First International Electronic Conference on Synthetic Organic Chemistry (ECSOC-1), www.mdpi.org/ecsoc/, September 1-30, 1997

[E0002]

Re-propose Organic and Inorganic Property Values and Group Electronegativity for Drug and Biological Molecules and Their Calculation through JavaScript and Application in QSAR Studies

Hanqing Wu
Department of Chemistry
University of Wisconsin-Milwaukee
Milwaukee, WI 53201
USA
Tel.: 414-332-8563, Fax: 414-332-8563
wuhanqing@geocities.com

Received: 13 July 1997 / Uploaded: 14 July 1997

Comments

During 1-30 September 1997, all comments on this poster should be sent by e-mail to ecsoc@listserv.arizona.edu with **E0002** as the message subject of your e-mail. After the conference, please send all the comments and reprints requests to the author(s).

Title page

Abstract

Introduction

Calculation

Applications

discussion

Conclusion

Acknowledgments

References

Appendix

First International Electronic Conference on Synthetic Organic Chemistry (ECSOC-1), www.mdpi.org/ecsoc/, September 1-30, 1997

Abstract

The inorganic and organic property values of general orgainc groups are re-proposed here. The inorganic (**I**) and organic (**O**) property values of drug and biological molecules or groups can be calculated based on their group values. The calculation can be easily done on line through Javascript. Similar calculation can be done for the drug and biological molecular group electronegativity according to the author's published paper. The calculation of lipophilicity (pi or logP) parameter of (macro)molecules (like proteins) can also be done on-line through Javascript. Two equations expressed with **I** and **O** are provided here to define the hydrophobicity of each amino acid. The correlations of inorganic and organic property values with other parameters are also discussed. These calculated papameters combined with other papameters can be used for QSAR studies in some drug molecules. Some applications are also disscussed in this paper.

Key words: Inorganic and organic properties, logP, drug and biological molecules, Javascript, QSAR, drug design.

Introduction

Quantitative structure-activity relationship (QSAR) correlations have been widely applied in biological activities over several decades, and many new descriptors (parameters) have been developed [1-5]. In the reference [1], six main types of molecular descriptors (parameters) are introduced, these are *constitutiional*, *geometrical*, *topological*, *electrostatic*, *quantum-chemical*, and *thermodynamic* classes of descriptors.The calculation of these descriptors are also introduced in the reference by many packages. Inorganic and organic property values [6] and group electronegativities [7] are also calculated solely on the basis of intrisic structural information of the molecular species under consideration. Obviously the group electronegativity belongs to the electrostatic descriptors,and it reflects the characteristics of the partial charge of the group. According to the author's previous work [7], it is highly correlated with proton chemical shift in X-H molecules (X is the group). The calculation of inorganic and organic property values of organic molecules as well as group electronegativities are calculated through Javascript [8], which is a powerful language used in internet and some authors have used it elsewhere [9]. Other parameters (like pi or logP value) can also be calculated through Javascript. The correlations of inorganic and organic property values of organic molecules or groups with other parameters are analyzed here, and the usage of these descriptors (parameters) are also discussed in QSAR studies of some drugs.

Calculation

Before calculation inorganic and organic property values, groups electronegativity through Javascript, one simple example is given first: It is common for organic chemists to calculate the molecule weight of organic molecules or some biological molecules (amino acids or nucleic acids), the common atoms in organic or biological molecules are carbon (**C**), hydrogen (**H**), oxygen (**O**), nitrogen (**N**), sulfur (**S**), phospholus (**P**), and others(like chloride (**Cl**), floride (**F**), bromide (**Br**), water (**W** or **w**) etc.). To calculate the molecular weigh of organic or biological molecules through Javascript, the users only need input the type of atoms (either lower case or upper case for the first symbol, such as: for chlorde, input "**Cl**" or "**cl**"; for carbon, input "**C**" or "**c**") and numbers of the atoms, there are total six rows to input maximun of six types of atoms at one calculation which is enough 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 six rows; for cobalt, input "**Co**" or "**co**" in one of six rows). The molecular weight calculation can be shown by "cricking" the following line. The logP has the similar calculation as molecular wight calculation. The symbols of groups can be seen from the references [10,11] and the appendix or the souce codes. Modification of the symbols can be done by the readers at their own convient. Due to the large and complicated groups of organic and drug molecules, the structure of the groups are shown on screem, and the numbers of groups are needed to calculate the organic or inorganic property values of the groups; both the numbers of groups or atoms and values of electronegativity of groups or atoms are needed to calculate the large group electronegativity. Some modification can be done for calculation above parameters through Javascript, the author does not provide all the possibilities. Other parameters can also be calculated through Javascript similar as calculation of logP.

