Introduction to Assignments #7 & #8: Writing a Scientific Paper

Science at its very core is "data-based, rational analysis" and the overwhelming majority of scientific papers contain original data. Aspects related to the acquisition of the original data are described in the "Materials and Methods" section and the original data are documented either in the paper or in an appendix. An authentic exercise in scientific writing must be concerned with the rational analysis of original data. Yet, there are obvious limits to original data generation in a writing class and the question is "How to write a scientific paper without original data?" To resolve this paradox, we will create a near-authentic experience by pretending that we have synthesized a number of new molecules and/or that we have measured one important property for one or more molecules and we will analyze the structure-activity relationships. Specifically, we will pretend that we have synthesized two new disubstituted benzoic acids, that we have measured the pK_a values of these new molecules in solution, and that we have measured the pK_a values of one other disubstituted benzoic acid. We further agree on the following points:

[1] A "disubstituted benzoic acid" contains 2 substituents in addition to the CO_2H.

[2] Synthetic details and data for the spectroscopic characterization of the 2 new molecules will be relegated to "Supporting Information."

[3] Experimental details of the pK_a measurements will be given in the M&M section.

[4] Your analysis should consider (i) benzoic acid itself, (ii) 3 disubstituted benzoic acids, (iii) all monosubstituted benzoic acids with the substituents occurring in any of the disubstituted benzoic acids, and (iv) no more that 10 benzoic acids overall.

[5] Assume that the pK_a values of all relevant monosubstituted benzoic acids are known.

Everybody will write a paper on ***Substituent Effects on Benzoic Acid Acidity***, either alone or together with another student, and this commonality makes the assignment manageable and guarantees well-educated peer review. And yet, the resulting papers will vary greatly because of the students' selections of "their casts". This selection needs to be made before starting to work on A07 and A08. Some changes can be made later, of course, but it will cost time and effort—better to plan ahead as much as possible.

Selection of "the Cast" for Assignments #7 & #8: Some Guidelines

The paper entitled "QSPR Prediction of pK_a for Benzoic Acids in Different Solvents" (Jover, J.; Bosque, R.; Sales, J. *QSAR Comb. Sci.* **2008**, *27*, 563-581) is provided and its Table 1 lists 519 experimental pK_a values of benzoic acids in a number of solvents (including aspirin).

[1] Select three disubstituted benzoic acids. The substituents in compound *n* are X_n and Y_n (*n* = 1 - 3) and there are many options for the identities of X and Y and also for their positions. You might consider X_n = Y_n = Cl (i.e., 2,3-, 2,4-, and 3,4-isomers). You might consider X = 4-Cl and vary 3-Y (i.e., F, Cl, Br). You might consider X_n = NO_2 and Y_n = Cl and look at three isomers….

[2] Include the "parent system" in your data set, i.e., benzoic acid itself.

[3] Select monosubstituted benzoic acids. For an analysis of the effects of double-substitution by X_n and Y_n (*n* = 1 - 3), you want to refer to the pK_a values of the monosubstituted benzoic acids with substituent X_n or Y_n in the same position. Doing so will add at most 6 molecules to your cast for a maximum of 10 molecules in your data set.

[4] Consider the solvent. You can select "to work in one solvent" and compare acidities in DMSO for example. The analysis might become more complicated (and more interesting) if you used data measured in different solvents.

[5] Consider the pK_a ranges. The K_a values vary by about 18 magnitudes! Do you want to argue about tiny changes or do you want to explain larger differences?

[6] Select the two "new" disubstituted benzoic acids. You will need to research and describe their syntheses and spectroscopic properties.

Outline the introduction to your paper while you make these choices. Why do acidities matter? Why would we want to know about acidities of substituted benzoic acids? Why would we want to know about acidities of disubstituted benzoic acids? Why did you want to know about the acidities of these disubstituted benzoic acids?

Assignments #7: Material, Methods and Appendix

Assignments #7 & #8 concern the writing of an original scientific paper on ***Substituent Effects on Benzoic Acid Acidity***. This paper will have two sections: The paper itself and its appendix. The appendix often is provided as a separate file. For our purposes, we will think of paper and appendix as two separate entities, but we will keep them together in one file for convenience. In Assignment #7 you will organize the outline of the paper, write the "Materials and Methods" section of the paper and tabulate your data, assemble the appendix, and provide all necessary bridges between paper and appendix. Your work on Assignment #7 will be evaluated by rubric-based peer-review. Your work in Assignments #8 - #10 will build on A#7 but the MMA section will not be evaluated again.

(a) Organize the Outline of Paper and Appendix. Open a Word file, set all margins to 1.25 inches, type the entire document in Times New Roman, 12 pt, with a line spacing of "at least 24 pt". On page 1 type the title in bold, provide the author line, and provide affiliation information. Starting on top of page 2, type the first-level headlines of your paper in bold (Introduction, Materials and Methods, Results and Discussion, Conclusion, References). References should be cited at the end of this first section. Create a section-break, start the second section on a new page, and restart page numbering in this section: the appendix goes in this section.

(b) Materials and Methods. Provide an overview of the syntheses of the two new compounds, provide a detailed description of the method(s) you used to measure the pK_a values of the three disubstituted benzoic acids, and create your data table.

Search the literature for syntheses of the two "new" compounds, pick the syntheses of your choice (if you find many), and "adopt" those synthetic methods as your own. Search the literature for the publications that originally reported the measurements of the pK_a values of the three disubstituted benzoic acids you selected and "adopt" those methods as your own. Search the literature for the publications that originally reported the measurements of the pK_a values of benzoic acid itself and of the monosubstituted benzoic acids you included in your cast. Create

your data table (i.e., Table 1) with one row for every member of your cast. Columns should list compound names, compound numbers, solvent(s), pK_a value, source of data, etc.

(c) Appendix. Provide a detailed description of the syntheses and provide the results of the comprehensive spectroscopic characterization of the two new compounds. Show spectra as images as much as possible (i.e., as with Assignment #5). On page 1 of the appendix, write "Supporting Information", and then provide the title of the paper, its author line, and the author affiliation information. On page 2 of the appendix, provide a "Table of Contents" for the appendix.

(d) Bridges. Page 1 of the Appendix provides the bridge from appendix to paper. The bridge from paper to appendix comprises several items: After "Conclusions" and before "References", insert a one-paragraph statement that begins with "**Supplemental Material Available**: The appendix contains…" to inform the reader that the paper comes with an appendix, to describe very briefly what type of information are contained in the appendix, and to guide the reader to the source of the appendix. In the "Materials and Methods" section, at the most logical place(s), add a statement that guides the detail-seeking reader to the appendix.

Submission & Target Dates: The assignment must be completed using MS Word 2007 with *JOC* formatting. Submit one Word file "A7_'your name'.docx" by Tuesday, 04/06/10, midnight. Dr. Glaser will review your submissions and provide feedback in class on Wednesday, 04/07/10, so that you can update your assignment before it is peer-reviewed. Bring one hardcopy to class on Friday, 04/09/10, for peer-review.

Assignments #8: Manuscript Preparation and Submission

Assignments #7 & #8 concern the writing of an original scientific paper on ***Substituent Effects on Benzoic Acid Acidity***. In Assignment #7 you organized the outline of the paper, you wrote the "Materials and Methods" section and created your main data table, you created the appendix, and you provided all necessary bridges between paper and appendix. Feel free to make minor adjustments to the "Materials and Methods" section as you write the paper. Now, in Assignment #8 it is your task to write/complete all other parts of the paper and these are: Final Title & Abstract, Introduction, Results and Discussion, and Conclusion.

(a) Introduction. Provide a three-paragraph introduction to explain acidity as an important concept and discuss its significance, to explain why the acidities of benzoic acids matter, and to explain your specific aims. The third paragraph must start with "Here we report the results of …." and it is in this paragraph where you state your hypothesis (i.e., expectation) and where you justify your effort (i.e., why the outcome of the experiments matters).

Much of the thinking that goes into "Introduction" occurred or should have occurred at the time you selected your cast. You might add one or two schemes to "Introduction". Refine and revise the introduction as the analysis of your data progresses.

(b) Results and Discussion. Use as many paragraphs as you see fit, but there should be three recognizable parts to "R&D" and they should appear in the standard sequence "RRD" (New Results, Reference Data, Discussion).

Provide an overview of what you actually did to obtain the new data listed in your main data table. Describe what existing data you will use in your analysis, provide/discuss the sources of these reference data, and add the most pertinent reference data to your main data table. The discussion consists in the analysis of the new data in the context of the reference data. Use as many Schemes and Figures as you see fit. Your analysis might include attempts to correlate the acidity data with other properties of the molecules / substituents (electronegativity,…). In that

case, you might add the additional parameters to the main data table or you may chose to present the parameters in an additional table or just in the text.

(c) Conclusion. Write two paragraphs. Summarize your accomplishments in the first paragraph, i.e., the work you did and that facts that you claim to have established beyond any reasonable doubt. In the second paragraph, explain the meaning and the significance of the new results and describe consequences that flow from your insights.

(d) Final Title & Abstract. "*Substituent Effects on Benzoic Acid Acidity*" is the working title of your paper and you should refine the title after you complete your analysis. The final title should reflect as much as possible any general conclusions you have reached. When all parts of the paper are complete and in final form, then write the abstract.

Submission & Target Dates: The assignment must be completed using MS Word 2007 with *JOC* formatting. Submit one Word file "A8_'your name'.docx" by Tuesday, 04/13/10, midnight. Dr. Glaser will review your submissions and provide feedback in class on Wednesday, 04/14/10, so that you can update your assignment before it enters peer review. Bring three hardcopies to class on Friday, 04/16/10, for peer-review.

QSAR & Combinatorial Science QCS

QSPR Prediction of pK_a for Benzoic Acids in Different Solvents

Jesús Jover, Ramón Bosque and Joaquim Sales*

Departament de Química Inorgànica, Universitat de Barcelona, Martí i Franquès, 1, 08028-Barcelona, Spain,
E-mail: joaquim.sales@ub.edu

Keywords: Benzoic acids, Different solvents, Multicomponent systems, Neural networks, pK_a, QSPR

Received: July 11, 2007; Accepted: September 23, 2007

DOI: 10.1002/qsar.200710095

Abstract

A Computational Neural Network (CNN) derived model is proposed for the pK_a prediction of benzoic acids in different solvents. The system studied contains 519 pK_a values corresponding to 136 benzoic acids determined in water and in 8 organic solvents. The benzoic acids were described by the usual molecular descriptors and the solvents by a number of physical properties and by several parameters of the most widely used polarity solvent scales. The model is composed of seven descriptors – five of them corresponding to the solute and the other two to the solvents – and was validated by an external prediction set. The three sets of values needed in the analysis, training, prediction, and cross-validation, have the same squared correlation coefficient (0.998) and Root-Mean Square Error (RMSE) (0.21) values. The robustness of the model is also given for the statistical results for small subsets, such as *ortho*- and non-*ortho*-substituted acids, those obtained in protic or aprotic solvents, those obtained in each solvent, and even those for the set of pK_a values of one specific acid in several solvents. The descriptors encode information about the chemical nature of the solutes and the solvents that is clearly related to the interactions present in the dissociation process in solution. The derived model also has the ability to predict pK_a values of a larger set of aromatic neutral acids, containing both phenols and benzoic acids.

