*Journal of Chromatography*, 559 (1991) 357–366
Elsevier Science Publishers B.V., Amsterdam

CHROM. 23 454

# Electrophoretic mobility modeling of proteins in free zone capillary electrophoresis and its application to monoclonal antibody microheterogeneity analysis

BRUCE JON COMPTON*
*Bristol-Myers Squibb Co., Industrial Division, P.O. Box 4755, Syracuse, NY 13221-4755 (USA)*

## ABSTRACT

A semi-empirical model for describing the electrophoretic mobility of proteins in free solution is derived. Protein mobility is found to be influenced, as dictated by the Debye–Hückel–Henry theory, by protein valence, size and shape, and by solution ionic strength, pH, viscosity and temperature. Protein valence, the most important mobility determining parameter intrinsic to the protein, is calculated for a given pH from its amino acid content using the Henderson–Hesselbalch equation. Electrostatic charge suppression causes actual valence to be less than that calculated. To equate the two an experimentally determined proportionality constant ($F_Z$) is introduced. Consequently, $F_Z$ can be applied to the calculated valence and mobility–pH titration curve for a protein, resulting in the actual mobility of the protein at any given pH. The model further predicts that the molecular weight ($M$) dependency of mobility should be a continuous function of $M^{-1/3}$ to $M^{-2/3}$, depending on the magnitude of the protein molecular weight and buffer ionic strength under investigation. Many aspects of the model are demonstrated by its application to the resolution of immunoglobulin G isoelectrotypes, normally only resolved using isoelectric focusing.

## INTRODUCTION

Capillary zone electrophoresis (CZE) in free solution has proved useful for the analysis of non-volatile thermally labile molecules such as proteins. The method is characterized by a high theoretical plate count (the generation of greater than $10^5$ plates/min [1]) and, in the free solution mode, operational simplicity. Application for the resolution of immunoglobulin isoelectrotypes or other proteins which exhibit microheterogeneity would appear logical, and CZE has been used, for example, in work on the analysis of proteins such as human growth hormone (rhGH), tissue plasminogen activator and CD4, tryptic digest of rhGH [2,3], and transferrin variants [4].

A question that comes to mind when considering therapeutic applications is whether it would be possible to predict conditions under which CZE in free solution is capable of distinguishing between a native protein and a variant (or degradant) which differs from the native form because of, for example, multiple deamidation events (Gln to Glu or Asn to Asp). To address this question a general model describing protein electrophoretic mobility is needed.

0021-9673/91/$03.50 

One of the benefits of working with recombinant or well characterized proteins is that either their primary sequence or at least their amino acid content is known or easily obtained from commercial data bases. In principle, this sequence data can be used to determine a theoretical titration curve for the protein, which gives the calculated protein valence ($Z_c$) as a function of pH. $Z_c$ should always be greater than actual protein valence ($Z_a$) because of electrostatic charge suppression [5,6], but a direct proportionality is postulated to exist between the two [7–10]. This proportionality constant ($F_Z$), is expected to be dependent on both the nature of the protein and its environment.

The molecular weight ($M$) of a protein can also be calculated from the sequence data and with some assumptions about the partial specific volume of the protein, can be used to calculate an equivalent radius ($r_e$), which is then equated to the protein's Stokes radius ($r$) after the application of a correction factor such as a frictional ratio. With this information, a classic model based on the Debye–Hückel–Henry theory (DHHT, [11–14]) is employed to predict electrophoretic mobility ($u$) of the protein.

The rationale for this approach is historically well developed, since much work has been done confirming the usefulness of the DHHT for describing the electrophoretic mobility of amphoteric species such as proteins [15,16]. In this previous work the critical parameter of protein valence ($Z$) was obtained either by direct titration or by measurements of membrane potentials, while size parameters were obtained through diffusion measurements [7–16]. The use of $Z_c$ and $r_e$ rather than $Z$ and $r$ to calculate $u$ sacrifices accuracy but eliminates the need to measure $Z$ and $r$ experimentally, a task requiring considerable amounts of material and time.

