.03)	0.15 (0.84)
2-Methyl-1-propanol	0 (0.02)	−0.04 (0.02)	0.28 (0.07)	0 (0)	0.41 (0.02)	1.01 (0.21)	2.41 (0.19)	0.09 (0.05)	0.2 (0.5)
2-Methyl-3-pentanol	0 (0.03)	0.08 (0.02)	0.75 (0.02)	0.01 (0)	0.59 (0.02)	0.7 (0.2)	2.34 (0.25)	0.03 (0.03)	−0.09 (0.84)
2-Pentanol	−0.01 (0.02)	0 (0.01)	0.34 (0.01)	0 (0)	0.45 (0.02)	1.03 (0.28)	1.85 (0.07)	0.07 (0.05)	−0.12 (0.79)
2-Propanol	0 (0.02)	−0.04 (0.03)	0.16 (0.05)	0 (0)	0.36 (0.03)	1.15 (0.17)	2.34 (0.29)	0.14 (0.04)	−0.04 (0.47)
3-Hexanol	0.01 (0.02)	0.01 (0.02)	0.51 (0.01)	0 (0)	0.53 (0.02)	0.87 (0.2)	1.63 (0.1)	0.06 (0.08)	−0.12 (0.74)
3-Methyl-1-butanol	−0.01 (0.04)	−0.04 (0.02)	0.36 (0.01)	0 (0)	0.42 (0.02)	0.9 (0.12)	1.85 (0.06)	0.03 (0.03)	−0.09 (0.73)
3-Pentanol	−0.01 (0.01)	−0.01 (0.02)	0.43 (0.01)	0 (0)	0.5 (0.03)	0.96 (0.31)	1.85 (0.1)	0.07 (0.04)	0.14 (0.71)
Ethanol	0.14 (0.03)	−0.03 (0.02)	−0.13 (0.03)	0.17 (0.01)	0.23 (0.02)	1.44 (0.25)	2.14 (0.1)	0.42 (0.04)	−0.11 (0.63)
Methanol	0.57 (0.03)	0.52 (0.05)	−0.01 (0.01)	0.62 (0.03)	0.21 (0.02)	1.58 (0.25)	2.78 (0.2)	0.27 (0.03)	0.11 (0.7)
Neopentanol (solid)	0 (0.03)	0 (0.02)	0.42 (0.01)	0 (0)	0.34 (0.02)	0.39 (0.2)	1.75 (0.07)	0.01 (0.03)	−0.21 (0.56)
Benzyl alcohol	−0.01 (0.02)	−0.04 (0.02)	0.11 (0.08)	0 (0)	0.17 (0.11)	0.34 (0.26)	1.36 (0.48)	0.01 (0.04)	−0.5 (0.76)
tert-Amyl alcohol	0 (0.02)	−0.03 (0.02)	0.47 (0.06)	0 (0)	0.46 (0.01)	0.74 (0.16)	2.26 (0.23)	0.08 (0.03)	0.09 (0.48)
1,3-Propanediol	−0.01 (0.04)	−0.03 (0.02)	0.02 (0.01)	0 (0)	0.01 (0.02)	0.09 (0.18)	0.09 (0.09)	0 (0.13)	0.04 (0.78)
1,4-Butanediol	−0.01 (0.02)	−0.03 (0.02)	0.03 (0.02)	0 (0.01)	0.04 (0.04)	0 (0.14)	0.13 (0.13)	0.01 (0.03)	−0.32 (0.87)

mono-alcohols having no other functionalities. We did not attempt to describe features of other functional groups in the QSAR, nor to evaluate the properties of other enzyme binding sites, so the results and QSAR modeling discussed herein are clearly limited to the alcohol/cytochrome P-450 system.

