# Computer Aided Olive Oil−Gas Partition Coefficient Calculations[§]

Gilles Klopman,*,[†] Congmei Ding, and Orest T. Macina[‡]

Department of Chemistry, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, and Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15238

Received September 27, 1996[⊗]

Our project aimed at developing a relation between the structure of diverse organic molecules and their ability to partition between olive oil and gas. Olive oil−gas partitions have been used as surrogates for blood−gas partition, which is believed to play a role in the pharmacokinetic profile of organic solvents. We found that linear relationships exist between olive oil−gas ($L_{oil}$) and hexadecane−gas ($L_{hex}$) partition coefficients within certain chemical classes. Log $L_{oil}$ models were also derived from a multivariate regression analysis based on log $L_{oil}$ values of organic molecules and a set of basic functional groups and correction factors found previously to be useful in the calculation of log *P*. Both cross-validation experiments and the calculation of an independent set of 36 compounds demonstrated that our methodology can give fast and accurate log $L_{oil}$ predictions for the classes of compounds represented within the models. Molecular weight was found to be an important factor in the determination of log $L_{oil}$.

## INTRODUCTION

The olive oil−gas partition coefficient ($L_{oil}$) is the ratio of the concentrations that a chemical achieves between olive oil and gas phases at equilibrium. It is an extremely useful property in that it relates to the blood−gas partition, which determines the absorption, distribution, and excretion of organic chemicals in the organism.[1] A good understanding of the olive oil−gas partition coefficient can provide some insight into the pharmacokinetic and toxicokinetic profile of organic drugs.[2] Sato et al. demonstrated that the partition coefficients of biologically active compounds between blood−gas and oil−gas are important factors in structure−activity relationship studies of the biological activity of some xenobiotics.[3] However, very few $L_{oil}$ have been reported in the literature because of the experimental difficulties encountered in measuring these values. Theoretical approaches for $L_{oil}$ are also scarce. It is therefore of great importance to develop methodologies capable of providing fast and reliable $L_{oil}$ estimates.

Abraham et al. found that hexadecane−gas partition coefficients ($L_{hex}$) are valuable in correlations involving many solvent−gas processes.[4] In this paper, possible relationships between $L_{oil}$ and $L_{hex}$ of different classes of chemicals were examined. Relationship were also sought between $L_{oil}$ and the presence of various substructures or groups of atoms. Group contribution is conceptually simple and can give fast and rather acurate estimation of physical-chemical properties, such as water solubility,[5] log *P*,[6] and boiling point.[7] Poor estimations, when encountered, are primarily due to intramolecular interaction between groups in these compounds.[6] The CASE[8,9] program developed in our group can automatically identify these correction factors. By using the summation of group contributions and the Computer Automated Structure Evaluation (CASE) methodology, our laboratory has developed a reliable and accurate log *P* estimation approach for organic compounds.[6] Encouraged by this success, we decided to use the group contribution methodology to develop a program for the *a priori* calculation of log $L_{oil}$. In this study, we assume that the set of basic functional groups and correction factors obtained in our previously described log *P* estimation model are also suitable for the log $L_{oil}$ calculation.

## DATABASES AND METHODS

A total of 119 log $L_{oil}$ values at 310 K and 215 log $L_{hex}$ values at 298 K were extracted from a publication by Abraham et al. (Table 1).[4] One hundred one compounds were given values in both databases. All experimental values are listed in columns 2 and 3 of Table 1. Linear regression analyses were evaluated between the log $L_{oil}$ and log $L_{hex}$ of these 101 molecules as well as for several subsets of similar molecules, see eqs 2−8 below. The correlation equations obtained thereby were then used to predict untested log $L_{oil}$ values from known log $L_{hex}$ values of compounds with structures similar to those used to derive the equations. The predicted results for molecules for which there are no experimental data are listed in column 4 of Table 1.

The molecular structures were encoded using the Klopman Line Notation (KLN).[10] A set of 94 basic group parameters and correction factors from the earlier log *P* estimation model were considered, as shown in Table 2,[6] and a program was developed to automatically identify the occurrence of each of the group parameters in the compounds. The basic group parameters belong to two fundamental types: (1) heavy atoms in their corresponding hybridization state and the number of attached hydrogen(s) and (2) chemical functionalities (OH, CHO, COOH, COO, $CONH_2$, CONH, CON, CON═, CO, CS, NO, $NO_2$, PO, SO, $SO_2$, $NH_2$, NH, CN, SH, etc.). Additional factors correcting for intramolecular group interactions were also identified by the CASE program. The log $L_{oil}$ values, occurrence of the group parameters, and

[†] Case Western Reserve University.
[‡] University of Pittsburgh.
[§] Keywords: CASE; group contribution approach; linear regression analysis; olive oil−gas partition coefficient.
[⊗] Abstract published in *Advance ACS Abstracts,* April 1, 1997.

