

CellOrganizer: Image-derived Models of Subcellular Organization over Time and Space

Robert F Murphy

External Senior Fellow, FRIAS

Ray & Stephanie Lane Professor of Computational Biology,
Carnegie Mellon University



RAY AND STEPHANIE LANE
Center for Computational Biology

Carnegie Mellon

Goal

- Represent the spatiotemporal patterns of all proteins in all cell types under all conditions for use in cell and tissue models/simulations

Assignment to classes

- Main current approach is description of subcellular location of proteins using words (e.g., GO terms)
- Can be directly assigned from images (Boland, Markey & Murphy, 1997) with better performance than humans (Murphy et al 2003)
- Can be predicted from sequence

From recognition to modeling

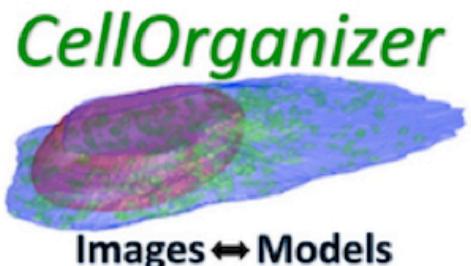
- Cells and cell components don't fall into distinct classes
 - sizes, shapes, patterns on a continuum
 - most have unique, overlapping distributions
- Solution: model subcellular patterns mathematically rather than verbally

Modeling choices

- Descriptive vs. generative
 - Conceptual vs. data-driven
-
- Generative models can be combined:
ideal for building higher level models
 - Building of data-driven models can be
automated, verified, and updated

Updated Goal

- Build generative models of the spatiotemporal patterns of all proteins in all cell types under all conditions



[Home](#)
[People](#)
[Publications](#)
[Downloads](#)

The **CellOrganizer** project provides tools for

- learning generative models of cell organization directly from images
- storing and retrieving those models in XML files
- synthesizing cell images (or other representations) from one or more models

Model learning captures variation among cells in a collection of images. Images used for model learning and instances synthesized from models can be two- or three-dimensional static images or movies.

Current components of **CellOrganizer** can learn models of

- cell shape
- nuclear shape
- chromatin texture
- vesicular organelle size, shape and position
- microtubule distribution.

These models can be *conditional* upon each other. For example, for a given synthesized cell instance, organelle position is dependent upon the cell and nuclear shape of that instance.

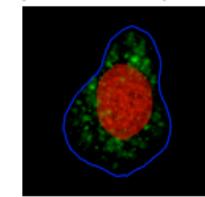
Cell types for which generative models for at least some organelles have been built include human HeLa cells, mouse NIH 3T3 cells, and Arabidopsis protoplasts. Planned projects include mouse T lymphocytes and rat PC12 cells.

RAY AND STEPHANIE LANE
Center for Computational Biology

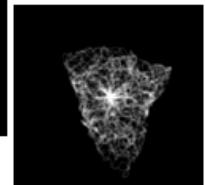
Carnegie
Mellon
University



Synthesized Cell Images
(click to view)



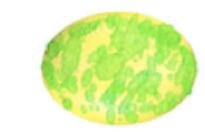
[2D HeLa
\(endosomes\)](#)



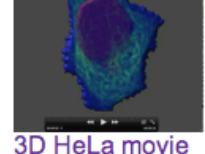
[3D HeLa
\(microtubules\)](#)



[3D HeLa
\(mitochondria\)](#)



[3D protoplast
\(chloroplasts\)](#)



[3D HeLa movie](#)

Support for **CellOrganizer** has been provided by grants GM075205 and GM090033 from the [National Institute of General Medical Sciences](#), by a Forschungspreis from the [Alexander von Humboldt Foundation](#), and by the [School of Life Sciences of the Freiburg Institute for Advanced Studies](#).



Alexander von Humboldt
Stiftung/Foundation



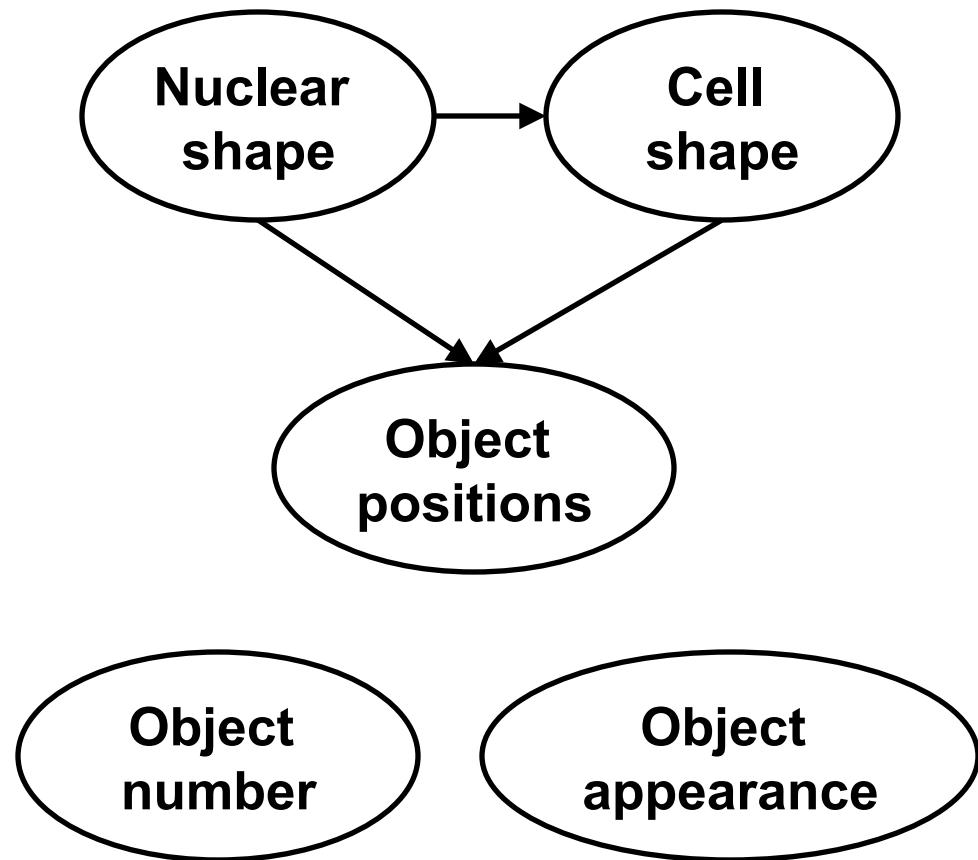
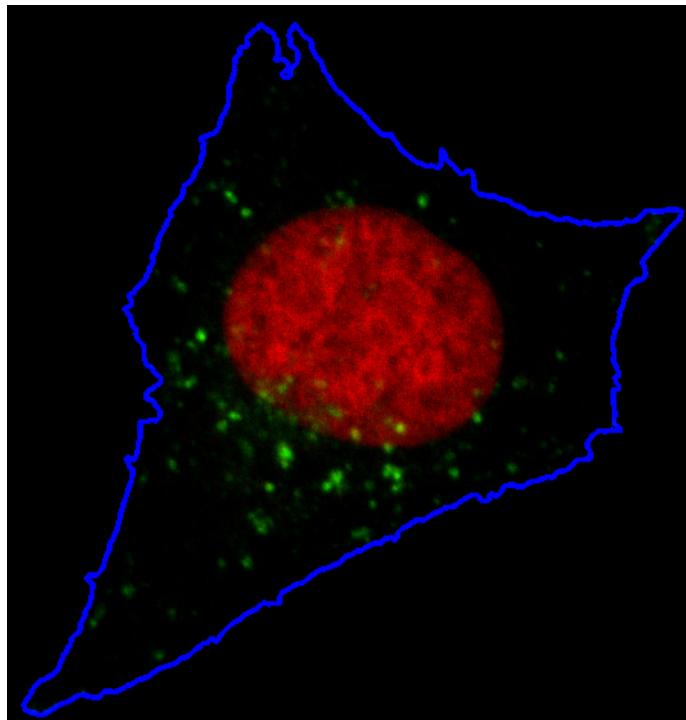
ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG



FRIAS
FREIBURG INSTITUTE
FOR ADVANCED STUDIES

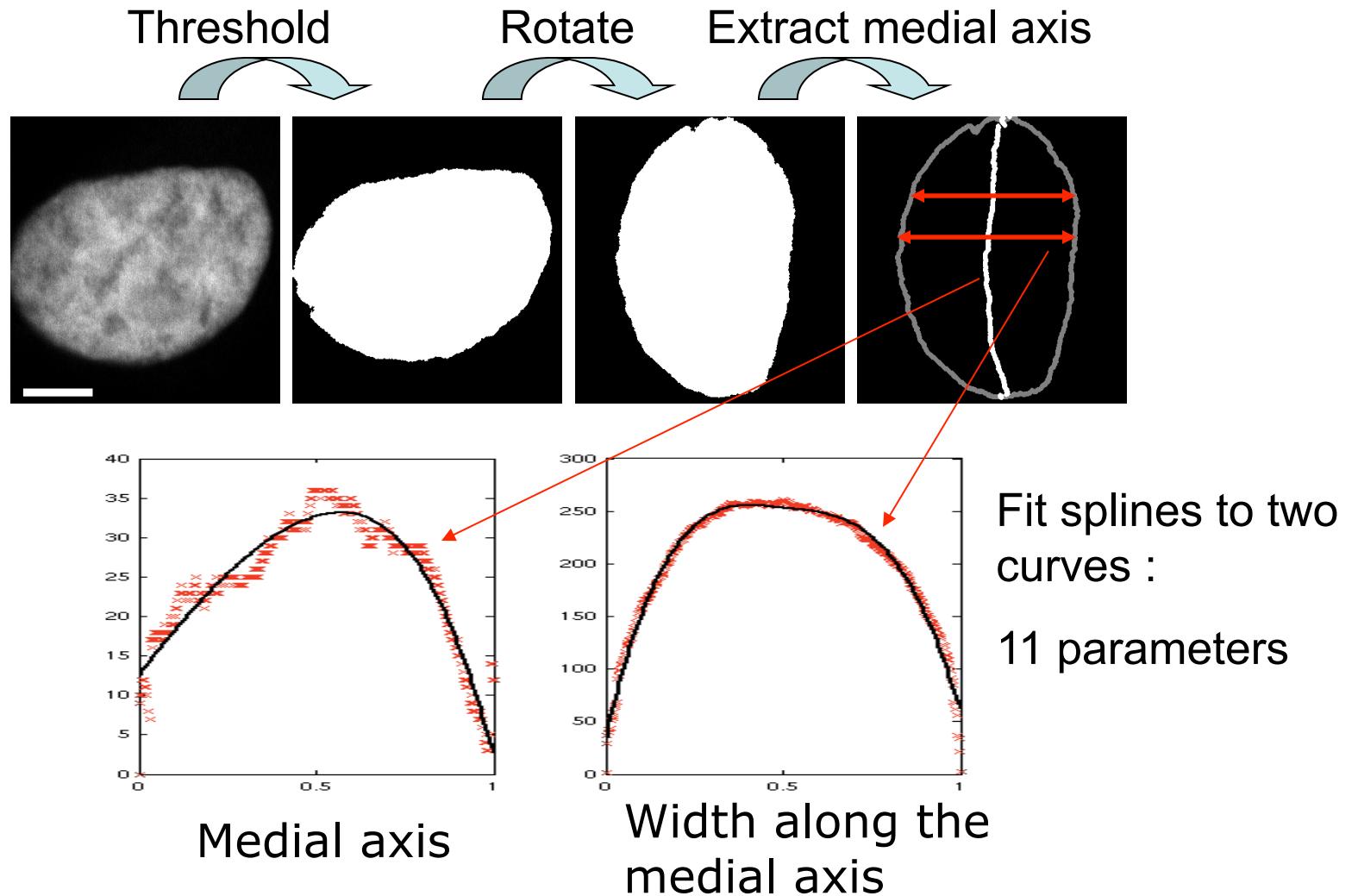
LIFE SCIENCES - LIFE.NET

Generative model structure



NUCLEAR SHAPE

Nuclear Shape - Medial Axis Model



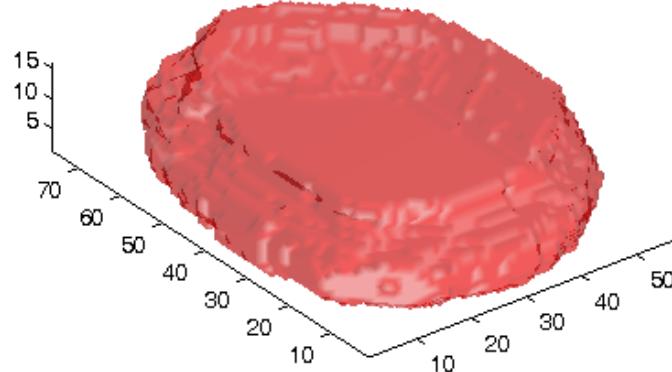
Shape generation

- 11 parameters for each object
 - 5 parameters for each curve
 - the length of the medial axis
- Learn the distribution of parameters over many nuclei
 - Assume multivariate normal
- Randomly sample parameters from distribution
- Construct nuclear shape using the sampled parameters

