ELSEVIER

International Journal of Pharmaceutics 164 (1998) 139–145

international journal of pharmaceutics

Precision tests of a pH-solubility profile computer program

William H. Streng *, Daniel H.-S. Yu

Hoechst Marion Roussel, P.O. Box 9627, Kansas City, MO 64134, USA

Received 2 May 1997; received in revised form 23 October 1997; accepted 2 December 1997

Abstract

The solubility of pharmaceutically important compounds as a function of pH is needed in order to optimize the formulation(s) and help to explain the observed behavior of the compounds during their development. The solubility of a weak acid or weak base is dependent upon its ionization constant, uncharged solubility and solubility product. Computer programs have been written to determine these parameters from experimental pH-solubility data and the parameters were found to be precise with random errors up to $\pm 40\%$ in $\{H\}$, ± 0.05 in pH, $\pm 40\%$ in solubility and $\pm 10\%$ in both solubility and $\{H\}$. Systematic errors in the pH-solubility data result in systematic errors in pK_a, pK_{sp} and uncharged solubility. These results demonstrate that pH-solubility data can be used to precisely estimate pK_as, pK_{sp}s and solubility of the neutral molecule. This technique may be a useful and practical alternative with low solubility compounds, particularly when there is no UV/VIS shift associated with species ionization.

Keywords: pH-solubility; Ionization constant; Compounds; Non-linear regression analysis

1. Introduction

The solubility of pharmaceutically important compounds as a function of pH is needed in order to optimize the formulation(s) and help to explain the observed behavior of the compounds during their development. The solubility of a weak acid or base as a function of pH has been shown to be dependent upon the pK_a, pK_{sp} and solubility of the uncharged species of the compound (Streng et al., 1984; Streng and Tan, 1985). Computer programs have been written to determine these parameters from pH-solubility data using non-linear regression analysis and, knowing the parameters, calculate a solubility profile. The programs are divided into four major parts:

1. an input program, I, to create a data file using the experimental pH-solubility data;
2. a non-linear regression program, II, used to calculate the values for the solubility of the uncharged species, pK_a and pK_{sp};

* Corresponding author.

0378-5173/98/$19.00
PII S0378-5173(97)00428-6

3. a solubility-profile program, III, to calculate the pH-solubility profile using the calculated parameters;
4. a graphing program, IV, to plot the original data and a calculated profile.

The purpose of this study was to determine the precision of the regression program. Values for the uncharged solubility, pK_a and pK_{sp}, were input into the pH-solubility profile program to generate a profile. Random errors and systematic errors were introduced into the calculated pH-solubility profile which were then input into the input and the regression programs to obtain new values for the uncharged solubility, pK_a and pK_{sp}. These values were compared to the original values to determine the precision of the regression program.

2. Theoretical section

The solubility profile of a monoprotic weak acid or base as a function of pH can be divided into two regions depending on which species is saturating the solution. The equations describing the solubility function in the two regions have been previously derived (Streng et al., 1984) and are given below for a weak base.

$$S_{1,0} = \frac{-\left([M^+] + \frac{\{H^+\}}{\gamma_{H+}} - \frac{K_w}{\{H^+\}\gamma_{OH-}}\right) + \left(\left([M^+] + \frac{\{H^+\}}{\gamma_{H+}} - \frac{K_w}{\{H^+\}\gamma_{OH-}}\right)^2 + \frac{4K_{sp}}{\gamma_{HB+}\gamma_{X-}}\right)^{0.5}}{2\left(\frac{\{H^+\}\gamma_B}{\{H^+\}\gamma_B + K_a\gamma_{HB+}}\right)} \quad (1)$$

$$S_{1,1} = \left(\frac{\{H^+\}}{K_a\gamma_{HB+}} + \frac{1}{\gamma_B}\right)\{B\}_s \quad (2)$$

where: K_a is the molar dissociation constant; K_w is the ionization constant of water; K_{sp} the molar solubility product; $\{B\}_s$ the molar activity of the uncharged weak base species; HB^+ the protonated weak base species; $\{H^+\}$ the hydrogen ion activity; $[M^+]$ the molar concentration of metal cation M^+ (Na^+, etc.); and γ_i is the activity coefficient of species *i*.

For a weak base, in region '0' the solubility is controlled by the pK_a and pK_{sp} while in region '1' the solubility is controlled by the pK_a and neutral species solubility.

