The bioinformatics of biological processes

The challenge of temporal data

Per J. Kraulis
CMCM, Tartu University

What is bioinformatics?

- "Information technology applied to the management and analysis of biological data"

  Attwood & Parry-Smith 1999

- "Collection, archiving, organization and interpretation of biological data"

  Thornton 2003

“Classical” bioinformatics

- Sequences
  - Nucleotide
  - Protein
- 3D structure
  - Protein

Classical bioinformatics

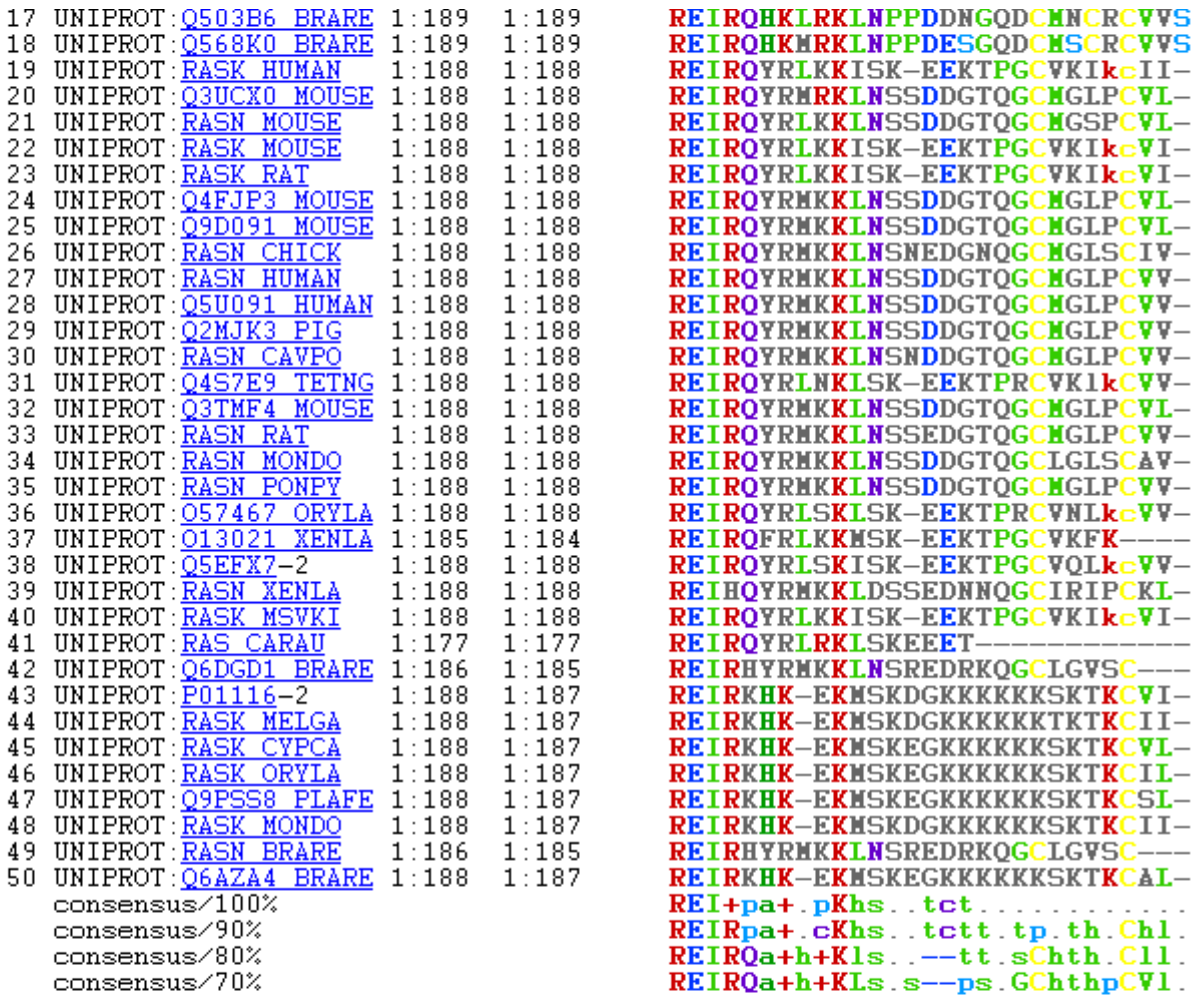
- Handling, storage
  - SwissProt, Ensembl, PDB, etc
- Analysis
  - BLAST, FASTA
  - Patterns, HMMs…
  - Many other methods…

