

Conceptual modeling in Systems Biology: before math



Outline

- Why conceptual modeling?
- SBGN – diagram as conceptual model
- SBGN-PD
- Exercise
- SBGN-ER
- Exercise
- SBGN-AF
- Exercise Solutions and Questions

Scientific modelling

- “A [model](#) in science is a physical, mathematical, or logical representation of a system of entities, phenomena, or processes.” (Wikipedia)
- “**Scientific modelling** is the process of generating [abstract](#), [conceptual](#), [graphical](#) and or [mathematical models](#)” (Wikipedia)

Why model always mathematical

- Formalised
 - Unambiguous
 - Precise
 - Convertible to machine-readable format
 - Analysable
 - Analytically
 - Numerically
 - Automatically
 - Manually
 - Simulatable

Why development mathematical of models are difficult?

- Formalised
 - Precise
 - Skills are required
 - Concept mapping is required
 - Dual expertise
 - Assumption validation

New language?

- Allow the representation of diverse biological objects and interactions
- Be semantically and visually unambiguous;
- Allow implementation in software that can aid the drawing and verification of diagrams;
- Have semantics that are sufficiently well defined that software tools can convert graphical models into mathematical formulas for analysis and simulation; and
- Be unrestricted in use and distribution, so that the entire community can freely use the notation without encumbrance or fear of intellectual property infractions.

SBGN

- Standard representation of essential biochemical and cellular processes
 - Set of symbols
 - Semantics
 - Syntax
- Ultimate goal:
 - Thousands way to draw
 - One way to read
- www.sbgn.org



"The goal of SBML is to help people to disagree as precisely as possible".

Ed Franck, Argonne National Laboratory

SBGN

- Community effort (about 30 contributors)
- Started at 2006 by Hiroaki Kitano
- First language release ICSB 2008 (SBGN-PD)
- Three languages (released independently)
 - *Process Diagram*: the causal sequences of molecular processes and their results
 - *Entity Relationship*: the interactions between entities irrespective of sequence
 - Activity flow: the flux of information going from one entity to another

SBGN

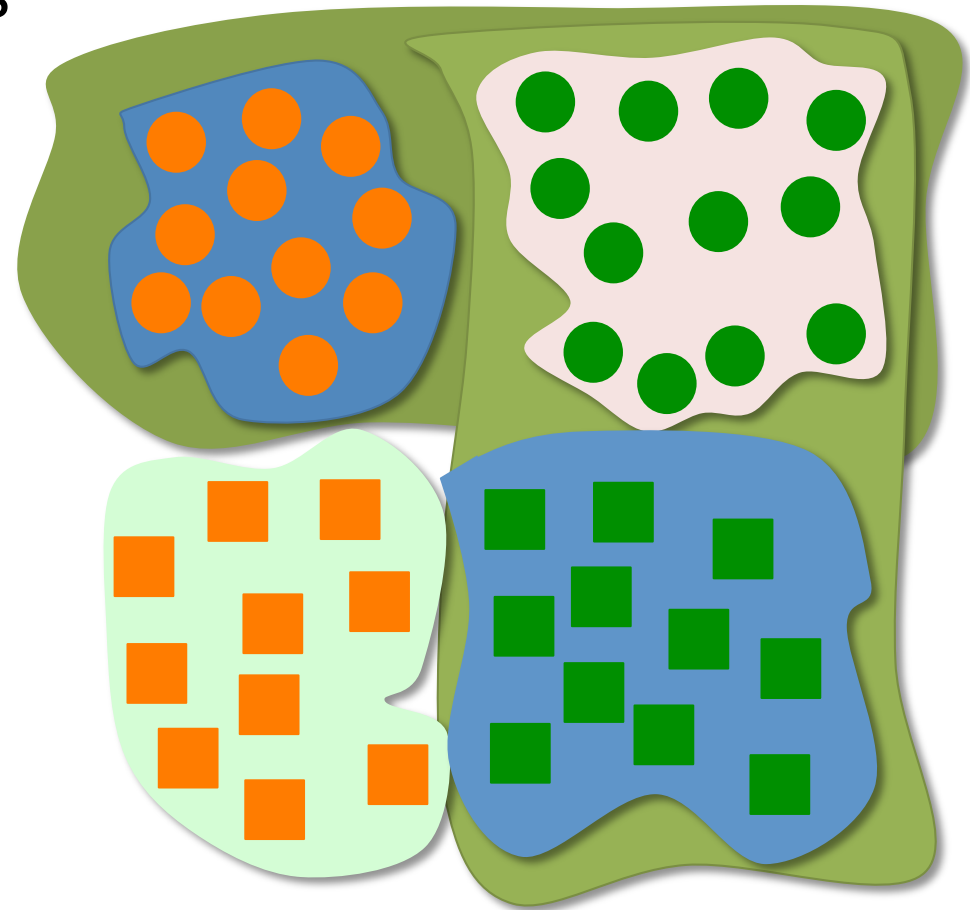
- Colour has no meaning
- Size has no meaning
- Meaning should be conserved upon
 - scaling
 - resolution
 - relayout

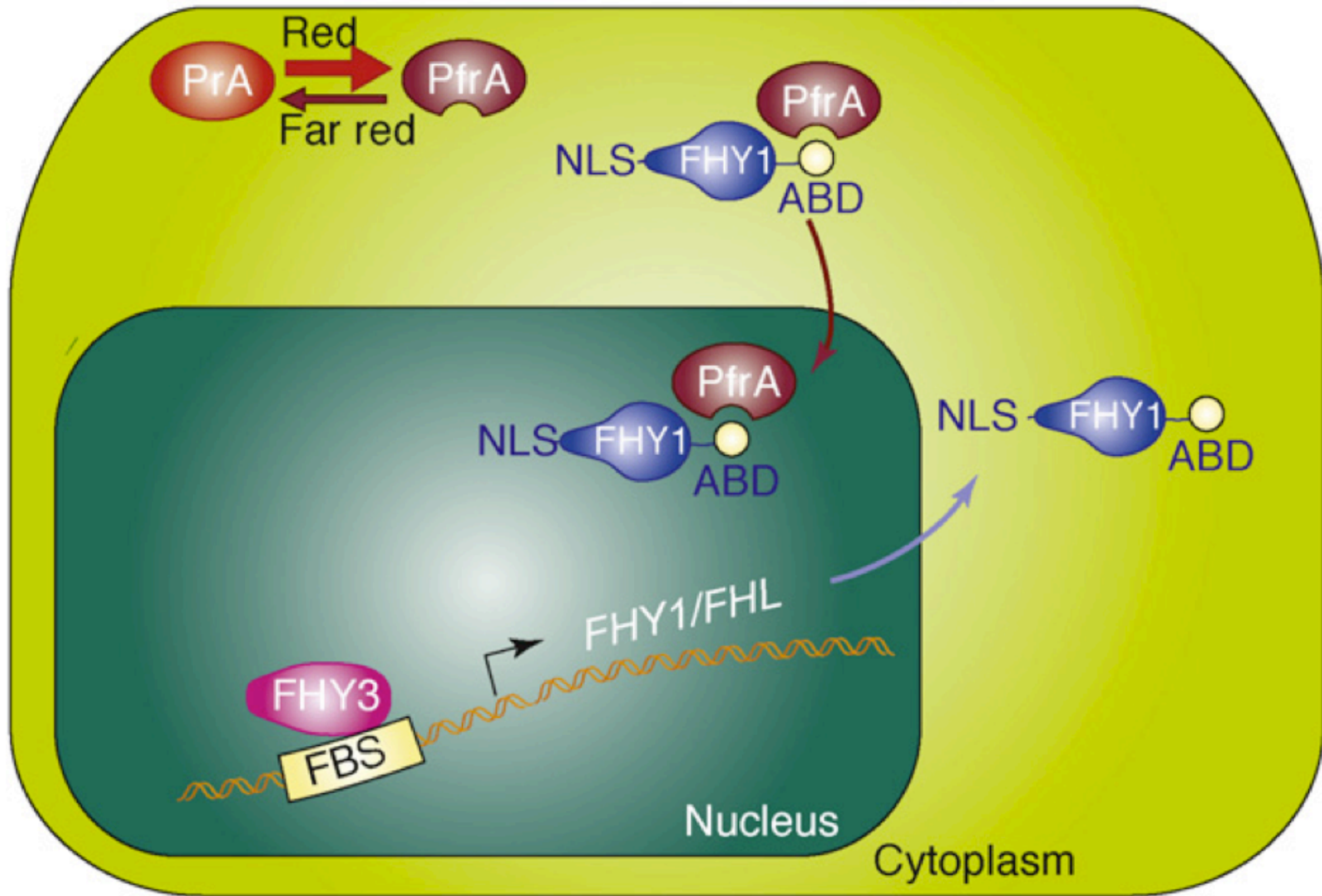
Process Description language

- Description of change
- Sequence of events is depicted
- Operates with pools rather than individuals
- Mapping to ODE model (SBML)

Pools of entities

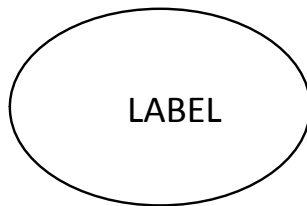
- Collection of molecules indistinguishable in some sense
- Non-overlapping
- characterized by concentration



