



Resources for Acid-Base Properties and Ionisation in the Context of the Regulation of Chemicals

Authors: Louise McLaughlin, Peter Fisk

Date: September 2007

Peter Fisk Associates Limited, ♦ Company Number 5758319
14 Darenth Close ♦ Herne Bay ♦ Kent ♦ UK ♦ CT6 7EX
Tel/fax: +44 (0)208 123 6265 ♦ Mobile: +44 (0)7795 517615
Web Site: <http://www.ecotoxchem.co.uk>

Acknowledgements

This report is based on an earlier web-based resource produced by the author in collaboration with the Laboratory of the Government Chemist, and funded by the UK Analytical Partnership.

Thanks to John Francis of LGC for his contributions to the original work and updates until November 2004.

Date of last literature search: August 2006

Contents:

Acknowledgements	i
Contents:.....	ii
Resources for Acid-Base Properties and Ionisation in the Context of the Regulation of Chemicals	3
1. Introduction	3
2. Definitions	4
2.1 Acids and Bases	4
2.2 Zwitterions	5
2.3 Salts	5
3. Recognition of whether ionisation could occur.....	6
4. pK _a measurement techniques	7
4.1 Acid-base Titration.....	7
4.2 UV-Spectrophotometric method	7
4.3 Conductometric method	7
4.4 Solubility and Partition.....	8
4.5 Capillary Electrophoresis	8
4.6 HPLC.....	8
5. Literature sources of pK _a data	9
6. Methods for prediction of pK _a	10
6.1 pKcalc	10
6.2 ACDLabs.....	11
6.3 QikProp	11
6.4 SPARC	11
6.5 ADMET.....	11
6.6 ADME Boxes	12
6.7 CSpKa	12
6.8 Pipeline Pilot™	12
6.9 ALOGPS 2.1	12
6.10 DISCON.....	12
6.11 Marvin	12
7. New research in pK _a prediction.....	13
8. pK _a in the context of REACH	14
References	15
Websites	16

Tables

Table 1 pKa values for acetic acid, phenol and derivatives	6
Table 2 Published sources of pK _a data	9

Resources for Acid-Base Properties and Ionisation in the Context of the Regulation of Chemicals

1. Introduction

The degree of ionisation of a substance can influence its environmental fate, due to its impact on other physicochemical parameters such as partition coefficients and solubility. Ionisation can also affect *in vivo* toxicological properties such as absorption through membranes. Therefore, from a regulatory point of view, the dissociation constant (pK_a) is an important property for substances containing [ionisable functional groups](#). Dissociation constant is a required REACH endpoint for substances produced or imported at 100 tonnes per annum and above, and the technical guidance (RIP 3.3) recommends that a predicted value should be obtained for ionisable substances above 10 tonnes per annum. It is also a SIDS (Screening Information Dataset) endpoint for high production volume chemicals.

In addition to the importance of ionisation in predicting environmental fate, and hence preparing accurate risk assessments, knowledge of pK_a is essential for ensuring that suitable test conditions are employed when measuring partition coefficients and for developing suitable analytical techniques. According to the EU Annex V test guidelines, octanol-water partition coefficient (K_{ow}) and adsorption/desorption coefficient (K_{oc}) should be determined for the neutral species, therefore these tests should be conducted at a pH at which the substance is in its unionised form. When using HPLC methods for determining these properties, it is important to eliminate the possibility of ionic interactions between the substance and the stationary phase so again, the mobile phase pH should be selected to ensure that the substance is unionised.

In terms of analytical method development, for the use of certain techniques, such as HPLC, capillary electrophoresis and other electrophoretic techniques the ionisation state of the analyte is important in identifying suitable conditions for the analysis. It is not always practicable to obtain a [measured \$pK_a\$ value](#) *in situ*, therefore [literature](#) or [predicted](#) data could prove invaluable.

