

A modified extended UNIQUAC model for proteins

João A.P. Coutinho^{a,*}, Fernando L.P. Pessoa^b

^a CICECO, Departamento de Química, Universidade de Aveiro, 3810-193 Aveiro, Portugal

^b Departamento de Eng. Química, Escola de Química, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21949-900 Rio de Janeiro, Brazil

Available online 8 July 2004

Abstract

A modification of the extended UNIQUAC model is proposed for the description of the non-ideality of protein solutions. Here, the Staverman–Guggenheim combinatorial contribution used in extended UNIQUAC model is replaced by the Flory–Huggins term to take into account the size differences between the protein and solvent. This new model allows an excellent description of the activity coefficients in protein systems, for a large range of pH and ionic strengths, with a reduced number of parameters.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Model; Activity coefficient; Protein

1. Introduction

With the development of biotechnology, the interest in production of proteins by microorganisms is quickly increasing. Proteins obtained by fermentation are produced in complex broths containing inorganic salts, sugars, organic acids and cells from where the recovery of the higher value compounds may prove difficult. A number of methods for protein purification, applicable at bench scale, have been developed [1]. However, the scale-up and optimisation of these techniques at industrial level is only possible if a mathematical description of the process is available. For that purpose, the thermodynamic description of the protein non-ideality in the broth and its dependency on the pH, ionic strength, and temperature on the presence of secondary compounds such as sugars (simple or polysaccharides) and polymers are essential.

A number of attempts to describe the behaviour of proteins in solution have been reported in the literature. Most approaches are based on the potential of mean force [2–4] or on equations of state based on hard sphere interaction potentials [5]. Excess Gibbs energy models have seldom been applied for the description of protein systems. Engineering local composition models such as UNIQUAC, UNIFAC or NRTL, although often used for amino acid systems with

some success [6–10], were only applied to proteins by Agena et al. [11,12] in collaboration with one of the authors. However, Agena's work is a very crude approach to the modelling of proteins in solution, as a conventional UNIQUAC model is used, that neither takes into account the large size differences between the protein and solvents, nor considers the long-range forces that arise from the electrostatic interactions between the protein and the solvent.

On this work, it will be shown that a new G^E model, developed by modifying the extended UNIQUAC model, can provide an excellent description of the activity coefficients of the protein in solution with a reduced number of parameters, and that these parameters can be used to predict the behaviour of the protein at pH's and ionic strengths other than those used for the parameter estimation. The protein activity coefficients derived from the osmotic pressure data reported by Haynes et al. [2] for the α -chymotrypsin are used to validate the proposed model.

2. Model

The extended UNIQUAC model has been proposed Sander et al. [13,14] for the description of classical electrolyte systems, i.e. aqueous systems of small ($\sim 4 \text{ \AA}$), inorganic ions with a charge of 1–3 such as Na^+ , K^+ , Cl^- or SO_4^{2+} . It consists of a combination of the original UNIQUAC model, accounting for the short-range interactions, with the Debye–Huckel model to account for the

* Corresponding author. Tel.: +351 234 401507;

fax: +351 234 370084.

E-mail address: jcoutinho@dq.ua.pt (J.A.P. Coutinho).

long-range electrostatic interactions. It has neither been applied to organometallic electrolytes nor to any sort of organic electrolytes. Nevertheless, the model proved to be very successful in the description of SLE and VLE for inorganic systems, as well as for a number of other properties such as osmotic coefficients, heat capacities and heats of solution [15,16]. For this reason, it was chosen among the different electrolyte models available for the description of behaviour of proteins in solution.

