

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges
Eigentum

Internationales Büro

(43) Internationales
Veröffentlichungsdatum
4. August 2016 (04.08.2016)



(10) Internationale Veröffentlichungsnummer
WO 2016/120361 A1

(51) Internationale Patentklassifikation:

G06F 19/12 (2011.01) G06F 19/26 (2011.01)

(21) Internationales Aktenzeichen: PCT/EP2016/051753

(22) Internationales Anmeldedatum:
28. Januar 2016 (28.01.2016)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
15153052.4 29. Januar 2015 (29.01.2015) EP

(71) Anmelder: BAYER TECHNOLOGY SERVICES
GMBH [DE/DE]; 51368 Leverkusen (DE).

(72) Erfinder: NEYMANN, Tobias; Margaretenhöhe 6a,
51465 Bergisch Gladbach (DE). HEBING, Lukas;
Plauenerstr. 60, 44139 Dortmund (DE). ENGELL,
Sebastian; Appendahl 22, 58300 Wetter (DE).

(74) Anwalt: BIP PATENTS; c/o Bayer Intellectual Property
GmbH, Alfred-Nobel-Str. 10, 40789 Monheim am Rhein
(DE).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für
jede verfügbare nationale Schutzrechtsart): AE, AG, AL,
AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW,
BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK,

DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM,
GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

(84) Bestimmungsstaaten (soweit nicht anders angegeben, für
jede verfügbare regionale Schutzrechtsart): ARIPO (BW,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST,
SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY, KG,
KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH,
CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE,
IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Erklärungen gemäß Regel 4.17:

— hinsichtlich der Berechtigung des Anmelders, ein Patent zu
beantragen und zu erhalten (Regel 4.17 Ziffer ii)

Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz
3)

(54) Title: COMPUTER-IMPLEMENTED METHOD FOR CREATING A FERMENTATION MODEL

(54) Bezeichnung : COMPUTERIMPLEMENTIERTES VERFAHREN ZUR ERSTELLUNG EINES
FERMENTATIONSMODELS

(57) Abstract: The invention relates to a computer-implemented method for creating a bioreaction model, in particular a
fermentation or whole cell catalytic process, using an organism based on measurement data.

(57) Zusammenfassung: Die Anmeldung betrifft ein computerimplementiertes Verfahren zur Erstellung eines Modells einer
Bioreaktion - Fermentations- oder Ganzzellkatalyseprozess - mit einem Organismus auf Basis von Messdaten.



WO 2016/120361 A1

Computerimplementiertes Verfahren zur Erstellung eines Fermentationsmodells

Die Erfindung betrifft ein Computerimplementiertes Verfahren zur Erstellung eines Modells einer Bioreaktion – insb. Fermentation oder Ganzzellkatalyse – unter Verwendung eines Organismus.

Organismus im Sinne der Anmeldung sind Kulturen aus pflanzlichen oder tierischen Zellen wie Säugetierzellen, Hefen, Bakterien, Algen, usw., die in Bioreaktionen eingesetzt werden.

Die sensorische Überwachung eines Fermentationsprozesses und Analyse von Proben aus einem Prozess z. B. mit Hilfe der Quality-by-Design Analytik-Automatisierungsplattform BaychromAT® der Bayer Technology Services GmbH liefert verschiedene Informationen über den Zustand des Prozesses in dem Bioreaktor in Echtzeit. Typischerweise werden Zellenanzahl, Zellvitalität, Konzentrationen von Substraten, wie Kohlenstoffquellen (z. B. Glukose), Aminosäuren oder O₂, Produkten und Nebenprodukten (z. B. Laktat oder CO₂), Prozessparameter wie Temperatur und/oder pH-Wert oder Produktmerkmale ermittelt. Diese Daten können durch berechnete Daten und/oder Extrapolationen z. B. aus dem Stand der Technik ergänzt werden. Zusammen bilden diese Daten die Messdaten oder das Prozesswissen im Sinne der Anmeldung.

Hintergrundwissen zu dem Organismus bedeutet im Sinne der Anmeldung Wissen über die biochemischen Reaktionen des Organismus - spezifische und unspezifische Reaktionen - und insbesondere die zellinternen Reaktionen, bzw. Makroreaktionen zur Beschreibung der organismusspezifischen Stoffwechselnetzwerke (auch SN oder metabolische Netzwerke genannt), die aus Substraten, Metaboliten (auch Knoten des metabolischen Netzwerks genannt), Produkten sowie den biochemischen Reaktionen zwischen ihnen bestehen. Diese biochemischen Reaktionen sind durch ihre:

- (a) Stöchiometrie
 - (b) Reversibilität (unter biologischen Bedingungen),
 - (c) Einbindung in ein stöchiometrisches Netzwerk,
- definiert.

Bisher werden die Messdaten hauptsächlich für die qualitative Überwachung des Prozesses eingesetzt. Im Folgenden wird eine Auswahl von technischen Fragestellungen dargestellt, für deren Lösung dynamische Prozessmodelle erforderlich sind.

Eine technische Verwendung des Prozesswissens im Sinne der Anmeldung bietet die modellbasierte Zustandsschätzung eines Prozesses in einem Bioreaktor. Methoden wie das „extended Kalman Filter“ ermöglichen eine kontinuierliche Abschätzung von Prozessgrößen, über die diskontinuierlich Messungen vorliegen [Welch G, Bishop G. 1995. An Introduction to the Kalman Filter. Chapel Hill, NC, USA: University of North Carolina at Chapel Hill.]. Auch der Verlauf nicht messbarer Größen kann aus anderen Messungen berechnet werden. Voraussetzung hierfür ist ein den zugrunde liegenden Prozess richtig beschreibendes Prozessmodell.

Eine weitere Anwendung ist die modellbasierte, optimale Prozessführung. Hierbei wird ein dynamisches Prozessmodell verwendet, um in einem modellbasierten, prädiktiven geschlossenen Regelkreis die Prozessführung hinsichtlich Produktmenge, Produktmerkmalen oder Bildung von Nebenprodukten oder anderen Zielgrößen zu optimieren. Dieses wird beispielsweise von Frahm et al. für eine Hybridoma Zellkultur gezeigt [Frahm B, Lane P, Atzert H, Munack A, Hoffmann M, Hass VC, Portner R. 2002. Adaptive, Model-Based Control by the Open-Loop-Feedback-Optimal (OLFO) Controller for the Effective Fed-Batch Cultivation of Hybridoma Cells. Biotechnol. Prog. 18(5):1095–1103].

Für beide genannten technischen Anwendungen ist es wichtig, dass das erstellte Prozessmodell eine möglichst geringe Komplexität, also eine begrenzte Anzahl an Zustandsvariablen und/oder Gleichungen, bei gleichzeitig guter Genauigkeit der Wiedergabe des Prozesses aufweist.

Neben den genannten Anwendungen zur Prozessführung können dynamische Prozessmodelle auch während der Prozessentwicklung genutzt werden um Experimente mit optimalem Informationsgewinn zu planen. Dieses Vorgehen wird als modellbasierte Versuchsplanung bezeichnet [Franceschini G, Macchietto S. 2008. Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science 63(19):4846–4872]. Neben den oben genannten Voraussetzungen für die Komplexität des Modells ist es für diese technische Anwendung erforderlich, dass ein dynamisches Prozessmodell bereits während der Entwicklungsphase vorhanden ist. Dieses sollte möglichst schnell aus bereits vorhandenem Prozesswissen erstellt werden können, um den Zeitaufwand der Prozessentwicklung gering zu halten.

Es bestand daher der Bedarf, ein Verfahren bereit zu stellen, dass die Erstellung eines dynamischen Prozessmodells unter der Verwendung von Hintergrundwissen und Prozesswissen ermöglicht. Um dieses Modell zum Beispiel für eine Zustandsschätzung, eine optimale Prozessführung oder zur modellbasierten Versuchsplanung nutzen zu können muss hierzu die Komplexität des Modelles gering sein. Abhängigkeiten, also Einflüsse der Prozessgrößen oder des Prozesszustandes auf das Prozessverhalten, sollen innerhalb des Design Space hinreichend genau quantifiziert sein. Alle

vorliegenden Informationen über den Prozesszustand sollen hierfür genutzt werden. Die modellbasierte Beschreibung von Produktmerkmalen soll bei Bedarf in das Modell integrierbar sein. Als Design Space wird der Bereich bezeichnet indem Prozesswissen vorliegt. Das Verfahren soll auf oben aufgeführte Bioreaktionen anwendbar sein und die Entwicklungszeit solcher dynamischen Modelle wesentlich verkürzen. Bisherige Ansätze zur Entwicklung von dynamischen Modellen benötigen Monate bis Jahre bis zur Fertigstellung eines Prozessmodells. Der vorliegende Ansatz reduziert erfahrungsgemäß die Entwicklungszeit auf wenige Wochen.

Typische Produktmerkmale im Sinne der Anmeldung sind beispielsweise Glykosylierungsmuster von Proteinen oder die Proteinintegrität, ohne sich darauf zu begrenzen. Dynamische Modelle, die in dem oben genannten Kontext verwendet werden, weisen diese Eigenschaft bislang nicht auf. Der vorliegende Ansatz ermöglicht eine einfache modellbasierte Integration von Produktmerkmalen.

Die modellbasierte Prozessführung von Fermentationen wird von Frahm et. al am Beispiel einer Hybridoma Zellkultur gezeigt (Frahm B, Lane P, Atzert H, Munack A, Hoffmann M, Hass VC, Portner R. 2002. Adaptive, Model-Based Control by the Open-Loop-Feedback-Optimal (OLFO) Controller for the Effective Fed-Batch Cultivation of Hybridoma Cells. *Biotechnol. Prog.* 18(5):1095–1103). Grundlegende Prozessgrößen werden hier modellbasiert gesteuert. Eine Integration von Produktmerkmalen findet hier nicht statt. Das mathematische Modell der Zelle wurde für diesen spezifischen Prozess entworfen und lässt sich nur mit großem Aufwand auf Prozesse mit dem gleichen oder anderen Organismen oder Stämmen des gleichen Organismus übertragen. Hintergrundwissen in Form von zellinternen Reaktionen wird im Modell nicht explizit berücksichtigt. Eine Integration weiterer Messgrößen in das Modell und somit eine vollständige Nutzung der Informationen über den Prozesszustand kann hier nur mit sehr großem Aufwand erfolgen. Der Ansatz stellt somit eine individuelle Lösung dar, die weder auf andere Prozesse übertragbar ist, noch die volle Nutzung der ermittelten Daten ermöglicht. Die genannte Methode löst aufgrund der zu erwartenden Entwicklungszeit des Modells und der aufwändigen Übertragbarkeit der Lösung auf andere Prozesse mit den gleichen oder mit anderen Organismen das oben genannte technische Problem nicht.

Eine weitergehende Modellierung, die auch Produktmerkmale wie die Glykosylierung mit einschließt, findet sich in den Veröffentlichungen von Kontoravdi et. al. Das Modell, das den Hauptstoffwechsel beschreibt, bezieht kein Hintergrundwissen in Form von zellinternen Reaktionen mit ein und lässt sich auch nicht auf andere Prozesse mit dem gleichen oder anderen Organismen übertragen. Eine Integration weiterer Messgrößen in das Modell kann hier nicht erfolgen [Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. 2007. Development of a dynamic model of monoclonal antibody

production and glycosylation for product quality monitoring. *Computers & Chemical Engineering* 31(5-6):392–400.]. Diese Methode ermöglicht auch keine vollständige Nutzung der Informationen über den Prozesszustand, erfordert eine lange Entwicklungszeit des Modells und ist auf andere Organismen oder Stämme nicht übertragbar. Diese Methode stellt somit keine Lösung des technischen Problems dar.

Die Modellierungen der Glykosylierung unter Einbindung des Nucleotidzuckerstoffwechsels von Jedrzejewski et al. und Jimenez et al. beziehen Hintergrundwissen in Form von Bilanzgleichungen interner Stoffwechselintermediate mit ein [Jedrzejewski PM, del Val, Ioscani Jimenez, Constantinou A, Dell A, Haslam SM, Polizzi KM, Kontoravdi C. 2014. Towards Controlling the Glycoform: A Model Framework Linking Extracellular Metabolites to Antibody Glycosylation. *International journal of molecular sciences* 15(3):4492–4522.; Jimenez del Val, Ioscani, Nagy JM, Kontoravdi C. 2011. A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnology progress* 27(6):1730–1743]. Bei einer Verwendung dieses Modells für die Prozessführung sind die Komplexität des gesamten Modells und die mangelnde Beobachtbarkeit zellinterner Stoffwechselintermediate aber nachteilig. Zudem lässt das Modell des Hauptstoffwechsels keine Übertragung auf andere Prozesse oder die vollständige Nutzung der Informationen über den Prozesszustand zu. Diese Methode stellt somit keine Lösung des technischen Problems dar.

Eine flexible Modellgenerierung für Bioprozesse wird von Leifheit et al. adressiert [Leifheit J, Heine T, Kawohl M, King R. 2007. Rechnergestützte halbautomatische Modellierung biotechnologischer Prozesse (Semiautomatic Modeling of Biotechnical Processes). *at – Automatisierungstechnik* 55(5)]. Die Modellgenerierung erfolgt unter Zuhilfenahme von Prozesswissen, allerdings ohne Hintergrundwissen. Die Prozedur ist für verschiedene Prozesse mit dem gleichen oder anderen Organismen verwendbar. Basis sind hier Makroreaktionen, die der Anwender selbst vorgibt. Deren genaue Stöchiometrien werden in dem Verfahren ermittelt. Die Methode wird für eine geringe Anzahl von Zustands- bzw. Messgrößen beschrieben. Eine Integration weiterer Zustands- bzw. Messgrößen wäre mit einem signifikanten Anstieg der Komplexität des Verfahrens verbunden. Bei einer Verwendung von umfassenden Datengrundlagen, wie sie beispielsweise durch die BaychromAT®-Plattform bereitgestellt wird, wäre diese Methode nicht mehr durchführbar. Eine Integration von Produktmerkmalen ermöglicht die Methode nicht. Sie stellt somit keine Lösung des oben genannten technischen Problems dar.

Die Verwendung des Hintergrundwissens in Form von Makroreaktionen, die als Elementary Modes (EM) aus den bekannten metabolischen (stöchiometrischen) Netzwerken eines Organismus gewonnen werden, wird von Provost beschrieben [Provost A. 2006. Metabolic design of dynamic bioreaction models. Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve, Louvain-la-Neuve, S. 81 ff., S.107 ff. S. 118 ff.]. Diese Methode ist für verschiedene Organismen oder Stämme des gleichen Organismus verwendbar. Die Auswahl der Makroreaktionen für das Prozessmodell erfolgt unter Verwendung von Prozesswissen. Jedoch werden Prozessabschnitte definiert für die eine vordefinierte Anzahl von Makroreaktionen separat zufällig ausgewählt werden. Das beschriebene Verfahren liefert hier eine von vielen möglichen Kombinationen von Elementary Modes. Die Anzahl der Makroreaktionen, und somit die Modellkomplexität, ist festgelegt und nicht veränderbar. Das Verfahren ergibt separate Modelle für jeden Prozessabschnitt. Die Auswahl der Kinetiken der einzelnen Makroreaktionen erfolgt unter Berücksichtigung der Stöchiometrie der gewählten Makroreaktionen. Die Parameter der Kinetiken (Modellparameter) werden jedoch nicht an die Prozessdaten angepasst. Stattdessen bildet die Verwendung der separaten Prozessabschnittsmodelle die Änderungen in den Prozessdaten ab. Die zufällige Auswahl der Reaktionen kann zwar auch auf Basis einer umfassenden Datengrundlage erfolgen, allerdings können durch den beschriebenen Ansatz zur Auswahl der Kinetiken und die ausgewählten Kinetiken den Verlauf des Prozesses bzw. das Verhalten des Organismus im Prozess nicht abbilden. Die Verwendung mehrerer Prozessabschnittsmodelle führt zudem zu einer unnötigen Komplexitätssteigerung des Prozessmodells. Die Abhängigkeiten, also Einflüsse von Prozessgrößen oder des Prozesszustandes auf das Prozessverhalten, werden mit dieser Methode nicht quantifiziert. Auch findet hier eine Integration von Produktmerkmalen nicht statt. Diese Methode stellt somit keine Lösung des o. g. technischen Problems dar.

Es bestand daher der Bedarf ein Verfahren bereit zu stellen, das die schnelle und effiziente Bereitstellung eines Modells auf Basis von Prozesswissen und -Messdaten und die Optimierung des Produktumsatzes und der kritischen Produktmerkmale unter Berücksichtigung von Hintergrundwissen ermöglicht und die oben genannten Nachteile nicht aufweist.

Die Aufgabe wurde durch ein Verfahren zur Erstellung eines Modells einer Bioreaktion mit einem Organismus in einem Bioreaktor wie folgend beschrieben gelöst.