Molecular Weight Calculating Spreadsheet

logP Calculating through Javascript

Inorganic and organic property values of groups.

Calculation of Inorganic property value of molecules or groups through JavaScript.

Calculation of Organic property value of molecules or groups through JavaScript.

Calculation of Group Electronegativity through JavaScript.

Results

The inorganic and organic property values of common organic or biological molecule groups are listed in Table 1. The author calculates the 50 common groups' inorganic and organic property values, and analyzed the correlations with other parameters, lipophilicity (p), polar constant (F), molar refractivity (MR), resonance constant (R), Hammett meta constant (s_m),para constant (s_p). All the parameters can be seen in Table 2, and the correlation among them can be seen in Table 3. The plot of inorganic property value vs organic property value of molecule groups is shown in Figure 1. It is clear that two types of group can be devided: one group is the hydrophobic molecule groups with $p > 0.1$, the other groups is the hydrophilic molecule groups with $p < 0.1$.

Table 3. The correlation among the nine descriptors

Descriptors	p	Mr	F	R	s_m	s_p	I	O
Mr	0.601							
F	-0.175	-0.251						
R	0.116	-0.104	0.243					
s_m	-0.105	-0.250	0.934	0.574				
s_p	0.007	-0.200	0.650	0.895	0.879			
I	-0.667	-0.109	0.234	-0.171	0.136	-0.025		
O	0.752	0.837	-0.030	0.206	0.050	0.144	-0.378	
X	-0.193	-0.245	0.637	-0.219	0.459	0.122	0.186	-0.244

$p = 0.9485 + 0.011901$ **I**

n = 50, r = 0.6670, s = 0.78, F = 38.47

$p = 0.8795 + 0.02583$ **O**

n = 50, r = 0.7521, s = 0.69, F = 62.5

$p = 0.1558 + 0.007963$ **I** $+ 0.02003$ **O**

n = 50, r = 0.8581, s = 0.5437, F = 65.60

p = 0.4221 + 0.001135 (**I-O**)

n = 50, r = 0.8202, s = 0.5993, F = 98.65

From Table 3, the author knows that there are high correlations among p and **I** or **O**, between Mr and **O** (r=0.837). The author also knows that F, R, s_m, s_p, and X have no correlation with **I** or **O**; X has some correlation with F (r=0.637). F has high correlation with s_m and s_p, s_p has high correlation with R and s_m (see Table 3).

The inorganic and organic property values of twenty amino acid are also calculated, the correlation between inorganic property and organic property values of twenty amino acid is only 0.054. The correlation of logP value of twenty amino acid and inorganic and organic property values (see Table 4) can be seen in the following equations:

Table 4. Amino acid residue logP, amino acid inorganic and organic property values and residue electronegativity values

Amino acid (residue)	log**P** value	Inorganic (**I**) value	Organic (**O**) value	Residue Electro-negativity (**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

---

logP = 5.0863 - 0.01618 **I**

n = 20, r = 0.7420, s = 1.0423, F = 22.04

logP = -1.5130 + 0.01710 **O**

n = 20, r = 0.5256, s = 1.3226, F 6.87

logP = 3.3275 - 0.01685 **I** + 0.01847 **O**

n = 20, r = 0.9337, s = 0.5729, F = 57.78

where: logP is the residue lipophilicity; **I** is the inoragnic property value of the amino acid; **O** is the organic property value of the amino acid.

Applications

The author knows that, like logP, most hydrophobicity scales listed in Table 3 of the reference are highly correlated with inorganic and organic property value of the twenty amino acid.

So, the hydrophobic scale can also be defined by I and O values, two equations can be obtained:

H1 = I/O -3, if H1 < 0, the amino acid is hydrophobic, otherwise hydrophilic;

H2 = I - O -160, if H2 < 0, the amino acid is hydrophobic, otherwise hydrophilic.

