September 2009 *Volume 5 Issue 3*

Trade Science Inc.

Organic CHEMISTRY

An Indian Journal

Full Paper

OCAIJ, 5(3), 2009 [290-295]

QSPR study of amino acids with the help of quantum chemical and energy descriptors

M.Ansari, Roli Tiwari, S.A.Khan, P.P.Singh*

Department of Chemistry, M.L.K. (P.G.) College, Balrampur, U. P., (INDIA)

E-mail : dr_ppsingh@sify.com; ansari.mohiuddin@rediffmail.com

Received: 8th June, 2009 ; Accepted: 18th June, 2009

ABSTRACT

Quantum chemical descriptors such as molecular weight, molar refractivity, ε HOMO & ε LUMO and energy descriptors such as heat of formation, total energy and dielectric energy based QSPR study of 15 amino acids is presented in this paper. Various descriptors have been calculated with the help of CAChe software using PM3 Hamiltonian. The dissociation constants of amino acids have been taken from literature. Multi Linear Regression Analysis has been made with the help of above descriptors in 90 combinations. Regression equations have been found to provide successful models as evident from the regression coefficient r^2 and the cross validation coefficient rCV^2 and also from the predicted value which is very close to the experimental value. Molecular weight in combination with molar refractivity has been noticed as the best quantum chemical descriptor. Similarly heat of formation in combination with total energy has been found to be the best energy descriptor. The dissociation constant of amino acids can be predicted reliably with the help of such regression models. © 2009 Trade Science Inc. - INDIA

KEYWORDS

QSPR;
MLR;
Energy descriptors;
Quantum Chemical descriptors;
Amino Acids;
pK1.

INTRODUCTION

Quantitative Structure Activity Relationship, (QSAR), which relates the biological activity of compounds with their molecular structure[1] has been extensively used as predicting tool in drug design.[2-5] The term activity in QSAR has been replaced by toxicity in QSTR to predict the toxicity of compounds.[6-9] Similarly the term activity in QSAR has been replaced by property in QSPR which allows one to predict diversified properties of compounds such as the reactivity sites[10], the hardness[11] and the boiling points[12] of compounds. Amino acids are biologically very vital molecules as they are the monomers of proteins[13,14] and enzymes. Different properties of amino acids have been studied by different people at different times and to mention one is the use of N- alkylated and N, N-dialkylated amino acids to make biologically important chelates with Cu(II) & Zn(II).[15-18] The stability constant values of copper complexes of amino acids have been experimentally obtained by Basalo[19] and Raos.[20] Raos et al. have also derived the stability constant values with the help of topological indices by QSPR study.[20] Literature survey reveals that attempts have never been made to study the QSPR to predict the pK1 of amino acids and its derivatives. We have taken this

task into consideration and proceeded accordingly and have presented QSPR study of 15 amino acids in this paper. Study is mainly based on two sets of descriptors. First set comprises of quantum chemical descriptors such as molecular weight, molar refractivity, HOMO energy and LUMO energy. Second set comprises of energy descriptors[21-24] such as heat of formation, total energy and dielectric energy.

MATERIALS & METHOD

15 amino acids have been taken as the study material and are given along with their pK1 values in the TABLE 1. The pK1 values of amino acids have been taken from the literature.[25]

TABLE 1 : List of amino acids along with their observed pK1

Chemical Sample	PK1	Chemical Sample	PK1
Gly[G]	2.4	Cys[C]	2.1
Ala[A]	2.4	Met[M]	2
Val[V]	2.2	Asp[D]	2.1
Leu[L]	2.3	Asc[N]	2.1
Ilu[I]	2.3	Arg[R]	2.2
Ser[S]	2.2	His[H]	2.2
Thr[T]	2.1	Phe[F]	2.2
Tyr[Y]	1.9		

TABLE 2 : Values of different quantum chemical descriptors along with amino acids.

