

The IUPAC aqueous and non-aqueous experimental pKa data repositories of organic acids and bases

Anthony Michael Slater

Received: 16 April 2014 / Accepted: 10 June 2014 / Published online: 22 June 2014
© Springer International Publishing Switzerland 2014

Abstract Accurate and well-curated experimental pKa data of organic acids and bases in both aqueous and non-aqueous media are invaluable in many areas of chemical research, including pharmaceutical, agrochemical, specialty chemical and property prediction research. In pharmaceutical research, pKa data are relevant in ligand design, protein binding, absorption, distribution, metabolism, elimination as well as solubility and dissolution rate. The pKa data compilations of the International Union of Pure and Applied Chemistry, originally in book form, have been carefully converted into computer-readable form, with value being added in the process, in the form of ionisation assignments and tautomer enumeration. These compilations offer a broad range of chemistry in both aqueous and non-aqueous media and the experimental conditions and original reference for all pKa determinations are supplied. The statistics for these compilations are presented and the utility of the computer-readable form of these compilations is examined in comparison to other pKa compilations. Finally, information is provided about how to access these databases.

Keywords IUPAC · pKa · Database · Aqueous pKa · Non-aqueous pKa · Organic acid · Organic base · pKa database · pKa compilation · pKa repository · Experimental pKa

Introduction

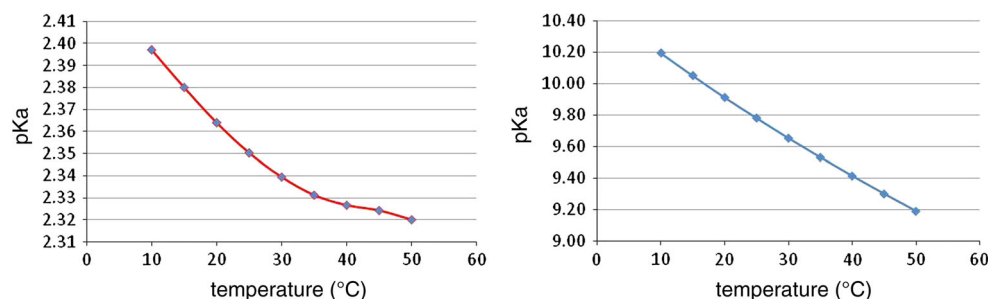
There are many compilations of pKa values of organic acids and bases in both aqueous and non-aqueous solvents in the literature which are of utility to chemists in many fields. Some may be accessed directly via the internet [1–4], while others are in the printed form [5–17]. They mainly cover “drug-like” molecules. These listings of pKa values are useful, but it is important to know the conditions under which the pKa measurements were made and the methods used in their determination. For example, Fig. 1 illustrates how pKa values can vary with temperature, so knowledge of the experimental temperature is of great significance. Ionic strength also influences measured pKa values, so it is vital to know if any corrections have been made to extrapolate to zero ionic strength, producing a thermodynamic pKa value. The importance of an assessment of data quality, which takes these sorts of considerations into account, was recognised many years ago by the International Union of Pure and Applied Chemistry (IUPAC) who sponsored their own pKa compilations in a series of five books, four for aqueous values [18–21] and one for non-aqueous values [22] in twelve organic solvents. It is these books that form the basis of the IUPAC pKa Databases [23]. Prankerd [14] provides an excellent and detailed evaluation of methods used to determine aqueous pKa values, and emphasises the great importance of the IUPAC assessment of data quality which he uses to good effect in his critical compilation.

