EPA/600/R-03/030
March 2003

PREDICTION OF CHEMICAL REACTIVITY PARAMETERS AND PHYSICAL PROPERTIES OF ORGANIC COMPOUNDS FROM MOLECULAR STRUCTURE USING SPARC

By

S.H. Hilal and S.W. Karickhoff
Ecosystems Research Division
Athens, Georgia

and

L.A. Carreira
Department of Chemisty
University of Georgia
Athens, GA

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

The computer program SPARC (SPARC Performs Automated Reasoning in Chemistry) has been under development for several years to estimate physical properties and chemical reactivity parameters of organic compounds strictly from molecular structure. SPARC uses computational algorithms based on fundamental chemical structure theory to estimate a variety of reactivity parameters. Resonance models were developed and calibrated on more than 5000 light absorption spectra, whereas electrostatic interaction models were developed using more than 4500 ionization pK_as in water. Solvation models (i.e., dispersion, induction, dipole-dipole, hydrogen bonding, etc.) have been developed using more than 8000 physical property data points on properties such as vapor pressure, boiling point, solubility, Henry's constant, GC retention times, K_{ow}, etc. At the present time, SPARC predicts ionization pK_a (in the gas phase and in many organic solvents including water as function of temperature), carboxylic acid ester hydrolysis rate constants (as function of solvent and temperature), $E_{1/2}$ reduction potential (as function of solvents, pH and temperature), gas phase electron affinity and numerous physical properties for a broad range of molecular structures.

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. 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 ERD-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.

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

TABLE OF CONTENTS

1. GENERAL INTRODUCTION

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 actually 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 (OPPTS) [1]. Hence, the need for physical and chemical constants of chemical compounds has greatly accelerated both in industry and government as assessments are made of potential pollutant exposure and risk.

Although a wide variety of approaches are commonly used in regulatory exposure and risk calculations, 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 [2-9]. Granted that considerable progress has been made in process elucidation and modeling for chemical processes [10-15], such as photolysis and hydrolysis, reliable estimates of the related fundamental thermodynamic and physicochemical properties (i.e., rate/equilibrium constants, distribution coefficient, solubility in water, etc.) 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 have been developed that allow estimation of some constants, such relationships are generally valid only within limited families of chemicals. Computer programs have been under development at the University of Georgia and U.S. Environmental Protection Agency for more than 12 years that predict a large number of chemical reactivity parameters and physical properties for a wide range of organic molecules strictly from molecular structure. This prototype computer program called SPARC (SPARC Performs Automated Reasoning in Chemistry) uses computational algorithms based on fundamental chemical structure theory to estimate a variety of reactivity parameters [16-26]. This capability crosses chemical family boundaries to cover a broad range of organic compounds. SPARC presently predicts numerous physical properties and chemical reactivity parameters for a large number of organic compounds strictly from molecular structure, as shown in Table 1.

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 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).

Although the primary emphasis in this report, and throughout the development of the SPARC program, has been aimed at supporting environmental exposure and risk assessments, the SPARC physicochemical models have widespread applicability (and are currently being used) in the academic and industrial communities. The recent interest in the calculation of physicochemical properties has led to a renaissance in the investigation of solute-solvent interactions. In recent ACS conferences, over one third of the computational chemistry talks have dealt with calculating physical properties and solvent-solute interactions.

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., a graduate class taught by the late Dr. Robert Taft at the University of California). Also, the SPARC calculator has been used for aiding industry (such as Pfizer, Merck, Pharmacia & Upjohn, etc.) in the areas of chemical manufacturing and pharmaceutical and pesticide design. The speed of calculation allows SPARC to be used for on-line control in many chemical engineering applications. SPARC can also be used for custom solvent and mixed solvent design to assist the synthesis chemist in achieving a particular product or yield.

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 SMILES (Simplified Molecular Input Line Entry System) notation, or via the CAS number, which will generate the SMILES notation. SPARC is programmed with the ALS (Applied Logic Systems) version of Prolog (PROgramming in LOGic).

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 local 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

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

2. SPARC COMPUTATIONAL METHODS

SPARC does not do a "first principles" computation; rather, SPARC seeks to analyze chemical structure relative to a specific reactivity query in much the same manner as an expert chemist would do. Physical 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 in SPARC that defines the role of structural constituents in affecting 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.

The SPARC computational approach is based on blending well known, established methods such as SAR (Structure Activity Relationships) [27, 28], LFER (Linear Free Energy Relationships) [29, 30] and PMO (Perturbed Molecular Orbital) theory [31, 32]. SPARC uses SAR for structure activity analysis, such as induction 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. In reality, every chemical property involves both quantum and thermal contributions and necessarily requires the use of all three methods for prediction.

A "toolbox" of mechanistic perturbation models has been developed that can be implemented where needed in SPARC for a specific reactivity query. Resonance perturbation models were developed and calibrated using light absorption spectra for more than 5000

compounds [1, 16], whereas electrostatic interaction perturbation models were developed using ionization pK_as in water for more than 4500 compounds [17-22]. Solvation perturbation models (i.e., dispersion, induction, H-bond and dipole-dipole) have been developed using physical properties data such as vapor pressure, boiling point, solubility, distribution coefficient, Henry's constant and GC chromatographic retention times for more than 8000 compounds [21, 23, 24]. Ultimately, these mechanistic components will be fully implemented for the aforementioned chemical and physical property models, and will be extended to additional properties such as hydrolytic and redox processes.

Any predictive method should be understood in terms of the purpose for which it is developed, and should be structured by appropriate operational constraints. SPARC's predictive methods were designed for engineering applications involving physical/chemical process modeling. More specifically, these methods provide:

1. an *a priori* estimate of the physicochemical parameters of organic compounds for physical and chemical fate process models when measured data are not available,
2. guidelines for ranking a large number of chemical parameters and processes in terms of relevance to the question at hand, thus establishing priorities for measurements or study,
3. an evaluation or screening mechanism for existing data based on "expected" behavior,
4. guidelines for interpreting or understanding existing data and observed phenomena.

3. CHEMICAL REACTIVITY PARAMETERS

Molecular structures are broken into functional units with known chemical properties called reaction centers, C. The intrinsic behavior of each reaction center is then "adjusted" for the

compound in question by describing mechanistically the effect(s) on reactivity of the molecular structure(s) appended to each reaction center using perturbation theory.

The SPARC chemical 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 macroscopic/microscopic ionization pK_a in water. The same reactivity models are used to estimate:

1. zwitterionic constant, isoelectric point, titration curve and speciation fractions as a function of the pH,
2. ionization pK_a in the gas phase,
3. ionization pK_a in non-aqueous solution,
4. gas phase electron affinity,
5. carboxylic acid ester hydrolysis rate constant in water and in non-aqueous solution.

3.1. Estimation of Ionization pK_a in Water

3.1.1 Introduction

A knowledge of the acid-base ionization properties of organic molecules is essential to describing their environmental transport and transformations, or estimating their potential environmental effects. For ionizable compounds, solubility, partitioning phenomena and chemical reactivity are all highly dependent on the state of ionization in any condensed phase. The ionization pK_a of an organic compound is a vital piece of information in environmental exposure assessment. It can be used to define the degree of ionization and resulting propensity for sorption to soil and sediment that, in turn, can determine a compound's mobility, reaction kinetics, bioavailability, complexation, etc. In addition to being highly significant in evaluating environmental fate and

effects, acid-base ionization equilibria provide an excellent development arena for electrostatic interaction perturbation models. Because the gain or loss of protons results in a change in molecular charge, these processes are extremely sensitive to electric field effects within the molecule.

Numerous investigators have attempted to predict ionization pK_a's using various approaches such as *ab initio* [33, 34] and semiempirical [35, 36] methods. The energy differences between the protonated and the unprotonated states are small compared to the total binding energies of the reactants involved. This presents a problem for *ab initio* computational methods that calculate absolute energy values. Computing the relatively small energy differences needed for the analysis of molecular chemical reactivity from the absolute energies requires extremely accurate calculations. Hence, the aforementioned calculation methods are generally limited to a small subclass of molecules. A more aggressive attempt was made by Klopman et. al., [37, 38]. They estimated the pK_a's for about 2400 molecules ($R^2 = 0.846$) based on QSAR using the Multi-CASE program. Despite the relatively large number of pK_a's estimated, their calculator was limited to only the first ionization site pK_a [38] for compounds processing multiple sites.

Unfortunately, up to now no reliable method has been available for predicting pK_a over a wide range of molecular structures, either for simple compounds or for complicated molecules such as dyes. The SPARC pK_a calculator has been highly refined and has been exhaustively tested. In this report, the calculation 'toolbox' will be described, along with testing results to date.

3.1.2. SPARC's Chemical Reactivity Modeling

Chemical properties describe molecules in transition, that is, the conversion of a reactant molecule to a different state or structure. For a given chemical property, the transition of interest may involve electron redistribution within a single molecule or bimolecular union to form a

transition state or distinct product. The behavior of chemicals depends on the differences in electronic properties of the initial state of the system and the state of interest. For example, a light absorption spectrum reflects the differences in energy between the ground and excited electronic states of a given molecule. Chemical equilibrium constants depend on the energy differences between the reactants and products. Electron affinity depends on the energy differences between the LUMO (Lowest Unoccupied Molecular Orbital) state and the HOMO (Highest Unoccupied Molecular Orbital) state.

For any chemical property addressed in SPARC, the energy differences between the initial state and the final state are small compared to the total binding energy of the reactants involved. Calculating these small energy differences by *ab initio* computational methods is difficult, if not impossible. On the other hand, perturbation methods provide these energy differences with more accuracy and with more computational simplicity and flexibility than *ab initio* methods. Perturbation methods treat the final state as a perturbed initial state and the energy differences between these two energy states are determined by quantifying the perturbation. For pK_a, the perturbation of the initial state, assumed to be the protonated form, versus the unprotonated final form is factored into the mechanistic contributions of resonance and electrostatic effects plus other perturbations such as H-bonding, steric contributions and solvation.

3.1.3. Ionization pK_a Computational Approach

Molecular structures are broken into functional units called the reaction center and the perturber. 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 perturber structure is assumed to be unchanged in the reaction. The pK_a of the reaction center

is either known from direct measurement or inferred indirectly from pK_a measurements. The pK_a of the reaction center is adjusted for the molecule in question using the mechanistic perturbation models described below.

Like all chemical reactivity parameters addressed in SPARC, pK_a is analyzed in terms of some critical equilibrium component:

$P\text{-}C_i \longleftrightarrow P\text{-}C_f$

where C_i denotes the initial protonated state, C_f is the final unprotonated state of the reaction center, C, and P is the "perturber". The pK_a for a molecule of interest is expressed in terms of the contributions of both P and C.

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

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. SPARC computes reactivity perturbations, $\delta_p(pK_a)_c$, that are then used to "correct" the ionization behavior of the reaction center for the compound in question in terms of the potential "mechanisms" for interaction(s) of P and C as

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

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 protonated and unprotonated states of C, respectively. Electrostatic interactions are derived from local dipoles or charges in P interacting with charges or dipoles in C. $\delta_{ele}pK_a$ represents the difference in the electrostatic interactions of the P with the two states. $\delta_{res}pK_a$ describes the change in the delocalization of π electrons of the two states due to P. This delocalization of π electrons is assumed to be into or out of the reaction center. Additional

perturbations include direct interactions of the structural elements of P that are contiguous to the reaction center such as H-bonding or the steric blockage of solvent access to C.

3.1.1.4. Ionization pK_a Modeling Approach

The modeling of the perturber effects for chemical reactivity relates to the structural representation S--$_i$R$_j$--C, where S--$_i$R$_j$ is the perturber structure, P, appended to the reaction center, C. S denotes substituent groups that "instigate" perturbation. For electrostatic effects, S contains (or can induce) electric fields; for resonance, S donates/receives electrons to/from the reaction center. R links the substituent and reaction center and serves as a conductor of the perturbation (i.e."conducts" resonant π electrons or electric fields). A given substituent, however, may be a part of the structure, R, connecting another substituent to C, and thus functions as a "conductor" for the second substituent. The i and j denote anchor atoms in R for S and C, respectively.

For each reaction center and substituent, SPARC catalogs appropriate characteristic parameters. Substituents include all non-carbon atoms and aliphatic carbon atoms contiguous to either the reaction center or a pi-unit. Some heteroatom substituents containing pi groups are treated collectively as substituents (e.g. -NO_2, -C≡N, -C=O, -CO_2H, etc.). The specification of these collective units as substituents is strictly facilitative. The only requisites are that they be structurally and electronically well-defined (charge and/or dipolar properties are relatively insensitive to the remainder of the perturber structure). Also, these units must be terminal with regard to resonance interactions (no pass-through conjugation). All hydrogen atoms are dropped and "bookkept" only through atom valence. An isoelectronic carbon equivalent plus an appended atom, Q, replace heteroatom substituents in these π units. For example -C=O- becomes C=C-Q, which is now treated in SPARC as perturbed ethylene.

In computing the contribution of any given substituent to $\delta_p(pK_a)_c$, the effect is factored into three independent components for the structural components C, S, and R:

1. substituent strength, which describes the potential of a particular S to "exert" a given effect. (Independent of the property, C and R),
2. molecular network conduction, which describes the "conduction" properties of the molecular structure R, connecting S to C with regard to a given effect, (Independent of the property, C and S), and
3. reaction center susceptibility, which rates the response of C to the effect in question (depends on the property, independent of S and R).

The contributions of the structural components C, S, and R are quantified independently. For example, the strength of a substituent in creating an electrostatic field effect depends only on the substituent regardless of the C, R, or property of interest. Likewise, the molecular network conductor R is modeled so as to be independent of the identities of S, C, or the property being estimated. The susceptibility of a reaction center to an electrostatic effect quantifies only the differential interaction of the initial state versus the final state with the electrical field. The susceptibility gauges only the reaction $C_{initial}$ - C_{final} and is completely independent of both R and S. *This factoring and quantifying of each structural component independently provides parameter "portability" and, hence, permits model portability to all structures and, in principle, to all types of reactivity.*

3.1.4.1. Electrostatic Effects Models

Electrostatic effects on reactivity derive from charges or electric dipoles in the appended perturber structure, P, interacting through space with charges or dipoles in the reaction center, C. Direct electrostatic interaction effects (field effects) are manifested by a fixed charge or dipole in a

substituent interacting through the intervening molecular cavity with a charge or dipole in the reaction center. The substituent can also "induce" electric fields in R that can interact electrostatically with C. This indirect interaction is called the "mesomeric field effect". In addition, electrostatic effects derived from electronegativity differences between the reaction center and the substituent are termed sigma induction. These effects are transmitted progressively through a chain of σ-bonds between atoms. For compounds containing multiple substituents, electrostatic perturbations are computed for each singly and summed to produce the total effect.

With regard to electrostatic effects, reaction centers are classified according to the electrostatic change accompanying the reaction. For example, monopolar reactions proceed with a change in net charge ($\delta q_c \neq 0$) at the reaction center and are denoted C_m; dipolar reactions, C_d, produce no net change in charge but involve a change in the dipole moment ($\delta\mu_c \neq 0$, $\delta q_c = 0$, etc.). The nature and magnitude of electrostatic change accompanying a reaction determine the "susceptibility" of a given reaction to electric fields existing in structure, P.

3.1.4.1.1. Field Effects Model

For a given dipolar or charged substituent interacting with the change in the charge at the reaction center, the direct field effect may be expressed as a multipole expansion

$$\delta(\Delta E)_{field} = \frac{\delta q_c q_s}{r'_{cs} D_e} + \frac{\delta q_c \mu_s \cos\theta_{cs}}{r_{cs}^{2} D_e} + \frac{\delta \mu_c q_s \cos\theta_{cs}}{r'^{2}_{cs} D_e} + \frac{\delta \mu_c \mu_s \cos\theta_{cs'} \cos\theta_{cs}}{r^{3}_{cs} D_e} + \ldots\ldots..$$

where q_s is the charge on the substituent, approximated as a point charge located at point, s'; μ_s is the substituent dipole located at point s (this dipole includes any polarization of the anchor atom i effected by S); q_c ($\delta\mu_c$) is the *change* in charge (dipole moment) of the reaction center

accompanying the reaction, both presumed to be located at point c; θ_{cs} is the angle the dipole subtends to the reaction center; D_e is the effective dielectric constant for the medium; and r_{cs} ($r_{cs}^{/}$) is the distance from the substituent dipole (charge) center to the reaction center.

In modeling electrostatic effects, only those terms containing the "leading" nonzero electric field change in the reaction center are retained. For example, acid-base ionization is a monopole reaction that is described by the first two terms of the preceding equation; electron affinity is described by only the second term, whereas the dipole change in H-bond formation is described by the third and fourth terms.

Once again, in order to provide parameter "portability" and, hence, effects-model portability to other structures and to other types of chemical reactivity, the contribution of each structural component is quantified independently:

$$\delta_{field}(pK_a)_c = \rho_{ele}\sigma_p = \rho_{ele}\sigma_{cs}F_S$$

where σ_p characterizes the field strength that the perturber exerts on the reaction center. ρ_{ele} is the susceptibility of a given reaction center to electric field effects that describes the electrostatic change accompanying the reaction. ρ_{ele} is presumed to be independent of the perturber. The perturber potential, σ_p, is further factored into a field strength parameter, F (characterizing the magnitude of the field component, charge or dipole, on the substituent), and a conduction descriptor, σ_{cs}, of the intervening molecular network for electrostatic interactions. This structure-function specification and subsequent parameterization of individual component contributions enables one to analyze a given molecular structure (containing an arbitrary assemblage of functional elements) and to "piece together" the appropriate component contributions to give the resultant reactivity effect. For

molecules containing multiple substituents, the substituent field effects are computed for each substituent and summed to produce the total effect as

$$\delta_{field}(pK_a)_c = \rho_{ele} \sum_{R=1}^{S} \sigma_{cs} F_s$$

The electrostatic susceptibility, ρ_{ele}, is a data-fitted parameter inferred directly from measured pK_as. This parameter is determined once for each reaction center and stored in the SPARC database. In parameterizing the SPARC electrostatic field effects models, the ionization of the carboxylic acid group was chosen to be the reference reaction center with an assigned ρ_{ele} of 1. For all the reaction centers addressed in SPARC, electrostatic interactions are calculated relative to a fixed geometric reference point that was chosen to approximate the center of charge for the carboxylate anion, r_{cj} = 1.3 unit, where the length unit is the aromatic carbon-carbon length (1.40A). The ρ_{ele} for the other reaction centers (e.g., OH, NR_2) reflect electric field changes for these reactions gauged relative to the carboxylic acid reference, but also subsumes any difference in charge distribution relative to the reference point, c.

With regard to the substituent parameters, each uncharged substituent has one field strength parameter, F_μ, characterizing the dipole field strength; whereas, a charged substituent has two, F_q and F_μ. F_q characterizes the effective charge on the substituent and F_μ describes the effective substituent dipole inclusive of the anchor atom i, which is assumed to be a carbon atom. If the anchor atom i, is a noncarbon atom, then F_μ is adjusted based on the electronegativity of the anchor atom relative to carbon. The effective dielectric constant, D_e, for the molecular cavity, any polarization of the anchor atom i affected by S, and any unit conversion factors for charges, angles, distances, etc. are included in the F's.

Initially, the distances between the reaction center and the substituent, r_{cs}, for both charges and dipoles are computed as the summation of the respective distance contributions of C, R and S as

$$r_{cs}^{o} = r_{cj} + r_{ij} + r_{is}$$

In some cases, such as in ring systems, this "zero-order" distance is adjusted (see below) for direct through-space interactions of S and C as opposed to interactions through the molecular cavity. However, these adjustments are significant only when C and S are ortho or perri (e.g., 1, 8-substituted naphthalene) to each other:

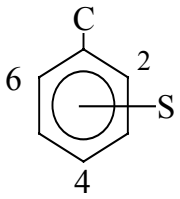
$$r_{cs} = A r_{cs}^{o}$$

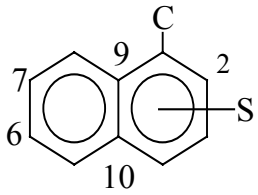
where A is an adjustment constant assumed to depend only on bond connectivity into and out of the R-π, unit (e.g., points i and j). For R-π units recognized by SPARC, "A factors" for each pair (i,j) are empirically determined from data (or inferred from structural similarity to other R-π units). The distance through R (r_{ij}) is calculated by summation over delineated units in the shortest molecular path from i to j. All aliphatic bonds contribute 1.1 unit; double and triple bonds contribute 0.9 and 0.8 units, respectively. For ring systems, SPARC contains a template listing distances between each constituent atom pair as illustrated in Table 2. The dipole orientation factors, $\cos\theta_{ij}$, are presently ignored (set to 1.0) except in those cases where S and C are attached to the same rigid R-π unit. In these latter situations, $\cos\theta_{ij}$s are assumed to depend solely on the point(s) of attachment, (i,j), and are pre-calculated and stored in SPARC databases.

The strength of the electrostatic interaction between S and C depends on the magnitude and relative orientation of the local fields of S and C and the dielectric properties and distances through the conducting medium. All uncharged dipole substituents and positively charged substituents will

increase the acidity of any acid, no matter what the charge, and hence, exert a +F. For a negatively charged substituent, the dipole field component tends to lower the pK_a, whereas the negative charge field component tends to raise the pK_a.

Table 2. Position on Ring and Geometry Parameters

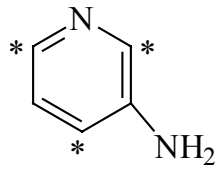




	Position on ring		Geometry parameters		
Molecule	Reaction Center	Substituent	r_{ij}	A_{ij}	$\cos\theta_{ij}$
benzene	1	2	1.0	0.25	0.53
	1	3	1.7	0.87	0.88
	1	4	2.0	1.00	1.00
naphthalene	1	2	1.0	0.25	0.53
	1	3	1.7	0.87	0.88
	1	4	2.0	1.00	1.00
	1	5	2.6	0.73	0.81
	1	6	3.0	0.63	0.83
	1	7	2.7	0.64	0.81
	1	8	1.7	0.47	0.77
	2	1	1.0	0.25	0.53
	2	3	1.0	0.25	0.53
	2	4	1.7	0.81	0.91
	2	5	3.0	0.63	0.83
	2	6	3.6	0.98	0.96
	2	7	3.4	0.80	0.84

3.1.4.1.2. Mesomeric Field Effects

As mentioned in the previous section, a substituent can also "induce" electric fields in the R that can interact electrostatically with C. This indirect interaction is called the "mesomeric field effect". For example, the amino group in the structure below exerts a +F direct effect that should normally lower the pK_a; however, the observed effect is exactly the opposite. The measured pK_a of m-amino pyridine is 6.1, and is greater than the pK_a of pyridine (5.2). In this case, the NH_2 induces charges ortho and para to the in-ring N. These charges interact indirectly with the dipole of the nitrogen in the ring and result in a net increase in the pK_a.



The contribution of the mesomeric field can be estimated as a collection of discrete charges, q_R, with the contribution of each described by the following equation. As is the case in modeling the direct field effects, the mesomeric effect components are resolved into three independent elements for S, R, and C as

$$\delta_{M_F}(pK_a)_c = \rho_{ele} q_R M_F$$

where M_F is a mesomeric field effect constant characteristic of the substituent S. It describes the ability or strength of a given substituent to induce a field in R_π. q_R describes the location and relative charge distributions in R, and ρ_{ele} describes the susceptibility of a particular reaction center to electrostatic effects. Since the reaction center can not discriminate the sources of the electric fields, ρ_{ele} is the same as that described previously in discussions of the direct field effects.

In modeling the mesomeric field effect, the intensity and the location of charges in R depend on both the substituent and the R_π network involved. The contributions of S and R_π are resolved by replacing the substituent with a reference probe or NBMO (NonBonded Molecular Orbital) charge source. This NBMO reference source for SPARC was chosen to be the methylene anion, $-CH_2^-$, for which the charge distribution in any arbitrary R_π network can be calculated.

The mesomeric substituent strength parameter describes the π-induction ability of a particular substituent relative to the CH_2^-. The magnitude of a given substituent M_F parameter describes the relative field strength, whereas the sign of the parameter specifies the positive (electron withdrawing such as NO_2) or negative (electron donating such as NR_2) character of the induced charge in R_π. The total mesomeric field effect for a given substituent is given by:

$$\delta_{M_F}(pK_a)_c = \rho_{ele} M_F \sum_k \frac{q_{ik}}{r_{kc}}$$

where q_{ik} is the charge induced at each atom k, with the reference probe attached at atom i, calculated using PMO theory. r_{kc} is the through-cavity distance to the reaction center as described previously for direct fields. Because induction does not change total molecular charge, the sum of all induced charges must be zero. This is achieved by placing, at the location of the substituent, a compensating charge, q_s, equal to but opposite to the total charge distributed within the R_π network.

3.1.4.1.3. Sigma Induction Effects Model

Sigma induction derives from electronegativity differences between two atoms. The electron cloud that bonds any two atoms is not symmetrical, except when the two atoms are the same and have the same substituents; hence, the higher electronegativity atom will polarize the

other. This effect is transmitted progressively between atoms, and dies off rapidly with distance, i.e. $\sim 0.4^n$, where n is the number of bonds through which the effect is transmitted.

The interaction energy of this effect depends on the difference in electronegativity between the reaction center and the substituent and on the number of substituents bonded to the reaction center. Sigma induction effects are resolved into two independent structural component contributions: that of the substituent, S, and that of the reaction center, C.

$$\delta_{Sigma} (pK_a)_c = \rho_{ele} \sum (\chi_s - \chi_c) NB$$

where ρ_{ele} is the susceptibility of a given reaction center to electric field effects. Once again, because the reaction center cannot discriminate the source of the electric fields, ρ_{ele} is the same as that described for the direct field effect. χ_c is the effective electronegativity of the reaction center. χ_s is the effective electronegativity of the substituent. NB is data-fitted parameter that depends on number of the substituents that are bonded directly to the reaction center. The electronegativity of reaction centers and substituents referenced to the electronegativity of the methyl group, chosen to be the reference group for this effect.

3.1.4.2. Resonance Effects Model

Resonance involves variations in charge transfer between the π system and a suitable orbital of the substituent. The interaction of the substituent orbital with a π-orbital of a reaction center can lead to charge transfer either to or from the reaction center. Electron withdrawing reaction centers will localize the charge over itself. As a result the acidic state will be stabilized more than the basic state making these compounds less acidic. For electron donating reaction centers, resonance will stabilize the basic state more than the acidic state and lower the pK_a.

Resonance stabilization energy in SPARC is a differential quantity, related directly to the extent of pi electron delocalization in the neutral state versus the ionized state of the reaction center. The source or sink in the perturber P, may be either the substituents or R-π units contiguous to the reaction center. As with the case of electrostatic perturbations, structural units are classified according to function. Substituents that withdraw electrons are designated S+ while electron donating groups are designated S-. The R-π units withdraw or donate electrons, or serve as a "conductor" of π electrons between resonant units. Reaction centers are likewise classified as C+ or C-, denoting withdrawal or donation of electrons, respectively.

In SPARC, the resonance interactions describe the delocalization of an NBMO electron or electron hole out of the initial state, (C_i) or final state, (C_f) into a contiguous R-π or conjugated substituent(s). To model this effect, a surrogate electron donor, CH_2^-, replaces the reaction center. The distribution of NBMO charge from this surrogate donor is used to quantify the acceptor potential for the substituent and the molecular conductor. The resonance perturbation of the initial state versus the final state for an electron-donating reaction center is given by:

$$\delta_{res}(pK_a)_c = \rho_{res}(\Delta q)_c$$

where $(\Delta q)_c$ is the fraction loss of NBMO charge from the surrogate reaction center calculated based on PMO theory (see Appendix). ρ_{res} is the susceptibility of a given reaction center to resonance interactions. ρ_{res} quantifies the differential "donor" ability of the two states of the reaction center relative to the reference donor CH_2^-. In the parameterization of resonance effects, resonance strength is defined for all the substituents (i.e., the ability to donate or receive electrons); resonance susceptibility is defined for all the reaction centers; and resonance "conduction" in R_π networks is

modeled so as to be portable to any array of R_π units or to the linking of any resonant source or sink group.

3.1.4.3. Solvation Effects Model

If a base is more solvated than its conjugate acid, its stability increases relative to the conjugate acid. For example, methylamine is a stronger base than ammonia, and diethylamine is stronger still. These results are easily explainable due to the sigma induction effect. However, trimethylamine is a weaker base than dimethylamine or methylamine. This behavior can be explained due to the differential hydration of the reaction center of interest and the reaction center.

The initial and the final states of the reaction center frequently differ substantially in degree of solvation, with the more highly charged moiety solvating more strongly. Steric blockage of the reaction center can be distinguished from steric-induced twisting of the reaction center in electron delocalization interaction models. Differential solvation is a significant effect in the protonation of organic bases (e.g., $-NH_2$, in-ring N, =N), but is less important for acidic compounds except for highly branched aliphatic alcohols.

In SPARC's reactivity models, differential solvation of the reaction center is incorporated in $(pK_a)_c$, ρ_{res} and ρ_{ele}. If the reaction center is bonded directly to more than one hydrophobic group or if the reaction center is *ortho* or *perri* to hydrophobic substituent, then $\delta_{solv}(pK_a)_c$ must be calculated. The $\delta_{solv}(pK_a)_c$ contributions for each reaction center bonded directly to more than one hydrophobic group are quantified based on the sizes and the numbers of hydrophobic groups attached to the reaction center and\or to the number of the aromatic bridges that are *approximate* to the reaction center using the following equation:

$$\delta_{solv}(pK_a)_c = \rho_{solv}(\nu_i + \nu_j + \nu_k)$$

where ρ_{solv} is the susceptibility of the reaction center to differential solvation due to steric blockage of the solvent, v are the solid angles occluded by the hydrophobic P that is bonded directly (i), *ortho* (j), or *perri* (k) to the reaction center, respectively.

3.1.4.4. Intramolecular H-Bonding Effects Model

Intramolecular hydrogen bonding is a direct site coupling of a proton donating (α) site with a proton accepting (β) site within the molecule. Reaction centers might interact with substituents through intramolecular H-bonding and thus impact the pK_a. The initial, C_i, and final, C_f, states of the reaction center frequently differ substantially in degree of hydrogen bonding strength with a substituent.

In aromatic, π-ring or π-aliphatic (i.e., diguanide) systems where the reaction center is contiguous to the substituent and where a stable 5 or 6 member ring may be formed, $\delta_{H\text{-}B}(pK_a)_c$ must be estimated. $\delta_{H\text{-}B}(pK_a)_c$ is a differential quantity that describes the H-bonding differences of the initial versus the final state of a reaction center with a substituent, and is given by:

$$\delta_{H-Bond}(pK_a)_{c-s} = HB_C S_i ML_s$$

where HB_c is the H-bond contribution for C-S when C and S adjacent to each other, S_i is a reduction factor for steric-induced twisting of C, and ML_s is either 1 or 0.7 for aromatic and π-ring systems, respectively. For a reaction center that might H-bond with more than one substituent, the H-bonding contribution for each substituent is calculated and the strongest contributor to H-bond is selected.

3.1.4.5. Statistical Effects Model

All the SPARC perturbation models presented thus far describe the ionization of an acid at a single site. If a molecule contains multiple equivalent sites, a statistical correction is required. For example, if a first ionization constant, K, is computed for a single site, and if the molecule has N such sites, then

$$\delta_{Sstat}\ (pK_a)_c = \log\frac{N_a}{N_b}$$

where a and b refer to the acid and conjugate base sites, respectively.

3.1.4.5.6. Temperature Dependence

For processes that can be modeled in terms of some equilibrium (or pseudo equilibrium component) the temperature dependance can be expressed by the Van't Hoff representation:

$$f(\Delta pK_a) = A_c + \delta_s(\Delta pK_a)_c + [B_c + \delta_H(\Delta pK_a)_c]/T$$

where A_c and B_c are the entropic and the enthalpic van't Hoff coefficients for the reaction center, and δ_H and δ_s are enthalpic and entropic perturbations, respectively. To date, all perturbations have been assumed to be predominantly enthalpic. The van't Hoff factors (A and B) can be derived from temperature data for the reaction center or inferred from simple structures with minimal perturbational contributions. An example of the temperature dependence of pK_a for the amino reaction center is shown in Figure 1. When the enthalpic perturbation cancels the B parameter as in the *para* nitroaniline example, little or no temperature dependence is observed. Some systems may have perturbations large enough to change the sign of the slope of the pK_a temperature dependence.

3.1.5. Results and Discussion

To date, the approach used in SPARC to predict chemical reactivity parameters has been applied to UV-visible spectra, pK_a in water, electron affinity and carboxylic acid ester hydrolysis rate constants. The computational algorithm is based on structure query. This involves simply combining perturbation potentials of perturber units with reaction susceptibilities of the reaction center. It is important to reemphasize that the reaction parameters describing a given reaction center (Table 4) are the same regardless of the appended molecular structures. Likewise, for substituents, the parameters in Table 3 are independent of the rest of the molecule. This structure factoring and function specification enables one to construct, for a given reaction center of interest, essentially any molecular array of appended units, and to compute the resultant reactivity.