Proteins differ from small inorganic ions in size (typically larger than 40 Å with molecular weights higher than 20,000), in complexity (they possess a large and diversified number of functional groups and elaborate structure), and in charge (proteins can change from a net charge of +60 to –60 within a few pH units). For such system, the excess Gibbs energy will consist of three contributions:

$$G^E = G_{\text{combinatorial}}^E + G_{\text{residual}}^E + G_{\text{Debye-Huckel}}^E \quad (1)$$

The combinatorial term will account for the entropic interactions arising from size and shape differences between the molecules. The extended UNIQUAC uses the Staverman–Guggenheim term to represent these interactions. However, it is known that for very asymmetric systems the Staverman–Guggenheim does not produce a good description of the non-ideality of the systems [17,18]. Thus, it was replaced by the Flory–Huggins term for a more adequate description of the very large size differences between the protein, the water and the other ions. Although the two terms are virtually identical for protein concentrations higher than 1 mol%, the Staverman–Guggenheim shows a very questionable behaviour for dilute solutions as shown in Fig. 1. The Flory–Huggins combinatorial term here used is given by

$$\frac{G_{\text{combinatorial}}^E}{RT} = \sum_i x_i \ln \left(\frac{\phi_i}{x_i} \right) \quad (2)$$

The combinatorial-free volume terms, usually used for polymer solutions [18], were not adopted since the water has a free volume similar to polymers [19]. Thus, it is not necessary to use a free volume contribution for polymer aqueous systems.

The UNIQUAC residual term that accounts for the energetic short-range interactions is [15]

$$\frac{G_{\text{residual}}^E}{RT} = - \sum_i x_i q_i \ln \left(\sum_k \theta_i \psi_{ki} \right) \quad (3)$$

where q_i are the UNIQUAC surface area parameters, θ_i the surface area fractions and the parameter ψ_{kl} is given by

$$\psi_{kl} = \exp \left(- \frac{u_{kl} - u_{ll}}{T} \right) \quad (4)$$

with the UNIQUAC interaction parameters $u_{kl} = u_{lk}$. The surface area parameter for the protein $q_p = 700.2723$ was obtained from a correlation by Avena et al. [11]. The parameters for water ($q_w = 1.400$) were obtained from [16].

Despite the change in the combinatorial term, it was assumed that the UNIQUAC parameters available in the extended UNIQUAC parameter table for the ion–ion, ion–water and water–water interactions can be used without reestimation. This was found acceptable in previous works, where UNIQUAC or UNIFAC interaction parameters, fitted to small molecules with a given combinatorial term were used for polymer systems with a new combinatorial term [18,20]. Support for this approach can be derived from Fig. 1. For concentrations higher than 1 mol% the combinatorial terms are essentially identical. Since most of the data available for simple systems belong to the region where the two combinatorial terms are identical, interaction parameters adjusted for the Flory–Huggins would be similar to the parameters obtained with the Staverman–Guggenheim combinatorial term. This approach allows that only a small number of interaction parameters must be fitted for new systems.

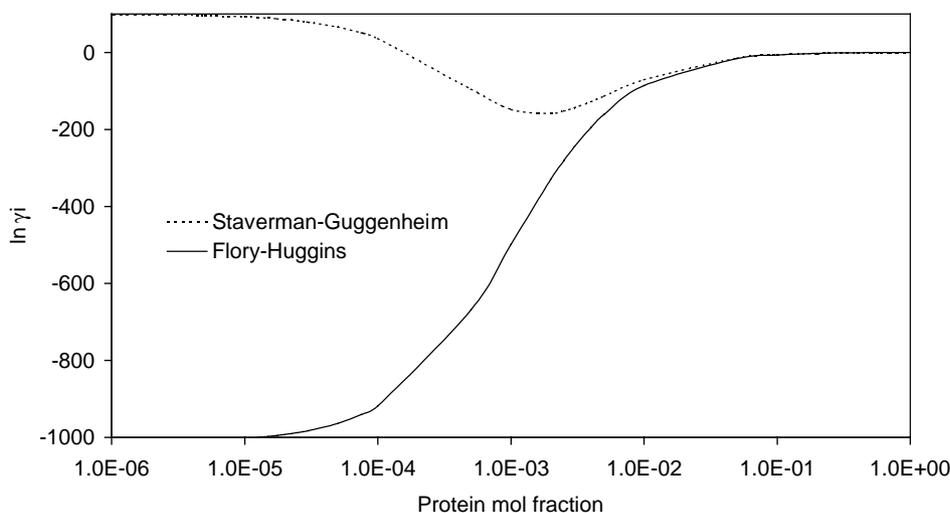


Fig. 1. Comparison between the Staverman–Guggenheim and Flory–Huggins combinatorial terms for α -chymotrypsin in aqueous solution.

