

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...

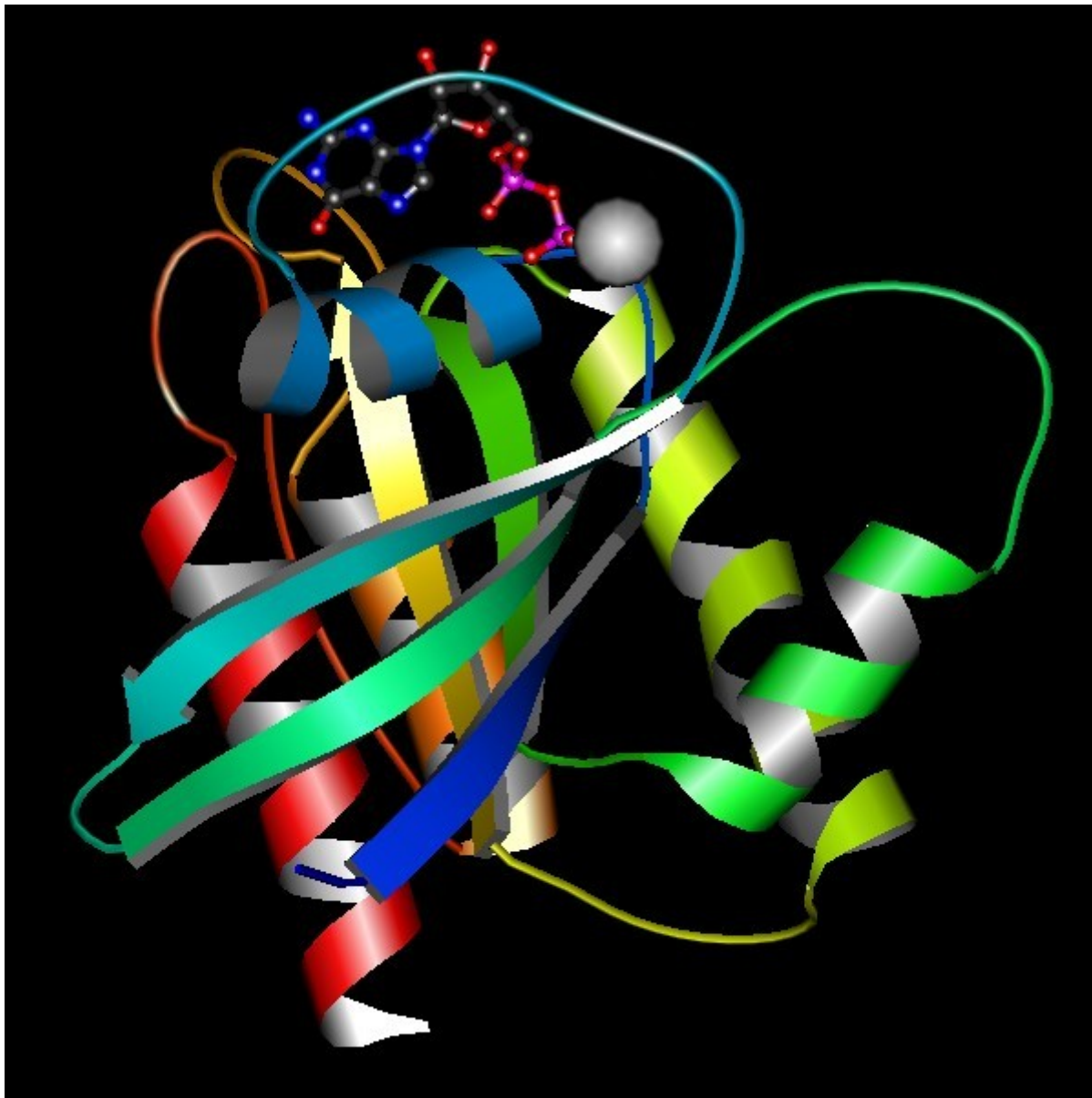
Thus far: Static structures

