

# Aqueous solubility of a simple (single-carbon) organic molecule as a function of its size & dipole moment

Kamal I. Al-Malah

Received: 20 May 2010 / Accepted: 6 July 2010 / Published online: 22 July 2010  
© Springer-Verlag 2010

**Abstract** The aqueous solubility of a single-carbon organic molecule as a function of its size & dipole moment was investigated. The molecular dipole moment was chosen to represent the polar character of a poly-atomic molecule. It is hypothesized here that at a given pH, temperature, and pressure, the solubility of a single-carbon organic molecule in water will be a function of its polar character; namely, dipole moment and of its molecular size. Different forms of the solubility function were tested; it was found that the solubility model, given by Eq. 1, which is based on the polar character and the molecular volume, adequately described the aqueous solubility of single-carbon organic moieties. The aqueous solubility of single-carbon organic solutes exhibits maximum at the condition of high polar character (large dipole moment) and low molecular volume. The general trend of the solubility of single-carbon organic solutes, based on the proposed model (Eq. 1) could be explained in terms of the trade-off between the driving force (degree of polar character of the solute) for solubilization versus the resistance to be solubilized as a result of the entropic effects which increase with increasing molecular volume of the organic moiety.

**Keywords** Aqueous solubility · Dipole moment · Entropic effects · Polar · Single-carbon organic solute · Van der Waals volume

## Introduction

Van Noort [1] proposed an estimation procedure for the aqueous solubility of some solid organic compounds; he derived a general compound-class specific estimation procedure for solid organic compound aqueous solubility and vapor pressure. The estimation procedure used a linear combination of simple molecular descriptors for the molecular structure variation within the compound class and a polynomial for the temperature of melting. His procedure was applied to the aqueous solubility of polycyclic aromatic hydrocarbons (PAHs), methylated PAHs, chlorinated benzenes, polychlorinated and poly-brominated biphenyls, chlorinated phenols, cresols, and chlorinated 2-methoxyphenols.

Ruckenstein and Shulgin [2] examined the effect of salts and organic additives on the solubility of proteins in aqueous mixed solvents. Their focus was on the correlation between the aqueous protein solubility and the osmotic second virial coefficient or the preferential binding parameter. The correlation of the aqueous protein solubility with the osmotic second virial coefficient when the co-solvent was an organic component required additional research. Their theory could predict the salting-in or -out effect of a co-solvent and good agreement was obtained between theoretical predictions and experimental results.

Tantishaiyakul [3] predicted the aqueous solubility of organic salts of diclofenac using computed molecular descriptors and multivariate partial least squares (PLS). The molecular descriptors included binding energy and surface area of salts. 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 permitted simple

K. I. Al-Malah (✉)  
Department of Chemical Engineering, University of Hail,  
Hail, Saudi Arabia  
e-mail: almalak61@hotmail.com

computational prediction of salt solubility of a same parent structure. Molecular weight of base was considered as 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 were found to have negative effect on salt solubility.

Ran et al. [4] used a general solubility equation (GSE) proposed by Jain and Yalkowsky [5] to estimate aqueous solubility of 1026 non-electrolytes. The parameters used in the GSE were melting points (MP) and octanol–water partition coefficients ( $K_{ow}$ ). No fitted parameters and no training set were employed in the GSE. The experimental aqueous solubility values were obtained from the AQUASOL dATABASE [6].

In the present article, the aqueous solubility of a single-carbon organic molecule as a function of its size & dipole moment is investigated.

## Theory

It is hypothesized here that at a given pH, temperature, and pressure, the solubility of a single-carbon organic molecule in water will be a function of its polar character; namely, dipole moment and of its molecular size. In a previous work [7], the aqueous solubility of a diatomic molecule was expressed as a function of its size and the electronegativity difference of the constituting atoms. Here, the electronegativity difference is replaced by the molecular dipole moment, for the latter is a better representation of the polar character of a poly-atomic molecule.

Even though the total charge on a molecule is zero, the nature of chemical bonds is such that the positive and negative charges do not completely overlap in most molecules. In other words, a non-uniform distribution of positive and negative charges on the various atoms, due to the difference in electronegativity of constituting atoms, forms the basis for existence of such permanent dipole moments. Such molecules are said to be polar because they possess a permanent dipole moment. A good example is the dipole moment of the water molecule where electron density is shared unequally between the constituting atoms hydrogen and oxygen.