Synthesized nuclear shapes

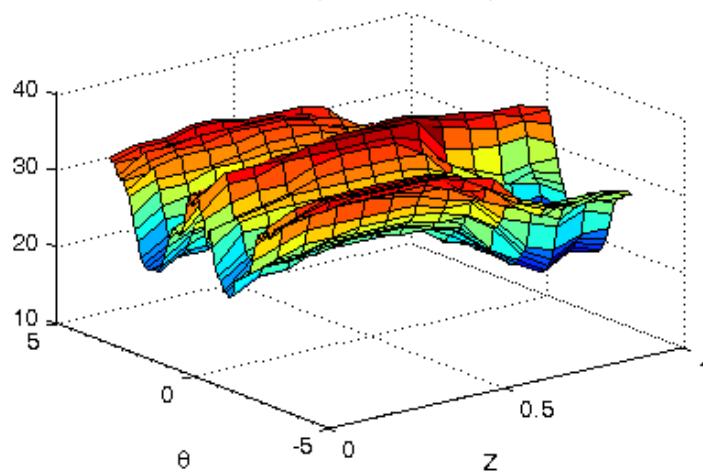


Extending to 3D: cylindrical spline surface

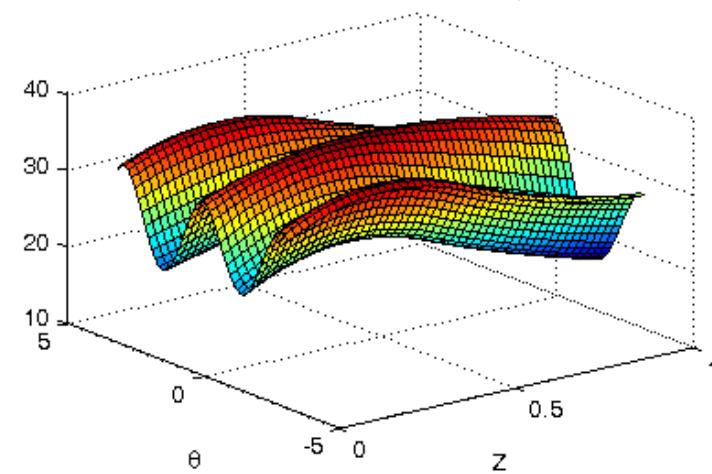


Tao
Peng

Resampled surface map

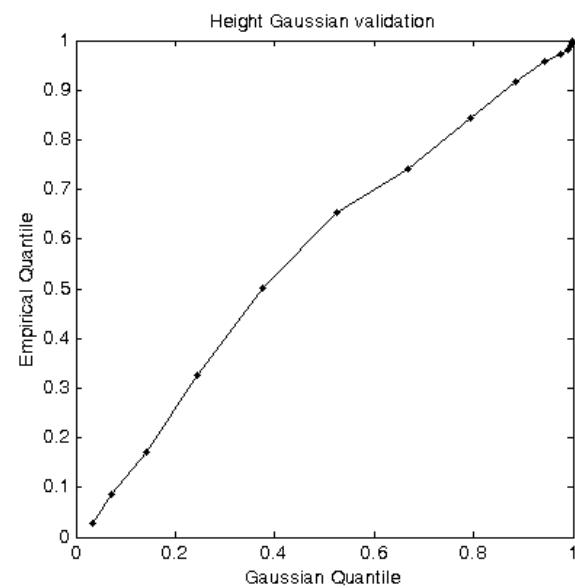


Reconstructed surface map



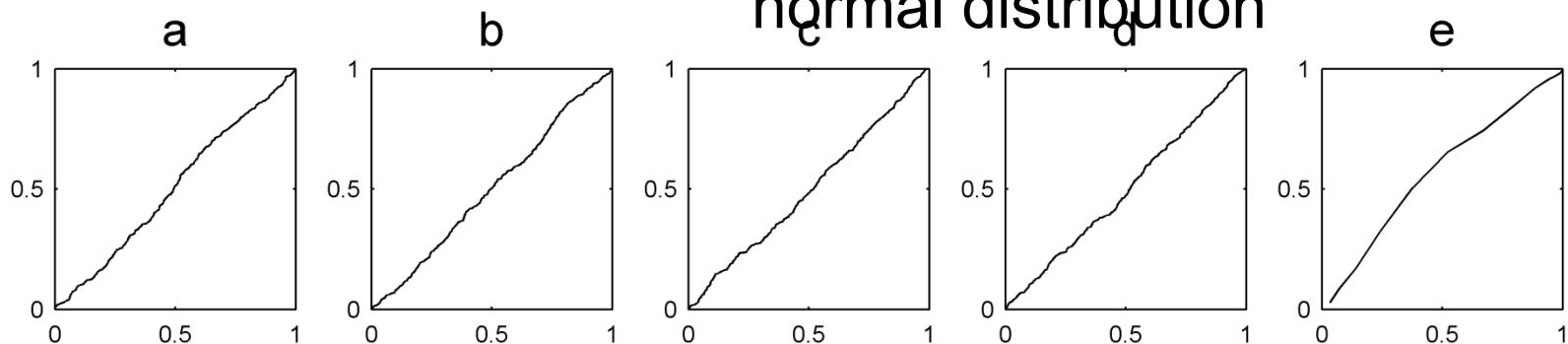
Spline surface model

- Model parameters for each nucleus
 - Coefficients: 32 parameters
 - Height: 1 parameter
- How do we capture the variation in these parameters for population of nuclei?
- Multivariate Gaussian?



Normality of the cell height

P-P plot of the distribution of thickness of cells vs. a normal distribution



Normality of the spline surface coefficients

P-P plot of the distribution of randomly selected coefficients vs. fitted normal distribution (based on 450 nuclear shapes)

Diffeomorphic analysis of nuclear shape

- Can use distance between shapes to characterize shape space instead of parameters of model

Gustavo
Rohde

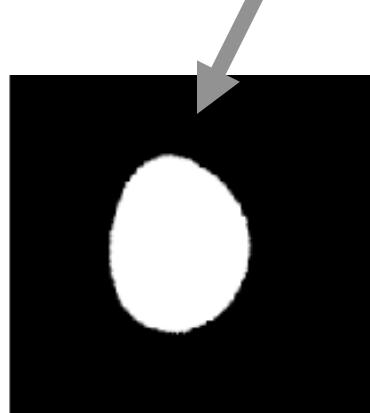


LDDMM – Large deformation diffeomorphic metric mapping: Miller and colleagues

- Goal: Find a function $g(x,t)$ which smoothly transforms an image I_n into an image I_m as t goes from 0 to T
- Choose $g(x,t)$ to minimize sum of
 - Total deformation in g from 0 to T
 - Distance between I_m and $I_n(g(x,T))$

Morphing one shape into another

Starting shape



Target shape



0

0.0165

0.0191

0.0194

0.0195

Distance



Shape space

- Can measure distance between all pairs of nuclei to construct a “shape space”
- Captures essential aspects of how shapes vary
- Can be applied not just to nuclei but to cells, organelles, etc.
- Can be applied in 2D, 3D, 4D and to more than one component at a time

CELL SHAPE

Conditional models

- To ensure that the proper relationship exists between a synthetic nucleus and a synthetic cell surface, the models must be *conditional*
- Model for cell shape captures how cell shape *depends on* nuclear shape

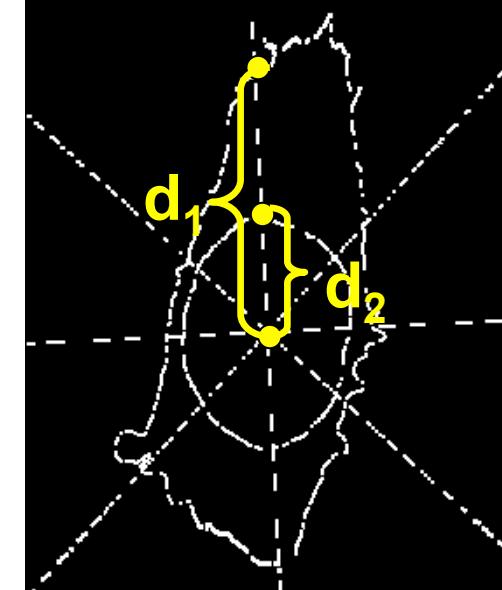
Cell shape: 2D eigenshape model

- Conditioned on nuclear shape: $r = d_1 / d_2$
 - Sample evenly around a circle to represent the shape by radius ratios

- Parameterization

$$\mathbf{r} \approx \bar{\mathbf{r}} + \sum_{i=1}^k b_i \Phi_i$$

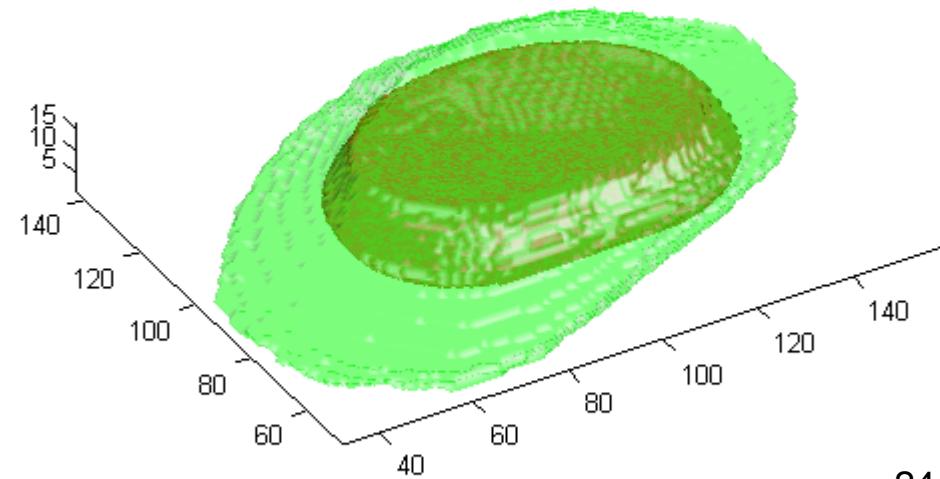
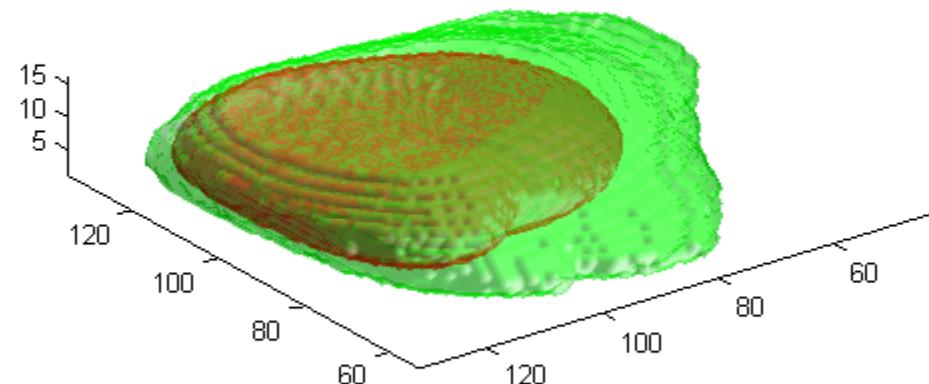
- Keep 10 principal components



