# molecules 

ISSN 1420-3049
http://www.mdpi.org

# Chemical Property Calculation through JavaScript and Applications in QSAR* 

Hanqing Wu<br>Department of Chemistry, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA<br>E-mail hanqing@csd.uwm.edu

* This paper was presented at ecsoc-1 at URL of http://www.unibas.ch/mdpi/ecsoc/e0002/e0002.htm Part of this work was performed in the Department of Applied Chemistry, University of Science and Technology of China, Hefei, Anhui, P. R. China (1984-90)

Editor's note: http://www.unibas.ch/mdpi/ecsoc/e0002/e0002.htm file will be included in the CDROM edition.

Received: 30 January 1998 / Accepted: 1 July 1998 / Published: 8 February 1999


#### Abstract

The inorganic property (I) and organic property ( $\mathbf{O}$ ) values of general organic groups are re-proposed here. Both $\mathbf{I}$ and $\mathbf{O}$ values of drug and biological molecules or groups can be calculated based on their common group values. The calculation can be performed easily on-line through JavaScript. Similar calculation can be done for the drug and biological molecular group electronegativity ( $\mathbf{X}$ ) according to the author's published paper. The calculation of lipophilicity ( $\pi$ or $\log \mathrm{P}$ ) parameter of (macro)molecules (like proteins) can also be performed on-line through JavaScript. Two equations expressed with $\mathbf{I}$ and $\mathbf{O}$ are provided here to define the hydrophobicity of each amino acid. The correlations of inorganic property and organic property values with other parameters are also discussed. These calculated parameters combined with other parameters can be used for QSAR studies in some drug molecules.


Keywords: Inorganic property and organic property, $\log \mathrm{P}$, bio-active molecules, JavaScript, QSAR, drug design.

## Introduction

Quantitative structure-activity relationship (QSAR) correlation has been widely applied in biological activities over several decades. Many new descriptors (parameters) have been developed © 1999 MDPI. All rights reserved.
[1-5]. Six main types of molecular descriptors were introduced: constitutional, geometrical, topological, electrostatic, quantum-chemical, and thermodynamic descriptors, the calculations of these descriptors using different packages were introduced by Katritzky and coworkers [1]. Inorganic property and organic property values [6,7], and group electronegativity [8] were also calculated solely on the basis of intrinsic structural information of the molecular species under consideration. Obviously the group electronegativity belongs to the electrostatic descriptors, and reflects the characteristics of the partial charge of the group. According to the author's previous work [8], group electronegativity is highly correlated with proton chemical shift in $\mathrm{R}-\mathrm{H}$ molecules ( R is the group). The calculation of inorganic property and organic property values of organic molecules as well as group electronegativity were calculated through JavaScript [9], which is a powerful language used in the Internet [10]. Other parameters (like $\pi$ or $\log P$ value) can also be calculated through JavaScript. The correlation of inorganic property and organic property values of organic molecules or groups with other parameters are analyzed here. The usage of these descriptors in QSAR studies of some drugs is also discussed.

## Calculation

Before introducing the calculation of inorganic property and organic property values and group's electronegativity using JavaScript, the author first gives a simple example: It is common for chemists to calculate the molecular weight of organic or biological molecules (amino acids or nucleic acids). The common atoms in organic or biological molecules are carbon $(\mathrm{C})$, hydrogen $(\mathrm{H})$, oxygen $(\mathrm{O})$, nitrogen $(\mathrm{N})$, sulfur ( S ), phosphorus ( P ), and others (like chloride ( Cl ), fluoride ( F ), bromide $(\mathrm{Br})$, water ( W or w) etc.). To calculate the molecular weight of organic or biological molecules using JavaScript, the users only need to input the type of atoms (either lower case or upper case for the first symbol of atoms, such as: for chloride, input " Cl " or "cl"; for carbon, input " C " or " "c") and the numbers of the atoms. There are a total of six rows to input a maximum of six types of atoms in one calculation, which are sufficient for common organic or biological molecules. The molecular weight of some metal organic molecules can also be calculated (for iron, input " Fe " or "fe" in one of the six rows; for cobalt, input "Co" or "co"). The molecular weight calculation is shown at the URL of http://www.unibas.ch/mdpi/ecsoc/e0002/mwcalcf.htm. The calculation for $\log \mathrm{P}$ is similar to that for molecular weight, the symbols of groups can be obtained from the references [11,12], and the appendix or the source codes. Modification of the symbols can be done by the readers at their own convenience. For large and complicated groups of organic and drug molecules, the structure of the groups can be shown on screen, and the numbers of groups are needed to input from the user to calculate the organic or inorganic property values of the groups. Both the numbers of sub-groups or atoms and the values of their electronegativity are needed to calculate the electronegativity of large groups. Other parameters can also be calculated using JavaScript, in a similar way to the calculation of log. .

