Modeling and calibration of preparative chromatography in gPROMS

Separations of protein monomers from dimers

Marcus Almqvist
Department of Chemical Engineering
Lund University, Sweden
January 2009

Abstract

A successful setup for calibration of a model for separation of bovine serum albumin monomer and its dimers is described. The model is calibrated using 3 different data sets and thereafter validated against two different experimental data sets. The objective function for the calibrations was based on a maximum likelihood approach, with a least square method for the first calibration and a constant variance method for the other calibration. When using gPROMS for fine calibration the results will be very good and the model will be able to predict other experiment within the same pH with high accuracy. It has been proven that some parameters are insensitive to the calibration of the data sets.

Keywords: BSA, Langmuir MPM, Homogenous, Calibration, Ion exchange.

Introduction

In this paper, a simulated model for adsorption of bovine serum albumin on a HP column is calibrated to experimental data on a homogenous, Langmuir mobile phase modulator, model. The modeling and calibration was done in gPROMS with 8 different parameters.

Theory

Ion exchange chromatography, IEC, is a chromatography technique where charged molecules are separated due to differences in ionic binding properties. The surface of the matrix is covered by charged ligands and the ligands have counter ions bound to them. The molecules, for example proteins, may have a stronger affinity for the ligands and the counter ions are disposed of and the proteins will bind to the ligands. The displaced counter ions will give rise to a higher salt concentration and molecules with the same charge will follow the mobile phase out of the column. Since proteins are large molecules it might be that the same protein is bound to several binding sites or even cover unused binding sites [1].

Mathematical model

Homogenous model

The homogenous model is a model with an intermediate level of complexity, where interaction is modeled as a reaction in which the reaction rate accounts for both all kinds of mass transfer resistance and interaction kinetics. The mobile phase was described by equation (1), which consists of a dispersion term that is described by diffusion, a convection term and finally a sorption term.

$$\frac{\partial c_i}{\partial t} = D_{Ax}\frac{\partial^2 c_i}{\partial x^2} - v_{int}\frac{\partial c_i}{\partial x} - \frac{1-\varepsilon_c}{\varepsilon}\frac{\partial q_i}{\partial t} \qquad (1)$$

D_{Ax} is the dispersion coefficient, $\frac{m^2}{s}$, v_{int} is the interstitial velocity, $\frac{m}{s}$, ε_c is the column void fraction, $\frac{m^3 m.p.}{m^3 column}$, ε is the total void, $\frac{m^3 total\ m.p.}{m^3 column}$, c_i is the concentration, $\frac{mol}{m^3}$, q_i is the concentration of the component in the stationary phase, $\frac{mol}{m^3 s.p.}$, i is an index and indicates component i, x is the axial coordinate along the column, m.

The boundary conditions that was used for this model was a Dirichlet at the inlet of column and a homogenous von Neumann at the end of the column, equation (2) and (3). The homogenous von Neumann condition comes from the assumption that there is no diffusion out of the column, only convection and that will give the mass balance according to equation (3).

$$c(t,0) = c_{in} \qquad at\ x = 0 \qquad (2)$$

$$\frac{\partial c_i}{\partial x} = 0 \qquad at\ x = L \qquad (3)$$

In order to estimate the contribution of the dispersion coefficient, the Peclet number was used as defined in equation (4).

$$Pe = \frac{v_{int} d_p}{D_{ax}} \qquad (4)$$

d_p is the particle diameter, m.

The interstitial velocity is the velocity experienced between the particles and is calculated in equation (5).

$$v_{int} = \frac{Q}{A \cdot \varepsilon_c} \qquad (5)$$

Q is the volume flow in the column, $\frac{m^3}{s}$ and A is the cross-sectional area of the column, m^2.

Equation (6) describes the sorption isotherm. The adsorption and desorption of proteins are regarded to be competitive processes. It is the inert salt, with $\frac{dq}{dt} = 0$, which affects the retention time of the protein.

$$\frac{\partial q_i}{\partial t} = k_{ads,i} \cdot c_i \cdot q_{max,i} \left(1 - \sum_{j=1}^{N} \frac{q_j}{q_{max,j}}\right) - k_{des,i} \cdot q_i \qquad (6)$$

$k_{ads,i}$ is the adsorption rate coefficient, $\frac{m^3}{mol\ salt}$, $k_{des,i}$ is the desorption rate coefficient, $\frac{m^3}{mol\ salt}$, $q_{max,i}$ is the maximum concentration of component i in the stationary phase, $\frac{mol}{m^3 s.p.}$.

The mobile phase modulators are defined by equation (7) and (8). As can be seen in equation (7), the desorption rate constant, $k_{des,i}$, is affected by the salt concentration, S, by the power of β_i.

$$k_{des,i} = k_{des0,i} \cdot S^{\beta_i} \qquad (7)$$

β_i describes the ion-exchange interaction while S is the concentration of salt, $\frac{mol}{m^3}$.

$$k_{ads,i} = k_{ads0,i} \cdot e^{\gamma_i \cdot S} \qquad (8)$$

γ_i is the hydrophobic interaction, $\frac{m^3}{mol}$. In this paper it is assumed that the contribution of the hydrophobic interaction was equal to zero.

