JOURNAL OF PHYSICAL ORGANIC CHEMISTRY, VOL. 7, 122–141 (1994)

# ESTIMATION OF IONIZATION CONSTANTS OF AZO DYES AND RELATED AROMATIC AMINES: ENVIRONMENTAL IMPLICATION

S. H. HILAL AND L. A. CARREIRA*

*Department of Chemistry, University of Georgia, Athens, Georgia 30602, U.S.A.*

G. L. BAUGHMAN AND S. W. KARICKHOFF

*Environmental Research Laboratory, US Environmental Protection Agency, Athens, Georgia 30613, U.S.A.*

AND

C. M. MELTON

*Artificial Intelligence Group, University of Georgia, Athens, Georgia 30602, U.S.A.*

**Ionization constants for 214 dye molecules were calculated from molecular structures using the chemical reactivity models developed in SPARC (SPARC Performs Automated Reasoning in Chemistry). These models used fundamental chemical structure theory to predict chemical reactivities for a wide range of organic molecules from molecular structure. The energy differences between the protonated state and the unprotonated state for a molecule of interest are factored into mechanistic components including the electrostatic and resonance contributions and any additional contributions to these energy differences. The RMS deviation was found to be less than 0·62 p$K_a$ units, which is similar to the experimental error.**

## INTRODUCTION

In recent years, the need for physico-chemical constants of chemical compounds has greatly accelerated in both industry and government. The impetus for this is the high cost of laboratory measurements and the need to examine the behavior of large numbers of diverse compounds. Among the latter is a requirement, under the US Toxic Substances Control Act, for environmental assessment of all new chemicals that are to be manufactured or used in the USA.

This situation has resulted in the development and widespread use of linear free energy relationships (LFER), structure–activity relationships (SAR) and other estimation methods particularly in the drug and environmental fields. Even so, mythologies and values are often not available for those parameters needed in the sophisticated mathematical models used for environmental exposure assessment.

Such is the case for upwards of 10% of dyes for which pre-market notifications (PMN) are received for review by the US Environmental Protection Agency (EPA).[1] Recently, the EPA has developed a computational procedure that is based on the use of artificial intelligence techniques to combine the results of both fundamental and empirical approaches much as a very knowledgeable chemist might.[2] The purpose of this paper is to demonstrate the utility of this procedure.

The p$K_a$ of an organic compound is vital to environmental exposure assessment because it can be used to define the degree of ionization and the propensity for sorption to soil and sediment by cation exchange. These processes, in turn, can determine mobility, reaction kinetics, bioavailability, complexation, etc.

Unfortunately, up to now no reliable method has been available for predicting p$K_a$ values over a wide range of molecular structures either for simple compounds or for complicated molecules such as dyes, at a level of accuracy that is within the experimental error. The object of this study was to demonstrate the application of SPARC (SPARC Performs Automated Reasoning in Chemistry)[2] to the prediction of p$K_a$ values for complex azo dyes and related aromatic amines that may be of environmental significance.

This new computer program (SPARC) will cost the user only a few minutes of computer time and will

* Author for correspondence.

CCC 0894–3230/94/030122–20


*Received 20 April 1993*
*Revised 11 October 1993*

provide 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 of interest to predict its $pK_a$. The user provides the program with the molecular structure either by direct entry as SMILES (Simplified Molecular Input Line Entry System) notation or via the molecular editor that will generate the structure and translate it into SMILES notation.

Dyes were chosen for this study for several reasons: (1) they are a severe test case; (2) a large number of new chemicals (PMN requests) are dyes; (3) many dyes, especially azo dyes, and their environmental transformation products are aromatic amines[3-5] and thus are of potential toxicological concern; (4) data on most new dyes, and also their products and precursor amines, are either unavailable or unmeasurable because of the solubility limitation; and (5) sufficient data are available to provide a comparison between measured and computed values.

## SPARC COMPUTATIONAL APPROACH

SPARC is a prototype computer program being developed to predict chemical reactivity and physical properties for a large number of organic molecules based on fundamental chemical structure theory. At the present stage of development, SPARC predicts ionization $pK_a$ values, electron affinities and numerous physical properties such as distribution coefficients, solubilities and vapour pressures.

The approach of SPARC is not to do 'first principles' computation; rather, it analyzes chemical structure relative to a specific reactivity query much as an expert chemist might. Hence, SPARC computation methods directly utilize the extensive knowledge base of organic chemistry. Organic chemists have established the types of structural groups or atomic arrays that impart certain types of reactivity and have described, in 'mechanistic' terms, the effects on reactivity of other structural constituents appended to the site of reaction.

The computational approaches in SPARC also blend conventional LFER,[6] SAR and Perturbed Molecular Orbital (PMO) methods.[7] In general, SPARC utilizes LFER to compute thermodynamic or thermal properties and PMO theory to describe quantum effects such as delocalization energies or polarizabilities of $\pi$ electrons. In reality, every chemical property involves both quantum and thermal contributions and necessarily requires the use of both perturbation methods for prediction.

For any chemical property addressed in SPARC, the energy differences between the initial state and the final state are small compared with the total binding energy of the reactant 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 extreme accuracy and with more computational simplicity and flexibility than *ab initio* methods. These 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 or solvation. Molecular structures are broken up 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 $pK_a$ 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.

## $pK_a$ COMPUTATIONAL PROCEDURE

SPARC computation begins by locating the reaction center within the molecule and the perturber. The perturber structure is assumed to be unchanged in the reaction. Like all chemical reactivity parameters addressed in SPARC, $pK_a$ is analyzed in terms of some critical equilibrium component:

$$P\text{–}C_i \xrightarrow{pK_a} P\text{–}C_f$$

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:

$$pK_a = (pK_a)_c + \delta_p(pK_a)_c \quad (1)$$

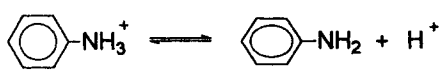
where $(pK_a)_c$ describes the ionization behavior of the reaction center and $\delta_p(pK_a)_c$ is the change in ionization behavior brought about by the perturber structure. SPARC computes reactivity perturbations, $\delta_p(pK_a)_c$, that are then used to 'correct' the ionization behavior of the reaction center for the compound in question in terms of potential 'mechanisms' for interaction of P and C as

$$\delta_p(pK_a)_c = \delta_{ele}pK_a + \delta_{res}pK_a + \delta_{sol}pK_a + \cdots \quad (2)$$

where $\delta_{res}pK_a$, $\delta_{ele}pK_a$ and $\delta_{sol}pK_a$ describe the differential resonance, electrostatic and solvation effects of P with 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 $\pi$ electrons of the two states due to P. This delocalization of $\pi$ 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 stearic blockage of solvent access to C.

In the ionization of aniline, $NR_2$ is the reaction center (denoted C) and the phenyl group is the perturber (denoted P):



The ionization equilibrium constant can be expressed as

$$pK_a = (pK_a)_c + \delta_{res}pK_a \quad (3)$$

where $(pK_a)_c$ is the $pK_a$ for the reaction center $NR_2$ and is equal to 8·93, and $\delta_{res}pK_a$ is the resonance contributions to $pK_a$.

Resonance effects models were developed and calibrated using light absorption spectra,[2] whereas electrostatic effects models were developed and calibrated using ionization $pK_a$ values.

## SPARC MODELING APPROACH

The modeling of the perturber effects for $pK_a$ relates to the structural representation S—$_iR_j$—C, where S—$_iR_j$ 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 from the reaction center. R links the substituent and reaction center and serves as a conductor of the perturbation ('conducts' resonant $\pi$ electrons or electric fields). The $i$ and $j$ denote anchor atoms in R for S and C, respectively. Perturbations are factored into three independent components for the structural components C, S and R: (1) substituent strength, which describes the potential of a particular substituent to 'exert' a given effect, (2) molecular network conduction, which describes the 'conduction' properties of the molecular structure R, connecting S to C with regard to a given effect; and (3) reaction center susceptibility, which rates the response of the reaction center to the effect in question.

The contributions of each structural component are quantified (i.e. parameterized independently). For example, the strength of the substituent's electrostatic field effect depends only on the substituent; likewise, the conduction of R is modeled to be independent of the specific identities of both the substituent and the reaction center. The susceptibility of C to the field effect quantifies the differential interaction of the initial state versus the final state with the electric field, but again this susceptibility gauges only the initial state versus the final state of the reaction center and is independent of both R and S. The rationale for the factoring is to remove, to the extent possible, both structural and reaction specificity from effects parameterization. This provides parameter 'portability' and, hence, effects-model portability to other structures and to other types of reactivity.

## ELECTROSTATIC EFFECTS MODELS

Electrostatic effects on $pK_a$ derive from charges or electric dipoles in the appended perturber structure, P, interacting through space with charges or dipole 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 'induce' electric fields in the 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 $\sigma$-bonds between atoms. For compounds containing multiple substituents, electrostatic perturbations are computed for each singly and summed to produce the total effect.

