The evolution and future of environmental partition coefficients

Journal:	<i>Environmental Reviews</i>
Manuscript ID	er-2015-0059.R1
Manuscript Type:	Review
Date Submitted by the Author:	23-Oct-2015
Complete List of Authors:	Mackay, Don; Trent University Celsie, Alena; Trent University, Chemistry Parnis, J; Trent University, Chemistry
Keyword:	Partition coefficients, QSARs, Octanol-water, Air-water, Octanol-air

1 **The Evolution and Future of Environmental Partition Coefficients** October 2015

2 Donald Mackay, Alena Celsie, and J. Mark Parnis

3 Chemical Partitioning Research Group, Department of Chemistry, Trent University, Peterborough

4 Ontario, Canada K9J 7B8

5

6 Corresponding author:

7 Donald Mackay, Trent University 1600 West Bank Drive, Peterborough, Ontario, K9J 7B8, Telephone:

8 705 740-2911, Fax: 705-748-1625, email: dmackay@trentu.ca

9 Word count: 8182 words

Draft

10 Graphical Abstract:

11
12 **The Evolution and Future of Environmental Partition Coefficients, by Don Mackay, Alena Celsie, and J.**
13 **Mark Parnis**

14 Abstract:

15 Partition or distribution coefficients (and increasingly referred to as partition ratios) are widely used in
16 environmental science to relate the concentration of a chemical solute in one phase to that in a second
17 phase between which equilibrium applies or is approached. The solutes include organic and inorganic
18 substances; the focus of this paper being on the former. The phases of interest include air, water, soils,
19 sediments, aerosols, and biotic phases such as lipids, blood and various tissues. Availability of reliable
20 partition coefficients for contaminants is essential for regulatory and scientific purposes, the general aim
21 being to understand and predict the distribution of the substances in multi-media environmental and
22 biological systems. The history of partition coefficients is reviewed, followed by a brief outline of their
23 theoretical basis and a discussion of methods for determining partition coefficients both empirically and
24 using a variety of predictive methods. It is suggested that ultimately a combination of empirical

25 measurements, quantitative structure-property relationships and computationally intensive, quantum
26 chemical molecular modeling techniques is required to provide accurate data for the large and
27 increasing number of chemicals of commerce that may enter the environment.

28

29 Key words: Partition coefficients, Partition ratios, QSARs, Octanol-water, Air-water, Octanol-air

Draft

30 Introduction:

31 Ratios of concentrations between two immiscible phases such as octanol and water or air and water are
32 traditionally called partition or distribution coefficients and are more recently referred to as partition
33 ratios. They play an important role in environmental science, especially when addressing the fate,
34 transport, exposure, and effects of relatively low concentrations of organic contaminants. They are also
35 widely used in medicinal and agro-chemistry, pharmacology, and in the design and simulation of
36 industrial chemical processes. Partition coefficients express the ratio of concentrations in two phases or
37 environmental media that are at equilibrium at a defined temperature and pressure, and usually at
38 dilute concentrations. They are thus thermodynamic functions independent of time and of chemical
39 reaction and transport rates. In this review we describe the evolution of partition coefficients from their
40 first conceptualization 200 years ago, with an emphasis on their applications, empirical measurement,
41 correlation, and prediction from fundamental chemical properties such as molecular structure.
42 Relationships to phase equilibrium criteria including chemical potentials, fugacity, and activity are
43 outlined. We conclude with a discussion of emerging methods of estimating these quantities using
44 computationally intensive molecular modeling techniques and recommendations for further
45 development of empirical and computational methods.

46

47 A Brief History:

48 The earliest partition coefficient can be regarded as the Henry's Law constant developed by William
49 Henry in 1803 to describe the relative concentrations of organic and inorganic gases between water and
50 air (Henry, 1803). Using modern units, this is expressed as:

$$51 \quad P_i = H \cdot C_w \quad (1)$$

Where P_i is partial pressure in air (Pa), C_w is concentration in water (mol/m^3) and H is the Henry's Law constant ($\text{Pa}\cdot\text{m}^3/\text{mol}$). At that time the Ideal Gas Law, defined in Equation 2, had not yet been fully formulated. The ideal gas law states that:

$$P_i V = nRT \quad (2)$$

Where P_i is the partial pressure (Pa), V is volume (m^3), n is number of moles, R is the ideal gas constant ($8.314\text{ Pa}\cdot\text{m}^3/\text{mol}\cdot\text{K}$), and T is the temperature (K). Equation 2 may be rearranged to express the corresponding gaseous concentration of species A (C_A) with units of mol/m^3 :

$$C_A = n/V = P_i/RT \quad (3)$$

Combining Equations 1 and 3, the dimensionless air-water partition coefficient (K_{AW}) may be expressed in terms of H as:

$$K_{AW} = C_A/C_w = H/RT \quad (4)$$

Later, in 1891 Nernst investigated the partitioning of benzoic acid between water and benzene, and proposed the Distribution Law that when a non-speciating chemical solute partitions between two immiscible liquids the ratio of concentrations at high dilution is constant (Nernst, 1891). He called this ratio the 'distribution coefficient'. The terms 'partition coefficient' and 'distribution coefficient' are now widely used. IUPAC (International Union of Pure & Applied Chemicals) has recently decreed that the preferred nomenclature should be partition ratios on the basis that the word 'coefficient' is inaccurate since a coefficient is a term for a factor in an equation. Here, we retain the old usage since much of the content is historical in nature.

The most frequently used partition coefficients from an environmental perspective are the air-water partition coefficient (K_{AW}), and the octanol-water partition coefficient (K_{OW}). They are routinely used to describe transport of chemicals between water and organic phases such as lipids and natural organic

carbon. Other partition coefficients that are also widely used include those for octanol-air (K_{OA}), soil-water, sediment-water, organic matter-water, lipid-water, and aerosol-air. In pharmacology analogous partition coefficients are employed between air, blood, and various tissues. Partition coefficients are also applied to whole organisms, calculated as fish-water partition coefficients.

The work of Henry (1801), Nernst (1891), and others established the concept that chemical solutes will partition predictably between liquid, gaseous, and solid phases. The common denominator or criterion for equilibrium can be regarded as the partial pressure that the substance exerts in each phase.

Following the work of Gibbs on chemical potentials (1876 & 1878), Lewis (1901) proposed the auxiliary function of fugacity as an equivalent criterion of equilibrium based on the mathematically rigorous derivations of functions based on the Second Law of Thermodynamics; namely entropy (ΔS , units of J/K·mol), free energy (ΔG , units of J/mol), and chemical potential (μ , units J/mol). Fugacity (f , units of Pa) was established as an accepted criterion of equilibrium between immiscible phases and became widely used in industrial chemical process design and simulation. For environmental systems at moderate temperatures and total pressures at low concentrations, partial pressure and fugacity become essentially equivalent and are readily interpreted as "escaping tendencies", where the chemical solute present in the phases at equilibrium exerts equal pressures in both phases. Lewis also introduced the concept of chemical activity as another criterion of equilibrium which is especially useful for addressing aqueous solutions of ionizing and non-ionizing solutes, as well as inorganic substances such as metals and ions.

The introduction of gaseous anaesthetics resulted in extensive studies of the mechanism by which chemicals exert anaesthesia and what is now termed reversible baseline narcosis. Overton (1901) and Meyer (1899) showed that narcosis tends to occur at a constant chemical concentration in the organism. The tendency of a chemical to partition from water into an organic phase, for example due to the

96 chemical's hydrophobicity, could be expressed using chemical-specific partition coefficients such as olive
97 oil -water.

98 Later, Hansch and Leo compiled an extensive database of octanol-water partition coefficients (K_{OW}) and
99 octanol became accepted as the more reproducible "standard" descriptor of hydrophobicity (Hansch *et*
100 *al.* 1995). This database enabled the effect of molecular structure on K_{OW} to be explored and correlated.
101 In 1939 Ferguson showed that narcosis tends to not only occur at a constant concentration in the
102 organism, but for a variety of chemicals and organisms at a common chemical activity in the gaseous or
103 liquid phase from which the organism respire (Ferguson 1939). Many environmental contaminants of
104 concern are hydrophobic and thus have a low, but measurable solubility in water. However, when the
105 solubility of a chemical is less than 1 mg/L or part-per-million, it can become challenging to measure the
106 partition coefficients experimentally because the solute may form colloids and it may adsorb to surfaces
107 and particles. The term 'hydrophobic' (water hating) and lipophilic (lipid loving) are often used
108 interchangeably. We suggest the term 'hydrophobic' is more descriptive because most organic
109 chemicals have solubilities in lipids or octanol of a few mol/L, but the solubilities in water vary from
110 mol/L to nmol/L. Thus differences in in water to lipid partitioning generally reflect differences in the
111 ability of water to accommodate the chemical, i.e. its hydrophobicity.

