

Kroll, P., et al. reported about a soft sensor for monitoring biomass subpopulations in mammalian cell culture processes (Biotechnol. Lett. 39 (2017) 1667-1673). The authors used a turbidity physical sensor to measure viable cell counts (VCC, equivalent to VCD) based on a linear model.

5

Summary

The present disclosure is based, at least in part, on the finding that by selecting certain, specific process variables from historical datasets, useful data-driven models can be obtained, which contain important parameters for the cultivation of CHO cells, such as, VCD (Viable Cell Density), VCV (Viable Cell Volume), glucose and lactate in real time. With the method according to the disclosure, it becomes possible to provide precise online-like values for the target variables over the entire course of the cultivation without sampling.

The object of the disclosure is a method for determining the viable cell density, and/or the viable cell volume and/or the glucose concentration in the cultivation medium and/or the lactate concentration in the cultivation medium for and during the cultivation of a CHO cell expressing an antibody, using exclusively on-line measured values from said cultivation by means of a model for the cultivation of this CHO cell, characterized in that the model based on a feature matrix comprising the features 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T.PV', 'ACO.PV', 'AO.PV', 'N2.PV', 'LGE.PV', 'CO2.PV', 'FED3T.PV', 'OUR' and 'PH .PV' is generated.

The features denote the parameters as follows:

feature	variable/value	unit
ACO.PV	CO2 concentration in off-gas flow	Volume %
ACOT.PV	CO2 total in off-gas flow	Volume %
AO.PV	O2 concentration in off-gas flow	Volume %
CHT.PV	Temperature of cooling element	°C
CO2.PV	CO2 inflow	ml/min
CO2T.PV	CO2 inflow cumulative	ml
DRZ.PV	Revolutions of stirrer	rpm
FED2T.PV	Feed 2 cumulative	ml
FED3T.PV	Feed 3 cumulative	ml
GEW.PV	Fermenter weight	g
LGE.PV	Base addition cumulative	g
N2.PV	N2 inflow	ml/min
O2.PV	O2 inflow	ml/min

feature	variable/value	unit
OUR	Oxygen utilization rate	mol/(l*h)
PH.PV	fermenter pH	-
PO2.PV	Oxygen concentration in fermenter liquid	%
Time	phase	Saturation
TEF.PV	Elapsed time since fermentation start	d
	Fermenter temperature	°C

Thus, in one aspect, the present disclosure provides a method for adjusting the glucose concentration to a target value during the mammalian cell cultivation, comprising the following steps

5 a) determine the current values at least for the bioreactor process variables:

- 'Time' which denotes elapsed time since fermentation start,
- 'CHT.PV' which denotes temperature of cooling element,
- 'ACOT.PV' which denotes CO₂ total in off-gas flow,
- 'FED2T.PV' which denotes feed 2 cumulative,
- 10 - 'GEW.PV' which denotes fermenter weight,
- 'CO2T.PV' which denotes CO₂ inflow cumulative,
- 'ACO.PV' which denotes CO₂ concentration in off-gas flow,
- 'AO.PV' which denotes O₂ concentration in off-gas flow,
- 'N2.PV' which denotes N₂ inflow,
- 15 - 'LGE.PV' which denotes base addition cumulative,
- 'CO2.PV' which denotes CO₂ inflow,
- 'FED3T.PV' which denotes feed 3 cumulative,

- 'OUR' which denotes oxygen utilization rate and

- 'PH.PV' which denotes fermenter pH,

in the cultivation,

b) determine the current glucose concentration in the cultivation medium using the measured values from a) by means of a data-driven model for the mammalian cell cultivation, which was generated using a feature matrix comprising the bioreactor process variables 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T.PV', 'ACO.PV', 'AO.PV', 'N2.PV', 'LGE.PV', 'CO2.PV', 'FED3T .PV', 'OUR' and 'PH.PV',

and

c) adding glucose until the target value is reached if the current glucose concentration from b) is lower than the target value, and thus adjusting the glucose concentration to a target value,

wherein the mammalian cell is a CHO cell.

In one example, the model is generated using the random forest method.

In one example, the training dataset comprises at least 10 cultivation runs, preferably at least 60 cultivation runs.

In one example, the model is obtained with a training dataset that includes cultivation runs of mammalian cells that express a complex IgG, i.e. an antibody that comprises a different form than a wild-type Y-shaped full length antibody, e.g. by comprising additional domains, such as one or more Fabs. In one example, the training dataset also contains cultivation runs of mammalian cells that express standard IgG, i.e. a Y-shaped wild-type-like antibody without additional or deleted domains.

In one example, approximately 80% of the datasets available for the model formation are used as training datasets and the remaining datasets are used as test datasets.

In one example

5 a) the datasets available for the modeling are randomly divided into training
 and test datasets in a ratio of 80:20,

b) the model is formed,

10 c) the mean and the standard deviation for the determination of the target
 parameter for the datasets are determined from the training dataset and the
 mean and the standard deviation for the determination of the target
 parameter for the records are determined from the test dataset,

d) steps a) to c) are repeated until comparable, i.e. within at most 10 %, preferably at most 5 % of each other, mean values and standard deviations are achieved regarding the division between test and training datasets.

In one example, missing data points in the datasets are supplemented by interpolation.

In one example, the datasets contain a data point for at least 60 minutes, preferably a data point for approximately every 5 to 10 minutes.

