

Strukturelle und funktionale Analyse von zellulären Netzwerken

Steffen Klamt

Modulbeschreibung

Lernziele und erworbene Kompetenzen:

Dieses Modul vermittelt verschiedene theoretische Ansätze und Methoden zur strukturellen und qualitativen Modellierung und Analyse zellulärer Netzwerke. Die Studenten bekommen ein allgemeines Verständnis für den strukturellen Aufbau und die Arbeitsweise unterschiedlicher Klassen von biochemischen Netzwerken und eignen sich verschiedene Methoden für die rechnergestützte Analyse dieser Netzwerke an. Die Verfahren kommen hauptsächlich aus dem Bereich der diskreten Mathematik (z.B. Graphen- und Hypergraphentheorie, Boolesche Netzwerke) und der linearen Algebra. Die Anwendung der theoretischen Methoden wird in Übungen mithilfe eines Softwarepakets und konkreten biologischen Beispielen gezeigt. Das interdisziplinäre (systembiologische) Denken der Studenten und das Verständnis für netzwerkweite Prozesse in der Zelle wird gefestigt. Außerdem werden Methoden zur Bestimmung strategischer Eingriffe und zur Rekonstruktion zellulärer Netzwerke vermittelt.

Inhalte:

- Einführung: zelluläre Netzwerke, Stoffflüsse und Signalflüsse, Datenbanken
- Graphentheoretische Analyse, statistische Kennzahlen, Netzwerk motive
- Metabolische Netzwerkanalyse: Erhaltungsrelationen, Stoffflussverteilungen, Flusskegel, Elementarmoden, Minimal cut sets, Metabolic Engineering
- Modellierung von regulatorischen und Signaltransduktionsnetzen mittels Interaktionsgraphen und Booleschen Netzwerken: Feedback loops, Cut sets, Abhängigkeitsmatrix, Minimale Interventionsstrategien
- Zusammenhänge zwischen Netzstruktur und qualitativer Dynamik
- Einführung in Methoden der Netzwerkrekonstruktion
- Einführung und Rechnerübungen mit dem MATLAB-Paket *CellNetAnalyzer*

Organisatorisches

- Voraussetzungen:
 - Grundlagen in Molekular- und Zellbiologie;
 - Lineare Algebra
 - Grundlagen der Modellierung mit DifferentialgleichungenVon Vorteil, aber nicht zwingend:
 - Graphentheorie, Algorithmen
 - MATLABFür die Übungen: Laptop mit MATLAB
- Termine:
Vorlesung+Übung: im kleinen Seminarraum am MPI (V0.05-1)
- Schein: Klausur
- Webseite: mit allgemeinen Informationen:
<http://www.mpi-magdeburg.mpg.de/klamt/lehre>
→ dort werden auch die Folien (=Skript) + Musterlösungen nach und nach als PDF abgelegt

Literatur

- Bücher (Systembiologie im Allgemeinen mit Bezügen zur Vorlesung)

Rene Thomas and Richard D'Ari: *Biological Feedback*.
CRC Press, Boca Raton, 1990.

Bernhard Palsson: *Systems Biology - Properties of Reconstructed Networks*.
Cambridge University Press: 2006. (ISBN 978-0-521-85903-5)

* E. Klipp, W. Liebermeister, C. Wierling, A. Kowald:
Systems Biology: A Textbook. Wiley-VCH: 2009. (ISBN 978-3527336364)
- weitere (spezifische) Literatur wird an entsprechender Stelle angegeben

Übersicht: Themen

- I: Netzwerke und Systembiologie: Einführung (1 V)
- II: Graphentheoretische und statistische Eigenschaften zellulärer Netzwerkarchitekturen (1,5 V)
- III: Metabolische/Stöchiometrische Netzwerkanalyse (4,5 V)
- IV: Regulatorische und Signaltransduktionsnetzwerke (4 V)
- V: Einführung in Methoden der Netzwerkrekonstruktion (2 V)

I. Netzwerke und Systembiologie

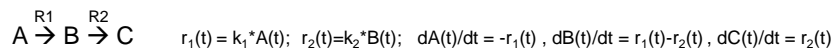
I.1 Netzwerke, Systeme, Systembiologie

- **Netzwerk**: Menge **K** von Elementen (Knoten) und Menge **R** von Relationen (Interaktionen) zwischen den Elementen: **N=(K,R)**



