

# Calculating physical properties of organic compounds for environmental modeling from molecular structure

S. H. Hilal · A. N. Saravananaraj · T. Whiteside ·  
L. A. Carreira

Received: 16 May 2007 / Accepted: 28 August 2007 / Published online: 8 November 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** Mathematical models for predicting the transport and fate of pollutants in the environment require reactivity parameter values – that is the value of the physical and chemical constants that govern reactivity. Although empirical structure–activity relationships have been developed that allow estimation of some constants, such relationships are generally valid only within limited families of chemicals. The computer program, SPARC, uses computational algorithms based on fundamental chemical structure theory to estimate a large number of chemical reactivity parameters and physical properties for a wide range of organic molecules strictly from molecular structure. Resonance models were developed and calibrated using measured light absorption spectra, whereas electrostatic interaction models were developed using measured ionization  $pK_{as}$  in water. Solvation models (i.e., dispersion, induction, H-bonding, etc.) have been developed using various measured physical properties data. At the present time, SPARC's physical property models can predict vapor pressure and heat of vaporization (as a function of temperature), boiling point (as a function of pressure), diffusion coefficient (as a function of pressure and temperature), activity coefficient, solubility, partition coefficient and chromatographic retention time as a function of solvent and temperature. This prediction capability crosses chemical

family boundaries to cover a broad range of organic compounds.

**Keywords** Physical properties · Molecular interaction · Vapor pressure · Activity coefficient · Partition coefficients · SPARC · SAR

## Introduction

As the evaluation and management of the risk to humans associated with the proliferation of anthropogenic chemicals in the environment becomes an increasingly urgent social priority, there accrues a corresponding demand on physical and biological scientists and engineers to provide effective techniques for quantifying the fate, potential exposure and environmental and health damage resulting from the production, release and use of such chemical compounds. Industrially developed nations have established government agencies to assess the risk to human health and the environment resulting from the use these chemicals. Agency risk assessors must identify and understand the key chemical and physical properties influencing the fate of organic chemicals in the environment, along with data relating to their toxicity and persistence, when conducting regulatory risk assessments.

Historically, health and safety regulators at the federal level in the United States have relied on field monitoring, toxicological test data and expert scientific knowledge to condemn or vindicate a given chemical. Recent emphasis, however, on more quantitative and comprehensive risk and benefit analyses has required the development of more sophisticated evaluation methods. These methods must be capable of forecasting pollutant behavior over wide ranges of environmental conditions, often without the benefit of

---

S. H. Hilal (✉)  
Ecosystems Research Division, National Exposure Research  
Laboratory, US Environmental Protection Agency, 960 College  
Station Road, Athens, GA, USA  
e-mail: hilal.said@epa.gov

A. N. Saravananaraj · T. Whiteside · L. A. Carreira  
Department of Chemistry, University of Georgia, Athens,  
GA, USA

measured data specific to the chemical or ecosystem in question. Although a wide variety of approaches are commonly used in regulatory exercises, a 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 Law constants [1–6]. Granted that considerable progress has been made in process elucidation and modeling for chemical processes [7–9], such as photolysis and hydrolysis, reliable estimates of the related fundamental chemical constants (i.e., rate and equilibrium constants) have been achieved only for a limited number of molecular structures. The values of parameters, in most instances, still must be derived from measurements or from the expert judgment of specialists in that particular area of chemistry. Trained technicians or adequate facilities are not likely to be made available for measuring thousands of chemicals. Therefore, reliable, cost-effective techniques for estimating the physical and chemical properties are needed for reducing the uncertainty in exposure and risk assessments. In response to this need, we have developed a predictive modeling computer program, SPARC (SPARC performs automated reasoning in chemistry). SPARC calculates a wide array of physical properties and chemical reactivity parameters for all classes of industrial organic chemicals based on the pollutant's molecular structure as function of environmental conditions.

## SPARC

The computational approach in SPARC is derived from a common set of core mechanistic perturbation models that describe the intra/intermolecular interactions between molecules as a function of environmental conditions. A “toolbox” of mechanistic perturbation models has been developed based on fundamental chemical structure theory to quantify the intra/intermolecular interaction effects on organic compounds that can be implemented whenever needed to predict a physiochemical property of interest. A classification scheme in SPARC defines the role of structural constituents in affecting the organic compound's 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 property of interest) and the selection and execution of the appropriate “mechanistic” models to quantify that property.

The SPARC mechanistic models are designed and parameterized so as to be portable, in principle, to any type of physical/chemical property and/or molecular structure. Hence, SPARC estimates numerous physical properties with greater accuracy, and with a wider scope than is possible with other estimation techniques. The SPARC mechanistic perturbation models have been developed and validated for a large number and wide variety of molecular structures. Resonance perturbation models were developed and calibrated using measured light absorption spectra [10]; electrostatic interaction perturbation models were developed using measured ionization  $pK_a$ s in water [11]. Self interaction (solute–solute) solvation perturbation models (i.e., dispersion, induction, H-bonding and dipole–dipole) have been developed using measured vapor pressures and boiling points [12]; the solute–solvent interaction models use measured activity coefficients [13]. Both the resonance and electrostatic models have been validated on measured ionization  $pK_a$  in water, hydrolysis rate constant, and gas-phase electron affinity [14, 15]. Likewise, the solvation models were validated for the prediction of heat of vaporization [12], solubility, distribution coefficient, Henry's constant and GC retention time [13, 16].

SPARC costs the user only a few minutes of computer time and provides greater accuracy and a broader scope than is possible with conventional estimation techniques. The user needs to know only the molecular structure of the compound to predict a property of interest. The user provides the program with the molecular structure either by direct entry in Simplified Molecular Input Line Entry System (SMILES) notation, or via the CAS number, which will generate the SMILES notation. SPARC is programmed with the SICSTUS version of PROgramming in LOGic (Prolog).

The SPARC web-based calculators have been used by many employees of various government agencies, academia and private chemical and pharmaceutical companies. SPARC has been in use in 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) model and LENS3, a multi-component mass balance model for application to oil spills) and in state agencies such as the Texas Natural Resource Commission. Also, the SPARC calculator has been used for aiding industry in the areas of chemical manufacturing and pharmaceutical and pesticide design. The SPARC web version has approximately 250,000 hits per month. The SPARC web version is available for public use (free of charge) and can be accessed at <http://sparc.chem.uga.edu>.

The majority of the calculations performed by outside users have been for the estimation of  $pK_a$  (speciation as a function of pH), vapor pressure, solubility, distribution coefficient and Henry's constant. The types of compounds for which the calculations are performed are unknown as SMILES strings are erased after use.

### Physical properties modeling approach

SPARC expresses the free energy change of a characteristic process such as the vapor pressure or activity coefficient (as well as all other physical properties addressed in SPARC) as

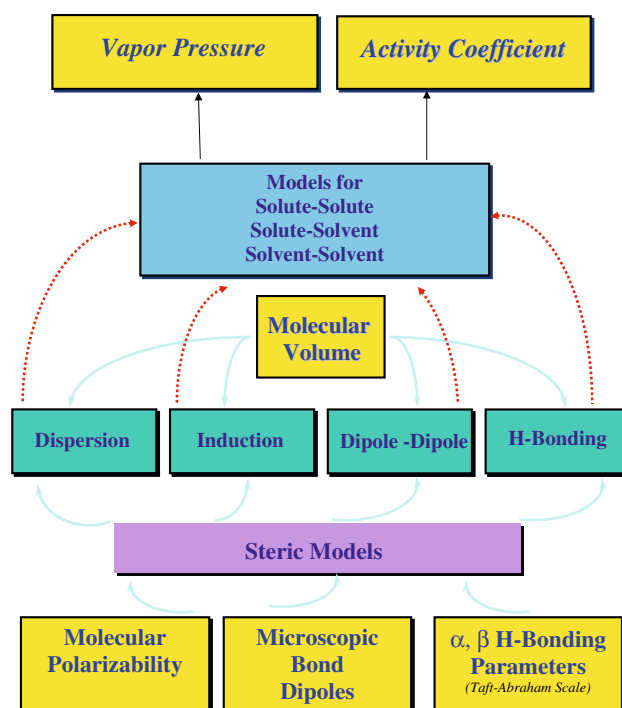
$$\Delta G_{\text{Process}} = \Delta G_{\text{Interaction}} + \Delta G_{\text{Other}} \quad (1)$$

where  $\Delta G_{\text{Interaction}}$  describes the change in the intermolecular interactions accompanying the process in question. For example, in the estimation of the vapor pressure of a solute molecule, i,  $\Delta G_{\text{Interaction}}$  describes the difference in the self (solute–solute, ii) intermolecular interactions in the gaseous versus liquid phases. Similarly, in quantifying the activity coefficient associated with placing a solute molecule, i, into the solvent, j,  $\Delta G_{\text{Interaction}}$  describes the difference in the self (solute–solute, ii and solvent–solvent, jj) and solute–solvent (ij) intermolecular interactions in the gaseous versus liquid phase. 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 relative to the melting point. The intermolecular interactions in the liquid phase are expressed as a summation over all the mechanistic components:

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

Dispersion interactions are present in all molecules, both polar and non-polar. 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  $\alpha$  represents the proton donation strength and  $\beta$  represents the proton acceptor strength. Each interaction mechanism is expressed in terms of four molecular descriptors; molecular volume, molecular polarizability, microscopic bond dipole, and H-bonding parameters as shown in Fig. 1.

