Effect of electrostatic energy on partitioning of proteins in aqueous two-phase systems

A. Olivera-Nappa*, G. Lagomarsino, B.A. Andrews, J.A. Asenjo

Centre for Biochemical Engineering and Biotechnology, Department of Chemical Engineering, Millenium Institute for Advanced Studies in Cell Biology and Biotechnology, University of Chile, Beauchef 861, Santiago, Chile

Abstract

An attempt has been made to adopt a different approach to evaluate the effect of a protein's charge on its partitioning behaviour in PEG/salt aqueous two-phase systems (ATPS). This has been done using a computer methodology (DelPhi) that allows the calculation of the electrostatic solvation energy that charged proteins present in a particular media such as aqueous polymer–salt systems. This calculation was done for the protein in each of the phases and a correlation was investigated that related the electrostatic energy difference of the protein in each of the phases and its partition coefficient in ATPS. Such correlation resulted in a statistical model that also included the effect of molecular weight and a shape factor at each particular pH. A global correlation which included the effect of pH was also found. All the correlations were statistically evaluated and gave good results.

Keywords: Aqueous two-phase systems; Partitioning; Electrostatic energy; Proteins

1. Introduction

Biotechnological applications of aqueous two-phase systems (ATPS) are clearly influenced by the ability to develop models and correlations that allow to understand how physico-chemical properties of proteins and their interaction with the properties of the surrounding polymer and salt phases affect partitioning in these systems. Such understanding would allow prediction of partitioning based on a protein's fundamental properties and could be of importance in the selection and design of characteristics of an appropriate ATPS that would optimize separation of a particular protein from another one, or of a product protein from one or more contaminants. Hence the ability to predict the partition coefficient of a protein in PEG/salt and PEG/dextran aqueous two-phase systems based on the molecular properties of the protein is clearly an important task.

The effect of a protein's hydrophobicity, charge, molecular weight as well as concentration on partitioning has been investigated [1–7] and it has been found that the main factor determining partitioning in PEG/salt systems is the hydrophobicity of the protein. Surface hydrophobicity of a protein has been measured by chromatography and precipitation [2,3]; and precipitation gave the best correlation with partitioning which has been later confirmed for a number of different proteins [7].

Many attempts have been made to try and correlate charge and molecular weight of proteins with their partition coefficients in ATPSs. Molecular weight has been quite satisfactorily correlated with partition coefficient, K, in PEG/polymer systems but not in PEG/salt ones [1,7]. To evaluate the effect of charge different approaches have been investigated such as pH change in the system [1], use of chemically or genetically modified proteins to alter its charge but no other properties [4,8] and also the use of a representative number of proteins of different charge where net charge, volume charge density and surface charge density were used to correlate with the value of K.

In PEG/salt systems such as the ones used in the present paper, the value of K of a protein increased as it became more negatively charged (higher pH) [1], and when a relatively large number of proteins were used a similar behaviour was observed, and when an additional salt (NaCl) was added to the system the trend changed and more positively charged proteins had higher values of K [7]. This

* Corresponding author.

behaviour was also observed when chemically modified proteins were used [4]. In all cases, but particularly when a range of different proteins where used, correlation coefficients have been very poor hence no good correlation has been found between properties related to a protein's charge and its partition behaviour in ATPSs.

For the reasons discussed above it was decided to investigate a methodology that allows the calculation of the electrostatic potential energy that a charged protein molecule has in a particular media such as an aqueous polymer or aqueous salt phase. This energy allows for a real measure of the electrostatic component of the solvation energy of a protein in a medium. This calculation can be done for the protein in each of the phases and the aim of this paper was to investigate if there is a function that correlates the electroststic energy difference of the protein in each of the phases and its partition coefficient in the ATPS.

