

Interactive 3D Whole Cell Reconstruction of *Escherichia coli*

Group 31: Yuxuan Xie^{1*}, Winnie Shi^{1*}, Rohan Grover^{1*}, Liangyu Zhao¹, Ruoqing Cheng¹, Kritin Karkare,
Edward Catoiu¹, Nathan Mih¹, Brett Barbaro², Bernhard Ø. Palsson¹

¹University of California San Diego, Department of Bioengineering ²Department of Integrative Structural and Computational Biology, The Scripps Research Institute



*These authors contributed equally to this work.

Background and Motivation

3D visualizations of whole cells are valuable education and research tools that contribute to the understanding of structural and systems biology. The use of these reconstructions helps us:

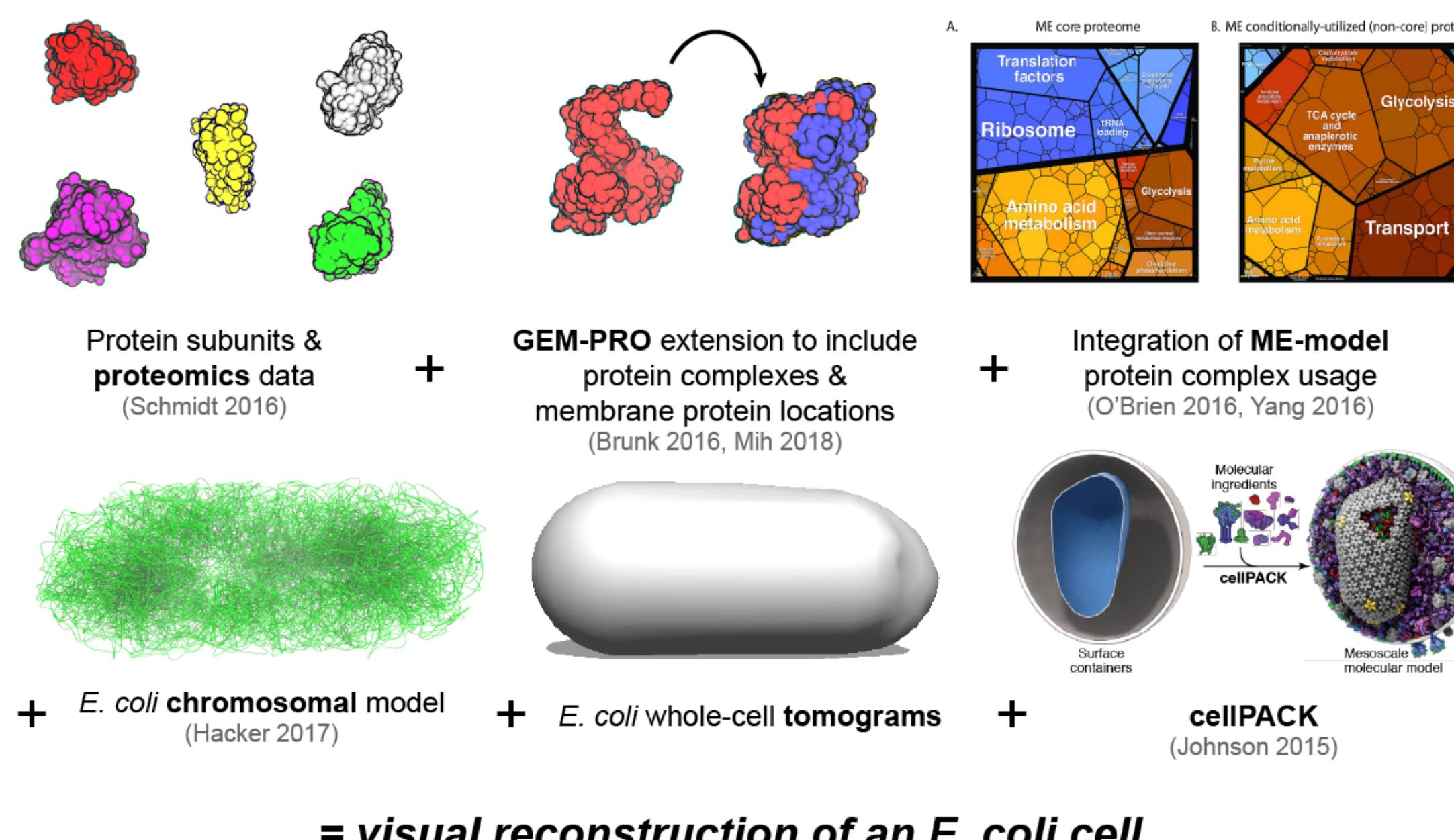
- identify missing and available information
- predict complex multi-network phenotypes
- provide insight into what future experiments should be conducted¹

