The Calculation of Physical Properties of Amino Acids Using Molecular Modeling Techniques (II)

Myung-Jae Lee† and Ui-Rak Kim*

Department of Chemistry, Keimyung University, Daegu 704-701, Korea
**Department of Natural Science (Chemistry), Medical College, Catholic University of Korea, Seoul 137-701, Korea*
Received May 10, 2004

Six physical properties (enthalpy, density, decomposition temperature, solubility in water, p*K*a values, and hydronium potential) were examined by molecular modeling techniques. The molecular connectivity index, Wiener distance index, and Ad hoc descriptor are employed as structural parameters to encode information about branching, size, cyclization, unsaturation, heteroatom content, and polarizability. This paper examines the correlation of the molecular modeling techniques parameters and the physicochemical properties of amino acids. As a results, calculated values were in agreement with experimental data in the above six physical properties of amino acids and the molecular connectivity index was superior to the other indices in fitting the calculated data.

Key Words : Physical properties of amino acid, Molecular modeling techniques, Correlation relationships

Introduction

Proteins are the most abundant macromolecules in living cells, they are found in all cells and all parts of cell. Moreover proteins have many different biological roles since they are the molecular instruments through which genetic information is expressed. All proteins, the most ancient lines of bacteria or the highest forms of life, are constructed from the basic 20 kinds of amino acids, Each of these amino acids have a distinctive side chain which lends its chemical individuality.

We have already investigated to find out relationship between the physical properties (molecular weight, heat capacity, side chain weight, side chain volume, standard entropy and partial molar volume) and structures of amino acids, peptides and their derivatives[1], various biological activities[2] (enzyme inhibitory, lipoxygenase inhibition, tadpole narcosis, vapor toxicity and heat of vaporization) and the physical properties for the organic compounds and their binary mixture[3] by the molecular descriptors of molecular modeling technique which can encode information about size, branching, cyclization, polarizability, unsaturation and heteroatom content.[4] The molecular connectivity indices,[5-7] Wiener distance indices, and Ad hoc descriptors.[11-13] have been known as the most useful molecular modeling parameters. These methods are currently being used by many chemists as an aid in the development of new drugs or chemical compounds.

In this paper, we examine the correlation of the molecular modeling technique's parameters and the physical properties of amino acids using least square method. As a result, we have obtained regression equations for six representative physical properties of amino acids. Such equations can be used predict values for unmeasured properties in the design of compounds that have properties suitable for special purposes. Also we have evaluated the relative merits of the descriptor sets in relating the physical properties of amino acids and derivatives.

*To whom all correspondence should be addressed. e-mail: kur@kmu.ac.kr

Calculation

Molecular Connectivity Index[5-7]. It is well known that chain isomers of a molecule have varying values of their physical and chemical properties. Then the molecular connectivity is described numerically on the basis of the branching in molecular skeletons.

We have dealt with structural fragments or subgraphs of various bond compositions.

It is possible to decompose a molecular structure into subgraphs of atoms. Thus the subgraph would be described by a single δ for each atoms which is to record at each atom how many neighboring atoms are bonded to it. We shall call these number δ (delta) values. We would compute the reciprocal square root of each δ_i and sum these to form an index we can call a zero-order index ${}^0\chi$.

$$ {}^0\chi = \Sigma(\delta_i)^{-1/2} \tag{1} $$

This index is directly related to atoms, not bonds. It decribles the number of atoms in a molecule and the branching in the molecule.

And the first order connectivity index, ${}^1\chi$ is the sum over all connections or edges in the hydrogen suppressed molecular frame. Atom i and j are formally bonded

$$ {}^1\chi = \Sigma 1/(\delta_i \cdot \delta_j)^{-1/2} \tag{2} $$

Similarly, different orders ${}^m\chi$ of the connectivity index are defined, according to the numbers of continuous bonds included as

$$^{2}\chi = \Sigma 1/(\delta_i\delta_j\delta_k)^{-1/2}$$
$$^{3}\chi = \Sigma 1/(\delta_i\delta_j\delta_k\delta_l)^{-1/2}$$
$$\vdots \quad \vdots$$
$$^{m}\chi = \Sigma 1/(\delta_1\delta_2\delta_3\ldots\delta_{m+1})^{-1/2} \quad (3)$$

Additional parameters for molecular connectivity index, $^{m}\chi_t$, can be obtained by summing analogous terms on substructural units involving path, cluster or pathcluster combinations of m bonds. It remained desirable to include multiple bonds and heteroatoms in the Randic method[5] in some natural way. For this purpose, Kier and Hall[6,7] originally proposed two form of δ assignment; one is the simple adjacency count, and the other is the adjacency over multiple paths in unsaturated bonds. This latter δ value is termed a valence δ ("valence delta") and is written δ^{v}. It is generalize by the expression $\delta^{v} = Z^{v} - hi$, where Z^{v} is the maximum valence and hi is the number of hydrogens on the carbon atom.