Cell shape: 3D eigenshape model

- Thickness of the cell varies
 - Resize along the z-axis to 18 slices
- Do the same radius ratio as for 2D
 - $18 \times 360 = 6480$ coefficients
 - Use 25 principal components

Synthesized cell and nuclear shapes



VESICULAR ORGANELLE MODELS

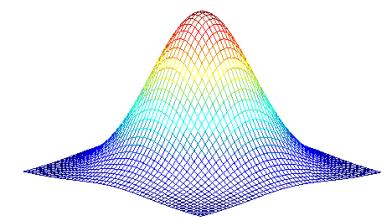
Protein object model

- Gaussian object model
 - “Vesicle”-like objects
 - Distributions of parameters

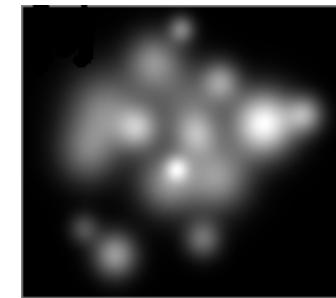
Number of objects in a cell N

Object size $\sigma_x, \sigma_y, \sigma_z$

Single object fluorescence F



- Aggregated objects: Gaussian mixture
 - Expectation-Maximization used for decomposition



Gaussian mixture

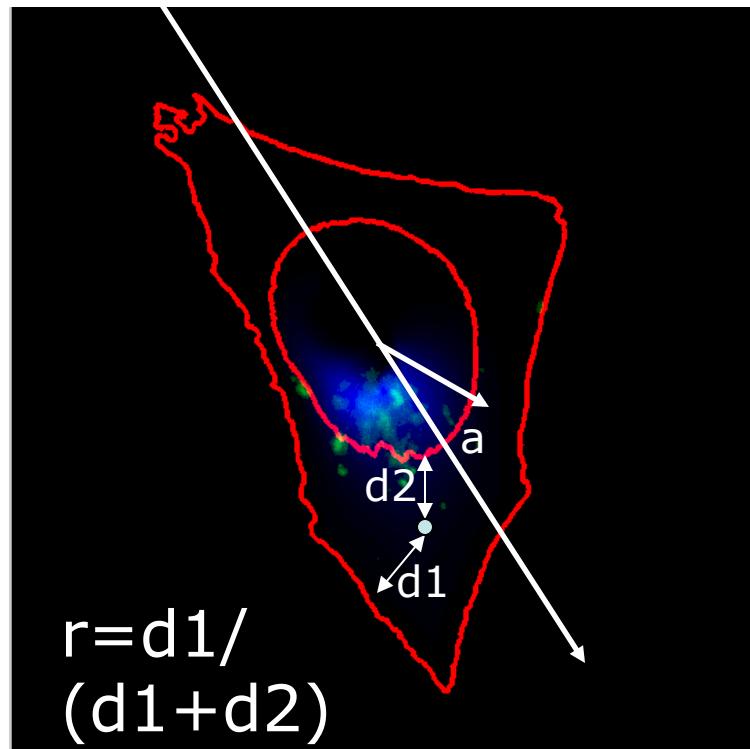
Protein Object Model: size, shape, intensity

- Size of each ellipsoid represented as distribution of length of major axis and conditional distributions of lengths of other axes relative to major axis
- Exponential distribution for intensity of each object

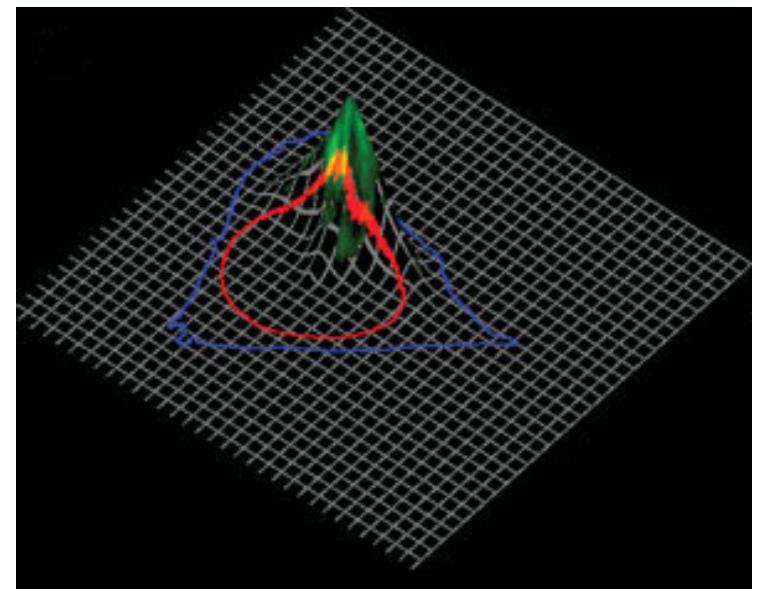
Protein Object Model: Position 2D

Potential: $P(r, a) \propto \frac{\tau(r, a)}{1 + \tau(r, a)}$

where $\tau(r, a) = \exp(\underline{\beta}_0 + \underline{\beta}_1 r + \underline{\beta}_2 r^2 + \underline{\beta}_3 \sin a + \underline{\beta}_4 \cos a)$



Normalized potential map:
A 2-d density function



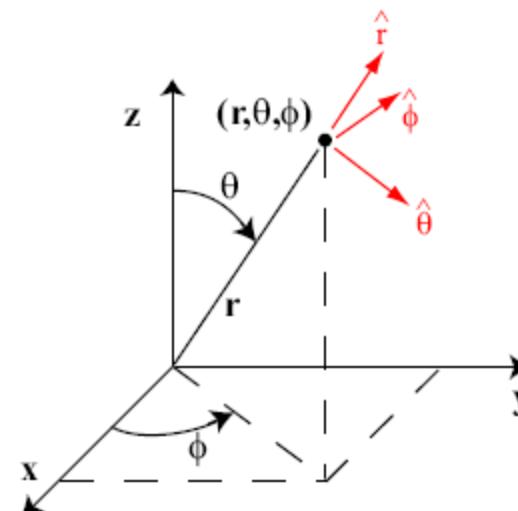
Protein Objects Model: Position 3D

$$P(r, a) = \frac{e^{\tau(r, \theta, \varphi)}}{1 + e^{\tau(r, \theta, \varphi)}}$$

$$\tau(r, \theta, \varphi) = \beta_0 + \beta_1 r + \beta_2 r^2$$

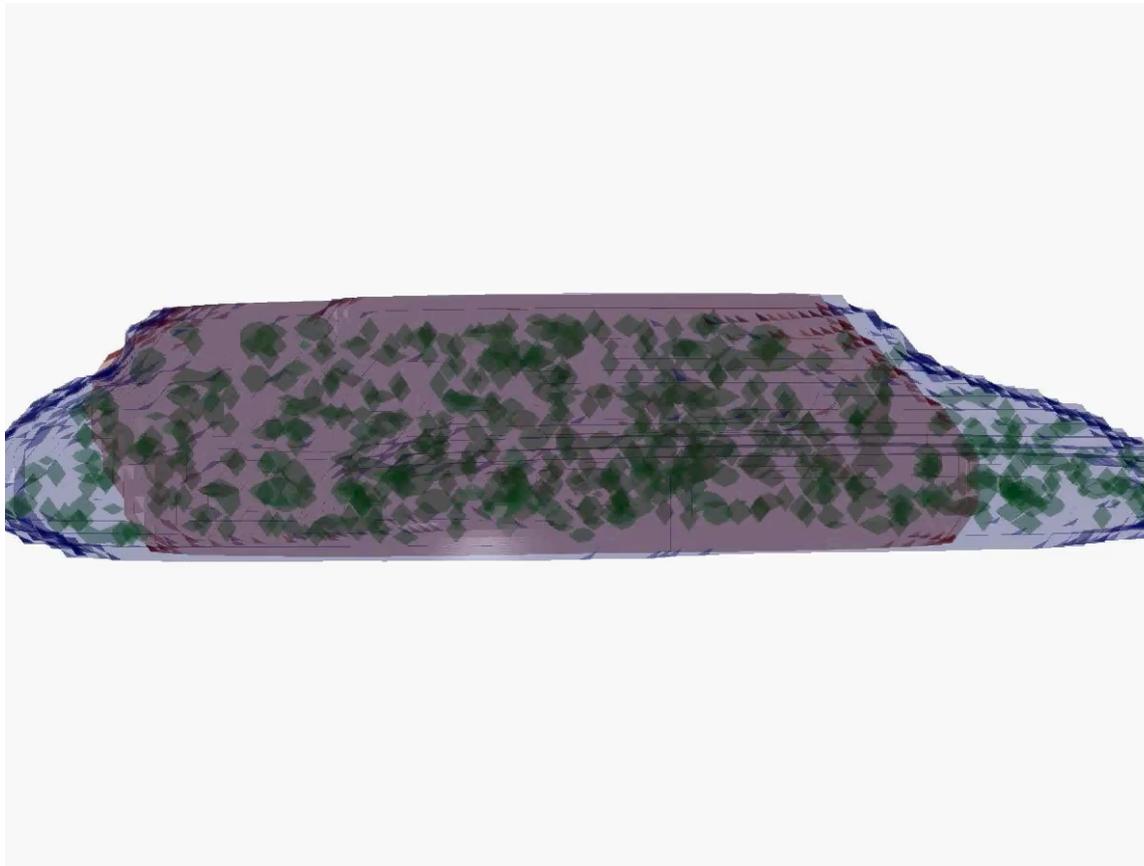
$$+ \beta_3 \sin \theta \cos \varphi + \beta_4 \sin \theta \sin \varphi + \beta_5 \cos \theta$$