## THEORY

### *Electrophoretic mobility model*

In electrolyte solution under unit field strength (1 V/m) the Debye–Hückle–Henry theory can be used to develop a general expression for protein mobility ($u$) such that

$$u = (Ze/6\pi nr)(1/1 + kr)\phi(kr) \tag{1}$$

where $Ze$, the product of the protein valence and $e$, the charge of an electron ($1.602 \cdot 10^{-19}$ C), gives protein net charge, $n$ is solution viscosity (0.000895 kg/ms, water at 25°C), $r$ is the Stokes radius (distance in m from the particle center to where solvent slippage occurs), $\phi(kr)$ is Henry's function [11], and $k$, the Debye parameter ($m^{-1}$), is $2^{1/2}Ne(e_0e'RT)^{-1/2}I^{1/2}$, where $N$ is Avogadro's number, $e_0$ is the fluid dielectric constant (78.54, water 25°C), $e'$ is permittivity of free space ($8.854 \cdot 10^{-12}$ $C^2/Jm$), $R$ is the gas constant (8.314 J/K mol), $T$ is absolute temperature, and $I$ is solution ionic strength ($mol/m^3$). $I$ is further defined as $\sum C_i/2z_i^2$, where $C_i$ is ion concentration ($mol/m^3$), and $z_i$ the valence of fluid ions. The magnitude of $\phi(kr)$ varies from 1.0 to 1.5 in a sigmoidal fashion as $kr$ increases from zero to infinity.

The Debye parameter ($k$) represents the reciprocal thickness of the ionic atmosphere surrounding the protein [17]. The expression $1/1 + kr$ indicates that the mobility of a protein has ionic strength ($I$) dependency, sometimes overshadowed in CZE by experimental complications [18]. This dependency is not a strict inverse

proportionality but can be expressed from eqn. 1 directly as

$$u = \frac{Ze\phi(kr)(e_0e'RT)^{1/2}}{6\pi nr(1 + 2^{1/2}Ner)I^{1/2}} \tag{2}$$

Eqn. 1 can be made more useful, but more approximate, by expressing the size dependency of mobility using protein molecular weight rather than radius, as discussed by Oncley [19], through the relationship

$$r_e = (3Mv/4\pi N)^{1/3}(f/fo) \tag{3}$$

where $r_0$ is equivalent radius (m), $v$ is protein partial specific volume and $(f/fo)$ is a frictional ratio discussed below. Substitution of eqn. 3 into eqn. 1 gives eqn. 4

$$u = \frac{K(1)Z}{K(2)\mathrm{M}^{1/3} + K(3)\mathrm{M}^{2/3}} \tag{4}$$

where terms $K(1) = e\phi(kr)$, $K(2) = 6\pi n(f/fo)(4\pi N/3v)^{-1/3}$ and $K(3) = 6\pi n(2^{1/2}Ne) \cdot (e_0e'RT)^{-1/2}(f/fo)^2(4\pi N/3v)^{-2/3}I^{1/2}$.

The equivalent radius is made equal to Stokes radius by the frictional ratio $(f/fo)$, such that $r_e(f/fo) = r$, which corrects for protein asymmetry and hydration [19]. The frictional ratio usually varies from 1.0 (ideal behavior) to 1.7 for globular proteins having non-spherical dimensions, to greater than 3 for a cylindrical protein such as myosin.

Eqn. 4 indicates that a protein's valence is its most important attribute with regard to determining its mobility. Molecular weight exhibits an inverse influence which varies continuously from $M^{1/3}$ to $M^{2/3}$ depending on the magnitude of the molecular weight and ionic strength [coefficient $K(3)$] range under examination.

*Calculation of protein valence*

As mentioned, one benefit of investigating therapeutic proteins is that they are usually well characterized with respect to their primary sequence, amino acid content and modifications (glycosylation, phosphorylation, etc.). With this information it is possible to calculate the protein's theoretical valence–pH titration curve using the Henderson–Hasselbalch equation. This gives its calculated net valence ($Z_c$) at a given pH. In this work the extended method of Sillero and Ribeiro [20] is used and the expression used to calculate $Z_c$ is as follows:

$$Z_c = \sum_{n=1-4} \frac{Pn}{1 + 10^{\mathrm{pH} - \mathrm{p}K(Pn)}} - \sum_{n=1-5} \frac{Nn}{1 + 10^{pK(Nn) - \mathrm{pH}}} \tag{5}$$

where pH refers to the buffer, $Pn$ and $Nn$ the integral number of each amino acid of that type, and p$K(Pn)$ and p$K(Nn)$ its respective ionization potential. The p$K$ values and the respective number of each charged amino acid type in native chimeric L6 are given in Table I.