2. Theory

2.1. Total electrostatic energy and its numerical determination using DelPhi

The electrostatic solvation energy is determined by the DelPhi program [9–11], which calculates the electrostatic potential inside and outside a protein molecule. To do this the Poisson–Boltzmann equation for continuous media is numerically solved, both in its non-linear (Eq. (1)) and linear form (Eq. (2)), this latter an approximation for low charge densities and ionic forces, using a finite differences method.

$$\nabla[\varepsilon(r)\,\nabla\phi(r)] - \bar{\kappa}_0^2 \sinh[\phi(r)] + 4\pi\rho_{\text{int}}(r) = 0 \tag{1}$$

$$\nabla[\varepsilon(r)\,\nabla\phi(r)] - \bar{\kappa}_0^2\phi(r) + 4\pi\rho_{\text{int}}(r) = 0 \tag{2}$$

where

$\bar{\kappa}_0 = 0$ inside the molecule; and

$$\bar{\kappa}_0 = \sqrt{\varepsilon_s \kappa} = \sqrt{\frac{8\pi e^2 N_A I}{1000kT}} \quad \text{outside the molecule.}$$

Given an electric charge distribution in the intramolecular space $\rho_{\text{int}}(r)$, dependent on the protein structure and its amino acid sequence, an ionic strength I outside the molecule and a dielectrical spatial function $\varepsilon(r)$, which has different values inside and outside the molecule, the numerical solution of Poisson–Boltzmann equation gives $\phi(r)$, the electrostatic potential value for every point in space, in units of kT/e. From this result and the charge distribution $\rho_{\text{int}}(r)$, DelPhi also calculates the total electrostatic energy of the charged molecule when suspended in the considered medium, in units of kT, which can be considered as the electrostatic component of solvation energy for the molecule in this medium [11]. When two different media are considered, the difference between the corresponding electrostatic energies can be subsequently taken.

3. Materials and methods

3.1. Materials

PEG with a molecular weight of 4000 (rel. mol. mass units) (Mr = 3.500–4.500) was purchased from Fluka Chemicals, Germany. All other chemicals were analytical grade.

3.1.1. Proteins

Table 1 shows the proteins used, their source and corresponding PDB file code.

3.2. Procedures and computational evaluation

3.2.1. Preparation of phase systems and partition experiments

Aqueous two-phase systems were prepared as described previously [1]. Briefly, systems were prepared from stock solutions of PEG (40% w/w) and phosphate (40% w/w). The phosphate stock solution consisted of a mixture of K_2HPO_4 and NaH_2PO_4 at the appropriate pH. Stock solutions were stored at 4 °C and were equilibrated by standing at room temperature before use. Partition experiments were done at 20 °C. Total protein was added to the systems at a final concentration of 1 g/l. Samples of top and bottom phase were assayed for protein concentration.

3.2.2. Protein assays

Total protein was measured using the modified Bradford dye-binding assay [12], interference from phase forming components was taken into account.

3.2.3. Titration curves

Titration curves were obtained in a PhastSystem (Pharmacia Amersham Biotechnology, Sweden) using PhastGel IEF 3–9.

Table 1
Proteins used, their source and corresponding PDB file code

Protein	Source	PDB code
Cytochrome *C*	Horse	1HRC
Ribonuclease A	Bovine	1AFU
Lactalbumin	Human	1A4V
Lysozyme	Chicken egg white	2LYM
Myoglobin	Horse heart	1YMB
Lactoglobulin	Bovine	1B0O
Trypsin inhibitor	Soy bean	1AVU
Thaumatin	Thaumatococcus danielli	1THU
Trypsin	Bovine pancreas	1BJU
Chymotrypsin	Bovine pancreas	4CHA
Chymotrypsinogen A	Bovine pancreas	2CGA
Subtilisin	Bacillus licheniformis	1AF4
Amylase	Bacillus licheniformis	1BLI
Conalbumin	Chicken egg white	1AIV

3.2.4. Computer analysis

A computational methodology (DelPhi**)** [9–11] was used to estimate the electrostatic solvation energy of the proteins. The program uses as input a Protein Data Bank file (PDB, http://www.rcsb.org/pdb) with the corresponding structural coordinates for atoms in the protein (Table 1).

3.2.5. Protein volume and surface estimation

DeepView/Swiss-pdbViewer 3.7 [13] was used on each molecular structure (http://www.expasy.org/spdbv).

