

# From Data to Models

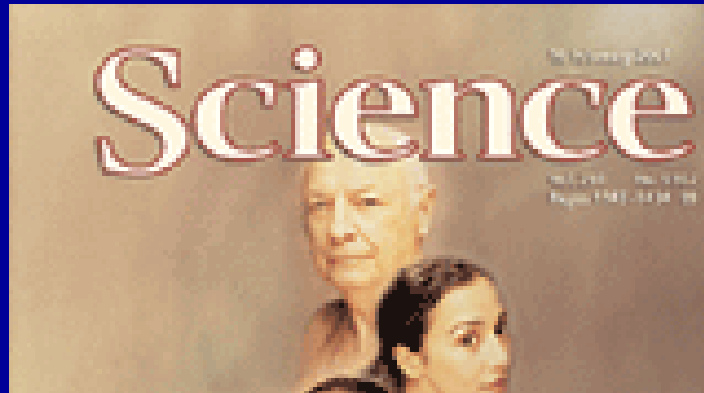
or...  
a biological  
“Theory of Everything”?

*Arthur Thomas*

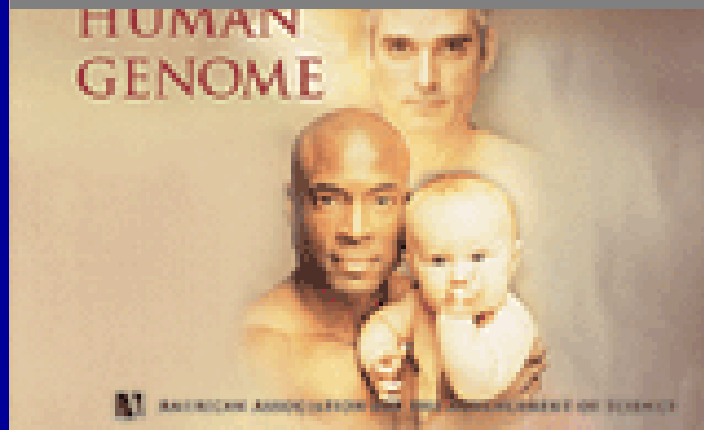
*Proteus Associates*

*ajt@proteus-associates.com*

February 15, 2001



"The end of the beginning"?



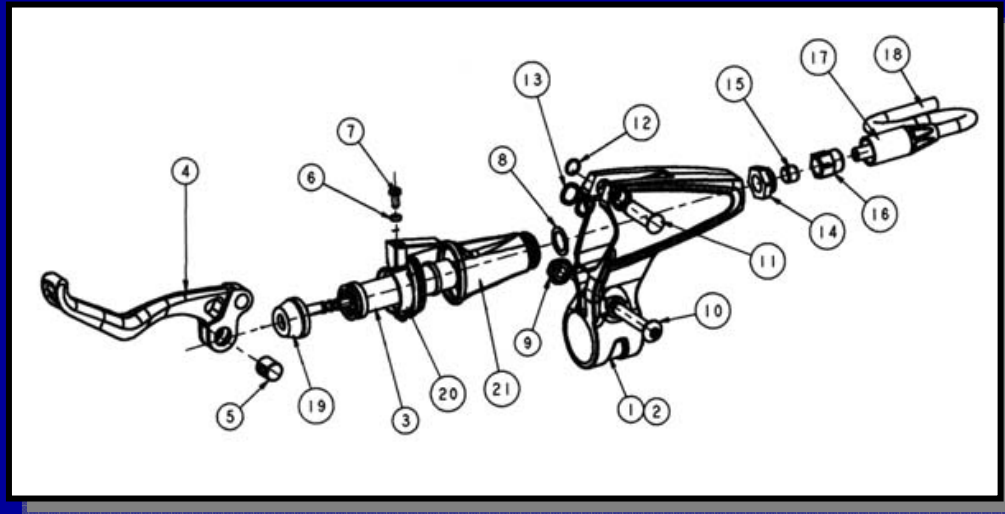
# The Challenge



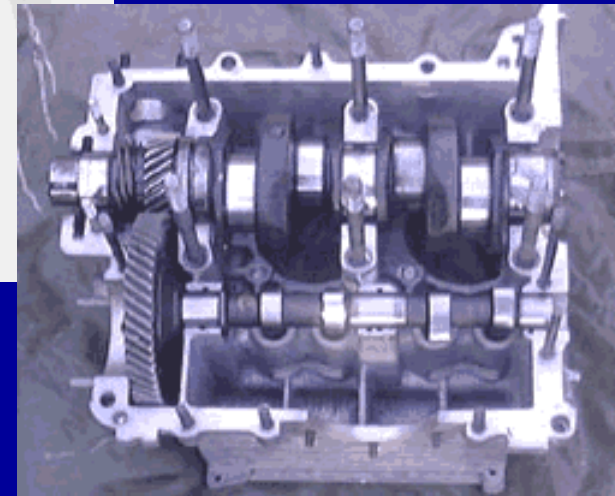
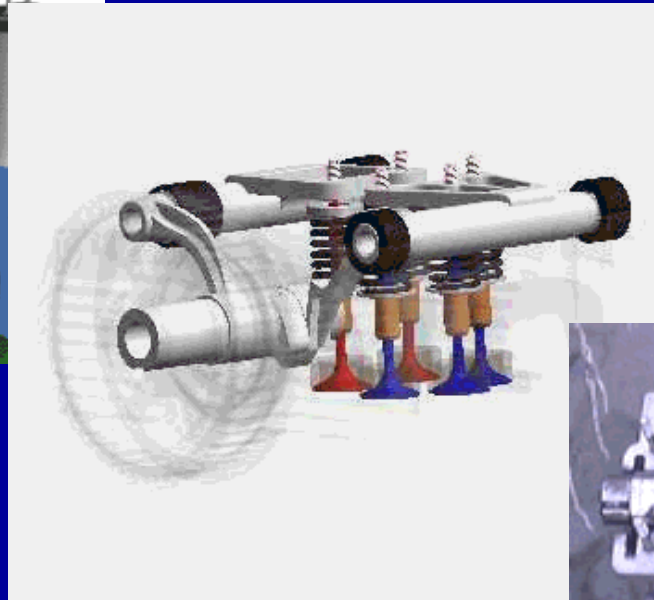
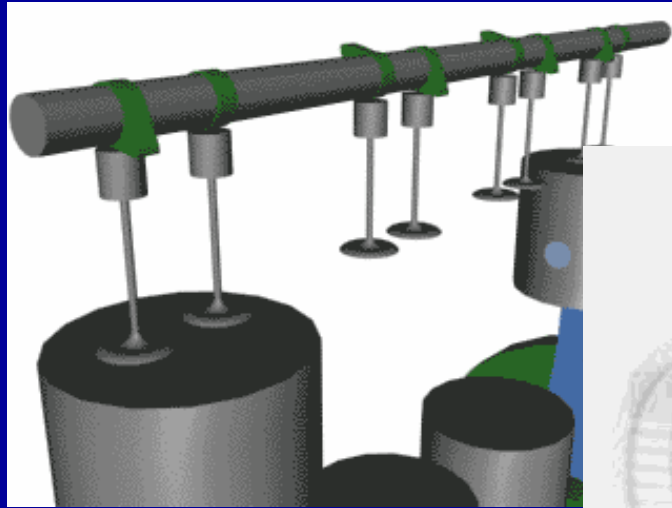
# It's even worse than that...

Getting from... To... To...

| <i>Part #</i> | <i>Description</i>             | <i>Label</i> |
|---------------|--------------------------------|--------------|
| 100-306       | Master Cylinder Body Left      | 1            |
| 100-223       | Master Cylinder Body Right     | 2            |
| 225-777       | Cartridge Assembly             | 3            |
| 100-876       | Lever                          | 4            |
| 101-888       | Adjustable Bushing             | 5            |
| 332-863       | Master Cylinder Bleeder O-ring | 6            |
| 444-754       | Master Cylinder Bleeder Screw  | 7            |
| 100-328       | Cartridge Gasket               | 8            |
| 274-000       | Clamp Nut                      | 9            |
| 273-875       | Clamp Screw                    | 10           |
| 273-876       | Lever Pivot Pin                | 11           |
| 700-201       | Lever Pivot Retaining Ring     | 12           |
| 700-202       | Wave Washer                    | 13           |
| 202-343       | Jam Nut                        | 14           |
| 209-876       | Compression Bushing            | 15           |
| 209-875       | Hose Nut                       | 16           |
| 209-874       | Nose Cone                      | 17           |
| 300-888       | Hose                           | 18           |
| 503-987       | Push Rod Seal                  | 19           |
| 503-972       | Bladder Retainer               | 20           |
|               |                                | 21           |



But, a static picture is not enough...

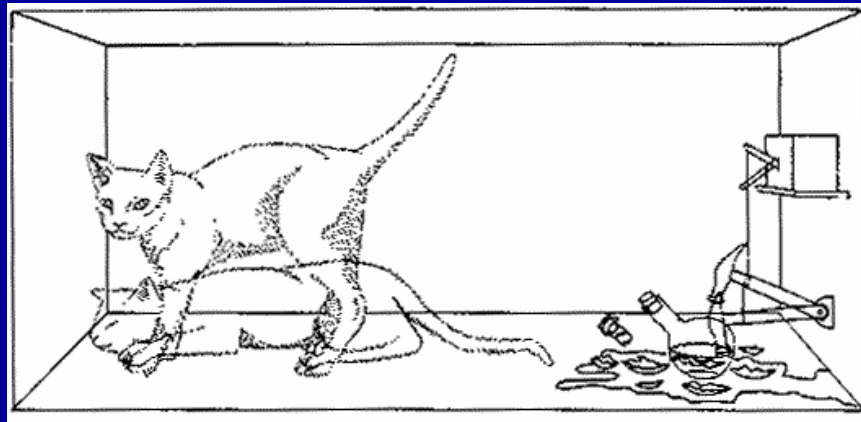


# Outline

- Steps towards a biological “Theory of Everything”
  - What would it take to get there?
    - Challenges to modelling: getting the data, doing the computations
  - How would we know we’d got there?
    - Completeness, accuracy
- Illustrative examples
  - Biological pathways as graphs
    - Inferring pathway graphs from data
    - Global properties of graphs: robustness and evolution
  - Integrative whole-organ modelling
    - The “Human Physiome Project”
    - Inferring whole-organ behaviour from molecular and other data
  - Super-organism modelling
- Some Implications
  - Drug discovery
  - “Wellness”
  - Environmental

Q: What's the difference between a live cat and a dead one?

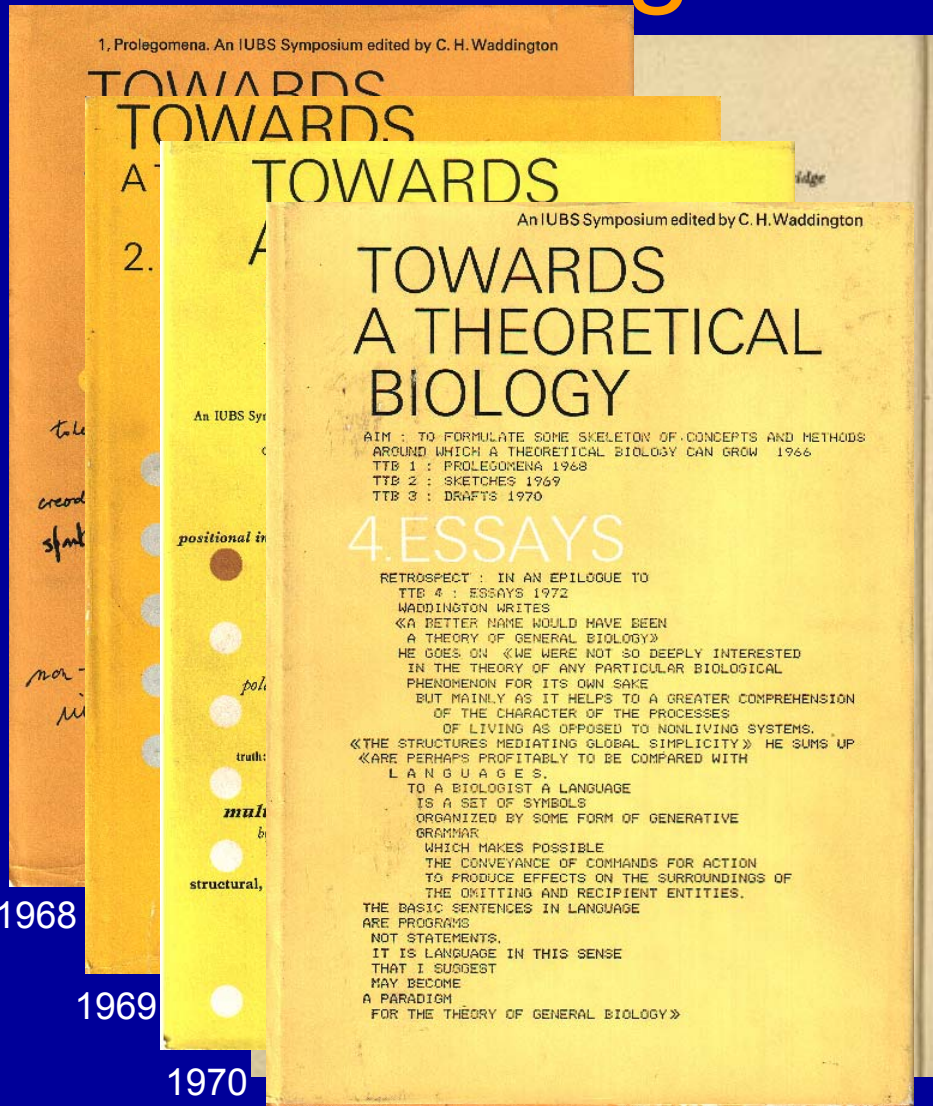
A1: an *alpha* particle



A2: Systems Biology!



# A long time coming...



1968

1969

1970

1972



2002



# Elements of Systems Biology

- Understanding of
    - Structure (components and structural relationships)
    - Behaviours and their phase spaces
    - Factors controlling states and behaviours
    - Methods for controlling systems and influencing/synthesizing desired behaviours
  - Main subjects
    - Genetic regulatory networks
    - Signal Transduction Networks
    - Metabolic networks
    - Organ/organism/super-organism *physiology*
      - biochemistry in structural context
      - inter-cell, inter-organism communication
- Pathway Models*

From Kitano (2000)

# Biology as an “Information Science”

“Genes can only specify the properties of the proteins they code for, and any integrative properties of the system must be *‘computed’ by their interactions.*”

S. Brenner (emphasis added)

“We need to design a theoretical framework that can take account of the *flow of information* through biological systems.”

S. Brenner (emphasis added)

# “So few genes, so many functions?”

- Gene function not directly specified in the code
- Genes may play multiple roles, and **code for multiple proteins**
  - splice variants, post-translational modifications, RNA silencing, ...
    - Extreme: 1 gene = 38,000 proteins in *Drosophila*
- Highly non-linear relationship between genome size and functional complexity: many functions result from **cooperation between multiple genes**;
  - for  $n$  genes and  $r$  genes/function there are  
 $nPr = n(n-1)(n-2) \dots (n-r+1) = n!/(n-r)!$  possible functions
    - 2 genes/function =  $0.5 \times 40000 \times 39999 = 799,980,000$  possible functions
    - 100 genes/function  $\sim 10^{300}$  possible functions<sup>(1)</sup>
- Re-use of functional “modules”: nature has only explored a **small sub-space**
- Functions depend on extra-genetic factors (water, etc.)

Feytmans, E., Noble, D. & Peitsch, M. “Genome Size and Numbers of Biological Functions,” *Trans. Comp. Systems Biology* 1:44-49 (2004)

# Basic Challenges

- Experimental:
  - Developing non-invasive techniques for measuring *in vivo* parameters down to the sub-cellular level, with appropriate accuracy
- Computational:
  - Multi-scale, multi-source, multi-physics modelling
  - Unifying frameworks (models, ontologies)
  - Interactive visualization of complex temporal data sets: “modelling at the speed of thought”

# Experimental Challenges

- Need for *tight (two-way) coupling* with experimental work
- *Accuracy* of measurements: PCR is noisy, low copy number, thermodynamic effects, stability, statistical criteria
- *Identifiability/Controllability*: what could we *possibly* observe/control?
- Importance of *structure* (cytoskeleton, compartments): need for high-spatial resolution intra-cellular measurements (e.g. *FISH*, reporter genes)
- Need kinetic data under uniform conditions

# How Much Data Do We Need?

- No framework for deciding when a model is complete or consistent with data
- How much accuracy is “enough”?



