

The bioinformatics of biological processes

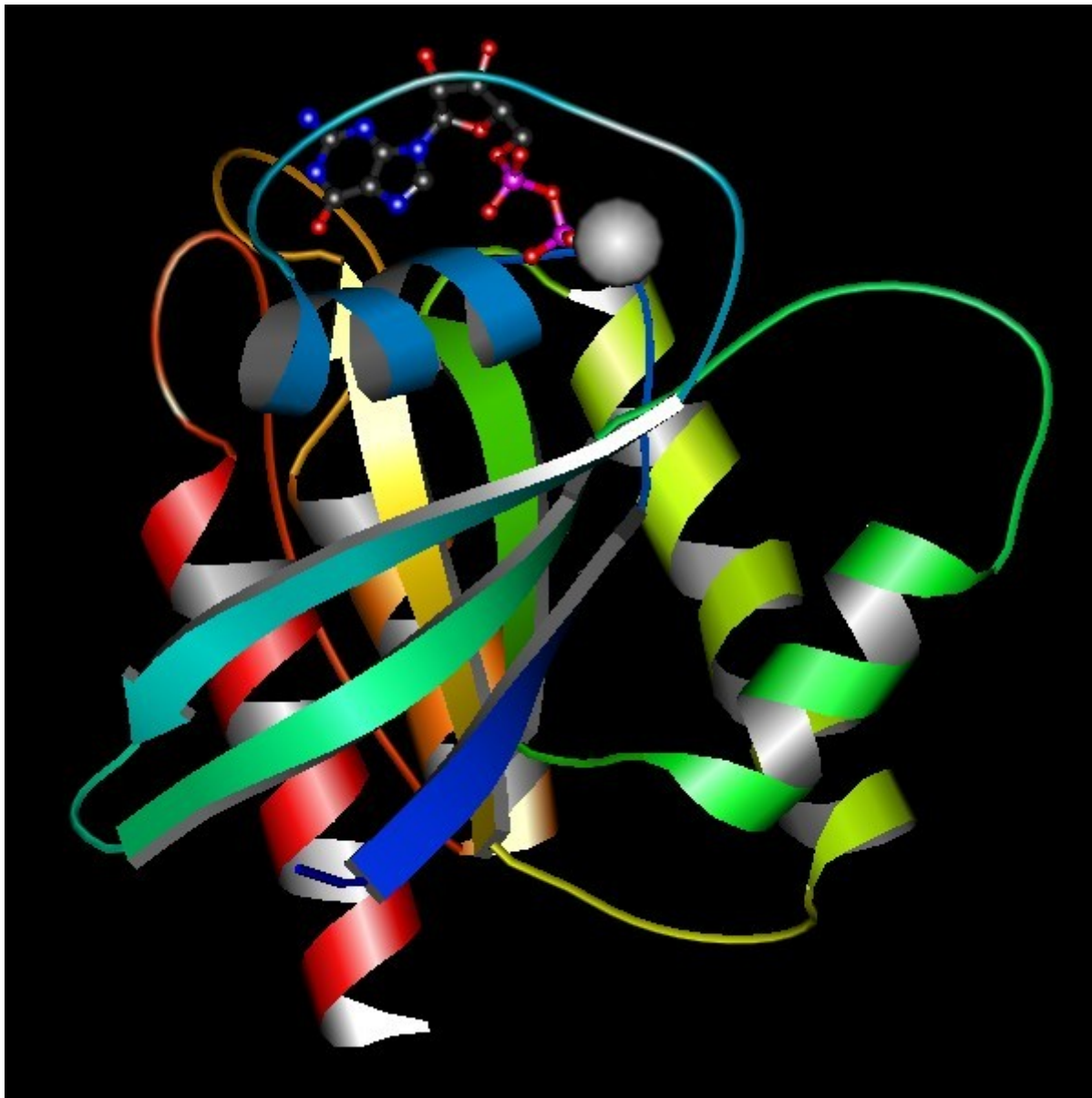
The challenge of temporal data

Per Kraulis

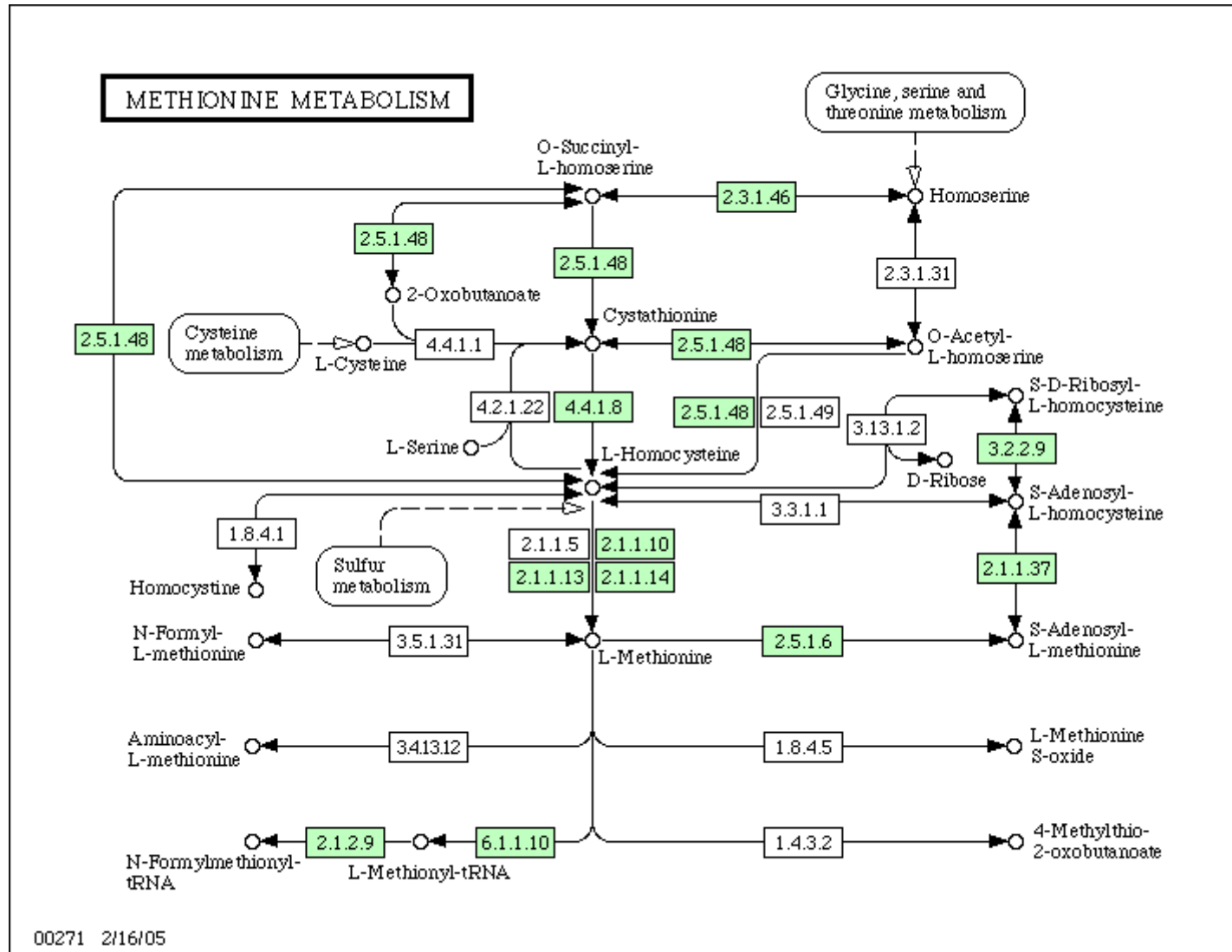
Avatar Software AB

17 UNIPROT: [Q503B6 BRARE](#) 1:189 1:189
 18 UNIPROT: [Q568K0 BRARE](#) 1:189 1:189
 19 UNIPROT: [RASK HUMAN](#) 1:188 1:188
 20 UNIPROT: [Q3UCX0 MOUSE](#) 1:188 1:188
 21 UNIPROT: [RASN MOUSE](#) 1:188 1:188
 22 UNIPROT: [RASK MOUSE](#) 1:188 1:188
 23 UNIPROT: [RASK RAT](#) 1:188 1:188
 24 UNIPROT: [Q4FJP3 MOUSE](#) 1:188 1:188
 25 UNIPROT: [Q9D091 MOUSE](#) 1:188 1:188
 26 UNIPROT: [RASN CHICK](#) 1:188 1:188
 27 UNIPROT: [RASN HUMAN](#) 1:188 1:188
 28 UNIPROT: [Q5U091 HUMAN](#) 1:188 1:188
 29 UNIPROT: [Q2MJK3 PIG](#) 1:188 1:188
 30 UNIPROT: [RASN CAVPO](#) 1:188 1:188
 31 UNIPROT: [Q4S7E9 TETNG](#) 1:188 1:188
 32 UNIPROT: [Q3TMF4 MOUSE](#) 1:188 1:188
 33 UNIPROT: [RASN RAT](#) 1:188 1:188
 34 UNIPROT: [RASN MONDO](#) 1:188 1:188
 35 UNIPROT: [RASN PONPY](#) 1:188 1:188
 36 UNIPROT: [Q57467 ORYLA](#) 1:188 1:188
 37 UNIPROT: [Q13021 XENLA](#) 1:185 1:184
 38 UNIPROT: [Q5EFX7-2](#) 1:188 1:188
 39 UNIPROT: [RASN XENLA](#) 1:188 1:188
 40 UNIPROT: [RASK MSVKI](#) 1:188 1:188
 41 UNIPROT: [RAS CARAU](#) 1:177 1:177
 42 UNIPROT: [Q6DGD1 BRARE](#) 1:186 1:185
 43 UNIPROT: [P01116-2](#) 1:188 1:187
 44 UNIPROT: [RASK MELGA](#) 1:188 1:187
 45 UNIPROT: [RASK CYPCA](#) 1:188 1:187
 46 UNIPROT: [RASK ORYLA](#) 1:188 1:187
