

# Utmaningar för bioinformatiken inom industri och akademi

Per Kraulis  
Biovitrum AB

Industriakademin, 27 okt 2005

# Landskapet 1

- Homo sapiens genomet klart (nåja)
- Flera mammalie-genom, i användbart skick
- 1000 bakterie-genom (snart)
- Genidentifiering, annotering
  - Proteinkodande gener identifierade
  - Sköts av stora centra: inget för mindre grupper
  - RNA-gener och reglering: mycket att göra

# Landskapet 2

- Expressionsanalys
  - Mognande teknik och analysmetoder
  - Nya applikationer/analyser?
- Proteomik
  - Stora dataset, nya typer av data
  - Ex: HPR [www.proteinatlas.org](http://www.proteinatlas.org)
  - Mycket att göra

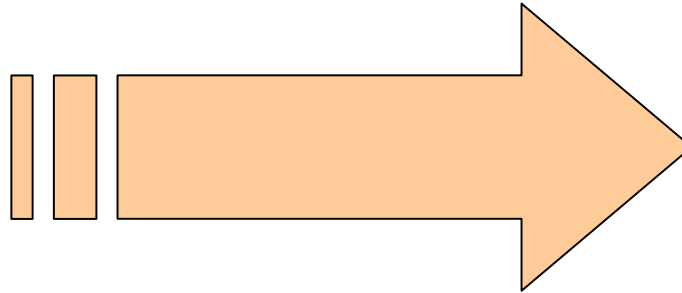
# Landskapet 3

- Nätverksanalys
  - Signalering
  - Metabola nätverk
  - Databaser, litteratur (text mining)
- Systembiologi
  - Modeller av mekanismer, simulering
  - Förklaring av förlopp
  - Men: Prediktion! Varför så skralt?

# Förändring av fokus

## Struktur

- genom
- gener
- proteinsekvenser
- proteindomäner

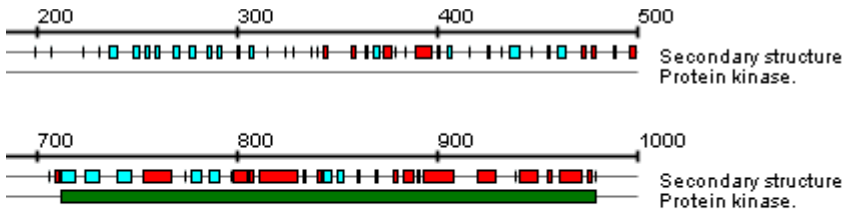


## Processer

- förlopp
- nätverk
- signallering
- metabolism

# Bioinformatics

## Integrated Pharmacology, Biovitrum

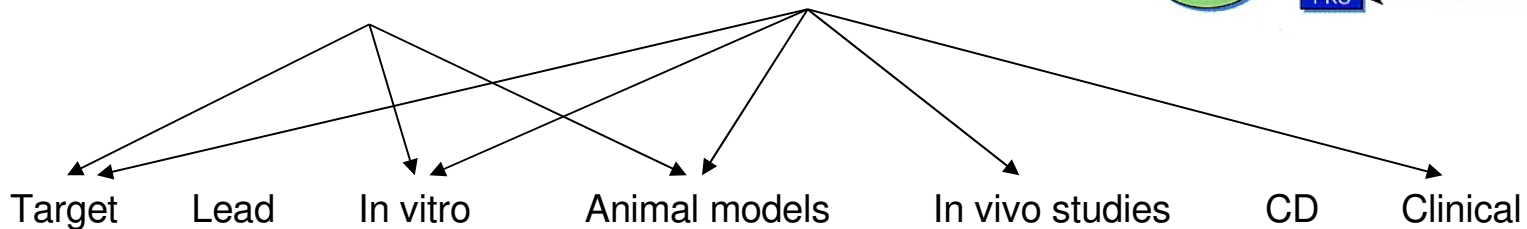
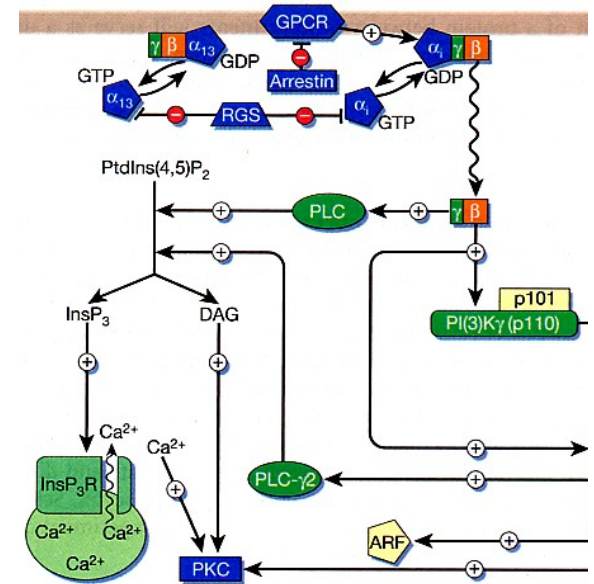


