

Methodology article

Open Access

Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control

Ana P Teixeira^{1,2}, Carlos Alves², Paula M Alves¹, Manuel JT Carrondo^{1,3} and Rui Oliveira*²

Address: ¹IBET/ITQB, Apartado 12, P-2781-901 Oeiras, Portugal, ²REQUIMTE, Laboratório de Engenharia Bioquímica, FCT/UNL P-2829-516 Caparica, Portugal and ³Laboratório de Engenharia Bioquímica, FCT/UNL P-2829-516 Caparica, Portugal

Email: Ana P Teixeira - ana.teixeira@dq.fct.unl.pt; Carlos Alves - c.alves@dq.fct.unl.pt; Paula M Alves - marques@itqb.unl.pt; Manuel JT Carrondo - mjtc@itqb.unl.pt; Rui Oliveira* - rui.oliveira@dq.fct.unl.pt

* Corresponding author

Published: 29 January 2007

Received: 23 October 2006

BMC Bioinformatics 2007, 8:30 doi:10.1186/1471-2105-8-30

Accepted: 29 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/30>

© 2007 Teixeira et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The progress in the "-omic" sciences has allowed a deeper knowledge on many biological systems with industrial interest. This knowledge is still rarely used for advanced bioprocess monitoring and control at the bioreactor level. In this work, a bioprocess control method is presented, which is designed on the basis of the metabolic network of the organism under consideration. The bioprocess dynamics are formulated using hybrid rigorous/data driven systems and its inherent structure is defined by the metabolism elementary modes.

Results: The metabolic network of the system under study is decomposed into elementary modes (EMs), which are the simplest paths able to operate coherently in steady-state. A reduced reaction mechanism in the form of simplified reactions connecting substrates with end-products is obtained. A dynamical hybrid system integrating material balance equations, EMs reactions stoichiometry and kinetics was formulated. EMs kinetics were defined as the product of two terms: a mechanistic/empirical known term and an unknown term that must be identified from data, in a process optimisation perspective. This approach allows the quantification of fluxes carried by individual elementary modes which is of great help to identify dominant pathways as a function of environmental conditions. The methodology was employed to analyse experimental data of recombinant Baby Hamster Kidney (BHK-21A) cultures producing a recombinant fusion glycoprotein. The identified EMs kinetics demonstrated typical glucose and glutamine metabolic responses during cell growth and IgG1-IL2 synthesis. Finally, an online optimisation study was conducted in which the optimal feeding strategies of glucose and glutamine were calculated after re-estimation of model parameters at each sampling time. An improvement in the final product concentration was obtained as a result of this online optimisation.

Conclusion: The main contribution of this work is a novel bioreactor optimal control method that uses detailed information concerning the metabolism of the underlying biological system. Moreover, the method allows the identification of structural modifications in metabolism over batch time.

Background

Knowledge of intracellular metabolic fluxes is crucial to understand how different pathways interact and their relative importance within the overall metabolic processes. Metabolic flux analysis (MFA) is an established methodology that allows the quantification of such intracellular fluxes. In MFA, intracellular fluxes are calculated by applying steady-state material balances around intracellular metabolites. In general the number of reactions exceeds the number of metabolites resulting in undetermined systems of algebraic equations [1]. Such systems can be solved after measurement of the missing fluxes, which are typically uptake rates of substrates and secretion rates of metabolites, and also intracellular fluxes when the former are not enough.

The determination of metabolic flux distribution in undetermined systems may also be obtained by flux-balance analysis (FBA) [2]. In FBA, unknown fluxes are determined by linear programming (LP), whereby a given objective function, related to a given cellular physiological state, is optimised. Typically, the maximisation of the growth flux defined in terms of biosynthetic requirements is used as the objective function [2-5]. In general, FBA provides flux distribution for a desirable physiological state, it is however uncertain that the provided solution is unique [6]. Frequently multiple optima are obtained which are a consequence of the existence of redundant pathways in the metabolic network conferring structural robustness to cells [7].

Metabolic Pathway Analysis (MPA) is another flux-based analysis method. MPA, unlike FBA, do not look only at the properties of solutions selected by the statement of an objective, but study the full range of achievable biochemical network states that are provided by the solution space. Network-based MPA has focused on two approaches, elementary modes (EMs) and extreme pathways (EPs) [8-10]. These approaches are very similar being EPs a subset of EMs. In certain network topologies the sets of EMs and EPs coincide. They are both unique for a given network and can be considered as nondecomposable steady state flux distributions using a minimal set of reactions. The difference is that EP analysis decouples all internal reversible reactions into two separate irreversible reactions (forward and backward directions) and EMs analysis accounts for reaction directionality. In this work we have adopted the EMs concept since it has broad application; EPs analysis can exclude important routes of the network giving misleading results [10]. MFA focuses on single flux distributions, but in a complex metabolic network of reactions there is a space of admissible flux distributions. The MPA allows the transition from a reaction based perspective to a pathway-oriented view of metabolism because each fea-

sible steady state flux distribution can be represented as a nonnegative combination of EPs or EMs [8,11].

Although flux-based analysis methods have been mainly used for metabolic engineering [1,12], they may also be useful in other phases of the bioprocess development cycle, namely for advanced bioreactor monitoring and control [13,14]. The EMs concept is particularly attractive since it reduces network complexity to a minimal set of reactions. Provost and Bastin [13] exemplified the use of the EMs concept for dynamic modeling of a CHO culture. The main objective of the present study is to derive an optimal control method that incorporates the knowledge of the metabolic network of the biological system under study using the EMs technique. Model-based off- and on-line control techniques are today well established in both theoretical and practical terms, and have been widely applied for bioprocess optimisation (e.g. [15,16]). The success of such methods is critically dependent on the quality of the supporting mathematical model. Not only accuracy in describing previously measured data but mainly the capacity to predict behaviour in unexplored states is the key for success. In previous studies [17,18], an iterative batch-to-batch optimisation scheme was developed and applied to the optimisation of recombinant BHK-21 cultures. The method is based on the premise that in general the biological system under consideration is only partially known or even poorly known in a mechanistic sense. Following this principle, a flexible hybrid parametric/nonparametric representation of the biological system was adopted to support the batch-to-batch optimisation scheme. It was verified that the model generalization capacity increases as more reliable mechanistic knowledge of cells is incorporated in the hybrid model. The algorithm presented here is essentially an extension to the previous batch-to-batch optimisation scheme whereby the knowledge concerning the metabolic network is incorporated in the optimisation algorithm. The methods will be exemplified through the application to a recombinant BHK-21 culture expressing the fusion glycoprotein IgG1-IL2.

Results and discussion

Proposed methodology

The proposed methodology for bioprocess monitoring and control is represented schematically in Figures 1, 2. The backbone of this methodology is the hybrid semiparametric model structure shown in Figure 1. The main design principle is flexible integration of knowledge concerning the metabolism, transport phenomena and empirical process data. The method contemplates the possibility of missing parts of the metabolism (e.g., the product metabolism) and of unknown reaction kinetics and underlying transduction mechanisms. Whenever knowledge is missing, empirical data modeling is called to fill