 47 UNIPROT: [Q9PSS8 PLAFE](#) 1:188 1:187
 48 UNIPROT: [RASK MONDO](#) 1:188 1:187
 49 UNIPROT: [RASN BRARE](#) 1:186 1:185
 50 UNIPROT: [Q6AZA4 BRARE](#) 1:188 1:187
 consensus/100%
 consensus/90%
 consensus/80%
 consensus/70%

REIROQHKLRLKLNPPDDNGQDCMNCRCVVS
REIROQHKLRLKLNPPDESQDCHS**CRCVVS**
REIROYRLKKISK-EEKTPGCVKIk**cII-**
REIROYRMKLNSSDDGTQGC**HGLPCVL-**
REIROYRLKKLNSSDDGTQGC**HGSPCVL-**
REIROYRLKKISK-EEKTPGCVKIk**cVI-**
REIROYRLKKISK-EEKTPGCVKIk**cVI-**
REIROYRMKLNSSDDGTQGC**HGLPCVL-**
REIROYRMKLNSSDDGTQGC**HGLPCVL-**
REIROYRMKLNNSNEDGNQGC**HGLSCIV-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRMKLNNSDDGTQGC**HGLPCVV-**
REIROYRLNKLSK-EEKTPRCVKIk**cVV-**
REIROYRMKLNSSDDGTQGC**HGLPCVL-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRMKLNSSDDGTQGC**HGLPCVV-**
REIROYRLSKLSK-EEKTPRCVNLK**cVV-**
REIROFRLKKMSK-EEKTPGCVKFK**----**
REIROYRLSKISK-EEKTPGCVQLK**cVV-**
REIHQYRMKKLDSSDNNQGC**IRIPCKL-**
REIROYRLKKISK-EEKTPGCVKIk**cVI-**
REIROYRLRKLKSKEET**-----**
REIRHYRMKLN**SREDRKQGC****LGVSC---**
REIRKHK-EKMSKDG**KKKKKKSKTK****CVI-**
REIRKHK-EKMSKDG**KKKKKKTKTK****CII-**
REIRKHK-EKMSKEG**KKKKKKSKTK****CVL-**
REIRKHK-EKMSKEG**KKKKKKSKTK****CIL-**
REIRKHK-EKMSKEG**KKKKKKSKTK****CSL-**
REIRKHK-EKMSKDG**KKKKKKSKTK****CII-**
REIRHYRMKLN**SREDRKQGC****LGVSC---**
REIRKHK-EKMSKEG**KKKKKKSKTK****CAL-**
REI+pa+ .pKhs . .tct
REIRpa+ .cKhs . .tctt .tp .th .Chl .
REIROa+h+Kls . .--tt .sChth .Cll .
REIROa+h+Kls .s--ps .GChthpCVL .



