Prediction of Drug Solubility by the General Solubility Equation (GSE)

Yingqing Ran* and Samuel H. Yalkowsky

Department of Pharmaceutical Sciences, College of Pharmacy, The University of Arizona, 1703 E. Mabel Street, Tucson, Arizona 85721

Received August 15, 2000

The revised general solubility equation (GSE) proposed by Jain and Yalkowsky is used to estimate the aqueous solubility of a set of organic nonelectrolytes studied by Jorgensen and Duffy. The only inputs used in the GSE are the Celsius melting point (MP) and the octanol water partition coefficient (K_{ow}). These are generally known, easily measured, or easily calculated. The GSE does not utilize any fitted parameters. The average absolute error for the 150 compounds is 0.43 compared to 0.56 with Jorgensen and Duffy's computational method, which utilitizes five fitted parameters. Thus, the revised GSE is simpler and provides a more accurate estimation of aqueous solubility of the same set of organic compounds. It is also more accurate than the original version of the GSE.

INTRODUCTION

The aqueous solubility of a drug plays an integral role in many industrial, pharmaceutical, and environmental applications.[1−7] There is a large number of methods to predict aqueous solubility. Unfortunately most of them need many descriptors and have to use regression or some other computational techniques to determine their values. Also different equations are frequently obtained for different sets of data.

The general solubility equation[3,7] has been used extensively in the pharmaceutical field. No fitted parameters are needed for this equation to estimate solubility of nonelectrolytes. It requires only two parameters: the Celsius melting point (MP) and the octanol−water partition coefficient (K_{ow}). Melting point values can be easily obtained from a variety of handbooks and the Website CHEMFINDER. Also melting points can be easily measured. Values of log K_{ow} can be experimentally measured or calculated from the software such as CLOGP. Jorgensen and Duffy[4] recently reported an alternative approach in which a Monte Carlo statistical mechanics simulation (MCSMS) is run for the solute in water. They developed a five-descriptor equation

$$\log S = 0.3158 \times \mathrm{ESXL} + 0.6498 \times \mathrm{HBAC} + 2.192 \times \mathrm{AMINE} - 1.759 \times \mathrm{NITRO} - 161.6 \times \mathrm{HBAC} \times \mathrm{HBDN}^{1/2}/\mathrm{SASA} + 1.181 \quad (1)$$

where ESXL is the Lennard-Jones interaction energy, SASA is the solvent-accessible surface area, HBDN and HBAC are the numbers of solute as donor and acceptor of hydrogen bonds, and AMINE and NITRO are the numbers of amino and nitro groups, respectively. The general solubility equation calculated the molar aqueous solubility by

$$\log S_w^{solid} = 0.5 - 0.01(\mathrm{MP} - 25) - \log K_{ow} \quad (2)$$

The objective of this study is to compare the GSE to the equation derived by Jorgensen and Duffy.[4]

* Corresponding author phone: (520)626-3847; fax: (520)626-4063; e-mail: ran@pharmacy.arizona.edu.

BACKGROUND

The solubility of a solid in water depends on two factors:[3] the crystallinity of the solute and the interaction of the solute with water. The ratio of the solubility of the crystal to the solubility of the liquid is called the ideal solubility, X^{ideal}.

According to the van't Hoff equation,[7−11] the ideal solubility of a solid depends on the Kelvin temperature (T), the Kelvin melting point of the solid (T_m), and on its molar heat of melting (ΔH_m), which is assumed to be independent of temperature.

$$\log X_u^{ideal} = -\Delta H_m \frac{(T_m - T)}{2.303 R T_m T} \quad (3)$$

Since the solid and the melt are in equilibrium at the melting point, $\Delta H_m/T_m$ can be replaced by ΔS_m, the entropy of melting. This gives

$$\log X_u^{ideal} = -\Delta S_m \frac{(T_m - T)}{2.303 R T} \quad (4)$$

which, at 298 K, becomes

$$\log X_u^{ideal} = -\Delta S_m \frac{(T_m - 298)}{5706} \quad (5)$$

According to Walden's rule,[8−11] ΔS_m=56.5 J/ degree mole for most organic compounds, so that at 25 °C eq 5 becomes

$$\log X_u^{ideal} = -0.01(T_m - 298) = -0.01(\mathrm{MP} - 25) \quad (6)$$

Equation 6 indicates the ideal solubility at room temperature as a simple function of the melting point. The ideal solubility can only be used to estimate the solubility of a solute in an ideal solvent, i.e., in a solvent of similar polarity, such as an ideal solvent.