### Sequence analysis

- Homologs
- Orthologs
- Splice variants
- Expression patterns

### Pathway analysis

- System effects
- Study design
- Readouts
- Biomarkers



Drug discovery pipeline

# Tes 1: Sekvensorienterad bioinformatik är rutin

- Annotering finns i publika DB
- Verktyg finns tillgängliga
- Få uppenbara möjligheter till “lyft”
- Undantag
  - RNA
  - Fylogenetiska jämförelser

## Tes 2: Vissa behov ej uppfyllda

- Saknas: Annoteringssystem för små grupper med intresse för specifika gener/system
- Existerande produkter är “imperialistiska”
- Idéer:
  - Modell: Dossier eller ‘best current view’
  - Editera: web eller specialverktyg
  - Läs: web eller PDF



# Tes 3: Nya 'drug targets' från biologi, funktion (inte sekvens)

- 'Drug target hunting' är passerat kapitel
- Tillbaka till cellbiologi, farmakologi, mm
- Hur kan bioinfo hjälpa experimentalisten?
  - Ordna, systematisera litteraturen
  - Designa experiment
  - Välja 'read-outs'
  - Handskas med data (DB motsv)

# Tes 4: (Bio)informatik krävs för systembiologi

- Mekanistiska modeller standard (nåja)
  - SBML, Reactome, KEGG, etc
- Men förloppen som ska simuleras?
  - Datamodeller/databaser saknas!
    - Initialvärden, randvillkor
    - Kontext
    - Dynamiska förändringar

# Tes 5: Bioinformatiken måste ta sig an biologiska förlopp

- Förlopp (processer) är biologins hjärta
- Den temporala aspekten är central
  - Ex: Vad händer när en cell stimuleras?
  - Ex: Cell-cykeln: vilka komponenter, processer?
- Få databaser/datamodeller!
- Jmf: Geographical Information Systems (GIS), temporal aspekter forskas kring sedan 15 år

# Proto-Systems biology?

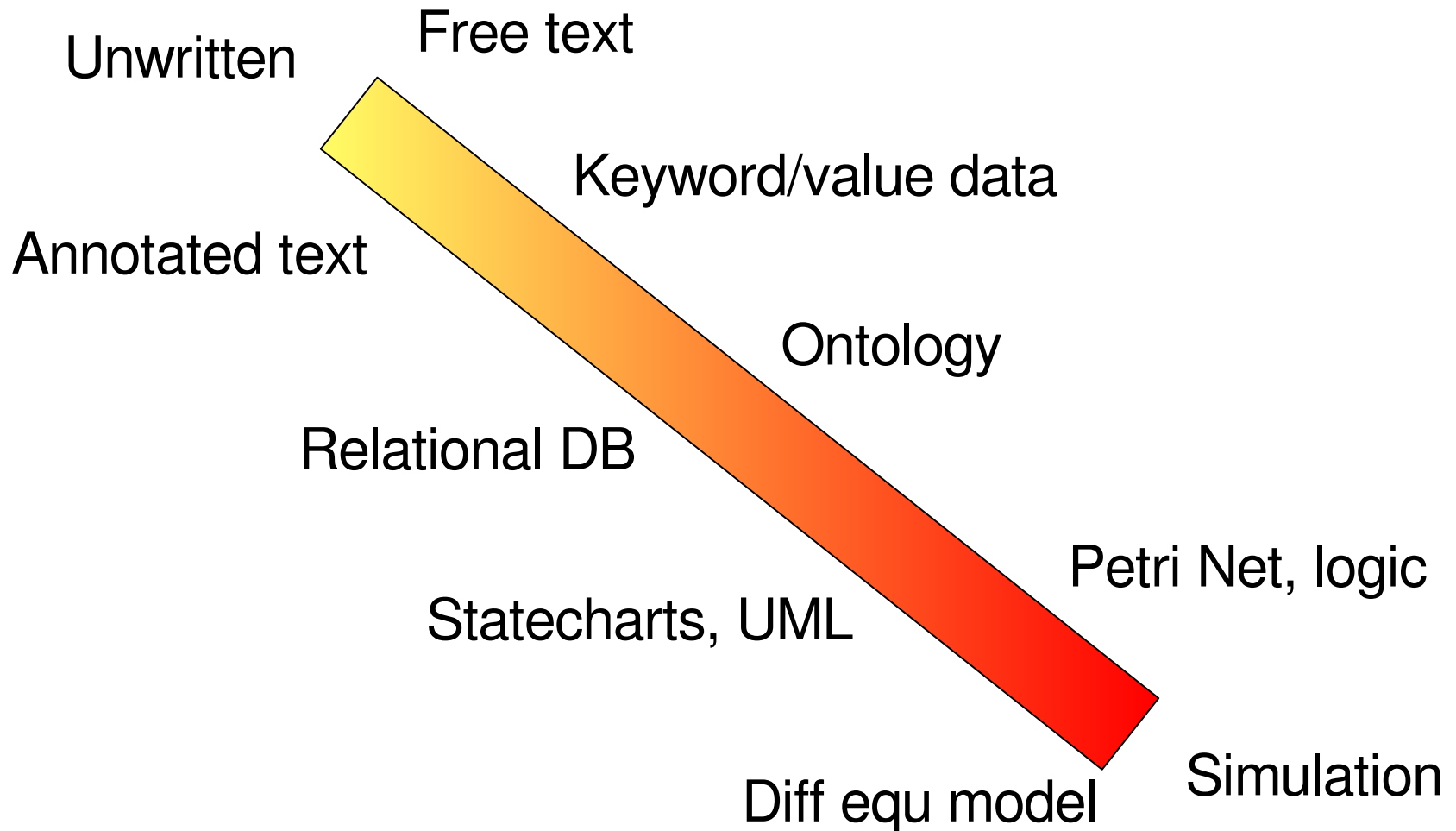
If sufficient regularity can be found between molecular entities and logical and informational outcomes to allow **appropriate databases** to be built, then genomic and post-genomic data could be interrogated more effectively.

