Journal of Chromatography, 630 (1993) 329-336
Elsevier Science Publishers B.V., Amsterdam

CHROM. 24 728

Protein and peptide mobility in capillary zone electrophoresis

A comparison of existing models and further analysis

Vincent J. Hilser, Jr.* and Gregory D. Worosila

Physical and Analytical Chemistry Department, Ciba-Geigy Corporation, Suffern, NY 10901 (USA)

Suzanne E. Rudnick

Department of Chemistry, Manhattan College, Riverdale, NY 10471 (USA)

(First received March 3rd, 1992; revised manuscript received November 2nd, 1992)

ABSTRACT

Capillary zone electrophoresis of peptide fragments from the tryptic digest of human recombinant insulin-like growth factor I (rhIGF-I) has been carried out and the observed mobilities used to compare the relative applicability of existing mobility models. In addition, the physical forces affecting electromigration have been systematically analyzed in order to more accurately describe the physical chemistry involved. Such an approach should further improve the ability to predict electrophoretic mobility in capillary zone electrophoresis.

INTRODUCTION

Capillary zone electrophoresis (CZE) is rapidly becoming a major laboratory method for the separation, analysis, and characterization of biomolecules. Several recent studies have focused on deriving a correlation between certain physical attributes of proteins and peptides and the observed electrophoretic mobilities in CZE. Such correlations would facilitate the prediction of electrophoretic mobilities, and thus the design of optimum experimental conditions.

Several treatments have been the subject of recent studies [1–4]. Grossman et al. [1] have proposed a semi-empirical model which correlates charge and polymer number to electrophoretic mobility [1]. In another study, Rickard *et al.* [2] used charge and molecular mass to establish a similar correlation. We have applied these treatments to peptides obtained in our laboratory from the tryptic digestion of rhIGF-I (human recombinant insulin-like growth factor I), in order to determine which approach best predicts mobility in capillary electrophoresis. We have also investigated a more accurate description of the mobility using a systematic analysis of the physical forces affecting electromigration.

THEORY

Electrostatic attraction between the cathode and positively charged peptides is the primary driving

Correspondence to: G. D. Worosila, Physical and Analytical Chemistry Department, Ciba-Geigy Corporation, Suffem, NY 10901, USA.

* Present address: Graduate Program in Biology, The Johns Hopkins University, Seeley Mudd Hall, Baltimore, MD 21218, USA.

0021-9673/93/$06.00

force in CZE at low pH. While electrostatic attractions will be affected by changes in pH, ionic strength, temperature, and dielectric constant. experimental conditions have been developed to minimize changes in these solvent characteristics during CZE separations. The pH and ionic strength can be easily controlled through the choice of buffer and its concentration. Temperature is also controlled by modern CZE instrumentation.

In CZE, the net velocity of an individual peptide will vary as a direct function of charge and as an inverse function of size. This is demonstrated in Stokes' equation for frictional drag in an electric field.

$$\mu = \frac{q}{6 \pi r \eta} \quad (1)$$

where q is the net charge, η is the viscosity of the separation buffer, and r is the radius of spherical species. The Stokes radius is a parameter which defines size and is intended to describe the mobility of uniform spherical ions, thus making its application to peptide mobility correlations of limited use. Therefore, an analogous value must be derived which will describe peptide size in solution and which will mathematically define the physical impedance observed in CZE separations.

In their recent study, Grossman et al. [1] attempted to characterize the contribution of local dielectric to the electrophoretic mobility of a series of heptapeptides which contained one variable amino acid at the fourth position in the chain with lysine residues at positions three and five. This treatment accounted for the observed mobility differences between the peptides as a function of the degree of solvent perturbation by each substituted residue in the microenvironment of the lysine prosthetic groups. The resultant change in local dielectric was proposed to affect the degree of ionization and, ultimately, the charge of the lysines at pH 2.5. In order to significantly affect the local dielectric in the region surrounded by two charged lysine residues, a structure capable of solvent perturbation must be generated [5–9]. Since it has been demonstrated that small peptides are generally unable to form more than very limited secondary structure [10,11], it would seem highly unlikely that the mobility differences observed by Grossman et *al.* [1] can be attributed to charge differences caused by shifts in local dielectric.