Gegenstand der Anmeldung ist ein computerimplementiertes Verfahren zur Erstellung eines Modells einer Bioreaktion – insb. Fermentation oder Ganzzellkatalyse - mit einem Organismus, das folgende Schritte umfasst:

- a. Ausgewählte Stoffwechselwege des Organismus, deren Stöchiometrie- sowie Reversibilitätseigenschaften werden als Hintergrundwissen in das Verfahren eingegeben. Mit anderen Worten werden ein oder mehrere metabolische Netzwerke des Organismus in das Verfahren eingegeben. Elementary Modes (EMs) werden aus dieser Eingabe berechnet.
- b. Die EMs werden in einer Matrix K zusammengefasst, wobei die EMs die Stoffwechselwege aus a) in Makroreaktionen zusammenfassen. Diese Matrix K enthält hiermit die Stöchiometrie und die Reversibilitätseigenschaften aller möglichen Makroreaktionen aus dem Hintergrundwissen.
- c. Die Messdaten (auch Prozesswissen genannt) zur Bioreaktion mit dem Organismus werden eingegeben.
- d. Mit Hilfe einer Interpolationsmethode werden auf Basis der eingegebenen Messdaten aus c) die für den Organismus spezifischen Raten - Ausscheide- und Aufnahmearten von einer oder mehreren Eingangsgrößen und Ausgangsgrößen - der eingegebenen Stoffwechselwege berechnet. Bevorzugt werden auch Wachstumsraten, besonders bevorzugt auch Absterberaten des Organismus berechnet.
- e. Relevante Makroreaktionen werden in Form eines Subsets der Elementary Modes aus b) ausgewählt durch
 - i. Datenunabhängige und/oder datenabhängige Vorreduktion der Anzahl der EMs aus b).
 - ii. Auswahl des Subsets aus der Vorreduktion aus e) i. mit den Messdaten aus c) und / oder einer oder mehrerer Raten aus d), bevorzugt mit dem Messdaten aus c) mittels eines Algorithmus nach einem mathematischen Gütekriterium und Zusammenfassung des Subsets in einer Matrix L .
 - iii. Optional erfolgt eine Darstellung des Subsets grafisch.
- f. Mit Hilfe einer Interpolationsmethode werden auf Basis der eingegebenen Messdaten aus c) und / oder der Raten aus d), die Reaktionsraten der Makroreaktionen des Subsets $\underline{r}(t)$ berechnet.
- g. Kinetiken der Makroreaktionen des Subsets aus e) ii. werden mit folgenden Zwischenschritten entworfen; dadurch werden die Modellparameter definiert.
 - i. Aus der Stöchiometrie der Makroreaktionen werden generische Kinetiken entworfen.
 - ii. Einflussgrößen auf die Makroreaktionen werden aus den Reaktionsraten aus f) ermittelt.

- iii. Die generischen Kinetiken aus g) i. werden um Terme erweitert, die die in g) ii. ermittelten Einflussgrößen quantifizieren.
- h. Optional erfolgt für die Kinetiken aus g) eine erste Anpassung der Modellparameterwerte für jede Makroreaktion separat an die berechneten Reaktionsraten aus f) und eine Überprüfung der Anpassungsgüte.
- i. Optional werden die Schritte g) und h) wiederholt bis eine vordefinierte Anpassungsgüte erreicht wird.
- j. Die Modellparameterwerte werden an die Messdaten aus c) angepasst.
- k. Die Matrix L , die Kinetiken aus g) und die Modellparameterwerte aus j) bilden das Modell und werden ausgegeben und / oder in ein Prozessführungs- oder Prozessentwicklungsmodul übertragen.

Typischerweise kommuniziert das Prozessführungsmodul on-line mit einem Prozessleitsystem, das üblicherweise zur Steuerung des Bioreaktors verwendet wird.

Typischerweise werden Prozessentwicklungsmodule zur off-line Optimierung des Prozesses oder zur Planung von Experimenten verwendet.

Die erfindungsgemäße Modellierung der Bioreaktion baut im Wesentlichen auf der Annahme repräsentativer Makroreaktionen auf, die interne Stoffwechselforgänge vereinfacht darstellen. Für die Auswahl der Reaktionen wird sowohl biochemisches Hintergrundwissen, als auch Prozesswissen benötigt.

Im ersten Schritt des Verfahrens erfolgt die Eingabe der Reaktionen des metabolischen Netzwerkes, deren Stöchiometrie und Reversibilitätseigenschaft durch den Nutzer über eine Nutzeroberfläche oder idealerweise automatisch durch die Selektion eines Organismus und seiner hinterlegten Stoffwechselwege aus einem Datenbankmodul, in dem das Hintergrundwissen zum Organismus gespeichert ist. Das metabolische Netzwerk (im Stand der Technik auch stöchiometrische Netzwerk benannt) und die Eigenschaften seiner einzelnen Reaktionen stellen das Hintergrundwissen des Organismus dar. Bevorzugt enthält das metabolische Netzwerk Reaktionen aus für den Organismus wichtigen Stoffwechselwegen, beispielsweise Reaktionen der Glykolyse. Besonders bevorzugt enthält die Auswahl externe Reaktionen. Eine externe Reaktion im Sinne der Anmeldung enthält mindestens eine Komponente außerhalb der Zelle, typischerweise mindestens eine Eingangsgröße und/oder mindestens eine Ausgangsgröße (Produkt, Nebenprodukt, usw.). Besonders bevorzugt enthält das metabolische Netzwerk Reaktionen, die das Zellwachstum beschreiben, z. B. in Form einer

vereinfachten Reaktion von internen Metaboliten zur externen Biomasse. Abb. 5 und Tabelle 1 im Beispiel beschreiben ein anwendbares metabolisches Netzwerk, ohne sich darauf zu begrenzen.

Dann werden aus den eingegebenen, in einem oder mehreren stöchiometrischen Netzwerken zusammengefassten, Stoffwechselwegen des Organismus Elementary Modes berechnet. Jeder Elementary Mode ist eine Linearkombination von Reaktionsraten aus den Stoffwechselwegen – d. h. internen und externen Reaktionen des metabolischen Netzwerkes, bei der sowohl die „steady state“ Bedingung für interne Metabolite erfüllt, als auch die Reversibilität bzw. Irreversibilität von Reaktionen berücksichtigt sind. Bei Linearkombinationen von Reaktionen, die die „steady state“ Bedingung für interne Metabolite berücksichtigen, können sich keine internen Metabolite akkumulieren.

Eine interne Reaktion im Sinne der Anmeldung erfolgt ausschließlich innerhalb der Zelle.

Durch eine Externalisierung einer internen Komponente, d. h. durch eine Klassifizierung einer eigentlich internen Komponente als Ein- oder Ausgangsgröße, ist es möglich, die mit der externalisierten, internen Komponenten verbundene interne Reaktion als externe Reaktion zu modellieren und somit die „steady-state“ Bedingung für interne Metabolite in diesem Fall zu umgehen. Eine Makroreaktion im Sinne der Anmeldung fasst alle Reaktionen zusammen, die von einer oder mehreren Eingangsgrößen zu einer oder mehreren Ausgangsgröße(n) führen. Jeder Elementary Mode beschreibt somit eine Makroreaktion. Im Vergleich zu der Methode von Leifheit, et al werden im Sinne der Anmeldung die Makroreaktionen auf Basis des eingegebenen Hintergrundwissen ermittelt.

Die Elementary Modes (EMs) werden in einer Matrix E zusammengefasst, vorzugsweise in einem Modul zum Matrixaufbau, das mit einem entsprechenden Algorithmus konfiguriert ist. Bekannte Algorithmen können zum Aufbau der Elementary Modes Matrix verwendet werden. METATOOL wird als Beispiel genannt, ohne sich darauf zu begrenzen: [Pfeiffer T, Montero F, Schuster S. 1999. METATOOL: for studying metabolic networks. Bioinformatics 15(3):251-257.]

Mit METATOOL wird eine erste Matrix E , die die eingegebenen internen und externen Reaktionen beschreibt, erzeugt.

In Schritt b) wird mit Hilfe der (externen) stöchiometrischen Matrix N_p aus der Matrix (E) eine Matrix bestehend aus möglichen Makroreaktionen K erzeugt.

$$K = N_p \cdot E \quad (\text{Formel 1})$$

Die Transformation der Matrix E nach K ist aus Provost bekannt [Provost A. 2006. Metabolic design of dynamic bioreaction models. Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve, Louvain-la-Neuve, S. 81].

Die Spaltenvektoren der Matrix K beschreiben die Makroreaktionen. Die Zeilenvektoren beschreiben die Komponenten der Makroreaktionen (Eingangs- und Ausgangsgrößen). In die Matrix K wird die Stöchiometrie der Makroreaktionen eingetragen.

Jede im Sinne des metabolischen Netzwerks mögliche Reaktionsrate kann als positive Linearkombination dieser Makroreaktionen dargestellt werden.

Die Verwendung von EMs als Basis eines Prozessmodells ist im Stand der Technik bekannt (z. B. aus Provost A. 2006. Metabolic design of dynamic bioreaction models. Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve, Louvain-la-Neuve S. 87, S. 118 ff. und Gao, J. et al. (2007). Dynamic metabolic modeling for a MAB bioprocess. *Biotechnology progress*, 23(1), 168-181).

In einem weiteren Schritt c) werden die verfügbaren Messdaten (Prozesswissen) zur Bioreaktion mit dem Organismus eingegeben. Typischerweise werden Zellenanzahl, Zellvitalität, Konzentrationen von Substraten, wie Kohlenstoffquellen (z. B. Glukose), Aminosäuren oder O_2 , Produkten und Nebenprodukten (z. B. Laktat oder CO_2), Prozessparameter wie Temperatur und/oder pH-Wert oder Produktmerkmale ermittelt. Diese Eingabe kann manuell durch den Nutzer oder automatisch erfolgen, wie z. B. durch Selektion aus einem Datenbankmodul zur Speicherung von Messdaten und Transfer der selektierten Daten in ein Datenanalysemodul, das mit dem Datenbankmodul verbunden ist.

Aus diesen Messdaten werden im Schritt d) die zellspezifischen Ausscheide- und Aufnahmearten von Substraten und (Neben)produkten - zusammen spezifische Raten $\underline{q}(t)$ - genannt, sowie optional die Wachstums- und Absterberaten des Organismus ($\mu(t)$, $\mu_d(t)$) berechnet. Voraussetzung für die Berechnung ist die Interpolation der vitalen Zellzahl, der Gesamtzellzahl und Medienkonzentrationen mit Hilfe einer Interpolationmethode. Aus diesen können zeitliche Änderungen der Messgrößen ermittelt werden. Die berechneten Raten $\underline{q}(t)$, $\mu(t)$, $\mu_d(t)$ geben Auskunft über das zu beobachtende dynamische Verhalten des Organismus über der Zeit.

Für die Berechnung der o. g. Raten kann eine oder mehrere unterschiedliche Interpolationsmethode(n) in Kombination verwendet werden.

Exemplarisch beschreibt Leifheit et al. die Ermittlung der zeitlichen Änderungen von Messgrößen - z. B. die der Gesamtzellzahl, die der vitalen Zellzahl oder die von anderen Medienkonzentrationen aus Messdaten mit Hilfe von Spline-interpolierten Messdaten [Leifheit, J., Heine, T., Kawohl, M., & King, R. (2007). Rechnergestützte halbautomatische Modellierung biotechnologischer Prozesse (Semiautomatic Modeling of Biotechnical Processes). *at-Automatisierungstechnik*, 55(5), 211-218]. Diese Methode wird hiermit per Referenz in die Anmeldung integriert.

Die o. g. Raten $q(t)$, $\mu(t)$, $\mu_d(t)$ werden aus diesen zeitlichen Änderungen berechnet:

Beispielsweise kann die Wachstumsrate des Organismus $\mu(t)$ aus Spline-interpolierten Werten der Gesamtzellzahl $X_t(t)$ und der Lebendzellzahl $X_v(t)$ sowie der daraus berechenbaren zeitlichen Änderung der Gesamtzellzahl $\frac{dX_t}{dt}(t)$ mit Formel 2 berechnet werden:

$$\frac{dX_t}{dt}(t) = -D(t) \cdot X_t(t) + \mu(t) \cdot X_v(t) \quad (\text{Formel 2})$$

Wobei $D(t)$ die Verdünnungsrate ist.

Die Absterberate $\mu_d(t)$ kann bei bekanntem Verlauf von $\mu(t)$ aus dem Verlauf von $X_v(t)$ und dem Verlauf der zeitlichen Änderung der vitalen Zellzahl $\frac{dX_v}{dt}(t)$ mit Formel 3 berechnet werden:

$$\frac{dX_v}{dt}(t) = -D(t) \cdot X_v(t) + (\mu(t) - \mu_d(t)) \cdot X_v(t) \quad (\text{Formel 3})$$

Die Berechnung der spezifischen Raten einer anderen Komponente i $q_i(t)$ kann aus Spline-interpolierten Werten der Lebendzellzahl $X_v(t)$ und der Konzentration der Komponente $C_i(t)$, sowie dem Verlauf der zeitlichen Änderung $\frac{dC_i}{dt}(t)$, der aus Spline-interpolierten Werten von $C_i(t)$ ermittelt werden kann, mit Formel 4 erfolgen:

$$\frac{dC_i}{dt}(t) = D(t) \cdot (C_{i,in} - C_i(t)) + q_i(t) \cdot X_v(t) \quad (\text{Formel 4})$$

In einer bevorzugten Ausführungsform des Verfahrens werden die Messdaten aus Schritt c) vor der ersten Interpolation folgendermaßen aufbereitet: Um alle nicht von den Zellen verursachten Konzentrationsänderungen zu berücksichtigen und aus den Messdaten einen stetigen Verlauf der Konzentrationsänderungen zu erhalten, werden die Messdaten verschoben (in der Anmeldung *shiften* genannt). Der Betrag $\Delta C_i(t)$, um den die Konzentrationsmessung verschoben wird, kann nach Formel 5 berechnet werden:

$$\Delta C_i(t) = \int_0^t D(\tau) \cdot (C_i(\tau) - C_{i,in}(\tau)) d\tau \quad (\text{Formel 5})$$

Wobei $D(\tau)$ die Verdünnungsrate ist. Der geshiftete Konzentrationsverlauf $C_{i,s}(t)$ ergibt sich dann nach Formel 6:

$$C_{i,s}(t) = C_i(t) - \Delta C_i(t) \quad (\text{Formel 6})$$

Die Differentialgleichung, die den Verlauf der geshifteten Konzentration $C_{i,s}(t)$ angibt, ergibt sich somit aus den Formeln 4 und 6 zu:

$$\frac{dC_{i,s}}{dt} = \frac{d(C_i(t) - \Delta C_i(t))}{dt} = q_i(t) \cdot X_v(t) \quad (\text{Formel 7})$$

Diese Aufbereitung (shiften) der Messdaten verhindert eine sprunghafte Änderung der berechneten spezifischen Raten beim An- oder Abstellen eines Feeds (Feedpeak) insbesondere in einem Fed-Batch Prozess.

Abbildung 1 zeigt das Aufbereitung/Shiften der Messdaten im Sinne dieser Anmeldung.

In einer besonderen Ausführungsform des Verfahrens werden die aufbereiteten Daten dann zur Berechnung eines Gradienten der Gesamtzellzahl $\left. \frac{dX_t}{dt} \right|_s(t)$ mit der Methode von Leifheit et al. verwendet. Dieser wird mit einer Spline-Interpolation gemäß Differentialgleichung 8 angenähert:

$$\left. \frac{dX_t}{dt} \right|_s(t) = \mu(t) \cdot X_v(t) \quad (\text{Formel 8})$$

Besonders bevorzugt wird die Lysis in der Berechnung anhand eines Lysis-Faktors K_l , einbezogen (Formel 9). Dieser kann z. B. als konstant über dem Prozessverlauf angenommen werden.

$$\left. \frac{dX_t}{dt} \right|_s = \mu(t) \cdot X_v(t) - K_l \cdot (X_t(t) - X_v(t)) \quad (\text{Formel 9})$$

Das Absinken der geshifteten Gesamtzellzahl $X_{t,s}(t)$ kann so durch Lysis erklärt werden, wodurch negative Werte für die Wachstumsrate $\mu(t)$ vermieden werden können.

Bevorzugt werden die aufbereiteten Daten auch für die Berechnung der Absterberate $\mu_d(t)$ verwendet.

Bevorzugt werden die möglichen Kombinationen spezifischer Raten $\underline{q}(t)$ in einem Flux-Map Diagramm wie beispielhaft in Abbildung 2 dargestellt. Diese Darstellung ermöglicht eine gute Übersicht über die berechneten spezifischen Raten $\underline{q}(t)$. Die Höhenlinien geben hier an, welche Bereiche physiologisch wichtig sind.

Wenn die spezifischen Raten $\underline{q}(t)$ und optional die weiteren Raten $\mu(t)$ und $\mu_d(t)$ unterschiedliche Größenordnungen und Einheiten besitzen, werden diese üblicherweise mit Hilfe von Vereinfachungen zu einem spezifischen Ratenvektor $\underline{\tilde{q}}(t)$ mit gleichen Einheiten zusammengefasst. Beispielsweise wird die spezifische Rate eines Makromoleküls, das in Gramm [g] gemessen wird, die Einheit $\left[\frac{g}{\text{Cell} \cdot h} \right]$ besitzen. Wenn die Zusammensetzung dieses Makromoleküls geschätzt wird, z. B. basierend auf

seinem C-mol-Gehalt, lässt sich seine spezifische Rate von $[g]$ auf $[C - mol]$ verändert darstellen, so dass die spezifische Rate die Einheit $\left[\frac{C-mol}{Cell \cdot h}\right]$ besitzt.