| Molecular Weight Calculating Spreadsheet |
| :---: |
| http://www.unibas.ch/mdpi/ecsoc/e0002/mwcalcf.htm |
| logP Calculating using JavaScript |
| http://www.unibas.ch/mdpi/ecsoc/e0002/logpcalc.htm |

# Inorganic property and organic property values of groups. <br> Calculation of Inorganic property value of molecules or groups using JavaScript. Calculation of Organic property value of molecules or groups using JavaScript. <br> Http://www.unibas.ch/mdpi/ecsoc/e0002/calinorg.htm 

## Calculation of Group Electronegativity using JavaScript. <br> Http://www.unibas.ch/mdpi/ecsoc/e0002/calelecg.htm

## Results

The inorganic property and organic property values of common organic or biological molecule groups are listed in Table 1. The author calculated the inorganic and organic property values of 50 common groups, and analyzed their correlation with other parameters, lipophilicity ( $\pi$ ), polar constant $(\mathbf{F})$, molar refractivity $(\mathbf{M r})$, resonance constant $(\mathbf{R})$, Hammett meta constant $\left(\sigma_{\mathrm{m}}\right)$, and para constant $\left(\sigma_{\mathrm{p}}\right)$. All the parameters can be seen in Table 2, and the correlations among them can be seen in Table 3. A plot of inorganic property value versus organic property value of molecule groups is shown in Figure 1. It is clear that two sets of molecule groups can be divided: one set is the hydrophobic molecule groups with $\pi$ (pi) $>0.1$, the other set is the hydrophilic molecule groups with $\pi$ (pi) <0.1.

From Table 3, it is shown that there are high correlations among $\pi$ and $\mathbf{I}$ or $\mathbf{O}$ (see the equations below), and between $\mathbf{M r}$ and $\mathbf{O}(\mathrm{r}=0.837)$. It is also shown that $\mathbf{F}$ has high correlation with $\sigma_{\mathrm{m}}$ and $\sigma_{\mathrm{p}}$; $\sigma_{\mathrm{p}}$ has high correlation with $\mathbf{R}$ and $\sigma_{\mathrm{m}} ; \mathbf{X}$ has some correlation with $\mathbf{F}(\mathrm{r}=0.637) ; \mathbf{F}, \mathbf{R}, \sigma_{\mathrm{m}}, \sigma_{\mathrm{p}}$, and $\mathbf{X}$ have less correlation with $\mathbf{I}$ or $\mathbf{O}$ (see Table 3).


Figure 1. Organic property value versus inorganic property value of 50 common groups.

Table 1. Inorganic property and organic property values of some inorganic groups and organic / inorganic groups $\dagger$

| Inorganic group | Value | Organic / inorganic group | Organic <br> Value | Inorganic value |
| :---: | :---: | :---: | :---: | :---: |
| $-\mathrm{SO}_{2}-\mathrm{NH}-$ | 240 | $-\mathrm{OSO}_{3} \mathrm{H}$ | 20 | 220 |
| $=\mathrm{N}-\mathrm{OH}$ | 210 | $>\mathrm{SO}_{2}$ | 40 | 170 |
| -NHCONH- |  |  |  |  |
| -CO-NH-* | 200 | >SO | 40 | 140 |
| $-\mathrm{COOH}$ | 150 | -SCN | 90 | 80 |
| $\mathrm{R}-\mathrm{CH}\left(\mathrm{CH}_{2}\right)_{\mathrm{n}}-\mathrm{C}=\mathrm{OO}_{4}$ | 120 | -NCS | 90 | 75 |
| -COO- (in a ring above) |  |  |  |  |
| -CO-O-CO- | 110 | $-\mathrm{NO}_{2}$ | 70 | 70 |
| Anthracene or | 105 | -As< | 40 | 70 |
| Phenanthrene nuclei |  | -CN |  |  |
| -OH | 100 | -P< | 20 | 70 |
| >Hg (covalent) | 95 | -O- [ $\left.\mathrm{CH}_{2} \mathrm{CH}_{2} \mathrm{O}\right]-\mathrm{CH}_{2}-$ | 30 | 60 |
| -NH-NH- | 80 | -NO | 50 | 50 |
| -O-CO-O- |  |  |  |  |
| -N < (-NH2, -NHPh,-NPh $)$ ) | 70 | $-\mathrm{O}-\mathrm{NO}_{2}$ | 60 | 40 |
| >CO | 65 | -NC | 40 | 40 |
| -COOPh | 60 | $-\mathrm{P}=\mathrm{P}-$ | 30 | 30 |
| naphthalene nuclei |  | -NCO |  |  |
| > $\mathrm{C}=\mathrm{NH}$ | 50 | -ONO | 40 | 20 |
|  |  | -SH |  |  |
|  |  | -S- |  |  |
| -O-O- | 40 | -I | 80 | 10 |
| $-\mathrm{N}=\mathrm{N}-$ | 30 | -Br | 60 | 10 |
| -O- | 20 | =S | 50 | 10 |
| Benzene nuclei (single aromatic ring) | 15 | -Cl | 40 | 10 |
| Ring (non aromatic single ring) | 10 | -F | 5 | 5 |
| Triple bond | 3 | Iso-** | -10 | 0 |
| Double bond | 2 | Tert-** | -20 | 0 |

$\dagger$ Reproduced from [7].

