Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids.

Benjamin G. Tehan[a], Edward J. Lloyd[a], Margaret G. Wong[b], Will R. Pitt[c], John G. Montana, David T. Manallack*[c] and Emanuela Gancia[c]

[a] Department of Medicinal Chemistry, Victorian College of Pharmacy, Monash University, 381 Royal Parade, Parkville, Australia.
[b] Department of Applied Chemistry, Swinburne University of Technology, John Street, Hawthorn, Australia.
[c] Celltech R&D Ltd., Granta Park, Great Abington, Cambridge, CB1 6GS, UK.

This paper is dedicated to the memory of Rich Green, who instigated this research project.

Abstract

The electronic properties of small molecules can be calculated quickly and with a reasonable degree of accuracy using semiempirical QM methods. In this study a set of QM properties derived from frontier electron theory have been used to produce a predictive model of the dissociation constants of phenols, benzoic acids and aliphatic carboxylic acids. The pK_a values and structures of nearly 500 compounds were extracted from the Physprop database for this purpose. Multiple linear regression was used to search for relationships between pK_a and the calculated QM properties. In most cases only a single independent variable, electrophilic superdelocalisability, was needed to produce a good model of pK_a. The advantages of our approach are in the speed of calculation and the simplicity of the resultant models. The merits of using semiempirical methods to predict pK_a are discussed in relation to previous studies.

1 Introduction

The pharmacokinetic properties of a compound can be strongly affected by its pK_a. This is particularly true for weak organic acids and bases since the protonation state of these molecules will be more readily affected at physiological pH levels. The protonation state will in turn influence the rate at which a compound diffuses across the many membranes and other physical barriers (e.g. the gut and BBB) that a drug may need to cross to reach its site of action. Active transport processes aside, the ability of a drug to diffuse across a membrane will depend upon its coefficient of partition between a lipid and an aqueous phase. In general, well absorbed drug molecules have high partition coefficients. In the case of an ionisable substance, its pK_a and the pH of the aqueous phase will affect the overall partition coefficient according to the following equation:

$$\log\left(\frac{P_N}{P_O} - 1\right) = pH - pK_a$$

where P_O is the observed overall partition coefficient and P_N the partition coefficient for the non-ionised form (note, this neglects the permeation of ions). More often than not, neutral molecules are more easily absorbed by membranes, while ionised molecules tend to remain in plasma or the gut.

A number of studies have been published reporting a correlation between pK_a and various biopharmaceutical parameters further demonstrating its relevance to ADME characteristics and biological properties. These studies include: the analgesic potency of a set of imidazoles [2];

* To receive all correspondence. Present address: De Novo Pharmaceuticals, Compass House, Vision Park, Histon, Cambridge, CB4 9ZR, UK.

Key words: frontier molecular orbital theory, quantum-mechanical descriptors, dissociation constants, pK_a.

Abbreviations: BBB, blood-brain barrier; ADMET, adsorption distribution metabolism excretion toxicity; SCRF, self consistent reaction field; FMO, frontier molecular orbital; HOMO, highest occupied molecular orbital; LUMO, lowest unoccupied molecular orbital; LFER, linear free energy relationships; QM, quantum mechanical; QSPR, quantitative structure-property relationships; FE, electrophilic frontier electron density; FN, nucleophilic frontier electron density; SE, electrophilic superdelocalisability; SN, nucleophilic superdelocalisability; SR, radical superdelocalisability; ALP, atom self-polarisability; AQ, atomic charge; MLR, multiple linear regression; r^2, squared correlation coefficient; r_{cv}^2, cross-validated squared correlation coefficient.

  0931-8771/02/0510-0457 $ 17.50+.50/0

the total clearance of six phenothiazine drugs [3]; the anticancer activity of a series of pyrrolidines, pyrrolines and piperidines [4]; the uncoupling potency of mitochondrial oxidative phosphorylation of non-steroidal anti-inflammatory drugs (NSAIDs) [5] and the mutagenic potency of six butenoic acids [6] to cite only a few representative examples.

While there are well established titrimetric methods [7] of measuring pK_a there is potentially great benefit to be derived from a method that can predict dissociation constants. The obvious advantage of predictive methods being that physical samples of compounds are not needed. A variety of theoretical prediction methods have been established based on LFER [8–11], semiempirical [12, 13] and *ab initio* [14–19] QM calculations or combinations of approaches [20]. LFER techniques such as those implemented in ACD Labs [8] pK_a predictor and pKalc [10] are very fast, although require the pre-derivation of many fragment constants and correction factors. If, however, molecules contain fragments not encountered in the training data, errors are likely to occur. The software package Jaguar [14] utilises *ab initio* QM calculations that employ self consistent reaction field (SCRF) continuum treatment of solvation and systematic corrections to predict pK_a. Recently, Gross and Seybold investigated the ability of *ab initio* QM parameters to model the pK_a of substituted anilines [15] and phenols [16] and compared the results to those achieved using Hammett constants. Whilst *ab initio* methods are very reliable, they are computationally expensive. Calculations can take up to half a day for a drug-sized molecule on a modern computer. Semiempirical QM methods have also been explored to study a set of phenols and carboxylic acids [12] using the energies of the HOMO and the LUMO, atomic charges and energy state differences. More recently, Citra [13] showed a correlation between Coulson charges and bond orders and the pK_a of sets of phenols, carboxylic acids and alcohols.

A recent study by Gancia and co-workers [21] employed parameters derived from frontier electron theory [22] to predict the hydrogen bond strengths of drug-like functional groups. Encouraged by these results, we envisaged that it should be possible to predict other related properties such as dissociation constants using the descriptors from frontier electron theory. We have applied semiempirical methods to the calculation of a set of molecular/atomic properties and these have been correlated with the pK_a of sets of phenols and carboxylic acids.

2 Methods

2.1 Datasets

Experimental pK_a values (unless otherwise stated) were taken from the Physprop database [23], which reports pK_a values for over 1,800 compounds (Tables 1–6). As our interest was in predicting the pK_a of molecules or fragments with relevance to the pharmaceutical industry, we applied a variety of filters in an attempt to remove non-druglike molecules from the dataset. These included: molecular weight cut-off of 700; removal of mixtures, salts and compounds with toxic functional groups. Finally only experimental data measured between 10° and 30 °C were used. This filtering resulted in a final set of 1,671 compounds from which three datasets were extracted consisting of phenols, benzoic acids and aliphatic acids. Care was taken to avoid ambiguities in pK_a assignment for compounds where more than one ionisable group was present. In these cases referral to the original references was undertaken. Similarly, compounds with two instances of the same functional group of interest were removed, except in cases where an identical electronic environment existed for both groups. A small number of compounds with unclear ionisation status of additional substituents have been included in our study (see footnotes in Tables 1-6). Although these structures may exist, at least partially, in a zwitterionic form, all the calculations have been performed on the neutral species.

2.2 Software

All the structures were initially extracted as 2D SDfiles and converted into 3D models using Corina [24]. The QM calculations were carried out with a modified version of Mopac 6.01 (Peter Bladon, Interchem Chemical Services, Glasgow), that is able to calculate frontier electron theory properties (keyword: PROPER) [25]. Sybyl [26] and Tsar [27] were used for data manipulation and statistical analyses, primarily using the default parameters.

2.3 Quantum Mechanical Descriptors

Each structure was fully optimised (EF routine, PRECISE), prior to any parameter calculation using the AM1 [28] Hamiltonian within Mopac. A set of properties, derived from frontier electron theory [22], were computed from the eigenvectors $c_{\alpha j}$ and the eigenvalues λ_j where α refers to the atomic orbital (i.e. s, p_x, p_y and p_z) and j to the molecular orbital. Given a molecule with N molecular orbitals, whose levels from 1 to m are occupied, and an atom p with q atomic orbitals, these properties are defined as follows.

Electrophilic frontier electron density (FE)

$$\mathrm{FE(p)} = \sum_{\alpha=1,\mathrm{q}} \mathrm{c}^2_{\alpha\mathrm{m}}$$

is the sum of all the squared eigenvectors of p on the HOMO.

Nucleophilic frontier electron density (FN)

$$\mathrm{FN(p)} = \sum_{\alpha=1,\mathrm{q}} \mathrm{c}^2_{\alpha(\mathrm{m}+1)}$$

is the sum of all the squared eigenvectors of p on the LUMO.

Electrophilic superdelocalisability (SE)

$$SE(p) = 2 * \sum_{j=1,m} \sum_{\alpha=1,q} \left(c_{\alpha j}^2 / \lambda_j \right)$$

where, the sum is over all the atomic orbitals of *p* and all the occupied molecular orbitals.

Nucleophilic superdelocalisability (SN)

$$SN(p) = 2 * \sum_{j=m+1,N} \sum_{\alpha=1,q} \left(c_{\alpha j}^2 / -\lambda_j \right)$$

where, the sum is over all the atomic orbitals of *p* and all the unoccupied molecular orbitals.

Radical superdelocalisability (SR)

$$SN(p) = \sum_{j=1,m} \sum_{\alpha=1,q} \left(c_{\alpha j}^2 / \lambda_j \right) + \sum_{j=m+1,N} \sum_{\alpha=1,q} \left(c_{\alpha j}^2 / -\lambda_j \right).$$

Atom self-polarisability (ALP)

$$ALP(p) = -4 * \sum_{j=1,m} \sum_{k=m+1,N} \sum_{\alpha=1,q} \left(c_{\alpha j}^2 * c_{\alpha k}^2 / \lambda_j - \lambda_k \right)$$

where, the sum is over all the atomic orbitals of *p* and all the occupied (*j*) and unoccupied molecular orbitals (*k*).

All of the above properties were calculated on and around the atoms in the functional group of interest (Figure 1). In addition, partial atomic charges (AQ) were obtained as well as the energies of the HOMO and LUMO.

3 Results

The three main datasets were analysed in the following manner. Correlation matrices between experimentally determined pK_a values and all the QM parameters were initially calculated and viewed within Tsar [27]. Scatter plots of experimental pK_a values versus the most highly correlated properties were then analysed and regression equations were derived for these single descriptors. Finally by applying MLR on all of the descriptors further equations were derived containing up to four descriptors. The most significant findings are reported here.