Chemical Sample	Molecular Weight	HOMO Energy (eV)	LUMO Energy (eV)	Molar Refractivity
Gly[G]	75.067	-9.218	-0.291	13.346
Ala[A]	89.094	-9.428	0.044	17.840
Val[V]	117.147	-9.414	0.138	26.836
Leu[L]	131.174	-9.422	0.140	31.514
Ilu[I]	131.174	-9.404	0.144	31.437
Ser[S]	105.093	-9.570	0.093	19.384
Thr[T]	119.120	-9.477	0.249	23.802
Tyr[Y]	181.191	-9.384	-0.444	44.153
Cys[C]	121.154	-7.594	0.999	25.511
Met[M]	149.207	-9.358	-0.096	35.175
Asp[D]	118.069	-5.380	4.520	18.492
Asc[N]	132.119	-9.413	-0.122	25.698
Arg[R]	175.210	-13.143	-4.013	45.477
His[H]	155.156	-9.140	-0.009	35.184
Phe[F]	165.191	-9.410	-0.441	42.459

For QSPR prediction, 3D modeling[26] and geometry optimization[27,28] of all the amino acids have been carried out with the help of PCMODEL software using semi empirical PM3 Hamiltonian.

The values of various quantum chemical descriptors have been calculated with the help of CAChe software using PM3 Hamiltonian. The results are reported in the TABLES 2 & 3. MLR analysis has been performed by the Project Leader program associated with the CAChe software.

TABLE 3 : Values of different energy descriptors along with amino acids.

Chemical Sample	Heat of Formation (kcal/mole)	Total Energy (Hartree)	Dielectric Energy (kcal/mole)
Gly[G]	-57.839	-47.602	-4.278
Ala[A]	-70.537	-54.756	-4.105
Val[V]	-81.000	-69.077	-3.775
Leu[L]	-86.240	-76.231	-3.780
Ilu[I]	-85.849	-76.233	-3.733
Ser[S]	-113.057	-66.941	-4.132
Thr[T]	-118.293	-74.084	-3.715
Tyr[Y]	-89.041	-103.167	-4.083
Cys[C]	-61.685	-63.637	-2.512
Met[M]	-72.294	-78.248	-4.255
Asp[D]	-177.662	-76.309	-7.778
Asc[N]	-111.791	-81.690	-4.174
Arg[R]	-70.362	-103.463	-6.848
His[H]	-41.693	-89.992	-4.548
Phe[F]	-43.923	-90.941	-3.884

THEORY

The values of various descriptors have been evaluated by solving the relevant equations as given below,

According to Koopman's theorem, the ionization potential is simply the eigen value of HOMO with change of sign and electron affinity is the eigen value of LUMO with of sign.[29]

Molar refractivity is used as a steric parameter[30] and measures the electronic effect also and may reflect dipole-dipole interaction at the active site,

Molar Ref. = $[(n^2-1)/(n^2+2)]mw/d$

Where n is the refractive index for the sodium D line, mw is the molecular weight and d is density of the compound.

Full Paper

Heat of formation is defined as:

$$\Delta H^{o}_{f} = E_{elec.} + E_{nuc.} - E_{isol.} + E_{atom}$$

where $E_{elec.}$ is the electronic energy, $E_{nuc.}$ is the nuclear-nuclear repulsion energy, $E_{isol.}$ is the energy required to strip off all the valence electrons of all the atoms in the system and E_{atom} is the total heat of atomization of all the atoms in the system.

Total energy of a molecular system is the sum of the total electronic energy, E_{ee} and the energy of internuclear repulsion, $E_{nr.}$

The total electronic energy of the system is given by[31]

$$E = P(H+F)/2$$

Where P = density matrix and H = one electron matrix. P, H and charges[32] on atoms were obtained from PM3 calculations.

RESULT AND DISCUSSION

Several different combinations of the descriptors have been tried and regression equations have been obtained but only 8 regression models were chosen whose regression coefficient r^2 were above 0.7. The predicted pK1 values of amino acids of these models are presented in TABLE 4. The QSPR models along with the regression equations of all the eight models are as follows-

TABLE 4 : Predicted value of pK1with the best eight models.