112 These pioneering concepts led to extensive efforts to measure octanol-water partition coefficients as
113 well as partition coefficients between other pairs of phases such as air-water, soil-water, and biota-
114 water. Successful correlations were developed between these partition coefficients, toxicity by narcosis,
115 bioconcentration, and pesticide fate in agricultural soils. These correlations also involved the pure
116 component properties of the solute, especially the melting and boiling point, vapour pressure and
117 solubility in water and other solvents. It was demonstrated that in soils and sediments, the primary
118 "solvent" phase for organic chemicals is organic matter, and by analogy the key solvent phase in biota is

often the lipid phase. As a result numerous correlations have been established between K_{ow} and partitioning from water into media such as soils, sediments, and lipid phases. It turns out that octanol is a particularly suitable surrogate solvent because it is readily available as a pure liquid and it has a similar ratio of C, H, and O atoms as are present in natural lipids (Mackay, 2001). Octanol is now used internationally in a regulatory context when assessing the fate and effects of chemicals. Predicting partition coefficients computationally was initially very expensive, however due to the general availability of low-cost and fast computing these methods have become more widely available both for development and use. This trend will likely continue implying that computational predictive methods will become increasingly feasible for obtaining partition coefficients.

There is now universal acceptance that chemical fate in the environment is primarily controlled by partition coefficients and irreversible reactivity (usually expressed as half-lives in a specific medium). Reliable partition coefficients and their relationships to fundamental pure solute properties, such as vapour pressure and solubility in water, are essential for evaluating chemical behaviour in the environment and in pharmacology. Many of these properties are required by regulatory programs world-wide.

Molecules can undergo chemical changes such as acid dissociation or complexation which pose additional challenges with respect to effectively describing their partitioning behaviour. For example, a neutral weak acid may show a partition coefficient that is 10,000 times that of the corresponding anion formed following acid dissociation. This indicates that the extent of partitioning of the molecule can vary over this range as a function of the extent of reversible chemical reactivity, in this case controlled by the pH. A common practice is to define a modified partition coefficient 'log D' instead of log K_{ow} in which D is the ratio of total ionized and non-ionized species in each phase. D is thus dependant on pH. Studies have been done on the effects of pH to partitioning or solubility behaviour including Kipton *et al.*

(1992), Pankow *et al.* (1997), Zhang *et al.* (2015), and others. Some estimation programs that predict log D include COSMO-RS (Klamt, 1995), ALOGPS (Tetko and Tanchuk, 2002), ACD/LogD (ACD/Labs, 2015), and Chemaxon (ChemAxon, 2015).

Temperature also has an effect on partition coefficients and this has been well documented (Mackay *et al.* 2006). Li *et al.* (2007) measured K_{AW} values at various temperatures and found these values increased with increasing temperature. Another study predicted H and K_{OA} values for a group of polycyclic aromatic hydrocarbons (PAHs) using the chemical property prediction software COSMOtherm (Parnis *et al.* 2015). Over a temperature range of -5°C to 40°C log K_{OA} values shifted by one log unit on average and log H shifted an average of 2 log units.

Measurement of Partition Coefficients:

In principle, all methods rely on direct or indirect measurement of concentration in the two phases of interest, such as water and octanol, preferably at low concentrations well below the solubility limits. In practice, it may be preferable to make the concentration measurement in only one phase and use mass balance techniques to deduce the concentration in the second phase. Experimentally, the two phases are brought into contact under static conditions (such as the shake-flask or slow stirring method) or under dynamic slow-flow conditions. The key requirement is that equilibrium is reached or at least closely approached. Partition coefficients vary enormously in magnitude, which necessitates the use of demanding and meticulous experimental techniques, often incurring high cost. The high cost of experimental determination has resulted in the development of a variety of predictive methods, outlined later. If the solute is subject to ionization, it is essential to define the aqueous phase pH. In

some cases partition coefficients can be estimated from pure solute properties such as vapour pressure and solubilities in phases such as water and octanol, however miscibility may prevent such estimations, for example ethanol in water.

A review of techniques for measuring or predicting partition coefficients are given in the Handbooks by Mackay *et al.* (2006) and in texts such as Boethling & Mackay (2000) and Lyman and Reehl, (1982). Many regulatory programs require the use of internationally accepted, standard experimental methods. International agencies have defined appropriate examples such as OECD test 123 (2006) for measuring K_{OW} values in the laboratory. As is discussed later, independent measurements of K_{AW} , K_{OW} and K_{OA} may not result in consistent data. A novel three-partition coefficient approach has been developed by Xu and Kropscott (2012) in which dilute solutions of cyclic siloxanes were brought to equilibrium between air, water and octanol. Using this technique K_{AW} , K_{OA} , and K_{OW} can be measured simultaneously and under the same conditions while ensuring consistency.

A striking feature of environmental partition coefficients is their high chemical-to-chemical variability. As is shown later, K_{AW} may vary over 7 orders of magnitude.. Similarly octanol-water coefficients can vary over 12 orders of magnitude. These variations translate into vast differences in environmental partitioning and fate as is illustrated later in the chemical space diagrams in Figures 2 & 3. A further implication is that for ionogenic substances such as the chlorophenols, the coefficients are sensitive to pH. For example for pentachlorophenol $\log K_{OW}$ varies from 3.0 at a pH of 7 to 4.8 at a pH of 4. Similarly chemical degradation products can have properties orders of magnitude different from their parent compound. An example illustrated in Figure 2 is the pesticide chlorpyrifos and its oxidised product the oxon in which a sulfur atom is replaced by oxygen.

Physico-Chemical Basis of Partition Coefficients

187 *Equilibrium Criteria*

188 If we consider three immiscible phases, such as air, water and octanol, all of which contain a dissolved
189 solute such as benzene that is at equilibrium and at low concentrations, the fundamental criterion of
190 phase equilibrium is that the overall ΔG of the system is minimized. At this minimum the incremental
191 changes in G with respect to mole number (dG/dn) of the solute are equal for each solvent pair, where n
192 is the moles of solute. These derivatives are of the Gibbs energy with respect to the number of moles of
193 the solute, and are both partial molar Gibbs free energies and the auxiliary functions of chemical
194 potential (μ) which is the fundamental criterion of establishing equilibrium of the solute between
195 phases. Calculating absolute values of chemical potential requires an absolute benchmark or standard
196 state to be defined. Furthermore, solute concentration is logarithmically related to chemical potential,
197 which proves to be inconvenient when calculating the magnitude of departures from equilibrium
198 between phases. There is also a conceptual difficulty arising from its unusual units of J/mol.

199 In 1901 G.N. Lewis addressed these problems by introducing another auxiliary function, fugacity (f , Pa):

200
$$d\mu = RT d \ln f \quad (5)$$

201 Definition also requires the limiting condition that $f/P \rightarrow 1$ as $P \rightarrow 0$, where P is pressure (Pa).

202 As previously stated, for most environmental purposes, at low temperature and pressure, fugacity can
203 be regarded as being equivalent to partial pressure. Lewis thus provided, in the form of fugacity, a
204 thermodynamically rigorous treatment of Henry's partial pressure as an equilibrium criterion.

205 It can be shown that a partition coefficient K expressed as a ratio of mole fractions is related to the ΔG
206 of the solute between the phases as follows with R and T :

207
$$\Delta G = RT \ln K_x \quad (6)$$

208 ΔG expresses the differences in solvation properties of the solute between the solvents and can be
209 related to molecular properties such as molar volume and interactions such as hydrogen bonding. The
210 result is that when correlating partition coefficients against solute and solvent properties it is preferable
211 to regress $\log K_x$ (ratio of mole fractions) or $\log K_c$ (ratio of concentrations such as mol/m^3) against these
212 properties. Simple examples of such structure-property relationships are illustrated in Figure 1.

213
214 Figure 1: Simple structure-property correlations between chemical properties of chlorobenzenes from
215 Mackay *et al.* (2006) A) $\log K_{ow}$ (octanol-water partition coefficient) vs. Le Bas molar volume at the (top
216 left) B) $\log S_L$ (subcooled solubility) vs molar volume (top right) C) $\log K_{aw}$ (air-water partition
217 coefficient) vs molar volume (bottom). Reproduced from Mackay *et al.* (2006) with permission from
218 Taylor and Francis, 2015.

219 In summary there are several criteria of inter-phase equilibrium, namely the original criterion of partial
 220 pressure introduced by Henry (1801), the rigorous criterion of chemical potential introduced by Gibbs
 221 (1876, 1878), the criterion of fugacity and activity introduced by Lewis (1901). These criteria, especially
 222 fugacity, provide the basis for defining all partition coefficients.

223 *Fugacity-concentration relationships.*

224 The relationships between fugacity (f_i) and mole fraction concentration are as follows:

225 For gases:
$$f_i = y_i \phi_i P_T \quad (7)$$

226 For liquids:
$$f_i = x_i \gamma_i f_i^o = x_i \gamma_i P_{Li} \quad (8)$$

227 In these equations y_i and x_i are mole fractions of solute in gases and liquids respectively, γ is an activity
 228 coefficient in the liquid phase whereas ϕ_i is an analogous fugacity coefficient in the vapour phase, f_i^o is
 229 the fugacity or vapour pressure of the pure liquid state chemical when x_i and γ_i are both equal to 1.0, P_T
 230 is the total pressure and P_{Li} is the liquid phase vapour pressure. At equilibrium the fugacities are equal
 231 and Equations 7 and 8 can be combined to give a partition coefficient K as a ratio of mole fractions,
 232 namely:

233
$$K_{yx} = y_i/x_i = \gamma_i P_{Li}/(\phi_i P_T) \quad (9)$$

234 For an ideal gas, as usually applies at low pressures, the fugacity coefficient ϕ_i can be assumed to be 1.0.
 235 If liquid phase ideality is also assumed then γ_i and ϕ_i are both 1.0, the equation reduces to Raoult's law,
 236 which states:

$$P_i = P_i^o X_i \quad (10)$$

237 Where P_i^o is the vapour pressure of the pure solvent and X_i is the mole fraction of solvent in solution
 238 (Ebbing and Gammon, 2005).

