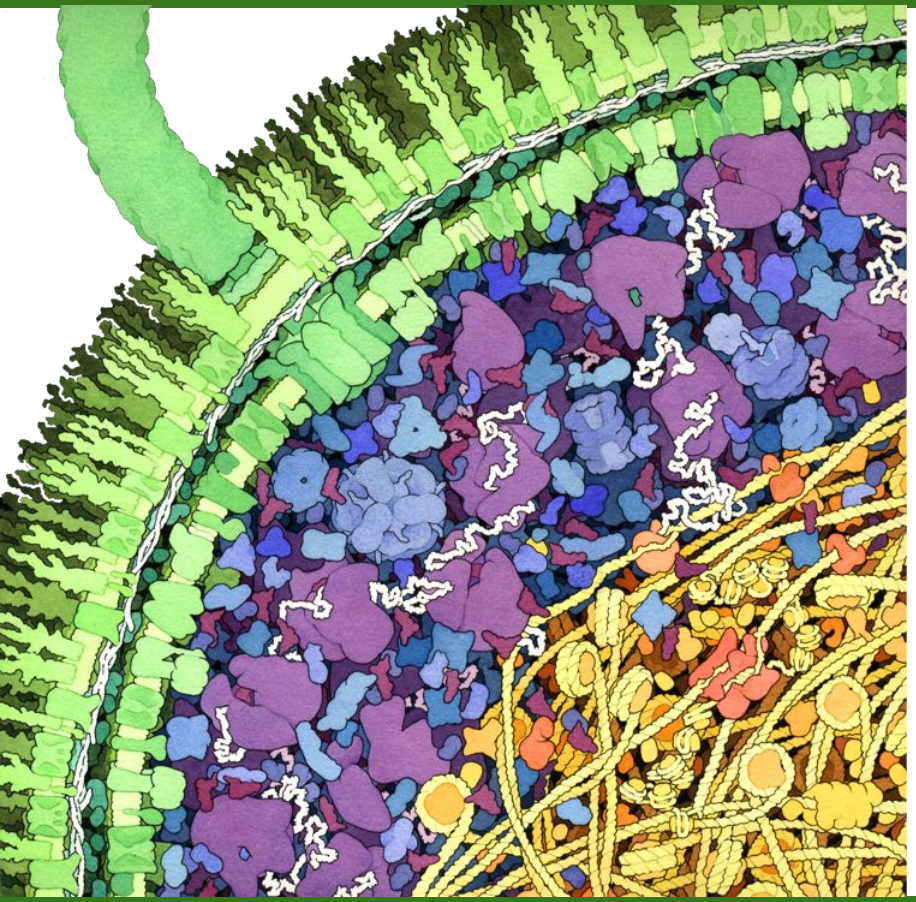
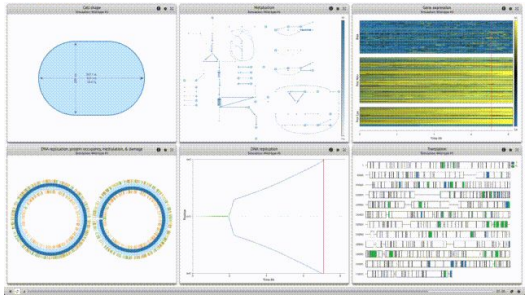


Group 31: An interactive, 3D visual reconstruction of a whole *E. coli* cell

Rohan Grover, Winnie Shi, Yuxuan Richard Xie

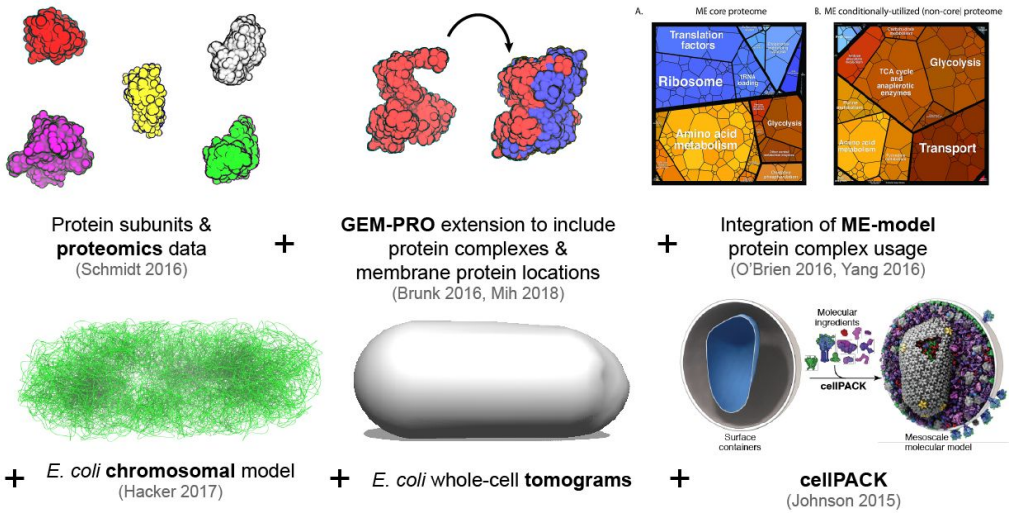
Whole Cell Models



- Help us understand:
  - Metabolic networks and expression networks
  - Relative concentrations and interactions of proteins
- Build more complex models in the future
- They are cool

Design Objectives

- Standardized pipelines for 3D reconstructions of *E. coli* and more
- Integrate high quality structural data along with:
  - Genomics
  - Proteomics
  - Metabolomics
- Interactive platform that enables easy search and manipulation

