Dynamic Hybrid Model for Nanobody-based Antivenom Production (scorpion antivenon) with *E. coli* CH10-12 and *E. coli* NbF12-10.

David Camilo Corralesa, Susana María Alonso Villelab, Balkiss Bouhaouala-Zaharc,d, Julien Cescuta, Fayza Daboussia,b, Michael O'donohueb, Luc Fillaudeaub, César Arturo Aceves-Larab\*

a INRAE, UMS (1337) TWB, 135 Avenue de Rangueil, 31077 Toulouse, France

b TBI, Université de Toulouse, CNRS, INRAE, INSA, 135 Avenue de Rangueil, 31077 Toulouse, France

c Laboratoire des Biomolécules, Venins et Applications Théranostiques (LBVAT), Institut Pasteur de Tunis, 13 Place Pasteur, BP-74, 1002 Le Belvédère, Tunis, Tunisia

d Faculté de Médecine de Tunis, Université Tunis El Manar, 15 rue Djebel Lakhdhar, 10007 Bab Saadoun Tunis, Tunisia

Corresponding author: aceves@insa-toulouse.fr

Abstract

Immunotherapy is a specific treatment for scorpion stings, with antibody fragments being used to neutralize scorpion neurotoxins. Nanobodies (VHH), fragments of camelid antibodies, were successfully produced intracellularly in *Escherichia coli WK6* in fed-batch cultures. Their production was further enhanced by dynamic modelling. Two dynamic modelling approaches to describe nanobody CH12-10 production as a function of the induction temperature are proposed in this work. The first one is a kinetic model and the second is a hybrid approach, coupling a mass balance with support vector machine (SVM). Both models were calibrated and validated with independent data sets. Results reveal that the hybrid model procures better predictions than the kinetic model. Finally, the hybrid model was improved by retraining the SVM Model, resulting in a Normed Root Mean Square Error (NRMSE) values between 0.1148 and 0.8523.

**Keywords:** Nanobodies, *Escherichia coli,* Hybrid Model*,* SVM, Dynamic Models

* 1. Introduction

Scorpion stings pose a serious health problem in tropical and subtropical countries (Chippaux and Goyffon, 2008). Serotherapy targets the neurotoxins in the scorpion venom by using fragments of equine antibodies (F(ab)’2). Engineered, toxin-specific Nanobodies (VHH), which are fragments of camelid antibodies, have been found to have better tissue penetration and lower molecular weight (15 kDa vs 100 kDa) than their equine counterparts (Bouhaouala-Zahar *et al*. 2011). These nanobodies are the best candidates to produce bispecific antibodies (Deffar *et al*. 2009). Their production in microbial hosts, such as *Escherichia coli*, is a frequent practice in the pharmaceutical industry to enhance antiserum production (Alonso Villela *et al*. 2023). However, defining optimal production conditions (induction, temperature, etc.) is a trial-and-error exercise. The production of recombinant proteins is usually modeled using Monod or Luedeking-Piret kinetics (Zheng *et al.* 2005; Hua *et al.* 2006), and rarely using dynamic models (Alonso Villela *et al.* 2021). Hybrid modeling approaches have recently gained importance (Badr *et al.* 2023, Kaya *et al.* 2023). Accordingly, in this paper we have compared the use of a kinetic model and a hybrid kinetic model integrating a Support Vector Machine (SVM).

* 1. Materials and Methods
		1. Microbial strains and culture medium

Cultures were carried out with ampicillin-resistant recombinant strains *E. coli* CH10-12 (Alonso Villela *et al.* 2021) and *E. coli* NbF12-10 (Hmila *et al*. 2010). The nanobodies were produced after induction with a 1 mM IPTG pulse. Cultures were initiated by spreading a small aliquot of a glycerol stock on LB (Lysogeny Broth) agar plates. After incubation, a single well-isolated colony was used to inoculate 15 mL LB medium. The bioreactor inoculum was prepared by diluting 1 mL of the LB culture in 100 mL of Minimal Medium (MM, 5 g/L glucose) and incubated at 37 °C. MM was also used for the bioreactor batch. However, for fed-batch and production phases, a modified MM with 300 g/L glucose was used, both described elsewhere (Alonso Villela *et al.* 2021).