...

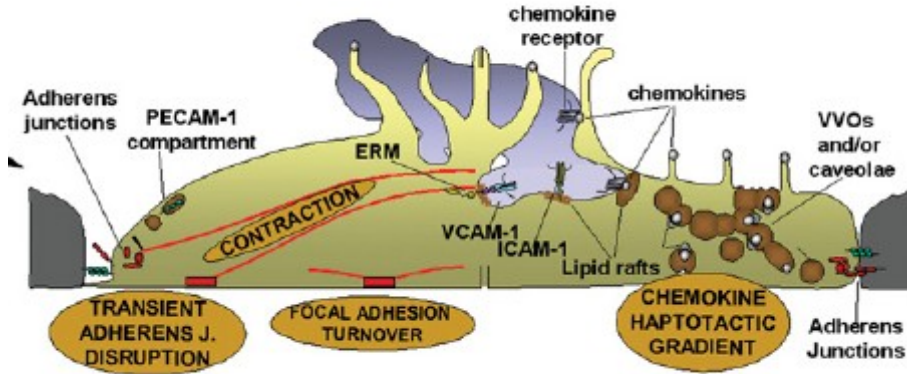
If successful, this approach **would not require detailed kinetic analyses** of all processes within cells, but rather rely on more cursory calculations to study phenomena of interest.

Paul Nurse, "Understanding cells", Nature (424) 2003, 883.

# Computable information



# Multiple levels and types in biology



Millán & Ridley (2005)

- Objects
  - Molecules
  - Complexes
  - Compartments
  - Cells
- Events
  - Reactions
  - Transport
  - Signals
  - Processes

Theory  
Explanation  
Prediction

Molecular  
properties



Macroscopic  
structure,  
dynamics

(Computable) information

# The predictions...

- Roles of uncharacterized components
- Behavior after perturbation
- Suggest points of intervention; **drug targets**