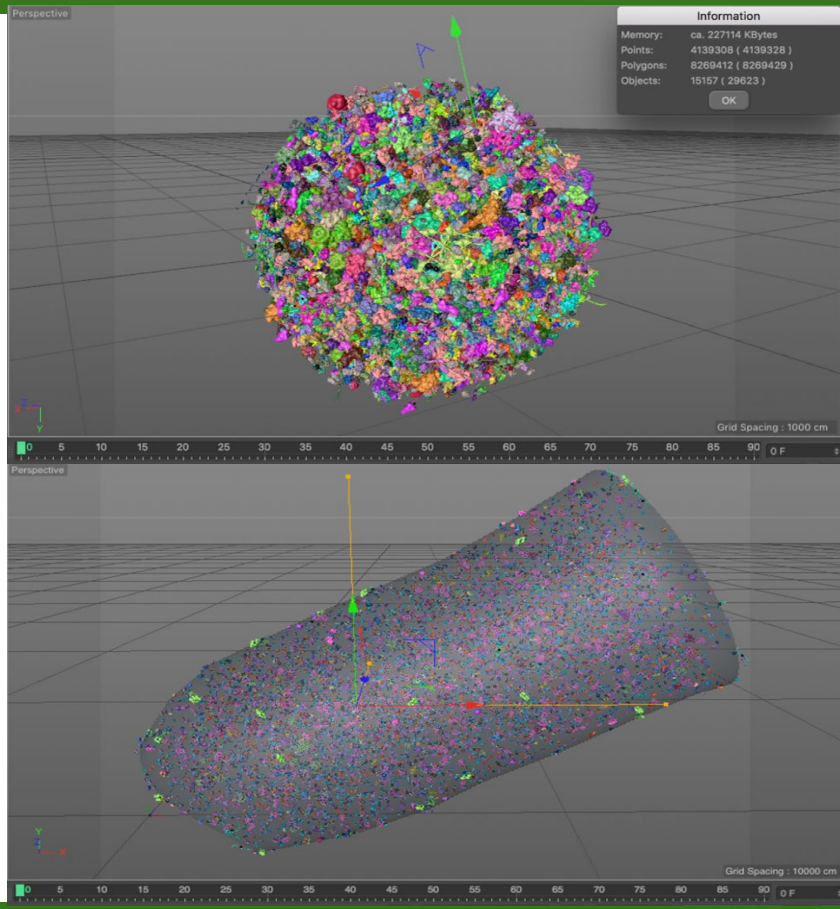


= visual reconstruction of an E. coli cell

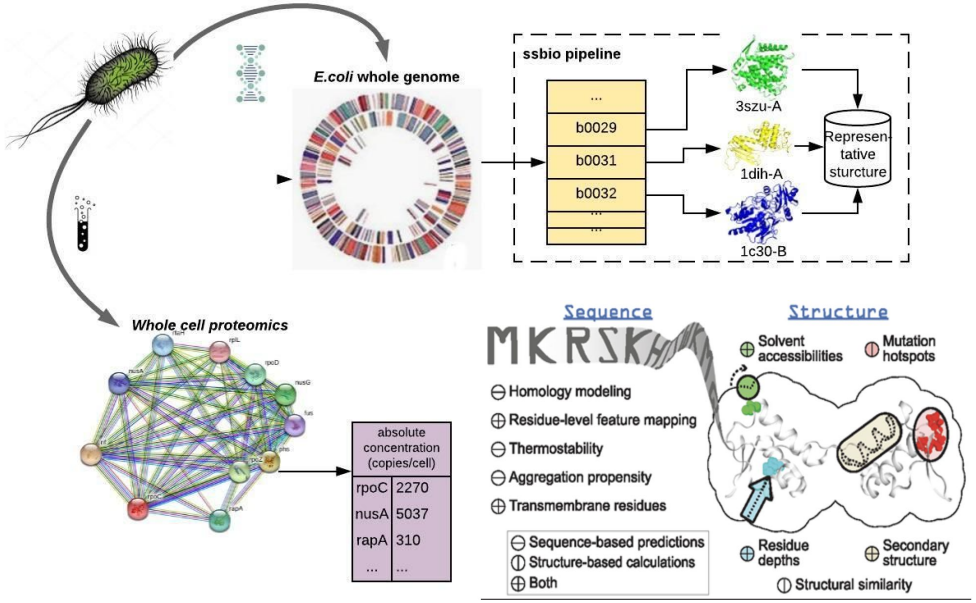
How do you curate and build a whole cell model of *E. coli*?

Prototype 1

Testing the feasibility of building a whole cell model for *E. coli*



Proteins and Structures



- ssbio (Python framework for **s**tructural **s**ystems **bio**logy): calculate and map structural information at genome scale
- Experimental protein concentrations from literature
- Building a repository of high quality protein structures with extensive annotations

What's in prototype 1?