The  $\Delta G_{\text{Other}}$  term describes all the non-interaction components, such as the entropy changes associated with mixing or expansion, volume changes, or changes in internal (vibrational, rotational) energies going from the



**Fig. 1** Development of SPARC basic physical properties (vapor pressure and activity coefficient) models from intermolecular interaction models. Each interaction mechanism is calculated from molecular descriptors that is molecular volume, molecular polarizability, microscopic bond dipole, and  $\alpha$ ,  $\beta$  H-bonding parameters. These interaction models are designed and parameterized so as to be portable to any type of physical/chemical property or molecular structure

initial state to the final state. The 'non-interaction' entropy components are process specific; it depends on the phase change involved and in the present application is presumed to depend only on the volume changes of the solute/solvent in each phase. It will be described briefly for the vapor pressure and activity coefficient models later.

### Molecular-level descriptors

The four molecular-level descriptors (molecular volume, molecular polarizability, microscopic bond dipole, and H-bonding parameters) are calculated strictly from molecular structure using SPARC. The computational approach for molecular descriptors is constitutive with the molecule in question being broken at each essential single bond and the property of interest calculated as a linear combination of fragment contributions as

$$\chi^o(\text{molecule}) = \sum_i (\chi_i^o - A_i) \quad (3)$$

where  $\chi_i^o$  are intrinsic fragment contributions (most cases are tabulated in SPARC databases) and  $A_i$  are adjustments

relating to steric or electrometric perturbations due to contiguous structural elements within the molecule in question. In some instances,  $\chi^{\circ}$  (molecule) is further adjusted for a specific process model or medium involved. Both  $\chi_i^{\circ}$  and  $A_i$  are empirically trained, either on direct measurements of the descriptor in question (e.g., liquid density-based volume) or on a directly related property (e.g., index of refraction) for which large, reliable data sets exist. For example, the molecular polarizability descriptor was trained on index of refraction, the molecular volume descriptor was trained on liquid density-based molecular volume at 25 °C, while the hydrogen bonding descriptor was trained on  $\alpha$  and  $\beta$  parameters (similar to Taft–Abraham H-bonding scale). The results of SPARC refractive index and the molecular volume calculators (directly related property) are shown in Table 1. For more details and for sample calculation of molecular polarizability, index of refraction and molecular volume are given in previous publications [17]

### Solute/solvent interactions

In SPARC all property estimations derive from a common set of core models describing intra/intermolecular

interactions, and require as user inputs molecular structure (solute and solvent(s)) and reaction conditions of interest (temperature, pressure, etc.). Models for self interactions between like molecules (solute–solute) and between solvent–solute molecules have been developed to calculate numerous physical properties from molecular structure. The interactions are either symmetrical (dispersion) or asymmetrical (Induction, dipole–dipole, H-bonding) interactions. For asymmetrical interactions that involve molecular orientation, an enthalpy change is involved in the disruption of the solute–solute and solvent–solvent interactions and their replacement by solute–solvent interactions. If the interactions between like molecules are smaller than the solute–solvent interactions,  $\Delta H_{\text{mix}}$  is negative,  $\Delta G_{\text{mix}}$  decreases and the solute particles will dissolve in the solvent; i.e., the solutions are miscible. On the other hand, if the interactions between like molecules are more favorable than the solute–solvent interactions,  $\Delta H_{\text{mix}}$  is positive, its contributions to  $\Delta G_{\text{mix}}$  may overcome the favorable entropy term, and the liquids will be immiscible. Based on the above, the solute particles form a cavity or sphere inside the solvent. The total interaction free energy change is then calculated as the interaction between like molecules combined with the interaction energy of the solute–solvent interactions in and out of the cavity as:

**Table 1** SPARC physical processes calculator statistical performance

Property	Units	Training		Testing/validating					Reaction conditions (temp/press/solvent)
		Total no	RMS	NPolar no	Polar no	Total no	RMS	R <sup>2</sup>	
Ref. index	NA	219	0.006	447	131	578	0.007	0.997	25 °C
Molecular volume	cm <sup>3</sup> /mole	950	1.79	625	815	1440	1.97	0.999	25 °C
Vapor pressure	log atm	291	0.11	447	300	747	0.15	0.994	25 °C
Boiling point	°C	1250	4.05	2150	1855	4005	5.71	0.999	0.1–1520 torr
Heat of vaporization	kcal/mol	611	2.11	841	422	1263	2.53	0.993	25 °C, boiling point
Heat of formation	kcal/mol	633	1.02	587	202	789	1.12	0.992	25 °C
Diffusion coefficient <sup>a</sup>	cm <sup>2</sup> /s	92	0.003	16	92	108	0.003	0.994	25 °C
Activity coefficient	log MF	415	0.078	1401	1246	2647	0.272	<b>0.953</b>	25 °C, 163 solvents
Solubility									
Single solvent	log MF	508	0.316	124	583	707	0.49	0.982	25 °C, 20 solvents
Binary mixture solvent		90	0.24	0	139	139	0.35	0.943	25 °C, 6 mixed solv.
Distribution coefficient	NA	623	0.43	105	593	698	0.44	0.978	25 °C, 6 solvents
Henry's constant	NA	165	0.28	265	505	750	0.41	0.982	25 °C, Water
		92	0.09	254	156	410	0.29	0.992	25, hexadecane
GC retention time	Kovtas	85	8.5	156	13	169	10.8	0.998	25 °C Squalane
		90	11.3	53	76	129	14.2	0.992	25 °C Apolane 87
		85	9.3	156	13	169	11.6	0.998	80 °C Squalane
		90	15.8	53	76	129	18.5	0.990	130°C Apolane 87
LC retention time <sup>b</sup>	No	NA	NA	86	69	155	0.10	0.994	25 °C, water/methanol
Entropy of fusion [18]	kcal/mol	204	1.53	187	717	904	3.07	0.89	
Enthalpy of fusion [18]	kcal/mol	272	4.12	187	717	904	1.21	0.93	

MF: mole fraction, Mols: Molecules, NA: Not Available, No: no training, a: [12], b: Hilal and Carreira (in preparation)

$$\Delta G_{ij} = \Delta G_{ii} + \delta G_{ij} \quad (4)$$

where  $\Delta G_{ii}$  is the self-energy (solute–solute or solvent–solvent) interaction.  $\delta G_{ij}$  describes the differential mixing of an 'isolated' solute molecule 'i' into solvent 'j', and is given as

$$\delta G_{ij} = W_c \delta G_{ij}^c + W_{nc} \delta G_{ij}^{nc} \quad (5)$$

where  $\delta G_{ij}^{gc}$  ( $\delta G_{ij}^c$ ) describes solvation with (without) solvent destructuring (outside and inside the cavity), respectively, and  $W_{nc}$  ( $W_c$ ) is the fraction of the solute molecules outside (inside) the cavity given by

$$W_c = 1 - W_{nc} \quad (6)$$

$$W_{nc} = 10^{\sum \delta G_{ij}^{nc}}$$

where the summation is over all dipole–dipole and hydrogen bonding interactions.

#### Dispersion interactions

The free energy associated with the differential energy for mixing solute i and solvent j is expressed as

$$\Delta G_{ij}(\text{Dispersion}) = \rho_{disp} (P_i^d - P_j^d)^2 V_i \quad \text{where} \quad (7)$$

$$P_i^d = \frac{\alpha_i^o + A_{disp}}{V_i}$$

$P_i^d$  and  $P_j^d$  are the effective polarizability densities of the solute and the solvent molecules, respectively.  $\rho_{disp}$  and  $V_i$  are the susceptibility to dispersion and molecular volume, respectively. Both effective polarizability densities ( $P_j^d$  and  $P_i^d$ ) are calculated the same way (Eq. 7) without any modification to the equation or differentiating between solute or solvent compounds.