MolScript: Per Kraulis 1991, 1997



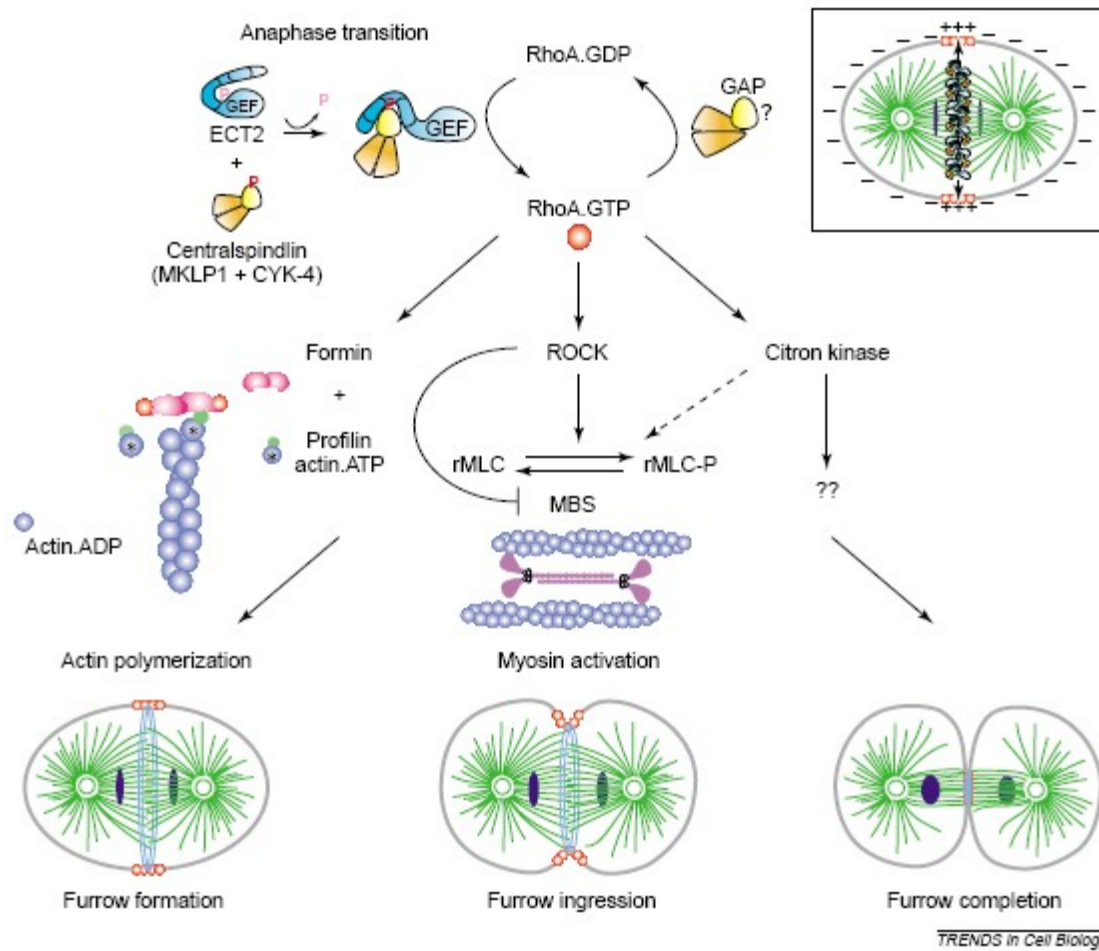
“Classical” Bioinformatics: Static structures

- Structural: Sequence, 3D coordinates
- Time is not involved

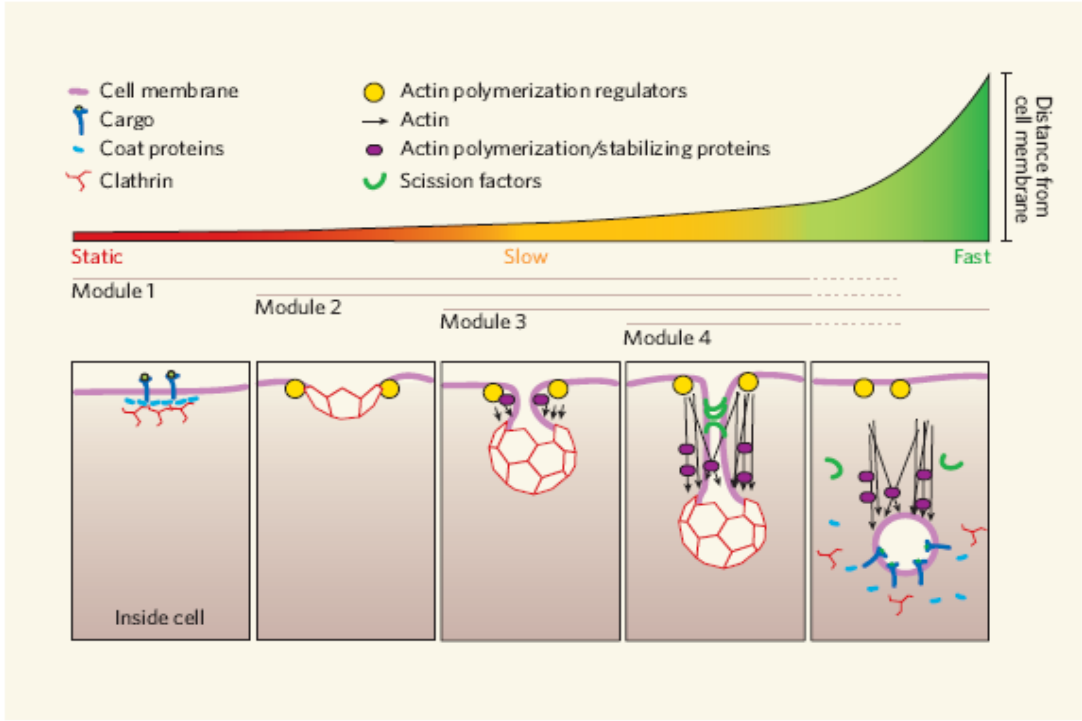
- Why? Easier to work with...
 - Experimentally
 - Conceptually?

Biology is temporal

- Processes are inherently temporal
 - Narrative descriptions in literature
 - Gene expression time series
 - Embryonal development
- Biological processes are goal-oriented
 - Cell cycle: produce another cell



Cytokinesis: Rho regulation
 Piekny, Werner, Glotzer 2005



Endocytic vesicle formation
 Duncan & Payne 2005

But: Few temporal databases?!

- Temporal: 't' as an essential parameter
- Temporal relationships
 - During
 - Before
 - After

"Can computers help to explain biology?"

...biological narratives of cause and effect are readily systematizable by computers.

