Articles

The Calculation of Physical Properties of Amino Acids using Molecular Modeling Techniques

Ui-Rak Kim*, Kyung-Sub Min, and Bong-Jin Jeong

Department of Chemistry, Keimyung University, Daegu 704-701
Received February 19, 1993

Six physical properties (molecular weight, heat capacity, side chain weight, side chain volume, standard entropy and partial molar volume) of amino acids, peptides and their derivatives 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 techique's parameters and the physicochemical properties of amino acids and their derivatives. As a result, calculated values were in agreement with experimental data in the above six physical properties of amino acids, peptides and their derivatives and the molecular connectivity index was superior to the other indices in fitting the calculated data.

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 investigated to find out relationship between the physical properties and structures of amino acids, peptides and their derivatives by molecular modeling techniques. The molecular modeling techniques[1] can encode information about size, branching, cyclization, polarizability, unsaturation and heteroatom content. The most successful of these molecular modeling parameters have been known to be Wiener distance indices[2-4], molecular connectivity indices[5-7] and ad hoc descriptors.[8-10] 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 and their derivatives using linear square method. As a result, we have obtained regression equations for six representative physical properties of amino acids, peptides and their derivatives. Such equations can be used to predict values for unmeasured properties of peptides and their derivatives and may be useful 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.

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.

To each atom of the hydrogen suppressed molecular graph a δ_i values is assigned corresponding to the number of non-hydrogen atoms bonded to it. The zeroth order connectivity index ${}^0\chi$ is then the sum over all atom i

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

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

$$ \begin{aligned} {}^2\chi &= \Sigma 1/(\delta_i\ \delta_j\ \delta_k)^{-1/2} \\ {}^3\chi &= \Sigma 1/(\delta_i\ \delta_j\ \delta_k\ \delta_l)^{-1/2} \\ &\vdots \qquad\quad \vdots \\ {}^m\chi &= \Sigma 1/(\delta_1\ \delta_2\ \delta_3 \cdots\ \delta_{m+1})^{-1/2} \end{aligned} \tag{3} $$

Additional parameters for molecular connectivity index, ${}^m\chi_t$, can be obtained by summing analogous terms oner substructural units involving path, cluster or path-cluster 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] originally proposed that the valence values of atoms be assigned according to $\delta^v = Z^v - h_i$, where Z^v is the number of valence electrons of atom i and h_i is the number of hydrogens bonded to it.

Table 1. δ and δ^v Values for Heteroatoms

Atom (group)	Type of molecule	δ	δ^v
-OH	alcohol	1	5
=O	carboxyl	1	6
=O-	ether	2	6
-NH	*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

Valence molecular connectivity indices χ^v are then calculated from valanced δ^v values instaed 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[2-4]. Wiener[2] 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.

In some case a reduced Wiener index, $W_r = W/N_c^2$, is used, where Nc is the number of carbon atoms. Wiener[5] also proposed a parameter P_3, equal to the number of pairs of carbon atoms separated by three bonds. P_3 parameter 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[10] 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[1] 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[8-10]. 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.[8]

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[10] 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, peptides and their derivatives using the above molecular modelling parameter calculation method are given in Table 2.

These parameter sets correlated with reported six representative values[12-20] of physical properties for amlino acids using the least square method.

The results are presented in Table 3 with the coefficients

Table 2. Selected Parameter Values for the Amino acids, Peptides and Their Derivatives

Compound	$^0\chi$	$^1\chi$	$^3\chi_p$	$^1\chi^v$	N_c	P_3	T_3	T_m	W	W_r	f
Glycine	4.2845	2.2701	0.8165	1.1895	2	0	0	0	18	4.500	8
Alanine	5.1548	2.6727	1.3333	1.6271	3	1	0	1	29	3.222	12
Valine	6.7321	3.5534	2.1031	2.5378	5	4	0	1	63	2,520	19
Leucine	7.4392	4.0366	1.9813	3.0209	6	5	1	2	96	2.667	20
Isoleucine	7.4392	4.0914	2.5931	3.0758	6	5	2	2	92	2.556	20
Serine	5.8618	3.1807	1.7820	1.7742	3	1	2	0	46	5.110	14
Threonine	6.7321	3.5533	2.1031	2.2186	4	2	0	1	63	3.937	18
Proline	5.9832	3.8045	2.3426	2.7669	5	3	1	0	62	2.480	20
Hydroxyproline	6.8535	4.1948	2.6782	2.6482	5	3	0	0	87	2.480	24
Phenylalanine	8.9748	5.6984	3.5679	3.7222	9	9	0	0	268	3.309	33
Tyrosine	8.8451	6.0922	3.9786	3.8665	9	9	0	0	212	2.617	30
Tryptophan	10.8367	7.1815	5.2942	4.7162	11	13	0	0	392	3.240	43
Aspartic acid	7.4392	4.0366	1.9813	2.2393	4	3	0	0	96	6.000	20
Glutamic acid	8.1463	4.5266	2.3374	2.7393	5	3	0	0	136	5.440	22