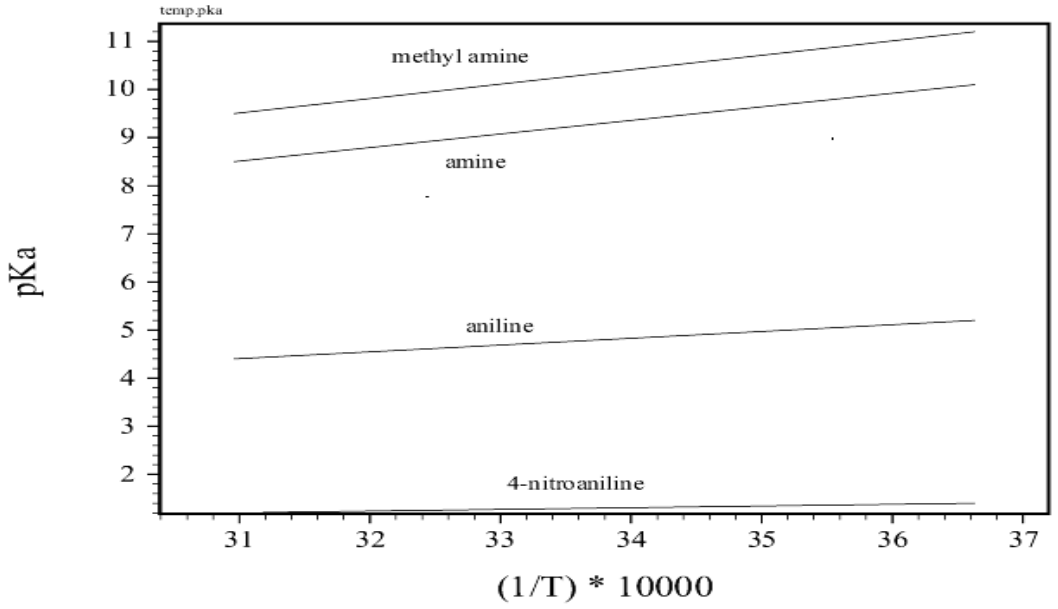


Figure 1. pK_a temperature dependence for selected molecules

Table 3. SPARC Substituent Characteristics Parameters

Substituent	F_s	F_q	M_F	E_r	r_{is}	χ_s
CO_2H	2.233	0.000	0.687	0.072	0.80	3.43
CO_2^-	1.639	-0.603	0.560	2.978	1.00	2.68
AsO_3H^-	0.300	-0.500	0.500	0.190	1.20	2.60
AsO_3^{-2}	0.600	-1.000	0.300	0.150	1.20	2.60
AsO_2H	1.000	-2.000	0.000	0.080	0.80	2.60
PO_3H^-	0.600	-0.786	0.400	0.220	1.20	3.32
PO_3^{-2}	0.600	-2.500	0.400	0.840	1.20	2.90
BO_2H_2	1.078	0.000	1.010	1.484	0.80	2.40
SO_3^-	6.315	-1.224	2.491	1.407	0.80	2.82
OH	1.506	0.000	-3.116	7.240	0.80	2.76
SH	2.931	0.000	-1.871	3.000	0.80	2.76
O^-	1.913	-1.566	-3.546	11.00	-0.50	3.01
S^-	1.727	-1.537	-1.437	9.368	-0.50	3.34
NR_2	1.190	0.000	-4.939	17.42	0.70	2.58
NR_2H^+	3.978	0.779	-2.505	21.70	0.50	3.23
CH_4	-1.10	0.000	-2.065	0.129	-0.63	2.30
NO_2	7.460	0.000	2.515	3.677	1.00	3.79
NO	6.714	0.000	4.127	1.691	1.00	3.80
CN	5.649	0.000	3.141	3.196	0.80	3.71
OR	2.138	0.000	-4.767	1.987	0.80	2.90
SR	2.323	0.000	-1.234	1.952	0.80	2.80
I	4.270	0.000	0.000	4.928	0.75	2.95
Br	3.756	0.000	-0.031	3.012	0.70	3.19
Cl	3.622	0.000	-0.066	1.498	0.65	3.37
F	3.164	0.000	-1.718	0.800	0.65	3.67
in-ring N	5.310	0.000	0.929	2.055	0.00	3.30
in-ring NH^+	1.379	3.785	6.995	8.708	0.00	3.80
SO_2	6.451	0.000	2.038	4.176	0.80	3.60
=N	1.533	0.000	0.544	4.918	0.00	3.80
$=NH^+$	2.000	1.000	2.800	2.600	0.00	3.80
=O	3.195	0.000	1.584	2.281	0.00	3.60

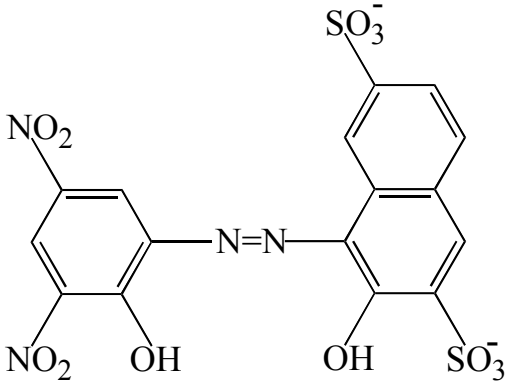
Table 4. Reaction Center Characteristics Parameters

C	ρ_{elc}	ρ_{res}	χ_c	$(pK_a)_c$
CO_2H	1.000	-1.118	2.60	3.75
AsO_2H	0.653	-0.817	2.22	6.63
PO_2H	0.489	-0.394	2.72	2.23
POSH	0.291	-0.402	2.69	1.55
PS_2H	0.101	-0.802	2.63	1.96
BO_2H_2	0.355	-0.050	3.04	8.32
SeO_3H	1.207	-0.400	2.30	4.64
SO_3H	0.451	-4.104	2.09	-0.10
OH	2.706	18.44	2.49	14.3
SH	2.195	4.348	2.76	7.40
NR_2	3.571	19.36	2.40	9.83
in-ring N	5.726	-11.279	2.31	2.28
=N	5.390	-4.631	2.47	5.33

Carbon and Nitrogen acid parameters are included in this table

The perturbations of some reaction centers such as oxy acids are small, whereas OH, NR_2, in-ring N and =N reaction centers have large perturbations. For example, the perturbation of the OH in the molecule below may be large as 12 pK_a units. The resonance and the electrostatic contributions of the two nitro and the =N groups substantially overcome the H-bond contributions of the OH with either the =N or the nitro groups making the pK_{a1} extremely acidic. On the other hand, the field effect of the negatively charged groups (SO_3^- and O^-) and the H-bond of the second OH with the =N will raise the second pK_{a2} and overcome the resonance contribution and the field effect of the uncharged groups.

	Observed	SPARC
pK_{a1}	2.0	1.7
pK_{a2}	12.2	12.1



3.1.6. Testing and Training of Ionization pK_a

The ionization pK_a calculator was trained on some 2400 compounds involving all the substituents and reaction centers shown in Table 3 and 4. The overall training set RMS deviation was 0.36 pK_a units. In addition, the SPARC pK_a calculator has been tested on 4338 pK_as (excluding carbon acid) for some 3685 compounds, including multiple pK_a's up to the sixth pK_a spanning a range of over 30 pK_a units as shown in Figure 2. The overall RMS deviation error for this large test set of compounds was found to be 0.37 pK_a units. While it is difficult to give a precise standard deviation of a SPARC calculated value for any 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. For complicated structures where a molecule has multiple ionization sites (N > 6) such as azo dyes, the expected SPARC error is ±0.65 pK_a units.

While the pK_a for simple structures can be measured to better than 0.1 pK_a units in the same laboratory. The interlaboratory RMS deviation error among the observed pK_a for simple organic molecules reported by IUPAC was not better than 0.3 pKa units even for simple carboxylic acid derivatives (see Table 5). For complicated structures, especially those with multiple ionization sites, the RMS deviation was much higher. For example, SPARC was used/tested to estimate 358 pK_a's for 214 azo dyes [18]. For these compounds, the SPARC calculated RMS deviation was 0.63 pK_a units. The experimental error reported by IUPAC for azo dyes was as high as 2 pK_a units [18]. The IUPAC reported RMS interlaboratory deviation between observed values of pK_a for azo dyes, where more than one measurement was reported was 0.64 [18]. Several examples of interlaboratory error for simple and relatively complicated molecules are shown in Table 5. We,

therefore, believe that the errors in SPARC-calculated values are comparable to experimental error, and perhaps better for these complicated molecules. We also note that the diversity and complexity of the molecules used for pK_a model development and testing has been drastically increased in the last few years in order to develop more robustness. A summary of the statistical parameters for the SPARC ionization pK_a in water calculator is shown in Table 6. For a sample hand calculation see reference 19.

In this rigorous test, almost all the organic molecules reported in the IUPAC series were included. The only compounds that were removed for this test were those that:

1. Form covalent hydrates. These include many of the multiple in-ring N compounds such as quinazoline and pteridine. See hydration rate section.
2. Are known to tautomerize, e.g., molecules such as methyl-substituted imidazole. See tautomeric constant section.
3. Carbon acid reaction center where the perturbations for this group are very large, and the measurement standard deviation is not better than 1 unit. For example, the pK_a's for methane, nitro-methane, tri-nitro-methane are 52, 10, 3.6, respectively. (SPARC calculates the pK_a for carbon acid within $\pm$ 1.3 pK_a units).
4. SPARC has not yet been designed to calculate, such as quaternary amines.

SPARC also may not be able to discriminate positional substituents effects for an oxy acid reaction center (where the perturbations are extremely small) in structures such as 3- or 4- S-C_6H_4-YC where Y is some side chain intervening between the benzene ring (e.g., Y = $(CH_2)_x$) and the reaction center, ($C{=}CO_2H$). SPARC can discriminate these effects for other reaction centers, C, such as NR_2 as shown in Table 7.

Table 5. Interlaboratory Measurements Error Range in pK_a For Simple and Relatively Complicated IUPAC Molecules

Molecule[a]	Range of Measurements
Phenol	pK_a 9.78 - 10.02
2-methylphenol	pK_a 10.10 - 10.33
3-methylphenol	pK_a 9.82 - 10.10
4-methylphenol	pK_a 10.02 - 10.28
Citric acid	pK_1 2.79 - 3.13
	pK_2 4.11 - 4.78
	pK_3 5.34 - 6.43
Dibutylpropanedioic acid	pK_1 1.89 - 2.64
	pK_2 7.19 - 7.70
Biphenyl-2-ol	pK_a 10.01 - 11.3
Uracil	pK_1 9.38 - 9.51
	pK_2 12.0 - 14.2
4-Dimethylamino-azobenzene	pK_1 3.2 - 3.50
	pK_2 -4.50 - (-1.3)
4-Dimethylamino-4'hydroxy- azobenzene	pK_1 -1.81 - 2.95
	pK_2 -2.3 - 3.40
2,4-Diamino-1,3,5-triazine	pK_a 3.90 - 5.88

a* In compiling pK_a's for this study, it was necessary to compile data from many laboratories. We used IUPAC-screened data, but even these data had relatively large variation, even for simple molecules as shown above.

Table 6. Statistical Parameters of SPARC pK_a Calculations

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

Table 7. Observed vs. Calculated pKa for 3 and 4-S-C_6H_4-$(CH2)_X$-C

C	NR_2				CO_2H			
S	**4-OC**		**3-OC**		**4-Cl**		**3-Cl**	
X	**Obs.**	**Cal.**	**Obs.**	**Calc.**	**Obs.**	**Cal.**	**Obs.**	**Cal.**
1	5.3	5.11	4.3	4.43	3.98	3.76	3.83	3.65
2	9.6	9.52	9.15	9.31	4.19	4.35	4.14	4.34
3	.....	9.92	....	9.77	4.65	4.59	4.58	4.59
4	.....	10.06	....	9.96	....	4.66	...	4.66
	.....	10.13	....	10.04	....	4.69	...	4.69

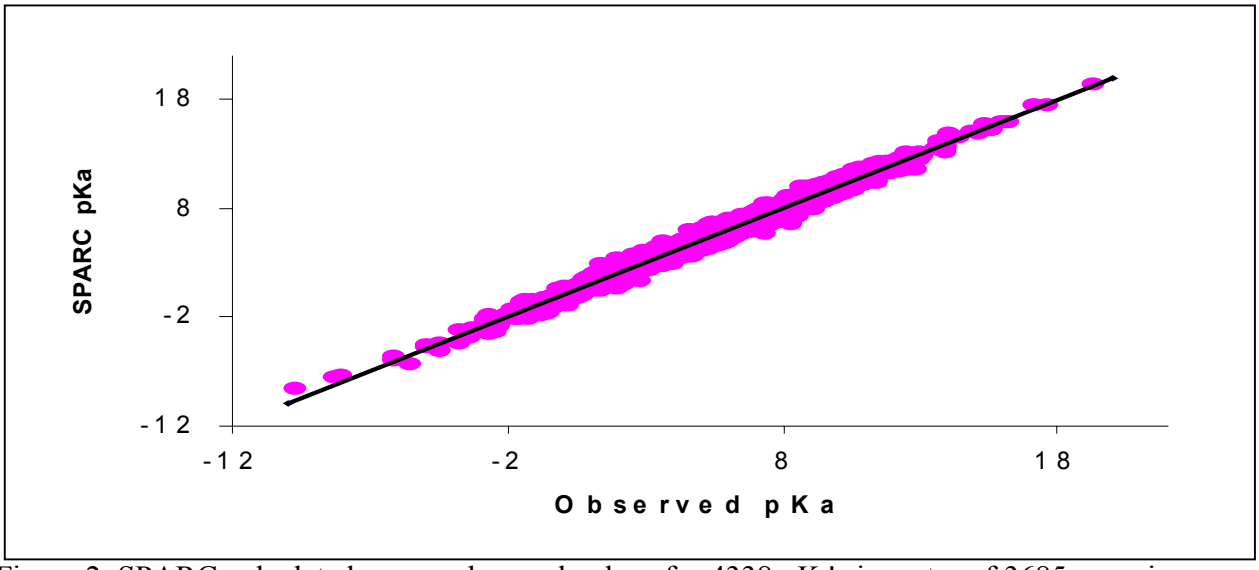


Figure 2. SPARC-calculated versus observed values for 4338 pK_a's in water of 3685 organic compounds. The overall RMS deviation was equal to 0.37. This test does not include carbon acid. The majority of the molecules are complex, e.g., some of the molecules have 8 different ionization sites (azo dyes).

3.1.7. Conclusion

The pK_a models are the most robust and most highly tested of the SPARC models. The models are fully implemented and are executing in code. However, the real test of SPARC does not lie in its predictive capability for pK_a's but is determined by *the extrapolatability of these models to other types of chemistry.* The SPARC chemical reactivity models used to predict ionization pK_a in water have been successfully extended to calculate many other properties (see next section).

3.2. Estimation of Microscopic, Zwitterionic Ionization Constants, Isoelectric Point and Molecular Speciation

3.2.1. Introduction

Determination of microscopic constants and zwitterionic ratios has played an important part in understanding the ionic composition of many biologically active molecules, particularly since all proteins fall into this class. The chemical and biological activities of these substances vary with the degree of ionization. For this reason, accurate knowledge of the ionization constants for zwitterionic substances is a prerequisite to an understanding of their mechanism of action in both chemical and biological processes.

Unfortunately, microscopic constants have been determined for less than 100 compounds, and for only a very few of these molecules has the zwitterionic constant been determined or calculated [39-43]. Moreover, determination or calculation of the fraction of the various microscopic species as a function of pH has been reported in the literature for less than a dozen molecules. Most of these measurements were restricted to aliphatic amino acid derivatives and only for simple, two ionization site molecules such as glycine and cysteine (where the CO_2H is already ionized). Benesch [40] calculated the relative concentration of the four microscopic forms for cysteine where the carboxylic acid group(s) was ionized in all the forms. He found that the concentration ratio of the $^{-}S\text{-}R\text{-}NH_3^{+}$ species to the $HS\text{-}R\text{-}NH_2$ species at any given pH was approximately 2 to 1 rather than 1 to 1 as suggested by Grafius [40]. This difference indicates the magnitude of the uncertainty involved in the various approximations made to calculate the microscopic constants and the relative concentration of the different species. As noted earlier, only a very few of the total number of microconstants needed to characterize the equilibria have been

measured or calculated. For example, only two microconstants have been determined for molecules with 4-ionizable sites such as DOPA and Epinephrin [42]. Estimation or measurement of the microscopic constants and relative concentration of the various species for such compounds is an extremely difficult task.

The SPARC pK_a calculator can be used to estimate the microscopic constants for almost any molecule of interest strictly from molecular structure. Hence, the microscopic ionization constants, the zwitterionic constant and the fraction of the various microscopic species as function of pH can be estimated without approximations such as limiting the number of species considered. The titration curves (charge versus pH) can also be calculated using the same reactivity models.

3.2.2. Calculation of Macroconstants

A Brönsted acid is defined as a proton donor and a Brönsted base as a proton acceptor. The acid-base ionization properties in solution are generally expressed in terms of ionization constants (pK_as) that describe the tendency for an acid to give up a proton to a solvent or the affinity of a base for a hydrogen ion. The strength of an acid in a solvent is measured by the ionization constant for the reaction. Many molecules of great importance in chemistry and biochemistry contain more than one acidic or basic site, and some macromolecules such as amino acids, peptides, proteins and nucleic acids may contain hundreds of such groups. These latter molecules may exist in a great number of distinct ionization states. The acidic groups are uncharged in strongly acidic solutions and negatively charged in sufficiently alkaline solutions. The basic groups are positively charged (protonated) in a strongly acidic solution and are uncharged in sufficiently alkaline solution. For a bifunctional acidic compound the ionization equilibria are usually written as

$H_2A \longleftrightarrow H^+ + HA^- \qquad K_1 = [HA^-][H^+]/[H_2A]$

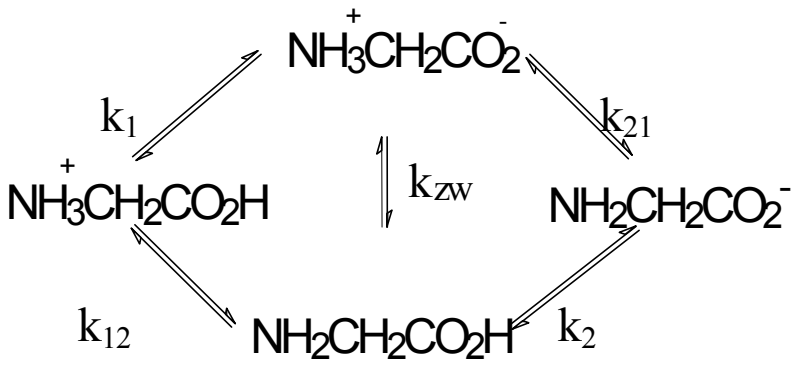
$HA^- \longleftrightarrow H^+ + A^{-2} \qquad K_2 = [A^{-2}][H^+]/[HA^-]$

where the constant concentration of the solvent has been absorbed in K. The pK_1 and pK_2 (pK_a = - log K_a) are commonly evaluated from a pH titration or spectroscopic measurement. These measured pK_a's are termed macroscopic constants because they often only describe a composite of the processes which are actually occurring in solution. The actual donor sites where the protons reside are not specified and may not be unique. Thus, a solution "species" such as H_2A may, in fact, consist of several H_2A species with protons occupying different basic sites in each of the species. On the other hand, microconstants are the equilibrium constants for equilibria involving individual species in solution. For these complex compounds, microconstants may or may not be capable of being either measured or determined distinctly.

3.2.3. Zwitterionic Equilibria: Microscopic Constant

Many molecules contain both an acid and a base functionality, but these sites are not able to ionize simultaneously. Such molecules are usually referred to as amphoteric. Amino phenols are good examples of amphoteric molecules. When the pH is very low, the cationic species predominates while at high pH the anionic species predominates. At intermediate pH's the molecule exists in the neutral form. Other substances contain both acid and base functionality where both the base and the acid sites may be simultaneously ionized to form an internal salt. These substances are referred to as zwitterionic or dipolar ions. The amino acids are an example of molecules that can exist as zwitterions. At low pH and high pH the cationic species and the anionic species predominate, respectively, as in the case of the amino phenol. But unlike the amphoteric amino phenol, the internal salt predominates as an intermediate species over a wide range of pH.

Actually, the zwitterion and the isomeric uncharged molecule are in equilibrium in aqueous solutions. The nature of this equilibrium depends on the acid and the base strength of the ionizing groups involved. For a molecule with two ionizable sites, such as glycine, this process can be represented diagrammatically below.



Each ionizable group has two microscopically different ionization pathways (k_1 and k_{12}, for loss of the first hydrogen and k_2 and k_{21} for loss of the second hydrogen). Each group has two constants associated with its ionization, one when the other group is ionized and one when the other group is not ionized.

When the pK_a's of the ionizing groups are arithmetically far apart (as those of glycine shown in Figure 3) knowledge of the two macroscopic constants, K_1 and K_2, is enough to calculate speciation as a function of pH. When the two pK_a values lie within 3 pK_a units of one another, such as in the case of N-phenylglycine (as shown in Figure 4), a more detailed survey of the problem becomes necessary. In such a case, the two macroscopic constants, K_1 and K_2, cannot fully describe the equilibria denoted by the four microscopic constants: k_1, k_{12}, k_{21}, k_2 and the zwitterionic constant k_{zw}. However, the macroscopic and microscopic equilibrium constants are closely related by the following equations:

$$K_1 = k_1 + k_{12}$$

$1 / K_2 = 1/ k_{21} + 1 / k_2$

The four microscopic ionization constants (k_i) involving the individual, microscopically distinct species describe precisely the acid-base chemistry of such a system at the molecular level while the two macroscopic pK_as provide an incomplete specification of the equilibria. It should be noted that the four constants are not independent but are subject to the relation $k_1 k_{12} = k_{21} k_2$. In order to calculate molecular speciation as a function of pH, k_{zw}, must be calculated. k_{zw} may be determined using the left or the right path of the above scheme. Since both thermodynamic paths give the zwitterion product, k_{zw} may be expressed as function of the microscopic constants within any loop as

$$k_{zw} = \frac{k_1}{k_{12}} = \frac{k_2}{k_{21}}$$

The integrity of the pK_a calculator in SPARC can be checked by calculating k_{zw} using the two different loops. The RMS deviation in calculated versus measured pk_{zw}s for the cases tested to date is 0.5 pK_a units [22]. This value is what one would expect from a calculation requiring two pK_a calculations (0.37 * √2). k_{zw}s calculated from different thermodynamic paths are averaged over the number of thermodynamic paths available.

3.2.4. Speciation of Two Ionizable Sites

In the pH range from 4 to 10, more than 99% of glycine in solution exists in a zwitterionic form where both the carboxylic and the amine groups are simultaneously charged. Over a wide pH range only 3 microscopic species have significant concentrations as is shown in Figure 3. The

concentration of the fourth species (neutral) is negligible (below 1%) over the entire pH range. The ratio of the zwitterionic concentration to the neutral concentration is about $4 * 10^5$. The macroconstants in a figure such as Figure 3 occur when the fraction of the ionizing group of interest is reduced to 50%. So, the left and the right-hand side where the fraction is equal to 50% are the macroscopic pK_{CO2H} and the pK_{NH2}, respectively. The microconstants interconnecting the species of interest occur where the species curves intersect. In the case where the two macroconstants are very far apart, the two macroconstants pK_1 and pK_2 are equal to the microconstants pk_1 and pk_{21}, respectively, as shown in Figure 3. Hence, the two macroconstants can satisfactorily describe the equilibrium in this instance.

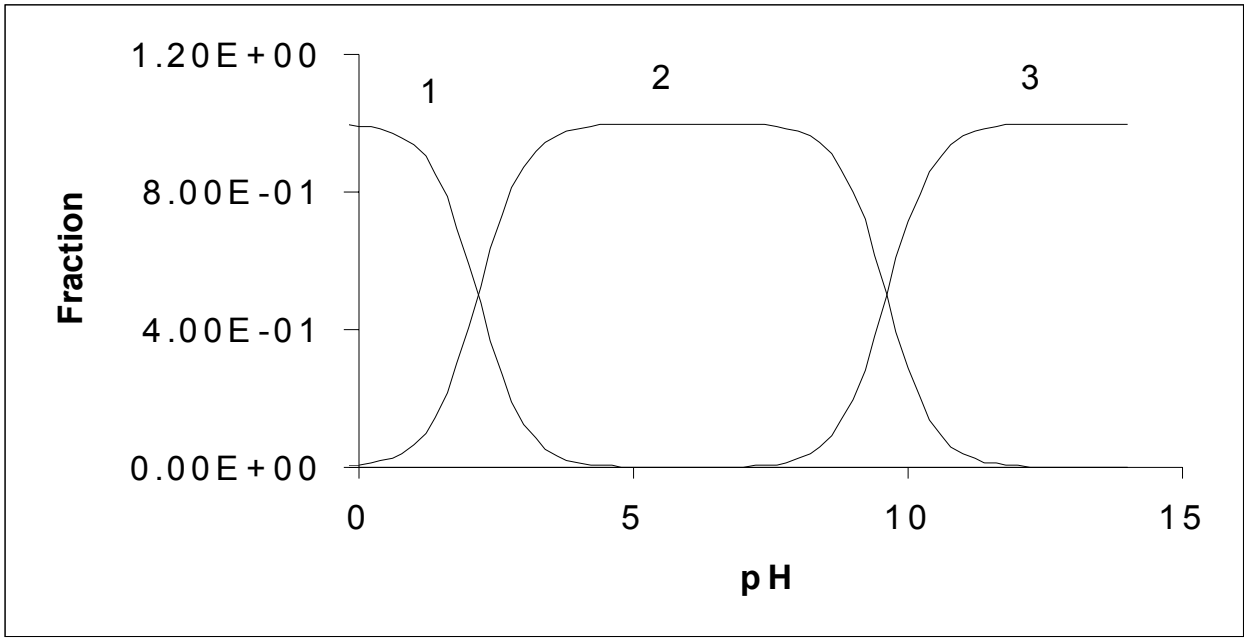


Figure 3. Fraction of the major microscopic species of glycine as a function of pH. The two macroscopic pK_a s are far apart and are equal to the microscopic constants (pk_{12} and pk_{21}). The number on the top of each curve corresponds to the microscopic species: 1 is positively charged; 2 is the zwitterion; and 3 is the negatively charged species.

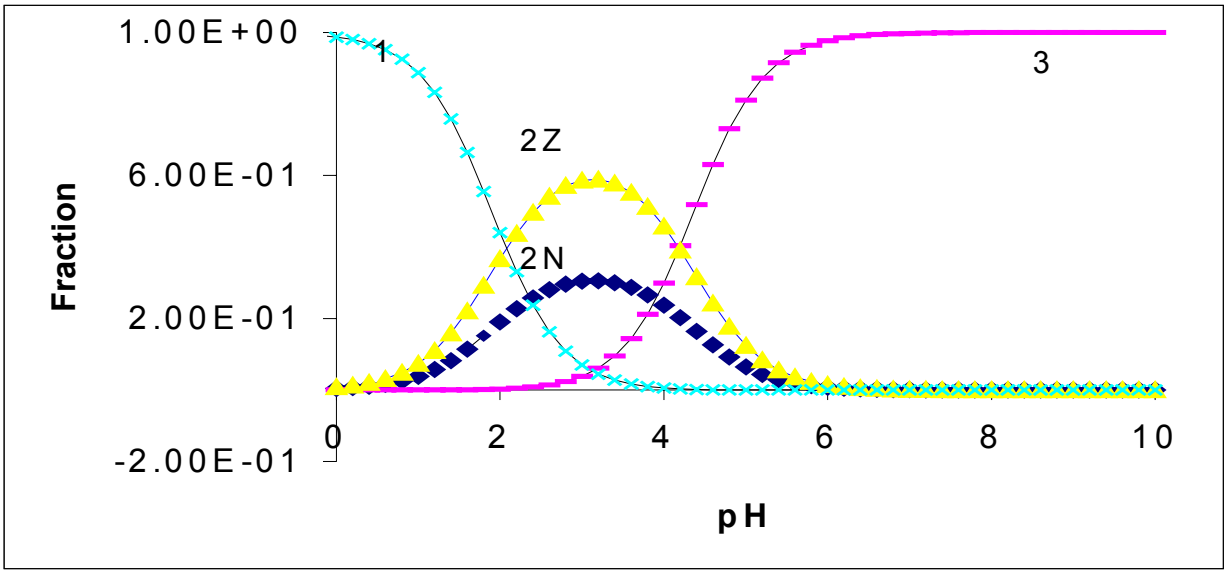


Figure 4. Fraction of the major microscopic species of N-phenylglycine as a function of pH The numbers on the top of each curve correspond to the microscopic species. 1, 3, 2N, 2Z corresponds to the positive, negative, neutral and the zwitter ion species, respectively.

On the other hand, N-phenyl substitution of glycine substantially lowers the macroscopic constant of the amine group due to resonance contributions. As a result, the amine and the carboxylic acid have comparable hydrogen ion affinities and both functional groups make important contributions to the hydrogen ion concentration (i.e., appreciable concentrations of the acidic and the basic forms of both functional groups are present in solution simultaneously). The macroconstants become more nearly equal (within 2 pK_a units) and the ratio of the zwitterionic species to the neutral species in solution decreases as indicated in Table 8. In this case the macroconstants are not equal to the microconstants (as is shown in Figure 4) and the equilibrium cannot be satisfactory described by pK_1 and pK_2. Substituents with a large dipole moment, such as a nitro or cyano group, will further decrease the zwitterionic ratio due to electrostatic and/or resonance effects. For example, m-nitro- and m-cyano-phenylglycines exist in aqueous solution predominantly in the non-zwitterionic form due to the large electrostatic effect of the dipolar

group. Compounds with weaker dipole substituents, such as methyl- and methoxy phenylglycines, exist largely in the zwitterionic form. Substituent effects on pK_a are illustrated in Figure 5 for glycine, N-(phenyl) glycine, N-(m-nitrophenyl) glycine and N-(m-methoxyphenyl) glycine where the charge lost in the molecule as function of the pH is plotted. The zwitterion ratio k_{zw} is very dependent on the nature of the substituent in these molecules. The proportion of zwitterions in aqueous solution is governed by the effect of the substituent on the pK_a's, and can vary substantially as shown in Table 8. Table 8 shows the observed versus SPARC-calculated microscopic ionization constants, pk_{ij}, and zwitterionic constants, pk_{zw}, for two molecules with ionizable site.

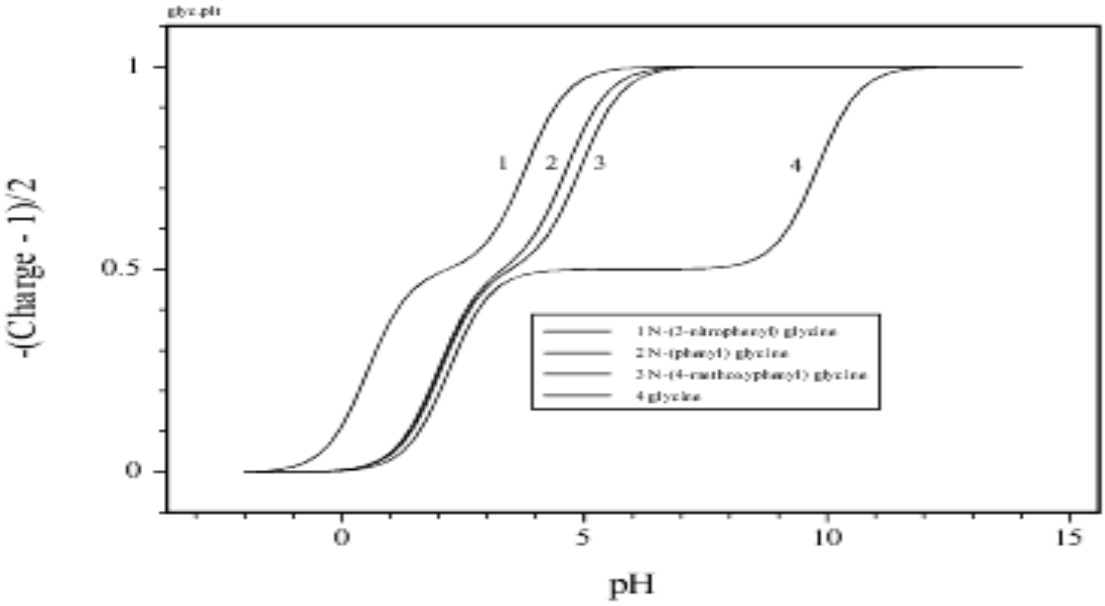


Figure 5. Total charge versus pH titration curves for (1) N-(m-nitrophenyl) glycine, (2) N-(phenyl) glycine, (3) N-(m-methoxyphenyl) glycine and (4) glycine.

Table 8. Observed vs. SPARC calculated microscopic constants for molecules with 2 ionizable site.

	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
Molecule	pk_{zw}		pk_1		pk_{12}		pk_2		pk_{21}	
Glycine	-5.60	-5.60	2.35	2.15	8.00	7.8	9.80	9.70	4.43	4.20
Phenylglycine	-0.26	-0.58	2.03	2.15	2.29	2.67	4.22	4.60	3.96	3.95
m-NO2-	...	1.35	...	2.00	0.06	0.50	..	2.50	3.75	3.80
m-CN-	...	1.10	...	2.00	0.28	0.80	..	2.70	3.78	3.80
m-Cl-	0.90	0.40	2.01	2.12	1.10	1.63	2.99	3.55	3.90	3.86
m-COMe-	0.85	0.07	2.03	2.11	1.20	1.90	3.05	3.87	3.87	3.87
p-Cl-	0.36	0.05	1.99	2.10	1.61	1.95	3.51	3.88	3.89	3.86
m-OMe-	-0.21	-0.27	2.09	2.16	1.89	2.33	3.74	4.26	3.95	3.90
m-Me-	-0.32	0.70	2.06	2.15	2.38	2.83	4.43	4.76	4.00	3.96
p-Me-	-0.70	-0.90	2.05	2.16	2.75	3.00	4.77	4.90	4.07	3.95
p-OMe-	-1.00	-0.94	2.12	2.19	3.11	3.10	5.07	5.00	4.07	3.98
4-Aminobenz. acid	0.93	0.87	3.40	3.71	2.47	2.83	3.90	3.80	4.83	4.61
4-Dimethaminobnz	0.62	0.42	3.28	3.74	2.66	2.98	4.28	4.39	4.9	4.51
3-Aminobenz. acid	-0.43	-0.72	3.22	3.40	3.65	4.00	4.66	4.79	4.23	4.02
Picolinic acid	-1.15	-1.2	1.04	1.37	2.21	2.67	5.29	5.28	4.12	4.12
Nicotinic acid	-1.00	-1.18	2.11	2.42	3.13	3.46	4.77	4.87	3.75	3.52
Isonicotinic acid	-1.40	-1.12	1.86	2.61	3.26	3.56	4.84	4.78	3.44	3.48
Tyrosine ethylester	....	1.59	9.63	9.10	7.33	7.3	...	8.35	...	9.75
5-Thiomethyl-Imid.	1.15	2.0	7.72	8.51	6.57	5.91	8.36	8.34	9.51	9.91
2-Thiomethyl-Imid.	-0.46	0.25	6.50	6.81	6.96	6.91	9.13	9.31	8.67	8.70
1-Me-2-thioMeImid.	-0.36	0.37	6.47	6.80	6.83	6.90	8.75	9.40	8.39	8.70
Tyramine	-0.4	-0.76	9.58	9.43	...	9.97	10.68	11.02	...	10.1
N,N-dimethyl-	-0.09	-0.35	9.03	9.4	...	9.56	10.31	10.68	...	10.0
DOPA	---	----	8.97	9.1	9.62	9.61	9.17	9.15	9.40	9.42
Dopamine	----	----	8.90	8.86	-----	----	10.1	9.93	---	----

3.2.5. Speciation of Multiple Ionization Sites

Martin, et. al [41] measured the 12 microscopic constants for tyrosine. They used various approximations to estimate the fraction of all the tyrosine species present in which the hydroxy group was ionized. They assumed that the macroscopic constant K_1 was equal to the microscopic constant k_1 for the CO_2H group and that the ionization of the hydroxy group was completely independent of the ammonium group. In addition, they assumed that the molar extinction coefficients for several species were identical. Martin [42] also used this approach to calculate the speciation of DOPA (l-3,4-dihydroxyphenylalanine), that is where the phenolic groups are ionized, as a function of pH. To the best of our knowledge, estimation of all the different microscopic species for molecules having four or more ionizable sites has not been reported.