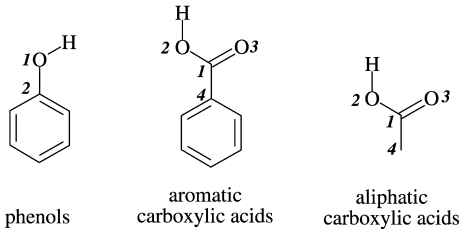

Figure 1. Numbering scheme used for each functional group.

3.1 Phenols

The final phenol dataset consisted of 175 compounds. These molecules contained a wide range of chemically diverse electron-withdrawing and electron-donating substituents, covering a pK_a range of more than 12 log units. The CAS numbers and names of all the phenols together with their experimental pK_a values are reported in Tables 1, 2 and 3.

A number of calculated properties showed a good correlation to pK_a, the highest correlation ($r^2 = 0.81$) being observed for the electrophilic superdelocalisability of the phenolic oxygen, SE_1 (atom 1, Figure 1). The scatter plot of observed pK_a versus SE_1 is shown in Figure 2. The best equations modelling phenols are reported in Table 7 (Eq. 1 and Eq. 2). Eq. 1 contains one term only, SE_1, whereas Eq. 2 was derived using four terms. An increase in both the squared correlation coefficient (r^2) and the cross – validated squared correlation coefficient (r_{cv}^2) is observed going from one to four descriptors. It is interesting to note that SE_1 was replaced by SE_2 in the second equation, but this is entirely feasible as the two terms are strongly correlated ($r = 0.97$).

As previous work [12] using QM descriptors to explain pK_a's had shown that improvements could be made in dividing the data into smaller subsets according to the positions of the substituents on the aromatic ring, we adopted a similar approach. The dataset was divided into a meta/para-substituted phenol dataset of 58 compounds and an ortho-substituted phenol dataset of 117 compounds. The ortho subset was further split into two smaller classes: *(i)* intra-molecularly hydrogen bonded compounds (26 molecules) and *(ii)* the remaining ortho compounds (91 molecules). Compounds were chosen for inclusion in the former class by visual inspection. Nitro groups were not included in those groups considered able to form intramolecular hydrogen bonds as these are poor hydrogen bond acceptors [29].

A clear advantage in splitting the phenols into subsets is that a single parameter is sufficient to model pK_a (Eqs. 3 – 5, Table 7). The three equations all have reasonable statistics (r^2 ranging from 0.83 to 0.92) and they are all based on the same parameter (SE_1) with slightly different coefficients. The plot of observed pK_a versus SE_1 for all the phenols, reported in Figure 2, demonstrates the separation between the subclasses and further justifies splitting the dataset. The pK_a values, calculated using Eq. 3, for the 58 meta/para-substituted phenols are reported in Table 1. All the calculated values are within 1 unit of the experimental pK_a. The calculated pK_a for the 26 ortho-substituted phenols able to form an intra-molecular hydrogen bond are listed in Table 2, and the calculated pK_a for the other 91 ortho-substituted phenols are listed in Table 3. There was only one compound with a residual greater than 2 log units: methyl salicylate (000119-36-8, no. **3**, Table 2) and no obvious explanation for this large error could be provided.

Table 1. Observed and calculated pK_a values for the meta/para-substituted phenols dataset using Eq. 3 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_1
1	000051-67-2	4-(2-aminoethyl)phenol†	9.77	9.88	– 9.55
2	000059-50-7	3-methyl-4-chlorophenol	9.20*	9.31	– 9.44
3	000080-46-6	4-tert-amylphenol	10.43	10.09	– 9.59
4	000088-30-2	3-trifluoromethyl-4-nitrophenol	6.07	6.18	– 8.84
5	000088-04-0	4-chloro-3,5-dimethyl phenol	9.70	9.47	– 9.47
6	000092-69-3	4-phenylphenol	9.55	9.73	– 9.52
7	000094-26-8	4-hydroxy butyl benzoate	8.47	8.58	– 9.30
8	000095-77-2	3,4-dichlorophenol	8.63	8.63	– 9.31
9	000095-65-8	3,4-dimethylphenol	10.36	10.14	– 9.60
10	000098-54-4	4-t-butylyphenol	10.39	10.09	– 9.59
11	000098-17-9	3-trifluoromethylphenol	8.95	8.48	– 9.28
12	000099-93-4	4-hydroxyacetophenone	8.05	8.74	– 9.33
13	000099-89-8	4-isopropylphenol	10.24	10.04	– 9.58
14	000100-83-4	3-hydroxybenzaldehyde	8.98	9.00	– 9.38
15	000100-02-7	4-nitrophenol	7.15	7.17	– 9.03
16	000103-90-2	n-(4-hydroxyphenyl)acetamide	9.38	9.20	– 9.42
17	000106-48-9	4-chlorophenol	9.41	9.15	– 9.41
18	000106-44-5	4-cresol	10.26	10.04	– 9.58
19	000106-41-2	4-bromophenol	9.17	8.94	– 9.37
20	000108-95-2	phenol	9.99	9.93	– 9.56
21	000108-68-9	3,5-dimethylphenol	10.19	10.14	– 9.60
22	000108-43-0	3-chlorophenol	9.12	9.15	– 9.41
23	000108-39-4	3-cresol	10.09	10.04	– 9.58
24	000120-47-8	4-hydroxybenzoic acid, ethyl ester	8.34	8.58	– 9.30
25	000121-71-1	3-hydroxyacetophenone	9.25	9.36	– 9.45
26	000123-30-8	4-aminophenol	10.45	10.51	– 9.67
27	000123-08-0	4-hydroxybenzaldehyde	7.61	8.58	– 9.30
28	000123-07-9	4-ethylphenol	10.00	10.04	– 9.58
29	000150-76-5	4-methoxyphenol	10.10	10.14	– 9.60
30	000150-19-6	3-methoxyphenol	9.65	9.78	– 9.53
31	000371-41-5	4-fluorophenol	9.91	9.26	– 9.43
32	000372-20-3	3-fluorophenol	9.21	9.05	– 9.39
33	000402-45-9	4-trifluoromethylphenol	8.68	8.11	– 9.21
34	000500-99-2	3,5-dimethoxyphenol	9.34	9.41	– 9.46
35	000540-38-5	4-iodophenol	9.21	8.94	– 9.37
36	000554-84-7	3-nitrophenol	8.36	7.80	– 9.15
37	000577-71-9	3,4-dinitrophenol	5.42	5.66	– 8.74
38	000580-51-8	3-phenylphenol	9.64	9.88	– 9.55
39	000585-34-2	3-(1,1-dimethylethyl)-phenol	10.12	10.14	– 9.60
40	000586-11-8	3,5-dinitrophenol	6.69	6.08	– 8.82
41	000591-35-5	3,5-dichlorophenol	8.18	8.53	– 9.29
42	000591-27-5	3-aminophenol	9.86	10.09	– 9.59
43	000591-20-8	3-bromophenol	9.03	9.15	– 9.41
44	000609-19-8	3,4,5-trichlorophenol	7.84	8.11	– 9.21
45	000618-45-1	3-isopropylphenol	10.16	10.09	– 9.59
46	000620-17-7	3-ethylphenol	9.90	10.04	– 9.58
47	000621-34-1	3-ethoxyphenol	9.65	9.83	– 9.54
48	000622-62-8	4-ethoxyphenol	10.13	10.25	– 9.62
49	000626-41-5	3,5-dibromophenol	8.06	8.48	– 9.28
50	000626-02-8	3-iodophenol	9.03	9.15	– 9.41
51	000645-56-7	4-propylphenol	10.34	10.04	– 9.58
52	000698-71-5	3-ethyl-5-methylphenol	10.10	10.14	– 9.60
53	000767-00-0	4-cyanophenol	7.97	8.32	– 9.25
54	000873-62-1	3-cyanophenol	8.61	8.68	– 9.32
55	001073-72-9	4-methiophenol	9.53	9.57	– 9.49
56	001470-94-6	5-indanol	10.32	10.19	– 9.61
57	002042-14-0	3-nitro-4-cresol	8.62	8.06	– 9.20
58	007339-87-9	hydroxyacetophenone	8.05	8.74	– 9.33

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.
* Value taken from Albert, A., and Serjeant, E.P., *The Determination of Ionization Constants: A Laboratory Manual,* 3rd ed., Chapman and Hall, London 1984.

Table 2. Observed and calculated pK_a values for the ortho-substituted phenols dataset (capable of forming internal hydrogen bonds) using Eq. 4 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_1
1	000065-45-2	2-hydroxybenzamide	8.89*	9.10	– 9.57
2	000087-17-2	salicylanilide	7.40	8.19	– 9.49
3	000119-36-8	methyl salicylate	9.87	7.74	– 9.45
4	000148-53-8	2-vanillin	7.91	6.83	– 9.37
5	001151-51-5	3,5,4′-trichloro salicylanilide	4.70	5.25	– 9.23
6	001697-18-3	2′-chloro salicylanilide	7.31	7.85	– 9.46
7	002389-37-9	5-nitro salicylanilide	3.03	2.76	– 9.01
8	002627-77-2	4′-bromo salicylanilide	7.31	7.40	– 9.42
9	003679-64-9	4′- chloro-5-bromo salicylanilide	6.00	5.81	– 9.28
10	003679-63-8	4′-chloro salicylanilide	7.30	7.51	– 9.43
11	004214-48-6	3,5-dichloro salicylanilide	4.70	5.70	– 9.27
12	004638-48-6	5-chlorosalicylanilide	6.17	6.72	– 9.36
13	025933-30-6	5- chloro-2′-methyl salicylanilide	6.60	6.83	– 9.37
14	037183-28-1	2′,4′-dichloro salicylanilide	7.14	7.29	– 9.41
15	037183-26-9	2′-nitro salicylanilide	6.91	7.06	– 9.39
16	037399-40-9	2′-nitro-4′-chloro salicylanilide	6.74	6.61	– 9.35
17	040912-87-6	5-bromo-2-hydroxy-n,3-dimethyl-benzamide	7.52	7.40	– 9.42
18	054850-02-1	3,5-dichloro-4′-fluoro salicylanilide	4.80	5.25	– 9.23
19	072699-09-3	3,5-dibromo-2′-nitro-4′-chloro salicylanilide	4.11	3.66	– 9.09
20	077067-91-5	2′-methyl-4′-chloro salicylanilide	7.43	7.63	– 9.44
21	077068-04-3	5-fluoro-2′-methyl-4′-bromo salicyanilide	7.10	6.27	– 9.32
22	077068-02-1	5-fluoro-2′-methyl-4′-chloro salicyanilide	7.30	6.49	– 9.34
23	079402-07-6	3,5-dibromo-2′,4′-difluoro salicylanilide	4.77	4.34	– 9.15
24	080033-99-4	3,5-dichloro-2′,4′-difluoro salicylanilide	4.77	5.02	– 9.21
25	090426-05-4	3,5,-4′-trichloro-4′-nitro salicylanilide	4.11	4.34	– 9.15
26	090426-03-2	3,5-dichloro-2′-methyl-4′-nitro salicylanilide	4.41	5.25	– 9.23

* Value taken from Albert, A., and Serjeant, E.P., *The Determination of Ionization Constants: A Laboratory Manual*, 3rd ed., Chapman and Hall, London 1984.

