

Mathematical Modelling of Protein Precipitation Based on the Phase Equilibrium for an Antibody Fragment from *E. coli* Lysis

Yu Ji and Yuhong Zhou*

Department of Biochemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK

Abstract

Precipitation is an important operation in biopharmaceutical purification yet the mechanism of protein precipitation in multi-component solutions is not well understood. Existing models lack fundamental understanding of the process. In this paper, a new model describing how the protein solubility changes in the protein precipitation is proposed and is based on the phase equilibrium of the light liquid phase and dense solid phase. The model structure is generic and robust. It adequately reflects the non-linearity of protein precipitation kinetics and thus provides new fundamental insights into the protein precipitation in multi-component, complex protein solution.

Two feed stocks of a pure fragment antigen-binding (Fab') solution obtained by chromatographic purification and a clarified Fab' homogenate solution from *E. coli* were used to examine the effect of ammonium sulphate concentrations and pH conditions on precipitation. It was found that the model can describe pure Fab' precipitation well, and identify the non-ideal behavior of Fab' precipitation in multi-component homogenates. Through statistical analysis, the model parameters have been further reduced from 8 to 4. The quality of the model is such that errors were within the acceptable statistical confidence limits, even when applied to multi-component impurity precipitation. The new model with fewer parameters is better than existing empirical models in reflecting the salting-in and salting-out effect of the protein precipitation. This demonstrated that the structure of the model is sound and over-fitting in the parameter estimation is avoided. The model can be applied directly to industrial processes for protein precipitation process design after appropriate calibration with the required operating conditions of pH and salt concentration.

Keywords: Precipitation; Model; Multi-component protein solution; Antibody fragment; Impurity

Nomenclature

a_1, b_1, c_1, d_1	constants in Equation (9)
a_2, b_2, c_2, d_2	constants in Equation (18)
$a_3, b_3, c_3, d_3, e_3, f_3$	constants in Equation (19)
$a_4, b_4, c_4, d_4, e_4, f_4$	constants in Equation (20)
g_4, h_4, i_4	constants in Equation (20)
$a_5, b_5, c_5, d_5, e_5, f_5$	constants in Equation (21)
A, B	constants in Equation (12)
C_d	protein molar concentration in the dense phase ($\text{mol} \cdot \text{L}^{-1}$)
C_i	other component molar concentration in the solution ($\text{mol} \cdot \text{L}^{-1}$)
C_l	protein molar concentration in the light phase ($\text{mol} \cdot \text{L}^{-1}$)
C_s	salt molar concentration ($\text{mol} \cdot \text{L}^{-1}$)
C_T	the maximum protein concentration in the solution ($\text{mol} \cdot \text{L}^{-1}$)
I	ionic strength ($\text{mol} \cdot \text{L}^{-1}$)
k_s	salt activity coefficient (-)
k_i	components activity coefficient (-)
m_3	the salt mole concentration ($\text{mol} \cdot \text{L}^{-1}$)
r_i	protein activity coefficient in the light phase (-)
r_d	protein activity coefficient in the dense phase (-)
Q_i	molar concentration of ion i ($\text{mol} \cdot \text{L}^{-1}$)
R_g	ideal gas constant ($\text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$)
R^2	coefficient of determination
S	Fab' concentration in the supernatant ($\text{mol} \cdot \text{L}^{-1}$)
S_0	Fab' concentration in the feedstock ($\text{mol} \cdot \text{L}^{-1}$)
T	the absolute temperature (K)
V_l	light liquid phase volume (L)
V_d	dense phase volume (L)
V_T	total solution volume (L)

w_1, w_2 and w_3	constants in Equation (11)
Y	Fab's concentration in the supernatant
Z	net charge of the protein (-)
Z_i	charge number of ion i
$\alpha, \beta, \chi, \delta$	lumped constants in Equation (9)
θ, λ	constants in Equation (17)
K	lumped constants in Equation (15)
λ_{scaled}	scaled value
λ_{real}	real value
λ_l	real value at low limit
λ_u	real value at upper limit
μ_l	chemical potential for liquid phase ($\text{J} \cdot \text{mol}^{-1}$)
μ_d	chemical potential for dense phase ($\text{J} \cdot \text{mol}^{-1}$)
μ_l^i	protein standard chemical potential in the light phase ($\text{J} \cdot \text{mol}^{-1}$)
μ_d^i	protein standard chemical potential in the dense phase ($\text{J} \cdot \text{mol}^{-1}$)
μ_s^i	protein standard chemical potential in the solution ($\text{J} \cdot \text{mol}^{-1}$)
μ_w^i	protein standard chemical potential in the water ($\text{J} \cdot \text{mol}^{-1}$)
σ, ν, ρ, ξ	constants in Equation (16)
ϕ, ψ, γ, η	lumped constants in Equation (14)

*Corresponding author: Yuhong Zhou, University College London, Department of Biochemical Engineering, Torrington Place, London WC1E 7JE, UK, Tel: +44 (0)207 679 3815; Fax: +44 (0)207 916 3943; E-mail: y.zhou@ucl.ac.uk

Received June 06, 2013; Accepted July 21, 2013; Published July 26, 2013

Citation: Ji Y, Zhou Y (2013) Mathematical Modelling of Protein Precipitation Based on the Phase Equilibrium for an Antibody Fragment from *E. coli* Lysis. J Bioprocess Biotech 3: 129 doi: [10.4172/2155-9821.1000129](https://doi.org/10.4172/2155-9821.1000129)

Copyright: © 2013 Ji Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Protein precipitation is a technique that utilizes the differences of protein solubility to precipitate proteins into the solid phase from the liquid phase. It has been used extensively to separate and purify proteins for sample preparation [1]. Ammonium sulphate is usually used to separate protein from complex solutions because it does not denature protein and has a very high salting-out effect [2-4]. Currently, with advanced fermentation technology, higher protein titres can be achieved upstream and it is now possible to produce multi-kilogram quantities of therapeutic monoclonal antibodies in a single batch [5]. However, this creates problems at the downstream purification stages. The high concentration of target protein plus impurities in the feedstock changes the physical properties of the protein solution. If such a complex biological material is applied directly to the chromatographic columns, they are susceptible to fouling and blockages [6-8] so significantly increasing the chromatographic processing time and cost. Therefore, a primary separation, such as protein precipitation, may be beneficial in the preparation of a relatively clearer and less contaminated solution for expensive high resolution steps.

During precipitation, the solubility of a protein depends primarily on process conditions including pH, salt concentration and temperature [9]. In order to optimize the precipitation process operation, a good understanding of the impact of these conditions on the behavior of the protein is needed. For industrial scale process engineering and design purpose, a protein precipitation model that directly links protein solubility with operating conditions would help support industrial process development e.g. scale-up, predict process optimal conditions and provide information for process control [10].