S0095-2338(96)00140-0 CCC: $14.00 

**Table 1.** Experimental, Calculated Log *L* Values, and Estimation Error[g]

| name | log $L_{hex}$[e] | log $L_{oil}$[e] | log $L_{oil}$[p] | log $L_{oil}$[c] | $\Delta$log $L_{oil}$ | name | log $L_{hex}$[e] | log $L_{oil}$[e] | log $L_{oil}$[p] | log $L_{oil}$[c] | $\Delta$log $L_{oil}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Aliphatic Hydrocarbons** | | | | | | | | | | | |
| ethane | 0.49 | 0.28[a] | | 0.61 | 0.33 | 2-methylhexane | 3.00 | | 2.46[a] | 2.54 | 0.08 |
| propane | 1.05 | 0.74[a] | | 1.03 | 0.29 | 3-methylhexane | 3.04 | | 2.50[a] | 2.54 | 0.04 |
| *n*-butane | 1.62 | 1.27[a] | | 1.44 | 0.17 | 2,2-dimethylpentane | 2.79 | | 2.28[a] | 2.19 | −0.09 |
| 2-methylpropane | 1.41 | 1.05[a] | | 1.29 | 0.24 | 2,4-dimethylpentane | 2.84 | | 2.32[a] | 2.39 | 0.07 |
| *n*-pentane | 2.16 | 1.67[a] | | 1.86 | 0.19 | 2,3-dimethylpentane | 3.02 | | 2.47[a] | 2.39 | −0.47 |
| *n*-hexane | 2.67 | 2.13[a] | | 2.27 | 0.14 | 3,3-dimethylpentane | 2.95 | | 2.41[a] | 2.19 | −0.22 |
| *n*-heptane | 3.17 | 2.59[a] | | 2.69 | 0.10 | 2,2,3-trimethylbutane | 2.85 | | 2.33[a] | 2.04 | −0.29 |
| *n*-octane | 3.68 | 3.04[a] | | 3.10 | 0.06 | 3-ethylpentane | 3.09 | | 2.54[a] | 2.54 | 0.00 |
| *n*-nonane | 4.18 | 3.48[a] | | 3.52 | 0.04 | 2,2,4-trimethylpentane | 3.12 | | 2.56[a] | 2.46 | −0.10 |
| *n*-decane | 4.69 | 3.92[a] | | 3.93 | 0.01 | cycloheptane | 3.53 | | 2.92[a] | 3.13 | 0.21 |
| *n*-undecane | 5.19 | 4.36[a] | | 4.35 | −0.01 | cyclooctane | 4.12 | | 3.43[a] | 3.54 | 0.11 |
| *n*-dodecane | 5.70 | 4.80[a] | | 4.76 | −0.04 | methylcyclopentane | 2.77 | | 2.26[a] | 2.26 | 0.00 |
| *n*-tridecane | 6.20 | 5.24[a] | | 5.18 | −0.06 | methylcyclohexane | 3.25 | | 2.68[a] | 2.67 | −0.15 |
| *n*-tetradecane | 6.70 | 5.69[a] | | 5.59 | −0.1 | adamantane | 4.77 | | | | |
| *n*-pentadecane | 7.21 | 6.13[a] | | 6.01 | −0.12 | propene | 0.95 | | | | |
| *n*-hexadecane | 7.71 | 6.57[a] | | 6.42 | −0.15 | but-1-ene | 1.49 | | | | |
| cyclopropane | 1.31 | 1.07[a] | | 1.49 | 0.42 | pent-1-ene | 2.01 | | | | |
| cyclopentane | 2.45 | 2.00[a] | | 2.31 | 0.31 | hex-1-ene | 2.55 | | | | |
| cyclohexane | 2.91 | 2.44[a] | | 2.72 | 0.28 | hept-1-ene | 3.06 | | | | |
| ethene | 0.29 | 0.10 | | 0.11 | 0.01 | oct-1-ene | 3.59 | | | | |
| ethyne | 0.15 | 0.24 | | 0.21 | −0.03 | buta-1,3-diene | 1.54 | | | | |
| 2-methylbutane | 2.01 | | 1.60[a] | 1.71 | 0.11 | cyclopentadiene | 2.22 | | | | |
| 2-methylpentane | 2.55 | | 2.07[a] | 2.12 | 0.05 | propyne | 1.02 | | | | |
| 3-methylpentane | 2.60 | | 2.11[a] | 2.12 | 0.01 | cyclohexene | | | | 2.61 | −0.05 |
| 2,3-dimethylbutane | 2.51 | | 2.03[a] | 1.97 | −0.06 | 2,3,4-trimethylpentane | | | | 2.65 | −0.17 |
| 2,2-dimethylbutane | 2.32 | | 1.87[a] | 1.78 | −0.09 | isoprene | | | | 1.14 | 0.20 |
| **Aromatic Hydrocarbons** | | | | | | | | | | | |
| benzene | 2.80 | 2.60[b] | | 2.64 | 0.04 | *p*-xylene | 3.86 | 3.53[b] | | 3.57 | 0.04 |
| toluene | 3.34 | 3.08[b] | | 3.10 | 0.02 | cumene | 4.10 | 3.79[b] | | 3.90 | 0.11 |
| ethylbenzene | 3.76 | 3.49[b] | | 3.58 | 0.09 | styrene | 3.91 | 3.68[b] | | 3.50 | −0.18 |
| *n*-propylbenzene | 4.22 | 3.99[b] | | 4.05 | 0.06 | allylbenzene | 4.23 | 3.91[b] | | 3.97 | 0.06 |
| *n*-butylbenzene | 4.69 | 4.46[b] | | 4.53 | 0.07 | *m*-methylstyrene | | | | 3.96 | −0.21 |
| *o*-xylene | 3.94 | 3.64[b] | | 3.57 | −0.07 | *p*-methylstyrene | | | | 3.96 | −0.18 |
| *m*-xylene | 3.86 | 3.52[b] | | 3.57 | 0.05 | | | | | | |
| **Alcohols, Ethers, and Phenols** | | | | | | | | | | | |
| methanol | 0.92 | 1.47[c] | | 1.43 | −0.04 | allyl alcohol | 2.00 | | 2.37[c] | 2.30 | −0.07 |
| ethanol | 1.48 | 1.96[c] | | 1.90 | −0.06 | cyclohexanol | 3.67 | | 3.86[c] | 3.66 | −0.20 |
