

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7092879号

(P7092879)

(45)発行日 令和4年6月28日(2022.6.28)

(24)登録日 令和4年6月20日(2022.6.20)

(51)国際特許分類

F I

G 1 6 B 5/00 (2019.01)

G 1 6 B 5/00

G 1 6 B 40/20 (2019.01)

G 1 6 B 40/20

G 0 1 N 33/50 (2006.01)

G 0 1 N 33/50

Z

C 1 2 M 1/00 (2006.01)

C 1 2 M 1/00

A

C 1 2 Q 1/02 (2006.01)

C 1 2 Q 1/02

請求項の数 20 (全74頁)

(21)出願番号 特願2020-534849(P2020-534849)

(86)(22)出願日 平成31年1月2日(2019.1.2)

(65)公表番号 特表2021-508872(P2021-508872
A)

(43)公表日 令和3年3月11日(2021.3.11)

(86)国際出願番号 PCT/EP2019/050006

(87)国際公開番号 WO2019/129891

(87)国際公開日 令和1年7月4日(2019.7.4)

審査請求日 令和2年8月20日(2020.8.20)

(31)優先権主張番号 17211217.9

(32)優先日 平成29年12月29日(2017.12.29)

(33)優先権主張国・地域又は機関
欧州特許庁(EP)

(73)特許権者 591003013

エフ・ホフマン-ラ ロシュ アーゲー
F. HOFFMANN-LA ROCH
E AKTIENGESELLSCHA
FTスイス・シーエイチ-4070バーゼル
・グレンツアーヘルストラツセ124

(74)代理人 100102978

弁理士 清水 初志

(74)代理人 100102118

弁理士 春名 雅夫

(74)代理人 100160923

弁理士 山口 裕孝

(74)代理人 100119507

弁理士 刑部 俊

最終頁に続く

(54)【発明の名称】 細胞培養物の代謝状態の予測

(57)【特許請求の範囲】

【請求項1】

以下の工程を含む、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するための方法：

- 該特定の細胞タイプの細胞の代謝モデル(402)を準備する工程(102)であって、該代謝モデルが、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックス(410)および細胞外フラックス(408)を含み、該代謝モデルが、該細胞内代謝産物(406)のうちの1つと該細胞外代謝産物(404)のうちの1つとの間の少なくとも1つの化学量論的関係を指定する化学量論式を含む、工程；

- 該細胞培養物の培養中の複数の時点のそれぞれにおける、

○該時点で測定された複数の測定値を受け取る工程(106)であって、該測定値が、該細胞培養物の培養培地における該代謝モデルの複数種の細胞外代謝産物の濃度と、該細胞培養物における該細胞の細胞密度測定値とを含む、工程；

○受け取った該測定値を、入力パラメータ値として、トレーニング済みの機械学習プログラムロジック(MLP)(218)に入力する工程(108)；

○受け取った該測定値を用いて、該MLPによって、将来の時点での細胞外代謝産物の細胞外フラックス(408)を予測する工程(110)であって、該将来の時点が、該測定値を受け取った時点より後の時点であり、該細胞外フラックスが、細胞内への細胞外代謝産物の取り込み速度および/または細胞から培地への細胞外代謝産物の放出速度である、工程；

○該細胞外代謝産物の該細胞外フラックスの予測値と該代謝モデルの該化学量論式とを用

いて、該将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程(112)。

【請求項2】

前記方法が、機械学習によりMLPを作成する工程をさらに含み、該作成する工程が、

- トレーニングデータセットを作成することであって、

○前記特定の細胞タイプの細胞の少なくとも1つのトレーニング細胞培養物であって、該トレーニングデータセットを作成するためにデータが収集される少なくとも1つのトレーニング細胞培養物の培養中の複数のトレーニング時点であって、トレーニングデータが収集される時点である複数のトレーニング時点のそれぞれにおいて、

■該トレーニング時点で測定された複数の測定値を受け取る工程であって、該測定値が、該少なくとも1つのトレーニング細胞培養物の培養培地における代謝モデルの複数種の細胞外代謝産物の濃度と、該少なくとも1つのトレーニング細胞培養物における細胞の細胞密度測定値を含む、工程、

■現在のトレーニング時点の時間表示を受け取る工程、および

■その時点で受け取った測定値と、それぞれの前の時点で受け取った測定値との関数として、該細胞外代謝産物の細胞外フラックスを計算する工程であって、該細胞外フラックスが、細胞内への該細胞外代謝産物の取り込み速度および/または培地への該細胞外代謝産物の放出速度である、工程

を含む、トレーニングデータセットを作成すること、

- 該MLPをトレーニングすることであって、

○該トレーニング時点のそれぞれにおいて受け取った測定値を、入力パラメータ値として該MLPに入力し、そのトレーニング時点の後続の各時点において該後続の時点について計算された細胞外フラックスを、該入力パラメータ値に関連する出力パラメータ値として該MLPに入力する工程；ならびに

○該MLPが、該入力パラメータ値に基づいて、それぞれの該関連する出力パラメータ値を予測するように学習するように、該MLPによる学習プロセスを実施する工程

を含む、該MLPをトレーニングすること、

- トレーニングされた該MLPを揮発性または非揮発性の記憶媒体に記憶すること

を含む、請求項1記載の方法。

【請求項3】

前記特定の細胞タイプの細胞の複数のトレーニング細胞培養物の培養中の前記複数のトレーニング時点のそれぞれにおいて、前記測定値および時間表示が受け取られ、かつ前記細胞外代謝産物の細胞外フラックスが計算されるように、前記トレーニングデータセットが作成され、ここで、該細胞培養物が、異なるタイプのバイオリアクターの中で培養され、該バイオリアクターのタイプが、フェドバッチバイオリアクター、バッチバイオリアクター、灌流リアクター、ケモスタット、およびスプリットバッチバイオリアクターを含むバイオリアクタータイプのうちの少なくとも2つの異なるバイオリアクタータイプを含む、請求項2記載の方法。

【請求項4】

前記MLPが、1つのニューラルネットワークであるか、またはいくつかのニューラルネットワークからなるシステムである、請求項1~3のいずれか一項記載の方法。

【請求項5】

前記複数種の細胞外代謝産物の濃度測定値がそれぞれ、

- 該代謝産物の体積当たりの含有量、特に、質量濃度もしくは物質濃度、または

- 該体積当たりの含有量と直線的もしくは問題となっている濃度範囲の少なくとも90%であるようにほぼ直線的に相関する値、特に、補正もしくは標準化された、該細胞外代謝産物の代謝産物濃度測定値もしくはフラックス測定値

である、請求項1~4のいずれか一項記載の方法。

【請求項6】

前記MLPが、いくつかのサブMLPからなるシステムであり、該システムに含まれる個々の

サブMLPがそれぞれ、単一の細胞外代謝産物の細胞外フラックスを予測するようにトレーニングされており、かつ、該単一の細胞外代謝産物の将来の時点での細胞外フラックスを予測するのに選択的に用いられる、請求項1～5のいずれか一項記載の方法。

【請求項7】

前記MLPが、前記細胞外代謝産物のうちの1つの細胞外フラックスを予測するために、複数種の細胞外代謝産物の濃度測定値を入力パラメータ値として使用し、ここで、細胞外フラックスを求めようとする該細胞外代謝産物のうちの少なくとも2つについて、入力パラメータ値として使用する該複数種の細胞外代謝産物が異なる、請求項1～6のいずれか一項記載の方法。

【請求項8】

- 全ての入力候補代謝産物の濃度を複数の時点にわたって測定する工程であって、該入力候補代謝産物が、前記特定のタイプの細胞培養物を有する参照バイオリアクター内で測定可能な程度に利用可能な全ての細胞外代謝産物を含むか、または前記代謝モデルの全ての細胞外代謝産物を含む、工程；

- 細胞外フラックスを予測しようとする単一の細胞外代謝産物のそれぞれについて、該単一の代謝産物の細胞外フラックスを予測するために入力パラメータ値として使用される複数種の細胞外代謝産物を特定するための選択プロセスを行う工程であって、該選択プロセスが、該単一の代謝産物それぞれに関して、

(a)細胞外代謝産物の第1のセットを定める工程であって、該第1のセットが全ての候補入力代謝産物を含む、工程；

(b)該第1のセットにおける各細胞外代謝産物の第1の関連性スコアを、その代謝産物の濃度測定値の関数として計算する工程であって、該第1の関連性スコアが、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの細胞外代謝産物の濃度の予測能力を示す、工程；

(c)該細胞外代謝産物のうち最も高い第1の関連性スコアを有するものだけを、該第1のセットから、まだ空である細胞外代謝産物の第2のセットに移し、この代謝産物を該第1のセットから除去する工程；

(d)該第1のセットにおける各細胞外代謝産物のさらなる関連性スコアを、その代謝産物の濃度測定値と、該第2のセットに含まれる全ての細胞外代謝産物の濃度測定値との関数として計算する工程であって、該さらなる関連性スコアが、既に該第2のセット含まれている代謝産物を考慮した、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの候補入力代謝産物の濃度の予測能力を示す、工程；

(e)該第1のセットの細胞外代謝産物のうち最も高いさらなる関連性スコアを有するものだけを該第2のセットに移し、この代謝産物を該第1のセットから除去する工程であって、この代謝産物を含めることによって、該単一の細胞外代謝産物の細胞外フラックスの予測に関する、第2のセットに含まれる代謝産物の情報冗長性の最大限度を第2のセットが超えない場合にだけ、この代謝産物が移される、工程；

(f)該第2のセットが該情報冗長性の最大限度を超えることなしでは、これ以上の代謝産物を該第1のセットから該第2のセットに移すことができなくなるまで、工程(d)および(e)を繰り返す工程；ならびに

(g)該単一の細胞外代謝産物の細胞外フラックスを予測するために、該第2のセットに移された代謝産物だけを選択的に入力パラメータ値として使用する工程

を含む、選択プロセスを行う工程

をさらに含む、請求項7記載の方法。

【請求項9】

- 前記第1のセットの代謝産物と、細胞外フラックスを予測しようとする前記単一の代謝産物との間の部分相互情報量スコア(PMISコア)として、前記第1の関連性スコアが計算される；

- 既に前記第2のセットに含まれている全ての代謝産物を考慮して、該第1のセットの代謝産物と、細胞外フラックスを予測しようとする該単一の代謝産物との間のPMISコアとし

10

20

30

40

50

て、第2の関連性スコアであって、該第1のセットにおける各細胞外代謝産物の前記さらなる関連性スコアである第2の関連性スコアが計算される、

請求項8記載の方法。

【請求項10】

前記MLPが、単一の代謝産物のそれぞれの細胞外フラックスを予測するために、複数種の細胞外代謝産物の濃度測定値を入力パラメータ値として使用し、該複数種の細胞外代謝産物が、少なくとも1種のアミノ酸、好ましくは少なくとも2種のアミノ酸を含む、請求項1～9のいずれか一項記載の方法。

【請求項11】

前記複数の時点が、10分～48時間、好ましくは1～24時間の時間間隔だけ隔てられている、請求項1～10のいずれか一項記載の方法。

10

【請求項12】

前記細胞タイプが、生体分子を得る目的でバイオリアクター内で維持されかつ/または増殖する遺伝子組換えされた細胞タイプである、請求項1～11のいずれか一項記載の方法。

【請求項13】

前記将来の時点のそれぞれにおいて、前記代謝モデルの細胞内フラックスのうちのいくつかまたは全ての計算が行われる、請求項1～12のいずれか一項記載の方法。

【請求項14】

- それぞれの参照値または参照値範囲から閾値を超えて逸脱している細胞内フラックス(410)計算値を有する、前記代謝モデルの1種または複数種の細胞内代謝産物を特定する工程；
- 細胞増殖または所望の生体分子の産生の制限要因として作用する細胞内フラックスを自動的に特定する工程

20

をさらに含む、請求項1～13のいずれか一項記載の方法。

【請求項15】

以下の工程を含む、特定の細胞タイプの細胞の細胞培養物を含むバイオリアクターをモニタリングおよび/または制御するための方法：

- 請求項1～14のいずれか一項記載の方法に従って将来の時点での該細胞の細胞内フラックスを計算する工程；
- 該細胞内フラックスの予測値を、それぞれの1種もしくは複数種の細胞内代謝産物の許容可能な細胞内フラックスの参照値もしくは参照値範囲と比較する工程；
- 少なくとも1つの細胞内代謝産物の細胞内フラックスの計算値の、そのそれぞれの参照値もしくは参照値範囲からの逸脱が限界値を超えた場合に、警告を発する工程；ならびに/または
- 該逸脱を小さくするように該バイオリアクターの状態もしくはその中に含まれる培地を変える工程を自動的に開始するための制御コマンドを該バイオリアクターに送る工程であって、該自動工程が、特に培養培地の量および/もしくは組成を変えることを含む、工程。

30

【請求項16】

- 請求項14に従って、細胞増殖もしくは所望の生体分子の産生の制限要因として作用する、細胞の代謝モデル内の反応を特定する工程；および
 - 細胞増殖もしくは該生体分子の産生または該生体分子の質が促進されるように、制限要因として作用する細胞内フラックスを改変する物質を選択的に自動的に追加するか、あるいは、このような追加の要求をユーザーインターフェイスを介して発行する、工程
- をさらに含む、請求項15記載の方法。

40

【請求項17】

以下の工程を含む、特定の細胞タイプの細胞の代謝的に好ましいクローンを特定するための方法：

- 複数のバイオリアクターの中で異なる細胞培養物を培養する工程であって、該異なる細胞培養物が、該特定の細胞タイプの細胞の異なるクローンを含む、工程；
- 各細胞クローンについて別々に複数の時点で、請求項1～16のいずれか一項に従って、1種または複数種の細胞内代謝産物の細胞内フラックスを計算する工程；

50

- 代謝的に最も好ましい、該1種または複数種の細胞内代謝産物の細胞内フラックス計算値を有する、該細胞クローンのうちの1つを特定する工程。

【請求項18】

- 1種または複数種の細胞外代謝産物の現在の細胞外フラックスを、現在の時点および以前の時点で測定された該細胞外代謝産物の濃度から計算する工程；
 - 該細胞外代謝産物の該現在の細胞外フラックスの計算値と代謝モデルの化学量論式とを用いて、さらなる代謝フラックス解析を行って、現在の時点での現在の細胞内フラックス(410)を計算する工程；ならびに
 - 該現在の細胞内フラックスの計算値を、細胞培養物の細胞の現在の代謝状態の特徴付けとして使用する工程
 をさらに含む、請求項1～17のいずれか一項記載の方法。

10

【請求項19】

- 前記細胞培養物の培養中の前記時点のそれぞれにおいて、該細胞培養物の培地中の乳酸デヒドロゲナーゼ(LDH)濃度の測定値がさらに受け取られ；
 - 前記将来の時点のそれぞれにおける前記細胞外代謝産物の細胞外フラックスの予測が、前記細胞密度測定値ではなく補正された細胞密度を用いて、前記MLPによって行われ、該補正された細胞密度の計算が、該時点のそれぞれにおける、
 ○該LDH濃度測定値の関数として、該細胞培養物の培地中の溶解細胞の密度を計算する工程であって、該関数が、該培地中のLDH濃度と前記特定の細胞タイプの溶解細胞の数との関係を表す、経験的に求められたヒューリスティックな一次関数である、工程；
 ○該培地中の該細胞密度測定値と該溶解細胞の密度計算値との和として、該補正された細胞密度を計算する工程
 を含む、
 請求項1～18のいずれか一項記載の方法。

20

【請求項20】

以下を備える、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するためのシステム(200)：

- 1つまたは複数のプロセッサ(202)、
 - 該細胞培養物を含むバイオリアクター(204、206、208)からの測定値を受け取るための第1のインターフェイス(210)、ならびに
 - 以下を備える、揮発性または非揮発性の記憶媒体(212)：
 ○該特定の細胞タイプの細胞の代謝モデル(402)であって、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックス(410)および細胞外フラックス(408)を含み、該細胞内代謝産物(406)のうちの1つと該細胞外代謝産物(404)のうちの1つとの間の少なくとも1つの化学量論的關係を指定する化学量論式を含む、代謝モデル(402)；
 ○トレーニング済みの機械学習プログラムロジック(MLP)(218)；
 ○該細胞培養物の培養中の複数の時点のそれぞれにおいて以下の工程を含む方法を行うように構成された、プログラムロジック(220)：
 ■第1のインターフェイスを介して該時点で測定された複数の測定値を受け取る工程(106)であって、該測定値が、該細胞培養物の培養培地における該代謝モデルの複数種の細胞外代謝産物の濃度と、該細胞培養物における該細胞の細胞密度測定値とを含む、工程；
 ■受け取った該測定値を入力パラメータ値として該MLPに入力する工程(108)；
 ■受け取った該測定値を用いて、該MLPによって、将来の時点での該細胞外代謝産物の細胞外フラックス(408)を予測する工程(110)であって、該将来の時点が、該測定値を受け取った時点より後の時点であり、該細胞外フラックスが、細胞内への該細胞外代謝産物の取り込み速度および/または細胞から培地への該細胞外代謝産物の放出速度である、工程；
 ■該細胞外フラックス予測値と該代謝モデルの該化学量論式とを用いて、該将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程(112)。

30

40

【発明の詳細な説明】

【技術分野】

50

【0001】

本発明は、細胞の代謝状態の予測、特に、細胞培養物中に維持されている細胞の代謝状態の予測に関する。

【背景技術】

【0002】

近年、製薬産業は、開発、生産、および品質保証におけるプロセスの効率性をもっと高めるために、プロセス解析、モニタリング、および制御の手段に強く焦点を合わせることで膨大な労力を費やしてきた。この風潮は、薬学的に関連する分子、特に、タンパク質などの巨大分子を産生するのに用いられる細胞培養リアクターの動作にも、ある特定の程度まで影響を及ぼす。

10

【0003】

しかしながら、薬学研究・開発という面での細胞代謝および発酵プロセスのモデリングは大きな技術的難題であることが判明している。すなわち、細胞（特に、真核細胞）の代謝は、解析によりシミュレートまたはモデル化するのが難しい非常に複雑な非線形的化学反応カスケードを特徴とする。

【0004】

例えば、代謝フラックス解析(MFA)を用いた代謝フラックスの解析がよく知られている。MFAは、特に、細胞内フラックス分布が時間によって変化しないプロセスに用いられる。これは、バッチプロセス（「バッチバイオリアクター」）での指数増殖期に、またはケモスタットでの培養に近似的になされる。しかしながら、現在、広く用いられているフェドバッチプロセスでは、細胞は、常に変化する環境条件に曝露される。従って、プロセス中に細胞内フラックスは変化する。従って、プロセス全体を通して細胞の状態を調査するためには、1回の代謝フロー解析を実施するだけでは十分でない。さらに、バイオリアクターの細胞が代謝的に好ましくない状態に既に入っていることが記述的にしか確かめられなければ、これらの標的プロセス制御に関して好ましくない。この状態を予測し、必要であれば初期段階で対策を取る方が良好だろう。代謝フラックスを解析するためのMFAの使用は、Ahn WS, Antoniewics MR (2012): 「Towards dynamic metabolic flux analysis in CHO cell cultures」, Biotechnology Journal 7, 61-74（非特許文献1）に記載されている。

20

【0005】

しかしながら、将来の時点の細胞培養物における細胞の状態を予測しようという試みは、いくつかの理由でも非常に難しい。通常、時間の経過を予測するために反応速度モデルが用いられる。これらは、物質濃度または物質量の経時変化を表した連立微分方程式からなる。通常、これらの式はメカニズムの知識に基づいており、例えば、ミカエリスメンテン速度式を用いてモデル化することができる。しかしながら、この反応速度の知識は入手することが難しい。速度式を用いて細胞内代謝ネットワーク全体を表すために、利用可能なメカニズムの情報は通常、十分でなく、推定されるパラメータの数は膨大である。

30

【0006】

しかしながら、代謝フラックス解析と反応速度の知識を組み合わせた反応速度ハイブリッドモデルは、実際には大部分が、細胞培養物の将来の挙動の信頼性の高い予測を行うことができなかった。これらのモデルの作成は極めて労働集約的であることが判明している。さらに、このように作成されたモデルはあまり融通性がなく、相当の手作業なしでは他の細胞タイプの代謝に合わせることができない。ハイブリッドモデルの使用は、例えば、Covert MW, Xiao N, Chen TJ, Karr JR (2008), 「Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli」, Bioinformatics Vol. 24 no. 18, 2044（非特許文献2）、およびNolan RP, Lee K (2011) 「Dynamic model of CHO cell metabolism」, Metabolic Engineering 13, 108（非特許文献3）に記載されている。

40

【先行技術文献】

【非特許文献】

50

【0007】

【文献】Ahn WS, Antoniewics MR (2012): 「Towards dynamic metabolic flux analysis in CHO cell cultures」, Biotechnology Journal 7, 61-74
Covert MW, Xiao N, Chen TJ, Karr JR (2008), 「Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli」, Bioinformatics Vol. 24 no. 18, 2044
Nolan RP, Lee K (2011) 「Dynamic model of CHO cell metabolism」, Metabolic Engineering 13, 108

【発明の概要】

【0008】

概要

これに関連して、上記の欠点が少なくとも部分的に回避される程度まで、細胞の代謝状態を予測するための改善された方法と、それに対応して改善されたシステムが必要とされている。

【0009】

本発明の主題は独立クレームにおいて述べられる。本発明の態様は従属クレームにおいて説明される。本明細書において説明される本発明の態様および実施例は相互排他的でなければ互いに自由に組み合わせられ得る。

【0010】

一局面において、本発明は、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するための方法に関する。当該方法は、特定の細胞タイプの細胞の代謝モデルを準備する工程を含み、該代謝モデルは、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックスおよび細胞外フラックスを含み、該細胞内代謝産物のうちの1つと該細胞外代謝産物のうちの1つとの間の少なくとも1つの化学量論的關係を指定する化学量論式を含む。

【0011】

さらに、本発明は、細胞培養物の培養中の複数の時点のそれぞれにおいて、以下の工程を実施することを含む：

- その時点で測定された複数の測定値を受け取る工程であって、測定値が、細胞培養物の培養培地における代謝モデルの複数種の細胞外代謝産物の濃度と、細胞培養物における細胞の細胞密度測定値を含む、工程；
- 受け取った測定値を、入力パラメータ値として、トレーニング済み機械学習プログラムロジック(MLP)に入力する工程；
- 受け取った測定値を用いて、MLPによって、将来の時点での細胞外代謝産物の細胞外フラックスを予測する工程であって、該将来の時点が、測定値を受け取った時点より後の時点であり、該細胞外フラックスが、細胞内への細胞外代謝産物の取り込み速度および/または細胞から培地への細胞外代謝産物の放出速度である、工程；
- 細胞外代謝産物の細胞外フラックス予測値と代謝モデルの化学量論式とを用いて、将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程。

【0012】

このことは有利であるかもしれない。なぜなら、このように、この知識は、例えば、化学量論反応式の形をとって、細胞の多くの重要な代謝プロセスについて文献中で既に利用可能であり、かつ代謝モデルの形で表すことができ、非常に有利なやり方でMLPの使用と組み合わせることができ、そのため高い融通性が得られるからである。いくつかの予備実験研究から、細胞の将来の代謝状態を、例えば、排出産物の濃度に純粋に基づいて予測するためのMLPに基づくアプローチの使用、例えば、ニューラルネットワークの使用は信頼できないことが分かっている。細胞代謝の複雑さと動態のために、洗練された現代のMLP方法でも、将来の細胞培養物の代謝状態を（特に、真核細胞を含むフェドバッチバイオリアクターの中での）現在の細胞濃度に純粋に基づいて信頼性をもって予測することは不可能なように思われる。一方で、このことは、内部の細胞状態が細胞外代謝産物の濃度測定値

10

20

30

40

50

によって非常に間接的かつ不十分にしか特徴付けられないということに起因するかもしれない。他方で、トレーニング期にMLPを作成している間にトレーニングデータに「オーバーフィッティング」するリスクもある。しかしながら、MLPに基づく方法はトレーニング形式について本明細書で説明されるように、少なくとも、あまり遠くない将来である時点の細胞外代謝産物の濃度とフラックスに関して非常に信頼性の高い予測を提供し得ることが示されている。この知識を、細胞内代謝産物と細胞外代謝産物との間の、および相互間の化学量論的依存性および輸送依存性の知識を組み合わせることによって、本発明の態様は細胞の代謝状態を個々の細胞内フラックスのレベルまで正確に予測することができる。

【0013】

細胞が排出した物質（例えば、乳酸）に基づいて細胞の代謝状態を大まかに特徴付けることは、これまで可能であったが、結局、細胞の内部プロセスは「ブラックボックス」のままである。現在知られているハイブリッドモデルでは、細胞代謝を個々の細胞内物質フラックスのレベルまで、きめ細かく予測することは不可能であった。

【0014】

本発明の別の有利な局面では、細胞代謝を予測するための融通性の高い方法が提供される。特定の細胞タイプの細胞を含む、いくつかの細胞培養物を培養することでも、その細胞タイプの細胞培養物について、将来の時点の代謝状態を信頼性をもって予測することができるMLPをトレーニングするために十分に大きなトレーニングデータセットが取得され得ることが分かっている。トレーニングデータの取得と、その後のMLPトレーニングは多くは自動化される可能性がある。対照的に、代謝フラックス解析に用いられる代謝モデルは広範囲にわたる文献研究と手作業による適合工程を必要とする場合があり、多くの場合、異なる細胞タイプの細胞に再使用され得る。従って、当技術分野において公知のハイブリッドモデルの使用とは対照的に、複雑な新モデル作成を必要とすることなく、単に、いくつかの細胞培養物を培養してトレーニングデータセットを作成し、次いで、このトレーニングデータセットに対してMLPをトレーニングすることによって、特定の細胞タイプの特定の代謝特徴を考慮に入れた予測方法を提供できることが多い。

【0015】

別の有利な局面では、MFAは、ここでは、細胞内フラックスを予測するのに首尾良く用いられる。これまで、MFAは主に、細胞内フラックス分布が時間によって変化しないプロセスに用いられてきた。これはバッチプロセスでの指数増殖期、またはケモスタットでの培養に対して近似的になされる。しかしながら、現在優勢に用いられている、細胞が保持されるフェドバッチプロセスおよび灌流に似たプロセスでは、細胞は、常に変化する環境条件に曝露される。従って、細胞内フラックスもプロセス中に変化する。プロセス全体を通して細胞の状態を調べようと試みてきたアプローチは失敗した。対照的に、（おそらく離散化された）プロセス時間の関数として細胞内フラックスを記述する動的代謝モデルを規定することによって、フェドバッチリアクターの代謝状態を少なくともこの先何時間および何日か正確に予測するアプローチが発見された。

【0016】

さらに有利な局面では、本発明に従う方法の態様を用いると、細胞培養物の培養中に予測の質と、この予測の間に得られる中間結果の質を管理することが可能になる。前記の方法は細胞外フラックスの予測を含む。これらは、培養培地試料のサンプリングと解析を繰り返すことによって発酵中に細胞外代謝産物の濃度変化から容易に求めることができる。従って、MLPの予測は、発酵中に、細胞外フラックスの予測値と測定値を繰り返し比較することによってチェックされてもよい。有意差が発生すれば計算はすぐに止められ、逸脱の原因が調査されてもよい。他の最先端の代謝フラックス予測方法では、細胞内マテリアルフローと細胞外マテリアルフローを予測するために定常状態での ^{13}C 測定データを使用する。従って、これらの方法は、細胞内フラックスおよび細胞外フラックスを直接予測するために ^{13}C 標識解析と定常状態条件下での実験を必要とする。 ^{13}C 解析は複雑であり、ファーマンターの動作中に定常状態とならない場合がある。さらに、細胞内フラックス予測値は経験的に検証することができず、そのため、このような手順には、測定された実デー

10

20

30

40

50

タと比較され得る中間工程が無い。

【0017】

本発明の態様によれば、MLPは、動的発酵条件下で将来の時点（次のサンプリング時間）での対応する細胞外フラックスを予測するために現在の時点での細胞外代謝産物の細胞外濃度の形で入力データを受け取る。細胞間MFAは、（将来のサンプリング時間、任意で、現在のサンプリング時間も）化学量論代謝モデルに基づいて計算される。これにより、細胞外代謝産物濃度測定値によって発酵中の予測を検証することが可能になる。

【0018】

従って、本発明の態様は、（細胞内代謝プロセスについて詳細な情報を提供し、細胞外フラックスに関する予測が、実際に測定された細胞外フラックスと非常によく対応することが実験により示されているので）正確であり、かつ、別の細胞タイプの細胞に簡単に適合され得るので、とても融通性のある、細胞培養物の代謝状態を予測するための方法を提供する。

10

【0019】

態様によれば、代謝モデルの細胞外フラックスおよび細胞内フラックスの少なくとも一部は、既知の代謝ネットワークをエレメンタリーフラックスモードに分解することに基づいていない。文献では、単純な細胞内ネットワークフラックスがMLPを用いて既に推定されている。最初に、ネットワークのエレメンタリーフラックスモードが作成された。エレメンタリーフラックスモードは、可能性のある他の全てのフラックス分布が異なる重み付けによって組み合わせられ得る一組の許容可能なフラックス分布を表している。MLPは、対応する重み付けを出力する。しかしながら、これらのエレメンタリーフラックスモードの数はほとんどの生化学的ネットワークにはかなり多いので、重み付けの有効かつロバストな推定は不可能である。

20

【0020】

しかしながら、本発明によれば、ニューラルネットワークまたは他のMLPを介して細胞外フラックスだけが推定される。次いで、代謝マテリアルフロー解析に結び付けることで細胞内フラックス分布との関係が確かめられる。

【0021】

従って、本発明の態様によれば、機械学習プログラムロジックは、細胞外フラックスだけを選択的に予測するが、細胞内フラックスを選択的に予測しないように訓練されている。

30

【0022】

このことは有利であるかもしれない。なぜなら、予測は、バイオリアクターを動作させている間でも簡単に測定できる値に制限され、その結果、細胞外フラックスは細胞内フラックスを予測するための中間値を表し、MLPの誤差を素早く特定するために真の測定データと簡単に比較できるからである。

【0023】

別の局面では、前記方法は以下の理由から有利であるかもしれない。すなわち、細胞の代謝はバイオリアクター内の状態に大きく左右される。pH値、pO₂値、圧力、および温度などの要因が一定に保たれていれば、反応培地中の代謝産物濃度は細胞の挙動に特に重要である。こういうわけで、これらをMLPの入力として選択した。将来の代謝産物フラックスの予測には、次のサンプリング時間での将来の代謝産物濃度の予測と比べて2つの大きな利点がある。

40

【0024】

一方では、このようにトレーニングされたMLPは時間間隔の選択に関して融通性が高い。代謝産物濃度を含むネットワークが出力としてトレーニングされた場合、このようなトレーニング済みMLPを用いて行われ得る予測は、常に、トレーニングデータセットと同じ時間間隔を指している。従って、既にトレーニングデータを作成している場合には、できるだけ均一に間隔が選択されることを確実にしなければならない。なぜなら、そうしないと不一致が生じることがあるからである。他方では、フローが予測される場合、これは、MLPトレーニング中に使用した時間間隔からの、現在の予測のための時間間隔の、ある一定

50

の独立をもたらす。従って、前記方法は、予測のための時間間隔の選択に関してロバストであり、かつ融通性が高い。

【0025】

他方では、代謝産物濃度は、場合によっては可変的に取り扱われることがある連続した、および/またはパルス様の栄養分投与に左右される。しかしながら、トレーニング済みMLPが、トレーニングデータセットにおいて別々に実施された追加用量を取り扱わず、やみくもに学習するのであれば、この可変性は予測に存在しない。細胞外代謝産物フラックスを使用することによって、将来の濃度は付属書(Appendix)の式4.4を用いて外挿され得る。フラックスがサンプリング間隔に対して強力な変動を受けなければ、時間間隔に関して、いくつかの融通性がある。好ましい態様によれば、細胞外フラックスの予測に使用した時点間の時間間隔は、トレーニングデータセットの作成に使用した時間間隔より短い、またはそれに等しいか、または最大で20%長い。