The sensors chosen for the model by the GFA are among those whose responses are most reproducible. Reproducibility was measured by examining the set of 10 responses of a given sensor to a given analyte. The value $S_{i,j}$ is defined as the standard deviation among the 10 responses of the jth sensor to the ith alcohol divided by the average of those responses. Each sensor has a set of 20 S values, one for each alcohol. A sensor's reproducibility can be gauged by the median of its set of S values. Four of the five sensors used in the model displayed median S values less than 0.063, ranking them among the best seven sensors. The only sensor outside this group, **15**, responded only to very polar analytes. Since its response to the majority of the analytes was quite small, its S value for those analytes is very large. However, for the analytes to which it did respond, for example methanol and ethanol, its S values are small, 0.040 and 0.041, respectively. The inclusion of **15** might be questioned if it were necessary only to model the activity of one analyte, namely the outlier methanol. To test the validity of including **15** in the QSAR, eq (2) was refit with the same set of sensors and all of the previously used alcohols, excluding methanol. In the new QSAR, the significance of **15**, as indicated by its t value, diminished to 98%. The decrease is large, but **15** remains significant. If the set of five sensor responses to methanol are substituted into the second QSAR equation, which was formed with no information about methanol, the predicted pI_{50} of methanol is −3.12, which is very close to its experimental value of −3.09. It appears that whatever molecular characteristics are probed by **15** are successfully extrapolated from the more moderately polar analytes to methanol. In other words, **15** is not just an indicator variable for methanol that is fit with an arbitrary coefficient.

A quantitative measure of the predictive power of the QSAR can be obtained by building a model using the biological and sensor response data from all the molecules except one, and then predicting the activity of the excluded molecule with that model. The procedure is

Table 2. Regression statistics for the coefficents of eq (2)[a]

	Coefficient	Standard error	*t* Statistic	*P*-value
Intercept	−1.29	0.27	−4.71	3.32E-04
3	0.51	0.07	6.93	6.98E-06
9	1.90	0.19	9.92	1.03E-07
13	−3.58	0.21	−17.13	8.70E-11
15	−2.14	0.27	−7.91	1.56E-06
18	−0.90	0.08	−11.34	1.94E-08

[a]The *t* statistics is equal to the value of the coefficient divided by its standard error; it is used to derive the *P* value, which indicates the significance of the coefficient.

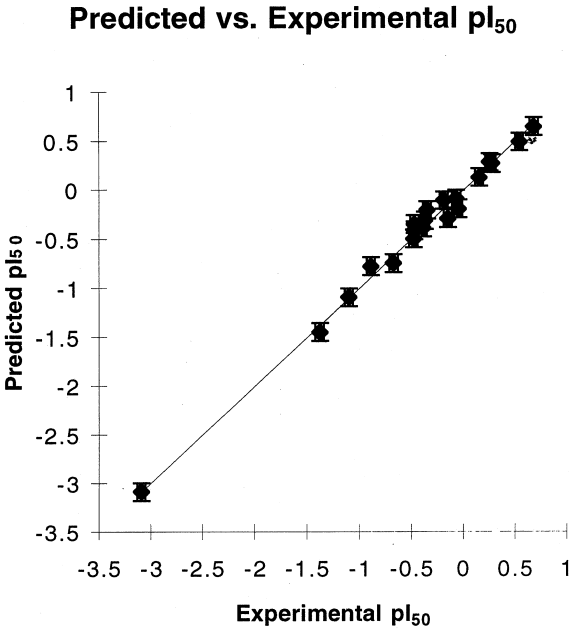

Figure 2. Plot of pI_{50} predicted by eq (2) versus the actual experimental values. Horizontal error bars represent an average experimental error as reported in the original paper,[4] and vertical error bars correspond to the standard error of eq (2). The line represents perfect agreement between experiment and prediction.

repeated for each molecule in the data set, and the predictive sum of squares (PRESS) is defined as the sum, over all analytes, of the squared differences between the predicted and actual biological activities. Using eq (2), the PRESS for the set of 20 alcohols is 0.221. This value can be compared to the residual sum of squares, RSS, in which one QSAR equation (fit to all samples) is used to calculate the predicted activities. As would be expected, the RSS of 0.117 is lower than the PRESS. More significantly, a large difference between the PRESS and RSS would imply that the model had used too many parameters and overfit the data, and this appears not to be the case.