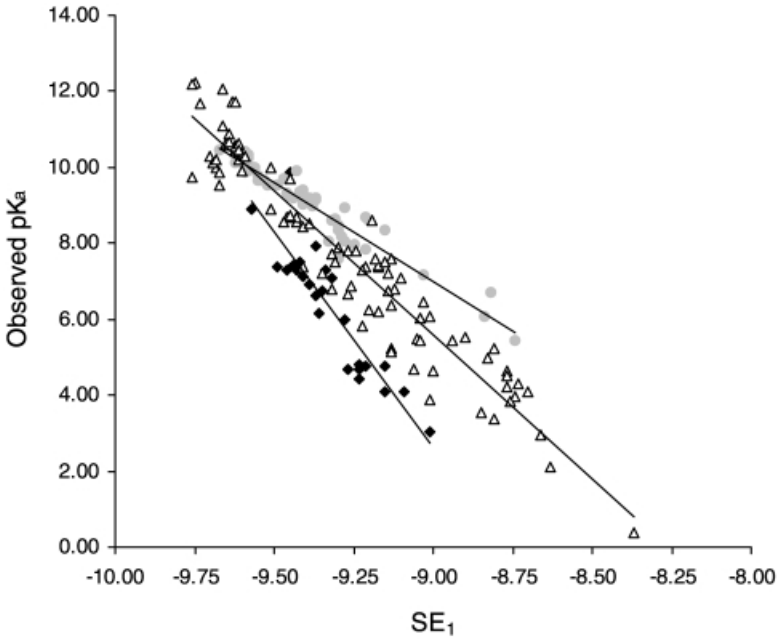

Figure 2. Plot of observed pK_a vs. SE_1 for the phenol dataset: meta/para-substituted phenols (●), phenols forming an internal hydrogen bond (◆) and other ortho-substituted phenols (▽).

Table 3. Observed and calculated pK_a values for the ortho-substituted phenols dataset (excluding those forming intra-molecular hydrogen bonds) using Eq. 5 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_1
1	000051-28-5	2,4-dinitrophenol	4.09	3.30	– 8.70
2	000058-90-2	2,3,4,6-tetrachlorophenol	5.22	6.56	– 9.13
3	000066-56-8	2,3-dinitrophenol	4.96	4.28	– 8.83
4	000087-86-5	pentachlorophenol	4.70	6.03	– 9.06
5	000087-65-0	2,6-dichlorophenol	6.79	7.99	– 9.32
6	000087-64-9	2-methyl-6-chlorophenol	8.69	9.05	– 9.46
7	000088-89-1	2,4,6-trinitrophenol	0.38	0.80	– 8.37
8	000088-87-9	4-chloro-2,6-dinitrophenol	2.96	3.00	– 8.66
9	000088-85-7	2-sec-butyl-4,6-dinitrophenol	4.62	3.83	– 8.77
10	000088-75-5	2-nitrophenol	7.23	6.63	– 9.14
11	000088-69-7	2-isopropylphenol	10.47	10.26	– 9.62
12	000088-18-6	2-t-butylphenol	10.28	10.11	– 9.60
13	000088-06-2	2,4,6-trichlorophenol	6.23	7.09	– 9.20
14	000089-83-8	thymol	10.62	10.49	– 9.65
15	000089-68-9	chlorothymol	9.98	9.43	– 9.51
16	000089-64-5	4-chloro-2-nitrophenol	6.46	5.80	– 9.03
17	000090-43-7	2-phenylphenol	9.92	10.11	– 9.60
18	000090-05-1	2-methoxyphenol	9.98	10.72	– 9.68
19	000090-00-6	2-ethylphenol	10.20	10.19	– 9.61
20	000093-51-6	4-methyl-2-methoxyphenol	10.28	10.87	– 9.70
21	000094-71-3	2-ethoxyphenol	10.11	10.79	– 9.69
22	000095-95-4	2,4,5-trichlorophenol	7.40	7.16	– 9.21
23	000095-87-4	2,5-dimethylphenol	10.41	10.26	– 9.62
24	000095-57-8	2-chlorophenol	8.56	8.83	– 9.43
25	000095-56-7	2-bromophenol	8.45	8.67	– 9.41
26	000095-55-6	2-aminophenol	9.75	11.32	– 9.76
27	000095-48-7	2-cresol	10.28	10.04	– 9.59
28	000096-76-4	2,4-di-t-butylphenol	11.72	10.34	– 9.63
29	000097-54-1	2-methoxy-4-(1-propenyl)phenol	9.88	10.64	– 9.67
30	000097-53-0	eugenol	10.19	10.72	– 9.68
31	000098-28-2	4-(tert-butyl)-2-chlorophenol	8.58	9.13	– 9.47
32	000098-27-1	4-(t-butyl)-2-cresol	10.59	10.26	– 9.62
33	000099-57-0	2-amino-4-nitrophenol	7.60	6.56	– 9.13
34	000099-28-5	2,6-dibromo-4-nitrophenol	3.39	4.13	– 8.81
35	000105-67-9	2,4-dimethylphenol	10.60	10.19	– 9.61
36	000118-79-6	2,4,6-tribromophenol	6.80	6.48	– 9.12
37	000119-34-6	phenol, 4-amino-2-nitro-	7.81	7.62	– 9.27
38	000119-33-5	4-methyl-2-nitrophenol	7.40*	6.86	– 9.17
39	000120-83-2	2,4-dichlorophenol	7.89	7.84	– 9.30
40	000121-33-5	vanillin	7.40	8.67	– 9.41
41	000128-39-2	2,6-di-t-butylphenol	11.70	11.10	– 9.73
42	000128-37-0	2,6-di-t-butyl-4-methylphenol (bht)	12.23	11.25	– 9.75
43	000131-89-5	2-cyclohexyl-4,6-dinitrophenol	4.52	3.83	– 8.77
44	000329-71-5	2,5-dinitrophenol	5.21	4.13	– 8.81
45	000367-12-4	2-fluorophenol	8.70	8.83	– 9.43
46	000446-36-6	5-fluoro-2-nitrophenol	6.07	5.65	– 9.01
47	000496-78-6	2,4,5-trimethylphenol	10.57	10.42	– 9.64
48	000526-75-0	2,3-dimethylphenol	10.54	10.26	– 9.62
49	000527-60-6	2,4,6-trimethylphenol	10.86	10.42	– 9.64
50	000533-58-4	2-iodophenol	8.51	8.52	– 9.39
51	000534-52-1	4,6-dinitro-o-cresol	4.31	3.53	– 8.73
52	000554-52-9	2-methyldopamine†	9.54	10.64	– 9.67
53	000573-56-8	2,6-dinitrophenol	3.97	3.60	– 8.74
54	000576-26-1	2,6-dimethylphenol	10.62	10.19	– 9.61
55	000576-24-9	2,3-dichlorophenol	7.70	7.99	– 9.32
56	000583-78-8	2,5-dichlorophenol	7.51	7.92	– 9.31
57	000603-86-1	6-chloro-2-nitrophenol	5.48	5.95	– 9.05
58	000608-71-9	pentabromophenol	4.62	5.57	– 9.00
59	000608-33-3	2,6-dibromophenol	6.67	7.62	– 9.27
60	000609-93-8	2,6-dinitro-p-cresol	4.23	3.83	– 8.77
61	000611-20-1	2-cyanophenol	6.86	7.54	– 9.26

Table 3. (cont.)

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_1
62	000611-07-4	5-chloro-2-nitrophenol	6.05	5.87	−9.04
63	000615-58-7	2,4-dibromophenol	7.79	7.39	−9.24
64	000618-80-4	2,6-dichloro-4-nitrophenol	3.55	4.44	−8.85
65	000619-08-9	2-chloro-4-nitrophenol	5.45	5.12	−8.94
66	000621-59-0	isovanillin	8.89	9.43	−9.51
67	000644-35-9	2-propylphenol	10.47	10.19	−9.61
68	000697-82-5	2,3,5-trimethylphenol	10.67	10.42	−9.64
69	000700-38-9	5-methyl-2-nitrophenol	7.41	6.86	−9.17
70	000731-92-0	2,4-dinitro-6-phenylphenol	3.85	3.75	−8.76
71	000732-26-3	2,4,6-tri(tert-butyl)phenol	12.19	11.32	−9.76
72	000771-61-9	pentafluorophenol	5.53	4.81	−8.90
73	000885-82-5	4-phenyl-2-nitrophenol	6.73	6.63	−9.14
74	000933-75-5	2,3,6-trichlorophenol	5.80	7.24	−9.22
75	000935-95-5	2,3,5,6-tetrachlorophenol	5.14	6.56	−9.13
76	000946-31-6	6-chloro-2,4-dinitrophenol	2.10	2.77	−8.63
77	001568-70-3	4-methoxy-2-nitrophenol	7.31	7.24	−9.22
78	001570-64-5	2-methyl-4-chlorophenol	9.71	8.98	−9.45
79	001689-84-5	bromoxynil	3.86	5.65	−9.01
80	001879-09-0	2-(1,1-dimethylethyl)-4,6-dimethylphenol	12.04	10.57	−9.66
81	002078-54-8	phenol, 2,6-bis(1-methylethyl)-	11.10	10.57	−9.66
82	002409-55-4	2-(tert-butyl)-4-methylphenol	11.72	10.26	−9.62
83	002423-71-4	2,6-dimethyl-4-nitrophenol	7.07	6.33	−9.10
84	002432-12-4	4-methyl-2,6-dichlorophenol	7.19	8.22	−9.35
85	003217-15-0	4-bromo-2,6-dichlorophenol	6.21	6.86	−9.17
86	003555-18-8	4-(sec-butyl)-2-nitrophenol	7.59	6.93	−9.18
87	003964-58-7	3-chloro-4-hydroxybenzoic acid	7.52	6.71	−9.15
88	004901-51-3	2,3,4,5-tetrachlorophenol	6.35	6.56	−9.13
89	005428-54-6	phenol, 2-methyl-5-nitro-	8.59	7.01	−9.19
90	006640-27-3	2-methyl-4-chlorophenol	8.74	8.98	−9.45
91	013181-17-4	bromofenoxim	5.46	5.87	−9.04

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.
* Value taken from Rapoport, M., Kinney Hancock, C., and Meyers, E.A. *J. Am. Chem. Soc. 83*, 3489–3494 (1961).