Not protein complexes, just the protein components themselves
(D, E in the figure, randomly distributed throughout the cell)

Protein

Structure

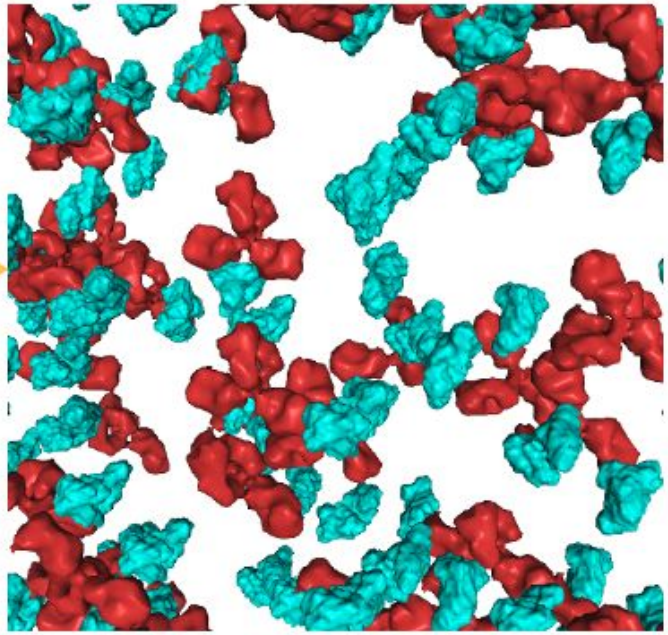
See:

- ssbio
- PDB
- Homology modeling

Proteomics measurements

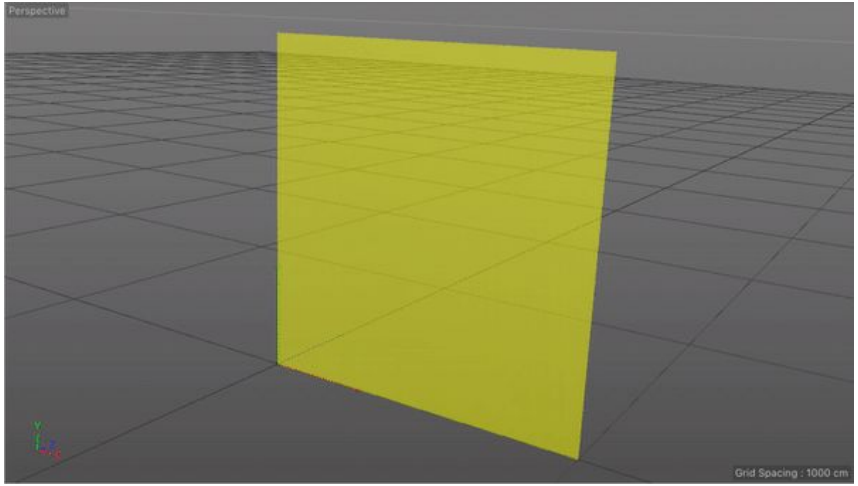
Copy number = 445

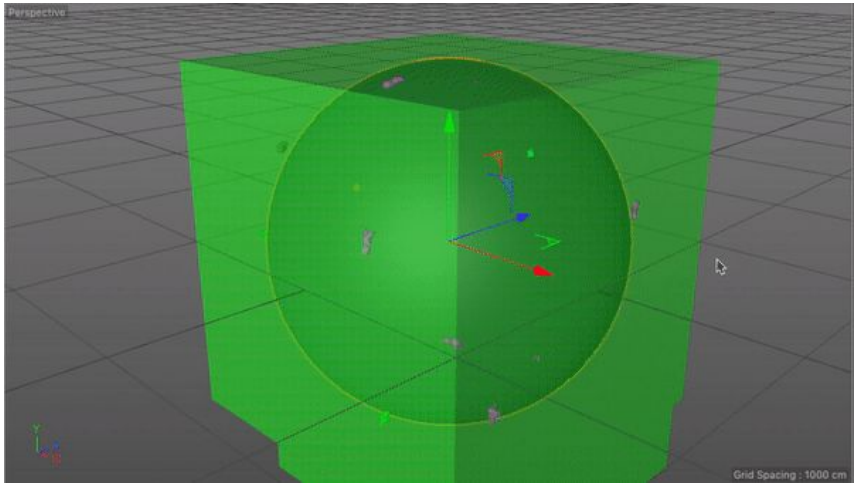
Concentration = Molarity = `(445/Avogadro's) / compartment_volume`



Gene	UniProt	NAME	Description	PDB	Structure_sourc	MOLARITY	Copy number	LOCALIZATION	INCLUDE
b0001	P0AD86	thrL	thr operon leader peptide					Cytosol	
b0002	P00561	thrA	Bifunctional aspartokinase	b0002_1_462_3c20-A_clean	SWISS-MODEL	2.59E-06	4375	Cytosol	x

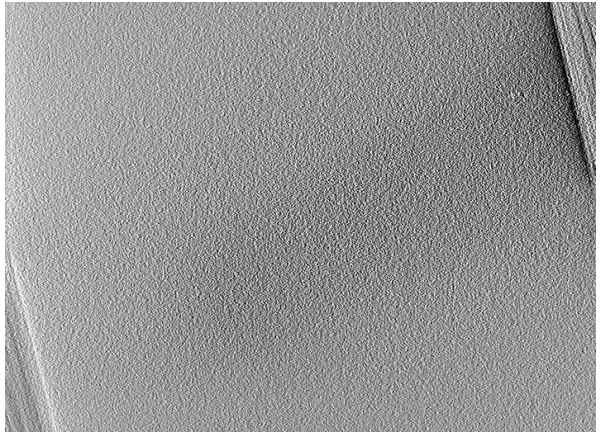
Integration Using cellPACK

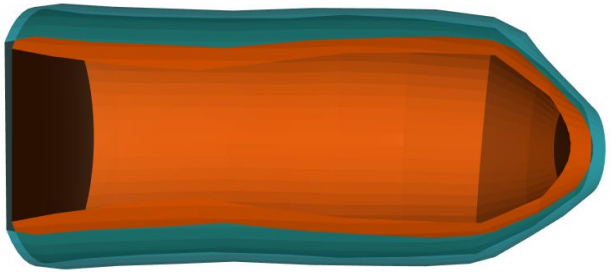




Membranes

- Built a rough object of the *E. coli* inner and outer membranes from tomograms using IMOD
- Is not a whole cell
- Volume of cell in tomograms small compared to average sizes from literature
    - Dimensions: ~1μm x ~0.4 μm



