

MODEL-DRIVEN DESIGN 4 SYNTHETIC SYSTEMS BIOLOGY

22TH APRIL 2015

Monika Heiner

Brunel University, Centre for Systems and Synthetic Biology

on sabbatical leave from

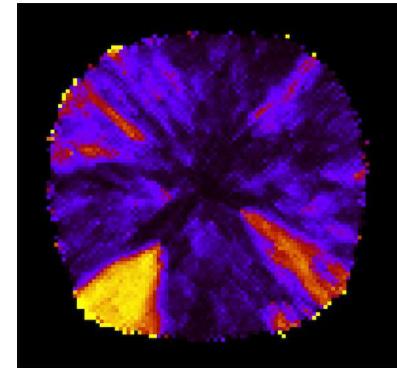
Brandenburg Technical University, Computer Science Institute

(partially supported by the SEED fund of the College of Engineering, Design and Physical Sciences)



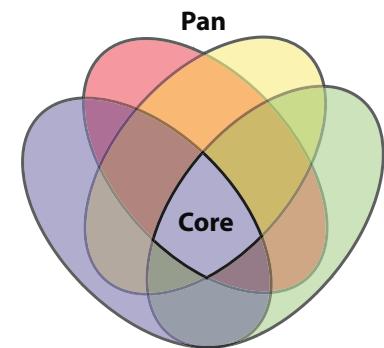
my background

-> *what I did so far*



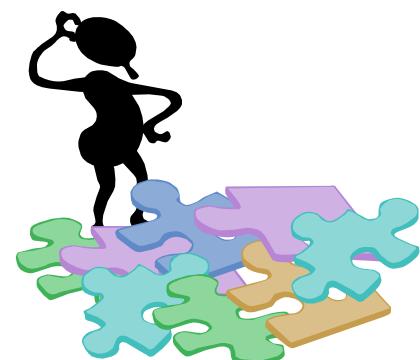
the project

-> *what I understood so far*

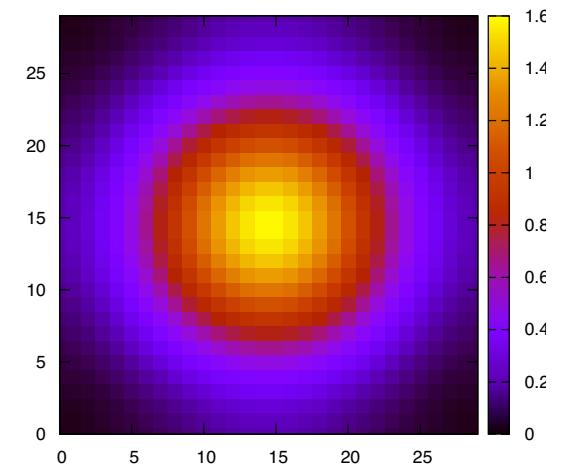
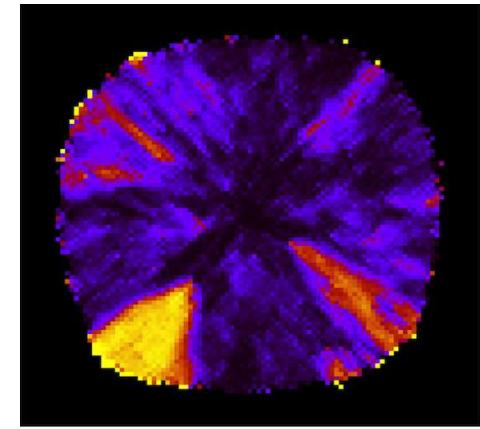
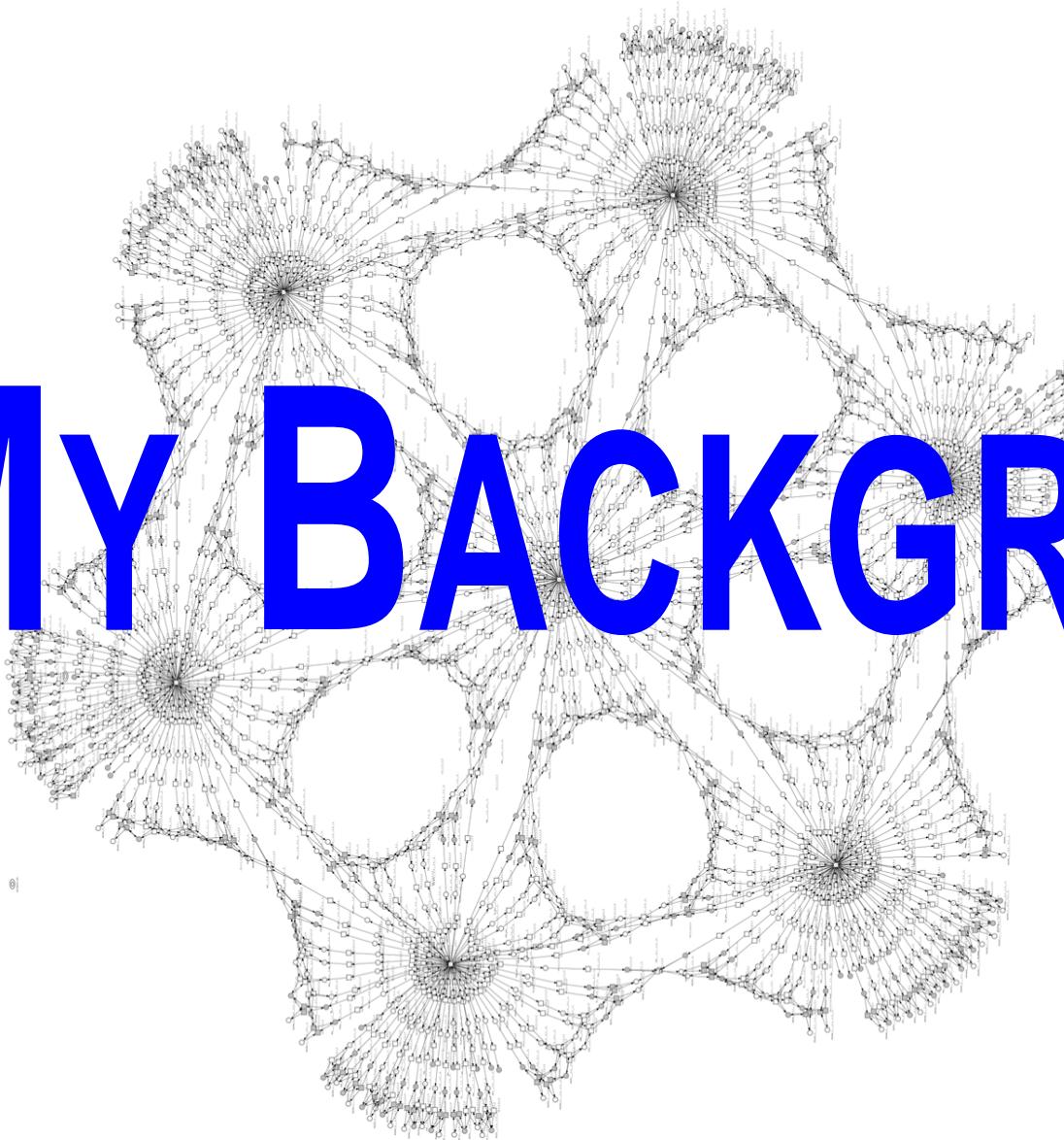


some tentative contributions

-> *what I could contribute, potentially*

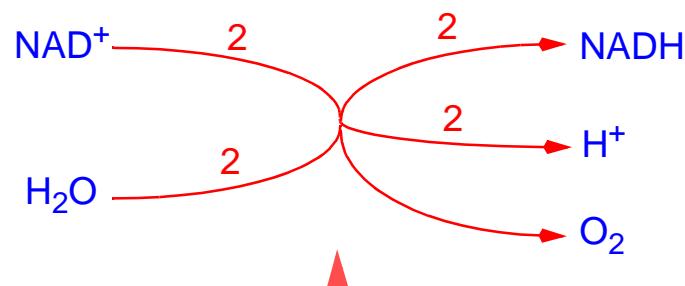
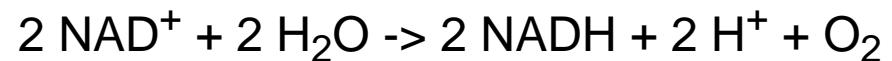


MY BACKGROUND

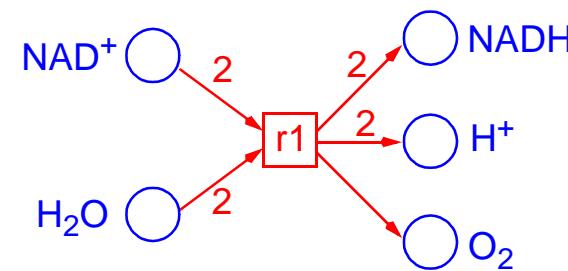


**ARE NETWORKS
OF BIOCHEMICAL
REACTIONS**

**NATURALLY
EXPRESSIBLE AS
PETRI NETS**

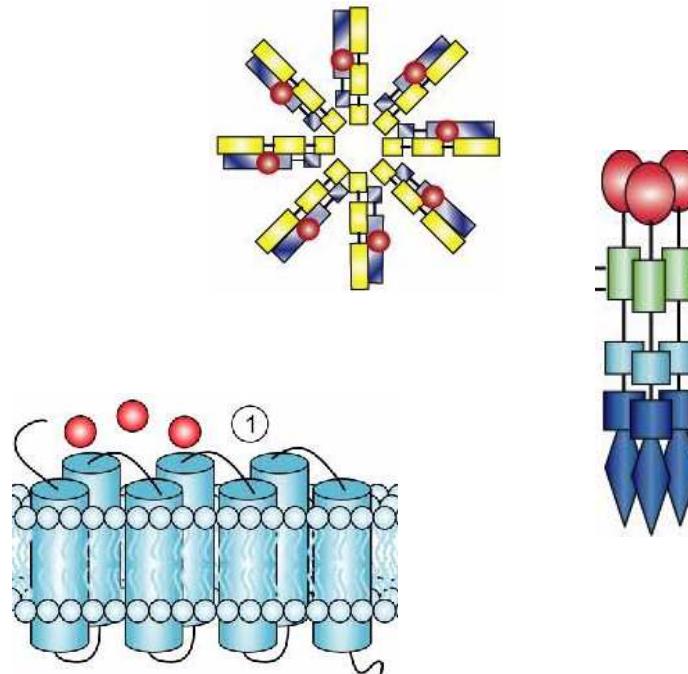


hyper-arcs



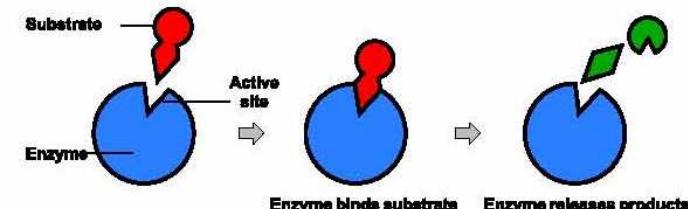
□ places → model variables

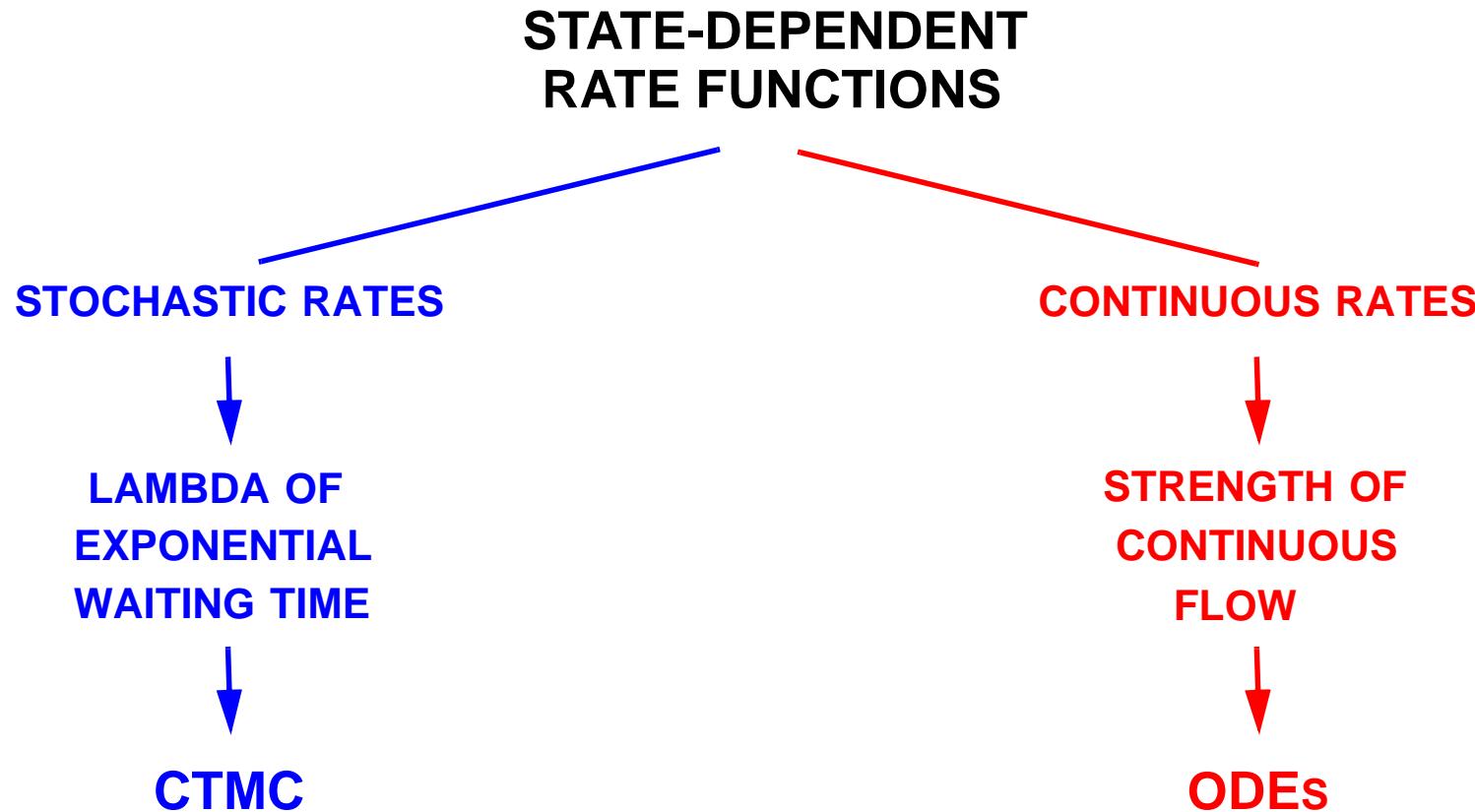
- > (bio-) chemical compounds
- > proteins
- > protein conformations
- > complexes
- > genes, . . . , etc.
- ... in different locations*



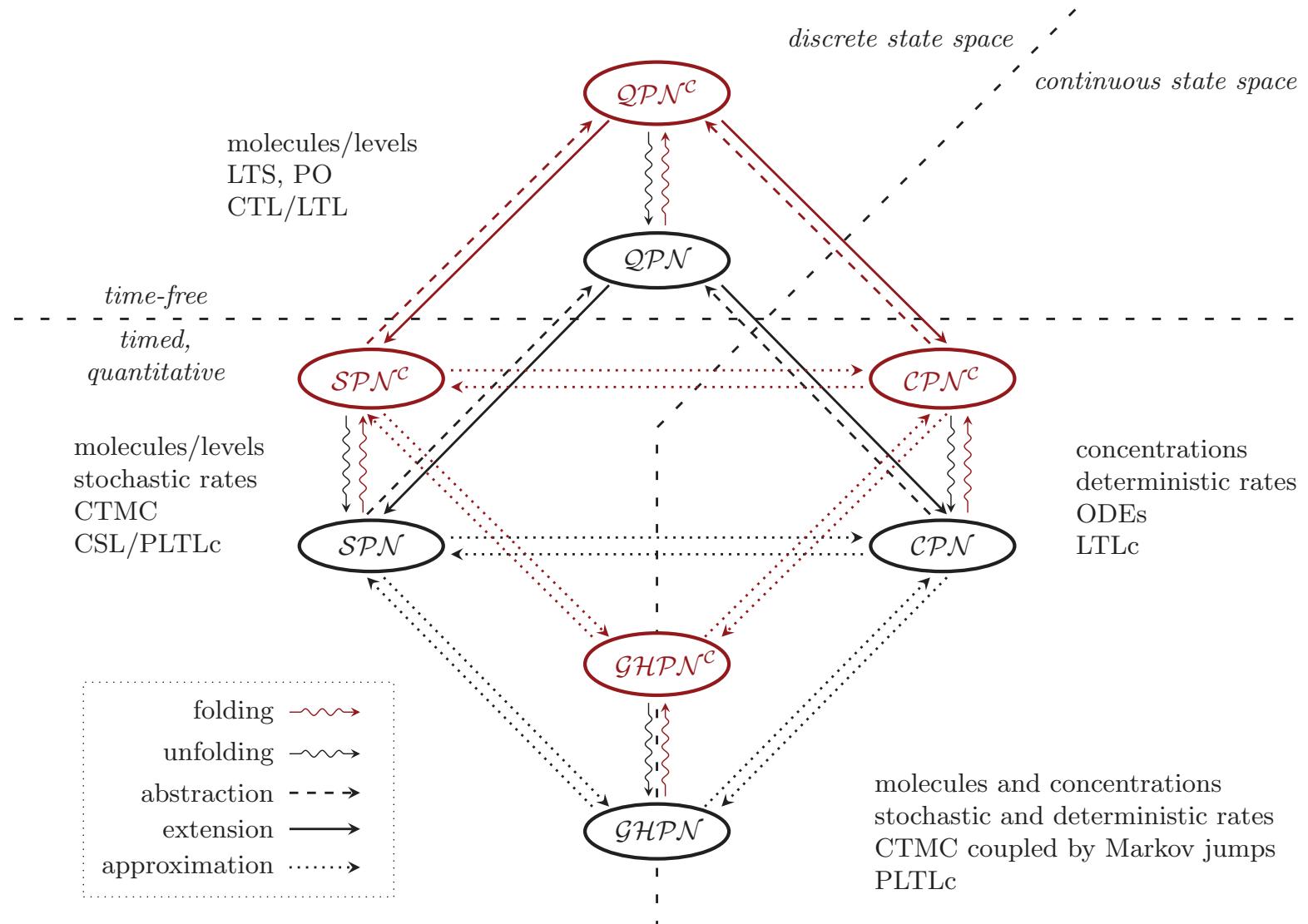
□ transitions → atomic events

- > (stoichiometric) chemical reaction
- > complexation / decomplexation
- > phosphorylation / dephosphorylation
- > conformational change
- > transport step, . . . , etc.
- ... in different locations*



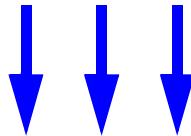


-> supported by, e.g., COPASI, Dizzy, ..., Snoopy



4X2

MODELS SHARING STRUCTURE



QUANTITATIVE MODEL = QUALITATIVE MODEL

+

**RATE FUNCTIONS
(KINETICS)**

□ SNOOPY

- > *modelling and animation/simulation of hierarchical graphs,*
e.g. various Petri net classes, e.g. PN, XPN, SPN, XSPN, CPN, ...

□ S4

- > *standalone, computational steering server*

□ CHARLIE

- > PN, XPN, Time/Timed Petri nets (TPN)
- > mostly standard analysis techniques of Petri net theory

□ MARCIE

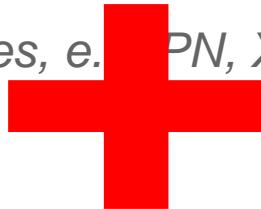
- > PN, XPN, SPN, XSPN, SRN
- > *symbolic and simulative model checking*

□ Patty

- > *animation via web browser*

SNOOPY

- > *modelling and animation/simulation of hierarchical graphs,
e.g. various Petri net classes, e.g. PN, XPN, SPN, XSPN, CPN, ...*