We therefore proposed that a more discriminating treatment of the real size of each peptide should facilitate explanation of the observed mobility differences. Initial attention focused on an analysis of the means by which size is accounted for in existing peptide mobility correlations. In the model described by Grossman *et al.* [1], the size of each peptide is defined by the number of constituent amino acids. Every peptide is viewed as a homopolymer. From a physical perspective, this homopolymeric treatment of peptides represents a random coil model. It relies on a one-dimensional separation mechanism, with the root-mean-square end-to-end distance being the only size factor considered.

The use of molecular mass, as described by Rickard *et al.* [2] is, in principle, an improvement. Theoretically, a molecular mass model should account for volume because individual amino acids possess characteristic length-to-mass ratios and some discrimination would occur as a result. Separation according to this model involves a component of length, as the peptide backbone is considered identical for all amino acids. The mass will increase with the addition of each amino acid residue, and the average excess molecular mass remaining after subtraction of the backbone molecular mass would represent the width component. This approach assumes that the overall average shape of each peptide is disk-like.

Therefore, we wanted to identify a mechanism that would properly account for the size of each amino acid residue and also agree with existing solution models. Since CZE separations occur in the absence of inter-phase mass transfer, the nature of the "size phenomena" should be a function of the thermodynamic interactions between the solvent and the sample. If the effective size of a sample is a function of such solute-solvent interactions, then solvation should greatly affect the effective size of the solute. Since the water ordering is a function of surface area, this is extremely useful in defining the "effective size" in CZE separations. The greater the surface area, the greater the degree of water ordering that occurs, and, consequently, the larger the hydration sphere that surrounds the residue. Hence, it is the degree of hydration which would, theoretically, define "effective size".

Comparison between existing models should prove useful in describing the factors contributing

to the "effective size" of residues. In addition, some insight should be gained into the relative contributions of forces involved in CZE separations.

EXPERIMENTAL

Reagents and materials

rhIGF-I was obtained from Ciba-Geigy Pharmaceuticals Division. A single batch of rhIGF-I was used to generate peptide fragments, and was chromatographically determined to have a purity > 98% by protein mass. HPLC-grade acetonitrile (Fisher, Pittsburgh, PA, USA) was used for HPLC analysis of tryptic fragments, optima-grade acetonitrile (Fisher) was utilized for all amino acid analysis experiments, and certified trifluoroacetic acid (TFA) (Fisher) was used for all HPLC analysis. In addition reagent-grade iodoacetamide (Sigma, St. Louis, MO, USA), gold label triethylamine (Aldrich, Milwaukee, WI, USA), and sequanal-grade phenylisothiocyanate (PITC) and constant boiling HCl (Pierce, Rockford, IL, USA) were used. All other chemicals used were analytical grade and were not further purified. Distilled, deionized Milli-S-Q water (Millipore, Bedford, MA, USA) was utilized in all experimentation. Trypsin, in various treated forms, was obtained from three sources (Sigma, Pierce, and ICN, Cleveland, OH, USA) and was used without further purification.

Instrumentation

Reversed-phase HPLC. All chromatographic separations were performed with a system composed of a Waters (Milford, MA, USA) 600E gradient pump and WISP Model 712 autosampler with a refrigeration unit, a Kratos (Foster City, CA, USA) Spectroflow 783 UV detector at 214 nm, and a Dionex (Sunnyvale, CA, USA) Eluant Degas Module. Data collection was performed with a Nelson (Danbury, CT, USA) 760 Series interface and a modified version of Nelson Analytical Software.

Capillary electrophoresis. CZE was performed on an Applied Biosystems 270A Capillary Electrophoresis System with a fused silica capillary (122 cm × 100 μm I.D.). Ultraviolet detection at 200 nm was used for peak analysis. Data collection and peak processing were performed as described for the HPLC analysis.

Amino acid analysis. Peptides were identified by amino acid analysis. The Waters Pica-Tag work station was utilized for the gas-phase hydrolysis of each peptide fragment and subsequent generation of phenylisdthiocyanate (PITC)-derivatized amino acids. Separation of derivatized amino acids, data collection, and peak processing were performed as described for HPLC.