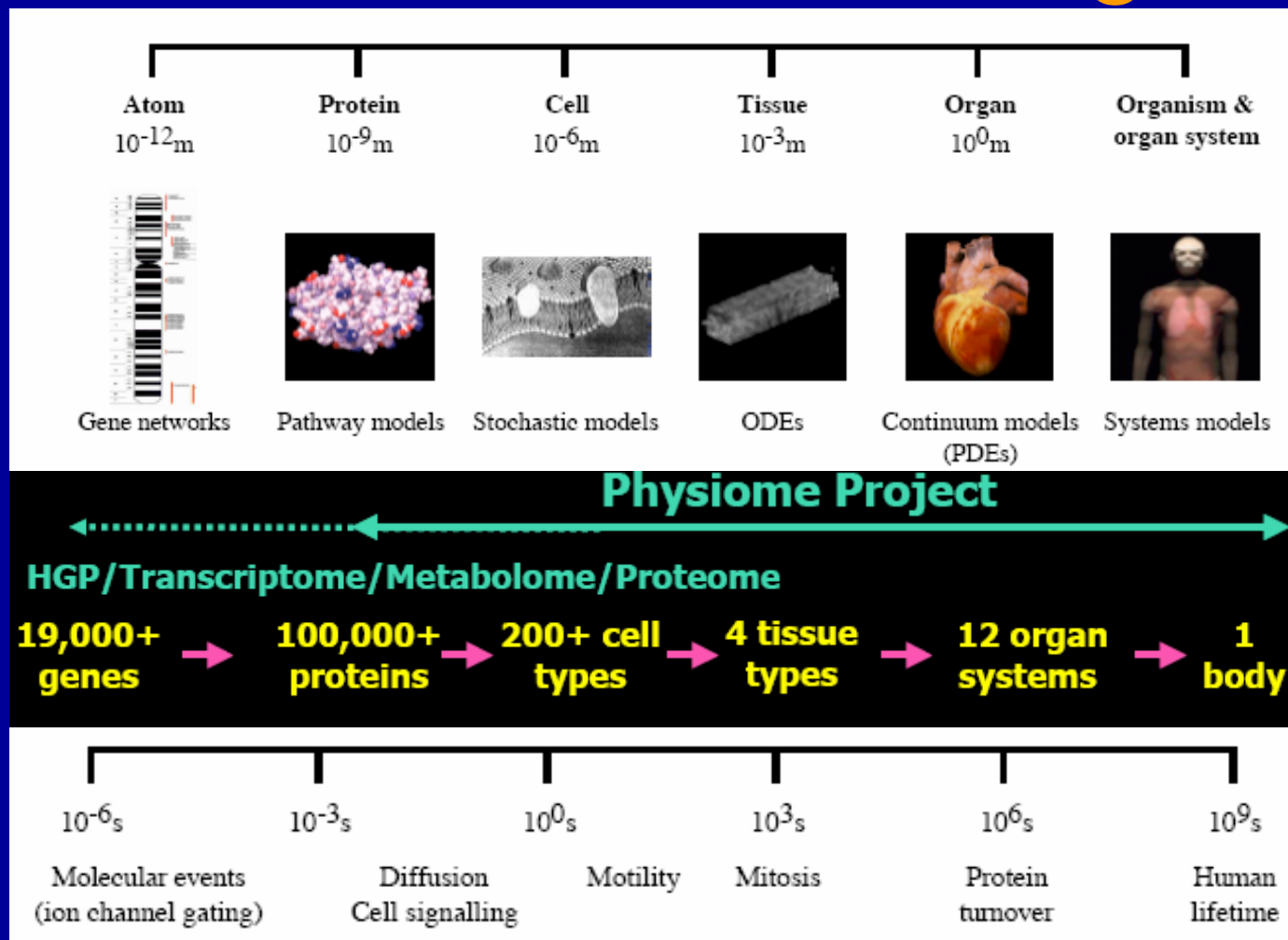
| Model                       | Scope   | Applicable Rules   |
|-----------------------------|---|--|
| Atomic                      | Cell $c$ at time $t$  | Physics  |
| Molecular                   | Cell $c$ at time $t$  | Chemistry  |
| Biomolecular (discrete)     | Cell $c$ at time $t$  | Molecular mechanics  |
| Biomolecular (statistical)  | Biochemically equivalent cells  | Chemical kinetics and thermodynamics described by differential equations |
| Biomolecular (steady-state) | Genetically equivalent cells, similar growth conditions, steady state | Flux balance, physical and chemical constraints                          |
| Boolean                     | Genetically equivalent cells  | Genetic and metabolic 'circuits'   |

| Organism             | N      | K | Estimated number of microarrays |             |
|----------------------|--------|---|---------------------------------|-------------|
|                      |        |   | Lower bound                     | Upper bound |
| <i>M. pneumoniae</i> | 688    | 1 | 10                              | 80          |
| <i>E. coli</i>       | 4,288  | 3 | 50                              | 40 000      |
| <i>H. sapiens</i>    | 50 000 | 4 | 100                             | 700 000     |

Upper and lower bounds on the number of microarrays (or equivalent transcriptome-wide experiments) to complete discrete transcriptional network models for various organisms, calculated according to Krupa [10]. N represents the number of nodes (genes in this example). K represents the maximum number of regulatory connections per node. The expression level of each gene is categorized as high, medium, or low ( $\xi = 3$ ). The lower bound (information-theoretic) is given by  $\xi^K + K \log_2(N/K)$ . The upper bound is given by  $\xi^{2K}(2K(\ln N + \ln \xi) + \ln C)$ , where the measurements fail to determine the model with probability  $1/C$ . Here we set  $1/C$  equal to 0.01. It is important to note that the upper bound estimate increases exponentially with K, making it the dominant parameter.

Selinger et al. *Trends in Biotech.* (2003)

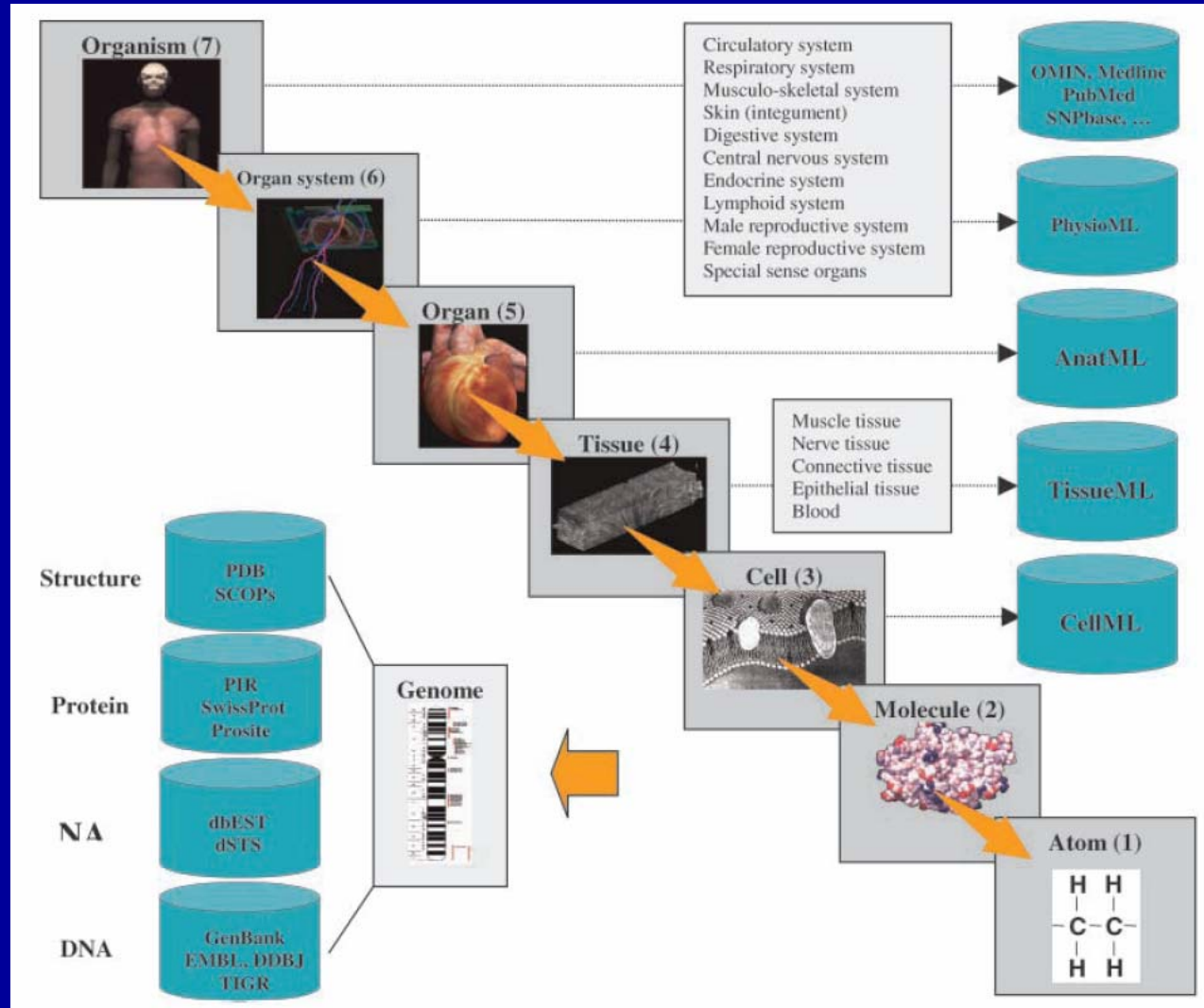
# Multi-scale Modelling



Adapted from P. Hunter, P. Robbins, D. Noble "The IUPS human physiome project" *Eur J Physiol* 445:1-9 (2002) and P. Hunter "An Update on the Human Physiome Project," in *Proc. IUPS Satellite Workshop on Computational Physiology, San Diego, CA (2005)*



# Multi-source Modelling

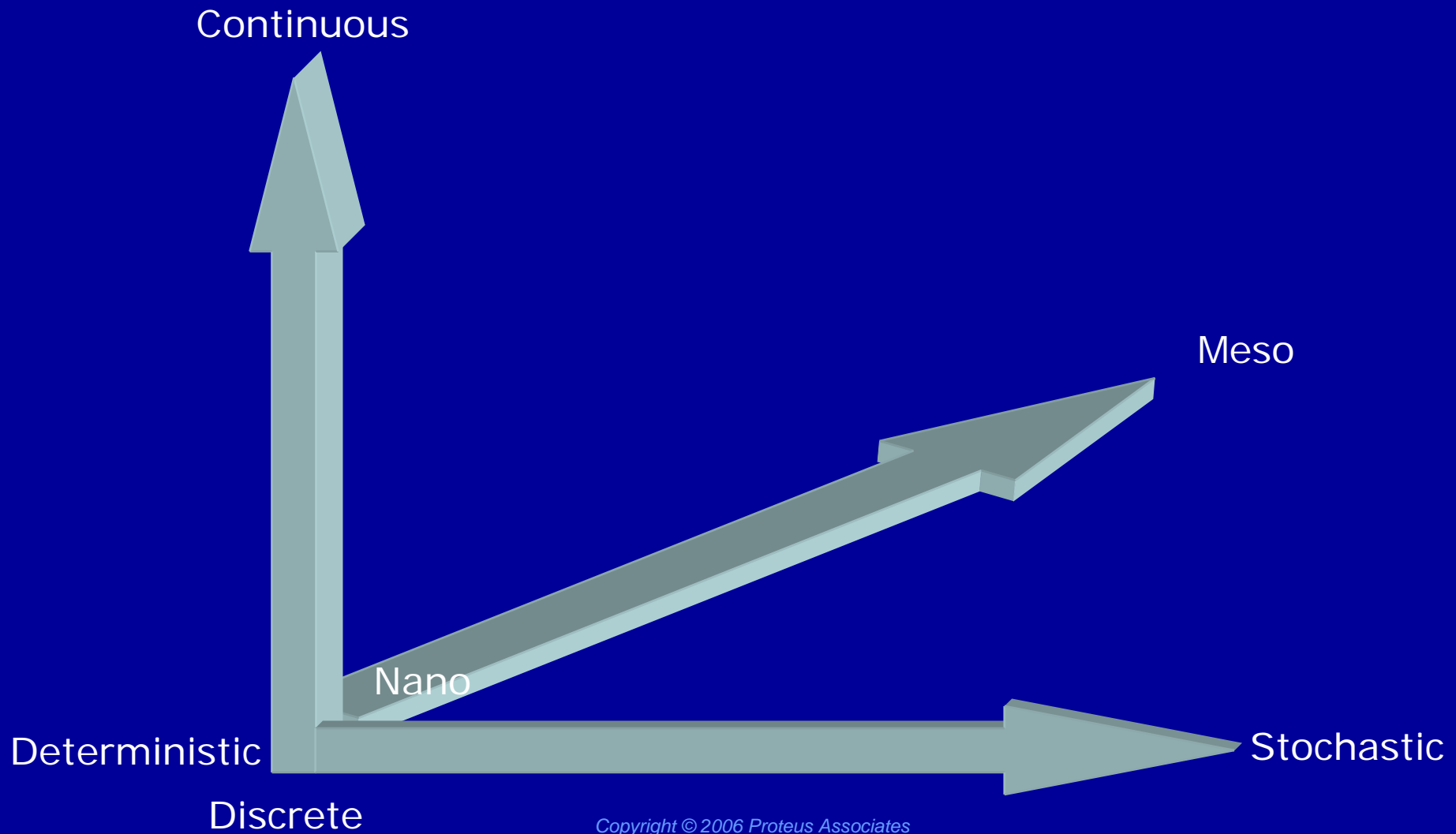


P. Hunter, P. Robbins, D. Noble "The IUPS human physiome project" *Eur J Physiol* 445:1-9 (2002)

# Multi-physics Modelling

|         | Genome   | Transcriptome  | Proteome/ Metabolome  | Cellome  | Physiome  |   |
|---------|--|--|---|--|---|---|
| Domain  | <ul style="list-style-type: none"> <li>• Genes</li> <li>• Promoters</li> </ul>   | <ul style="list-style-type: none"> <li>• Gene expression</li> <li>• Genetic networks</li> </ul>  | <ul style="list-style-type: none"> <li>• Post-translational modification</li> <li>• Protein-protein interactions</li> </ul> | <ul style="list-style-type: none"> <li>• Pathways</li> <li>• Enzyme kinetics</li> </ul>                          | <ul style="list-style-type: none"> <li>• Compartments</li> <li>• Transport</li> <li>• Signal transduction</li> </ul>      | <ul style="list-style-type: none"> <li>• Whole organ models</li> </ul>  |
| Methods | <ul style="list-style-type: none"> <li>• Sequence homology</li> <li>• Pattern recognition</li> <li>• Markoff models</li> </ul> | <ul style="list-style-type: none"> <li>• Clustering</li> <li>• SVD/PCA</li> <li>• Random Boolean Nets</li> <li>• Petri nets</li> <li>• Graph theoretic</li> <li>• Bayesian nets</li> <li>• Mutual information</li> </ul> | <ul style="list-style-type: none"> <li>• Graph-theoretic</li> <li>• Petri nets</li> <li>• Monte Carlo</li> </ul>            | <ul style="list-style-type: none"> <li>• Graph-theoretic (DAGs)</li> <li>• Ontologies</li> <li>• ODEs</li> </ul> | <ul style="list-style-type: none"> <li>• ODEs</li> <li>• PDEs</li> <li>• Graph-theoretic</li> <li>• Ontologies</li> </ul> | <ul style="list-style-type: none"> <li>• ODEs</li> <li>• PDEs</li> <li>• Reaction-Diffusion</li> <li>• Finite element/boundary</li> <li>• Ontologies</li> </ul> |

# Modelling Methods

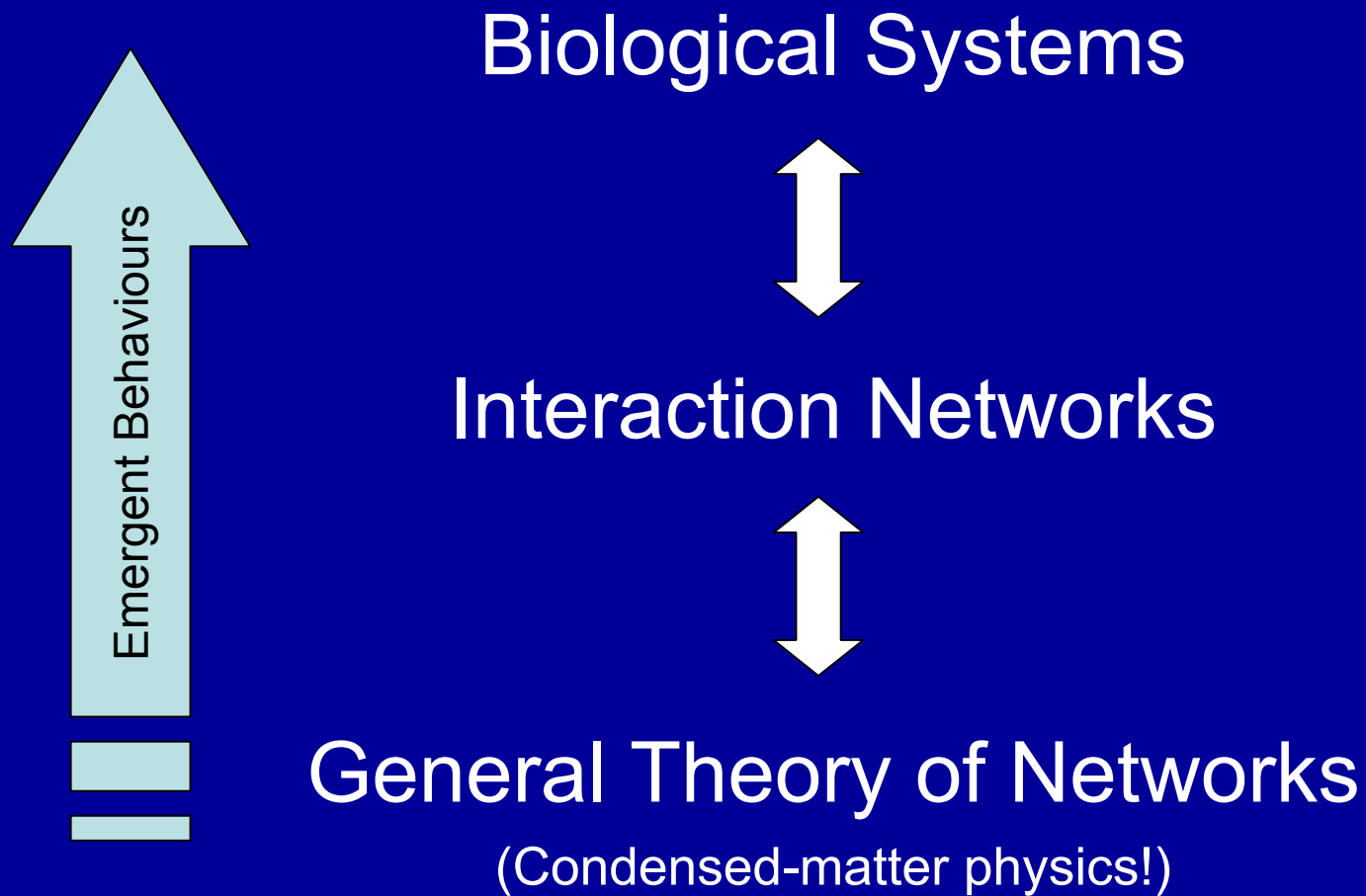


# Networks

“The emergent theory that describes how non-biological networks constitute themselves, how they react dynamically to perturbations and how their behaviour translates into predictable and measurable properties of the system... should provide a *general theoretical framework for systems biology*.”