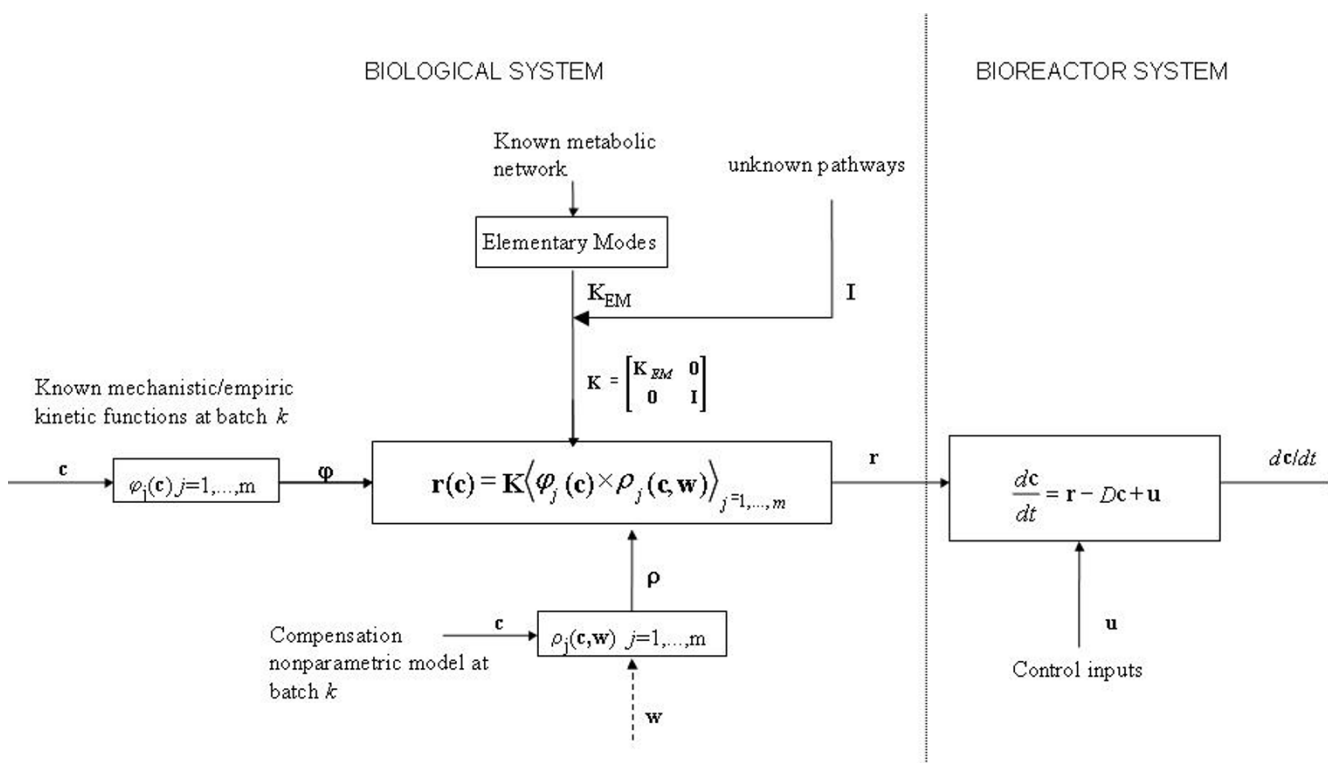


Figure 1
General hybrid structure for bioprocesses. This hybrid model structure integrates knowledge concerning the metabolism, transport phenomena and empirical process data. The bioreactor dynamics are then described by the material balance equations of each component occurring in the EMs. The EMs kinetics are identified with data from exploratory experiments, using chemometric techniques.

the gaps. The starting point is the establishment of the metabolic network of the biological system under study. Firstly, the metabolic network is analyzed using the elementary modes technique. The overall network is decomposed into structural subunits, the EMs, which are the simplest paths connecting substrates with end-products. This structural analysis identifies all compounds (substrates, metabolites and products) taken up and/or secreted to the abiotic phase, which essentially define the system state space vector. The bioreactor dynamics are subsequently described by the material balance equations of each component occurring in the EMs. The EMs kinetics are identified with data from exploratory experiments, using chemometric techniques.

Once the model is properly validated, it can be used for on-line intracellular flux distribution monitoring and for on-line process performance optimisation (Figure 2). So the next step is the on-line implementation of the newly developed hybrid model for process monitoring and/or optimisation in the sense of maximizing the process performance by manipulating the control inputs, i.e., the optimal control problem [16,18-20]. The performance

function includes a penalty term that accounts for the risk of model unreliability, i.e., extrapolation outside the model trust region. The empirical parameters are re-estimated, followed by the re-optimisation of the future process time course whenever new measurements of the process state are performed. In the lines below we describe in detail the steps involved.

Elementary modes analysis

We consider a generic metabolic network with m metabolites and q reactions such as the network represented in Figure 3. Assuming balanced growth and negligible dilution, the fundamental steady state mass balance equations on intracellular metabolites are expressed as follows:

$$\begin{cases} \mathbf{N}\mathbf{v} = \mathbf{0} \\ v_k > 0 \end{cases} \quad (1)$$

with $\mathbf{N} = \{n_{ij}\}$ a $m \times q$ stoichiometric matrix and $\mathbf{v} = \{v_j\}$ the vector of q metabolic fluxes with v_j denoting the net specific rate of reaction j . Some of the q reactions are irreversible, thus the respective fluxes must be nonnegative: v_k

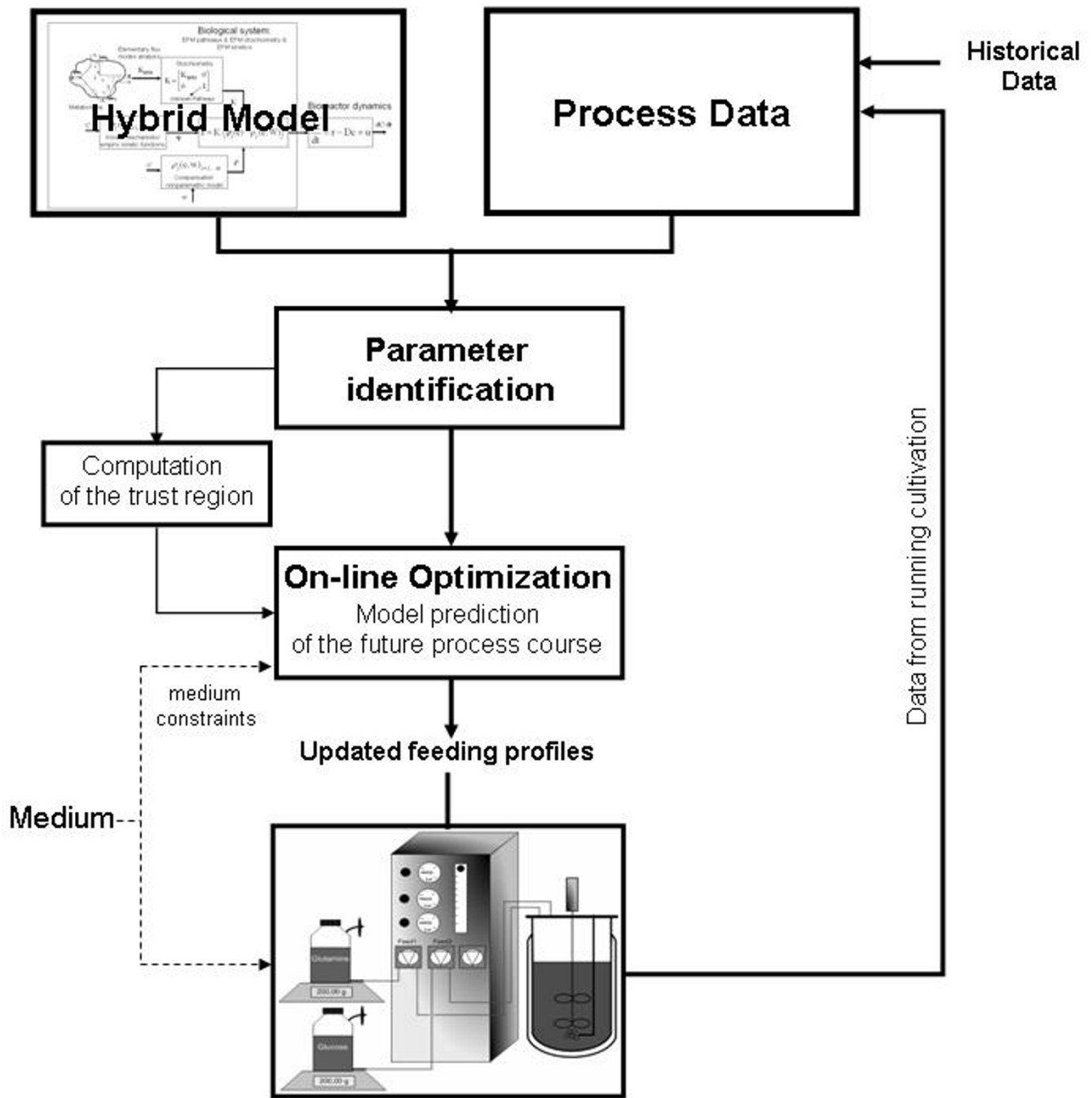


Figure 2
Proposed model-based optimisation scheme. On-line optimisation supported by the hybrid model. The process performance function includes a penalty term that accounts for the risk of model unreliability, i.e., extrapolation outside the model trust region. The parameter estimation as well as the optimisation of the future process course occurs every time a new measurement becomes available.