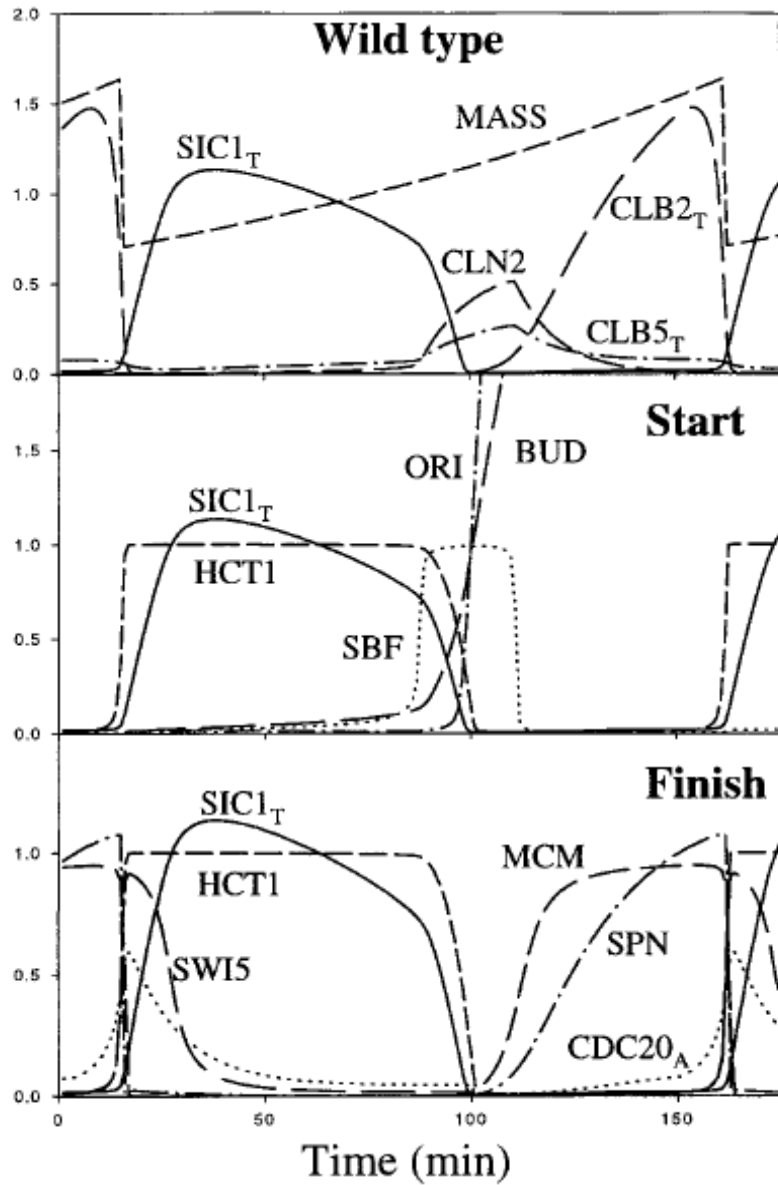
Happily, there is considerable interest in wanting to build one element of biological semantics — the passage of time — into information theory.

[This] might help biologists to go beyond quantifying reaction rates and molecular species of biological systems to understand their dynamic behaviour.

R. Brent & J. Bruck, *Nature* (440) 23 March 2006, 416-417

Computable temporal data

- Required for simulation
 - Initial values
 - To test model against
- Appropriate data model
 - Events during a process
 - Context, preconditions
 - Duration
 - Property = $f(t)$

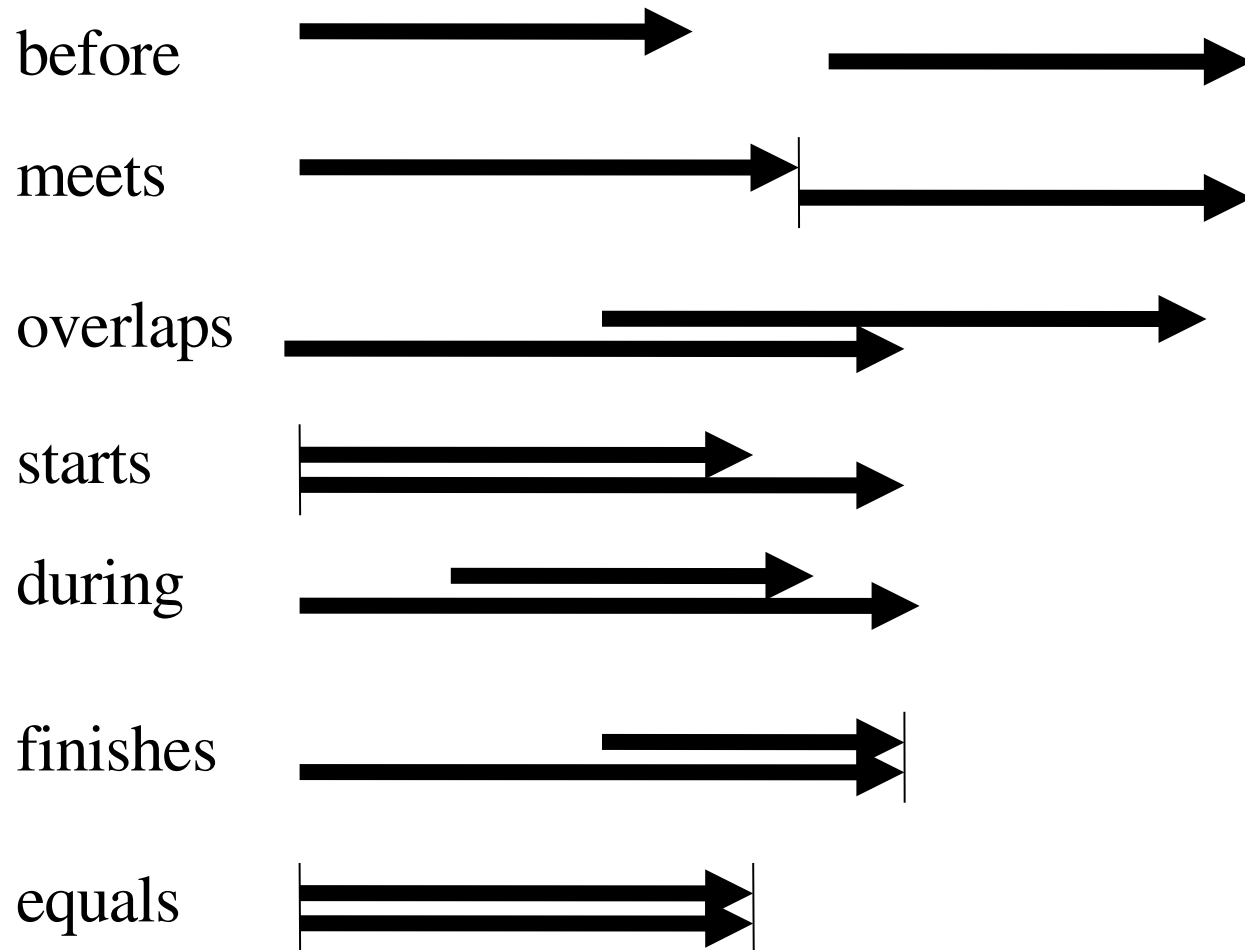


Kinetic analysis of budding yeast cell cycle: Chen et al 2000

Work in other fields

- Geographical Information Science, GIS
- Artificial Intelligence
 - Knowledge Representation
 - Temporal Logic
 - Automated planning, scheduling
- Temporal databases
- Project management