Table 1. δ and δ^v Values for Heteroatoms

Atom (group)	Type of molecule	δ	δ^v
-OH	alcohol	1	5
=O	carboxyl	1	6
=O-	ether	2	6
-NH	*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

Valence molecular connectivity indices χ^v are then calculated from valanced δ^v values instaed 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[2-4]. Wiener[2] 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.

In some case a reduced Wiener index, $W_r = W/N_c^2$, is used, where Nc is the number of carbon atoms. Wiener[5] also proposed a parameter P_3, equal to the number of pairs of carbon atoms separated by three bonds. P_3 parameter 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[10] 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[1] 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[8-10]. 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.[8]

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[10] 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, peptides and their derivatives using the above molecular modelling parameter calculation method are given in Table 2.

These parameter sets correlated with reported six representative values[12-20] of physical properties for amlino acids using the least square method.

The results are presented in Table 3 with the coefficients

Table 2. Selected Parameter Values for the Amino acids, Peptides and Their Derivatives

Compound	$^0\chi$	$^1\chi$	$^3\chi_p$	$^1\chi^v$	N_c	P_3	T_3	T_m	W	W_r	f
Glycine	4.2845	2.2701	0.8165	1.1895	2	0	0	0	18	4.500	8
Alanine	5.1548	2.6727	1.3333	1.6271	3	1	0	1	29	3.222	12
Valine	6.7321	3.5534	2.1031	2.5378	5	4	0	1	63	2,520	19
Leucine	7.4392	4.0366	1.9813	3.0209	6	5	1	2	96	2.667	20
Isoleucine	7.4392	4.0914	2.5931	3.0758	6	5	2	2	92	2.556	20
Serine	5.8618	3.1807	1.7820	1.7742	3	1	2	0	46	5.110	14
Threonine	6.7321	3.5533	2.1031	2.2186	4	2	0	1	63	3.937	18
Proline	5.9832	3.8045	2.3426	2.7669	5	3	1	0	62	2.480	20
Hydroxyproline	6.8535	4.1948	2.6782	2.6482	5	3	0	0	87	2.480	24
Phenylalanine	8.9748	5.6984	3.5679	3.7222	9	9	0	0	268	3.309	33
Tyrosine	8.8451	6.0922	3.9786	3.8665	9	9	0	0	212	2.617	30
Tryptophan	10.8367	7.1815	5.2942	4.7162	11	13	0	0	392	3.240	43
Aspartic acid	7.4392	4.0366	1.9813	2.2393	4	3	0	0	96	6.000	20
Glutamic acid	8.1463	4.5266	2.3374	2.7393	5	3	0	0	136	5.440	22