Die spezifischen Raten $\underline{\tilde{q}}(t)$ bilden eine der Grundlagen für den weiteren Schritt e) des Verfahrens, nämlich die Auswahl der relevanten Makroreaktionen.

Im Schritt e) wird ein Subset (L) der EMs auf Basis der Daten ausgewählt, mit dem die spezifischen Raten $\underline{\tilde{q}}(t)$ aus d) und / oder die Messdaten aus c) nach einem mathematischen Gütekriterium gut abgebildet werden können. Die Anzahl der EMs in dem Subset (L) soll möglichst klein sein, um eine möglichst geringe Komplexität des Prozessmodells zu gewährleisten. Das Subset L soll aber eine gute Beschreibung des Prozesswissens sicherstellen.

Die Selektion von EMs verkleinert den Lösungsraum im Vergleich zum original EMs-Set (K) aus a), aber enthält weiterhin den ermittelten physiologisch wichtigen Bereich der Zellen.

Abbildung 3 zeigt eine Darstellung des Lösungsraums, wobei das originale Set von EMs (K) zu einem Subset (L) reduziert wird.

Für den Schritt e) werden die berechneten spezifischen Raten $\underline{\tilde{q}}(t)$ sowie die Messdaten aus c) üblicherweise in ein Modul zur Auswahl der relevanten Makroreaktionen transferiert, das mit entsprechenden Algorithmen konfiguriert ist.

Im Schritt e) i) erfolgt eine datenunabhängige und/oder eine datenabhängige Vorreduktion der Matrix K in beliebiger Reihenfolge:

Die datenunabhängige Vorreduktion erfolgt vorzugsweise durch eine geometrische Reduktion. Dabei werden für einen zufällig ausgewählten EM alle Kosinus-Ähnlichkeiten zu allen anderen Modes berechnet. Der EM mit der höchsten Ähnlichkeit wird aus dem Set entfernt. Diese Vorgehensweise wird wiederholt, bis eine vordefinierte Anzahl an EMs erreicht ist. Die gewünschte Anzahl wird üblicherweise für das Verfahren im Vorfeld definiert. Als Kontrollvariable kann das Volumen des Lösungsraumes verwendet werden. Überraschenderweise wurde festgestellt, dass eine deutliche Reduktion der Anzahl der Makroreaktionen unter Beibehaltung von 90 bis 98 %, bevorzugt 92 bis 95% des aufgespannten Volumens, im Vergleich zum Originalvolumen möglich ist.

Die datenabhängige Vorreduktion kann durch den Vergleich von Ausbeutekoeffizienten der EMs (Y^{EM}) zu den Ausbeutekoeffizienten, die aus den spezifischen Raten $\underline{\tilde{q}}(t)$ aus d) berechnet werden (Y^m) erfolgen. Der Ausbeutekoeffizient des k-ten EMs ($Y_{i,j}^{EM,k}$) wird nach Formel 10 durch die

Division der entsprechenden stöchiometrischen Koeffizienten der externen Metaboliten i und j ermittelt. Für den k -ten EM sind dies die Matrixeinträge $K_{i,k}$ und $K_{j,k}$.

$$Y_{i,j}^{EM,k} = \frac{K_{i,k}}{K_{j,k}} \quad (\text{Formel 10})$$

Ist der stöchiometrische Koeffizient $K_{j,k} = 0$, kann der Ausbeutekoeffizient nicht ermittelt werden.

Der Ausbeutekoeffizient $Y_{i,j}^m(t)$ gibt das Verhältnis zweier gemessener oder nach d) berechneter zellspezifischen Raten ($\tilde{q}_i(t), \tilde{q}_j(t)$) zueinander nach Formel 11 an:

$$Y_{i,j}^m(t) = \frac{\tilde{q}_i(t)}{\tilde{q}_j(t)} \quad (\text{Formel 11})$$

Aus den Ausbeutekoeffizienten Y^m können für jede mögliche Kombination zweier externer Komponenten i und j eine obere und eine untere Grenze bestimmt werden. Beispielsweise kann der kleinste Ausbeutekoeffizient zweier externer Metabolite i und j $Y_{i,j}^m(t)$ als untere Grenze und der größte Wert von $Y_{i,j}^m(t)$ als obere Grenze verwendet werden, andere Grenzen sind jedoch auch möglich. EMs, deren Ausbeutekoeffizienten $Y_{i,j}^{EM}$ oberhalb der oberen Grenze oder unterhalb der unteren Grenze liegen, werden aus der Matrix K entfernt. Wenn der Ausbeutekoeffizient eines EMs $Y_{i,j}^{EM}$ nicht ermittelt werden kann, verbleibt dieser in der Matrix K . Bevorzugt kann auch die auf Seite 15 beschriebene, erfindungsgemäße Methode „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen mit NNLS“ zur datenabhängigen Vorreduktion verwendet werden. Eine Erläuterung der Methode im Kontext der Vorreduktion findet sich dort. Das Heranziehen der Prozessdaten in der datenabhängigen Vorreduktion in dem Verfahren hat den Vorteil, dass die Reduktion prozessbezogen und damit effektiver und fokussierter erfolgt.

Im Schritt e) ii. wird ein Subset aus Makroreaktionen mit einem Algorithmus ausgewählt: Für die Auswahl wird ein Gütekriterium, mit dem quantifiziert werden kann, wie gut die spezifischen Raten $\tilde{q}(t)$ aus d) und / oder die Messdaten aus c) mit einem Subset (L) abgebildet werden können, sowie ein Algorithmus zur Auswahl des Subsets benötigt.

Als Gütekriterium für die Abbildung von berechneten spezifischen Raten $\tilde{q}(t)$ mit einem Subset L kann nach Soons et al. die Summe der quadrierten Residuen der spezifischen Raten (SSR_q) nach Formel 12 verwendet werden [Soons, Z. I. T. A., Ferreira, E. C., Rocha, I. (2010). Selection of Elementary Modes for Bioprocess Control. 11th International Symposium on Computer Applications in Biotechnology, Leuven, Belgium, July 7-9, 2010, 156-161].

$$SSR_q = \sum_{i=1}^{N_t} \left\| L \cdot \underline{r}(t_i) - \underline{\tilde{q}}(t_i) \right\| \quad (\text{Formel 12})$$

Der Wert für SSR_q soll möglichst klein sein.

Für die Minimierung von SSR_q muss zuvor für jeden betrachteten Zeitpunkt t_i der Vektor $\underline{r}(t_i)$ mit Hilfe eines non-negative-Least-squares Algorithmus so bestimmt werden, dass gilt:

$$\min_{\underline{r}(t_i)} \left(\left\| L \cdot \underline{r}(t_i) - \underline{\tilde{q}}(t_i) \right\| \right) \quad (\text{Formel 13})$$

mit der zusätzlichen Randbedingung:

$$\underline{r}(t_i) \geq 0 \quad (\text{Formel 14})$$

Vorteil dieser Methode ist, dass die Berechnungen nach Formel 12 – 14 auch für sehr große Subsets mit vielen EMs durchgeführt werden können. Ein signifikanter Nachteil ist, dass für diese Berechnung die berechneten spezifischen Raten $\underline{\tilde{q}}(t)$ erforderlich sind. Da diese aus interpolierten Messwerten gewonnen werden, liegen diese mit großer Unsicherheit bezüglich ihren wahren Werte vor. Messungenauigkeiten können sich unter Umständen sehr stark auf die berechneten spezifischen Raten $\underline{\tilde{q}}(t)$ auswirken. Das Gütekriterium SSR_q kann folglich auch nur unter großer Unsicherheit bestimmt werden. Neben der Information über die Güte der Abbildung wird mit dieser Methode auch ein geschätzter Verlauf der Reaktionsraten $\underline{r}(t)$ des Subsets L als Ergebnis der Minimierung nach Formel 13 und 14 erhalten.

Leighty, R. et al. beschreibt eine weitere Methode in der die Messwerte (Konzentrationsmessungen) direkt durch eine lineare Schätzung von volumetrischen Reaktionsraten über der Zeit approximiert werden. Durch Lösen eines linearen Optimierungsproblems mit einem Linear Least-Squares Solver kann der Verlauf der Reaktionen schnell geschätzt werden, unter der Annahme, dass dieser linear zwischen Stützstellen verläuft [Leighty, R. W., & Antoniewicz, M. R. (2011), Dynamic metabolic flux analysis (DMFA): a framework for determining fluxes at metabolic non-steady state. *Metabolic engineering*, 13(6), 745-755]. Diese Methode bezieht sich lediglich auf reversible Makroreaktionen (wie den in der Quelle genannten „Free Fluxes“), zudem können Verdünnungseffekte (also Konzentrationsänderungen, die nicht durch die Zellen verursacht werden) nicht berücksichtigt werden. Stimmen die Dimensionen der Makroreaktionen und der Messwerte nicht überein können diese Messwerte nicht zur Schätzung von Reaktionsraten verwendet werden. Dies ist zum Beispiel der Fall wenn das Zellwachstum in Form von Bildung externer Biomasse Teil der Makroreaktionen ist und aus den Messwerten nur die Zelltrockenmasse bekannt ist. In dieser Form ist diese Methode daher nicht auf die Anwendung von irreversiblen Makroreaktionen und Fed-Batch Prozessen geeignet.

Verwendet man das Konzept von Leighty et al., das hiermit per Referenz in diese Anmeldung integriert wird, mit den erfindungsgemäß aufbereiteten (geshifteten) Daten, lässt sich diese Methode jetzt auch auf Fed-Batch Prozesse anwenden. Darüber hinaus kann durch das Hinzufügen einer unteren Grenze für die Reaktionsraten der Makroreaktionen als Randbedingung des linearen Optimierungsproblems die Methode auch für irreversible Reaktionen – wie die Elementary Modes - verwendet werden. Stimmen die Dimensionen der Makroreaktionen und der Messwerte nicht überein kann über geeignete Korrelationen die Dimension der Messwerte an die der Makroreaktionen angepasst werden. Diese Kombination der linearen Schätzung nach Leighty et al. mit den Verbesserungen aus diesem Anspruch wird im Folgenden als „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ bezeichnet.

Es lässt sich so überprüfen, ob sich mit den gewählten Makroreaktionen eines Subsets L des originalen EM-Sets K die gemessenen Daten hinreichend darstellen lassen. Die hier errechnete finale Summe der quadrierten Residuen SSR_C nach Formel 15 zwischen den mit der Methode ermittelten geshifteten Konzentrationen $\hat{C}_s(t)$ und den geshifteten Konzentrationen $\underline{C}_s(t)$ gibt an wie gut die Messdaten mit dem Subset abgebildet werden können.

$$SSR_C = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} (C_{j,s}(t_i) - \hat{C}_{j,s}(t_i))^2 \quad (\text{Formel 15})$$

Je kleiner der Wert von SSR_C , desto besser ist das Subset L . Diese Methode wird gegenüber der Methode von Soons et al., insbesondere für die Modellierung von Fed-Batch Prozessen bevorzugt, da eine Überprüfung der Güte eines Subsets schnell auch ohne eine eventuell fehlerbehaftete vorherige Bestimmung der spezifischen Raten möglich ist. Durch die Annahme, dass die geschätzten Reaktionsraten linear zwischen Stützstellen verlaufen, wirken sich Messabweichungen sehr wenig auf die Abschätzung der Reaktionsraten aus. Der Nachteil dieser Methode ist, dass die Größe des zu untersuchenden Subsets L durch die Lösung des linearen Optimierungsproblems begrenzt ist. Die maximale Anzahl der Reaktionen in dem Subset ist gleich der Anzahl an vorhandenen Messungen geteilt durch die Anzahl der Stützstellen.

Neben der Information über die Güte der Abbildung wird mit dieser Methode auch ein geschätzter Verlauf der Reaktionsraten des Subsets $\underline{r}(t)$ erhalten.

In einer bevorzugten Ausführung der erfindungsgemäßen „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ wird zur Lösung des linearen Optimierungsproblems statt eines Linear Least-Squares Solvers der Non-negative Least-Squares Solver (NNLS) von Lawson et al verwendet

[Lawson, C.L. and R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, 1974, Chapter 23, p. 161.]. Dadurch ist es möglich, auch die Güte von größeren Subsets mit der Methode zu überprüfen. Die maximale Anzahl der Makroreaktionen kann hierbei auch signifikant größer als die Anzahl an vorhandenen Messungen geteilt durch die Anzahl der Stützstellen sein. Diese Kombination der „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ mit der Verwendung des Non-negative Least-Squares Solvers wird im Folgenden als „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen mit NNLS“ bezeichnet.

Die erfindungsgemäße Methode der „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen mit NNLS“ kann zusätzlich als weitere datenabhängige Methode zu Vorreduktion der EMs in Schritt e) i) verwendet werden. Hierfür kann hier ein sehr großes Set K von Makroreaktionen verwendet werden. Ergebnis der Methode ist zum einen der Wert für SSR_C und zum anderen der Verlauf der Reaktionsraten $\underline{r}(t)$. EMs mit kleinen Werten der dazugehörigen Rate $\underline{r}(t)$ werden aus der Matrix K entfernt. Diese Vorgehensweise wird wiederholt, bis eine vordefinierte Anzahl an EMs erreicht ist oder der Wert von SSR_C einen festgelegten Grenzwert überschreitet.

Algorithmen zur Auswahl des Subsets sind z. B. aus Provost et al. und Soons et al. bekannt [Provost A. 2006. Metabolic design of dynamic bioreaction models. Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve, Louvain-la-Neuve ; Soons, Z. I. T. A., Ferreira, E. C., Rocha, I. (2010). Selection of Elementary Modes for Bioprocess Control. 11th International Symposium on Computer Applications in Biotechnology, Leuven, Belgium, July 7-9, 2010, 156-161].

Soons et al. beschreiben die Bildung eines EM-Subsets in einem zweistufigem Optimierungsverfahren. Für verschiedene, zufällig ausgewählte EMs werden jeweils die Werte für SSR_q wie oben beschrieben minimiert. Das Set mit dem kleinsten minimierten SSR_q Wert wird ausgewählt. Bei einer großen Zahl von EMs ist jedoch die zufällige Auswahl ineffektiv, da die Anzahl der möglichen Kombinationen sehr stark wächst. Beispielsweise gibt es bei der Auswahl von 10 Reaktionen aus einem Set von 20.000 EMs mehr als $2,8 \cdot 10^{36}$ Kombinationen. Die Wahrscheinlichkeit, dass die optimale Kombination gefunden wird ist hier sehr gering. Durch die Verwendung des Gütekriteriums SSR_q ist diese Methode anfällig gegenüber Messunsicherheiten und Messabweichungen.

Provost beschreibt einen alternativen Algorithmus, bei dem für verschiedene spezifische Werte von $\underline{\tilde{q}}(t_i) \ i = 1, \dots, n$ alle möglichen positiven Linearkombinationen von Elementary Modes ermittelt

werden bei denen gilt: $SSR_q(t = t_i) = 0$. Von diesen zahlreichen möglichen Kombinationen wird anschließend eine Kombination zufällig ausgewählt. Dieses Verfahren verwendet jeweils nur einen Vektor $\underline{\tilde{q}}(t_i)$ und nicht den gesamten Verlauf über der Zeit. Eine Auswahl von EMs für den gesamten Prozess ist deshalb nicht möglich. Mit der zufälligen Auswahl lässt sich zwar der Vektor $\underline{\tilde{q}}(t_i)$ darstellen, inwiefern der Rest des Prozesses hiermit dargestellt werden kann, wird jedoch nicht ermittelt. Ein weiterer Nachteil dieses Verfahrens ist, dass ein Vektor $\underline{\tilde{q}}(t_i)$, der nicht im Lösungsraum aller EMs liegt, nicht verwendet werden kann. Eine Näherungslösung lässt sich nicht ermitteln. Dies ist vor allem bei mit Unsicherheit behafteten Messungen und spezifischen Raten ein großer Nachteil. Durch die Verwendung des Gütekriteriums SSR_q ist diese Methode ebenfalls anfällig gegenüber Messunsicherheiten und Messabweichungen.

In einer weiteren und bevorzugten Ausführungsform des Verfahrens wird daher zur Auswahl der relevanten Makroreaktionen, d. h. zur Auswahl des EM-Subsets L ein evolutionärer, insbesondere ein genetischer, Algorithmus verwendet. Ein solcher Algorithmus ist z. B. aus Baker et al. bekannt [Syed Murtuza Baker, Kai Schallau, Björn H. Junker. 2010. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J. Integrative Bioinformatics*:1-1.]. Besonders bevorzugt kann ein genetischer Algorithmus verwendet werden, in dessen Zielfunktion für verschiedene Kombinationen von EMs der jeweilige Wert SSR_C mit der Methode „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ berechnet wird. Alternativ kann auch eine zufällige Auswahl verwendet werden. Nach Abschluss des Schritt ii) beinhaltet die Matrix L die notwendigen Makroreaktionen (Schritt iii).