The plot of **O** vs **I** can also be seen at Figure 2, the same as in Figure 1, two groups can be devided, one group is more hydrophilic with logP < 0.0, the other is more hydrophobic with logP > 0.0.

The group electronegativity has been proposed on the reference [7], and the comparison with other group electronegativity scales has been discussed. The author will not discuss it here.

The application of group electronegativity, inorganic and organic property values of drug molecules have been successfully used in QSAR studies of some drug molecules [12-16]. To test futher the usefulness of group electronegativity, inorganic and organic property values, the author selects the common data set used by recent published papers (the detail data set and descriptions can be seen in the reference [17,18]). In summary, Maddalena and Johnston [17] used the ten final descriptors out of 6*7=42 descriptors, which are $^{p}7$, MR1, MR2, MR6, F7, F2, R1, $^{s}_{m}3$, $^{s}_{p}8$, mi1, gave high value of correlation coefficients for both training (0.938) and cross-validation (0.896) by the methods of artificial neural networks; and Sung-Sau So and Martin Karplus [18] used six descriptors, which are $^{p}7$, F7, MR1, ^{s} 2, $^{p}6$, MR8, gave the

m

QSAR results as good or even better than those with higher dimentions by genetic neural networks for quantitative structure-activity relationships [19]. The author does not focus on the methods for QSARs, but focus on the parameters proposed here. Each of six groups' electronegativities are calculated, and the sum of the inorganic and organic property values of the six groups, sum of the lipophilicity (p) values are also calculated, combined with the thirteen parameters used in the references [17,18] which are mentioned above. Total twenty-two descriptors are used here for those fifty-seven benzodiazepines QSAR studies by the Minitab program, a common statistical progam. By using forward selcetion, and using of upto seven descriptors, none of the parameters proposed here are selected, but most of them are among the best ten alternative descriptors, which means that they can be substituted with other descriptors used if they are not available; by using a backward elimination strategy, the parameters proposed here are significant useful in nine descriptors-QSAR equations. Table 5 gives the summary. The correlation r^2 vs number of descriptors can be seen in Figure 6. The nine descriptors equation can be seen in the following:

$$\log \mathbf{IC}_{50} = 2.9561 - 0.64431X_{R7} - 0.34616\ X_{R2'} + 0.6575\ X_{R8} - 0.012646\ \mathbf{O} - 0.30804\ p7 + 0.16273\ MR1 + 1.1648\ p6 + 0.10732\ MR8 + 2.6698\ R1$$

$$n = 57,\ r^2 = 0.878,\ s = 0.2779,\ F = 37.58$$

Discussion

From the application of inorganic and organic property values of orgaic molecules or groups in QSAR studies of some drugs, the author knows that they are good intrinsic descriptors, and these values reflect the inorganic and organic property values of the organic and biological molecules or groups. Only carbon hydrates give the pure organic property values. Other organic molecules and biological molecules with nitrogen and oxygen atoms have both inorganic and organic property values. Urea $(NH_2)_2CO$, and carbon dioxide 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 is never found. According to the inorganic and organic property values of the common groups in Table 1 the common groups, even the sulfur (S) or chloride (Cl), also has partial organic property values. The correlation of inorganic 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 of inorganic and organic property values or the ratio of inorganic and organic proterty values of the amino acids can be used to identify the hydrophobic or non-hydrophobic amino acids. The disadvantage of usage of the inorganic or organic property values is some of their values are not provided in Table 1, such as N_3^- group, has not their values, and the author assigns the inorganic and organic value as 10 and 50 respectively. The advantage of usage of group electronegativity is that all the group electronegativity can be calculated.

The usage of inorganic and organic property values of drug and biological molecules is that they can replace the Pi or Mr descriptors, which may be useful in some QSAR studies [20]. The group electronegativity can provide additional descriptor for each variable groups in the drug molecules (for example, in benzodiazepine/$GABA_A$ receptors, there are six variable groups (R_7, R_1, $R_{2'}$, $R_{6'}$, R_3, R_8), and the six group electronegativities X_i can be used as additional descriptors for QSAR studies).