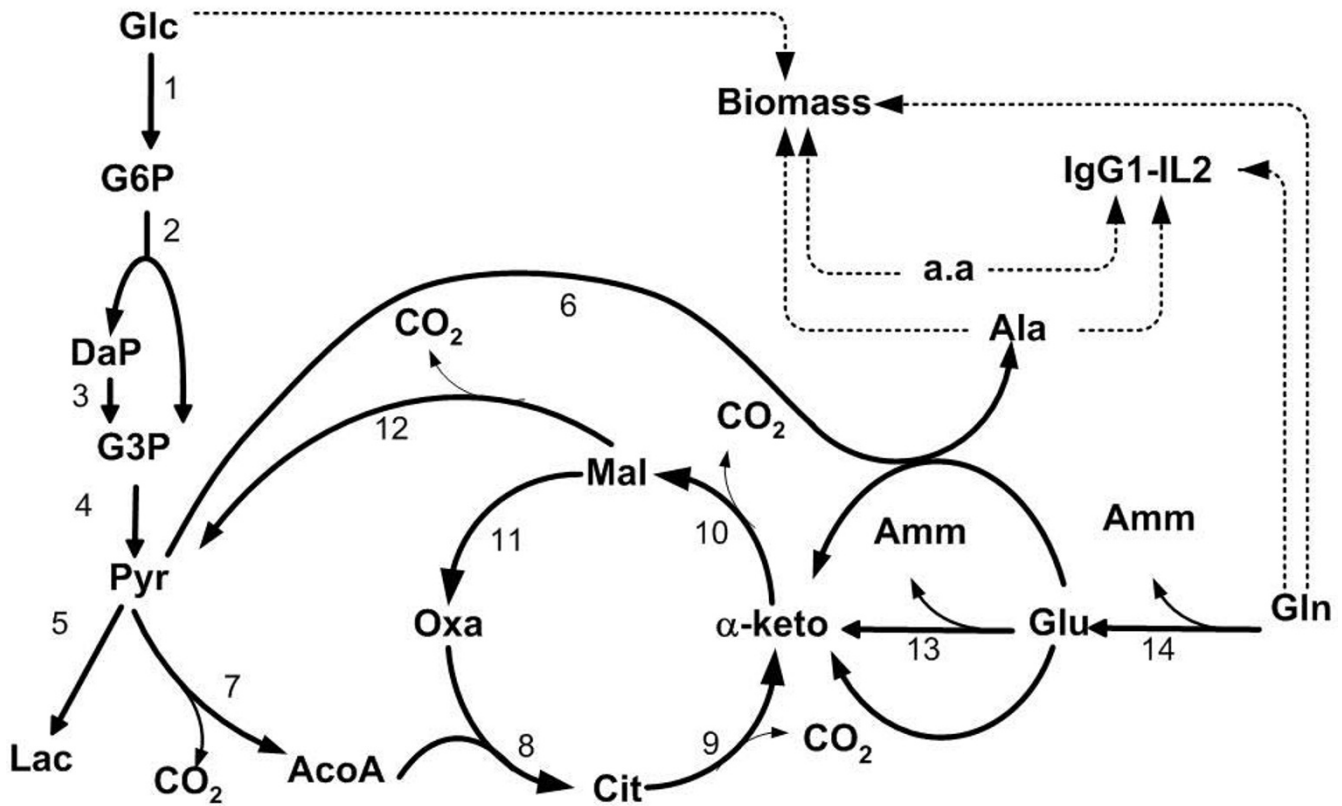


Figure 3
BHK cells metabolic network. The figure shows important pathways in the central metabolism of BHK cells. The dashed arrows indicate lumped pathways towards biomass and desired product synthesis.

> 0 with k denoting the irreversible reactions in the metabolic network. The universe of solutions of system (1) forms a convex polyhedral cone in the solution space [21,22]. It is a property of this system that the infinite set of solutions forming the convex polyhedral cone may be expressed as nonnegative linear combinations of a finite set of n fundamental vectors e_i called elementary modes (EMs):

$$v = \sum_{i=1}^n \lambda_i e_i \quad \forall v \in \mathfrak{R}^q \quad (2)$$

These elementary modes $e = \{e_i\}$ obey to constraints (1) and additionally to the elementary constraint stating that there is no other non-null flux vector involving proper subsets of the reactions participating in that particular EM [22]. In the context of the present work, there are

two most important features of the EMs analysis: it allows to identify all possible pathways for the conversion of substrates into products, and opens the way to the quantification of the relative importance of pathways at a given process stage. The non-null elements in each elementary mode, e_i , define a subset of active reactions of the overall metabolic network N , which are essentially pathways for the conversion of substrates into products.

Hybrid dynamical model formulation

The knowledge acquired from the EMs analysis is integrated in the hybrid model structure represented in Figure 1. This structure allows the introduction of the *a priori* knowledge concerning the metabolic reactions and intracellular kinetics, but it is also open to the possibility of existing missing parts in both cases. The EMs analysis provides a stoichiometric matrix, K_{EM} , that may be aug-

mented (to a \mathbf{K} matrix) if important parts of the metabolism are missing. For instance, the energetic and biosynthetic metabolism may be known, but product metabolism may be unknown. In this case the product concentration or other compounds and respective reactions kinetics are included through matrix \mathbf{I} independently of the EMs.

Even though information about cellular components is growing rapidly, enzymes concentrations and intracellular kinetic data are difficult to obtain. In this work the EMs kinetics are defined by the product of two functions of the system state: a mechanistic/empirical function (in case it is available) and a nonparametric function that represents what is extracted from data without mechanistic interpretation. This model structure can be formulated mathematically by the following two equations, which may be regarded as a general hybrid model for ideal bioreactor systems [18]:

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

$$\mathbf{r}(\mathbf{c}, \mathbf{w}) = \mathbf{K}(\varphi_j(\mathbf{c}) \times \rho_j(\mathbf{c}, \mathbf{w}))_{j=1, \dots, m} \quad (3b)$$

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

Identification of EMs kinetics

The reaction rate of elementary mode j is defined by the product $\varphi_j(\mathbf{c}) \times \rho_j(\mathbf{c}, \mathbf{w})$. The function $\varphi_j(\mathbf{c})$ represents "known" mechanisms whereas $\rho_j(\mathbf{c}, \mathbf{w})$ represents unknown mechanisms. Redundancy and degeneracy are common problems in the determination of fluxes of biological networks [23,24]. It is very important to define *a priori* the conditions under which metabolic fluxes are identifiable. A rank of matrix \mathbf{K} equal to the number of unknown EMs and an equivalent number of measured states are necessary conditions for the identifiability of system (3). If identifiable, the unknown functions $\rho_j(\mathbf{c}, \mathbf{w})$ can be extracted from data using chemometric techniques such as multilinear regression, partial least squares, artificial neural networks and many other. In the frame of hybrid modeling, neural networks have been the most widely used technique for reaction kinetics modeling in biosystems [17,18,25-28]. We used a backpropagation neural network with a single hidden layer to define $\rho_j(\mathbf{c}, \mathbf{w})$:

$$\rho(\mathbf{c}, \mathbf{w}) = \rho_{\max} s(\mathbf{w}_2 s(\mathbf{w}_1 \mathbf{c} + \mathbf{b}_1) + \mathbf{b}_2) \quad (4a)$$

$$s(x) = \frac{1}{1 + e^{-x}} \quad (4b)$$

with $\rho = (\rho_j(\mathbf{c}, \mathbf{w}))$ a vector of m unknown reaction rates, ρ_{\max} a vector of scaling factors with $\dim(\rho_{\max}) = m$, \mathbf{w}_1 , \mathbf{b}_1 , \mathbf{w}_2 and \mathbf{b}_2 are parameter matrices associated with connections between the nodes of the network, \mathbf{w} is a vectored form of \mathbf{w}_1 , \mathbf{b}_1 , \mathbf{w}_2 , \mathbf{b}_2 , and $s(\cdot)$ the sigmoid activation function. A batch neural network training method was adopted, whereby the parameters \mathbf{w} are estimated in the sense of least squares employing a quasi-Newton optimiser with gradients calculated by the sensitivities method [25,28,29] as described in the methods section.

Dynamic optimisation of culture operation: optimal control problem

In the dynamic optimisation step the process performance is optimised with respect to control inputs. This problem may be formulated mathematically as follows:

$$\max_{t_b, \mathbf{c}(0), \mathbf{u}(t)} \left\{ J = f(\mathbf{c}(t_b)) + \int_0^{t_b} g(\mathbf{c}(\tau), \mathbf{u}(\tau)) d\tau \right\} \quad (5)$$

with J the performance index, t_b the batch time, $f(\cdot)$ a terminal performance function and $g(\cdot)$ a time-dependent performance function. The algorithm used was the micro-genetic algorithm [30] coded by Carroll [31]. For simplicity, a piecewise constant approximation of the control inputs \mathbf{u} was adopted. The optimisation (5) is subject to the constraint defined by the hybrid dynamical model (3)–(4) (and indirectly by the metabolic network (1)) but possibly also by other equality and inequality constraints regarding process states, cellular states and control inputs. Due to the use of nonparametric functions, namely of the neural network function (4), it is important to evaluate the unreliability risk during the optimisation. After the EM identification step, the measured input space is clustered by ellipsoidal functions (see the methods section). The clustered input space forms the trust region, wherein the function (4) is considered to be reliable. Optimisation (5) is then further constrained by the risk of function (4) inputs being outside the trust region. The technique is described in detail in the methods section.

Case study: optimisation of recombinant BHK cultures

Process description

To illustrate the methodology described above it will be applied to a recombinant Baby Hamster Kidney (BHK-21A) culture expressing a fusion glycoprotein (an antibody type 1 linked to an interleukin type 2, IgG1-IL2) intended for cancer therapy [32]. The experiments were carried out in serum free and protein free medium (SMIF6, Life Technology, Glasgow, UK). The batch cultures were set up in a 2 l reactor volume and the fed-batch

cultures were set up at 3 different volume scales (2, 8 and 24 l). Sparger aeration was employed. Dissolved oxygen concentration was set at 15% of air saturation. Agitation rate used was 60 rpm; pH was set as 7.2 and controlled through the addition of CO₂. Experimental data of viable cells concentration and six extracellular species (glucose, glutamine, lactate, ammonia, alanine and desired product) were collected. Analytical techniques are described elsewhere [17].