1 Introduction

The acid dissociation constant, K_a, which describes the extent to which a compound dissociates in the gas phase or in solution, is a fundamental property of any chemical compound. It is a key feature, which governs the chemical reactivity of the substances with other compounds in any solvent, and also the interaction with the solvent itself. In aqueous solution, the pK_a is responsible for several pharmacokinetic properties. Jointly with integrity, lipophilicity, solubility, and permeability, pK_a has been considered one of the five key physiochemical profiling screens to provide an early understanding of the key properties that affect ADME characteristics [1]. Although most of the experimental pK_a values have been determined in water, nowadays pK_a values of different types of organic and inorganic compounds in several solvents are known, as a consequence of the growing use of non-aqueous solvents in organic synthesis as well as in inorganic and organometallic chemistry. A precise knowledge about the acidity and basicity of molecular compounds in organic solvents is of crucial significance for the understanding of mechanisms in a wide variety of chemical reactions and synthetic processes; moreover, these properties also play a fundamental role in many analytical procedures such as acid–base titrations, solvent extraction, complex formation, and ion transport. Thanks to the development of computational chemistry, the acidity and basicity of chemical compounds in the gas phase can now be calculated with equivalent or greater accuracy than that obtained experimentally. However, the situation is less satisfactory in solution, since it is not easy to describe the solvation effects that exert profound influence on proton transfer processes. Using dielectric continuum methods, however, it is possible to calculate the acidity of various organic molecules in water, with a precision of about 0.5–2.2 pK_a units [2]. Recently, these methods have been applied with good results to several organic acids in Acetonitrile (AN) [3] and in dimethylsulfoxide (DMSO) [2–4], demonstrating the increasing interest in the chemistry of non-aqueous solvents. The Quantitative Structure–Property Relationship (QSPR) approach is a widely used method to predict different physical and chemical properties of chemical compounds, from numerical descriptors which are derived from the molecular structure. Several studies are known on the estimation of pK_a of different types of organic compounds in aqueous solution; thus, recently pK_a values for carboxylic acids and alcohols have been predicted with empirical atomic charge

descriptors [5]. Regarding benzoic acids, a three-term equation obtained by multiple linear regression, has been derived from a set of 31 benzoic acids and a correlation coefficient (R^2) of 0.85 was obtained [6]. Another QSPR study on a set of 99 benzoic acids has been published; the model contains four quantum mechanical descriptors and gives an R^2 of 0.66 and a standard deviation of 0.57 [7]. In non-aqueous solutions, other approaches have been applied; thus, a comparative study of Hammett – Taft and Drago models for the prediction of pK_a of benzoic acids, carboxylic acids, phenols, and protonated amines, in methanol has been published [8]. On the other hand, Pytela and coworkers in a series of papers, have developed the AISE theory (Alternative Interpretation of Substituent Effects) and applied it to the analysis of pK_a values of benzoic acids in different solvents [9], through a detailed study on the effect of the nature and position of the substituents in the aromatic ring on the pK_a values. In nearly all the cases reported, the QSPR approach has been applied to the prediction of physical and chemical properties that depend only on the chemical nature of the set of compounds analyzed and not on the characteristics of other components, such as the solvent. Thus, the references given above on the pK_a prediction of acids only consider water as solvent. In contrast, studies reported on more complex systems in which the property is determined by varying the solvents or the physical conditions are extremely scarce. In these cases, descriptors of the different elements of the system must be used simultaneously and, due to the fact that the feature usually depends nonlinearly on all of these conditions, it is convenient to use Computational Neural Networks (CNNs) in the derivation of the QSPR model. Temperature is, in general, the additional descriptor used in the prediction of several temperature dependent physical properties, such as vapor pressure, viscosity, and density of organic compounds [10 – 12]. On the other hand, another study on the kinetics of the acid hydrolysis of carboxylic acid esters at various temperatures and in different solvents has also been reported [13].

In a previous paper, we applied the QSPR approach to a complex multicomponent system, in order to predict the pK_a values of a set of phenols in ten solvents. Using descriptors of the solutes and the solvents simultaneously, we proposed a seven-descriptor model, CNN derived, which gave good correlation coefficients ($R^2 \approx 0.98$), and Root-Mean Square Errors (RMSEs) lower than one unit of pK_a (≈ 0.8) [14]. Now, we report here the results obtained in the prediction of pK_a values of a wide set of benzoic acids determined in water and in eight organic solvents (set A), using descriptors of the two components. The derived model is also able to estimate the pK_a values of a larger set containing both types of neutral aromatic acids: benzoic acids and phenols (set B).

2 Data and Computational Methods

2.1 Data Set

The data set A was comprised of 519 pK_a values, corresponding to 136 benzoic acids in nine solvents, four protic: isopropanol, ethanol, methanol, and water, and five aprotic: *N*,*N*-dimethylacetamide (DMA), *N*,*N*-dimethylformamide (DMF), acetone, AN, and DMSO. The experimental pK_a values have been taken mainly from the series of papers by Pytela *et al.* [15 – 20], Bosch and coworkers [21, 22], the Izutsku compilation [23], Kolthoff and coworkers [24, 25], and others [7, 26]. The 136 benzoic acids contain different substituents in the aromatic ring, and besides the benzoic acid itself there are 49 mono-, 68 di-, 14 tri-, 3 tetra-, and 1 penta-substituted compounds. Twenty-four types of substituents with very different electronic and steric effects, such as halogens, alkyl, alkoxy, hydroxy, nitro, amino, cyano, acetamido, *etc.* are present in *ortho*-, *meta*-, and *para*-positions of the aromatic ring. Two hundred eighty-one values belong to acids with only substituents in *para*- and/or *meta*-positions, and the remaining 238 values correspond to *ortho*-substituted benzoic acids. The number of solutes in each solvent is: methanol (107), water (105), DMSO (84), AN (63), DMF (58), acetone (55), ethanol (23), isopropanol (18), and DMA (6). Thus, to sum up, there are 253 values in protic solvents and 266 in aprotic ones. The pK_a values ranged from 0.65 for the 2,4,6-trinitrobenzoic acid in water to 21.29 for 3-methyl-4-methoxybenzoic acid in AN, with a mean value of 10.41. The full data set of 519 values was divided randomly in three subsets: the training set consists of 379 values (72%), the prediction set is composed of other 98 values (19%), and the cross-validation set contains the remaining 42 pK_a values (9%). The random selection was done to ensure that the prediction and the cross-validation sets contain values in all the solvents studied and with all the types of the substituents. The training set is used exclusively to derive the model. The prediction set, formed by pK_a values that were not included in the model development, is used to probe the predictive ability of the model. The third set, the cross-validation set, is used to determine when to stop training the neural network, in order to prevent its overtraining and to ensure that the network would have good and general predictive capacity. Table 1 contains all the experimental and calculated pK_a values.

2.2 Solute Descriptors

The generation of the structural descriptors of the benzoic acids was performed with the CODESSA program [27]. The structures of the compounds were drawn with HyperChem Lite and the geometries were fully optimized, without symmetry restrictions, using the semi-empirical method AM1 [28] implemented in the MOPAC 6.0 program [29]. In all cases, frequency calculations were performed in

Table 1. Experimental and calculated pK_a for the training, prediction, and cross-validation sets.

No.	Set[a]	Compound	Solvent	Experimental	Calculated
1	c	2,3,4,5,6-Pentamethylbenzoic acid	DMSO	11.31	11.08
2	t	2,3,4,5,6-Pentamethylbenzoic acid	MeOH	8.66	8.25
3	p	2,3,4,5-Tetramethylbenzoic acid	DMSO	11.58	11.71
4	t	2,3,4,5-Tetramethylbenzoic acid	MeOH	9.24	9.38
5	t	2,3,4,6-Tetramethylbenzoic acid	DMSO	11.17	11.12
6	t	2,3,4,6-Tetramethylbenzoic acid	MeOH	8.61	8.71
7	t	2,3,4-Trimethylbenzoic acid	DMSO	11.26	11.15
8	c	2,3,4-Trimethylbenzoic acid	MeOH	9.19	8.97
9	t	2,3,5,6-Tetramethylbenzoic acid	DMSO	10.89	11.22
10	t	2,3,5,6-Tetramethylbenzoic acid	MeOH	8.51	8.57
11	t	2,3,5-Trimethylbenzoic acid	DMSO	11.24	11.69
12	p	2,3,5-Trimethylbenzoic acid	MeOH	9.05	9.50
13	t	2,3,6-Trichlorobenzoic acid	Water	1.50	1.44
14	t	2,3,6-Trimethylbenzoic acid	DMSO	11.07	11.03
15	t	2,3,6-Trimethylbenzoic acid	MeOH	8.42	8.55
16	c	2,3-Dihydroxybenzoic acid	Water	2.91	2.81
17	t	2,3-Dimethoxybenzoic acid	Water	3.98	4.00
18	t	2,3-Dimethylbenzoic acid	DMSO	11.09	11.51
19	t	2,3-Dimethylbenzoic acid	Water	3.74	4.33
20	t	2,3-Dimethylbenzoic acid	MeOH	8.98	9.48
21	t	2,4,5-Trimethylbenzoic acid	DMSO	11.58	11.89
22	t	2,4,5-Trimethylbenzoic acid	MeOH	9.60	9.75
23	t	2,4,6-Trimethylbenzoic acid	AN	20.55	20.33
24	t	2,4,6-Trimethylbenzoic acid	DMSO	10.88	10.77
25	t	2,4,6-Trimethylbenzoic acid	Water	3.45	3.52
26	t	2,4,6-Trimethylbenzoic acid	MeOH	8.57	8.57
27	t	2,4,6-Trinitrobenzoic acid	Water	0.65	0.88
28	p	2,4-Dichlorobenzoic acid	AN	18.40	18.82
29	t	2,4-Dichlorobenzoic acid	DMSO	9.12	9.06
30	p	2,4-Dichlorobenzoic acid	Water	2.76	2.63
31	t	2,4-Dichlorobenzoic acid	MeOH	7.80	7.81
32	t	2,4-Dimethoxybenzoic acid	Water	4.36	4.22
33	t	2,4-Dimethylbenzoic acid	DMSO	11.42	11.49
34	c	2,4-Dimethylbenzoic acid	Water	4.22	4.35
35	p	2,4-Dimethylbenzoic acid	MeOH	9.50	9.53
36	t	2,4-Dinitrobenzoic acid	Water	1.43	1.81
37	t	2,4-Dinitrobenzoic acid	MeOH	6.45	6.81
38	p	2,5-Dichlorobenzoic acid	Water	2.47	2.42
39	t	2,5-Dihydroxybenzoic acid	Water	2.95	2.77
40	t	2,5-Dihydroxybenzoic acid	MeOH	8.04	7.95
41	p	2,5-Dimethylbenzoic acid	DMSO	11.36	11.43
42	t	2,5-Dimethylbenzoic acid	Water	3.98	4.29
43	p	2,5-Dimethylbenzoic acid	MeOH	9.29	9.45
44	t	2,6-Dibromobenzoic acid	Isopropanol	8.61	8.68
45	t	2,6-Dibromobenzoic acid	Acetone	15.89	15.83
46	c	2,6-Dibromobenzoic acid	AN	18.04	17.90
47	p	2,6-Dibromobenzoic acid	DMF	9.72	9.75
48	t	2,6-Dibromobenzoic acid	DMSO	8.61	8.37
49	c	2,6-Dibromobenzoic acid	EtOH	7.93	8.00
50	t	2,6-Dibromobenzoic acid	MeOH	7.14	7.27
51	t	2,6-Dibutoxybenzoic acid	Isopropanol	11.48	11.28
52	t	2,6-Dibutoxybenzoic acid	Acetone	19.00	18.89
53	t	2,6-Dibutoxybenzoic acid	AN	21.13	21.10
54	t	2,6-Dibutoxybenzoic acid	DMF	12.90	12.90
55	t	2,6-Dibutoxybenzoic acid	DMSO	11.45	11.38
56	t	2,6-Dibutoxybenzoic acid	EtOH	10.19	10.28
57	t	2,6-Dibutoxybenzoic acid	Water	3.60	3.69
58	t	2,6-Dibutoxybenzoic acid	MeOH	9.13	9.12
59	p	2,6-Dichlorobenzoic acid	Isopropanol	8.71	8.60
60	c	2,6-Dichlorobenzoic acid	Acetone	15.68	15.99
61	t	2,6-Dichlorobenzoic acid	AN	18.16	18.16

