EPA/600/R-03/033
March 2003

VERIFICATION AND VALIDATION OF THE SPARC MODEL

By

S.H. Hilal and S.W. Karickhoff
Ecosystems Research Division
U.S. Environmental Protection Agency
Athens, Georgia

and

L.A. Carreira
Department of Chemistry
University of Georgia
Athens, Georgia

National Exposure Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

DISCLAIMER

The United States Environmental Protection Agency through its Office of Research and Development partially funded and collaborated in the research described here under assistance agreement number 822999010 to the University of Georgia. It has been subjected to the Agency peer and administration review process and approved for publication as an EPA document.

ABSTRACT

SPARC (SPARC Performs Automated Reasoning in Chemistry) chemical reactivity models were validated on more than 5000 ionization pK_as (in the gas phase and in many organic solvents including water as a function of temperature), 1200 carboxylic acid ester hydrolysis rate constants (as a function of solvent and temperature), 350 $E_{1/2}$ reduction potential (as a function of solvents, pH and temperature), and 250 gas phase electron affinities. Physical properties have been developed and validated using more than 8000 physical property data points on many properties such as vapor pressure (as a function of temperature), boiling point (as a function of pressure), solubility, activity coefficients, Henry's constant and K_{ow} (as a function of solvent and temperature) , etc. However, the true validity of the SPARC property models is the ability of the SPARC basic models to be extended to calculate numerous chemical/physical properties (as a function of solvent, temperature, pressure, pH, etc.) without modification or extra parameterization to any of the SPARC basic models.

FOREWORD

Recent trends in environmental regulatory strategies dictate that EPA will rely heavily on predictive modeling to carry out the increasingly complex array of exposure and risk assessments necessary to develop scientifically defensible regulations. In response to this need, researchers at ERD-Athens have developed a predictive modeling system SPARC (SPARC Performs Automated Reasoning in Chemistry) that calculates a large number of physical and chemical properties from pollutant molecular structure across all classes of industrial organic chemicals. SPARC execution involves the classification of molecular structures and the selection and execution of appropriate "mechanistic" models, such as induction, resonance, and field effects to quantify reactivity. The basic mechanistic models in SPARC were designed and parameterized to be portable to any type of chemistry or organic chemical structure. This expanded prediction capability allows one to choose, for exhaustive validating, the reaction parameters for which large and reliable data sets do exist to validate against. The SPARC models have been validated on more than 12,000 data points for many properties. The verification and validation of the SPARC models will be presented in this report.

Rosemarie C. Russo, Ph.D.
Director
Ecosystems Research Division
Athens, Georgia

SUMMARY

The major differences among behavioral profiles of molecules in the environment are attributable to their physicochemical properties. For most chemicals, only fragmentary knowledge exists about those properties that determine each compound's environmental fate. A chemical-by-chemical measurement of the required properties is not practical because of expense and because trained technicians and adequate facilities are not available for measurement efforts involving thousands of chemicals. In fact, physical and chemical properties have only been measured for about 1 percent of the approximately 70,000 industrial chemicals listed by the U.S. Environmental Protection Agency's Office of Prevention, Pesticides and Toxic Substances [1].

Although a wide variety of approaches are commonly used in regulatory exposure and risk assessments, knowledge of the relevant chemistry of the compound in question is critical to any assessment scenario. For volatilization, sorption and other physical processes, considerable success has been achieved in not only phenomenological process modeling but also *a priori* estimation of requisite chemical parameters, such as solubilities and Henry's constant. Granted that considerable progress has been made in process elucidation and modeling for chemical processes, such as photolysis and hydrolysis, reliable estimates of the related fundamental physicochemical properties (i.e., rate and equilibrium constants) have been achieved for only a limited number of molecular structures. The values of these latter parameters, in most instances, must be derived from measurements or from the expert judgment of specialists in that particular area of chemistry.

Mathematical models for predicting the transport and fate of pollutants in the environment require reactivity parameter values--that is, the physical and chemical constants that govern reactivity. Although empirical structure-activity relationships that allow estimation of some constants have been available for many years, such relationships generally hold only within very limited families of

chemicals. On the other hand, we are developing computer programs that predict chemical reactivity strictly from molecular structure for virtually all organic compounds. Our computer system called SPARC (SPARC Performs Automated Reasoning in Chemistry) uses computational algorithms based on fundamental chemical structure theory to estimate a large array of physical/chemical parameters. See Table 1 for current SPARC physical property and chemical reactivity parameter estimation capabilities.

In every aspect of SPARC development, from choosing the programming environment to building model algorithms or rule bases, system validation and verification were important criteria. The basic mechanistic models in SPARC were designed and parameterized to be portable to any type of chemistry or organic chemical structure. This extrapolatability impacts system validation and verification in several ways. First, as the diversity of structures and the chemistry that is addressable increases, so does the opportunity for error. More importantly, however, in verifying against the theoretical knowledge of reactivity, specific situations can be chosen that offer specific challenges. This is important when verifying or validating performance in areas where existing data are limited or where additional data collection may be required. Finally, this expanded prediction capability allows one to choose, for exhaustive validating, the reaction parameters for which large and reliable data sets do exist to validate against. The SPARC models have been validated on more than 12,000 data points. The verification and validation of the SPARC models will be presented in this report.

Table 1. SPARC current physical and chemical properties estimation capabilities

Physical Property & Molecular Descriptor	Status	Reaction Conditions
Molecular Weight	Yes	
Polarizability	Yes	Temp
α, β H-bond	Yes	
Microscopic bond dipole	Yes	
Density	Yes	Temp
Volume	Yes	Temp
Refractive Index	Yes	Temp
Vapor Pressure	Yes	Temp
Viscosity	Mixed	Temp
Boiling Point	Yes	Press
Heat of Vaporization	Yes	Temp
Heat of formation	UD	Temp
Diffusion Coefficient in Air	Mixed	Temp, Press
Diffusion Coefficient in Water	Mixed	Temp
Activity Coefficient	Yes	Temp, Solv
Solubility	Yes	Temp, Solv
Gas/Liquid Partition Gas/Solid Partition Liquid/Liquid Partition Liquid /Solid Partition	Yes Mixed Yes Mixed	Temp, Solv Temp, Solv Temp, Solv Temp, Solv
GC Retention Times LC Retention Times	Yes Mixed	Temp, Solv Temp, Solv
Chemical Reactivity		
Ionization pK_a in Water Ionization pK_a in non-Aqueous Solution. Ionization pK_a in Gas phase Microscopic Ionization pK_a Constant Zwitterionic Constant Molecular Speciation Isoelectric Point	Yes Mixed Mixed Yes Yes Yes Yes	Temp, pH Temp, Solv Temp Temp, Solv, pH Temp, Solv, pH Temp, Solv, pH Temp, Solv, pH
Electron Affinity	Mixed	
Ester Carboxylic Hydrolysis Rate Constant	Yes	Temp , Solv
Hydration Constant	Mixed	Temp , Solv
Tautomer Constant	Mixed	Temp, Solv, pH
$E_{½}$ Chemical Reduction Potential	Mixed	Temp, Solv, pH

Mixed: Some capability exists but needs to be tested more, automated and/or extended.
Yes: Already tested and implemented in SPARC
UD: Under Development at this time
Press: Pressure, Temp: Temperature, Solv: Solvent
α: proton-donating site, β: proton-accepting site.

TABLE OF CONTENTS

INTRODUCTION

Recent trends in environmental regulatory strategies dictate that EPA will rely heavily on predictive modeling to carry out the increasingly complex array of exposure and risk assessments necessary to develop scientifically defensible regulations. The pressing need for multimedia, multistressor, multipathway assessments, from both the human and ecological perspectives, over broad spatial and temporal scales, places a high priority on the development of broad new modeling tools. However, as this modeling capability increases in complexity and scale, so must the inputs. These new models will necessarily require huge arrays of input data, and many of the required inputs are neither available nor easily measured. In response to this need, researchers at NERL-Athens have developed the predictive modeling system SPARC which calculates a large number of physical and chemical parameters from pollutant molecular structure and basic information about the environment (media, temperature, pressure, pH, etc.). Currently, SPARC calculates a wide array of physical properties and chemical reactivity parameters for organic chemicals strictly from molecular structure. See Table 1.

SPARC has been in use in the Agency programs for several years, providing chemical and physical properties to program offices (e.g., Office of Water, Office of Solid Waste and Emergency Response, Office of Prevention, Pesticides and Toxic Substances) and Regional Offices. Also, SPARC has been used in Agency modeling programs (e.g., the Multimedia, Multi-pathway, Multi-receptor Risk Assessment (3MRA) and LENS3, a multi-component mass balance model for application to oil spills) and to state agencies such as the Texas Natural Resource Commission. The SPARC web-based calculators have been used by many employees of various government agencies, academia and private chemical/pharmaceutical companies

throughout the United States. The SPARC web version performs approximately 50,000-100,000 calculations each month. (See the summary of usage of the SPARC web version in the Appendix).

PREVIOUS PEER REVIEWS OF THE SPARC SYSTEM

Over the lifespan of its development, the SPARC computer system has undergone numerous (and various types of) reviews that have helped to establish its validity. For example, we have published 10 journal articles on the SPARC computer system, each of which has undergone extensive peer-review (See references 2-11). Also, the SPARC computer system underwent an EPA Science Advisory Board (SAB) review in 1991, which was relatively early in its development stage. This multi-day review gave the SPARC development team the opportunity to demonstrate the system, and to discuss its modeling philosophy with experts in environmental science. Their comments on the system were very favorable, and they provided important input on further system development. Following is a brief excerpt from the SAB's written report. "*For a program still in development, progress is excellent. Resources should be made available to complete the documentation and conduct extensive testing of the model"* [12].

Also, the SPARC computer system has been included, along with other projects at our Division, in several major peer reviews, once in 1997 and again in 2000. These reviews were conducted by "blue-ribbon" panels of scientists from outside of EPA. Again, the comments on SPARC were laudatory and they provided important input to model development. For example, the following is an excerpt from the 1997 peer review panel's report on the SPARC project. *"The review panel is extremely impressed with the quality and productivity of this broad project, as well as with the presentation, the speaker, and the body of work that he summarized in such a remarkable fashion. This effort represents a central cog in the entire ERD program in environmental chemistry, as well as a key component of the Division's programmatic support to the broader agency. Moreover, this body of work represents a truly impressive service to the*

larger environmental chemistry community. EPA must find ways of providing permanent, long-term support for the commitment to this effort. It should also work toward making this service fully available via the internet. This project truly represents a key component of the NERL/Athens scientific endeavor" [13].

Furthermore, the SPARC developers have frequently engaged in informal consultation with leaders in relevant fields of science throughout the SPARC model design. These scientists include the late Dr. Robert Taft of the University of California, Dr. John Garst and Dr. Bruce King of the University of Georgia, Dr. Ralph Dougherty of the Florida State University and Dr. Samuel Yalkowski of the University of Arizona.

In summary, the SPARC system has now been extensively reviewed by many renowned scientists outside of EPA and in many different peer-review processes. Reviewer comments have always been favorable and the suggestions of these scientists have always been used to improve further model development. This type of "open communication" with leaders in various fields of science improves and helps to establish the validity of the SPARC models.