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

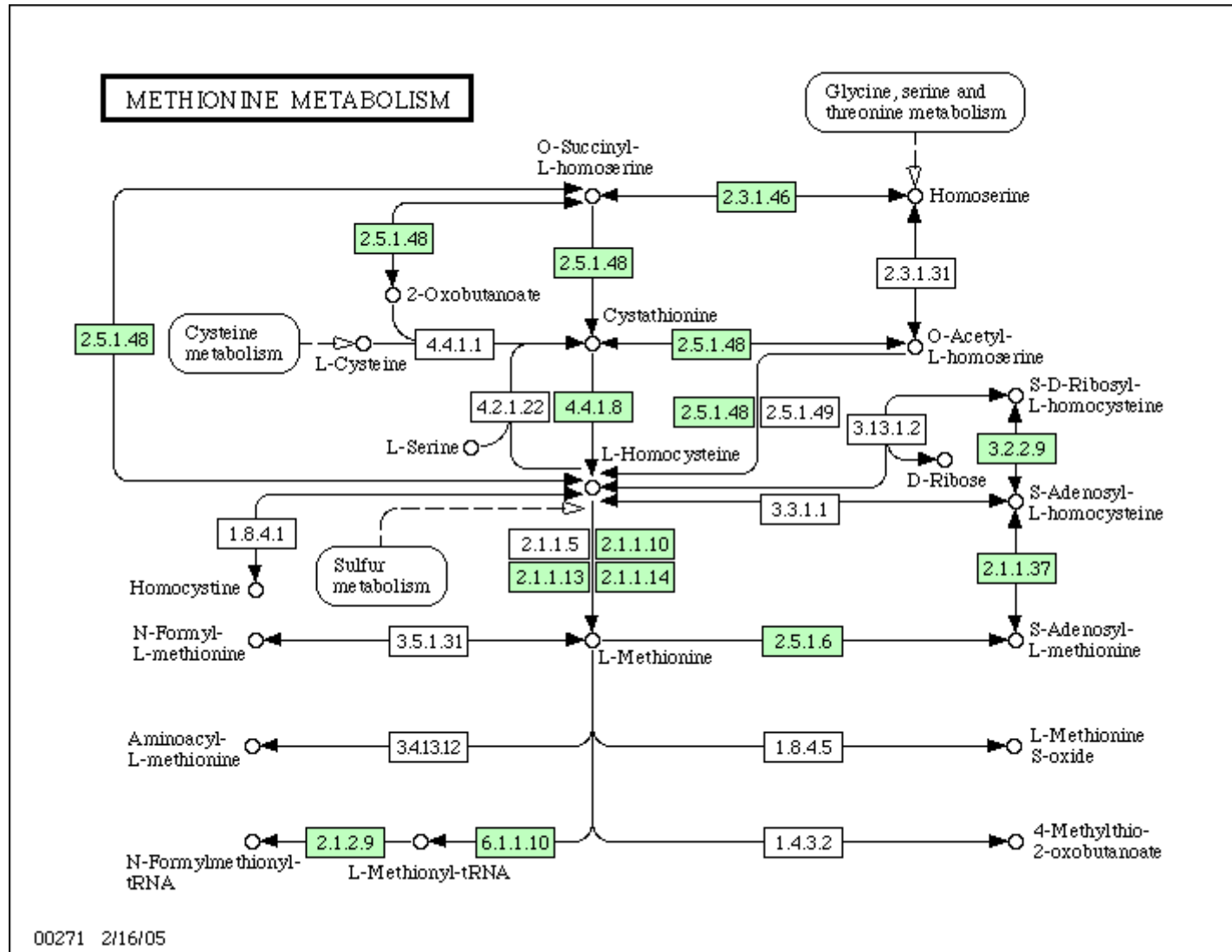
- Easier to work with?
 - Experimentally
 - Conceptually?