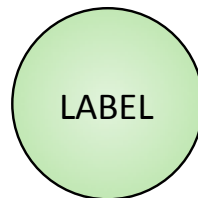


Entity types

Unspecified
Entity



Simple
chemical



Macromolecule

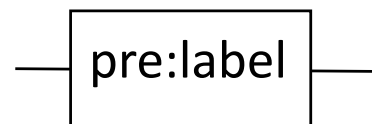


Nucleic acid
feature



Material type of molecule

- Unit of information
- Controlled vocabulary



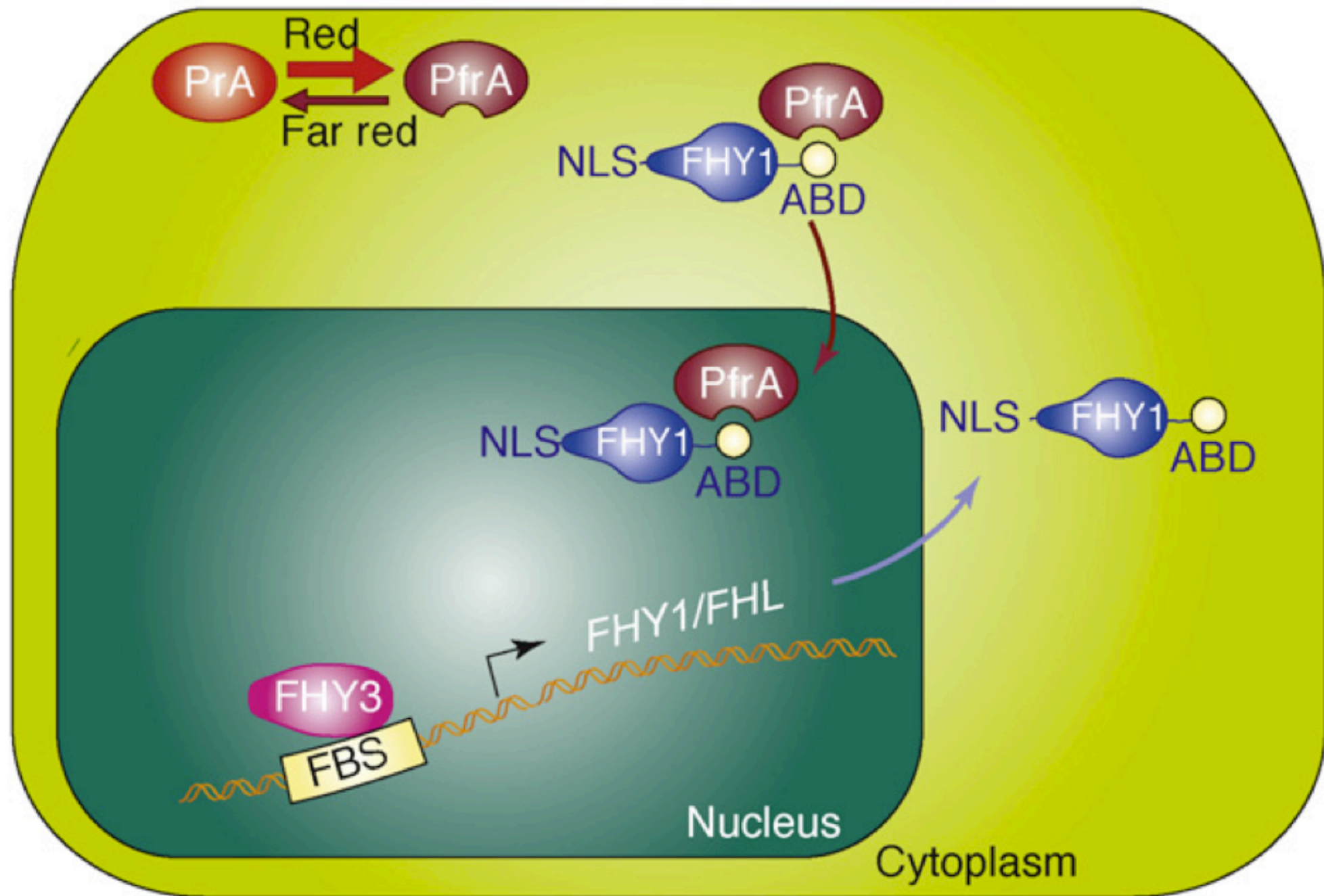
Name	Label
Non-macromolecular ion	mt:ion
Non-macromolecular radical	mt:rad
Ribonucleic acid	mt:rna
Deoxyribonucleic acid	mt:dna
Protein	mt:prot
Polysaccharide	mt:psac

Conceptual type of nucleic acid feature

Name	Label
Gene	ct:gene
Transcription start site	ct:tss
Gene coding region	ct:coding
Gene regulatory region	ct:grr
Messenger RNA	ct:mRNA

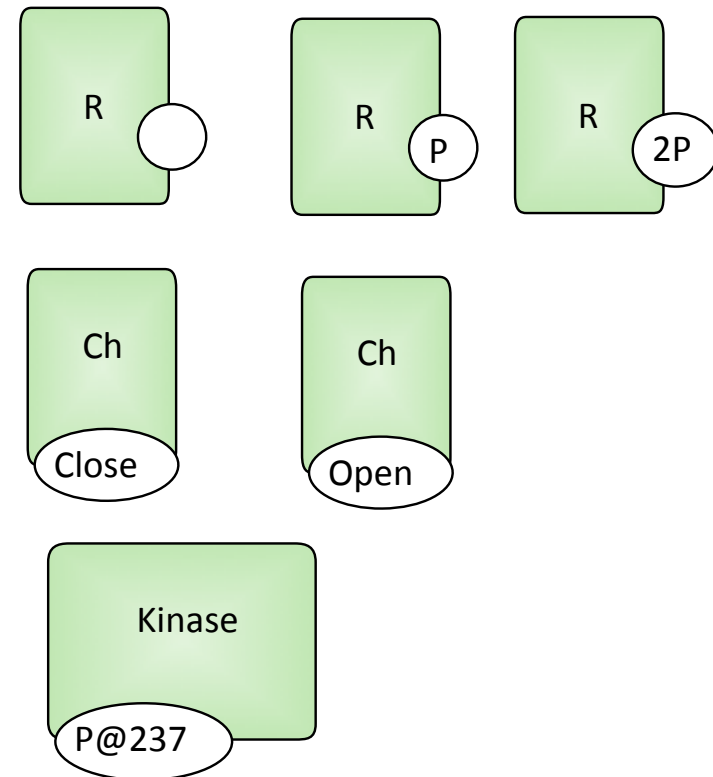
mt:prot

PhyA



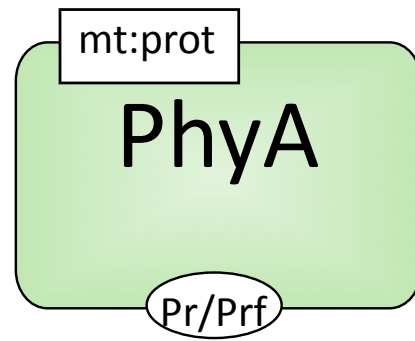
Macromolecular pools: state variables

- Pools is set of molecules somehow undistinguishable
- Molecules can be in different state
 - (Non)phosphorylated
 - Open/close channel
 - Modified at some state



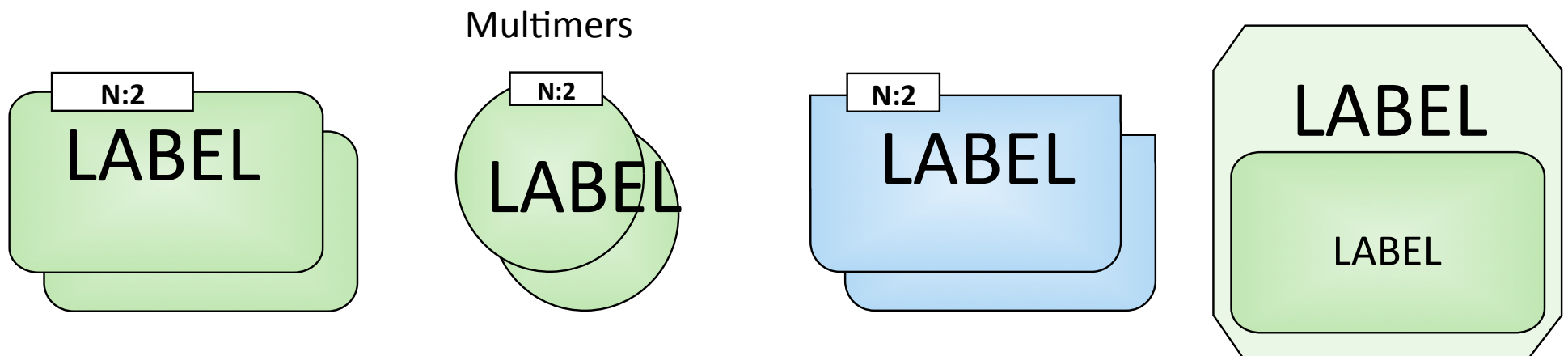
Stateless and state-full entity types

- Not all entities can have states: Stateless
 - Simple chemicals
 - Unspecified entity
- State-full entities
 - Macromolecule
 - Nucleic acid feature
 - Complex
- State is defined as combination of state values
- Once defined state variable should be always visible



Complex and multimer

- Represents complexes of molecules held together by non-covalent bonds
- Multimer require cardinality
- Can have State variables
 - In multimer it means that all monomers
 - Use complex instead



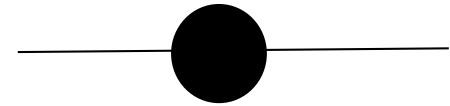
PhyA
Pr

N:2
PhyA
Pr

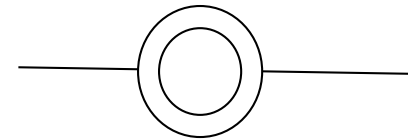
Key concept: Process

- Process: conversion of element of one pool to another
- Special cases:
 - Non-covalent binding
 - Association
 - Dissociation
 - Incompleteness
 - Uncertain process
 - Omitted process

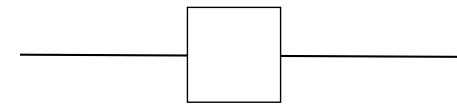
Association



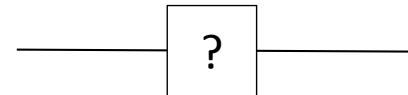
Dissociation



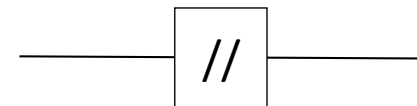
Process

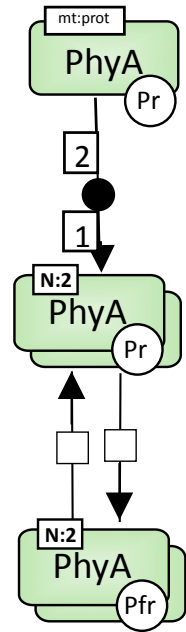


Uncertain process



Omitted process





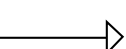
Arcs

- Using pools by process
 - Consumption/production
 - Stoichiometry (optional)
- Regulating process rate
 - Stimulation
 - Inhibition
 - Catalysis
- Requirement for process
 - Necessary stimulation


