Comments on the Article "Evaluation of pK_a Estimation Methods on 211 Druglike Compounds"

John C. Shelley,* David Calkins, and Arron P. Sullivan

Schrödinger, Inc., 101 SW Main Street, Portland, Oregon 97204, United States

Received August 28, 2010

The recent article "Evaluation of pK_a Estimation Methods on 211 Druglike Compounds" (Manchester, J.; et al. *J. Chem Inf. Model.* **2010**, *50*, 565−571) reports poor results for the program Epik. Here, we highlight likely sources for the poor performance and describe work done to improve the performance. Running Epik in the mode intended to calculate pK_a values for sequentially adding/removing protons, as needed to reproduce the experimental conditions, improves the root mean squared error (RMSE) from 3.0 to 2.18 for the 85 public compounds available from the paper. Despite this improvement, there are still other programs in the Manchester paper that outperform Epik. The primary reason is that the public portion of the data set is not diverse and Epik is missing a few key functional groups in this data set that are heavily represented. We show that incorporation of these missing functional groups into the Epik training set improves the RMSE for the public compounds to 1.04. Furthermore, these enhancements help improve the overall performance of Epik on a large druglike test set.

INTRODUCTION

The Manchester paper[1] compares the predicted pK_a values from a number of pK_a software packages, including Epik (v2.0.109),[2] with 378 experimental pK_a values for 211 druglike molecules. A single acid or base classification was used for each experimental pK_a value, and the output from each program was used to determine whether the predicted values should be assigned to a single acid or base classification. Only values with the same classification were matched up. Overall, Epik is reported to have a root mean squared error (RMSE) vs experimental pK_a values of 3.0 and to provide predictions for only 345 of the experimental values (i.e., 33, roughly 9%, were missing). Manchester et al. note that "Epik was developed primarily for rapidly assigning protonation states to (large) databases of virtual compounds prior to virtual screening, an application at which it excels". However, the RMSE and the number of missing pK_a values reported for Epik are large enough that if these results were a true representation of Epik's accuracy, Epik would in fact not be a useful tool for producing protonation states for virtual screening studies. However, in our own studies, we find that the use of Epik to prepare compound libraries for virtual screening significantly improves enrichments (data not shown). Epik was designed primarily as a tool for preparing large ligand libraries for use in virtual screening and does not provide the information Manchester et al. required in their comparison with experimental pK_a values. Below, we describe the procedure required to extract the necessary information from the Epik output files to make the assignments used in the evaluation described by Manchester et al. and discuss how the lack of diversity in the compounds used in that study adversely affects Epik. Finally, we outline improvements that have been made to the latest released version of Epik (v2.1.206), which when combined with the proper protocol for comparing to experimental pK_a values, results in an RMSE of 1.04 for the Manchester public data set.

* To whom correspondence should be addressed. Phone: 503-299-1150. Fax: 503-299-4532. E-mail: John.Shelley@schrodinger.com.

ANALYSIS OF THE MANCHESTER DATA SET USING EPIK

We used the 85 public compounds with 142 pK_a values disclosed in the Supporting Information of Manchester et al.[1] for the current analysis. The structures of the other 126 molecules are proprietary and thus are not available to us. While this introduces some uncertainty in explaining the published results for the full test set, the authors point out that the "physical property distributions of the subset in the Supporting Information parallel those of the full set as do the relative proportions of the types of titratable groups, with the exception of tertiary amines (only 19 are represented) and benzodiazepines (11 of the total of 13 are contained in the Supporting Information)". The experimental data used in the Manchester paper were produced using capillary migration electrophoresis in buffered solutions for a range of pH values with subsequent curve fitting to determine pK_a values. In addition to the pK_a values themselves, the corresponding changes in net charge state of the molecules are indicated. Using a convention in which functional groups that are neutral when protonated are considered acids while those that are neutral when deprotonated are considered bases, the authors identified B1 (the base with the highest pK_a value), B2 (the base with the second highest pK_a value), and A (the acid with the lowest pK_a value) for each molecule. Most molecules do not have all three values. We note that this convention conflicts with the IUPAC definitions of acids and bases and requires special processing of the Epik output in order to make a direct comparison with the experiment (see below). The authors classified the pK_a values predicted

10.1021/ci100332m
Published on Web 12/07/2010

by each of the software packages in this manner, and RMSE values were then calculated between the corresponding B1, B2, and A experimental and calculated values.

