

# Vorlesung “Modellierung von Zellprozessen”

Wolfram Liebermeister, WS 2009/2010

Vorläufige Version des Skriptums (Oktober 2009)

## Zusammenfassung

Die Vorlesung behandelt grundlegende Konzepte und Methoden der mathematischen Modellierung von Zellen. Von den experimentellen Daten zum fertigen Modell ist es meist ein langer Weg: die Modellstruktur soll die entscheidenden Prozesse in der Zelle wiedergeben und unwichtiges Beiwerk ausklammern, Parameter müssen geschätzt und das Verhalten des Modells analysiert werden.

## 1 Einführung

### 1.1 Wozu Modelle?

**Definition** Ein Modell ist ein vereinfachtes Abbild der Wirklichkeit.

#### Beispiele

- Flugzeugmodell
- Puppe
- mathematisches Modell

#### Beispiele für Zellmodelle/biologische Modelle:

- **Vorgänge:** Zellstoffwechsel, Zellzyklus, Signalwege, Transkription und Translation, Kalziumdynamik, Krebswachstum, Neuronenaktivität, Pharmokokinetik, Epidemien
- **Mathematisch:** Molekülbewegung, Molekülanzahlen in diskreten Kompartimenten, partielle Differentialgleichungen, Konzentrationen

#### Und wozu das Ganze?

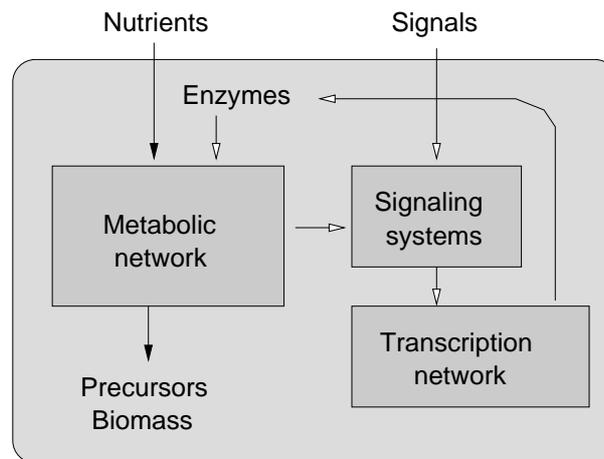
- Modelle reduzieren Wirklichkeit auf Teilaspekte
- diese Teilaspekte sind bestimmt durch unser Verständnis, was wesentlich ist
- Vereinfachung soll Komplexität begreifbar machen
- Modelle sollten in der Lage sein testbare Voraussagen zu machen

#### Verständnis komplexer biologischer Systeme

- Grundeinheiten der Genetik: einzelne Gene und ihre Produkte

- Nur im Rahmen des Gesamtsystems gewinnen die einzelnen Gene ihren Sinn: wie funktioniert ihr Zusammenspiel, das den Phänotyp bestimmt?
- Netzwerke: nur aussagekräftig mit dynamischer Beschreibung
- Biologische Prinzipien: Evolution, Robustheit, Modularität, Optimierung

**Paradigma der Systembiologie** Mathematische Modelle können Laborexperimente im Vorhinein *in silico* testen, um möglichst aussagekräftige Experimente zu entwickeln und so Zeit und Kosten sparen.



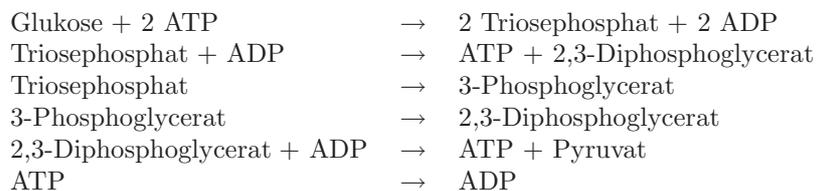
## 1.2 Wie simuliert man Zellen? Kinetisches Modell der Glykolyse als Beispiel

Glykolyse: Glukose  $\rightarrow$  ATP, Pyruvat (anaerober Vorgang)

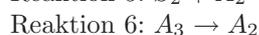
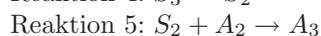
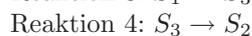
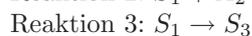
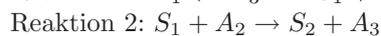
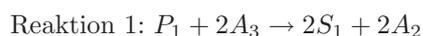
Bild der Glykolyse im Gesamtstoffwechsel (Boehringerkarte)

