PETRI NET BASED MODEL VALIDATION IN SYSTEMS BIOLOGY

Monika Heiner

Brandenburg University of Technology Cottbus

Dep. of CS

Ina Koch

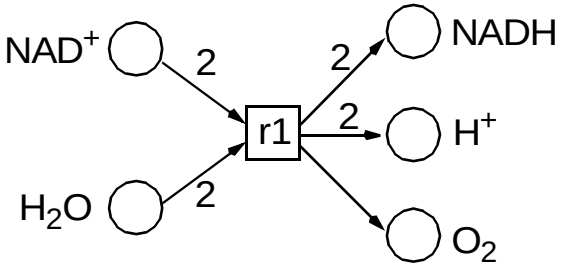
Technical University of Applied Sciences Berlin

Dep. of Bioinformatics

- **chemical reactions -> atomic actions -> Petri net transitions**

$$2\ NAD^+ + 2\ H_2O \rightarrow 2\ NADH + 2\ H^+ + O_2$$

input compounds

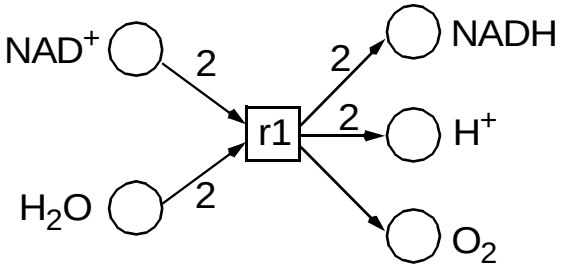


output compounds

❑ **chemical reactions -> atomic actions -> Petri net transitions**

$$2\ NAD^+ + 2\ H_2O \rightarrow 2\ NADH + 2\ H^+ + O_2$$

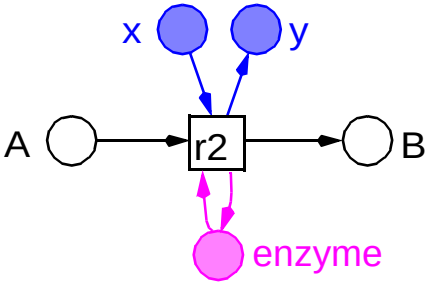
input compounds



output compounds

❑ **chemical compounds -> Petri net places**

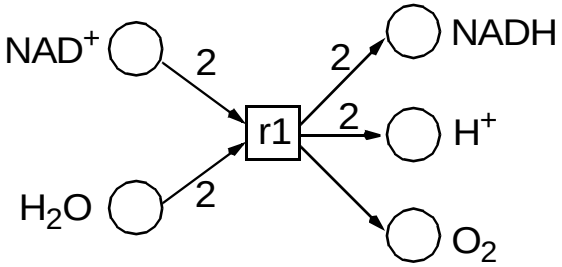
- *primary compounds* - *metabolites*
- *auxiliary compounds, ubiquitous -> fusion nodes* - *e. g. electron carrier*
- *catalyzing compounds* - *enzymes*



- **chemical reactions -> atomic actions -> Petri net transitions**

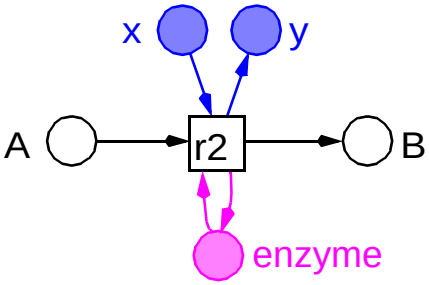
$$2\ NAD^+ + 2\ H_2O \rightarrow 2\ NADH + 2\ H^+ + O_2$$

input compounds



output compounds

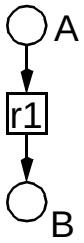
- **chemical compounds -> Petri net places**
  - *primary compounds* - *metabolites*
  - *auxiliary compounds, ubiquitous -> fusion nodes* - *e. g. electron carrier*
  - *catalyzing compounds* - *enzymes*