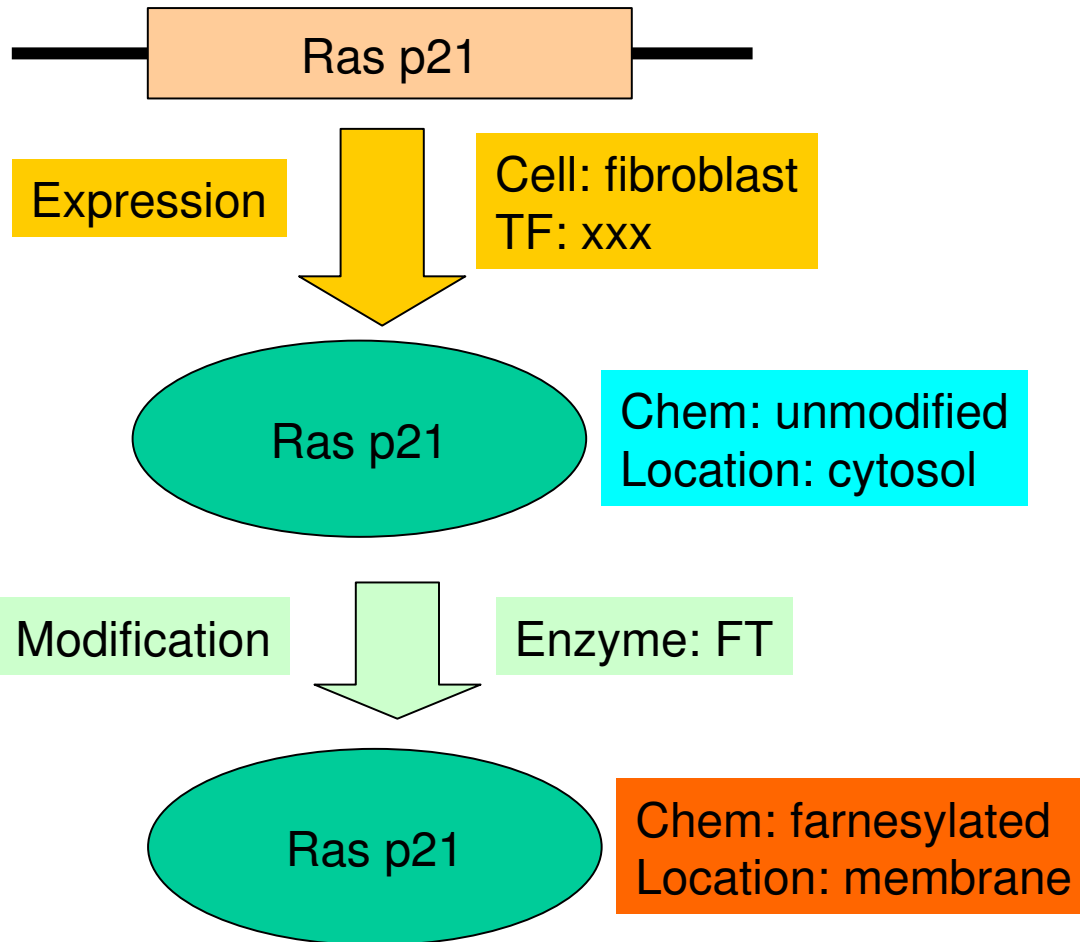
# Part 1: Molecular data

- Molecular components
  - Genomics, transcriptomics, proteomics, etc
- Molecular events
  - Interactions
  - Modifications
  - Localisation
- Explicit data model required for DB!

# Part 2: Macroscopic processes

- Describe macroscopic processes
  - Simulations must be compared with something
  - Goal-oriented description?
    - What is required to achieve a specific state?
- Life processes as projects
  - Goals, milestones
  - Resource usage; scheduling
  - Subprojects, tasks

# GeneCV concepts

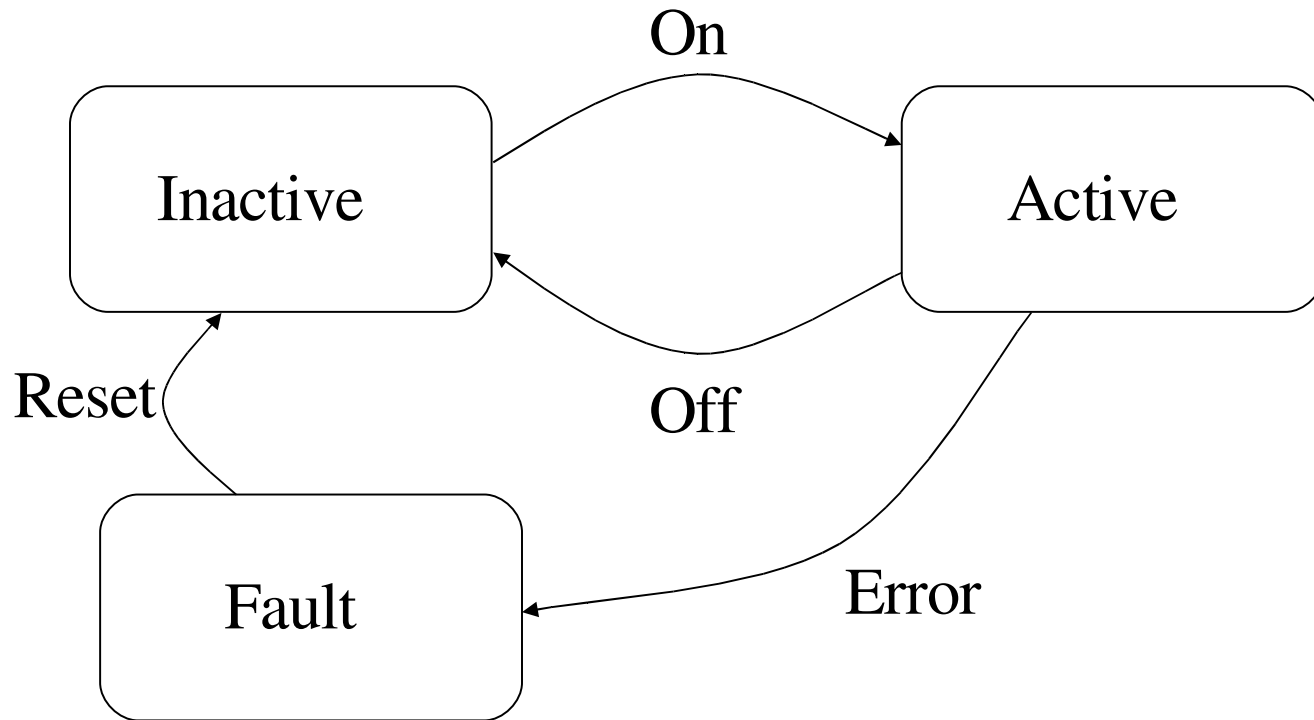


- Entities
  - Genes
  - Proteins
  - Molecules
  - Complexes
- States
  - Complexes, member of
  - Modifications
  - Location
- Transitions
  - Creation
  - Destruction
  - Interactions
  - Regulation
  - Transport