The Debye–Hückel term for the long-range electrostatic interactions is [15]

$$\frac{G_{\text{Debye-Hückel}}^E}{RT} = -x_w M_w \frac{4A}{b^3} \left[\ln(1 + b\sqrt{I}) - b\sqrt{I} + \frac{b^2 I}{2} \right] \quad (5)$$

where x_w is the mole fraction, M_w the molar mass of water and A is a constant that is in the 273.15–373.15 K range can be approximated by [16]

$$A = 1.131 - 1.335 \times 10^{-3} (T - 273.15) + 1.164 \times 10^{-5} (T - 273.15)^2 \quad (6)$$

with T in Kelvin and b being a constant that depends on the size of the ions. For inorganic ions ($\sim 4 \text{ \AA}$), it can be taken as $1.5 (\text{kg mol}^{-1})^{1/2}$. For the protein with a size of $\sim 40 \text{ \AA}$, b was taken as $15 (\text{kg mol}^{-1})^{1/2}$. I is the ionic strength given by

$$I = \frac{1}{2} \sum_i m_i z_i^2 \quad (7)$$

where m_i is the molality of ion i and z_i its charge.

Unlike the charge of simple ions, the protein net charge and its dependency on the pH is not easy to obtain with accuracy, neither experimentally nor by calculation. The approach used in this work was to treat the protein net charge (z_{protein}) as a fitting parameter. A discussion on the physical sense of the fitted protein charges is presented in the next section.

The activity coefficient for species i , γ_i is obtained by partial molar differentiation of the excess Gibbs energy according to

$$\ln \gamma_i = \left[\frac{\partial(nG^E/RT)}{\partial n_i} \right]_{P,T,n_{j \neq i}} \quad (8)$$

As usual with electrolytes, the asymmetrical convention is adopted in this work. The activity coefficients presented are thus asymmetrical activity coefficients with an infinite dilution value equal to unity.

3. Results

The description of the activity coefficients for aqueous solutions of α -chymotrypsin obtained from osmotic pressure measurements [2] was used to evaluate the ability of the proposed model proposed. The asymmetric molal activity coefficients are calculated from the osmotic pressures following the approach of Wills et al. [21]. From the osmotic pressure data, Π , over the molar protein concentration, c_p , it was possible to obtain the virial coefficients, B_k , using the equation:

$$\frac{\Pi}{RT} = c_p(1 + B_2 c_p + B_3 c_p^2 + \dots) \quad (9)$$

Solute molal activity coefficients dependency with the molal composition, m , are given by

$$\ln \gamma_p = 2C_2 m_p + \frac{3}{2} C_3 m_p^2 \quad (10)$$

The coefficients, C_k , are related to the virial coefficients, B_k , by

$$C_2 = (B_2 - v_p M_p) \rho_s \quad (11)$$

$$C_3 = (B_3 - 2B_2 v_p M_p + (v_p M_p)^2) \rho_s^2 \quad (12)$$

where v_p and M_p represent the partial specific volume and the molecular weight of the protein with values of $0.736 \text{ cm}^3/\text{g}$ [11] and 25651 g/mol [22], and ρ is the solvent density.

The data used have protein concentrations ranging from 0 to 9 g/L and pH's ranging from 3 to 12 in 0.1 M potassium sulphate buffer ($I = 0.3 \text{ M}$), except for two cases, for which the buffer ionic strength, at pH 3, was 0.03 and 0.15 M. Information about the α -chymotrypsin composition and structure were obtained from [22]. The ionisable groups present are reported in Table 1.

3.1. The pH dependency of the protein activity coefficient

The pH dependency of the protein solubility is well known and can be found in any biochemistry textbook [23]. It follows a U-shaped curve with a minimum solubility at the isoelectric point, increasing for both higher and lower pH's. The asymmetrical molal activity coefficients obtained from osmotic pressure measurements also show this behaviour. The activity coefficients have values close to 1 near the isoelectric point decreasing with pH to both sides of the isoelectric point.