Lysine	7.9831	4.6809	2.4011	4.6807	6	4	0	0	143	3.972	20
Arginine	9.5605	6.5116	2.8290	3.6002	6	3	0	0	247	6.861	24
Histidine	8.2677	5.1984	3.3179	3.1594	6	3	0	0	165	4.583	28
Asparagine	7.4392	4.0366	1.9813	2.3043	4	2	0	0	96	6.000	20
Glutamine	8.1463	4.5366	2.3374	2.8043	5	3	0	0	136	5.440	22
α-amino butyric acid	5.7432	3.2137	1.0372	2.1651	4	2	1	1	46	2.875	14
β-amino butyric acid	5.8618	3.1259	0.9428	2.1002	4	2	1	1	48	3.000	14
γ-amino butyric acid	5.6987	3.2701	1.1350	2.1895	4	2	0	0	52	3.250	12
Dialanine	9.0165	4.9473	3.0197	3.2359	6	2	0	2	173	4.806	26
Trialanine	12.8783	7.2518	4.7347	4.8447	9	3	0	3	494	6.099	40
Tetralanine	16.7401	9.5562	6.4497	6.4536	12	4	0	4	1076	7.472	54
Diglycine	7.2760	4.1639	1.9337	2.3508	4	0	0	0	104	6.500	18
Triglycine	10.2676	6.0577	3.0860	3.5120	6	0	0	0	306	8.500	28
Tetraglycine	13.2591	7.9516	4.2383	4.6732	8	0	0	0	664	10.375	38
Pentaglycine	16.2507	9.8454	5.3905	5.8345	10	0	0	0	1252	12.520	48
Diserine	10.4307	6.0233	3.7571	3.5302	6	2	0	0	266	7.389	30
Triserine	14.9996	8.8658	5.7607	5.2861	9	3	0	0	764	9.430	46
5-amino pentanoic acid	6.4058	3.7701	1.3850	2.6895	5	3	0	0	79	3.160	14
6-amino hexanoic acid	7.1127	4.2701	1.6350	3.1895	6	4	0	0	114	3.167	16
7-amino heptanoic acid	7.8200	4.7701	1.8850	3.6895	7	5	0	0	158	3.224	18
8-amino octanoic acid	8.5271	5.2701	2.1350	4.1895	8	6	0	0	212	3.313	20
9-amino nonanoic acid	9.2342	5.7701	2.3850	4.6895	9	7	0	0	277	3.420	22
10-amino decanoic acid	9.9413	6.2701	2.6350	5.1895	10	8	0	0	354	3.540	24
11-amino undecanoic acid	10.6484	6.7701	2.8850	5.6895	11	9	0	0	444	3.669	26

Table 3. Summary of Results for One, Two and Three Parameter Models for the Six Physical Properties in Amino Acids

Property	Molecular connectivity index			Ad hoc descriptor			Wiener distance index		
	Parameter	r	S.E.	Parameter	r	S.E.	Parameter	r	S.E.
Molecular	$^1\chi$	0.9855	5.192	N_c	0.8978	13.502	f	0.9379	11.158
Weight	$^1\chi$ $^3\chi_p$	0.9865	5.003	N_c $1/N_c$	0.9158	10.215	f W/N_c	0.9850	5.229
	$^1\chi$ $^3\chi_p$ $^0\chi$	0.9980	1.927	N_c $1/N_c$ T_m	0.9460	9.906	f W/N_c W_r	0.9925	3.716
Heat	$^1\chi^v$	0.9463	2.718	N_c	0.8725	4.111	W/N_c	0.8061	5.004
Capacity	$^1\chi^v$ $^2\chi^v$	0.9786	1.739	N_c P_3	0.9193	3.354	W/N_c W_r	0.9510	2.615
	$^1\chi^v$ $^2\chi^v$ $^3\chi^v$	0.9788	1.724	N_c P_3 T_3	0.9313	3.065	W/N_c W_r W	0.9586	2.346
Standard	$^1\chi^v$	0.9351	3.315	N_c	0.8369	4.988	W/N_c	0.8004	5.462
Entropy	$^1\chi^v$ $^3\chi_p$	0.9641	2.419	N_c P_3	0.8994	3.982	W/N_c W_r	0.9291	3.393
	$^1\chi^v$ $^3\chi_p$ $^0\chi$	0.9808	1.778	N_c P_3 T_3^2	0.9191	3.591	W/N_c W_r W	0.9538	2.727
Side-chain	$^0\chi$	0.9934	3.621	N_c	0.8927	13.766	f	0.9358	10.766
Weight	$^0\chi$ $^1\chi$	0.9970	2.383	N_c T_m	0.9155	12.812	f W/N_c	0.9854	6.648
	$^0\chi$ $^1\chi$ $^3\chi_p$	0.9980	1.957	N_c T_m $1/N_c$	0.9446	10.034	f W/N_c W_r	0.9925	3.729
Side-chain	$^1\chi^v$	0.9359	8.713	N_c	0.9347	9.937	f	0.8715	10.842
Volume	$^1\chi^v$ $^0\chi$	0.9645	6.533	N_c $1/N_c$	0.9670	6.303	f W/N_c	0.9615	6.782
	$^1\chi^v$ $^0\chi$ $^2\chi^v$	0.9731	5.587	N_c $1/N_c$ P_3	0.9695	6.313	f W/N_c W_r	0.9654	6.446
Partial	$^1\chi^v$	0.9930	4.772	N_c	0.9517	12.545	W_{mod}	0.8612	22.062
Molar	$^1\chi^v$ $^2\chi^v$	0.9943	4.423	N_c P_3	0.9837	7.347	W_{mod} W	0.9392	14.091
Volume	$^1\chi^v$ $^2\chi^v$ $^3\chi^v$	0.9972	3.458	N_c P_3 T_3	0.9847	7.121	W_{mod} W W/N_c	0.9402	13.920

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 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.