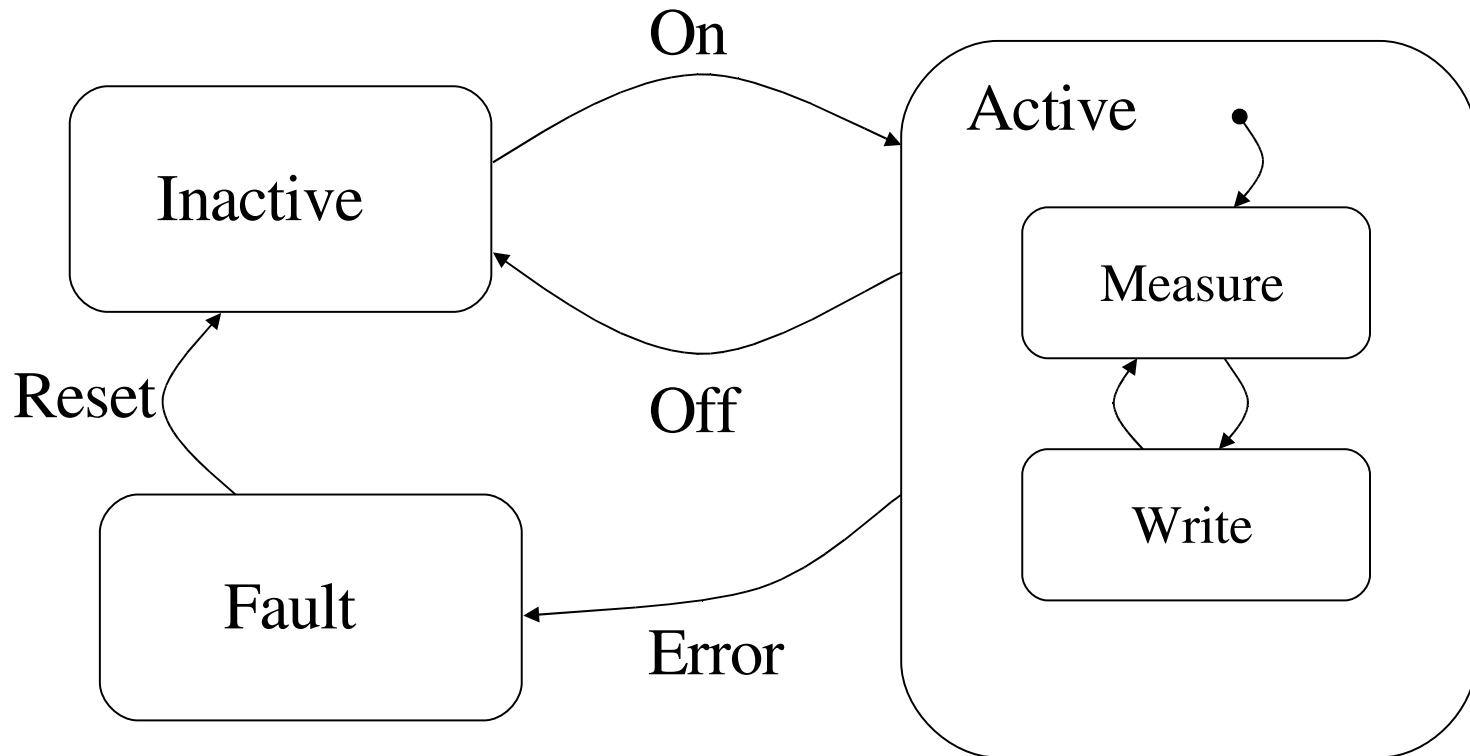
# Statecharts

- David Harel, 1987
- Describe reactive computer systems
  - Event-driven
  - Responding to external and internal stimuli
- State-transition diagrams extended with:
  - Hierarchy
  - Orthogonality
  - Communication
- Now part of UML