- **System**: Netzwerk mit dynamischem Verhalten: Elemente ändern zeitlich (+ eventuell auch räumlich) ihren Zustand (z.B. Konzentration) durch Wirken der Relationen: **S=(K,R,f)**

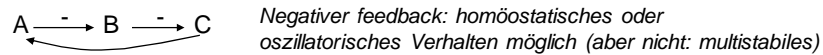
f: Beschreibung wie Relationen quantitativ wirken und dadurch zeitliches Systemverhalten bestimmen (oft: Differentialgleichungen)



Strukturelle + Dynamische Eigenschaften

Strukturelle Netzwerkanalyse und Systemtheorie

- Netzwerke sind die Wirkstrukturen von Systemen
- **statische/strukturelle Netzwerkanalyse**: untersucht topologische Aspekte und parameterfreie (qualitative) Eigenschaften von Systemen
 - + basiert nur auf (oft bekanntem) qualitativem/strukturellem Wissen
 - + wichtige Systemmerkmale (z.B. Feedback loops) oft in Netzstruktur fixiert

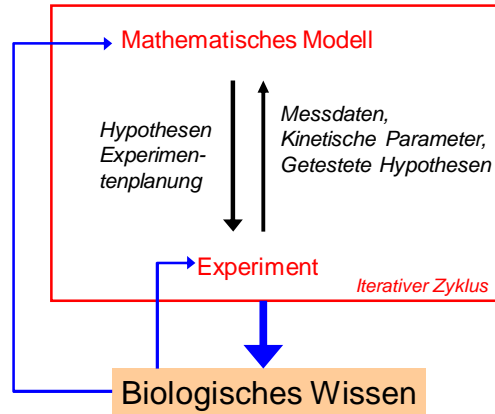
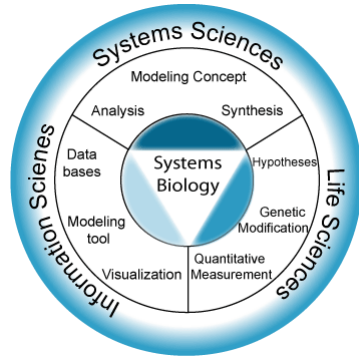


- aber: Prädiktionen des quantitativen und dynamischen Verhaltens limitiert

- **Systemtheorie** untersucht dynamisches Verhalten des Systems
 - + mit Differentialgleichungen kann breites Arsenal an quantitativen und dynamischen Systemeigenschaften untersucht werden
 - + hohes Potential an deskriptiver and prädiktiver Power
 - aber: umfangreiches mechanistisches Wissen + quantitative Daten und Parameter nötig (welche in der Biologie oft nicht vorhanden sind)



Systembiologie

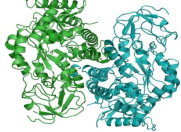


Wichtige Technologien für die Systembiologie:

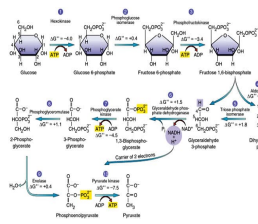
- Hochdurchsatzmessverfahren (Transkriptomics, Proteomics, Metabolomics, ...);
- Standards (Experimente+Modelle)
- spezielle theoretische Methoden (z.B. Parameterschätzung, Netzwerkrekonstruktion),
- spezielle Software;

Systembiologie: kleine und große zelluläre Netzwerke ...