4. Results and discussion

4.1. Experimental data

A number of proteins (14) with a wide range of physico-chemical properties were chosen to evaluate the effect of charge and electrostatic potential energy on partitioning. These are shown in Table 1 with their source and PDB file code. The PEG/salt aqueous two-phase system used was PEG 4000/phosphate as a representative system. The selected proteins did not form any precipitate at the interface.

The titration curves and partition coefficients for all proteins were measured and the net charge of the proteins was evaluated at pH 5, 7 and 9 as previously reported [7,14]. Partition coefficients, K, were also measured at these pHs. For most proteins the value of K increased with increasing pH. Only for thaumatin (pI 8.5), lysozyme (pI 10.5) and myoglobin (pI 7.4) it decreased slightly and for cytochrome C (pI 9.7) and chymtrypsinogen A (pI 8.6) it was almost constant. Fig. 1 shows the relationship between $\log K$ and net surface charge.

From this figure, it is not possible to infer any clear relationship between partition coefficient and net surface charge determined by the titration curve for each protein, and though a little tendency can be observed in the behaviour, it is not sufficient to derive a valid model. This could be due to net surface charge not taking into account solvent characteristics and how solvent interacts with individual charged amino acid molecules, making this parameter not accurate enough to describe the studied behaviour.

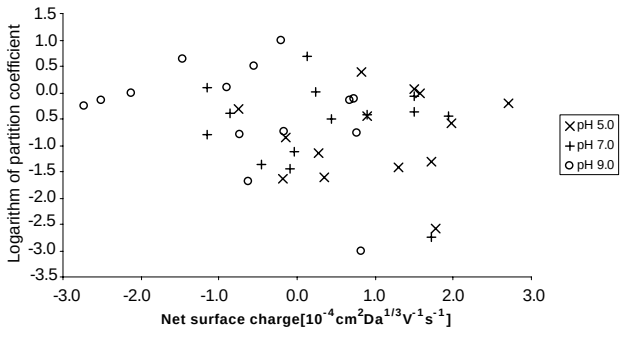

Fig. 1. Relationship between net surface charge measured on experimental titration curves and partitioning behaviour for each protein studied at different pH values in an aqueous two-phase PEG4000/phosphate system.

4.2. Determination of total electrostatic energy using DelPhi

In the program runs, we used default values for the ionic exclusion layer of the molecule and a water probe radius of 1.8 Å for molecular surface calculations. For each protein, the corresponding PDB archive containing structural data was used as input for the calculation, along with standard PARSE charge and size data for each atom type [15]. The ionization of particular residues in the protein was changed at each different pH value, which was used to reflect their titration behaviour. As a first approach, exposed histidines were considered to change their formal charges, from +1 at pH 5.0 to 0 at pH 7.0 and 9.0, exposed tyrosines were considered as neutral (0) at pH 5.0 and 7.0 and negative (−1) at pH 9.0, and free exposed cysteines were treated in the same way as tyrosines. The other residues were considered neutral, except for the exposed acid residues (Asp, Glu), which were always considered as negatively charged (−1), and the exposed basic residues (Arg, Lys), which were always assigned a formal positive charge (+1).

The dielectric constants used in the program runs were those estimated from values given in the literature for the upper and lower phases as 40 and 80 respectively [16,17], outside the molecule, and a constant value of 2 inside the molecule.

The ionic strength for each phase was determined using their salt concentration, which was calculated from the location of the operation point on the respective tie-line, in mol/l. The values used were 14.211 mol/l for the lower phase and 0.14 mol/l for the upper phase.

For each protein, the program was run three times. A first calculation was made with 30% grid occupancy and a dipolar border condition, where potential at grid borders is approximated by the Debye–Hückel potential of a dipole equivalent to the molecular charge distribution. The following two calculations were made with 50 and 90% occupancy respectively, using as border conditions the potentials generated by the previous calculation in the considered spatial positions. This kind of approach allowed us to perform calculations iteratively from the outside to the inside of the grid, increasing accuracy by using a finer grid each time. The results of the third iteration were used to calculate the total electrostatic energy of the protein in the respective medium (top and bottom phases).