The most successful single parameter are seen to be the molecular connectivity indices $^1\chi$ and $^1\chi^v$, the number of carbon atoms N_c, and f parameter. At the three term level, three approaches tent to show similar multiple correlation coefficients. Most of the physical properties are reasonably well predicted for by the present parameters at the three term level.

Table 4. Multiple Regression Equations for the Physical Properties of amino Acids Using Molecular Connectivity Index

Molecular weight = $10.3093^{0}\chi + 6.1533^{1}\chi - 48.8266\ 1/^{1}\chi + 3.8856^{3}\chi_{p} + 33.9043$
$n = 19$ $r = 0.9987$ S.E. = 1.578

Heat capacity = $17.5683^{1}\chi^{v} + 1.0594^{2}\chi^{v} - 15.0576^{3}\chi^{v} + 7.3611$
$n = 19$ $r = 0.9788$ S.E. = 1.724

Standard entropy = $1.6004^{0}\chi - 49.8305\ 1/^{0}\chi - 6.2209^{3}\chi_{p} + 12.2576^{1}\chi^{v} + 20.6643$
$n = 19$ $r = 0.9826$ S.E. = 1.723

Side chain weight = $11.0942^{0}\chi + 5.4455^{1}\chi - 49.6664\ 1/^{1}\chi + 3.4811^{3}\chi_{p} + 14.586$
$n = 19$ $r = 0.9989$ S.E. = 1.601

Side chain volume = $7.6839^{0}\chi + 3.3190^{3}\chi_{p} + 11.3455^{1}\chi^{v} + 27.0146^{2}\chi^{v} - 3.1764$
$n = 19$ $r = 0.9748$ S.E. = 5.576

Partial molar volume = $27.3082^{1}\chi^{v} + 19.4242^{2}\chi^{v} - 20.5163^{3}\chi^{v} - 0.2100$
$n = 36$ $r = 0.9972$ S.E. = 3.458

Table 5. Multiple Regression Equations for the Physical Properties of Amino Acids Using Ad Hoc Descriptor

Molecular weight = $13.3701\ N_c - 3.6753\ P_3 - 15.4602\ T_m - 109.3216\ 1/N_c + 3.3722\ T_3^2 + 105.12$
$n = 19$ $r = 0.9505$ S.E. = 9.467

Heat capacity = $8.3389\ N_c - 3.7146\ P_3 + 4.7843\ T_3 - 3.0865\ T_m + 11.0042$
$n = 19$ $r = 0.9348$ S.E. = 3.079

Standard entropy = $6.3901\ N_c - 3.1256\ P_3 + 3.9842\ T_m - 32.2009\ 1/N_c + 2.9007\ T_3^2 + 29.0634$
$n = 19$ $r = 0.9340$ S.E. = 3.257

Side chain weight = $12.5895\ N_c - 2.2705\ P_3 - 16.3776\ T_m - 113.9979\ 1/N_c + 3.855\ T_3^2 + 91.900$
$n = 19$ $r = 0.9491$ S.E. = 9.634

Side chain volume = $13.0911\ N_c - 3.9716\ P_3 - 75.7293\ 1/N_c + 2.4570\ T_3^2 + 46.0334$
$n = 19$ $r = 0.9760$ S.E. = 5.383

Partial molar volume = $18.5606\ N_c - 4.3538\ P_3 + 4.3445\ T_3 + 7.7031\ 1/N_c + 6.8264$
$n = 36$ $r = 0.9850$ S.E. = 7.050

Table 6. Multiple Regression Equations for the Physical Properties of Amino Acids Using Wiener Distance Index

Molecular weight = $-6.1051\ W_r - 3.6071\ W + 3.8613\ W/N_c + 3.4321\ W_{mod} + 1.1471\ f + 62.3038$
$n = 19$ $r = 0.9973$ S.E. = 2.259