For a molecule that has N ionizable sites, there are N macroscopic ionization constants that can be measured. There are, however, $2^{N-1} * N$ microscopic ionization constants and 2^N microscopically different species or states. For example, tyrosine in a strongly acidic solution contains 3 ionizable protons attached to a carboxyl, aromatic hydroxyl and ammonium group. Since each of the three groups may exist in either of two states, tyrosine may exist in 8 (2^3) microscopically different forms. The most positive of these 8 states is the cation, with net charge Z =1; the most negative is the divalent anion, with Z = -2. Each of the two intermediate states of net charge Z =0 and Z = -1, respectively, can have three microscopically different forms. Each of the ionizable groups in tyrosine is characterized by four microconstants, since the tendency of each group to accept or donate a proton depends on the ionization state of the other two groups. Hence, there are 12 ($3 * 2^2$) microscopic ionization constants connecting the 8 species. Three macroscopic ionization constants (K_1, K_2, K_3) for tyrosine have been determined experimentally from titration and spectroscopic data [43].

For tyrosine, only 5 of the 8 species ever have appreciable concentration [22]. The fraction of each of the microscopic species formed by a molecule with three ionizable sites (e.g. tyrosine or cysteine) can be expressed as function of pH in terms of the microconstants. If we start from the neutral species (uncharged species) rather than from the positively charged species, the fraction of any microscopic species for a molecule having N ionizable sites can be expressed in general as $D_{ij...k}/D_T$ where D_T can be expressed [22] as

$$D_T = \frac{1}{0!} + \frac{\sum_i k_i [H^+]^{L_i}}{1!} + \frac{\sum_i \sum_{j \neq i} k_i k_{ij} [H^+]^{L_{ij}}}{2!} + \ldots + \frac{\sum_i \sum_{j \neq i} \cdots \sum_{k \neq i,j\ldots} k_i k_{ij} \cdots k_{ij\ldots k} [H^+]^{L_{ij\ldots k}}}{N!}$$

and $L_{ij..k}$ is the charge of the final state (ij..k state). The factorial is the number of different thermodynamic paths that lead to the ij..k state and $D_{ij...k}$ is one of terms in the denominator. For example, the fraction of neutral species would be $1/D_T$ and the fraction of a singly ionized species would be $k_i * [H^+]^{Li}/D_T$.

The fraction of any distinct species as function of pH (fraction-species curve) can be determined from the Equation above. Whenever the total net charge of two (or more) charged species are equal, the maximum of the corresponding fraction-species curves will occur at the same pH. This can be shown by estimating the ratio of the fraction of any two equally charged species using the above equation. The H ion dependence will cancel and the ratio of the two fractions will be totally independent of pH. In addition, the titration curve (charge curve) can determined by multiplying the fraction-species curve by the charge on the species and summing over all species (Figure 5). The macroscopic pK_as can be determined by taking the first derivative of the titration curve [22]. Table 9 shows the observed versus SPARC calculated microscopic constants for several systems containing 3 ionizable sites. The notation for the macroconstants follows the scheme first proposed by Hill [44] and used later by R. Martin, et. al [43]. The ionizing group of interest is indicated in the microscopic pk by the last number in the subscript. Any number

preceding this in the subscript denotes another specified group in the molecule that already exists in the basic form when the ionization under consideration is taking place. Thus, pk_{32} denotes the pk value for the ionization of the OH group when the NH_3^+ group has already been converted to the conjugate base $-NH_2$. Since the number 1 does not appear in the subscript 32, its absence denotes that group 1, the carboxyl, is still in the un-ionized form during the reaction corresponding to the pk value in question.

Table 10 shows the observed versus the SPARC-calculated microscopic pk_is for glutathione where two CO_2H groups, an SH and a NH_3^+ can be ionized simultaneously in solution. Since each of the four groups may exist in either of two states (acidic or basic) the molecule may exist in 2^4 states. To describe the population of the possible 16 microscopic species, 32 microscopic ionization constants are required (see reference 22). However, only 8 of these constants have been measured for glutathione. In general, for N possible ionizable sites in a molecule there are NI microconstants that lead to a molecular state of ionization of IS where IS is the number (≤N) of sites that are ionized. NI may expressed as

$$NI = \frac{N!}{(IS-1)!(N-IS)!}$$

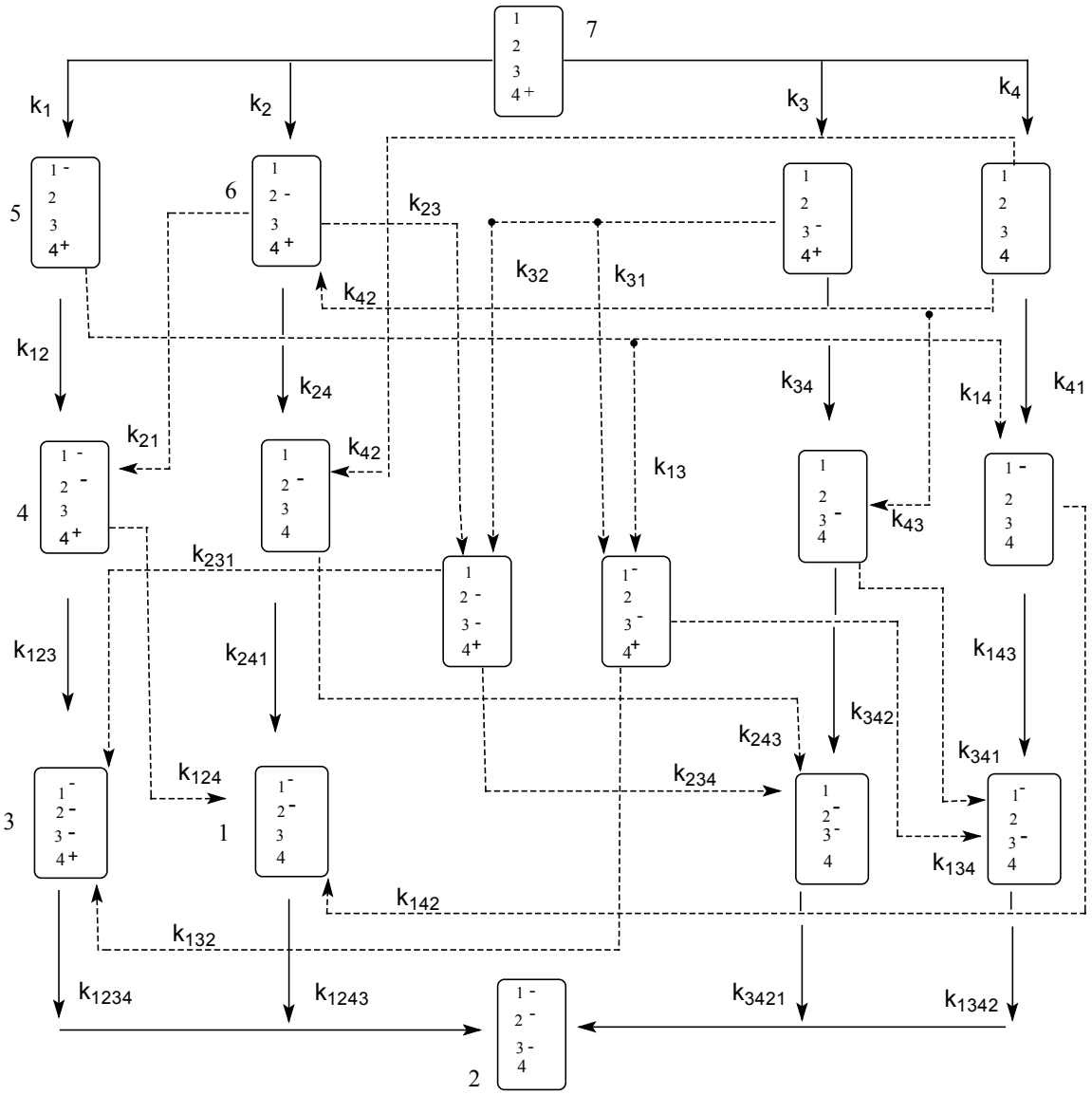
For example, in the glutathione case N = 4 and there are 4 microscopic constants leading to both one (IS = 1) and four ionized sites (IS = 4) and 12 microconstants for each of the two and three (IS = 2 and 3) ionized species. Only 7 microscopic species have an appreciable concentration between pH 0-14 [22]. The complete ionization relational scheme for glutathione microscopic species and the corresponding microconstants are illustrated in Figure 6.

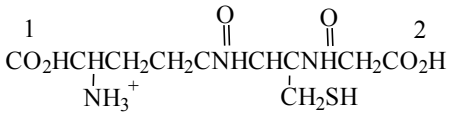
Table 9. Observed vs. SPARC-calculated value for the microscopic pk_is for 3 ionizable site molecules.

	Tyrosine		Cysteine		Cysteine glycine		Cysteine ethylester		Glutamic acid	
	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
pk_1	2.21	2.00	1.71	1.80	..	3.17	...	...	2.15	2.13
pk_{21}	2.61	2.30	2.79	2.40	..	3.40	...	...	2.62	2.30
pk_{31}	4.37	3.90	3.80	3.80	..	3.50	...	...	4.30	4.10
pk_{231}	4.77	4.20	4.74	4.40	..		...	...	4.74	4.20
pk_2	9.31	9.30	7.45	7.80	..	7.10	7.45	7.08	3.85	4.20
pk_{12}	9.71	9.60	8.53	8.20	7.87	7.40	...	...	4.32	4.30
pk_{32}	9.91	10.0	9.50	9.09	..	8.90	9.09	8.88	4.65	4.60
pk_{132}	10.3	10.3	10.0	10.0	9.45	9.20	...	...	5.09	4.80
pk_3	7.19	7.30	6.77	6.70	..	6.50	6.77	6.29	7.04	7.87
pk_{13}	9.35	9.60	8.60	8.86	7.14	6.88	...	...	9.19	9.50
pk_{23}	7.79	8.40	8.41	8.60	..	8.43	8.41	8.38	7.84	8.40
pk_{123}	9.95	10.6	10.36	10.5	8.75	8.80	...	...	9.96	10.0

Table 10. Observed vs. SPARC calculated microscopic ionization constants for glutathione.

	CO_2H (1)			CO_2H (2)			SH (3)			NH_3^+ (4)	
pk_{ijk}	Obs.	Calc.	pk_{ijk}	Obs.	Calc.	pk_{ijk}	Obs.	Calc.	pk_{ijk}	Obs.	Calc.
pk_1	2.09	1.92	pk_2	3.12	3.26	pk_3	....	7.94	pk_4	....	7.04
pk_{21}	2.33	1.98	pk_{12}	3.36	3.31	pk_{13}	....	8.14	pk_{14}	....	8.65
pk_{31}	....	2.06	pk_{32}	....	3.50	pk_{23}	....	8.21	pk_{24}	....	7.31
pk_{41}	....	3.84	pk_{42}	....	3.35	pk_{43}	....	8.24	pk_{34}	....	7.64
pk_{241}	....	3.91	pk_{132}	....	3.56	pk_{123}	8.93	8.37	pk_{124}	9.13	8.88
pk_{231}	....	2.12	pk_{142}	....	3.41	pk_{243}	....	8.49	pk_{134}	....	9.26
pk_{341}	....	4.10	pk_{342}	....	3.60	pk_{143}	....	8.43	pk_{234}	....	7.86
pk_{3421}	...	4.10	pk_{1342}	....	3.65	pk_{1243}	9.08	8.91	pk_{1234}	9.28	9.50





Glutathione

Figure 6. The 16 microscopic states of glutathione and the 32 microscopic ionization constants that interrelate them. See ref. 22 for more details.

Another example, hemimellitic acid (1, 2, 3-benzenetricarboxlic acid) presents unusual ionization behavior. The micro constants and the observed macroscopic constants are not identical. The first ionization step favors leaving the molecule ionized at the 2 position and is stabilized by hydrogen bonding with the carboxylic groups in positions 1 and 3. The second step is more complicated. Here, the most stable di-anion is the species ionized at positions 1 and 3. This minimizes electrostatic interactions. Going from the molecule ionized at the 2 position to a di-anion ionized at the 1 and 3 positions is not a simple one proton loss. This process involves three protons as shown in Figure 7. Figure 7 also shows the calculated microscopic species distribution of hemimellitic acid as function of pH. The thermodynamic steps for the different two ionization paths, the SPARC calculated results for each step (the micro constants), the first and the final step (the observed macro constants) are shown in Figure 8.

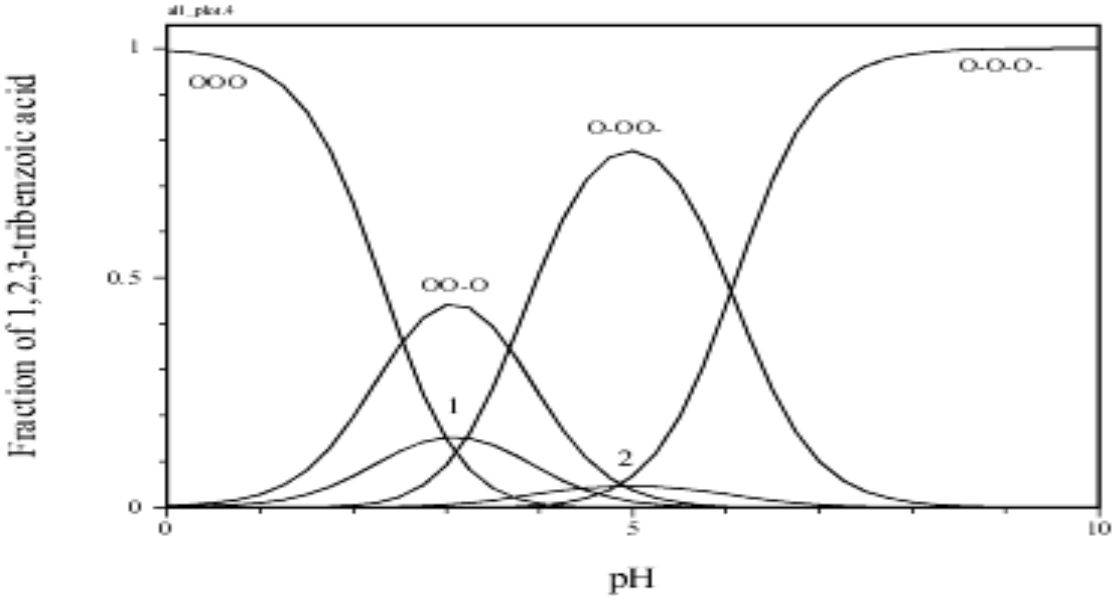


Figure 7. The calculated fraction of the eight microscopic species of Hemimellitic acid versus pH. 4 of the microscopic species are shown on the top of the graph. Graphs 1 & 2 show the other 4 microscopic species; each having two different symmetrical species lying on the top of each other: O^-OO/OOO^- and $O^-O^-O/O\ O^-\ O^-$, respectively

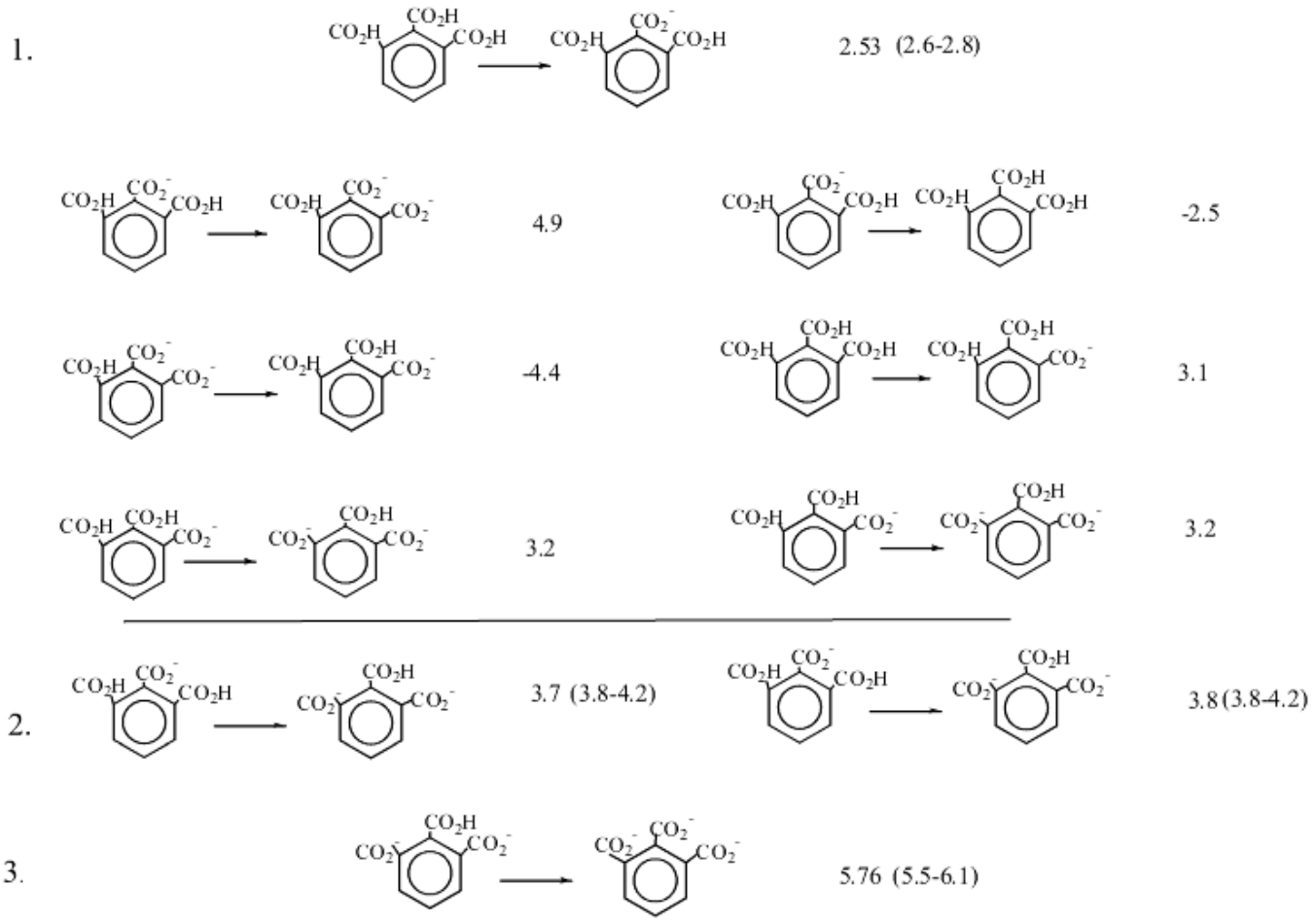


Figure 8. The three ionization macroscopic pK_as for hemimellitic acid. The second macroscopic pK_a involves three different microscopic constants that are shown to the side of each step. The Two paths are calculated within 0.1 pK_a units of each other. The observed macroscopic are the numbers between the bracket. The others are SPARC-calculated pK_as.

3.2.6. Isoelectric Points

Many molecules, such as amino acids, peptides, and proteins, contain both acidic and basic groups. The acidic sites for these molecules are uncharged in strongly acidic solutions and are negatively charged in sufficiently alkaline solutions. In contrast, the basic groups are positively charged in strongly acid solution, and the conjugate bases are uncharged in sufficiently alkaline

solution. Therefore, in an electric field such a molecule migrates as a cation in strongly acid solution and as an anion in strongly alkaline solution. At some intermediate pH value, the mean net molecular charge, Z, must attain the value zero, and the molecule will remain stationary in an electric field. The pH value at which this occurs is known as the isoelectric point. Edsall and Wyman [45] points out "for most simple ampholytes of this type, such as glycine or phenylglycine, pK_1 and pK_2 are so far apart that there is not merely an isoelectric point, but a broad zone of pH values in which the ampholyte is practically isoelectric". However, Edsall showed that the theoretical ioselectric point (pH_I) for a two ionizable sites molecule, such as those mentioned above, is given by

$$pH_I = \frac{pK_1 + pK_2}{2}$$

For polyvalent ampholytes, Edsall showed that for molecules containing 3 ionizing groups where one of the macroscopic pK_a's is far apart from the other two pK_as, a reasonable approximation can be made to calculate the isoelectric point. For example the isoelectric point for lysine may simply be expressed as $(pK_2 + pK_3)/2$. Unfortunately, for more complicated systems, a very rough approximation has to be made. With SPARC, the isoelectric point can be estimated by plotting the fraction of neutral or zwitterionic species versus pH (e.g. Figure 3, 4). The pH at the middle of the zwitterionic (or any other species where the total net charge of the molecule is zero) range is labeled as the isoelectric point. The observed versus SPARC-calculated isoelectric points for several molecules are shown in Table 11.

The procedure for calculating the zwitterionic equilibrium constant has been automated by allowing the user to specify all the zwitterionic acid-base pairs directly at the level of SMILES input. For example, the input for ß-amino propionic acid would be N@+CC(=O)O@-. The SPARC molecular parser now recognizes the @ as a request to perform an automatic calculation of

the zwitterionic constant along with the four relevant pK_as. The user can also generate plots of the relative concentrations of neutral, cationic, anionic and zwitterionic species as a function of pH.

Table 11. SPARC-Calculated Isoelectric Points.

No	Molecule	Obs.	Calc.
1	Glycine	6.0	5.8
2	Cysteine	5.1	5.0
3	Lysine	10	9.8
4	Glutamic acid	3.2	3.2
5	Penicillamine	4.9	5.0
6	Phenylglycine	3.1	3.2
7	m-NO2-	1.9	2.0
8	m-CN-	2.0	2.1
9	m-Cl-	2.5	2.6
10	m-COMe-	2.5	2.6
11	p-Cl-	2.7	2.8
12	m-Ome-	2.9	3.0
13	m-Me-	3.2	3.3
14	p-Me-	3.4	3.4
15	p-OMe-	3.6	3.4
16	Thiazolidine-4-carboxylic acid	3.9	4.4
17	2-Methyl-	4.4	4.6
18	2,2-Dimethyl	4.2	4.7
19	5,5-Dimethyl	4.2	4.4
20	2,5,5-Trimethyl	4.1	4.5
21	2,2,5,5-Tetramethyl	4.2	4.7
22	2-Ethyl-2-methyl-	5.2	4.7
23	2-Ethyl-2,5,5-trimethyl	5.2	4.7
34	Niflumic acid	3.30	3.1

3.2.7. Conclusion

The SPARC chemical reactivity models used to estimate ionization pK_a in water can also predict zwitterionic and microscopic ionization constants, pk_i, of organic molecules with multiple ionization sites that are as reliable as most experimental measurements. The corresponding complex speciation for these molecules as a function of pH and the titration curve can also be estimated using the same models without modification. The chemical reactivity models are fully implemented and are executing in code.

3.3. Estimation of Gas Phase Electron Affinity

3.3.1. Introduction

One of the fundamental properties of gaseous negative ions is the lowest energy required to remove an electron. This energy is called the electron affinity (EA). The electron affinity of a molecule plays an important role not only in gas-phase ions, but also in condensed-phase and charge-transfer complexes in chemistry, biology and physics. Although gaseous negative ions in molecules were first observed around 1900 and have been studied extensively since then, the first reliable EA was not obtained until the early 1960's for O_2. Since that time, numerous methods have been utilized to predict and measure electron affinity. Despite the availability of a large number of methods and the fundamental importance of electron affinity values, the complexity of molecular negative ions and the inherent difficulties in determining electron affinity have prevented the determination of this important property for many molecules of interest. Wide disagreement also exists among reported values.

3.3.2. Electron Affinity Computational Methods

The electron affinity property of a molecule describes the conversion of the neutral molecule to a molecular negative ion when both the neutral molecule, E, and the negative ion, E^-, are in their most stable state. Electron affinity is defined as the difference in energy between a neutral molecule plus an electron at rest at infinity and the molecular negative ion when both the neutral molecule and the negative ion are in their ground electronic, vibrational and rotational states.

The added electron enters the LUMO (Lowest Unoccupied Molecular Orbital), which in the negative ion becomes the HOMO (Highest Occupied Molecular Orbital). The lower the energy of

the LUMO, the greater will be the electron affinity and *vice versa*. The energy differences between the LUMO state and the HOMO state are small compared to the total binding energy of the reactant involved; hence, perturbation theories can be used to calculate the energy differences between these two states. Perturbation theory treats the final state as a perturbed initial state and the energy difference between these two energy states determined by quantifying the perturbation. The perturbation of the HOMO state versus the LUMO state is factored into mechanistic components of resonance, field and sigma induction contributions.

The successful application of the SPARC chemical reactivity models developed for pK_a to the calculation of electron affinities demonstrates the power of the molecular toolbox approach. Knowledge and models developed in the arena of ionization pK_a were directly applied to another chemical reactivity problem. Similar to SPARC's approach for calculating ionization pK_a, molecular structures are broken into functional units having known intrinsic electron affinities $(EA)_c$. The intrinsic behavior is then adjusted for the molecule in question using the mechanistic perturbation models described for ionization pK_a.

3.3.3. Electron Affinity 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). All reactions to be addressed in SPARC are analyzed in terms of critical equilibrium components:

$P\text{-}C_I \longleftrightarrow P\text{-}C_f$

where C_i and C_f denote the initial state and the final state or the LUMO state and the HOMO state of the reaction center, respectively; P is the structure that is presumed unchanged by the reaction; and EA denotes the electron affinity reaction. Electron affinity is expressed as a function of the energy required to add an electron to the LUMO state. It represents the energy difference between the neutral molecular state and the molecular negative ion state. To model this energy difference, electron affinity is expressed in terms of the summation of the contributions of all the components, perturber(s) and reaction center(s), in the molecule:

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

where the summation is over n, which is defined as the number of subunits that could form a molecular negative ion or simply as the number of reaction centers in the molecule. $(EA)_c$ is the electron affinity for the reaction center. The electron affinity of the reaction center is assumed to be unperturbed and independent of P. $\delta_p(\Delta EA)_c$ is a differential quantity that describes the change in the EA behavior affected by the perturber structure. $\delta_p(\Delta EA)_c$ is factored into mechanistic contributions as

$$\delta_p(\Delta EA)_c = \delta(\Delta EA_{field}) + \delta(\Delta EA_{res}) + \delta(\Delta EA_{sig})$$

where ΔEA_{field} describes the difference in field interactions of P with the two states, ΔEA_{res} describes the change in the delocalization of pi electrons of the two states due to P (this delocalization of π electrons is assumed to be into or out of the reaction center), and ΔEA_{sig} describes the change in sigma induction of P with the two states.

3.3.1.1. Field Effects Model

The modeling of the substitents field effects relates to the structural representation S--$_i$R$_j$--C, where S--$_i$R$_j$ is the perturber structure, P, appended to the reaction center C, i and j denote the anchor atoms in R for the substituent S and reaction center C, respectively.

Field effects derive from charges or electrical dipoles in the appended structure P, interacting with the charges or the dipoles of the reaction center, C, through space. The molecular conductor, R, acts as a low dielectric conductor. This effect follows from the fact that the bonds between most atoms are not completely covalent, but possess a partial ionic character that imposes electrical asymmetry either in the substituent or the reaction center bonds. The field interaction between S and C depends on the magnitude and the relative orientation of the local fields of S and C, the dielectric properties of the conduction medium and the distances through the molecular cavity. The field effect of a given S is given as

$$\delta(\Delta EA)_{field} = \frac{\delta q_c \mu_s \cos\theta_{cs}}{r_{cs}^{\,2} D_e}$$

where μ_s is the substituent dipole located at point S; δq_c is the change in charge of the reaction center accompanying the reaction, presumed to be located at point C; θ_{sc} gives the orientation of the substituent dipole relative to the reaction center; r's are the appropriate distances of separation; and D_e gives the effective dielectric constant for the intervening conduction medium.

Field effects are resolved into three independent structural component contributions representing the change in dipole field strength of S, a conduction factor of R, and the change in the charge at C as

$$\delta(\Delta EA)_{field} = \rho_{ele}\sigma_P = \rho_{ele}\sigma_R F_S$$

where σ_p describes the potential of P to "create" an electric field irrespective of C, and ρ_{ele} is the susceptibility of a given reaction center to electric field effects that describes the change in the electric field of the reaction center accompanying the reaction. The perturber potential, σ_p, is further factored into the field strength parameter, F_S, (describing the magnitude of the dipole field component on the substituent) and a conduction descriptor, σ_R, of the intervening molecular network for the electrostatic interactions.

The ρ_{ele} for the electron affinity of the reaction centers are data-fitted parameters that are inferred directly from measured electron affinity data. The field strength parameter, F_S, for each substituent is inferred from measurements of ionization pK_a. The distances among the various components and the orientation angle are calculated from geometry models and stored in the SPARC database as previously explained for pK_a.

3.3.2. Sigma Induction Model

Sigma induction derives from electronegativity differences between two atoms. This effect is transmitted progressively through a chain of π-bonds among atoms. This is a short-range interaction that is strong when the two atoms are bonded to each other, and any effect beyond the second atom is negligible. As is the case in modeling field effects, sigma induction effects are resolved into the three independent structural component contributions of S, R and C, characterizing the change in the difference of the electronegativity between the substituent and the reaction center, a conduction factor of R, and the change in the electrostatic effects of C.

$$\delta_{Sigma}\,(\Delta EA)_c \quad = \quad \rho_{ele} \sum\,(\chi_s - \chi_c)\,NB$$

where ρ_{ele} as indicated previously, is the susceptibility of a given reaction center to electric field effects. χ_c and χ_s are the effective electronegativity of the reaction center and the substituent,

respectively. NB is data-fitted parameter that depends on number of the substituents that are bonded directly to the reaction center, C. Both NB and χ_s are the same as those used for ionization pK_a calculations.

3.3.3. Resonance Model

Resonance involves the delocalization of pi electrons into or out of the reaction center. This long-range interaction is transmitted through the π -bond network. The resonance reactivity perturbation, $\rho_{res}(\Delta EA)$, is the differential resonance stabilization of the initial versus final state of the reaction center. It is a differential quantity, related directly to the extent of electron delocalization in the neutral state versus the molecular ion state of the reaction center. The source or sink in P may be the substituents or R-pi units contiguous to the reaction center.

As explained in the estimation of ionization pK_a, a surrogate electron donor, CH_2^-, replaces the reaction center. The distribution of NBMO charge from this surrogate donor is used to quantify the acceptor potential for the perturber structure, P. The reactivity perturbation is given by:

$$\delta(\Delta EA)_{res} = \rho_{res}(\Delta q)_c$$

where $(\Delta q)_c$ is the fraction loss of NBMO charge from the surrogate reaction center, and the susceptibility, ρ_{res}, of a given reaction center to resonance quantifies the differential "donor" ability of the two states of the reaction center relative to the reference donor CH_2^-.

3.3.4. Results and Discussion

In modeling any property in SPARC, the contributions of the structural components C, S, and R are quantified (parameterized) independently. For example, the strength of a substituent in creating an electrostatic field effect depends only on the substituent regardless of the C, R, or

property of interest. Likewise, the molecular network conductor R is modeled so as to be independent of the identities of S, C, or the property being estimated. Hence, S and R parameters for electron affinity are the same as those for pK_a. The susceptibility of a reaction center to an electrostatic effect quantifies only the differential interaction of the initial state versus the final state with the electrostatic fields. The susceptibility gauges only the reaction $C_{initial}$ -C_{final} and is completely independent of both R or S. For instance, for electron affinity the electrostatic susceptibility reflects the electrostatic perturbations of the LUMO state versus the HOMO state, which once again is totally independent of C or R. Thus, no modifications in pK_a models or extra parameterization for either S or R are needed to calculate electron affinity from pK_a models, other than inferring the electronegativity and susceptibility of electron affinity reaction centers to electrostatic and resonance effects.

Figure 9 shows a sample calculation of EA for 4-chloronitrobenzene. SPARC first computes the resonance and electrostatic perturbations of the appended substituent Cl para to the reaction center NO_2 through the molecular conductor of the benzene ring. Next SPARC computes the perturbation of the appended substituent NO_2 para to the reaction center Cl through the benzene ring. Finally, SPARC sums these perturbations with the base electron affinities for NO_2 and Cl. The susceptibility of the NO_2 and Cl reaction center for resonance and electrostatic effects are shown in Table 12. The substituent parameters and the distance between NO_2 and Cl are obtained as described for pK_a.

Benzene is the progenitor of the other aromatic compounds. The added electron enters the LUMO state, which in the negative ion becomes the SOMO (Singly Occupied Molecular Orbital). Since the LUMO energy is still high, a stable negative ion is not formed in the gas phase. The LUMO energy can be lowered and stabilized either by expansion of the π conjugation or by

introduction of electron-withdrawing substituents. For unsubstituted benzene, naphthalene and anthracene, the electron affinity is seen to increase significantly due to increase in the resonance contributions in going from benzene to anthracene. A positive electron affinity (a stable negative ion) is observed only for anthracene.