4.3. Development of statistical model

The electrostatic energy difference of the proteins at the three pH values (3, 7 and 9) was calculated with DelPhi using the conditions stated above. The values obtained are shown in Table 2 and have been plotted as a function of $\log K$ in Fig. 2. The electrostatic energy at each pH depends

Table 2
Electrostatic energy difference between phases ($\delta E(kT)$) at three pH values

Protein	pH 5	pH 7	pH 9
Amylase	177.533	161.689	158.311
Chymotrypsin	107.630	117.746	122.086
Cytochrome *C*	75.533	83.393	84.287
Conalbumin	466.134	454.960	518.492
Chymotrypsinogen A	103.821	111.939	114.062
Lactalbumin	88.050	83.961	80.213
Lisozyme	80.252	75.273	72.311
Mioglobin	73.301	77.326	74.881
Lactoglobulin	113.832	118.725	126.527
Ribonuclease A	84.571	78.652	78.884
Subtilisin	79.630	93.728	109.155
Thaumatin	96.000	96.302	102.664
Trypsin	99.396	108.007	107.796
Trypsin inhibitor	158.532	122.228	178.533

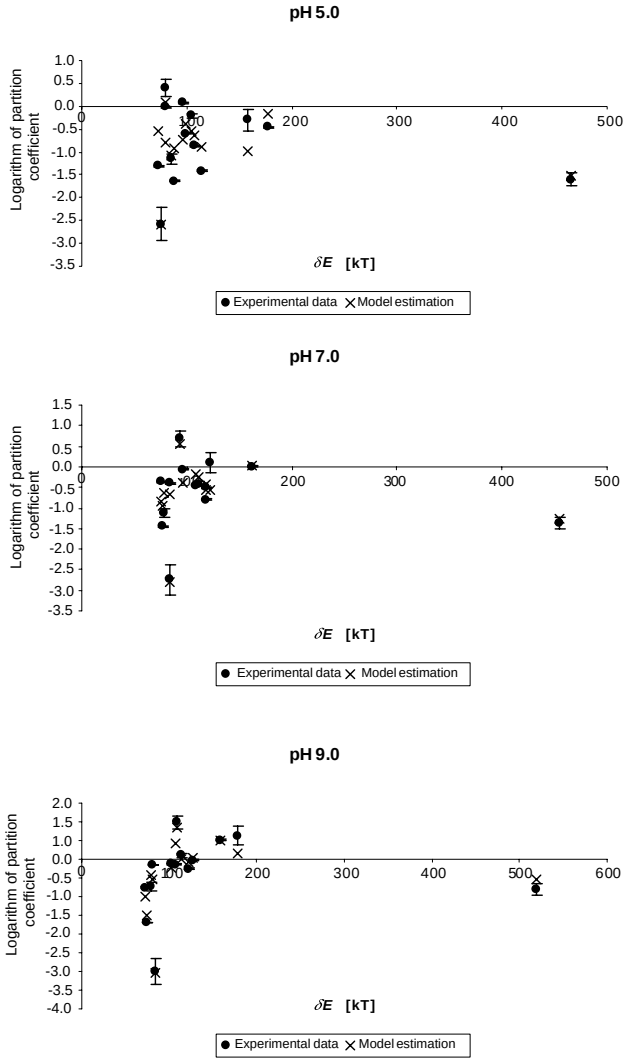

Fig. 2. Relationship between electrostatic energy difference (δE) and partitioning behaviour for each studied protein in an aqueous two-phase PEG4000/phosphate system. Each graph shows the experimental data and the predicted estimation of the respective model for each different pH value.

not only on the net number of total charges of the protein, but also on the location of such charges. The variation of the electrostatic energy with pH also depends on the order in which such charged groups are ionised or de-ionized with the change in pH. Also, the dielectric constant of the media, its ionic strength and the surface shape of the protein affect the value of the electrostatic energy in such media. The experimental data for the three pH values shown in Fig. 2 show a similar trend, ascending at low values of δE ($100kT$ and below) and descending above $100kT$.

