Development of Soft Sensors for Online Biomass Prediction in Production of Hepatitis B Vaccine

Ahmad Beiroti 1, Seyed Nezamedin Hosseini 1,*, Dariush Norouzian 2, Mohammad Reza Aghasadeghi 3,*, Leila Hatami Giklou Jajan 1

1. Department of Recombinant Hepatitis B Vaccine, Pasteur Institute of Iran, Tehran, Iran
2. Department of Pilot Nano-Biotechnology, Pasteur Institute of Iran, Tehran, Iran
3. Department of Hepatitis and AIDS, Pasteur Institute of Iran, Tehran, Iran
* Correspondence seyyednezam@yahoo.com

Received: 19.10.2021; Accepted: 15.11.2021; Published: 3.04.2022

Abstract: Neural networks can provide highly accurate and robust solutions for complex non-linear kinetic like bioprocess reactions. It is applied as a soft sensor in the biotechnology industry for measuring and analyzing parameters that can hardly be reached online, as well as the reproducibility of the product without significant deviation. This study attempted to obtain the best neural network structure for online estimation of P. pastoris yeast biomass, which is used to express the hepatitis B surface antigen (HBsAg). During the fed-batch fermentation process, the CO2 evolution rate (CEO), ammonia consumption rate (ACR), and methanol consumption rate (MCR) were considered as inputs and biomass (WCW) as ANN output parameter. The results showed that after training the neural network structure using 4 fed-batch fermentation batches performed in the laboratory, biomass estimation (WCW) was obtained with high accuracy at specific times of the fed-batch fermentation process. R-Squared and RSME between actual and ANN estimations were 0.999 and 9.57, respectively.

Keywords: soft sensor; artificial neural network; online biomass estimation; fed-batch fermentation; recombinant Pichia pastoris; HBsAg vaccine.

© 2022 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution [CC BY] license [https://creativecommons.org/licenses/by/4.0/].

1. Introduction

In the last 4 decades, DNA recombinant technology has grown significantly in the biopharmaceutical industry to make therapeutic proteins [1]. The expression of recombinant therapeutic proteins with biological hosts such as bacteria, yeast, and animal cells is another key process in the pharmaceutical industry [2–6]. The market for these recombinant proteins in the United States in 2011 was more than $ 99 billion and will cause their sales to increase in the coming years, including biological drugs. Meanwhile, biomedical drugs play a major market share in the treatment sector [7]. Approved biological drugs by FDA and EMA from 2004 to 2013 showing that most of them are from animal cells (56%), E. coli (26%), yeast (13%), plants, animals transgenic (3%), and insect cells (4%) [8].

The five main types of hepatitis virus that cause liver infection are: Hepatitis A, B, C, D, and E. Hepatitis B virus, with a size of 42 nm, is one of the causes of liver disease worldwide. Its nucleus contains a viral genome consisting of 3,200 base pairs of single-stranded DNA that bind to its protein nucleus and DNA polymerase. The virus nucleus is surrounded by a phospholipid compound that carries most of the determinant factors of surface antigen [9]. It is estimated that approximately 250 to 300 million people are chronically infected with this...
virus each year [10]. This has led to the use of recombinant hepatitis B vaccine and its development worldwide to prevent the spread of liver diseases, including liver cancer. *Pichia pastoris* is a methylotrophic yeast and is one of the host cells that has been given special attention for the production of biological drugs [4, 11]. High expression level, low cost, protein-free growth medium, easy scaling up, low risk of bacteriophage contamination, and non-production of viral DNA or endotoxin are some of the important advantages of this eukaryotic system [12–14]. Recombinant hepatitis B surface antigen (r-HBsAg) is one of the most well-known approved biopharmaceuticals produced in the yeast *Pichia pastoris* [15, 16]. In recombinant *P. pastoris* yeast, the Mut* and MutS* phenotypes are commonly used to produce hepatitis B surface antigen [4, 17, 18]. In both phenotypes, methanol induction plays a role in antigen expression. However, in the MutS* phenotype, the lack of AOX1 as the primary promoter for methanol induction will lead to lower methanol consumption rates [18, 19]. Although the benefits and losses of using these two phenotypes to produce recombinant proteins are still debated, *P. pastoris* Mut* yeast has been considered for producing hepatitis B antigen on an industrial scale [18–20].