On the contrary, molecules with mirror symmetry like oxygen, nitrogen, carbon dioxide, and carbon tetrachloride have no permanent dipole moments. Even if there is no permanent dipole moment, it is possible to induce, however, a dipole moment by the application of an external electric field. This is called polarization and the magnitude of the dipole moment induced is a measure of the polarizability of the molecular species.

A popular proverb used for predicting solubility is "like dissolves like" [8]. This statement indicates that a solute will dissolve best in a solvent that has a similar polarity to itself. This view is rather simplistic, since it ignores many solvent-solute interactions, but it is a useful rule of thumb. For example, a very polar (hydrophilic) solute such as urea is very soluble in highly polar water, less soluble in fairly polar methanol, and practically insoluble in non-polar solvents such as benzene. In contrast, a non-polar or hydrophobic solute such as naphthalene is insoluble in water, fairly soluble in methanol, and highly soluble in non-polar benzene [9]. Liquid solubilities also generally follow this rule. Hydrophobic plant oils, such as olive oil and palm oil, dissolve in non-polar solvents such as alkanes, but are less soluble in polar liquids such as water. As aqueous solubility of alkanes is very scarce, it is usually expressed in part of solute per million parts of solution (ppm), or even part per billion (ppb).

The physical chemist Peter J. W. Debye was the first scientist to study molecular dipoles extensively, and, as a consequence, dipole moments are measured in units named debye (D) in his honor. The debye is a CGS unit (a non-SI metric unit) of electric dipole moment, which is defined as  $1 \times 10^{-18}$  statcoulomb\*centimeter (esu\*cm) or  $3.336 \times 10^{-30}$  Coulomb\*meter. Historically, the debye was defined as the dipole moment resulting from two charges of opposite sign but an equal magnitude of  $10^{-10}$  statcoulomb (or, called electrostatic unit, e.s.u.), which were separated by 1 ångström ( $10^{-8}$  cm). The dipole moment involves the first moment of the electric charge density of the compound. Dipoles can be characterized by their dipole moment, a vector quantity. For the simplest form of electric dipole moment, it would point from the negative charge toward the positive charge, and have a magnitude equal to the strength of each charge times the separation between the charges.

On the other hand, van der Waals' volume represents the envelope volume occupied by the atoms making up the molecule, which takes into account the atomic volumetric contribution of all constituting atoms corrected by the presence of bonds and the aromatic and non-aromatic rings. In other words, it represents the volume enclosed by the van der Waals surface that may be visualized as a wrapping film surrounding the molecule from all sides.

Liquids (and solids) differ from gases in that the molecules of the liquid (or solid) are held together by a certain amount of intermolecular stickiness. For a solution to occur, the solvent molecules must overcome this intermolecular stickiness in the solute and find their way between and around the solute molecules. At the same time, the solvent molecules themselves must be separated from each other by the molecules of the solute. This is accomplished best when the attractions between the molecules of both components are similar. If the attractions

**Table 1** The dipole moment (D), the aqueous solubility ( $\frac{\text{g solute}}{\text{kg water}}$ ), and the van der Waals' volume ( $\text{cm}^3 \text{mol}^{-1}$ ) for single-carbon organic solutes<sup>+</sup>