In order to explain this bimodal behaviour a function of the type

$$\text{Log } K = Af(\delta E) + Bf(\delta E) + C \tag{3}$$

was evaluated. δE is the electrostatic energy difference. Furthermore we investigated whether such a correlation could be improved by introducing an additional property in the correlation. The properties evaluated were pI, hydrophobicity measured as discussed by Andrews et al. [7] and Asenjo et al. [1], molecular weight, protein volume, protein surface area and a shape factor (sphericity factor).

The only properties that significantly improved the correlation given in Eq. (3) were molecular weight and sphericity, F.

Thus it was considered appropriate to include these two properties in the model giving an equation of the following form

$$\text{Log } K = A(\delta E)^{a}(\text{MW})^{b}(F)^{c} + B(\delta E)^{d}(\text{MW})^{e}(F)^{f} + C \tag{4}$$

where MW is the molecular weight and F is the sphericity factor given by Eq. (5) where S is the surface area and V the volume of the protein molecule.

$$F = \frac{S^3}{36\pi V^2} \tag{5}$$

This is a sphericity factor that has a minimum value of 1, when the molecule is a perfect sphere, and goes above unity when the molecule's shape departs from a sphere and its surface-volume ratio increases.

Eq. (4) was fitted to the experimental data using MS Excel (quadratic fitting) at each different pH and the following equations were obtained:

- pH 5.0

$$\log(K) = -4.107\frac{\delta E^{0.08507}F^{0.1962}}{\text{MW}^{0.1884}} - 15.95\frac{\delta E^{45.34}}{F^{107.7}\text{MW}^{16.09}} + 4.218 \tag{6}$$

- pH 7.0

$$\log(K) = 416,400\frac{\text{MW}^{0.7849}}{\delta E^{1.330}F^{4.031}} - 1,596,000\frac{\text{MW}^{0.7195}}{F^{4.513}\,\delta E^{1.422}} - 1.255 \tag{7}$$

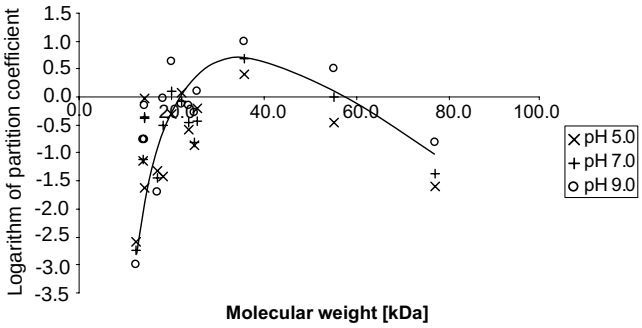

Fig. 3. Relationship between molecular weight and partitioning behaviour for each studied protein in an aqueous two-phase PEG4000/phosphate system at different pH values.

- pH 9.0

$$\log(K) = 512,100\frac{\mathrm{MW}^{0.7170}}{\delta E^{0.09112} F^{6.883}} - 1,065,000\frac{\mathrm{MW}^{0.7158}}{F^{7.012}\,\delta E^{0.2067}} - 0.5254 \qquad (8)$$

In Fig. 2 the values calculated using Eqs. (6)–(8) are also shown as well as the error bars of the experimental data. A very good fit was obtained particularly at pH 7 and 9.

Given that the range of molecular weights of the proteins was relatively large (12,400–70,000) the values of $\log K$ were also plotted as a function of MW (Da). This is shown in Fig. 3.

The volume and molecular surface of the proteins were analyzed together as factors that could influence protein–solvent interactions and phase affinity not directly related to electrostatic energy. Volume can be partially represented in the model by molecular weight, which could take into account, for instance, the size exclusion behaviour described as important for the stability of solutes in a liquid solvent. Surface can take into account the contributions of interfacial tension to protein solvation, which are directly related to the exposed surface of the molecule and partially include also size exclusion behaviours. Both molecular volume and surface have been incorporated into the model as the molecular shape factor defined in Eq. (5).

Fig. 3 shows that molecular weight and energy differential follow a similar behaviour, which indicates that this variable has an important influence over protein partitioning. From this observation, and being the molecular weight only very indirectly related to electrostatic energies calculations (by way of the molecular boundary of molecules, which is related to size, or molecular weight), it can be inferred that molecular weight also plays an important role in protein partitioning in PEG/salt aqueous two-phase systems and has thus been included in the final model.