Table 12. Reaction center characteristic parameters

Reaction Center	ρ_{ele}	ρ_{res}	χ_{cs}	$(EA)_c$
$-NO_2$	-0.05	2.0	3.28	0.51
-C≡N	-0.03	1.2	----	0.29
-C=O	-0.02	3.2	----	-0.39
in-ring N	-0.16	0.76	----	0.20
$-NR_2$	0.02	-1.01	----	0.17
-OH	0.11	-1.80	----	0.42
-OR	0.06	-0.67	----	0.24
$-CH_3$	0.01	-0.14	2.30	0.01
-F	-0.01	-0.11	4.35	0.16
-Cl	-0.02	-0.09	4.15	0.25
-Br	-0.05	-0.07	3.95	0.26

Introducing π-electron-withdrawing substituents such as cyano, nitro, aldehyde and ketone strongly perturbs the benzene and raises the electron affinity significantly. The LUMO states for these substituents are close to the degenerate π^* benzene LUMOS. This leads to a strong interaction between the LUMO states of these substituents and one of the degenerate LUMO states of benzene, which in turn, lowers and stabilizes the energy of the LUMO states of these molecules. The order of the resonance stabilizing effect of the LUMO state of these substituents is -C=O > $-NO_2$ > -C≡N > in-ring N (aromatic nitrogen, such as pyridine, quinoline, and acridine). For benzene substituted by any electron-donating groups like F, Cl, Br, NR_2, OH and OR, the perturbations of the LUMO state versus the HOMO state are extremely small. Hence, the electron affinity for these molecules

will be close to the electron affinity of benzene (-1.2 eV). The same trend was found for all of these groups when attached to any other aromatic or ethylenic compound. The case for in-ring N is similar. The substitution of N for one of the carbons in a benzene ring decreases the electron density in the π^*-type SOMO state, stabilizing the LUMO state and increasing the electron affinity for pyridine by 0.50 eV. The LUMO energy for pyridine is still relatively high, so that its electron affinity is -0.7 eV and no stable negative ion is formed in the gas phase state. The LUMO energy in n-aromatic molecules can be lowered and stabilized by expansion of the π conjugation. Similar to carbon-aromatic systems, the electron affinity is seen to increase significantly in the order of pyridine, quinoline, and acridine. A positive electron affinity is expected for quinoline and acridine.

$E.A_{(4\text{-chloronitrobenzene})}$	=	$\delta(EA)_{NO2}$	+ $\delta(EA)_{Cl}$
Reaction center		NO_2	Cl
Substituent		Cl	NO_2
Molecular network		benzene	benzene

$$\delta_{Sub}(EA) = EA_{Sub} + \delta_{Sub}\,\Delta EA_{field} + \delta_{Sub}\,\Delta EA_{res}$$

$$\delta(EA)_{Sub} = EA_{Sub} + \frac{\rho_{ele}\ x\ F_s\ x \cos\theta}{(r_{cj} + r_{ij} + r_{is})^2} + \rho_{res}\ x\ \Delta q$$

$$\delta(EA)_{No2} = 0.51 + \frac{-0.05\ x\ 3.622\ x\ 1.0}{(1.3 + 2.0 + 0.65)^2} + (2.0\ x\ 0.236) = 0.962$$

$$\delta(EA)_{cl} = 0.25 + \frac{-0.02\ x\ 7.46\ x\ 1.0}{(1.3 + 2.0 + 1)^2} + (-0.09\ x\ 0.273) = 0.217$$

$EA_{4\text{-chloronitrobenzene}}$ = 0.962 + 0.217 = 1.18

Figure 9. Sample calculation of 4-chloronitrobenzene EA.

Benzoquinone has a high electron affinity (1.9 e.V.). Expansion of the pi system from benzoquinone to naphthoquinone to anthraquinone leads to a decrease in the electron affinity. Both

Hückel MO and STO-3G calculations predict a lower LUMO energy for benzoquinone relative to naphthoquinone. This agrees with experimentally measured electron affinity and the order is opposite to that observed for benzene, naphthalene and anthracene. Introduction of electron withdrawing groups increases the electron affinity of benzoquinone compared to benzene. On the other hand, methyl or alkyl group substitution leads to a decrease in the electron affinity.

Cyclic, unsaturated dicarbonyls such as maleic anhydrides, maleimides and cyclo-pentenedione form a long-lived negative ion in the gas phase. Similar to benzo-, naphthoquinone, the extra electron in these systems enters the LUMO, which is a π^* orbital resulting from a combination of $\pi^*_{c\text{-}c}$ and π^*_{co}. Their LUMO energies are higher, which leads to lower EA for these compounds relative to the quinones. Thus, the electron affinity of maleic anhydride is lower than that for benzoquinone. The electron affinity of the oxy compounds is larger than that of the NH and CH_2 bridged structures. The EA decreases as the electronegativity of the bridging atom decreases, i.e. in the order of OH, NH_2, and CH_2. Substitution of a methyl group destabilizes the LUMO's of the quinones and the anhydrides. The substitution of electron-withdrawing substituents leads to significant increases in electron affinity.

The SPARC-calculated electron affinity for alkene compounds is close to -2.0 eV, the same as the electron affinity for ethylene. Methyl substitution effects on electron affinity are small; in general, electron affinity will decrease by almost 0.04 eV per substituent. 1,2-Dicyanoethylene and tetracyanoethylene (TCNE) are compounds of high electron affinity that are often involved as electron acceptors in charge-transfer complexes. The additional electron in the negative ion enters the LUMO, which is the π^* orbital of ethylene lowered by conjugation with the electron-withdrawing cyano groups.

p-Quinodimethane is expected to have a negative electron affinity similar to benzene. Cyano substitution lowers the LUMO state substantially and increases the electron affinity, hence, tetracyanoquinodimethane (TCNQ) electron affinity is as high as 3 eV. Fluoro substitution in TCNQ will lower the LUMO state even more resulting in higher electron affinity, e.g., tetrafluoro-tetracyanoquinodimethane at 3.2 eV. Pyrrole and furan both have a large negative electron affinity. The methoxy and the amine groups raise the LUMO state thereby lowering the electron affinity for these compounds more than the ethylene electron affinity. Figure 10 shows SPARC-calculated versus observed electron affinities. The RMS deviation was found to be 0.14 eV, which is about equal to the measurement error in charge transfer experiments [17].

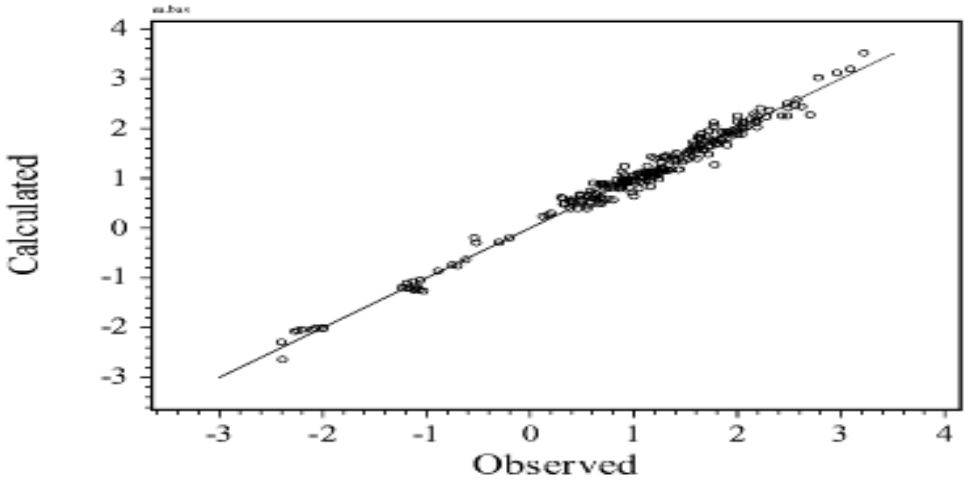


Figure 10. SPARC-calculated versus observed gas phase electron affinities in eV. The RMS deviation for was 0.14 eV.

3.3.5. Conclusion

The SPARC electron affinity models have been tested on all of the reliable (charge transfer equilibria) data available in the literature. The models are fully implemented and are executing in code. These models have been tested to the maximum extent possible given the limited set of direct observations.

3.4. Estimation of Ester Carboxylic Acid Hydrolysis Rate Constants

3.4.1. Introduction

Hydrolysis of organic chemicals is a transformation process in which an compound, RX, reacts with water, forming a new carbon-oxygen bond and the cleaving of a carbon-X bond in the original molecule:

$$R\text{-}X \xrightarrow{H_2O} R\text{-}OH + X^- + H^+$$

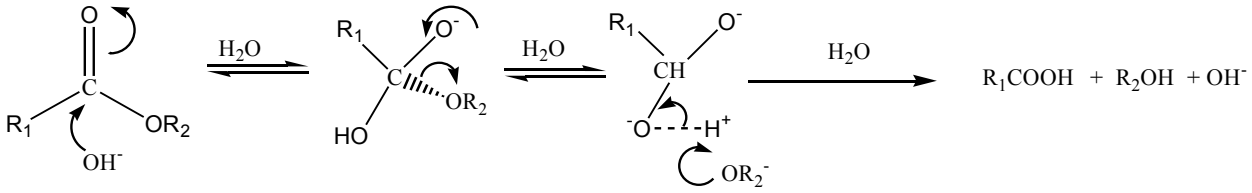
Hydrolysis is one of the most important reactions of organic molecules in aqueous environments, and is a significant environmental fate process for many organic chemicals. It is actually not one reaction but a family of reactions involving compound types as diverse as alkyl halides, carboxylic acid esters, phosphate esters, carbamates, epoxides, nitriles, etc. This study seeks to apply SPARC chemical reactivity models to estimate hydrolysis rate constants for carboxylic acid esters strictly from molecular structure. In the near future, these same models will be used to predict hydrolysis rate constants for other groups such as alkyl halides and phosphate esters.

The general structure for carboxylic acid esters is represented by $R_1C(=O)OR_2$, where R_1 and R_2 are organic substituents. These R substituents can be substituted alkyl chains, phenyl groups or heteroatoms. Carboxylic acid esters are used industrially to make flavors, soaps, herbicides, pesticides, etc. Carboxylic acid esters undergo hydrolysis through three different mechanisms; base, acid and general base-catalyzed ester hydrolysis.

3.4.1.1. Base-Catalyzed Hydrolysis

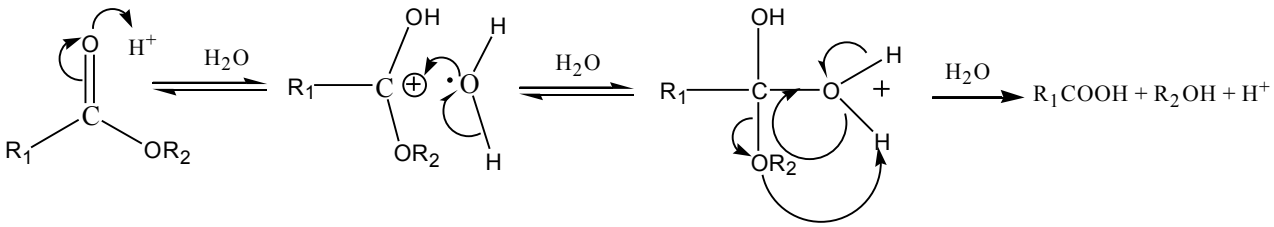
The base-catalyzed or alkaline hydrolysis of esters generally takes place via a B_{AC2} mechanism as shown below. B_{AC2} stands for base-catalyzed, acyl-oxygen fission and bimolecular reaction. It is similar to the SN_2 reaction, and occurs when the hydroxide ion attacks the carbonyl

carbon of an ester to give the carboxylic acid and alcohol. In addition, alkaline hydrolysis of esters may also occur through other mechanisms, such as B_{AC1} (base-catalyzed, acyl-oxygen fission, unimolecular), B_{AL1} (base-catalyzed, alkyl-oxygen fission, unimolecular) and B_{AL2} (base-catalyzed, alkyl-oxygen fission, bimolecular). However, B_{AC2} is the most common mechanism for alkaline hydrolysis of esters and it usually masks all the other plausible mechanisms.



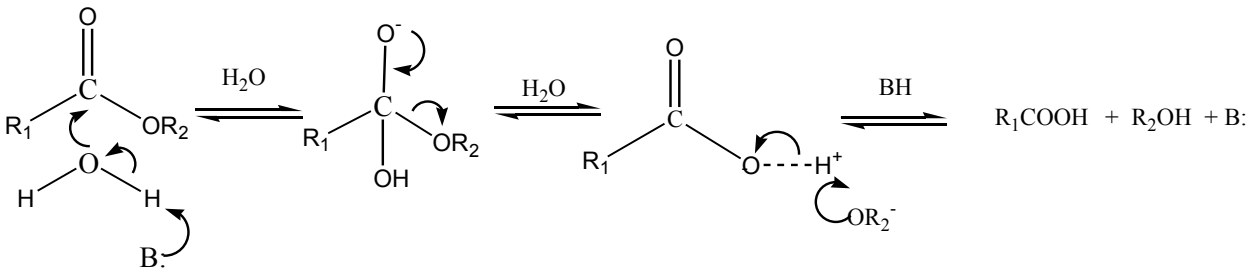
3.4.1.2. Acid-Catalyzed Hydrolysis

The acid catalyzed hydrolysis of esters takes place via A_{AC2} mechanism as shown below. A_{AC2} stands for acid-catalyzed, acyl-oxygen fission and bimolecular reaction. It is also similar to the SN_2 reaction. It occurs when a positive hydrogen ion catalyzes the ester and the water molecule attacks the carbonyl carbon of the ester to give the carboxylic acid and alcohol. In addition, the acid-catalyzed hydrolysis of esters may also take place by other mechanisms, such as A_{AC1} (acid-catalyzed, acyl-oxygen fission, unimolecular), A_{AL1} (acid-catalyzed, alkyl-oxygen fission, unimolecular) and A_{AL2} (acid-catalyzed, alkyl-oxygen fission, bimolecular). However, A_{AC2} is the general mechanism for acid-catalyzed hydrolysis of esters and it usually masks all the other possible mechanisms.



3.4.1.3. General Base-Catalyzed Hydrolysis

The general base-catalyzed hydrolysis of esters takes place via B_{AC2} mechanism as shown below. B_{AC2} stands for base-catalyzed, acyl-oxygen fission, bimolecular reaction. It is too similar to the SN_2 reaction. It occurs when a base (B:) abstracts a hydrogen atom from a water molecule thereby releasing a hydroxide ion that eventually attacks the carbonyl carbon of the ester to yield the carboxylic acid and alcohol. The base, B:, stands for any base, such as ammonia, acetate ion, imidazole and so on. In case of neutral hydrolysis, B: represents the water molecule.



3.4.2. SPARC Modeling Approach

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 (initial) and transition (final) states. SPARC therefore follows the modeling approach described previously for ionization pK_a, in water. For these molecules/chemicals to be modeled, reaction centers with known intrinsic reactivity are identified and the reaction rate constants expressed (ion energy terms) by perturbation theory as

$$\log k = \log k_c + \Delta_p \log k_c$$

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 Δ_p log k_c denotes the perturbation of the log rate constant due to appended structure. These rate constants are expressed in the appropriate second order form inclusive of catalytic effects. A given reaction center may have two or more appendages or perturbing units. For example, a carboxylic acid ester has two appendages and a phosphate ester three. With the exception of steric effects, the perturbations associated with these appendages are modeled independently and simply summed. Each perturbation is factored into the mechanistic components of resonance, steric and electrostatic effects. All that is required for the perturbation calculations are data-fitted susceptibilities of the log k rates to these mechanisms, and knowledge of the electrostatic change at the reaction center (monopole, dipole, etc.) associated with the formation of the transition state to determine the drop off ($1/r^2$) of electrostatic interactions. The latter can be inferred from knowledge of the reaction mechanism or can be determined by an optimized data fit. The log k_c for the reaction center can be either measured directly or can be data-fitted.

3.4.3. Hydrolysis Computational Models

The computational model for hydrolysis of carboxylic acid esters is categorized into three sub models, namely, reference rate, internal perturbation and external perturbation models. The reference rate model calculates the hydrolysis rate constant for the smallest ester compound, which excludes internal perturbation and steric effects. The internal perturbation model calculates the perturbation of the reference hydrolysis rate constant due to the internal perturbation interactions between the reaction center and the substituents. The internal perturbation interactions include steric, resonance and electrostatic effects. Finally, the external perturbation model calculates the solvation contributions to the hydrolysis rate constant due to solute-solvent interactions. The

hydrolysis rate constant contributions from these three models are then added to give the total calculated hydrolysis rate constant as

$$\log k_{Hydrolysis} = \log k_c + \delta_{IP} \log k_c + \delta_{EP} \log k_c$$

where log k_c describes the hydrolysis behavior of the reaction center "reference rate", in this study the hydrolysis rate constant for an unperturbed methyl formate. δ_{IP} log k_c is the change in hydrolysis behavior brought about by the perturber structure. SPARC computes the internal reactivity perturbations, δ_{IP} log k_c, that is then used to "correct" the hydrolysis behavior of the reaction center for the compound in question in terms of the potential "mechanisms" for the interaction of P and C. The last term describes the external perturbation of the solvent with both the initial and the final state. Specifically, δ_{EP} log k_c describes the change in the solvation of the initial state versus the transition state due to H-bond and field stabilization effects of the solvent.

3.4.3.1 Reference Rate Model

The reference rate, log k_c, is the hydrolysis rate constant for the smallest ester compound, that resembles the structure of reaction center C(=O)O. The reference rate is free of any internal perturbation interactions, such as resonance and electrostatic effects. Neither does it show steric effects. However, it is dependent upon the temperature. As the temperature increases, the reference hydrolysis rate increases. SPARC expresses the reference rate, log k_c, as function of the temperature and enthalpic and entropic contributions as

$$\log k_c = A + \log T_k + Ref_1 + Ref_2/T_k$$

where A is the log of the pre-exponential factor, T_k is the temperature in Kelvin, Ref_1 is the entropic contribution to the rate and Ref_2 is the enthalpic contribution. The A_1, Ref_1 and Ref_2 are data-fitted parameters that are assumed to be the same for all molecules, solvents and temperatures.

3.4.3.2 Internal Perturbation Models

Like all chemical reactivity properties addressed in SPARC, molecular structures are broken into functional units called reaction center and the perturber. The reaction center, C, is the smallest subunit that has the potential to hydrolyze. The perturber, P, is the molecular structure appended to the reaction center, C. The perturber structure is assumed to be unchanged in the hydrolysis reaction. The reference hydrolysis rate of the reaction center is adjusted for the molecule in question using the mechanistic perturbation models described below. The perturbation of the log of hydrolysis rate for a molecule of interest is expressed in terms of mechanistic perturbations as :

$$\delta_{IP}\log k_c = \delta_{elec}\log k_c + \delta_{res}\log k_c + \delta_{steric}\log k_c$$

where the $\delta_{ele}\log k_c$ and $\delta_{res}\log k_c$ terms describe the electrostatic and resonance perturbations of the initial and the transition state (activation energy), respectively, caused by molecular substituents (P). Electrostatic interactions are derived from local dipoles or charges in P interacting with charges or dipoles in C. $\delta_{res}\log k_c$ describes the change in the delocalization of π electrons of the two states due to P. This delocalization of π electrons is assumed to be into or out of the reaction center. $\delta_{steric}\log k_c$ describes the change in the steric blockage of solvent access to C of the initial state versus the transition state.

As we described in detail earlier, the modeling of the perturber effects for chemical reactivity relates to the structural representation S-R-C, where S-R is the perturber structure, P, appended to the reaction center, C. S denotes a substituent group that "instigates" the perturbation. For electrostatic effects, S contains (or can induce) electric fields; for resonance, S donates/receives electrons to/from the reaction center. R links the substituent (S) and reaction center (C) and serves as a conductor of the perturbation (i.e., "conducts" resonant π electrons or electric fields).

3.4.3.2.1. Electrostatic Effect Models

The electrostatic effect is a phenomenon of interaction of dipoles or charges of the substituent (S) with the dipoles or charges of the reaction center (C). The types of electrostatic effects that can occur between the substituent (S) and the reaction center (C) are direct field, mesomeric field , R-π, and sigma induction effects.

3.4.3.2.1.1. Direct Field Effect Model

The direct field effect occurs when the electrical dipole/charge of the substituent (S) interacts with the dipole/charge of the reaction center (C) through space. Since the transition states for base and general base-catalyzed hydrolysis of esters are negatively charged, a dipole will increase the hydrolysis rate constant for these reactions. In contrast, the transition state for acid hydrolysis of esters is positively charged and a dipole will decrease the hydrolysis rate constant in this situation. Field effects are resolved into three independent structural component contributions representing the change in dipole field strength of S, a conduction factor of R, and the change in the charge at C as:

$$\delta_{field} \log k_c = \rho_{ele} \sigma_p = \rho_{ele} \sum_S \sigma_{cs} F_S$$

where, $\delta_{ele} \log K_c$ is the direct electrostatic effect due to the interaction between the dipole of the perturber and the reaction center. ρ_{elec} is the susceptibility of the reaction center to electrostatic effects and is presumed to be independent of the perturber. σ_p is the field strength that the perturber exerts on the reaction center, which can be calculated as shown previously for ionization pK_a. The perturber potential, σ_p, is further factored into a field strength parameter, F_S, (describing the magnitude of the dipole field for the substituent) and a conduction descriptor, σ_{cs}, of the intervening

molecular network for the electrostatic interactions. The effective dielectric constant, D_e, for the molecular cavity, any polarization of the anchor atom affected by S, and any unit conversion factors are included in the field strength parameter, F_S. The distances among the various components and orientation angle are calculated as described earlier in this report.

3.4.3.2.1.2. Mesomeric Field Effect Model

A mesomeric field (M_F) is generated when an electron-withdrawing or donating group induces charges on the conductor R. An electron-withdrawing group creates positive charges on the conductor, while an electron-donating group creates negative charges. Since the transition states of base and general base-catalyzed hydrolyses are negatively charged, the electron-withdrawing groups will increase the hydrolysis rate constant because the induced positive charges on the conductor will stabilize the negative charges of the reaction center. On the other hand, induced negative charges will have an opposite effect. The M_F effect due to interaction between either the electron withdrawing or donating group with the reaction center is given as.

$$\delta_{MF} \log k_c = \rho_{ele} M_F \sum q_{ik} / r_{kc}$$

where $\delta_{MF} \log k_c$ is the mesomeric field effect due to the interaction between the substituent induced charges on the molecular conductor and the reaction center. ρ_{ele} is the susceptibility of the reaction center, which is assumed to be independent of the substituent. M_F is the mesomeric field constant characteristic of the substituent. It describes the ability of the substituent to induce charges. Substituent M_F values have been calculated for ionization pK_a and were shown in Table 3. q_{ik} is the charge induced at each atom k on R, and is calculated using PMO theory. r_{kc} is the through-cavity distance between the induced charge on atom k and the reaction center.

3.4.3.2.1.3. Sigma Induction Effect Model

Sigma induction occurs due to the difference in electronegativity between the reaction center and the substituents. For base and general base-catalyzed hydrolyses, the reaction center has a large electronegativity and methyl substituents, for example, will move charge or electrons into the reaction center and decrease the hydrolysis rate constant. The acid-catalyst hydrolysis reaction center, on the other hand, is less electronegative and the induced perturbations are always quite small. The sigma induction is a short range effect. We have calculated these effects up to two atoms from the reaction center and consider effects further a way to be negligible. The sigma induction due to an electronegativity difference between the reaction center and the substituent is calculated using the following equation.

$$\delta_{Sigma} \log k_c = \rho_{elec} \sum (\chi_s - \chi_c)\, NB$$

where ρ_{elec} is the susceptibility of the reaction center to electrostatic. χ_c and χ_s are the electronegativity of the reaction center and the substituents, respectively. NB is data-fitted parameter that depends on number of the substituents that are bonded directly to the reaction center, C. Both NB and χ_s values are the same as those used for ionization pK_a calculations.

3.4.3.2.1.4. R_π Effect Model

The R_π effect is similar to sigma induction, except it involves π-electrons instead of σ-electrons. The magnitude of the reactivity perturbation, $\delta_\pi \log k_c$, depends upon the difference in the electronegativity of the atom of the π group and that of the reaction center to which it is attached. Since the differential induction capability is highly correlated with ρ_{ele}, SPARC uses a

simple model to quantify the effect, requiring a minimum computation and only one extra parameter:

$$\delta_{\pi} \log k_c = \rho_{ele} \sigma_{\pi}$$

where σ_π is a data-fitted parameter. The reaction center is classified as a C+ group (at the carbon of the carbonyl) that withdraws electrons and a C- group (at the acyl oxygen) that donates electrons to a reference point. If the π-system is attached to the C+ group, the R_π effect contributes negatively or lowers the hydrolysis rate constant. In contrast, if the π-system is attached to the C- group, the R_π increases the hydrolysis rate constant.

3.4.3.2.2. Resonance Effect Model

Resonance is a phenomenon of π-electrons moving in or out of the reaction center. Resonance stabilization energy in SPARC is a differential quantity, related directly to the extent of electron delocalization in the initial state versus the transition state of the reaction center. The source or sink in P may be substituents or R-π units contiguous to the reaction center. Substituents that withdraw electrons from a reference point, e.g., CH_2^-, are designated S+ and those that donate electrons are designated S-. The R-π units withdraw or donate electrons or may serve as "conductors" of π-electrons between resonance units. Reaction centers are likewise classified as C+ (carbonyl carbon) and C- (acyl oxygen) denoting the withdrawing and donating of electrons, respectively. The distribution of NBMO charge from a surrogate donor, CH_2^-, is used to quantify the acceptor potential for the perturber structure, P. The resonance perturbation is given by:

$$\delta_{res} \log k_c = \rho_{res} \Delta q_c$$

where ρ_{res} is the susceptibility of the reaction center to resonance interactions. That is, quantifies "donor" ability of the two states of C relative to CH_2^- . Δq_c is the fraction loss of NBMO

(nonbonding molecular orbital) charge from the surrogate reaction center calculated based on PMO theory. The reaction center has two different ρ_{res}s one is for the oxygen of the ester when it is attached to π-networks and the other is for the carbonyl when it is attached to π-networks. Resonance plays two important roles in ester hydrolysis. The major impact is that of resonance stabilization of the leaving group. Thus, π-networks attached to the oxygen of the ester reaction center have a pronounced effect and greatly increase the hydrolysis rate. π-networks attached to the carbonyl tends to destabilize the leaving group and reduce the hydrolysis rate.

3.4.3.2.3. Steric Effect Model

The normal trend for steric effect is that as the bulkiness of a substituent increase, the steric effect increases. Thus, the steric effect always decreases the hydrolysis rate constant. Comparing the steric effect on hydrolysis rate constant in various solvents, we observe the trend of lesser steric effect in pure water than in other mixed solvents. The reason for this trend is that pure water solvates the substituents more and aligns the structure of the esters in suitable position for the attacking hydroxide ion or water molecule. On the other hand, the mixed aqueous solvents only partially solvate the substituents and thus deform the structure of esters, creating a hindrance to attack from the hydroxide ion or water molecule. Therefore, the reaction does not proceed at a normal rate and the hydrolysis rate constant decreases. Steric effects will include blockage of reactant access and strain in achieving the transition state. SPARC expresses the steric contributions as:

$$\delta_{steric} \log k_c = \frac{\rho_{steric} (V_s + V_{ex} - V_{thresh})}{T_k D_e}$$

where V_s is the sum of the appendage sizes, V_{thresh} is a threshold size for onset of steric effects, and V_{ex} is an excluded (cavity) volume between pairs of appendages (zero for formates, 2 vex for acetates, 3 vex for trialkyl phosphates). ρ_{steric} is the steric susceptibility, D_e is the dielectric constant for the solvent and T_k is the reaction temperature.

3.4.3.3. External Perturbation Models

3.4.3.3.1. Solvation Effect Model

Solvation effects for ester hydrolysis include both hydrogen bonding and field stabilization effects. Hydrogen bonding gauges the hydrogen acceptor effect (alpha) and hydrogen donor effect (beta) of the ester, while the field stabilization interaction describes the effect of dielectric constant of the solvent on the hydrolysis rate constant.

3.4.3.3.1.1. Hydrogen Bonding

Hydrogen bonding 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 is resolved into a proton-donating site, α, and proton-accepting site, β, which in the SPARC models are presumed to be independently quantifiable. If the transition state is more solvated or stabilized by the hydrogen bonding than the initial state, the hydrolyses rate constant increases. The negatively charged transition states of base and general base catalyzed hydrolysis are strongly solvated or stabilized by solvent alphas, while the betas play a minor rule. Thus, one might conclude that strong alpha solvent should increase the base-catalyzed hydrolysis rate constant. However, the strong alpha solvents not only solvate the transition states, but also solvate the attacking hydroxide ion. This tends to stabilize the initial state more than the transition state as shown in the Figure 11. Therefore, strong alpha solvents actually

tend to decrease the base-catalyzed hydrolysis rate constants. On the other hand, the strong beta solvents interact with the alpha sites freeing-up the hydroxide ions to react with the esters and tending to increase the base-catalyzed hydrolysis rate constant. For acid-catalyzed hydrolysis, both the alpha and beta sites stabilize the initial state more than the transition state. Therefore, hydrogen bonding decreases the acid-catalyzed hydrolysis rate constant. Furthermore, the alpha and beta impacts on the hydrolysis rate constant in various mixed solvents depend on the relative amount of alpha and beta sites available in those solvents. Water has an almost equal strength of alpha and beta, whereas mixed solvents generally have weaker alpha and stronger beta. Consequently, the alpha contribution from pure water to hydrolysis rate constant in base and general base catalyzed hydrolysis should be lower or more negative than from the mixed solvents. The beta contribution, on the other hand, should be higher in mixed solvents than in pure water. SPARC expresses alpha and beta H-bond contribution as:

$$alpha = \frac{\rho_A \alpha (1 - F_v Vol)}{T_k} \qquad beta = \frac{\rho_B \beta}{T_k}$$

where alpha (beta) is the hydrogen accepter (donor) effect of the solute ester. α (β) is the hydrogen donating (accepting) value of the solvent. ρ_A, and ρ_B are the susceptibility for α and β of the solvent, respectively. These are data-fitted parameters. F_v is another data-fitted parameter for the volume of alpha of the solvent. Vol is the volume of the solvent and T_k is the temperature in Kelvin. Both α and β of the solvent are calculated as pseudo pK_a's, with the electrostatic component treated as a dipole transition

3.4.3.3.1.2. Field Stabilization Effect

The field stabilization effect is given as

$$\delta_{FS} \log k_c = \frac{\rho_{FS}}{T_k (D_e + Damp)}$$

where ρ_{FS} is the susceptibility of the solvent to solvation due to the dielectric constant of the solvent and is a data-fitted parameter. D_e is the dielectric constant of the solvent and Damp is a damping adjusting factor. The dielectric constant impacts both the solvation of initial and transition states of the reactants in the hydrolysis reaction. Hence, dielectricity of the solvent solvates or stabilizes the initial state more than the transition state. Thus, the field stabilization effect decreases the hydrolysis rate constant. Comparing the dielectricity effect on hydrolysis rate constant in different mixed solvents, we observe that the dielectricity or field stabilization effect reduces the rate constant less in pure water than in mixed organic aqueous solvents. The reason for this phenomenon is that the high dielectricity of pure water stabilizes the transition state more than mixed aqueous organic solvents which have less overall dielectricity.

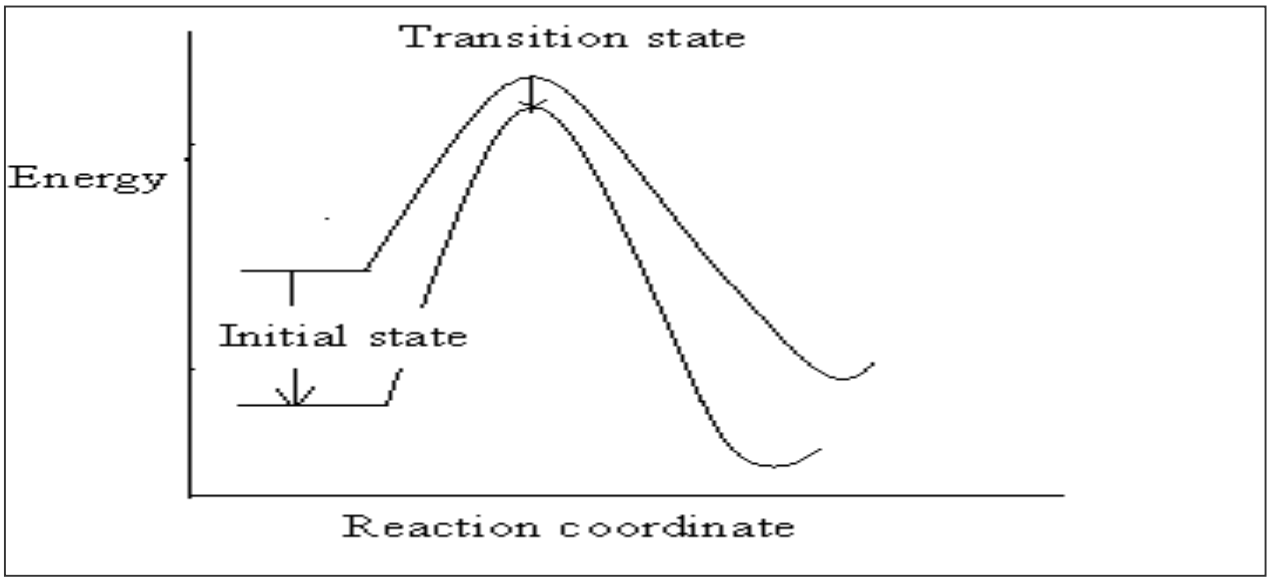


Figure 11. The effect of alpha sites on the initial and transition states. Alpha sites tremendously solvate the hydroxide ion and stabilize the initial state more than the transition state. As a result, alpha sites decrease the hydrolysis rate constant.

3.4.3.3. Temperature Effect

Steric, hydrogen bonding and field stabilization effects decrease with increasing temperature, and the rate of hydrolysis of organic compounds increases with temperature. The quantitative relationship between the hydrolysis rate constant and temperature is frequently expressed by the Arrhenius equation as

$$k = A\, e^{-\left(\frac{E_a}{RT}\right)}$$

where k is the hydrolysis rate constant, A is the pre-exponential factor or the frequency factor, E_a is the activation energy (the minimum energy required to form a product from the reactants), R is the gas constant and T is the absolute temperature. As shown previously, temperature dependence was also incorporated in to the field stabilization, alpha/beta and steric effect relationships. For example, recall that, as the temperature increases, the steric effect decreases. The reason for this is that as the temperature increases the reactants tend to mimic gas phase structure, producing minimal or null steric effect for hydrolysis of esters.