The logarithm of the octanol−water partition coefficient can be used to account for the difference between an ideal solution and an aqueous solution. The octanol−water partition coefficient is defined as

10.1021/ci000338c CCC: $20.00
Published on Web 01/12/2001

Table 1. Aqueous Solubilities for the Selected Compounds

name	MP	ClogP calcd	MlogP obsd	log $S^{1,2,5}$ obsd	log *S* (GSE) CLOGP	log *S* (GSE) MLOGP	log *S* MCSMC
diethylstilbestrol	169	4.96	5.07	−4.07	−5.90	−6.01	−4.39
fenclofenac	136	4.71	4.80	−3.85	−5.32	−5.41	−4.38
camphor	177	2.18	2.38	−1.96	−3.20	−3.40	−2.24
ethylamine	25	−0.13	−0.13	2.06	0.63	0.63	2.25
2,3-dichlorophenol	59	2.84	2.84	−1.30	−2.68	−2.68	−2.27
salicylic acid	159	2.19	2.26	−1.82	−3.03	−3.10	−0.91
p−*tert*-butylphenol	98	3.30	3.31	−2.41	−3.53	−3.54	−2.72
phenol	41	1.47	1.47	0.00	−1.13	−1.13	−1.02
trimethylamine	25	0.02	0.16	1.32	0.48	0.34	1.13
pyridine	25	0.64	0.65	0.76	−0.14	−0.15	−0.48
2-naphthol	122	2.65	2.70	−2.28	−3.12	−3.17	−2.25
m-nitrobenzoic acid	142	1.84	1.83	−1.68	−2.51	−2.50	−2.14
benzoic acid	122	1.88	1.87	−1.55	−0.27	−2.34	−0.90
p-cresol	33	1.97	1.94	−0.73	−1.55	−1.52	−1.39
flurbiprofen	111	3.75	4.16	−3.74	−4.11	−4.52	−3.62
desipramine	25	4.47	4.90	−3.66	−3.97	−4.40	−3.99
indole	53	2.13	2.14	−1.21	−1.91	−1.92	−2.37
caffeine	238	−0.06	−0.07	−0.88	−1.57	−1.56	0.08
propanal	25	0.30	0.59	0.58	0.20	−0.09	0.31
morpholine	25	−0.41	−0.86	1.97	0.91	1.36	1.92
adenine	363	−0.29	−0.09	−2.18	−2.59	−2.79	−1.22
phenobarbitol	174	1.90	1.90	−2.29	−2.89	−2.89	−2.10
theophylline	272	−0.06	−0.02	−1.39	−1.91	−1.95	−0.86
acetamide	81	−1.11	−1.09	1.58	1.05	1.03	−0.12
ethyl-*p*-hydroxybenzoate	117	2.51	2.47	−2.35	−2.93	−2.89	−1.94
phenytoin	295	2.08	2.28	−3.99	−4.28	−4.48	−3.13
2-methyl-2-propanol	26	0.47	0.35	0.63	0.03	0.15	−0.45
indomethacin	158	4.18	4.27	−4.62	−5.01	−5.10	−4.81
lidocaine	68	1.95	2.26	−1.71	−1.88	−2.19	−2.22
diuron	159	2.68	2.68	−3.05	−3.52	−3.52	−3.81
tetrahydrofuran	25	0.53	0.47	0.49	−0.03	0.03	0.14
acetaminophen	169	0.49	0.51	−1.02	−1.43	−1.45	−1.60
codeine	155	0.98	1.14	−1.52	−1.78	−1.94	−1.98
barbital	190	0.66	0.65	−1.42	−1.81	−1.80	−1.97