TABLE I

LIST OF IONIZATION CONSTANTS OF AMINO ACIDS

| Amino acid type | | $pK_a$ | Charged amino acid content of cL6 |
|---|---|---|---|
| P1 | HIS | 6.4 | 22 |
| P2 | $tNH_2$ | 8.2 | 4 |
| P3 | LYS | 10.4 | 100 |
| P4 | ARG | 12.0 | 30 |
| N1 | tCOOH | 3.2 | 4 |
| N2 | ASP | 4.0 | 46 |
| N3 | GLU | 4.5 | 64 |
| N4 | CYS | 9.0 | 32 |
| N5 | TYR | 10.0 | 62 |

As mentioned in the introduction, the magnitude of $Z_c$ is expected to be greater then actual protein charge ($Z_a$) due to electrostatic charge suppression. Though the ionization constants used by Sillero and Ribeiro are taken from peptides, $Z_c$ is a simple sum of the individual charges of amino acids and its calculation does not account completely for electrostatic interactions which can modify an amino acid's (side chain or terminal) ionization potential by 2–3 units [5,6]. A similar bias has been described in the past for studies which attempted to equate protein valances obtained from titration, membrane potentials, and mobility measurements. Valence measurements from membrane potentials gave results consistent with mobility measurements while titration results overestimated protein valence [9].

Because of electrostatic considerations, the differences between $Z_c$ and $Z_a$ are most pronounced at electrolyte pH values at which the protein has an appreciable charge, *i.e.* far from the isoelectric point (p*I*) of the protein. Consequently, titration curves for proteins flatten out at pH values far from the p*I* of the molecule, at pHs where the proteins should still be gaining appreciable charge [5].

Some additional evidence for this discrepancy can be seen in the recent work of the derivation of an empirical expression for predicting peptide mobility as a function of sequence and size [21]. It was found that a non-linear relationship [$\ln(Z_c + 1)$] between calculated charge and mobility existed. This relationship must be valid since it empirically relates $Z_c$ to $u$. However, a contradiction arises since it is well documented that a direct relationship exists between $u$ and $Z_a$ as predicted from fundamental principles and determined experimentally by a variety of means, for such diverse entities as ions and particles. This discrepancy can be accounted for when one considers that the difference between $Z_c$ and $Z_a$ increases at pH values far from the p*I* of the peptide and that this difference was not accounted for when deriving the empirical expression.

The proportionality that exists between $Z_c$ and $Z_a$ for any given protein under a particular set of conditions such as solution pH [5,6], is expressed by $Z_c = Z_a F_Z$. Since $F_Z$ is independent of pH, it can be determined for one solution pH and, along with the theoretical titration curve and eqn. 4, be used to calculate $u$ for other pH conditions. This in turn allows optimization of conditions for resolving protein variants.

One problem associated with this approach is that in the calculation of $Z_c$ no consideration is given for variations in amino acid ionization potentials as a function of buffer type and ionic strength. These differences are factored into $F_Z$ and accordingly, a given value for $F_Z$ is expected to be valid only for the experimental conditions —*i.e.*, solution ionic strength, dielectric constant and temperature— under which it is determined.

## EXPERIMENTAL

CZE was conducted on a Bio-Rad HPE 100 system (Bio-Rad Labs., Richmond, CA, USA) with a 25 $\mu$m I.D. coated capillary, with 0.20 m between electrode reservoirs and 0.172 m from the injection reservoir to detector. Reagent-grade sodium orthophosphate buffer was used (Fisher Scientific, Pittsburg, PA, USA). Detection was by ultraviolet absorption at 200 nm. Sample concentrations were generally 100 $\mu$g/ml, made up with a diluent that gave a final sample solution of 10% ionic strength of the electrophoresis buffer used in any particular experiment. Most experiments were conducted with +12 000 V at the injector reservoir which, based on the lack of current variation during the initiation of a run, did not cause appreciable capillary heating. All samples were loaded by electromigration, the potential and duration of which varied with experiment. All experimental conditions were studied using triplicate sample introductions and reported as mean results of the main electrophoretic band. The use of coated capillaries reduced electroendoosmosis to non-measurable levels [22].

