Modelling & Optimization of Two Connected Steps in Biopharmaceutical Industry

Bruno Otero

Department of Chemical Engineering, Lund Institute of Technology

Abstract

PEGylation is a very common method to improve the pharmacokinetic properties of protein drugs. The enzymatic reaction of this model protein gives two forms of PEGylated protein, one of which has no commercial interest. A model is constructed and calibrated in MATLAB from experimental data for the reaction and purification steps of the process. These two coupled steps can be simulated and optimized together in different scenarios to obtain a deeper understanding of the system.

Three modes of operation are simulated: Without recycling, with 5 recycles and with infinite recycling. Optimizing these cases to achieve maximum productivity or maximum yield shows the differences between their behaviors. It is observed that to achieve optimal productivity and yield performance, both reactor and column cannot be optimized separately. Also, non-recycle modes of operations are more productive but entail lower global yields in the process with the consequent waste of native protein, as opposed to recycle modes of operation.

Keywords: Process design, chromatography, optimization, coupled steps.

Introduction

PEGylation is the reaction of attachment of polyethylene glycol (PEG) polymer to another molecule. One of the most important targets for the PEGylation is therapeutic proteins due to the modification of their pharmacokinetic properties, which results in an improvement in the delivering in the body. This method reduces renal clearance and supposes a more sustained absorption, which translates in increased clinical effectiveness when the desired effects are concentration-dependent, due to more constant and sustained plasma concentrations. Obtaining a stable linkage, enhanced water solubility or decreased clearance are other goals for these coupled polymers to protein drugs. [1] [2]

Due to the high cost of producing therapeutic un-PEGylated proteins, achieving high purity of PEGylated forms at the expense of yield may not be an economically suitable option, thus making it necessary to improve the reaction specificity and efficiency in purification processes.

The reaction step gives as products the mono and di-PEGylated forms of the protein –named BC and D, respectively whereas the native protein is called A-, of which only the first has economic interest. As can be seen in *Figure 1*, longer reaction times will yield higher di-PEGylated protein concentrations.

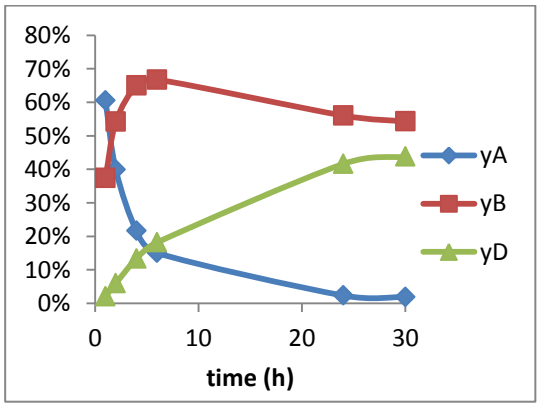

Figure 1. Reaction yields for components A, BC and D with 1.5 substrate molar ratio and 10g/l of initial native protein concentration.

The downstream processing is performed in a cation exchange column based in the different charges of the PEGylated forms.

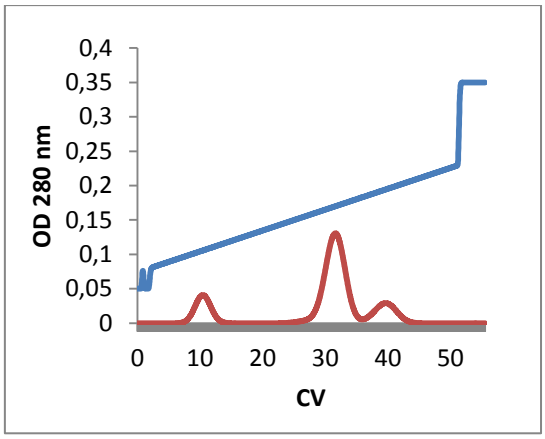

Figure 2. Chromatogram of a A,BC and D mixture for a 0.2 CV load. Di-PEGylated forms of the protein elute at lower salt concentration, followed by the mono-PEGylated and unreacted protein.

The chromatogram in *Figure 2* shows baseline separation between the PEGylated forms and a slower elution time for the unreacted protein that elutes at higher salt concentration.