The first attempt to model the protein solubility was by Cohn [11]. His log-linear equation, discussed later, gives a simple empirical relationship between the soluble protein concentration and ionic strength in the solution over a narrow salt concentration range. Melander and Horvath [12] then improved Cohn's empirical equation by linking the hydrophobic effect with thermodynamic parameters such as the hydrophobic surface. Unfortunately the improved model often sheds little light on the bioprocess operation and design as the linkage between the operating conditions and the hydrophobic surface cannot be established. The universal quasi chemical (UNIQUAC) model describes protein solubility by protein activity coefficients and a polynomial relationship between protein activity coefficients and osmotic second virial coefficients can be used to model protein behavior [13,14] where experiments to obtain protein activity data are required. The theoretical thermodynamic equations to predict protein solubility with molecular radius and surface parameters [15-17] worked quite well in a simple and defined system in which all physical properties are known. However, such thermodynamic-based models are of limited use for process design and control because the thermodynamic properties for complex multi-component processing materials are unknown.

Modified empirical exponential models that describe the traditional sigmoid shape of the precipitation curve directly link predictions with process conditions [18,19], but these models provide little fundamental understanding. Despite pH having been reported as a strong factor on protein precipitation, pH was not considered in these models. Temperature is another variable that often strongly influences protein precipitation. However, as most proteins are sensitive to temperature, a fixed temperature is applied during the industrial precipitation process, typically a low temperature (~ 4°C), to prevent protein denature.

The goal of this paper is to develop and validate a protein

precipitation model that uses bioprocessing conditions as inputs to predict the protein solubility for complex multi-component materials. The model will be based on theoretical phase equilibrium to achieve an improved process understanding.

Antibody fragments expressed intracellularly in *E. coli* as next generation therapeutics is cheaper to produce by fermentation than antibodies from mammalian cell culture because of shorter culture time and less expensive media. It also has better selectivity than antibodies but is more difficult to purify due to high level of impurities.

Two different feed stocks, a pure fragment of antibody (Fab') solution and a clarified Fab' solution from *E. coli* homogenate, will be used to examine the generality of the model. The model will be validated by experimental data, and then statistical tests will be used to evaluate the quality of the model. The predictions of the model will be compared with four existing models where pH will be introduced as an extra variable [11,18,19].

Materials and Methods

Sodium monobasic phosphate, sodium dibasic phosphate, sodium acetate and ammonium sulphate were purchased from Sigma Chemical Co. Ltd. (Dorset, UK). All chemicals were reagent grade. Fab' producing strain *E. coli* W3110 was kindly provided by UCB (Slough, UK) and the cell paste was provided by the Fermentation Group, Department of Biochemical Engineering in University College London.

Precipitation material preparation

E. coli cells were suspended in a 10 mM pH 7.0 phosphate buffer at 40% (wt) and homogenized in an APV Lab 40 Homogenizer at 750 bar. The homogenized solution was centrifuged in an Eppendorf Centrifuge 5810R at 12,000 rpm for 2 hours with supernatant collected as the stock solution for further study. Pure Fab' solution was prepared from the collected supernatant. An AKTA Basic HPLC (GE Healthcare, Sweden) and Hitrap Mabsselect 5 ml HPLC column (GE Healthcare, Sweden) were used to purify Fab'. The eluate was buffer exchanged to 10 mM pH 7.0 phosphate buffer and stored at 4°C.

Fab' and impurity concentration analysis

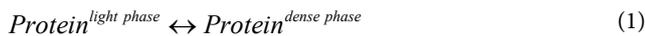
Fab' concentration and total protein concentration were analyzed by HPLC Agilent 1200 (Agilent Technologies, UK) with a Hitrap Protein G 1ml HPLC column (GE Healthcare, Sweden). Fab' concentration was calculated from the peak area according to a calibration curve, which was obtained using pure Fab' after Protein A and size exclusion chromatography. The impurity concentration analysis method was the same as Fab' except that the feedstock was used as the standard and the impurity area monitored.

Fab' precipitation by microwell scale high throughput experimentation

The Fab' precipitation was carried out in the ABgene's 96 deep microwell plates by a Packard MultiPROBE II HT EX (Packard BioScience Company, Meriden, U.S.A.). The experimental conditions were selected as follows: pH from 4.5 to 8.0 with intervals of 0.5, ammonium sulphate concentration from 0 to 3.0 mol/L, with intervals of 0.2 mol/L for pure Fab' solution and with intervals of 0.3 mol/L for clarified homogenized solution. The total volume of precipitate supernatant was 1.8 ml. The precipitation plate was shaken on an Eppendorf thermomixer at 600 rpm for 2 hours and then centrifuged at 4000 rpm for 15 minutes. The clear supernatant was transferred to an Agilent 96 HPLC microwell plate and analyzed in triplicate.

Hybrid model derivation and parameter estimation

Phase equilibrium based protein precipitation model: Protein precipitation has been thermodynamically regarded as a pure crystallisation process because the solution has only protein and salt [15-17]. However, for proteins in a real fermentation broth or with other complex biological materials, the precipitation will not form pure crystal but an amorphous mixture [20,21]. The precipitation can be treated as a distribution between a light liquid phase (supernatant) and a dense liquid phase (precipitate). Therefore, the proposed model in this paper is based on the phase equilibrium for the target protein in a multi-component solution:



When the two phases are in equilibrium, the chemical potentials of the proteins must be equal:

$$\mu_l = \mu_d \quad (2)$$

and

$$\mu_l^0 + RT \ln C_l r_l = \mu_d^0 + RT \ln C_d r_d \quad (3)$$

where μ_l is the chemical potential for the liquid phase and μ_d the chemical potential for the dense phase, C_l the protein mole concentration in the light phase, C_d the protein mole concentration in the dense phase, r_l the protein activity coefficient in the light phase, r_d the protein activity coefficient in the dense phase, R the ideal gas constant, T the absolute temperature, μ_l^0 the protein standard chemical potential in the light phase and μ_d^0 the protein standard chemical potential in the dense phase.

Equation (3) can be rearranged to

$$C_l/C_d = (r_d/r_l) \exp(-(\mu_l^0 - \mu_d^0)/RT) \quad (4)$$

Suppose V_l is the light liquid phase volume, V_d the dense phase volume, C_T the maximum protein concentration in the solution and V_T the total solution volume with the approximation that there is no volume change during precipitation. Then,

$$V_T = V_l + V_d \quad (5)$$

$$C_l V_l + C_d V_d = C_T V_T \quad (6)$$

Introducing equations (5) and (6) into equation (4) we obtain

$$C_l/C_T = 1 / \left(\frac{V_l}{V_T} + \left(\frac{V_d}{V_T} \right) (r_l/r_d) \exp(-(\mu_l^0 - \mu_d^0)/RT) \right) \quad (7)$$

