Workflows for Parameter Studies of Multi-Cell Modeling

Randy Heiland, Maciek Swat, Benjamin Zaitlen,
James Glazier, Andrew Lumsdaine

INDIANA UNIVERSITY
PERVASIVE TECHNOLOGY INSTITUTE

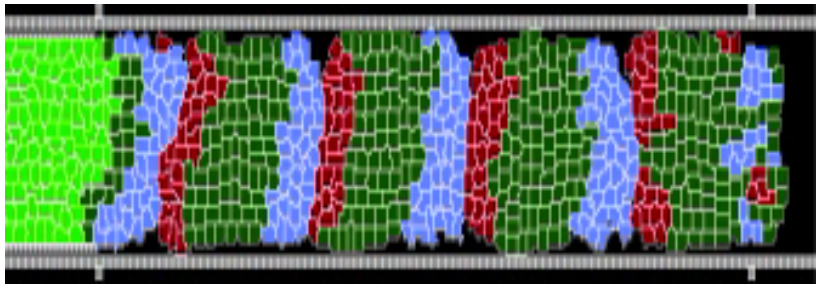
the biocomplexity institute

Outline

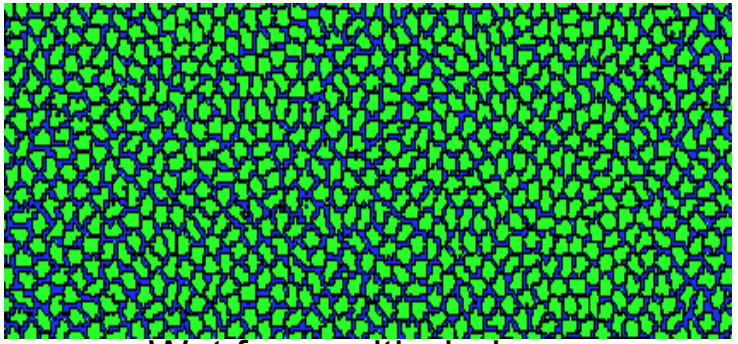
- Overview of CompuCell3D (for multi-cell modeling)
- Overview of VisTrails (for workflows)
- Overview of TeraGrid (for HPC; parameter study)
- Implicitly promote Python and FOSS
- Use case of parameter study for biological cell sorting

Multi-Cell Modeling: CompuCell3D

- Indiana University (started at Notre Dame)
- Software framework/application
  - Framework: general-purpose cell-based modeling
  - Application: standalone app w/ GUI; run batch (no GUI)
- Cells: biological or other (e.g. foams, grains)



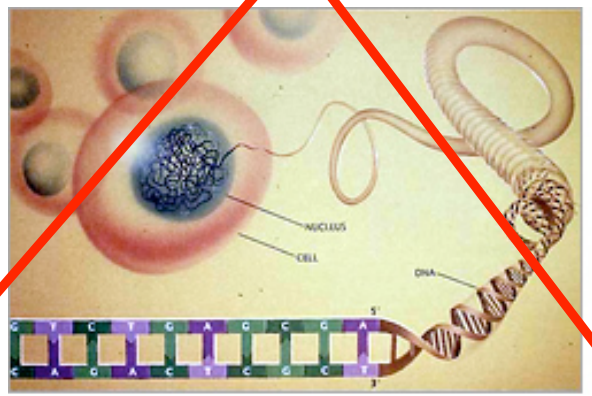
Somite sim (Julio Belmonte, IU)



Wet foam with drainage

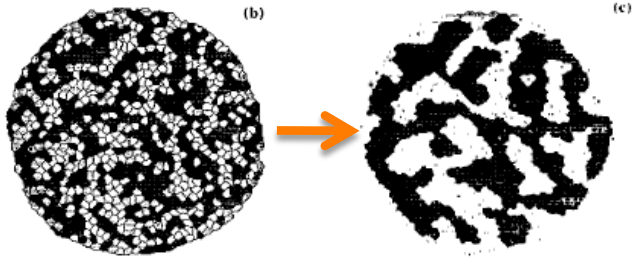
CompuCell3D: what it isn't/is

- Intracellular model
- Genetic mechanisms: genes, proteins, …



from genome.gov

- Intercellular model
- Physical mechanisms: elasticity, mitosis, apoptosis,chemotaxis, …



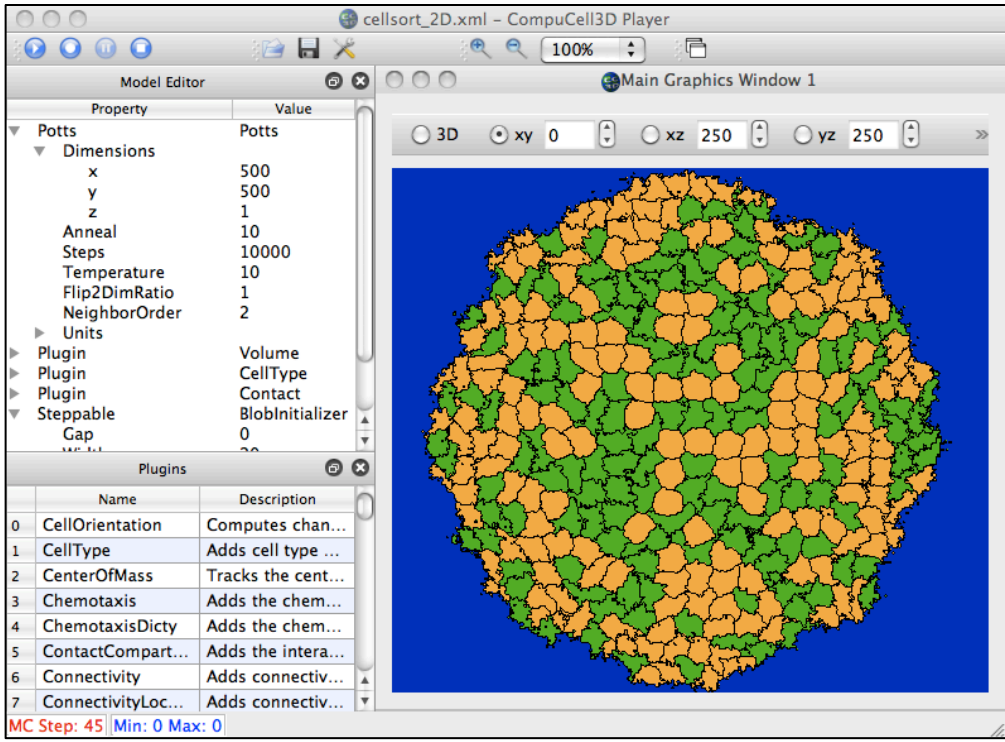
(rf. D'Arcy Thompson: On Growth and Form)