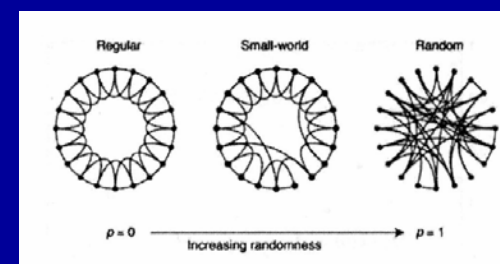
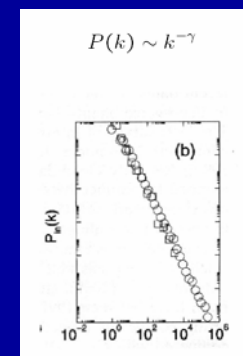
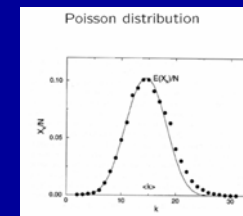
*Rudi Aebersold (2005)*

# Hierarchies of Networks



# Properties of Biological Networks

- Degree distribution
  - frequency distribution  $P(k)$  of nodes of degree  $k$
- Random:  $P(k)$  follows a Poisson law
- Scale-free:  $P(k)$  follows a power law
- “Small world”: average separation between nodes is small compared to the overall size of the network



# Properties of Biological Networks

- “Rich get richer”: likelihood of attachment to a node is proportional to the node’s degree
- Evolution of “hubs”
  - Trade-off between diversity and robustness
- Scale free networks are
  - Robust against **random failure** but not **loss of hubs**
  - Low congestion/efficient transport/energetics
- Move away from looking at single protein functions to studying networks of interactions (protein complexes)



# Robustness & Evolution

- Uneven degree connectivity makes networks resilient to mutations
- Importance of **feed-back/-forward loops** in network stability
  - **But**, sometimes rare perturbations lead to catastrophic failure



Not robust to turbulence

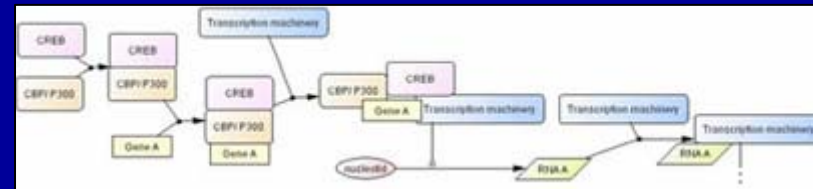
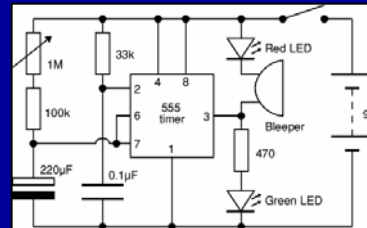
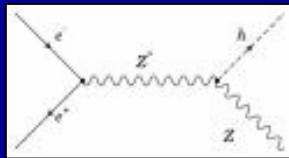


Robust to turbulence, but fragile to power failure!

- **Modularity**: maintain function by localizing damage

# Importance of Notation

- Visual



www.sbgm.org

- Computational

$$SU(3) \times SU(2) \times U(1)/Z_6$$

$$\begin{aligned} \nabla \cdot \mathbf{D} &= 4\pi\rho \\ \nabla \times \mathbf{H} &= \frac{4\pi}{c} \mathbf{J} + \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} \\ \nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned}$$

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level1"
level="1" version="2">
  <model name="gene_network_model">
    <listOfUnitDefinitions>
      ...
    </listOfUnitDefinitions>
    <listOfCompartments>
      ...
    </listOfCompartments>
    <listOfSpecies>
      ...
    </listOfSpecies>
    <listOfParameters>
      ...
    </listOfParameters>
    <listOfRules>
      ...
    </listOfRules>
    <listOfReactions>
      ...
    </listOfReactions>
  </model>
</sbml>
```

www.sbml.org

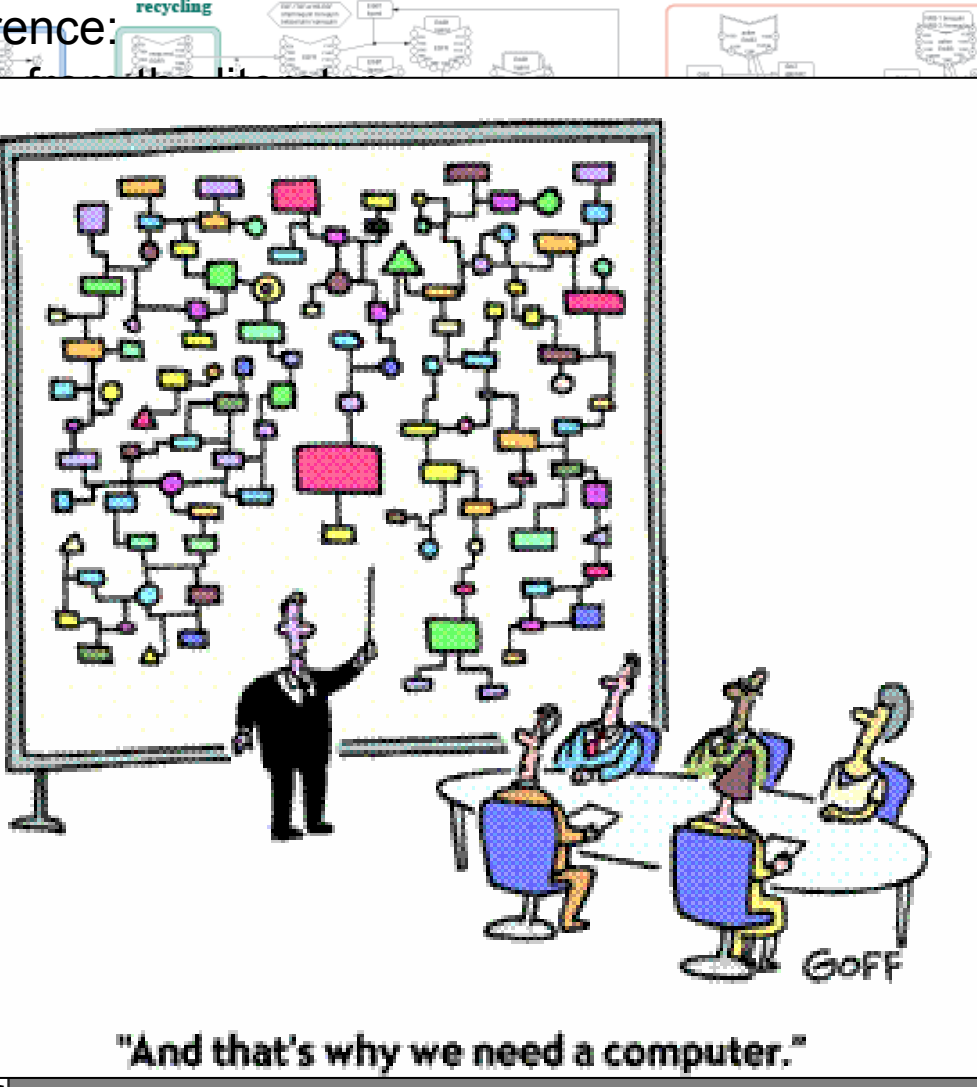
| Reaction                            | Rate   |
|-------------------------------------|--|
| $src \rightarrow RNAP$              | $V_i / (1 + P/K_i)$  |
| $RNAP \rightarrow waste$            | $V_{kd} \cdot RNAP$  |
| $RNA_{nuc} \rightarrow mRNA_{nuc}$  | $\frac{V_{m1} \cdot RNAP \cdot RNA_{nuc}}{K_{m1} + RNA_{nuc}}$ |
| $mRNA_{nuc} \rightarrow mRNA_{cyt}$ | $k_1 \cdot mRNA_{nuc}$   |
| $mRNA_{cyt} \rightarrow RNA_{cyt}$  | $\frac{V_{m2} \cdot mRNA_{cyt}}{mRNA_{cyt} + K_{m2}}$          |
| $RNA_{cyt} \rightarrow RNA_{nuc}$   | $k_2 \cdot RNA_{cyt}$  |