Conclusion

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. The Javascript is an easiest tool for organic chemists to calculate the molecular weight, inorganic and organic property values, group electronegativity of organic or biological molecules, and other types of descriptor (like logP). The descriptors proposed here are useful in QSAR studies of high dimentional and large sample systems..

Acknowledgments

This article is dedicated to my father **Shuren Wu** who has given years and years of encouragement .

Reference

[1]. Katritzky, Alan A. et al., 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, *Journal of computer-aided molecular design* 1997, **11**(1), 79

[3]. Estrada, Ernesto; Ramirez, Alain, Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications *J. Chem. Inf. Comput. Sci.* 1996, **36**(4), 837

[4]. Karelson, Mati; Lobanov, Victor S.; Katritzky, Alan R., Quantum-Chemical Descriptors in QSAR/QSPR Studies *Chemical reviews* 1996, **96**(3),1027

[5]. Katrizky, Alan P.; Gordeeva, Elkaterina V., Traditional Topological Indices vs. Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR REsearch, *J. Chem. Inf. Comput. Sci.* 1993, **33**(6), 835

[6]. Fujita, M "Organic Analysis", Tokyo, Kakoya bookstore, 33(1930) (in Japanese).

[7]. Wu, H. Study of group electronegativities in biological and drug molecules. *Zhongguo Kexue Jishu Daxue Xuebao*, 1990, **20**(4), 517-24(Ch).

[8]. JavaScript. See http://home.netscape.com/eng/mozilla/3.0/handbook/javascript/index.html.

[9]. David H. Mosley, David H.; and André, Jean-Marie Use of JavaScript in Simple Quantum Chemical Applications. ECCC3, November, 1996 at http://hackberry.chem.niu.edu/ECCC3/.

[10]. Roelof F. Rekker, The Hydrophobic Fragmental Constant, Its Derivation and Application A Means of Characterizing Membrane Systems, Elsevier Scientific Publishing Company, Amsterdam-Oxford-New York, 1977.

[11]. Black, Shaun D.; Mould, Diane R. Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications, *Analytical Biochemistry* 1991, **193**, 72-82

[12]. Wu, H.; Wen, Y. Computer Aided Anthracycline Anticancer Drugs Design, Second WATOC World Congress, Toronto, 8-14 July 1990

[13]. 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.

[14]. 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.

[15]. 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.

[16]. 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.

[17]. Maddalena, Desmond J.; Johnston, Graham A. R. Prediction of Receptor Properties and Binding Affinity of Ligands to Benzodiazepine/$GABA_A$ Receptors Using Articial Neural Networks, *J. Med. Chem.* 1995, **38**, 715-724

[18]. So, Sung-Sau; Karplus, Martin 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

[19]. So, Sung-Sau; Karplus, Martin Evolution Optimization in Quantitative Structure-Activity Relationship: An Application of Genetic Neural Networks *J. Med. Chem.* 1996, **39**, 1521-1530

[20]. Gao, Hua; Hansch, Corwin QSAR of P450 Oxidation: On the Value of Comparing Kcat and Km with Kcat/Km *Drug Metabolism Reviews* 1996, **28**(4), 513-526

Appendix

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

Structure of the groups	Alternative structure	p value
c6h5	ph1	1.886
c6h4	ph2	1.688
c6h3	ph3	1.431
ch3	me1	0.702
ch2	me2	0.530
ch	me3	0.235
c(quart.)	c	0.15
ch2dch	ch2//ch	0.935
chtc	ch///c	0.73
h	H	0.175
h(neg)	H(neg)	0.462
(al)cooh	cooh	-0.954
(ar)cooh	cooh/	-0.093
(al)coo	coo	-1.292
(ar)coo	coo/	-0.431
(al)co	co	-1.703
(ar)co	co/	-0.842
(al)o	o	-1.581