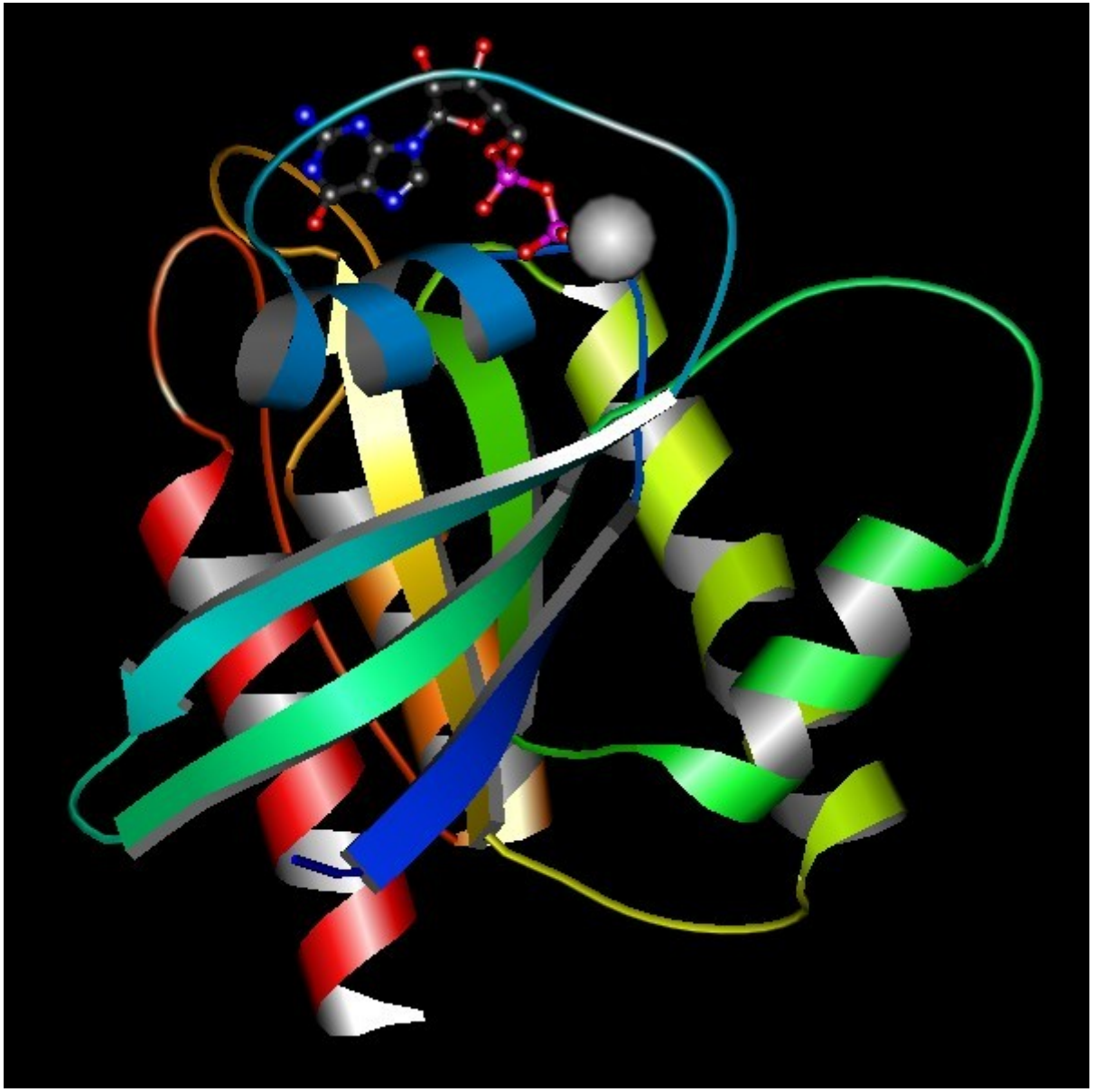
Thus far: Static structures

- Structural: sequence, 3D coordinates