As discussed above, the interaction parameters u_{kl} , for ions and water already available for the extended UNIQUAC [16] were used in this work without reestimation in order to minimize the number of parameters to fit. A parameter table for the extended UNIQUAC can be found in the work by Thomsen et al. [16]. To prevent interferences of the electrostatic contribution on the optimised parameter values, the UNIQUAC interaction parameters for the protein (protein–protein, protein–water, protein– K^+ and protein– SO_4^{2+}) were fitted to the data available at the isoelectric point (pH 8.25). With the available data for the protein in a single salt, it was not possible for the model to distinguish between the cation and the anion interaction

Table 1
Acid and basic groups on α -chymotrypsin [22]

Basic groups		Acid groups	
Arginine	4	Tyrosine	4
Histidine	2	Cysteine	10
Lysine	14	Ac. glutamic	5
NH ₂ terminal	1	Ac. aspartic	9
		COOH terminal	1

Table 2
UNIQUAC interaction parameters, u_{kl} , for α -chymotrypsin, estimated using data at the isoelectric point

Interaction parameter	Parameter value (K)
$u_{\alpha\text{-chymotrypsin}/\alpha\text{-chymotrypsin}}$	-518.22
$u_{\alpha\text{-chymotrypsin}/\text{water}}$	-161.47
$u_{\alpha\text{-chymotrypsin}/\text{K}^+}$	883.10
$u_{\alpha\text{-chymotrypsin}/\text{SO}_4^{2+}}$	883.10

with the protein. For this reason, the interaction parameters for both ions with the protein were taken as identical. The parameters obtained are reported in Table 2. Since the short-range interactions described by the residual term will remain the same over the entire range of pH, the set of parameters fitted at the isoelectric point was used to describe the energetic interactions for all pH values.

The long-range interactions generated by the protein net charge, and their dependency with the pH, are described by the Debye–Huckel term. As mentioned before, the protein net charge is not easily available with the exception of the isoelectric point and the very high and low pH region. For this reason, the protein net charge, z_i in Eq. (7), was used as a fitting parameter in the model for the description of the activity coefficient dependency with the pH of the system. The fitted values for the protein net charge are presented in Fig. 2. A comparison between the experimental and model results for the activity coefficients, reported in Fig. 3a and b, shows the adequacy of the model for the description of the non-ideality of protein solutions. The proposed model, compared to previous works [11,12], uses lower number of parameters to describe the activity coefficient data. These are the UNIQUAC interaction parameters reported in Table 2, adjusted at the isoelectric point, plus the protein net charge fitted to each pH value.

In Fig. 2, the fitted values of the protein net charge (PNC) are compared with the values calculated by assuming that

the pK_a 's of the amino acids on the protein are the same as the free amino acids according to the following equation:

$$\text{PNC} = \sum_{i+} n_{i+} \frac{10^{pK_{i+}-\text{pH}}}{10^{pK_{i+}-\text{pH}} + 1} + \sum_{i-} n_{i-} \frac{10^{\text{pH}-pK_{i-}}}{10^{\text{pH}-pK_{i-}} + 1} \quad (13)$$

where $i+$ are the basic groups and $i-$ the acid groups described in Table 1. This approach is not exact. The pK_a 's for amino acids in proteins may differ from their free-form values by as much as three pH units [24], and they are also dependent on the salts present in solution [23]. Nevertheless, this approach provides a fair and simple estimate of the protein net charge dependency with pH. The comparison between fitted and calculated values shows an excellent agreement between the two curves, except for pH values between 4 and 7. This is probably due to the position of some acidic amino acids in the protein structure that make them less available to the solvent and consequently more difficult to ionise.

Since the UNIQUAC interaction parameters are pH independent, the combinatorial and residual term contributions to the activity coefficient are always identical to the values at the isoelectric point (pH 8.25). The main contribution to the activity coefficient at pH's removed from the isoelectric point is the Debye–Huckel term, as shown in Fig. 4 where the three individual contributions are plotted for 4 pH values. Only the Debye–Huckel term shows a change with the pH. This importance of the Debye–Huckel term is due to the large protein net charge and its variation with the pH of the medium.