Immunoglobulin G (IgG) monoclonal antibody chimeric L6 (human-murine, estimated p*I* 10.2) was either produced at the Bristol-Myers Squibb Syracuse site or obtained from Oncogen (Seattle, WA, USA). Samples were stored at −70°C in phosphate buffered saline at 5 mg/ml. For use, samples were thawed on a rotary shaker at 32°C for 2 h, diluted appropriately, and analyzed as above. All other samples and materials were purchased from Sigma (St. Louis, MO, USA).

## RESULTS AND DISCUSSION

Fundamental models for calculating the electrophoretic mobility of proteins require charge and size information obtainable only through tedious membrane potential and diffusion studies. Because buffer type and ionic strength affect protein p*I*, these studies must be carried out under conditions identical to the electrophoretic studies [10,16]. This experimentally derived information often allows exact calculation of protein mobility from first principles [9,12], and as such is useful for demonstrating the validity of these principles. However, this approach is not practical for optimizing protein resolution in CZE, therefore, an effort has been made to use readily available sequence information to calculate protein valence and size. This results in a calculated valence–pH titration curve for the protein which, when corrected for by experimentally determined $F_Z$, allows calculation of a protein's actual valence and mobility at all pH values.

This approach has been developed specifically to address the CZE resolution of protein isoelectrotypes often encountered when developing therapeutic proteins such as antibodies and growth factors. This is a special case of a protein which exhibits microheterogeneity with regard to charge, but not size, presumably due to variations in

primary sequence, for instance, from protein degradation though deamidation, or alternatively in charged carbohydrate portions of the molecule [2–4,23]. In either case, this semi-empirical model is intended to allow quantitative prediction on how best to optimize a CZE resolution of isoelectrotypes.

The electrophoretic mobility model developed here has resulted in eqn. 4, which indicates that $u$ is directly proportional to protein charge $Z$ and inversely proportional to molecular weight in a continuous fashion from $M^{1/3}$ to $M^{2/3}$, depending on the magnitude of $M$ and the solution ionic strength ($I$). Protein charge is the most sensitive parameter governing mobility, and is directly controlled through adjustments to solution pH, such that achievement of resolution of similar sized native proteins in CZE is most practically brought about by controlling this parameter. Additional selectivity can be achieved through the specific interactions of various ions on specific proteins.

The calculated mobility and valence–pH relationship of a protein is obtained using eqns. 4 and 5 and values in Table I, examples for which are shown in Fig. 1 and 2 for IgG chimeric L6. Fig. 1 further illustrates the constant bias postulated between experimental mobility results and $u$ calculated using the DHHT and $Z_c$. This bias is accounted for by modifying $Z_c$ using $F_Z$; in this case $F_Z$, which is determined by directly ratioing calculated and measured mobility for a solution pH of 2.5, is 5.66. Once this correction is made to $Z_c$, mobility is seen to be directly proportional to protein valence. As previously mentioned, this bias between experimental and calculated mobility and direct proportionality between mobility and $Z$, as determined using titration and membrane potentials, has been previously demonstrated for a number of proteins.

The titration curves further indicate the appropriate choice of buffer pH for optimum mobility discrimination between isoelectrotypes, as shown in Fig. 2 for hypothetical desamido forms of cL6. The figure indicates that resolution of these

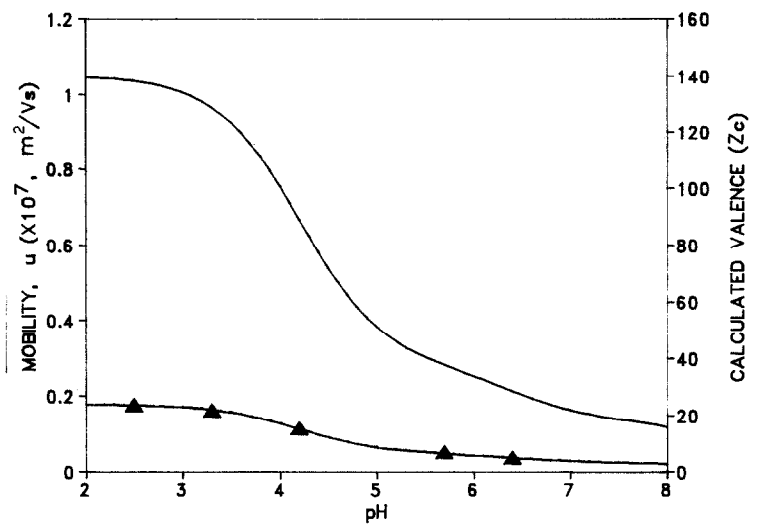


Fig. 1. Calculated (solid line) and experimental (triangles) mobilities for cL6. The calculated results were obtained from eqns. 4 and 5 using constants and parameters given in Table I. The calculated values, when corrected for a $F_Z$ of 5.66, gave the mobilities shown by the dashed line.

