3D Virtual Cell meeting

Agenda is here: http://itsb.ucsd.edu/conference-toward-3d-virtual-cell-agenda

- Day 1: Thursday, 13 December.
  - Session 1: WHOLE CELL MODELING
    - Ramesh Rao
    - Phil Bourne
    - Andrew McCulloch
    - Tom Knudsen
    - Wah Chiu
    - Steven Peltier
    - Marcel Oberlaender
    - James Glazier
    - Panel discussion
  - Session 2
    - Leslie Loew
    - Bernhard Palsson
    - Martin Meier-Schellersheim
    - Peter Karp
    - Blake Borgeson
    - Daniel Russel
    - Panel session
  - Session 3
    - Eugene Kolker
    - Jessie Dixon
    - Arthur Olson
    - David Goodsell
    - Ross Whitaker
    - Chandrajit Bajaj
    - Panel
  - Open Space Sessions
    - Purple 1 - If you can't find it on google...
    - Blue 2 - Literate Models
    - Green 3
    - Red 3
- Day 2: Friday, 14 December
  - Session 1A: Whole Cell Modeling track
    - Peter Hunter

- Terry Sejnowski
- Maciej Swat
- Markus Covert
- Graham Johnson
- Natarajan Kannan
- Panel

Session 1B: ANALYSIS
- James Sluka
- Herbert Sauro
- Yolanda Gil
- Tim Clark
- Nigam Shah
- Nancy Wilkins-Diehr
- Panel

Session 2A: STRUCTURE BASED MODELING
- Susan Taylor
- Pieter Dorrestein
- Wei Wang
- Don Blumenthal
- Theodore Alexandrov
- Panel

Session 2B: EDUCATION/DISSEMINATION
- Philip Compeau
- Erich Huang
- Maryann Martone
- Anita de Waard
- Lucila Ohno-Machado
- Panel:

Workgroups
- Plenary Session & Introduction to Workgroups
- Group 1 (green)
- Group 2 (yellow)

FInal topics
- Infrastructure
- Sociology

Closing Remarks

Day 1: Thursday, 13 December.

Session 1: WHOLE CELL MODELING

Ramesh Rao

http://www.calit2.net/

Phil Bourne

http://www.sdsc.edu/pb/

- cartoon
- Science is very inefficient
- Reward system and contributions are not well correlated
- Resources are siloed, science is not
- Social networking has yet to impact science in a big way
- RCSB PDB paper has been cited 14k times, no one has read it
    - Community of experimentalists, but it could be more
    - The idea of a science App Store
- How can we do better and establish a vibrant community?

Approach

- 10 min per speaker
- Cartoonists are drawing images of what is being said. Images are being tweeted
- Web cast
- Panel Discussion
    - Open Space - discuss solutions
    - Drill down and lay groundwork for a grant submission - working groups
- End product: Crowd sourced grant application
- 10 min presentations:

Andrew McCulloch

http://cmrg.ucsd.edu/

- expect to hear about the potential of integration
    - Information integration
    - Functional Integration
    - Structural Integration (aka multi scale modelling)
    - we need all 3 of these approaches
    - No reason to stop at level of cell
    - Software engineering is important
    - We need next generation of interdisciplinary training
    -
- Problem: Cardiac myocyte model
    - Use Markov Models to bridge molecular to network models
        - uses Molecular Dynamics

        - use models to investigate defects

Tom Knudsen

- cartoon
- Virtual Embryo http://www.epa.gov/ncct/v-Embryo/
- Predictive Toxicology
- we know a lot about pathways, but less about cellular network function
- predict how system responds

Wah Chiu

http://ncmi.bcm.edu/ncmi

- cartoon
- Cryo-Electron Tomography of Cells
- From small complexes to whole cells
- Annotated views of tomograms
- challenge: how to present dynamics of cells

Steven Peltier

http://ncmir.ucsd.edu/

- cartoon
- multiple terrabytes per week
- Image segmentation - via algorithms, curate results of computation via crowd sourcing
- Cell browser - webgl
- @bgood **#crowdsourcing** strategy for **#mturk** in image segmentation at **#vcell** pic.twitter.com/4Y5bdR6Z

Marcel Oberlaender

http://www.kyb.tuebingen.mpg.de/

- cartoon
- from subcellular to behavioral levels
- Whisker movements are part of decision making
- Look at 1/2 million of neurons
- simulate synapses

James Glazier

http://www.compucell3d.org/

- cartoon

- developmental biology: how to go from cell to organism without plan
- you need to understand how cells interact to understand how the disease works
- Models as searchable data
- Experimental data as model
- Consistent treatment of experimental and simulation data

Panel discussion

- cartoon
- What else do we need?
    - need better tools and infrastructure
    - enable community to have an appstore / itunes
    - we are a long way from having an understanding of the problems
    - in app store - interoperability is limited, well defined set of questions
        - here we have not way to describe cell behaviour, there is no language
        - how to describe subcellular components?
        - We can build interoperability once we have these components
    - successful multi scale integrations
        - genomics ... but that is a limited dataset
        - geographical information systems, perhaps biggest success story
    - we need the middle layer and community agreement
    - how do we change science behaviour so people care about reproducibility?
    - Workflows - people don't agree what they mean
        - workflow developer agree on some services that are needed
        - -> this provides the base layer for an app store
    - apps that use well defined APIs will be well used and cited
    - problem: people build monolithic non-reusable code
    - some groups think that reproducibility is not an issue
    - data: we should not ask what data can do for us, but what we can do for data
    - many people do lots of measurements, but some crucial ones are missing, that are needed to build a model
    - nobody will use tools, unless they provide immediate benefit
    - lower the level of participation

Session 2

Leslie Loew

- cartoon
-
- The virtual cell project VCell http://bit.ly/XXVLWG Virtual Cell Modeling & Analysis Software
    - 4k users ran models
    - 350 models available

- model took 10 days to simulate
- things to consider: morphology, mechanics, intercellular trafficking
    - experimental details to constraint models

Bernhard Palsson

http://systemsbiology.ucsd.edu/

- genome scale models
- bacterial models
- Roger chang's structure to pathway mappings
- Genome scale models will play an important role moving forward

Martin Meier-Schellersheim

http://www.niaid.nih.gov/labsandresources/labs/aboutlabs/lsb/computationalbiology/Pages/default.aspx

- challenges
    - make experimental data available for quantitative modeling
        - data acquisition, measurement technologies, standardization, data encoding
- most data lack temporal and spatial resolution
- develop tools for efficient modeling
- Need better validation/statistics of model-data fit
- How to build a community:
    - need to find a driving biological problem with well specified goals
- problem areas on various scales that can be tackled by specific groups

Peter Karp

http://www.ai.sri.com/pkarp/

- Programs are "works of literature" --Don Knuth
- solve human task, not just writing code
- we don't know how to document and explain programs
- the cost of a company understanding its own company... huge
- Need to write clearly about programs, and design APIs that are easy to understand
- Now: literate modelling
- readily understandable models
- link to other resources (genomes, literature, etc)
- graphical interfaces
- Fuse model organism databases with system biology models, EcoCyc has done this.
- Common curation efforts