Sequence analysis

17	UNIPROT:	Q503B6 BRARE	1:189	1:189	REIRQHKL RKLNPDDNGQDCMNCRCVVS
18	UNIPROT:	Q568K0 BRARE	1:189	1:189	REIRQH KMRKLNPPDESGQDCMNSCRVVS
19	UNIPROT:	RASK HUMAN	1:188	1:188	REIRQYRL LKKISK-EEKTPGCVKIk c II-
20	UNIPROT:	Q3UCX0 MOUSE	1:188	1:188	REIRQYRM RKLNSSDDGTQGC H GLPC V L-
21	UNIPROT:	RASN MOUSE	1:188	1:188	REIRQYRL LKKLNSSDDGTQGC H GS P CVL-
22	UNIPROT:	RASK MOUSE	1:188	1:188	REIRQYRL LKKISK-EEKTPGCVKIk c VI-
23	UNIPROT:	RASK RAT	1:188	1:188	REIRQYRL LKKISK-EEKTPGCVKIk c VI-
24	UNIPROT:	Q4FJP3 MOUSE	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V L-
25	UNIPROT:	Q9D091 MOUSE	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V L-
26	UNIPROT:	RASN CHICK	1:188	1:188	REIRQYRM KKLN S NEDGNQGC H GLSC I V-
27	UNIPROT:	RASN HUMAN	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V V-
28	UNIPROT:	Q5U091 HUMAN	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V V-
29	UNIPROT:	Q2MJK3 PIG	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V V-
30	UNIPROT:	RASN CAVPO	1:188	1:188	REIRQYRM KKLN S NDGTQGC H GLPC V V-
31	UNIPROT:	Q4S7E9 TETNG	1:188	1:188	REIRQYRL NKLSK-EKTPRC V KIk c VV-
32	UNIPROT:	Q3TMF4 MOUSE	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V L-
33	UNIPROT:	RASN RAT	1:188	1:188	REIRQYRM KKLN S SEDGTQGC H GLPC V V-
34	UNIPROT:	RASN MONDO	1:188	1:188	REIRQYRM KKLNSSDDGTQGC L GLSC A V-
35	UNIPROT:	RASN PONPY	1:188	1:188	REIRQYRM KKLNSSDDGTQGC H GLPC V V-
36	UNIPROT:	Q57467 ORYLA	1:188	1:188	REIRQYRL SKLSK-EKTPRC V NLk c VV-
37	UNIPROT:	Q13021 XENLA	1:185	1:184	REIRQFRL LKKMSK-EEKTPGCVK F K----
38	UNIPROT:	Q5EFX7-2	1:188	1:188	REIRQYRL SKISK-EEKTPGCVQ L k c VV-
39	UNIPROT:	RASN XENLA	1:188	1:188	REIHQYRM KKLD S SEDNNQGC I RIP C KL-
40	UNIPROT:	RASK MSVKI	1:188	1:188	REIRQYRL LKKISK-EEKTPGCVKIk c VI-
41	UNIPROT:	RAS CARAU	1:177	1:177	REIRQYRL RKLSKEE E T-----
42	UNIPROT:	Q6DGD1 BRARE	1:186	1:185	REIRHYRM KKLN S REDRKQGC L GV S C----
43	UNIPROT:	P01116-2	1:188	1:187	REIRKHK -EKMSKDGKKKKKK S KT K CVI-
44	UNIPROT:	RASK MELGA	1:188	1:187	REIRKHK -EKMSKDGKKKKKK T KT K CI I-
45	UNIPROT:	RASK CYPCA	1:188	1:187	REIRKHK -EKMSKEGKKKKKK S KT K CVL-
46	UNIPROT:	RASK ORYLA	1:188	1:187	REIRKHK -EKMSKEGKKKKKK S KT K CI L-
47	UNIPROT:	Q9PSS8 PLAFE	1:188	1:187	REIRKHK -EKMSKEGKKKKKK S KT K CS L -
48	UNIPROT:	RASK MONDO	1:188	1:187	REIRKHK -EKMSKDGKKKKKK S KT K CI I-
49	UNIPROT:	RASN BRARE	1:186	1:185	REIRHYRM KKLN S REDRKQGC L GV S C----
50	UNIPROT:	Q6AZA4 BRARE	1:188	1:187	REIRKHK -EKMSKEGKKKKKK S KT K CA L -
		consensus/100%			REI+pa+.pKhs..tct.....
		consensus/90%			REIRpa+.cKhs..tctt.tp.th.Chl.
		consensus/80%			REIRQa+h+Kls.--tt.sChth.Cll.
		consensus/70%			REIRQa+h+Kls.s--ps.GChthpCVL.



