

Mesoscope: A Web-based Tool for Mesoscale Data Integration and Curation

L. Autin¹ , M. Maritan¹ , B. A. Barbaro¹ , A. Gardner¹ , A. J. Olson¹ , M. Sanner¹ , and D.S. Goodsell^{1,2} 

¹Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

²RCSB Protein Data Bank and Center for Integrative Proteomics Research, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, USA

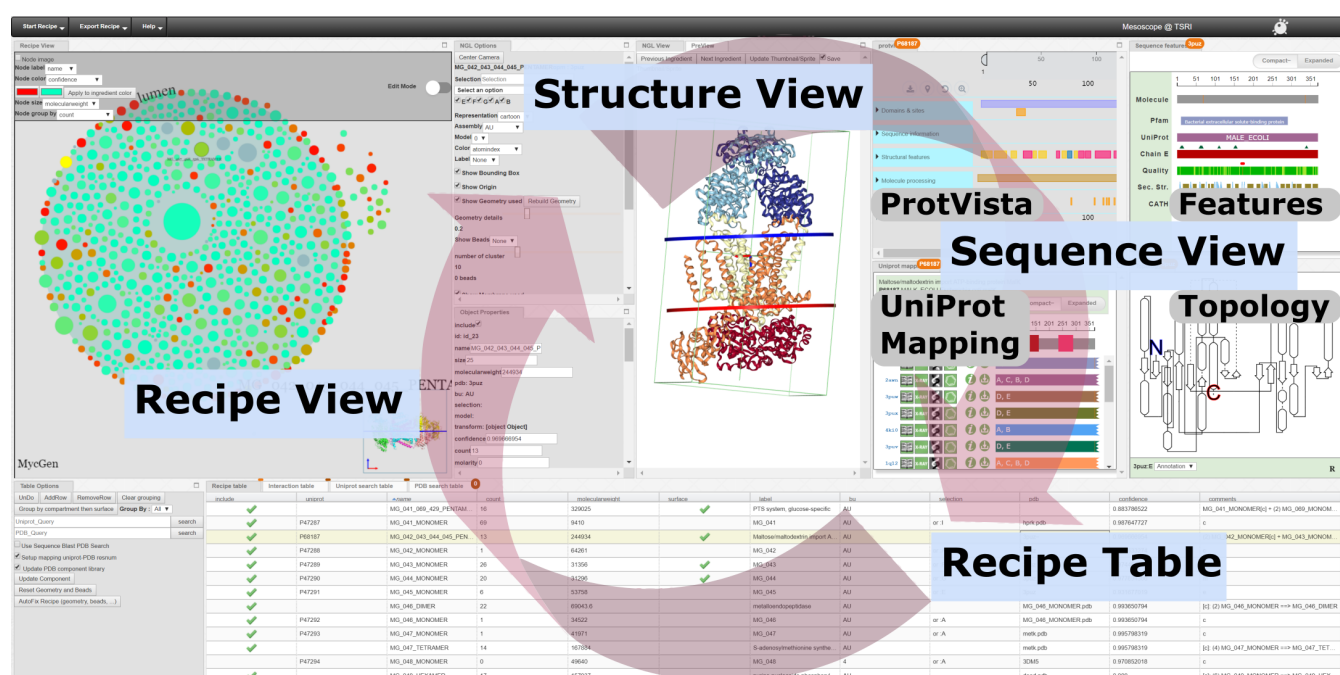


Figure 1: Mesoscope Web Interface

Abstract

Interest is growing for 3D models of the biological mesoscale, the intermediate scale between the nanometer scale of molecular structure and micrometer scale of cellular biology. However, it is currently difficult to gather, curate and integrate all the data required to define such models. To address this challenge we developed Mesoscope (mesoscope.scripps.edu/beta), a web-based data integration and curation tool. Mesoscope allows users to begin with a listing of molecules (such as data from proteomics), and to use resources at UniProt and the PDB to identify, prepare and validate appropriate structures and representations for each molecule, ultimately producing a portable output file used by CellPACK and other modeling tools for generation of 3D models of the biological mesoscale. The availability of this tool has proven essential in several exploratory applications, given the high complexity of mesoscale models and the heterogeneity of the available data sources.

CCS Concepts

• Applied computing → Bioinformatics; • Computing methodologies → Graphics systems and interfaces;

1. Introduction

Structural modeling of entire cells is a grand challenge actively being addressed by the structural biology community. Given the magnitude and complexity of this challenge, it is essential to take an integrative approach, bringing together data from multiple experimental modalities that address multiple levels of temporal and spatial scale. Already, this approach has generated detailed atomic models of synaptic vesicles [THS*06] and a synaptic bouton [WMT*14], HIV in its physiological context [JGA*14], and bacterial cells [GAO18, HLE17]. These efforts are the first steps towards a comprehensive understanding of living systems at the cellular mesoscale, which spans from the atomic level to the cellular level. In many respects, the overarching goal of biological research is to explore and understand all steps from genotype to phenotype, and use this knowledge to identify key bottlenecks to target with medically-relevant interventions. This is the grand challenge of mesoscale modeling: to fill in all of the gray areas in our understanding spanning molecular biology to cell biology, such that we can then apply this knowledge to health and welfare [AM18, SMW*18, FS19].