3.2 Aromatic Carboxylic Acids

The benzoic acid dataset consisted of 99 compounds, covering a pK_a range of about 5 log units. The CAS numbers, names and corresponding experimental pK_a values are reported in Tables 4 and 5.

As in the previous set, the electrophilic superdelocalisability calculated on atoms 1, 2 and 3 (Figure 1) showed a good correlation to pK_a. The order of correlation with pK_a was $SE_1 > SE_3 > SE_2$ with values of $0.81 > 0.78 > 0.72$, respectively. Further analyses have employed the electrophilic superdelocalisability of atom 3 and it should be noted that SE_1 and SE_3 are highly correlated ($r = 0.95$).

As for the phenol dataset, the aromatic carboxylic acids were divided into a meta/para-substituted set and an ortho-substituted set of 46 and 53 compounds, respectively. In the case of benzoic acids, the ortho substituents able to form a hydrogen bond did not seem to fall into a separate class (this finding is in agreement with that reported in Ref. [12]). These compounds were therefore studied together with the other ortho acids. Both the meta/para and ortho subsets could be modelled reasonably well by using the SE_3 parameter. The SE_3 equation for meta/para-substituted benzoic acids gave an r^2 of 0.75 (Eq. 8, Table 7), while the best model for ortho-substituted acids gave an r^2 of 0.79 (Eq. 10, Table 7). The scatter plot of observed pK_a versus SE_3 is reported in Figure 3 and the calculated pK_a values are listed in Tables 4 and 5.

All the calculated pK_a's of the meta/para benzoic acids are within 1 unit of the measured values with the exception of 3,4-di-amino-benzoic acid (000619-05-6, no. **39** in Table 4) which had a residual value of −1.01. Removal of this compound, improved the r^2 to 0.86 (Eq. 9, Table 7).

3.3 Aliphatic Carboxylic Acids

The aliphatic carboxylic acid dataset consisted of 185 compounds, covering a pK_a range of just over 5 log units. The correlation matrix for all of the aliphatic carboxylic acids showed that there were only poor correlations between pK_a and any of the descriptors. There were tenuous correlations with both SE_2 and SE_3 with a correlation coefficient around 0.5-0.6. The best model that could be derived for all 185 compounds was a four-descriptor model (Eq. 11, Table 7) indicating that the set may be too diverse for analysis on its own and may benefit from further

Table 4. Observed and calculated pK_a values for the meta/para-substituted benzoic acids dataset using Eq. 8 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pKa	SE_3
1	000051-44-5	3,4-dichlorobenzoic acid	3.64	3.79	– 10.31
2	000051-36-5	3,5-dichlorobenzoic acid	3.54	3.75	– 10.29
3	000057-66-9	probenecid	3.40	3.65	– 10.24
4	000062-23-7	p-nitrobenzoic acid	3.44	3.33	– 10.08
5	000065-85-0	benzoic acid	4.19	4.21	– 10.52
6	000074-11-3	4-chlorobenzoic acid	3.98	3.99	– 10.41
7	000093-09-4	2-naphthoic acid	4.17	4.23	– 10.53
8	000093-07-2	3,4-dimethoxybenzoic acid	4.36	4.30	– 10.57
9	000098-73-7	4-(tert-butyl)- benzoic acid	4.40	4.32	– 10.58
10	000099-96-7	p-hydroxybenzoic acid	4.54	4.28	– 10.56
11	000099-94-5	p-toluic acid	4.37	4.30	– 10.57
12	000099-50-3	3,4-dihydroxybenzoic acid	4.48*	4.15	– 10.49
13	000099-34-3	3,5-dinitrobenzoic acid	2.82	2.81	– 9.818
14	000099-10-5	3,5-dihydroxybenzoic acid	4.04	3.97	– 10.4
15	000099-05-8	3-aminobenzoic acid†	4.74	4.28	– 10.56
16	000099-04-7	m-toluic acid	4.27	4.25	– 10.54
17	000100-09-4	p-methoxybenzoic acid	4.47	4.36	– 10.6
18	000121-92-6	m-nitrobenzoic acid	3.46	3.43	– 10.13
19	000121-34-6	4-hydroxy-3-methoxybenzoic acid	4.51	4.23	– 10.53
20	000149-91-7	3,4,5-trihydroxybenzoic acid	4.21**	3.93	– 10.38
21	000150-13-0	4-aminobenzoic acid†	4.85	4.62	– 10.73
22	000455-38-9	m-fluorobenzoic acid	3.86	3.91	– 10.37
23	000456-22-4	p-fluorobenzoic acid	4.14	4.01	– 10.42
24	000528-45-0	3,4-dinitrobenzoic acid	2.82	2.77	– 9.796
25	000530-57-4	4-hydroxy-3,5-dimethoxybenzioc acid	4.34	4.15	– 10.49
26	000535-80-8	m-chlorobenzoic acid	3.81	3.97	– 10.4
27	000536-66-3	cumic acid	4.35	4.30	– 10.57
28	000585-76-2	m-bromobenzoic acid	3.81	3.95	– 10.39
29	000586-89-0	p-acetylbenzoic acid	3.70	3.87	– 10.35
30	000586-76-5	p-bromobenzoic acid	4.00	3.93	– 10.38
31	000586-38-9	m-methoxybenzoic acid	4.09	4.19	– 10.51
32	000618-51-9	3-iodobenzoic acid	3.85	3.97	– 10.4
33	000619-86-3	p-ethoxybenzoic acid	4.45***	4.40	– 10.62
34	000619-66-9	4-formylbenzoic acid	3.77	3.85	– 10.34
35	000619-65-8	p-cyanobenzoic acid	3.55	3.71	– 10.27
36	000619-64-7	4-ethylbenzoic acid	4.35	4.28	– 10.56
37	000619-58-9	4-iodobenzoic acid	4.00	3.91	– 10.37
38	000619-21-6	3-formylbenzoic acid	3.84	3.85	– 10.34
39	000619-05-6	3,4-diamino-benzoic acid†	3.49	4.50	– 10.67
40	001132-21-4	3,5-dimethoxybenzoic acid	3.97	4.19	– 10.51
41	001877-72-1	m-cyanobenzoic acid	3.60	3.75	– 10.29
42	002215-77-2	p-phenoxybenzoic acid	4.52	4.28	– 10.56
43	003739-38-6	m-phenoxybenzoic acid	3.92	4.09	– 10.46
44	004052-30-6	p-methylsulfonylbenzoic acid	3.64	3.49	– 10.16
45	005438-19-7	4-propoxybenzoic acid	4.46***	4.40	– 10.62
46	007496-53-9	4-[(acetylamino)amino]-benzoic acid	4.20	4.52	– 10.68

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.
* Value taken from Shorter, J., and Stubbs, F.J. *J. Chem. Soc.* 1180 (1949).
** Value taken from http://www.sirius-analytical.com.
*** Value taken from Brown, H.C. *et al.*, in *Determination of Organic Structures by Physical Methods*; E. A. Braude, F. C. Nachod (Eds.); Academic Press, New York 1995; Cavill, G. W. K., Gibson, N. A., and Nyholm, R. S., *J. Chem. Soc.* 2466 (1949).

subdivision. Unlike the phenol and benzoic acid datasets, which could be split according to the position of ring substituents, the aliphatic acids set did not present simple and sensible ways of dividing the dataset. Moreover, the inspection of the scatter plots of each single QM parameter versus pK_a did not suggest any obvious way of dividing the aliphatic acids into sub-classes (c.f. phenols). As the amino acids seemed to be grouped separately in some of the plots and due to their zwitterionic character, we removed these compounds to see whether we could improve the correlation for the remaining 143 acids. The CAS numbers, names and experimental pK_a values for these compounds are reported in Table 6.