- Static: time is not involved

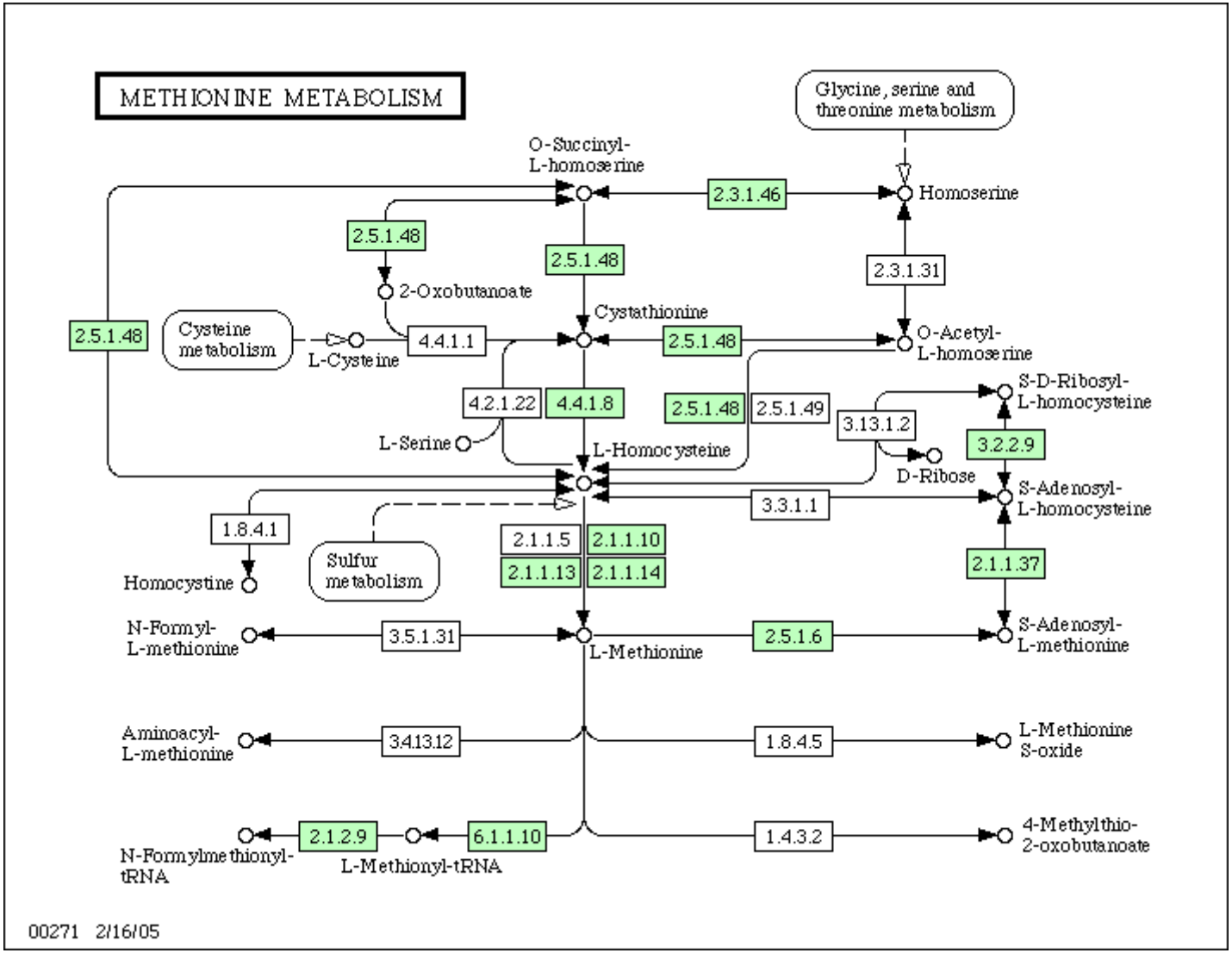
- Easier to work with?
  - Experimentally
  - Conceptually?