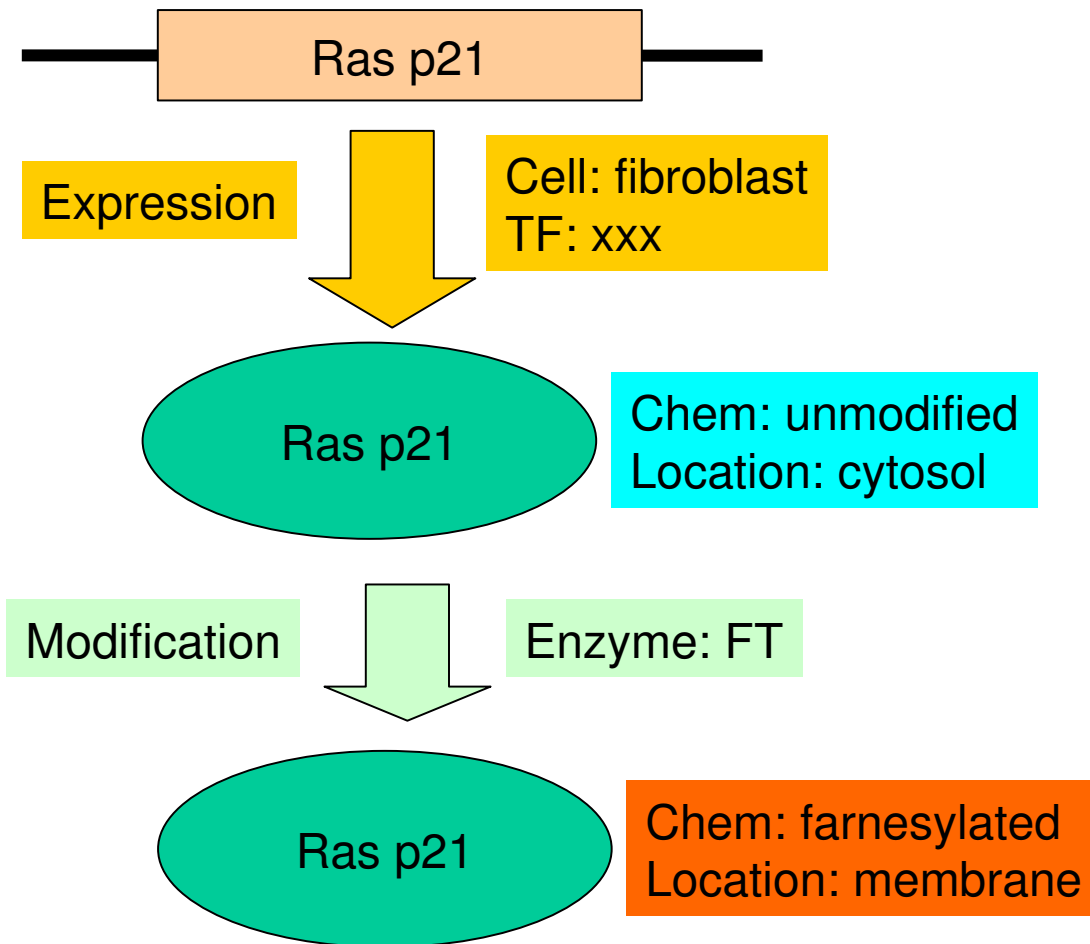
Allen's temporal relationships



Project: BioChronicle

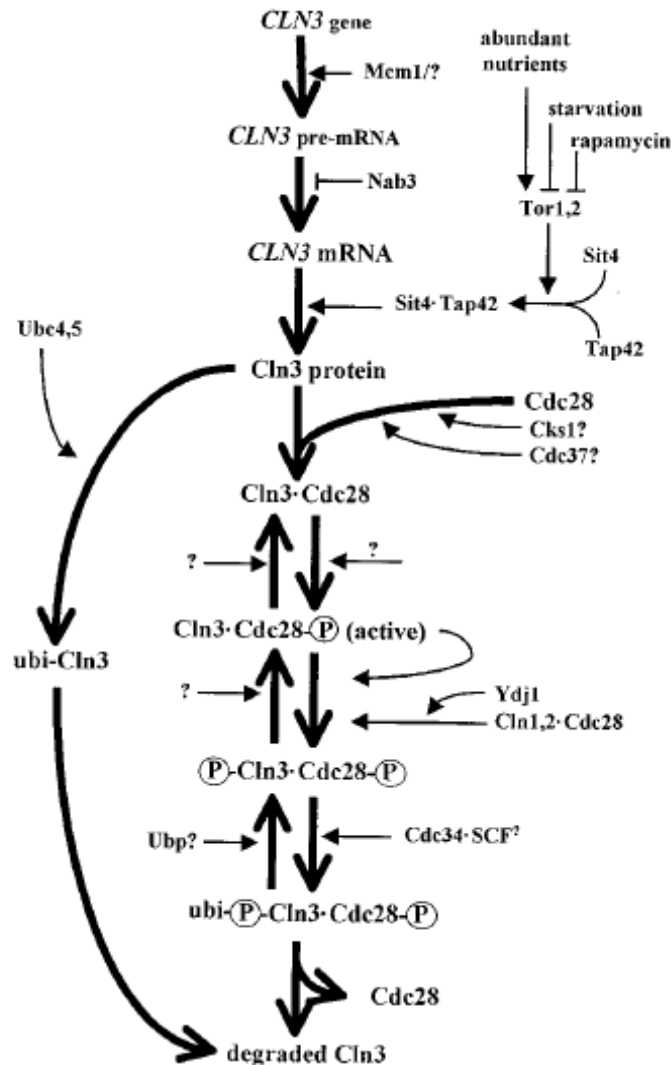
- Data model, implementation
- Handle temporal biological information
 - Events, subevents
 - Relationships
 - Duration
 - Properties, dynamics
 - Preconditions, context
- Test case: Cell cycle