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
62	t	2,6-Dichlorobenzoic acid	DMF	9.71	9.81
63	t	2,6-Dichlorobenzoic acid	DMSO	8.31	8.45
64	t	2,6-Dichlorobenzoic acid	EtOH	8.03	7.90
65	t	2,6-Dichlorobenzoic acid	Water	1.82	1.67
66	t	2,6-Dichlorobenzoic acid	MeOH	7.26	7.18
67	t	2,6-Diethoxybenzoic acid	Isopropanol	11.09	11.10
68	p	2,6-Diethoxybenzoic acid	Acetone	18.88	18.73
69	t	2,6-Diethoxybenzoic acid	AN	20.97	20.93
70	t	2,6-Diethoxybenzoic acid	DMF	12.97	13.00
71	c	2,6-Diethoxybenzoic acid	DMSO	11.38	11.46
72	p	2,6-Diethoxybenzoic acid	EtOH	10.03	10.05
73	c	2,6-Diethoxybenzoic acid	MeOH	9.05	8.91
74	t	2,6-Difluorobenzoic acid	Isopropanol	8.93	8.98
75	t	2,6-Difluorobenzoic acid	Acetone	16.03	15.96
76	t	2,6-Difluorobenzoic acid	AN	18.30	18.24
77	t	2,6-Difluorobenzoic acid	DMF	9.91	9.93
78	t	2,6-Difluorobenzoic acid	DMSO	8.59	8.57
79	t	2,6-Difluorobenzoic acid	EtOH	8.22	8.24
80	t	2,6-Difluorobenzoic acid	MeOH	7.39	7.45
81	p	2,6-Dihydroxybenzoic acid	Water	1.22	1.43
82	c	2,6-Diiodobenzoic acid	Isopropanol	9.18	8.79
83	t	2,6-Diiodobenzoic acid	Acetone	16.34	16.44
84	t	2,6-Diiodobenzoic acid	AN	18.27	18.27
85	p	2,6-Diiodobenzoic acid	DMF	10.13	10.47
86	p	2,6-Diiodobenzoic acid	DMSO	8.75	8.85
87	t	2,6-Diiodobenzoic acid	EtOH	8.22	8.12
88	p	2,6-Diiodobenzoic acid	MeOH	7.29	7.42
89	t	2,6-Diisopropoxybenzoic acid	Isopropanol	11.61	11.46
90	t	2,6-Diisopropoxybenzoic acid	Acetone	19.06	19.20
91	t	2,6-Diisopropoxybenzoic acid	AN	21.17	21.23
92	t	2,6-Diisopropoxybenzoic acid	DMF	12.95	13.37
93	p	2,6-Diisopropoxybenzoic acid	DMSO	11.50	11.80
94	t	2,6-Diisopropoxybenzoic acid	EtOH	10.34	10.34
95	t	2,6-Diisopropoxybenzoic acid	MeOH	9.13	9.08
96	p	2,6-Dimethoxybenzoic acid	Isopropanol	10.81	11.08
97	t	2,6-Dimethoxybenzoic acid	Acetone	18.65	18.49
98	p	2,6-Dimethoxybenzoic acid	AN	20.87	20.73
99	c	2,6-Dimethoxybenzoic acid	DMF	12.72	12.79
100	t	2,6-Dimethoxybenzoic acid	DMSO	11.29	11.23
101	t	2,6-Dimethoxybenzoic acid	EtOH	9.84	10.09
102	t	2,6-Dimethoxybenzoic acid	Water	3.44	3.56
103	p	2,6-Dimethoxybenzoic acid	MeOH	8.57	8.99
104	t	2,6-Dimethylbenzoic acid	Isopropanol	10.32	10.50
105	t	2,6-Dimethylbenzoic acid	Acetone	17.74	17.99
106	t	2,6-Dimethylbenzoic acid	AN	20.31	20.59
107	t	2,6-Dimethylbenzoic acid	DMF	11.96	12.01
108	t	2,6-Dimethylbenzoic acid	DMSO	10.68	10.94
109	t	2,6-Dimethylbenzoic acid	EtOH	9.42	9.66
110	t	2,6-Dimethylbenzoic acid	Water	3.25	3.80
111	t	2,6-Dimethylbenzoic acid	MeOH	8.57	8.78
112	t	2,6-Dinitrobenzoic acid	Isopropanol	7.59	7.34
113	t	2,6-Dinitrobenzoic acid	Acetone	14.28	14.37
114	t	2,6-Dinitrobenzoic acid	AN	16.67	16.63
115	t	2,6-Dinitrobenzoic acid	DMF	8.26	8.33
116	t	2,6-Dinitrobenzoic acid	DMSO	6.72	6.89
117	p	2,6-Dinitrobenzoic acid	EtOH	7.15	6.84
118	t	2,6-Dinitrobenzoic acid	water	1.14	1.07
119	t	2,6-Dinitrobenzoic acid	MeOH	6.49	6.37
120	t	2,6-Dipropoxybenzoic acid	Isopropanol	11.43	11.47
121	t	2,6-Dipropoxybenzoic acid	Acetone	18.98	18.77
122	t	2,6-Dipropoxybenzoic acid	AN	21.08	20.91

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
123	p	2,6-Dipropoxybenzoic acid	DMF	12.92	12.99
124	t	2,6-Dipropoxybenzoic acid	DMSO	11.52	11.50
125	t	2,6-Dipropoxybenzoic acid	EtOH	10.19	10.34
126	t	2,6-Dipropoxybenzoic acid	MeOH	9.19	9.02
127	t	2-Acetamidobenzoic acid	Water	3.67	3.75
128	t	2-Acetylbenzoic acid	Acetone	17.67	17.41
129	t	2-Acetylbenzoic acid	AN	20.51	20.10
130	t	2-Acetylbenzoic acid	DMF	11.68	11.49
131	t	2-Acetylbenzoic acid	EtOH	10.07	9.73
132	t	2-Acetylbenzoic acid	Water	4.13	3.83
133	t	2-Acetylbenzoic acid	MeOH	9.28	8.84
134	t	2-Aminobenzoic acid	Water	4.87	4.41
135	p	2-Bromobenzoic acid	DMSO	9.57	9.24
136	t	2-Bromobenzoic acid	EtOH	8.98	9.03
137	p	2-Bromobenzoic acid	Water	2.85	2.85
138	p	2-Bromobenzoic acid	MeOH	8.19	8.12
139	p	2-Chloro-3-nitrobenzoic acid	Water	2.02	1.84
140	t	2-Chloro-4-nitrobenzoic acid	DMSO	7.96	7.85
141	t	2-Chloro-4-nitrobenzoic acid	Water	1.92	2.14
142	t	2-Chloro-4-nitrobenzoic acid	MeOH	7.12	7.21
143	p	2-Chloro-5-nitrobenzoic acid	DMSO	7.95	7.92
144	t	2-Chloro-5-nitrobenzoic acid	Water	2.17	2.24
145	c	2-Chloro-6-methylbenzoic acid	Water	2.75	2.98
146	p	2-Chloro-6-nitrobenzoic acid	Acetone	15.12	15.27
147	t	2-Chloro-6-nitrobenzoic acid	AN	17.53	17.67
148	t	2-Chloro-6-nitrobenzoic acid	DMF	9.14	9.03
149	t	2-Chloro-6-nitrobenzoic acid	DMSO	7.66	7.66
150	t	2-Chloro-6-nitrobenzoic acid	Water	1.34	1.70
151	t	2-Chloro-6-nitrobenzoic acid	MeOH	6.86	6.90
152	p	2-Chlorobenzoic acid	AN	19.00	19.17
153	t	2-Chlorobenzoic acid	DMA	9.60	9.51
154	t	2-Chlorobenzoic acid	DMSO	9.16	9.51
155	c	2-Chlorobenzoic acid	EtOH	9.08	9.21
156	t	2-Chlorobenzoic acid	Water	2.92	3.06
157	t	2-Chlorobenzoic acid	MeOH	8.31	8.23
158	p	2-Cyanobenzoic acid	Acetone	16.65	16.72
159	c	2-Cyanobenzoic acid	AN	19.07	19.02
160	p	2-Cyanobenzoic acid	DMF	10.37	10.75
161	p	2-Cyanobenzoic acid	Water	3.08	2.99
162	t	2-Cyanobenzoic acid	MeOH	8.15	8.09
163	t	2-Fluorobenzoic acid	EtOH	9.52	9.35
164	t	2-Fluorobenzoic acid	Water	3.57	3.36
165	t	2-Fluorobenzoic acid	MeOH	8.41	8.51
166	t	2-Formylbenzoic acid	Acetone	17.39	17.41
167	t	2-Formylbenzoic acid	AN	20.38	20.48
168	t	2-Formylbenzoic acid	DMF	11.21	11.11
169	t	2-Formylbenzoic acid	Water	4.55	4.05
170	t	2-Formylbenzoic acid	MeOH	9.19	8.91
171	t	2-Hydroxy-3,5-dinitrobenzoic acid	MeOH	5.88	6.03
172	c	2-Hydroxy-5-bromobenzoic acid	Water	2.54	2.58
173	p	2-Hydroxy-5-bromobenzoic acid	MeOH	7.76	7.70
174	t	2-Hydroxy-5-chlorobenzoic acid	Water	2.59	2.64
175	c	2-Hydroxy-5-chlorobenzoic acid	MeOH	7.81	7.80
176	p	2-Hydroxy-5-fluorobenzoic acid	Water	2.70	2.62
177	t	2-Hydroxy-5-fluorobenzoic acid	MeOH	7.91	7.89
178	p	2-Hydroxy-5-iodobenzoic acid	Water	2.62	2.64
179	t	2-Hydroxy-5-nitrobenzoic acid	Water	2.05	1.94
180	t	2-Hydroxy-5-nitrobenzoic acid	MeOH	7.25	7.04
181	t	2-Hydroxybenzoic acid	EtOH	8.45	8.50
182	t	2-Hydroxybenzoic acid	Water	2.77	2.88
183	t	2-Hydroxybenzoic acid	MeOH	7.92	8.03

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
184	p	2-Iodobenzoic acid	DMSO	9.90	9.54
185	t	2-Iodobenzoic acid	EtOH	9.09	8.99
186	t	2-Iodobenzoic acid	Water	2.96	2.86
187	p	2-Iodobenzoic acid	MeOH	8.24	8.10
188	t	2-Methoxy-6-chlorobenzoic acid	Acetone	17.24	17.30
189	t	2-Methoxy-6-chlorobenzoic acid	AN	19.33	19.50
190	t	2-Methoxy-6-chlorobenzoic acid	DMF	11.24	11.35
191	t	2-Methoxy-6-chlorobenzoic acid	DMSO	9.76	9.96
192	t	2-Methoxy-6-chlorobenzoic acid	MeOH	8.06	8.11
193	t	2-Methoxy-6-nitrobenzoic acid	Acetone	16.65	16.72
194	t	2-Methoxy-6-nitrobenzoic acid	AN	18.98	18.87
195	t	2-Methoxy-6-nitrobenzoic acid	DMF	10.65	10.49
196	t	2-Methoxy-6-nitrobenzoic acid	DMSO	9.16	8.85
197	t	2-Methoxy-6-nitrobenzoic acid	MeOH	7.65	7.60
198	p	2-Methoxybenzoic acid	EtOH	10.21	9.79
199	p	2-Methoxybenzoic acid	Water	4.09	3.86
200	t	2-Methoxybenzoic acid	MeOH	9.26	9.13
201	p	2-Methyl-6-chlorobenzoic acid	Acetone	16.74	16.60
202	t	2-Methyl-6-chlorobenzoic acid	AN	19.17	19.11
203	t	2-Methyl-6-chlorobenzoic acid	DMF	10.89	10.63
204	c	2-Methyl-6-chlorobenzoic acid	DMSO	9.48	9.49
205	t	2-Methyl-6-chlorobenzoic acid	Water	2.75	2.71
206	t	2-Methyl-6-chlorobenzoic acid	MeOH	7.91	7.99
207	p	2-Methyl-6-methoxybenzoic acid	Acetone	18.24	18.24
208	t	2-Methyl-6-methoxybenzoic acid	AN	20.52	20.65
209	t	2-Methyl-6-methoxybenzoic acid	DMF	12.34	12.20
210	p	2-Methyl-6-methoxybenzoic acid	DMSO	10.95	10.71
211	t	2-Methyl-6-methoxybenzoic acid	Water	3.46	3.98
212	t	2-Methyl-6-methoxybenzoic acid	MeOH	8.70	9.21
213	c	2-Methyl-6-nitrobenzoic acid	Acetone	16.13	16.16
214	t	2-Methyl-6-nitrobenzoic acid	AN	18.63	18.64
215	p	2-Methyl-6-nitrobenzoic acid	DMF	10.13	10.01
216	t	2-Methyl-6-nitrobenzoic acid	DMSO	8.85	8.63
217	t	2-Methyl-6-nitrobenzoic acid	Water	2.21	2.50
218	p	2-Methyl-6-nitrobenzoic acid	MeOH	7.56	7.59
219	t	2-Methylbenzoic acid	DMSO	11.07	11.15
220	t	2-Methylbenzoic acid	EtOH	10.23	10.18
221	t	2-Methylbenzoic acid	Water	3.91	4.24
222	c	2-Methylbenzoic acid	MeOH	9.31	9.40
223	t	2-Methylesterbenzoic acid	Acetone	17.34	17.53
224	p	2-Methylesterbenzoic acid	AN	19.78	19.99
225	t	2-Methylesterbenzoic acid	DMF	11.33	11.60
226	c	2-Methylesterbenzoic acid	Water	3.18	3.62
227	t	2-Methylesterbenzoic acid	MeOH	8.47	8.55
228	t	2-Nitrobenzoic acid	Acetone	15.98	15.64
229	t	2-Nitrobenzoic acid	AN	18.54	18.21
230	c	2-Nitrobenzoic acid	DMF	9.96	9.46
231	p	2-Nitrobenzoic acid	DMSO	8.18	8.09
232	p	2-Nitrobenzoic acid	EtOH	8.27	8.23
233	t	2-Nitrobenzoic acid	Water	2.19	2.38
234	t	2-Nitrobenzoic acid	MeOH	7.59	7.44
235	t	3,4,5-Trihydoxybenzoic acid	Water	4.21	4.00
236	t	3,4,5-Trimethoxybenzoic acid	DMSO	10.30	10.37
237	t	3,4,5-Trimethylbenzoic acid	DMSO	11.54	11.49
238	t	3,4,5-Trimethylbenzoic acid	MeOH	9.66	9.57
239	t	3,4-Dichlorobenzoic acid	Isopropanol	9.80	10.14
240	p	3,4-Dichlorobenzoic acid	Acetone	16.70	16.76
241	t	3,4-Dichlorobenzoic acid	AN	19.42	19.47
242	t	3,4-Dichlorobenzoic acid	DMF	10.79	10.84
243	t	3,4-Dichlorobenzoic acid	DMSO	9.60	9.51
244	t	3,4-Dichlorobenzoic acid	Water	3.64	3.56