An optimum fit (as judged by the LOF parameter) was found to require five descriptors; no equation with a different number of descriptors formed as significant a model. The best 1, 2, 3, 4, 5, and 6 sensor fits are given in Table 3. As expected, the correlation coefficient increased with each added descriptor. However, the five-sensor model produced a minimum in the PRESS and a maximum in the significance of the equation. The best four-sensor QSAR, consisting of sensors **1**, **13**, **16** and **17**, has an $R = 0.984$, $s = 0.163$, and $F = 114$, indicating an overall significance at the level of $1-(5\times10^{-11})$. The PRESS for this model had a value of 0.991, so addition of the fifth descriptor clearly enhanced the predictive ability of the QSAR. However, if **4** is added to eq (2) to form the best six-sensor equation, certain key statistics point to a diminished model relative to the five-sensor fit. As would be expected with an additional parameter, R increases, from 0.995 to 0.996. Additionally, the standard error decreases from 0.0916 to 0.0834, the RSS decreases from 0.117 to 0.090, and the F statistic increases from 297 to 300. However, the significance of the fit, represented by the F statistic, decreases from $1-(1.08\times10^{-13})$ to $1-(3.66\times10^{-13})$. The PRESS increases from 0.221 to 0.253. Thus, although the six-sensor model fits the set of 20 alcohols better than the five-sensor model, the six-sensor model is worse at predicting the activity of an alcohol that was not included in the fit, indicating that the six-sensor model has overfit the data.

As described above, the cytochrome P-450 *p*-hydroxylation inhibition activities of all the aliphatic monoalcohols investigated in this work could be quite accurately predicted from a model that was constructed without the use of any information about the molecular structure of the alcohols for which the predictions are made. This indicates that the resistance data output of the electronic nose contains implicit information on most of the chemical factors that control the interactions of the enzyme with the alcohols. These resistance data reflect the binding interactions between the alcohols and a collection of polymers having a diverse collection of chemical attributes. Selection of the polymers used in the current version of the electronic nose is based upon experience in our laboratory and in other laboratories[19,20] aimed at obtaining a set of diverse polymers having a broad distribution of chemical properties, such as polarity, hydrogen bonding interactions, hydrophobicity, etc., that are capable of probing a variety of both physicochemical (log *P*, etc.) and stereochemical (isomeric, enantiomeric) features of analytes. The differential sorption properties of analytes into these polymers therefore clearly contain chemical information on relatively subtle differences between analytes, and allow classification of such analytes and correlation of their electronic nose sensor patterns with other physicochemical properties, such as pI_{50} values in this initial test case. However, the set of polymers that we have used in this initial test study is not unique nor exhaustive, and it is certainly possible that improvements in the diversity of the polymers in the sensor array, and/or inclusion of a more exhaustive set of polymers that encompasses still further types of interactions with the analytes of concern, would provide yet improved performance in formulating a QSAR model for the alcohol/cytochrome P-450 system. This would be especially true if certain aspects of the binding to the active site in the system being evaluated had not been included in the conducting polymer composite sensors used to formulate the QSAR model. For example, enantiomerically pure polymers would certainly be required if the

Table 3. The best QSAR equations using 1–6 sensors and their regression statistics

# of sensors	Equation	*R*	*S*	*F*	Significance	PRESS
1	−0.105 − 0.343(**8**)	0.777	0.526	27	5.58E-05	7.73
2	0.706−1.01(**1**)−0.847(**10**)	0.917	0.342	45	1.57E-07	2.66
3	0.599 − 4.72(**6**) − 1.87(**13**) − 2.91(**15**)	0.969	0.22	81	6.72E-10	1.65
4	−0.473 − 0.547(**1**) − 4.86(**13**) + 7.64(**16**) − 1.46(**17**)	0.984	0.163	114	4.75E-11	0.991
5	−1.29 + 0.511(**3**) + 1.90(**9**)−3.58(**13**)−2.14(**15**)−0.901(**18**)	0.995	0.092	297	1.08E-13	0.221
6	−1.39 + 0.401(**3**) − 0.486(**4**) + 2.14(**9**) − 3.93(**13**) − 2.06(**15**) − 0.791(**18**)	0.996	0.083	300	3.66E-13	0.253

binding site under evaluation were chiral and the analytes were enantiomerically pure. The present diversity of polymers seems to be adequate for the present study, because the resulting QSAR model captures most of the variance in the cytochrome P-450/alcohol pI_{50} data and the model has a good predictive value for new members of this test set of aliphatic alcohols. The main point is that it is not necessary that an individual polymer probe specifically and exclusively one such descriptor of the analyte–substrate interaction, because the desired information can apparently be obtained through analysis of the collective response of an array of broadly responsive sensors to the analytes.