Theory

The system has been modeled in two different spaces, the first a stirred-tank reactor with first order kinetics with regard to most components and a reaction dispersive model for the dispersion and adsorption in the column. [3] [4] [5]

Reaction model

$$\frac{\partial c_A}{\partial t} = -k_1 \cdot c_A^{n1} \cdot c_S \cdot c_E \qquad (1)$$

$$\frac{\partial c_B}{\partial t} = k_1 \cdot c_A^{n1} \cdot c_S \cdot c_E - k_3 \cdot c_B^{n3} \cdot c_S \cdot c_E \qquad (2)$$

$$\frac{\partial c_S}{\partial t} = -k_1 \cdot c_A^{n1} \cdot c_S \cdot c_E - k_3 \cdot c_B^{n3} \cdot c_S \cdot c_E \qquad (3)$$

$$\frac{\partial c_D}{\partial t} = k_3 \cdot c_B^{n3} \cdot c_S \cdot c_E \qquad (4)$$

$$\frac{\partial c_E}{\partial t} = k_5 \cdot Ca^{n5} \cdot c_E - (c_{E0} - c_E) \qquad (5)$$

Column model

$$\frac{\partial q}{\partial t} = k_{kin} \cdot (H \cdot c \cdot (1 - \sum_i \frac{q_i}{q_{\max}}) - q) \qquad (6)$$

$$\frac{\partial c}{\partial t} = \underbrace{D_{ax} \cdot \frac{\partial^2 c}{\partial z^2}}_{dispersion} - \underbrace{\frac{\nu}{\varepsilon} \cdot \frac{\partial c}{\partial z}}_{convection} - \underbrace{\frac{1-\varepsilon_c}{\varepsilon} \cdot \frac{\partial q}{\partial t}}_{adsorption} \qquad (7)$$

The boundary conditions at the inlet and outlet of the mobile phase are described by a Dirichlet condition and a von Neumann condition.

$$\frac{\partial c_i}{\partial z} = c_i \qquad \text{at z=0}$$

$$\frac{\partial c_i}{\partial z} = 0 \qquad \text{at z=L}$$

The system is discretized in N grid points for which the mathematical model will be solved.

Materials and Methods

Simulation

Firstly, two separate models for the reactor and column are constructed using an ode solver function for the equation systems previously described. Secondly, the two models for reactor and column are coupled so that the volume injected in the column has the final concentrations in the reactor. A recycle scenario is also simulated, in which the fraction rich in unreacted protein is mixed with fresh feed and reintroduced in the next reaction step.

Calibration

The calibration of the reactor is performed with the MATLAB function lsqcurvefit. This function compares two sets of data and tries to minimize the difference between them by modifying the parameters specified.

Table 1. Calibrated parameters for the reactor

Kinetic constant	Reaction order
k_1	n_1
k_3	n_3
k_5	n_5

The calibration of the column consists of manual modification of the parameters and a simulation of the column model. A visual comparison of the chromatogram plot between the experimental data and the simulated data will allow the proper fitting of the model.

Table 2. Calibrated parameters for the column

Henry constant	Kinetic constant
H_{0A}	k_A
H_{0BC}	k_{BC}
H_{0D}	k_D

An increase in the Henry constant will translate in a later elution from the column and vice versa; an increase in the kinetic constant k_{kin} will sharpen the peak.

Optimization

The optimization consists in the modification of a number of decision variables under a set of constraints. The model is simulated with the different variables with the goal of minimizing the value of an optimization objective. This process can be optimized to either maximize the yield of mono-PEGylated protein or to maximize the productivity. The first one optimizes the use of the expensive native protein and the second one optimizes the time efficiency and the fixed costs.

$$y = \frac{amount_{BC}}{A_0} \qquad (9)$$

$$\mathrm{Pr} = \frac{amount_{BC}}{T \cdot V_C} \qquad (10)$$

The decision variables are chosen so they affect the process significantly and the appropriate behavior of the system can be analyzed by the change in these parameters.

Table 3. Design variables for the different models optimized.

