

recombinant CHO-K1 producer clones and a host cell line and were able to derive an extended set of clone performance criteria that not only captured growth and product formation, but also incorporated information on intracellular clone physiology and on metabolic changes during the process. Using these criteria
5 allowed a quantitative clone ranking and allowed to identify metabolic differences between high-producing CHO-K1 clones yielding comparably high product titers.

WO 2011/140093 disclosed a method of assessing the severity of nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and/or liver fibrosis in a subject which includes obtaining a bodily sample from a subject and determining a level of the at
10 least one oxidized fatty acid product in the sample when compared to the sample of a healthy individual.

WO 2011/136515 disclosed that only recently, genome-scale technologies enabled a system-level analysis to elucidate the complex biomolecular basis of protein production in mammalian cells promising an increased process understanding and the deduction of knowledge-based approaches for further process optimization. The
15 document described a method for a rational cell culturing process using such a knowledge-based approach.

Paul, W., et al. (<https://dc.engconfintl.org/ccexvi/161/>) disclosed new approach in metabolic/process modeling and results. They have found that hybrid metabolic models enable a new approach for intracellular metabolic flux calculation and
20 prediction of the metabolic status of the cell of tomorrow. These models enable metabolic driven process control. Artificial neural network succeeded metabolic flux estimation values with high confidence and low root mean squared error on average (RMSE) ~20% (POC).

Current metabolic model based methods rely on manually inserted data into developed models. Corrupt data input can occur and harm the whole model output. When performed at all, outlier detection methods rely on erratic data structure but not on biological relevance and cross-validation of data.
25

Summary

One aim of the current disclosure was to provide methods for the identification or
30 determination of cell cultures affected by a problem, i.e. the alignment and/or consistency control of experimentally determined data, using in silico modelling and metabolic flux analysis. By determining the goodness of fit between the model and

the experimental data cultivations affected by a problem can be identified. The problem can either be a technical problem or a biological problem. A technical problem is based on a failure in the hardware used for performing the cultivation. A biological problem is based on the cell as such, e.g. resulting from bacterial or fungal contamination of the cultivation. Preferably the problem is a technical problem associated with the hardware, i.e. probes, vessels, electronic, devices, analytics etc., used for performing and/or monitoring the cultivation.

Thus, herein are provided methods for verifying cultivation performance by determining the goodness of fit (GoF) of experimental data to a (established) metabolic model.

All methods according to the current disclosure comprise either the following first set of steps:

- fitting the process data acquired during the cultivation of a cell clone expressing a recombinant, heterologous polypeptide, preferably an antibody, using a metabolic model generated for the same cell expressing the same recombinant, heterologous polypeptide, and
- determining that the cultivation is affected by a problem when the obtained fit shows an offset with respect to the raw data of more than 10 %;

or the following second set of steps:

- receiving process data of a cultivation of a cell clone, wherein the cell clone produces a polypeptide, preferably an antibody, heterologous to said cell,
- fitting the data using a metabolic model established for said cell and characterizing the fit by a χ^2 (also written as χ^2) value determined by a statistical correlation method, preferably a Pearson's chi-squared test, and
- identifying the process data to be affected by a problem if the χ^2 value is more than 5.

In one aspect, the present disclosure provides a method for determining if process data acquired during the cultivation of a mammalian or bacterial cell is affected by a problem comprising the following steps:

- fitting the process data acquired during the cultivation of the mammalian or bacterial cell expressing a recombinant, heterologous polypeptide using a metabolic model generated for the same mammalian or bacterial cell expressing the same recombinant, heterologous polypeptide,

- 5 - determining that the cultivation is affected by a problem if
- (i) the modeled fit shows an offset with respect to the raw data of more than 10 %, or
- (ii) the modeled fit has a χ^2 value determined by a Pearson's chi-squared test of more than 5.

10 In another aspect, the present disclosure provides a method for selecting a cell expressing a heterologous polypeptide, wherein the method comprises the following steps:

a) separately cultivating a multitude of mammalian or bacterial cell clones that produce the same heterologous polypeptide, whereby during the cultivating temporal process data is recorded,

15

b) fitting the process data acquired in step a) of each clone individually using the same metabolic model, which had been generated for the same mammalian or bacterial cell,

c) determining that a cultivation of step a) is affected by a problem if the fit obtained in step b) for the process data of said cultivation obtained in step a) in said metabolic model (i) shows an offset with respect to the raw data of more than 10 %, or (ii) the χ^2 value determined by a Pearson's chi-squared test for the fit is 5 or more,

20

d) repeating steps a) to c) with the clones that had a problem in the cultivation as determined in step c)

25

or

if no clone had a problem in the cultivation as determined in step c) selecting the clone from the multitude of clones as cell expressing a heterologous polypeptide that has (i) the highest titer, and/or (ii) the highest level of intended product quality attribute(s), and/or (iii) the preferred metabolic phenotype/highest rang in metabolic performance indicators.

30