| propan-1-ol | 2.10 | 2.50[c] | | 2.38 | −0.12 | *sec*-butyl alcohol | 2.34 | | | | |
| propan-2-ol | 1.82 | 2.16[c] | | 2.15 | −0.01 | *o*-cresol | 4.24 | | | | |
| butan-1-ol | 2.60 | 2.94[c] | | 2.85 | −0.09 | *m*-cresol | 4.33 | | | | |
| *tert*-butyl alcohol | 2.02 | 2.27[c] | | 1.73 | −0.54 | *p*-cresol | 4.31 | | | | |
| pentan-1-ol | 3.11 | 3.38[c] | | 3.33 | −0.05 | 2-isopropylphenol | 4.92 | | | | |
| hexan-1-ol | 3.61 | 3.82[c] | | 3.74 | −0.08 | diethyl ether | 2.06 | 1.81 | | 1.99 | 0.18 |
| heptan-1-ol | 4.11 | 4.26[c] | | 4.21 | −0.05 | diisopropyl ether | 2.60 | 2.15 | | 2.64 | 0.49 |
| octan-1-ol | 4.62 | 4.70[c] | | 4.69 | −0.01 | di-*n*-butyl ether | 4.00 | 3.42 | | 3.89 | 0.47 |
| nonan-1-ol | 5.12 | 5.15[c] | | 5.16 | 0.01 | THF | 2.53 | 2.39 | | 2.45 | 0.06 |
| decan-1-ol | 5.63 | 5.59[c] | | 5.63 | 0.04 | 1,4-dioxane | 2.80 | 2.83 | | 2.77 | −0.06 |
| benzyl alcohol | 4.44 | 4.73 | | 4.72 | −0.01 | 1,2-dimethoxyethane | 2.65 | 2.55 | | 2.26 | −0.29 |
| phenol | 3.86 | 4.29 | | 3.61 | −0.68 | methylal | | 1.96 | | 1.78 | −0.18 |
| isobutyl alcohol | 2.40 | | 2.73[c] | 2.70 | 0.03 | divinyl ether | | 1.78 | | 1.78 | 0.00 |
| pentan-2-ol | 2.84 | | 3.12[c] | 3.10 | −0.02 | di-*n*-propyl ether | 2.99 | | | | |
| hexan-2-ol | 3.34 | | 3.57[c] | 3.57 | 0.00 | propylene oxide | 1.77 | | | | |
| heptan-2-ol | 3.84 | | 4.01[c] | 4.05 | 0.04 | anisole | 3.93 | | | | |
| octan-2-ol | 4.34 | | 4.46[c] | 4.45 | −0.01 | isopentyl alcohol | | | | 3.10 | 0.10 |
| decan-2-ol | 5.36 | | 5.36[c] | 5.40 | 0.04 | | | | | | |
| **Aldehydes and Ketones** | | | | | | | | | | | |
| methanal | | 1.41 | | 0.76 | −0.65 | heptan-2-one | 3.76 | 3.83[d] | | 3.75 | −0.08 |
| ethanal | 1.23 | | | | | cyclopentanone | 3.12 | 3.20[d] | | 2.61 | −0.59 |
| propanal | 1.81 | | | | | hexan-3-one | 3.31 | | 3.30[d] | 3.28 | −0.02 |
| butanal | 2.27 | | | | | heptan-3-one | 3.81 | | 3.77[d] | 3.75 | −0.02 |
| pentanal | 2.77 | | | | | heptan-4-one | 3.82 | | 3.78[d] | 3.75 | −0.03 |
| hexanal | 3.37 | | | | | MeCOBu-*t* | 2.89 | | 2.90[d] | 2.79 | −0.11 |
| propanone | 1.76 | 1.92[d] | | 1.85 | −0.07 | octan-2-one | 4.26 | | 4.19[d] | 4.16 | −0.03 |
| butanone | 2.29 | 2.36[d] | | 2.33 | −0.03 | octan-3-one | 4.31 | | 4.24[d] | 4.23 | −0.01 |
| pentan-2-one | 2.76 | 2.70[d] | | 2.80 | 0.1 | nonan-2-one | 4.75 | | 4.66[d] | 4.64 | 4.66 |
| pentan-3-one | 2.81 | 2.72[d] | | 2.80 | 0.08 | cyclohexanone | 3.62 | | 3.58[d] | 3.08 | −0.50 |
| hexan-2-one | 3.26 | 3.21[d] | | 3.28 | 0.07 | acetophenone | 4.48 | | | | |
| MeCOBu-*i* | 3.05 | 2.97[d] | | 3.13 | 0.16 | | | | | | |
| **Esters** | | | | | | | | | | | |
| methyl formate | 1.46 | 1.56[e] | | 1.49 | −0.07 | butyl propanoate | 3.86 | 3.67[e] | | 3.73 | 0.06 |
| ethyl formate | 1.90 | 1.96[e] | | 1.97 | 0.01 | *n*-hexyl acetate | 4.38 | | 4.03[e] | 4.14 | 0.11 |
| *n*-butyl formate | 2.92 | 2.87[e] | | 2.91 | 0.04 | methyl propanoate | 2.46 | | 2.44[e] | 2.31 | −0.13 |
| methyl acetate | 1.96 | 2.02[e] | | 1.83 | −0.19 | methyl butanoate | 2.94 | | 2.84[e] | 2.78 | −0.06 |
| ethyl acetate | 2.38 | 2.36[e] | | 2.31 | −0.05 | ethyl butanoate | 3.38 | | 3.20[e] | 3.26 | 0.06 |
| *n*-propyl acetate | 2.88 | 2.78[e] | | 2.78 | 0.00 | methyl pentanoate | 3.44 | | 3.25[e] | 3.26 | 0.01 |
| *n*-butyl acetate | 3.38 | 3.20[e] | | 3.26 | 0.06 | methyl hexanoate | 3.98 | | 3.70[e] | 3.73 | 0.03 |
| *n*-pentyl acetate | 3.88 | 3.48[e] | | 3.73 | 0.25 | *n*-propyl formate | | 2.42 | | 2.44 | 0.02 |
| isopropyl acetate | 2.63 | 2.79[e] | | 2.63 | −0.16 | isobutyl acetate | | | | 3.11 | 0.00 |
| ethyl propanoate | 2.88 | 2.71[e] | | 2.78 | 0.07 | isopentyl acetate | | | | 3.58 | 0.11 |