Although development of the SPARC program has been aimed at use in environmental assessments, these physicochemical models have widespread applicability in the academic and industrial communities For example, the SPARC program has been used at several universities as an instructional tool to demonstrate the applicability of physical organic models to the quantitative calculation of physicochemical properties (e.g., graduate class taught by the late Dr. Robert Taft at the University of California). Also, the SPARC calculator has great potential for aiding industry (such as Pfizer, Merck, Pharmacia & Upjohn, etc.) in the areas of chemical manufacturing and pharmaceutical and pesticide design.

ISSUES REGARDING VERIFICATION AND VALIDATION OF SPARC

To adequately convey the importance of verification and validation of SPARC models, it is necessary to first describe briefly, and in general terms, the SPARC modeling approach and philosophy. Indeed, it is useful to compare and contrast the SPARC approach to that of more conventional models for predicting physicochemical parameters.

Most models that predict a given physicochemical property (e.g., solubility, boiling point, etc.) are based, in a very direct way, on experimental data for that property for a limited training set of chemicals. Model development involves finding the best correlations between various descriptors of chemical structure and the observed property values. These descriptors are subsequently used to construct a model that adequately "recalculates" the training (or calibration) data set. Then, to validate, one must demonstrate that the empirical model also accurately predicts property values for chemicals not included in the training set, but whose experimental values are known. These data are often called the validation set. In order to predict a new physicochemical property (e.g., octanol/water partition coefficient), the entire process must be repeated, requiring new training and validation data sets for each new property.

On the other hand, with SPARC, experimental data for physicochemical properties (such as boiling point) are not used to develop (or directly impact) the model that calculates that particular property. Instead, physicochemical properties are predicted using a few models that quantify the underlying phenomena that drive all types of chemical behavior (e.g., resonance, electrostatic, induction, dispersion, H-bonding interactions, etc.). These mechanistic models were parameterized using a very limited set of experimental data, but not data for the end-use properties that will subsequently be predicted. After verification, the mechanistic models were

used in (or ported to) the various software modules that calculate the various end-use properties (such as boiling point). It is critical to recognize that the same mechanistic model (e.g., H-bonding model) will appear in all of the software modules that predict the various end-use properties (e.g., boiling point) for which that phenomenon is important. Thus, any comparison of SPARC-calculated physicochemical properties to an adequate experimental data set is a true model validation test -- there is no training (or calibration) data set in the traditional sense for that particular property. The results of validation tests on the various SPARC property models are presented below in the sections devoted to each property.

The unique approach to SPARC modeling also impacts our strategy for module verification. For example, when a mechanistic model is updated or improved by incorporating new knowledge, the impact on all of the various end-use parameters must be assessed. Toward this end, we have developed quality assurance software that executes each quarter. This software runs the various property modules for a large number of chemicals (4200 data-point calculations) and compares the output to historical results obtained over the life-span of the SPARC program. (Note that, early in our developmental stage, output of all SPARC modules were compared to hand calculations with selected chemicals to the extent possible. Satisfactory results were obtained prior to proceeding with further development). In this way, we ensure that existing parameter models still work correctly after new capabilities and improvements are added to SPARC. This also ensures that the computer code for all property and mechanistic models are fully operational. Since the same approach to verification was taken for all property modules, and since they are all driven by the same verified mechanistic models, we will not discuss verification in the following sections devoted to each property.

SPARC COMPUTAIONAL APPROACH

SPARC does not do "first principles" computation; rather, it analyzes chemical structure relative to a specific reactivity query much as an expert chemist might. SPARC utilizes directly the extensive knowledge base of organic chemistry. Organic chemists have established the types of structural groups or atomic arrays that impact certain types of reactivity and have described, in "mechanistic" terms, the effects on reactivity of other structural constituents appended to the site of reaction. To encode this knowledge base, a classification scheme was developed that defines the role of structural constituents in affecting or modifying reactivity. Furthermore, models have been developed that quantify the various "mechanistic" descriptions commonly utilized in structure-activity analysis, such as induction, resonance and field effects. SPARC execution involves the classification of molecular structure (relative to a particular reactivity of interest) and the selection and execution of appropriate "mechanistic" models to quantify reactivity. In brief, the SPARC model consists of a set of core models describing intra/intermolecular interactions that are linked by the appropriate thermodynamic relationships to provide estimates of reactivity parameters under desired conditions such as temperature, pressure and pH. The details of SPARC computational methods are presented in a companion U.S. E.P.A report, "Prediction of Chemical Reactivity Parameters and Physical Properties of Organic Compounds from Molecular Structure Using SPARC" [14]. Hence, only an overview will be given here.

For physical properties, intermolecular interactions are expressed as a summation over all the interaction forces between molecules (i.e., dispersion, induction, dipole and H-bonding). Each of these interaction energies is expressed in terms of a limited set of molecular-level descriptors (volume, molecular polarizability, molecular dipole, and H-bonding parameters) that, in turn, are

calculated from molecular structure. For chemical reactivity, molecular structure is broken into functional units. Reaction centers with known intrinsic reactivity are identified and the impact on reactivity of appended molecular structure is quantified using mechanistic perturbation models.

A "toolbox" of mechanistic perturbation models has been developed that can be implemented where needed for a specific reactivity query. Resonance models were developed and validated on more than 5000 light absorption spectra [1, 2], whereas electrostatic interaction models were developed and validated on more than 4500 ionization pK_as in water [3-8]. Solvation models (i.e., dispersion, induction, H-bond and dipole interactions) have been developed and validated on more than 8000 physical property data points on properties such as vapor pressure, boiling point, solubility, Henry's constant, GC chromatographic retention times, Kow, etc [3, 9, 10]. The SPARC computational approach is based on blending well known, established methods such as SAR (Structure Activity Relationships) [15, 16], LFER (Linear Free Energy Relationships) [17, 18] and PMO (Perturbed Molecular Orbital) theory [19, 20]. SPARC uses SAR for structure activity analysis, such as induction, resonance and field effects. LFER is used to estimate thermodynamic or thermal properties and PMO theory is used to describe quantum effects such as charge distribution delocalization energy and polarizability of the π electron network.

SPARC PHYSICAL PROPERTIES MODELS

For all physical properties (e.g., vapor pressure, boiling point, activity coefficient, solubility, partition coefficients, GC/LC chromatographic retention times, diffusion coefficients, etc.), SPARC uses one master equation to calculate characteristic process parameters:

$$\Delta G_{\text{Process}} = \Delta G_{\text{Interaction}} + \Delta G_{\text{Other}} \quad (1)$$

where $\Delta G_{Interaction}$ describes the change in the energy associated with the intermolecular interactions accompanying the process in question. For example, in liquid to gas vaporization, $\Delta G_{Interaction}$ describes the difference in the energy associated with intermolecular interactions in the gaseous phase versus that associated with interactions in the liquid phase. The intermolecular interaction forces between the molecules are assumed to be additive. The ΔG_{Other} lumps all non-interaction energy components such as entropy changes associated with mixing or expansion, and changes in internal molecular (vibrational, rotational) energies. At the present time, the intermolecular interactions in the liquid phase are modeled explicitly, interactions in the gas phase are ignored, and molecular interactions in the crystalline phase are extrapolated from the subcooled liquid state using the melting point. The 'non-interaction' entropy components are process specific and will be described later, in the vapor pressure and the activity coefficient models. The intermolecular interactions in the liquid phase are expressed as a summation over all the mechanistic components:

$$\Delta G_{Interaction} = \Delta G_{Dispersion} + \Delta G_{Induction} + \Delta G_{Dipole-dipole} + \Delta G_{H-Bond} \quad (2)$$

Each of these interaction mechanisms is expressed in terms of a limited set of pure component descriptors (liquid density-based volume, molecular polarizability, microscopic bond dipole, and hydrogen bonding parameters), which in turn are calculated strictly from molecular structure [3, 9].

Dispersion interactions are present in all molecules, including polar and non-polar molecules. Induction interactions are present between two molecules when at least one of them has a local dipole moment. Dipole-dipole interactions exist when both molecules have local dipole moments. H-bonding interactions exist when $\alpha_i \beta_j$ or $\alpha_j \beta_i$ products are non zero, where α

represents the proton donation strength and β represents the proton acceptor strength. In SPARC, all the physical property estimations derive from a common set of core models describing intra/intermolecular interactions, and require as user inputs molecular structure (both solute and solvent(s)) and reaction conditions of interest (temperature, pressure, etc.). The self-term, ΔG_{ii} (solute-solute) interaction model is used to describe the vapor pressure at 25° C. The self terms, ΔG_{ii} and ΔG_{jj} (solvent-solvent) plus the cross term, ΔG_{ij} (solute-solvent) interactions, are required to describe the solute, i, activity coefficient in any solvent, j at 25° C.

Like the chemical reactivity models, the $\Delta G_{ii,}$ ΔG_{ij} and ΔG_{jj} models have been extended and validated on numerous physical properties under different reaction conditions such as temperature, pressure and solvents. The self-term interaction model has been tested on a large number of vapor pressures, boiling points, diffusion coefficients and heat of vaporization. Likewise, the solute-solvent interaction model has been validated on activity coefficients, solubilities, partition coefficients and GC/LC chromatographic retention times in any solvent at any temperature.

Validation of the SPARC Refractive Index Model

The molecular polarizability and volume can be related to the index of refraction (n) using the Lorentz-Lorenz equation. For our units of cm^3/mole for volume (V) and $Å^3$/molecule for polarizability (P), the Lorentz-Lorenz equation can be written as

$$\frac{n^2-1}{n^2+2} = \frac{4\pi(0.6023P)}{3V} \quad (3)$$

The refractive index output was initially verified by comparing the SPARC prediction value to hand calculations for selected key molecules. The refractive index calculator was trained on 325 non-polar and polar organic compounds at 25^{o} C then validated on 578 organic compounds at 25^{o} C [9, 10] as shown in Figure 1. The statistical performance for the SPARC refractive index calculator is shown in Table 2. See reference 9 for sample hand calculations.