- **stoichiometric relations -> Petri net arc multiplicities**

- **compounds distribution -> marking**

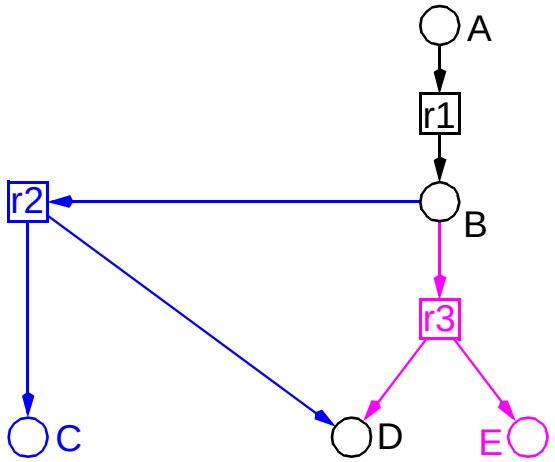
r1: A -> B



r1: A -> B

r2: B -> C + D

r3: B -> D + E



-> alternative reactions

r1: A -> B

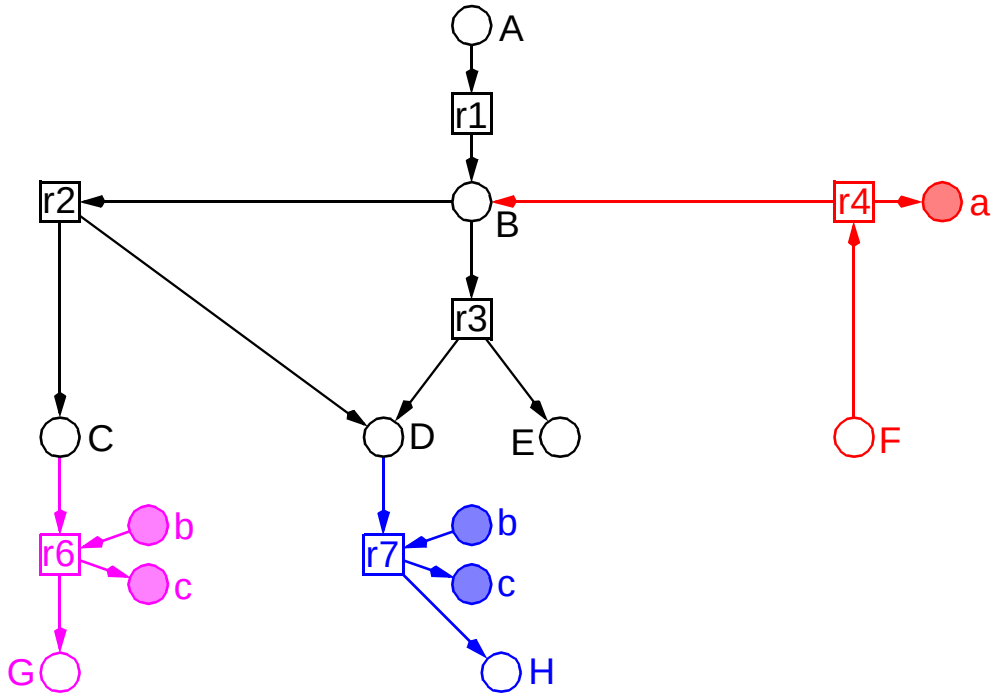
r2: B -> C + D

r3: B -> D + E

r4: F -> B + a

r6: C + b -> G + c

r7: D + b -> H + c



-> concurrent reactions

r1: A -> B

r2: B -> C + D

r3: B -> D + E

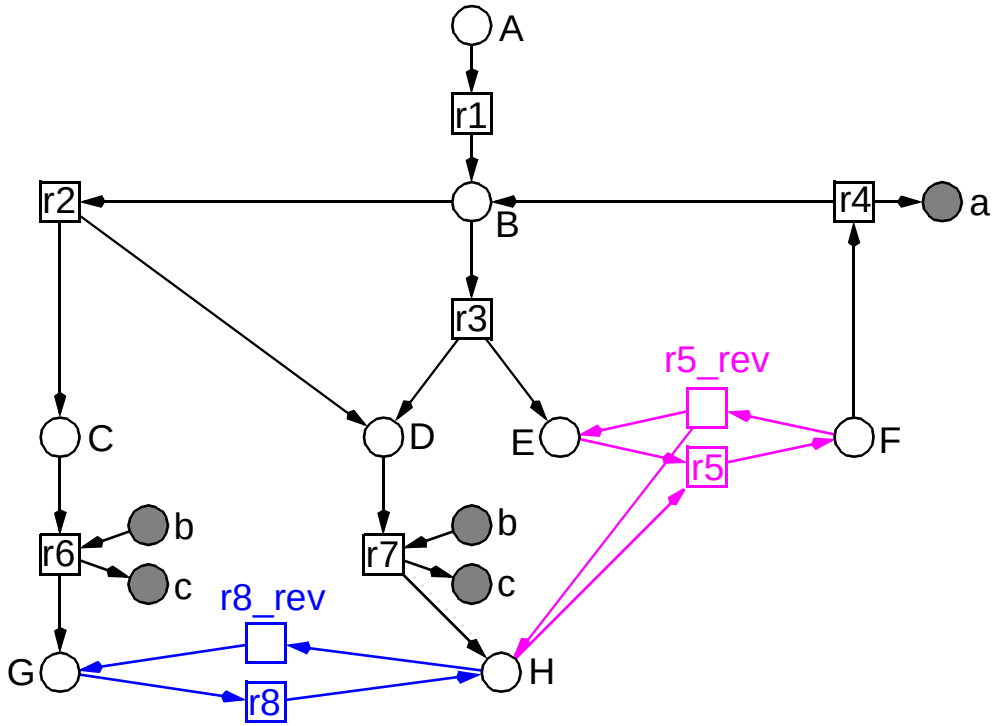
r4: F -> B + a

r5: E + H <-> F

r6: C + b -> G + c

r7: D + b -> H + c

r8: H <-> G



-> reversible reactions

r1: A -> B

r2: B -> C + D

r3: B -> D + E

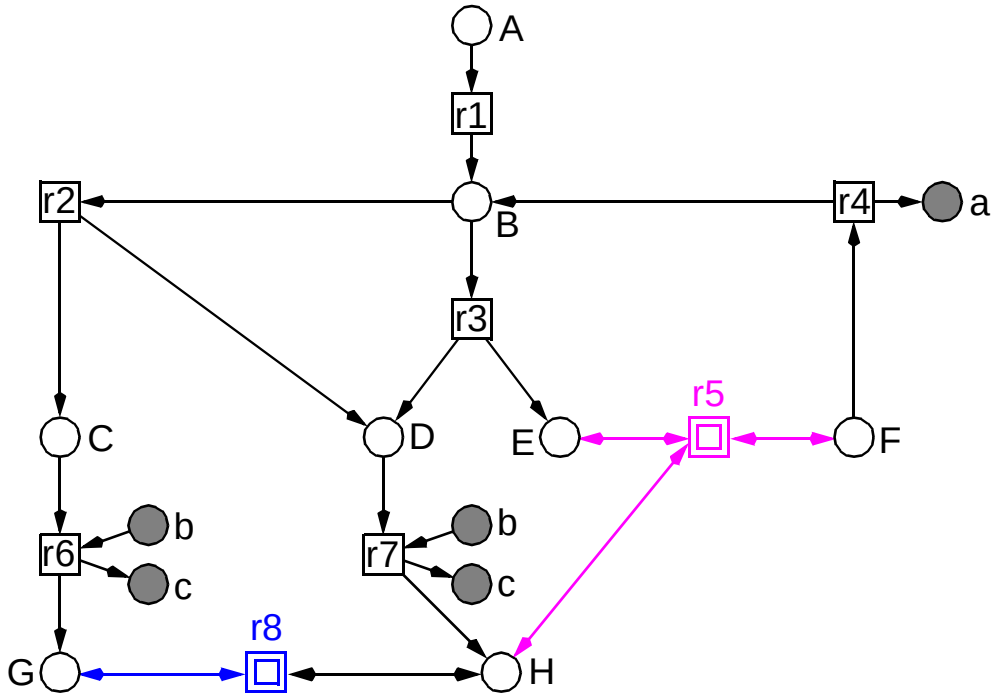
r4: F -> B + a

r5: E + H <-> F

r6: C + b -> G + c

r7: D + b -> H + c

r8: H <-> G



-> reversible reactions

- hierarchical nodes

r1: A -> B

r2: B -> C + D

r3: B -> D + E

r4: F -> B + a

r5: E + H <-> F

r6: C + b -> G + c

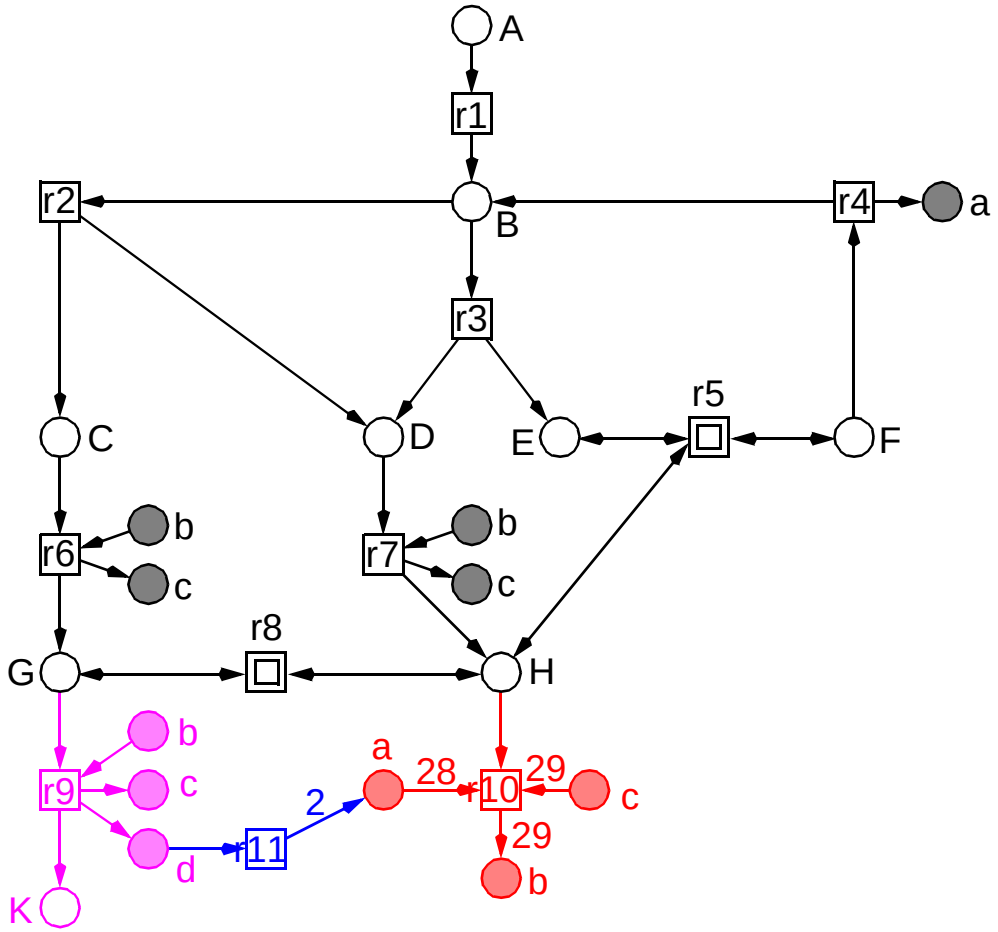
r7: D + b -> H + c

r8: H <-> G

r9: G + b -> K + c + d

r10: H + 28a + 29c -> 29b

r11: d -> 2a



r1: A -> B

r2: B -> C + D

r3: B -> D + E

r4: F -> B + a

r5: E + H <-> F

r6: C + b -> G + c

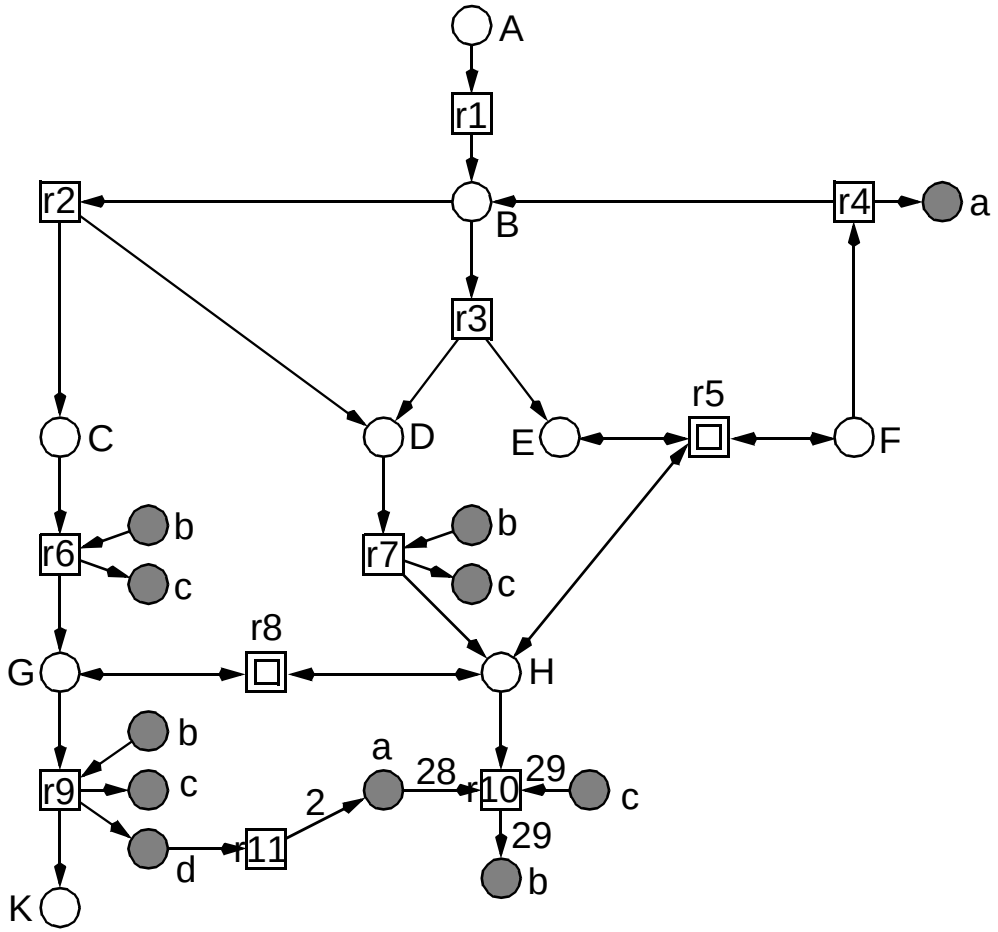
r7: D + b -> H + c

r8: H <-> G

r9: G + b -> K + c + d

r10: H + 28a + 29c -> 29b

r11: d -> 2a



r1: A -> B

r2: B -> C + D

r3: B -> D + E

r4: F -> B + a

r5: E + H <-> F

r6: C + b -> G + c

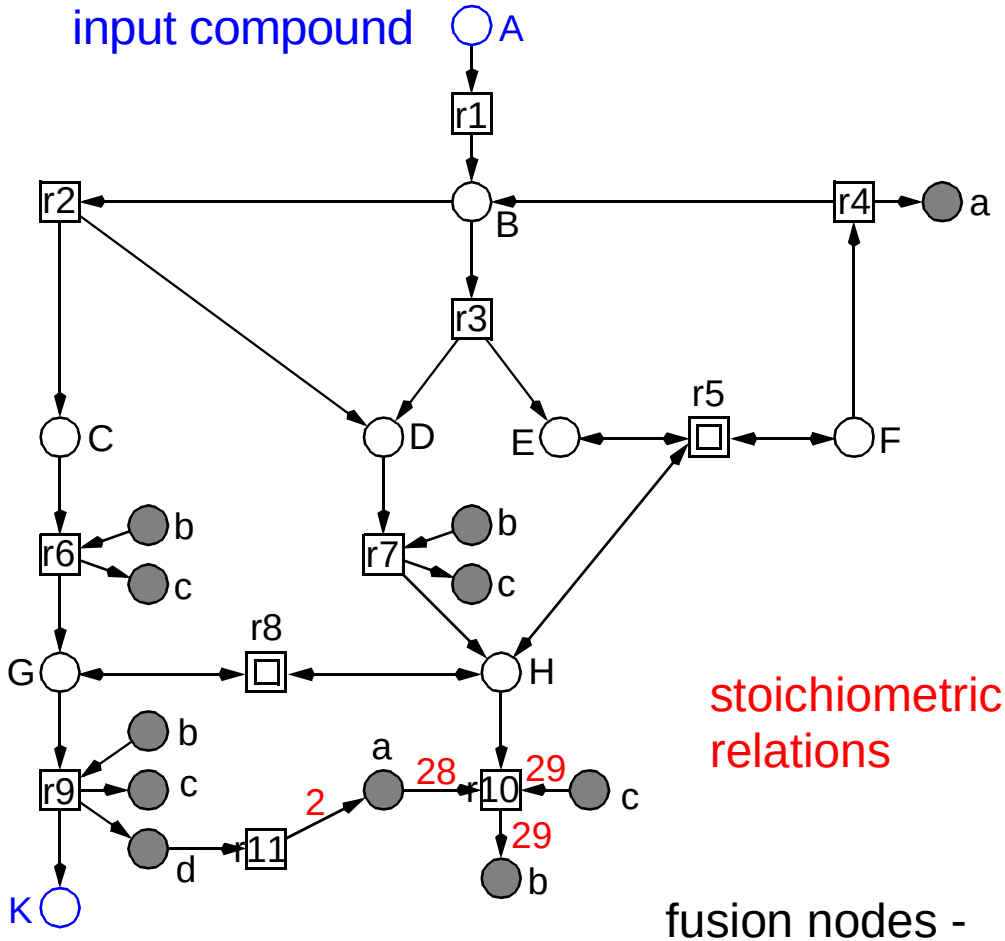
r7: D + b -> H + c

r8: H <-> G

r9: G + b -> K + c + d

r10: H + 28a + 29c -> 29b

r11: d -> 2a



- **networks of chemical reactions**

- **biologically interpreted Petri net**

  -> *partial order sequences of chemical reactions*
  - *transforming input into output compounds*
  - *respecting the given stoichiometric relations*

- **networks of chemical reactions**

- **biologically interpreted Petri net**

  -> *partial order sequences of chemical reactions*
  - *transforming input into output compounds*
  - *respecting the given stoichiometric relations*

- **typical (structural) properties**

```
INA
ORD HOM NBM PUR CSV SCF CON  SC  Ft0 tF0 Fp0 pF0 MG  SM FC  EFC ES
 N   N   N   Y   N   N   Y   N   N   N   Y   Y   N   N   N   N   N
DTP CPI CTI  B  SB  REV DSt BSt DTr DCF  L   LV L&S
 N   N   N   Y   Y   ?   ?   ?   ?   ?   N   ?   N
```

- **observation**

  -> *tend to grow fast*

- **networks of chemical reactions**

- **biologically interpreted Petri net**

  -> *partial order sequences of chemical reactions*
  - *transforming input into output compounds*
  - *respecting the given stoichiometric relations*

- **typical (structural) properties**

INA																		
ORD	HOM	NBM	PUR	CSV	SCF	CON	SC	Ft0	tF0	Fp0	pF0	MG	SM	FC	EFC	ES		
N	N	N	Y	N	N	Y	N	N	N	Y	Y	N	N	N	N	N		
DTP	CPI	CTI	B	SB	REV	DSt	BSt	DTr	DCF	L	LV	L&S						
N	N	N	Y	Y	?	?	?	?	?	N	?	N						