Molecular formula	Material name	Dipole moment (D)	Aqueous sol. ( $\frac{\text{g solute}}{\text{kg water}}$ )	Van der Waals volume ( $\text{cm}^3/\text{mol}$ )
CBrF <sub>3</sub>	bromotrifluoromethane	0.65	0.3	35.33
CClF <sub>3</sub>	chlorotrifluoromethane	0.51	0.09	33.57
CClN	cyanogen chloride	2.82	30	26.32
CCl <sub>2</sub> F <sub>2</sub>	dichlorodifluoromethane	0.51	0.3	39.81
CCl <sub>2</sub> O	phosgene	1.17	6.8255	34.9
CCl <sub>3</sub> F	trichlorofluoromethane	0.45	1.1	46.05
CCl <sub>4</sub>	carbon tetrachloride	0	0.7934	52.3
CF <sub>4</sub>	carbon tetrafluoride	0	0.016	27.33
CHBr <sub>3</sub>	tribromomethane	0.99	3.11	49.98
CHClF <sub>2</sub>	chlorodifluoromethane	1.42	2.77	31.02
CHCl <sub>2</sub> F	dichlorofluoromethane	1.29	18.8	35.74
CHCl <sub>3</sub>	chloroform	1.01	7.5	43.5
CHF <sub>3</sub>	trifluoromethane	1.65	0.9	23.94
CHI <sub>3</sub>	triiodomethane	0.9	0.1	64.32
CHN	hydrogen cyanide	2.98	1000	18.13
CH <sub>2</sub> BrCl	bromochloromethane	1.66	15	37.07
CH <sub>2</sub> Br <sub>2</sub>	dibromomethane	1.43	12	39.43
CH <sub>2</sub> ClF	chlorofluoromethane	1.82	10.52	27.57
CH <sub>2</sub> Cl <sub>2</sub>	dichloromethane	1.6	19.4	34.71
CH <sub>2</sub> F <sub>2</sub>	difluoromethane	1.96	4.4	21.67
CH <sub>2</sub> I <sub>2</sub>	diiodomethane	1.22	1.24	50.93
CH <sub>2</sub> O	formaldehyde	2.33	1000	18.6
CHOOH	formic acid	1.42	1000	22.74
CH <sub>3</sub> Br	methyl bromide	1.81	13.4	28.07
CH <sub>3</sub> Cl	methyl chloride	1.87	5.9	25.29
CH <sub>3</sub> I	methyl iodide	1.62	26.1	32.85
CH <sub>3</sub> NO	formamide	3.73	1000	25.46
CH <sub>3</sub> NO <sub>2</sub>	nitromethane	3.46	10	30.47
CH <sub>4</sub>	methane	0	0.0244	17.05
CH <sub>3</sub> OH	methanol	1.7	1000	21.71
CH <sub>3</sub> SH	methyl mercaptan	1.52	24	28.48
CH <sub>3</sub> NH <sub>2</sub>	methylamine	1.31	519.2	24.21
CO	carbon monoxide	0.11	0.02377	16.2
COS	carbonyl sulfide	0.71	1.15	25.9
CO <sub>2</sub>	carbon dioxide	0	1.95	19.7
CS <sub>2</sub>	carbon disulfide	0	1.879	31.2

<sup>+</sup>Data were taken from Yaws [9].

are sufficiently different, the strongly attracted molecules will cling together, excluding the weakly attracted molecules, and immiscibility (not able to be mixed) will result. Oil and water do not mix because the water molecules, strongly attracted to each other, will not allow the weakly attracted oil molecules between them. Such sticky forces between molecules are called *van der Waals forces* (after Johannes van der Waals who first described them in 1873).

Tolls et al. [10] examined the aqueous solubility–molecular size relationships. For a given n-alkane in

aqueous solution, the all-trans conformation is preferred over folded geometries. Within alkanes, molecular size is the primary determinant of their solubility in water, and increasing molecular size results in a decrease in water solubility mainly due to the increased free energy penalty (i.e.,  $\Delta G$  becomes more positive with increasing volume of the solute) for cavity formation in water.

Zhao et al. [11] proposed a new method for calculating van der Waals volume based on Bondi radii. The method, termed atomic and bond contributions of van der Waals

**Table 2** Statistical parameters associated with the solubility model (Eq. 1) as well as other models, for all single-carbon organic molecules shown in Table 1. Solubility data are @ 25°C