 consumption

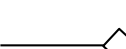
 production

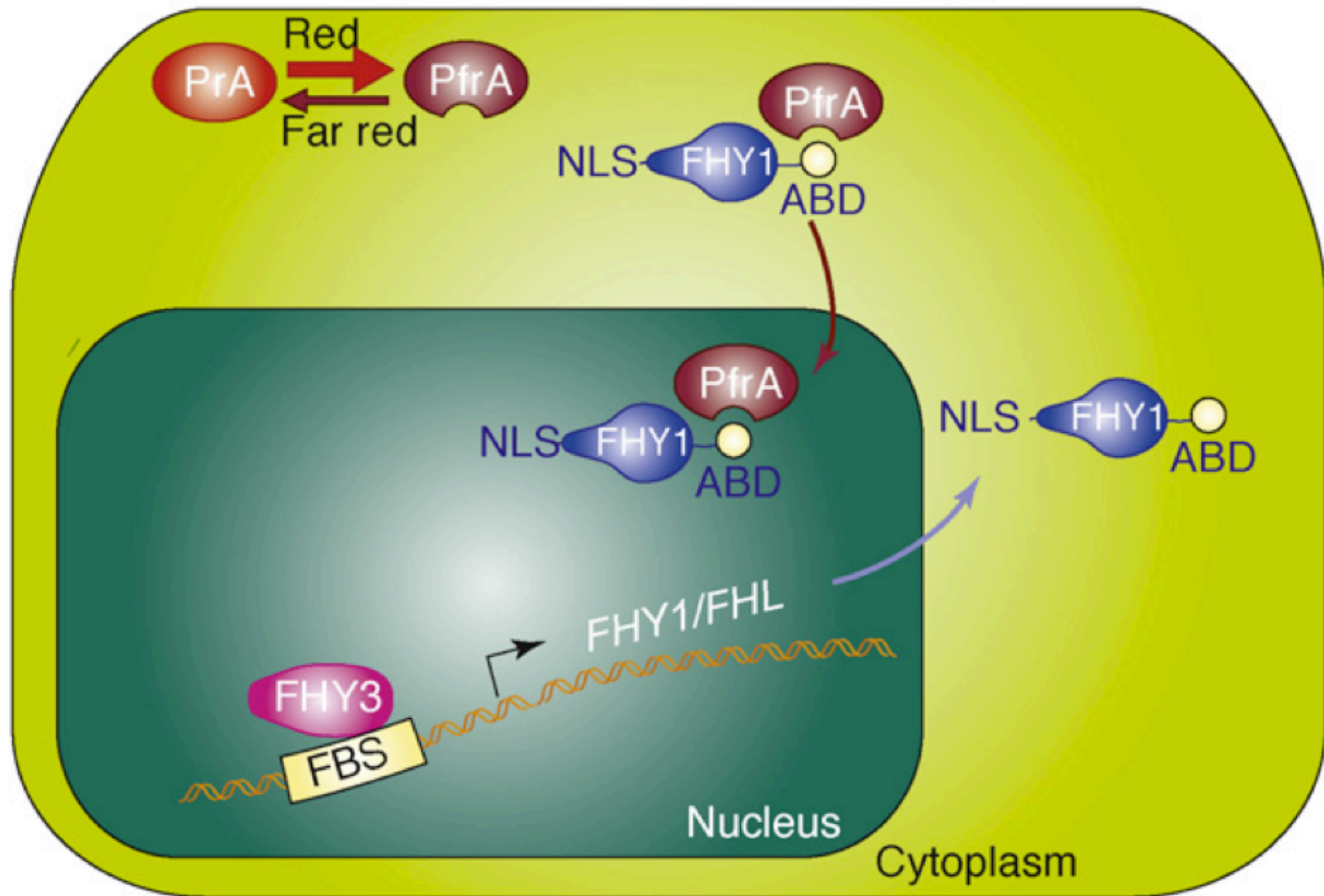
 catalysis

 stimulation

 inhibition

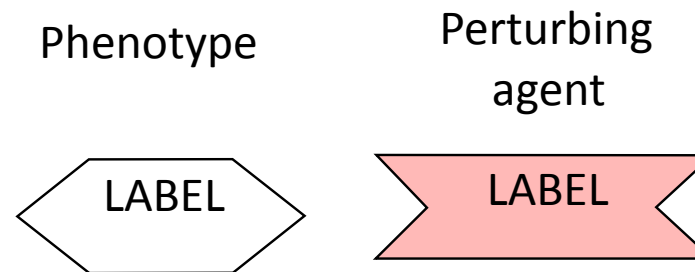
 necessary
stimulation

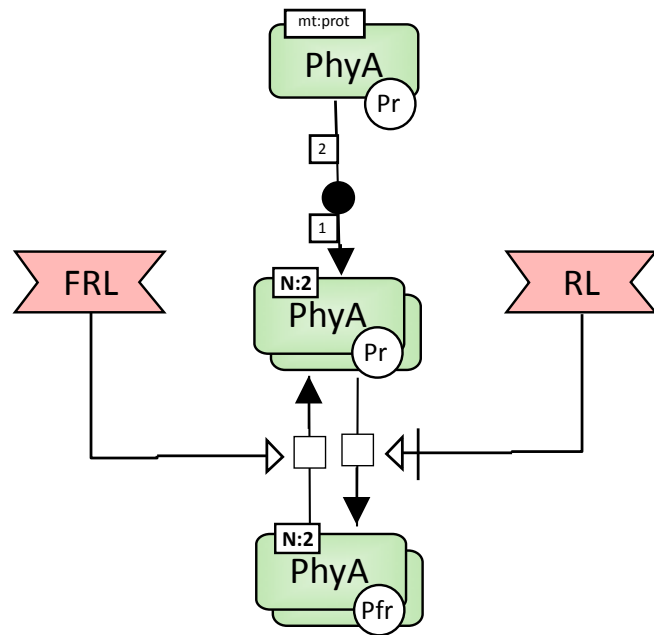
 modulation



Environmental influence

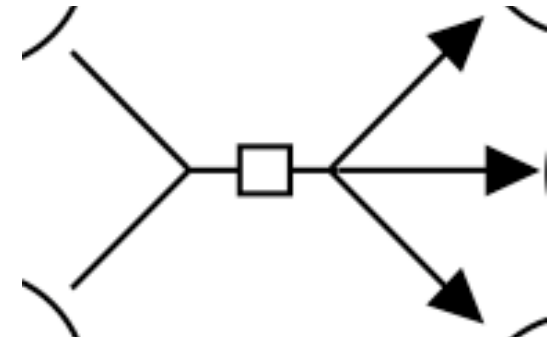
- External influences: Perturbing agent
 - Light
 - Temperature change
 - Mutation/disease
- System manifestation: Phenotype
 - Apoptosis
 - Phenotype





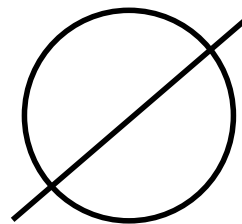
Laying out process arcs

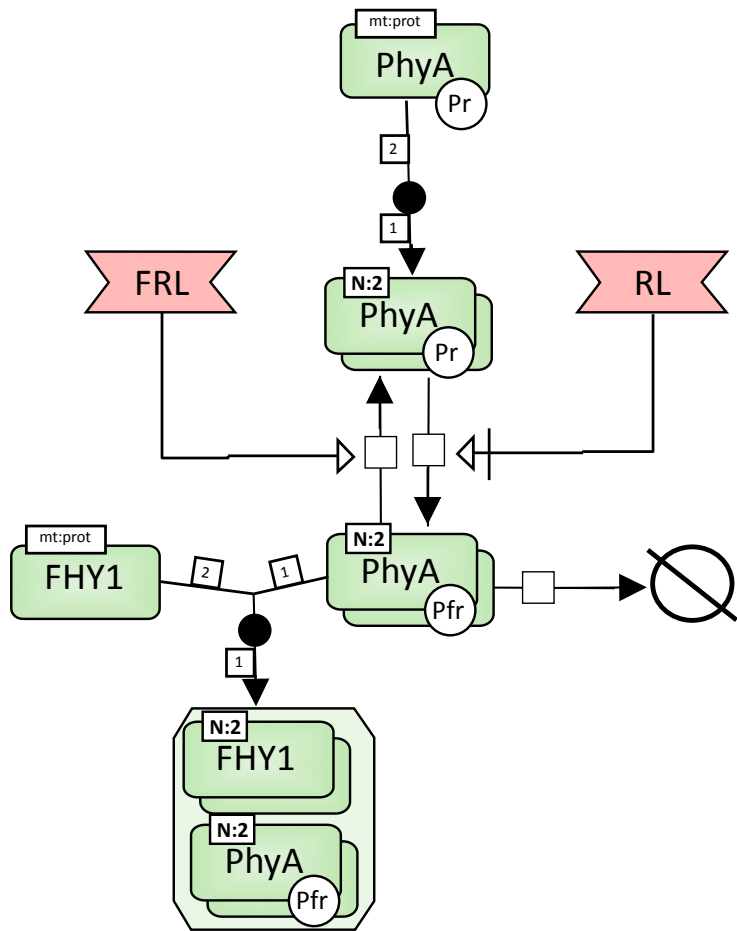
- Production can represent consumption:
 - Reversible process
 - Substrates and products should come to opposite sides of process shape (two connectors)
- Regulatory arcs should come to other two sides of the process
- If you have separate regulation of forward and backward process, you have to split

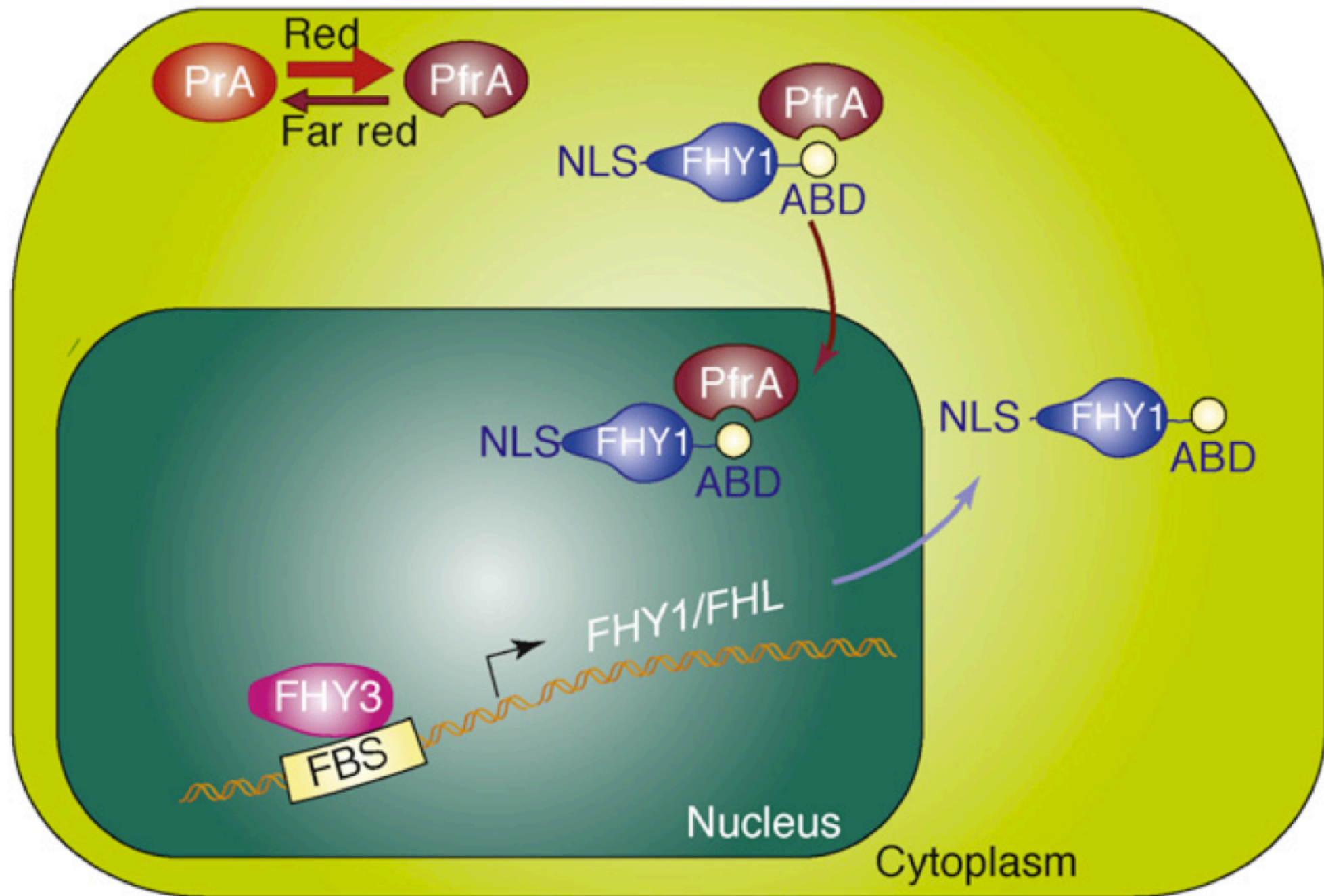


Sink/source: creation and destruction

- We need represent creation and destruction of entities
- We cannot omit consumption and production arcs
- We need shape to represent source of materials and sink of degraded entities

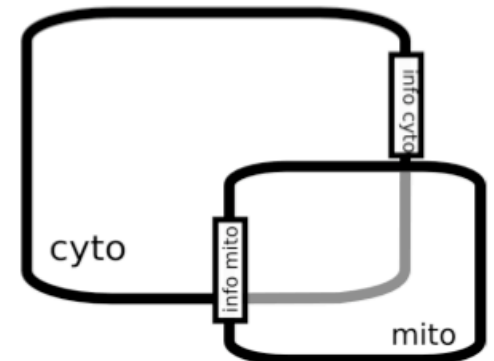
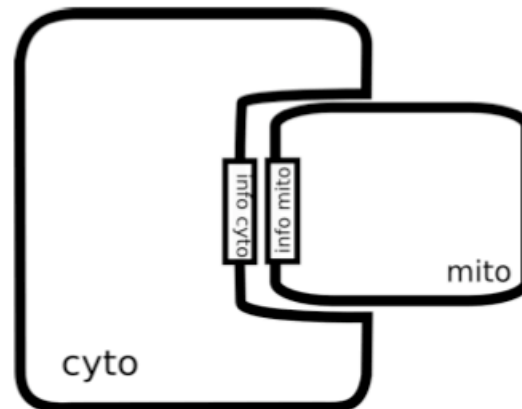