Table 5. Observed and calculated pK_a values for the ortho-substituted benzoic acids dataset using Eq. 10 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_3
1	000050-85-1	4-methylsalicylic acid	3.40	3.21	– 10.44
2	000050-84-0	2,4-dichlorobenzoic acid	2.68	2.85	– 10.34
3	000050-79-3	2,5-dichlorobenzoic acid	2.47	2.82	– 10.33
4	000050-78-2	acetylsalicylic acid	3.49	3.31	– 10.47
5	000050-31-7	2,3,6-trichlorobenzoic acid	1.50	2.33	– 10.19
6	000050-30-6	2,6-dichlorobenzoic acid	1.59	2.64	– 10.28
7	000059-07-4	2-ethoxy-4-aminobenzoic acid†	5.09	4.96	– 10.94
8	000061-68-7	mefenamic acid	4.20	3.87	– 10.63
9	000065-49-6	p-aminosalicylic acid†	3.66	3.84	– 10.62
10	000069-72-7	salicylic acid	2.97	3.07	– 10.40
11	000083-40-9	3-methylsalicylic acid	2.95	3.10	– 10.41
12	000088-67-5	2-iodobenzoic acid	2.93	2.68	– 10.29
13	000088-65-3	o-bromobenzoic acid	2.88	3.10	– 10.41
14	000089-86-1	2,4-dihydroxybenzoic acid	3.11	3.21	– 10.44
15	000089-56-5	5-methylsalicylic acid	3.15	3.14	– 10.42
16	000089-55-4	5-bromosalicylic acid	2.66	2.19	– 10.15
17	000089-52-1	n-acetyl o-aminobenzoic acid	3.40	3.00	– 10.38
18	000091-52-1	2,4-dimethoxybenzoic acid	4.36	4.40	– 10.78
19	000091-40-7	n-phenyl o-aminobenzoic acid	3.99	3.77	– 10.60
20	000092-70-6	2-naphthalenecarboxylic acid, 3-hydroxy-	2.79	2.82	– 10.33
21	000096-97-9	5-nitrosalicylic acid	2.12	1.70	– 10.01
22	000099-60-5	2-chloro-4-nitro-benzoic acid	2.14	1.73	– 10.02
23	000118-92-3	2-aminobenzoic acid	4.95	4.08	– 10.69
24	000118-91-2	2-chlorobenzoic acid	2.89	3.21	– 10.44
25	000118-90-1	o-toluic acid	3.98	3.63	– 10.56
26	000119-30-2	2-hydroxy-5-iodo-benzoic acid	2.62	2.61	– 10.27
27	000129-66-8	2,4,6-trinitrobenzoic acid	0.65	0.75	– 9.74
28	000133-91-5	3,5-diiodosalicylic acid	2.30	2.22	– 10.16
29	000133-90-4	3-amino-2,5-dichlorobenzoic acid†	3.40	2.78	– 10.32
30	000303-38-8	2,3-dihydroxybenzoic acid	2.91	2.71	– 10.30
31	000303-07-1	2,6-dihydroxybenzoic acid	1.05	2.50	– 10.24
32	000321-14-2	5-chlorosalicylic acid	2.65	2.64	– 10.28
33	000445-29-4	2-fluorobenzoic acid	3.27	3.28	– 10.46
34	000490-79-9	2,5-dihydroxybenzoic acid	2.95	2.89	– 10.35
35	000552-16-9	2-nitrobenzoic acid	2.17*	2.26	– 10.17
36	000577-56-0	o-acetylbenzoic acid	4.13	3.00	– 10.38
37	000579-75-9	o-methoxybenzoic acid	3.90	4.15	– 10.71
38	000609-99-4	2-hydroxy-3,5-dinitro-benzoic acid	0.70	0.60	– 9.70
39	000610-30-0	2,4-dinitrobenzoic acid	1.42	1.09	– 9.84
40	000632-46-2	2,6-dimethylbenzoic acid	3.35	3.59	– 10.55
41	000652-32-4	2,3,5,6-tetrafluoro-4-methyl-benzoic acid	2.00	2.33	– 10.19
42	000947-84-2	[1,1′-biphenyl]-2-carboxylic acid	3.46	3.21	– 10.44
43	001466-76-8	2,6-dimethoxybenzoic acid	3.44	4.19	– 10.72
44	001521-38-6	2,3-dimethoxybenzoic acid	3.98	3.52	– 10.53
45	001918-00-9	3,6-dichloro-2-methoxybenzoic acid	1.97	2.64	– 10.28
46	002243-42-7	o-phenoxybenzoic acid	3.53	3.84	– 10.62
47	002438-04-2	o-isopropylbenzoic acid	3.63	3.42	– 10.50
48	002516-96-3	2-chloro-5-nitrobenzoic acid	2.17	1.80	– 10.04
49	003970-35-2	2-chloro-3-nitrobenzoic acid	2.02**	1.77	– 10.03
50	004727-29-1	n-phenylphthalamic acid	2.50	2.12	– 10.13
51	005344-49-0	2-chloro-6-nitro-benzoic acid	1.34	2.26	– 10.17
52	021327-86-6	2-chloro-6-methyl-benzoic acid	2.75	2.96	– 10.37
53	25784-02-5	2-[(acetylamino)amino]-benzoic acid	4.20	4.05	– 10.68

† This compound may be partially in a zwitterionic form, however, it has been modelled in the neutral form.
* Value taken from Albert, A., Serjeant, E.P. T*he Determination of Ionization Constants: A Laboratory Manual*; 3rd ed., Chapman and Hall, London 1984.
** Value from Dippy, J.F.J., and Hughes, S.R.H., *Tetrahedron, 19*, 1527 (1963).

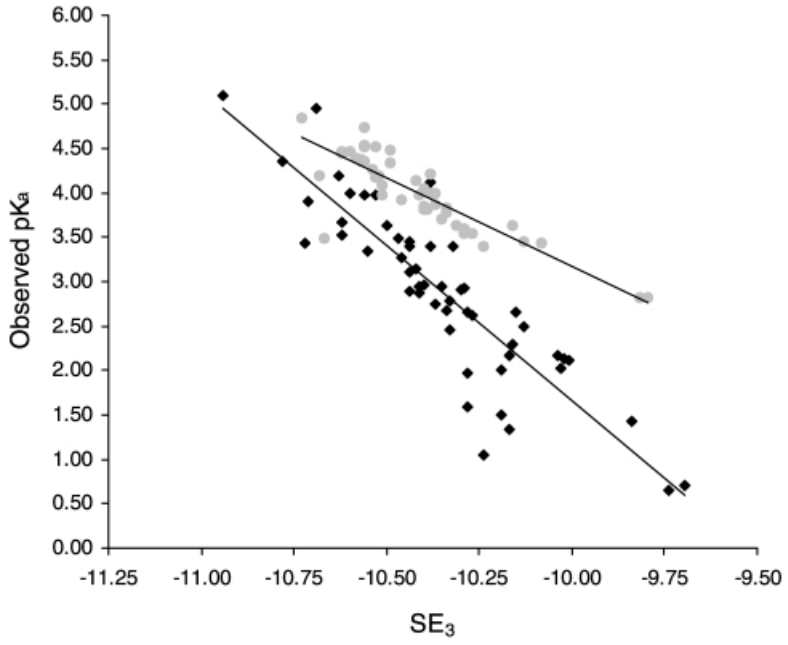

Figure 3. Plot of observed pK_a vs. SE_1 for the aromatic acid dataset: meta/para-substituted acids (●) and ortho-substituted acids (◆).

Table 6. Observed and calculated pK_a values for the aliphatic carboxylic acids dataset (excluding amino acids) using Eq. 12 (Table 7).

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_3	ALP_3
1	000050-21-5	lactic acid	3.86	3.89	−10.27	−13.66
2	000053-86-1	indomethacin	4.50	4.70	−10.49	−13.60
3	000061-78-9	glycine, n-(4-aminobenzoyl)-	3.80	3.78	−10.29	−13.71
4	000061-33-6	benzylpenicillin	2.74	2.52	−10.07	−13.91
5	000061-32-5	methicillin	2.77	3.14	−10.21	−13.84
6	000064-69-7	iodo-acetic acid	3.15	3.17	−10.21	−13.83
7	000064-19-7	acetic acid	4.76	4.43	−10.45	−13.65
8	000068-11-1	mercaptoacetic acid	3.55	3.62	−10.23	−13.71
9	000075-99-0	2,2-dichloro-proprionic acid	1.79	1.89	−9.99	−14.03
10	000075-98-9	2,2-dimethyl propanoic acid	5.03	4.60	−10.50	−13.64
11	000076-93-7	benzilic acid	3.05	3.94	−10.30	−13.67
12	000076-05-1	trifluoroacetic acid	0.52	0.04	−9.50	−14.18
13	000076-03-9	trichloroacetic acid	0.51	1.01	−9.77	−14.12
14	000077-06-5	gibberellic acid	4.00	3.76	−10.20	−13.64
15	000079-43-6	dichloroacetic acid	1.26	1.68	−9.92	−14.04
16	000079-31-2	isobutyric acid	4.84	4.54	−10.48	−13.64
17	000079-14-1	hydroxyacetic acid	3.83	3.72	−10.22	−13.67
18	000079-11-8	chloroacetic acid	2.87	2.84	−10.15	−13.88
19	000079-09-4	propionic acid	4.88	4.46	−10.46	−13.65
20	000079-08-3	bromoacetic acid	2.89	2.81	−10.14	−13.88
21	000081-25-4	cholic acid	4.98	4.74	−10.54	−13.63
22	000085-34-7	(2,3,6-trichlorophenyl)acetic acid	3.70	3.85	−10.37	−13.76
23	000086-87-3	naphthaleneacetic acid	4.23	4.18	−10.43	−13.71
24	000087-51-4	indole-3-acetic acid	4.75	4.87	−10.48	−13.54
25	000087-08-1	phenoxymethylpenicillin	2.79	2.73	−10.11	−13.88
26	000088-09-5	2-ethylbutyric acid	4.71	4.72	−10.52	−13.62
27	000090-64-2	a-hydroxyphenylacetic acid	3.41	3.70	−10.25	−13.70
28	000093-76-5	2,4,5-trichlorophenoxyacetic acid	2.83	3.11	−10.19	−13.83
29	000093-72-1	2-(2,4,5-trichlorophenoxy)propionic acid	2.84	3.10	−10.22	−13.86
30	000094-82-6	4-(2,4-dichlorophenoxy)butyric acid	4.95	3.84	−10.31	−13.71
31	000094-81-5	4-(4-chloro-o-tolyloxy)butyric acid (MCPB)	6.20	3.90	−10.32	−13.70
32	000094-75-7	2,4-dichlorophenoxyacetic acid	2.73	3.32	−10.24	−13.81
33	000094-74-6	2-methyl-4-chlorophenoxyacetic acid	3.13	3.42	−10.24	−13.78
34	000097-61-0	2-methyl- pentanoic acid	4.79	4.57	−10.49	−13.64
35	000098-89-5	cyclohexanecarboxylic acid	4.90	4.57	−10.49	−13.64

Table 6. (cont.)

