Microbial Cell Factories

Poster Presentation

HYBMFA: a bioinformatics' tool for batch-to-batch bioprocess optimisation supported by elementary flux analysis

Ana Teixeira*[1], Carlos Alves[1], Paula Alves[2], Manuel Carrondo[1,2] and Rui Oliveira[1]

Address: [1]REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, P-2829-516 Caparica, Portugal and [2]IBET/ITQB Instituto de Biologia Experimental e Tecnologia/Instituto de Tecnologia Química e Biológica, Apartado 12, P-2781-901 Oeiras, Portugal

* Corresponding author

Background

The central metabolic pathways of many biological systems with industrial interest are currently known. Knowledge of intracellular fluxes is crucial to understand cell metabolism. Bioreactor dynamic optimisation schemes could profit from the incorporation of this knowledge [1,2]. A number of methods have been developed to study the structure of biochemical networks. The elementary flux modes (EFMs) method is particularly attractive since it allows to reduce network complexity to a minimal set of reactions [3].

In previous studies [4], a bioprocess batch-to-batch optimisation scheme supported by a hybrid model was developed and applied to the optimization of a BHK culture expressing the fusion glycoprotein IgG1-IL2. The main contribution of the present study is to improve the previous method by incorporating the knowledge of the metabolic network. The incorporation of the metabolic network in the form of EFMs, may increase the generalization properties of the model and may thus contribute to the increase of the rate of success of the optimization method.

Results

The proposed methodology is based on the premise that the biological system under consideration is only partially known in a mechanistic sense. Following this principle, a hybrid parametric/nonparametric representation of the biological system was adopted to support a batch-to-batch optimization scheme (Figure 1).

In the first step, the metabolic network structure of the biological system under study is analyzed using the elementary flux modes technique. Elementary flux modes are the simplest paths within a network that connect substrates with end-products [3], thus they define the minimum set of n species that must be considered for modelling and how they are connected in a simplified reaction mechanism. The EFM analysis of a given biosystem results in m elementary flux modes and the corresponding $n \times m$ stoichiometric matrix K, with n the number of compounds that must be considered for modelling. The BHK metabolic network analyzed in this work considers the most relevant pathways involving the two main nutrients (glucose and glutamine) within the central metabolism of BHK cells. The FluxAnalyzer software [3] was used to determine the EFMs of BHK metabolic network. There are seven EFMs describing the BHK metabolic network. Assuming the balanced growth condition it is possible to eliminate the intermediate metabolites from each EFM resulting in a set of simplified reactions connecting extracellular substrates (glucose and glutamine) with end-products (lactate, ammonia, alanine, carbon dioxide, purine and pyrimidine). Furthermore, some assumptions concerning the fluxes were made based on literature, resulting in five EFMs. The following stoichiometric matrix was obtained
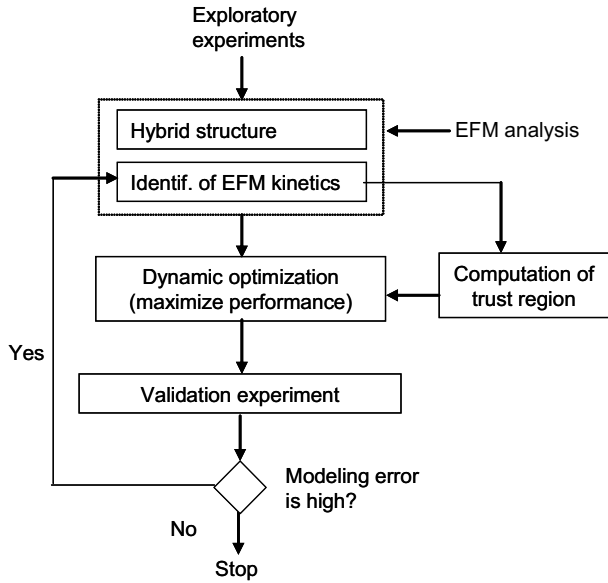
Figure 1
Proposed optimisation scheme.

$$
K = \begin{matrix}
\mu - k_d \ r_1 & r_2 & r_3 & r_4 & r_5 & r_{IgG} \\
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & -1 & 0 & 0 & -2 & 0 \\
0 & 0 & 0 & -1 & -1 & -5 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 2 & 2 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\begin{matrix}
X_V \\ Glc \\ Gln \\ Lac \\ Amm \\ Ala \\ IgG
\end{matrix}
\end{matrix}
\qquad (1)
$$

Note that K also accounts for cell growth and product formation as completely independent fluxes since the stoichiometry of theses reactions is not accurately known.