There is a need for:

- **3D visual reconstructions at the scale of a whole single prokaryotic cell**
- **Organized compilation of information about *Escherichia coli***, a relatively large prokaryote, with a number of proteins, protein complexes, and metabolic networks
- **Structural detail and protein-protein interaction involved in complex formation**
- **Pipeline that allows for the integration of “omics” data** (genomic, proteomic, metabolomics, etc.) into such reconstruction

Design and Objectives

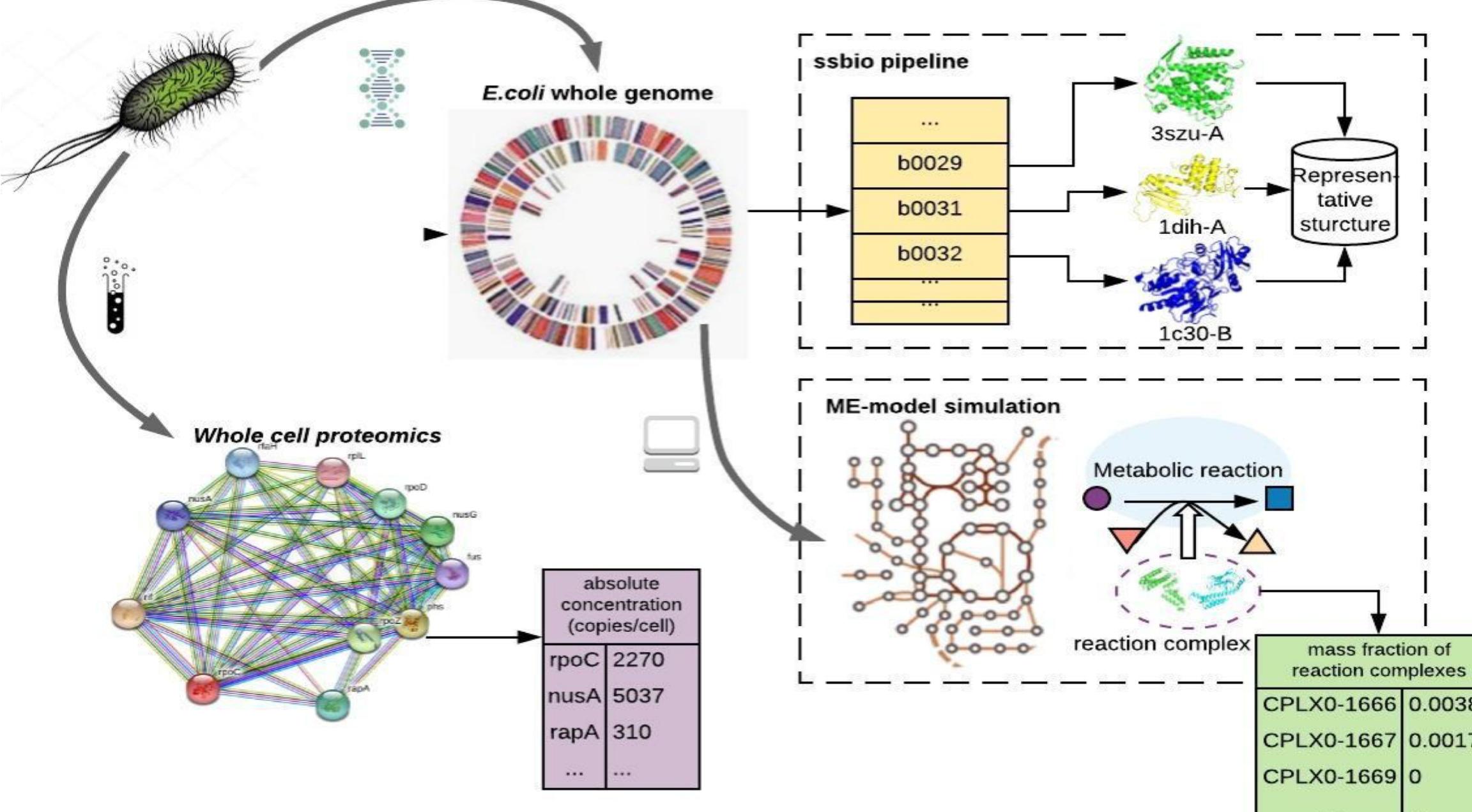
- Pipeline for visual reconstruction of *E. coli* and its expansion to other organisms
- Structural, large-scale, 3D visualization containing the requisite proteomic, genomic, and metabolic data
- Clean, fast, easy-to-use website for end users to interact with the model and access *E. coli* datasets