In einem optionalen Schritt iii) wird die Validität des EM-Subsets L grafisch überprüft. Es kann hier die Flux Map aus Schritt d) als Projektion des EM-Subsets L herangezogen werden. Abbildung 4 zeigt die Flux Map mit der Projektion eines Subsets von sechs EMs. Ist der EM-Subset L valide, befinden sich die Messdaten weiterhin innerhalb des EM-Subsets L . Diese Darstellung ermöglicht eine schnelle grafische Kontrolle der Validität der Auswahl.

In einem weiteren Schritt f) werden mit den spezifischen Raten $\underline{\tilde{q}}(t)$ aus d) und / oder den Messdaten aus c) die Reaktionsraten der Makroreaktionen des Subsets L berechnet. Die Berechnung von $\underline{r}(t)$ kann anhand der spezifischen Raten $\underline{\tilde{q}}(t)$ wie in e) beschrieben nach Soons et al. erfolgen [Soons, Z. I. T. A., Ferreira, E. C., Rocha, I. (2010). Selection of Elementary Modes for Bioprocess Control. 11th International Symposium on Computer Applications in Biotechnology, Leuven, Belgium, July 7-9,

2010, 156-161], bevorzugt erfolgt die Berechnung von $\underline{r}(t)$ anhand der Messdaten aus c) mit der erfindungsgemäßen „linearen Schätzung von Reaktionsraten ausgewählter Makroreaktionen“.

Im Schritt g) des Verfahrens werden die Kinetiken der Makroreaktionen entworfen. Die ermittelten Kinetiken sollen die dynamischen Einflüsse des Prozesszustands auf die jeweiligen Reaktionsraten \hat{r}_k quantifizieren:

$$\hat{r}_k = f(\underline{C}, pH, T, \dots) \quad (\text{Formel 16})$$

Aus den Kinetiken ergeben sich die zu bestimmenden Modellparameter.

Die generischen Kinetiken werden in Schritt g) i. aus der Stöchiometrie der Makroreaktionen entworfen. Für Substrate der Makroreaktion i wird eine Limitierung vom Monod Typ angenommen. Die maximale Reaktionsrate wird mit den verschiedenen Limitierungen \tilde{r}_i multipliziert:

$$\hat{r}_k(t) = r_{k,max} \cdot \prod_{i=1}^{N_l} \tilde{r}_i(t) = r_{k,max} \cdot \prod_{i=1}^{N_l} \left(\frac{C_i(t)}{K_{m,k,i} + C_i(t)} \right)^{n_i} \quad (\text{Formel 17})$$

Wobei die Monod-Konstanten $K_{m,k,i}$ und die Hill-Koeffizienten n_i die Parameter der Gleichung darstellen, deren erste Werte manuell eingegeben werden. Üblicherweise werden die Monod-Konstanten $K_{m,k,i}$ auf ein Zehntel der jeweiligen maximalen gemessenen Konzentrationen und die Hill-Koeffizienten n_i auf den Wert 1 gesetzt. Die Ermittlung von generischen Kinetiken aus den Reaktionsstöchiometrien wird von Provost oder von Gao et al. beschrieben [Provost A. 2006. Metabolic design of dynamic bioreaction models. *Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve*, S. 126 ; [Gao, J., Gorenflo, V. M., Scharer, J. M., & Budman, H. M. (2007). Dynamic metabolic modeling for a MAB bioprocess. *Biotechnology progress*, 23(1), 168-181]. Diese Methoden werden hiermit in die Anmeldung per Referenz integriert. In diesen Methoden werden der Substratlimitierungen des Monod-Typs für die jeweiligen Substrate einer Reaktion verwendet. Obwohl Provost oder Gao dies nicht beschreiben, sind Inhibierungen durch toxische Produkte ebenfalls mit dieser Methode aus der Reaktionsstöchiometrie ableitbar.

Im Schritt g) ii. werden die Einflussgrößen auf die in f) ermittelten Reaktionsraten $\underline{r}(t)$ ermittelt. Dabei werden alle Größen, die den Prozesszustand beschreiben (also auch Bioreaktionsbedingungen wie z. B. der pH-Wert, die Reaktortemperatur, Partialdrücke, die nicht aus der Stöchiometrie der Makroreaktion ableitbar sind) betrachtet. Die Einflussgrößen können manuell beispielsweise mit Hilfe einer statistischen Methode wie Partial-Least-Squares ermittelt werden. Hierzu wird die Korrelation zwischen dem Prozesszustand (der in einer Matrix zusammengefasst wird) und den Reaktionsraten $\underline{r}(t)$ aus f) ermittelt.

In einem Schritt g) iii. werden dann die in g) ii. ermittelten Einflüsse quantifiziert und die Kinetik aus i. um entsprechende Terme erweitert. Ein in g) ii. gefundener Einfluss von einer Größe des Prozesszustandes auf eine Reaktionsrate kann dann mit einem Term \tilde{r}_i belegt werden. Der Term \tilde{r}_i ist eine beliebige Funktion, die in Abhängigkeit des Prozesszustandes einen Wert zwischen 0 und 1 ergibt. Die in g) i. aufgestellte generische Kinetik der Reaktion wird dann mit diesem Term multipliziert.

Beispielsweise deutet eine gefundene negative Korrelation zwischen der Konzentration einer Komponente i und der Reaktion k auf eine Beeinflussung der Reaktionsrate k durch die Konzentration von i (C_i) hin. Dieses kann beispielsweise mit einer Inhibierungskinetik nach Haldane belegt werden:

$$\tilde{r}_i(t) = \frac{K_{I,k,i}}{K_{I,k,i} + C_i(t)} \quad (\text{Formel 18})$$

Wobei $K_{I,k,i}$ die Inhibierungskonstante bezeichnet und einen weiteren Modellparameter darstellt, dessen erster Wert manuell eingegeben wird und üblicherweise auf ein Zehntel der jeweiligen maximalen gemessenen Konzentrationen gesetzt wird.

In einem optionalen Schritt h) werden die Modellparameterwerte \underline{p} der Kinetiken an die in f) ermittelten Reaktionsraten der Makroreaktionen $\underline{r}(t)$ angepasst:

$$\min_{\underline{p}} \sum_{k=1}^{N_t} (\hat{r}_k(\underline{p}) - r_k)^2 \quad (\text{Formel 19})$$

Dieses wird im Folgenden als Modellparameterwertschätzung bezeichnet. Auf ein numerisches Lösen einer oder mehrerer Differentialgleichungen nach den Formeln 2 bis 4 in diesem Schritt kann verzichtet werden; die Modellparameterwerte können in unabhängigen Gruppen mit üblicherweise 3 bis 10 Parametern separat für jede Makroreaktion k angepasst werden. Die Anpassung erfolgt durch eine übliche Methode wie der Gauß-Newton Methode [Bates DM, Watts DG. 1988. Nonlinear regression analysis and its applications. New York: Wiley. xiv, 365.].

Diese für jede Makroreaktion separate Modellparameterwertschätzung ist für die Schritte i) und j) besonders vorteilhaft, da sie einerseits schnell durchführbar ist und andererseits verbesserte Startwerte für die Anpassung der Modellparameterwerte an Messdaten aus c) in Schritt j) liefert.

Die Anpassungsgüte wird beispielsweise mit der Summe der quadrierten Residuen SSR_r nach Formel 20 berechnet:

$$SSR_r = \sum_{k=1}^{N_t} \left(\hat{r}_k(\underline{p}) - r_k \right)^2 \quad (\text{Formel 20})$$

Je kleiner der Wert für SSR_r , desto besser ist die Anpassung. Alternativ erfolgt die Überprüfung der Anpassungsgüte durch einen graphischen Vergleich von \hat{r}_k und r_k .

In einem optionalen Schritt i) werden die in g) gewählten Kinetiken der Makroreaktionen auf ihre Anpassungsgüte hin überprüft. Grundlage ist der in Schritt h) berechnete Wert SSR_r , der die Anpassungsgüte der Modellparameterwertschätzung quantifiziert. Bei einer nicht zufriedenstellenden Anpassungsgüte können die Schritte g) und h) wiederholt werden, bis eine vordefinierte Anpassungsgüte erreicht wird.

In einem weiteren Schritt j) kann die Anpassung der Parameterwerte der Kinetiken aus g) an die Messdaten aus c) nach einer für Anpassungen üblichen Methode erfolgen. Bevorzugt werden für diese Anpassung die Startwerte aus Schritt h) verwendet. Die Modellparameterwertanpassung erfolgt unter Einbeziehung der o. g. Differentialgleichungen (Formeln 2 bis 4), z. B. mit Hilfe der Gauß-Newton Methode [Bates DM, Watts DG. 1988. Nonlinear regression analysis and its applications. New York: Wiley. xiv, 365.] oder mit einem multiple Shooting Algorithmus [Peifer M, Timmer J. 2007. Parameter estimation in ordinary differential equations for biochemical processes using the method of multiple shooting. Systems Biology, IET 1(2):78–88.].

Bevorzugt können Produktmerkmale in das Modell integriert werden. Besonders bevorzugt kann dies für Produktmerkmale, die von der Konzentration von Neben- oder Zwischenprodukten abhängen, eingeführt werden. Konzentrationen von Nebenprodukten, die externe Komponenten des in a) eingegebenen metabolischen Netzwerkes sind, sind bereits in das Modell integriert und können berechnet werden. Bei Bedarf lassen sich allerdings auch andere Neben- oder Zwischenprodukte in einem oder mehreren separaten metabolischen Netze zusammenfassen. Dies ist vorteilhaft, wenn die erwarteten Ausscheide- oder Aufnahmearten sich in unterschiedlichen Größenordnungen bewegen oder bestimmte Stoffwechselforgänge in unterschiedlichen Detailgraden betrachtet werden sollen. Alternativ zur einem integrierten Modell kann mit den Schritten a) bis j) ein separates Modell zur Berechnung der Produktmerkmale erzeugt werden, das den Prozessverlauf der externen Komponenten des separaten metabolischen Netzwerkes ebenfalls mit einem Satz von Makroreaktionen mit eigenen Kinetiken beschreibt. Neben- oder Zwischenprodukte, die nicht außerhalb des Organismus vorliegen, aber deren zellinterne Akkumulation sich auf ein- oder mehrere Produktmerkmale auswirkt können in Schritt a) und b) bei der Berechnung der EMs und der Formulierung der Makroreaktionen externalisiert, also als externe Komponenten klassifiziert, werden. Die Einbindung von

Produktmerkmalen, die von zellinternen oder zellexternen Konzentrationen abhängig sind, kann dann über die zusätzliche Integration von quantitativen oder qualitativen Zusammenhängen zwischen Konzentrationen und Produktmerkmalen erfolgen.

Weitere Gegenstände sind ein Computerprogramm oder eine Software zur Durchführung des erfindungsgemäßen Verfahrens.

Das mit dem erfindungsgemäßen Verfahren bereitgestellte Modell kann zur Prozessführung oder Planung der Prozessführung sowie Untersuchung des Verfahrens im Reaktor verwendet werden.

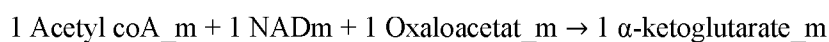
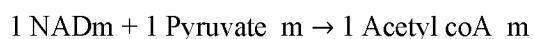
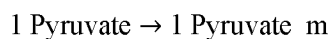
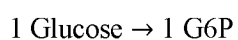
Eine besondere Ausführungsform des erfindungsgemäßen Verfahrens wird exemplarisch beschrieben, ohne sich darauf zu beschränken. Mit diesem Verfahren wurde ebenfalls exemplarisch ein Modell einer Fermentation mit Hybridoma-Zellen erstellt und dessen Validität wie beschrieben geprüft.

Beispiel: Modellierung einer Hybridoma-Zellkultur

1 Schritt a)

Das Hintergrundwissen in Form eines metabolischen Netzwerkes wurde aus der Veröffentlichung von Niu et al. (Metabolic pathway analysis and reduction for mammalian cell cultures—Towards macroscopic modeling. Chemical Engineering Science (2013) 102, S. 461–473. DOI: 10.1016/j.ces.2013.07.034.) entnommen. Das hier beschriebene metabolische Netzwerk einer tierischen Zelle enthält 35 Reaktionen, die 37 interne und externe Metaboliten verknüpfen (siehe Abbildung 5, siehe Tabelle 1).

Tabelle 1: Reaktionen des Metabolischen Netzwerkes nach Niu et al. (Metabolic pathway analysis and reduction for mammalian cell cultures—Towards macroscopic modeling. Chemical Engineering Science (2013) 102, S. 461–473. DOI: 10.1016/j.ces.2013.07.034.)



1 -ketoglutarate_m + 1 NADm → 1 Succinyl CoA_m
 1 FADm + 1 Succinyl CoA_m → 1 Fumarate
 1 Fumarate → 1 Malat_m
 1 Malat_m + 1 NADm → 1 Oxaloacetat_m
 1 Glutamine → 1 Glutamate + 1 NH3
 1 Glutamate + 1 NADm → 1 α-ketoglutarate_m + 1 NH3
 1 Malat_m → 1 Malate
 1 Malate + 1 NAD → 1 Pyruvate
 1 Glutamate + 1 Pyruvate → 1 α-ketoglutarate_m + 1 Alanine
 1 Glutamate + 1 Oxaloacetat_m → 1 α-ketoglutarate_m + 1 Aspartate
 1 Arginine + 2 NADm → 1 Glutamate + 3 NH3
 1 Asparagine → 1 Aspartate + 1 NH3
 2 Glycine + 1 NADm → 1 NH3
 1 Histidine + 1 NADm → 1 Glutamate + 2 NH3
 1 Isoleucine + 2 NADm → 1 Acetyl coA_m + 1 NH3 + 1 Succinyl CoA_m
 1 Leucine + 3 NADm → 3 Acetyl coA_m
 1 Lysine + 6 NADm → 2 Acetyl coA_m
 1 Methionine + 4 NADm → 1 NH3 + 1 Succinyl CoA_m
 1 NADm + 1 Phenalanine → 1 Tyrosine
 1 Serine → 1 NH3 + 1 Pyruvate
 1 NADm + 1 Threonine → 1 NH3 + 1 Succinyl CoA_m
 19 NADm + 1 TRP → 3 Acetyl coA_m
 5 NADm + 1 Tyrosine → 2 Acetyl coA_m + 1 Fumarate
 5 NADm + 1 Valine → 1 NH3 + 1 Succinyl CoA_m
 1 NADm → 1 NAD
 0.5 Oxygen(O2) → 1 NADm
 1 NADm → 1 FADm
 0.0156 Alanine + 0.0082 Arginine + 0.0287 Aspartate + 0.0167 G6P + 0.0245 Glutamine + 0.0039
 Glutamate + 0.0196 Glycine + 0.0038 Histidine + 0.0099 Isoleucine + 0.0156 Leucine + 0.0119
 Lysine + 0.0039 Methionine + 0.0065 Phenalanine + 0.016 Serine + 0.0094 Threonine + 0.0047
 Tyrosine + 0.0113 Valine → 1 X (**Biomasse**) + 0.0981 NAD
 0.01101 Alanine + 0.005033 Arginine + 0.007235 Asparagine + 0.0081787 Aspartate + 0.010381
 Glutamine + 0.010695 Glutamate + 0.01447 Glycine + 0.0034602 Histidine + 0.005033 Isoleucine +

0.014155 Leucine + 0.01447 Lysine + 0.0028311 Methionine + 0.007235 Phenalanine + 0.026738 Serine + 0.016043 Threonine + 0.0084932 Tyrosine + 0.018874 Valine → 1 IgG (**Antibody**)

In der Veröffentlichung ist die Reversibilität der Reaktionen nicht explizit angegeben. Stattdessen wurden die Daten zur „Metabolischen Flussanalyse“ aus derselben Veröffentlichung ausgewertet und zur Identifizierung der irreversiblen Reaktionen verwendet.

Mit der stöchiometrischen Matrix N , die die Stöchiometrie, d. h. die stöchiometrischen Koeffizienten, der internen Metabolite enthält, und den Informationen über die Reversibilität der Reaktionen wurden mit Hilfe von METATOOL 5.1 (Pfeiffer et al. METATOOL: for studying metabolic networks, Bioinformatics 1999 15 (3), S. 251–257.) alle Elementary Modes (EMs) des Netzwerkes berechnet. Die Zahl der EMs liegt hier bei über 300.000.

2 Schritt b)

Die Matrix mit den berechneten EMs E wurde in Schritt a) gewonnen. Analog zur Matrix N , enthält die Matrix N_p die Stöchiometrie, d. h. die stöchiometrischen Koeffizienten, der externen Metabolite. Mögliche Makroreaktionen des stöchiometrischen Netzwerkes wurden in der Matrix K mit Formel 21 zusammengefasst:

$$K = N_p \cdot E \quad (\text{Formel 21})$$

3 Schritt c)

Die Messdaten des Prozesses wurden aus Baughman et al. entnommen, der verschiedene Messgrößen einer Fermentation von Hybridoma-Zellen über den Verlauf eines Batch-Prozess (vgl. Abbildung 6) angibt [On the dynamic modeling of mammalian cell metabolism and mAb production. In: Computers & Chemical Engineering (2010) 34 (2), S. 210–222.]. Die Messdaten wurden in das Verfahren eingegeben.