The AOX1 promoter is strongly suppressed by carbon sources such as glucose, glycerin, ethanol, and many other sources do not express proteins and are induced in the presence of methanol. The alcohol oxidase enzyme converts methanol to aldehyde in the oxidation pathway [13, 21–24]. Although the AOX1 promoter is a strong promoter for protein expression in the presence of methanol, it also has limitations, including oxygen supply in *P. pastoris* semi-continuous culture method with high methanol concentration in culture medium, high cell density, and high heat of biochemical reaction, which are important issues. This carbon source is flammable and toxic so it can cause problems in the upstream and downstream parts [23]. To overcome this problem, feeding should be based on cell growth in the fermenter to prevent methanol accumulation in the fermenter. According to the significant development in artificial intelligence technologies in recent years, advances in control system design are not dependent on accurate mathematical descriptions of the process. In this regard, the ANN model has been a highly usable method in recent years. In the last decade, ANN has been shown to be a powerful and reliable tool for monitoring and controlling bioprocess systems. ANN does not require knowledge of the process model but requires a lot of data to learn ANN algorithms. Therefore, ANN is optimally used for the online estimation of variables such as cell mass concentration and specific growth rate. Controllers with a combination of ANN with neurodynamic programming can predict and control recombinant proteins production by the semi-continuous growth method in the fermenter. The optimum feed source curve of the carbon source can be obtained by using the desired control systems. The artificial neural network model is a mathematical estimate of a biological neural network [3].

In this study, first, glycerol was used as a carbon source to increase *P. pastoris* cell mass [25, 26]. After the cell mass concentration reaches its maximum value, methanol is added as a carbon source in a specific concentration to express protein and cell mass growth [27–31]. During the methanol induction phase, two basic points must be considered. First, during the addition of methanol, the concentration of glycerol in the medium should be minimized to prevent inhibition of protein expression [23, 31, 32]; otherwise, substrate inhibition occurs, which prevents the growth of cell mass and subsequently recombinant hepatitis B antigen. Usually, the fermentation process of recombinant *P. pastoris* to produce hepatitis B antigen is performed in a fed-batch mode with the carbon source of glycerol and methanol [11, 33–36]. In this feeding method, to obtain high cell density, the preliminary approach is increasing of
productivity and titration of HBsAg protein [27, 37]. Depending on the scale of the process and the design specifications of the fermenter, the cell density should not exceed a certain level so that the amount of required oxygen and the generated heat by the reaction do not adversely affect physiological and metabolic reactions and ultimately decrease the process efficiency. [11, 31, 38]. For this purpose, during fed-batch fermentation, the amount of feed rate is controlled based on cell density. As a result, this method optimizes the cell density of \textit{P. pastoris} and prevents the amount of excess methanol in the culture medium [23, 39]. Online specific growth rate monitoring has clear advantages over offline measurements. Therefore, combining the operation unit with the feed control system is possible to adjust and manipulate this online factor to correct any deviations [14, 40, 41].

The specific growth rate is a factor that is not directly measured. Its indirect calculation with changes in cell mass or biomass can be done by online monitoring by applying sensors or soft-sensors [42–44]. Although there is progress in hardware sensors [45, 46], there is still a series of uncertainties about the proper performance of such sensors in measuring biomass, especially on an industrial scale and over a long process time. As an alternative, special importance is given to soft-sensor for monitoring and controlling variables in fermentation processes [42, 45–47].

2. Materials and Methods

2.1. Materials.

The materials required for the buffers and culture medium were analytical grade and were provided by Merck Company of Germany.

Mut+ yeast strain \textit{Pichia pastoris} His4 (GS115) was used to express HBsAg under the control of the AOX1 promoter [29, 48–50]. Culture medium: Contains minerals ((NH4)2SO4, KH2PO4, MgSO4) with glycerol and vitamins (Biotin, riboflavin) and is enriched with trace elements, KI, CuSO4 × 5H2O, FeCl2.

Operating conditions: temperature 30 °C, pH 5, in aeration conditions 1vvm and stirring at 500 rpm continuously. Laptop programming and computing (Intel® Core ™ i7-2630QM @ 2.00 GHz, 4 GB RAM, MS-Windows 10 homw) which is connected to a 10-liter fermenter (Model Winpact FS01-VB-L) and also Matlab software R2016b (version 9.1.0) was used (Figure 1). The MC-160 methanol analyzer sensor from PTI Instrument (Co., US) and the FerMac 368 fermenter gas analyzer from Gloucestershire, UK (Electrolab Biotech) were used during the fermentation process. Lab type peristaltic pump (HEID_28010, Germany) Heidolph and VWR digital scales (B2T, USA), Eppendorf laboratory centrifuge (Germany) were also used [29, 44].

2.2. Methods.

The fermentation process is performed in four stages to obtain the maximum cell mass [4, 30, 51].