1. Proteomic and Genomic Raw Data Compilation

Proteins and Complexes

- Information about structures, localizations, and stoichiometry were determined from BiGG models, EcoCyc, SWISS-MODEL, and literature²
- Unknown structures were
 - determined computationally using homology models
 - or represented by spheres of similar volume
- Membrane orientation data used to align membrane proteins correctly
- Complex concentrations within the cell calculated using mass fraction ratios from the ME-model



Membranes

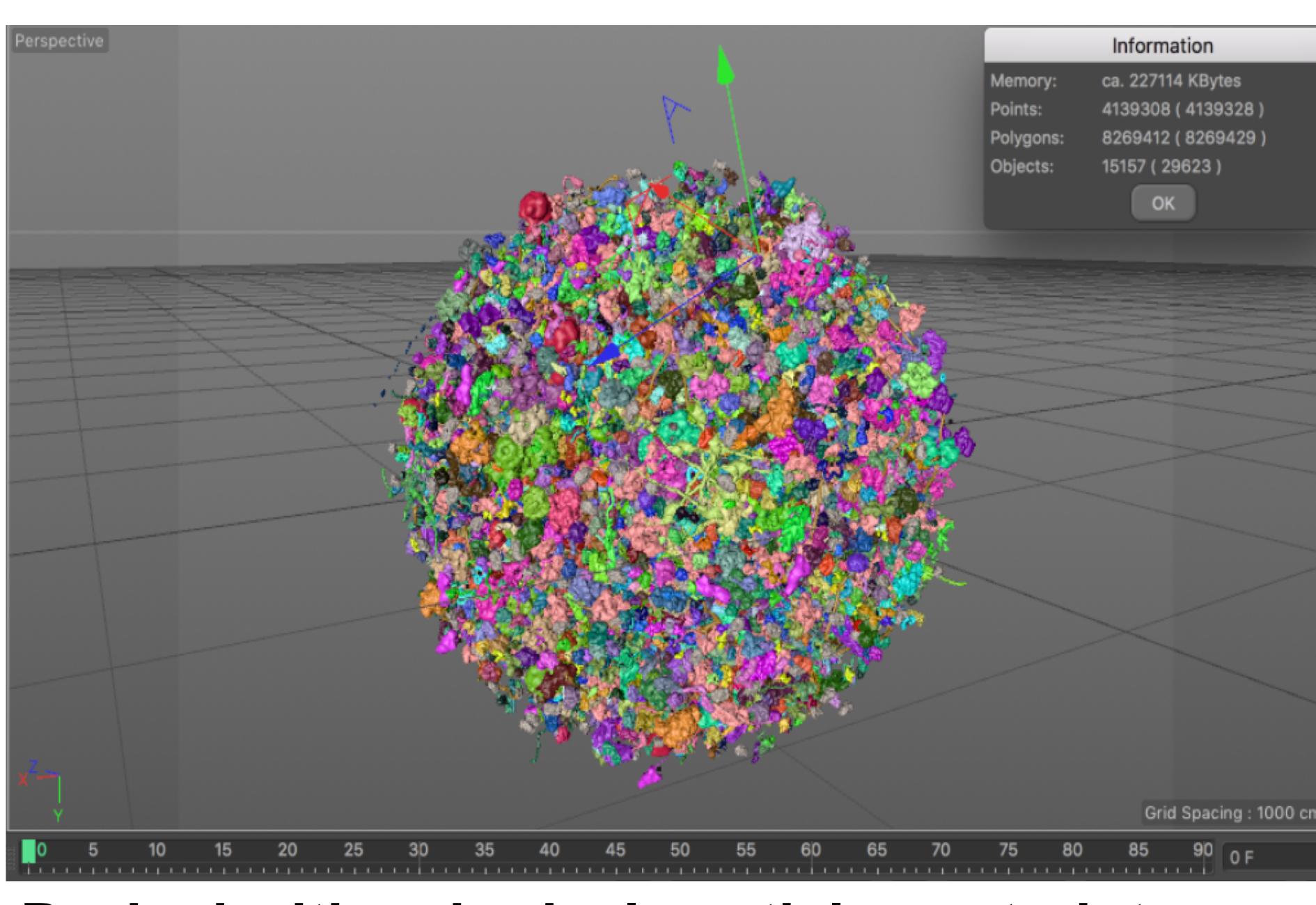
- Tomograms were used as reference for the overall shape of the membranes
- Volume and dimension of the cell determined from literature³