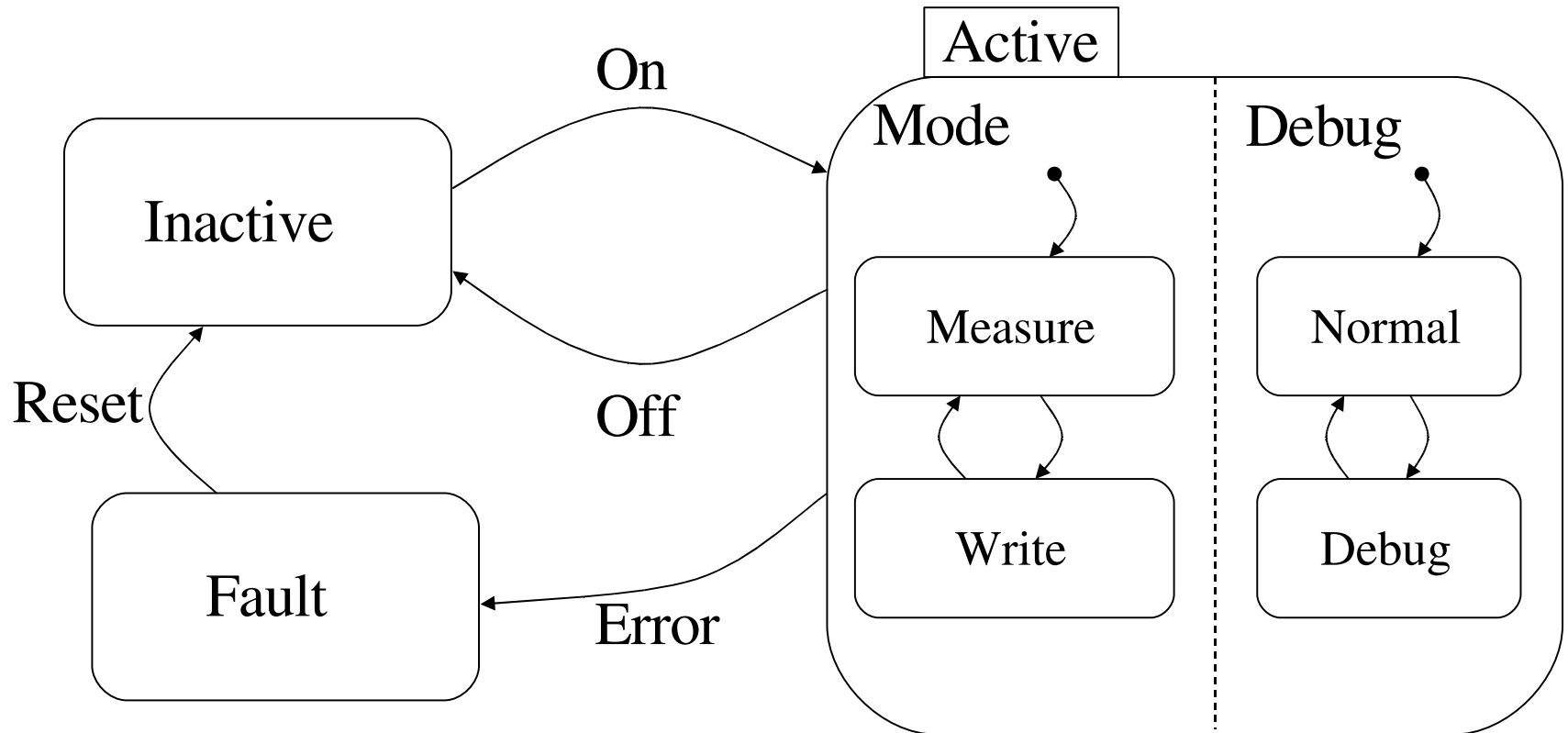
# Statecharts: states and events



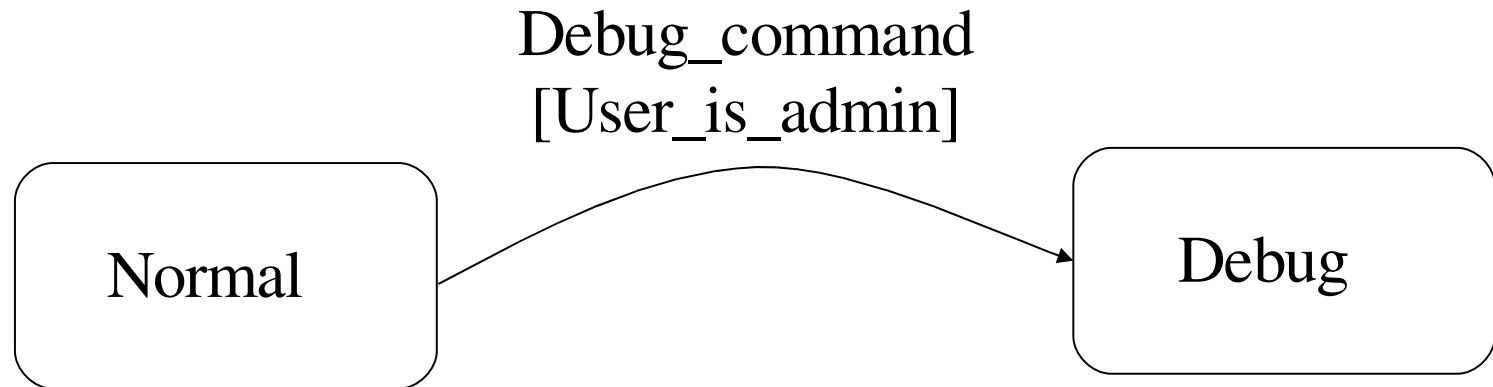
# Statecharts: state hierarchy



# Statecharts: state orthogonality



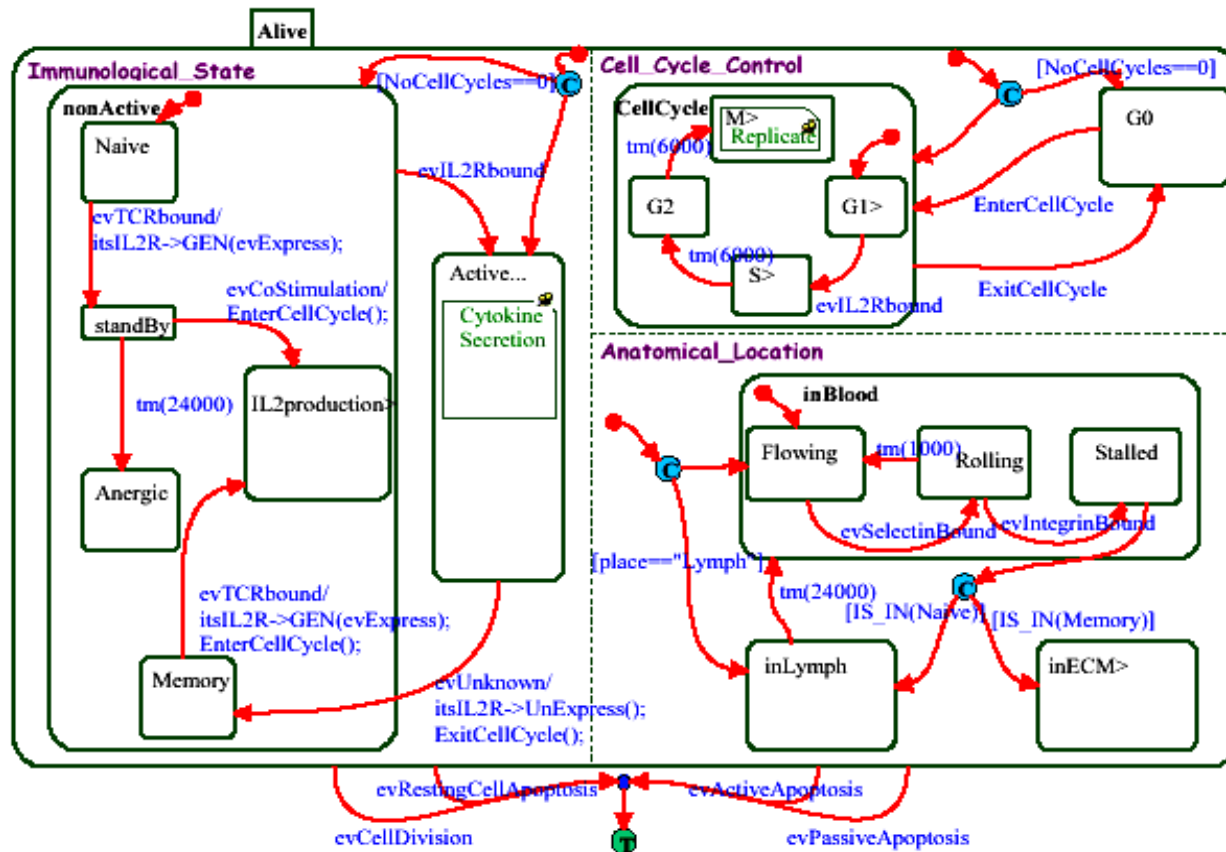
# Statecharts: conditions



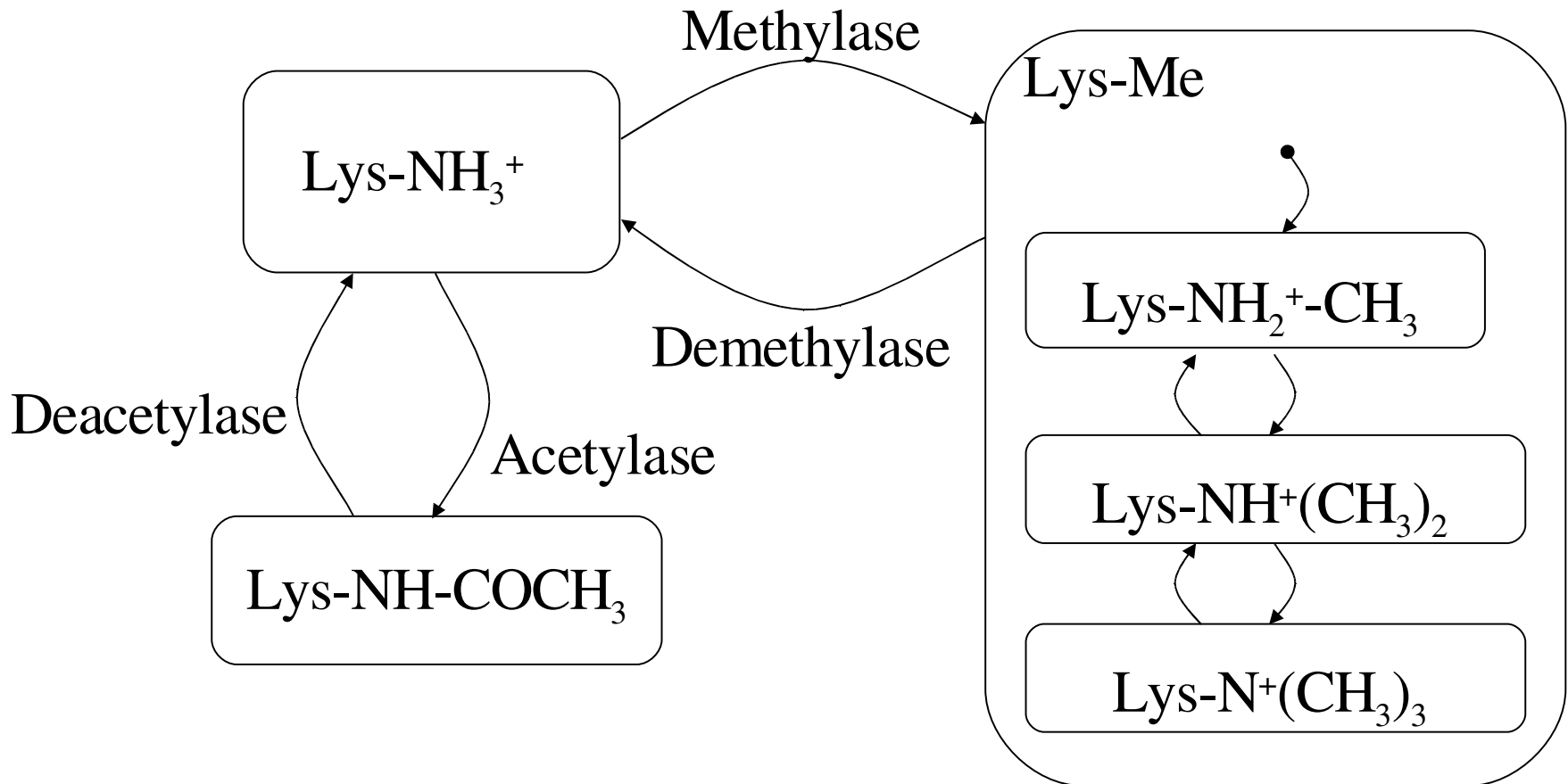


# Modeling T-cell transformations

Kam, Cohen, Harel 2001