1-propanol	25	0.29	0.25	0.62	0.21	0.25	−0.05
succinic acid	188	−0.53	−0.59	−0.19	−0.60	−0.54	−0.28
nevirapine[6]	248	2.35	1.81	−3.19	−4.08	−3.54	−4.29
limonene	74	4.35	4.35	−4.00	−4.34	−4.34	−4.02
p-chloroacetanilide	178	2.13	2.12	−2.84	−3.16	−3.15	−2.23
butanone	25	0.32	0.29	0.52	0.18	0.21	0.20
diethyl ether	25	0.87	0.89	−0.09	−0.37	−0.39	−0.09
dexamethasone	263	1.75	2.01	−3.59	−3.63	−3.89	−4.50
ethanol	25	−0.24	−0.31	1.10	0.74	0.81	0.29
methanol	25	−0.76	−0.77	1.56	1.26	1.27	0.55
cocaine	98	2.57	2.30	−2.25	−2.80	−2.53	−1.51
hexamethylbenzene	164	4.99	4.61	−5.23	−5.88	−5.50	−4.46
prostaglandin E2	67	2.01	2.82	−2.47	−1.93	−2.74	−4.01
perphenazine	97	4.32	4.20	−4.16	−4.54	−4.42	−3.90
quinoline	25	2.03	2.03	−1.30	−1.53	−1.53	−2.67
acetanilide	114	1.16	1.16	−1.33	−1.55	−1.55	−1.58
benzamide	129	0.65	0.64	−0.96	−1.19	−1.18	−1.18
procaine	61	2.54	2.14	−1.78	−2.40	−2.00	−2.24
p-chloroaniline	73	1.91	1.88	−1.66	−1.89	−1.86	−1.85
ethyl acetate	25	0.71	0.73	−0.04	−0.21	−0.23	0.13
dimethoxymethane	25	−0.43	0.18	0.48	0.93	0.32	0.62
ketoprofen	94	2.76	3.12	−3.16	−2.95	−3.31	−3.35
trichloroethene	25	2.63	2.61	−1.96	−2.13	−2.11	−2.01
methyl acetate	25	0.18	0.18	0.46	0.32	0.32	0.63
atropine	116	1.32	1.83	−2.12	−1.73	−2.24	−2.03
testosterone	155	3.22	3.32	−4.02	−4.02	−4.12	−5.02
dimethyl sulfide	25	0.84	1.05	−0.45	−0.34	−0.55	−0.62
DDT	109	6.76	6.91	−7.15	−7.10	−7.25	−7.38
ephedrine	38	0.89	0.93	−0.47	−0.52	−0.56	−0.72
aspirin	135	1.02	1.19	−1.72	−1.62	−1.79	−1.21
1-chloropropane	25	1.99	2.04	−1.47	−1.49	−1.54	−0.84
chlorpromazine	57	5.80	5.35	−5.10	−5.62	−5.17	−5.41
thioridazine	73	6.50	5.90	−5.82	−6.48	−5.88	−6.20
2-methylnevirapine[6]	235	2.73	2.73	−4.27	−4.33	−4.33	−4.38
atrazine	175	2.50	2.61	−3.55	−3.50	−3.61	−3.02
benzonitrile	25	1.57	1.56	−1.00	−1.07	−1.06	−1.12
thiophene	25	1.79	1.89	−1.33	−1.29	−1.39	−1.54
uracil	335	−1.06	−1.07	−1.49	−1.54	−1.53	−0.71
nevirapine analogue 1[6]	250	3.43	3.43	−5.15	−5.18	−5.18	−4.47
furan	25	1.32	1.34	−0.82	−0.82	−0.84	−0.56
toluene	25	2.64	2.73	−2.21	−2.14	−2.23	−2.15
trifluoperazine	25	5.21	5.03	−4.52	−4.71	−4.53	−4.23