**Table 1** (Continued)

| name | log $L_{hex}^{e}$ | log $L_{oil}^{e}$ | log $L_{oil}^{p}$ | log $L_{oil}^{c}$ | $\Delta$log $L_{oil}$ | name | log $L_{hex}^{e}$ | log $L_{oil}^{e}$ | log $L_{oil}^{p}$ | log $L_{oil}^{c}$ | $\Delta$log $L_{oil}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Halogenated Hydrocarbons | | | | | |
| $CH_3F$ | | 0.06 | | 0.38 | 0.32 | $CF_3CH_2Cl$ | | 1.38 | | 1.30 | −0.08 |
| $C_2H_5F$ | | 0.58 | | 0.85 | 0.27 | $CHClF_2$ | | 0.64 | | 1.08 | 0.44 |
| *n*-$C_3H_7F$ | | 0.92 | | 1.33 | 0.41 | $CF_3CHFBr$ | | 1.46 | | 1.74 | 0.28 |
| *i*-$C_3H_7F$ | | 1.09 | | 1.17 | 0.08 | $CHCl{:}CF_2$ | | 1.15 | | 1.32 | 0.17 |
| perfluoropentane | 0.69 | | | | | $CHF_2CF_2CH_2Br$ | | 2.51 | | 2.21 | −0.30 |
| perfluoroheptane | 1.12 | | | | | $CF_3CHClBr$ | 2.18 | 2.29 | | 2.44 | 0.15 |
| perfluorononane | 1.77 | | | | | $CCl_2FCF_2Cl$ | 2.12 | | | | |
| $CH_2Cl_2$ | 2.02 | 2.14[f] | | 1.85 | −0.29 | $CFBr_3$ | 3.21 | | | | |
| $CHCl_3$ | 2.48 | 2.58[f] | | 2.48 | −0.10 | $CCl_2{:}CH_2$ | 2.11 | | | | |
| $CCl_4$ | 2.82 | 2.53[f] | | 2.91 | 0.38 | allyl chloride | 2.11 | | | | |
| $C_2H_5Cl$ | 1.68 | 1.55[f] | | 1.55 | 0.00 | allyl bromide | 2.51 | | | | |
| $CH_2ClCH_2Cl$ | 2.57 | 2.61[f] | | 2.33 | −0.28 | benzyl chloride | 4.29 | | | | |
| $CHCl_2CH_3$ | 2.35 | 2.27[f] | | 2.17 | −0.10 | hexafluorobenzene | 2.53 | | | | |
| $CCl_3CH_3$ | 2.69 | 2.47[f] | | 2.61 | 0.14 | *p*-difluorobenzene | 2.77 | | | | |
| $CHCl_2CHCl_2$ | 2.83 | 4.12[f] | | 3.58 | −0.54 | chlorobenzene | 3.64 | 3.45 | | 3.92 | 0.47 |
| *n*-$C_3H_7Cl$ | 2.00 | 2.08[f] | | 2.02 | −0.06 | *o*-dichlorobenzene | 4.40 | 4.60 | | 4.72 | 0.12 |
| *n*-$C_4H_9Cl$ | 2.72 | 2.46[f] | | 2.50 | 0.04 | bromobenzene | 4.03 | 4.14 | | 4.67 | 0.53 |
| *n*-$C_5H_{11}Cl$ | 3.22 | 2.99[f] | | 2.97 | −0.02 | *m*-dichlorobenzene | | 4.43 | | 4.72 | 0.29 |
| *trans*-CHCl:CHCl | 2.35 | 2.28[f] | | 2.35 | 0.07 | methyl chloride ($CH_3Cl$) | 1.16 | | | 1.07 | 0.14 |
| $CHCl{:}CCl_2$ | 3.00 | 2.79[f] | | 2.71 | −0.08 | difluoromethane | | | | 0.46 | −0.22 |
| $CCl_2{:}CCl_2$ | 3.58 | 3.22[f] | | 3.07 | −0.15 | fluorochloromethane | | | | 1.15 | −0.20 |
| $CHCl_2CH_2Cl$ | | 3.36 | | 2.95 | −0.41 | bromochloromethane | | | | 2.36 | −0.20 |
| $CCl_3CH_2Cl$ | | 3.63 | | 3.39 | −0.24 | dibromomethane | | | | 2.87 | −0.11 |
| $CH_3CHClCH_2Cl$ | | 2.87 | | 2.65 | −0.22 | chlorodibromomethane | | | | 3.50 | 0.07 |
| $(CH_3)_3CCl$ | 2.22 | | | | | pentachloroethane | | | | 4.02 | 0.19 |
| $CH_3CHClCH_3$ | 1.97 | | | | | hexachloroethane | | | | 4.45 | 0.75 |
| $C_2H_5Br$ | 2.02 | | | | | 1-bromo-2-chloroethane | | | | 2.84 | 0.08 |
| *n*-$C_4H_9Br$ | 3.10 | | | | | 2-chloropropane | | | | 1.87 | 0.03 |
| $CH_2BrCl$ | 2.44 | | | | | *n*-propyl bromide | | | | 2.53 | 0.10 |
| $CH_2Br_2$ | 2.85 | | | | | isopropyl bromide | | | | 2.38 | 0.17 |
| $CHBrCl_2$ | 2.93 | | | | | vinyl chloride | | | | 1.47 | 0.08 |
| $CHBr_2Cl$ | 3.34 | | | | | 1,1-dichloroethylene | | | | 1.83 | 0.02 |
| $CHBr_3$ | 3.75 | | | | | vinyl bromide | | | | 1.98 | 0.23 |
| $CBrCl_3$ | 3.27 | | | | | allyl chloride | | | | 1.94 | −0.10 |
| $CH_2BrCH_2Br$ | 3.40 | 3.56 | | 3.35 | −0.21 | | | | | | |
| | | | | | | Amines and Nitriles | | | | | |
| ethylamine | 1.68 | | | | | diethylamine | 2.40 | | | | |
| *n*-propylamine | 2.14 | | | | | di-*n*-propylamine | 3.37 | | | | |
| *n*-butylamine | 2.62 | | | | | diisopropylamine | 2.89 | | | | |
| *tert*-butylamine | 2.49 | | | | | trimethylamine | 1.62 | | | | |
| *n*-pentylamine | 3.09 | | | | | triethylamine | 3.08 | 2.83 | | 2.69 | −0.14 |
| *n*-hexylamine | 3.56 | | | | | *N*,*N*-dimethylaniline | 4.75 | 4.66 | | 4.08 | −0.58 |
| methyl-*n*-propylamine | 2.49 | | | | | aniline | 3.99 | | | | |
| methylisopropylamine | 2.29 | | | | | acetonitrile | 1.56 | | | | |
| methyl-*n*-butylamine | 3.05 | | | | | propiononitrile | 1.94 | | | | |
| | | | | | | Pyridines | | | | | |
| pyridine | 3.00 | 3.20 | | 3.21 | 0.01 | 3-methylpyridine | 3.60 | 3.73 | | 3.67 | −0.06 |
| 2-methylpyridine | 2.44 | 3.54 | | 3.67 | 0.13 | 4-methylpyridine | 3.59 | 3.75 | | 3.67 | −0.08 |
| | | | | | | Nitro Hydrocarbons | | | | | |
| nitromethane | 1.89 | 2.44 | | 2.20 | −0.24 | 2-nitropropane | 2.55 | | | 3.00 | 0.19 |
| nitroethane | 2.37 | 2.75 | | 2.68 | −0.07 | nitrobenzene | 4.46 | | | | |
| 1-nitropropane | 2.71 | | | 3.15 | 0.12 | | | | | | |
| | | | | | | Multifunctional Compounds | | | | | |
| $CF_3CH_2OH$ | 1.22 | | | | | *m*-dimethoxybenzene | 5.02 | | | | |
| $(CF_3)_2CHOH$ | 1.39 | | | | | *p*-dimethoxybenzene | 5.04 | | | | |
| 3-fluorophenol | 3.84 | | | | | ethyl chloroacetate | 2.56 | | | | |
| 2-nitrophenol | 4.68 | | | | | isoflurane | | | | 1.91 | 0.01 |
| 2,6-difluorophenol | 3.69 | | | | | 1-methoxy-2-propanol | | | | 2.89 | 0.05 |
| methoxyflurane | 2.86 | 2.93 | | 2.59 | −0.34 | 2-methoxyethanol | | | | 2.65 | −0.07 |
| isoflurane | 1.58 | 1.98 | | 1.91 | −0.07 | 2-ethoxyethanol | | | | 3.12 | 0.14 |
| enflurane | 1.65 | 2.02 | | 1.91 | −0.11 | 2-isopropoxyethanol | | | | 3.45 | 0.24 |
| fluroxene | | 1.68 | | 1.61 | −0.07 | 2-ethoxyethyl acetate | | | | 3.53 | −0.16 |
| *o*-dimethoxybenzene | 4.97 | | | | | 2-butoxyethanol | | | | 3.92 | 0.18 |

[a] Indicates that the chemical is used to derive the correlation equation for aliphatic hydrocarbons or that the log $L_{oil}$ value is predicted from the correlation equation for aliphatic hydrocarbons. [b] Indicates that the chemical is used to derive the correlation equation for aromatic hydrocarbons or that the log $L_{oil}$ value is predicted from the correlation equation for aromatic hydrocarbons. [c] Indicates that the chemical is used to derive the correlation equation for alcohols or that the log $L_{oil}$ value is predicted from the correlation equation for alcohols. [d] Indicates that the chemical is used to derive the correlation equation for ketones or that the log $L_{oil}$ value is predicted from the correlation equation for ketones. [e] Indicates that the chemical is used to derive the correlation equation for esters or that the log $L_{oil}$ value is predicted from the correlation equation for esters. [f] Indicates that the chemical is used to derive the correlation equation for chlorinated aliphatic hydrocarbons or that the log $L_{oil}$ value is predicted from the correlation equation for chlorinated aliphatic hydrocarbons. [g] Log $L_{hex}^{e}$ and log $L_{oil}^{e}$ are the experimental values. log $L_{oil}^{p}$ is the value predicted from correlation eqs 3, 5, 6, or 7. log $L_{oil}^{c}$ is the value calculated from eq 10. $\Delta$log $L_{oil}$ is the estimation error of eq 10 (log $L_{oil}^{c}$ − log $L_{oil}^{e}$ or log $L_{oil}^{c}$ − log $L_{oil}^{p}$).

corresponding correction factors for all of the compounds of the database were then submitted to a multivariate linear regression analysis. An equation relating the log $L_{oil}$ values to the basic group parameters and correction factors was

