Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Pharmaceutics 275 (2004) 133–139
international journal of pharmaceutics
www.elsevier.com/locate/ijpharm

Prediction of aqueous solubility of organic salts of diclofenac using PLS and molecular modeling

Vimon Tantishaiyakul*

Department of Pharmaceutical Chemistry, Prince of Songkla University, Hat-Yai, Songkhla 90112, Thailand

Received 7 August 2003; received in revised form 9 December 2003; accepted 22 January 2004

Abstract

Organic salts of diclofenac were predicted by using computed molecular descriptors and multivariate partial least squares (PLS). The molecular descriptors including binding energy and surface area of salts were calculated by the use of Hyperchem and ChemPlus QSAR programs for Windows. Other physicochemical properties such as hydrogen acceptor for oxygen atoms, hydrogen acceptor for nitrogen atoms, hydrogen bond donors, hydrogen bond-forming ability, molecular weight, and log partition coefficient (log *P*) of bases were also used as descriptors. Good statistical models were derived that permit simple computational prediction of salt solubility of a same parent structure. The final models derived had R^2 value = 0.96 and root mean square error for prediction (RMSEP) values ranging from 0.021 to 0.054 (log scale). Preferably all utilized descriptors in the final models can readily obtain from the chemical structure of salt and base. Molecular weight of base is one of the important factors associated with salt solubility. While increased molecular weight of base, surface area of salt and hydrogen bonding ability of base increase solubility, and increased binding energy and log *P* of base have negative effect on salt solubility.

Keywords: Salt; Solubility; Partial least squares; Diclofenac; Binding energy

1. Introduction

Aqueous drug solubility is one of the most important factors in the process of drug discovery and development from molecular design to pharmaceutical formulation and biopharmacy. For weak electrolyte drugs, salt formation that is a much simpler method than complex molecular modifications, is a common approach to improve solubility. The use of different counter-ions in salt formation can result in salts with different physicochemical properties that are not always predictable without experimentation. Accordingly various organic and inorganic salts of acidic and basic drugs have been prepared and their physicochemical properties investigated for salt selections (O'Connor and Corrigan, 2001; Forbes et al., 1995).

The ability to predict the aqueous solubility of salts can speed up the process of drug development. Various predictions of aqueous solubility of organic compounds using quantitative structure–property relationship (QSPR) have been reported (Yaffe et al., 2001; Chen et al., 2002; Gao et al., 2002; Jorgensen and Duffy, 2002). However, QSPR has rarely been used in the prediction of salts. Recently, the correlation of aqueous solubility of salts of benzylamine with experimental parameters was investigated by Parshad et al. (2002). The most significant descriptors included Charton's steric parameter, Hansch hydrophobic parameter, molecular weight and intrinsic solubility.

* Tel.: +66-74-288864; fax: +66-74-428239.
E-mail address: tvimon@ratree.psu.ac.th (V. Tantishaiyakul).

0378-5173/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpharm.2004.01.028

It was reported that no correlation was found between diclofenac salt solubility and any one of the parameters pK_a, hydrophilicity, or melting point of counter-ions (O'Connor and Corrigan, 2001). One important factor that might govern aqueous solubility of salt/ion pair is the electrostatic interaction between cationic and anionic species of the ion pair. It is thus interesting to investigate the significance of binding energy of the ion pair and employ it as one of the descriptors for predicting salt solubility. Diclofenac, an acidic drug (pK_a 3.80 at 25 °C) with very low aqueous solubility (6×10^{-5} M at 25 °C) in the unionized form (Chiarini et al., 1984), is used as a model drug. The calculated binding energy and surface area of salts and other physicochemical properties of organic bases were related to diclofenac salt solubility using the partial least squares or projection to latent structures (PLS) regression.