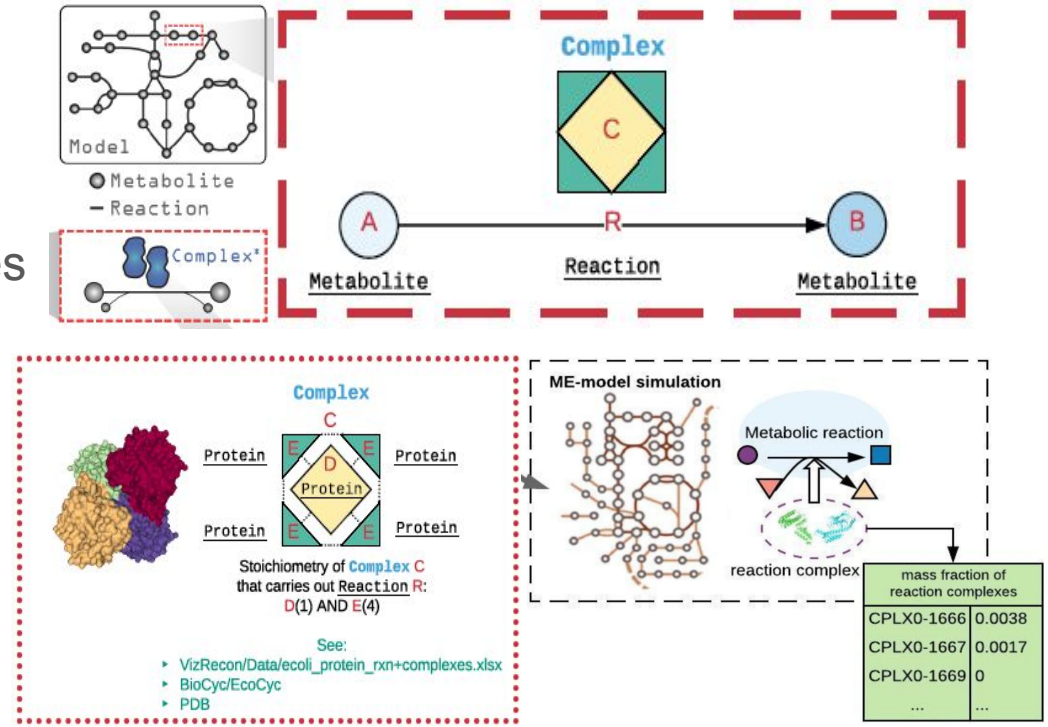


Prototype 2

A more complete representation of *E. coli* and final model for this senior design project

Addition of Complexes

- Macromolecule Expression model (ME-model) determines stoichiometry of complex structures
- Formation rate (flux) of these complexes are obtained by ME-model simulations
- Homology models used for unknown structures

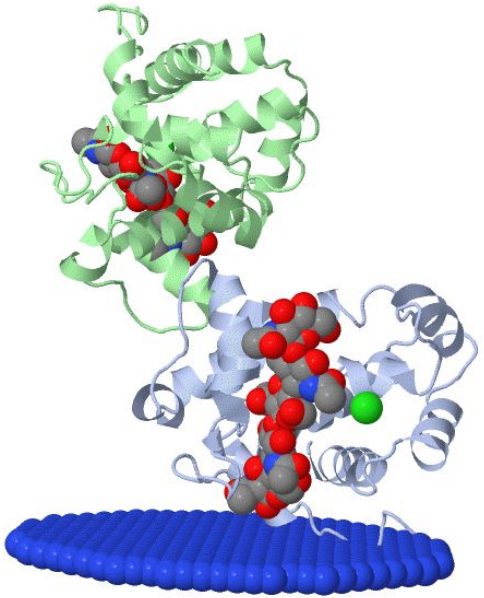


Computation of Complex Copy Numbers

- Complex copy numbers were estimated using ME-model mass fraction ratios
- Copy Numbers were determined using these assumptions:
    1. Proteins want to be in complexes
    2. There is a “limiting” protein
- The “limiting protein” is arbitrarily chosen to best maximize total number of complexes made

Membrane Proteins and Complexes

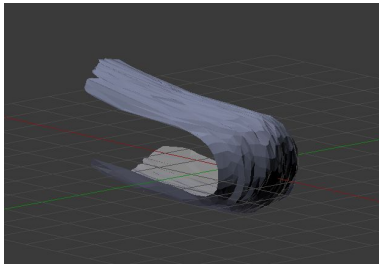
- cellPACK can pack membrane proteins by locking parameters
- Orientations of Proteins in Membranes (OPM) contains high-quality annotations for proteins
  - Alternatives: Transmembrane Helix Prediction using Hidden Markov Models (TMHMM), UniProt
- Lab member developed algorithm to confirm quality