Epik's "Sequential pK_a values" mode predicts pK_a values by first adjusting the protonation state to the one with the highest mole fraction at pH 7 and then successively adding and removing protons while recalculating the pK_a values after each change of protonation state in which the pK_a values for sequentially adding/removing protons are calculated. This mode is the appropriate mode to use in comparisons with pK_a values from titration experiments. From subsequent communications with the authors, it is our understanding that, for the values reported in their publication, they used Epik's "Query only" mode, which gives pK_a values specific to protonation state as provided (i.e., the pK_a values that one would map onto the B2 ionization would not be shifted by the protonation corresponding to B1). Epik uses the IUPAC nomenclature for protonic acid/base equilibria, which is the Brønstead−Lowry definition of acids ("a molecular entity or chemical species capable of donating a proton")[3] and bases ("a chemical species or molecular entity having an available pair of electrons capable of forming a chemical bond with a proton").[4] The acid/base designations for each pK_a value in Epik's log file conform to these definitions and thus will change when a proton is added or removed, while those used by Manchester et al. do not. Therefore, in order to correctly map the IUPAC acid/base designations onto B1, B2, and A, one needs to know the formal charge of the molecule for a given proton addition/removal event in addition to the information in the Epik log file. The following procedure can be used to classify pK_a values produced by Epik's "Sequential pK_a values" mode in the desired manner:

1. Determine acid/base designations as used in the Manchester paper.
   a. If the pK_a value is higher than 7, the site is a conjugate acid. If the formal charge on the protonated heavy atom is positive, consider it a base; otherwise, consider it an acid.
   b. If the pK_a value is less than 7, the site is a conjugate base. If the formal charge on the deprotonated heavy atom is negative, consider it a base; otherwise, consider it an acid. An exception is if the formal charge is 0 and the heavy atom is a conjugated heteratom (e.g., nitrogen from a tetrazole ring) in which one of the other atoms in the system carries a negative charge; in this case, it is classified as an acid.
2. Epik will calculate low pK_a values for protonating conjugate acids to prevent them from protonating a second time, so it is possible for two pK_a values to be assigned to the same heavy atom from an acid; when this happens the lower value is discarded.
3. Identify Epik's B1 (the base with the highest pK_a value), B2 (the base with the second highest pK_a value), and A (the acid with the lowest pK_a value).
4. Calculate the RMSE by matching up the B1, B2, and A values of Epik with experimental values.

Using this procedure, Epik v2.0.109 produces 141 pK_a predictions out of the 142 with an overall RMSE of 2.18. While we have asked Manchester et al. to apply this analysis on the internal data, we had not received a response at the time of the submission of this work. However, on the basis of the statement in the paper that the public and private data sets are similar, we can expect a similar improvement in the RMSE and percent of unpredicted pK_a values (<1%) for the entire data set.

Table 1. Predominant Types of Acid/Base Groups As Classified by Epik 2.1027

type of ionizable group	number of matches (% in public data set)
amine ethylamine	26 (18%)
pyridinium ion N^+	24 (17%)
benzylaminium, ring substituted	20 (14%)
secondary aminium ions	20 (14%)
tertiary aminium ions	13 (9%)

DIVERSITY OF IONIZABLE GROUPS IN THE TEST SET

The molecules in the public data set do not have the level of diversity one would expect in a druglike database, nor do they represent a uniform coverage of important functional groups. For instance, 50 compounds contain piperidines and 44 of those are 3-aminopiperidines or 4-aminopiperidines, which both have closely coupled pairs of amine sites. Epik predictions have an RMSE of 1.92 for these 50 compounds. Furthermore, Epik has large RMSE values of 3.09 and 3.83 for two other classes of compounds that are derivatives of 2,7-dihydropyrazolo[3,4-*d*]pyrimidine-4,6-dione with a 2-pyrrolyl or a 1,2,4-triazolyl substituent at the 3 position (present in six compounds) and 3*H*-pyrido[2,3-*e*][1,4]diazapin-2-amine (eight compounds). All three classes of compounds have little or no representation in Epik's training set and only the 4-aminopiperidines have sufficient data in Reaxys[5] to support reliable fitting for pK_a prediction. However, because this small number of poorly parametrized functional groups represents such a large fraction of the public test set—more than 78% of the pK_a values—Epik's overall performance is significantly worse that we observe for more diverse molecules.