### Field effects model

The field effect is expressed as a multiple expansion. For a dipolar substituent, the field effect may be expressed as

$$\delta_{field}(pK_a)_c = \frac{q_c\mu_s \cos\theta_{cs}}{D_e r_{cs}^2} \quad (4a)$$

where $q_c$ is the change in charge on the reaction center, $\mu_s$ is the local dipole of the substituent, $\theta_{cs}$ is the angle the dipole subtends to the reaction center, $D_e$ is the effective dielectric constant for the medium and $r_{cs}$ is the distance from the substituent dipole center to the reaction center. If the substituent has a charge, $q_s$, then the corresponding equation becomes

$$\delta_{field}(pK_a) = \frac{q_c q_s}{D_e r_{cs}} \quad (4b)$$

Once again, in order to provide parameter 'portability' and, hence, effect-model portability to other structures and to other types of chemical reactivity, the contribution of each structural component is quantified (i.e. parameterized) independently:

$$\delta_{field}(pK_a)_c = \rho_{ele}\sigma_p = \rho_{ele}\sigma_{cs}F_s \quad (5)$$

where $\sigma_p$ characterizes the field strength that the perturber exerts on the reaction center; $\rho_{ele}$ is the susceptibility of a given reaction center to electric field effects that describe the electrostatic charge accompanying the reaction, and is presumed to be independent of the perturber. The perturber potential, $\sigma_p$, 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, $\sigma_{cs}$, of the intervening molecular network for electrostatic interactions. For molecules containing multiple substituents, the substituent field effects are computed for each substituent and summed to produce the total effect as

$$\delta_{field}(pK_a)_c = \rho_{ele} \sum \sigma_{cs} F_s \quad (6)$$

The electrostatic susceptibility, $\rho_{ele}$, is a data-fitted parameter inferred directly from measured $pK_a$ values. This parameter is determined once for each reaction center and stored in the SPARC database. In parameterizing the electrostatic field effects models, the ionization of the carboxylic acid group is chosen to be the reference reaction center with an assigned $\rho_{ele} = 1$. For all the reaction centers addressed in SPARC, electrostatic interactions are calculated relative to a fixed geometric reference point that is chosen to approximate the center of charge for the carboxylate anion, $r_{cj} = 1 \cdot 3$ unit, where the length unit is the aromatic carbon–carbon length (1·40 Å). The $\rho_{ele}$ for other reaction centers reflects electric field changes for these reactions gauged relative to the reference reaction center.

With regard to the substituent parameters, each uncharged substituent has one field strength parameter, $F_\mu$, characterizing the dipole field strength, whereas a charged substituent has two, $F_q$ and $F_\mu$. $F_q$ characterizes the effective charge on the substituent and $F_\mu$ 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 non-carbon atom, then $F_\mu$ 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 $F_s$.

The distance between the reaction center and the substituent, $r_{cs}$, for both charges and dipoles is computed as a summation of the respective distance contributions of C, R and S as

$$r^0_{cs} = r_{cj} + r_{ij} + r_{is} \quad (7)$$

This zero-order distance is adjusted for ring systems to correct for electric field interactions through space and those involving either S or C units. These adjustments are significant only when C and S are *ortho* to each other:

$$r_{cs} = A r^0_{cs} \quad (8)$$

where $A$ is an adjustment constant and is assumed to depend only on bond connectivity into and out of the R–$\pi$ unit (e.g. points $i$ and $j$). For R–$\pi$ units recognized by SPARC, $A$ factors for each pair $(i,j)$ are empirically determined from data (or inferred from structural similarity to other R–$\pi$ units) as shown in Table 1. 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 1·0 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 1.

The dipole orientation factors, $\cos \Theta_{ij}$, are at present ignored (set to 1·0) except in those cases where S and C are attached to the same rigid R–$\pi$ unit. In these situations, they are assumed to depend solely on the point(s) of attachment, $(i,j)$, and are pre-calculated and stored in SPARC databases.

### Mesomeric field effects

The electric field derived from substituent-induced polarization of $\pi$ electrons is termed the mesomeric field. This field will result in an indirect interaction between the induced charges in $R_\pi$ with charges or dipoles in the reaction center.

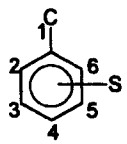
The contribution of the mesomeric field can be estimated as a collection of discrete charges, $q_R$ with the contribution of each described by equation (4). As is the case in modeling the direct field effects, the mesomeric effect components are resolved into three independent components, S, R, and C, and as

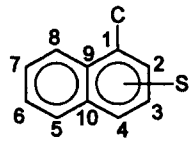
$$\delta_{M_F}(\Delta E)_c = \rho_{ele} q_R M_F \quad (9)$$

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

In modeling the mesomeric field effect, the intensity and the location of charges in R depend on both the substituent and the $R_\pi$ network involved. The contributions of S and $R_\pi$ are resolved by replacing the reaction center with the surrogate electron donor $CH_2^-$ as a non-bonded molecular orbital (NBMO) charge source. The NBMO charge distribution from this surrogate donor is calculated from PMO theory.[2,7] The mesomeric substituent strength parameter describes the $\pi$-induction ability of a particular substituent relative to the $CH_2^-$. The magnitude of a given $M_F$ parameter describes the relative field strength, whereas the sign of the parameter specifies the positive or negative

Table 1. Geometry parameterization for selected ring systems




| Molecule | Position on ring | | Geometry parameters | | |
|---|---|---|---|---|---|
| | 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 | 1·0 | 1·0 |
| 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 |
| | 2 | 8 | 2·7 | 0·64 | 0·81 |

character of the induced charge in $R_\pi$. For $pK_a$, the mesomeric field effect for a given substituent is given by

$$\delta_{M_F}(pK_a)_c = \rho_{ele} M_F \sum_k \frac{q_{ik}}{r_{kc}} \quad (10)$$

where $q_{ik}$ is the charge inducted at atom $k$, with the reference probe attached at atom $i$ calculated from on PMO theory.[2-7] $r_{kc}$ is the through-space distance to the reaction center as described previously for direct field.

### 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. The effect is believed to be transmitted progressively between atoms. The substituent electronegativity effect acts importantly only at the atom to which the substituent is attached and any effect beyond the second atom is negligible.

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 of S and C:

$$\delta_{sig}(pK_a)_c = \rho_{ele} \sum [\chi_c - \chi_s] \quad (11)$$

where $\rho_{ele}$ is the susceptibility of a given reaction center to electric field effects. Once again, because the reaction center does not discriminate the source of the electric fields, $\rho_{ele}$ is the same as described for the field effect; $\chi_c$ is the effective electronegativity of the reaction center and $\chi_s$ is the effective electronegativity of the substituent.

## RESONANCE EFFECTS MODEL

Resonance stabilization energy in SPARC is a differential quantity, related directly to the extent of electron delocalization in the neutral state versus the ionizable 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 with the case of electrostatic perturbations, structural units are classified according to function. Substituents that withdraw electrons from a reference point are designated S+ while electron donating groups are designated S−. The R–$\pi$ units withdraw or donate electrons, or serve as a 'conductor'

of $\pi$ electrons between resonant units. Reaction centers are likewise classified as C+ and C−, denoting withdrawal and donation of electrons, respectively.

In SPARC, the resonance interactions describe the delocalization of an NBMO out of $C_i$ or $C_f$ into a contiguous R−$\pi$ or a conjugated S+ substituent. To model this effect, the reaction center is replaced by a surrogate electron donor, $CH_2^-$. 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

$$\delta_{res}(pK_a)_c = \rho_{res}(\Delta q)_c \quad (12)$$

where $(\Delta q)_c$ is the fractional loss of NBMO charge from the surrogate reaction center calculated based on PMO theory; $\rho_{res}$ is the susceptibility of a given reaction center to resonance interactions, and quantifies the differential 'donor' ability of the two states of the reaction center relative to the reference donor $CH_2^-$.

## SOLVATION EFFECTS MODEL

For acid–base ionization equilibria in aqueous solutions, $C_i$ and $C_f$ frequently differ substantially in degree of solvation, with the more highly charged moiety solvating more strongly. Thus steric blockage of the reaction center is distinguished from the steric-induced twisting of the reaction center incorporated in electron delocalization interactions. 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.

To model this effect, differential solvation of the reaction center is incorporated in $(pK_a)_c$, $\rho_{res}$ and $\rho_{ele}$. If the reaction center is bonded directly to more than one hydrophobic group (e.g. alkane or aromatic systems) or if the reaction center is *ortho* to an aromatic bridge, then $\delta_{sol}(pK_a)$ must be calculated. The $\delta_{sol}(pK_a)$ 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 *ortho* to the reaction center.

## INTRAMOLECULAR H-BONDING EFFECTS MODEL

Reaction centers that are *ortho* or *peri* to substituents in ring systems might interact with those substituents through intramolecular H-bonding and thus affect the $pK_a$. For each reaction center that is *ortho* or *peri* to a substituent, SPARC calculates the H-bonding contributions for each reaction center with each substituent $\delta_{H\text{-}B}(pK_a)$. $\delta_{H\text{-}B}(pK_a)$ describes the H-bonding differences of the initial state versus the final state of a reaction center with a substituent. For reaction centers that might H-bond with more than one substituent, the H-bonding contribution for each substituent is calculated and the stronger contributor to this effect is selected.

## 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, but the molecule has $n$ such sites, then

$$\delta_{stat}(pK_a)_c = \log(n_a/n_b) \quad (13)$$

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

## RESULTS AND DISCUSSION

Figure 1 shows SPARC-calculated versus observed values of ionization equilibrium constants in water at 25 °C for $CO_2H$, $AsO_2H$, $PO_2H$, $BO_2H_2$, $SeO_3H$, OH and SH as acid reaction centers and $NR_2$, aromatic in-ring N and =N as base reaction centers. The $pK_a$ calculator was first parameterized (trained) using measured ionization constant for more than 775 compounds. The root mean square (RMS) deviation for the set was found to be equal to 0·22 $pK_a$ units (for $pK_a$ sample calculations and performance of the SPARC $pK_a$ calculator for simple molecules, see Ref. 2). The reaction center $pK_a$s $(pK_a)_c$ for $CO_2H$, OH, SH and $NR_2$ were measured values, whereas the rest of the reaction centers were trained values inferred directly from $pK_a$ measurements and stored in SPARC database. Tables 2 and 3 show substituent and reaction center parameters, respectively.