CompuCell3D

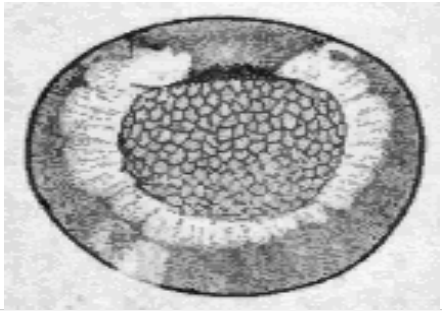
- Origin: Cellular Potts Model

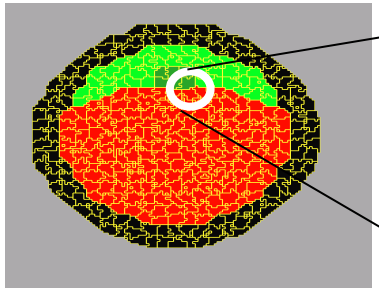
(now Glazier-Graner-Hogeweg model)

- Energy minimization formalism
- Lattice-based domain (2D or 3D)
- Stochastic algorithm
- C++ code base
- Python interface
- Open source
- *Download & try the Demos*

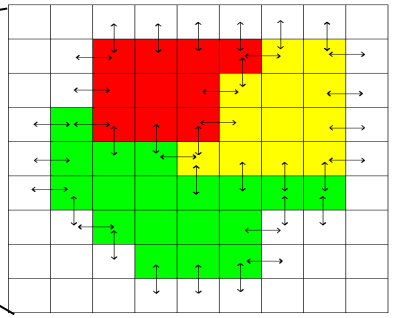


Cell defined on a regular lattice

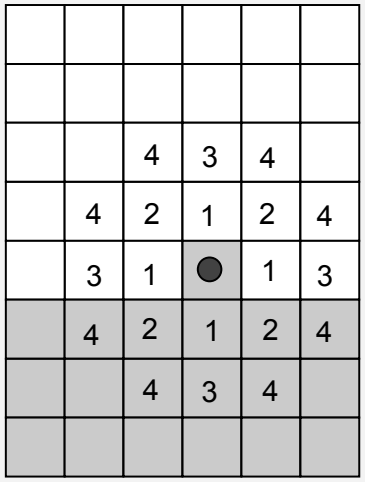


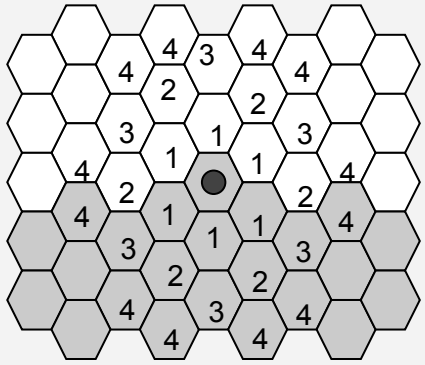


x 20



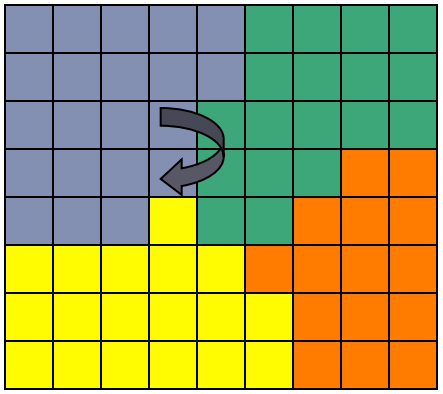
Cell types (colored)



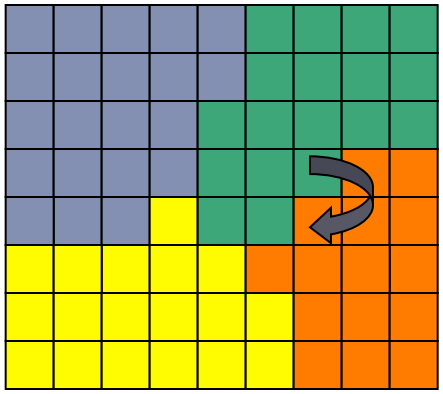


2D: Square or Hex lattice
3D: Cube or Rhombic Dodecahedron
(showing neighbors at different distances)