**Table 2.** The Basic Group Sets and the Correction Factors[a]

| | parameter | | parameter |
|---|---|---|---|
| 1 | $-CH_3$ groups or $CH_4$ | 48 | acyclic-NH |
| 2 | acyclic $-CH_2-$ groups | 49 | acyclic $-N<$ |
| 3 | acyclic $-CH<$ groups | 50 | cyclic $-NH-$ |
| 4 | acyclic $>C<$ groups | 51 | cyclic $-N<$ |
| 5 | $=CH_2$ groups, excluding CO and CS | 52 | c-C≡N |
| 6 | acyclic $=CH-$ groups, excluding CO and CS | 53 | C≡N excluding c-C≡N |
| 7 | acyclic $=C<$ groups, excluding CO and CS | 54 | acyclic $=N-$ and $=NH$, excluding amides and S=C=N− |
| 8 | acyclic $=C=$ groups | 55 | cyclic $=N-$, excluding amides |
| 9 | HC≡C− groups or HC≡CH | 56 | $NO_2$-c |
| 10 | $-C\equiv$ groups excluding $-C\equiv N$ and $-C\equiv CH$ | 57 | $NO_2$, excluding $NO_2$-c |
| 11 | cyclic $-CH_2-$ groups | 58 | $-SH$ |
| 12 | cyclic $-CH<$ groups | 59 | acyclic $-S-$ |
| 13 | cyclic $>C<$ groups | 60 | cyclic $-S-$ |
| 14 | cyclic $=CH-$ groups, excluding CO and CS | 61 | $-CS$ |
| 15 | cyclic $=C<$ groups, excluding CO and CS | 62 | acyclic $>SO_2$ |
| 16 | F-c | 63 | cyclic $>SO_2$ |
| 17 | F atoms excluding F-c | 64 | $-P<(sp^3)$, $-P=(sp^2)$ with no O atoms attached, and P(+5) w/one double bond |
| 18 | Cl-c | 65 | ncOH or N=COH |
| 19 | Cl atoms excluding Cl-c | 66 | nc-COOH and N=CCOOH |
| 20 | Br-c | 67 | $-NH-N=CHX$ and n(H)nc(H)X, where X<> ($-N=$, n) |
| 21 | Br atoms excluding Br-c | 68 | $-CH(=O)X$, where <> C |
| 22 | I-c | 69 | NC(=O)N, where N is any N except $NH_3$ |
| 23 | I atoms excluding I-c | 70 | OH(0,1)C(=O)N, where N is any N except $NH_3$ |
| 24 | $OHCH_2R$, where R = any group or atom including H | 71 | C(=O)NH(0,1)C(=O) |
| 25 | $OHCHR_2$, where R<>H | 72 | CH(0−3)NHCH(0−3) |
| 26 | $OHCR_3$, where R<>H | 73 | CH(0−3)OCH(0−3) |
| 27 | OH-c | 74 | N=C(NH(1,2))N= and nc(NH(1,2))n |
| 28 | all OH excluding OHC($sp^3$) and OH-c | 75 | OHC=CCOOH, where only C=C are in a ring and OH-cc-COOH |
| 29 | cyclic −CO, excluding esters | 76 | OHC=CCH(1,0)(=O), where only C=C are in a ring and OHccCH(1,2)(=O) excluding COOH |
| 30 | acyclic $-CO-$, excluding esters | 77 | CH(1,2)-NH(1,2)-COOH |
| 31 | CO in aldehydes | 78 | NH(0,1)N=NN= and NH(0,1)N=NCH(0,1)= and NH(0,1)N=CH(0,1)N= and nnnn- and cH(0,1)nnn- and ncH(0,1)nn- and nnnnH and cH(0,1)nnnH and ncH(0,1)nnH |
| 32 | c-COOH | 79 | $CON(NH_2)C=N-$ |
| 33 | COOH, excluding c-COOH | 80 | $OCH_2COOH$ |
| 34 | acyclic CO in esters | 81 | $NH(1,2)CH_2CH(1,2)OH$ |
| 35 | cyclic CO in esters | 82 | NH(0−2)CONNO and $NH(0-2)CONNO_2$ |
| 36 | CO in $-CONH_2$ | 83 | ncH(0,1)cH(0,1)cOH, where n<>nH or n- |
| 37 | CO in $-CONH-$ | 84 | $NO_2$-ccH(0,1)cH(0,1)c-$NH_2$ and $NO_2$-ccH(0,1)c-$NH_2$ |
| 38 | CO in $-CON<$ | 85 | $NO_2$-ccH(0,1)cH(0,1)c-OH and $NO_2$-ccH(0,1)c-OH |
| 39 | CO in $-CON=$ | 86 | $NH_2$-ccH(0,1)cH(0,1)c-COOH |
| 40 | acyclic CO in ketones | 87 | NH(1,2)-ccH(0,1)cH(0,1)c-$SO_2$NH(1,2) |
| 41 | cyclic CO in ketones | 88 | CONH-ccH(0,1)cH(0,1)c-OH |
| 42 | $-NO$ | 89 | 1-*n*-hexyl group not in saturated hydrocarbon compound |
| 43 | $-SO$ | 90 | (0/1) $OH(CH_2)$n-pyridine, $n > 2$ |
| 44 | $NH_2CH_2R$, where R = any group or atom including H | 91 | (0/1) $NH_2(CH_2)$n-pyridine, $n > 2$ |
| 45 | $NH_2CHR_2$ and $NH_2CR_3$, where R<>H | 92 | (0/1) $NH_2CO(CH_2)n$-pyridine, $n > 1$ |
| 46 | $NH_2$-c | 93 | total no. of atoms in acyclic saturated and unicyclic saturated hydrocarbons |
| 47 | $NH_2$ excluding $NH_2C(sp^3)$ and $NH_2$-c | 94 | (0/1) all polycyclic, unsaturated, or aromatic hydrocarbons |

[a] Lowercase (aromatic ring atoms). n-(aromatic atom with nonmoveable bond system, e.g., *N*-methylpyrrole).

obtained in the form of eq 1

$$\log L = \alpha + \sum b_i B_i \qquad (1)$$

where $\alpha$ is a constant, $b_i$ is the contribution coefficient of the *i*th basic group or correction factor, and $B_i$ is the number of occurrences of the *i*th basic group or correction factor identified by the CASE program. The statistical significance of each parameter and the resulting equation were evaluated by consideration of the respective *t*- and *F*-values.

## RESULTS AND DISCUSSION

**Statistical Analysis.** The possibility of estimating $\log L_{oil}$ values of chemicals from the corresponding experimental known $\log L_{hex}$ values was tested by linear regression analyses. $\log L_{oil}$ values were modeled in terms of $\log L_{hex}$ values. Figure 1 is the scatter graph between $\log L_{oil}$ and $\log L_{hex}$ of the 101 compounds for which both values were known. Given possible variability in olive oil composition, the results are remarkably good.

The line in Figure 1 resulted from the following linear regression equation (eq 2):

$$\log L_{oil} = 0.30975 + 0.86186 * \log L_{hex} \qquad (2)$$

$$r^2 = 0.9385, \quad n = 101, \quad SD = 0.3014$$

where $r^2$ is the square of the correlation coefficient and Sd is the root-mean-square error of the fit.

Even though the correlation between olive oil−gas and hexadecane−gas partition coefficients was good, the scatter pattern based on all values have two boundaries. Examination of the chemicals in the boundaries revealed that most of the chemicals in the upper boundary are alcohols, whereas the chemicals in the lower line boundary are exclusively aliphatic hydrocarbons. The use of a single linear equation to represent the relationship for all classes of chemicals is