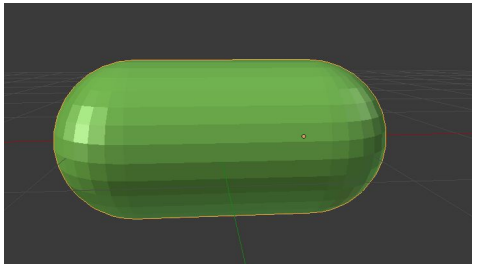


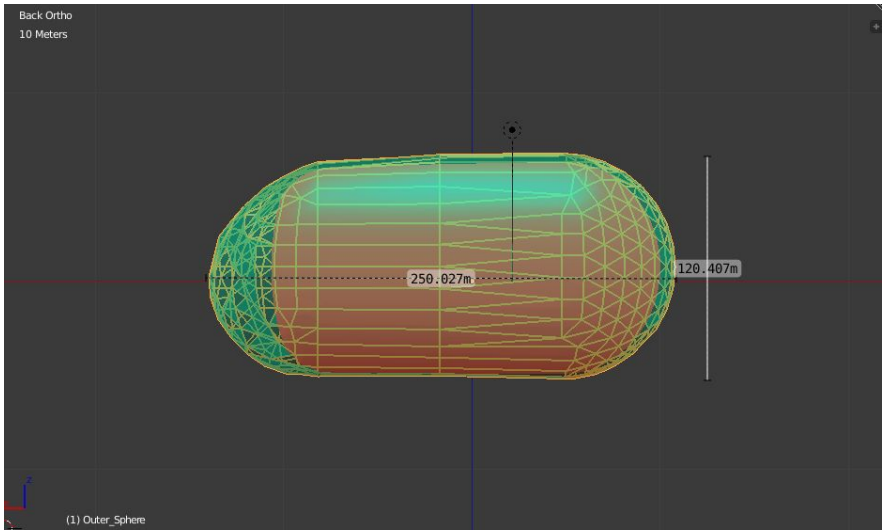
FirstGlance in Jmol

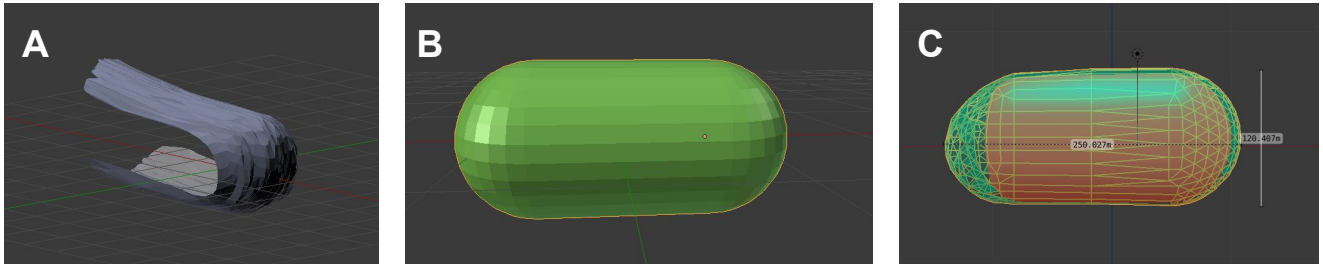
Membrane Improvements

- Made “pill” from dimensions in literature
- Total Volume: 2.2μm^3
- Drawbacks:
  - Periplasm Volume to whole cell volume: 19% (compared to 8% from experimental data)



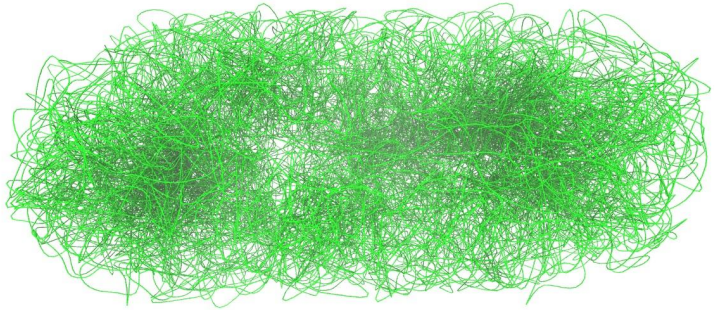


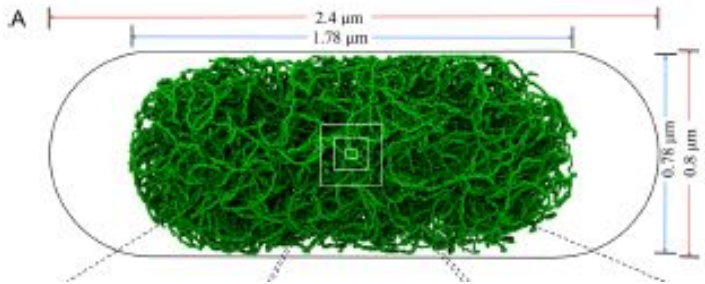



A
B
C
250.027m
120.407m

Addition of Chromosome

- 500 base pairs per bead (bpb) model of an *E. coli* chromosome
- Mid-cell origin of replication
- Shared with us by William Hacker and Professor Adrian Elcock