5

Specific embodiments of the invention

1. A method for determining one or more process variables during the cultivation of a mammalian cell, characterized in that

the process variable(s) are determined solely

10

i) by means of a data-driven model of the cultivation of the mammalian cell, which has been generated with a feature matrix comprising the process variables 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T.PV', 'ACO.PV', 'AO.PV', 'N2.PV', 'LGE.PV', 'CO2.PV', 'FED3T.PV', 'OUR' and 'PH.PV',

and

15

ii) by using solely/only on-line measured values from the cultivation.

2. The method according to embodiment 1, characterized in that the on-line measured values are used at least for the process variables 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T. PV', 'ACO.PV', 'AO.PV', 'N2.PV', 'LGE.PV', 'CO2.PV', 'FED3T.PV', 'OUR' and 'PH.PV' of the cultivation.

20

3. A method for adjusting the glucose concentration to a target value during the cultivation of a mammalian cell, comprising the following steps

25

a) determining the current values at least for the process variables 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T.PV', 'ACO.PV', 'AO.PV', 'N2.PV', 'LGE.PV', 'CO2.PV', 'FED3T.PV', 'OUR' and 'PH.PV' of the cultivation,

30

b) determining the current glucose concentration in the cultivation medium using the values as determined in a) by means of a data-driven model for the cultivation of the mammalian cell, which is generated using a feature matrix comprising the process variables 'Time',

Claims

1. A method for adjusting the glucose concentration to a target value during the mammalian cell cultivation, comprising the following steps

5 a) determine the current values at least for the bioreactor process variables:

- 'Time' which denotes elapsed time since fermentation start,
- 'CHT.PV' which denotes temperature of cooling element,
- 'ACOT.PV' which denotes CO₂ total in off-gas flow,
- 'FED2T.PV' which denotes feed 2 cumulative,
- 10 - 'GEW.PV' which denotes fermenter weight,
- 'CO2T.PV' which denotes CO₂ inflow cumulative,
- 'ACO.PV' which denotes CO₂ concentration in off-gas flow,
- 'AO.PV' which denotes O₂ concentration in off-gas flow,
- 'N2.PV' which denotes N₂ inflow,
- 15 - 'LGE.PV' which denotes base addition cumulative,
- 'CO2.PV' which denotes CO₂ inflow,
- 'FED3T.PV' which denotes feed 3 cumulative,
- 'OUR' which denotes oxygen utilization rate and
- 'PH.PV' which denotes fermenter pH,
- 20 in the cultivation,

b) determine the current glucose concentration in the cultivation medium using the measured values from a) by means of a data-driven model for the mammalian cell cultivation, which was generated using a feature matrix comprising the bioreactor process variables 'Time', 'CHT.PV', 'ACOT.PV', 'FED2T.PV', 'GEW.PV', 'CO2T.PV', 'ACO.PV',

25

‘AO.PV’, ‘N2.PV’, ‘LGE.PV’, ‘CO2.PV’, ‘FED3T .PV’, ‘OUR’ and ‘PH.PV’,

and

5 c) adding glucose until the target value is reached if the current glucose concentration from b) is lower than the target value, and thus adjusting the glucose concentration to a target value,

wherein the mammalian cell is a CHO cell.

2. The method according to claim 1, characterized in that the method is carried out without sampling and exclusively using on-line measured values from this cultivation.

10

3. The method according to one of claims 1 through 2, characterized in that the data-driven model is generated by means of machine learning.

4. The method according to one of claims 1 through 3, characterized in that the data-driven model is generated with a random forest method.

15 5. The method according to one of claims 1 through 4, characterized in that the data-driven model is generated with a training dataset, which comprises at least 10 cultivation runs.

6. The method according to claim 5, characterized in that

20 a) the datasets available for modeling are randomly divided into training and test datasets in a ratio between 70:30 and 80:20,

b) the model is formed,

25 c) the mean value and the standard deviation for the determination of the process variable for the datasets from the training dataset is determined and the mean value and the standard deviation for the determination of the process for the datasets is determined from the test dataset,

d) steps a) to c) are repeated until comparable mean values and standard deviations with regard to the division between test and training dataset are achieved, wherein the division obtained under a) is different with each new run.

7. The method according to one of claims 5 through 6, characterized in that the datasets, which are used to generate the data-driven model, each contain the same number of data points.
- 5 8. The method according to one of claims 5 through 7, characterized in that the data points in the datasets, which are used to generate the data-driven model, are each for the same times of the cultivation.
9. The method according to one of claims 5 through 8, characterized in that missing data points in the datasets are supplemented by interpolation.
- 10 10. The method according to one of claims 5 through 9, characterized in that the datasets contain a data point at least every 144 minutes.
11. The method according to any one of claims 1 through 10, characterized in that the mammalian cell is a CHO-K1 cell.
12. The method according to any one of claims 1 through 11, characterized in that the mammalian cell expresses and secretes an antibody.
- 15 13. The method according to any one of claims 1 through 12, characterized in that the data-driven model is generated with a training dataset that contains complex IgG cultivation runs and standard IgG cultivation runs,

wherein the complex IgG is an antibody that comprises a different form than a wild-type Y-shaped full length antibody, and

20 wherein the standard IgG is a Y-shaped wild-type-like antibody without additional or deleted domains.
14. The method according to any one of claims 1 through 13, characterized in that the cultivation volume is 300 mL or less.