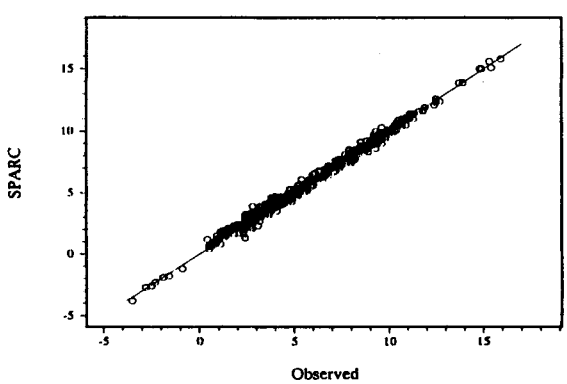



Figure 1. Observed versus SPARC-calculated $pK_a$ values for IUPAC organic compounds

Table 2. p$K_a$ substituent characteristic parameters[a]

| Species | $F_\mu$ | $F_q$ | $M_F$ | $E_r$ | $\chi_s$ |
|---|---|---|---|---|---|
| $CO_2H$ | 1·524 | 0·000 | 1·077 | 0·073 | 3·21 |
| $CO_2^-$ | 0·900 | −1·030 | 4·723 | 0·800 | 2·85 |
| $PO_2H_2$ | 1·100 | 0·000 | 0·700 | 0·080 | 2·70 |
| $BO_2H_2$ | 1·686 | 0·000 | 1·500 | 0·000 | 2·40 |
| $SO_3^-$ | 5·037 | −0·544 | 3·752 | 2·040 | 3·20 |
| OH | 1·448 | 0·000 | −4·712 | 14·97 | 4·87 |
| SH | 5·476 | 0·000 | −0·873 | 12·00 | 2·76 |
| $O^-$ | 5·584 | −3·064 | −3·673 | 7·577 | 3·10 |
| $S^-$ | 6·482 | −2·882 | −1·418 | 10·38 | 3·34 |
| $NR_2$ | 1·060 | 0·000 | −5·852 | 27·47 | 2·62 |
| $NR_2H^+$ | 6·543 | 0·176 | −1·272 | 15·00 | 3·80 |
| CH3 | 0·000 | 0·000 | −1·912 | 0·129 | 2·30 |
| $NO_2$ | 8·305 | 0·000 | 1·992 | 2·330 | 2·10 |
| C≡N | 7·056 | 0·000 | 1·445 | 2·418 | 3·09 |
| OR | 1·897 | 0·000 | −2·985 | 5·637 | 2·99 |
| SR | 2·007 | 0·000 | −0·830 | 3·094 | 2·80 |
| I | 3·924 | 0·000 | 0·000 | 4·928 | 3·12 |
| Br | 4·100 | 0·000 | −0·050 | 3·012 | 3·46 |
| Cl | 4·070 | 0·000 | −0·332 | 1·498 | 3·64 |
| F | 4·100 | 0·000 | −0·834 | 0·800 | 3·75 |
| in-ring N | 6·468 | 0·000 | 0·775 | 2·080 | — |
| in$H^+$-ring N | 6·520 | 3·156 | 4·200 | 9·007 | 3·80 |
| $SO_2$ | 7·116 | 0·000 | 2·779 | 3·547 | 3·60 |
| =N | 6·068 | 0·000 | 2·101 | 0·098 | — |
| $=NH^+$ | 0·600 | 1·000 | 8·800 | 4·600 | — |
| =O | 4·973 | 0·000 | 4·000 | 2·339 | — |
| PO | 3·910 | 1·000 | 0·000 | 0·800 | — |
| AsO | 2·910 | 0·000 | 0·000 | 0·600 | — |

[a] $F_\mu$ = dipole direct field parameter; $F_q$ = charge direct field parameter; $M_F$ = mesomeric effect parameter; $E_r$ = resonance parameter; $\chi_s$ = electronegativity parameter.

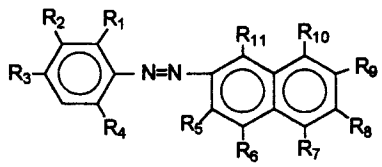
Table 3. SPARC p$K_a$ reaction center parameters

| Reaction center | $(pK_a)_c$ | $\rho_{ele}$ | $\rho_{res}$ | $\chi_c$ |
|---|---|---|---|---|
| $CO_2H$ | 3·75 | 1·000 | −1·100 | 2·591 |
| $SO_3H$ | −0·5 | 0·890 | −3·200 | — |
| $AsO_2H$ | 6·99 | 0·618 | 0·000 | 2·210 |
| $PO_2H$ | 2·96 | 0·403 | 0·000 | 2·792 |
| $BO_2H_2$ | 8·26 | 0·798 | −0·050 | — |
| $SeO_3H$ | 4·64 | 0·714 | −0·400 | 2·300 |
| OH | 14·3 | 2·260 | 18·65 | 2·512 |
| SH | 7·34 | 2·058 | 3·769 | 2·793 |
| $NR_2$ | 9·83 | 3·282 | 19·328 | 2·422 |
| in-ring N | 5·03 | 5·548 | −6·204 | — |
| =N | 5·06 | 4·051 | −6·236 | — |

The p$K_a$ calculator was then tested on data for *ca* 3000 compounds from the International Union of Pure and Applied Chemistry (IUPAC).[8,9] The RMS deviation for this large set of compounds was found to be 0·35 p$K_a$ units, which was approximately the same as experimental error. A report on this p$K_a$ performance test for IUPAC-approved organic compounds is in preparation.[10]

SPARC was also used to estimate 358 p$K_a$s for 214 azo dyes and a number of related aromatic amines (Tables 4–12). The results of this test are shown in Figure 2. For these compounds, the RMS deviation was 0·62 p$K_a$ units. The experimental error in the measured p$K_a$s for some of these dyes can be as high as 2 p$K_a$ units. Consider the case of molecule 99 where the p$K_a$s for the first $PO_3H_2$ group are 2·5 and 7·3 whereas those for the second $PO_3H_2$ group are 1·5 and 5·5. This molecule is symmetric and the $PO_3H_2$ groups are well removed and it is expected that the p$K_a$s for the two $PO_3H_2$ groups should be the same within a statistical term of 0·3. In addition, the p$K_a$s for the $PO_3H_2$ in molecule 22 are 1·9 and 7·3, which indicates that the second 4-chloro-2-phosphonophenylazo in molecule 99 has no effect on the first $PO_3H_2$ group. The two $PO_3H_2$ groups are too far away from each other to be affected electrostatically. Hence, the differences in the p$K_a$s are due to the statistical factor as shown in the calculated

Table 4. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 1–26




| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $R_9$ | $R_{10}$ | $R_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $CO_2H$ | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 2 | $OCH_3$ | — | — | — | — | $SO_3H$ | — | — | — | — | OH |
| 3 | — | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 4 | OH | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | $NH_2$ | OH |
| 5 | OH | — | — | — | — | — | $SO_3H$ | — | $SO_3H$ | $NH_2$ | OH |
| 6 | OH | $SO_3H$ | — | Cl | — | — | — | — | — | — | OH |
| 7 | OH | — | Cl | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 8 | OH | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 9 | OH | — | OH | — | — | $SO_3H$ | — | — | — | — | OH |
| 10 | — | — | $SO_3H$ | — | $SO_3H$ | — | — | $SO_3H$ | — | — | OH |
| 11 | — | — | Cl | — | $SO_3H$ | — | — | — | — | — | OH |
| 12 | — | — | Cl | — | — | $SO_3H$ | — | — | — | — | OH |
| 13 | $AsO_3H_2$ | — | — | — | — | $SO_3H$ | — | — | — | — | $NH_2$ |
| 14 | $NO_2$ | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 15 | — | $NO_2$ | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 16 | — | — | $NO_2$ | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 17 | — | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 18 | — | — | $SO_3H$ | — | $SO_3H$ | — | — | — | — | $SO_3H$ | OH |
| 19 | $SO_3H$ | — | — | — | $SO_3H$ | — | — | — | — | OH | OH |
| 20 | — | — | $SO_3H$ | — | $SO_3H$ | — | — | — | — | OH | OH |
| 21 | — | $SO_3H$ | — | — | $SO_3H$ | — | — | — | — | OH | OH |
| 22 | $PO_3H_2$ | — | Cl | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 23 | $CO_2H$ | — | — | — | — | $SO_3H$ | — | — | — | — | $NH_2$ |
| 24 | $C_2O_3H_3$ | — | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 25 | — | $C_2O_3H_3$ | — | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |
| 26 | — | — | $C_2O_3H_3$ | — | $SO_3H$ | — | — | $SO_3H$ | — | OH | OH |

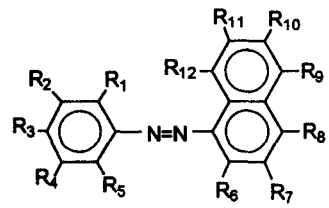
*continued*

Table 4. (*Continued*)

| Mol. No.[b] | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $1^{14}$ | 3·5 | $4·1^{R_1}$ | 10·2 | $9·0^{R_{10}}$ | | | | | | | | |
| $2^{16}$ | 8·8 | $9·2^{R_{10}}$ | | | | | | | | | | |
| $3^{16}$ | 9·2 | $8·4^{R_{10}}$ | | | | | | | | | | |
| $4^{11}$ | 8·4 | $8·1^{R_{11}}$ | 11·9 | $11·7^{R_1}$ | | | | | | | | |
| $5^{11}$ | 7·4 | $7·5^{R_{11}}$ | 11·6 | $12^{R_1}$ | | | | | | | | |
| $6^{16}$ | 6·0 | $6·0^{R_1}$ | 9·5 | $11^{R_{11}}$ | | | | | | | | |
| $7^{12}$ | 8·0 | $7·4^{R_{11}}$ | 10·5 | $10·7^{R_1}$ | 11·9 | $12·6^{10}$ | | | | | | |
| $8^{13}$ | 7·6 | $7·4^{R_{11}}$ | 9·3 | $11·1^{R_1}$ | 12·4 | $12·7^{10}$ | | | | | | |
| $9^{8}$ | 7·0 | $6·8^{R_{11}}$ | 9·2 | $9·6^{R_3}$ | | | | | | | | |
| $10^{8}$ | 10·4 | $10·8^{R_{11}}$ | | | | | | | | | | |
| $11^{8}$ | 11·3 | $10·7^{R_{11}}$ | | | | | | | | | | |
| $12^{8}$ | 7·8 | $9·4^{R_{11}}$ | | | | | | | | | | |
| $13^{8}$ | 3·9 | $4·4^{AsO_3H_2}$ | 6·4 | $7·4^{AsO_3H^{-1}}$ | | | | | | | | |
| $14^{8}$ | 9·0 | $8·9^{d}$ OH | | | | | | | | | | |
| $15^{8}$ | 8·6 | $8·7^{R_{10}}$ | | | | | | | | | | |
| $16^{8}$ | 8·8 | $8·9^{R_{10}}$ | | | | | | | | | | |
| $17^{8}$ | 9·2 | $8·9^{R_{10}}$ | | | | | | | | | | |
| $18^{8}$ | 9·4 | $11·0^{R_{10}}$ | | | | | | | | | | |
| $19^{8}$ | 9·3 | $9·3^{R_{10}}$ | | | | | | | | | | |
| $20^{8}$ | 8·9 | $8·9^{R_{10}}$ | | | | | | | | | | |
| $21^{8}$ | 8·9 | $9·1^{R_{10}}$ | | | | | | | | | | |
| $22^{8}$ | 0·6 | $0·5^{d}$ $SO_3H$ | 0·8 | $1·1^{d}$ $SO_3H$ | 1·9 | $2·2^{PO_3H_2}$ | 7·3 | $7·1^{PO_3H^{-1}}$ | 11·8 | $10·8^{R_{10}}$ | 15·2 | $13·8^{R_{11}}$ |
| $23^{8}$ | 4·0 | $3·9^{R_1}$ | | | | | | | | | | |
| $24^{8}$ | 3·0 | $4·2^{R_1}$ | 9·8 | $8·3^{R_{10}}$ | | | | | | | | |
| $25^{8}$ | 9·5 | $8·4^{R_{10}}$ | | | | | | | | | | |
| $26^{8}$ | 3·0 | $4·2^{R_2}$ | 9·0 | $8·4^{R_{10}}$ | | | | | | | | |

