Accepted Manuscript

Title: Proton dissociation properties of arylphosphonates: Determination of accurate Hammett equation parameters

Authors: Gergő Dargó, Adrienn Bölcskei, Alajos Grün, Szabolcs Béni, Zoltán Szántó, Antal Lopata, György Keglevich, György T. Balogh

PII: S0731-7085(17)30868-3
DOI: http://dx.doi.org/doi:10.1016/j.jpba.2017.05.038
Reference: PBA 11292

To appear in: *Journal of Pharmaceutical and Biomedical Analysis*

Received date: 7-4-2017
Revised date: 19-5-2017
Accepted date: 21-5-2017

Please cite this article as: Gergő Dargó, Adrienn Bölcskei, Alajos Grün, Szabolcs Béni, Zoltán Szántó, Antal Lopata, György Keglevich, György T.Balogh, Proton dissociation properties of arylphosphonates: Determination of accurate Hammett equation parameters, Journal of Pharmaceutical and Biomedical Analysishttp://dx.doi.org/10.1016/j.jpba.2017.05.038

Proton dissociation properties of arylphosphonates: Determination of accurate Hammett equation parameters

Gergő Dargó[a,b], Adrienn Bölcskei[b], Alajos Grün[b], Szabolcs Béni[c], Zoltán Szántó[d], Antal Lopata[d], György Keglevich[b], György T. Balogh[a*]

[a] *Compound Profiling Laboratory, Gedeon Richter Plc., Gyömrői út. 19-21., 1107 Budapest, Hungary*

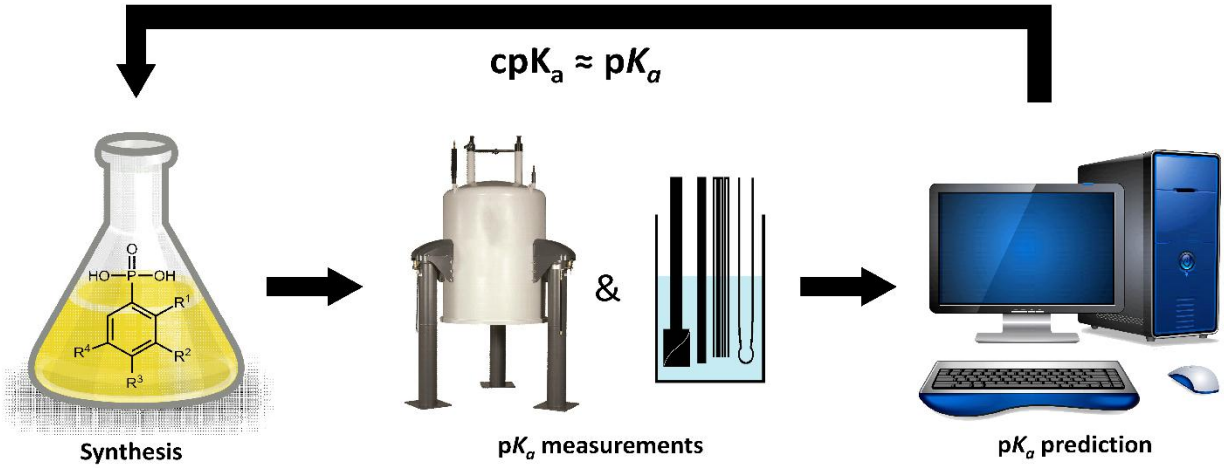
[b] *Department of Organic Chemistry and Technology, Budapest University of Technology and Economics, Műegyetem rakpart 3., 1111 Budapest, Hungary*

[c] *Department of Pharmacognosy, Semmelweis University, Üllői út 26, 1085 Budapest, Hungary*

[d] *CheMicro Ltd., Károly körút 1., 1075 Budapest, Hungary*

[*] Corresponding author at Compound Profiling Laboratory, Gedeon Richter Plc., Gyömrői út 19-21, H-1475 Budapest, Hungary. E-mail: gy.balogh@richter.hu Phone: +36 1 431 4855 Fax: +36 1 889 8782.

Graphical Abstract



Highlights

- Determination of the proton dissociation constants of arylphosphonic acid derivatives
- Comparison of our experimental values to predicted and formerly measured pK_a values
- Hammett equations of proton dissociation of arylphosphonic acid derivatives were refined

Abstract

Determination of the proton dissociation constants of several arylphosphonic acid derivatives was carried out to investigate the accuracy of the Hammett equations available for this family of compounds. For the measurement of the pK_a values modern, accurate methods, such as the differential potentiometric titration and NMR-pH titration were used. We found our results significantly different from the pK_a values reported before (pK_{a1}: MAE= 0.16 pK_{a2}: MAE=0.59). Based on our recently measured pK_a values, refined Hammett equations were determined that might be used for predicting highly accurate ionization constants of newly synthesized compounds (pK_{a1} = 1.70 – 0.894σ, pK_{a2} = 6.92 – 0.934σ).

Keywords: Arylphosphonic acids; ; ; ; , Hammett equation, pH-metric pK_a measurements, NMR-pH titration, pK_a predictors

1 Introduction

Arylphosphonates are widespread intermediates in the synthesis of arylphosphonic acids [1] and many of these compounds possess biological activity. In recent years, several potential pharmaceutical applications of biologically active molecules containing arylphosphonic acid unit and its derivatives have been reported, like metabotropic glutamate receptor antagonists [2–5], tyrosine kinase inhibitors [6], protein tyrosine phosphatase inhibitors [7], carbonic anhydrase inhibitors [8], growth factor receptor bound protein 2 SH2 domain inhibitors [9], metallo-ß-lactamase inhibitors [10] and eIF4E inhibitors [11].(Fig. 1.)

Proton-dissociation capability has a great impact on both pharmacodynamics and pharmacokinetic properties of drug molecules. The majority of contemporary therapeutic substances possess acid/base character according to a 2013 chemogenomic analysis [12]. Quantitative description of ionization state, as described by proton dissociation constant (pK_a), has therefore a crucial role in drug discovery. Moreover, pK_a value also reflects the ionization state of active pharmaceutical ingredients in various physiologically relevant media. Ionization has a characteristic impact on ADMET (A: absorption, D: distribution, M: metabolism, E: excretion, T: toxicity) properties, including solubility, lipophilicity, dissolution rate, membrane permeability, plasma protein binding, CNS penetration, P-gp efflux, hERG inhibition and cytochrome P450 inhibition [13]. The calculation of distribution coefficient (log$D$) also requires accurate pK_a values. Since ionization is a crucial factor among the drug-like properties, the measurement of the pK_a value(s) of new chemical substances is required by both the FDA (Food and Drug Administration) and the OECD (Organizations for Economic Cooperation and Development).

The pK_a values can be determined by experimental methods or predicted by *in silico* calculations. There are a number of experimental techniques available for the evaluation of proton dissociation constant, such as potentiometric and spectrophotometric titrations, capillary electrophoresis (CE), HPLC and NMR-pH titration [14–16]. Among them the most extensively used is the pH-potentiometric method, in the industrial drug discovery settings [15], due to the robust generally applicable methodology, which provides reliable pK_a values even in medium throughput mode. Although experimental pK_a values are essential for the characterization of drug discovery compounds, effective design of new drug-like molecules also requires efficient *in silico* tools for prediction of ionization constants. The *in silico* methods of pK_a prediction can be classified into two major groups: empirical methods and methods based on quantum chemical calculations. The empirical methods can be further divided into three subgroups based on the approach they use: methods utilizing empirical relations of Hammett and Taft (Linear free-energy relationships (LFER) methods), methods correlating calculated structural descriptors with pK_a (Quantitative structure–property relationships (QSPR) methods) and database lookup methods that calculate the pK_a based on experimental pK_a values of structurally similar molecules of a predetermined database[17].