239 In environmental systems, however, γ_i is rarely 1.0 and values can reach into the millions for
 240 hydrophobic chemicals, such as polychlorinated biphenyls (PCBs) or dichloro-diphenyl-trichloroethane
 241 (DDT) in aqueous solutions.

242 Lewis also introduced a related auxiliary function activity (a)

$$243 \quad a = x_i \gamma_i = \phi_i y_i = f_i / f_i^\circ = f_i / P_L \quad (11)$$

244 Here, f_i / f_i° and f_i / P_L are the fractions of saturation.

245 The fugacity equations provide the thermodynamic basis of partition coefficients that are usually
 246 expressed in volumetric concentrations (C) such as mol/m³ instead of mole fractions x and y . For liquids,
 247 using water as an example:

$$248 \quad C_{iW} = x_i / v_W = f_i / (\gamma_i P_{Wi} v_W) \quad (12)$$

249 Where v_W is the molar volume (mol/m³) of the water solution ($\cong 18 \times 10^{-6}$ m³/mol) and x is equal to the
 250 molar mass divided by the phase density (eg $x = 18.02$ g/mol $\div 10^6$ g/m³ for water).

251 For gases obeying the ideal gas law such as air, the air concentration C_{iA} may be expressed in terms of
 252 the chemical's partial pressure P_i , which is essentially equal to the its fugacity under environmental
 253 conditions:

$$254 \quad C_{iA} = (y_i P_T) / RT = P_i / RT = f_i / RT \quad (13)$$

255 Note that in the case of an ideal gas, C_i is both n_i / V and P_i / RT . At equilibrium, f_i is equal in both phases
 256 and by combining Equations 12 and 13 and cancelling the fugacities out K_{AW} can be deduced as in
 257 Equation 14:

$$258 \quad C_{iA} / C_{iW} = (\gamma_i P_{Li} v_L) / RT = K_{AW} = H / RT \quad (14)$$

This provides a definition of H from Equation 4 in terms of its fundamental physical-chemical determinants.

$$H = \gamma_i P_{Wi} v_W \quad (15)$$

A high value of H and thus of K_{AW} is therefore caused by a high activity coefficient in water (which corresponds to high hydrophobicity) and a high vapour pressure. Substances such as PCBs can thus have relatively high values of K_{AW} of the order of 0.01 because the very low vapour pressure is compensated by the high activity coefficient in water.

For liquid-liquid systems such as octanol-water, the fugacity Equation 8 for liquids is applied twice to account for both liquids. Recalling that at saturation the fugacity equals the reference fugacity or the liquid vapor pressure, thus $x \cdot \gamma$ is 1.0 and C is mole fraction /molar volume:

$$C_2/C_1 = (x_2/x_1)(v_1/v_2) = (\gamma_1 v_1)/(\gamma_2 v_2) = K_{21} \quad (16)$$

A high partition coefficient K_{21} is thus caused primarily by the high activity coefficient in the second phase. For example for octanol-water it is the value of γ_2 in water that is the primary determinant of a high K_{OW} and thus of hydrophobicity.

Z values

Most environmental phases such as soils, sediments, and biota have indeterminate molar masses and molar volumes. In such cases the fundamental equations just discussed cannot be applied directly, but a simple expedient is to define the concentration as a linear function of fugacity:

$$C = Zf \quad (17)$$

Here Z is a 'fugacity capacity' with units of $\text{mol}/\text{m}^3 \cdot \text{Pa}$. For a gaseous phase Z_A is $1/RT$, and for a pure liquid such as water, Z_W is $1/H$ or $1/(\gamma_i P_{Wi}^s v_W)$. Z values reflect the solvent capacity or affinity of the

chemical towards the liquid solvent; a high Z value reflects a high affinity and solubility. Z is essentially the reciprocal of the Henry's Law constant for the substance in the liquid. For a two-solvent system at equilibrium, the fugacities are equal and the partition coefficient is given by:

$$C_1/C_2 = Z_1 f_1 / Z_2 f_2 = Z_1 / Z_2 = K_{12} \quad (18)$$

A Z value is essentially 'half' of a partition coefficient. Measured partition coefficients such as those for soil-water or biota-water can then be used to deduce Z values in these environmental phases when both γ and v are uncertain.

The fugacity concept is particularly convenient for establishing multimedia mass balance models as discussed by Mackay (2001). The calculated fugacities can be converted to activities by dividing by the saturation liquid-state vapour pressure. Concentrations (C) can be calculated as Zf (units of mol m⁻³), and to amounts in terms of moles calculated as VZf where V is the compartment volume (m³).

Solid solutes and the fugacity ratio.

It is noteworthy that it is the vapour pressure of the liquid state chemical that is used to calculate the fugacity. This is termed the reference fugacity ($P_L = f^\circ$). For solid chemicals such as naphthalene the liquid vapour pressure and solubility cannot be measured experimentally, it is then necessary to estimate the subcooled liquid properties. The liquid state vapour pressures and solubilities are greater than the solid values, the reason being that at temperatures at and above the melting point the solute precipitates as a crystalline solid as a result of the lower energy state of the solid. The ratio of the solid and liquid vapour pressures (P_S/P_L) is termed the 'fugacity ratio' (F) and is estimated from the entropy of fusion, the ambient temperature T (K) and the melting point T_M (K) expressed by the following relationship (Mackay, 2001) :

$$F = \exp \left[-6.79 \left(\frac{T_M}{T} - 1 \right) \right] \quad (19)$$

302 An important implication is that solid solutes in solution cannot achieve a fugacity exceeding either $F \cdot P_L$
303 or P_s , nor can the activity (a) exceed F .

304 In summary, environmental partition coefficients can be measured directly and correlated with other
305 chemical properties, they can be used to calculate phase-specific Z values and in some cases they can be
306 deduced from fundamental physico-chemical properties, such as solubilities in water and other solvents,
307 and vapour pressures. As is discussed later, considerable effort has been devoted to correlating these
308 partition coefficients to molecular descriptors (such as molar mass or volume) and to the presence of
309 specific atoms, functional groups, or solvent parameters, as detailed below.

310

311 *The octanol-water-air triangle and chemical space diagrams*

312 Consideration of the numerous and complex environmental phases or media and the large and
313 increasing number of chemicals of concern provides an incentive for displaying these partitioning
314 properties in a simple, reproducible and transparent format. The environment can be treated as being
315 approximately simulated by three phases, namely air, water and organic matter. Mineral phases
316 generally play a minor role except under very dry conditions. Organic matter can be viewed as being
317 somewhat similar in solvent properties to *n*-octanol, resulting in convenient correlations between, for
318 example, K_{OW} and soil organic carbon-water partition coefficients (K_{OC}). This simplification has the merit
319 that the three partition coefficients K_{AW} , K_{OW} and K_{OA} can be measured reproducibly in the laboratory at
320 a specified temperature and the values used internationally when assessing chemical fate in the
321 environment and the resulting wildlife and human exposures and risks of adverse effects.

322 A useful method of displaying these partition coefficients is to plot them on a “chemical space” diagram
323 as in Figure 2, usually employing log-log scales. Two partition coefficients can be used and the third

corresponds to a diagonal, in which case diagonal lines will be marked on the graph. In Figure 2 the x axis is $\log K_{OW}$ and the y axis is $\log K_{AW}$; thus the diagonals are $\log K_{OA}$. Only two partition coefficients (for example K_{OW} and K_{AW}) need to be known in order to plot values and once plotted the third partition coefficient (in this case K_{OA}) value can be observed. If the third partition coefficient is known these values can be compared for consistency and possible errors in one or more of the reported values. Chemicals that partition predominantly into the atmosphere tend to lie to the upper left whereas those that concentrate in water or soils tend to lie to the lower left and to the lower right respectively. If volumes of the phases are defined then the percentages of chemical masses in each phase can be estimated using a Level II fugacity model (Mackay, 2001). Of particular interest are the multimedia chemicals that lie near the middle of the diagram and have the ability to partition into all media and may undergo inter-media transfers. A classic case is DDT that is regarded as an involatile solid but can experience water to air transfer and tends to accumulate to high concentrations in organic phases such as sediments, soils and biota such as fish and birds. The resulting adverse effects on birds and other wildlife were first recognised by Rachel Carson in her seminal book "Silent Spring" (1962).

Figure 2: Chemical space plot of $\log K_{AW}$ vs $\log K_{OW}$ with $\log K_{OA}$ as diagonals, illustrating the wide variation in partition coefficients and showing the locations of selected chemicals. The bold lines represent estimated fractions of the chemical mass partitioning into air, water and octanol for assumed volumes of these solvents, typical of environmental proportions. Values come from Mackay *et al.* (2006), EPISuite, and Nowosielski (1997). Taken from Mackay (2001) and modified with permission from Taylor and Francis, 2015.

Wania (2003) has plotted similar diagrams of $\log K_{AW}$ vs $\log K_{OA}$ with $\log K_{OW}$ on the diagonal as shown in Figure 3. This is especially useful for illustrating how chemicals are likely to behave in the context of global transport in air and oceans. The “flyers” are located to the upper left, “swimmers” to the lower left and sorbers to the right. A homologous series of chemicals such as the PCBs can form a group that extends across this diagram between regions. Finally triangular diagrams used for describing ternary phase equilibrium systems can be used, but they are most illuminating when the scales are linear rather than logarithmic and they are not readily applied when partition coefficients vary over orders of magnitude (Wania, 2003).