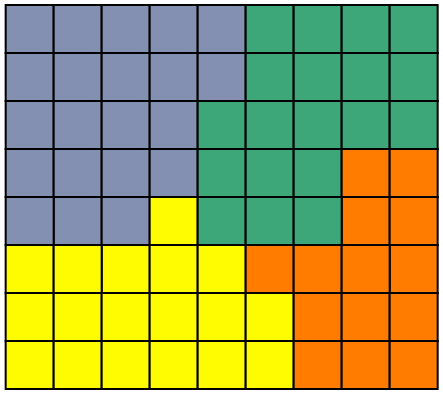
invalid attempt



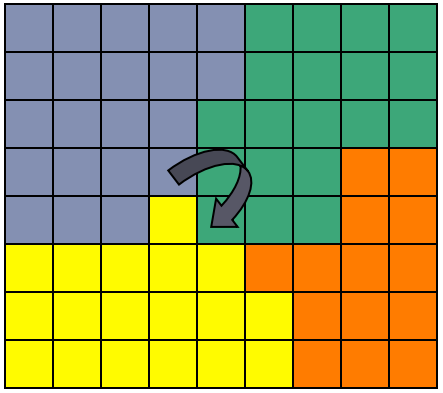
valid attempt



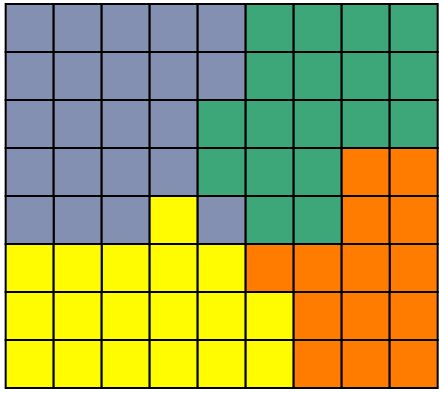
accept



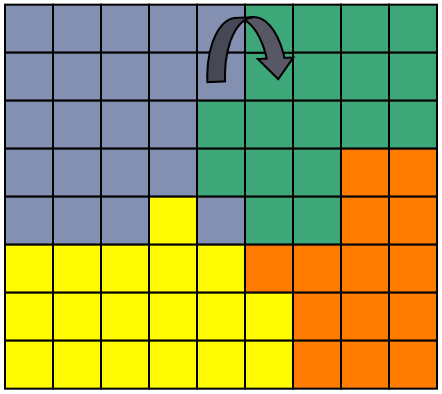
valid attempt



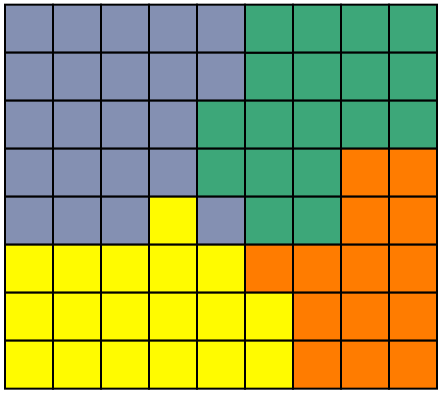
accept



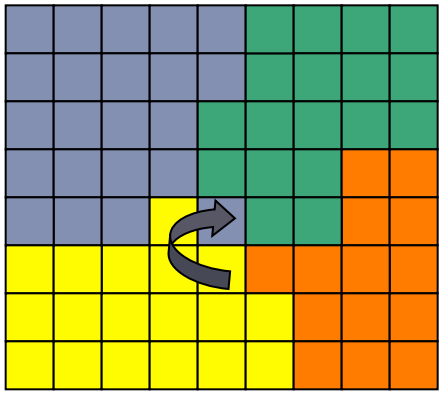
valid attempt



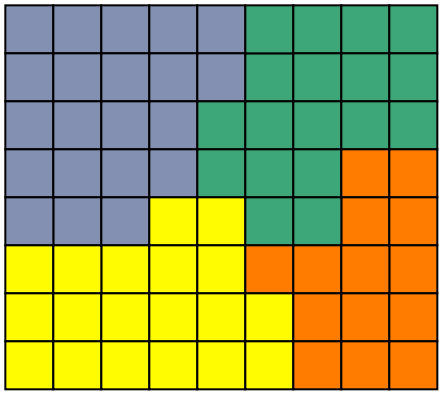
reject



valid attempt

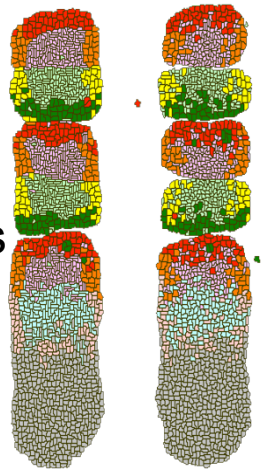


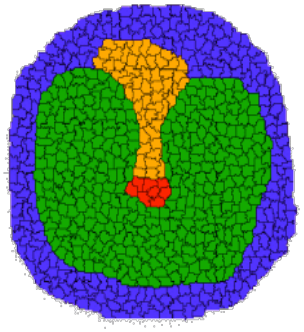
accept



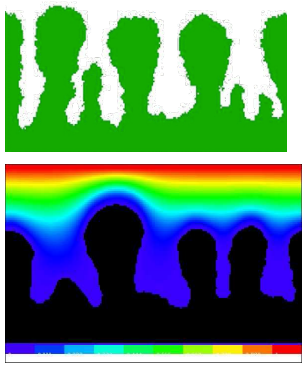