Scientists from Advanced Chemistry Development presented a poster at the HPLC 2004 conference in Philadelphia in June, and again at CPSA in October of that year, "Application of Structure-based pK_a Prediction in Reversed-Phase Chromatographic Method Development". The poster contained an overview of how ionisation of an analyte affects chromatographic retention time peak shape, thus impacting upon system suitability requirements, resolution of components and detection limits. The full poster can be viewed at: http://www.acdlabs.com/download/publ/2004/cpsa04_pka.pdf

2. Definitions

2.1 Acids and Bases

OECD Test Guideline 112 (Dissociation Constants in Water; 1981) sets out the necessary definitions of the dissociation constant of an acid RH,



and

$$K_a = \frac{[R^-][H^+]}{[RH]}$$

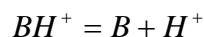
These values are usually expressed as

$$pK_a = pH - \log \frac{[R^-]}{[RH]}$$

where p is $-\log_{10}$

Strong acids have low pK_a values i.e. < 3 , weak acids have values in the range 3 to 10, and values above 10 indicate essentially a non-acidic substance in the aqueous environment.

Base strength is expressed as the acidity of the 'conjugate acid'. For the base B and the conjugate acid BH^+



and

$$pK_a = pH - \log \frac{[B]}{[BH^+]}$$

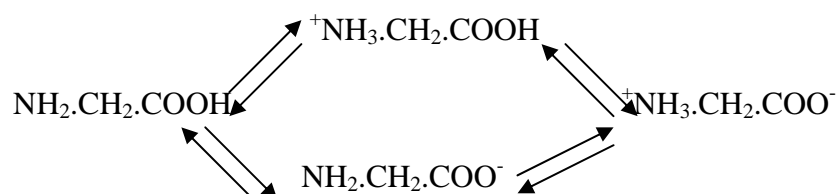
Therefore strong bases have high pK_a and weak bases have low pK_a . Note that the term 'strong' should not be used to imply 'of high concentration', which is a common usage in older literature.

The term K_b has long been superseded and should not be used under any circumstances, even if apparently reliable sources use it. K_b was superseded so that acidity and basicity would be expressed on a single scale.

A substance can have more than one acidic function, and the expressions for the K_a values are as above for each step. The same applies to bases.

2.2 Zwitterions

Substances may possess both an acid and a basic group. For example, consider glycine, an amino-acid. In aqueous solution we may envisage the following processes



The 'zwitterion' form $^+\text{NH}_3\text{CH}_2\text{COO}^-$ is the usual form in aqueous solution near neutral pH, because the NH_2 is a moderately strong base ($\text{pK}_a = 9.8$) and the COOH is a moderately strong acid ($\text{pK}_a = 2.4$). Therefore the proportion of $\text{NH}_2\text{CH}_2\text{COOH}$ in a solution of glycine (aqueous) is very low. At a point approximately mid-way between the two pK_a values the amount of zwitterion is maximised, which is termed the iso-electric point, since the net charge on the molecule is zero for the highest proportion. In an organic solvent the dissolved form may be non-ionised rather than zwitterionic.

It is important to note that a substance containing a weakly acidic and a weakly basic group will be largely non-ionised at near neutral pH and thus is not zwitterionic, nor of particular concern in the present discussion. For example, 4-aminophenol, which has an acid pK_a of around 8 and a basic pK_a around 3.

2.3 Salts

Salts are reaction products of acids and bases that retain ionic character; they may be prepared from any combination of acid and base. Their solubility depends upon the properties of the crystalline form, but when dissolved in water (at moderate concentration) the ionisation state depends upon pH and the pK_a of each half.

3. Recognition of whether ionisation could occur

This is not as straightforward as might first be thought, and is part of the reason why *a priori* estimation of pK_a is so difficult. The focus of interest here is on pH conditions of relevance to mammalian toxicology or the environment, and corresponding analytical method development. Consequently weak acidity or basicity that can be exploited in chemical synthesis, under non-aqueous conditions, is ignored.

In deciding whether ionisation could occur it is important to recognise firstly that the presence of well-known functional groups is the most important indicator.

Secondly it is probable that scientists working with a substance may have picked up clues about its nature, and be aware of possible acid-base properties. This could come from reactivity properties, physicochemical properties, or spectroscopic data.

To recognise acid-base properties some chemical knowledge is required, looking for the presence of, for example, strong electron-withdrawing or donating groups (see Table 1 for examples of how this applies in two chemical series – acetic acid derivatives and phenol derivatives). However, it is possible to state that the absence of indicative chemical functionality is a sign that no acid-base properties can be expected. Should it still be necessary to seek further information, then the simplest way to get some insight is to qualitatively compare the [solubility](#) of the substance in solutions buffered at pH values around 2, 7 and 12.