[a] Observed values might have more than one value, depending on the source.
[b] Superscripts are reference numbers.
[c] Superscripts are the ionizable reaction centers.
[d] $pK_a$ values for two or more reaction centers are identical within 0·1 $pK_a$ unit.

Table 5. Observed[a] versus SPARC-calculated $pK_a$ values for compounds 27–69




| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $R_9$ | $R_{10}$ | $R_{11}$ | $R_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | — | Cl | — | $SO_3H$ | OH | OH | $CO_2H$ | — | — | — | — | — |
| 28 | — | Cl | — | $SO_3H$ | OH | OH | $SO_3H$ | — | — | $SO_3H$ | — | — |
| 29 | $CO_2H$ | — | — | — | — | OH | — | — | — | $SO_3H$ | — | — |
| 30 | — | — | — | — | OH | OH | — | — | — | — | — | — |
| 31 | — | — | — | — | — | OH | — | — | — | $SO_3H$ | — | $SO_3H$ |
| 32 | — | — | — | — | — | OH | — | — | — | — | — | — |
| 33 | $CH_3$ | — | $SO_3H$ | — | — | OH | — | — | — | — | — | — |
| 34 | — | — | — | — | $CO_2H$ | OH | — | — | — | — | — | — |
| 35 | — | $C_6H_6$ | — | — | OH | OH | — | — | — | — | — | — |
| 36 | — | $SO_3H$ | — | — | OH | OH | — | — | — | — | — | — |
| 37 | — | $CH_3$ | — | — | OH | OH | — | $SO_3H$ | — | — | — | — |
| 38 | — | $NO_2$ | — | — | — | — | — | OH | — | $SO_3H$ | — | — |
| 39 | — | — | — | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 40 | — | — | $SO_3H$ | — | — | — | — | OH | — | — | — | — |
| 41 | — | $OCH_3$ | — | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 42 | — | — | $OCH_3$ | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 43 | — | $NO_2$ | — | $NO_2$ | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 44 | — | — | $NO_2$ | — | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 45 | — | $SO_3H$ | — | $NO_2$ | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 46 | — | $NO_2$ | — | $SO_3H$ | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 47 | — | — | — | — | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 48 | — | — | $NH_2$ | — | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 49 | — | $NO_2$ | — | Cl | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 50 | — | — | Cl | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 51 | — | Cl | — | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 52 | — | — | $C_2H_3O$ | — | — | — | — | OH | — | — | $SO_3H$ | — |
| 53 | — | — | — | $C_2H_3O$ | — | — | — | OH | — | — | $SO_3H$ | — |
| 54 | — | — | — | — | $SO_3H$ | OH | — | — | — | — | — | — |
| 55 | Cl | — | — | — | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 56 | — | Cl | — | $SO_3H$ | OH | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 57 | — | — | — | — | $AsO_3H_2$ | OH | $SO_3H$ | — | — | — | $SO_3H$ | — |
| 58 | — | — | — | — | $NO_2$ | — | $CO_2H$ | OH | — | — | — | — |
| 59 | — | — | — | $NO_2$ | — | — | $CO_2H$ | OH | — | — | — | — |
| 60 | — | — | — | — | — | — | $CO_2H$ | OH | — | — | — | — |
| 61 | — | — | Cl | — | — | — | $CO_2H$ | OH | — | — | — | — |
| 62 | — | — | I | — | — | — | $CO_2H$ | OH | — | — | — | — |
| 63 | $CH_3$ | — | — | — | OH | OH | — | $SO_3H$ | — | — | — | — |
| 64 | — | — | — | — | $CH_3$ | — | $CO_2H$ | OH | — | — | — | — |
| 65 | — | — | — | $CH_3$ | — | — | $CO_2H$ | OH | — | — | — | — |
| 66 | — | — | $CH_3$ | — | — | — | $CO_2H$ | OH | — | — | — | — |
| 67 | — | — | $SO_3H$ | — | — | OH | — | $SO_3H$ | — | — | — | $SO_3H$ |
| 68 | — | — | $CH_3$ | — | — | — | — | OH | — | — | $CO_2H$ | — |
| 69 | — | — | $OCH_3$ | — | — | — | $CO_2H$ | OH | — | — | — | — |

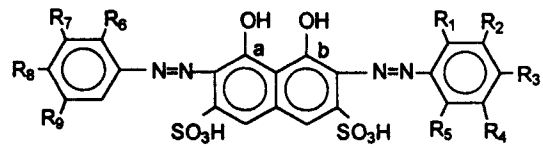
*continued*

Table 5. (*Continued*)

| Mol. No.[b] | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. |
|---|---|---|---|---|---|---|---|---|
| 27[15] | 7·3 | $6·7^{R_5}$ | 13·5 | $13·4^{R_6}$ | | | | |
| 28[15] | 6·0 | $6·5^{R_5}$ | 12·5 | $12·7^{R_6}$ | | | | |
| 29[14] | 4·2 | $4·0^{R_1}$ | 12·2 | $11·4^{R_6}$ | | | | |
| 30[11] | 7·7 | $7·0^{R_6}$ | 12·4 | $12·3^{R_5}$ | | | | |
| 31[16] | 11·5 | $11·2^{R_6}$ | | | | | | |
| 32[16] | 11·5 | $11·2^{R_6}$ | | | | | | |
| 33[16] | 11·8 | $11·2^{R_6}$ | | | | | | |
| 34[11] | 12·0 | $12·0^{R_6}$ | | | | | | |
| 35[11] | 8·0 | $7·3^{R_6}$ | 11·8 | $11·3^{R_5}$ | | | | |
| 36[12] | 7·0 | $6·9^{R_5}$ | 13·0 | $11·8^{R_6}$ | | | | |
| 37[12] | 8·1 | $8·3^{R_6}$ | 12·4 | $12·1^{R_5}$ | | | | |
| 38[8] | 7·5 | $7·5^{R_8}$ | | | | | | |
| 39[8] | 7·3 | $7·7^{R_8}$ | | | | | | |
| 40[8] | 8·2 | $8·0^{R_8}$ | | | | | | |
| 41[8] | 7·5 | $7·6^{R_8}$ | | | | | | |
| 42[8] | 7·3 | $7·7^{R_8}$ | | | | | | |
| 43[8] | 2·0 | $1·7^{R_5}$ | 12·2 | $12·1^{R_6}$ | | | | |
| 44[8] | 6·6 | $6·0^{R_5}$ | 13·0 | $12·2^{R_6}$ | | | | |
| 45[8] | 4·0 | $3·1^{R_5}$ | 12·6 | $12·7^{R_6}$ | | | | |
| 46[8] | 4·3 | $4·2^{R_5}$ | 11·9 | $12·3^{R_6}$ | | | | |
| 47[8] | 7·30 | $6·2^{R_6}$ | 12·6 | $12·6^{R_5}$ | | | | |
| 48[8] | 7·2 | $6·2^{R_6}$ | 12·5 | $12·8^{R_5}$ | | | | |
| 49[8] | 3·6 | $3·8^{R_5}$ | 11·8 | $12·0^{R_6}$ | | | | |
| 50[8] | 7·4 | $7·6^{R_8}$ | | | | | | |
| 51[8] | 7·5 | $7·6^{R_8}$ | | | | | | |
| 52[8] | 8·0 | $7·6^{R_8}$ | | | | | | |
| 53[8] | 7·6 | $7·6^{R_8}$ | | | | | | |
| 54[8] | 11·4 | $11·4^{R_6}$ | | | | | | |
| 55[8] | 7·5 | $7·4^{R_6}$ | 12·5 | $12·0^{R_5}$ | | | | |
| 56[8] | 6·6 | $6·0^{R_6}$ | 12·5 | $12·7^{R_5}$ | | | | |
| 57[8] | 2·4 | $0·9^{R_7}$ | 4·4 | $4·7^{AsO_3H_2}$ | 8·3 | $8·4^{AsO_3H^{-1}}$ | 11·2 | $12·7^{R_6}$ |
| 58[8] | 11·7 | $11·7^{R_8}$ | | | | | | |
| 59[8] | 10·8 | $10·6^{R_8}$ | | | | | | |
| 60[8] | 11·4 | $11·8^{R_8}$ | | | | | | |
| 61[8] | 11·3 | $11·7^{R_8}$ | | | | | | |
| 62[8] | 11·3 | $11·7^{R_8}$ | | | | | | |
| 63[8] | 8·1 | $8·2^{R_6}$ | 12·4 | $12·7^{R_5}$ | | | | |
| 64[8] | 11·6 | $11·9^{R_8}$ | | | | | | |
| 65[8] | 11·4 | $11·9^{R_8}$ | | | | | | |
| 66[8] | 11·5 | $11·9^{R_8}$ | | | | | | |
| 67[8] | 10·5 | $10·6^{R_6}$ | | | | | | |
| 68[8] | 7·30 | $8·30^{R_8}$ | | | | | | |
| 69[8] | 11·6 | $11·9^{R_8}$ | | | | | | |

[a–c] See Table 4.