* Applied to non-ring compound.
** Applied to end of the compound.

Table 2. Inorganic and organic property and other parameters of 50 common groups.

| Group | $\pi$ | Mr | F | R | $\sigma_{\text {m }}$ | $\sigma_{p}$ | I | 0 | X |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| F | 0.14 | 0.92 | 0.43 | -0.34 | 0.34 | 0.06 | 5 | 5 | 3.98 |
| Cl | 0.71 | 6.03 | 0.41 | -0.15 | 0.37 | 0.23 | 10 | 40 | 3.16 |
| Br | 0.86 | 8.88 | 0.44 | -0.17 | 0.39 | 0.23 | 10 | 50 | 2.96 |
| I | 1.12 | 13.94 | 0.4 | -0.19 | 0.35 | 0.18 | 10 | 60 | 2.66 |
| H | 0 | 1.03 | 0 | 0 | 0 | 0 | 0 | 0 | 2.20 |
| OH | -0.67 | 2.85 | 0.29 | -0.64 | 0.12 | -0.37 | 100 | 0 | 3.03 |
| SH | 0.39 | 9.22 | 0.28 | -0.11 | 0.25 | 0.15 | 20 | 40 | 2.45 |
| $\mathrm{NH}_{2}$ | -1.23 | 5.42 | 0.02 | -0.68 | -0.16 | -0.66 | 70 | 0 | 2.70 |
| NHOH | -1.34 | 7.22 | 0.06 | -0.4 | -0.04 | -0.34 | 170 | 0 | 2.87 |
| $\mathrm{NHNH}_{2}$ | -0.88 | 8.44 | 0.17 | -0.71 | -0.02 | -0.55 | 140 | 0 | 2.80 |
| $\mathrm{CF}_{3}$ | 0.88 | 5.02 | 0.38 | 0.19 | 0.43 | 0.54 | 15 | 35 | 3.16 |
| $\mathrm{OCF}_{3}$ | 1.04 | 7.86 | 0.38 | 0 | 0.38 | 0.35 | 35 | 35 | 3.35 |
| CN | -0.57 | 6.33 | 0.51 | 0.19 | 0.56 | 0.66 | 70 | 40 | 2.76 |
| NCS | 1.15 | 17.24 | 0.51 | -0.09 | 0.48 | 0.38 | 75 | 90 | 2.85 |
| SCN | 0.41 | 13.4 | 0.36 | 0.19 | 0.41 | 0.52 | 80 | 90 | 2.64 |
| NHCN | -0.26 | 10.14 | 0.26 | -0.18 | 0.21 | 0.06 | 140 | 40 | 2.82 |
| CHO | -0.65 | 6.88 | 0.31 | 0.13 | 0.35 | 0.42 | 65 | 20 | 2.75 |
| $\mathrm{CO}_{2} \mathrm{H}$ | -0.32 | 6.93 | 0.33 | 0.15 | 0.37 | 0.45 | 150 | 20 | 2.87 |
| $\mathrm{CH}_{2} \mathrm{Br}$ | 0.79 | 13.39 | 0.1 | 0.05 | 0.12 | 0.14 | 10 | 80 | 2.51 |
| $\mathrm{CH}_{2} \mathrm{Cl}$ | 0.17 | 10.49 | 0.1 | 0.03 | 0.11 | 0.12 | 10 | 60 | 2.54 |
| $\mathrm{CH}_{2} \mathrm{I}$ | 1.5 | 18.6 | 0.09 | 0.03 | 0.1 | 0.11 | 10 | 100 | 2.47 |
| NHCHO | -0.98 | 10.31 | 0.25 | -0.23 | 0.19 | 0 | 135 | 20 | 2.81 |
| $\mathrm{CONH}_{2}$ | -1.49 | 9.81 | 0.24 | 0.14 | 0.28 | 0.36 | 135 | 20 | 2.83 |
| $\mathrm{CH}=\mathrm{NOH}$ | -0.38 | 10.28 | 0.25 | -0.13 | 0.22 | 0.1 | 220 | 20 | 2.64 |
| $\mathrm{CH}_{3}$ | 0.56 | 5.65 | -0.04 | -0.13 | -0.07 | -0.17 | 0 | 20 | 2.40 |
| $\mathrm{NHCONH}_{2}$ | -1.3 | 13.72 | 0.04 | -0.28 | -0.03 | -0.24 | 105 | 20 | 2.83 |
| $\mathrm{OCH}_{3}$ | -0.02 | 7.87 | 0.26 | -0.51 | 0.12 | -0.27 | 20 | 20 | 3.09 |
| $\mathrm{CH}_{2} \mathrm{OH}$ | -1.03 | 7.19 | 0 | 0 | 0 | 0 | 100 | 20 | 2.52 |
| $\mathrm{NHCH}_{3}$ | -0.47 | 10.33 | -0.11 | -0.74 | -0.3 | -0.84 | 70 | 20 | 2.74 |
| - $\mathrm{C} \equiv \mathrm{CH}$ | 0.4 | 9.55 | 0.19 | 0.05 | 0.21 | 0.23 | 3 | 40 | 2.50 |
| $\mathrm{NHCOCF}_{3}$ | 0.08 | 14.3 | 0.36 | -0.21 | 0.3 | 0.12 | 150 | 40 | 2.84 |
| $\mathrm{CH}_{2} \mathrm{CN}$ | -0.57 | 10.11 | 0.21 | -0.18 | 0.16 | 0 | 70 | 60 | 2.48 |
| $-\mathrm{CH}=\mathrm{CH}_{2}$ | 0.82 | 10.99 | 0.07 | -0.08 | 0.05 | -0.02 | 2 | 40 | 2.46 |
| $\mathrm{COCH}_{3}$ | -0.55 | 11.18 | 0.32 | 0.2 | 0.38 | 0.5 | 65 | 40 | 2.78 |
| $\mathrm{OCOCH}_{3}$ | -0.64 | 12.47 | 0.41 | -0.07 | 0.39 | 0.31 | 85 | 40 | 3.22 |
| $\mathrm{CO}_{2} \mathrm{CH}_{3}$ | -0.01 | 12.87 | 0.33 | 0.15 | 0.37 | 0.45 | 85 | 40 | 2.88 |
| $\mathrm{NHCOCH}_{3}$ | -0.97 | 14.93 | 0.28 | -0.26 | 0.21 | 0 | 135 | 40 | 2.82 |
| $\mathrm{NHCO}_{2} \mathrm{CH}_{3}$ | -0.37 | 16.53 | 0.14 | -0.28 | 0.07 | -0.15 | 155 | 40 | 2.84 |
| $\mathrm{CH}_{2} \mathrm{CH}_{3}$ | 1.02 | 10.3 | -0.05 | -0.1 | -0.07 | -0.15 | 0 | 40 | 2.43 |
| $\mathrm{CH}_{2} \mathrm{OCH}_{3}$ | -0.78 | 12.07 | 0.01 | 0.02 | 0.02 | 0.03 | 20 | 40 | 2.53 |