The classic empirical method used for predicting pK_a values is based on the Hammett equation. In 1935, Hammett found a linear relationship between the pK_a values and the substitution parameters (Hammett σ) of the substituents in the case of benzoic acid and its meta- and para-substituted derivatives [18,19].

$$\log(K_{a,S}/K_{a,0})=\sigma_S \qquad \text{Eq. (1)}$$

A similar relationship was found in the case of aliphatic and ortho-substituted aromatic compounds by Taft et al. in 1952 [20–22]. Based on these observations the initial equation can be extended for multiple substituents and classes of compounds leading to the following general equation:

$$pK_{a,S} = pK_{a,0} - \rho\sum\sigma_S \quad \text{Eq. (2)}$$

referring to compounds/reactions where the pK_a proton dissociation constants are influenced by electronic substituent effects only, and steric effects do not occur. In Eq. (2) p$K_{a,0}$ is the proton dissociation constant of the parent molecule of the class of compounds, p$K_{a,S}$ is the proton dissociation constant of the substituted derivative, σ_S is the electronic substituent constant, characteristic to a given substituent, ρ is the reaction constant, or sensitivity constant, which describes the susceptibility of the reaction to substituent effects, compared to the ionization of benzoic acid. Since the pK_a values of phosphonic acid derivatives depend on the substituents of the aromatic ring, the use of a Hammett-type equation to predict the pK_a values in this class of compounds is valid.

There are several publications on determining the parameters of Hammett equations and Taft equations for arylphosphonic acids in the literature [23–25]. However, the experimental pK_a values used for this purpose are highly variable due to varying experimental environment (ionic strength, temperature) and outdated measurement techniques. As a consequence, the Hammett equations of arylphosphonic acids reported before might describe the correlation between substituents and pK_a values less accurately. Therefore, here we report the pK_a values of several arylphosphonic acids measured by highly accurate methods (differential pH-metric titration, NMR-pH titration), compare the results with *in silico* predicted values (cpK_a) and those measured by Nagarajan et al. [26] based on the goodness of the correlation (R^2, s) and the mean absolute error (MAE) of the compared pK_a values, and determine a more accurate Hammett equation for arylphosphonic acid derivatives.

2 Materials and methods

2.1 Materials

The arylphosphonic acids were synthetized at the Department of Organic Chemistry and Technology at Budapest University of Technology and Economics and 4 further arylphosphonic acids were purchased from Sigma-Aldrich Co. LLC (St. Louis, MO, The United States). The titration reagents (0.5 M potassium hydroxide, hydrochloride acid, and potassium chloride), phosphoric acid solution (85%), cc. acetic acid, tetramethylsilane (TMS) were purchased from Sigma-Aldrich Co. LLC (St. Louis, MO, The United States), methanol was purchased from Merck Millipore.

2.2 pH-metric pK_a measurements

In the case of compounds ***1a-g*** and ***1o-s***, the ionization constants of arylphosphonic acids were measured by pH-metric titration method as described earlier [27].

In the case of compounds ***1h-n*** and ***1t-z***, the SiriusT3™ automated pK_a analyser (Sirius Analytical Instruments Ltd., Forest Row, UK) fitted with combination Ag/AgCl pH electrode was used for determination of dissociation constants. The pK_a values were calculated by SiriusT3Refine™ software (Sirius Analytical Instruments Ltd., Forest Row, UK). Methodologies used by the software have been described in earlier publications [28,29].

In each experiment, 1.50 mL of a 1-5 mM aqueous solution of sample was titrated by the same method, under the same circumstances as in the case of the measurement with the GLpKa apparatus. Three parallel measurements were carried out and the pK_a values of samples were calculated by SiriusT3Refine™ software.

In the case of compounds where we could not measure the proton dissociation constants in aqueous medium due to poor aqueous solubility, the cosolvent dissociation constants (p_sK_a values) were determined in various MeOH–water mixtures (between 15:85 and 70:30, w/w). The same titration protocol was performed as above. Each sample was measured in minimum six different MeOH–water mixtures. To obtain the aqueous pK_a value from p_sK_a data Yasuda–Shedlovsky extrapolation method has been used. This method establishes a correlation with the dielectric constant (ε) using the following equation:

$$p_sK_a + \log[H_2O] = a \cdot \varepsilon + b \qquad \text{Eq. (3)}$$

where $\log[H_2O]$ is the molar water concentration of the given solvent mixture. This method is the most widely used procedure in cosolvent techniques[30,31].

2.3 NMR-pH pK_a measurements

The ^{31}P, ^{13}C, and ^{1}H NMR spectra were obtained on a Bruker DRX-500 spectrometer operating at 202.4, 125.7, and 500 MHz, respectively. Chemical shifts are downfield relative to 85% H_3PO_4 or TMS. The couplings are given in Hz.

For the determination of the low pK_a (< 2.5) values of the aromatic phosphonic acid derivatives ^{1}H NMR-pH titrations were applied. Depending on the solubility, a 0.1–2 mM analyte solution was prepared containing 1 mM dichloroacetic acid as an NMR-pH indicator and 0.05 M 2,2-dimethyl-2-silapentane-5-sulfonate as a chemical shift reference in a 5% D_2O 95% H_2O solution. The presence of such a minute amount of D_2O in the sample, results in less than 0.02 pH unit difference in the pH scale. The ionic strength of the solutions was set to 0.15 M with KCl. A minimum of 15 titration points (i.e. 15 various pH solutions) were prepared in the pH range of 0.6–4.0.

The NMR measurements were carried out on 600 MHz Varian DDR NMR spectrometer equipped with a 5 mm inverse-detection gradient (IDPFG) probehead at 25.0 ± 0.2 °C. ^{1}H NMR spectra of the solutions were recorded with a double pulse field gradient spin echo pulse sequence to suppress the water signal and the spectra were processed by a VnmrJ 3.2 C/Chempack 5.1 software. The low pK_a values of phosphonic acids were determined by the Perrin–Fabian method without the direct measurement of the pH during the titrations [32]. Fitting the ^{1}H chemical shifts of the analyte protons as a function of the ^{1}H NMR chemical shift of the NMR-pH indicator dichloroacetic acid singlet, resulted in the pK_a difference (ΔpK) between dichloroacetic acid and the analyte.

$$\delta L_{obs} = \delta L + \frac{(\delta HL - \delta L)(\delta I_{obs} - \delta I)}{(1 - 10^{\Delta pK})(\delta I_{obs} - \delta I) + 10^{\Delta pK}(\delta HI - \delta I)} \qquad \text{Eq. (4)}$$

where δL_{obs} is the observed chemical shift of the analyte, δI_{obs} is the measured chemical shift of the dichloroacetic acid, δHL and δL are the limiting chemical shifts of the protonated and

non-protonated species while δHI and δI stand for the limiting chemical shifts of the protonated and the anionic forms of dichloroacetic acid, respectively.

The chemical shifts of the aromatic protons in a given compound were fitted simultaneously by the OriginPro 8 software. During the fitting, the limiting chemical shifts of dichloroacetic acid were kept constant at δHI = 6.341 and δI = 6.051. These limiting chemical shift values were determined by Szakacs et al. [33] at 0.15 M ionic strength. The low pK_a values of the phosphonic acid derivatives were calculated from the ΔpK values and the pK_a of dichloroacetic acid (pK_a = 1.14).

2.4 HRMS measurements

HRMS analyses were performed on a LTQ FT Ultra (Thermo Fisher Scientific, Bremen, Germany) system. The ionization method was ESI and operated in positive ion mode. The protonated molecular ion peaks were fragmented by CID at a normalized collision energy of 45%. The samples were solved in methanol-water (1:1, v/v) acidified with 1 V/V% cc. AcOH. Data acquisition and analysis were accomplished with Xcalibur software version 2.0 (Thermo Fisher Scientific).