Table 1. (Continued)

name	MP	ClogP calcd	MlogP obsd	log $S^{1,2,5}$ obsd	log *S* (GSE) CLOGP	log *S* (GSE) MLOGP	log *S* MCSMC
chlorobenzene	25	2.86	2.89	−2.38	−2.36	−2.39	−1.82
7-methylpteridine	197	−0.36	−0.36	−0.85	−0.86	−0.86	−1.80
trifluoromethylbenzene	25	3.03	3.01	−2.51	−2.53	−2.51	−2.58
1,1,1-trichloroethane	25	2.48	2.49	−2.00	−1.98	−1.99	−2.00
aniline	25	0.91	0.90	−0.41	−0.41	−0.40	−1.30
benzene	25	2.14	2.13	−1.64	−1.64	−1.63	−1.54
menthone	25	2.83	2.83	−2.35	−2.33	−2.33	−2.60
2-propanol	25	0.07	0.05	0.43	0.43	0.45	−0.18
progesterone	127	3.77	3.87	−4.42	−4.29	−4.39	−4.54
methyl butyrate	25	1.24	1.29	−0.82	−0.74	−0.79	−0.58
fluorobenzene	25	2.28	2.27	−1.80	−1.78	−1.77	−1.73
chloramphenicol	151	1.28	1.14	−1.94	−2.04	−1.90	−3.47
hydrocortisone	218	1.70	1.61	−3.09	−3.13	−3.04	−4.08
bromobenzene	25	3.01	2.99	−2.55	−2.51	−2.49	−2.62
allopurinol	350	−0.88	−0.55	−2.26	−1.87	−2.20	−1.05
propionitrile	25	0.13	0.16	0.28	0.37	0.34	0.85
naproxen	153	2.82	3.34	−4.20	−3.60	−4.12	−2.98
1,2-dichloroethane	25	1.46	1.47	−1.06	−0.96	−0.97	−0.53
pent-1-yne	25	1.98	1.98	−1.64	−1.48	−1.48	−1.10
N,*N*-dimethylacetamide	25	−0.80	−0.77	1.11	1.30	1.27	0.09
cyclohexane	25	3.35	3.44	−3.10	−2.85	−2.94	−2.34
bromazepam	238	1.69	1.69	−3.48	−3.32	−3.32	−4.45
phenacetin	135	1.77	1.58	−2.35	−2.37	−2.18	−2.49
promazine	33	4.90	4.55	−4.30	−4.48	−4.13	−4.40
estradiol	173	3.78	3.86	−5.03	−4.76	−4.84	−4.71
acetophenone	25	1.58	1.58	−1.28	−1.08	−1.08	−1.37
2,3,4,5,6-pentachlorobiphenyl	124	7.09	7.09	−7.78	−7.58	−7.58	−6.64
benzaldehyde	25	1.50	1.47	−1.19	−1.00	−0.97	−0.87
perchlorobiphenyl	306	9.92	8.27	−10.80	−12.23	−10.58	−9.20
cyclohexene	25	2.87	2.86	−2.59	−2.37	−2.36	−2.14
methyl benzoate	25	2.11	2.12	−1.85	−1.61	−1.62	−1.58
corticosterone	181	2.32	1.94	−3.24	−3.38	−3.00	−4.62
anisole	25	2.06	2.11	−1.85	−1.56	−1.61	−2.00
naphthalene	81	3.32	3.30	−3.60	−3.38	−3.36	−3.21
imipramine	25	5.04	4.44	−4.19	−4.54	−3.94	−4.83
ibuprofen	76	3.68	3.50	−3.76	−3.69	−3.51	−3.35
triflupromazine	25	6.11	5.54	−5.30	−5.61	−5.04	−5.39
diazepam	125	3.16	2.99	−3.75	−3.66	−3.49	−3.98
p-toluenesulfonamide	139	0.80	0.82	−1.74	−1.44	−1.46	−0.44
bifonazole	142	4.77	4.99	−5.95	−5.44	−5.66	−5.67
lorazepam	167	2.36	2.39	−3.60	−3.28	−3.31	−3.65
pentane	25	3.34	3.39	−3.18	−2.84	−2.89	−2.16
benzocaine	89	1.92	1.86	−2.32	−2.06	−2.00	−2.64
alanine	316	−3.12	−2.98	0.25	0.72	0.58	1.38
biphenyl	70	4.03	4.01	−4.35	−3.98	−3.96	−4.08
dibenzofuran	82	4.09	4.12	−4.60	−4.16	−4.19	−3.89
fluorene	116	4.07	4.18	−5.00	−4.48	−4.59	−4.33
oxazepam	205	2.29	2.24	−3.95	−3.59	−3.54	−3.72
cytosine	323	−1.85	−1.73	−1.16	−0.63	−0.75	−0.91
prednisone	234	1.62	1.47	−3.48	−3.21	−3.06	−2.93
hexane	25	3.87	3.90	−3.84	−3.37	−3.40	−2.46
griseofulvin	220	1.75	2.18	−4.07	−3.20	−3.63	−2.31
nitrobenzene	25	1.88	1.85	−1.80	−1.38	−1.35	−0.94
pyrene	156	4.95	4.88	−6.18	−5.76	−5.69	−5.27
lindane	113	3.75	3.72	−4.60	−4.13	−4.10	−5.16
nitroethane	25	0.25	0.18	−0.22	0.25	0.32	0.03
carvone	25	2.01	2.01	−2.06	−1.51	−1.51	−2.22
nitromethane	25	−0.28	−0.35	0.26	0.78	0.85	0.55
fenoxycarb	54	4.46	4.30	−4.70	−4.25	−4.09	−4.88
indoprofen	214	2.74	2.77	−4.82	−4.13	−4.16	−4.22
fluconazole	139	0.47	0.47	−1.80	−1.11	−1.11	−2.35
warfarin	161	2.89	2.70	−4.26	−3.75	−3.56	−5.28
morphine	254	0.59	0.76	−3.28	−2.38	−2.55	−2.76
desmedipham	120	3.40	3.39	−4.63	−3.85	−3.84	−4.71
benzidine	128	1.58	1.34	−2.70	−2.11	−1.87	−3.07
sulindac	184	3.16	3.05	−5.00	−4.25	−4.14	−3.98
nevirapine analogue 12[6]	131	1.17	1.17	−2.62	−1.73	−1.73	−3.06
nifedipine	173	3.41	2.86	−4.76	−4.39	−3.84	−4.01
fenbufen	186	3.14	3.20	−5.30	−4.25	−4.31	−4.08
nifuroxime	162	1.10	0.30	−2.19	−1.97	−1.17	−2.46
mannitol	167	−2.05	−2.65	0.06	1.13	1.73	−1.25
guanine	360	−1.28	−0.96	−3.58	−1.57	−1.89	−1.99
nitrofurantoin	268	−0.47	−0.47	−3.38	−1.46	−1.46	−3.01
dimethylbarbiturate	123	−0.78	−0.78	−1.74	0.31	0.31	−1.08
absolute av error					0.45	0.43	0.56
root-mean-square error					0.62	0.52	0.70