- **observation**

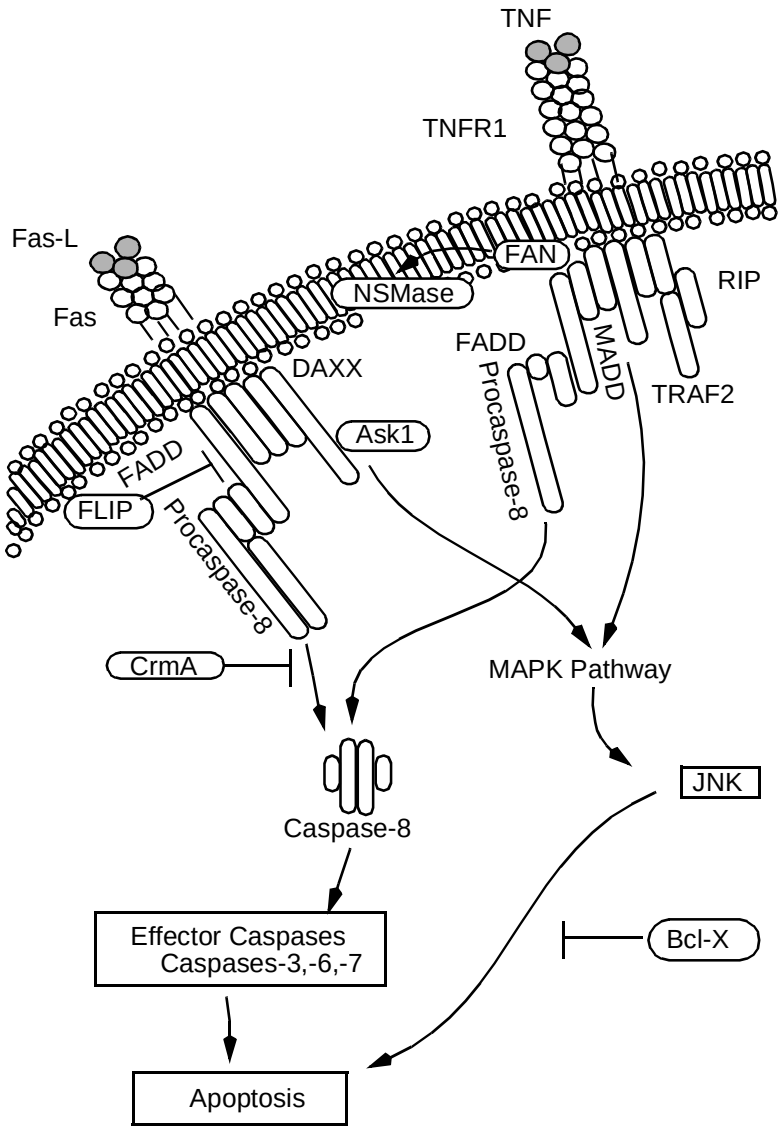
  -> *tend to grow fast*

BIONETWORKS, SOME PROBLEMS

- **network structure** -> ***PROBLEM 1***
  - -> *dense, apparently unstructured*
  - -> *hard to read*
- **knowledge** -> ***PROBLEM 2***
  - -> *uncertain*
  - -> *growing, changing*
  - -> *distributed over independent data bases, papers, journals, . . .*
- **various, mostly ambiguous representations** -> ***PROBLEM 3***
  - -> *verbose descriptions*
  - -> *diverse graphical representations*
  - -> *contradictory and / or fuzzy statements*