Sequence analysis





MolScript: Per Kraulis 1991, 1997



KEGG: Kanehisa 2004

Functional genomics

- The function of genes
  - Catalytic activity
  - Biological function
  - Biological process

- Links between genes/proteins
  - Pathways (metabolic, signaling)
  - Genetic relationships

Gene Ontology

- Annotation
  - Statement of properties
  - Keyword/phrases, controlled vocabulary
  - Comparison, analysis
- Ontology
  - Activity; chemical
  - Localization; spatial
  - Process; functional

Biology is temporal

- Processes are inherently temporal
  - Narrative descriptions in lit
  - Gene expression time series
  - Embryonal development (FunGenes!)

- The goals of biological processes
  - Cell cycle: produce another cell

But: No temporal databases?!

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

Computable temporal data

- Database needed
- Appropriate data model
  - Events during a process
  - Context, preconditions
  - Temporal relationships
  - Duration
  - Property = f(t)

"Can computers help to explain biology?"

Like information in Geographical Information Systems, which also have a limited vocabulary, 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

Work in other fields

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

Knowledge representation I

- Logic: Formal rules of inference
- Ontology: The types of entities
- Computation: Automated analysis

Sowa 2000

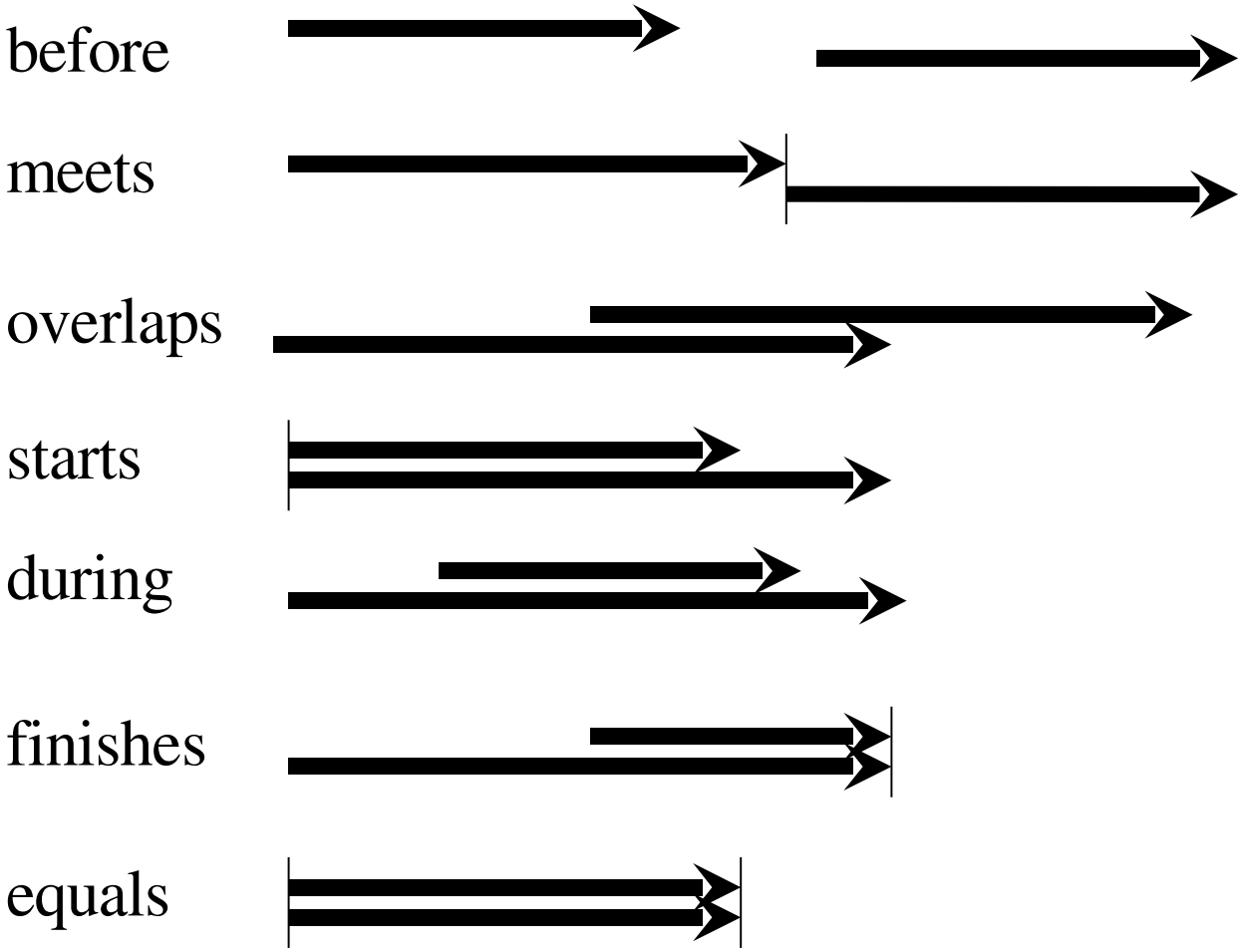
Knowledge representation II

- Artificial Intelligence, AI
  - Reasoning
  - Planning, scheduling
- Philosophy: Aristotle, Kant, Peirce, Whitehead…
- Holy grail: Well-structured and natural
- Fundamental for data model
  - EcoCyc
  - Gene Ontology (GO)

Knowledge Representation

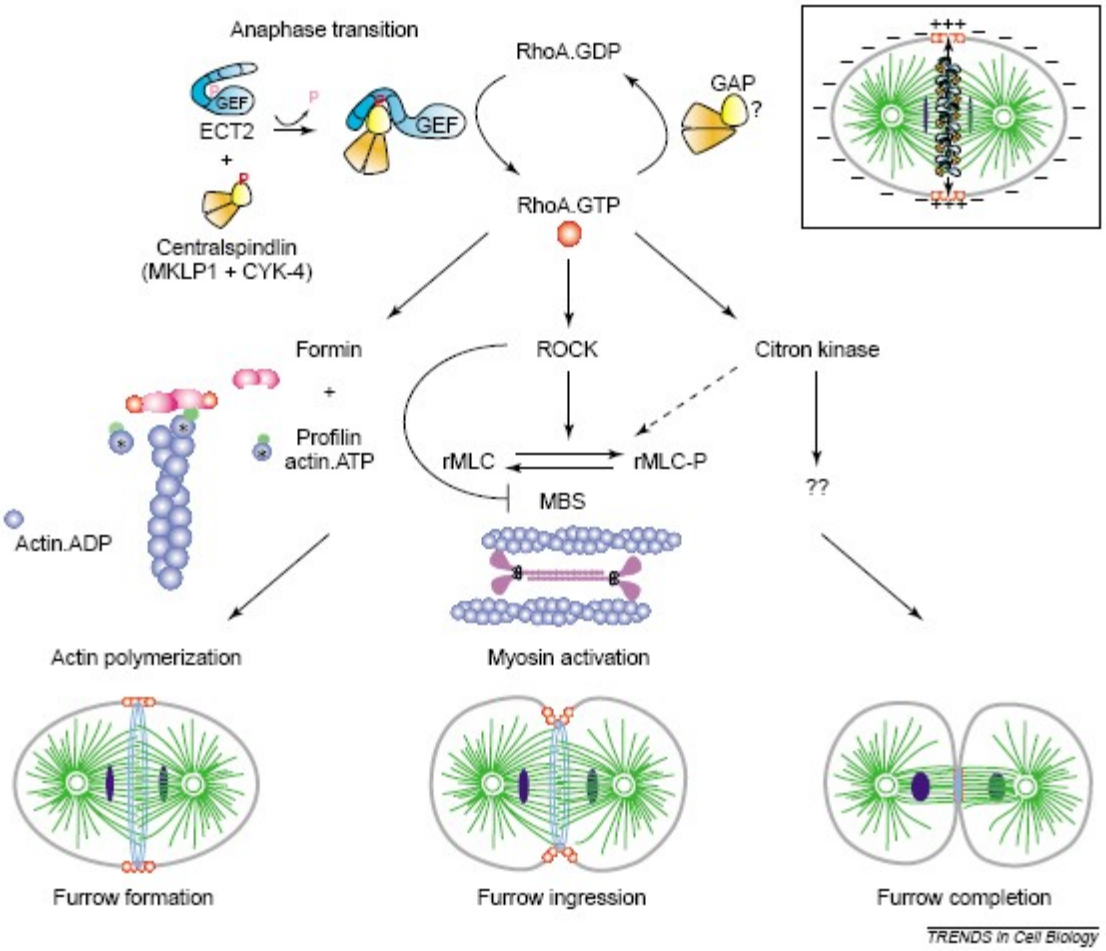
- Philosophy
  - Ontology: what exists
  - Logic
- Artificial Intelligence, AI
  - Database design
  - Reasoning systems
  - Planning, scheduling