The pH at which the function changes is the pH of maximum solubility, $\{H^+\}_{max}$:

$$\{H^+\}_{max} = \frac{-[M^+] + \left([M^+]^2 + 4\left(\frac{1}{\gamma_{H+}} + \frac{\{B\}_s}{K_a\gamma_{HB+}}\right)\left(\frac{K_w}{\gamma_{OH-}} + \frac{K_aK_{sp}}{\{B\}_s\gamma_{X-}}\right)\right)^{0.5}}{2\left(\frac{1}{\gamma_{H+}} + \frac{\{B\}_s}{K_a\gamma_{HB+}}\right)} \quad (3)$$

3. Computer programs

The following is a listing of the FORTRAN source codes which have been written along with a brief description:

1. SOL_DATA.FOR: a program used to enter the data used in the calculation of the solubility parameters;
2. SOL_Y.FOR: a SUBROUTINE used to calculate the activity coefficients of the singly charged species in a saturated solution of a weak acid or a weak base using either Davies' equation or Debye-Huckel theory;
3. C_ROOT.FOR: a double precision function which solves for the roots of a cubic expression and returns a value closest to an estimated value;
4. S3P_FUNC.FOR: a double precision function used to determine the values for the parameters pK_a, pK_{sp} and p(UNCH);

5. S3P_DATA.FOR: a SUBROUTINE used to transfer the experimental data file into the nonlinear regression program;
6. S3P_COM.FOR: an INCLUDE module used in the regression program which determines the pK_a, pK_{sp} and p(UNCH);
7. SOL_PROF.FOR: a double precision program used to calculate the solubility profile of a weak acid or base knowing values for the p(UNCH), pK_{sp} and pK_a, [p(UNCH) equals: $-\log$(uncharged species solubility)];
8. SOL_PLOT.FOR: a program which calls the graphics subroutine to plot the solubility profile;
9. GRAPHER.FOR: a subroutine which plots the solubility data and calculated profile.

The source codes for the above programs to be used on a WINDOWS 95® platform are available upon request[1].

The computer programs were run on a desk-top personal computer[2] under FORTRAN environment[3]. Four programs were created, compiled and linked to form one executable file. Any regression program which allows the input of a module containing the model can be used to fit the experimental data. The above function (S3P_FUNC.FOR) was specifically written for use with PCNONLIN4[4].

The first program (I), the data input program, SOL_DATA.FOR, is used to create a data file containing the experimental data which is needed by the regression program. The following information needs to be provided as part of the input data:

1. molecular weight (MW) of the compound;
2. temperature;
3. pK_w;
4. dielectric constant of the solution;
5. type of the compound (weak acid or weak base);
6. form of the bulk material (protonated or deprotonated form);
7. the activity coefficient model (Davies or Debye-Huckel). If the Debye-Huckel model is used, the ion size parameters need to be input. It is not always possible to obtain estimates for the ion size parameters and ease of use makes Davies equation a preferred choice. Although there will be differences in the calculated parameters, the differences are usually not significant;
8. a normalized convergence function to use in the regression analysis. Function 1 is the sum of the absolute differences in the base ten logarithms of the experimental and calculated solubility and function 2 is the sum of the squares of the quantity of the difference between the experimental and calculated value divided by the calculated value. For a particular data point, both of these functions return a value of zero when the calculated value equals the experimental value. Small differences in the calculated parameters will be observed between the two functions. It is not possible to know a priori which function will give the 'best' fit to the data, however, a decision can usually be made by comparing plots of the calculated curves with the experimental data;
9. number of the data sets used in the calculation;
10. data sets including pH, solubility, salt concentration, acid concentration, base concentration, undissolved solid concentration (moles of excess solid divided by the total solution volume).

The second program (II), the non-linear regression program, which includes S3P_FUNC.FOR, S3P_DATA.FOR, SOL_Y.FOR, S3P_COM.FOR, and C_ROOT.FOR, is used to determine the values of p(UNCH), pK_a and pK_{sp} based on the pH-solubility experimental data. The following information needs to be provided to run the program:

1. data file name;
2. estimated values of p(UNCH), pK_a and pK_{sp} used to initialize the regression process;
3. initial step sizes for the parameters.