S4

- > *standalone, computational steering server*

CHAKRA

- > *PN, XPN, Timed Petri nets (TPN),*

- > *mostly standard analysis techniques of Petri net theory*

MARCIE

- > *PN, XPN, SPN, XSPN, SRM*

- > *symbolic and simulative model checking*

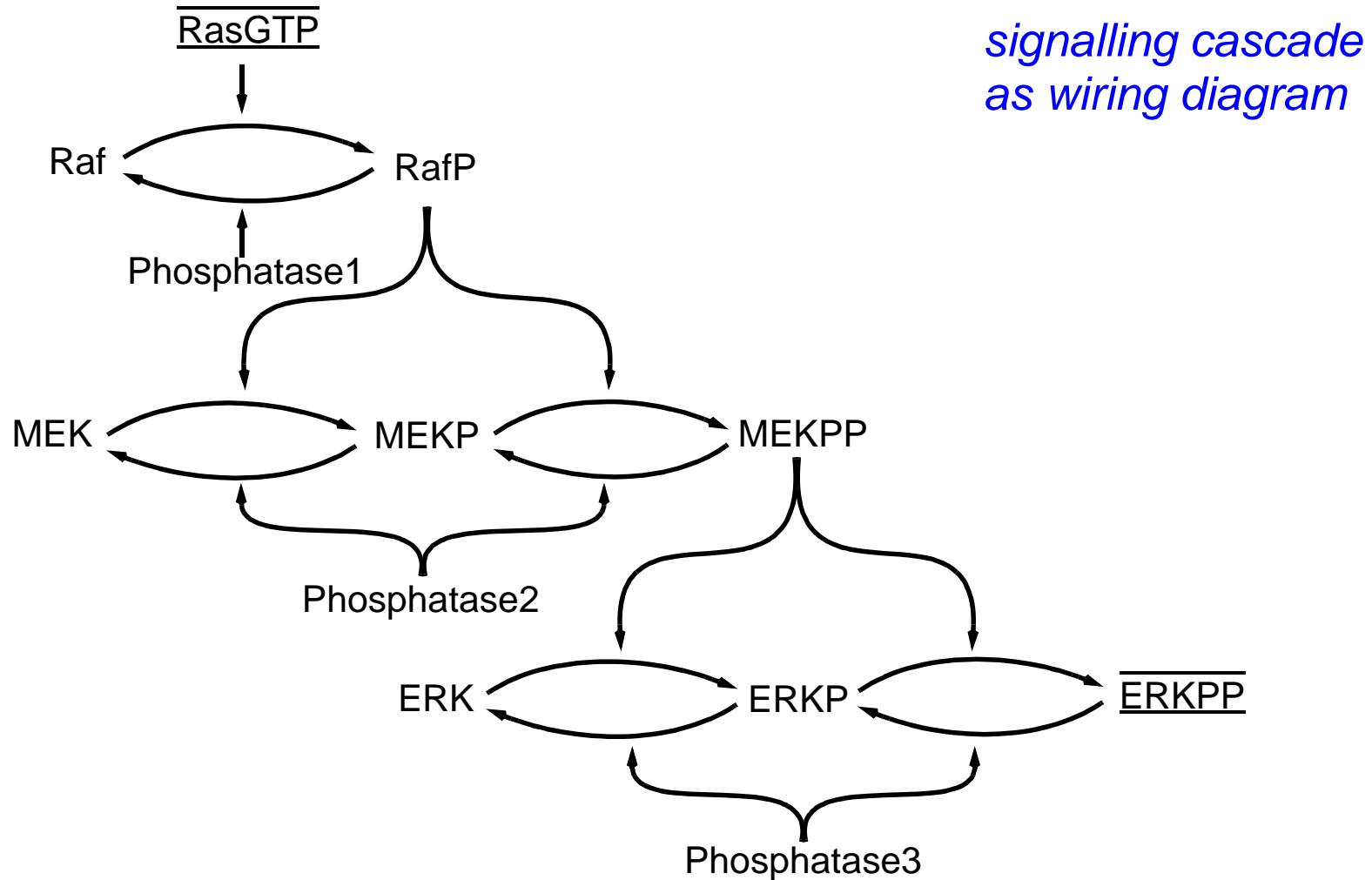
Patti

- > *animation via web browser*

SBML import/export

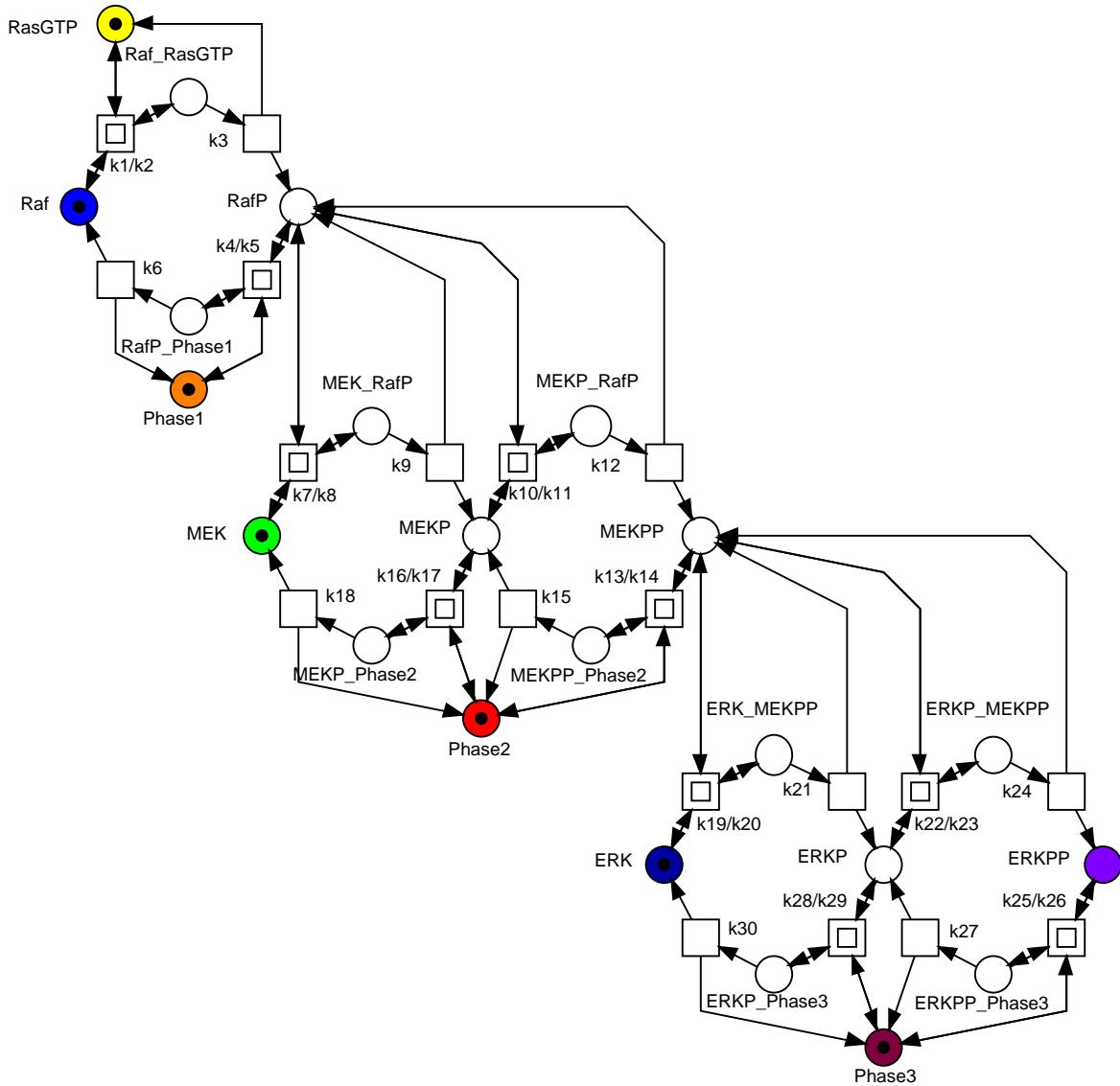
EXPORT TO MATLAB AND

MANY OTHER TOOLS



CASE STUDIES, Ex1 - SIGNALLING PATHWAYS

PN & Bacterial BioEngineering



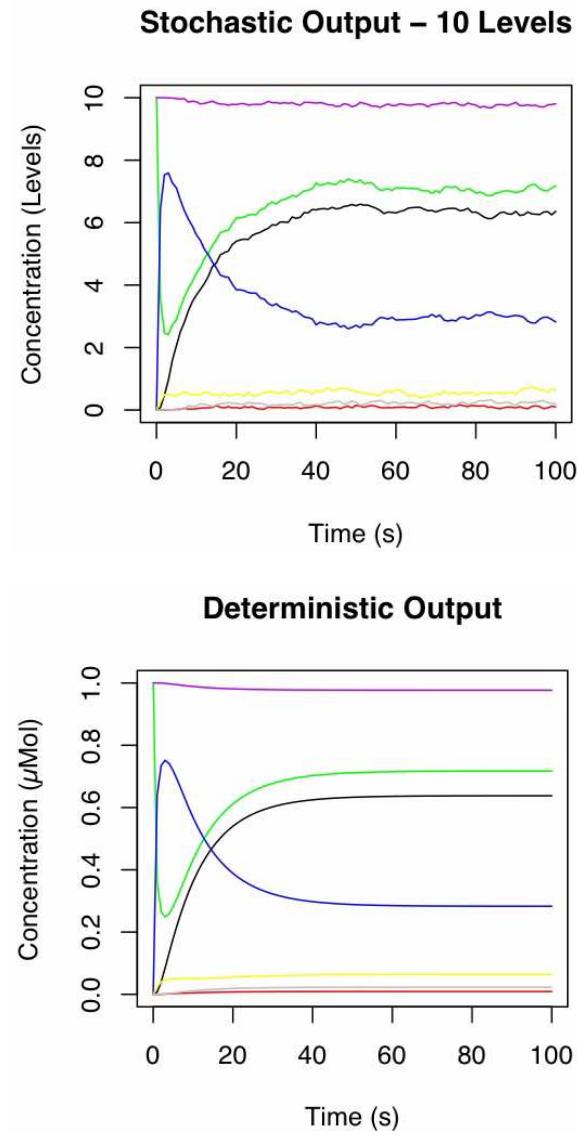
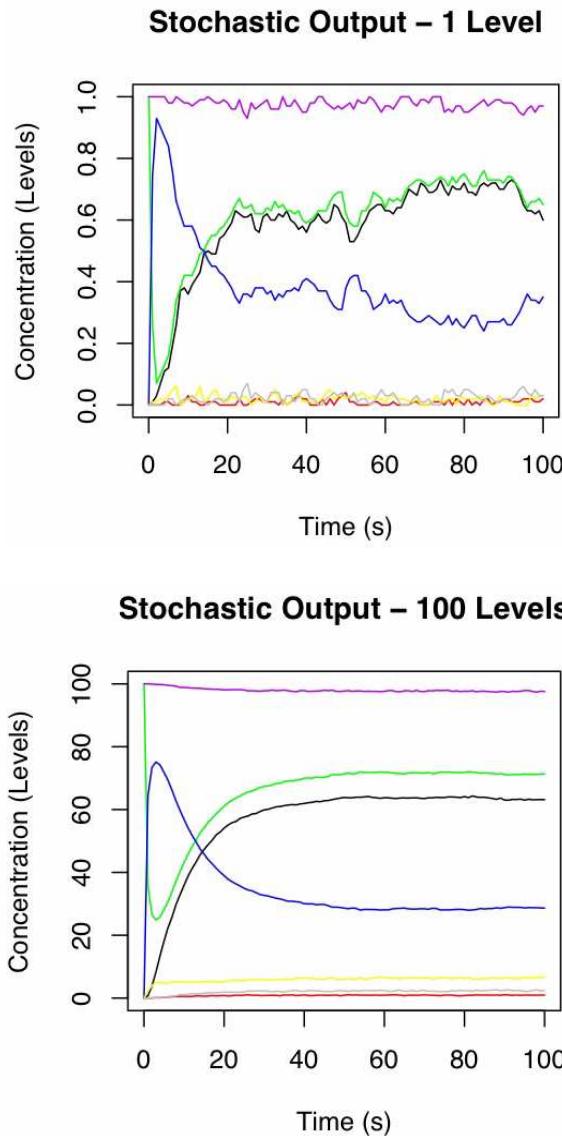
*signalling cascade
as wiring diagram*

[GILBERT,
HEINER,
LEHRACK 2007]

[HEINER,
GILBERT,
DONALDSON 2008]

CASE STUDIES, Ex1 - SIGNALLING PATHWAYS

PN & Bacterial BioEngineering



*signalling cascade
as wiring diagram*

[GILBERT,
HEINER,
LEHRACK 2007]

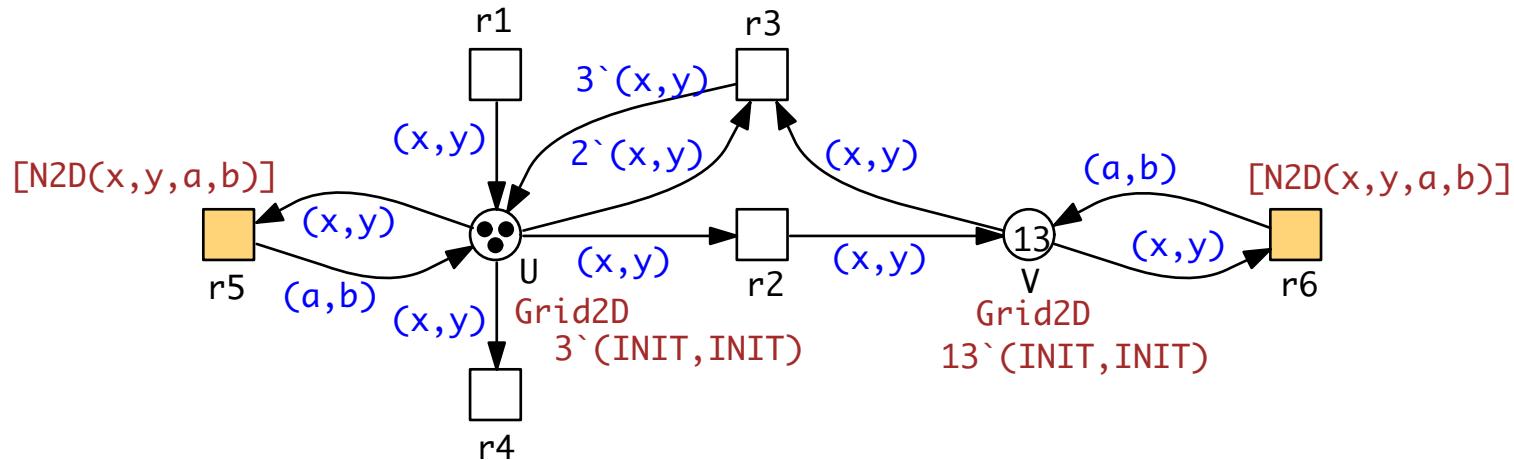
[HEINER,
GILBERT,
DONALDSON 2008]

CASE STUDIES, Ex2: TURING PATTERNS

PN & Bacterial BioEngineering

$r1: \quad \text{-} \rightarrow U$
 $r2: \quad U \rightarrow V$
 $r3: \quad 2U + V \rightarrow 3U$
 $r4: \quad U \rightarrow$

diffusion:
 $r5: \quad U_{xy} \xrightarrow{1/h^2} U_{ab}$
 $r6: \quad V_{xy} \xrightarrow{D/h^2} V_{ab}$



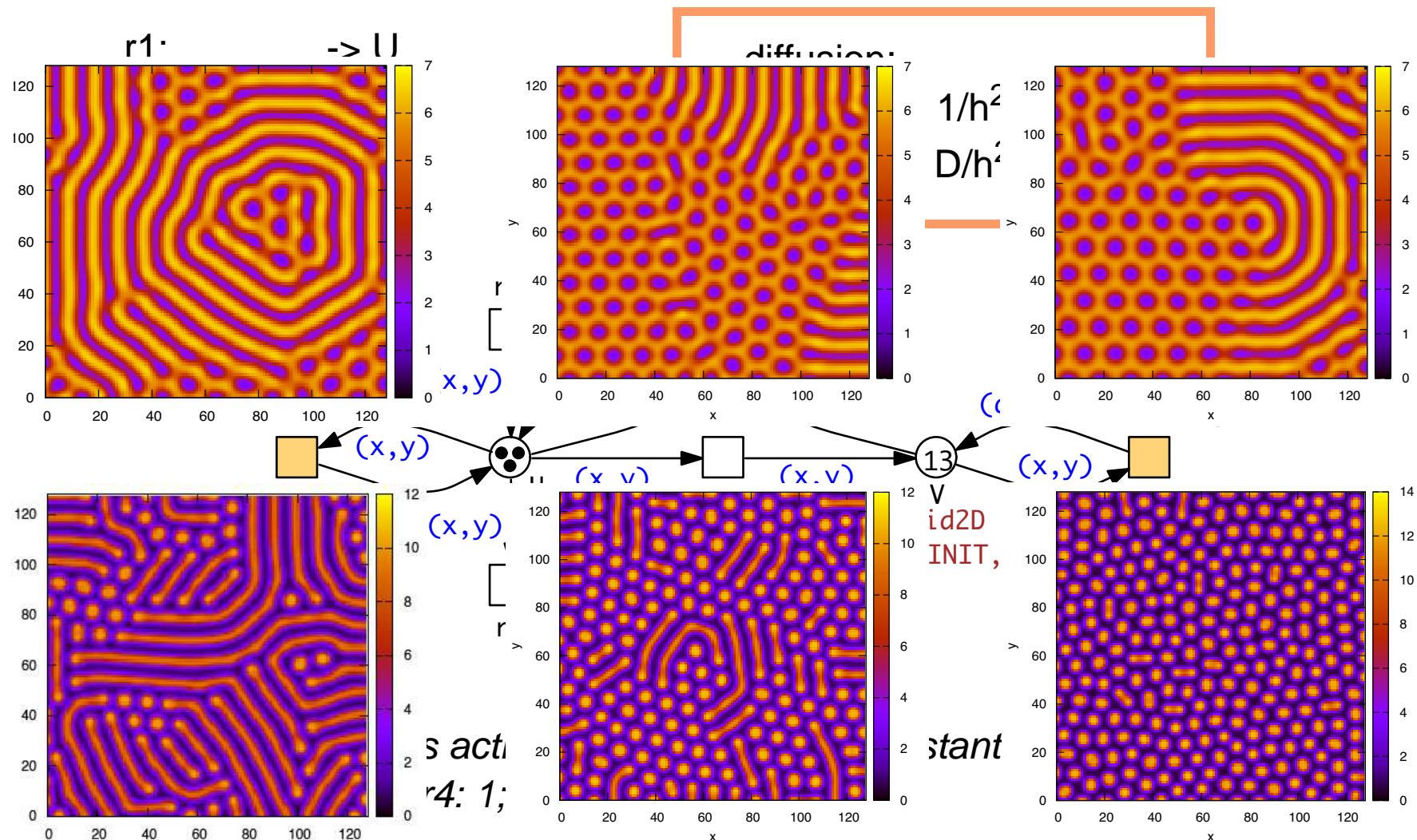
$r1 - r4$ follow mass action kinetics with rate constants:

$r1: a, r2: b, r3: 1, r4: 1;$

[LIU ET AL 2014]

CASE STUDIES, Ex2: TURING PATTERNS

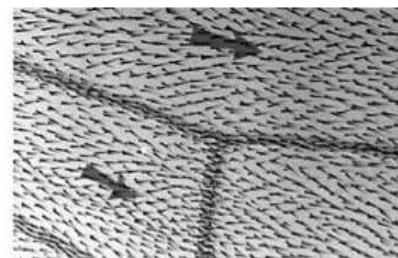
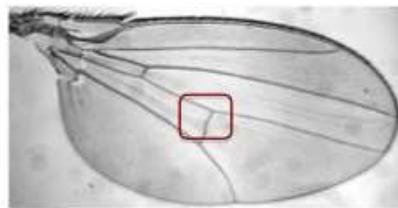
PN & Bacterial BioEngineering



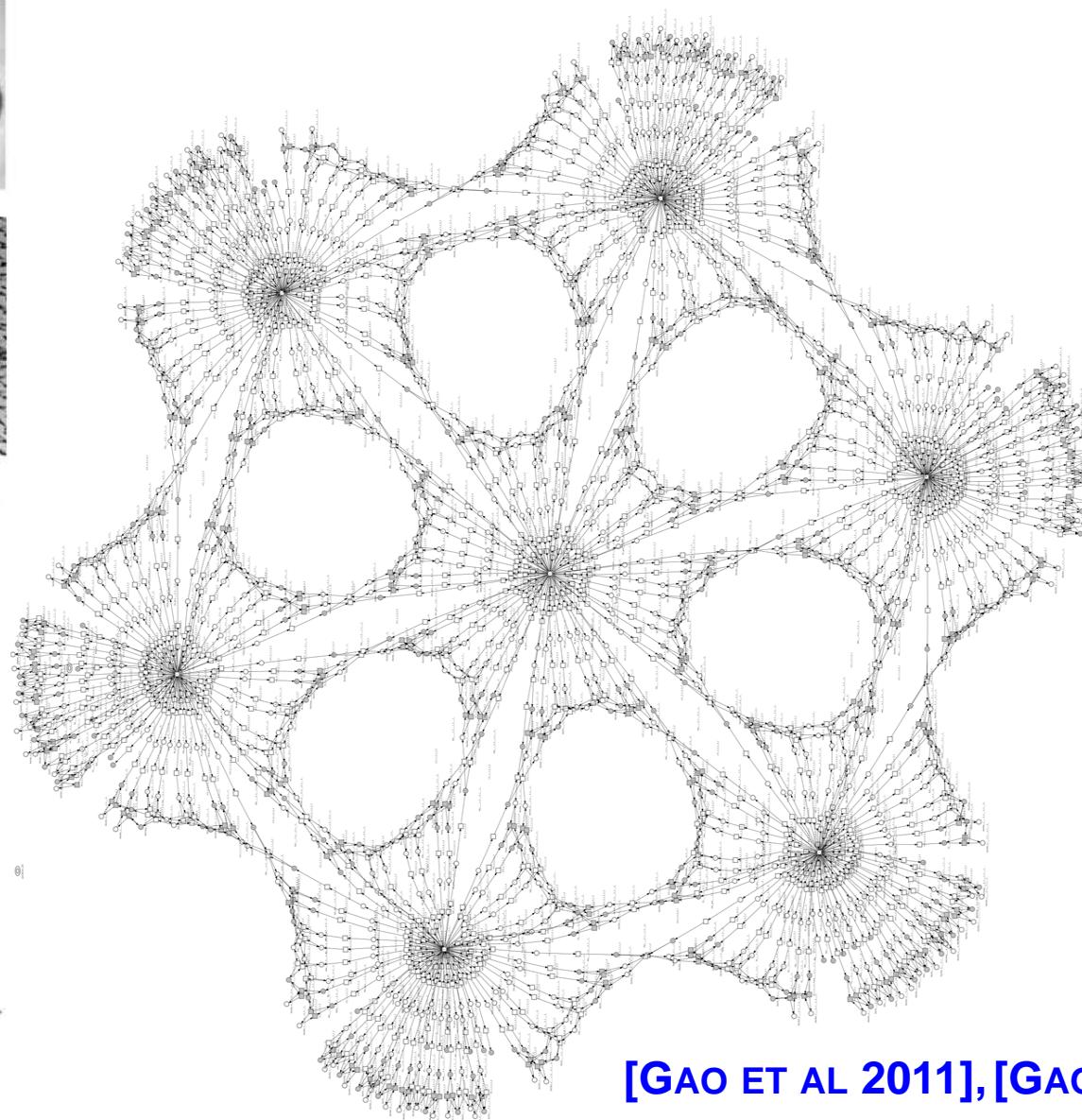
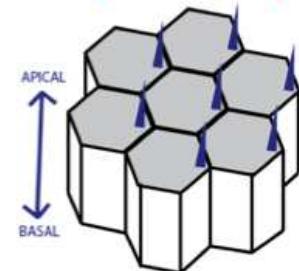
[LIU ET AL 2014]

CASE STUDIES, Ex3 - PLANAR CELL POLARITY

PN & Bacterial BioEngineering



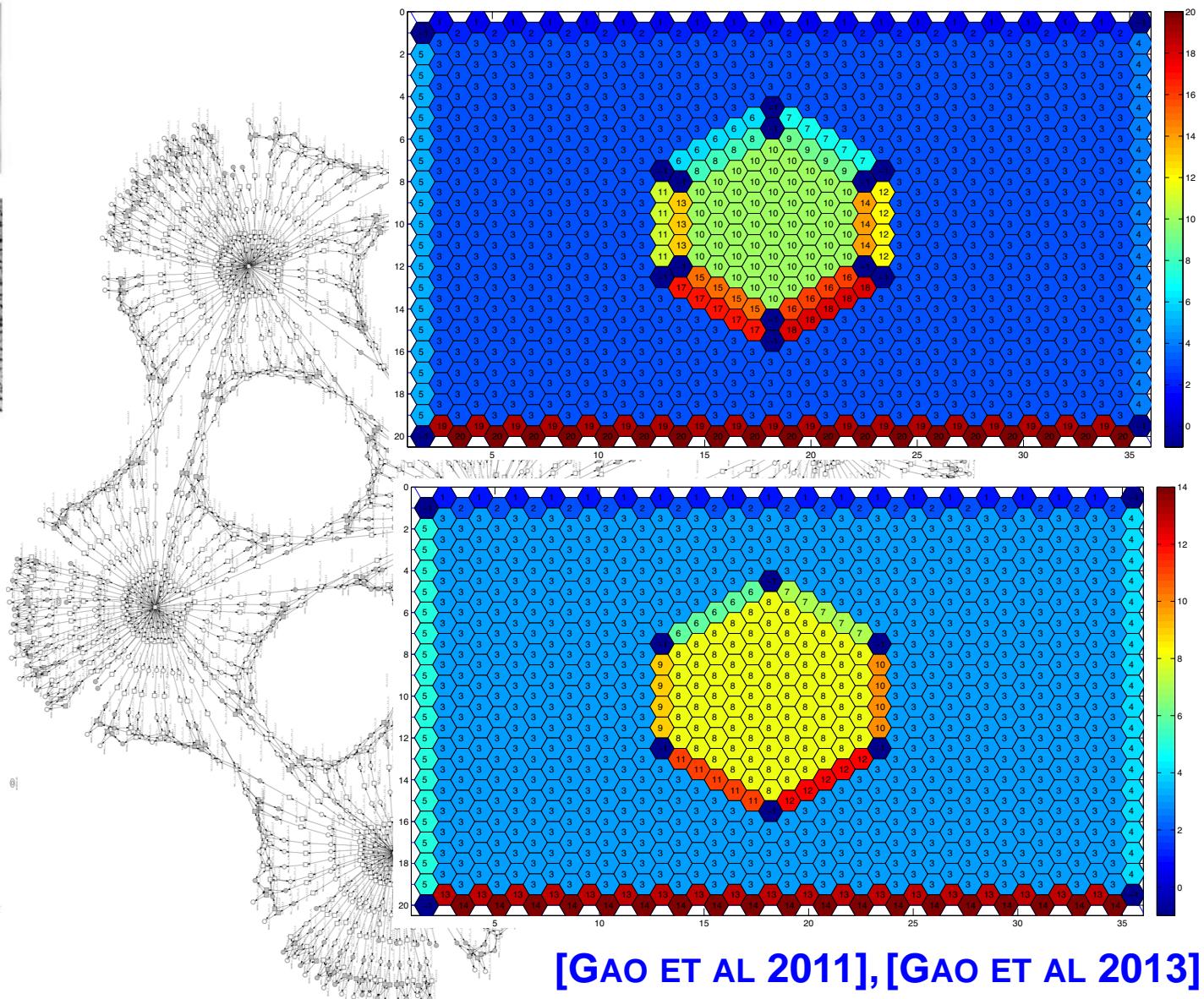
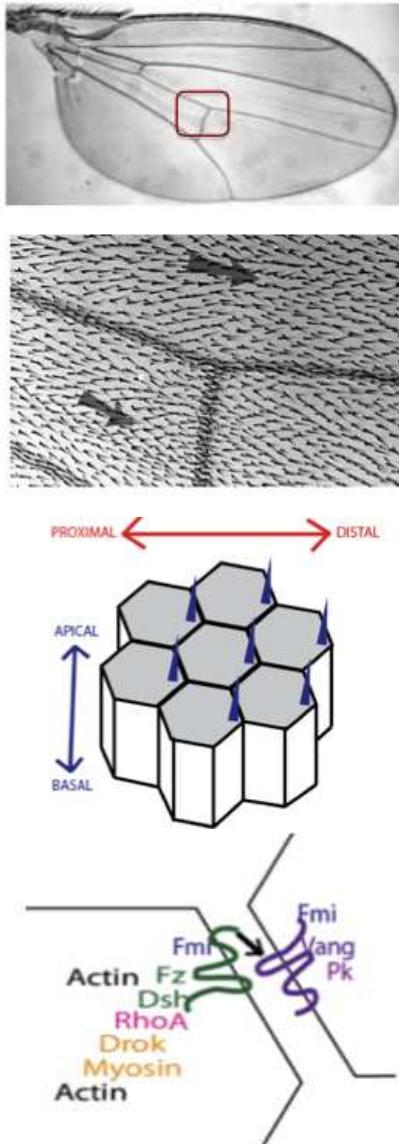
PROXIMAL ← → DISTAL



[GAO ET AL 2011], [GAO ET AL 2013]

CASE STUDIES, Ex3 - PLANAR CELL POLARITY

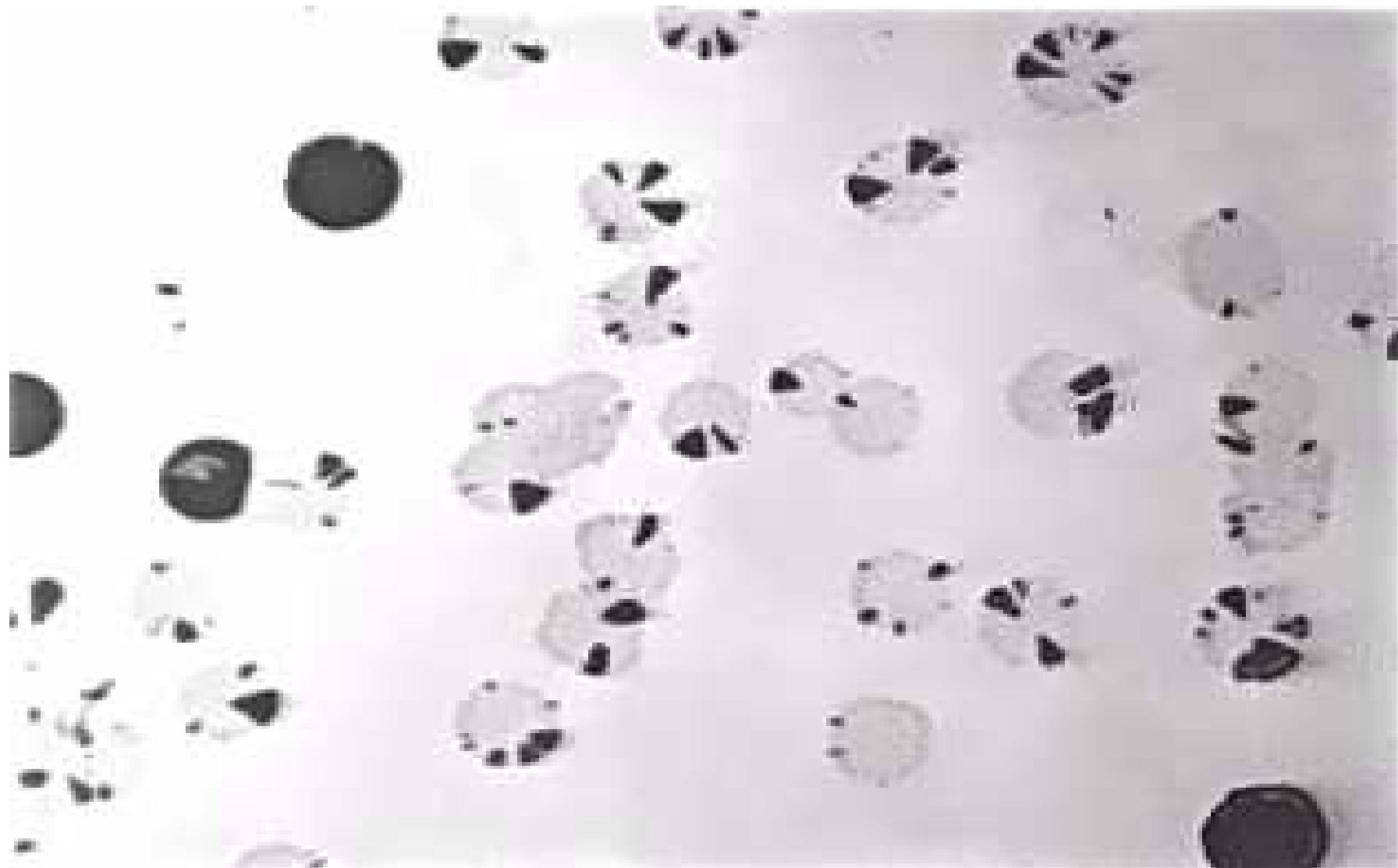
PN & Bacterial BioEngineering



[GAO ET AL 2011], [GAO ET AL 2013]