MolScript: Per Kraulis 1991, 1997



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

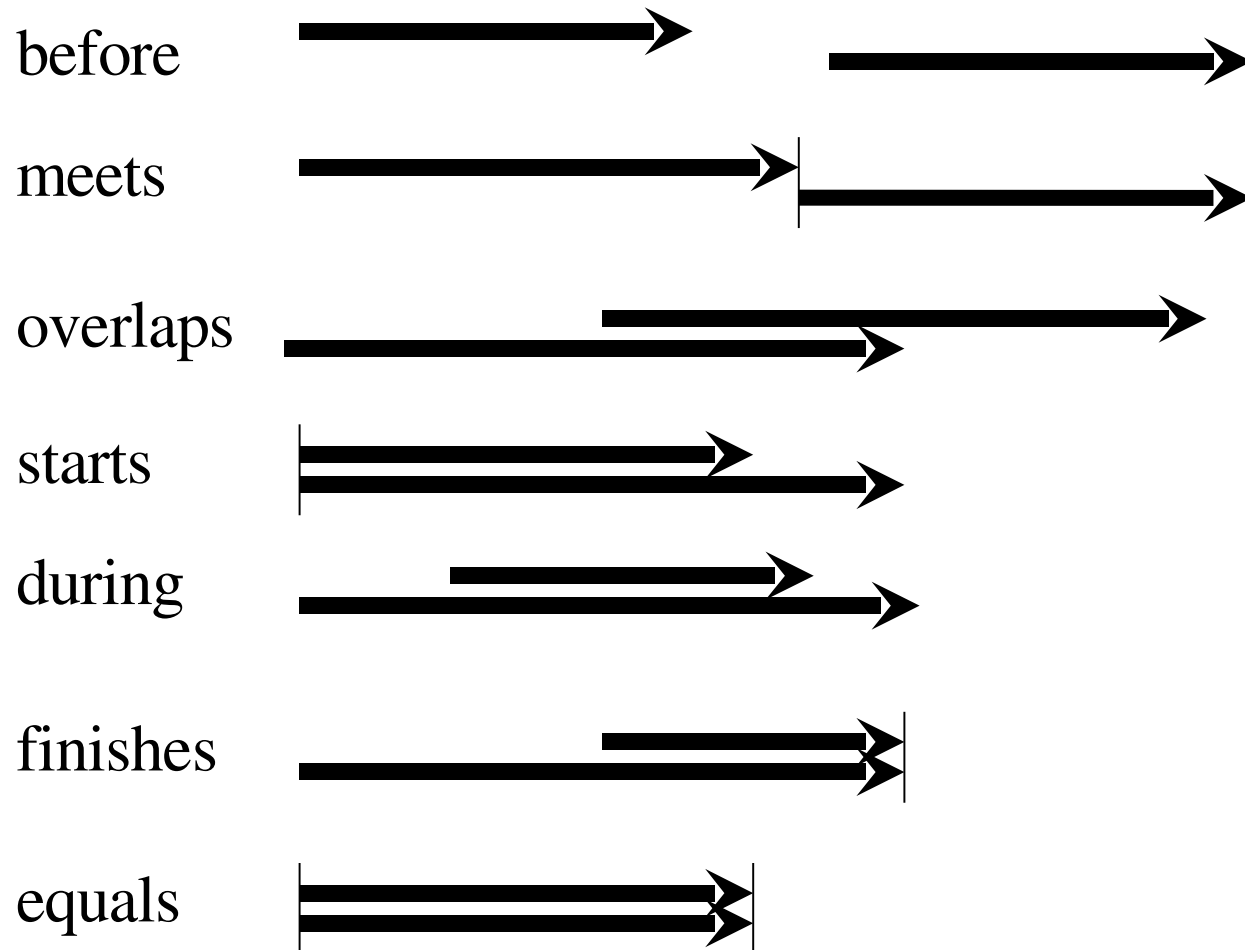
Knowledge representation II

- 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