Table 2. SPARC Physical and Chemical Properties Calculator Statistical Performance versus Observations

Property	Units	Total # Molecule	RMS	R^2	Reaction Conditions Temp/Solvent
Refractive Index	N/A	578	0.007	0.997	25
Volume	g/cm^3	1440	1.97	0.999	25
Vapor Pressure	log atm	747	0.15	0.994	25
Boiling Point	^oC	4000	5.71	0.999	0.1-1520 torr
Heat of Vaporization[3]	Kcal/mole	1263	0.301	0.993	25, Boiling Point
Diffusion Coefficient in Air[4]	cm^2/s	108	0.003	0.994	25
Activity Coefficient	log MF[5]	491	0.064	0.998	25, 41 solvents
Solubility	log MF	647	0.40	0.987	25, 21 solvents
Distribution Coefficient	N/A	623	0.43	0.983	25 Octanol, Toluene CCl_4, Benzene, Cyclohexane, Ethyl Ether
Henry's Constant	M/L[6]	286 271	0.34 0.10	0.990 0.997	25, Water 25, Hexadecane
GC Retention Time[2]	Kovtas	295	10	0.998	25-190, Squalane, B18
LC Retention Time	Kovtas	125	0.095	0.992	25, Water/Methanol
Gas pK_a[3] Non-aqueous pK_a[3] pK_a in water	Kcal Kcal Kcal/1.36	400 300 4338	2.25 1.90 0.356	0.999 0.960 0.994	 25 , Alcohols, Aceteonitrile, Acetic acid, DMF[1], THF[1], pyridine 25-100, Water
Electron Affinity	e.V.	260	0.14	0.98	Gas
Ester Carboxylic Hydrolysis Rate	$M^{-1}s^{-1}$	1470	0.37	0.968	25-130, Water, Acetone, Alcohols, Dioxane, Aceteonitrile
Tautomer Constant[3] Hydration Constant[3]	Kcal/1.36 Kcal/1.36	36 27	0.3 0.43	0.950 0.744	25, Water
$E_{½}$ Chemical Reduction[3]	e.V	352	0.18	0.95	25, Water, Alcohols, DMF[1] Aceteonitrile, DMSO[1]

1 DMF: N,N'-dimethylforamide
DMSO: Dimethyl sulfoxide
HF: Tetrahydrofuran

2. GC retention times in SE-30, OV-101 and PEG-20M liquid stationary liquid phase is not included in this report.
3 See the companion SPARC report [14]
4. Models were developed after the HWIR exercise.
5. MF: mole fraction
6. M/L unit is (mole/L)/(mole/L); unitless

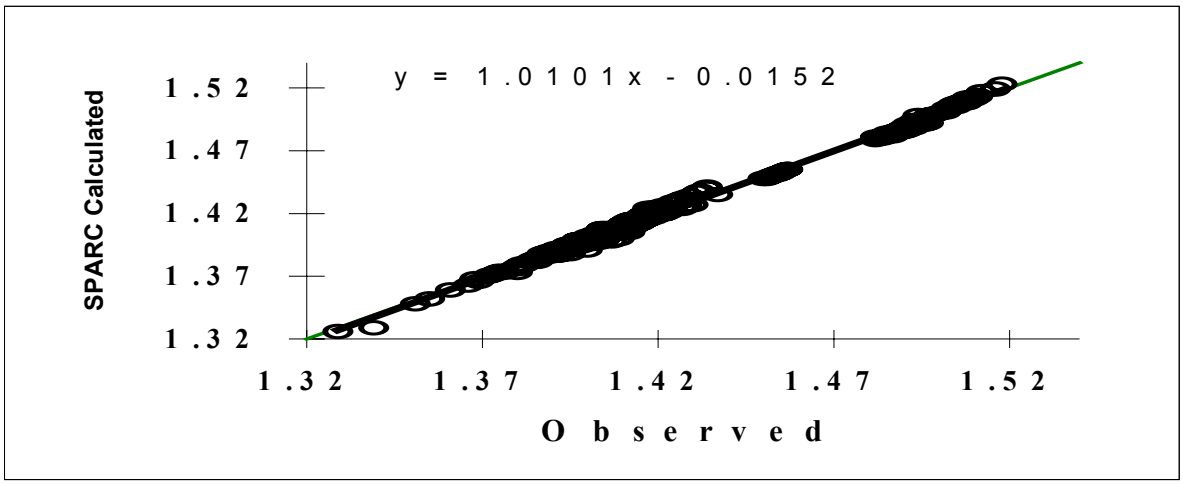

Figure 1. SPARC-calculated versus observed refractive index at 25° C. The RMS (Root Mean Square) deviation was 0.007 and R^2 was 0.997.

Validation of the SPARC Molecular Volume Models

The zero order density-based molecular volume at 25° C is expressed as

$$V_{25}^{o} = \sum_{i}(V_i^{frag} - A_i) \tag{4}$$

where V_i^{frag} is the volume of the i^{th} molecular fragment and A_i is a correction to that volume based on both the number and size of fragments attached to it. The V_i^{frag} are determined empirically from measured liquid-density based volumes, and then stored in the SPARC database. The zero order volume at 25° C is further adjusted for changes resulting from dipole-dipole and hydrogen bonding intermolecular interactions:

$$V_i = V_{25}^{o} + A_{dipole-dipole} \frac{\sum_i D_i}{V_{25}^{o}} + A_{H-bond} \frac{\alpha_i \beta_i}{V_{25}^{o}} \tag{5}$$

where D_i is the weighted sum of the local dipole for the molecule, and α and β are the H-bonding parameters of potential proton donor and proton acceptor sites within the molecule, respectively. $A_{dipole\text{-}dipole}$ and $A_{H\text{-}bond}$ are adjustment constants due to dipole-dipole and H-bonding, respectively. The final molecular volume at any temperature T is then expressed as a polynomial expansion in (T-25) corrected for H- bonding, dipole density and polarizability density interactions [9, 14].

The molecular volume can be calculated within 2 cm^3 $mole^{-1}$ for most organic molecules. Figure 2 shows the SPARC-calculated versus observed molecular volumes for both polar and non-polar compounds at 25° C. The statistical performance for the volume calculator is in Table 2. See reference 9 for sample hand calculations.

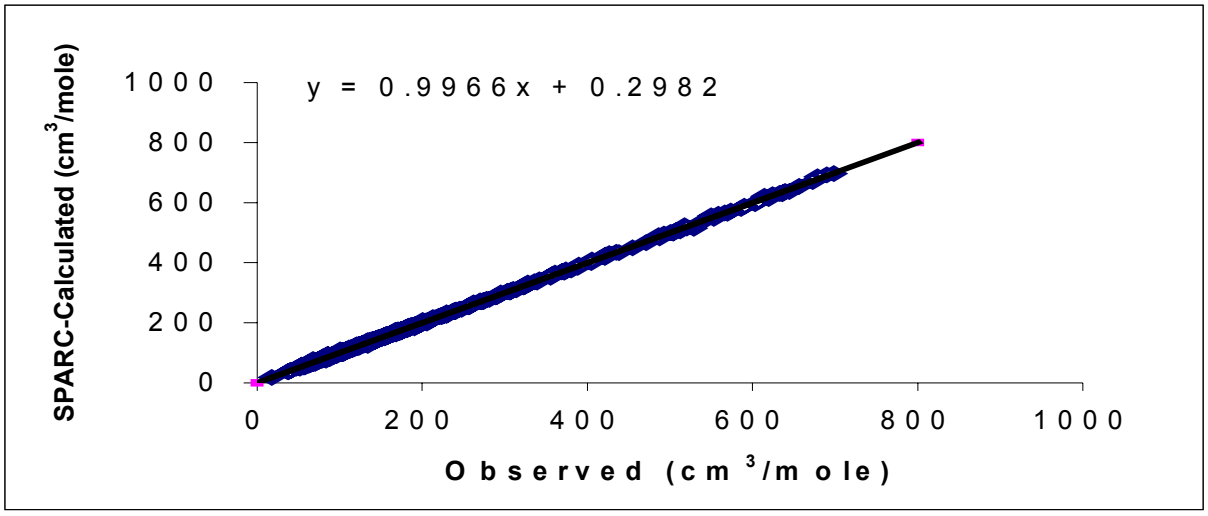

Figure 2. SPARC-calculated vs. observed-liquid density based volume at 25° C for 1440 organic molecules. The RMS deviation was 1.97 cm^3 $mole^{-1}$ and R^2 was 0.999.

Validation of the SPARC Vapor Pressure Model

The saturated vapor pressure is one of the most important physiochemical properties of pure compounds. By 1978, vapor pressure data (as a function of temperature) were available for more

than 7000 organic compounds [21]. Despite the frequency of reporting in the published literature, the number of compounds where the vapor pressure was truly measured and not extrapolated to 25° C from higher temperature measurements, is limited. Most of the measured 25° C vapor pressure data are for compounds that are either pure hydrocarbons or molecules that have relatively small dipole moments and/or weak hydrogen bonds. There is a pressing need to predict the vapor pressure of those compounds that have not been measured experimentally. In addition to being highly significant in evaluating a compound's environmental fate, the vapor pressure at 25° C provides an excellent arena for developing and testing the SPARC self interaction physical process models.

The vapor pressure, vp^{o}_{i} of a pure solute, i, can be expressed as function of all the intermolecular interaction mechanisms, ΔG_{ii} (interaction), as

$$\log vp_i^o = \frac{-\Delta G_{ii}(\mathit{Interaction})}{2.303RT} + LogT + C \qquad (6)$$

where log (T) + C describes the change in the entropy contribution associated with the volume change in going from the liquid to the gas phase. The crystal energy term (given in reference 14), CE, must be added to equation 6 for molecules that are solids at 25° C, the CE contribution becomes important, especially for rigid structures such as aromatic or ethylenic molecules that have high melting points [14].

The vapor pressure computational algorithm output was initially verified by comparing the SPARC prediction of the vapor pressure at 25° C to hand calculations for key molecules. Since the SPARC self interactions model, ΔG_{ii}, was developed initially on this property, the vapor pressure model undergoes the most frequent validation tests. The calculator was trained on 315 non-polar

and polar organic compounds at 25° C. Figure 3 presents the SPARC-calculated vapor pressure at 25° C versus measured values for 747 compounds. The SPARC self-interactions model can predict the vapor pressure at 25° C within experimental error over a wide range of molecular structures and measurements (over 8 log units). For simple structures, SPARC can calculate the vapor pressure to better than a factor of 2. For complex structures such as some of the pesticides and pharmaceutical drugs where dipole-dipole and/or hydrogen bond interactions are strong, SPARC calculates the vapor pressure within a factor of 3-4. The statistical performance for the vapor pressure calculator is shown in Table 2. See references 9 and 14 for sample hand calculations. The vapor pressure model was also tested on the boiling point and heats of vaporization [9, 14].

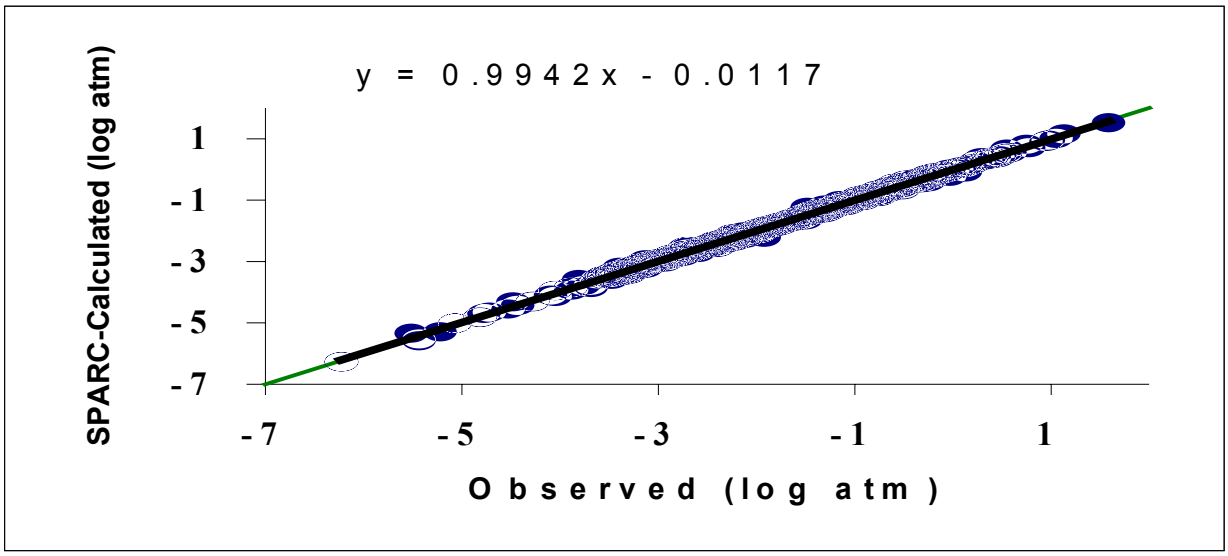

Figure 3. SPARC-calculated vs. observed log vapor pressure for 747 organic molecules at 25° C. The figure includes all the vapor pressure measurements (real not extrapolated) we found in the literature. The RMS deviation error was 0.15 log atm and R^2 was 0.994.