Equation type	SSE♦	R <sup>2</sup>	# of Observations	D.o.F*	RMSE <sup>+</sup>
$Z = a + b^*X + c^*Y + d^*(Y/X)$	5.594e+005	0.8716	36	32	132.2
$Z = a + b^*X + c^*Y + d^*(Y)^2$	2.956e+006	0.3217	36	32	304
$Z = a + b^*X + c^*(X)^2 + d^*Y$	2.59e+006	0.4057	36	32	284.5
$Z = a + \frac{b}{X} + c^*Y + d^*(Y)^2$	2.699e+006	0.3807	36	32	290.4
$Z = a + b^*X + c^*(X)^2 + d/Y$	3.054e+006	0.2994	36	32	308.9
$Z = a + \frac{b}{X} + c^* \log(Y) + d^* \log(Y)^2$	2.683e+006	0.3845	36	32	289.6
$Z = a + \frac{b}{X} + c^*Y + d^*(Y)^2 + e^*(Y)^3$	2.553e+006	0.4143	36	31	287
$Z = a + b^*X + c^*Y + d^*(X)^2 + e^*Y^2$	2.59e+006	0.4058	36	31	289
$Z = a + \frac{b}{X} + c^*Y + \frac{d}{(X)^2} + e^*Y^2$	2.659e+006	0.3900	36	31	292.9
$Z = a + b^*X + c^* \log(Y) + d^*X^2 + e^* \log(Y)^2$	2.573e+006	0.4096	36	31	288.1

♦SSE: Sum of Square Errors =  $\sum_{i=1}^N (Z_{\text{exp}} - Z_{\text{fit}})^2$

D.o.F.: Degree of Freedom = # of data points - # of model parameters

<sup>+</sup> RMSE: Root Mean Square Error =  $\sqrt{\frac{\text{SSE}}{\text{D.o.F.}}}$

volume (VABC), is very simple and fast. The only information needed for calculating VABC is atomic contributions and the number of atoms, bonds, and rings. Then, the van der Waals volume ( $\text{\AA}^3/\text{molecule}$ ) can be calculated from the following formula:  $V_{\text{vdW}} = \sum$  all atom contributions  $-5.92N_{\text{B}} - 14.7R_{\text{A}} - 3.8R_{\text{NR}}$  ( $N_{\text{B}}$  is the number of bonds,  $R_{\text{A}}$  is the number of aromatic rings, and  $R_{\text{NA}}$  is the number of non-aromatic rings). The number of bonds present ( $N_{\text{B}}$ ) can be simply calculated by  $N_{\text{B}} = N - 1 + R_{\text{A}} + R_{\text{NA}}$  (where  $N$  is the total number of atoms). A simple Excel spread sheet has been made to calculate van der Waals volumes for a wide range of 677 organic compounds, including 237 drug compounds.

Table 1 shows the dipole moment ( $D$ ), the aqueous solubility ( $\frac{\text{g solute}}{\text{kg water}}$ ), and the van der Waals' volume ( $\text{cm}^3 \text{mol}^{-1}$ ) for single-carbon organic solutes. The data were taken from [9].

## Results and discussion

The MATLAB<sup>®</sup> surface fitting tool was used to fit a multi-dimensional, non-linear regression problem as is the case here. The general formula for curve-fitting is:  $Z = f(X, Y)$  where  $Z$  is the dependent variable and  $X$  &  $Y$  are the independent variables.

**Table 3** Curve-fitted parameters of the solubility model (Eq. 1), accompanied by the 95% confidence interval for single-carbon organic molecules. Data points are @ 25°C

a±95% C.I.	b±95% C.I.	c±95% C.I.	d±95% C.I.
-142.8 ±193	5.225±5.5	-437±171	1.43e+004±4110

The solubility ( $Z$ ) was fitted as a function of both the van der Waals' volume ( $X$ ) and the molecular dipole moment ( $Y$ ) as given in Table 1. Different models were attempted and the best result is shown below. In general, the following model was found to be the most adequate in terms of model goodness:

$$Z = a + b^*X + c^*Y + d^*(Y/X). \quad (1)$$

Table 2 shows that the correlation coefficient,  $R^2$ , is high for Eq. 1, compared with other models.

The curve-fitted parameters of the solubility model (Eq. 1), being applied to single-carbon organic molecules while accompanied by the 95% confidence interval, are shown in Table 3.

Figure 1 shows the aqueous solubility of single-carbon organic solutes as a function of the dipole moment and the van der Waals' molar volume.