10

【0026】

態様によれば、細胞外フラックス計算値は、物質添加(例えば、グルコース添加)に応じて調整されるように計算される。従って、1種または複数種の細胞外フラックスを予測する場合には、特定の細胞外代謝産物が次の時間間隔で添加されるという情報が、それに応じて、この細胞外代謝産物の濃度の予測を調整するのに用いられ得る。この細胞外代謝産物の欠乏が予想されることをMLPが予測するのであれば、予測に基づいて適宜、供給を調整することもできる。しかしながら、追加用量はリアクター内の濃度を変えるので、一般的に平均フラックスに影響を及ぼし得ることに留意しなければならない。投与プロファイルがトレーニングデータから有意に逸脱すれば、不正確さが生じることがある。

20

【0027】

本発明の態様によれば、代謝状態を予測しようとする細胞培養物に代謝産物を供給する時間プロファイルおよび定量プロファイルは、1つまたは複数のトレーニングバイオリアクターを用いてトレーニングデータを作成するのに使用した供給プロファイルと同一になるか、または類似するように選択され、MLPは、これらのトレーニングデータに対してトレーニングされている。

【0028】

本発明の態様によれば、測定データの取得ならびに細胞外フラックスの予測のための測定がリアルタイムで、すなわち、細胞培養物を含有するバイオリアクターを動作させている間に行われる。測定データの収集と細胞代謝状態の予測との間の時間は典型的には短く、数秒または数分の範囲内、典型的には15分未満であるのに対して、個々の予測のための時間間隔は典型的には1~48時間、特に6~24時間の範囲内である。

30

【0029】

従って、さらなる有利な局面では、本発明の態様はリアルタイムで細胞培養物の代謝状態を予測することができる。なぜなら、予測が基盤とする情報ベースは、既に定義された代謝モデル、既に存在しているトレーニング済みMLP、およびリアルタイムで簡単に収集され得る測定データからなるからである。例えば、培養培地から定期的に試料を採取し、この試料中の細胞数と代謝産物濃度を求めることによって、予測を実施するのに必要な測定データが取得され得る。

40

【0030】

本発明の態様によれば、細胞培養物の代謝状態を予測するための方法は、将来の時点で、細胞培養物を含むバイオリアクターの動作中にリアルタイムで連続して行われる。

【0031】

態様によれば、前記方法は機械学習によるMLP作成をさらに含む。

【0032】

MLPの作成は、トレーニングデータセットを作成する工程を含み、トレーニングデータセットの作成は、特定の細胞タイプの細胞の少なくとも1つのトレーニング細胞培養物の培養中の複数のトレーニング時点のそれぞれにおいて、以下の動作を行うことを含む：

・ 前記トレーニング時点で測定された複数の測定値を受け取る工程であって、前記測定値

50

が、前記少なくとも1つのトレーニング細胞培養物の培養培地における代謝モデルの複数種の細胞外代謝産物の濃度と、前記少なくとも1つのトレーニング細胞培養物における細胞の細胞密度測定値を含む、工程；

- 現在のトレーニング時点の時間表示を受け取る工程；ならびに
- その時点で受け取った測定値と、それぞれの前の時点で受け取った測定値との関数として細胞外代謝産物の細胞外フラックスを計算する工程であって、細胞外フラックスが、細胞内への細胞外代謝産物の取り込み速度および/または培地への細胞外代謝産物の放出速度である、工程；
- MLPをトレーニングすることであって、
 - トレーニング時点のそれぞれにおいて受け取った測定値を入力パラメータ値としてMLPに入力する工程、および、そのトレーニング時点の後続の各時点において該後続の時点について計算した細胞外フラックスを、その入力パラメータ値に関連する出力パラメータ値としてMLPに入力する工程；ならびに
 - MLPが、入力パラメータ値に基づいて、それぞれの関連する出力パラメータを予測するように学習するように、MLPによる学習プロセスを実施する工程を含む、MLPをトレーニングすること；
- トレーニングされた該MLPを揮発性または非揮発性の記憶媒体に記憶すること。

【0033】

本発明の態様に従うMLPの使用および作成は有利であるかもしれない。なぜなら、一方では、適切な機械学習アルゴリズムを選択した場合には、高精度の予測を成し遂げることができ、他方では、他の細胞タイプにおける代謝状態への適合を非常に簡単に、かつ多大な手作業による労力も文献研究もなく行うことができるからである。上記の測定データとこれから計算した細胞外フラックスの連続した収集と記録を伴う、いくつかのトレーニング細胞培養物の動作は、トレーニング細胞培養物中で使用した細胞の細胞タイプ専用、どのMLPがトレーニングおよび作成される得るかに基づいてトレーニングデータセットを提供するのに十分である。

【0034】

本発明の態様によれば、特定の細胞タイプの細胞の複数のトレーニング細胞培養物の培養中の複数のトレーニング時点のそれぞれにおいて測定値と時間表示が受け取られ、細胞外代謝産物の細胞外フラックスが計算されるようにトレーニングデータセットは作成される。この目的のために、細胞培養物は、好ましくは、異なるタイプのバイオリアクターの中で培養される。好ましくは、これらのバイオリアクタータイプは、以下のセット:フェドバッチバイオリアクター、バッチバイオリアクター、灌流リアクター（細胞が保持される変形を含む）、ケモスタット、およびスプリットバッチバイオリアクターから少なくとも2つの異なるタイプのバイオリアクターを含む。

【0035】

トレーニングデータセットの作成において異なるタイプのバイオリアクターを使用することは有利であるかもしれない。なぜなら、さらに広範なデータ基礎が作成され、トレーニング中のMLPの「オーバーフィッティング」が回避または低減できるからである。さらに、これにより、多くの異なるタイプのバイオリアクターの中での細胞培養物の将来の代謝状態を首尾良く予測するために同じMLPを使用することが可能になる。好ましくは、トレーニングデータセットは、異なるバイオリアクタータイプの中で培養されたトレーニング細胞培養物に基づいて収集され、バイオリアクタータイプは少なくとも1つのフェドバッチバイオリアクターおよび/または少なくとも1つの灌流リアクターを含む。このことは有利であるかもしれない。なぜなら、これらのリアクタータイプは、実際に、ますます頻繁に用いられており、これらのリアクタータイプの動態が大きいため、これらのリアクタータイプの中にある細胞培養物の代謝状態の表示は、これまで特に難しかったからである。

【0036】

好ましくは、将来の代謝状態を予測しようとする細胞培養物は、トレーニングデータセットを作成するのにも使用したタイプのバイオリアクターの中で培養される。

【0037】

態様によれば、特定の細胞タイプの細胞のいくつかのトレーニング細胞培養物の培養中の複数のトレーニング時点のそれぞれにおいて測定値と時間表示が受け取られ、細胞外代謝産物の細胞外フラックスが計算されるようにトレーニングデータセットは作成され、細胞培養物は同じタイプまたは異なるタイプのバイオリアクターの中で培養され、全てのバイオリアクターがバッチバイオリアクタータイプに属さない。例えば、全てのバイオリアクターがフェドバッチタイプのバイオリアクターでもよい。

【0038】

動作中に追加の培養培地が連続添加またはパルス添加されるために、フェドバッチバイオリアクター内にある細胞の将来の代謝状態の予測は、かなり技術的な難題であることが判明している。本発明の態様は、フェドバッチバイオリアクターの使用という面に関して特に有利である。なぜなら、このタイプのリアクターの中で培養される細胞培養物の代謝が複雑であることにもかかわらず、本発明の態様に従う予測は正確なことが示されているからである。

【0039】

一部の態様によれば、MLPはサポートベクトルマシンまたはいくつかのサポートベクトルマシンのシステムである。

【0040】

他の態様によれば、MLPは、1つのニューラルネットワークであるか、またはいくつかのニューラルネットワーク(NN)からなるシステムである。

【0041】

いくつかの初期試験から、サポートベクトルマシンとニューラルネットワークに加えて他のMLP方法が用いられ得ることが示唆される。しかしながら、ニューラルネットワークを使用した時に特に良好な予測結果が得られており、トレーニング期ならびに適用期にニューラルネットワークの簡単な取り扱いを可能にする、様々なネットワークアーキテクチャの広範なソフトウェアソリューションが既に利用可能である。

【0042】

本発明の態様によれば、MLPは、いくつかのサブMLP（特に個々のNN）からなるシステムであり、このシステムに含まれる個々のサブMLPはそれぞれ、個々の細胞外代謝産物の細胞外フラックスを予測するようにトレーニングされたことがあり、将来の時点での個々の細胞外代謝産物の細胞外フラックスを予測するために選択的に使用される。

【0043】

このことは有利であるかもしれない。なぜなら、細胞外代謝産物の濃度測定値の予測能力は他の細胞外代謝産物の個々の細胞外フラックスに関して異なることが示されているからである。予測の質は、それぞれが、出力パラメータ値である細胞外代謝産物の細胞外フラックスに関して特定の入力パラメータ値セットに基づいている、個々のサブMLP（例えば、個々のニューラルネットワーク）をトレーニングすることによって改善され得る。1つまたは複数種の細胞外代謝産物の細胞外フラックスが全体の結果として戻されるように、個々のサブMLPの結果は、より高度のMLPまたは他のプログラムロジックによって結び付けられ得る。「サブMLP」は、この「サブMLP」の出力と、さらなる「サブMLP」の出力がより高度のプログラムロジック（特に、さらなる、より高度のMLP）によって組み合わされて（例えば、集められて）全体の結果が形成されるように、1つまたは複数のさらなる「サブMLP」と機能的に結び付けられるMLPである。

【0044】

本発明の態様によれば、MLPは、細胞外代謝産物のうちの1つの細胞外フラックスを予測するために、数種の細胞外代謝産物の濃度測定値を入力パラメータ値として使用する。この点に関して、細胞外フラックスを求めようとする細胞外代謝産物のうちの少なくとも2つについて、入力パラメータ値として使用した複数種の細胞外代謝産物が異なる。

【0045】

これは予測の精度をさらに高める可能性があるので有利かもしれない。

10

20

30

40

50

【0046】

本発明の態様によれば、前記方法は、全ての入力候補代謝産物の濃度をいくつかの時点にわたって測定する工程をさらに含む。特に、この測定は代謝濃度プロファイルを経時的に確立するのに役立ち得る。入力候補代謝産物のセットは、特定のタイプの細胞培養物を有する参照バイオリアクター内で度量衡学的に利用可能な全ての細胞外代謝産物または代謝モデルの全ての細胞外代謝産物を含む。例えば、参照リアクターは、1つまたは複数のトレーニングバイオリアクター、すなわち、MLP作成用のトレーニングデータを収集するのに使用したリアクターでもよい。または、参照リアクターはまた、現在モニタリングされているバイオリアクターと同じタイプの細胞培養物を有する別のリアクターでもよい。従って、モデルの個々の細胞外代謝産物の細胞外フラックスの予測に関してMLP用の入力パラメータ値として役立つ細胞外代謝産物の決定は好ましくはMLPトレーニングの間または前に行われる。

10

【0047】

細胞外フラックスを予測しようとする単一の細胞外代謝産物のそれぞれについて、該単一の代謝産物の細胞外フラックスを予測するために入力パラメータ値として用いられる複数種の細胞外代謝産物を特定するための選択プロセスが行われる。選択プロセスは、該単一の代謝産物に関して、それぞれの場合で、

(a)細胞外代謝産物の第1のセットを定める工程であって、第1のセットが全ての候補入力代謝産物を含む、工程；

(b)第1のセットにおける各細胞外代謝産物の第1の関連性スコアを、その代謝産物の濃度測定値の関数として計算する工程であって、第1の関連性スコアが、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの細胞外代謝産物の濃度の予測能力を示す、工程；

20

(c)細胞外代謝産物のうち最も高い第1の関連性スコアを有するものだけを、第1のセットから、まだ空である細胞外代謝産物の第2のセットに移し、この代謝産物を第1のセットから除去する工程；

(d)第1のセットにおける各細胞外代謝産物のさらなる関連性スコアを、その代謝産物濃度の測定値と、第2のセットに含まれる全ての細胞外代謝産物濃度の測定値との関数として計算する工程であって、さらなる関連性スコアが、既に第2のセット含まれている代謝産物を考慮した、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの候補入力代謝産物の濃度の予測能力を示す、工程；

30

(e)第1のセットの細胞外代謝産物のうち最も高いさらなる関連性スコアを有するものだけを第2のセットに移し、この代謝産物を第1のセットから除去する工程であって、この代謝産物を含めることによって、該単一の細胞外代謝産物の細胞外フラックスの予測に関する、第2のセットに含まれる代謝産物の情報冗長性の最大限度を第2のセットが超えない場合にだけ、この代謝産物に移される、工程；

(f)第2のセットが情報冗長性の最大限度を超えることなしでは、これ以上の代謝産物を第1のセットから第2のセットに移すことができなくなるまで、工程d)およびe)を繰り返す工程；ならびに

(g)該単一の細胞外代謝産物の細胞外フラックスを予測するために、第2のセットに移された代謝産物だけを選択的に入力パラメータ値として使用する工程を含む。

40

【0048】

態様によれば、第1の関連性スコアは、第1のセットの代謝産物と、細胞外フラックスを予測しようとする単一の代謝産物との間の部分相互情報量スコア(PMISコア)として計算される。第2の関連性スコアは、既に第2のセット含まれている全ての代謝産物を考慮して、第1のセットの代謝産物と、細胞外フラックスを予測しようとする単一の代謝産物との間のPMISコアとして計算される。

【0049】

このことは有利であるかもしれない。なぜなら、各細胞外代謝産物またはそのフラックス

50

について特に高い予測能力を有する入力パラメータ値を決定および使用できる可能性があるからである。トレーニング中のオーバーフィッティング、従って、不十分な予測の質が回避される得る。

【0050】

細胞外代謝産物の特定の細胞外フラックスについて最も高い有意性(「関連性」または「予測関連性」)を有する代謝産物を特定することによって、選択された入力パラメータ値が、優れた予測能力をもつMLPの作成を可能にすることが保証される。予測関連性は、好ましくは、特定の細胞外代謝産物の代謝産物濃度プロファイル測定値と、細胞外フラックスがMLPの出力パラメータ値として役立つ(すなわち、細胞外フラックスを予測しようとする)代謝産物の代謝産物濃度プロファイルとの相関の程度を求めることによって確かめられる。予測関連性は、本発明の態様について下記で説明するように、様々な方法、例えば、主成分分析またはPMI(「部分的相互情報量」)によって確かめられ得る。この代謝産物の濃度プロファイルと強く相関する(この代謝産物の高度の情報冗長性を意味し、この代謝産物は既に第2のセット含まれていることを意味する)濃度プロファイルを有する代謝産物を、第2のセットがまだ含んでいない場合にだけ、代謝産物が第2のセットに含まれるということは、冗長な情報と強く相関し、従って、冗長な情報を第2のセットに導入する濃度プロファイルを有する2種以上の代謝産物のグループを第2のセットも含むことを防止する。第2のセットに情報冗長代謝産物が多くあるとオーバーフィッティング効果が生じるだろう。

10

【0051】

例えば、さらなる代謝産物を第1のセットから第2のセットに含めるために、通常、代謝産物の一部しか第1のセットから第2のセットに送られないように情報冗長性の最大値の到達が定義されてもよく、別の終了基準が定義されてもよい。

20

【0052】

従って、本発明の態様によれば、入力パラメータ(濃度が測定され、MLPに入力される細胞外代謝産物)の選択はMLPから完全に独立していてもよい。

【0053】

しかしながら、選択プロセスは、付属書の4.7.3章に記載のように「ラッパー」の形で、例えば、ニューラルネットワークによって提供される機能として行われる可能性もある。

【0054】

態様によれば、この単一の細胞外代謝産物に関して第1のセットにある各代謝産物の第1の関連性スコアは、この代謝産物と、それぞれの場合で細胞外フラックスを予測しようとする単一の代謝産物との間のPMIスコアとして計算される。

30

【0055】

第1のセットに残っている各代謝産物のさらなる関連性スコアを計算するために、さらなる、反復して実行される工程では、PMIスコア(より単純な相互情報量(MI)とは対照的に)も使われることがある。さらなるPMIスコアは、細胞外代謝産物に対する、この代謝産物の予測関連性を示し、この関連性は、既に第2のセット含まれている全ての代謝産物に対する、第1のセットから試験しようとする代謝産物の情報冗長性を含む。従って、代謝産物が最終的に第2のセットに含められる前に、第2のセットに既にある代謝産物を考慮して、既に第2のセット含まれている代謝産物より多く、かつそれ以上に「予測付加値」を有するかどうかチェックされる。この代謝産物の濃度プロファイル測定値が、第2のセットに既に存在する代謝産物と強く相関するのであれば、これは否定される。この場合、第2のセットには含められない。

40

【0056】

試験データセットは、入力パラメータの選択の質を推定し、MLPアーキテクチャ(例えば、ニューラルネットワークの層の数)を最適化するのに使用され得る。

【0057】

従って、PMIによる出力パラメータ代謝産物に対する代謝産物の予測関連性が確かめられたら、可能性のある入力変数と、出力変数に対する可能性のある、それぞれの入力変数間

50

の依存関係の測定または推定が可能になる。依存関係が強ければ強いほど、入力変数に基づいて予測され得る出力変数が良くなり、入力値代謝産物の関連性スコアが高くなり得る。これにより、第1のセットの代謝産物の関連性スコアの計算と、スコアに基づいたソーティングが可能になる。

【0058】

代謝産物が第2のセットに含められるかどうかというPMIに基づく決定では、PMIは、既に第2のセットにある各代謝産物への、この代謝産物の依存性を確かめるために使用される。冗長性を回避するために、第2のセットに既にある代謝産物と比較して十分な関連性のある新情報を含む代謝産物だけが第2のセットに含められる。

【0059】

従って、第2のセットにおいて次の関連性のある代謝産物を選択する時に、第2のセットにおいて既に選択されている代謝産物も考慮される。

【0060】

PMI（「部分的相互情報量」）の形での細胞外代謝産物間の依存関係の計算は、例えば、第1の量の代謝産物を第2の量に含めるべきかどうか決断を下す間に、PMI計算値とPMI基準、例えば、依然として許容可能な程度の依存性のPMI限界値を比較するために行われ得る。しかしながら、PMI基準との比較は、通常、限界値との単純な比較からならず、統計検定または別の選択プロセス、例えば、付属書、例えば、付属書の式2.22~2.27および4.7.3章に説明されるような「ラッパー」からなる。付属書に記載の態様によれば、「最も関連性のある」、すなわち、最も高いPMI値を有するものが常に第1のセットに送られる。このように、全ての代謝産物の順序が作成される。しかしながら、最終的には、全ての代謝産物は第1のセットにある。後の工程(ラッパー)では、次に、最も関連性のある代謝産物のうちいくつか、予測に用いられる代謝産物の「第2のリスト」に含められるかが決定される。

【0061】

態様によれば、MLPでは、細胞外代謝産物のそれぞれの細胞外フラックスを予測するために入力パラメータ値として数種の細胞外代謝産物の濃度測定値を使用し、数種の細胞外代謝産物は少なくとも1種の、好ましくは少なくとも2種のアミノ酸を含む。その濃度が入力パラメータ値として使用される代謝産物は、フラックスが出力パラメータ値として使用される代謝産物を含むことが多いが、必ずそうなるとは限らない。

【0062】

MLPまたはMLPトレーニングのために入力パラメータ値として培地中のアミノ酸濃度を使用することは有利であるかもしれない。なぜなら、濃度を求める確立した方法が既に存在し、バイオリクター内で多くの標的タンパク質を効率的に合成するにはアミノ酸が十分に存在することが必要なことが多いからである。

【0063】

既にトレーニング中に、MLPをトレーニングするために、この代謝産物サブセットの濃度だけが選択的に用いられた可能性がある。

【0064】

関連性に従って（例えば、PMIに従って）第1のセットの入力パラメータ値がソートされたら、最後には、この「セット」に、利用可能な全ての入力が含まれるが、関連性を示す順序で含まれる。次に、このリストから、将来の入力として使用される入力パラメータ（濃度）の数を決定するために、さらなる基準が使用される。本発明の態様によれば、選択はトレーニングの間に行われた（様々な数の入力パラメータが予測能力の点で比較され、当該能力は試験データセットを用いて評価された）。

【0065】

好ましくは、代謝状態を予測しようとする細胞培養物の中で測定された、この代謝産物サブセットの濃度だけが、既にトレーニング済みのMLPの入力パラメータ値として選択的に使用される。（細胞外代謝産物の全ての代謝的に利用可能な濃度の代わりに）予測の上で関連性があり、かつ独立した入力変数の選択的使用が有利であるかもしれない。なぜなら

10

20

30

40

50

、これは「オーバーフィッティング」の問題を小さくし、データ収集を簡略化し（モデルの全ての細胞外代謝産物濃度を測定することは必要でないかもしれない）、評価するのに必要な入力パラメータ値が少なくなるので予測のためのコンピュータリソースの必要性を小さくするからである。

【0066】

態様によれば、それぞれの細胞内代謝産物の流入フラックスの和がその代謝産物の流出フラックスの和に等しくなるように、モデルの細胞内代謝産物の代謝モデルは細胞内代謝産物の量が一定のままであるという定常状態仮定を表している。

【0067】

態様によれば、（現在培養されている細胞培養物の細胞外フラックスを予測するのに使用した）複数の時点は、10分～48時間、好ましくは1～24時間の時間間隔だけ隔てられる。好ましい態様によれば、細胞外フラックスの予測に使用した時点間の時間間隔は、トレーニングデータセットの作成に使用した時間間隔より短い、またはそれに等しい、または最大で20%長い。

【0068】

態様によれば、複数の時点は、細胞培養作業の期間にわたる等しい長さの時間間隔または細胞培養作業の終わりに近づくにつれて長さが短くなる時間間隔だけ隔てられる。好ましい態様により、細胞外フラックスの予測に使用した時点間の時間間隔の変化のプロファイルは、トレーニングデータセットの作成に使用した時間間隔の変化のプロファイルと同一であるか、または非常に似ている。

【0069】

特定の細胞タイプは原核生物でも真核生物でもよい。

【0070】

特に、特定の細胞タイプは真核細胞タイプでもよい。

【0071】

真核生物代謝プロセスの複雑さが高いのにもかかわらず、本発明の態様は、細胞の代謝状態、特に、数時間～1日～2日先の将来の期間の細胞の代謝状態を正確に予測できることが分かっている。

【0072】

例えば、特定の細胞タイプは哺乳動物細胞タイプ、例えば、HELA細胞などでもよい。

【0073】

一態様によれば、特定の細胞タイプはチャイニーズハムスター卵巣(CHO)細胞である。

【0074】

態様によれば、特定の細胞タイプは、生体分子を得る目的でバイオリアクター内で維持されかつ/または増殖する遺伝子組換えされた細胞タイプである。例えば、特定の細胞タイプは、特定のタンパク質、例えば、酵素もしくは特定の抗体を発現する、および/または特定のタンパク質を特に多い量で発現する遺伝子組換えされた細胞株でもよい。

【0075】

態様が決定された後に、細胞内フラックス計算値が妥当性および一貫性について、ならびに/またはさらなる品質基準に関して評価され、かつ必要であれば、化学量論式を追加、除去、もしくは変更することによって改変される。次いで、測定および/または予測された細胞外代謝産物フラックスは、改変された代謝モデルに渡され、細胞内フラックスは再計算され、妥当性および/または一貫性について再評価される。従って、代謝モデルの質は改善されることがあり、必要であれば、特定の細胞タイプまたは細胞クローンに適合されることがある。これらの妥当性基準を用いて、細胞内フラックスの妥当な参照値も、いくつかの実験的試験に基づいて本発明の態様に従って取得され得る。

【0076】

態様によれば、将来の時点のそれぞれでの、1種または複数種の細胞内代謝産物の細胞内フラックスの計算は代謝モデルのいくつかの、または好ましくは全ての細胞内フラックスの計算を含む。好ましくは、モデルの全ての細胞内フラックスが計算される。例えば、妥

10

20

30

40

50

当性の評価において、考慮される細胞内フラックスが多ければ多いほど、予測の信頼性が高くなる。

【0077】

態様によれば、前記方法は、限界値を超えて、それぞれの参照値または参照値範囲から逸脱した全ての細胞内フラックスの特定を含む。参照値または参照値範囲は、例えば、経験的に得られてもよい、および/または文献から導き出されてもよい。前記方法は、細胞増殖または所望の生体分子の産生の制限要因として作用する細胞内フラックスの自動特定をさらに含む。

【0078】

多くの代謝産物について、これらの近似的な細胞内フラックスは、特定の代謝状態という面に関して、例えば、反応速度モデルによって、¹³C標識基質および代謝産物のアイソトポマーのNMRを介した定量によって、または細胞タンパク質などのアミノ酸組成によって既知である。従って、これらの参照範囲からの強い逸脱は、細胞培養物における細胞の代謝が好ましくない状態または少なくとも予想外の状態であることを示している。参照値を介して細胞の代謝状態について結論を出すために、本発明の態様が細胞内フラックスを参照値と比較し、例えば、細胞外代謝産物の濃度を参照値と比較しないことは有利である。なぜなら、これにより、生理学的によくあるか、または好ましい参照値からの細胞の生理学的状態の（概して望ましくない）逸脱をきめ細かく、かつより良好に決定できる可能性があるからである。

【0079】

さらなる局面において、本発明は、特定の細胞タイプの細胞の細胞培養物を含むバイオリアクターをモニタリングおよび/または制御するための方法に関する。前記方法は、本明細書に記載の細胞の代謝状態を予測するための方法の態様および実施例に従って、将来の時点での細胞内フラックスを計算する工程を含む。前記方法は、細胞内フラックス予測値を、それぞれの1種または複数種の細胞内代謝産物の許容可能な細胞内フラックスの参照値または参照値範囲と比較する工程をさらに含む。

【0080】

前記方法はバイオリアクターをモニタリングするのに使用される場合があり、警告を発する工程を含む場合があり、警告は、細胞内フラックス計算値の、そのそれぞれの参照値もしくは参照値範囲からの逸脱が限界値を超えたら発せられる。警告は、例えば、グラフィカルユーザーインターフェイスを介して人間に、および/または別のインターフェイスを介して、逸脱をログ記録する機械もしくはソフトウェアプログラムに送られてもよい。

【0081】

さらに、もしくは代わりに、前記方法はバイオリアクターを制御するのに使用される場合があり、制御コマンドをバイオリアクターに送る工程を含む場合がある。制御コマンドは、逸脱を小さくするように、バイオリアクターの状態もしくはその中に含まれる培地を変える工程を自動的に開始するように送信される。例えば、制御コマンドはリアクター内のバルブ、ポンプ、または他のアクチュエーターに向かい、培養培地、微量元素、酸素、CO₂、pHを調節する酸もしくは塩基の添加、またはそれに対応した添加のスロットリングを引き起こしてもよい。特に、自動工程は培養培地の量または組成の変化を伴うことがある。従って、例えば、培養培地中の特定の糖、塩、および/もしくはアミノ酸の量を変えるために、または予測されるフラックスに応じてバイオリアクターへの培養培地の供給速度を下げるか、もしくは上げるために、制御コマンドは、培養培地のミキサーに、または培養培地の供給経路にあるスロットリングユニットに送られることがある。

【0082】

例えば、前記方法の態様によれば、特定の細胞内代謝経路が、予想よりも、または望んだものよりもかなり弱い（少ない細胞内フラックスにより示される）ことが確かめられるか、または予測されることがある。この代謝経路は、多くの場合、培地中の特定のビタミンまたは微量元素、例えば、鉄の量によって制限されることが分かっている場合がある。従って、この特定の細胞内フラックスが予想よりも少ないことが検出されれば、物理的パラ

10

20

30

40

50

メータまたは細胞外パラメータ、例えば、温度、pH、グルコース濃度を一定に保つようにすることだけ注意が払われる場合よりも、かなり特異的に、バイオリアクターへの的を絞った鉄添加が相殺されるかもしれない。

【0083】

態様によれば、バイオリアクターをモニタリングおよび/またはを制御するための方法は、本明細書に記載の細胞の代謝状態と、いくつかの細胞内フラックスを予測するための方法の態様および実施例に従って、細胞増殖または所望の生体分子の産生の制限要因として作用する、細胞の代謝モデル内の反応を特定する工程を含む。前記方法は、細胞増殖もしくは生体分子の産生または生体分子の質が促進されるように、制限要因として作用する細胞内フラックスを（独占的に、または特に）変える物質（特に、酵素、補酵素、微量元素、または栄養分）を選択的に自動的に追加する工程をさらに含む。さらに、もしくは代わりに、前記方法は、ユーザーインターフェイスを介して、このような追加の要求を出力する工程を含む。

【0084】

このことは有利であるかもしれない。なぜなら、培地中で測定される、いくつかのパラメータを含むバイオリアクターの動作パラメータを一定に保とうとすることだけを試みることに基づく当技術分野において公知の方法を用いた場合よりも、さらに詳細なモニタリングまたはバイオリアクターを制御するための非常に具体的な制御手段の採用が可能になるからである。例えば、アミノ酸の取り込みおよび代謝が多いことは、望ましい標的タンパク質を合成するために細胞もアミノ酸を利用することを必ず意味するとは限らない。細胞の細胞内フラックスの状態に応じて、吸収されたアミノ酸が完全に異なる目的で代謝されることもあり得る。しかしながら、本発明によれば、これは、MLP予測を使用する、細胞の特定の代謝モデルに基づいた代謝フラックス解析によって成し遂げられ得る。

【0085】

前記方法の態様によれば、細胞外フラックス予測値および細胞内フラックス予測値もモデルの質を試験するのに使用される。例えば、特定のCHO細胞クローンについて定式化されたモデルが別のクローンまたは別の細胞株（他の組織もしくは動物種）とはかなり異なるのであれば、フラックス予測値（MFAにより細胞内、MLPにより細胞外）は濃度変化速度の測定値または妥当性基準（非現実的な、もしくは非生理的に高いフラックスなどがない）と比較され得る。予測値が、測定されたフラックスまたは妥当なフラックスから大きく逸脱している場合、このモデルは調整されるか、またはどのモデルが補正されたかに基づいて代謝モデルに誤差が存在することを示すものとみなされる。

【0086】

フラックス予測値を、測定されたフラックスまたは妥当なフラックスと比較することによって、細胞外代謝産物またはバイオマスの濃度の決定における測定誤差が、態様に従って、例えば、統計検定によって特定される。

【0087】

さらなる局面において、本発明は、特定の細胞タイプの細胞の代謝的に有利なクローンを特定するための方法に関する。前記方法は、

- いくつかのバイオリアクターの中で異なる細胞培養物を培養する工程であって、異なる細胞培養物が、特定の細胞タイプの細胞の異なるクローンを含む、工程；
- 本明細書に記載の細胞の代謝状態を予測するための方法の態様および実施例に従って、細胞クローンのそれぞれについて別々にいくつかの時点で、1種または複数種の細胞内代謝産物の細胞内フラックスを計算する工程；
- 代謝的に最も好ましい、1種または複数種の細胞内代謝産物の細胞内フラックス計算値を有する、細胞クローンのうちの1つを特定する工程を含む。

【0088】

このことは有利であるかもしれない。なぜなら、薬学的合成プロセスの状況では、代謝的に有利な細胞クローンを特定し、選択的に増殖させることが必要なことが多いからである

。細胞を遺伝子組換えする多くの既知の方法では、合成しようとする標的タンパク質をコードする特定の遺伝子が挿入されるかどうか、および合成しようとする標的タンパク質をコードする特定の遺伝子が細胞ゲノムのどの位置に挿入されるかは完全に制御できない。ゲノム内での位置に応じて、発現速度が変化することがある。例えば、ウイルスを用いた細胞のトランスフェクションは、多くの異なる細胞クローンが作り出され、このうちの一部が、望ましい遺伝子が挿入された全ての位置および他の位置に望ましい遺伝子を含まず、ゲノム内の異なる位置に望ましい遺伝子を含む。本発明に従って、いくつかのバイオリアクターを異なる細胞クローンと共にリアルタイムで並列で動作させることによって、標的タンパク質をコードする遺伝子が細胞ゲノムに組み込まれており、細胞内で標的タンパク質がかなり合成されていることが細胞内フラックスから分かるかどうか明らかになるかもしれない。例えば、標的タンパク質のアミノ酸組成を、個々のアミノ酸を合成または分解するための細胞内フラックスと比較することによって、標的タンパク質が組み込まれたかどうか分かるかもしれない。さらに、もしくは代わりに、細胞内フラックスから、特定の細胞クローンが十分に速く再生し、生きているかどうかについての情報と、毒性の、または他の点で望ましくない代謝産物などの形成速度が遅いかどうかについての情報が得られるかもしれない。

