

S2I2 Institute for Translational Systems
Biology

30D
VIRTUAL CELL

Philip E. Bourne
UC San Diego
www.3dvcell.org



My Agenda

- Discuss the 3D Virtual Cell Project
- Provide some opinions on software and data sustainability through community engagement



My Perspective

- Built computing infrastructure
- Computational biologist but NOT a modeler
- 15 years with a community resource – PDB
- Establishing communities – PLOS, FORCE11, *DELSA*, *NIF*
- University administrator
- Numerous advisory boards



The Structural Biology of HIV

Viral Enzymes

Accessory Proteins

Structural Proteins

How Do DRUGS Work?

Examples from the PDB archive

Antibiotics & Antivirals

Anticancer Chemotherapy

Drug Metabolism

Design of Signaling Proteins

Lifestyle Drugs

Virus Structures

Viruses attack cells and force them to make more viruses, often killing the cell in the process. Some viruses are very simple, such as the poliovirus that caused a short outbreak of polio in the US. More complex viruses target the cell with a viral genome capable of encoding proteins to fight the cell's defenses.

Most viruses are much smaller than cells. The ones shown here are all shown at approximately 100,000 magnification, ranging from 100 nm to 100 micrometers in diameter (1 micrometer = 1,000 nanometers).

Designer liver virus

Scientists have designed a virus that can target liver cells and cause liver damage. This virus is a model for understanding how viruses can be used to study liver disease.

Respiratory syncytial virus

Respiratory syncytial virus (RSV) is a common respiratory virus that causes the common cold and bronchiolitis. It is a model for understanding how viruses can be used to study respiratory disease.

Foot and mouth disease virus

Foot and mouth disease virus (FMDV) is a highly contagious virus that causes foot and mouth disease. It is a model for understanding how viruses can be used to study infectious disease.

Fallovirus

Fallovirus is a virus that causes falvovirus disease. It is a model for understanding how viruses can be used to study infectious disease.

Poliovirus

Poliovirus is a virus that causes polio. It is a model for understanding how viruses can be used to study infectious disease.

Hepatitis virus

Hepatitis virus is a virus that causes hepatitis. It is a model for understanding how viruses can be used to study infectious disease.

Complex virus

Complex virus is a virus that causes complex virus disease. It is a model for understanding how viruses can be used to study infectious disease.

Enveloped virus

Enveloped virus is a virus that causes enveloped virus disease. It is a model for understanding how viruses can be used to study infectious disease.

What Got Me Thinking

- At PDB40 Jane Richardson described the early hand drawing of proteins and the emergence of the icon ribbon diagram to aid conceptualization
- In subsequent years molecular graphics emerged to automate this process
- David Goodsell described how he determines cell contents by literature review and draws the contents
- Automating that conceptualization would seem a logical next step



Thinking on Software back in 2008..

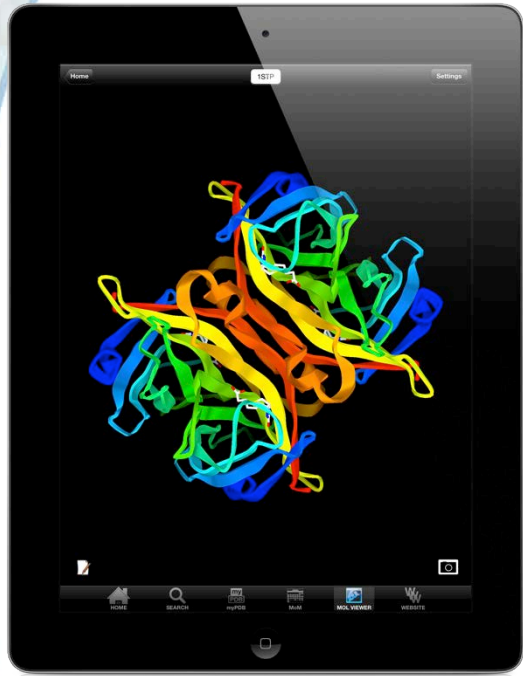
- Costs too much
- Is located in silos
- Does not foster reproducibility
- Is poorly maintained – is unsustainable
- Does not meet the needs of 21st century biology



Computational Biology Resources Lack Persistence and Usability. *PLOS Comp. Biol.* 2008 . 4(7): e1000136



What Got Me Thinking More



- Software development in science has improved thanks to open source, github etc. but for the most part remains arcane
- Software (and data) atrophy is a problem
- There is much we can learn from the app model
 - Consistent user interface – intuitive
 - Common calling interface
 - App store – ratings commentary etc.