| $\mathrm{OCH}_{2} \mathrm{CH}_{3}$ | 0.38 | 12.47 | 0.22 | -0.44 | 0.1 | -0.24 | 20 | 40 | 3.10 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathrm{~N}^{\left(\mathrm{CH}_{3}\right)_{2}}$ | 0.18 | 15.55 | 0.1 | -0.92 | -0.15 | -0.83 | 70 | 40 | 2.78 |
| $\mathrm{CH}\left(\mathrm{CH}_{3}\right)_{2}$ | 1.53 | 14.96 | -0.05 | -0.1 | -0.07 | -0.15 | 0 | 50 | 2.46 |
| $\mathrm{C}_{3} \mathrm{H}_{7}$ | 1.55 | 14.96 | -0.06 | -0.08 | -0.07 | -0.13 | 0 | 60 | 2.43 |
| $\mathrm{C}_{4} \mathrm{H}_{9}$ | 2.13 | 19.61 | -0.06 | -0.11 | -0.08 | -0.16 | 0 | 80 | 2.43 |
| $\mathrm{C}_{4}\left(\mathrm{CH}_{3}\right)_{3}$ | 1.98 | 19.62 | -0.07 | -0.13 | -0.1 | -0.2 | 0 | 60 | 2.49 |
| $\mathrm{C}_{6} \mathrm{H}_{5}$ | 1.96 | 25.36 | 0.08 | -0.08 | 0.06 | -0.01 | 15 | 120 | 2.51 |
| $\mathrm{C}_{5} \mathrm{H}_{11}$ | 2.67 | 24.26 | -0.06 | -0.08 | -0.08 | -0.16 | 0 | 100 | 2.43 |
| $\mathrm{OC}_{6} \mathrm{H}_{5}$ | 2.08 | 27.68 | 0.34 | -0.35 | 0.25 | -0.03 | 35 | 120 | 3.13 |
| $\mathrm{~N}\left(\mathrm{C}_{2} \mathrm{H}_{5}\right)_{2}$ | 1.18 | 24.85 | 0.01 | -0.91 | -0.23 | -0.9 | 70 | 80 | 2.80 |