Project: GeneCV



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

GeneCV



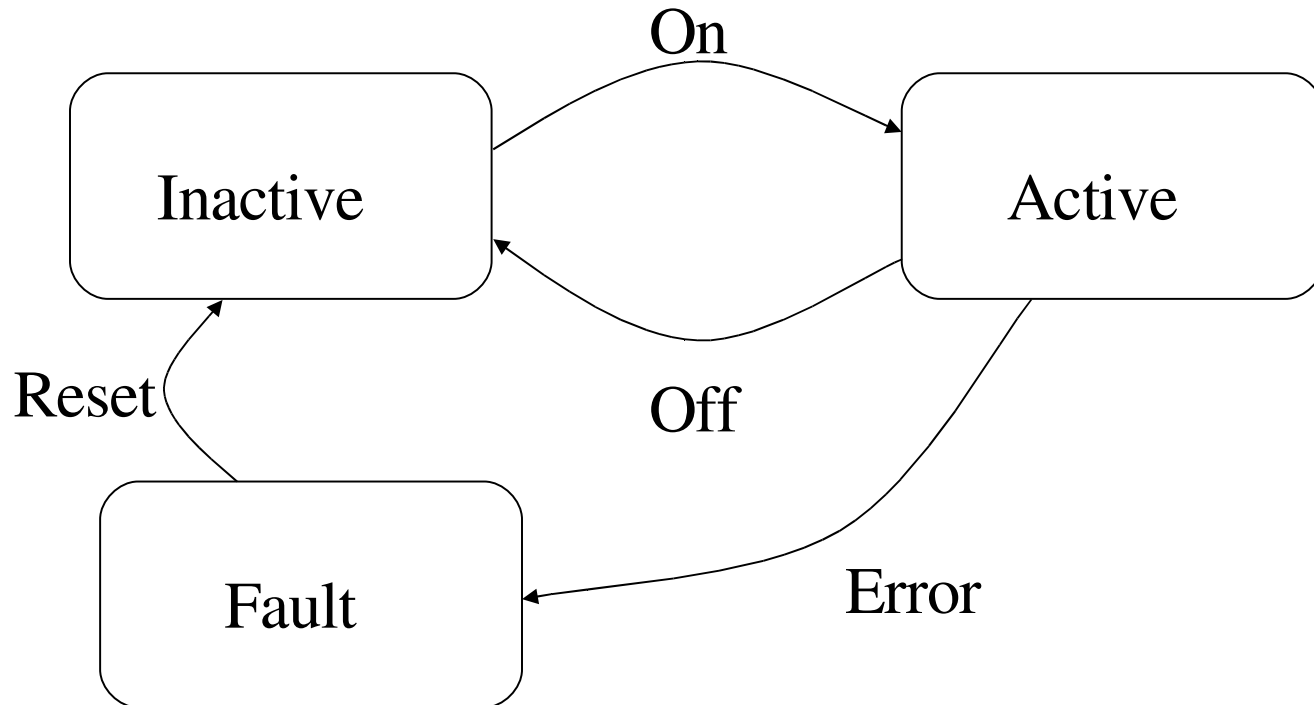
Mendenhall & Hodge 1998

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

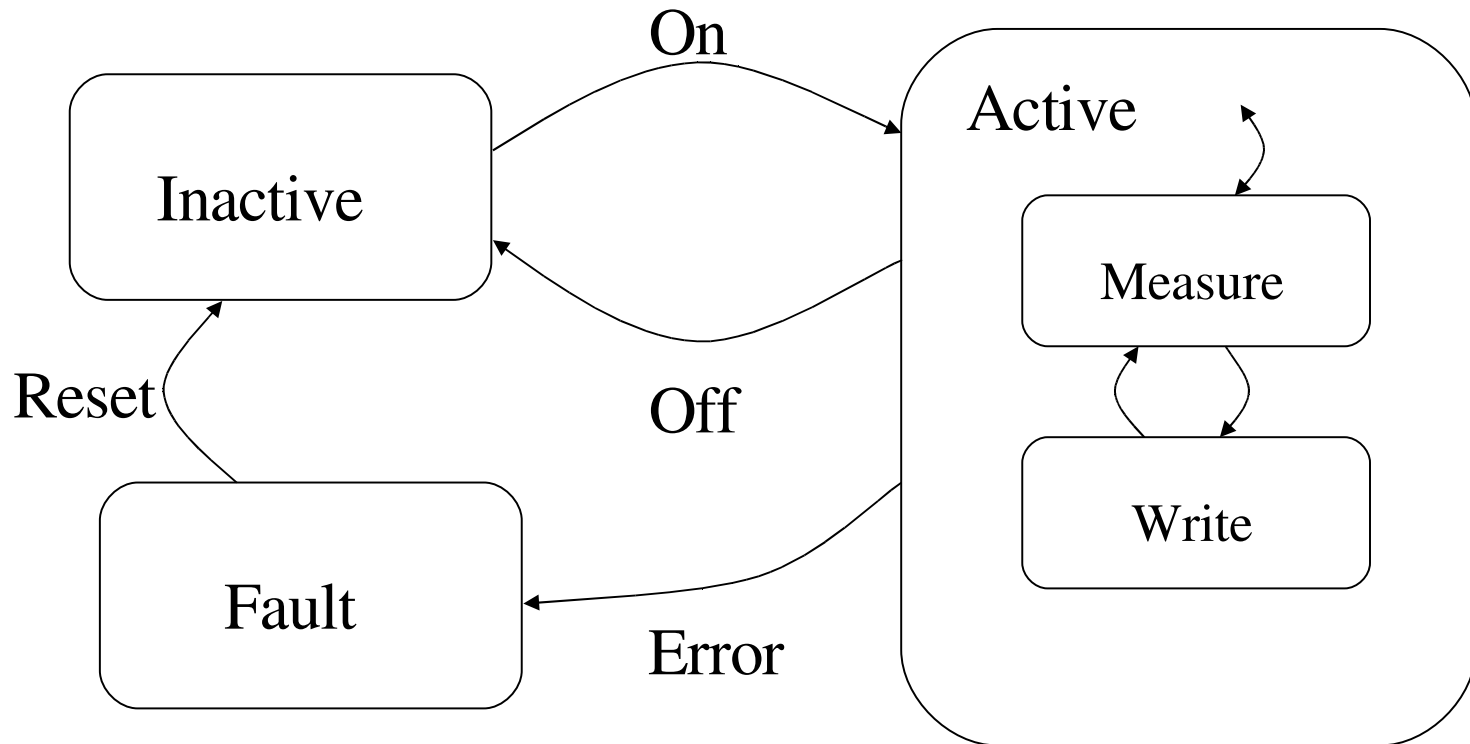
Statecharts