Gallery

somitogenesis

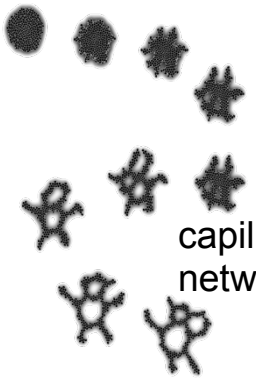




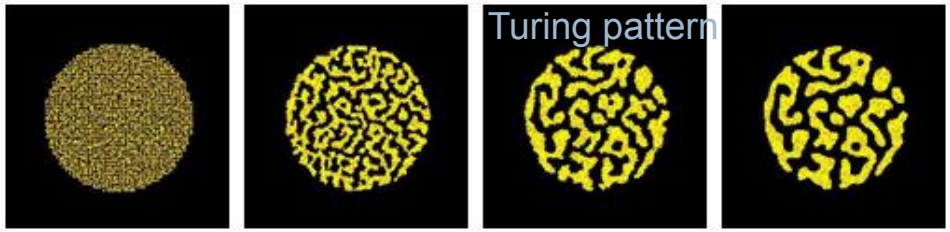
gastrulation



biofilms

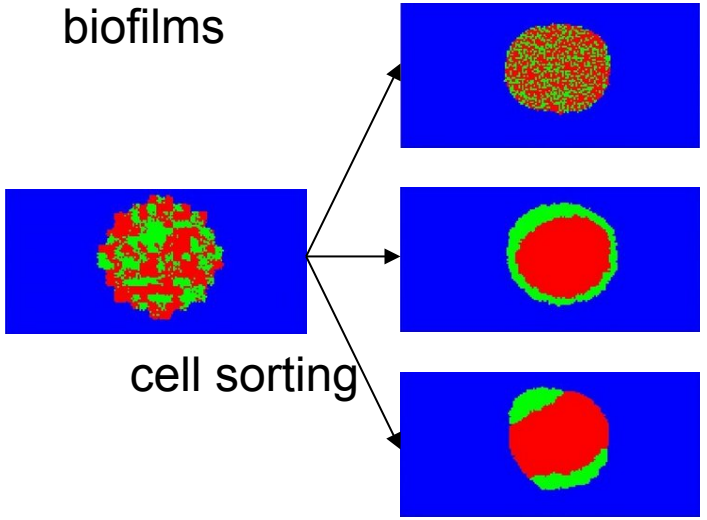


capillary network



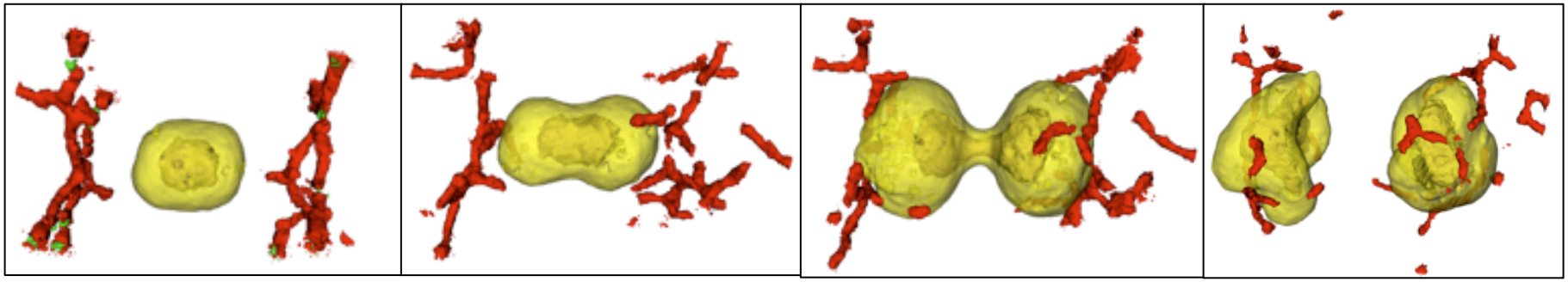
Turing pattern

dictyostelium



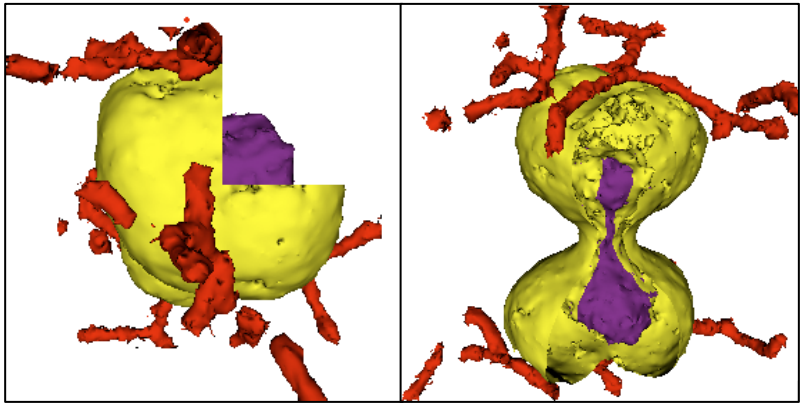
cell sorting

Tumor growth & vascularization



(see accompanying movies)

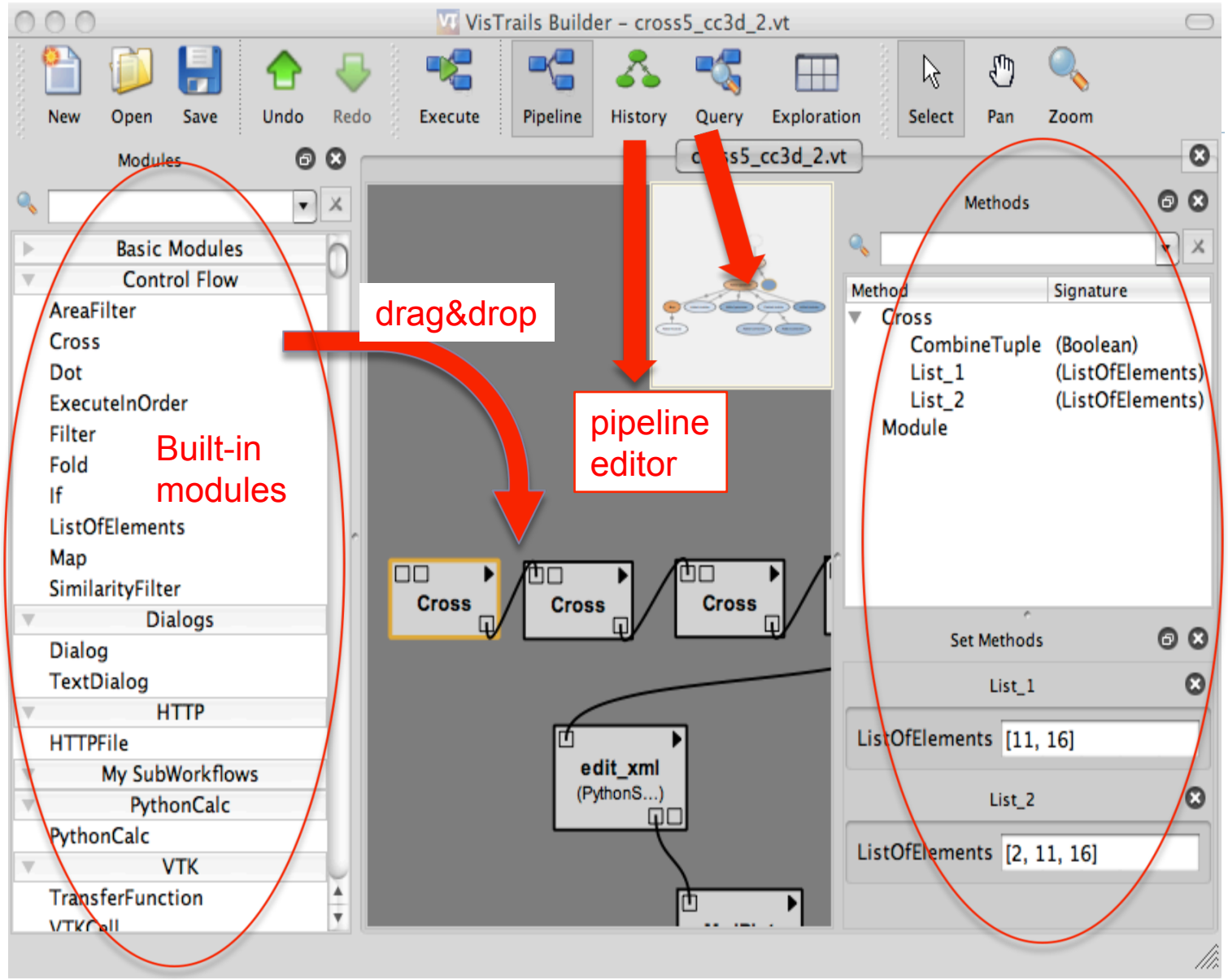
"3D Multi-Cell Simulation of Tumor Growth and Angiogenesis", Shirinifard, Abbas *et al.* *PLoS ONE*, Oct. 2009