Reactor	**Column**
t_R (h)	t_{load} (s)
A_0 (g/l)	c_{Sf} (g/l)
[Ca] (mol/l)	t_{slope} (h)
$ratio_{S/A}$	

The function outline for the optimization as shown in *Figure 3* consists of a function that contains an optimizer, *fminsearchbnd.* This optimizer calls an objective function that calculates the reactor and column models with the design parameters given. The optimizer function will then try to find the set of decision variables that minimize the objective function defined applying a Nelder-Mead based algorithm that approximates a local minimum when the objective function varies smoothly and is unimodal [6].

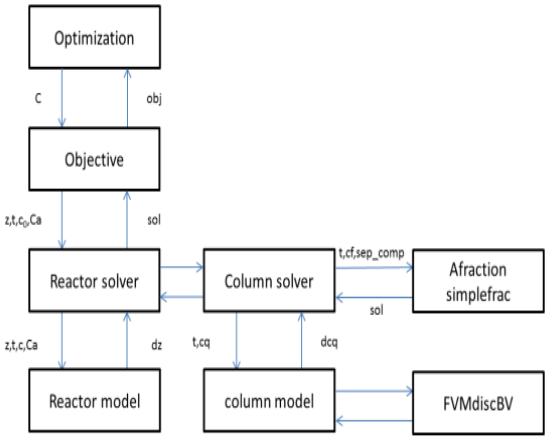

Figure 3. Function outline for the optimization

This optimization will be carried out in different case scenarios. The first will be run the simplest operation mode, the batch mode, followed by a model with 5 recycles and an infinite recycling.

Results

Batch mode

The chromatogram of the optimization allows observing the behavior of both the reactor and column. By looking at the size of the peaks it is possible to estimate the concentrations in the reaction outlet and the position and shape of the peaks shows the separation profile.

The optimization of the yield results unfeasible since the optimum amounts are extremely low and highly unproductive. Optimizing for productivity and productivity·yield gives similar results.

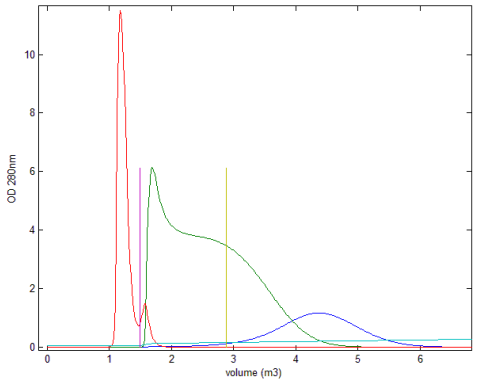

Figure 4. Optimized chromatogram for the productivity objective with fixed ratio.

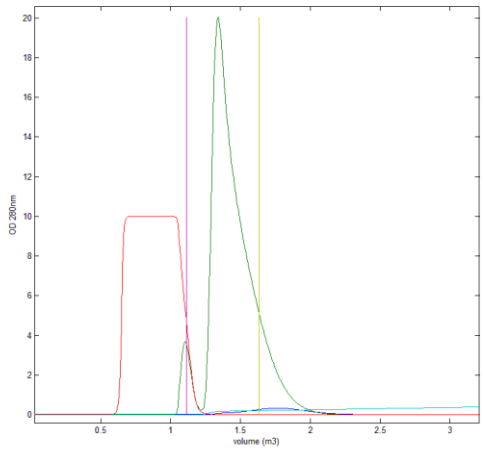

Figure 5. Optimized chromatogram for the productivity objective.

When simulating with the ratio as a new decision variable, the optimizer increases its value in order to improve the column separation. With a higher substrate ratio almost all the native protein disappears at the expense of lower reaction yield, since di-PEGylated protein production also increases due to polymer substrate excess. As it can be observed in *Figure 5*, there is almost no native protein left which allows retrieving higher amount of BC compound in the fraction.

Recycle Batch mode

Optimization of the recycle gave a different chromatogram shape. Separation of BC and A is not so important since it is going to be recycled and the reaction yield is lowered.