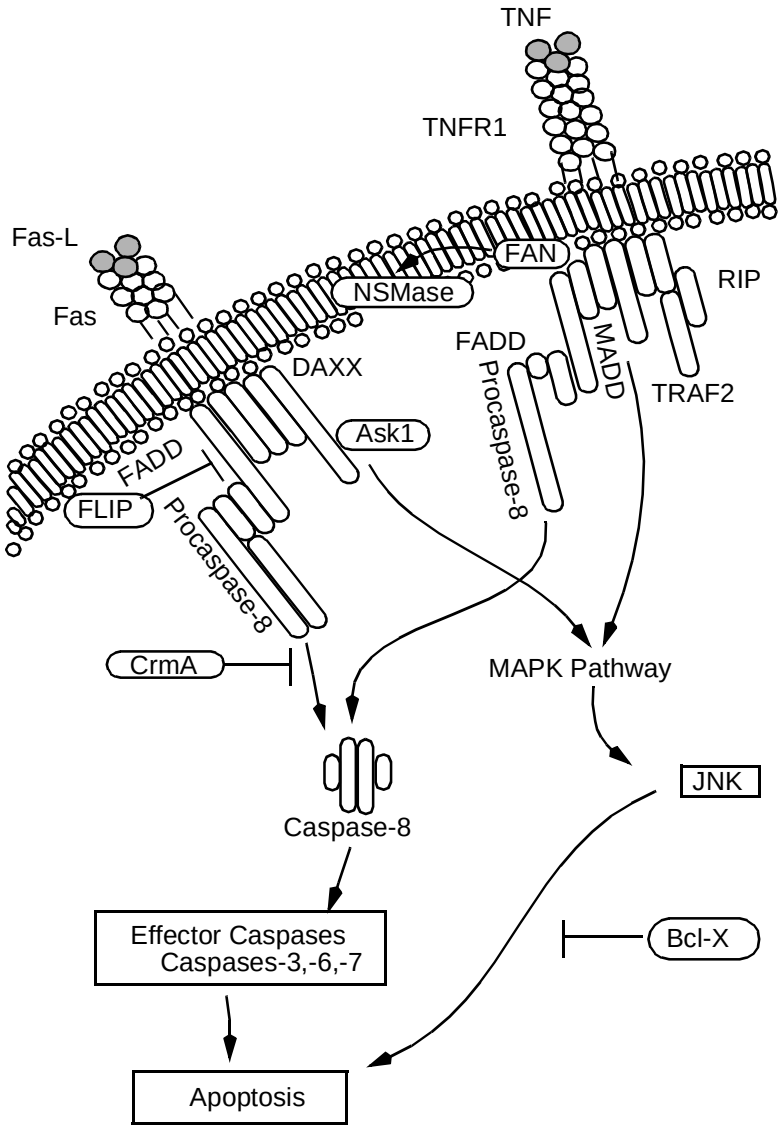
  -> *some examples*

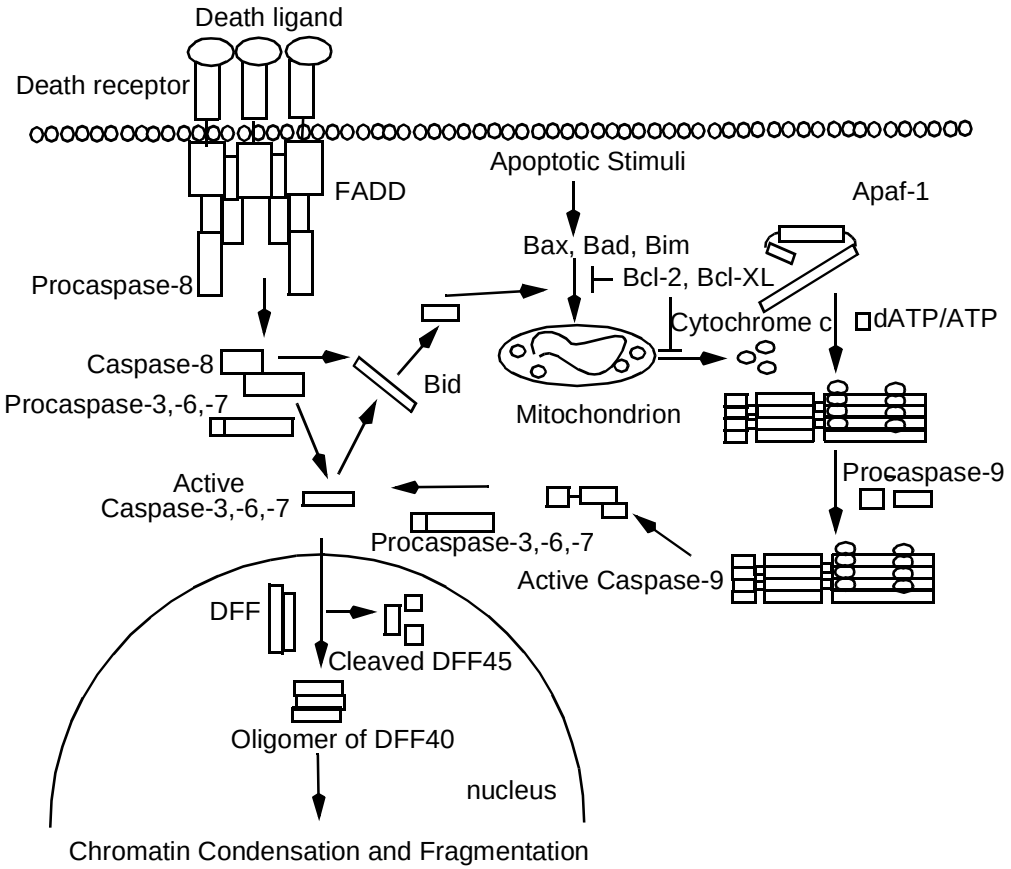
NETWORK REPRESENTATIONS, EX1



-> CASE STUDY 1

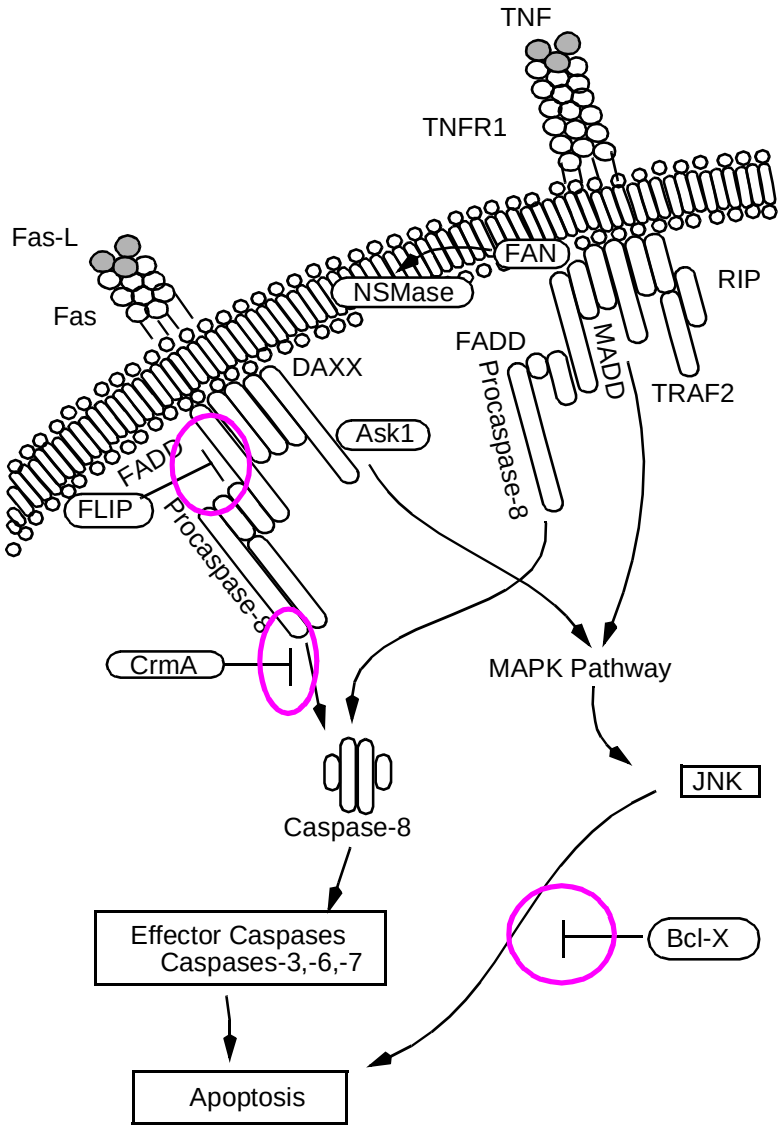
NETWORK REPRESENTATIONS, EX1 + EX2

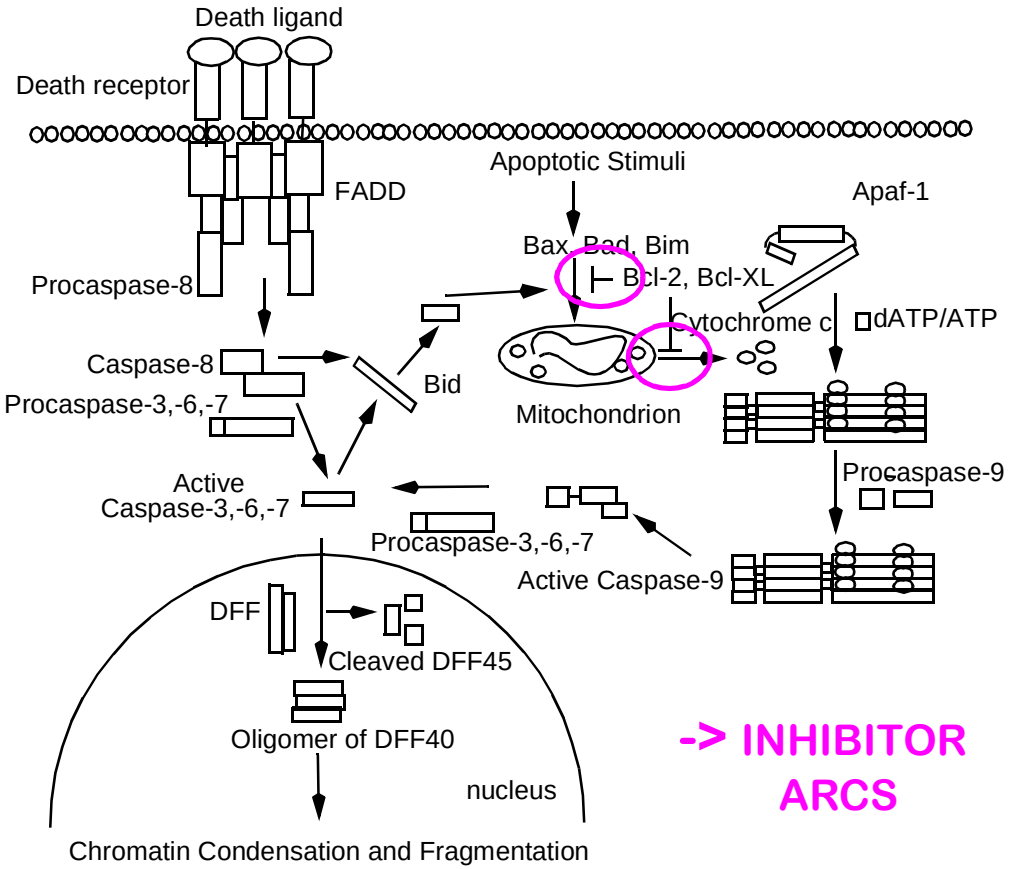




-> CASE STUDY 1

NETWORK REPRESENTATIONS, EX1 + EX2

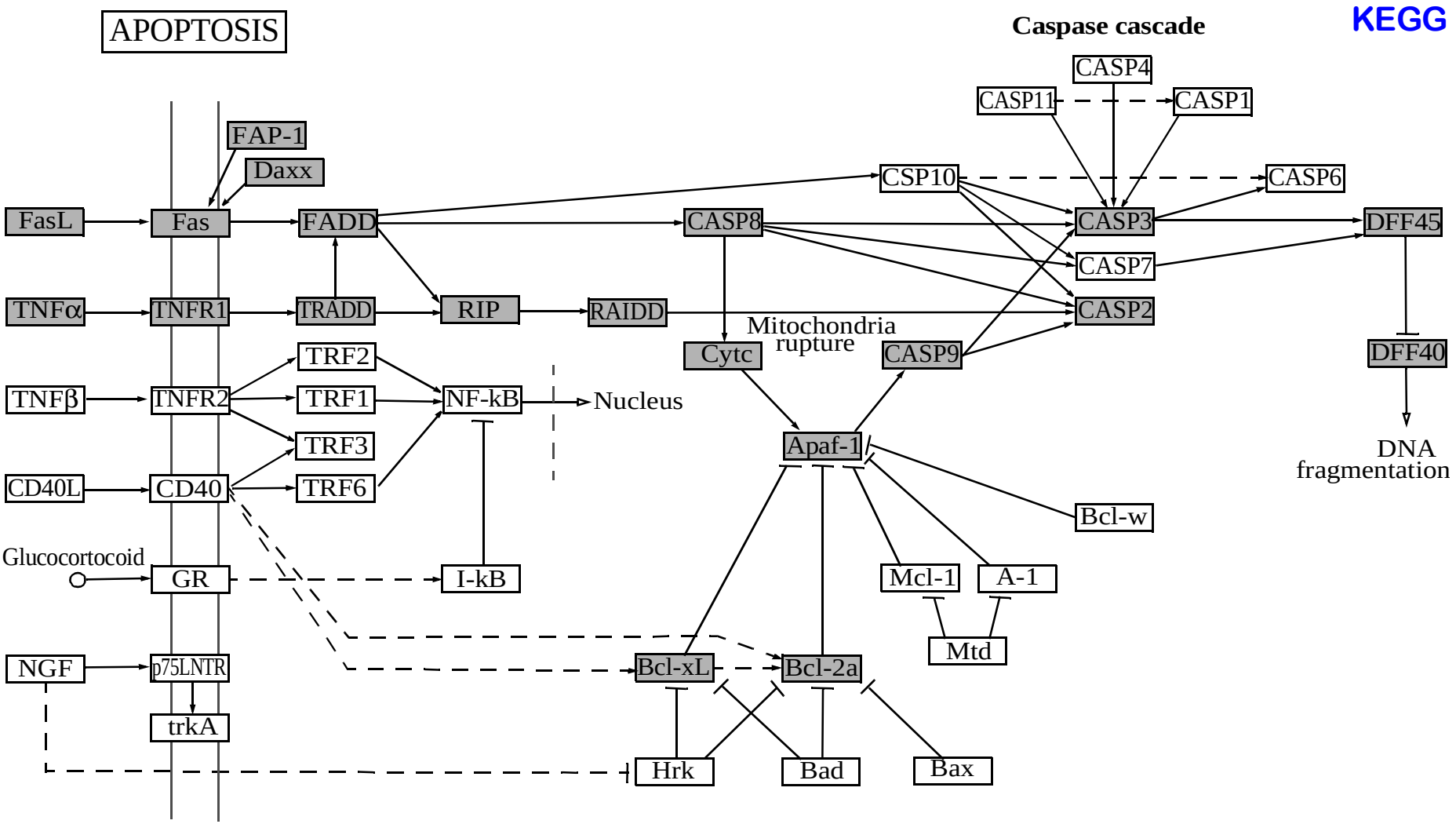




-> INHIBITOR ARCS

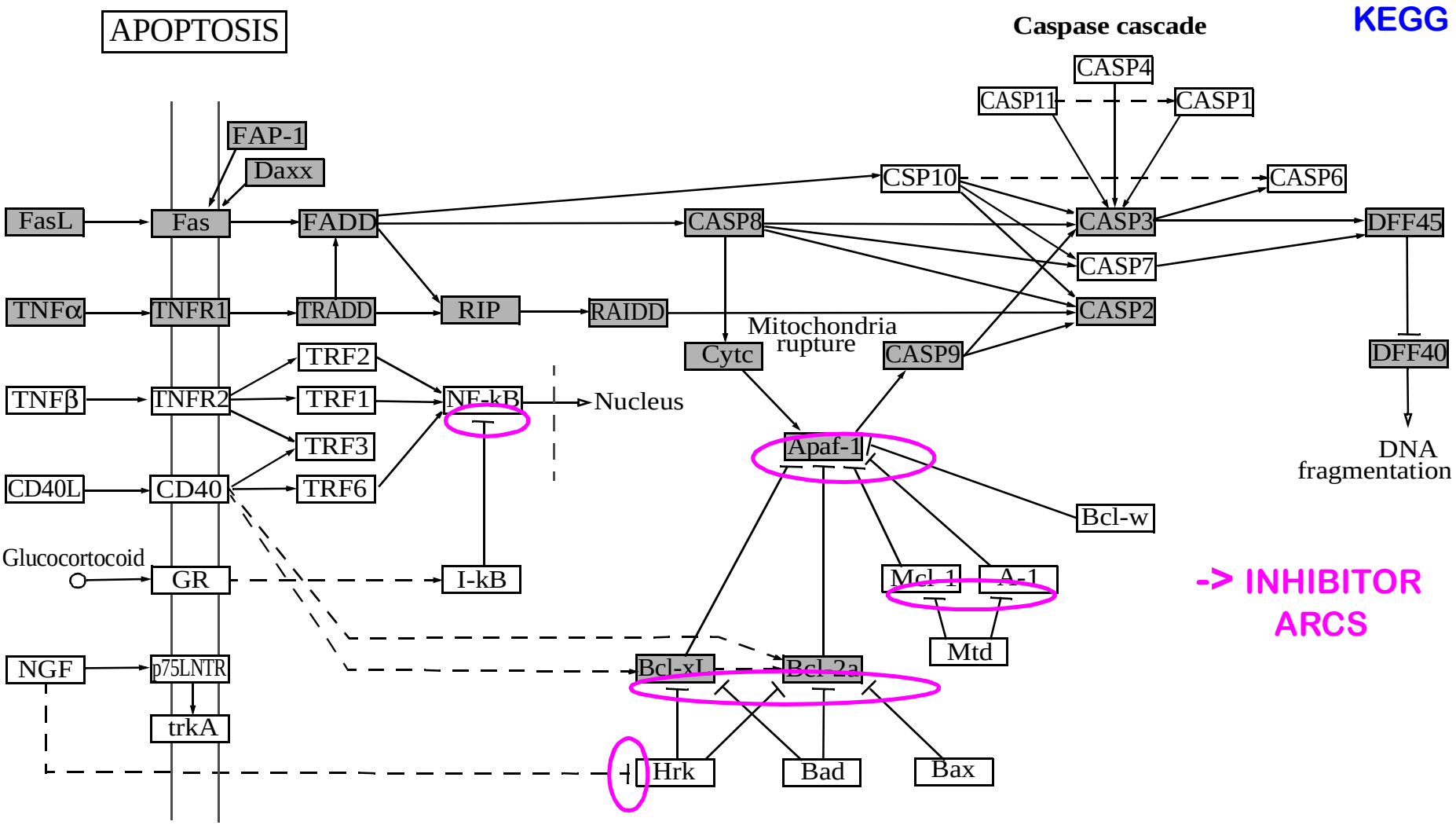
-> CASE STUDY 1

NETWORK REPRESENTATIONS, EX3

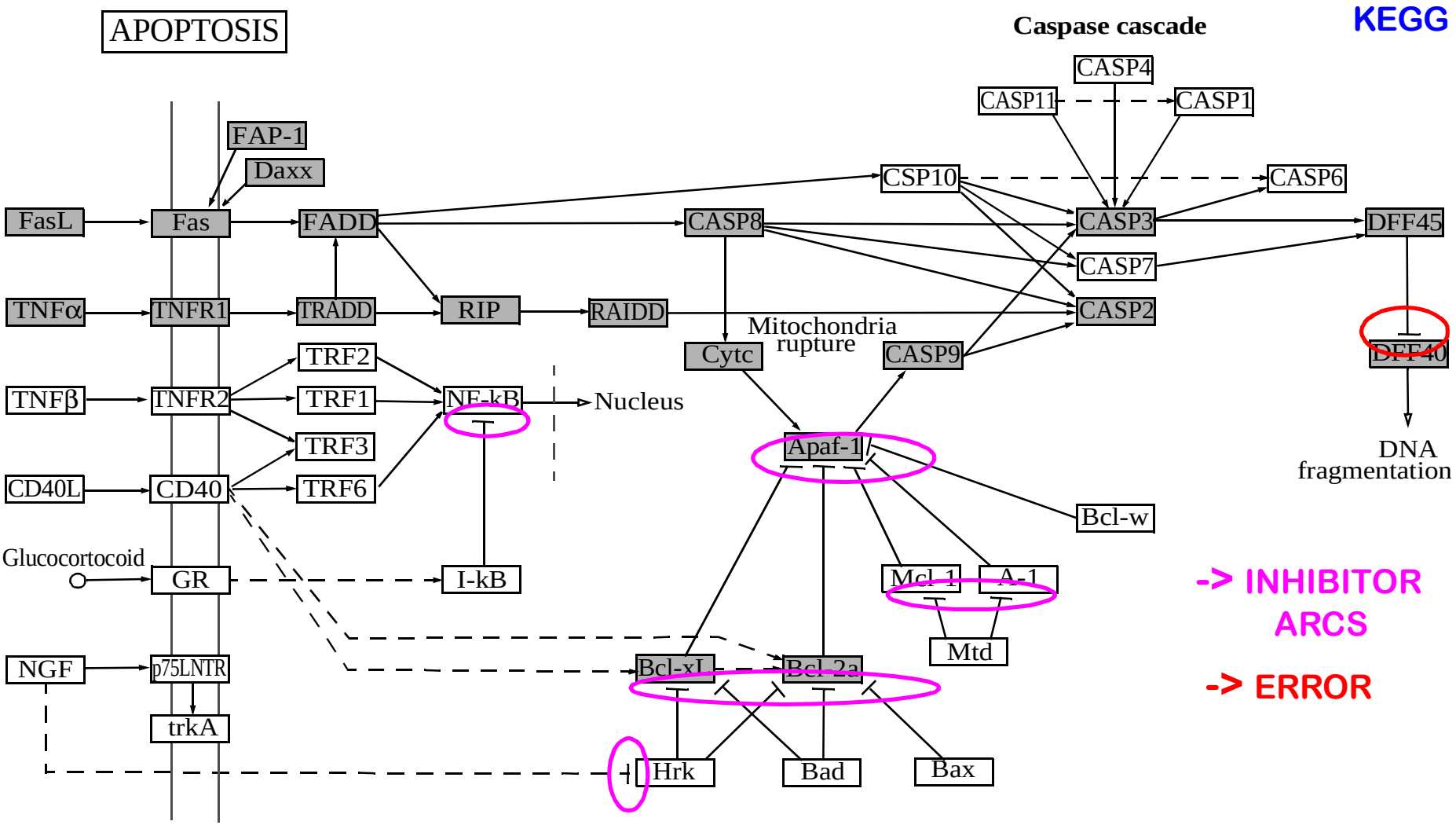


-> CASE STUDY 1

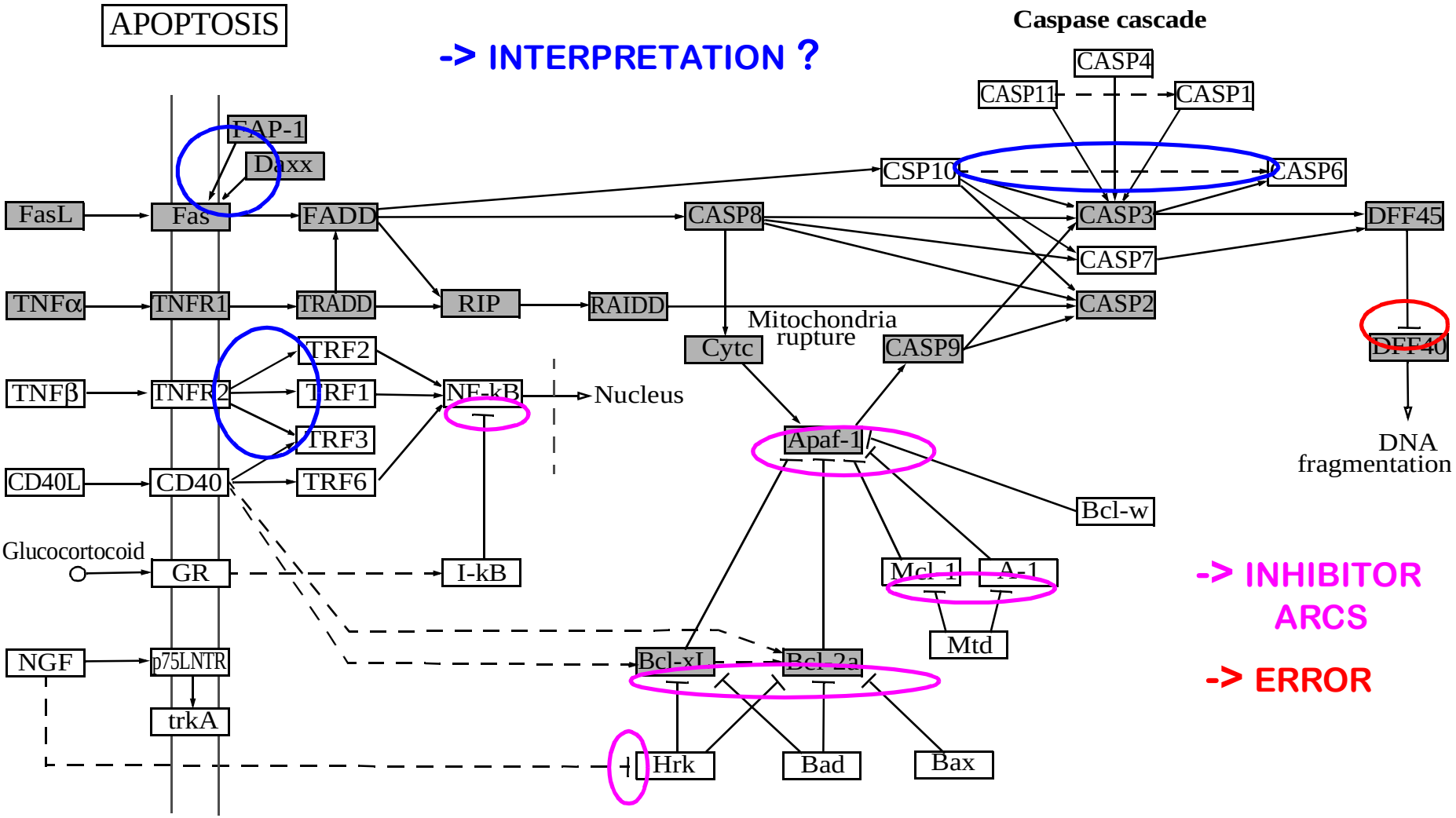
NETWORK REPRESENTATIONS, EX3



NETWORK REPRESENTATIONS, EX3

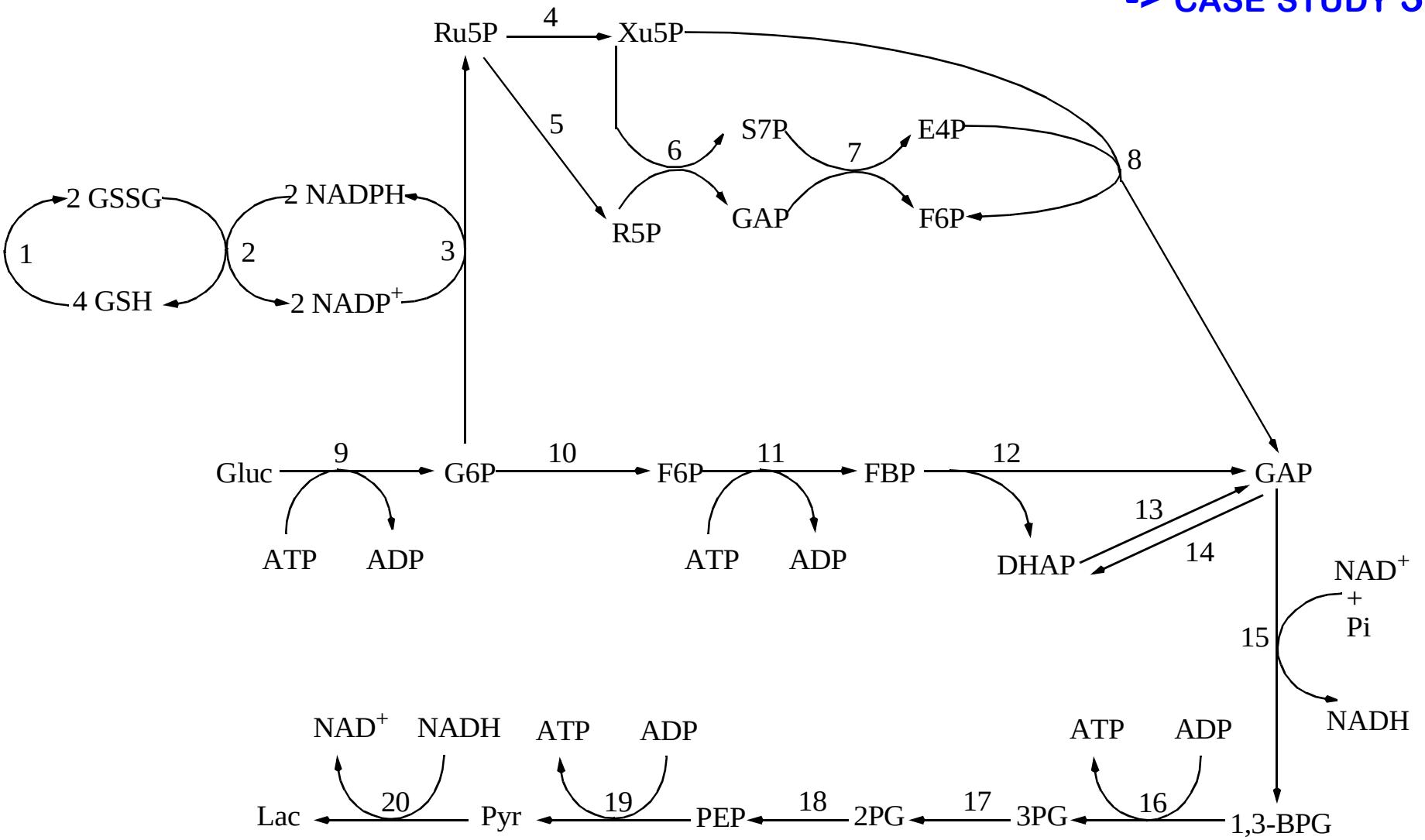


NETWORK REPRESENTATIONS, EX3



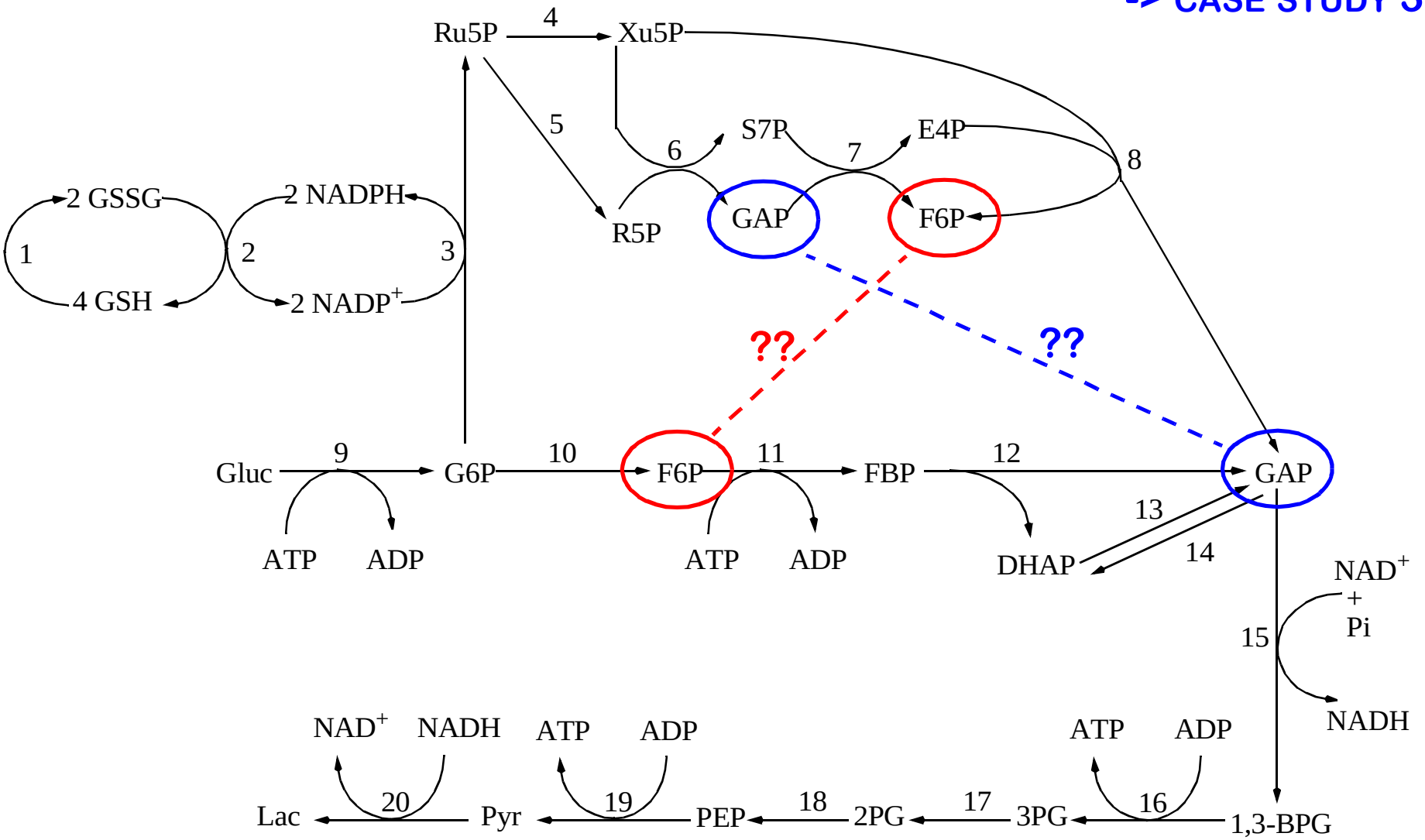
NETWORK REPRESENTATIONS, EX4

-> CASE STUDY 3



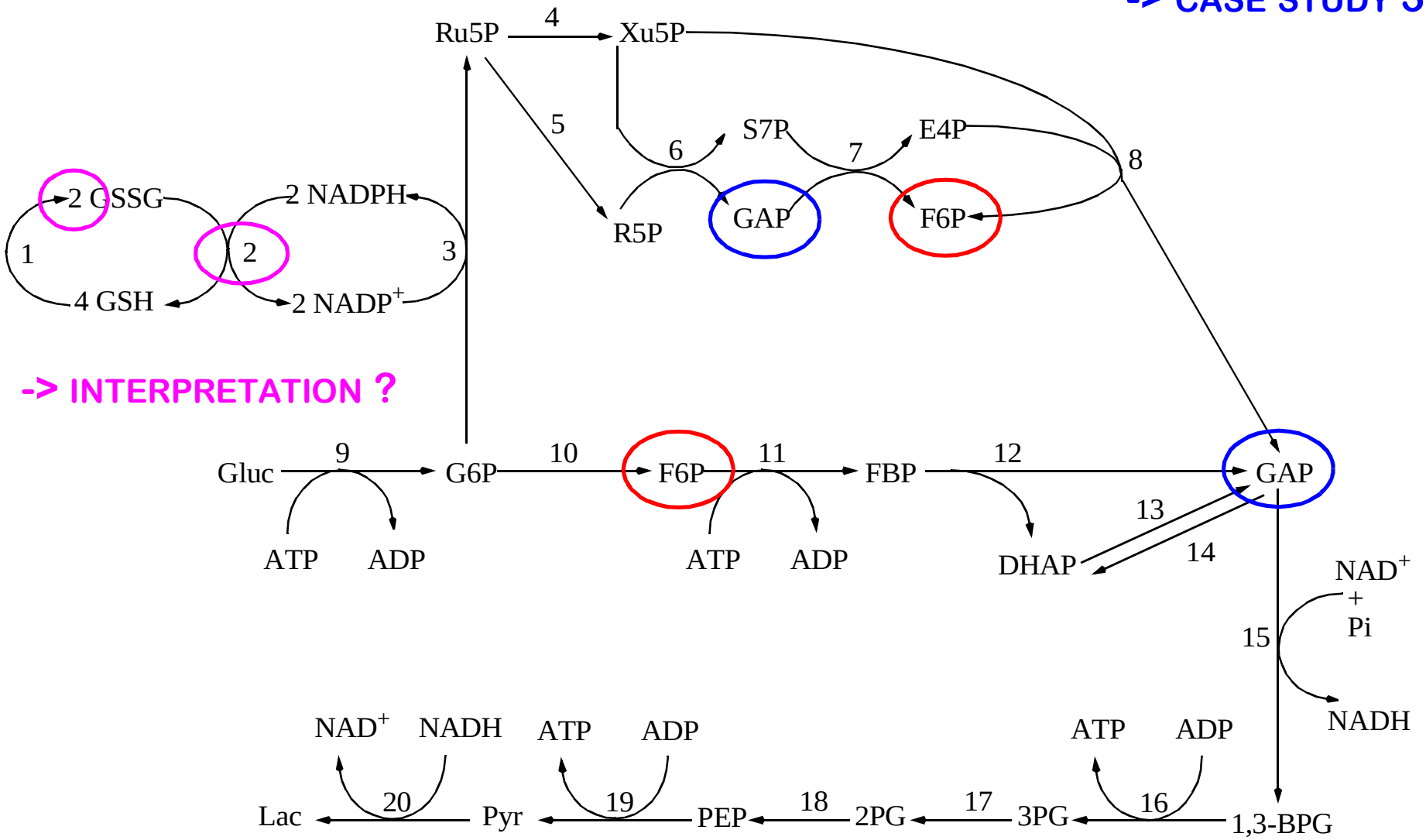
NETWORK REPRESENTATIONS, EX4

-> CASE STUDY 3



NETWORK REPRESENTATIONS, EX4

-> CASE STUDY 3