Table 3. The correlations among the nine descriptors.

| Descriptors | $\boldsymbol{\pi}$ | $\mathbf{M r}$ | $\mathbf{F}$ | $\mathbf{R}$ | $\boldsymbol{\sigma}_{\boldsymbol{m}}$ | $\boldsymbol{\sigma}_{\boldsymbol{p}}$ | $\mathbf{I}$ | $\mathbf{O}$ |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{M r}$ | 0.601 |  |  |  |  |  |  |  |
| $\mathbf{F}$ | -0.175 | -0.251 |  |  |  |  |  |  |
| $\mathbf{R}$ | 0.116 | -0.104 | 0.243 |  |  |  |  |  |
| $\boldsymbol{\sigma}_{\mathbf{m}}$ | -0.105 | -0.250 | 0.934 | 0.574 |  |  |  |  |
| $\boldsymbol{\sigma}_{\mathrm{p}}$ | 0.007 | -0.200 | 0.650 | 0.895 | 0.879 |  |  |  |
| $\mathbf{I}$ | -0.667 | -0.109 | 0.234 | -0.171 | 0.136 | -0.025 |  |  |
| $\mathbf{O}$ | 0.752 | 0.837 | -0.030 | 0.206 | 0.050 | 0.144 | -0.378 |  |
| $\mathbf{X}$ | -0.193 | -0.245 | 0.637 | -0.219 | 0.459 | 0.122 | 0.186 | -0.244 |

$$
\begin{aligned}
& \pi=0.9485+0.011901 \mathbf{I} \\
& \mathrm{n}=50, \mathrm{r}=0.6670, \mathrm{~s}=0.78, \mathrm{~F}=38.47 \\
& \pi=0.8795+0.02583 \mathbf{O} \\
& \mathrm{n}=50, \mathrm{r}=0.7521, \mathrm{~s}=0.69, \mathrm{~F}=62.5 \\
& \pi=0.1558+0.007963 \mathbf{I}+0.02003 \mathbf{O} \\
& \mathrm{n}=50, \mathrm{r}=0.8581, \mathrm{~s}=0.5437, \mathrm{~F}=65.60 \\
& \pi=0.4221+0.001135(\mathbf{I}-\mathbf{O}) \\
& \mathrm{n}=50, \mathrm{r}=0.8202, \mathrm{~s}=0.5993, \mathrm{~F}=98.65
\end{aligned}
$$

The inorganic property and organic property values of twenty amino acids are also calculated (see Table 4). The correlation between inorganic and organic property values of these 20 amino acids is only 0.054 . The correlation of $\log P$ value of twenty amino acid and inorganic property and organic property values can be seen from Table 4 and from the following equations:

$$
\begin{aligned}
& \log P=5.0863-0.01618 \mathbf{I} \\
& n=20, r=0.7420, \mathrm{~s}=1.0423, F=22.04 \\
& \log P=-1.5130+0.01710 \mathbf{O} \\
& n=20, r=0.5256, \mathrm{~s}=1.3226, F=6.87
\end{aligned}
$$

$$
\begin{aligned}
& \log \mathrm{P}=3.3275-0.01685 \mathbf{I}+0.01847 \mathbf{O} \\
& \mathrm{n}=20, \mathrm{r}=0.9337, \mathrm{~s}=0.5729, \mathrm{~F}=57.78
\end{aligned}
$$

Where $\log \mathrm{P}$ is the residue lipophilicity, $\mathbf{I}$ is the inorganic property value of the amino acid, $\mathbf{O}$ is the organic property value of the amino acid.

Table 4. Amino acid residue $\log \mathrm{P}$, amino acid inorganic property and organic property values and residue electronegativity values.

| Amino acid <br> (residue) | logP Value | Inorganic (I) <br> Value | Organic (O) <br> Value | Residue Electronegativity <br> (X) |
| :---: | :---: | :---: | :---: | :---: |
| A (Ala) | 0.702 | 220 | 60 | 2.40 |
| R (Arg) | -2.061 | 432 | 120 | 2.43 |
| N (Asn) | -1.003 | 355 | 60 | 2.49 |
| D (Asp) | -1.935 | 370 | 60 | 2.50 |
| C (Cys) | 0.987 | 240 | 100 | 2.44 |
| Q (Gln) | -0.936 | 355 | 80 | 2.49 |
| E (Glu) | -1.868 | 370 | 80 | 2.44 |
| G (Gly) | 0.184 | 220 | 40 | 2.20 |
| H (His) | -1.321 | 375 | 120 | 2.45 |
| I (Ile) | 2.167 | 220 | 120 | 2.46 |
| L (Leu) | 2.167 | 220 | 110 | 2.44 |
| K (Lys) | -0.790 | 290 | 120 | 2.43 |
| M (Met) | 1.246 | 240 | 140 | 2.44 |
| F (Phe) | 2.423 | 235 | 180 | 2.45 |
| P (Pro) | 1.128 | 230 | 100 | $/$ |
| S (Ser) | -0.453 | 320 | 60 | 2.52 |
| T (Thr) | -0.042 | 320 | 80 | 2.45 |
| W (Trp) | 1.878 | 350 | 220 | 2.45 |
| Y (Tyr) | 1.887 | 335 | 180 | 2.45 |
| V (Val) | 1.640 | 220 | 90 | 2.46 |