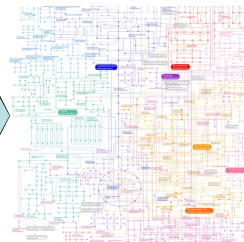
Moleküle



Stoffwechselwege



Netzwerke



Komponenten, Interaktionen

Dynamische Modellierung
(<30 Elemente)

- Differentialgleichungen
- Systemtheoretische Analysen

Strukturelle/Qualitative Modellierung
(>30 Elemente)

- Strukturmodelle
- Qualitative/Topologische Eigenschaften

Thema dieser Vorlesung

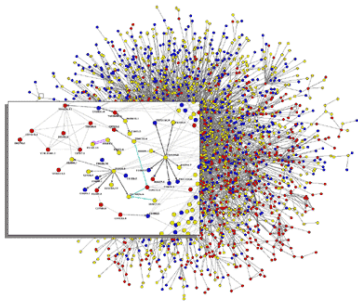
Molekulare und quantitative Details

I.2. Klassen von zellulären Netzwerken

(1) Protein – Protein - Interaktionsnetzwerke



z.B. Proteine A, B und C binden zu einem Protein-Komplex



Beispiel: C. elegans network

Quelle: http://focus.hms.harvard.edu/2004/Feb6_2004/research_briefs.html

Netzwerkmodell: ungerichtete Graphen

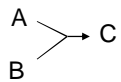
Knoten: Proteine

Relationen:

Typ: ungerichtete Kanten:

Bedeutung: physikalische Interaktionen, Assoziationen

(2) Metabolische Netzwerke (Stoffwechselnetze)



$A + B \rightarrow C$ (Metabolite A und B werden zu C umgesetzt)

Netzwerkmodell:

Stöchiometrische Reaktionsnetzwerke (= gerichtete Hypergraphen), seltener: bipartite Graphen

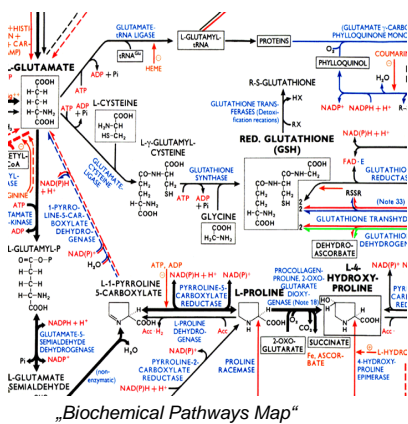
Knoten: Metabolite

Relationen:

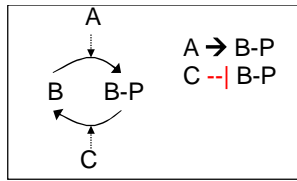
Typ: gerichtete (Hyper)Kanten;

Bedeutung: biochemische Reaktionen (durch Enzyme katalysiert)

Metabolische Netzwerke sind durch **Stoffumwandlungen und Massenflüsse** gekennzeichnet: in jeder Reaktion werden die Edukte verbraucht und die Produkte neu synthetisiert.



(3) Signaltransduktionssysteme



(Protein A aktiviert (z.B. phosphoryliert) Protein B)
(Protein C inhibiert (z.B. dephosphoryliert) Protein B)

Netzwerkmodell:

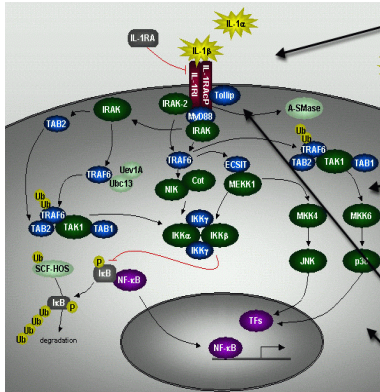
Gerichtete Vorzeichengraphen (rechts),
Boolesche / logische Netzwerke
(selten: stöchiom. Reaktionsnetzwerke (links))

Knoten: Proteine, Metabolite

Relationen:

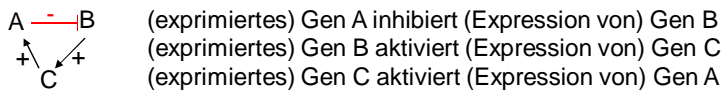
Typ: gerichtete Kanten mit Vorzeichen
oder logische Gatter

Bedeutung: Aktivierungs- und
Deaktivierungsprozesse



Signaltransduktionsnetzwerke sind gekennzeichnet durch **Signal- und Informationsflüsse**.
(Beteiligte Partner in einer Interaktion werden nicht verbraucht oder synthetisiert sondern nur (leicht) modifiziert, z.B. phopsphoryliert.)

