

IV117: Úvod do systémové biologie

David Šafránek

24.9.2008

Obsah

Modelové organismy

Získávání biologických dat

Modely a simulace in silico

Obsah

Modelové organismy

Získávání biologických dat

Modely a simulace in silico

Escherichia Coli

- často studovaný prokaryotický organismus
- objeven Theodorem Escherichem roku 1885
 - *Escherichia Coli* (*E. coli*)
- bakterie žijící v trávicím ústrojí savců
- většina kmenů neškodných
 - součástí střevní flóry
 - zabraňuje usazení patologických bakterií
 - produkuje vitamin K_2
- některé kmeny mohou způsobit otravu
 - O157:H7, O121 a O104:H21
 - produkují toxin
 - mají dobré rezistentní schopnosti vůči lékům

E. coli jako modelový organismus

- kultivované kmeny (např. *E. coli* K12)
 - dobře adaptovatelné pro laboratorní prostředí
 - postrádají schopnost uchycení v zažívacím traktu
- detailní znalosti dostupné
- možnost experimentální manipulace v laboratořích
- experimenty s *E. coli* vedly k založení biotechnologie
 - technologie rekombinantní DNA
 - výroba rekombinantních proteinů pomocí *E. coli* (např. insulin)
- systémové studie *E. coli* zobecnitelné na složitější organismy

Shluk kolonie E. coli

Kvantitativní parametry E.coli

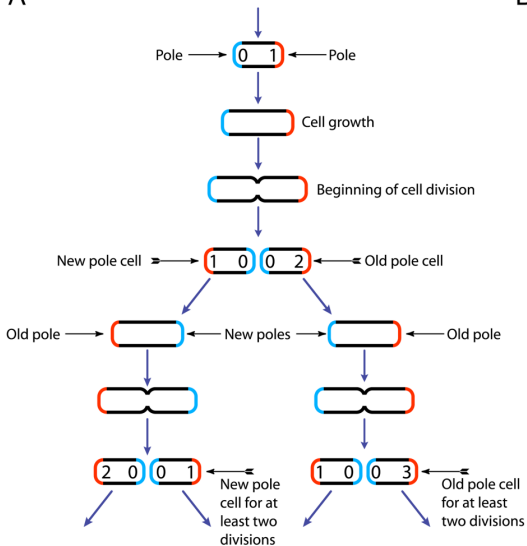
Experimentálně zjištěný parametr	Hodnota v E.Coli
Velikost buňky	$1\mu m^3$
Počet molekul proteinů v buňce	$4 \cdot 10^6$
Velikost molekuly proteinu	$5nm$
Koncentrace jednoho proteinu v buňce	$1nM$
Podíl proteinů v obsahu buňky	18%
Doba difúze proteinu v buňce	0,1s
Doba difúze ostatních molekul	1msec
Počet genů v buňce	4500

Časové dimenze v životě *E.coli*

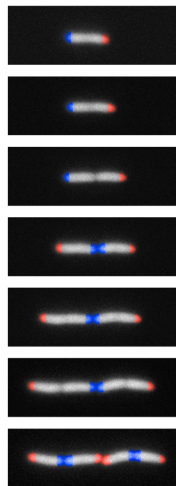
Experimentálně zjištěný parametr	E.Coli
Vazba molekuly signálu na transkripční faktor vedoucí ke změně aktivity faktoru	1msec
Vazby aktivního faktoru na operon DNA	1sec
Transkripce	1min
Translace	2min
Životnost mRNA	2-5min
Trvání jedné generace buňky	min. 30min