【0089】

しかしながら、本発明に従う方法の態様は細胞の将来の代謝状態を予測するのに用いられるだけでなく、細胞の現在の代謝状態を表すのにも用いられることがある。

【0090】

態様によれば、前記方法は、

- 現在の時点および以前の時点で測定された細胞外代謝産物濃度から、細胞外代謝産物のうちの1つまたは複数の現在の細胞外フラックスを計算する工程；
- さらに代謝フラックス解析を行って、細胞外代謝産物の現在の細胞外フラックス計算値と代謝モデルの化学量論式を用いて現在の時点での現在の細胞内フラックスを計算する工程；および
- 細胞培養物の細胞の現在の代謝状態を表すものとして現在の細胞内フラックス計算値を使用する工程を含む。

【0091】

このことは有利であるかもしれない。なぜなら、細胞培養物における細胞の現在の代謝状態が個々の細胞内フラックスのレベルまで非常に正確に評価されるからである。

【0092】

態様によれば、測定値はまた乳酸デヒドロゲナーゼ(LDH)濃度も含み、細胞培養物の培養中の時点のそれぞれで、細胞培養物の培地中の測定されたLDH濃度が受け取られる。将来の時点のそれぞれでの、細胞外代謝産物の細胞外フラックスの予測は、細胞密度測定値ではなく、補正された細胞密度を用いてMLPによって行われる。

【0093】

好ましくは、それぞれの時点について、補正された細胞密度の計算は、LDH濃度測定値の関数として、細胞培養物の培地中の溶解細胞の密度を計算する工程を含む。この関数は、特に、その特定の細胞タイプの溶解細胞の数に対する、培地中のLDH濃度の依存性を表す経験的に求められたヒューリスティックな一次関数でもよい。次いで、補正された細胞密度は、培地中の細胞密度計算値と溶解細胞の密度計算値の和として計算される。

【0094】

このことは有利であるかもしれない。なぜなら、細胞密度を求める光学的方法では、完全に溶解された細胞または部分的に溶解された細胞を検出できないか、または不十分にしか検出できないことが多く、溶解細胞は、溶解のすぐ前まで細胞外代謝産物の濃度に影響を及ぼしてきた可能性があるからである。細胞密度を求める現行の方法では、構造が依然として完全な状態の細胞しか検出しない。発酵のもっと遅い時点で起こる傾向がある細胞腐敗が起これば、これにより、全細胞密度の測定が変化する。場合によっては、態様によっ

10

20

30

40

50

て本明細書に記載の方法を用いたフラックス予測において系統誤差が観察されたことが観察され、その結果、フラックス予測値は、測定可能なフラックスと一致せず、系統誤差を示した。この誤差は培地中のLDH濃度と相関することが観察された。LDH濃度は、溶解細胞が培地中に存在することの指標である。この酵素は無傷の細胞によって培地に放出されない。従って、反応培地中にLDHが検出されると細胞が破壊されたことが分かる。

【0095】

バッチ発酵を除いて、LDH濃度と溶解細胞数との同様の、ほぼ直線的な関係は多くの発酵アプローチにおいて証明されている。態様によれば、培地中のLDH濃度は、補正された細胞密度を計算するための、さらなる測定値として使用される。

【0096】

本発明の態様によれば、補正された細胞密度は、トレーニングの間、またはトレーニング済みMLPを適用している間の入力パラメータ値として使用される。

【0097】

態様によれば、MLPに基づく細胞外フラックス予測では、細胞外代謝産物の濃度に加えて、LDH濃度が、モデルの細胞外フラックスの少なくとも一部を予測するための入力変数として使用される。態様によれば、細胞外代謝産物の濃度の加えて、または細胞外代謝産物の濃度の代わりに、細胞密度のLDH補正值が出力変数として使用される。従って、本発明の態様によれば、LDH濃度測定値によって補正された細胞密度が、MLPトレーニングにおいて細胞外フラックスを計算するために使用される、および/またはLDH濃度が、さらなる入力パラメータ値として、もしくは細胞外代謝産物の濃度として使用される。

【0098】

別の局面において、本発明は、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するためのシステムに関する。前記システムは、1つまたは複数のプロセッサと、細胞培養物を含有するバイオリアクターから測定値を受け取るための第1のインターフェイスと、揮発性または非揮発性の記憶媒体とを備える。

【0099】

記憶媒体は特定の細胞タイプの細胞の代謝モデルを含み、代謝モデルは、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックスおよび細胞外フラックス(408)を含み、細胞内代謝産物のうちの1つと細胞外代謝産物のうちの1つとの間の少なくとも1つの化学量論的関係を指定する化学量論式を含む。さらに、記憶媒体は、トレーニング済み機械学習プログラムロジック(MLP)と、細胞培養物の培養中の複数の時点のそれぞれにおける細胞の代謝状態を予測するための方法を行うように訓練されたプログラムロジックを含む。この方法は、

- 第1のインターフェイスを介して、その時点で測定された複数の測定値を受け取る工程であって、前記測定値が、細胞培養物の培養培地における代謝モデルの複数種の細胞外代謝産物の濃度と、細胞培養物における細胞の細胞密度測定値を含む、工程；
- 受け取った測定値を入力パラメータ値としてMLPに入力する工程；
- 前記受け取った測定値を用いて、前記MLPによって将来の時点での前記細胞外代謝産物の細胞外フラックスを予測する工程であって、前記将来の時点が、前記測定値を受け取った時点より後の時点であり、前記細胞外フラックスが、細胞への前記細胞外代謝産物の取り込み速度および/または細胞から前記培地への前記細胞外代謝産物の放出速度である、工程；
- 細胞外フラックス予測値と代謝モデルの化学量論式とを用いて、将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程を含む。

【0100】

代謝モデルは、例えば、MatLabにおいて定義された一組の化学量論反応式の形で記憶媒体に記憶されてもよい。記憶媒体はコンピュータのメインメモリでもよく、電磁的または光学的な記憶媒体、例えば、ハードディスクでもよい。記憶媒体はまた、いくつかのハードウェア記憶装置の分散システム、例えば、クラウドサービスによって提供されるクラウド

10

20

30

40

50

ドストレージ領域または研究室のITインフラストラクチャーでもよい。前記システムは、例えば、標準的なコンピュータとして実行されてもよく、ユーザーのノートブックまたは携帯型モバイル装置として実行されてもよい。前記システムは、検査情報システム(Laboratory Information System)(LIS)に組み込まれていてもよいし、動作可能に接続されていてもよい。しかしながら、前記システムはまた1つもしくは複数のバイオリアクター用の制御コンピュータでもよく、単一のバイオリアクターに（例えば、当該バイオリアクターの重要な部分として）可逆的または不可逆的につながられた制御モジュールでもよい。

【0101】

例えば、第1のインターフェイスは、バイオリアクターの1つまたは複数のセンサーとの無線接続または有線接続を確立するネットワークインターフェイスでもよい。さらに、もしくは代わりに、第1のインターフェイスはまた、対応する測定値の手作業による入力用のインターフェイスでもよい。例えば、測定値は、特定の時点に、手作業による方法、半自動方法、または完全自動方法でバイオリアクターから試料を採取することによって得られ、次いで、試料は1つまたは複数のさらなる分析装置に輸送され、次いで、細胞外代謝産物の細胞密度および/または濃度が求められる可能性がある。そして次に、このように得られた測定値は、第1のインターフェイスを介して、これらの分析装置から自動的に前記システムに送られてもよく、または、ユーザーは、グラフィカルユーザーインターフェイスとして設計された第1のインターフェイスを介して測定値を前記システムに手作業で入力してもよい。

【0102】

態様によれば、前記システムは、1つもしくは複数のバイオリアクターをモニタリングおよび/もしくは制御するための制御装置であるか、またはこのような制御装置に動作可能に接続されている。好ましくは、前記システムは、細胞内フラックス計算値をユーザーに出力するためのユーザーインターフェイスをさらに備える。例えば、フロー予測値が表の形で表示されてもよく、フローが動画によって、例えば、細胞内フロー予測の基礎をなす代謝モデルのグラフ表示に基づく動画によって動的に視覚化されてもよい。

【0103】

前記システムの態様によれば、前記システムはまた、制御コマンドをバイオリアクターに送るための第2のインターフェイスも備える。このプログラムロジックは、

- 細胞内フラックス予測値を、それぞれの細胞内代謝産物の許容可能な細胞内フラックスの参照値もしくは参照値範囲と比較する；
 - 細胞内代謝産物の少なくとも1つの細胞内フラックス計算値の、そのそれぞれの参照値もしくは参照値範囲からの逸脱が限界値を超えたら、ユーザーインターフェイスを介して警告を発する；および/または
 - 第2のインターフェイスを介して制御コマンドをバイオリアクターに送る
- ように訓練されており、制御コマンドは、逸脱が小さくなるように、バイオリアクターの状態もしくはその中に含まれる培地を変えるように訓練されている。

【0104】

ユーザーインターフェイスは、例えば、グラフィカルユーザーインターフェイス(GUI)および/または音響ユーザーインターフェイスでもよい。例えば、警告音が音響ユーザーインターフェイスを介して発せられてもよく、文章の形をとった警告、好ましくは、逸脱が検出された細胞内フラックスの条件付きの表示のある文章の形をとった警告がGUIを介して画面に表示されてもよい。画面は、バイオリアクターとつながった画面でもよく、ネットワークを介して1つまたは複数のバイオリアクターと接続したコンピュータ、例えば、デスクトップコンピュータ、サーバー、または移動通信装置、例えば、ユーザーのスマートフォンの画面でもよい。

【0105】

第2のインターフェイスは、例えば、前記システムとバイオリアクターまたはバイオリアクターアクチュエーターとの間の無線接続または有線接続として物理的に、適合されていてもよい。アクチュエーターは、ポンプ、様々な栄養分、緩衝液、pH調節液、微量元素、

10

20

30

40

50

気体のバルブ、および/またはスターラーもしくは温度コントローラでもよい。

【0106】

「トレーニング時点」とは、トレーニング記録を作成している間の時点である。対照的に、請求項1記載の「時点」とは、現在モニタリングおよび/または制御されているバイオリアクターの代謝状態を予測するために、トレーニングデータセットに対してトレーニングされたMLPが適用される、もっと後の時点を目指す。

【0107】

「代謝産物」とは、この用語の狭い意味では生化学代謝経路における中間産物(中間体)である。しかしながら、本発明の意味の「代謝産物」は、抽出物、産物、または中間体の形で細胞の生化学反応に関与する、どの物質よりも広いものとする。特に、代謝産物は、アミノ酸、糖、脂肪、ペプチド、抗体、タンパク質、クエン酸回路成分、解糖成分、タンパク質合成または分解経路の成分、および類似する物質でもよい。

10

【0108】

「細胞外代謝産物」とは、例えば、特定の細胞タイプの細胞によって細胞培養物の培地に分泌されるか(例えば、乳酸)、または培地もしくは培養培地の成分として細胞培養物に添加されるために(例えば、グルコース)、研究中の細胞タイプの代謝モデルに従って細胞培養物の培地中で生じることが知られているか、または想定される代謝産物だと本明細書で理解される。

【0109】

「細胞内代謝産物」は、例えば、細胞培養物の培地から取り入れられるか、または細胞によって産生されるために、研究中の細胞タイプの代謝モデルに従って特定の細胞タイプの細胞内で生じることが知られているか、または想定される代謝産物と本明細書で定義される。

20

【0110】

「フラックス」とは、特定の輸送または反応プロセスによる特定の体積中での時間あたりに変化する物質の量である。従って、フラックスとは「反応速度」または「輸送速度」とも呼ばれる。いくつかのプロセスが特定の体積中で同時に起こる場合、ある物質の正味の濃度は、いくつかのプロセスの反対方向に回転する性質により、例えば、物質変換、取り込み、または放出により変化しない可能性がある。

【0111】

態様によれば、フラックス表示が指す体積は、細胞内代謝産物ならびに細胞外代謝産物については細胞の体積である。

30

【0112】

特定のプロセスによって引き起こされる物質の量の変化は細胞または細胞体積と関連するので、フラックスは、この特定の反応または輸送プロセスによって引き起こされる、特定のプロセスによって引き起こされる、この体積中での、時間あたりのこの物質の濃度の変化も暗黙的に示す。ある特定の体積の場合、物質のフラックスは、特定のプロセスによって引き起こされる、この体積中での、この物質の濃度の変化も暗黙的に示し、かつ逆もまた同じであるので、本明細書中でのフラックスの概念は、ある体積中での時間あたりの物質の量の変化ならびに時間あたりの濃度の量の変化を等しく含むはずである。

40

【0113】

好ましくは、本発明の態様に従って用いられる代謝モデルは、細胞内代謝産物の量または濃度がほぼ一定のままであると仮定する。しかしながら、このことは、細胞内フラックスが一定のままであることを意味しない。もっと正確に言うと、サイトゾル中で細胞内代謝産物濃度が急増した細胞は、この急増を、例えば、この代謝産物を代謝する1つまたは複数の化学反応の反応速度を速めることによって補償し得る。

【0114】

「細胞外フラックス」は、細胞外代謝産物が特定の輸送プロセスを介して細胞によって取り入れられるか、または放出される速度だと本明細書で理解される。もっと正確には、「細胞外フラックス」とは、この特定の輸送プロセスによって細胞1個あたり、および時間

50

あたりの細胞に取り入れられる代謝産物の量、または細胞によって周囲の培地に放出される代謝産物の量である。本発明の態様によれば、時間 t での、ある成分の細胞外フラックス v は、リアクター培地中の対応する代謝産物の濃度の変化の測定値に基づいて求められ、これは生細胞密度を用いて基準化される。

【0115】

態様によれば、例えば、培養培地中での細胞外代謝産物濃度の変化は時間間隔にわたって測定される。これは、（リアクター内の培地の体積を用いて）培地中の代謝産物量の絶対的变化を計算するのに用いられることがある。培地中の生細胞数が測定されることがある。細胞外代謝産物の量の絶対的变化を細胞数に対して基準化することによって、細胞外フラックスが比の（バイオマス関連）量とされることがある。培養細胞タイプの1細胞の平均体積は、通常、文献から分かっているか、または測定され得るので、本発明の態様に従って、測定された濃度の変化が平均細胞体積に変換される。

10

【0116】

例えば、代謝フラックス解析（「フロー解析」または「マテリアルフロー解析」ともいう）に用いられる細胞代謝の代謝モデルは細胞外化学量論反応式を含むことがあり、これらはそれぞれ細胞外フラックスと関連付けられている。従って、細胞外フラックスは、特に、細胞内代謝産物として後で機能するような、細胞培養物の細胞内への細胞外代謝産物の取り込み速度と、細胞外代謝産物として後で機能するような、周囲の培地への細胞培養物の細胞の細胞外代謝産物の放出速度を記述する。

【0117】

細胞内外への輸送フラックスを表す、これらの細胞外フラックスを介して、純粋に細胞内のフラックスは代謝モデルにおいて細胞外代謝産物の濃度および濃度変化と結び付けられ、このモデルにおいて指定された化学量論と、経験的に求められたものに基づいて妥当性基準の決定を可能にするか、または化学量論と、妥当性基準と、このモデルにおいて指定された培地中の細胞外代謝産物の経験的に求められたか、もしくは予測された濃度変化を用いた細胞外フラックスに基づいて細胞の現在の細胞内フラックスの記述的評価および/もしくは細胞の将来の細胞内フラックスの予測を可能にする。

20

【0118】

従って、細胞内フラックスと細胞外フラックスは1種または複数種の細胞内代謝産物を介して代謝モデルにおいて一緒に結び付けられる。なぜなら、一部の物質は、1種もしくは複数種の細胞内代謝産物から形成されるか、または1種もしくは複数種の細胞内代謝産物に代謝される速度が述べられている細胞内代謝産物と、細胞に輸送されるか、細胞から放出される速度がモデルの細胞外フラックスによって述べられている細胞内代謝産物の両方として生じるからである。

30

【0119】

細胞外代謝産物濃度の変化、従って、細胞外フラックスの変化はまた、バイオリアクターのいくつかのタイプでは、細胞培養物培地への、この代謝産物の外部添加または供給によって引き起こされ得る。細胞外フラックスは、例えば、最初に、2つの連続した時点で測定された代謝産物濃度の絶対差を求め、次に、この差を細胞密度測定値に変換することによって近似的に計算され得る。細胞密度が多ければ多いほど、細胞1個あたりのフラックスは少なくなる。なぜなら、代謝産物濃度の測定された絶対的变化が、さらに多数の細胞に分配されるからである。従って、将来の時点での細胞外代謝産物の取り込みフラックスは、時間間隔の長さで割った、2つの時点間の時間間隔において測定された代謝産物濃度の差として計算され、次に、この結果は細胞密度測定値によって割られることがある。

40

【0120】

「細胞内フラックス」は、細胞内で反応が起こる速度（時間間隔と関連する量）だと本明細書で理解される。この反応は、細胞内区画間の輸送（例えば、サイトゾルからミトコンドリアへの輸送。逆もまた同じである）または細胞内での1種もしくは複数種の細胞内代謝産物（抽出物）から1つもしくは複数の他の代謝産物（産物）への変換からなってもよい。細胞内フラックスの場合、全ての抽出物および産物が細胞内代謝産物である。

50

【0121】

細胞内フラックスは、上記のカテゴリの可逆的反応または不可逆的反応（細胞区画間の細胞内輸送、代謝変換）を指定する化学量論式の代謝モデルにおいて定式化される。

【0122】

「測定不可能なフラックス」とは、発酵データから容易に求められないことがあるフラックスである。一般的に、細胞内フラックスは細胞内代謝産物の観察が難しいために測定不可能である。

【0123】

本発明の好ましい態様によれば、提供された細胞外フラックスに基づいて細胞内フラックスを求めるMFAは細胞内代謝産物の濃度が変化しないという仮定に基づいている。細胞内代謝産物の濃度は測定が非常に難しいか、または不可能である。本発明の態様は、MFAでは、（濃度が未知の）細胞内代謝産物の入力フローと出力フローの和が同一であり、従って、その濃度が変化しないという仮定に基づいている。リアクターの動作モードに応じて、これは全くそのとおりというわけでもないが、細胞代謝がすぐに化学平衡に戻るので、本明細書で説明された時間間隔に関して少なくともほぼ十分に正しい。

【0124】

態様によれば、MFAにおいて用いられる代謝モデルは、少なくとも10、好ましくは少なくとも20の化学量論式を含む。このモデルは、好ましくは、細胞内での代謝プロセスを完全に、またはほぼ完全に記述するように適合されている。

【0125】

できるだけ包括的に細胞代謝を含む代謝モデルを使用することは有利であるかもしれない。なぜなら、特定の細胞クローンを選択するには、できるだけ、調査された各細胞クロンの個々の代謝活性および特徴の全体像を掴むことが重要だからである。MLPによって個々の化学量論式によって予測された細胞外フラックスの純粋な妥当性チェックは、一般的に、特定の細胞クロンの有利な代謝特性があるために、このモデルに基づいて特定の細胞クローンを選択するのに十分なデータを提供しない。

【0126】

別の有利な局面では、MFAにおける代謝モデルと多数の化学量論式の使用は細胞培養物をモニタリングするのにも使用されることがあり、セットポイント範囲から1種または複数種の代謝産物フラックスが逸脱した場合は、供給される栄養分溶液の量および/もしくは組成物または他の制御パラメータ（温度、pH、酸素分圧、CO₂分圧、スターラー速度など）の特定の変化を自動的におよび/または手作業で加えるようも使用されることがある。

【0127】

「代謝モデル」は、特定の細胞タイプの細胞の現在の代謝状態の記述モデルだと本明細書で理解される。代謝モデルは、好ましくは、細胞内で起こる真の反応と比較して複雑度が低く、それぞれの用途について特に関心が高い細胞代謝部分に限定される。好ましくは、このモデルは、いくつかの細胞外代謝産物および細胞内代謝産物、化学量論因子を用いた反応式および輸送式、ならびに細胞外フラックスおよび細胞内フラックスを含む。細胞の代謝状態は、少なくとも部分的には、特定の時点での個々の細胞内フラックスのレベルを示すことによって特徴付けられ得る。好ましくは、細胞内の代謝産物量の時間的变化は、モデルの細胞外反応式および細胞内反応式において指定されるような流入マテリアルフローと流出マテリアルフローとの均衡化に起因する。このモデルは、代謝フラックス解析を行うのに適したMFAモデルであり、いわゆる、細胞内代謝産物の量が一定のままであるという定常状態仮定に基づいている。この定常状態仮定とは、それぞれの細胞内代謝産物の流入フラックスの和が流出フラックスの和に等しいことを意味する。

【0128】

トレーニング済みMLPの入力パラメータ値として測定および使用される細胞外代謝産物の濃度は本発明の態様によれば厳密な意味の代謝産物濃度である。この用語の狭い意味の代謝産物濃度は、体積に関連する含有率の指定だと本明細書で理解される。従って、濃度は、参照体積（例えば、細胞培養物培地）にどれくらいの代謝産物が存在するかを示してい

10

20

30

40

50

る。代謝産物の濃度は、例えば、単位g/lの質量濃度として示されてもよいし、単位mol/lの物質濃度として示されてもよい。

【0129】

他の態様によれば、トレーニング済みMLPの入力パラメータ値として測定および使用される細胞外代謝産物の濃度は広い意味の代謝産物濃度である。広い意味の代謝産物濃度は、本明細書では、測定値と、それぞれの代謝産物の狭い意味の代謝産物濃度と直線的に、または少なくともほぼ(問題となっている濃度範囲の少なくとも90%)直線的に相関することが知られている、測定値から導き出された値を意味する。例えば、細胞外代謝産物の細胞外フラックスは広い意味の代謝産物濃度と理解され得る。細胞外フラックス測定値とは、過去の時点、例えば、代謝産物濃度の前回の測定値と、測定値が現在収集されており、MLPの入力として用いられる現在の時点との間の期間中での、培地中の代謝産物の濃度の変化を指す。測定されたフラックスと細胞外濃度との間の本質的に直線的な関係は、代謝産物の細胞外フラックスが特定の量分だけ増加すると、それに対応して、細胞外培地中の代謝産物の濃度が変化するという事実起因する。さらに、本発明の態様によれば、代謝産物濃度測定値は様々なやり方で補正係数および標準化係数で相殺することによって修正されることがあり、その結果、これらの修正値は最終的には広い意味の代謝産物濃度も表す、すなわち、最初に測定された値と直線的に相関しているが、同一ではない。実際に測定された代謝産物濃度と広い意味の代謝産物濃度は互いに直線的に相関するか、または代謝産物濃度が狭い意味の代謝産物濃度測定値から導き出され得るので、両タイプの代謝産物濃度データは最終的にはMLPの入力として等しく用いられ得る。

【0130】

「記述モデル」とは、細胞の現在の静的代謝状態および動的代謝状態を記述するモデルである。

【0131】

「予測モデル」とは、細胞の将来の静的代謝状態および動的代謝状態を予測するモデルである。態様によって、フラックス解析に用いられる代謝モデルは記述代謝モデルである。

【0132】

「代謝ハイブリッドモデル」とは、反応速度の知識を表す機構モデルと、バイオリアクターからの測定値と細胞の代謝状態とを関連付ける経験モデルの組み合わせである。

【0133】

「代謝マテリアルフロー解析」(代謝フラックス解析、MFA)とは、細胞外フラックスから細胞内フラックスを推定するのに使用され得る計算方法である。細胞外マテリアルフローは、細胞によって吸収または放出された物質の量を経時的に表す。細胞内マテリアルフローは、細胞内代謝産物に変換または形成される反応速度を表す。MFAは、生物学的単位内での代謝反応速度を求めるための様々なアプローチを含む。代謝は動的プロセスであり、MFAによって説明される、ある種の「細胞表現型」だとみなされ得る。

【0134】

「バッチバイオリアクター」とは、「バッチプロセス」で動作するか、またはバッチプロセスで動作するように適合されたバイオリアクターである。バッチプロセスは、接種される前に、全ての基質(特に、糖およびアミノ酸)がバイオリアクター内に提供されることを特徴とする。他方で、気体およびpH調整剤もプロセス中に前記系内に導入される。しかしながら、概してバッチプロセスは、一定の反応体積がある閉じた系だとみなされ得る。細胞が基質を徐々に消費するため、初期の無制限の増殖期(いわゆる対数期)の後に、細胞の増殖と死が平衡状態になる停滞期(定常期)が生じる。この後に、栄養分が極度に不足した結果として細胞密度が減少する死滅期が続く。通常のバイオリアクター動作方法の中でバッチプロセスは一番安価であり、汚染を引き起こす可能性が一番低い。しかしながら、通常、バッチプロセスは、最適な産物収率をもたらさない。プロセス期間は基質の消費によって制限される。

【0135】

「フェドバッチバイオリアクター」とは、「フェドバッチプロセス」で動作するか、また

はフェドバッチプロセスで動作するように適合されたバイオリアクターである。フェドバッチプロセスでは、プロセス中に（多くの場合、初回バッチ期後にだけ）追加の培養培地が添加される。供給は連続でもよく、1回または複数回の高濃度大量投与（すなわち、パルス）の形をとってもよい。バッチ方法と比較して使用済みの基質を再投与できるので、プロセス時間を延長することができる。さらに、もっと良いプロセス制御が実現する可能性がある。例えば、抽出物濃度を連続して低く保つことによって阻害現象と毒性副産物形成を食い止めることができる。全体的に見て、フェドバッチプロセスを用いると、バッチ方法よりもかなり高い細胞密度と産物収率が得られる可能性がある。

【0136】

「ケモスタット」（「連続リアクター」）とは、一定の培養培地の流れが供給され、細胞および産物を含有する反応培地が同程度まで抜き取られるバイオリアクターである。従って、発酵中に反応体積は変化しない。体積流量が適切に選択されればリアクター内で定常状態平衡が確立され、細胞密度および栄養分濃度は一定のままである。

【0137】

「灌流バイオリアクター」または「灌流リアクター」とは、連続した培養培地の流れが添加され、連続した（通常、無細胞）反応培地の流れが抜き取られるバイオリアクターである。連続した細胞含有反応培地の流れはまた、方向をもった「ブリーディング」によって取り出されることもある。体積流量の選択は1つまたは複数のプロセス制御パラメータと結び付けられ、リアクター内の反応体積が一定のままになるように調整される。細胞密度および栄養分濃度はプロセスデザインに応じて一定でもよく（細胞密度に関連する）、可変でもよい（フェドバッチ動作に類似する）。

【0138】

「スプリットバッチバイオリアクター」とは、培地のかなりの部分、例えば、10%超または30%超が、それに含まれる細胞を回収する目的でリアクターから1回または複数回取り出されるように動作するバッチタイプまたはフェドバッチタイプのバイオリアクターである。

【0139】

「PMI」（部分的相互情報量）とは、2つのパラメータの非直線的な直接依存性を定量するデータ値である。いくつかの機械学習アプローチ、例えば、ニューラルネットワークの状況では、PMIは、既に選択されている入力を考慮に入れた、入力確率変数Xと出力確率変数Yとの間の依存性の尺度である。2つの変数のPMIを計算するための様々なアプローチが知られている。例えば、Sharma A (2000): 「Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 - A strategy for system predictor identification」, Journal of Hydrology Vol. 239, Issues 1-4, 232-239。

【0140】

従って、「PMI基準」とは、決定を下すためにPMIに関して明らかにされている特色または特徴を指す。例えば、PMI基準は限界値の場合があり、限界値を上回るか、または限界値を下回ると方法の経過が影響を受ける。

【0141】

本発明のいくつかの実施例が、本願に添付された付属書においてさらに詳細に説明され、付属書の開示内容は本願の一部である。明細書と付属書との間の一貫性を保証するために、付属書の変数一覧表において指定された変数の意味が維持される。変数の意味に関して、付属書にある変数のリストについて言及がなされる。実施例および態様ならびに付属書に記載されたさらなる説明は、出願文書に記載の態様、実施例、および特徴と、これらが相互排他的でなければ自由に組み合わせることができる。

[本発明1001]

以下の工程を含む、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するための方法：

- 該特定の細胞タイプの細胞の代謝モデル(402)を準備する工程(102)であって、該代謝

10

20

30

40

50

モデルが、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックス(410)および細胞外フラックス(408)を含み、該代謝モデルが、該細胞内代謝産物(406)のうちの1つと該細胞外代謝産物(404)のうちの1つとの間の少なくとも1つの化学量論的關係を指定する化学量論式を含む、工程；

- 該細胞培養物の培養中の複数の時点のそれぞれにおける、

○該時点で測定された複数の測定値を受け取る工程(106)であって、該測定値が、該細胞培養物の培養培地における該代謝モデルの複数種の細胞外代謝産物の濃度と、該細胞培養物における該細胞の細胞密度測定値とを含む、工程；

○受け取った該測定値を、入力パラメータ値として、トレーニング済みの機械学習プログラムロジック(MLP)(218)に入力する工程(108)；

○受け取った該測定値を用いて、該MLPによって、将来の時点での細胞外代謝産物の細胞外フラックス(408)を予測する工程(110)であって、該将来の時点が、該測定値を受け取った時点より後の時点であり、該細胞外フラックスが、細胞内への細胞外代謝産物の取り込み速度および/または細胞から培地への細胞外代謝産物の放出速度である、工程；

○該細胞外代謝産物の該細胞外フラックスの予測値と該代謝モデルの該化学量論式とを用いて、該将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程(112)。

[本発明1002]

前記方法が、機械学習によりMLPを作成する工程をさらに含み、該作成する工程が、

- トレーニングデータセットを作成することであって、

○前記特定の細胞タイプの細胞の少なくとも1つのトレーニング細胞培養物の培養中の複数のトレーニング時点のそれぞれにおいて、

■該トレーニング時点で測定された複数の測定値を受け取る工程であって、該測定値が、該少なくとも1つのトレーニング細胞培養物の培養培地における代謝モデルの複数種の細胞外代謝産物の濃度と、該少なくとも1つのトレーニング細胞培養物における細胞の細胞密度測定値を含む、工程、

■現在のトレーニング時点の時間表示を受け取る工程、および

■その時点で受け取った測定値と、それぞれの前の時点で受け取った測定値との関数として、該細胞外代謝産物の細胞外フラックスを計算する工程であって、該細胞外フラックスが、細胞内への該細胞外代謝産物の取り込み速度および/または培地への該細胞外代謝産物の放出速度である、工程

を含む、トレーニングデータセットを作成すること、

- 該MLPをトレーニングすることであって、

○該トレーニング時点のそれぞれにおいて受け取った測定値を、入力パラメータ値として該MLPに入力し、そのトレーニング時点の後続の各時点において該後続の時点について計算された細胞外フラックスを、該入力パラメータ値に関連する出力パラメータ値として該MLPに入力する工程；ならびに

○該MLPが、該入力パラメータ値に基づいて、それぞれの該関連する出力パラメータ値を予測するように学習するように、該MLPによる学習プロセスを実施する工程を含む、該MLPをトレーニングすること、

- トレーニングされた該MLPを揮発性または非揮発性の記憶媒体に記憶すること

を含む、本発明1001の方法。

[本発明1003]