3 Results and discussion

To further investigate the substituent effects on the pK_a values of arylphosphonic acids, beside the formerly synthesized molecules (***1a-g*** and ***1o-s***), we synthesized 10 more arylphosphonic acid derivatives (***1h-j*** and ***1t-z***) using the method published before [27], and carried out structural identification based on NMR and HRMS measurements (See the Supplementary material). We also purchased further 4 derivatives (***1k-n***) from Sigma Aldrich Co.. Altogether we carried out the pK_a measurements of 26 compounds (Table 1.) using up-to-date methods and determined accurate pK_a values of the proton dissociation equilibria of arylphosphonic acid derivatives (Fig. 2.). The pK_a values of 5 compounds, ***1t-x*** have not been reported before.

pK_{a2} values of the second dissociation step of arylphosphonic acid derivatives and p$K_{a,COOH}$ values of carboxyl groups were measured by differential pH-metric titrations. In the case of pK_{a1} values (the first dissociation step of arylphosphonic acid) we used NMR-pH titration technique since most pK_a values were outside of the operational pH range of the pH-metric method (pH 1.8-12.0).[34] (Table A.1-A.2, see Appendices)

The measured pK_a values were compared with those reported by Nagarajan et al. [26] and also with calculated pK_a values (cpK$_a$) predicted by the Chemaxon/Marvin Sketch 16.10.17 plugin [35], as well as the ACD/Percepta pK_a predictor module of Advanced Chemistry Development, Inc. (ACD/Labs) using the Classic and GALAS models [36]. (Table A.1-A.2, see Appendices) Finally, the new Hammett equations were determined based on the in-house results.

3.1 Comparison of the results measured by us with those of Nagarajan et al.

We compared our experimental pK_a values with those measured by Nagarajan et al. [26] using linear regression analysis and the mean absolute error (Table 2.). Altogether 16 compounds were involved in this comparison as shown by Table A.1-A.2.

These data show that the correlation in the case of the first proton dissociation step (pK_{a1}) is slightly better with an acceptable MAE, while in the case of the second proton dissociation step of arylphosphonic acids (pK_{a2}) all three values indicate a worse correlation and a significant MAE, exceeding the margin of error (0.59 > 0.50 [37–39]). This might be caused by the following potential sources of measurement error Nagarajan's experimental pK_a values suffer from:

a) they used a less accurate, now obsolete direct potentiometric method,
b) most of the reported pK_a values are out of the standardization range of the pH meter they used (pH 2.0 to 7.0),
c) they corrected their values to zero ionic strength, while modern commercial potentiometric pK_a analysers work based on the constant ionic medium reference state method, using a 0.15 M KCl background electrolyte to improve the measurement's precision and to mimic the physiological salt level [40].

3.2 Comparison of predicted and experimental pK_a values

As can be seen in Table 3. and Figure A.1. (see Appendices), a good correlation was found for the full in-house data set between measured and *in silico* predicted pK_a values when using the ACD/Percepta pK_a predictor with the Classic model (pK_{a1}: R^2=0.851; pK_{a2}: R^2=0.819), while poor correlations were found in case of the other two predictors. The better results of the former pK_a predictor might be caused by the fact that it uses Hammett-type equations and electronic substituent constants (σ) when calculating the pK_a values. The other two predictors using microconstants for predicting the pK_a values seem to have problems with predicting accurate values for the set of compounds we studied.

In the next step the predicted and experimental pK_a values were compared for the overlapping points in Nagarajan's data set and our data, i.e. altogether 16 compounds were involved in this comparison. In general, better correlations were found for pK_{a1} values and for Nagarajan's data set, however, in two cases the MAE calculated for this data set exceeds the acceptable value (>0.50 [37–39]) (Table 4., Fig. A.2., see Appendices). It is also notable, that in the case of the ACD/Percepta pK_a predictor using the Classic model, the MAE values are significantly better for Nagarajan's set, suggesting that the predictor's database contains similar pK_a values and Hammett equations to those reported by Nagarajan et al. The relatively high MAE of prediction in the case of the in-house pK_{a2} values might also be a consequence that this predictor's database contains inaccurate pK_a values and Hammett equations.

3.3 Hammett sensitivity constants

To determine the sensitivity constant (ρ) for arylphosphonic acid derivatives Eq. 2. was rearranged to the following form:

$$pK_{a,0} - pK_{a,S} = \rho \sum \sigma_S \qquad \text{Eq. (5.)}$$

Then the sensitivity constant ρ can be determined by plotting the difference of pK_a values of the parent compound (***1a***) and its *meta* and *para* substituted derivatives (***1c-e, 1g-r, 1t-z***) against the sum of Hammett σ values of the $\mathbf{R^2}$ to $\mathbf{R^4}$ substituents of (***1***), obtained from lookup tables in reference [41] and fitting a line onto the data points using linear regression analysis with no intercept (regression through the origin). The sensitivity constant is given by the slope of the regression line.

The *ortho* substituted derivatives (***1b, 1f, 1s***) should be omitted from the calculation because in the case of *ortho* substituents steric effects also occur, and therefore Eq. (2) does not describe such compounds.

There are further two derivatives that need special attention: they are compounds (***1y***) and (***1z***) containing a carboxyl group in the *para* and *meta* position of (***1***), respectively. Table A.2. shows that the differences p$K_{a,COOH}$ - pK_{a1} and p$K_{a,COOH}$ - pK_{a2} (where p$K_{a,\,COOH}$ describes the proton dissociation of the carboxyl group of compounds ***1y*** and ***1z***) are in the range of 2.2 to 3 pK_a units. This means that in the pH range used in the course of determining the experimental values of pK_{a1} and pK_{a2} both the unionized COOH and ionized COO^- form of the carboxyl group of compounds (***1y***) and (***1z***) exist. Of course, it can be stated that in the pH range used when determining pK_{a1} the unionized COOH form will dominate, whereas in the pH range used when determining pK_{a2} the ionized COO^- form will dominate.

It follows that compounds (***1y***) and (***1z***) should be characterized by $\sigma_{p,COOH}$ and $\sigma_{m,COOH}$, respectively, in the correlation of (p$K_{a1,0}$ - p$K_{a1,S}$) with $\sum\sigma_S$ and by σ_{p,COO^-} and σ_{m,COO^-}, respectively, in the correlation of (p$K_{a2,0}$ - p$K_{a2,S}$) with $\sum\sigma_S$.

The results of regression analyses performed for altogether 22 data points on the basis of Eq. (5) and the above considerations are shown in Table 5. It can be seen that much better correlation was found for pK_{a2} than pK_{a1} (R^2=0.917 and 0.862, respectively). An analysis of the residuals (differences between experimental and predicted ΔpK_{a1} values) revealed that the worse correlation is due to compound (***1z***) being an outlier in the linear regression. This is probably caused by the fact that this compound was characterized with the substituent constant $\sigma_{m,COOH}$ as if only the unionized form of its carboxyl group were present, although we know that the ionized carboxylate form also exists in the pH range used when determining the pK_{a1} value of this compound. This may be the case to a lesser extent with compound (***1y***) for pK_{a1} and with compounds (***1y***) and (***1z***) for pK_{a2}. Therefore, the above regression analysis was performed again without compounds (***1y***) and (***1z***), i.e. for altogether 20 data points. The results are shown in Table 5 and Figure 3. These regression results are considered the final ones.

Table 5 shows that the sensitivity constants (ρ) determined for pK_{a1} and pK_{a2} values are 0.894 and 0.934, respectively. Using the new ρ values, the Hammett equations describing the first and second dissociation step of arylphosphonic acids are given in Table 5.