Heat capacity = $-5.1467\ W_r - 2.1989\ W + 1.9881\ W/N_c + 2.1256\ W_{mod} - 0.8844\ f + 440.5717$
$n = 19$ $r = 0.9804$ S.E. = 1.658

Standard entropy = $-5.8896\ W_r - 2.3709\ W + 2.4251\ W/N_c + 2.2416\ W_{mod} - 0.6108\ f + 37.5719$
$n = 19$ $r = 0.9679$ S.E. = 2.834

Side chain weight = $-6.0438\ W_r - 3.7359\ W + 3.9497\ W/N_c + 3.5499\ W_{mod} + 1.1567\ f + 43.1373$
$n = 19$ $r = 0.9977$ S.E. = 2.088

Side chain volume = $-16.688\ W_r - 5.9921\ W + 5.5557\ W/N_c + 5.7127\ W_{mod} - 1.2334\ f + 75.9418$
$n = 19$ $r = 0.9857$ S.E. = 4.173

Partial molar volume = $-13.3814\ W_r - 3.4753\ W + 2.9574\ W/N_c + 3.4189\ W_{mod} + 82.9228$
$n = 36$ $r = 0.9541$ S.E. = 12.236

Table 7. Correlation Between Molecular Modeling Indices and Molecular Weight of Amino Acids

Compounds	Molecular weight			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	75.09	73.69	75.60	77.20
Alanine	89.10	90.35	92.03	89.66
Valine	117.15	119.57	120.07	123.32
Leucine	131.17	131.01	131.11	131.31
Isoleucine	131.17	133.88	129.91	131.31
Serine	105.09	105.44	101.75	105.12
Threonine	119.12	119.57	115.57	111.83
Proline	115.13	115.24	117.43	139.08
Hydroxyproline	131.13	129.14	130.85	139.08
Phenylalanine	165.19	166.76	165.48	180.23
Tyrosine	181.19	180.31	178.93	180.23
Tryptophan	204.21	203.56	205.47	194.48
Aspartic acid	133.10	131.01	134.78	120.25
Glutamic acid	142.13	144.10	145.85	139.08
Lysine	146.19	143.88	141.73	152.42
Arginine	174.21	176.00	175.91	156.09
Histidine	155.16	154.59	154.05	156.09
Asparagine	132.12	131.01	131.35	123.92
Glutamine	146.15	144.10	145.86	139.08

*Taken from reference 12. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

In a comparision of topological indices, we found the connectivity indices to be the most effective in representing several properties of amino acids. More exact regression equations for the six physical properties are shown in Table 4 with the molecular connectivity indices, in Table 5 with the ad hoc descriptors and in Table 6 with the Wiener distance indices. With an aim of simplicity, these equations have been arbitrarily limited to no more than four or five parameters.

Table 7-12 show that relationship between observed and calculated six physical properties values from multiple regression equations in Table 4-6, where 1, 2, 3 in the parenthesis are using the molecular connectivity indices, ad hoc descriptors and Wiener distance indices, respectly. A very significant liner correlation were found between observed values and calculated values using parameter sets.

The molecular connectivity index is superior to the other method in fitting the calculated values for the present set of amino acids properties. All the correlation coefficients of six properties are greater than 0.98 and the standard errors are smaller than 3.5.

Table 8. Correlation Between Molecular Modeling Indices and Heat Capaicty of Amino Acids

Compounds	Heat capacity [Cal K^{-1} mol^{-1}]			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	27.71	26.26	26.91	27.68
Alanine	29.22	31.28	32.59	32.31
Valine	40.35	39.07	39.73	39.54
Leucine	48.03	47.46	44.56	45.86
Isoleucine	45.00	40.56	44.10	45.86
Serine	32.40	31.99	31.13	32.31
Threonine	35.20	35.92	35.33	38.63
Proline	36.13	37.62	36.60	41.55
Hydroxyproline	36.79	38.73	36.03	41.55
Phenylalanine	48.52	51.22	50.99	52.62
Tyrosine	51.73	52.36	53.05	52.62
Tryptophan	56.92	57.31	55.62	54.44
Aspartic acid	37.09	37.62	39.06	33.22
Glutamic acid	41.84	42.62	43.60	41.55
Lysine	–	57.38	47.84	46.18
Arginine	55.80	53.24	55.14	49.89
Histidine	–	45.62	41.17	49.89
Asparagine	38.30	37.96	36.93	36.93
Glutamine	44.02	43.35	43.60	41.55