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
245	t	3,4-Dichlorobenzoic acid	MeOH	8.64	8.60
246	t	3,4-Dihydroxybenzoic acid	Water	4.48	4.18
247	t	3,4-Dimethoxybenzoic acid	Acetone	18.59	18.61
248	t	3,4-Dimethoxybenzoic acid	AN	20.95	20.95
249	t	3,4-Dimethoxybenzoic acid	DMF	12.65	12.41
250	t	3,4-Dimethoxybenzoic acid	DMSO	11.40	11.18
251	p	3,4-Dimethoxybenzoic acid	Water	4.44	4.43
252	t	3,4-Dimethoxybenzoic acid	MeOH	9.54	9.60
253	t	3,4-Dimethylbenzoic acid	Isopropanol	11.60	11.11
254	t	3,4-Dimethylbenzoic acid	Acetone	18.71	18.61
255	t	3,4-Dimethylbenzoic acid	AN	21.05	20.89
256	t	3,4-Dimethylbenzoic acid	DMF	12.70	12.64
257	p	3,4-Dimethylbenzoic acid	DMSO	11.46	11.33
258	t	3,4-Dimethylbenzoic acid	Water	4.40	4.37
259	t	3,4-Dimethylbenzoic acid	MeOH	9.63	9.55
260	t	3,4-Dinitrobenzoic acid	Acetone	15.07	15.00
261	p	3,4-Dinitrobenzoic acid	AN	17.50	17.37
262	t	3,4-Dinitrobenzoic acid	DMF	9.07	9.09
263	t	3,4-Dinitrobenzoic acid	DMSO	7.88	7.77
264	t	3,4-Dinitrobenzoic acid	Water	2.82	2.57
265	t	3,4-Dinitrobenzoic acid	MeOH	7.44	7.64
266	t	3,5-Dibromobenzoic acid	Acetone	16.23	16.47
267	c	3,5-Dibromobenzoic acid	AN	18.86	19.22
268	p	3,5-Dibromobenzoic acid	DMF	10.11	10.54
269	t	3,5-Dibromobenzoic acid	DMSO	8.88	9.21
270	t	3,5-Dibromobenzoic acid	MeOH	8.29	8.47
271	t	3,5-Dichlorobenzoic acid	AN	18.75	19.12
272	t	3,5-Dichlorobenzoic acid	DMF	10.43	10.39
273	t	3,5-Dichlorobenzoic acid	DMSO	8.81	9.10
274	t	3,5-Dichlorobenzoic acid	Water	3.10	3.29
275	t	3,5-Dichlorobenzoic acid	MeOH	8.26	8.32
276	t	3,5-Dihydroxybenzoic acid	Water	4.04	3.98
277	t	3,5-Dimethoxybenzoic acid	Acetone	18.02	18.05
278	t	3,5-Dimethoxybenzoic acid	AN	20.39	20.59
279	t	3,5-Dimethoxybenzoic acid	DMF	12.04	11.92
280	t	3,5-Dimethoxybenzoic acid	DMSO	10.71	10.73
281	t	3,5-Dimethoxybenzoic acid	Water	3.97	4.17
282	t	3,5-Dimethoxybenzoic acid	MeOH	9.27	9.29
283	t	3,5-Dimethylbenzoic acid	Acetone	18.54	18.60
284	p	3,5-Dimethylbenzoic acid	AN	20.95	20.90
285	t	3,5-Dimethylbenzoic acid	DMF	12.57	12.62
286	c	3,5-Dimethylbenzoic acid	DMSO	11.29	11.34
287	p	3,5-Dimethylbenzoic acid	Water	4.30	4.32
288	c	3,5-Dimethylbenzoic acid	MeOH	9.59	9.46
289	t	3,5-Dinitrobenzoic acid	Isopropanol	8.30	8.50
290	p	3,5-Dinitrobenzoic acid	Acetone	14.87	15.14
291	t	3,5-Dinitrobenzoic acid	AN	17.36	17.42
292	t	3,5-Dinitrobenzoic acid	DMF	8.76	9.10
293	t	3,5-Dinitrobenzoic acid	DMSO	7.38	7.67
294	t	3,5-Dinitrobenzoic acid	Water	2.67	2.37
295	t	3,5-Dinitrobenzoic acid	MeOH	7.64	7.39
296	c	3,6-Dichloro-2-methoxybenzoic acid	Water	1.97	2.09
297	t	3-Acetamidobenzoic acid	Water	4.06	4.17
298	t	3-Acetamidobenzoic acid	MeOH	9.25	9.28
299	t	3-Acetylbenzoic acid	DMSO	10.22	10.50
300	t	3-Acetylbenzoic acid	Water	3.83	4.07
301	t	3-Acetylbenzoic acid	MeOH	8.87	9.25
302	t	3-Amino-2,5-dichlorobenzoic acid	Water	3.40	2.82
303	t	3-Aminobenzoic acid	DMSO	11.60	11.08
304	t	3-Aminobenzoic acid	Water	4.74	4.25
305	t	3-Bromo-4-methoxybenzoic acid	Acetone	17.82	17.93

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
306	p	3-Bromo-4-methoxybenzoic acid	AN	20.38	20.42
307	c	3-Bromo-4-methoxybenzoic acid	DMF	11.83	11.98
308	t	3-Bromo-4-methoxybenzoic acid	DMSO	10.61	10.63
309	t	3-Bromo-4-methoxybenzoic acid	MeOH	9.27	9.34
310	c	3-Bromo-4-methylbenzoic acid	Acetone	17.59	17.51
311	t	3-Bromo-4-methylbenzoic acid	AN	20.03	20.11
312	t	3-Bromo-4-methylbenzoic acid	DMF	11.49	11.62
313	c	3-Bromo-4-methylbenzoic acid	DMSO	10.30	10.29
314	t	3-Bromo-4-methylbenzoic acid	Water	3.29	3.64
315	p	3-Bromo-4-methylbenzoic acid	MeOH	9.06	9.10
316	t	3-Bromobenzoic acid	Isopropanol	10.11	10.26
317	t	3-Bromobenzoic acid	AN	19.50	19.76
318	t	3-Bromobenzoic acid	DMF	11.20	11.16
319	t	3-Bromobenzoic acid	DMSO	9.68	9.84
320	t	3-Bromobenzoic acid	EtOH	9.40	9.65
321	c	3-Bromobenzoic acid	Water	3.81	3.78
322	t	3-Bromobenzoic acid	MeOH	8.80	8.89
323	t	3-Chloro-4-nitrobenzoic acid	Acetone	15.78	15.98
324	p	3-Chloro-4-nitrobenzoic acid	AN	18.28	18.52
325	t	3-Chloro-4-nitrobenzoic acid	DMF	10.07	9.99
326	t	3-Chloro-4-nitrobenzoic acid	DMSO	8.61	8.65
327	t	3-Chloro-4-nitrobenzoic acid	MeOH	8.09	7.91
328	t	3-Chloro-5-nitrobenzoic acid	Acetone	15.57	15.77
329	t	3-Chloro-5-nitrobenzoic acid	AN	17.98	18.30
330	t	3-Chloro-5-nitrobenzoic acid	DMF	9.60	9.68
331	t	3-Chloro-5-nitrobenzoic acid	DMSO	8.37	8.30
332	t	3-Chloro-5-nitrobenzoic acid	Water	3.13	2.66
333	t	3-Chloro-5-nitrobenzoic acid	MeOH	8.03	7.67
334	p	3-Chlorobenzoic acid	DMA	9.80	10.05
335	p	3-Chlorobenzoic acid	DMF	10.95	11.39
336	p	3-Chlorobenzoic acid	DMSO	9.51	10.05
337	t	3-Chlorobenzoic acid	Water	3.80	3.90
338	p	3-Chlorobenzoic acid	MeOH	8.83	8.99
339	t	3-Cyanobenzoic acid	DMSO	9.44	9.69
340	t	3-Cyanobenzoic acid	Water	3.60	3.64
341	t	3-Cyanobenzoic acid	MeOH	8.53	8.75
342	p	3-Fluorobenzoic acid	Water	3.88	3.71
343	t	3-Fluorobenzoic acid	MeOH	8.87	8.74
344	t	3-Formylbenzoic acid	Water	3.84	3.97
345	t	3-Hydroxybenzoic acid	DMA	10.75	10.45
346	t	3-Hydroxybenzoic acid	Water	4.03	3.97
347	p	3-Hydroxybenzoic acid	MeOH	9.58	9.04
348	t	3-Iodobenzoic acid	DMSO	10.16	9.96
349	t	3-Iodobenzoic acid	Water	3.82	3.86
350	t	3-Iodobenzoic acid	MeOH	8.89	9.02
351	t	3-Methoxy-4-chlorobenzoic acid	Acetone	17.39	17.32
352	t	3-Methoxy-4-chlorobenzoic acid	AN	20.02	19.97
353	t	3-Methoxy-4-chlorobenzoic acid	DMF	11.58	11.42
354	p	3-Methoxy-4-chlorobenzoic acid	DMSO	10.39	10.22
355	p	3-Methoxy-4-chlorobenzoic acid	MeOH	8.98	8.88
356	p	3-Methoxy-4-methylbenzoic acid	Acetone	18.32	18.27
357	t	3-Methoxy-4-methylbenzoic acid	AN	20.76	20.69
358	t	3-Methoxy-4-methylbenzoic acid	DMF	12.39	12.30
359	t	3-Methoxy-4-methylbenzoic acid	DMSO	11.06	11.10
360	t	3-Methoxy-4-methylbenzoic acid	Water	4.13	4.23
361	c	3-Methoxy-4-methylbenzoic acid	MeOH	9.49	9.34
362	t	3-Methoxy-4-nitrobenzoic acid	Acetone	16.21	16.57
363	t	3-Methoxy-4-nitrobenzoic acid	AN	18.77	19.35
364	t	3-Methoxy-4-nitrobenzoic acid	DMF	10.33	10.80
365	t	3-Methoxy-4-nitrobenzoic acid	DMSO	9.22	9.58
366	t	3-Methoxy-4-nitrobenzoic acid	MeOH	9.28	9.33

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
367	t	3-Methoxy-5-bromobenzoic acid	Acetone	17.03	17.25
368	p	3-Methoxy-5-bromobenzoic acid	AN	19.73	19.85
369	t	3-Methoxy-5-bromobenzoic acid	DMF	11.28	11.32
370	t	3-Methoxy-5-bromobenzoic acid	DMSO	9.79	10.03
371	t	3-Methoxy-5-bromobenzoic acid	MeOH	8.79	8.80
372	t	3-Methoxy-5-nitrobenzoic acid	Acetone	16.31	16.03
373	c	3-Methoxy-5-nitrobenzoic acid	AN	18.72	18.71
374	p	3-Methoxy-5-nitrobenzoic acid	DMF	10.37	10.17
375	p	3-Methoxy-5-nitrobenzoic acid	DMSO	9.01	8.90
376	t	3-Methoxy-5-nitrobenzoic acid	MeOH	8.40	8.57
377	t	3-Methoxybenzoic acid	DMSO	11.10	10.72
378	t	3-Methoxybenzoic acid	Water	4.12	4.06
379	t	3-Methoxybenzoic acid	MeOH	9.30	9.16
380	p	3-Methyl-4-chlorobenzoic acid	Acetone	17.59	17.55
381	t	3-Methyl-4-chlorobenzoic acid	AN	20.27	20.14
382	c	3-Methyl-4-chlorobenzoic acid	DMF	11.70	11.67
383	t	3-Methyl-4-chlorobenzoic acid	DMSO	10.54	10.39
384	t	3-Methyl-4-chlorobenzoic acid	Water	4.07	3.95
385	p	3-Methyl-4-chlorobenzoic acid	MeOH	9.12	9.07
386	t	3-Methyl-4-methoxybenzoic acid	Acetone	19.01	18.90
387	t	3-Methyl-4-methoxybenzoic acid	AN	21.29	21.09
388	t	3-Methyl-4-methoxybenzoic acid	DMF	12.84	12.87
389	t	3-Methyl-4-methoxybenzoic acid	DMSO	11.71	11.59
390	p	3-Methyl-4-methoxybenzoic acid	Water	4.35	4.47
391	t	3-Methyl-4-methoxybenzoic acid	MeOH	9.81	9.69
392	t	3-Methyl-4-nitrobenzoic acid	Acetone	16.55	16.38
393	t	3-Methyl-4-nitrobenzoic acid	AN	19.21	18.97
394	t	3-Methyl-4-nitrobenzoic acid	DMF	11.03	10.55
395	t	3-Methyl-4-nitrobenzoic acid	DMSO	9.48	9.29
396	c	3-Methyl-4-nitrobenzoic acid	Water	3.65	3.31
397	t	3-Methyl-4-nitrobenzoic acid	MeOH	8.54	8.40
398	t	3-Methyl-5-chlorobenzoic acid	Acetone	17.40	17.40
399	p	3-Methyl-5-chlorobenzoic acid	AN	19.85	19.99
400	t	3-Methyl-5-chlorobenzoic acid	DMF	11.35	11.50
401	t	3-Methyl-5-chlorobenzoic acid	DMSO	10.04	10.18
402	t	3-Methyl-5-chlorobenzoic acid	MeOH	8.94	8.95
403	t	3-Methyl-5-methoxybenzoic acid	Acetone	18.26	18.26
404	t	3-Methyl-5-methoxybenzoic acid	AN	20.66	20.71
405	p	3-Methyl-5-methoxybenzoic acid	DMF	12.32	12.26
406	t	3-Methyl-5-methoxybenzoic acid	DMSO	10.93	11.08
407	t	3-Methyl-5-methoxybenzoic acid	MeOH	9.44	9.27
408	p	3-Methyl-5-nitrobenzoic acid	Acetone	16.62	16.37
409	t	3-Methyl-5-nitrobenzoic acid	AN	19.18	18.98
410	t	3-Methyl-5-nitrobenzoic acid	DMF	10.92	10.51
411	c	3-Methyl-5-nitrobenzoic acid	DMSO	9.36	9.22
412	t	3-Methyl-5-nitrobenzoic acid	MeOH	8.55	8.43
413	t	3-Methylbenzoic acid	DMSO	11.26	10.98
414	t	3-Methylbenzoic acid	Water	4.24	4.21
415	t	3-Methylbenzoic acid	MeOH	9.47	9.36
416	p	3-Nitro-4-chlorobenzoic acid	Isopropanol	9.30	9.52
417	c	3-Nitro-4-chlorobenzoic acid	Acetone	15.98	15.99
418	t	3-Nitro-4-chlorobenzoic acid	AN	18.57	18.67
419	t	3-Nitro-4-chlorobenzoic acid	DMF	9.98	10.03
420	p	3-Nitro-4-chlorobenzoic acid	DMSO	8.74	8.68
421	p	3-Nitro-4-chlorobenzoic acid	Water	3.29	3.14
422	t	3-Nitro-4-chlorobenzoic acid	MeOH	8.24	8.16
423	t	3-Nitro-4-methoxybenzoic acid	Acetone	17.04	17.01
424	t	3-Nitro-4-methoxybenzoic acid	AN	19.51	19.70
425	t	3-Nitro-4-methoxybenzoic acid	DMF	11.18	11.12
426	t	3-Nitro-4-methoxybenzoic acid	DMSO	9.85	9.79
427	p	3-Nitro-4-methoxybenzoic acid	Water	3.72	4.07