Blake Borgeson

- Protein complexes in animals are mostly undiscovered
- How can we fix that?
- map complexes at a time
- using mass spec

Daniel Russel

http://www.integrativemodeling.org/

- Assembly Modeling
- Integrative structure modeling
- modeling of 26s proteasome
- available software: IMP (integrative modeling platform)
- big challenge: visualisation

Panel session

- Knuth: need both a human, but also a machine readable information of the modeled biology
- dissemination of models
- it is students who use models
- data sharing plans are never enforced
- lack of collaboration, lack of enforcement
- how to find what is there
- Search engines can't find what's inside databases

Session 3

Eugene Kolker

- Delsa Data-Enabled Life Science Alliance http://delsaglobal.org/
- 11% reproducibility rate (amgen) and 85% research waste (Chalmers)

Jessie Dixon

(Bing Ren lab http://bioinformatics-renlab.ucsd.edu/rentrac/)

- Use High-C to look at genome-wide chromatin structure
- strong local chromatin interaction
- Topological domains (megabase syntenic blocks) are conserved between species & between experimental conditions

- Q: How reproducible is high-C between samples?
- Topological domains often share regulatory mechanism (aka TAD topologically associated domains)
- Using high-C data, find best-fit 3D model
    - Q: but data is from many cells, and could potentially be in different conformations? How much of the data is explained by 3D model/howmuch error is there in the structure
    - See some properties that correlate with 3D model, eg membrane-binding proteins localize to one size of the helical ch2 model

Arthur Olson

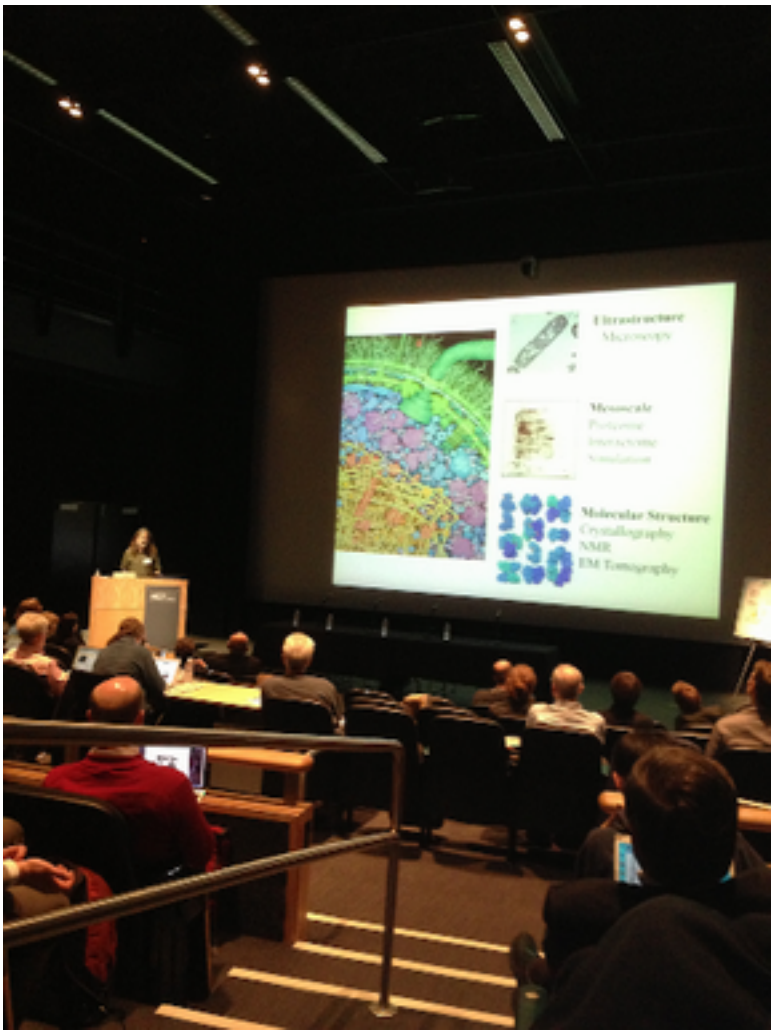
http://mgl.scripps.edu/

- should we make bigger and bigger communities to solve larger and larger problems?
- There is a continuity, everyting acts together
- Physical models promote many other types of interactions (perceptual, social, ...)
    - Virtualization loses some power of understanding
- Physical models not sufficient
    - Augmented reality
        - eg physically docking models with overlaid electrostatic information

- Procedurally produce Goodsell-style dense 3D volumes
    - nice graphics of cell complexity
    - benchmark for segmentation algorithms

