



Correlating Polymer-Carbon Composite Sensor Response with Molecular Descriptors

Abhijit V. Shevade,^{*z} Margie L. Homer, Charles J. Taylor, Hanying Zhou,
April D. Jewell, Kenneth S. Manatt, Adam K. Kisor,
Shiao-Pin S. Yen, and Margaret A. Ryan

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

We report a quantitative structure-activity relationships (QSAR) study using genetic function approximations to describe the activities of a polymer-carbon composite chemical vapor sensor using a novel approach to selecting a molecular descriptor set. The measured sensor responses are conductivity changes in polymer-carbon composite films upon exposure to target vapors at parts-per-million concentrations. The descriptor set combines the basic analyte descriptor set commonly used in QSAR studies with descriptors for sensing film-analyte interactions. The basic analyte descriptors are obtained using a combination of empirical and semiempirical quantitative structure-property relationships methods. The descriptors for the sensing film-analyte interactions are calculated using molecular modeling and simulation tools. A statistically validated QSAR model was developed for a training data set consisting of 17 analyte molecules. The applicability of this model was also tested by predicting sensor activities for three test analytes not considered in the training set.

© 2006 The Electrochemical Society. [DOI: 10.1149/1.2337771] All rights reserved.

Manuscript submitted June 29, 2005; revised manuscript received June 30, 2006. Available electronically September 6, 2006.

An electronic nose is an array of weakly specific chemical sensors, controlled and analyzed electronically, which mimics the action of the mammalian nose by recognizing patterns of response using a data analysis algorithm.¹⁻⁵ Polymer-carbon composite films are used as conductometric sensors to detect organic vapors and other environmental contaminants in the JPL Electronic Nose (ENose). There are also electronic noses using inorganic-based sensors. Investigations involving polymer-only sensing films have been performed using transducers such as surface acoustic wave sensors,^{6,7} or mass-uptake sensors such as quartz crystal microbalances.⁸ There have been several studies to elucidate and predict the response of polymer-film-based sensors, primarily on pure polymer films. Response in polymer films has been modeled with linear solvation energy relationships (LSER),^{6,7} and solubility parameters,⁸ and good correlation has been obtained between calculated and measured responses. Previous efforts to model response in polymer-carbon composites using molecular modeling^{9,10} have taken into account only the effect of polymer-analyte interactions and assumed that neither carbon nor analyte in the film plays a role in sorbing analyte molecules or in contributing to the response of the film. This model of sensor response may not represent a complete picture of response in polymer-carbon composite sensors, especially at concentrations of single to tens of parts-per-million (ppm) analyte in air.

We have developed an approach which relates sensing material and analyte to sensor activity by using experimental and theoretical data. This approach takes into account the interactions between the analyte and all of the components of the composite sensor. We have used a multivariate statistical approach, quantitative structure-activity relationship (QSAR), to correlate sensor activity with sensing film and analyte physical and chemical properties. This technique has been used extensively in biochemical, medical, and environmental remediation fields for drug-receptor screening and in evaluating phenomenological models.¹¹⁻¹⁷

Our goal is to develop a representative equation for each polymer-carbon composite sensor in the JPL ENose sensing array. The current QSAR studies for a single sensor in the JPL ENose will have a limited training set. The small set size is a result of the application of the JPL ENose; it is being developed as an event monitor for spacecraft air and the set of target analytes is 20–25 compounds. This results in a limited training data set, unlike other

QSAR studies for pharmaceutical and biological systems, where training sets may include more than 100 compounds.

Statistical analysis methods available in QSAR include data analysis and regression analysis methods.¹⁸ Methods such as principal component analysis (PCA) and cluster analysis methods (e.g., hierarchical cluster analysis) are included in data analysis. PCA aims at representing large amounts of multidimensional data as a more intuitive, low-dimensional representation. Cluster analysis methods are aimed toward partitioning a data set into classes or categories consisting of elements of comparable similarity. Regression methods include simple and multiple linear regression methods (MLR), stepwise multiple linear regression using genetic function approximation (GFA) methods, and partial least-square methods (PLS).

GFA is a powerful method that has been used extensively to investigate structure-activity relationships in biological data.¹⁴ The advantage of GFA is that during evolution, thousands of candidate models are created and tested, and only the superior models survive. These models are then used as “parents” for the creation of the next-generation candidate models. GFA can thus select the optimum number of descriptors in linear regression analysis automatically; it also constructs multiple linear regression models with any possible combinations of terms (linear, higher order polynomials, splines, and Gaussians). The multiple models produced and subsequently analyzed by GFA are in contrast to most statistical methods, such as MLR and PLS, that focus on a single “best statistical” model.

The descriptors commonly used in QSAR studies describe intrinsic chemical and physical analyte properties. These descriptors are predicted using empirical and semiempirical predictive methods such as quantitative structure-property relationships (QSPR). We carried out some trial QSAR runs to determine the performance of these default analyte descriptors to describe the sensor activity. Although the intrinsic analyte descriptors provide a statistically significant fit, we were concerned that the default descriptors might not be sufficient to describe the sensor response. Polymer-carbon composite sensor responses are measured as a change in resistance, and the response can be attributed to the swelling of the polymer film; other mechanisms may also contribute to the sensor response. Previously, we used the LSER methods⁶ to model responses of three polymer-composite sensors to six different analytes.¹⁸ Comparison to experimental data showed LSER to be a poor predictor for sensor response. This is not surprising as the equation considers only polymer-analyte interactions and does not account for adsorption of analyte on the carbon conductive medium dispersed in the film, for analyte molecule diffusion within the film, film thickness, or hydration of the film. Accordingly, the descriptor sets used in this QSAR

* Electrochemical Society Active Member.

^z E-mail: Abhijit.shevade@jpl.nasa.gov

Table I. Analyte list and concentration range tested in parts-per-million (ppm) for ENose operation (760 Torr, 23 °C). Data are taken at a constant humidity of 5000 ppm water (approximately 18% relative humidity).

Analyte	Concentration tested Low - High (ppm)
1. Acetone	64–600
2. Ammonia	6–60
3. Chlorobenzene	3–30
4. Dichloromethane	10–150
5. Ethanol	200–6000
6. Isopropanol	30–400
7. Xylenes (mixed)	33–300
8. Tetrahydrofuran	13–120
9. Trichloroethane	7–200
10. Acetonitrile	1–25
11. Ethylbenzene	20–180
12. Freon113	15–500
13. Hexane	15–150
14. Methyl ethyl ketone	15–150
15. Methane	1600–50000
16. Methanol	6–100
17. Toluene	5–50
18. Benzene	10–100
19. Indole	25–450
20. Dichloroethane	10–100

study combine descriptors that describe the intrinsic analyte properties as well as interactions between sensing film and analytes.