Methods

Trypsin digestion of rhIGF-I. The hydrolysis conditions utilized were a modification of those described by Worosila [12]. An aliquot of 10 μl of trypsin solution (10 mg/ml in 0.1 m*M* HCl) was added to the rhIGF-I solution (250 μl of a 10 mg/ml solution in water) initially and again after 3 h [trypsin–rhIGF-I (1:25)]. After an 18-h incubation at 37°C, the reaction was quenched with 100 μl of 10% (v/v) trifluoroacetic acid (TFA) in water.

Reversed-phase HPLC. Reversed-phase chromatographic analyses of rhIGF-I tryptic digests were performed using a Vydac Protein & Peptide C_{18} column (15.0 cm × 4.6 mm I.D.) (Vydac, Hesperia, CA, USA) at ambient temperature. The flow-rate was 0.8 ml/min. Mobile phases consisted of (A) 0.1% TFA in water and (B) 0.08% TFA in acetonitrile–water (80:20). A gradient was employed which ran according to the following program: 100% A at 3 min, 65% A at 38 min, 0% A at 55 min. Final gradient conditions were kept for 5 min before being returned to the initial conditions. A 15-min equilibration time was utilized between each run. The injection volume utilized for optimum resolution was 20 μl, and the injection volume for fraction collection was 30 μl.

Peak identification of peptides. Peptides were identified by amino acid analysis. HPLC fractions were pooled from ten runs and lyophilized in acid washed vials. Peak purity was determined chromatographically using the gradient elution described for peak analysis. The lyophilisate was then redissolved in 200 μl of distilled water and 30 μl used for amino acid analysis using a modification of the Waters Pica-Tag PITC method.

Capillary zone electrophoresis. CZE was performed on all peptide peaks isolated from reversed-phase chromatography. Samples were dissolved in 10 m*M* sodium citrate, pH 2.5 to an approximate concentration of 0.5 mg/ml, as determined by calculation of recovery from the reversed-phase separa-

tions. The CZE mobile phase utilized was 20 m*M* sodium citrate, pH 2.5. CZE was performed at a temperature of 30°C, with a constant voltage of 30 kV. Samples were introduced by a l-s vacuum injection onto a 122-cm capillary (99.7 cm to the detector). A proprietary mobility standard was utilized to facilitate the calculation of the peptide mobilities. Electrophoretic mobilities of samples were calculated according to the equation:

$$\mu = \left(\frac{L_D L_t}{V}\right)\left(\frac{1}{t} - \frac{1}{t_s}\right) + \mu_s \quad (2)$$

where:

L_D = length of capillary to detector, in cm
L_t = total length of capillary, in cm
V = system voltage
t = retention time of sample peak
t_s = retention time of standard peak
μ_s = mobility value for standard peak

Charge calculations. All peptide charges were calculated with the Henderson-Hasselbach equation using the pK_a values in ref. 2.

RESULTS AND DISCUSSION

Earlier work with peptides to determine a mobility model in CZE was performed using synthesized peptides [1]. As discussed previously, a model was derived which utilizes the polymer number of the sample peptides to define "size". Mathematically, this was justified by combining the classic mobility (eqn. 1) with the root-mean-square end-to-end distance equation to yield the semi-empirical relationship

$$\mu \propto \frac{\ln (q + 1)}{p.43} \quad (3)$$

The charge term, $\ln (q + 1)$, in the equation was obtained from a systematic' analysis of charge changes within a system of identically sized peptides. The deviation from the square root function for the polymer number (N) was empirically determined to give the most linear relationship. The mobility relationship which was derived from molecular mass (M_r) values [2] was determined entirely from empirical observations. The molecular mass-dependent equation was determined to be

$$\mu \propto \frac{q}{M_r^{2/3}} \quad (4)$$

One obvious difference between the two equation forms is the natural log dependence for the charge ***(q)*** in eqn. 3. Since this function was determined by systematic analysis, we predicted that the polymer-based equation would result in a more accurate treatment of charge effects. With respect to the "size" component of the samples, both equations represent a different approach to the physical impedance imposed on the samples. Ideally, linearity would be achieved with the polymer equation if (1) all amino acids were the same size, (2) the peptide backbone were fluid, and (3) there were no steric interactions between different parts of the peptide chain. The physical model for this equation assumes only minor contributions of the width dimension to the overall size of the amino acids. The molecular mass equation represents an attempt to account for the second dimension, width. Although this model does not take into account solvation properties and differences between chemical moieties, it does theoretically account for differences in mobility between peptides with identical polymer numbers.