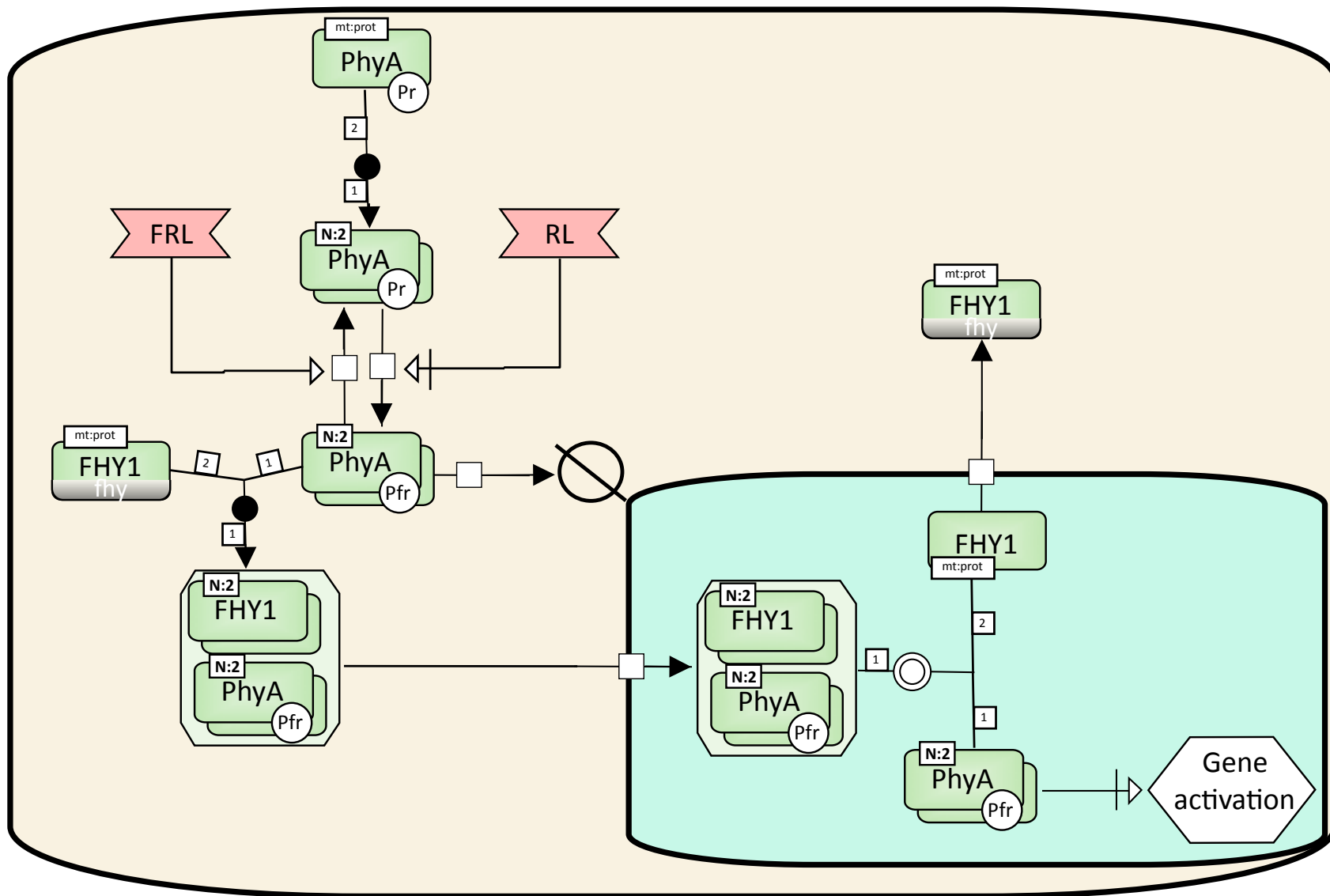




Compartments

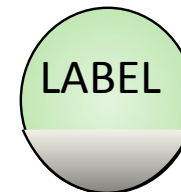
- Container to represent physical or logical structure
 - Free form
 - Visually thicker line
- The same entity pools in different compartments are different
- Compartments are independent
- Overlapping do not mean containment

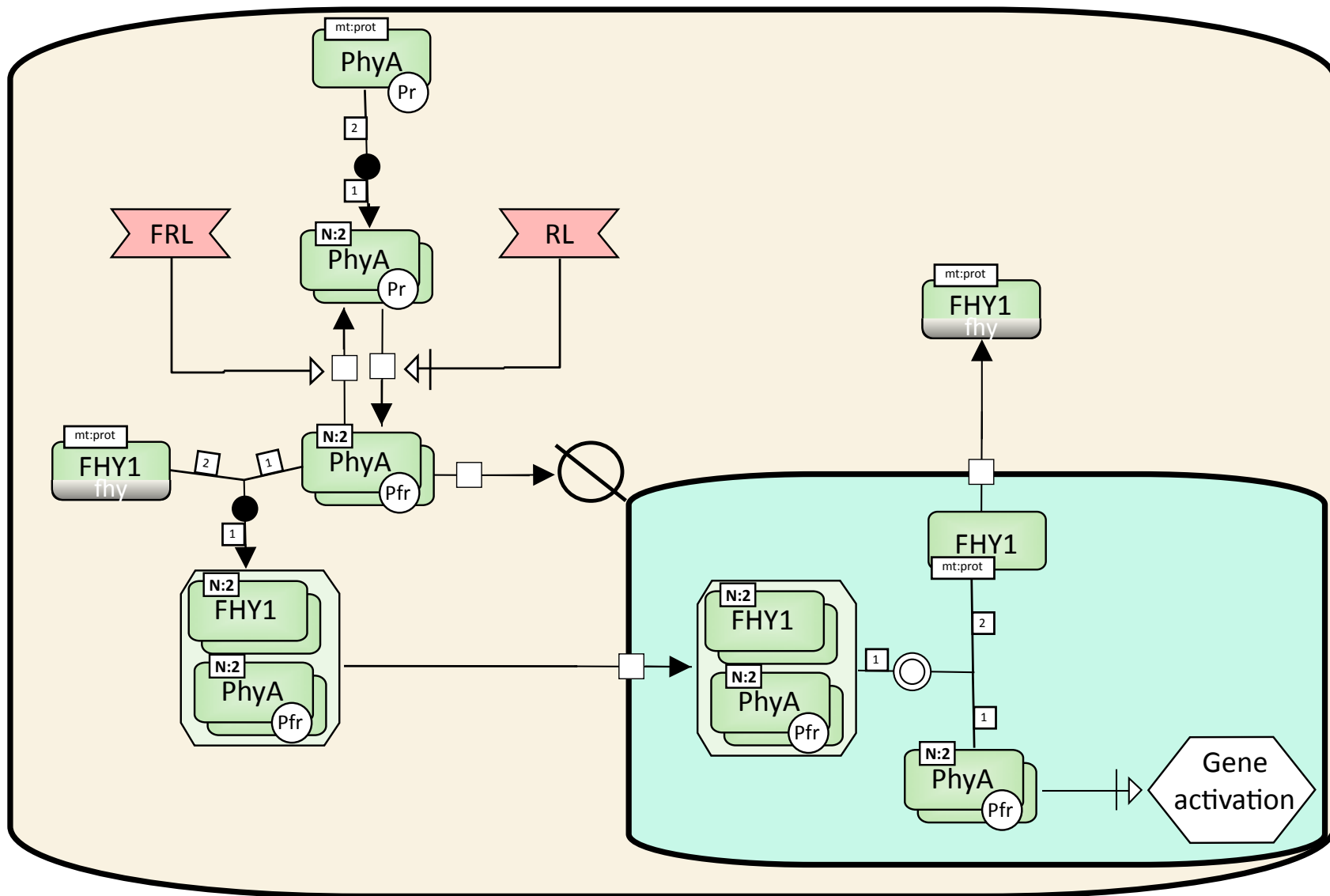




Clone marker

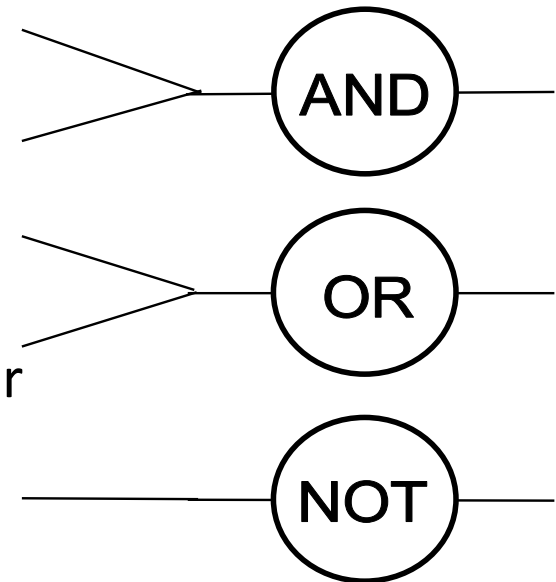
- Each entity pool represents only once on the map
- Layout problems
- Clone marker as visual indicator of duplication
 - Stateless nodes carry unnamed marker
 - State-full nodes carry named marker to simplify recognition





Logical gates

- Encode of network logic
 - to simplify layout
 - 20 activators for the process
 - to include uncertain information
 - Combination of TF with unknown or combinatorial binding kinetics
- Three main logic operations
 - AND: all are required
 - OR: any combination is required
 - NOT: prevent influence



Strength and weakness of SBGN-PD

Strength

- Easy convert to math
 - Natural mapping to SBML
- A lot of information in DB
 - KEGG
 - Panther
- Timeline is easily extractable

Weakness

- Full explicit definition of state
 - Combinatorial complexity
 - Additional assumption to include uncertain information
- Laborious creation

FIRST EXERCISE

Entity-Relationship language

- Draw influences of entities
- States are independent
- There is no time sequence
- Logical or probabilistic description of system
- Naturally map narrative description of the system

Model to draw

- Process of Polymerase Chain Reaction (PCR):
 - Sense and antisense DNA strands bind to each other
 - Polymerase enzyme recreate missing parts of dsDNA based upon ssDNA as template
 - There are two short primers to initiate synthesis of new DNA
 - Once heated DNA melts and primers become able to bind ssDNA and prevent sense and antisense DNA strands bind to each other back

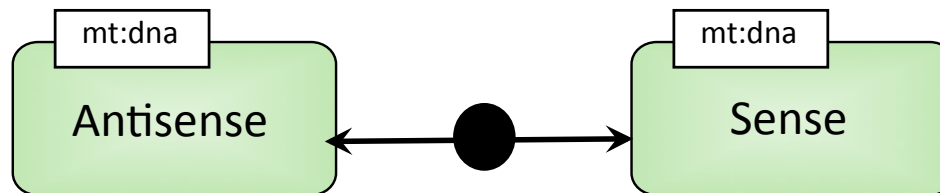
Entity

- There is only one shape for entity
 - There is no difference between Macromolecule and Gene
 - Material and/or conceptual type could be represented as Unit-of-Information
- Entity is something that can exist
 - Molecule
 - Gene
 - Allele