4 Schritt d)

Mit Hilfe von Spline-interpolierten Messwerten aus c) (C^{int}) wurden die Wachstums- und Absterberaten sowie die spezifischen Aufnahme- bzw. Ausscheideraten berechnet (vgl. Abbildung 7). Die Lysis wurde mit einem vordefinierten und in das Verfahren eingegebenen, über den Prozesszeitraum konstanten Lysis-Faktor $K_l = 0.1$ einbezogen. Ein Shiften der Messdaten war nicht

notwendig, da es sich hier um Daten eines Batch-Prozesses ohne weitere Zugaben handelt. Die Daten zeigen dementsprechend einen steigenden Verlauf da alle Konzentrationsänderungen von den Zellen- und nicht von Zugaben verursacht sind.

Zur Berechnung der Raten \tilde{q} werden zusätzliche Informationen herangezogen. So konnte mit Hilfe der ebenfalls im Datensatz aus Baughman et al. angegebenen gesamten Biomasse (BM in $\left[\frac{C-mol}{l}\right]$) und der gesamten Zellzahl ein durchschnittlicher C-mol Gehalt von $f_{C-mol, X_v} = 18.41 \left[\frac{C-mol}{10^9 cells}\right]$ errechnet werden. Die C-mol bezogene Wachstumsrate konnte nun mit Formel 22:

$$\tilde{\mu} \left[\frac{C-mol}{h \cdot 10^9 cells} \right] = \mu \left[\frac{1}{h} \right] \cdot f_{C-mol, X_v} \left[\frac{C-mol}{10^9 cells} \right] \quad (\text{Formel 22})$$

berechnet werden. Analog kann die C-mol bezogene Bildungsrate des Antikörpers abgeschätzt werden. Hierzu wurde die molare Zusammensetzung des Antikörpers zu $CH_{1.58}O_{0.31}N_{0.27}S_{0.004}$ mit einer formellen molaren Masse von $M_{mAb, C-mol} = 22.45 \frac{g}{mol}$ geschätzt. Hier wird angenommen, dass die molare Zusammensetzung einer durchschnittlichen molaren Zusammensetzung von Proteinen wie von Villadsen et al. [Bioreaction engineering principles (2011), Kapitel 3, Elemental and Redox Balances, S. 73, Springer Verlag, ISBN: 978-1-4419-9687-9] angegeben entspricht. Die molare Masse des gesamten Antikörpers wurde zu $M_{mAb} = 150.000 \frac{g}{mol}$ geschätzt. Die Bildungsrate des Antikörpers ergab sich dann aus der Formel:

$$\tilde{q}_{mAb} \left[\frac{C-mol}{h \cdot 10^9 cells} \right] = q_{mAb} \left[\frac{10^{-4} mol}{h \cdot 10^9 cells} \right] \cdot \frac{M_{mAb, C-mol}}{M_{mAb}} \cdot 10^4 \quad (\text{Formel 23})$$

Der zeitliche Verlauf der Raten $\tilde{q}(t)$ konnte dann zur Auswahl der Makroreaktionen herangezogen werden.

5 Schritt e)

Im Schritt e) wurde ein EM-Subset von Makroreaktionen erzeugt, mit dem der Datensatz bestmöglich wiedergegeben wurde. Hierzu wurde die Matrix K aus Schritt b) benötigt. Da die Anzahl von über 300.000 Makroreaktionen zu einer zu großen Anzahl von möglichen Kombinationen geführt hätte, wurde zuerst eine datenabhängige Vorreduktion durchgeführt:

Hierzu wurden die in Schritt d) ermittelten Raten $\tilde{q}(t)$ zur Berechnung der Ausbeutekoeffizienten Y^m für alle Kombinationen zweier externer Metabolite verwendet. Die untere Grenze eines Ausbeutekoeffizienten $Y_{i,j}$ wurde so gewählt, dass 99% der ermittelten Ausbeutekoeffizienten $Y_{i,j}^m(t)$ über diesem Wert liegen. Die obere Grenze wurde so gewählt, dass 99% der ermittelten

Ausbeutekoeffizienten $Y_{i,j}^m(t)$ unter diesem Wert liegen. Exemplarisch sind einige ermittelten Grenzen sowie der Anteil der EMs, dessen Ausbeutekoeffizienten $Y_{i,j}^{EM}$ innerhalb dieser Grenzen liegen in Tabelle 2 angegeben. Insgesamt konnte die Anzahl der EMs so auf ca. 3000 reduziert werden.

Tabelle 2: Externe Metabolite, deren maximale und minimale Ausbeutekoeffizienten $Y_{i,j}$ sowie der Anteil der EMs, dessen Ausbeutekoeffizienten innerhalb der angegebenen Grenzen liegen

Externe Komponenten	Untere Grenze	Obere Grenze	Anteil der EMs innerhalb der Grenzen
Ala : Asn	-14.6811	-0.1752	99.9134 %
Asn : Glc	0.0293	16.1603	64.0488 %
Asp : Ala	-1.1130	-2.1011	74.0839 %

Im Anschluss an die datenabhängige Reduktion wurde anschließend noch eine datenunabhängige Reduktion durchgeführt. Hierbei wurde ein Maximalwert für die Kosinus-Ähnlichkeit zweier EMs von 0,995 definiert. Beginnend mit der ersten Reaktion wurden so alle Makroreaktionen aus der Matrix K entfernt, die diesen Wert überschritten. Es verblieben ca. 500 Makroreaktionen aus der Matrix K (auch reduzierte Matrix K genannt), die weiterhin mehr als 95% des Volumens des durch die ca. 3000 EMs aufgespannten Lösungsraum abdecken.

Vor dem Auswahlprozess erfolgte zudem ein Abgleich der im metabolischen Netzwerk nach Niu et al. angegebenen Komponenten (die den externen Metaboliten des metabolischen Netzwerkes aus a) entsprechen) und den gemessenen Komponentenkonzentrationen aus c). Bis auf Prolin werden alle von Baughman et al. gemessenen Konzentrationen auch im metabolischen Netzwerk nach Niu et al. berücksichtigt. Um die Messung der Prolinkonzentration nutzen zu können wäre entweder ein anderes vereinfachtes Netzwerk, das Prolin als externen Metaboliten enthält, verwendbar oder eine Erweiterung des bestehenden metabolischen Netzwerkes möglich.

Komponenten, die zwar in den berechneten Makroreaktionen vorkamen, aber von denen keine Daten verfügbar waren, wurden ebenfalls im Folgenden ignoriert. Die entsprechenden Zeilen der Matrix K wurden dementsprechend aus der Matrix gestrichen. Das Streichen der entsprechenden Zeilen bedeutet nicht, dass diese Ein- bzw. Ausgänge von der Zelle nicht genutzt werden. Sie bestehen weiterhin im metabolischen Netzwerk, nur fehlen hierzu Messungen, die diese bilanzierbar machen können. In diesem Beispiel wurden die Ein- bzw. Ausgänge von Arginine, Glutamate, Glycin, Histidin, Leucin, Lysine, Methionin, Ammonium, Sauerstoff, Phenylalanin, Serin, Threonin, Tryptophan, Tyrosin und Valin vernachlässigt.

In der folgenden Schritten des Verfahrens wird nun die reduzierte Matrix K - die die das Hintergrundwissen abbildet - und die Raten $\tilde{q}(t)$ aus d) sowie die Messdaten aus c) - die das Prozesswissen bilden - dazu verwendet, ein möglichst kleines Subset L der Makroreaktionen aus K zu gewinnen.

Es wurde die erfindungsgemäße „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ als Gütekriterium verwendet.

Analog zu den Raten $\tilde{q}(t)$ wurden die Messwerte der Zellzahl und des Antikörpers hier auf C-mol normiert. Dies ist erforderlich damit die Dimension der Makroreaktionen mit denen der Messwerte übereinstimmt.

Die Auswahl des Subsets erfolgte mit einem genetischen Algorithmus. In der Berechnung der Zielfunktion dieses genetischen Algorithmus wurde hierbei das bei der „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ angesprochene lineare Optimierungsproblem gelöst. Die hier errechnete finale Summe der kleinsten Fehlerquadrate des linearen Optimierungsproblems war zugleich der Wert der Zielfunktion für die jeweilige Auswahl der Makroreaktionen.

Zur Auswahl der Größe des Subsets L aus K wurde die Optimierung mit wiederholt mit einer unterschiedlichen Anzahl an Makroreaktionen in L durchgeführt. Die Anzahl stellt einen Kompromiss zwischen der Modellkomplexität und der Wiedergabegenauigkeit dar. Um zu ermitteln, wie viele Reaktionen ausreichend für die Wiedergabe sind, kann entweder die Auswahl des Subsets L für eine variierende Anzahl von Makroreaktionen wiederholt werden oder ein Strafterm für die Reaktionsanzahl wird direkt der Zielfunktion des genetischen Algorithmus hinzugefügt werden. In diesem Fall wurden mehrere Optimierungen mit einer vordefinierten Anzahl von Makroreaktionen (10, 7, 5, 4 und 3) durchgeführt. Die mit dem genetischen Algorithmus gefundene kleinste Summe der Fehlerquadrate ist in Abbildung 9 gegenüber der Zahl der Makroreaktionen aufgetragen. Es zeigte sich, dass in diesem Fall weniger als sieben Makroreaktionen zu wenig sind um den Prozessverlauf hinreichend gut darzustellen. Die gewählten Makroreaktionen sind in Tabelle 3 angegeben.

Tabelle 3: Gewähltes Subset der Makroreaktionen (L). Nicht unterstrichene Komponenten werden in dem Modell nicht berücksichtigt, da hierzu keine Messungen vorliegen.

0.474 Alanine + 0.474 Methionine
 → 0.158 Asparagine + 0.316 Aspartate + 0.632 Glycine
 + 0.158 Tryptophan

0.015 Alanine + 0.00789 *Arginine* + 0.0304 Asparagine + 0.0161 Glucose
+ 0.0236 Glutamine + 0.00375 *Glutamate* + 0.00366 *Histidine*
+ 0.00953 *Isoleucine* + 0.015 *Leucine* + 0.112 *Methionine*
+ 0.00626 *Phenalanine* + 0.0154 *Serine* + 0.0109 *Valine*
→ 0.963 X (Biomass) + 0.00276 Aspartate + 0.24 *Glycine*
+ 0.0208 *Tryptophan*

0.295 Asparagine + 0.147 *Glutamate*
→ 0.295 Aspartate + 0.885 *Glycine* + 0.147 Lactate

0.00753 *Arginine* + 0.113 Asparagine + 0.0603 Glucose + 0.0225 Glutamine
+ 0.0824 *Histidine* + 0.00909 *Isoleucine* + 0.00597 *Phenalanine*
+ 0.0216 *Tryptophan* + 0.00431 *Tyrosine* + 0.0104 *Valine*
→ 0.918 X (Biomass) + 0.061 Alanine + 0.0865 Aspartate
+ 0.343 *Glycine* + 0.0631 *Methionine*

0.0654 *Arginine* + 0.412 Aspartate + 0.00991 Glucose + 0.0145 Glutamine
+ 0.554 *Glycine* + 0.00226 *Histidine* + 0.00588 *Isoleucine*
+ 0.00926 *Leucine* + 0.00706 *Lysine* + 0.0649 *Phenalanine*
+ 0.0095 *Serine* + 0.00671 *Valine*
→ 0.594 X (Biomass) + 0.049 Alanine + 0.395 Asparagine
+ 0.0503 *Threonine* + 0.0388 *Tryptophan*

$$\begin{aligned}
&0.0077 \textit{ Arginine} + \underline{0.179 \textit{ Aspartate}} + \underline{0.0157 \textit{ Glucose}} + \underline{0.104 \textit{ Glutamine}} \\
&\quad + 0.216 \textit{ Glycine} + 0.00357 \textit{ Histidine} + 0.00929 \textit{ Isoleucine} \\
&\quad + 0.0146 \textit{ Leucine} + 0.0112 \textit{ Lysine} + 0.038 \textit{ Tyrosine} + 0.0106 \textit{ Valine} \\
&\rightarrow \underline{0.939 X (\textit{Biomass})} + \underline{0.0624 \textit{ Alanine}} + \underline{0.152 \textit{ Asparagine}} \\
&\quad + 0.0183 \textit{ Tryptophan}
\end{aligned}$$

$$\begin{aligned}
&0.0342 \textit{ Arginine} + \underline{0.211 \textit{ Aspartate}} + \underline{0.00762 \textit{ Glucose}} + \underline{0.0195 \textit{ Glutamine}} \\
&\quad + 0.244 \textit{ Glycine} + 0.00452 \textit{ Histidine} + 0.0546 \textit{ Isoleucine} \\
&\quad + 0.0185 \textit{ Leucine} + 0.0171 \textit{ Lysine} + 0.00406 \textit{ Methionine} \\
&\quad + 0.0178 \textit{ Tyrosine} + 0.0203 \textit{ Valine} \\
&\rightarrow \underline{0.457 X (\textit{Biomass})} + \underline{0.804 \textit{ IgG (\textit{Antibody})}} + \underline{0.185 \textit{ Asparagine}} \\
&\quad + 0.0153 \textit{ Tryptophan}
\end{aligned}$$

In den gezeigten Makroreaktionen sind alle externen Metaboliten des metabolischen Netzwerkes aus a) angegeben. Teil des Modells sind jedoch nur die unterstrichenen externen Metabolite, da nur für diese Messdaten aus c) vorliegen.

6 Schritt f)

Für das gewählte Set der Makroreaktionen wurden die Reaktionsraten über der Zeit ermittelt. In diesem Beispiel wurden mit der erfindungsgemäßen Methode „lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ die in Abbildung 10 gezeigten Messwerte durch eine Schätzung der Reaktionsraten $\underline{r}(t)$ approximiert. Das Ergebnis der Methode ist ein stückweise linearer Verlauf der einzelnen (volumetrischen) Reaktionsraten. Durch Division mit dem interpoliertem Verlauf der vitalen Zellzahl $X_v(t)$ wurden die zellspezifischen Reaktionsraten $\underline{r}(t)$ der in Tabelle 3 gezeigten Makroreaktionen gewonnen. Die so erhaltenen Reaktionsraten $\underline{r}(t)$ sind in Abbildung 10 abgebildet.

7 Schritt g)

Für alle in Tabelle 3 gezeigten Makroreaktionen wurden generische Kinetiken gemäß Formel 24 angenommen:

$$\hat{r}_k(t) = r_{k,max} \cdot \prod_{i=1}^{N_i} \tilde{r}_i(\underline{C}(t), \underline{p}, \dots) \quad (\text{Formel 24})$$

In diesem Fall wurden sie durch eine Monod-Kinetik realisiert D. h. das bei jeder Reaktion k für jedes Substrat i eine Limitierung gemäß Formel 25:

$$\tilde{r}_i(t) = \left(\frac{C_i(t)}{K_{m,k,i} + C_i(t)} \right)^{n_i} \quad (\text{Formel 25})$$

eingeführt wurde. Hierbei ist $r_{k,max}$ die maximale Reaktionsgeschwindigkeit, N_i die Anzahl der berücksichtigten Limitierungen, C_i die Konzentration der Komponente i , $K_{m,k,i}$ die dazugehörigen Monod-Konstanten und n_i der Hill-Parameter für die Reaktionsordnung darstellen. Deren Werte werden in den Schritten h) und j) angepasst.

Weitere Terme ergeben sich aus der Analyse der Reaktionsraten $\underline{r}(t)$ aus f). In diesem Beispiel wurden neben Substratlimitierungen auch Inhibierungen gemäß Formel 26 berücksichtigt.

$$\tilde{r}_i = \left(\frac{K_{I,k,i}}{K_{I,k,i} + C_i} \right)^{n_i} \quad (\text{Formel 26})$$

Auch für diese Limitierung mussten die Werte der Parameter $K_{I,k,i}$ und n_i angepasst werden. Die verwendeten kinetischen Terme der Reaktionen sind in Tabelle 4 angegeben.