An attempt was made to include all data shown at the three different pH values in a single statistical model that would include all experimental data found. Evidently this model also includes pH as a variable and is given by Eq. (9).

$$\log(K) = -529.1\frac{F^{0.004276}}{\mathrm{pH}^{0.002191}\,\delta E^{0.001901}\mathrm{MW}^{0.0001971}} - 137.4\frac{\delta E^{1.161}\mathrm{pH}^{0.5815}}{\mathrm{MW}^{1.537} F^{4.155}} + 526.1 \qquad (9)$$

The incorporation of a logarithmic-scale variable such as pH into the model also has the effect of concentrating each group of parameters *A–B–C*, *a–b–c* and *d–e–f* in a single order of magnitude, instead of the large span they cover in the isolated, pH-independent models as seen in Eqs. (6)–(8). This validates the choice of pH as the remaining variable to explain the observed behaviour (Table 3).

The equations found at pH 5, 7 and 9 as well as the overall model were statistically analyzed and the quality of the fit can be seen in Fig. 4. In these Figures, it can be observed that all points fall between the prediction limit of the models, which indicate the appropriateness and standalone capacity of the model to represent the behaviour of the experimental data. Moreover, the majority of the points fall between the 99% confidence level interval, which gives an idea of the prediction accuracy of the model.

The statistic significance of the models has been assessed by means of different statistic tests, which results are summarized in Table 3. They were very satisfactory.

Table 3
Statistical significance estimators for the models

Model	Correlation coefficient (%)	Explained variability (%)	Standard deviation of the residues	Durbin–Watson test value[a]	*P*-value	Student's *t* test value[b]
pH 5.0	77.7	60.4	0.54	2.39	0.0011	−1.56
pH 7.0	90.5	81.9	0.37	1.98	0.0000	2.13
pH 9.0	96.0	92.1	0.30	1.96	0.0000	−2.28
Global	86.1	74.1	0.46	2.33	0.0000	−2.63

[a] Values between 1.4 and 2.5 indicate that model errors are not auto-correlated.
[b] Indicates how many standard deviations departs the model predictions from the estimated data when the worst point is set aside from the model. Values between −3 and 3 indicate that the model explains all the experimental points.