Chromosome

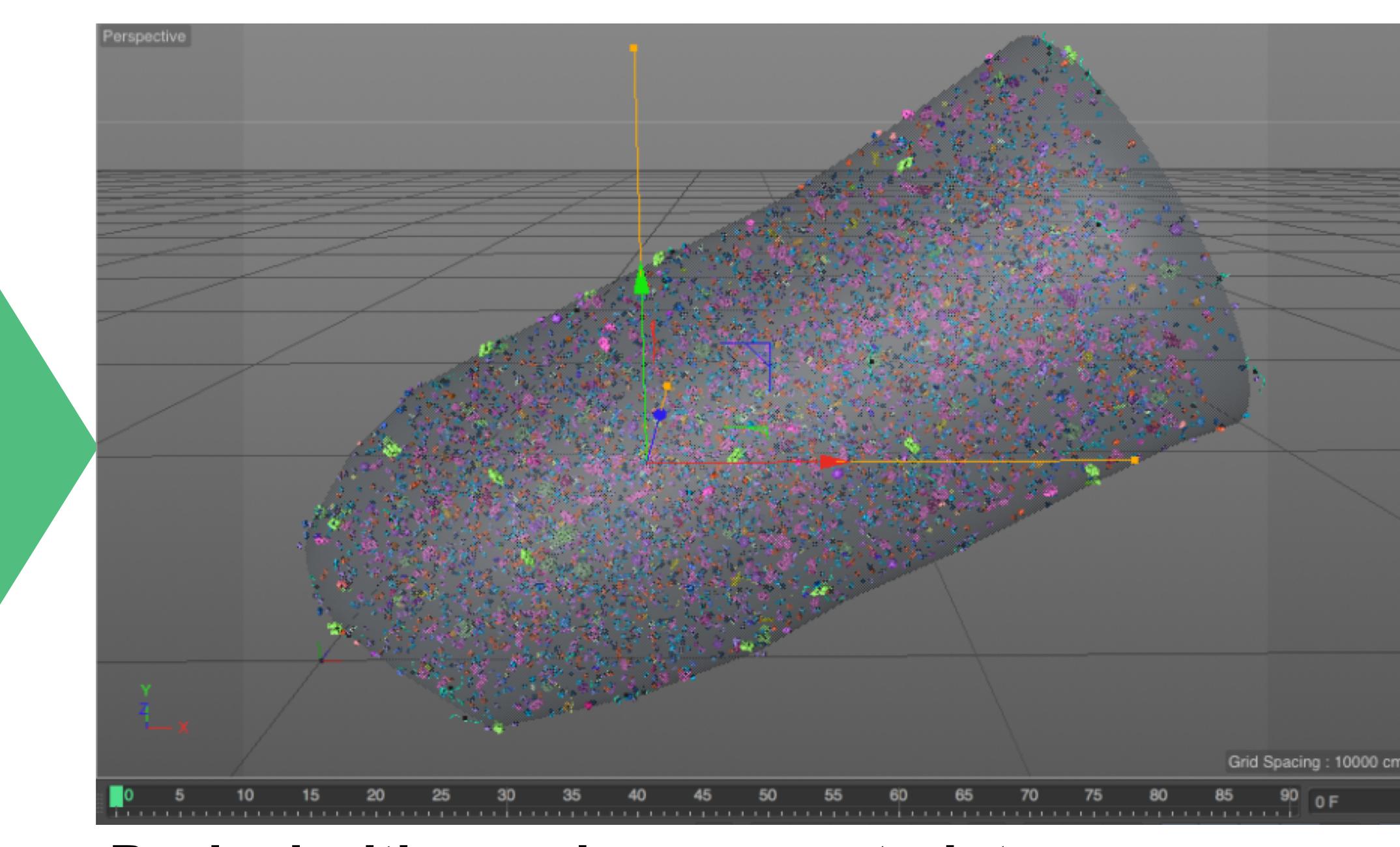
- 500 beads per base pair with midcell origin of replication
- Shared with us by Prof. Adrian Elcock and William Hacker

2. Packing with cellPACK

- cellPACK packs the proteins into the volume determined by the membrane
- Differences in protein concentration between the periplasm and cytoplasm were not considered