The dense phase volume in protein precipitation cases will increase with salt concentration as more proteins will be precipitated and reach a nearly constant level at high salt concentration. It is often very small compared to the total solution volume because the total protein concentration is low in biopharmaceutical processing material, so it is reasonable to assume that $V_l/V_T \approx 1$. As we know, V_d/V_T depends on the protein properties and pH probably with an apparent isoelectric point (pI). Our preliminary experimental results shown in Figure 1 illustrate that the Fab' concentration increased linearly with pH in the range 4-8. Hence pH having a linear effect was approximated and the effects of salt on V_d/V_T followed similarly to a Michaelis-Menten relationship. Thus, the empirical equation to represent the effect of pH and salt concentration on V_d/V_T is proposed as:

$$V_d/V_T = a_1 + b_1 (pH - c_1) C_s / (d_1 + C_s) \quad (8)$$

or

$$V_d/V_T = (\alpha + \beta C_s + \chi pH C_s) / (\delta + C_s) \quad (9)$$

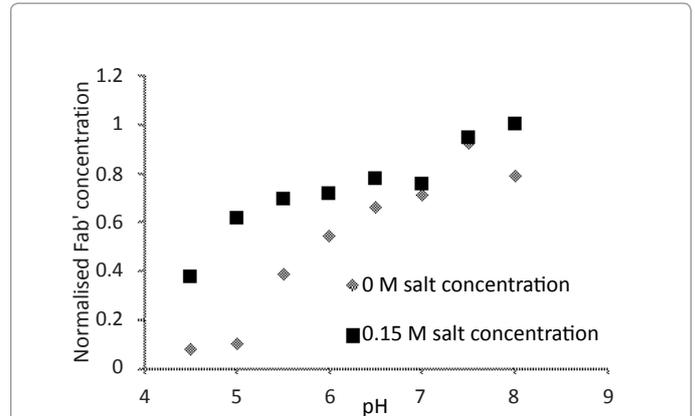


Figure 1: Preliminary result showing the linear dependence between Fab' protein precipitation and pH.

where C_s is the salt concentration, a_1, b_1, c_1, d_1 are constants, and $\alpha = a_1 d_1, \beta = a_1 - b_1 c_1, \chi = b_1, \delta = d_1$, are the lumped constants. With different protein solutions, these lumped parameters may vary and hence should be estimated from real experimental data.

In 1943, Kirkwood [22] defined the protein activity coefficient in a multi-component solution as a simple function of the concentrations of all solute species. Long and Mcdevit [23] assumed that the protein activity coefficient can be represented by a log-linear function based on fundamental chemical thermodynamics:

$$\log r_p = k_s C_s + \sum k_i C_i \quad i=1, \dots, n, \quad (10)$$

where k_s is the salt activity coefficient, k_i the components activity coefficient and C_i the other component concentrations in the solution.

In a multi-component solution containing biomolecules and salt, the protein activity coefficient is dominated by the salt concentration; the other effects caused by biomolecules can be regarded as constant due to their very low concentrations. Therefore in the liquid phase, the second part of equation (10) can be represented by a constant. In the dense phase, the concentration of salt is considered as not significantly changing while the other molecules still have no or little effect; hence the overall protein activity coefficient can be regarded as a constant. Therefore

$$r_l/r_d = \exp(k_s C_s + w_1) / \exp(w_2) = \exp(k_s C_s + w_3) \quad (11)$$

where w_1, w_2 and w_3 are constants.

In some cases, the protein property and its main interaction with salt will depend on the type of salt, so equation (11) may need a second order or even a higher order term of the salt concentration [23]. In this study, only the first order term was used in the model. A higher order model needs only to be considered if the first order model fails.

In 1985, Arakawa and Timasheff [20] published a theoretical protein precipitation model, which gave the following theoretical chemical potential equation:

$$(\mu_2^0 - \mu_{2,w}^0) / RT = AZ^2 (I)^{1/2} (1 + B(I)^{1/2}) - (1/2.303RT) \int_0^m (\partial \mu_2 / \partial m_3)_{T, m_2} dm_3 \quad (12)$$

where μ_2^0 is the protein standard chemical potential in the solution, $\mu_{2,w}^0$ the protein standard chemical potential in the water, Z the net charge of the protein, I the ionic strength, m_3 the salt mole concentration with A and B the coefficients. The second differential

term can be approximated by a first order term of salt concentration because $(\partial\mu_i/\partial m_i)_{T,m_2}$ is an empirical constant over a wide range of salt concentrations [20]. The temperature was regarded as a constant in this study.

The ionic strength is defined as:

$$I = (1/2) \sum Q_i (Z_i)^2 \quad (13)$$

where Q_i is the molar concentration of ion i , and Z_i is the charge number of that ion. For a neutral salt such as ammonium sulphate the ionic strength is linearly proportional to salt concentration. As there is no general mathematical model for protein surface net charge as a function of pH, we will approximate Z^2 in equation (12) by a second order pH polynomial equation. Therefore, equation (12) can be simplified into a function of salt concentration and pH by:

$$(\mu_i^0 - \mu_{2,w}^0)/RT = \phi(pH - \gamma)^2 (C_s)^{1/2} / (\eta + (C_s)^{1/2}) + \phi C_s \quad (14)$$

where ϕ , ϕ , γ , η are lumped constants.

Under the assumption that the salt concentration in the dense phase is very small and does not change significantly, the value of equation (14) for dense phase protein will be considered as a constant, so

$$(\mu_i^0 - \mu_d^0)/RT = \phi(pH - \gamma)^2 (C_s)^{1/2} / (\eta + (C_s)^{1/2}) + \phi C_s + \tau \quad (15)$$

where τ is a lumped constant.

The second term essentially describes the protein salting-in effect at low salt concentration. To simplify the calculation, the second term is approximated by the pH effect for the low concentration range because from our experiments the pH effect dominated at low salt concentration. It was then described by a simplified second order polynomial function, while the salt effect was separated from this term and lumped into the first term on the right hand side of equation (15) to give:

$$(\mu_i^0 - \mu_d^0)/RT \approx \sigma pH^2 + \nu pH + \rho + \xi C_s \quad (16)$$

where σ , ν , ρ , ξ are constants. At high salt concentration the salting-in phenomena does not occur or the effect is very small compared to the first term in equation (15). Therefore, the coefficients in equation (16) were relatively insignificant for the overall model prediction at a high concentration range. It also kept the model mathematically consistent throughout the salt range.

Combining equations (9), (11) and (16), equation (7) becomes:

$$C_i/C_T = 1 / (1 + ((\alpha + \beta C_s + \chi pH C_s) / (\delta + C_s)) \exp(k_s C_s + w_3 + \sigma pH^2 + \nu pH + \rho + \xi C_s)) \quad (17)$$

or

$$C_i/C_T = 1 / (1 + ((\alpha + \beta C_s + \chi pH C_s) / (\delta + C_s)) \exp(\theta C_s + \nu pH + \sigma pH^2 + \lambda)) \quad (18)$$

where $\theta = k_s + \xi$, $\lambda = w_3 + \rho$

This model is able to describe the strong non-linearity of the precipitation surface due to its sigmoid structure. All the parameters in equation (18) are lumped and so it is difficult to predict their values or limit their ranges. However, according to the modelling assumptions, parameters θ and δ should have positive values and the term $\alpha + \beta C_s + \chi pH C_s$ will be positive. At low salt concentrations, the exponential expression in the model is not a dominant effect and thus the decrease of dense phase volume caused by the salting-in effect, which makes the value of V_d/V_T smaller, explains the protein concentration increase. At high salt concentration, the second term which contains an exponential expression will have a value much larger than 1.0, so we can neglect the value of 1.0 in the denominator. Thus this model is similar to the exponential structure of Cohn's equation, discussed below.