Comparison with other QSARs

Cohen and Mannering fit the activity of 11 of the unbranched 1- and 2-alcohols (excluding methanol) to a one parameter equation using log *P*.[9] A modified version, using updated log *P* values and fit to only 10 alcohols (excluding methanol and ethanol), was given later by Shusterman (eq (3)).[13]

$$pI_{50} = 0.43 \log P - 0.53$$
$$n = 0 \quad R = 0.954 \quad s = 0.128 \tag{3}$$

However, Shusterman also showed[13] that, for a larger set of alcohols, a simple fit to log *P* was inadequate to describe most of their activity; a fit of 19 of the alcohols yielded eq (4), which has rather poor regression statistics.

$$pI_{50} = 0.35 \log P - 0.71$$
$$n = 19 \quad R = 0.505 \quad s = 0.468 \tag{4}$$

In a second equation using two descriptors, log *P* and $(\log P)^2$, Cohen and Mannering fit 17 of the alcohols with an *R* of 0.98 (eq. (5)).[9]

$$pI_{50} = 1.50 \log P - 0.36 (\log P)^2 + 1.75$$
$$n = 17 \quad R = 0.98 \quad s = 0.44 \tag{5}$$

Although this was a better fit, it used more descriptors. Additionally, it is evident from inspection of the data that there are factors besides hydrophobicity that determine an alcohol's activity. Four subsequent QSARs have therefore been used to model the data set more fully,[10–13] and some aspects of these models are discussed below.

A more complex, three-parameter QSAR was based upon log *P*, a calculated electronic parameter (ϵ_{HOMO}), and a steric parameter ($BULK_{lat}$) (eq. (6)).[10]

$$pI_{50} = 16.2 \log P - 16.0 \log (\beta P + 1) - 1.35 \tag{6}$$
$$BULK_{lat} + 0.381\, \varepsilon_{HOMO} + 22.5$$
$$n = 21 \quad R = 0.982 \quad s = 0.170 \quad \log \beta = 1.05$$

Shusterman and Johnson,[13] however, pointed out that the use of ε_{HOMO} as a parameter was unjustified since it was necessary only to fit benzyl alcohol, and becomes an insignificant parameter (as indicated by its *t* value) when benzyl alcohol is excluded from the data set. Similarly, these authors showed[13] that the bilinear dependence of pI_{50} upon log *P* of eq (6) was necessary only to fit a single data point, methanol.

Another QSAR, based on a choice of molecular connectivity indices,[21] has also been used to model the activity of 20 of the alcohols (benzyl alcohol and *tert*-amyl alcohol were excluded) (eq (7)).[11]

$$pI_{50} = -6.88\left(1/{}^{0}\chi^{v}\right) - 1.14\left({}^{4}\chi_{PC}\right) + 1.85 \tag{7}$$
$$n = 20 \quad R = 0.983 \quad s = 0.156$$

The parameter ${}^{0}\chi^{v}$, the zero-order valence molecular connectivity index, basically corresponds to molecular size, and therefore hydrophobicity, for this set of molecules. Hence, the inverse of the index has a negative coefficient in eq (7). The parameter ${}^{4}\chi_{PC}$, the fourth-order path/cluster molecular connectivity index, correlates with the degree of branching in the molecule, and therefore also has a negative coefficient in eq (7).

A third QSAR, which relies entirely upon calculated electronic parameters as descriptors, has been constructed and used to fit all 22 alcohols.[12] Shusterman[13] noted problems with the QSAR. For example, it was asserted that the α-carbon of the alcohols was acting as an electron acceptor from the enzyme, because a correlation between activity and Q_CL, the electron density on the α-carbon in the LUMO, was found. Q_CL is correlated with log *P* ($R = 0.747$), to some extent explaining the fit. Two alcohols, 3-methyl-1-butanol and 2,4-dimethyl-3-pentanol, were poorly fit, and no rationalization was presented for why the correlation with Q_CL would not apply to these two substrates as well.[12]

Finally, Shusterman created a QSAR based on log *P* and two steric parameters, *M* and *A*, which were used to describe the branching of the alcohols.[13] *M* is the number of carbons beyond the methyl substituent in Figure 3; thus, 1- and 2-alcohols have an $M = 0$, while *M* for 3-pentanol would be 1, and *M* for 2,4-dimethyl-3-pentanol