David Goodsell

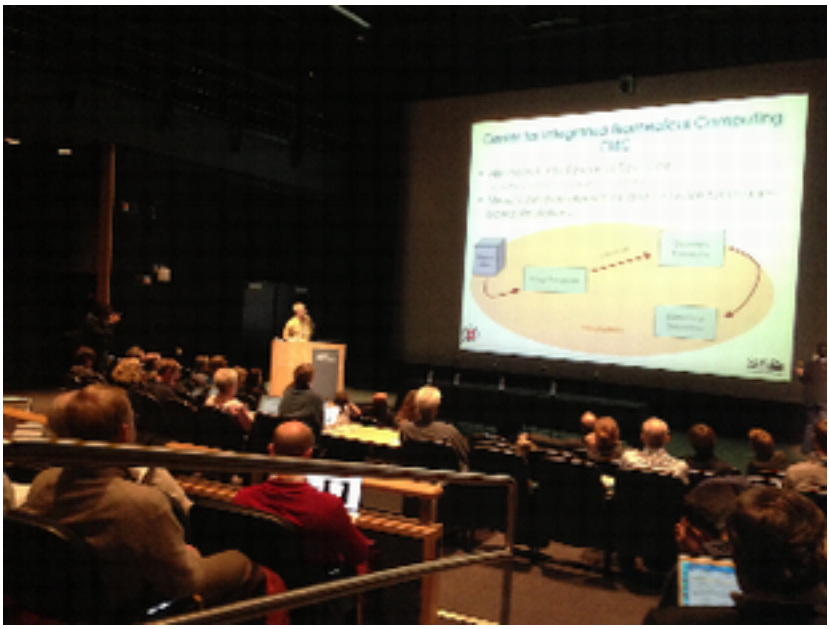
http://mgl.scripps.edu/people/goodsell



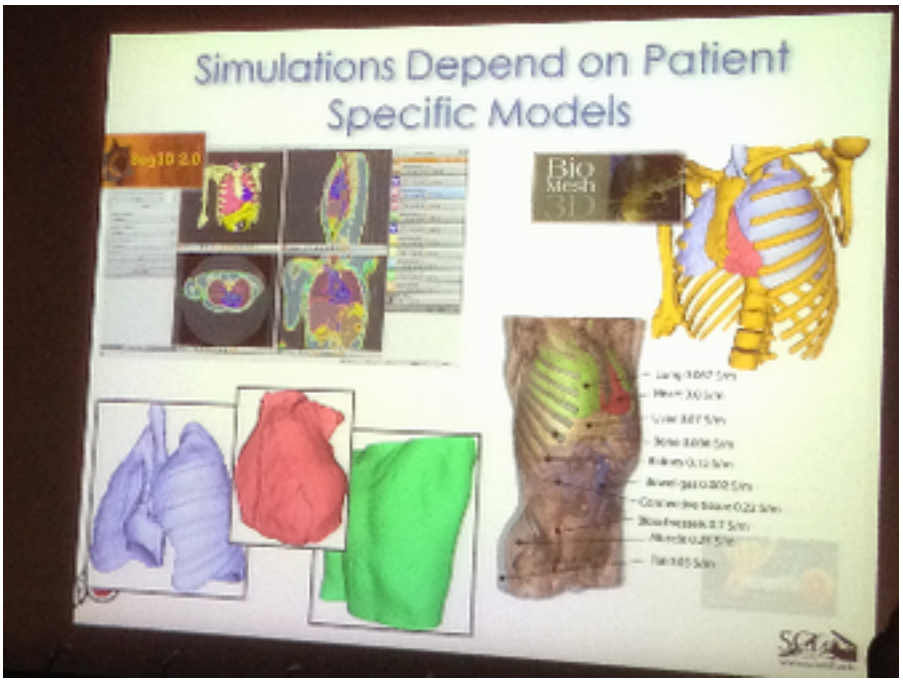
- What are the shapes to be visualized?
- For years has been creating images by hand (paintings) of the structural features of cells
- Enormous infrastructure required for conceptually simple actions
    - Eg synapse function: function is neurotransmitter release, receptor binding, and reabsorption, but also relies on structural proteins, membranes, signal transmission, etc etc etc
- What are concentrations? (some of the data is from the 1980s)
- What are the interactions? Find diagrams that show what is bound to what
- Publication illustrations are helpful for organization, but are not drawn by structural biologists and can be not quite what you need
- Articles he has published in BAMBED may be useful to identify items for which data is missing at this point.
- Use aesthetics to enhance understandability (color by function, molecular orientations)
    - Visually informative may conflict with biological accuracy

Ross Whitaker

http://www.sci.utah.edu/cibc.html



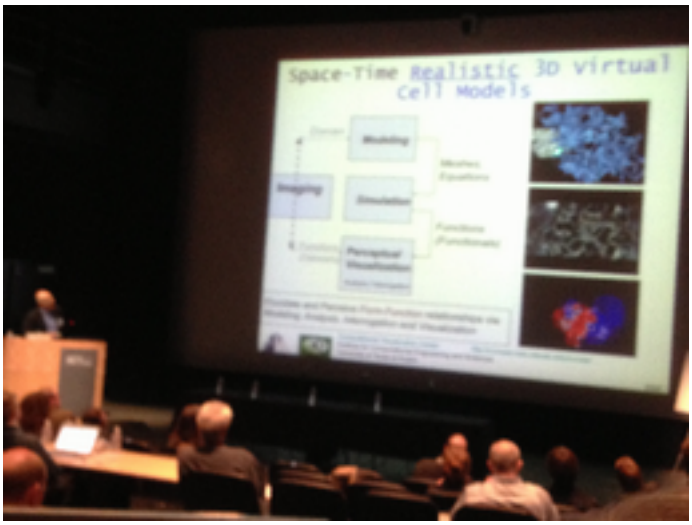
- Algorithms
- Software
- Applications
- Cardiac Defibrillators, off the shelf hardware not suited
- Do a CT, partition CT into tissue types, body fitting
- =>Patient specific models
    - Seg3D for patient model
    - Skirun for electrocardio simulations



- Biomechanics - Hip pathoanatomy
    - FEBio software
- validate simulations in cadavers
- Conclusion: many pieces of software need to interoperate

Chandrajit Bajaj

http://cvcweb.ices.utexas.edu/cvcwp/?page_id=81



- Try to model neurons, mapping function to individual neurons
- scale requires large compute
- solution: use multi scale models and try to solve from reduced model at different scale

- “In texas we do things big”
- "we built the largest ipad" -about multi-screen touch-sensitive wall
- Projecting MD out to 2020, predicts that large complexes still only tractable at very coarse resolution
    - Solution may be multiscale methods to reduce complexity

Panel

- EK: What are the new technologies that we will use to address the challenges?
- CB: For multiscale models, need to determine when high resolution (and large computation) is needed, and when the coarse model is good enough
- Attempt to establish distinction of terms used in this discussion:
    - Model: in-silico representation of what has been observed in an experiment
    - Simulation: prediction, and what is done to the model
- Q: Need visualizations for 3D objects (cells, molecules, organs, etc), but don't we also need better visualizations for abstract graphs, figures, models?
    - Comes down to analysis tools, both automatic and manual
- “Communication between people is hard, communication between models is even more difficult”
- How do we deal with variability? Between individual cells, tissues, timepoints, etc
- “Models come out of databases is a limited set of models”
- "Modeling is a creative exercise"
- “The model is not the thing it is modeling. It is an abstract view”
- a model is a tool, not an end for itself
- Understanding is the purpose, however we get there. We can take different routes

Open Space Sessions

(A professional facilitator, Eric, to work on how to really build community. Breaks out into small groups to discuss.)

Objective: How do we establish a vibrant, open and **collaborative** community to support whole cell simulation?

Think of topics

- Require *passion* for that topic
- Require *responsibility* for moderating the idea
- Don't require expert

Time	Blue	Orange	Green	Purple	Red	Yellow
25 min	Education & awareness	Competition & Principles of Community	Reusable Software	If you can't find it on google it doesn't exist	infrastructure, standards, semantics	Data & Models Journal
25min						
25 min	Openness & Incentives	Parallel processing & Infrastructure				

{This is Mary Mangan, btw} I wish I had a way to get input into those discussion groups (education + awareness). Alas. But I would love to suggest that they "embed" a science journalist in some of the projects, for a week at a time, to help them understand the personalities

and the work. And they can create the "wonder" and excitement type of outreach that can reach a lot of young people and science enthusiasts in the public.
{Spencer Bliven} I'll be transcribing one of them here. Not the best way to contribute, but better than nothing.
{Andreas Prlic} Should this be discussed as part of the discussion groups?
{Mary} Eh, if it ends up in the notes that's fine, maybe someone will see it and pick it up at some point. Strikes me now that the virtual conference folks like me need a way into this part...in the future.
{SB} Yeah. I think they were hoping the forums would get some activity along those lines.
{Mary} Oh, yeah, I saw the forums and tried to point people to them in my blog post--but they were all at zero then. Who wants to be first :) ? But a good idea for other confs would be to have a virtual attendee guide/monitor/something to encourage that. Edit: Grrr--I was going to put my idea in the education forum, but it refuses to accept the login I created yesterday.

Purple 1 - If you can't find it on google...