## Applications

Like $\log P$, most hydrophobicity scales described in reference [12] are highly correlated with inorganic property and organic property values of the twenty amino acids. So, the hydrophobic scale can also be defined by $\mathbf{I}$ and $\mathbf{O}$ values. Two equations can be obtained:

H1 = I/O-3, if H1 < 0, the amino acid is hydrophobic, otherwise hydrophilic; $\mathrm{H} 2=\mathrm{I}-\mathrm{O}-160$, if $\mathrm{H} 2<0$, the amino acid is hydrophobic, otherwise hydrophilic. http://www.unibas.ch/mdpi/ecsoc/e0002/aa_io.htm

A plot of $\mathbf{O}$ vs $\mathbf{I}$ can be seen from Figure 2. As shown in Figure 1, two sets of amino acids can be seen; one set is more hydrophilic with $\pi<0.0$, the other set is more hydrophobic with $\pi>0.0$.


Figure 2. Organic property value versus inorganic property value of amino acids.
A new group electronegativity scale has been proposed in the reference [8], and its comparison with other scales of group electronegativity has been discussed. The application of group electronegativity, inorganic property and organic property values of drug molecules have been successfully used in QSAR studies of some drug molecules [13-17]. To further test the usefulness of group electronegativity, inorganic property and organic property values, the common data set used in recent published papers was selected and tested (the detail data set and descriptions can be seen in literature [18,19]). Maddalena and Johnston [18] used ten final descriptors out of $6 \times 7=42$ descriptors ( $\pi 7$, Mr1, Mr2, Mr6, F7, F2, R1, $\sigma_{\mathrm{m}} 3, \sigma_{\mathrm{p}} 8, \mu 1$ ), found that using artificial neural network methods which gave a high value of correlation coefficients for both training ( 0.938 ) and cross-validation ( 0.896 ) to perform similar study. Sung-Sau So and Martin Karplus [19] used six descriptors ( $\pi 7$, F7, Mr1, $\sigma_{\mathrm{m}} 2$, $\pi 6, \mathrm{Mr} 8$ ), which gave QSAR results as good or even better than those calculated using higher dimensions by genetic QSAR neural networks [20]. The author does not focus on the methods for QSARs, but focuses on the parameters proposed here. The electronegativity for each of six groups was calculated, as well as the sum of the inorganic property and organic property values, and the sum of the lipophilicity $(\pi)$ value. The thirteen parameters used in the references $[18,19]$ were also used here. Therefore, in total, twenty-two descriptors were used here for fifty-seven benzodiazepines QSAR studies, using the Minitab program (a common statistical program). By using forward selection method and selecting up to seven descriptors, no proposed parameters were selected, but most of them are among the best ten alternative descriptors, which means that they can be substituted for other descriptors if they are not available. By using a backward elimination method, the proposed parameters are useful in nine-descriptor-QSAR equation (see below). Table 5 gives the summary, and Figure 3 shows the plot of the correlation ( $\mathrm{r}^{2}$ ) versus number of descriptors used in the equation.

$$
\begin{array}{rl}
\log \mathbf{I C} & 50 \\
= & 2.9561-0.64431 \mathbf{X}_{\mathrm{R} 7}-0.34616 \mathbf{X}_{\mathrm{R} 2}+0.6575 \mathbf{X}_{\mathrm{R} 8}-0.012646 \mathbf{O}-0.30804 \pi 7 \\
& +0.16273 \mathbf{M R} 1+1.1648 \pi 6+0.10732 \mathbf{M R} 8+2.6698 \mathbf{R} 1 \\
\mathrm{n}=57, \mathrm{r}^{2} & =0.878, \mathrm{~s}=0.2779, \mathrm{~F}=37.58
\end{array}
$$