The state space vector is formed by the *n* concentrations of compounds of the final reactions set and additionally, the concentrations of viable cells, X_v, and product, IgG:

$$\mathbf{c} = [Xv, Glc, Gln, Lac, Amm, Ala, IgG]^T. \qquad (2)$$

Once a reaction mechanism has been established using the EFM method, the next step is the identification of the EFM kinetics from data. Here we adopted a hybrid parametric/nonparametric model structure assuming that that reaction kinetics of EFM are partially known or even completely unknown. This model structure can be formulated mathematically by the following two equations [4]:

$$\frac{d\mathbf{c}}{dt} = \mathbf{r}(\mathbf{c}, \mathbf{w}) - D\mathbf{c} + \mathbf{u} \qquad (3a)$$

$$\mathbf{r}(\mathbf{c}, \mathbf{w}) = \mathbf{K} < \phi_j(\mathbf{c}) \times \rho_j(\mathbf{c}, \mathbf{w})>_{j=1,...,m} \qquad (3b)$$

with **r** a vector of *n* volumetric reaction rates, **K** a $n \times m$ coefficients matrix obtained from the elementary flux modes analysis, $\varphi,(\mathbf{c})$ are *m* kinetic functions established from mechanistic knowledge, $\rho_j(\mathbf{c},\mathbf{w})$ are *m* unknown kinetic functions, **w** a vector of parameters that must be estimated from data, *D* is the dilution rate, **u** is a vector of *n* volumetric input rates (control inputs).

For the system under study the vector of known kinetic functions is given by:

$$\varphi(c) = [X_v \ X_v Glc \ X_v Glc \ X_v Gln \ X_v Gln \ X_v GlnGln \ X_v]^T, \qquad (4)$$

whereas the vector of unknown kinetics is given by:

$$\rho = [\mu - kd \ r_1 \ r_2 \ r_3 \ r_4 \ r_5 \ r_{IgG}]^T = \rho(Glc, Gln, Amm, w). \qquad (5)$$

A backpropagation neural network with a single hidden layer was used for the identification of $\rho_i(c, w)$:

$$\rho(\mathbf{c}, \mathbf{w}) = \rho_{max} s(\mathbf{w}_2 s(\mathbf{w}_1 c + \mathbf{b}_1) + \mathbf{b}_2) \qquad (6)$$

with ρ_{max} a vector of scaling factors with $\dim(\rho_{max}) = m$, $\mathbf{w}_1, \mathbf{b}_1, \mathbf{w}_2, \mathbf{b}_2$ are parameter matrices associated with connections between the nodes of the network, **w** is a vectored form of $\mathbf{w}_1, \mathbf{b}_1, \mathbf{w}_2, \mathbf{b}_2$ and s(.) the sigmoid activation function defined as follows:

$$s(x) = \frac{1}{1 + e^{-x}} \qquad (7)$$

Finally, the last term in eq. (3a), the control input vector is $\mathbf{u} = [0 \ F_{Glc} \ F_{Gln} \ 0 \ 0 \ 0 \ 0]$ with F_{Glc} and F_{Gln} the volumetric feeding rates of glucose and glutamine respectively.

Off-line measurements of the seven state variables from five experiments were used for model training and validation. The neural network had three inputs: glucose and glutamine, the main limiting nutrients, and ammonia, the main toxic by-product. The output vector was formed by the seven unknown specific kinetics: μ-k_d, r_1, r_2, r_3, r_4, r_5, r_{IgG}. The criterion to stop the training was the minimum modelling error of the validation data set. The best result was obtained with five hidden nodes. Figure 2 presents the hybrid modelling results for one of the training and one of the validation data sets. A relevant result is the fact that the hybrid model was able to describe simultaneously all five batches with high accuracy.
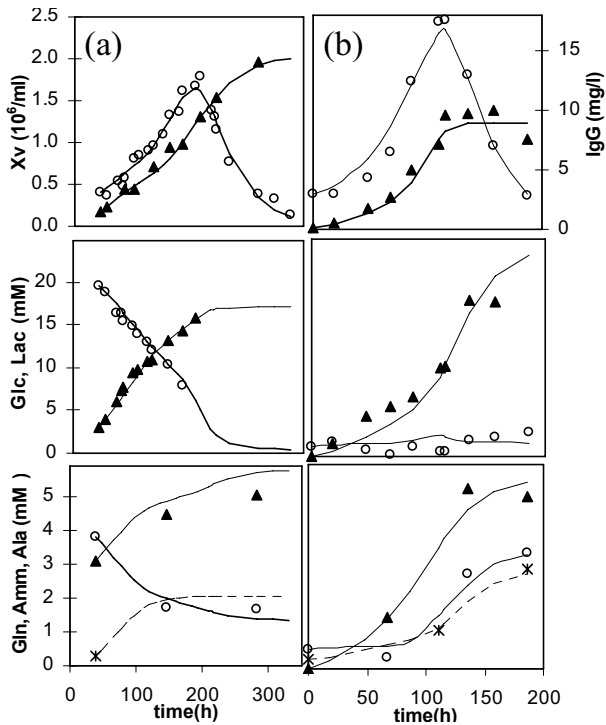
Figure 2
Hybrid model results for a training data set (a) and a validation data set (b).