```
<ode(Xr1, time) = 50.0 {dimensionless}/
</ode>
(1.0 {dimensionless}) * exp(-(V-5.0) * (m1 / (1 + V * T))) *
exp(-(V-5.0) * (m1 / (1 + V * T)))
```

$$\frac{dXr1}{dt} = \frac{50.0}{1.0 + e^{\frac{-(V-5.0)}{9.0}}} \left( 1.0 - Xr1 \right) - 0.05 \cdot e^{\frac{-(V-20.0)}{15.0}}$$

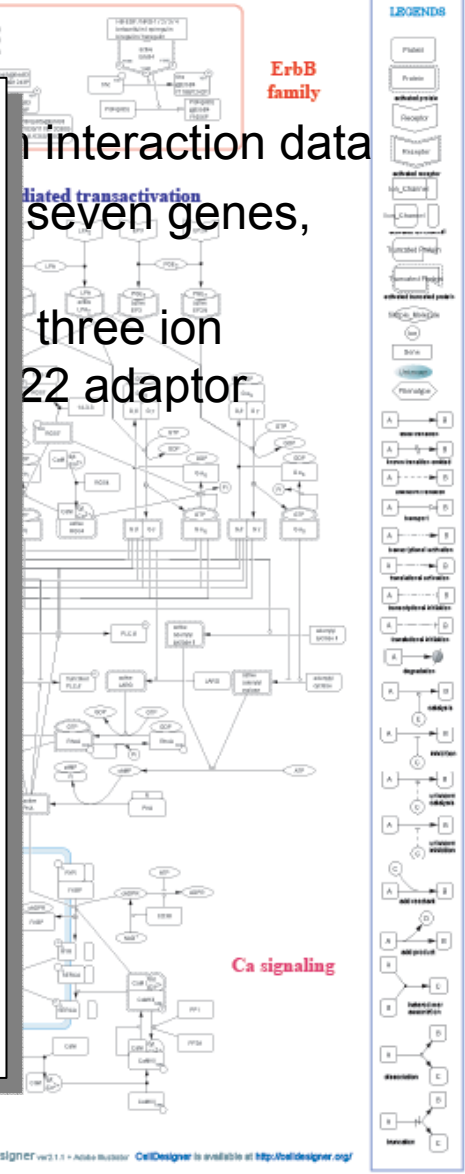
• Pathway inference:

- Manually
- Computed
- 202 proteins, and seven RNAs
- 10 ligands, 10 channels, 10 transcription factors, 10 proteins.
- Reactions catalyzed by
  - 131 state transitions
  - 34 transcription factors
  - 32 associated enzymes
  - 11 dissociation reactions
  - two transcription factors
- 247 interactions
  - 206 catalyzed by enzymes
  - 9 unknown
  - 16 inhibitory
  - 12 transcriptional
  - 4 transcriptional inhibitions.



"And that's why we need a computer."

• interaction data  
 • seven genes,  
 • three ion channels,  
 • 22 adaptor proteins

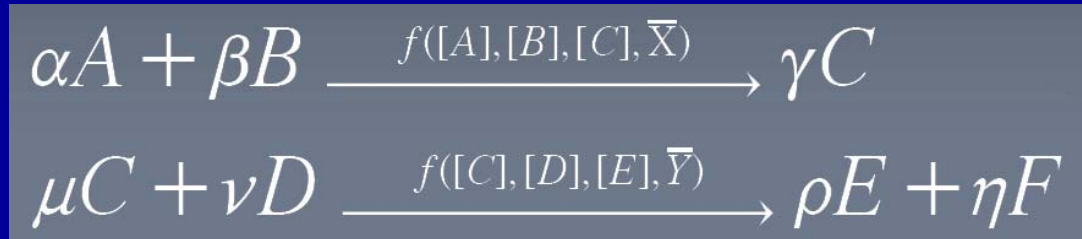




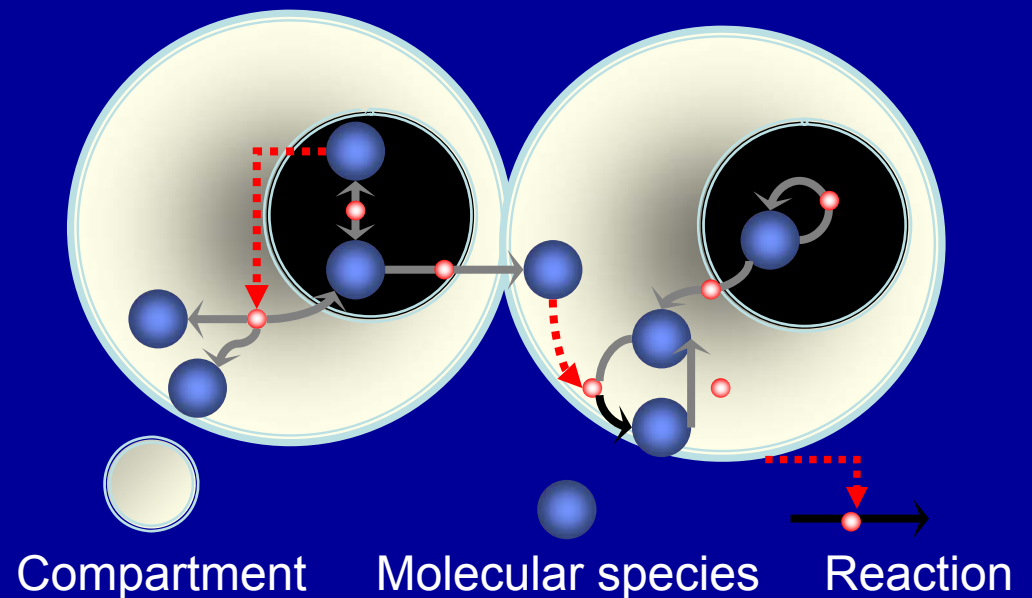
- Machine-readable, application-neutral format for representing computational models in systems biology
  - Expressed in XML
  - Intended for software tools
- Supported by >90 modelling packages
- API, Conversion tools (KEGG, CellML), Mathematica/Matlab plugins

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

- Focus: systems of biochemical reactions



- Also include:
  - Compartments
  - Rules/constraints
  - Discrete events



Source: M. Hucka, CalTech



- Level 3 Model (forthcoming) adds:
  - Graphical layout of models
  - Model composition (submodels)
  - Multistate complex species
  - Arrays of elements
  - 2-D & 3-D geometry

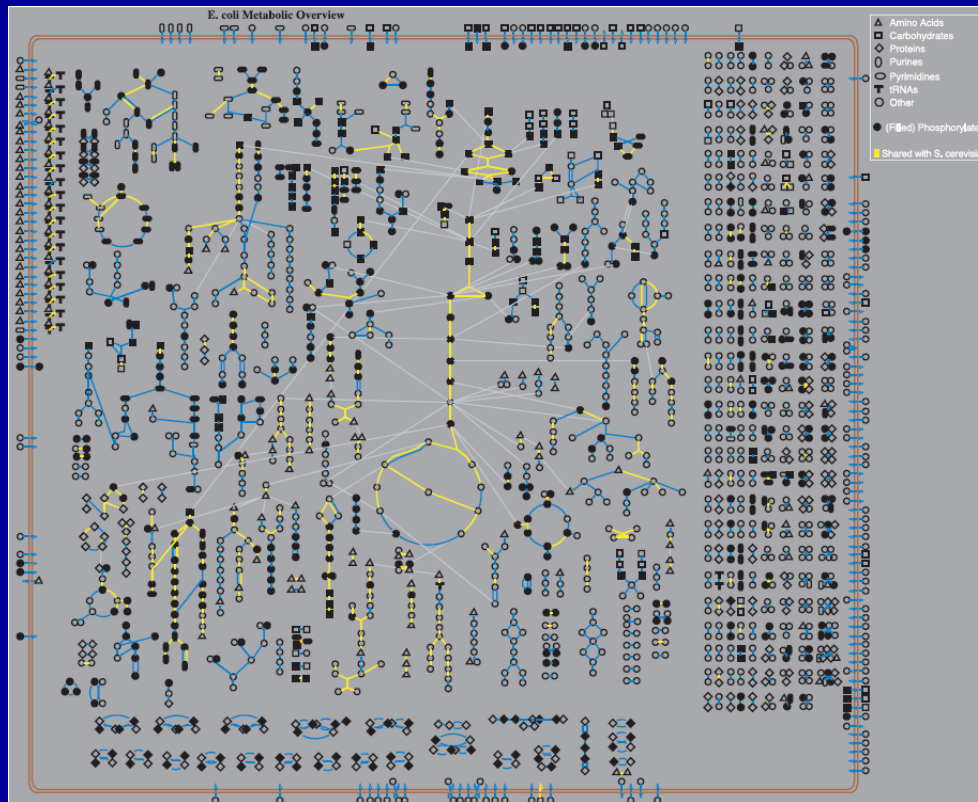
# Computational Pathway Inference

- ANOVA, hierarchical and k-means clustering, discriminant analysis, SOMs, SVMs
- Different methods give divergent results
- Data quality a *major* issue
- Mutual Information Methods
  - Systematic analysis of mutual information in
  - Boolean network state transition tables to extract minimal network architectures (“reverse engineering”)
  - Performs well for “low- $k$ ” networks (small number of inputs per gene); being extended to higher- $k$
  - Impact of measurement errors?
- Correlation Metric Methods
  - Use a time-lagged correlation metric as a measure of distance between reacting species



# Pathway Analysis: *BioCyc*

206 pathway databases, derived from literature and by computation [www.biocyc.org](http://www.biocyc.org)

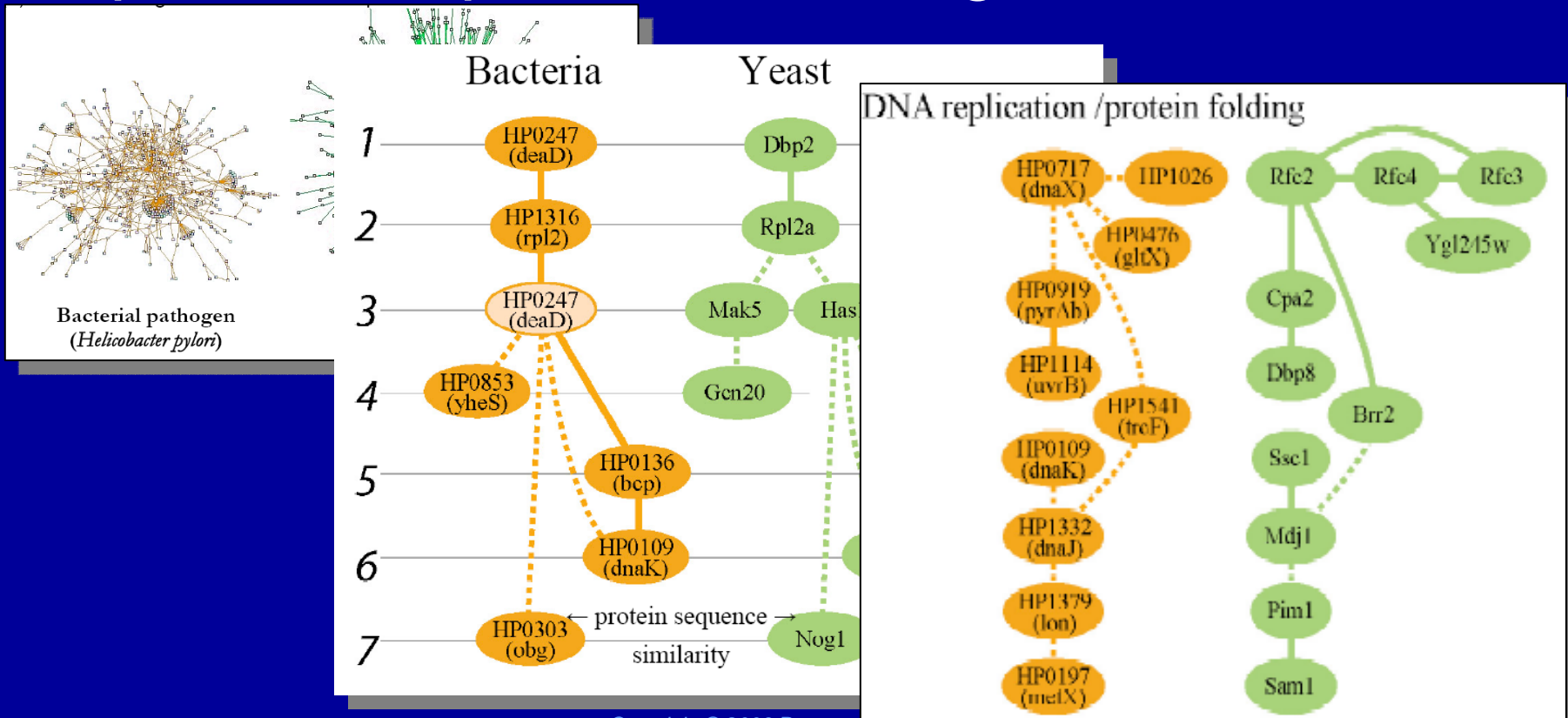


P. Karp, *Science* 293 (2001)

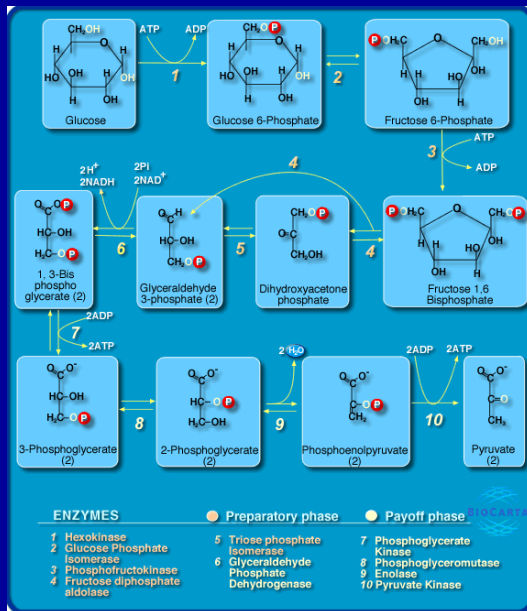
- **Qualitative simulation by symbolic computation on the network:**
  - represent network as a set of “production rules”  $A \wedge B \rightarrow X$
  - answer questions such as “can this network produce X?”
- **PathoLogic: pathway inference from whole genome data**
  - two inputs
    - annotated genome sequence that includes locations and predicted functions of genes within the genome,
    - reference pathway DB.
  - output
    - new PGDB that includes a set of pathways predicted to be present
  - match enzymes in the annotated genomes against enzymes in the *MetaCyc* DB
  - compute a score for presence of different pathways on basis of number of matching enzymes, and their positions within the pathway

# Pathway Analysis: *PathBlast*

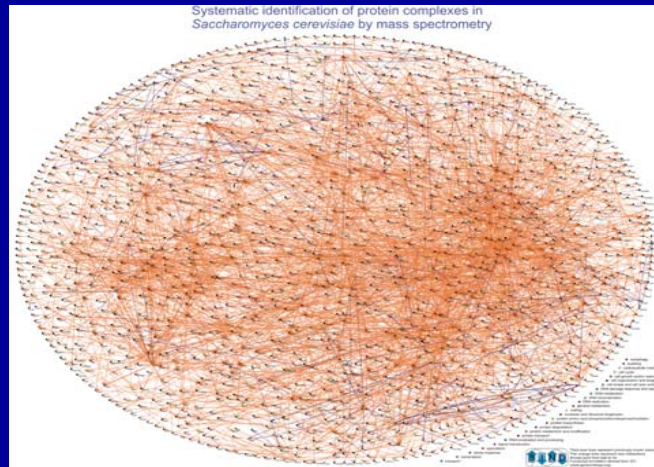
- Function prediction by cross-species protein sequence matching



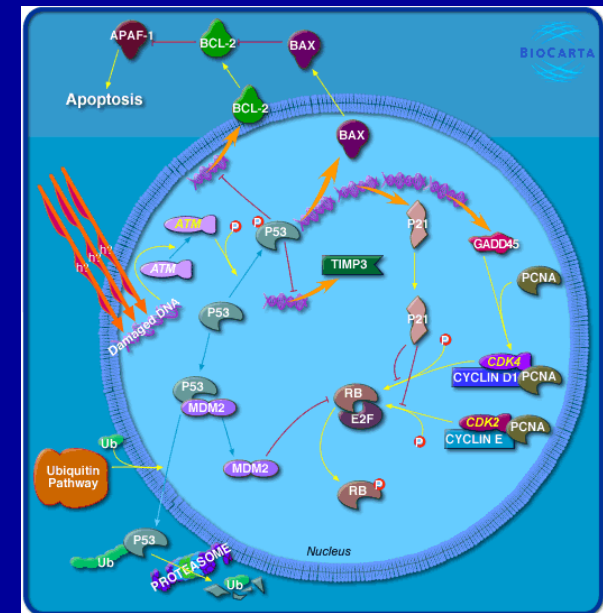
# BioPAX: ontology and data exchange format for pathways



Metabolic Pathways



Molecular Interaction Networks



Signaling Pathways



# BioPAX: so many databases...

## 138 Biological Pathway Databases

**Protein-Protein Interactions**  
 ABCdb - Archaea and Bacteria ABC transporter database  
 AFCS - Alliance for Cellular Signalling Molecule Pages Database  
 AIFuse - Functional Associations of Proteins in Complete Genomes  
 ASEdb - Alanine Scanning Energetics Database  
 ASPD - Artificial Selected Proteins/Peptides Database  
 BID - Binding Interface Database  
 BIND - Biomolecular Interaction Network Database  
 BindingDB - The Binding Database  
 BRITE - Biomolecular Relations in Information Transmission and Expression  
 DDIB - Database of Domain Interactions and Bindings  
 DIP - Database of Interacting Proteins  
 DRC - Database of Ribosomal Crosslinks  
 DSM - Dynamic Signaling Maps  
 FIMM - Functional Molecular Immunology  
 FlyNets - FlyNets  
 FusionDB - Prokaryote Gene Fusion Events  
 GPCR-PD - G protein-coupled receptors protein database  
 GRID - General Repository for Interaction Datasets  
 GroEL PPI - Proteins that interact with GroEL and factors that affect their release  
 HIVMID - HIV Molecular Immunology Database  
 HPRD - Human Protein Reference Database  
 HSV1 PPI - Protein-Protein Interactions Table for Human herpesvirus 1  
 HumanPSD - Human Proteome Survey Database  
 ICBS - Inter-Chain Beta-Sheets  
 IntAct - IntAct  