2.3. Offline and online analyses.

According to the protocol in the industrial production of the hepatitis B vaccine, the fresh biomass weight was measured every 24 hours by centrifuging the cell solution and separating the biomass at 3000 rpm for 20 minutes [52, 53]. The concentration of hepatitis B
antigen at the end of the semi-continuous process was measured by the sandwich ELISA method, which contains antibodies to hepatitis B antigen coated in plate wells and conjugated with Radish-Horse peroxidase [54–56].

The initial starting point for the experiments was the methanol induction phase in semi-continuous fermentation. Details of previous stages of fermentation, including Erlenmeyer cultivation, discontinuous fermentation of glycerol, and adaptation phase under methanol conditions, are given in research papers [30, 57, 58].

The process of fed-batch fermentation was started by feeding methanol as a limiting carbon source in the presence of *P. pastoris* with an initial cell wet weight (WCW) density of 212 g / l. The initial volume of the culture medium was considered to be 4 liters. The fermentation process at 30°C, and pH is fixed at 5 by adding 20% by volume of ammonia during the process. Agitation and dissolved oxygen content (saturation 20% <DO) under cascade operation conditions were done (Figure 1). A methanol sensor was used during the fed-batch process to determine methanol concentration in the culture medium [31].

### 2.4. Artificial neural network architecture.

The aim of this study is to design a dynamic neural network to simulate and predict cell mass in the fed-batch fermentation process of the hepatitis B vaccine production line using experimental data from a series of different batches. The input variables are the percentage of carbon dioxide evolution rate in the off-gas, the amount of methanol consumed, and the ammonia consumption rate in the fermenter.

Four batches of the fermentation process were considered to estimate and predict the amount of biomass using the model. The relationship between the biomass and the amount of data obtained from the gas analyzer and the amount of consumed methanol and ammonia data from the digital scale is obtained for each experiment. For this purpose, a simple neural network was used due to the dependence of online biomass and the dependence between previous
biomass and the next. The model must be validated. Also, the model's percentage of error in predicting biomass should be minimized, as shown in Figure 2.

2.5. Model input variables.

Model input variables are methanol consumption rate (MCR) in milliliters, CO2 production percentage (CER), and ammonia consumption rate (ACR) in milliliters.

2.6. Output variable of the model.

The model output variable is biomass in grams.

Figure 2. The basic structure of the neural network in this design.

The network training step is performed by experimental data obtained from the Levenberg Marquardt method based on the propagation algorithm [59]. ANN was considered to determine the non-linear relationship between the three input vectors, including CER, ACR, and MCR, and the output parameter, which is WCW. We used different topologies from ANN 3: x: 1 to obtain the best prediction model in this study. As shown in Figure 2, the input and output layers consist of 3 and 1 nodes and coordinate the independent and dependent variables, respectively. A hidden layer between the input and output layers connects these two layers to weight and bias data and the transfer function. In this study, we used many variables of hidden layers between 2 and 9 nodes, for which the tansig transfer function was used. Finally, according to Table 1, 8 nodes in the hidden layer were optimal. According to Table 1, considering the hidden layer with 8 nodes is the best, and according to Figures 6 and 7, regression of network input and output data using fermentation batches has good results.

Table 1. The optimal number of hidden layer nodes.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>MSE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1024.6</td>
<td>0.9955</td>
</tr>
<tr>
<td>3</td>
<td>837.67</td>
<td>0.9959</td>
</tr>
<tr>
<td>4</td>
<td>315.42</td>
<td>0.9979</td>
</tr>
<tr>
<td>5</td>
<td>421.38</td>
<td>0.9954</td>
</tr>
<tr>
<td>6</td>
<td>542.08</td>
<td>0.9969</td>
</tr>
<tr>
<td>7</td>
<td>345.75</td>
<td>0.9972</td>
</tr>
<tr>
<td>8</td>
<td>325.55</td>
<td>0.9980</td>
</tr>
<tr>
<td>9</td>
<td>553.63</td>
<td>0.9971</td>
</tr>
</tbody>
</table>

2.7. Training, validating, and self-testing procedure.

As a training algorithm, the propagation method based on the Levenberg Marquardt algorithm was used (Figure 3). In this training algorithm, the error between the predicted biomass value and the biomass value obtained through the experiment was measured, and feedback was given from the output layer to the hidden layer and finally to the input layer.
through the weight of the layers. During network training, weights and bias values were adjusted by the algorithm to bring the predicted biomass response closer to the experimental biomass response. 30% of the data obtained in the experimental experiment were used to validate and test the neural network. A network validation test was performed during training to minimize out-of-range data. ANN goes to the validation test when the mean error square is smaller than the previously specified value. Network training stops when the validation error and training are close enough. In the self-testing phase, the ability to predict the trained network was evaluated by a group of data that did not participate in the network training phase (Figure 3).