Table 5. Stepwise regression of $\operatorname{logIC}_{50}$ on 23 predictors, with $\mathbf{n}=57$

| Step <br> Constant | $\begin{aligned} & 1 \\ & 1.757 \end{aligned}$ | $\begin{array}{\|l\|} \hline 2 \\ 1.444 \\ \hline \end{array}$ | $\begin{aligned} & 3 \\ & 1.506 \end{aligned}$ | $1.850$ | $\begin{aligned} & 5 \\ & 2.098 \\ & \hline \end{aligned}$ | $\begin{array}{\|l\|} \hline 6 \\ 2.145 \end{array}$ | $\begin{aligned} & \hline 7^{*} \\ & 1.998 \\ & \hline \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\sigma_{\mathrm{m}} \mathbf{2}$ | -2.54 | -2.31 | -1.91 | -1.78 | -1.79 | -1.64 | -1.68 |
| T-ratio | -5.44 | -5.50 | -5.36 | -5.84 | -6.20 | -6.03 | -6.42 |
| MR1 |  | 0.067 | 0.093 | 0.106 | 0.100 | 0.126 | 0.131 |
| T-ratio |  | 3.97 | 6.25 | 8.15 | 7.93 | 8.72 | 9.33 |
| $\pi$ (total) |  |  | -0.363 | -0.352 | -0.234 | -0.221 | -0.290 |
| T-ratio |  |  | -5.08 | -5.78 | -3.23 | -3.29 | -4.07 |
| F7 |  |  |  | -1.15 | -1.71 | -1.85 | -1.76 |
| T-ratio |  |  |  | -4.55 | -5.38 | -6.19 | -6.09 |
| $\pi 7$ |  |  |  |  | -0.209 | -0.239 | -0.203 |
| T-ratio |  |  |  |  | -2.69 | -3.29 | -2.85 |
| R1 |  |  |  |  |  | 2.33 | 2.36 |
| T-ratio |  |  |  |  |  | 3.06 | 3.22 |
| MR6 |  |  |  |  |  |  | 0.116 |
| T-ratio |  |  |  |  |  |  | 2.30 |
| s | 0.593 | 0.527 | 0.436 | 0.372 | 0.352 | 0.326 | 0.313 |
| $\mathrm{r}^{2}$ | 0.3496 | 0.4968 | 0.6615 | 0.7578 | 0.7879 | 0.8213 | 0.8387 |
| F | 29.56 | 26.65 | 34.53 | 40.68 | 37.89 | 38.29 | 36.39 |

The next best alternative variables are $\pi 6, \mathbf{X}_{\mathrm{R} 7}, \mathbf{X}_{\mathrm{R} 8}, \sigma_{\mathrm{p}} 8, \mathbf{M R} 8, \mathbf{O}, \mathbf{X}_{\mathrm{R} 1}, \mathbf{F} 2, \mathbf{X}_{\mathrm{R} 6}$.


Figure 3. The correlation $r^{2}$ versus number of descriptors.

## Discussion

It is known that inorganic property and organic property values are good intrinsic descriptors, which reflect the inorganic property and organic property of the organic and biological molecules or groups. Only carbon hydrates give pure organic property values. Other organic molecules and biological molecules with nitrogen and oxygen atoms have both inorganic and organic property values. Urea $\left(\mathrm{NH}_{2}\right)_{2} \mathrm{CO}$ and carbon dioxide $\mathrm{CO}_{2}$ still have organic property values of 20. An organic molecule or a biological molecule with pure inorganic property value and without organic property value has never been found. According to the inorganic property and organic property values of the common groups shown in Table 1, it is found that most groups, even sulfur ( S ) or chloride ( Cl ), also have partial organic property values. The correlation of inorganic property and organic property values of 20 amino acids with other amino acid or residue parameters (like logP) are also extensively studied by the author. The difference between the inorganic property and organic property values, or their ratios, of the amino acids, can be used to identify which amino acids are hydrophobic or hydrophilic. Table 1 cannot cover all the possible groups (e.g. $\mathrm{N}_{3}$ group). The author assigns the inorganic property and organic property values of this $\mathrm{N}_{3}^{-}$group as 10 and 50 , respectively. The advantage of using group electronegativity is that all the group electronegativities can be calculated.

The usage of inorganic property and organic property values of drug and biological molecules is that they can replace the $\pi$ or $\mathbf{M r}$ descriptors, which may be useful in some QSAR studies [21]. The group electronegativity can provide an additional descriptor for each variable group in the drug molecules (for example, in benzodiazepine/ $\mathrm{GABA}_{A}$ receptors, there are six variable groups ( $\mathrm{R}_{7}, \mathrm{R}_{1}, \mathrm{R}_{2}$, $\mathrm{R}_{6}, \mathrm{R}_{3}, \mathrm{R}_{8}$ ), and these six group electronegativities $\mathbf{X}_{i}$ can be used as additional descriptors for QSAR studies).

## Conclusions

The inorganic and organic property values reflect the inorganic and organic properties of the organic or biological molecules or groups, which mainly reflect the hydrophilic or hydrophobic characteristics. Group electronegativity reflects the electrostatic properties of the groups. JavaScript is an easy tool for chemists to calculate the molecular weight, inorganic property and organic property values, and the group electronegativity of organic or biological molecules, as well as other types of descriptor (like logP). The proposed descriptors are useful in QSAR studies of high-dimensional and large-number-sample systems.