Dispersion is a short range interaction involving surface or near surface atoms, and  $A_{disp}$  is an adjustment that subtracts from the total polarizability a portion of the contributions of sterically occluded atoms in the molecular lattice. Presently, SPARC only corrects for access judged to be less than that afforded by a linear array of atoms. Branched atoms in a non-polar or polar compound will therefore lose a small part of their intrinsic molecular polarizability, depending on the size and number of appended group, and the proximity of other branched carbons. Similarly, carbons in saturated or aromatic rings may lose their intrinsic polarizability contributions depending on ring size and the presence of a ring appendage.

#### Induction interactions

$\Delta G_{ii}$  is the self-energy (solute–solute or solvent–solvent) interaction given as

$$\Delta G_{ii}^c(\text{induction}) = \rho_{Ind} P_i' D_i V_i \quad (8)$$

$$\delta G_{ij}^c(\text{induction}) = \rho_{Ind} [-D_i P_j' - P_i' D_j' + P_j' D_j'] V_i \quad (9)$$

$$\begin{aligned} \delta G_{ij}^{nc}(\text{induction}) \\ = \rho_{Ind} \left[ -(D_i P_j' + P_i' D_j') + \left( \frac{V_i + V_j}{2} \right) + P_j' D_j' V_j \right] \end{aligned} \quad (10)$$

where

$$P_i' = \frac{\alpha_i^o + A_{Ind}}{V_i} \quad D_i = \frac{\mu_i}{V_i}$$

and  $\rho_{ind}$  is the 'susceptibility' to induction and  $P_i'$  and  $D_i$  are polarizability density (adjusted for induction) and the dipole density of the molecule-in-question, respectively.  $\mu_i$  is the effective microscopic dipole strength calculated from molecular descriptors and  $A_{ind}$  is a polarizability adjustment for induction.  $\delta G_{ij}^{nc}$  ( $\delta G_{ij}^c$ ) describes solvation with (without) solvent destructuring (outside and inside the cavity), respectively.  $P_j'$ ,  $D_j$  and  $D_j'$  are polarizability density, the dipole density of the solvent, and the dipole density of the solvent inside the cavity, respectively. The dipole density of the solvent,  $D_j$  inside the cavity is adjusted as:

$$D_j' = f_d D_j \quad (11)$$

where  $f_d$  gauges reduction in solute–solvent dipole–dipole interaction inside versus outside of the cavity sphere solvation.

Induction describes molecular polarization affected by a point dipole on the surface, averaged over all orientations of the molecule. Inductive polarization interactions 'propagate' deeply within conjugated systems, but only one or two atoms deep in a non-conjugated array of atoms. SPARC adjusts the molecular polarizability algorithmically, utilizing electron withdrawing and releasing substituent parameters derived from  $pK_a$  models.

#### Dipole–dipole interactions

Similar to induction, the free energy change of the dipole–dipole interaction is given by

$$\Delta G_{ii}(\text{dipole} - \text{dipole}) = \rho_{d-d} D_i D_i V_i \quad (12)$$



$$\delta G_{ij}^{nc}(\text{dipole} - \text{dipole}) = \rho_{Id-d} \left[ -2D_i D_j \left( \frac{V_i + V_j}{2} \right) + D_j^2 V_j \right] \quad (13)$$

$$\delta G_{ij}^c(\text{dipole} - \text{dipole}) = \rho_{Id-d} \left[ -2D_i D_j' + D_j'^2 \right] V_j \quad (14)$$

where  $\rho_{d-d}$  is the susceptibility to dipolar interactions;  $D_i^d$  and  $V_i$  are the effective dipole density and molar volume of the molecule-in-question, respectively. SPARC adjusts the effective bond dipole strength,  $\mu_i$  for steric blockage due to an appended molecular structure. Since dipole–dipole interactions depend on the position of the polar molecule with respect to its neighbor. SPARC adjusts the dipole–dipole interaction as a function of the number and magnitude of the microscopic bond dipole moments in the molecule. In addition, SPARC adjusts  $\Delta G_{ii}$  for the ability of one dipole to align the dipole in the other molecule into a favorable arrangement, and if the two dipoles can interact with each other through H-bonding interactions.

### H-bonding interactions

Hydrogen bonding interaction is a direct site coupling of a proton-donating site of one molecule with a proton-accepting site of another molecule. The H-bond energy ( $\Delta G_{H-Bond}$ ) is resolved into a proton donating site  $\alpha$  and proton accepting site  $\beta$ , which in our models are presumed to be independently quantifiable. For single site interactions, the self interaction free energy change of a proton donating site interacting with proton accepting site is given by

$$\Delta G_{ii}(H - B) = \rho_{HB} S_{ii} \alpha_i \beta_i \quad (15)$$

$$\delta G_{ij}^c(H - B) = \rho_{HB} \left( -S_{ij} \alpha_i \beta_j' - S_{ji} \beta_i \alpha_j' + S_{jj} \alpha_j' \beta_j' \frac{V_i}{V_j} \right) \quad (16)$$

$$\delta G_{ij}^{nc}(H - B) = \rho_{HB} (-S_{ij} \alpha_i \beta_j - S_{ji} \beta_i \alpha_j + S_{jj} \alpha_j \beta_j) \quad (17)$$

where  $\rho_{HB}$  is the susceptibility to H-bonding interactions,  $S_{ii}$  ( $S_{ij}$ ) is a steric reduction factor for the solute–solute (solute–solvent) [13]. Similar to the dipole–dipole interaction, the  $\alpha$  and  $\beta$  of the solvent inside the cavity is adjusted as shown in Eq. 18. The  $f_{HB}$  factor gauges reduction in solute–solvent H-bonding for outer versus inner sphere solvation. The solvent–solvent term is the cavity creation energy in the absence of solvent destructuring.

$$\alpha_j' = f_{HB} \alpha_j \quad \beta_j' = f_{HB} \beta_j \quad (18)$$

### SPARC physical models

SPARC solvation models are described in this section. Results of these solvation models in estimating solute activities and 'activity based' properties (solubilities, vapor pressures, and distribution coefficients) are given for a wide range of solutes and solvents. A prototypical set of solutes and solvents has been selected that covers a wide range of interaction forces both in type and strength. Model extensions to more complex molecules are described along with some calculated properties. Any chemical model should be understood in terms of the purpose for which it is conceived and its prescribed usage. The models described herein are intended for what might be characterized as engineering applications in environmental assessments; the target user has minimal chemistry/computer skills; the target computer a standard PC. The process modeling goal is to optimize physical and chemical integrity yet achieve both the requisite range of prediction capability (physical/chemical processes, 'environmental' conditions, and molecular structures) and 'accessibility' for the target audience

### Vapor pressure model

The vapor pressure is among the most frequently measured and reported physical property of organic compounds. However, the number of compounds where the vapor pressure was actually measured and not extrapolated to 25 °C from higher temperature measurements is less than 1000. In addition to being highly significant in evaluating 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. SPARC expresses the vapor pressure  $VP_i^o$  of a pure solute,  $i$ , at 25 °C as function of all the intermolecular interaction mechanisms as

$$\log VP_i^o = \frac{-\Delta G_{ii}(\text{Interaction})}{2.303RT} - \log T + C \quad (19)$$

where  $\log(T) + C$  describes the change in the entropy contribution associated with the volume change in going from the liquid phase to the gas phase [12]. For molecules that are solids above 25 °C, the crystal energy contribution becomes important and can be calculated using SPARC [18].

## Normal boiling point

SPARC calculates the boiling point for a 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. Since the same factors that affect the boiling point of a compound affect the vapor pressure, the dipole–dipole and H-bond interactions become less important and decrease significantly with increasing temperature. Since, the number of compounds where the vapor pressure at 25 °C was directly measured and not extrapolated from higher temperature measurements is limited [19], the vapor pressure calculator was also validated against 4005 boiling points measured at different pressures ranging from 0.05 to 1520 torr spanning a range of over 800 °C as shown in Table 1.