*Taken from reference 15. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

Table 9. Correlation Between Molecular Modeling Indices and Standard Entropy of Amino Acids

Compounds	Standard entropy [Cal K^{-1} mol^{-1}]			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	24.74	25.39	25.68	25.74
Alanine	30.88	30.89	38.20	30.39
Valine	42.75	42.06	42.50	40.99
Leucine	50.62	50.57	47.25	50.04
Isoleucine	49.71	47.44	46.81	50.04
Serine	35.65	32.20	32.41	34.38
Threonine	36.50	38.15	37.92	39.24
Proline	39.20	41.25	39.53	45.20
Hydroxyproline	41.19	40.16	40.09	45.20
Phenylalanine	51.06	52.90	53.72	54.87
Tyrosine	51.15	53.87	55.67	54.87
Tryptophan	60.00	58.28	57.11	55.79
Aspartic acid	40.66	40.99	42.54	37.20
Glutamic acid	44.98	46.61	47.20	45.20
Lysine	–	69.63	50.24	49.53
Arginine	59.90	57.28	58.34	52.66
Histidine	–	45.95	45.51	52.66
Asparagine	41.70	41.79	40.29	40.32
Glutamine	46.62	47.41	47.20	45.20

*Taken from reference 12. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

Table 10. Correlation Between Molecular Modeling Indices and Side-Chain Weight of Amino Acids

Compounds	Side chain weight			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	57.05	55.47	57.45	60.08
Alanine	71.08	72.37	73.92	71.93
Valine	99.13	101.97	102.16	106.04
Leucine	113.16	113.69	113.28	114.25
Isoleucine	113.16	116.28	112.06	114.25
Serine	87.08	87.52	84.06	88.30
Threonine	101.10	101.97	97.80	94.50
Proline	97.11	96.78	99.41	121.94
Hydroxyproline	113.13	110.97	131.10	121.94
Phenylalanine	147.17	148.87	147.61	162.21
Tyrosine	163.17	162.68	161.07	162.21
Tryptophan	186.20	185.42	187.33	176.21
Aspartic acid	115.09	113.69	117.67	103.65
Glutamic acid	129.11	126.85	128.56	121.94
Lysine	128.17	126.39	124.05	134.95
Arginine	156.19	158.32	158.79	138.32
Histidine	137.14	136.60	136.46	138.32
Asparagine	114.10	113.69	114.12	107.02
Glutamine	128.13	126.85	128.56	121.94

*Taken from reference 12. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

Table 11. Correlation Between Molecular Modeling Indices and Side-chain Volume of Amino Acids

Compounds	Side chain volume [ml mol^{-1}]			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	36.30	35.10	35.95	34.35
Alanine	52.60	52.67	58.69	56.09
Valine	85.10	83.68	87.70	82.91
Leucine	102.00	95.82	97.42	101.93
Isoleucine	102.00	88.46	96.68	101.93
Serine	54.90	58.09	51.44	56.09
Threonine	72.20	73.70	69.36	73.98
Proline	73.60	72.86	78.59	84.43
Hydroxyproline	–	81.42	77.76	84.43
Phenylalanine	113.90	106.55	118.32	119.70
Tyrosine	116.20	117.97	122.01	119.70
Tryptophan	135.40	138.92	131.58	131.52
Aspartic acid	68.40	76.82	74.80	67.55
Glutamic acid	84.70	86.80	88.28	84.43
Lysine	105.10	113.44	100.30	96.07
Arginine	109.10	104.39	111.46	100.04
Histidine	91.10	95.19	88.08	100.04
Asparagine	72.40	65.89	69.08	71.52
Glutamine	92.70	88.02	88.28	84.43

*Taken from reference 12. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

Table 12. Correlation Between Molecular Modeling Indices and Partial Molar Volume of Amino Acids