Table 6. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 70–101




| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ | $R_9$ |
|---|---|---|---|---|---|---|---|---|---|
| 70 | $CO_2H$ | — | — | — | — | $NO_2$ | — | — | — |
| 71 | $CO_2H$ | — | — | — | — | — | $NO_2$ | — | — |
| 72 | $CO_2H$ | — | — | — | — | — | — | $NO_2$ | — |
| 73 | $CO_2H$ | — | — | — | — | — | — | — | — |
| 74 | $CO_2H$ | — | — | — | — | — | $SO_3H$ | — | — |
| 75 | $CO_2H$ | — | — | — | — | — | — | $SO_3H$ | — |
| 76 | $CO_2H$ | — | — | — | — | $CH_3$ | — | — | — |
| 77 | $CO_2H$ | — | — | — | — | — | $CH_3$ | — | — |
| 78 | $CO_2H$ | — | — | — | — | — | — | $CH_3$ | — |
| 79 | $CO_2H$ | — | — | — | — | — | — | $OCH_3$ | — |
| 80 | — | $NO_2$ | — | — | — | — | $NO_2$ | — | — |
| 81 | — | — | $NO_2$ | — | — | — | — | $NO_2$ | — |
| 82 | — | — | $SO_3H$ | — | — | — | — | $NO_2$ | — |
| 83 | — | — | — | — | — | — | — | — | — |
| 84 | — | — | — | — | — | — | — | OH | — |
| 85 | — | — | OH | — | — | — | — | OH | — |
| 86 | $SO_3H$ | — | — | — | — | — | — | — | — |
| 87 | — | — | $SO_3H$ | — | — | — | — | — | — |
| 88 | — | — | $SO_3H$ | — | — | — | — | $SO_3H$ | — |
| 89 | — | — | $CH_3$ | — | — | — | — | $CH_3$ | — |
| 90 | — | — | $OCH_3$ | — | — | — | — | $OCH_3$ | — |
| 91 | $AsO_3H_2$ | — | — | — | — | OH | $NO_2$ | — | $NO_2$ |
| 92 | $CO_2H$ | — | — | — | — | $CO_2H$ | — | — | — |
| 93 | $CO_2H$ | — | — | — | — | — | $CO_2H$ | — | — |
| 94 | $CO_2H$ | — | — | — | — | — | — | $CO_2H$ | — |
| 95 | — | — | $CO_2H$ | — | — | — | — | $CO_2H$ | — |
| 96 | $SO_3H$ | — | — | — | — | $SO_3H$ | — | — | — |
| 97 | OH | $SO_3H$ | — | Cl | — | OH | $SO_3H$ | — | Cl |
| 98 | $PO_3H_2$ | — | — | — | — | $PO_3H_2$ | — | — | — |
| 99 | $PO_3H_2$ | — | Cl | — | — | $PO_3H_2$ | — | Cl | — |
| 100 | $PO_3H_2$ | — | Cl | $CH_3$ | — | $PO_3H_2$ | — | Cl | $CH_3$ |
| 101 | $AsO_3H_2$ | — | — | $CO_2H$ | — | $AsO_3H_2$ | — | — | $CO_2H$ |

*continued*

Table 6. (*Continued*)

| Mol. No.[b] | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. |
|---|---|---|---|---|---|---|---|---|
| 70[8] | 9·0 | 9·0[a] | 14·5 | 14·2[b] | | | | |
| 71[8] | 9·8 | 9·8[a] | 14·4 | 14·2[b] | | | | |
| 72[8] | 9·9 | 9·5[a] | 14·5 | 14·2[b] | | | | |
| 73[8] | 10·3 | 10·4[d] OH | 15·2 | 14·2[d] OH | | | | |
| 74[8] | 10·4 | 9·7[a] | 14·3 | 14·3[b] | | | | |
| 75[8] | 10·6 | 9·7[a] | 14·4 | 14·4[b] | | | | |
| 76[8] | 11·2 | 9·8[d] OH | 14·2 | 14·5[d] OH | | | | |
| 77[8] | 10·3 | 9·6[a] | 14·8 | 14·2[b] | | | | |
| 78[8] | 10·3 | 9·7[a] | 14·7 | 14·3[b] | | | | |
| 79[8] | 10·5 | 9·5[a] | 14·8 | 14·1[b] | | | | |
| 80[8] | 8·4 | 8·4[d] OH | 13·7 | 13·3[d] OH | | | | |
| 81[8] | 8·1 | 8·3[d] OH | 13·5 | 13·5[d] OH | | | | |
| 82[8] | 8·5 | 8·9[a] | 13·5 | 13·5[b] | | | | |
| 83[8] | 8·9 | 8·9[d] OH | 14·3 | 14·0[d] OH | | | | |
| 84[8] | 10·3 | 9·7[a] | | | | | | |
| 85[8] | 10·3 | 10·5[d] OH | | | | | | |
| 86[8] | 9·9 | 9·1[d] OH | 14·7 | 14·1[d] OH | | | | |
| 87[8] | 9·6 | 9·1[d] OH | 14·2 | 14·0[d] OH | | | | |
| 88[8] | 9·4 | 8·8[d] OH | 14·0 | 14·1[d] OH | | | | |
| 89[8] | 9·2 | 8·9[d] OH | 14·3 | 14·1[d] OH | | | | |
| 90[8] | 9·4 | 8·9[d] OH | 14·6 | 14·1[d] OH | | | | |
| 76[8] | 2·0 | 2·0[a] | 12·0 | 11·5[b] | | | | |
| 92[8] | 10·1 | 10·6[d] OH | 14·6 | 15·1[d] OH | | | | |
| 93[8] | 10·3 | 9·9[d] OH | 14·6 | 14·6[d] OH | | | | |
| 94[8] | 10·0 | 9·9[d] OH | 14·6 | 14·6[d] OH | | | | |
| 95[8] | 9·3 | 9·9[d] OH | 14·2 | 14·5[d] OH | | | | |
| 96[8] | 0·9 | 0·9[d] $SO_3H$ | 1·9 | 1·6[d] $SO_3H$ | 2·3 | 2·0[d] $SO_3H$ | 2·8 | 2·6[d] $SO_3H$ |
| | 11·6 | 9·8[d] OH | 14·4 | 14·2[d] OH | | | | |
| 97[8] | 1·3 | 1·2[d] $SO_3H$ | 2·5 | 1·8[d] $SO_3H$ | 7·1 | 6·9[d] $R_1,R_6$ | 9·7 | 7·2[d] $R_1,R_6$ |
| | 11·9 | 10·9[d] OH | 14·5 | 14·5[d] OH | | | | |
| 98[8] | 0·3 | 0·9[d] $SO_3H$ | 0·6 | 1·4[d] $SO_3H$ | 1·7 | 2·4[d] $PO_3H_2$ | 4·5 | 3·1[d] $PO_3H_2$ |
| | 7·2 | 7·0[d] $PO_3H^{-1}$ | 9·7 | 7·7[d] $PO_3H^{-1}$ | 11·3 | 11·0[d] OH | 14·6 | 15·6[d] OH |
| 99[8] | 0·6 | 1·0[d] $SO_3H$ | 0·8 | 1·1[d] $SO_3H$ | 1·5 | 2·4[d] $PO_3H_2$ | 2·5 | 2·6[d] $PO_3H_2$ |
| | 5·5 | 6·7[d] $PO_3H^{-1}$ | 7·2 | 7·3[d] $PO_3H^{-1}$ | 12·5 | 10·9 | 15·3 | 15·2 |
| 100[8] | 0·3 | 0·8[d] $SO_3H$ | 1·6 | 1·5[d] $SO_3H$ | 1·6 | 1·1[d] $PO_3H_2$ | 4·3 | 1·8[d] $PO_3H_2$ |
| | 7·2 | 6·8[d] $PO_3H^{-1}$ | 9·4 | 7·3[d] $PO_3H^{-1}$ | 11·2 | 11·0[d] OH | 14·6 | 15·0[d] OH |
| 101[8] | 3·3 | 3·5[d] $CO_2H$ | 4·0 | 4·1[d] $CO_2H$ | 5·0 | 4·7[d] $AsO_3H_2$ | 6·5 | 5·3[d] $AsO_3H_2$ |
| | 6·5 | 8·3[d] $AsO_3H^{-1}$ | 9·0 | 8·9[d] $AsO_3H^{-1}$ | 11·5 | 11·1[d] OH | 14·7 | 16·0[d] OH |

[a–d] See Table 4.

values. The same case applies for molecules such as 101 and 100. The reported RMS interlaboratory deviations between the different observed values for azo dyes and related aromatic amines where more than one measurement was reported is 0·64.[8,9] We therefore believe that the errors in our calculated values are comparable to experimental error for these complicated molecules. The utility of such data is illustrated by the following analysis.