It is possible to describe the equilibrium by using the contribution of the mobile phase modulator as can be seen in equation (9) [2].

$$K_i = \frac{k_{ads0,i} \cdot e^{\gamma_i \cdot S}}{k_{des0,i} \cdot S^{\beta_i}} = K_{i,MPM} \cdot \frac{e^{\gamma_i \cdot S}}{S^{\beta_i}} \qquad (9)$$

Calibration

The purpose of parameter estimation is to calibrate and validate the implemented model to experimental data. It is supposed to be a tool to predict physical behaviors, optimization and robustness analysis. A prediction of an experimental point, with some kind of precise measure, can only be generated by interpolation in the calibration interval. The accuracy of the model that is required is related to the purpose for which the model is intended and the operating window in which the model is supposed to be used. A large operating window might reduce the accuracy [2].

The potential cost and time savings by using a mathematical model must be weighed against the fact that the model only imitates reality and does not incorporate all features of the real system. The process of interest contains information not available or maybe not even valid in the model [3].

When designing the experiments it is very important that every experiment highlights certain vital process parameters. Some important process parameters are the load volume, gradient length and pH. By varying the load volume it is possible to predict the maximum capacity of the stationary phase. The variation of the gradient length will make it possible to estimate the ion exchange interaction, β_i, or the hydrophobic interaction, γ_i. The desorption rate coefficient, k_{kin}, is estimated by measuring the peak width. By varying pH it will be possible to predict the pH dependency of the proteins and the effect on its net charge.

The parameters that matters the most in calibration of the homogenous IEC model are K_{eq}, β, k_{kin}, q_{max}. By altering K_{eq} the estimator can find the

position of the peaks and it also affects the peak height. β, k_{kin} are more concerned with the zone width and the spacing of the peaks. q_{max} is a parameter that is important to know, since the given value from the manufacturer can differ from what is really experienced. It is not even certain that all of the proteins of interest have been tested on the particular stationary phase by that manufacturer.

Himmelblau, [3] writes that when fitting a mathematical model to experimental data it is important to keep in mind that the data sets must be equal or greater than the number of parameters that is going to be estimated. A very common method when calibrating a model with data sets is the least square method which minimizes the sum of the squares of the errors between the predicted and the experimental values of the dependent variable y for each data point x. The Parameter estimation in this thesis was mainly made by the least squares method.

Objective function

To calibrate a model is not different from optimisation in the sense that there is an objective function which is supposed to be minimized, equation (19). The parameter estimation attempts to determine values for the unknown parameters in order to maximize the probability that the mathematical model will predict the values obtained from the experiments. Equation 28 presupposes that the measurement errors,ϵ_{ijk}, are normally distributed with zero means and standard deviations, σ_{ijk}.

$$\phi = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\min_{\theta}\left\{\sum_{i=1}^{NE}\sum_{j=1}^{NV_i}\sum_{k=1}^{NM_{ij}}\left[\ln\left(\sigma_{ijk}^2\right) + \frac{\left(\tilde{z}_{ijk}-z_{ijk}\right)^2}{\sigma_{ijk}^2}\right]\right\} \quad (10)$$

ϕ is the objective function which is supposed to be minimized in order to maximize the probability that the mathematical model will predict the values obtained from the experiments. N is the total number of measurements taken during all the experiments. NE is the number of experiments performed. NV_i is the number of variables measured in the ith experiment. NM_{ij} is the numbers of measurements of the jth variable in the ith experiment. σ_{ijk}^2 is the variance of the kth measurement of variable j in experiment i. $\tilde{z}_{ijk}$ is the kth measured value of variable j in experiment i, z_{ijk} is the kth (model-) predicted value of variable j in experiment i [4].

Results and discussions

The loading volume, gradient length and pH for the different data sets are presented in table 1. The results from both calibrations can be seen in table 2. For calibration data sets 1, 2 and 3 were used while for validation, data set 4 and 5 were used.

Table 1. The properties for the different experimental data.

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
Gradient length (CV)	20	30	40	40	60
Loading volume (ml)	2	2	5	0.5	2
pH	8.4	8.4	8.4	8.4	8.4

As can be seen in figure 1, the best overall fit to experimental data was obtained when using a least squares method. Even though the constant variance method showed a much better fit for exp. 3, the fit for exp. 1 and 2 were not as good.