## Enthalpy of vaporization

In SPARC, the enthalpic contribution for any physical process is estimated from the corresponding free energy process. For example, since the heat of vaporization can be determined from the vapor pressure, the enthalpy contribution for each intermolecular interaction that contributes to the free energy process can be expressed as

$$\log \Delta H_{ii}(\text{vap}) = \frac{-\Delta G_{ii}^H(\text{Interaction})}{2.303RT} - \log T + C^H \quad (20)$$

where  $\Delta G_{ii}^H$  is the enthalpic contribution to the free energy change of the self interactions as explained in the following. Similar to the vapor pressure model, the  $\log(T) + C^H$  term describes the change in entropy associated with the change in the molecular volume going from the liquid to gas phase upon vaporization. However, unlike in the vapor pressure model  $C^H$  is independent of temperature and represents the Clausius–Clapeyron integration constant. For the  $\Delta H_{(\text{vaporization})}$  contribution, SPARC modifies the susceptibility of each molecular interaction mechanism as:

$$\rho_{\text{Mechanism}}^{\Delta H} = \Omega_{\text{Mechanism}} \rho_{\text{Mechanism}} \quad (21)$$

where  $\Omega_{\text{Mechanism}}$  is dependent on the interaction mechanism (dispersion, induction, dipole–dipole and H-bond) involved, and is data-fitted at 25 °C and stored in the SPARC database. Likewise, the susceptibility,  $\rho_{\text{Mechanism}}$ , as indicated in previous sections also depends on the type of the interaction mechanism involved, such as dispersion, induction, etc.

## Activity coefficient model

The activity coefficient in water is one of the most important physical properties of a chemical compound, having numerous applications in the estimation of its fate and effects in the environment. In addition to being highly significant in evaluating compound's environmental fate, the activity coefficient is an excellent arena for developing and validating the SPARC solute–solvent interaction physical process models in water as well as in any other solvent. In SPARC, the solubility as well as Liquid/Liquid, Liquid/Solid, Gas/Liquid is 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. For a solute molecule  $i$  in a liquid phase,  $j$ , at infinite dilution, SPARC expresses the activity coefficient as

$$-RT \log \gamma_{ij}^\infty = \sum \Delta G_{ij}(\text{Interaction}) + RT \left( \log \frac{V_i}{V_j} + \frac{\left( \frac{V_i}{V_j} - 1 \right)}{2.303} \right) \quad (22)$$

where the last term is the Flory–Huggins [20, 21] excess entropy of mixing contribution in the liquid phase due to placing a solute molecule  $i$  in the solvent  $j$ . It is an entropy contribution that accounts for differences in the molar volumes of the solvent and solute molecules. The Flory–Huggins term is damped-out by orienting interactions that reduce the randomness of placement.

## Solubility

Solubility is the maximum amount of a compound that will dissolve in pure solvent at a given temperature. SPARC does not calculate the solubility from first principles, but rather uses the activity coefficient model described previously. Specifically, SPARC estimates molecular solubility from a calculation of the infinite dilution activity coefficient,  $\gamma^\infty$ . When  $\log \gamma^\infty$  is greater than 2, the mole fraction solubility can be reliably estimated as  $\chi^{\text{sol}} = 1/\gamma^\infty$ . However, when the  $\log \gamma^\infty$  is calculated to be less than 2, this approximation fails. In these cases,  $\gamma^\infty$  is greater than  $\gamma^{\text{sol}}$  and SPARC would underestimate the solubility. In order to overcome these limitations, SPARC employs an iterative calculation. SPARC sets the initial guess of the solubility as  $\chi_{\text{guess}} = 1/\gamma^\infty$ . SPARC then 'prepares' a mixed solvent that is  $\chi_{\text{guess}}$  in the solute and  $(1-\chi_{\text{guess}})$  in the solvent. SPARC recalculates  $\gamma^\infty$  for the 'new' solvent. This process is continued until recalculated  $\gamma^\infty$  values converge or equal to 1 (miscible).

## Mixed solvents system

SPARC uses the same molecular descriptors to describe solvents as it does solutes. The only solvent 'known' to SPARC at start up is water. All other solvents must be entered as SMILES strings and processed by the system. The user may declare the molecule as a solvent to be remembered for future calculations. SPARC then stores the molecular descriptors that it has calculated in memory. Any molecule that SPARC can run as a solute may be declared a solvent, so essentially any organic solvent may be specified. SPARC can handle solvent mixtures for a virtually unlimited number of components. Speed and memory requirements usually limit the number of components to less than 20. 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.

Solvent descriptors that are essentially bulk in nature (e.g. polarizability) are volume fraction averaged when employed in the interaction models described earlier. Solvent descriptors that are site interactions (e.g. hydrogen bonding) are mole fraction weighted when used in the interaction models, and the interactions are summed over all solvent components. To date, we have tested the SPARC prediction capability on 139 solubilities in six mixed solvents systems kindly provided by Dr. Yalkowsky. The RMS deviation in mixed binary solvent system (at mole fraction 0.1, 0.3, 0.5, 0.7 solvent–water system, where the solvents are methanol, ethanol, isopropanol, dimethyl formamide, acetone and acetonitrile) was 0.35. There is no modification or extra parameterization to any of the SPARC physical processes models to predict the solubility in mixed solvents.

## Partitioning coefficients

Thermodynamically a partition constant is defined as

$$\Delta G_{ij} = -RT \ln K_{ij} \quad (23)$$

where  $K_{ij}$  is defined as the ratio of the activity coefficients in  $i$  and  $j$ . All partitioning (Liquid/Liquid, Liquid/Solid, Gas/Liquid and Gas/Solid) coefficients addressed in SPARC are determined by calculating the activity of the molecular species in each of the phases without any modification or extra parameterization to any of the SPARC process models.

## Liquid/liquid partitioning

SPARC calculates the liquid-liquid partition coefficients such as the octanol/water distribution coefficient,  $K_{ow}$ , by

simply calculating the activity at infinite dilution of the molecular species of interest in each of the liquid phases as

$$\log K_{liq1/liq2} = \log \gamma_{liq2}^{\infty} - \log \gamma_{liq1}^{\infty} + \log R_m \quad (24)$$

where the  $\gamma^{\infty}$ 's are the activities at infinite dilution of the compound of interest in the two phases and  $R_m$  is the ratio of the molecularities of the two phases ( $\text{mole}_1/L_1$ )/( $\text{mole}_2/L_2$ ). Although octanol–water partition coefficients are widely used and measured, the SPARC system does not limit itself to only this calculation. SPARC can calculate a compound's liquid–liquid partition coefficient for any two immiscible phases; the phases can be mixed solvents. In fact, when calculating an octanol/water partition coefficient, SPARC calculates the activity in water and the activity in wetted octanol, i.e., a 5% water 95% octanol (by volume) mixture. The water in the octanol phase makes this a more cohesive solvent than pure octanol [13].

## Gas/liquid (Henry's constant) partitioning

Henry's constant characterizes the equilibrium partitioning of molecules between air and solvent as

$$H_x = VP_i^o \gamma_{ij}^{\infty} \quad (25)$$

where  $VP_i^o$  is the vapor pressure of pure solute  $i$  (liquid or subcooled liquid) and  $\gamma_{ij}^{\infty}$  is the activity coefficient of solute ( $i$ ) in the liquid phase ( $j$ ) at infinite dilution. The vapor pressure coupled with the activity coefficient models were used to estimate Henry's law constants in water and hexadecane strictly from molecular structure without modification or additional parameterization to any of the SPARC physical processes models.

## Kov'ats indices

Despite some limitations, the Kov'ats index (KI) has found much greater use than all other specialized retention specification schemes. The KI is the only retention value in gas–liquid chromatography (GC) in which two fundamental quantities, the relative retention and the specific retention volume are united [22]. Moreover, a series of explicit relationships between retention indices and a number of physicochemical quantities related to GC have been developed. Many different linear relationships between the KI value for a molecule and other fundamental molecular properties such as carbon number, boiling point and refractive index have been derived [22, 23]. In this study, we report the result of extending the Henry's constant calculator to estimate GC in squalane and Apolane 87.



There is no modification or extra parameterization to any of the SPARC physical processes models.

The KI of the compound of interest, X, can be expressed in terms of Henry's law constant [17] as

$$KI = 100 \times \left( \frac{\log H_{N_x} - \log H_{N_z}}{\log H_{N(z+1)} - \log H_{N_z}} + CN \right) \quad (26)$$

where CN is the number of carbon atoms in the first n-alkane standard.  $H_{N_x}$ ,  $H_{N_z}$ , and  $H_{N(z+1)}$  are Henry's constant for the compound of interest X, first n-alkane standard,  $H_{N_z}$ , and the second n-alkane standard,  $H_{N(z+1)}$ , respectively.