CASE STUDIES, Ex4 - PHASE VARIATION

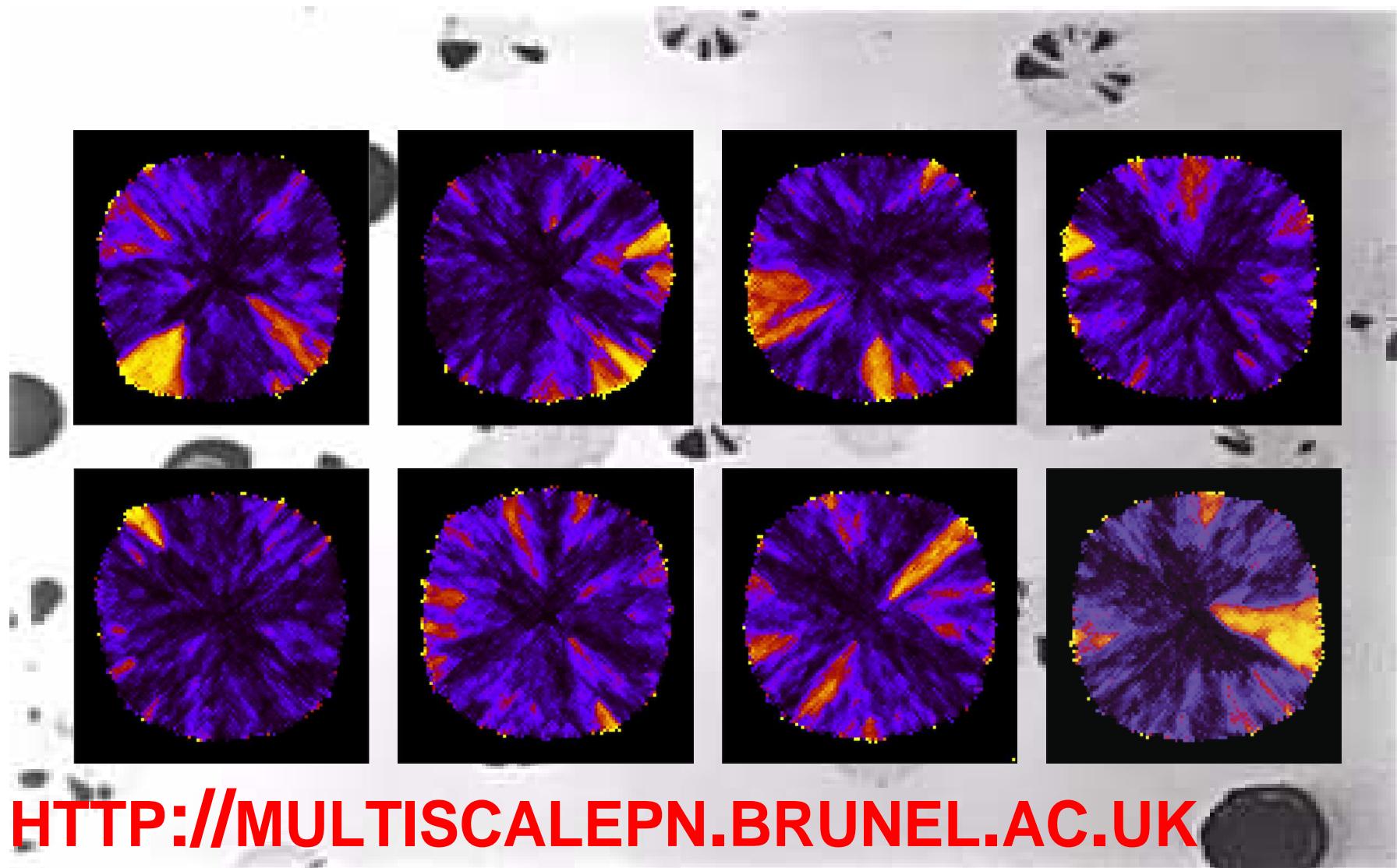
PN & Bacterial BioEngineering



[GILBERT ET AL 2013], [PARVU ET AL 2013], [PARVU ET AL 2015]

CASE STUDIES, Ex4 - PHASE VARIATION

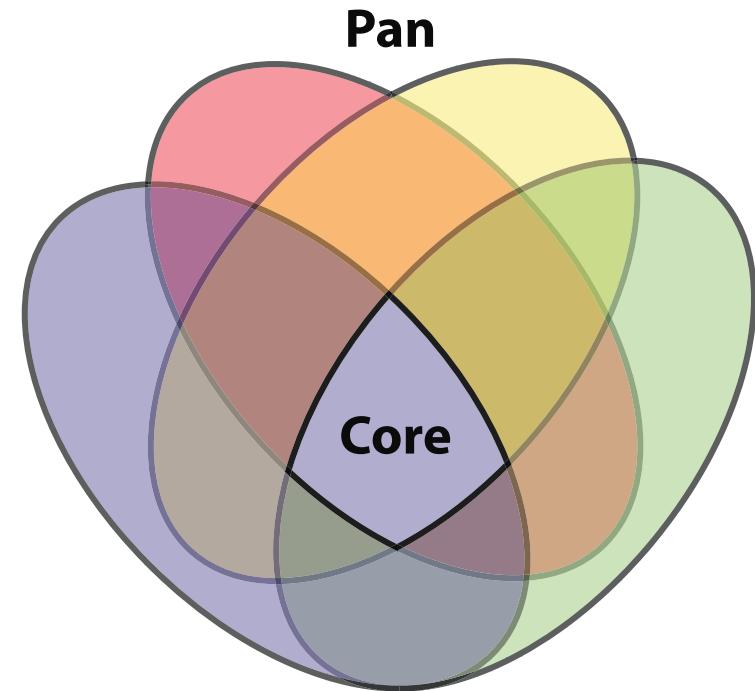
PN & Bacterial BioEngineering



HTTP://MULTISCALEPN.BRUNEL.AC.UK

[GILBERT ET AL 2013], [PARVU ET AL 2013], [PARVU ET AL 2015]

THE PROJECT



CENTRE FOR SYSTEMS AND SYNTHETIC BIOLOGY
DAVID GILBERT, NIGEL SAUNDERS

□ **synthetic systems biology**

- > *knowledge-based design*
- > *creation of new living systems (in our case: bacteria)*
- > *with novel exploitable/translational applications*

□ **bacteria can be engineered to act as little factories for**

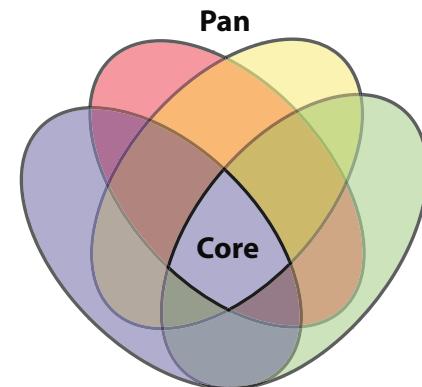
- > *energy production*
- > *drug production*
- > *immune system booster (probiotics)*
- > *pollution clean up, and*
- > *environmental sensors*

□ **bacterial engineering**

- > *genetic engineering*
- > *target-driven genome modification*
- > *metabolic engineering*
- > *focus on metabolism*

□ **model organisms**

- 1885 discovered by German-Austrian pediatrician Theodor Escherich
 - > named after him in 1919
- can be grown and cultured easily and inexpensively in a laboratory setting
 - > under favourable conditions, it takes only 20 minutes to reproduce
- the most widely studied prokaryotic (single-cell) model organism
 - > EcoliWiki
- one of the most diverse bacterial species
 - > only 20% of the genome is common to all strains
- strain
 - > a species' subgroup with unique characteristics that distinguish it from other strains
 - > EcoCyc: scientific database for the bacterium *Escherichia coli K-12 MG1655*



- 1997, first complete DNA sequence

- today

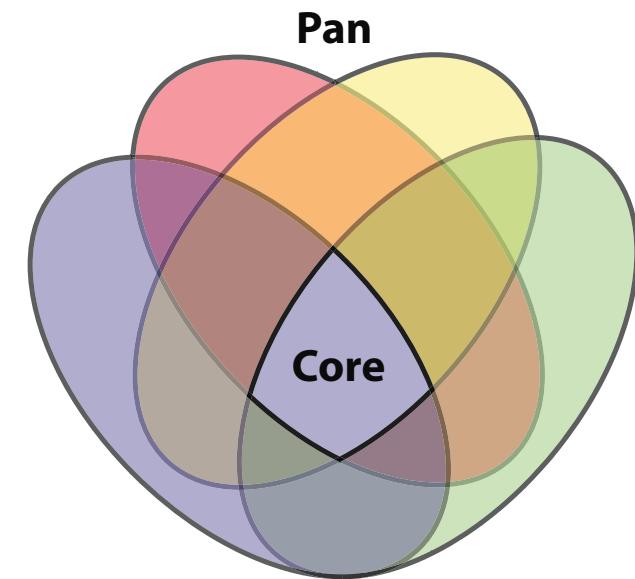
- > several hundred 'complete' genomic sequences
- > each individual genome: 4,000 - 5,500 genes
(protein genes, RNA genes)
- > *How many protein genes control metabolism?*

- core genome: 800 - 1,100 genes

- > the genome common to all strains
- > compare: genome of all humans differ by about 1%

- pangenome: exceeds 16,000 genes

- > total number of different genes among all of the sequenced *E. coli* strains
- > possible explanation: horizontal gene transfer



- 1997, first complete DNA sequence

- today
 - > several hundred 'complete' genomic sequences
 - > each individual genome: 4,000 - 5,500 genes (protein genes, RNA genes)
 - > How many protein genes control metabolism?

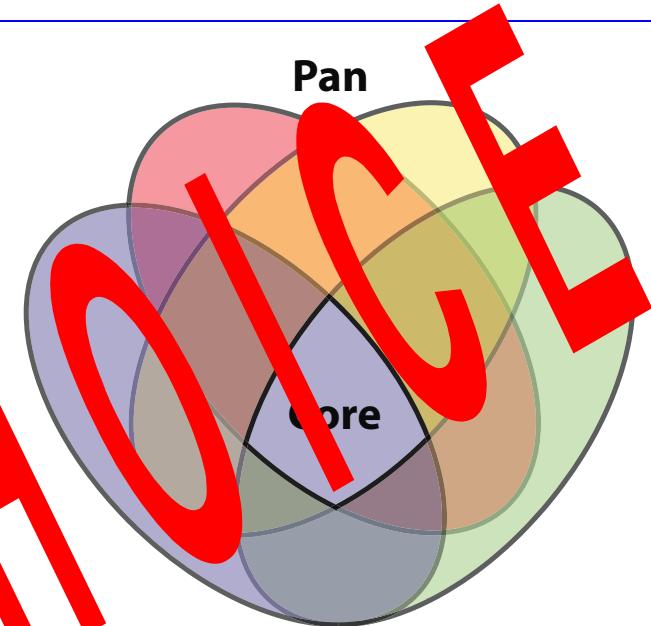
- core genome: 800 - 1,100 genes

- > the genome common to all strains
 - > compare: genome of all humans differ by about 1%

- pangenome: exceeds 16,000 genes

- > total number of different genes among all of the sequenced *E. coli* strains
 - > possible explanation: horizontal gene transfer

SPOL



- research subject since about 15 years
- basically two categories of models and related analysis techniques
 - > static (structural) modelling (no kinetic information required)
 - > dynamic (kinetic) modelling -> computational models
- the fast majority of research follows first approach, related analysis techniques
- standard graph algorithms
 - > eg., structural (linear) paths going from input A to output B, while avoiding or passing specified intermediates C
- linear programming techniques + steady state assumption
 - > all minimal flows (EM - elementary modes, T-invariants, ...)
 - > flux balance analysis = „all minimal flows + target function“
 - > optimum flux distributions

□ identification

- > *gene candidates for overexpression,
if they control the production pathway*
- > *gene candidates for downregulation or knockout,
if they control competing pathways and don't contradict the cell's viability*
- > *minimal cut sets to find minimal deletion strategies*

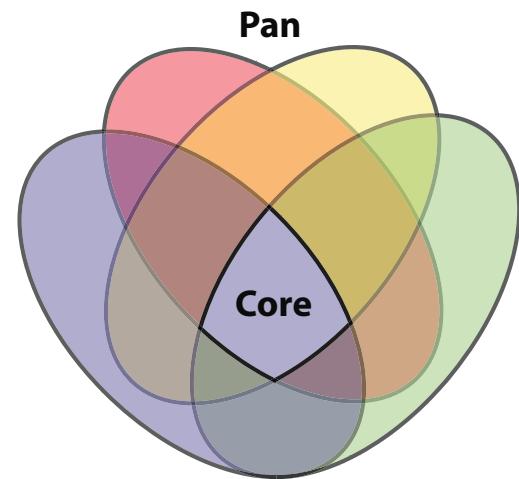
□ CONS

- > *the core metabolism of E. coli (about 100 reactions)*
 - *around 272 million EMs*
 - *requires a machine with at least 153 GB of RAM [Zanghellini 2013]*
- > *99% of the EMs in the core model are biologically infeasible [Zanghellini 2013]*
 - *enhance metabolic models by regulatory networks, thermodynamical and/or biological data to constrain the available solution space*

-> engineering of single strains

- **How to generate strain-specific models ?**

- > *computational metabolic models*
- > *estimated model size:*
1,800 reactions, 1,700 metabolites [Monk 2013]



- **How to (semi-automatically) generate models for new strains?**

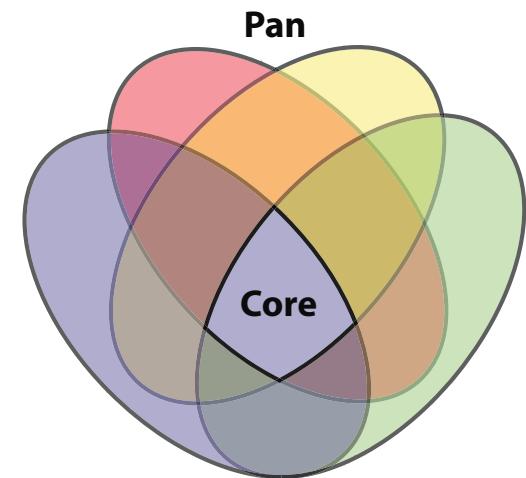
- **How to deal with sets of models ?**

- > *to rank according to target behaviour observed in the wetlab*
- > *to identify genes crucial for different performance*

- **How to select ?**

- > *chassis strain: target of gene transfer*
- > *donor strains: source of gene transfer*

- How to generate strain-specific models ?
 - > computational metabolic models
 - > estimated model size:
1,800 reactions, 1,700 metabolites [Monk 2013]
- How to (semi-automatically) generate models for new strains?
- How to deal with sets of models ?
 - > to refine according to target behaviour observed in the wet lab
 - > to identify genes crucial for different performance



COMPUTATIONAL TOOLS FOR MODEL-DRIVEN DESIGN 4 SYNTHETIC SYSTEMS BIOLOGY

SOME TENTATIVE CONTRIBUTIONS



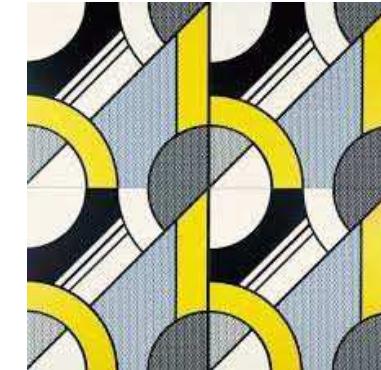
SOME TENTATIVE CONTRIBUTIONS

Modular modelling

- > *layout options*
 - > *structural protein knowledge*

□ Atomic fragments

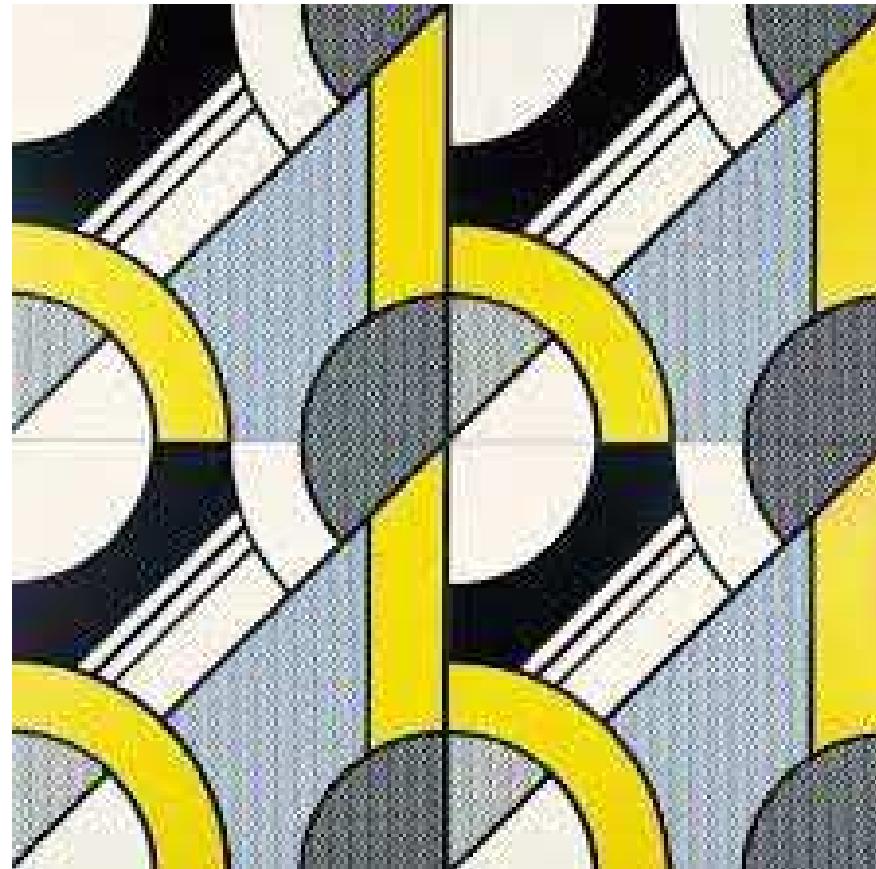
- > static (constraint-based) analysis
 - > T -invariants / elementary flax modes



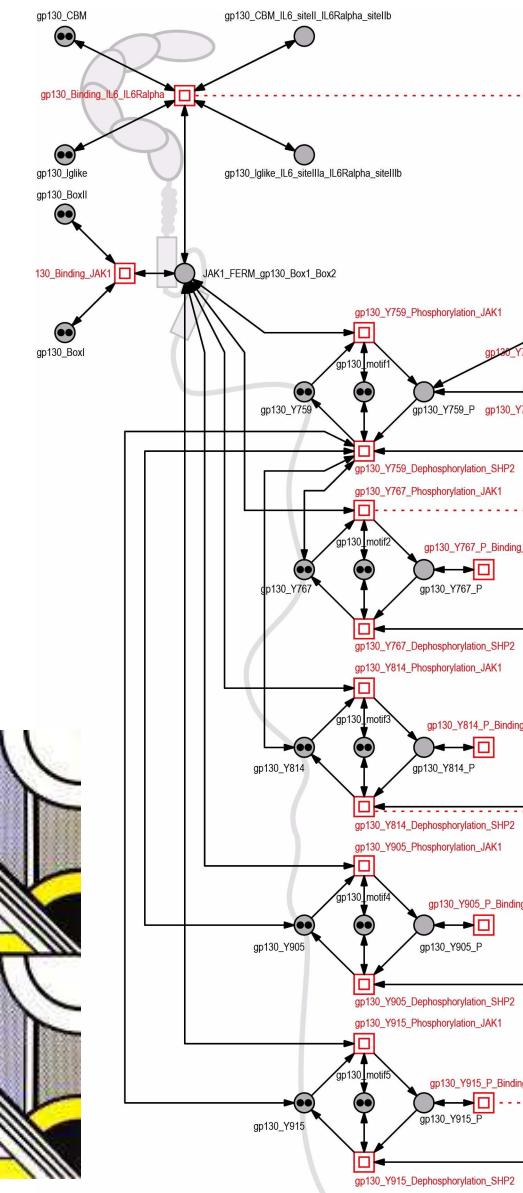
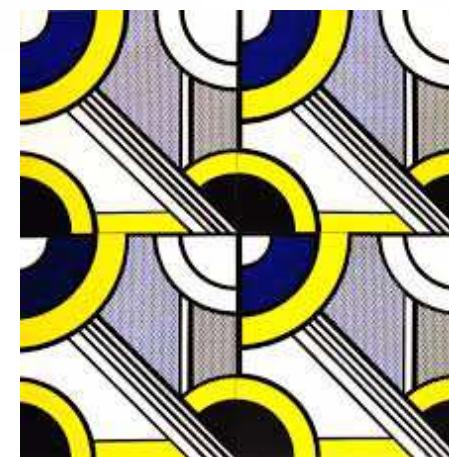
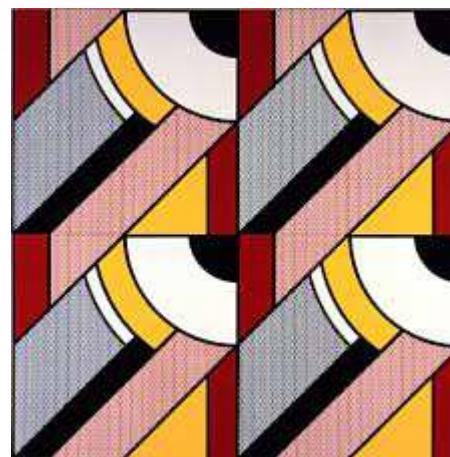
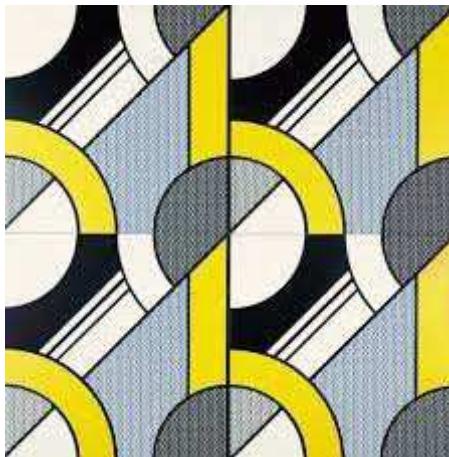
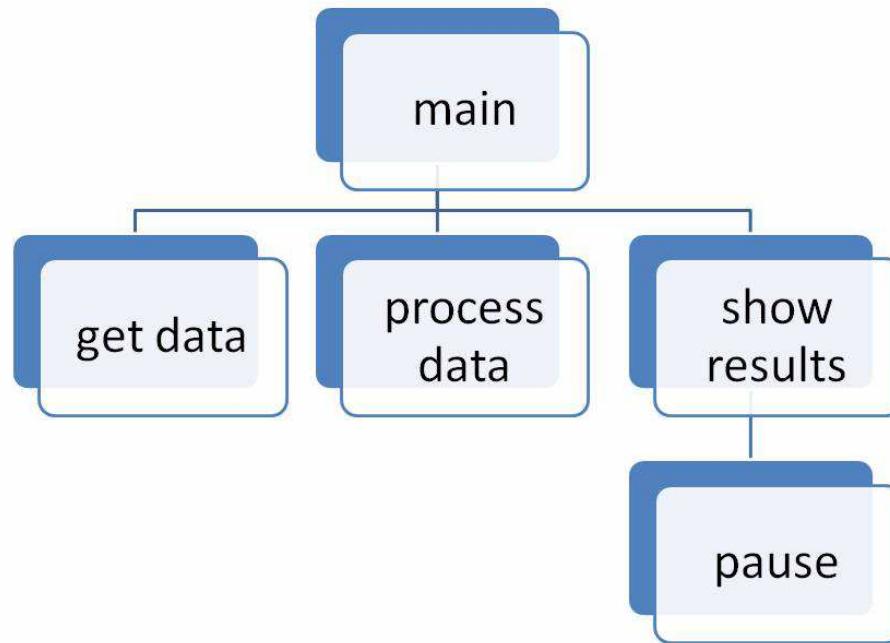
□ Parameter-free Simulation

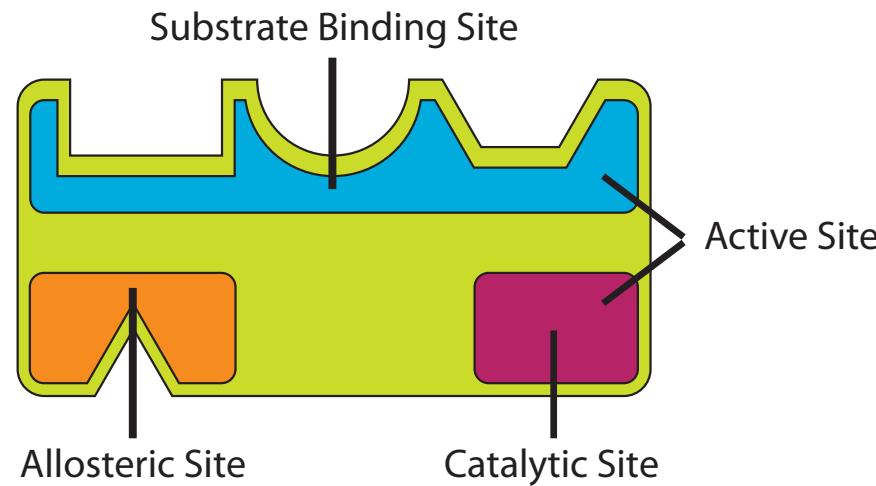
- > efficient simulation without kinetic parameters