Figure 3: Chemical space diagram from Wania (2003) containing partition coefficient predictions for certain classes of chemicals including chlorobenzenes (CBZs), PCBs, polychlorinated dibenzo-p-dioxins (PCDDs), and polychlorinated dibenzofurans (PCDFs). Each letter represents a chemical with specific physico-chemical properties: A is a particle-sorbed chemical, B is a water-soluble chemical, C is a volatile

water soluble chemical, D is a multimedia chemical, and E is a very hydrophobic semivolatile chemical. Reprinted with permission from Wania, F. 2003. Assessing the potential of persistent organic chemicals for long-range transport and accumulation in polar regions. *Environmental Science & Technology*, 37, 1344-1351. Copyright 2015 American Chemical Society.

Three solubility approach and the air-water-octanol triangle

These partition coefficients are constrained by a thermodynamic cycle as shown in Figure 4, thus, for example, K_{OA} should equal K_{OW}/K_{AW} and only two of three can be defined independently. Ideally these partition coefficients should be measured at dilute conditions well below the saturation conditions corresponding to the solubilities. It is possible, and indeed likely, that the activity coefficients are concentration dependent thus the infinite dilution activity coefficient of, for example, benzene in water may not equal its value at the higher concentration corresponding to saturation. If the solubility is low this may not be an issue and the activity coefficient can be estimated from the solubility. From Equation 14 if the fugacity f_i is equated to the reference fugacity P_i as applies at saturation then the solubility is given by C_s and equals $1/(\gamma_i v_L)$.

Mackay and Cole (2000) suggested a 'three solubility' approach in which the three 'solubilities' in air, water and octanol are defined and the three partition coefficients are then defined as ratios of these solubilities as illustrated in Figure 4. These solubilities can be real and measureable or hypothetical and unmeasurable when the solute is miscible with the liquid phase. Benzene has a real solubility in water but ethanol has only a 'pseudo-solubility' in water because of miscibility. More fundamentally these solubilities can be expressed in terms of activity coefficients and molar volumes as discussed earlier. The reason that three solubilities are defined but only two partition coefficients need be defined is that the solubilities contain additional information about saturation conditions. This thermodynamic triangular

cycle can also be applied to thermophysical properties such as enthalpies of phase transfer that are used to estimate the temperature dependence of partition properties over a range of environmental temperatures.

Figure 4: An illustration showing variables that are related to one another, including partition coefficients (K), solubilities (C^s), liquid vapour pressure (P^s_L), fugacity (f) and activity (γ), for a liquid chemical, at a given temperature.

The solubilities (mol/m^3) are respectively P/RT in air, $1/(\gamma_W v_W)$ in water and $1/(\gamma_O v_O)$ in octanol. Thus

$$K_{OW} = C_O/C_W = \gamma_W v_W / \gamma_O v_O \quad (20)$$

$$K_{AW} = P \gamma_W v_W / RT \quad (21)$$

$$K_{OA} = RT / P \gamma_O v_O \quad (22)$$

397 *Consistency tests*

398 Partition coefficients can be measured directly from ratios of low concentrations at equilibrium at sub-
399 saturated conditions or they can be estimated from solubilities. It is useful to test the consistency of
400 experimentally determined or estimated partition coefficients and solubilities as depicted in Figure 4. If,
401 as is usually the case, empirical measurements are not exactly consistent, the properties can be adjusted
402 taking into account perceived accuracy to select a set of internally consistent values. For example, K_{AW}
403 may not be equal to K_{OA}/K_{OW} or to H/RT when H is estimated from vapour pressure and solubility in
404 water. This estimation is suspect if the solubility is high (i.e. >1 g/L), which occurs with more polar
405 organic chemicals. Many environmental chemicals are hydrophobic and sparingly soluble in water, in
406 which case the above assumption is acceptable. In some cases, the solubility of a chemical in octanol can
407 be measured and K_{OW} calculated by dividing this solubility by the solubility in water. K_{OA} can then be
408 estimated as K_{OW}/K_{AW} . Often the solubility in octanol cannot be determined because of miscibility, in
409 which case only a "pseudo-solubility" can be estimated. Compounds that especially benefit from such
410 consistency tests are those that have unusual, extreme, or difficult-to-measure properties, such as
411 extremely low solubilities in a particular medium, or vanishingly small vapour pressures (Eg. salts).
412 Highly volatile substances such as methylsiloxanes and involatile substances such as multi-ring polycyclic
413 aromatics such as anthracene are also examples of compounds with extreme properties.

414 Techniques have been suggested for these consistency tests and adjustment techniques have been
415 developed by Hung *et al.* (2013), Wania (2003), Beyer *et al.* (2002), Li *et al.* (2003), Schenker *et al.*
416 (2005), Shen and Wania, (2005), and Xiao *et al.* (2004); a least squares approach has been developed by
417 Aberg *et al.* (2008) and Schenker *et al.* (2005). Li *et al.* (2003) have used this approach to select
418 "favoured adjusted values". Inconsistencies may arise from analytical error, mistakes in reporting, multi-
419 level referencing, and failure to properly reference the original source (Shen and Wania 2005). Li *et al.*

(2003) derived internally consistent properties for PCBs and selected organochlorine pesticides.

Solubility, vapour pressure, Henry's Law constant, and K_{OW} were among the selected parameters for the internal consistency checks.

The current trend in regulatory practice for environmental chemicals is to require acquisition of accurate property data over the relevant range of environmental temperatures using established techniques and subjecting the data to consistency tests. Regression plots for homologous series of organic solutes as illustrated in Figure 1 are useful for checking the consistency of reported empirical results and for interpolation, but they should not be used for extrapolations.

Partial miscibility of octanol and water

Octanol and water are somewhat miscible, so a problem may arise when determining K_{OW} because the octanol in the octanol-water medium becomes saturated with water, whereas the octanol in K_{OA} may remain pure. Octanol has a very low solubility in water (4.5×10^{-3} M), so the molar volume of water is usually assumed to be that of pure water. Water has a solubility of 2.3 M in octanol, and due to this relatively high solubility it is more accurate to use water-saturated octanol (or 'wet' octanol) as the second phase in the solvent (Sangster 1989).

The result may be a deviation of the K_{OW} value from its expected value, and the extent of this deviation depends on the extent of miscibility of octanol and water. Chiou *et al.* (1982) calculated this deviation for both K_{OW} and solubility, and it was found that this deviation became systematically larger with decreasing water solubility of the solute. The effect of dissolved octanol on water solubility is most pronounced for the solutes which are least soluble in water and vice versa; the presence of dissolved water in octanol decreases the solubility of octanol towards other solutes (Chiou *et al.* 1982).

An alternative method is to use an adjustment procedure can be done as described by Beyer *et al.* (2002) and Li *et al.* (2003). The adjustment procedure is a simple regression that relates the experimentally determined K_{OW} to the ratio of the solubilities in both pure octanol and pure water. Alternatively some estimation programs, such as COSMOtherm (COSMOthermX, 2013), have the option of calculating K_{OW} for either wet or dry octanol.

Figure 5 shows related variables including ΔG , mole fractions (x), molar volumes (v), molar concentrations (C) and thermodynamic partition coefficients (K). These variables can be applied to calculate Gibbs free energies between any of the three environmental media (air, water or octanol) and a dielectric medium. ΔG is a thermodynamics term which refers to the amount of free energy involved in the phase transition. Partition coefficients and concentrations can be calculated as shown, and are applied to environments such as the ones indicated on the outside of Figure 5.

Figure 5: A representation showing related variables relevant to environmental partitioning. The central quantities in black font are the fundamental thermodynamic parameters, those in red are the measured physico-chemical properties and those in green are the related quantities of environmental interest.

A note on the difference between K' and K : from a fundamental physical-chemical or thermodynamic viewpoint the key partition coefficient is the ratio of mole fractions since it is directly related to ΔG . It is often more convenient to express the coefficient as a ratio of concentrations such as mol/L or mol/m³ and most reported values are in these mol/volume units. The conversion is readily done using the molar volumes of the two phases as shown in Figure 5. This usually assumes that the molar volume of the pure solvent phase applies to the dilute solutions. For many environmental purposes the molar volume is uncertain or cannot be determined, examples being soils or fish in which case a mole fraction coefficient is indeterminate.

Correlations Using Atom, Fragment, and Group Contribution Methods

Numerous methods have been developed in which partition coefficients such as K_{ow} are regressed (usually as $\log K_{ow}$) against the number and type of atoms or 'fragment groups' that are present in the solute molecule. These methods rely on empirical data. Table 1 lists several predictive methods and associated error as reported by Mannhold *et al.* (2009).

Table 1: Various methods commonly used to predict K_{ow} reported by Mannhold *et al.* (2009). Methods chosen were based on highest RMSE values while representing all method types. RMSE is root-mean-square error, n is number of test chemicals.