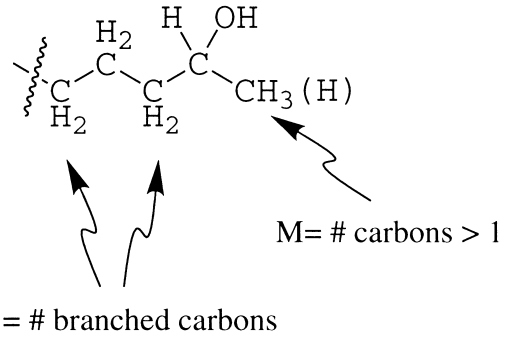

Figure 3. Diagram illustrating the M and A steric parameters of Shusterman and Johnson.[9]

is 2. The second parameter, A, refers to the number of branched carbons in the main chain; $A = 1$ for 2-methyl-1-butanol and 2 for neopentyl alcohol. A fit of 19 of the alcohols (benzyl alcohol, *tert*-amyl alcohol, and methanol were excluded) yielded eq (8). The negative coefficients for M and A indicate the loss of activity with branching.

$$pI_{50} = 0.48 \log P - 0.65M - 0.31A - 0.60 \quad (8)$$

$$n = 19 \quad R = 0.955 \quad s = 0.71$$

To compare the electronic nose QSAR to those of Sabljic[11] and Shusterman,[13] one must use statistics that take into account the number of descriptors used. Table 4 lists some appropriate statistics for comparison. Because the electronic nose QSAR model uses more parameters, it is inappropriate to compare just either the correlation coefficients, standard error, or residual sum of squares of the models. To some extent, the PRESS should be independent of the number of parameters in a model, since the model is tested upon molecules about which it has no information. The PRESS of the electronic nose QSAR model is significantly lower than the other two models of interest. Finally, the F statistic gauges the overall significance of the fit while accounting for the number of parameters used. By this measure, the electronic nose QSAR is approximately as significant as Sabljic's[11] and more significant than Shusterman's.[13]

We also performed our own QSAR modeling based on calculated parameters. A large number of descriptors were calculated for the 20 molecules of this study using Cerius[2]. These descriptors were obtained for a specific, energy-minimized molecular geometry. The results of the best 1–5 descriptor equations are given in Table 5. Perhaps because of the number of descriptors (>100) from which the utilized ones were chosen, very good QSARs were discovered. The best three-descriptor equation yielded a fit similar to that of the five-sensor equation (eq (2)). The 'charged partial surface area' descriptors of Jurs[22] are strikingly common in this set of equations. This is not unexpected because these descriptors were developed specifically to better describe polar intermolecular interactions, which are presumably important in the interaction of the alcohols with the binding site of cytochrome P-450.

The goal of this work was not to compare the performance of calculated descriptors versus experimentally-obtained ones, to decide which approach would be superior on the test system of concern, but instead to ascertain whether it would be possible to describe successfully the chemical binding interactions involved in an enzyne inhibition event from the measured responses of a broadly-responsive array of sensors. The results clearly imply that the important chemical interactions involved in the partitioning of the aliphatic alcohols into the enzyme binding site are probed by the array responses. If only physicochemical factors such as log P were dominant in determining the pI_{50} values, then the earlier QSAR models would not require inclusion of additional parameters to satisfactorily model the variation in cytochrome P-450 alcohol inhibition constants. Furthermore, the differences between response patterns of these alcohols on the electronic nose sensor array seem to be successfully capturing these other chemically based factors, otherwise it would not be possible to correlate the differential response patterns with the experimentally observed inhibition data.

Obtaining chemical insight into the nature of the dominant binding forces involved in the reaction being modeled would require a complete understanding of the chemical factors that determine the analyte partitioning into each polymer in the electronic nose. It is possible to choose descriptors for each analyte/polymer sorption event and to project these descriptors onto the sensor response data, thereby obtaining an indication of the fractional contribution of each chosen descriptor to the overall QSAR model. For example, each sensor could be characterized in terms of its response to analyte polarity, surface area, volume, branching factor, hydrogen bonding properties, etc., and the weighted sensor responses in the resulting cytochrome P-450 pI_{50} QSAR model could be used to evaluate the contribution of

Table 4. Comparison of selected regression statistics from the QSAR of Sabljic,[6] Shusterman,[8] (eq (2)), and the QSAR created when the coefficients of eq (2) were fit to the 19 alcohols beside methanol[a]