Comparing the new equations with the formerly reported ones, we can observe that our ρ values are similar to those determined by Nagarajan's group. However, there are significant differences in the parent compound's pK_a values (the intercept of the Hammett equation), which might cause substantial errors during pK_a predictions, especially in the case of pK_{a2} values (Eq. 7. vs. Eq. 9.), where the difference exceeds 0.5 pK_a units. It is even more unfortunate, because the knowledge of accurate pK_{a2} values is essential since they are in the physiologically significant pH range. The standard errors of measured p$K_{a,0}$ values (SD < 0.005) have no considerable effect on the goodness of prediction of pK_{a1} and pK_{a2} values (Eq. 8-9.) compared to the standard errors of Hammett ρ parameters.

4 Conclusions

In this study, we carried out a thorough analysis of pK_a values of 26 different arylphosphonic acid derivatives of which 5 compounds' pK_a values were not reported before. Comparing our experimental pK_a values to predicted ones we found a good correlation with predicted pK_a values calculated by the Hammett-type equation based ACD/Percepta pK_a predictor module using the Classic model, while the microconstant based predictors proved to be less accurate. Comparing our results to experimental values previously reported by Nagarajan et al. we found significant divergence, especially in case of pK_{a2}. It might stem from the different accuracy of the methods used by Nagarajan (direct potentiometry, zero ionic strength) and the up-to-date methods used in this study (differential potentiometry, NMR-titration, constant ionic strength). Based on our results we refined the Hammett equation for arylphosphonic acids, the new parameters, together with our measured pK_a values might be used for more accurate pK_a prediction of arylphosphonic acid derivatives. The refinement of the key parameters of pK_a prediction provides a reliable physico-chemical characterization of novel drug-like molecules, and refreshes pK_a databases, which can lead to more efficient early-stage drug discovery processes.

5 Acknowledgement

Gergő Dargó thanks the Gedeon Richter Talentum Foundation for the financial support. Szabolcs Béni gratefully acknowledges the financial support from NKFIH PD109373 and from ÚNKP-16-4 New National Excellence Program of the Ministry of Human Capacities. The authors thank ACD/Labs for providing a free evaluation version of the ACD/Percepta software.

6 References

[1] L.D. Quin, A Guide to Organophosphorus Chemistry, Wiley, 2000.

[2] F. Gasparini, W. Inderbitzin, E. Francotte, G. Lecis, P. Richert, Z. Dragic, R. Kuhn, P.J. Flor, (+)-4-Phosphonophenylglycine (PPG) a new group III selective metabotropic glutamate receptor agonist, Bioorg. Med. Chem. Lett. 10 (2000) 1241–1244. doi:10.1016/S0960-894X(00)00197-9.

[3] H. Bräuner-Osborne, J. Egebjerg, E. Nielsen, U. Madsen, P. Krogsgaard-Larsen, Ligands for glutamate receptors: Design and therapeutic prospects, J. Med. Chem. 43 (2000) 2609–2645. doi:10.1021/jm000007r.

[4] S.J. Conway, J.C. Miller, P.A. Howson, B.P. Clark, D.E. Jane, Synthesis of phenylglycine derivatives as potent and selective antagonists of group III metabotropic glutamate receptors, Bioorganic Med. Chem. Lett. 11 (2001) 777–780. doi:10.1016/S0960-894X(01)00052-X.

[5] I. Collado, C. Pedregal, A.B. Bueno, A. Marcos, R. González, J. Blanco-Urgoiti, J. Pérez-Castells, D.D. Schoepp, R. a Wright, B.G. Johnson, A.E. Kingston, E.D. Moher, D.W. Hoard, K.I. Griffey, J.P. Tizzano, (2S,1'S,2'R,3'R)-2-(2'-Carboxy-3'-hydroxymethylcyclopropyl) glycine is a highly potent group 2 and 3 metabotropic glutamate receptor agonist with oral activity., J. Med. Chem. 47 (2004) 456–66. doi:10.1021/jm030967o.

[6] N. Kawahata, M.G. Yang, G.P. Luke, W.C. Shakespeare, R. Sundaramoorthi, Y. Wang, D. Johnson, T. Merry, S. Violette, W. Guan, C. Bartlett, J. Smith, M. Hatada, X. Lu, D.C. Dalgarno, C.J. Eyermann, R.S. Bohacek, T.K. Sawyer, A Novel Phosphotyrosine Mimetic 4'-Carboxymethyloxy-3'-phosphonophenylalanine (Cpp): Exploitation in the Design of

Nonpeptide Inhibitors of pp60Src SH2 domain, Bioorganic Med. Chem. Lett. 11 (2001) 2319–2323. doi:10.1016/S0960-894X(01)00446-2.

[7] C.K. Lau, C.I. Bayly, J.Y. Gauthier, C.S. Li, M. Therien, E. Asante-Appiah, W. Cromlish, Y. Boie, F. Forghani, S. Desmarais, Q. Wang, K. Skorey, D. Waddleton, P. Payette, C. Ramachandran, B.P. Kennedy, G. Scapin, Structure based design of a series of potent and selective non peptidic PTP-1B inhibitors, Bioorganic Med. Chem. Lett. 14 (2004) 1043–1048. doi:10.1016/j.bmcl.2003.11.076.

[8] J.Y. Winum, A. Innocenti, V. Gagnard, J.L. Montero, A. Scozzafava, D. Vullo, C.T. Supuran, Carbonic anhydrase inhibitors. Interaction of isozymes I, II, IV, V, and IX with organic phosphates and phosphonates, Bioorganic Med. Chem. Lett. 15 (2005) 1683–1686. doi:10.1016/j.bmcl.2005.01.049.

[9] S. Oishi, R.G. Karki, S.U. Kang, X. Wang, K.M. Worthy, L.K. Bindu, M.C. Nicklaus, R.J. Fisher, T.R. Burke, Design and synthesis of conformationally constrained Grb2 SH2 domain binding peptides employing α-methylphenylalanyl based phosphotyrosyl mimetics, J. Med. Chem. 48 (2005) 764–772. doi:10.1021/jm0492709.

[10] P. Lassaux, M. Hamel, M. Gulea, H. Delbrück, P.S. Mercuri, L. Horsfall, D. Dehareng, M. Kupper, J.M. Frére, K. Hoffmann, M. Galleni, C. Bebrone, Mercaptophosphonate compounds as broad-spectrum inhibitors of the metallo-ß-lactamases, J. Med. Chem. 53 (2010) 4862–4876. doi:10.1021/jm100213c.

[11] X. Chen, D.J. Kopecky, J. Mihalic, S. Je, X. Min, J. Heath, J. Deignan, S. Lai, Z. Fu, C. Guimaraes, S. Shen, S. Li, S. Johnstone, S. Thibault, H. Xu, M. Cardozo, W. Shen, N. Walker, F. Kayser, Z. Wang, Structure-Guided Design, Synthesis, and Evaluation of Guanine- Derived Inhibitors of the eIF4E mRNA − Cap Interaction, (2012).

[12] D.T. Manallack, R.J. Prankerd, G.C. Nassta, O. Ursu, T.I. Oprea, D.K. Chalmers, A Chemogenomic Analysis of Ionization Constants-Implications for Drug Discovery, ChemMedChem. 8 (2013) 242–255. doi:10.1002/cmdc.201200507.

[13] M.P. Gleeson, Generation of a Set of Simple, Interpretable ADMET Rules of Thumb, J. Med. Chem. 51 (2008) 817–834. doi:10.1021/jm701122q.

[14] H. Wan, J. Ulander, High-throughput pKa screening and prediction amenable for ADME profiling, Expert Opin. Drug Metab. Toxicol. 2 (2006) 139–155. doi:10.1517/17425255.2.1.139.

[15] A.C. Lee, G.M. Crippen, Predicting pKa, J. Chem. Inf. Model. 49 (2009) 2013–2033. doi:10.1021/ci900209w.