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

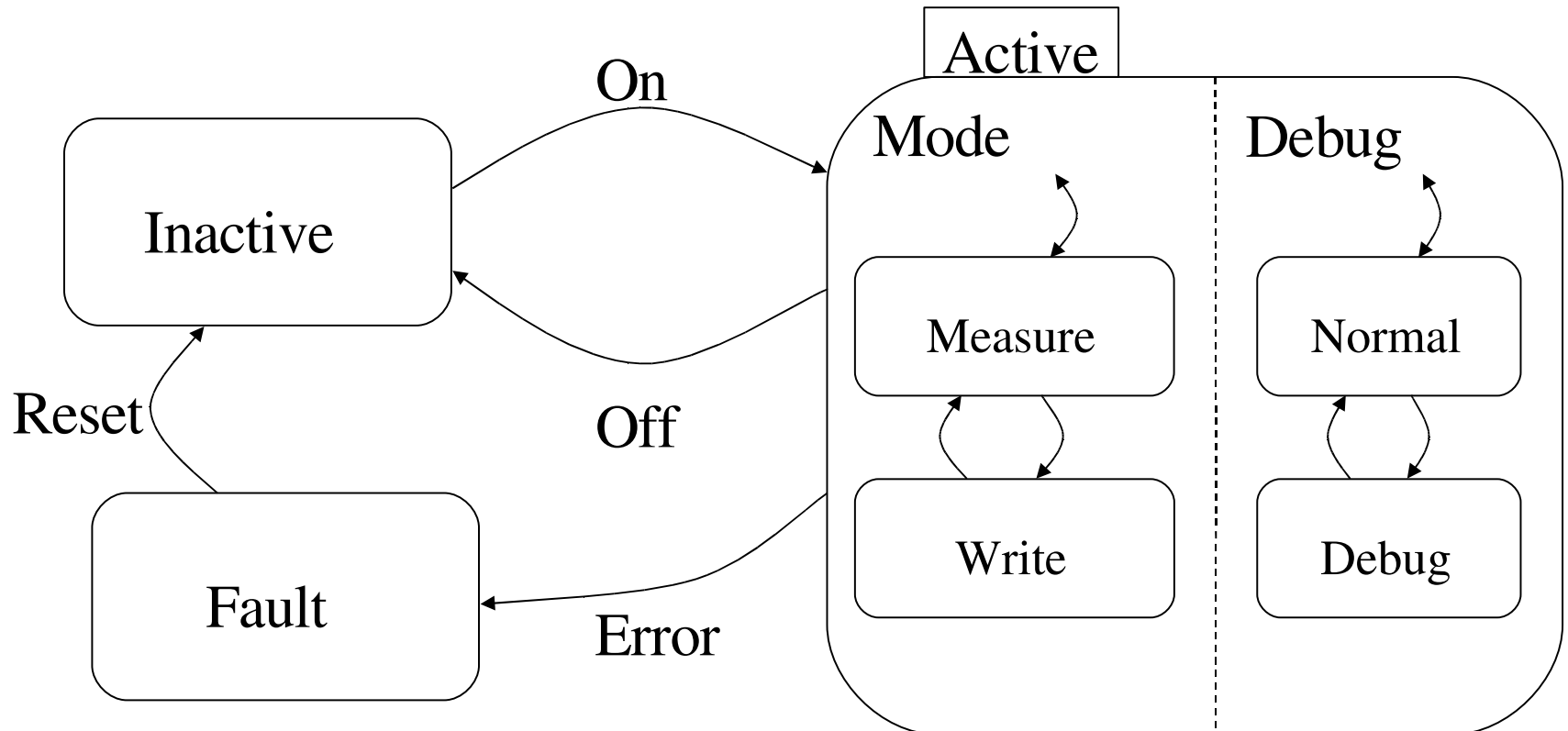
Statecharts: states and events



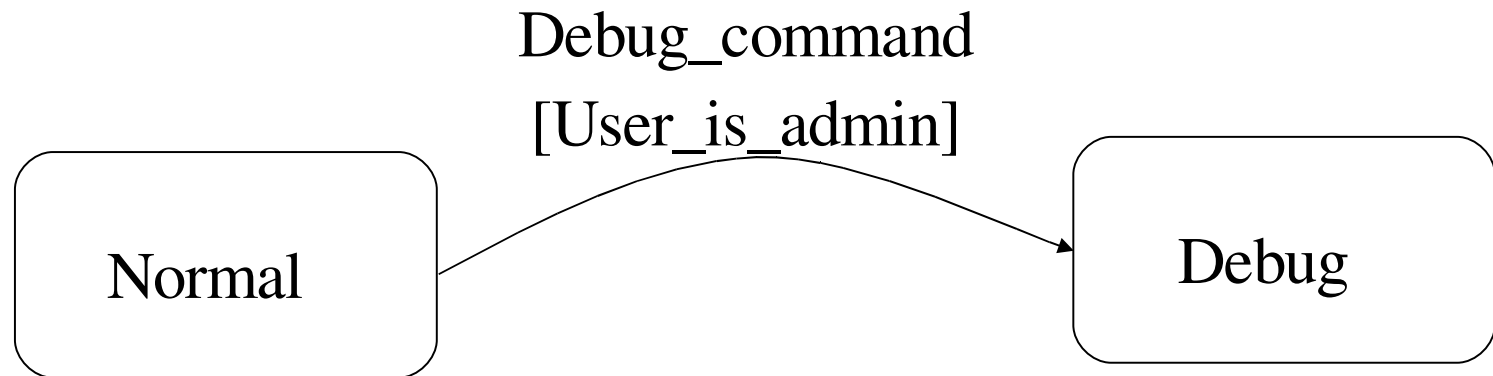
Statecharts: state hierarchy



Statecharts: state orthogonality

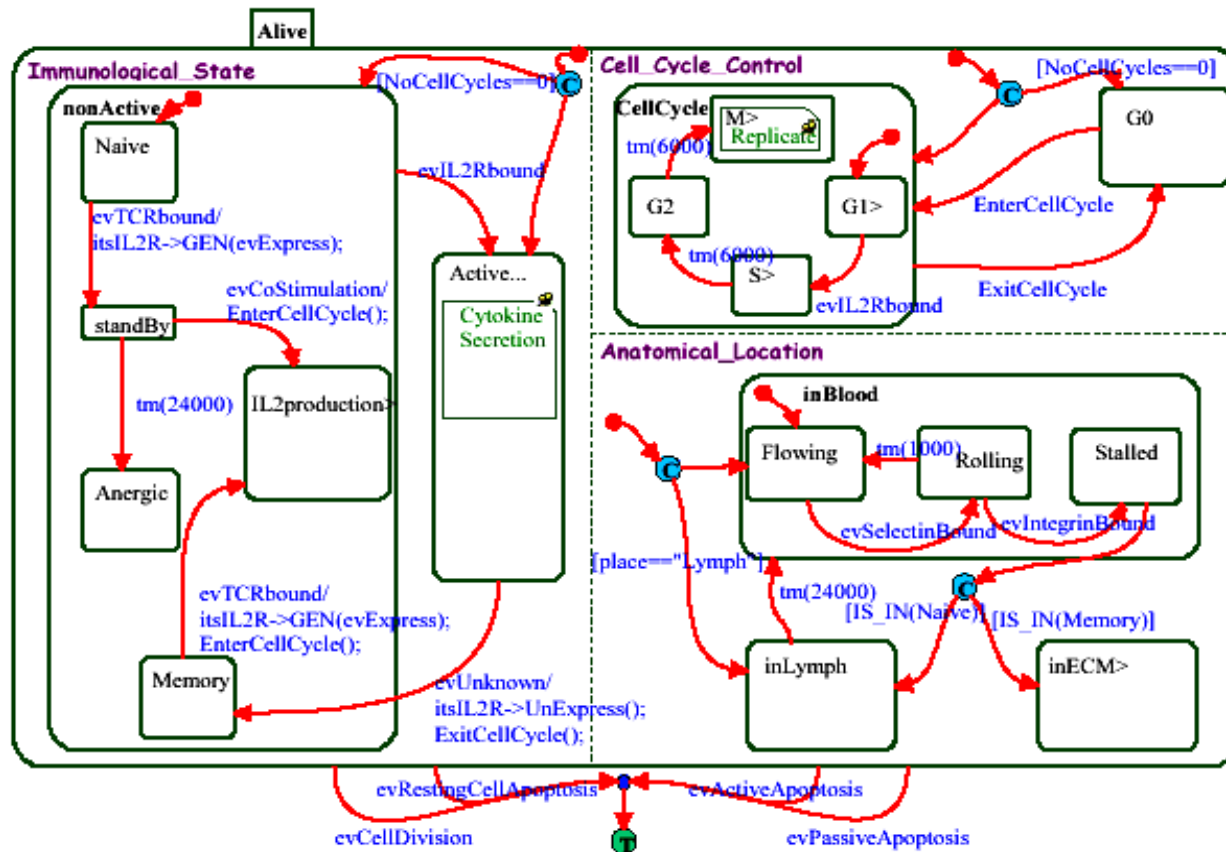


Statecharts: conditions