Scribe: Spencer Bliven

1) If you can't find it on google, it doesn't exist (Jim)
2) Google Cell Wiki available (Brett)

Jim Sluke: If you can't find it on google, you haven't met the NIH's standard of openness
Mark Ellisman: But our datasets are so large, it needs to be separately indexed
Need annotation with googlable engines
JS: Can google parts for a 68 chevy, since manufacturers add appropriate metadata. There are more parts in a automobile than in the human genome

How do we get ranking? Crosslinks between major providers.

Jeffrey Grethe: Provide central wiki/site for tools and databases. Need community annotated list of resources

Databases should be stored with flat-file structures to enable indexing.

Shouldn't worry about data relevance until we have easy access.

Goodrelations: microformat for product data

Klaus R

Apply friend-of-friend concept to genes, with crosslinks. Allows existing social networking tools to crawl genes and work on genes.

Brett Barbaro: Include concept of 'like' button to annotate genes & networks
Lilia Iakoucheva: But biased by popularity, eg P53 connected to every gene

Need color codes for reliability of data.
Some inconsistencies come from real variability (different conditions), some from prediction failures

Blake Borgeson: Google Cell Wiki is data repository. How do we include lots of types of data, while still having structure

JS: Need a stake in the sand. For instance, many things wrong with Goodsell drawings, but still useful. Having some reference model is useful, even if known wrong.

Goodsell models would be even more useful if they were annotated with references, metadata.

JG: Googling a restaurant usually returns a yelp first hit. Thus aggregating small resources and making them available can increase access.
JS: Also improves persistence of small

Blue 2 - Literate Models

Unofficial Scribe: Spencer Bliven (official notes are elsewhere)

Herbert: all data is in latex file to generate figures & paper
Some data too big to store, point to it online
Jim Sluke: Cache data in multiple places, for persistence

Need to render data in multiple ways
- human (html)
- machine (SBML, XML, etc)

Either program contains docs, or docs contain program

How do I convince my PI that it's worth annotating my model?
- You can't (laughter)

Enforcement
- model journal
    - policies become more strict over time
- no current standards

    - Andrew McCulloch: use domain-specific language as input to a code generator/ JIT compiler. Self-defined standard

Herbert S: Not separate model journal, but rather PLoS Comp Bio or other existing, plus stricter standards

- could have optional standards with gold/silver/bronze levels of compliance
- Issue ratings to authors based on compliance

Yolanda Gill: How do we get interoperability to integrate with others' models?

- different types of relationships between models (parallel, serial, independent)
- no consensus answer

Combining models requires thorough understanding of the models being merged. Thus automated methods usually make mistakes. Literate models are better documented, so easier to merge.

Ex of incompatible standards: HTML, SBML, and CellML all used to represent kinetic models

- interconversion requires loss of information

Green 3

Scribe: Spencer Bliven

Big data often published, but small plots often only published as JPEG figure

Need data to be published, eg SBML, which can be imported into lots of packages.

Don't only need data standards, but also need to standardize data acquisition (or at least annotate the acquisition process)

Problem: pressure to have short methods section.
Journals should require detailed methods in supplemental section

Methods are continually developed and changed, but keeping up with changing acquisition methods is very tough..

Without a common way of describing data, how do we start? Many kinds of data (eg binary reactions vs timecourse vs etc in kinetic modelling)

Microscopy looks like animated simulations, but very different formats, so can't use same tools.

Evolving so fast: in past, never would have expected synapse-level neural images, but now trivial. Hard to be flexible enough to include all future data

Sometimes DON'T want to keep all data (eg simulation intermediates)

Some very basic annotation still lacking, eg units (eg mM for "millimolar" wrong)

How do we correct errors? 1) peer review, or 2) crowd source
Journals should assume responsibility since they are paid for it

Is there a need for post-publication annotation of models? When models make incorrect predictions, how do we link that evidence? Can models be reused passed a single PhD thesis?
Kinetic models especially susceptible to version rot due to fast evolution

Since biologists usually responsible for simulation software, often use outdated software methods.
Need to use standard libraries for things like ODE solvers

NIH could provide API for basic simulation processes. Then everyone would use the most up-to-date simulator. Plus, could cache results, provide versioning, publication, etc.
But sometimes don't need to save models, since specific to one condition
For instance many basic systems like hemoglobin binding are general, but so old that they aren't in modern formats

Ideal: Siri, how does this change the cell?

Multi-scale modelling: How do we relate parameters at different scales?
Use bayesian parameters, do sensitivity analysis

Relate parameters to intrinsic variability

Red 3

What are Minimum requirements for a 3D virtual cell

Maciek Swat
Peter Rose
Andreas Prlic
Blake Borgeson
Ross Whitaker
Andrew McCulloch

What are the minimum requirements to make something useful/reusable

Software components and public API to build a model

What is a model (in the sense of simulation)?

Cell simulation of microplasma (one second intervals) – minimum requirements “we have simulated a cell” => Getting this for a human cell

No information where one protein is, but higher order interactions

What is our goal?
Do you simulate one cell cycle?
Do you provide a fly-through?

The 3D virtual cell is more than just one model, but it is a useful umbrella for many different types of models. This is a young field, the paradigms are few. Some established types of models are
Interaction diffusion model
Transcriptional networks
Metabolic networks
Putting together the different approaches is making the idea of this community most interesting. Each of the models needs experimental data as input, needs to verify the models.

Define a set of components, and how they interact.

Define cell behaviors without using model assumptions (they grow, divide, stick to each other, etc.)

We need ontologies describing the components and standards for the components to interact.

We need a registry to register components
We need tools / software / to do something with the components.
How can I find tools that help me to do something?
We need libraries to build tools.

Should NIH come on board to help fund the infrastructure? NSF has the research grants? There is examples from data visualisation.

in cell based models you have many models. you are logical, rather than mathematical. describe biologically what each cell does. How this is translated into code, who knows.
There are frameworks for rule based models (they are horrible) . These frameworks can lead to conclusions. Some critical comments about that.

Are there other type of models? “relational grounding” CS has come up with concepts that are valuable for modeling. Programming languages have been designed for that. Idea that models can be described independent from their implementation.

There are cases where the code is the model

You need some language to express some models.

Day 2: Friday, 14 December

Phil Bourne intro: quick summary of yesterday’s activities. Discussion about how to proceed, and other projects that didn’t succeed. Cell Signalling alliance project as example of prior one that didn’t deliver as desired.