The third program (III), the solubility-profile calculation program, which includes SOL_

[1] Please provide a 3.5 inch floppy disk along with a stamped self addressed envelope. If a copy of the compiled program is wanted, the authors must receive a statement that a version of PCNONLIN is available.

[2] A GATEWAY2000 4DX2-66V computer.

[3] Microsoft FORTRAN visual workbench, version 1.0.

[4] PCNONLIN is software provided by SCI Software, Lexington, KY. The program used the Nelder Mead method during the regression analysis.

PROF.FOR, SOL_Y.FOR, and C_ROOT.FOR, is used to generate the pH-solubility profile using predetermined values of p(UNCH), pK_a and pK_{sp}. In addition to these parameters, the following information needs to be provided when running the program:

1. molecular weight (MW) of the bulk compound;
2. MW of the precipitated protonated compound;
3. MW of the precipitated unprotonated compound;
4. type of the compound (weak acid or base);
5. form of the bulk material (protonated or unprotonated form);
6. Is there a common ion between the precipitated salt and pH adjusting acid or base?
7. Is there a common ion between the precipitated salt and additional salt? If the answer is yes, what is the concentration?
8. Is there any salt present without a common ion? If the answer is yes, what is the concentration?
9. temperature;
10. dielectric constant of the solution;
11. pK_w;
12. the activity coefficient model (Davies' or Debye-Huckel). If the Debye-Huckel model is used, the ion size parameters need to be provided;
13. initial pH;
14. final pH;
15. pH increments.

A plot routine, program IV, can be run which will graph the molar concentrations of the experimental data and calculated profile versus pH.

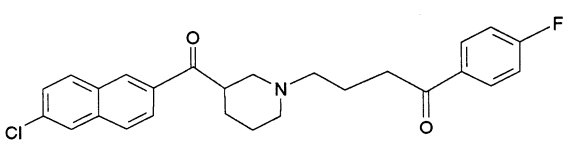
To test for precision, values for p(UNCH), pK_a and pK_{sp} were input into the solubility profile program, (III), and the pH-solubility profile generated. Selected profile points were taken and were changed by either introducing random errors or systematic errors to the values. These modified points were input using program I, to provide a data file for the regression program. New values of p(UNCH), pK_a and pK_{sp} were obtained by running the regression program, (II), which were then compared to the original values of p(UNCH), pK_a and pK_{sp} to estimate the precision of the programs.

Table 1
Input data for compound I to obtain the pH-solubility profile

Compound name	M17214
Solvent system	50% ethanol–water
MW of bulk compound	316.82
MW of the ppt protonated compound	353.32
MW of the ppt unprotonated compound	316.82
p (solubility of uncharged species)	4.125
pK_a	7.81
pK_{sp}	4.51
Type of the compound	Weak base
Form of the bulk material	B (unprotonated)
A common ion between the ppt salt and pH adjusting agent?	Yes
A common ion between the ppt salt and additional salt? If yes, what is the concentration?	No
Any salt present without a common ion? If yes, what is the concentration?	No
Temperature	25.0
Dielectric constant (50% ethanol)	51.4
pK_w	14.8717
Activity coefficient model used	Davies
Initial pH	2.0
Final pH	9.0
pH increments	0.1
Data file generated	M17214.SOL

4. Results and discussion

The pH-solubility profiles of compound I have been reported (Streng et al., 1984) and the parameters were used in this study as a model



Scheme 1. Chemical structure of compound I.

[5] 4-[4-[(6-chloro-2-naphthalenyl)carbonyl-1-piperidinyl] × -1-(4-fluorophenyl)-1-butanone.