2. Methods

2.1. Data set selection

Aqueous solubility data for 23 diclofenac salts were taken from Fini et al. (1996). These values were converted from milligram per milliliter to logarithm of salt solubility in millimoles (log S). These salts were

Table 1
Salt solubilities, physicochemical and molecular parameters of salts and bases

Number	Salt	S[a]	log S[b]	Bind E[c]	HAO[d]	HAN[e]	HD[f]	HB[g]	Sarea[h]	c log P B[i]	MW B[j]	MP B[k]
1	META	94.98	1.98	121.43	0	1	2	3	529.12	−0.664	31.05	173.00
2	DMETA	132.17	2.12	116.69	0	1	1	2	569.46	−0.518	45.05	178.50
3	TMETA	166.44	2.22	113.87	0	1	0	1	572.89	0.048	59.11	212.50
4	TTMETA	198.11	2.30	104.03	0	0	0	0	582.33	–	74.14	242.50
5	DETA	198.11	2.30	105.58	0	1	1	2	605.91	0.54	73.13	158.00
6	DPRA	254.70	2.41	117.31	0	1	1	2	670.37	1.598	101.19	168.50
7	DBUA	303.88	2.48	115.25	0	1	1	2	713.38	2.656	129.24	145.50
8	P	193.68	2.29	117.55	0	1	1	2	597.45	0.106	71.12	111.00
9	PP	223.35	2.35	112.27	0	1	1	2	624.4	0.665	85.14	184.50
10	M	233.34	2.37	115.07	2	1	1	4	608.96	−0.646	87.12	171.00
11	PZ	225.30	2.35	119.11	0	2	2	4	618.05	−0.8	86.13	205.00
12	MEP	223.35	2.35	112.57	0	1	0	1	624.97	0.842	85.14	209.50
13	MEPP	250.91	2.40	108.94	0	1	0	1	614.44	1.401	99.17	86.00
14	MEM	254.57	2.41	114.43	2	1	0	3	614.92	0.132	101.14	106.50
15	MEPZ	252.67	2.40	111.06	0	2	1	3	625.63	−0.022	100.16	163.50
16	MEA	171.02	2.23	113.28	2	1	3	6	555.54	−1.295	61.08	145.00
17	DEA	262.04	2.42	118.01	4	1	3	8	616.53	−1.463	105.13	125.50
18	TEA	335.03	2.53	114.47	6	1	3	10	609.64	−1.586	149.19	133.50
19	TRIS	290.33	2.46	119.04	6	1	5	12	608.81	−2.908	121.13	195.00
20	HEP	280.05	2.45	110.61	2	1	1	4	649.47	0.263	115.17	103.00
21	HEPP	303.78	2.48	108.72	2	1	1	4	643.78	0.822	129.2	123.00
22	HEM	307.01	2.49	110.49	4	1	1	6	615.26	−0.447	131.17	91.00
23	HEPZ	305.40	2.48	113.99	2	2	2	6	620.82	−0.601	130.18	103.00

[a] Salt solubility (mM).
[b] Log of salt solubility.
[c] Binding energy of salt.
[d] Number of hydrogen bond acceptor oxygen atoms.
[e] Number of hydrogen bond acceptor nitrogen atoms.
[f] Number of hydrogen bond donor atoms.
[g] Hydrogen bond formation ability.
[h] Surface area of salts (Å^2).
[i] Calculated log P of bases.
[j] Molecular weight of bases.
[k] Melting point of bases.

prepared from linear alkylamine: methyl (META), dimethyl (DMETA), trimethyl (TMETA), diethyl (DETA), dipropyl (DPRA), dibutyl (DBUA) amines, tetramethyl ammonium (TTMETA); cyclic alkylamine: pirrolidine (P), piperidine (PP), morpholine (M), piperazine (PZ), *N*-methyl pyrrolidine (MEP), *N*-methyl piperidine (MEPP), *N*-methyl morpholine (MEM), *N*-methyl piperazine (MEPZ); hydroxy alkylamine: monoethanol (MEA), diethanol (DEA), triethanol (TEA), *tris*(hydroxyethyl) aminomethanol (TRIS), *N*-(2-hydroxyethyl) pyrrolidine (HEP), *N*-(2-hydroxyethyl) piperidine (HEPP), *N*-(2-hydroxyethyl) morpholine (HEM), *N*-(2-hydroxyethyl) piperazine (HEPZ). All of these organic salts and their solubilities are presented in Table 1.