A valid starting point for an analysis of mobility correlations is a comparison between the two published empirical models [1,2]. A comparison of the predictive ability of each equation can be used to assess the relative contribution of charge and size toward mobility. In addition, analysis of deviant peptides should also facilitate the recognition of contributing forces. In order to generate peptides for our use, trypsin was used to cleave rhIGF-I. In addition, some non-tryptic fragments of rhIGF-I were obtained and used in this work. The primary sequence of rhIGF-I is shown in Fig. 1, along with the predicted tryptic cleavage sites and the tryptic fragments obtained. The peptide fragments isolated from the tryptic digestion of rhIGF-I were identified by amino acid analysis and eqn. 2 was used to calculate the respective electrophoretic mobilities. To avoid variations due to drift, mobilities were calculated based on that of a proprietary standard (μ_s = 3.95 10^{-4} cm^2/Vs). In addition, the charge of each fragment was calculated and the molecular mass and polymer number determined. However, as discussed by Rickard ***et al.*** [2], the assignment of pK_a values is critical in charge calculations. Rickard *et* ***al.*** [2] compared pK_a values of free amino acids with values obtained from non-electrophoretic

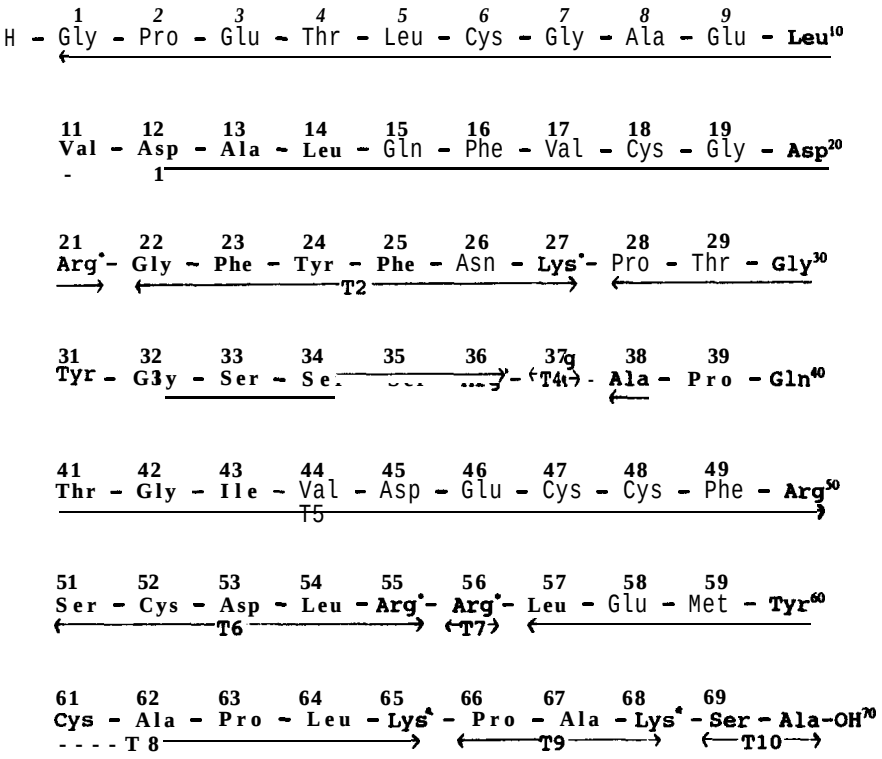

Fig. 1. The predicted tryptic cleavage sites of rhIGF-I. Arrows with (*) represent cleavage sites at the carboxyl terminals of Arg and Lys residues. Tryptic fragments are labeled according to their position within the rhIGF-I sequence.