This model has a thermodynamic base assisted by empirical relationships. The model structure involves eight parameters to allow a proper expression of the solubility surface. This is different from models from purely experimental observation which are too simplified and so loose the necessary details.

Model comparison: In order to evaluate the capability of this new model, it is useful to compare it with three published models, Cohn's [11], Niktari's [18] and Habib's [19] models plus a polynomial model. For process design purposes, all the selected models were modified to contain a pH factor by introducing a second order polynomial expression of pH to substitute for the model coefficients without changing the model structure, in order to link protein solubility directly to operating variables.

Cohn's equation is an empirical equation. In Cohn's papers, it has been shown that the second constant is only associated with the salt effect and the first constant is associated with all of the other effects, therefore the second constant d is not related to pH. Hence, the Cohn's equation will only be expanded in the first constant:

$$\ln(S/S_0) = a_2 + b_2 pH + c_2 pH^2 - d_2 C_s \quad (19)$$

where S/S_0 is the ratio of Fab' concentration in the supernatant to the Fab' concentration in the feedstock. The expanded Niktari's equation is:

$$y = 1 / (1 + (C_s / (a_3 + b_3 pH + c_3 pH^2))^{d_3 + e_3 pH + f_3 pH^2}) \quad (20)$$

The expanded Habib's sigmoid model is:

$$y = 1 - (a_4 + b_4 pH + c_4 pH^2) (1 - \exp(-(C_s / (d_4 + e_4 pH + f_4 pH^2))^{g_4 + h_4 pH + i_4 pH^2})) \quad (21)$$

A second order polynomial equation with interaction terms is used to represent the conventional two factors polynomial model:

$$y = a_5 + b_5 C_s + c_5 pH + d_5 C_s^2 + e_5 pH + f_5 C_s pH^2 \quad (22)$$

where y is Fab' concentration in the supernatant, and others are parameters.

Although the number of parameters involved in these models is similar, the structure of our model is very different.

Data analysis and parameter estimation: In order to eliminate the errors in the model parameter estimation caused by different orders of variables, the experimental conditions of pH and salt concentration were normalized to a 0-1 range by the scaling:

$$\lambda_{scaled} = (\lambda_{real} - \lambda_L) / (\lambda_U - \lambda_L) \quad (23)$$

where λ_{scaled} is the scaled value, λ_{real} the real value, λ_L the real value at low limit and λ_U the real value at upper limit.

The Fab' concentration and the impurity concentration in both feed stocks were also normalized. The initial concentration was not preferred in the normalization because there was a salting-in effect. The maximum concentration of Fab' and impurities during salting-in were regarded as the true maximum protein concentration in the solution. In reality, this concentration was difficult to obtain so the maximum concentration from the experimental data was used as the approximation to the true maximum concentration, and also as the upper limit for equation (22). The model parameters in equations (18) - (22) were estimated by the nonlinear least-squares regression method in the Matlab Tool box (MathWorks, USA).

Results and Discussion

The Fab' solubility in pure Fab' solution and clarified homogenate

was described by equation (18) with regards to pH and ammonium sulphate concentration. As the focus of this study is on the development of an accurate precipitation model, a large data set is needed to test the model accuracy and regression fitting. A nonlinear least squares method was used to estimate the parameters in the model. The accuracies of the model, equation (18), and the other models, equations (19), (20), (21) and (22), were measured by R^2 . In addition, the statistical F-test was used to evaluate the new model. The Wilcoxon test and paired t-test for individual parameters in the model were used to validate the model.

Protein precipitation modelling for pure Fab', Fab' in clarified homogenate and impurities

119 experiments for pure Fab' solution and 79 experiments for clarified homogenate were carried out. The pure Fab' concentration, as shown in Figure 2a, slightly increased during the salting-in phase and then gradually decreased with salt concentration, similar to many previous pure protein precipitation curves[11, 20]. For the clarified homogenate shown in Figure 2b, when the salt concentration was low, the Fab' concentration was significantly affected by other components

in the solution and its concentration was altered compared to that of pure Fab' solution. Under low pH and at low salt concentration, the Fab' concentration was only 40% of the highest concentration. The same phenomena occurred with the impurity solubility, shown in Figure 2c. An explanation may be that the low Fab' concentration at certain conditions, while it was not a pure solution, was probably caused by the co-precipitation between the Fab' molecule and other impurity proteins, the solubility of which were significantly changed by pH at low ionic strength. The effect of salting-out dominated when the salt concentration increased and thus the solubility was nearly the same as that of pure Fab' solution.

These experimental data sets were then used to develop the pure Fab' precipitation model based on equation (18). The estimated parameters are shown in Table 1 and the R^2 value was 0.975. The F-test value of the model fitting was 624.81; indicating 95% confidence model accuracy was achieved. It is clear that the model prediction satisfactorily described both the salting-in and salting-out features of the concentrations without the cost of losing accuracy for any phases.

The generality of the model structure was then assessed by applying the model to Fab' precipitation in a clarified homogenate where multi-components exist. The parameters of the pure Fab' model were used as the initial guess for the parameter estimation. The R^2 value for Fab' in clarified homogenate was 0.972 as shown in Table 1, which was very similar to that of the pure Fab' model. The F-test value of this model was 320.36, which was smaller than that in the pure Fab' model. Nevertheless, this suggests that the model was accurate to within 95% confidence. The predicted Fab' concentration surface is shown in Figure 2b. It can be seen that the Fab' solubility in the clarified homogenate was in general predicted well by the model. Nevertheless, there is a slight discrepancy between predicted solubility and experimental data at a low pH range as well as very low salt concentrations.