前記特定の細胞タイプの細胞の複数のトレーニング細胞培養物の培養中の複数のトレーニング時点のそれぞれにおいて、前記測定値および時間表示が受け取られ、かつ前記細胞外代謝産物の細胞外フラックスが計算されるように、前記トレーニングデータセットが作成され、ここで、該細胞培養物が、異なるタイプのバイオリアクターの中で培養され、該バイオリアクターのタイプが、フェドバッチバイオリアクター、バッチバイオリアクター、灌流リアクター、ケモスタット、およびスプリットバッチバイオリアクターを含むバイオリアクタータイプのうちの少なくとも2つの異なるバイオリアクタータイプを含む、本

10

20

30

40

50

発明1002の方法。

[本発明1004]

前記MLPが、1つのニューラルネットワークであるか、またはいくつかのニューラルネットワークからなるシステムである、前記本発明のいずれかの方法。

[本発明1005]

前記複数種の細胞外代謝産物の濃度測定値がそれぞれ、

- 該代謝産物の体積当たりの含有量、特に、質量濃度もしくは物質濃度、または
- 該体積当たりの含有量と直線的もしくは少なくともほぼ直線的に相関する値、特に、補正もしくは標準化された、該細胞外代謝産物の代謝産物濃度測定値もしくはフラックス測定値

である、前記本発明のいずれかの方法。

[本発明1006]

前記MLPが、いくつかのサブMLPからなるシステムであり、該システムに含まれる個々のサブMLPがそれぞれ、単一の細胞外代謝産物の細胞外フラックスを予測するようにトレーニングされており、かつ、該単一の細胞外代謝産物の将来の時点での細胞外フラックスを予測するのに選択的に用いられる、前記本発明のいずれかの方法。

[本発明1007]

前記MLPが、前記細胞外代謝産物のうちの1つの細胞外フラックスを予測するために、複数種の細胞外代謝産物の濃度測定値を入力パラメータ値として使用し、ここで、細胞外フラックスを求めようとする該細胞外代謝産物のうちの少なくとも2つについて、入力パラメータ値として使用する該複数種の細胞外代謝産物が異なる、前記本発明のいずれかの方法。

[本発明1008]

- 全ての入力候補代謝産物の濃度を複数の時点にわたって測定する工程であって、該入力候補代謝産物が、前記特定のタイプの細胞培養物を有する参照バイオリアクター内で測定可能な程度に利用可能な全ての細胞外代謝産物を含むか、または前記代謝モデルの全ての細胞外代謝産物を含む、工程；

- 細胞外フラックスを予測しようとする単一の細胞外代謝産物のそれぞれについて、該単一の代謝産物の細胞外フラックスを予測するために入力パラメータ値として使用される複数種の細胞外代謝産物を特定するための選択プロセスを行う工程であって、該選択プロセスが、該単一の代謝産物それぞれに関して、

- (a)細胞外代謝産物の第1のセットを定める工程であって、該第1のセットが全ての候補入力代謝産物を含む、工程；

- (b)該第1のセットにおける各細胞外代謝産物の第1の関連性スコアを、その代謝産物の濃度測定値の関数として計算する工程であって、該第1の関連性スコアが、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの細胞外代謝産物の濃度の予測能力を示す、工程；

- (c)該細胞外代謝産物のうち最も高い第1の関連性スコアを有するものだけを、該第1のセットから、まだ空である細胞外代謝産物の第2のセットに移し、この代謝産物を該第1のセットから除去する工程；

- (d)該第1のセットにおける各細胞外代謝産物のさらなる関連性スコアを、その代謝産物の濃度測定値と、該第2のセットに含まれる全ての細胞外代謝産物の濃度測定値との関数として計算する工程であって、該さらなる関連性スコアが、既に該第2のセット含まれている代謝産物を考慮した、該単一の細胞外代謝産物の細胞外フラックスに関する、それぞれの候補入力代謝産物の濃度の予測能力を示す、工程；

- (e)該第1のセットの細胞外代謝産物のうち最も高いさらなる関連性スコアを有するものだけを該第2のセットに移し、この代謝産物を該第1のセットから除去する工程であって、この代謝産物を含めることによって、該単一の細胞外代謝産物の細胞外フラックスの予測に関する、第2のセットに含まれる代謝産物の情報冗長性の最大限度を第2のセットが超えない場合にだけ、この代謝産物に移される、工程；

10

20

30

40

50

(f)該第2のセットが該情報冗長性の最大限度を超えることなしでは、これ以上の代謝産物を該第1のセットから該第2のセットに移すことができなくなるまで、工程(d)および(e)を繰り返す工程；ならびに

(g)該単一の細胞外代謝産物の細胞外フラックスを予測するために、該第2のセットに移された代謝産物だけを選択的に入力パラメータ値として使用する工程を含む、選択プロセスを行う工程をさらに含む、本発明1007の方法。

[本発明1009]

- 前記第1のセットの代謝産物と、細胞外フラックスを予測しようとする前記単一の代謝産物との間の部分相互情報量スコア(PMIスコア)として、前記第1の関連性スコアが計算され；

- 既に前記第2のセットに含まれている全ての代謝産物を考慮して、該第1のセットの代謝産物と、細胞外フラックスを予測しようとする該単一の代謝産物との間のPMIスコアとして、第2の関連性スコアが計算される、本発明1008の方法。

[本発明1010]

前記MLPが、単一の代謝産物のそれぞれの細胞外フラックスを予測するために、複数種の細胞外代謝産物の濃度測定値を入力パラメータ値として使用し、該複数種の細胞外代謝産物が、少なくとも1種のアミノ酸、好ましくは少なくとも2種のアミノ酸を含む、前記本発明のいずれかの方法。

[本発明1011]

前記複数の時点が、10分～48時間、好ましくは1～24時間の時間間隔だけ隔てられている、前記本発明のいずれかの方法。

[本発明1012]

前記細胞タイプが、生体分子を得る目的でバイオリアクター内で維持されかつ/または増殖する遺伝子組換えされた細胞タイプである、前記本発明のいずれかの方法。

[本発明1013]

前記将来の時点のそれぞれで、前記代謝モデルの細胞内フラックスのうちのいくつかまたは全ての計算が行われる、前記本発明のいずれかの方法。

[本発明1014]

- それぞれの参照値または参照値範囲から閾値を超えて逸脱している細胞内フラックス(410)計算値を有する、前記代謝モデルの1種または複数種の細胞内代謝産物を特定する工程；

- 細胞増殖または所望の生体分子の産生の制限要因として作用する細胞内フラックスを自動的に特定する工程をさらに含む、前記本発明のいずれかの方法。

[本発明1015]

以下の工程を含む、特定の細胞タイプの細胞の細胞培養物を含むバイオリアクターをモニタリングおよび/または制御するための方法：

- 本発明1001～1014のいずれかの方法に従って将来の時点での該細胞の細胞内フラックスを計算する工程；

- 該細胞内フラックスの予測値を、それぞれの1種もしくは複数種の細胞内代謝産物の許容可能な細胞内フラックスの参照値もしくは参照値範囲と比較する工程；

- 少なくとも1つの細胞内代謝産物の細胞内フラックスの計算値の、そのそれぞれの参照値もしくは参照値範囲からの逸脱が限界値を超えた場合に、警告を発する工程；ならびに/または

- 該逸脱を小さくするように該バイオリアクターの状態もしくはその中に含まれる培地を変える工程を自動的に開始するための制御コマンドを該バイオリアクターに送る工程であって、該自動工程が、特に培養培地の量および/もしくは組成を変えることを含む、工程。

[本発明1016]

- 本発明1014に従って、細胞増殖もしくは所望の生体分子の産生の制限要因として作用する、細胞の代謝モデル内の反応を特定する工程；および

- 細胞増殖もしくは該生体分子の産生または該生体分子の質が促進されるように、制限要因として作用する細胞内フラックスを改変する物質を選択的に自動的に追加するか、あるいは、このような追加の要求をユーザーインターフェイスを介して発行する、工程をさらに含む、本発明1015の方法。

[本発明1017]

以下の工程を含む、特定の細胞タイプの細胞の代謝的に好ましいクローンを特定するための方法：

- 複数のバイオリアクターの中で異なる細胞培養物を培養する工程であって、該異なる細胞培養物が、該特定の細胞タイプの細胞の異なるクローンを含む、工程；

- 各細胞クローンについて別々に複数の時点で、本発明1001～1016のいずれかに従って、1種または複数種の細胞内代謝産物の細胞内フラックスを計算する工程；

- 代謝的に最も好ましい、該1種または複数種の細胞内代謝産物の細胞内フラックス計算値を有する、該細胞クローンのうちの1つを特定する工程。

[本発明1018]

- 1種または複数種の細胞外代謝産物の現在の細胞外フラックスを、現在の時点および以前の時点で測定された該細胞外代謝産物の濃度から計算する工程；

- 該細胞外代謝産物の該現在の細胞外フラックスの計算値と代謝モデルの化学量論式とを用いて、さらなる代謝フラックス解析を行って、現在の時点での現在の細胞内フラックス(410)を計算する工程；ならびに

- 該現在の細胞内フラックスの計算値を、細胞培養物の細胞の現在の代謝状態の特徴付けとして使用する工程

をさらに含む、前記本発明のいずれかの方法。

[本発明1019]

- 前記細胞培養物の培養中の前記時点のそれぞれにおいて、該細胞培養物の培地中の乳酸デヒドロゲナーゼ(LDH)濃度の測定値がさらに受け取られ；

- 前記将来の時点のそれぞれにおける前記細胞外代謝産物の細胞外フラックスの予測が、前記細胞密度測定値ではなく補正された細胞密度を用いて、前記MLPによって行われ、該補正された細胞密度の計算が、該時点のそれぞれにおける、

○該LDH濃度測定値の関数として、該細胞培養物の培地中の溶解細胞の密度を計算する工程であって、該関数が、該培地中のLDH濃度と前記特定の細胞タイプの溶解細胞の数との関係を表す、経験的に求められたヒューリスティックな一次関数である、工程；

○該培地中の該細胞密度測定値と該溶解細胞の密度計算値との和として、該補正された細胞密度を計算する工程

を含む、

前記本発明のいずれかの方法。

[本発明1020]

以下を備える、特定の細胞タイプの細胞の細胞培養物の代謝状態を予測するためのシステム(200)：

- 1つまたは複数のプロセッサ(202)、

- 該細胞培養物を含むバイオリアクター(204、206、208)からの測定値を受け取るための第1のインターフェイス(210)、ならびに

- 以下を備える、揮発性または非揮発性の記憶媒体(212)：

○該特定の細胞タイプの細胞の代謝モデル(402)であって、複数種の細胞内代謝産物および細胞外代謝産物ならびに複数種の細胞内フラックス(410)および細胞外フラックス(408)を含み、該細胞内代謝産物(406)のうちの1つと該細胞外代謝産物(404)のうちの1つとの間の少なくとも1つの化学量論的關係を指定する化学量論式を含む、代謝モデル(402)；

○トレーニング済みの機械学習プログラムロジック(MLP)(218)；

○該細胞培養物の培養中の複数の時点のそれぞれにおいて以下の工程を含む方法を行う

ように訓練された、プログラムロジック(220)：

■第1のインターフェイスを介して該時点で測定された複数の測定値を受け取る工程(106)であって、該測定値が、該細胞培養物の培養培地における該代謝モデルの複数種の細胞外代謝産物の濃度と、該細胞培養物における該細胞の細胞密度測定値とを含む、工程；

■受け取った該測定値を入力パラメータ値として該MLPに入力する工程(108)；

■受け取った該測定値を用いて、該MLPによって、将来の時点での該細胞外代謝産物の細胞外フラックス(408)を予測する工程(110)であって、該将来の時点が、該測定値を受け取った時点より後の時点であり、該細胞外フラックスが、細胞内への該細胞外代謝産物の取り込み速度および/または細胞から培地への該細胞外代謝産物の放出速度である、工程；

■該細胞外フラックス予測値と該代謝モデルの該化学量論式とを用いて、該将来の時点での細胞内フラックスを計算するために代謝フラックス解析を行う工程(112)。

【図面の簡単な説明】

【0142】

以下では、本発明の態様が例示的にさらに詳細に説明され、それぞれが本発明の態様またはこれらの態様の個々の局面を表す図面について言及がなされる。

【0143】

【図1】細胞の代謝状態を予測するための方法のフローチャートを示す。

【図2】様々な装置およびデータソースを用いて、いくつかの段階で情報を取得するプロセスの一例を示す。

【図3A】図3は、1つまたは複数のバイオリアクターをモニタリングおよび/または制御するのに使用され得る、細胞の代謝状態を予測するためのシステムのブロック図を示す。

【図3B】図3Aの説明を参照。

【図4A】図4は、細胞内代謝産物および細胞外代謝産物の複数種の細胞内フラックスおよび細胞外フラックスがある細胞の代謝モデルを示す。

【図4B】図4Aの説明を参照。

【図5】バイオリアクターを動作させている間の、いくつかの連続した時点での細胞内フラックスの計算を示す。

【図6】図4に示した代謝モデルに従う異なる代謝産物のフラックスを示す。

【図7】生物学的知識を生み出すための方法のうまくいった使用を例示した、いくつかの代謝産物フローを示す。

【図8】生物学的知識を生み出すための方法のうまくいった使用を例示した、いくつかの代謝産物フローを示す。

【図9】乳酸フラックスおよびグルタミン濃度を用いたグラフを示す。

【図10】細胞内フラックスの時間経過を示す。

【図11A】図11は、細胞培養物の培養中の異なる時点でのいくつかの細胞内フラックスおよび細胞外フラックスを示す。

【図11B】図11Aの説明を参照。

【図11C】図11Aの説明を参照。

【図11D】図11Aの説明を参照。

【図12】記述的MFAによって現在の時点について計算した細胞内フラックスと、MLPとMFAの組み合わせによって将来の時点について予測した細胞内フラックスの強い相関を持つ経過を示す。

【図13】NNの入力パラメータ値および出力パラメータ値を示す。

【図14】12のフェドバッチ発酵動作における細胞内フラックスを対象にした、得られたRMSEのヒストグラムを示す。

【図15】12のフェドバッチ発酵動作における細胞外フラックスを対象にした、得られたRMSEのヒストグラムを示す。

【図16】それぞれに、2つの異なるバイオリアクター（フェドバッチおよびスプリットバッチ）の中での同一の細胞クローンを対象にした1つの細胞外代謝産物フラックス予測

10

20

30

40

50

値と2つの細胞外代謝産物フラックス測定値がある12個のグラフを示す。

【図17】それぞれに2つの曲線がある11個のグラフを示す。全てのグラフが、2つの異なる計算方法を用いたフェドバッチバイオリアクターについて得られた。

【図18A】図18は、それぞれ細胞クローンZK1の細胞外フラックス計算値の2つの曲線がある、11個のグラフを示す。

【図18B】図18Aの続きを示す。

【図19A】図19は、それぞれ細胞クローンZK2の細胞外フラックス計算値の2つの曲線がある、11個のグラフを示す。

【図19B】図19Aの続きを示す。

【図20A】図20は、それぞれ細胞クローンZK1の細胞内フラックス計算値の2つの曲線がある、11個のグラフを示す。

【図20B】図20Aの続きを示す。

【図21A】図21は、それぞれ細胞クローンZK1の細胞内フラックス計算値の2つの曲線がある、11個のグラフを示す。

【図21B】図21Aの続きを示す。

【発明を実施するための形態】

【0144】

図1は、態様に従ってCHO細胞の細胞培養物の代謝状態を予測するための方法のフローチャートを示す。この方法は他の細胞タイプに同じように適している。

【0145】

工程102:モデル作成

バイオリアクター内のプロセスは数式で表すことができる。最初に、反応培地中の関連性のある物質濃度または量の時間変化（例えば、基質量、産物量、細胞密度の経過）の写像を考慮しなければならない。式は、物質収支に基づいており、物質の反応を表す項と、リアクター内外への任意の物質フローを含む対流項からなる。これは、一般式（付属書の[51]、セクション4.2を参照されたい）：

$$\begin{array}{ccccccc} \text{系内の成分の} & & \text{時間とともに} & & \text{時間とともに} & & \text{時間とともに} \\ \text{物質量の} & = & \text{系に加えられた} & - & \text{系から取り出された} & + & \text{系内で変換された} \\ \text{経時変化} & & \text{成分の量} & & \text{成分の量} & & \text{成分の量} \\ & & \underbrace{\hspace{10em}} & & \underbrace{\hspace{10em}} & & \\ & & \text{対流項} & & \text{反応項} & & \end{array}$$

または式：

$$\frac{dm}{dt} = \dot{V}_{zu} \cdot c_{zu} - \dot{V}_{ab} \cdot c_{ab} + Q \quad (2.1)$$

で成り立つ。

式中、mは反応培地中の物質の量、tはプロセス時間

$\dot{V}_{zu}$

または

$\dot{V}_{ab}$

はそれぞれ内外の流量体積、 c_{zu} または c_{ab} はそれぞれ内外の対応する物質の濃度、ならびにQは時間および体積あたりに変換される物質の量である。細胞外代謝産物の量が考慮される場合、反応項は主として細胞による物質の取り込みまたは放出を含む。ファーマンター内の細胞密度が記述される場合、これは細胞の形成と死を含む。

【0146】

態様によれば、代謝モデルは、物質の量の時間経過が微分可能だという上記の式に従う仮定に基づいている。これは、全ての流入フラックスおよび流出フラックスが連続していれ

ば正当化される。発酵中に大量投与供給またはサンプリングする場合、結果は区分的連続曲線である。次に、上記の式は不連続点間の領域において成り立つ。

【0147】

この微分方程式では、プロセスの説明、制御、および最適化のために非常に多くの数学モデルが指定されることがある。これらは、費用と時間がかかる実験室実験を削減しながら、バイオプロセスをより良く理解し、インシリコで改善するという高まる欲求により、ますます大きな役割を果たしている。このような物質収支にだけ基づき、細胞内プロセスを記述しないモデルはブラックボックスモデルとして知られている。このようなモデルは、考慮されているプロセス間の依存性を機構的に説明できないことがある。機構的に説明するために、異なる細胞外物質を結び付けるものとして代謝をモデル化することが必要であろう。

10

【0148】

最終的には、代謝モデルは代謝マテリアルフロー解析が行われるのを可能にしなければならず、その結果、細胞外フラックスから細胞内フラックスに結論を下すのに使用され得る。一般的に、細胞密度を求め、細胞外物質濃度を測定するための使い易い方法が確立しているが、細胞内反応速度の観察はかなり複雑である。このような実験を回避するために、細胞外フラックスから細胞内フラックスを推定するのに使用され得る計算方法である代謝マテリアルフロー解析（代謝フラックス解析、MFA）が開発されている。細胞外マテリアルフローは、経時的に、細胞によって吸収または放出される物質の量を記述する。細胞内マテリアルフローは、時間当たりの細胞1個当たりの細胞内反応によって変換される物質の量である。

20

【0149】

従って、代謝マテリアルフロー解析を実施できるようにするために、最初に、検討されている生物の（好ましくは、または通常、簡素化された）生化学的・化学量論的代謝ネットワークが作成される。これは最も重要な細胞内反応と細胞外反応を含む。細胞外反応はフラックスに類似し、代謝産物が細胞によって取り入れられるか、または放出される反応である。

【0150】

ネットワークは、 k 種類の反応（ k_m は測定可能であり従って既知である（これらは通常、細胞外反応である））および l 種類の細胞内代謝産物からなる。次に、これらの反応は、化学量論係数（個々の反応の抽出物についてはマイナス、産物についてはプラス）が入力される化学量論行列

30

$$A \in \mathbb{R}^{l \times k}$$

で記録することができ、行は様々な代謝産物に対応し、列は反応に対応する。細胞外代謝産物は省略されている。具体例を付属書のセクション2に示した。

【0151】

j 番目の反応のマテリアルフローを v_j で示した。さらに、 m_i は、1つの細胞の中にある i 番目の細胞内代謝産物の量である。フラックスと代謝産物量が組み合わされてベクトル v または m になった場合、以下が成り立つ：

40

$$\frac{dm}{dt} = Av \quad (2.2)$$

【0152】

この式は、細胞における代謝産物量の時間的変化が流入マテリアルフローと流出マテリアルフローの均衡化に起因することを示している。

【0153】

次いで、MFAは、いわゆる、細胞内代謝産物の量が一定のままであるという定常状態仮定に基づいている。この定常状態仮定は、それぞれの細胞内代謝産物の流入フラックスの和が流出フラックスの和に等しいことを意味する。これは、式(2.2)を

50

$$0 = Av \quad (2.3)$$

に簡約する。

【 0 1 5 4 】

ベクトル v を、既知の（測定可能な）フラックスのサブベクトル

$$v_m \in \mathbb{R}^{k_m}$$

と、未知のフラックスのサブベクトル

$$v_u \in \mathbb{R}^{k-k_m}$$

に分け、それに対応して、行列 A を部分行列

$$A_m \in \mathbb{R}^{l \times k_m}$$

および

$$A_u \in \mathbb{R}^{l \times (k-k_m)}$$

に分けると、 $Av = A_m v_m + A_u v_u$ になり、式(2.3)が

$$A_u v_u = -A_m v_m \quad (2.4)$$

になる。

【 0 1 5 5 】

従って、MFAによる未知のフラックス v_u の決定は、連立一次方程式を解くことによって実施され得る。

【 0 1 5 6 】

次の工程は、前記の連立方程式を分類する工程である。前記の連立方程式の解を得るための自明の式

$$v_u = -A_u^{-1} A_m v_m$$

は通常、成り立たない。なぜなら、通常、行列 A_u は可逆でないからである。このような場合、解空間は無限になることがあるか、または解が存在しないように矛盾が生じることがある。van der Heijden et al.(1994)によって以下の項が導入され、これは、可解性および一貫性の基準に従って前記の連立方程式またはまたはマテリアルフローを分類し[17, 39,40]、本発明の態様に従ってモデルを作成するのに使用される。

【 0 1 5 7 】

決定:

$\text{Rang}(A_u) < k - k_m$ が成り立つ場合、前記の式(2.4)は過小決定されている。この場合、全ての未知のフラックスが明確に計算され得るわけではない。なぜなら、代謝ネットワークは、あまりに少ない制限を含むからである。 $\text{Rang}(A_u) = k - k_m$ の場合、前記の式は最大で1つの解を有し、決定されたと呼ばれる。

【 0 1 5 8 】

冗長性:

$\text{Rang}(A_u) < l$ の場合、前記の式は冗長である。これは、線形従属線が A_u に存在することを意味する。 v_m の決定に測定誤差があるために、または代謝ネットワークモデルに不正確さがあるために、これは、通常、解が存在しない一貫性のない式につながり（

$$\text{Rang}(A_u) < \text{Rang}((A_u | -A_m v_m))$$

が成り立つ）、後の方の項は拡大係数行列を表す。 $\text{Rang}(A_u) = l$ であれば、前記の式は冗長でなく、従って常に一貫している。

【 0 1 5 9 】

計算可能性(Kalkulierbarkeit):

フラックス v_u は、式(2.4)を用いて明確に計算されれば計算可能とされ、そうでなければ計算可能でない。一貫性のある式が本明細書において想定される。前記の式が過小決定されたら、計算可能でない少なくとも1つのフラックスがある。

【0160】

均衡可能性(Balancierbarkeit):

フラックス v_m は、その値が前記の式の一貫性に影響を及ぼすのであれば均衡可能だと呼ばれ、そうでなければ均衡可能でない。均衡可能なフローは冗長な式でしか生じない。

【0161】

上述したように、MFA式は、多くの場合、過小決定される、および/または冗長である（だが、ある式は同時に、過小決定され、かつ冗長になることがある）。次に、ムーア・ペンローズ型疑似逆行列(Moore-Penrose pseudo-inverse)

$$A_u^\#$$

を用いて解を式：

$$v_u = -A_u^\# A_m v_m \quad (2.5)$$

で表すことができる。これは全ての A_u について定義される。

【0162】

過小決定の場合、この式から、連立方程式に対して無限に多くの解の1つが得られる。不一致の場合、最小二乗解が得られる。

【0163】

以下において、過小決定の場合に、未知のフラックスのどれが計算可能かどうか、冗長性の場合にフラックス測定値のどれが均衡化できるかどうかチェックされ得る方法を示す。

【0164】

計算可能なフラックスの特定

Klamtらにより発表された、計算可能なフラックスを特定するための方法は[17]に示されている。

$$T \in \mathbb{R}^{(k-k_m) \times (k-k_m-\text{Rang}(A_u))}$$

が、 A_u のコアの基底を形成する列を有する行列である場合、

$$A_u T = 0$$

が成り立つ。

【0165】

それぞれのベクトル

$$p \in \text{Kern}(A_u)$$

は基底ベクトルの線形結合として表すことができる。その結果、

$$a \in \mathbb{R}^{k-k_m-\text{Rang}(A_u)}$$

となり、

$$p = Ta$$

が成り立つ。

【0166】

式(2.4)は、以下：

$$A_u v_u = -A_m v_m + A_u T a$$

のように展開することができ、式中、

$$a \in \mathbb{R}^{k-k_m-\text{Rang}(A_u)}$$

は任意である。次に、疑似逆行列の左辺適用により

$$v_u = -A_u^\# A_m v_m + T a$$

となる。

【0167】

ベクトル a の変化により連立式的(2.4)の体系の解空間が得られる。この式が過小決定されているのであれば、 T は0より大きなランクを有する。これから、計算可能なフラックス v_u は、 a の変化が影響を及ぼさないものであると結論を下すことができる。これは、行列 T の対応する行が0行である場合である。

【0168】

好ましくは、本発明の態様によれば、次の工程は、計算可能なフラックスの特定である。

【0169】

冗長でない式の場合、(2.5)を(2.4)に挿入すると、以下の式で表される。

$$Rv_m = 0 \quad (2.6)$$

式中、冗長性行列 R は

$$R := A_m - A_u A_u^\# A_m$$

である（付属書の参考文献[17,39]を参照されたい）。

【0170】

しかしながら、式が冗長の場合、式(2.6)は、特定の v_m についてしか満たされず、これは、式(2.4)の可解性に等しい。列の表記

$$Rv_m = r_1 v_{m,1} + \dots + r_{k_m} v_{m,k_m} = 0$$

は、対応する列ベクトル R がヌルベクトルであれば、フラックス測定値が式(2.4)の可解性に影響を及ぼさないことを示している。上記の式において、 r_j は R の j 番目の列ベクトルを示し、 $v_{m,j}$ は j 番目のフラックス測定値を示す。

【0171】

以下では、過小決定されたおよび冗長なシステムまたは対応する代謝モデルを本発明の態様に従って処理するためのオプションが示される。

【0172】

過小決定された式の処理

式が過小決定されている場合、すなわち、無限の数の解を有する場合、ユニークな解に到達するために問題を変更するいくつかの手法がある。1つの選択肢は、可能であれば、結果として生じた制限を、代謝ネットワークモデルを拡大することによって、従って、決められた式を作り出すことによって厳しくすることであろう。さらに、さらなる実験定量化によって既知のフラックスの数を増やすことが試みられるかもしれない。細胞内フラックスの決定は ^{13}C 標識実験によって成し遂げられ得る（付属書の文献[43]を参照されたい）。

【0173】

この実験作業を回避したければ、広く行き渡ったフラックス均衡解析(Flux Balance Analysis)(FBA)方法が良い選択肢である。ここでは、最適化しようとする標的関数が定義される。これは生物学的妥当性に従って選択され、マテリアルフローに依存する。例えば、宿主生物は増殖速度が最大になるようにマテリアルフローを向けると想定することができる。なぜなら、これは大きな進化上の利点であるからである。標的関数が F と指定されれば、FBAにより、

一般式:

$$\begin{aligned} \max \quad & F(v) \\ \text{s.t.} \quad & Av = 0 \end{aligned}$$

が得られる。

【0174】

これを式の制約を用いて最適化問題として定式化することによって、通常、明らかなフラックス分布が解として提供される。反応の不可逆性についての情報が入手可能であれば、

許容可能な範囲（許容可能な範囲:最適化問題の全ての制約が満たされている点集合）を、さらなる不等条件

$$v_{irrev} \geq 0$$

によってさらに制限することができる。式中、 v_{irrev} は全ての不可逆的フラックスのベクトルである。

【0175】

FBAにおける主な問題は、解が依存する目的の関数を正確に選択することである。発酵中に細胞が生物学的標的を変える可能性は十分にある（付属書の文献[33]を参照されたい）。

【0176】

冗長な式の管理

冗長な連立方程式の場合、通常、均衡化することができるフローに矛盾があるために、式(2.4)を解くフラックス分布は無い。完全に正確に定義されたネットワークモデルでも、 v_m の決定における測定誤差によって引き起こされる不一致が通常、生じる。実際に測定されたフロー値は、真の測定可能な値とはっきりと区別するために、以下において

$$\bar{v}_m = (\bar{v}_{m,1}, \dots, \bar{v}_{m,k_m})^T$$

と指定すべきである。文献には、冗長なMFA問題の近似解

$$\hat{v} = [\hat{v}_m, \hat{v}_u]$$

を計算するいくつかの方法がある。

【0177】

1つの可能性は、

$$\hat{v}_m = \bar{v}_m$$

を設定し、ベクトル

$$A_u \hat{v}_u$$

と

$$-A_m \bar{v}_m$$

間の平均二乗距離が最小になるようにベクトル

$$\hat{v}_u$$

を選択することであろう。次に、最適化問題としての式は

$$\min_{v_u} \|A_u v_u + A_m \bar{v}_m\|^2 \quad (2.7)$$

である。

【0178】

対応する解

$$\hat{v}_u = -A_u^\# A_m \bar{v}_m$$

だけが定常状態仮定を近似的に満たす。式が決定されたら、次に、

$$A_u^\# = (A_u^T A_u)^{-1} A_u^T$$

が成り立つ。

【0179】

第2の可能性は、フラックス測定値との相対二乗距離を最小化し（付属書の文献[39]を参照されたい）、定常状態状態を満たすベクトル v の最小二乗解を決定することである。この

10

20

30

40

50

式は

$$\min_{\mathbf{v}} \quad \frac{1}{2} \sum_{j=1}^{K_m} \left(\frac{v_{m,j} - \bar{v}_{m,j}}{\bar{v}_{m,j}} \right)^2 \quad (2.8)$$

$$\text{s.t.} \quad A\mathbf{v} = \mathbf{0}$$

である。

【0180】

さらなる制約として不可逆性も含まれることがある。

【0181】

第2の方法は、主に、その解が定常状態仮定を満たすという点で第1の方法と異なる（次に、フラックス測定値は均衡化されたと呼ばれる）。このために、解における測定可能なフラックスの値は、実際に測定された値に近似的にしか対応しない。定常状態仮定が、常に誤差を受けるフラックスの測定値よりも信頼性が高いとみなされたら、これは妥当であるとみなされ得る。(2.8)の解では、均衡化することができるフラックス測定値は全て調整される。均衡化できないものは変化しないままである。

【0182】

均衡可能なフラックスの調整について、ちょうど説明した方法の一般化は、付属書[39,40]に引用した文献において説明した、さらに統計学的に動機づけられたアプローチに起因する。これは加重最小二乗アプローチに基づいている。

【0183】

$\bar{\mathbf{v}}_m$

を、測定した、誤りのあるフラックスのベクトルとし、 $\mathbf{v}_m$ を、対応する真の値のベクトルとする。測定誤差ベクトル δ は真の値と測定値との差を示す。

$$\delta := \mathbf{v}_m - \bar{\mathbf{v}}_m$$

δ は正規分布していると想定される。期待ベクトル $\mathbf{0}$ および共分散行列

$$C_{\bar{\mathbf{v}}_m} := \mathbb{E}[\delta\delta^T]$$

【0184】

$\mathbf{v}_m$ は未知であるので、共分散行列の妥当な、経験に基づく推定を、さらなる計算に使用しなければならない。

【0185】

この目的は、共分散についての情報を考慮に入れた、定常状態仮定を満たすと同時に、

$\bar{\mathbf{v}}_m$

に近い $\mathbf{v}_m$ の推定である。これはマハラノビス距離を用いることで成し遂げられる。このための最適化問題は、

$$\min_{\mathbf{v}_m} \quad \delta^T C_{\bar{\mathbf{v}}_m}^{-1} \delta \quad (2.9)$$

$$\text{s.t.} \quad R\mathbf{v}_m = \mathbf{0}$$

$$\delta = \mathbf{v}_m - \bar{\mathbf{v}}_m$$

である。

【0186】

最適化問題の解

$\hat{\mathbf{v}}_m$

は、式

$$\hat{\mathbf{v}}_m = (\mathbf{I} - \mathbf{C}_{\bar{\mathbf{v}}_m} \mathbf{R}'^T (\mathbf{R}' \mathbf{C}_{\bar{\mathbf{v}}_m} \mathbf{R}'^T)^{-1} \mathbf{R}') \bar{\mathbf{v}}_m \quad (2.10)$$

