

Cite this: *CrystEngComm*, 2012, 14, 6362–6365

www.rsc.org/crystengcomm

COMMUNICATION

Acid–base crystalline complexes and the pK_a rule

Aurora J. Cruz-Cabeza*^{ab}

Received 3rd July 2012, Accepted 26th July 2012

DOI: 10.1039/c2ce26055g

Differences in the predicted aqueous pK_a values (ΔpK_a) have been calculated for 6465 crystalline complexes containing ionised and non-ionised acid–base pairs in the Cambridge Structural Database. A linear relationship between ΔpK_a and the probability of proton transfer between acid–base pairs has been derived for crystalline complexes with ΔpK_a between -1 and 4 . The pK_a rule is validated and quantitated.

Drugs are often delivered in the form of salts^{1,2} and, more recently, also in the form of cocrystals.^{3,4} Multicomponent crystals can provide better drug performance than single component crystals,^{1–4} e.g. improved solubility^{5,6} and physico-chemical stability^{7,8} or tailored pharmacokinetics.^{1–4} It has been estimated that around half of the commercial drugs are formulated as salts.¹ Salt formation requires active pharmaceutical ingredients (APIs) with ionisable groups whereas, in principle, all drug molecules can be formulated as cocrystals provided the right cocrystal formers are selected. The range, nature and volume of the pharmaceutically acceptable cocrystal formers possible is much wider than that of the counter ions and the formulation of cocrystals might also present some advantages over that of salts, such as a more predictable structure and stoichiometry.⁹

For salt selection, the counter ions are often chosen using the pK_a rule. It is accepted that when the pK_a difference between a cocrystallising acid and base is greater than 2 or 3 ($\Delta pK_a = pK_a[\text{protonated base}] - pK_a[\text{acid}] > 2$ or 3), salt formation is expected.¹ This pK_a rule of thumb has been derived after years of experience in salt formulation within the pharmaceutical industry and it has also been exemplified by a limited number of studies, at least in the published literature, on small datasets of multicomponent complexes.^{10,11} The origin and universality of this rule has not been formally proven, even though it is, without doubt, a rule widely used in salt design.

With the growing interest in the design and formulation of cocrystals, an extension of the pK_a rule has recently been proposed for cocrystals containing an acid–base pair. Nangia *et al.*¹² studied a series of cocrystals and salts of carboxylic acids and pyridines and concluded that, at least for these types of complexes, negative pK_a differences between the cocrystallising acid–base pairs will

almost always afford cocrystals but not salts. It has also been highlighted that in domains of ΔpK_a values between 0 and 3, if an acid and a base cocrystallise, the prediction of cocrystal or salt observation based on the pK_a values of the acid and the base is no longer possible.¹³ In this region of ΔpK_a , the crystal structure of the complex itself has a strong influence on the location of the acid proton and, sometimes, the experimental determination of the proton position can be challenging.

It is the purpose of the present study to generate diverse and sufficient data in order to verify and quantify the pK_a rule. With the growing number of crystal structures of multicomponent crystals available in the Cambridge Structural Database (CSD)¹⁴ and the development of fast and reasonably accurate pK_a calculators, we are now in a unique position to verify and quantify such an important and widely used rule in the design of salts.

The CSD best R factor list (Nov 2010) was searched for a filtered set of multicomponent crystal structures using Conquest.¹⁵ Only the most common elements in organic molecules were allowed (C, H, N, O, S and halogens) and crystal structures with more than one component were retrieved. Crystal structures showing disorder, crystallographic errors, polymeric structures and structures containing zwitterions or quaternary ions were excluded. The Conquest search option “no ions” was used to separate the dataset into two subsets: crystal structures containing ions and crystal structures containing multiple non-ionised components. Structures were then analysed using Pipeline Pilot.¹⁶ The pK_a component was used to identify acid and bases using a predefined set of 62 acid and base queries (broadly classified as aliphatic, aromatic and phenolic acids as well as aliphatic, aromatic and heterocyclic bases). In addition to these, the inorganic acids HCl, HBr, HI, and HNO₃ and HF were also considered because of their frequent use in salt synthesis,^{17,18} and water (as both, acid and base). Only crystal structures containing at least one acid–base (AB) pair were retained. Aqueous pK_a values were calculated for the acids and the protonated bases using two different models: (1) the pK_a calculator component implemented in Pipeline Pilot¹⁶ and (2) the pK_a plugin implemented in Marvin.¹⁹† The range of pH sampled for the pK_a calculations was -10 to 20 . Only crystal structures in which one single proton was transferred, or the potential transfer of the most acidic proton of the acid to the most basic atom of the base, were considered (first ionisation constant only). Final pK_a values (pK_a [protonated base], pK_a [acid] and ΔpK_a) were calculated averaging the predictions from both models. In order to minimise the errors due to the uncertainties of the pK_a calculations, the data were filtered using the following

^aCambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK

^bVan't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. E-mail: aurorajosecruz@gmail.com; Tel: (+31) 20 525 8264

Table 1 Number of crystal structures with different number of components in the AB and A⁻B⁺ subsets

Number of components	AB	A ⁻ B ⁺
2	2301 (88.6 %)	2768 (71.6 %)
3	282 (10.9 %)	982 (25.4 %)
4	11 (0.4 %)	112 (2.9 %)
> 4	3 (0.1 %)	6 (0.1 %)

criterion: only structures with a maximum deviation between the final ΔpK_a and the $\Delta pK_a[\text{model}]$ of 1 were retained. For the analysis, the data were binned in ΔpK_a intervals of 1 ΔpK_a (eg. bin at $\Delta pK_a = 0$ contains crystal structures with calculated ΔpK_a values from -0.5 to 0.5).

The final set of acid–base complexes (6465) contains 2597 crystal structures with non-ionised and 3868 with ionised acid–base pairs (AB and A⁻B⁺ subsets, respectively). The AB subset consists of unhydrated/unsolvated cocrystals but also hydrates, solvates and hydrated/solvated cocrystals and their polymorphs. The A⁻B⁺ subset consists of unhydrated/unsolvated salts but also hydrated/solvated salts, cocrystals of salts and hydrated/solvated cocrystals of salts and their polymorphs.²⁰ Table 1 summarises the number of crystal structures containing a different number of components in both subsets. Ionised acid–base pairs, A⁻B⁺, are more likely to crystallise with extra components (28.4 %) than non-ionised pairs, AB (11.4 %), an observation that has been discussed in detail elsewhere.⁹

The number of occurrences of AB and A⁻B⁺ per ΔpK_a bin are presented in Fig. 1. Three zones of ΔpK_a domains were defined and analysed separately because of their different behaviour (Fig. 1 and Table 2): zone 1 ($\Delta pK_a < -1$), zone 2 ($-1 \leq \Delta pK_a \leq 4$) and zone 3 ($\Delta pK_a > 4$).

In zone 1, crystal structures containing non-ionised AB pairs are almost exclusively observed (99.1 %). The dataset in zone 1 is dominated by hydrates of acids and bases of varying strength in which water acts as either an extremely weak base or acid. A varied set of phenols (acids, $pK_a \sim 9$), pyrazine and caffeine

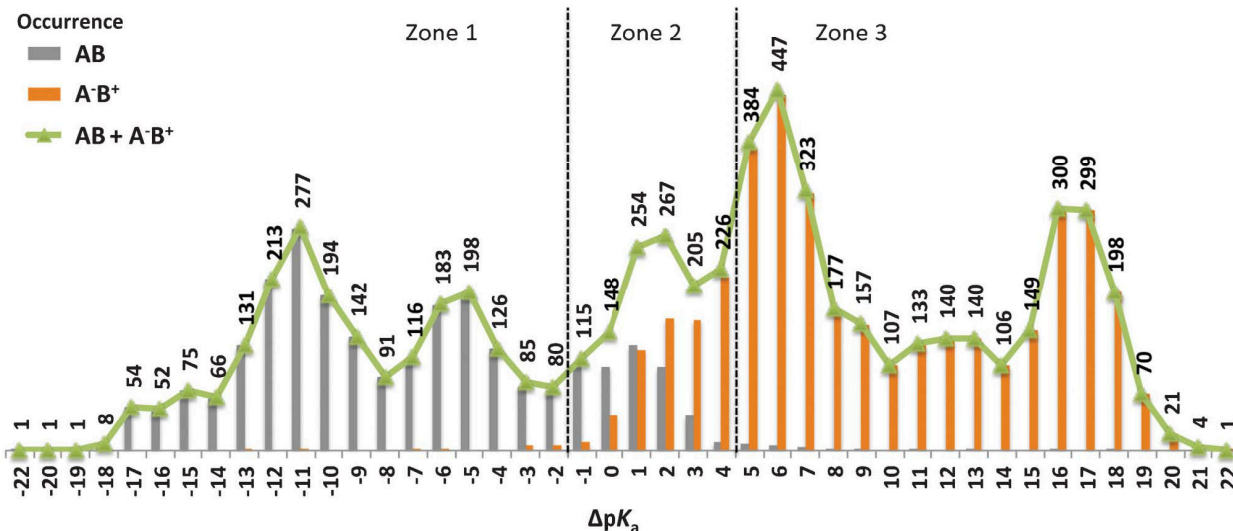
Table 2 Number of crystal structures and percentages of AB and A⁻B⁺ within the 3 different ΔpK_a zones

Zone	ΔpK_a domain	AB	A ⁻ B ⁺
1	< -1	2076 (99.1 %)	18 ^a (0.9 %)
2	$[-1, 4]$	495 (40.7 %)	720 (59.3 %)
3	> 4	26 ^a (0.8 %)	3130 (99.2 %)

^a Mostly structures with assignments of the acid hydrogen atom positions.

