Asian Journal of Chemistry Vol. 20, No. 2 (2008), 1148-1152

Predicting Electrophoretic Mobility of Amino Acids and Small Peptides using Computational Descriptors

A. JOUYBAN*, H. AHMADI† and S. SOLTANI‡
Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz 51664, Iran
Fax: (98)(411)3344798; E-mail: ajouyban@hotmail.com

A simple two parameters linear model was proposed to calculate absolute electrophoretic mobility of amino acids and small peptides. The accuracy of the model was evaluated using percentage deviation (PD) of calculated and experimental values and the mean PD (± SD) was 4.8 (± 5.6) %.

Key Words: Capillary zone electrophoresis, Electrophoretic mobility, Amino acids, HyperChem descriptors, Prediction.

INTRODUCTION

Analytical and physico-chemical properties of amino acids provide interesting information for the analytical and biological scientists. Amino acids are characterized by the presence of hydrophilic groups (*i.e.* $–COO^-$ and $–NH_3^+$) and hydrophobic backbone (-R-) with a number of polar groups for some amino acids. The exsitance of charge on these analytes make them suitable for separation methods like capillary electrophoresis. Their oligomers and polymers among monomers' electrophoretic mobilities could be used in analytical method development stage using capillary electrophoresis. To develop an analytical method to separate the analytes using capillary electrophoresis, the common method is the trial and error and depends on analyst's intuition. It takes time and is also costly. The electrophoretic mobility of analytes under investigation is the most important parameter governing the separation efficiency and possibility of a successful separation using capillary electrophoresis. Any computational method for predicting the mobility of analytes could provide valuable information for the analysts. To the best of our knowledge, no pure predictive method is available so far, however, using developed mathematical models, prediction using minimum number of experimental data points are possible[1]. The aim of this communication is to present a simple least squares model using computational physico-chemical properties of the analytes, calculated based on their chemical structure of analytes by HyperChem software[2].

†Drug Applied Research Center, Tabriz University of Medical Sciences, Tabriz 51664, Iran.
‡Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

Based on classical theories, the electrophoretic mobility of mono-charged analytes is a function of hydrodynamic and dielectric functions as:

$$\mu_{0,i} = f\left(f_{i,\,hydrodynamic}, f_{i,\,dielectric}\right) \quad (1)$$

In the simplest mathematical form, eqn. 1 could be written as:

$$\mu_{0,i} = A_0 + A_1 f_{i,\,hydrodynamic} + A_2 f_{i,\,dielectric} \quad (2)$$

where A_0-A_2 are the model constants computed using linear least squares method. Different parameters have been used to represent hydrodynamic and dielectric frictions and the most known parameters are size and ionization parameters (pK values) developed by Fu *et al.*[3] presented by a non-linear model as:

$$\mu_{0.i} = \frac{C_0}{V^{C_1} + C_2 pK} \quad (3)$$

where C_0-C_2 are the model constants computed using non-linear least squares method, V is the van der Waals molecular volume and pK is –log of acid/base dissociation constants. Accuracy of equation 3 has been evaluated using mobility data of mono-amines and the produced percentage error was *ca.* 5 %[3].

Determinations of pK values of analytes are not straight forward for all analytes under investigation. As an example, it is somehow confusing when the analytes possessing multi-functional groups (such as amino acids) are considered and this is a major limitation of eqn. 3. In this work, two computational parameters are proposed for mathematical representation of mobility of amino acids. The molar volume of the analytes computed by HyperChem was converted to $V^{2/3}$ and assumed as a function of size of the analytes[1]. By assuming spherical anlaytes, it could be considered as effective surface area (SA) which is a representative of hydrodynamic function. The next term considered as dielectric function is the dipole moment (DM) of the anlaytes computed using the HyperChem software. It should be noted that from physico-chemical properties computed by the software, DM is the most relevant property for representing dielectric function in capillary electrophoresis. The proposed terms were replaced to eqn. 2 as:

$$\mu_{0,i} = B_0 + B_1 SA + B_2 DM \quad (4)$$

where B_0-B_2 are the model constants.