を有する（付属書の参考文献[23]を参照されたい）。 $\mathbf{I}$ は単位行列である。 $\mathbf{R}'$ は $\mathbf{R}$ の誘導形であり、線形従属線を除くことによって作成され、従って、最大ランクを有する（これはユニークでない）。これは、対応する線形変換：

$$\mathbf{R}' = \mathbf{\Gamma} \mathbf{R} \quad (2.11)$$

を行う、非正方行列 $\mathbf{\Gamma}$ との掛け算によって作成され得る。

【0187】

これらの均衡化された値は、未知のフラックスを計算するために式(2.5)において使用され得る。

【0188】

フラックスの測定が独立しており（すなわち、共分散行列

$$\mathbf{C}_{\bar{\mathbf{v}}_m}$$

が対角であり）、

$$\bar{v}_{m,j}$$

の標準偏差が測定値

$$v_{m,j} - \bar{v}_{m,j}$$

の大きさと比例し、一定の比例定数 $b > 0$ であると仮定して、最適化問題(2.9)は(2.8)に等しい。以下：

$$\delta^T \mathbf{C}_{\bar{\mathbf{v}}_m}^{-1} \delta = (\mathbf{v}_m - \bar{\mathbf{v}}_m)^T \frac{1}{b^2} \begin{pmatrix} \frac{1}{\bar{v}_{m,1}^2} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\bar{v}_{m,k_m}^2} \end{pmatrix} (\mathbf{v}_m - \bar{\mathbf{v}}_m) = \frac{1}{b^2} \sum_{j=1}^{k_m} \left(\frac{v_{m,j} - \bar{v}_{m,j}}{\bar{v}_{m,j}} \right)^2$$

が成り立つ。

【0189】

式(2.9)は、(2.8)と比べて、共分散行列が測定データの質に融通をきかして適合され得るという利点をもたらす。測定値が信頼できないと分類されたフラックスを均衡化した時、これにより、おそらく、もっと正確に測定されたフラックスを用いた場合よりも値が大きく変わる。

【0190】

バイオプロセスエンジニアリング問題におけるMFAの適用を例示する前に、以下のセクションにおいて最初に、本発明の態様に従って行われるような代謝モデルの統計学的検証を扱う。なぜなら、これは、ちょうど説明した考慮すべき事項を基礎としているからである。

【0191】

生化学的代謝モデルの検証

推論された生化学的ネットワークモデルの質は以前の論評において扱われたことがない。しかしながら、その定式化における質が十分でないことでMFA結果に重大な欠陥が生じ得ることは明らかである。高い有意性と、できるだけ大きな単純化の中を取るものとして、意味のあるモデルを作り出すために検証方法が必要とされる。1994からのvan der Heijdenらの刊行物において、統計学的に動機づけられた検定が示されており、可能性のある系統的な誤差発生源の検出が説明されている（付属書の文献目録[40]を参照されたい）。これらの研究は、均衡化することができるフローの解析と、モデルにおける不一致への影響に基づいている。従って、これらの研究は、冗長な式が存在する場合にしか成り立たない。

10

20

30

40

50

【0192】

不一致を評価するための検定

前のセクションの式(2.11)では、冗長性行列の誘導形 R' が既に導入された。残差ベクトルは、

$$\varepsilon := R' \bar{v}_m$$

によって定義される。

【0193】

冗長な式の場合、概して

$$\varepsilon \neq 0$$

が成り立つ。 ε の共分散行列 C_ε は、

$$C_\varepsilon := R' C_{\bar{v}_m}^{-1} R'^T$$

によって計算され得る。従って、これはフラックス測定値の共分散行列に依存し、測定値の不確実性を考慮に入れている。検定のために、検定統計量 H_ε が使用され、その観測値は、

$$h_\varepsilon := \varepsilon^T C_\varepsilon^{-1} \varepsilon$$

によって示される。

【0194】

検定統計量は χ^2 分布を受けることが示されるかもしれない（付属書の[40]を参照されたい）。自由度は C_ε のランクに対応する。

【0195】

全体的に見て、以下の仮説検定が得られる。

検定

H_0 : 考慮された代謝モデルの不一致は有意でない

対

H_1 : 考慮されている代謝モデルの不一致は有意である

有意水準 α で H_0 を棄却する $\Leftrightarrow h_\varepsilon > \chi_{\text{Rang}(C_\varepsilon), 1-\alpha}^2$ (2.12)

$$\chi_{\text{Rang}(C_\varepsilon), 1-\alpha}^2$$

は、 $(1-\alpha) \cdot \chi^2$ 分布の100%分位数を示す。 $\text{Rang}(C_\varepsilon)$ は自由度である。

【0196】

可能性のある系統的な誤差発生源の検出

前もって定義された検定が不一致を示すのであれば、これは、行列

$$C_{\bar{v}_m}$$

に反映され、従って、検定の結果に影響を及ぼす測定ノイズの過小評価によるものかもしれない。3つのさらなる、可能性のある誤差発生源が付属書の文献[40]において議論されている。

【0197】

系統的測定誤差:j番目のフラックス

$$\bar{v}_{m,j}$$

の測定は系統誤差

$$\pi = \bar{v}_{m,j} - v_{m,j}$$

を受ける。

【0198】

代謝ネットワークに重要反応が無い:(k+1)番目の重要な反応がネットワークモデルに欠けている;化学量論行列Aを別の列 a_{k+1} によって拡大しなければならないだろう。対応するフラックスも v_{k+1} として示される。

【0199】

代謝ネットワークにおける反応の定義が不正確である:j番目の反応が不正確に定義されている;ベクトル $a_j + \Delta a_j$ を、化学量論行列の列ベクトル a_j の代わりに使用しなければならない。

【0200】

誤差の研究は残差ベクトル ε の構造に基づいている。すなわち、上記のそれぞれの誤差について、特徴的な比較ベクトル v が定義され得る。その方向は、誤差発生源が実際に存在すれば ε の方向とほぼ同じである。ベクトル方向間の類似性を評価する統計検定も示される。 ε の長さは誤差 s の大きさを示している。以下の表に、対応する比較ベクトルを列挙する。 r'_j は、ここでは、 R' のj番目の列ベクトルを示している。比較ベクトルの導出は、ここでは、列挙された誤差発生源の最初のものしか証明されないものとする。さらなる事例については[40]を参照されたい。

【0201】

$$r'_1 \bar{v}_{m,1} + r'_2 \bar{v}_{m,2} + \dots + r'_{k_m} \bar{v}_{m,k_m} = \varepsilon$$

が成り立つ。

【0202】

正確に定義されたネットワークモデルの場合、
 $E[\varepsilon] = 0$

が成り立つ。j番目のフラックスの真の値が系統誤差 π 分だけ測定値と異なる、すなわち、
 $v_{m,j} = \bar{v}_{m,j} - \pi$

である場合、

$$\begin{aligned} E[\varepsilon] &= E[r'_1 \bar{v}_{m,1} + r'_2 \bar{v}_{m,2} + \dots + r'_j (\bar{v}_{m,j} + \pi) + \dots + r'_{k_m} \bar{v}_{m,k_m}] \\ &= E[r'_1 \bar{v}_{m,1} + r'_2 \bar{v}_{m,2} + \dots + r'_j v_{m,j} + \dots + r'_{k_m} \bar{v}_{m,k_m}] + E[r'_j \pi] \\ &= r'_j \pi \end{aligned}$$

が成り立つ。

【0203】

従って、 ε と r'_j は同じ方向を有すると予想される。

【0204】

(表1) 3つの異なる誤差発生源と対象にした比較ベクトルと、関連する誤差サイズ

誤差発生源	比較ベクトル v	誤差サイズ s
j番目のフラックスの測定値 $\bar{v}_{m,j}$ が間違っている	r'_j	π
(k+1)番目の反応が欠けている	$\Gamma(A_u A_u^\# - I) a_{k+1}$	v_{k+1}
j番目の反応が不正確に定義されている	$\Gamma(A_u A_u^\# - I) \Delta a_j$	v_j

【0205】

ε と v との類似性を評価するために、検定統計量

$$\Delta^2 = \varepsilon^T C_\varepsilon^{-1} \varepsilon - \frac{(\varepsilon^T C_\varepsilon^{-1} v)^2}{v^T C_\varepsilon^{-1} v}$$

10

20

30

40

50

が使用される。これは、 $\text{Rang}(C_\varepsilon) - 1$ の自由度で χ^2 分布している。

【0206】

以下の仮説検定は ε と v との類似性を評価する。

検定

H_0 :ベクトル ε と v は類似する

対

H_1 :これらのベクトルは類似しない

有意水準 α で H_0 を棄却する $\Leftrightarrow \Delta^2 > \chi^2_{\text{Rang}(C_\varepsilon)-1, 1-\alpha}$ (2.13)

10

【0207】

統計的導出は付属書の参考文献[40]において見ることができる。

【0208】

態様によれば、本発明の態様に従って作成される代謝モデルはネットワークを含み、このネットワークは中心細胞内マテリアルフローを含まなければならないが、可能な限り小さな複雑度を有さなければならない。一例として本明細書で説明されるモデルは、本質的には、以下の刊行物:Altamirano C, Illanes A, Becerra S, Cairo JJ, Godia F (2006):

「Considerations on the lactate consumption by CHO cells in the presence of galactose», Journal of Biotechnology 125, 547 - 556; Llaneras F, Pico J (2007): 「A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient», BMC Bioinformatics 8:421;およびNolan RP, Lee K (2011): 「Dynamic model of CHO cell metabolism», Metabolic Engineering 13, 108-124において提唱されたネットワーク化学量論に基づいている。

20

【0209】

しかしながら、細胞区画は考慮されなかった。数が多いため、酸化還元およびエネルギー当量に関与する全ての反応が代謝モデルに含まれるとは限らない。従って、NAD(P)HおよびATPを化学量論の定式化に含めなかった。さらに、いくつかの代謝分岐を詳細に考慮しなかったが、バイオマス形成（例えば、五炭糖リン酸経路）に組み込んだ。ほとんどの場合で、定式化された反応は、分岐の無い、いくつかの連続した生化学的反応をまとめたものであり、これらは定常状態仮定に従って同一のマテリアルフローを有するはずである（例えば、解糖またはクエン酸回路のほんのわずかな中間体しか明示的に列挙していない）。

30

【0210】

バイオマス平衡ならびに生細胞密度および全細胞密度から単位mol/lへの変換をNolan(2011)による上記の刊行物から採用した。標的タンパク質のアミノ酸組成の結果として、産物形成のための化学量論の定式化が後から付いてくる。上記の刊行物は反応の可逆性を確かめるのにも使用した。

【0211】

結果として生じた代謝モデルを図4に詳細に示す。これは、図4aに、CHO細胞の細胞内マテリアルフローおよび細胞外マテリアルフローの生化学的ネットワークモデルを示す。これは、グルコース(Glc)、乳酸(Lac)、アラニン(Ala)、グルタミン酸(Glu)、グルタミン(Gln)、アンモニア(NH₃)、アスパラギン酸(Asp)の輸送および変換であり、アスパラギン(Asn)、セリン(Ser)、グリシン(Gly)、全細胞密度(BIO)、産物(Prod)、グルコース-6-リン酸(G6P)、ピルビン酸(Pyr)、 α -ケトグルタル酸(AKG)、リンゴ酸(Mal)、およびオキサロ酢酸(Oxa)を含む代謝の中心フラックスをモデル化している。可逆性は矢印の形状によって示される。

40

【0212】

図4bの表は代謝モデルの個々の化学量論反応を列挙している。これらのうち、反応1、3、9、11、13、14、16、18、20、21、22、および23は細胞外である。このモデルは、定常状態仮定が適用される13種の細胞内代謝産物を含む。添え字「e」はここでは細胞

50

外物質を指す。可逆性/反応の可逆性は反応矢印によって示される。

【0213】

代謝モデルに含まれるネットワークモデルは中心の細胞内マテリアルフローを含むが、できるだけ単純でなければならない。この論文で選択される定式化は、本質的には、Altamirano et al(2006)、Llaneras et al(2007)、およびNolan et al(2011)による上記の刊行物において提唱されたネットワーク化学量論に基づいている。しかしながら、細胞の区画は考慮されなかった。数が多いため、酸化還元およびエネルギー当量に關与する全ての反応が代謝モデルに含まれるとは限らない。従って、NAD(P)HおよびATPを化学量論の定式化に含めなかった。さらに、いくつかの代謝分岐を詳細に考慮しなかったが、バイオマス形成（例えば、五炭糖リン酸経路）に組み込んだ。ほとんどの場合で、定式化された反応は、分岐の無い、いくつかの連続した生化学的反應をまとめたものであり、これらは定常状態仮定に従って同一のマテリアルフローを有するはずである（例えば、解糖またはクエン酸回路のほんのわずかな中間体しか明示的に列挙していない）。

10

【0214】

バイオマス平衡ならびに生細胞密度および全細胞密度から単位mol/lへの変換をNolan(2011)による上記の刊行物から採用した。標的タンパク質のアミノ酸組成の結果として、産物形成のための化学量論の定式化が後から付いてくる。

【0215】

反応の可逆性に関しても上記の刊行物を調べた。

【0216】

上記の表に示した代謝ネットワークについて化学量論行列Aを定式化した。既知の（この場合、細胞外）マテリアルフローに対応する列を組み合わせる部分行列A_mを形成し、残りを組み合わせる部分行列A_uを形成した。

20

【0217】

代謝ネットワークの特徴付けは、付属書に図4.2として示し、説明したスキームに従って実施され得る。

【0218】

代謝ネットワークは検証され、必要であれば、付属書に記載のように変更され得る。

【0219】

従って、図4に示したようにCHO細胞の代謝モデル402が提供されている。この代謝モデルは、様々な細胞内フラックス410および細胞外フラックス408を含み、細胞内代謝産物406と細胞外代謝産物404との間に少なくとも1つの化学量論的關係を指定する。

30

【0220】

以下の工程106~112は、バイオリアクター内での細胞培養物の培養中の複数の時点にわたって行われる。実際に測定した細胞外マテリアルフローと、(規定された長さ、例えば、24時間の間隔の後に)次の時点について予測された細胞外フラックスのプロファイルが作成され得る。これらの2つのプロファイルの互いからの逸脱は予測の質を示している。

【0221】

工程106: 測定値を受け取る工程

一態様では、試料は、細胞培養物のバイオリアクター208の中で、細胞培養物の培養中のいくつかの時点で採取され、図2に示したように1つまたは複数の分析装置250に自動的に、または手作業で移される。分析装置は、1つまたは複数の分析器のシステム、例えば、Thomasチャンバーまたは細胞密度を求める光学的計数ステーションでもよい。例えば、高速液体クロマトグラフまたは当技術分野において公知の他の適切な方法が個々のアミノ酸の濃度を求めるのに使用され得る。試料は、例えば、24時間間隔で採取されてもよい。このように得られた測定データはデータ処理システム252に伝送される。データ処理システム252は、例えば、制御装置として1つまたは複数のバイオリアクターをモニタリングおよび/または制御するコンピュータでもよい。

40

【0222】

一部の態様によれば、測定値の少なくとも一部、例えば、細胞密度は、バイオリアクター

50

208それ自体の対応するセンサーでも求められ、データ処理システム252に伝送される。

【0223】

将来の時点での、この細胞外代謝産物または別の細胞外代謝産物の濃度およびフラックスに関して、ある特定の予測能力を有することが分かっているか、または期待される選択された細胞外代謝産物の濃度に加えて、他の入力パラメータ値、特に、現在の時間、現在の細胞密度、適宜、細胞密度測定に含められない溶解細胞に対する補正係数として使用され得る他のパラメータ、例えば、LDH濃度も求められることがある。決められた時点で、このように経験的に得られた測定データは、以下の工程に記載のように、MLPによる、後の時点、例えば、翌日での細胞外フラックス予測に用いられる場合がある。

【0224】

工程108:トレーニング済みMLPへの測定値の入力

データ処理システム252は、決められた時点で測定された入力パラメータ値（特に、細胞外代謝産物の濃度および細胞密度）に基づいて代謝モデル254の1種または複数種の細胞外フラックスを予測または推定するためにトレーニングされたことのあるMLP、例えば、ニューラルネットワーク(NN)またはいくつかのニューラルネットワークの協働システムを含む。例えば、データ処理システム252は、入力として、ある時点で得られた測定データを、将来の時点（例えば、翌日）での代謝状態を予測しようとする細胞培養物の細胞と同じタイプの細胞の細胞培養物から得られた試験データセットに対してトレーニングされたMLPに自動的に渡すプログラムロジックを含んでもよい。

【0225】

工程110:MLPの将来の取り入れ速度および放出速度の予測

ニューラルネットワークを使用して、時点 $t^{(n)}$ で現在測定されている細胞外代謝産物濃度

$$\bar{c}_{m,j}^{(n)}$$

に基づいて、一段階予測において、自由に選択された将来の時点の時点 $t^{(n+1)}$ での細胞外フラックスが予測または評価される。従って、この工程では、測定された入力パラメータ値がMLPに入力されたことに応じて、MLPは将来の時点の1種または複数種の細胞外フラックスを計算し、戻す。

【0226】

任意:細胞外代謝産物の濃度のMLP予測

ニューラルネットワークを介して推定された細胞外フラックスはまた、本発明の態様に従って、その時点 $t^{(n)}$ での現在の濃度

$$\bar{c}_{m,j}^{(n)}$$

に基づいて、任意に選択された将来の時点 $t^{(n+1)}$ での代謝産物濃度を一段階予測するのにも使用される。

【0227】

この目的で、付属書の式(4.1)を c_2 に従って解いた。将来の時点での生細胞密度 $VCD^{(n+1)}$ が未知であるので、現在測定されている細胞密度、必要であれば、補正された細胞密度によって交換される。再定式化された式は以下の通りである。

$$\bar{c}_{m,j}^{(n+1)} = \bar{v}_{m,j}^{(n)}(t^{(n+1)} - t^{(n)}) \frac{VCD^{(n)}(V^{(n)} + V^{(n+1)})}{2V^{(n+1)}} + \frac{\bar{c}_{m,j}^{(n)}V^{(n)} + c_{zu,m,j}^{(n)}V_{zu}^{(n)}}{V^{(n+1)}} \quad (4.4)$$

【0228】

供給がバイオリアクターを動作させている間に行われれば、好ましくは、

10

20

30

40

50

$$c_{zu_{m,j}}^{(n)}, V_{zu}^{(n)} \text{ および } V^{(n+1)}$$

において考慮すべきである。

【0229】

工程112:MFAの実行

次に、将来の時点についてMLPによって予測された細胞外代謝産物の取り込みと放出速度（細胞外フラックス）に基づいて、代謝フラックス解析が本発明の態様に従って行われる。これも、代謝モデルにおいて定式化されたように細胞内フラックスと化学量論式を組み込んでいる。この代謝モデルでは細胞外フラックスと細胞内フラックスは1種または複数種の細胞内代謝産物を介して互いにつながられているので、細胞外フラックス予測値に基づいて将来の時点での細胞内フラックスを予測することも数学的に可能である。代謝フラックス解析を行うための、対応するプログラムルーチンはMatlabおよび市販されている他のソフトウェアソリューションにおいて実行することができる。

【0230】

細胞内フラックスの予測は、動的経験的データに対してトレーニングされたMLPの予測と、代謝モデルにおいて指定された化学量論的關係および反応式の知識を含むので、この予測工程はハイブリッドモデル關係の予測としても記述され得る。

【0231】

例えば、MLP予測の結果と、マテリアルフロー解析の経過における代謝モデルの情報は以下の通りに結び付けられ得る。

【0232】

ニューラルネットワークを用いて以下の時間間隔の細胞外フラックスを予測した後に、次の時間間隔のフラックス分布を推定するために代謝マテリアルフロー解析が行われる。

【0233】

最初に、代謝モデルのフラックスの共分散行列が推定される。

$$C_{\check{v}_m} = \mathbb{E}[(\mathbf{v}_m - \check{\mathbf{v}}_m)(\mathbf{v}_m - \check{\mathbf{v}}_m)^\top]$$

【0234】

式

$$\begin{aligned} & \mathbb{E}[(\mathbf{v}_m - \check{\mathbf{v}}_m)(\mathbf{v}_m - \check{\mathbf{v}}_m)^\top] \\ &= \mathbb{E}[(\mathbf{v}_m - \bar{\mathbf{v}}_m + \bar{\mathbf{v}}_m - \check{\mathbf{v}}_m)(\mathbf{v}_m - \bar{\mathbf{v}}_m + \bar{\mathbf{v}}_m - \check{\mathbf{v}}_m)^\top] \\ &= \underbrace{\mathbb{E}[(\mathbf{v}_m - \bar{\mathbf{v}}_m)(\mathbf{v}_m - \bar{\mathbf{v}}_m)^\top]}_{=C_{\bar{\mathbf{v}}_m}} - 2\mathbb{E}[(\mathbf{v}_m - \bar{\mathbf{v}}_m)(\check{\mathbf{v}}_m - \bar{\mathbf{v}}_m)^\top] + \mathbb{E}[(\bar{\mathbf{v}}_m - \check{\mathbf{v}}_m)(\bar{\mathbf{v}}_m - \check{\mathbf{v}}_m)^\top] \end{aligned}$$

は、測定値の質と、ニューラルネットワークによる測定値の推定の質が共分散行列に組み込まれていることを示している。しかしながら、リワーディングが推定を容易にしない可能性がある。なぜなら、差

$$(\mathbf{v}_m - \bar{\mathbf{v}}_m) \text{ および } (\check{\mathbf{v}}_m - \bar{\mathbf{v}}_m)$$

は右辺の真ん中の項において独立確率変数と期待値を表さず、従って、右辺の真ん中の項にある期待値は未知だからである。従って、以下で説明される、付属書の式4.5に従う、この論文において使用した式は実際の共分散行列の概算にすぎない。

【0235】

態様によれば、共分散行列は、将来の細胞内フラックスを予測する目的で対角行列として選択される。なぜなら、異なる代謝産物フラックスは別々のネットワークにわたって推定

され、従って、誤差はおおむね互いに独立して考慮され得るからである。定義により、対角成分は、推定されるフラックスの誤差のばらつきを反映しているか、または少なくともこれに比例するはずで（比例因子は、付属書の式2.9に示したようにMFA問題を解くのに役割を果たさない）。

【0236】

対照的に、細胞の現在の代謝状態の記述的代謝物質フラックス解析のために本発明の態様に従うと、共分散行列

$$C_{v_m}$$

が用いられ、対角成分はフラックスの量に依存しない。

10

【0237】

n番目の時間間隔におけるj番目の測定された予測フラックスを

$$\check{v}_{m,j}^{(n)}$$

とする。共分散行列は対角行列として定式化され、構造:

$$C_{\check{v}_m} = \begin{pmatrix} \psi_1 & & 0 \\ & \ddots & \\ 0 & & \psi_{k_m} \end{pmatrix} \quad (4.5)$$

20

を有する。

【0238】

対角成分

$$\psi_j$$

を量

30

$$\{(\bar{v}_{m,j}^{(n)} - \check{v}_{m,j}^{(n)})^2 \mid \text{トレーニングデータセットにおけるn番目の時間間隔}\} \quad (4.6)$$

の中央値として選択した。この態様では、「トレーニングデータセットにおける時間間隔」は、現在の予測に使用される時間間隔と同一であるか、または類似すると想定される。

【0239】

次に、当技術分野においてそれ自体が公知のように、フロー解析が、この共分散行列に基づいて行われる。共分散行列は、付属書の式(2.10)に従ってフラックスを均衡化するのに使用される。付属書の式(2.10)は記述的MFAを指しており、異なる共分散行列を付属書に記載の態様に従って使用した。しかしながら、MFAによるフラックスの均衡化または計算は予測MFAの場合と同様に行われる。

40

【0240】

任意で、例えば、ガウス誤差伝播(Gaussian error propagation)に基づくモデルの誤差解析が付属書のセクション4.7.8に記載のように実施され得る。

【0241】

ニューラルネットワークの作成およびトレーニング

将来のマテリアルフローを予測するための反応速度モデルの仕様と比較して、トレーニング済みMLPの使用には、その作成が、通常、半自動方法では簡単であり、かつ速いという利点がある。トレーニングによってNNの形でMLPがどのように作成され得るかという例の1つを下記で説明する。

50

【0242】

a)いくつかのトレーニング細胞培養物の培養

MLPトレーニングのために、できるだけ広いデータベースを得るために、いくつかのトレーニング培養物を、好ましくは、いくつかのバイオリアクターの中で培養する。好ましくは、これらのバイオリアクターは、1つまたは複数のフェドバッチリアクターと、他のリアクタータイプからの1つまたは複数の追加のバイオリアクターを備える。

【0243】

一態様では、組換えCHO細胞クローンの8回の発酵が1リットルスケールで実施された。初期条件（体積、培地組成物、接種物質濃度）を各バイオリアクターについて全く同じに選択したが、異なる動作モードを使用した。

【0244】

● あるバイオリアクターを、細胞の生存率が50%を割り込むまでバッチモードで動作させた。

● 第2のバイオリアクターでも最初にバッチ方法を使用した。指数増殖期の終わりに近づくにつれて部分回収を行い、リアクターを新鮮な培地で満たした。その結果、（体積および接種物質濃度の点で）発酵の始まりに似た条件を実現した。その後、バッチ方法を継続した（いわゆるスプリットバッチ方法）。

● 残りの6つのリアクター内での発酵を初期バッチ期ではフェドバッチとして行った。発酵の終わりに近づくにつれて連続供給と、パルス様の栄養分添加を両方とも行った。フェドバッチ発酵は培地中のグルコース濃度の点で異なり、グルコース濃度を異なる供給戦略によって調整した。以下の説明では、フェドバッチアプローチに番号が付いていることが多い。この番号に従って、最初の2つのアプローチでは、プロセスの後半に短期間の完全グルコース制限を行った後に、大量投与添加を行った。第3および第4のアプローチには同じ制限があったが、その後の大量投与は、さらに高いグルコース濃度を設定した。第5および第6のアプローチでは、プラスの最小グルコース濃度が常にあった。

【0245】

温度、 pO_2 値、およびpH値を発酵の間ずっと一定に保った。プロセス全体を通して定期的なサンプリングを行った。試料を生細胞密度および全細胞密度に関して、生細胞と死細胞を区別する染色方法を用いて調べた。溶解細胞は記録しなかった。さらに、反応培地中の様々な物質の含有率を、COBAS INTEGRA分析装置または高速液体クロマトグラフィーを用いて調べた。これらは、グルコース、乳酸、およびアンモニウム、ならびにアミノ酸アラニン、グルタミン、グルタミン酸、アスパラギン、アスパラギン酸、セリン、グリシン、ならびに酵素乳酸デヒドロゲナーゼを含んだ。産物濃度も求めた。

【0246】

サンプリング時点で、バイオリアクターの中にある液体の体積を、ファーマンタースケールを用いて測定した。

【0247】

b)ネットワークアーキテクチャの決定

バイオリアクター内の現在の状態から、現在のサンプリング時間と次のサンプリング時間との間の細胞外代謝産物の平均フラックスを推定するために、ニューラルネットワークを作成した（いわゆる一段階予測）。一部の態様によれば、出力変数として、それぞれの細胞外マテリアルフローを得るために別々のネットワークをトレーニングした。現在主流となっている、いろいろな細胞外代謝産物濃度が入力変数として役立った。

【0248】

ニューラルネットワークは、出力層に線形活性化関数があり、隠れ層にシグモイド活性化関数がある2層パーセプトロンからなった（付属書の図2.4を参照されたい）。後者については、シグモイド関数と双曲線正接を両方とも検定した。シグモイド関数を用いて得られた結果に対して有意差は無かった。トレーニングは、（付属書の式(2.16)、(2.17)、および(2.18)に例示されるように）逐次的勾配降下法に従って行った。

【0249】

c)入力変数の選択

本発明の好ましい態様によれば、入力パラメータ値を選択するために、代謝モデルにおいて述べた、いくつかのまたは全ての細胞外代謝産物は、それぞれのフラックスの予測のために関連性に従ってソートされる。好ましくは、冗長な情報内容を有する細胞外代謝産物は考慮されない。

【0250】

本発明の態様に従って入力パラメータ値を選択するための詳細な説明を図13の説明に示した。

【0251】

d)トレーニングパラメータの選択、重みの初期化、およびデータの標準化

イテレーションおよび隠れニューロンの数を決定した後に、それぞれのネットにおける初回学習工程のために出力層に値 $\eta=0.1$ 、隠れ層に $\eta=0.02$ を選択する。総イテレーション数の $1/10$ になるたびに、ステップサイズを初期値の $1/10$ にする。重みの初期値を

$$\left[-\frac{1}{10}, \frac{1}{10}\right]$$

均等に分布する乱数にした。調整された値が経験的平均0と経験的標準偏差0.5を有するように、入力トレーニングデータと出力トレーニングデータを別々に代謝産物によって標準化した。トレーニングデータセットの平均値および標準偏差を用いて試験データを同じように変換した。

【0252】

e)トレーニングデータセットおよび試験データセットの選択

推定は、トレーニングデータセットおよび試験データセットへのデータのグループ化の点で異なる3つの異なるニューラルネットワークによって行った。

1. ネットワーク1:第1のネットワークのトレーニングデータセットは、フェドバッチ発酵およびバッチ発酵のうちの3つからのデータからなり、試験データセットは、他の3つのフェドバッチ発酵およびスプリットバッチ発酵からのデータからなった。

2. ネットワーク2:第2のネットワークは、トレーニングデータセットにフェドバッチ発酵のうちの3つからのデータと、試験データセットにフェドバッチ発酵、バッチ発酵、およびスプリットバッチ発酵のうちの残りの3つからのデータを有した。

3. ネットワーク3:第3のネットワークのトレーニングデータセットは、フェドバッチ発酵のうちの3つからのデータを含み、試験データセットは残りの3つのフェドバッチ発酵からのデータを含んだ。

【0253】

f)細胞外代謝産物濃度の一段階予測

この目的は、細胞代謝の最も正確な予測を可能にするトレーニング済みMLPを作成することである。

【0254】

トレーニングデータセット作製中に低頻度データが作成されたら、データの平均化/フィルタリングによって、あまりにも大きな情報損失が偶然に生じるかもしれない。従って、この場合、2つの連続した測定点間の平均細胞外フラックスを近似しなければならない。この計算は、式

$$\dot{m} = \dot{V}_{zu} \cdot c_{zu} - \dot{V}_{ab} \cdot c_{ab} + Q$$

に基づいており、以下で説明される。

【0255】

バッチプロセスおよびフェドバッチプロセスが考慮され、従って、液体排出は無いので、どんな場合でも、

$$\dot{V}_{ab} = 0$$

が成り立つ。時間 t での、ある成分の細胞外フラックス v は、付属書のセクション2.2.2において既に述べられているように、時間あたりの細胞によって吸収または放出された物質の量である。従って、これは、

$$v(t) := \frac{Q(t)}{VC(t)}$$

によって示される。式中、 $VC(t)$ は、時間 t での反応培地中の生細胞の数である。従って、付属書の式2.1によれば、2つの不連続点間の領域において、

$$v(t) = \frac{\dot{m}(t) - \dot{V}_{zu}(t) \cdot c_{zu}}{VC(t)}$$

10

が成り立つ。

【0256】

供給中の物質の濃度 c_{zu} は経時的に一定であるか、またはほぼ一定であると想定される。

【0257】

不安定さがサンプリング時間で、および大量投与添加中に生じる場合がある。

【0258】

最初に、栄養分の添加は常に連続しており、従って、量 m および V_{zu} は2つの測定点間で区別できると想定されるはずである。2つの連続した測定点 t_1 と t_2 の間で平均フラックス v_{avg} を求めたかったら、最初にみたところささいなものでも、