[16] M. Rupp, R. Korner, I. V. Tetko, Predicting the pKa of Small Molecules, Comb. Chem. High Throughput Screen. 14 (2011) 307–327. doi:10.2174/138620711795508403.

[17] C. Liao, M.C. Nicklaus, Comparison of Nine Programs Predicting pKa Values of Pharmaceutical Substances, J. Chem. Inf. Model. 49 (2009) 2801–2812. doi:10.1021/ci900289x.

[18] L.P. Hammett, Some Relations between Reaction Rates and Equilibrium Constants., Chem. Rev. (1935) 125–136. doi:10.1021/cr60056a010.

[19] L.P. Hammett, The Effect of Structure upon the Reactions of Organic Compounds. Benzene Derivatives, J. Am. Chem. Soc. 59 (1937) 96–103. doi:10.1021/ja01280a022.

[20] R.W. Taft, Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters, J. Am. Chem. Soc. 74 (1952) 3120–3128. doi:10.1021/ja01132a049.

[21] R.W. Taft, Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters, J. Am. Chem. Soc. 74 (1952) 2729–2732. doi:10.1021/ja01131a010.

[22] R.W. Taft, Linear Steric Energy Relationships, J. Am. Chem. Soc. 75 (1953) 4538–4539. doi:10.1021/ja01114a044.

[23] H.H. Jaffé, L.D. Freedman, G.O. Doak, The Acid Dissociation Constants of Aromatic Phosphonic Acids. I. Meta and Para Substituted Compounds, J. Am. Chem. Soc. 75 (1953) 2209–2211. doi:10.1021/ja01105a054.

[24] H.H. Jaffé, L.D. Freedman, G.O. Doak, The Acid Dissociation Constants of Aromatic Phosphonic Acids. I. Meta and Para Substituted Compounds - Correction, J. Am. Chem. Soc. 75 (1953) 6361–6361. doi:10.1021/ja01120a637.

[25] H.H. Jaffé, L.D. Freedman, G.O. Doak, The Acid Dissociation Constants of Aromatic Phosphonic Acids. II. Compounds with Ortho Substituents, J. Am. Chem. Soc. 76 (1954) 1548–1552. doi:10.1021/ja01635a022.

[26] K. Nagarajan, K.P. Shelly, R.R. Perkins, R. Stewart, Arylphosphonic acids. I. Substituent effects on their first and second dissociations, Can. J. Chem. 65 (1987) 1729–1733. doi:10.1139/v87-290.

[27] G. Keglevich, A. Grün, A. Bölcskei, L. Drahos, M. Kraszni, G.T. Balogh, Synthesis and Proton Dissociation Properties of Arylphosphonates: A Microwave-Assisted Catalytic Arbuzov Reaction with Aryl Bromides, Heteroat. Chem. 23 (2012) 574–582. doi:10.1002/hc.21053.

[28] A. Avdeef, pH-metric log P. II: Refinement of Partition Coefficients and Ionization Constants of Multiprotic Substances, J. Pharm. Sci. 82 (1993) 183–190. doi:10.1002/jps.2600820214.

[29] R.I. Allen, K.J. Box, J.E.A. Comer, C. Peake, K.Y. Tam, Multiwavelength spectrophotometric determination of acid dissociation constants of ionizable drugs, J. Pharm. Biomed. Anal. 17 (1998) 699–712. doi:10.1016/S0731-7085(98)00010-7.

[30] A. Avdeef, J.E.A. Comer, S.J. Thomson, pH-Metric log P. 3. Glass electrode calibration in methanol-water, applied to pKa determination of water-insoluble substances, Anal. Chem. 65 (1993) 42–49. doi:10.1021/ac00049a010.

[31] K. Takács-Novák, K.J. Box, A. Avdeef, Potentiometric pKa determination of water-insoluble compounds: validation study in methanol/water mixtures, Int. J. Pharm. 151 (1997) 235–248. doi:10.1016/S0378-5173(97)04907-7.

[32] C.L. Perrin, M.A. Fabian, Multicomponent NMR titration for simultaneous measurement of relative pKaS, Anal. Chem. 68 (1996) 2127–34. doi:10.1021/ac960117z.

[33] Z. Szakács, S. Béni, Z. Varga, L. Örfi, G. Kéri, B. Noszál, Acid-base profiling of imatinib (Gleevec) and its fragments, J. Med. Chem. 48 (2005) 249–255. doi:10.1021/jm049546c.

[34] Z. Szakács, G. Hägele, Accurate determination of low pK values by 1H NMR titration, Talanta. 62 (2004) 819–825. doi:10.1016/j.talanta.2003.10.007.

[35] Calculator Plugins were used for structure property prediction and calculation, Marvin 16.2.15, 2016, ChemAxon, https://www.chemaxon.com/products/calculator-plugins/property-predictors/.

[36] ACD/Percepta, v2015, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2015.

[37] J. Manchester, G. Walkup, O. Rivin, Z. You, Evaluation of pKa Estimation Methods on 211 Druglike Compounds, J. Chem. Inf. Model. 50 (2010) 565–571. doi:10.1021/ci100019p.

[38] J.C. Shelley, D. Calkins, A.P. Sullivan, Comments on the Article “Evaluation of pKa Estimation Methods on 211 Druglike Compounds,” J. Chem. Inf. Model. 51 (2011) 102–104. doi:10.1021/ci100332m.

[39] G.T. Balogh, Á. Tarcsay, G.M. Keserű, Comparative evaluation of pKa prediction tools on a drug discovery dataset, J. Pharm. Biomed. Anal. 67–68 (2012) 63–70. doi:10.1016/j.jpba.2012.04.021.

[40] A. Avdeef, Absorption and Drug Development : Solubility, Permeability, and Charge State, John Wiley & Sons, 2012.

[41] A. Williams, Free Energy Relationships in Organic and Bio-Organic Chemistry, Royal Society of Chemistry, Cambridge, 2003. doi:10.1039/9781847550927.