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
428	p	3-Nitro-4-methoxybenzoic acid	MeOH	8.83	9.22
429	t	3-Nitro-4-methylbenzoic acid	Acetone	16.75	16.59
430	p	3-Nitro-4-methylbenzoic acid	AN	19.36	19.28
431	t	3-Nitro-4-methylbenzoic acid	DMF	10.93	10.73
432	t	3-Nitro-4-methylbenzoic acid	DMSO	9.45	9.41
433	p	3-Nitro-4-methylbenzoic acid	Water	3.62	3.64
434	c	3-Nitro-4-methylbenzoic acid	MeOH	8.61	8.74
435	t	3-Nitrobenzoic acid	AN	19.29	19.06
436	t	3-Nitrobenzoic acid	DMSO	9.17	9.12
437	p	3-Nitrobenzoic acid	Water	3.46	3.43
438	t	3-Nitrobenzoic acid	MeOH	8.32	8.50
439	t	3-Trifluoromethylbenzoic acid	Water	3.75	4.07
440	t	3-Trifluoromethylbenzoic acid	MeOH	8.69	8.94
441	t	4-(Dimethylamino)benzoic acid	Water	5.03	4.86
442	t	4-(*N*,*N*-dimethylaminocarbonyl)benzoic acid	Acetone	17.83	17.70
443	t	4-(*N*,*N*-dimethylaminocarbonyl)benzoic acid	AN	20.55	20.21
444	p	4-(*N*,*N*-dimethylaminocarbonyl)benzoic acid	MeOH	9.00	9.24
445	p	4-Acetamidobenzoic acid	Water	4.30	4.52
446	t	4-Acetamidobenzoic acid	MeOH	9.57	9.78
447	p	4-Acetylbenzoic acid	Acetone	17.32	17.49
448	t	4-Acetylbenzoic acid	AN	19.95	20.07
449	t	4-Acetylbenzoic acid	DMF	11.38	11.65
450	t	4-Acetylbenzoic acid	Water	3.74	3.89
451	t	4-Acetylbenzoic acid	MeOH	8.84	9.02
452	t	4-Aminobenzoic acid	Water	4.82	4.56
453	t	4-Aminobenzoic acid	MeOH	10.25	9.84
454	t	4-Bromobenzoic acid	AN	20.30	19.75
455	t	4-Bromobenzoic acid	DMSO	10.20	9.79
456	t	4-Bromobenzoic acid	Water	3.99	3.72
457	p	4-Bromobenzoic acid	MeOH	8.93	8.84
458	t	4-Chlorobenzoic acid	DMA	10.30	10.21
459	p	4-Chlorobenzoic acid	DMSO	10.10	10.58
460	t	4-Chlorobenzoic acid	Water	4.00	4.13
461	t	4-Chlorobenzoic acid	MeOH	9.09	9.26
462	t	4-Cyanobenzoic acid	Acetone	16.69	16.95
463	t	4-Cyanobenzoic acid	AN	19.34	19.62
464	t	4-Cyanobenzoic acid	DMF	11.02	11.05
465	t	4-Cyanobenzoic acid	DMSO	9.27	9.73
466	t	4-Cyanobenzoic acid	Water	3.53	3.56
467	t	4-Cyanobenzoic acid	MeOH	8.49	8.66
468	c	4-Ethoxybenzoic acid	DMSO	11.50	11.52
469	t	4-Ethoxybenzoic acid	Water	4.80	4.44
470	p	4-Ethylbenzoic acid	Water	4.35	4.27
471	t	4-Fluorobenzoic acid	Water	4.15	3.99
472	p	4-Fluorobenzoic acid	MeOH	9.23	9.13
473	t	4-Formylbenzoic acid	Acetone	17.10	17.38
474	t	4-Formylbenzoic acid	AN	19.74	20.04
475	t	4-Formylbenzoic acid	DMF	11.19	11.49
476	t	4-Formylbenzoic acid	Water	3.69	3.68
477	t	4-Formylbenzoic acid	MeOH	8.94	8.68
478	t	4-Hydroxy-3,5-dimethoxybenzoic acid	Water	4.34	4.38
479	t	4-Hydroxy-3-methoxybenzoic acid	Water	4.51	4.31
480	t	4-Hydroxybenzoic acid	AN	20.80	20.77
481	t	4-Hydroxybenzoic acid	DMA	11.25	11.15
482	t	4-Hydroxybenzoic acid	Water	4.55	4.38
483	t	4-Hydroxybenzoic acid	MeOH	9.99	9.59
484	t	4-Iodo-3-nitrobenzoic acid	DMSO	8.65	8.43
485	p	4-Iodobenzoic acid	Water	3.98	3.70
486	t	4-Iodobenzoic acid	MeOH	9.04	8.84
487	t	4-Methoxybenzoic acid	DMSO	11.80	11.37
488	t	4-Methoxybenzoic acid	Water	4.25	4.40

Table 1. (cont.)

No.	Set[a]	Compound	Solvent	Experimental	Calculated
489	t	4-Methoxybenzoic acid	MeOH	9.79	9.62
490	p	4-Methylbenzoic acid	DMSO	11.42	11.20
491	c	4-Methylbenzoic acid	Water	4.38	4.37
492	p	4-Methylbenzoic acid	MeOH	9.61	9.55
493	t	4-Methylesterbenzoic acid	Acetone	17.37	17.23
494	t	4-Methylesterbenzoic acid	AN	19.91	19.82
495	p	4-Methylesterbenzoic acid	DMF	11.45	11.50
496	t	4-Methylesterbenzoic acid	Water	3.74	3.90
497	t	4-Methylesterbenzoic acid	MeOH	8.84	9.07
498	t	4-Nitrobenzoic acid	Isopropanol	9.60	9.60
499	t	4-Nitrobenzoic acid	Acetone	16.42	16.16
500	t	4-Nitrobenzoic acid	AN	19.08	18.78
501	c	4-Nitrobenzoic acid	DMF	10.80	10.28
502	t	4-Nitrobenzoic acid	DMSO	9.04	8.98
503	t	4-Nitrobenzoic acid	EtOH	8.90	8.99
504	c	4-Nitrobenzoic acid	Water	3.43	3.13
505	p	4-Nitrobenzoic acid	MeOH	8.36	8.20
506	t	4-Propoxybenzoic acid	Water	4.46	4.59
507	t	4-*tert*-Butylbenzoic acid	Water	4.36	3.93
508	t	4-*tert*-Butylbenzoic acid	MeOH	9.61	9.11
509	t	Acetylsalicylic acid	DMF	11.30	11.26
510	t	Acetylsalicylic acid	Water	3.69	3.73
511	p	Benzoic acid	Isopropanol	10.71	10.62
512	t	Benzoic acid	Acetone	18.20	18.00
513	c	Benzoic acid	AN	20.70	20.49
514	t	Benzoic acid	DMA	11.00	10.74
515	p	Benzoic acid	DMF	12.27	12.05
516	t	Benzoic acid	DMSO	11.00	10.73
517	t	Benzoic acid	EtOH	10.25	9.99
518	t	Benzoic acid	Water	4.21	4.07
519	c	Benzoic acid	MeOH	9.41	9.25

[a] Training set (t); prediction set (p); cross-validation set (c).

order to ensure that all the calculated geometries correspond to true minima. The MOPAC output files were used by CODESSA to calculate some hundreds of molecular descriptors which can be classified in: constitutional, topological, geometrical, electrostatic, quantum-chemical, and Charge Partial Surface Area (CPSA) types.

2.3 Solvent Descriptors

The solvents were characterized by several physical properties (molecular weight, density, molar volume, refractive index, dielectric constant, dipole moment, polarizability, refractivity, standard molar vaporization enthalpy, standard internal energy of vaporization, and Hildebrand's solubility parameter), and also by parameters present in the most popular empirical model-dependent solvents scales [30–36]. In total, 25 descriptors were calculated for each solvent.

The heuristic multilinear regression procedures of the CODESSA program were used to make the first reduction pool of the descriptors; the initial number of 629 was reduced, after the application of the collinearity control methods, to 240 descriptors: 221 of the benzoic acids and 19 of the solvents.

2.4 CNN Methods (ADAPT)

The computations were performed with the ADAPT (Automated Data Analysis and Pattern recognition Toolkit) program [37, 38], including feature selection routines (genetic algorithm [39] and simulated annealing [40]) and CNN procedures [41]. The CNNs used for this analysis are three-layered, fully connected, feed-forward networks, and they have been described in detail by Jurs and coworkers [41, 42]. The number of neurons of the input layer corresponds to the number of descriptors in the model. The number of hidden neurons controls the flexibility of the network and has to be adjusted until the optimal network architecture is achieved. This step is done by means of a building-up procedure that consists in starting with a lesser number of hidden neurons and increasing it one unit until the results achieved with that architecture are not better than those obtained with the previous one. In this work, we built architectures from 5 to 18 hidden neurons, stop-

ping at this point because the results obtained did not improve those derived from a 7-17-1 architecture. So, the architecture selected for the subsequent calculations will be 7-17-1. The output layer contains one neuron representing the pK_a value. The 240 descriptors selected with the previous heuristic method were imported to the ADAPT program, these descriptors were then subjected to the objective feature selection procedures of this program, and a reduced pool of 82 descriptors was obtained: 73 from the solutes and 9 from the solvents. Full CNNs were developed using a genetic algorithm descriptor selection routine with a CNN for evaluating the fitness of each subset of descriptors selected. The fitness of descriptor subsets was calculated as $COST = TSET + 0.4|TSET - CVSET|$, where TSET and CVSET denote rms errors for the training and cross-validation sets, respectively. Models chosen with this quality factor performed better than models chosen with just training set rms error as the quality factor. That is, CNNs that produce training and cross-validation set errors that are low and similar in magnitude tend to perform well in predicting the properties of interest for compounds not used in the training process. A quasi-Newton method BFGS (Broyden – Flectcher – Golfarb – Shanno) [42] is used to train the network. It should be noted that the ratio of training set observations to adjustable parameters should be kept above 2.0 to avoid overtraining [43]. The number of adjustable parameters is computed as $AP = [(IL+1)HL] + [(HL+1)OL]$, where IL, HL, and OL denote the number of neurons in the input layer, hidden layer, and output layer, respectively. With this architecture, the network contains 154 adjustable parameters, corresponding to a ratio of 2.43 for training observations (379) to adjustable parameters, above the minimum acceptable ratio of 2.0. The ten models derived by ADAPT were analyzed and evaluated by the external prediction set and the best one was selected.