* + 1. Experimental setup

Cultures were conducted in a 5 L bioreactor (Biostat® B-DCU, Sartorius) with a working volume of 2 L. The bioreactor was equipped with dissolved oxygen (pO2), pH, temperature (T), and pressure (P) sensors, and two six-flat-blade Rushton turbines a three-blade marine impeller. Temperature was set to 37 °C for the batch and fed-batch modes, pO2 kept over 15 %, and the pH was regulated at 7 for the entire culture. Induction of *E. coli* CH10-12 was carried at temperatures 28 °C, 29 °C, 30 °C, 32 °C, 33 °C, and 37 °C, for a duration of 6 h, and up to 35 h. *E. coli* NbF12-10 was carried at 29 °C for up to 35 h. The batch phase was started with 100 mL of inoculum in 1.5 L MM. Upon glucose depletion, the feed was started (µ of 0.38 h-1) by the BioPAT® MFCS (Sartorius) software, and during the production phase the feed was set to 4 g/h of glucose.

* + 1. Biomass quantification

The optical density (OD) of the cell culture was measured by spectroscopy at 600 nm. Biomass cell dry weight (cdw) was quantified by gravimetry. During the batch phase an aliquot (5 to 10 mL) of culture was removed and filtered in pre-weighted polyamide filters. During the fed-batch and production phases 1.5 mL of culture were centrifuged in a pre-weighted Eppendorf tube.

* + 1. Glucose quantification

Glucose was quantified both by enzymatic analysis and high-performance liquid chromatography (HPLC) coupled with a photodiode array (UV)

* + 1. Nanobody quantification

Periplasmic proteins were extracted by osmotic shock and purified by IMAC. The eluates were separated using SDS-PAGE in reducing conditions and later quantified by image densitometry according to (Alonso Villela *et al.*, 2020).

* + 1. Gas analysis

Inlet and outlet gas were analyzed for carbon dioxide (CO2) and oxygen (O2) using photoacoustic spectroscopy and magneto-acoustic spectroscopy, respectively.

* 1. Model description

The bioreactor is assumed to be an infinitely mixed culture without any transfer limitation. The mass balances for biomass, glucose, proteins and volume are given by Eq. (1) – (4).

|  |  |
| --- | --- |
| $$\frac{dX}{dt}=r\_{X}-X\*\frac{F\_{in}}{V}$$ | (1) |
| $$\frac{dS}{dt}=\frac{F\_{in}}{V}\left(S\_{in}-S\right)-r\_{S}-S\*\frac{F\_{in}}{V}$$ | (2) |
| $$\frac{dP}{dt}=r\_{P}-P\*\frac{F\_{in}}{V}$$ | (3) |
| $$\frac{dV}{dt}=F\_{in}$$ | (4) |

*X* and *S* are the concentrations of biomass and glucose, in g/L, and *P* is the concentration of nanobody proteins, in mg/L. *Fin* is the glucose flowrate, in L/h, *V* is the volume of the liquid phase in the bioreactor, in L, *Sin* is the concentration of glucose in the feed, in g/L, and *rX* is the production rate of biomass, in g/h, and defined by Eq. (5). *rS* is the consumption rate of glucose, in g/h, as defined by Eq. (6).

|  |  |
| --- | --- |
| $$r\_{X}=µ\*X$$ | (5) |
| $$r\_{S}=-\frac{1}{Y\_{SX}}\*µ\*X$$ | (6) |

With µ as the specific growth rate, in h-1, as defined by a Monod kinetics in Eq. (7), and YSX as the yield coefficient of substrate, in g cdw/g S.

|  |  |
| --- | --- |
| $$µ=\frac{µ\_{max}\*S}{K\_{S}+S}$$ | (7) |
| $$µ\_{max}=µ\_{max}^{'}\*T+b$$ | (8) |

µmax is the maximum specific growth rate in h-1 and KS is the substrate saturation coefficient, in g/L. µ’max is a temperature-dependent factor in h-1, T is the temperature of the bioreactor, in °C, and b is a temperature-independent factor, in h-1. The production rate of nanobody proteins, *rP*, is in mg/h, and defined as follows:

|  |  |
| --- | --- |
| $$r\_{P}= \left\{\begin{matrix}\begin{matrix}0 \\q\_{P}\*X \end{matrix}&\begin{matrix}if there is not induction\\if there is an induction\end{matrix}\end{matrix}\right.$$ | (9) |

Where *qp* is the specific productivity of the nanobody production (mgprotein/g cdw/h). Two models were proposed to predict the specific productivity of the nanobody production. In the first one, *qP* is modelled using a temperature-dependent equation, and is described in section *3.1. Dynamic kinetic Model*. In the second one, *qP* is modelled with a hybrid model using a Support Vector Machine (SVM), described in section *3.2. Hybrid Model approach*.