This study compares different descriptor sets that include the default analyte descriptor sets as well as film-analyte descriptors that we have modeled. The sensing film-analyte interactions are calculated with molecular modeling tools using two approaches. To test the validity of the approach of correlating sensor activity with molecular descriptors, we have used experimental data from one sensor used in the JPL ENose sensor array.^{2,3} The sensor is a film of polyethylene oxide that has been loaded with 12% (wt/wt) carbon black. QSAR equations that describe sensor activities were developed for a training set of 17 analytes using GFA. If the statistically most significant equation produced by QSAR did not contain the sensing film-analyte interaction term, the most significant QSAR equation which did was selected. Generally the statistical significance was not different for the above cases. The QSAR equation which was selected to represent the sensor was then used to predict the activity of test analytes not included in the QSAR training set.

Experimental

Sensor activity measurements.—The JPL ENose consists of an array made up of 16 polymer-carbon composite films.^{2,3} Response from the polyethylene oxide-carbon (PEO-CB) sensing films was selected as a model system for this study. The choice of the current system is to add to our understanding of the behavior of the PEO-CB sensing films by developing a model that relates the sensor activity to molecular descriptors. Our previous experimental studies on the PEO-CB sensing films¹⁹⁻²³ have focused on evaluating factors affecting sensing film formulation, such as carbon content, solvent type, polymer molecular weight, temperature, and measuring the PEO-CB sensor response to a suite of analytes under various operating conditions, including humidity and temperature. The present work provides insight into the sensing mechanism by developing a relationship between the sensor activity and molecular descriptors. The analytes and concentration ranges considered for this study are shown in Table I.

QSAR conditions.—To develop data for QSAR studies, experimental responses of a PEO-CB sensor were used. Measured concentrations of analyte with controlled humidity were delivered to the sensor several times over a range of concentrations. The response of

the sensor is expressed as a normalized change in sensor resistance and is plotted against the delivered concentration. Response data for each analyte and each sensor are fit to an equation of the form $y = A_1x + A_2x^2$, where x is the analyte concentration and y is the normalized change in resistance.²³ Figure 1 shows experimental data and the fitted concentration-response equations for a PEO-CB sensor and three representative analytes. As can be seen in the figure, the coefficient for the quadratic term, A_2 , is generally three orders of magnitude smaller than the coefficient of the linear term, A_1 . By selecting coefficients from the sensor response equation as the activity to be used in QSAR studies, we can develop an approach to calculating response which is concentration independent. For the purpose of this study, to determine whether a QSAR approach can be used to predict analyte response in a sensor, only the coefficient, A_1 is used as the activity in this study. Analytes no. 1–17 (Table I) are used in the training set to develop the QSAR equations, and the remaining analytes (no. 18–20) are used as the test set. Test molecules were selected based on experimental data. The sensor response to these test analytes showed larger fluctuations under varying humidity conditions than other analytes.

QSAR.—Descriptors: analyte properties and sensor response.—The goal of the current investigation is to correlate coefficient A_1 of the JPL ENose sensor response with physico-chemical molecular descriptors. The physical descriptors that are considered for the QSAR study include those that describe intrinsic analyte properties as well as sensor-analyte interactions.

The default descriptors in QSAR studies which describe intrinsic analyte properties fall into electronic, spatial, structural, thermodynamic, and topological categories. These default analyte descriptors are shown in Table II.

Sensor response at a molecular level is described by interactions between the sensing film and the analyte. Descriptors for sensor response include interaction energies of the sensing film components, polymer, and carbon-black, with analyte molecules and water.

Two approaches to development of a sensor response descriptor set were investigated in this work. These approaches represent two different molecular models to calculate analyte interactions with the polymer-carbon sensing film. Approach I is based on cluster calculations (nonperiodic) of various sensing film components with the analyte. Approach II involves performing sorption studies of analytes in a periodic model of the polymer-carbon composite film. Approach I is computationally less intensive than approach II; these approaches are discussed below.

Approaches for calculating sensing film-analyte interaction energies.—In approach I, the individual components of the polymer-carbon composite system are used to calculate the binding energies between analyte molecules and between analyte molecules and the polymer-carbon composite sensing film components. In approach II, the interaction of the analyte with the polymer-carbon composite film is described by calculating the isosteric heat of sorption of analyte molecules in the polymer-carbon composite. To calculate the heat of sorption, a model of the polymer-carbon composite film was developed based on the sensing film formulation process.

The pair interaction energies²⁴ considered to represent the sensor response descriptors in approach I are polymer-analyte, carbon black-analyte, polymer-water, and carbon black-water, analyte-analyte, and analyte-water. The general notation that is used here to represent these interaction descriptors is of the form E_{xy} . Depending on the type of interactions considered, the suffixes x and y could be polymer (p), carbon black (cb), analyte (a), or water (w). For example interaction energy between the polymer (p) and the target analyte (a) is denoted by E_{pa} . The sensor response descriptor set using approach I is shown in Table III. It includes the sensing film-analyte descriptors E_{pa} , E_{pw} , E_{p-cb} , E_{cb-a} , E_{cb-w} , E_{cb-cb} , E_{aa} , and E_{aw} .

For approach II, the isosteric heat of sorption (H_{sorp}) (Ref. 25) of the analyte in the polymer-carbon composite is the descriptor that represents the combined interactions of the analyte with the polymer and carbon components. This term incorporates the combined effect