(4) Genregulatorische Netzwerke



(exprimiertes) Gen A inhibiert (Expression von) Gen B
(exprimiertes) Gen B aktiviert (Expression von) Gen C
(exprimiertes) Gen C aktiviert (Expression von) Gen A

Netzwerkmodell:

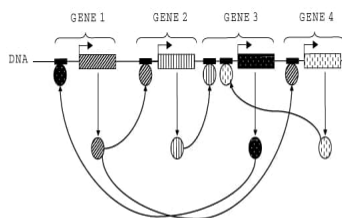
Gerichtete Vorzeichengraphen (rechts),
Boolesche / logische Netzwerke

Knoten: Gene (oder deren kodierte Proteine)

Relationen:

Typ: gerichtete Kanten mit Vorzeichen
oder logische Gatter

Bedeutung: Aktivierungs- und
Deaktivierungsprozesse



■ transcription factor binding site
in promoter region
→ coding DNA
● transcription factor

Representation of a simple, fictional transcription factor network. All genes shown encode transcription factors that control the activity of genes encoding transcription factors.
Schlitt et al. BMC Bioinformatics 2007 8(Suppl 6):S9 doi:10.1186/1471-2105-8-S6-S9

Genregulatorische Netzwerke sind ebenfalls durch **Signal- und Informationsflüsse** gekennzeichnet.

- die verschiedenen Typen von zellulären Netzwerken benötigen unterschiedliche topologische Modellierungsansätze
- auch bei identischem Modellierungsansatz (z.B. Graph) stehen hinter den Relationen oft ganz verschiedene Interaktionen
- Netzwerke sind nur im einfachsten Fall Graphen!
Graphen können nur 1:1- (aber nicht n:m-) Relationen abbilden

I.3. Ausgewählte Datenbanken für zelluläre Netzwerke

Übersicht (>250 Datenbanken!): <http://www.pathguide.org/>

Metabolische Netzwerke

- KEGG: <http://www.genome.jp/kegg/>
- MetaCyc/EcoCyc: <http://www.metacyc.org/>
- BRENDA: <http://www.brenda-enzymes.info/>
- BiGG (Metabolische Netzwerkmodelle): <http://biqq.ucsd.edu/>

Signaltransduktionsnetze:

- Wikipathways: <http://www.wikipathways.org/index.php/WikiPathways> (auch andere Netze)
- Cancer Genome Anatomy Project: https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways
- [KEGG]

Protein-Protein-Interaktionsnetzwerke

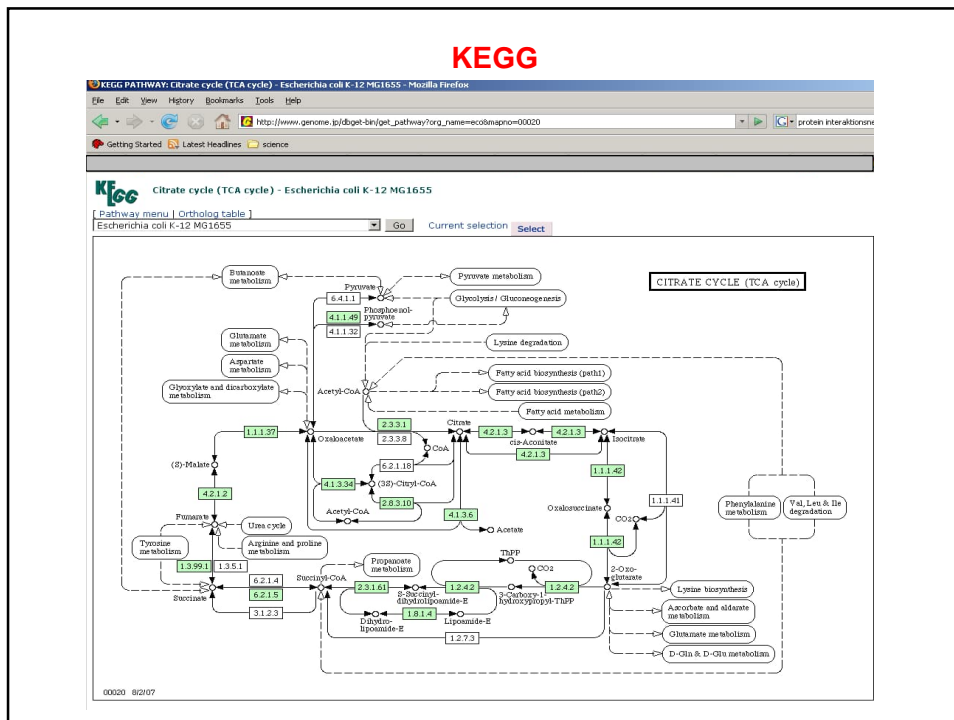
- MIPS: <http://mips.gsf.de/proi/ppi/>
- ihop: <http://www.ihop-net.org> (literature mining)