measurements. These adjusted pK_a values improved their linearity and were used in their correlations. As our experiments were done at pH 2.5, ionization constants for the carboxyl terminal residues and for aspartic and glutamic acid must be accurately characterized to assure correct charge assignments. To determine the relative accuracy of the two mobility equations, we felt it necessary to use identical charge values for each. Therefore, linearities were developed for both functions using pK_a values obtained from the free amino acids (AAp*K* values) and from the adjusted values (Pp*K* values) of Rickard *et al.* [2]. Table I lists the data used to plot the mobilities of the individual peptide fragments in eqns. 3 and 4.

When the molecular mass function of eqn. 4, using both AAp*K*- and PpK-derived charges, was plotted against the corresponding mobility for each peptide, a linear trend was established (Figs. 2A and B). The correlation coefficients indicated marginally linear relationships (r^2 = 0.870 and 0.851, respectively). When the same mobility values were plotted against the polymer function in eqn. 3, a linear trend was again established for functions using AAp*K*- and PpK-derived charges (Figs. 3A and B). The correlation coefficients for these functions (r^2 = 0.908 and 0.878, respectively) demonstrate an improvement over the molecular mass-derived functions.

Interestingly, when each function is examined separately, the correlations made with the AAp*K*-derived charges differed very little from correlations made with the Pp*K* charges. However, these small differences in the correlation coefficients were accompanied by large changes in the distribution of the data points from AAp*K*- to PpK-derived functions, as seen by comparing Figs. 2A and B or Figs. 3A and B. This indicates that the degree of ion-

TABLE I

VALUES USED IN CZE MOBILITY CALCULATIONS

Peptide fragment	Designation	Charge (q) AApK	Charge (q) PpK	Polymer number (N)	Molecular mass	Mobility (μ) ($\times 10^4$)
Leu 57-Tyr 60	A	0.31	0.82	4	554.7	1.31
Gly 22-Tyr 24	B	0.33	0.83	3	385.4	1.58
Gly1–Leu 14"	C	0.31	0.71	14	1443.6	0.90
Gln 15–Phe 16	D	0.43	0.83	2	293.3	1.74
Gly 22-Phe 25	E	0.43	0.83	4	532.6	1.41
Gly 1–Arg 21"	F	1.06	1.61	21	2305.6	1.18
Val17–Arg 21"	G	1.16	1.73	5	604.6	2.25
Ser 51-Arg 55"	H	1.16	1.73	5	648.7	2.23
Leu 57-Lys 65"	I	1.30	1.82	9	1123.4	1.91
Arg 56–Tyr 60	J	1.31	1.82	5	710.9	2.14
Pro 66–Lys 68	K	1.32	1.83	3	314.4	2.60
Cys 61-Lys 65"	L	1.32	1.83	5	586.7	1.86
Phe 25-Tyr 31	M	1.33	1.83	7	825.9	1.97
Arg 37-Arg 50"	N	2.13	2.73	14	1706.9	1.94
Arg 56–Lys 65"	0	2.31	2.82	10	1279.6	2.31
Leu 57-Ala 70"	P	2.39	2.82	14	1577.9	2.03
Phe 25-Arg 36	Q	2.20	2.83	12	1300.4	2.21
Arg 56-Lys 68"	R	3.30	3.82	13	1576.0	2.46

[a] These fragments contain a carboxymethylated cysteine. The molecular masses of the fragments reflect this.
[b] These fragments resulted from the hydrolysis of rhIGF-I

ization at pH 2.5 is not the primary contributing factor to the deviation from linearity for either function.