RE1	RE2	RE3	RE4	RE5	RE6	RE7	RE8
2.418	2.413	2.397	2.39	2.4	2.427	2.418	2.416
2.367	2.37	2.361	2.363	2.363	2.372	2.371	2.363
2.275	2.284	2.299	2.299	2.289	2.269	2.276	2.264
2.234	2.252	2.276	2.275	2.258	2.227	2.241	2.218
2.232	2.248	2.273	2.272	2.255	2.224	2.237	2.217
2.236	2.224	2.219	2.226	2.238	2.201	2.2	2.189
2.187	2.173	2.186	2.192	2.198	2.14	2.141	2.14
1.997	2.001	2.029	2.022	2.019	1.976	1.986	1.975
2.115	2.085	2.089	2.084	2.093	2.148	2.13	2.239
2.046	2.069	2.025	2.037	2.046	2.127	2.131	2.129
2.045	2.092	2.085	2.079	2.052	2.1	2.113	2.057
2.097	2.068	2.068	2.072	2.092	2.048	2.04	2.052
2.237	2.224	2.168	2.172	2.193	2.231	2.212	2.179
2.088	2.061	2.065	2.063	2.066	2.072	2.058	2.11
2.125	2.137	2.16	2.152	2.138	2.138	2.145	2.151

First QSPR Model

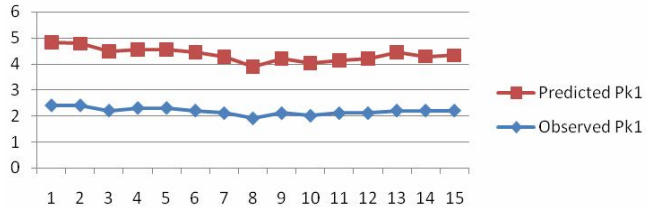
In this QSPR model, we have taken heat of formation as first descriptor, molecular weight as second, total energy as third and molar refractivity as fourth descriptor. Regression equation is as follows-

RE1=0.000565863*ΔHf-0.0179748*MW-0.0111618*TE+0.0287269* MR+2.88484

rCV^2=0.454928

r^2=0.775522

As the regression coefficient is about 0.80, this QSPR model provides very good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 1.



Graph 1 : Graphical presentation of observed pK1 and predicted pK1(RE1)

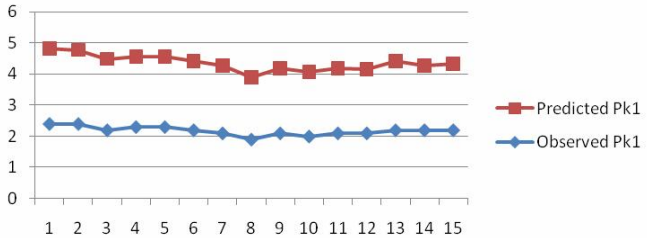
Second QSPR Model

In this QSPR model, we have taken molecular weight as first descriptor, total energy as second, molar refractivity as third and dielectric energy as fourth descriptor.

RE2=-0.0193358*MW-0.00855701*TE+0.0379332*MR-0.0199319*DE+2.86528rCV^2=0.507315

r^2=0.775395

As the regression coefficient is about 0.80, this QSPR model provides very good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 2.



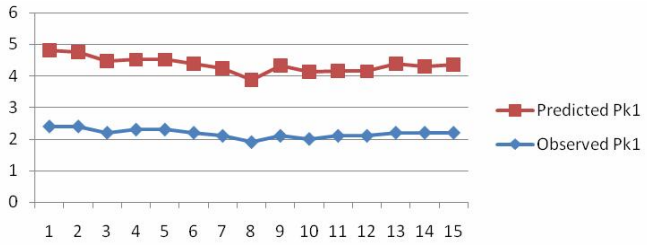
Graph 2 : Graphical presentation of observed pK1 and predicted pK1(RE2)

Third QSPR Model

In this QSPR model, we have taken molecular weight as first descriptor, total energy as second, HOMO energy as third and molar refractivity as fourth descriptor.