Session 1A: Whole Cell Modeling track

Peter Hunter

The VPH/Physiome Project

- http://www.abi.auckland.ac.nz/uoa/physiome-project
- SBML and CellML framework into cell + tissue types, scaling to organs and organ systems
- Using standards and ontologies is crucial for robustness
- SID-ML needed to be developed
- Need to pressure journals to adopt standards and models must have them
- 500-600 curated CellML models curated but not yet annotated
- Heart now modelled from cell (nm) scale through whole-body scale (m)
- Need more annotation of models to allow API access
- Need better minimum information standards for experiments
-
-

Terry Sejnowski

http://cnl.salk.edu/ MCell

- Particle-based modelling *can* scale to cell-level (although hybrid methods still needed)
- Each neuron has 1000-10000 synapses. 10^14-10^15 in brain
- Their target is the synapse, to focus on the system. To explain changes over time.
- Scales:
    - 1Å. Protein structures of secretion system & signalling pathways
- Very small volume, so concentration meaningless (only 7±3 Ca2+ ions)
- cooperativity leads to huge number of states, so ODE impractical
- Simulate repeated synaptic excitations by tracking individual neurotransmitter release as particles
    - anatomy + physiology + biochemistry combined in the model, demo, with dramatic music

Maciej Swat

http://biocomplexity.indiana.edu/ CompuCell3D

- Some recent success at simulations of single cells (particle-based), including some that predict phenotype
- But going from single cell to organism is scale problem, 100 supercomputers to simulate one eukaryotic cell (back of envelope calc)
    - human has 10^14 cells, each 100x more complex than *Mycoplasma genitalium* (current state of art)
- Liver sinusoid example
- Need to develop more techniques to bridge scales from cell to organism
- Why do people use Matlab, not OSS?
    - easy scripting
    - many toolkits
    - widespread use
    - nice UI
    - support and training
- Need to copy these traits in any virtual cell software
    - Scalable & reusable components
    - easy integration between components
    - User friendly
    - Open source (not GPL)
    - python wrappers for easy scripting

Markus Covert

http://covertlab.stanford.edu/

- "Never trust an experiment that is not supported by a good theory" --Monod
    - Need to selectively ignore poor experiments in order to build strong model
- Built whole-cell model of *M. genitalium* with every gene & molecule, although not full 3D model
- What do we do with a whole-cell model? How to get the biological insights out of this?

- Comparing wild type and knockouts, giving qualitative predictions
- Now moving toward quantitative data to compare w predictions
- Molecular Pathology - use full simulation data to analyze incorrect predictions
- Based on incorrect predictions, can determine which parameters were probably incorrect and improve model
- Iterating between simulation and experiment hopefully converge to accurate model.

Graham Johnson

http://www.grahamj.com/ CellPACK

- Build whole cells in molecular detail
    - Starting point for simulations
    - Fill in missing data
- "mesoscale" between light microscopy and atomic resolution
- Includes packing algorithm to fill a volume
- packing algorithms must account for multiple types of structures, globular, fibrous, etc
- Specify distributions, densities
- recipies for blood serum, HIV, vesicles,generic blood serum
- allows domain experts to come in and work on that
- Online viewing, software, modeling, multiple levels of access available for the community
- Contest for 3D artists to improve HIV models autopack.org for more info on contest

Natarajan Kannan

http://esbg.bmb.uga.edu/

- kinase structures are available for many of these, but still don't understand because of the complexity at many levels
- requires data integration from many sources--sequence, structure, pathway, function, disease, etc
- Ontologies as a solution to the data integration challenge
- ProKinO protein kinase ontology browser

Panel

- Identifying modules can be a challenge in developing multiscale models
- MS: "Automatic model reduction currently a fantasy"
- MC: Case of experimental error: measured amount of DNA off by factor of 5 from calculated amount based on sequence. So sometimes model more accurate than data
- TS: SimCell game to explore a 3D cell
    - GJ: already in progress
-
- TS: Finding collaborations is like a marriage. Requires similar adaptation and alignments.
- What can be done in 5 years?
    - TS: whole-cell models, hybrid methods

    - MS: Need standard, easy-to-install libraries for standard stuff like ODE solvers
    - GJ: HIV model contest is a way to test out how a project could be set up, framework for integration http://autopack.cgsociety.org/autopack
    - PH: full 3d virtual cells for at least a few cell types
        - heart failure requires knowledge of small-scale physiological defects
- Susan Taylor: Keep biology and modelling integrated
-

Session 1B: ANALYSIS

James Sluka

- When writing paper the in between is not recorded, only observation and the result. the model and a lot of data is getting lost
- best practises for cross-scale modeling
    built on standard components and apis
    needs to biologically annotated
    model needs to run by itself
    specifications should be wrappable components
- there are data creators and data consumers

Herbert Sauro

- How to build a sustainable community
- Synthetic Biology Open Language
- Web site, active mailing list, twitter feed industry buy in
- two workshops per year
- What does SBML produce?
    - specifications, use case documentation, software/APIs, publications, outreach
- invite postdocs and graduates
- build ownership and responsibiilty
- all participants can contribute
-

Yolanda Gil

- Focus on the process rather than the data

- nsf workshop on discovery informatics discoveryinformaticsinitiative.org
- Text extraction in Hanalyzer
- Model exploration in Eureqa
- 80-20 rule, 80 percent of the time you spend on getting data ready. 20 % to get the actual work done.
-

Tim Clark

- Where is the evidence for the statement?
- Problems of reproducibility and citation distortion
- Prevent this by Direct citation of data and methods
- Only 11% of research can be reproduced

some serious sound problems in this room! talk got disrupted...

- Data citation - data needs to be a first class object
- Method citation - they are more reusable than data
- Figure needs to cite primary data, needs to be machine resolvable
- Domeo web annotation toolkit alpha version released next month

Nigam Shah

- goals for coming here: you need ontologies for building a virtual cell
- NCBO: rest.bioontology.org
- various services for working with ontologies
- Need support for pathways

Nancy Wilkins-Diehr

- Science Gateway Institute
- shows video.
- The goal is to make projects less transient

- If you build it they will NOT come
- … unless they are part of the conversation

Panel

- what do you need to do to make publications more reproducible?
- it is not the community that is the goal but the science. the community will allow you to get to the science faster
- who knows who landed first on the moon?
- Who knows who was responsible for the life feed documenting the moon landing?
- without evidence, who knows that you landed on the moon?

Session 2A: STRUCTURE BASED MODELING

Susan Taylor

- Human Kinome
- One phosphate can make a big difference in kinases
- many fundamental questions the kinases are important for
- Hydrophobic Spines through core important to understanding function
- PKA is made as an active enzyme, its activity depends on cAMP
- cyclic nucleotide binding domain is widely conserved
- PKA symmetry and allostery