### Liquid/solid partitioning

SPARC calculates liquid/solid partitioning in a manner similar to liquid/liquid partitioning, except that for the solid phase the self-self interactions,  $\Delta G_{jj}$ , are dropped from the calculation. The capability of mixed-solvent/solid partitioning is applied to the calculation of liquid chromatographic retention times. There, the mobile phase is a water–methanol mixture and the stationary phase is octadecane/surface-water. At the present time, we are validating, testing and refining liquid/solid partitioning models. Hence, we will report only the results of our calculations of liquid chromatographic retention times as shown in Table 1

### Temperature dependence of physical processes

SPARC calculates a physical property of interest at 25 °C. In addition to the inherent temperature dependence described previously, the susceptibility of dispersion at temperature T is modeled as a function of the polarizability density and the effective polarizability density. In effect, this describes the small temperature dependence of enthalpy. For an “activity-driven” process  $\rho_{\text{disp}}$  as a function of temperature is given by:

$$\rho_{\text{Mechanism}}^T = \left[ 1 + \left( 1 - \sum_n a_n \left( \frac{298}{T} \right)^n \right) f(P^d, D^d, HB) \right] \rho_{\text{Mechanism}}^{25}$$

where  $a_n$  are quantified from physical properties measurement, mainly on boiling points measured at different pressures, heat of vaporizations (at the boiling point) and on GC chromatographic retention times at different temperatures ranging from 30 to 190 °C.

For interactions that involve molecular orientation, SPARC includes explicit temperature dependence

associated with the molecular orientation requirements for dipole–dipole coupling and hydrogen bonding interactions. To accomplish this, initially, SPARC modifies both interaction susceptibilities for the dipole–dipole and H-bonding as follows:

$$\rho_M = \rho_M^{25} \left[ \frac{298}{T} \Omega_M + (1 - \Omega_M) \right] \frac{298}{T}$$

where  $\Omega_M$  are data fitted parameters (for dipole–dipole and/or H-bonding). Both  $\Omega_{\text{dispersion}}$  and  $\Omega_{\text{induction}}$  are set to be equal to 1. The  $\Delta H$  and  $\Delta S$  temperature dependence are described by the first and the second term, respectively. SPARC assumes that the  $\Delta H/\Delta S$  contribution to  $\Delta G$  is constant at 25 °C. The multiplier of the  $(298/T)$  in both equations is the temperature dependence factor associated with molecular orientation. That is why the dipole–dipole and H-bonding interaction will drop out faster than either dispersion or induction as the temperature increases.

## Result and discussions

In SPARC, experimental data for physicochemical properties (such as boiling point, partition coefficients) 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., induction, dispersion and H-bonding interactions). These mechanistic models were parameterized using a very limited set of experimental data, but as clearly stated, 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 summary of the statistical parameters of the validation tests of some of the SPARC physical property models are presented in Table 1.

The SPARC physicochemical calculator was trained on vapor pressure (as a function of temperature), boiling point (as a function of pressure), diffusion coefficient (as a function of pressure and temperature), heat of vaporization (as a function of temperature), activity coefficient (as a function of solvent and temperature), solubility (as a function of

solvent and temperature), GC retention time (as a function of stationary liquid phase and temperature) and partition coefficient (as a function of solvent). The overall SPARC physical property training set contains 3760 calculation of all the above mentioned physical properties. There were only trained 211 parameters that describe the solute–solute and solute–solvent dispersion, induction, dipole, H-bonding interactions and only 11 parameters to describe the entropic contributions. Results of the aforementioned solvation model tests relative to estimating solute activity coefficients and 'activity based' properties (such as solubilities, vapor pressures, and distribution coefficients) are given for a wide range of solutes and solvents in Table 1. For these tests, prototypical set of solutes and solvents were selected that covered a wide range of interaction forces, both in type and strength. The validating set includes more than 163 different pure solvents as well as 18 mixed solvent systems. The list of solvents and total number of solute molecules in each solvent on which we tested our activity coefficient and solubility models is shown in Table 2.

The vapor pressure calculator at 25 °C was tested against 747 measured data points spanning more than 9 log atm units [12]. Since the SPARC mechanistic models are designed and parameterized so as to be portable, in principle, to any type of physical property or molecular structure, the SPARC vapor pressure model was also validated for predicted boiling points at different pressures ranging from 0.05 to 1520 torr spanning a range of over 800 °C [12], for predicted heats of vaporization at 25 °C and at the boiling points spanning a range of over 27 kcal/mole, and for predicted diffusion coefficients as a function of temperature and pressure [12]. The activity coefficient model was tested on 2647 relatively simple polar and non-polar organic compounds at 25 °C in 163 different solvents, spanning a range of over 7 log mole fraction units [13]. The RMS deviation for this test was 0.272 log mole fraction with an  $R^2$  of 0.940. The activity coefficient model was also validated on solubilities spanning a range of more than 13 log mole fraction units in 20 different solvents, for predicted liquid/liquid distribution coefficients in carbon tetrachloride, benzene, cyclohexane, ethyl ether, octanol and toluene spanning a range of more than 15 log unitless values. In addition, Henry's law constant estimates were validated (see Table 1) for a wide range of molecular structures spanning a range of more than 11 log units in water and in hexadecane, these estimates were made by coupling the vapor pressure and activity coefficient models [16].

From our study, we see that SPARC can reliably estimate numerous physical properties of organic compounds using the same interaction models without modifications or additional parameterization either to the vapor pressure and/or the activity coefficient estimators (Fig. 2).

Indeed, the SPARC physical properties calculator predictions are as reliable as most of the experimental measurements for these properties. For simple structures, SPARC can calculate a property/parameter of interest within a factor of 2, or even better, as was the case for the estimated physical properties of ethanol in Table 3. For complex structures, where dipole–dipole and/or H-bond interactions are strong, SPARC estimates were within a factor of 3–4 as illustrated for the strong dipole/H-bonding molecules, such as cyanazine and caffeine in Table 2. In addition, set of the observed versus SPARC-calculated log mole fraction solubilities for 707 compounds at 25 °C yielded an RMS deviation of 0.487 log mole fraction unit with an  $R^2$  of 0.982 log mole fraction (see Table 1). Furthermore, the RMS deviation for the 158 liquid compounds of the test set was 0.21 log mole fraction units with an  $R^2$  of 0.998 mole fraction units; the remaining 549 solid compounds had an RMS deviation of 0.541 log mole fraction units with an  $R^2$  of 0.980 mole fraction units [13]. The RMS deviation for the solid compounds was about 2 1/2 times greater than that for the liquid compounds for two reasons. One reason is that crystal energy contribution for these compounds that are liquid at 25 °C is zero while for the solid compounds 25 °C the crystal energy could be as large as 1–3 log mole fraction units. Another reason is that most of the non-polar compounds in the test data set were simple aliphatic, ethylenic and aromatic hydrocarbon structures (other than the poly-nuclear aromatic hydrocarbons such as perylene, pentacene and coronene) whereas most of the polar compounds that are solids at 25 °C were either large pharmaceutical drug or complicated pesticide molecules such as fipronil, nitrofurantoin, sulfauinoxaline and flumetopir-methyl. Some of the solid compounds 25 °C we tested had more than ten strongly dipole and/or H-bonding substituents.

However, the real test of the SPARC physical property models does not lie in their demonstrated predictive capability for vapor pressure or activity coefficient, but is determined by the extrapolatability of these same models to other types of chemistry, process or properties, such as extrapolatability was demonstrated in that the vapor pressure model was extended to estimate boiling points, heats of vaporization and diffusion coefficients. Similarly, the activity coefficient model was extended to predict (and validated on) solubilities, partition coefficients and chromatographic retention time for single and mixed solvent systems. Indeed, both the vapor pressure and activity coefficient models successfully were extended and combined to calculate many other properties without any modifications to either model or any extra parameterization.