IUPAC pKa databases

While these various literature compilations are of importance, their utility can be greatly enhanced by providing the

A. M. Slater (✉)
pKaData Limited, 18 Church Lane, Reepham,
Lincoln LN3 4DQ, UK
e-mail: tony@pkadata.com

Fig. 1 Variation of aqueous pKa with temperature for glycine using “reliable” data (vide infra) from the IUPAC Aqueous Database. Left graph (red line) is for the acidic ionisation and the right graph (blue line) is for the basic ionisation



data in computer-readable form, searchable by substructure, similarity etc. The four IUPAC books of aqueous pKa measurements [18–21] and the IUPAC book of non-aqueous pKa measurements [22] have been carefully translated into computer-readable form, adding value in the process, for example:-

- for every molecule, a SMILES string [24] has been assigned allowing for computer recognition of each structure
- tautomer enumeration [25] has been performed for enhanced searching capability, so that the user does not need to know about tautomer preferences in order to obtain the required hits (encoding preferences is outside the scope of this work)
- ionisation assignments have been added (defining each ionisation to be acidic or basic, but with no definition of the substructural site of ionisation) for enhanced searching capability and the calculation of logD values [26]
- pKb values have been converted to pKa values using the method of Bandura and Lvov [27]

Other computer-readable databases have been produced [28, 29], but only the IUPAC Databases include the added value listed above, together with all the features and search capabilities described below:-

- IUPAC critical data quality assessment for Aqueous Database and authors' estimate of precision for Non-aqueous Database
- full primary reference
- method description for each determination
- good range of organic chemistry with applicability to pharmaceutical, agrochemical and specialty chemicals research, as well as pKa prediction research
- Databases can be merged with existing in-house data, with the IUPAC-sourced data clearly identified
- very flexible searching due to careful field assignment, e.g.
 - substructure, similarity, exact match
 - search for basic pKa with $6.5 < \text{pKa} < 7.5$
 - search for molecule using any relevant tautomer

- use only the highest quality data (Aqueous Database)
- search for pKa values in dimethyl sulfoxide (Non-aqueous Database)
- search for data where $35^\circ\text{C} \leq \text{temp} \leq 40^\circ\text{C}$
- any combination of the above

IUPAC data quality assessment

Two types of experimental conditions were considered by the IUPAC compilers in order to reach a conclusion about the reliability of the data. Firstly, the experimental variables were studied. These included the nature of the system under study, the reproducibility of the measurements, the purity and stability of the materials being investigated (e.g. isomeric content) and of the solvents used, the constancy of the temperature, etc. Secondly, the method of measurement and the procedures for evaluating the quality of the measurement were inspected. For example, the range of concentrations used and the graphical or mathematical methods involved in the extrapolation of the results to zero ionic strength were examined.

The IUPAC definitions of their data quality assessment vary in a very minor way in some of the books of aqueous pKa values. For example, some have extra ranges, but the key definitions are for “reliable”, “approximate” and “uncertain” as illustrated in Table 1.

The IUPAC textual descriptions of these data quality assessments appear to contrast with the tight definitions provided in Table 1. For example, the assessment of “approximate” has an estimated uncertainty in the pKa measurement of $\leq \pm 0.04$, so a description of “approximate” in this context seems somewhat strict. It must be remembered, however, that pKa is a logarithmic quantity, so a change of 0.04 in pKa represents an almost 10 % change in Ka.

By collapsing the ranges in Table 1 to end up with just four, “reliable”, “approximate”, “uncertain” and “very uncertain”, the percentage of pKa measurements falling into these categories is illustrated in Fig. 2. Note that

Table 1 Definitions of IUPAC data quality assessment across all four books [18–21] of aqueous pKa values

| Base 1 ^a | Base 1 ΔpKa | Acid 1 ^b | Acid 1 ΔpKa | Base 2 ^c | Base 2 ΔpKa | Acid 2 ^d | Acid 2 ΔpKa |
|---------------------|-------------|---------------------------|-------------|---------------------|-------------|---------------------|-------------|
| | | Very reliable | ~ ±0.0005 | | | | |
| | | Very reliable to reliable | | | | | |
| Reliable | ≤ ±0.005 | Reliable | ~ ±0.005 | Reliable | ≤ ±0.005 | Reliable | ≤ ±0.005 |
| | | Reliable to approximate | | | | | |
| Approximate | ≤ ±0.04 | Approximate | ≤ ±0.04 | Approximate | ≤ ±0.04 | Approximate | ≤ ±0.04 |
| | | Approximate to uncertain | | | | | |
| Uncertain | > ±0.04 | Uncertain | > ±0.04 | Uncertain | > ±0.04 | Uncertain | > ±0.04 |
| Very uncertain | | | | | | Very uncertain | Very great |