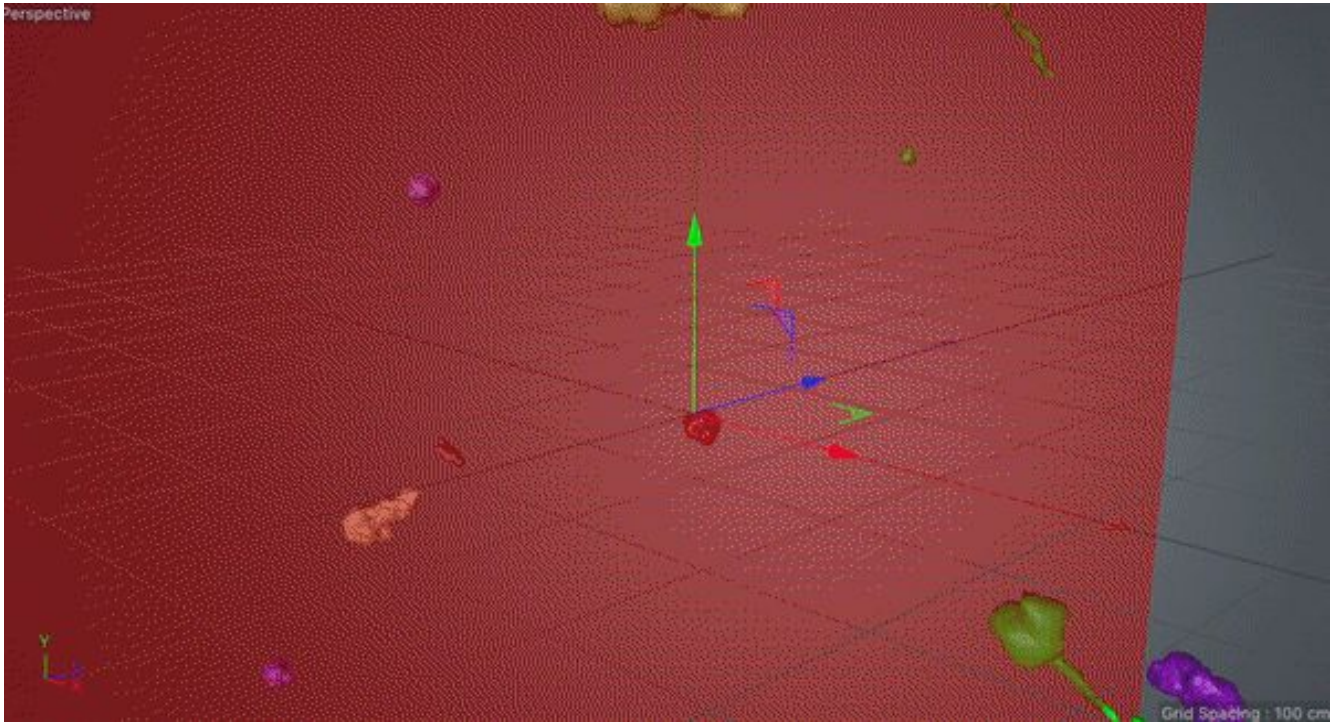




Final “Menu” and Recipe

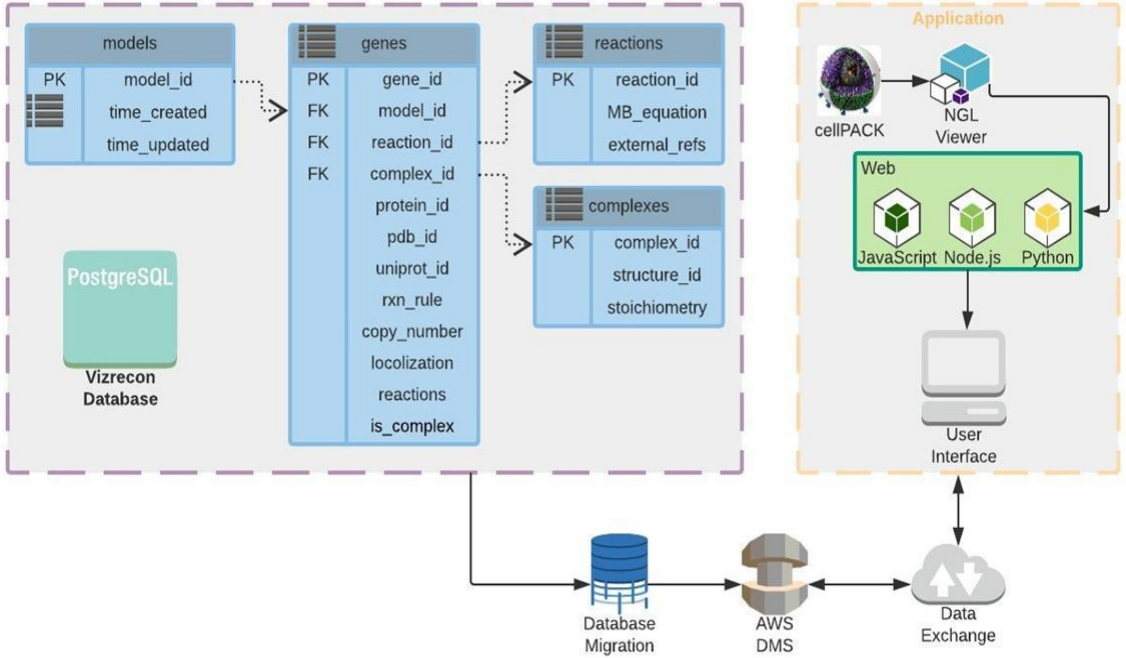
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	UniProt	NAME	Description	Structure_ID	PDB	Structure_sc	MW	MOLARITY	Copy_numb	Localization	Localization_	INCLUDE	COLOR	NOTES	model_
2	P0AD86	thrL	thr operon leader	REP-E00001	E00001_clean-X	metaTASSER	2138.4426			Cytosol	EchoLOCATION	predicted			iML151
3	P00561	thrA	Bifunctional aspa	REP-E00002	E00002_clean-X	metaTASSER	89119.2202	4.11E-06	4375	Cytosol	iJL1678	x			iML151
4	P00547	thrB	Homoserine kina	REP-1_308_4rpf	b0003_1_308_4r	SWISS-MODEL	33623.2905	0	844	Cytosol	iJL1678				iML151
5	P00934	thrC	Threonine syntha	REP-1vb3	1vb3-A_clean.pd	Protein Data Ban	47113.2536	1.28E-05	13676	Cytosol	iJL1678	x			iML151
6	P75616	yaaX	Uncharacterized	REP-E00005	E00005_clean-X	metaTASSER	11354.7994			Periplasm	EchoLOCATION	predicted			iML151
7	P0A8I3	yaaA	UPF0246 protein	REP-5caj	5caj-A_clean.pdb	Protein Data Ban	29585.4538	5.75E-07	613	Cytosol	EchoLOCATION	x			iML151
8	P30143	yaaJ	Uncharacterized	REP-E00007	E00007_clean-X	metaTASSER	51662.291			Inner_Membrane	iJL1678				iML151
9	P0A870	talB	Transaldolase B	REP-4s2b	4s2b-A_clean.pd	Protein Data Ban	35218.8107	8.68E-06	9247	Cytosol	iJL1678	x			iML151
10	P0AF03	mog	Molybdopterin ad	REP-1di6	1di6-A_clean.pdb	Protein Data Ban	21222.105	1.45E-07	155	Cytosol	iJL1678	x			iML151
11	P0AC98	satP	Succinate-acetat	REP-E00010	E00010_clean-X	metaTASSER	20070.6303			Cytosol	Arbitrary assignment				iML151