Validation of the SPARC Boiling Point Model

SPARC estimates the boiling point for any molecular species by varying the temperature at which a vapor pressure calculation is done. When the vapor pressure equals the desired

pressure, then that temperature is the boiling point at that pressure. The normal boiling point is calculated by setting the desired pressure to 760 torr. Boiling points at a reduced pressure can be calculated by setting the desired pressure to a different value.

SPARC temperature dependence models were developed initially on the boiling point. The boiling point calculator was trained on 1900 boiling points for a wide range of non-polar and polar organic compounds. The calculator was validated against 4000 boiling points measured at different pressures ranging from 0.05 to 1520 torr spanning a range of over 800° C as shown in Figure 4.

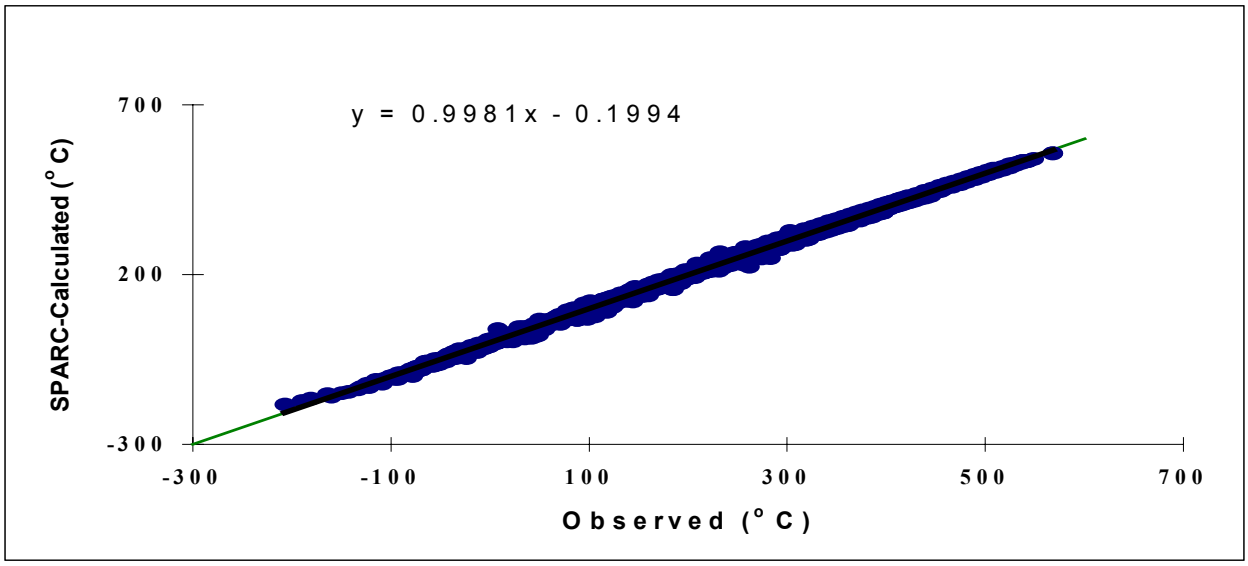

Figure 4. SPARC-calculation versus observed 4000 boiling points for pressure ranging from 0.1 to at 1520 torr. The Total RMS deviation was 5.71° C. The RMS deviation for polar molecules was 8.2° C and R^2 was 0.9988, while for non-polar molecules the RMS was 2.6° C and R^2 was 0.9995.

Validation of the SPARC Activity Coefficient Model

For a solute, i, in a liquid phase, j, at infinite dilution, SPARC expresses the activity coefficient as

$$-RT\ \log\gamma_{ij}^{\infty} = \sum\Delta G_{ij}\,(Interaction) + RT\ (\log\frac{V_i}{V_j} + \frac{(\frac{V_i}{V_j} - 1)}{2.303}) \qquad (7)$$

where V_i and V_j are the molecular volume of the solute and the solvent, respectively. The last term is the Flory-Huggins excess-entropy-of-mixing contribution in the liquid phase for placing a solute molecule in the solvent [3, 14].

The activity coefficient computational algorithm output was initially verified by comparing the SPARC prediction to hand calculations for key molecules. The SPARC activity coefficient calculator was trained on 211 activities for a wide range of organic molecules. Figure 5 presents the validation for SPARC-calculated log activity coefficients versus measured values for 491 compounds at 25^o C in 41 different solvents. The SPARC activity coefficient test statistical parameters are shown in Table 2. The activity coefficients calculator was also tested on the solubility in more than 20 different solvents and partition coefficients in more than 18 different solvents. See following sections for more details.

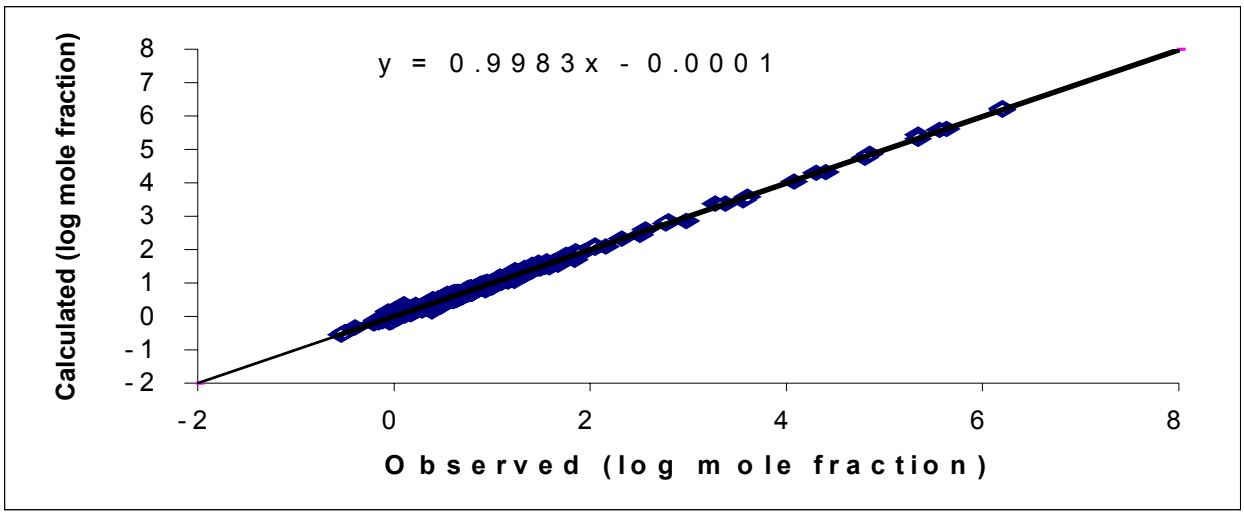

Figure 5. SPARC-calculated versus observed log activity coefficients at infinite dilution for 491 compounds in 41 solvents including water. Only 15% of these compounds have strong dipole-dipole and/or H-bond interactions. The RMS deviation was 0.064 log mole fraction and with an R^2 of 0.998.

Validation of the SPARC Solubility Model

SPARC does not calculate the solubility from first principles, but rather from the infinite dilution activity coefficient model discussed previously. SPARC first calculates the infinite dilution activity coefficient, γ^{∞}; when log γ^{∞} is greater than 2, the mole fraction solubility can be reliably estimated as $\chi^{sol} = 1/\gamma^{\infty}$. However, when the log γ^{∞} is calculated to be less than 2, this approximation fails. In these cases, γ^{∞} is greater than γ^{sol} and SPARC would underestimate the solubility using the inverse relationship. In order to overcome this limitation, SPARC employs an iterative calculation. SPARC sets the initial guess of the solubility as $\chi^{sol}_{guess} = 1/\gamma^{\infty}$. SPARC then 'prepares' a mixed solvent that is χ^{sol}_{guess} in the solute and (1- χ^{sol}_{guess}) in the solvent. SPARC then recalculates γ^{∞} in the 'new' solvent and the corresponding χ^{sol}_{guess}. This process is continued until γ^{∞} converges to 1 (miscible). The solubility calculator spans more than 12 log mole fraction as shown in Figure 6. The RMS deviation was 0.40 log mole fraction, which was close to the experimental error. SPARC estimates the solubility for simple organic molecules to better than a factor of 2 (0.3 log mole fraction) and within a factor of 4 (0.6 log mole fraction) for complicated molecules like pesticides and pharmaceutical drugs. The RMS deviation for the solids compounds is 3 times greater than the RMS deviation for liquids compounds due to the crystal energy contributions. For more details see reference 14. The statistical parameters for calculated log solubility for 647 organic molecules in 21 different solvents including water at 25^{o} C are shown in Table 2.

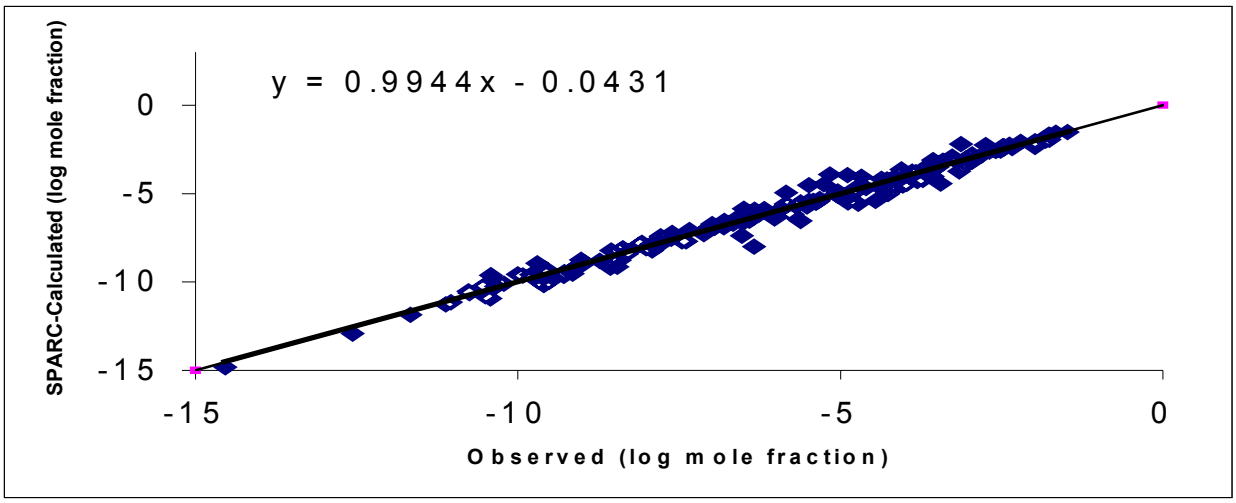

Figure 6. Test results for SPARC calculated log solubilites for 260 compounds. The RMS deviation is 0.321 and R^2 is 0.991. The RMS deviation for 119 liquid soluibilties is 0.135 and R^2 is 0.997 while for the 141 solids compounds the RMS deviation is 0.419 and R^2 is 0.985.