Valence molecular connectivity indices χ^{v} are then calculated from valanced δ^{v} values instead of δ values in the same manner as above equations. More recently, Kier and Hall[7] have suggested the formular

$$\delta^{v} = \frac{Z^{v} - h_i}{Z - Z^{v}} \quad (4)$$

where Z is the total number of electrons, to account for elements not in the second row of the periodic table. Now we calculated for valenced δ values using this formular. Then δ^{v} values for heteroatoms are shown in Table 1.

Wiener Distance Index[8-10]. Wiener[8] proposed an index of molecular structure based on path distances between carbon atoms.

The Wiener distance index, W, is the sum of all unique, shortest path distances, in terms of C-C bonds, of hydrogen suppressed molecular frame. The index W is calculated by forming a table of shortest path distances between all carbon atoms and summing its elements.

The Wiener distance index increased with the number of carbon atoms and it is lower for branched isomers than for more extended isomers.

Table 1. δ and δ^{v} values for heteroatoms

Atome (group)	Type of molecule	δ	δ^{v}
-OH	alcohol	1	5
=O	carbonyl	1	6
-O-	ether	2	6
$-NH_2$	*pri*-amine	1	3
-NH-	*sec*-amine	2	4
>N-	*tert*-amine	3	5
≡N	nitrile	1	5
-N=	pyridine	2	5
|| -N-	nitro	3	5

In some case a reduced Wiener index, $W_r = W/N_c^2$, is used, where N_c is the number of carbon atoms. Wiener[10] also proposed a parameter P_3, equal to the number of pairs of carbon atoms separated by three bonds. P_3 paramter is described as a polarizability parameter, because it is more properly related to steric aspects of structure. In many application the modified Wiener index, $W_{mod} = W+P_3$, has been used. Also Platt[13] introduced an additional structural parameter, f, which was calculated by taking the number of adjustment C-C bonds summed over all C-C bonds of the molecule.

It is desirable to extend the Wiener scheme to include heteroatom such as *N*, *O*, *S* etc. It defined that an atomic site index[10] S_i which is the sum of all shortest distance from atom i to other atoms.

The inverse of W and Wr are also included in these parameter sets.

Ad hoc descriptor[11-13]. Ad hoc descriptors, which are not so sophisticated indices, are also good modeling techniques for characterized molecular structures.

The first parameter, used in Ad hoc descriptor, N_c is the number of carbon atoms, which is simplest representative of molecular mass and volume.

Also, each branch of a hydrocarbon being terminates as a methyl group, the number of terminal methyl groups, T_m, is a crude measure of branching or compactness.[11]

Randic[5] introduced a steric parameter T_3, which is the number of terminal methyl separated by 3 bonds in molecular frame.

For aliphatic alcohols a convenient additional index[13] is the number of carbons bonded to the alpha carbon, C_α.

And polarizability parameter, P_3, has been identified above in connection with the Wiener scheme.

The inverse $1/N_c$ and the quadratic terms N_c^2, T_3^2 and T_m^2 were also included in these parameter sets.

Results and Discussion

Values for the most significant parameters according to the molecular modeling techniques for various amino acids using the above molecular modelling parameter calculation method are given in Table 2.

We can write down two forms of δ assignment: one, previously described, in the simple adjacency count, and the other is the adjacency over multiple paths in unsaturated bonds. This latter δ value is termed a valence δ ("valence data") and in written δ^{v}.

The index calculated using δ^{v} valence is thus a valence connectivity index, or valence χ, and is written $^{m}\chi^{v}$.