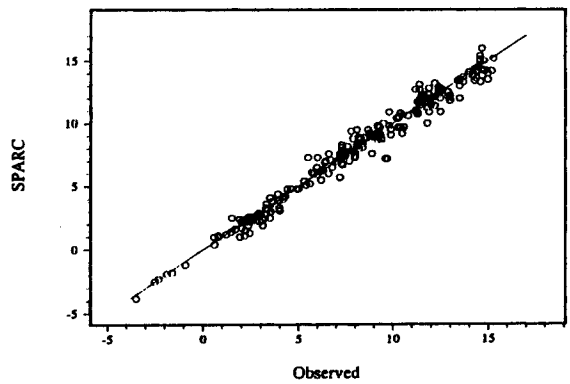
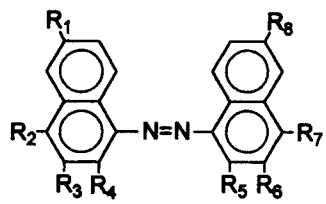



Figure 2. Observed versus SPARC-calculated p$K_a$ values for dye compounds

Table 7. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 102–110

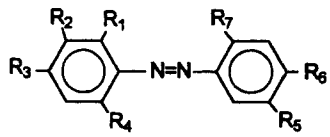



| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ | $R_8$ |
|---|---|---|---|---|---|---|---|---|
| 102 | $SO_3H$ | — | $SO_3H$ | OH | OH | — | — | $SO_3H$ |
| 103 | $SO_3H$ | — | $SO_3H$ | OH | OH | $SO_3H$ | — | $SO_3H$ |
| 104 | $SO_3H$ | — | $SO_3H$ | OH | OH | $CO_2H$ | — | — |
| 105 | $SO_3H$ | — | $SO_3H$ | OH | OH | OH | — | — |
| 106 | $SO_3H$ | — | $SO_3H$ | OH | OH | OH | — | $SO_3H$ |
| 107 | — | — | $CO_2H$ | OH | OH | — | $SO_3H$ | — |
| 108 | $SO_3H$ | — | $SO_3H$ | OH | OH | — | $SO_3H$ | — |
| 109 | — | $SO_3H$ | — | OH | OH | — | — | — |
| 110 | $NO_2$ | $SO_3H$ | — | OH | OH | — | — | — |

| Mol. No.[b] | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. |
|---|---|---|---|---|
| 102[15] | 5·7 | 6·1$^{R_4}$ | 13·4 | 13·4$^{R_5}$ |
| 103[15] | 5·7 | 6·1$^{d\ OH}$ | 13·5 | 13·0$^{d\ OH}$ |
| 104[15] | 9·3 | 9·3$^{R_4}$ | 14·0 | 13·9$^{R_5}$ |
| 105[15] | 6·0 | 6·0$^{R_4}$ | | |
| 106[15] | 5·8 | 5·8$^{R_4}$ | | |
| 107[15] | 9·3 | 9·1$^{R_5}$ | 13·7 | 13·7$^{R_4}$ |
| 108[15] | 6·1 | 7·0$^{R_4}$ | 12·9 | 12·5$^{R_5}$ |
| 109[12] | 6·8 | 6·6$^{R_4}$ | 13·5 | 12·0$^{R_5}$ |
| 110[12] | 6·2 | 5·5$^{R_4}$ | 13·0 | 12·0$^{R_5}$ |

[a–d] See Table 4.

Table 8. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 111–127

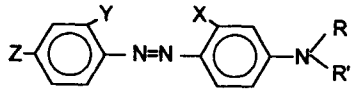



| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ | $R_7$ |
|---|---|---|---|---|---|---|---|
| 111 | — | — | OH | — | — | $N_2C_6H_5$ | — |
| 112 | OH | — | — | — | — | — | OH |
| 113 | OH | — | — | — | — | OH | OH |
| 114 | $CO_2H$ | — | — | — | $CH_3$ | — | OH |
| 115 | OH | — | OH | — | $C_6H_6$ | — | OH |
| 116 | — | $CO_2H$ | OH | — | — | Cl | — |
| 117 | — | $CO_2H$ | OH | — | — | Br | — |
| 118 | — | — | OH | — | — | $PO_3H_2$ | — |
| 119 | — | — | OH | — | — | $AsO_3H_2$ | — |
| 120 | — | — | — | — | — | — | — |
| 121 | — | — | $NH_2$ | — | — | — | — |
| 122 | — | — | $NH_2$ | — | $CH_3$ | — | — |
| 123 | — | — | $NH_2$ | — | — | $CH_3$ | — |
| 124 | $CO_2H$ | — | — | — | — | OH | — |
| 125 | — | — | $CO_2H$ | — | — | OH | — |
| 126 | — | — | OH | — | — | — | — |
| 127 | — | — | $NO_2$ | — | — | — | — |

| Mol. No.[b] | p$K_a$[c] | | p$K_a$[c] | | p$K_a$[c] | |
|---|---|---|---|---|---|---|
| | Obs. | Calc. | Obs. | Calc. | Obs. | Calc. |
| 111[16] | 8·1 | $8·2^{R_3}$ | | | | |
| 112[11] | 7·8 | $8·0^{d\ OH}$ | 11·5 | $11·4^{d\ OH}$ | | |
| 113[11] | 6·6 | $7·0^{R_7}$ | 8·7 | $9·2^{R_6}$ | 12·2 | $11·7^{R_1}$ |
| 114[11] | 11·4 | $11·8^{R_1}$ | | | | |
| 115[11] | 6·7 | $7·0^{R_7}$ | 8·1 | $9·2^{R_3}$ | 11·4 | $11·3^{R_1}$ |
| 116[9] | 11·6 | $11·5^{R_3}$ | | | | |
| 117[9] | 11·5 | $11·5^{R_3}$ | | | | |
| 118[9] | 6·4 | $6·9^{R_6}$ | | | | |
| 119[9] | 7·9 | $7·8^{R_6}$ | | | | |
| 120[9] | −2·5 | $-2·5^{d\ =N}$ | | | | |
| 121[9] | 2·8 | $2·7^{R_3}$ | | | | |
| 122[9] | 2·9 | $2·7^{R_3}$ | | | | |
| 123[9] | 3·0 | $2·8^{R_3}$ | | | | |
| 124[9] | −1·3 | $-1·6^{=N}$ | 8·2 | $8·6^{R_6}$ | | |
| 125[9] | −1·6 | $-1·8^{=N}$ | 7·9 | $8·8^{R_6}$ | | |
| 126[9] | 8·2 | $8·4^{R_3}$ | | | | |
| 127[9] | −3·5 | $-3·8^{=N}$ | | | | |

[a–d] See Table 4.

Table 9. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 128–159

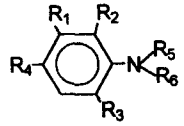



| No. | X | Y | Z | R | R' |
|---|---|---|---|---|---|
| 128 | — | — | — | $C_2H_5OH$ | $C_2H_5OH$ |
| 129 | $CH_3$ | — | — | $C_2H_5OH$ | $C_2H_5OH$ |
| 130 | — | $OCH_3$ | — | $C_2H_5OH$ | $C_2H_5OH$ |
| 131 | — | $CH_3$ | — | $C_2H_5OH$ | $C_2H_5OH$ |
| 132 | — | — | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 133 | — | $OCH_3$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 134 | — | $CH_3$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 135 | — | Cl | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 136 | — | $NO_2$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 137 | $CH_3$ | — | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 138 | $CH_3$ | $OCH_3$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 139 | $CH_3$ | $CH_3$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 140 | $CH_3$ | Cl | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 141 | $CH_3$ | $NO_2$ | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 142 | Cl | — | $NO_2$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 143 | — | — | $OCH_3$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 144 | — | — | $CH_3$ | $C_2H_5OH$ | $C_2H_5OH$ |
| 145 | — | — | F | $C_2H_5OH$ | $C_2H_5OH$ |
| 146 | — | — | Cl | $C_2H_5OH$ | $C_2H_5OH$ |
| 147 | — | — | Br | $C_2H_5OH$ | $C_2H_5OH$ |
| 148 | — | — | CN | $C_2H_5OH$ | $C_2H_5OH$ |
| 149 | — | — | $NO_2$ | $C_2H_5$ | $C_2H_5OH$ |
| 150 | — | Cl | $NO_2$ | $C_2H_5$ | $C_2H_5OH$ |
| 151 | — | — | — | — | $C_2H_5OH$ |
| 152 | — | — | — | $C_2H_5$ | $C_2H_5OH$ |
| 153 | — | — | — | $C_2H_5$ | $C_2H_4CN$ |
| 154 | — | — | — | $C_6H_5$ | — |
| 155 | — | — | OH | $C_2H_5$ | $C_2H_5$ |
| 156 | — | — | — | $CH_3$ | $CH_3$ |
| 157 | — | — | OH | $CH_3$ | $CH_3$ |
| 158 | — | — | $CO_2H$ | $CH_3$ | $CH_3$ |
| 159 | — | — | $C_2O_2H_3$ | $CH_3$ | $CH_3$ |

| Mol. No.[b] | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. |
|---|---|---|---|---|---|---|
| 128[17] | 2·6 | $2·2^{NR_2}$ | | | | |
| 129[17] | 3·5 | $3·5^{NR_2}$ | | | | |
| 130[17] | 2·3 | $2·3^{NR_2}$ | | | | |
| 131[17] | 2·1 | $2·8^{NR_2}$ | | | | |
| 132[17] | 2·4 | $1·3^{NR_2}$ | | | | |
| 133[17] | 2·2 | $1·5^{NR_2}$ | | | | |
| 134[17] | 1·7 | $1·6^{NR_2}$ | | | | |
| 135[17] | 1·5 | $1·4^{NR_2}$ | | | | |
| 136[17] | 1·9 | $1·0^{NR_2}$ | | | | |
| 137[17] | 3·5 | $2·5^{NR_2}$ | | | | |
| 138[17] | 3·2 | $2·5^{NR_2}$ | | | | |
| 139[17] | 2·7 | $2·6^{NR_2}$ | | | | |
| 140[17] | 2·6 | $2·4^{NR_2}$ | | | | |
| 141[17] | 3·1 | $1·9^{NR_2}$ | | | | |
| 142[17] | 1·2 | $1·2^{NR_2}$ | | | | |
| 143[17] | 2·4 | $2·2^{NR_2}$ | | | | |
| 144[17] | 2·6 | $2·3^{NR_2}$ | | | | |
| 145[17] | 2·4 | $2·2^{NR_2}$ | | | | |
| 146[17] | 2·4 | $2·2^{NR_2}$ | | | | |
| 147[17] | 2·4 | $2·4^{NR_2}$ | | | | |
| 148[17] | 2·4 | $1·9^{NR_2}$ | | | | |
| 149[17] | 3·1 | $2·4^{NR_2}$ | | | | |
| 150[17] | 2·2 | $2·3^{NR_2}$ | | | | |
| 151[17] | 2·8 | $2·3^{NR_2}$ | | | | |
| 152[17] | 3·3 | $3·2^{NR_2}$ | | | | |
| 153[17] | 2·2 | $1·2^{NR_2}$ | | | | |
| 154[9] | 1·0 | $0·8^{NR_2}$ | | | | |
| 155[9] | −1·9 | $-1·9^{=N}$ | | | | |
| 156[9] | 3·3 | $3·3^{NR_2}$ | | | | |
| 157[9] | −2·3 | $-2·3^{=N}$ | 3·4 | $3·6^{NR_2}$ | 8·6 | $8·9^{OH}$ |
| 158[18] | 2·4 | $3·5^{NR_2}$ | — | $3·9^{CO_2H}$ | | |
| 159[9] | −3·1 | $-2·8^{=N}$ | 2·6 | $2·7^{NR_2}$ | | |

[a–c] See Table 4.