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_3	ALP_3
36	000099-66-1	valproic acid	4.60	4.69	−10.52	−13.63
37	000102-32-9	3,4-dihydroxyphenylacetic acid	4.25	4.31	−10.35	−13.60
38	000103-82-2	phenylacetic acid	4.31	4.44	−10.43	−13.63
39	000104-03-0	p-nitrophenylacetic acid	3.85	3.08	−10.13	−13.79
40	000104-01-8	p-methoxyphenylacetic acid	4.36	4.56	−10.45	−13.61
41	000107-94-8	3-chloropropionic acid	3.99	3.77	−10.25	−13.68
42	000107-92-6	butyric acid	4.82	4.48	−10.47	−13.65
43	000116-53-0	2-methyl- butanoic acid	4.81	4.57	−10.49	−13.64
44	000117-34-0	diphenylacetic acid	3.94	4.51	−10.42	−13.60
45	000120-36-5	2-(2,4-dichlorophenoxy)propanoic acid	3.10	3.27	−10.27	−13.85
46	000122-88-3	p-chlorophenoxyacetic acid	3.10*	2.83	−10.18	−13.91
47	000122-59-8	phenoxyacetic acid	3.17	3.06	−10.25	−13.90
48	000123-76-2	levulinic acid	4.64	4.33	−10.38	−13.62
49	000141-82-2	malonic acid	2.85	2.74	−10.08	−13.85
50	000141-76-4	3-iodopropionic acid	4.09	3.85	−10.28	−13.68
51	000144-49-0	fluoroacetic acid	2.59	2.78	−10.14	−13.89
52	000300-85-6	β-hydroxybutyric acid	4.41	4.33	−10.30	−13.55
53	000305-03-3	chlorambucil	5.75	4.39	−10.45	−13.66
54	000306-08-1	4-OH-3-methoxy- benzeneacetic acid	4.41	4.43	−10.38	−13.59
55	000327-97-9	chlorogenic acid	2.66	3.38	−10.17	−13.73
56	000331-25-9	m-fluorophenylacetic acid	4.13	4.09	−10.33	−13.65
57	000348-10-7	o-fluorophenoxyacetic acid	3.08	3.49	−10.30	−13.81
58	000372-09-8	cyanoacetic acid	2.45	2.59	−10.05	−13.87
59	000404-98-8	m-fluorophenoxyacetic acid	3.13	3.61	−10.24	−13.72
60	000405-79-8	p-fluorophenoxyacetic acid	3.13	3.61	−10.24	−13.72
61	000405-50-5	p-fluorophenylacetic acid	4.24	4.15	−10.34	−13.64
62	000462-60-2	n-(aminocarbonyl)glycine	3.89	2.94	−10.06	−13.77
63	000467-69-6	flurenol	1.09	4.18	−10.34	−13.63
64	000473-81-4	glyceric acid	3.55	3.35	−10.18	−13.75
65	000501-52-0	beta-phenylpropionic acid	4.66	4.40	−10.44	−13.65
66	000503-74-2	isovaleric acid	4.77	4.57	−10.49	−13.64
67	000503-66-2	hydracrylic acid	4.51	3.99	−10.33	−13.68
68	000515-30-0	α-hydroxy-α-methyl benzeneacetic acid	3.53	3.94	−10.30	−13.67
69	000516-05-2	methyl malonic acid	3.12	2.86	−10.11	−13.84
70	000539-35-5	mycobacidin	5.10	4.17	−10.37	−13.66
71	000581-96-4	2-naphthaleneacetic acid	4.25	4.38	−10.42	−13.64
72	000588-32-9	m-chlorophenoxyacetic acid	3.07	2.92	−10.19	−13.89
73	000588-22-7	3,4-dichlorophenoxyacetic acid	2.92	2.65	−10.13	−13.92
74	000594-61-6	α-hydroxy-i-butyric acid	3.61	4.07	−10.31	−13.64
75	000595-46-0	dimethylmalonic acid	3.15	3.22	−10.18	−13.79
76	000595-37-9	2,2-dimethyl butyric acid	5.03	4.66	−10.51	−13.63
77	000598-78-7	2-chloropropionic acid	2.80	3.17	−10.20	−13.82
78	000598-72-1	α-bromopropionic acid	2.97	3.17	−10.21	−13.83
79	000601-75-2	ethylmalonic acid	2.96	3.33	−10.21	−13.78
80	000614-61-9	o-chlorophenoxyacetic acid	3.05	3.55	−10.31	−13.80
81	000616-62-6	propylpropanedioic acid	2.99	3.22	−10.16	−13.77
82	000617-31-2	2-hydroxy- pentanoic acid	3.89**	3.95	−10.29	−13.66
83	000622-47-9	p-methylphenylacetic acid	4.37	4.49	−10.45	−13.63
84	000646-07-1	4-methylpentanoic acid	4.84	4.54	−10.48	−13.64
85	000689-13-4	hadacidin	3.50	3.32	−10.10	−13.69
86	000940-64-7	p-methylphenoxyacetic acid	3.21	3.86	−10.34	−13.73
87	001643-15-8	m-methylphenoxyacetic	3.20	3.08	−10.26	−13.90
88	001759-53-1	cyclopropanecarboxylic acid	4.83	4.63	−10.50	−13.63
89	001798-99-8	m-bromophenoxyacetic acid	3.09	3.57	−10.26	−13.75
90	001798-11-4	p-nitrophenoxyacetic acid	2.89	2.16	−9.99	−13.95
91	001798-06-7	p-iodophenylacetic acid	4.18	4.06	−10.32	−13.65
92	001821-12-1	4-phenylbutyric acid	4.76	4.39	−10.45	−13.66
93	001877-75-4	p-methoxyphenoxyacetic acid	3.21	3.89	−10.34	−13.72
94	001877-73-2	m-nitrophenylacetic acid	3.97	3.56	−10.14	−13.65
95	001878-94-0	p-iodophenoxyacetic acid	3.16	2.74	−10.16	−13.92
96	001878-93-9	m-iodophenoxyacetic acid	3.13	3.57	−10.26	−13.75
97	001878-92-8	o-iodophenoxyacetic acid	3.17	3.42	−10.23	−13.77

Table 6. (cont.)

N	CAS no.	Chemical name	Exp pK_a	Calc pK_a	SE_3	ALP_3
98	001878-91-7	p-bromophenoxyacetic acid	3.13	2.69	−10.12	−13.90
99	001878-88-2	m-nitrophenoxyacetic acid	2.95	3.05	−10.11	−13.78
100	001878-87-1	o-nitrophenoxyacetic acid	2.90	3.01	−10.14	−13.82
101	001878-85-9	o-methoxy phenoxyacetic acid	3.23	4.01	−10.44	−13.77
102	001878-82-6	p-cyanophenoxyacetic acid	2.93	2.54	−10.10	−13.93
103	001878-69-9	m-iodophenylacetic acid	4.16	4.12	−10.33	−13.64
104	001878-68-8	p-bromophenylacetic acid	4.19	4.06	−10.32	−13.65
105	001878-66-6	p-chlorophenylacetic acid	4.19	4.12	−10.34	−13.65
106	001878-65-5	m-chlorophenylacetic acid	4.14	4.15	−10.34	−13.64
107	001878-49-5	o-methylphenoxyacetic acid	3.23	3.67	−10.32	−13.77
108	001879-58-9	m-cyanophenoxyacetic acid	3.03	3.41	−10.19	−13.74
109	001879-56-7	o-bromophenoxyacetic acid	3.13	3.52	−10.30	−13.80
110	002088-24-6	m-methoxyphenoxyacetic acid	3.14	3.05	−10.26	−13.91
111	002270-20-4	5-phenylpentanoic acid	4.88	4.43	−10.45	−13.65
112	002976-75-2	(1-naphthalenyloxy)-acetic acid	3.20	3.00	−10.22	−13.89
113	003813-05-6	4-chloro-2-oxo-3(2H)-benzothiazoleacetic acid	3.04	3.50	−10.28	−13.79
114	005292-21-7	cyclohexylacetic acid	4.80	4.57	−10.49	−13.64
115	006324-11-4	o-hydroxyphenoxyacetic acid	3.02	3.58	−10.16	−13.66
116	010502-44-0	2-hydroxy-2-(4-methoxyphenyl)acetic acid	3.42	3.79	−10.27	−13.69
117	014387-10-1	4-ethyl- benzeneacetic acid	4.37	4.49	−10.45	−13.63
118	015307-86-5	diclofenac	4.15	4.16	−10.46	−13.74
119	015687-27-1	ibuprofen	4.45***	4.61	−10.48	−13.62
120	016484-77-8	(R)-2-(4-chloro-o-tolyloxy)propionic acid	3.68	3.53	−10.27	−13.77
121	016563-41-0	3-(1-naphthalenyloxy)-propanoic acid	4.00	3.95	−10.29	−13.66
122	018046-21-4	fentiazac	3.60	4.02	−10.34	−13.68
123	018698-96-9	2-iodo- benzeneacetic acid	4.04	4.24	−10.36	−13.63
124	020225-24-5	2-ethylpentanoic acid	4.71	4.66	−10.51	−13.63
125	022071-15-4	ketoprofen	4.45	4.38	−10.40	−13.62
126	022131-79-9	alcofenac	4.29****	4.30	−10.37	−13.62
127	022204-53-1	naprosyn	4.15	4.52	−10.46	−13.63
128	031879-05-7	fenoprofen	4.50	4.38	−10.41	−13.63
129	032857-63-9	4-(1,1-dimethylethyl)-benzeneacetic acid	4.42	4.52	−10.46	−13.63
130	036330-85-5	fenbufen	4.51	4.42	−10.40	−13.61
131	038194-50-2	sulindac	4.70	4.30	−10.38	−13.63
132	040828-46-4	suprofen	3.91	4.11	−10.35	−13.66
133	040843-25-2	2-[4-(2,4-dichlorophenoxy)phenoxy]propanoic acid	3.43	2.96	−10.17	−13.86
134	053808-88-1	lonazolac	4.30	4.27	−10.37	−13.63
135	055335-06-3	3,5,6-trichloro-2-pyridyloxyacetic acid	2.68*****	2.83	−10.17	−13.90
136	055863-26-8	tiopinac	3.71	4.20	−10.38	−13.66
137	058667-63-3	N-benzoyl-N-(3-chloro-4-fluorophenyl)-DL-alanine	3.72	3.55	−10.16	−13.67
138	069335-91-7	2-[4-[[5-(trifluoromethyl)-2-pyridinyl]oxy]phenoxy]propanoic acid	3.12	3.66	−10.28	−13.74
139	069806-34-4	2-[4-[[3-chloro-5-(trifluoromethyl)-2-pyridinyl]oxy]phenoxy]propanoic acid	2.90	2.90	−10.16	−13.87
140	074103-06-3	ketorolac	3.49	3.84	−10.32	−13.72
141	089894-13-3	phenoxyacetic acid, 4-chloro-3-nitro	2.96	2.77	−10.01	−13.78
142	104273-73-6	4-(cyclopropylhydroxymethylene)-3,5-dioxocyclohexanecarboxylic acid	5.32	3.79	−10.27	−13.69
143	–	4-methyl-umbelliferyl β-d-glucuronide	2.82	2.75	−10.07	−13.84

* Value taken from CRC Handbook of Chemistry and Physics; 81st ed., Lide, D. R., and Lide, Jr. (Eds.), Chapman and Hall, London 2001.
** Value from Brown, H. C. et al., Braude, E. A., Nachod, F. C. (Eds.), Determination of Organic Structures by Physical Methods, Academic Press, 1955, pp. 567–662.
*** Value from Avdeef A., Box, K. J., Comer, J. E., Hibbert, C., and Tam, K. Y. *Pharm. Res. 15*, 209–215 (1998); Balon, K., Riebesehl, B. U., and Muller, B. W. *Pharm. Res. 16*, 882–888 (1999).
**** Value from Hansch, C. (Ed.), *Comprehensive Medicinal Chemistry*; Vol. 6.; Pergamon Press: Oxford, 1990.
***** Value from *Martindale - The Extra Pharmacopoeia*, 31st ed., The Pharmaceutical Press, London 1996.