3.4.3.4. Results and Discussions

For any property estimated by SPARC, the contributions of the structural components C (reaction center), S (substituent), and R (molecular conductor) are quantified independently. For example, the strength of a substituent (S) in creating an electrostatic field effect is assumed to depend only on the substituent regardless of the C, R, or property of interest. Likewise, R is modeled so as to be independent of the identities of S, C, or the property being estimated. Hence, S and R parameters for hydrolysis rate are the same as those for ionization pK_a in water. The susceptibility of a C to an electrostatic effect quantifies only the differential interaction of the initial

state versus the final state with the electrostatic fields. The susceptibility gauges only the reaction $C_{initial} - C_{transition\ state}$ and is completely independent of both R or S. Thus, no modifications in any of the pK_a models, or extra parameterization for either substituent (S) or the molecular conductor (R) are needed to calculate hydrolysis rate constant using the ionization pK_a models in water, other than inferring the electronegativity and susceptibilities of the carboxylic acid ester hydrolysis rate constant to electrostatic, resonance, steric and solvation effects. A sample calculation for the acid catalyzed hydrolysis rate constant for p-nitrophenyl acetate in water at 25^o C is shown in Figure 12. Figures 13-15 show SPARC-calculated versus observed values for hydrolysis rate constants of esters undergoing base, acid and general base catalyzed hydrolysis, respectively in six different solvents and various temperatures, respectively. These sets represent 321, 416 and 50 unique esters in base, acid and general base catalyzed hydrolysis of esters respectively. Because several of the esters were measured under different conditions (solvents, temperatures, etc) there were 653, 667 and 150 base, acid and general base catalyzed calculations performed. The RMS deviation of the SPARC-calculated versus measured carboxylic ester hydrolysis rate constant values for these three hydrolysis mechanisms were 0.37, 0.37 and 0.39 log units, respectively.

Reaction Center	ρ_{ele}	ρ_{res}	χ_c	Const	A	Ref_1	Ref_2	ρ_{FS}	ρ_{Steric}	ρ_α	ρ_β	Damp
Acid Hydrolysis	-0.87	-0.4, 1.11	1.8	1.7	3.36	3.886	-3070.8	$-0.051 * 10^7$	-310935	- 1821	944.6	32.6
Base Hydrolysis	4.5	-0.8, 3.9	2.55	1.7	3.36	-3.522	-1443.4	$0.18 * 10^7$	-328642	4205	-2119 .2	32.6
GBase Hydrolysis	5.2	-0.6, 2.6	2.65	1.7	3.36	-5.249	-3479.7	$-0.15 * 10^7$	-114969	-373	5598.3	32.6

$\log k_c$ = ***3.36*** + ***log (298.15)*** + ***3.886*** + ***(-3070.8 / 298.15)*** = ***- 0.59***

ρ_{elec} χ_c NO2 $\cos\theta_{cs}$ NBMO Charges **Constant** ρ_{res}

$$\delta_{IP}\log k_c = -0.87\left[3.45\frac{(2.3-1.8)}{(1.3+1)^2} + \frac{7.46\times 1}{(1.3+2+1)^3} + 0.008 + \left(2\times\frac{0.078}{(1.73+1+1.3)^2} + \frac{0.078}{(1.3+1)^2} - \frac{3\times 0.078}{(1.3+2+1)^2}\right)2.515\times -1.7\right] 0.28\times 1.11 = -0.075$$

NB r_{ic} χ_s r_{is} σ_π r_{13} r_{14} MF_{NO2}

Sigma Field R_π Mesomeric Resonance

ρ_{FS} ρ_{Steric} V's ρ_α **α** Vol ρ_β **β**

$$\delta_{EP}\log k_c = \frac{-0.051\times 10^7}{298.15\,(78.54+32.6)} + \frac{-310935\times 0.0687}{298\,(78.54+32.6)} + \frac{-1821\times 0.384(1-0\times 0.07)}{298} + \frac{944.6\times 0.382}{298} = -3.333$$

$D_e(H_2O)$ Damp

Field Stabilization Steric Hydrogen bonding

Log $k_{hydrolysis}$ = -0.59 + (-0.075) + (-3.33) = 3.995 (Observed is 3.9)

Figures 12. Sample calculations of hydrolysis rate constant acid catalyzed media for p-nitrophenyl acetate in water at 25° C. Only the reaction center parameters are trained on hydrolysis rate constant (in bold and they are showing at the top of the Figure). The substituent, molecular conductor parameters and distances between the various components (indicated by lines) are the same as of ionization pK_a and they are shown in Table 2 and 3, see text. The α and β are for the solvent, water, and they are calibrated using SPARC physical process models. See next section.

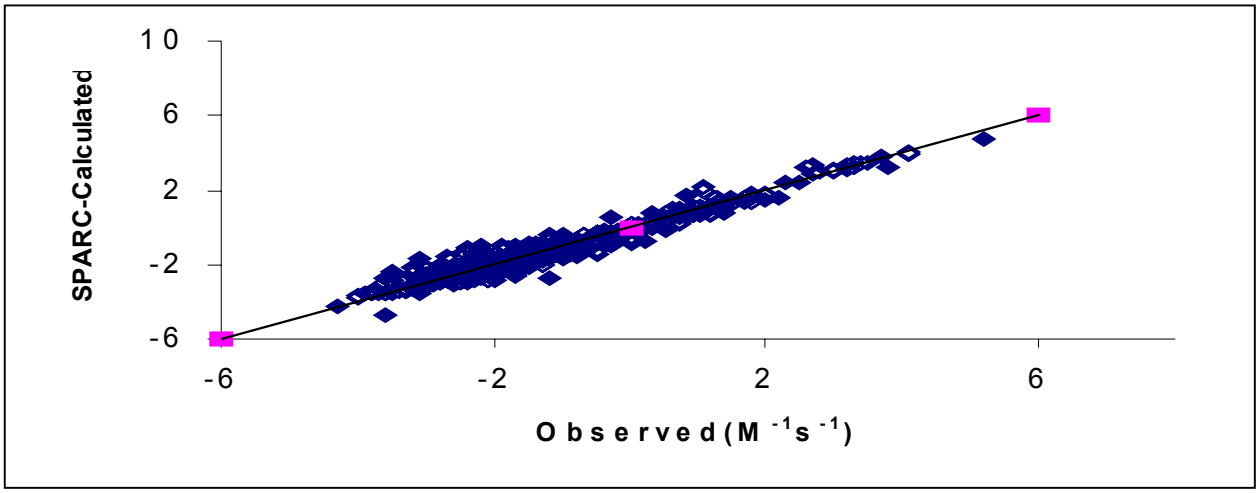


Figure 13. SPARC-calculated versus observed log hydrolysis rate constants for alkaline hydrolysis in six different solvents, respectively. The RMS deviation error is 0.37 and R^2 is 0.96.

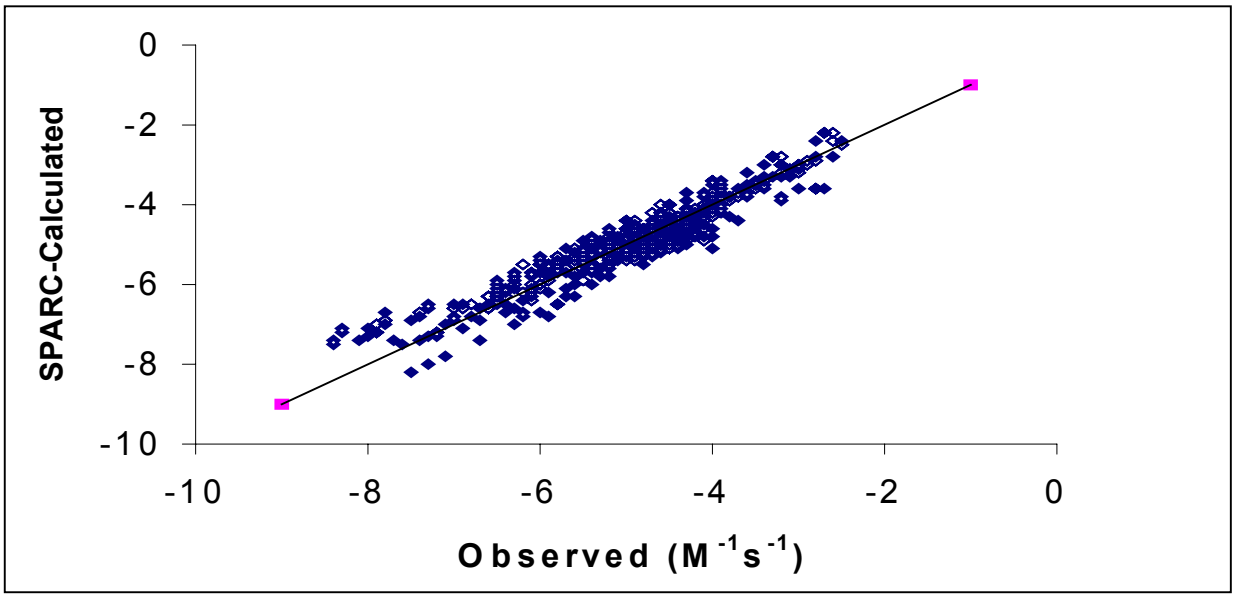


Figure 14. SPARC-calculated versus observed log hydrolysis rate constants for acid hydrolysis in six different solvents and at different temperatures. The RMS deviation is 0.37 and R^2 is 0.97.

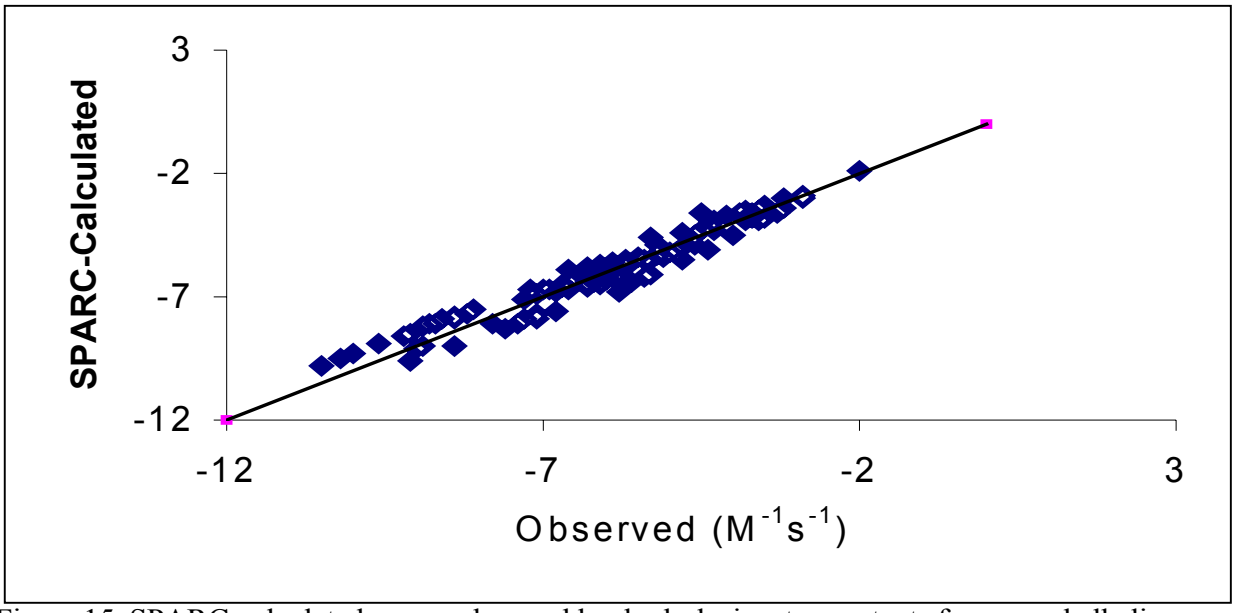


Figure 15. SPARC-calculated versus observed log hydrolysis rate constants for general alkaline base hydrolysis in six different solvents and at different temperatures. The RMS deviation is 0.39 and R^2 is 0.97**.**

Table. 13 Statistical Parameters of SPARC Hydrolysis Rate Constant ($M^{-1}s^{-1}$)

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

3.4.5. Conclusion

SPARC's reactivity models, used to calculate both ionization pKa and electron affinity, have been successfully extended to calculate hydrolysis rate constants for carboxylic acid esters. These models have been tested to the maximum extent possible as function of temperature and for single and mixed solvent systems on all the reliable data available in the literature. Further extension of these reactivity models is currently under development to calculate hydrolysis rate constants for phosphate ester compounds.

4. PHYSICAL PROPERTIES

4.1. Estimation of Physical Properties

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 (solute and solvent(s)) and reaction conditions of interest (temperature, pressure, etc.). SPARC solvation models are described in this section. Results of these solvation models in estimating solute activities and 'activity based' properties (solubilities, vapor pressures, 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 and chemical processes, 'environmental' conditions, and molecular structures) and 'accessibility' for the target audience.

Intermolecular interactions are expressed as a summation over all the interaction forces between molecules (i.e., dispersion, induction, dipole-dipole and hydrogen bonding). Each of these interaction energies is expressed in terms of a limited set of molecular-level descriptors (density-based volume, molecular polarizability, molecular dipole, and hydrogen bonding parameters) that, in turn, are calculated strictly from molecular structure.

4.2. Physical Properties Computational Approach

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

$$\Delta G_{process} = \Delta G_{Interaction} + \Delta G_{Other}$$

where $\Delta G_{Interaction}$ describes the change in the intermolecular interactions accompanying the process in question. For example, in liquid to gas vaporization, $\Delta G_{Interaction}$ describes the difference in the intermolecular interactions in the gaseous versus the liquid phase. The intermolecular interaction forces between the molecules are assumed to be additive. The ΔG_{Other} lumps all non-interaction components, such as excess entropy changes associated with mixing or expansion, and changes in internal (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. The intermolecular interactions in the liquid phase are expressed as a summation over all the mechanistic components:

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

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.

4.3. SPARC Molecular Descriptors

The computational approach for molecular-level descriptors is constitutive with the molecule in question being broken at each essential single bond and the property of interest being expressed as a linear combination of fragment contributions as

$$\chi^o(molecule) = \sum_i (\chi_j^o - A_i)$$

where χ_j^o are intrinsic fragment contributions (which in most cases are tabulated in SPARC databases) and A_i are adjustments relating to steric or electrometric perturbations from contiguous structural elements for the molecule in question. In some instances, the χ^o (molecule) is further adjusted for a specific process model or medium involved. Both χ_j^o and A_i are empirically trained, either on direct measurements of the descriptor in question (e.g., liquid density based molecular volume) or on a directly related property (e.g., index of refraction, which can be related to polarizability) for which large reliable data sets exist. This partition of molecular descriptors into intrinsic fragment contributions enables one to construct, for any given molecule-of-interest, essentially any molecular array of appended units, and thereby to estimate the descriptors of interest for any molecular structure.

4.3.1. Average Molecular Polarizability

Molecules are composed of positively charged nuclei and negatively charged electrons. When molecules are subjected to an electric field, the electrons are attracted toward the positive plate and the positive nuclei are displaced from their ordinary position toward the negative plate. The result is an electric distortion or polarization of the molecules producing electric dipoles. As

mentioned previously, molecular structures are broken at each essential single bond with known intrinsic atomic polarizability. The molecular polarizability of any molecule-of-interest is calculated as the linear combination of all the fragment polarizabilities, which in turn are estimated from intrinsic atomic polarizability contributions, $\chi_{j,}$. The polarizability of fragment i is expressed as

$$\overline{\alpha}_i = \frac{1}{N_i}[\sum_j \chi_j]^2$$

where the summation is over all the atoms in fragment i, χ_j is the intrinsic atomic hybrid polarizability contribution, and N_i is the number of electrons in fragment i. The χ_j are empirically determined from measured polarizabilities and stored in the SPARC database (with the exception of hydrogen, which is calculated from the measured polarizability of H_2). The average molecular polarizability, α^o, is calculated as the sum over all i fragment:

$$\alpha^o = \sum_i (\alpha_i - A_i)$$

where α_i is the polarizability of fragment i and A_i are adjustments for the molecule in question. The only adjustment, A_i, currently implemented in SPARC is a 10% reduction in α_i for hydrocarbon fragments with an attached polar group or atom. The partition of polarizability into atomic contributions enables estimates to be made of molecular polarizabilities for any given molecular structure. The molecular polarizability can be calculated to better than 1% of measured values for a wide range of organic molecules.

4.3.1.1. Refractive Index

Many physical properties depend upon polarizability; the most familiar is the refraction of light. The passage of a light wave is accompanied by an oscillating electric field at right angles to

the direction of the light propagation producing a corresponding oscillating dipole in nearby molecules. This interaction reduces the velocity of propagation of the light wave, which is to say that the refractive index, n, of the material medium is greater than 1. Index of refraction is thus a good way to check the polarizability density for a molecule. The molecular polarizability and volume can be related to the index of refraction using the Lorentz-Lorenz equation. For our units of cm^3/mole for volume and $Å^3$/molecule for polarizability, the Lorentz-Lorenz equation can be written as

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

where n is the index of refraction, P is the molecular polarizability and V is the molecular volume.

The refractive index calculator was trained on 325 non-polar and polar organic compounds at 25° C then validated on 578 organic compounds at 25° C as shown in Figure 16. The statistical performance for the SPARC refractive index calculator is shown in Table 14. See reference 23 for sample hand calculations.

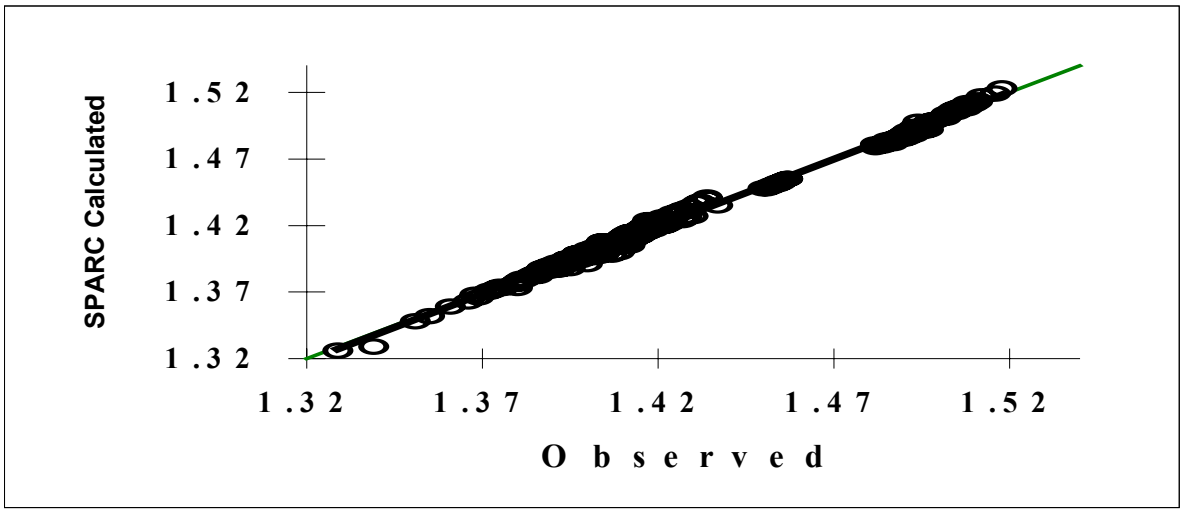


Figure 16. SPARC-calculated versus observed refractive index at 25° C. The RMS (Root Mean Square) deviation error was 0.007 and R^2 was 0.997

Table 14. SPARC Physical/Chemical Properties Statistical Parameters

Property	Units	Total # Molecule	RMS	R^2	Reaction Conditions Temp/Solvent
Refractive Index		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	Kcal/mole	1263	0.301	0.993	25, Boiling Point
Diffusion Coefficient in Air	cm^2/s	108	0.003	0.994	25
Activity Coefficient	log MF[3]	491	0.064	0.998	25, 41 solvents
Solubility	log MF	647	0.40	0.987	25, 21 solvents
Distribution Coefficient		623	0.43	0.983	25 Octanol, Toluene CCl_4, Benzene, Cyclohexane, Ethyl Ether
Henry's Constant	N/A	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	N/A	125	0.095	0.992	25, Water/Methanol
Gas pK_a Non-aqueous pK_a 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 Hydration Constant	Kcal/1.36 Kcal/1.36	36 27	0.3 0.43	0.950 0.744	25 , Water 25, water
$E_{½}$ Chemical Reduction	e.V	352	0.18	0.95	25, Water, Alcohols, DMF[1] Aceteonitrile, DMSO[1]

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

2. GC retention times in SE-30 and PEG-20M liquid stationary liquid phase is not included in this report.

3. MF: mole fraction

4.3.2. Molecular Volume

The "zero order" liquid density-based molecular volume is expressed as

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

where V_i^{frag} is the volume of the i^{th} 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}s are determined empirically from measured volumes and then stored in the SPARC database. This zero order volume at 25° C is further adjusted for shrinkage resulting from dipole-dipole and H-bonding interactions by the following equation:

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

where D_i is the bond dipole of the ith fragment, and α_i, β_i is the H-bonding parameters of potential proton donor and proton acceptor sites within the molecule, respectively. The product $\alpha_i * \beta_i$ is the largest H-bonding interaction contribution in the molecule. $A_{dipole-dipole}$ and A_{H-bond} are adjustment constants due to dipole-dipole and H-bonding, respectively. The final molecular volume at temperature T is then expressed as a polynomial expansion in (T-25) corrected for H-bonding (HB), dipole density (D^d) and polarizability density (P^d) interactions as

$$V_T = V_{25}[1 + f(P^d, D^d, HB) \sum_n a_n (T - 25)^n]$$

where a_n are trainable parameters. The SPARC molecular volume calculator was tested on more than 1440 compounds at 25° C and the RMS deviation error between calculated and measured values was 1.97 volume units as shown in Figure 17.

SPARC calculates the density at 25° C directly from the molecular volume calculator result using the simple equation Density = Molecular Weight/Volume. The accuracy of the SPARC density calculation depends purely on the accuracy of the calculated molecular volume.

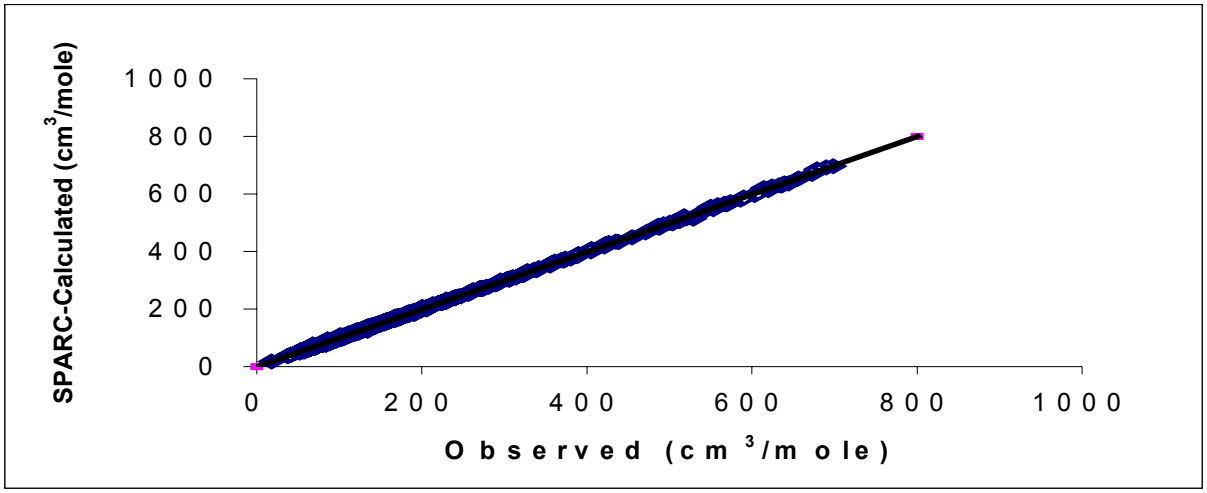


Figure 17. SPARC-calculated vs. observed-liquid density based volume at 25° C for 1440 organic molecules. The RMS deviation error was 1.97 cm^3 $mole^{-1}$ and R^2 was 0.999.

4.3.3. Microscopic Bond Dipole

For induction and dipole-dipole interactions, the effective microscopic dipole μ is computed from the bond dipole contributions of individual substituents μ_i^o. μ_i^o describes the effective substituent bond dipole strength when it is attached to either a methyl, ethylenic or aromatic group. The effective substituent bond dipoles, μ_i^o, used in SPARC, are derived from tabulated substituent dipoles. Specifically, for each substituent, S, an S-methyl, S-ethylenic and S-aromatic dipole value is stored in the SPARC database based on the dipole moments reported by McClellan [46]. For any given molecule, a simple algebraic summation over all i fragments is employed to calculate the effective microscopic dipole for the molecule as

$$\mu_i = \sum_i S_i \mu_i^o$$

S_i is a reduction factor for steric blockage due to an appended molecular structure to substituent i given as

$$S_i = \rho_i \upsilon_i$$

where ν_i is the solid angle occluded by the appended molecular structure. The value of ν_i depends on the size and number of the groups that are appended to i. ρ_i is the steric 'susceptibility' of each substituent dipole-in-question. The reductions factors, S_i, are usually small (< 5% for most molecules) and will be ignored in this report.

For cases where multiple substituents are 'clustered', i.e., are attached to one of the following: (1) the same aromatic ring or ethylenic unit or; (2) the same (or adjacent) aliphatic atom(s), a 'local' vector sum is invoked, wherein each individual dipole is reduced by a fraction of the resultant vector field of the other dipoles in the cluster. Details of the vector models will not be given here except for two cases where the corrections are large. When the atom to which multiple substituents are bound is also an intrinsic component of the individual dipoles (e.g., halogen atoms attached to the same carbon), 'vectorial' reduction is substantial. For the chlorinated methane series CCl_x (x=1,2...4), the reduced individual dipole components are 1.8, 0.9, 0.4, 0.05 Debye, respectively.

4.3.4. Hydrogen Bonding

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 interaction ($\Delta G_{H\text{-}Bond}$) is resolved into a proton donating site α and proton accepting site β, which in our models are presumed to be independently quantifiable. The α and β for both the solute and solvent are calculated as pseudo pK_a's, with the electrostatic component treated as a dipole transition. Details of the pK_a computational methods were given earlier in this report; only a brief description for the H-bond calculation is given here. Molecular structures are broken into functional units called reaction centers and perturbers. The potential sites for reactions-of-interest are designated as the reaction

centers, C. These are the smallest subunits that have potential to be a proton donating (accepting) site. The perturbers, P, are the molecular structures appended to the reaction centers and are assumed to be unchanged in the reaction. The α_c and β_c of the reaction center, C, are adjusted for the molecule in question by

$$\alpha = \alpha_c + \delta\alpha_p \qquad \beta = \beta_c + \delta\beta_p$$

where α_c and β_c denote the intrinsic behavior of the reaction center. Both α_c and β_c of the reaction center are inferred indirectly from physical process measurements. $\delta\alpha_p$ and $\delta\beta_p$ denote changes in reactivity effected by the perturber structure, P. SPARC computes both $\delta\alpha_p$ and $\delta\beta_p$, which are then used to "correct" the donating (accepting) site behaviors of each reaction center for the molecule of interest in terms of potential "mechanisms", by

$$\delta\alpha_p = \delta\alpha_{ele} + \delta\alpha_{res} \qquad \delta\beta_p = \delta\beta_{ele} + \delta\beta_{res}$$

where $\delta\alpha_{ele}$ and $\delta\alpha_{res}$ ($\delta\beta_{ele}$ and $\delta\beta_{res}$) describe the differential electrostatic and resonance effects of P on the initial state versus the final state for the proton donating (accepting) sites, respectively.

The H-bond calculations are exactly as described earlier in the pK_a models except for the electrostatic multipole terms. Whereas ionization pK_a is a monopole process resulting in a net change in charge in the reaction center; hydrogen bond formation on the other hand is dipolar, resulting in the change in bond dipole(s) in the reaction center. As was the case in the ionization pK_a calculation, electrostatic interaction terms up through substituent dipole are retained, but the radial and angular dependence of electrostatic conduction differs as previously described for the multipole model. The net result is a faster die-off of field effects with intermolecular distance and increased sensitivity to dipole-dipole alignment for dipolar substituents. Also, because H-bonding does not involve forming or breaking of covalent bonds, differential resonance stabilization of the reaction center is small, which means resonance effects are small.

4.4. SPARC Interaction Models

The differences in strength of the intermolecular interaction forces are reflected in the physical property behavior of the compounds. For example, the boiling points for ethane, chloromethane and ethanol are -88, -24 and 65.4 C, respectively. Dispersion interactions are present in all 3 molecules, but dipole-dipole and induction interactions elevate the boiling points of the chloromethane and ethanol. In addition, H-bond interactions exist only for ethanol and raise its boiling point even higher. In general, compounds whose molecules interact through H-bonding have higher boiling points than molecules of the same molecular weight, volume and dipole moment where hydrogen bonding is not present. SPARC's interaction models are built on a limited set of molecular-level descriptors (volume, polarizability, molecular dipole and hydrogen bonding parameters) as described earlier in this report. These interaction models are for dispersion, induction, dipole-dipole and H-bonding. Dispersion interactions are present for all molecules, including 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 donator strength and β represents the proton acceptor strength.

4.4.1. Dispersion Interactions

Dispersion interactions occur between all molecules as a result of very rapidly varying dipoles formed between nuclei and electrons at zero-point motion of the molecules, acting upon the polarizability of other molecules to produce an induced dipole in the phase. The free energy associated with these self-interactions is expressed as

$$\Delta G_{ii}(\text{Dispersion}) = \rho_{disp}(P_i^d)^2 V_i \qquad \text{where} \qquad P_i^{\,d} = \frac{\bar{\alpha}_i + A_{disp}}{V_i}$$

P_i^d is the effective polarizability density of molecule i; ρ_{disp} is the susceptibility to dispersion; V_i and $\bar{\alpha}_i$ are the molar volume and average molecular polarizability, respectively. 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 (i.e., for branched structures or rings small enough to prohibit intra penetration of the solvent). Branched (ternary or quaternary) atoms in an alkane structure will lose a small part of their intrinsic molecular polarizability depending on the size and number of appended groups, and the proximity of other branched carbons. Similarly, carbons in rings may lose their intrinsic polarizability contributions depending on ring size and the presence of a ring appendage.

4.4.2. Induction Interactions

Induction or dipole-induced dipole interactions occur between molecules where one or both contain a permanent dipole. The dipole moment of a polar molecule has the effect of polarizing a second molecule. This induced dipole moment can then interact with the dipole moment of the first molecule. The magnitude of this effect depends on both the strength of the dipole moment of the first molecule and the polarizability of the second one. For self interactions, the free energy change due induction effects may given as

$$\Delta G_{ii}(Induction) = \rho_{ind} P_i^{d'} D_i^d V_i$$

where ρ_{ind} is the 'susceptibility' to induction and $P_i^{d'}$, D_i^d and V_i are polarizability density (adjusted for induction), dipole density and molar volume of the molecule-in-question, respectively.

$$D_i^d = \frac{\mu_i}{V_i} \qquad \text{and} \qquad P_i^{d'} = \frac{\bar{\alpha}_i + A_{Ind}}{V_i}$$

where μ_i is the effective microscopic dipole described previously and A_{ind} is a polarizability adjustment for induction. Induction describes molecular polarization effected 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 nonconjugated array of atoms. SPARC adjusts the molecular polarizability algorithmically, utilizing electron withdrawing/releasing substituent parameters derived from pK_a models; these models will not be presented here, but a few simple rules will be given that capture all major adjustments. No significant adjustments need to be made for unsubstituted systems. A polar substituent attached to a π unit (aromatic ring or ethylene unit) reduces its intrinsic 'induction' polarizability by ~25 percent; this results in a polarizability reduction of ~1.3 cubic angstroms for a singly substituted ethylene and ~2.75 cubic angstroms for a substituted benzene or condensed ring hydrocarbon. Heteroatoms within a given π unit reduce the intrinsic polarizability by ~75 percent. Adjustments for multiple substituents are additive. Aliphatic hydrocarbon units have a $P_i^{d'}$ approximating that of ethane or 0.025 $Å^3/cm^3$.

4.4.3. Dipole-Dipole Interaction

Dipole-dipole interactions occur between molecules containing permanent dipoles. The dipole aligns itself with other dipoles in a head-to-tail fashion resulting in a dipole-dipole attraction

between these molecules. Between like molecules the free energy change of the dipole-dipole interaction is given by

$$\Delta G_{ii}(dipole-dipole) = \rho_{d-d} \left(D_i^d\right)^2 V_i$$

where $\rho_{d\text{-}d}$ 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, which are calculated as described earlier in this report.

Since dipole-dipole interactions depend on the position of the polar molecule with respect to its neighbor, the interaction forces is not additive in nature. SPARC adjusts the dipole-dipole interaction as a function of number and magnitude of the microscopic bond dipole moment in the molecule. In addition, SPARC adjusts ΔG_{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-bond interactions.

4.4.4. Hydrogen Bonding Interactions

For single site interactions between like molecules, the free energy change of the interaction is given by

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

where ρ_{HB} is the susceptibility to hydrogen bonding interactions and S_{ii} is a steric reduction factor given by

$$1 \geq S_{ii} = 1 - \left[\rho_{S_{ii}} Size\alpha_{ii} + \rho_{S_{ii}} Size\beta_{ii} - thresh\right]$$

and ρ_{Sii}, Sizeα and Sizeβ are the susceptibility to steric effects, and the steric sizes of the molecules looking back from the α and ß sites, respectively. No steric reduction is applied if the sum of these

sizes does not exceed a threshold value. The threshold value is inferred from physical properties data. The H-bond parameters are data-fitted on measured physical properties and stored in SPARC databases. It should be pointed out that these parameters are process independent; the $\rho_{H\text{-}Bond}$ and steric threshold value are also molecule independent.

4.4.5. Solute-Solvent Interactions

For symmetrical interactions (dispersion) the differential energy for mixing solute i and solvent j is given by

$$\Delta G_{ij}\ (dispersion) = \rho_{disp} \left(P_i^d - P_j^d\right)^2 V_i$$

where i and j designate the solute and solvent molecules, respectively. 'Symmetrical' connotes independence of order-of-mixing (i.e., 'i' into 'j' versus 'j' into 'i') in differential energy density. For interactions that involve molecular orientation (dipolar or H-bonding),

$$\Delta G_{ij} = \Delta G_{ii} + \delta G_{ij}$$

where ΔG_{ii} is the solute self-energy described previously and ΔG_{ij} describes the differential mixing of an 'isolated' solute molecule 'i' into solvent 'j' and

$$\delta G_{ij} = w_c\ \delta G_{ij}^c + w_{nc}\ \delta G_{ij}^{nc}$$

with δG_{ij}^{nc} (δG_{ij}^c) describing solvation with (without) solvent destructuring; these two components might also be termed 'outer' and 'inner' sphere solvation, respectively. w_c and w_{nc} are given by

$$w_c = 1 - w_{nc}$$

$$w_{nc} = 10^{\Sigma \delta G_{ij}^{nc}}$$

where the summations are over all dipole and H-Bonding interactions.