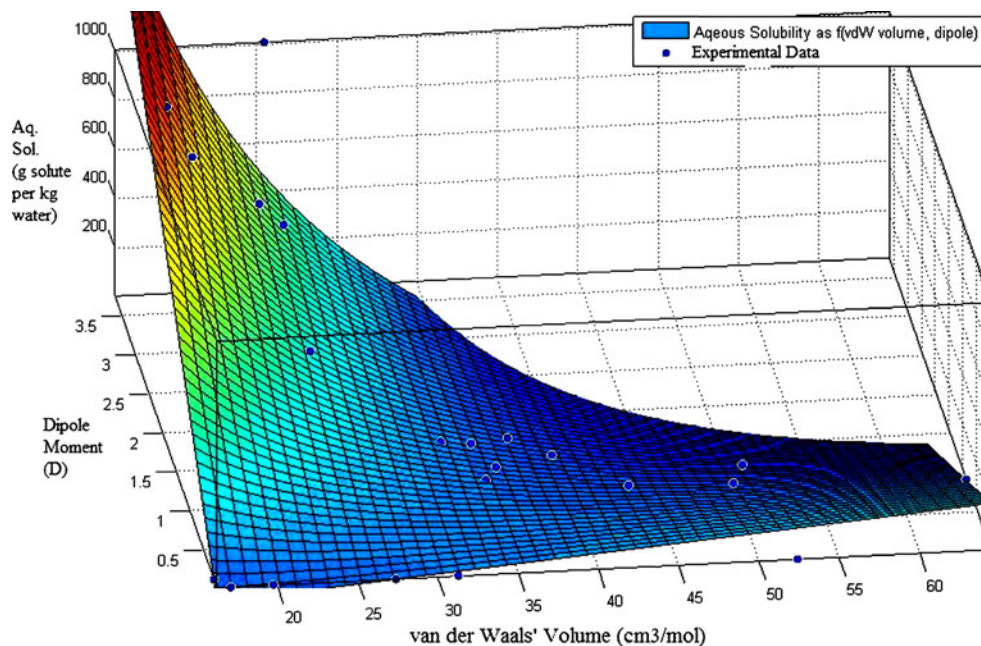
Figure 2 shows a two-dimensional contour surface for the aqueous solubility of single-carbon organic solutes as a function of van der Waals' volume and the dipole moment for single-carbon organic molecules. The red zone means maximum solubility and the dark blue means minimum.

As shown in Figs. 1 and 2, the aqueous solubility of single-carbon organic molecules exhibits maximum at the condition of large dipole moment and small molecular volume (left top corner of Fig. 2). As we move rightward or downward, the general trend of the solubility surface is that we go downhill (see Fig. 1). However, the distribution of experimental data points is such that they lie at or below the diagonal line as we move off the left top corner. No experimental data points were found in the right top quartile (region).

Starting at the left top corner, as the van der Waals' molar volume is low and the dipole moment is high, then there is a significant driving force for the solubilization or



**Fig. 1** A plot for the aqueous solubility, at 25°C, as a function of van der Waals' volume and the dipole moment for single-carbon organic molecules



hydration of an organic moiety as a result of its polar character. On the other hand, less entropic forces are involved which may otherwise interfere with the solubilization process.

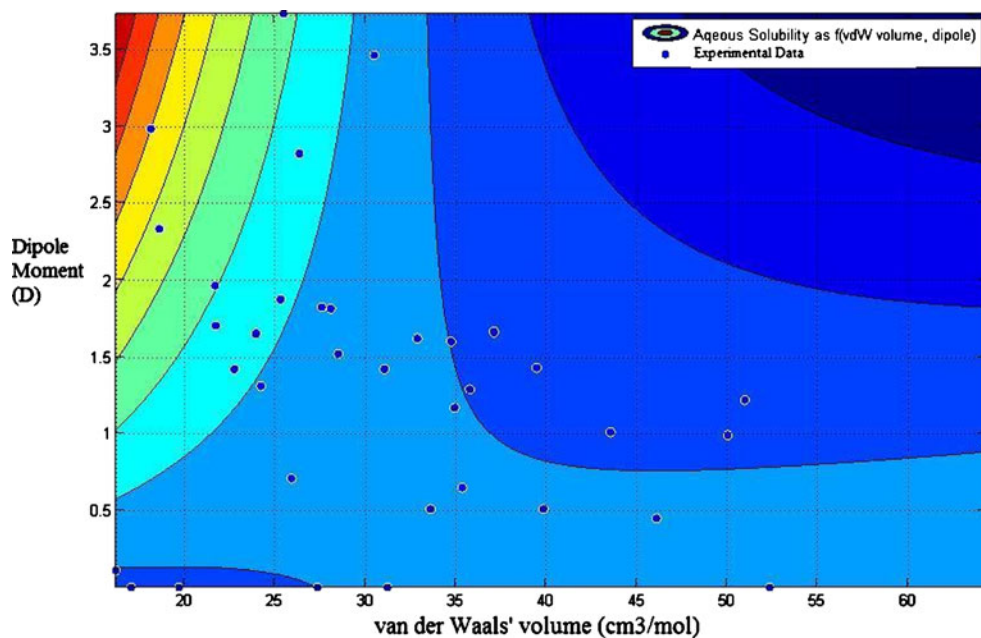
Moving diagonally off this corner in the direction of the abundance of experimental data points means that the organic solute starts losing its polar character (i.e., permanent dipole moment decreases), while at the same time, the van der Waals' volume of the organic solute starts increasing; hence, the entropic effects increase. Consequently, the driving force for solubilization gets weaker and the resistance to solubilization becomes stronger; these two