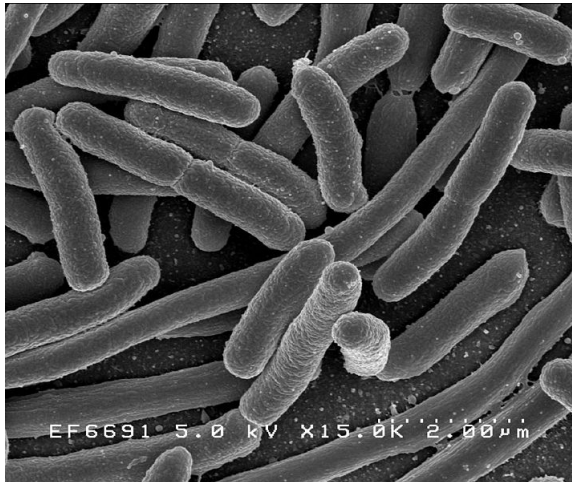
Životní cyklus *E. coli*

A

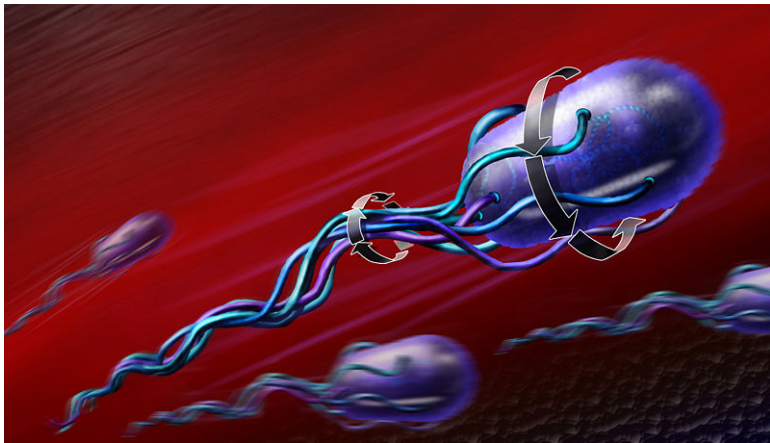


B



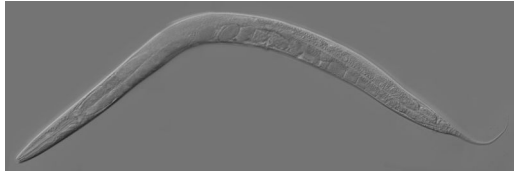
E. coli biofilm

E. coli – pohybové ústrojí (bičík)



Caenorhabditis elegans

- háďátko obecné
- volně žijící půdní helmint z kmene hlístic
- první mnohobuněčný eukaryotický organismus s plně osekvenovaným genomem (1998)
- vhodný k modelovému výzkumu
 - buněčná diferenciac
 - hermafroditní rozmnožování
 - jednoduchá nervová soustava

Caenorhabditis elegans

Caenorhabditis elegans

- genom obsahuje 97 miliónů nukleotidů
- 18841 genů kódujících různé proteiny
- geny mají v DNA *C. elegans* singulární výskyty
- většina genů podobných lidským
- organismus složen z necelého tisíce buněk
- třetina buněk tvoří nervový systém

Obsah

Modelové organismy

Získávání biologických dat

Modely a simulace in silico

Původ biologických dat

- tradiční biochemie a genetika
 - výsledky redukcionistického období biologie
 - forma: vědecké články
 - získání dat: dolováním znalostí z textů (text mining)
- low-throughput technologie
 - např. separace proteinů prostřednictvím antigenů
- high-throughput technologie
 - sekvenování genomu
 - měření exprese genů na DNA mikročipu
 - hmotnostní spektrometrie
 - proteomika a metabolomika

Databáze odborných článků

- PubMed Entrez – vyhledávání v odborných biologických/lékařských článcích

`http://www.ncbi.nlm.nih.gov/sites/entrez`

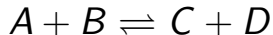
- Maintained by U.S. National Library of Medicine
- obsahuje okolo 15 mil. referencí
- zahrnuje články více než 4000 časopisů
- slovník pojmů MeSH

Ontologie genů a genových produktů

- GO – <http://www.geneontology.org/>
- ontologie představuje hierarchický pohled na data včetně vazeb mezi nimi
- GO spojuje několik databází genů a proteinů
- každý pojem jednoznačně identifikován tzv. GO-termem (kódem ve tvaru GO:xxxxxxx)
- sjednocení tří ontologií - buněčné komponenty, biologické procesy, molekulární funkce