Interaction

- Entities can interact if they exists
- Interaction is statement
 - If (when) two entities interacts, then..
- We have new entity, which is actualization of
THEN:
 - Outcome

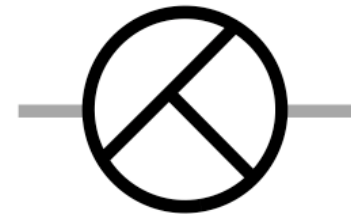


State variables

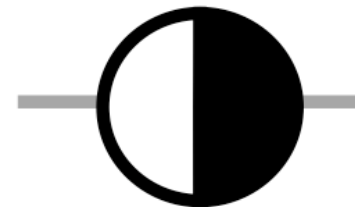
- State of Entity could be described with state variables
- Unlike PD state variables
 - do not require to have value
 - Are independent
 - Can be assigned
- Two special variable types
 - Existence
 - Location

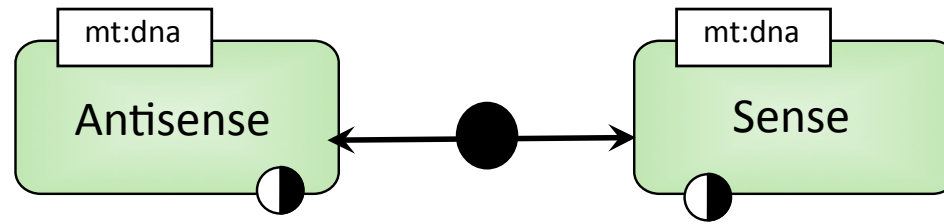


Location



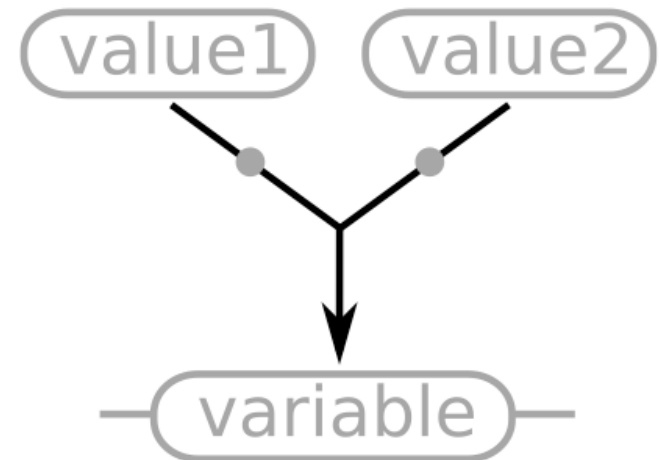
Existence

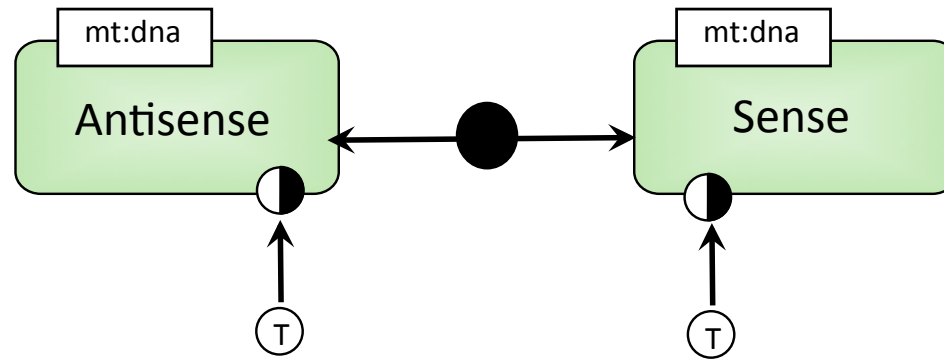




State value assignment

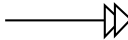
- Another type of statement
 - If (when) state variable acquire a value
 - Site is phosphorylated
 - Entity deleted
 - existence assigned FALSE
 - Entity moved to nucleus
 - location assigned value 'nucleus'
 - then ...
 - Another outcome
- Selector:
 - More than one value of variable

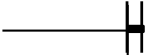


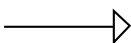


Influences

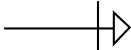
- Arc to represent influence of entity (outcome) to a relationship (interaction or assignment)
- Logic rules to connect statements

 absolute stimulation

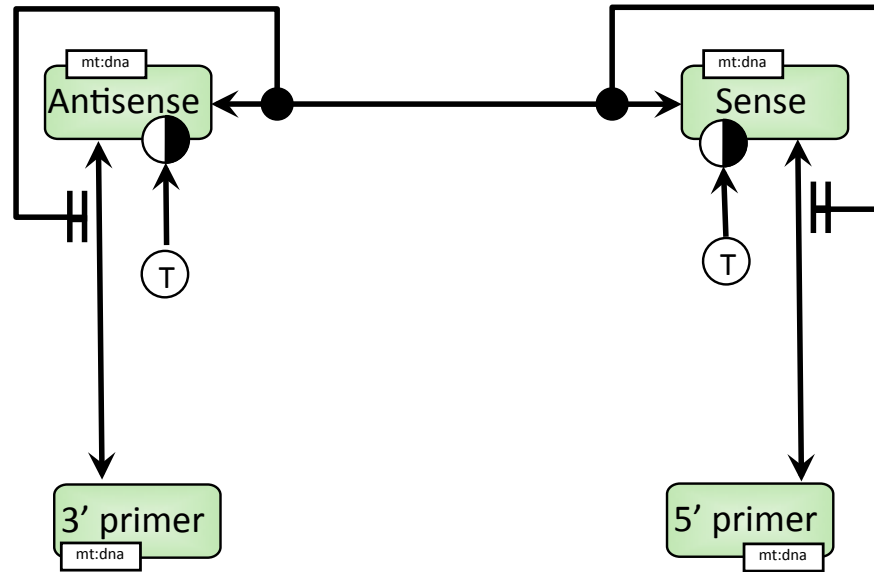
 absolute inhibition

 stimulation

 inhibition

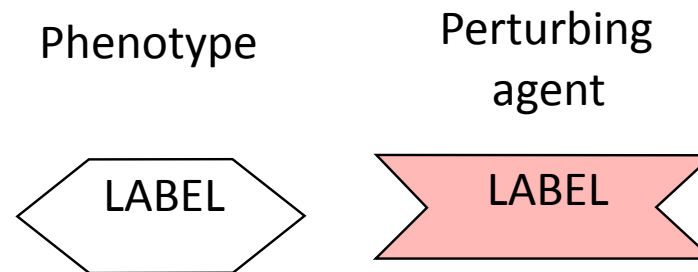
 necessary stimulation

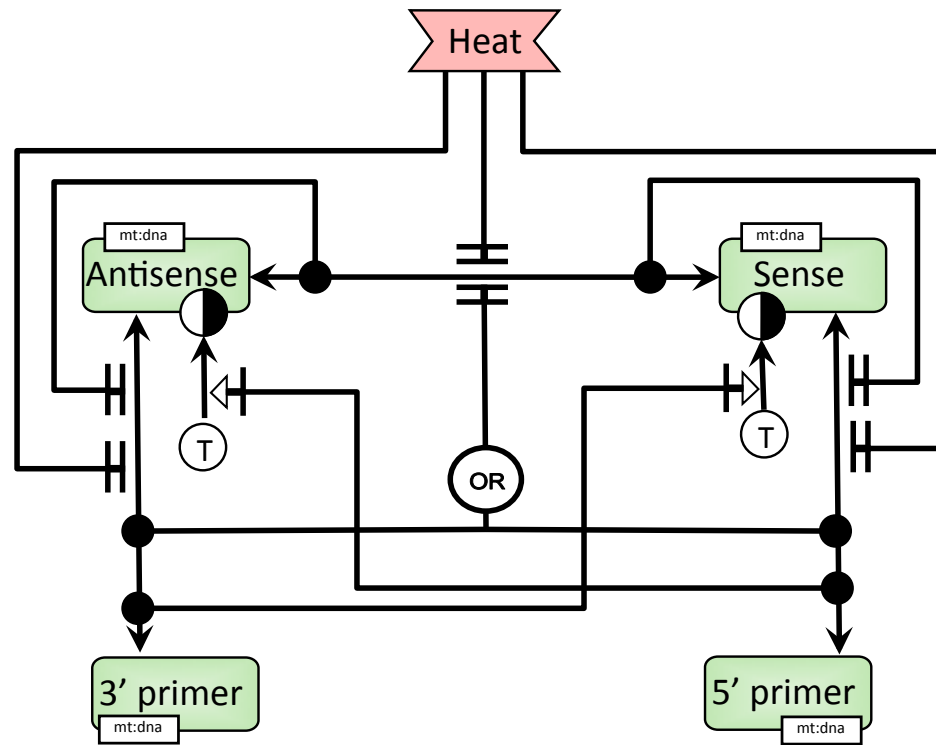
 modulation



Environmental influence

- External influences: Perturbing agent
 - Light
 - Temperature change
 - Mutation/disease
- System manifestation: Phenotype
 - Apoptosis
 - Phenotype



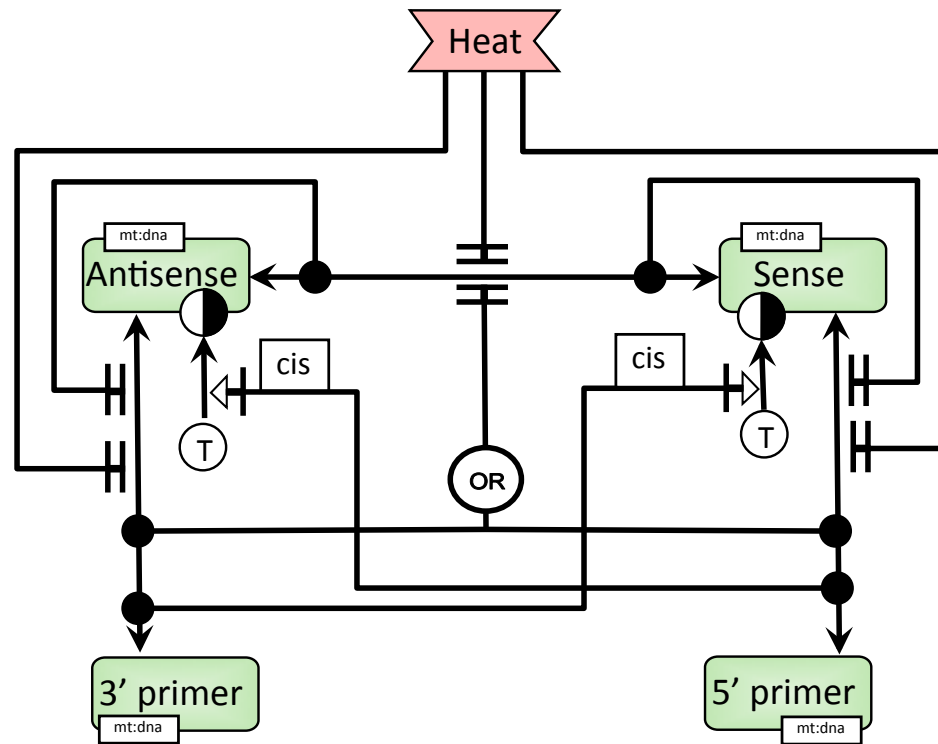


Cis and Trans interaction

- Working with individuals we need to distinguish between two cases
 - Changing itself (Cis)
 - Internal phosphorylation of RTK after dimerisation
 - Changing neighbours of the same type (Trans)
 - Activation kinase to phosphorylate another proteins
- Shown in the same way as stoichiometry in PD

cis

trans



Strength and weakness of SBGN-ER

Strength

- Handle combinatorial complexity naturally
- Close mapping to rule-based modelling
 - BNGL, kappa
- Statement based nature of ER helps in text annotation