For a single site H-bonding interaction,

$$\delta G_{ij}^{c}(H\text{–}B) = \rho_{HB}\left(-S_{ij}\ \alpha_i \beta_{j'} - S_{ji}\beta_i \alpha_{j'} + S_{jj}\alpha_{j'}\beta_{j'}\frac{V_i}{V_j}\right)$$

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

where

$$\alpha_{j'} \equiv f_{HB}\ \alpha_j$$

$$\beta_{j'} \equiv f_{HB}\ \beta_j$$

The f_{HB} 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. For multiple hydrogen bonds the algebra becomes much more complex; each αß product is of the following form:

$$\delta G_{ij} = \sum_{solute}^{i} [\rho_{HB}\left(\sum_{solute}^{i} S_{ik}^{\alpha}(1+\log(\frac{1}{Z_{ik}^{\alpha}})W_{ik}^{\alpha}\alpha_i\beta_k\right) + \sum_{solvent}^{k} S_{ik}^{\alpha}(1+\log(\frac{1}{Z_{ik}^{\beta}})W_{ik}^{\beta}\alpha_k\beta_i]$$

where W_{ik} is the probability that the $\alpha_k ß_i$ bond will form, and Z_{ik} represents a statistical factor for the interaction. W_{ik} and Z_{ik} are defined as

$$Z_{ik}^{\alpha} = \frac{\alpha_i\beta_k}{\sum_{solute}^{m}\alpha_m\beta_k} \qquad W_{ik}^{\alpha} = \frac{10^{\alpha_i\beta_k}}{O_{intra}^{i} + O_{intra}^{k} + \sum_{solute}^{m} 10^{\alpha_m\beta_k}}$$

where O_{intra} are any intramolecular hydrogen bonds that can compete with the interaction of interest.

Similarly for dipole-dipole interactions,

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

$$\delta G_{ij}^{c}(dipole-dipole) = \rho_{d-d}\ (-2\ D_i\ D_j^{'} + D_{j'}^2)V_i$$

where the dipole D_j is adjusted as $D_{j'} = f_d\ D_j$

The f_d gauges reduction in solute-solvent dipole interaction for outer versus inner sphere solvation.

Likewise for induction interactions,

$$\delta G_{ij}^{nc}\ (induction) = \rho_{Ind}\left[-(D_i P_j^i + P_i^j D_j) + \left(\frac{V_i + V_j}{2}\right) + P_j^i D_j V_j \right]$$

$$\delta G_{ij}^{c}\ (induction) = \rho_{Ind}\left[-D_i P_j^i - P_i^i D_{j'} + P_j^i D_{j'} \right] V_i$$

4.5. Solvents

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. SPARC has a small 'common name' to SMILES string database in memory and will use the common name if it finds it. If the name is not found, the user must supply a name by which the solvent will be recognized. Any molecule that SPARC can run as a solute may be declared a solvent, so essentially any organic solvent may be specified.

4.6. Physical Process Models

All physical process models are built directly from the molecular interaction models described above.

4.6.1. Vapor Pressure Model:

The saturated vapor pressure is one of the most important physiochemical properties of pure compounds. Actually, the vapor pressure is among the most frequently measured and reported physical properties. According to Dykyj et al [47], by the end of 1970's, vapor pressure data (as a function of temperature) were available for more than 7000 organic compounds. 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 pressures 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 pressures 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

$$\log vp_i^o = \frac{-\Delta G_{ii}(Interactio\ n)}{2.303\,RT} + LogT + C$$

where log(T) + C describes the change in the entropy contribution [48] associated with the volume change in going from the liquid to the gas phase. For molecules that are solids at 25° C, the crystal

energy contribution becomes important, especially for rigid structures such as aromatic or ethylenic molecules that have high melting points (greater than 50 ^{o} C). Each intermolecular interaction (dispersion, induction, dipole-dipole and H-bonding) is assumed to have a different, but constant, ratio of enthalpic to entropic contributions to the free energy process at 25^{o} C. SPARC estimates the crystal energy contributions assuming that at the melting point $\Delta G_{ii} = \Delta H_{ii} + T\Delta S_{ii} = 0$. See the Crystal Energy Model section for more details.

The vapor pressure computational algorithm output was initially verified by comparing the SPARC prediction of the vapor pressure at 25^{o} C to hand calculations for key molecules. For sample hand calculations see Figure 18. 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^{o} C. Figure 19 presents the SPARC-calculated vapor pressure at 25^{o} C versus measured values for 747 compounds. The SPARC self-interactions model can predict the vapor pressure at 25^{o} 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 14. The vapor pressure model was also tested on the boiling point and heats of vaporization. See later sections.

Figure 18. Sample hand calculations of the vapor pressure at 25° C for hexane and 1-chlorohexane

Molecular Descriptors	n-C_6H_{14}	$C_6H_{13}Cl$
Volume	131.56	138
Polarizability	11.77	13.73
Dipole	0	1.404
H-Bond	0	0
Dispersion Interaction	$-2.56\left(\frac{11.77}{131.6}\right)^2 131.6 = -2.69$	$-2.56\left(\frac{13.73}{138}\right)^2 138 = -3.49$
Induction Interaction	0.00	$-2.522\left(\frac{13.73}{138}\right)\left(\frac{1.404}{138}\right) 138 = -0.32$
Dipole-Dipole Interaction	0.00	$-2.837\left(\frac{1.404}{138}\right)^2 138 = -0.04$
H-Bond Interaction	0.00	0.00
Entropic Term	log (298) - 0.457 = 2.05	log (298) - 0.457 = 2.05
Total (log vp)	-0.64 atm (observed: -0.7)	-1.8 atm (observed: -1.9)

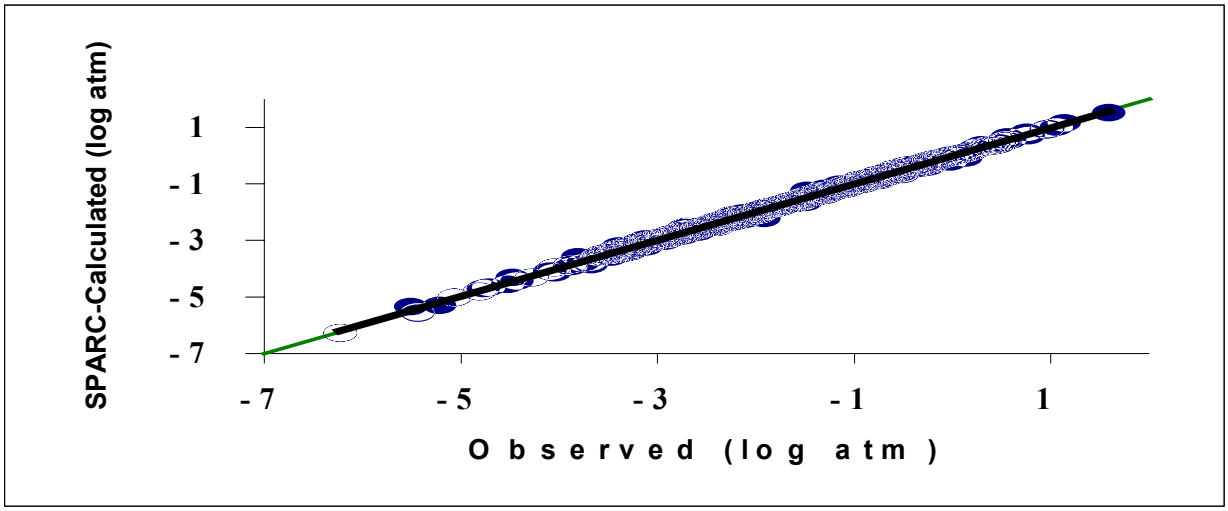


Figure 19. SPARC-calculated vs. observed log vapor pressure for 747 organic molecules at 25° C. The figure includes all the vapor pressure measurements (real not extrapolated) we found in the literature. The RMS deviation was 0.15 log atm and R^2 was 0.994.

4.6.2. Activity Coefficient Model

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

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

where the last term is the Flory-Huggins [49, 50], excess entropy of mixing contribution in the liquid phase for placing a solute molecule in the solvent. The Flory-Huggins term is damped out by orientation interactions (e.g. hydrogen bonding) that reduce the randomness of placement. When the solute and solvent have the same molecular volume, the Flory-Huggins term will go to zero. It should be noted that the negative log activity coefficient for small alkanes in squalane is a consequence of the large Flory-Huggins contributions [22]. The activity follows the Raoult's Law convention (i.e. γ_{ij} -> 1 as χ_i -> 1). The crystal energy is calculated the same way as for vapor pressure.

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 20 presents the validation for SPARC-calculated log activity coefficients versus measured values for 491 compounds at 25° C in 41 different solvents. The SPARC activity coefficient test statistical parameters are shown in Table 14. 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.

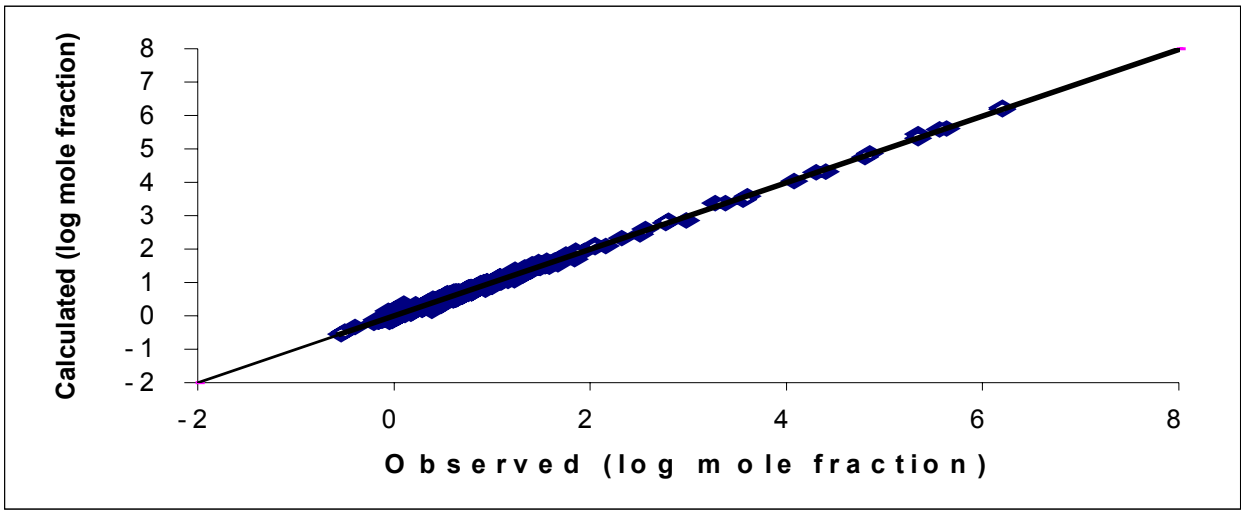


Figure 20. Calculated versus observed log activity coefficients at infinite dilution for 491 compounds in 41 solvents including water. Only 15% of these compounds have strong dipole-dipole and/or H-bond interactions. The RMS deviation was 0.064 log mole fraction R^2 was 0.998.

4.6.3. Crystal Energy Model

For aromatic compounds, the SPARC entropy energy calculator is very similar to that used by Yalkowski [5, 10]. The SPARC calculator first estimates what fraction of the molecule is conjugated (aromatic) and designate that value as F_a. At the melting point, T_{mp}, the ΔG in going from a crystal to liquid, ΔG_{xstal}, is zero, and

$\Delta G_{xstal} = \Delta H_{xstal} - T_{mp} \Delta S_{xstal} = 0$

If we also assume that ΔH_{xstal} and ΔS_{xstal} don't vary significantly with temperature, T, we can write

$\Delta G_{xstal}(T) \approx \Delta H_{xstal} - T \Delta S_{xstal}$

and

$\Delta G_{xstal}(T) \approx T_{mp} \Delta S_{xstal} - T \Delta S_{xstal} = (T_{mp} - T) \Delta S_{xstal}$

Based on the vapor pressure equation given previously in terms of total interaction energy ($\Delta G = -RT\ 2.303 \log vp^o_i$), the contribution of the $\log vp^o_i$ from the crystal energy is given by

$$\log vp_i^o = \frac{(T_{mp} - T)\ \Delta S_{xstal}}{-2.303RT}$$

SPARC uses the Yalkowski value for $\Delta S_{xstal} = 13.5$ eu. for aromatics. SPARC then calculates a first order correction as

$\Delta S_{xstal} = 13.5\ F_a + (1 - F_a)\ N_a + LC + AAR$

where F_a is the aromatic fraction, N_a is a derived non-aromatic contribution of the crystal entropy (this value is small) and LC is the coiling entropy that SPARC uses to estimate entropy change associated with alkane chains of length greater than Y (this is linear in length Y). AAR is the aromatic-aromatic bond rotation entropy change that is associated with the very low frequency internal rotation in molecules having aromatic-aromatic bonds (e.g. biphenyl). This model works fine for compounds with melting points less than 300° C. When the melting point gets much above 300° C, the model breaks down. This could be due to several factors, including our assumption that ΔH_{xstal} are ΔS_{xstal} are independent of temperature. To compensate for this, SPARC incorporates a second order non-linear contribution to $\Delta S_{xstal.}$ The second order correction term looks like

$$\Delta S_{xstal2} = \Pi \times F_a \frac{(T_{mp} - T)^2}{T^2}$$

where Π is data fitted contribution. The corrected crystal component of the log vapor pressure term then becomes

$$\log vp_i^o = \frac{\left[(13.5) F_a + (1 - F_a) N_a + LC + AAR\right]}{-2.3030 RT} (T_{mp} - T) + \frac{\Pi F_a}{T^2} (T_{mp} - T)^2$$

4.6.4. Enthalpy of Vaporization

The heat of vaporization, ΔH_v, is sometimes referred to as the enthalpy of vaporization. It is the difference between the enthalpy of the saturated vapor and that of the saturated liquid at the same temperature. ΔH_v is related to the slope of the vapor pressure versus temperature curve by the Clausius-Clapeyron equation. Many estimation methods for ΔH_v are simply based on either the Clausius-Clapeyron equation or the law of corresponding states. However, 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

$$Log\ \Delta H_{ii}(\text{vap}) = \frac{-\Delta G_{ii}^H (Interactio\ n)}{2.303 RT} - LogT + C^H$$

where ΔG_{ii}^H is the free energy change of the self interactions modified for the enthalpic contribution 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 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 [51]. The crystal energy calculation is the same as in the vapor pressure model. For the $\Delta H_{(vaporization)}$ contribution, SPARC modifies the susceptibility of each molecular interaction mechanism as:

$$\rho^{\Delta H}_{Mechanism} = \Omega_{Mechanism} \quad \rho_{Mechanism}$$

where $\Omega_{Mechanism}$ is dependent on the interaction mechanism (dispersion, induction, dipole-dipole and H-bond), is data-fitted at 25° C and stored in the SPARC database. Likewise, the susceptibility, $\rho_{Mechanism}$, depends on the type of the interaction mechanism (dispersion, induction, etc) and is the same as explained earlier. Figure 21 shows the performance of the SPARC calculator for heat of vaporization at 25° C and at the boiling point. The test statistical parameters are shown in Table 14.

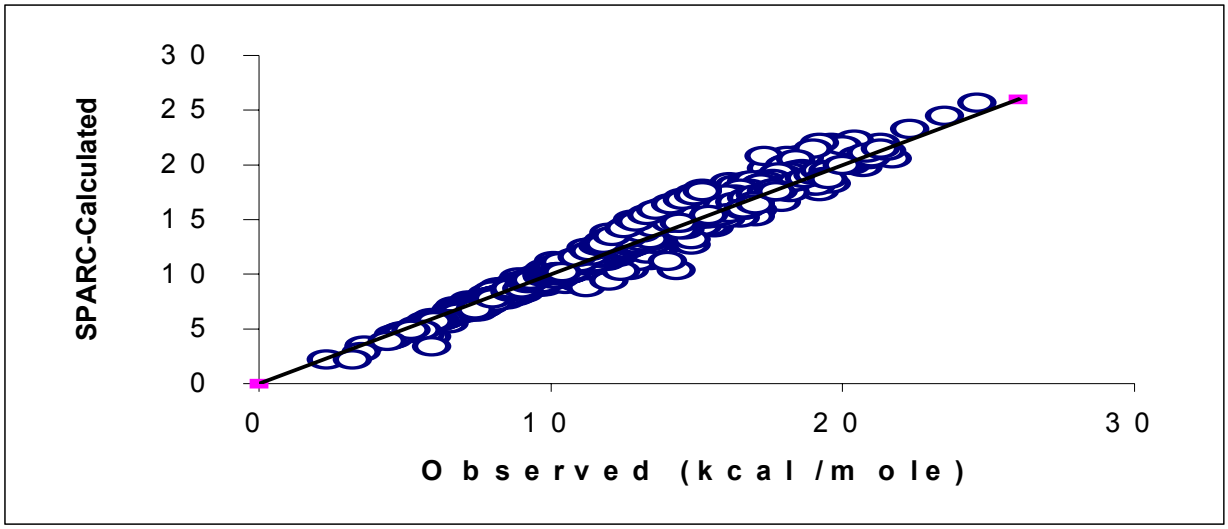


Figure 21. SPARC-calculated vs. observed heat of vaporization. The RMS deviation was 0.302.

4.6.5. Temperature Dependence of Physical Process Models

The temperature dependence of some physical process and molecular volume models was included in the previous discussion. In addition to the normal free energy temperature dependence,

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 the susceptibilities for dipole-dipole and hydrogen bonding as follows:

$$\rho_{\text{dipole-dipole}} = \rho^{25}_{\text{dipole-dipole}} \left[\frac{298}{T}\Omega_{\text{dipole-dipole}} + (1-\Omega_{\text{dipole-dipole}})\right] \quad \frac{298}{T}$$

$$\rho_{\text{H-Bond}} = \rho^{25}_{\text{H-Bond}} \left[\frac{298}{T}\Omega_{\text{H-Bond}} + (1-\Omega_{\text{H-Bond}})\right] \quad \frac{298}{T}$$

where $\Omega_{\text{H-Bond}}$ and $\Omega_{\text{dipole-dipole}}$ are data fitted parameters stored in SPARC database. Both $\Omega_{\text{dispersion}}$ and $\Omega_{\text{induction}}$ are set to be equal to 1. The ΔH and ΔS temperature dependence are described by the first and the second term, respectively. SPARC assumes that the $\Delta H/\Delta S$ contribution to ΔG 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. Further, the enthalpic term 298/T is then expanded as a polynomial function of all the interaction forces. In general, for any "activity-driven" process in SPARC, the susceptibility, ρ, of a given interaction at temperature T is modeled as function of the H-bonding (HB), dipole density (D^d) and polarizability density(P^d) is given by

$$\rho^{T}_{Mechanism} = \left[1 + (1 - \sum_{n}^{5} a_n \left(\frac{298}{T}\right)^n) \;\; f(P^d, D^d, HB)\right] \rho^{25}_{Mechanism}$$

where $\rho_{\text{Mechanism}}$ is dependent on the interaction mechanism (dispersion, induction, dipole-dipole and H-bond). When T = 25° C, the two susceptibilities are equal to each other. a_n are trainable parameters quantified from physical properties measurement, mainly on 4000 boiling points measured at different pressures, 600 heat of vaporizations (at the boiling point) and on more than 600 GC chromatographic retention times at different temperatures ranging from 30 to 190° C.

However, this is a small correction to any physical property such as ΔH_v, vapor pressure and boiling point. It describes the small temperature dependence of the enthalpic contribution. For example, for non-polar molecules such as alkanes, alkenes and aromatics, this correction amounts to only a few degrees in the boiling point calculator. For molecules that contain a small dipole moment, such as aromatic or aliphatic halogens, this correction might be 3-6 degrees in boiling point estimation. Molecules that have a large dipole moment, such as nitrobenzene, or the capability to H-bond with each other, such as phenol, might produce a correction between 6-12 C in boiling point estimation.

4.6.6. Normal Boiling Point

If a liquid is heated in an open container, its temperature rises only until its vapor pressure equals the external pressure. At this point, the liquid changes completely into vapor at constant temperature. This temperature is known as the normal boiling point of the liquid. 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. 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 above the boiling point. The SPARC boiling point calculator was tested against 4000 boiling points measured at different pressures ranging from 0.05 to 1520 torr spanning a range of over 800^o C as shown Figure 22 while the statistical parameters are shown in Table 14.

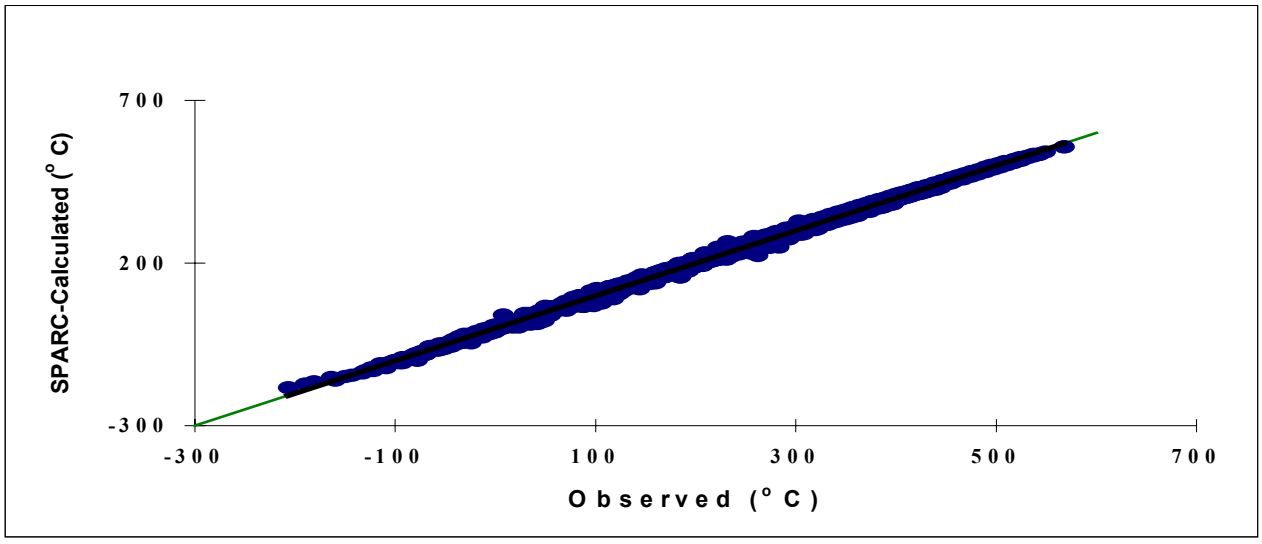


Figure 22. SPARC-calculation vs. observed 4000 boiling points for pressure ranging from 0.1 to at 1520 torr. The Total RMS deviation was 5.71° C. The RMS deviation for polar molecules was 8.2° C and R^2 was 0.9988, while for non-polar molecules the RMS was 2.6° C and R^2 was 0.9995

4.6.7. Solubility (Activity Coefficient as a Function of Concentration)

Solubility is the maximum amount of a compound that will dissolve in pure solvents at a given temperature. SPARC does not calculate the solubility from first principles, but from the activity coefficient model described previously. SPARC estimates molecular solubility from a calculation of 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. In order to overcome these limitations, SPARC employs an iterative calculation. SPARC sets the initial guess of the solubility as $\chi_{guess} = 1/\gamma^{\infty}$. SPARC then 'prepares' a mixed solvent that is χ_{guess} in the solute and $(1-\chi_{guess})$ in the solvent. SPARC recalculates γ^{∞} in the 'new' solvent. This process is continued until γ^{∞} converges or goes to 1 (miscible). Using this technique, SPARC correctly calculates the solubilities of the aliphatic alcohol series and shows propanol to be miscible and butanol to be very soluble in water. This technique also works for

mixed solvent systems. For example, the log mole fraction solubility of toluene in a water(30)/ethanol(70) mixture is observed to be -1.47. The initial "guess" from the SPARC calculator was -1.87. This value converged to -1.50 after three iterations. Figure 23 shows display a result of SPARC calculated log solubilites of 260 compounds versus observed values at 25° C. The RMS deviation was 0.321 with an R^2 of 0.991. The RMS deviation for 119 liquid compounds was 0.135 with an R^2 of 0.997, while for 141 solids log mole fraction solubilties, the RMS deviation was 0.419 with an R^2 of 0.985. The RMS deviation for the solids compounds is 3 times greater than that for the liquid compounds due to the crystal energy contributions.

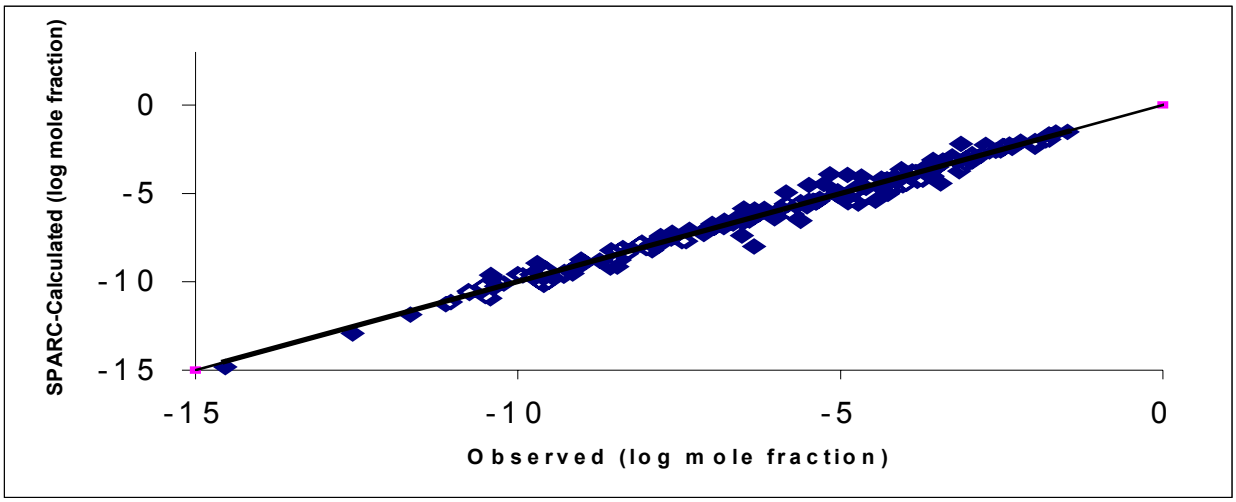


Figure 23. A test results for SPARC calculated log mole fraction solubilites for 260 compounds at 25° C versus observed values. The RMS deviation is 0.321 and R^2 is 0.991. The RMS for 119 liquid solubilties is 0.135 and R^2 is 0.997 while for the 141 solids the RMS is 0.419 and R^2 is 0.985.

4.6.8. Mixed Solvents

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 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.

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 summed over all solvent components. SPARC calculation of solubility of organic molecules in binary solvent mixtures has been tested and appears to work well. Most of the binary mixture data available is in the form of solubilities. Figure 24 shows SPARC-calculated versus observed log activities in mixed methanol/water medium.

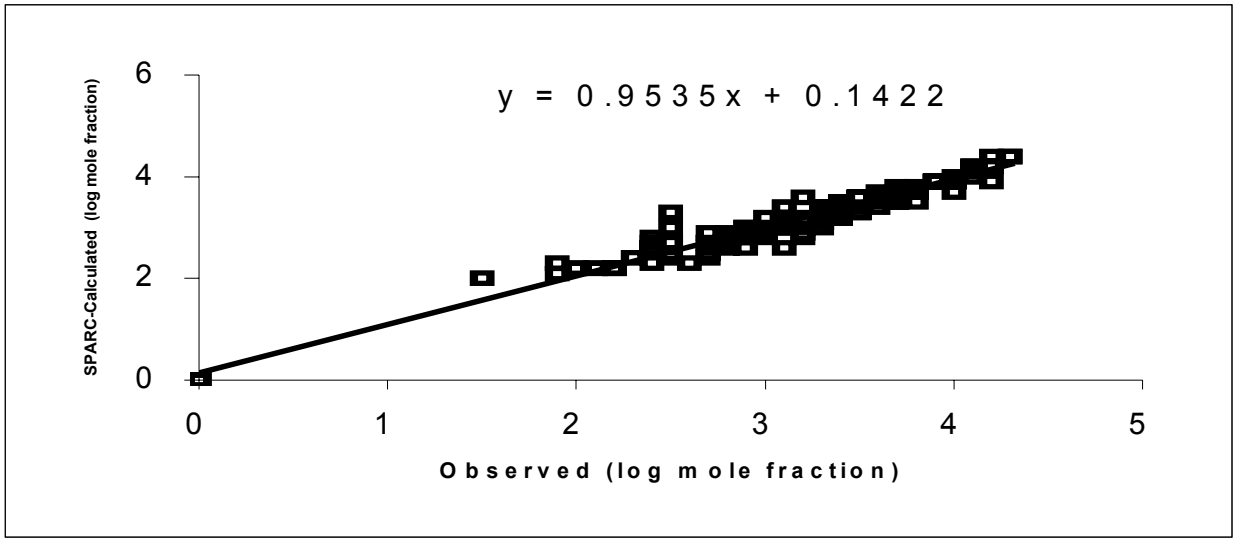


Figure 24. SPARC-calculated versus observed log activities for 120 compounds in water/methanol mixed solvent at 25° C. The RMS deviation error was 0.18 and the R^2 was 0.980.

4.6.9. Partitioning Constants

All partitioning (Liquid/Liquid, Liquid/Solid, Gas/Liquid and Gas/Solid) constants are determined in SPARC by calculating the activity of the molecular species in each of the phases without any modification to the activity models.

4.6.9.1 Liquid/Liquid Partitioning

SPARC calculates the liquid-liquid partition constant 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

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

where the γ^{∞}'s are the activities at infinite dilution of the compound of interest 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 only this calculation. SPARC can calculate a compound's liquid-liquid partition coefficient for any two immiscible phases. The phases may also 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. The SPARC-calculated K_{ow}'s are not greatly different than those calculated assuming dry or pure octanol when the molecule of interest is small and/or has a large hydrogen-bonding interaction. However, the differences can be significant (~0.8 log units) when the molecule is large and hydrophobic, as in the case of large polynuclear aromatics (PNA's) e.g., coronene. Figure 25 shows the current performance of SPARC for log $K_{solvent/water}$, where the solvents are carbon tetrachloride, benzene, cyclohexane, ethyl ether, octanol and toluene. Figure 26 displays a comparison of the EPA Office of Water 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.

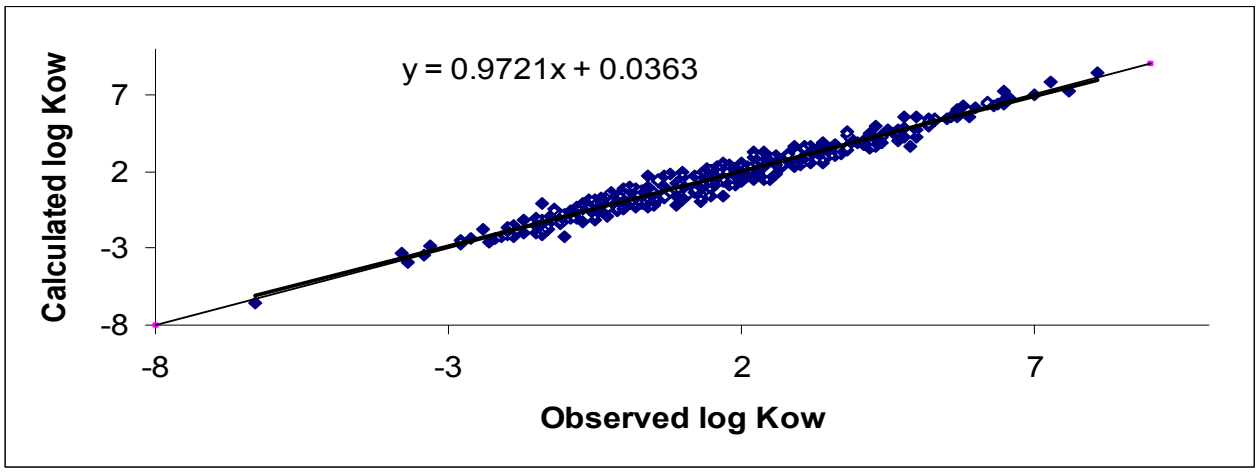


Figure 25. SPARC-calculated versus observed log distribution coefficients $K_{solvent/water}$ for 623 organic compounds in carbon tetrachloride, benzene, cyclohexane, ethyl ether, octanol and toluene at 25° C. The RMS deviation was 0.38 and R^2 was 0.983.

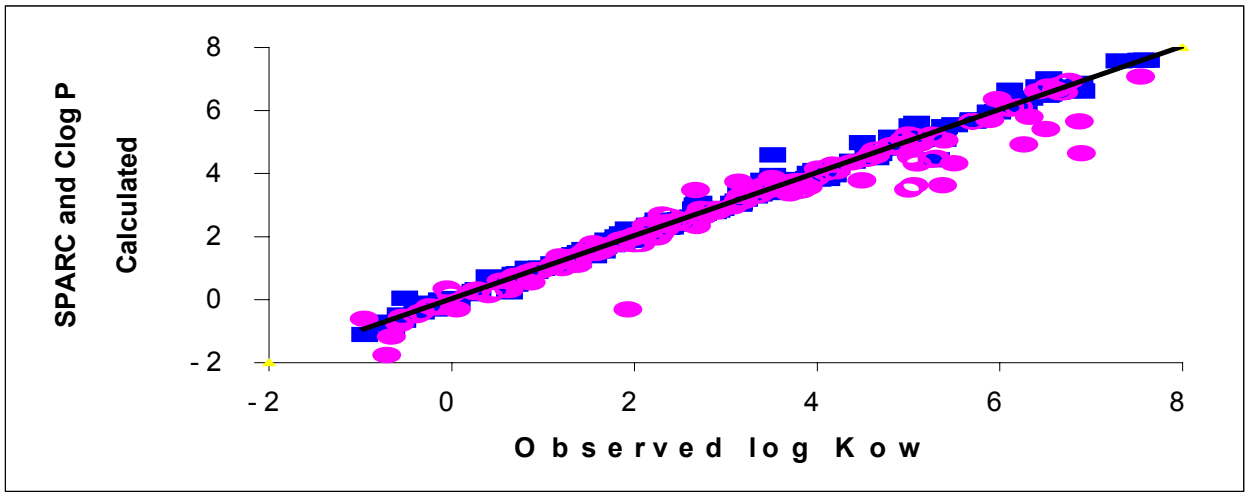


Figure 26. Test for EPA OWPP calculated $K_{octanol/water}$ versus recommended measured values. Squares are SPARC calculate values, circles are ClogP calculate values. The RMS deviation and R^2 values were is 0.18 and 0.996 respectively for SPARC and 0.44 and 0.978 respectively for ClogP calculated values

4.6.9.2. 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, ΔG_{jj}, are dropped from the calculation. Later in this report, 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.