2.2. Molecular modeling and molecular property calculation

Molecular modeling calculations were performed using HyperChem 5.1 for Windows (Hypercube, FL, USA). The MM+ molecular mechanics force field was first run to get close to the optimized geometry. Molecular mechanics calculations treat atoms as Newtonian particles interacting through a potential energy function. The potential energies depend on bond lengths, bond angles, torsion angles, and nonbonded interactions (including van der Waals forces, electrostatic interactions, and hydrogen bonds). In these calculations, the forces on atoms are functions of atomic position. The conformation obtained from molecular mechanics was subjected to a refined geometry optimization using the PM3 semiempirical quantum chemistry. Semiempirical calculations solve the Schrödinger equation to describe the electronic properties of atoms and molecules. To simplify and shorten these calculations, semiempirical methods make many simplifications, calculating only for valence electrons; neglecting the integrals for certain interactions; using standard, non-optimized, electron orbital basis functions. It was reported that calculation result of PM3 semiempirical molecular method and HF/6-31G base function of ab initio molecule orbital method are very close (Lu et al., 1998b). The PM3 method has been previously used for geometry optimization in various studies (Lu et al., 1998a; Huibers, 1999; Tantishaiyakul, 2001). To shorten the time to calculate on a computer, binding energy of diclofenac salt was computed in vacuo using the PM3 method.

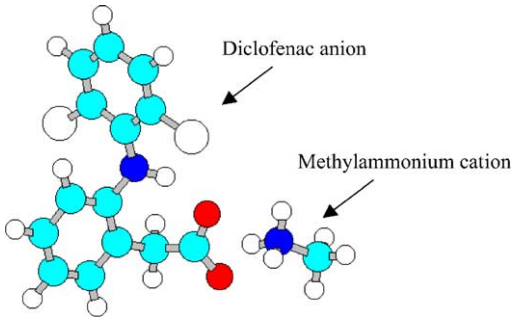

Fig. 1. Salt/ion pair formation constituted by diclofenac anion and methylammonium cation.

Binding energy was calculated as described by Aleman and Zanuy (2000) and Madhan et al. (2001). In brief, the salt/ion pair constituted by the diclofenac anion and the positively charged base (Fig. 1) was calculated to obtain the total energy of ion pair ($TE_{ion-pair}$). The interaction energy of the ion pair ($E_{interaction}$) was calculated as the difference between the total energy of the ion pair and the sum of the energy of diclofenac ($E_{diclofenac}$) and organic base (E_{base}). The negative of the interaction energy is termed the binding energy ($E_{binding}$):

$$E_{interaction} = TE_{ion-pair} - [E_{diclofenac} + E_{base}]$$

$$E_{binding} = -E_{interaction}$$

The ChemPlus QSAR Properties 1.5 (Hypercube, FL, USA) was used for further calculation of surface area of the salt/ion pair.

Hydrogen bond-forming ability (HB) of the whole molecule of base is the sum of hydrogen bond numbers of various groups including oxygen–hydrogen bonding acceptor (HAO), nitrogen–hydrogen bonding acceptor (HAN), and hydrogen bonding donor (HD) and was calculated as described by Xia et al. (1998). The log *P* value of base was calculated using the *c*log *P* program (Biobyte, CA, USA) which calculates directly from the molecular structure.

2.3. Statistical analysis

A principal components analysis (PCA) was performed on the data set consisting of all descriptors. A

PCA is a projection method for extracting the information contained in the descriptor matrix. The principle of PCA is to find the directions in space along which the distance between data points is the largest. This can be translated as finding the linear combinations of the initial variables that contribute most to making the samples different from each other. These directions are called principal components (PCs).

The relationship between the experimentally determined solubility of organic salts of diclofenac and the descriptors was subsequently determined using PLS1 analysis. PLS is a bilinear modeling method where information in the descriptor matrix X is projected onto a small number of underlying ("latent") variables called PLS components. These PLS components are similar to principal components, and will also be referred to as PCs. The matrix Y is simultaneously used in estimating the "latent" variables in X that will be most relevant for predicting the Y variables.