factors together will result in a decrease in the aqueous solubility of a single-carbon organic solute.

## Conclusions

1. In general, the solubility model, given by Eq. 1, which is based on the polar character and the molecular volume, adequately describes the aqueous solubility of single-carbon organic moieties.
2. The aqueous solubility of single-carbon organic solutes exhibits maximum at the condition of high polar

**Fig. 2** Two-dimensional (2-D) contour plot for the aqueous solubility, at 25°C, as a function of van der Waals' volume and the dipole moment for single-carbon organic molecules. The red zone means maximum solubility and the dark blue means minimum



character (large dipole moment) and low molecular volume.

3. The general trend of the solubility of single-carbon organic solutes, based on the proposed model (Eq. 1) could be explained in terms of the trade-off between the driving force (degree of polar character of the solute) for solubilization versus the resistance to be solubilized as a result of the entropic effects which increase with increasing molecular volume of the organic moiety.

## References

1. van Noort PCM (2009) Compound-class specific estimation of solid organic compound vapor pressure and aqueous solubility from simple molecular structure descriptors and the temperature of melting. *Chemosphere* 77:838–841. doi:10.1016/j.chemosphere.2009.07.069
2. Ruckenstein E, Shulgin IL (2006) Effect of salts and organic additives on the solubility of proteins in aqueous solutions. *Adv Colloid Interface Sci* 123–126:97–103. doi:10.1016/j.cis.2006.05.018
3. Tantishaiyakul V (2004) Prediction of aqueous solubility of organic salts of diclofenac using PLS and molecular modeling. *Int J Pharm* 275:133–139. doi:10.1016/j.ijpharm.2004.01.028
4. Ran Y, He Y, Yang G, Johnson JLH, Yalkowsky SH (2002) Estimation of aqueous solubility of organic compounds by using the general solubility equation. *Chemosphere* 48:487–509, PII: S0045-6535(02)00118-2
5. Jain N, Yalkowsky SH (2000) Estimation of the aqueous solubility I: Application to organic non-electrolytes. *J Pharm Sci* 90:234–252
6. Yalkowsky SH, He Y (2002) <http://www.pharmacy.arizona.edu/outreach/aquasol/index.html>
7. Al-Malah KI (2010) Aqueous solubility of a diatomic molecule as a function of its size & electronegativity difference. *J Mol Model*. doi:10.1007/s00894-010-0729-1
8. Fahlman BD (2007) *Materials Chemistry*. Springer, Dordrecht, Netherlands
9. Yaws CL (1999) *Chemical Properties Handbook*. McGraw-Hill, New York
10. Tolls J, van-Dijk J, Verbruggen EJM, Hermens JLM, Loeprecht B, Schüürmann G (2002) Aqueous solubility–molecular size relationships: A mechanistic case study using C<sub>10</sub>- to C<sub>19</sub>-alkanes. *J Phys Chem A* 106(11):2760–2765
11. Zhao YH, Abraham MH, Zissimos AM (2003) Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds. *J Org Chem* 68(19):7368–7373. doi:10.1021/jo034808o