Validation of the SPARC Mixed Solvents Model

SPARC can handle solvent mixtures for a large number of components. However, speed and memory requirements usually limit the number of solvent components to less than twenty on a PC. The user specifies the name and volume fraction for each solvent component. Each of the solvent components must have been previously initialized as a solvent. SPARC will allow the user to specify a name for the mixture so that it can be used later as a 'known' solvent. The activity coefficients (or solubility) of molecules in binary solvent mixtures have been tested and appear to work well. Figure 7 shows the calculated log γ in a water/methanol mixture versus measured values. For more details see reference 14.

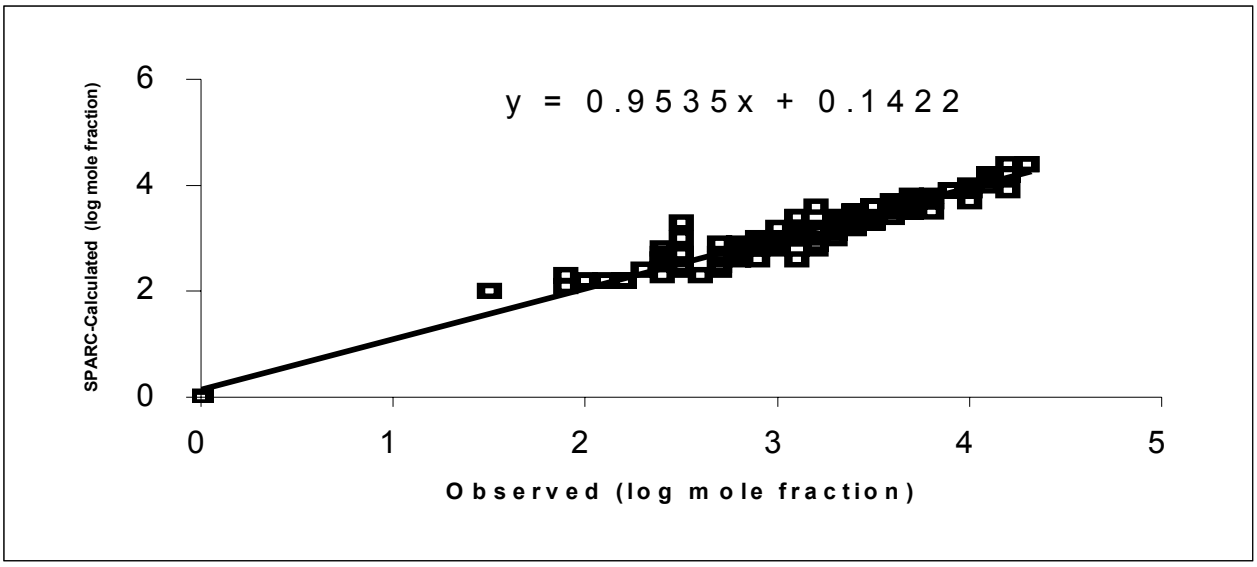

Figure 7. SPARC-calculated versus observed log activities for 120 compounds in water/methanol mixed solvent at 25^{o} C. The RMS deviation error was 0.18 and the R^2 was 0.980.

Validation of the SPARC Partition Constants Models

All partition (Liquid/Liquid, Liquid/Solid, Gas/Liquid, Gas/Solid) constants are determined by calculating the activity coefficient of the molecular species of concern in each of the phases without modification or extra parameterization to the activity coefficient model.

Gas/liquid (Henry's Constant) Model

Henry's constant may be expressed as

$$H_x = vp_i^o \ \gamma_{ij}^{\infty} \quad (8)$$

where vp_i^o is the vapor pressure of pure solute i (liquid or subcooled liquid) and γ_{ij}^{∞} is the activity coefficient of solute (i) in the liquid phase (j) at infinite dilution. SPARC vapor pressure and activity coefficient models can be used to calculate the Henry's constant for any solute out of a

mixed solute-solvent liquid phase. An application of Henry's law constant for the prediction of gas-liquid chromatography retention time is given in the companion SPARC report [14].

Liquid/Liquid Partitioning Model

SPARC calculates the liquid/liquid partition constant, such as the octanol/water distribution coefficient, by simply calculating the activity of the molecular species in each of the liquid phases as

$$\log K_{liq1/liq2} = \log \gamma^{\infty}_{liq2} - \log \gamma^{\infty}_{liq1} + \log R_m \quad (9)$$

where the γ^{∞}s are the infinite dilution activities in the two phases and R_m is the ratio of the molecularites of the two phases (M_1/M_2). Although octanol/water partition coefficients are widely used and measured, the SPARC system does not limit itself to this calculation. SPARC can calculate the liquid/liquid partition coefficient for any two immiscible phases.

Gas/Solid Partitioning Model

SPARC calculates gas/solid partitioning in a manner similar to gas/liquid partitioning. For the solid phase, the solvent self-self interactions, ΔG_{jj}, are dropped from the calculation when one of the phases is solid. This type of modeling will be useful for calculating retention times for capillary column gas chromatography.

Liquid/Solid Partitioning Model

SPARC calculates liquid/solid partitioning in a manner similar to liquid/liquid partitioning. For the solid phase, the solvent self-self interactions, ΔG_{jj}, are dropped from the calculation.

The gas/liquid models have been extensively tested against observed Henry's constant measurements. The two largest data sets are air/water and air/hexadecane systems. The liquid/solid and gas/solid partitioning models are implemented in code but have not been extensively tested. The liquid/liquid partitioning models are the most extensively tested partitioning models due to the large octanol/water data sets available. The statistical parameters for SPARC-calculated partition constants in many solvents at 25° C are shown in Table 2. Figure 8 shows calculated versus observed Henry's constant for compounds dissolved in hexadecane. Figure 9 shows the current general performance of SPARC for log $K_{solvent/water}$, where the solvents are carbon tetrachloride, benzene, cyclohexane, ethyl ether, octanol and toluene. Figure 10 displays a comparison of the EPA Office of water (OW) recommended observed octanol-water distribution coefficients versus SPARC and C log P calculated values. The RMS deviation and R^2 values were is 0.18 and 0.996 respectively for SPARC and 0.44 and 0.978 respectively for ClogP calculated values [22].

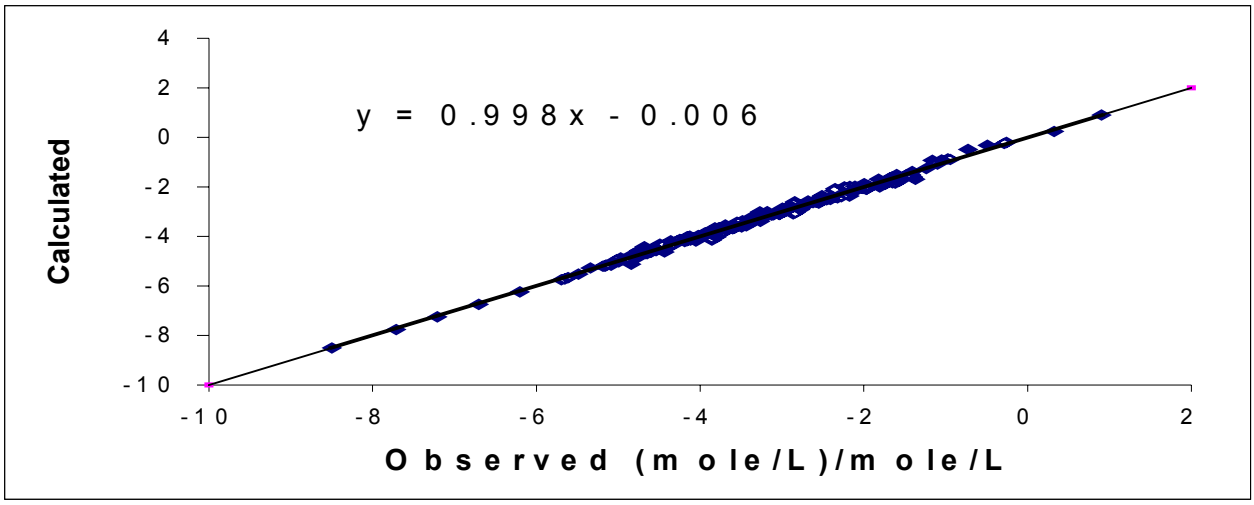

Figure 8. Observed vs. SPARC-calculated Henry's constants for 271 organic compounds in hexadecane. The RMS deviation was 0.1, while the R^2 was 0.997.

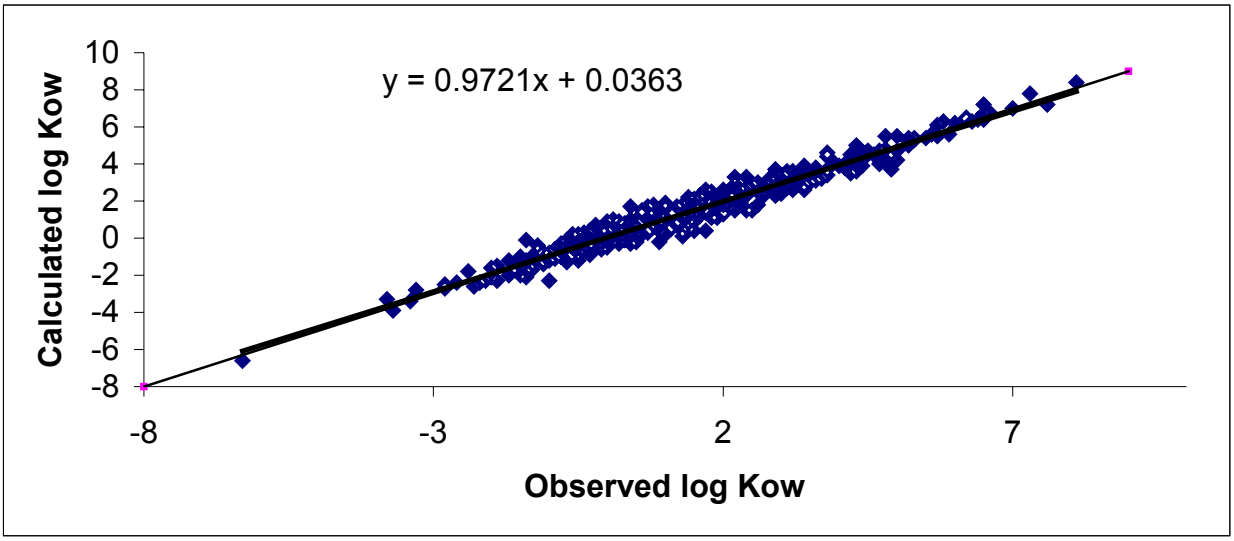

Figure 9. SPARC-calculated versus observed log distribution coefficients $K_{solvent/water}$ for 623 organic compounds in six solvents at 25° C. The RMS deviation was 0.38 and R^2 was 0.983.