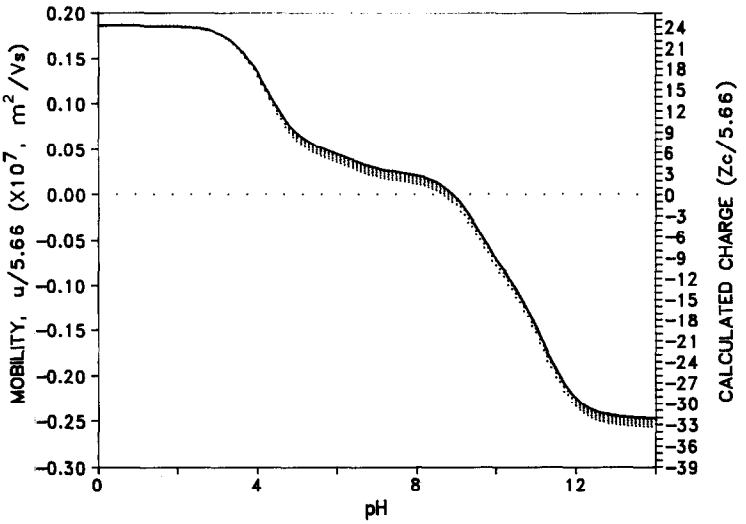


Fig. 2. Calculated mobility and valence titration curve for native and hypothetical deamidated cL6 from eqns. 4 and 5 and Table I. The solid line shows the mobility for the native form while each subsequent dashed line is the mobility for deamidated cL6 occurring at intervals of 5 (by the sequential substitution of 5 Glu for 5 Gln residues per curve).

degradants is best achieved at pH values between 4 and 8, or above 12. The expected time of migration ($t$) of native and deamidated cL6 under the experimental conditions can be calculated from $t = (L_t L_d)/Vu$, where $L_t$ and $L_d$ are total capillary length and length to detector, respectively.

The validity of this prediction is shown by the electropherograms in Fig. 3 for

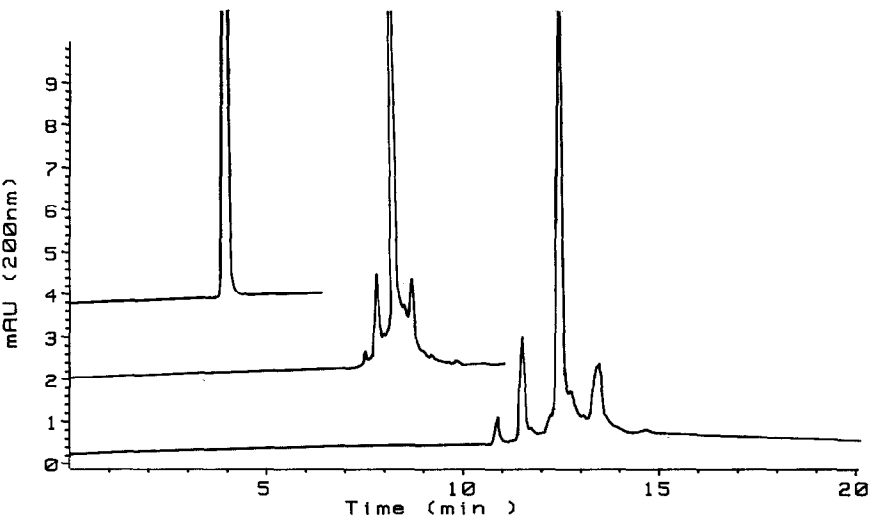


Fig. 3. Electropherograms of cL6 at (left-to-right) pH 2.5, 5.6 and 6.4 showing resolution of isoelectrotypes as predicted in Fig. 2.

native cL6. This information has been used to develop a method for determining the lot-to-lot reproducibility of various cL6 lots as well as how cL6 isoelectrotypes vary with production and storage conditions. This method is comparable to rapid isoelectric focusing and analytical ion-exchange chromatographic methods previously described for highly basic antibodies [23].