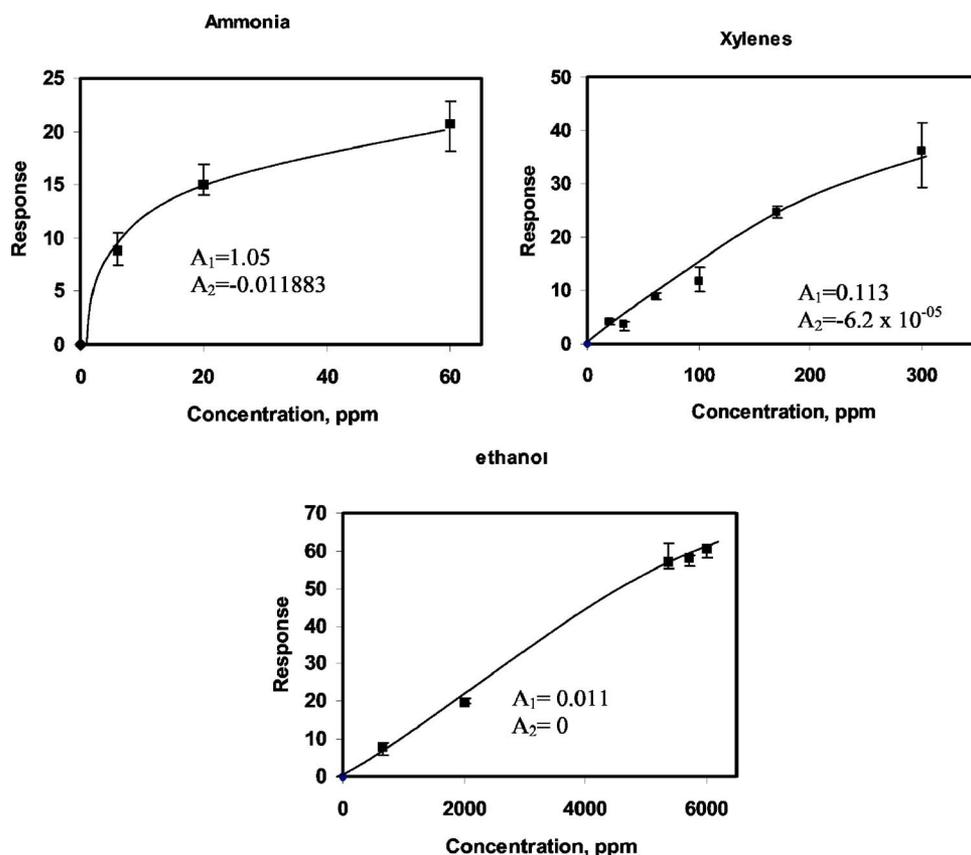


Figure 1. Plots for three gases (ammonia, xylenes, ethanol) showing the experimental data and fitted concentration-response equation ($y = A_1x + A_2x^2$) for the PEO-CB sensing film. The lines are drawn to guide the eye.

of the separately used terms in approach I: E_{pa} and E_{cb-a} . The H_{sorp} values of the analyte molecules in the PEO-CB composite are shown in Table III. The sensor response descriptor set using approach II for the QSAR study includes H_{sorp} , E_{aa} , and E_{aw} . These interactions also play an important role apart from the interactions of the analyte molecule with the polymer-carbon composites, reflected in the H_{sorp} value.

A combined descriptor set that includes the default analyte descriptors along with sensor response descriptors calculated using the two approaches (approach I and II) was used in two separate QSAR studies to correlate the sensor coefficients A_1 with the molecular descriptors. The combined descriptor set based on the approaches I and II are referred to here as descriptor set *D1* and *D2*, respectively.

Table II. Default analyte descriptor set.

Default analyte descriptors	Description
A_{pol}	Sum of atomic polarizabilities
Dipole-mag, Dipole-X, Y, Z	Dipole moment magnitude and X, Y, and Z components
R_G	Radius of gyration
Area	Molecular surface area
MW	Molecular weight
V_m	Molecular volume
Density	Density
PMI-mag, PMI-X, Y, Z	Principal moment of inertia magnitude and X, Y, and Z components
Rotlbonds	Number of rotatable bonds
HB_A	Number of hydrogen bond acceptors
HB_D	Number of hydrogen bond donors
AlogP	Log of the octanol/water partition coefficient
MR	Molar refractivity

Methodology.—Descriptor calculations.—Default analyte descriptors were predicted by empirical and semiempirical QSPR using the commercial software²⁶ Cerius² on a Silicon Graphics O2 workstation.

Molecular models used in approach I were developed using the Cerius² graphical models and are shown in Fig. 2. The polymer was modeled using its basic unit, the monomer. Carbon black was modeled as naphthalene rings with no hydrogen (small graphite sheets) and zero charge on the carbon atom.²⁷ A cluster of 32 naphthalene molecules was used to represent the CB. Analyte models except water were constructed using the drawing tools in the software, and all atoms were assigned charges and equilibrated according to the methodology discussed below. The single point charge model was used for water.²⁸ Charges on the monomer and analyte atoms were assigned by the charge equilibration method (Q_{eq}).²⁹ The Dreiding force field³⁰ was used for the polymer and analyte molecules, and graphite parameters were assigned to carbon black atoms.³¹ Equilibration was achieved by molecular mechanics and then by molecular dynamics simulations at 300 K.

Details for the polymer-carbon composite molecular model for approach II (Fig. 3) are described elsewhere.²⁷ In brief, the polymer-carbon composite model was developed by inserting naphthalene rings (carbon black) in a polymer matrix, followed by a combination of molecular mechanics (MM) and molecular dynamics (NPT-MD and NVT-MD) techniques for equilibration. Force field parameters used in approach II are the same as those used in approach I.

To calculate the sensor response descriptors in approach I, the sensing film-analyte interaction energies were calculated using Monte Carlo simulation techniques. Interaction energy between the polymer and an analyte molecule, E_{pa} , was calculated by fixing the polymer structure in space and sampling the analyte molecule around the polymer, then averaging the energy calculated over all the random analyte configurations generated around the polymer. In this study we used 10^5 configurations. E_{pw} , E_{p-cb} , E_{cb-a} , E_{cb-w} ,

Table III. QSAR study table showing sensing film-analyte interaction energy descriptors used in approaches I and II. Sensing film-analyte interactions energies (kcal/mol) are calculated using molecular modeling techniques. The standard deviations for E_{xy} and H_{sorpt} values are shown in parentheses.