Elementary modes

BHK-21A cells use glucose and glutamine as major sources of carbon and energy, and produce lactate and ammonia as toxic by-products. A reduction of this waste production will improve both cellular growth and glycoprotein (IgG1-IL2) synthesis. Figure 3 shows the metabolic network adopted in this work [33,34]. As catabolic routes, the network includes the glucose and glutamine fluxes through glycolysis, glutaminolysis and TCA cycle. The amino acids metabolism was not considered; it was assumed that all of them are provided by the culture medium. The elementary modes of the 14 reactions that compose the catabolism were calculated using the *FluxAnalyzer* program [22,35]. This system has five EMs, each one consisting of collections of reactions steps (Figure 4). The hypothesis of balanced growth allows the elimination of the intermediate metabolites resulting in a simplified set of reactions (see Table 1) connecting extracellular sub-

strates (glucose and glutamine) with end-products (lactate, ammonia, alanine and carbon dioxide). The first elementary mode corresponds to the glucose flux converted into lactate; the second is the complete oxidation of glucose via TCA cycle (the most energetic pathway involving glucose); the third mode is the complete oxidation of glutamine (the most energetic pathway involving glutamine) and the fourth and fifth modes are partial oxidations of glutamine in alanine and lactate, respectively.

Biomass and product synthesis

In addition to the catabolism, the anabolism and product synthesis must also be considered. For the sake of simplicity, the anabolism (biomass synthesis) was represented as a lumped equation combining the precursors of the main cellular building blocks (glucose, required for the synthesis of carbohydrates, lipids and nucleotides, and amino acids required for the synthesis of cellular proteins and some of them also for the synthesis of nucleotides). The stoichiometry established by Wei-Shou Hu and coworkers [36] for an hybridoma cell line was adopted in this work.

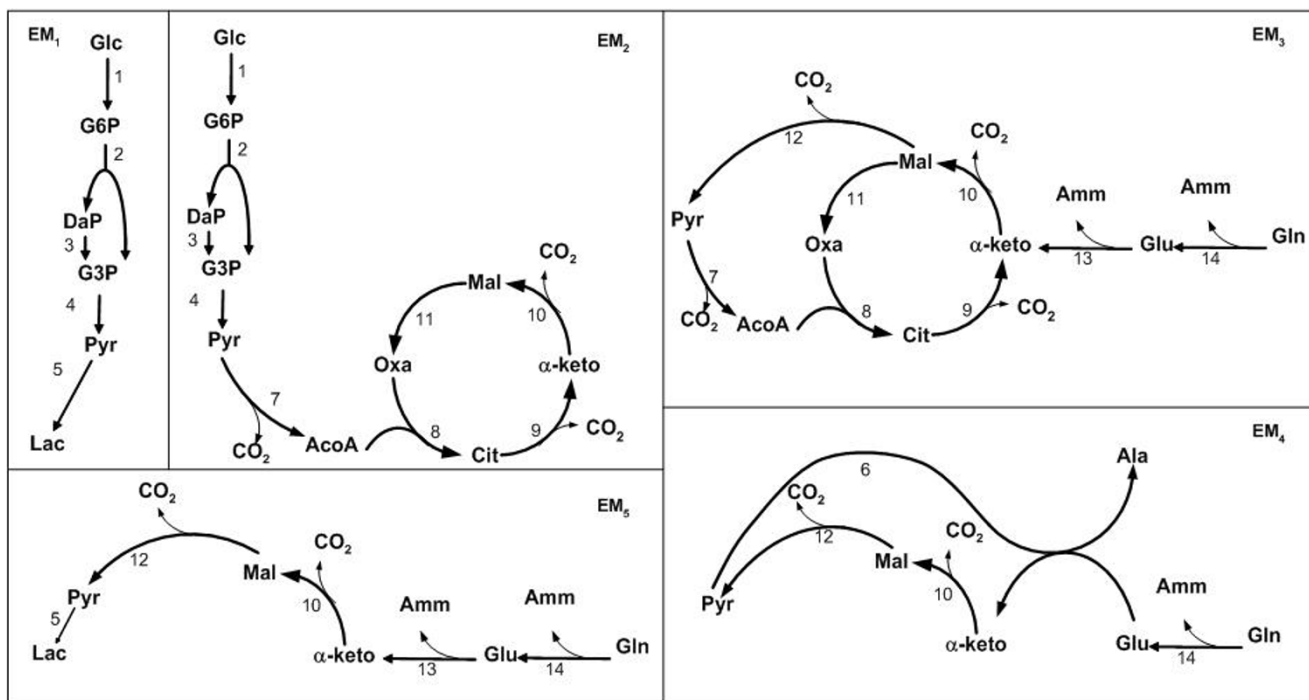
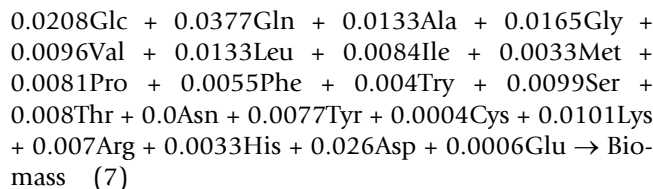


Figure 4
Elementary modes of the metabolic network considered.

Table 1: Elementary modes of the metabolic network considered.

EM ₁ :	Glucose → 2 Lactate
EM ₂ :	Glucose → 6 CO ₂
EM ₃ :	Glutamine → 5 CO ₂ + 2 Ammonia
EM ₄ :	Glutamine → 2 CO ₂ + Ammonia + Alanine
EM ₅ :	Glutamine → Lactate + 2 CO ₂ + 2 Ammonia

The IgG1-IL2 synthesis was also represented as a lumped equation as follows:

$$0.0104\text{Gln} + 0.0112\text{Ala} + 0.0139\text{Gly} + 0.0163\text{Val} + 0.0182\text{Leu} + 0.0061\text{Ile} + 0.0029\text{Met} + 0.0147\text{Pro} + 0.0072\text{Phe} + 0.0037\text{Try} + 0.0243\text{Ser} + 0.0163\text{Thr} + 0.0088\text{Asn} + 0.0077\text{Tyr} + 0.053\text{Cys} + 0.0136\text{Lys} + 0.0061\text{Arg} + 0.0043\text{His} + 0.0083\text{Asp} + 0.0096\text{Glu} \rightarrow \text{IgG1-IL2} \quad (8)$$

This equation is based on the amino acid composition of the antibody IgG1 plus the interleukin IL2 (both amino acid sequences are available at [37]). The carbohydrate content of this fusion glycoprotein was neglected since the glucose contribution is extremely small compared to overall glucose consumption.

Hybrid model structure

The EMs analysis provides a simplified reaction mechanism based on which the following hybrid model structure (equivalent to system 3a-b) is formulated:

$$\frac{d}{dt} \begin{bmatrix} X_v \\ \text{Glc} \\ \text{Gln} \\ \text{Lac} \\ \text{Amm} \\ \text{Ala} \\ \text{IgG} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & -0.0208 & 0 \\ 0 & 0 & -1 & -1 & -0.0377 & -0.0104 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -0.0133 & -0.0112 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_1(\text{Glc, Gln, Amm}) \cdot (X_v \text{Glc}) \\ \rho_2(\text{Glc, Gln, Amm}) \cdot (X_v \text{Glc}) \\ \rho_3(\text{Glc, Gln, Amm}) \cdot (X_v \text{Gln}) \\ \rho_4(\text{Glc, Gln, Amm}) \cdot (X_v \text{Gln}) \\ \mu(\text{Glc, Gln, Amm}) \cdot (X_v) \\ \pi(\text{Glc, Gln, Amm}) \cdot (X_v \text{Gln}) \end{bmatrix} - D \begin{bmatrix} X_v \\ \text{Glc} \\ \text{Gln} \\ \text{Lac} \\ \text{Amm} \\ \text{Ala} \\ \text{IgG} \end{bmatrix} + \begin{bmatrix} 0 \\ F_{\text{Glc}} \\ F_{\text{Gln}} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (9)$$

The state space vector is formed by the concentrations of compounds that intervene in the final reactions set (glucose, Glc, glutamine, Gln, lactate, Lac, ammonia, Amm, alanine, Ala) and additionally the concentrations of viable cells, X_v, and product, IgG:

$$c = [X_v, \text{Glc}, \text{Gln}, \text{Lac}, \text{Amm}, \text{Ala}, \text{IgG}]^T. \quad (10)$$

Carbon dioxide was excluded because its concentration was not measured and because it doesn't interfere with the dynamics of the remaining variables (given that pH is controlled). The stoichiometric matrix, K, is established from the elementary modes of Table 1, but it also accounts for cell growth (5th column) and product formation (6th column) as lumped equations of glucose, glutamine and alanine. It should be noted that the 5th EM was not included in K because preliminary results showed that this EM has negligible flux. This observation is in agreement with some published works [38-40], stating

that lactate is mainly produced from glucose, being the percentage coming from glutamine very low (less than 10%). The volumetric reaction rates of the EMs were defined on the basis of the following assumptions:

- (i) all reaction rates are specific (proportional to the concentration of viable cells),
- (ii) the metabolic reactions considered are all irreversible (in this particular problem) and therefore the respective reaction rates are nonnegative
- (iii) uncertainty in relation to kinetic constants and possible unknown saturation and inhibition effects.

(iv) only the concentrations of glucose, glutamine and ammonia have a significant impact on the specific reaction kinetics [41]. Lactate never reaches inhibitory levels in our experiments.