4.6.9.3. Gas/liquid (Henry's constant) Partitioning

For solutions that are so dilute that each solute molecule is surrounded only by solvent molecules, small changes in the solute concentration will not affect the composition of the nearest neighbor molecules. In this case, the intermolecular interactions the solute molecule experiences will not change with concentration, the vapor pressure will be proportional to the mole fraction of the solute and Henry's constant may be expressed as

$$H_x = vp_i^{o} \gamma_{ij}^{\infty}$$

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 are used to calculate the Henry's constant for a any solute out of a given solvent liquid phase as shown in Figure 27. An application of SPARC-calculated Henry's law constants for the prediction of gas-liquid chromatography retention times in polar and non-polar stationary liquid phases is presented later in this report.

4.6.9.4. Gas/Solid Partitioning

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 (i.e., surface interactions; no dissolution of the solute in the solid). This type of modeling is useful for calculating retention times for capillary column gas chromatography.

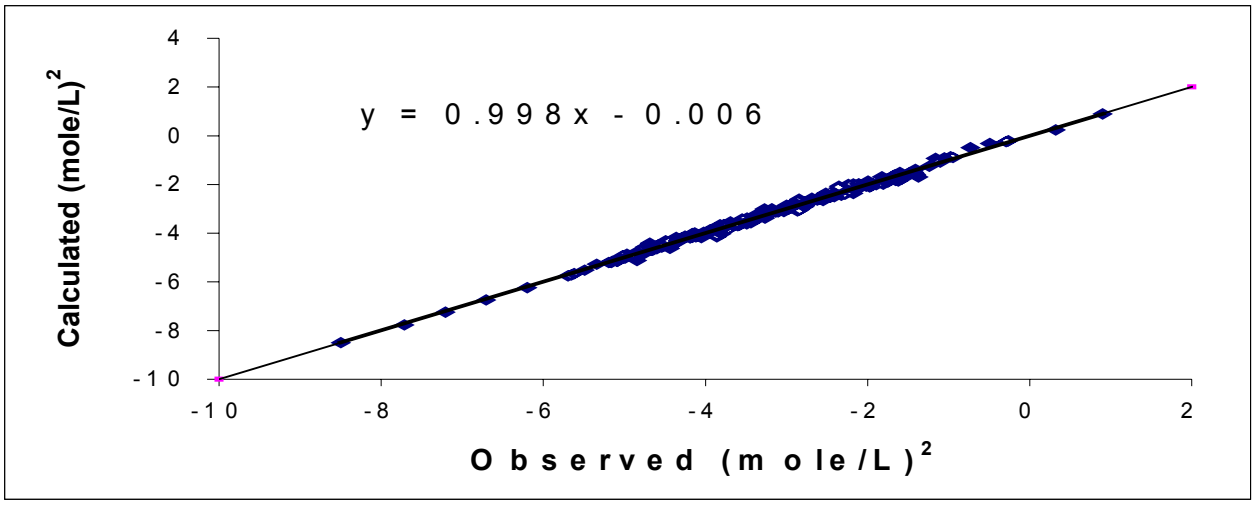


Figure 27. Observed vs. SPARC-calculated Henry's constants for 271 organic compounds in hexadecane. The RMS deviation was 0.1 $(mole/L)^2$ with an R^2 was 0.997.

4.6.10. Gas Chromatography

Despite some limitations, the Kov'ats index has found much greater use than all other specialized retention specification schemes. The Kov'ats index is the only retention value in gas-liquid chromatography (GLC) in which two fundamental quantities, the relative retention and the specific retention volume are united [52]. Moreover, a series of explicit relationships between retention indices and a number of physicochemical quantities related to GLC have been developed. Also, many different linear relationships between the Kov'ats index value for a molecule and other fundamental molecular properties such as carbon number, boiling point and refractive index have been derived [52, 53].

The Kov'ats [54] index expresses the retention of a compound of interest relative to a homologous series of n-alkanes examined under the same isothermal conditions. The Kov'ats index for a particular compound of interest is defined as the carbon number (CN) multiplied by 100 of a hypothetical n-alkane having exactly the same net retention volume characteristics of the compound of interest measured under the same conditions:

$$KI = 100 * \left(\frac{\log V_{Nc} - \log V_{Nx}}{\log V_{Nc} - \log V_{N(cn+1)}} + CN\right)$$

where KI is the Kov'ats Index of the compound of interest, X, X is a compound with a retention between that of the first n-alkane and second n-alkane standard, CN is the number of carbon atoms in the first n-alkane standard, cn +1 is the number of carbon atoms in the second n-alkane standard, V_{Nx} is the net retention volume of the compound of interest X, V_{Nc} is the net retention volume of the first n-alkane standard, and $V_{N(c+1)}$ is the net retention of the second n-alkane standard.

Numerous investigators have attempted to calculate or predict KI using physicochemical descriptors like boiling point, density, dipole moment, etc. Unfortunately, all of the correlations of retention indices and the various physicochemical properties are either relatively limited in scope or their application is restricted to a particular chemical class. Other attempts to predict retention indices for a wide range of molecular structures using molecular bond length, molecular bond angle, topological indices [52, 53, 55], or other molecular characteristics have been only marginally successful. Most of these studies also were restricted to a particular class of molecules on a specific stationary liquid phase.

Despite all the attempts to predict Kov'ats index, no realistic scheme with widespread application for different classes of compounds on different polarity stationary liquid phases is available. The following is a discussion of SPARC models for the Henry's constant and Kov'ats index applied to branched hydrocarbons on squalane. Our goal, however, is to develop general mathematical models to calculate the Kov'ats index at any temperature for a wide range of different classes of compounds on different polar and non-polar stationary liquid phases using the SPARC Henry's constant calculator described earlier.

SPARC vapor pressure and activity coefficient models are used to calculate the Henry's constant for a solute in a squalane liquid phase. Henry's constant can be related to the net retention volume, V_N, by

$$H_i = \frac{R\,T\;\;V_L}{M\;\;V_N}$$

where M is the molecular weight of the solvent, and V_L is the volume of the stationary phase. Substituting in the previous equation (previous page), we get

$$KI = 100\;\;x\;\;\left(\frac{\log H_{Nx} - \log H_{Nz}}{\log H_{N(z+1)} - \log H_{Nz}} + CN\right)$$

where H_{Nx}, H_{Nz}, and $H_{N(z+1)}$ are Henry's constant for the compound of interest X, first n-alkane standard and the second n-alkane standard, respectively.

4.6.10.1. Calculation of Kov'ats Indices

Retention indices may be reproduced within a laboratory using modern instrumentation with considerable precision over finite time periods. Reproducibility of 0.1 units was reported by Schomburg and Dielmann [56] in 1973. However, squalane columns produced reproducible results for only for a few hours and, therefore, need to be continually replaced. For routine operation a reproducibility of about 1 Kov'ats index unit might be expected with a squalane liquid phase. Unfortunately, inter-laboratory reproducibility remains unsatisfactory, except for a few cases. The actual discrepancies between experimental values of retention indices for identical compounds obtained at different laboratories in routine analysis is assumed to be up to ± 10 Kov'ats units or even more [53].

4.6.10.2. Unified Retention Index

The unified retention index developed by Dimov [57, 58] has been used to explain the variations in the retention index of simple hydrocarbons on Squalane liquid phase. The temperature dependence of the retention index is well known, the function d(KI)/dT being hyperbolic. A statistical treatment using simple regression analysis of the data allows computation of the unified retention index (UI_T) as

$$UI_T = UI_o + \left(\frac{dUI}{dT} \right) T$$

where UI_o is the Kov'ats index at 0° C and dUI/dT the temperature dependence where -dUI/dT is the slope of the plotted data. The UI_T is a statistically obtained value and, hence, it is more reliable than any individual KI_{exp} value. Also dUI/dT is a more reliable value than d(KI)/dt for estimation of the temperature dependence of retention indices. The UI_T and dUI/dt served as the observed values for the optimization of SPARC dispersion parameters for prediction of the retention indices. Figure 28 shows the SPARC-calculated versus observed [57, 58] Kov'ats index at 25° C for 156 organic compounds in a squalane liquid phase. The RMS deviation was less than 7 Kov'ats units, a value that approximates interlaboratory experimental error. The SPARC physical properties and the temperature dependence models were also tested on GC chromatographic retention times in non-polar liquid phase such as squalane and B-18, and polar liquid phase such as SE-30, OV-101 and PEG-20M at various temperatures. The RMS deviation for the Kov'ats index at 80° C in squalane and at 130 ° C in B-18 for139 organic compounds was 9.3 and 12 Kov'ats units, respectively. See Table 14 for some of these results.

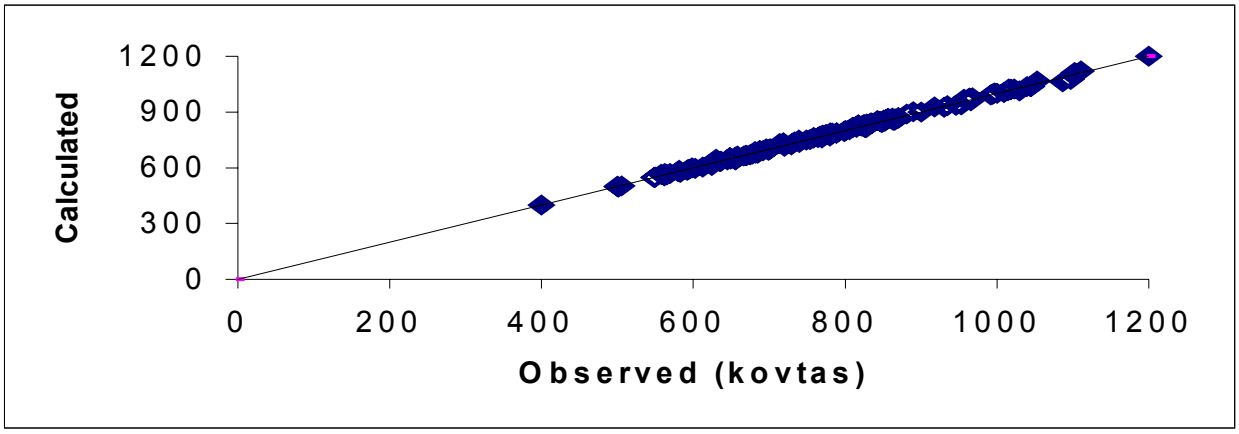
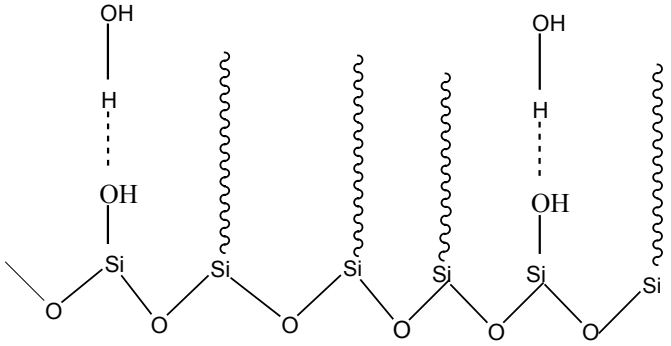


Figure 28 SPARC-calculated versus observed values for the GC chromatographic retention time in squalane liquid phase25^{o} C for 156 organic compounds. The RMS deviation was 7 kovats.

4.6.11. Liquid Chromatography

Just as the gas-liquid partition constant (Henry's constant) can be used above to model GLC retention time, SPARC uses liquid-solid partition constant to model liquid chromatography. To date we have only looked at one reversed phase separation. This work was a very precise measurement of the k prime values for a broad range of solutes on a C_{18} reversed phase column using several different mobile phase compositions. Our major observation in modeling the retention times was that the data could not be modeled without including a polar, hydrogen bonding species as part of the stationary phase. The activities of several of the molecules in the data set had been measured in hexadecane (very close to C_{18}) and in mixed solvents. These measurements were not consistent with the observed LC retention times whenever the molecule of interest had strong hydrogen bonding sites. We modeled the LC retention times using a three phase model. The mobile phase was modeled as a mixed solvent as described previously with no further refinements. The stationary phase was modeled as two stationary phases. The C_{18} phase bonded to the silica was treated as a C_{18} molecule. The second stationary phase was modeled as an unknown phase whose polarizability density, dipole density and hydrogen bonding characteristics were inferred by SPARC

from the observed retention times. The values for these later molecular descriptors were within a few percent of what would be expected for water sitting on isolated sites on the surface.



Wetted Silica on a C_{18} Chromatographic Column

Using surface water as the third phase, SPARC models LC retention times (relative to one of the molecules) as

$$\log K_{rel} = \log K_{stationary/mobile} + C_{ref}$$

where the stationary/mobile phase K is expressed as

$$K_{stationary/mobile} = F\,K_{C_{18}/mobile} + (1-F)\,K_{surface\ water/mobile}$$

where F is the fraction surface coverage of C_{18}. The two partition coefficients are calculated as

$$\log K_{C_{18}/mobile} = \log \gamma^{\infty}_{mobile} - \log \gamma^{\infty}_{C_{18}} + \log R_{C_{18}/mobile}$$

$$\log K_{surface\ water/mobile} = \log \gamma^{\infty}_{mobile} - \log \gamma^{\infty}_{surface\ water} + \log R_{water/mobile}$$

where the log R values are the molecularity corrections to convert the activity based K values to concentration based K's. Using this approach the best fit to the data was found with a stationary phase composed of 95% C_{18} and 5% isolated surface water, both presumably bound to silica. The following figure is a fit of the data.

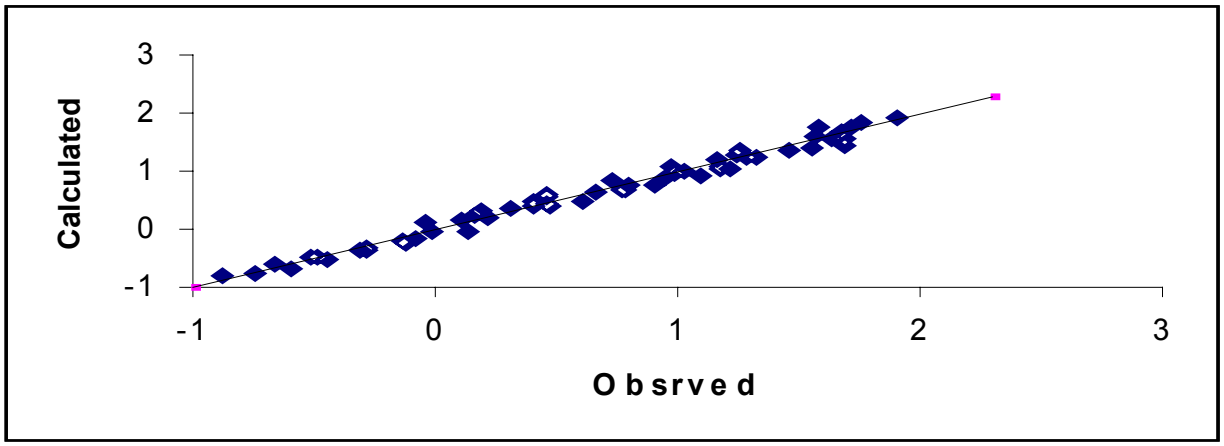


Figure 29. SPARC-calculated versus observed log $K_{surface}$ for LC retention time in water methanol mixture at 25° C. The RMS deviation was 0.095 and R^2 is 0.992.

4.6.12. Diffusion Coefficient in Air

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 for a wide variety of molecules. This equation is that of Wilke and Lee [59], which for a general binary diffusion coefficient is expressed as:

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

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, 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 [60] where $T^* = kT/\varepsilon_{AB}$. The term σ is determined from the liquid molar volume (calculated by SPARC as a function of T) as, $\sigma_{AB} = 1.18V^{1/3}$. The terms in T^* are given by $\varepsilon/k = 1.15\ T_B$, where T_B is the normal boiling point for the molecule in K. Given the temperature and pressure, SPARC can calculate the volume of the molecule at that temperature, and then its the normal boiling point, T_B,

for the molecule. The coefficients of the Neufeld equation are stored in the SPARC database. The Wilke-Lee approach predicts gas phase diffusion coefficients to better than 6%. Figure 30 compares observed to SPARC calculated gas phase diffusion coefficients at 25° C.

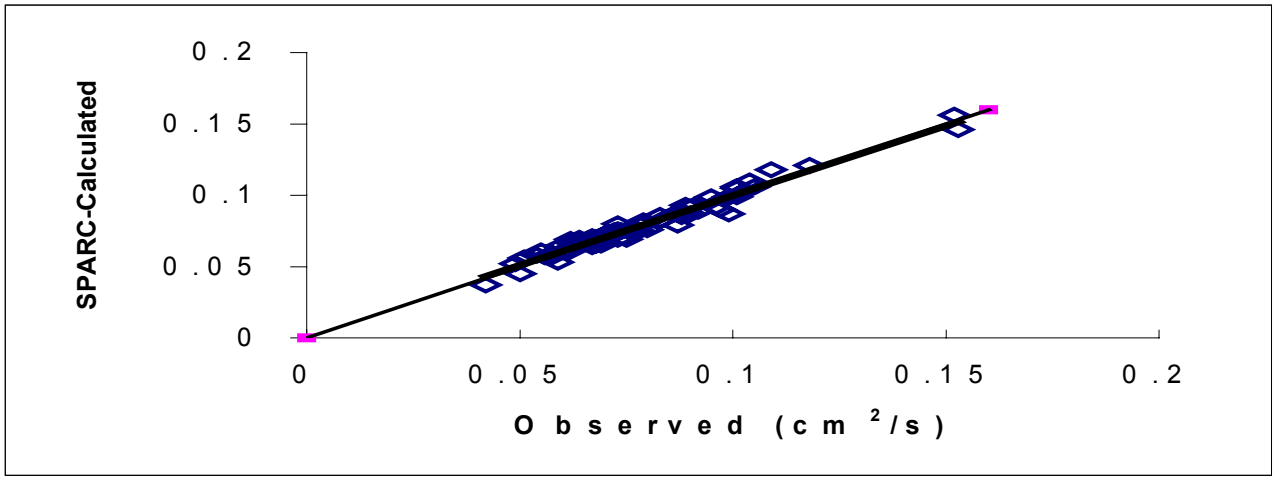


Figure 30. The SPARC-calculated versus observed diffusion coefficients in Air. The RMS deviation was 0.0034 cm^2/s.

4.6.13. Diffusion Coefficient in Water

Several engineering equations exist that do a very respectable job of calculating molecular diffusion coefficients in water. The equation most compatible with the SPARC calculator is

$$D_w = \frac{1.4\times10^{-4}}{Vis_{water}^{1.1} V_i^{0.6}}$$

where V_i is the molar volume and Vis_{water} is the viscosity of water equal to 1.004 at 20° C.

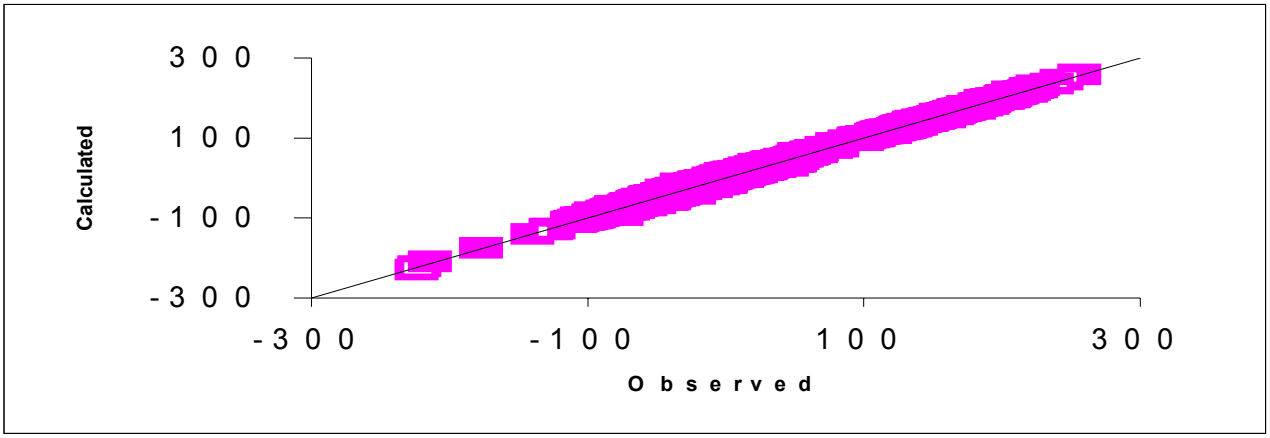


Figure 31.Observed vs. SPARC training set for 2400 calculations. The RMS is 0.29 and R^2 is 0.997.

4.7. Conclusion

A composite SPARC training set output is shown in Figure 31. 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 15) as well as 18 mixed solvent systems.

Table 15. 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 can reliably estimate numerous physical properties of compounds using the same interaction models without modifications or additional parameterization. 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 of interest within a factor of 2, or even better. For complex structures, where dipole-dipole and/or H-bond interactions are strong, SPARC calculations are within a factor of 3-4.

5. PHYSICAL PROPERTIES COUPLED WITH CHEMICAL REACTIVITY MODELS

SPARC models have been extended to ionic organic species by incorporating monopole electrostatic interaction models into SPARC's physical properties toolbox. These ionic models play a major role in modeling the activity and solubility of ionic species in any solvent system. These capabilities (ionic activity), in turn, allow SPARC to calculate gas phase pK_a. Likewise, the calculation of gas phase pK_a will allow SPARC to estimate ionization pK_a, zwitterionic equilibria, ionic partitioning and $E_{1/2}$ chemical reduction potential in any solvent system.

5.1. Henry's Constant (Gas/liquid Partition Coefficient) for Charged Compounds

Recently, experimentalists have been able to carryout reasonably accurate measurements of proton transfer equilibria in the gas phase. These measurements provide a direct measure of relative gas phase acidity. Several international meetings have been held with the purpose of developing a coherent absolute scale for gas phase acidity. This scale is now relatively stable, and Professor Taft at U.C. Irvine has kindly provided us with these screened datasets. The combination of absolute and relative pKa's in both the gas phase and in water were used to develop the SPARC ionic interaction models. The following thermodynamic cycles were used in this development.

AH_{gas}	→	$A^-_{gas} + H^+_{gas}$	Gas pK_a
A^-_{gas}	→	$A^-_{water\ (solv)}$	- Henry's Constant
H^+_{gas}	→	$H^+_{water\ (solv)}$	- Henry's Constant
$AH_{water\ (solv)}$	→	AH_{gas}	Henry's Constant
$AH_{water\ (solv)}$	→	$A^-_{water\ (solv)} + H^+_{water\ (solv)}$	pK_a in water $_{(solv)}$

Steps 1, 4 and 5 are (or are related to) the gas phase pKa for AH, the Henry's constant of AH out of water (solvent) and the pKa of AH in water (solvent), respectively. These values are either known or can be calculated by SPARC. Step 3 represents the Henry's constant for a proton, and will be the same (or a constant) for all molecules AH. This value, along with those for most counter ions in water, have been estimated in the literature. ΔG for step 2 can be inferred from the other steps. SPARC monopole interaction models were developed to calculate the inferred step 2 values.

5.1.1. Microscopic Monopole

The effective molecular monopole (microscopic monopole) was computed as the sum of monopole contributions of individual substituents, modified for steric blockage by appended molecular structures. Individual monopoles were summed algebraically. Monopole moments for each charged substituent were estimated. For each charged substituent, S+/-, an S+/--methyl, S+/--aromatic and S+/--ethylenic dipoles were inferred from data and stored in the SPARC database.

5.1.2. Induction-Monopole Interaction

Induction or monopole-induced dipole interactions will occur between molecules where one (or both) contain a monopole. Between like molecules

$$\Delta G_{ii}\ (monopole\ induction) = \rho_{ind}\ P_i^i\ m_i\ V_i$$

where ρ_{ind} is the susceptibility to induction and P^i_i, m_i and V_i are polarizability density (adjusted for induction), monopole density and molar volume of the molecule-in-question, respectively.

$$P_i^i = \frac{\bar{\alpha}_i + A_{Ind}}{V_i} \qquad and \qquad m_i = \frac{MS_i}{V_i}$$

where MS_i is the microscopic monopole strength described above, A_{ind} is a polarizability adjustment for induction and $\bar{\alpha}_i$ is the average molecular polarizability. Induction describes molecular polarization effected by a monopole on the surface, averaged over all orientations of the molecule. Inductive polarization interactions 'propagate' effectively within conjugated systems, but only one or two atoms deep in a nonconjugated array of atoms. SPARC adjusts the molecular polarizability algorithmically, utilizing electron withdrawing/releasing substituent parameters derived from pK_a models as described previously.

5.1.3. Monopole-Monopole Interaction

Monopole-monopole interactions occur between molecules, each containing a monopole. Between like molecules

$$\Delta G_{ii}(monopole - monopole) = \rho_{m-m} m_i^2 v_i$$

where ρ_{m-m} is the susceptibility to monopole-monopole interactions; m_i and V_i are monopole density and molar volume of the molecule-in-question, which are calculated by SPARC as described in the previous section .

5.1.4. Dipole-Monopole Interaction

Dipole-monopole interactions occur between molecules containing a permanent dipole and a monopole. Between like molecules

$$\Delta G_{ii}(dipole-monopole) = \rho_{d-m} D_i m_i V_i$$

where $\rho_{d\text{-}m}$ is the susceptibility to monopole-dipole interactions, d_i, m_i and V_i are dipole density, monopole density and molar volume of the molecule-in-question which are calculated by SPARC described previously.

51.5. Hydrogen Bonding Interaction

An α and a ß associated with each of the charged substituents are calculated as described earlier in this report and the same SPARC hydrogen bonding interaction models are used

5.2. Estimation of Ionization pKa in the Gas Phase and in non-Aqueous Solution

The pK_a in the gas phase was calculated after these monopole ion models were stable, using the same thermodynamic loop above with water as the solvent. Likewise, the pK_a in any solvent can be calculated by using the same thermodynamic loop except changing the solvent from water to the solvent of interest. Figure 32 and 33 show the SPARC calculated versus observed values for the pK_a's in the gas phase and in 9 different non-aqueous solvents.

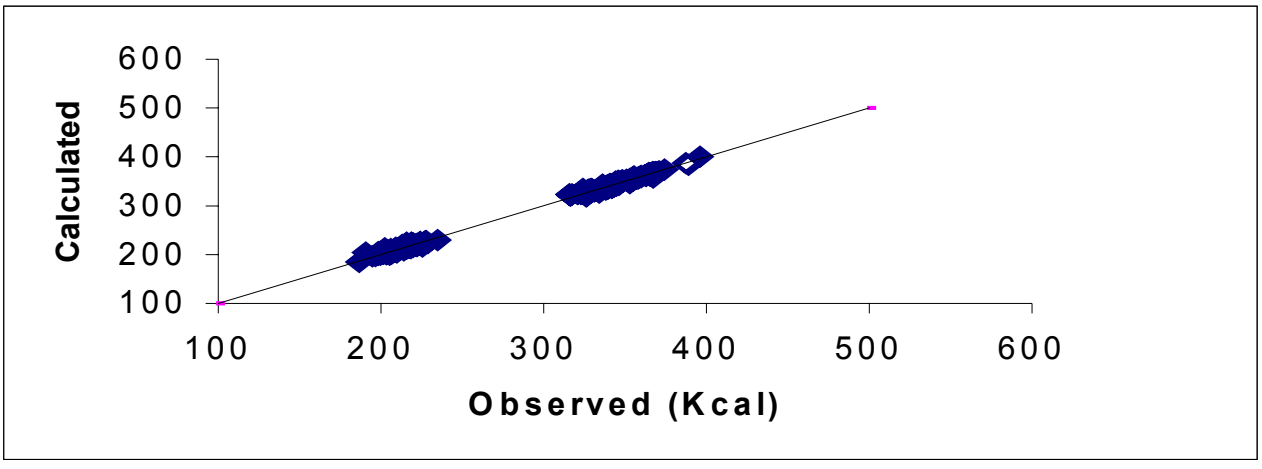


Figure 32. SPARC-calculated versus calculated ionization pK_a in the gas phase for 400 organic compounds. The RMS deviation was 2.25 Kcal with an R^2 of 0.998

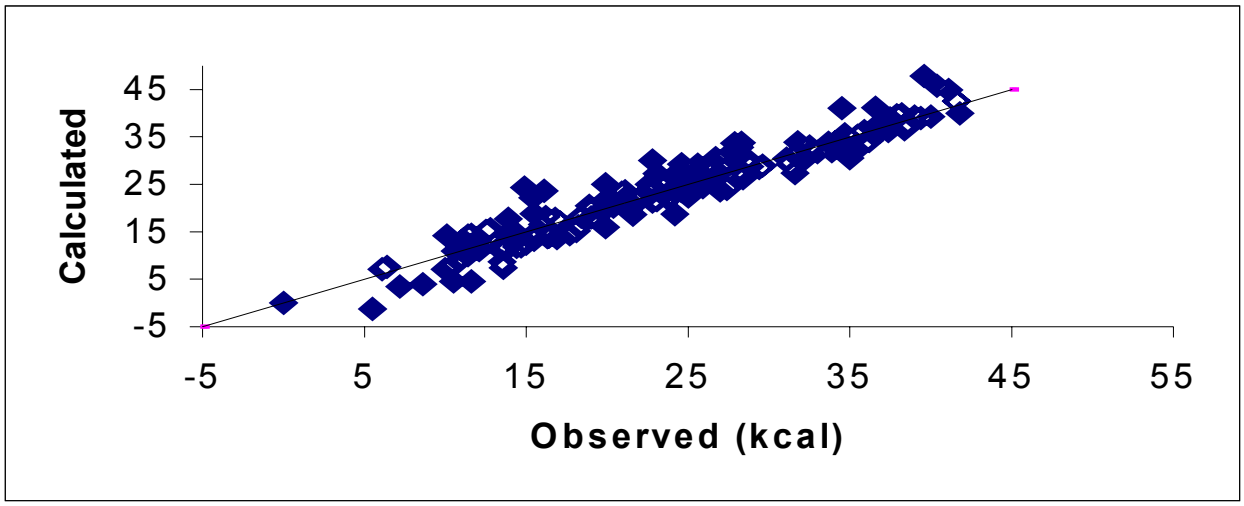


Figure 33. SPARC-calculated versus calculated ionization pK_a in non-aqueous solvents. Solvents were DMSO, THF, DMF, 3 alcohols, aceteonitrile, pyridine and acetic acid. The RMS deviation was 1.9 Kcal with an R^2 of 0.996.

5.3. $E_{1/2}$ Chemical Reduction Potential

The original electron affinity (EA) calculator models were first refined to better integrate with the new models that were used to estimate one electron reduction in the condensed phase. As was the case for estimating gas phase and non-aqueous pK_a, SPARC uses the following thermodynamic cycles:

$$M_{gas} + e^-_{gas} \rightarrow M^-_{gas} \qquad \Delta G_{EA}$$

$$M_{solvent} \rightarrow M_{gas} \qquad -\Delta G^M_H$$

$$e^-_{solvent} \rightarrow e^-_{gas} \qquad -\Delta G^e_H$$

$$M^-_{gas} \rightarrow M^-_{solvent} \qquad -\Delta G^{M^-}_H$$

$$M_{solvent} + e^-_{solvent} \rightarrow M^-_{solvent} \qquad \Delta G_{E_{1/2}}$$

The sum of the first four steps leads to the fifth, the desired half reduction potential for a compound of interest in an arbitrary solvent. The change in internal energy for the addition of an electron (step 1) has already been modeled (electron affinity section). Steps 2, 3 and 4 are Henry's

constant calculations for the three species. Steps 1 and 2 are already implemented in SPARC. Step 3 exists in the literature [64] for water, but is not modeled for arbitrary solvents or conditions in the present SPARC calculator. Step 3 was taken as a constant term in the SPARC model and was data-fit for several solvent systems. Step 4 represented a new model for the SPARC system.

This overall new SPARC model was tested against the compendia of measured one electron reduction potentials for clean systems. In these comparisons, the differential sorption terms were ignored and equal access to all the substituents assumed. Of all the solvent systems studied, water was the most problematic. Non-aqueous data in the literature was readily available and reasonably consistent across measurements from several laboratories, whereas reported water measurements often varied by as much as one electron volt. These SPARC models were initially developed and tested using the non-aqueous data. The solvent systems used were picked to represent a wide variety of hydrogen bonding, dipolar and inductive environments. Once the ion-transfer (step 4) models were in place and tested for a large number of molecules in a variety of solvents, our efforts were focused on unraveling the problems with modeling the aqueous reduction system. Steps 1, 2 and 4 were in place and step 3 was well estimated in the literature. The major problem ultimately encountered with the aqueous reduction measurements was the lack of consistency in data reporting, in particular, the reporting of 'effective' reduction potentials that had incorporated into them the effect of pH. pH-dependent SPARC models were then developed and implemented to calculates aqueous reduction potentials as a function of pH. Once these aqueous models were in place, a final refinement of the models was undertaken using both aqueous and non-aqueous reduction data. The refined models are now in place for a limited number of molecular structures (i.e., for only electron-withdrawing groups such as NO_2, C≡N, etc) and available for estimation of one-electron reduction potentials with the SPARC calculator.