Optimisation results

With the hybrid model just developed, the process performance (described as the glycoprotein quantity at the end of the bioreaction) is optimized with respect to control inputs F_{Glc} and F_{Gln}, using a micro-genetic algorithm [5]

$$\max_{u} \quad J = C_{IgG1-IL2}\left(t_f\right)V\left(t_f\right) \qquad (8)$$

The optimization (8) is constrained by the hybrid dynamical model and by the risk of ANN inputs being outside the trust region. The optimisation results are presented in Figure 3 showing the optimal trajectories of viable cells, glucose, glutamine and product concentrations. The final product titre is 25 mg/l representing a 67% improvement of performance obtained in the fed-batch experiments so far.

According to the iterative batch-to-batch optimisation scheme shown in Fig. 1, the next step is to perform a new experiment to validate this optimization results. If measured data and predicted optimal process trajectories deviate considerably, additional iterations are performed until convergence of model and process performance is achieved.

Conclusion

A bioinformatic tool was developed that integrates classical optimal control and elementary flux analysis tools. A hybrid parametric/nonparametric modelling framework was adopted that does not require detailed knowledge of intracellular kinetics. A dynamic optimisation method is employed constrained by the risk of nonparametric components unreliability. The method was applied to a recombinant BHK-21 cell line expressing the fusion glycoprotein IgG2-IL1. The final hybrid model was then used to optimise conditions that favour product formation showing that high productivity increments are likely for the process at hand.
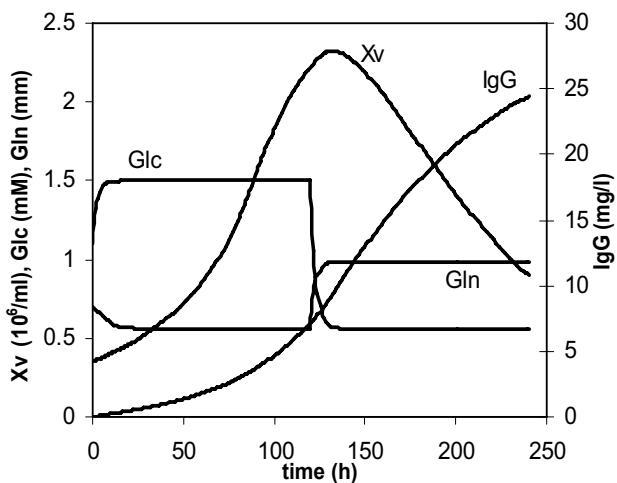
Figure 3

References
1. Provost A, Bastin G: **Dynamic metabolic modeling under balanced growth condition.** *J Process Control* 2004, **14:**717-728.
2. Mahadevan R, Burgard A, Famili I, Van Dien S, Schilling C: **Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals.** *Biotechnol Bioprocess* 2005, **10:**408-417.
3. Klamt S, Stelling J, Ginkel M, Gilles E: **FluxAnalyser: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps.** *Bioinformatics* 2003, **19:**261-269.
4. Teixeira A, Cunha A, Clemente J, Moreira J, Cruz H, Alves P, Carrondo M, Oliveira R: **Modelling and optimisation of a recombinant BHK-21 cultivation process using hybrid grey-box systems.** *J Biotechnol* 2005, **118:**290-303.
5. Krishnakumar K: **Micro-Genetic Algorithms for Stationary and Non-Stationary Function Optimization.** In *SPIE: Intelligent Control and Adaptive Systems Volume 1196. Philadelphia, PA.*; 1989.