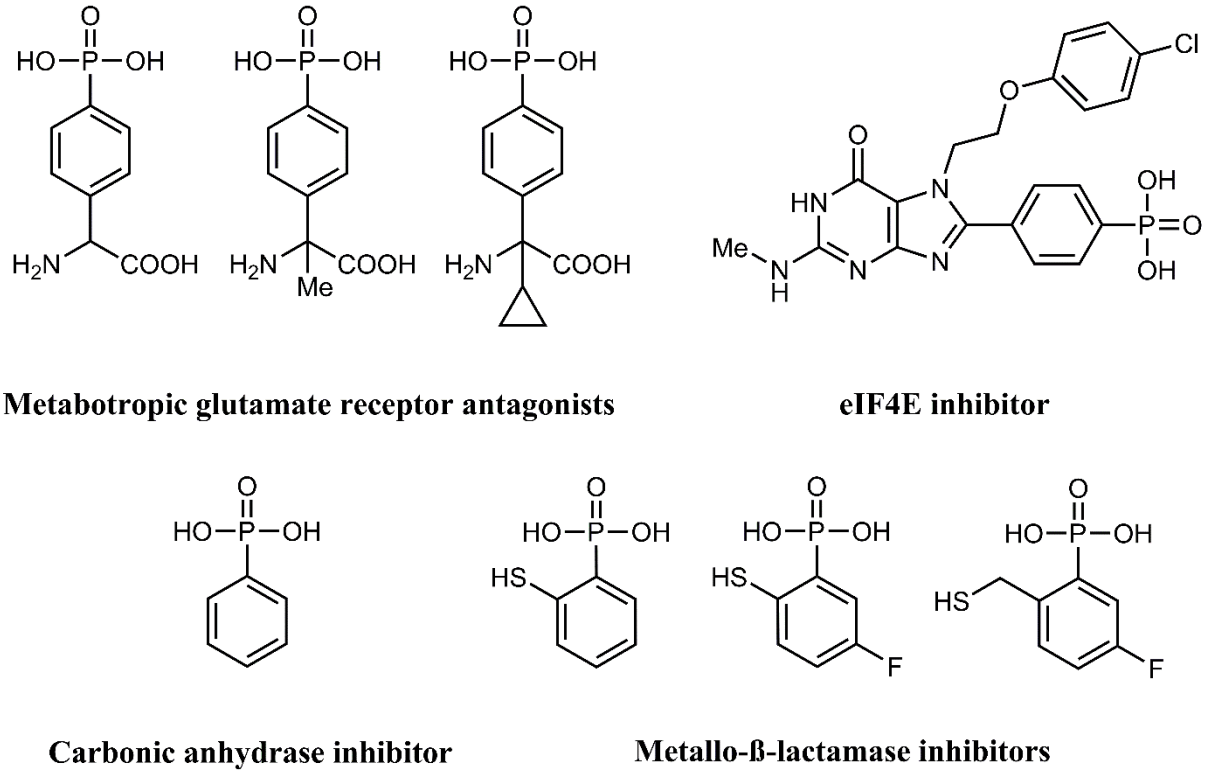


Figure 1. Examples of biologically active arylphosphonic acid derivatives

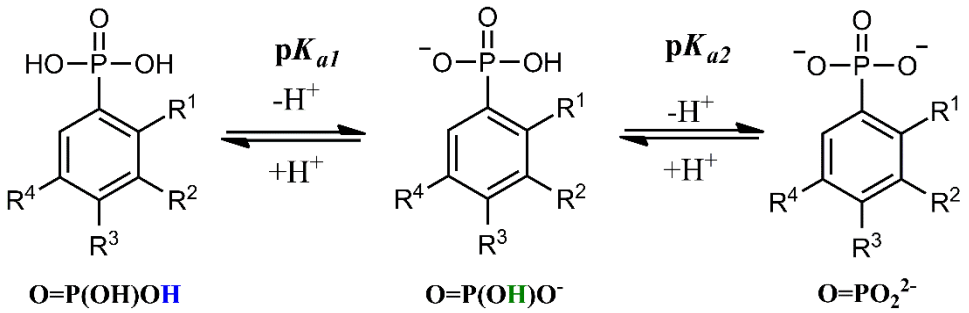


Figure 2. Proton dissociation steps of arylphosphonic acid derivatives (*1*)

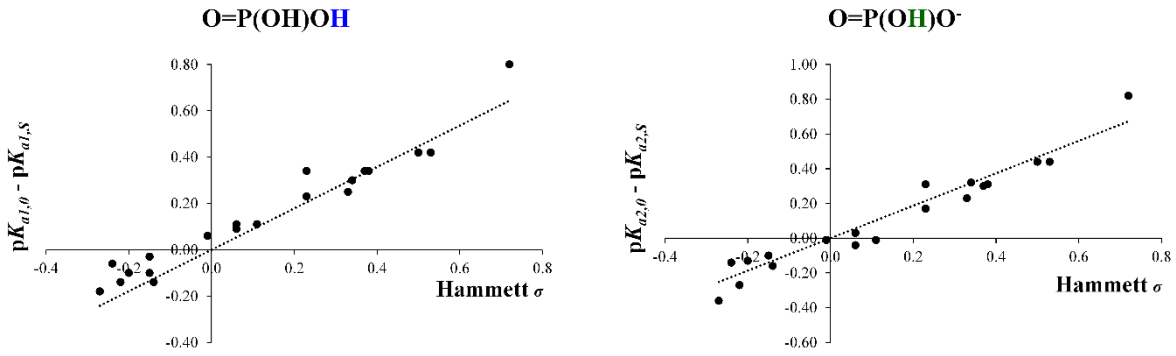


Figure 3. Determination of Hammett sensitivity constants (n=20 data points, *1c-e, 1g-r, 1t-x*)

Table 1. Substituents of the investigated arylphosphonic acids (*1a-z*)

1	**R^1**	**R^2**	**R^3**	**R^4**
a	H	H	H	H
b	Me	H	H	H
c	H	H	Me	H
d	H	H	OCF_3	H
e	H	H	OMe	H
f	F	H	H	H
g	H	F	H	H
h	H	OMe	H	H
i	H	H	Cl	H
j	H	Cl	H	H
k	H	H	Br	H
l	H	NO_2	H	H
m	H	H	Et	H
n	H	H	EtO	H
o	H	H	$C(O)CH_3$	H
p	H	H	CF_3	H
q	H	H	cyclohexyl	H
r	H	H	F	H
s	F	H	H	F
t	H	H	Ph	H
u	H	Ph	H	H
v	H	H	nPr	H
w	H	$C(O)CH_3$	H	H
x	H	H	tBu	H
y	H	H	COOH	H
z	H	COOH	H	H

Table 2. Results of comparison of in-house and formerly reported pK_a values (n=16 compounds, *1a-n, 1y-z*, see Table A.1-A.2)

Parameters*	pK_{a1}	pK_{a2}
R^2	0.899	0.842
s	0.087	0.123
MAE	0.16	0.59

*R^2 and **s** are the squared correlation coefficient and the standard error of estimate, respectively, of the linear regression p$K_{a,\ in\text{-}house}$ = a * p$K_{a,\ Nagarajan}$ + b, and **MAE** is the mean absolute error calculated as follows: MAE= (1/n) * $\Sigma|$p$K_{a,\ in\text{-}house}$ – p$K_{a,\ Nagarajan}|$

Table 3. Comparison of predicted and measured pK_a values for the full in-house set (n= 26 compounds, *1a-z*, see Table A.1-A.2) [*]

In-house pK_{a1}	ACD Classic	ACD GALAS	Marvin Sketch
R^2	0.851	0.746	0.603
s	0.103	0.134	0.168
MAE	0.16	0.21	0.24

In-house pK_{a2}	ACD Classic	ACD GALAS	Marvin Sketch
R^2	0.819	0.500[#]	0.573
s	0.127	0.211	0.195
MAE	0.46	0.22	0.20

[*] **R^2** and **s** are the squared correlation coefficient and the standard error of estimate, respectively, of the linear regression $pK_{a,\,experimental} = a * pK_{a,\,predicted} + b$, and **MAE** is the mean absolute error calculated as follows: $MAE = (1/n) * \Sigma|pK_{a,\,experimental} - pK_{a,\,predicted}|$.

[#] The particularly low correlation is due to the outlying points belonging to compounds ***1y-z,*** containing carboxyl functional group. By excluding them the correlation improves significantly (R^2=0.756, n=24)

Table 4. Comparison of predicted and measured pK_a values for the overlapping sets of experimental data (n=16 compounds, *1a-n, 1y-z*, see Table A.1-A.2) [*]

pK_{a1}	ACD Classic		ACD GALAS		Marvin Sketch	
	In-house	Nagarajan	In-house	Nagarajan	In-house	Nagarajan
R^2	0.768	0.885	0.752	0.831	0.658	0.803
s	0.132	0.091	0.137	0.110	0.160	0.119
MAE	0.17	0.04	0.23	0.12	0.17	0.21

pK_{a2}	ACD Classic		ACD GALAS		Marvin Sketch	
	In-house	Nagarajan	In-house	Nagarajan	In-house	Nagarajan
R^2	0.746	0.696	0.379[#]	0.661	0.607	0.677
s	0.156	0.186	0.244	0.196	0.194	0.192
MAE	0.47	0.16	0.21	0.67	0.18	0.59

[*] **R^2** and **s** are the squared correlation coefficient and the standard error of estimate, respectively, of the linear regression $pK_{a,\,experimental} = a * pK_{a,\,predicted} + b$, and **MAE** is the mean absolute error calculated as follows: $MAE = (1/n) * \Sigma|pK_{a,\,experimental} - pK_{a,\,predicted}|$.