In the reaction rates of eq. (9) the term in parenthesis represents the *a priori* knowledge concerning the kinetics of the particular reaction (points (i) and (ii)) whereas the ρ_i, μ and π terms represent the uncertainty concerning the reaction kinetics (point (iii)) and are functions of three state variables (point (iv)). With this particular formulation, the vector of known kinetic functions is given by:

$$\varphi(c) = [X_v \text{Glc} \ X_v \text{Glc} \ X_v \text{Gln} \ X_v \text{Gln} \ X_v X_v \text{Gln}]^T, \quad (11)$$

whereas the vector of unknown kinetics is given by:

$$\rho = [\rho_1 \ \rho_2 \ \rho_3 \ \rho_4 \ \mu \ \pi]^T = \rho(\text{Glc, Gln, Amm}, w) \quad (12)$$

The last term in eq. (9) is the control input vector u that in our case accounts for the volumetric feeding of glucose, F_{Glc}, and glutamine, F_{Gln}.

Identification of the EM kinetics

An important point for the identification of unknown flux functions (12) is that the rank of K is 6, thus the measurement of (10) (dim(c) = 7 > rank(K)) is sufficient for the observability of the EM kinetics. The other relevant point is the availability of sufficiently "rich" measurements to identify the "true" fluxes. Preliminary simulation tests showed that, for the system structure of eq (9), the "true" fluxes can be identified under typical fed-batch conditions (results not shown).

Experimental data of seven experiments (three batch and four fed-batch cultures) were used for the identification of the EM kinetics. Data of 5 experiments were used for parameter calibration whereas data of 2 experiments were used for model validation. The concentrations in the state space vector (eq. 10) were analyzed off-line according to

methods published elsewhere [17]. The neural network has three inputs, $[Glc, Gln, Amm]^T$ and 6 outputs as defined by eq. (12). The number of hidden nodes was tuned heuristically in the sense of minimizing the error of the validation data set. The best result was obtained with 5 hidden nodes giving a total number of network parameters equal to 56. The output scaling factors reflect the maximum kinetic rates and were defined as $\rho_{max} = [0.11 \ 0.30 \ 0.05 \ 0.05 \ 0.09 \ 0.11]^T$.

The hybrid modeling results in terms of measured and predicted state variables are presented in Figure 5 for both training and validation data sets. Examples of EM kinetics identification are provided in Figure 6. The hybrid model was able to describe simultaneously all seven experiments with acceptable accuracy. In particular, the results with the validation data set strengthen the predictive potential of the model.

Metabolic interpretation

Figure 6 illustrates the EM kinetics identified by the hybrid model for two experiments, one batch and one fed-batch culture. The analysis of the flux distribution over the time course of bioreaction provides valuable information concerning the evolution of BHK metabolism and how to control the flux distribution through the feeding of glucose and glutamine.

In general, the metabolic activity of BHK cells during the cell growth phase is higher in the batch culture than in the fed-batch culture. The EM fluxes in the fed-batch experiments seem to be much more controlled than in the batch experiments. The high levels of glucose and glutamine at the beginning of the batch culture are mostly directed toward the overflow metabolism, i.e., waste production of lactate and alanine (EM_1 and EM_4). The fed-batch culture, which had lower glucose and glutamine concentrations during the cell growth phase, started with substantially lower consumption rates of these nutrients, while maintaining the flux to biomass.

Glucose is consumed for biomass synthesis and is metabolized through elementary modes EM_1 and EM_2 . The consumption of this nutrient differs significantly from batch to fed-batch cultures. Cells use glucose in a more efficient way in fed-batch than in batch cultures particularly at the beginning of the culture, since the glucose metabolized to lactate (EM_1) is much higher in the batch experiment. These results are in agreement with several published works for other mammalian cells [42-44], where it is reported that high glycolytic activity of animal cells results from high residual glucose.

Glutamine is consumed for product and biomass synthesis and is metabolized through elementary modes EM_3

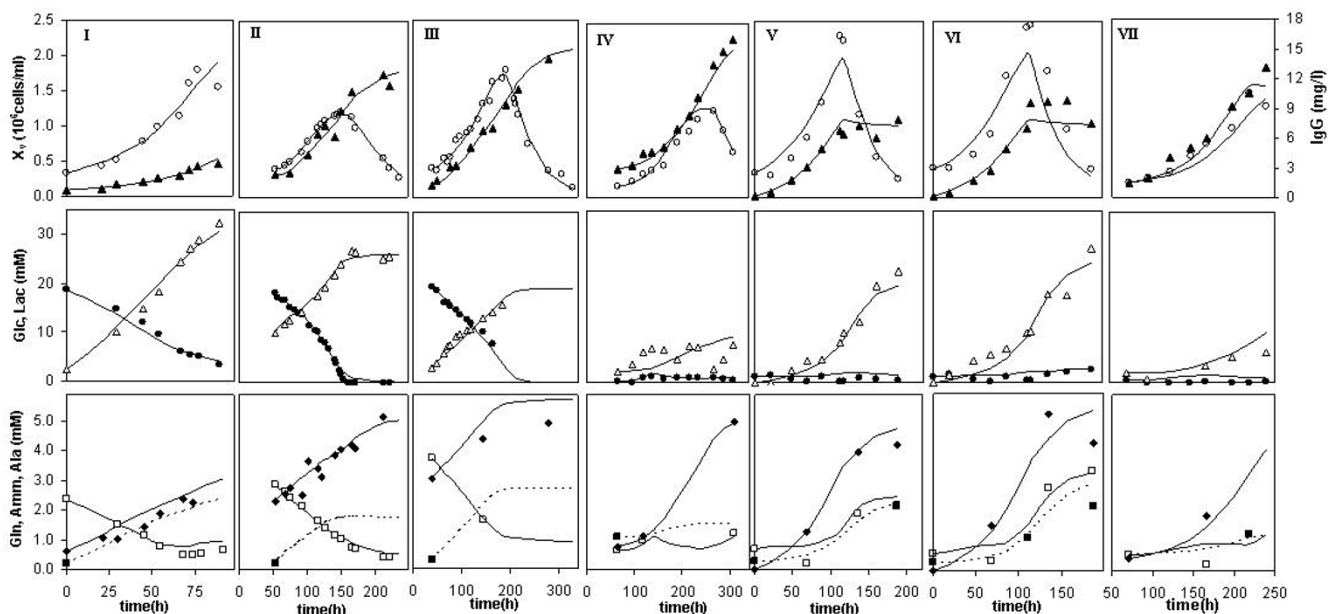


Figure 5
Hybrid modeling results. Modeling results of all seven state variables for both training (I-V) and validation data sets (VI-VII). Experiments I-III are Batch cultures and IV-VII are Fed-batch cultures.

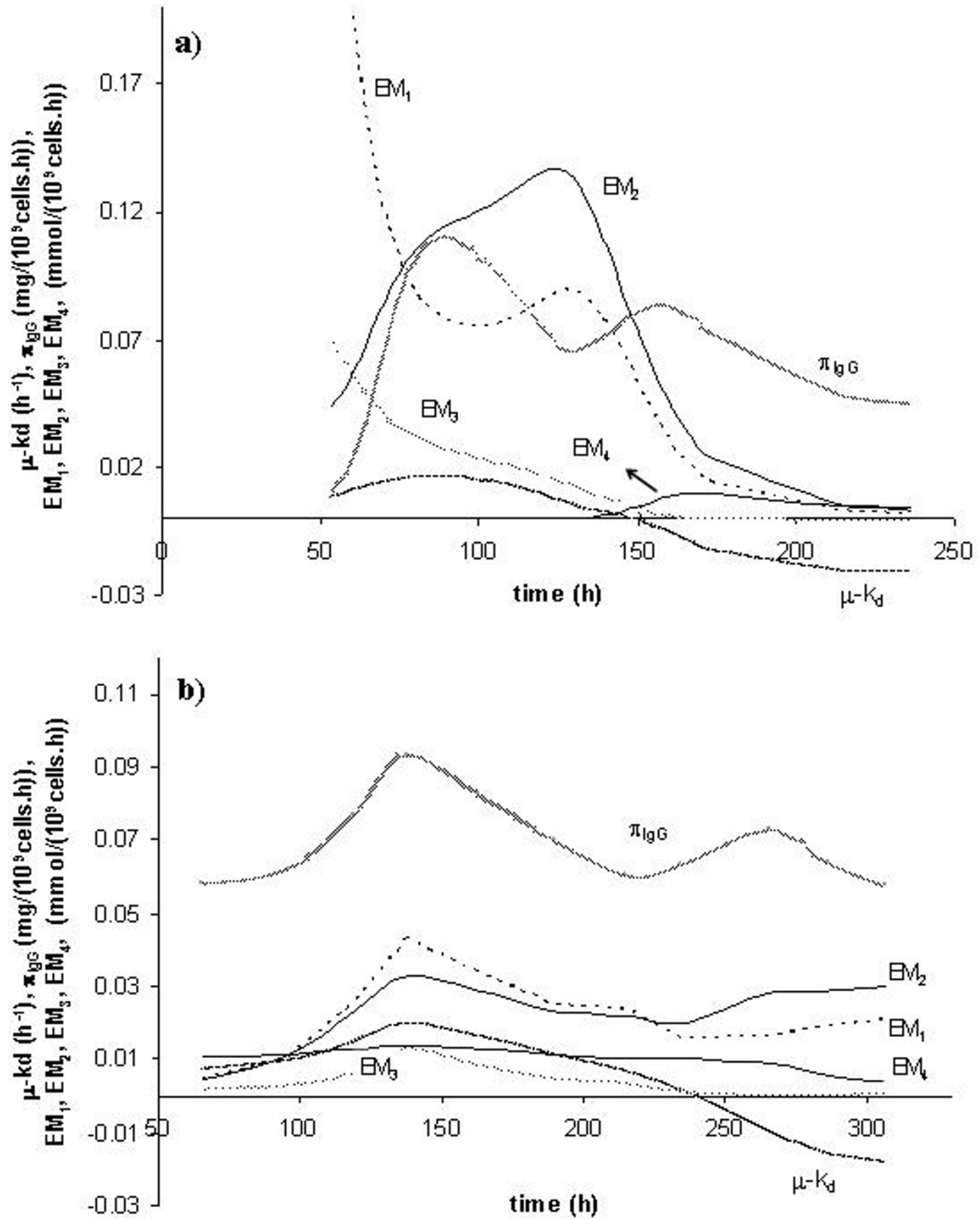


Figure 6
Kinetic rates identified by the hybrid model. The kinetic rates over the time course of bioreaction can provide valuable information concerning the evolution of BHK metabolism, (a) Batch culture; (b) Fed-batch culture.

