Introduction

The success of the virtual screening is fundamentally influenced by the optimal tautomer form and ionization state that underlines the importance of accurate \(pK_a\) prediction for even large libraries. Moreover, ADME and \(P_k\) optimization supported by in silico methods also requires information on the ionization state, therefore, fast and accurate prediction of \(pK_a\) values is in high demand. Testing commercially available in silico tools using unbiased compound set provides information about overall performance and uncovers typical errors that might need further attention. Our test set involved ~ 200 experimental \(pK_a\) values of ~100 compounds that were predicted by different \(pK_a\) predictors including Marvin, ACD, Epik, Pallas and Pharma Algorithm (ADMEBox). Statistical analysis and specific examples will be discussed evaluating the predictive power of different commercially available \(pK_a\) predictors.

Materials and Methods

Potentiometric \(pK_a\) measurement

GlpKa automated \(pK_a\) analyzer (Sirius Analytical Instrument Ltd., Forest Row, UK) fitted with combination Ag/AgCl pH electrode was used for determination of dissociation constants of Gedeon Richter Plc.’s in-house compounds.

Potentiometric titration in aqueous medium. In general, 10.00 mL of a 1 mM aqueous solution of sample was pre-acidified to pH 2.0 with 0.5 M HCl, and then titrated with 0.5 M KOH to an appropriately high pH, usually 12. Titrations were carried at constant ionic strength (I=0.15 M KCl) and temperature (t=25.0±0.5 °C), was pre-acidified to pH 2.0 with 0.5 M HCl, and then titrated with 0.5 M KOH to an appropriately high pH, usually 12. Titrations were carried at constant ionic strength (I=0.15 M KCl) and temperature (t=25.0±0.5 °C), and under nitrogen atmosphere. A minimum of three parallel measurements were carried out and the \(pK_a\) values of samples were calculated by RefinementPro software (Sirius Analytical Instrument Ltd, Forest Row, UK).

Titrations in MeOH-water mixture. The co-solvent dissociation constants (\(pK_a\) values) of the compounds with relatively low aqueous solubility were also determined in various MeOH-water mixtures between 15 and 65 wt%.

UV-pH titration

In general, 10.00 mL of an aqueous solution of sample was titrated in an ultra-violet spectrometer (Sirius Analytical Instrument Ltd., Forest Row, UK) attachment for the measurement analyzer (Sirius Analytical Instrument Ltd., Forest Row, UK) fitted with combination Ag/AgCl pH electrode and glass–electrode (Hanna Instruments, UK). The sample was pre-acidified to pH 2.0 with 0.5 M HCl, and then titrated with 0.5 M KOH to an appropriately high pH, usually 12. Titrations were carried at constant ionic strength (I=0.15 M KCl) and temperature (t=25.0±0.5 °C), was pre-acidified to pH 2.0 with 0.5 M HCl, and then titrated with 0.5 M KOH to an appropriately high pH, usually 12. Titrations were carried at constant ionic strength (I=0.15 M KCl) and temperature (t=25.0±0.5 °C), and under nitrogen atmosphere. A minimum of three parallel measurements were carried out and the \(pK_a\) values of samples were calculated by RefinementPro software (Sirius Analytical Instrument Ltd, Forest Row, UK).

Results

Performance of the prediction was evaluated in terms of the number of non-calculated (no predicted \(pK_a\) or 0-predicted \(pK_a\)) cases, correlation between predicted and experimental data, Fischer-F value (F), standard error of estimate (SEE) and mean absolute error (MAE), results are summarized in Table 2. Marvin was found to predict the highest amount of the \(pK_a\) values, while Pallas could not calculate the \(pK_a\) in 40 cases. Considering the r², SEE and MAE no significant difference can be observed. Marvin and ACD slightly outperformed Epik and Pharma, while Pallas gave somewhat less accurate prediction (see Figure 3).

Impact of multiple ionization constants on the reliability of the prediction was also investigated. The MAE in function of the number of ionizable groups is shown in Figure 4. No correlation between the number of ionizable groups and the MAE was observed except in case of Pallas, while the MAE increased with the increasing number of ionizable groups. Next, the effect of \(pK_a\) range on the MAE was analyzed (see Figure 5.). It can be concluded, that the used tools had their highest accuracy in the interval of \(pK_a\) 7-10.

Conclusions

Finally, we investigated the cases, where all predictors made mistakes resulting error above 0.5 \(pK_a\) unit. These mistakes are shown in Figure 6. The non-predicted cases are shown in Figure 7.

The highest performance was observed in case of Marvin, while ACD, PharmaAlgorithm and Epik had moderate accuracy and Pallas proved somewhat less accurate results. It is noteworthy to state that among miscalculated cases weak acidic NH and oxime moieties were marked to be common types.

References


Figure 1. \(pK_a\) distribution

Figure 2. Molecular weight distribution

Figure 3. Evaluation of \(pK_a\) prediction capabilities made by the investigated tools

Figure 4. Impact of multiple ionization on the performance.

Figure 5. Mean absolute error dependence on the \(pK_a\) range

Figure 6. Common outliers identified by all the five \(pK_a\) prediction tools. (NP = non-predicted, red is the worst and green is the best predictor)

Figure 7. Further non-predicted functions by at least two \(pK_a\) prediction tools