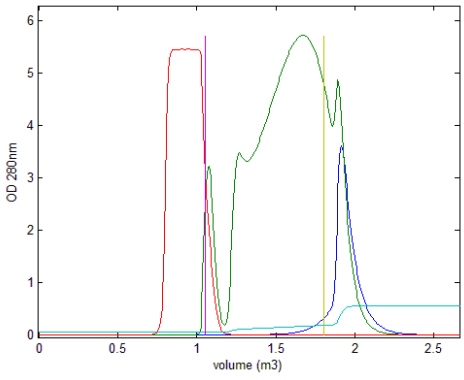

Figure 6. Optimized chromatogram for a particular cycle in the 5 recycle mode.

The recycle model is run with the same decision variables for every cycle and the product between yield and productivity as objective function. This is the only approach studied because it is the only realistic optimization for this method. A yield optimization would give an optimum separation in the column that would be highly unproductive. A productivity optimization would try to decrease the volume recycled to zero since it doesn't take into consideration the economic value of the native protein and the recycle stream is poorer in protein than the fresh feed.

Infinite Recycle mode

It was seen that the finite recycle mode tended to reach a steady state depending on the decision variables. It was hence calculated when this step was reached in every case, step that was later optimized.

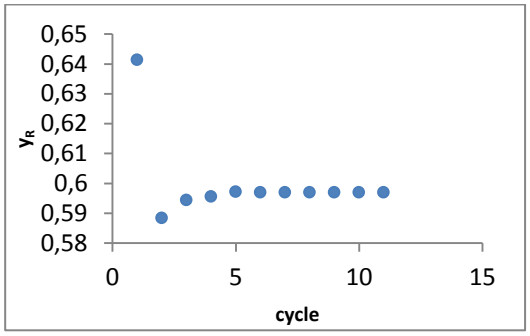

Figure 7. Reactor yield versus cycle number.

The number of cycles until the steady state is reached is generally low, giving very similar results to the 5 recycle mode. In this case the steady state was reached in the sixth cycle, for which both the reactor and the column yield remained constant.

Table 4. Design variables for the different models optimized.

	Batch	**Recycle batch**	**Infinite recycle**
$\mathbf{y_R}$	0.61	0.60*	0.59
$\mathbf{t_R}$	1.43 h	0.97 h	3.67 h
$\mathbf{A_0}$	16.97 g/l	5.61 g/l	6.49 g/l
[Ca]	$7.08 \cdot 10^{-5}$ mol/l	$3.05 \cdot 10^{-6}$ mol/l	$1.59 \cdot 10^{-6}$ mol/l
ratio	1.33	1.46	1.20
$\mathbf{y_C}$	0.75	0.71*	0.78
$\mathbf{t_{slope}}$	1.01 h	4.36 h	3.16 h
$\mathbf{c_{Sf}}$	0.31 mol/l	0.70 mol/l	0.55 mol/l
$\mathbf{t_{load}}$	2,256 s	5,866 s	6,425 s

The previous table shows big differences between the non-recycle mode and both the recycle mode, which have very similar settings, presumably because the stable cycle is reached quickly. The main differences for the non-

recycle mode are the bigger protein load and the faster elution in the column, which are not so important when the recycling is present. It can also be observed that in the recycle cases the loading times are bigger in order to decrease the size of the column to counter the drop in productivity.

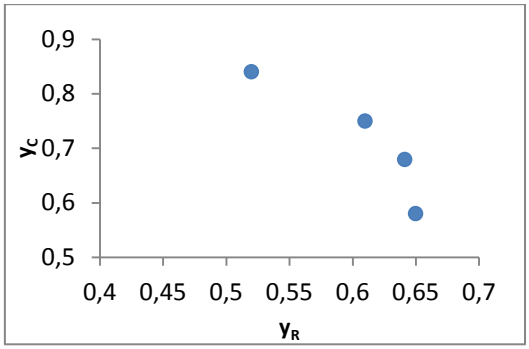

Figure 8. Reaction yield versus column yield for different optimizations.

Comparing the yields of the different steps for the batch optimizations it can be seen a correlation between them. The optimal reactor yield requires a feed and a substrate ratio that difficult the column separation and, hence, lowers the column yield, implying that it is necessary the optimization of both processes together.