Table 7. Results for the phenol and carboxylic acid datasets.

Data set	N	Equations & Statistics	
Phenols			
All	175	$pK_a = -7.46*SE_1 - 61.65$ $r = 0.90$, $r^2 = 0.81$, $r_{cv}^2 = 0.81$, $F = 736.12$, $s = 1.003$	(1)
	175	$pK_a = -45.20*AQ_1 + 9.77*FE_1 + 3.39*ALP_2 - 6.34*SE_2 + 27.21$ $r = 0.97$, $r^2 = 0.93$, $r_{cv}^2 = 0.93$, $F = 594.74$, $s = 0.599$	(2)
Meta/para	58	$pK_a = -5.21*SE_1 - 39.87$ $r = 0.96$, $r^2 = 0.92$, $r_{cv}^2 = 0.91$, $F = 644.74$, $s = 0.315$	(3)
H-bonded	26	$pK_a = -11.33*SE_1 - 99.31$ $r = 0.91$, $r^2 = 0.83$, $r_{cv}^2 = 0.81$, $F = 119.92$, $s = 0.688$	(4)
Remaining ortho	91	$pK_a = -7.57*SE_1 - 62.56$ $r = 0.96$, $r^2 = 0.92$, $r_{cv}^2 = 0.91$, $F = 1003.6$, $s = 0.753$	(5)
Aromatic carboxylic acids			
All	99	$pK_a = -9.60*SE_1 - 41.74$ $r = 0.81$, $r^2 = 0.66$, $r_{cv}^2 = 0.64$, $F = 185.08$, $s = 0.567$	(6)
	99	$pK_a = -0.86*LUMO - 7.73*SE_1 + 3.04*ALP_4 - 2.33*SE_4 + 16.35$ $r = 0.93$, $r^2 = 0.87$, $r_{cv}^2 = 0.85$, $F = 153.99$, $s = 0.357$	(7)
Meta/para	46	$pK_a = -1.98*SE_3 - 16.69$ $r = 0.86$, $r^2 = 0.75$, $r_{cv}^2 = 0.72$, $F = 131.61$, $s = 0.228$	(8)
	45	$pK_a = -2.14*SE_3 - 18.27$ $r = 0.93$, $r^2 = 0.86$, $r_{cv}^2 = 0.85$, $F = 274.94$, $s = 0.167$	(9)
Ortho	53	$pK_a = -3.51*SE_3 - 33.43$ $r = 0.89$, $r^2 = 0.79$, $r_{cv}^2 = 0.77$, $F = 188.53$, $s = 0.463$	(10)
Aliphatic carboxylic acids			
All	185	$pK_a = 4.29*ALP_1 - 41.77*AQ_2 - 30.04*AQ_3 + 0.71*FE_3 + 56.06$ $r = 0.83$, $r^2 = 0.69$, $r_{cv}^2 = 0.67$, $F = 101.28$, $s = 0.564$	(11)
Excluding amino acids	143	$pK_a = 3.24*ALP_3 - 2.80*SE_3 + 19.43$ $r = 0.84$, $r^2 = 0.70$, $r_{cv}^2 = 0.69$, $F = 165.80$, $s = 0.510$	(12)
	141	$pK_a = 3.53*ALP_3 - 2.65*SE_3 + 25.01$ $r = 0.90$, $r^2 = 0.80$, $r_{cv}^2 = 0.80$, $F = 282.85$, $s = 0.394$	(13)

N indicates the number of molecules in each dataset.

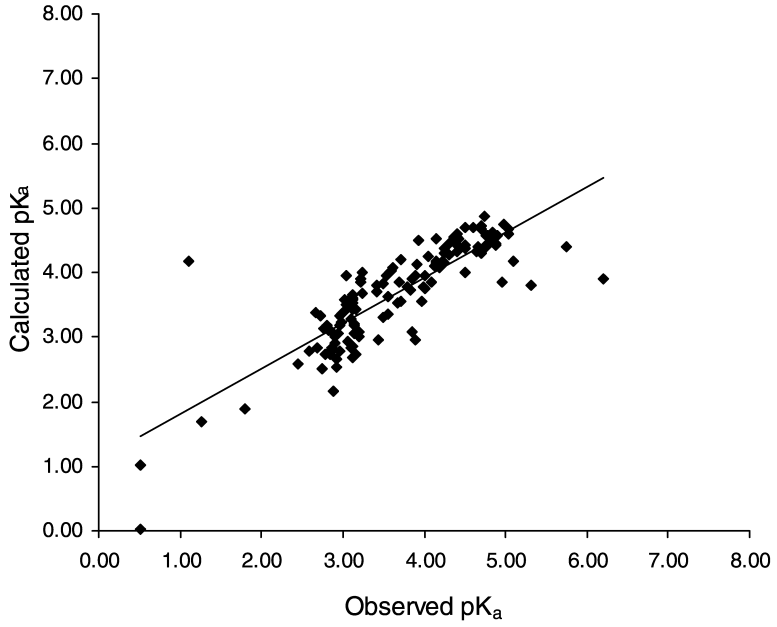

Figure 4. Plot of calculated vs. observed pK_a for the aliphatic acid dataset, using Eq. 12 from Table 7.

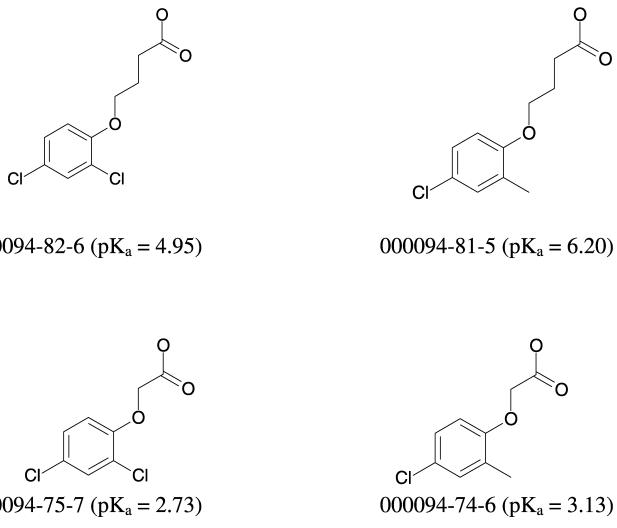

Figure 5. Structures and observed pK_a values for compounds 000094-82-6 (no. **30**, Table 6), 000094-81-5 (no. **31**, Table 6), 000094-75-7 (no. **32**, Table 6) and 000094-74-6 (no. **33**, Table 6).

We found that two parameters (SE_3 and ALP_3) (Eq. 12, Table 7) could model the 143 acids, with an r^2 of 0.70. It was encouraging to note that SE_3 was used previously to model the benzoic acids (Eqs. 8 – 10, Table 7). In our opinion, this observation together with the use of only two terms, makes Eq. 12 more reliable than Eq. 11. The calculated pK_a for the 143 aliphatic acids are reported in Table 6 and a plot of calculated versus observed pK_a is shown in Figure 4. Two compounds appeared to be clear outliers: flurenol (000467-69-6, no. **63**, Table 6) and 4-(4-chloro-o-tolyloxy)butyric acid (MCPB) (000094-81-5, no. **31**, Table 6). While it was not clear why flurenol was an outlier, its pK_a may be difficult to predict given the proximity of the carboxylate group to the two aromatic rings and that it contains an α-hydroxy group. However, when the pK_a of flurenol is compared to 000076-93-7 (**11**, Table 6) then the value quoted is clearly anomalous. With respect to the second compound, we observed that some analogues of MCPB are present in the dataset. In particular the pair 000094-82-6/000094-81-5 (**30**, **31** Table 6) are very closely related to the pair 000094-75-7/ 000094-74-6 (**32**, **33** Table 6. See Figure 5). Both sets differ only in the replacement of a methyl group with a chlorine atom. In the first case this substitution results in an increase in the pK_a of 1.25 units, while in the second pair the same substitution results in an increase of only 0.4 units, thus making this experimental data difficult to explain. Moreover, when the pK_a of 001821-12-1 (**92**, Table 6) is considered then it further brings into question the values listed for both 000094-82-6/000094-81-5. Following the removal of MCPB and flurenol from the training set, we obtained a very similar model in terms of coefficients, but with an improved r^2 (0.80, Eq. 13, Table 7). There were no other compounds with an unsigned error greater than 2 log units in Table 6. The small pK_a range covered by the amino acids (about 2 log units) prevented us from being able to model them separately with any confidence.

4 Discussion

We have described the application of QM parameters, derived from frontier electron theory, to the prediction of the acidity of phenols and carboxylic acids. In particular, the property describing electrophilic superdelocalisability (SE) was consistently found to be highly correlated with pK_a. This finding is in agreement with the results from a recent paper, where SE was found superior to other QM calculated quantities in modelling the hydrogen bonding ability of common chemical groups [21].