The number of significant PCs for the PLS algorithm is determined using the cross-validation method. With cross-validation, some samples are kept out of the calibration and used for prediction. The process is repeated so that all samples are kept out once. The values for the left out compound is then predicted and compared with the known value. The prediction error sum of squares (PRESS) obtained in the cross-validation is calculated each time that a new PC is added to the model. Haaland and Thomas criterion (1988) was applied for the selection of the optimum number of PCs. This involves the comparison of the PRESS from models (h models) with the model which involves the number of PCs yielding the minimum PRESS (h* model). The F-statistic is used to make the significance determination with a value of $\alpha = 0.25$.

A descriptor selection was determined according to variables important in the projection and loading plot. Insignificant descriptors were left out of the model and their importance for predictivity determined by a cross-validation procedure. If the predictivity of the model increased, the descriptor in question was removed from the model, otherwise the descriptor was kept in the model.

The software package used for conducting both PCA and PLS analysis was Unscrambler 6.01 (Computer-Aided Modelling A/S, Trondheim, Norway). All variables were centered prior to PLS processing. Four analyses were performed: one on a training set of 16 salts and three others on the entire data set of 23 salts.

3. Results and discussion

The data set used consists of 23 organic salts of diclofenac with solubility ranging from 94.98 to 335.03 mM. All the descriptors used and salt solubility are listed in Table 1.

In PCA analysis, the two first PCs account for approximately 91% of the variance. Salts are uniformly distributed in the four quadrants of a score plot, suggesting that molecular diversity of the salts is reasonable. Molecular weight of bases and surface area of salts are the most important variables.

The goodness-of-fit of PLS model can be expressed as a root mean square error (RMSE)

$$\text{R.M.S.E.} = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}$$

where $\hat{y}_i$ and y_i represent the calculated and the experimental value of salt solubility, respectively, and n is the number of samples. The value of RMSE is an indication of the average error in the analysis for each set. The results of all models built from PLS analyses are summarized in Table 2. The values for the PLS coefficients are presented in Table 3. Initially, the 23 salts were used as a training set and all 9 descriptors were included to build the model (model 1). The resulting PLS analysis yielded a statistically acceptable model containing 3 PCs with $R^2 = 0.969$, $Q^2 = 0.937$. In general, the use of the smallest possible number of significant descriptors that yields an acceptable model is recommended. To obtain a model containing fewer descriptors, HAN, the least important variable was thus excluded from the model. The resulting model (model 2) shows the overall statistics similar to model 1. To focus on a few important variables, the descriptors in model 2 were further reduced. Removing melting point of base and HD resulted in model 3 with a slight statistical improvement with respect to cross-validated RMSE value. However, model 3 is the most applicable model since it contains fewer variables and all the utilized descriptors can be obtained directly from the molecular structure of salts and bases.

Table 2
PLS statistics of the derived PLS models

Model	R^{2}[a]	Q^{2}[b]	N_{PC}[c]	N_{TR}[d]	F[e]	P	$RMSE_{TR}$[f]	$RMSE_{CV}$[g]	N_{TE}[h]	RMSEP.[i]
1	0.969	0.937	3	23	197.97	<0.001	3.16E–02	4.47E–02		
2	0.969	0.937	3	23	197.97	<0.001	3.15E–02	4.47E–02		
3	0.968	0.940	2	23	302.50	<0.001	3.18E–02	4.39E–02		
4	0.963	0.914	2	16	169.18	<0.001	3.55E–02	5.40E–02	7	2.13E–02

[a] Calibration correlation coefficient.
[b] Cross-validated correlation coefficient.
[c] Number of principal components.
[d] Number of salts in the training set.
[e] *F*-value.
[f] Root mean square error for calibration.
[g] Root mean square error for validation.
[h] Number of salts in the test set.
[i] Root mean square error for test set.