A further analysis of the charge and its contribution to both eqn. 2 and eqn. 3 was performed. Inspection of Table I reveals three distinct groups of charge values; Group I (0 to + 1.0), Group II (+ 1 .O to + 2.0), and Group III (+ 2.0 to + 3.0). At pH 2.5, peptides in Group I contain one protonated amino terminal group contributing a charge of + 1.0, and one partially ionized carboxyl group. In addition to the amino and carboxyl terminal

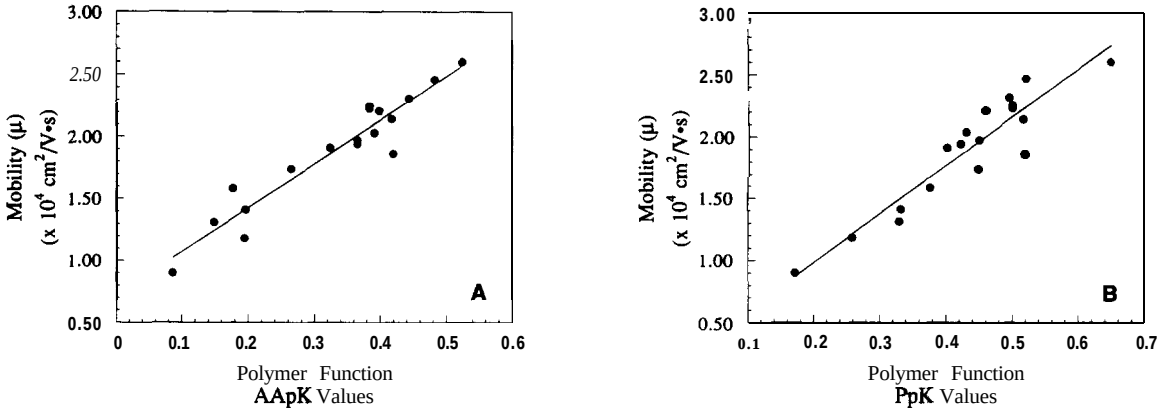

Fig. 2. Plots of the molecular mass function (eqn. 4) for the peptides in Table I *versus* the experimentally observed mobilities. (A) This plot was obtained using pK_a values of free amino acid residues (AApK), and (B) used the adjusted pK_a values (PpK) described in ref. 2.

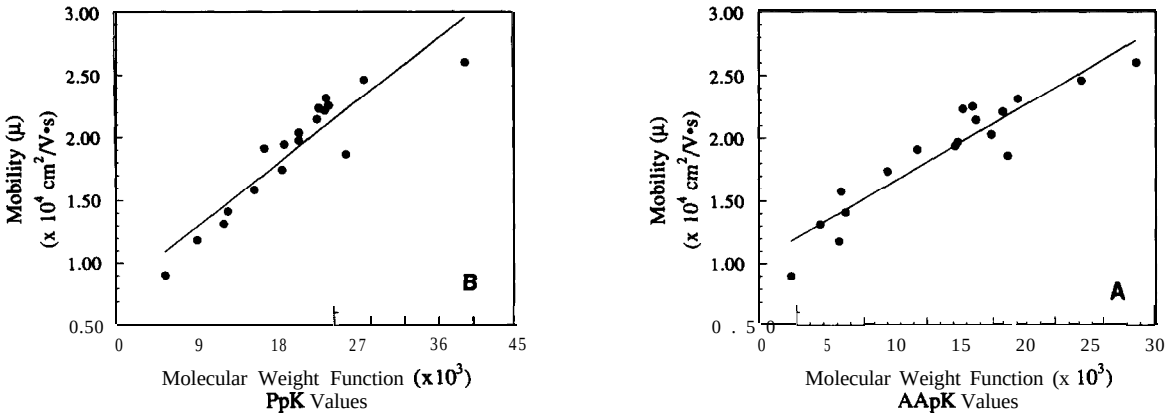

Fig. 3. Plots of the polymer function (eqn. 3) for the peptides in Table I versus the experimentally observed mobilities. (A) This plot was obtained using pK_a values of free amino acid residues (AAp*K*), and (B) used the adjusted pK_a values (Pp*K*) described in ref. 2.

groups, peptides in Groups II and III contain basic prosthetic groups, adding a charge of + 1 .O for each group. By generating linearities for each respective charge group, the relative contributions were established. Table II lists the relative contributions for both functions.