Genregulatorische Netzwerke

- Regulon DB (für E. coli): <http://regulondb.ccg.unam.mx/>
- Transfac: www.transfac.de (kommerziell)

- verschiedene Ontologien; sortierbar nach Organismen, Molekülen, Pathways etc.

KEGG



Übungen

- 1.) Teste einige der vorn angegebenen Datenbanken!
(Insbesondere: KEGG, BRENDA; STKE, BioCarta, RegulonDB, MIPS)

- 2.) Finden Sie über eine Datenbank heraus, ob das Bakterium *Rhodobacter sphaeroides* 2.4.1 das metabolische Enzyme „Pyruvatkinase“ (pyruvate kinase) besitzt; welche Reaktion(sgleichung) wird durch dieses Enzym katalysiert? In welchen Stoffwechselwegen ist es in diesem Bakterium beteiligt?

Übungen

Einführung MATLAB

MATLAB ist eine kommerzielle Software des Unternehmens The MathWorks zur Lösung mathematischer Probleme und zur grafischen Darstellung der Ergebnisse. Matlab ist primär für numerische Berechnungen mithilfe von Matrizen ausgelegt, woher sich auch der Name ableitet: MATrix LABoratory.

Beispiel für gute und kurze Einführung (oder selber im Internet suchen)
<http://www1.uni-hamburg.de/W.Wiedl/Skripte/Matlab/>

Dokumentation der MATLAB Funktionen
http://www.mathworks.com/support/functions/alpha_list.html

In MATLAB kann man Kommandos

- * einzeln (im Kommandofenster)
- * mittels Skripte („Batch“ von mehreren Kommandos)
- * oder kompakt als Funktion

starten.

- Variablen: werden deklariert bei erstem Aufruf; lokale + globale Variablen
- Skalar, Vektor, Matrix, Matrix*Vektor, „:“ Operator
- Ablaufstrukturen: for, if, while, switch, ... ; Operatoren: ==,&&,||
- wichtige eingebaute MATLAB-Funktionen (help <funcname> zeigt Hilfetext)
size, find, sum, any, disp, zeros, ...

Übungen

Beispiel für eine Funktion *inci2adja*, die eine Inzidenzmatrix eines gerichteten Graphen in eine Adjazenzmatrix umwandelt
(Syntaxregeln bzw. MATLAB Kommandos sind hervorgehoben)

```
function adja = inci2adja(inci) % Deklaration der Funktion inci2adja, welche als Input die Inzidenzmatrix
                                % bekommt und als Output die dazugehörige Adjazenzmatrix liefert
specs=size(inci,1); % Anzahl Knoten = Anzahl Zeilen in inci
edges=size(inci,2); % Anzahl Kanten = Anzahl Spalten in inci
adja=zeros(specs,specs); % Initialisiere eine Adjazenzmatrix mit Nullen
for i=1:edges % Von i=1 bis Anzahl Spalten
    s=find(inci(:,i)<0); % Finde für die i-te Kante (Spalte) den Startknoten (-1)
    z=find(inci(:,i)>0); % Finde für die i-te Kante (Spalte) den Endknoten (1)
    adja(s,z)=adja(s,z)+1; % Zähle Anzahl Kanten S->Z eins hoch
end % Ende von for
```

Schreiben Sie eine Funktion, die eine Adjazenzmatrix eines gerichteten Graphen in eine Inzidenzmatrix umwandelt! (Self-loops sollen abgefangen werden!)