At pH values above 7, cL6 did not migrate sufficiently to be detected within 30 min. Electrolyte conditions above pH 6 were avoided to minimize hydrolysis of the capillary coating and electroendoosmosis. Fig. 3 also indicates that the valence–pH relationship can be further exploited such that at the p$I$ of the IgG isoelectrotypes, essentially infinite resolution occurs as the more acidic isoelectrotypes undergo charge reversal and migrate in an opposite direction to those with higher p$I$.

For an exact solution, the mobility model requires input of the protein partial specific volume ($v$), frictional ratio ($f/fo$), an indirect measure of protein asymmetry, and Henry's function $\phi(kr)$. The protein partial specific volume falls between 0.745 and 0.750 [19,24,25] for natural proteins and $f/fo$ is between 1.0–1.7 for globular proteins [19]. Literature values for IgGs indicate $v = 0.74$, irrespective of species origin [26]. The magnitude of $f/fo$ is taken here to be 1.0, though it has been noted from light scattering studies that IgGs are highly assymmetric [27]. For this work the magnitude of Henry's function $\phi(kr)$ is taken to be 1.05, as calculated from Henry's work [11]. As indicated by the notation for the function, $\phi(kr)$ varies as a function of solution ionic strength and protein radius.

Besides changing solution pH, two other means of controlling $u$, as seen in eqn. 4, are by varying fluid viscosity ($n$) and ionic strength ($I$); the former has a general effect on $u$ and the latter modifies the Debye–Hückel parameter and protein charge by changing both $F_Z$ and the protein's p$I$. While increasing fluid viscosity decreases $u$ by increasing the frictional factor of the protein and thus tends to accentuate differences in protein size, little enhancement of resolution is expected since the size dependency of $u$ is rather weak. For instance, as mentioned previously, various empirical models have been developed to explain electrophoretic results [21,28] and indicate a molecular weight dependency ranging from $M^{-1/3}$ to $M^{-2/3}$. The $M^{-1/3}$ dependency is an ideal lower limit for small molecules in low-ionic-strength buffer. Larger molecules in high ionic strength buffer will exhibit a $M^{-2/3}$ dependency as seen in Offord's model [28]. For intermediate sized molecules in medium-strength buffer, a $M^{-1/2}$ dependency is seen in Lauer's model. The transition from a $M^{-1/3}$ to $M^{-2/3}$ dependency is shown in Fig. 4 for various ionic strengths of aqueous solutions at 25°C, by showing how the ratio of the coefficients $[K(3)/K(2)]$ varies with these parameters. For any given experimental condition, eqn. 4 reduces to $u = K(4)ZM^{-n}$, where $n$ varies from 1/3 to 2/3 and $K(4)$ is an aggregate of $K(1–3)$.

In conclusion, the semi-empirical model presented here indicates that a proteins primary sequence data and $F_Z$ can be used to calculate $u$ for a protein at a variety of pH conditions. The model utilizes the observation that the calculated valence $Z_c$ of a particular protein is proportional to its actual valency $Z_a$ by $F_Z$. In this application $F_Z$ also incorporated the frictional ratio $f/fo$ which was taken to be unity. In reality, the model indicates that the mobility of a protein, and variants of that protein, at any given pH can be calculated directly from its theoretical titration curve if its mobility is measured at any one pH since mobility is directly proportional to protein valence.

One of the main benefits of the model is that it shows that the molecular weight

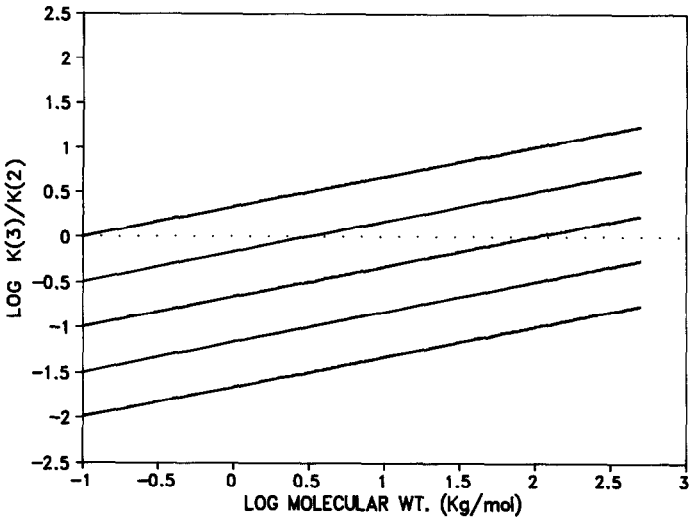


Fig. 4. Plot of the ratio of the $M^{-2/3}$ and $M^{-1/3}$ coefficients [$K(3)/K(2)$] from eqn. 4 showing the molecular weight dependency of $u$ as a function of both $M$ and $I$, at $I$ = 0.1, 1.0, 10, 100 and 1000 mol/m$^3$ (bottom-to-top).