Weakness

- Difficult to read
- No timeline
- Difficult for validation and reasoning

SECOND EXERCISE

Abstraction and decomposition

Decomposition

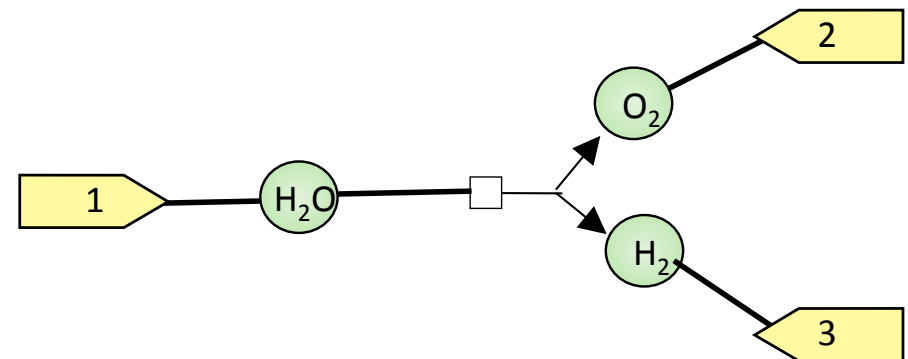
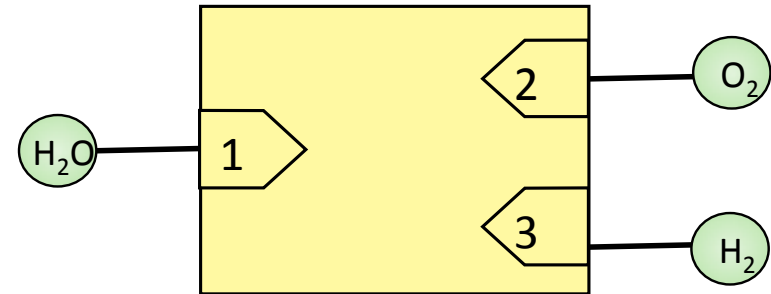
- Identify boundaries in the model
 - Scales
 - Time
 - Concentration
 - Function
 - Topology
- Split model into set of simple modules
- PD submap

Abstraction

- New concepts
- Less details
- Higher levels
- AF

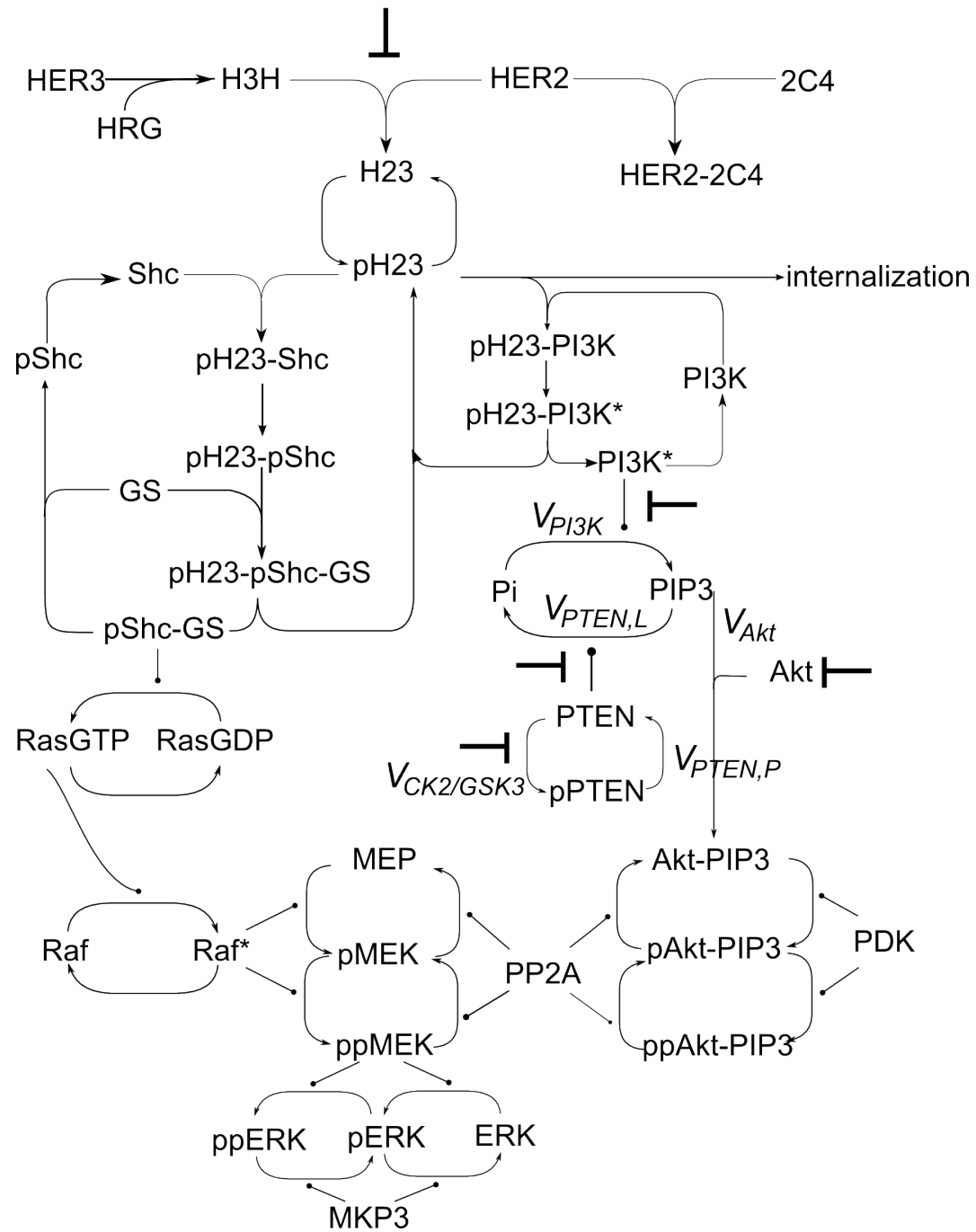
Back to PD: submaps

- Encapsulate modules
- Provide connection between main map and module
- Share the namespace
- There is no way to define role of elements within submap



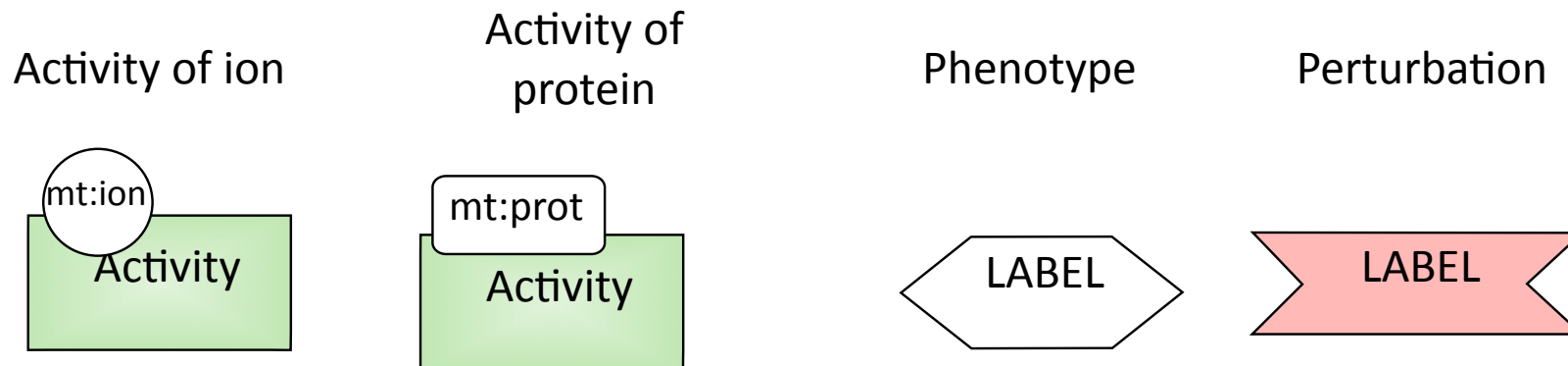
Activity Flow: abstraction

- Main concept is **Biological Activity**
 - Each node represents an activity, but not the entity.
 - Multiple nodes can be used to represent activities from one entity, e.g., receptor protein kinase.
 - One node can be used to represent activities from a group of entities (e.g., a complex).



Material and conceptual types in AF

- Activity node is rectangular to emphasize similarity to reaction
- Unit of information has shape according to node type
- Unit of information can carry name of entity, which has the activity



Regulatory arcs

- Operates on activities
- Shows influences
 - Positive
 - Catalysis
 - Stimulation
 - Negative
 - Inhibition
 - Required
 - Necessary stimulation