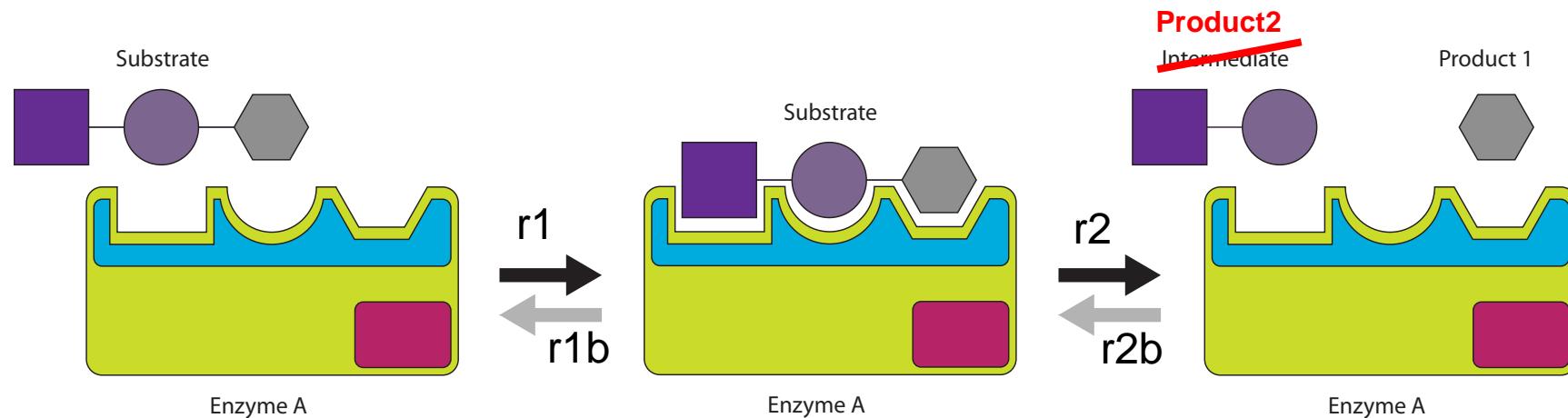
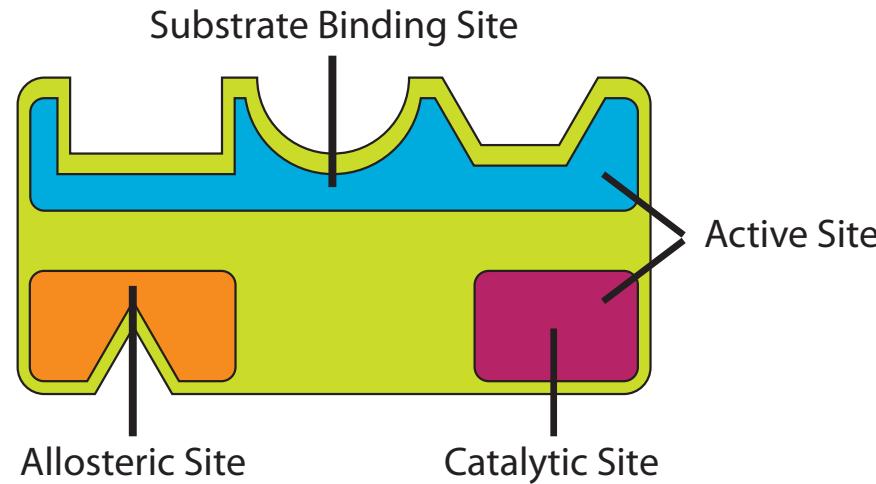
MODULAR MODELLING



**JOINT WORK WITH
MARY ANN BLÄTKE, WOLFGANG MARWAN**

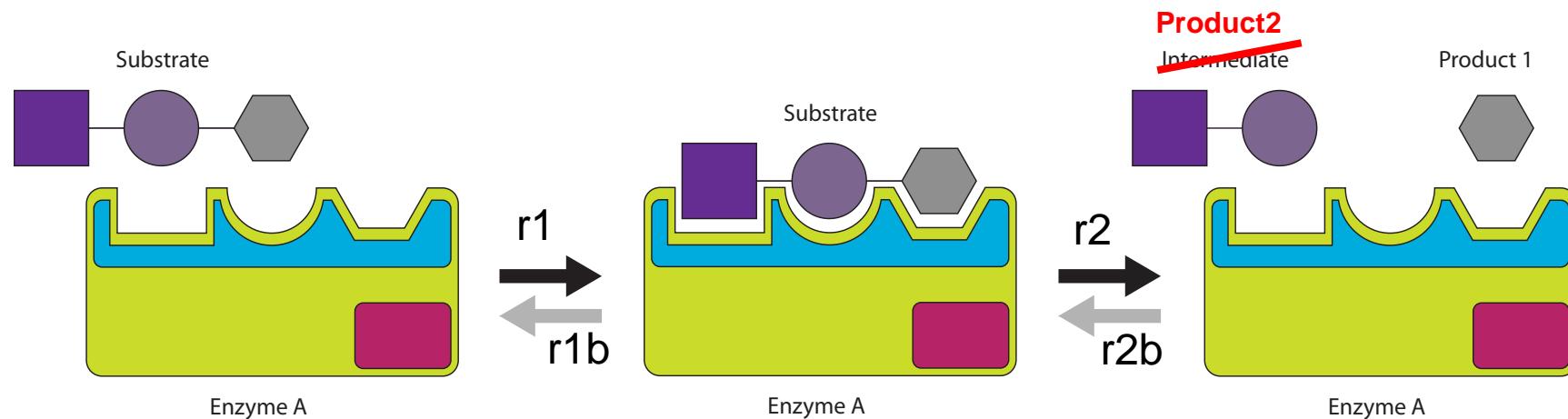




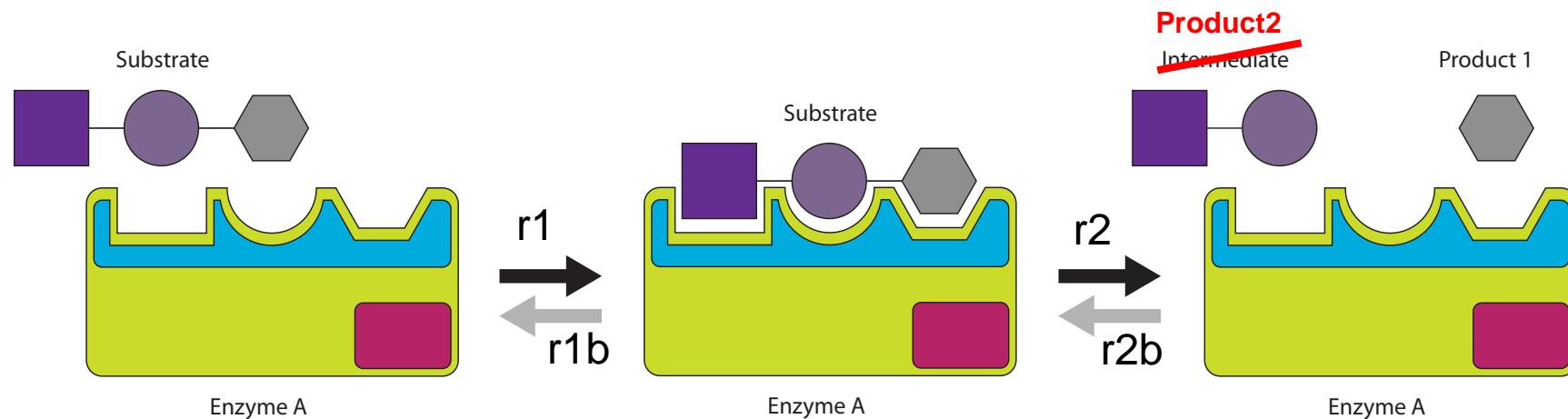


ENZYMATIC REACTIONS

PN & Bacterial BioEngineering

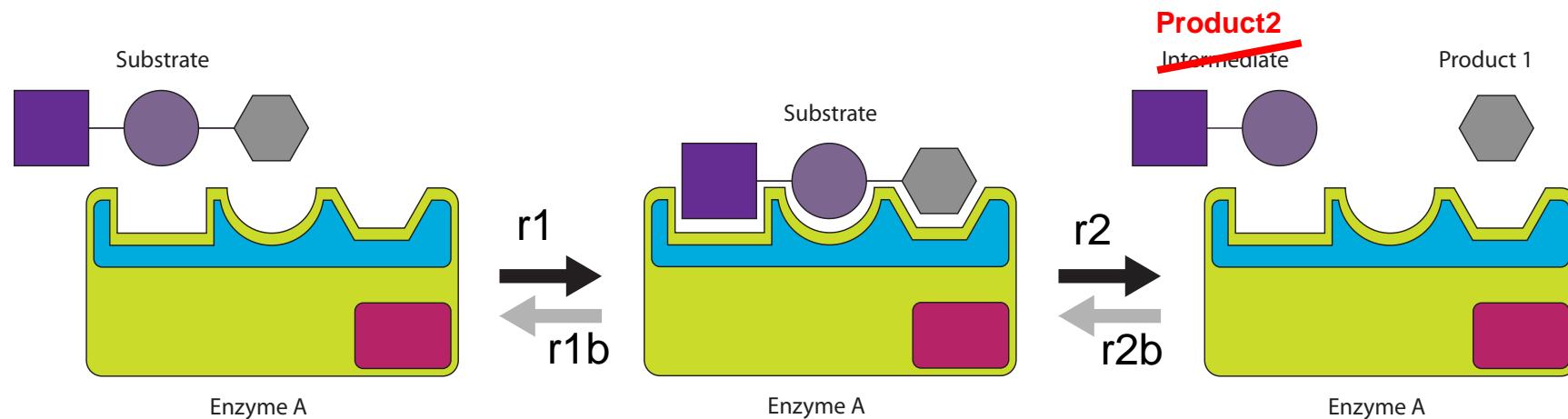


HOW TO CONSTRUCT THE MODEL?



HOW TO CONSTRUCT THE MODEL?

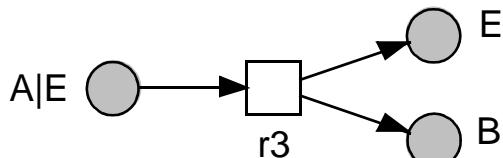
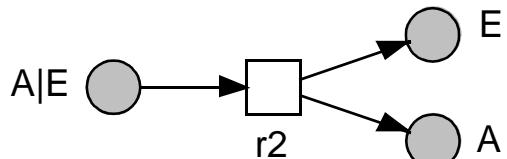
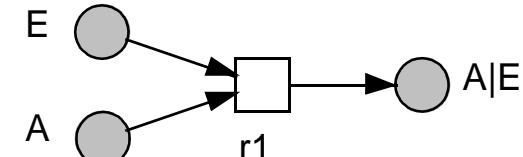
HOW TO DESIGN THE MODEL?



HOW TO DESIGN THE MODEL - LAYOUT OPTIONS

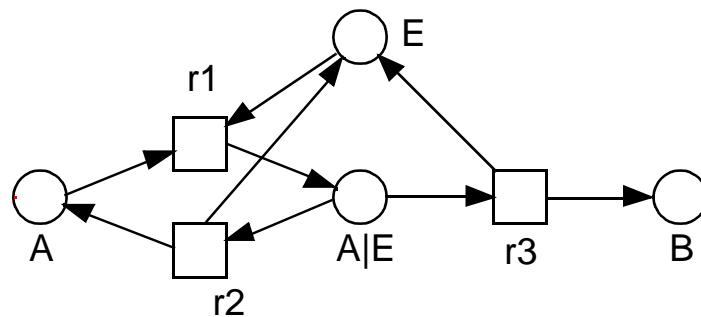
PN & Bacterial BioEngineering

reaction-centred view

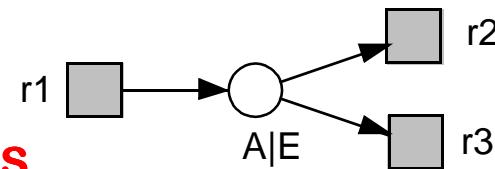
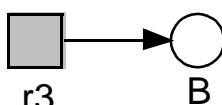
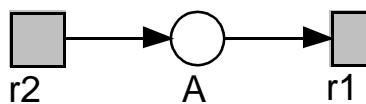


LOGICAL PLACES

process-oriented view



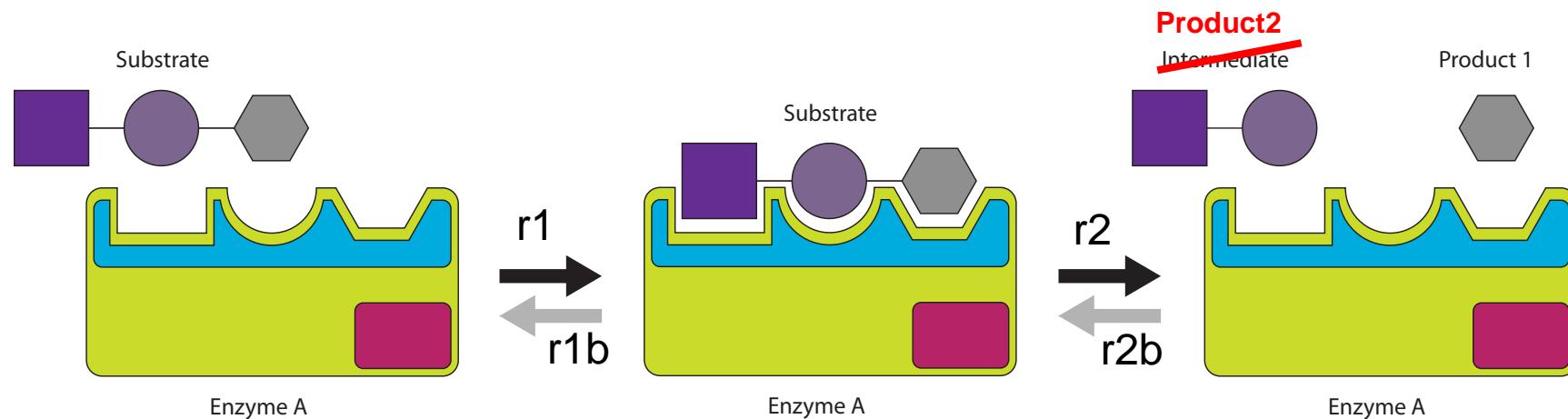
species-centred view



LOGICAL TRANSITIONS

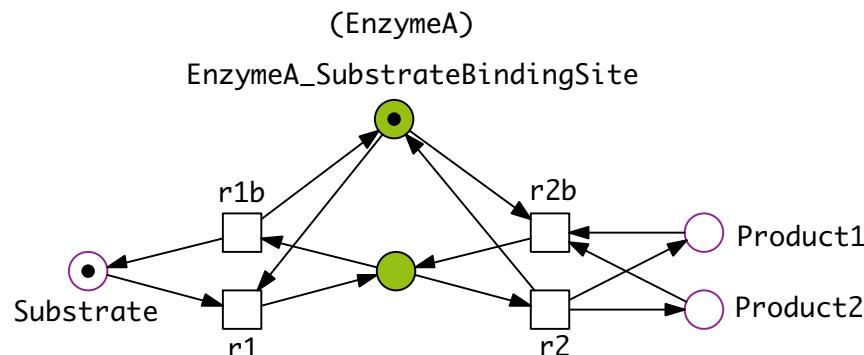
ENZYMATIC REACTIONS, A FIRST PETRI NET MODEL

PN & Bacterial BioEngineering

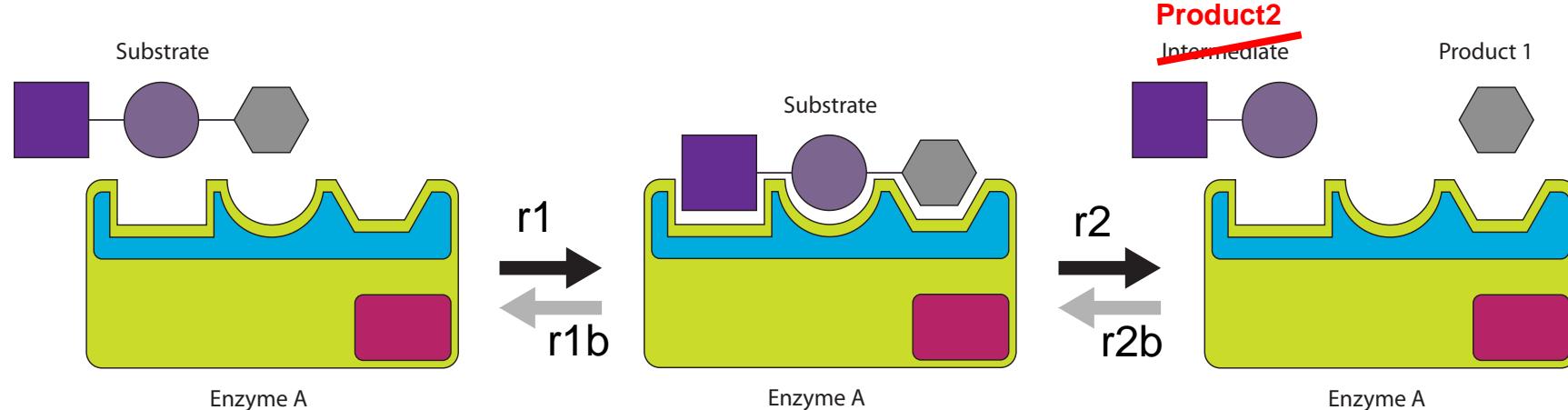


ENZYMATIC REACTIONS, A FIRST PETRI NET MODEL

PN & Bacterial BioEngineering

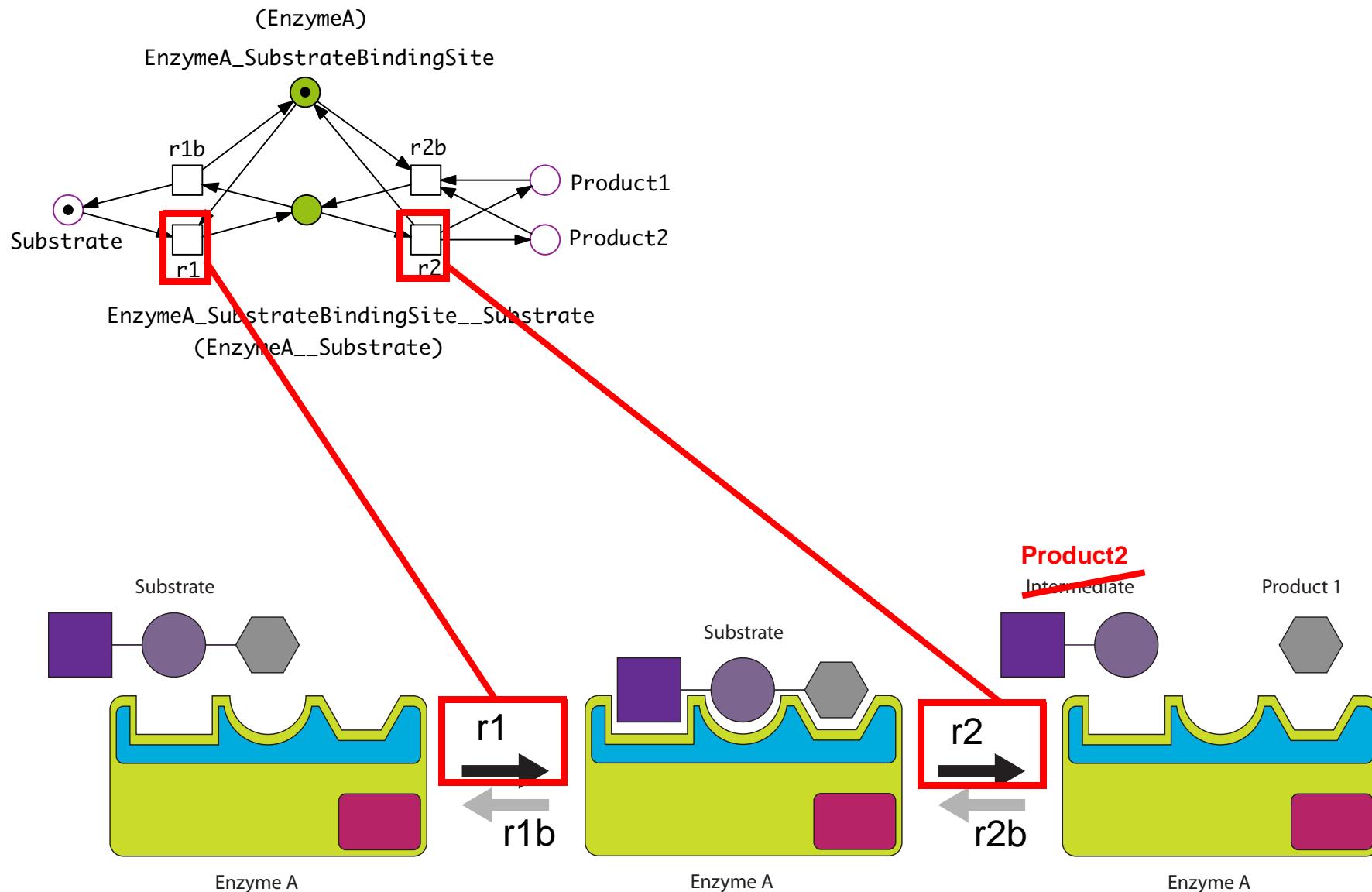


EnzymeA_SubstrateBindingSite__Substrate
(EnzymeA__Substrate)



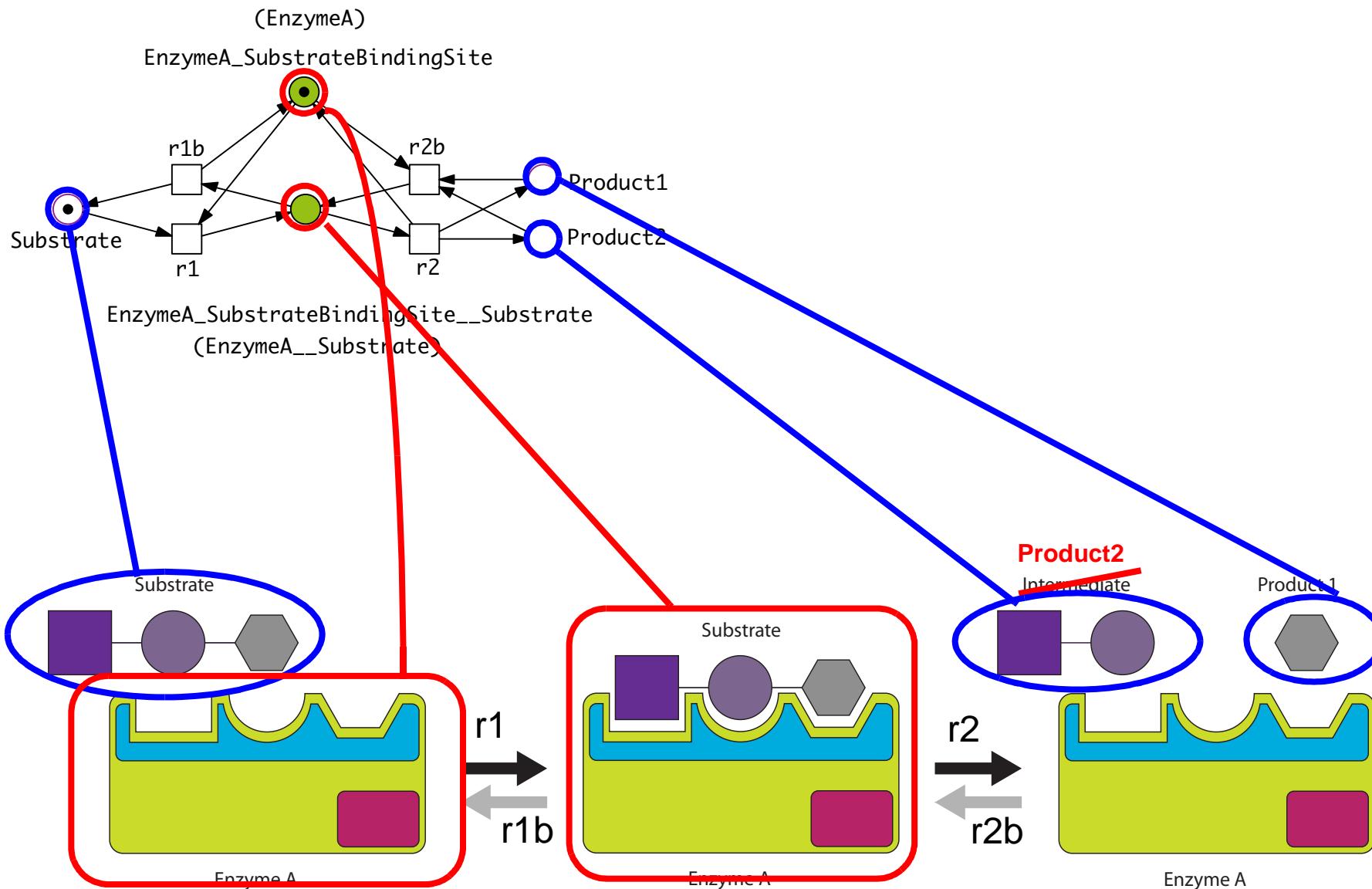
ENZYMATIC REACTIONS, A FIRST PETRI NET MODEL

PN & Bacterial BioEngineering



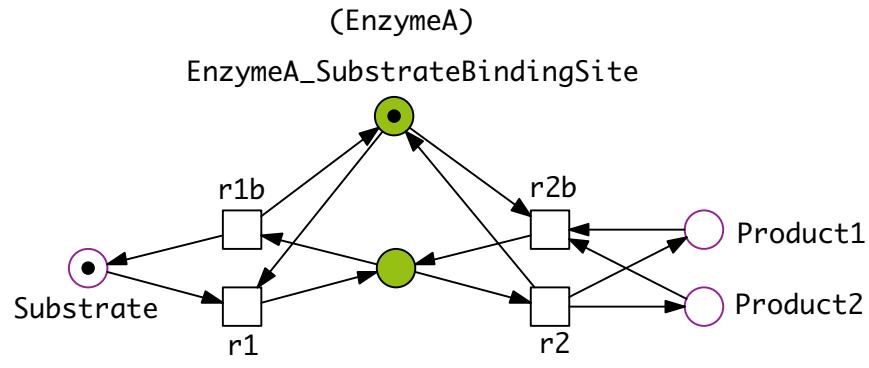
ENZYMATIC REACTIONS, A FIRST PETRI NET MODEL