Tabelle 4: Kinetische Terme der gewählten Makroreaktionen aus L

$$\hat{r}_1(t) = r_{1,max} \cdot \left(\frac{[Ala](t)}{K_{m,Ala,1} + [Ala](t)} \right) \cdot \left(\frac{[Glc](t)}{K_{m,Glc,1} + [Glc](t)} \right)^2 \cdot \left(\frac{K_{I,Asn,1}}{K_{I,Asn,1} + [Asn](t)} \right)^2$$

$$\hat{r}_2(t) = r_{2,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,2} + [Glc](t)} \right) \cdot \left(\frac{[Gln](t)}{K_{m,Gln,2} + [Gln](t)} \right) \cdot \left(\frac{[Asn](t)}{K_{m,Asn,2} + [Asn](t)} \right) \cdot \left(\frac{[Ala](t)}{K_{m,Ala,2} + [Ala](t)} \right)^2$$

$$\hat{r}_3(t) = r_{3,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,3} + [Glc](t)} \right) \cdot \left(\frac{[Asn](t)}{K_{m,Gln,3} + [Asn](t)} \right) \cdot \left(\frac{K_{I,Lac,3}}{K_{I,Lac,3} + [Lac](t)} \right)^2$$

$$\hat{r}_4(t) = r_{4,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,4} + [Glc](t)} \right) \cdot \left(\frac{[Gln](t)}{K_{m,Gln,4} + [Gln](t)} \right) \cdot \left(\frac{[Asn](t)}{K_{m,Asn,4} + [Asn](t)} \right) \cdot \left(\frac{[X_t](t)}{K_{m,Xt,4} + [X_t](t)} \right)$$

$$\hat{r}_5(t) = r_{5,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,5} + [Glc](t)} \right) \cdot \left(\frac{[Gln](t)}{K_{m,Gln,5} + [Gln](t)} \right) \cdot \left(\frac{[Asp](t)}{K_{m,Asp,5} + [Asp](t)} \right) \cdot \left(\frac{K_{I,Asp,5}}{K_{I,Asp,5} + [Asp](t)} \right)^3 \cdot \left(\frac{K_{I,Asn,5}}{K_{I,Asn,5} + [Asn](t)} \right)$$

$$\hat{r}_6(t) = r_{6,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,6} + [Glc](t)} \right) \cdot \left(\frac{[Gln](t)}{K_{m,Gln,6} + [Gln](t)} \right) \cdot \left(\frac{[Asp](t)}{K_{m,Asp,6} + [Asp](t)} \right)^2$$

$$\hat{r}_7(t) = r_{7,max} \cdot \left(\frac{[Glc](t)}{K_{m,Glc,7} + [Glc](t)} \right) \cdot \left(\frac{[Gln](t)}{K_{m,Gln,7} + [Gln](t)} \right) \cdot \left(\frac{[Asp](t)}{K_{m,Asp,7} + [Asp](t)} \right)$$

8 Schritt h)

Für jede Reaktionsrate konnte mit der in Tabelle 4 angegebenen Kinetik und den interpolierten Werten der in der Kinetik berücksichtigten Konzentrationen $\underline{C}^{int}(t)$ der Verlauf der Reaktionsrate $\hat{r}_i(\underline{p}, \underline{C}^{int}(t))$ algebraisch berechnet werden.

Die Anpassung der Parameter dieser Kinetiken erfolgte separat für jede Reaktion i an die in Schritt f) ermittelte Reaktionsrate $r_i(t)$. Die Zielfunktion für eine Optimierung der in Reaktion i vorkommenden Parameter lautete in diesem Beispiel:

$$\min_{\underline{p}_k} \left(\sum_{l=0}^T \left(\hat{r}_k(\underline{p}_k, \underline{C}^{int}(t_l)) - r_k(t_l) \right)^2 \right) \quad (\text{Formel 27})$$

Die so angepassten Verläufe aller berechneten $\hat{r}_k(\underline{p}_k, \underline{C}^{int}(t))$ sind zusammen mit den entsprechenden $r_k(t)$ in Abbildung 11 dargestellt. Die Verläufe der ersteren sind gestrichelt, die der letzteren durchgehend dargestellt. Es ist erkennbar, dass der Verlauf qualitativ übereinstimmt. Dies bedeutet, dass mit den gewählten Kinetiken auch die Dynamik des Prozesses zufriedenstellend wiedergegeben werden kann. Diese Information ist in diesem Modellierungsschritt sehr hilfreich, da bei einer nicht zufriedenstellenden Wiedergabe die schnell durchzuführenden Schritte g) (Auswahl anderer Kinetiken) und h) (Parameterwertschätzung) wiederholt werden können, bis der gewünschte Anpassungsgrad erreicht ist. Schritt i) war hier also nicht erforderlich.

9 Schritt j)

Die weiterführende Anpassung der Modellparameterwerte \underline{p} erfolgte mit den Messdaten aus c). Hierfür wurden alle Parameter zugleich optimiert. Zudem wurden die bisher nicht betrachteten Vorgänge Apoptose und Lysis mit einbezogen. Diese werden in den Differentialgleichungen, die die Entwicklung der vitalen- und gesamten Zellzahl beschreiben, benötigt:

$$\frac{dX_v}{dt} = (\mu_x - \mu_d) \cdot X_v \quad (\text{Formel 28})$$

$$\frac{dX_t}{dt} = \mu_x \cdot X_v - K_l \cdot (X_t - X_v) \quad (\text{Formel 29})$$

Die gewählte Kinetik für die Beschreibung der Apoptose lautete:

$$\mu_d(t) = \mu_{d,max} \cdot \frac{([Lac](t) - C_{Lac,cr})}{K_{d,Lac} + ([Lac](t) - C_{Lac,cr})}, [Lac] \geq C_{Lac,cr} \quad (\text{Formel 30})$$

$$\mu_d(t) = 0, [Lac] < C_{Lac,cr} \quad (\text{Formel 31})$$

Die Lysisrate K_l wurde als konstant über dem Prozess angenommen. Neben den Parametern der Reaktionsraten wurden in diesem Schritt die durch die Apoptose und Lysis eingeführten Parameter $C_{Lac,cr}$ (kritische Laktat Konzentration), $\mu_{d,max}$ (maximale Absterberate), $K_{d,Lac}$ (Monod-Parameter zur Beschreibung des Einflusses der Laktat-konzentration auf die Absterberate) und K_l (Lysisrate) bestimmt. In dem Beispiel wurde ausgehend von den Startwerten des Datensatzes durch numerisches Lösen des ODE-Systems der Verlauf der geschätzten Konzentrationen $\hat{C}(t)$ ermittelt. Die Differenz zwischen den gemessenen Konzentrationen $\underline{C}^m(t)$ und den geschätzten Konzentrationen $\hat{C}(t)$ wurde dabei mit üblichen Methoden mit folgender Zielfunktion minimiert:

$$\min_{\underline{p}} \left(\sum_{i=1}^{n_{comp}} \left(\sum_{l=0}^T \left(\hat{C}_i(\underline{p}, t_l) - C_i^m(t_l) \right)^2 \right) \right) \quad (\text{Formel 32})$$

Mit insgesamt 33 Parametern \underline{p} ist diese Optimierung i. d. R. schwierig durchzuführen, da die Zielfunktion viele lokale Optima besitzt. Startet man einen deterministischen Optimierungsalgorithmus, wie z. B. den Levenberg-Marquardt Algorithmus an den aus Schritt h) bekannten Startwerten der Parameter ist die Erfolgsaussicht hingegen stark erhöht. Der angepasste Prozessverlauf ist in Abbildung 12 abgebildet. Die angepassten Parameter sind in Tabelle 5 abgebildet.

Tabelle 5: Parameter der Kinetiken sowie der Apoptose und Lysis

$K_{m,Glc,1}$	14,6	$K_{m,Gln,7}$	0,0187
$K_{m,Ala,1}$	3,41	$K_{m,Asp,7}$	0,872
$K_{m,Glc,2}$	0,0508	$r_{1,max}$	9,47
$K_{m,Gln,2}$	0,00881	$r_{2,max}$	9,91
$K_{m,Asn,2}$	1,38	$r_{3,max}$	57,6
$K_{m,Ala,2}$	2,19	$r_{4,max}$	21,7
$K_{m,Glc,3}$	7,13	$r_{5,max}$	0,345
$K_{m,Asn,3}$	6,84	$r_{6,max}$	49,4
$K_{m,Xt,4}$	0,0315	$r_{7,max}$	3,03
$K_{m,Glc,4}$	1,29	$K_{I,Asn,1}$	16,1
$K_{m,Gln,4}$	2,19	$K_{I,Lac,3}$	0,681
$K_{m,Asn,4}$	1,68	$K_{I,Asp,5}$	9,74
$K_{m,Glc,5}$	100	$K_{I,Asn,5}$	1,10
$K_{m,Gln,5}$	28,2	$\mu_{d,max}$	0,125
$K_{m,Asp,5}$	102	$K_{d,Lac}$	1,01
$K_{m,Glc,6}$	0,0451	$C_{Lac,cr}$	1,22
$K_{m,Gln,6}$	0,791	K_l	0,00843
$K_{m,Asp,6}$	1,06	$K_{m,Glc,7}$	0,0145

10 Schritt k)

Das Modell, bestehend aus der Matrix L , den Kinetiken aus Tabelle 4 sowie den Kinetiken der Apoptose mit den mit den dazugehörigen Parameterwerten aus Tabelle 5 wurde ausgegeben.

Symbolverzeichnis

<u> </u> (Unterstrich)	Bezeichnet einen Vektor
i (Index i)	Bezeichnet das i -te Element eines Vektors
k (Index k)	Bezeichnet das k -te Element eines Vektors
[]	Bezeichnet die Konzentration der in der Klammer stehenden Komponente
C	Konzentration

ΔC	Konzentrationsdifferenz
C^{Int}	Interpolierte Konzentration
\hat{C}	Geschätzte Konzentration (z.B. durch das Lösen einer Differentialgleichung)
C_s	Geshiftete Konzentration
C_{cr}	Kritische Konzentration
C^m	Gemessene Konzentration
D	Verdünnungsrate
q	Ermittelte zellspezifischen Ausscheide- und Aufnahmerate
\tilde{q}	Ermittelte zellspezifischen Ausscheide- und Aufnahmerate, die von einer beliebigen Einheit auf $\left[\frac{\text{Stoffmenge}}{\text{Zeit} \cdot \text{Zellzahl}}\right]$ umgerechnet wurde
r	Ermittelte Reaktionsrate
\hat{r}	Geschätzte Reaktionsrate (z.B. durch das Berechnen einer Reaktionskinetik)
\tilde{r}	Limitierung einer Kinetik
r_{max}	Parameter einer Reaktionskinetik
N	Stöchiometrische Matrix
N_p	Externe stöchiometrische Matrix
K	Matrix, die Makroreaktionen enthält
E	Matrix, die alle Elementary Modes enthält
X_t	Gesamtzellzahl
X_v	Vitale Zellzahl
μ	Wachstumsrate
μ_d	Absterberate
$\tilde{\mu}$	Wachstumsrate, die von einer beliebigen Einheit auf $\left[\frac{\text{Stoffmenge}}{\text{Zeit} \cdot \text{Zellzahl}}\right]$ umgerechnet wurde
K_d	Lysis-rate
K_I	Parameter einer Inhibierungs-Limitierung
K_M	Parameter einer Substratlimitierung
n	Hill-Parameter einer Inhibierungs- oder Substratlimitierung
L	Subset der Makroreaktionen, das für das Modell verwendet wird
p	Modellparameter
S	Substrat

SSR_q	Summe der quadrierten Residuen der spezifischen Aufnahme- oder Abgaberaten
SSR_C	Summe der quadrierten Residuen der Konzentration
SSR_r	Summe der quadrierten Residuen der Reaktionsraten

Beschreibung der Abbildungen:

Abbildung 1 zeigt das Shiften von Messdaten: Dargestellt ist der tatsächliche Verlauf einer gemessenen Größe ($C_i(t)$), der sich bei Änderungen der Verdünnungsrate ($D(t)$) sprunghaft ändert. Der geshiftete Verlauf ($C_{i,s}(t)$) kommt nur durch von der Zelle verursachte Änderungen zu Stande.

Abbildung 2 zeigt die Flux Map zweier spezifischer Raten q_1 und q_2 . Die Höhenlinien geben die Häufigkeit an, mit der die jeweilige Kombination der Raten in den gemessenen Daten vorkommt.

Abbildung 3 zeigt eine dreidimensionale Darstellung des Lösungsraumes, der durch eine positive Linearkombination von EMs aufgespannt wird. In schwarz ist der Lösungsraum des gesamten Sets, in grau der eines Subsets dargestellt.

Abbildung 4 zeigt die Flux Map zweier spezifischer Raten q_1 und q_2 . Als Vektoren sind die 2-dimensionalen Projektionen der Makroreaktionen eines Sets L dargestellt.

Abbildung 5 zeigt eine schematische Darstellung des Metabolischen Netzwerkes aus Niu et al. Hierbei ist die Begrenzung der Zelle als Kasten gezeigt. Die zellinterne Abgrenzung des Mitochondriums ist mit einer gestrichelten Linie gekennzeichnet. Externe Komponenten sind mit dem Index „xt“ gekennzeichnet. Die Pfeile und gepunkteten Pfeile kennzeichnen Reaktionen.

Abbildung 6 zeigt die Messdaten einer Fermentation mit Hybridoma-Zellen aus Baughman et al. Die Gesamtzellzahl (total Cells) ist hierbei aus der Summe der lebenden Zellen (vital Cells) und toten Zellen (dead Cells) berechnet. Die Abkürzungen GLC, GLN, ASP, ASN, LAC, ALA und PRO bezeichnen das Substrat Glucose und die Aminosäuren Glutamin, Asparaginsäure, Asparagin, Alanin und Prolin sowie das Stoffwechselprodukt Laktat. Die Abkürzung MAB bezeichnet das Produkt monoklonaler Antikörper und BM die Biomasse.

Abbildung 7 zeigt die Wachstums- und Absterberaten sowie die zellspezifischen Aufnahme- und Ausscheideraten. Alle zellspezifischen Raten bis auf q_{MAB} sind in $\left[\frac{mM}{h \cdot 10^9 \text{ Cells}}\right]$ angegeben. Die Rate q_{MAB} ist in $\left[\frac{10^{-4} mM}{h \cdot 10^9 \text{ Cells}}\right]$ angegeben.

Abbildung 8 zeigt die mit der „linearen Schätzung von Reaktionsraten ausgewählter Makroreaktionen“ mit dem gewählten Reaktionsset approximierten Konzentrationen. Die Gesamtzellzahl (X_t) sowie die Antikörperkonzentration (MAB) wurden dafür auf C-mol umgerechnet.

Abbildung 9 zeigt die kleinste ermittelte Summe der Fehlerquadrate ("Minimum error") aufgetragen über der Zahl der Makroreaktionen im Subset (n_R).

Abbildung 10 zeigt die mit der erfindungsgemäßen Methode "lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen" ermittelten Reaktionsraten der Makroreaktionen $\underline{r}(t)$.

Abbildung 11 zeigt die mit der erfindungsgemäßen Methode "lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen" ermittelten Reaktionsraten der Makroreaktionen $\underline{r}(t)$ (durchgehende Linie) zusammen mit den algebraisch berechneten Reaktionsraten $\hat{\underline{r}}(\underline{p}, \underline{C}^{int}(t))$ (gestrichelte Linie)

Abbildung 12 zeigt einen Vergleich der gemessenen Konzentrationen $\underline{C}^m(t)$ (Punkte) und dem simuliertem Prozessverlauf $\hat{\underline{C}}(t)$ (durchgehende Linie). Die Konzentrationen sind in $[mM]$ angegeben. Ausnahmen sind die vitale und gesamte Zellzahl (X_v/X_t in $[10^9 \text{ cells} / l]$) und die Konzentration des Antikörpers (mAb in $[10^{-4} mM]$).

Ansprüche:

1. Computerimplementiertes Verfahren zur Erstellung eines Modells einer Bioreaktion mit einem Organismus, das folgende Schritte umfasst:
 - a. Ausgewählte Stoffwechselwege des Organismus, deren Stöchiometrie- sowie Reversibilitätseigenschaften werden als Hintergrundwissen in das Verfahren eingegeben und Elementary Modes werden aus dieser Eingabe berechnet.
 - b. Die Elementary Modes aus a) werden in einer Matrix K zusammengefasst, wobei die Elementary Modes die Stoffwechselwege aus a) in Makroreaktionen zusammenfassen und die Matrix K die Stöchiometrie und die Reversibilitätseigenschaften aller Makroreaktionen enthält.
 - c. Die Messdaten zur Bioreaktion mit dem Organismus werden eingegeben.
 - d. Mit Hilfe einer Interpolationsmethode werden auf Basis der eingegebenen Messdaten aus c) die für den Organismus spezifischen Raten - Ausscheide- und Aufnahmearten von einer oder mehreren Eingangsgrößen und Ausgangsgrößen - der eingegebenen Stoffwechselwege berechnet.
 - e. Relevante Makroreaktionen werden in Form eines Subsets der Elementary Modes aus a) ausgewählt durch
 - i. datenunabhängige und / oder datenabhängige Vorreduktion der Anzahl der Elementary Modes aus a),
 - ii. Auswahl des Subsets aus der Vorreduktion aus e) i. mit den Messdaten aus c) und / oder einer oder mehrerer Raten aus d) mittels eines Algorithmus nach einem mathematischen Gütekriterium und Zusammenfassung des Subsets in einer Matrix L ,
 - iii. Optional erfolgt eine Darstellung des Subsets grafisch.
 - f. Mit Hilfe einer Interpolationsmethode werden auf Basis der eingegebenen Messdaten aus c) und / oder der Raten aus d), die Reaktionsraten der Makroreaktionen des Subsets $\underline{r}(t)$ berechnet.
 - g. Kinetiken der Makroreaktionen des Subsets aus e) ii. werden mit folgenden Zwischenschritten entworfen; dadurch werden die Modellparameter definiert.
 - i. Aus der Stöchiometrie der Makroreaktionen werden generische Kinetiken entworfen.