$$K_{ow} = \frac{C_o}{C_w} \tag{7}$$

where C_o and C_w are the equilibrium dilute molar concentrations of the solute in octanol and water, respectively. If the effects of mutual saturation are small, this can be approximated as

$$K_{ow} = \frac{S_o}{S_w} \tag{8}$$

or

$$\log S_w = \log S_o - \log K_{ow} \tag{9}$$

where S_o and S_w are the solubilities of the substance in octanol and water, respectively. Jain and Yalkowsky[3] showed that most un-ionized organic liquids are completely miscible with octanol and that complete miscibility is expressed as $X_o = 0.5$, so the molar solubility, S_o, of the solute in octanol is

$$S_o^{liquid} = \frac{6.3}{2} = 3.15 \text{ or } \log S_o^{liquid} = 0.5 \tag{10}$$

where 6.3 is the molarity of pure octanol. Substituting eq 10 into eq 9 gives an expression for the estimation the aqueous solubility of a liquid S_w^{liquid}

$$\log S_w^{liquid} = 0.5 - \log K_{ow} \tag{11}$$

The solubility of the crystal in water ($S_w^{crystal}$) is equal to the solubility of the liquid in water multiplied the ideal solubility.

$$\log S_w^c = \log\left(S_w^{liquid}\frac{X^{crystal}}{X^{liquid}}\right) = \log S_w^1 + \log X_u^{ideal} \tag{12}$$

Inserting eq 11 and eq 6 into eq 12 gives the general solubility equation, i.e., and eq 2. This equation can be easily used to predict the water solubility of most organic compounds. The term (MP-25) must be set equal to zero for compounds that are liquid at room temperature, so that eq 2 reduces to eq 11. (Note that the original GSE utilized a constant of 0.8 which corresponds to complete miscibility being described by unit mole fraction, i.e., by $S_o^{liquid} = 0.630$.)