—▶ Positive influence

—| Negative influence

—▶ Necessary stimulation

—◇ Unknown influence

Logical gates

- Three main logic operations

- AND: all are required

- OR: any combination is required

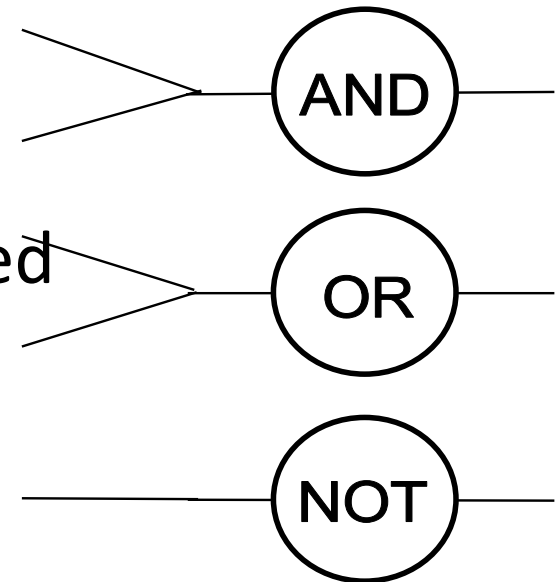
- NOT: prevent influence

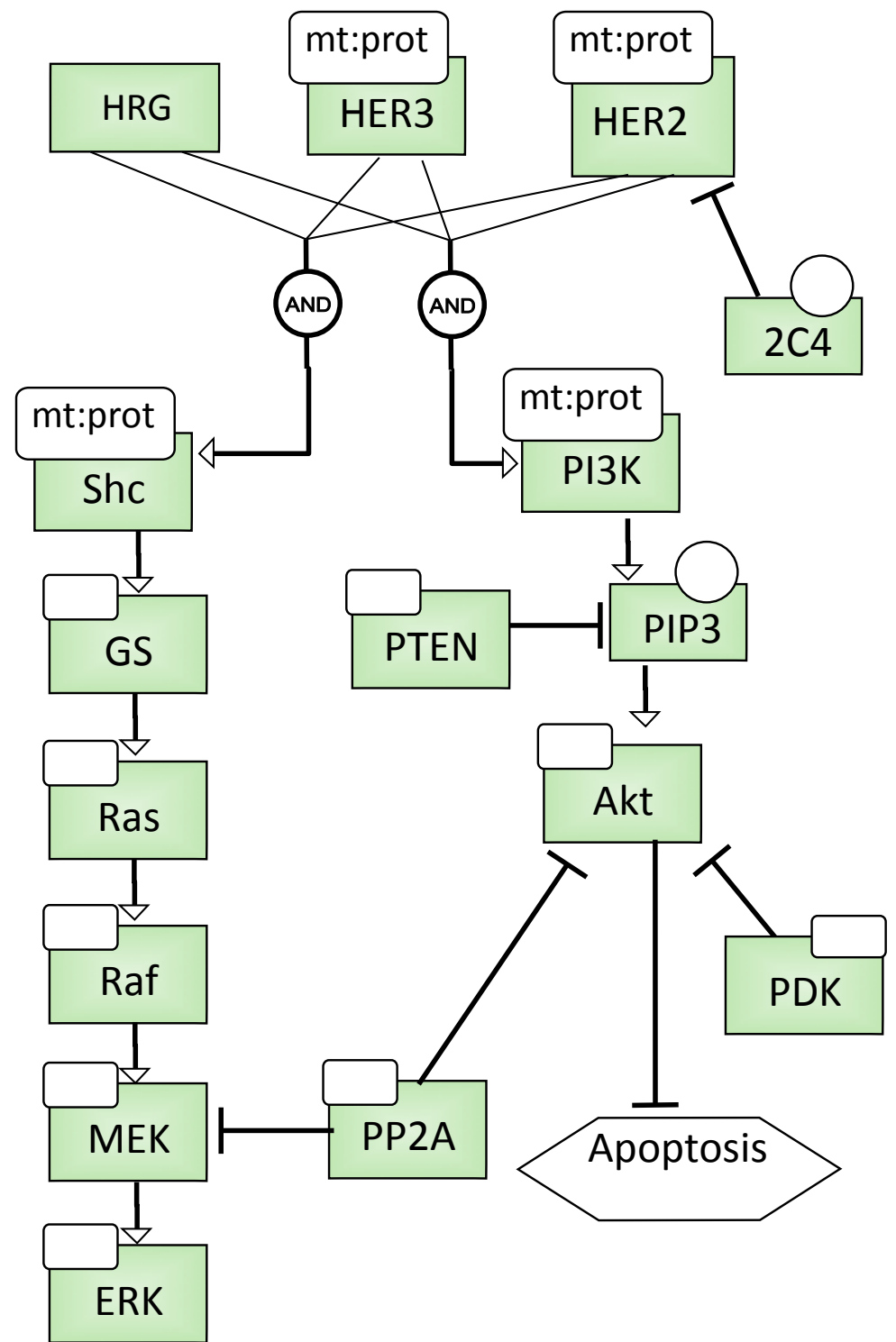
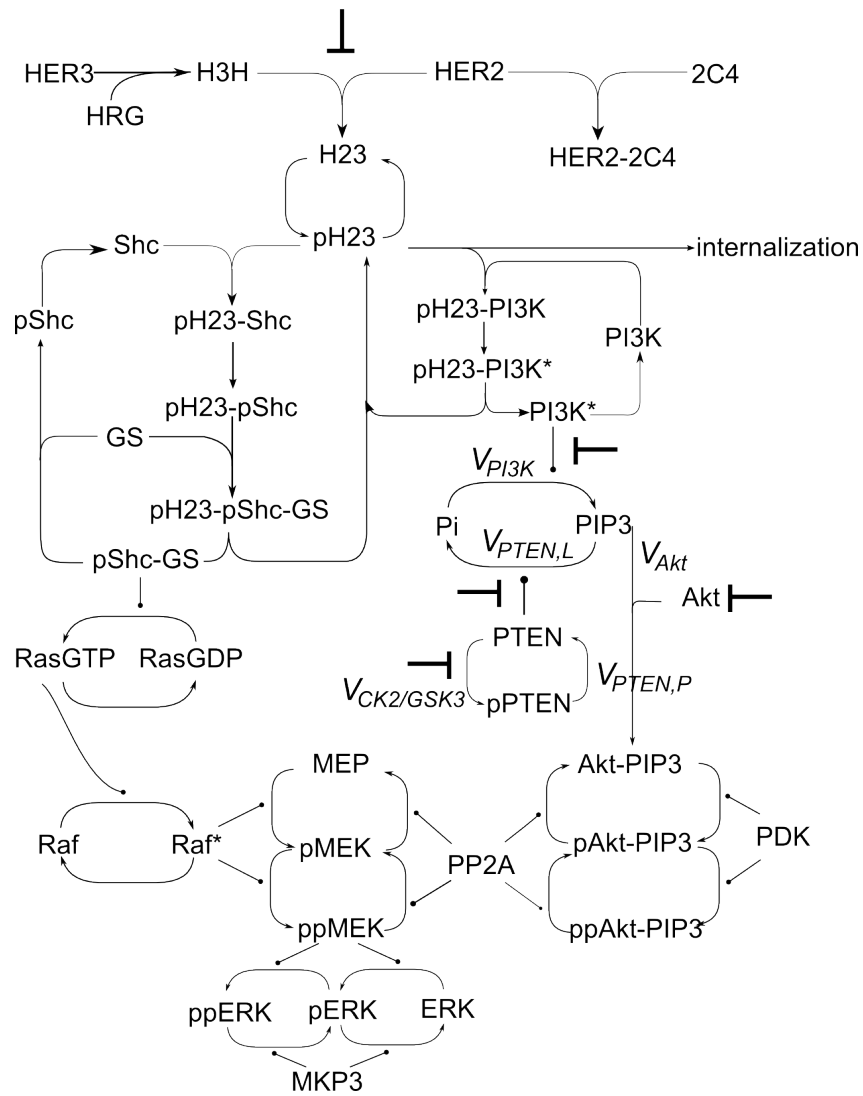
- Crucial for AF

- No complex

- No outcome

- No modifications





Strength and weakness of SBGN-AF

Strength

- Similar to biological sketch drawings
- Compact

Weakness

- Ambiguous
- Requires text or other diagram

THIRD EXERCISE

Specifications

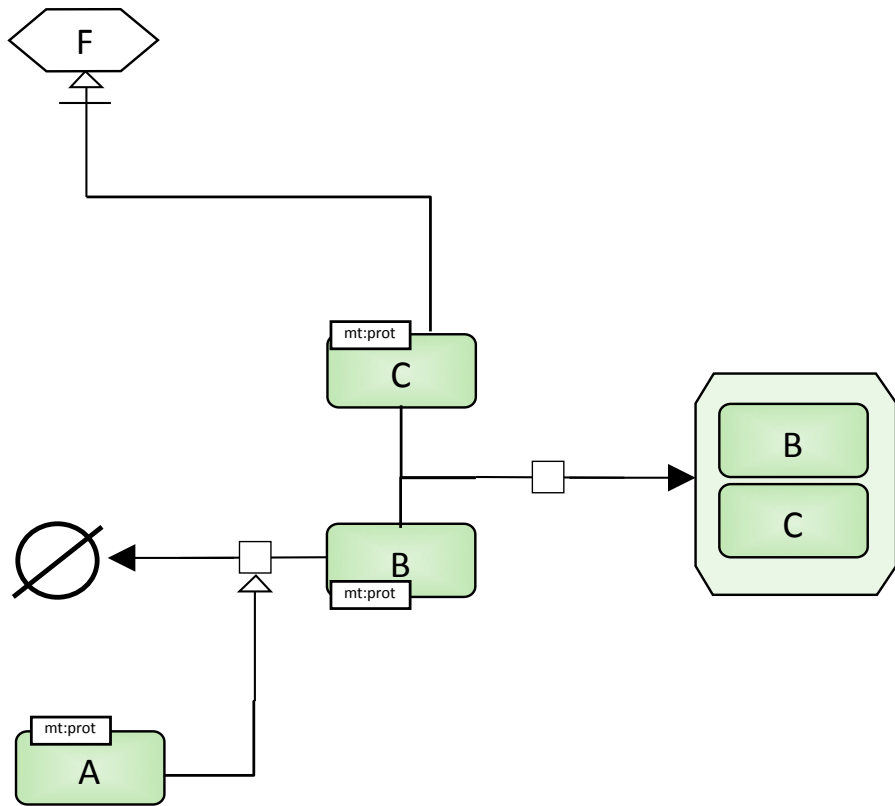
- Process Description ([doi:10.1038/npre.2009.3721.1](https://doi.org/10.1038/npre.2009.3721.1))
 - Molecular pools and reactions
- Activity Flow ([doi:10.1038/npre.2009.3724.1](https://doi.org/10.1038/npre.2009.3724.1))
 - Functions and their cross-coupling
- Entity Relationship ([doi:10.1038/npre.2009.3719.1](https://doi.org/10.1038/npre.2009.3719.1))
 - Molecules and their interactions
- www.sbggn.org
- SBGN-discuss mailing list

Software support of SBGN

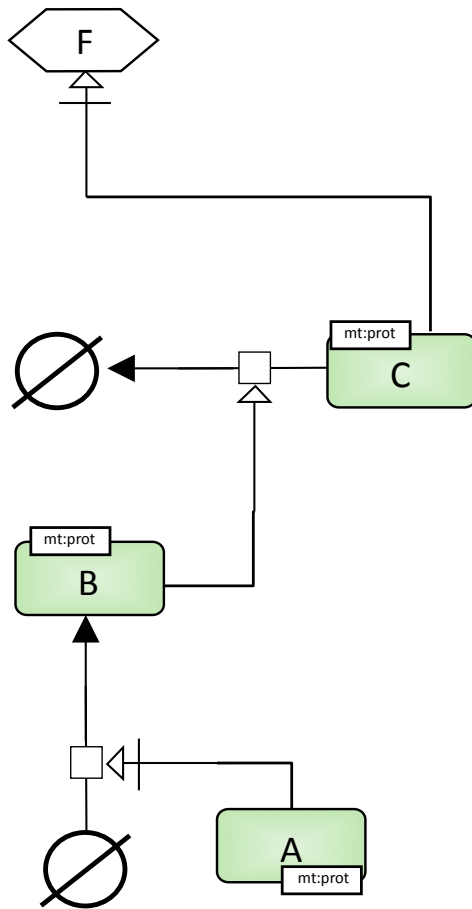
- Edinburgh Pathway Editor
 - www.pathwayeditor.org
- Vanted
 - vanted.ipk-gatersleben.de
- CellDesigner
 - www.celldesigner.org
- Arcadia
 - www.arcadiapathways.sf.net