Pieter Dorrestein

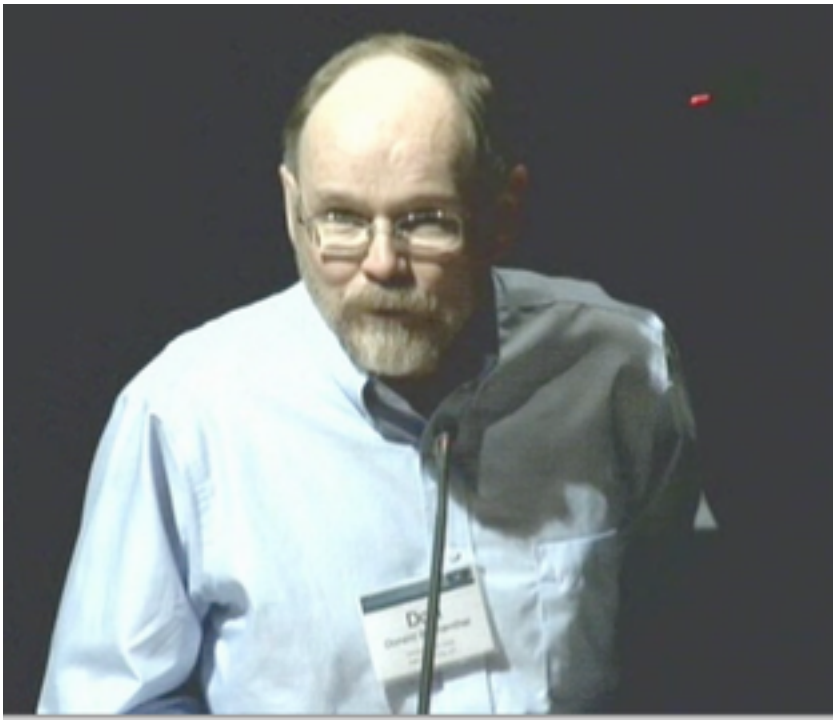
- Not good at even a basic inventory of cell components
- wants a google map view of a cell
- “Streetview”
- Use mass spec to get signature of surface molecules
    - bacterial colonies
    - student's finger!
- bar codes for molecules; MS/MS spectrum = bar code
- cluster spectra
- Use 3D network visualization for clusters

Wei Wang

- another post translational modification

- methylated lysine
- 30 chromodomains in human proteome
- What non-histone proteins are methylated and bound by chromodomains?
    - chromatin mod. enzymes
    - regulation of transcript.
    - Rna processing

Don Blumenthal



- small angle scattering
- Compare predicted scattering with exp. data
- often no structures available
    - => do ab initio models
- if structs of domains are available, put them together (rigid body model)
- neutron scattering gives information about relative positions of subunits in complex
    - Flexible structures require ensemble methods
- use small angle scattering to look at PKA isoforms
    - very different sizes and shapes, although similar weight and domain organisation
- Using different D2O/H20 mixtures, can differentiate between subunits
- how can we use 3Dvcell?
- animations for how GPCRs can work
- shows conformational changes
- what is needed to build a 3Dvcell?
    - macrom. struct
    - pharmacodynamic data
    - info about org of structs in cell

    - struct dynamics

Theodore Alexandrov

- Capturing a snapshot of the cell, organ, or organism at one point of time
- Use imaging mass spec
    - Can image 2D location of 1000s of m/z values (molecules) at a time
    - Can get 3D reconstructed from serial sections
- Got spectra for spots across the whole body
- Challenges:
    - single-cell resolution
    - in vivo imaging
    - identification
    - creating molecular atlas
    - integration with other -omics and portals

Panel

- ST: the section was more on the chemical definition of processes
- lots of techniques have only been developed recently and the amount of data is amazing
- Is the use of virtual cell models the next step for structure based drug design?
- personalized health care is a challenge
- WW: network medicine. How to model the combinatorial effect of several drugs. For this a 3D virtual cell can be really useful.
- ST:with advances of microscopy you can see location of molecules
- ST:you need to confirm that what you see in crystal structure is also what is happening in solution
-

Session 2B: EDUCATION/DISSEMINATION

Maryann Martone, Moderator

Philip Compeau

http://rosalind.info/problems/list-view/ Rosalind

- Named after Rosalind Franklin
- Programming is the literacy need of the 21st century
- MOOCs also pick up programming teaching
- Scientists need *specialized* CS education
- Currently: list of 94 computational problems w/ bio introduction, glossary items, figures
- forum for questions and interactions and solutions
- Can use any language
- Problems are collected into a learning tree, building in complexity
- Or can explore by topic
- Implemented tutorial of Python so students don't have to leave site
- stats show huge range of countries
- earn badges and achievements

- working on faculty function to use in universities, and private Rosalind interface for your specific class
- 3600 active users 6 continents in 6 months
- people approach them for expansion to other fields, Rosalind spin offs
- want to move to more actively teaching/interacting w students

Erich Huang

http://blog.clearscience.io/

- Comes from clinician/bench sci perspective
- Relates experience with a patient on breast cancer treatments
- How to make decisions based on the current science, boil down 30yrs of knowledge in the field to make a call for the patient
- Open: OpenSource, OpenData, OpenAccess. Does this really mean all the science is really clear to us? So much of what's out there is not reproducible.
- we really need to continually validate, readapt--paper is not a done/end
- Reimagine scientific communication
- make every component of project web accessible
- variety of levels
- APIs
- enable people to reproduce the whole sci environment

Maryann Martone

http://force11.org/ http://neuinfo.org/

- 2 projects: NIF neuroscience information framework and Force11
- wanted to create 3D cells, integrating from many types of expts
- spatial, semantic frameworks + federated system required
- data is still an issue, but so is dissemination
- lack of interoperability is a barrier
- NIF: catalog and integrate resources in neuroscience
- thousands of resources from NIH funding, but hard to find and know what they are
- NIF has tackled integration of hundreds, understand the problems even if they haven't solved them all
- Powerful trends: many data, many eyes, many algorithms, many analytics
- Many vehicles, including the FORCE11 communication efforts

Anita de Waard

http://elsatglabs.com/labs/anita/

- Coming from non-biologist perspective, came from physics
- problem: biology is messy, diverse, complex, big problems for models + expts
- Statistics to the rescue?
- collaboratories: work together with other labs, but can enable incidental collaboratories with access to data + metadata

- bio research can be quite insular though
- Biology is small (compared to galaxies), messy (requires benchwork), competitive
- objections + rebuttals to data sharing
- one of the biggest hurdles is reward system to move from competition to shared mission
- workflow environment to store, share, report
- social barriers are the hardest to cross

Lucila Ohno-Machado

http://idash.ucsd.edu/

- Sharing software and data in the context of the iDASH center
- barriers, little incentive to share
- resources limited, lawyers involved, constraints challenging
- Nat Centers funded to study and improve data sharing
- allow scientists for focus on science, provide infrastructure to enable them to share by 3rd party
- Cloud storage, Cloud computing
- iDASH: access controls, DUA data use agreement, some QA and tools for that
- 3 sharing models requiring different features and legal agreements: traditional downloads; users compute within the environment; export computation but keep data on site (popular for medical/health info and international collaborators)
- Patient-centered data sharing being developed as well
- One Model Does Not Fit All = main message

Panel:

- Try to wrangle discussion to 3D virtual cell with the panel
- MM: is it possible to use app store model to make things work sort of the same way
- LOM: can't have one way of doing things, won't work, data sets vary widely
- EH: not recommending taking a top down approach to that, diversity of solutions and increasing computational literacy are going to help us figure it out--don't be too stringent w standards at this point because we don't know enough what we'll need
- ADW: need to be clear on the goal: is it to connect different research, or educational tool, or really enable new science? Probably no answer to this, but need to think about that
- audience comment: importance of things like NIF; Royal society model was useful and effective at the time; app store will become equal to/important like PubMed. Won't know ahead of time how useful/how to use. Generalization of app store useful way to think
- [missed a bunch in here, sorry]
- EH: would suggest different metaphor from app store; apps is hardening things too much already, GitHub is a better model currently
- ADW: concerned that GitHub/Apps don't necessarily work together
- MM: there are many models currently, …