12	P75617	yaaW	UPF0174 protein	REP-E00011	E00011_clean-X	metaTASSER	26664.5763			Cytosol	EchoLOCATION	predicted			iML151
13	P28697	mbiA	Uncharacterized protein MbiA				17637.1282			Cytosol	Arbitrary assignment				iML151
14	P28696	yaaI	UPF0412 protein	REP-E00012	E00012_clean-X	metaTASSER	14481.418			Periplasm	EchoLOCATION	predicted			iML151
15	P0A6Y8	dnaK	Chaperone prote	REP-4jne	4jne-A_clean.pdb	Protein Data Ban	69114.1284	1.70E-05	18133	Cytosol	iJL1678	x			iML151
16	P08622	dnaJ	Chaperone prote	REP-4_348_4j80	b0015_4_348_4j	SWISS-MODEL	41099.8758	7.02E-07	748	Cytosol	iJL1678	x			iML151
17	P0CF91	insL1	Putative transpos	REP-INSL2_ECC	INSL2_ECOLI_m	I-TASSER	40908.3273			Cytosol	Arbitrary assignment				iML151
18	P33236	mokC	Regulatory protei	REP-E00016	E00016_clean-X	metaTASSER	7742.206			Periplasm	EchoLOCATION	predicted			iML151
19	P13738	nhaA	Na(+)/H(+) antipo	REP-1zcd	1zcd-A_clean.pd	Protein Data Ban	41355.1815			Inner_Membrane	iJL1678				iML151
20	P0A9G2	nhaR	Transcriptional a	REP-6_298_3k1	b0020_6_298_3k	SWISS-MODEL	34284.0399			Cytosol	EchoLOCATION	predicted			iML151
21	P0CF25	insB1	Insertion element	IS1 1 protein InsB			19564.3535			Cytosol	Arbitrary assignment				iML151
22	P0CF07	insA1	Insertion element	IS1 1 protein InsA			9868.2797			Cytosol	Arbitrary assignment				iML151
23	P0A7U7	rpsT	30S ribosomal pr	REP-5afi	5afi-t_clean.pdb	Protein Data Ban	9684.2814	2.19E-05	31222	Cytosol	iJL1678	x			iML151
24	P75620	yaaY	Uncharacterized	REP-E00023	E00023_clean-X	metaTASSER	7890.2841			Inner_Membrane	EchoLOCATION	predicted			iML151