ITSB/3DVC

Scientific **community and infrastructure** to accelerate interdisciplinary science through *in silico* modeling and simulation of the action of a living cell.



Community Driven

Information Hub

Scientific Collaborations

Interdisciplinary Science

Bridging Scientific Gaps

Model Development



3D Virtual Cell Project

Education and Training

Publications

Rewards and Incentives

Outreach

INSTITUTE CONCEPTUAL DESIGN



Some Impediments

- “Hubs” are a curiosity not mainstream
- Education is still very much a “what” rather than a “how”
- The metric of success is still the paper
 - Software and data are undervalued
 - Software and data scientists are undervalued
 - Improved modes of comprehension remain sparse



CONCEPTUALIZATION + analysis of a

3oD VIRTUAL CELL

VISIT WWW.3oDCELL.ORG

★ The IN SILICO MODELING and SIMULATION of a living cell at MULTIPLE LEVELS is **within our grasp.**



The GRAND CHALLENGE is the CURRENT STATE of TODAY'S CYBERINFRASTRUCTURE



- we need more coordinated software development, better visualization tools, more emphasis on people, and MORE.

the way forward is the **APP MODEL**



the APP MODEL is PROVIDED BY SCIENTISTS for a community of users - educators, scientists, students - ALL with a COLLECTIVE and SHARED WISDOM.

the END RESULT is the ACCURATE PREDICTION of cellular function ACCESSIBLE by ALL STAKEHOLDERS.



NEW MODES + DISSEMINATION
NEW REWARD SYSTEM
CHANGED SOCIOLOGY



ENERGY PRODUCTION (biofuels)
ACCELERATED DRUG DISCOVERY
ENVIRONMENTAL CLEAN-UP
ENHANCED FOOD PRODUCTION
...and MORE!

THE JOURNEY

PHASE 1

3DVC Conference

Smaller Group Meetings

Community Surveys



Community Website

Resource Catalog

Outreach

<http://www.apachenitro.com>



COMMUNITY INPUT



HOW DO WE
ESTABLISH A
VIBRANT
& OPEN
COLLABORATIVE
COMMUNITY
to SUPPORT "3D"
VIRTUAL CELL
SIMULATION?



CAN WE DO BETTER?



IT HAS TO BE
FUN

ENGAGE
"OUTSIDERS"

ANSWER
the
UNANSW
QUESTION

START WITH
A SIMPLE
POINT of
ENTRY

HERE



HOW DO YOU MEASURE?

INCENTIVES:

CROWD
SOURCE



"A LATERAL
POINT of
INFRASTRUCTURE
=> PEOPLE TO
ACCELERATE
DISCOVERY"

CREATIVE
COMMONS



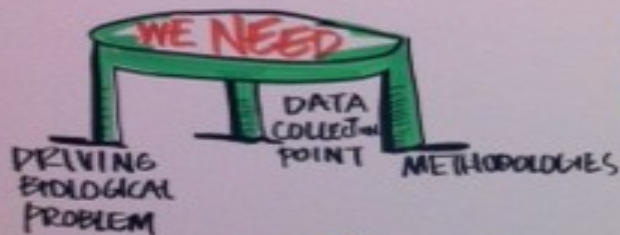
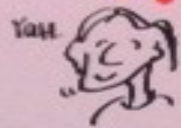
USABLE BY
OTHERS?
FOR THE 'GOOD' of
HUMANITY?



WE NEED

A WELL DEFINED
BIOLOGICAL
PROBLEM

BUILDING
TRUST IS
KEY!