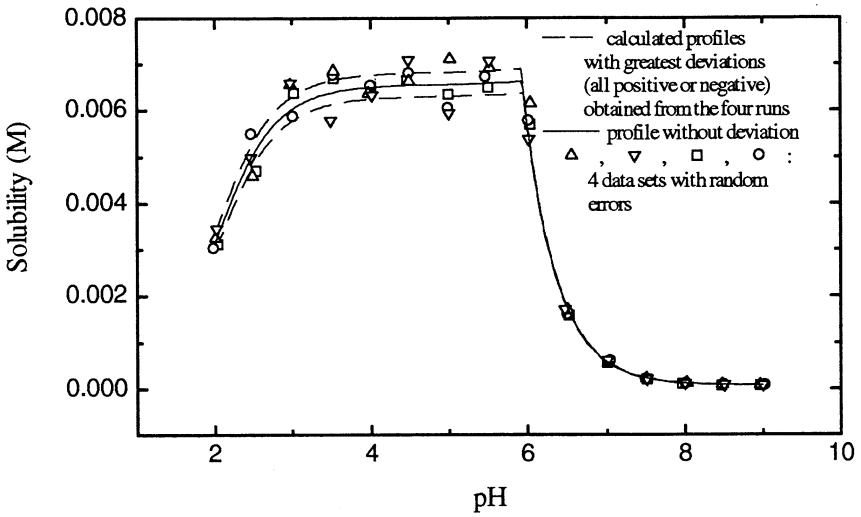


Fig. 1. The pH-solubility profile and data for pH and solubility having 10% random errors.

compound for a weak base. The chemical structure of compound I[5] is shown above and the input data used to obtain the pH-solubility profile are listed in Table 1 Scheme 1.

The pH-solubility profile obtained after running the solubility profile program is shown in Fig. 1. Fifteen selected data points were chosen to validate the computer programs. These 15 data points were evenly distributed over the pH range studied and can be reasonably obtained during an experiment.

The original data calculated for the solubility profile was modified by introducing random errors. Numbers between O and 1 were randomly generated[6] using a uniformly distributed function. Using a normally distributed function, which may or may not be correct, would only make the results tighter since more random values would be obtained near the original value. The random error is defined as:

$$\text{random error} = [(\text{numbers generated between 0 and 1}) - 0.5] \times 2 \times \text{level} \quad (4)$$

In Eq. (4), level is the maximum percentage error, for example, a 5% random error level indicates the error is between -5 and 5%.

4.1. Random errors in solubilities

The first precision test was to introduce random errors in the solubilities. Different random errors were generated and used to modify the original 15 solubility data according to the following formula:

$$\text{solubility with random error} = \text{solubility} \times (1 + \text{random error}) \quad (5)$$

The greatest absolute value of deviation for each parameter was less than 0.073 for random errors in solubilities up to $\pm 40\%$ indicating the program is precise in this variable. Normal experimental errors in the solubility determinations are usually less than 2%.

4.2. Random errors in {H}

The second precision test was to introduce errors in $\{H\}$. Different random errors were generated and used to modify the original 15 pH values according to the following formula:

[6] RANDOM function in Microsoft Excel 4.0 for Windows.

pH with random error =

$$-\log_{10}[\{H\} \times (1 + \text{random error})] \quad (6)$$

The greatest absolute value of deviation for each parameter is less than 0.027 for random errors up to 40% indicating the program is precise in this variable.

4.3. Random errors in pH

Random errors may also be generated when the pH is measured in an experiment. The third precision test was to introduce random errors in pH. The largest change in pH was set at 0.05. Different random errors were generated and used to modify the original 15 pH values according to the following formula:

$$\text{pH with random error} = \text{pH} + \text{random error} \quad (7)$$

The greatest absolute value of deviation for each parameter was less than 0.040 with random error in pH up to 0.05 indicating the program is precise in this variable. Normal experimental random errors in pH are not greater than 0.01.

4.4. Random errors in both solubilities and {H}

The fourth precision test was to introduce random errors in both the solubility and $\{H\}$ values (a greater random error effect was observed on $\{H\}$ than on pH). Different levels of random errors were introduced to the 15 solubility and $\{H\}$ data sets using Eqs. (5) and (6).

The results indicate the programs are precise. For 5% random errors in both the solubilities and the $\{H\}_s$ for the four runs, the greatest absolute values of deviation for p(UNCH), pK_a and pK_{sp} are 0.019, 0.034 and 0.027, respectively, and for 10% random errors these values are 0.032, 0.035 and 0.040, respectively. Fig. 1 shows the data and the pH-solubility profiles for 10% random error.

4.5. Systematic errors in pH

The fifth precision test was to introduce a systematic shift in pH. The original 15 pH data were modified by introducing shifts of -0.05, -0.02, -0.01, 0.01, 0.02 and 0.05 according to the following formula:

$$\text{pH with systemic error} = \text{pH} + \text{pH shift} \quad (8)$$

Systematic pH shifts result in a deviation of the pK_a and pK_{sp}. Although the values of p(UNCH) remain the same for all pH shifts and the deviation of pK_{sp} is very small, the value for the pK_a deviates the same amount as the pH shift. For example, when pH shifts -0.05 U, the pK_a shifts -0.05 U and when pH shifts 0.02 U the pK_a shifts 0.02 U.