Method	Approach	RMSE	% of	% of	% of	Reference
--------	----------	------	------	------	------	-----------

			Molecules with Error Range <0.5	Molecules within Error Range of 0.5-1	Molecules with Error Range >1	
AB/LogP	Fragment	0.41	84	12	4	Japertas <i>et al.</i> (2002)
S+LogP	Neural network	0.45	76	22	3	(ADMET Predictor, 2015)
ACD/logP	Fragment	0.50	75	17	7	Petrauskas and Kolovanov (2000)
CLOGP	Fragment	0.52	74	20	6	Hansch <i>et al.</i> (1995)
XLOGP3	Atom	0.62	60	30	10	Cheng <i>et al.</i> (2007)
KowWIN	Fragment	0.64	68	21	11	Meylan and Howard (2000) EPISuite (2015)
ABSOLV	Whole- molecule	0.75	53	30	17	(Abraham, 1993)
COSMOFrag	Fragment	0.84	48	26	19	Hornig and Klamt (2005)

UNIQUAC & UNIFAC

Among the group-contribution methods that have been developed over the past decades, the family of UNIFAC methods are of special note. Initially, the predictive software UNIQUAC (UNiversal QUAsi-

Chemical) was designed for prediction of non-ideal liquid-liquid or liquid-vapour phase mixtures (Abrams and Prausnitz, 1975). This method uses activity coefficients calculated from excess enthalpy, which is modeled as a function of the chemical composition. This model uses only two adjustable parameters per molecular pair and can be used to predict partitioning behaviour, activity coefficients, and excess enthalpy.

This model was later modified by adding a functional group component to the calculation, the new method named UNIFAC (UNIQUAC Functional-group Activity Coefficients) (Fredenslund *et al.* 1975). UNIFAC is a group-contribution method in which the independent variables are the concentrations of the functional groups instead of the molecules themselves. The UNIFAC model requires two molecular size-based parameters derived from experimental data per pair of functional groups. The model makes predictions for non-ideal solutions or mixtures and accounts for temperature with additional parameters. The most recent versions of UNIQUAC and UNIFAC software can be obtained online through the Dortmund Data Bank Software & Separation Technology (DDBST) website (DDBST GmbH, 2015). These methods are most powerful for determination of solution-phase mixture properties with molecules for which reliable and detailed experimental data are available.

Whole-Molecule Approach

Whole-molecule calculation approaches are based on the properties of an entire molecule, such as charge densities, molecular surface area, volume and free energy (Mannhold *et al.* 2009). These methods are advantageous in that missing/inaccurate fragment values are avoided. Computational methods are used to employ this approach due to the complexity of the calculations. Whole-molecule approaches can be divided into two classes: empirical methods or methods that employ 3D-structure representation, and methods that use topological descriptors (Mannhold *et al.* 2009).

503 Methods which involve empirical approaches or 3D structure representations are based on free energy
504 calculations. The base 10 log of the partition coefficient such as log K is proportional to the energy
505 transfer that occurs between octanol and water:

$$506 \quad -2.303RT\log K_{OW} = \Delta G_{oct} - \Delta G_w = \Delta G_{ow} \quad (23)$$

$$507 \quad = \Delta G_{ow}^{el} + \Delta G_{ow}^{nel} \quad (24)$$

508 In the equation above G_w and G_{oct} are the solvation-free energies of the solute in water and octanol,
509 respectively. G_{ow}^{el} and G_{ow}^{nel} correspond to electrostatic and non-polar terms of solute-solvent
510 interactions.

511 Experimental methods are based on the assumption that 1) the solute creates a cavity in the solvent,
512 which requires energy, and 2) that within the cavity there will be various solute-solvent interactions.
513 Examples of whole-molecule calculation methods which are based on empirical data include ABSOLV
514 (Platts *et al.* 1999; available from ACD/ADME, 2009) and SPARC (Hilal *et al.* 1994; available from
515 ARChem, 2015)

516 Semi-empirical calculation methods are used to increase the speed of calculation because these
517 calculations are not as in depth and rely more on known data and assumptions. QLOGP is such a method
518 that uses semi-empirical calculations. QLOGP is a molecular sized-based model which uses molar
519 volume, interaction potentials, and minimal structure-based parameters to predict the chemical
520 potential of the chemical (Buchwald and Bodor, 1998; available from Buchwald, 2015). The chemical
521 potential is then used to predict certain physico-chemical properties such as partition coefficients.
522 Another approach is to use a continuum solvation model, which measures the electrostatic solvation
523 transfer energy using the Generalized Born (GB) approximation, essentially approximating the dielectric
524 response of the solvation cavity with a data-parametrized expression for the interaction energy. This is

the basis of the program GBLOGP that predicts K_{ow} . Other methods are based on molecular dynamics calculations, *ab initio* calculations, connectivity indices, molecular lipophilicity potentials, and lattice energy calculations.

Methods based on topological descriptors have the advantage of being hundreds to thousands of times faster than intensive *ab initio* or molecular dynamics calculations. Such methods include Dragon 6 which used to be called MLOGP (Moriguchi *et al.* 1992; available from Talete SRL, 2015), TLOGP (Junghans and Pretsch, 1997; available from Upstream Solutions, 2015), methods which use energy-state indexes such as ALOGPS (available from VCC Labs, 2015) which uses the neural networking method, and methods based on diverse descriptors, such as S+LogP (ADMET Predictor, 2015) which uses uncorrelated 2D descriptors.

Whole Molecule Approach - Linear Free Energy Relationships

A linear free energy relationship (LFER) is a linear correlation between the logarithm of an equilibrium constant which is proportional to the standard free energy change. These relationships arise because at a constant pressure and temperature the logarithm of an equilibrium constant and of a rate constant are both linear functions of the free energy of activation. LFERs have been applied to molecular properties, however it is assumed that contributions of structural components or property descriptors of a molecule to the overall properties of that molecule are additive and independent. Abraham (1993) has developed a particularly effective and popular estimation method based LFERs for the calculation of partition coefficients. It is generally available and is conveniently implemented in the ABSOLV program.

ABSOLV: Abraham's Linear Free Energy Method (available from ACD/Labs, 2015)

546 Using a Linear Solvation Energy Relationship (LSER) approach, this model was designed by Abraham
547 using solute descriptors and are termed Abraham solute descriptors (Abraham, 1993). ABSOLV is a
548 computational program which uses solvation parameters to calculate various solvation-associated
549 properties using Abraham type equations. There are 6 solute descriptors in the program with values
550 available for solutes ranging from biochemical interactions to environmental pollutants. These
551 descriptors are combined into the following linear free energy relationships (Abraham *et al.* 20010):

$$552 \quad \log P_w = c + eE + sS + aA + bB + vV \quad (25)$$

$$\log P_g = c + eE + sS + aA + bB + lL \quad (26)$$

553 These two relationships are used by the ABSOLV program for partitioning between water and solvents
554 (Equation 25) and partitioning between the gas-phase and solvents (Equation 26).

555 The descriptors are:

- 556 • E = excess molar refraction, which is calculated using the refractive index of the liquid at 20 °C,
557 can be calculated as addition of fragment pieces (i.e., substructures).
- 558 • S = dipolarity / polarizability; obtained from water-solvent partition coefficients or
559 experimentally determined by gas liquid chromatographic measurements on stationary polar
560 phases. Can be obtained also by fragment addition or by analogy to another compound.
- 561 • A = H-Bonding acidity parameter; overall effective hydrogen bond acidity which can be obtained
562 by fragment addition or by analogy to another compound, or by measurement of
563 physicochemical properties such as log P for a number of water-solvent systems.
- 564 • B= H-Bonding basicity parameter; overall effective hydrogen bond basicity, which can be
565 obtained also by fragment addition or by analogy to another compound, or by measurement of
566 physicochemical properties such as log P for a number of water-solvent systems.

- V = McGowan volume, which requires only the molecular formula, number of bonds, and number of rings in a molecule. It is calculated using the algorithm of Abraham, and has units of $(\text{cm}^3 \cdot \text{mol}^{-1})/100$ (Abraham and McGowan, 1987).
- L is the solute gas to hexadecane partition coefficient at 25°C

In general, the solute descriptors represent the influence of the solute over various solute-solvent interactions, and the coefficients (represented by small letters) correspond to the complementary effect of the phases on these interactions. The coefficients are calculated by ABSOLV using linear regression prior to calculation of K_{OW} . The ABSOLV program currently contains over 15 000 compounds with experimental log P values (Mannhold *et al.* 2009).

For example, the Abraham solute descriptors for octanol-water systems are shown below (Abraham, 1993):

$$K_{OW} = 0.088 + 0.562E - 1.054S + 3.814V + 0.034A - 3.460B \quad (27)$$

$$n = 613 \quad r = 0.9974, \quad SD = 0.116,$$

Poly-parameter Linear Free Energy Relationship-Composition-Based-Model

Goss *et al.* (2013) have proposed a biotic partitioning model which is based on poly-parameter linear free energy relationships (PP-LFER), but is unique in that it also uses tissue composition parameters for individual organisms. It is based on the following relationship:

590 $K_{\text{tissue/water}} = f_{\text{storage lipid}}K_{\text{storage lipid/water}} + f_{\text{membrane}}K_{\text{membrane/water}} + f_{\text{albumin}}K_{\text{albumin/water}} +$
591 $f_{\text{protein}}K_{\text{protein/water}} + f_{\text{water}}$ (28)

592
593 Where f denotes volume fraction and K denote partition coefficients. This method was compared to
594 both experimental partition coefficients in various tissue and to literature values, with satisfactory
595 results. The root mean square error ranged from 0.32-0.53 log units, without any additional fitting. This
596 model also takes sorption capacities into consideration in order to predict the contribution of different
597 biomolecules (Eg. protein) to various tissue types (Eg .muscle). The model can assist in assessing toxicity
598 and for interpreting biomonitoring results.