$^{0}\chi^{v}$, $^{1}\chi^{v}$, $^{2}\chi^{v}$, $^{3}\chi^{v}$ .... $^{m}\chi^{v}$ of all moleculer was calculated by the Table 1 and equation (1), (2), (3), and (4). These parameter sets correlated with reported six representative values of physical properties for amino acids[14-23] using the least square method. The results are presented with the coefficients of correlation (r) and standard errors (S.E) for the most successful one, two and three parameter models. All three structural parameter sets tested were reasonably

Table 2. Parameter values for various compounds using modeling technique

Compounds	$^0\chi^v$	$^1\chi^v$	$^2\chi^v$	$^3\chi^v$	$^4\chi^v$	W	Wr	f	P_3	T_3	T_m	N_c
Glycine	2.639	1.189	0.597	0.174	0.000	18	4.500	8	0	0	0	2
Alanine	3.510	1.627	1.126	0.389	0.000	29	3.222	12	1	0	1	3
Valine	5.087	2.537	2.108	1.003	0.285	63	2.520	19	4	0	1	5
Leucine	5.794	3.020	2.574	1.042	0.608	96	2.667	20	5	1	2	6
Isoleucine	5.794	3.075	2.249	1.542	0.497	92	2.556	20	5	2	2	6
Serine	3.664	1.774	1.127	0.513	0.078	46	5.110	14	1	2	0	3
Treonine	4.534	2.218	1.604	0.804	0.206	63	3.937	18	2	0	1	4
Aspartic acid	4.572	2.239	1.542	0.711	0.275	76	6.000	20	3	0	0	4
Glutamic acid	5.279	2.739	1.877	0.986	0.402	136	5.440	22	3	0	0	5
Lysine	5.915	3.366	2.229	1.286	0.632	143	3.972	20	4	0	0	6
Argine	6.708	3.600	2.427	1.324	0.668	247	6.861	24	3	0	0	6
Asparagine	4.702	2.304	1.615	0.738	0.304	96	6.000	20	2	0	0	4
Glutamine	5.409	2.804	1.949	1.018	0.421	136	5.440	22	3	0	0	5

Table 3. Multiple regression equations for the physical properties of amino acids using molecular connectivity index

Enthalpy = $0.2476 + 1.9435\ ^0\chi^v - 3.0188\ ^1\chi^v - 1.4934\ ^2\chi^v - 0.1237\ ^0\chi^v + 3.0655\ ^4\chi^v$
n = 13 r = 0.98291 S.E = 0.47

Decomposition temperature = $452.316 - 53.758\ ^0\chi^v - 130.674\ ^1\chi^v - 181.968\ ^2\chi^v - 33.811\ ^3\chi^v + 140.458\ ^4\chi^v$
n = 13 r = 0.98762 S.E = 0.56

Solubility in water = $7.719 + 2.483\ ^0\chi^v + 3.944\ ^1\chi^v - 5.722\ ^2\chi^v - 4.454\ ^3\chi^v - 22.405\ ^4\chi^v$
n = 13 r = 0.9870 S.E = 0.27

p*Ka* = $2.601 - 0.551\ ^0\chi^v + 0.561\ ^1\chi^v + 0.588\ ^2\chi^v + 0.275\ ^3\chi^v + 1.028\ ^4\chi^v$
n = 13 r = 0.982 S.E = 0.02

Density = $0.0939\ ^1\chi^v + 1.6352\ ^2\chi^v - 1.2267\ ^3\chi^v$
n = 10 r = -0.9867 S.E = 0.3562

Potential of Hydronium = $2.0901 - 7.781\mathrm{N}\ ^0\chi^v + 19.873\ ^1\chi^v + 0.426\ ^2\chi^v - 0.924\ ^3\chi^v - 11.815\ ^4\chi^v$
n = 13 r =0.9932 S.E = 0.2745

Table 4. Multiple regression equations for the physical properties of amino acids using Ad hoc desciptor

Enthalpy = $0.4259 + 0.0013\ W + 0.1674\ Wr - 0.04725\ f + 0.0126 P_3$
n = 13 r = 0.9589 S.E = 0.5742

Decomposition temperature = $361.357 - 0.062W - 3.712Wr - 7.2835\ f + 20.705 P_3$
n = 13 r = 0.929 S.E = 0.3720

Solubility in water = $44.622 + 0.122W - 5.980\ Wr - 0.036\ f - 7.556 P_3$
n = 13 r = 0.9572 S.E = 0.2242

p*Ka* = $2.967 + 0.018W - 0.241\ Wr + 0.0425\ f - 0.300 P_3$
n = 13 r = 0.932 S.E = 0.443

Density = $9.83171 + 14.1489W - 1.262\ Wr - 0.0226\ f - 0.5200\ P_3$
n =10 r = 0.896 S.E = 0.377

Potential of hydronium = $13.272 + 0.063W - 1.546Wr - 0.2062\ f - 0.960 P_3$
n = 13 r = 0.948 S.E = 0.5621

Table 5. Multiple regression equations for physical properties of amino acids using Wiener distance index