The results presented show that the proposed model can be used to describe the activity coefficients for proteins, if experimental data at the isoelectric point is available to fit the short-range interaction parameters u_{kl} . If data at other pH values is available, the model can also be used to

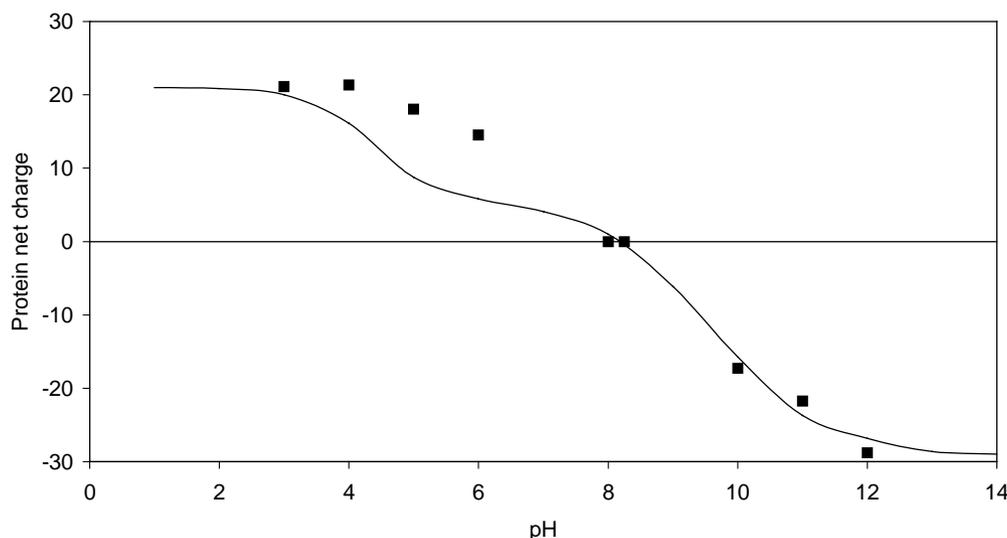


Fig. 2. Comparison between the fitted and estimated protein net charge pH dependency.

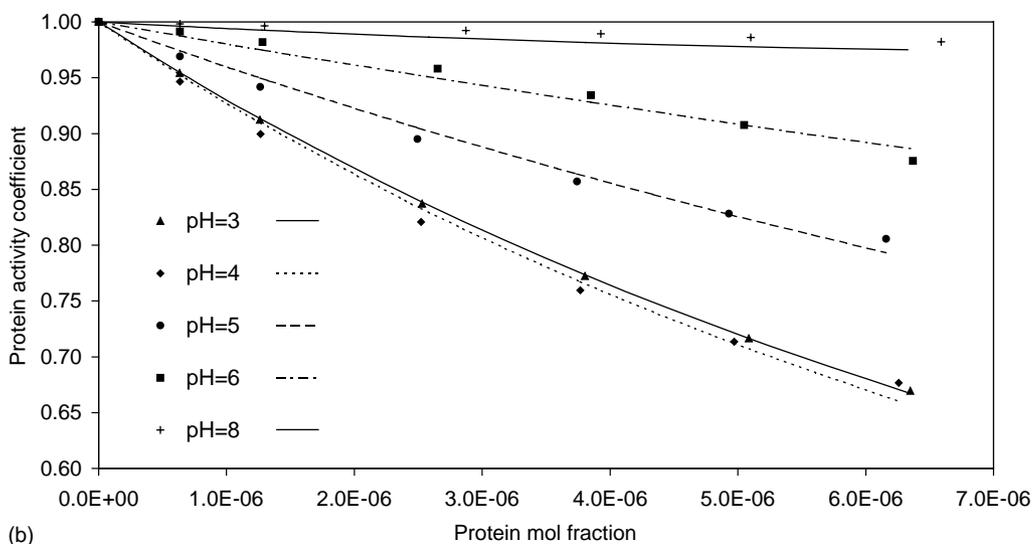
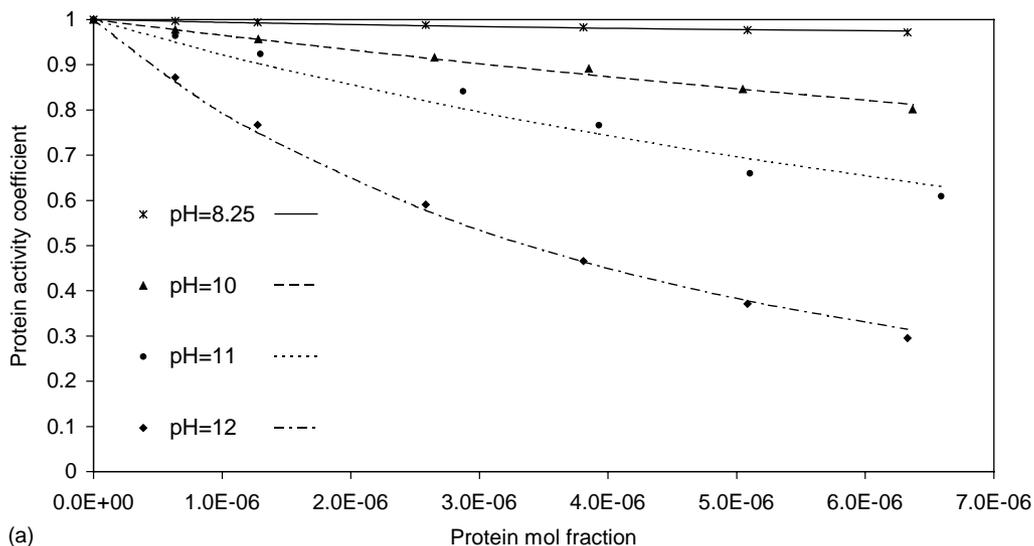