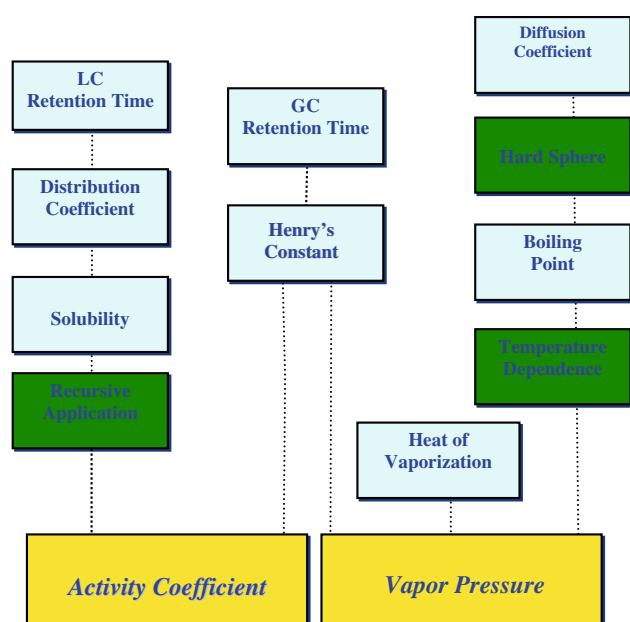
**Table 2** The RMS deviation error for SPARC-calculated (a) activity coefficients and (b) solubilities in multiple solvents

Name of the solvent	No	RMS	Name of the solvent	No	RMS
(a) Activity coefficients					
Benzene	34	0.13	Propionitrile	20	0.192
1-Heptene	5	0.18	Pyridine	15	0.115
1-Hexadecene	11	0.06	Quinoline	33	0.123
1-Iodohexadecane	8	0.17	Tertbutylmethylether	1	0.041
1-Methylnaphthalene	24	0.09	THF	14	0.2
1-Phenyldodecane	1	0.13	THF-2-carboxylicacidmethylester	11	0.438
Cyclohexane	20	0.10	THF-3-carboxylicacidmethylester	10	0.422
Decalin	4	0.09	1,2,4-trichlorobenzene	13	0.067
Decane	16	0.04	1-Bromohexadecane	12	0.043
Heptadecane	17	0.08	1-Chlorohexadecane	6	0.031
Heptane	32	0.11	1-Chloronaphthalene	14	0.049
Hexadecane	46	0.14	1-Chlorooctadecane	22	0.125
Hexane	33	0.07	Benzyl chloride	7	0.363
Isooctane	32	0.08	Bromo-4-fluorobenzene	6	0.124
Nonane	32	0.03	Bromobenzene	18	0.179
Nonene	1	0.03	Bromonaphthalene	17	0.155
Octadecane	22	0.10	Carbon tetrachloride	36	0.106
Octadecene	14	0.09	Chlorobenzene	18	0.151
Octane	25	0.17	Ethylbromide	12	0.151
p-Xylene	31	0.10	3-Methyl-1-butanol	17	0.087
Pentane	15	0.12	4-Nitrocyclohexene	5	0.147
Perhydrosqualane	45	0.16	4-Phenyl-2-butanone	7	0.223
Squalene	15	0.19	N,N-dibutyl-2-ethylhexanamide	11	0.34
Tetracosane	1	0.01	N,N-diethylacetamide	7	0.207
Toluene	33	0.10	N,N-dimethylacetamide	30	0.351
Indene	7	0.08	N,N-dimethylformamide	43	0.27
1,2,3,5-tetrafluoro-4-nitrobenzene	7	0.51	N,N-dimethylpropionamide	7	0.493
1,2,4-trifluoro-3-nitrobenzene	7	0.45	N-Ethylaceticacidamide	8	0.794
1-Cyclohexyl-2-propanone	7	0.10	N-Methyl-2-pyrrolidone	30	0.141
1-Dodecanol	17	0.12	N-Methylacetamide	7	1.002
1-Fluoro-2-nitrobenzene	7	0.13	N-Methylformamide	18	0.382
1-Nitropropane	7	0.21	N-Methylpropionamide	8	0.833
1-Phenyl-2-propanone	7	0.33	Acetic acid	16	0.154
2,3,4,6-tetrafluoroaniline	7	0.43	Acetone	21	0.197
2,4-difluoroaniline	7	0.32	Acetonitrile	51	0.196
2,6-lutidine	6	0.22	Acetophenone	34	0.148
2-Butanone	31	0.21	$\alpha$ -Aminotoluene	5	0.175
2-Cyanofuran	11	0.65	Aminocyclohexane	5	0.091
2-Cyanotetrahydrofuran	11	0.23	Aniline	43	0.278
2-Dodecanone	6	0.24	Anisole	31	0.32
2-Fluoroaniline	7	0.32	Benzaldehyde	5	0.125
2-Furancarboxylicacidmethylester	11	0.58	Benzonitrile	25	0.101
2-Furylmethylketone	11	0.20	Benzyl alcohol	20	0.117
2-Mercaptoethanol	10	0.87	Benzylacetate	17	0.19
2-Methyl-2-propanol	16	0.17	Benzylether	6	0.302
2-Methyl-3-nitrofurane	11	0.40	Bromoanisole	17	0.177
2-Methyl-5-nitrofurane	9	0.60	Butanol	31	0.169

**Table 2** continued

Name of the solvent	No	RMS	Name of the solvent	No	RMS
2-Methylquinoline	13	0.13	Butylacetate	3	0.053
2-Nitrofuran	11	0.82	Butylether	15	0.147
2-Nitropropane	17	0.23	Butyronitrile	14	0.146
2-Nitrotoluene	1	0.03	Ethylenechloride	31	0.208
2-Pentanone	3	0.41	Hexafluorobenzene	6	0.284
2-Phenylethanol	5	0.14	Iodobenzene	7	0.176
2-Picoline	15	0.12	Methylene chloride	16	0.24
2-Propanol	15	0.16	Methyleneiodide	17	0.754
3-Furaldehyde	11	0.69	n-Butylchloride	11	0.067
3-Furancarboxylicacidmethylester	11	0.50	Perfluorohexane	17	0.5
4-Methylquinoline	21	0.07	Perfluorooctane	11	0.127
Nitrobenzene	37	0.14	Chloroform	15	0.198
Nitrocyclohexane	5	0.16	Cyanocyclohexane	5	0.182
Nitroethane	26	0.55	Cycloheptanol	3	0.177
Nitromethane	19	0.34	Cyclohexanol	6	0.127
Nonanenitrile	20	0.13	Cyclohexanone	34	0.173
Octanol	37	0.13	Cyclopentanol	2	0.169
Pentadecane nitrile	17	0.14	Diethylcarbitol	17	0.242
Perfluorotributylamine	11	0.76	Diethyleneglycol	6	0.347
Phenol	7	0.34	Diethyloxalate	7	0.404
Phenylacetoneitrile	7	0.15	Dimethylmalonate	7	0.358
Propanol	16	0.17	Dimethylsulfoxide	22	0.334
Tetrahydrofurfurylcohol	11	0.33	Diocetyl ether	18	0.23
Triethylamine	10	0.27	Dioxane	16	0.334
Trifluoroethanol	15	0.27	Dipropylsulfoxide	6	0.285
Water	205	0.29	Ethanol	22	0.21
m-Cresol	12	0.33	Ethoxybenzene	17	0.202
Methanol	19	0.15	Ethyl acetate	35	0.22
Methoxyethanol	16	0.33	Ethylbenzoate	1	0.1
Ethyleneglycol	6	0.36	Ethyl ether	13	0.064
Furfural	12	0.82	Isopropylether	16	0.078
Furfurylcohol	11	0.25	Isoquinoline	26	0.081
Hexafluoro-2-propanol	16	0.78	Methylbenzoate	5	0.195
Hexahydroacetophenone	5	0.16	Methylhexahydrobenzoate	5	0.108
(b) Solubilities					
Name of the solvent	No	RMS	Name of the solvent	No	RMS
Water	507	0.49	Isopropanol	4	0.317
N,N-dimethylformamide	3	0.62	Methanol	10	0.619
Acetone	20	0.52	Pyridine	4	0.332
Benzene	18	0.49	Carbon tetrachloride	5	0.384
Chloroform	14	0.63	Methylene chloride	3	0.186
Dioxane	4	0.55	Cyclohexane	3	0.261
Dimethylsulfoxide	2	0.10	Hexane	11	0.64
Ethanol	30	0.53	p-Xylene	4	0.516
Ethyl acetate	2	0.72	Pentane	3	0.789
Ethyl ether	19	0.57	Toluene	3	0.48

No: Number of the solute compounds



**Fig. 2** Estimation of other physical properties based on vapor pressure and activity coefficient models. The vapor pressure model was extended to estimate boiling points, heat of vaporization and diffusion coefficient. The activity coefficient model was extended to estimate solubility and distribution coefficient. Henry's constant and GC retention time are calculated by coupling the vapor pressure and activity coefficient models.