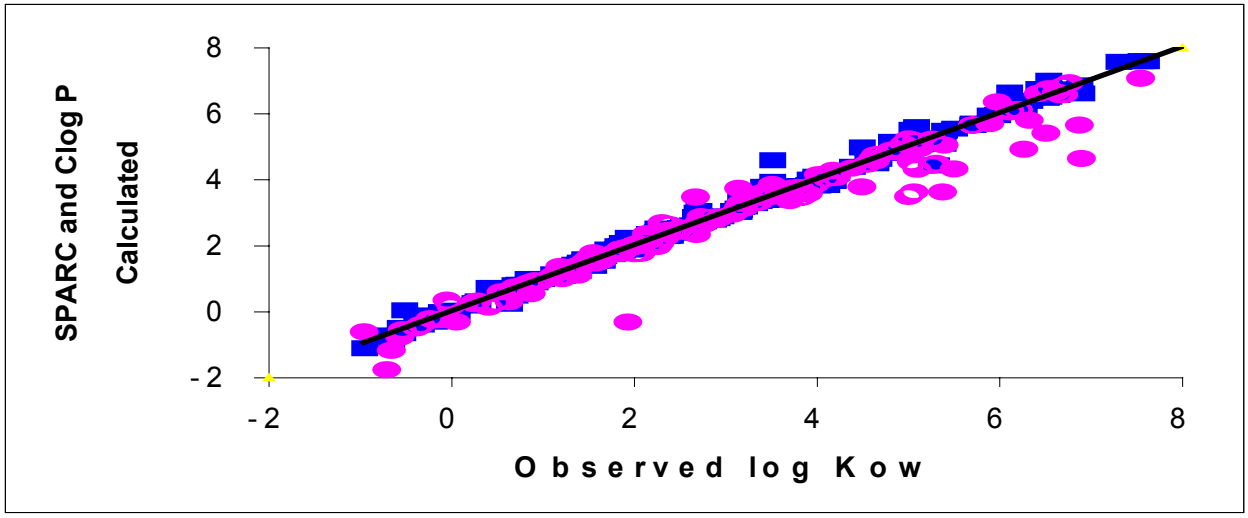

Figure 10. Test of OW for calculated $K_{octanol/water}$ versus measured values. Squares are SPARC calculate values, circles are ClogP calculate values. The RMS deviation and R^2 values were 0.18 and 0.996 respectively for SPARC and 0.44 and 0.978 respectively for ClogP calculated values

Validation of the SPARC Diffusion Coefficient in Air Model

Several engineering equations exist that do a very respectable job of calculating molecular diffusion coefficients in air over wide ranges of temperature and pressure. The equation most compatible with the SPARC calculator is also the relationship that seems to perform the best over

a wide variety of molecules. This equation is that of Wilke and Lee [23], which for binary diffusion coefficient is expressed as:

$$D_{AB} = [3.03 - (0.98 / M_{AB}^{1/2})](10^{-3}) \frac{T^{3/2}}{P M_{AB}^{1/2} \sigma_{AB}^{2} \Omega_D} \quad (10)$$

where D_{AB} is the binary diffusion coefficient in cm^2/s, T is the temperature in K, M_A and M_B are the molecular weights of A and B in g/mol and M_{AB} is $2[(1/M_A) + (1/M_B)]^{-1}$ and P is the pressure in bar. The Ω_D is a complex function of T^* and has been accurately determined by Neufeld [24].

SPARC predicts gas phase binary diffusion coefficients at any temperature and pressure to better than 6% as shown in Figure 11. The statistical parameters are in Table shown 2.

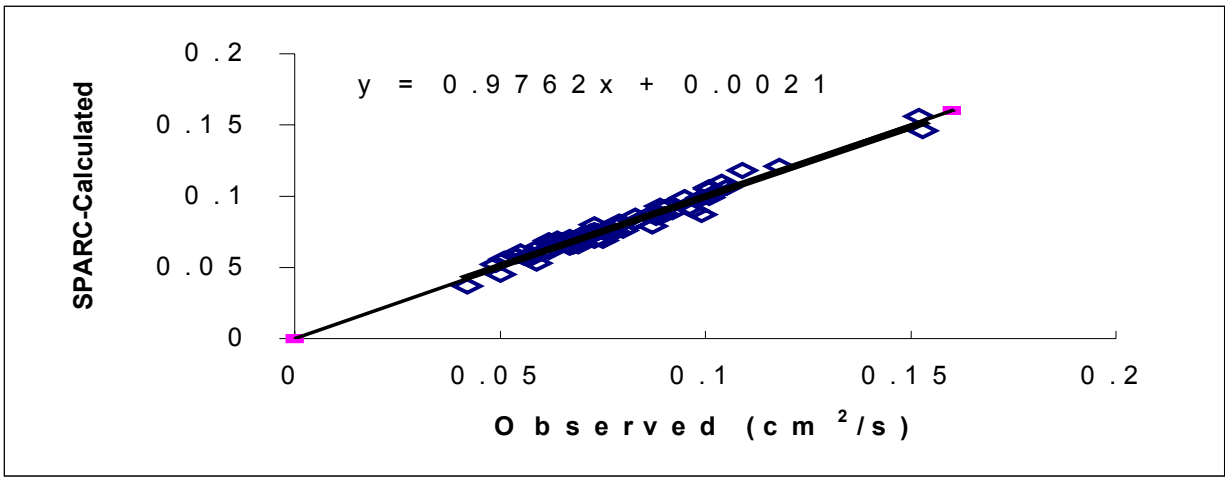

Figure 11. SPARC-calculated vs. observed diffusion coefficient. The RMS deviation was 0.003.

The overall SPARC physical properties training set output is shown in Figure 12. The training set includes vapor pressure (as a function of temperature), boiling point (as a function of pressure), diffusion coefficients (as a function of pressure and temperature), heat of vaporization (as function of temperature), activity coefficient (as a function of solvent), solubility (as a function of

solvent and temperature), GC retention times (as a function of stationary liquid phase and temperature) and partition coefficients (as a function of solvent). This set includes more than 50 different pure solvents (see Table 3) as well as 18 mixed solvent systems. The observed measured values for the training and validations sets were from many sources such as references 26-34.

For other SPARC physical properties models such as GC/LC retention time in polar and non-polar liquid phase, heat of vaporization and diffusion coefficient in water, see reference 14.

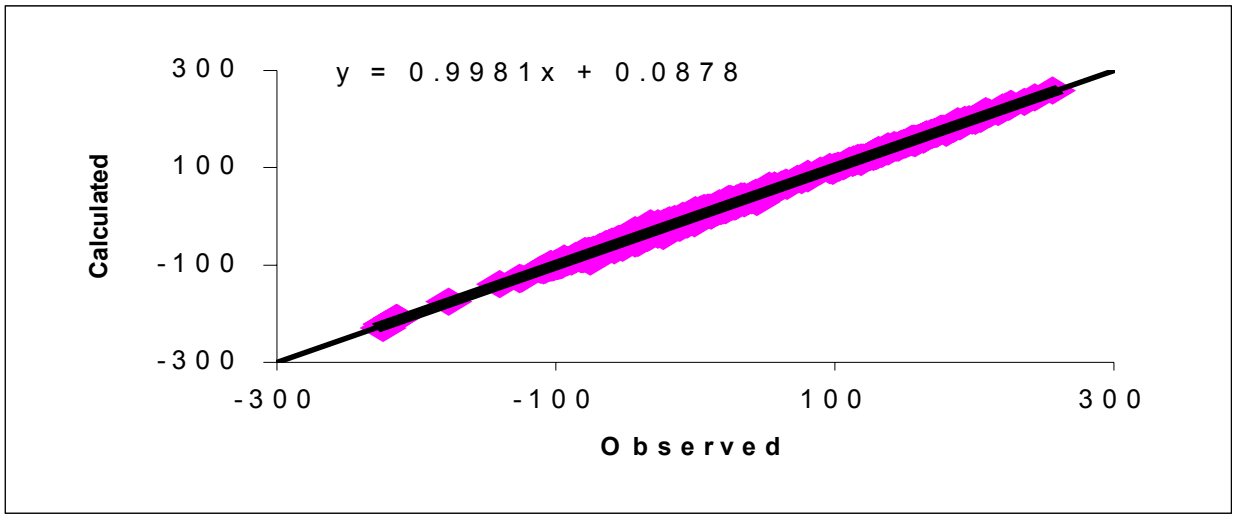

Figure 12. SPARC-calculated vs. 2400 observed training set physical property values. The aggregate RMS is 0.29 and R^2 is 0.997. For more details see text.

Table 3. Solvents that have been tested in SPARC

Chloroform	1-butanol	1-chloro hexadecane	1-dodecanol	OV-101
1-propanol	butanone	1-nitro propane	2-dodecanone	isopropanol
isobutanol	acetone	2-nitro propane	aceteonitrile	PEG-20M
benzyl ether	benzene	benzylchloride	benzonitrile	SE-30
cyclohexane	decane	bromobenzene	butronitrile	pyridine
cyanohexane	ethanol	dioctyl ether	cyano cyclohexane	water
heptane	hexane	hexadecane	heptadecane	squalane
methanol	nonane	1-butyl chloride	nitrobenzene	1-me naphthalene
nitroethane	octane	nitro cyclohexane	nitro methane	2-me naphthalene
nonanenitrile	squalene	pentadecane nitrile	isoquinoline	m-cresol
quinoline	phenol	1,2,4 trichlorobenzene	hexafluorobenzene	p-xylene

SPARC CHEMICAL REACTIVITY MODELS

SPARC reactivity models have been designed and parameterized to be portable to any chemical reactivity property and any chemical structure. For example, chemical reactivity models are used to estimate ionization pK_a, zwitterionic constant, isoelectric point and speciation fractions as a function of pH. The same reactivity models are used to estimate gas phase electron affinity and ester hydrolysis rate constants in water and in non-aqueous solutions.

Validation of the SPARC pK_a in water Models

Like all chemical reactivity parameters addressed in SPARC, molecular structures are broken into functional units called the reaction center and the perturber in order to estimate pK_a in water. The reaction center, C, is the smallest subunit that has the potential to ionize and lose a proton to a solvent. The perturber, P, is the molecular structure appended to the reaction center, C. The pK_a of the reaction center is adjusted for the molecule in question using the mechanistic perturbation models. The pK_a for a molecule of interest is expressed in terms of the contributions of both P and C.

$$pK_a = (pK_a)_c + \delta_p(pK_a)_c \quad (11)$$

where $(pK_a)_c$ describes the ionization behavior of the reaction center, and $\delta_p(pK_a)_c$ is the change in ionization behavior brought about by the perturber structure given as

$$\delta_p(pK_a)_c = \delta_{ele}\,pK_a + \delta_{res}\,pK_a + \delta_{sol}\,pK_a + \ldots \quad (12)$$

where $\delta_{res}pK_a$, $\delta_{ele}pK_a$ and $\delta_{sol}pK_a$ describe the differential resonance, electrostatic and solvation effects of P on the initial and final states of C, respectively.