Table 10. Observed[a] versus SPARC-calculated p$K_a$ values for compounds 160–193

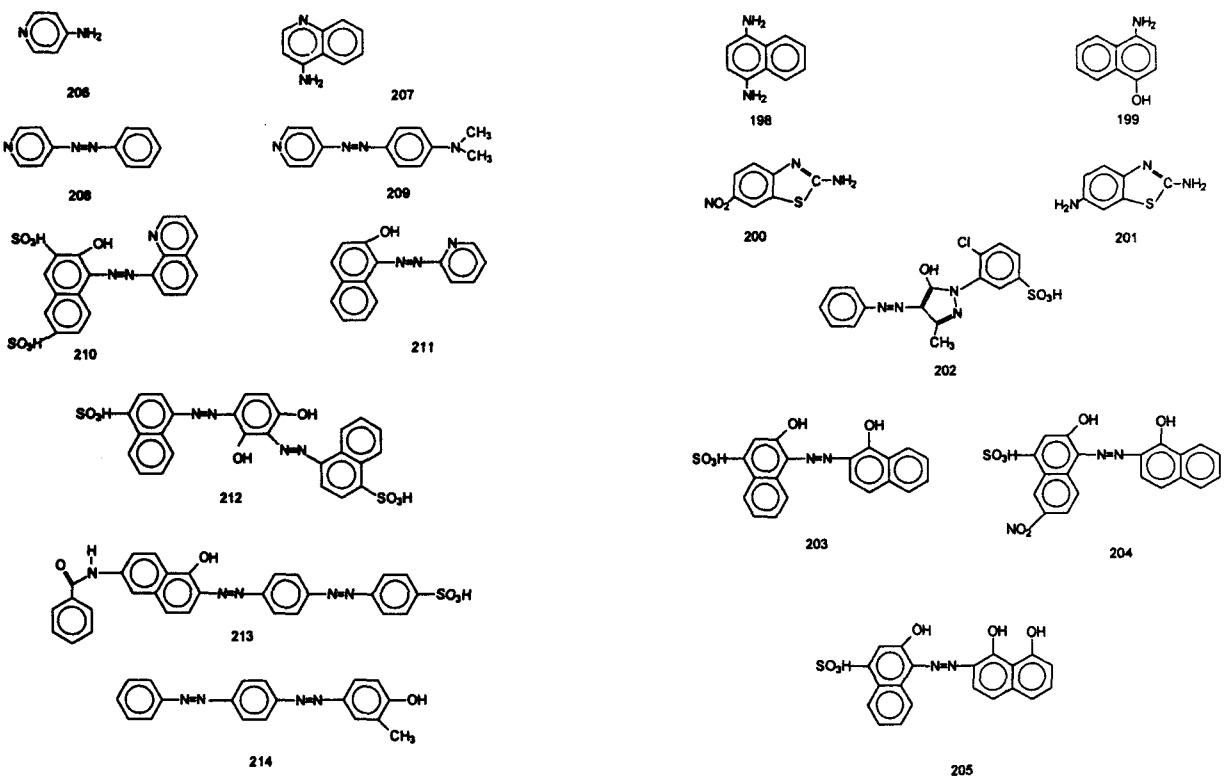



| No. | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ | $R_6$ |
|---|---|---|---|---|---|---|
| 160 | — | — | — | — | — | — |
| 161 | — | — | — | $NH_2$ | — | — |
| 162 | — | — | — | OH | — | — |
| 163 | — | Br | $NO_2$ | $NO_2$ | — | — |
| 164 | — | Br | $NH_2$ | $NO_2$ | — | — |
| 165 | — | Cl | Cl | $NO_2$ | — | — |
| 166 | — | Br | $NH_2$ | $NH_2$ | — | — |
| 167 | — | Cl | — | $NH_2$ | — | — |
| 168 | — | Cl | Cl | $NH_2$ | — | — |
| 169 | — | $CH_3$ | — | $NH_2$ | — | — |
| 170 | — | Br | $NO_2$ | $NH_2$ | — | — |
| 171 | — | $NH_2$ | — | $NH_2$ | — | — |
| 172 | — | — | — | $NH_2$ | $CH_3$ | — |
| 173 | — | — | — | $NH_2$ | $CH_3$ | $CH_3$ |
| 174 | — | — | — | $NH_2$ | $C_2H_5$ | $C_2H_5$ |
| 175 | — | — | — | $NH_2$ | $C_2H_5$ | $C_2H_4OH$ |
| 176 | — | — | — | $NH_2$ | $C_2H_5$ | $C_2H_4CN$ |
| 177 | — | — | — | $NH_2$ | $C_2H_4OH$ | $C_2H_4OH$ |
| 178 | — | — | — | $NH_2$ | — | $C_2H_4OH$ |
| 179 | — | — | — | $NH_2$ | $C_2H_4CN$ | $C_4O_2H_7$ |
| 180 | — | — | — | $NH_2$ | $C_2H_4OH$ | $C_2H_4CN$ |
| 181 | — | — | — | $NH_2$ | $C_2H_4CN$ | $C_2H_4CN$ |
| 182 | Cl | — | — | $NH_2$ | $C_2H_4OH$ | $C_2H_4OH$ |
| 183 | — | — | — | — | $CH_3$ | — |
| 184 | — | — | — | — | $CH_3$ | $CH_3$ |
| 185 | — | — | — | — | $C_2H_5$ | $C_2H_5$ |
| 186 | — | — | — | — | $C_2H_5$ | $C_2H_4OH$ |
| 187 | — | — | — | — | $C_2H_5$ | $C_2H_4CN$ |
| 188 | — | — | — | — | $C_2H_4OH$ | $C_2H_4OH$ |
| 189 | — | — | — | — | — | $C_2H_4OH$ |
| 190 | — | — | — | — | $C_2H_4OH$ | $C_2H_4CN$ |
| 191 | — | — | — | — | $C_2H_4CN$ | $C_2H_4CN$ |
| 192 | — | — | — | $N(CH_3)_2$ | $CH_3$ | $CH_3$ |
| 193 | — | — | — | $N(C_2H_5)_2$ | $C_2H_5$ | $C_2H_5$ |

| Mol. No.[b] | p$K_a$[c] Obs. | p$K_a$[c] Calc. | p$K_a$[c] Obs. | p$K_a$[c] Calc. |
|---|---|---|---|---|
| 160[9] | 4·6 | $4·9^{NR_2}$ | | |
| 161[9] | 3·3 | $3·1^{d\ NR_2}$ | 6·0 | $6·1^{d\ NR_2}$ |
| 162[9] | 5·3 | $5·4^{NR_2}$ | 10·4 | $10·8^{OH}$ |
| 163[9] | −6·6 | $-5·4^{NR_2}$ | | |
| 164 | — | $-1·5^{NR_2}$ | — | $1·4^{R_3}$ |
| 165[9] | −2·6 | $-3·2^{NR_2}$ | | |
| 166 | — | $1·3^{NR_2}$ | — | $4·5^{R_4}$ |
| 167 | — | $1·8^{NR_2}$ | — | $4·9^{R_4}$ |
| 168 | — | $-1·1^{NR_2}$ | — | $3·7^{R_4}$ |
| 169 | — | $3·3^{NR_2}$ | — | $6·0^{R_4}$ |
| 170 | — | $0^{R_4}$ | — | $3·2^{NR_2}$ |
| 171[9] | 3·7 | $3·5^{NR_2}$ | 6·1 | $6·1^{R_4}$ |
| 172 | — | $3·4^{d\ NR_2}$ | — | $6·1^{d\ NR_2}$ |
| 173[9] | — | $3·4^{R_4}$ | 6·6 | $6·4^{NR_2}$ |
| 174[9] | — | $3·9^{R_4}$ | 8·0 | $7·3^{NR_2}$ |
| 175 | — | $3·4^{R_4}$ | — | $6·4^{NR_2}$ |
| 176[17] | 2·7 | $3·9^{R_4}$ | — | $5·8^{NR_2}$ |
| 177 | — | $0·7^{R_4}$ | — | $5·4^{NR_2}$ |
| 178 | — | $2·8^{R_4}$ | — | $6·0^{NR_2}$ |
| 179 | — | $-0·1^{d\ NR_2}$ | — | $5·2^{d\ NR_2}$ |
| 180 | — | $0·6^{d\ NR_2}$ | — | $5·4^{d\ NR_2}$ |
| 181 | — | $-1·4^{NR_2}$ | — | $5·0^{R_4}$ |
| 182 | — | $1·9^{R_4}$ | — | $4·4^{NR_2}$ |
| 183[9] | 4·9 | $4·8^{NR_2}$ | | |
| 184[9] | 5·2 | $5·3^{NR_2}$ | | |
| 185[9] | 6·5 | $6·1^{NR_2}$ | | |
| 186[17] | 5·6 | $5·1^{NR_2}$ | | |
| 187[17] | 3·7 | $3·2^{NR_2}$ | | |
| 188[17] | 4·2 | $4·1^{NR_2}$ | | |
| 189[17] | 4·2 | $4·2^{NR_2}$ | | |
| 190[17] | 2·5 | $2·2^{NR_2}$ | | |
| 191 | — | $0·0^{NR_2}$ | | |
| 192[9] | 2·9 | $3·8^{d\ NR_2}$ | 6·3 | $6·3^{d\ NR_2}$ |
| 193 | — | $4·7^{d\ NR_2}$ | — | $7·4^{d\ NR_2}$ |

[a–d] See Table 4.