COLLABORATION
CAN GET
BETTER
SCIENCE
FASTER!!

IF IT'S OPEN
and EXTENSIBLE,
CELL TYPE WILL
NOT MATTER.

NEW
KIND
P.I.

INTERDISCIPLINARY
SCIENTISTS

APP
STORE



Its All About Trust



Its All About Trust

- Trust is like compound interest
- Comes from listening
- Comes from engaging the community in every aspect of the process
- Comes from data consistency and level of annotation
- Comes from responsiveness
- Comes from the quality of the delivery service



Data Quality Begats Trust

- About 25% of our budget has been spent on data remediation
- Support for versioning hence the copy of record
- Our ontology/data model has been a critical component of our workflow and data accuracy
- Until recently the same data model was too complex to facilitate wide adoption by others that use our data



Modeling Examples



MARKUS COVERT STANFORD UNIVERSITY

A Computational Whole-Cell Model Predicts Genotype from Phenotype

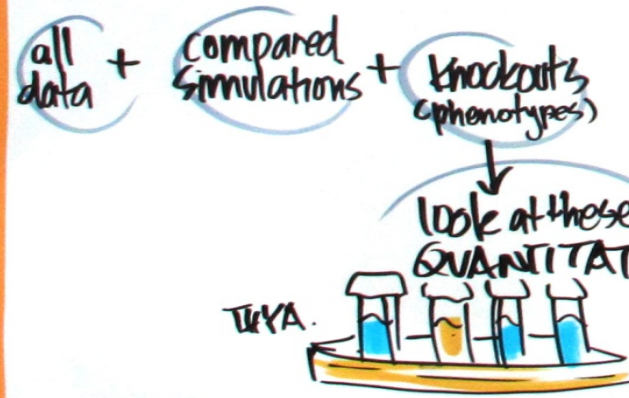
"NEVER TRUST a GOOD EXPERIMENT that doesn't have a GOOD THEORY"

sometimes you have to IGNORE DATA

Don't fall in LOVE with your model.

TEACH HOW TO HAVE the EXPERIMENT and the THEORY TRUST EACH OTHER

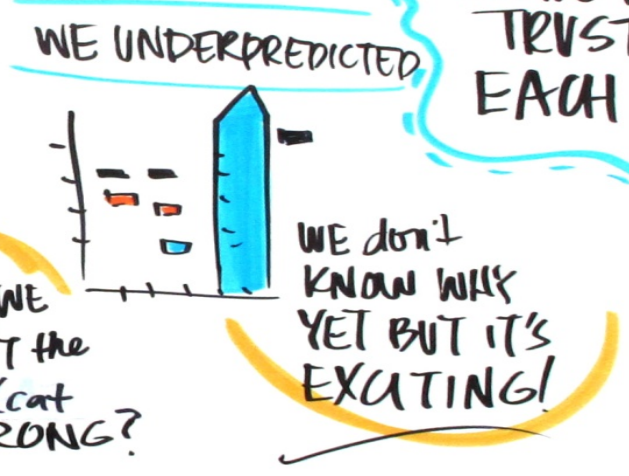
"WHAT DO WE DO WITH THIS THING?"



WHY DID WE GET PREDICTIONS WRONG?