PN & Bacterial BioEngineering



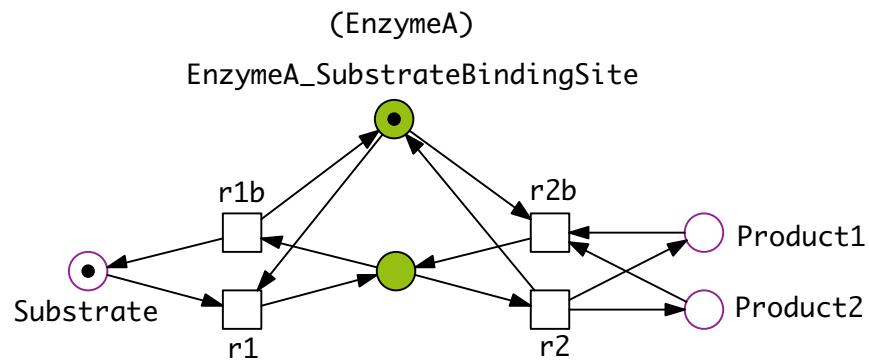
ENZYMATIC REACTIONS, LAYOUT OPTIONS

PN & Bacterial BioEngineering

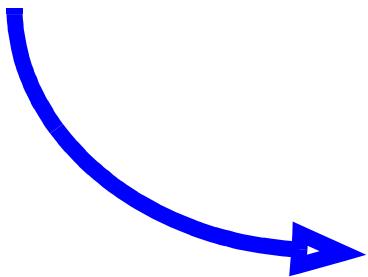


ENZYMATIC REACTIONS, LAYOUT OPTIONS

PN & Bacterial BioEngineering

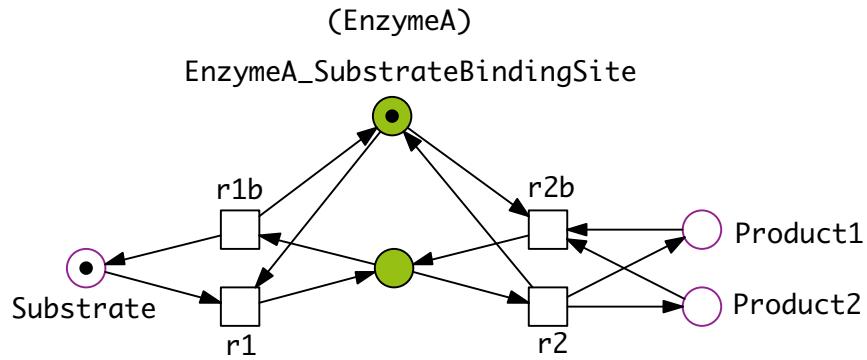


EnzymeA_SubstrateBindingSite__Substrate
(EnzymeA__Substrate)

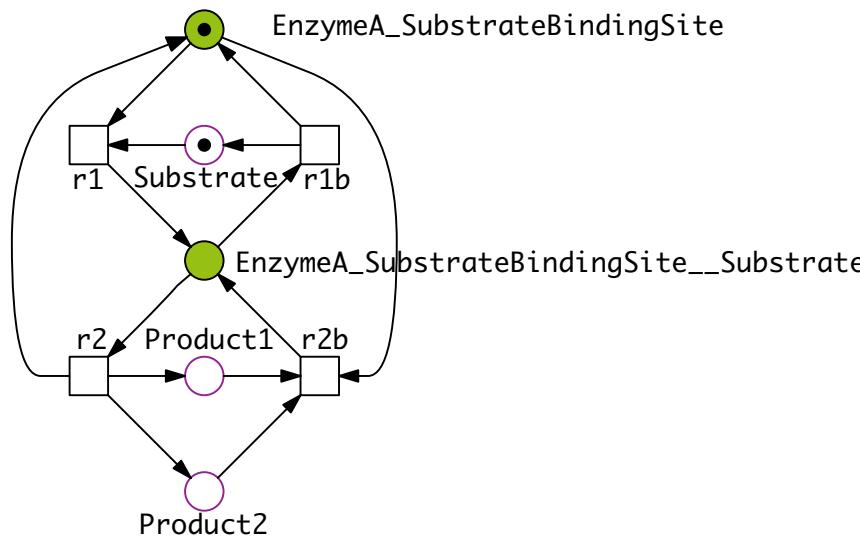


ENZYMATIC REACTIONS, LAYOUT OPTIONS

PN & Bacterial BioEngineering



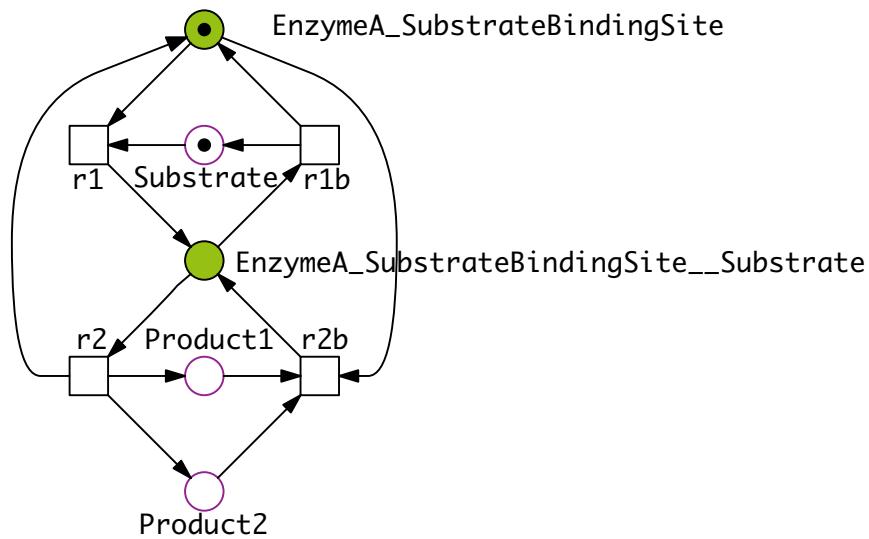
EnzymeA_SubstrateBindingSite__Substrate
(EnzymeA__Substrate)