20

$$v_{avg} = \frac{\frac{\Delta m}{\Delta t} - \frac{\Delta V}{\Delta t} c_{zu}}{VC_{avg}} = \frac{m_2 - m_1 - m_{zu}}{(t_2 - t_1) \frac{VC_1 + VC_2}{2}} = \frac{c_2 V_2 - c_1 V_1 - c_{zu} V_{zu}}{(t_2 - t_1) \frac{VCD_1 V_1 + VCD_2 V_2}{2}} \quad (4.1)$$

によって推定し得る。

【0259】

ここでは、指数の付いた変数は時点 t_1 または t_2 での値を示し、 m_{zu} は、考慮された時間間隔での供給を介して反応培地に供給された物質の量であり、 VCD は生細胞密度を示す。演算子 Δ は、 t_1 と t_2 の間の対応する量の差を表す。

30

【0260】

上記の推定(4.1)は以下において数学的に証明されるものとする。これは時間間隔 (t_1, t_2) にわたって積分することによって得られる。個々の測定された量は、2つのサンプリング時間の間に、かつさらなるテイラー展開によって一次補間される。

【0261】

一次補間を用いて、

$$\dot{m}(t) = \frac{m_2 - m_1}{t_2 - t_1}$$

40

$$\dot{V}_{zu} = \frac{V_2 - V_1}{t_2 - t_1}$$

$$VC(t) = VC_1 + (VC_2 - VC_1) \frac{t - t_1}{t_2 - t_1} = \frac{(t_2 - t) VC_1 + (t - t_1) VC_2}{t_2 - t_1}$$

が、全ての $t \in (t_1, t_2)$ において成り立つ。従って、

50

$$\begin{aligned}
v_{\text{avg}} &= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{Q(t)}{VC(t)} dt \\
&= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{\dot{m}(t) - \dot{V}_{zu}(t) \cdot c_{zu}}{VC(t)} dt \\
&= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{\frac{m_2 - m_1}{t_2 - t_1} - c_{zu} \frac{V_2 - V_1}{t_2 - t_1}}{VC(t)} dt \\
&= \frac{m_2 - m_1 - c_{zu}(V_2 - V_1)}{t_2 - t_1} \cdot \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{1}{VC(t)} dt \\
&= \frac{m_2 - m_1 - m_{zu}}{t_2 - t_1} \cdot \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{1}{VC(t)} dt
\end{aligned}$$

【 0 2 6 2 】

残りの積分においては、

$$\begin{aligned}
\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{1}{VC(t)} dt &= \int_{t_1}^{t_2} \frac{1}{(t_2 - t)VC_1 + (t - t_1)VC_2} dt \\
&= \int_{t_1}^{t_2} \frac{1}{t_2 VC_1 - t_1 VC_2 + t(VC_2 - VC_1)} dt \\
&= \frac{1}{VC_2 - VC_1} [\ln |t_2 VC_1 - t_1 VC_2 + t(VC_2 - VC_1)|]_{t_1}^{t_2} \\
&= \frac{\ln((t_2 - t_1)VC_2) - \ln((t_2 - t_1)VC_1)}{VC_2 - VC_1} \\
&= \frac{\ln(VC_2) - \ln(VC_1)}{VC_2 - VC_1}
\end{aligned}$$

が成り立つ。

【 0 2 6 3 】

一次までの

$$\frac{VC_1 + VC_2}{2}$$

による対数のテイラー展開が行われる。これにより、

$$\begin{aligned}
\frac{\ln(VC_2) - \ln(VC_1)}{VC_2 - VC_1} &= \frac{\ln(\frac{VC_1 + VC_2}{2} + \frac{VC_2 - VC_1}{2}) - \ln(\frac{VC_1 + VC_2}{2} + \frac{VC_1 - VC_2}{2})}{VC_1 - VC_2} \\
&\approx \frac{\ln(\frac{VC_1 + VC_2}{2}) + \frac{VC_2 - VC_1}{2} \frac{1}{\frac{VC_1 + VC_2}{2}} - \ln(\frac{VC_1 + VC_2}{2}) - \frac{VC_1 - VC_2}{2} \frac{1}{\frac{VC_1 + VC_2}{2}}}{VC_1 - VC_2} \\
&= \frac{\frac{2VC_2 - 2VC_1}{VC_1 + VC_2}}{VC_1 - VC_2} \\
&= \frac{1}{\frac{VC_1 + VC_2}{2}}
\end{aligned}$$

が得られる。

【0264】

上記の積分について、この式を使用すると、上記の式4.1または付属書に従ってフラックスが推定される。

【0265】

2つのサンプリング時点 t_1 と t_2 との間に、検討されている物質を含有する大量投与が時間 t_B で添加された場合、 t_1 と t_B の間および t_B と t_2 の間の平均フラックスは同じであると想定される。次に、この場合にも式(4.1)が成り立ち得ることが簡単に示され得る。大量投与を介して添加された物質の量は式 m_{zu} に含まれる。

【0266】

従って、トレーニングバイオリアクターからの現在のサンプリングの時点で、以前のサンプリングにおける細胞外代謝産物濃度と、最後のサンプリングから計算した細胞外フラックスに基づいて計算した細胞外フラックス計算値は両方とも既知であり、トレーニングしようとするMLPに参照値量として一緒に渡され得る。これによって、MLPは、最後のサンプリングにおいて計算した細胞外代謝産物濃度に基づいて、細胞外フラックス計算値からの逸脱ができるだけ最小になるように細胞外代謝産物計算値を予測するように学習する。

【0267】

トレーニング済みMLPまたはトレーニング済みニューラルネットワークは記憶され、その時点 $t(n)$ での現在の濃度

$\hat{c}_{m,j}^{(n)}$

に基づいて、任意の選択された将来の時点 $t(n+1)$ 、例えば、翌日の細胞外代謝産物濃度を一段階予測するために使用され得る。

【0268】

要約すると、ニューロンネットワークをトレーニングするという考えは、細胞外代謝産物濃度を簡単に測定することができ、このことから、少なくとも振り返ると細胞外フラックスを経験的に求めることができる事実に基づいている。細胞外代謝産物の濃度差に基づいて、この現在の時点と将来の時点との間の時間間隔にわたる細胞外フラックスを出力パラメータ値として計算することができる場合、入力パラメータ値として、決められた時点で測定された細胞外代謝産物濃度と、細胞外フラックスを使用することによって、ニューラルネットワークまたは他の機械学習アルゴリズムは、少なくとも、ある将来の時点の細胞外フラックスを予測するようにトレーニングされ得る。細胞密度を求めることによって、培地中の全濃度差を、培地に含まれる細胞培養物の個々の細胞に割り当てることが可能になる。

【0269】

細胞外代謝産物の取り込み速度/排出速度（反応項）は、好ましい態様に従って、バイオリアクター内にある物質-バイオリアクターから添加/除去された物質の全濃度変化の差（対流項）から計算される。細胞外代謝産物の反応項は主として細胞内への、または細胞を通る取り込みまたは放出から求められ、その結果、細胞外代謝産物の濃度変化測定値は、細胞内への、または細胞からの、これらの細胞外代謝産物の取り込み速度または放出速度と本質的に等しくなり得ることが見出された。しかしながら、ある特定の細胞外代謝産物がバイオリアクター動作中に多量に供給または除去される一部の態様では、これらの細胞外代謝産物の取り込み速度または除去速度を計算する時に、培養培地への、または培養培地からの細胞外代謝産物の外部供給または除去による濃度変化の部分を濃度変化測定値から差し引くことによって補正計算が行われ得る。しかしながら、この補正は、フェドバッチプロセスでも必ず行わなければならないとは限らない。フェドバッチプロセスでも流入フローおよび流出フローは連続している。発酵中に大量投与供給またはサンプリングが行われる場合、長距離にわたって連続しており、従って、微分可能な曲線が得られる。このことは、フェドバッチリアクターについても微分可能な濃度変化経過があるという仮定を正当化する。

【0270】

しかしながら、トレーニング中に、および/またはトレーニング済みMLPを用いた細胞外フラックスの予測において細胞密度測定値を補正することが好ましい。バッチ発酵を除いて、発酵アプローチにおいてLDH濃度と細胞密度の差の間には類似するほぼ直線的な関係があるように思われる。

【0271】

本発明のデザインによれば、ファーマンターの全細胞密度測定値が連続して記録され、グラフに示される。並行して、細胞密度が計算される。予測は、例えば、さらなる出力パラメータ値として細胞濃度測定値を用いて本発明の態様に従ってトレーニングされたトレーニング済みMLPによって計算され得る。この「予測された」細胞密度はグラフにもプロットされる。従って、本発明の態様によれば、ある特定の細胞タイプの細胞培養物の細胞密度測定値の第1の時間プロファイルが経験的に求められ、代謝モデルと細胞外代謝産物濃度に基づいて細胞密度予測値の第2の時間プロファイルも求められる。従って、細胞密度測定値と細胞密度予測値の時間プロファイルと、互いとの逸脱を含むグラフが得られる。

【0272】

細胞密度測定値と、MLPによって測定された細胞密度との間には、特に、細胞培養物の終わりに近づくにつれて、かなりの逸脱があることが多いと示されている。細胞密度測定値と細胞密度予測値との間の相違は以下では「密度相違プロファイル」と呼ばれ、任意でグラフにも図示され得る。密度相違プロファイルは、細胞として測定できないが、依然として細胞外代謝産物濃度に影響を及ぼしている溶解細胞が発生する時間プロファイルを表している。次に、経験的関数が、例えば、2つのパラメータを特徴とする密度相違プロファイルにわたって直線補償線の助けを借りて作成される。これは、決められた時点で測定された培地中のLDH濃度と直線関係にある密度相違プロファイルに従って溶解細胞の密度を設定する。この関数は、LDH濃度から密度差への、従って、溶解細胞への近似的換算を可能にする。従って、補正された細胞密度は、細胞密度測定値と、LDH濃度測定値に基づいて一次関数によって計算した溶解細胞数の和である。言い換えると、細胞密度測定値

$$\bar{c}_{m,11}^{(n)}$$

は、この

$$\bar{c}_{lys,11}^{(n)} := \bar{c}_{m,11}^{(n)} + a_1 \bar{c}_{LDH}^{(n)} + a'_0$$

を用いた密度差によって補足される。式中、

$$\bar{c}_{LDH}^{(n)}$$

n番目の時点でのLDH濃度が指定され、 n_0 が発酵を開始した時点であれば、

$$\bar{c}_{lys,11}^{(n_0)} = \bar{c}_{m,11}^{(n_0)}$$

が成り立つように、 a'_0 が選択される。

【0273】

これらの補正された細胞密度値は、好ましい実行形式に従って、補正されたバイオマスフラックス

$$\bar{v}_{lys,11}^{(n)}$$

を計算するために使用される。これには、MFAの結果としての、一致しないフラックス分布の割合が大幅に低下するという利点がある。

【0274】

図2は、異なる装置およびデータソースを用いた、いくつかの段階での情報取得プロセスの一例を示す。

【0275】

将来の時点での代謝状態を予測しようとする細胞は、バイオリアクター208、例えば、フェドバッチファーマンターの中に保持される。ファーマンターは、いくつかのセンサーを備えてもよいし、いくつかのセンサー（例えば、細胞密度を求めるセンサー）に動作可能に接続されていてもよい。これらのセンサーの代わりに、またはこれらのセンサーに加えて、試料が細胞培養物から定期的に採取され、1つまたは複数の分析機器250に移され得る。そこで細胞外代謝産物の濃度が測定される。細胞密度測定値と、細胞外代謝産物濃度測定値と、適宜、代謝産物の外部供給（例えば、グルコースボラス投与）の時点および量がデータ処理システム252に送られる。このシステム252は、細胞の代謝モデルと、現在使用されている細胞タイプの現在測定されている細胞外代謝産物濃度に基づいて細胞外フラックスを予測するようにトレーニングされたMLPを備える。決められた時点で受け取られ測定された測定値はトレーニング済みMLPへの入力として伝送され、次に、トレーニング済みMLPは将来の時点での細胞外フラックスを予測する。次のMFAの間に、細胞外フラックスと、このモデルの化学量論式は将来の時点の細胞内フラックスを予測するのにも使用される。予測された細胞外フラックスと細胞内フラックスを含む完全なモデルはグラフィカルユーザーインターフェイスを介して細胞の代謝状態の「スナップショット」画像254として表示および/または記憶され得る。

【0276】

バイオリアクター208が、トレーニング細胞培養物（すなわち、トレーニングデータセットを作成するためにデータが長期間にわたって定期的に収集される細胞培養物）を含む場合、データ処理システム252は、細胞外代謝産物の濃度測定値のグラフに基づいて、将来の時間間隔（すなわち、現在の時点から、代謝状態の予測が行われる次の時点までの時間間隔）の細胞外フラックスを計算し、トレーニング中の出力パラメータ値として、これをMLPに渡すように、さらに適合される。

【0277】

図3は、1つまたは複数のバイオリアクターをモニタリングおよび/または制御するのに使用され得る、細胞の代謝状態を予測するためのシステムのブロック図を示す。

【0278】

システム200は様々なタイプのデータ処理システムでもよい。例えば、システム200は、デスクトップコンピュータ、サーバーコンピュータ、ノートブックまたはユーザーの携帯型モバイル装置でもよい。システム200は、バイオリアクター、または複数のバイオリアクターを備えるバイオリアクタープラントの一部であるか、またはこれに接続された制御モジュールでもよい。ここで示した態様では、前記システムは3つのバイオリアクター204、206、208につなげられ、それぞれの細胞培養物における細胞の代謝状態をリアルタイムでモニタリングし、将来の時点、例えば、将来の12時間または24時間の時点を予測するように構成されている。バイオリアクターはそれぞれ、細胞密度および/もしくは代謝産物濃度を求めるための1つまたは複数の測定装置もしくはセンサー、または代謝産物濃度が他の装置によって試料分析によって求められるように試料の命名を可能にする機構を有してもよい。好ましくは、204~208のバイオリアクターは、前記システムにつなげられ、必要であれば前記システムから制御コマンドを受け取って実行し得る様々な制御装置、例えば、バルブ、ポンプ、大量投与用の投与装置、スターラーなどを有する。

【0279】

システム200は、1つまたは複数のプロセッサ202ならびに1つまたは複数のバイオリアクターから測定データを受け取るための第1のインターフェイス210を備える。インターフェイス210は、バイオリアクターへのダイレクトインターフェイスとして、またはバイオリアクターからの試料を分析する分析装置へのインターフェイスとして、またはユーザーが、取得した測定データを手作業で、もしくは他の手段によって入力することを可能にす

るグラフィカルユーザーインターフェイスへのインターフェイスとして適合されてもよい。

【0280】

さらに、前記システム201は、揮発性もしくは非揮発性の記憶媒体212を備えるか、またはこれにつながられる。記憶媒体は、例えば、メインメモリ、ハードディスク、またはクラウドサービスのメモリ、またはネットワークストレージ、または上記のタイプのストレージの組み合わせでもよい。記憶媒体は、バイオリアクター内で保持および増殖された細胞の代謝モデル214、例えば、図4に示したモデルを備える。

【0281】

記憶媒体は、将来の時点で受け取ったいくつかの細胞外代謝産物の濃度測定値から、将来の時点での1種または複数種の細胞外フラックスを予測するように訓練された、トレーニング済みMLP 218、例えば、トレーニング済みニューラルネットワークを含む。

10

【0282】

さらに、記憶媒体は、細胞外フラックスの予測を行うために、受け取った測定値をMLP 218に渡すように訓練されたプログラムロジック220を含む。さらに、プログラムロジックは、将来の時点の細胞内フラックスを予測するために、代謝モデル214と細胞外フラックス予測値に基づいてリアルタイム代謝フラックス解析(MFA)を行うように訓練されている。プログラムロジック220は任意のプログラミング言語、例えば、C++、Java、Matlabで実行されてもよく、互いに相互運用可能な、異なるプログラミング言語または同じプログラミング言語で、いくつかのプログラムモジュールの形で実行されてもよい。

【0283】

任意で、記憶媒体は、いくつかの参照値および参照値範囲を含んでもよい。これらの参照値または参照値範囲は異なる細胞内代謝産物の許容可能な、または望ましい細胞内フラックスを示す。細胞内フラックス予測値と参照値216とのリアルタイム比較によって、プログラムロジック220はバイオリアクターの1つまたは複数の中にある細胞が望ましくない代謝状態に向かっているかどうか検出し、必要であれば、予測された傾向を打ち消すために、第2のインターフェイス222を介して適切な制御コマンドを、それぞれのリアクターに発行することによって対策を取ることがある。代わりに、またはさらに、この場合、ユーザーインターフェイス224、例えば、ディスプレイ装置、例えば、LCDディスプレイを介して警告がユーザーに出されることがある。ディスプレイ装置は、ユーザーに、予測された細胞外フラックスおよび細胞内フラックスを知らせ、望ましい参照範囲からの、これらのフラックスのあらゆる予測された逸脱も知らせる。

20

【0284】

図5は、バイオリアクターを動作させている間の、いくつかの連続した時点での細胞内フラックスの計算を示す。上部のグラフ502は、6つの測定点(1日につき1回の測定)で測定された細胞外代謝産物の濃度のプロファイルを示す。真ん中のグラフ504は、トレーニング済みMLPが、これらの測定値を使用して将来の時点の1種または複数種の細胞外フラックスを予測することを示す。上部および真ん中のグラフにある点の位置を比較すると、測定データを収集した時点と、細胞外フラックスを予測した時点は約半日離れていることが分かる。このことは、例えば、細胞外代謝産物の濃度を12時正午に毎日測定したら、これらのデータを使用して夜の12時に細胞外フラックスを予測することを意味する。下部のグラフ506は、これらの将来の時点のそれぞれでMFAを行うことによって、細胞外フラックス予測値が細胞内フラックス予測値によって補足されたことを示す。

30

40

【0285】

図6は、図4に示した代謝モデルに表した異なるフラックスを示す。これらのフラックスは、ある時間間隔にわたる代謝産物濃度変化の測定値と細胞密度測定値に基づいて計算され、異なる代謝産物のフラックスが非常に異なる、かつ部分的に特徴的な経過を有することを示す。

【0286】

図7は、生物学的知識を生み出すための方法のうまくいった使用を示す。左上にあるグルコースフラックスのグラフは、約0.65の時点でのグルコースフラックスのプロファイルに

50

においてグルコースフローは実質的に停止することを示す。このグルコース欠乏の影響は、アラニン、セリン、およびグリシンについて観察することができる。すなわち、アラニンは制限の場合に高速で吸収される。これは、おそらく、もっと多くのピルビン酸を供給するのに役立つ。さらに、セリンからグリシンへの強力な変換が起こり、これはアンモニウム形成の増加と関連する。

【0287】

グルコース曲線および産物フロー曲線を比較すると、ある特定の類似性が明らかになる。すなわち、両曲線とも、発酵の初期の時点でのフラックスの一時的減少と、グルコースが培地から一時的に無くなった、もっと後の崩壊を示す。この一時的低下は、グルコース制限のないファーメンターの産物フローでは認められない。同様に、培地中のグルコース濃度に対するグルコースフラックスの感度を示すグルコースボーラス投与の影響も産物形成において認められる。明らかに、これらのフラックス間には非常に密接した結び付きがある。発酵プロセスの終わりに、産物中にある物質の量は、グルコース制限の無いバイオリアクターで最も多い。従って、グルコースの利用可能性は有効な産物形成に必要な不可欠なように思われ、不足は厳しく回避しなければならない。

【0288】

図8は、生物学的知識を生み出すための方法の別のうまくいった使用を示す。

【0289】

いわゆる乳酸シフトは、乳酸フローがプラスからマイナスに符号を変える、細胞培養物の培養において頻繁に観察される作用を指す。文献には乳酸シフトを説明する非常に多くの試みがある。Mulukutlaらは、例えば、乳酸シフトが、乳酸阻害の増加によって始動する調節機構の結果であると推論している。この生物学的仮説は、本発明に従う方法を用いて、いくつかの測定値にわたって、いくつかのバイオリアクターの中の乳酸フラックスを確かめることによって、および細胞外乳酸濃度を測定することによって試験された。4つのバイオリアクターの対応する結果を図9に示した。乳酸の正味のフロー方向の逆転が観察された時点で、異なるバイオリアクター内の乳酸濃度が異なることが示された。従って、この時点まで文献において推論された機構は乳酸シフトを担っていないように思われる。もっと正確に言うと、補遺の図6.2に示したように、このシフトはグルタミン代謝に依存すると思われる。いくつかの異なるファーメンタータイプにおけるグルタミン濃度と乳酸フローの経過を比較すると、乳酸シフトは常に完全なグルタミン消費を伴うことが分かる。

【0290】

この観察はまた、決められたCHOクローンについて述べた文献でも報告されており、この場合、グルコースが消費された後だけシフトが起こった(Zagari F, Jordan M, Stettler M, Broly H, Wurm FM (2013): 「Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity」, New Biotechnology, Vol.30, No. 2)。従って、本発明によれば、細胞内フラックスを繰り返し予測し、これらのフラックスを他の細胞内フラックスおよび/または細胞外代謝産物の濃度と比較することによって細胞代謝に関する仮説を首尾良く検定し、必要であれば細胞代謝に関する仮説を棄却し、個々の細胞クローンの代謝特性を特定することができる。この分析結果に基づけば、一般的なCHOクローンにおいて形成される乳酸の大部分がグルタミノリシスに由来することは明らかである。興味深いことに、乳酸シフトの後、グルタミノリシスは初め停止している。すなわち、これ以上グルタミンは取り込まれない。もっと正確に言うと少量のグルタミンが形成され、そのためバイオリアクター内での濃度は上昇する。細胞は基質グルコースに限定して代謝を調整したように思われる。グルコースが非常に低い濃度に達したらすぐに、第2のグルタミン取り込み期が観察され得ることに注目すべきである。同時に、短期の新たな乳酸産生も観察することができる。

【0291】

図9は、異なるタイプの4つのバイオリアクターの乳酸フラックスとグルタミン濃度を用いたグラフを示す。これにより、決められた細胞培養物の中でのグルタミン濃度と乳酸フロ

一の経過を比較することが可能になる。グルタミン濃度を実線で示し、乳酸フラックスを点線で示した。比較性を高めるために乳酸フラックスと、個々のフラックスおよび代謝産物の濃度を前もって基準化した。矢印は、乳酸フラックスの徴候の逆転を示している。

【0292】

図10は、現在の時点について代謝モデルに基づいて、MFAを用いて記述的モデリングによって計算した細胞内フラックスの時間経過を示す（すなわち、記述的であり、予測ではない）。

【0293】

付属書のセクション2.2.2には、冗長な代謝モデルを処理するための異なる方法を示す。フラックスを計算するために、最適化問題（付属書の式2.9）を解き、従って、加重最小二乗解を得た。これにより、最適化問題を解くことによって得られた細胞外フラックスは、実験データから直接計算したフラックスと相違した。

【0294】

付属書の式(2.10)を使用して、均衡化された細胞外フラックスを計算し、細胞内フラックスを付属書の式(2.5)、この場合も

$$\hat{\mathbf{v}}_m = (\mathbf{I} - \mathbf{C}_{\mathbf{v}_m} \mathbf{R}'^T (\mathbf{R}' \mathbf{C}_{\mathbf{v}_m} \mathbf{R}'^T)^{-1} \mathbf{R}') \bar{\mathbf{v}}_m$$

および

$$\hat{\mathbf{v}}_u = -\mathbf{A}_u^{\#} \mathbf{A}_m \hat{\mathbf{v}}_m$$

によって求めた。

【0295】

共分散行列

$$\mathbf{C}_{\mathbf{v}_m}$$

を定式化するために、好ましくは、モデル検証の結果を含める。これにより、どの標準偏差が、おおむね一貫したモデルにつながるか、場合によっては、どの測定値が信頼できないと分類すべきかが分かった。

【0296】

その後、フラックス計算値を視覚化した。一方では、時間経過を代謝産物に従って別々にプロットし、他方では、選択された時点での全体のフラックス分布を視覚化した。

【0297】

図11は、CHO細胞の細胞培養物の培養中の異なる時点での細胞内フラックスおよび細胞外フラックスのスナップショットを示す。

【0298】

経時的な個々のマテリアルフローの経過に加えて、個々の時間間隔での全体の細胞内フラックス分布および細胞外フラックス分布のスナップショットが考慮されることがある。図11は、第4のフェドバッチ発酵からの細胞の4つの代謝状態の一例を示す。これらは第2、第11、第14、および第21の時間間隔に由来する。4つの図は、フェドバッチ発酵中のフラックス分布を示している。第2の時間間隔でのグルコース(gluc39ose)フローは参照として役立つ。産物フローに10,000を掛けた。

【0299】

第2の時間間隔では、すなわち、初期発酵期(図11Aに示した区分の第I期)ではグルコースは依然として過剰であり、急速に細胞内に輸送され、クエン酸回路に入る。グルタミンも取り入れられ、グルタミン酸への変換が行われ、さらに、フラックス v_8 を介してピルビン酸と一緒にクエン酸回路の代謝産物とアラニンになる。乳酸とアンモニアは培地中に多量に放出される。

【0300】

図11Bに示した第11の時間間隔（第III期）では、乳酸シフトが既に生じている。すなわち、乳酸が培地に吸収され、ピルビン酸に代謝され、次にクエン酸回路に入る。グルコース

10

20

30

40

50

消費は低下し、グルタミンはもはや吸収されない。アンモニウムはもはや放出されない。バイオマス産生は減少しているが、産物形成は増加している。

【0301】

発酵開始と比較して、クエン酸回路はほとんど変わらない強度で回転するが、リンゴ酸からピルビン酸への補充反応は低下している。

【0302】

図11Cに示した14番目の時間間隔（これも第III期）では培地中の乳酸濃度は非常に低く、その結果、取り入れも制限される。従って、アラニンが吸収され、ピルビン酸に変換される。アンモニウムは、ますます培地に再放出される。産物形成はさらにもっと増加している。

【0303】

図11Dに示した発酵の終わりに近づくにつれて（第IV期）、グルコースが大量投与のようなやり方で添加される。従って、取り込みは再増加し、クエン酸回路を通り抜けるフラックスは特に強い。しかしながら、他の全ての反応は極端に低下している。

【0304】

図12は、記述的MFAを用いて現在の時点について計算した細胞内フラックスと、MLPおよびMFAの組み合わせを用いて将来の時点について予測した細胞内フラックスの強い相関を持つ経過を示す。「現在の時点を対象にした記述的MFA」では、たった今経過した時間間隔にわたる細胞外代謝産物のフラックス測定値から代謝モデルの細胞外フラックスを計算し、これらの細胞外フラックスをMFAの入力として使用して現在の時点の細胞内フラックスを計算する。対照的に、細胞内フラックスの「予測」決定は、トレーニングされたMLP予測を使用して将来の時点の細胞外フラックスを予測し、このように予測された細胞外フラックスをNSAの入力として使用して将来の時点の細胞内フラックスを予測することに基づいている。

【0305】

図13は、細胞外フラックスを予測するために入力値として濃度が使用される細胞外代謝産物の一部を示す。

【0306】

トレーニング済みMLPによって1回予測され、トレーニング中に出力パラメータ値として渡された、それぞれの細胞外代謝産物フラックス（表の一番上の行に示した）について、下にある縦列は、「出力代謝産物」フラックスを推定するために入力代謝産物を関連性の上位から順に列挙している。ここでの「Bio」は、バイオマス測定値、好ましくは、細胞密度の点で指定したバイオマス測定値を意味する。「TZD」は、LDH濃度について補正した値を意味する。この関連性を求めるために行ったPMI計算は、好ましくは、実施した全てのトレーニング発酵からの値に基づいている。

【0307】

この表はまた、入力変数の数、隠れ変数Hの数、およびイテレーションの数を決定した交差検証の結果も含んでいる。濃度が入力パラメータ値としてニューラルネットワークに入力される入力代謝産物1504を黄色で強調した。

【0308】

いくつかの(例えば、3つの)フェドバッチ発酵からの値ならびに1つのバッチアプローチからの値をトレーニングデータセットとして使用した。同じ細胞タイプを用いた他のファーマンターからのデータが試験データセットを形成した。さらに、他のトレーニング/試験データセット分割を用いた交差検証が行われる場合がある。

【0309】

しかしながら、この表に示したリストは、必ず、生物学的観点から直感で分かるような順序に対応するとは限らない。例えば、グルコース濃度はここでは補助的な役割を果たしているが、CHO細胞の最も重要な基質が、いくつかの物質の結果に大きな影響を及ぼすことは明らかであろう。しかしながら、下記のPMIに基づく配置には、限られた生物学的重要性しかない。これは、以下の事実に起因するかもしれない。すなわち、代謝産物濃度の経

10

20

30

40

50

過は、ある程度まで代謝を介して相関付けられる。従って、ある代謝産物を選択した後、他の多くの代謝産物が、推定のための関連性を失う可能性がある。なぜなら、これらに含まれる情報の多くが第1の代謝産物によって既に記述されているからである。しかしながら、選択された代謝産物が、機構上の観点から、実際に物質フローに影響を及ぼすものでない可能性があり、単に、強い相関を持っているのにすぎない。実際には、PMI値を計算するのに使用した、データ選択における白色変化の試料は、場合によっては、異なる配置の原因となったことが観察されたことがある。場合によっては、例えば、代謝を介して密接に関連するグルタミン酸およびグルタミンの位置は逆になった。しかしながら、生物学的直感に基づく選択は、いくつかの刊行物において現在行われているように、かなり悪い予測をもたらす時もあった。これは、生物学的関係が、いつも分かっているわけではないことが多く、冗長な情報が選択されるという事実によるものかもしれない。さらに、推定された生物学的関連性に基づいて、適切な入力パラメータ値を選択することを目的とした文献研究は多くの時間を必要とする。

【0310】

最初に、それぞれの代謝産物フローについて、可能性のある入力を、部分的相互情報量(PMI)を用いて関連性に従ってソートした。付属書の式2.19を参照されたい。PMIを、付属書の式2.22に従って離散化を用いて計算した。使用したコア密度推定量は都市ブロック関数に基づいており、ナダラヤ・ワトソン推定量(Nadaraya-Watson estimator) (付属書の式2.27を参照されたい)を使用して残差を計算した。このためにトレーニングデータセットを使用した。本件と同様に、これは8つの異なるバイオリアクター内での8つの細胞培養計画の測定データを含んでいる可能性がある。

【0311】

j番目の細胞外代謝産物フラックスを推定するための入力の命令を以下のアルゴリズムに従って行った。

1. セット χ に、可能性のある全ての入力変数、ここでは、全ての現在の細胞外代謝産物濃度をまとめる。このセット u は、既に選択されている全ての入力変数を含んでいる（最初は空である）。 Y は、現在のサンプリング時間と将来のサンプリング時間との間の出力変数、すなわち、j番目の細胞外代謝産物フラックスである。
2. 所定の標準化されたデータセットに基づいて、 u における要素を考慮して、 χ と Y にある可能性のある、それぞれの入力変数間の部分トランスインフォメーションのために近似値を計算する。例えば、標準化により、全ての入力変数と出力変数の平均値は0となり、標準偏差は0.5となり、従って、異なる桁のせいで変数の関連性をゆがめる効果を取り除かれる。好ましくは、この標準化は、トレーニングデータセットの測定値に関してニューラルネットワークをトレーニングする前に、ならびに現在得られた測定値に関してトレーニング済みMLPを用いて現在の測定値を検定手順に入力する時に行われる。
3. 最も高いPMI値を有する、 χ にある変数を記録する。これを u に追加し、 χ から除去する。
4. χ が空になるまで工程2および3を繰り返す。