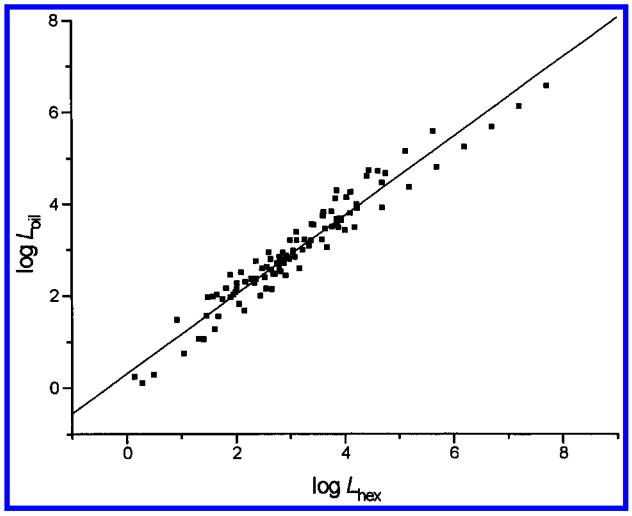


**Figure 1.** Relationship between log $L_{oil}$ and log $L_{hex}$ of all values.

not satisfactory and could result in substantial deviations. Therefore we divided the 101 chemicals into several classes based on their structure characteristics: aliphatic hydrocarbons[a], aromatic hydrocarbons[b], alcohols[c], ketones[d], esters[e], chlorinated aliphatic hydrocarbons[f] (the chemicals labeled as $^{a-f}$ in Table 1), and others. For each of the groups, a correlation was established between their olive oil−gas and hexadecane−gas partition coefficients. The resulting correlation equations are

aliphatic hydrocarbons:

$$\log L_{oil} = -0.1525 + 0.87051 * \log L_{hex} \quad (3)$$

$$r^2 = 0.9999, \quad n = 19, \quad SD = 0.0305$$

aromatic hydrocarbons:

$$\log L_{oil} = -0.19927 + 0.98104 * \log L_{hex} \quad (4)$$

$$r^2 = 0.992, \quad n = 11, \quad SD = 0.0454$$

alcohols:

$$\log L_{oil} = 0.60402 + 0.88695 * \log L_{hex} \quad (5)$$

$$r^2 = 0.9988, \quad n = 12, \quad SD = 0.0510$$

ketones:

$$\log L_{oil} = 0.18045 + 0.9413 * \log L_{hex} \quad (6)$$

$$r^2 = 0.9771, \quad n = 8, \quad SD = 0.0953$$

esters:

$$\log L_{oil} = 0.41181 + 0.82485 * \log L_{hex} \quad (7)$$

$$r^2 = 0.9819, \quad n = 11, \quad SD = 0.0924$$

chlorinated hydrocarbons:

$$\log L_{oil} = 0.04116 + 0.95224 * \log L_{hex} \quad (8)$$

$$r^2 = 0.9033, \quad n = 14, \quad SD = 0.1953$$

Except for the chlorinated aliphatic hydrocarbons, correlations between the two partition coefficients are excellent. Based on these correlations we predicted the log $L_{oil}$ of 40 compounds for which this value was not known and proceeded then to evaluate the possibility of creating a log $L_{oil}$ model based on group contributions. The estimated log $L_{oil}$ values of these 40 chemicals are listed in column 4 of Table 1, along with the experimental values for the other 119 chemicals.

**Group Contribution Approach.** The occurrence of each group parameter within the olive oil−gas learning database (119 experimental + 40 estimated log $L_{oil}^{p}$ values) was determined. Thirty-six of the 94 group parameters selected as potential predictors of log $L_{oil}$ had nonzero occurrences. Among them, 13 parameters occurred in only one, two, or three compounds. The derived regression coefficients for these 13 parameters are not considered reliable due to the low number of occurrences within the data base. Therefore, they were removed in order to obtain a more general correlation. Regression analysis indicated that the remaining parameters were significant enough to be included in the correlation. The following equation was found, correlating these parameters with log $L_{oil}$ value (eq 9).

$$\begin{aligned}\log L_{oil} = {} & 0.9608 - 0.0001L_1 + 0.4706L_2 + 0.7265L_3 + \\ & 0.9577L_4 - 0.0101L_5 + 0.4368L_6 + 0.3896L_{11} + \\ & 0.4284L_{14} + 0.9096L_{15} - 0.4270L_{17} + 0.3436L_{19} + \\ & 0.7891L_{21} + 0.5198L_{24} + 0.5431L_{25} + 0.0911L_{30} + \\ & 0.8844L_{34} + 0.9121L_{40} + 0.0914L_{55} - 0.3485L_{68} + \\ & 0.1128L_{73} - 0.0883L_{89} - 0.0249L_{93} - 0.9110L_{94} \quad (9)\end{aligned}$$

$$n = 159, \quad r^2 = 0.9377, \quad SD = 0.2954$$

A testing set of 36 compounds with experimentally known log $L_{oil}$ values was obtained from the literature in order to evaluate the prediction performance of eq 9.[1,2,11,12] These values had not been used in the derivation of eq 9 except for three of them, hexachloroethane, methylcyclohexane, and 2,2,4-trimethylpentane.

The experimental and predicted log $L_{oil}$ values based on eq 9 as well as the estimation errors of the 36 compounds are listed in Table 3. There were only three compounds showing deviations greater than 0.6, two of them containing the undetermined parameter $-NO_2$.

This test set was subsequently added to the learning set so that a new model could be derived. This model, using the additional two parameters 7 and 57 gave a correlation coefficient of $r^2$ of 0.9609 and standard deviation of 0.2267. Compared with the first model, the new model contains two additional parameters, but the standard deviation has been reduced by 0.07 log unit, and the $r^2$ improved by 0.02.

By studying the log $L_{oil}$ values, we found that the molecular weight may be a contributing factor for the determination of log $L_{oil}$. Within each class of compounds, usually an increase in molecular weight results in a corresponding increase in log $L_{oil}$. Therefore we modified our model to include the molecular weight and found a new estimation model for log $L_{oil}$ (eq 10) with $r^2 = 0.9678$ and SD = 0.2056. The frequency, standard error, and *t*-value of each parameter are listed in Table 4 and for the corresponding regression coefficient see eq 10. Both the $r^2$ and the standard deviation of the log $L_{oil}$ were thus found to improve. The comparison of the distribution of estimation error with and without molecular weight as a parameter are shown in Figure 2. The number of compounds with a large deviation