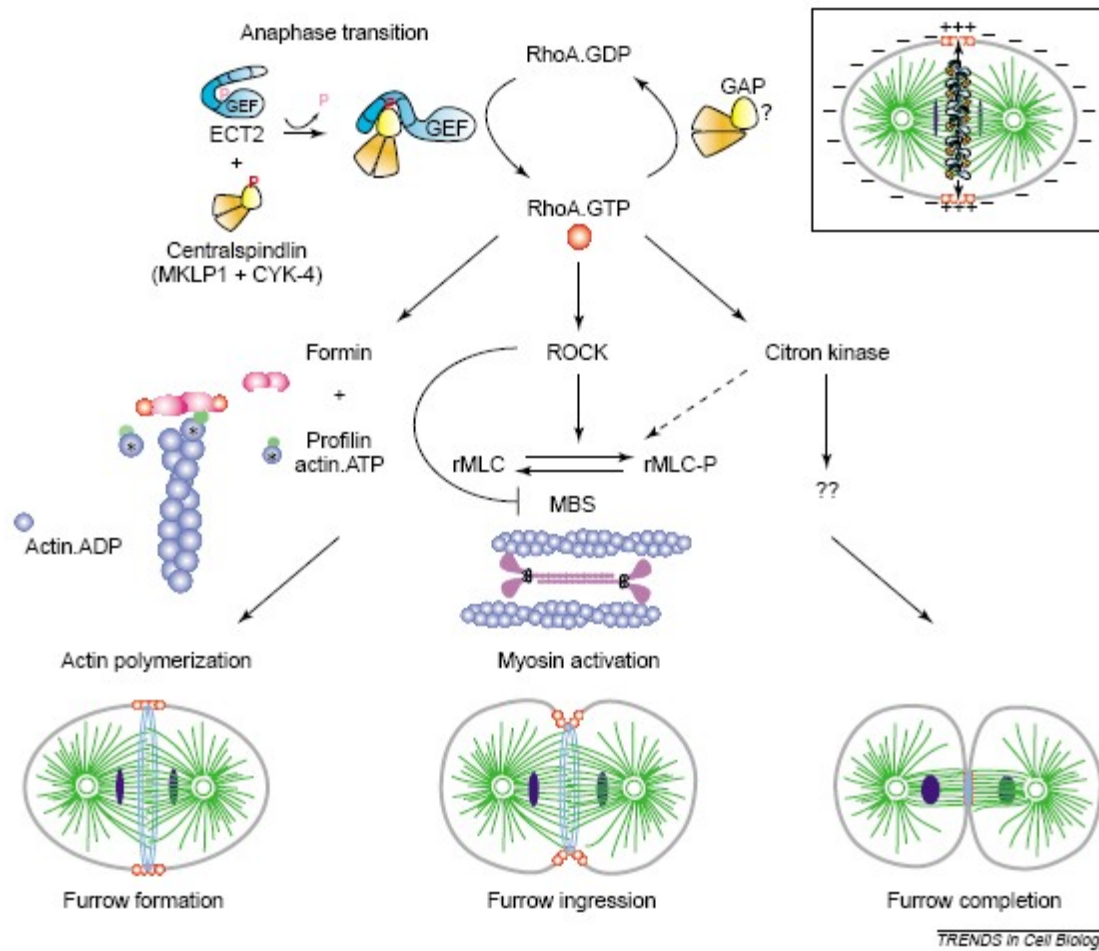
- Philosophy
 - Ontology: what exists
 - Logic
- Artificial Intelligence, AI
 - Database design
 - Reasoning systems
 - Planning, scheduling