## Appendix

## Appendix I. Hydrophobic Fragmental Constant of Groups and Amino Acid Residues

The Hydrophobic Fragmental Constant of Groups and Amino Acid Residues can be seen at the URL http://www.unibas.ch/mdpi/ecsoc/e0002/app.htm.

## References

1. Katritzky, A. A.; Rachwal, P. Law, K. W.; Karelson, M.; Lobanov, V. S. Prediction of polymer glass transition temperatures using a general quantitative structure-property relationship treatment. J. Chem. Inf. Comput. Sci. 1996, 36, 879-884.
2. Bravi, G.; Gancia, E.; Zaliani, A. MS-WHIM, New 3D Theoretical Descriptors Derived from Molecular Surface Properties: A Comparative 3D QSAR Study in A Series of Steroids. J. Computer-aided Mol. Design 1997, 11, 79.
3. Estrada, E.; Ramirez, A. Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications. J. Chem. Inf. Comput. Sci. 1996, 36, 837.
4. Karelson, M; Lobanov, V.S.; Katritzky, A.R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. Chem Rev. 1996, 96, 1027.
5. Katrizky, A.P.; Gordeeva, E.V. Traditional Topological Indices vs. Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR Research. J. Chem. Inf. Comput. Sci. 1993, 33, 835.
6. Fujita, M. Organic Analysis; Kakoya bookstore: Tokyo, 1930; p. 33 (in Japanese).
7. Yuan, L.B.; Ding Y. Organic Conceptional Diagram and Its Application in Surface Chemistry, Hиaxue Tongbao 1988, (2), 19-21.
8. Wu, H. Study of Group Electronegativities in Biological and Drug Molecules. Zhongguo Kexue Jishu Daxиe Xuebao 1990, 20, 517-24.
9. JavaScript. See http://home.netscape.com/eng/mozilla/3.0/handbook/javascript/index.html.
10. Mosley, D.H.; André, J-M. Use of JavaScript in Simple Quantum Chemical Applications. ECCC3, November 1996, at http://hackberry.chem.niu.edu/ECCC3/.
11. Rekker, R.F. The Hydrophobic Fragmental Constant, Its Derivation and Application, A Means of Characterizing Membrane Systems; Elsevier: Amsterdam, 1977.
12. Black, S.D.; Mould, D.R. Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications. Anal. Biochem. 1991, 193, 72-82.
13. Wu, H.; Wen, Y. Computer Aided Anthracycline Anticancer Drugs Design, 2nd WATOC World Congress, Toronto, 8-14 July 1990.
14. Wu, H.; Wen, Y. Quantitative Structure-Activity Relationship Studies of Lincomycin Derivates. 2nd Symposium on Molecular Mechanics and Drug Design of China, Shanghai, P.R. China, November 1989.
15. Wu, H.; Wen, Y. Quantitative Structure-Activity Relationship Studies of Leucomycin and Clindamycin Derivates. 2nd Symposium on Molecular Mechanics and Drug Design of China, Shanghai, P.R. China, November 1989.
16. Wu, H.; Wen, Y. Structure Activity Studies of Podophyllotoxin Derivates. 2nd Symposium on Molecular Mechanics and Drug Design of China, Shanghai, P.R. China, November 1989.
17. Wu, H.; Wen, Y. Structure Activity Studies of Erythromycin Derivates. 2nd Symposium on Molecular Mechanics and Drug Design of China, Shanghai, P.R. China, November 1989.
18. Maddalena, D.J.; Johnston, G.A.R. Prediction of Receptor Properties and Binding Affinity of Ligands to Benzodiazepine/GABA ${ }_{\mathrm{A}}$ Receptors Using Articial Neural Networks. J. Med. Chem. 1995, 38, 715-724.
19. So, S-S.; Karplus, M. Genetic Neural Networks for Quantitative Structure-Activity Relationships: Improvements and Application of Benzodiazepine Affinity for Benzodiazepine/GABA $A$ Receptors. J. Med. Chem. 1996, 39, 5246-5256.
20. So, S-S.; Karplus, M. Evolution Optimization in Quantitative Structure-Activity Relationship: An Application of Genetic Neural Networks. J. Med. Chem. 1996, 39, 1521-1530.
21. Gao, H.; Hansch, C. QSAR of P450 Oxidation: On the Value of Comparing $\mathrm{K}_{\text {cat }}$ and $\mathrm{K}_{\mathrm{m}}$ with $\mathrm{K}_{\mathrm{ca}} / \mathrm{K}_{\mathrm{m}}$. Drug Metabolism Rev. 1996, 28, 513-526.

Samples Availability: Not available.
© 1999 by Molecular Diversity Preservation International (MDPI, http://www.mdpi.org)