Analyte	Experimental activity	E_{pa}	E_{pw}	E_{aw}	E_{aa}	E_{cba}	E_{p-cb}	E_{cb-cb}	H_{sorpt}
Acetone	0.006	-0.50 (0.56)	-0.28 (0.74)	-0.33 (0.75)	-0.65 (0.54)	-0.95 (0.36)	-0.79 (0.30)	-1.68 (0.74)	9.96 (0.10)
Ammonia	1.05	-0.16 (0.34)	-0.28	-0.11 (0.51)	-0.05 (0.23)	-0.28 (0.10)	-0.79	-1.68	3.65 (0.65)
Chlorobenzene	0.229	-0.69 (0.75)	-0.28	-0.43 (0.82)	-1.14 (1.08)	-1.28 (0.58)	-0.79	-1.68	14.72 (2.00)
Dichloromethane	0.025	-0.57 (0.63)	-0.28	-0.36 (0.83)	-0.77 (0.76)	-1.09 (0.41)	-0.79	-1.68	11.54 (1.72)
Ethanol	0.0112	-0.42 (0.51)	-0.28	-0.25 (0.67)	-0.41 (0.59)	-0.78 (0.30)	-0.79	-1.68	9.88 (1.27)
Isopropanol	0.054	-0.48 (0.53)	-0.28	-0.28 (0.66)	-0.49 (0.70)	-0.90 (0.33)	-0.79	-1.68	10.58 (1.37)
o-xylene	0.113	-0.62 (0.42)	-0.28	-0.30 (0.52)	-0.94 (0.63)	-1.20 (0.55)	-0.79	-1.68	13.79 (0.44)
Tetrahydrofuran	0.029	-0.55 (0.52)	-0.28	-0.32 (0.63)	-0.53 (0.65)	-1.19 (0.53)	-0.79	-1.68	11.72 (0.68)
Trichloroethane	0.0024	-0.65 (0.60)	-0.28	-0.43 (0.79)	-1.27 (0.77)	-1.26 (0.49)	-0.79	-1.68	13.97 (2.55)
Acetonitrile	0.134	-0.44 (0.34)	-0.28	-0.29 (0.47)	-0.46 (0.25)	-0.83 (0.32)	-0.79	-1.68	7.97 (1.13)
Ethylbenzene	0.445	-0.63 (0.42)	-0.28	-0.37 (0.46)	-0.91 (0.58)	-1.19 (0.53)	-0.79	-1.68	13.03 (1.40)
Freon113	0.007	-0.65 (0.67)	-0.28	-0.45 (0.55)	-0.17 (1.70)	-1.32 (0.50)	-0.79	-1.68	12.67 (0.86)
Hexane	0.0253	-0.55 (0.40)	-0.28	-0.32 (0.42)	-0.69 (0.40)	-1.06 (0.43)	-0.79	-1.68	12.05 (0.90)
Methylethyl ketone	0.0265	-0.52 (0.57)	-0.28	-0.33 (0.73)	-0.64 (0.73)	-1.01 (0.39)	-0.79	-1.68	11.43 (1.12)
Methane	0.0001	-0.32 (0.10)	-0.28	-0.17 (0.10)	-0.23 (0.06)	-0.61 (0.22)	-0.79	-1.68	3.72 (0.41)
Methanol	0.031	-0.34 (0.57)	-0.28	-0.23 (0.82)	-0.29 (0.60)	-0.65 (0.25)	-0.79	-1.68	8.57 (0.87)
Toluene	0.061	-0.62 (0.39)	-0.28	-0.36 (0.45)	-0.90 (0.57)	-1.17 (0.53)	-0.79	-1.68	11.46 (2.56)

E_{cb-cb} , E_{aa} , and E_{aw} descriptors were calculated similarly. The BLENDS module in the Cerius² software performs the calculations based on the methodology described above.

The sensor response descriptor calculations for approach II involve performing sorption simulations of the analyte molecule in the polymer-carbon composite to calculate the isosteric heat of sorption (H_{sorpt}). H_{sorpt} is a positive value²⁵ and is the differential heat liberated by the isothermal desorption of a differential amount of the adsorbate (analyte). It is calculated from the integral enthalpy of desorption relative to a perfect-gas reference state. In our case, H_{sorpt} depends on the interaction energies of the sorbed analyte molecule in the polymer-carbon composite. The sorption simulations for this work were performed at constant analyte loading of one molecule. The SORPTION module in the Cerius² software was used to accomplish the task. The program generates random points in the polymer and tries to insert an analyte molecule. Insertion attempts that involve the overlapping of the analyte molecule with the polymer

structure are discarded. After the insertion step, each subsequent configuration is generated by either a random translation or rotation of the analyte molecule in the polymer matrix, taken in the usual Metropolis Monte Carlo manner. The isosteric heat of sorption value is calculated at the end of the run. Sorption simulations were performed using two trajectories extracted from the NVT-MD simulations. These trajectories were recorded during the NVT-MD simulations corresponding to PEO-CB composite structure at 50 and 100 ps time. For each trajectory, two independent sorption simulations at a fixed analyte loading of one molecule were carried out at 300 K for $\sim 1-3$ million iterations. This was to allow the analyte molecule to access different structural and spatial amorphous environments in the polymer composite. The software calculates a running average value of the isosteric heat of sorption at the end of each of run. The heat of sorption value for each analyte used for the current work represents an average value over the trajectories.

QSAR equation: terms and functional form selection.— The first step in the QSAR model development process was to use the combined descriptor set to investigate the number of terms (N_{term}) and the functional forms, linear or linear quadratic, to be used in the

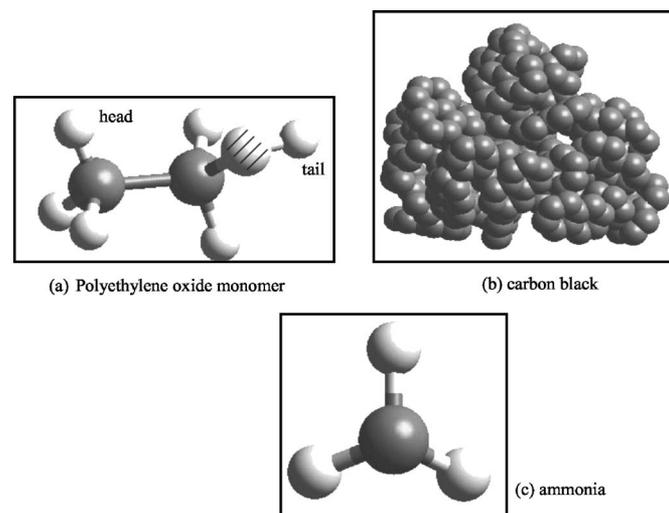


Figure 2. Molecular models for approach I to calculate sensing film-analyte interactions. (a) Polyethylene oxide monomer unit, (b) carbon black, (c) ammonia molecule (representative model). The shaded atom in the polyethylene oxide monomer unit represents oxygen.

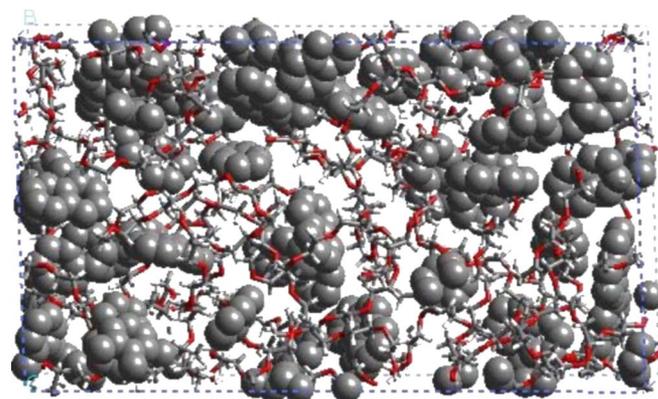


Figure 3. Molecular model for approach II to calculate sensing film-analyte interactions. Shown is a polyethylene oxide-carbon composite model with a density of 0.90 g/cm³. The clusters represent carbon black and the cylindrical chains represent the polymer.