RE3=0.0251829*MW0.0157555*
TE+0.0207193*εHOMO+0.0464828*
MR+3.10843rCV^2=0.629882
r^2=0.773318

As the regression coefficient is about 0.80, this QSPR model provides very good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 3.



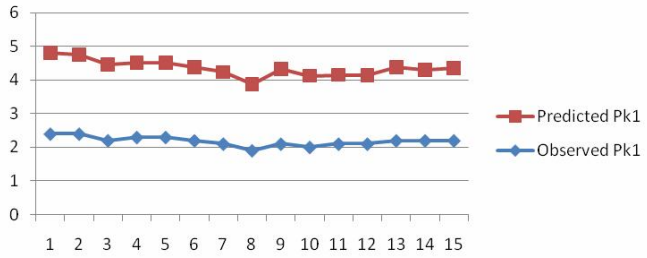
Graph 3 : Graphical presentation of observed pK1 and predicted pK1(RE3)

Fourth QSPR Model

In this QSPR model, we have taken molecular weight as first descriptor, total energy as second, LUMO energy as third and molar refractivity as fourth descriptor.

RE4=0.0234059*MW-0.0139539*TE+
0.0145478*εLUMO+0.0435874*MR
+2.90572
rCV^2=0.637676
r^2=0.766665

As the regression coefficient is about 0.80, this QSPR model provides very good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 4.



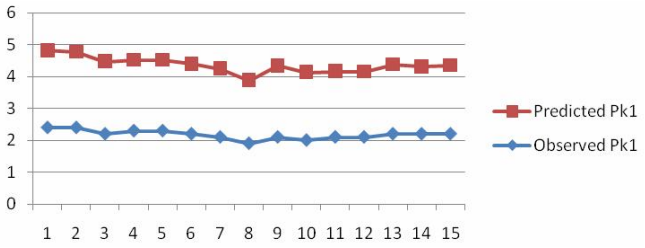
Graph 4 : Graphical presentation of observed pK1 and predicted pK1(RE4)

Fifth QSPR Model

In this QSPR model, we have taken molecular weight as first descriptor, total energy as second and molar refractivity as third descriptor.

RE5=-0.0204993*MW-0.0120431*
TE+0.0365181*MR+2.87822
rCV^2=0.721186
r^2=0.755753

As the regression coefficient is about 0.75, this QSPR model provides good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 5.



Graph 5 : Graphical presentation of observed pK1 and predicted pK1(RE5)

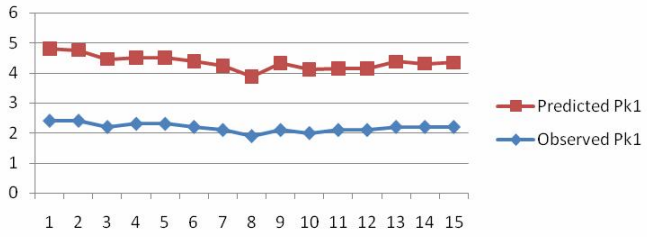
Sixth QSPR Model

In this QSPR model, we have taken heat of formation as first descriptor, molecular weight as second, molar refractivity as third and dielectric energy as fourth descriptor.

RE6=0.000354916*ΔHf-0.0126008*
MW+0.0293405*MR-0.0305223
*DE+2.87093
rCV^2=0.237051
r^2=0.746382

As the regression coefficient is about 0.75, this QSPR model provides good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 6.

Full Paper



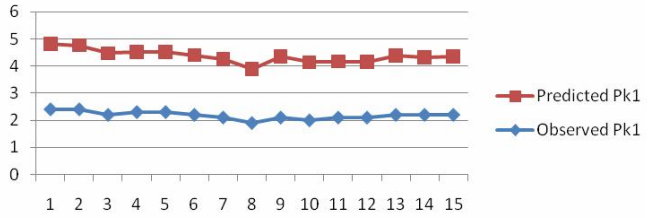
Graph 6 : Graphical presentation of observed pK1 and predicted pK1(RE6)

Seventh QSPR Model

In this QSPR model, we have taken molecular weight as first descriptor, molar refractivity as second and dielectric energy as third descriptor.