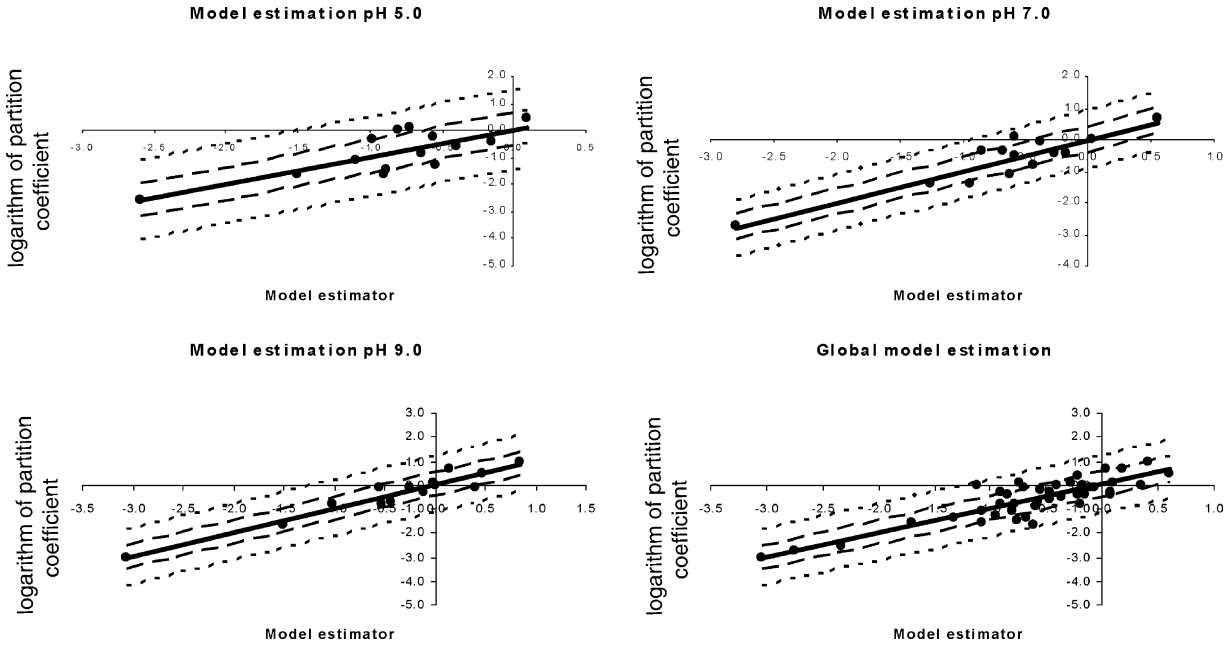

Fig. 4. Deviation from linearity of predicted results vs. experimental data for the logarithm of partition coefficient, for the three models at every different pH values and for the total model. Dashed lines indicate 99% confidence level interval and dotted lines represent the prediction limit for the model.

5. Conclusions

A computer methodology (DelPhi), that allows the calculation of the electrostatic energy that charged proteins present in a particular media such as aqueous polymer–salt systems (top and bottom phases of a PEG/phosphate ATPS), has been used to evaluate the effect of a protein's charge on its partitioning behaviour.

This methodology includes the effect of the charges of each amino acid of the protein and their distribution.

The investigation resulted in statistical correlations that also took into account the effect of molecular weight and a shape factor and was estimated at three different pH values. A global correlation which included the effect of pH was also found. All correlations were statistically analysed and had a high significance level.

Acknowledgements

Financial support from Conicyt (Fondecyt 1010142) and the Millennium Scientific Initiative (ICM 99-031-F) is gratefully acknowledged. We also acknowledge the assistance of Dr. Reynaldo Lopez-Planes.

References

[1] J.A. Asenjo, A.S. Schmidt, F. Hachem, B.A. Andrews, J. Chromatogr. B 668 (1994) 47.
[2] F.M. Hachem, B.A. Andrews, J.A. Asenjo, Enzyme Microb. Technol. 19 (1996) 507.
[3] T.T. Franco, A.T. Andrews, J.A. Asenjo, Biotechnol. Bioeng. 49 (1996a) 300.
[4] T.T. Franco, A.T. Andrews, J.A. Asenjo, Biotechnol. Bioeng. 49 (1996) 309.
[5] A.S. Schmidt, B.A. Andrews, J.A. Asenjo, Biotechnol. Bioeng. 50 (1996) 617.
[6] B.A. Andrews, J.A. Asenjo, J. Chromatogr. B 685 (1996) 15.
[7] B.A. Andrews, A.S. Schmidt, J.A. Asenjo, Biotechnol. Bioeng., submitted for publication.
[8] J.R. Luther, C.E. Glatz, Biotechnol. Bioeng. 44 (1994) 147.
[9] A. Nicholls, B. Honig, J. Comp. Chem. 12 (1991) 435.
[10] B. Honig, A. Nicholls, Science 268 (1995) 1144.
[11] W. Rocchia, E. Alexov, B. Honig, J. Phys. Chem. B 105 (2001) 6507.
[12] J.J. Sedmak, S.E. Grossberg, Anal. Biochem. 79 (1977) 544.
[13] N. Guex, M.C. Peitsch, Electrophoresis 18 (1997) 2714.
[14] E. Watanabe, S. Tsoka, J.A. Asenjo, Ann. N.Y. Acad. Sci. 721 (1994) 348.
[15] D. Sitkoff, K.A. Sharp, B. Honig, J. Phys. Chem. 98 (1994) 1978.
[16] P.F. Stanbury, A. Whitaker, S.J. Hall, Principles of Fermentation Technology, Pergamon/Elsevier Science Ltd., New York, 1995.
[17] P.-Å. Albertsson, Partition of Cell Particles and Macromolecules, Wiley, New York, 1986.