We also assessed the model by applying it to impurity precipitation, where a mixture of proteins was treated as an assumed pseudo-single molecule with average characteristics of all the proteins in the solution, e.g. average electronic charges and hydrophobic behaviour. 79 experimental data points from an impurity precipitation in a clarified homogenate were used for parameter estimation. The results are given in Table 1 and the R^2 value was 0.945, which is slightly lower than that of the Fab' models. The F-test value of the model was 172.60. These measures showed that the model was accurate to 95% confidence. The predicted impurity concentration surface is shown in Figure 2c. The geometrical pattern for the real data points and the model predicted surface were slightly different, especially in the high salt concentration region. We believe the difference was caused by the simplification of a mixture of proteins into a pseudo-single protein. Although the impurities were regarded as a pseudo-single molecule with an average value of all the mixture, in reality, different proteins will have a different sensitivity to pH and salt concentrations. Conditions may significantly affect one protein with no great effect on other molecules. Thus, the

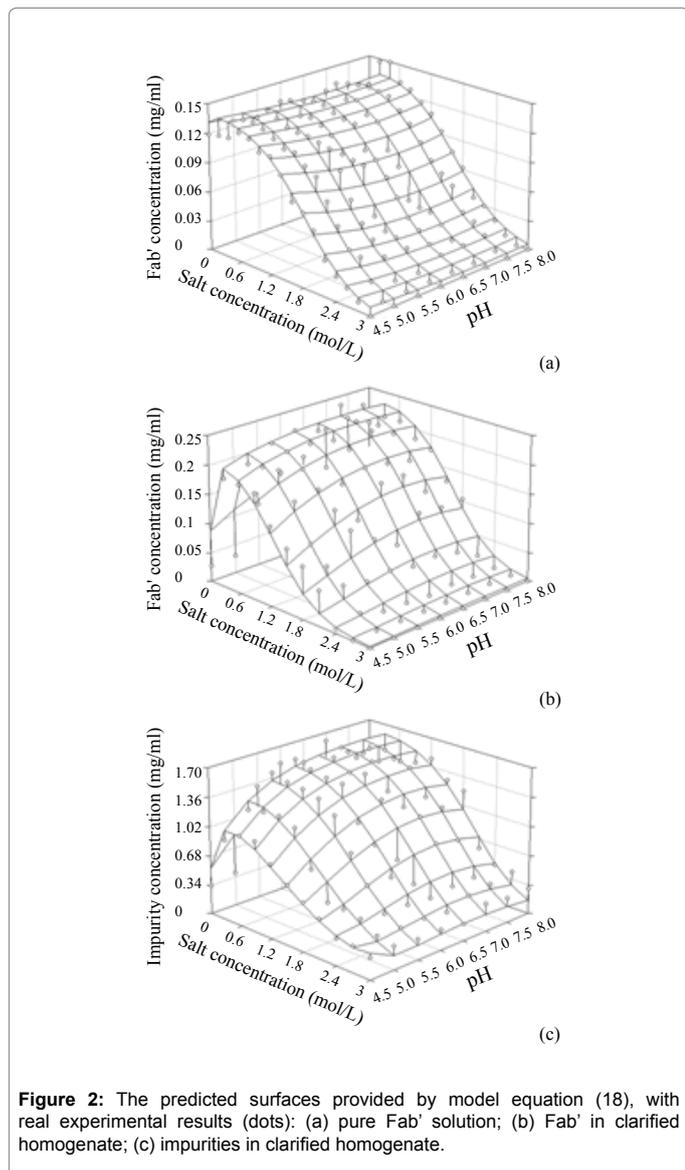


Figure 2: The predicted surfaces provided by model equation (18), with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

	Parameters								R^2	F-test value
	θ	ν	σ	δ	α	β	χ	λ		
Pure Fab'	7.97	1.62	0.53	-5.30	1.15	-1.05	0.16	0.03	0.975	624.81
Fab' in clarified homogenate	7.61	-6.60	4.34	-4.27	1.04	-0.75	21.12	0.002	0.972	320.36
Impurities	5.51	-7.88	5.03	-3.43	1.32	-0.41	20.60	0.01	0.945	172.68

Table 1: Parameters, F-test value and R^2 value of developed model using equation (18) for pure Fab' precipitation, Fab' precipitation in clarified homogenate and impurity precipitation.

assumed average properties of a pseudo-single molecule are not constant under all conditions, especially for extreme conditions, e.g. low pH or high salt concentration.

Model modification

It has been shown in Figures 2 that the new model predictions agreed well with the experimental data. However, there are eight parameters in equation (18) and the model exhibits a high level of nonlinearity. A simpler model is more useful for the processes operation and design. It is also beneficial for the parameter estimation since a higher number of parameters have the potential to over-fit the data which could result in less accuracy of the model. The t-test for individual parameter in the model can be used to evaluate and decide if a parameter is necessary. If a parameter fails the t-test, it is either inaccurate or not needed, or both. After carrying out the t-test for each parameter in model (18), the parameters α , β , and χ failed in all three models, i.e., pure Fab', Fab' in clarified homogenate and impurities. Therefore a simplified model is proposed:

$$C_i / C_T = 1 / (1 + (1 / (\delta + C_s)) \exp(\theta C_s + \nu pH + \sigma pH^2 + \lambda)) \quad (24)$$

The simplified models were developed for pure Fab', Fab' in clarified homogenate and impurities in clarified homogenate based on equation (24) by using the experimental data sets again. The parameters are shown in Table 2. The R^2 values were evaluated and were a little lower than previous, but all tests showed that the models had excellent statistical confidence. All parameters passed t-tests with 95% confidence.

Using the values in Table 2, all three models showed similar predictions as in Figure 2. When the salt concentration is low (<0.2 mol/L), the value of the exponential term is small and changes little while the linear term dominates and changes rapidly. It describes the salting-in phenomenon at low salt concentration better. When the salt concentration is high, e.g. in the salting-out range, the value of the exponential term dominates due to the large values of parameter θ , while the effect of the linear term is small. The values of the parameters associated with pH vary relatively little. The parameter of the second order pH term for impurities, σ , is the largest with the fact that impurity concentration was influenced most by the pH in all three materials. The effect of pH around the neutral point is very small due to its small parameter value and its second order structure. However, according to the models there will be large effects describing the protein concentration changes in the experiments at the extreme pH conditions.

The most difficult operation conditions in precipitation to determine are the cutting points either in salting-in at the low salt concentration and salting-out at the high salt concentration ranges. As equation (24) is derived from thermodynamic phase equilibrium, the term $1/(\delta + C_s)$ is strongly related to the salting-in effect, and δ significantly influences the magnitude of the salting-in effect. Term $(1/\delta)\exp(\lambda)$ indicates the potential increase in protein concentration. As shown in Table 2, $(1/\delta)$ values in the models of Fab' in clarified homogenate and impurity are much larger than that of pure Fab', so the salting-in effect is significant

	Parameters					R^2	F-test value
	θ	ν	σ	δ	λ		
Pure Fab'	8.21	1.49	-1.08	-5.34	0.03	0.973	1037.82
Fab' in clarified homogenate	9.46	-3.16	2.19	-4.60	0.005	0.937	308.28
Impurities	7.02	-5.01	3.70	-3.51	0.02	0.914	195.48

Table 2: Parameters, F-test value and R^2 value of simplified model using equation (24) for pure Fab' precipitation, Fab' precipitation in clarified homogenate and impurities precipitation.