 INTERACT - INTERACT Protein-protein interaction database  
 InterDom - Database of Interacting Domains  
 Interolog - Interolog/Regulog Database  
 JenPep - JenPep Peptide Binding Database  
 MHCPEP - Database of MHC binding peptides  
 MINT - Molecular Interactions Database  
 MIPS CYGD - MIPS Comprehensive Yeast Genome Database  
 MPID - MHC-Peptide Interaction Database  
 MycoPathPD - Human Fungal Pathogens Proteome Database  
 NetPro - Molecular Connections NetPro  
 ooTFD - Object Oriented Transcription Factors Database

Pathcalling - Curagen Pathcalling  
 PathDB - Pathways Database  
 Phospho.ELM - Post-translational phosphorylation database  
 PhosphoBase - Database of phosphorylation sites  
 PhosphoSite - Cell Signaling Technology's PhosphoSite Database  
 PIMdb - Drosophila Protein Interaction Map Database  
 PIMRider - Protein Interaction Map - Hybrigenics  
 PINdb - Proteins Interacting in the Nucleus database  
 PombePD - Schizosaccharomyces pombe Proteome Database  
 PPIID - Protein-Protein Interaction Database  
 PPIV (FANTOM) - RIKEN FANTOM Protein Protein Interaction Viewer  
 PQS - Protein Quaternary Structure database  
 Predictome - Predictome  
 ProChart - ProChart database of signal transduction pathway information  
 Prolinks - Prolinks  
 ProMesh - ProMesh Protein-Protein Interaction Database  
 ProNet - Protein-protein interaction Database  
 PubGene - PubGene  
 S/MARL\_DB - The S/MAR transaction DataBase  
 Scansite - Scansite  
 SPID - Subtilis Protein interaction Database  
 SPIN-PP - Surface Properties of Interfaces - Protein-Protein Interfaces  
 STRING - Search Tool for the Retrieval of Interacting Genes/Proteins  
 SYFPEITHI - Database of MHC Ligands and Peptide Motifs  
 WormPD - Caenorhabditis elegans Proteome Database  
 YPD - Yeast Proteome Database

**Metabolic Pathways**  
 aMAZE - Protein Function and Biochemical Pathways Project  
 BioCyc - BioCyc Knowledge Library  
 BRENDA - Comprehensive Enzyme Information System  
 CellML Repository - CellML Model Repository  
 EcoCyc - Encyclopedia of E. coli Genes and Metabolism  
 EMP - Enzymes and Metabolic Pathways Database  
 ENZYME - Enzyme nomenclature database  
 GOLD.db - Genomics of Lipid-associated Disorders

Indigo - Gene Neighborhoods and Codon Usage  
 IntEnz - Integrated relational Enzyme database  
 KEGG - Kyoto Encyclopedia of Genes and Genomes  
 LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways  
 Malaria - Malaria Parasite Metabolic Pathways  
 MetaCore - MetaCore pathway database  
 MetaCyc - Metabolic Pathway Database  
 MIPS CYGD - MIPS Comprehensive Yeast Genome Database  
 MPB - Metabolic Pathways of Biochemistry  
 NetBiochem - Medical Biochemistry Resource  
 PathArt - Pathway Articulor  
 PathDB - Pathways Database

**Signaling Pathways**  
 AFCS - Alliance for Cellular Signalling Molecule Pages Database  
 aMAZE - Protein Function and Biochemical Pathways Project  
 CellML Repository - CellML Model Repository  
 CMAP Pathway - CMAP Pathway Interaction Database  
 COPE - Cytokines Online Pathfinder Encyclopedia  
 CSNDB - Cell Signalling Networks Database  
 DOQCS - Database of Quantitative Cellular Signaling  
 DSM - Dynamic Signaling Maps  
 eMIM - Electronic Molecular Interaction Map  
 GOLD.db - Genomics of Lipid-associated Disorders  
 MetaCore - MetaCore pathway database

ROSPath - Reactive Oxygen Species related Signaling Pathway  
 SBML Model Repository - SBML Model Repository  
 Sentra - Sentra Signal Transduction Database  
 SigPath - Signaling Pathway Information System  
 SPAD - Signaling Pathway Database  
 STCDB - Signal Transduction Classification Database  
 STKE - Signal Transduction Knowledge Environment  
 TRANSPATH - Signal Transduction Browser  
 TRRD - Transcription Regulatory Regions Database

**Pathway Diagrams**  
 BBID - Biological Biochemical Image Database  
 BioCarta - BioCarta Pathway Diagrams  
 BMPH - Boehringer Mannheim Biochemical Pathways  
 Well Chart  
 CMAP Pathway - CMAP Pathway Interaction Database  
 DSM - Dynamic Signaling Maps  
 eMIM - Electronic Molecular Interaction Map  
 HPRD - Human Protein Reference Database  
 KEGG - Kyoto Encyclopedia of Genes and Genomes  
 KMIM - Kohn Molecular Interaction Maps  
 Malaria - Malaria Parasite Metabolic Pathways  
 MIPS CYGD - MIPS Comprehensive Yeast Genome Database  
 MPB - Metabolic Pathways of Biochemistry  
 NetBiochem - Medical Biochemistry Resource  
 PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base  
 SPAD - Signaling Pathway Database

**Transcription Factors / Genetic Regulatory Networks**  
 COMPTEL - Database on Composite Regulatory Elements  
 DBTBS - Database of Bacillus subtilis Promoters and Transcription Factors  
 DPInteract - DNA-Protein Interactions Database  
 GeNet - Gene Networks Database  
 HoxPro - HOX Pro  
 Interolog - Interolog/Regulog Database  
 JASPAR - JASPAR Transcription Factor Binding Profile Database  
 ooTFD - Object Oriented Transcription Factors Database  