Radial
preference

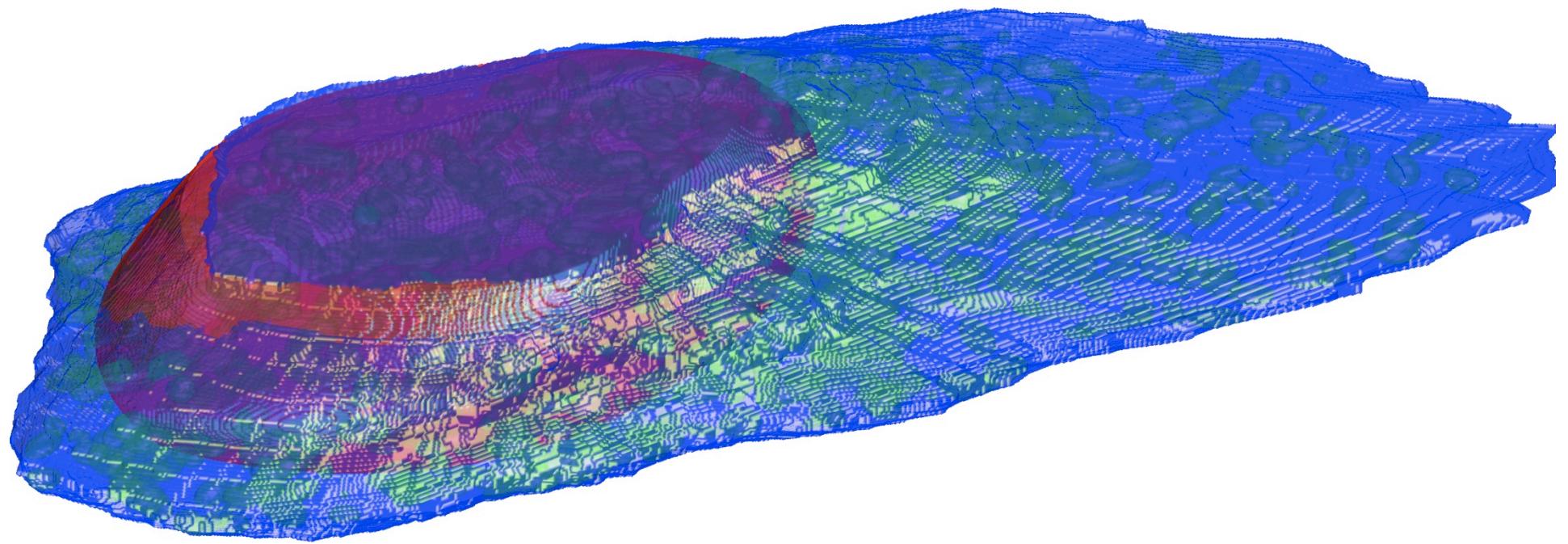


Angular
preference

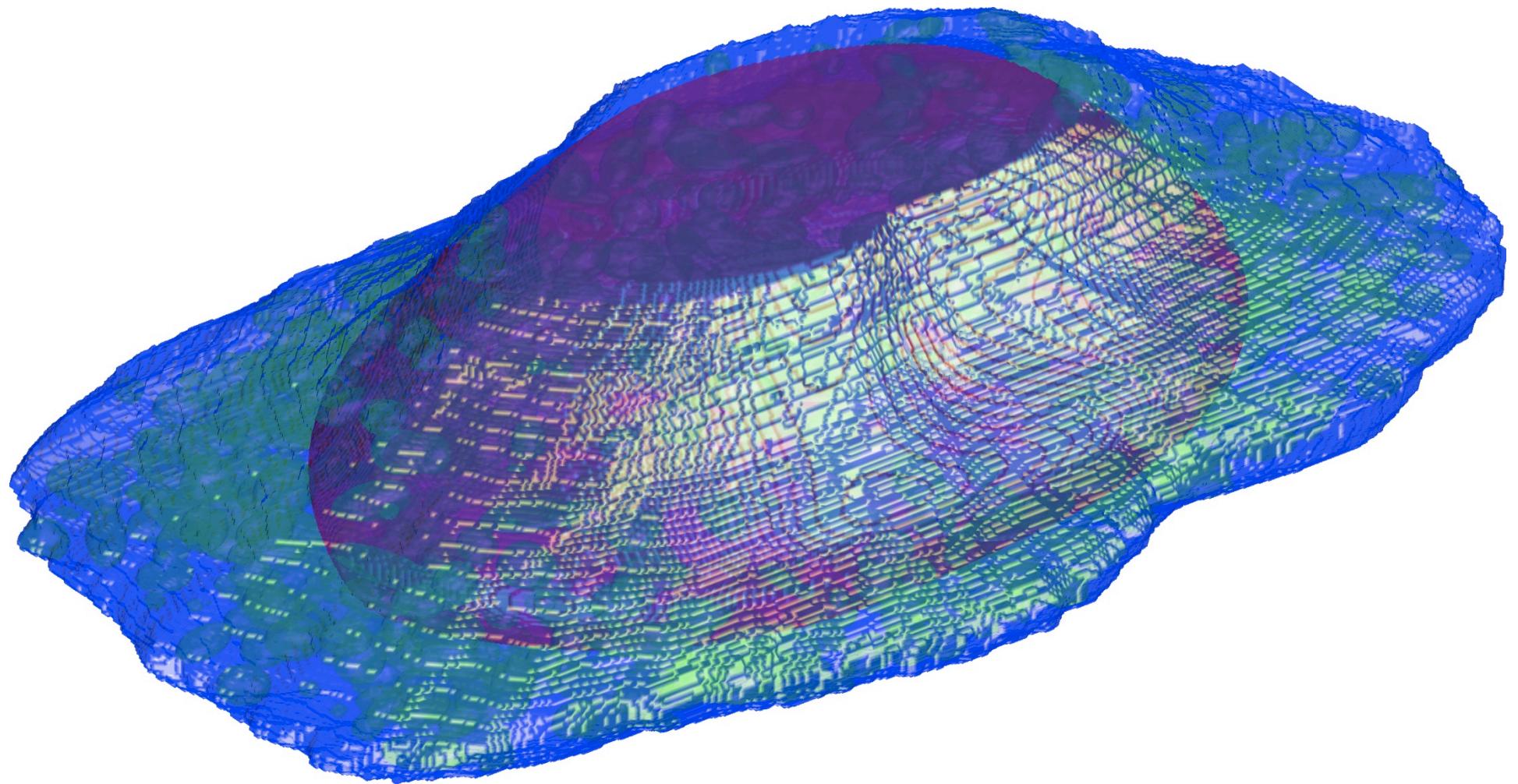
3D Endosome Movie



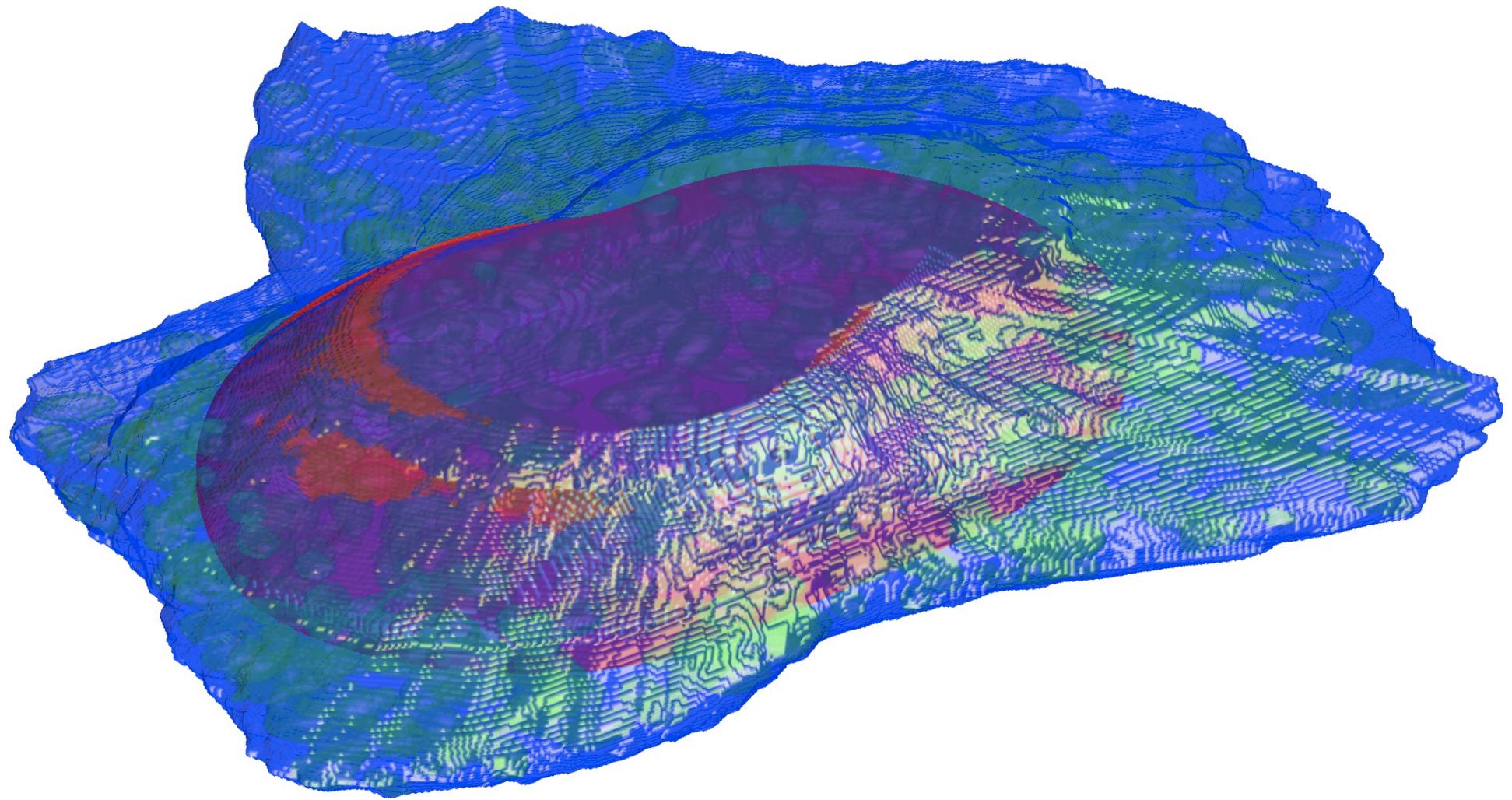
Endosomes



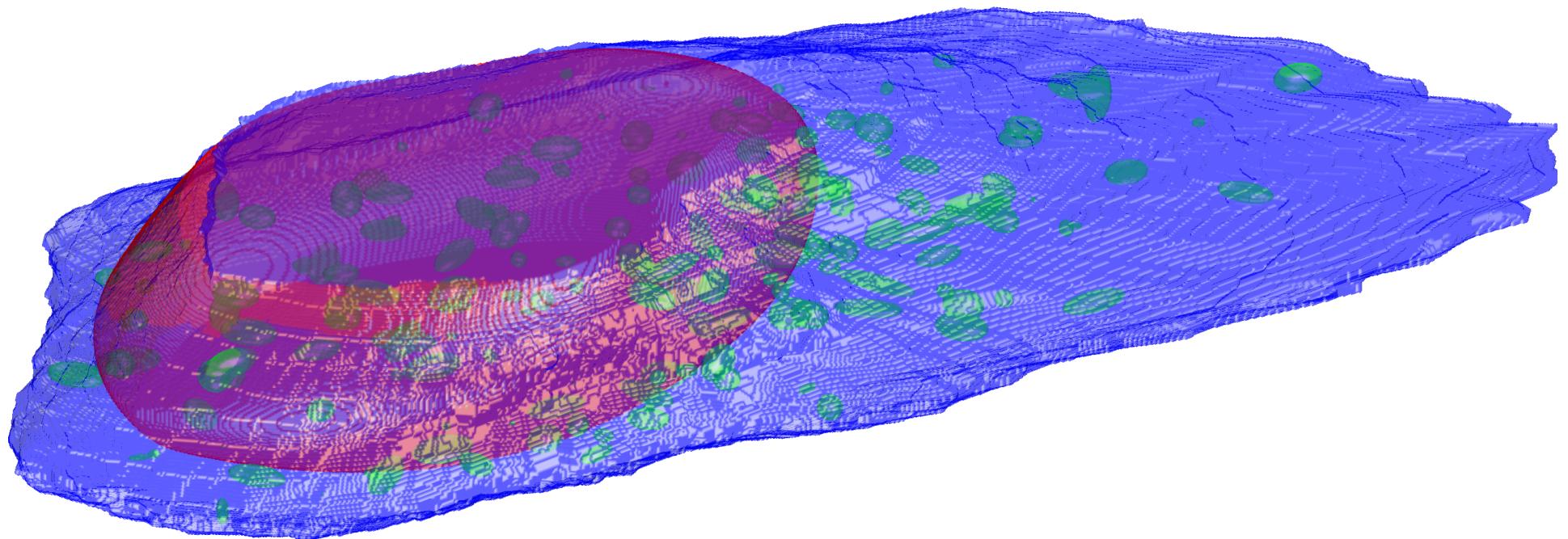
Endosomes



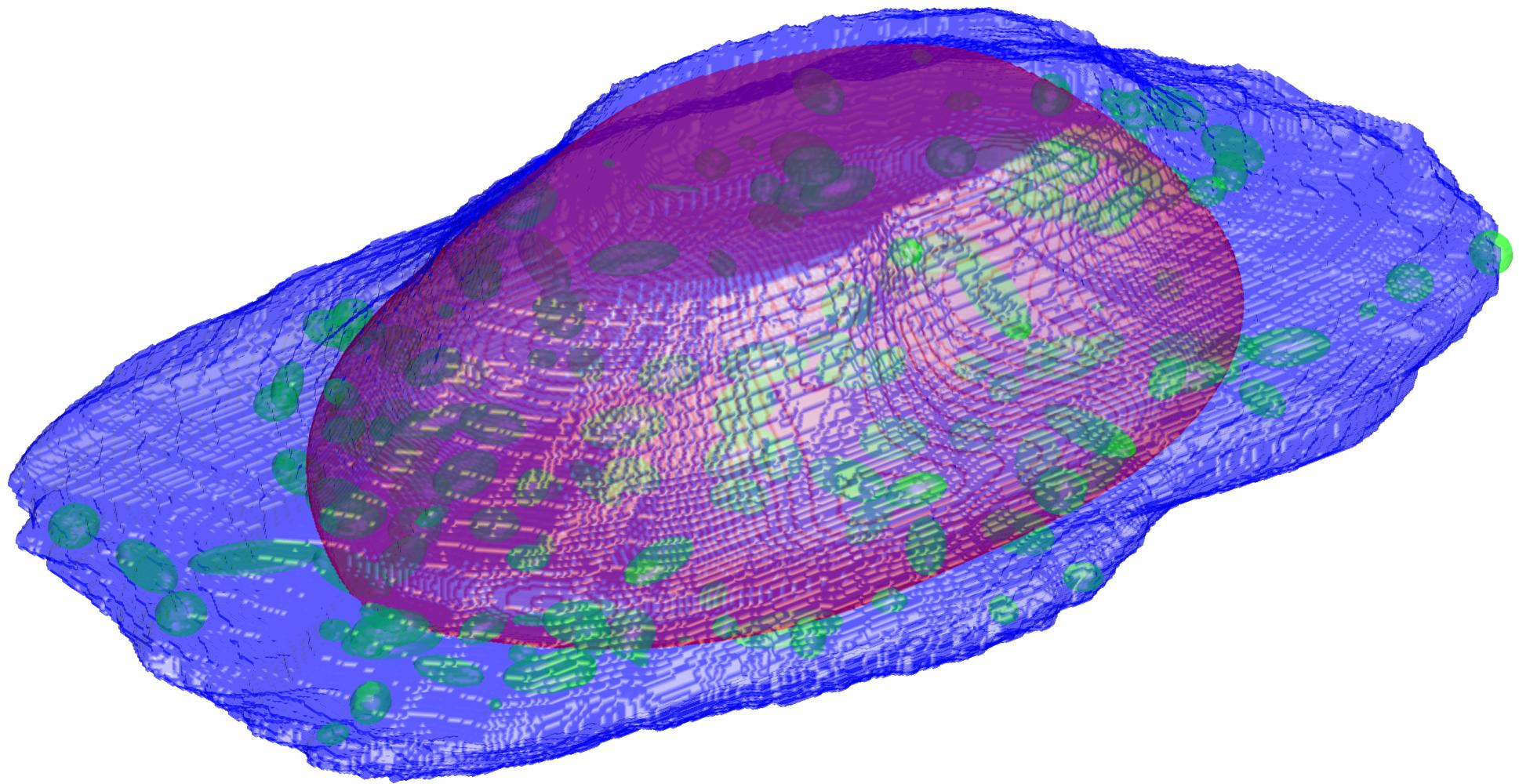
Endosomes



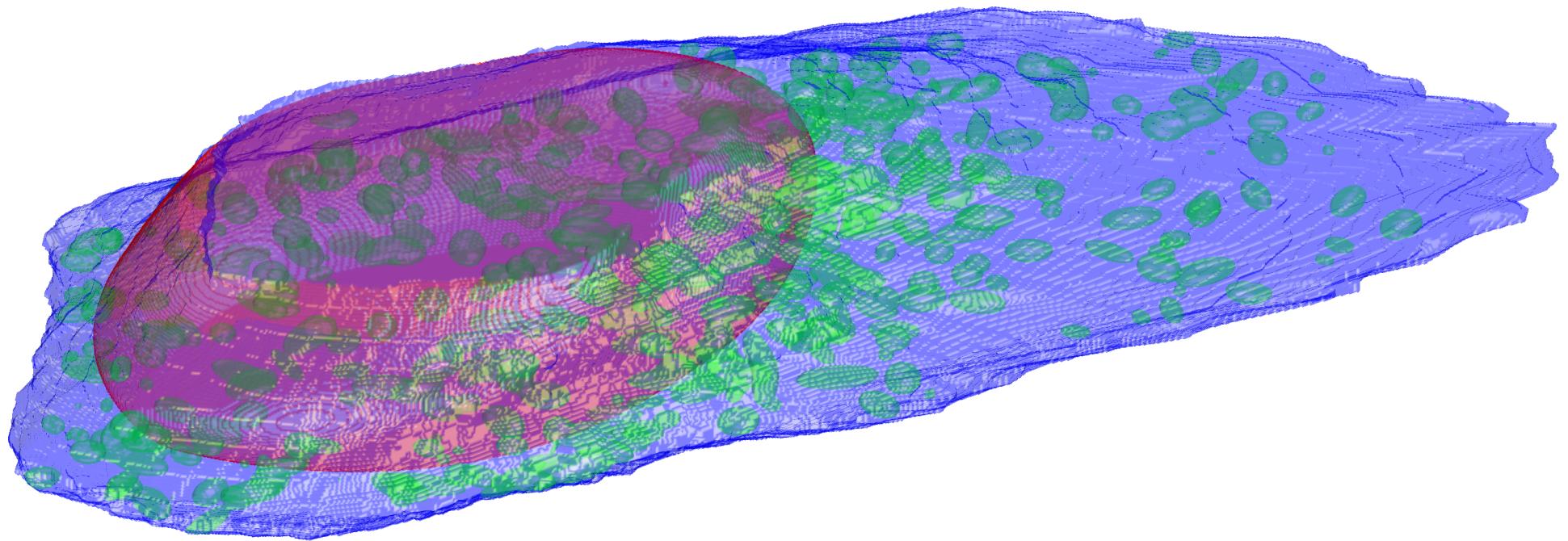
Lysosomes



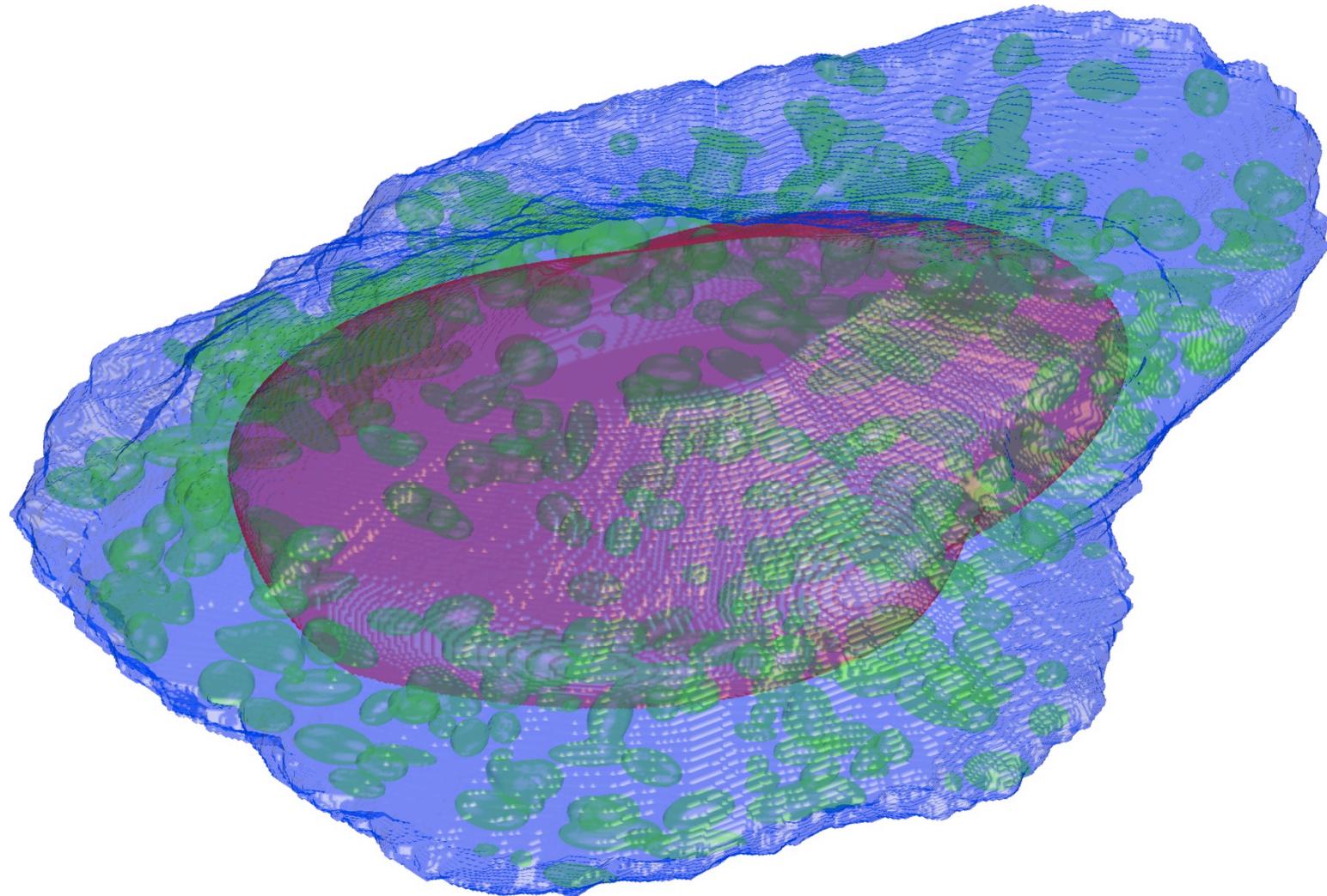
Lysosomes



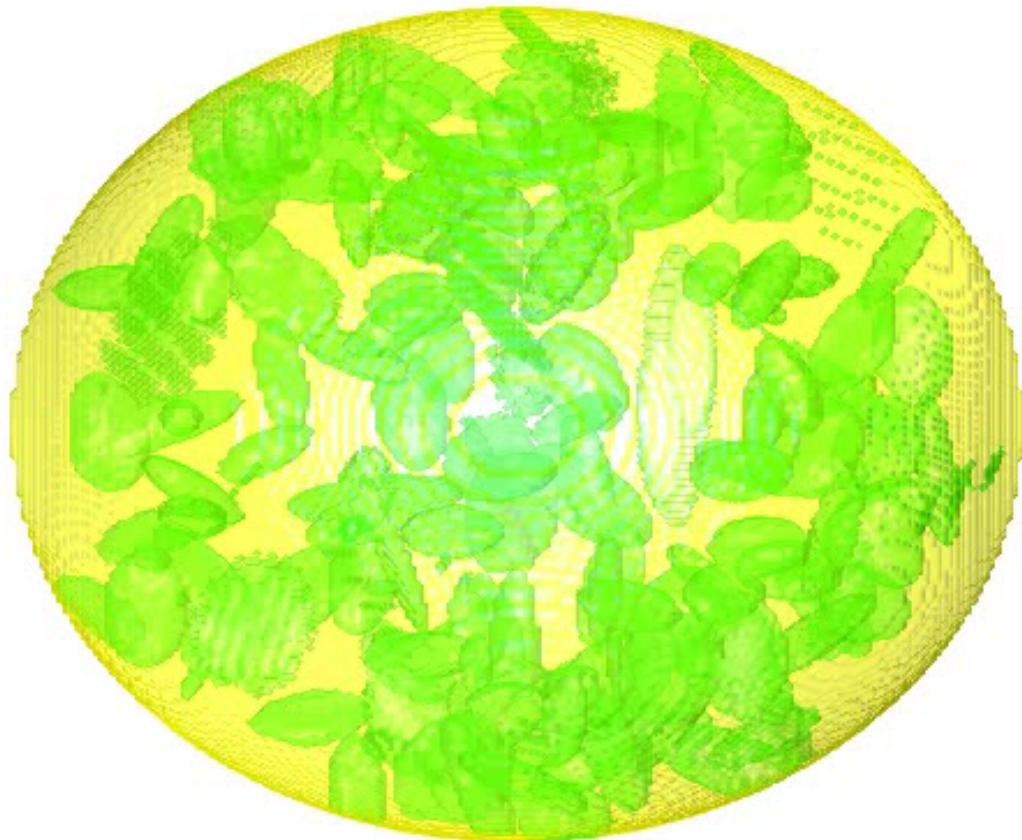
Mitochondria



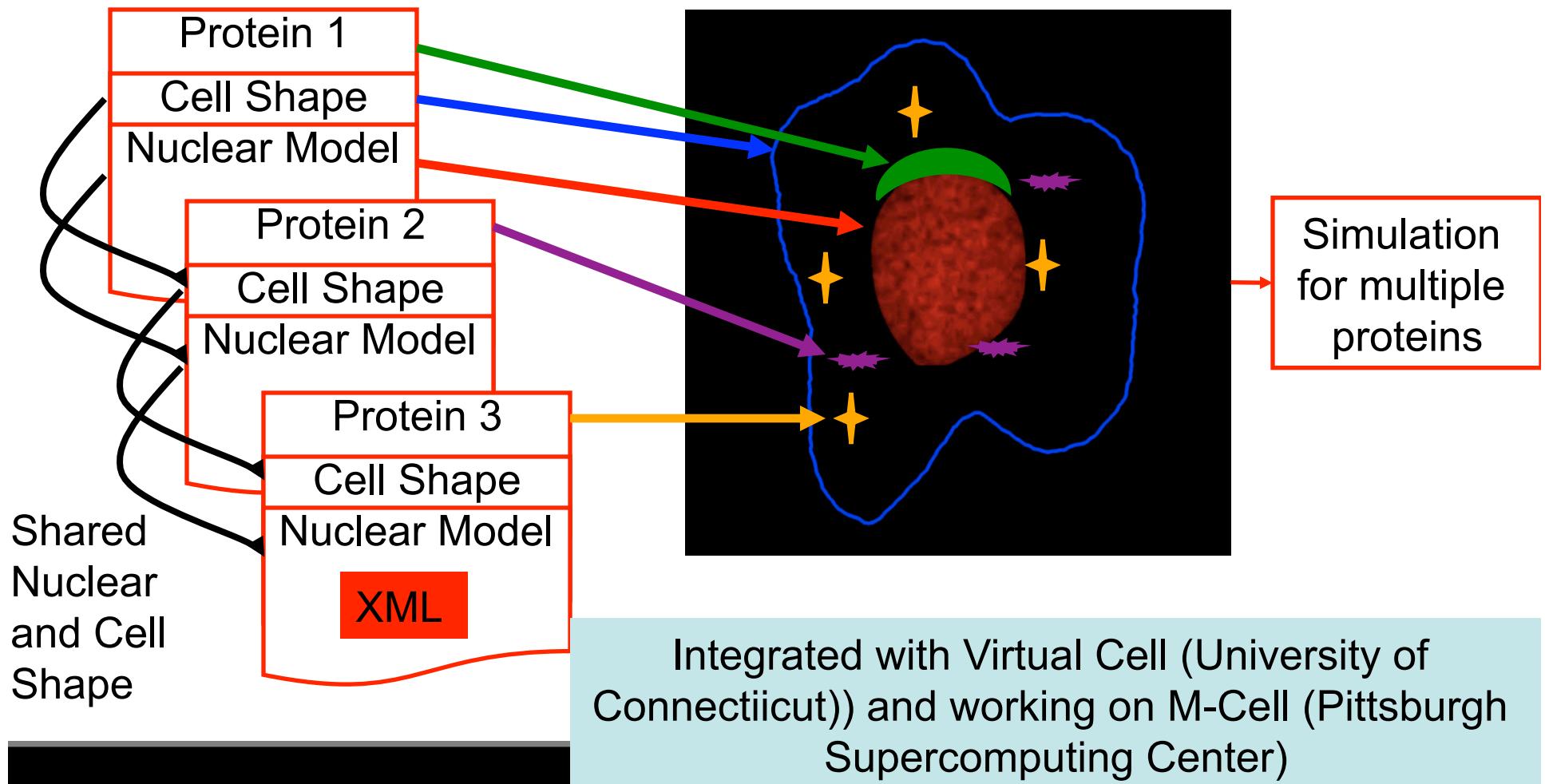
Mitochondria



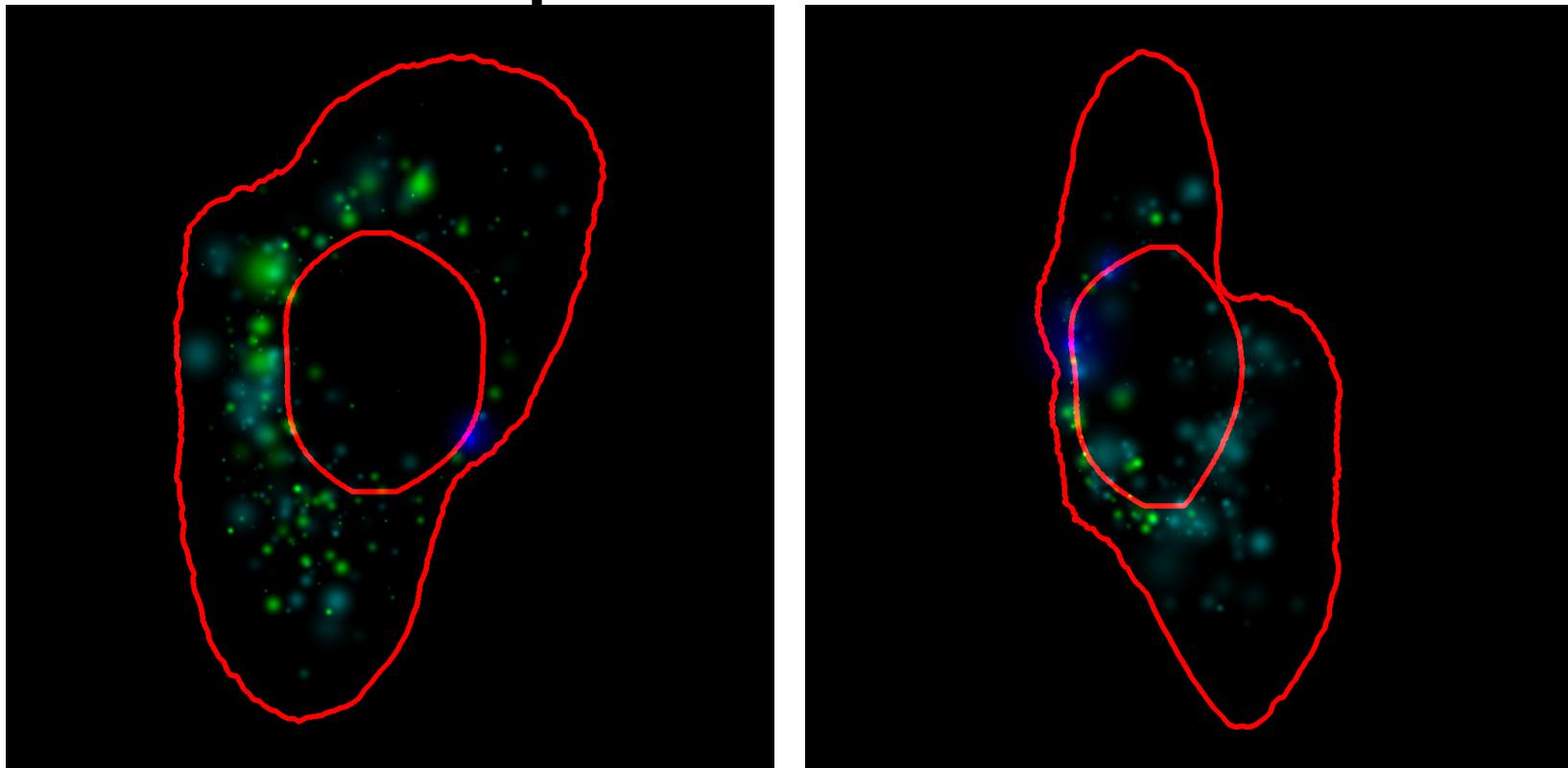
Chloroplasts



Combining Models for Cell Simulations



Example combination