cellPACK

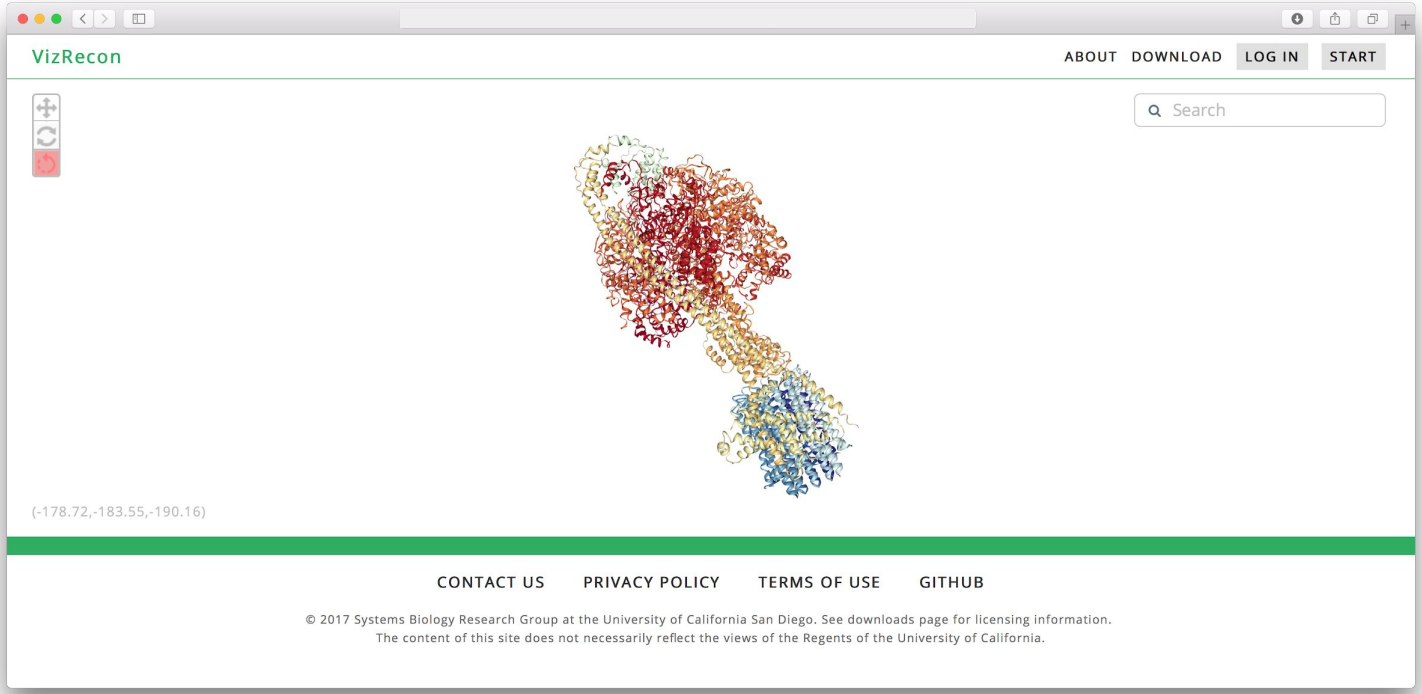


How do we display the pipeline and model?

Integration of the components



Building an interactive platform



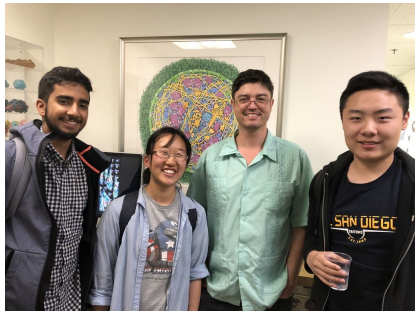
Conclusions and Future Goals

- Created pipeline to link structural and spatial info with “omics” data
- Able to generate appropriate complexes and membrane structures
- Still in progress of packing final model
- Working on finding an efficient way of displaying the model using NGL
- Improve the detail of visualization (i.e. interactions between components, more constraints for localizations, thermodynamics, kinetics, etc.)
- Update model to include more experimentally determined components
- Apply the pipeline and create reconstructions to other organisms
- Integrate metabolic networks to simulate directly on the 3D model with the aid of molecular dynamics (MD)

Acknowledgements

- Dr. Nathan Mih
- Edward Catoiu
- Ruoqing Cheng
- Liangyu Zhao
- Kritin Karkare
- Dr. Bruce Wheeler
- Gladys Ornelas
- Anjulie Agrusa
- Marissa Keller

- Dr. Brett Barbaro
- Dr. Alex Rose
- Dr. Bernhard Ø. Palsson



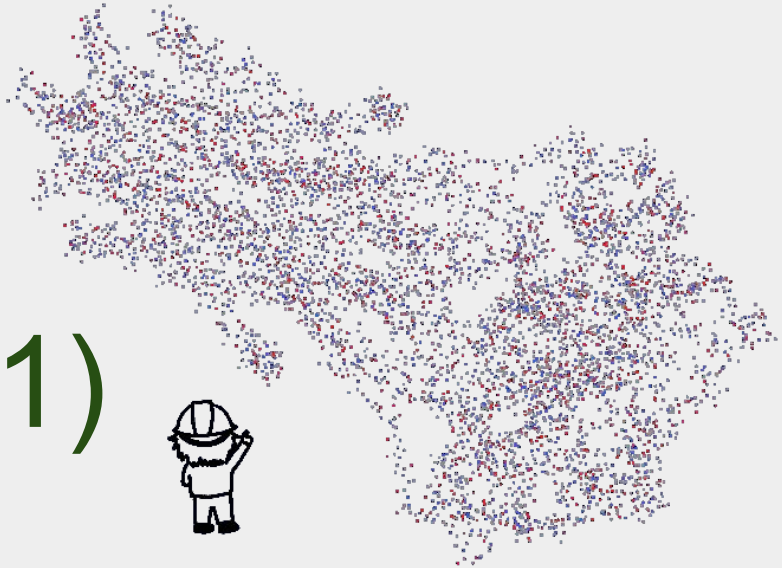
THE SCRIPPS RESEARCH INSTITUTE

50 YEARS 1966 2016 UC SAN DIEGO BIOENGINEERING Opportunities for Life

SYSTEMS BIOLOGY RESEARCH GROUP

Questions?
(we are Group 31)

((for those who are wondering))



References

- **ssbio: a Python framework for structural systems biology 2018**, Nathan Mih, Elizabeth Brunk, Ke Chen, Edward Catoiu, Anand Sastry, Erol Kavvas, Jonathan M Monk, Zhen Zhang, Bernhard O Palsson
- **The quantitative and condition-dependent *Escherichia coli* proteome 2016**, Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold & Matthias Heinemann
- **cellPACK: a virtual mesoscope to model and visualize structural systems biology 2015**, Graham T Johnson, Ludovic Autin, Mostafa Al-Alusi, David S Goodsell, Michel F Sanner& Arthur J Olson