### 1.2.1 Skelettmodell (Heinrich et al, 1987)

#### Summenformeln



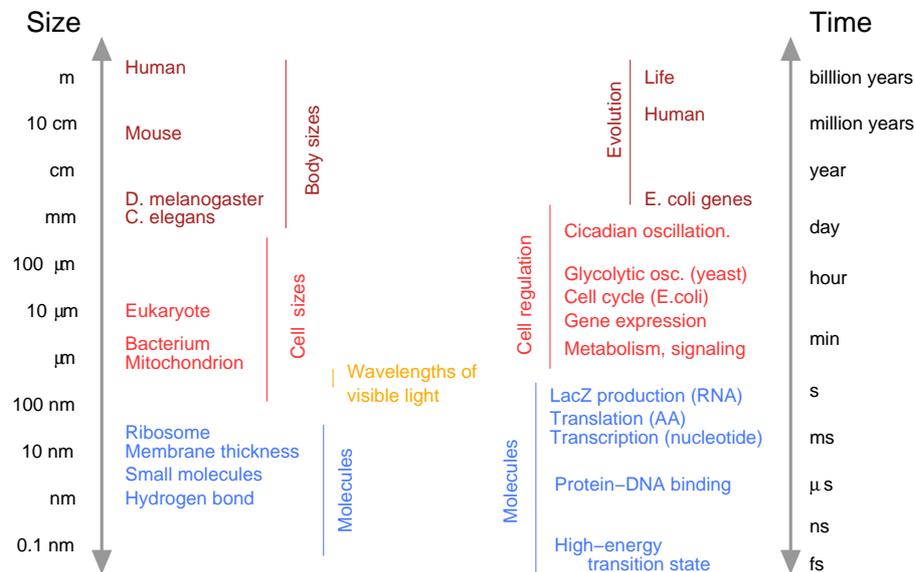
Mit **Abkürzungen:** Glukose ( $P_1$ ), Triose-Phosphat ( $S_1$ ), 2,3-Diphosphoglycerat ( $S_2$ ), 3-Phosphoglycerat ( $S_3$ ), ADP ( $A_2$ ), ATP ( $A_3$ )



#### Bedeutung der Bilanzgleichungen

- Bilanz einzelner Reaktionsereignisse

- Reaktionsrichtung ist Konventionssache
- Stöchiometrische Koeffizienten



### 1.2.2 Kinetisches Modell (wichtigste Modellvariante für biochemische Vorgänge)

Menge von Stoffen  $S_i$  in Kompartiment mit Volumen  $\Omega$ :

**Molekülzahl**  $X_i \rightarrow$  Menge  $a_i = X_i/N_A$  (in mol) mit Avogadrozahl  $N_A =$  Zahl der Moleküle / Mol

**Konzentration**  $c_i = a_i/\Omega$  in mM = mmol/l = mol/m<sup>3</sup>)

(kann reelle Werte annehmen, Methoden der Analysis sind anwendbar, z.B. Differentialgleichungen)

**Reaktionsgeschwindigkeit**  $v_r =$  durchschnittliche Zahl von Reaktionsereignissen (Zeitmittel) / ( $N_A\Omega$ )

Einheit: 1 mol/s, manchmal auch mM/s

**Bilanzgleichung** (z.B. für Triosephosphat): Produktion - Abbau

$$d/dtc_1 = 2v_1 - v_2 - v_3 \quad (1)$$

- Stoffkonzentrationen werden durch chemische Reaktionen verändert (und sonst durch nichts!)
- Transport gilt als chemische Reaktion
- erzeugende Reaktion:  $+v$ , verbrauchende Reaktion  $-v$
- Stöchiometrische Koeffizienten
- wenn alle  $v_r(c)$  und  $c_i(0)$  bekannt sind  $\rightarrow$  numerische Lösung  $c_i(t)$
- Abstraktion: kontinuierliche Konzentrationen und Vorgänge

### 1.2.3 Bilanzgleichung für Stoffkonzentrationen

Allgemeine Form der Differentialgleichungen (hier mit Reaktionsraten in mM/s)

$$dc_i(t)/dt = \sum_r n_{ir}v_r(c(t)) \quad (2)$$

In Matrixform:

$$dc(t)/dt = Nv(c(t)) \quad (3)$$

**Stöchiometrische Matrix**

$$N = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 2 & -1 & 0 & 0 & -1 & 1 \\ -2 & 1 & 0 & 0 & 1 & -1 \end{pmatrix} \quad (4)$$