Red = nuclear membrane, plasma membrane

Blue = Golgi

Green = Lysosomes

Cyan = Endosomes

Combining Patterns in One Cell

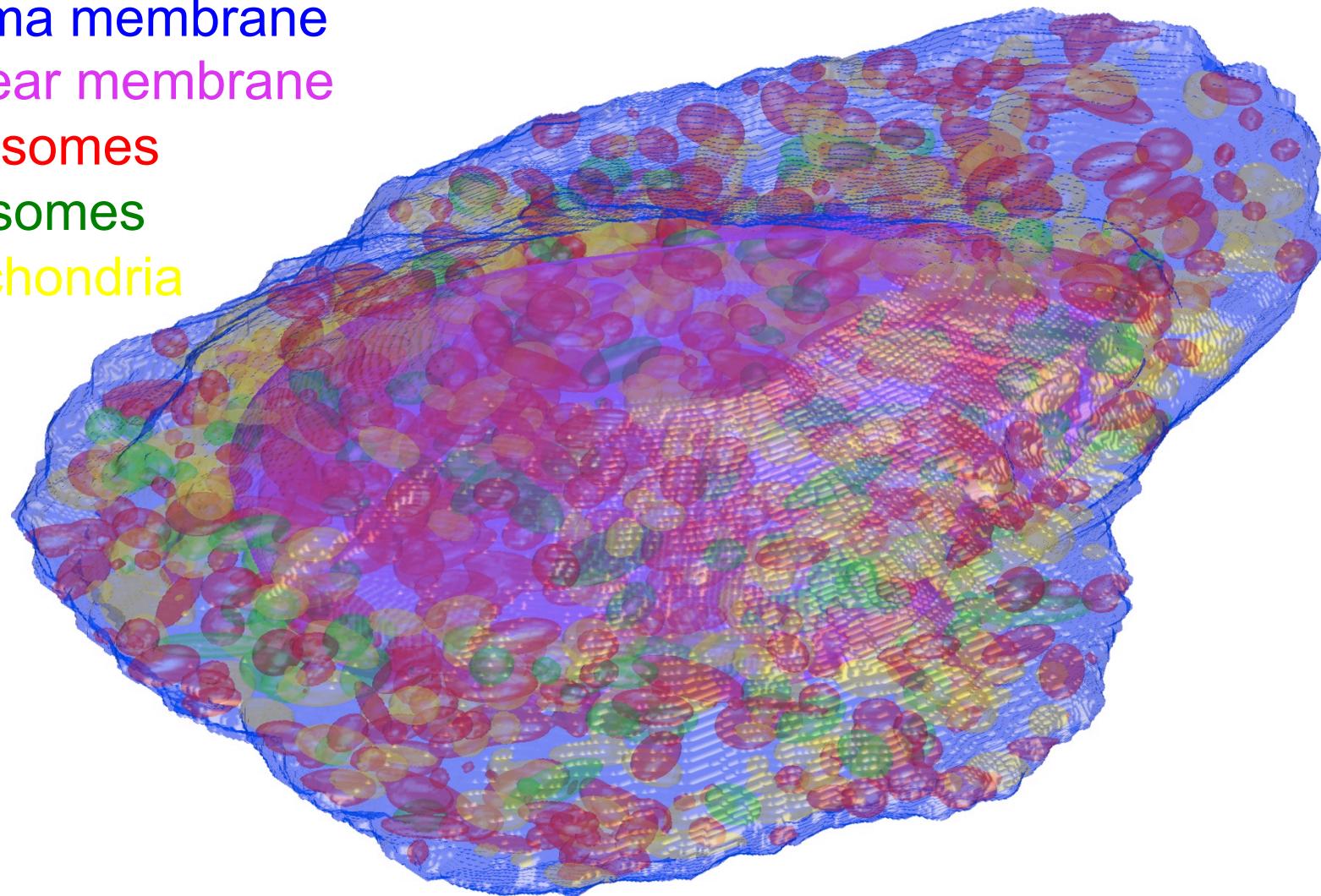
Plasma membrane

Nuclear membrane

Endosomes

Lysosomes

Mitochondria



MICROTUBULE DISTRIBUTIONS

Direct vs Indirect Generative Modeling

- Direct Method:
 - Estimate model parameters directly from images
 - Hard for microtubule networks
- Indirect Method:
 - Compare images with synthetic images generated from model to find parameters

Aabid Sharif



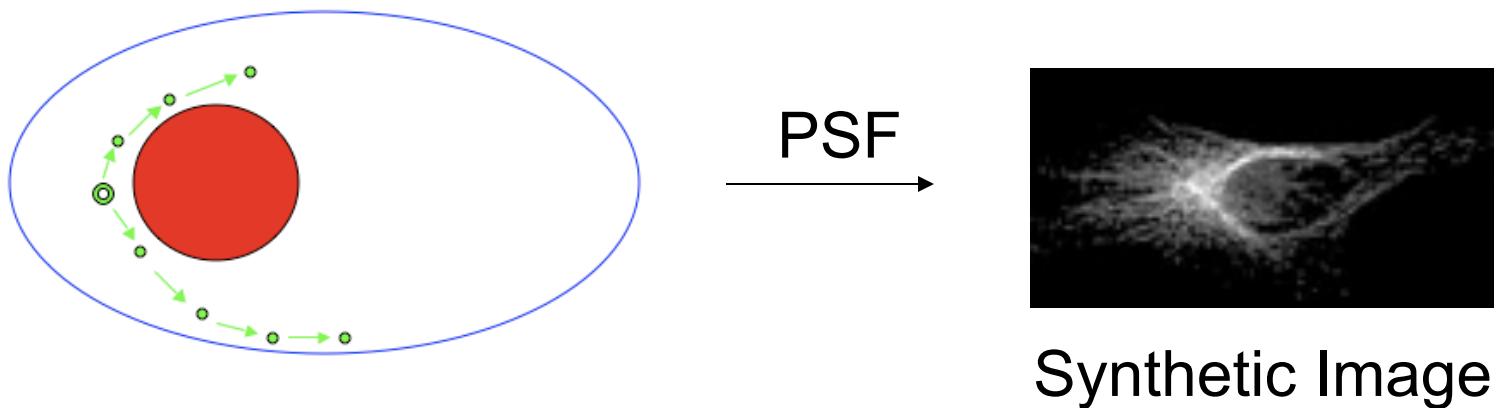
Gustavo Rohde



Microtubule Generative Model

- Inputs
 - Cell geometry
 - Nucleus
 - Cell membrane