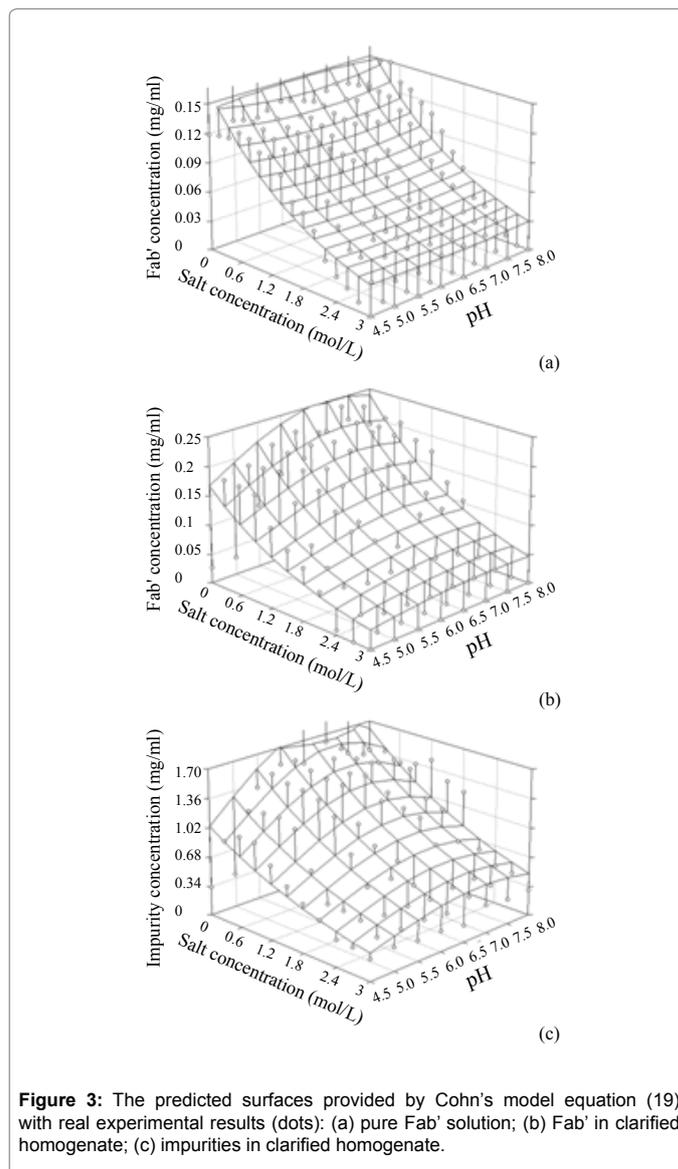


Figure 3: The predicted surfaces provided by Cohn's model equation (19) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

in these multi-component cases. During salting-out, χ , the coefficient for the interaction between pH and salt, and θ , the coefficient for the salt, play an important role. In the clarified homogenate case, the values of χ in the Fab' model and the impurity model are nearly two hundredfold of that in the pure Fab' model, showing that the impact of multi-components on the salting-out are significant. Together with λ , these three parameters dominate the salting-out effect.

Model validation

Nine independent experiments were carried out and the results used to validate the model. When validating bioprocess models, it is not recommended to use error percentage to evaluate the models because the range of bioprocess data may be very wide, even after scaling, which will introduce mathematical errors. Thus, statistical tests are needed to validate any new model, no matter how good the fitting of the data is in the regression step.

There are however several unusual problems for bioprocess model validation. First of all, the number of samples used for validation is

normally small, e.g. nine samples in this case, due to the high cost and long experimental time as well as limited protein solution materials available. Secondly, the distribution of most bioprocess data is normally unknown or the data does not conform to any known distribution, e.g. standard normal distribution. Statistically, normal distribution can only be assumed when the number of samples is large, normally more than 30. Therefore, for a small validation group with an unknown distribution, it is risky to use the paired t-test due to a high probability to fail. Two possible solutions are a Wilcoxon signed-rank test [24] for few samples or to analyse validation samples together with previous regression data by the paired t-test. For a Wilcoxon test, when 2-tailed significance > 0.05, it is regarded as a validation pass. A paired t-test, when significance (the p-value associated with the correlation) > 0.05, can be considered as the null hypothesis, i.e. no difference between the experimental data and model calculated value. Table 3 shows the test results for modified model equation (24) with paired t-test significance > 0.05 and Wilcoxon 2-tailed significance > 0.05 from the Statistical Package for the Social Sciences (SPSS, an IBM Company) calculated for all three materials. It demonstrates that the simplified model passed the validation criteria with strong statistical confidence.

Model comparison

The experimental data sets were also used to estimate the parameters for the four models described by equations (19) to (22). Table 4 presents the R^2 and the F-test values for all four models in different feed stocks. The predicted Fab' and impurity concentration surfaces for the four models are shown in Figures 3 to 6. A large difference between predicted surface and experimental data was observed for all three models based on Cohn's equation in Figures 3a-3c. However as all statistical tests failed with all R^2 values below 0.80, Cohn's equation failed to describe the protein salting-in effect at low salt concentration in multi-component solution as well as salting-out effect due to its linear model structure. Niktari's expansion model and Habib's expansion model can fit quite well for pure Fab' precipitation, as shown in Figure 4a and Figure 5a, with R^2 values of 0.956 and 0.971, F-test values of 402.72 and 400.30 respectively. However, the R^2 for the Fab' precipitation model in multi-components solution and impurity precipitation model were decreased below 0.9 and failed to describe the salting-in effect, as seen in Figure 4b, 4c and Figure 5b, 5c. The R^2 values and F-test values of these two models show poor model fitting and less predicting capability. Besides, both models do not consider the pH effect originally and a second-order polynomial expression for pH expansion may not effectively describe the pH impact.

Theoretically, the polynomial model has the most flexibility to fit

	Pure Fab'	Fab' in clarified homogenate	Impurities
t-test value	1.201	1.220	-0.198
sig. (2-tailed)	0.232	0.227	0.843
Wilcoxon 2-tailed sig.	0.767	0.086	0.515

Table 3: Results with 9 samples Wilcoxon signed-rank test and all samples paired t-test results for modified model using equation (24).

	Our model		Cohn's equation		Niktari's model		Habib's model		Polynomial model	
	R^2	F-test	R^2	F-test	R^2	F-test	R^2	F-test	R^2	F-test
Pure Fab'	0.973	1037.82	0.795	148.53	0.956	402.72	0.971	400.30	0.950	433.74
Fab' in clarified homogenate	0.937	308.28	0.621	40.88	0.795	46.55	0.864	48.78	0.825	68.90
Impurities	0.914	195.48	0.721	64.50	0.859	73.03	0.877	54.48	0.858	88.00

Table 4: R^2 and F-test values for all four existing models.

the data. However, the same results were obtained as for the previous two models as shown in Figure 6. The model in pure Fab' precipitation was quite good, with R^2 of 0.95, but the models for Fab' precipitation in multi-components were less good, as R^2 was less than 0.90 for this non-ideal solution. Moreover, at high salt concentration, the predicted concentration surface of the polynomial model was below zero, which also occurred in Habib's model. Therefore, both models were quite misleading as the predicted value below zero had no physical meaning. The negative value can be manually eliminated by constraining the parameters during regression but at the cost of overall accuracy. The polynomial model was not considered adequate for fitting in clarified homogenate precipitation. The higher order polynomial model may be more accurate but will inevitably introduce more parameters.