Table 1 pK_a values for acetic acid, phenol and derivatives¹

Substance	pK_a
Acetic acid	4.76
Chloroacetic acid	2.85
Trichloroacetic acid	0.70
Phenol	9.89
4-Chlorophenol	9.18
2,4-Dinitrophenol	3.96

¹ Data from <http://www.zirchrom.com/organic.htm>

4. pK_a measurement techniques

The most detailed source on this subject is Albert (1984). The basic principle of pK_a measurement is very simple: it is necessary to be able to obtain information about the relative concentrations of the various ionised forms that may be present in aqueous solution. The dissociation constant will vary somewhat with the composition of the aqueous medium used, and so it is usual to try to use the most non-invasive conditions possible. This point will be returned to below.

OECD Test Guideline 112 describes titration, spectrophotometric and conductometric methods.

4.1 Acid-base Titration

The titration of an acid by a strong base is a very familiar chemistry concept that is described in standard textbooks. However, it is less useful than might first appear, because it is only really useful when:

- Sufficient sample of high purity material is available, which might not be the case for research samples, since at least 100mg will be needed.
- The sample is soluble in water
- The titration curve is clear enough to be interpreted.

The elegant simplicity of looking at an acid-base titration curve does however make this technique appealing.

4.2 UV-Spectrophotometric method

This method is applicable when a substance possesses a UV-visible spectrum. It depends on comparing the spectra of the substance obtained over a range of pH values, and requires that the ionised and unionised forms have significantly different absorbances at the chosen wavelength. Because spectral response is a function of concentration, the relative proportions of the ionised forms can be determined. It is a very useful method when it can be used, because it is quick and needs very little substance. It can also be used for substances of low water solubility.

In other cases, other spectrophotometric techniques can be used, but are not widely applicable.

4.3 Conductometric method

The conductometric method is applied by measuring the conductivity of a 0.1 M solution of the test substance in distilled water, followed by measurements on a series of sequentially diluted solutions.

4.4 Solubility and Partition

Ionised forms of substances are usually much more soluble in water than non-ionised forms. Therefore, if studies of the solubility of a substance in a range of pH-buffered aqueous solutions are undertaken, interpretation of the results can yield accurate pK_a values if solubilities are quantified.

On a similar theme, study of the partition of a substance in a two-phase solvent-water system can provide pK_a values. This method is demanding in terms of analytical chemistry, but requires relatively little substance, and is particularly useful when sample is in short supply.

4.5 Capillary Electrophoresis

Recently published articles (Wan *et al.*, 2003 and Foulon *et al.*, 2004) describe methods for pK_a determination using capillary electrophoresis (CE). This technique has several advantages over conventional methods including increased sensitivity and hence requiring only a small sample size; greater selectivity; suitability for non-UV absorbing substances. Wan *et al.* report that the method can also be applied for substances with poor aqueous solubility, by using a co-solvent such as DMSO. CE also lends itself to automation, making rapid screening of compounds possible.

4.6 HPLC

Markuszewski *et al.* (2004) recently described a method for determining pK_a using a reverse phase HPLC method, by maintaining a constant proportion of organic modifier and using a pH gradient in the aqueous phase. pK_a values derived using this new technique are reported to correlate well with literature values determined using conventional titration methods. An advantage of this technique is that it permits simultaneous determinations for mixtures of analytes such as those produced by combinatorial synthesis.

5. Literature sources of pK_a data

Measured pK_a data are available from many published sources. Some commonly cited sources are listed in Table 2.

Table 2 Published sources of pK_a data

Author(s)	Title	Reference	ISBN
Martell, Smith	Critical Stability Constants (Volumes 1 to 5)		0306352117 (Vol 1) 0306352125 (Vol 2) 0306352133 (Vol 3) 0306352141 (Vol 4) 0306410052 (Vol 5)
Serjeant, Dempsey	Ionization constants of organic acids in aqueous solution	IUPAC Chemical Data Series No 23, Pergamon, 1979 ²	0080223397
Perrin	Ionization constants of organic bases in aqueous solution	IUPAC Chemical Data Series, Pergamon, 1965 and supplement 1972	
Albert	Ionization Constants of Heterocyclic Substances	Physical Methods in Heterocyclic Chemistry, Katritzky, A. R., Ed., Academic Press, New York, 1963	
Lide	Handbook of Chemistry and Physics, 84th Edition	CRC Press 2003	0849304849
O'Neil	The Merck Index: An Encyclopedia of Chemicals, Drugs and Biologicals	14 th Edition, October 2006	978-0-911910-00-1