and EM₄. The most energetic pathway involving glutamine, EM₃, is practically inactive during the cell growth phase of the batch culture, being glutamine preferentially converted into alanine. On the opposite, this elementary flux mode is an important pathway in the course of the fed-batch culture, representing 50% of the total glutamine consumption. The higher production rate of alanine at high glutamine levels (such as the ones present at the beginning of batch cultures) is consistent with observations made by Doverskog et al. [44] and Vriezen et al. [45].

Concerning the specific formation rate of glycoprotein, π_{IgG} , the product formation is consistently more stable in the fed-batch culture than in the batch culture. The product synthesis rate oscillates between 0,06 and 0,07 mg 10⁻⁹ cells h⁻¹ in the former case whereas in the latter case the reaction is much lower in the beginning. This appears to be correlated with the overflow metabolism in the batch experiments, which seems to be detrimental for product synthesis.

On-line culture optimisation

The hybrid model was further used for on-line optimisation of a fed-batch BHK culture. The model parameters were re-tuned on-line using the data from the running cultivation. A batch training scheme was adopted, whereby the data of the running cultivation along with the data of historical experiments were used for model adjustment (Figure 2). Some variables, namely glucose, glutamine, lactate and viable cells concentration, can be measured off-line, with the results available after a short period of time (about 10 minutes). Therefore, at sampling times of 6 to 12 h, these measurements were stored in the training data set and then used for parameter identification using the same strategy previously described in the *Identification of the EM kinetics* section. The only difference was that the initial parameter values were those obtained in the off-line training procedure. The ammonia, alanine and IgG concentrations were excluded from this model adjustment step because they were quantified only at the end of the experiment.

After the parameters retuning step, the glucose and glutamine feeding rates were re-optimised ($\mathbf{u} = [F_{\text{GLC}}, F_{\text{GLN}}]^T$) in the sense of maximising the total amount of IgG1-IL2 produced at the end of the experiment (13).

$$\begin{aligned} \max_{\mathbf{u}} \quad & J = C_{\text{IgG-IL2}}(t_f)V(t_f) \\ & \text{equation (9)} \\ & 0 \leq F_{\text{glu}} \leq F_{\text{glu,max}} \\ & 0 \leq F_{\text{gln}} \leq F_{\text{gln,max}} \\ & (t_f - t_0)^{-1} \int_{t_0}^{t_f} \text{RISK}[c(\tau)] d\tau \leq \text{RISK}_{\text{max}} \end{aligned} \quad (13)$$

The optimisation (13) is subjected to the constraints of the process dynamics (equation 9), upper and lower bounds for the glucose and glutamine feeding rates, and the maximum risk of model unreliability, RISK_{max} (see methods section). The increase in volume is negligible in this problem, thus it was not considered in optimisation (13). The risk constraint states that the average risk must not exceed a given maximum level defined by the user. This restricts the feasibility domain to low risk regions and is essential for process optimisation supported by hybrid models since the black-box model (4a,b) predictions degrade in regions of the input space with sparse measurements.

The micro-genetic algorithm coded by [31] was used to solve optimisation (13). The population size and number of generations was 5 and 2000, respectively. A maximum risk level of 35% ($\text{RISK}_{\text{max}} = 0.35$) was adopted during the on-line optimisation experiment. At each sampling time, the flow rates of the feeding pumps were updated according to the re-optimised feeding profiles of glucose and glutamine.

Figure 7 shows the optimised trajectories and corresponding measurements for the main state variables (viable cells, glucose, glutamine and product) at cultivation times of 0 h (i.e., *a priori* optimised trajectories), 46 h, 75 h, 92 h and 195 h. The comparison of predicted and measured concentrations shows a very satisfactory performance for the on-line optimisation. Furthermore, although the measured product concentration was only available at the end of the experiment, the predicted time course of this variable follows closely the product measurements. The final product titre obtained was 16.4 mg/l corresponding to a 10% increase in relation to previous experiments carried out with the same medium (initial glucose and glutamine concentrations) and without on-line optimisation.

Medium optimisation

Higher productivities are likely if the initial medium composition is optimised. The optimisation results of initial glucose and glutamine concentrations along with the corresponding feeding strategies are shown in Figure 8a. The medium should have initially low levels of glucose and

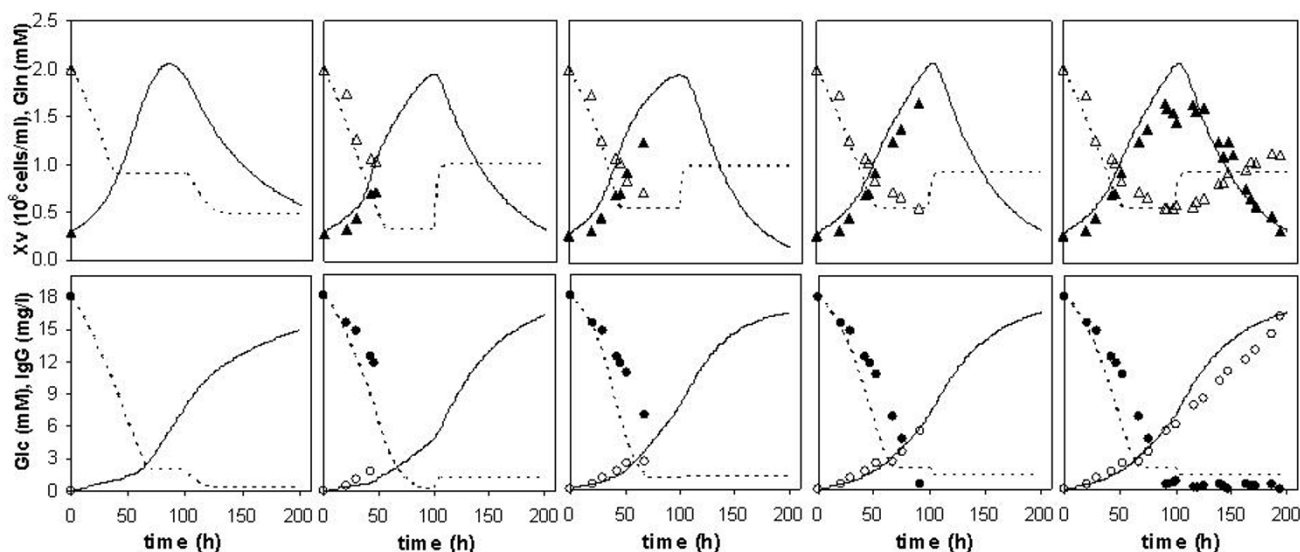


Figure 7

Optimal trajectories during a fed-batch on-line optimization. Optimised trajectories of process variables for five periods of cultivation: 0, 46, 75, 92 and 195 h (lines are model predictions and symbols are experimental data).

glutamine. As for the feeding strategies, low levels of glucose and glutamine are optimal during the cell growth phase (glutamine concentration at 0.6 mM and glucose at 1 mM), whereas during the cell death phase, glutamine should slightly increase and glucose should decrease to a concentration of 0.4 mM. A significant increase of product titre was predicted under such cultivation conditions.

The kinetics of the elementary modes for the optimal strategy is presented in Figure 8b. The optimal strategy can be interpreted by means of intracellular flux distribution. The results were not much different from the elementary flux distribution of the fed-batch culture (Figure 6b). The main difference arises from the fact that the specific product formation (π_{IgC}) increases during the cell death phase as a consequence of the increase of the glutamine concentration.