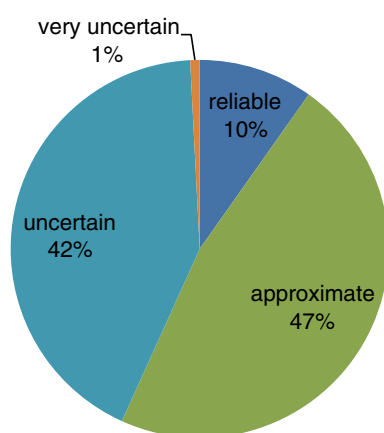
The ΔpKa columns define the estimated uncertainty in the pKa measurements

^a Reference [18]

^b Reference [20]

^c Reference [19]

^d Reference [21]

**Fig. 2** Distribution of data quality assessments in the IUPAC Aqueous database

“reliable” now includes “very reliable”, “very reliable to reliable” and “reliable” for example.

Statistics for the IUPAC pKa databases

(1) Aqueous Database

Base 1 [18]

Dissociation Constants of Organic Bases in Aqueous Solution, by D. D. Perrin

3,775 molecules, 8,766 pKa values

Acid 1 [20]

Dissociation Constants of Organic Acids in Aqueous Solution, by G. Kortum, W. Vogel and K. Andrussov

1,063 molecules, 2,893 pKa values

Base 2 [19]

Dissociation Constants of Organic Bases in Aqueous Solution, Supplement 1972, by D. D. Perrin

4,275 molecules, 7,844 pKa values

Acid 2 [21]

Ionisation Constants of Organic Acids in Aqueous Solution, by E. P. Serjeant and Boyd Dempsey

4,584 molecules, 10,912 pKa values

The complete IUPAC Aqueous Database comprises 13,697 molecules (12,189 unique) and 30,415 pKa values (29,946 unique).

(2) Non-Aqueous Database

Acid–Base Dissociation Constants in Dipolar Aprotic Solvents, by K. Izutsu [22]

1,956 molecules (1,165 unique), 2,945 pKa values in the following 12 solvents:-

Acetone
Acetonitrile
N,N-Dimethylacetamide
N,N-Dimethylformamide
Dimethyl sulfoxide
Hexamethylphosphoric triamide
Isobutyl methyl ketone
1-Methyl-2-pyrrolidinone
Nitromethane
Propylene carbonate
Pyridine
Sulfolane

Access to the IUPAC pKa Databases

There are three methods of accessing the Databases, the second of which was being worked on at the time of writing the manuscript:-

1. Obtain the Aqueous and Non-Aqueous Databases for inclusion into an in-house database management system (via the pKaData website [23])
2. Search the Databases directly on the pKaData website via a secure link [23]
3. Use the sophisticated search software, pKa Prospector [30], from OpenEye Scientific Software

Acknowledgments The author would like to thank Peter Kenny, Andrew Grant, David Cosgrove, Thorsten Nowak, Anthony Nicholls, Joe Corkery, Huw Jones, Martin Harrison, Nicola Colclough, Philip Jewsbury, Andrew Poirrette, Neil Hales and Sandra McLaughlin for discussions, help and encouragement before and during the conversion of the IUPAC books into computer-readable form.