(bases, $pK_a \sim 0$) are also weak acids and bases commonly found in AB complexes in zone 1. In zone 3, crystal structures containing ionised A⁻B⁺ pairs are almost exclusively observed (99.2%). The A⁻B⁺ dataset in zone 3 mainly consists of salts of very strong acids (hydrogen halides, nitric and methanesulfonic acids with $pK_a < 0$) and/or very strong bases (ammonia and various primary, secondary and tertiary amines with $pK_a \sim 10$). The 18 ionised A⁻B⁺ complexes in zone 1 and the 26 non-ionised AB complexes in zone 3 (Table 2) correspond, mostly, to structures with dubious assignments of the acid hydrogen atom positions and, in a lesser extent, to deviations of the calculated ΔpK_a values. Hybridisation assignments and hydrogen atom positions were taken as reported by the crystallographers.

In zone 2, crystal structures containing both ionised A⁻B⁺ and non-ionised AB pairs are observed. Zone 2 is perhaps the most interesting domain of ΔpK_a values with regards to the design of salts and cocrystals. The dataset consists of acid–base complexes with acids and bases of similar strengths (Fig. 2). Most acids in zone 2 have pK_a values between 3 and 4 whilst most protonated bases have pK_a values between 4 and 7. From strong to weak, the acids in zone 2 include activated carboxylic acids ($pK_a \sim 1$), aliphatic dicarboxylic acids ($pK_a \sim 2$ to 3), benzoic acid derivatives ($pK_a \sim 3$ to 5), barbituric acid derivatives ($pK_a \sim 4$), aliphatic monocarboxylic acids ($pK_a \sim 5$) and activated phenols ($pK_a \sim 7$). From weak to strong, bases in zone 2 include phenazine derivatives ($pK_a \sim 2$), aniline derivatives ($pK_a \sim 3$ to 4), pyridine, 2/5-aminopyrimidines and imidazole derivatives ($pK_a \sim 4$ to 7) and

**Fig. 1** Occurrence of AB (grey) and A⁻B⁺ (orange) as a function of the calculated ΔpK_a (6465 acid–base complexes).

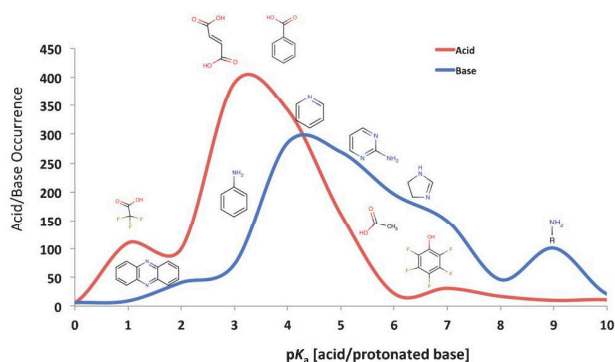


Fig. 2 Occurrence of acids (red) and bases (blue) of different strength (pK_a values) in zone 2.

amines ($pK_a \sim 9$). Complexes containing carboxylic acids and pyridine derivatives are the most common in both ionised and non-ionised datasets (293 AB vs. 190 A^-B^+) in zone 2.

In zone 2, the relative occurrence† of A^-B^+ complexes increases linearly with increasing ΔpK_a values (from $\sim 10\%$ to $\sim 95\%$) whilst that of AB complexes decreases (Fig. 3). At $\Delta pK_a = 1$, the occurrences of AB and A^-B^+ are practically equal (Fig. 1 and 3). The relative occurrences of AB and A^-B^+ (within zone 2 only, crossed circles in Fig. 3) were fitted to two lines using linear regression (dotted lines in Fig. 3) with an extremely good fit ($R^2 = 0.996$). Given the pK_a values of an acid and a base cocrystallising together, the probability of them forming a non-ionised or an ionised acid–base complex can now be estimated using the equations:

$$P_{\text{obs}}(\text{AB}, \%) = -17 \Delta pK_a + 72 \text{ for } -1 \leq \Delta pK_a \leq 4$$

$$P_{\text{obs}}(A^-B^+, \%) = 17 \Delta pK_a + 28 \text{ for } -1 \leq \Delta pK_a \leq 4$$