-> INTERPRETATION ?

BIONETWORKS, SOME PROBLEMS, CONSEQUENCES

- **patchwork**
  -> *likely to be erroneous*
- **long-term purpose**
  -> *(quantitative) behaviour prediction*
- **necessary intermediate step**
  -> *model validation* -> *validation criteria ?*

BIONETWORKS, SOME PROBLEMS, CONSEQUENCES

- **patchwork**
  -> *likely to be erroneous*
- **long-term purpose**
  -> *(quantitative) behaviour prediction*
- **necessary intermediate step**
  -> *model validation* -> *validation criteria ?*
- **steady state behaviour -> pathways**
  -> *all possible flows preserving the given compounds distribution*
  -> *elementary modes = minimal T-invariants*
- **consistency criteria -> pathways analysis**
  -> *CTI*
  -> *no minimal T-invariant without biological interpretation*
  -> *no known biological behaviour without corresponding T-invariant*

BIONETWORKS, SOME PROBLEMS, CONSEQUENCES

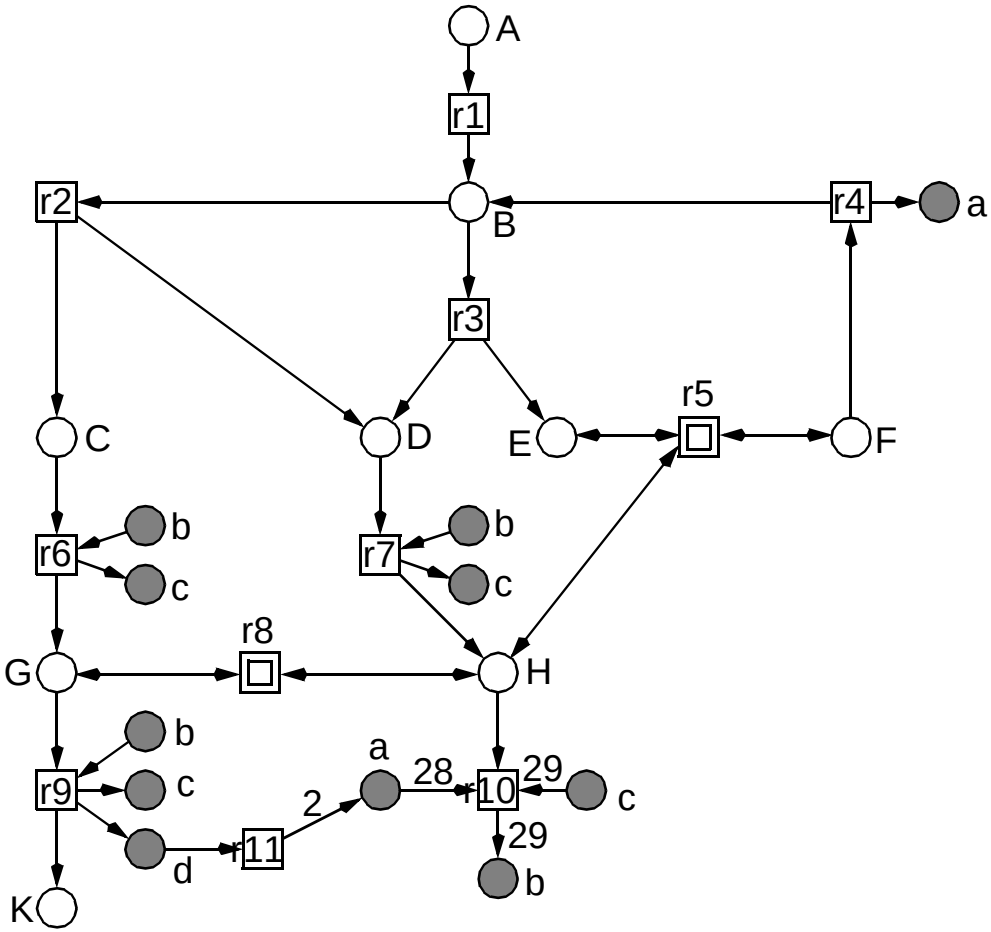
- **patchwork**
  -> *likely to be erroneous*
- **long-term purpose**
  -> *(quantitative) behaviour prediction*
- **necessary intermediate step**
  -> *model validation* -> *validation criteria ?*
- **steady state behaviour -> pathways**
  -> *all possible flows preserving the given compounds distribution*
  -> *elementary modes = minimal T-invariants*
- **consistency criteria -> pathways analysis**
  -> *CTI*
  -> *no minimal T-invariant without biological interpretation*
  -> *no known biological behaviour without corresponding T-invariant*

T-INVARIANTS, REFRESHMENT

- **Lautenbach, 1973**
- **T-invariants** -> *multisets of transitions*
  -> *integer solutions of* $Cx = 0, x \neq 0, x \geq 0$
- **minimal T-invariants**
  -> *there is no T-invariant with a smaller support* -> *sets of transitions*
  -> *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*
- **Covered by T-Invariants (CTI)**
  -> *each transition belongs to a T-invariant*