To further highlight the lack of diversity in the data set, we note that five acid/base patterns account for 73% of the matches with experimental pK_a values in the public data set. As noted earlier, the frequency at which functional groups are in close proximity and that carry a positive charge when protonated is an unusual feature of this data set. For comparison, 14 different acid/base patterns were required to match with 74% of the experimental pK_a values and the top five patterns only accounted for 40% of the pK_a values reported in another recent publication comparing programs that predict pK_a values.[6] Interestingly, for this more diverse data set Epik (v1.6) had a mean absolute error of 0.89 for 257 of the 261 experimental values tested (2% miss rate), which is much more consistent with our internal findings on diverse data sets.

IMPROVING EPIK'S PERFORMANCE

For the primary application of Epik—preparing protonation states for virtual screening—an RMSE for pK_a values below 2 is essential in order to produce appropriate protonation states. Consistent with this goal, we have made adjustments to the Epik parameters in order to address issues for the molecules in this data set. To cover the functional groups for which Epik did not perform well while not fitting directly to this data, we obtained information from other sources

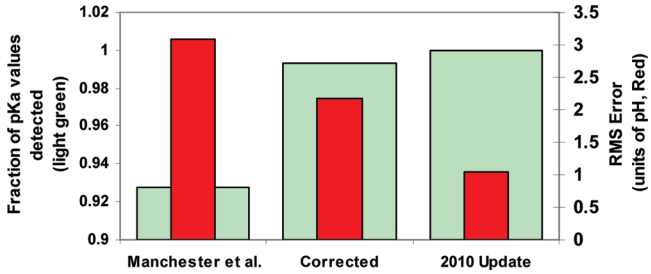

Figure 1. RMSE and number of pK_a values predicted. Red bars represent the RMSE vs experiment while green bars represent the fraction of the pK_a values detected by Epik. The results for Manchester et al. are as reported in ref 1 for the whole data set. The corrected results are for our analysis of the public data set using the same version of Epik as Manchester et al. The 2010 update results are for Epik version 2.1.207 using the same analysis procedure for the public data set. The scales have been selected to mimic those in Figure 2 from ref 1.

including experimental values, Jaguar pK_a calculations, and Jaguar calculations of the relative energies of tautomers in water. With these new parameters, along with the correct protocol for comparing the results with these experiments (as described above), the RMSE for the public data set dropped to 1.04 for Epik v2.1.207. Furthermore, these changes along with many other unrelated changes improved the performance on a training set of 798 druglike compounds to 1.21, as compared to 1.39 for the previous released version (v2.0.109).

SUMMARY

The performance of Epik for pK_a predictions on the public data set from Manchester et al. was presented and we have provided likely explanations for why the results are worse than those in previously published studies that included Epik. First, using a protocol described here for correctly assigning acids and bases in a manner consistent with the experiments resulted in an improvement of the RMSE from 3.08 to 2.18. Furthermore, using this correct protocol improved the number of predicted pK_a values from 135 to 141 (out of 142) for the public data set. Another factor contributing to the poor results of Epik is that the compounds used by Manchester et al. are not diverse relative to a typical data set of druglike molecules. Five patterns covered 74% of the pK_a sites in the public set and Epik does not perform well on some of these functional groups. Improvements to Epik along with the correct assignment of acids and bases have led to a drop in the RMSE to 1.04 for the public data set, which represents a significant improvement compared to the value of 3.0 published by Manchester et al.

REFERENCES AND NOTES

(1) Manchester, J.; Walkup, G.; Rivin, O.; You, Z. Evaluation of pK_a estimation methods on 211 druglike compounds. *J. Chem. Inf. Model.* **2010**, *50*, 565–571.

(2) (a) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput.-Aided. Mol. Des.* **2007**, *21*, 681–691. (b) *Epik*, version 2.0109; Schrödinger, Inc.: New York, NY, 2009.

(3) Nic, M.; Jirat, J.; Kosata, B. IUPAC Gold Book, Acid;http://goldbook.iupac.org/A00071.html (accessed July 8, 2010).

(4) Nic, M.; Jirat, J.; Kosata, B. IUPAC Gold Book, Base; http://goldbook.iupac.org/B00601.html (accessed July 8, 2010).

(5) *Reaxys*, version 1.48; Elsevier Properties, SA: New York, 2010.

(6) Liao, C.; Nicklaus, M. C. Comparison of nine programs predicting pK(a) values of pharmaceutical substances. *J. Chem. Inf. Model.* **2009**, *49*, 2801–2812.

CI100332M