- [audience comment] Issues w app/github analogies: app store exists because there is an iPhone. There is no iPhone for scientists. Also, incentives. Makes money for platform owner and app writer. Missing in science. Writers to consumers ratio is 1:10000... 50000. These create a false sense of structure for thinking about this as technology problem rather than an incentive problem
- LOM: agree with the incentives problem, but can be turned around with other strategies for citations, or altmetrics. Levels of communities out there.
- ADW: cell as a metaphor? needs to consume at least by one other entity? so as not to silo things
- MM: to only target people comfortable with github would miss the biologists; there are a lot of biologists who want to learn to code too; incentives can be as simple as a spiderman badge [missed audio from PC on this and Rosalind]
- ADW: if there was a way to cross-reward somehow
- [audience comment] some journals implementing badge stuff; apps idea--papers really are an app, maybe develop that more; Kell paper that was more interactive than in the past. Move 18th century idea to 21st.
- ADW: paper doesn't carry workflow and all the data though. Commenter says they are moving that way with workflow too.
- MM: Beyond the PDF is going there, will be live streamed. Don't necessarily want to retrofit papers to go there, need to broaden scholarly communication.
- EH: platform improvements are coming to journals to take it out there in new ways
- Peter Karp: wants to know what they are finding with new models. What do you want to count instead of papers?
- MM: part of Force11 is to see what is needed, how to move there. Altmetrics.
- ADW: types of altmetrics
- PK: wants to embed ways to count usage in your software
- ADW: new metrics ideas are being discussed, incl microattribution
- PK: ontology of attributions? Group says yes, 5 exist.
- MM: there are great ideas out there already
- [audience] Gaming google scholar, ways to game altmetrics, an issue
- Phil Bourne: gaming is an issue people are aware of. Reward systems are partly an educational strategy that we have to do for reviewers. Push upward on the awareness.

Workgroups

Plenary Session & Introduction to Workgroups

- Why create a virtual cell? What is the mission?
- Groups will discuss the question, return and discuss

Strawman (Why do we need a 3D cell) -> form Groups -> Discuss -> come up with statements for the purpose -> Review->Topic

Synapsis from what came out of the discussions yesterday afternoon

There needs to be a driving biological problem

today: gather a community statement

Infrastructure.Methods
Education/Dissemination
Sociology
Community Statement
To describe the multi scale structure and associated biological processes that enable comprehensive modeling of a simple cell type that can be disseminated and easily used by others for discovery.

Group 1 (green)

Linda Nye
Don Blumenthal
Peter Karp
Tom Bartol
Eugene Kolker
Anita

Want mission statement for community, not statement for infrastructure

A: Where's the moon? How do we know when we've reached the 3d virtual cell?

TB: A cell moving around and interacting with the environment. All causal relationships.
PK: Should be broader and include many types of models.
A: Existing statement lacks clear goal

EK: need some engineering of infrastructure to get detailed model of a non-trivial cell

DB: Should enable broad use of the model, eg for education, SimCell, etc. Want wide-scale participation (eg high school students should be able to contribute)

A: Kids learn well from games, and are actually patient about tweaking parameters to reach a goal. We should better support that.
DB: Like the HIV model competition for animators
A: Known successes, eg foldit

1) *To describe the multi scale structure and associated biological processes that enable comprehensive modeling of a simple cell type that can be disseminated and easily used by*

*others for **discovery**.*

2) The mission of this community is to describe the multiscale structure and associated biological processes that enable comprehensive modeling of a cell. This structure should be easy to disseminate, testable, and available to the community for discovery.

- Not simple cell type, but multiple cell types
- falsifiable/testable
- multiple environments (gaming, education, science)
    - those are the path, not the goal
- free & libre
- discovery!
- challenge!

EK: Should we include list of benefits (scientific, educational)
DB: "This is a grand challenge, not a puny challenge"

Crowd sourcing is powerful

3) Our mission is to describe the biological processes and associated multiscale structure that enable comprehensive modeling, simulation and visualization of a cell. This virtual cell should be predictive, testable, easy to disseminate, and available to the community to enable discovery.

Final:
Our mission is to develop models of entire living cells that are testable, readily understandable, and accessible to the community, broadly defined. They must engage and challenge the community, enable discovery and accelerate scientific insights.

Group 2 (yellow)

Herbert Sauro
Friedrich Herzberg
Arthur Olsen
Andreas Prlic
Ingrid Stuiver
Shane Hutson
Blake Borgson
Roger Higdon
Chris McClendon

Susan Taylor

What is the purpose of this group
What do WE want?
Why do we think this is important?
Do we agree on the approach?
What do we want to model? Is it a single cell? not necessarily
With structure genomics - goal was as man structures as possible, no sense of biology. we don;t that.
What is our favorite system? do we want to work with many systems?

Biol process is something a cell does. It can be generic to all cells. One can model the fundamental mechanical and dynamic aspects, You may not need a community for that. Using computation to make physical models on a molecular level.

Building physical models using computational models.

Protein synthesis or microtubule assembly are specific systems. there are biol. questions that can be asked beyond that.

We want a model where we can really ask some major biological questions.

People are already doing smaller things. Some things have not been done yet. What has not been done yet? What is in common for the community? Nobody has simulated E. coli. Nobody has simulated a human cell. Can we come up with something that simulated both?

If you have a model of a simple cell. What questions do you want to ask?

- How does the cell respond to drug ?
- Protein aggregation in a cell.
- Development
- How do cardiac myocytes work?
- Stem cells

For many of these questions we don’t need to model the entire cell.

If we want to look at multi-pathway interactions we need more than the one or two pathway approaches that are common currently.

80% of all drugs go through the liver. What goes on in a liver cell?

Proteins behave differently in whole cells than in cultured cells.

Need to find a justification why we need a whole cell model.

How is ATP distributed amongst different processes ? If I had an entire cell I could model that. “how does the economy of a cell work”

Mitochondrial dynamics. How is it influenced by ATP ? How does it impact the entire cell?

We want to ask testable biological questions. Just doing one thing is not enough.
Cells living or dying. Many things are involved into that. Currently we have do western blots and other things to figure out what is going on. What is actually different? cellular pathology could become possible with a good model.

You can age progress a cell.

Apollo program stimulated many other technologies.

There is new things in biology, will push forward new approaches in computing, not just for cell biology. Has influence on cancer research, drug development.

like Airplane design nowadays.

why has nobody come up with a systems biology center for the red blood cell. (liver cell)

What about malaria? (focus does not have to be disease). If we have to chose a topic, we could chose red blood cells.lots of work already being done on neurons. What you can’t do with red blood cells. You can’t use gene expression.

in proteomics different groups of scientists (per continent) took different tissues. One might have several cell systems distributed around the world.