Data (bio)chemických reakcí

- vstupy reakce – substráty (reaktanty)
- výstupy reakce – produkty

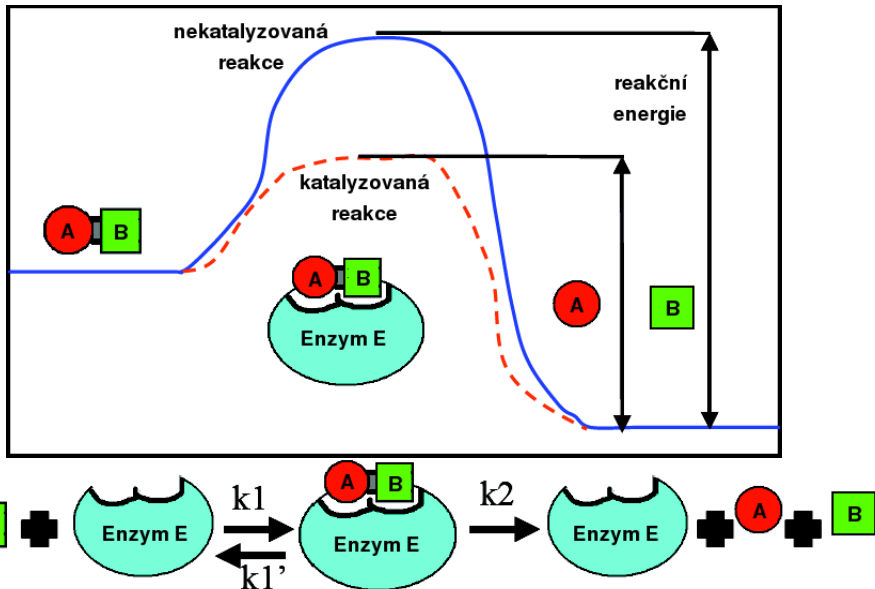


- řídicí látky (specifické proteiny — enzymy, koenzymy)
 - katalýza – urychlení/zpomalení reakce (aktivátory/inhibitory)
 - zesilování katalytických účinků enzymů koenzymy
 - např. katalytické urychlení rozpadu:



- pilířem je komplex katalytických reakcí – metabolismus
- metabolismus umožňuje zpracování a přeměnu energie

Energetický průběh reakce



Složitost metabolismu

- v e.coli je více než 1000 reakcí
- opakující se motivy
 - zpracování ATP (adenozintrifosfát) – zisk energie
 - omezené množství aktivních meziproductů
 - jen cca 100 molekul má zásadní funkci
 - omezené množství typů reakcí
- individuální reakce jsou typicky jednoduché (omezené množství substrátů/productů)

Klasifikace reakcí

- každá reakce identifikována EC číslem
- definováno dle názvosloví enzymů

<http://www.chem.qmw.ac.uk/iubmb/enzyme>

EC 3.2.1.23

typ reakce

druh substrátu

popis akceptoru

subidentifikace

Klasifikace reakcí

- celkem 8 typů reakcí (enzymů)
 - oxydoreduktázy, transferázy, hydrolázy, ...
- EC 3.2.1.23 – β -galaktocidáza
 - hydrolýza inhibující tvorbu laktózy, součást galaktózové metabolické dráhy
 - http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/enzymes/GetPage.pl?ec_number=3.2.1.23