- ii. Einflussgrößen auf die Makroreaktionen werden aus den Reaktionsraten aus f) ermittelt.
 - iii. Die generischen Kinetiken aus g) i. werden um Terme erweitert, die die in g) ii. ermittelten Einflussgrößen quantifizieren.
 - h. Optional erfolgt aus den Kinetiken aus g) eine erste Anpassung der Modellparameterwerte für jede Makroreaktion separat an die berechneten Reaktionsraten aus f) für jede Makroreaktion separat.
 - i. Optional werden die Schritte g) und h) wiederholt bis eine vordefinierte Anpassungsgüte erreicht wird.
 - j. Die Modellparameterwerte werden an die Messdaten aus c) angepasst.
 - k. Die Matrix L, die Kinetiken aus g) und die Modellparameterwerte aus j) bilden das Modell und werden ausgegeben und / oder in ein Prozessführungs- oder Prozessentwicklungsmodul übertragen.
2. Computer-implementiertes Verfahren nach Anspruch 1, wobei im Schritt d) auch Wachstumsraten, besonders bevorzugt auch Absterberaten des Organismus berechnet werden.
3. Computer-implementiertes Verfahren nach einem der Ansprüche 1 oder 2, wobei im Schritt g) eine individuelle Anpassung der Kinetiken basierend auf einer Analyse der Reaktionsraten aus f) erfolgt.
4. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 3, wobei im Schritt h) die Anpassung der Parameterwerte der Kinetiken aus g) durch Kombination von mehreren Anpassungsmethoden erfolgt.
5. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 4, wobei im Schritt e) ii. zur Auswahl des Subsets von Makroreaktionen eine lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen durchgeführt wird.
6. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 5, wobei im Schritt e) ii. zur Auswahl des Subsets von Makroreaktionen eine lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen in Kombination mit einem evolutionären Algorithmus durchgeführt wird.

7. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 6, wobei die Messdaten vor Anwendung der Interpolationsmethode im Schritt d) geshiftet werden, um die Beschreibung eines konstanten Verbrauchs ohne Feedpeaks zu erreichen.
8. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 7, wobei im Schritt f) eine lineare Schätzung von Reaktionsraten ausgewählter Makroreaktionen durchgeführt wird.
9. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 8, wobei im Schritt e) i. eine datenabhängige Vorreduktion durchgeführt wird und für diese die Methode der linearen Schätzung von Reaktionsraten ausgewählter Makroreaktionen mit NNLS verwendet wird.
10. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 9, wobei im Schritt e) iii. die Validität der Auswahl des Subsets von Makroreaktionen mit Hilfe einer Flux-Map geprüft wird.
11. Computer-implementiertes Verfahren nach einem der Ansprüche 1 bis 10, wobei im Schritt e) ii. die Auswahl aus der Vorreduktion aus e) i. mit den Messdaten aus c) erfolgt.
12. Computerprogramm zur Durchführung der Verfahrensschritte nach einem der Ansprüche 1 bis 10.
13. Software zur Durchführung der Verfahrensschritte nach einem der Ansprüche 1 bis 11.