Enthalpy = $0.9543 - 0.1368T_3 - 0.0944T_m - 0.0839N_c$
n = 13 r =0.9743 S.E = 0.245

Decomposition temperature = $288.339 - 14.516\ T_3 - 50.401T_m - 10.519N_c$
n = 13 r = 0.39 S.E = 0.237

Solubility in water = $25.621 - 4.208T_3 - 4.566T_m - 4.046N_c$
n = 13 r = 0.897 S.E = 0.426

p*Ka* = $2.362 - 0.047T_3 + 0.196T_m - 0.044N_c$
n = 13 r = 0.9660 S.E = 0.282

Density = $4.562 - 0.057T_3 + 0.327T_m - 0.027N_c$
n = 10 r = 0.998 S.E = 0.398

Potential of hydronium = $3.029 - 0.074\ T_3 - 0.490T_m + 0.743N_c$
n = 13 r = 0.994 S.E = 0.479

Table 6. Correlation between molecular modeling indices and enthalpy of amino acids

Compounds	Enthalpy (Cal K^{-1} mol^{-1})			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	0.94	0.91	0.90	0.89
Alanine	0.58	0.54	0.45	0.61
Valine	0.17	0.15	0.08	0.44
Leucine	0.36	0.30	0.22	0.23
Isoleucine	-0.02	0.03	0.09	-0.02
Serine	0.32	0.47	0.57	0.52
Treonine	0.37	0.45	0.41	0.51
Aspartic acid	0.80	0.83	0.65	0.74
Glutamic acid	0.17	0.18	0.31	0.15
Lysine	0.03	0.03	0.38	0.45
Argine	0.98	0.95	0.88	0.90
Asparagine	0.74	0.80	0.74	0.67
Glutamine	0.67	0.65	0.61	0.64

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

Table 7. Correlation between molecular modeling indices and decomposition temperature of amino acids

Compounds	Decomposition Temperature (°C)			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	292	290	285	267
Alanine	297	289	281	307
Valine	315	309	299	286
Leucine	337	336	303	312
Isoleucine	284	280	304	298
Serine	228	221	258	228
Treonine	253	261	253	297
Glutamic acid	249	242	237	236
Aspartic acid	270	267	250	256
Lysine	224	223	275	225
Arginine	238	202	208	225
Asparagine	236	250	230	246
Glutamine	185	223	235	236

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

Table 8. Correlation between molecular modeling indices and solubility in water of amino acid

Compounds	Solubility in Water (g/100cc)			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	25	22.78	20.62	19.72
Alanine	16.5	15.72	20.91	18.05
Valine	8.85	8.10	7.33	9.52
Leucine	2.32	1.89	1.92	3.20
Isoleucine	4.12	3.84	3.07	3.07
Serine	5.0	6.29	8.62	5.07
Treonine	20.5	18.25	18.01	19.78
Aspartic acid	0.5	0.62	-0.20	7.35
Glutamic acid	0.84	0.75	0.52	0.54
Lysine	9.01	2.42	3.04	1.34
Arginine	15.0	9.29	12.22	13.47
Asparagine	3.1	3.21	4.62	4.20
Glutamine	185	180.20	82.60	160.25

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

successful in accounting for the physical properties of these amino acids. But the molecular connectivity indices and Wiener distance indices generally provided the most favorable correlation.

Conclusion

We have applied molecular modeling techniques to six representative physical properties (solubility in water, p*Ka*, hydronium potential, enthalpy, density, decomposition temperature) of amino acids, Most of the properties were well modeled by the molecular modeling techniques.

We obtained the regression equation for these properties using Wiener distance indices, molecular connectivity

Table 9. Correlation between molecular modeling indices and pKa values of amino acid

Compounds	pKa values (25 °C) (α-COOH group)			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	2.34	2.12	3.22	2.27
Alanine	2.34	2.35	3.02	2.42
Valine	2.32	2.34	2.42	2.34
Leucine	2.36	2.29	2.28	2.44
Isoleucine	2.36	2.37	2.30	2.40
Serine	2.21	2.30	2.19	2.14
Treonine	2.71	2.47	2.46	2.4
Aspartic acid	1.88	1.89	2.16	2.18
Glutamic acid	2.16	2.19	2.19	2.14
Lysine	2.18	2.25	2.26	2.09
Arginine	2.17	2.03	2.13	2.09
Asparagine	2.02	2.14	2.11	2.18
Glutamine	2.17	2.19	2.19	2.14