each growth rate has 0.5 GB of data

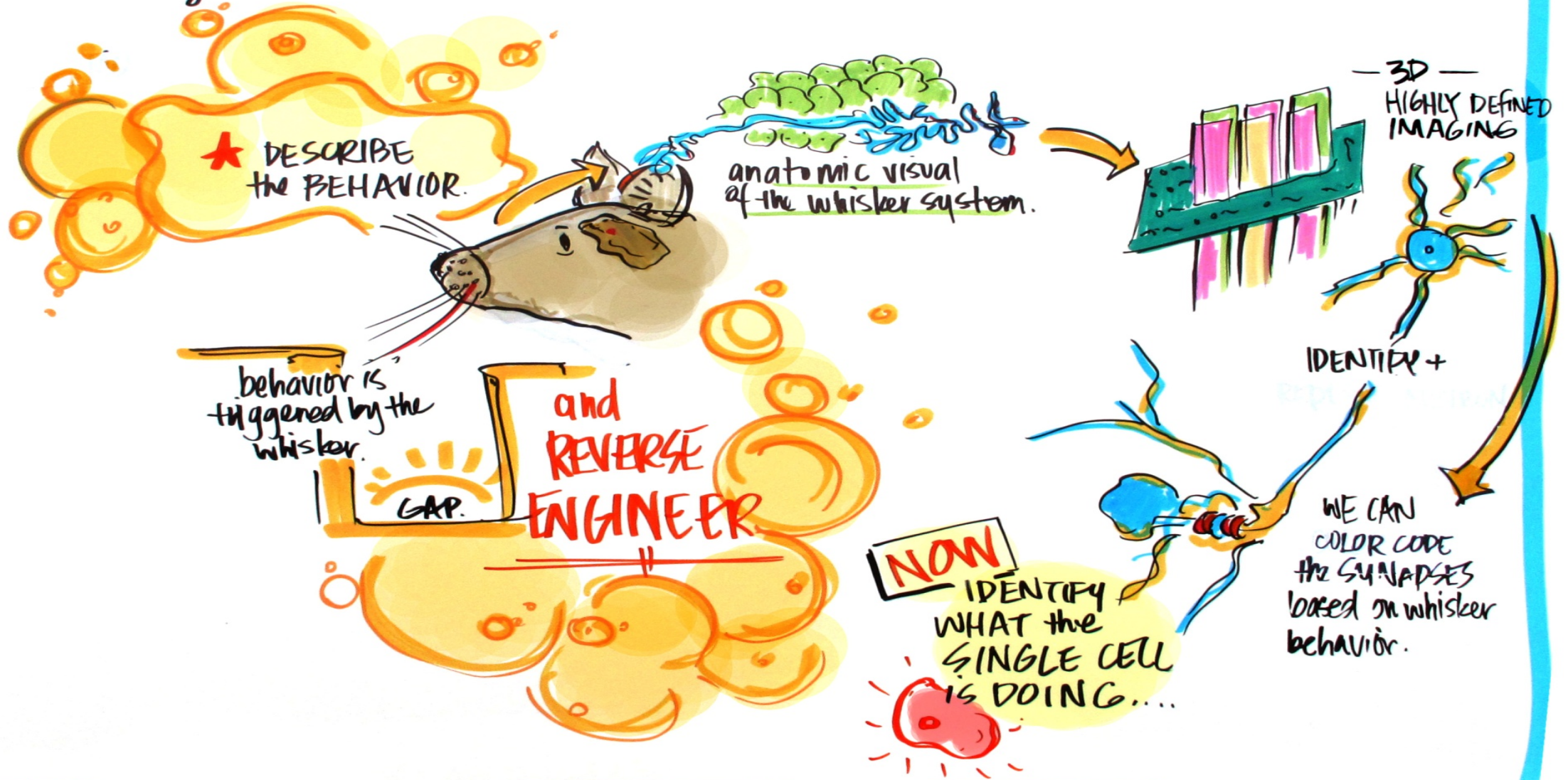


<http://www.3dvcell.org/conference-toward-3d-virtual-cell-videos>

MARCEL OBERLAENDER

MAX PLANCK INSTITUTE FOR
BIOLOGICAL CYBERNETICS

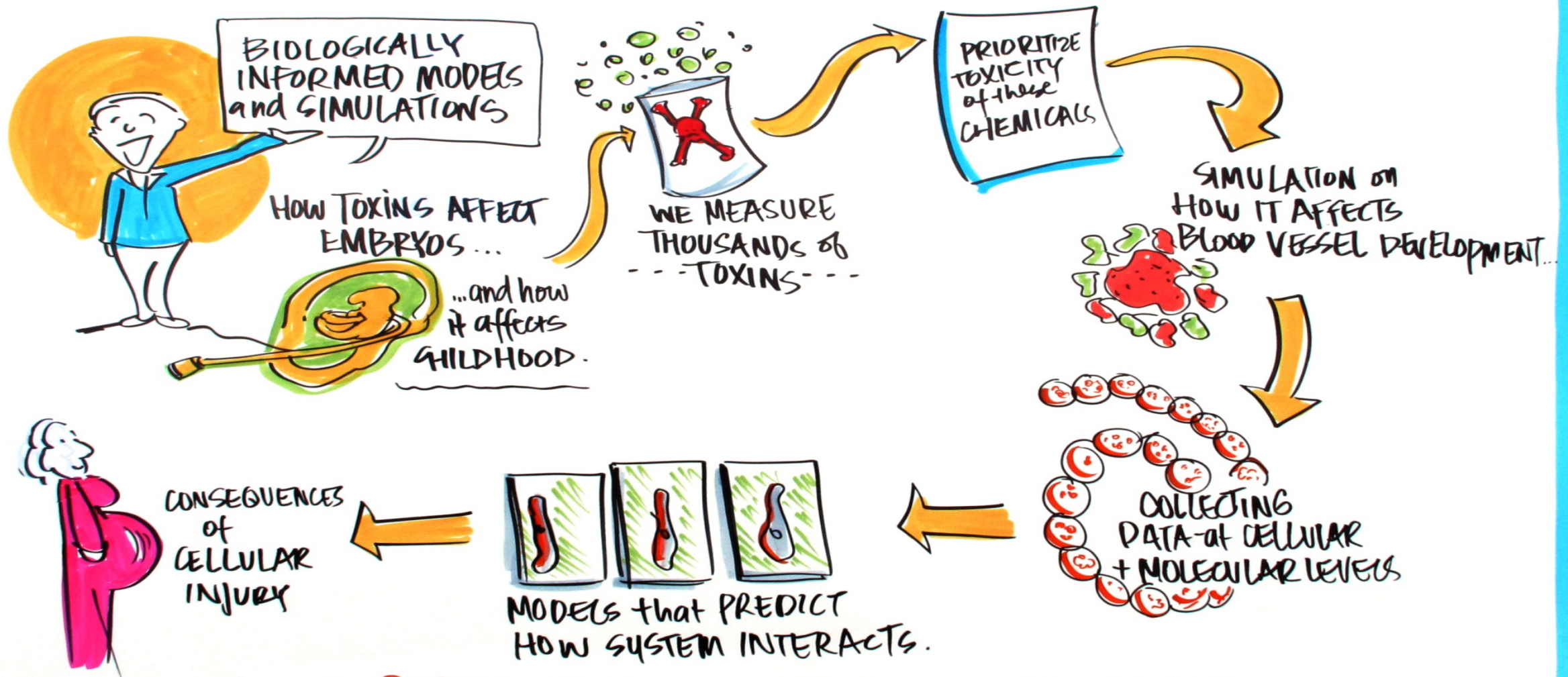
Simulation of Sensory-Evoked Signal Flow in 3D Reconstructions of Anatomically Realistic Network Models



TOM KNUDSEN

ENVIRONMENTAL
PROTECTION
AGENCY

Virtual Embryo: Computational Embryology + Predictive Toxicology



*** FIND SWEET SPOT TO ENABLE the BIOLOGY...**

ANDREW McCULLOCH

U.C.
San Diego

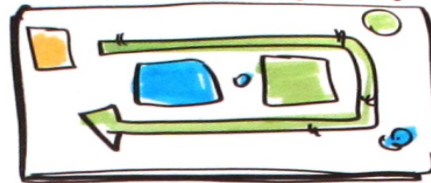
Multi-scale Modeling of the Cardiac Myocyte

WE NEED ALL 3...

1. BIOINFORMATICS
2. SYSTEMS BIOLOGY
3. STRUCTURAL INTEGRATION



• 3D MODELING
to help us develop
STRUCTURES
to work with.



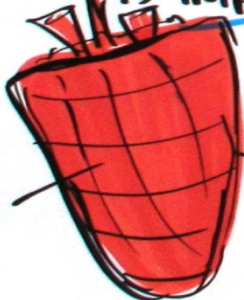
• and BRIDGING
MOLECULAR to
NETWORK MODELS.