Cut-away view to reveal the necrotic core

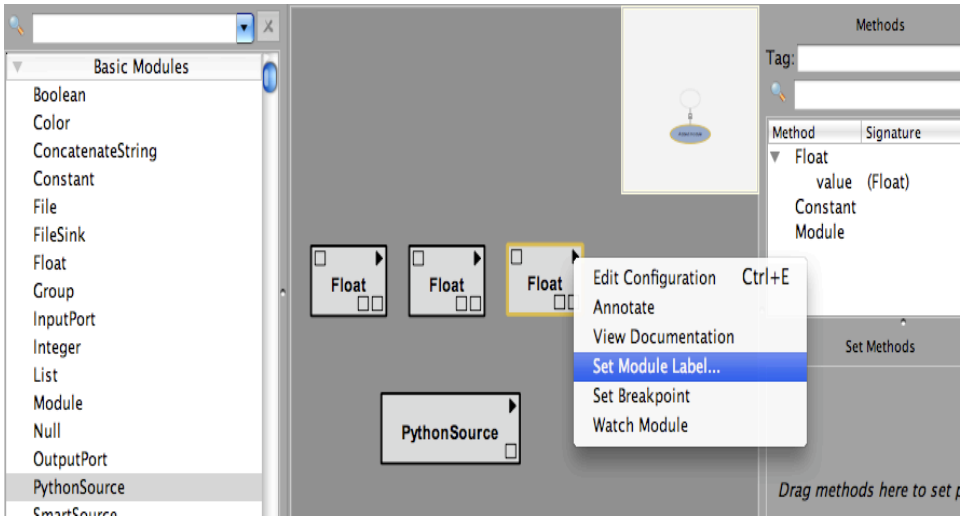
Workflows: VisTrails

- Scientific workflow and provenance mgmt system
  - Workflow: visual representation (pipeline) of dataflow
  - Provenance: recording/annotating data
- Drag & drop pipeline editor, built-in modules (visualization, control, etc)
- Written in Python
- Open source (vistrails.org)
- University of Utah (SCI Institute)

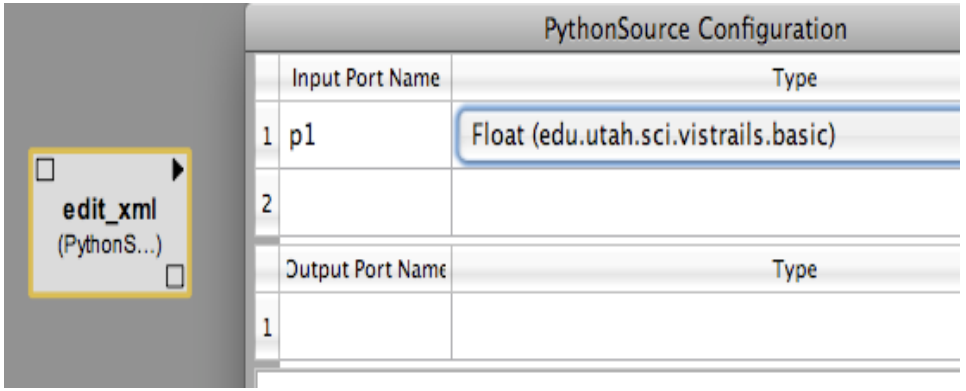


VisTrails GUI

VisTrails: Easy to use and Extensible



Can create/edit
Input/output ports



VisTrails → parameter sweeps → CC3D

- Models have many parameters
- Not uncommon to have single simulation run several hours on a workstation
- Let's farm out the simulations to a distributed compute resource
- Where/How?

TeraGrid

- Distributed resources (HPC, storage, etc) w/ high-speed networking
- Common software stack
- NSF-funded
- Open science