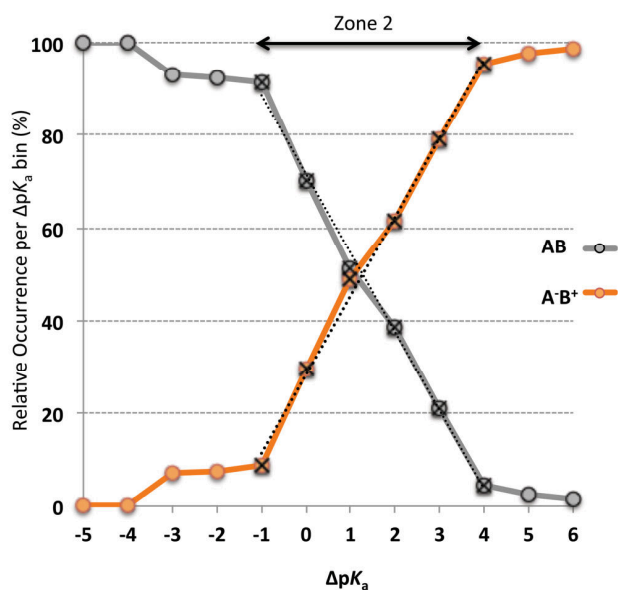


Fig. 3 Relative occurrences of AB (grey) and A^-B^+ (orange) as a function of the calculated ΔpK_a .

where ΔpK_a is the difference between the pK_a of the protonated base and the pK_a of the acid and P_{obs} is the probability of observing AB or A^-B^+ (in %).

In summary, given an acid and a base forming a crystalline complex, if their $\Delta pK_a < -1$ the exclusive observation of a non-ionised AB complex is expected, if their $\Delta pK_a > 4$ the exclusive observation of an ionised AB complex is expected. Between $-1 \leq \Delta pK_a \leq 4$, every increase in one ΔpK_a difference increases the probability of proton transfer between the acid and the base (salt formation) by 17% from $\sim 10\%$ at $\Delta pK_a = -1$ to $\sim 95\%$ at $\Delta pK_a = 4$. These observations are in good agreement with previous formulations of the pK_a rule that can now be quantified. When a cocrystallising acid and base have a $\Delta pK_a > 2$, > 3 or > 3.75 ,¹⁰ the herein derived equations estimate ~ 60 , 80 and 90% likelihood of salt formation, respectively. At negative values of ΔpK_a , on the other hand, the observation of non-ionised acid–base complexes clearly dominates over that of salts (89 vs. 11% at $\Delta pK_a = -1$), which is in agreement with recent observations in acid–base cocrystals.¹²

Around $\Delta pK_a \sim 1$, predicting the location of the acid proton based on aqueous ΔpK_a data alone is not possible: the model predicts a very similar likelihood of observing ionised/non-ionised acid–base complexes. Small changes in the crystalline environment can influence the location of the acid proton. It was noticed that, for structures with ΔpK_a from 0 to 2, carboxylates are involved in 3 or more hydrogen bonds in 80% of A^-B^+ structures whilst carboxylic acids are only involved in 3 or more hydrogen bonds in 10% of the AB structures. Secondary hydrogen bond interactions can certainly influence the stabilisation of the ionised acid–base pairs in the crystal as well as in solution.²¹ Complex behaviour and proton disorder has also been observed at $\Delta pK_a \sim 1$ in what has been referred to as the “salt–cocystal continuum”.¹³ The herein studied dataset, however, does not contain structures with disorder.

Finally, it is acknowledged that pK_a values change in different solvents and that the choice of solvent is extremely important in the synthesis of organic salts.²² The data studied here exclusively concerns aqueous pK_a values, which are the most common values measured experimentally and those for which most pK_a predictors have been developed. It would be interesting, however, to derive similar data for different solvents as this might help the actual selection of the solvent for salt synthesis.

Conclusions

The occurrences of crystal structures containing neutral and ionised acid–base pairs have been presented together with their calculated ΔpK_a values for a large and varied set of acid–base multicomponent crystal structures in the CSD. The data presented validates previous observations of the pK_a rule in the literature that are based on much smaller datasets of crystal structures. Ionised acid–base complexes are observed exclusively for $\Delta pK_a > 4$ and non-ionised acid–base complexes are observed exclusively for $\Delta pK_a < -1$. For acid–base complexes whose ΔpK_a values lie between -1 and 4 , a linear relationship between the probability of salt formation and their ΔpK_a value has been derived. This relationship has allowed us to quantify previous formulations of the pK_a rule and can be a valuable tool for scientists involved in cocrystal design and salt selection strategies.