【0312】

本発明の態様によれば、パラメータ X 、例えば、決められた細胞外代謝産物のPMIを、以下のように、パラメータ Y 、例えば、別の細胞外代謝産物に関して計算する。

$$I'(X, Y) := \int \int g_{x', y'}(x', y') \ln \left(\frac{g_{x', y'}(x', y')}{g_{x'}(x') g_{y'}(y')} \right) dx' dy' \quad (2.19)$$

残差

$$x' := x - \mathbb{E}[x|U] \quad (2.20)$$

および

$$y' := y - \mathbb{E}[y|U] \quad (2.21)$$

式中、 g は周辺分布または共通分布の密度関数である。残差は、 U にまだ含まれていない X と Y の情報しか含んでいない。 I' の値が大きければ大きいほど、依存性は強くなる。

【0313】

式2.19の近似的な別個のバージョンは以下の通りである。

$$I'(X, Y) = \frac{1}{N} \sum_{n=1}^N \ln \left(\frac{g_{x', y'}(x^{(n)'}, y^{(n)'})}{g_{x'}(x^{(n)'}) g_{y'}(y^{(n)'})} \right) \quad (2.22)$$

10

$(x^{(n)}, y^{(n)})$ 、 $n=1, \dots, N$ 、 X と Y の標本対

$(x^{(n)'}, y^{(n)'})$ は、関連付けられた $g_{x', y'}$ 分布の残差対である。そして次に、普通知られていない密度関数はコア密度推定量によって近似され得る。これもまた N 個の標本からの情報を使用する。これらの推定量から、簡単に言うと、標本のヒストグラムに似た連続密度関数が得られる。これは N 個のコア関数の重みつき重ね合わせによって生じる。そして次に、これらは、ベル形であり、かつ標本値のうちの1つについて対称的な密度関数である。

【0314】

特に、PMIを計算するために今日まで刊行物ではガウスコアおよび都市ブロック関数が使用されてきた。

20

【0315】

コア関数 κ とバンド幅 μ を用いた、標本 $x^{(1)}, \dots, x^{(N)}$ の q 次元ランダムベクトル X の密度の推定は、

$$\hat{g}_x(x) = \frac{1}{N} \sum_{n=1}^N \mathcal{K}_\mu(x - x^{(n)}) = \frac{1}{N} \sum_{n=1}^N \frac{1}{\mu^q} \mathcal{K} \left(\frac{x - x^{(n)}}{\mu} \right)$$

である。

30

【0316】

コア関数として都市ブロック関数を用いると、これは、

$$\hat{g}_x(x) = \frac{1}{N} \sum_{n=1}^N \frac{1}{\mu^q} \prod_{j=1}^q \left(\frac{1}{2} e^{-\frac{|x_j - x_j^{(n)}|}{\mu}} \right) = \frac{1}{N(2\mu)^q} \sum_{n=1}^N e^{-\frac{1}{\mu} \sum_{j=1}^q |x_j - x_j^{(n)}|} \quad (2.23)$$

になる。

【0317】

2つのランダムベクトル X と Y の共通密度分布は、推定量と産物コア

$$\mathcal{K}_{\mu_x, \mu_y}(x, y) = \mathcal{K}_{\mu_x}(x) \cdot \mathcal{K}_{\mu_y}(y)$$

40

を用いて定式化することができる。

$$\hat{g}_{x, y}(x, y) = \frac{1}{N} \sum_{n=1}^N \mathcal{K}_{\mu_x}(x - x^{(n)}) \cdot \mathcal{K}_{\mu_y}(y - y^{(n)})$$

【0318】

範囲の選択は推定の質に大きな影響を及ぼす。バンド幅が大きければ大きいほど、密度近

50

似が滑らかになるが、精緻でなくなる。いくつかの研究および本件において、選択肢

$$\mu = \left(\frac{4}{N(q+2)} \right)^{\frac{1}{q+4}}$$

はうまくいくことが判明している。

【0319】

式2.20および2.21に従って冗長性を計算するために、2つのランダムベクトルXおよびUの条件付き期待値

$$\mathbb{E}[X|U = u]$$

は、おおまかに推定されるはずである。この目的にナダラヤ・ワトソン推定量が使用されることがある。これは、以前に適用された原理に基づいており、以下のように導出することができる。

$$\begin{aligned} \mathbb{E}[X|U = u] &= \int \frac{x \cdot g_{x,u}(x, u)}{g_u(u)} dx \\ &\approx \frac{\int x \sum_{n=1}^N (\mathcal{K}_{\mu_x}(x - x^{(n)}) \mathcal{K}_{\mu_u}(u - u^{(n)})) dx}{\sum_{n=1}^N \mathcal{K}_{\mu_u}(u - u^{(n)})} \\ &= \frac{\sum_{n=1}^N \mathcal{K}_{\mu_u}(u - u^{(n)}) \int x \mathcal{K}_{\mu_x}(x - x^{(n)}) dx}{\sum_{n=1}^N \mathcal{K}_{\mu_u}(u - u^{(n)})} \\ &= \frac{\sum_{n=1}^N x^{(n)} \mathcal{K}_{\mu_u}(u - u^{(n)})}{\sum_{n=1}^N \mathcal{K}_{\mu_u}(u - u^{(n)})} \end{aligned}$$

【0320】

近似工程では、密度をコア密度推定量によって近似した。最後の等号は、

$$\mathcal{K}_{\mu_x}(\cdot - x^{(n)})$$

という事実起因する。

【0321】

本発明の態様によれば、選択はラッパーの原理に従って行われる。すなわち、それぞれの出力変数について、(PMIに従って) 1~7個の最も関連性のある入力と1~10個の隠れニューロンのある70のネットを、最初、1000回のイテレーションでトレーニングした。この目的のために、8つのモニタリングされたトレーニング発酵から形成された全トレーニングデータセットの一部からトレーニングデータセットを作り出した。残りのデータは試験データとして役立った。それぞれのトレーニングセッションについて、試験誤差Eの値をイテレーション回数にわたって記録し、その最小値と、対応するイテレーション回数を決定した。その後、70のネットを、全最小試験誤差を用いて比較した。これにより、それぞれの代謝産物フラックスを推定するための入力変数の組み合わせと、対応するイテレーション回数および隠れニューロンの数を決定した。

【0322】

様々な出力パラメータ値(「出力」、「出力変数」)に対する、本明細書では「入力」ま

たは「入力変数」とも呼ぶ選択された入力パラメータ値のリストを、それに対応して図13に示した。

【0323】

好ましい態様によれば、トレーニング済みMLPのトレーニング用または供給用の入力パラメータ値として濃度を使用する細胞外代謝産物（入力代謝産物）を、各出力代謝産物に対して個々に、すなわち予測しようとする各細胞外フラックスに対して個々に、純粋に統計学的基準に従って選択した。

【0324】

1. 「関連性」に従う入力代謝産物のランク付け:

最初に、全ての測定可能で利用可能な入力代謝産物または少なくとも、細胞外代謝産物として代謝モデルにおいて生じた全ての代謝産物を「第1のセット」に移し、関連性に従ってソートする。すなわち、PMI基準を使用することによって、どの代謝産物が、出力のための最大の表現/予測能力（細胞外フラックスを予測しようとする細胞外代謝産物-「出力代謝産物」の速度）を持つかを確かめる。この代謝産物を「第2のセット」に移す。次に、「第2のセット」は実際の入力パラメータ値を提供する。この第1のソーティング工程において確かめた「関連性」または予測能力は第2のセットにある代謝産物に依存しない。

【0325】

2. 真の入力代謝産物の決定:

好ましくは、MLPトレーニングまたはMLP適用のための入力として全ての入力代謝産物を使用するわけではない（これは、オーバーフィッティングにより非常に不十分な予測能力をもたらす）。従って、最も関連性のある入力代謝産物のうちどれだけ使用すべきかを確かめる。これは、試験データセットから「x」個の最も関連性のある入力代謝産物の予測能力を確かめ、xを変え、次に、予測能力が最も良い数xを選択することによって行われる。

【0326】

第1のソーティング工程と、第1セットから、最も関連性のある代謝産物を、実際の入力代謝産物（その濃度がMLPの入力パラメータとして提供される）の「第2のセット」へと送った後に、第2のセットの中身を考慮して、第1のセットの残りの代謝産物中にある、決められた出力代謝産物のフラックスに関して予測能力が最も高い入力代謝産物を繰り返し特定する。第1のセットの残っているメンバー内で予測関連性が最も高い代謝産物の濃度プロファイルが、既に第2のセット含まれている代謝産物と強く相関するのであれば、通常、この代謝産物は第2のセットに移されない。なぜなら、その予測能力が高い可能性があるが、その濃度プロファイルは、既に第2のセット含まれている代謝産物以上に多大な貢献をしないからである。もっと正確に言うと、それを取り込んでも、第2のセットにおいて冗長な情報の量が増加するだけであろう。従って、これらの理由で、第1のセットにある代謝産物が第2のセットに含まれないことがあれば、第1のセットの次の最も高い関連性スコアを有し、第2のセットの代謝産物の濃度プロファイルの情報内容の冗長性を過剰に増やさない代謝産物を、第1のセットから第2のセットに移す。

【0327】

従って、代謝産物は、第2のセットに移されずに、ある特定の出力代謝産物を出力する速度にとって非常に意味があるかもしれない。特に、この代謝産物の濃度プロファイルが、既に第2のセット含まれている代謝産物の濃度プロファイルと非常に強い相関関係がある場合に、移すのを省略されることがある。その結果、MLPをトレーニングする時に、その後でトレーニング済みMLPを使用する時でも、入力パラメータ値として2つのその濃度のうちの1つだけ使用することで十分である。これは、2つのうちの1つが「関連性のある」ものとして選択される場合、他方はその意義を失うことを意味する。これはPMI基準によって認識される。予測能力の点で関連性があり、かつ十分に互いに独立している全ての入力代謝産物が第2のセットに含められるまで、このように続ける。

【0328】

例えば、Gluフラックス（図13に示した表における4列目）の場合、現在のバイオマスと現在のグリシン濃度についての情報が、入力としてバイオマスしか使用されない場合（情

10

20

30

40

50

報が少なすぎる) またはグルタミン、アスパラギン酸などについての追加情報がニューラルネットワークに供給される場合(情報が多すぎる。オーバーフィッティング) よりも優れた予測を提供し得ることが見出されている。

【0329】

図14は、いくつかのフェドバッチ発酵動作について得られた細胞内フラックスの、得られた「二乗平均平方根誤差」(RMSE)のヒストグラムを示す。12のフェドバッチリアクターを準備し、それぞれが、他の11のリアクターの細胞クローンとは異なるCHO細胞クローンを含んだ。12全てのクローンが同じ産物、すなわち、二重特異性抗体を産生するように、12の細胞クローンを遺伝子組換えした。クローンを生成するトランスフェクションの間に同じDNA配列を使用した。これらのクローンは、組み込まれたDNA配列の挿入遺伝子座が異なることで、および/またはコピー数が異なることで代謝の違いを示す。MLP(ここでは、ニューラルネットワークモデル)をトレーニングするために10のフェドバッチを使用し、モデルを検定するのに2つのフェドバッチを使用した。これらの2つのフェドバッチをRMSE計算に使用した。

【0330】

MLPとMFAの組み合わせによって予測した細胞内フラックスと、細胞外フラックス測定値から計算した細胞内フラックスとの差から細胞内フラックスのRMSEを計算する。RMSEは一度もマイナスにならず、値0(実際にはほとんどない)は予測データと測定データの完全に一致することを示している。一般的に、大きなRMSEより小さなRMSEの方が良い。RMSEは、二乗した誤差の平均の平方根である。RMSEに及ぼす、それぞれの誤差の影響は、二乗した誤差の大きさに比例する。そのため、大きな誤差はRMSEに不釣り合いなほどの影響を及ぼす。

【0331】

二重特異性抗体を産生するために、同じタイプの培地と同じ培養培地を12のフェドバッチリアクターに使用した。しかしながら、この培地と培養培地は、付属書の図5.9に示した代謝フットプリントをもつバイオリアクターまたは細胞培養物に使用した培地および培養培地とは異なった。

【0332】

MLP、ここではニューラルネットワーク(NN)を、付属書の図5.9に示したバイオリアクターから取得したデータに対して既に1回トレーニングされたことがある新たなデータセット、すなわち、NNに対して再較正し、二重特異性抗体産生について12のフェドバッチリアクターからのデータに対して「再トレーニング」、すなわち新たにトレーニングした。これらのモデルは、機械ライブラリー-サイキット・ラーン(Scikit learn)を用いてPythonで作成した。様々なNNモデルのハイパーパラメータをグリッドサーチ関数を用いて最適化した。データと最もマッチしたモデルおよびハイパーパラメータを記憶し、12のバイオリアクターの中にある細胞培養物の細胞外フラックスならびに細胞内フラックスの将来の予測のために「再トレーニング済み」NNの形で使用した。

【0333】

図15は、図14で述べた12のフェドバッチ発酵動作について得られた細胞外フラックスの、得られたRMSEのヒストグラムを示す。全てのRMSE値を、代謝産物グルコースについて得られた誤差に対して基準化した。細胞外フラックスのRMSE値は、外部フラックス測定値と、MLP、特にニューラルネットワークとMFAの組み合わせによる細胞外フラックス予測値との間の差から計算した。

【0334】

二重特異性抗体産生用の12のフェドバッチリアクターについて得られた細胞内フラックスと細胞外フラックスの両方のRMSEが他の細胞培養物または他の細胞クローンについて得られたRMSEと同じ範囲にあったことが観察されるかもしれない(修士論文「付属書」67頁、図5.9を参照されたい)。この細胞培養物のタンパク質産物は抗体融合タンパク質である。12の細胞培養物はCHO細胞培養物である。(予測されたものと比べて測定された)外部フラックスおよび内部フラックスのRMSEは10~35%であった。

【0335】

図16は、異なるバイオリアクター（フェドバッチバイオリアクターおよびスプリットバッチバイオリアクター）の中にある2つの同一の細胞クローンの異なる代謝産物（アミノ酸）を対象にした、細胞外代謝産物フラックス予測値（黒色の線）と2つの細胞外代謝産物フラックス測定値（2本の灰色の線）との比較を示す。両バイオリアクターの中にある細胞クローンは、特定の抗体融合タンパク質を合成する組換えCHO細胞（モノクローナル）である。これらの曲線は、本発明の態様に従う方法が、フェドバッチ(fb)およびスプリットバッチ(b)の両方について細胞外代謝産物のフラックスを非常に正確に予測できることを示している。詳細には、図16は、トレーニングデータ/試験データセットの異なる組み合わせでの細胞外フラックス測定値と細胞外フラックス予測値との比較を示す。「fb/fb+b」という表現は、トレーニングデータとしてのフェドバッチデータセットと、試験データセットとしてのフェドバッチ+バッチを意味する。次に、試験データセットからの1回の動作を、それぞれの場合で示す。

10

【0336】

図17は、それぞれに2つの曲線がある、いくつかのグラフを示す。全てのグラフが、抗体融合タンパク質産生用の細胞クローンを用いたフェドバッチバイオリアクターについて得られた。点線からなる曲線（「記述的MFA」）は、代謝フラックス解析(MFA)の入力として細胞外代謝産物の濃度測定値を使用して、それぞれが11のグラフの1つに相当する異なる細胞外フラックスを計算することによって得られた。実線曲線（「NN-MFA」）は、MLP（例えば、NN）の入力として、決められた時点t0での細胞外代謝産物の濃度測定値を使用して将来の時点t1での細胞外フラックスを予測し、代謝フラックス解析(MFA)の入力として、これらの将来の細胞外フラックス予測値を使用することによって得られた。従って、11のグラフのそれぞれの2つの曲線を比較すると、将来の時点についてMLP+MFAによって予測した値は、現在の細胞外代謝産物の濃度測定値を用いて静的MFAモデルによって得た値と非常によく一致していることが分かる。

20

【0337】

図18は、それぞれに2つの曲線がある、いくつかのグラフを示す。全てのグラフが、二重特異性抗体産生用の第1の細胞クローンZK1を用いたフェドバッチバイオリアクターについて得られた。点線からなる曲線（「NN+記述的細胞外MFA」）は、MLP（例えば、NN）の入力として、ある特定の時点t0での細胞外代謝産物のフラックス測定値を使用して将来の時点t1での細胞外フラックスを予測し、代謝フラックス解析(MFA)の入力として、これらの将来の細胞外フラックス予測値を使用することによって得られた。交差した線の曲線（「記述的細胞外MFA」）は、代謝フラックス解析(MFA)の入力として、現在測定されている細胞外代謝産物の細胞外フラックスを使用して、それぞれが図18に示したグラフの1つに対応する異なる細胞外フラックスを計算することによって得られた。各グラフについて、毎日、細胞外フラックスをグルコース濃度の点から基準化した。従って、細胞外代謝産物濃度測定値の代わりに、現在の代謝産物濃度と、過去の時点での過去の代謝産物濃度に基づいて計算した現在の細胞外フラックスをMLP入力として使用した。従って、広い意味の代謝産物濃度を入力として使用した。または、代謝産物濃度測定値を入力として使用した場合に、MLPの予測結果が最終的に本質的に同一であることが観察された。従って、狭い意味の濃度測定値と広い意味の代謝産物濃度は等しくMLPの入力として役立ち得る。

30

40

【0338】

従って、図18のグラフのそれぞれの2つの曲線を比較すると、将来の時点についてMLP+MFAによって予測した値は、現在の細胞外代謝産物の濃度測定値を用いて静的MFAモデルによって得られた値と非常によく一致していることが分かる。しかしながら、詳細には逸脱は起こり得る。

【0339】

図18のグラフは、a)（細胞外）フラックス測定値の進行を、MLPにより予測した対応物と比較したものである。全てのフラックスを、基準化した値-フラックス測定値に対して基準

50

化した値で表した。従って、グルコースフラックスは常に1である（グラフとして示さなかった）。b) MFAにより計算した外部フラックス測定値からの内部フラックスの進行を、NNにより計算した外部フラックスからMFAにより計算した外部フラックスからの内部フラックスと比較したものである。これも、全てのフラックスをグルコースフラックスに対して基準化した。

【0340】

個々の代謝産物について、フラックス測定値は予測値から逸脱する時もあった。しかしながら、一方では、スケーリングはグルコース基準化による非常に高い「解像度」と一致し、代謝産物の総量を考慮した時には逸脱はかなり小さかったことに留意すべきである。さらに、データがオーバーフィッティングする、ある特定の傾向が観察された。これは、通常、データセットのサイズを増やすことによって補正することができる。

【0341】

11種の異なる細胞外フラックスについて作成した図18のグラフは細胞クローンZK1の「代謝指紋」を表している。以下では、この指紋は、同じ産物を産生する（二重特異性抗体を産生するが、このタンパク質をコードする配列を異なる数のコピーで、または異なるゲノム位置で含む）他の細胞クローンの指紋と相当異なる場合があることが示される。

【0342】

図19は、それぞれに2つの曲線がある、いくつかのグラフを示す。全てのグラフが、二重特異性抗体産生用の第2の細胞クローンZK2を用いた別のフェドバッチバイオリアクターBR2について得られた。これらの曲線は、図18について説明したように求めたが、第2の細胞クローンZK2からのデータに基づいた。図18および図19に示したグラフは、同じタイプの培地と培養培地を用いて動作させた2つのバイオリアクターBR1、BR2からのデータに基づいている。BR1、BR2は、本質的に、2つのバイオリアクターが異なる細胞クローンZK1およびZK2を保持している点でしか差がない。細胞クローンZK1およびZK2は両方とも同じ産物（二重特異性抗体）を合成するが、クローンの代謝に影響を及ぼす可能性がある遺伝子の違いも有する。例えば、これらのクローンは、新たなDNA配列セグメントを細胞ゲノムに組み込む位置および/または組み込まれる配列セグメントの数を完全に制御しない方法によって産生されている可能性がある。従って、使用した出発細胞は遺伝的に同一であるが、新たな遺伝子またはDNA配列を組み込んでいる間（例えば、トランスフェクションをしている間のランダムな組み込みの場合）、二重特異性抗体の軽鎖(LC)および重鎖(HC)の遺伝子配列は異なるコピー数および異なるゲノム遺伝子座で挿入される可能性がある。そして次に、これにより、代謝の違いが生じ、特定の細胞クローンの生産性および生命力の違いも生じる可能性がある。これらは、本発明の態様に従って、MLPとMFAの組み合わせの代謝予測によって容易に識別可能および区別可能である。第2のバイオリアクターBR2は、クローンZK1の細胞培養物に使用したバイオリアクターBR1と構造の点で同一であった。

【0343】

図18および図19はそれぞれ、特定の細胞クローンZK1、ZK2について、ニューラルネットワークNNと代謝フラックス解析MFAの組み合わせによって、純粋にMFAと実際の時間に基づく細胞外フラックスの予測とは異なる、かつ最終的にそれより正確な予測が得られることを示す。さらに、この組み合わせには、予測の質をフラックス測定値によって繰り返し検証できるという利点がある。細胞内フラックスの直接予測では、¹³C標識代謝産物を用いた検証しかできず、これは、特にハイスループットの場合または生産発酵の場合、実用的でも経済的でもない。さらに、例えば、ZK1については図18、ZK2については図19に示したように特定の細胞クローンZK1、ZK2について細胞外フラックス予測値の計算量は特定の細胞クローンの「代謝指紋」を表す。これは、増殖速度、アンモニア分解速度、セリン産生速度などなどの決められたパラメータに関して特に有益だと思われる特定の細胞クローンを特定するために分析されることがある。代謝「指紋」はまた、代謝類似性に関して2つ以上の異なる細胞クローンを選択または特徴付けるために使用されることがある。

【0344】

図20は、それぞれに2つの曲線がある、いくつかのグラフを示す。全てのグラフが、細胞クローンZK1を用いたBR1バイオリアクターについて得られ、それぞれのグラフが細胞内フラックス計算値を表す。図20は、細胞外フラックス計算値ではなく細胞内フラックス計算値を示したという違いはあるが図18に対応する。

【0345】

図21は、それぞれに2つの曲線がある、いくつかのグラフを示す。全てのグラフが、細胞クローンZK2を用いたBR2バイオリアクターについて得られ、それぞれのグラフが細胞内フラックス計算値を表す。従って、図21は、細胞外フラックス計算値ではなく細胞内フラックス計算値を示したという違いはあるが図19に対応する。従って、図19および図21にあるグラフの全体が、細胞クローンの代謝特徴付けに使用され得る代謝指紋を表している。多数の異なるクローンと発酵条件について、これらの多数の「指紋」からデータを収集することにより、ならびに個々のクローンの生命力および/または生産性を記録することにより、例えば、相関解析によって、有利な代謝指紋を特定し、将来のクローニングに有利なクローンを特定する目的で参照値として有利であると突き止められた、これらの代謝指紋を使用し、発酵計画のために選択することができる。

【0346】

符番のリスト

102~112 工程

200 システム

202 プロセッサ

204 バイオリアクター

206 バイオリアクター

208 バイオリアクター

210 測定値インターフェイス

212 記憶媒体

214 代謝モデル

216 参照値

218 機械学習ロジック

220 プログラムロジック

222 制御インターフェイス

224 ユーザーインターフェイス

250 代謝産物の濃度を求めるための装置

252 計算および予測のためのコンピュータシステム

254 フラックスのある代謝モデル

256 細胞外代謝産物の濃度測定値

402 代謝モデル

404 細胞外代謝産物

406 細胞内代謝産物

408 細胞外フラックス

410 細胞内フラックス

502 細胞外代謝産物の濃度測定値の経過

504 細胞外フラックス予測値の経過

506 MFAによって計算した細胞内フラックスの経過

802 フェドバッチバイオリアクター1のグラフ

804 フェドバッチバイオリアクター5のグラフ

806 バッチバイオリアクターのグラフ

808 スプリットバッチバイオリアクターのグラフ

1502 出力パラメータ値

1504 入力パラメータ値

10

20

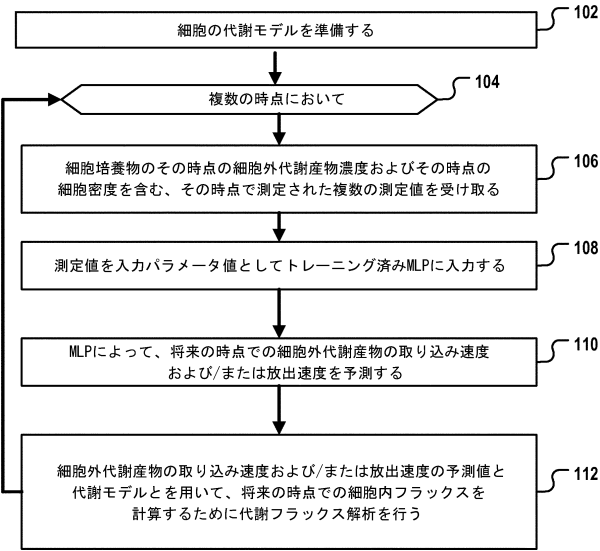
30

40

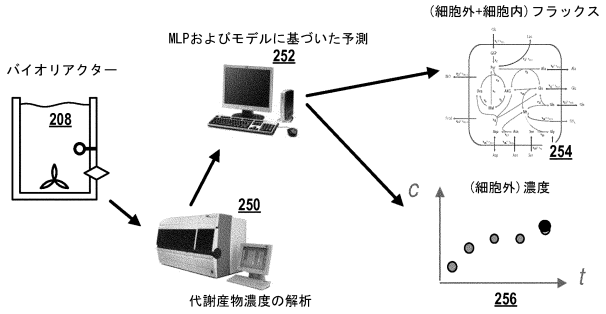
50

【図面】

【図 1】

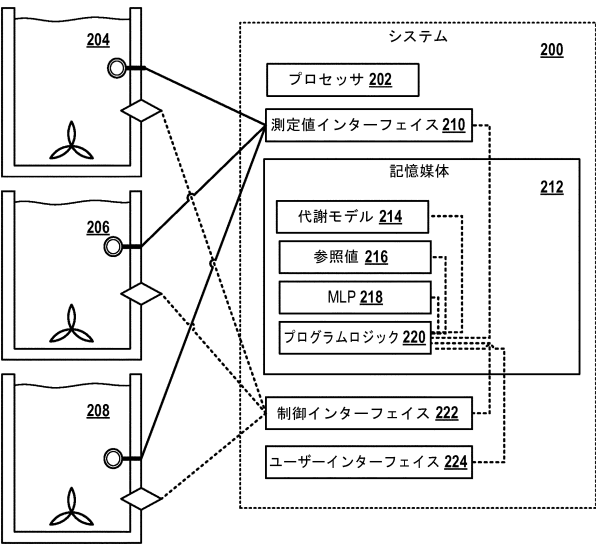


【図 2】

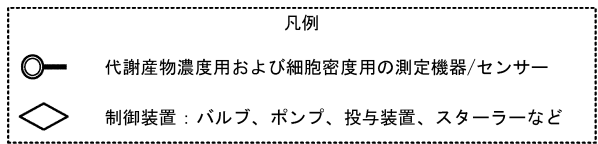


10

【図 3 A】



【図 3 B】



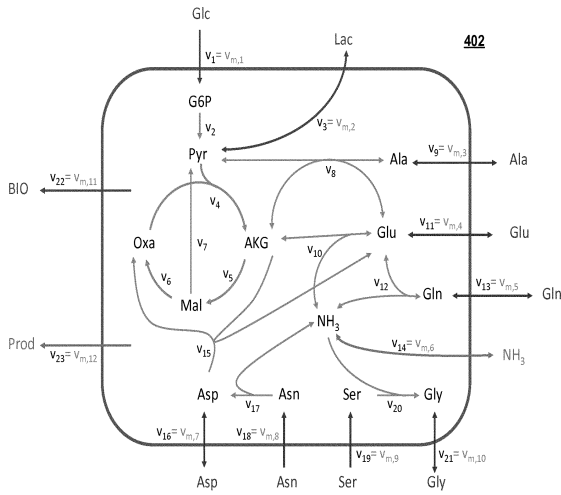
20

30

40

50

【図 4 A】



凡例

404 12種の細胞外代謝産物：
Glc, Lac, Ala, Glu, Gln, NH₃, Gly, Ser, Asn, Asp, Prod, BIO

406 13種の細胞内代謝産物：
G6P, Pyr, Ala, Glu, Gln, NH₃, Gly, Ser, Asn, Asp, Mal, Oxa, AKG

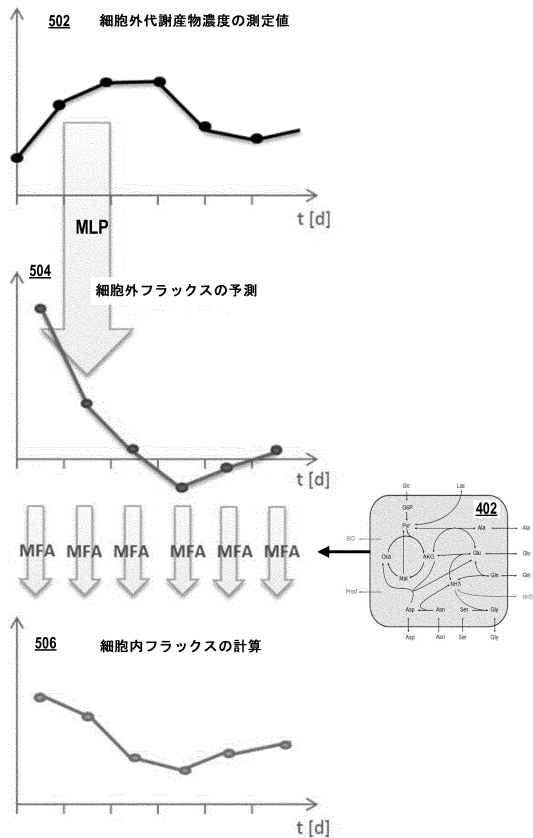
408 12種の細胞外フラックス：
v1, v3, v9, v11, v13, v14, v21, v19, v18, v16, v23, v22

410 11種の細胞内フラックス：
v2, v4, v5, v6, v7, v8, v10, v12, v15, v17, v20
(Altamirano et al. (2006)、Llaneras et al. (2007)、Nolan et al. (2011)に基づく改良モデル)

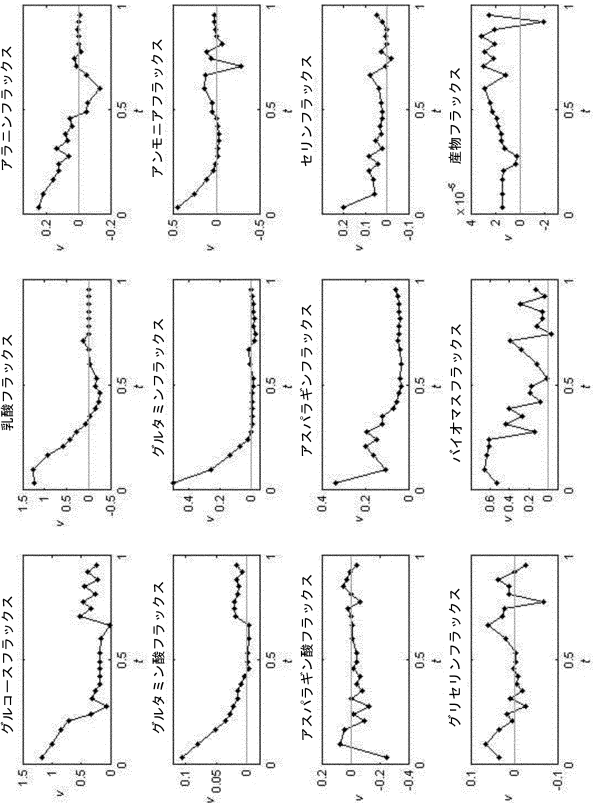
【図 4 B】

反応番号	化学量論
1	Glc _e → G6P
2	G6P → 2 Pyr
3	Pyr ↔ Lac _e
4	Pyr + Oxa → AKG
5	AKG → Mal
6	Mal → Oxa
7	Mal → Pyr
8	Pyr + Glu ↔ Ala + AKG
9	Ala ↔ Ala _e
10	AKG + NH ₃ ↔ Glu
11	Glu ↔ Glu _e
12	Gln ↔ Glu + NH ₃
13	Gln _e ↔ Gln
14	NH ₃ ↔ NH _{3e}
15	AKG + Asp → Oxa + Glu
16	Asp ↔ Asp _e
17	Asn → Asp + NH ₃
18	Asn _e → Asn
19	Ser _e → Ser
20	Ser + NH ₃ → 2 Gly
21	Gly ↔ Gly _e
22	0.084 Ala + 0.087 Gln + 0.080 Asp + 0.041 Asn + 0.096 Ser + 0.056 Gly + 0.452 G6P + 0.427 Oxa → Bio + 0.008 Glu + 0.445 Mal + 0.209 Pyr
23	138 Ala + 104 Glu + 106 Gln + 80 Asp + 72 Asn + 316 Ser + 204 Gly → Prod