The developed salt solubility model was evaluated by applying an external test set. The training set consisted of 16 salts was selected to cover salts that span the variations both in variable descriptors and solubility. The remaining 7 salts were used as an external test set to investigate the predictive power of the derived model. The built model (model 4) is statistically satisfactory with $R^2 = 0.963$, $Q^2 = 0.914$, $F = 169.18$, RMSE = 0.036, and $P < 0.001$. The predictive ability of model 4 is favorable with RMSE for the test set of 0.021. The experimental and calculated/predicted solubility values for models 3 and 4 are summarized in Table 4.

According to the PLS analyses, the most important descriptor influencing the model is molecular weight of the base. In contrast to the solubility of single compounds, in this study salt solubility surprisingly increases with the increasing size of bases. To gain an understanding of this consequence, the interactions of salt solubility should be considered. Typically, when salts dissolve in a solvent, two processes take place. Bonds are broken between the ions, and the detached solute species are then dispersed throughout the solvent medium. The solute species becomes solvated. Thus, for dissolution to occur the forces that bind the pure substance together must be overcome by the forces of solvent–solute interactions. The steric effect from molecular size of the base might be one factor that decreases the binding force of the salt, thereby increasing the solubility. In addition to molecular weight of the base, surface area of the salt also yields the same effect on salt solubility. Binding energy of the

Table 3
Regression coefficients of PLS models

Descriptors	Model 1	Model 2	Model 3	Model 4
Molecular weight of base	3.67E−03	3.67E−03	3.60E−03	3.48E−03
Binding energy of salt	−3.27E−04	−3.27E−04	−2.58E−04	−3.66E−04
HAO[a]	2.69E−04	2.69E−04	2.54E−04	2.19E−04
HAN[b]	4.62E−06	–	–	–
HD[c]	7.23E−05	7.23E−05	–	–
HB[d]	3.45E−04	3.45E−04	3.25E−04	2.58E−04
$c\log P$ of base	−1.48E-04	−1.48E−04	−1.06E−04	−9.49E−05
Surface area of salt	5.65E−04	5.65E−04	5.81E−04	5.86E−04
Melting point of base	9.57E−05	9.57E−05	–	–

[a] Number of hydrogen bond acceptor oxygen atoms.
[b] Number of hydrogen bond acceptor nitrogen atoms.
[c] Number of hydrogen bond donor atoms.
[d] Hydrogen bond formation ability.

Table 4
Experimental, calculated and predicted solubility of salts

No	Salt	Experimental[a]	PLS model of solubility (mM)[b]		
			3	4	
			Calculated	Calculated[c]	Predicted[d]
1	META	1.98	2.08	2.08	
2	DMETA	2.12	2.15		2.16
3	TMETA	2.22	2.21	2.21	
4	TTMETA	2.30	2.27	2.27	
5	DETA	2.30	2.28		2.28
6	DPRA	2.41	2.42	2.41	
7	DBUA	2.48	2.54	2.54	
8	P	2.29	2.26	2.26	
9	PP	2.35	2.33	2.33	
10	M	2.37	2.33	2.33	
11	PZ	2.35	2.33	2.33	
12	MEP	2.35	2.33	2.33	
13	MEPP	2.40	2.38		2.37
14	MEM	2.41	2.38	2.38	
15	MEPZ	2.40	2.39		2.38
16	MEA	2.23	2.21	2.21	
17	DEA	2.42	2.40		2.40
18	TEA	2.53	2.56	2.55	
19	TRIS	2.46	2.46		2.45
20	HEP	2.45	2.46	2.45	
21	HEPP	2.48	2.50	2.50	
22	HEM	2.49	2.50		2.49
23	HEPZ	2.48	2.49	2.49	

[a] Log of experimental solubility (mM).
[b] PLS models: 3 = based on all salts, 4 = based on the selected training set of salts.
[c] Log of calculated/fitted solubility (mM) for the training set of salts.
[d] Log of predicted solubility (mM) for test set of salts.

ion pair of a salt, however, contributes negatively to salt solubility. As the binding energy of the ion pair increases, higher forces are needed for breaking bonds between the ions, resulting in lower salt solubilities. Hydrogen bond descriptors from bases such as HAO or HB can increase the binding interaction between acidic anion and its cation; however, they can also interact with water molecules. The result obtained from PLS regression indicates the positive effect of HAO and HB on the solubility of salt. Water, which is a polar solvent, can better interact with polar solutes, the lipophilicity of base ($c \log P$) therefore has a negative influence on the salt solubility as expected.