	Data pts fit	Descriptors used	R	s	RSS	F	Significance	PRESS
Sabljic[11]	20	2	0.983	0.156	0.414	250	2.51E-13	0.872
Shusterman[13]	19	3	0.956	0.17	0.436	53	3.34E-08	0.786
Eq (2)	20	5	0.995	0.092	0.117	297	1.08E-13	0.221
Eq (2), no methanol	19	5	0.988	0.095	0.117	109	3.89E-10	0.243

[a]R is the correlation coefficient, s is the standard error, and RSS and PRESS are explained in the text. The final column is the overall significance of the regression equation.

Table 5. The best QSAR equations using 1–5 calculated descriptors from Cerius[2] and their regression statistics[a,b]

# of descriptors	Equation	*R*	*s*	*F*	Significance	PRESS
1	19.60 − 2.98(LUMO)	0.855	0.433	49	1.55 E-06	5.73
2	3.07 − 0.0908(PMI-X) − 11.7(Jurs-RPCG)	0.974	0.194	159	9.85 E-12	0.97
3	0.808 − 1.52(Jurs-WNSA-3) − 23.8(Jurs-FNSA-1) − 0.0747(V-ADJ-mag)	0.994	0.093	479	7.18E-16	0.211
4	4.57 − 0.130(MolRef) + 27.6(Jurs-FNSA-2) − 1.71(Jurs-WNSA-3) −8.52(Jurs-RNCG)	0.998	0.057	962	7.01E-18	0.083
5	−3.92 − 0.154(Jurs-RNCS) − 1.89(Jurs-WNSA-3) + 0.0764(Jurs-PNSA-2) −0.519(Foct) − 0.590(CHI-V-O)	0.999	0.048	1076	1.41E-17	0.061

[a]For further information and references pertaining to the descriptors, see ref 23.
[b]CHI-V-O, zero order molecular connectivity index; Foct, 1-octanol desolvation free energy; Jurs-FNSA-1, fractional charge weighted negative surface area; Jurs-FNSA-2, fractional atomic charge weighted positive surface area; Jurs-PNSA-2, total charge weighted negative surface area; Jurs-RNCG, relative negative charge; Jurs-RNCS, relative negative surface area; Jurs-RPCG, relative positive charge; Jurs-WNSA-3, surface weighted atomic charge negative surface area; LUMO, LUMO energy (eV) calculated at the CNDO/2 level; MolRef, molar refractivity; PMI-X, principal moment of inertia along *x*-axis; V-ADJ-mag, vertex adjacency/magnitude.

each of these descriptors to the correlation model of the data. However, using such a process itself requires a choice of descriptors, and is in essence similar to the modeling that has been done previously based on purely computational QSAR formulations of the pI_{50} data in the test system of concern. The electronic nose approach is a 'black-box' in the sense that one need not choose descriptors nor make assumptions regarding which computational parameters to use to capture the variation in the binding constant data, provided that the array of polymers is diverse enough that it can capture the important chemical effects in its response data, as is apparently the case for this test system with the present sensor array.

The present work therefore suggests that, at least in certain cases, a library of electronic nose data might be useful in predicting binding affinities for members of a library of molecules towards a number of different binding sites, once sufficient data points have been collected to construct a statistically significant predictive model for the binding site/analyte set of concern. The gas/polymer partitioning that produces the electronic nose sensor response patterns likely yields more direct information on physicochemical properties such as log *P*, etc., so that properties such as passive transmembrane diffusion, bioavailability, or even binding to albumin (where the binding site is hydrophobic and relatively indiscriminate), where the correlations at present are empirical and non-structure based and predominantly reflect the sorption properties of the species of concern, might be readily correlated with the electronic nose sensor response data and might ultimately represent areas in which correlations of biological activity with the electronic nose signatures would be the most valuable. The present work suggests that, in certain cases, even more subtle steric/chemical interaction data is available from the differential sorption pattern data produced by the electronic nose and that, at least in this specific initial test case, such response data can be correlated with the inhibition properties of a test set of alcohols in the cytochrome P-450 hydroxylation process, which probes to some extent the steric factors of the inhibitor. It remains to be evaluated whether such correlations can be extended to other systems that display yet more discriminating steric binding selectivity. The limitations of this method will of course depend on the breadth of the class members represented in the library, the interactions that are probed by the polymers chosen for the electronic nose array, and the specificity of the binding interactions. Studies to evaluate the tradeoffs between these various factors in other scenarios of interest are underway at present.