EXPERIMENTAL SECTION

All 150 compounds used by Jorgensen and Duffy[4] were used to test the GSE. Melting points were obtained from The Merck Index and CHEMFINDER. Measured and calculated partition coefficients (MLOGP and CLOGP, respectively) were obtained from the CLOGP software. Solubilities were calculated from both MLOGP and CLOGP by the general solubility equation. The absolute average error for each calculation was determined by

$$AAE = \frac{\sum|\log S_{calc} - \log S_{obs}|}{150} \tag{13}$$

and the root-mean-square errors by

$$RSME = \frac{\sqrt{\sum(\log S_{calc} - \log S_{obs})^2}}{150} \tag{14}$$

RESULTS

The values of MP, CLOGP, MLOGP, and log *S* are given in Table 1, along with the values calculated by the GSE (eq 2) and by the MCSMC equation of Jorgensen and Duffy (eq 1). The absolute average errors and root-mean-square errors are given in the last two lines of Table 1.

It is clear from Table 1 that the GSE (with only two unfitted parameters) provides more accurate prediction of aqueous solubility than the Monte Carlo based on eq 1 (which utilitizes five regression fitted parameters). It is also clear that, for the present data set, the GSE gives similar results whether experimentally measured or calculated partition coefficients are used. (The use of the original version of GSE, which predicts solubilities that are 2-fold higher than predicted by eq 2, gives an average absolute error of 0.49 and root-mean-square error of 0.62.) The success of the GSE is due to the fact that the partition coefficient is inversely proportional to the solubility of an organic liquid and that the melting point term accurately accounts for the effects of solute crystallinity. It is interesting to note that equations (such as eq 1) which do not contain a melting point term can often give reasonable solubility estimates. This is because melting points and the partition coefficient can be somewhat intercorrelated for small data sets. However, with large and diverse data set such as used by Jain and Yalkowsky,[3] the importance of the melting point term becomes apparent.

CONCLUSIONS

Based on the data set selected by Jorgensen and Duffy,[4] it appears that the general solubility equation is simpler and easier to use than the Monte Carlo simulation equation for estimations of the aqueous solubility of un-ionized drugs.

REFERENCES AND NOTES

(1) Abraham, M. H.; Le, J. The Correlation and Prediction of the Solubility of Compounds in water Using an Amended Solvation Energy Relationship. *J. Pharm. Sci.* **1999**, *89*, 868−880.
(2) Huuskonen, J.; Salo, M.; Taskinen, J. Aqueous Solubility Prediction of Drugs Based on Molecular Topology and neural Network Modeling *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 450−456.
(3) Jain, N.; Yalkowsky, S. H. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J. Pharm. Sci.* **2001**, *90*(2), 234−252.
(4) Jorgensen, W. L.; Duffy, E. M. Prediction of Drug Solubility from Monte CarloSimultaions *Bioorg. Medicinal Chem. Lett.* **2000**, *10*, 1155−1158.
(5) Mitchell, B. E.; Jurs, P. C. Prediction of Aqueous Solubility of Organic Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489−496.
(6) Morelock, M. M.; Choi, L. L.; Bell, G. L.; Wright, J. L. Estimation and Correlationof Drug Water Solubility with Pharmacological Parameters requried for Biological Activity. *J. Pharm. Sci.* **1994**, *83*, 948−951.
(7) Yalkowsky, S. H. *Solubility and solubilization in Aqueous Media*; Oxford University: Oxford, 1999.
(8) Yalkowsky, S. H.; Valvani, S. C. Solubility and Partitioning I: Solubility ofNonelectrolytes in Water. *J. Pharm. Sci.* **1980**, *69*, 912−922.
(9) Yalkowsky, S. H.; Sinkula, A. A.; Valvani, S. C. *Physical Chemical Properties of drugs*; Marcel Dekker: New York, 1980.
(10) Yalkowsky, S. H.; Pinal, R. Estimation of the Aqueous Solubiltiy of ComplexOrganic Molecules. *Chemosphere* **1993**, *26*(7), 1239−1261.
(11) Yalkowsky, S. H.; Mishra, D. S. Ideal Solubility of a Solid Solute: Effect of HeatCapacity Assumptions. *Pharm. Res.* **1992**, *9*, 958−959.

CI000338C