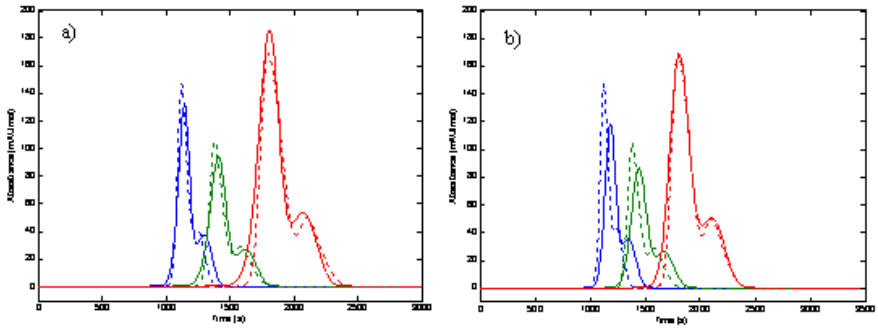

Figure 1. The results from calibration. Left is exp. 1, middle exp. 2 and to the right exp. 3. a) are the results from using a least squares method. b) are the results from using a constant variance method.

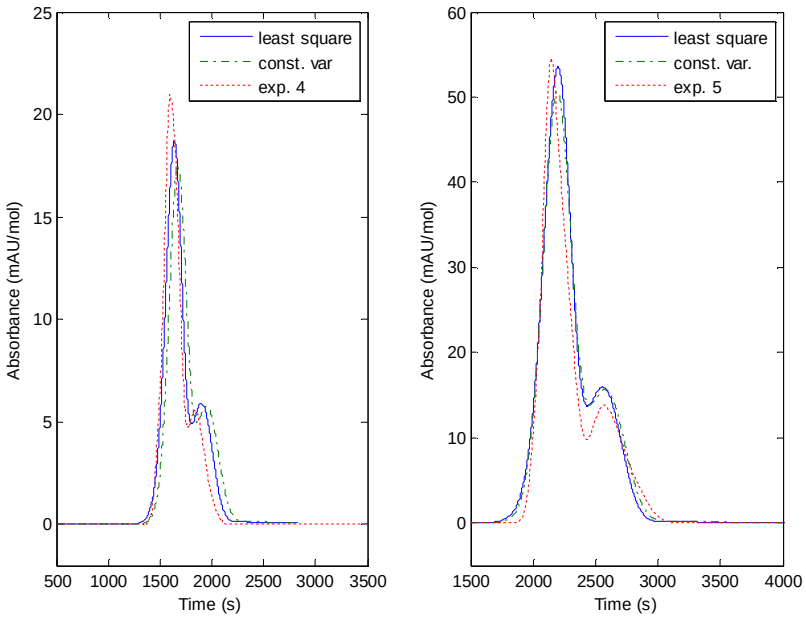

Figure 2. The results from the calibrations of the two models, the left figure is for exp. 4 and the right figure is for exp. 5.

Considering the results in table 2, there are some differences that are peculiar. The maximum capacity, $q_{i,max}$, differ the most between the calibrations and is an indication that the parameters are insensitive to the calibration. It would have been better to calibrate with Henry's constant, which is a merging of the equilibrium constant and the maximum capacity. When comparing the desorption rate coefficients and comparing the UV-curves in figure 1, it seems like the least square method found the better zone width, even though the fit for exp. 3 was worse. The least square method also found a better relation between $k_{i,des}$ and $K_{i,MPM}$ which decides the positioning and the peak height.

Table 2. The results from calibration with 3 data sets.

Parameter	Least square	Constant variance
β_{BSA}	4.85	3.39
$\beta_{BSA2mer}$	4.50	3.38
$k_{BSA,des}$	1.53e-13	$5.41 \cdot 10^{-10}$
$k_{BSA2mer,des}$	2.47e-13	$1.79 \cdot 10^{-10}$
$K_{BSA,MPM00}$	28.17	1.01
$K_{BSA2mer,MPM00}$	1.04	0.25
$q_{BSA,max}$	135.49	1.85
$q_{BSA2mer,max}$	499.93	4.86

The validation of both calibrations can be seen in figure 2. Both models will give more than just an ok prediction of other experiments. Exp. 4 had the same gradient length as exp. 3 and it was surprising that the fit was not better than it was. The spacing was almost correct for both models, but the desorption coefficient was a bit wrong, the zone width was a bit too wide. Exp. 5 was an extrapolation with a gradient length of 60 CVs. The fit for that experiment was excellent, even though the fit for the dimer was not as good as hoped for. The spacing was excellent and the zone width almost perfect. When comparing both models with both validation experiments it is actually the least square method that gives the best overall fit.

Conclusions

The model should be adjusted and replace $K_{i,MPM}$ and $q_{i,max}$ with Henry's constant instead, since the maximum capacity was insensitive to the estimation. gPROMS offers an excellent tool for parameter estimation and for the available data sets, the least square method gave the best fit.

References

[1] Garfin D. E., Ahuja S., Separation Methods in Proteomics", Kirk-Othmer encyclopedia of chemical technology, 4th ed. Found on http://elin.lub.lu.se/ 2008-08-28

[2] Jakobsson N., A model-based approach to robustness analysis and optimization of chromatography processes, Department of Chemical Engineering Lund University, 2006

[3] Edgar T. F., Himmelblau D. M., Optimization of chemical processes, ISBN 0-07-100415-7, Chapter 2.

[4] gPROMS Advanced User Guide, release 2.3, February 2004