The comparison studies demonstrate that the structure of the model is crucial. If the structure is not right, the model will not predict the performance well even though there are high number of parameters in the model e.g. 9 parameters in Habib's model.

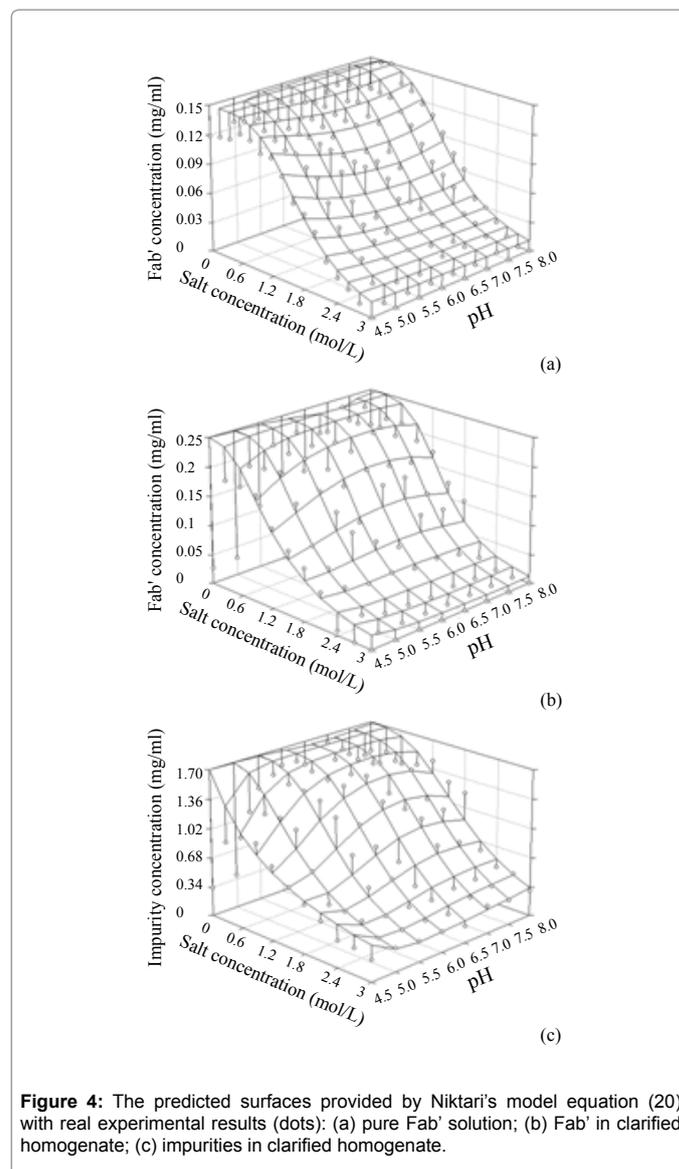
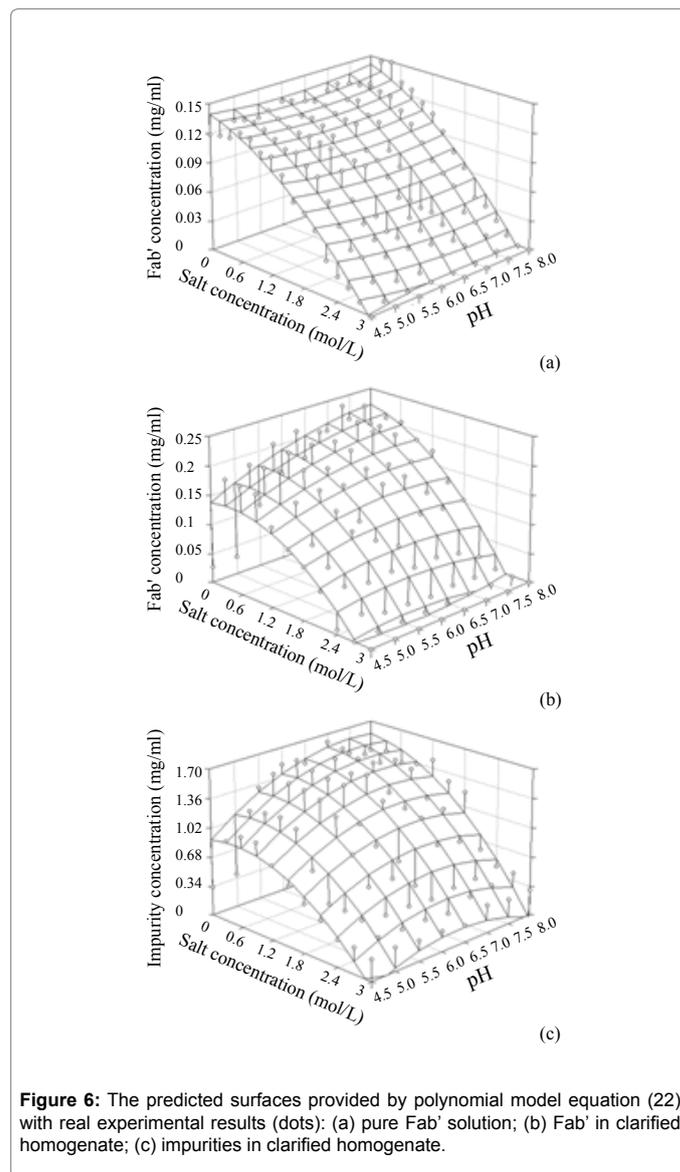
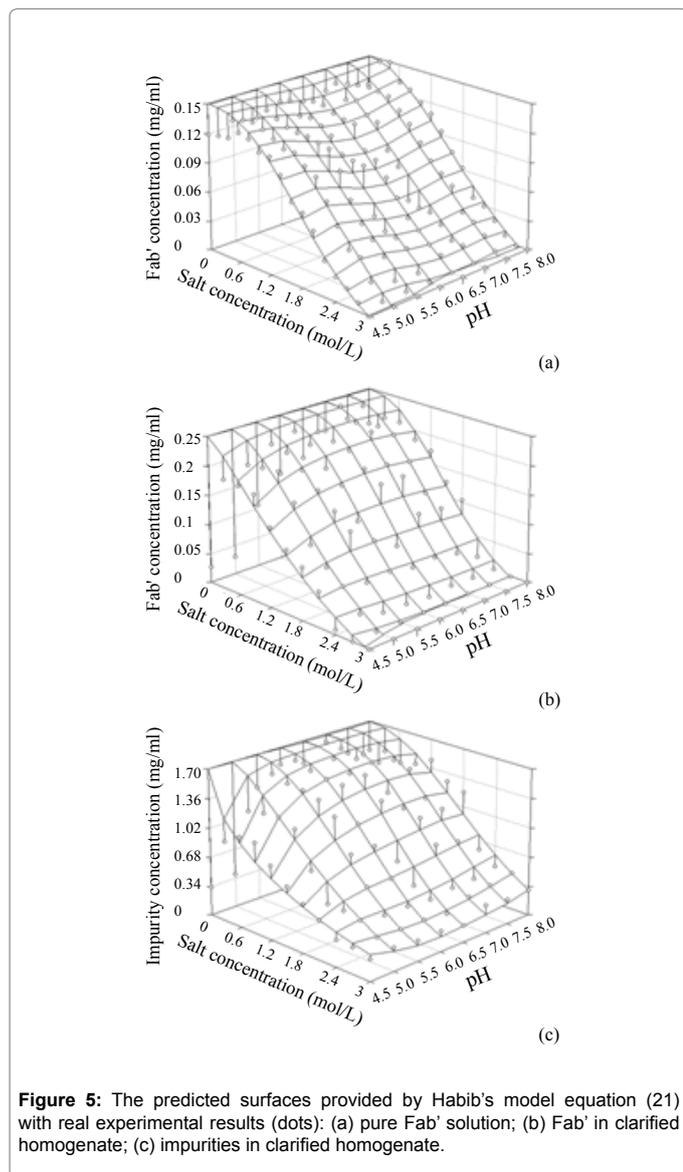