Experimental absolute mobility data of analytes taken from the work of Wronski[4]. The 2D structure of each compound was drawn and converted to 3D using HyperChem 7.0[2] and pre-minimized by Polak-Ribiere geometry optimization using MM^+. The structures were found by MM^+, used as the starting point for re-minimization by Polak-Ribiere optimization using AM_1 semi-empirical and also quantum mechanical methods. The complete energy optimized molecules were used to compute molecular descriptors, *i.e.* molar volume (V) and dipole moment.

The accuracy of the theoretically calculated mobilities was then examined with respect to the percentage deviations (PD), which were computed from the expression:

$$PD = 100\left(\frac{|\text{calculated} - \text{observed}|}{\text{observed}}\right)$$

The mean of PD is then calculated as an overall criterion for the comparison of the models.

Table-1 showed a list of amino acids, absolute electrophoretic mobilities, the computational parameters of the analytes and PDs of the numerical analyses. All data points were used in the training process of the eqn. 3 and 4 and the back-calculated mobilities were compared with the corresponding experimental values. The PD values were shown in Table-1. The mean of PD values (± SD) are 4.8 (± 5.6), 4.4 (± 5.5) and the mean difference is not statistically significant (paired t-test, $p > 0.48$). The least square based models and their computed model constants are:

$$\mu_{0.i} = \frac{452.314}{V^{0.450} + 0.029pK} \quad (5)$$

$$\mu_{0,i} = 43.476 - 0.264SA + 0.257DM \quad (6)$$

TABLE-1
ABSOULTE ELECTROPHORETIC MOBILITY (μ_0 10^{-9} $m^2 V^{-1} s^{-1}$) OF AMINO ACIDS TAKEN FROM A REFERENCE (4). THE MOLAR VOLUME (V), DIPOLE MOMENT (DM) COMPUTED USING HYPERCHEM, ACID DISSOCIATION CONSTANTS (pKa) OF ANLYTES AND PERCENTAGE DEVIATION (PD) OF EQNS. 5 AND 6

Analyte	μ_0	V	DM	pKa	PD eqn.5	PD eqn.6
Ala	31.0	333.03	1.29	9.9	5.3	0.5
AlaAla	27.0	535.58	3.25	8.5	1.9	0.3
AlaAlaAla	22.2	735.92	3.62	8.3	3.6	3.1
AlaAsn	25.5	615.70	4.38	8.5	2.4	0.0
AlaGly	28.8	481.06	3.18	8.4	3.5	2.4
AlaGlyGly	25.0	635.29	1.89	8.3	1.8	2.2
AlaLeu	23.9	676.63	3.42	8.5	0.2	0.4
AlaLeuGly	21.3	808.63	3.23	8.3	3.5	0.3
AlaMet	24.2	692.23	2.12	8.5	2.4	3.5
AlaPhe	23.9	741.14	3.38	8.5	4.1	5.0
AlaSer	26.2	548.04	3.80	8.3	0.1	2.3
AlaVal	25.2	623.93	3.55	8.5	1.8	0.3
Arg	32.1	589.39	14.56	9.1	21.0	10.4