**Matrixmultiplikation** (Skalarprodukt aus Zeilen und Spalten):

$$C = AB \Leftrightarrow c_{il} = \sum_k a_{ik} b_{kl} \quad (5)$$

**ÜA:** Wie müssen demnach die Matrixdimensionen aussehen?

**Erhaltungsbedingung:** ATP+ADP=const (siehe Schema bzw. Matrix)

Setze darum  $a = [\text{ATP}] + [\text{ADP}] = a_2 + a_3$ , also  $a_2 = a - a_3$

### 1.2.4 Reaktionsgeschwindigkeiten (=Raten)

Annahme: Raten abhängig von den Konzentrationen (und nur von diesen!!! Räumliche Unterstrukturen der Zelle und andere Einflüsse werden vernachlässigt)

**Bedingungen:** (Beispielreaktion  $A \rightarrow B$ )

- Substratkonzentration  $a = 0 \Rightarrow v \leq 0$
- Produktkonzentration  $b = 0 \Rightarrow v \geq 0$
- Massenwirkungsgesetz: Im chemischen Gleichgewicht ( $v = 0$ ) gilt  $b/a = \text{const} = k^{eq}$  (Gleichgewichtskonstante) unabhängig von den Anfangsbedingungen.

**Massenwirkungskinetik:**  $v = k_+ a - k_- b$

- Vorwärtsreaktion - Rückwärtsreaktion (entspricht mikroskopischem Bild)
- Je mehr Stoff vorhanden, desto häufigere mikroskopische Reaktionsereignisse
- Mikroskopisch:  
Geschwindigkeit = Anzahl der Moleküle · Reaktionswahrscheinlichkeit / (Zeit und Molekül)
- Die Massenwirkungskinetik erfüllt die Richtungsbedingungen:

$$\begin{aligned} b/a < k^{eq} &\Rightarrow v > 0 \\ b/a > k^{eq} &\Rightarrow v < 0 \end{aligned} \quad (6)$$

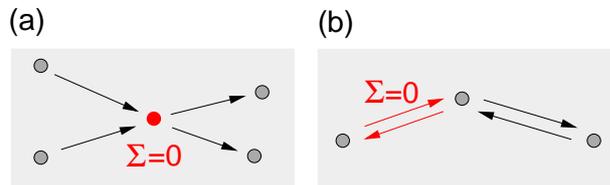
**Bimolekulare Reaktion:**  $A + B \rightarrow C + D$

$$v = k_+ a b - k_- c d \quad (7)$$

Mikroskopisch: Geschwindigkeit = Anzahl der Stöße/Zeit · Reaktionswahrscheinlichkeit / Stoß

Enzymkatalysierte Reaktionen haben unter Umständen komplizierte kinetische Gesetze.  
Die Massenwirkungskinetik ist oft eine gute Näherung für kleine Substratmengen.

### 1.2.5 Stationäre Zustände



Stationärer Zustand ((a), Fließgleichgewicht) und Gleichgewicht (b)

#### Gleichgewicht

- Definition: ein Zustand, in dem keinerlei (makroskopische) Reaktionen mehr ablaufen.
- Das bedeutet: Für jede Reaktion heben sich Vorwärts- und Rückwärtsrate genau auf; das System ist tot
- Mathematisch: Paar  $(c, v)$  (Konzentrationen und Flüsse), sodass  $v(c) = 0$
- Analogie: ruhiger See

#### Stationärer Zustand (=Fließgleichgewicht)

- Definition: Ein Zustand, in dem sämtliche Variablen zeitlich konstant sind.
- Das bedeutet: für jeden Stoff heben sich erzeugende und verbrauchende Reaktionen genau auf; ein solches System kann leben.
- Mathematisch:  $(c, v)$ , sodass  $dc/dt = Nv(c) = 0$
- Analogie: ruhiger Fluss

Stationärer Zustand makroskopisch und mikroskopisch?

**Stabilität eines stationären Zustandes:** Ist ein System zum Zeitpunkt  $t$  in einem stationären Zustand  $c(t) = s$ , so bleibt es für immer in diesem Zustand. Was passiert, wenn es sich bei  $s + \Delta c$  in der Nähe von  $c$  befindet?