Allen's temporal relationships

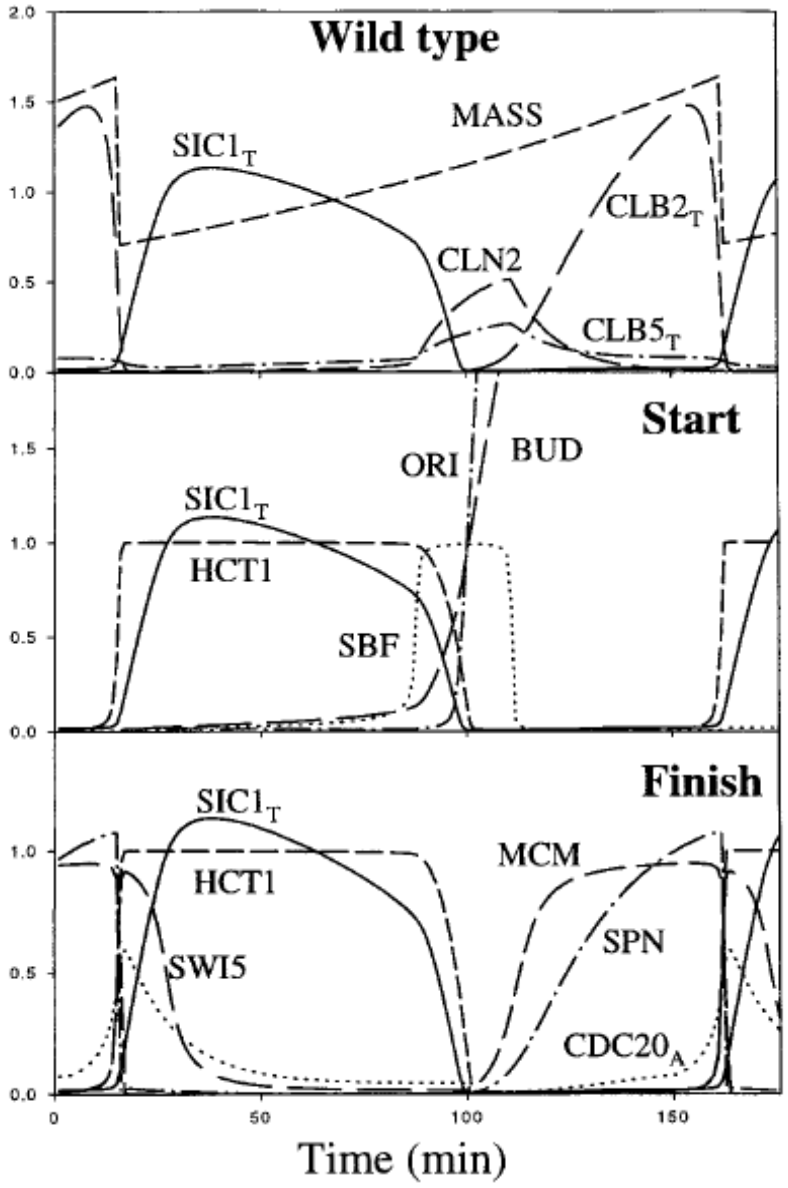


Temporal data in GIS

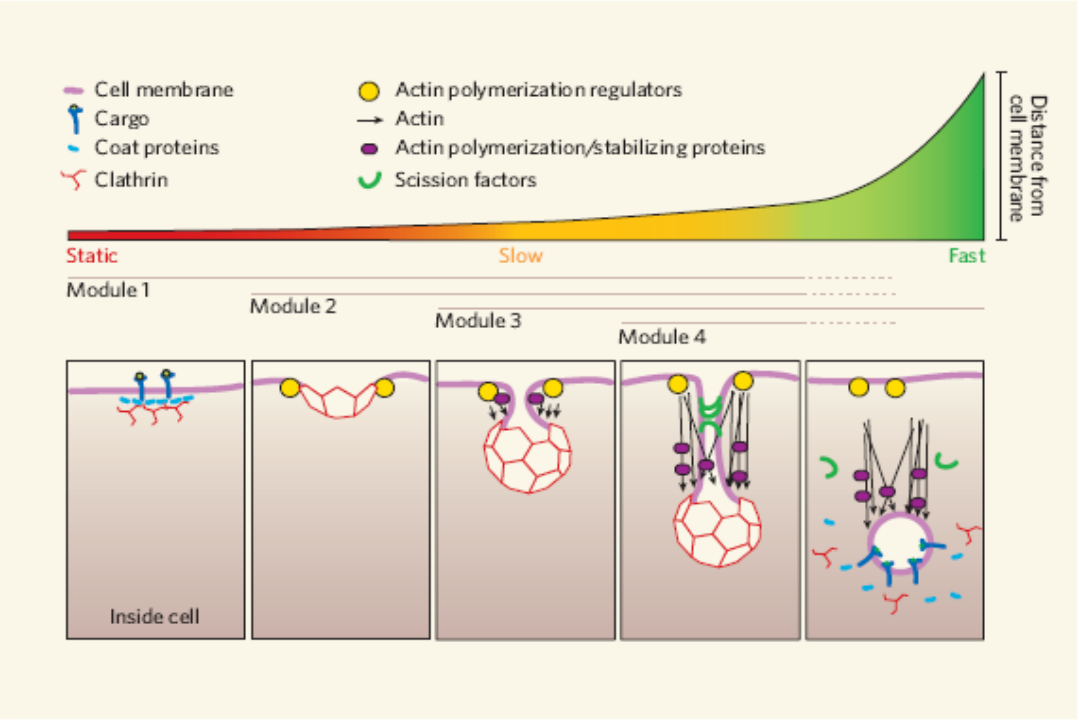
- Galton's distinctions



Cytokinesis: Rho regulation
Piekny, Werner, Glotzer 2005