【図 5】



【図 6】



10

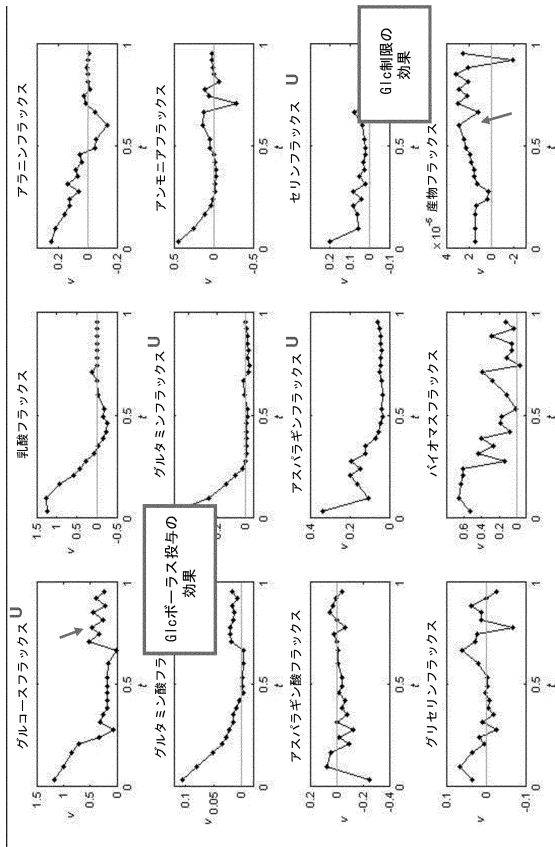
20

30

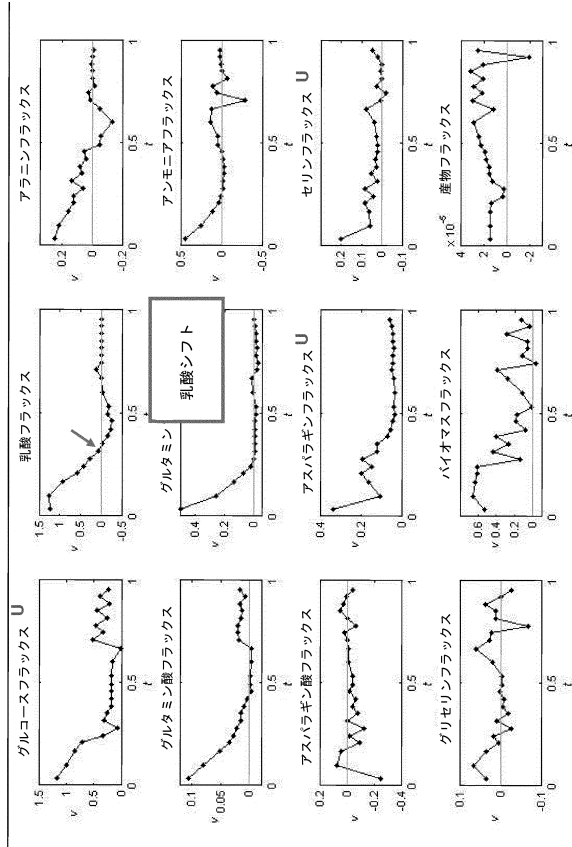
40

50

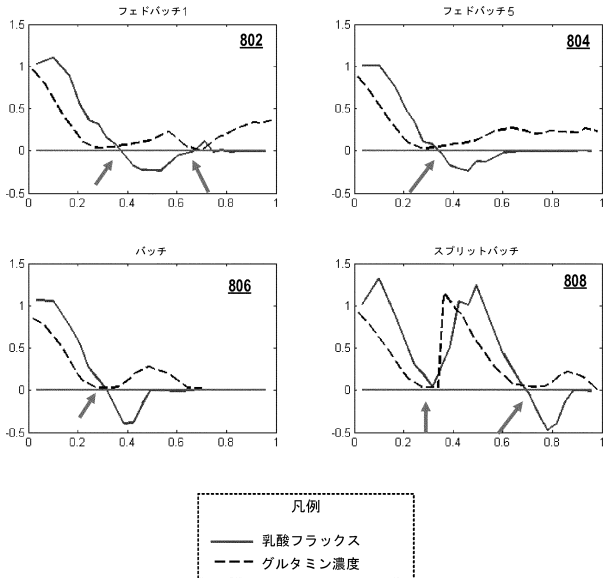
【図 7】



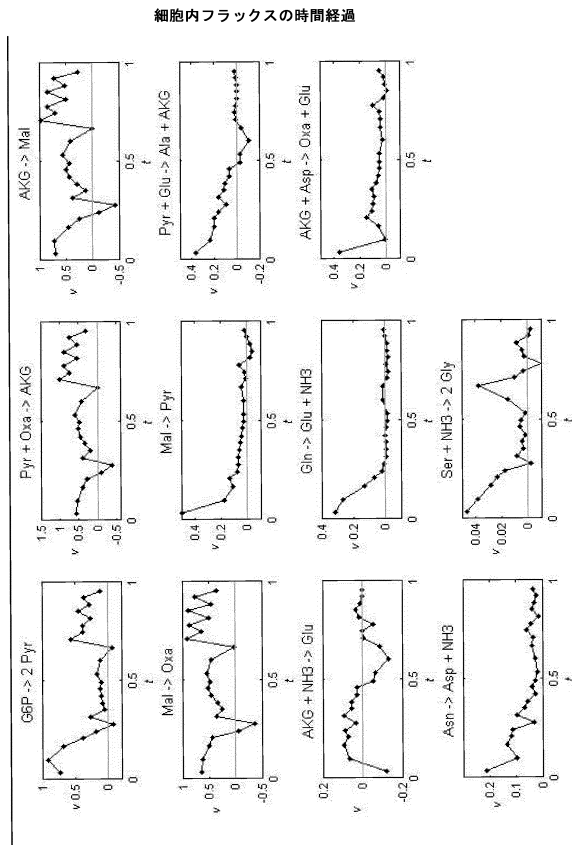
【図 8】



【図 9】



【図 10】



10

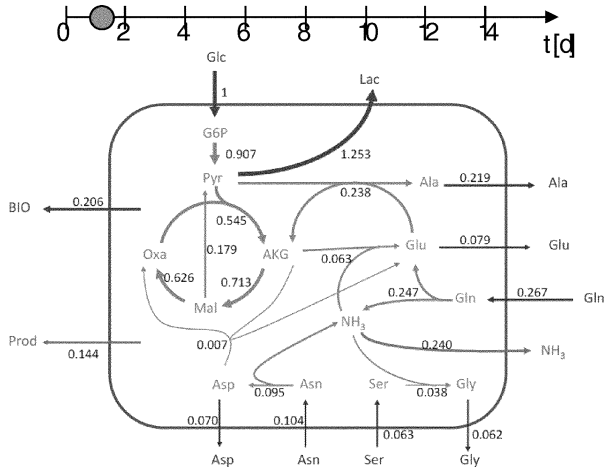
20

30

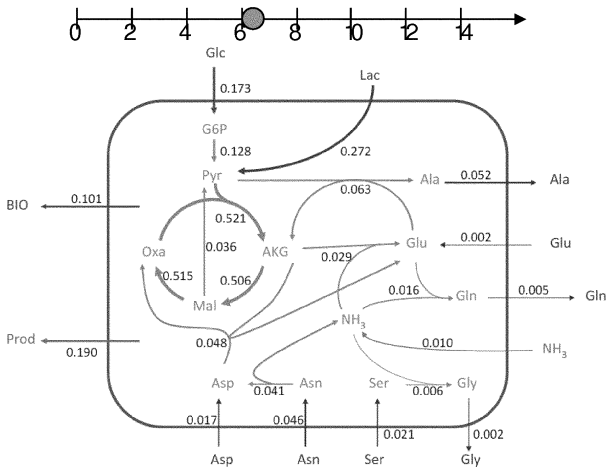
40

50

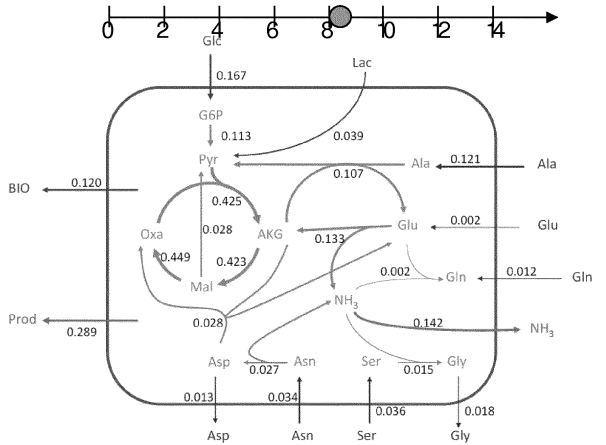
【図 1 1 A】



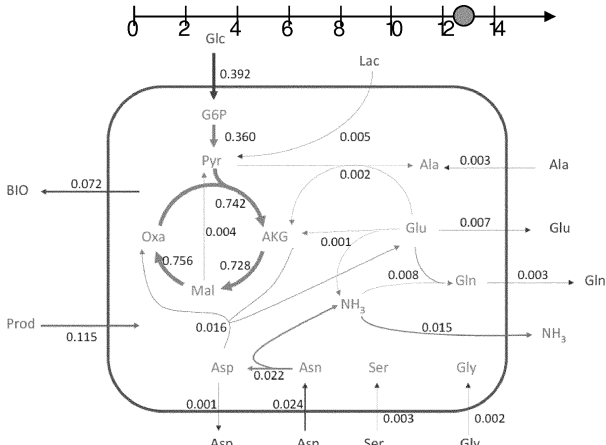
【図 1 1 B】



【図 1 1 C】



【図 1 1 D】



10

20

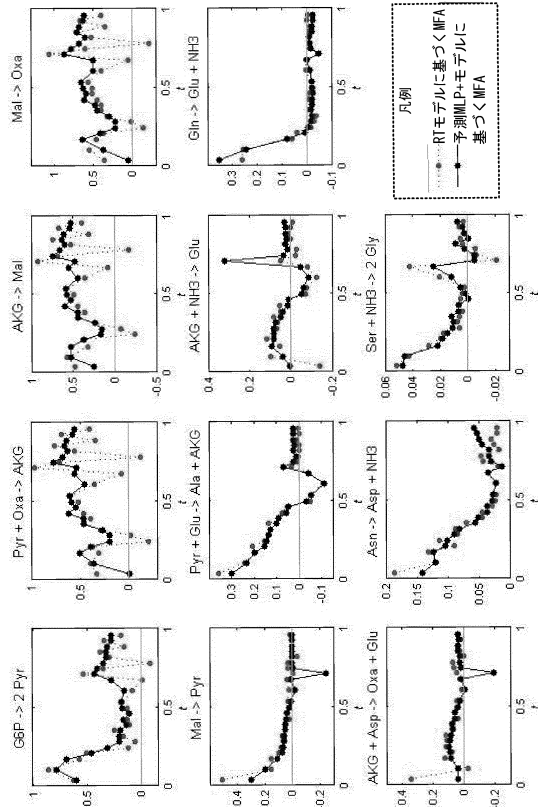
30

40

50

【図 1 2】

RTモデルに基づくMFAと予測MLP+モデルに基づくMFAとの比較

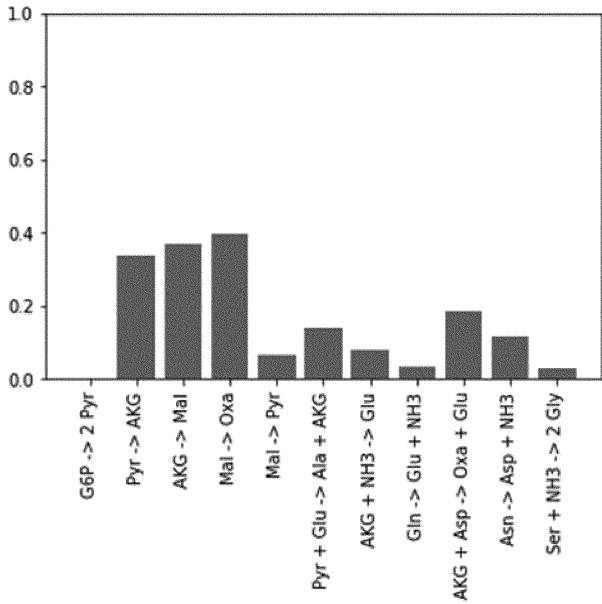


【図 1 3】

出力	Glc	Lac	Ala	Glu	Gln	NH ₃	Asp	Asn	Ser	Gly	TZD	Prod
入力	Bio	Asn	Gln	Bio	Gln	Gln	Lac	Asn	Asn	Gln	Bio	Prod
	NH ₃	Ala	Gln	Gly	Asn	NH ₃	Lac	Asp	Asp	LDH	Gly	Ala
	Gln	Gln	NH ₃	Gln	Asp	Ser	Asn	Gln	Ser	Ser	Ser	Asp
	Ser	Lac	Lac	Asp	Glu	Gly	Asp	Ala	LDH	Ala	NH ₃	Ser
	Ala	Ser	Ser	Ala	Asn	LDH	Ser	Ala	Asp	Asp	LDH	LDH
	Asp	Bio	Ala	Asn	Prod	Glu	Ala	Glu	Glu	Asn	Glu	Glu
	Asn	NH ₃	Asn	Glu	Lac	Glu	Gly	Prod	Prod	Prod	NH ₃	NH ₃
	Lac	Asp	Bio	LDH	Ala	Glc	Prod	Prod	Gly	Gly	Gln	Gln
	Prod	Glu	Prod	NH ₃	NH ₃	Ala	Glu	Gly	Gly	Glu	Asn	Asn
	Glu	Prod	Asp	Prod	Bio	Prod	Bio	Lac	Ala	Asp	Glc	Glc
	Gly	Gly	Glu	Ser	Ser	Asn	Glc	NH ₃	Bio	NH ₃	Prod	Bio
	LDH	LDH	Gly	Lac	Gly	Bio	LDH	LDH	NH ₃	Bio	Ala	Gly
	Glc	Glc	LDH	Glc	Glc	Asp	NH ₃	Glc	Lac	Lac	Lac	Lac
H	8	4	6	8	1	2	6	4	2	3	1	4
イテレーション	916	1000	1000	299	1000	999	770	761	169	764	1000	500

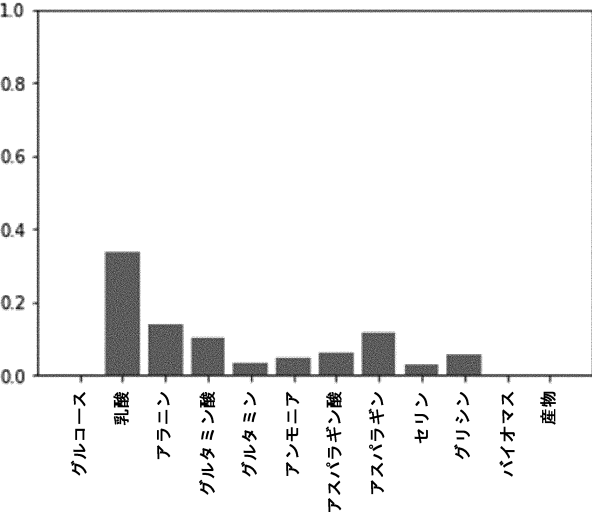
【図 1 4】

RMSE細胞内フラックス



【図 1 5】

RMSE細胞外フラックス



10

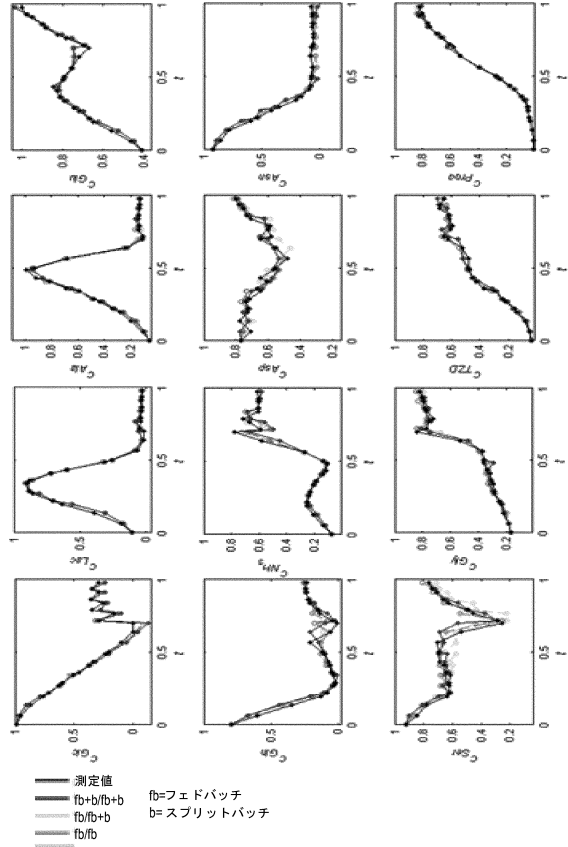
20

30

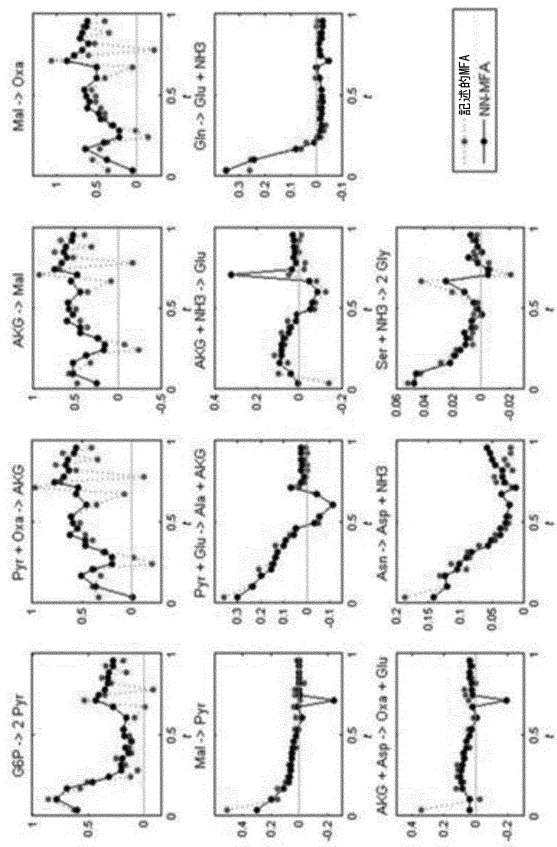
40

50

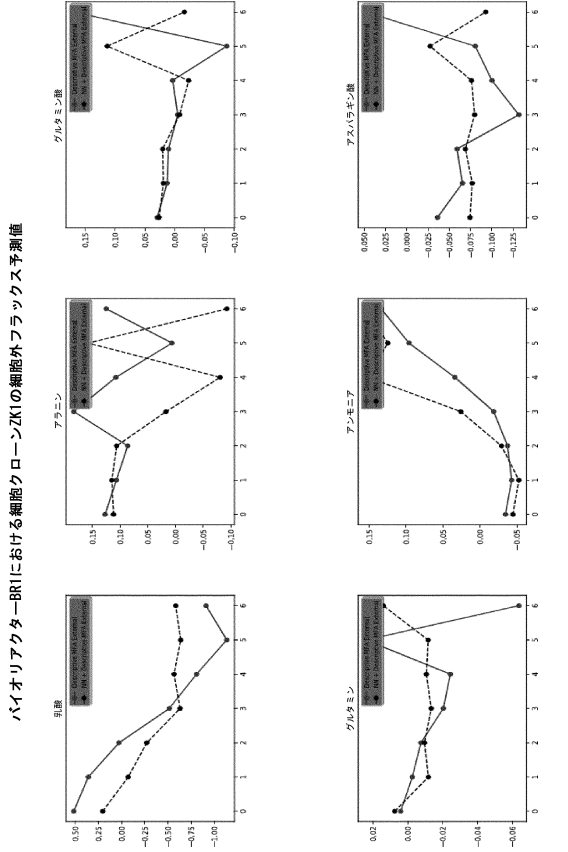
【図 16】



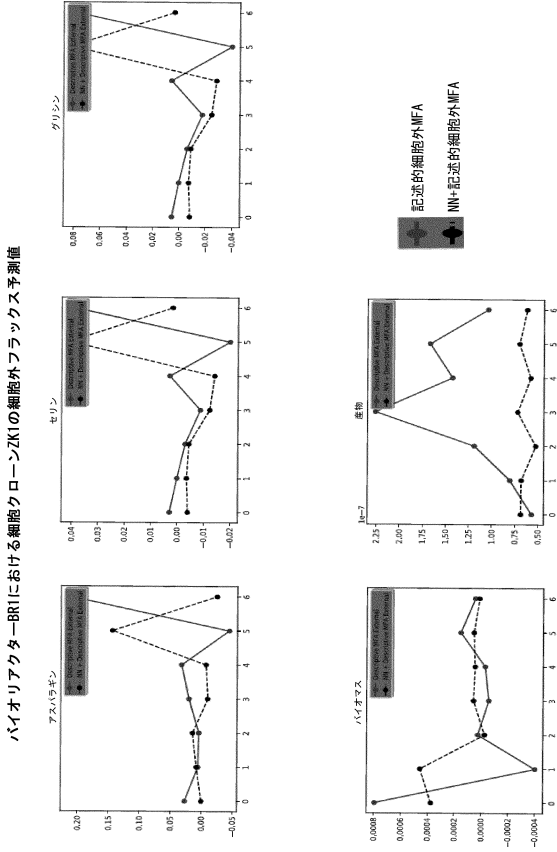
【図 17】



【図 18 A】



【図 18 B】



バイオリアクター-BRIにおける細胞クローンZK1の細胞外フラックス予測値

バイオリアクター-BRIにおける細胞クローンZK1の細胞外フラックス予測値

10

20

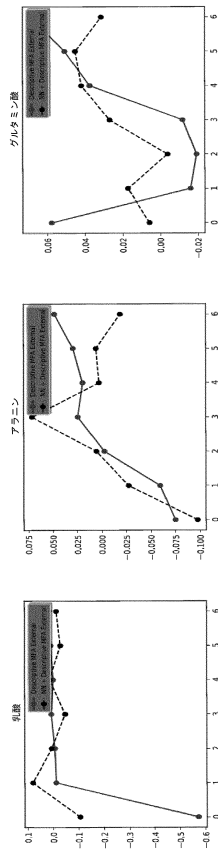
30

40

50

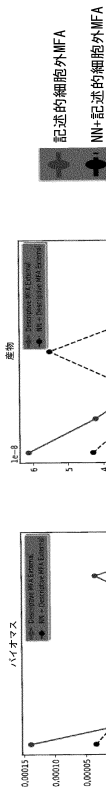
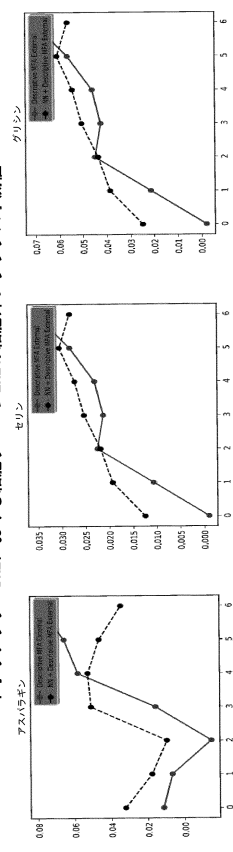
バイオリアクター-BR2における細胞クローンZK2の細胞外フラックス予測値

【図 19 A】



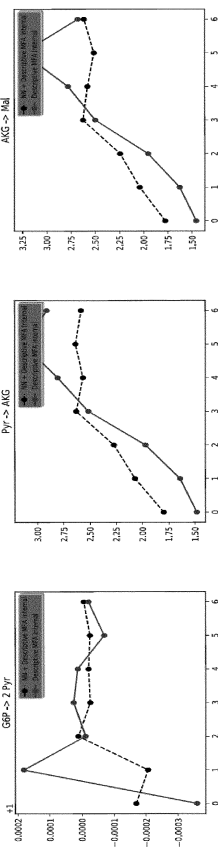
バイオリアクター-BR2における細胞クローンZK2の細胞外フラックス予測値

【図 19 B】



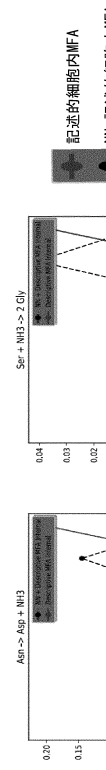
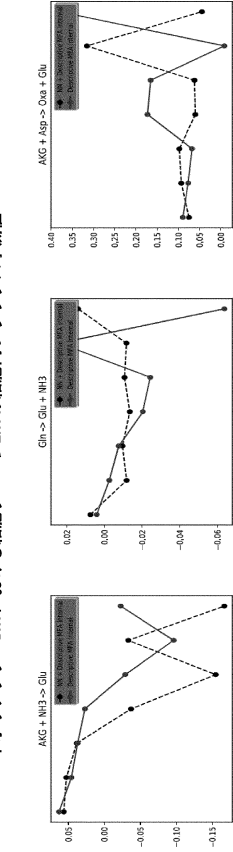
バイオリアクター-BR1における細胞クローンZK1の細胞内フラックス予測値

【図 20 A】



バイオリアクター-BR1における細胞クローンZK1の細胞内フラックス予測値

【図 20 B】



10

20

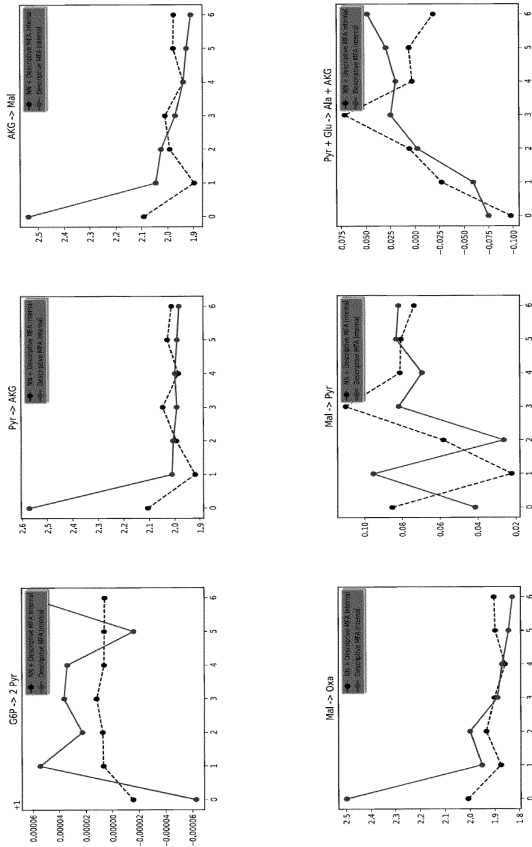
30

40

50

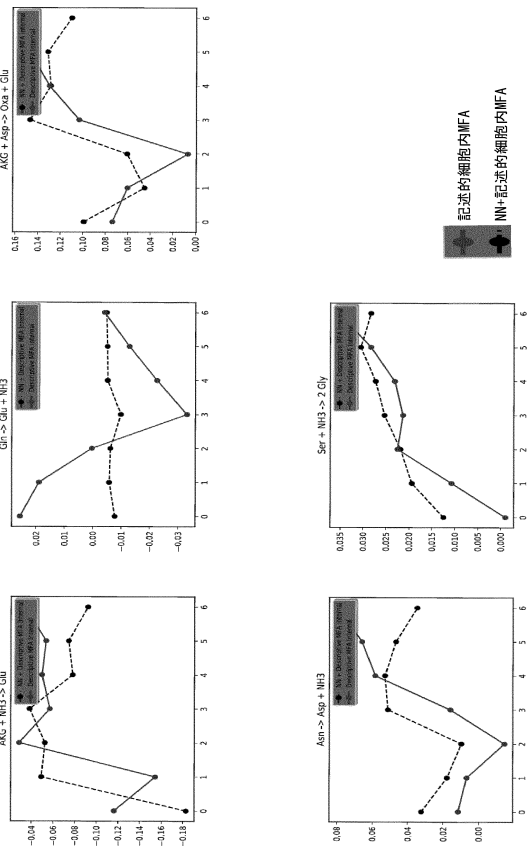
バイオリアクターBR2における細胞クローンZK2の細胞内フラックス予測値

【図 2 1 A】



バイオリアクターBR2における細胞クローンZK2の細胞内フラックス予測値

【図 2 1 B】



NN-記述的細胞内MFA
NN-記述的細胞内MFA

10

20

30

40

50

フロントページの続き

- (74)代理人 100142929
弁理士 井上 隆一
- (74)代理人 100148699
弁理士 佐藤 利光
- (74)代理人 100128048
弁理士 新見 浩一
- (74)代理人 100129506
弁理士 小林 智彦
- (74)代理人 100205707
弁理士 小寺 秀紀
- (74)代理人 100114340
弁理士 大関 雅人
- (74)代理人 100121072
弁理士 川本 和弥
- (72)発明者 ポール ヴォルフガング
ドイツ連邦共和国 8 2 3 7 7 ペンツベルク ノンネンヴァルト 2 ロシュ ダイアグノスティッ
クス ゲーエムベーハー内
- (72)発明者 モーラル アーサー
ドイツ連邦共和国 8 2 3 7 7 ペンツベルク ノンネンヴァルト 2 ロシュ ダイアグノスティッ
クス ゲーエムベーハー内
- (72)発明者 ホルチュ サユリ
ドイツ連邦共和国 8 0 8 0 5 ミュンヘン ホランドシュトラッセ 2
- 審査官 藤澤 美穂
- (56)参考文献 特開 2 0 0 7 - 0 4 7 9 9 4 (J P , A)
特開 2 0 0 5 - 0 3 4 1 4 9 (J P , A)
特開 2 0 1 3 - 0 8 5 5 1 6 (J P , A)
特開 2 0 0 5 - 2 9 3 3 9 4 (J P , A)
Wu S.G. et al., Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming, PLOS Computational Biology, 2016年04月19日, p.1-22, 検索日: 令和3年10月20日, URL: <https://doi.org/10.1371/journal.pcbi.1004838>, DOI: 10.1371/journal.pcbi.1004838
Nolan R.P. et al., Dynamic model of CHO cell metabolism, Metabolic Engineering, 2010年10月07日, 検索日: 令和3年10月20日, URL: <https://doi.org/10.1016/j.ymben.2010.09.003>, DOI: 10.1016/j.ymben.2010.09.003
Panagiotou, G. et al., Monitoring novel metabolic pathways using metabolomics and machine learning: induction of the phosphoketolase pathway in *Aspergillus nidulans* cultivations, Metabolomics, 2007年12月, Vol. 3, No. 4, p.503-516, 検索日: 令和3年10月20日, URL: <https://doi.org/10.1007/s11306-007-0061-7>, DOI: 10.1007/s11306-007-0061-7
- (58)調査した分野 (Int.Cl., D B名)
G 1 6 B 5 / 0 0 - 9 9 / 0 0
G 1 6 C 1 0 / 0 0 - 9 9 / 0 0
G 0 1 N 3 3 / 5 0 Z
C 1 2 M 1 / 0 0 A
C 1 2 Q 1 / 0 2