 - Centrosome position
 - Microtubule growth simulation
 1. Number of microtubules (N)
 2. Mean of the Normal Distribution for length (μ)
 3. Standard Deviation of the length (σ)
 4. Collinearity

Converting Model to Image

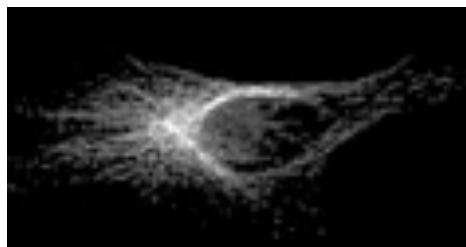


Convolve the microtubule points with point spread function G

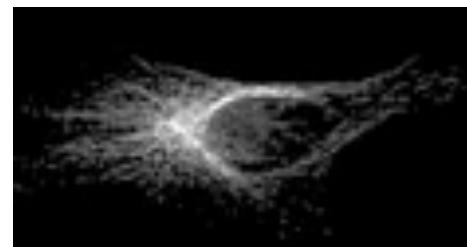
$$f(x) = \frac{1}{M} \sum_{i=0}^{M-1} G(x - X_i)$$

Finding Best Match

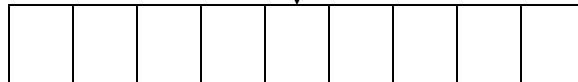
Synthetic Image



Real Image



Extract
Texture
Features



Feature Vector S

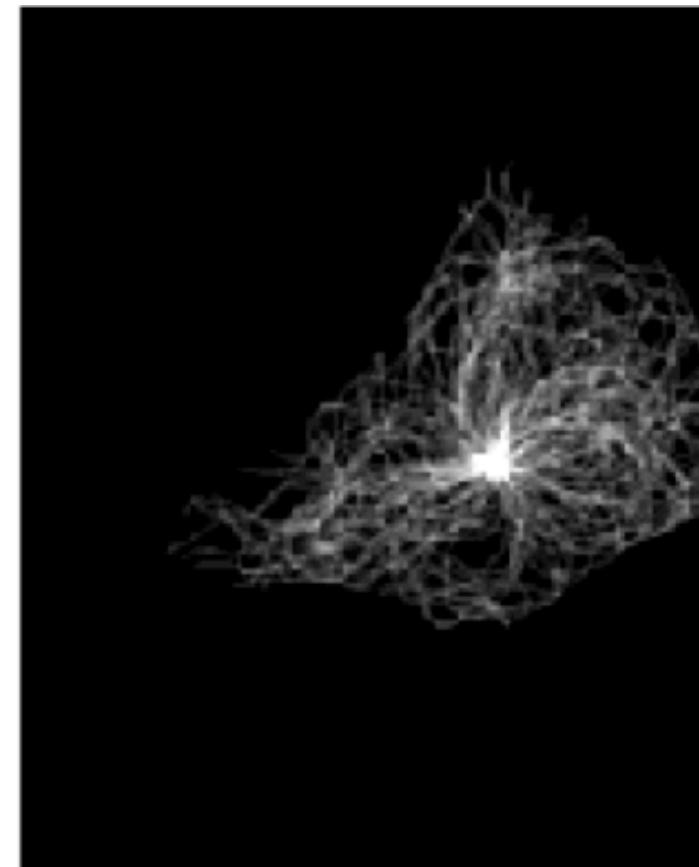
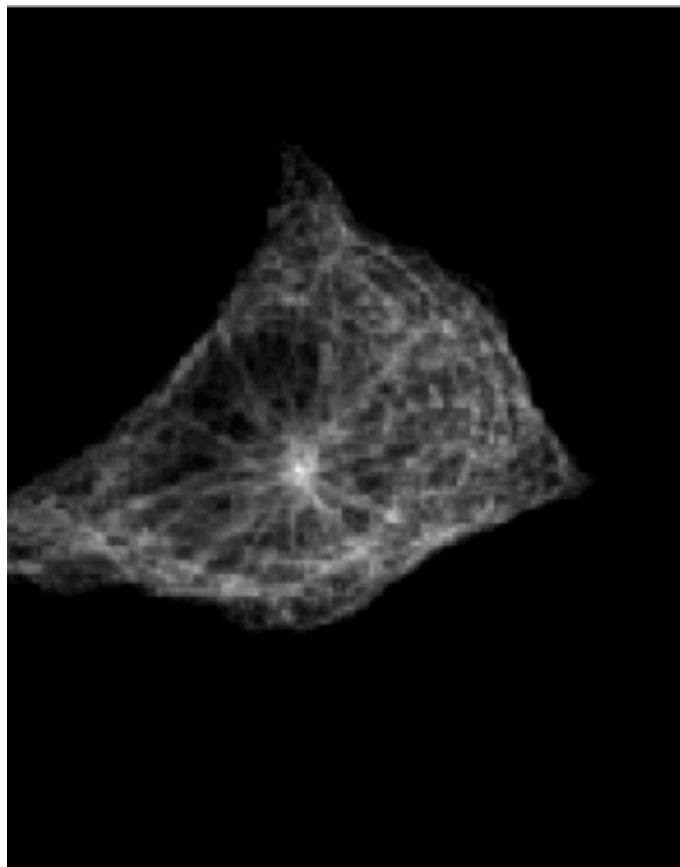
Extract
Texture
Features



Feature Vector R

Generate library of synthetic images for many parameter values
Find best match of real image using Mahalanobis distance

Real and Synthetic Tubulin Patterns



CellOrganizer GUI

- Simple interface for training models and generating instances
- Supports reading images from OMERO databases
- Saves models in SLML (Subcellular Location Markup Language)
- Saves instances as images (TIFF)

Issues for incorporation into systems models

- Generative models of geometry vs. instances of those models
- Geometries that capture molecule concentrations
- Distributions of molecule distributions over compartments