#### Näherung der Differentialgleichung

$$dc/dt = f(c) = f(s + \Delta c) \approx f(s) + df/dc \Delta c = M \Delta c \quad (8)$$

mit der Jacobimatrix

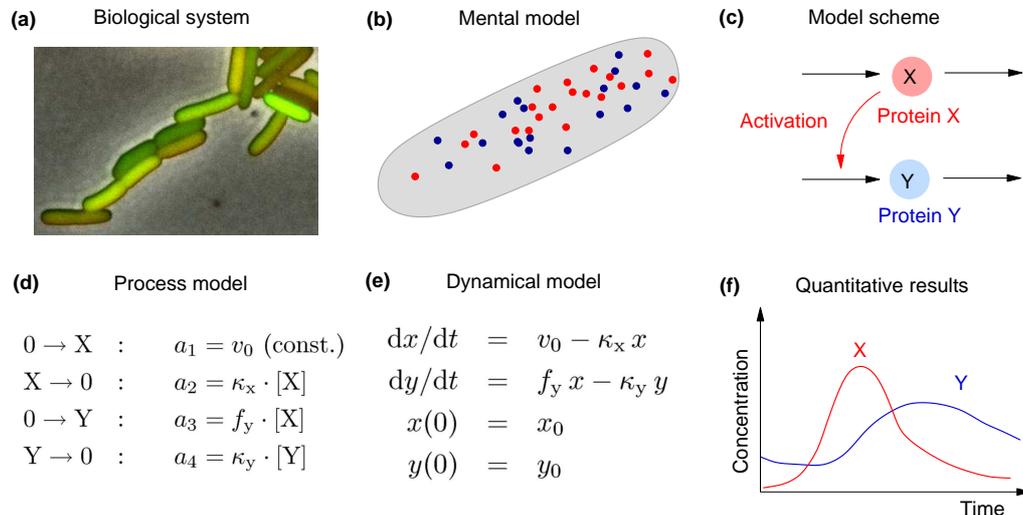
$$M_{il} = df_i/dc_l = \sum_r n_{ir} dv_r/dc_i = \sum_r n_{ir} \varepsilon_{ri} \quad (9)$$

Die Größen  $\varepsilon_{ri} = dv_r/dc_i$  (Ableitungen der Ratengleichungen) heißen unskalierte *Reaktionselastizitäten*.

**Stabilitätskriterium:** Ein Zustand heißt asymptotisch stabil, wenn alle Trajektorien, die in der Nähe des Zustandes starten, gegen diesen konvergieren. Die Stabilität eines stationären Zustandes lässt sich im Allgemeinen an den Eigenwerten der Jacobimatrix ablesen: sind die Realteile aller Eigenwerte negativ, so ist der Zustand stabil.

**Eigenwertgleichung:**  $Mx = \lambda x$  mit Vektor  $x$  and Skalar  $\lambda$ .

## 1.3 Der Modellierungsprozeß



### 1.3.1 Der Modellierungsprozeß in 8 Schritten

1. Beobachtung / Messung / Daten oder allgemein: ein **Problem**
2. **Hypothese** über die zugrundeliegenden Prozesse oder Mechanismen → **Gedankenmodell**
3. Präzisierung: das **Wortmodell**
  - Welche Frage soll das Modell beantworten?
    - ⇒ Auswahl der **Komponenten** und deren **Interaktionen** → System
    - ⇒ Systemgrenzen festlegen (zeitlich, räumlich)
  - Das Wortmodell sollte als **Diagramm** festgehalten werden, welches alle Komponenten und deren Interaktion zeigt.
  - Beispiele
    - Für Oszillationen braucht man in der Regel eine Art Feedback, d.h. der Ausgang (**output**) muß den Eingang (**input**) des Systems beeinflussen.
    - Für metabolische Prozesse im Fließgleichgewicht (**steady state**) können Enzymkonzentrationen oft als konstant angenommen werden. Innerhalb bestimmter zeitlicher Grenzen können also Transkription und Translation vernachlässigt werden.
4. Formalisierung: das **mathematische Modell**
  - Auswahl des Formalismus, so zum Beispiel
    - deterministisch (Differentialgleichungen)
    - stochastisch (Gillespie, Monte-Carlo, Markov)
    - regelbasiert (Petri-Netze, zelluläre Automaten)
    - bedingungs basiert (constraint based – basierend auf der stoichiometrischen Matrix und Nebenbedingungen, z.B. **Flussbilanzanalyse** [FAB])
    - Logisches Modell (Boolesches Netz)
    - Bayessche Netze
5. **Verifizierung**
  - qualitative Analyse des Modells
  - Ist das Modell im Prinzip in der Lage die gestellten Fragen zu beantworten?