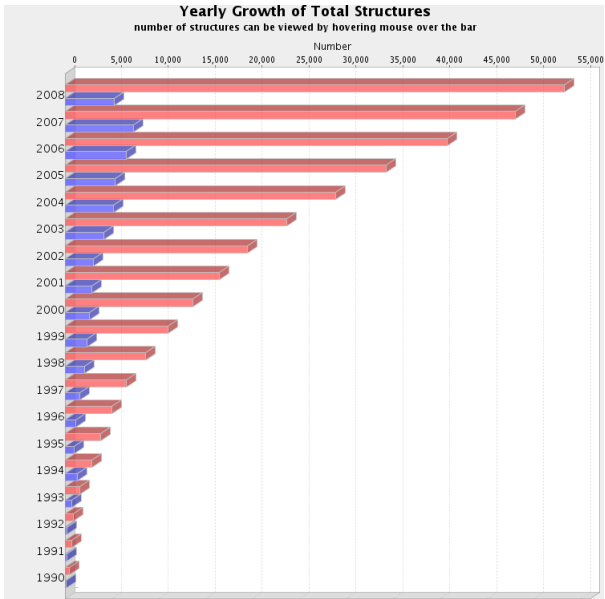
Databáze enzymů a metabolických drah

- BRENDA — <http://www.brenda.uni-koeln.de/>
- IntEnz — <http://www.ebi.ac.uk/intenz/>
- KEGG — <http://www.genome.jp/kegg/>
<http://www.genome.jp/kegg/pathway.html>
- EMBL-EBI — <http://www.ebi.ac.uk/>
- MetaCyc — <http://metacyc.org/>
- BioCyc — <http://biocyc.org>
- EcoCyc — <http://ecocyc.org> - E. coli K12

Proteinová databanka PDB

- databáze 3D strukturálních dat proteinů a nukleových kyselin
- založena 1971
- veřejný přístup
- data poskytují biologové a biochemici
- klíčový zdroj pro strukturní biologii a strukturní genomiku
- <http://www.rcsb.org/pdb/>
- PDB formát souborů –
<http://www.wwpdb.org/documentation/format30>
- primární struktura (sekvence aminokyselin) – FASTA formát
- vyhledávání proteinů dle sekvencí – BLAST –
<http://www.ebi.ac.uk/Tools/blast2/>

Proteinová databanka PDB



Obsah

Modelové organismy

Získávání biologických dat

Modely a simulace in silico

Proč dělat model in silico?

- nestačí jen znalost o chemických látkách a genech
- simulace a analýza chování při daných iniciálních podmínkách
 - numerické metody
 - diskrétní metody (analýza stavového prostoru)
- predikce chování – inspirace pro experimenty
 - tvorba hypotéz
- detailní analýza důvodů (ne)platnosti hypotézy

In silico model

- abstraktní model
- teoretický (idealizovaný) obraz skutečného organismu
- sestává z množiny proměnných a množiny logických a kvantitativních relací mezi proměnnými
- umožňuje simulovat funkční stavy organismu za určitých podmínek

Simulace modelu

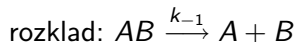
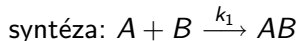
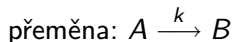
- simulace představuje “spuštění” modelu pro dané výchozí nastavení proměnných v daném prostředí
- simulace imituje skutečné chování modelovaného objektu na určité úrovni abstrakce
- simulace umožňuje predikci hypotéz
- simulace zobecňuje a doplňuje in vivo/in vitro experimenty
 - predikce má smysl pouze pro validovaný model!

Tvorba in silico modelu

- pilířem jsou biochemické reakce
- model reprezentován staticky biologickou sítí
- nezávislý na výpočetních a analyzačních nástrojích
- obecný popisovací jazyk — Systems Biology Markup Language (SBML) – <http://www.sbml.org>
- analýza spočívá v simulaci dynamiky reakcí
 - vývoj substrátů v čase
 - substrát = proměnná
- různé přístupy k modelování dynamiky, aproximace
 - spojité/diskrétní
 - deterministické/stochastické

Modely dynamiky – základní vlastnosti

- chemické reakce — syntéza a rozklad látek v čase
- **stav** — situace pozorovaná v určitém okamžiku
- kinetika — změna stavu za limitní jednotku času
 - rychlost reakce určuje konstanta (reaction rate, *mol/s*)
- ekvilibrium — rovnovážný stav
 - vyčerpání energie (zdrojů)
 - reakce vyrovnávají vzájemný účinek