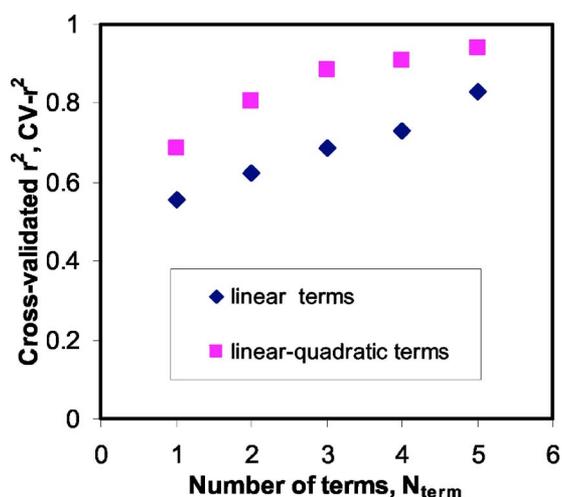


Figure 4. (Color online) A plot of cross-validated r^2 ($CV-r^2$) vs number of descriptor (N_{term}) for the training analyte set using the combined descriptor set (analyte and sensor response descriptors).

QSAR equations. Spline terms are not considered because of the small size of the training set (17 analyte molecules).

QSAR studies were performed using $D1$ descriptors with no constant by varying the number of terms from 1 to 5, with linear and quadratic functionality. QSAR studies for the PEO-CB sensor were performed using GFA for the training set, analytes 1–17. QSAR equations were developed using 5000 crossovers.

Statistical parameters such as correlation coefficient (r^2) and F (Ref. 18) were used to determine the reliability and significance of QSAR models. The r^2 value describes the goodness of fit of the data to QSAR model. The F value is a ratio of explained and unexplained variance. As the F value increases, the significance of the QSAR equation becomes greater.

The predicting ability of the QSAR equations can be estimated using the leave-one-out cross-validation technique. In this procedure, new regression coefficients are generated for a given model after systematically removing one analyte sample at a time from the training data set. This new regression model is then used to predict the activity of the removed sample. This procedure is performed for each of the 17 analytes. The series of predictions is then used to calculate a new value for the cross-validated r^2 ($CV-r^2$).

To determine the number of terms and equation type to be used in our model, we plotted $CV-r^2$ against the number of terms for both linear and linear-quadratic equation types as shown in Fig. 4. It can be seen that the linear-quadratic terms have a greater $CV-r^2$ compared to the linear terms. Figure 4 also shows that, for a linear-quadratic equation, no substantial increase in $CV-r^2$ is achieved by increasing the number of terms from 3 to 4. Because the training data set is limited to only 17 analytes, any increase in number of terms in the model would be an overfit to the data. Even in traditional QSAR studies for biological systems, using 50–100 training data points,^{14–17} the number of terms considered for model development is generally 3–4 terms.

QSAR equation: sensor activity representation.—The response of a polymer carbon sensing film to a given analyte molecule is based on how the sensing film components (polymer and carbon black) in the polymer-carbon composite interact with the analyte molecule. Therefore, the QSAR equation that we have chosen to represent a given sensor is selected from a set of cross-validated equations generated by the GFA algorithm. The selected equation is the statistically most significant one (largest r^2 value) of the equation set, which also contains the polymer-analyte (E_{pa}) descriptor for the $D1$ descriptor set or heat of sorption (H_{sorp}) term for the $D2$ descriptor set. The goal of this modeling effort is to develop an equation for

Table IV. QSAR generated partial equation set for approaches I and II, containing E_{pa} and H_{sorp} terms. The statistically significant equation is the one chosen to represent the sensor activity.

Calculated Activity (A_1)=	r^2	F
Approach I		
$0.15207 E_{pa} + 0.116727 HB_D^2 + 0.000241 MR^2$	0.86	40.6
$-0.283563 E_{pa}^2 + 0.114330 HB_D^2 + 0.000256 MR^2$	0.86	38.7
$0.286937 E_{pa}^2 + 0.182699 HB_D^2 - 0.200424 HB_D$	0.84	34.6
Approach II		
$-0.008856 H_{sorp} + 0.118308 HB_D^2 + 0.000264 MR^2$	0.87	44.4
$-0.000801 H_{sorp}^2 + 0.115485 HB_D^2 + 0.000284 MR^2$	0.87	43.1
$-0.022602 H_{sorp} + 0.120321 HB_D^2 + 0.013894 MR$	0.86	40.3

each sensor in the ENose array, each equation describing the response of that particular sensor to analytes. The interaction descriptors E_{pw} , E_{p-cb} , and E_{cb-w} are the same for all analytes and hence will not appear in the QSAR equations.

Results and Discussion

QSAR using $D1$ descriptor set (approach I).—As discussed in the Methodology section, we have used a three term linear-quadratic form for the QSAR study. A set of cross-validated QSAR equations generated by GFA, containing the polymer-analyte (E_{pa}) term, is shown in Table IV. The statistically most significant equation containing the descriptor E_{pa} is

$$\text{Calculated activity-1} = 0.15207 E_{pa} + 0.116727 HB_D^2 + 0.000241 MR^2 \quad [1]$$

($r^2 = 0.86$, $F = 40.6$)

The best statistical QSAR equation obtained using descriptor set $D1$ has $r^2 = 0.88$ and $F = 44.8$; it does not contain an E_{pa} term. There is no substantial statistical difference between the best equation obtained and Eq. 1. The calculated activity is plotted vs the experimental values in Fig. 5.