- Sind zum Beispiel Oszillation möglich?
- Prüfen, ob Fixpunkte usw. Sinn ergeben.
- In der Praxis: ausrechnen, aber in der Regel rumspielen mit Parametern und Struktur

## 6. Validierung

- Quantitative Analyse des Modells
- Kann das Modell die Daten reproduzieren?
- Parameter anpassen
- oft schwierig und langwierig, daher lohnt es sich vorher zu verifizieren

## 7. Sensitivitätsanalyse

- systematisches Testen der Modelleigenschaften
- Einfluß von Anfangs- und Eingangswerten
- Parametersensitivitäten
  - *sensitive Parameter* haben bei kleinen Änderungen große Auswirkungen
  - sollten genau gemessen sein
  - *unsensitive* (bzw. komplett unbestimmbare) Parameter legen Vereinfachungen nahe
  - sensitive Parameter oft genau bestimmbar
  - Problem: In der Natur sind Parameter oft insensitiv, d.h. **robust**. Für Modelle sind sensitive Parameter aber angenehmer, da besser bestimmbar.
- Standardmaß für Sensitivität (einfache Parameter)

$$S_p^Z = \frac{\frac{\Delta Z}{Z_0}}{\frac{\Delta p}{p_0}}, \quad Z \triangleq \text{betrachtete Zielfunktion, -größe, } p \triangleq \text{Parameter}$$

wobei die Zielfunktion  $Z$  zum Beispiel steady-state-Konzentrationen, das Maximum aller Zielfunktionen  $\max_t Z$  oder die Summe aller Konzentrationen über ein Zeitintervall  $[0, t]$  also  $\int_0^t c(t)dt$  sein kann.

- In der Praxis: Ändere Parameter  $\pm 10\%$ ,  $50\%$  und betrachte Abweichung auf bestimmte Größe  $Z$ .
- Daumenregel:

$$S_p^Z \ll 1 \Rightarrow \text{insensitiv} \quad S_p^Z \gg 1 \Rightarrow \text{sensitiv}$$

- Metabolische Kontrollanalyse
  - betrachte infinitesimale Änderung des steady-state  $\rightarrow$  Antwortkoeffizienten, z.B. Konzentrationskontrollkoeffizient

$$c_k^0 = \frac{p_k}{c_i} \frac{\partial c_i}{\partial p_k} = \frac{\partial \ln c_i}{\partial \ln p_k}$$

- Sensitivitäten können zeitlich veränderlich sein

$$R_p^c = dc(t, p)/dp$$

- multiple Parametersensitivität
  - $S_p^Z$  hängt von  $p_0$  und von den restlichen konstant gehaltenen Parametern ab.
  - $S_p^Z$  für alle möglichen Parameterkombinationen in der Regel nicht berechenbar.  
Lösung: Monte-Carlo-Analyse: Aus Zufallsstichproben aus Parameterverteilungen, wird Verteilung von  $Z$  ermittelt.

## 8. Vorhersage

- Ist das Modell in der Lage, neue Beobachtungen korrekt vorherzusagen?

**Variante: Kreuzvalidierung.** Um schon mit den gegebenen Daten eine Vorhersage durchzuspielen, werden die Daten in ein Trainingsset und ein Testset eingeteilt. Das Modell wird an die Trainingsdaten angepasst und alsdann werden die Modellvorhersagen mit dem Testset verglichen. Diese Prozedur wird mehrfach mit verschied gewählten Trainings- und Testsets durchgeführt.

Modellierung ist kein streng iteratives Schema (idealerweise eine zirkuläre Abfolge von Experiment, Hypothesenbildung, Modellierung, Vorhersage usw.; in Wirklichkeit aber oft eher eine Küche ...)

### 1.3.2 Direktes und inverses Problem

#### **Definition**

Das *direkte Problem* besteht darin, aus einem Modell Daten abzuleiten, d.h. in der Regel simulieren.

Das *inverse Problem* besteht darin, aus Daten ein Modell abzuleiten.

**Bemerkung** Das direkte Problem ist höchstens technisch schwierig. Das inverse Problem ist im Allgemeinen schwierig. Der Modellierungsprozeß ist ein iterativer Prozeß des direkten und des inversen Problems.