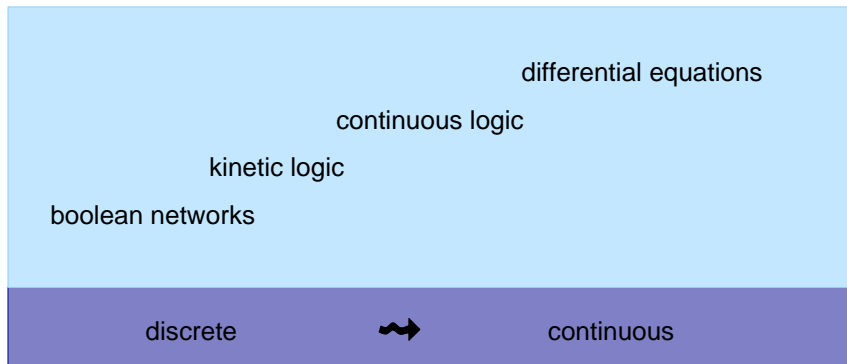
Deterministické modely dynamiky

- makropohled — předpoklad vysoké molární koncentrace látek
 - stav \equiv vektor aktuálních koncentrací látek
 - vyžadují nejvíce kvantitativních znalostí
- diskrétní modely abstrahující čas a hodnoty koncentrací
 - boolovské sítě a kinetická logika
 - stav zachycuje určitou diskrétní úroveň koncentrací
- diskrétní modely abstrahující čas a průběh interakcí
 - stav definován jako u spojitých modelů
 - dynamika řídicích interakcí diskretizována

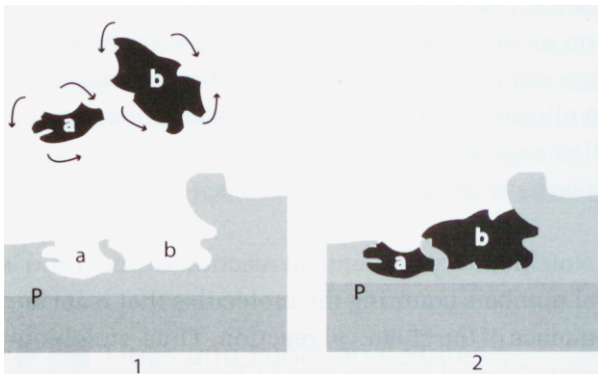
Stochastické modely dynamiky

- mikropohled — interakce individuálních molekul [Gillespie]
- stav zachycuje počty molekul jednotlivých látek
 - může být přímo diskrétní konfigurace
 - nebo distribuční funkce
- vzhledem k čase diskrétní i spojité modely
 - markovovy řetězce
 - stochastické diferenciální rovnice

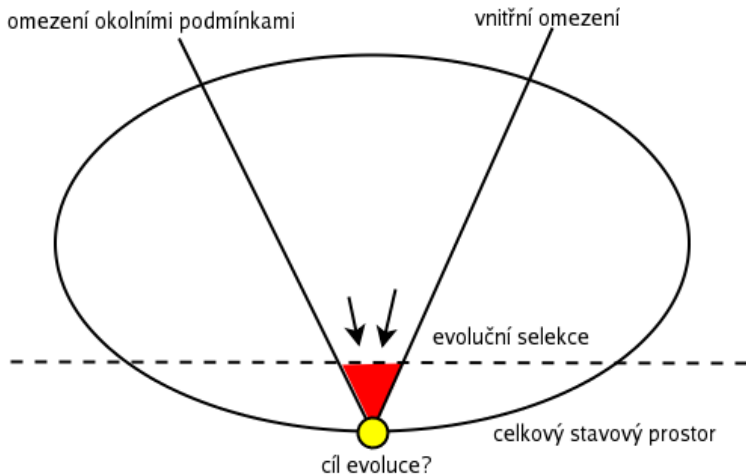
Deterministické modely dynamiky



Koncept vymezení relevantního podprostoru



Koncept vymezení relevantního podprostoru



Rizika systémové biologie

- dogmata nesmí být přeceňována
 - “le programme génétique” [Monod, Jacob]
 - genetický determinismus
 - “the selfish gene” [Dawkins]
- hypotézy musí být řádně testovány
- etická rizika

Shrnutí

- biologický systém definován interakcemi mezi jeho komponentami
- interakce jsou omezeny základními zákony chemie ale i evolučním vývojem
- syntaxí organismu-systému je síť komponent
- sémantikou organismu-systému je jeho funkce (dynamika)
- základní koncepty systémové biologie
 - modely a simulace in silico
 - důraz na interakci, součinnost
 - hierarchie (např. separace časových škál)
 - vymezení relevantního podprostoru možností

Shrnutí