Figure 4: The predicted surfaces provided by Niktari's model equation (20) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.



Conclusions

Phase equilibrium based models have been developed then validated using statistical tests. The model equation (24) can precisely describe the precipitation salting-in and salting-out effects for pure Fab', Fab' in clarified homogenate and impurity with salt concentration and pH change. The model structure, based on a single protein, takes the multi-component factor into consideration and can be applied to a protein in a multi-component solution or a pseudo-single molecule. The estimated parameters in the model passed regression statistical tests proving that the models were accurate with 95% confidence.

Compared to thermodynamic based models, this new model conveniently predicts the precipitation results from operation conditions rather than thermodynamic parameters. The comparison between this new model and four existing empirical models showed that it was superior backed with the statistical tests. As our model structure is derived from fundamental phase equilibrium, it exhibits good prediction even though there are only 4 parameters. Such studies also show the challenge in multi-component protein precipitation

modeling when strong nonlinearity exists. Our model structure is able to reflect the non-linearity of the protein precipitation in both salting-in and salting-out better than existing empirical models.

Acknowledgements

This research was funded by the UK Government Overseas Research Student Scholarship and a University College London Graduate Student Research Scholarship. The provision of *E. coli* W3110 by UCB and cell paste from the upstream fermentation group, Department of Biochemical Engineering, UCL, was highly appreciated.

References

- Shahina S, Dalby PA (2011) Thermodynamic parameters for salt-induced reversible protein precipitation from automated microscale experiments. *Biotechnol Bioeng* 108: 322-332.
- Cheng YC, Lobo RF, Sandler SI, Lenhoff AM (2006) Kinetics and equilibria of lysozyme precipitation and crystallization in concentrated ammonium sulfate solutions. *Biotechnol Bioeng* 94: 177-188.
- Cohn EJ, Strong LE, Hughes WL, Mulford DJ, Ashworth JN, et al. (1946) Preparation and Properties of Serum and Plasma Proteins IV. A System for

- the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids. *J Am Chem Soc* 68: 459-475.
4. Foster PR, Dunnill P, Lilly MD (1976) The Kinetics of Protein Salting-Out: Precipitation of Yeast Enzymes by Ammonium Sulfate. *Biotechnol Bioeng* 18: 545-580.
 5. Aldington S, Bonnerjea J (2007) Scale-up of monoclonal antibody purification processes. *J Chromatogr B* 848: 64-78.
 6. Shukla AA, Hubbard B, Tressel T, Guhan S, Low D (2007) Downstream processing of monoclonal antibodies-application of platform approaches. *J Chromatogr B Analyt Technol Biomed Life Sci* 848: 28-39.
 7. Gottschalk U (2008) Bioseparation in antibody manufacturing: the good, the bad and the ugly. *Biotechnol Prog* 24: 496-503.
 8. Jiang C, Liu J, Rubacha M, Shukla AA (2009) A mechanistic study of Protein A chromatography resin lifetime. *J Chromatogr A* 31: 5849-5855.
 9. Knevelman C, Davies J, Allen L, Titchener-Hooker NJ (2010) High-throughput screening techniques for rapid PEG-based precipitation of IgG4 mAb from clarified cell culture supernatant. *Biotechnol Prog* 26: 679-705.
 10. Bernaerts K, Van Impe J (2004) Data-driven approaches to the modelling of bioprocesses. *T I Meas Control* 26: 349-372.
 11. Cohn EJ (1925) The Physical Chemistry of the Proteins. *Physiol Rev* 5: 349-437.
 12. Melander W, Horvath C (1977) Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: An interpretation of the lyotropic series. *Arch Biochem Biophys* 183: 200-215.
 13. Ruppert S, Sandler S, Lenhoff AM (2001) Correlation between the osmotic second virial coefficient and the solubility of proteins. *Biotechnol Prog* 17: 182-187.
 14. Winzor DJ, Carrington LE, Harding SE (2001) Analysis of thermodynamic non-ideality in terms of protein solvation. *Biophys Chem* 93: 231-240.
 15. Avena SM, Bogle ID, Pessoa FL (1997) An activity coefficient model for proteins. *Biotechnol Bioeng* 55: 65-71.
 16. Chiew YC, Kuehner D, Blanch HW, Prausnitz JM (1995) Molecular thermodynamics for salt-induced protein precipitation. *AIChE J* 41: 2150-2159.
 17. Kuehner DE, Blanch HW, Prausnitz JM (1996) Salt-induced protein precipitation: Phase equilibria from an equation of state. *Fluid Phase Equilib* 116: 140-147.
 18. Niktari M, Chard S, Richardson P, Hoare M (1990) The Monitoring and Control of Protein Purification and Recovery Processes. *Separations for Biotechnology* 2: 622-634.
 19. Habib G, Zhou Y, Hoare M (2000) Rapid monitoring for the enhanced definition and control of a selective cell homogenate purification by a batch-flocculation process. *Biotechnol Bioeng* 70: 131-142.
 20. Arakawa T, Timasheff SN (1984) Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* 23: 5912-5923.
 21. Shih YC, Prausnitz JM, Blanch HW (1992) Some characteristics of protein precipitation by salts. *Biotechnol Bioeng* 40: 1155-1164.
 22. Kirkwood JG (1943) The theoretical interpretation of the properties of solutions of dipolar ions: Protein, amino acids and peptides as ions and dipolar ions. Cohn EJ, Edsall JT, Reinhold Publishing, New York, USA.
 23. Long FA, McDevit WF (1952) Activity Coefficients of Nonelectrolyte Solutes in Aqueous Salt Solutions. *Chem Rev* 51: 119-169.
 24. Wilcoxon F (1945) Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1: 80-83.