Conclusion

The current work presents a hybrid modeling approach for bioprocesses that integrates information concerning intracellular metabolic fluxes, bioreactor transport phenomena and measured data. First, elementary flux analysis is used to reduce the biosystem metabolic network into a set of macroscopic reactions relating extracellular components only. A bioreactor dynamical model is then established from material balance equations of the compounds which intervene in the final reaction set. Nonparametric techniques are used for identification of the elementary modes kinetics from measured data. The method was successfully applied to a recombinant BHK-

21A cell line producing a fusion glycoprotein. A significant result is the achievement of the flux distribution over the runtime of a bioprocess. The so obtained information can be used to identify conditions that favour product formation. A fed-batch BHK culture performed with on-line optimisation supported by the proposed methodology allowed a 10% increase in the final productivity. Higher productivities are expected if starting nutrients concentrations are optimised. The developed tool promises to be advantageous for optimising the productivity of fed-batch biochemical processes since the transfer and adaptation to different cell lines is reasonably straightforward

Methods

Neural network training procedure

A batch least squares criteria of residuals in concentrations was adopted

$$\min_{\mathbf{w}} \left\{ e = \frac{1}{P \times n} \sum_{t=1}^P \|\bar{\mathbf{c}}(t) - \mathbf{c}(t, \mathbf{w})\|^2 \right\}, \quad (M1)$$

with P the number of measured patterns, n the number of state variables, $\bar{\mathbf{c}}(t)$ and $\mathbf{c}(t, \mathbf{w})$ the measured and model state vectors. The concentrations were previously scaled by the measurements standard deviations. A quasi-Newton algorithm with conjugate gradient with line search (MATLAB™ optimisation toolbox) was adopted.

The gradients $\partial e / \partial \mathbf{w}$ can be evaluated using the sensitivities equations, which can be obtained by differentiating

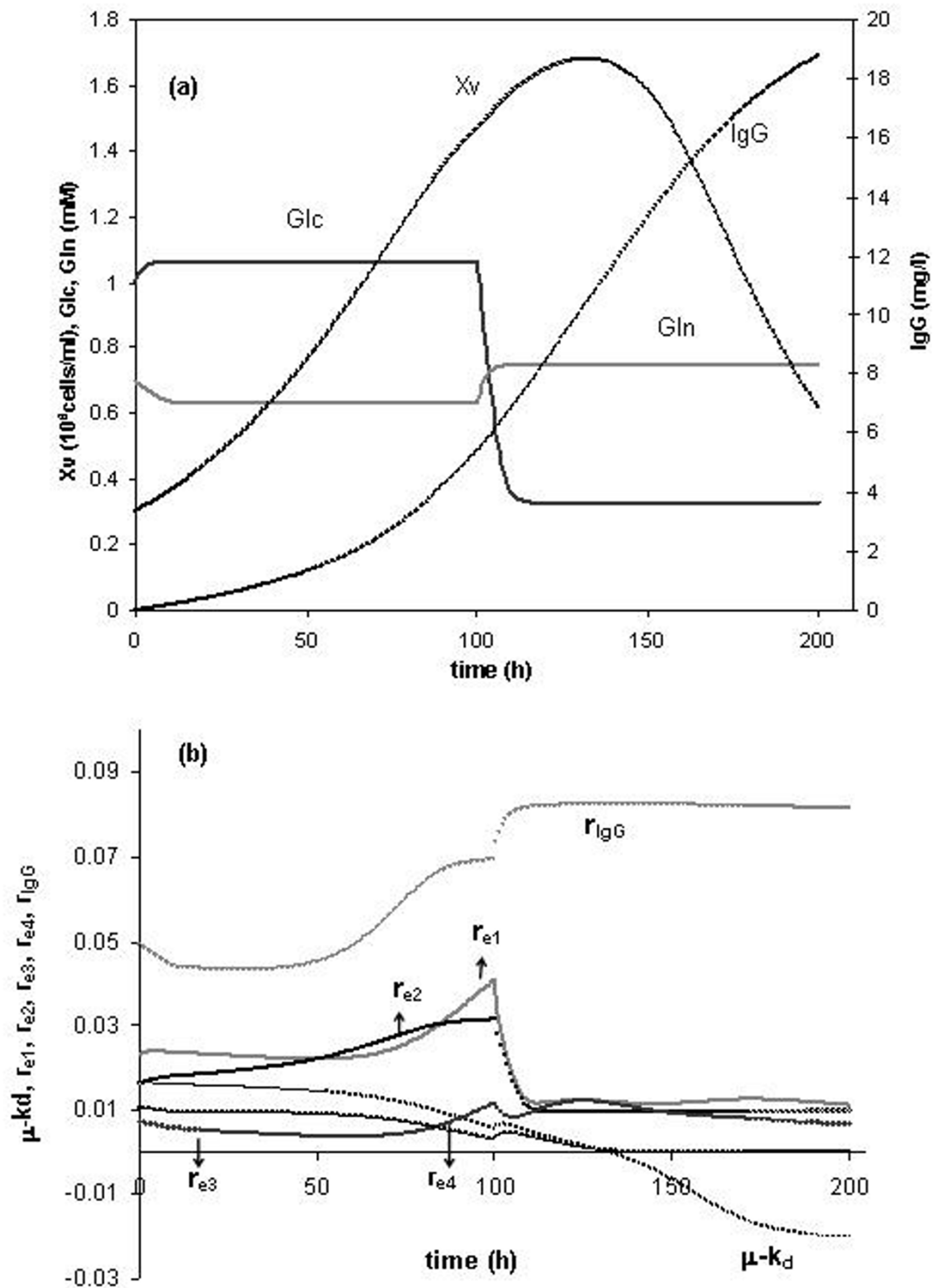


Figure 8
Optimisation results with optimised medium. (a) Predicted optimal trajectories of viable cells, glucose, glutamine and product concentrations starting with low levels of glucose and glutamine. (b) Distribution of intracellular elementary modes over the time course of the process for the optimal strategy.

eq. (3a) in order to \mathbf{w} , yielding, after some manipulations, the following linear time-varying system:

$$\frac{d}{dt} \left(\frac{\partial \mathbf{c}}{\partial \mathbf{w}} \right) = \mathbf{A} \frac{\partial \mathbf{c}}{\partial \mathbf{w}} + \mathbf{B} \quad (\text{M2})$$

with

$$\mathbf{A} = \mathbf{K} \left\langle \rho_i \frac{\partial \varphi_j}{\partial \mathbf{c}} + \varphi_j \frac{\partial \rho_j}{\partial \mathbf{c}} \right\rangle - \mathbf{DI}_n, \quad (\text{M3})$$

$$\mathbf{B} = \mathbf{K} \left\langle \varphi_j \frac{\partial \rho_j}{\partial \mathbf{w}} \right\rangle, \quad (\text{M4})$$

These equations must be integrated along with hybrid model eqs. (3)–(4). The initial value is $(\partial \mathbf{c} / \partial \mathbf{w})_{t=0} = \mathbf{0}$ because the initial state is independent of parameters \mathbf{w} . The evaluation of matrices \mathbf{A} and \mathbf{B} require the sensitivities $\partial \varphi_j / \partial \mathbf{c}$, $\partial \rho_j / \partial \mathbf{w}$ and $\partial \rho_j / \partial \mathbf{c}$. The first term is obtained by analytical derivation of known functions (in our case, by the derivation of eq. (9)). The other two matrices are obtained by backpropagating the identity matrix through the neural network. The backpropagation of a given identity matrix column 'i' results in the evaluation of vectors $\partial \rho_i / \partial \mathbf{w}$ and $\partial \rho_i / \partial \mathbf{c}$.

Evaluation of prediction risk

The trust region is the subspace of the input domain, where the model was properly validated with experimental data, showing low modeling error. Model predictions, \mathbf{c}^* , outside the trust region may have a high risk, $\text{RISK}(\mathbf{c}^*)$, of being unreliable. The use of unreliable model predictions for process control should be avoided. For this reason, the value of the risk is used as a constraint to the optimisation (13). Here, the trust region was defined by nc ellipsoidal clusters of the form:

$$y_C(\mathbf{c}, \mathbf{m}_j, \Sigma) = e^{-0.5(\mathbf{c}-\mathbf{m}_j)^T \Sigma^{-1}(\mathbf{c}-\mathbf{m}_j)} \quad (\text{M5})$$

with \mathbf{m}_j the cluster centres and $\Sigma = \text{diag}\{\sigma_i^2\}$ the diagonal covariance matrix. The choice of the clusters number (nc) is done by trial and error according to the criteria of goodness of measurement density representation. The rule $nc = P/3$, with P the number of measured patterns normally provides acceptable results. The interpolation tolerances (IT) defines the distance to measured pattern such that $y_C > 0.5$. The clusters width is then given by:

$$\sigma = [-2 \ln(0.5/n)]^{-1/2} \text{IT} \quad (\text{M6})$$

The k-means algorithm was used to calculate \mathbf{m}_j . The final set of clusters forms a continuous density function $f: \mathbf{c} \rightarrow v$ by applying the maximum operator:

$$v(\mathbf{c}) = \max_j \{y_C(\mathbf{c}, \mathbf{m}_j, \Sigma)\}. \quad (\text{M7})$$

The output $v(\mathbf{c})$ is a scalar between 0 and 1 that can be interpreted as the degree of membership of \mathbf{c} to the data set used for training the black-box model (4). Low v values (i.e. \mathbf{c} vectors out of the interpolation tolerance) are an indication of high risk of the black box model outputs being unreliable. Finally, the risk of black box model unreliability is given by: $\text{RISK}(\mathbf{c}) = 1 - v(\mathbf{c})$ (M8)

Abbreviations

BHK – Baby Hamster Kidney cells

CHO – Chinese Hamster Ovary cells

EM – Elementary Flux Mode

EP – Extreme Pathway

FBA – Flux Balance Analysis

IgG1-IL2 – Immunoglobulin type one linked to interleukin type two

LP – Linear Programming

MFA – Metabolic Flux Analysis

MPA – Metabolic Pathway Analysis

Authors' contributions

All authors read and approved the final manuscript. RO developed the software. APT, CA and RO participated in the model implementation. APT performed experimental work. PA, MJTC and RO designed and coordinated the study. APT and RO drafted the manuscript.