	(ar)o	o/	-0.433
	(al)oh	oh	-1.491
	(ar)oh	oh/	-0.343
	(ar)coh	coh/	-0.38
	(al)och2cooh	och2cooh	-1.155
	(ar)och2cooh	och2cooh/	-0.581
	(al)cf3	cf3	0.757
	(ar)cf3	cf3/	1.331
	(al)ccl3	ccl3	1.79
	(al)f	f	-0.462
	(ar)f	f/	0.399
	(al)cl	cl	0.061
	(ar)cl	cl/	0.922
	(al)br	br	0.270
	(ar)br	br/	1.131
	(al)i	i	0.587
	(ar)i	i/	1.448
	c5h4n	c5nh4	0.526
	c3h3n2	c3n2h3	-0.119
	(al)nh2	nh2	-1.428
	(ar)nh2	nh2/	-0.854
	(al)nh	nh	-1.825
	(ar)nh	nh/	-0.964
	(al)n	n	-2.16
	(ar)n	n/	-1.012
	(ar)so2nh2	so2nh2/	-1.530
	(ar)so2nh	so2nh/	-1.992
	(ar)so2n	so2n/	-2.454
	(al)conh2	conh2	-1.970

(ar)conh2	conh2/	-1.109
(al)con	con	-2.894
(al)nhcoo	nhcoo	-1.943
(ar)nhcoo	nhcoo/	-0.795
(al)oocnh2	oocnh2	-1.481
(al)no2	no2	-0.939
(ar)no2	no2/	-0.078
(al)ctn	ctn	-1.066
(ar)ctn	ctn/	-0.205
(ar)cdn	cdn/	-1.88
(ar)chdchno2	chdchno2/	0.395
(ar)chdcno2	chdcno2/	0.220
(ar)chdchcoo	chdchcoo/	0.042
(ar)chdchconh	chdchconh	-1.1
(al)sh	sh	0.0
(ar)sh	sh/	0.62
(al)s	s	-0.51
(ar)s	s/	0.11
(al)s-s	ss	0.37
(al)so	so	-2.75
(ar)so	so/	-2.05
(ar)so2	so2/	-1.87
pe(1)	pe1	0.861
pe(2)	pe2	0.574
cm(oct)	cm_oct	0.268
cm(s)	cm_s	0.268
rho1	r2	1.00
rho2	r1	1.00
ala	ALA	0.702

cys	CYS	0.987
asp	ASP	-1.935
glu	GLU	-1.868
phe	PHE	2.423
gly	GLY	0.184
his	HIS	-1.321
ile	ILE	2.167
lys	LYS	-0.790
leu	LEU	2.167
met	MET	1.246
asn	ASN	-1.003
pro	PRO	1.128
gln	GLN	-0.936
arg	ARG	-2.061
ser	SER	-0.453
thr	THR	-0.042
val	VAL	1.640
trp	TRP	1.878
tyr	TYR	1.887
asx	ASX	-1.469
glx	GLX	-1.402
gla	GLA	-4.281
pca	PCA	0.164
hyp	HYP	0.304
hyl	HYL	-1.614
ack	ACK	1.436
mek	MEK	-1.022
m2k	M2K	-1.384
m3k	M3K	-1.526

msx	MSK	-0.994
mso	MSO	-3.234
hse	HSE	-0.386
hcy	HCY	1.054
css	CSS	1.172
ogs	OGS	-3.498
ogt	OGT	-3.087
asg	ASG	-34.586
ac1	AC1	-0.028
ac2	AC2	0.490
ac3	AC3	6.814
ac4	AC4	7.993
nh3	NH3	-0.271
coo	COO	-0.858
pep	peptide	-2.882

* In the structures in the table, double bond uses "d" to represent it, triple bound uses "t" to represent it; in the alternative structure, (al) is omitted, (ar) is also omitted, but add "/" at the end of alternative structure. All the data are come from Roelof F. Rekker, The Hydrophobic Fragmental Constant, Its Derivation and Application aA Means of Characterizing Membrane Systems, Elsevier Scientific Publishing Company, Amsterdam-Oxford-New York, 1977; Shaun D. Black and Diane R. Mould, Development of Hydrophobicity Parameters to Analyze Proteins Which Bear Post- or Cotranslational Modifications, *Analytical Biochemistry*, 193, 72-82 (1991).

Appendix II. QSAR studies on fifty-seven benzodiazepines

The total twentytwo descriptors and logIC_{50} values used here for those fifty-seven benzodiazepines QSAR studies and some results obtained from by the use of Minitab program can be downloaded from the zip file here.