599
600 Connectivity Indices

601 Connectivity indices were developed by exploiting the polarizability of atoms in a molecule, which can
602 be done because polarizability is a fundamental property which ultimately relates to biological response
603 (Kier *et al.*, , 1975). These indices can be used to calculate various physicochemical and thermodynamic
604 properties using LFERs (Boethling and Mackay, 2000). The mathematical derivation of connectivity
605 indices are simple and are based on discrimination of bond types in a molecule according to the number
606 non-hydrogen neighbours. The individual molecular connectivity values (χ) reflect the molecular
607 connectivity of a molecule, and is expressed as an expansion of weight bond contributions. This method
608 can be advantageous because the indices include information on topological properties. A disadvantage
609 to this approach is the need to deduce the indices, as well as the lack of physical meaning associated
610 with the indices, both of which could lead to errors.

611

612

613 COSMO-RS (COnductor-like Screening MOdel for Real Solvents):

614 COSMO-RS is a solvation theory based on the determination of solvation energy in a realistic solvent,
615 thermodynamically referenced to a perfect or ideal solvation environment. The theory is built on the
616 earlier COSMO (Conductor-Like Screening Model) theory, extended to realistic solvation, hence the
617 name COSMO-RS. COSMO is based on solvation theory in which quantum-mechanical calculations are
618 used to predict by energy minimization the preferred configuration and electron distribution of a
619 molecule within a an electrostatically responsive solvation environment. The response of the solvent in
620 the COSMO-RS calculation is set to the ideal or perfect solvent (i.e., infinite dielectric constant) giving a
621 thermodynamic reference state that can be used to determine relative Gibbs free energies of solvation
622 in real solvents. COSMO-RS theory then calculates the relative energy of interaction of the molecule
623 with its surrounding 'solvent' with respect to this ideal COSMO state (Klamt, 2005). This stands in
624 contrast to traditional quantum-mechanical approaches that use the non-interacting gas phase state of
625 the molecule as a thermodynamic reference state. The COSMO state is much closer to a realistic
626 solvation state, so the error in using it as a reference for real solvents is much smaller than that
627 associated with the gas phase.

628 The COSMO-RS methodological approach (COSMOthermX, 2013) has the potential to revolutionize the
629 prediction of thermochemical properties of solvated molecules, since it is essentially a “first-principles”
630 approach that relies on only a few optimized, universal parameters. The scope of the COSMO-RS
631 approach is such that no data are needed for the particular family of molecule in question in order to
632 obtain a predicted value of K_{ow} or other thermodynamic parameter with equally good accuracy
633 compared to its application with known and well-characterized molecular families. Thus, the problem of
634 predicting properties of new chemicals for which no related family of experimentally characterized

635 molecules exists has been solved by dealing directly with the properties of the molecule itself to a
636 degree of accuracy that is independent of how much is known about the molecule experimentally.

637 The COSMO-RS method involves a two-step approach to determination of K_{OW} and other
638 physicochemical properties of molecules in solution. The first part, COSMO, treats the molecule of
639 interest as if it were embedded in a continuous environment (the dielectric), which responds
640 electrostatically to the way in which the molecule's electrons are distributed over the molecule's nuclear
641 backbone. In a COSMO-RS calculation, the dielectric or solvation environment is a perfect conductor,
642 providing exactly the same amount of charge (with opposite sign) to the molecule's surface at all points.
643 Thus, the molecule's electron distribution is exactly matched by the dielectric, giving a "perfectly
644 solvated" molecular species. The key point is that both the geometry of the molecule and the electron
645 distribution over the molecule are optimized under the influence of the solvating dielectric, giving
646 modeled geometries and charge distributions that are realistic and close to the likely form of the
647 solvated molecule in a real solvent. Once this optimized solvated geometry and charge distribution is
648 determined, it may be used in any further computation of COSMO-RS-derived properties.

649 The second part of the COSMO-RS computation is a probability-based determination of the extent to
650 which the solvents in question (for example, octanol and water) deviate from a perfect solvent (i.e. from
651 a conducting dielectric). The extent of interaction between each solvent and solute is determined
652 separately by representing the solvent and solute as a series of charged surface areas that cover the
653 molecules at a distance from the backbone that is approximately that of intermolecular contact
654 distances (the Van der Waals distance). Then the probability that the solvent and solute will meet with
655 any given set of surface charges during a random collision is determined statistically, and the averaged
656 sum of all such collisional interactions is used to determine the deviation of the solvent in question from
657 the perfectly conducting ideal solvent. The change in energy from the ideal solvent may be determined.

This energy difference is easily converted to chemical activity, chemical potential, and ultimately partition coefficients. To do such calculations for essentially all organic chemicals, one requires only 5-10 intrinsic parameters and about 10 element-specific parameters, two per element, (Klamt *et al.* 1998). These optimized parameters are physically based on measurable physical interactions but actually determined by fitting to a large reliable experimental data set. Beyond this initial parameterization, no further fitting to experimental data is required in the application of COSMO-RS theory to most chemical partitioning problems involving realistic molecules. Figure 6 is a graph of experimental vs. COSMO-RS calculated $\log K_{OW}$ values reported by Klamt (1995).

Figure 6: (Klamt, 1995) Experimental vs. predicted $\log K_{OW}$ values using COSMO-RS method for a variety of well-studied organic molecules. This first-principles method calculates energy differences using several molecular properties, including charge densities and hydrogen bonding parameters. $r=0.96$, $s=0.28$. Reprinted with permission from Klamt, A. 1995. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* 99(7), 2224-2235. Copyright 2015 American Chemical Society.

The COSMO-RS approach is similar to the UNIFAC approach discussed above, in that it has as its basis pairwise surface contacts between solvent and solute. For example, Xue *et al.* (2012) compared several

675 UNIFAC-based approaches with two early secondary implementations of COSMO-RS theory, and found
676 the former to give better results with chemical species from the UNIFAC parameterization database.
677 Klamt (2012) has pointed out that this is not a fair comparison, and noted that with the more developed
678 primary implementation of COSMOtherm, COSMO-RS performs much better. However, as expected,
679 the modified UNIFAC(Dortmund), which is a modern and extensively improved version of UNIFAC,
680 outperforms COSMOtherm with simple, structurally-related compounds for which it was designed.
681 Being based only on a priori calculations, COSMO-RS cannot compete with theories that are
682 parametrized with extensive experimental data, but will generally outperform other methods for
683 unusual molecular system and systems for which little or no experimental data are available.

684

685 Present Practice and Future Developments

686 In this review we have sought to outline the history of partition coefficients from the early recognition
687 by Henry (1803) and later Nernst (1891) that solutes partition with defined and measurable
688 concentration ratios between immiscible phases or solvents. The thermodynamic criteria underlying
689 partitioning were later developed by Gibbs (1876, 1878) and Lewis (1901). The incentive for
690 experimental measurements of partitioning was largely driven by interest in anaesthesia and toxicity
691 and led to the pioneering studies of narcosis by Overton (1901) and Meyer (1899). The compilation and
692 interpretation of extensive databases on K_{ow} by Hansch and Leo (2005) resulted in quantitative
693 structure-activity relationships (QSARs) and quantitative structure-property relationships (QSPRs) in
694 which partition coefficients were correlated with, and predicted from, a variety of molecular descriptors.
695 The Abraham poly-parameter LFERs represent the most sophisticated QSPRs currently available.

696

These QSAR predictions have relied on experimental determinations. As confidence in predictions increased and the number of chemicals of environmental and biological concern has grown there has been a trend to rely increasingly on predictive methods and to confirm the reliability of these methods using experimental data. The advent of low cost and high speed computing has facilitated the development of fundamental quantum mechanical methods of calculating solute-solvent free energy interactions thus leading to “first principles” estimation methods that require no experimental data, except for validation.

We suggest that the evolution of the science of partition coefficients in the future will involve three complementary efforts. First is continued development of novel and more accurate experimental methods for determining partition coefficients, especially for substances with extreme and unusual properties including ionogenic substances. These data are inherently valuable and can serve to validate predictive methods.

Second is continued development and dissemination of QSARs for both scientific and regulatory purposes with emphasis on readily applied and validated methods treating a wide range of molecular structures.

Third is continual development of “first principles” computational methods such as COSMO-RS with emphasis on validation, application to less studied chemicals and mixtures, complex solvents, especially complex and biotic in nature, and dissemination of simpler software that can be more readily employed by non-experts.

These efforts should contribute to a growth in understanding of environmental partitioning phenomena and thus to improved evaluation of the environmental risks and benefits of chemical substances.