In QSAR Eq. 1, we observe that the analyte descriptors that appear along with the polymer-analyte interaction term (E_{pa}) are hydrogen bond donor site (HB_D) and molar refractivity (MR). The PEO monomer has one hydrogen bond acceptor site; thus, it is logical that a descriptor that represents the hydrogen bond donor nature of the analyte appears in the equation. The analyte descriptor MR is a combined measure of molecule size and polarizability, and is calculated from the refractive index, molecular weight, and density of

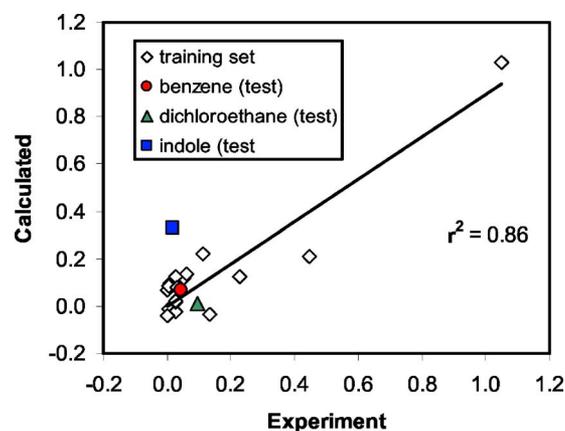


Figure 5. (Color online) A plot of QSAR calculated vs experimental sensor activity for the training and test analyte set using combined descriptor set $D1$. The calculated values for both the training and test analyte set are obtained using Eq. 1. The r^2 value refers to the correlation between the calculated vs the experimental values obtained for the training data set.

the analyte. As swelling in the polymer-carbon composite film is one mechanism of sensor response, it is logical that molecular size of the analyte will appear in the equation describing sensor response.

The next task was to use QSAR Eq. 1, developed using the training set to predict the sensor coefficients for test analytes that were not included in the training set. The aim of this exercise is to determine whether the equation can be used to predict sensor activity without performing exhaustive and time-consuming experiments, for example if new analytes are added to the target list. The analytes considered for this case are analytes 18–20 of Table I, benzene, dichloroethane, and indole. The calculated values for the training set and the test analytes are also shown in Fig. 5 along with the training data set. The model works satisfactorily for benzene and dichloroethane but overpredicts the activity for indole. Indole is the least volatile analyte on our list. It is also the only aromatic on our list that has a hydrogen bond donor site. As discussed below, the A_1 coefficient for indole may be better predicted with the inclusion of vapor pressure in the molecular descriptor set.

Effect of polymer chain length on sensor activity equation.— The polymer-analyte interaction energies in Eq. 1 were calculated considering a monomer representation of the polymer. We also investigated the effect of polymer chain length on the QSAR model formation. A polymer chain model consisting of 20 PEO monomer units was built. The detailed methodology to construct the polymer chain model is discussed elsewhere.²⁷ The statistically most significant QSAR equation containing the descriptor E_{pa} is

$$\text{Calculated activity-1} = 0.074309 E_{pa} + 0.11555 \text{HB}_D^2 + 0.000222 \text{MR}^2 \quad [2]$$

($r^2 = 0.86$, $F = 40.0$)

On comparing the sensor activity Eq. 1 and 2 for approach I, representing the monomer and polymer cases, respectively, note that the statistically best equation with E_{pa} contains the same other descriptors: HB_D and MR , with similar coefficients. The statistical parameters, r^2 and F , are also similar. This leads us to believe that using monomer or polymer case will give similar results.

QSAR using D2 descriptor set (approach II).— QSAR studies using descriptor set $D2$ were performed using the three-term linear-quadratic form. The statistically most significant equation generated which set contained the heat of sorption term (H_{sorpt}) along with other descriptor terms was chosen from a set of cross-validated models generated in the QSAR studies. The best statistical QSAR equation obtained using descriptor set $D2$ has $r^2 = 0.88$ and $F = 46.3$. A set of cross-validated QSAR equations generated by GFA, containing the H_{sorpt} term, is shown in Table IV. The statistically most significant of the QSAR equations obtained containing H_{sorpt} term was

$$\text{Calculated activity-2} = -0.008856 H_{\text{sorpt}} + 0.118308 \text{HB}_D^2 + 0.000264 \text{MR}^2 \quad [3]$$

($r^2 = 0.87$, $F = 44.4$)

Equation 3 was used to predict the sensor coefficient A_1 for the test analytes: benzene, dichloroethane, and indole. A plot of QSAR calculated vs experimental sensor activities for the descriptor set $D2$ is shown in Fig. 6. As seen in Fig. 6, similar to the previous case, the model works satisfactorily for benzene and dichloroethane but overpredicts indole.

The r^2 value of the calculated vs experimental fit using Eq. 3 is not significantly different from that of Eq. 1. On visual comparison of Fig. 5 and 6, we can see that the data fit using $D2$ descriptor set appears to be marginally better than the data fit from $D1$. This improvement is also reflected in a marginal improvement in the F value. On comparing Eq. 1 and 2, we see similarity in the descriptors that have appeared in the equations as well as the functional form. There is quadratic dependence on the HB_D and MR terms, and a linear dependence on the sensor response descriptor that appears in

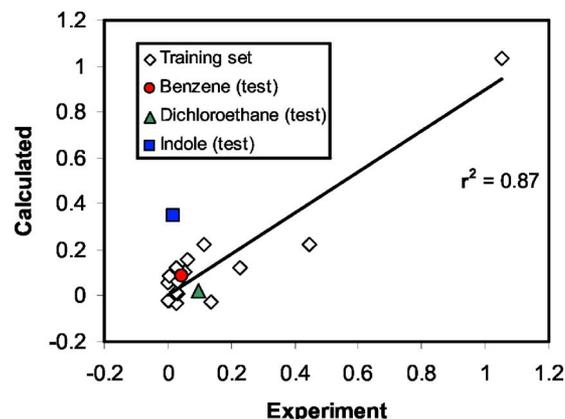


Figure 6. (Color online) A plot of QSAR calculated vs experimental sensor activity for the training and test analyte using combined descriptor set $D2$. The calculated values for both the training and test analyte set are obtained using Eq. 3. The r^2 value refers to the correlation between the calculated vs the experimental values obtained for the training data set.

the selected equation, E_{pa} , for approach I and H_{sorpt} for approach II. The sensor response descriptors have similar functional nature.

Alternative descriptors and approaches.— Note that the analyte descriptor terms MR and HB_D , which appear in our calculated activity models, also appear in the LSER approach.^{6,32} In addition, the partition coefficient of a sensing film, K , is correlated with a linear combination of analyte solubility descriptors (solvation parameters) in LSER. The regression coefficients obtained for the LSER models characterize the properties of the sensing film. The analyte property terms that appear in the LSER equation are the excess molar refractivity, dipolarity-polarizability, hydrogen bond acidity and basicity parameters, and gas-liquid partition coefficient. The LSER approach has been used for sorption studies of vapors in a polymer film only³² and also for sorption in graphite-fullerene coatings.³³ The sensing film used in our studies is a polymer-carbon composite film; our previous efforts¹⁹ to use LSER for polymer-carbon composite have resulted in poor correlation between calculated and measured sensor response. The QSAR approach used in this work is different from LSER, as it includes polymer-analyte and carbon black-analyte interactions as well as contributions from analyte-water interactions that represent experimentation conditions in the descriptor set.