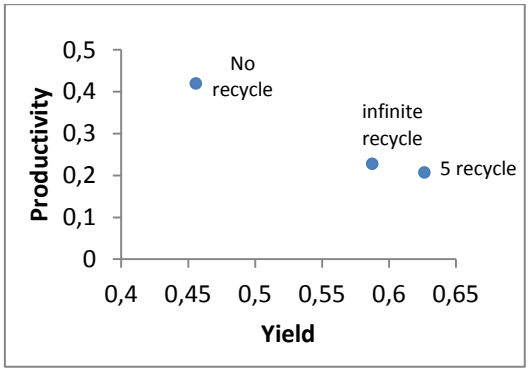

Figure 9. Productivity vs global yield for the productivity·yield optimizations.

The previous plot shows the global productivity and the global yield for the different modes of operation. It can be observed that, as expected, the no recycle mode is more productive at the expense of yield and the recycling implies a higher yield, but is more unproductive since the recycle stream is poorer in protein than the fresh feed. The fact that 5 recycle gives a higher yield than the infinite recycle is probably due to the optimization of the product between productivity and yield that was carried out.

Conclusions

A model that could fit to the given experimental data was successfully coded and studied. This model could simulate efficiently a reactor and column chromatography for the protein giving data that could help understand how the process works.

The calibration gave a good adjustment to the experimental data, which was the purposed established previously. Nonetheless, in the calibration of the reactor the relations between different factors such as A_0 and [Ca] are neglected, and some assumptions are made for the sake of the model's simplicity.

The optimization showed the behavior of the design parameters studied. Reaction time tends to be low in all the simulations, especially compared to the experimental reaction. Protein load and substrate ratio have proved to be key in the process design. These parameters control the amount of native protein in the final mixture that enters the column, which has a high influence in the difficulty to extract the product.

Loading time will determine the size of the column by being inversely proportional, so the size will try to be minimized as long as the productivity decrease allows it. The gradient slope determines the elution speed and since the main difficulty is the separation of the native and mono-PEGylated proteins, closer ratios to the optimum –which suppose a higher native protein amount- will slow down the elution.

The three modes of operation studied are valid ways of running the system and the decision on which to use will most likely rely on economic reasons. The importance that the productivity or the global yield have on the industrial scale will ultimately determine the most convenient approach.

Table of Symbols

$c =$ Concentration in mobile phase $\left(mol/m^3\right)$

k_{kin} = Kinetic constant $\left(s^{-1}\right)$

D_{ax} = Dispersion coefficient $\left(m^2/s\right)$

ε_c = Porosity (void) of particle bed

ε = Total porosity

q = Adsorbed amount $\left(mol/l\right)$

$q_{\max}$ = Maximum adsorbed amount $\left(mol/l\right)$

t = Time $\left(s\right)$

v_{int} = Linear velocity inside the column $\left(m/s\right)$

$H =$ Henry constant $\left(l/mol\right)$

Table of subscripts

$A =$ Component A (Protein)

$BC =$ Components B and C (Mono-PEGylated protein)

$D =$ Component D (Di-PEGylated protein)

$S =$ Substrate (PEG)

$E =$ Enzyme

References

1. J. Milton Harris, Nancy E. Martin and Marlene Modi. Pegylation, a novel process for modifying pharmacokinetics. Clin Pharmacokinet 2001; 40 (7): 539-551.
2. Conan J. Fee, James M. Van Alstine. PEG-proteins: Reaction engineering and separation issues. Chemical Engineering Science, 2006; 61: 924 – 939.
3. Karlsson D, Jakobsson N, Peter Axelsson J, Zacchi G, Nilsson B. Using computer simulation to assist in the robustness analysis of an ion-exchange chromatography step. 2005; 1063: 99-109.
4. Carlsson F. Mathematical modelling and simulation of fixed-bed chromatographic processes. Chemical Engineering 1; 1994.
5. Schmidt-Traub H. Preparative Chromatography of Fine Chemicals and Pharmaceutical Agents Weinheim: Wiley-VCH; 2005
6. The MathWorks, Inc. Global Optimization Toolbox User's Guide. 2004.