**Figure 3.** Flowchart of model-based training method for training, validation, and testing neural network and predicting the amount of cell mass. The loop is repeated until the grid covers the error function up to the value of the trip.

**Figure 5.** The results of validation performance of ANN with 8 nodes.

### 3. Results and Discussion

Fed-batch fermentation is not in a steady-state compared to continuous fermentation; therefore, due to the complex nature of the system, in this operating mode, predicting the behavior of the fermentation process and key variables such as biomass concentration is very challenging. In this study, we considered three parameters of methanol consumption rate
(MCR), CO$_2$ evolution rate (CER), and ammonia consumption rate (ACR) as input vectors for the artificial neural network.

**Figure 6.** Comparison between output and target in (A) Training; (B) Validation; (C) Test; and (D) all used to process data in ANN with 8 nodes.

**Figure 7.** The Comparison of target and output of ANN for four batches of (A) fermentation; (B) R-squared; (C) RMSE; and (D) $\mu$, $\delta$ values.
As in previous research, the CER value is usually one of the key factors in predicting biomass concentration [29, 47, 59]. In this study, the feed rate of methanol is another input parameter that has been considered, and therefore its feed rate during different processes of fed-batch fermentation has varied depending on the concentration of biomass at any time [60–62]. The fermentation profile of P. pastoris, especially Mut+, strongly depends on methanol concentration in the culture medium. Although increasing the concentration of methanol to a certain level accelerates the growth of P. pastoris biomass after increasing it from the limited level in the culture medium, the inhibitory property of methanol at such a concentration prevents the growth of biomass and production. The ACR parameter was selected as another input parameter in this study, used to control the pH that estimates the increase or decrease of biomass. For fermentation processes, other parameters such as initial biomass concentration, temperature, and pH are precisely controlled and kept constant, as increasing data changes can have a negative impact on the quality of biomass estimation using neural networks [18, 19, 63]. Nevertheless, in the presence of fluctuations, ANN can still be used as an estimator of the amount of biomass at a high level of accuracy, given a large number of presented experimental data.

The results of fed-batch fermentation with a carbon methanol source from four experiments are shown in Figure 4. To obtain the best neural network topology for this study, we tested the accuracy of different neural network structures with a hidden layer of 2 to 9 nodes to estimate the WCW of P. pastoris by trial and error. For this purpose, the data obtained from four batches of fed-batch fermentation were used separately to train these neural networks. By training, validating, and testing these neural networks, the feed-forward neural network with eight nodes in the hidden layer showed that it is the best neural network for accurate estimation. To teach this network, a backpropagation algorithm was considered to correct the synaptic weights attached to the neurons in the adjacent layers to achieve the desired degree of convergence. To obtain the best neural network topology, all four sets of data obtained in a fed-batch fermentation process over time were used to train the neural network. The neural network training and testing results are shown in Figures 5 and 6. Figures 5 and 6 show that the best validation performance occurs in epoch 20 with a mean square root of error (RMSE) of 18.26.

![Figure 8. Comparison between predicted values by ANN and actual (offline analyzing) values about total wet cell weight for the 5th fed-batch fermentation, non-given datasets for ANN training.](https://biointerfaceresearch.com/)
The regression coefficient calculated in the network training and testing phase, according to Figure 6, were 0.997 and 0.999, respectively, which shows the high efficiency of the neural network in the training process. In the final and critical phase, we tested the accuracy of the trained neural network to estimate *P. pastoris* biomass in real-time for new fed-batch fermentation experiments according to three neural network online input parameters. As shown in Figure 8, the test results are in accordance with the calculated values. The predicted values were very close to the actual offline obtained values. The R-squared and RMSE values were 0.999 and 9.57, respectively. The results show that ANN with the desired topology and input variables can predict the fresh weight (WCW) of recombinant *P. pastoris*, which express the surface antigen of intracellular hepatitis B, with high accuracy in the process fed-batch fermentation. In other words, it is possible to generate a defined ANN to estimate and control the amount of biomass in the rHBsAg production industry, which is used recombinant *P. pastoris* as a host cell.

**Figure 9.** Four fed-batch fermentation experiments for training the artificial neural network (A–D).
4. Conclusions

In this study, ANN (3: 8: 1) was used to predict the biomass (WCW) of *P. pastoris* yeast to produce HBsAg. By comparing real biomass and predicted biomass based on ANN, it was shown that a reliable soft-sensor could be estimated and control the biomass of this recombinant methylo trophic yeast. This approved toolbox can be critical in the biopharmaceutical industry, such as the industrial production of *P. pastoris* hepatitis B vaccine based in different countries.

References