The SPARC pK_a calculator was trained on 2500 organic molecules, then validated on 4338 pK_a's (4550 including carbon acid) in water as shown in Figure 13 and Table 4. The calculator was tested for multiple ionization's up to the 6th (simple organic molecules) and 8th (azo dyes) for molecules with multiple ionization sites. In addition, the pK_a models were tested on all the literature values we found for zwitterionic constants (12 data points), the thermodynamic microscopic ionization constants, pk_i, of molecules with multiple ionization sites (120 measurement data points, the RMS deviation error is 0.5), the corresponding complex speciation as a function of pH and the isoelectric points (29 measurement data points) in water. The diversity and complexity of the molecules used was varied over a wide range in order to develop more robust models during the last few years. Hence, the SPARC pK_a models are now very robust and highly tested against almost all the available experimental literature data.

While it is difficult to give a precise standard deviation of a SPARC calculated value for any given individual molecule, in general SPARC can calculate the pK_a for simple molecules such as oxy acids and aliphatic bases and acids within ±0.25 pK_a units; ±0.36 pK_a units for most other organic structures such as amines and acids; and ±0.41 pK_a units for =N and in-ring N reaction centers and for complicated structures. Where a molecule has more than six ionization sites (n > 6), the expected SPARC error is ±0.65 pK_a units. For more details see reference 14.

Table 4. Statistical Parameters of SPARC pK_a Calculations

Set	Training	R^2	RMS	Test	R^2	RMS
Simple organic compounds	793	0.995	0.235	2000	0.995	0.274
Azo dyes compounds	50	0.991	0.550	273	0.990	0.630
IUPAC compounds[1]	2500	0.994	0.356	4338[2]	0.994	0.370

1. Observed values are from many ref. such as 35-36
2. Carbon acid pK_as are not included

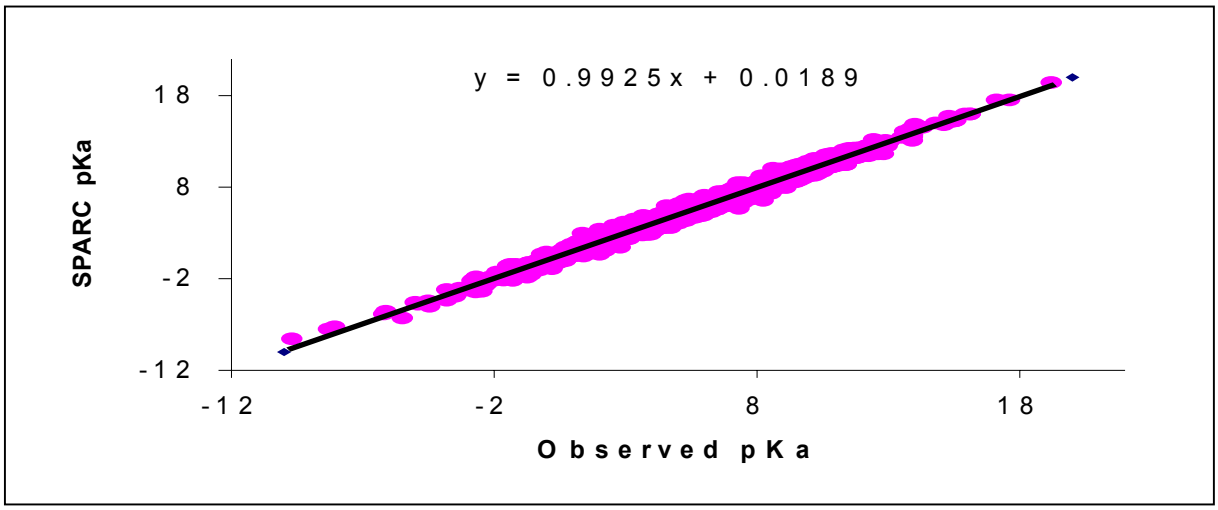

Figure13. SPARC-calculated versus observed for 4338 pK_a's of 3685 organic compounds. The RMS deviation was equal to 0.37. This test does not include carbon acid reaction center. The majority of the molecules are complex compounds. Some of the molecules such as azo dyes have 8 different ionization sites.

Validation of the SPARC Carboxylic Acid Ester Hydrolysis Rate Constant Models

Reaction kinetics were quantitatively modeled within the chemical equilibrium framework described previously for ionization pK_a in water. It was assumed that a reaction rate constant could be described in terms of a pseudo equilibrium constant between the reactant and transition states. SPARC therefore follows the modeling approach described for pK_a. For these chemicals, reaction centers with known intrinsic reactivity are identified and the reaction rate constants expressed by perturbation theory as

$$\log k = \log k_c + \Delta_p \log k_c \qquad (13)$$

where log k is the log of the rate constant of interest; log k_c is the log of the intrinsic rate constant of the reaction center and $\Delta_p \log k_c$ denotes the perturbation of the log rate constant due to the appended structure.

The ester hydrolysis rate constant models have been tested to the maximum extent possible as function of temperature and solvent. The RMS deviation error for 1470 hydrolysis rate constants in 6 solvents and at different temperature was 0.37 as shown in Figure 14. In this test, there were 653, 667 and 150 base, acid and general base catalyzed calculations performed as shown in Table 5 [14, 25].

Table 5. Statistical Parameters of SPARC Calculated Hydrolysis Rate Constants ($M^{-1}s^{-1}$)

Solvent	Base			Acid			Gbase		
	No	RMS	R^2	No	RMS	R^2	No	RMS	R^2
Water	142	0.39	0.98	383	0.36	0.98	51	0.34	0.98
Acetone/Water	143	0.34	0.83	208	0.33	0.96	73	0.36	0.96
Ethanol/Water	105	0.29	0.83	39	0.17	0.98	9	0.1	0.99
Methanol/Water	150	0.36	0.78	22	0.22	0 .95	N/A		
Dioxnae/Water	90	0.47	0.75	15	0.16	0.87	17	0.47	0.67
Aceteonitrile/Water	24	0.3	0.97	N/A			N/A		
Total Molecules	654	0.37	0.96	667	0.37	0.97	150	0.39	0.97

The observed-measured values are from many references such as 37-40

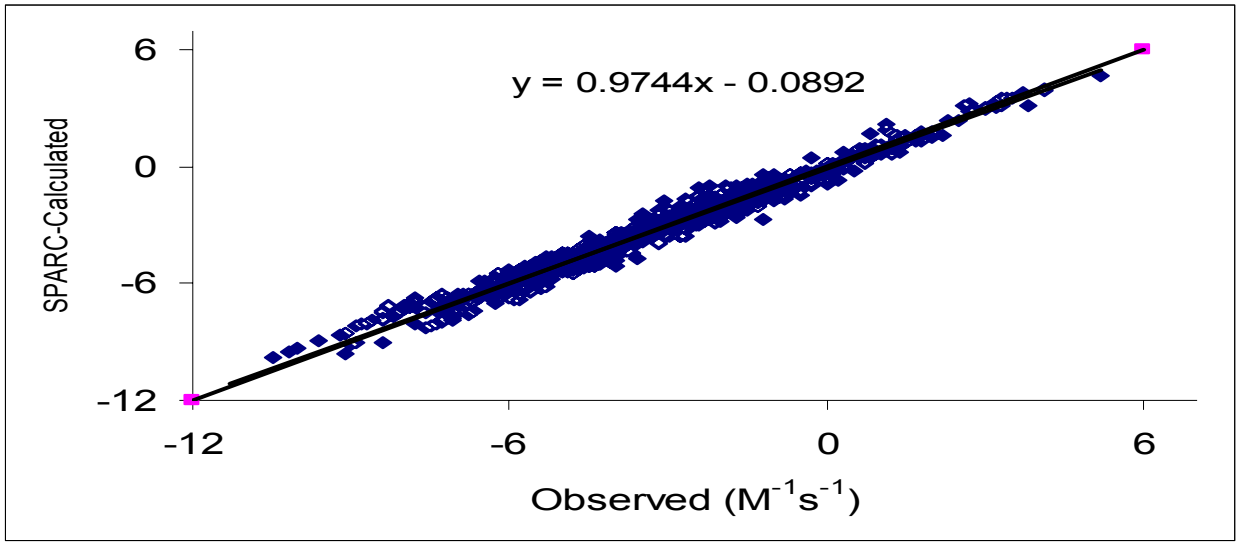

Figure 14. SPARC-calculated versus observed hydrolysis rate constants for base, acid and general base in six different solvents and at different temperatures. The aggregate RMS was 0.37.

Validation of the SPARC Electron Affinity (EA) Models

As was the case for pK_a, the SPARC computational procedure starts by locating the potential sites within the molecule at which a particular reaction of interest could occur. In the case of EA these reaction centers, C, are the smallest subunit(s) that could form a molecular negative ion. Any molecular structure appended to C is viewed as a "perturber" (P). EA as expressed in terms of the summation of the contributions of all the components, perturber(s) and reaction center(s), in the molecule:

$$EA = \sum_{c=1}^{n} [(EA)_c + \delta_p(\Delta EA)_c] \qquad (14)$$

where the summation is over n, which is defined as the number of reaction centers in the molecule. $(EA)_c$ is the electron affinity for the reaction center. $\delta_p(\Delta EA)_c$ is a differential quantity that describes the change in the electron affinity behavior affected by the perturber structure.

In the estimation of EA, there was no modifications to any of the pK_a models or any extra parameterization for P to calculate electron affinity from ionization pK_a models other than inferring the electronegativity and the electron affinity susceptibility of the reaction centers (C) to electrostatic and resonance effects [4].

The EA models have been tested to the maximum extent possible on all the gas phase electron affinity measurements reported by Kebarle, McIver and Wentworth [4]. The RMS deviation for the 260 EA's was 0.14 e.V. and R^2 was 0.98 as shown in Figure 15. The statistical parameters are shown in Table 2.

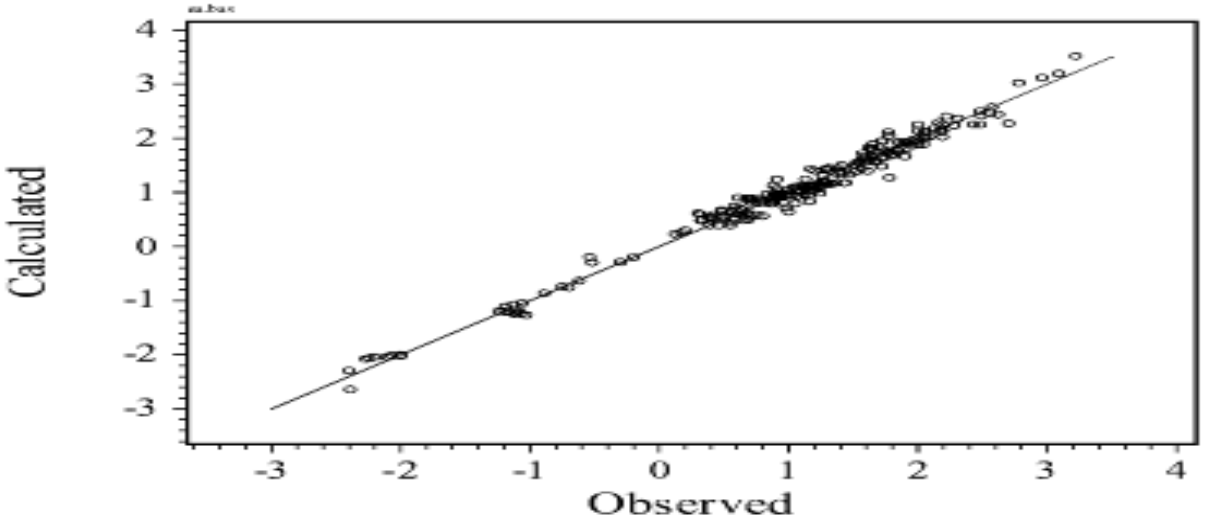

Figure 15. SPARC-calculated versus observed electron affinity for 260 organic compounds. The RMS deviation was 0.14 e.V. and R^2 was 0.98.