Conclusions

We have created a QSAR based on the relative differential resistance responses of an array of conducting polymer composites that models the inhibition of cytochrome P-450 *p*-hydroxylation of aniline by a set of aliphatic alcohols. The QSAR is at least as successful as two previously published QSARs, having a correlation coefficient of 0.995, an *F* statistic of 297, and an overall statistical significance of the fit at a level of $1–10^{-13}$. Coefficients for all sensors in the model are significant far beyond the 99.9% level. The electronic nose QSAR model is of interest because it implies that the different response patterns of the various alcohols on the electronic nose are apparently sufficiently sensitive to steric and other molecular factors to allow correlation with the variation of inhibition activities for this particular enzymatic reaction.

Experimental

Materials

Poly(4-vinylpyridine), poly(vinylchloride), poly(stryrene-*co*-allyl alcohol), poly(vinylacetate), ethyl cellulose, poly(ethylene-*co*-acrylic acid) (15% acrylic acid), 1,2-poly(butadiene), poly(butadiene) (36% *cis* and 55% *trans* 1–4), poly(methyloctadecylsiloxane), poly(2,4,6-tribromostyrene), and poly(styrene-*co*-acrylonitrile) were purchased from Scientific Polymer Products. Poly(4-vinylphenol), poly(methylvinylether-*co*-maleic anhydride), poly(vinylbutyral), and poly(ethylene glycol) were purchased from Polysciences. Poly(ethylene oxide), poly (ethylene-*co*-vinyl acetate) (18% vinyl acetate), poly(vinylpyrrolidone), poly(styrene), and poly(methylmethacrylate) were purchased from Aldrich. The carbon

black was Black Pearls 2000 from Cabot Corporation. Alcohols were purchased from Aldrich, except 2-methyl-3-pentanol (Lancaster), methanol and 2-propanol (EM Science), ethanol (Quantum), and 3-methyl-1-butanol (Acros).

Sensors and instrumentation

Polymers were generally dissolved in tetrahydrofuran, except for poly(4-vinylpyridine) and poly(vinylpyrrolidone), which were dissolved in ethanol, and poly (ethylene-*co*-vinyl acetate) (18% vinyl acetate), 1,2-poly(butadiene), and poly(butadiene) (36% *cis* and 55% *trans* 1–4), which were dissolved in toluene. Each polymer (160 mg) was dissolved in its respective solvent (20 mL) either at room temperature or by heating to 35–40 °C for several hours. Carbon black (40 mg) was added and the suspension sonicated for at least 20 min.

Corning microscope slides were cut into 10×25 mm pieces to provide substrates for the sensors. A 7–8 mm gap across the middle of each piece was masked while 300 nm of chromium and then 500 nm of gold was evaporated onto the ends of the slides to form the electrical contacts. Sensors were formed by spin-coating polymer/carbon black suspensions onto the prepared substrates. The resulting films were then allowed to dry overnight.

Measurements

The instrumentation and apparatus for resistance measurements and for the delivery of vapors have been described previously.[5] Eight bubblers for generation of vapors were available, so the 22 alcohols and two diols were divided into three groups of eight, as indicated in Table 1. To pre-condition the sensors, prior to each of the three runs, the sensors were subjected to 40 exposures, five to each of the eight analytes. Data collection then consisted of a set of 10 exposures to the eight analytes, with the 80 exposures performed in randomized order to eliminate systematic errors from any history effects. In the third run, bubbler 2 was replaced by a Pyrex tube 37 cm in length with a 1 cm inner diameter. This tube was loaded with ~25 cm of grannular, solid neopentanol. Flow rates were calculated to give 100 mL/min of saturated vapor from the bubblers, which were of sufficient path length to provide saturated vapors.[3] The background air flow was 1900 mL/min, so that the analyte concentration delivered to the sensors was 5% of the analyte's saturated vapor pressure at room temperature. The ability of the vapor delivery system to provide the expected analyte concentrations based on the input and control settings to the mass flow controllers was verified using a calibrated flame ionization detector that sampled several test analyte gas streams being delivered to the sensor chamber.