SOLUTIONS AND QUESTIONS

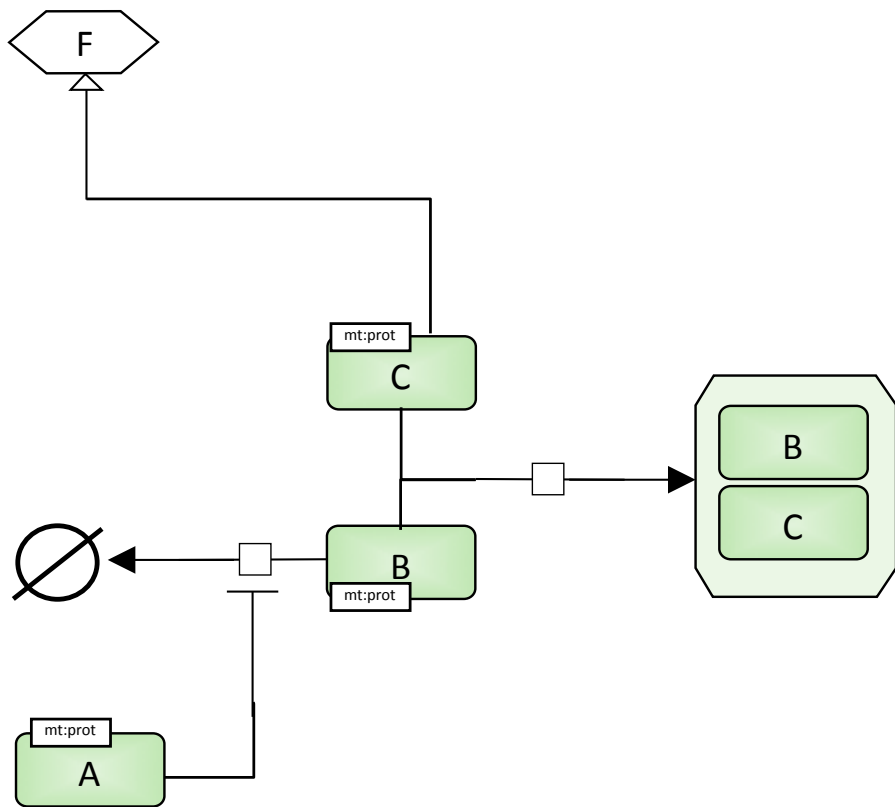
SBGN PD



- Protein C activity in triggering phenotype F is promoted by Protein A through marking its inhibitor B for degradation
- Protein A inhibits phenotype F by triggering gene B, which product degrade protein C that required for F to take place.
- Protein A protects protein B from degradation and that makes protein C being sequestered by B and prevents stimulation of phenotype F

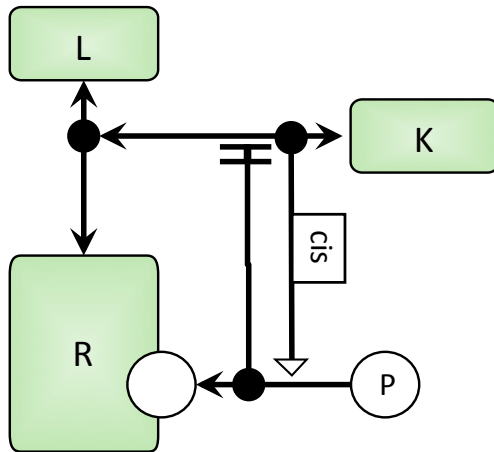


- Protein C activity in triggering phenotype F is promoted by Protein A through marking its inhibitor B for degradation
- Protein A inhibits phenotype F by triggering gene B, which product degrade protein C that required for F to take place.
- Protein A protects protein B from degradation and that makes protein C sequestered by B and prevents stimulation of phenotype F

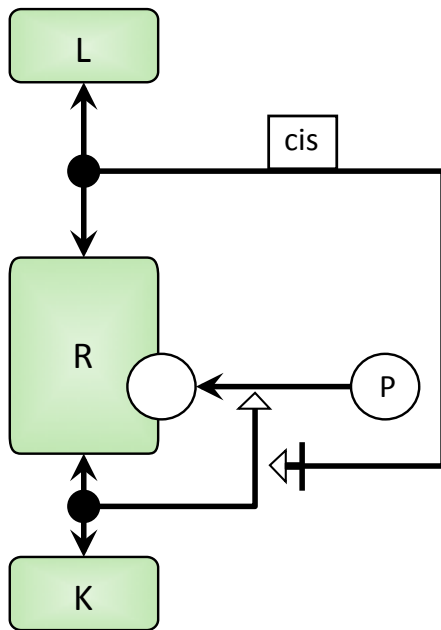


- Protein C activity in triggering phenotype F is promoted by Protein A through marking its inhibitor B for degradation
- Protein A inhibits phenotype F by triggering gene B, which product degrade protein C that required for F to take place.
- **Protein A protects protein B from degradation and that makes protein C sequestered by B and prevents stimulation of phenotype F**

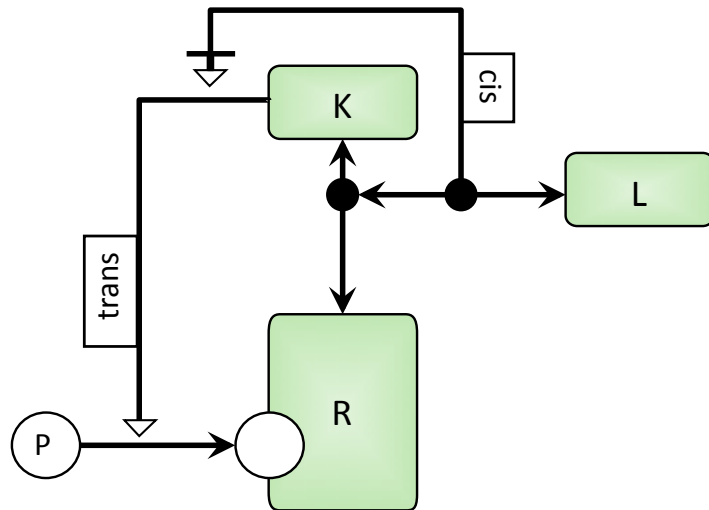
SBGN ER



- Binding of ligand L to receptor R sequesters kinase K to the complex and causes phosphorylation of receptor. Kinase can not bind to phosphorylated receptor and leaves the complex
- Kinase K and ligand L can bind receptor independently but only in the presence of ligand kinase is able to phosphorylate receptor
- Ligand L binding to receptor-kinase complex (RK) makes possible phosphorylation of another molecule of receptor R by kinase K.

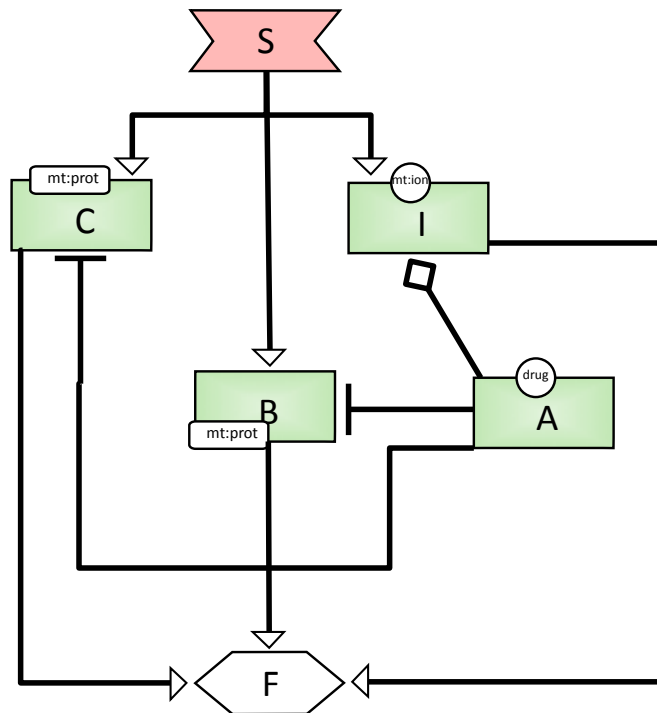


- Binding of ligand L to receptor R sequesters kinase K to the complex and causes phosphorylation of receptor. Kinase can not bind to phosphorylated receptor and leaves the complex
- Kinase K and ligand L can bind receptor independently but only in the presence of ligand kinase is able to phosphorylate receptor
- Ligand L binding to receptor-kinase complex (RK) makes possible phosphorylation of another molecule of receptor R by kinase K.



- Binding of ligand L to receptor R sequesters kinase K to the complex and causes phosphorylation of receptor. Kinase can not bind to phosphorylated receptor and leaves the complex
- Kinase K and ligand L can bind receptor independently but only in the presence of ligand kinase is able to phosphorylate receptor
- Ligand L binding to receptor-kinase complex (RK) makes possible phosphorylation of another molecule of receptor R by kinase K.

SBGN AF

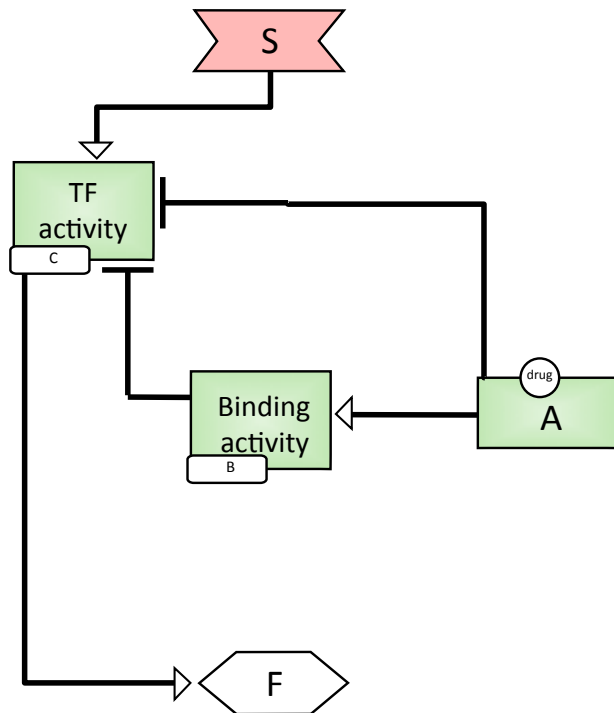


- Drug A inhibits phenotype F associated with stress S by modulating intracellular level of ion I and inhibiting protein B and C activation

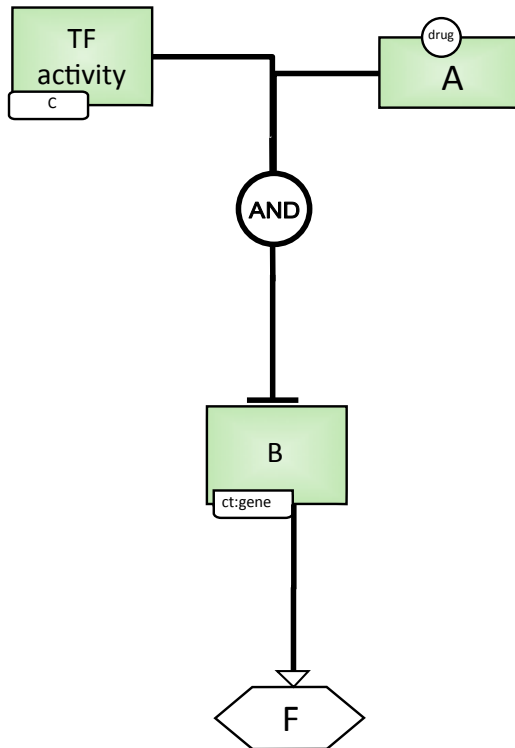
- Drug A was shown to reduce response of the system to stress S by inducing protein B and C association, which reduces protein C activity.

Simultaneously drug A block protein C trans-activation domain that also inhibits phenotype F

- Drug A was shown to reduce transcriptional activation and expression of gene B in protein C dependant manner which might contribute to inhibition of phenotype F.



- Drug A inhibits phenotype F associated with stress S by modulating intracellular level of ion I and inhibiting protein B and C activation
- Drug A was shown to reduce response of the system to stress S by inducing protein B and C association, which reduces protein C activity. Simultaneously drug A block protein C trans-activation domain that also inhibits phenotype F
- Drug A was shown to reduce transcriptional activation and expression of gene B in protein C dependant manner which might contribute to inhibition of phenotype F.



- Drug A inhibits phenotype F associated with stress S by modulating intracellular level of ion I and inhibiting protein B and C activation
- Drug A was shown to reduce response of the system to stress S by inducing protein B and C association, which reduces protein C activity. Simultaneously drug A block protein C trans-activation domain that also inhibits phenotype F
- Drug A was shown to reduce transcriptional activation and expression of gene B in protein C dependant manner which might contribute to inhibition of phenotype F.