Fig. 3. Correlation of the activity coefficients with extended UNIQUAC for pH's (a) above pI and (b) below pI.

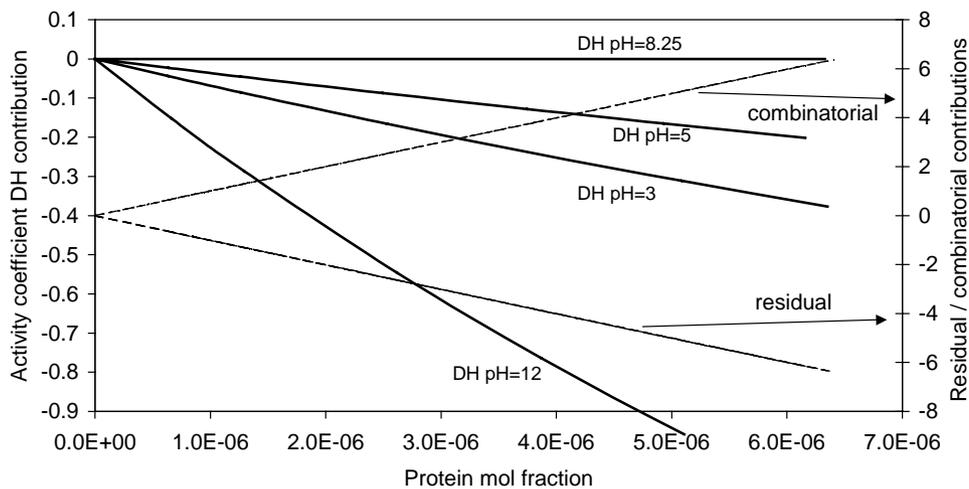


Fig. 4. Values for the asymmetric residual, combinatorial and Debye–Huckel contributions for the activity coefficient of the protein at pH values ranging from 3 to 12.