- Overkill for our purpose, but convenient

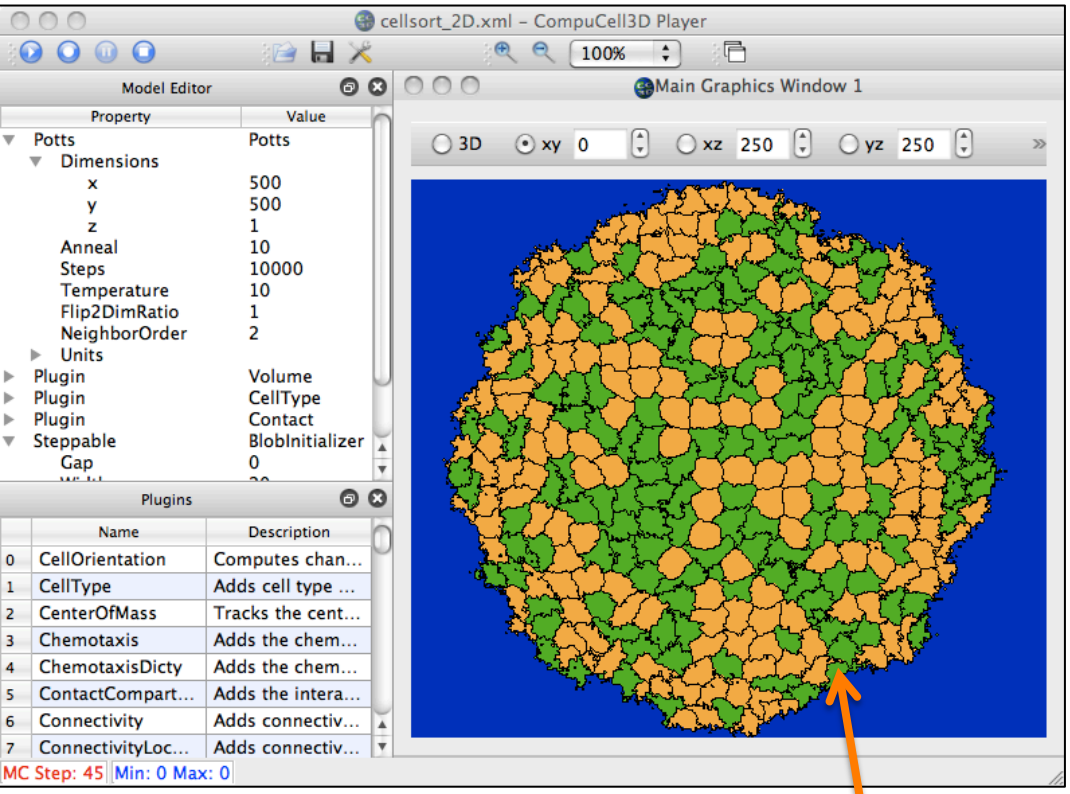
Indiana Univ: a TeraGrid Resource Provider

- IBM PowerPC “Big Red” cluster

Fall 2006 – Mar. ~~2010~~ 2011

- s/w stack includes Globus
  - Grid certificate
  - Remote job submission

Use case: 2D biological cell sorting



CompuCell3D

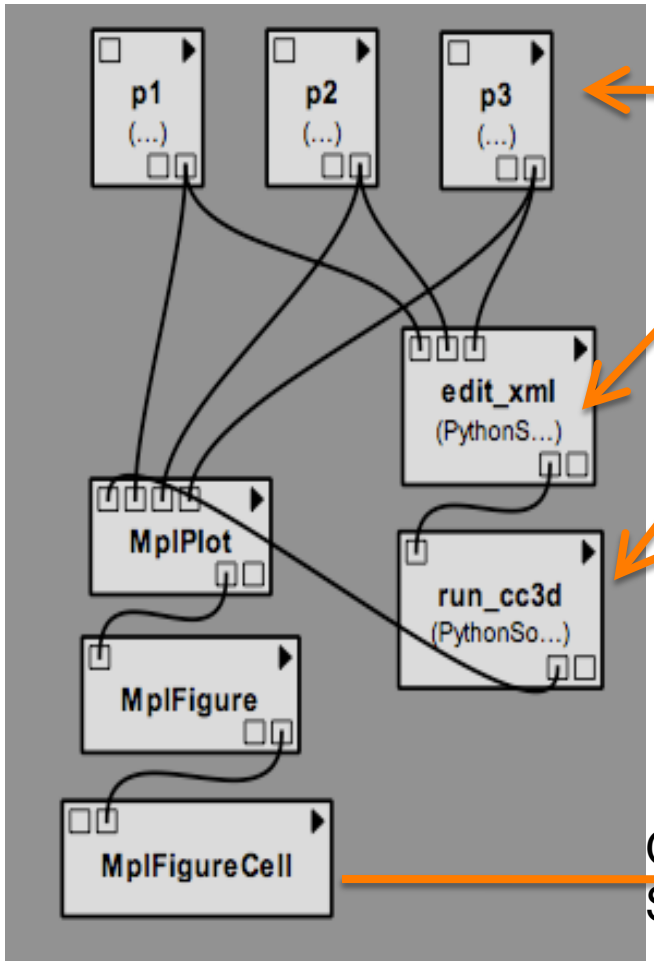
- Different cell types tend to adhere (or not) to neighbor cells