## Conclusion

The SPARC mechanistic models are designed and parameterized so as to be portable, in principle, to any type of physical/chemical property or molecular structure. Hence, SPARC estimates numerous physical properties with greater accuracy, and with a wider scope than is possible with other estimation techniques. SPARC physical property models have been rigorously tested against all available measurement data found. These data cover a wide range of reaction conditions to include solvent, temperature, pressure, pH and ionic strength. The diversity and complexity of the molecules used in the tests during the last few years were drastically increased in order to develop more robust models. A complete tabulation of the SPARC physical property test results can be obtained from the Authors.

SPARC is an important tool in designing chemicals, in preliminary evaluation of the risks associated with the release of new chemical compounds into the environment, and in general chemical fate-modeling applications in the absence of experimentally measured values. The SPARC physical process models will meet some of the needs of EPA, whose long term research task agenda for chemical toxicology assessment emphasizes that EPA

**Table 3** Observed<sup>a</sup> versus SPARC calculated of ethanol physical properties

No.	Property	Solvent	T or P	Observed	SPARC
1	Refractive index	NA	20 °C	1.3611	1.374
2	Volume	NA	25 °C	58.63	56.95
3	Vapor pressure	NA	25 °C	−1.10	−1.03
4	Boiling point	NA	760 torr	78.29	81.03
5	Boiling point	NA	100 torr	34.23	32.3
6	Boiling point	NA	10 torr	−2.84	−4.8
7	Heat of vaporization	NA	78 °C	9.22	8.65
8	Heat of vaporization	NA	25 °C	10.11	9.86
9	Diffusion coefficients	NA	25 °C	0.12	0.121
10	Solubility	Water	25 °C	Very sol.	0.62
11	Activity coefficient	1-Butanol	25 °C	−0.03	−0.09
12	Activity coefficient	2-Methyl-2-propanol	25 °C	−0.18	0.09
13	Activity coefficient	3-Methyl-1-butanol	25 °C	−0.06	−0.08
14	Activity coefficient	Acetic acid	25 °C	−0.08	−0.13
15	Activity coefficient	Benzyl chloride	25 °C	1.06	1.02
16	Activity coefficient	Dioctyl ether	25 °C	0.90	0.59
17	Activity coefficient	Ethyl acetate	25 °C	0.53	−0.5
18	Activity coefficient	m-Cresol	25 °C	−0.54	−0.48
19	Activity coefficient	N,N-dimethylformamide	25 °C	−0.13	0.1
20	Activity coefficient	p-Xylene	25 °C	1.19	1.17
21	Activity coefficient	Pentadecane nitrile	25 °C	0.64	0.3
22	Activity coefficient	Acetone	25 °C	0.38	−0.12
23	Activity coefficient	Acetonitrile	25 °C	0.59	0.09



**Table 3** continued

No.	Property	Solvent	T or P	Observed	SPARC
24	Activity coefficient	Aniline	25 °C	0.38	−0.08
25	Activity coefficient	Anisole	25 °C	0.98	0.83
26	Activity coefficient	Benzene	25 °C	1.18	1.21
27	Activity coefficient	Benzonitrile	25 °C	0.66	0.62
28	Activity coefficient	Benzylether	25 °C	0.81	0.23
29	Activity coefficient	Bromobenzene	25 °C	1.24	1.33
30	Activity coefficient	Chloroform	25 °C	0.72	1.09
31	Activity coefficient	Cyclohexane	25 °C	1.79	1.53
32	Activity coefficient	Dioxane	25 °C	0.40	0.21
33	Activity coefficient	DMSO	25 °C	−0.28	−0.16
34	Activity coefficient	Hexafluorobenzene	25 °C	1.39	1.26
35	Activity coefficient	Isopropanol	25 °C	−0.05	−0.02
36	Activity coefficient	Methanol	25 °C	0.07	0.07
37	Activity coefficient	Nitrobenzene	25 °C	0.96	0.99
38	Activity coefficient	Nitroethane	25 °C	0.82	0.36
39	Activity coefficient	Nitromethane	25 °C	0.92	0.27
40	Activity coefficient	Octanol	25 °C	0.06	0.01
41	Activity coefficient	Pyridine	25 °C	−0.10	0.46
42	Activity coefficient	Squalane	25 °C	1.32	1.41
43	Activity coefficient	Toluene	25 °C	1.19	1.18
44	Activity coefficient	Triethylamine	25 °C	0.29	0.96
45	Activity coefficient	Trifluoroethanol	25 °C	−0.62	−0.30
46	Distribution coefficient	Water, cyclohexane	25 °C	−1.96	−1.68
47	Distribution coefficient	Water, octanol	25 °C	−0.25	−0.15
48	Henry's constant	Water	25 °C	−3.62	−3.51
49	Henry's constant	Hexadecane	25 °C	−1.43	−1.43

a: units are the same as in Table 1. The RMS error is 0.33 and  $R^2$  is 0.94 for ethanol activity/distribution coefficients and Henry's constant in various solvents

**Table 4** Observed versus SPARC-calculated physical properties for selected molecules

Property	3-Methyl hexane		Benzene		2-Butanon		1,3,5-trinitro benzene <sup>c</sup>		Fenuron		Caffeine <sup>c</sup>	
Volume	147.3	146.7	89.4	89.5	90.1	86.7	144 <sup>d</sup>	139.3	NA	143	NA	180.6
Ref. index <sup>a</sup>	NA	1.385	1.501	1.500	1.379	1.38	NA	1.61	NA	1.59	NA	1.451
Vapor pressure	−1.09	−1.02	−0.92	−0.86	−0.92	−0.88	NA	−7.89	−7.3	−7.41	NA	−7.55
Boling point	91.8	90.3	80.1	77.8	79.6	80.2	315	327	NA	304	NA	271.6
Heat of vap.	8.39	7.82	8.09	7.65	8.32	7.98	NA	24.6	NA	21.28	NA	17.38
Heat of vap. <sup>b</sup>	7.36	7.28	7.35	7.01	7.48	7.40	NA	18.67	NA	16.80	NA	13.97
Henry's const.												
Hexadecane	−3.04	−2.95	−2.8	−2.79	−2.25	−2.31	NA	−7.20	NA	−6.90	NA	−5.79
Water	2.03	1.98	−0.61	−0.72	−2.63	−2.58	NA	−6.76	−7.1	−6.94	NA	−7.13
Kow	NA	4.31	2.13	1.92	0.29	0.55	1.18	1.45	0.98	1.39	−0.07	−0.37
Solubility												
Hexane	NA	Mis	NA	−0.06	−0.66	−0.66	NA	−3.14	−4.0	−2.65	NA	−4.37
Benzene	NA	−0.33	Mis	Mis	−0.1	−0.21	NA	−1.20	−2.8	−1.21	−2.3	−1.75
Chloroform	NA	−0.27	NA	−0.12	NA	−0.03	NA	−0.86	−1.02	−1.01	−1.03	−1.40
Carbon tetrachloride	NA	−0.11	−0.04	−0.01	−0.31	−0.47	NA	−2.31	NA	−2.45	NA	−3.31