[#] The particularly low correlation is due to the outlying points belonging to compounds ***1y-z,*** containing carboxyl functional group. By excluding them the correlation improves significantly (R^2=0.850, n=14)

Table 5. Results of regression analyses based on Eq. (5)[*]

Dependent variable	**($pK_{a1,0}$ - $pK_{a1,S}$)**	**($pK_{a2,0}$ - $pK_{a2,S}$)**
No. of data points	22	22
R^2	0.862	0.917
s	0.090	0.080
ρ	0.837(0.060)	0.928(0.057)
No. of data points	20	20
R^2	0.911	0.938
s	0.075	0.072
ρ	0.894(0.054)	0.934(0.052)

Hammett equation based on Nagarajan et al.'s work[#]		
$pK_{a1} = 1.84 - 0.856 \sum\sigma_S$	**R^2** = 0.975	Eq. (6)
$pK_{a2} = 7.48 - 0.980 \sum\sigma_S$	**R^2** = 0.956	Eq. (7)
Hammett equation based on this work		
$pK_{a1} = 1.70 - 0.894(0.054) \sum\sigma_S$	**R^2** = 0.911	Eq. (8)
$pK_{a2} = 6.92 - 0.934(0.052) \sum\sigma_S$	**R^2** = 0.938	Eq. (9)

[*] **R^2** and **s** are the squared correlation coefficient and the standard error of estimate, respectively, and the values in parentheses are the standard errors of the *ρ* sensitivity constants

[#] The corresponding equations reported in Nagarajan et al.'s work appear to be incorrect, i.e. the sign of the intercept in the original publication should be negative.

Appendices

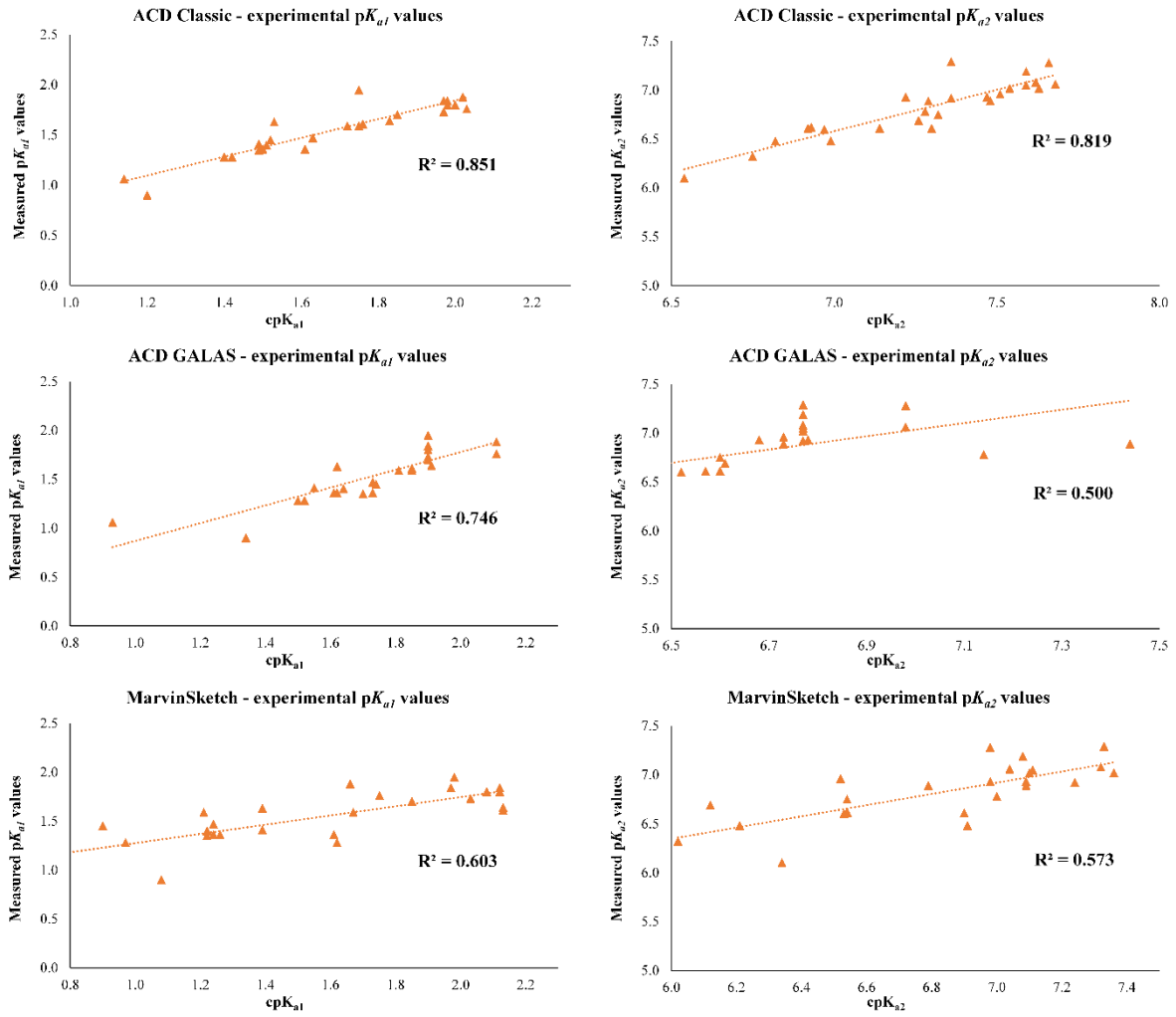


Figure A.1. Correlation between predicted and measured pK_a values for the full in-house dataset (n=26 compounds, *1a-z*)