In addition to the sources presented in Table 2, several on-line resources are available, including:

PhysProp Database, Syracuse Research Corporation. The database contains experimental and calculated data for over 18,000 substances and is free to search on-line by substructure (www.syrres.com/esc/physdemo.htm). The database can also be purchased as a whole from SRC.

An on-line database of pK_a values for over 600 compounds can also be found at www.zirchrom.com/organic.htm

The Merck Index can also be purchased on CD-ROM or as an on-line Internet edition (<http://www.merckbooks.com/mindex/>).

² This is a supplement to an earlier volume edited by Kortum, Dissociation constants of organic acids in aqueous solution, 1961.

6. Methods for prediction of pK_a

For relatively simple structures, methods for predicting pK_a values are well documented, and fully described by Perrin (1981). The main technique relies on correlation of 'sigma constants' of functional groups with pK_a .

For novel structures, some calibration with measurements will be needed. Certain commercial software packages use this method.

It might have been expected that modern computer-based molecular modelling techniques (e.g. molecular orbital calculations) could predict pK_a values. In fact, it is a difficult objective, for two reasons:

- It requires consideration of the presence of many molecules of solvent (water) not just the structure of interest
- The pK_a represents the energy difference of ionised and non-ionised forms, and as such requires very accurate calculations.

However, where laboratories have them, molecular orbital calculations can give insights as to whether ionisation is possible.

There are several software packages available that include modules for prediction of pK_a . Some of these are outlined below, although a comprehensive comparison of performance has not been carried out.

6.1 pKalc

The [PALLAS 3.4](#) suite of programs is available to purchase from CompuDrug and includes estimation of pK_a using pKalc, the latest version of which was launched in February 2006. The program database contains measured data on over 10,000 compounds. Prediction equations are based on Hammett and Taft equations. In addition, a pK_a wizard is built into the program to allow development of new equations for pK_a . For most substances, errors in the prediction are reported as within 0.25 pK_a units.

Limitations of the program include difficulties in predicting multiple pK_a values for small molecules (e.g. citric acid); inability to distinguish between cis and trans isomers and absence of the relevant information on ionising groups or substituent from the program database.

A free software demo can be downloaded from the website. The Pallas Net package includes the pKalc module in a web-based software package, which can be run using intranet or internet networks.

6.2 ACDLabs

ACD/pK_a DB (version 9.0) from [Advanced Chemistry Development, Inc.](#) contains a searchable database of over 16,000 compounds. The program allows prediction of single or multiple pK_a values for any combination of dissociated and non-dissociated centres in a given molecule. Other features include the ability to create a user database of measured pK_a values, which can also be used to train the prediction equations and hence improve the accuracy of prediction for a substance-type. The software can be purchased from ACD Labs.

A new algorithm implemented in version 9.0 allows more accurate predictions for molecules with 3 or more ionisable centres.

Prediction of analyte pK_a is incorporated into the ACD/LC and ACD/Method Development programs, allowing optimisation of chromatographic conditions.

Further information on the latest version, and previous versions, of the software is available on the website.

6.3 QikProp

QikProp is a commercial software package from [Schrödinger](#) which is designed for the prediction of ADME (Absorption, Distribution, Metabolism and Excretion) properties of pharmaceuticals. Various physicochemical properties, including pK_a, are predicted using algorithms based on 3-D molecular structure. A trial version of the software can be requested.

6.4 SPARC

[SPARC](#) is an on-line property prediction service operated by the University of Georgia. pK_a can be predicted for single or multiple ionisations and a searchable database of measured data is available. The service is free to use.

Also available on the website is a list of published references where the software has been applied.

6.5 ADMET

ADMET is an ADME property prediction package from [SimulationsPlus Inc.](#) It predicts macroscopic pK_a for eight types of ionisable functional group (hydroxyacids, acidic amines, acids based on aromatic –NH, thioacids, carboxylic acids, amines, bases based on aromatic N and N-oxides). The method uses an artificial neural network for each substance type, taking into account a database of over 8500 compounds and a theory of microionisation equilibria.