 PRODORIG - Prokaryotic database of gene regulation  
 RegulonDB - Database on Transcriptional Regulation and Genome Organization

SCPD - The Promoter Database of Saccharomyces cerevisiae  
 TRANSFAC - Transcription Factor Database  
 TRRD - Transcription Regulatory Regions Database

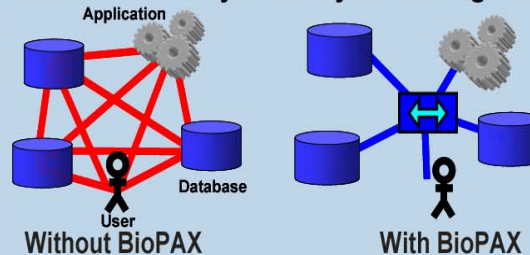
**Protein-Compound Interactions**  
 HetPDB\_Navi - Het-PDB Navi  
 MDB - Metalloprotein Database  
 NRR - Nuclear Receptor Resource  
 ORDB - Olfactory Receptor Database  
 PDSP - Psychoactive Drug Screening Program Kit Database  
 Reibase - Protein-ligand database query tool  
 TTD - Therapeutic Target Database

**Genetic Interaction Networks**  
 BIND - Biomolecular Interaction Network Database  
 GeneNet - Genetic Networks  
 GenePath - GenePath  
 GRID - General Repository for Interaction Datasets  
 KNIFE - Drosophila pattern formation knowledge base  
 MIPS CYGD - MIPS Comprehensive Yeast Genome Database

**Protein Sequence Focused**  
 InBase - The Intein Database  
 MEROPS - MEROPS Peptidase Database  
 NRR - Nuclear Receptor Resource  
 ORDB - Olfactory Receptor Database  
 PhosphoBase - Database of phosphorylation sites  
 REBASE - Restriction Enzyme Database  
 Sentra - Sentra Signal Transduction Database  
 TGDB - Tumor Gene Database

**Other**  
 SELEX\_DB - Randomized DNA/RNA sequence database  
 AARSDB - Aminoacyl-tRNA Synthetase Database  
 MedGene - MedGene  
 AANT - Amino Acid-Nucleotide Interaction Database  
 ProNIT - Thermodynamic Database for Protein-Nucleic Acid Interactions  
 TCDB - Transport Classification Database  
 TransportDB - TransportDB  
 DPIDB - DNA-Protein Interaction Database

### Motivation: Easy Pathway Data Integration



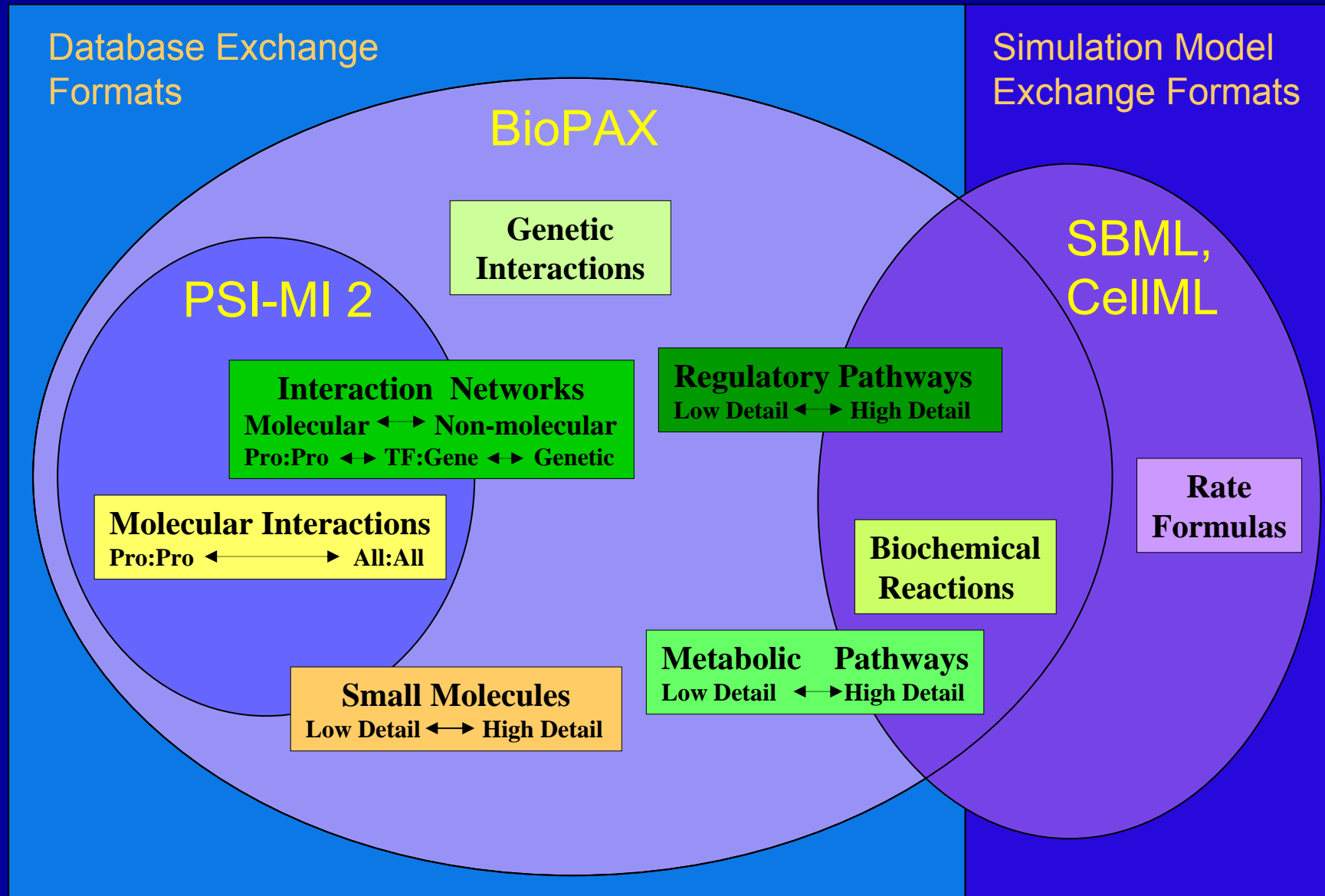
PATIKA - Pathway Analysis Tool for integration and Knowledge Acquisition  
 PDS - Pathways Database System  
 PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base  
 Reactome - Reactome KnowledgeBase  
 SBML Model Repository - SBML Model Repository  
 SoyBase - USDA-ARS Soybean Genetics and Genome Database  
 UMSB - Biocatalysis/Biodegradation Database  
 WIT - What Is There?

PathArt - Pathway Articulor  
 PathDB - Pathways Database  
 Pathways Knowledge Base - Ingenuity Pathways Knowledge Base  
 PATIKA - Pathway Analysis Tool for integration and Knowledge Acquisition  
 PDS - Pathways Database System  
 PhosphoSite - Cell Signaling Technology's PhosphoSite Database  
 Reactome - Reactome KnowledgeBase

Pathway Resource List <http://cbio.mskcc.org/prl/>

Database Exchange  
Formats

Simulation Model  
Exchange Formats



# Aggregation, Integration, Inference

1. Multiple kinds of pathway databases
  - metabolic
  - molecular interactions
  - signal transduction
  - gene regulatory
2. Constructs designed for integration
  - DB References
  - XRefs (Publication, Unification, Relationship)
  - Synonyms
  - Provenance (not yet implemented)
3. OWL DL – to enable reasoning

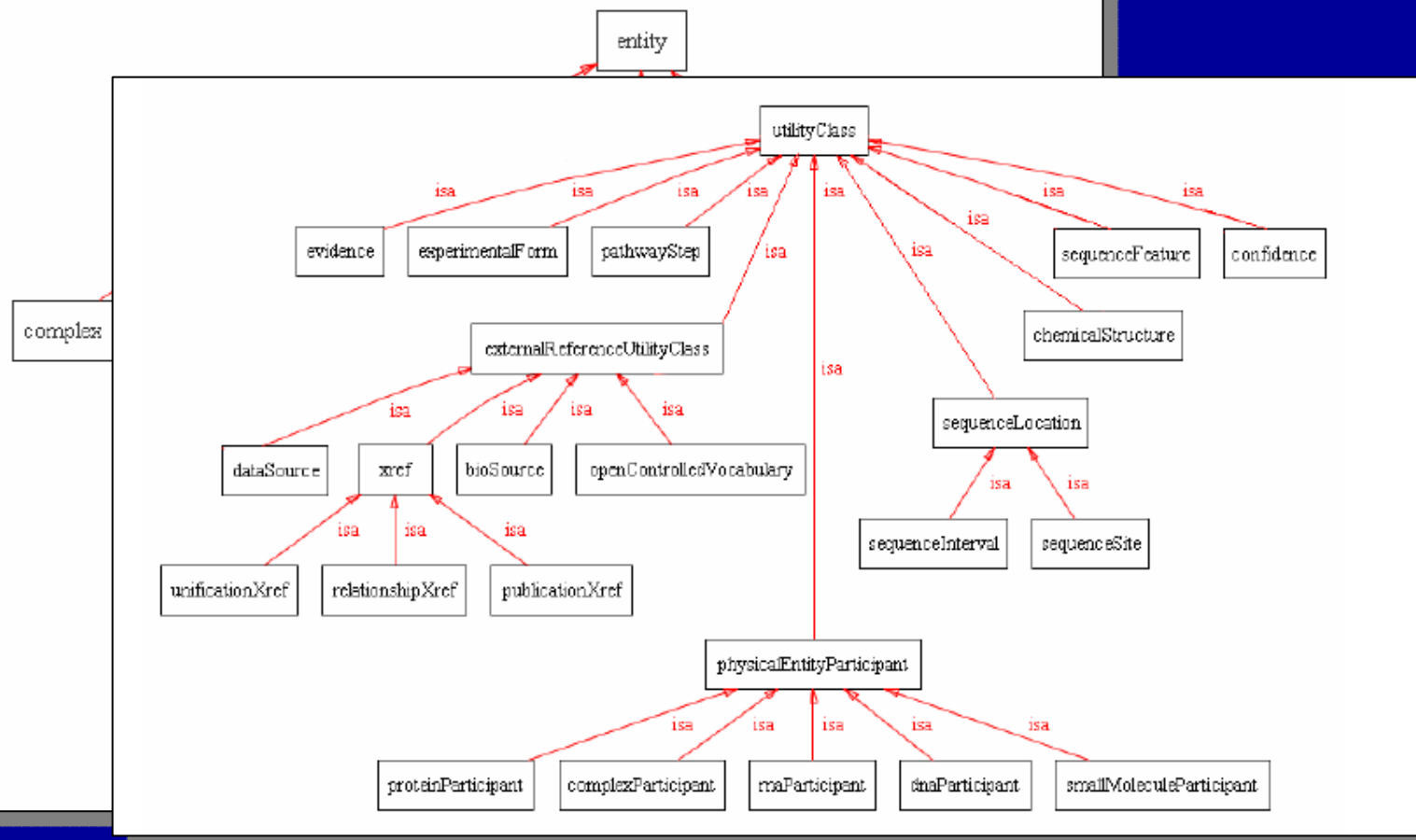
# BioPAX: use of other ontologies

- Conceptual framework based upon existing DB schemas:
  - *aMAZE*, *BIND*, *EcoCyc*, *WIT*, *KEGG*, *Reactome*, etc.
  - Allows wide range of detail, multiple levels of abstraction
- Uses pointers to existing ontologies to provide supplemental annotation where appropriate
  - Cellular location → *GO Component*
  - Cell type → *Cell.obo*
  - Organism → *NCBI taxon DB*
- Incorporate other standards where appropriate
  - Chemical structure → *SMILES*, *CML*, *INCHI*
- Interoperate with existing standards (*RDF/OWL*, *LSID*, *SBML*, *PSI*, *CellML Metadata Standard*)



# BioPAX Ontology

## Summary of BioPAX Class Structure



# BIOMODELS.NET

- Store, search and retrieve published mathematical models of biological interests
- Annotated and linked to relevant data resources, such as publications, databases of compounds and pathways, controlled vocabularies, etc.
- *MIRIAM: Minimum information requested in the annotation of biochemical models*

|                    | Release 1<br>(April 2005) | Release 2<br>(June 2005) | Release 3<br>(July 2005) | Release 4<br>(January 2006) |
|--------------------|---------------------------|--------------------------|--------------------------|-----------------------------|
| <b>Models</b>      | 20                        | 30                       | 44                       | 50                          |
| <b>Species</b>     | 322                       | 425                      | 596                      | 761                         |
| <b>Reactions</b>   | 631                       | 736                      | 943                      | 1163                        |
| <b>Annotations</b> | 1084                      | 1609                     | 2373                     | 3126                        |

| Topic                     | # models |
|---------------------------|----------|
| <b>Cell communication</b> | 24       |