**Table 3.** Experimental, Predicted Log $L_{oil}$, and the Estimation Error of the Test Set[a]

| | name | log $L_{oil}^e$ | log $L_{oil}^p$ | $\Delta$log $L_{oil}$ | |
|---|---|---|---|---|---|
| 1 | methylcyclohexane | 2.82 | | | 2.68[b] |
| 2 | cyclohexene | 2.66 | 2.47 | −0.19 | |
| 3 | methyl chloride ($CH_3Cl$) | 0.93 | 1.30 | 0.37 | |
| 4 | difluoromethane | 0.68 | 0.58 | −0.10 | |
| 5 | fluorochloromethane | 1.35 | 1.35 | 0.00 | |
| 6 | bromochloromethane | 2.56 | 2.56 | 0.00 | |
| 7 | dibromomethane | 2.98 | 3.01 | 0.03 | |
| 8 | chlorodibromomethane | 3.43 | 3.61 | 0.18 | |
| 9 | pentachloroethane | 3.83 | 4.36 | 0.53 | |
| 10 | hexachloroethane | 3.70 | 4.94 | 1.24 | |
| 11 | 1-bromo-2-chloroethane | 2.76 | 3.04 | 0.28 | |
| 12 | 2-chloropropane | 1.84 | 2.03 | 0.19 | |
| 13 | *n*-propyl bromide | 2.44 | 2.69 | 0.25 | |
| 14 | isopropyl bromide | 2.22 | 2.48 | 0.26 | |
| 15 | 1-nitropropane | 3.03 | 1.90 | −1.13 | |
| 16 | 2-nitropropane | 2.81 | 1.69 | −1.12 | |
| 17 | 2,3,4-trimethylpentane | 2.82 | 2.50 | −0.32 | |
| 18 | 2,2,4-trimethylpentane | 2.56 | | | 2.56[b] |
| 19 | vinyl chloride | 1.39 | 1.73 | 0.34 | |
| 20 | 1,1-dichloroethylene | 1.81 | 1.64 | −0.17 | |
| 21 | vinyl bromide | 1.75 | 2.18 | 0.43 | |
| 22 | *m*-methylstyrene | 4.17 | 4.01 | −0.16 | |
| 23 | *p*-methylstyrene | 4.14 | 4.01 | −0.13 | |
| 24 | isoflurane | 1.90 | 1.78 | −0.12 | |
| 25 | allyl chloride | 2.04 | 2.20 | 0.16 | |
| 26 | isoprene | 0.95 | 0.47 | −0.48 | |
| 27 | 1-methoxy-2-propanol | 2.84 | 2.91 | 0.07 | |
| 28 | 2-methoxyethanol | 2.72 | 2.63 | −0.09 | |
| 29 | 2-ethoxyethanol | 2.98 | 3.10 | 0.12 | |
| 30 | 2-isopropoxyethanol | 3.21 | 3.35 | 0.14 | |
| 31 | 2-ethoxyethyl acetate | 3.69 | 3.46 | −0.23 | |
| 32 | 2-butoxyethanol | 3.74 | 3.82 | 0.09 | |
| 33 | isobutyl acetate | 3.11 | 3.04 | −0.07 | |
| 34 | isopentyl acetate | 3.47 | 3.51 | 0.04 | |
| 35 | isobutyl alcohol | 2.67 | | | 2.73[b] |
| 36 | isopentyl alcohol | 3.00 | 3.17 | 0.17 | |
| | | av | | 0.271 | |

[a] Log $L_{oil}^e$ is the experimental value. log $L_{oil}^p$ is the predicated value based on eq 9. $\Delta$log $L_{oil}$ is the estimation error. [b] Represent the log $L_{oil}$ values estimated from the correlation eqs 3−7.

decreased significantly in the model that includes the molecular weight.

$$\log L_{oil} = 0.2541 + 0.0057L_1 + 0.2399L_2 + 0.3233L_3 + 0.2153L_4 - 0.0678L_5 + 0.2688L_6 + 0.0847L_7 + 0.2357L_{11} + 0.2598L_{14} + 0.4826L_{15} - 0.4539L_{17} - 0.0328L_{19} - 0.2673L_{21} + 0.6331L_{24} + 0.5531L_{25} - 0.0456L_{30} + 0.3261L_{34} + 0.6163L_{40} + 0.3297L_{55} + 0.9196L_{57} - 0.1014L_{68} + 0.0478L_{73} - 0.0678L_{89} - 0.0198L_{93} - 0.4828L_{94} + 0.1675L_{95} \quad (10)$$

* parameter 95 is the molecular weight

$$n = 192, \quad r^2 = 0.9678, \quad \mathrm{SD} = 0.2056$$

The newly calculated log $L_{oil}$ values of the 192 compounds as well as the estimation errors are also summarized in Table 1. The olive oil−gas learning set can be divided into 10 classes based on their chemical structures. The estimation error for each class of compounds are listed in Table 5. As can be seen, the final log $L_{oil}$ model results in reliable values for every class of compound.

In order to further test the reliability of the final log $L_{oil}$ model, cross-validation tests were performed. To do so, about 10% of the compounds of the original database were randomly selected as a test set, while the remaining 90% were used as a learning database. A model was developed based on the learning set using only those parameters predicted from the reduced set of compounds in the learning

**Table 4.** Number of Compounds and Occurrences of Each Fragment in the Learning Set, Standard Error and *t*-Value[a]

| parameter | no. of compds | freq of occurrence | standard error | *t*-value |
|---|---|---|---|---|
| 1 | 133 | 260 | 0.0497 | 0.1145 |
| 2 | 112 | 352 | 0.0429 | 5.5848 |
| 3 | 49 | 59 | 0.1059 | 3.0542 |
| 4 | 21 | 22 | 0.1761 | 1.2222 |
| 5 | 13 | 16 | 0.0905 | −0.7494 |
| 6 | 14 | 16 | 0.0796 | 3.3756 |
| 7 | 5 | 6 | 0.1347 | 0.6289 |
| 11 | 13 | 64 | 0.0370 | 6.3727 |
| 14 | 25 | 113 | 0.0401 | 6.4828 |
| 15 | 22 | 29 | 0.1051 | 4.5935 |
| 17 | 17 | 45 | 0.0542 | −8.3698 |
| 19 | 36 | 72 | 0.0714 | −0.4595 |
| 21 | 11 | 14 | 0.1862 | −1.4354 |
| 24 | 17 | 17 | 0.0837 | 7.5606 |
| 25 | 9 | 9 | 0.1050 | 5.2659 |
| 30 | 17 | 19 | 0.1063 | −0.4294 |
| 34 | 21 | 21 | 0.1353 | 2.4109 |
| 40 | 14 | 14 | 0.1060 | 5.8134 |
| 55 | 4 | 4 | 0.1344 | 2.4538 |
| 57 | 4 | 4 | 0.1445 | 6.366 |
| 68 | 4 | 4 | 0.1395 | −0.7268 |
| 73 | 17 | 20 | 0.0948 | 0.5043 |
| 89 | 10 | 10 | 0.0976 | −0.6946 |
| 93 | 38 | 908 | 0.0038 | −5.1874 |
| 94 | 17 | 17 | 0.1083 | −4.4591 |
| 95 | 192 | | 0.0281 | 5.9663 |

[a] Contribution values are listed in eq 10.

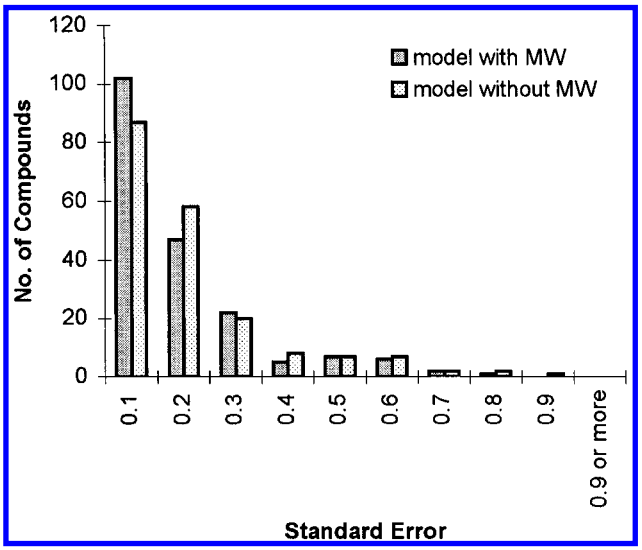


**Figure 2.** Distribution of estimation error of log $L_{oil}$.

**Table 5.** Estimation Result for a Different Class of Compounds of the Olive Oil−Gas Database

| | no. of compds | std dev |
|---|---|---|
| aliphatic hydrocarbons | 42 | 0.14 |
| aromatic hydrocarbons | 13 | 0.11 |
| alcohols ethers, and phenols | 31 | 0.22 |
| aldehydes and ketones | 17 | 0.24 |
| esters | 20 | 0.08 |
| halogenated hydrocarbons | 48 | 0.26 |
| amines and nitriles | 2 | unavailable |
| pyridines | 4 | 0.10 |
| nitro hydrocarbons | 4 | 0.13 |
| multifunctional compounds | 11 | 0.15 |
| all | 192 | 0.16 |