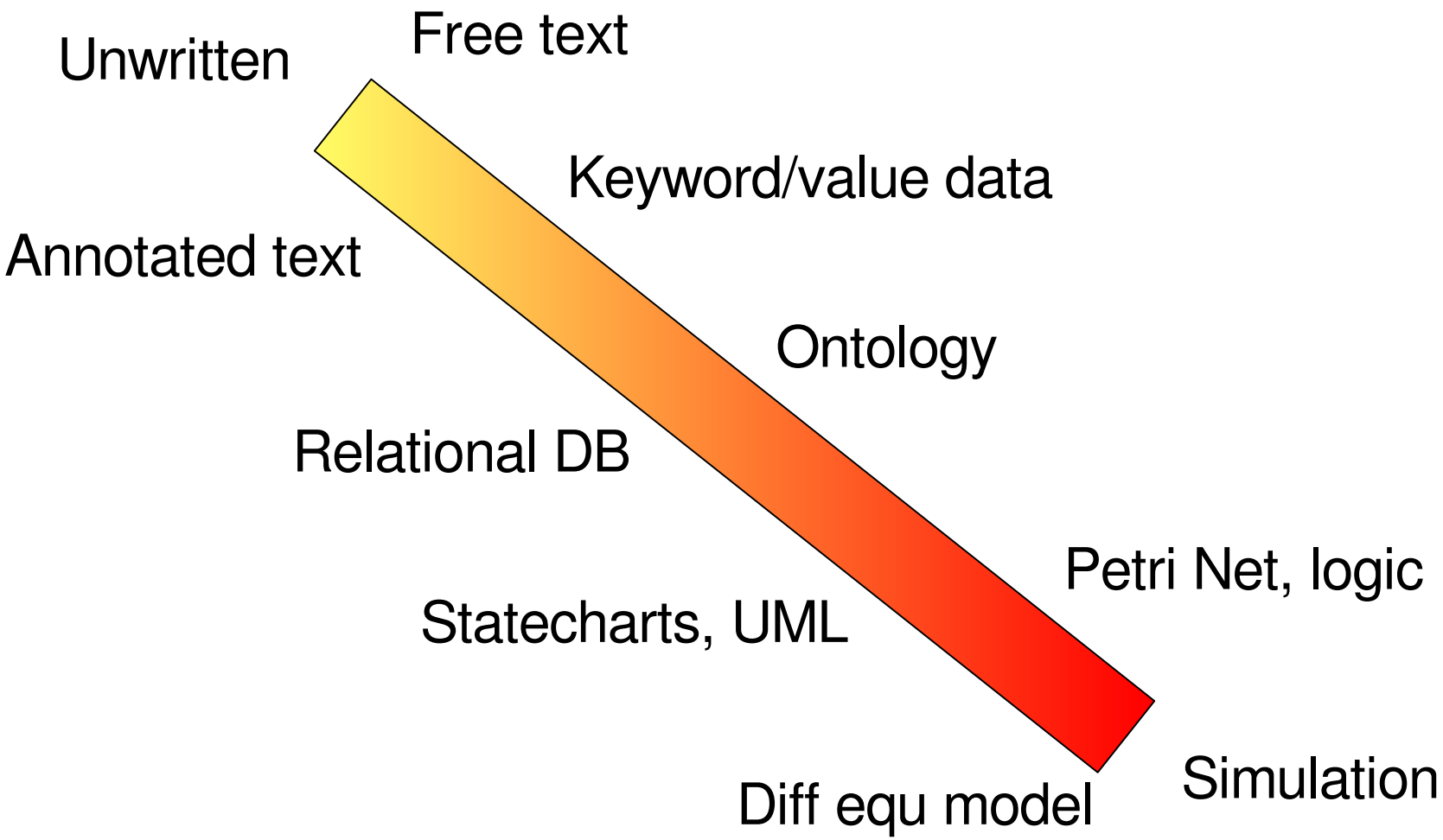


Kinetic analysis of budding yeast cell cycle: Chen et al 2000



Endocytic vesicle formation
Duncan & Payne 2005

Computable information



Design issues

- Main entities:
  - Continuant (thing, object)
  - Occurrent (event, process, happening)

BioChronicle

- Handle temporal biological information
  - Events, subevents
  - Relationships
  - Duration
  - Property values
  - Preconditions, context
- Database
- Test case: Cell cycle