  → cell sorting

- Parameter study of "contact energies" between cells

2 cell types
surrounded by ECM

Cell sorting workflow: version 1

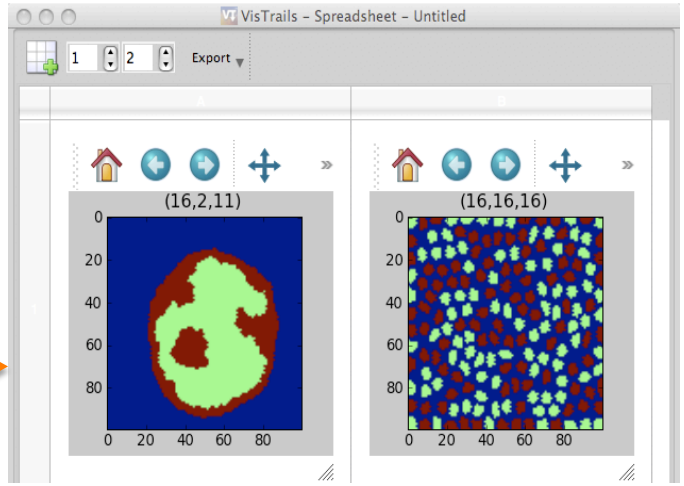


Manually enter 3 param values

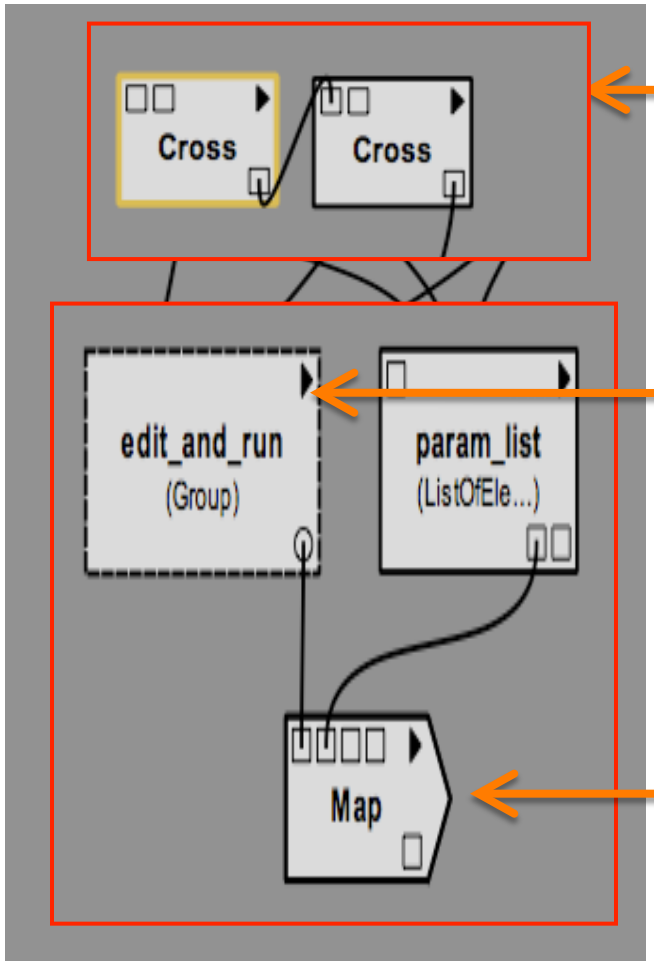
Python script to edit XML input

Run locally

Output to Spreadsheet



Cell sorting workflow: version 2 (part A)



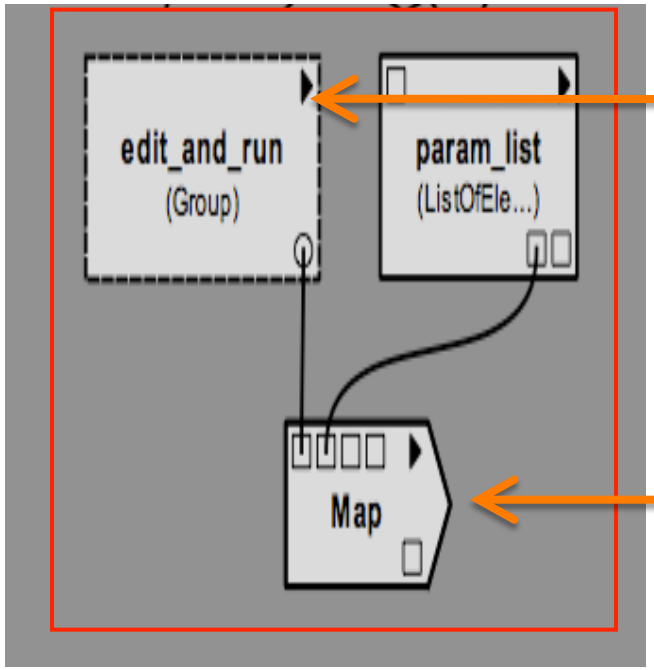
Use Cross product modules for params

Python script to edit XML input and

Submit remote jobs (via Globus cmds)

Use Map module: param set → job

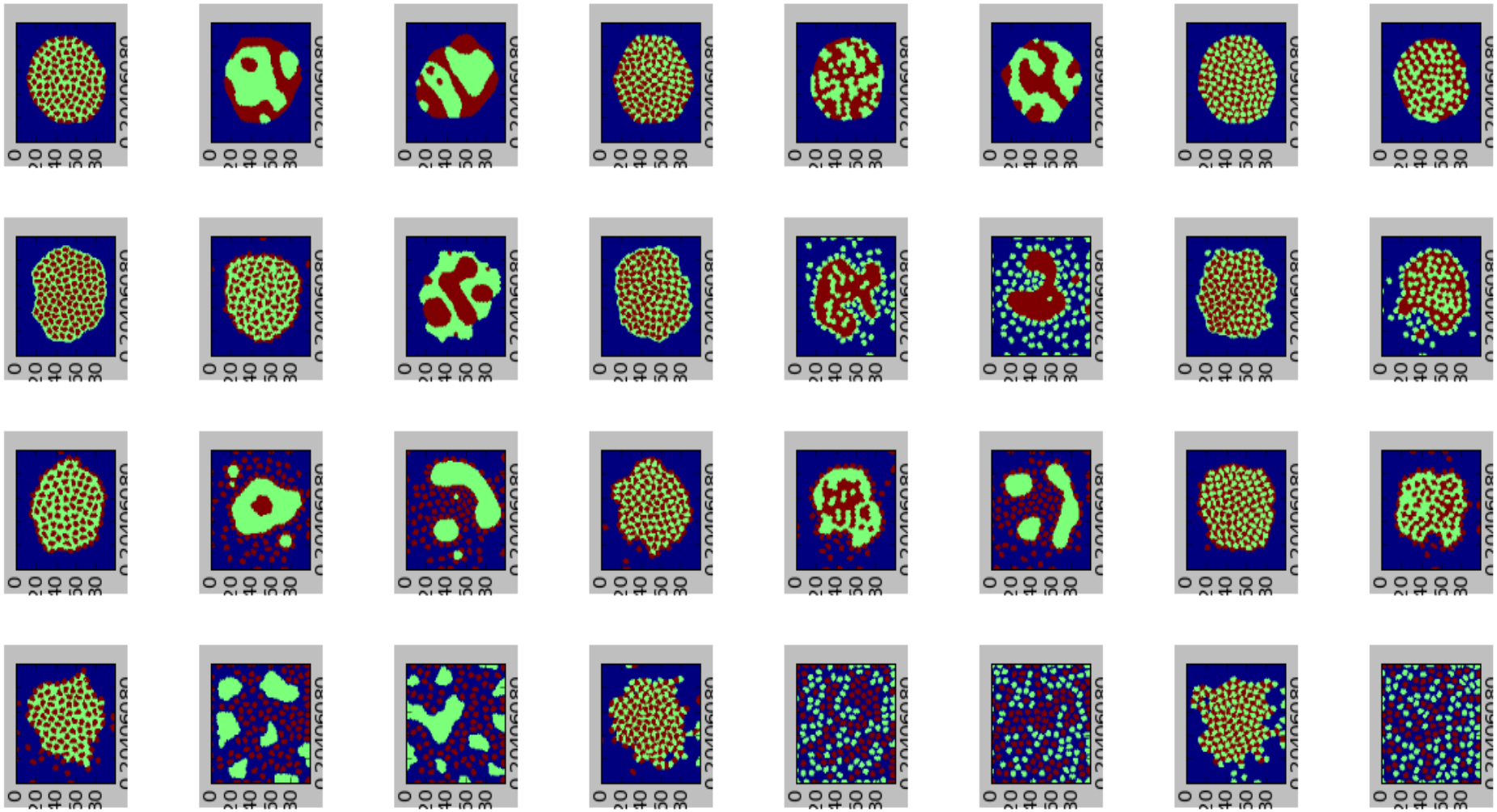
Cell sorting workflow: version 2 (part B)