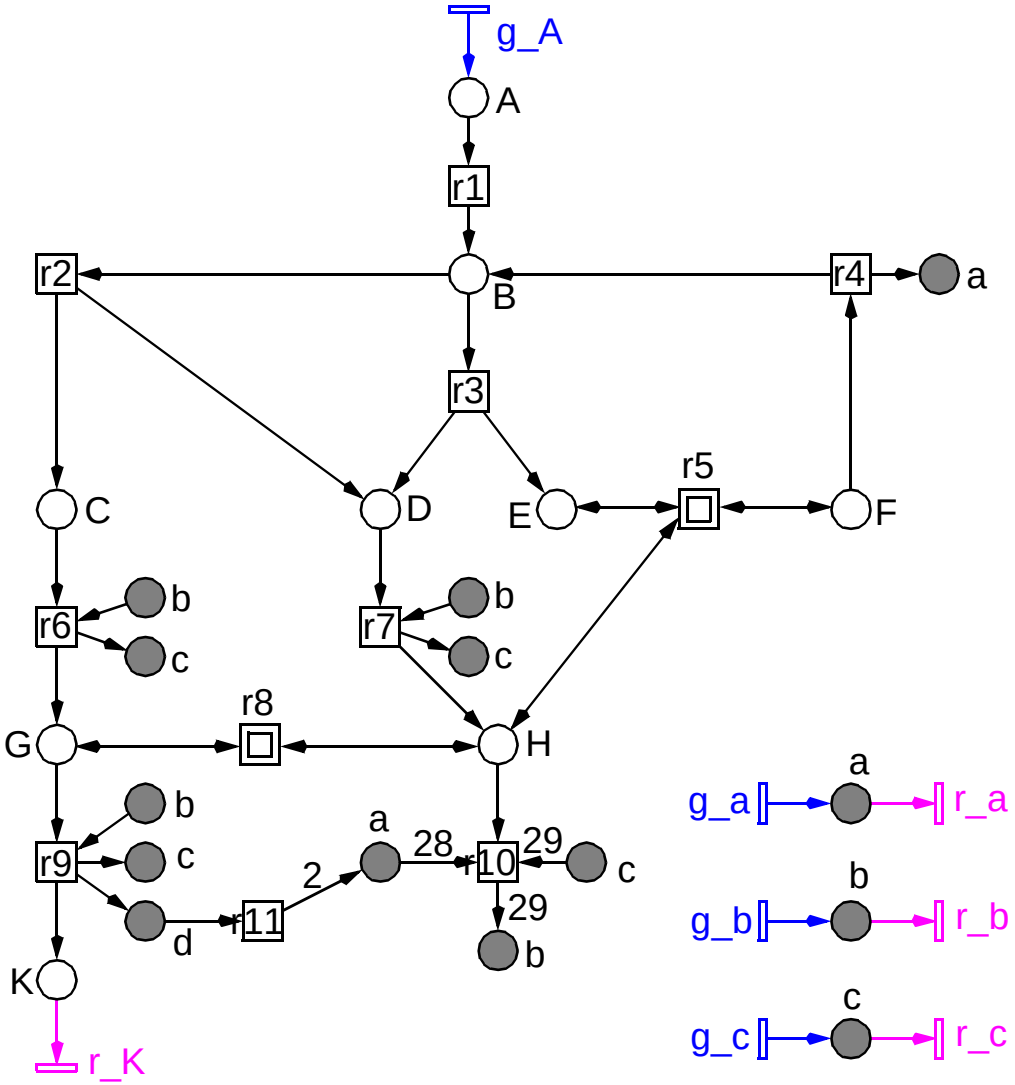
NETWORK NEEDS ENVIRONMENT BEHAVIOUR

- **to animate the model**
  - -> *infinite substance flow*
  - -> *deeper insights*
- **to validate the model**
  - -> *consistency criteria*
- **steady flow**
  - -> *input substances*
  - -> *output substances*
- **auxiliary substances**
  - -> *as much as necessary*
- **minimal assumptions**



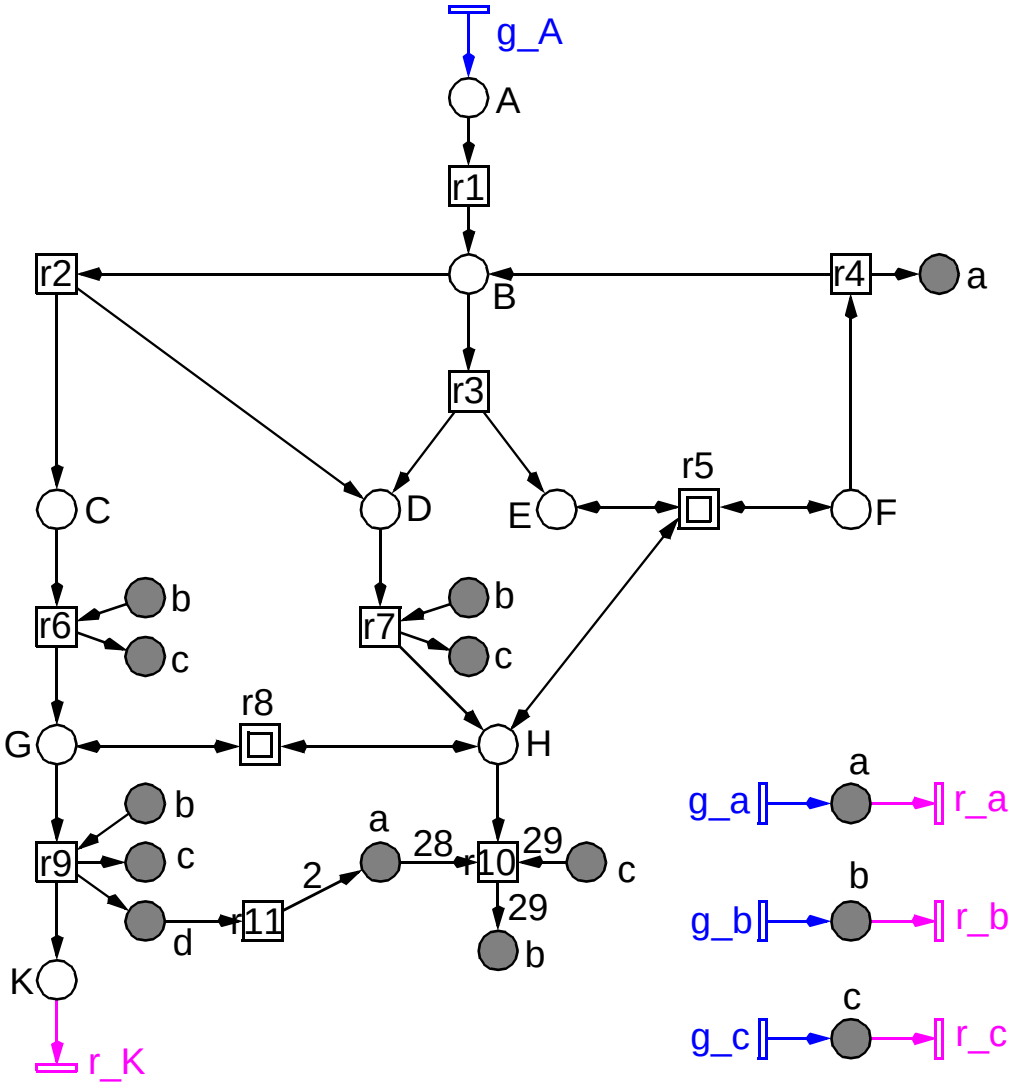
NETWORK NEEDS ENVIRONMENT BEHAVIOUR

- **to animate the model**
  - -> *infinite substance flow*
  - -> *deeper insights*
- **to validate the model**
  - -> *consistency criteria*
- **steady flow**
  - -> *input substances*
  - -> *output substances*
- **auxiliary substances**
  - -> *as much as necessary*
- **minimal assumptions**



NETWORK WITH ENVIRONMENT BEHAVIOUR

- **input substances**
  -> *generating pre-transitions*
- **output substances**
  -> *consuming post-transitions*
- **auxiliary substances**
  -> *both*
- **no boundary places, but boundary transitions**
- **transitions without pre-places**
  -> *live*
  -> *all post-places are unbounded*
- **steady state behaviour**
  -> *empty marking reproduction*



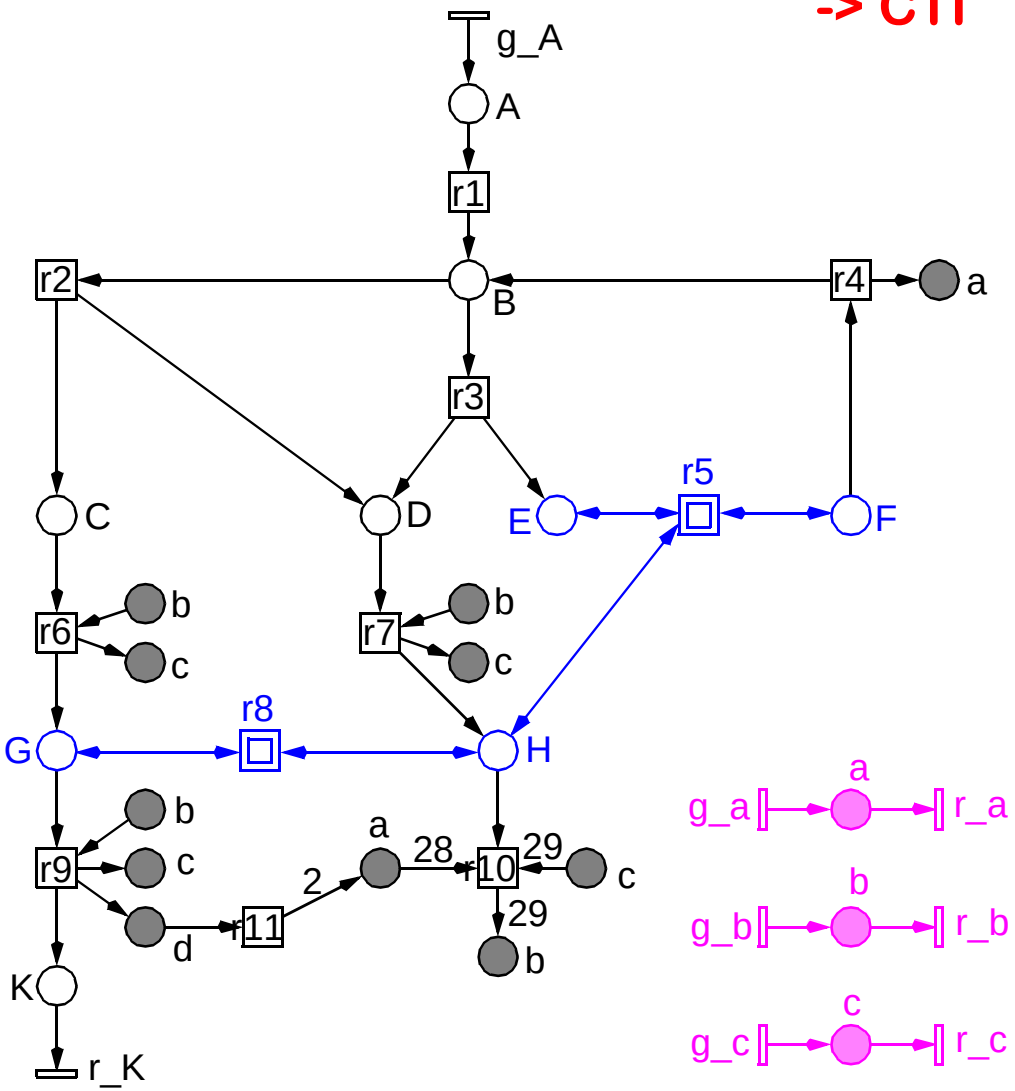
trivial min. T-invariants (5)

-> CTI

- boundary transitions of auxiliary compounds
  -> *(g_a, r_a), (g_b, r_b), (g_c, r_c)*
- reversible reactions
  -> *(r5, r5_rev), (r8, r8_rev)*

non-trivial min. T-invariants (7)

- covering boundary transitions of input / output compounds
  -> *i/o-T-invariants*
- inner cycles