In conclusion, derived PLS models with good predictability for organic salt of the same parent compound have been developed in this study. These statistical models are based on simple computed molecular descriptors which account for three aspects of compounds, namely electronic, steric and hydrophobic effects. These obtainable descriptors consist of both typical variables for predicting of single compound solubility, such as size and hydrophobicity/hydrophilicity, and also variables from two species including binding energy and surface area of salt.

References

Aleman, C., Zanuy, D., 2000. A quantum mechanical study of the ionic interactions in model compounds of polyelectrolite-surfactant complexes derived from polypeptides. Chem. Phys. Lett. 319, 318–326.

Chen, X., Cho, S.J., Venkatesh, S., 2002. Prediction of aqueous solubility of organic compounds using a quantitative structure-property relationship. J. Pharm. Sci. 91, 1838–1852.

Chiarini, A., Tartarini, A., Fini, A., 1984. pH-solubility relationship and partition coefficients for some anti-inflammatory arylaliphatic acids. Arch. Pharm. (Weinheim.) 317, 268–273.

Fini, A., Fazio, G., Hervas, M.F., Holgado, M.A., Rabasco, A.M., 1996. Factors governing the dissolution of diclofenac salts. Eur. J. Pharm. Sci. 4, 231–238.

Forbes, R.T., York, P., Davidson, J.R., 1995. Dissolution kinetics and solubilities of *p*-aminosalicylic acid and its salts. Int. J. Pharm. 126, 199–208.

Gao, H., Shanmugasundaram, V., Lee, P., 2002. Estimation of aqueous solubility of organic compounds with QSPR approach. Pharm. Res. 19, 497–503.

Haaland, D.M., Thomas, E.V., 1988. Partial Least-Squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal. Chem. 60, 1193–1202.

Huibers, P.D.T., 1999. Quantum-chemical calculations of the charge distribution in ionic surfactants. Langmuir 15, 7546–7550.

Jorgensen, W.L., Duffy, E.M., 2002. Prediction of drug solubility from structure. Adv. Drug. Deliv. Rev. 54, 355–366.

Lu, L., Chen, C., Sun, K., 1998a. Theoretical study of fullerene derivatives: $C_{28}H_4$ and $C_{28}X_4$ cluster molecules. Int. J. Quant. Chem. 67, 187–197.

Lu, L., Chen, C., Sun, K., 1998b. Theoretical study of fullerene derivatives: $C_{40}H_4$ and $C_{40}X_4$ cluster molecules. Int. J. Quant. Chem. 68, 273–284.

Madhan, B., Thanikaivelan, P., Subramanian, V., Rao, J.R., Nair, B.U., Ramasami, T., 2001. Molecular mechanics and dynamics studies on the interaction of gallic acid with collagen-like peptides. Chem. Phys. Lett. 346, 334–340.

O'Connor, K.M., Corrigan, O.I., 2001. Preparation and characterisation of a range of diclofenac. Int. J. Pharm. 226, 163–179.

Parshad, H., Frydenvang, K., Liljefors, T., Larsen, C.S., 2002. Correlation of aqueous solubility of salts of benzylamine with experimentally and theoretically derived parameters: a multivariate data analysis approach. Int. J. Pharm. 237, 193–207.

Tantishaiyakul, V., 2001. Prediction of caco-2 cell permeability using partial least squares multivariate analysis. Pharmazie 56, 407–411.

Xia, C.Q., Yang, J.J., Ren, S., Lien, E.J., 1998. QSAR analysis of polyamine transport inhibitors in L1210 cells. J. Drug Target. 6, 65–77.

Yaffe, D., Cohen, Y., Epinosa, G., Arenas, A., Giralt, F., 2001. A fuzzy ARTMAP based on quantitative structure-property relationships (QSPRs) for predicting aqueous solubility of organic compounds. J. Chem. Inf. Comput. Sci. 41, 1177–1207.