718 **Acknowledgements:**

719 We thank Trent University and DMER Ltd for financial support.

Draft

720 References:

- 721 Åberg, A., Macleod, M., and Wiberg, K. 2008. Physical-chemical property data for Dibenzo-p-dioxin (DD),
722 Dibenzofuran (DF), and Chlorinated DD/Fs: A Critical Review and Recommended Values. J. Phys. Chem.
723 Ref. Data. **37**, 1997-2008.
- 724 Abraham, M.H. 1993. Scales of solute hydrogen bonding: Their construction and application to
725 physicochemical and biochemical processes. Chem. Soc. Rev., **22**, 73–83.
- 726 Abraham, M.H., Smith, R.E., Luchtefeld, R., Boorem, A.J., Luo, R., and Acree Jr., W.E. 2010. Prediction of
727 solubility of drugs and other compounds in organic solvents. J. Pharm. Sci. **99**(3), 1500-1515.
- 728 Abraham, M.H., and McGowan, J.C. 1987. The use of characteristic volumes to measure cavity terms in
729 reversed phase liquid chromatography. Chromatographia **23**, 243-246.
- 730 Abrams, D.S., and Prausnitz, J.M. 1975. Statistical thermodynamics of liquid mixtures: a new expression
731 for the excess Gibbs energy of partly or completely miscible systems. AIChE Journal. **21**(1), 116-128.
- 732 ACD/Labs, ABSOLV version 2015. 2015. Advanced Chemistry Development, Inc., Toronto, ON, Canada,
733 www.acdlabs.com [Accessed October 2015]
- 734 ACD/Labs, ACD/LogD. 2015. <http://www.acdlabs.com/products/percepta/predictors/logd/> [Accessed
735 October 2015]
- 736 ADMET Predictor(TM) version 8.0, Simulations Plus, Inc. 2015. Lancaster, CA, USA.
737 <http://www.simulations-plus.com> [Accessed October 2015]
- 738 ARChem, Automated reasoning in chemistry. 2015 <http://www.archemcalc.com/sparc.html> [Accessed
739 October 2015]

- 740 Beyer, A., Wania, F., Gouin, T., Mackay, D., and Matthies, M. 2002. Selecting internally consistent
741 physicochemical properties of organic compounds. *Environ. Toxicol. Chem.* **21**, 941-953.
- 742 Boethling, R., and Mackay, D. 2000. *Handbook of Property Estimation Methods for Chemicals*. Lewis
743 Publishing. pp 93-105.
- 744 Buchwald, P., and Bodor, N. 1998. Molecular size-based model to describe simple organic liquids. *J.*
745 *Phys. Chem. B.* **102**, 5715-5726.
- 746 Buchwald, P. 2015. pbuchwald@miami.edu
- 747 Carson, Rachel L. 1962. *Silent Spring*. Boston: Houghton-Mifflin.
- 748 ChemAxon. <https://www.chemaxon.com/about-us/>. [Accessed October 2015]
- 749 Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., and Lai, L. 2007. Computation of
750 octanol water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.*
751 **47**, 2140–2148.
- 752 Chiou, C.T., Schmedding, D.W., and Manes, M. 1982. Partitioning of Organic Compounds in Octanol-
753 Water Systems. *Environ. Sci. Tech.* **16**, 6-9.
- 754 COSMOthermX VC30_1401, COSMOlogic GmbH & Co. KG, V4. Leverkusen, Germany,
755 cosmotherm@cosmologic.de [Accessed October 2015]
- 756 Dortmund Data Bank Software & Separation Technology (DDBST) GmbH. 2015. Center for Applied
757 Thermodynamics. Marie-Curie-Str. 10. D-26129 Oldenburg, Germany. Available from
758 <http://www.ddbst.com/contact.html> [Accessed August 2015].

- 759 Ebbing, D.E. and Gammon, S.D. 2005. General Chemistry. Houghton Mifflin Co. New York, Boston.
760 pp497.
- 761 EPISuite. 2000-2012. U.S. Environmental Protection Agency (EPA). Developed by EPA's office of pollution
762 prevention toxics and Syracuse Research Corporation (SRC). Available from
763 <http://www.epa.gov/opptintr/exposure/pubs/episuite.htm> [Accessed August 2015].
- 764 Ferguson, J. 1939. The use of chemical potentials as indices of toxicity. Proc. R. Soc. London. Ser. B. **127**,
765 387-404.
- 766 Fredenslund, A., Jones, R.L., and Prausnitz, J.M. 1975. Group contribution estimation of activity
767 coefficients in nonideal liquid mixtures. AIChE Journal, **21**(6), 1086-1099.
- 768 Gibbs, J.W., 1876. On the equilibrium of heterogeneous substances Part I. The transactions of the
769 Connecticut Academy of Arts and Sciences. **3**, 108-248.
- 770 Gibbs, J.W., 1878. On the equilibrium of heterogeneous substances Part II. The transactions of the
771 Connecticut Academy of Arts and Sciences. **3**, 343-524.
- 772 Goss, K.U., Brown, T., and Brown, T. 2013. General model for estimating partition coefficients to
773 organisms and their tissues using biological compositions and PPLFERS. Env. Sci. Tech. **47**, 6630-6639.
- 774 Hansch, C., Leo, A., and Hoekman, D. 1995. Hydrophobic, electronic, and steric constants. Washington,
775 DC: American Chemical Society.
- 776 Hansch, C., Hoekman, D., Leo, A., Zhang, L., and Li, P. 2005. The expanding role of quantitative structure-
777 activity relationships (QSARs) in toxicology. Toxicol. Lett. **79**, 45-53.
- 778 Henry, W. 1803. Experiments on the quantity of gases absorbed by water, at different temperatures and
779 different pressures. Philosophical Transactions with the Royal Society of London. **93**, 29-46

- 780 Hilal, S.H., Carreira, L.A., Karickhoff, S.W., and Melton, C.M. 1994. Estimation of gas-liquid
781 chromatographic retention times from molecular structure. *J. Chromatogr. A.* **662**, 269–280.
- 782 Hornig, M., and Klamt, A. 2005. COSMOfrag: A novel tool for high-throughput ADME property prediction
783 and similarity screening based on quantum chemistry. *J. Chem. Inf. Model.* **45**, 1169–1177.
- 784 Hung, H., MacLeod, M., Guardans, R., Scheringer, M., Barra, R., Harner, T., and Zhang, G. 2013. Toward
785 the next generation of air quality monitoring: Persistent organic pollutants. *Atmos. Environ.* **80**, 591-598.
- 786 Japertas P, Didziapetris R, and Petrauskas A. 2002. Fragmental methods in the design of new
787 compounds. Applications of the Advanced Algorithm Builder. *Quant. Struct-Activ. Rel.* **21**, 23–37.
- 788 Junghans, M., and Pretsch, E. 1997. Estimation of partition coefficients of organic compounds: Local
789 database modeling with uniform-length structure descriptors. *Fresenius J. Anal. Chem.* **359**, 88– 92.
- 790 Kier, L., Hall, L., and Murray, W. 1975. Molecular Connectivity III: Relationship to Partition Coefficients. *J.*
791 *Pharm. Sci.* **64**(12), 1978-1981.
- 792 Kipton, H., Powell, J., and Town, R.M. 1992. Solubility and fractionation of humic acid: effect of pH and
793 ionic edium. *Analytica Chemica Acta.* **267**, 47-54.
- 794 Klamt, A. 1995. Conductor-like screening model for real solvents: a new approach to the quantitative
795 calculation of solvation phenomena. *J. Phys. Chem.* **99**(7), 2224-2235.
- 796 Klamt, A. 2005. COSMO-RS: From quantum chemistry to fluid phase thermodynamics and drug design.
797 Elsevier. Leverkusen, Germany.
- 798 Klamt, A., Jonas, V., Bürger, T., and Lorenz, J.C.W. 1998. Refinement and Parametrization of COSMO-RS.
799 *J. Phys. Chem. A*, **102**, 5074-5085.

- 800 Klamt, A. 2012. Comment on “Comparison of the a priori COSMO-RS models and group contribution
801 methods: original UNIFAC, modified UNIFAC(Do), and modified UNIFAC(Do) consortium”. Ind. Eng.
802 Chem. Res. **51**, 13538-13540.
- 803 Lewis, G.N. 1901. The law of physico-chemical change. Proceedings of the American academy of arts and
804 sciences. **37**(3), 49-69.
- 805 Li, H., Ellis, D., and Mackay, D. 2007. Measurement of low air-water partition coefficients of organic
806 acids by evaporation from a water surface. J. Chem. Eng. Data, **52**, 1580-1584.
- 807 Li, N.Q., Wania, F., Lei, Y.D., and Daly, G.L. 2003. A comprehensive and critical compilation, evaluation,
808 and selection of physical chemical property data for selected polychlorinated biphenyls. J. Phys. Chem.
809 Ref. Data **32**, 1545-1590.
- 810 Lyman, W.J., and Reehl, W.F. 1982. Handbook of chemical property estimation methods. McGraw-Hill
811 Book Company, New York.
- 812 Mackay, D. 2001. Multimedia Environmental Models: The Fugacity Approach. CRC Press, 83-116.
- 813 Mackay, D., and Cole, J. 2000. Correlating environmental partitioning properties of organic compounds:
814 the three solubility approach. Environ. Toxicol. Chem. **19**, 265-270.
- 815 Mackay, D., Shiu, W.Y., Ma, K.C., and Lee, S.C. 2006. Physical chemical properties and environmental
816 fate for organic chemicals. Talyor & Francis publishing, CRC group. Boca Raton.
- 817 Mannhold, R., Poda, G.I., Ostermann, C., and Tetko, IV. 2009. Calculation of molecular lipophilicity: state
818 of the art and comparison of logP methods on more than 90 000 compounds. J. Pharm. Sci. **98**(3), 861-
819 893.
- 820 Meyer, H.H. 1899. Zur Theorie der Alkoholnarkose. Arch. Exp. Pathol. Pharmacol. **42**, 109–118.