References

1. Aqueous pKa compilation by Williams R (document compiled by Jencks WP, added to by Westheimer FH). http://research.chem.psu.edu/brpgroup/pKa_compilation.pdf. Accessed 11 Mar 2014
2. Aqueous pKa compilation with some dimethyl sulfoxide values by Ripin DH, Evans DA. http://www2.lsddiv.harvard.edu/labs/evans/pdf/evans_pKa_table.pdf. Accessed 11 Mar 2014
3. pKa values in dimethyl sulfoxide compilation by Bordwell FG. <http://www.chem.wisc.edu/areas/reich/pkatable/index.htm>. Accessed 12 Mar 2014
4. Links to tables of pKa values in non-aqueous solvents taken from literature by Leito I. http://tera.chem.ut.ee/~ivo/HA_UT/ Accessed 12 Mar 2014
5. Albert A (1963) Ionization constants of heterocyclic substances. In: Katritzky AR (ed) Physical methods in heterocyclic chemistry. Academic Press, New York
6. Ritschel WA (1972) pKa values and some clinical applications. In: Francke D, Whitney H (eds) Perspectives in clinical pharmacy, 1st edn. Drug Intelligence Publications, Hamilton, pp 325–367
7. Smith S, Rawlins M (1973) Appendix C. Variability in human drug response, 1st edn. Butterworths, London, pp 154–165
8. Speight T (1987) Avery's drug treatment, 3rd edn. Publishing Sciences Group Inc, Littleton, pp 1352–1380
9. Hoover J (1976) In: Dispensing of Medication, 8th edn. Mack Publishing, Easton, PA, p. 230, 247, 418–426, 468–634.
10. Newton D, Kluza R (1978) pKa values of medicinal compounds in pharmacy practice. Drug Intell Clin Pharm 12:546–554
11. Craig P (1990) In: Hansch C, Sammes P, Taylor J (eds) Compendium of drugs, in comprehensive medicinal chemistry, 1st edn. Pergamon Press, New York, pp 237–965
12. Williams D (1995) Appendix A-1. In: Foye W, Lemke T, Williams D (eds) Principles of medicinal chemistry, 4th edn. Williams and Wilkins, Baltimore, pp 948–961
13. Delgado J, Remers W (2000) pKas of drugs and reference compounds. In: Delgado J, Remers W (eds) Wilson and Gisvold's textbook of organic medicinal and pharmaceutical chemistry, 8th edn. Lippincott-Raven, Philadelphia
14. Pranker RJ (2007) In: Brittain HG (ed) Profiles of drug substances, excipients and related methodology volume 33, 1st edn. Academic Press, London
15. Liao C, Nicklaus M (2009) Comparison of nine programs predicting pKa values of pharmaceutical substances. J Chem Inf Model 49:2801–2812
16. Manchester J, Walkup G, Rivin O, You Z (2010) Evaluation of pKa estimation methods on 211 druglike compounds. J Chem Inf Model 50:565–571
17. Cox BG (2013) Acids and bases solvent effects on acid–base strength, 1st edn. Oxford University Press, Oxford
18. Perrin DD (1965) Dissociation constants of organic bases in aqueous solution, 1st edn. Butterworths, London
19. Perrin DD (1972) Dissociation constants of organic bases in aqueous solution, supplement, 1st edn. Butterworths, London
20. Kortum G, Vogel W, Andrussov K (1961) Dissociation constants of organic acids in aqueous solution, 1st edn. Butterworths, London
21. Serjeant E, Dempsey B (1979) Ionisation constants of organic acids in aqueous solution, 1st edn. Pergamon Press, Oxford
22. Izutsu K (1990) Acid–base dissociation constants in dipolar aprotic solvents, 1st edn. Blackwell Scientific Publications, Oxford
23. IUPAC books of experimental pKa values of organic acids and bases in both aqueous and non-aqueous solvents as searchable pKa databases by Slater AM. <http://pkadata.com>. Accessed 13 March 2014
24. SMILES—A Simplified Chemical Language. <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>. Accessed 14 March 2014
25. Tautomer enumeration using QuacPac by OpenEye Scientific Software. <http://www.eyesopen.com/quacpac>. Accessed 14 March 2014
26. Partition coefficient definition. http://en.wikipedia.org/wiki/Partition_coefficient. Accessed 14 March 2014
27. Bandura AV, Lvov SN (2006) The ionization constant of water over a wide range of temperatures and densities. J Phys Chem Ref Data 35:15–30
28. ACD Labs internal pKa database. <http://www.acdlabs.com/products/percepta/predictors/pka/>. Accessed 14 March 2014
29. BioByte raw data. http://www.biobyte.com/bb/Prices_2.html. Accessed 14 March 2014
30. pKa Prospector, search software from OpenEye Scientific Software. <http://www.eyesopen.com/pka-prospector>. Accessed 14 March 2014