6.6 ADME Boxes

ADME Boxes version 3.5 is another ADME property prediction software package from [Pharma Algorithms](#). A training set of over 7000 substances is biased towards pharmacologically relevant compounds. A trial version of the software can be downloaded via the website.

6.7 CSpKa

The CSpKa software from [ChemSilico](#) calculates macroscopic dissociation constants using a neural network approach for four chemical classes: N-heteroaromatics; acids (including phosphonic acids); alcohols and amines). A trial version can be requested.

6.8 Pipeline Pilot™

The [Chemistry Collection](#) module of Pipeline Pilot™ from Scitegic is a suite of calculation programs for physicochemical properties, including pK_a. It uses a proprietary method for calculating structural fingerprints (Extended Connectivity Fingerprints).

6.9 ALOGPS 2.1

[ALOGPS 2.1](#) is a free on-line property predictor from the Virtual Computational Chemistry Laboratory (VCCLAB). The user can create their own compound library for use with the program.

6.10 DISCON

The DISCON program is available as part of the MolPro program package developed by the Institute of Biomedical Chemistry, Russian Academy of Medical Sciences. It operates using a database of over 16,000 pK_a values. Prediction is based on a search for a substance's nearest neighbour in the database and can show the distribution of all molecular species for an ionisable compound at all pHs. The software can be purchased from [Tim Tec](#).

6.11 Marvin

Chemaxon provide [Marvin](#), a Java-based chemical software application that can be used to build intranet or internet sites or in software development. As well as structure-drawing capabilities, plug-ins are available for estimation of a number of chemical properties including pK_a. The pK_a estimation methods were updated in March 2004. Marvin applications are freely available for use on internet sites or scientific publications available on the internet.

Results of tests on measured versus predicted values (implemented in version 3.5 of the software) are available on the website.

7. New research in pK_a prediction

There is currently a great deal of interest in developing new and improved models for pK_a prediction, with various research groups around the world publishing new approaches using molecular modelling and statistical methods. These seem to offer hope that prediction based on structure alone, to +/- 0.5 pK_a units, might be possible. Two very distinct methods are those of Tehan *et al.* (2002) and Xing *et al.* (2003). The former approach uses regression analysis to search for relationships between pK_a and calculated quantum mechanical properties of certain classes of substance. The result is a set of equations based on electrophilic superdelocalisability which are applicable to various subsets within chemical classes, for example ortho, meta and para substituted phenols.

The approach of Xing *et al.* (2003), on the other hand, is to develop an accurate method which does not require knowledge of the 3-D structure of a molecule and does not employ quantum mechanics or molecular dynamics in the calculations. Rather than the more conventional “fragment” approach used for many QSARs, the model uses an atom-by-atom approach, with only a few key functional groups included. A “connection tree” is constructed, starting from the acidic or basic centre and working outward in order to characterise the ionising centre, and the pK_a is determined from equations based on the number of occurrences of a particular type of atom or group.

Li and Cui (2003) have published a proposal for a new approach to pK_a prediction using a hybrid quantum mechanical/molecular mechanical potential and the free energy perturbation technique. Researchers at Wayne State College, USA (Kounovsky *et al.*, 2003), reported investigations into pK_a predictions using Density Functional Theory computational methods and various solvation methods.

Szegezdi and Csizmadia (2004) reported a prediction method based on micro-constants which is implemented in the pK_a plug-in of the Marvin software. Further information on the method is available at the [Chemaxon](#) website.

Several papers have also been published concerning development of pK_a databanks and web-based access to prediction models. Jallas *et al.* (2001) have created an electronic database containing pK_a values for 20,000 organic and inorganic substances, compiled from scientific journals from 1890 to 1990. Their system is searchable by chemical structure and contains information on experimental conditions such as temperature and pressure, as well as full literature references.

[CompuDrug](#) and [Chemaxon](#) both offer packages which can be used for web-based property prediction on intranet or internet sites.