| <b>Cell cycle</b>         | 6        |
| <b>Metabolism</b>         | 10       |
| <b>Circadian rhythm</b>   | 7        |
| <b>Other</b>              | 3        |

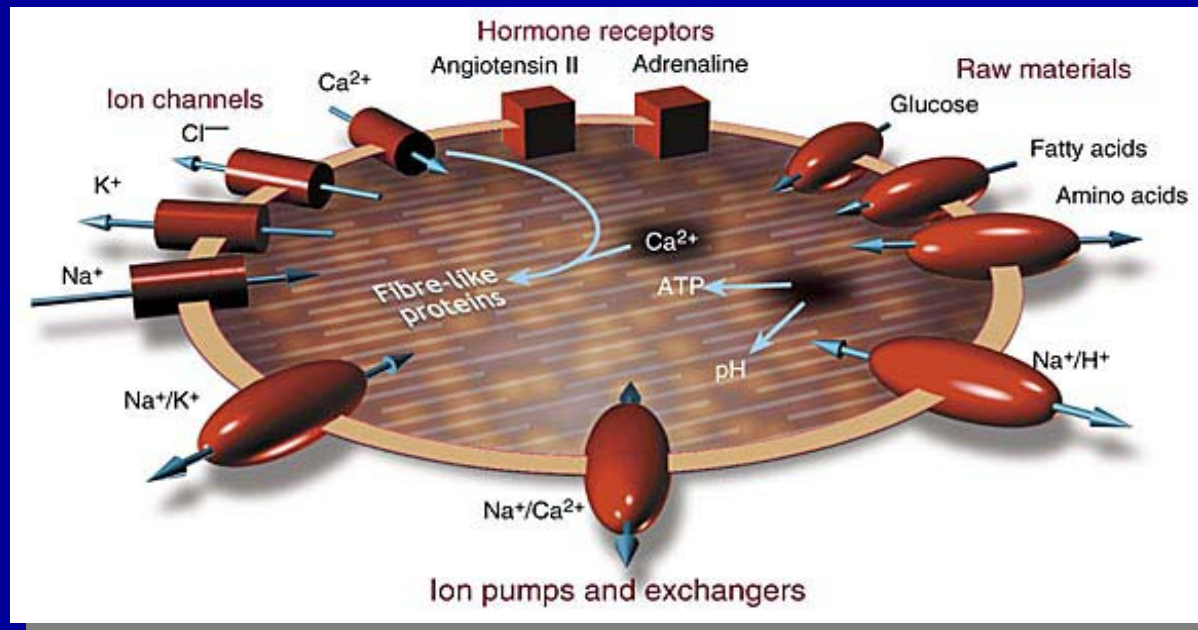
# Whole-organ Models: “Physiomics”



# Cardiac Modelling Domains

- Cardiac Myocyte Models
  - Membrane currents
  - Intracellular  $\text{Ca}^{++}$  transport
  - Excitation-contraction coupling
  - Interval/force relationships
- Micro/Meso-scale **Anatomical Structure**
  - Histological models
  - DTMR imaging
  - Electrical bundle/node geometry
  - Vascular system dynamics

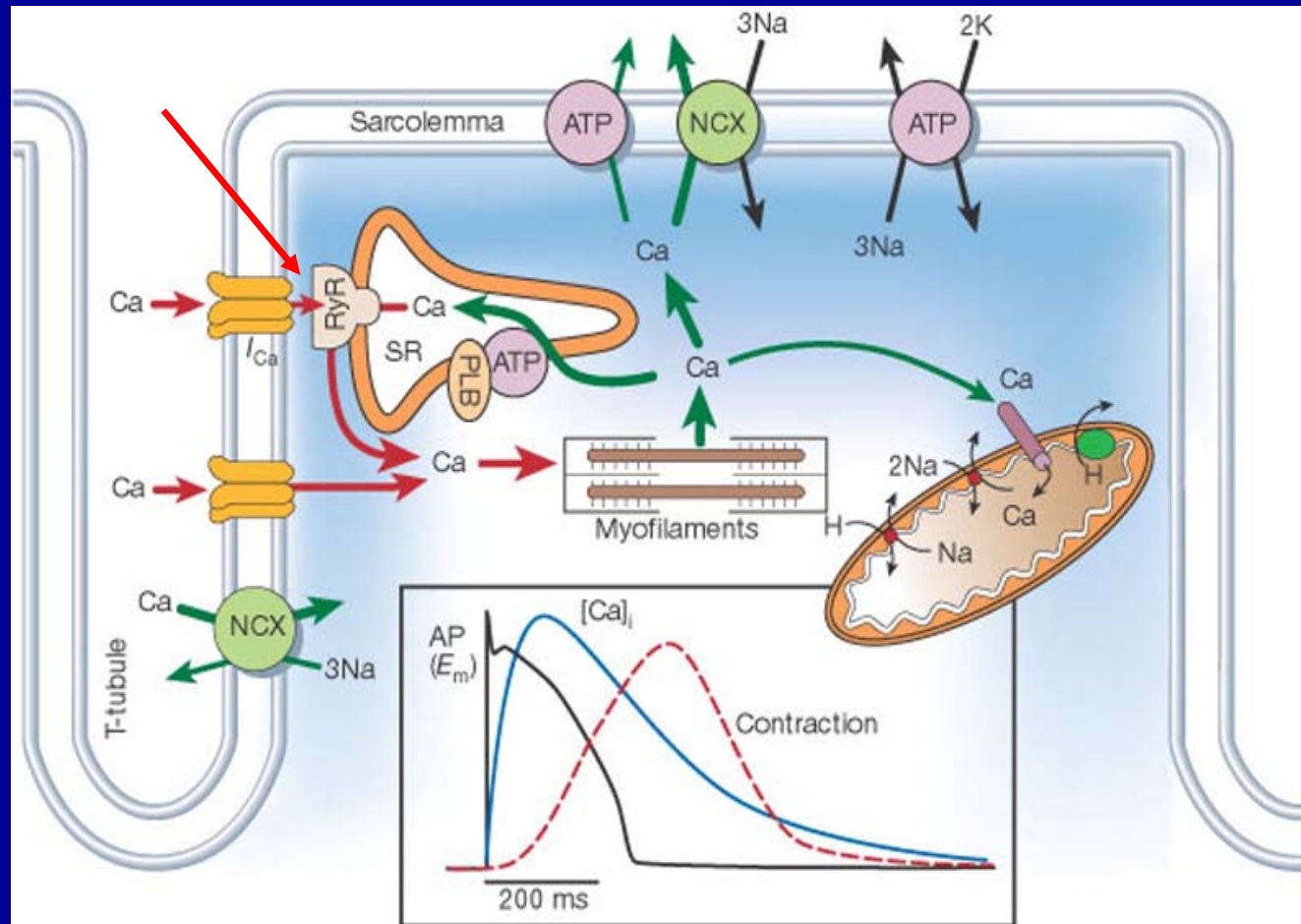
# Cardiac Myocyte Model



Source: *New Scientist*

- Model of voltage-gated membrane currents, membrane transporters,  $\text{Ca}^{2+}$  sequestration, intracellular  $\text{Na}^+$ ,  $\text{Ca}^{2+}$  and  $\text{K}^+$
- Voltage- and time-dependent behaviour described by systems of non-linear ODEs

# Cardiac Myocyte Model



Bers *Nature* 415: 198-205 (2002)

# Modelling Ventricular Myocytes

Single Cell

$$dV/dt = -\frac{I_{ion} + I_{stim}}{C_m}$$

2-D continuous sheet of cells

$$\partial V/\partial t = -\frac{I_{ion} + I_{stim}}{C_m} + \frac{1}{\rho_x S_x C_m} \left( \frac{\partial^2 V}{\partial x^2} \right) + \frac{1}{\rho_y S_y C_m} \left( \frac{\partial^2 V}{\partial y^2} \right)$$

where  $I_{stim}$  is the externally applied stimulus current,  $C_m$  is the cell capacitance per unit surface area,  $\rho_x$  and  $\rho_y$  are the cellular resistivity in the  $x$  and  $y$  direction,  $S_x$  and  $S_y$  are the surface to volume ratio in the  $x$  and  $y$  direction, and  $I_{ion}$  is the sum of all transmembrane ionic currents given by the following equation:

$$I_{ion} = I_{Na} + I_{K1} + I_{to} + I_{Kr} + I_{Ks} + I_{CaL} + I_{NaCa} + I_{NaK} + I_{pCa} + I_{pK} + I_{bCa} + I_{bNa}$$

Ten Tusscher K. H.W. J., Noble D., Noble P. J. and Panfilov A. V. "A model for human ventricular tissue," *Am J Physiol Heart Circ Physiol* (December, 2003)



| Membrane Current  | Model  |
|---|--|
| <i>Fast Na<sup>+</sup> current: I<sub>Na</sub></i>  | $I_{Na} = G_{Na} m^3 h j (V - E_{Na})$   |
| <i>L-type Ca<sup>2+</sup> current: I<sub>CaL</sub></i>  | $I_{CaL} = G_{CaL} d f f_{Ca} A \frac{VF^2 Ca_i e^{2VF/RT} - 0.341 Ca_o}{RT e^{2VF/RT} - 1}$   |
| <i>Transient outward current: I<sub>to</sub></i>  | $I_{to} = G_{to} r s (V - E_K)$  |
| <i>Slow delayed rectifier current: I<sub>Ks</sub></i>   | $I_{Ks} = G_{Ks} x_s^2 (V - E_{Ks})$   |
| <i>Rapid delayed rectifier current: I<sub>Kr</sub></i>  | $I_{Kr} = G_{Kr} \sqrt{\frac{K_o}{5.4}} x_{r1} x_{r2} (V - E_K)$   |
| <i>Inward rectifier K<sup>+</sup> current: I<sub>K1</sub></i>   | $I_{K1} = G_{K1} \sqrt{\frac{K_o}{5.4}} x_{K1\infty} (V - E_K)$  |
| <i>Na<sup>+</sup>/Ca<sup>2+</sup> exchanger current, Na<sup>+</sup>/K<sup>+</sup> pump current, plateau and background currents</i> | $I_{NaCa} = k_{NaCa} \frac{e^{\frac{VF}{RT}} Na_i^3 Ca_o - e^{\frac{(\gamma-1)VF}{RT}} Na_o^3 Ca_i \alpha}{(K_{mNa}^3 + Na_o^3)(K_{mCa} + Ca_o)(1 + k_{sat} e^{\frac{(\gamma-1)VF}{RT}})}$ $I_{NaK} = P_{NaK} \frac{K_o}{K_o + K_{mK}} \frac{Na_i}{Na_i + K_{mNa}} \frac{1}{1 + 0.1245 e^{-0.1VF/RT} + 0.0353 e^{-VF/RT}}$ |
| <i>I<sub>pCa</sub> I<sub>pK</sub></i>   | $I_{pCa} = G_{pCa} \frac{Ca_i}{Ca_i + K_{pCa}} \quad I_{pK} = G_{pK} \frac{V - E_K}{1. + e^{(25-V)/5.98}}$   |
| <i>background sodium and calcium leakage currents</i>   | $I_{bNa} = G_{bNa} (V - E_{Na})$ $I_{bCa} = G_{bCa} (V - E_{Ca})$  |

# Intracellular Ion Dynamics

## Calcium

$$I_{leak} = V_{leak}(Ca_{SR} - Ca_i)$$

$$I_{up} = \frac{V_{maxup}}{1 + K_{up}^2/Ca_i^2}$$

$$I_{rel} = (a_{rel} \frac{Ca_{SR}^2}{b_{rel}^2 + Ca_{SR}^2} + c_{rel})dg$$

$$Ca_{ibufc} = \frac{Ca_i \times Buf_c}{Ca_i + K_{bufc}}$$

$$dCa_{itotal}/dt = -\frac{I_{CaL} + I_{bCa} + I_{pCa} - 2I_{NaCa}}{2V_cF} + I_{leak} - I_{up} + I_{rel}$$

$$Ca_{srbufsr} = \frac{Ca_{sr} \times Buf_{sr}}{Ca_{sr} + K_{bufsr}}$$

$$dCa_{srtotal}/dt = \frac{V_c}{V_{sr}}(-I_{leak} + I_{up} - I_{rel}),$$

## Na<sub>i</sub> and K<sub>i</sub>

$$dNa_i/dt = -\frac{I_{Na} + I_{bNa} + 3I_{NaK} + 3I_{NaCa}}{V_cF}$$

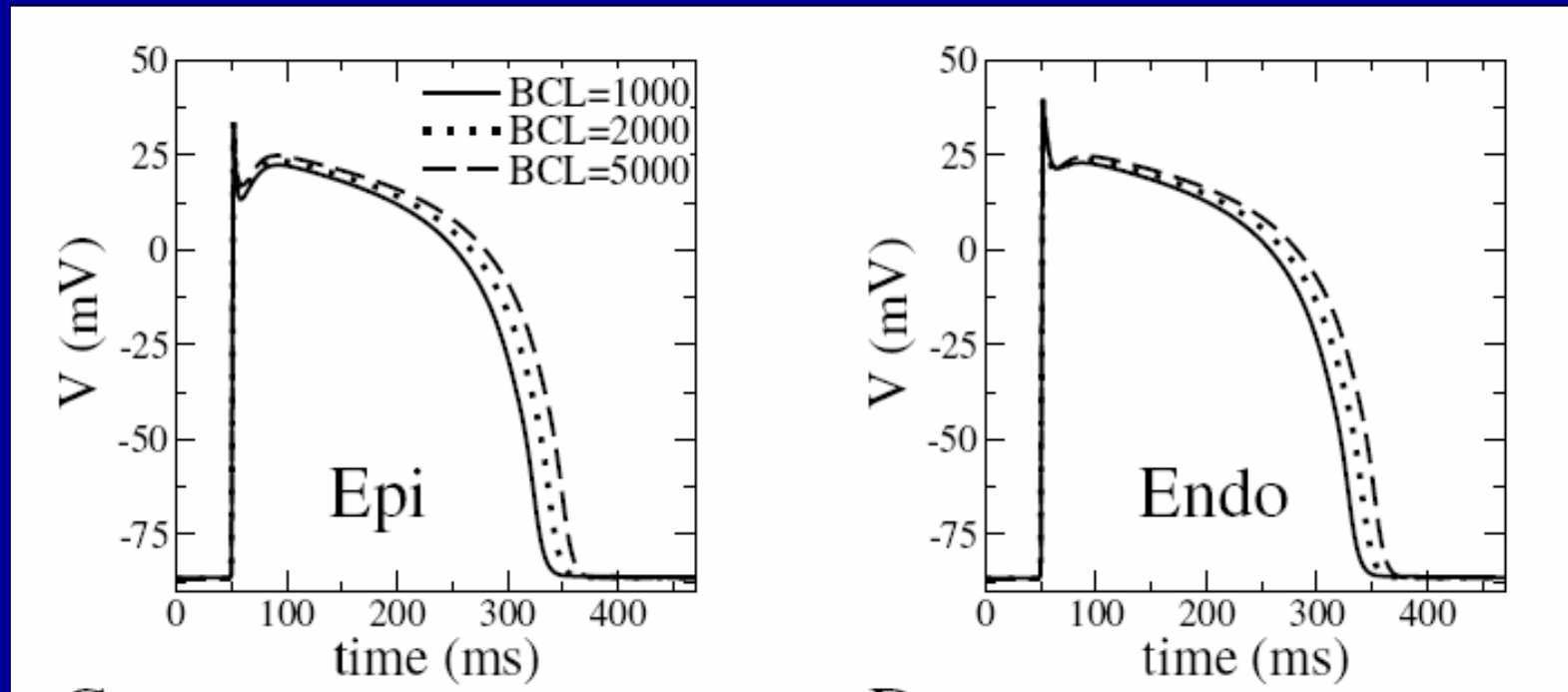
$$dK_i/dt = -\frac{I_{K1} + I_{to} + I_K - 2I_{NaK} + I_{pK} + I_{stim} - I_{ax}}{V_cF}$$

Ten Tusscher K. H.W. J., Noble D., Noble P. J. and Panfilov A. V. "A model for human ventricular tissue," *Am J Physiol Heart Circ Physiol* (December, 2003)

# Myocyte Model Parameters

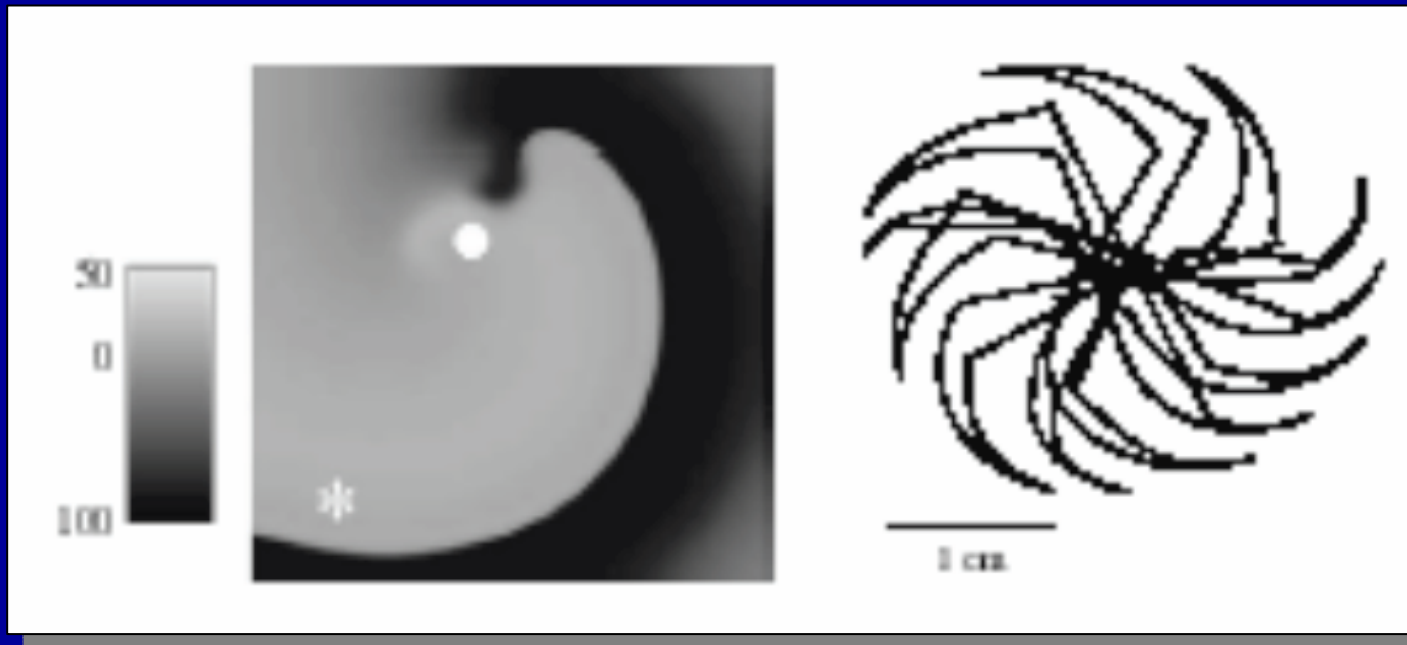
- $3 \times 10^7$  myocytes/heart, each unique in geometry, orientation
- 34 coupled non-linear ODEs/myocyte
- Integrated using modified Runge-Kutta 4<sup>th</sup>-order adaptive step algorithm
  - Maximum step size = 100  $\mu\text{sec}$
  - Maximum error tolerance =  $10^{-6}$
- Some processes have very rapid kinetics, so use a stiff integrator (DVODE) to improve performance

# Myocyte Model: the calculated Action Potential



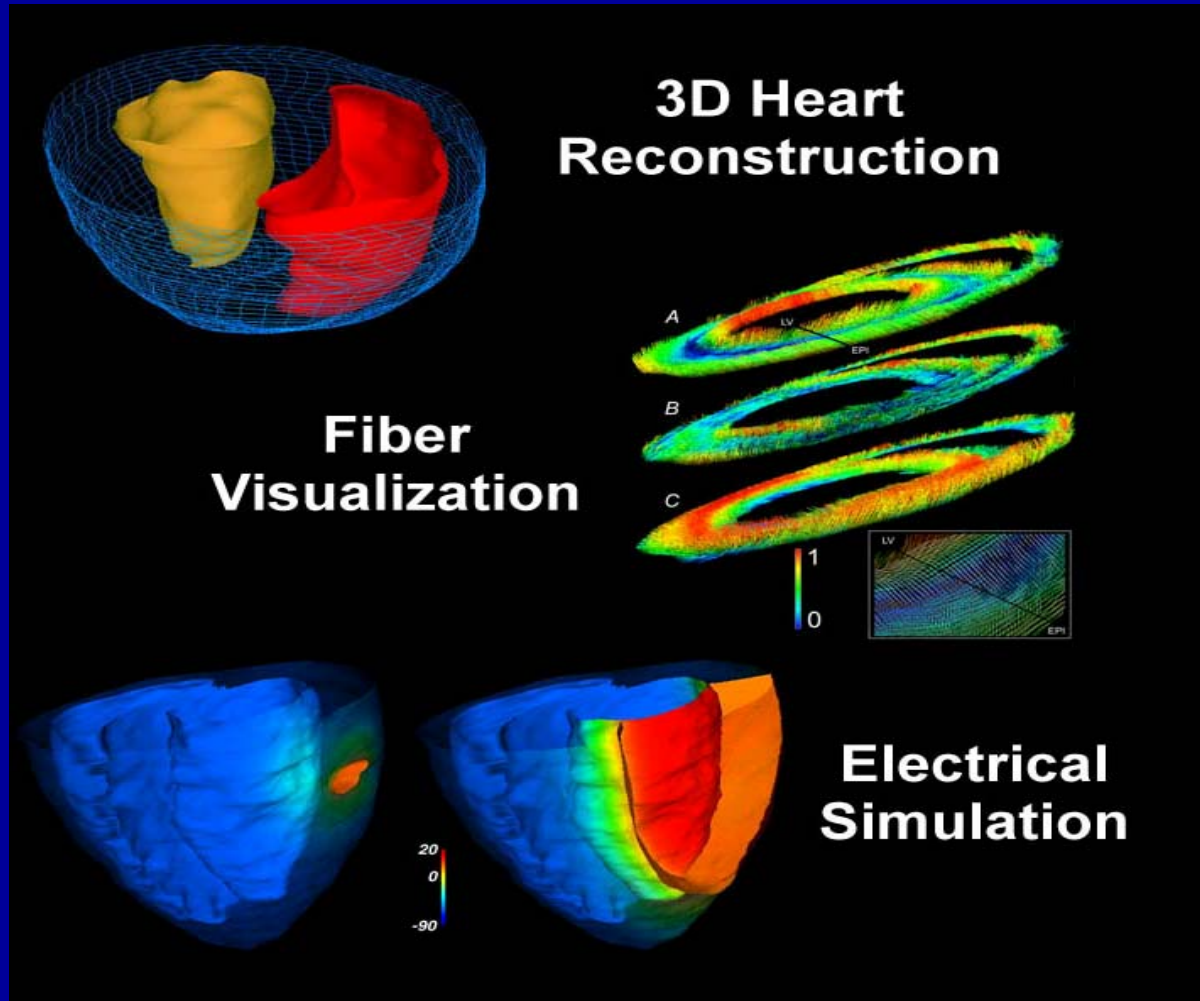