One could also pick e.coli. (might be the best characterised) - or one of the minimal genomes from C. Venter

Best to use simplest cell, but does it bring as much ? Perhaps better to take a more interesting cell type...

Human cells- we can use pharmacological tools. e.g. Fibroblasts, standard tissue cells.

Synthetic sciences vs. a descriptive science. We can think about this here as well. The payoff is not to make another living organism, but making a nano robot. Start as simple as possible and build up complexity.

This community wants to model at multiple scales. Molecular level upwards. Model whole pathways, signalling pathways. It is hard to get data at every level. Community can interact at each level.

There is no simple cell type

Generating tools with certain level of granulatiry.

Develop tools for different scales and cell types that will advance research and cell related technology platforms and open up new avenues of inquiry and address previously unanswerable questions.

Push technology and development of tools that cross scales and that advance interdisciplinary sciences. Science that will drive the tool development is essential. It will allow us to answer questions that we can't answer at the moment.

one of the end products: can we synthesize a living cell from components

all you need to do is feed it ATP and it does something.

what is the science that is required
computing, maths, new standards, numerical algorithms

FInal topics

Infrastructure

Education/Dissemination

Sociology
- principle of community
- community is not created, they are found and nurtured - how?
- Reward incentive systems. Why annotate models, why deposit, why do anything but papers
- the role of crowd sourcing

Infrastructure

Scribe: Spencer Bliven
Chris McClendon
Andrew McCulloch

Tim Clark
Tony Hunt
Wilfred Li
Larissa
Tom B
Yolanda Gill
Steven Peltier
Anita
Jim S
Blake
Fritz Herberg
Susan Tayler
Maryann Martone
...

- Infrastructure should unify experiment and theory (simulation)
    - Impossible to find all parameters for complex models, but often can estimate them just by physical constraints from geometry, physics, etc
- How to make modeling part of the experiment
- Need to go beyond cytoscape
- Structures and models as data - errors persistence
- Models need to be portable between formats, annotatable and already documented (interoperable, literate)
- Reproducibility - does it matter?
    - Replication (exact rerun) vs reproducibility vs inspectability
        - reproducible is less stringent but more important
    - version control infrastructure important
    - not all models equally valuable in perpetuity (need to discard data eventually)
    - Just OS changes and version rot obsoletes some data
        - But while not replicable, should be reproducible
    - Expensive studies generally not repeated, but may be reproduced
    - Should expect reproducibility for some time period (eg 5 years)
        - Community should describe timeframe
        - Longest time frame for your data is your lifetime
    - Modulated by cost of data storage
    - Where data not stored, should at least be reproducible
- A coat rack where people hang their data
- Continuum of scales - enter at any point and understand concept and content in both directions

YG: What are CS challenges?
TB: Frustrated by tiny steps

Many different user groups, many of whom prefer UI to compiling code, despite need for OSS

We should look for problems where its 1964 and we can reach the moon. Or where sputnik just flew over and we need to get moving.

TB: "Scientists have 2 important traits: curiosity and persistence"

AM: With existing tools and theory, what problems should we have solved but haven't? "Entire cell" may take away from biology problem, since most biologists really care about some specific problem. Need to map specific problems against existing technologies
JS: but whole cell interesting since includes localization problems
CM: Get info about molecular pathology

YG: What existing baby steps are out there?
TB: See Terry's talk this morning

MM: Lots of things change when scale changes (coordinate system)
We have some spacial knowledge, but it's incomplete
B: Determining localization of cell components could be a **tractable first step.**

Just putting model together can show holes in knowledge. How much space in the cell is unknown?

TH: biology is about behaviour. Infrastructure should enable adding behaviour.

AM: Biologists are interested in what happens and why.
L: Virtual cell should allow virtual experiments. Predict data

MM: Very difficult to find parameters (literature search), and no progress on that
Impossible to experimentally get parameters

TB: Useful to visualize protein structures in many ways at many resolutions.

With 3D part of 3D virtual cell can make many fundamental predictions

CM: Want to do coarse simulation, then take a snapshot to initialize fine simulation

Tim & Tom: Infrastructure should support "virtual witnessing" (steven chapen?)–doctors would sit in operating theater and all observe the experiment/operation.
SB: how differs from Openness?
Witnessing = Simultaneous observation

Sociology

paricipants
Sharon Franks
Janet Paluh
Andreas Prlic
Maciek Swat
Shane Hutson
Graham Johnnson
Herbert Sauro
Peter Karp
Blake Borgeson

need another question:
What defines a community?

idea: micro-attribution in papers: more than acknowledgment
Should you make the author of some software that you use a co-author?
what makes you a co-author?
.. only if there is personal communication and additional work involved...

community .. keep the PIs out of it, they have a different agenda.

“boy scouts for science”

community: they know each other well enough to be able to argue things out

What do you do to ensure sustained funding?

How do you pay for meetings?
    some pay for themselves

For people who do models, do you ever go to neuro science meetings?

Everybody can be part of the community, not only developers, also the child that is interested in.

What is the difference to education/dissemination?
It is both

citizen science

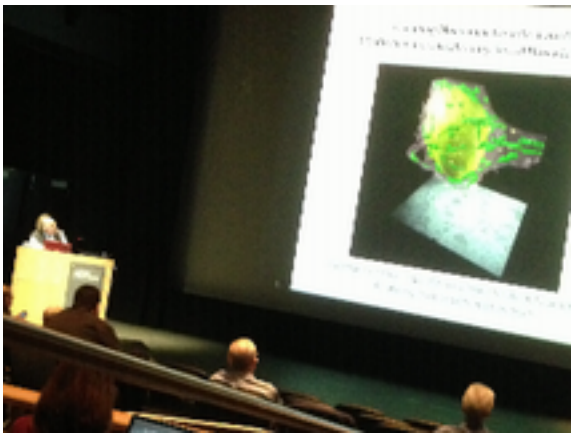
Crowd sourcing, fold-it

Do you contributed or do you use something that is produced by the community. There are many different roles.

What are rewards/incentives?
    fame?

Closing Remarks

Mark Ellisman - NCMIR



- Mesoscale
    - see microtubules: 25nm
    - see molecules: 1-2 nm
    - needs 10^21 voxels for storage
- Need a way to incorporate many people's data into single model/db
- Workflows for image processing (segmentation, meshing, simulation)
- Visual microscopy
    - requires florophor, cell resolution
- EM
    - expensive, unique microscopes needed for highres/tomography
    - lower resolutions accessible to avg laboratory
- Strategy is to make all 20+ neuroimaging labs to release data post-publication
- mitotic cell packed with ER
- Simulating synapse showed that neurotransmitters are released outside synaptic regions and then diffuse to receptors (predicted by simulation, later experimentally verified)
- Can identify proteins based on genetic tags

- Can move towards molecular scale by taking multiple EM pictures or fitting xray structures
- Now use game software & google body to display models on commodity hardware: **whole brain catalog**
- Allows microscopy images & simulations to be superimposed on the whole brain catalog
- Link out to other structures
- "Learning is the brain's response to the injury of experience"