The results we present in this study on phenols compare very well with those reported in the literature. Grüber and Buss [12] obtained good results for predicting the pK_a of phenols using a six-term equation based on partial atomic charges and HOMO energy. Their equation ($r^2 = 0.94$, $n = 99$) is comparable to our Eq. 2 ($r^2 = 0.93$, $n = 175$). The results of the phenol subsets (Eqs. 3 – 5) are also comparable to those of Grüber and Buss in term of statistics, with the exception of the H-bonded phenols, which gave an r^2 of 0.83. It should be kept in mind, however, that the pK_a range covered by the H-bonded set is lower than the range covered by our other two subsets. Our equations can be considered to be easier to interpret as only one descriptor is required (SE). In a recent study, Citra [13] has used bond orders and Coulson partial atomic charges to predict the pK_a of phenols. It should be mentioned that Citra performed multiple geometry optimisations, so that the values of the descriptors used in the regression analyses are the weighted average of the QM properties computed for all optimised conformations. For their set of 101 phenols extracted from the Syracuse database a three term equation with an r^2 of 0.96 was derived. Unfortunately we were not able to reproduce these results for our set of 175 phenols, using a single conformation (data not shown). We have calculated the Coulson charges and found that the correlation between Coulson charges and pK_a ($r = -0.81$) was lower than the correlation between SE_1 and pK_a ($r = -0.91$) for the 175 phenols. This may be due to either the use of a single conformation or the larger number of compounds in our study. To further investigate this we have considered only those phenols common to the study by Citra and this present work. Once again using a single conformation on this set we were unable to achieve the same level of statistical parameters found by Citra. This work suggests that there is a conformational dependency on the relationship between Coulson charges and pK_a.

Our set of benzoic acids contained 99 compounds. For this class, Grüber and Buss [12] reported a two-term equation ($r^2 = 0.67$, $n = 52$). Citra [13] reported a three-term equation ($r^2 = 0.89$, $n = 31$). We obtained an r^2 of 0.87 ($n = 99$, Eq. 7, Table 7), which is comparable to the Citra model. By splitting the set into ortho and meta/para subset, one parameter only (SE_3) was sufficient to obtain similarly significant models (Eqs. 8 – 10, Table 7). The slightly lower squared coefficients compared to the phenols are due to the smaller range of pK_a covered by benzoic acids (5 versus 12

log units). These predictions are reasonable given that the majority of compounds had an unsigned error less than 1 unit (four compounds had an error between 1 and 2 units).

The set of aliphatic carboxylic acids were more difficult to model, partly because it was not obvious how to split them into separate classes and partly due to the their higher flexibility. As the frontier electron theory properties are sensitive to the molecular geometry, particularly in the case of flexible structures, it may be advisable to perform a conformational analysis and use descriptors that are averaged on all the low energy conformations (c.f. Citra [13]). Much smaller datasets of aliphatic carboxylic acids have been studied in the past, making the comparison with our results quite difficult. Grüber and Buss developed a three-term equation, with an r^2 of 0.80 using a training set of 32 compounds. Citra reported a three-term equation as well with an r^2 of 0.84 for 56 acids. We obtained a four-parameter equation with an r^2 of 0.69 for 185 acids. The r^2 could be improved to 0.80 when amino acids were removed as well as two outliers, resulting in a much simpler two-term equation.

5 Conclusions

This study has shown that parameters derived from semiempirical QM calculations can be applied to estimate the pK_a of a variety of acidic compounds (phenols, benzoic acids and aliphatic carboxylic acids), as an alternative to LFER methods or more computationally demanding *ab initio* calculations. Semiempirical QM calculations are relatively fast and can be easily automated to handle hundreds or even thousands of structures. Current trends are such that it is becoming more important to determine the pK_a and other physico-chemical properties associated with a drug, before synthetic work is undertaken, with the aim of avoiding the synthesis of compounds that are predicted to have poor biopharmaceutical characteristics. In this context, our methodology offers a step forward towards the incorporation of pK_a predictions into the computer-based filtering tools which are now routinely being used in the pharmaceutical industry. Future work in this area may need to explore how transferable this methodology is to structural classes beyond those studied here. In addition, improvements in the accuracy of the method need to be investigated if reasonable estimates are needed for properties such as the fraction of ionisation at specified pH values.

To make our models more generally applicable, we have extended our approach to other nitrogen containing chemical classes (amines, anilines and various heterocylic compounds). These results are reported in the following paper.

Acknowledgements

The authors would like to thank Dr. Peter Ertl (Novartis, Switzerland) for his stimulating discussions over the application of the frontier electron theory descriptors. We also extend our thanks to the referees of this paper whose comments lead to significant improvements to this study.

References

[1] Bowman, W. C., and Rand, M. J., *Textbook of Pharmacology*, 2nd ed., Blackwell Scientific Publications, Oxford 1980.

[2] Jayasekhar, P., and Kasture, A. V., Physicochemical Activity Relationship of Some Substituted Benzimidazoles, *Bull. Chim. Farm. 138*, 489 – 492 (1999).

[3] Jones, T., and Taylor, G., Quantitative Structure – Pharmacokinetic Relationships Amongst Phenothiazine Drugs, *Proc. – Eur. Congr. Biopharm. Pharmacokinet. 3rd 2,* 181 – 90 (1987).

[4] Sosnovsky, G., and Bell, P., In the Search for New Anticancer Drugs. 29. A Study on the Correlation of Lipophilicities, Ionization Constants and Anticancer Activities of Aminoxyl Labeled TEPA Congeners, *Life Sci. 62*, 639 – 648 (1998).

[5] Mahmud, T., Rafi, S. S., Scott, D. L., Wrigglesworth, J. M., and Bjarnason, I., Nonsteroidal Antiinflammatory Drugs and Uncoupling of Mitochondrial Oxidative Phosphorylation, *Arthritis Rheum. 39*, 1998 – 2003 (1996).

[6] Ahmed, S., Owen, C. P., James, K., Patel, C. K., and Patel, M., Acid Dissociation Constant, a Potential Physicochemical Factor in the Inhibition of the Enzyme Estrone Sulfatase (ES), *Bioorg. Med. Chem. Lett. 11*, 899 – 902 (2001).

[7] Albert, A., and Serjeant, E. P., *The Determination of Ionization Constants*, 2nd ed., Chapman and Hall, London 1971.

[8] ACD/Labs *ACD Labs*, version 3. 133 Richmond St. West Suite 605, Toronto, ON, M5H 2L3, Canada.

[9] Metaphorics LLC, Sante Fe, New Mexico, USA. http://www.daylight.com/meetings/emug00/Sayle/pkapredict.html.

[10] pKalc, Compudrug International 705 Grandview Drive, South San Francisco, CA 94080, USA.

[11] Perrin, D. D., Dempsey, B., and Serjeant, E. P., pKa Prediction for Organic Acids and Bases, Chapman and Hall, New York 1981.

[12] Grüber, C., and Buss, V., Quantum-Mechanically Calculated Properties for the Development of Quantitative Structure Activity Relationships (QSARs). pKa Values of Phenols and Aromatic and Aliphatic Carboxylic Acids, *Chemosphere 19*, 1595 – 1609 (1989).

[13] Citra, M. J., Estimating the pKa of Phenols, Carboxylic Acids and Alcohols from Semi-Empirical Quantum Chemical Methods, *Chemosphere 38*, 191 – 206 (1999).

[14] Schrödinger *Jaguar*, http://www.schrodinger.com/Products/jaguar.html 1500 SW First Ave, Suite 1180, Portland, Oregon, 97201, USA.

[15] Gross, K. C., and Seybold, P. G., Comparison of Quantum Chemical Parameters and Hamett Constants in Correlating pKa Values of Substituted Anilines, *J. Org. Chem. 66*, 6919 – 6925 (2001).

[16] Gross, K. C., and Seybold, P. G., Substituent Effects on the Physical Properties and pK_a of Phenols, *Int. J. Quant. Chem. 85,* 569 – 579 (2001).

[17] Richardson, W. H., Peng, C., Bashford, D., Noodleman, L., and Case D. A., Incorporating Solvation Effects Into Density Functional Theory: Calculation of Absolute Acidities, *Int. J. Quant. Chem. 61*, 207 – 217 (1997).

[18] Kallies, B., and Mitzner, R., pKa Values of Amines in Water from Quantum Mechanical Calculations Using a Polarized Dielectric Continuum Representation of the Solvent, *J. Phys. Chem. B 101*, 2959 – 2967 (1997).

[19] Liptak, M. D., and Shields, G. C., Accurate pK(a) Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods, *J. Am. Chem. Soc. 123*, 7314 – 7319 (2001).
[20] Schüürmann, G., Modelling pKa of Carboxylic Acids and Chlorinated Phenols, *Quant. Struct.-Act. Relat. 15*, 121 – 132 (1996).
[21] Gancia, E., Montana, J. G., and Manallack, D. T., Theoretical Hydrogen Bonding Parameters for Drug Design, *J. Mol. Graph. Mod. 19*, 349 – 362 (2001).
[22] Fukui, K., Yonezawa, T., and Nagata, C., Theory of Substitution in Conjugated Molecules, *Bull. Chem. Soc. Japan 27*, 423 – 427 (1954).
[23] Howard, P., and Meylan, W., Physical/Chemical Property Database (PHYSPROP), 1999 ed., Syracuse Research Corporation, Environmental Science Center, North Syracuse NY, 1999.
[24] Sadowski, J., and Gasteiger, J., From Atoms and Bonds to three-Dimensional Atomic Coordinates: Automatic Model Builders, *Chem. Rev. 93*, 2567 – 2581 (1993).
[25] Stewart, J. J. P., *Mopac program package, quantum chemistry program exchange no. 455*.
[26] Tripos Inc., 1699 South Hanley Road, St. Louis, MO 63144, USA.
[27] Accelerys *Tsar*, version 3.21, The Medawar Centre, Oxford Science Park, Oxford, OX4 4GA, UK.
[28] Dewar, M. J. S., Zoebish, E. G., Healy, E. F., and Stewart, J. J. P., Development and Use of Quantum Mechanical Molecular Models. AM1: a New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc. 107*, 3902 – 3909 (1985).
[29] Abraham, M. H., Duce, P. P., Prior, D. V., Barratt, D. G., Morris, J. J. et al., Hydrogen bonding. Part 9. Solute Proton Donor and Proton Acceptor Scales for Use in Drug Design, *J. Chem. Soc. Perkin Trans. II* 1355 – 1375 (1989).

Received on April 8, 2002; accepted on July 8, 2002