Allen's temporal relationships

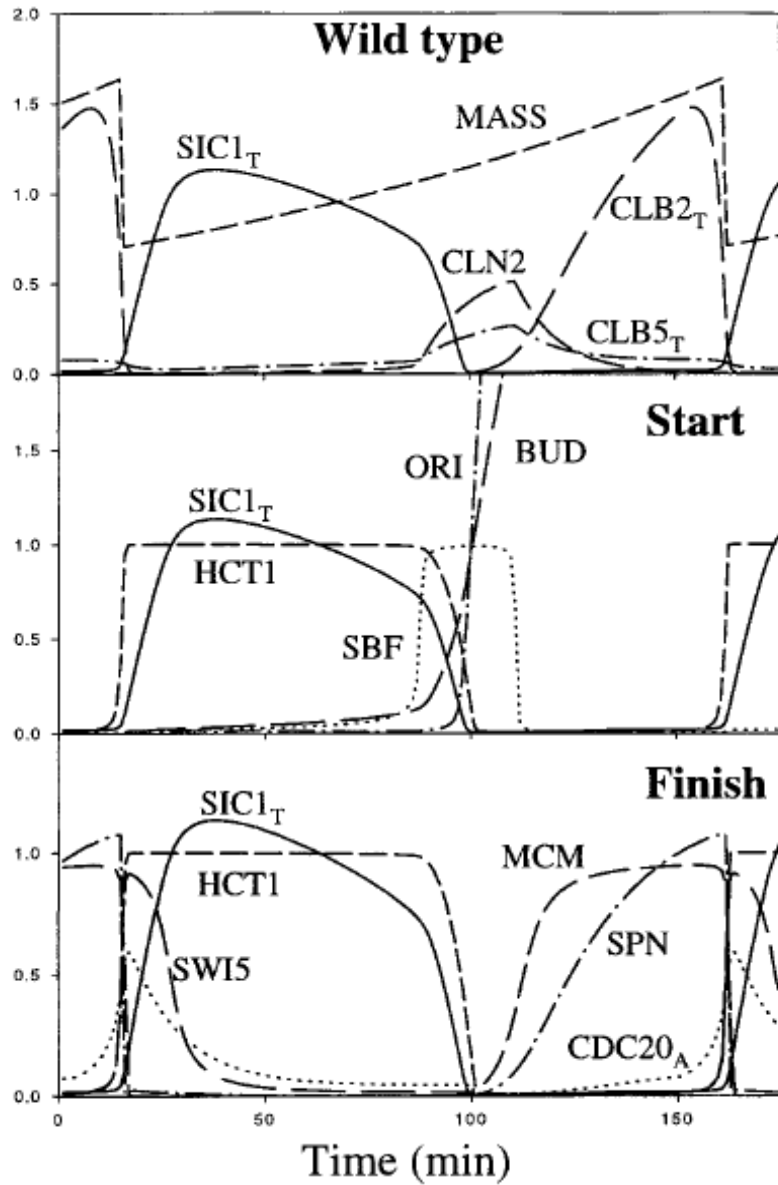


Temporal data in GIS

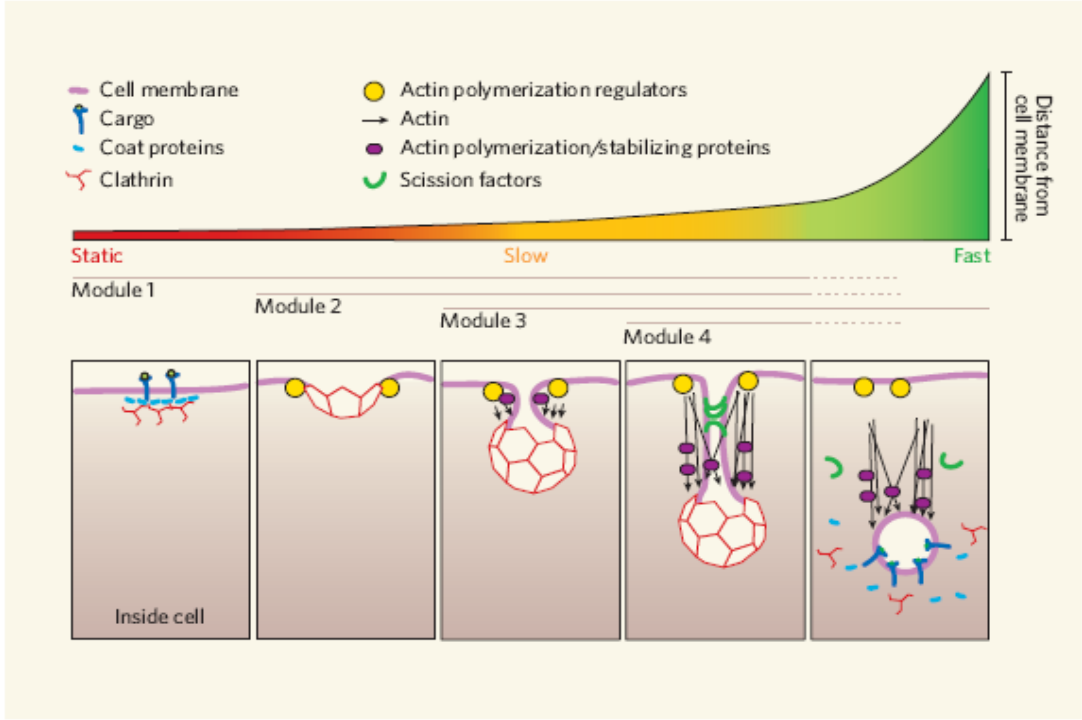
- Galton's distinctions



Cytokinesis: Rho regulation