influence on mobility is a complex function of both the magnitude of protein molecular weight and the buffer ionic strength. This reconciles differences in previous empirically derived models, which is a necessity if the model developed here is to be considered valid.

## REFERENCES

1 B. Nickerson and J. W. Jorgenson, *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 11 (1988) 553.
2 J. Frenz, S.-L. Wu and W. S. Hancock, *J. Chromatogr.*, 480 (1989) 379.
3 S.-L. Wu, G. Teshima, J. Cacia and W. S. Hancock, *J. Chromatogr.*, 516 (1990) 115.
4 F. Kilàr and S. Hjertén, *J. Chromatogr.*, 480 (1989) 351–357.
5 J. B. Matthew, *Ann. Rev. Biophys. Biophys. Chem.*, 14 (1985) 387.
6 C. Tanford and J. G. Kirkwood, *J. Am. Chem. Soc.*, 79 (1957) 5333.
7 J. Th. G. Overbeek, in H. Mark and E. J. W. Verweys (Editors), *Advances in Colloid Science*, Vol. 3, Wiley-Interscience, New York, 1950, pp. 97–134.
8 L. G. Longsworth, *Ann. N.Y. Acad. Sci.*, 41 (1941) 267.
9 G. S. Adair and M. E. Adair, *Trans. Faraday. Soc.*, 36 (1940) 23.
10 R. K. Cannan, A. Kibrick and A. H. Palmer, *Ann. N.Y. Acad. Sci.*, 41 (1941) 243.
11 D. C. Henry, *Proc. R. Soc. London, Ser. A*, 133 (1931) 106–124.
12 A. Tiselius and H. Svensson, *Trans. Faraday Soc.*, 36 (1940) 16.
13 G. J. Wei and V. A. Bloomfield, *Anal. Biochem.*, 101 (1980) 245.
14 R. A. Mosher, D. Dewey, W. Thormann, D. A. Saville and M. Bier, *Anal. Chem.*, 61 (1989) 362.
15 H. A. Abramson, *J. Gen. Physiol.*, 15 (1931–32) 575.
16 H. A. Abramson, M. H. Gorin and L. S. Moyer, *Chem. Rev.*, 24 (1939) 345.
17 W. J. Moore, *Physical Chemistry*, Prentice-Hall, NJ, 4th ed., 1972, p. 455.
18 H. T. Rasmussen and H. M. McHair, *J. Chromatogr.*, 516 (1990) 223.
19 J. L. Oncley, *Ann. N.Y. Acad. Sci.*, 41 (1941) 121–150.
20 A. Sillero and J. M. Ribeiro, *Anal. Biochem.*, 179 (1989) 319–325.
21 P. D. Grossman, J. C. Colburn and H. H. Lauer, *Anal. Biochem.*, 179 (1989) 28.

22 M. Zhu, D. L. Hansen, S. Burd and F. Gannon, *J. Chromatogr.*, 480 (1989) 311.
23 B. J. Compton, J. S. Gerald, D. A. Lowe and R. P. Elander, *Biotech. Tech.*, 3 (1989) 349.
24 L. D. Ward and D. J. Winzor, *Arch. Biochem. Biophys.*, 209 (1981) 650.
25 A. A. Zamyatnin, *Ann. Rev. Biophys. Bioeng.*, 13 (1984) 145–165.
26 I. Pilz, G. Puchwein, O. Kratky, M. Herbst, O. Haager, W. E. Gall and G. M. Edelman, *Biochemistry*, 9 (1970) 211.
27 J. Yguerabide, H. F. Epstein and L. Stryer, *J. Mol. Biol.*, 51 (1970) 573.
28 R. E. Offord, *Nature (London)*, 211 (1966) 591.