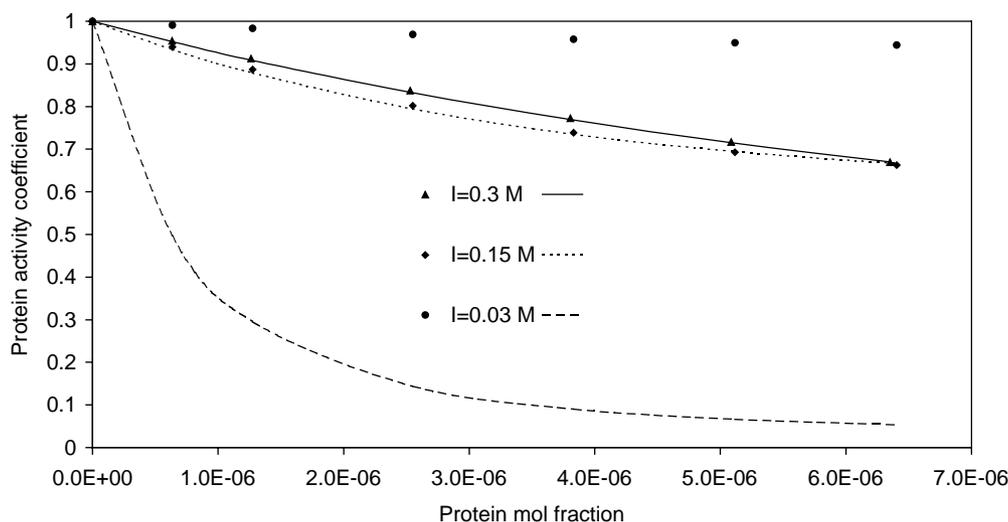


Fig. 5. Experimental and predicted ionic strength dependency of the activity coefficients.

describe the pH dependency of the activity coefficients by fitting the protein net charge. Yet, if this data is not available, the protein net charge assessed from the pK_a 's of the free amino acids can be used to predict the activity coefficients for the protein at different pH values with a reasonable degree of confidence. This is particularly valid for the extreme values of pH and around the isoelectric point, where the calculated protein net charge is less prone to be affected by groups with pK_a 's substantially different from their amino acids free-form values, due to their location within the protein.

3.2. Ionic strength dependency of the protein activity coefficient

As for the pH, the ionic strength dependency of the protein activity coefficients is well established [23]. For very low salt concentrations, the protein solubility increases with increasing salt content, creating the *salting-in* region, while for higher salt concentrations, the increase in salt content promotes the precipitation of the protein in a phenomenon known as *salting-out*. This originates the typical dependency of the protein solubility curves with ionic strength with the shape of a reverse U. In terms of activity coefficients a decrease is found in the salting-in region followed by an increase in the region of salting-out.

Predictions on the effect of the ionic strength on the solubility of the protein at pH 3 were done using the net charge and interaction parameters previously estimated. The results shown in Fig. 5 indicate that experimentally there is a decrease in the activity coefficients between the ionic strengths of 0.03 and 0.15M corresponding to a salting-in region followed by an increase between 0.15 and 0.3M characteristic of a salting-out region. However, the model describes a continuous decrease of the activity coefficients as the ionic strength is reduced from 0.3 to 0.03M. The 0.3 and

0.15M systems are well described. It must be emphasized, that the behaviour at 0.15M is a pure prediction. Yet, the model cannot describe the salting-in region and predicts activity coefficients for ionic strengths of 0.03M much below the experimental values. This limitation may be related to the simplifications adopted for the protein-ion interactions discussed above. Studies to improve the model for simultaneous description of the salting-in and salting-out regions are being undertaken.

3.3. Temperature dependency of the protein activity coefficient

Proteins may show various temperature-dependent solubilities. Cristopher et al. [25] present results for a number of proteins with temperature-dependent solubilities including a number of proteins with retrograde solubility (the α -chymotrypsinogen may have such behaviour) and some proteins whose solubility does not seem to be significantly affected by the temperature. The data for the system under study is available only at 25 °C, thus the temperature dependency of the protein behaviour in solution cannot be investigated or correlated for this system. For systems where data is available adjusting temperature-dependent parameters to the modified extended UNIQUAC model [13–16]:

$$u_{ij} = u_{ij}^0 + u_{ij}^t(T - 298.15) \quad (14)$$

should allow the description of the temperature dependency of the activity coefficients within a reasonable temperature range.