MONOPOLE MODELS (IONIC SPECIES)

The SPARC models were extended to ionic organic species by incorporating monopole (charge) electrostatic interaction models to SPARC's physical properties toolbox. These ionic models play a major role in modeling and estimating Henry's constant for charged (ionic) species in any solvent system. These capabilities (ionic activity) in turn allow SPARC to calculate gas phase pK_a, and non-aqueous ionization pK_a and $E_{1/2}$ chemical reduction in any solvent system.

Validation of the SPARC Monopole Models

The SPARC monopole models have been tested on all the available data for Henry's constant for charged molecules in water, unfortunately there was only 12 data points. However, the SPARC Ionization pK_a in water coupled with Henry's constant for charged molecules was used to estimate 400 pK_a's in the gas phase and 300 pK_a's in non-aqueous solvents. Also, SPARC electron affinity calculator coupled with Henry's constant for charged molecules was used to

estimate 352 $E_{1/2}$ chemical reduction data measurements. See Table 2 and for more details see reference 14.

QUALITY ASSURANCE

A quality assurance (QA) plan was developed to recalculate all the aforementioned physical and chemical properties and compare each calculation to an originally-calculated value stored in the SPARC databases. Every quarter, two batch files that contain more than 3000 compounds (4200 calculations) recalculate various physical and chemical properties. QA software compares every single "new" output to the SPARC originally-calculated-value dating back to 1993-1999. This ensures the integrity of the SPARC model as new features are added.

CONCLUSION

The strength of the SPARC chemical reactivity parameters and physical properties calculator is the ability to estimate numerous properties for a wide range of organic compounds within an acceptable error, especially for molecules that are difficult to measure. The SPARC physical properties/chemical reactivity parameters calculator prediction is as reliable as most of the experimental measurements for these properties. For simple structures, SPARC can calculate a property of interest within a factor of 2 or even better. For complex structures where dipole-dipole and/or H-bond interactions are strong, properties can generally be calculated within a factor of 3-4.

The true validity of the SPARC physical/chemical property models does not lie in the models' predictive capability for pK_a, or solubility, but is determined by the extrapolatability of these same models to other types of chemistry. The ability of SPARC models to be extended to various chemical/physical properties without modification or extra parameterization to any of the basic models, provides great confidence in this powerful calculation tool.

APPENDIX

Summary of usage of the SPARC-web version

Two months back-to-back report, which represents the usage of the SPARC calculator in October and November, 2002. November was the highest while October was the lowest usage to date.

Summary of Activity for Report

October 2002	**November 2002**
Hits Entire Site (Successful) 56,875 Average Number of Hits per day on Weekdays 2,153 Average Number of Hits for the entire Weekend 1,297 Most Active Day of the Week Thu Least Active Day of the Week Sat Most Active Day Ever October 24, 2002 Number of Hits on Most Active Day 4,963 Least Active Day Ever October 05, 2002 Number of Hits on Least Active Day 7 **URL's of most active users** 207.168.147.52 463 p120x183.tnrcc.state.tx.us 3,986 141.189.251.7 1,720 198.137.21.14 455 57.67.16.50 327 gateway.huntingdon.com 6,823 aries.chemie.uni-erlangen.de 1,487 p120x226.tnrcc.state.tx.us 67 thompson.rtp.epa.gov 413 webcache.crd.GE.COM 143	**Hits Entire Site (Successful) 95,447** Average Number of Hits per day on Weekdays 4,146 Average Number of Hits for the entire Weekend 842 Most Active Day of the Week Wed Least Active Day of the Week Sun Most Active Day Ever November 13, 2002 Number of Hits on Most Active Day 15,450 Least Active Day Ever November 02, 2002 Number of Hits on Least Active Day 7 **URL's of most active users** 141.189.251.7 1,223 gw.bas.roche.com 1,821 gateway.huntingdon.com 3,729 p120x183.tnrcc.state.tx.us 737 hwcgate.hc-sc.gc.ca 660 p120x226.tnrcc.state.tx.us 379 thompson.rtp.epa.gov 563 chen.rice.edu 966

SPARC is online and can be used at http://ibmlc2.chem.uga.edu/sparc

REFERENCES

1. S. W. Karickhoff, V. K. McDaniel, C. M. Melton, A. N. Vellino, D. E. Nute, and L. A. Carreira., US. EPA, Athens, GA, EPA/600/M-89/017.

2. S. W. Karickhoff, V. K. McDaniel, C. M. Melton, A. N. Vellino, D. E. Nute, and L. A. Carreira., *Environ. Toxicol. Chem*. 10 1405 1991.

3. S. H. Hilal, L. A. Carreira and S. W. Karickhoff, "Theoretical and Computational Chemistry, Quantitative Treatment of Solute/Solvent Interactions", Eds. P. Politzer and J. S. Murray, Elsevier Publishers, chapter 9, 1994.

4. S. H. Hilal, L. A. Carreira, C. M. Melton and S. W. Karickhoff, *Quant. Struct. Act. Relat,* 12 389 1993.

5. S. H. Hilal, L. A. Carreira, C. M. Melton, G. L. Baughman and S. W. Karickhoff, *J. Phys. Org. Chem.* 7, 122 1994.

6. S. H. Hilal, L. A. Carreira and S. W. Karickhoff., *Quant. Struct. Act. Relat.* 14 348 1995.

7. S. H. Hilal, L. A. Carreira, S. W. Karickhoff , M. Rizk, Y. El-Shabrawy and N. A. Zakhari, *Talanta,* 43 , 607 1996.

8. S. H. Hilal, L. A. Carreira and S. W. Karickhoff, *Talanta.,* 50 827 1999.

9. S. H. Hilal, L. A. Carreira, S. W. Karickhoff, *J. Chromatogr.,* 269 662 1994.

10. S. H. Hilal, L. A. Carreira and S. W. Karickhoff, *Accepted, Quant. Struct. Act. Relat.*

11. S. H. Hilal, J.M Brewer, L. Lebioda and L.A. Carreira, *Biochem. Biophys. Res. Com.,* 607 211 1995

12. SAB Report, *Evaluation of EPA's Research on Expert Systems to Predict the Fate and Effects of Chemicals*, November 1991.

13. Peer Review of the Research Programs of the Ecosystems Research Division. U.S. EPA, NERL, Athens, Ga, June, 1997.

14. S. H. Hilal, a companion U.S. E.P.A report, *"Prediction of Chemical Reactivity Parameters and Physical Properties of Organic Compounds from Molecular Structure Using SPARC".*

15. J. E. Lemer and E.Grunwald, *Rates of Equilibria of Organic Reactions,* John Wiley & Sons, New York, NY, 1965.

16. Thomas H. Lowry and Kathleen S. Richardson, *Mechanism and Theory in Organic Chemistry*. 3ed ed., Harper & Row, New York, NY, 1987.

17. L. P. Hammett, *Physical Organic Chemistry*, 2nd ed. McGraw Hill, New York, NY, 1970.

18. R.W. Taft, *Progress in Organic Chemistry,* Vol.16, John Wiley & Sons, New York, NY, 1987.

19. M. J. S. Dewar, *The Molecular Orbital Theory of Organic Chemistry,* McGraw Hill, New York, NY, 1969.

20. M. J. S. Dewar and R. C. Doughetry, *The PMO Theory of Organic Chemistry,* Plenum Press, New York, NY, 1975.

21. J. Dykyj, M. Repas and J. Anmd Svoboda., *Vapor Pressure of Organic substances.* VEDA, Vydavatel Stvo, Slovenskej Akademie Vied, Bratislava, 1984.

22. S. W. Karickhoff and MacArthur Long, *US. EPA Internal Report*, April 10 1995.

23. C. R. Willke and C. Y. Lee, *Ind. Eng. Chem*. 47 1253 1955.

24. P. D. Neufeld, A. R. Janzen and R. A. Aziz, J. *Chem. Phys*. 57 1100 1972.

25. S. H. Hilal, L. A. Carreira and S. W. Karickhoff, *To be Submitted*.

26. R. C. Reid, J. M. Prausnitz and J. K. Sherwood, *The Properties of Gases and Liquids*, 3ed ., McGraw-Hill Book Co., 1977.

27. R. R. Dreisbach *Physical Properties of Chemical Compounds: Advanced in Chemistry Series,* Dow Chemical Co., ACS, Washington, D.C., (A) Volume I, 1955, (B) Volume II, 1959, (C) Volume III, 1961.

28. R. C. Wilhoit and B. J. Zwolinski, *J. Phys. Chem. Ref. Data*, 2 1, 1973. Supplement No. 1.

29. T. E. Jordan, *The Vapor Pressure of Organic Compounds*, Interscience Publisher Inc, Philadelphia, Pennsylvania, 1954.

30. R. Weast and M. Astle, CRC *Handbook of Chemistry and Physics, 79th ed.,* CRC Press Inc., West Palm Beach, 1999.

31. D. Mackay, W. Y. Shiu and K.C Ma*, Illustrated Handbook of Physical/Chemical Properties and Environmental Fate of Organic Chemicals*, Lewis Publishers, volume I, II, III, 1993.

32. Douglas Hartley and Hamish Kidd, *The Agro Chemical Handbook,* Royal Society of Chemistry, University of Nottingham, England, 1983.

33. S. R. Heller, D.W. Bigwood, P. Laster and K. Scott, *The ARS Pesticide Properties Database*, Maryland, U.S.A.

34. H. A. Hornsby, D. R. Wauchope and E. Albert Herner, *Pesticide Properties in the Environment,* Springer, New York, NY, 1996.

35. E. P. Serjeant and B. Dempsey, *Ionization Constants of Organic Acids in Aqueous Solution*, Pergamon Press, Oxford, 1979.

36. (A) D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution*, Butterworth & Co, London, 1965 & Supplement 1972. (B) Supplement 1972.

37. N. B. Chapman, *J. Chem. Soc*., 1291 1963.

38. L. W. Deady and R. A. Shanks, *Aust. J. Chem.*, 25, 2363 1972.

39. M. L. Bender and Robert J. Thomas, *J. Am. Chem. Soc*., 83 4189 1961.

40. DeLos DeTar and Carl J. Tenpas, *J. Am. Chem. Soc*., 7903 1976.