2.5 CNN Methods (SVM)

The ADAPT program is restricted to work with not more than 500 compounds, thus in the management of the complete set containing the benzoic acids and the phenols (set B), we have used Support Vector Machine (SVM) methods. This set B is composed of the 793 pK_a values corresponding to the benzoic acids studied here (set A) and to the phenols reported in our previous paper [14]. The SVM method has been developed by Vapnik [44]. A detailed description of the theory of SVM can be found in some books and tutorials [45, 46]. In support vector regression, the vectors of descriptors are mapped into a higher dimensional feature space, with the use of a kernel function, in which the model is developed. The generalization performance of SVR depends on the kernel function selected because it defines the distribution of the training set samples in the high dimensional feature space. The LibSVM [47] package, which was used in our work, includes four types of kernel functions for SVR studies: linear, polynomial, sigmoid, and Gaussian. The Gaussian (or RBF) kernel functions are often chosen for support vector regression due to their effectiveness and speed in the training process. The parameters that should be optimized in an SVR procedure with RBF kernel are: the width of the Gaussian functions γ, the regularization constant C, and the ε of ε-insensitive loss function. The parameter γ is a measure of the interpolation capacity of the Gaussian function. C is a regularization parameter that controls the trade-off between fitting the training data and the generalization capability of the model. If C is small, insufficient stress will be placed in fitting the data; on the other hand, if C is too large, overfitting may happen. Liu *et al.* [48] showed that the prediction error of the model is not heavily influenced by the value of C, and a high value of this parameter allows to stabilize the model, *i.e.*, $C = 100$. The parameter ε of ε-insensitive loss function characterizes the prediction error penalty; the optimal value of this parameter is related with the type of noise present in the data, which is usually unknown. This value is critical because it directly affects the number of support vectors used to establish the model; as ε gets bigger, lesser number of support vectors are selected and worse prediction capability for the model should be expected. The optimization of these three parameters: γ, C, and ε is done by means of recursive leave-one-out procedures of minimization of training set cross-validation error. In the present paper, all calculation implementing SVM with RBF as kernel functions [49, 50] are written in an M-file based on the MATLAB program and all the models are constructed using the software LibSVM [47]. The prediction set used to test the model contains 153 values and was formed by the addition of the corresponding prediction sets of the benzoic acids and the phenols. On the other hand, the training set was composed of the 640 pK_a values contained in the training and cross-validation sets, of the benzoic acids and phenols, used in the ADAPT program.

3 Results and Discussion

The derived model for set A contains the seven descriptors shown in Table 2: five descriptors are of the solutes and the other two are of the solvents. Concerning the solute descriptors, the maximum partial charge for a hydrogen atom is an electrostatic descriptor calculated by the method proposed by Zefirov *et al.* [51] based on the Sanderson scale of electronegativity. The hydrogen atom involved in all the cases corresponds to the carboxylic group. The second descriptor, the Fractional Partial Positive Surface Area (FPSA-2) belongs to the CPSA type [52], which combines the surface area of the whole molecule and the charge distribution in the molecule, and therefore is responsible for polar interactions between molecules. In general, these CPSA descriptors are calculated from the con-

Table 2. Seven descriptors forming the model.

Descriptor
Maximum partial charge for a H atom (solute)
Fractional partial positive surface area, FPSA-2 (solute)
Maximum electron–electron repulsion for an O atom (solute)
Minimum resonance energy for O–H bond (solute)
Minimum valency of a C atom (solute)
Hydrogen-bond donation ability (Kamlet and Taft), α (solvent)
Standard internal energy of vaporization or cohesive energy ($\Delta_{vap}U^{\circ}$) (solvent)

tributions related to the atomic partial charges and the molecular solvent-accessible surface area. In this case, FPSA-2 corresponds to the ratio between the total charged weighted Partially Positively Charged Surface Area (PPSA-2) and the Total Molecular Surface Area (TMSA). The other three descriptors are of the quantum-chemical type; the third descriptor is the maximum electron–electron repulsion for an O atom. The electron–electron repulsion energy for a given atomic species describes the electron repulsion-driven processes in the molecule and may be related to the atomic reactivity in the molecule. The extreme values (maximum or minimum) of this energy can be considered as global characteristics of a compound [53]. In this case, the extreme value corresponds to the oxygen atom of the carboxylic group; thus, it specifies the particular chemical activity of this atom. The fourth descriptor is the minimum resonance energy for the O–H bond and is related with the energy of this bond. The fifth descriptor is the minimum valency of a C atom; the free valency of a given atomic species or its extreme value in the molecule can be used as molecular stability descriptors, and they are useful for predicting reactive centers in the molecule [54]; the carbon atom involved belongs to the carboxylic group. The other two descriptors completing the model correspond to the solvent; the hydrogen-bond donation ability, α, belongs to the Kamlet–Taft scale [30], which is one of the most used scales for the hydrogen bonding ability of the solvents; it is based on solvatochromic parameters and measures the ability to donate hydrogen bonds of the solvent molecules to a general solute. The last descriptor is the standard internal energy of vaporization ($\Delta_{vap}U^{\circ}$) [55], and it measures the energy necessary to bring the molecules of the liquid from their equilibrium distances to an infinite distance.

The statistical results obtained with this model are very good, for the training set the correlation coefficient, R^2, is 0.998 and the RMSE is 0.21; the same R^2 and RMSE values are obtained for the prediction and cross-validation sets, showing the robustness and the high prediction capacity of the model. Figure 1 shows a plot of the calculated *versus* experimental pK_a for all the compounds studied (training, prediction, and cross-validation sets). Figure 2 shows the distribution of the predicted errors for the 519 pK_a values. For 75% of these values, the prediction errors

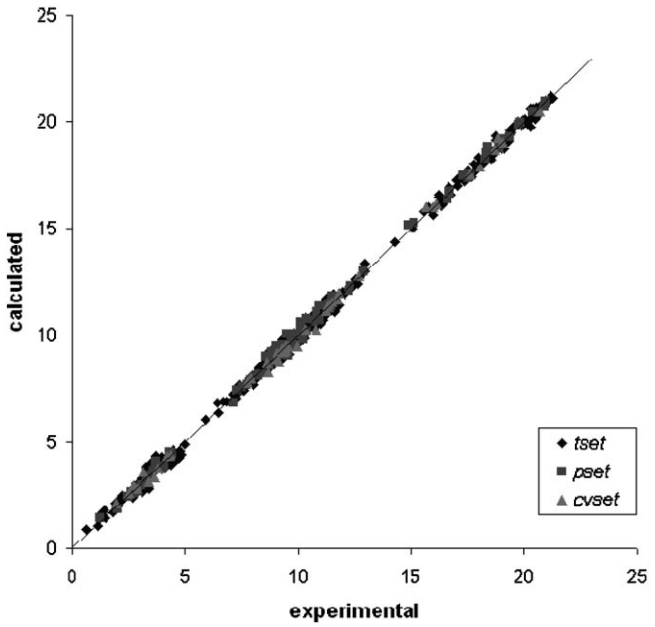

Figure 1. Plot of calculated *vs.* experimental pK_a values for the training, prediction, and cross-validation sets of benzoic acids (set A).

are within 0.25 units, and only 2.5% have errors larger than 0.5 units. To ensure that the model developed in this QSPR study was not due to chance effects, Y-randomization experiments were conducted. The average R^2 and the RMSE for the five experiments performed are 0.1 and 6, respectively, showing that chance correlation did not play any significant role in this study. We have repeated the neural network analysis five more times varying in each case the splitting of the original data set into training, cross-validation, and prediction subsets. The statistical parameters obtained in each case are very similar to those reported earlier. The average values for R^2 and RMSE are 0.997 and 0.22, respectively, showing the goodness of the proposed model.

As is well known, the dissociation reaction in solution is a complex process that is highly dependent on the nature of the solvent involved, since most of the solvents have acidic or basic properties themselves. This process can be represented by an equilibrium containing the remaining non-dissociated solvated acid molecules and the solvated conjugated base and proton formed. On the other hand, the intermolecular solute/solvent interactions are also very complex and have been conventionally divided into two types [56]. The first is associated to non-specific effects. These effects are related to the bulk of the solvent and include the solvent dielectric polarization in the field of the solute molecule, the isotropic dispersion interactions, and the solute cavity formation. The second type of solvent effects, the specific ones, involves the formation of chemical bonds and other anisotropic interactions between the solute and solvent molecules in the solution. In most cases,

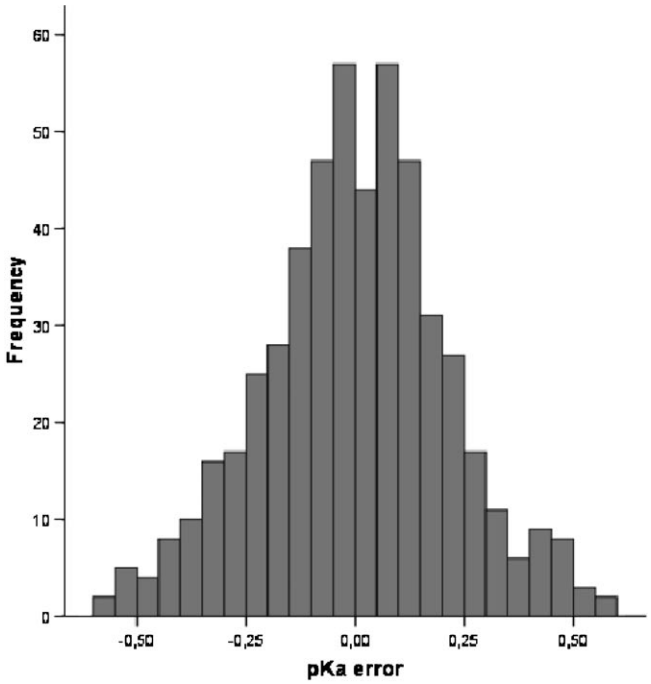

Figure 2. Distribution of pK_a error for benzoic acids (set A).

these specific effects are related to the Lewis donor–acceptor interactions or to hydrogen bonding between the solute and the solvent molecules. From the pioneer works of Hammett and Taft, the pK_a values of benzoic acids and other organic molecules have been analyzed by the steric and electronic effects of the substituents present and their position in the phenyl ring; however, it is now known that the influence of these steric and electronic effects of the substituents is also highly dependent on the solvent used in the pK_a determination [20]. Thus, for example, the stabilization of conjugate bases of substituted derivatives is aided by the solvation of substituents that can form hydrogen bonds; thus, the solvent has an important task in the solvation of substituted salicylic acids and aminobenzoic acids. Whereas in protic solvents the hydroxy groups present are significantly solvated, in aprotic solvents, the formation of intramolecular hydrogen bonds between OH or NH_2 groups and the reaction center is observed. As, in general, aprotic solvents do not solvate well the conjugate base, the effect of substituents on its stabilization is more significant. According to these arguments, the pK_a values of benzoic acids are the result of different delicate interactions that are consequences of the nature and position of the substituents in the ring and of the characteristics of the solvent used.

Although one of the drawbacks of the neural networks is the matter of interpretability, in spite of the complexity of the dissociation process mentioned before, the descriptors of the proposed model allow some explanations of the dissociation process studied. Thus, four solute descriptors of the model are clearly related to the stability, reactivity, or the tendency to dissociate the carboxylic group. The maximum partial charge for a hydrogen atom descriptor reflects the polarity of the O–H bond that is cleaved in the dissociation process, and consequently the facility of the compounds to give protons and benzoate anions. Also, the minimum resonance energy for the O–H bond descriptor is related to the energy of the bond that dissociates and the maximum electron–electron repulsion of oxygen and the minimum valency of carbon atom descriptors account for the reactivity of these atoms, respectively. On the other hand, the fifth solute descriptor, FPSA-2, represents a density of charge of the solute and explains the intermolecular interactions, in this case those between the benzoic acid and the solvent. All these five solute descriptors can be associated to the non-specific solute/solvent interactions. Of the two solvent descriptors, the standard internal energy of vaporization or cohesive energy, $\Delta_{vap}U^{\circ}$, indicates the energy involved in the solute cavity formation in the bulk of the solvent in the dissolution process and can also be associated to the non-specific solute/solvent interactions. The other solvent descriptor, the solvent acidity, α, is obviously associated to the specific hydrogen-bonding interactions and stands for the hydrogen-bond donor properties of the solvents.

A simple method to measure the relative importance of the descriptors in CNN derived models has been proposed [57]. The first input descriptor is randomly scrambled, and then the neural network model is used to predict the property. Because the values of the descriptor have been scrambled, the correlation between descriptor values and property values is obscured. As a result, the RMSE for these new predictions should be larger than the RMSE of the model, the so-called base RMSE. The difference between this RMSE value and the base RMSE indicates the importance of the descriptor to the predictive ability of the model. The values of these increases in RMSE range from 6.30 to 0.23; the solvent descriptors – the hydrogen-bond acidity, α, and the $\Delta_{vap}U^{\circ}$, with values of 6.30 and 3.70, respectively – being the most significant. This result reflects the experimental fact that the variation in the pK_a values is larger for one benzoic acid in two solvents than between two benzoic acids in the same solvent.