• MOLECULAR DYNAMICS
to WHOLE ORGAN SYSTEMS

SOFTWARE
DEVELOPMENT
is IMPORTANT.

data
collector



shareable
+ reusable

GRADUATE
TRAINING
PROGRAM.

this is a
BIG OPPORTUNITY!

I want
this course.



Numerical Analysis
for Multi-Scale
Biology

Communities



HERBERT SAURO

UNIVERSITY of WASHINGTON

Building Sustainable > Vibrant Communities



SBML

SBOL:



4 YEARS Ago

FUNDED by MICROSOFT

SYNTHETIC BIOLOGY OPEN LANGUAGE



WHY DO THIS?



ITS IMPOSSIBLE TO RECREATE MODELS

WE LEARNED:



INVITE POST-DOCS & P.I.s



MEETINGS ARE DISCUSSION WORKSHOPS

GOVERNANCE



EDITORIAL BOARD:

WEBSITE CONFERENCES

A LOT of INDUSTRY SUPPORT



TODAY:

INTERNATIONAL COMMUNITY

CORE + WORKING GROUPS

WE PRODUCE

- SPECIFICATIONS
- USE CASES
- SOFTWARE (LIBRARIES)
- PUBLICATION
- OUTREACH

SUSTAINABLE

THANK to INVOLVING YOUNG SCIENTISTS

Its All About People The Global Personalities



Its NOT All About Institutions

- As far as I am aware no data standards body has directly influenced anything we have done in 15 years of running the PDB
- The structural biology community created a very successful data sharing plan long before funding bodies did



It is About Openness

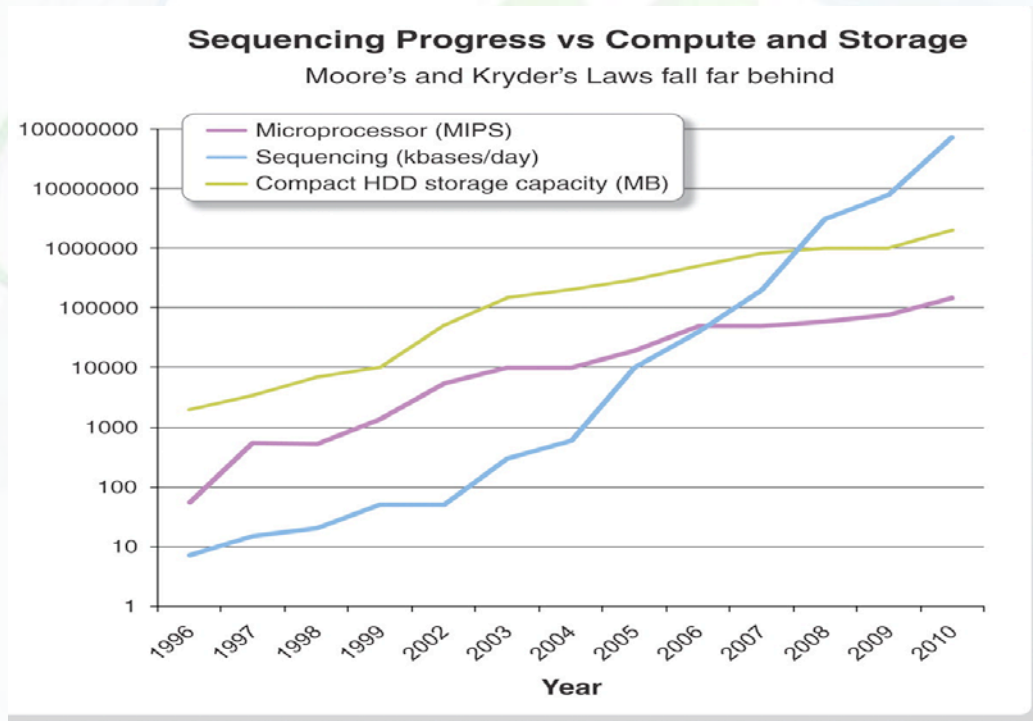
- There are no restrictions on the usage of the data beyond attribution
- The PDB runs exclusively on open source software
- We maintain and contribute to the Biojava repository
- We need to be transparent about data usage



So What Needs to Change re Data?