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

Table 10. Correlation between molecular modeling Indices and density of amino acid

Compounds	Density (g/cc)			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	1.58	1.47	1.27	1.42
α-alanine	1.42	1.53	1.33	1.32
β-alanine	1.40	1.40	1.28	1.08
Valine	1.32	1.12	1.54	1.56
Leucine	1.32	1.29	1.43	1.45
Serine	1.54	1.48	1.50	1.38
Tyrosine	1.46	1.79	1.48	1.47
Arginine	1.10	1.43	1.23	1.32
Aspartic acid	1.66	1.59	1.45	1.50
Glutamic acid	1.46	1.42	1.34	1.28

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

Table 11. Correlation between molecular modeling indices and potential of hydronium of amino acids

Compounds	Potential of Hydronium			
	obsed[a]	Calc(1)[a]	Calc(2)[b]	Calc(3)[c]
Glycine	5.97	5.58	5.81	4.52
Alanine	6.01	5.68	6.69	4.77
Valine	5.96	5.52	5.61	6.46
Leucine	5.98	5.80	6.31	6.44
Isoleucine	6.02	5.61	6.23	6.36
Serine	5.68	5.87	4.44	5.11
Treonine	6.16	5.18	5.55	5.51
Aspartic acid	2.77	2.92	3.08	6.00
Glutamic acid	3.24	4.64	6.06	6.75
Lysine	9.82	10.11	8.23	7.49
Arginine	10.76	10.48	8.06	7.49
Asparagine	5.41	4.751	4.03	6.00
Glutamine	5.65	5.57	6.06	6.75

Taken from reference[14-23] [a]calculated with the equation in Table 3. [b]calculated with the equation in Table 4. [c]calculated with the equation in Table 5.

indices and Ad hoc descriptors.

The molecular connectivity indices is superior to the other indices in fitting the data for the present set of amino acids. In this case, all the correlation coefficient (r) for the prediction of the six properties are greater than 0.98 and the average percentage deviation are smaller than 1%.

References

1. Kim, U. R.; Min, K. S.; Jeong, B. J. *Bull. Korean Chem. Soc.* **1994**, *15*, 106.
2. Kim, U. R.; Min, K. S.; Kim, J. T.; Jeong, B. J. *J. Korean Chem. Soc.* **1993**, *37*, 68.
3. Kim, U. R.; Min, K. S.; Lee, M. J.; Kim, S. H.; Jeong, B. J. *J. Korean Chem. Soc.* **1992**, *36*, 485.
4. Needham, P. E.; Chim, I.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186.
5. Ranic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
6. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley: New York, 1986.
7. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
8. Wiener, H. *J. Chem. Phys.* **1947**, *15*, 766.
9. Wiener, H. *J. Phys. Chem.* **1948**, *52*, 425.
10. Wiener, H. *J. Am. Chem. Soc.* **1947**, *17*, 2636.
11. Seybold, P. G.; May, M. A.; Bogal, U. A. *J. Chem. Soc.* **1987**, *64*, 575.
12. Seybold, P. G.; May, M. A.; Gargas, M. L. *Acta Pharm. Jugosl.* **1986**, *36*, 253.
13. Platt, J. R. *J. Phys. Chem.* **1952**, *556*, 328.
14. *Handbook of Chemistry and Physics*, 70th ed.; CRC: Boca Roton, F.C. 1989-1990; p 426.
15. *The Merck Index*, 11th ed.; Merck Co. Inc.: Rahway, N. J. 1989; p 244.
16. Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M. *Data for Biochemical Research*, 3th ed.; Clarendon Press: 1992; p 198.
17. *Handbook of Biochemistry*; The Chemical Rubber Co. Inc.; Rahwaay, N. J. 1989; p 156.
18. Stryr, L. *Biochemistry*; W. H. Freeman and Company: San Francisco, 1993.
19. Lepori, L.; Molocav, *J. Phys. Chem. (Neue Folge)* **1980**, *123*, 51.
20. Millieo, F. J.; Losurdo, A.; Schin, C. *J. Phys. Chem.* **1978**, *82*, 784.
21. Jolicoeur, C.; Boileu, J. *Cand. J. Chem.* **1978**, *56*, 1707.
22. Ahluwalia, J. C.; Ostigoy, C.; Penon, G.; Desnoyers, J. E. *Cand. J. Chem.* **1977**, *55*, 3364.
23. Shahidi, F.; Farrell, P. G. *J. Chem. Soc. Faraday Trans.* **1978**, *74*, 858.