* + 1. Dynamic kinetic model

The dynamic kinetic model uses a kinetic for *qP* that is function of temperature as presented in Eq. (11) and established by (Alonso Villela *et al.*, 2021).

|  |  |
| --- | --- |
| $$q\_{P}=-α\*µ+γ\_{1}\*exp(-Ap\_{1}/T)-γ\_{2}\*exp(-Ap\_{2}/T)+β$$ | (11) |

With *α* as the growth-dependent factor, in mg/g cdw, *γ1* and *γ2* as the pre-exponential coefficients, in mg/g cdw/h, *Ap1* and *Ap2* as the activation and inactivation coefficient, respectively in °C, and *β* as the temperature-independent coefficient in mg/g cdw/h.

* + 1. Dynamic hybrid model approach

The hybrid model integrates a SVM learner to compute qP. The SVM was trained with experimental data provided by (Alonso Villela *et al.*, 2021).

|  |  |
| --- | --- |
| $q\_{P}=f(T,µ)$ | (12) |

Four kernel functions (linear, quadratic, cubic and Gaussian) were tested using Matlab2020a. The Gaussian kernel obtained the best performance using 5 k-fold cross validation.

* + 1. Statistical analysis

Models prediction quality for protein dynamic production were analyzed using the normed root mean square error (NMRSE) defined as:

|  |  |
| --- | --- |
| $$NRMSE=\sqrt{\frac{\sum\_{i=1}^{T}(Y\_{exp}-\hat{Y})^{2}}{n}}/(Y\_{exp, max}-Y\_{exp,  min})$$ | (13) |

With 𝑛 the number of experimental data, *Yexp* the experimental data corresponding to time *ti*, $\hat{Y}$ the data simulated at *ti*, *Y𝑚𝑖𝑛*, and *Y𝑚𝑎𝑥* the minimum and maximum data values.

* 1. Results
		1. Model calibration

The models were calibrated using *E. coli* CH10-12 fed-batch cultures induced at 28, 32, 33 and 37 °C using the *fmincon* function Matlab2020a. Table 1 shows model calibrated parameters. *µmax* and *Ysx* have similar values (0.67*<µmax<*0.87 and 0.29<*Ysx*<0.47 respectively) to those reported in literature (Alonso Villela et al., 2021).

**Table 1**. Model Parameters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameters | $$µ\_{max}$$ | *Ks* | *Ysx* | $$µ'\_{max}$$ | *b* |
| Value | 0.7853 | 0.0001 | 0.2361 | 0.3538 | 0.0726 |

Figure 1 shows both kinetic model and hybrid model simulations for two data sets (28 °C and 32 °C). Both models correctly estimate the final production of biomass and substrate.

|  |  |
| --- | --- |
|  |  |

**Figure 1**. Experimental data of *E. coli* CH10-12 induced at 28 °C (left) and 32 °C (right) used for calibration of glucose (o), biomass (<) and proteins (o), kinetic model (- -), and hybrid model (–).

The novel hybrid model provides a better estimation of the dynamics of protein production than the kinetic model. Models were validated with three independent data sets of *E. coli* CH10-12 fed-batch cultures induced at temperatures 29 and 30 °C, and *E. coli* NbF12-10 induced at 29 °C. Figure 2 shows both model simulations and experimental data for two data sets of *E. coli* CH10-12 (29 °C and 30 °C). Once again, the hybrid dynamic model has a better prediction for proteins than the kinetic model.

|  |  |
| --- | --- |
|  |  |

**Figure 2.** Experimental data of *E. coli* CH10-12 induced at 29 °C (left) and 30 °C (right) used for validation of glucose (o), biomass (<) and proteins (o), kinetic model (- -), and hybrid model (–).

|  |  |
| --- | --- |
|  |  |
|  |  |

**Figure 3.** Experimental data for protein concentration (o) in *E. coli* CH10-12 induced at 28 °C (upper-left), 33 °C (upper-right), 29 °C (lower-left), and 30 °C (lower-right), hybrid model for calibration (- -), Case 1 (\*-), Case 2 (·-) and Case 3 (–).

The hybrid model, SVM, was retrained using experimental data other than calibration to improve protein simulations. Four cases were compared:

* Case 1: CH10-12 culture data used for calibration.
* Case 2: CH10-12 culture data used for calibration and induced at 30 °C.
* Case 3: CH10-12 culture data used for calibration and induced at 29 °C, and 30 °C.
* Case 4: CH10-12 culture data used for calibration and induced at 29 °C, 30 °C, and NbF12-10 data induced at 29 °C.

Table 2 reveals that the hybrid approach achieved better accuracy than the kinetic approach. Case 2 and Case 3 have similar NRMSE values (Table 2 and Figure 3). These results show that the hybrid model could be improved by retraining the experimental data. Finally, the hybrid model is able to simulate the production of the protein correctly for two different *E. coli* strains (Table 2).