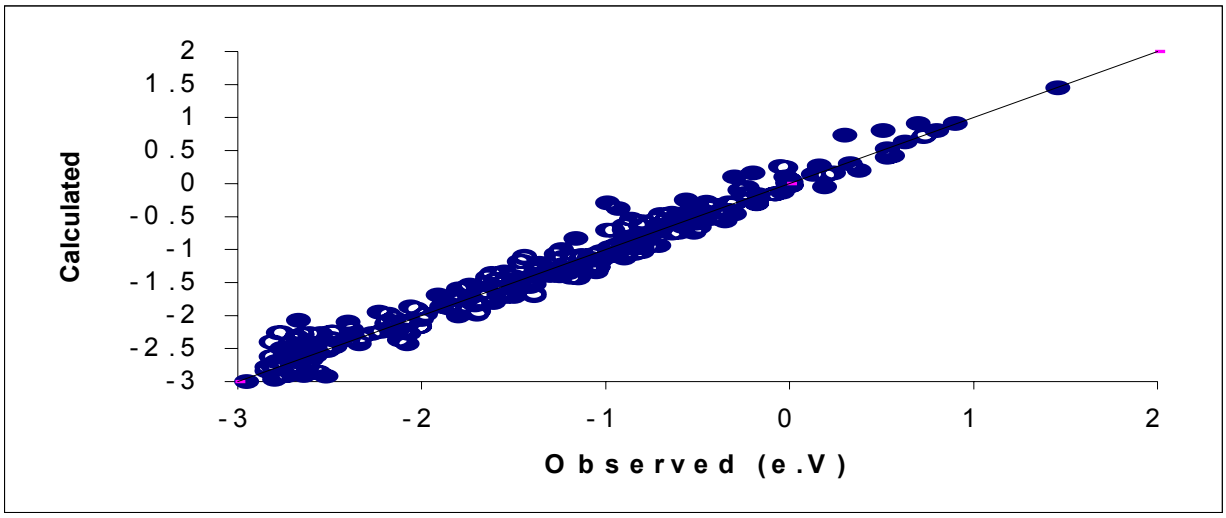


Figure 36. SPARC-calculate versus observed $E_{1/2}$ chemical reduction potential in water, 3 alcohol's, DMF, THF, DMSO, aceteonitrile at 25° C. The RMS was 0.35 e.v.

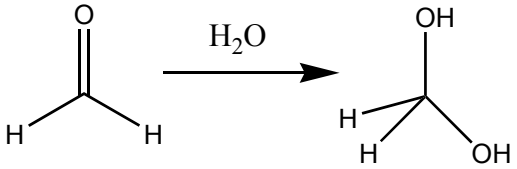
5.4. Chemical Speciation

Complex chemical solutes may exist in solvents in multiple forms or species (ions, zwitterions, tautomers, hydrates) that differ dramatically in their chemical and physical properties. The distribution of a given chemical among its various speciates forms depends on the system conditions (temperature, pH, ionic strength) and medium composition (gas, liquid, solid components). Although much is known about the existence of such species, with the exception of simple ionization, very little data exist for quantifying these speciation processes. This is particularly true for complex (poly-functional) molecules and for aqueous systems. Another difficulty in studying or modeling speciation processes is that they are frequently coupled (e.g., ionization may occur with synchronous tautomerization or hydration). As described herein, SPARC speciation models for ionization are fully developed and tested and models for tautomerization are operational, but only minimally trained and tested at this writing ongoing

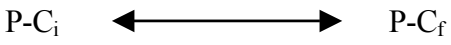
research will complete and integrate these existing models into SPARC and then develop, test, and integrate hydration models.

5.5. Hydration

Hydration in the SPARC context applied herein, is the reversible addition of water across a 'pi-electron functional group'. The two structural units where this is known to occur are the carbonyl and imine functional groups. In each case, a hydroxyl group attaches to the base carbon and a hydrogen atom to the heteroatom.



In the SPARC modeling approach, these functional groups are reaction centers, C, and any molecular structure(s) appended thereto are designated perturber, P, structure.



In the case of hydration, differential solvation of the two species involved will play a major role. In this case we started with the following thermodynamic cycles to model the reaction.

P-C=O (g)	$\rightarrow$ P-$C(OH)_2$ (g)	$\Delta G_{hydration}$(g)	
P-$C(OH)_2$ (g)	$\rightarrow$ P-$C(OH)_2$ (l)	$\Delta G_{transfer}$(O)	Henry's constant
P-C=O (l)	$\rightarrow$ P-C=O (g)	$-\Delta G_{transfer}$(=O)	Henry's constant

---

P-C=O (l) $\rightarrow$ P-$C(OH)_2$ (l) $\Delta G_{hydration}$(l)

The top reaction was modeled using the usual SPARC perturbation approach (see chemical reactivity section) as

$$\Delta G_{hydration} = (\Delta G_{hydration})_c + \delta_p(\Delta G_{hydration})_c$$

where the reaction center ΔG (in this case formaldehyde) is perturbed by appended molecular structure. The perturbation was further factored into mechanistic components such as:

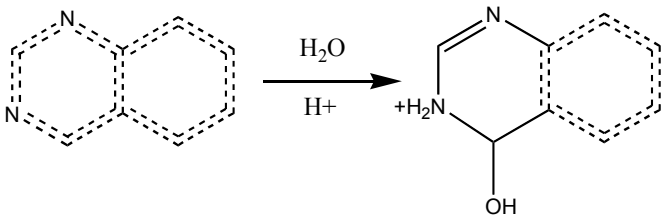
$$\delta_p(\Delta G_{hydration})_c = \delta_{ele}\Delta G_{hydration} + \delta_{res}\Delta G_{hydration} + \delta_{steric}\Delta G_{hydration} + \ldots$$

From structure theory of organic chemistry, it is known that nucleophilic addition reactions across π bonds are sensitive to inductive and steric effects from atoms contiguous to the π group. Also, it is known that functional groups containing non-bonded electrons (-OH, $-OCH_3$, $-NR_2$) attached to the base carbon will prohibit hydration (via induction and resonance). With this model, we can confirm the failure of esters, amides, urea, and carboxylic acids to hydrate, and can project other structures that are readily hydrated. The biggest perturbations were found to be direct field effects (increase), sigma induction (decrease), resonance (decrease) and steric (decrease).

The literature was scoured for measurements of carbonyl hydration. Surprisingly, hydration data for only 37 molecules have been reported in the literature. However, the chemistry represented in these 37 structures was considerably varied, and the data set displayed large effects for all the SPARC mechanistic models. We feel that the basic SPARC models for resonance, field effects (direct and indirect), differential electronegativity and steric environment were well represented by these 37 molecules. The extensive work done in SPARC pK_a and property modeling provided all the needed parameters for the substituents. Although the hydration data set was very limited, only four new parameters were needed to describe the hydration constants for these molecules. These four were the hydration susceptibility of the reaction center to a) resonance effects, b) field effects,

c) steric occlusion and d) the effective electronegativity of the reaction center (C=O) in gauging sigma inductive effects. The available data were least squares fitted using these parameters, and the log hydration constants were estimated to better than 0.3 log pK_a units as shown in Figure 34. This predictability is about as good as the experimental error.

For imines, several compounds are known to readily hydrate, but we have found no measured hydration constants in the literature. Most aliphatic imines readily degrade in water. It is reasonable to assume that hydration may be involved in these processes. Imines that are stable in water are highly sterically hindered, which blocks any potential hydration. Imine-like structures within aromatic rings are known to form stable hydrates. For example, quinazoline (and several quinazoline derivatives) are known to form stable hydrates in the cationic form. This is readily apparent from the increased observed basicities for these compounds. The 'observed pKa's' for these compounds are really mixed constants, representing concurrent protonation and hydration.



The number of quinazoline and pteridine hydration constants found in the literature was much greater than that for the carbonyl. For these aromatic molecules, several researchers employed stopped-flow techniques and pH jump experiments to sort out the individual components in the observed mixed constant pK_a measurements. For example, the pK_a constants for direct protonation of quinazoline was measured to be 1.8, very close to the SPARC calculated value of 1.9. The same approach used in SPARC to model the hydration of C=O was used to model quinazoline and pteridine hydration. These models were further tested by calculating the pKa of the hydrated form

and comparing them to the values inferred from pH jump experiments. For example, the pK_a of the hydrated form of quinazoline was measured to be ~7.5 and the SPARC-calculated pK_a was 7.0. This agreement is good considering the difficulties and assumptions made in both measurement and calculation. Both the measurement and calculation were complicated by the possibility of tautomeric conversion. The observed quinazoline and pteridine hydration constants were compared to the SPARC-calculated values using the models described above. The RMS deviation was 0.43 as shown in Figure 35. Again, the prediction errors are on the order of the experimental error. The hydration models are now fully integrated into the SPARC calculation system and available for use.

5.6. Process Integration

For chemicals that can speciate or exist in multiple forms (ions, zwitterions, tautomers, hydrates), observed chemical behavior may reflect integration over several discrete chemical species or processes. It is convenient to designate as 'macro' macroscopic/observed equilibrium or kinetic constants, and designate as 'micro', a constant for a single species or speciation event (which may or may not be resolved experimentally). As an example in ionization, a micro constant would describe the loss or gain of a proton at a specific site whereas a macro constant may involve poly-protonic events relating to (1) loss or gain of protons from different sites on separate molecules that are integrated in the measurement, or (2) synchronous loss/gain of protons from different sites on the same molecule resulting in one unit change in total charge (e.g., gain of one and loss of two protons). These concepts will be utilized in the following discussion on tautomeric equilibria.

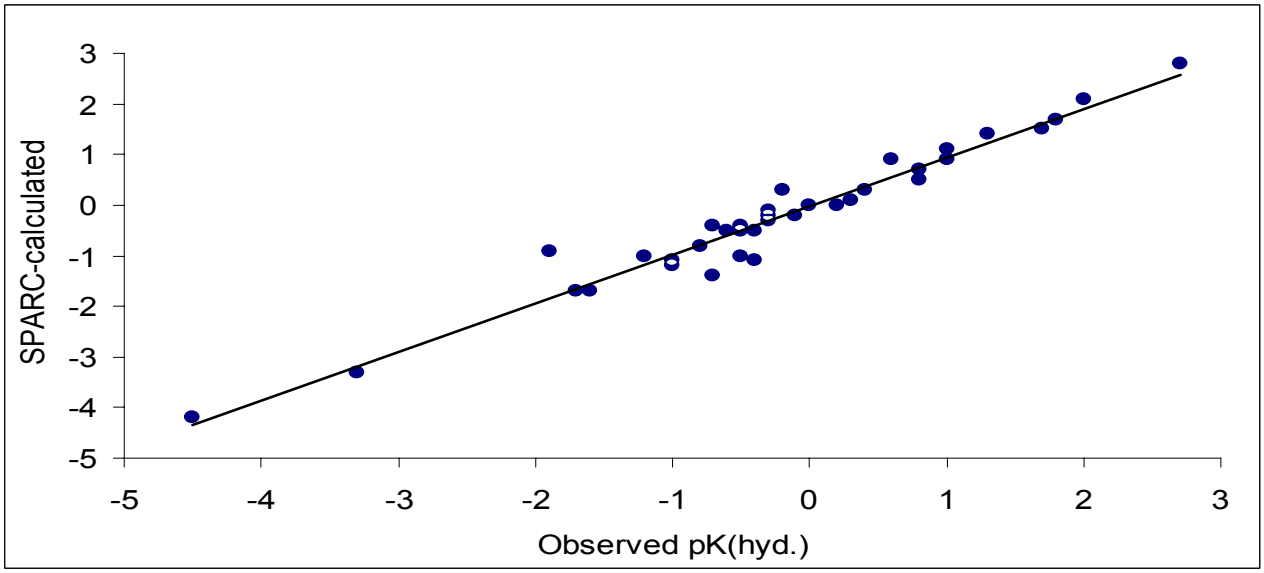


Figure 34. SPARC-calculated versus observed log hydration equilibrium constants for hydration of 36 aldehydes and ketones in water at 25° C. The RMS deviation was 0.3 with an R^2 of 0.95

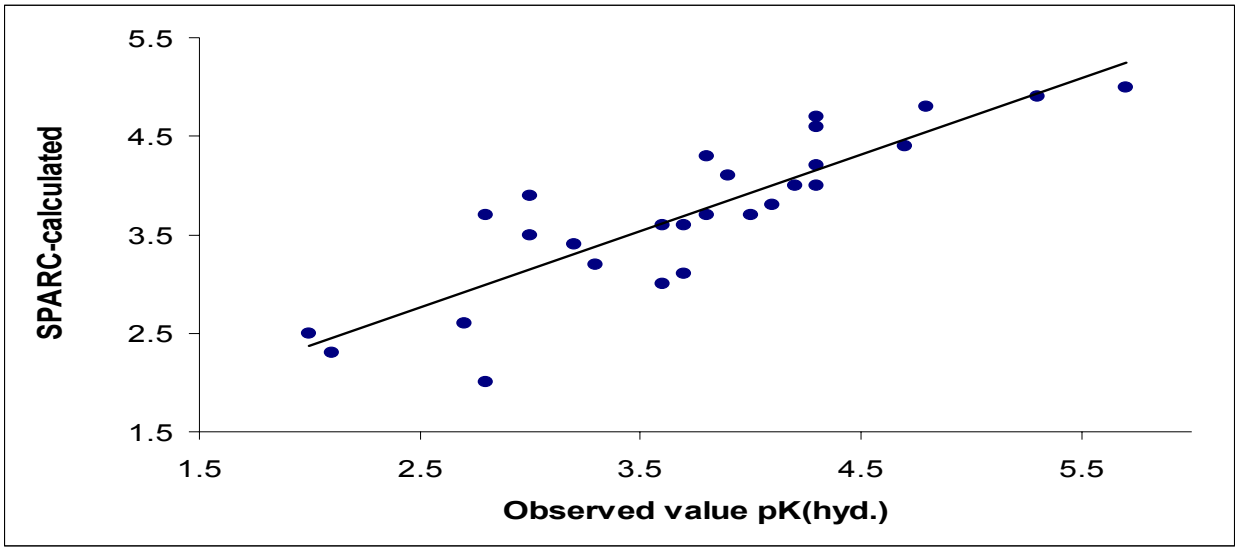


Figure 35. SPARC-calculated versus observed log hydration equilibrium constant for 27 unique quinazolines in water at 25° C. The RMS deviation was 0.43 with an R^2 of 0.774

5.7. Tautomeric Equilibria

For tautomeric processes, a micro constant describes a discrete tautomeric event whereas a macro constant may involve multiple events similar to those described above for ionization or may involve synchronous tautomerization and ionization. The SPARC system must consider all possible

tautomeric and ionization events in order to generate these synchronous processes. In the case of synchronous ionization and tautomerization, the SPARC system first calculates all possible neutral tautomers. This process is currently a rule-driven search for possible tautomeric flips. The current system analyzes the molecule and generates all possible *neutral* tautomers starting with the molecule entered by the user. Once all the possible tautomeric forms have been identified and their SMILES representations generated, the system proceeds to estimate each form's abundance relative to the form entered by the user. The system currently moves 'a bond at a time', and keeps track not only of the final species but the molecular path to get to that form as sequential tautomeric flips occur. The tautomeric calculator uses a combination of the pK_a calculator and the physical property calculator to generate an estimate of the tautomeric equilibrium constant.

The sum of the reactions in Figure 36 leads to the pK_{taut} for a particular bond flip. There may be further tautomerization possible out of this state. An equation similar to that for sequential ionization was developed for all possible neutral tautomeric forms. The k's in the equation for sequential ionization were replaced with $k_{tautomer}$ and the relative abundances of each form calculated. The assigned relative weight for the original starting structure is 1, and to each of the tautomers a weight T_i. In order to capture synchronous ionization and tautomeric events, the system then does a full speciation calculation for each possible neutral tautomer similar to that for ionization previously via the following equation:

$$D_T = \frac{1}{0!} + \frac{\sum_i k_i [H^+]^{L_i}}{1!} + \frac{\sum_i \sum_{j \neq i} k_i k_{ij} [H^+]^{L_{ij}}}{2!} + \ldots + \frac{\sum_i \sum_{j \neq i} \cdots \sum_{k \neq i,j \ldots} k_i k_{ij} \cdots k_{ij \ldots k} [H^+]^{L_{ij \ldots k}}}{N!}$$

where D_{Ti} is the sum of the relative concentrations of all ionized species with the molecular structure of the T_i^{th} tautomer. The fraction of any particular ionization state at a given pH is

expressed as one of the individual terms in the above equation divided the weighted sum of all D_{Ti} givin by:

$$D_{total} = \sum_{i=1}^{\#tautomers} T_i D_{T_i}$$

where T_i is the fraction of the i^{th} tautomer relative to starting structure. For example, the fraction of the starting molecule at a given pH would be simply $1/D_{total}$ and the fraction of the i^{th} tautomer compound having only its j^{th} state ionized would be $T_i \bullet k_j^i[H^+]^{Lj}/D_{total}$ etc. The need to develop intelligent filters were extremely important since the number of calculations grows geometrically with both the number of ionizable sites and the number of possible tautomeric forms. The neutral tautomeric relative concentration cannot be the only factor. For the simple simultaneous ionization/tautomerization scheme shown above, the neutral endo form is predominant (~100/1). In this case, the basicity of the exo form (pK_a ~11) drives the equilibrium and stabilizes the tautomeric form as a cation. The observed apparent K_a (pK_a ~8) is the product of the tautomeric K_{taut} and the pK_a of the tautomeric form. Incorporation of tautomeric re-arrangements is now fully implemented in the SPARC system and is available for use.

5.8. Conclusion

SPARC estimates gas phase and non-aqueous ionization pK_a and $E_{1/2}$ chemical reduction potential (only for electron withdrawing group) in any solvent within experimental error. Hydration and tautomeric constants also can be calculated using the same physical and chemical models. Further testing and refining of the SPARC models for these properties is needed. Integration of speciation, tautomer and hydration models are underway at this time.

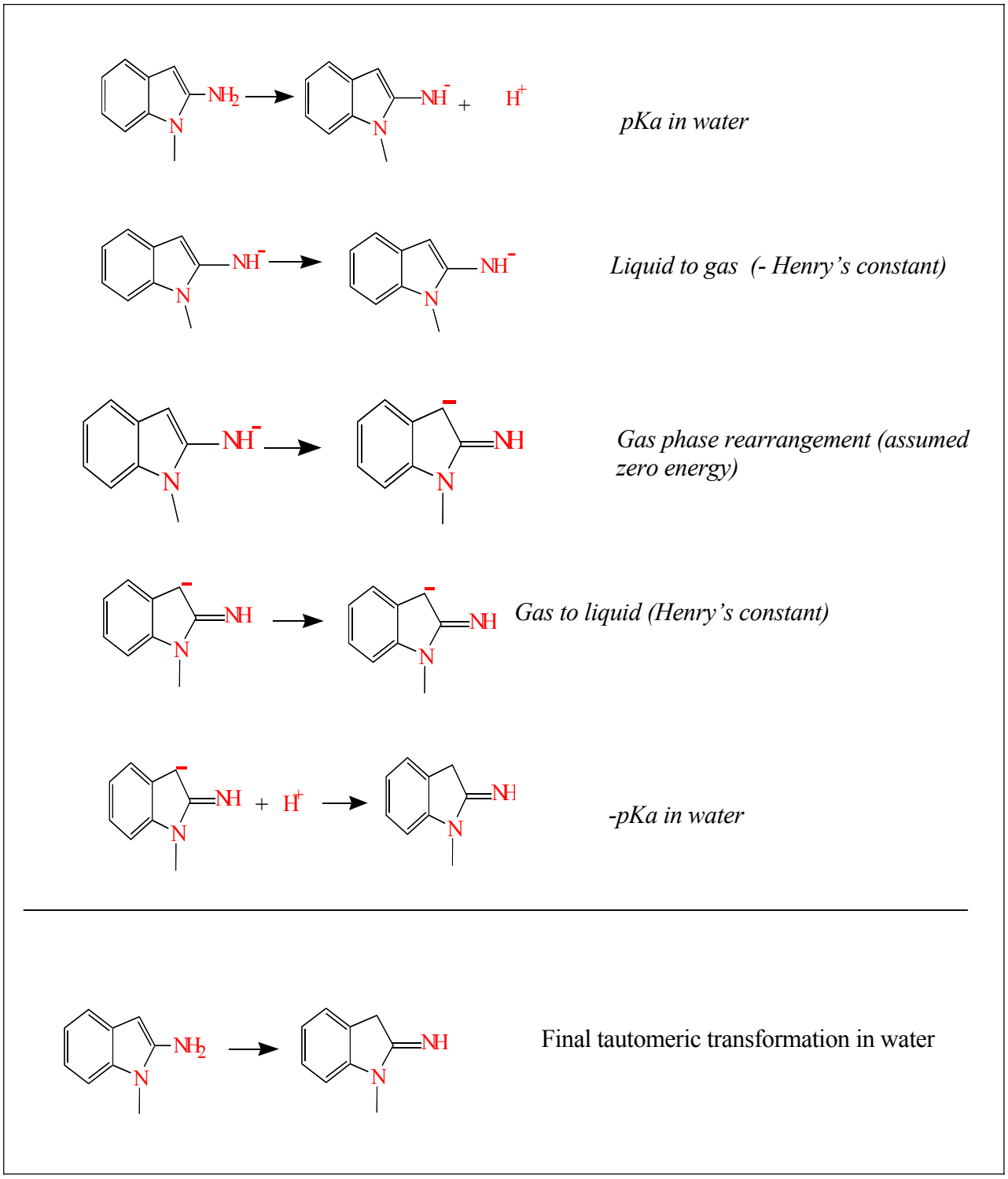


Figure 36. The thermodynamic cycle for the tautomerization of methyl-H-Indol-2-amine

6. MODEL VERIFICATION AND VALIDATION

In chemistry, as with all physical sciences, one can never determine the "validity" of any predictive model with absolute certainty. This is a direct consequence of the empirical nature of science. Because SPARC is expected to predict reaction parameters for processes for which little data exists, "validity" must drive the efficiency of the models constructs in "capturing" or reflecting the existing base of chemical reactivity. 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.

Hence, in SPARC, the 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 SPARC models have been validated on more than 10,000 data points as shown in Table 14.

7. TRAINING AND MODEL PARAMETER INPUT

All quantitative chemical models requires, at some point, calibration or parameterization. The quality of computational output necessarily reflects the quality of the calibration parameters. For this reason, a self-training complement (TRAIN) to SPARC was developed. Although a detailed description of TRAIN will not be given at this time, the following is a general review. For a given set of targeted model parameters, the program takes initial "guesstimates" (and the appropriate boundary constraints) together with a set of designated training data and provides an optimizes set of model parameters. TRAIN cycles once or iteratively through Jacobian optimization procedure that is basically a non-linear, least square matrix method. TRAIN sets up and executes the optimization specifics according to user prescription.

8. 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/chemical properties. QA software compares every single "new" output to the SPARC originally-calculated-value dated back to 1993-1999. 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.

9. SUMMARY

SPARC estimates numerous physical and chemical properties for a wide range of organic compounds strictly from molecular structure. SPARC physical property and chemical reactivity 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. For simple structures SPARC can predict the properties of interest within a factor of 2 or even better. For complicated structures, where hydrogen bond and/or dipole interactions are strong, SPARC can estimate a property of interest within a factor of 3-4 depending on the type of property.

The strength of the SPARC calculator is its ability to estimate the property of interest for almost any molecular structure within an acceptable error, especially for molecules that are difficult to measure. However, the real test of SPARC does not lie in testing the predictive capability for pK_a's, vapor pressure or activity coefficient but is determined by, the extrapolatability of these models to other types of chemistry.

For chemical reactivity models:

The ionization pK_a models in water have been extended to calculate many other properties to include:

1. Estimation of the thermodynamic microscopic ionization constants of molecules with multiple ionization sites, zwttierionic constants and the corresponding complex speciation as a function of pH and the isoelectric points in water.
2. Estimation of gas phase electron affinity.
3. Estimation of ester hydrolysis rate constants as function of solvents and temperature.

For physical property models

The vapor pressure and the activity coefficient models have been extended to calculate many other properties using the solute-solute and solute-solvent models without any modifications to any of these models or any extra parameterization to include:

4. The SPARC self-interactions (solute-solute) model can predict the vapor pressure within experimental error for a wide range of molecular structures over a wide range of measurements. This model has been extensively tested on boiling points, heat of vaporization and diffusion coefficients.
5. The solute/solvent interactions model can predict the activity coefficient within experimental error for a wide range of molecular structures in any solvent. This model was extended and tested on solubilities, partition coefficients (liquid/liquid, liquid/solid, gas/liquid) and GC/LC chromatographic retention times in any single or mixed solvent systems at any temperature.

For Coupled physical property and chemical reactivity models:

Henry's constant for charge and neutral molecules and chemical reactivity models were coupled and extended to calculate many other properties:

6. Ionization pK_a in the gas phase and in non-aqueous solutions.
7. Thermodynamic microscopic ionization, zwitterionic, hydration, and tautomeric equilibrium constant in water or any other solvent.
8. $E_{1/2}$ chemical reduction in water and in many other solvent systems.

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

10. 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. M. M. Miller, S. P. Wasik, G. L Huang, W. Y. Shiu and D. Mackay. *Environ. Sci. & Technol*. 19 522 1985

3. W. J. Doucette and A.W. Andres. *Environ. Sci. Technol.* 21 821 1987

4. R. F. Rekker, *The Hydrophobic Fragment Constant,* Elsevier, Amsterdam, Netherlands 1977.

5. S. Banerjee, S. H. Yalkowski and S. C. Valvani. *Environ. Sci. and Toxicol.* 14 1227 1980.

6. M. J. Kamlet, R. M. Dougherty, V. M. Abboud, M. H. Abraham and R.W. Taft. J. Pharm. Sci. 75(4) 338 1986

7. W. J. Lyman, W. E. Reehl and D. H. Rosenblatt. *Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Chemicals.* McGraw-Elill, NewYork, NY. 1982.

8. G. Shuurmann. *Quant. Struct. Act Relat.* 9 326 1990.

9. A. J. Leo. *Structure Activity Correlations in Studies of Toxicity and Bio-concentrations with Aquatic Organisms*. (Veith,G.D.,ed.), International Joint Commission, Windsor, Ontario 1975.

10. D. MacKay, A. Bobra, W. Y. Shiu and S. H. Yalkowski. *Chemosphere,* 9 701 1980.

11. R. G. Zepp. *Handbook of Environmental Chemistry* (Hutzinger,O.,ed.), vol.2(B) Springer-Verlag, New York, NY, 1982.

12. R. G. Zepp and D. M. Cline. *Environ Sci & Techno* 11 359 1977.

13. N. L. Wolfe, R.G. Zepp, J. A. Gordon, G. L. Baughman and D. M. Cline. *Environ Sci & Tecno* 11 88 1977.

14. J. L. Smith, W. R. Mabey, N. Bohanes, B. B. Hold, S.S. Lee, T.W. Chou, D.C. Bomberger and T. Mill. *Environmental Pathways of Selected Chemicals in Fresh Water Systems: Part 11,* U.S. Environmental Protection Agency, Athens, GA 1978.

15. H. Drossman, H. Johnson and T. Mill. *Chemoshpere* 17(8) 1509 1987.

16. 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.

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

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

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

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

21. 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, 1994.

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

23. S. H. Hilal, L. A. Carreira, C. M. Melton and S. W. Karickhoff., *J. Chromatogr.,* 269 662 1994.

24. S. H. Hilal, L. A. Carreira and S. W. Karickhoff, *Quant. Struct. Act. Relat., In Press.*

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

26. S. H. Hilal, L. A. Carreira, and S. W. Karickhoff., *To be submitted*

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

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

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

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

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

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

33. C. Lim, D. Bashford, and M. Karplus, *J. Phys. Chem.*, 95 5610 1991.

34 W. L. Jorgensen and J. M. Briggs, *J. Am. Chem Soc.,* 111 4190 1989.

35. C. Grüber, and V. Buss, *Chemosphere*, 19 1595 1989.

36. K. Ohta., *Bull. Chem. Soc. jpn*., 65 2543 1992.

37. G. Klopman, *Quant. Struct. Act. Relat.,* 11 176 1993.

38. G. Klopman, D. Fercu, *J. Comp Chem*., 15 1041 1994.

39. A. E. Martell and R. J. Motekaities, *The determination and use of stability constants*, Weinhiem, New York, VCH publisher, 1988.

40. R. E. Benesch and R. Benesch, *J. Am. Chem. Soc.*, 5877 77 1955.

41. M. A. Grafius and J. B. Neilands , *J. Am. Chem. Soc.*, 3389 77 1955.

42. R. B. Martin, *J. Phys. Chem.*, 2657 75 1971.

43. R. B. Martin, J. T. Edsall, D. B. Wetlaufer and B.R. Hollingworth, *J. Biol. Chem.*, 1429 233 1958.

44. T. L. Hill, *J. Phys. Chem.*, 101 48 1944.

45. J. T. Edsall and J. Wyman, J., *Biophysical Chemistry, Vol. 1,* Academic Press Inc. New York, 1958.

46. A. L. McClellan*, Tables of Experimental Dipole Moments*. W.H. Freeman and Co., London, 1963.

47. J. Dykyj, M. Repas and J. Anmd Svoboda. *Vapor Pressure of Organic Substances.* VEDA, Vydavatel' stvo, Slovenskej Akademie Vied, Bratislava, 1984.

48. K. A. Sharp, A. Nicholls, R. Friedman and B. Honig, *Biochemistry*, 30 9686 1991.

49. P. J. Flory, Chem. Phys., 10 51 1942.

50. M. L. Huggins, J. Am. Chem. Soc., 64 1712 (1942)

51. R. C. Reid, J. M. Prausnitz and J. K. Sherwood, The Properties of Gasses and Liquids, 3ed ., McGraw-hill book Co.,1977

52. G. Tarjan, I. Timar, J. M. Takacs, S. Y. Meszaros, Sz. Nyiredy, M. V. Budahegyl, E. R. Lombosi and T. S. Lombosi., *J. Chromatogr.*, 271 213 (1982)

53. J. K. Haken and M. B. Evans., *J. Chromatogr*., 472 93 (1989).

54. E.sz. Kov'ats, *Adv. Chrommatogr*., 1 31A (1965).

55. L. S. Anker and P. C. Jurs., *Anal. Chem*., 62 2676 (1990).

56. G. Schomburg and G. Dielmann., *J. Chromatogr. Sci*. 11 151 (1973).

57. N. Dimov, *J. Chromatogr.*, 347 366 (1985).

58. D. Papazova and N. Dimov, *J. Chromatogr.,* 356 320 (1986).

59. C. R. Wilke and C. Y. Lee, *Ind. Eng. Chem.*, 47 1253 (1955)

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

61. R. A. Larson and E.J. Weber, *Reaction mechanisms in Environmental Organic Chemistry*"., Chelsea, MI, Lewis Publishers 1994.

62. E. J. Weber and R. L Adams, *Environ. Sci. Technol*., 29 1163 1995.

63. T. M. Vogel, C. S. Criddle and P.L. McCarty, *Environ. Sci. Technol.* 21 (8) 722 1987.

64. J. Saveant, Adv. Phys. Org., 26 1 1990.

65. S. W. Karickhoff and J. MacArthur long, *US EPA, Internal Report,* Athens, GA.

66. Haim Shalev and Dennis Evans, *J. Am. Chem. Soc.,* 111 7 1989.

11. GLOSSARY

1. SPARC = SPARC Performs Automated Reasoning in Chemistry
2. EA = Electron Affinity
3. SAR = Structure Activity Relationships
4. QSAR = Quantitative structure Activity Relationship
5. LFER = Linear Free Energy Relationships
6. PMO = Perturbed Molecular Orbital Theory
7. IUPAC = International Union of Pure and Applied Chemistry
8. MO = Molecular Orbital
9. LUMO = Lowest Unoccupied Molecular Orbital
10. HOMO = Highest Occupied Molecular Orbital
11. NBMO = Non Bonded Molecular Orbital
12. C = Reaction Center
13. P = Perturber
14. S = Substituent
15. R = Molecular conductor connecting S to C
16. R_{π} = A rigid fully conjugated π structure (such as benzene)
17. C_i = Initial state
18. C_f = Final state
19. Δq_c = Fraction of NBMO charge
20. v = Solid angle occluded by P
21. S_i = Reduction factor for steric blockage
22. A_c, B_c = Entropic and the enthalpic van't Hoff coefficients of C, respectively.

23. $(pK_a)_c$, EA_c = pK_a and EA of the reaction center (reference point), respectively.

24. $\delta_p(pK_a)_c$, $\delta_p(EA)_c$ = Change in the pK_a and EA due to P, respectively.

25. k_{zw} = Zwitterionic ionization constant

26. K, k = Macroscopic and microscopic equilibrium constants, respectively

27. D_i = Fraction of the i^{th} microscopic species

28. D_T = Sum of the relative concentration of all the ionizes species

29. T_i = Fraction of the i^{th} tautomer species

30. N = Number of the ionizable sites in a molecule

31. N_i = Number of the electrons in fragment i

32. NI = Total number of the microconstants

33. IS = Number (≤N) of sites that are ionized

34. pH_I = Isoelectric point

35. A = log of the pre-exponential factor,

36. T_k = Temperature in Kelvin,

37. Ref_1, Ref_2 = Entropic and enthalpic contribution to the rate, respectively

38. log k_c = Hydrolysis behavior of the reaction center "reference rate"

39. δ_{IP} log k_c = Change in the hydrolysis behavior brought about by the perturber structure.

40. δ_{EP} log k_c = Change in the solvation of C_i vs the transition state due to H-bond and field stabilization effects of the solvent.

41. F_s, F_q, F_μ = Substituent field strength, charge strength and dipole strength, respectively. F_s = F_μ

42. M_F = Substituent mesomeric strength

43. χ = Electronegativity

44. E_r = Substituent resonance strength

45. α = Proton donating site

46. β = Proton accepting site

47. NB = Data-fitted parameter that depends on number of the substituents that are bonded directly to the reaction center for sigma induction

48. $\bar{\alpha}_I$ = Average molecular polarizability

49. P_i^d = Effective polarizability density of molecule i

50. D^d_i = Effective dipole density of molecule i

51. ρ_i = Susceptibility to a mechanistic mechanism

52. V_i = Molar volume

53. μ_i = Effective microscopic dipole

54. A_{disp} = Polarizability adjustment for dispersion

55. A_{ind} = Polarizability adjustment for induction

56. CN = Carbon number

57. KI = Kov'ats index

58. UI_o = Kov'ats index at 0^o C

59. m_i = Monopole density

60. R_m = Ratio of the molecularites of the two phases

61. RMS = Root Mean Square

12. 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