Ten Tusscher K. H.W. J., Noble D., Noble P. J. and Panfilov A. V. "A model for human ventricular tissue," *Am J Physiol Heart Circ Physiol* (December, 2003) [BCL = "basic cycle length"]

# Myocyte Model: complex non-linear phenomena



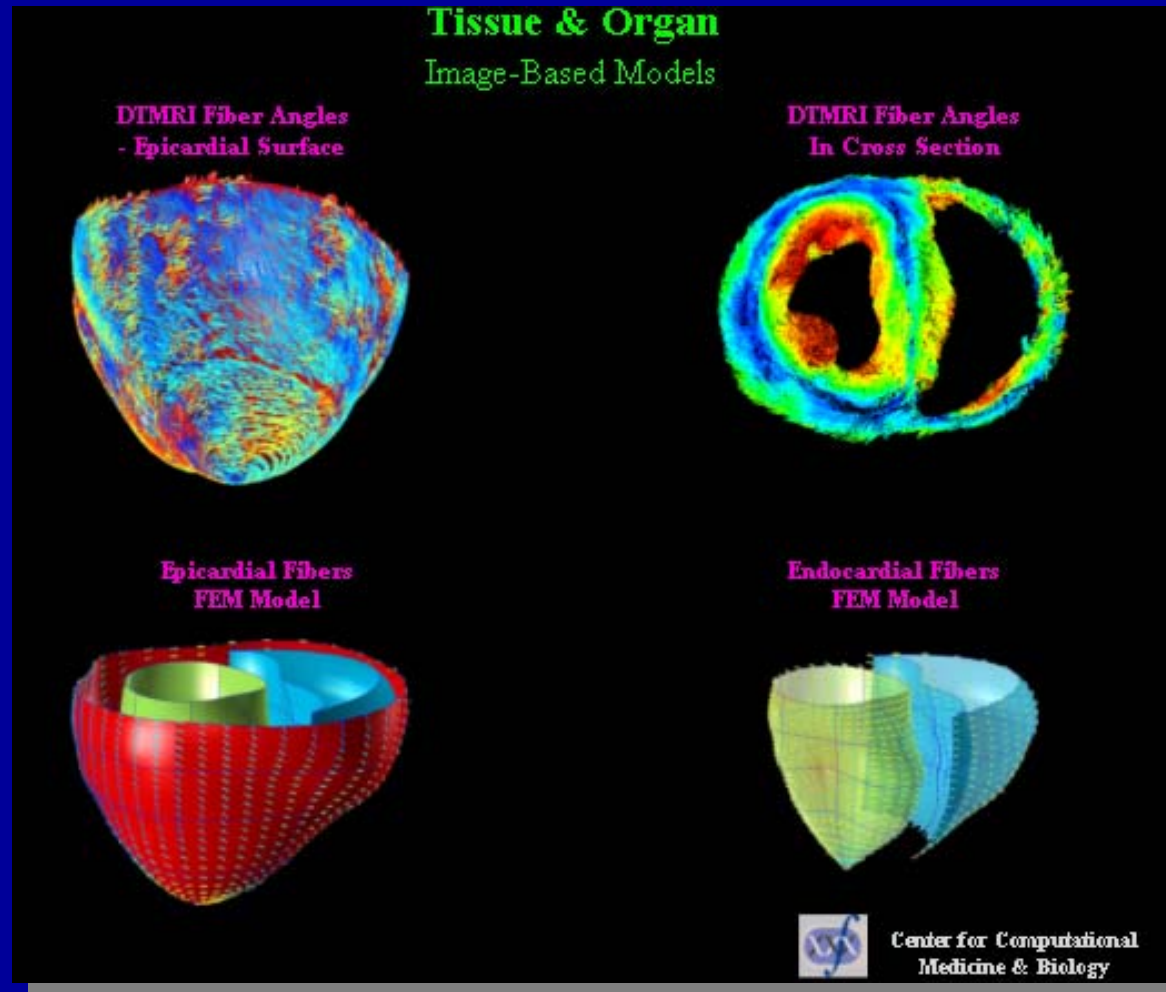
Ten Tusscher K. H.W. J., Noble D., Noble P. J. and Panfilov A. V. "A model for human ventricular tissue," *Am J Physiol Heart Circ Physiol* (December, 2003) [BCL = "basic cycle length"]

# Whole-heart Modelling



R. Winslow JHU

# Image-based Models



R. Winslow JHU

# Integrated Excitation Model

- Treat each myocardial domain as a continuum: use average conductivity, trans-membrane voltages
- Coupled parabolic and elliptic equations which must be satisfied by the myocardium and the surrounding medium (can be reduced to a parabolic reaction-diffusion equation)

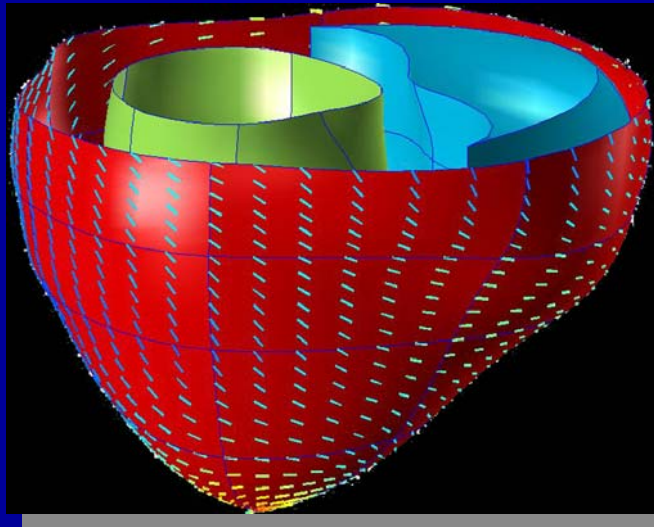


# Integrated Excitation Model

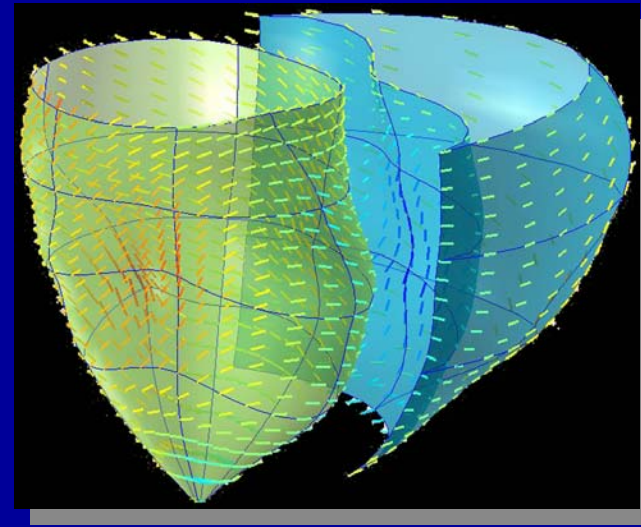
- Parallelized and solved on a 4-node x 8-processor IBM SP Power<sup>3</sup> using OpenMP
- ~250  $\mu\text{m}$  mesh =  $3 \times 10^6$  lattice points =  
     $3 \times 10^7$  coupled non-linear ODEs;  
    2.5  $\mu\text{s}$  time step; 30 ms cycle
- Solved using:
  - Forward Euler method
  - 4<sup>th</sup>-order non-adaptive Runge-Kutta method
- Parallelized by creating one sub-grid for each processor

# Finite Element Modelling

**Epicardial Fibers – FEM Model**

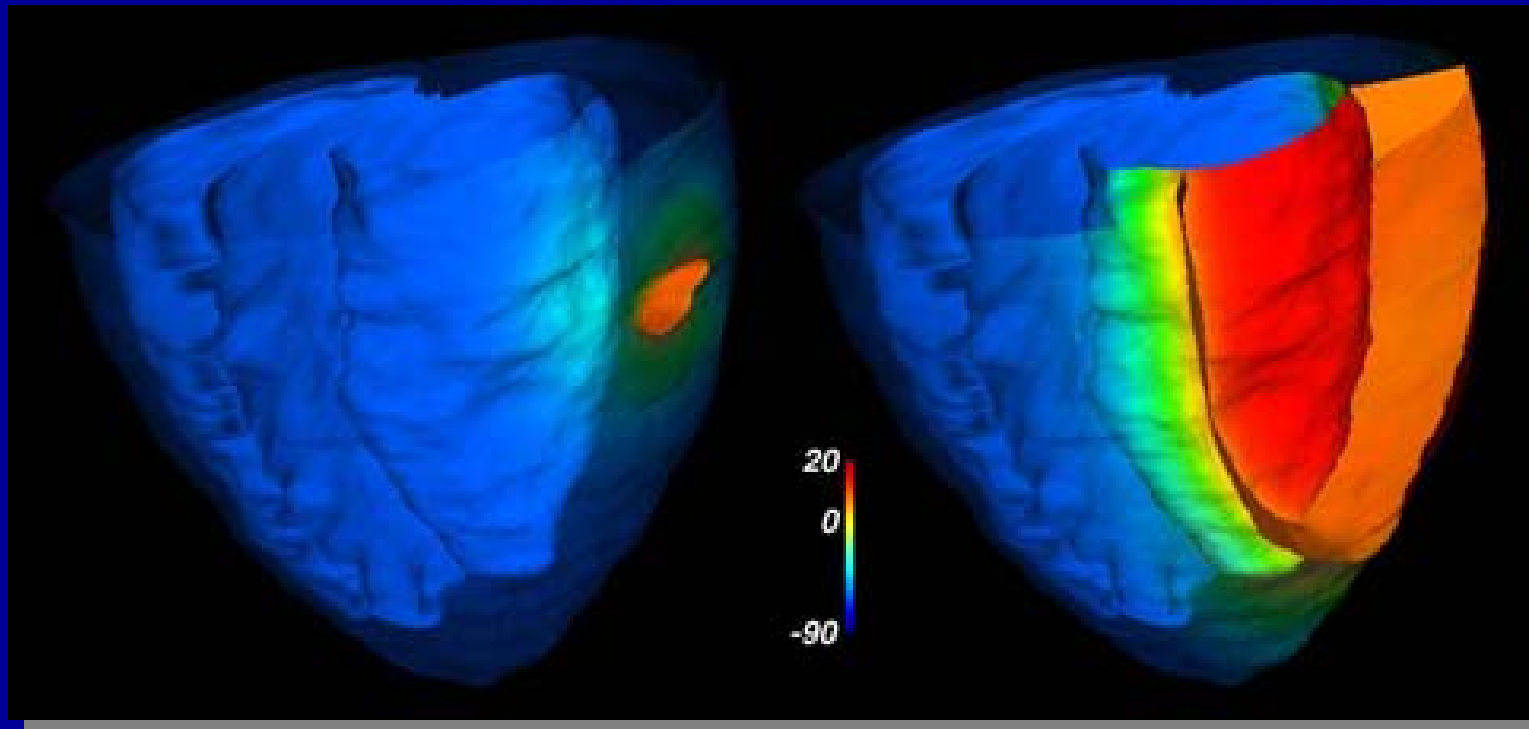


**Endocardial Fibers – FEM Model**



R. Winslow, Johns Hopkins Univ.

# Electrical Model



Geometry and fiber structure from DTMRI converted to a computational domain over which PDEs governing the spread of current in the myocardium were solved

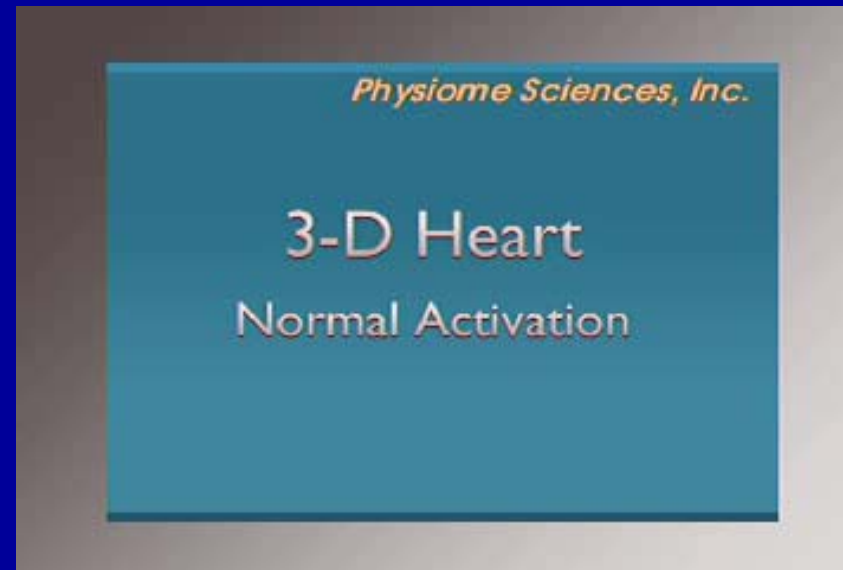
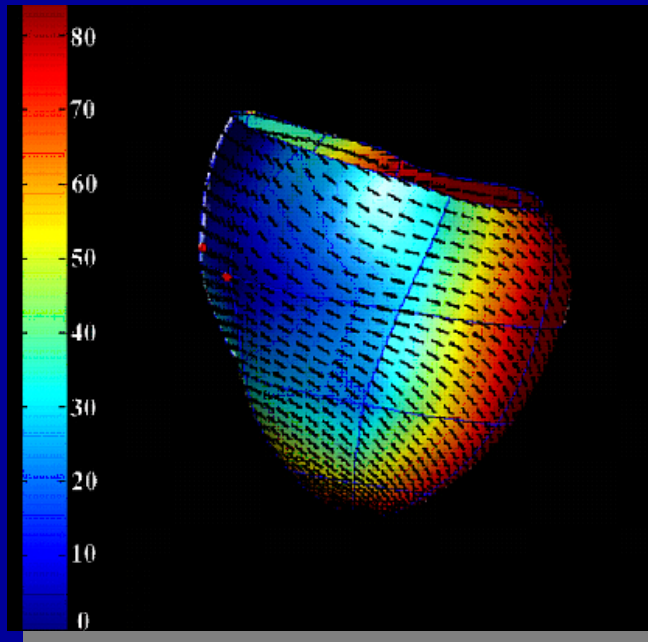
R. Winslow (Johns Hopkins)

## Reaction-Diffusion Equation

$$\frac{\partial v(\underline{x}, t)}{\partial t} = \frac{1}{C_m} \left[ \underbrace{-I_{ion}(v(\underline{x}, t)) - I_{app}(\underline{x}, t)}_{\text{From Ionic Models}} + \frac{1}{\beta} \left( \frac{\kappa}{\kappa + 1} \right) \underbrace{\nabla \cdot (M_i(\underline{x}) \nabla v(\underline{x}, t))}_{\text{From DTMRI}} \right], \dots \forall \underline{x} \in H$$

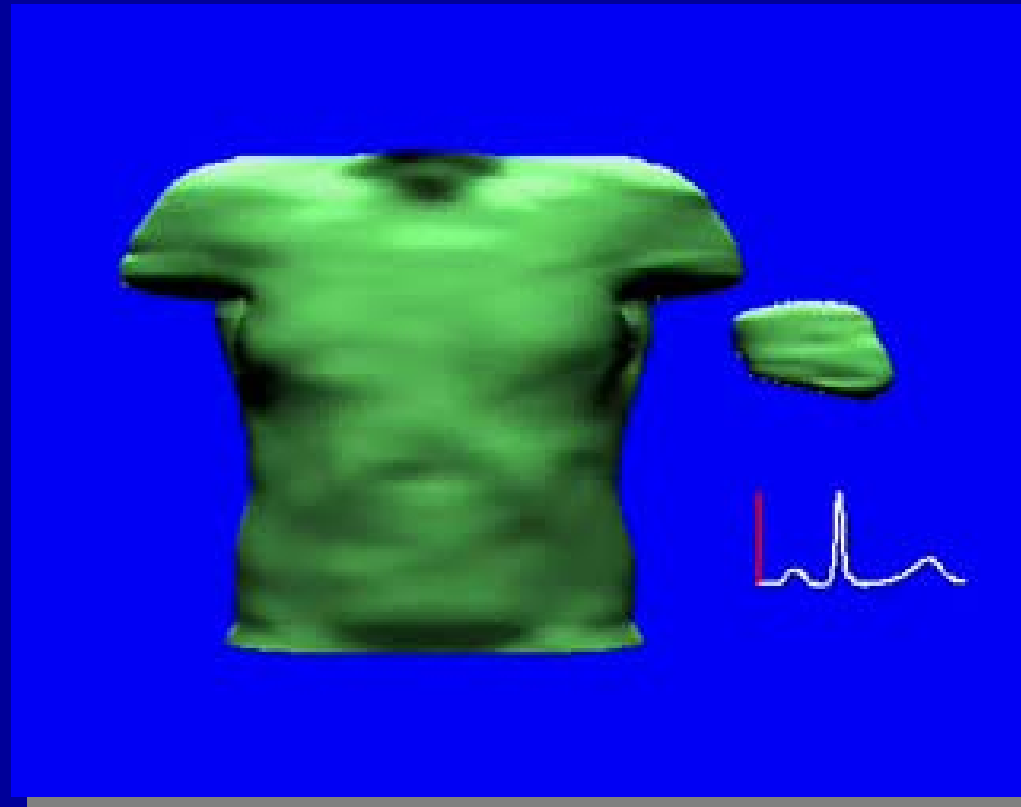
From Ionic Models

From DTMRI



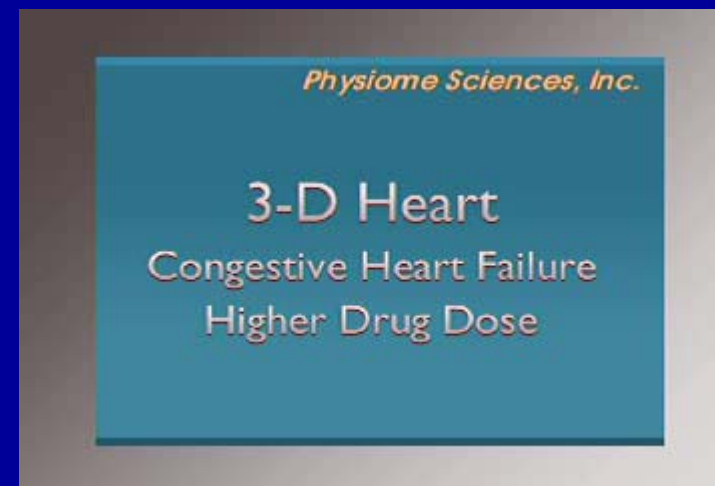
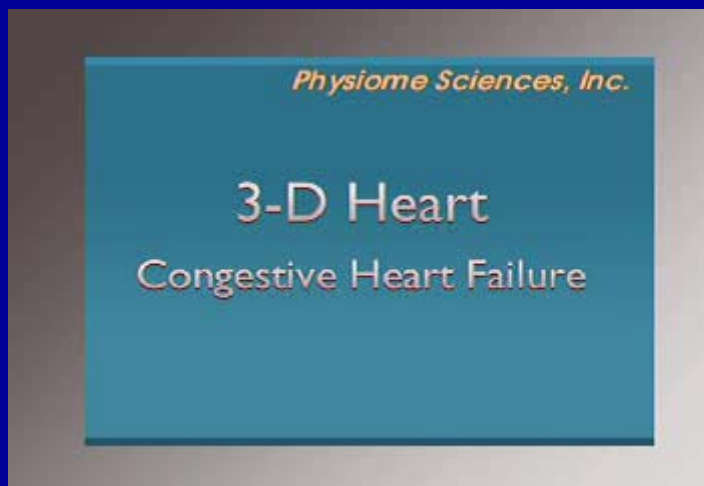
Courtesy of R. Winslow/Physiome Sciences

# Heart in torso model



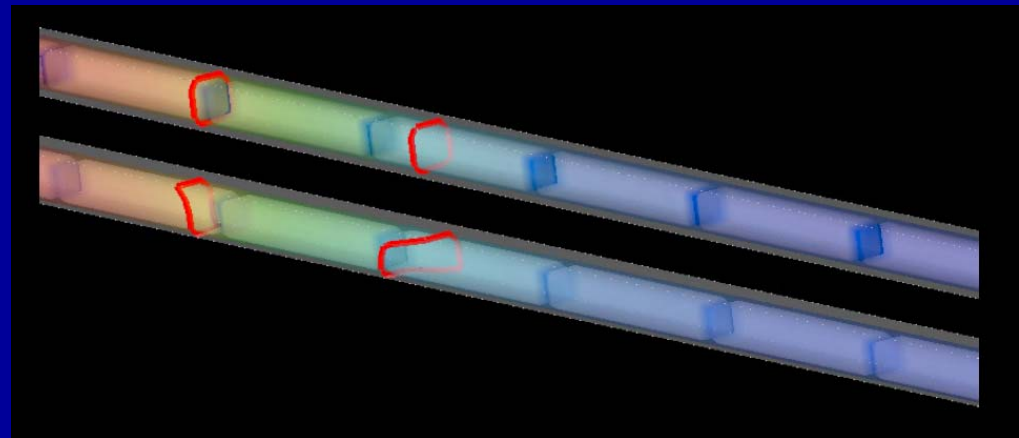
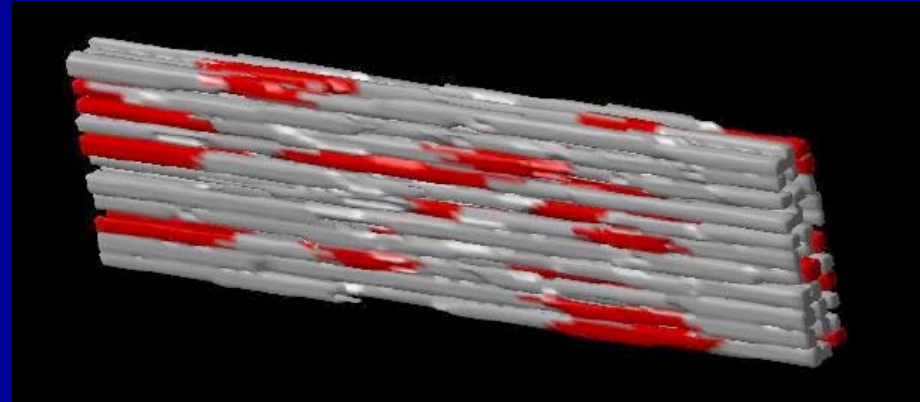
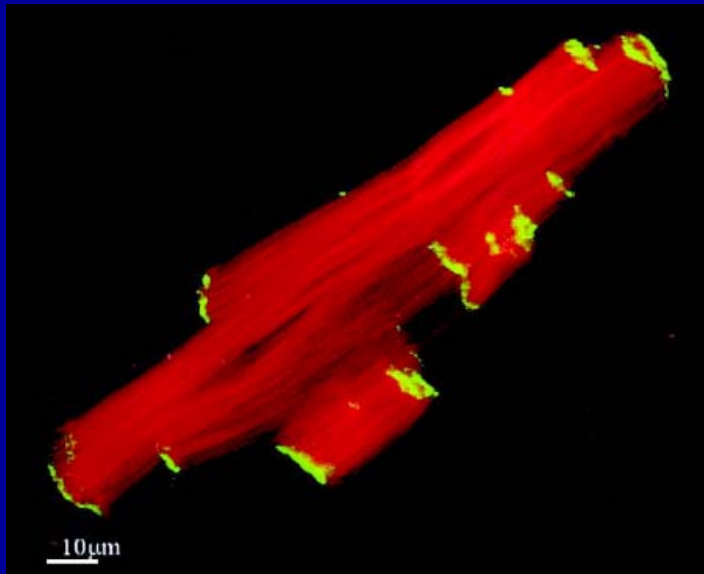
*R. Winslow et al.*

# Modelling Arrhythmias



R. Winslow, JHU/Physiome Sciences

# But reality is more complex...

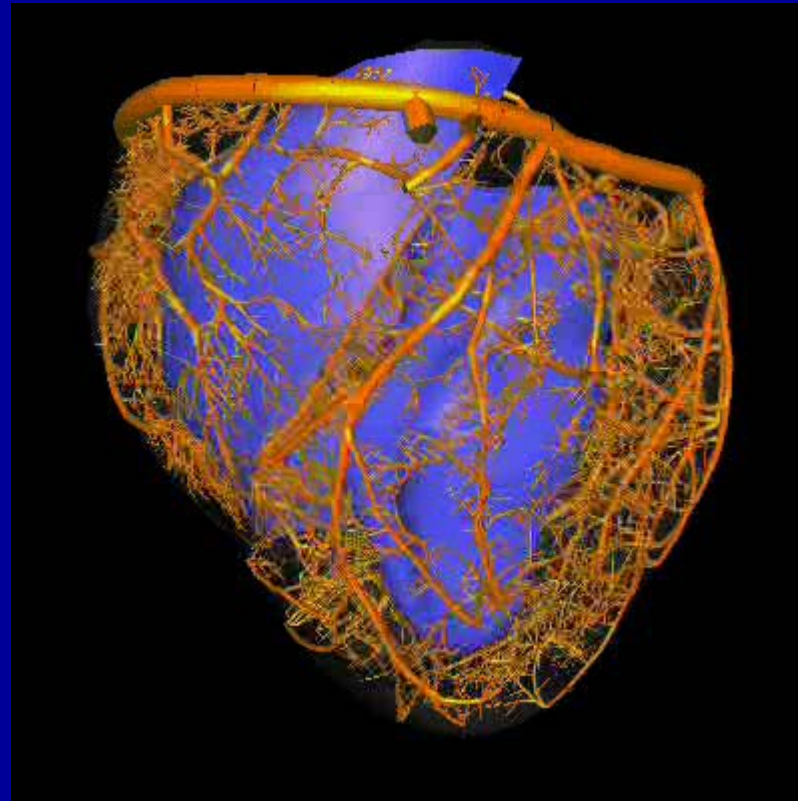


- Cell to Cell Coupling
- Cell Size
- Action potential Heterogeneity
- Fibrosis

*C.S. Henriquez, S.F. Roberts "3D Microstructural Modeling of Cardiac Tissue," in Proc. IUPS Satellite Workshop on Computational Physiology, San Diego, CA (2005)*

# Coronary Vasculature

- 1mm spatial resolution for larger vessels
- fractal model for smaller vessels



*Source: N. Smith and P. Hunter 1998*



# Coronary Vasculature: Simulating Ischaemia

Fig. 3. Flow calculations coupled to the deforming myocardium. The color coding represents transmural pressure acting on the coronary vessels from the myocardial stress (dark blue, zero pressure, red, peak pressure). The deformation states are (from left to right) zero pressure, end-diastole, early systole, and late systole (26).

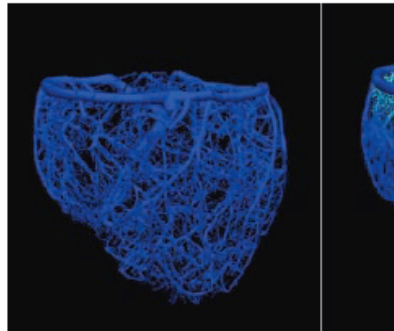
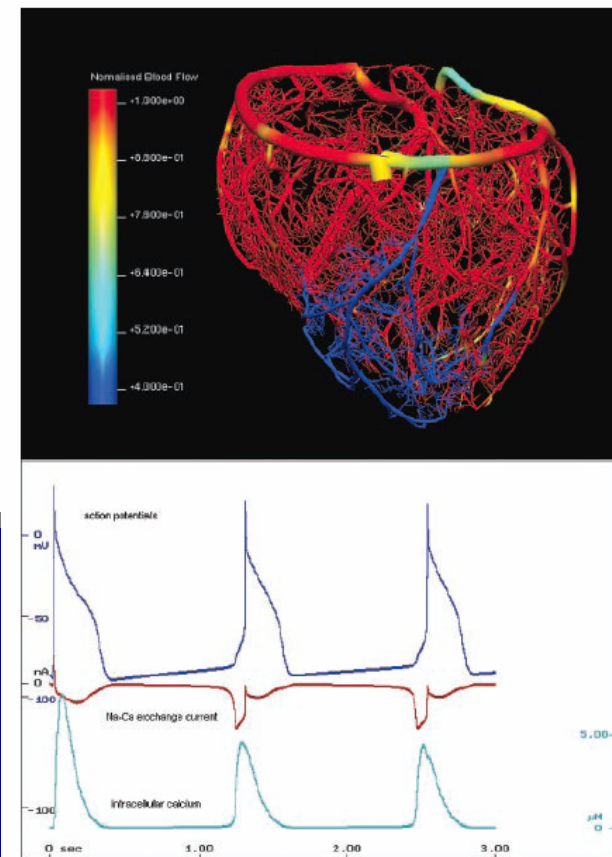
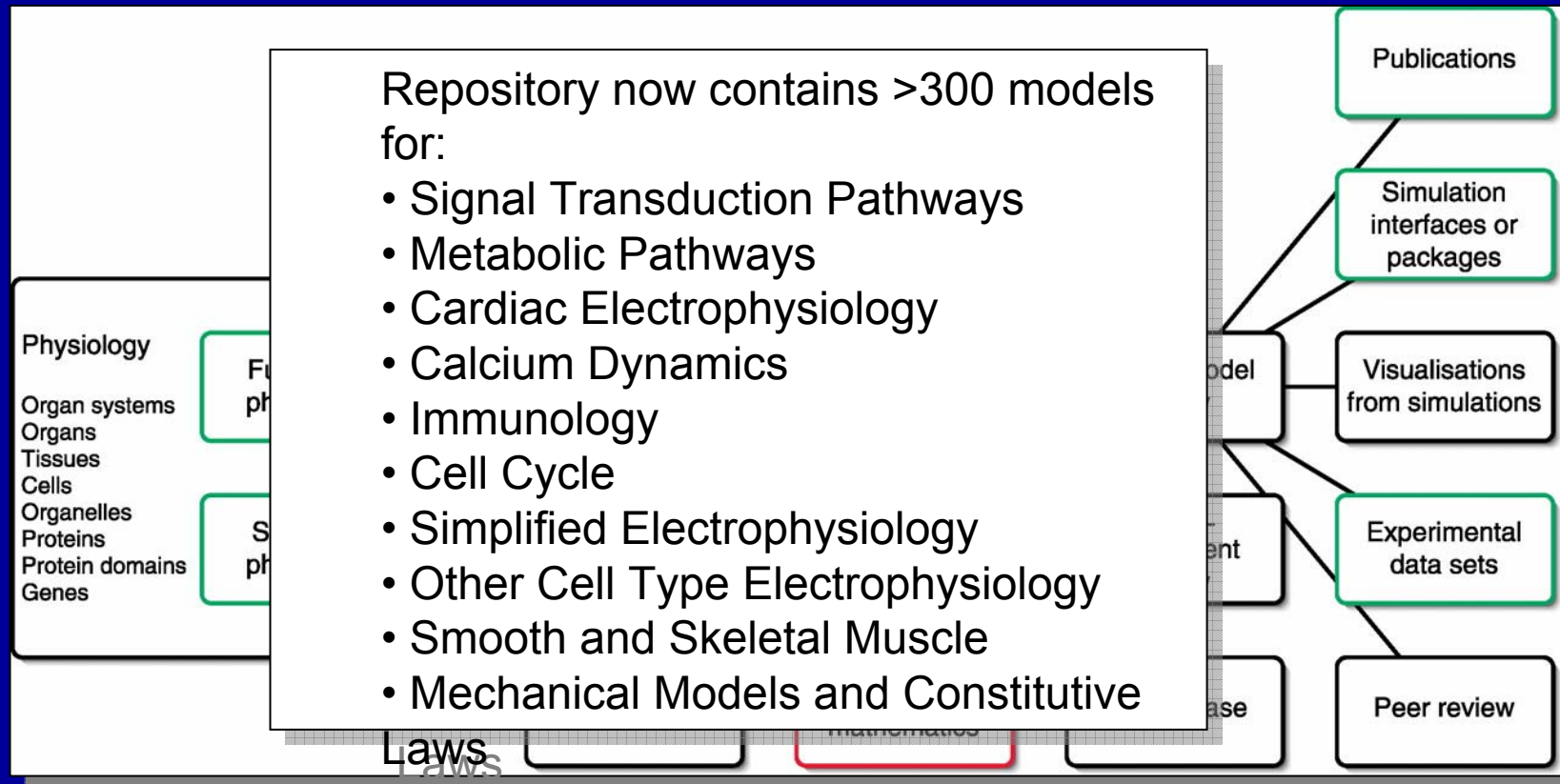


Fig. 4. Left, the coronary circulation model shown in Fig. 3 has been subjected to a constriction of one of the main branches leading to blocked blood flow in the regions colored blue. Right, simulation of ectopic beats in a Purkinje fiber model in conditions of calcium overload of the kind that occurs in ischemic tissue. Oscillatory calcium changes (bottom) induce inward sodium-calcium exchange current (middle) leading to initiation of action potentials (above).



Noble, D. "Modeling the Heart: from Genes to Cells to the Whole Organ,"  
*Science* 295: 1678 (2002)



# Some Implications

- Drug Discovery and “Wellness”



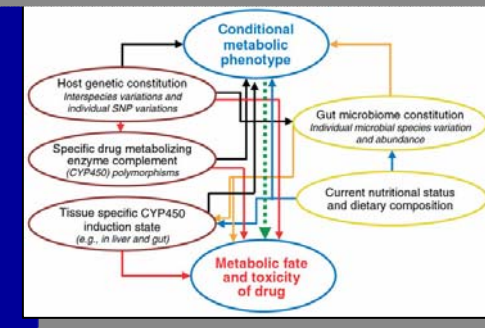
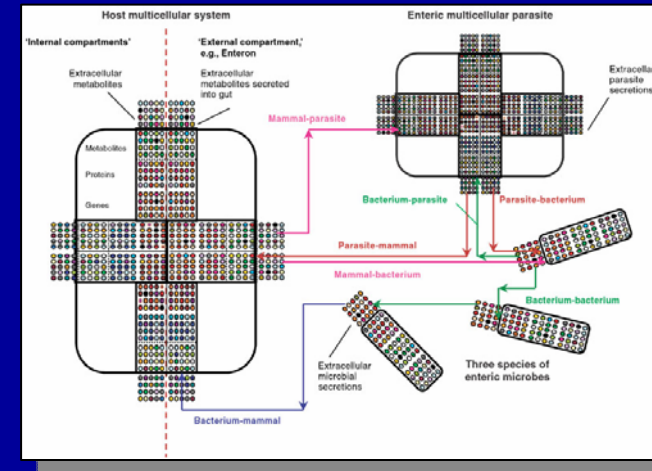
- Metagenomics: “super-organisms”

- Intestinal **microbiome**

- 1000+ species, 1Kg =  $10^{11}$  cells
    - Milk a “molecular cocktail”?

- Drug interactions with multiple proteins e.g. arrhythmia

- Look for small molecules which can alter behaviour of **entire networks**, not single proteins



Nicholson et al *Nature Biotech* 22:2268 (2004)

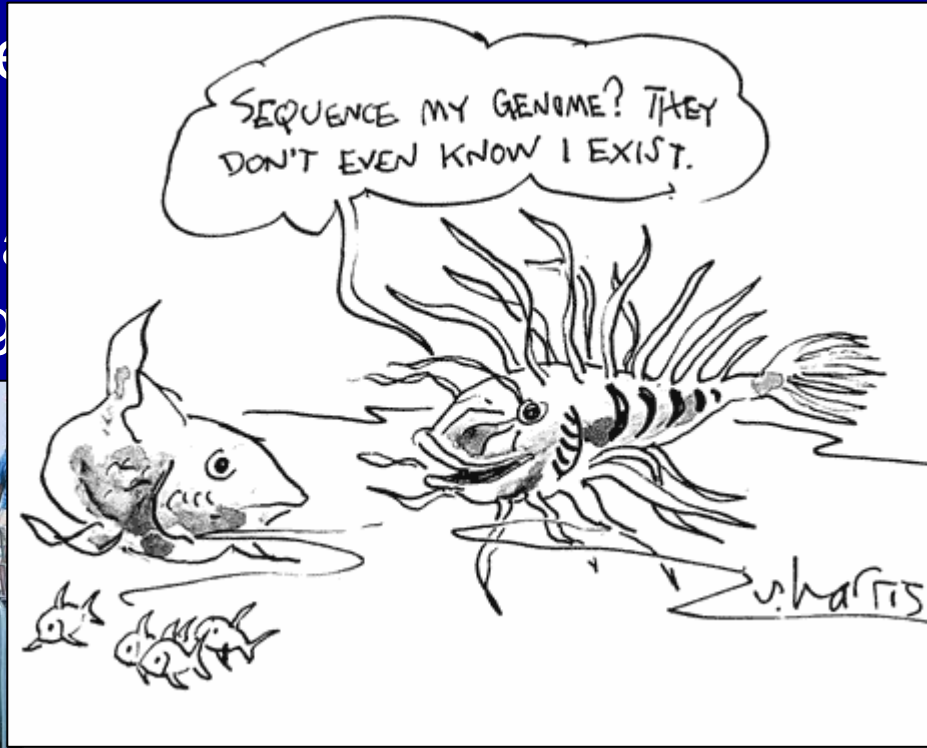
# Some Implications

- Understanding the planet

- Meta-genomics

- 6000

- 1 ocean  
new g



ents

crobes?

s, 1.2 million



# Not quite there yet...



Thank You!

Merci!

Danke schön!

Grazie!

БЛАГОДАРНОСТЬ!

由於

Diolch!