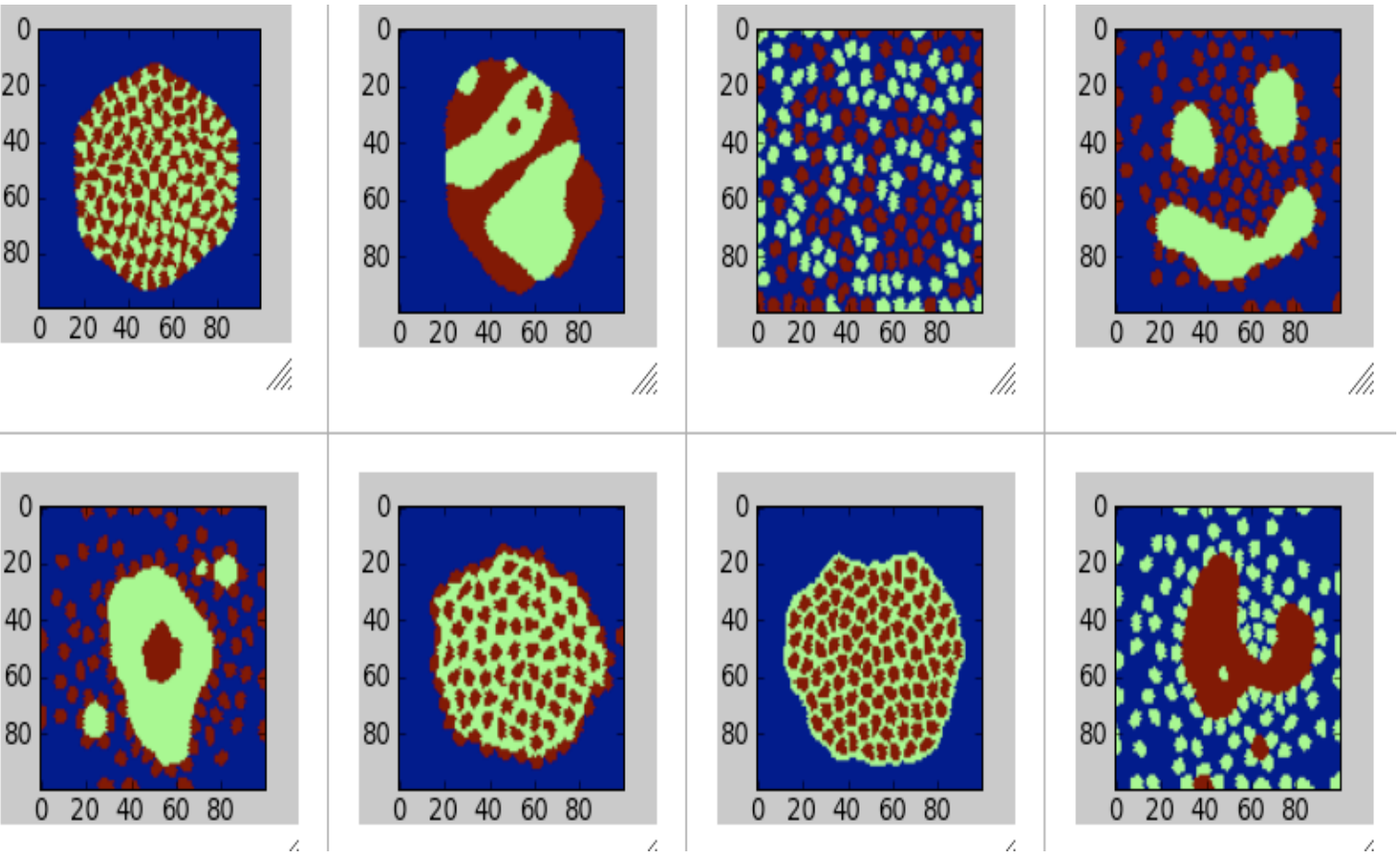
Retrieve remote results

Map module: output file → render

Subset of cell sorting results (32/72)



Qualitatively different outcomes



Caution:
NOT steady-state solutions, but…

Summary: Basic recipe

in 6 EASY steps!

1. Install VisTrails (on your workstation)
2. Get TeraGrid/Big Red account and Grid certificate
3. Install CompuCell3D (remotely; locally for local tests)
4. Install Globus client s/w on your workstation
5. Workflow for parameter sweep (remotely via Globus)
6. Workflow to retrieve results and display in VisTrails

Credits

- NIH NIGMS 1R01 GM076692-01
- Biocomplexity Institute, Indiana University
- Pervasive Technology Institute, Indiana University
- Lilly Endowment, Inc.

Thanks for your attention!

(If anyone has interest/experience in bifurcation analysis of stochastic processes, please email me)