# Example: Lysine post-translational modifications



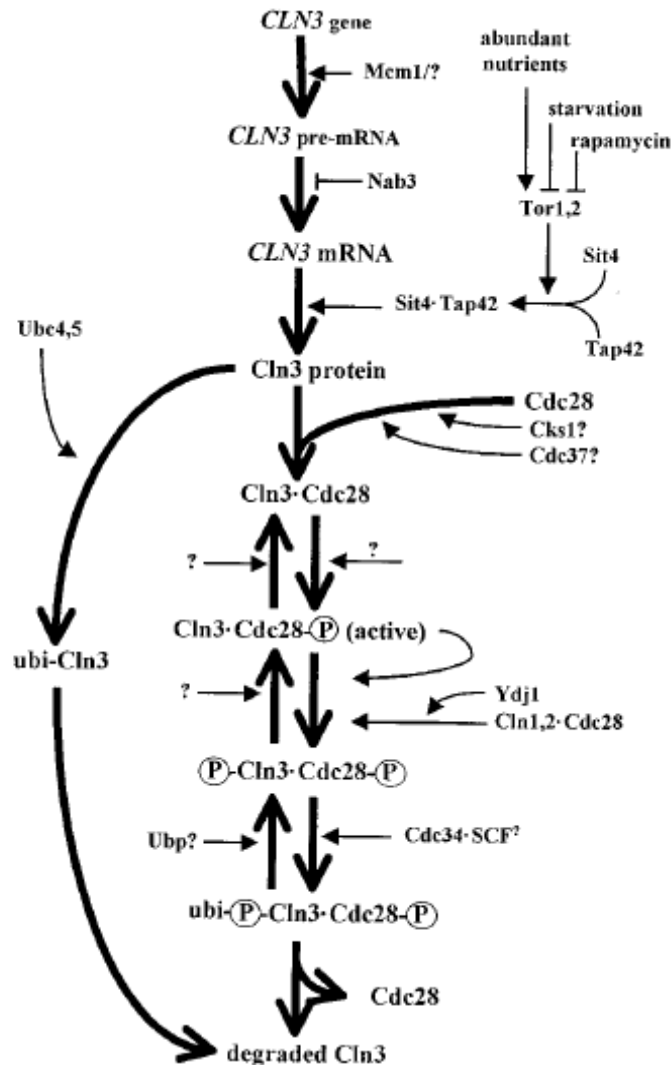
[www.reactome.org](http://www.reactome.org)

- CSHL, EBI, GO collaboration
- Entities
  - Generic/concrete
  - No explicit state; no hierarchy of states
- Events
  - Hierarchy
  - Molecular as well as macroscopic (processes)

# [www.signaling-gateway.org](http://www.signaling-gateway.org)

- Alliance for Cell Signaling, AfCS
- Molecules
  - Proteins
- States
  - No hierarchy
  - Molecular only; complexes are states
  - Location is not state
- Transitions
  - Conditions?

# GeneCV



- The life of a biomolecule
- Molecular data only!
- Creation
- Maturation
- Transport
- Interactions
- Destruction

Mendenhall & Hodge 1998