**Table 6.** Cross-Validation Test Results for the 192 Compounds of Olive Oil−Gas Database

| | learning set | | | test set | | |
|---|---|---|---|---|---|---|
| | no. of compds | $r^2$ | SD[a] | no. of compds | $r^2$ | SD[a] |
| 1 | 173 | 0.963 | 0.214 | 19 | 0.988 | 0.163 |
| 2 | 173 | 0.970 | 0.205 | 19 | 0.900 | 0.241 |
| 3 | 173 | 0.969 | 0.207 | 19 | 0.941 | 0.222 |
| 4 | 173 | 0.964 | 0.209 | 19 | 0.985 | 0.179 |
| 5 | 173 | 0.967 | 0.207 | 19 | 0.971 | 0.224 |
| 6 | 173 | 0.971 | 0.197 | 19 | 0.907 | 0.366 |
| 7 | 173 | 0.969 | 0.203 | 19 | 0.954 | 0.231 |
| 8 | 173 | 0.970 | 0.199 | 19 | 0.925 | 0.300 |
| 9 | 173 | 0.970 | 0.200 | 19 | 0.933 | 0.284 |
| 10 | 171 | 0.972 | 0.196 | 21 | 0.914 | 0.276 |
| mean | | 0.969 | 0.204 | | 0.942 | 0.249 |

[a] SD represents standard deviation.

set, and predictions were made for the molecules of the test set. The $r^2$ and standard deviation are calculated for both the learning and test sets. The results for ten such cross-validation experiments for the olive oil−gas database are shown in Table 6.

The $r^2$ and the standard deviation values remain almost the same for all the learning sets. As is often the case, they are not as stable for the test sets. The results for some of the tests sets are extremely good, even better than their corresponding learning sets, while the results for a few are sometimes just acceptable. While it is to be expected that the results obtained for data not used in the correlation will be somewhat inferior to those obtained for retrofitting the data, the distance between the mean values of the correlation coefficients indicate the existence of gaps in the knowledge generated by individual learning sets. The main factor responsible for this condition is that the size of the database is too small to be representative of every class of organic chemical. Additionally there are some compounds which contain undefined parameters, e.g., in the initial learning set of log $L_{oil}$, only nitromethane and nitroethane contain the nitro group. The resulting model could therefore not properly represent molecules containing a nitro group. By adding two more nitro hydrocarbons to the updated database, the estimation errors of the two compounds decreased significantly from −1.48 to −0.24 or −0.25 and from −1.32 to −0.07, respectively.

Overall, we find that the correlation equations for different classes of compounds obtained from linear regression analyses gave excellent log $L_{oil}$ predictions when the log $L_{hex}$ values of the corresponding classes of compounds are available. Our strategy of incorporating chemical groups and correction factors as previously found useful for the log $P$ model was found to be applicable to the determination of log $L_{oil}$ values as well. The log $L_{oil}$ models established from the group contribution approach can provide accurate estimates of log $L_{oil}$ predictions for molecules whose structure class is represented in the learning sets. Our model is however limited in that it will produce large deviations when attempting to predict log $L_{oil}$ values for molecules containing still undefined parameters. This can be overcome in the future by updating the database as new data becomes available. Another shortcoming of our methodology is that it does not distinguish between subtle structural differences and predicts identical log $L_{oil}$ value for geometrical isomers as well as some structural isomers.

## CONCLUSION

The olive oil−gas partition coefficients correlate well with the hexadecane-gas partition coefficients. The olive oil−gas partition coefficients, not easily obtainable experimentally, can be predicted accurately from a set of equations linking log $L_{oil}$ and log $L_{hex}$. With a regression analysis using molecular weight and the 94 parameters obtained from our previously published log $P$ estimation model, we have successfully developed reliable log $L$ estimation models (olive oil−gas: $n = 192$, $r^2 = 0.9678$, SD = 0.2056). Both cross-validation and the testing of an independent set of 33 compounds have proved that our methodology can give fast and accurate log $L_{oil}$ values for compounds represented within the model. Inclusion of molecular weight significantly increases the prediction capability of the models. Further refinements to our models will be forthcoming as additional data become available.

## ACKNOWLEDGMENT

This investigation was made under contract with Concurrent Technologies Corporation/National Defense Center for Environmental Excellence in support of the U.S. Department of Defense (Contract No. DAAA21-93-C-0046). The CASE program and the ToxAlert program (that contain log $L$ estimate calculations) can be obtained from MULTICASE Inc., P.O. Box 22517, Cleveland, OH 44122.

## REFERENCES AND NOTES

(1) Olivier-Droz, P.; Fernandez, J. Solubility of Organic Solvents. I. Gas Chromatographic Determination of Olive Oil−Gas Partition Coefficients. *Helv. Chim. Acta* **1977**, *60*, 454−458.

(2) Johanson, G.; Dynesius, B. Liquid/Air Partition Coefficients of Six Commonly Used Glycol Ethers. *Br. J. Ind. Med.* **1988**, *45*, 561−564.

(3) Sato, A.; Nakajima, T. Partition Coefficients of Some Aromatic Hydrocarbons and Ketones in Water, Blood and Oil. *Br. J. Ind. Med.* **1979**, *36*, 231−234.

(4) Abraham, M. H.; Grellier, P. L.; McGill, R. A. Determination of Olive Oil−Gas Partition Coefficients, and Calculation of the Corresponding Olive Oil−Water and Hexadecane-Water Partition Coefficients. *J. Chem. Soc., Perkin Trans. II* **1987**, 797−803.

(5) Klopman, G.; Wang, S.; Balthasar, D. M. Estimation of Aqueous Solubility of Organic Molecular by the Group Contribution Approach. Application to the Study of Biodegradation. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 475−482.

(6) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M. Computer Automated Log P Calculation Based on an Extended Group Contribution Approach. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752−781.

(7) Wang, S.; Milne, G. W. A.; Klopman, G. Graph Theory and Group Contribution in the Estimation of Boiling Points. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1242−1250.

(8) Klopman, G. Artificial Intelligence Approach to Structure-Activity Studies. Computer Automated Structure Evaluation of Biological Activity of Organic Molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.

(9) Klopman, G. MULTICASE 1. A Hierarchical Computer Automated Structure Evaluation Program. *Quant. Struct.-Act. Relat.* **1992**, *11*, 176−184.

(10) Klopman, G.; McGonigal, M. Computer Simulation of Physical-Chemical Properties of Organic Molecules. 1. Molecular System Identification. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 48−52.

(11) Gargas, M. L.; Burgess, R. J.; Voisard, D. E.; Cason, G. H.; Andersen, M. E. Partition Coefficient of Low-Molecular-Weight Volatile Chemicals in Various Liquids and Tissues. *Toxicol. Appl. Pharmacol.* **1989**, *98*, 87−99.

(12) Kaneko, T.; Wang, P.-Y.; Sato, A. Partition Coefficients of Some Acetate Esters and Alcohols in Water, Blood, Olive oil, and Rat Tissues. *Occup. Environ. Med.* **1994**, *51*, 68−72.

CI960140L