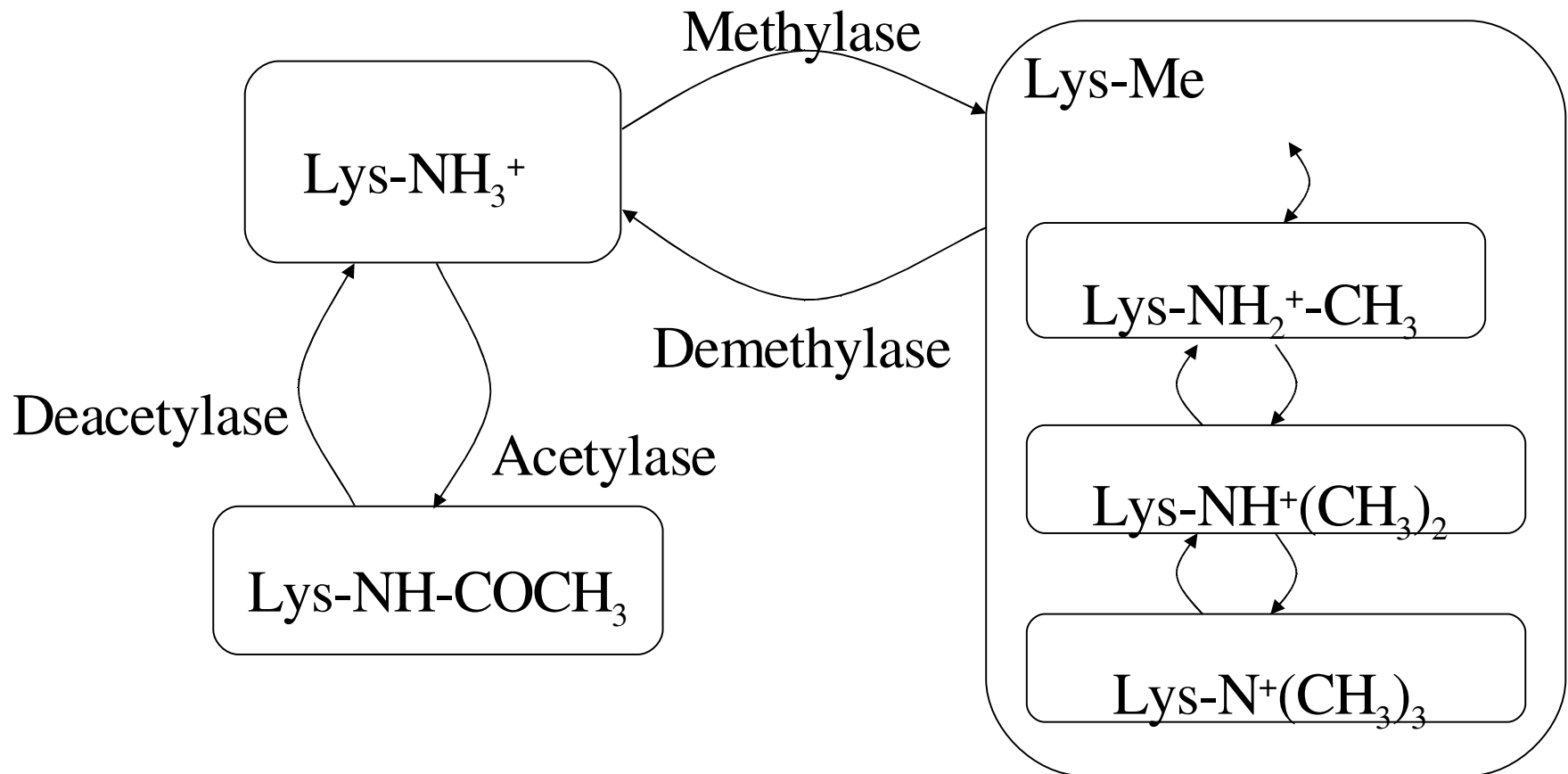


Modeling T-cell transformations

Kam, Cohen, Harel 2001



Example: Lysine post-transl mod's



www.reactome.org

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

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?

Computable information