RE7=-0.0140726*MW+0.0346558*MR-0.0347153*DE+2.86381
rCV^2=0.574686
r^2=0.739637

As the regression coefficient is about 0.7, this QSPR model provides good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 7.



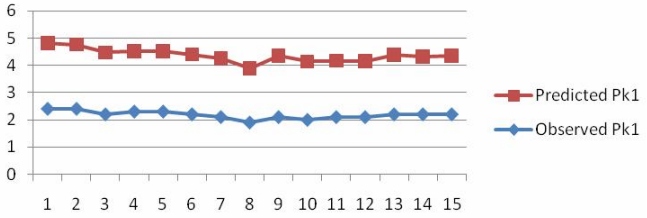
Graph 7 : Graphical presentation of observed pK1 and predicted pK1(RE7)

Eighth QSPR Model

In this QSPR model, we have taken heat of formation as first descriptor, molecular weight as second, LUMO energy as third and molar refractivity as fourth descriptor.

RE8=0.00120855*ΔHf-0.00974959*MW+0.0206717*εLUMO+0.0205929*MR+2.948
rCV^2=0.570402
r^2=0.704961

As the regression coefficient is about 0.7, this QSPR model also provides good predictive power and can be used efficiently to predict the value of pK1 of any amino acid with the help of the descriptors used in regression equation. Graph between observed and predicted values of pK1 is shown in Graph 8.



Graph 8 : Graphical presentation of observed pK1 and predicted pK1(RE8)

CONCLUSION

The values of cross validation coefficient rCV^2 and regression coefficient r^2 and their sum of the most suited regression models is given in the TABLE 5. Analysis of results indicate the following points-

TABLE 5 : List of Cross validation Coefficient and Regression Coefficient with various Descriptor Combinations.

Descriptor Combination	Cross validation Coefficient	Regression Coefficient	Sum of rCV^2 and r^2
RE1	0.454928	0.775522	1.23045
RE2	0.507315	0.775395	1.28271
RE3	0.629882	0.773318	1.4032
RE4	0.637676	0.766665	1.404341
RE5	0.721186	0.755753	1.476939
RE6	0.237051	0.746382	0.983433
RE7	0.574686	0.739637	1.314323
RE8	0.570402	0.704961	1.275363

1. The best regression model is RE1 in which the descriptors are heat of formation, molecular weight, total energy and molar refractivity.
2. The second best regression model is RE2 in which the descriptors are molecular weight, total energy, HOMO energy and molar refractivity.
3. The third best regression model is RE3 in which the descriptors are molecular weight, total energy, HOMO energy and molar refractivity.
4. The best quantum chemical descriptor is molecular weight in combination with molar refractivity as indicated by the regression models RE1, RE2 and RE3.
5. The best energy descriptor is total energy as indicated by regression models RE1, RE2 and RE3.

Thus the pK1 of any amino acid as well as its derivatives can be predicted reliably with the best QSPR model RE1.

ACKNOWLEDGEMENT

Authors are grateful to Dr. R. B. Srivastava, Reader, Mathematics Department, M. L. K. (P. G.) College, Balrampur, U. P., for providing computational data.