Acknowledgements

Support from Sartorius BBI and Merck KGaA is gratefully acknowledged. Financial support for this work was provided by the Portuguese *Fundação para a Ciência e Tecnologia* through project POCTI/BIO/57927/2004 and PhD grant SFRH/BD/13712/2003.

References

- Stephanopoulos G: **Metabolic fluxes and metabolic engineering.** *Metab Eng* 1999, **1**(1):1-11.
- Varma A, Palsson BO: **Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use.** *Bio-Technology* 1994, **12**(994-998):994-998.
- Bonarius HPJ, Schmid G, Tramper J: **Flux analysis of underdetermined metabolic networks: the quest for the missing constraints.** *Trends in Biotechnology* 1997, **15**(8):308-314.
- Pramanik J, Keasling JD: **Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass.** *Biotech and Bioeng* 1997, **56**(4):398-421.
- Edwards JS, Ibarra RU, Palsson BO: **In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.** *Nat Biotechnol* 2001, **19**(2):125-130.

6. Edwards JS, Covert M, Palsson B: **Metabolic modelling of microbes: the flux-balance approach.** *Environ Microbiol* 2002, **4(3)**:133-140.
7. Lee DY, Fan LT, Park S, Lee SY, Shafie S, Bertok B, Friedler F: **Complementary identification of multiple flux distributions and multiple metabolic pathways.** *Metab Eng* 2005, **7(3)**:182-200.
8. Schuster S, Dandekar T, Fell DA: **Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.** *Trends Biotechnol* 1999, **17(2)**:53-60.
9. Schuster S, Fell DA, Dandekar T: **A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.** *Nat Biotechnol* 2000, **18(3)**:326-332.
10. Klamt S, Stelling J: **Two approaches for metabolic pathway analysis?** *Trends Biotechnol* 2003, **21(2)**:64-69.
11. Schilling CH, Letscher D, Palsson BO: **Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective.** *Journal of Theoretical Biology* 2000, **203(3)**:229-248.
12. Edwards JS, Palsson BO: **How will bioinformatics influence metabolic engineering?** *Biotechnol Bioeng* 1998, **58**:162-169.
13. Provost A, Bastin G: **Dynamic metabolic modeling under balanced growth condition.** *J Process Control* 2004, **14**:717-728.
14. Mahadevan R, Burgard AP, Famili I, Van Dien S, Schilling CH: **Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals.** *Biotech and Bioprocess Eng* 2005, **10(5)**:408-417.
15. Dhir S, Morrow KJ Jr, Rhinehart RR, Wiesner T: **Dynamic optimization of hybridoma growth in a fed-batch bioreactor.** *Biotechnol Bioeng* 2000, **67(2)**:197-205.
16. Mahadevan R, Doyle IF: **On-line optimization of recombinant product in a fed-batch bioreactor.** *Biotechnol Prog* 2003, **19(2)**:639-646.
17. Teixeira A, Cunha AE, Clemente JJ, Moreira JL, Cruz HJ, Alves PM, Carrondo MJ, Oliveira R: **Modelling and optimization of a recombinant BHK-21 cultivation process using hybrid grey-box systems.** *J Biotechnol* 2005, **118(3)**:290-303.
18. Teixeira AP, Clemente JJ, Cunha AE, Carrondo MJ, Oliveira R: **Bioprocess iterative batch-to-batch optimization based on hybrid parametric/nonparametric models.** *Biotechnol Prog* 2006, **22(1)**:247-258.
19. Bonvin D: **Optimal operation of batch reactors: a personal view.** *J process Control* 1998, **8**:355-368.
20. Banga JR, Balsa-canto E, Moles CG, Alonso AA: **Dynamic optimization of bioreactors: a review.** *Indian Natn Sci Acad* 2003, **69**:359-401.
21. Rockafellar R: **Convex analysis.** Princeton, Princeton University Press; 1970.
22. Gagneur J, Klamt S: **Computation of elementary modes: a unifying framework and the new binary approach.** *BMC Bioinformatics* 2004, **5**:175-196.
23. Gadkar KG, Gunawan R, Doyle FJ 3rd: **Iterative approach to model identification of biological networks.** *BMC Bioinformatics* 2005, **6**:155.
24. Stephanopoulos G, Aristidou AA, Nielsen J: **Metabolic Engineering.** Elsevier Science 1998.
25. Oliveira R: **Combining first principles modelling and artificial neural networks: a general framework.** *Computers & Chemical Engineering* 2004, **28(5)**:755-766.
26. Thompson ML, Kramer MA: **Modeling chemical processes using prior knowledge and neural networks.** *Process Systems Engineering* 1994, **40(8)**:1328-1340.
27. van Can HJ, te Braake HA, Bijman A, Hellinga C, Luyben KC, Heijnen JJ: **An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances: Part II.** *Biotechnol Bioeng* 1999, **62(6)**:666-680.
28. Psychogios D, Ungar LH: **A hybrid neural network-first principles approach to process modeling.** *AIChE Journal* 1992, **38(10)**:1499-1511.
29. Schubert J, Simutis R, Doors M, Havlik I, Lübbert A: **Hybrid Modelling of Yeast Production Processes.** *Chem Eng Technol* 1994, **17**:10-20.
30. Krishnakumar K: **Micro-Genetic Algorithms for Stationary and Non-Stationary Function Optimization.** In *SPIE: Intelligent Control and Adaptive Systems* Philadelphia, PA; 1989:1196.
31. Carroll DL: **Chemical Laser Modeling with Genetic Algorithms.** *AIAA J* 1996, **34(2)**:338-346.
32. Cruz HJ, Conradt HS, Dunker R, Peixoto CM, Cunha AE, Thomaz M, Burger C, Dias EM, Clemente J, Moreira JL, et al.: **Process development of a recombinant antibody/interleukin-2 fusion protein expressed in protein-free medium by BHK cells.** *J Biotechnol* 2002, **96(2)**:169-183.
33. Follstad BD, Balcarcel RR, Stephanopoulos G, Wang DI: **Metabolic flux analysis of hybridoma continuous culture steady state multiplicity.** *Biotechnol Bioeng* 1999, **63(6)**:675-683.
34. Gódia F, Cairó JJ: **Metabolic engineering of animal cells.** *Bioprocesses and Biosystems Engineering* 2002, **24**:289-298.
35. Klamt S, Stelling J, Ginkel M, Gilles ED: **FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps.** *Bioinformatics* 2003, **19(2)**:261-269.
36. Gambhir A, Korke R, Lee J, Fu PC, Europa A, Hu WS: **Analysis of cellular metabolism of hybridoma cells at distinct physiological states.** *J Biosci Bioeng* 2003, **95(4)**:317-327. [<http://www.rcsb.org>].
37. Reitzer L, Vice BM, Kennel D: **Evidence that glutamine, not sugar, is the major energy source for cultured HeLa Cells.** *The Journal of Biological Chemistry* 1979, **254(8)**:2669-2676.
39. Ozturk SS, Palsson BO: **Growth, metabolic, and antibody production kinetics of hybridoma cell culture: 1. Analysis of data from controlled batch reactors.** *Biotechnol Prog* 1991, **7(6)**:471-480.
40. Ozturk SS, Palsson BO: **Growth, metabolic, and antibody production kinetics of hybridoma cell culture: 2. Effects of serum concentration, dissolved oxygen concentration, and medium pH in a batch reactor.** *Biotechnol Prog* 1991, **7(6)**:481-494.
41. Cruz HJ, Freitas CM, Alves PM, Moreira JL, Carrondo MJ: **Effects of ammonia and lactate on growth, metabolism, and productivity of BHK cells.** *Enzyme Microb Technol* 2000, **27(1-2)**:43-52.
42. Xie L, Wang DIC: **Material balance studies on animal cell metabolism using a stoichiometrically based reaction network.** *Biotechnology and Bioengineering* 1996, **52**:579-590.
43. Bonarius HPJ, Hatzimanikatis V, Meesters KPH, Gooijer CD, Schmid G, Tramper J: **Metabolic flux analysis of hybridoma cells in different culture media using mass balances.** *Biotechnology and Bioengineering* 1996, **50**:299-318.
44. Doverskog M, Ljunggren J, Ohman L, Haggstrom L: **Physiology of cultured animal cells.** *J Biotechnol* 1997, **59(1-2)**:103-115.
45. Vriezen N, Romein B, Luyben KCAM, van Dijken JP: **Effects of glutamine supply on growth and metabolism of mammalian cells in chemostat culture.** *Biotech and Bioeng* 1997, **54(3)**:272-286.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