ENZYMATIC REACTIONS AS MODULES

PN & Bacterial BioEngineering

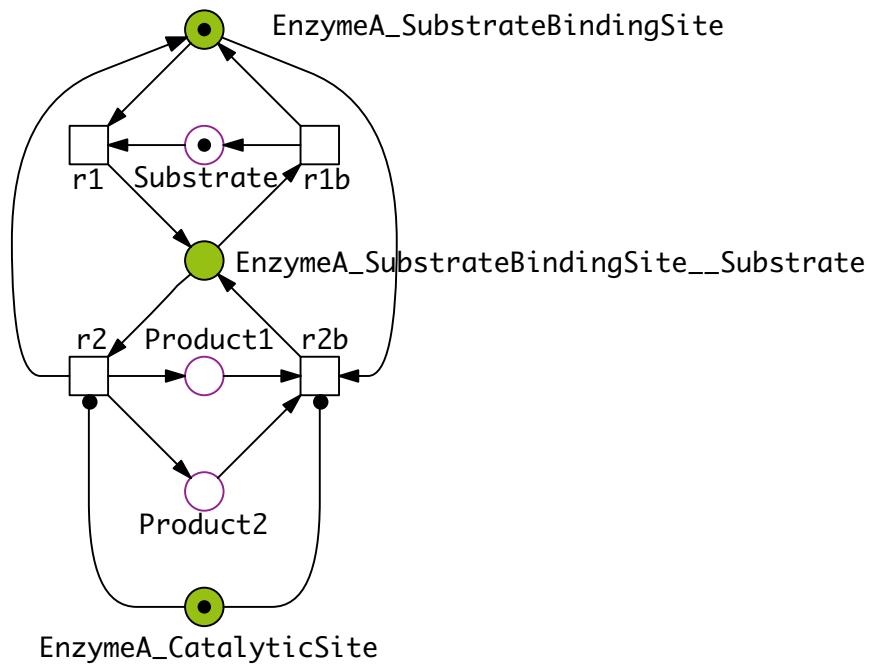
Module: Enzyme A



ENZYMATIC REACTIONS AS MODULES

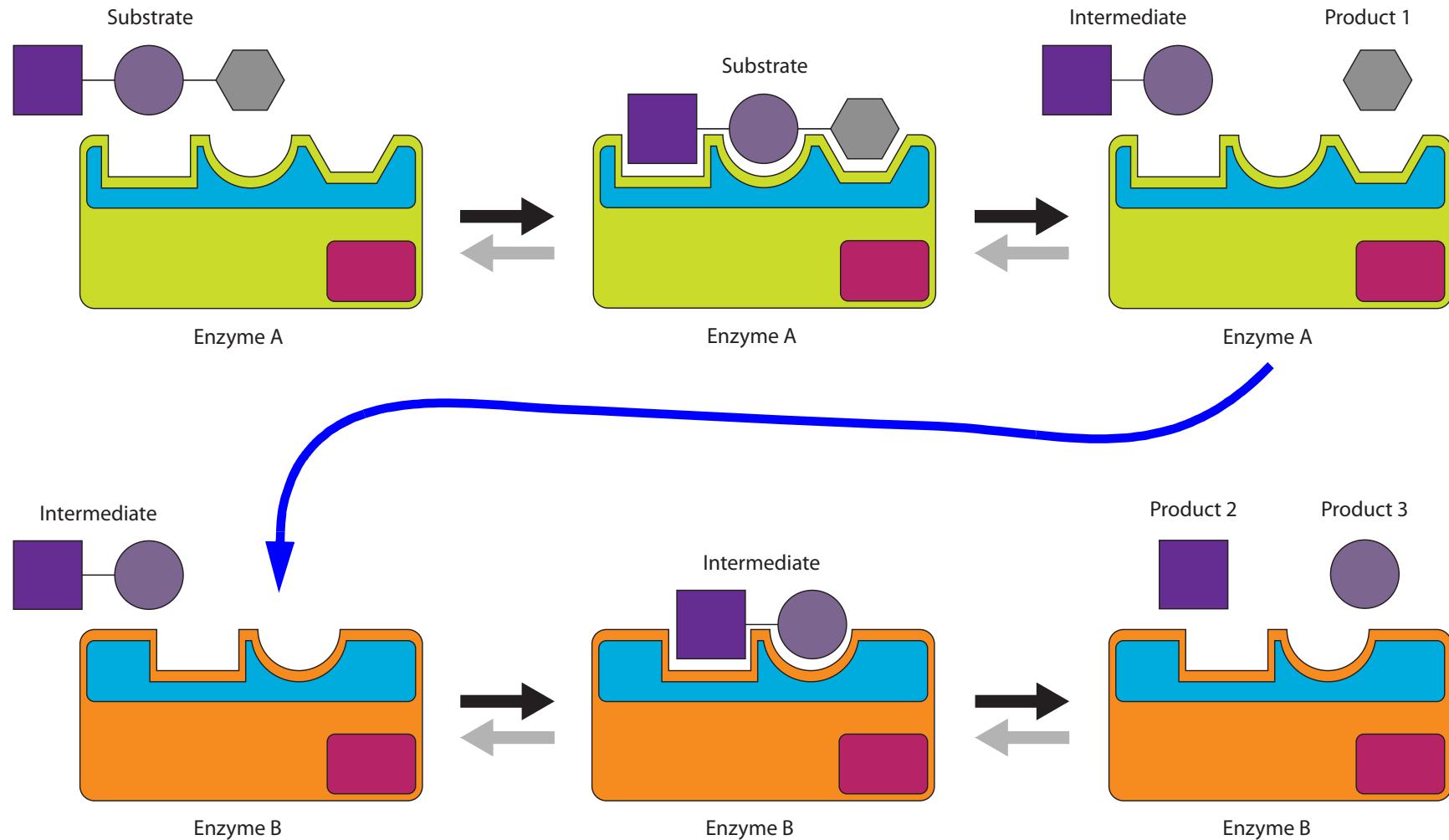
PN & Bacterial BioEngineering

Module: Enzyme A



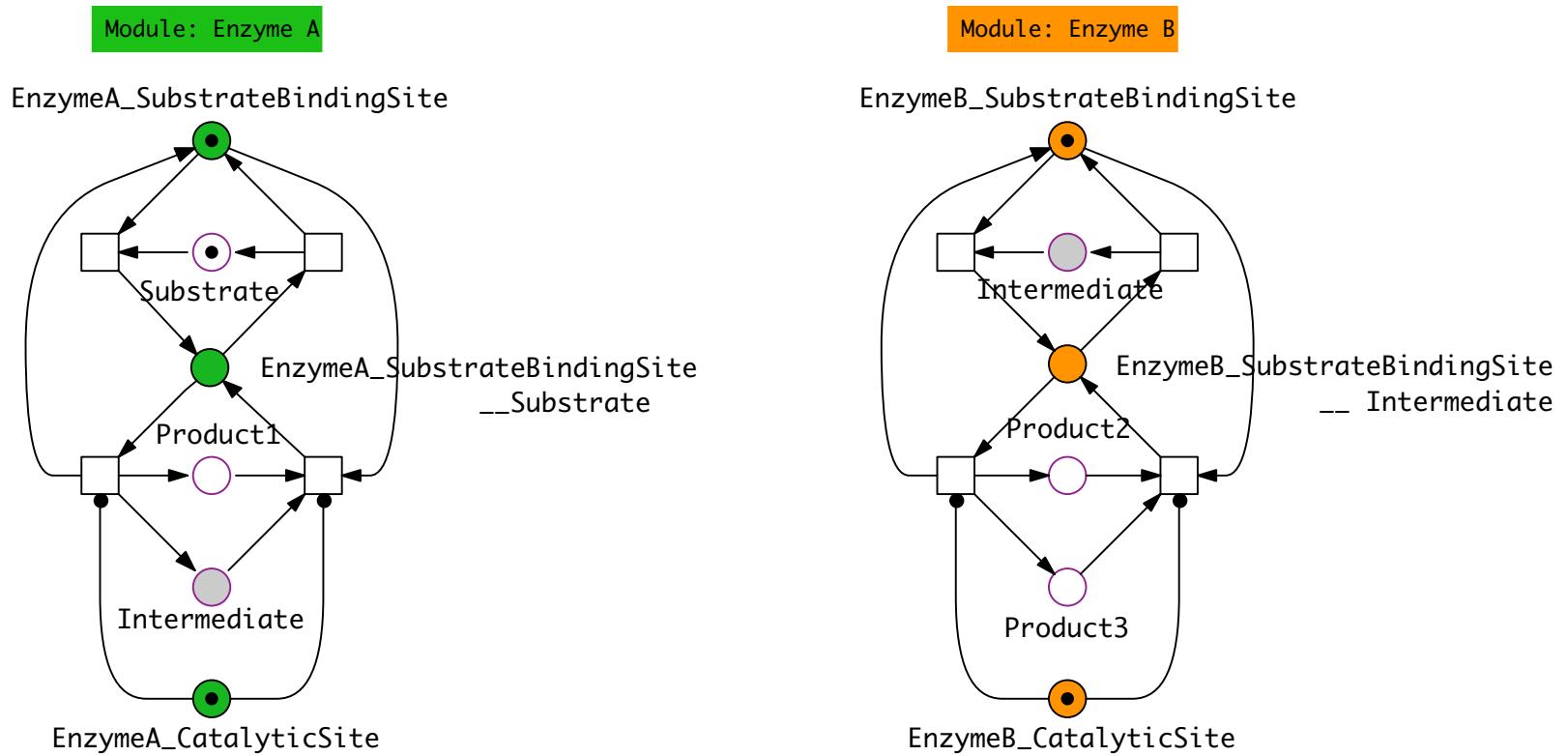
ENZYMATIC REACTIONS IN A CHAIN

PN & Bacterial BioEngineering



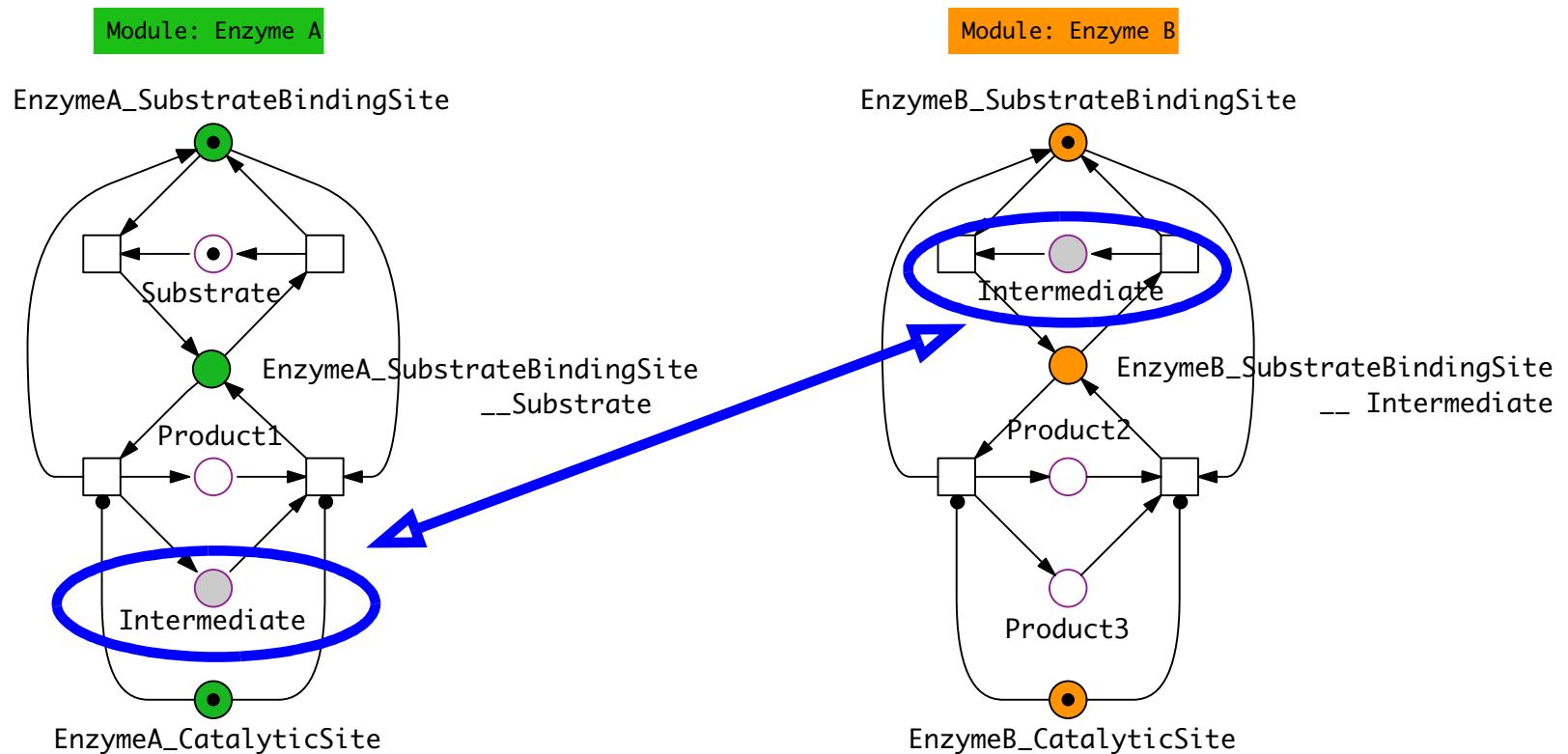
ENZYMATIC REACTIONS IN A CHAIN

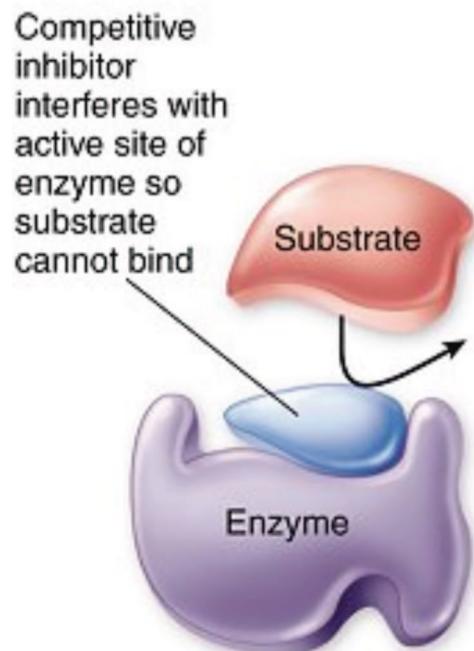
PN & Bacterial BioEngineering



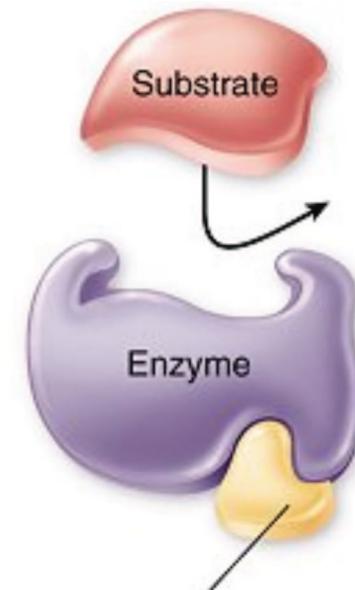
ENZYMATIC REACTIONS IN A CHAIN

PN & Bacterial BioEngineering





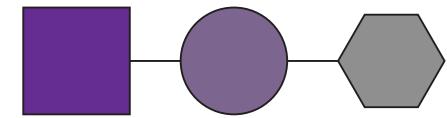
(a) Competitive inhibition



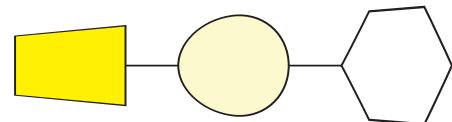
(b) Noncompetitive inhibition

INHIBITION OF ENZYMATIC REACTIONS

PN & Bacterial BioEngineering



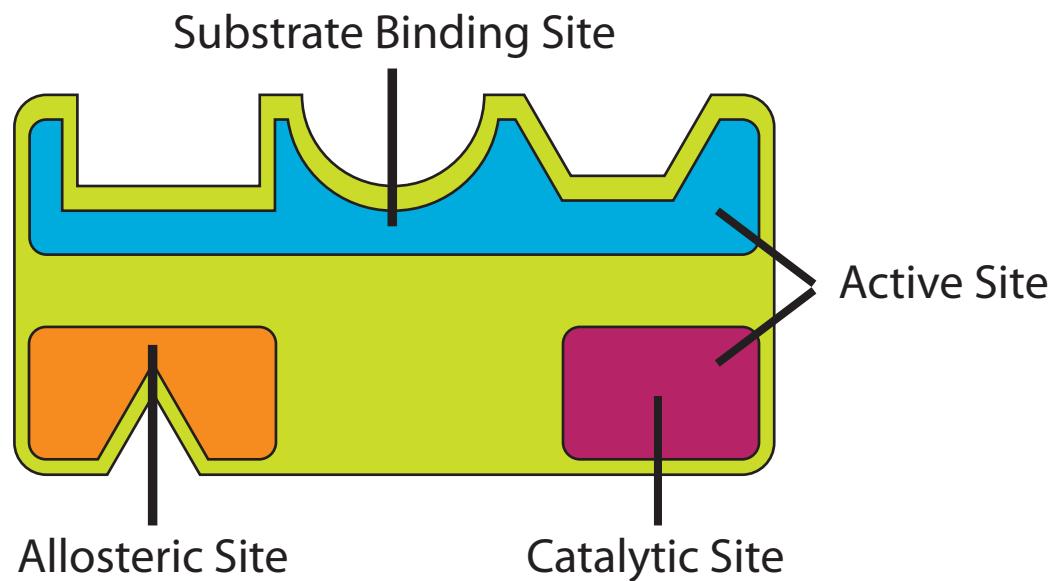
Substrate



competitive Inhibitor

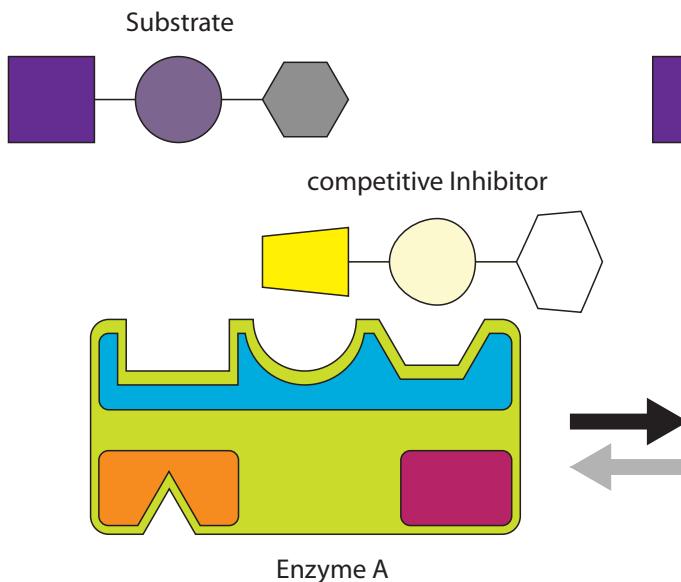


non-competitive Inhibitor

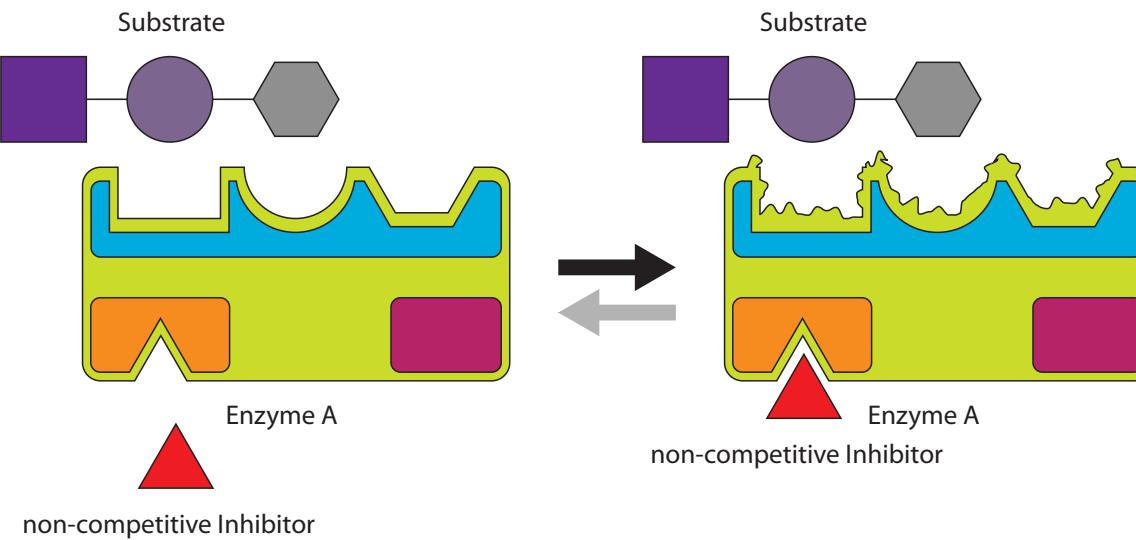


INHIBITION OF ENZYMATIC REACTIONS

PN & Bacterial BioEngineering

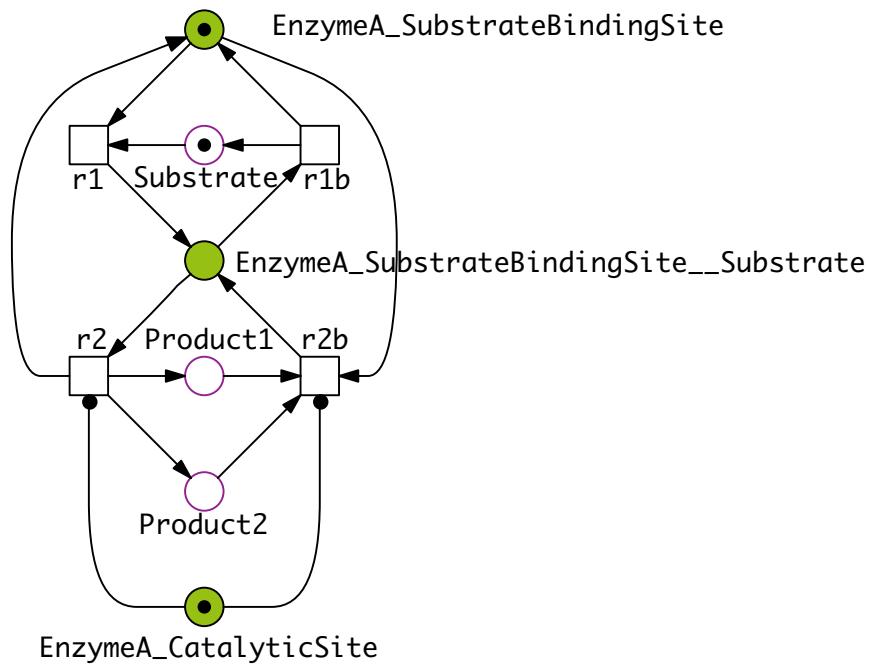


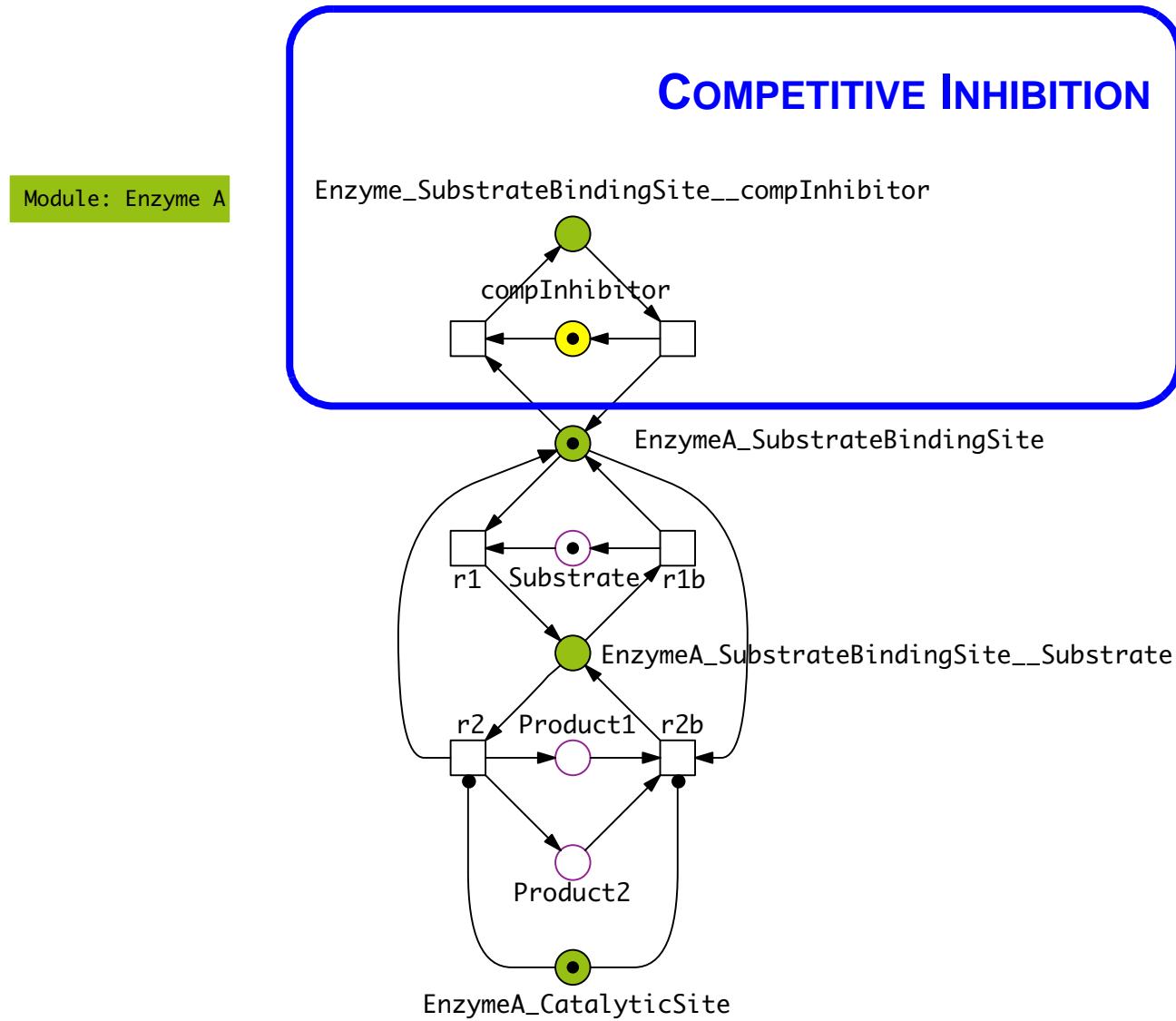
COMPETITIVE INHIBITION



NON-COMPETITIVE INHIBITION

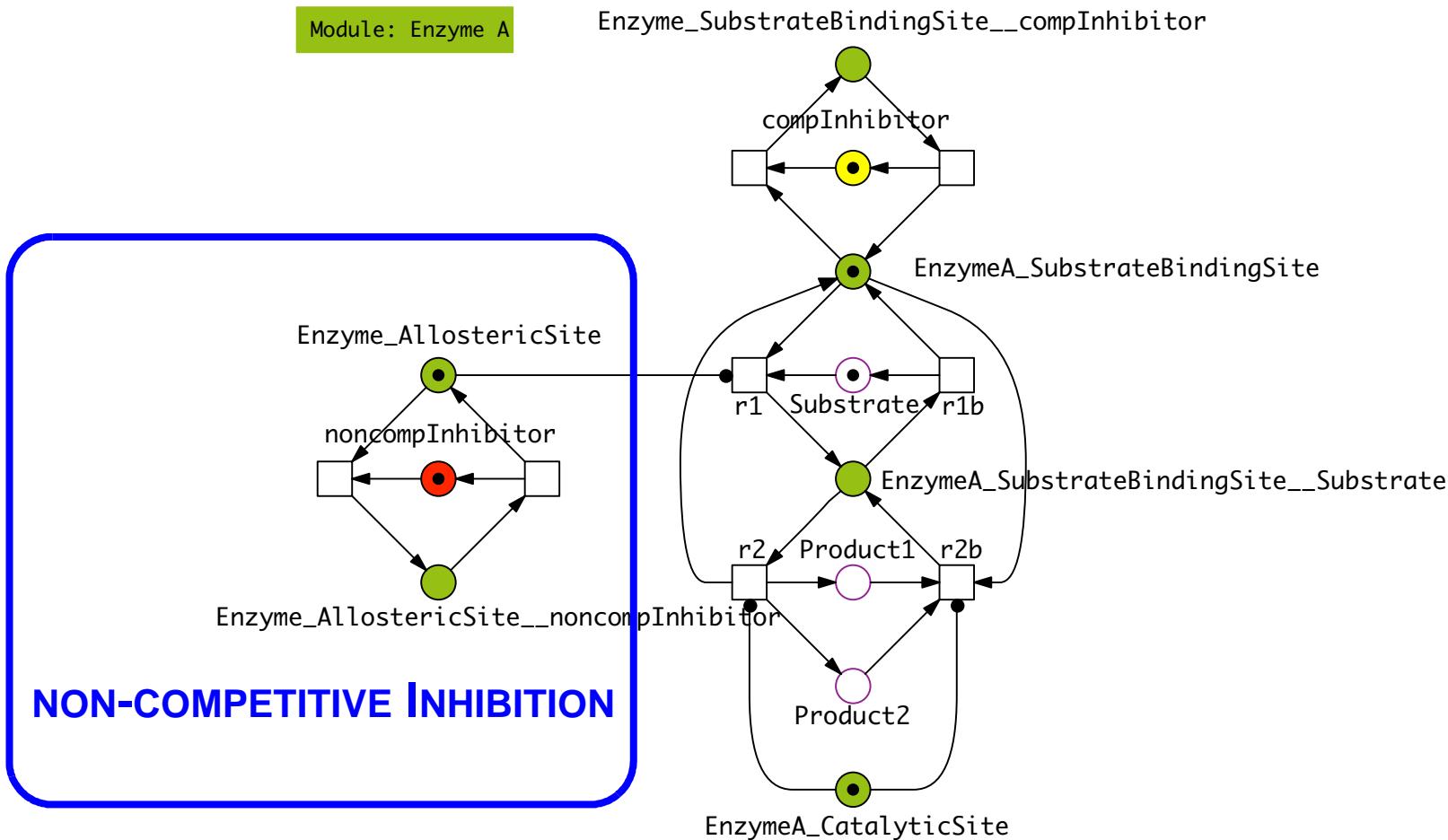
Module: Enzyme A





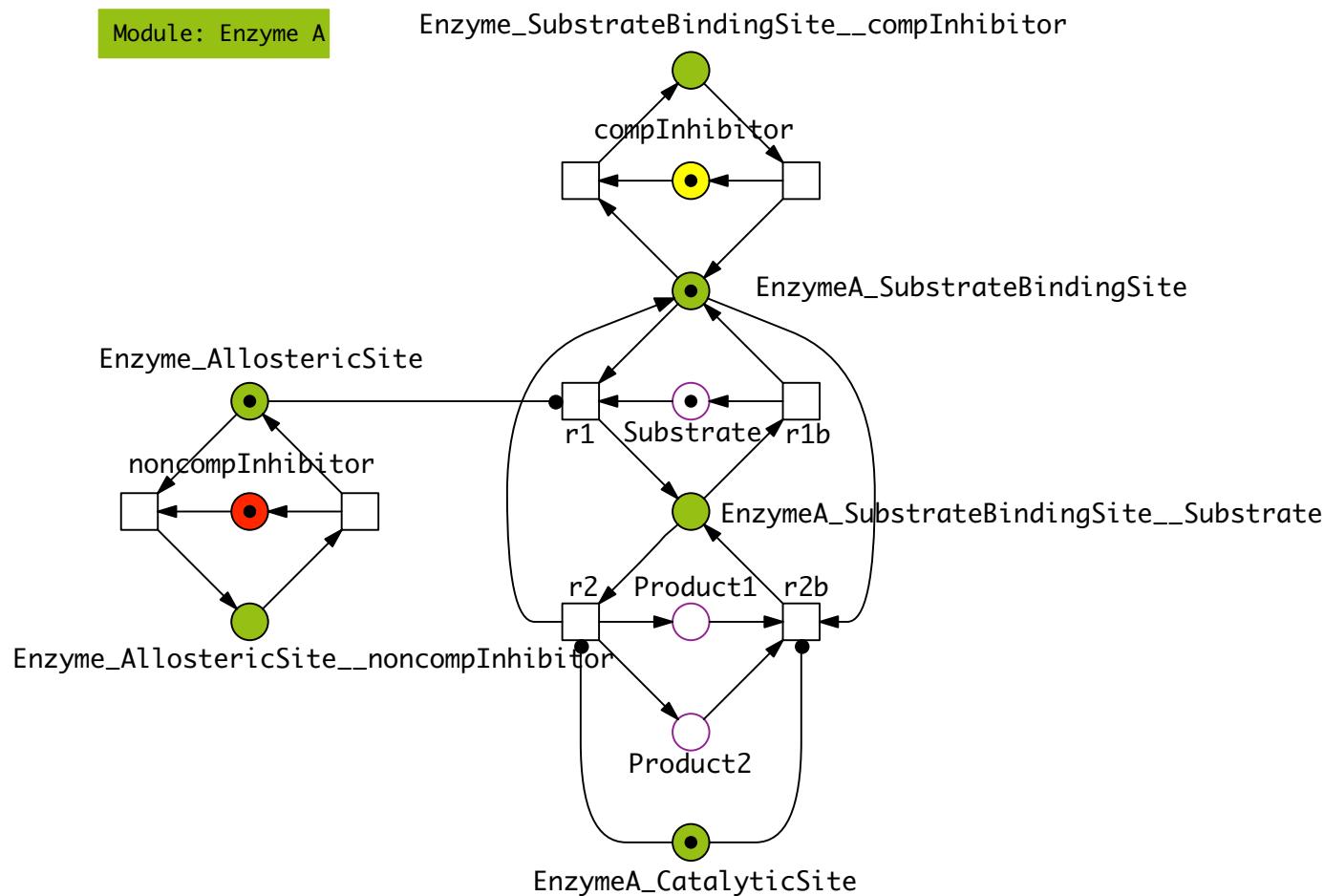
ENZYME MODULE WITH INHIBITION 2

PN & Bacterial BioEngineering



ENZYME MODULE WITH INHIBITION 2

PN & Bacterial BioEngineering



□ **JAK-STAT pathway in IL-6 signalling [BLÄTKE ET A. 2013]**

- > *Anna Dietrich, Fred Scharper, OvG University Magdeburg*
- > *constraints: induced by IL-6 cytokine, one isoform*
- > *7 protein modules: JAK1, STAT3, SOCS3, SHP2, gp130, IL-6, IL-6R + SOCS3 biosynthesis module*
- > *in total: 92 places, 102 transitions, 58 pages, nesting depth 4*

□ **phosphate regulatory network [BLÄTKE ET A. 2012]**

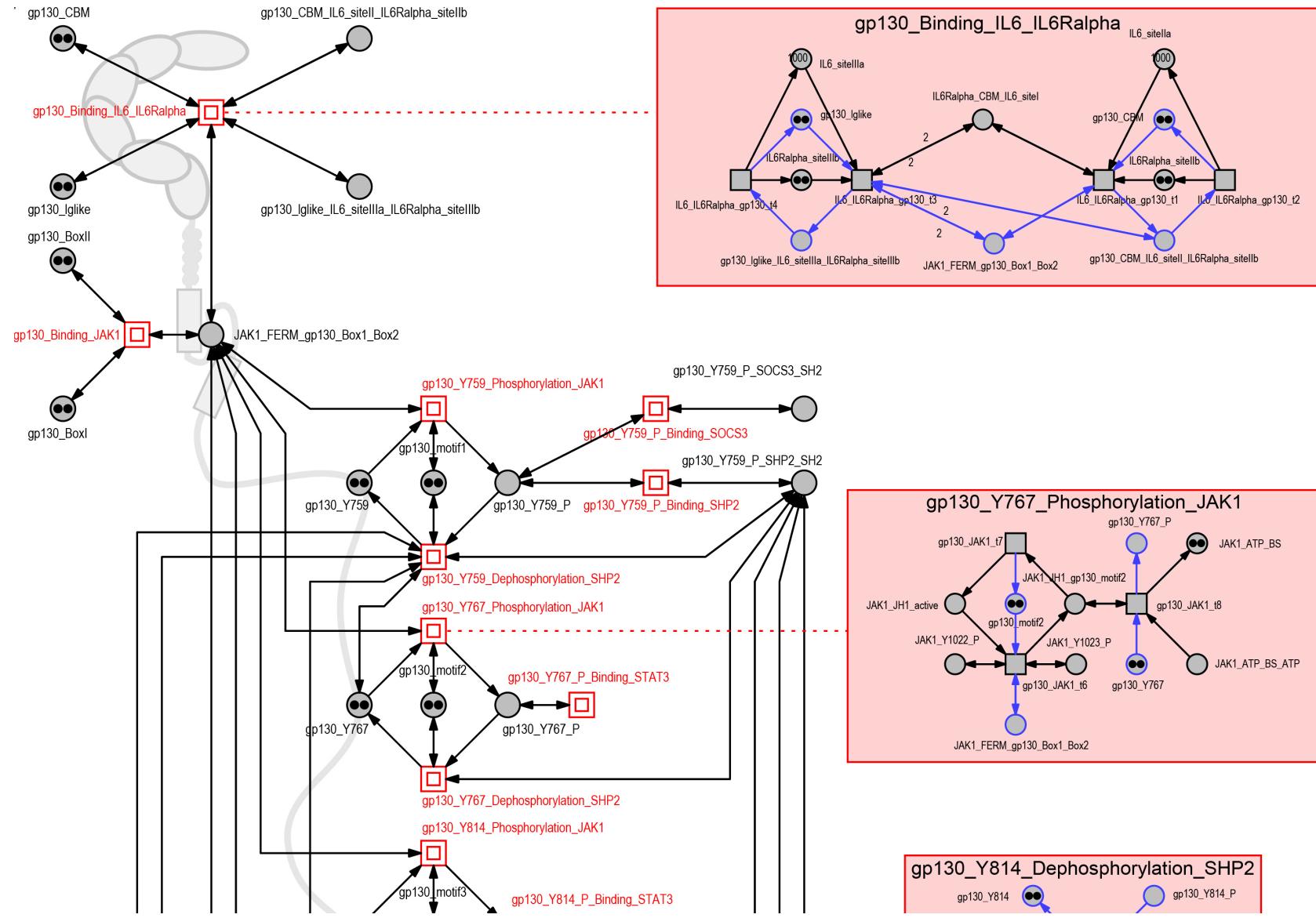
- > *basics demonstrated in CMSB 2012*

□ **nociceptive network in pain signalling [HUCHO ET AL., IN PREPARATION]**

- > *running BMBF project with 7 consortium partners*
- > *(so far) 38 protein modules, among them several membrane receptors, kinases, phosphatases and ion channels*
- > *in total (so far): 713 places, 775 transitions, 325 pages, nesting depth 4*

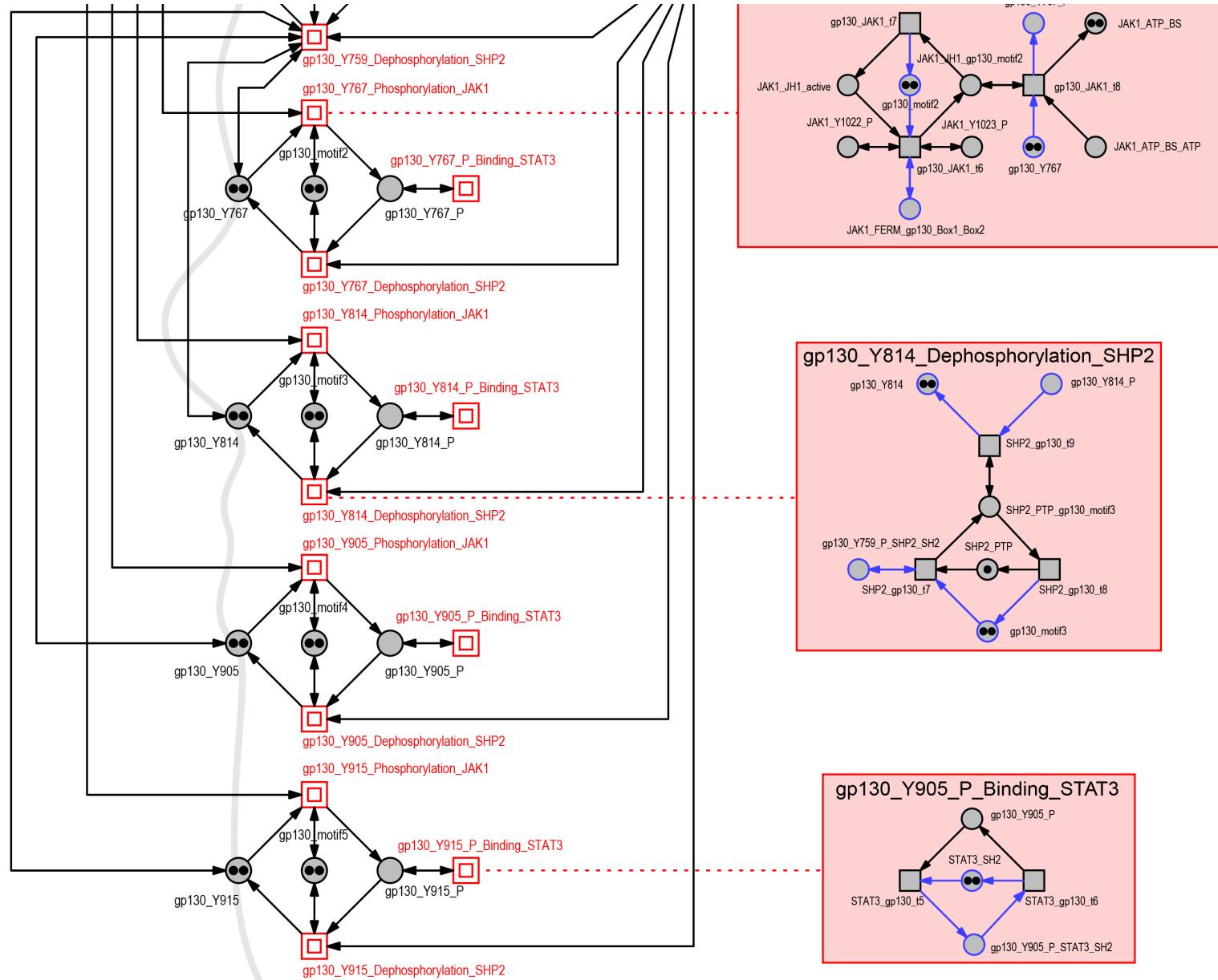
MODULE GP130 TRANSMEMBRANE RECEPTOR

PN & Bacterial BioEngineering



MODULE GP130 TRANSMEMBRANE RECEPTOR

PN & Bacterial BioEngineering

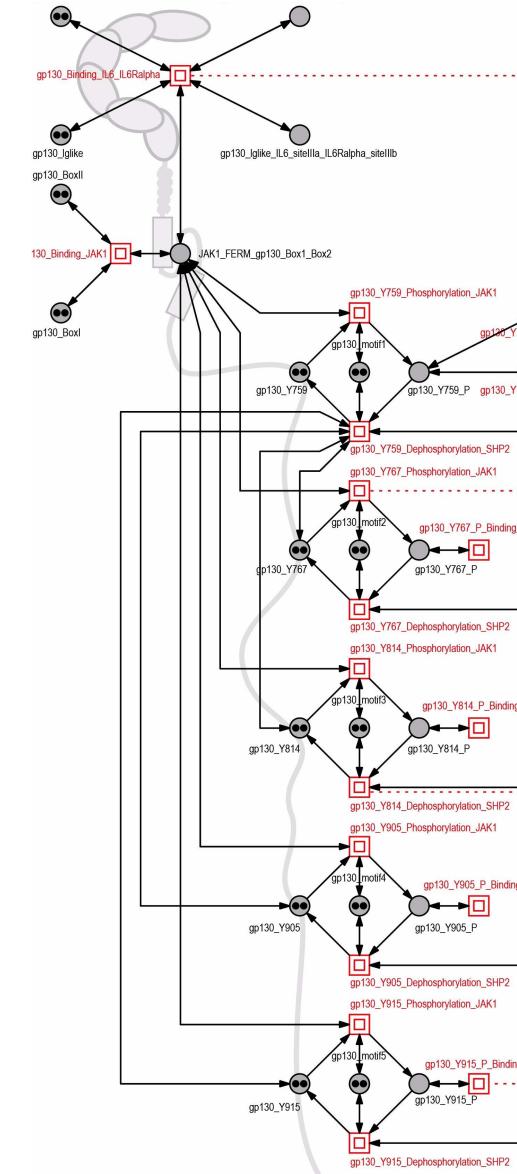


□ monolithic models

- > *process-oriented view*
- > *cons: it's difficult to*
 - reuse network components
 - combine networks

□ modular models

- > *a priori modular network design*
- > *object-oriented view*
 - object = protein
 - well-defined object interfaces = compositionality
- > *pros*
 - step-wise incorporation of structural knowledge
 - step-wise construction of larger networks
 - automatic mutation of modules/networks
- > *personalised medicine (Horizon 2020 -PHC)*



ATOMIC FRAGMENTS

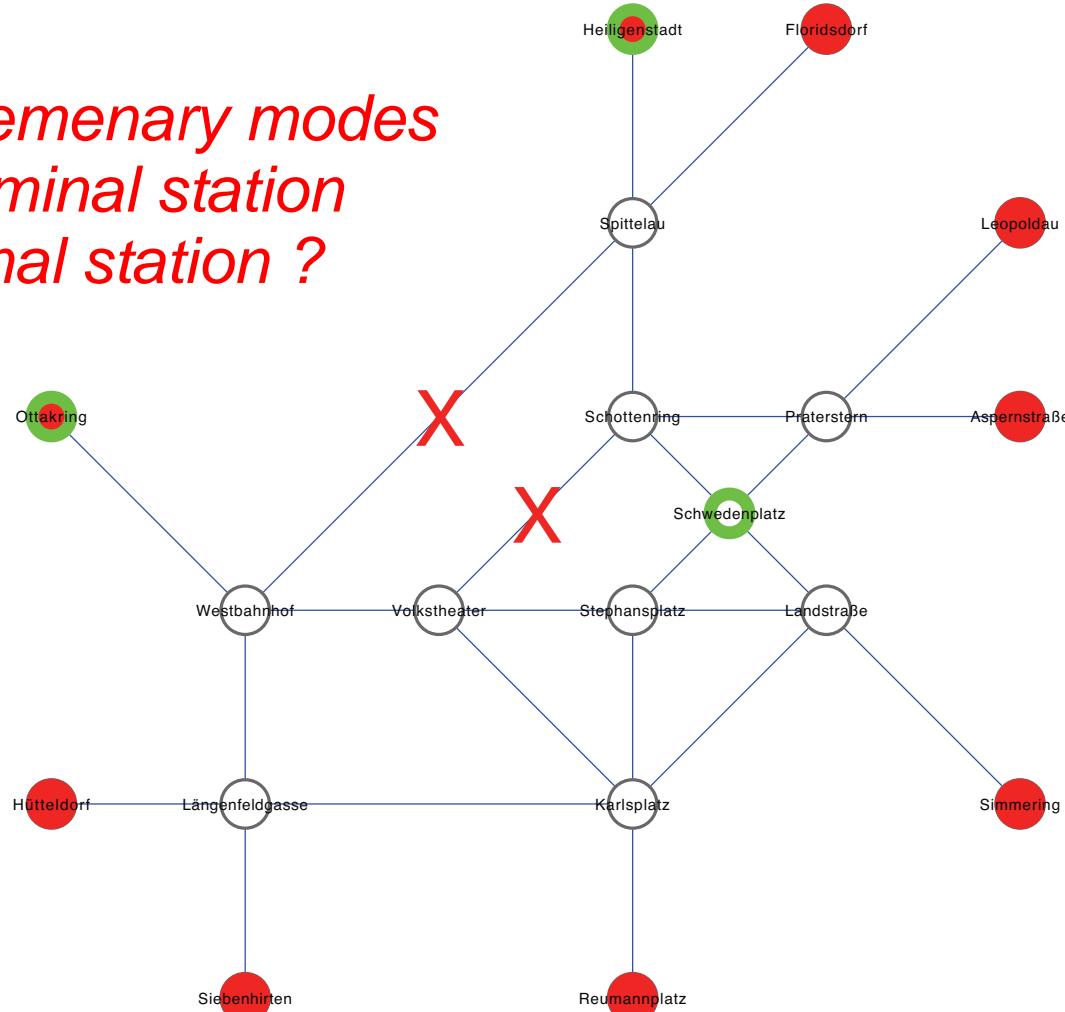
T- INVARIANTS
 (ELEMENTARY (FLUX) MODES)
 (EXTREME PATHWAYS)
 (GENERIC PATHWAYS)