Analyte	μ_0	V	DM	pKa	PD eqn.5	PD eqn.6
b-AlaHis	24.4	654.75	8.50	9.7	0.9	5.7
Cys	29.0	387.20	2.88	8.6	5.5	4.3
Glutamic acid	28.9	460.38	1.64	4.4	1.5	2.5
Glutamine	27.0	471.13	4.86	4.3	4.4	6.6
GlyAla	28.8	466.87	3.47	8.4	2.3	1.0
GlyAsn	27.5	557.10	2.73	8.4	5.4	4.3
Glycine	33.7	281.77	1.47	9.8	4.3	3.4
GlyGly	31.5	417.56	3.41	8.4	6.1	5.9
GlyGlyGly	26.1	553.29	3.74	8.1	0.1	2.2
GlyGlyIle	21.9	742.25	3.95	8.1	4.6	4.3
GlyGlyLeu	21.9	743.35	4.11	8.1	4.5	4.4
GlyGlyPhe	21.9	819.40	3.85	8.0	0.1	2.6
GlyGlyVal	22.6	694.07	4.31	8.1	4.5	5.7
GlyHisGly	22.5	759.24	13.37	8.3	0.8	11.1
GlyIle	25.2	595.82	3.77	8.4	0.2	2.2
GlyLeu	25.1	615.46	3.54	8.4	0.8	0.8
GlyLeuAla	21.1	805.01	4.52	8.3	4.7	3.2
GlyLeuTyr	21.0	1039.55	4.60	8.4	6.2	16.6
GlyPhe	24.7	662.03	3.40	8.3	2.4	1.7
GlyPhePhe	19.7	819.40	3.85	8.2	11.3	8.3
GlyPro	27.8	536.39	4.38	8.8	4.8	2.2
GlyProAla	22.5	735.26	4.40	8.5	2.2	2.6
GlySer	28.1	496.25	4.12	8.4	2.5	0.3
GlyThr	26.3	542.24	4.73	8.3	0.2	3.3
GlyTrp	23.6	727.85	2.44	8.4	2.1	3.7
GlyTyr	19.7	684.17	4.11	8.2	20.6	22.0
GlyVal	26.0	564.20	3.47	8.4	0.5	1.4
i-Leucine	26.7	468.37	1.34	9.8	5.1	4.6
Leucine	26.7	474.71	1.39	9.7	4.5	4.1
LeuGlyGly	21.5	763.37	1.89	8.0	5.2	1.8
LeuGlyPhe	19.3	1039.70	1.87	7.9	2.2	13.0
Phe	26.1	726.05	3.24	9.3	11.5	12.0
Proline	29.5	400.24	2.04	10.6	1.9	0.7
Serine	33.6	349.63	2.60	9.3	4.8	7.5
Threonine	30.9	401.64	3.15	9.2	2.7	3.1
Tryptophane	25.4	615.66	2.81	9.6	2.1	1.2
Tyrosine	20.0	562.88	1.64	9.1	29.4	29.5
			Mean of PD		4.4	4.8

The model constants of both models possess physical meaning. As shown in eqn. 6 there is a reciprocal relationship between effective surface area and the mobility. The wider the effective surface area the slower is the mobility of the analyte. This is expected since the effective surface area term reflects the effect of hydrodynamic friction on mobility. It is also the case for dipole moment term which is a representative of dielectric function, as the more dipole moment the more polarity and therefore the faster mobility is expected. For eqn. 5, the minimum percentage deviation of 0.1 % was observed for GlyGlyPhe and the maximum percentage deviation of 29.4 % for Tyr. The corresponding percentage deviation values for eqn. 6 was 0.0 (*i.e.* < 0.05 %) for AlaAsn and 29.5 % for Tyr. Both models produced high percentage deviation for Tyr and this could suggest the possibility of an outlier point for Tyr data. By excluding this point, mean percentage deviations were 4.0 and 4.3 %, respectively for eqns. 5 and 6.

In conclusion, the proposed model provided reasonable accurate mobility predictions and its descriptors could be computed using HyperChem software and the required calculations are straightforward.

ACKNOWLEDGEMENT

The financial support from the Research Office, Tabriz University of Medical Sciences (Grant No. 5/4/2039) was gratefully acknowledged.

REFERENCES

1. A. Jouyban and E. Kenndler, *Electrophoresis*, **27**, 992 (2006).
2. HyperChem 7.0, Molecular Mechanics and Quantum Chemical Calculations Package, HyperCube Inc., Ontario, Canada (2002).
3. S. Fu, D. Li and C.A. Lucy, *Analyst*, **123**, 1487 (1998).
4. M. Wronski, *J. Chromatogr. A*, **657**, 165 (1993).

(*Received*: 1 February 2007; *Accepted*: 1 October 2007) AJC-5950