Approach I used in this work to calculate the sensor response descriptors is a rapid approach that takes into account the chemical nature of the individual components, i.e., thermodynamic and electronic characteristics of the monomer, carbon black, and the analyte. Approach II represents a more complete representation of sensing film-analyte interactions, and takes into account the structural aspect of the polymer composite film as well as the sorption process that occur, which mimics the process that happens during experiments. The performance of these two approaches is similar, but we might see differences in performance of these approaches for our future work with other sensing films used in the formation of the JPL-ENose sensor array based on different polymer types being used.

The above studies suggest that the descriptor set used in this work does not describe the sensor response fully; it is possible that additional descriptors could be considered in the descriptor list. In addition, it is recognized that the partition coefficient of the analyte in the polymer correlates to polymer-carbon black sensor response.^{34,35} The sensing film response for gas-phase detection of a target analyte molecule is a function of the equilibrium partition coefficient, K , of the analyte molecule in the sensing film.^{32,36,37} This equilibrium constant is defined as the ratio of the equilibrium concentrations of the analyte in the sensing film (C_s) to the bulk

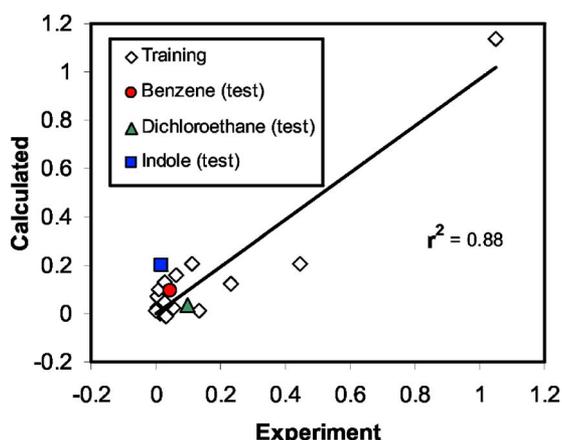


Figure 7. (Color online) A plot of QSAR trial run results, with vapor pressure (VP) descriptor added to the combined descriptor set $D2$. The calculated values for both the training and test analyte set are obtained using the equation, $0.000207 MR^2 + 1.00E - 06 VP^2 - 0.000344 H_{sorp}^2$. The r^2 value refers to the correlation between the calculated vs the experimental values obtained for the training data set.

analyte concentration (C_v) the sensing film is exposed to. The analyte concentration in the sensing film, C_s , is normally measured using piezoelectric techniques. The bulk analyte concentration in a carrier gas (air or nitrogen) depends on the vapor pressure of the target analyte at a given temperature. As evaluating partition coefficients by measuring experiments for mass uptake by the polymer-carbon composite film is not within the scope of this work, we decided to run a trial QSAR study by adding the vapor pressure³⁸ of the analyte molecules (calculated at 300 K) to the descriptor set $D2$. Preliminary results are shown in Fig. 7. There is no substantial improvement in the fit of the training data set, seen from $r^2 = 0.88$ and $F = 48.5$ values. However, the QSAR equation predictions for the test analytes appear superior, particularly for indole, which was poorly predicted by the previous two descriptor sets. The descriptor that replaces the HB_D term of Eq. 2 is vapor pressure (VP). Future studies will determine whether descriptors related to analyte partitioning in the sensing film should be included as a variable in the descriptor list.

Conclusions

The similarity of the results for approaches I and II depends on the correlation between the calculated coefficients, A_1 , and the molecular descriptor set. The key descriptor used for sensor activity representation for the current investigation is the sensing film-analyte interaction, in particular E_{pa} or H_{sorp} . The calculations of E_{pa} or H_{sorp} values depend on the physicochemical and morphological properties of the polymer/polymer-carbon composite matrix; polymers with different functional groups from the PEO used in this study may lead to significantly different results for approaches I and II. For example, in approach II, which involves building a polymer-carbon composite model, the flexibility of the polymer could also influence the microstructure of the polymer-carbon composite, and thus the sorption location of the analyte and the heat of sorption value. Also, for PEO, the same correlation is observed between the calculated coefficients A_1 and the molecular descriptors. The 16 polymers used on the JPL ENose represent various categories of chemical functionality: hydrogen-bond acidic, hydrogen-bond basic, dipolar, and hydrogen-bond basic, etc. The similarity of the two approaches for other polymer-carbon sensing film systems, for polymers in these categories (e.g., ethyl cellulose, polypyrrolidone) is currently being investigated.

We do not expect to calculate the exact value of the coefficient A_1 using the QSAR approach described here. It is likely that there are effects influencing sensor response which are not accounted for

in the models of interaction energy or in the physicochemical properties of the analyte. For example, any addition of pathways for conduction, such as ions in the polymer matrix, would result in a decrease in resistance, but such a decrease would not be accounted for in either the calculated interaction energies or in the default descriptor set. In addition, because the sensor response is expressed as a quadratic equation, and this study does not consider the coefficient A_2 , there will necessarily be some degree of error in the calculated A_1 . As the coefficient A_2 is generally three orders of magnitude smaller than A_1 , the quadratic term does not contribute strongly to sensor response at low concentration, but it does contribute. This study demonstrates that this QSAR approach is a promising approach to predicting sensor response to a new analyte, and will allow us to determine whether an analyte is likely to induce a weak or strong response in selected sensors. Future work will focus on incorporating both A_1 and A_2 into the model, on expanding the sensor set beyond the PEO-carbon composite film used in this study, and on selection of descriptors for full description of activity.

Experimental data for a polyethylene oxide-carbon black sensor in the JPL ENose were correlated using QSAR with intrinsic analyte properties and molecular interaction energy terms. The model developed showed good correlation for the entire analyte set as well as analyte subsets. The descriptors that predict the polymer-carbon sensor response indicate that the polymer-analyte interaction is not the only important interaction to consider. In addition to predicting sensor response, it may be possible to elucidate the sensing mechanisms using this approach. The approach will be extended to other polymer composite sensors used in the JPL ENose system.

The ability to predict sensor responses accurately will be of great help in characterizing sensing materials. Developing a response library (training), an array for a given set of analytes, and a given set of environmental conditions (temperature, pressure, and humidity) is time-consuming; in addition, developing training sets and calibration information may impinge on the useful lifetime of the sensors.² Our goal is to develop one representative equation for each sensor material (polymer-carbon sensing film) in the array. In the future, the development of “ n ” equations to describe the ENose sensing array will facilitate the generation of virtual training sets for any given sensor array for analytes that may not easily be tested, such as highly toxic or explosive compounds. The predictions can also be used to generate parameters for the identification and quantification software. Subsequently, fewer experimental tests will need to be run on any given sensor array.