SBML spatial package Geometry Definition

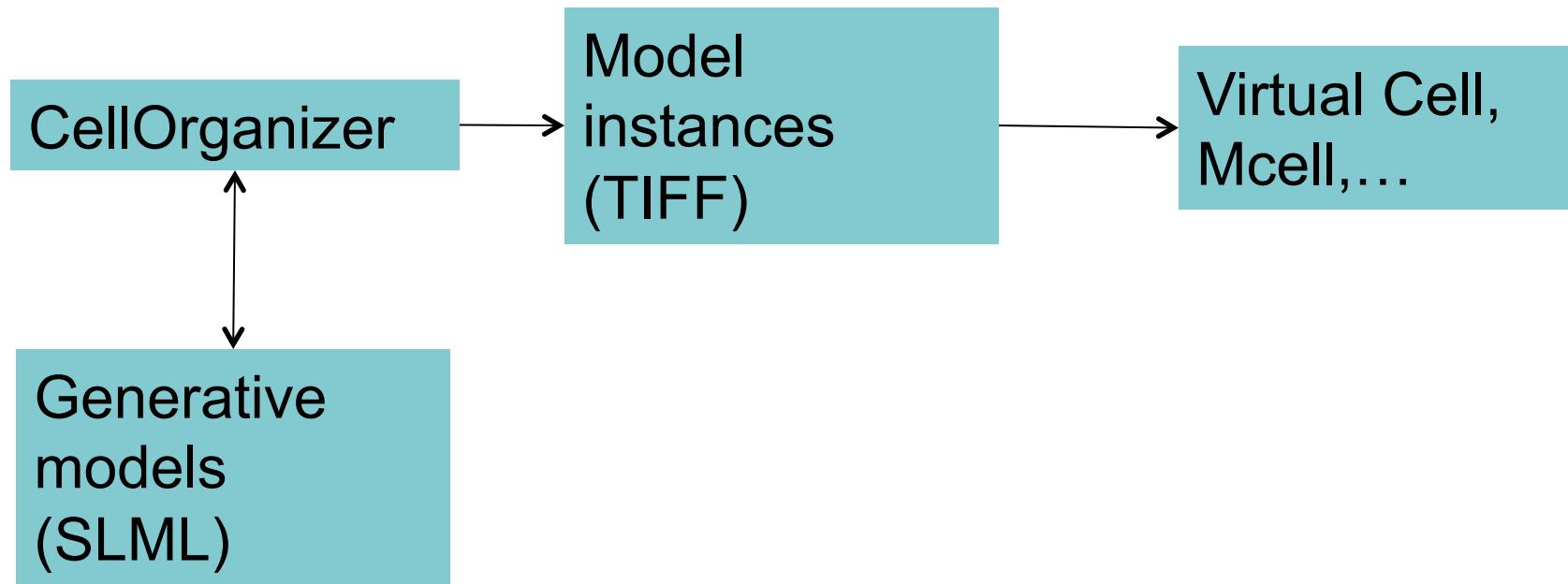
- **AnalyticGeometry**
- **SampledFieldGeometry**
- **CSGeometry (Constructive Solid)**
- **ParametricGeometry**

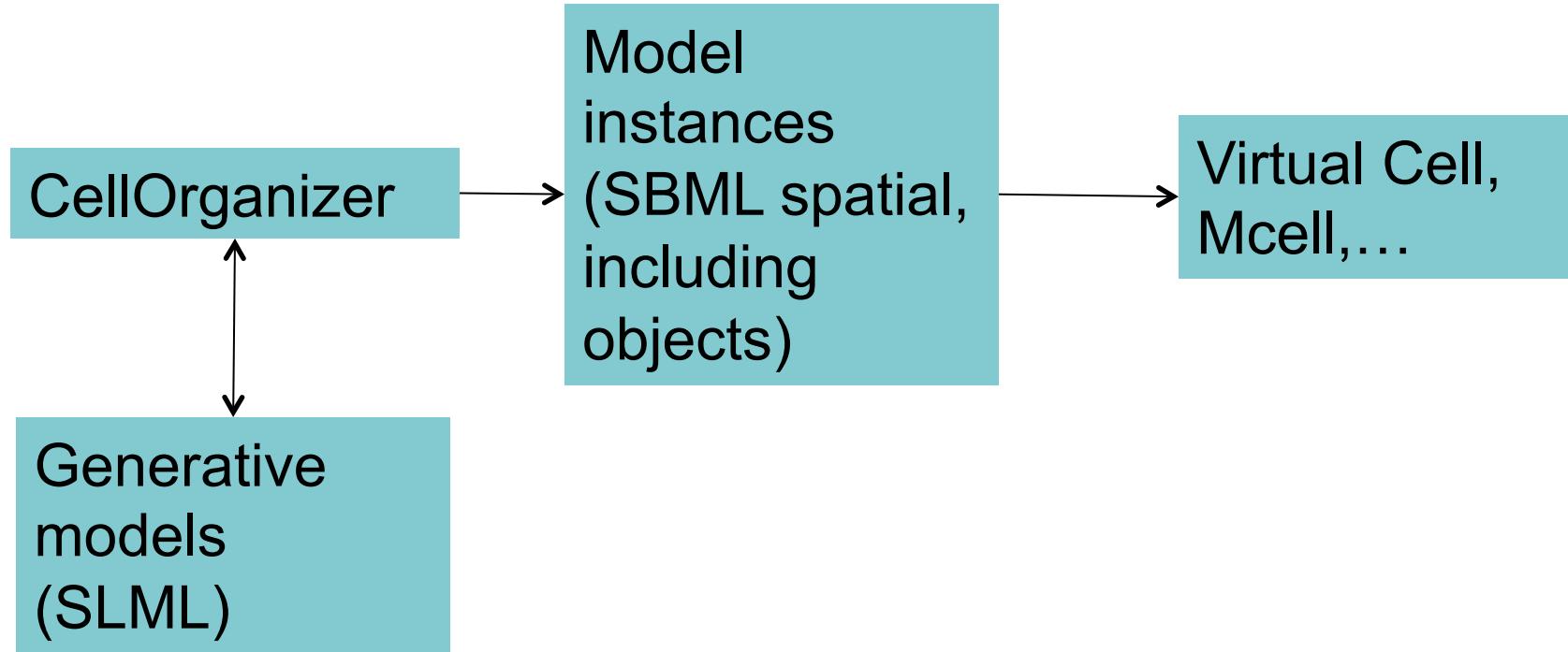
Current shape model instance types

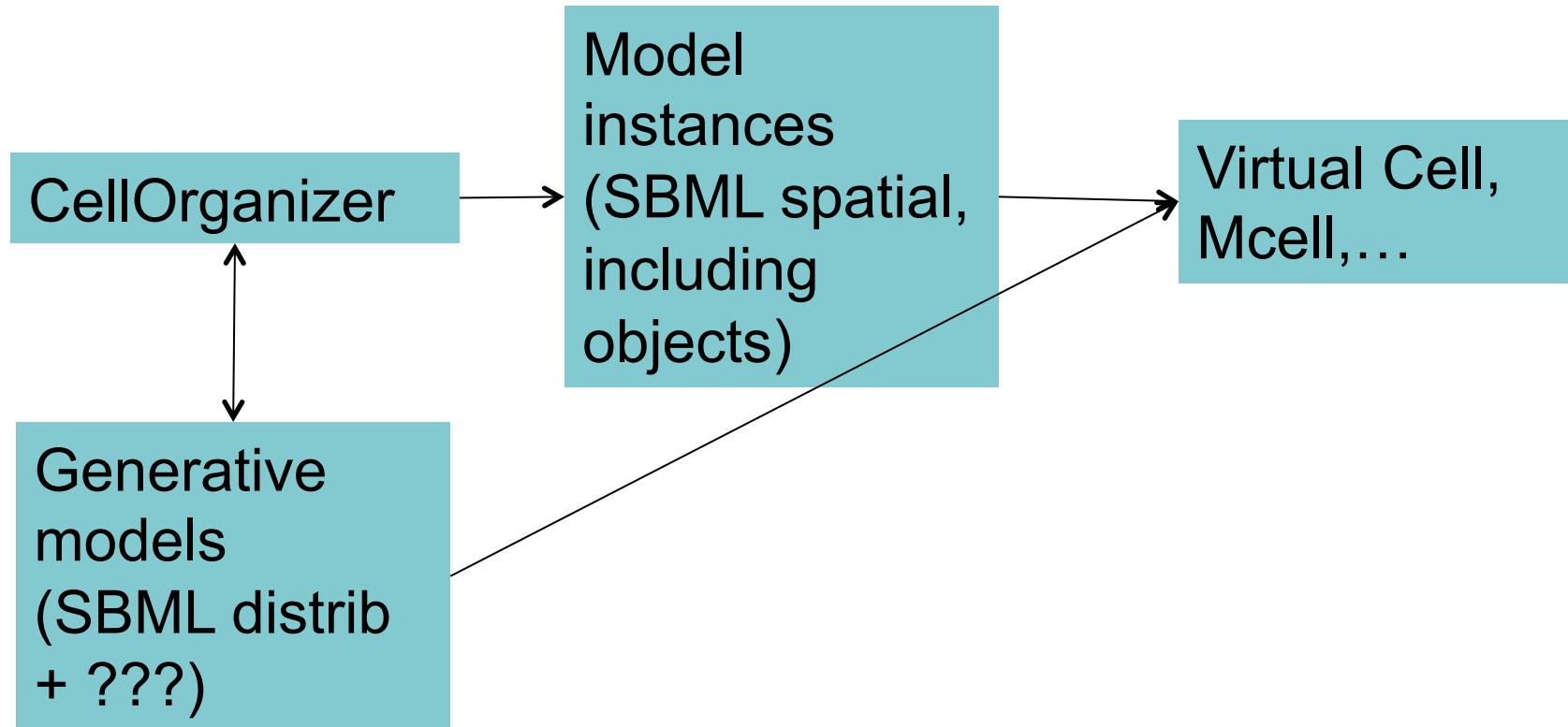
- Medial Axis Spline (2D)
- Cylindrical Spline Surface (3D)
- Cell/Nuclear Ratio (2D,3D) [conditional upon nuclear]
- Gaussian object mixture (2D, 3D)
- Microtubule growth model

Current shape generative model types

- Medial Axis Spline Multivariate Gaussian (2D)
- Cylindrical Spline Surface Principal Components (3D)
- LDDMM Principal Components (2D, 3D)
- Gaussian Object Axis Length Conditional Gaussian (2D, 3D)
- Normalized object probability density (2D, 3D)
[conditional upon nuclear and cell]
- Microtubule growth model parameter Gaussian (2D, 3D)







Conclusions

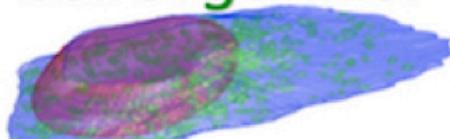
- Generative models provide means of representing location needed for *in silico* models of cell behavior
- Generalizable, transportable means for comparing and communicating effects of perturbagens across experiments, laboratories, cell types, and technologies

Conclusions

- Need new generation of cell simulation tools that can combine molecular behaviors with generative models of how cell organization changes over time
- Need ever better methods for building generative models directly from images
 - collaborators welcome!

Cellorganizer.org

CellOrganizer



Images ↔ Models

People

[Home](#)

[People](#)

[Publications](#)

Project leaders: Robert F. Murphy and Gustavo Rohde

[Downloads](#)

Major collaborators: Klaus Palme, Hauke Busch, Jörn Dengjel, Melanie Boerries, Christoph Wülfing

Current team members: Taraz Buck, Ivan Cao-Berg, Jieyue Li, Baek Hwon Cho

Past contributors: Ting Zhao, Tao Peng, Wei Wang, Aabid Sharif, Joshua Kangas, Jianwei Zhang, Alexander Dovzhenko, Rüdiger Trojok

RAY AND STEPHANIE LANE
Center for Computational Biology

Carnegie
Mellon
University



Support for *CellOrganizer* has been provided by grants GM075205 and GM090033 from the [National Institute of General Medical Sciences](#), by a ForschungPreis from the [Alexander von Humboldt Foundation](#), and by the [School of Life Sciences of the Freiburg Institute for Advanced Studies](#).



ALBERT-LUDWIGS-
UNIVERSITÄT FREIBURG



FRIAS
FREIBURG INSTITUTE
FOR ADVANCED STUDIES