4. Conclusions

A modified version of the extended UNIQUAC model is proposed for proteins and other polyelectrolytes in aqueous

solution. The activity coefficients for α -chymotrypsin, obtained from osmotic pressure measurements in a wide range of pH and ionic strength, were used to validate the proposed model. The interaction parameters for water and ions were taken from extended UNIQUAC tables and the protein interactions were fitted to data at the isoelectric point. The protein net charge was used as a fitting pH-dependent parameter, but it was shown to be in close agreement to the calculated protein net charge. Due to the large protein net charge, the Debye–Hückel is the key term in the activity coefficient model. A good description of the activity coefficients is achieved over the entire pH region studied and the estimated parameters allow the predictions of the protein behaviour on the salting-out region.

Acknowledgements

This work was developed in the framework of a project financed by GRICES (Portugal) and CAPES (Brazil). Thanks are due to Professor Carlos M. Silva (UA) and Professor Maria Alice Z. Coelho (UFRJ) for fruitful discussions.

References

- [1] R.K. Scopes, Protein Purification: Principles and Practice, 3rd ed., Springer-Verlag, 1994.
- [2] C.A. Haynes, K. Tamura, H.R. Korfer, H.W. Blanch, J.M. Prausnitz, J. Phys. Chem. 96 (1992) 905–912.
- [3] D.E. Kuehner, H.W. Blanch, J.M. Prausnitz, Fluid Phase Equilibr. 116 (1996) 140–147.
- [4] F.W. Tavares, S.I. Sandler, AIChE J. 43 (1997) 218–231.
- [5] T. Hino, J.M. Prausnitz, AIChE J. 45 (1999) 622–632.
- [6] S.P. Pinho, C.M. Silva, E.A. Macedo, Ind. Eng. Chem. Res. 33 (1994) 1341–1347.
- [7] M.T. Gude, L.A.M. van der Wielen, K.Ch.A.M. Luyben, Fluid Phase Equilibr. 116 (1996) 110–117.
- [8] M.K. Khoshkbarchi, J.H. Vera, Ind. Eng. Chem. Res. 37 (1998) 3052–3057.
- [9] J.C. Bellot, R.V. Tarantino, J.S. Condoret, AIChE J. 45 (1999) 1329–1341.
- [10] A. Soto, A. Arce, M.K. Khoshkbarchi, J.H. Vera, Fluid Phase Equilibr. 158–160 (1999) 893–901.
- [11] S.M. Agena, D.L. Bogle, F.L. Pessoa, Biotech. Bioeng. 55 (1997) 65–71.
- [12] S.M. Agena, M.L. Pusey, D.L. Bogle, Biotech. Bioeng. 64 (1999) 44–50.
- [13] B. Sander, A. Fredenslund, P. Rasmussen, Chem. Eng. Sci. 41 (1986) 1171–1183.
- [14] B. Sander, A. Fredenslund, P. Rasmussen, Chem. Eng. Sci. 41 (1986) 1197–1202.
- [15] H. Nicolaisen, P. Rasmussen, J.M. Sørensen, Chem. Eng. Sci. 48 (1993) 3149–3158.
- [16] K. Thomsen, P. Rasmussen, R. Gani, Chem. Eng. Sci. 51 (1996) 3675–3683.
- [17] I. Kikic, P. Alessi, P. Rasmussen, A. Fredenslund, Can. J. Chem. Eng. 58 (1980) 253–257.
- [18] G.M. Kontogeorgis, A. Fredenslund, D. Tassios, Ind. Eng. Chem. Res. 32 (1993) 362–372.
- [19] D. Rasmussen, P. Rasmussen, Chem. Eng. Prog. 85 (2) (1989) 50–56.
- [20] J.A.P. Coutinho, E.H. Stenby, Ind. Eng. Chem. Res. 35 (1996) 918–925.
- [21] P.R. Wills, W.D. Comper, J.D. Winzor, Arch. Biochem. Biophys. 300 (1993) 206–212.
- [22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000) 235–242.
- [23] D. Voet, J.G. Voet, Biochemistry, 2nd ed., John Wiley, 1995.
- [24] D.T. Haynie, Biological Thermodynamics, Cambridge University Press, Cambridge, 1999.
- [25] G.K. Christopher, A.G. Phipps, R.J. Gray, J. Cryst. Growth 191 (1998) 820–826.