Acknowledgments

This research was funded by the Advanced Environmental Monitoring and Control Program of NASA. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The authors thank Anton Hopfinger and David Rogers for their valuable suggestions and the Caltech Material and Process Simulation center (MSC) for use of computing facilities.

California Institute of Technology assisted in meeting the publication costs of this article.

References

1. P. N. Bartlett and J. W. Gardner, *Electronic Noses: Principles and Applications*, Chap. 7, pp. 140–183, Oxford University Press, Oxford (1999), and references therein.
2. M. A. Ryan, H. Zhou, M. G. Buehler, K. S. Manatt, V. S. Mowrey, S. P. Jackson, A. K. Kisor, A. V. Shevade, and M. L. Homer, *IEEE Sens. J.*, **4**, 337 (2004).
3. M. A. Ryan, A. V. Shevade, H. Zhou, and M. L. Homer, *MRS Bull.*, **29**, 714 (2004).
4. M. S. Freund and N. S. Lewis, *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 2652 (1995).
5. S. V. Patel, M. W. Jenkins, R. C. Hughes, W. G. Yelton, and A. J. Ricco, *Anal. Chem.*, **72**, 1532 (2000).
6. J. W. Grate, S. J. Patrash, S. N. Kaganove, M. H. Abraham, B. M. Wise, and N. B. Gallagher, *Anal. Chem.*, **73**, 5247 (2001).
7. A. Hierlemann, E. T. Zellers, and A. J. Ricco, *Anal. Chem.*, **73**, 3458 (2001).
8. M. P. Eastman, R. C. Hughes, G. Yelton, A. J. Ricco, S. V. Patel, and M. W. Jenkins, *J. Electrochem. Soc.*, **146**, 3907 (1999).
9. W. A. Goddard III, T. Cagin, M. Blanco, N. Vaidehi, S. Dasgupta, W. Floriano, M.

- Belmares, J. Kua, G. Zamanakos, S. Kashihara, M. Iotov, and G. H. Gao, *Comput. Theor. Polym. Sci.*, **11**, 329 (2001).
10. M. Belmares, M. Blanco, W. A. Goddard III, R. B. Ross, G. Caldwell, S.-H. Chou, J. Pham, P. M. Olofson, and C. Thomas, *J. Comput. Chem.*, **25**, 1814 (2004).
 11. A. Kurup, S. B. Mekapati, R. Garg, and C. Hansch, *Curr. Med. Res. Opin.*, **10**, 1679 (2003).
 12. J. W. Raymond, T. N. Rogers, D. R. Shonnard, and A. Kline, *J. Hazard. Mater.*, **84**, 189 (2001).
 13. R. Perkins, H. Fang, W. D. Tong, and W. Welsh, *Envir. Toxicol. Chem.*, **22**, 1666 (2003).
 14. D. Rogers and A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.*, **34**, 854 (1994).
 15. S. Brocchini, K. James, V. Tangpasuthadol, and J. Kohn, *J. Biomed. Mater. Res.*, **42**, 66 (1998).
 16. J. R. Smith, A. Seyda, N. Weber, D. Knight, S. Abramson, and J. Kohn, *Macromol. Rapid Commun.*, **25**, 127 (2004).
 17. S. D. Abramson, G. Alexe, P. L. Hammer, and J. Kohn, *J. Biomed. Mater. Res.*, **73A**, 116 (2005).
 18. D. Livingstone, *Data Analysis for Chemists*, Oxford University Press Inc., New York (1995).
 19. M. A. Ryan, M. L. Homer, H. Zhou, K. S. Manatt, and A. Manfreda, *Proceedings of the International Conference on Environmental Systems*, 2001-01-2308, Orlando, FL (2001).
 20. M. L. Homer, J. R. Lim, K. Manatt, A. Kisor, L. Lara, A. D. Jewell, A. V. Shevade, and M. A. Ryan, *Proceedings of IEEE Sensors*, **1**, 144 (2003).
 21. M. L. Homer, J. R. Lim, K. Manatt, A. Kisor, A. M. Manfreda, L. Lara, A. D. Jewell, S.-P. S. Yen, H. Zhou, A. V. Shevade, and M. A. Ryan, *Proceedings of IEEE Sensors*, **2**, 877 (2003).
 22. A. M. Manfreda, "Elucidating Humidity Dependence of the Jet Propulsion Laboratory's Electronic Nose Polymer-Carbon Composite Sensors," Research report submitted to the California State Polytechnic University-Pomona (2002).
 23. H. Zhou, M. L. Homer, A. V. Shevade, and M. A. Ryan, *Sensors*, **6**, 118 (2006).
 24. M. Blanco, *J. Comput. Chem.*, **12**, 237 (1991).
 25. F. R. Siperstein and A. L. Myers, *AIChE J.*, **47**, 1141 (2001).
 26. Cerius² v 4.2, Accelrys Inc., San Diego, CA (2000).
 27. A. V. Shevade, M. A. Ryan, M. L. Homer, A. M. Manfreda, H. Zhou, and K. Manatt, *Sens. Actuators B*, **93**, 84 (2003).
 28. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular Forces*, B. Pullman, Editor, p. 331 (Reidel, Dordrecht, The Netherlands, 1981).
 29. A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, **95**, 3358 (1991).
 30. S. L. Mayo, B. D. Olafson, and W. A. Goddard, *J. Phys. Chem.*, **94**, 8897 (1990).
 31. W. A. Steele, *The Interaction of Gases with Solids Surfaces*, Clarendon Press, Oxford (1974).
 32. J. W. Grate, S. N. Kaganove, and V. R. Bhethanabotla, *Anal. Chem.*, **70**, 199 (1998).
 33. J. W. Grate, M. H. Abraham, C. M. Du, R. A. McGill, and W. J. Shuely, *Langmuir*, **11**, 2125 (1995).
 34. E. J. Severin, B. J. Doleman, and N. S. Lewis, *Anal. Chem.*, **72**, 658 (2000).
 35. B. J. Doleman, E. J. Severin, and N. S. Lewis, *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 5442 (1998).
 36. M. C. Lonergan, E. J. Severin, B. J. Doleman, S. A. Beaver, R. H. Grubbs, and N. S. Lewis, *Chem. Mater.*, **8**, 2298 (1996).
 37. M. Janghorbani and H. Freund, *Anal. Chem.*, **45**, 325 (1973).
 38. NIST/TRC Vapor Pressure Database, version 2001.