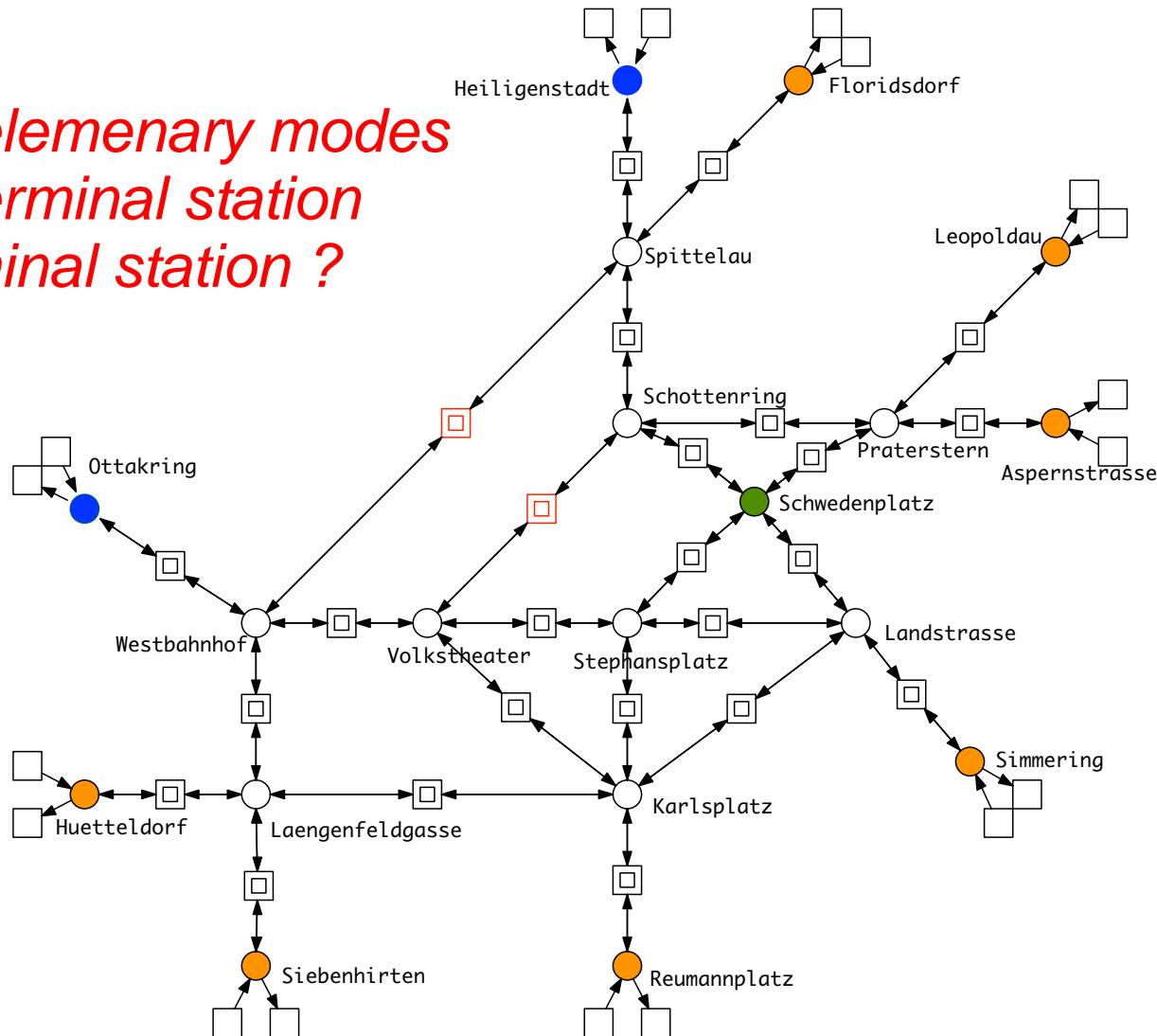
[Zanghellini 2013]

*How many elementary modes
from any terminal station
to any terminal station ?*



[Zanghellini 2013]

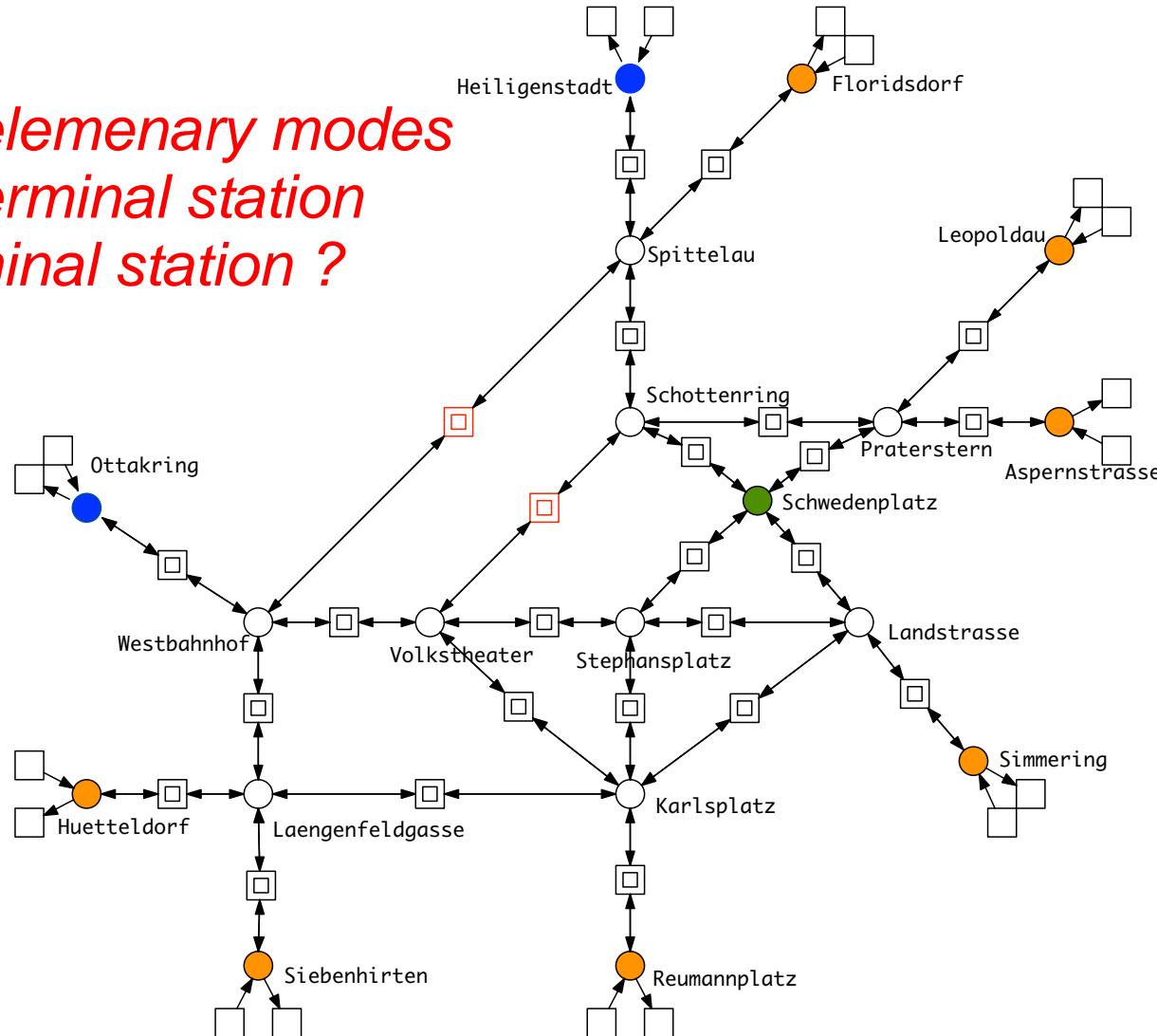
*How many elemenary modes
from any terminal station
to any terminal station ?*



*How many elemenary modes
from any terminal station
to any terminal station ?*

-> 2,800

- 128
**CLOSED
CYCLES**



STOICHIOMETRIC MATRIX

PN & Bacterial BioEngineering

- a representation of the net structure

=> incidence matrix

P \ T	t1	...	tj	...	tm
p1					
pi			cij		
:			Δt_j		
pn					

$$c_{ij} = (p_i, t_j) = F(t_j, p_i) - F(p_i, t_j) = \Delta t_j(p_i)$$

$$\Delta t_j = \Delta t_j(*)$$

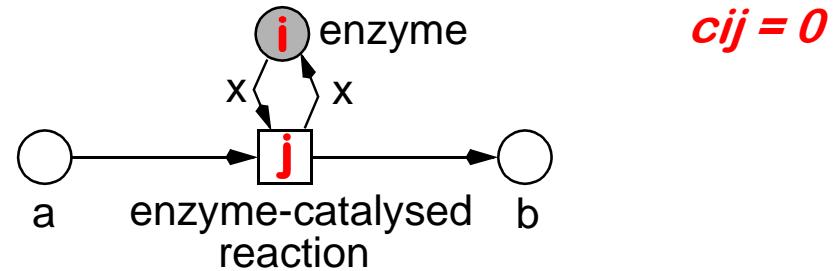
- matrix entry c_{ij} :

token change in place p_i by firing of transition t_j

- matrix column Δt_j :

vector describing the change of the whole marking by firing of t_j

- side-conditions are neglected



- Lautenbach, 1973
- T-invariant x
 - > integer solution of $Cx = 0, x \neq 0, x \geq 0$
- support of a T-invariant x -> $\text{supp}(x)$
 - > set of transitions involved, i.e. $x(i) \neq 0$
- minimal T-invariants
 - > there is no T-invariant with a smaller support
 - > gcd of all entries is 1
- any T-invariant is a non-negative linear combination of minimal ones
 - > multiplication with a positive integer
 - > addition
 - > Division by gcd

$$kx = \sum_i a_i x_i$$

- T-invariants = (multi-) sets of transitions = Parikh vector

- > zero effect on marking
- > reproducing a marking / system state

- two interpretations

1. *partially ordered transition sequence* -> behaviour understanding
of transitions occurring one after the other
-> substance / signal flow

2. *relative transition firing rates* -> steady state behaviour
of transitions occurring permanently & concurrently
-> steady state behaviour

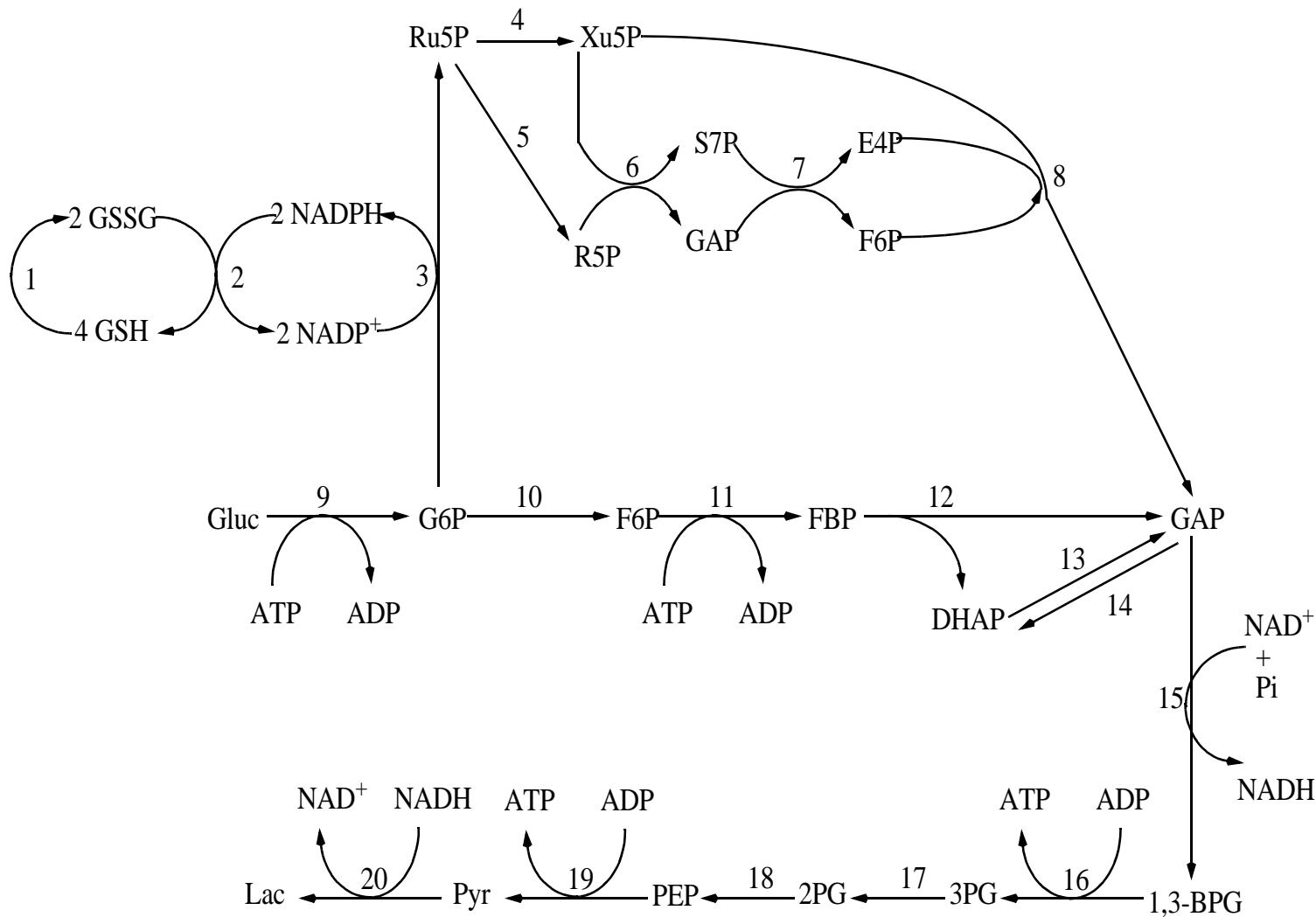
- a minimal T-invariant defines a connected subnet

- > the T-invariant's transitions (the support),
+ all their pre- and post-places
+ the arcs in between
- > pre-set of support = post-set of support

Ex1 - Glycolysis and Pentose Phosphate Pathway

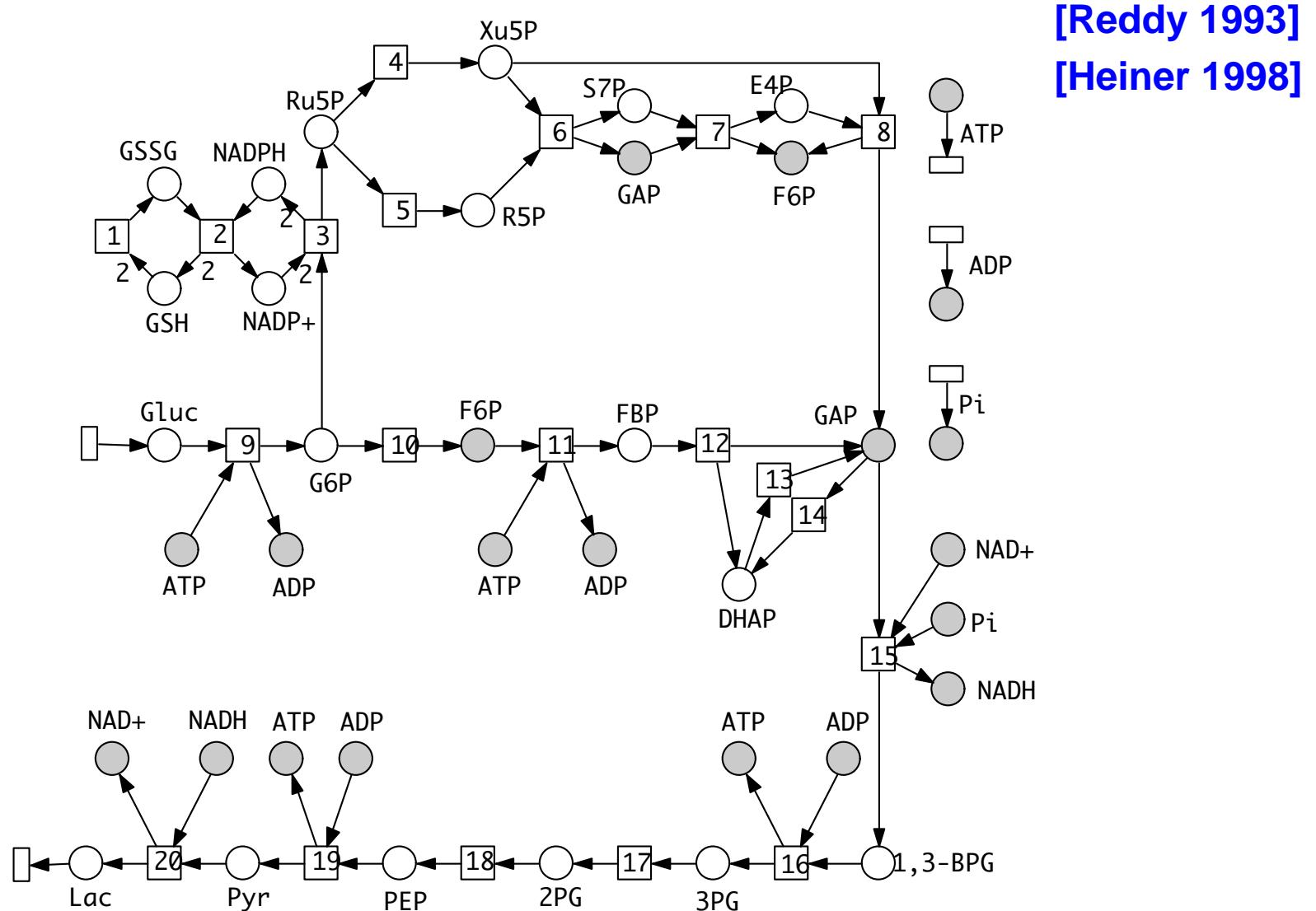
PN & Bacterial BioEngineering

[Reddy 1993]



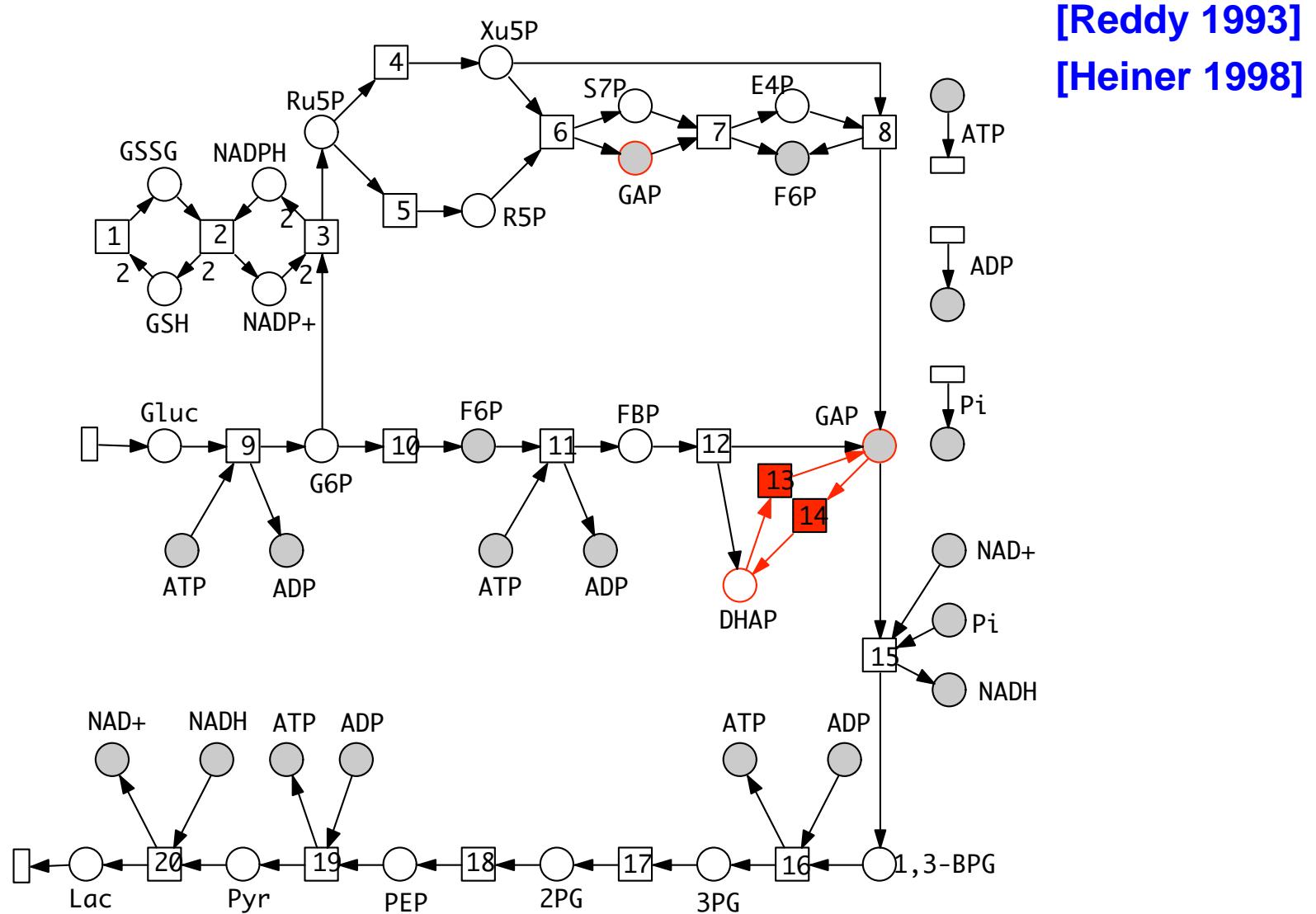
Ex - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



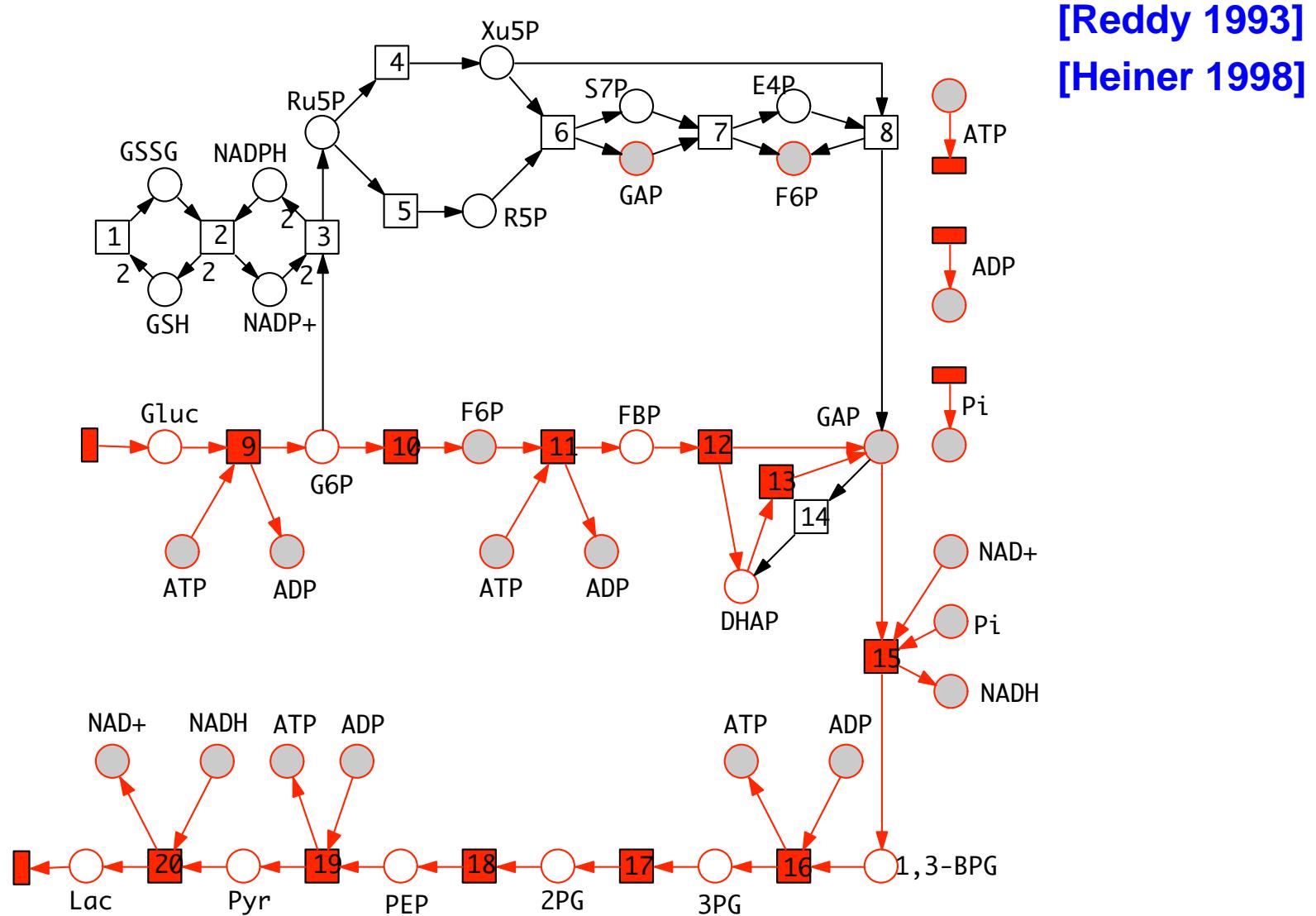
Ex - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