4.6. Systematic errors in solubilities

The sixth precision test was to introduce systematic errors into the solubility. The systematic errors tested were -20, -10, 10 and 20% in the solubility according to the following formula:

$$\text{solubility with systemic error} = \text{solubility} \times (1 + \text{solubility shift}) \quad (9)$$

The extent of deviation was found to be: pK_{sp} > p(UNCH) > pK_a. Also, the values of UNCH, K_a and K_{sp}, their deviations and percentage deviations were calculated and are shown in Fig. 2. The percentage deviation of UNCH is the same as the percentage solubility shifts. For example, when solubility shifts -20%, UNCH shifts -20% and when solubility shifts 20%, UNCH shifts 20%. The figure shows the deviation for pK_a is very small.

This program was also tested for precision when the compound of interest was a weak acid. Identical results were obtained for all of the above sources of random and systematic error.

5. Conclusions

The computer programs were found to be precise for the determination of p(UNCH), pK_a and pK_{sp} using the pH-solubility profile data for random errors in $\{H\}$ (up to ± 40% tested), in

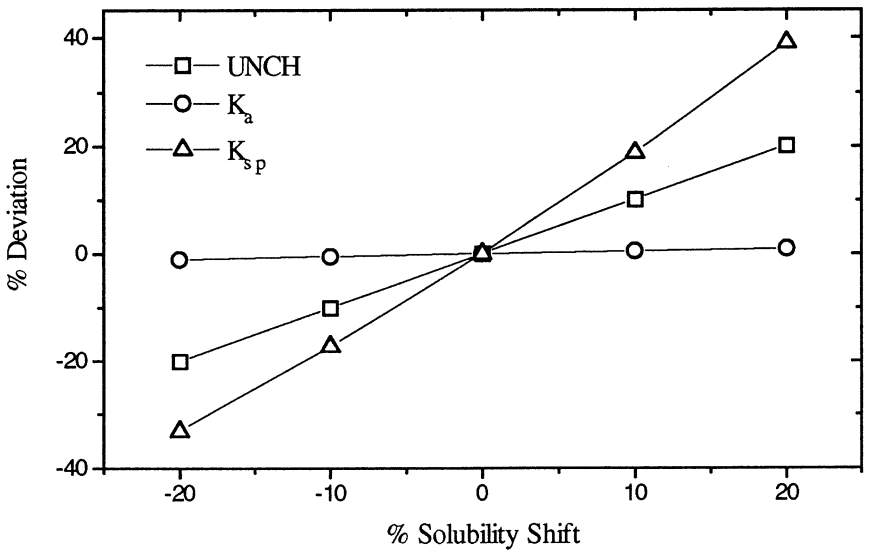


Fig. 2. The percentage deviation of UNCH, K_a and K_{sp} vs. percentage solubility shift for a weak base.

pH (up to ± 0.05 tested), in solubility (up to $\pm 40\%$ tested) and in both $\{H\}$ and solubility (up to $\pm 10\%$ tested). When there are systematic errors in the pH-solubility data, the values obtained have systematic deviations. When there is a systematic shift in pH, the values for the pK_a shift the same amount while there is no effect on p(UNCH) and little effect on pK_{sp}. When there is a systematic shift in solubility, the values for p(UNCH) and pK_{sp} have systematic deviations while there is little effect on pK_a; the values for UNCH shifts the same amount as the solubility shift.

Acknowledgements

The authors would like to thank Chengyue Zhu for writing the initial screen module making it possible to run the software on the Windows 95 platform. He also modified the programs to use default values, transfer common data between the programs and plot the data and calculated curves.

References

Streng, W.H., Swee, K.H., Helms, P.E., Tan, H.G.H., 1984. General treatment of pH-solubility profiles of weak acids and bases and the effects of different acids on the solubility of a weak base. J. Pharm. Sci. 73 (12), 1679–1684.

Streng, W.H., Tan, H.G.H., 1985. General treatment of pH-solubility profiles of weak acids and bases. II. Evaluation of thermodynamic parameters from the temperature dependence of solubility profiles applied to a Zwitterionic compound. Int. J. Pharm. 25, 135–145.