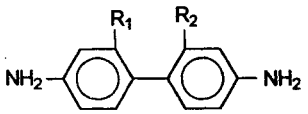
Table 12. Observed[a] versus SPARC-calculated $pK_a$ values for compounds 198–214



| Mol. No.[b] | $pK_a$[c] Obs. | $pK_a$[c] Calc. | $pK_a$[c] Obs. | $pK_a$[c] Calc. |
|---|---|---|---|---|
| 198[9] | 2·8 | $2·3^{d\ NR_2}$ | 5·9 | $5·7^{d\ NR_2}$ |
| 199[9] | — | $4·4^{NR_2}$ | — | $10·4^{OH}$ |
| 200 | — | $0·1^{=N}$ | | |
| 201 | — | $3·2^{=N}$ | — | $3·6^{NR_2}$ |
| 202[16] | 8·1 | $8·3^{OH}$ | | |
| 203[12] | 6·2 | $6·5^{OH}$ | 12·5 | $12·5^{OH}$ |
| 204[12] | 6·3 | $6·5^{OH}$ | 11·6 | $11·5^{OH}$ |
| 205[20] | 8·9 | $7·5^{OH}$ | 12·9 | $12·3^{OH}$ |
| 206[9] | 9·1 | $8·6^{\pi\text{-Ring}}$ | | |
| 207[9] | 9·1 | $8·8^{\pi\text{-Ring}}$ | | |
| 208[9] | 3·5 | $3·2^{\pi\text{-Ring}}$ | | |
| 209[9] | 3·4 | $3·0^{\pi\text{-Ring}}$ | 5·4 | $4·2^{NR_2}$ |
| 210[8] | 11·5 | $11·7^{OH}$ | | |
| 211[8] | 2·9 | $2·8^{\pi\text{-Ring}}$ | 11·5 | $11·2^{OH}$ |
| 212 | — | $12·1^{OH}$ | | |
| 213[16] | 11·5 | $11·8^{OH}$ | | |
| 214 | — | $8·2^{OH}$ | | |

[a–d] See Table 4.

Table 11. Observed[a] versus SPARC-calculated $pK_a$ values for compounds 194–197




| No. | $R_1$ | $R_2$ |
|---|---|---|
| 194 | H | H |
| 195 | $CH_3$ | $CH_3$ |
| 196 | Cl | Cl |
| 197 | $OCH_3$ | $OCH_3$ |

| Mol. No.[b] | $pK_a2_{NR_2}$ | | $pK_a1_{NR_2}$ | |
|---|---|---|---|---|
| | Obs. | Calc. | Obs. | Calc. |
| 194[9] | 3·7 | 3·8 | 5·1 | 5·0 |
| 195[19] | 4·0 | 3·7 | 5·3 | 4·9 |
| 196 | — | 2·5 | — | 3·9 |
| 197[19] | — | 3·6 | — | 4·9 |

[a,b] See Table 4.

## ENVIRONMENTAL IMPLICATIONS FOR DYES

Although sorption of inorganic ions by soil and sediment has been studied extensively, rigorous methods are not available for quantitatively predicting the extent of such equilibria for organic ions in aquatic systems. It has been shown, however, that aromatic amines sorb in a fashion characteristic of cation exchange and that sorption decreases with increasing pH above the $pK_a$.[21–23] Hence it usually is assumed that compounds will sorb strongly if they have a $pK_a$ below or near the pH of natural water, i.e. *ca* 5–7. This is, of course, in addition to hydrophobic sorption of the unprotonated species that can be predicted from the compound's octanol–water partition coefficient or its water solubility.[24]

Table 10 gives previously unavailable $pK_a$s for a number of disperse azo dyes and, as might be expected, most of the $pK_a$s are 3 or less. Hence it is reasonable to expect hydrophobic effects to play an important role, or even dominate, sediment sorption of such compounds.

Importantly, disperse dyes of the type shown in Table 9 are reduced in anoxic sediments with formation of amine products of the types shown in Tables 10 and 12.[3,5,25] Although the $pK_a$ of most of these compounds is not available from the literature, the SPARC $pK_a$ estimates are about 5–6. Hence these amines are likely to be sorbed strongly by cation exchange. Further, this conclusion would not change even if the constants are in error by the amount expected for SPARC.

Similarly, the carcinogenic benzidine moiety is expected to result from sediment transformation of many direct dyes and pigments.[4,26] It has been shown that benzidine (compound 194)[21] and dichlorobenzidine (compound 196)[23] undergo the above-mentioned pH-sensitive sorption by soil and sediment. Even though the $pK_a$ has not been measured for dichlorobenzidine, the estimated data in Table 11 clearly support the strong sorption observed experimentally.[23]

Specifically, the data show that, for dyes, many of the toxic amines resulting from environmental transformation, most likely in the benthic sediments, are also likely to be sorbed strongly to sediment or soil. However, it should be noted that this generalization cannot be applied to compounds that ionize through proton loss, i.e. anions.

## CONCLUSION

The SPARC model predicts $pK_a$ values that are as reliable as most experimental measurements for a wide range of molecular structures. Further, the model permits the prediction of $pK_a$ values for many compounds that are not amenable to experimental measurement.

Application of the model to azo dyes and their degradation products shows that most disperse dyes will probably sorb by a hydrophobic mechanism rather than by ion exchange. The data also suggest that the reverse is probably true for the aromatic amine products that result from reductive cleavage of the azo bonds.

### ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support provided by the US Environmental Protection Agency. We also acknowledge the support for this research from the US Environmental Research Laboratory at Athens, Georgia.

### REFERENCES

1. J. Houk, M. J. Doa, M. Dezube and J. M. Rovinski, in *The Design and Synthesis of Organic Dyes and Pigments* edited by A. T. Peters and H. S. Freeman. pp. 135–150. Elsevier Applied Science, London (1991).
2. S. W. Karickhoff, V. K. McDaniel, C. M. Melton, A. N. Vellino, D. E. Nute, and L. A. Carreira, *Environ. Toxicol. Chem.*, **10**, 1405–1416 (1991).
3. E. J. Weber, *Fate of Textile Dyes in Aquatic Environment: Degradation of Disperse Blue 79 in Anaerobic Sediment–Water Systems*, EPA-600/x-88-146. US Environmental Protection Agency, Athens, GA, 1988.
4. O. Yoshida, M. Miyakawa, Y. Okada, K. Ohshiro, T. Harada, S. Machida and T. Kato, in *Decomposition of Toxic and Nontoxic Organic Compounds in Soils*, edited by M. R. Overcash, pp. 227–231. Ann Arbor Science Publishers, Ann Arbor, MI (1981).

5. G. L. Baughman, E. J. Weber, R. L. Adams, M. S. Brewer, *Fate of Colored Smoke Dyes*, National Technical Information Center Report No. AD-A249073. US Army, Fort Detrick, MD (1982).
6. L. P. Hammett, *Physical Organic Chemistry*, 2nd edn. McGraw-Hill, New York (1970).
7. M. J. S. Dewar and R. C. Doughetry, *The PMO Theory of Organic Chemistry*. Plenum Press, New York (1975).
8. E. P. Serjeant and B. Dempsey, *Ionization Constants of Organic Acids in Aqueous Solution*. Pergamon Press, Oxford (1979).
9. D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution*. Butterworth, London (1965); D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution: Supplement*. Butterworth, London (1972).
10. S. H. Hilal, L. A. Carreira, S. W. Karickhoff and C. M. Melton, in preparation.
11. H. Diehl and J. Ellingboe, *Anal. Chem.* **32**, 1120–1123 (1960).
12. G. Schwarzenbach and H. Flaschka, *Complexometric Titration*, 2nd edn, translated by H. M. N. H. Irving. Marcel Dekker, New York, (1969).
13. C. N. Reilley, A. J. Barnard, Jr, and R. Püschell, in *Handbook of Analytical Chemistry*, edited by L. Meites, pp. 76–226. McGraw-Hill, New York (1963).
14. D. G. McMinn, and B. Kratochvil, *Can. J. Chem.* **55**, 3909–3914 (1955).
15. H. Wada and G. Nakagawa, *Anal. Chim. Acta* **121**, 265–273 (1980).
16. W. R. Haag and T. Mill, *Environ. Toxicol. Chem.* **6**, 359–369 (1987).
17. T. Shibusawa, T. Saito and T. Hamayose, *Nippon Kagaku Kaishi*, **2**, 264–271 (1977).
18. K. M. Tawarah and H. M. Abu-Shamleh, *Dyes Pigm.* **17**, 203–215 (1991).
19. I. M. Korenman and B. A. Nikolaev, *Zh. Fiz. Khim.* **48**, 2545–2549 (1974).
20. E. Bishop, *Indicators*, 1st edn. Pergamon Press, Oxford (1972).
21. D. L. Zierath, J. J. Hassett and W. L. Banwart, *Soil Sci.* **129**, 277–281 (1980).
22. J. M. Zachara, C. C. Ainsworth, L. J. Felice and C. T. Resch, *Environ. Sci. Technol.* **20**, 620–627 (1986).
23. H. C. Sikka, H. T. Appleton and S. Banerjee, *Fate of 3,3'-Dichlorobenzidine in Aquatic Environments*, EPA-600/3-78-068. US Environmental Protection Agency, Athens, GA (1978).
24. S. W. Karickhoff, *J. Hydraul. Eng.* **110**, 707–735 (1984).
25. C.-P. C. Yen, T. A. Perenich and G. L. Baughman, *Environ. Toxicol. Chem.* **10**, 1009–1017 (1991).
26. E. J. Weber, *Environ. Toxicol. Chem.* **10**, 609–618 (1991).