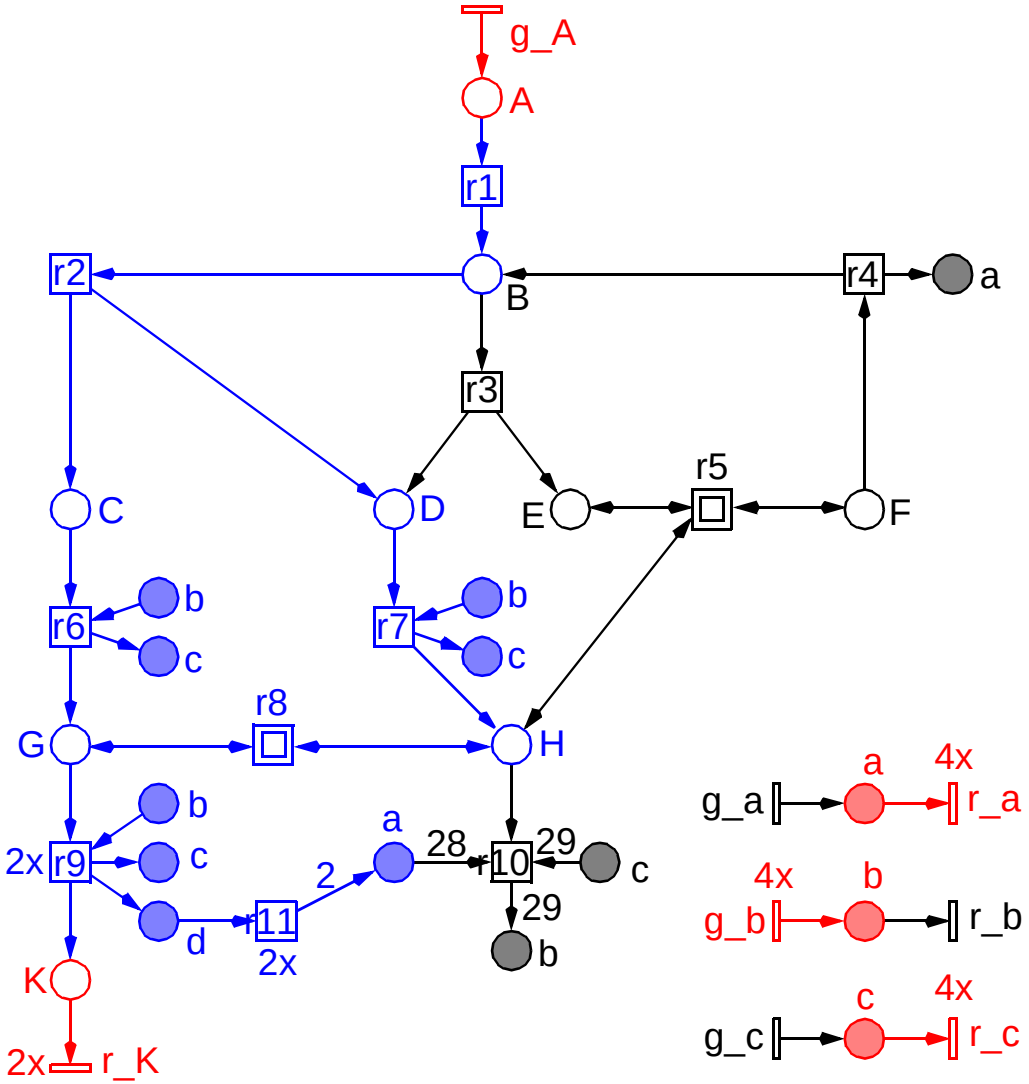
NETWORK, I/O-T-INVARIANT

- **i/o-T-invariant, example**

12 | *0.r1* : *1*
| *1.r2* : *1,*
| *3.r8_rev* : *1,*
| *4.r6* : *1,*
| *5.r7* : *1,*
| *9.r9* : *2,*
| *12.r11* : *2,*
| *13.g_A* : *1,*
| *14.r_K* : *2,*
| *15.g_b* : *4,*
| *18.r_c* : *4,*
| *20.r_a* : *4*

- **sum equation**

A + 4b -> 2K +4a + 4c



ENVIRONMENT BEHAVIOUR, THREE STYLES

- **STYLE 1A**
  - -> *weak assumptions*
  - -> *infinite flow into/out the network*
- **STYLE 1B**
  - -> *firm assumptions*
  - -> *finite, but sufficient reservoir of auxiliary compounds*
- **STYLE 2**
  - -> *strong assumptions*
  - -> *finite, but sufficient reservoir of auxiliary compounds*
  - -> *quantitative relations of input/output compounds*

INCREASING STRENGTH

ENVIRONMENT BEHAVIOUR, STYLE 1A

- **weak assumptions**

- **input compounds**
  -> *generating pre-transitions*

- **output compounds**
  -> *consuming post-transitions*

} *no assumptions about quantitative relations of input/output compounds*

- **auxiliary compounds**
  -> *generating pre-transitions & consuming post-transitions*
  -> *infinite reservoir*

- **typical properties**

INA																
ORD	HOM	NBM	PUR	CSV	SCF	CON	SC	Ft0	tF0	Fp0	pF0	MG	SM	FC	EFC	ES
N	N	N	Y	N	N	Y	N	Y	Y	N	N	N	N	N	N	N
DTP	CPI	CTI	B	SB	REV	DSt	BSt	DTr	DCF	L	LV	L&S				
?	N	Y	N	N	?	N	?	N	n	Y	Y	N				

ENVIRONMENT BEHAVIOUR, STYLE 1A

- **weak assumptions**

- **input compounds**
  -> *generating pre-transitions*

- **output compounds**
  -> *consuming post-transitions*

no assumptions about quantitative relations of input/output compounds

- **auxiliary compounds**
  -> *generating pre-transitions & consuming post-transitions*
  -> *infinite reservoir*

- **typical properties**

INA																
ORD	HOM	NBM	PUR	CSV	SCF	CON	SC	Ft0	tF0	Fp0	pF0	MG	SM	FC	EFC	ES
N	N	N	Y	N	N	Y	N	Y	Y	N	N	N	N	N	N	N
DTP	CPI	CTI	B	SB	REV	DSt	BSt	DTr	DCF	L		LV	L&S			
?	N	Y	N	N	?	N	?	N	n	Y		Y	N			

ENVIRONMENT BEHAVIOUR, STYLE 1B

- **firm assumptions**

- **input compounds**
  -> *generating pre-transitions*

- **output compounds**
  -> *consuming post-transitions*

no assumptions about quantitative relations of input/output compounds

- **auxiliary compounds**
  -> *neither generating pre-transitions nor consuming post-transitions*
  -> *finite reservoir* -> *P-invariants (= traps = co-traps)*
  -> *Which (minimal) initial token distribution makes the net live ?*

- **typically expected properties**
  -> *see style 1*

ENVIRONMENT BEHAVIOUR, STYLE 2

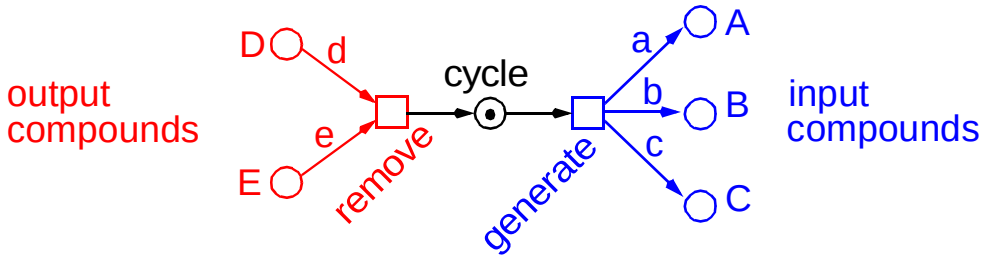
- **strong assumptions**
  - -> *quantitative relations of input/output compounds*
  - -> *finite reservoir of auxiliary compounds*

- **typically expected properties**

INA																
ORD	HOM	NBM	PUR	CSV	SCF	CON	SC	Ft0	tF0	Fp0	pF0	MG	SM	FC	EFC	ES
N	N	N	Y	N	N	Y	Y	N	N	N	N	N	N	N	N	N
DTP	CPI	CTI	B	SB	REV	DSt	BSt	DTr	DCF	L	LV	L&S				
?	Y	Y	Y	Y	?	N	?	N	Y	Y	Y	N				

- **additional model component**
  - -> *network sum equation*

aA + bB +cC -> dD + eE



- **How to compute ? For alternative paths ?**

ENVIRONMENT BEHAVIOUR, STYLE 2

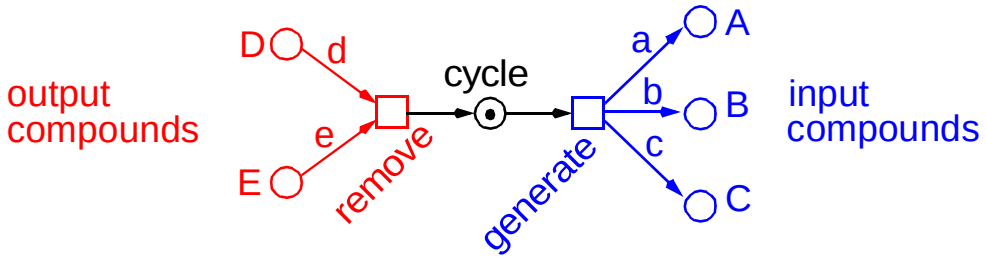
- **strong assumptions**
  - -> *quantitative relations of input/output compounds*
  - -> *finite reservoir of auxiliary compounds*

- **typically expected properties**

INA																
ORD	HOM	NBM	PUR	CSV	SCF	CON	SC	Ft0	tF0	Fp0	pF0	MG	SM	FC	EFC	ES
N	N	N	Y	N	N	Y	Y	N	N	N	N	N	N	N	N	N
DTP	CPI	CTI	B	SB	REV	DSt	BSt	DTr	DCF	L	LV	L&S				
?	Y	Y	Y	Y	?	N	?	N	N	Y	Y	N				

- **additional model component**
  - -> *network sum equation*

aA + bB +cC -> dD + eE



- **How to compute ? For alternative paths ?**

CASE STUDIES

- **case study 1**
  - -> *signal-transduction network - apoptosis*
  - -> *no stoichiometric relations - ordinary place/transition net*
  - -> *many read arcs, resolved for analysis*
  - -> *environment behaviour, style 1A*

- **case study 2**
  - -> *metabolic network - carbon metabolism in potato tuber*
  - -> *stiochiometric relations known - non-ordinary place/transition net*
  - -> *many reversible reactions*
  - -> *environment behaviour, style 1B*

- **case study 3**
  - -> *metabolic network - glycolysis and pentose phosphate metabolism*
  - -> *coloured nets (Design-CPN)*
  - -> *environment behaviour, style 2, computed by prototype tool SY / Genrich*

CASE STUDIES

- **case study 1** **-> 37 P / 45 T -> 10 t-inv's**
  - -> *signal-transduction network - apoptosis*
  - -> *no stoichiometric relations - ordinary place/transition net*
  - -> *many read arcs, resolved for analysis*
  - -> *environment behaviour, style 1A*
- **case study 2** **-> 17 P / 25 T -> 19 t-inv's / 3 p-inv's**
  - -> *metabolic network - carbon metabolism in potato tuber*
  - -> *stiochiometric relations known - non-ordinary place/transition net*
  - -> *many reversible reactions*
  - -> *environment behaviour, style 1B*
- **case study 3** **-> 32 P / 22 T -> 1 t-inv / 39 p-inv's**
  - -> *metabolic network - glycolysis and pentose phosphate metabolism*
  - -> *coloured nets (Design-CPN)*
  - -> *environment behaviour, style 2, computed by prototype tool SY / Genrich*

CHALLENGES

- **extensions**
  - -> *read arcs*
  - -> *inhibitor arcs !?*
- **efficient computation of minimal invariants**
  - -> *exponential complexity*
  - -> *compositional approach ?*
- **analysis of bounded, but not safe non-ordinary nets with inhibitor arcs**
  - -> *huge state spaces, beyond exponential growth (?)*
  - -> *smaller, bounded version of case study 2* $\geq 10^{10}$ *states (IDD-based mc tool)*
- **analysis of unbounded nets**
  - -> *besides T-invariant analysis ?*
- **model checking**
  - -> *relevant properties ?*

FURTHER CASE STUDIES

- blood clotting
  (hemostasis versus fibrinolysis)
- the whole E. coli pathway
- the whole potato tuber pathway
- detailed glycolysis in humans
- G1/S - phase in mammalian cells
- lipoprotein metabolism (liver)

FURTHER CASE STUDIES

- blood clotting (hemostasis versus fibrinolysis) -> Gerry Neumann / BTU
- the whole E. coli pathway -> Nina Kramer / TFH
- the whole potato tuber pathway -> Nina Kramer / TFH
- detailed glycolysis in humans -> Thomas Runge / BTU
- G1/S - phase in mammalian cells -> Thomas Kaunath / TFH
- lipoprotein metabolism (liver) -> Daniel Schrödter / BTU

SUMMARY

- **representation of bionetworks by Petri nets**
  - -> *unifying view*
  - -> *animation*
  - -> *model validation against consistency criteria*
  - -> *qualitative/quantitative behaviour prediction*
- **three styles of environment description**
- **steady state behaviour**
  - -> *pathways* -> *T-invariants*
  - -> *elementary modes* -> *minimal T-invariants*
- **consistency criteria**
  - -> *CTI*
  - -> *T-invariant <-> biological interpretation*
- **many challenging questions for analysis techniques**

GRAZIE

PER LA VOSTRA

ATTENZIONE !