- 821 Meylan, W.M., and Howard, P.H. 2000. Estimating logP with atom/fragments and water solubility with
822 log P. *Perspect. Drug. Discov. Des.* **19**, 67– 84.
- 823 Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I., and Matsushita, Y. 1992. Simple method of calculating
824 octanol/water partition coefficient. *Chem. Pharm. Bull.* **40**, 127–130.
- 825 Nernst, W. 1891. Distribution of a substance between two solvents, etc. *J. Chem. Soc.* **8**, 110-139.
- 826 Nowosielski, B.E. 1997. Experimental study of octanol-water partition coefficients for 2,4,6-
827 trichlorophenol and pentachlorophenol: derivation of an empirical model of chlorophenol partitioning
828 behaviour. M.Sc. thesis, department of Earth and Planetary Sciences, McGill University, Montreal, QC.
- 829 OECD Guidelines for testing of chemicals, section 1. Test No. 123: Partition coefficient (1-octanol/water):
830 Slow-stirring method. [http://www.oecd-ilibrary.org/environment/test-no-123-partition-coefficient-1-](http://www.oecd-ilibrary.org/environment/test-no-123-partition-coefficient-1-octanol-water-slow-stirring-method_9789264015845-en;jsessionid=2gqv5qrme4y2b.x-oecd-live-03)
831 [octanol-water-slow-stirring-method_9789264015845-en;jsessionid=2gqv5qrme4y2b.x-oecd-live-03](http://www.oecd-ilibrary.org/environment/test-no-123-partition-coefficient-1-octanol-water-slow-stirring-method_9789264015845-en;jsessionid=2gqv5qrme4y2b.x-oecd-live-03)
832 [Accessed October 2015]Overton, C.E. 1901. Studien über die Narkose zugleich ein Beitrag zur
833 allgemeinen Pharmakologie. Gustav Fischer, Jena, Switzerland.
- 834 Pankow, J., Mader, B.T., Isabelle, L.M., Luo, W., and Pavlick, A., Liang, C. 1997. Conversion of nicotine in
835 tobacco smoke to its volatile and available free-base form through the action of gaseous ammonia.
836 *Environ. Sci. Technol.* **31**, 2428-2433.
- 837 Parnis, J.M., Mackay, D., and Harner, T. 2015. Temperature dependence of Henry's law constants and
838 K_{OA} for simple and heteroatom-substituted PAHs by COSMO-RS. *Atmos. Environ.* **110**, 27-35.
- 839 Petrauskas, A.A, and Kolovanov, E.A. 2000. ACD/LogP method description. *Perspect. Drug. Discov. Des.*
840 **19**, 99–116.

- 841 Platts, J.A., Butina, D., Abraham, M.H., and Hershey, A. 1999. [Estimation of molecular linear free energy](#)
842 [relation descriptors using a group contribution approach](#). J. Chem. Inf. Comput. Sci. **39**, 835-845.
- 843 Sangster, J., 1989. Octanol-water partition coefficients of simple organic compounds. J. Phys. Chem Ref.
844 Data. **18**(3), 1111-1226.
- 845 Schenker U., MacLeod M., Scheringer M., and Hungerbühler K. 2005. Improving data quality for
846 environmental fate models: a least-squares adjustment procedure for harmonizing physicochemical
847 properties of organic compounds. Environ. Sci. Technol. **39**, 8434-8431.
- 848 Shen L., and Wania F. 2005. Compilation, evaluation, and selection of physical-chemical property data
849 for organochlorine pesticides. J. Chem. Eng. Data **50**, 742-768.
- 850 Talete SRL. 2013. Milano, Italy. <http://www.talete.mi.it> [Accessed October 2015]
- 851 Tetko, I.V., and Tanchuk, V.Y. 2002. Application of associative neural networks for prediction of
852 lipophilicity in ALOGPS 2.1 program. J. Chem. Inf. Comput. Sci. **42**, 1136–1145.
- 853 Upstream Solutions. 2015. <http://www.upstream.ch> [Accessed October 2015]
- 854 Virtual Computational Chemistry Laboratory, 2011. <http://www.vcclab.org/lab/alogps/> [Accessed
855 October 2015]
- 856 Wania, F. 2003. Assessing the potential of persistent organic chemicals for long-range transport and
857 accumulation in polar regions. Environ. Sci. Technol. **37**, 1344-1351.
- 858 Xiao H., Li N., and Wania F. 2004. Compilation, evaluation, and selection of physical-chemical property
859 data for α -, β -, and γ -hexachlorocyclohexane. J. Chem. Eng. Data. **49**, 173-185.

- 860 Xu, S., and Kropscott, B. 2012. Method for simultaneous determination of partition coefficients for cyclic
861 volatile methylsiloxanes and dimethylsilanediol. *Anal. Chem.* **84**, 1948-1955.
- 862 Xue, Z., Mu, T., and Gmehling, J. 2012. Comparison of the a priori COSMO-RS models and group
863 contribution methods: original UNIFAC, modified UNIFAC(Do), and modified UNIFAC(Do) consortium.
864 *Ind. Eng. Chem. Res.* **51**, 11809-11817.
- 865 Zhang, K., Pereira, A.D.S, and Martin, J.W. 2015. Estimates of octanol-water partitioning for thousands of
866 dissolved organic species in oil sands process-affected water. *Environ. Sci. Technol.* **49**(14), 8907-8913.

Draft

867 **Tables:**

868 Table 1: Various methods commonly used to predict K_{ow} reported by Mannhold *et al.* (2009). Methods
 869 chosen were based on highest RMSE values while representing all method types. RMSE is root-mean-
 870 square error, n is number of test chemicals.

Method	Approach	RMSE	% of Molecules with Error Range <0.5	% of Molecules within Error Range of 0.5-1	% of Molecules with Error Range >1	Reference
AB/LogP	Fragment	0.41	84	12	4	Japertas <i>et al.</i> (2002)
S+LogP	Neural network	0.45	76	22	3	(ADMET Predictor, 2015)
ACD/logP	Fragment	0.50	75	17	7	Petrauskas and Kolovanov (2000)
CLOGP	Fragment	0.52	74	20	6	Hansch <i>et al.</i> (1995)
XLOGP3	Atom	0.62	60	30	10	Cheng <i>et al.</i> (2007)
KowWIN	Fragment	0.64	68	21	11	Meylan and Howard (2000) EPISuite (2015)
ABSOLV	Whole- molecule	0.75	53	30	17	(Abraham, 1993)
COSMOFrag	Fragment	0.84	48	26	19	Hornig and Klamt (2005)

871

872 **Figure captions:**

873 Figure 1: Simple structure-property correlations between chemical properties of chlorobenzenes from
874 Mackay *et al.* (2006) A) Log K_{OW} (octanol-water partition coefficient) vs. Le Bas molar volume at the (top
875 left) B) Log S_L (subcooled solubility) vs molar volume (top right) C) Log K_{AW} (air-water partition
876 coefficient) vs molar volume (bottom). Reproduced from Mackay *et al.* (2006) with permission from
877 Taylor and Francis, 2015.

878 Figure 2: Chemical space plot of log K_{AW} vs log K_{OW} with log K_{OA} as diagonals, illustrating the wide
879 variation in partition coefficients and showing the locations of selected chemicals. The bold lines
880 represent estimated fractions of the chemical mass partitioning into air, water and octanol for assumed
881 volumes of these solvents, typical of environmental proportions. Values come from Mackay *et al.*
882 (2006), EPISuite, and Nowosielski (1997). Taken from Mackay (2001) and modified with permission from
883 Taylor and Francis, 2015.

884 Figure 3: Chemical space diagram from Wania (2003) containing partition coefficient predictions for
885 certain classes of chemicals including chlorobenzenes (CBZs), PCBs, polychlorinated dibenzo-p-dioxins
886 (PCDDs), and polychlorinated dibenzofurans (PCDFs). Each letter represents a chemical with specific
887 physico-chemical properties: A is a particle-sorbed chemical, B is a water-soluble chemical, C is a volatile
888 water soluble chemical, D is a multimedia chemical, and E is a very hydrophobic semivolatile chemical.
889 Reprinted with permission from Wania, F. 2003. Assessing the potential of persistent organic chemicals
890 for long-range transport and accumulation in polar regions. *Environmental Science & Technology*, 37,
891 1344-1351. Copyright 2015 American Chemical Society.

892 Figure 4: An illustration showing variables that are related to one another, including partition
893 coefficients (K), solubilities (C^S), liquid vapour pressure (P^S_L), fugacity (f) and activity (γ), for a liquid
894 chemical, at a given temperature.

895 Figure 5: A representation showing related variables relevant to environmental partitioning. The central
896 quantities in black font are the fundamental thermodynamic parameters, those in red are the measured
897 physico-chemical properties and those in green are the related quantities of environmental interest.

898 Figure 6: (Klamt, 1995) Experimental vs. predicted log K_{ow} values using COSMO-RS method for a variety
899 of well-studied organic molecules. This first-principles method calculates energy differences using
900 several molecular properties, including charge densities and hydrogen bonding parameters. $r = 0.96$, s
901 $= 0.28$. Reprinted with permission from Klamt, A. 1995. Conductor-like screening model for real solvents:
902 a new approach to the quantitative calculation of solvation phenomena. J. Phys. Chem. 99(7), 2224-
903 2235. Copyright 2015 American Chemical Society.

904

Draft

A

B

C