 Piekny, Werner, Glotzer 2005

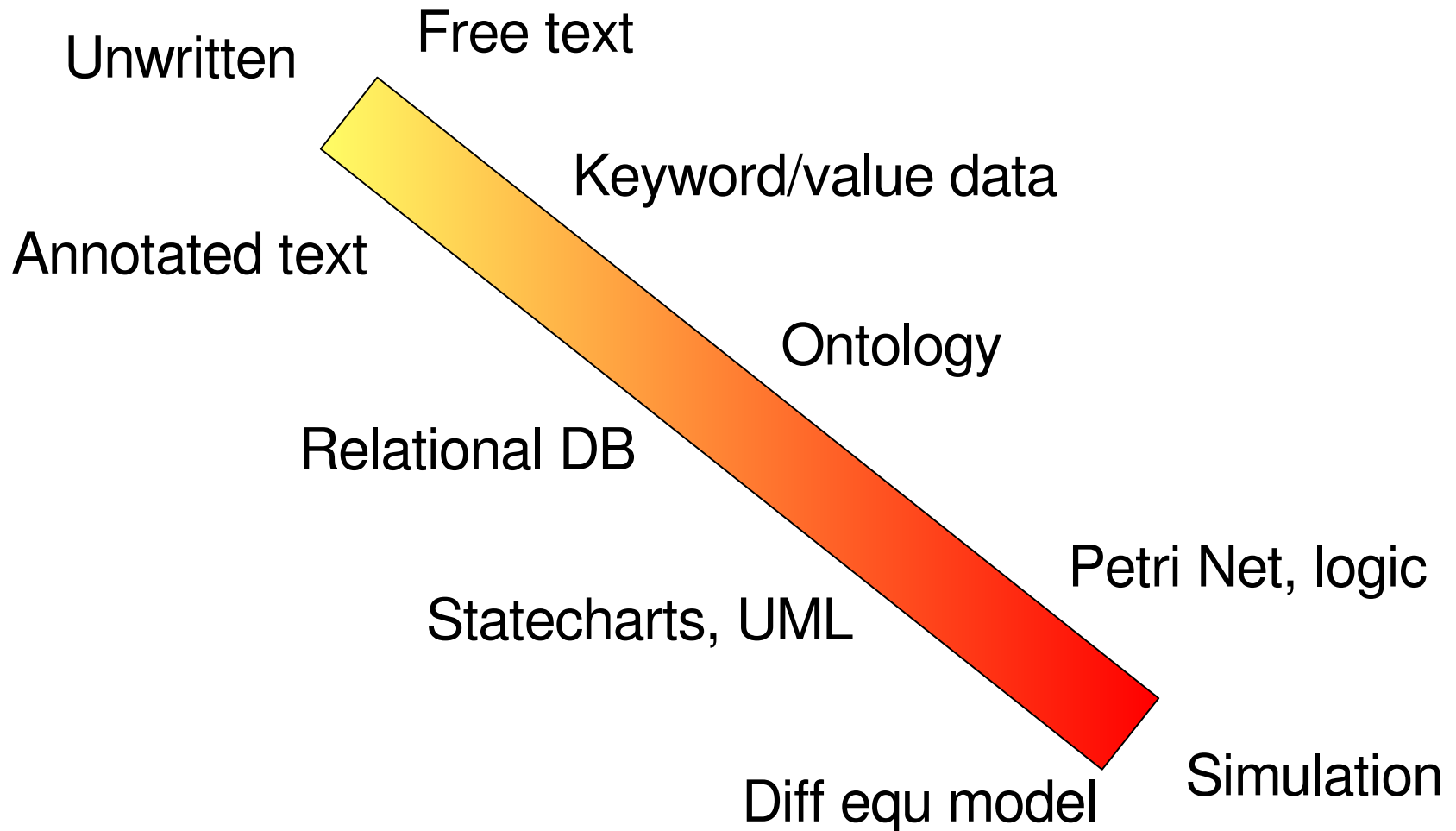


Kinetic analysis of budding yeast cell cycle: Chen et al 2000



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