**Table 2.** NRMSE analysis for proteins simulation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Strain | Culture | Kinetic | Case 1 | Case 2 | Case 3 | Case 4 |
| CH10-12 | 28 °C | 3.109 | 0.577 | 0.911 | 0.153 | 0.152 |
| CH10-12 | 33 °C | 7.427 | 0.118 | 0.534 | 0.195 | 0.170 |
| CH10-12 | 37 °C | 18.011 | 0.635 | 1.049 | 0.453 | 0.429 |
| CH10-12 | 32 °C | 29.802 | 0.860 | 0.168 | 0.141 | 0.538 |
| CH10-12 | 30 °C | 3.107 | 0.209 | 0.0879 | 0.115 | 0.125 |
| CH10-12 | 29 °C | 3.2463 | 2.102 | 4.132 | 0.254 | 0.421 |
| NbF12-10 | 29 °C | 12.775 | 4.010 | 8.6452 | 0.696 | 0.852 |

* 1. Conclusions

The data from recombinant protein production produced in *E. coli* CH10-12 fed-batch cultures were used to build two dynamic models to optimize the nanobody production for future cultures. In this study, two dynamic models were built, a kinetic approach and a hybrid approach based on SVM. The hybrid approach reached a better accuracy to predict nanobody production compared to the kinetic approach. The protein concentration predictions were better by retraining the hybrid model using new experimental data with NRMSE values between 0.11 and 0.85. Finally, this approach shows that it is possible to accurately simulate protein production from two different *E. coli* strains.

Acknowledgments

This work was funded (or co-funded) by the European Union under the Horizon Europe project Bioindustry 4.0, grant n. 101094287

References

S.M. Alonso Villela, H. Kraïem, B. Bouhaouala‐Zahar, C. Bideaux, C.A. Aceves Lara, L. Fillaudeau, 2020. A protocol for recombinant protein quantification by densitometry. MicrobiologyOpen, 9, 1175–1182.

S.M. Alonso Villela, H. Ghezal-Kraïem, B. Bouhaouala-Zahar, C. Bideaux, C.A. Aceves Lara, L. Fillaudeau, 2021. Effect of temperature on the production of a recombinant antivenom in fed-batch mode. Appl Microbiol Biotechnol, 105, 1017–1030.

S.M. Alonso Villela, H. Kraïem-Ghezal, B. Bouhaouala-Zahar, C. Bideaux, C.A. Aceves Lara, L. Fillaudeau, 2023. Production of recombinant scorpion antivenoms in *E. coli*: current state and perspectives. Appl Microbiol Biotechnol, 107, 4133–4152

S. Badr, K. Oishi, K. Okamura, S. Murakami, H. Sugiyama, 2023. Hybrid modelling and data-driven parameterization of monoclonal antibody cultivation processes: Shifts in cell metabolic behavior, Comput Aided Chem Eng, 52, 985-990.

B. Bouhaouala-Zahar, R. Ben Abderrazek, I. Hmila, N. Abidi, S. Muyldermans, M. El Ayeb, 2011. Immunological aspects of scorpion toxins: current status and perspectives. Inflamm Allergy Drug Targets 10, 358–368

J.P. Chippaux, M. Goyffon, 2008. Epidemiology of scorpionism: A global appraisal. Acta Trop, 107, 71–79.

K. Deffar, H. Shi, L. Li, X. Wang, X. Zhu, 2009. Nanobodies - the new concept in antibody engineering. Afr J Biotechnol, 8, 2645–2652.

I. Hmila, D. Saerens, R.B. Abderrazek, C. Vincke, N. Abidi, Z. Benlasfar, J. Govaert, M.E. Ayeb, B. Bouhaouala‐Zahar, S. Muyldermans, 2010. A Bispecific Nanobody to Provide Full Protection against Lethal Scorpion Envenoming. FASEB j, 24 (9), 3479–3489

X. Hua, D. Fan, Y. Luo, X. Zhang, H. Shi, Y. Mi, X. Ma, L. Shang, G. Zhao, 2006. Kinetics of High Cell Density Fed-batch Culture of Recombinant *Escherichia coli* Producing Human-like Collagen. Chin J Chem Eng, 14, 242–247.

S. Kaya, H. Kaya, A.W. Rogersa, D. Zhang, 2023. Integrating hybrid modelling and transfer learning for new bioprocess predictive modelling, Comput Aided Chem Eng, 52, 985-990.

Z.Y. Zheng, S.J. Yao, D.Q. Lin, 2005. Using a kinetic model that considers cell segregation to optimize hEGF expression in fed-batch cultures of recombinant *Escherichia coli*. Bioprocess Biosyst Eng, 27, 143–152.