8. pK_a in the context of REACH

As mentioned in the introduction (Section 1), dissociation constant is a required endpoint under REACH for substances produced or imported at 100 tonnes per annum. A measured value of pK_a is not required under the following circumstances:

- Absence of ionisable functional groups
- Hydrolysis half-life less than 12 hours or the substance is readily oxidisable
- Measurement is not scientifically possible (e.g. lack of a suitably sensitive analytical method).

In such cases, a suitable remark should be included in the Chemical Safety Assessment.

Where a substance contains more than one ionisable functional group, the pK_a value for each dissociated species should be reported.

Before new measurements are made, it is acceptable to take published data on the substance and close structural analogues into account. Read-across, supported by appropriate predictions using one of the methods described in Section 6, can be used.

It is also recommended that literature or predicted values should be obtained for ionisable substances at 10 tonnes per annum to allow proper understanding and interpretation of the physicochemical, environmental fate and toxicological properties of the substance.

Predictive methods are also invaluable for complex or multi-constituent substances where it is not possible to measure pK_a values for individual constituents.

References

- OECD Test Guideline no. 112. "Dissociation Constants in Water" adopted 1981. Freely downloadable from OECD website at <http://titania.sourceoecd.org/vl=1126100/cl=32/nw=1/rpsv/ij/oecdjournals/1607310x/v1n1/s12/p1>
- Albert, A; The Determination of Ionization Constants, 2nd ed 1984, possibly out of print, ISBN 0412103001
- D.D. Perrin, Boyd Dempsey and E.P. Serjeant, pK_a Prediction for Organic Acids and Bases, Chapman and Hall, 1981, ISBN 041222190X
- "High throughput screening of pK_a values of pharmaceuticals by pressure-assisted capillary electrophoresis and mass spectrometry", Wan *et al*, Rapid Commun. Mass Spectrom. 2003; 17: 2639 – 2648
- "Determination of ionisation constants of N-imidazole derivatives, aromatase inhibitors, using capillary electrophoresis and influence of substituents on pK_a shifts", C. Foulon, C. Danel, C. Vaccher, S. Yous, J.-P. Bonte and J. -F. Goossens. Journal of Chromatography, A (2004), 1035(1), 131 – 136
- "High-throughput evaluation of lipophilicity and acidity by new gradient HPLC methods", Markuszewski, MJ, Wiczling, P and Kaliszan, R, Combinatorial Chemistry and High Throughput Screening (2004), 7(4), 281-289.
- "Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 1: Application to Phenols and Carboxylic Acids", Tehan, Lloyd, Wong, Pitt, Montana, Manallack and Gancia, Quant. Struct.-Act. Relat., (2002), 21(5), 457-472 (Manallack is the correspondence author)
- "Estimation of pK_a Using Semiempirical Molecular Orbital Methods. Part 2: Application to amines, anilines and various nitrogen containing heterocyclic compounds", Quant. Struct.-Act.Relat.,(2002), 21(5), 473-485
- "Predicting pK_a by Molecular Tree Structure Fingerprints and PLS", Xing, Glen and Clark, J. Chem. Inf. Comput. Sci. 2003, 43, 870 - 879
- "pK_a Calculations with QM/MM Free Energy Perturbations", Li and Cui, J. Phys. Chem (2003) 107(51), 14521-14528
- "Prediction of pK_as of acids and bases from *ab initio* calculation utilizing various solvation methods", Kounovsky, Hale, Melvin and D'Souza, Conference Abstracts, 38th Midwest Regional Meeting of the American Chemical Society, November 2003
- "Prediction of dissociation constant using microconstants", Szegezdi and Csizmadia, Conference Abstracts, 227th ACS National Meeting, March 2004
- "A Computerized Storage and Retrieval System of pK_a Values of Hydrogen Acids", Jallas, Palm and Tenno, Proc. Estonian Acad. Sci.Chem, 2001, 50, 3,-186-192

Websites

www.zirchrom.com/organic.htm
www.syrres.com/esc/physdemo.htm
www.merckbooks.com/mindex/
www.compudrug.com
www.acdlabs.com
www.schrodinger.com
http://ibmlc2.chem.uga.edu/sparc/
www.simulationsplus.com
www.ap-algorithms.com
www.chemsilico.com
www.scitegic.com
www.vcclab.org
http://software.timtec.net/discon.htm
www.chemaxon.com/marvin/index.html