Figures

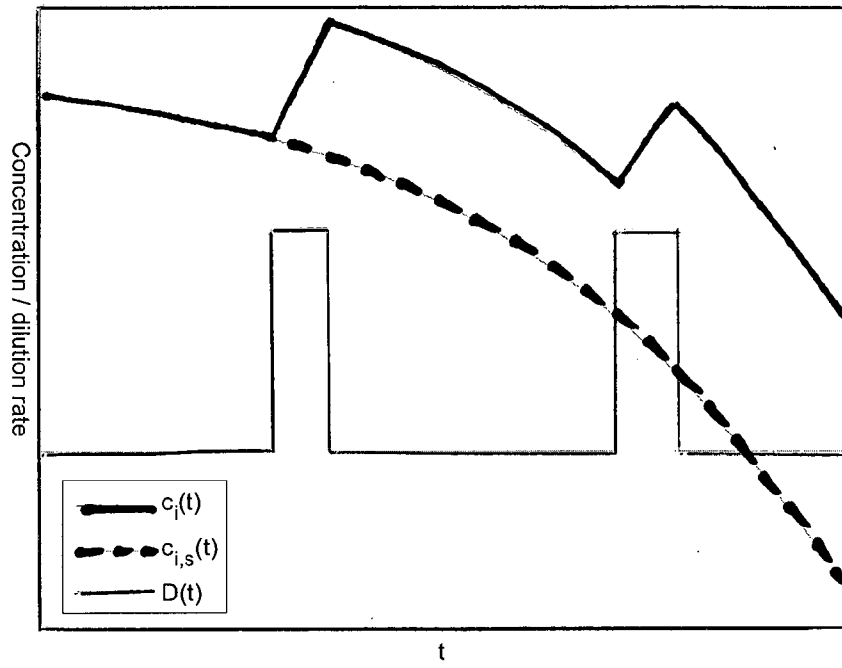


Figure 1

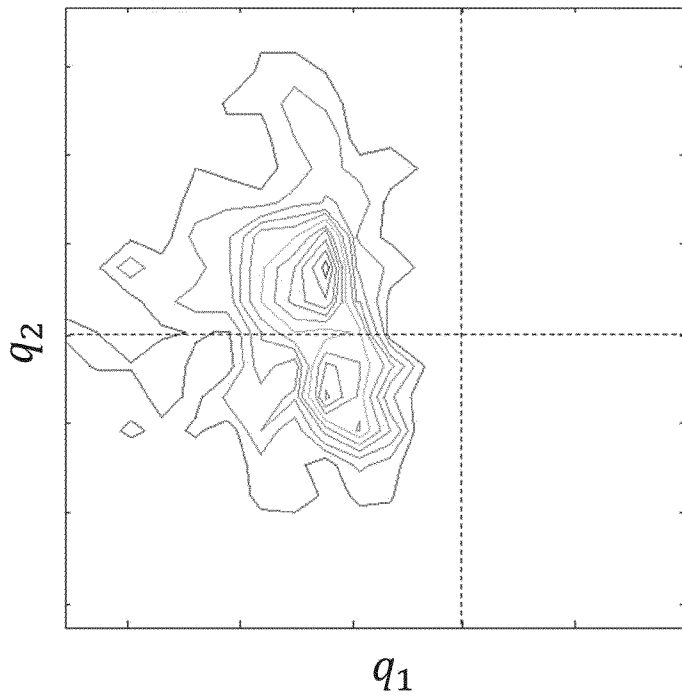


Abbildung 2

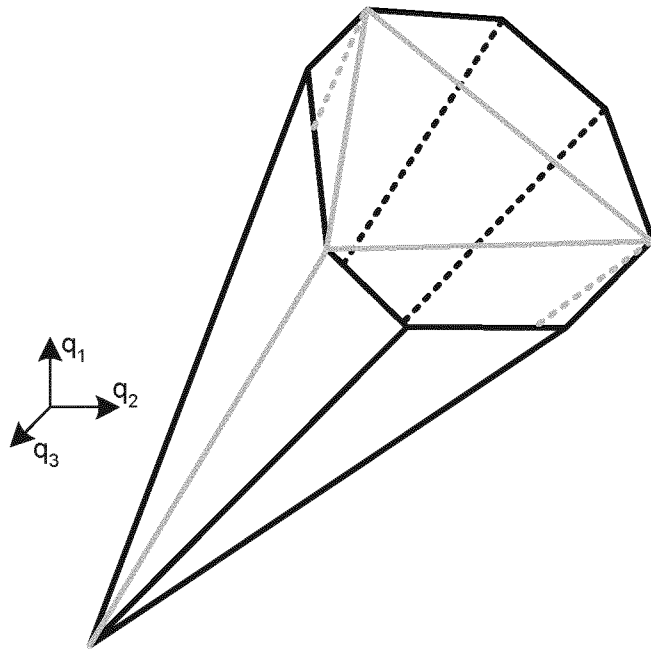


Abbildung 3

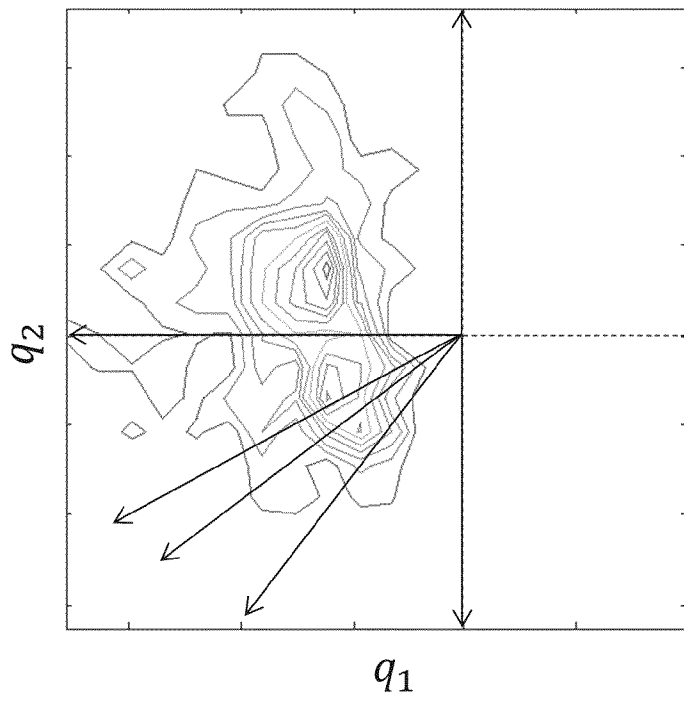


Abbildung 4

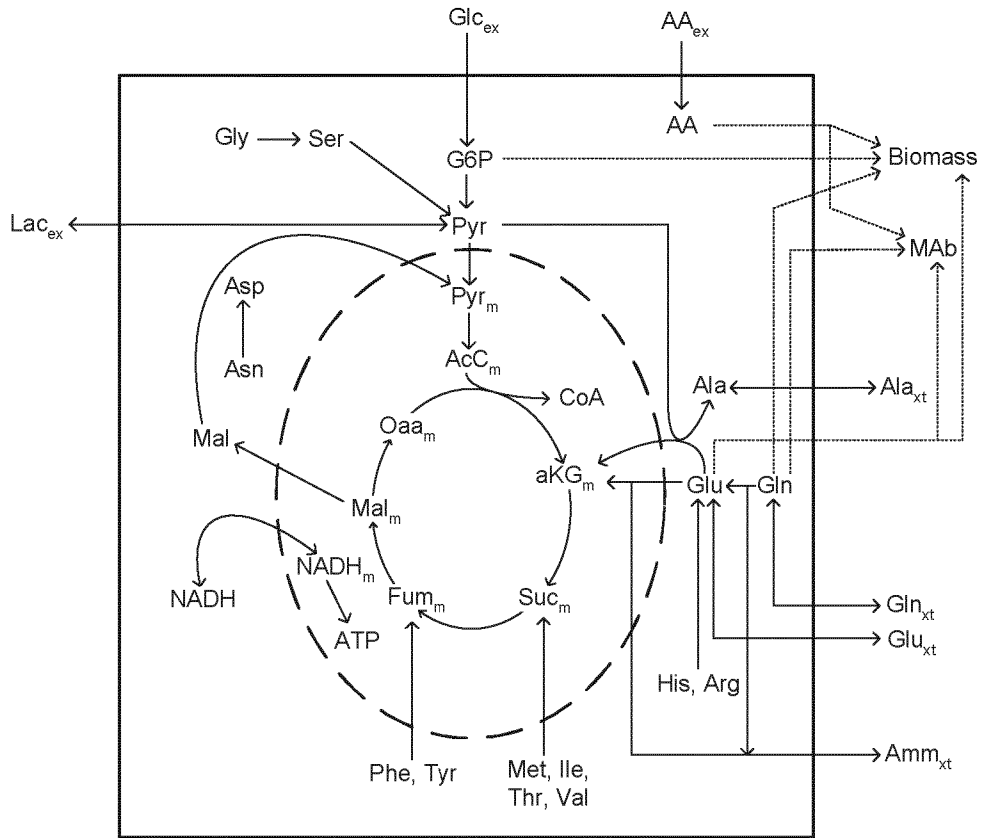


Abbildung 5

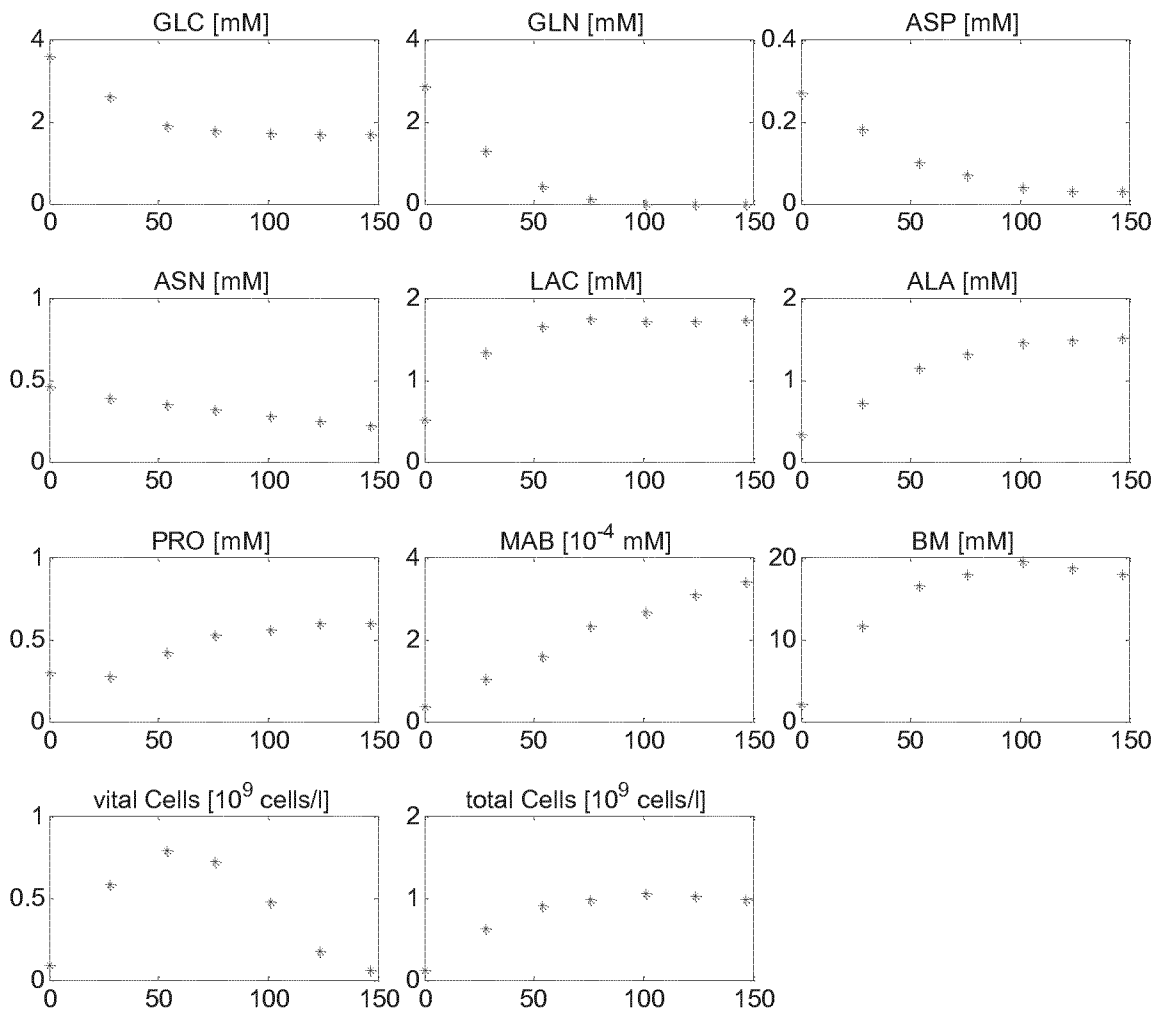


Abbildung 6

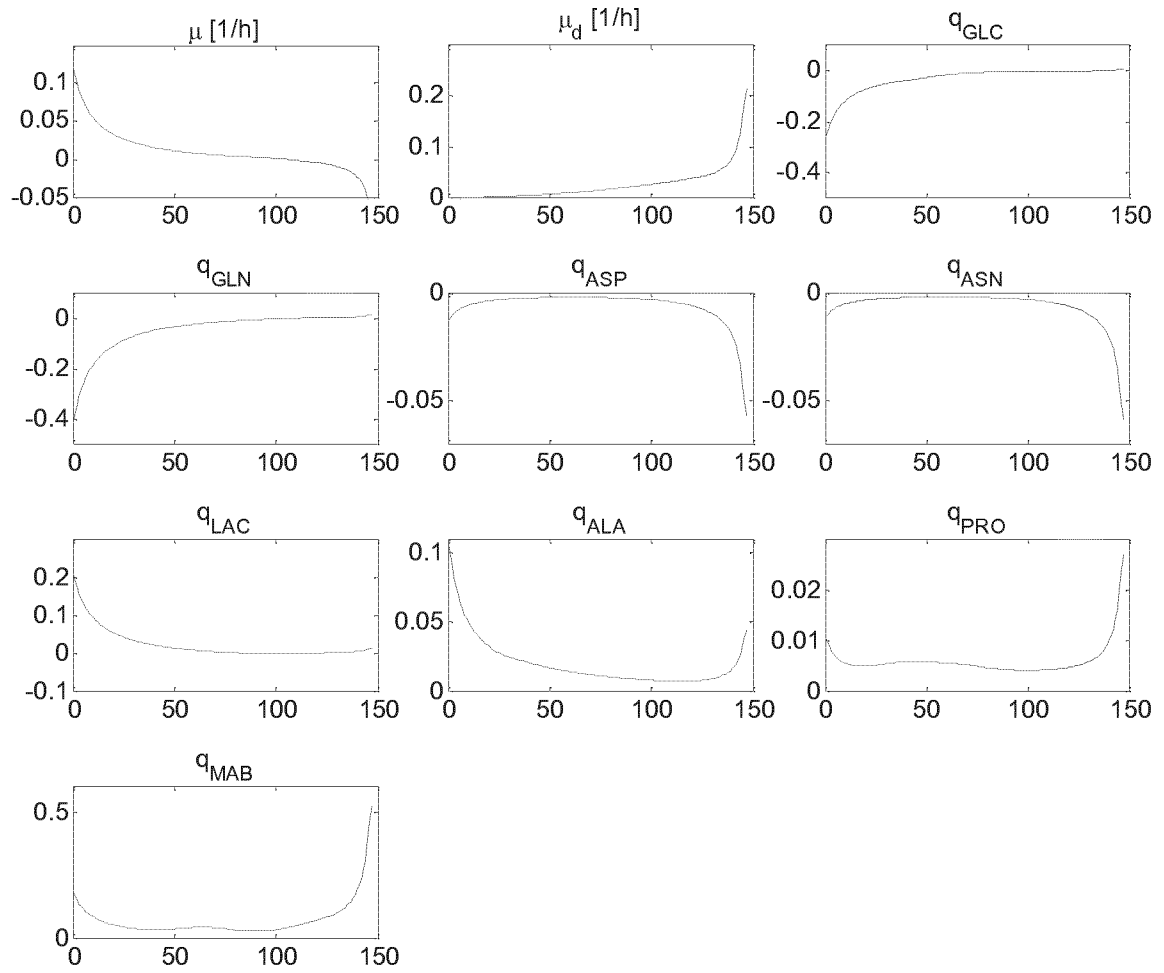


Abbildung 7

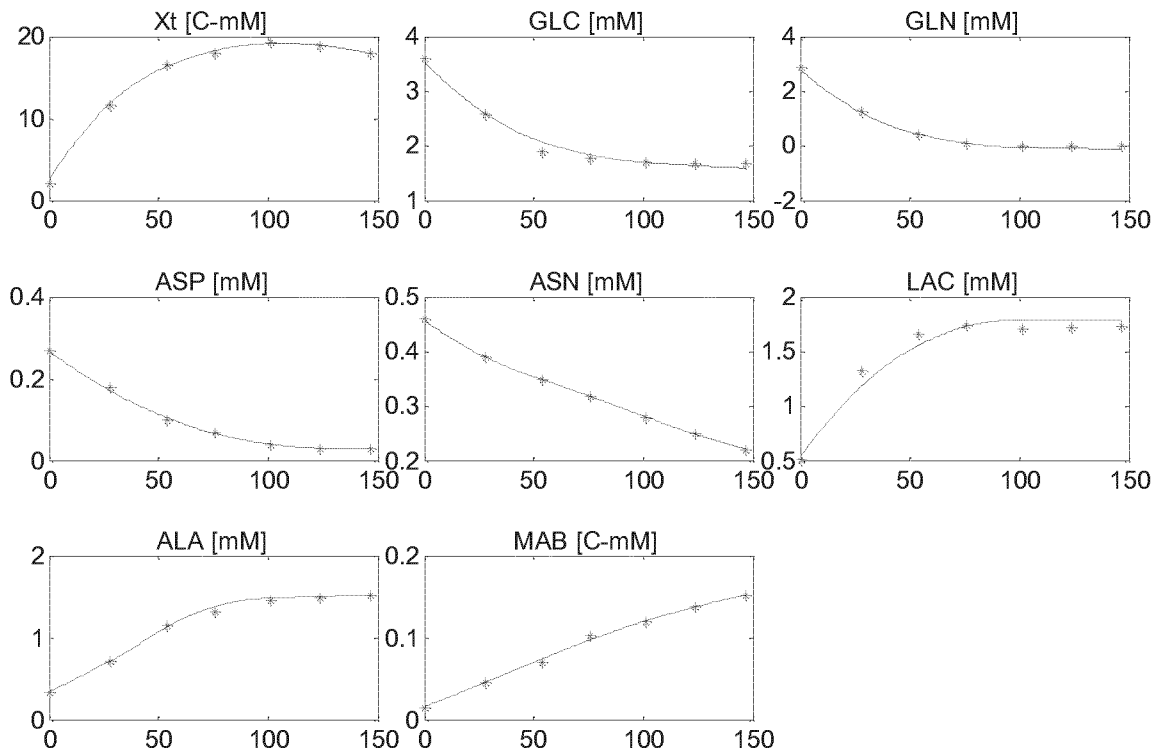


Abbildung 8

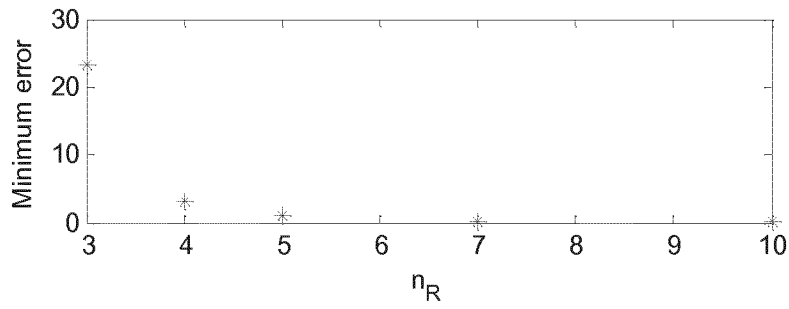


Abbildung 9

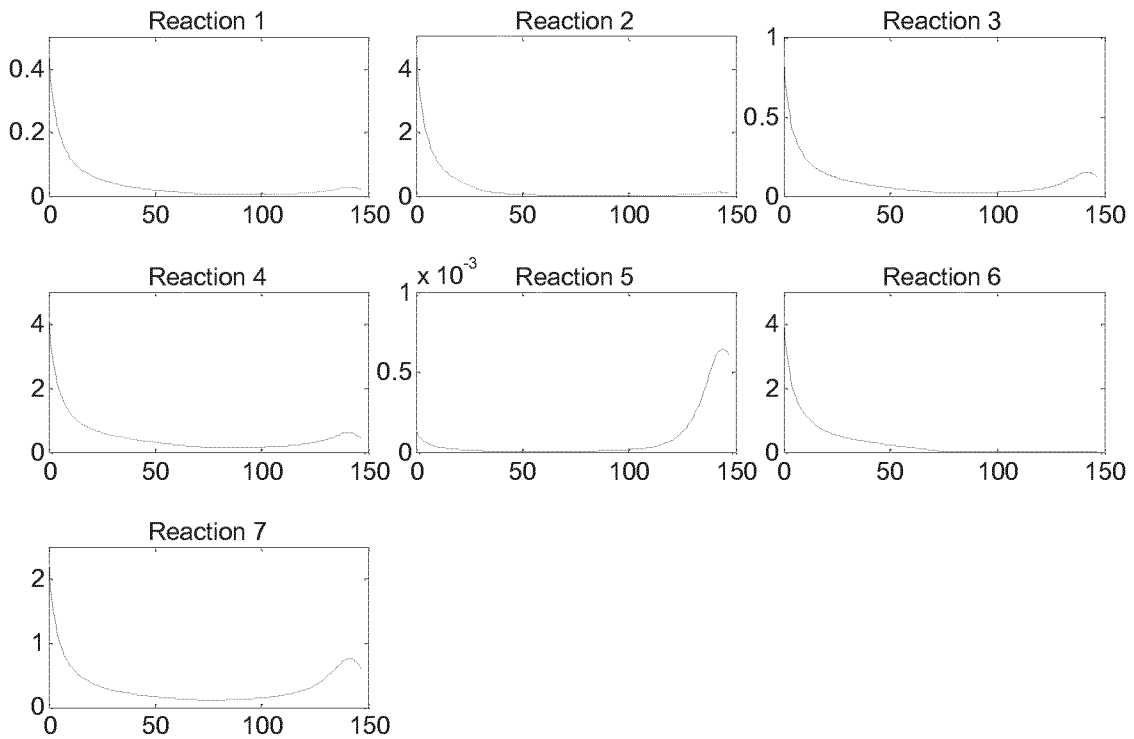


Abbildung 10

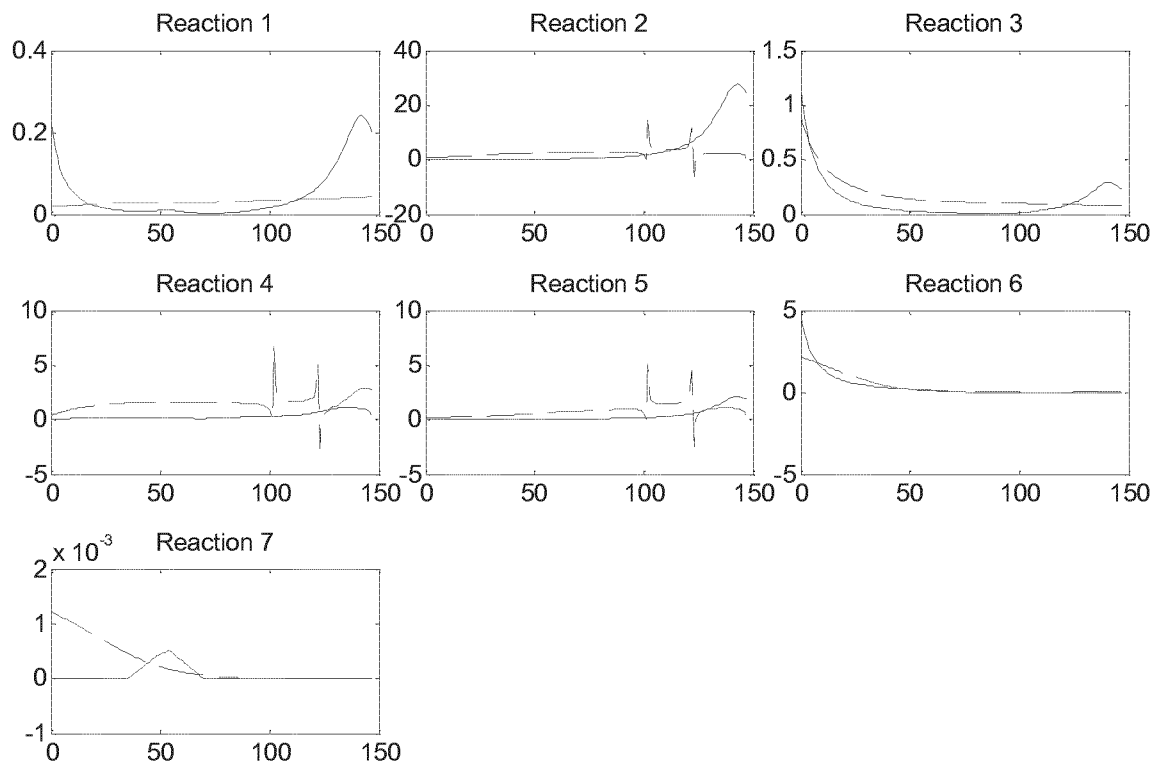


Abbildung 11

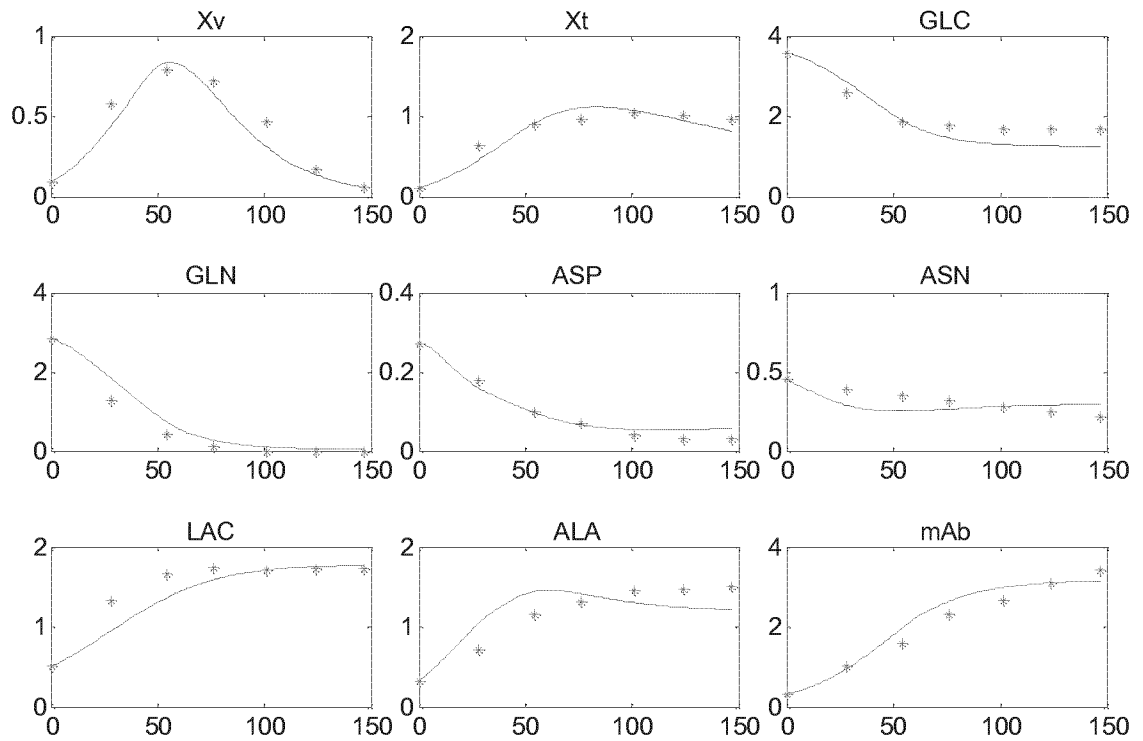


Abbildung 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/051753

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/12
ADD. G06F19/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	J. GAO ET AL: "Dynamic Metabolic Modeling for a MAB Bioprocess", BIOTECHNOLOGY PROGRESS, vol. 23, no. 1, 2 February 2007 (2007-02-02), pages 168-181, XP055199374, ISSN: 8756-7938, DOI: 10.1021/bp060089y abstract page 169, column 1, paragraph 3 - page 170, column 1, paragraph 1 page 175, column 1, paragraph 1 - paragraph 3 ----- -/--	1-13

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier application or patent but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search 21 March 2016	Date of mailing of the international search report 01/04/2016
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Schmitt, Constanze

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/051753

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>PROVOST A ET AL: "Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells", BIOPROCESS AND BIOSYSTEMS ENGINEERING, SPRINGER, BERLIN, DE, vol. 29, no. 5-6, 30 September 2006 (2006-09-30), pages 349-366, XP019461015, ISSN: 1615-7605, DOI: 10.1007/S00449-006-0083-Y abstract page 350, column 2, paragraph 1 - page 352, column 2, paragraph 5 figures 1,6 page 171, column 2, paragraph 3 - page 173, column 1, paragraph 1 -----</p>	1-13

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

INV. G06F19/12

ADD. G06F19/26

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

G06F

Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	J. GAO ET AL: "Dynamic Metabolic Modeling for a MAB Bioprocess", BIOTECHNOLOGY PROGRESS, Bd. 23, Nr. 1, 2. Februar 2007 (2007-02-02), Seiten 168-181, XP055199374, ISSN: 8756-7938, DOI: 10.1021/bp060089y Zusammenfassung Seite 169, Spalte 1, Absatz 3 - Seite 170, Spalte 1, Absatz 1 Seite 175, Spalte 1, Absatz 1 - Absatz 3 ----- -/--	1-13



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

"E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

21. März 2016

Absendedatum des internationalen Recherchenberichts

01/04/2016

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Schmitt, Constanze

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	<p>PROVOST A ET AL: "Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells", BIOPROCESS AND BIOSYSTEMS ENGINEERING, SPRINGER, BERLIN, DE, Bd. 29, Nr. 5-6, 30. September 2006 (2006-09-30), Seiten 349-366, XP019461015, ISSN: 1615-7605, DOI: 10.1007/S00449-006-0083-Y Zusammenfassung Seite 350, Spalte 2, Absatz 1 - Seite 352, Spalte 2, Absatz 5 Abbildungen 1,6 Seite 171, Spalte 2, Absatz 3 - Seite 173, Spalte 1, Absatz 1 -----</p>	1-13