The calculated pK_a values (Table 1) show that the highest residual is -0.59, lower than three times the RMSE; the 31 entries with residuals greater than twice the RMSE are collected in Table 3, ranging from -0.59 for the 2,3-dimethylbenzoic acid to $+0.58$ for the 3-amino-2,5-dichlorobenzoic acid, the solvent being water in both cases. No specific features derived from the molecular structure of the benzoic acids, nor has the nature of the solvents been observed for these residuals. Thus, these 31 values correspond to 24 different benzoic acids in 6 solvents; there are 15 positive and 16 negative residuals, 17 of which were determined in protic solvents and 14 in aprotic ones. On the other hand, the position of the substituents in the aromatic ring is not significant, since there are 12 values due to *or-*

Table 3. Predictions with residuals larger than twice the RMSE.

Benzoic acid	Solvent	Entry	pK_a exp.	pK_a calc.	Residual
2,3-Dimethylbenzoic acid	Water	19	3.74	4.33	−0.59
3-Methoxy-4-nitrobenzoic acid	AN	363	18.77	19.35	−0.58
2,6-Dimethylbenzoic acid	Water	110	3.25	3.80	−0.55
3-Chlorobenzoic acid	DMSO	336	9.51	10.05	−0.54
2-Methyl-6-methoxybenzoic acid	Water	211	3.46	3.98	−0.52
2-Methyl-6-methoxybenzoic acid	MeOH	212	8.70	9.21	−0.51
2,3-Dimethylbenzoic acid	MeOH	20	8.98	9.48	−0.50
4-Chlorobenzoic acid	DMSO	459	10.10	10.58	−0.48
3-Methoxy-4-nitrobenzoic acid	DMF	364	10.33	10.80	−0.47
4-Cyanobenzoic acid	DMSO	465	9.27	9.73	−0.46
2,3,5-Trimethylbenzoic acid	MeOH	12	9.05	9.50	−0.45
2,3,5-Trimethylbenzoic acid	DMSO	11	11.24	11.69	−0.45
3-Chlorobenzoic acid	DMF	335	10.95	11.39	−0.44
2-Methylesterbenzoic acid	Water	226	3.18	3.62	−0.44
3,5-Dibromobenzoic acid	DMF	268	10.11	10.54	−0.43
4-*tert*-Butylbenzoic acid	Water	507	4.36	3.93	0.43
4-Methoxybenzoic acid	DMSO	487	11.80	11.37	0.43
2-Acetylbenzoic acid	MeOH	133	9.28	8.84	0.44
2-Aminobenzoic acid	Water	134	4.87	4.41	0.46
3-Chloro-5-nitrobenzoic acid	Water	332	3.13	2.66	0.47
3-Methyl-4-nitrobenzoic acid	DMF	394	11.03	10.55	0.48
3-Aminobenzoic acid	Water	304	4.74	4.25	0.49
3,4-Dimethylbenzoic acid	Isopropanol	253	11.60	11.11	0.49
2-Formylbenzoic acid	Water	169	4.55	4.05	0.50
4-*tert*-butylbenzoic acid	MeOH	508	9.61	9.11	0.50
2-Nitrobenzoic acid	DMF	230	9.96	9.46	0.50
3-Aminobenzoic acid	DMSO	303	11.60	11.08	0.52
4-Nitrobenzoic acid	DMF	501	10.80	10.28	0.52
3-Hydroxybenzoic acid	MeOH	347	9.58	9.04	0.54
4-Bromobenzoic acid	AN	454	20.30	19.75	0.55
3-Amino-2,5-dichlorobenzoic acid	Water	302	3.40	2.82	0.58

tho-substituted benzoic acids and 19 to *meta*- and/or *para*-substituted ones. Table 4 shows the statistical results found in the predicted values of pK_a for several subsets of benzoic acids The overall set of 519 values was separated into subsets according to the presence or absence of the substituents in *ortho*-position, and depending on the protic or aprotic nature of the solvents; even small subsets composed of non-*ortho*- or *ortho*-substituted acids in protic or aprotic solvents, have also been analyzed. The results given in Table 4 show the robustness of the model, since the R^2 (0.99) and the RMSE (0.21) are nearly equal for each of these eight subsets, and also equal to the values obtained for the global set. The results obtained for the specific subsets in each solvent, show that the values of R^2 are slightly smaller (0.94 – 0.98) and that of RMSE are similar (0.23 – 0.16) to those of the training and the prediction sets, thus confirming the goodness of the model and its capacity to predict the pK_a for a wide set of solvents. A detailed analysis of the results given in Table 4 shows that for the protic solvents, mainly methanol and water, although the regression coefficient is slightly worse ($\approx$0.94), the standard deviation is the same as for the larger sets. We also evaluated the subsets formed by the pK_a values of specific benzoic acids in several solvents. Table 5 shows the results for the 38 benzoic acids whose pK_a are known in at least six different solvents, and the set includes *ortho*- and non-*ortho*-derivatives. The statistical reports are excellent again; the average for R^2 and the standard deviation are 0.999 and 0.14, respectively. The global analysis of these results, collected in Tables 3 – 5, proves that the proposed model is very robust and has a high capacity of prediction, and that the descriptors contained are able to discriminate between the subtle reasons that produce different pK_a values depending on the position, for example, *ortho* or *para*, of the substituents, since no significant difference in the accuracy of the predicted values of these groups of compounds is observed.

In a previous paper [14], we studied a set of 276 pK_a values corresponding to 94 phenols with a wide variety of substituents in the ring, in ten solvents; a seven-descriptor model was derived which gives a correlation factor, R^2 = 0.98, and an RMSE of 0.8. The similarity between the descriptors contained in that model and the one derived here for the benzoic acids is noteworthy. In both cases, the models contain two descriptors of the solvent and five of the solute. Of the two solvent descriptors, the hydrogen-bond donation ability, α, is common in both cases. In the phenols set dipole moment, μ, appears instead of the $\Delta_{vap}U^{\circ}$,

Table 4. Statistics of pK_a estimation for different subsets.

Subset	n	R^2	RMSE
Training	379	0.998	0.21
Prediction	98	0.998	0.21
Cross-validation	42	0.998	0.21
Protic	253	0.994	0.22
Aprotic	266	0.997	0.21
Ortho	238	0.998	0.21
Non-*ortho*	281	0.998	0.22
Protic and *ortho*	130	0.995	0.22
Aprotic and *ortho*	108	0.998	0.19
Protic and non-*ortho*	123	0.994	0.21
Aprotic and non-*ortho*	158	0.997	0.22
Water	105	0.937	0.22
MeOH	107	0.939	0.20
EtOH	23	0.963	0.19
Isopropanol	18	0.967	0.23
DMSO	84	0.966	0.23
AN	63	0.963	0.21
Acetone	55	0.980	0.16
DMF	58	0.962	0.23
DMA	6	0.925	0.17

but these two descriptors give details about non-specific solute/solvent interactions. Moreover, these two solvent descriptors are the most important in both sets. Of the five solute descriptors, there is one electrostatic, one CPSA, and three quantum-chemical, in each model. The electrostatic descriptor, the maximum partial charge for a hydrogen atom, is common to both sets; in the phenol set, the CPSA descriptor is the relative positive charged surface area, RPCS. The model contains the quantum chemical descriptor, the maximum e – n attraction energy for a C–O bond, which analogously to the minimum resonance energy for the O–H bond in the benzoic acids set, is related to the energy of the functional group that dissociates in each case. The other two quantum descriptors are the polarizability and the LUMO + 1 energy of the phenols. Given the type of solute/solvent interactions involved, in both cases it is possible to recognize non-specific interactions, of polar type, and specific interactions of the hydrogen-bond type. In the phenols set, the LUMO + 1 energy of the solutes is clearly linked to Lewis acid – base interactions. Due to the similarity between the descriptors contained in both models, we made a new complete set (set B) containing the pK_a values of the phenols and the benzoic acids. The model derived here from the benzoic acids was then applied to set B. As indicated above, the ADAPT program does not make it possible to work with more than 500 compounds; thus, we employed neural networks within the SVM methodology. In order to ensure that both types of neural networks, the three-layer, fully connected, feed-forward networks used in ADAPT, and the radial basis functions employed in SVM, gave similar predicted values, we used the SVM programs to calculate the 519 pK_a values of set A. The training procedure of the SVM model is made

Table 5. Statistics of pK_a estimation for specific benzoic acids in several solvents.

Benzoic acid	n[a]	R^2	RMSE
2,6-Dibromobenzoic acid	7	0.999	0.13
2,6-Dibutoxybenzoic acid	8	1.000	0.09
2,6-Dichlorobenzoic acid	8	0.999	0.14
2,6-Dietoxybenzoic acid	7	1.000	0.09
2,6-Difluorobenzoic acid	7	1.000	0.03
2,6-Diiodobenzoic acid	7	0.997	0.25
2,6-Diisopropoxybenzoic acid	7	0.998	0.22
2,6-Dimethoxybenzoic acid	8	0.999	0.16
2,6-Dimethylbenzoic acid	8	0.999	0.14
2,6-Dinitrobenzoic acid	8	0.999	0.17
2,6-Dipropoxybenzoic acid	7	0.999	0.12
2-Acetylbenzoic acid	6	1.000	0.10
2-Chloro-6-nitrobenzoic acid	6	0.999	0.18
2-Chlorobenzoic acid	6	0.999	0.18
2-Methyl-6-chlorobenzoic acid	6	1.000	0.13
2-Methyl-6-methoxybenzoic acid	6	0.998	0.13
2-Methyl-6-nitrobenzoic acid	6	0.999	0.18
2-Nitrobenzoic acid	7	0.999	0.17
3,4-Dichlorobenzoic acid	7	0.999	0.16
3,4-Dimethoxybenzoic acid	6	1.000	0.14
3,4-Dimethylbenzoic acid	7	0.999	0.17
3,4-Dinitrobenzoic acid	6	0.999	0.17
3,5-Dimethoxybenzoic acid	6	1.000	0.14
3,5-Dimethylbenzoic acid	6	1.000	0.08
3,5-Dinitrobenzoic acid	7	0.998	0.26
3-Dromo-4-methylbenzoic acid	6	1.000	0.12
3-Dromobenzoic acid	7	1.000	0.11
3-Chloro-5-nitrobenzoic acid	6	1.000	0.13
3-Methoxy-4-methylbenzoic acid	6	1.000	0.09
3-Methyl-4-chlorobenzoic acid	6	1.000	0.06
3-Methyl-4-methoxybenzoic acid	6	1.000	0.08
3-Methyl-4-nitrobenzoic acid	6	1.000	0.14
3-Nitro-4chlorobenzoic acid	7	1.000	0.12
3-Nitro-4-methoxyenzoic acid	6	0.999	0.21
3-Nitro-4-methylbenzoic acid	6	1.000	0.11
4-Cyanobenzoic acid	6	0.999	0.17
4-Nitrobenzoic acid	8	0.998	0.21
Benzoic acid	9	1.000	0.06

[a] Number of solvents.

by the optimization of the parameters of the kernel function selected. Since we used radial basis functions as kernel, the parameters that should be optimized are the width of the Gaussian functions γ, the regularization constant C and the ε of ε-insensitive loss function. The optimization of these three parameters is done by means of recursive leave-one-out minimization of training set cross-validation error. The starting values for γ, C, and ε are 0.1, 0.1, and 100, respectively; and the optimized final values for these parameters are 0.9, 0.09, and 100. The established model contains 308 support vectors, extracted from the training set compounds. The obtained values show clearly that the results are equivalent, since the correlation coefficient ($R^2 = 0.998$) and RMSE (0.21), in the fit of the experimental *versus* SVM derived values, are the same as those obtained with the ADAPT program. This demonstrates that

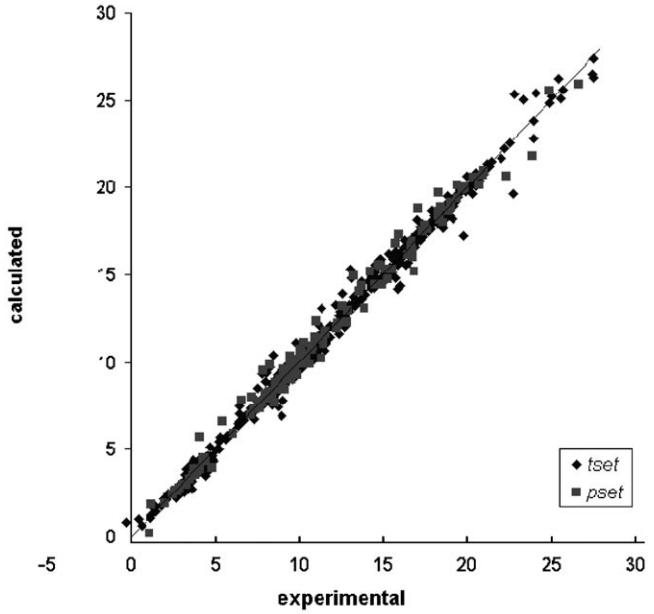

Figure 3. Plot of calculated *vs.* experimental pK_a values for the training and prediction sets of benzoic acids and phenols (set B).

Table 6. Statistics of pK_a estimations for different subsets of benzoic acids and phenols (set B).