Acknowledgements

The Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) is acknowledged for a VENI fellowship.

References

† The pK_a component in Pipeline Pilot estimates the pK_a data of a given molecule by calculating its molecular fingerprint and then applying different learning models derived from ~12 000 experimental pK_a values of organic molecules in water at 25 °C. The pK_a calculator implemented in Marvin predicts pK_a data based on calculations of partial charges of atoms in the molecules.

‡ Relative occurrence of $[AB] = (AB \text{ observations}) / (AB + A^-B^+ \text{ observations})$; relative occurrence of $[A^-B^+] = (A^-B^+ \text{ observations}) / (AB + A^-B^+ \text{ observations})$

- 1 P. H. Stahl, C. G. Wermuth, *Handbook of Pharmaceutical Salts: Properties, Selection, and Use*, Wiley-vch, 2002.
- 2 S. M. Berge, L. D. Bighley and D. C. Monkhouse, *J. Pharm. Sci.*, 1977, **66**, 1–19.
- 3 P. Vishweshwar, J. A. McMahon, J. A. Bis and M. J. Zaworotko, *J. Pharm. Sci.*, 2006, **95**, 499–516.
- 4 N. Shan and M. J. Zaworotko, *Drug Discov. Today*, 2008, **13**, 440–446.
- 5 A. T. M. Serajuddin, *Adv. Drug Deliv. Rev.*, 2007, **59**, 603–616.
- 6 D. J. Good and N. Rodriguez-Hornedo, *Cryst. Growth Des.*, 2009, **9**, 2252–2264.
- 7 N. Schultheiss and A. Newman, *Cryst. Growth Des.*, 2009, **9**, 2950–2967.
- 8 A. V. Trask, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2005, **5**, 1013–1021.
- 9 C. B. Aakeröy, M. E. Fasulo and J. Desper, *Mol. Pharmaceutics*, 2007, **4**, 317–322.
- 10 S. L. Johnson and K. A. Rumon, *J. Phys. Chem.*, 1965, **69**, 74–86.
- 11 W. Q. T. Tong and G. Whitesell, *Pharm. Dev. Technol.*, 1998, **3**, 215–223.
- 12 B. R. Bhogala, S. Basavoju and A. Nangia, *CrystEngComm*, 2005, **7**, 551–562.
- 13 S. L. Childs, G. P. Stahly and A. Park, *Mol. Pharmaceutics*, 2011, **4**, 323–338.
- 14 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.
- 15 I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 389–397.
- 16 *SciTegic Pipeline Pilot*, Accelrys Inc.: San Diego, CA 2008.
- 17 D. A. Haynes, W. Jones and W. D. S. Motherwell, *J. Pharm. Sci.*, 2005, **94**, 2111–2120.
- 18 G. S. Paulekuhn, J. B. Dressman and C. Saal, *J. Med. Chem.*, 2007, **50**, 6665–667.
- 19 Marvin 5.10.1, 2012, ChemAxon (<http://www.chemaxon.com>).
- 20 S. Aitipamula, R. Banerjee, A. K. Bansal, K. Biradha, M. L. Cheney, A. R. Choudhury, G. R. Desiraju, A. G. Dikundwar, R. Dubey, N. Duggirala, P. P. Ghogale, S. Ghosh, P. K. Goswami, N. R. Goud, R. R. K. R. Jetti, P. Karpinski, P. Kaushik, D. Kumar, V. Kumar, B. Moulton, A. Mukherjee, G. Mukherjee, A. S. Myerson, V. Puri, A. Ramanan, T. Rajamannar, C. M. Reddy, N. Rodriguez-Hornedo, R. D. Rogers, T. N. G. Row, P. Sanphui, N. Shan, G. Shete, A. Singh, C. C. Sun, J. A. Swift, R. Thaimattam, T. S. Thakur, R. Kumar Thaper, S. P. Thomas, S. Tothadi, V. R. Vangala, N. Variankaval, P. Vishweshwar, D. R. Weyna, and M. J. Zaworotko, *Cryst. Growth Des.*, 2012, **12**, 2147–2152.
- 21 V. Balevicius, R. Bariseviciute, K. Aidas, I. Svoboda, H. Ehrenberg and H. Fuess, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3181–3189.
- 22 S. N. Black, E. A. Collier, R. J. Davey and R. J. Roberts, *J. Pharm. Sci.*, 2007, **96**, 1053–1068.