That All Data Are Created Equal Must End



- We need to understand how data are used
- Sustainability is not more money from the funding agencies its about business models
- Reductionism is not a dirty word – Reference Data!
- We need to do more with the long tail

On the Future of Genomic Data

Science 11 February 2011:
vol. 331 no. 6018 728-729



Institutions That Generate Data Must Play a Greater Role

- We need institutional data sharing plans
- We need data scientists to be better recognized by institutions – its not all about papers – this implies new metrics



POTENTIAL PHASE 2

Model Repository

Standards and Best Practices

Data Accessibility

Ontologies



Virtual Cell Animations App Store

Software Development

Shared Software

Data Analysis

Science App (sAPP+) Models

MODEL REPOSITORY & sAPP+





LOGIN Search



- HOME
- ABOUT
- CATALOG
- NEWSROOM
- MULTIMEDIA
- EVENTS
- CONTACT

3DVC CATALOG

The 3DVC Catalog has been built to enable scientists to add, edit, find and rate existing resources within the cell biology in silico modeling community.

We encourage viewing, editing and adding to the catalog. Thank you.

SEARCH

Search all or choose a category ▾

+ ADD TO CATALOG

+ OR SEARCH BY KEYWORD: [\(click here to view tags\)](#)

This project has been funded by a grant (#1216893) from the National Science Foundation

UC San Diego has been awarded this grant for planning activities leading to a design for an NSF S2I2 Institute for Translational Systems Biology (ITSB) that emphasizes biological 3-dimensional (3D) structures from molecule to cell and addresses the sustainability challenge.



UC San Diego

SITE BY Calit2



POTENTIAL PHASE 3

Education

Training

New Reward System

New Incentive Program



<http://swissnexsanfrancisco.org>

Sustainability

Scholarly Communication

Collaborative Science

Open Access



CHANGING SCIENCE



OUTCOMES

Accurate Prediction of Cellular Function

New Modes of Dissemination

Changed Sociology

Accelerated Drug Discovery



Diverse Discipline Cross Training

Public/Private Partnerships

Open Access

?



TOWARD THE FUTURE



SPONSORED BY...



SUPPORTED BY...



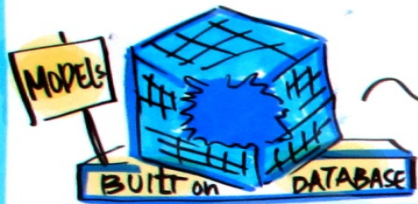
Back Pocket Slides



LESLIE LOEW

UNIVERSITY of CONNECTICUT
HEALTH CENTER

— The Virtual Cell Project — funded since 1998 by NIH



563 ONLINE MODELS...

$$\frac{\partial C_i}{\partial t} = -\text{div} f_i + R_i$$



Types of problems WE CAN SOLVE!



physiology

applications

POWERFUL because it allows you to create physiological models.

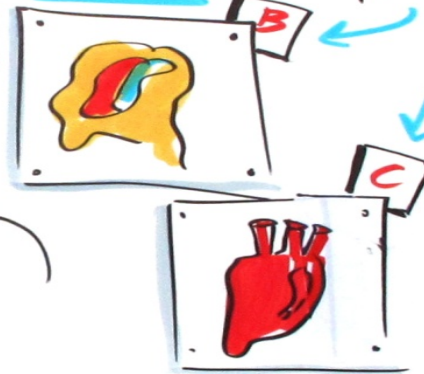
database

create pathways save access models

OUR

CHALLENGES

WE CAN RUN DIFFERENT MODELS FROM THE SAME PHYSIOLOGY



WE HAVE a LONG WAY TO GO....

need experimental data to constrain models, drive models, and validate models.

need analysis tools

need software + data that is ACCESSIBLE

→ and develop at the same time.