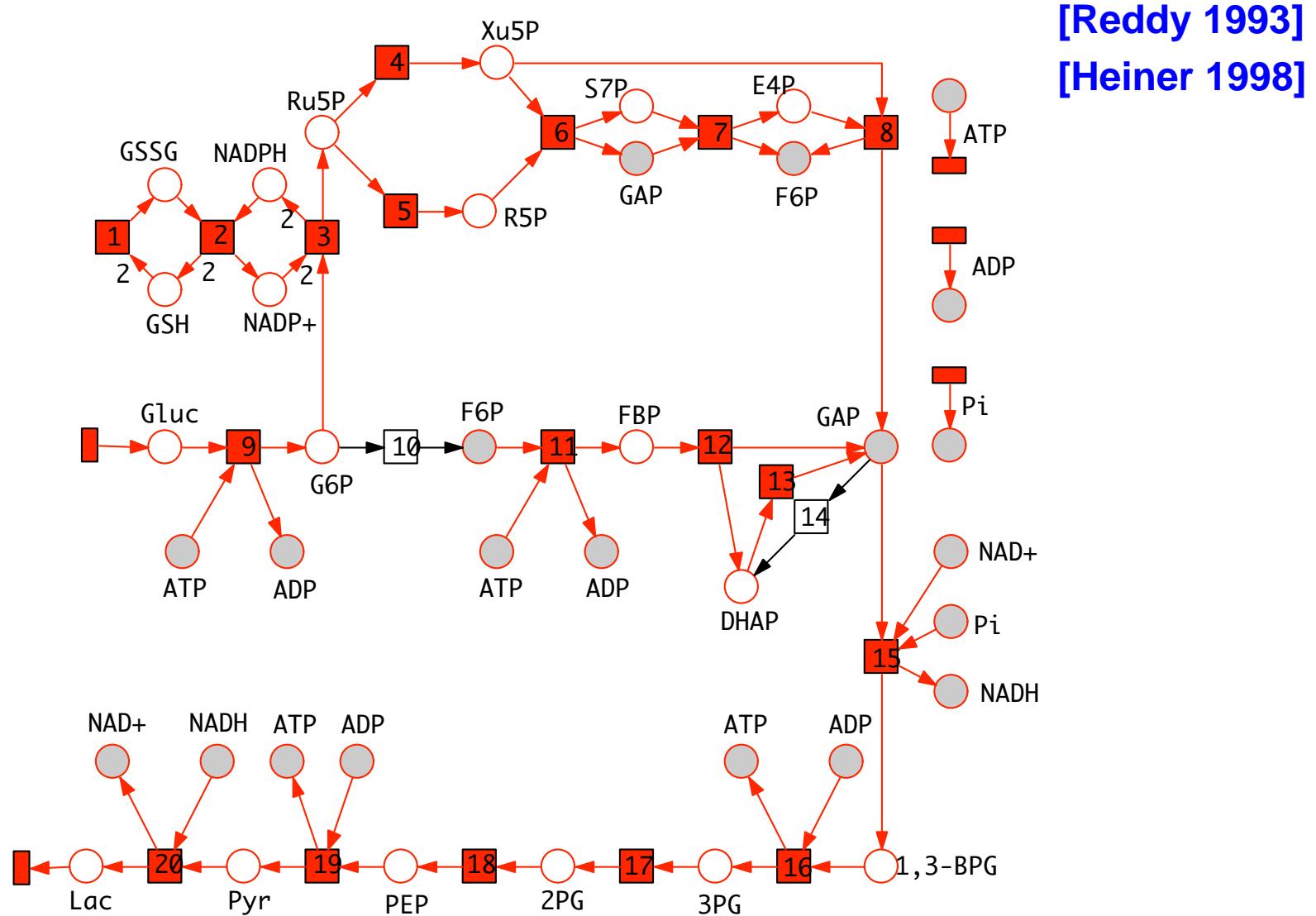
Ex - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



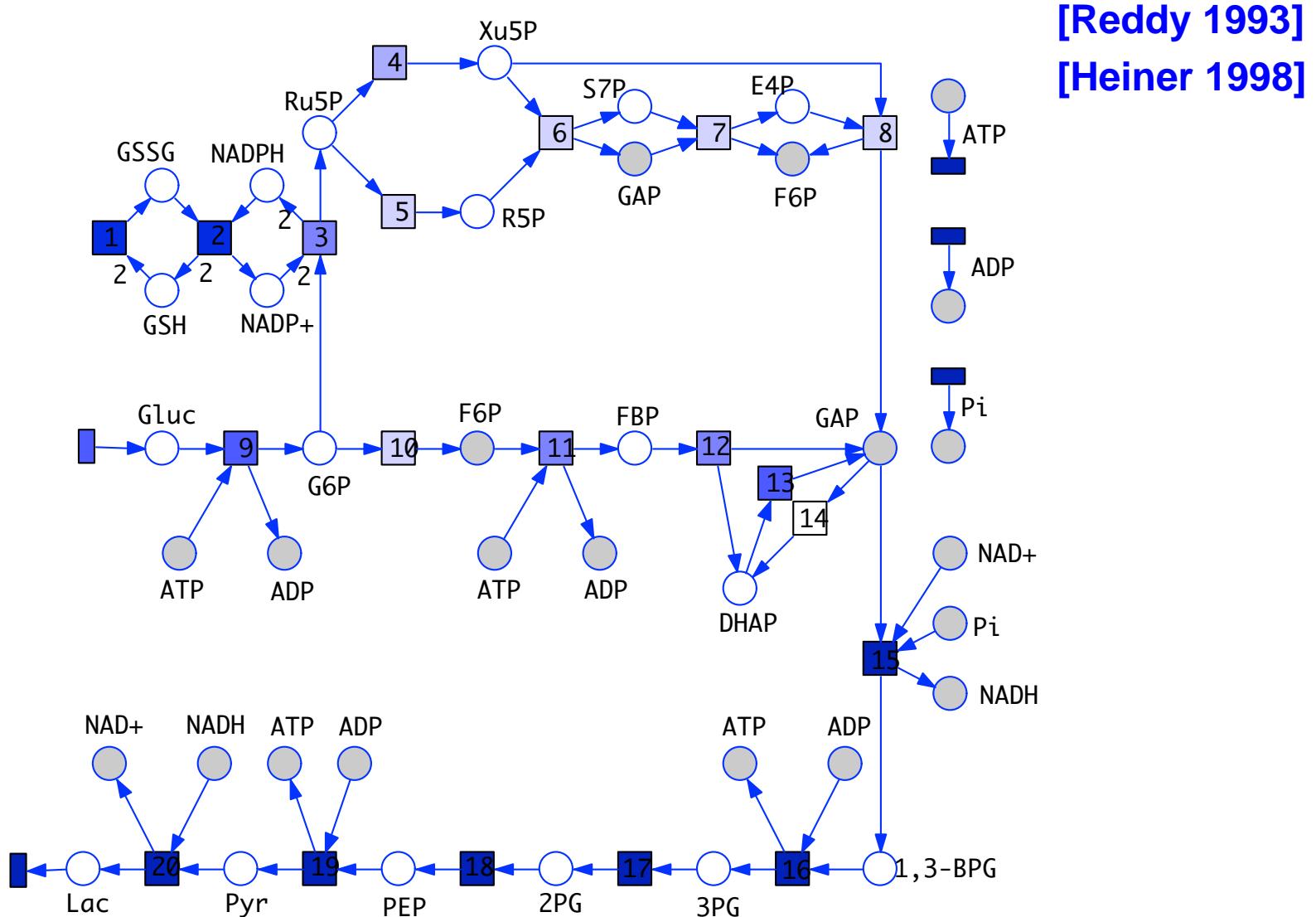
Ex - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



Ex - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



□ minimal T-invariants

- > connected, but *generally overlapping subnets*
- > computation: *NP-complete*
- > worst case: *exponentially many (in number of transitions)*

□ maximal abstract dependent transition sets (ADT sets)

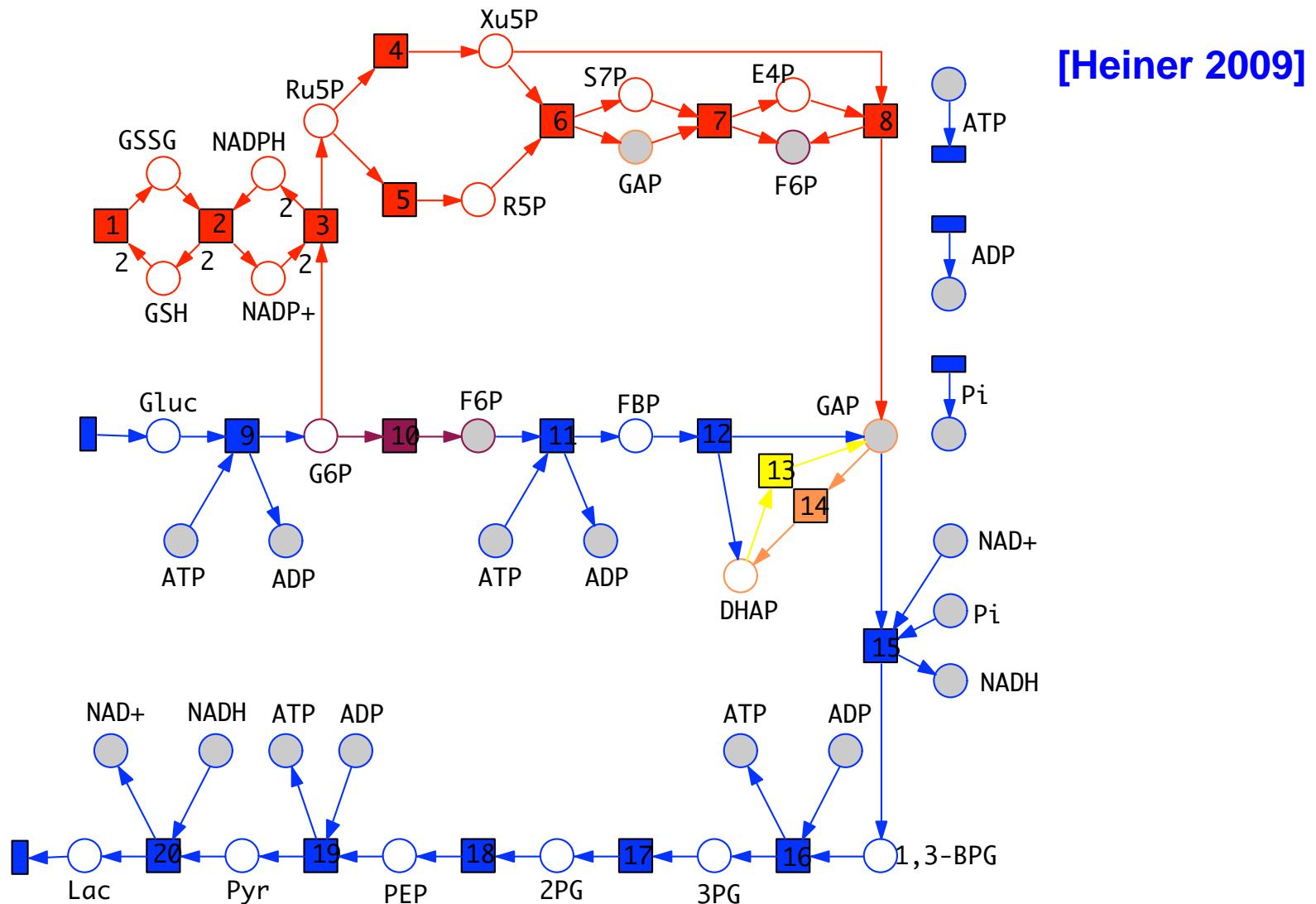
- > unique partitioning of the transition set
- > disjunctive, but *not necessarily connected subnets*
- > direct computation: polynomial - *DIANE, PETIA (lpsolve)*
- > worst case: linear (in number of transitions)

□ maximal connected ADT sets -> atomic fragments

- > disjunctive and connected subnets
- > coarse (hierarchical) network structure
 - ... better understanding, highlighting of crucial parts, ...
- > possibly according to primary compounds only,
i.e., neglecting connections by auxiliary compounds

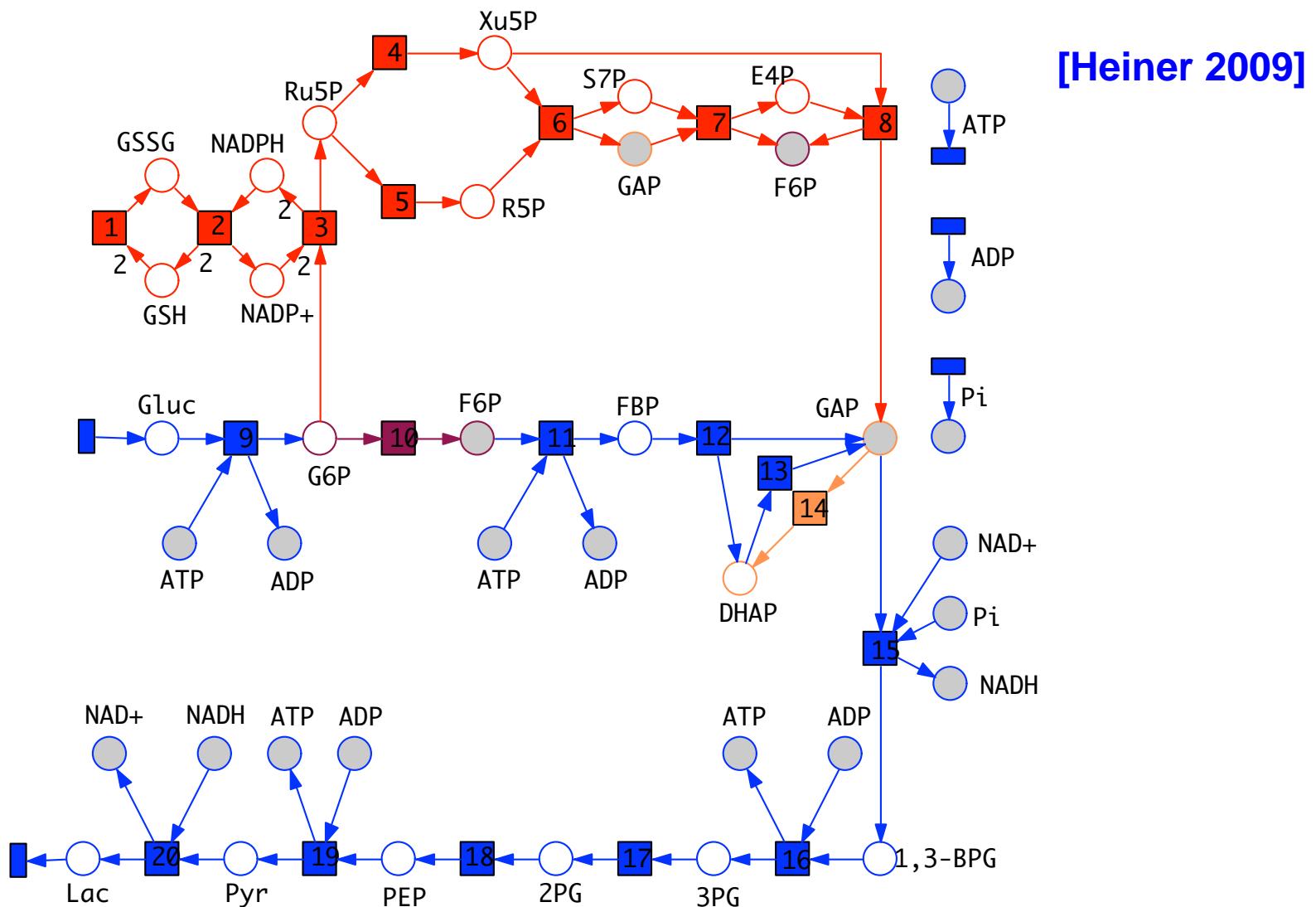
Ex1 - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



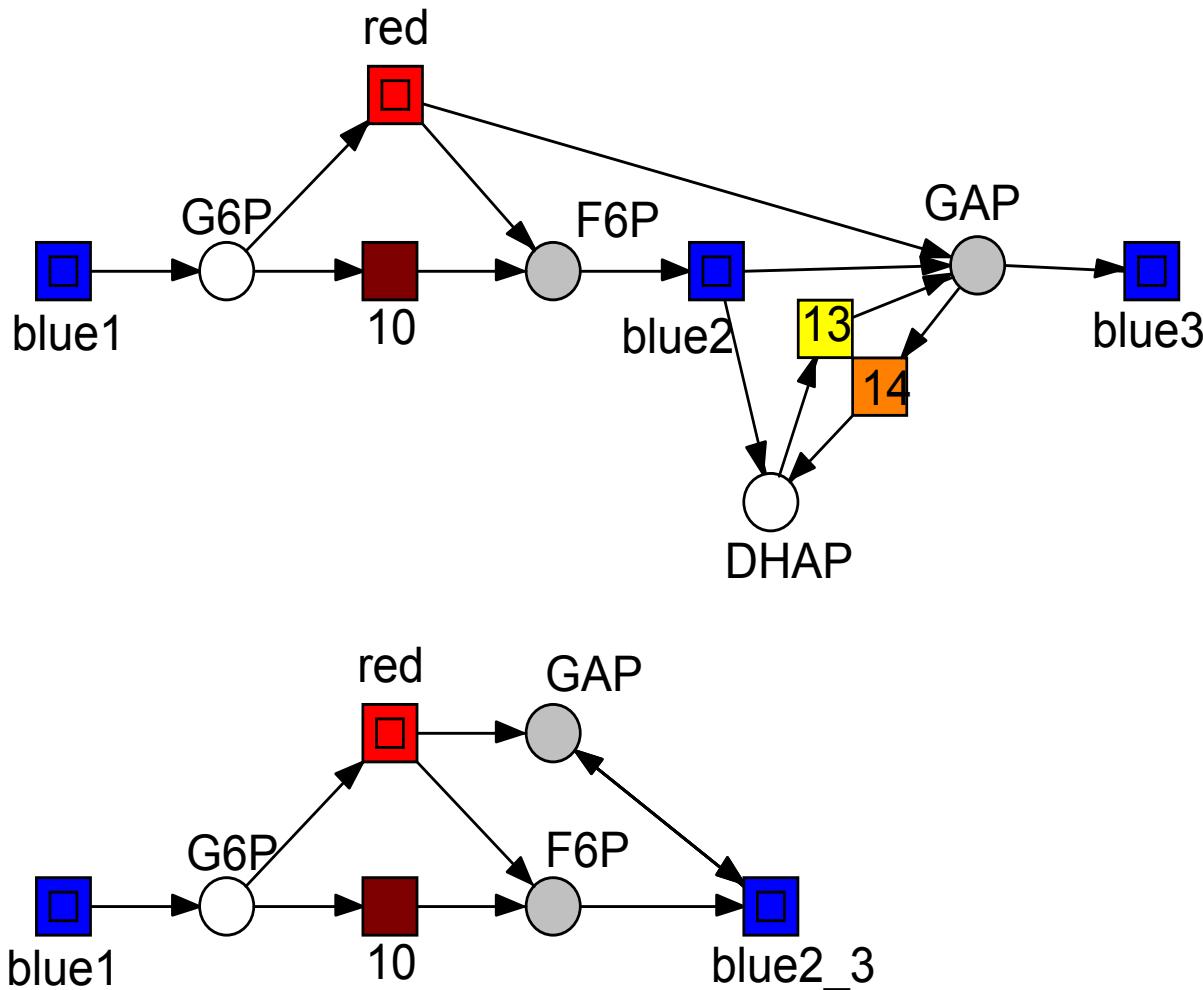
Ex1 - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering



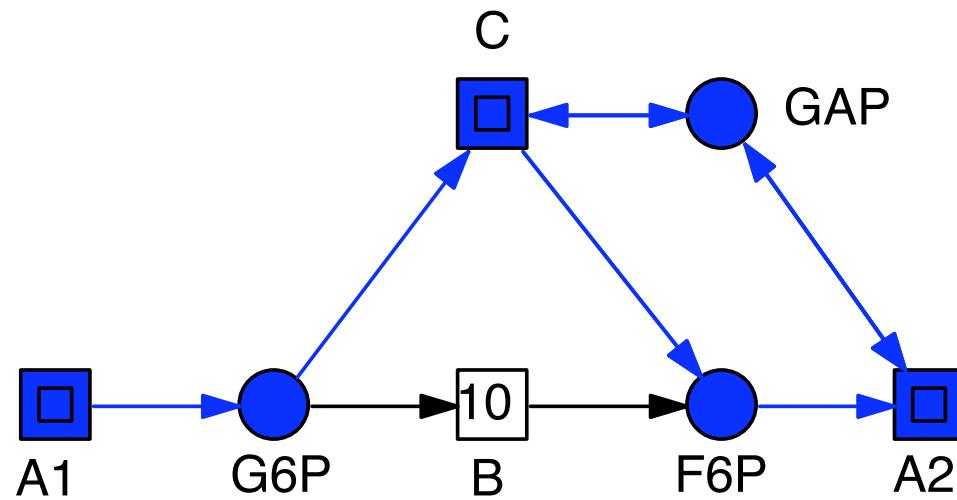
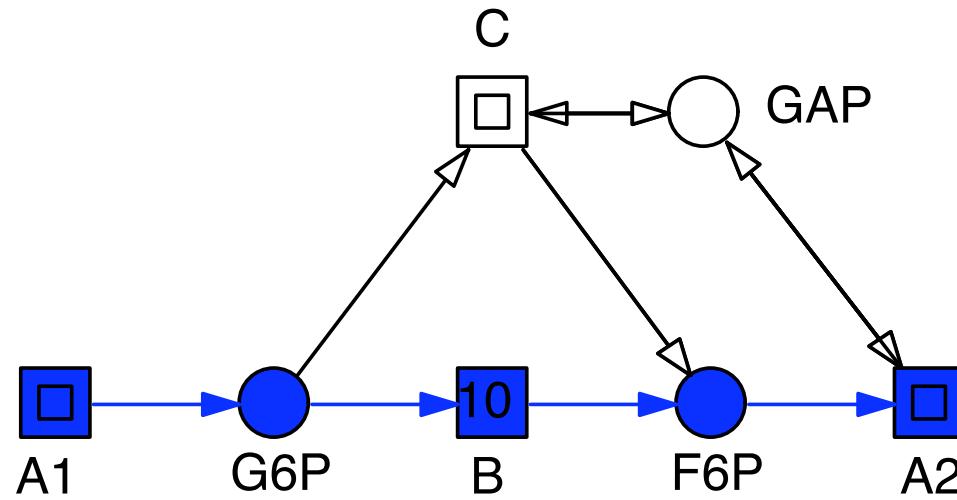
Ex1 - Glycolysis and Pentose Phosphate Pathway

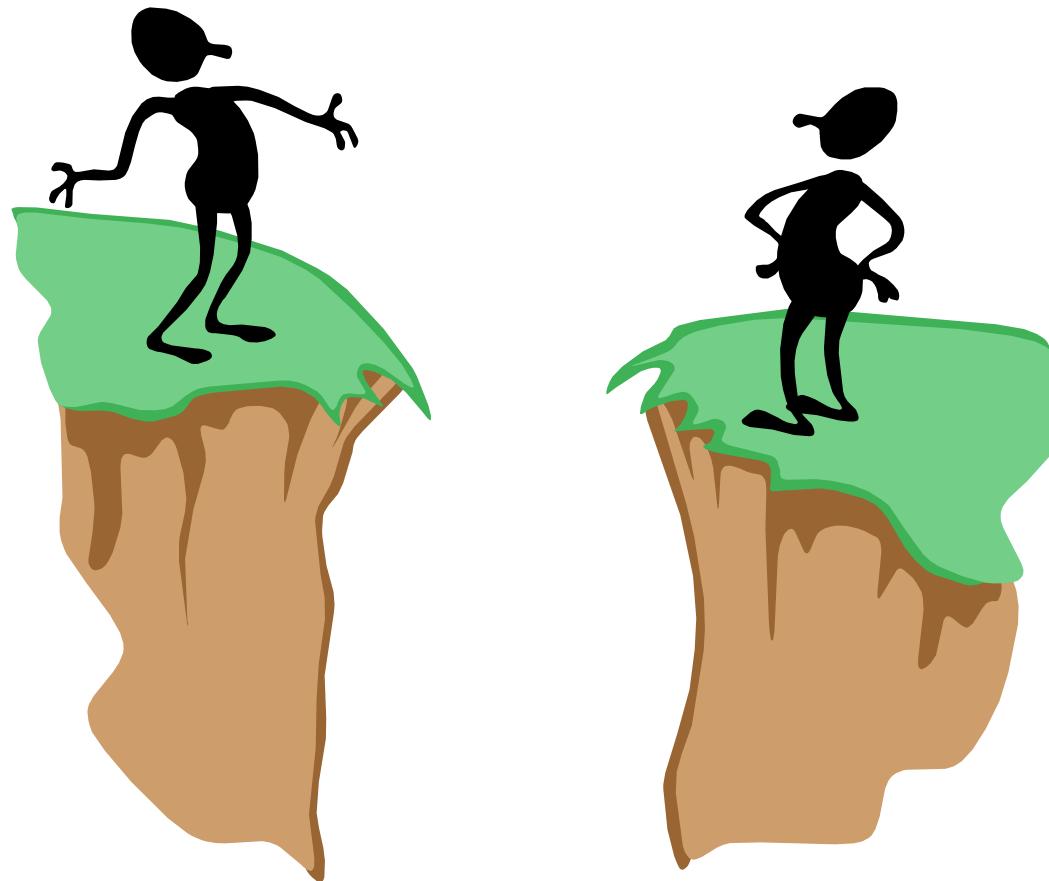
PN & Bacterial BioEngineering



Ex1 - Glycolysis and Pentose Phosphate Pathway

PN & Bacterial BioEngineering





... TO BE CONTINUED