As determined by the r^2 values, there was no charge related variability in any of the functions. Further inspection of the trends in Table II, however, reveals that in every instance the linearity is substantially improved when adjusted pK_a values (Pp*K*) are used for the correlation.

In addition to charge-related trends in the mobility correlation, an analysis of size was performed using the eqn. 3. Unlike the trend seen with charge, there was no statistically significant distribution of peptides based on any component of size in the polymer equation. Efforts were made to incorporate surface area data, volume, hydrophobicities, and partition coefficients into the equation to improve linearity. While minor improvements were seen in the *r* values obtained using these altered equations, the general distribution of points remained virtually identical to the original form.

CONCLUSIONS

Both the homopolymer function (eqn. 3) and the molecular mass function (eqn. 4) represent semi-empirical correlation models for mobility. By utilizing an experimental situation, mobility analysis of the tryptic fragments of rhIGF-I, the relative ap-

TABLE II

CORRELATION COEFFICIENTS (r^2) FOR COMPARISON OF MOBILITY EQUATIONS

Groups (charge values)	Molecular mass function		Polymer function	
	AAp*K* (Fig. 2A)	pp*K* (Fig. 2B)	AAp*K* (Fig. 3A)	pp*K* (Fig. 3B)
I (0 to +1.0)	0.883	0.991	0.875	0.979
II (+ 1.0 to +2.0)	0.729	0.765	0.824	0.864
III (+ 2.0 to + 3.0)	0.869	0.949	0.862	0.942
Overall function	0.870	0.851	0.908	0.878

plicability of both treatments was assessed. In this treatment, all available experimental data were used. All data points were treated equally, and no data points were eliminated because we found no suitable criteria which would justify differential treatment. Using these data and this treatment, the homopolymer function [1] displays better linearity than the molecular mass function [2]. In addition, the adjusted pK_a values (PpK) appear to be the more accurate, and therefore some other factor or combination of factors must be the source of the deviations from linearity observed. We believe that a large percentage of that deviation is due to the inadequate treatment of size in both equations. For a size function to be ideal it must define the space a molecule occupies in solution, the ability of the solute to hydrogen bond with the solvent, the degree of ordering that the solute imposes on the solvent, and the effect of charge on these parameters. We therefore believe that size must ultimately be defined by a more complex thermodynamic function which involves entropic and enthalpic contributions of both polar and apolar groups. Such a treatment should further improve our ability to predict electrophoretic mobility in CZE.

NOTE ADDED IN PROOF

Another mobility model has been proposed and appears in refs. 3 and 4. The implications of the referenced model are not fully explored in this manuscript, since it appeared in print after the submission of this paper.

REFERENCES

1 P. D. Grossman, J. C. Colburn and H. Lauer, *Anal. Biochem.*, 179 (1989) 28.
2 E. C. Rickard, M. M. Strohl and R. G. Nielsen. *Anal. Biochem.*, 197 (1991) 197.
3 B. J. Compton, *J.* Chromatogr., 559 (1991) 357.
4 B. J. Compton and E. A. O'Grady, *Anal.* Chem., 63 (1991) 2597.
5 S. M. Parsons and M. A. Raftery. *Biochemistry,* 11 (1972) 1623.
6 D. E. Schmidt, Jr. and F. H. Westheimer, *Biochemistry,* 10 *(1971) 1249.*
7 J. G. Voet, J. Coe, J. Epstein, V. Matossian and T. Shipley, *Biochemistry, 20* (1981) 7182.
8 S. Karplus, G. H. Snyder and B. D. Sykes, *Biochemistry, 12 (1973) 1323.*
9 F. A. Johnson, S. D. Lewis and J. A. Shafer, *Biochemistry, 20 (1981) 44.*
10 H. J. Dyson, M. Rance, R. A. Houghten, R. A. Lerner and P. E. Wright, *J. Mol. Biol.*, 201 (1988) 161.
11 H. J. Dyson, M. Rance, R. A. Houghten, P. E. Wright and R. A. Lerner, *J.* Mol. *Biol.*, 201 (1988) 201.
12 G. Worosila, *Ph. D. Thesis,* Rutgers University. New Brunswick, NJ, 1985, p. 45.