REFERENCES

[1] C.Hansch; Acc.Chem.Res., **2**, 232 **(1969)**.
[2] H.K.Srivastava, F.A.Pasha, P.P.Singh; Int.J.Quantum Chem., **103**, 237 **(2005)**.
[3] F.A.Pasha, H.K.Srivastava, P.P.Singh; Mol.Div., **19**, 215 **(2005)**.
[4] F.A.Pasha, H.K.Srivastava, P.P.Singh; Bioorg.Med.Chem., **13**, 6823 **(2005)**.
[5] F.A.Pasha, H.K.Srivastava, P.P.Singh; Int.J.Quantum Chem., **104**, 87 **(2005)**.
[6] R.K.Singh, A.K.R.Khan, V.K.Sahu, P.P.Singh; Int.J.Quantum Chem., **109**, 185 **(2009)**.
[7] D.R.Roy, R.Parthasarathi, B.Maiti, V.Subramanian, P.K.Chattaraj; Bioorg.Med.Chem., **13(10)**, 3405 **(2005)**.
[8] F.A.Pasha, H.K.Srivastava, A.Srivastva, P.P.Singh; QSAR & Comb.Sci., **26**, 69 **(2007)**.
[9] C.Hansch, T.Fujita; J.Am.Chem.Soc., **86**, 1616 **(1964)**.
[10] P.P.Singh, F.A.Pasha, H.K.Srivastava; QSAR & Combi.Sci., **22**, 843 **(2003)**.
[11] P.P.Singh, S.Saxena, Abhay Srivastava, J.P.Singh; A.I.J.Inorg.Chem., **2(2)**, 93 **(2007)**.
[12] Hemant Kumar Srivastava; Bull.Korean Chem.Soc., **29(1)**, 67 **(2008)**.
[13] H.B.Vickery, C.L.A.Scmidt; Chem.Rev., **9**, 169 **(1931)**.
[14] H.Neurath, K.Bailey; The Proteins. Academic Press, New York, **I**, Part A, P. 99 **(1953)**.
[15] N.Paulic, N.J.Raos; Coord.Chem., **31**, 180 **(1994)**.
[16] V.A.Davankov, S.V.Roghozim, A.A.Kurganov; Izu.Akad.Nauk.SSSR.Ser.Khim., 204A, **(1971)**.
[17] V.A.Davankov, S.V.Roghozim, A.A.Kurganov, L.Ya.Zhuchkova; J.Inorg.Nucl.Chem., **37**, 369 **(1975)**.
[18] V.I.Simen, O.A.Weber; Croat.Chem.Acta., **38**, 161 **(1966)**.
[19] F.Basalo, Y.T.Chen; J.Am.Chem.Soc., **76**, 953 **(1954)**.
[20] B.Grgas, S.Nikolic, N.Paulic, N.Raos; Croat.Chem.Acta., **72**, 885 **(1992)**.
[21] B.W.Clare, C.T.Supuran; J.Pharm.Sci., **83**, 768 **(1994)**.
[22] N.Bodor, Z.Gabanyi, C.K.Wang; J.Am.Chem.Soc., **111**, 3783 **(1989)**.
[23] B.W.Clare; Aust.J.Chem., **48**, 1385 **(1995)**.
[24] P.G.De Benedetti, M.C.Menziani, M.Cocchi, C.Frassineti; Quant.Struct.Act.Rela., **6**, 51 **(1987)**.
[25] H.R.Horton, L.A.Moran, R.S.Ochs, J.Rawn, K.G.Scrimgeour; Principles Of Biochemistry, 2nd ed. Prentice Hall, New Jersey, P. 62 **(1996)**.
[26] J.J.P.Stewart; J.Comp.Chem., **10**, 209 **(1989)**.
[27] M.C.Flanigan, A.Komornicki, J.W.Mecver; Modern Theoretical Chemistry, Plenum Press, NewYork, **8**, **(1977)**.
[28] M.J.S.Dewar, H.S.Rzepy; J.Am.Chem.Soc., **100**, 777 **(1978)**.
[29] R.G.Parr, R.A.Donelly, M.Levy, W.E.Palke; J.Chem.Phy., **68**, 3801 **(1978)**.
[30] C.Hansch, A.Leo, S.H.Unger, K.H.Kim, D.Nikaitani, E.J.Lien; J.Med.Chem., **16**, 1207 **(1973)**.
[31] R.G.Parr, L.V.Szentpaly, S.J.Liu; J.Am.Chem.Soc., **121**, 1922 **(1999)**.
[32] J.P.P.Stewart; J.Comp.Chem., **10**, 209 **(1989)**.