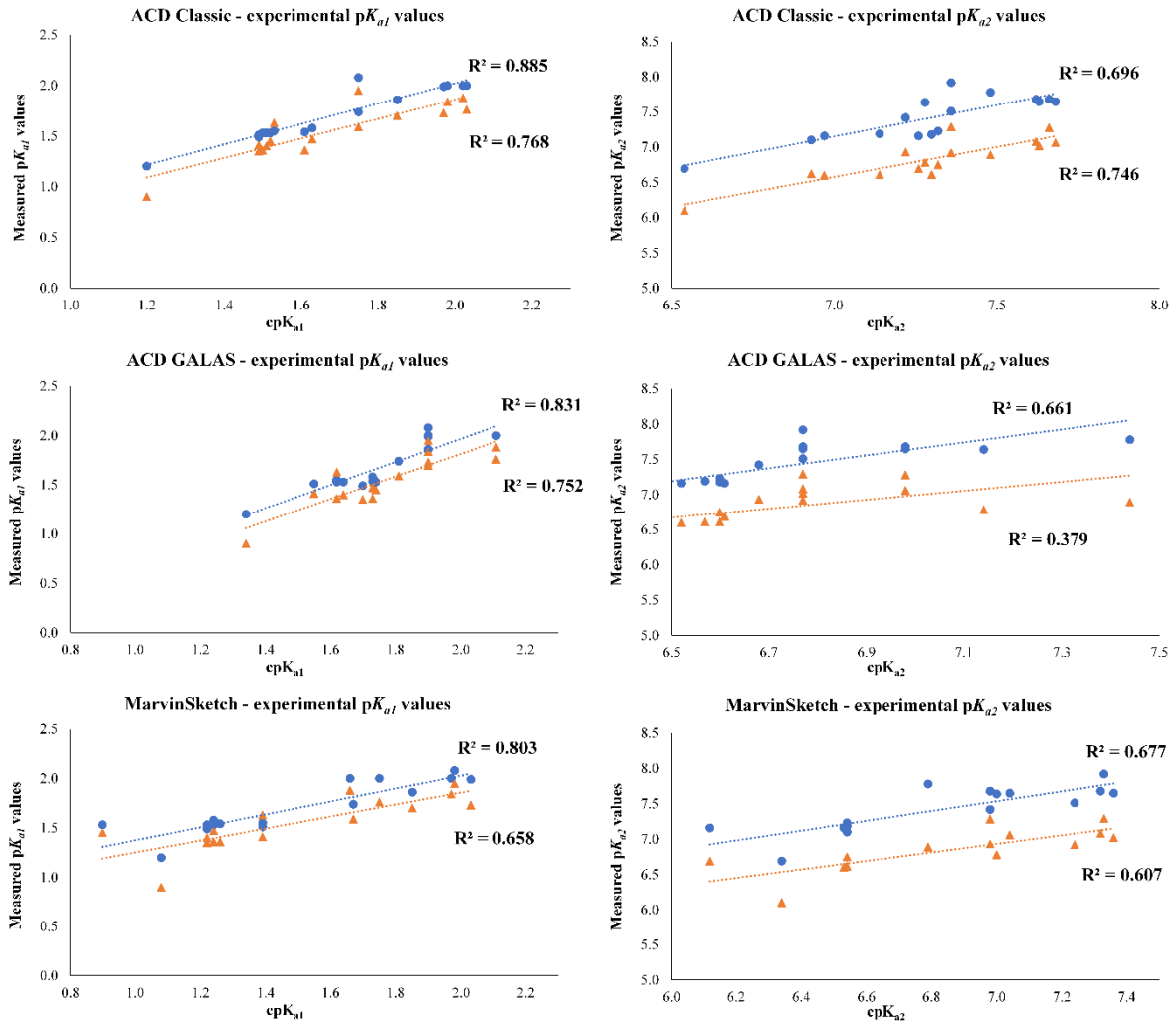


Figure A.2 Correlation between predicted and measured pK_a values of the overlapping sets (n=16 compounds, *1a-n, 1y-z*; blue circles: pK_a values reported by Nagarajan et al., orange triangles: in-house measured pK_a values)

1 **Table A.1. Measured and predicted pK_a values of arylphosphonic acid derivatives (*1a-x*)**

1	Hammett σ [41]	ACD Classic		ACD GALAS		Marvin Sketch		Nagarajan et al.		In-house NMR-pH titration		In-house pH-metric titration		In-house vs. Nagarajan's values	
		cpK_{a1}	cpK_{a2}	cpK_{a1}	cpK_{a2}	cpK_{a1}	cpK_{a2}	pK_{a1}	pK_{a2}	pK_{a1}	SD	pK_{a2}	SD	$\lvert\Delta pK_{a1}\rvert$	$\lvert\Delta pK_{a2}\rvert$
a	0	1.85	7.36	1.90	6.77	1.85	7.24	1.86	7.51	1.70	0.00	6.92	0.00	0.16	0.59
b	-	1.75	7.36	1.90	6.77	1.98	7.33	2.08	7.92	1.95	0.00	7.29	0.00	0.13	0.63
c	-0.14	1.98	7.62	1.90	6.77	1.97	7.32	2.00	7.68	1.84	0.00	7.08	0.00	0.16	0.60
d	0.33	1.52	7.26	1.74	6.61	0.90	6.12	1.53	7.16	1.45	0.00	6.69	0.01	0.08	0.47
e	-0.27	2.02	7.66	2.11	6.98	1.66	6.98	2.00	7.68	1.88	0.00	7.28	0.00	0.12	0.40
f	-	1.49	7.14	1.70	6.57	1.22	6.54	1.49	7.19	1.35	0.00	6.61	0.00	0.14	0.58
g	0.34	1.51	6.97	1.64	6.52	1.22	6.53	1.53	7.16	1.40	0.00	6.60	0.00	0.13	0.56
h	0.11	1.75	7.22	1.81	6.68	1.67	6.98	1.74	7.42	1.59	0.01	6.93	0.00	0.15	0.49
i	0.23	1.63	7.32	1.73	6.60	1.24	6.54	1.58	7.23	1.47	0.01	6.75*	0.01	0.11	0.48
j	0.37	1.50	6.93	1.62	6.49	1.24	6.54	1.53	7.10	1.36	0.02	6.62*	0.01	0.17	0.48
k	0.23	1.61	7.30	1.73	6.60	1.26	6.54	1.54	7.18	1.36	0.01	6.61	0.00	0.18	0.57
l	0.72	1.20	6.54	1.34	6.21	1.08	6.34	1.20	6.69	0.90	0.02	6.10	0.01	0.30	0.59
m	-0.15	1.97	7.63	1.90	6.77	2.03	7.36	1.99	7.65	1.73	0.00	7.02	0.01	0.26	0.63

n	-0.24	2.03	7.68	2.11	6.98	1.75	7.04	2.00	7.65	1.76	0.01	7.06	0.00	0.24	0.59
o	0.50	1.42	6.82	1.52	6.39	1.62	6.91	-	-	1.28	0.00	6.48	0.00	-	-
p	0.53	1.40	6.99	1.50	6.37	0.97	6.21	-	-	1.28	0.01	6.48	0.00	-	-
q	-0.22	1.97	7.59	1.90	6.77	2.12	7.08	-	-	1.84	0.00	7.19*	0.01	-	-
r	0.06	1.72	7.51	1.85	6.73	1.21	6.52	-	-	1.59	0.00	6.96	0.00	-	-
s	-	1.14	6.75	0.93	5.81	0.76	6.02	-	-	1.06	0.00	6.32	0.01	-	-
t	-0.01	1.83	7.47	1.91	6.78	2.13	7.09	-	-	1.64	0.01	6.93	0.05	-	-
u	0.06	1.76	7.29	1.85	6.73	2.13	7.09	-	-	1.61	0.01	6.89	0.01	-	-
v	-0.15	1.98	7.54	1.90	6.77	2.08	7.10	-	-	1.80	0.01	7.02	0.01	-	-
w	0.38	1.50	6.92	1.61	6.49	1.61	6.90	-	-	1.36	0.01	6.61	0.00	-	-
x	-0.20	2.00	7.59	1.90	6.77	2.12	7.11	-	-	1.80	0.00	7.05*	0.01	-	-

*determined from cosolvent p_sK_a values measured in MeOH-water mixtures

Table A.2. Measured and predicted pK_a values of arylphosphonic acid derivatives containing carboxyl functional groups (*1y-z*)

1	Hammett *σ* [41]	ACD Classic			ACD GALAS			Marvin			Nagarajan et al.		NMR-pH titration		Differential pH-metric titration				In-house vs. Nagarajan's values	
		cp K_{a1}	cp K_a coo H	cp K_{a2}	cp K_{a1}	cp K_a coo H	cp K_{a2}	cp K_{a1}	cp K_a coo H	cp K_{a2}	pK_{a1}	pK_{a2}	pK_{a1}	SD	pK_a CO OH	SD	pK_{a2}	SD	|ΔpK_{a1}|	|ΔpK_{a2}|
y	0.44*	1.49	3.94	7.28	1.55	4.38	7.14	1.39	3.95	7.00	1.51	7.64	1.41	0.02	3.82	0.00	6.78	0.00	0.10	0.86
z	0.35#	1.53	4.00	7.48	1.62	4.67	7.44	1.39	3.96	6.79	1.55	7.78	1.63	0.05	3.88	0.00	6.89	0.00	0.08	0.89

*This Hammett ***σ*** value refers to the unionized carboxyl group of compound (***1y***), while the ***σ*** value referring to the ionized carboxylate group is 0.00.

#This Hammett ***σ*** value refers to the unionized carboxyl group of compound (***1z***), while the ***σ*** value referring to the ionized carboxylate group is -0.10.