An exposure consisted of 300 s of background air flow, followed by 300 s of flow of analyte at 5% of its vapor pressure, followed by 300 s of background air. The dc resistance of each sensor was measured at intervals of approximately 6 s using a multiplexing ohmmeter. The baseline resistance of a sensor was taken as an average of all measurements of the resistance of that sensor acquired over a 60 s period that started between 60 and 66 s prior to the start of the exposure to an analyte. The exact initiation time of this baseline resistance measurement was different for each sensor, due to small variations in the time interval required to read the set of 20 resistance values through the multiplexing ohmmeter. The resistance response for each sensor to an analyte was taken as an average of all measurements for that sensor in a 60 s period that started between 234 and 240 s after the beginning of the presentation of the vapor to the sensors, with the exact initiation time for each sensor channel staggered similarly to that of the baseline resistance readings. A response was taken to be the change in resistance of a sensor, ΔR, divided by its baseline resistance, R.[3] All differential resistance values ($\Delta R/R$) used in the data analysis represented, or very closely approximated, the steady-state resistance readings obtained from the sensors during exposure to the analyte of interest.

Data analysis

Initial raw data manipulation and calculation of responses was performed using Microsoft Excel. Multiple linear regression (MLR) was performed using either Excel or the QSAR module of the Cerius[2] program (Molecular Simulations, Inc.) on a Silicon Graphics O_2 computer. Many possible MLR models were created, compared, cross-bred, and evolved by the genetic function approximation on Cerius[2].[17]

Acknowledgement

We acknowledge the Army Office of Research, MURI grant DAAG55-98-1-0266, for support of this work.

References

1. Hansch, C.; Hoekman, D.; Gao, H. *Chem. Rev.* **1996**, *96*, 1045.
2. Hansch, C.; Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society: Washington, DC, 1996.
3. Lonergan, M. C.; Severin, E. J.; Doleman, B. J.; Beaber, S. A.; Grubbs, R. H.; Lewis, N. S. *Chem. Mater* **1996**, *8*, 2298.
4. Slater, J. M.; Paynter, J. *Analyst* **1994**, *119*, 191.
5. Doleman, B. J.; Lonergan, M. C.; Severin, E. J.; Vaid, T. P.; Lewis, N. S. *Anal. Chem.* **1998**, *70*, 4177.
6. Doleman, B. J.; Severin, E. J.; Lewis, N. S. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5442.
7. Vaid, T. P.; Lewis, N. S. unpublished results.
8. Severin, E. J.; Sanner, R. D.; Doleman, B. J.; Lewis, N. S. *Anal. Chem.* **1998**, *70*, 1440.
9. Cohen, G. M.; Mannering, G. J. *Mol. Pharmacol.* **1973**, *9*, 383.
10. Testa, B. *Chem.–Biol. Interactions* **1981**, *34*, 287.
11. Sabljic, A.; Protic-Sabljic, M. *Mol. Pharmacol.* **1983**, *23*, 213.
12. Lewis, D. F. V. *Chem.–Biol. Interactions* **1987**, *62*, 271.
13. Shusterman, A. J.; Johnson, A. S. *Chem.–Biol. Interactions* **1990**, *74*, 63.

14. Livingstone, D. *Data Analysis for Chemists: Applications to QSAR and Chemical Product Design*. Oxford University Press: New York, 1995.
15. Burns, J. A.; Whitesides, G. M. *Chem. Rev.* **1993**, *93*, 2583.
16. Wang, Z.; Hwang, J.-N.; Kowalski, B. R. *Anal. Chem.* **1995**, *67*, 1497.
17. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
18. Friedman, J. *Multivariate Adaptive Regression Splines. Technical Report No. 102*. Stanford University: Laboratory for Computational Statistics, Department of Statistics, 1988 (revised August 1990).
19. McGill, R. A.; Abraham, M. H.; Grate, J. W. *Chemtech.* **1994**, *24*, 27.
20. Zellers, E. T.; Batterman, S. A.; Han, M.; Patrash, S. *J. Anal. Chem.* **1995**, *67*, 1092.
21. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*. Academic Press: New York, 1976.
22. Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
23. Anon. *Cerius*^{2} *QSAR+*; Molecular Simulations Inc.: San Diego, 1997.