**Table 4** continued

Property	3-Methyl hexane		Benzene		2-Butanon		1,3,5-trinitro benzene <sup>c</sup>		Fenuron		Caffeine <sup>c</sup>	
Methanol	NA	−1.39	−0.87	−0.96	NA	−0.46	−2.02	−2.01	NA	−2.31	NA	−1.40
Ethanol	NA	−1.01	−0.74	−0.81	−0.39	−0.31	−2.26	−2.13	−1.50	−1.92	−2.26	−1.40
Ethyl ether	NA	−0.15	NA	−0.18	−0.26	−0.62	−2.26	−1.50	−2.57	−1.57	−2.23	−1.84
Acetone	NA	−1.16	−0.21	−0.59	−0.03	−0.10	NA	−0.84	−1.55	−1.22	−2.1	−1.40
Acetonitrile	NA	−1.42	−0.49	−0.02	−0.08	−0.03	NA	−0.84	NA	−1.15	NA	−1.40
Ethyl acetate	NA	−0.44	NA	−0.30	NA	Mis	NA	−0.84	NA	−0.96	Sol	−1.40
Dichloromethane	NA	−0.40	NA	−0.06	NA	Mis	NA	−0.84	NA	−0.75	NA	−1.40
p-Xylene	NA	−0.17	NA	−0.002	NA	−0.27	NA	−1.47	NA	−1.47	NA	−2.07
Water	−6.04 <sup>j</sup>	−6.21	−3.39 <sup>h</sup>	−3.26	−1.4 <sup>h</sup>	−1.49	−4.52	−4.28	−3.41	−3.58	−2.69	−3.47
GC retent. time												
Apolane 87	NA	667	644	629	538	525	NA	1428	NA	1475	NA	1321
Squalane	675	668	630	637	NA	537	NA	1477	NA	1452	NA	1238
Property	Cyanazine <sup>c</sup>		Mefluidide <sup>c</sup>		Lindane <sup>e,f</sup>		4,4 DDT <sup>g</sup>		Beniocarb <sup>h</sup>		9,10-Anthraquinone <sup>i</sup>	
Volume	NA	197.7	NA	224.6	NA	178.6	NA	240.1	NA	176.2	NA	168.7
Ref. index <sup>a</sup>	NA	1.541	NA	1.504	NA	1.55	NA	1.625	NA	1.55	NA	1.681
Vapor pressure	−11.6	−11.5	< −7	−13.3	−6.28	−6.63	−9.7	−9.65	−8.3	−8.20	< −4	−10.8
Boiling point	NA	376.3	NA	360	323.1	328	NA	379	NA	371	377	374.1
Heat of vap.	NA	29.5	NA	35.7	NA	21.3	NA	26.3	NA	25.4	NA	28.54
Heat of vap. <sup>b</sup>	NA	21.92	NA	25.9	NA	14.2	NA	19.53	NA	23.73	NA	20.66
Henry's const.												
Hexadecane	NA	−8.26	NA	−8.67	NA	−7.1	NA	−10.5	NA	−7.52	NA	−8.73
Water	−8.84	−9.04	< −8.4	−10.7	−4.1	−4.05	−3.46	−3.27	−5.78	−7.22	−6.0 <sup>k</sup>	−6.13
Kow	2.22	1.94	2.08	2.21	3.72	3.65	6.91	6.83	1.7	2.08	3.39	3.09
Solubility												
Hexane	−2.3	−4.9	NA	−8.03	NA	−2.22	NA	−10.59	−3.68	−3.68	NA	−4.47
Benzene	−2.3	−2.94	−4.1	−4.62	−1.1	−0.75	−0.75	−0.58	−1.78	−2.14	−2.95	−2.84
Chloroform	−1.3	−2.93	NA	−4.14	−0.98	−0.73	−0.95	−0.58	−1.02	−1.73	−2.46	−2.63
CCl <sub>4</sub>	NA	−4.12	NA	−6.12	−1.4	−1.53	−0.70	−1.07	NA	−3.11	NA	−3.75
Methanol	NA	−2.78	−1.46	−2.87	−2.1	−2.26	−2.49	−3.54	NA	−2.47	NA	−4.51
Ethanol	−2.94	−2.75	NA	−2.93	−1.95	−2.03	−2.11	−3.25	−2.03	−2.32	−4.4	−4.63
Ethyl ether	NA	−2.61	−3.2	−3.55	−1.32	−1.22	−1.21	−1.41	NA	−1.94	−3.4	−3.38
Acetone	NA	−1.24	−1.25	−1.63	−1.06	−0.79	−1.03	−1.75	−1.28	−1.43	NA	−2.18
Acetonitrile	NA	−1.63	−1.9	−1.62	NA	−0.93	NA	−1.83	NA	−1.25	NA	−3.02
Ethyl acetate	NA	−1.24	−1.84	−1.61	−1.03	−0.78	NA	−1.47	NA	−1.34	NA	−2.06
Dichloromethane	NA	−2.31	−3.2	−3.31	NA	−0.73	−0.63	−0.77	−0.9	−1.19	NA	−2.48
p-Xylene	NA	−3.02	−4.51	−4.65	−1.12	−1.13	−0.77	−0.58	−2.1	−2.17	NA	−3.08
Water	−4.89	−4.92	−4.97	−5.59	−6.3	−6.12	−9.92	−9.65	−4.6	−4.62	NSol	−7.42
GC retent. time												
Apolane 87	NA	1872	NA	1872	NA	1478	NA	2186	NA	1583	NA	1755
Squalane	NA	1825	NA	1747	NA	1451	NA	2179	NA	1557	NA	1794

\*Units are the same as in Table 1. a: at 20 °C, b: at the boiling point, c: [24], d: at 152 °C, e: [25]. f: solubilities are at 20 °C, g: [24–26], h: [25, 26], i: [24], j: [27], k: [28], K<sub>ow</sub>: octanol/water distribution coefficient, Heat of vap: Heat of vaporization, NA: Not available. NSol: not soluble, Sol: Soluble. Mis: Miscible, log mole fraction solubility is 0

will rely heavily on predictive modeling to carry out the increasingly complex array of exposure and risk assessments necessary to develop scientifically defensible regulations.

### Disclaimer

The information in this document has been funded by the United States Environmental Protection Agency. It has

been subjected to the Agency's peer and administrative review, and has been approved for publication. Mention of trade names of commercial products does not constitute endorsement or recommendation for use.

## References

1. Miller MM, Wasik SP, Huang GL, Shiu WT, Mackay D (1985) *Environ Sci Technol* 19:522
2. Rekker RF (1977) The hydrophobic fragment constant. Elsevier, Amsterdam
3. Banerjee S, Yalkowsky SH, Valvani SC (1980) *Environ Sci Toxicol* 14:1227
4. Doucette WJ, Andren AW (1987) *Environ Sci Technol* 21:821
5. Lyman WJ, Reehl E, Rosenblatt D (1990) *Handbook of Chemical Property Estimation Methods*. ACS (ed), Washington
6. Leo AJ (1975) In: Veith GD (ed) *Structure activity correlations in studies of toxicity and bio-concentrations with aquatic organism*. International Joint Commission, Windsor, p 151
7. Wolfe NL, Zepp RG, Gordon JA, Baughman GL, Cline DM (1977) *Environ Sci Technol* 11:88
8. Zepp RG, Cline DM (1977) *Environ Sci Technol* 11:359
9. Zepp RG (1982) *Handbook of environmental chemistry*. Springer Verlag, New York
10. Karickhoff SW, McDaniel VK, Melton C, Vellino AN, Nute DE, Carreira LA (1991) *Environ Tox Chem* 10:1405
11. Hilal SH, Karickhoff SW, Carreira LA (1995) *Quant Struct Act Relat* 14:348
12. Hilal SH, Karickhoff SW, Carreira LA (2003) *QSAR Comb Sci* 22:565
13. Hilal SH, Carreira LA, Karickhoff SW (2004) *QSAR Comb Sci* 23:709
14. Hilal SH, Carreira LA, Karickhoff SW, Melton CM (1993) *Quant Struct Act Relat* 12:389
15. Whiteside TS, Hilal SH, Carreira LA (2006) *QSAR Comb Sci* 25:123
16. Hilal SH, Carreira LA (2007) To be submitted
17. Hilal SH, Carreira LA, Karickhoff SW, Melton CM (1994) *J Chromatogr* 662:269
18. Whiteside TS, Carreira LA, Hilal SH, Brenner A (2007) To be submitted
19. Dykyj J, Repas M, Svoboda AJ (1984) Vapor pressure of organic substances. VEDA, Vydavatel' stvo, Slovenskej Akademie Vied, Bratislava
20. Flory PJ (1941) *J Chem Phys* 9:660
21. Huggins ML (1941) *J Chem Phys* 9:440
22. Tarjan G, Timar I, Takacs JM, Meszaros SY, Nyiredy S, Budahegyi MV, Lombosi ER, Lombosi TS (1982) *J Chromatogr* 271:213
23. Haken JK, Evans MB (1989) *J Chromatogr* 93:472
24. Budavari S, O'Neil M, Smith A, Heckelman PE, Kinneary JF (1996) *The Merck index*. Merck & CO., Inc., Whitehouse Station
25. Heller SR, Bigwood DW, Laster P, Scott K, *Ars pesticide properties database*. USDA
26. Hartley D, Kidd H (1983) *The agro chemical handbook*. Royal Society of Chemistry, Nottingham
27. Yalkowsky SH, He Y (2006) *AQUASOL dATABASE of aqueous solubility*. The University of Arizona
28. Meylan WM, Howard PH (2004) SRC's EPI suite: *PHYSPROP database*. Syracuse Research Corporation