Compounds	Partial molar volume [ml mol^{-1}]			
	Obsed*	Calc (1)[a]	Calc (2)[b]	Calc (3)[c]
Glycine	36.30	35.10	35.95	34.35
Alanine	52.60	52.67	58.69	56.09
Valine	85.10	83.68	87.70	82.91
Leucine	102.00	95.82	97.42	101.93
Isoleucine	102.00	88.46	96.68	101.93
Serine	54.90	58.09	51.44	56.09
Threonine	72.20	73.70	69.36	73.98
Proline	73.60	72.86	78.59	84.43
Hydroxyproline	–	81.42	77.76	84.43
Phenylalanine	113.90	106.55	118.32	119.70
Tyrosine	116.20	117.97	122.01	119.70
Tryptophan	135.40	138.92	131.58	131.52
Aspartic acid	68.40	76.82	74.80	67.55
Glutamic acid	84.70	86.80	88.28	84.43
Lysine	105.10	113.44	100.30	96.07
Arginine	109.10	104.39	111.46	100.04
Histidine	91.10	95.19	88.08	100.04
Asparagine	72.40	65.89	69.08	71.52
Glutamine	92.70	88.02	88.28	84.43
α-Amino butyric acid	36.30	35.10	35.95	34.35
β-Amino butyric acid	52.60	52.67	58.69	56.09
γ-Amino butyric acid	85.10	83.68	87.70	82.91
Dialanine	102.00	95.82	97.42	101.93
Trialanine	102.00	88.46	96.68	101.93
Tetralanine	54.90	58.09	51.44	56.09
Diglycine	72.20	73.70	69.36	73.98
Triglycine	73.60	72.86	78.59	84.43
Tetraglycinene	–	81.42	77.76	84.43
Pentaglycine	113.90	106.55	118.32	119.70
Diserine	116.20	117.97	122.01	119.70
Triserine	135.40	138.92	131.58	131.52
5-Amino pentanoic acid	68.40	76.82	74.80	67.55
6-Amino hexanoic acid	84.70	86.80	88.28	84.43
7-Amino heptanoic acid	105.10	113.44	100.30	96.07
8-Amino octanoic acid	109.10	104.39	111.46	100.04
9-Amino nonanoic acid	91.10	95.19	88.08	100.04
10-Amino decanoic acid	72.40	65.89	69.08	71.52
11-Amino undecanoic acid	92.70	88.02	88.28	84.43

*Taken from reference 12. [a] Calculated with the equation in Table 4. [b] Calculated with the equation in Table 5. [c] Calculated with the equation in Table 6.

Conclusion

We have applied molecular modeling techniques to six representative physical properties (molecular weight, side chain weight, standard entropy, side chain volume, heat capacity and partical molar volume) of amino acids, peptides and their derivatives. Most of the properties were well modeled by the molecular modeling techniques.

We obtained the regression equations for these properties using Wiener distance indices, molecular connectivity 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 coefficients (r) for the prediction of the six properties are greater than 0.98 and the average percentage deviation are smaller than 3%.

References

1. Needham. P. E.; Chim, I.; Seybold, P. G. *J. Am. Chem. Soc.* **1988**, *110*, 4186.
2. Wiener, H. *J. Chem. Phys.* **1947**, *15*, 766.
3. Wiener, H. *J. Phys. Chem.* **1948**, *52*, 425.
4. Wiener, H. *J. Am. Chem. Soc.* **1947**, *17*, 2636.
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 Cemistry and Drug Research*; Academic: New York, 1976.
8. Seybold. P. G.; May, M. A.; Bogal, U. A. *J. Chem. Soc.* **1987**, *64*, 575.
9. Seybold. P. G.; May, M. A.; Gargas, M. L. *Acta. Pharm. Jugosl.* **1986**, *36*, 253.
10. Platt, J. R. *J. Phys. Chem.* **1952**, *556*, 328.
11. Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1807.
12. *Handbook of Chemistry and Physics*, 70th ed.; CRC, Boca Roton, F. C. **1989-1990**.
13. *The Merck Index*, 11th ed.; Merck Co. Inc.: Rahway, N. J. **1989**.
14. Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M. *Data for Biochemical Research*, 3th ed.; Clarendon Press:
15. *Handbook of Biochemistry*; The Chemical Rubber Co. Inc.: Rahway, N. J. **1989**.
16. Lepori, L.; Molicav, *J. Phys. Chem. (Neue Folge)* **1980**, *123*, 51.
17. Millieo, F. J.; Losurdo, A.; Schin, C. *J. Phys. Chem.* **1978**, *82*, 784.
18. Jolicoeur, C.; Boileu, J. *Cand. J. Chem.* **1978**, *56*, 1707.
19. Ahluwalia, J. C.; Ostigoy, C.; Penon, G.; Desnoyers, J. E. *Cand. J. Chem.* **1977**, *55*, 3364.
20. Shahidi, F.; Farrell, P. G. *J. C. S. Faraday Trans.* **1978**, *74*, 858.