Packed with spherical spatial constraints



Packed with membrane constraints

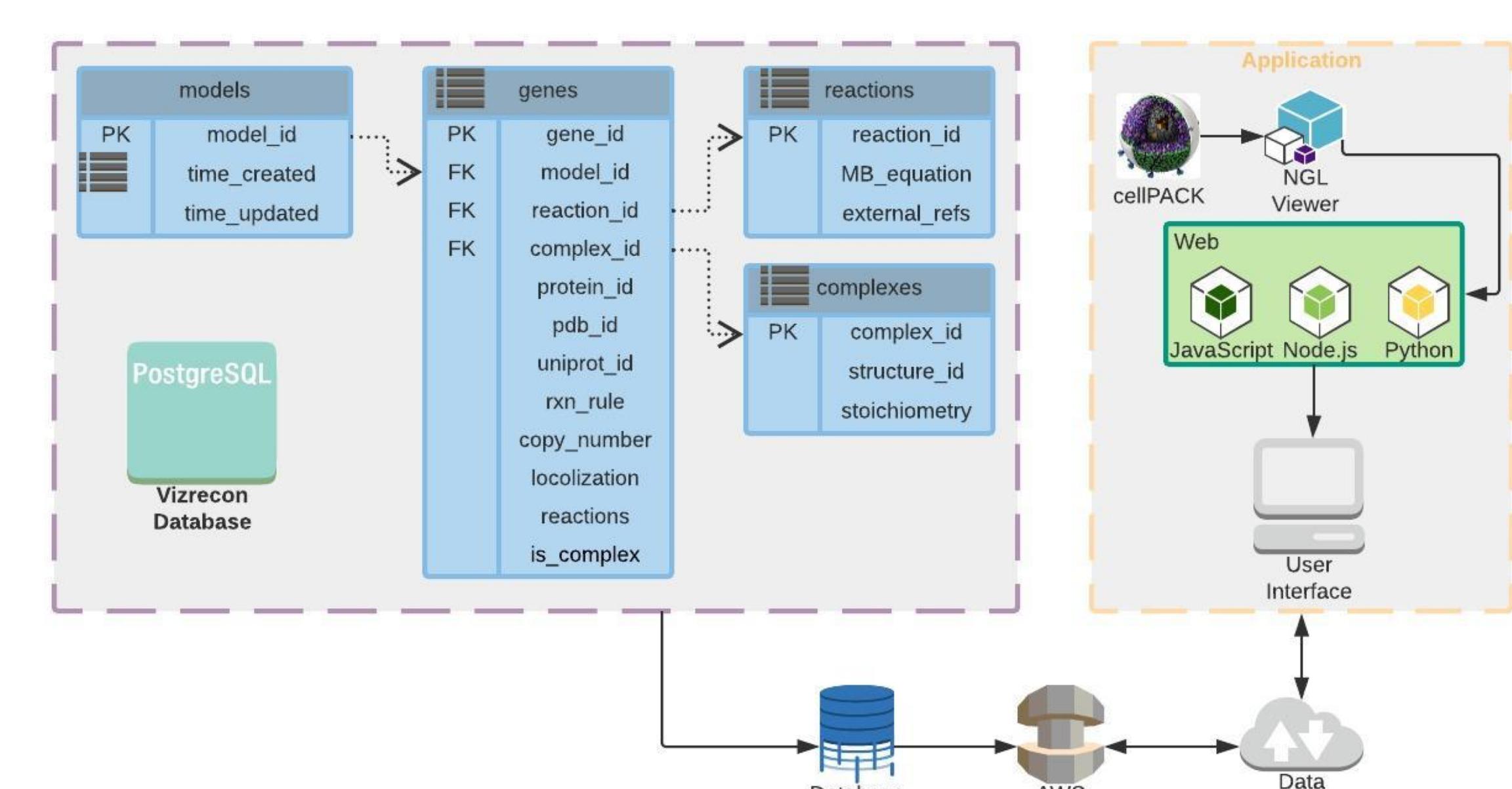
3. Display on Website

Backend

- Database is required to serve detailed data, better for searching than JSONs

Frontend

- NGL adaptations for added functionality
- User is able to access information stored in database by searching for specific proteins or clicking on a protein in the model



Conclusions

- Large size of *E. coli* cell remains a computational challenge
- Use points on NGL to represent proteins to reduce loading speed
- For the future:
 - Update with new experimental data to increase accuracy
 - Scale the pipeline to more organisms
 - Incorporate simulation results from to visualize changes in cell environment in response to various perturbations
 - Better represent interactions within the cell

Acknowledgements & References

1. Carrera, J. et al. Why Build Whole Cell Models?. *Trends Cell Biol.* **25**, 719 (2015)
2. Schmidt, A. et al. The quantitative and condition-dependent *E. coli* proteome. *Nat. Biotechnol.* **34**, 104-110 (2016)
3. Volkmer, B. et al. Condition-Dependent Cell Volume and Concentration of *E. coli* to Facilitate Data Conversion for Systems Biology Modeling. *PLoS One.* **6**, e23126 (2011)



SYSTEMS BIOLOGY
RESEARCH GROUP