Subset	*n*	R^2	RMSE
Training	640	0.993	0.44
Prediction	153	0.988	0.59
Benzoic acids	519	0.997	0.26
Phenols	274	0.982	0.72
Protic	407	0.988	0.40
Aprotic	386	0.988	0.54
Ortho	377	0.992	0.46
Non-*ortho*	418	0.992	0.47
Water	188	0.976	0.45
MeOH	146	0.982	0.32
Isopropanol	34	0.975	0.41
EtOH	23	0.970	0.16
tert-BuOH	16	0.962	0.72
DMSO	127	0.971	0.50
AN	88	0.960	0.53
DMF	87	0.966	0.48
Acetone	65	0.922	0.54
DMA	10	0.990	0.25
NM	9	0.968	0.91

if the ADAPT program could be used for the complete set B we would obtain equivalent values of pK_a.

Set B contains 793 pK_a values corresponding to 228 compounds: 136 benzoic acids and 92 phenols. Concerning the solvents, the pK_a values of the phenols were determined in the ten following solvents: water, methanol, isopropanol, *tert*-butanol, DMSO, DMF, AN, acetone, and NM. That is, of the total of eleven solvents used, eight are common to the two families of compounds (water, methanol, isopropanol, DMSO, DMF, AN, acetone, and DMA). However, three of these solvents only have one type of compound: ethanol includes only benzoic acids and *tert*-butanol and nitromethane contain only phenols. The optimization of the three parameters, in the training of the SVM model, is also done by means of recursive leave-one-out minimization of training set cross-validation error. The starting values for γ, C, and ε are 0.1, 0.1, and 100, respectively; and the optimized final values for them are 1.95, 0.06, and 105. The model built with these optimized values for the radial basis functions parameters has 562 support vectors, selected from the 640 compounds of the training set. For the training set, the correlation coefficient is 0.993 and the RMSE is 0.44, and a good statistical result was also obtained for the prediction set of 153 compounds: $R^2 = 0.988$, RMSE = 0.58. Figure 3 gives the plot of calculated *versus* experimental pK_a values for these two sets. Table 6 gives the distribution of these values in the usual subsets: protic, aprotic, *ortho*-non-*ortho*-substituted, and in each solvent, and also for the benzoic acids and the phenols separately. Although the values of R^2 and the RMSE for the small sets of the specific solvents are slightly worse, the figures are good enough to confirm the validity and goodness of the QSPR approach for this type of studies.

4 Conclusions

The present study confirms the high capacity of the QSPR approach to predict pK_a values of benzoic acids in different solvents. The set of experimental values is large, the benzoic acids studied have very different chemical characteristics and the number and types of the solvents is also significant. All these facts are reflected in the large range of the analyzed experimental values, which is larger than 20 units of pK_a. The statistical results are good, with high correlation coefficients and low RMSE. Moreover, the R^2 and the RMSE are equal for the training, prediction, and cross-validation sets. In the fit of a large set of experimental values, with several reasonable descriptors, often it is possible to derive a model that gives satisfactory R^2 and RMSE values. But if the set of experimental values is rather heterogeneous, for instance, if the compounds belong to different types of chemical substances, or the studied property has been taken in different conditions (solvent, concentration, temperature, *etc.*), then the statistical results from small subsets of values, cannot be as good as the complete set. In these cases, it is convenient to analyze the predicted values for small subsets corresponding to families of compounds or values obtained in similar experimental conditions. In the present case, the statistical parameters obtained for several small subsets, protic or aprotic solvents, *ortho* or non-*ortho* substituted acids, and even the smaller subsets of each solvent are all very similar between them as well as to those achieved with the training and

prediction sets. The validity of the model is also reinforced by the good statistics parameters found with the set of values of one specific benzoic acid in several solvents. The comparison of the results presented here with those previously reported clearly shows the higher predictive capacity of our model. Thus, QSPR methods applied to sets of 30 and 90 benzoic acids in aqueous solutions [6, 7] with quantum descriptors, give correlation coefficients of 0.8, while the RMSE value is 1. The comparison with the results derived from the application of high-level theoretical calculations through continuum methods is not straightforward, however, since the sets studied are rather heterogeneous and contain diverse kinds of organic compounds. On the other hand, the reported studies in water, DMSO, and AN [2–4] offer precisions of 1–2 pK_a units. So, in general the predicted values by these top-level methods are less accurate than those derived by the QSPR approach presented here. In addition, the capacity of the model, which is CNN derived, should be noted to relate the descriptors with the complex dissociation process. Another gratifying aspect to emphasize is the similarity between the descriptors capable of modeling the pK_a values of both benzoic acids and phenols. Consequently, the application of the model derived from the benzoic acid values to the complete set B, which also contains the values for the phenols, gives excellent results, as can be seen from the statistics of the corresponding training and prediction sets and of the other small subsets analyzed. This set contains pK_a values of an important number of benzoic acids and phenols, which hold a wide variety of substituents in all the positions of the phenyl ring, in 11 solvents. The good results obtained show that the model can be used in the pK_a prediction of benzoic acids or phenols with the same type of substituents and in the solvents indicated.

To sum up, the QSPR approach is a satisfactory methodology to analyze quite complex multicomponent systems, such as the pK_a values of neutral aromatic acids in water and in several organic solvents. The precision of the values obtained can be favorably compared with other approaches, including high-level quantum calculations. On the other hand, the model derived highlights two substantial aspects of QSPR methods: their capacity to predict properties and their capacity to correlate these properties with the molecular structure of the compounds.

Acknowledgements

The authors thank Professor Peter C. Jurs for giving us access to the ADAPT program. We are grateful to the Catalan Government for financial support (Grant 2005 SGR 00184) and J. J. to the University of Barcelona for a BRD grant.

References

[1] H. Wang, J. Ulander, *Expert Opin. Drug Metab. Toxicol.* **2006**, *2*, 139–155.
[2] Y. Fu, L. Liu, R.-Q. Li, R. Liu, Q.-X. Guo, *J. Am. Chem. Soc.* **2004**, *126*, 814–822.
[3] D. M. Chipman, *J. Phys. Chem. A* **2002**, *106*, 7413–7422.
[4] G. I. Almerindo, D. W. Tondo, J. R., Pliego, *J. Phys. Chem. A* **2004**, *108*, 166–171.
[5] J. Zhang, T. Kleinöder, J. Gasteiger, *J. Chem. Inf. Model.* **2006**, *46*, 2256–2266.
[6] M. J. Citra, *Chemosphere* **1999**, *38*, 191–206.
[7] B. G. Tehan, E. J. Lloyd, M. G. Wong, W. R. Pitt, E. Gancia, J. G. Montana, D. T. Manallack, *Quant. Struc. Act. Relat.* **2002**, *21*, 473–485.
[8] E. Bosch, F. Rived, M. Rosés, J. Sales, *J. Chem. Soc., Perkin Trans. 2* **1999**, 1953–1958.
[9] J. Picha, R. Cibulka, F. Liska, P. Parik, O. Pytela, *Collect. Czech. Chem. Commun.* **2004**, *69*, 2239–2252, and references therein.
[10] A. J. Chalk, B. Beck, T. Clark, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1053–1059.
[11] T. Suzuki, R.-U. Ebert, G. Schüürmann, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 776–790.
[12] N. M. Halberstam, I. I. Baskin, V. A Palyulin, N. S. Zefirov, *Doklady Chem.* **2002**, *384*, 140–143.
[13] N. M. Halberstam, I. I. Baskin, V. A. Palyulin, N. S., Zefirov, *Mendeleev Commun.* **2002**, 185–186.
[14] J. Jover, R. Bosque, J. Sales, *QSAR Comb. Sci.* **2007**, *26*, 385–397.
[15] O. Pytela, J. Kulhánek, *Collect. Czech. Chem. Commun.* **2002**, *67*, 596–608.
[16] O. Pytela, J. Kulhánek, M. Ludwig, V. Ríha, *Collect. Czech. Chem. Commun.* **1994**, *59*, 627–638.
[17] O. Pytela, J. Kulhánek, M. Ludwig, *Collect. Czech. Chem. Commun.* **1994**, *59*, 1637–1644.
[18] O. Pytela, J. Kulhánek, *Collect. Czech. Chem. Commun.* **1995**, *60*, 829–840.
[19] O. Pytela, O. Prusek, *Collect. Czech. Chem. Commun.* **1999**, *64*, 1617–1628.
[20] O. Pytela, J. Kulhánek, *Collect. Czech. Chem. Commun.* **1997**, *62*, 913–923.
[21] F. Rived, M. Rosés, E. Bosch, *Anal. Chim. Acta* **1998**, *374*, 309–324.
[22] F. Rived, I. Canals, E. Bosch, M. Rosés, *Anal. Chim. Acta* **2001**, *439*, 315–333.
[23] K. Izutsu, *Acid-Base Dissociation Constants in Dipolar Aprotic Solvents*, Blackwell Scientific Publications, Oxford **1990**.
[24] M. K. Chantooni, I. M. Kolthoff, *Anal. Chem.* **1979**, *51*, 133–140.
[25] I. M. Kolthoff, M. K. Chantooni, *J. Am. Chem. Soc.* **1971**, *93*, 3843–3849.
[26] J. Kulhánek, O. Exner, *J. Chem. Soc. Perkin Trans. 2* **1998**, 1397–1401.
[27] A. R. Katritzky, V. S. Lovanov, M. Karelson, *CODESSA, Reference Manual V 2.13*, Semichem and the University of Florida **1997**.
[28] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. P. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[29] J. P. P. Stewart, *MOPAC 6.0 Quantum Chemistry Program Exchange*, QCPE, No 455, Indiana University, Bloomington, IN **1989**.

[30] R. W. Taft, M. J. Kamlet, *J. Am. Chem. Soc.* **1976**, *98*, 2886 – 2894.
[31] K. Dimroth, C. Reichardt, T. Siepmann, F. Bohlmann, *Liebigs Ann. Chem.* **1963**, *661*, 1 – 37.
[32] I. A. Koppel, V. A. Palm, in: N. B. Chapmann, J. Shorter (Eds.), *The Influence of Solvent on Organic Reactivity in Advances in Linear Free Energy Relationships*, Plenum Press, London **1972**, Chapter 5.
[33] V. Gutmann, E. Wychera, *Inorg. Nucl. Chem. Lett.* **1966**, *2*, 257 – 260.
[34] E. M. Kosower, *J. Am. Chem. Soc.* **1958**, *80*, 3253 – 3260.
[35] J. Catalán, V. López, P. Pérez, R. Martín-Villamil, J. G. Rodríguez, *Liebigs Ann.* **1995**, 241 – 252.
[36] R. S. Drago, *Applications of Electrostatic-Covalent Models in Chemistry*, Surfside, Gainsville **1994**.
[37] P. C. Jurs, J. T. Chow, M. Yuan, in: E. C. Olson, R. E. Christorffersen (Eds.), *Computer-Assisted Drug Design*, The American Chemical Society, Washington, DC **1979**, pp. 103 – 129.
[38] A. J. Stuper, W. E. Brugger, P. C. Jurs, *Computer-Assisted Studies of Chemical Structure and Biological Function*, Wiley, New York **1979**.
[39] B. T. Luke, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1279 – 1287.
[40] J. M. Sutter, S. L. Dixon, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 77 – 84.
[41] L. Xu, J. W. Ball, S. L. Dixon, P. C. Jurs, *Environ. Toxicol. Chem.* **1994**, *13*, 841 – 851.
[42] M. D. Wessel, P. C. Jurs, *Anal. Chem.* **1994**, *66*, 2480 – 2487.
[43] D. J. Livingstone, D. T. Manallack, *J. Med. Chem.* **1993**, *36*, 1295 – 1297.
[44] V. N. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, New York **1998**.
[45] N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK **2000**.
[46] B. Schölkopf, A. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA **2002**.
[47] C.-C. Chang, C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, **2001.** Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[48] H. X. Liu, R. S. Zhang, X. J. Yao, M. C. Liu, Z. D. Hu, B. T. Fan, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1288 – 1296.
[49] M. J. L. Orr, *Introduction to Radial Basis Function Networks*, Centre for Cognitive Science, Edinburgh University **1996**.
[50] M. J. L. Orr, *Matlab Routines for Subset Selection and Ridge Regression in Linear Neural Networks*, Centre for Cognitive Science, Edinburgh University **1996**.
[51] N. S. Zefirov, M. A. Kirpichenok, F. F. Izmailov, M. I. Trofimov, *Dokl. Akad. Nauk SSSR* **1987**, *296*, 883 – 887.
[52] D. T. Stanton, P. C. Jurs, *Anal. Chem.* **1990**, *62*, 2323 – 2329.
[53] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York **2000**.
[54] D. R. Armstrong, P. G. Perkins, J. J. P. Stewart, *J. Chem. Soc. Dalton Trans.* **1973**, 2273.
[55] J.-L. M. Abboud, R. Notario, *Pure Appl. Chem.* **1999**, *71*, 645 – 718.
[56] C. Reichardt, *Solvent and Solvent Effects in Organic Chemistry*, 3rd edn., VCH, Weinheim **2003**.
[57] R. Guha, P. C. Jurs, *J. Chem. Inf. Comput. Model.* **2005**, *45*, 800 – 806.