Integrative mesoscale structural biology poses several challenges that are testing the current limits of technology and understanding. The intrinsic complexity of mesoscale biology poses the first major challenge. Mesoscale environments are highly crowded, often with 20-30% of the space filled with macromolecular components [Ful82], and these components are highly heterogeneous in size and shape, with functional modes of flexibility and highly specific modes of interaction. The complexity of mesoscale systems also strains the limits of current hardware and software, and modalities of archiving, dissemination and reproducibility. Mesoscale models pose exciting new challenges for conceptualization and analysis, and current tools for modeling, simulation and visualization, which have typically been designed for use on individual macromolecular structures or image/volume-based cellular data, often fail when dealing with systems of this size and complexity.

The heterogeneity of biological data poses an orthogonal challenge to this biological complexity. This data is heterogeneous in both the intrinsic form of the data and the ways that it is disseminated. Many sources of data are available, often independent and with their own API, formats, and data standards. For example, structural and sequence data is becoming increasingly interconnected through mature global resources such as the PDB (www.rcsb.org) and UniProt (www.uniprot.org), but proteomics and interactomics data are still scattered across multiple databases. Effective progress in mesoscale modeling will necessarily require effective methods for gathering, curating and integrating these heterogeneous sources of information.

Mesoscope (mesoscope.scripps.edu/beta) addresses the challenge of data gathering and integration (figure 1). It is a web-based tool that allows users to build and rapidly curate a database of molecules, and define their concentration, location, and local interactions within the mesoscale environment. It produces a json recipe file for convenient use in methods for generating mesoscale models, such as CellPACK [JAAA*15] and CellPAINT [GAB*18].

2. Mesoscope Overview

We chose a web approach for the development of the interface as it offers the fastest development release cycle, is free of installation hurdles, and is platform independent. The whole platform uses vanilla javascript and is hosted on a virtual machine at our institutes. The javascript ecosystem offers a large and diverse range of tools, so most of the features described below leverage third party javascript libraries. While relying on third party service gives a lot of flexibility and permits dynamic development, we lose some control and durability. To minimize potential problems, we use frozen copies of each library to avoid broken links and broken libraries.

A typical project performed within Mesoscope involves several tasks. It begins with an overall list of components, or ingredients, often taken from a proteomics study. Preliminary assignments for atomic structures may also be associated with list, as well as locations and concentrations, which may be input with this list of components. Then the user interactively curates the list and associated structures, validating that each structure is appropriate and that locations and concentrations are reasonable. During this process, various third-party tools allow the user to search existing database resources to obtain any information that is missing or replace information that is faulty. After the list is curated, Mesoscope processes the structural, location and concentration data to create a recipe that includes representations at multiple levels of details of the structures, orientation information for membrane-bound proteins, connectivity information for fibers, and a host of other metadata that is needed for model generation. Mesoscope also allows interactive visualization and curation of this process, to correct cases where problems occur. Finally, the recipe is output, which includes a detailed list of protein and macromolecule structures, their molarity or copy number and their localisation in a given compartment. The recipe may then be used in modeling and visualization software such as CellPACK to create 3D instances of models that are consistent with the information in the recipe.

Figure 1 shows an overview of the default layout of Mesoscope being used to curate information for a model of an entire bacterial cell. The recipe of soluble, membrane-bound, and extracellular molecules is represented schematically at upper left in the Recipe View and in a detailed spreadsheet at lower right, the Recipe Table. At center, the Structure View, based on NGL [RH15], is used to display atomic structures of any selected molecule. The Structure View is tightly connected to the Sequence View which offers a variety of curation tools center on sequence annotation.

2.1. Dynamic Layout and User Interaction

Mesoscope presents the user with a collection of sub-windows that perform various tasks, including: windows that access the overall recipe, windows that visualize different aspects of the components, and windows that allow access to external databases and information for curation and incorporation of missing information.

These sub-windows can be dragged around, allowing users to customize their layout depending on the task. This dynamic layout is facilitated through the use of the goldenLayout library (www.golden-layout.com). Most Mesoscope tasks are performed interactively within this dynamic layout, but power users can use

the console to do scripting and advanced manipulation not available in the interface.

2.2. Recipe Setup and Visualisation

Recipes for current mesoscale models, such as the vesicles and small bacterial cells described below, typically have hundreds of components arrayed in one or more compartments. One main goal of Mesoscope is to allow nimble navigation through these components, to facilitate effective curation and validation. The current version of Mesoscope provides two mechanisms for navigation through the recipe: a graphical interface that gives a visual abstract of the entire recipe, and a table interface that contains a comprehensive list of information associated with each component.

2.2.1. Recipe View

We designed a visual abstraction of the recipe based on the D3 library, with three main features: tree structure (hierarchical node-link structure), circle packing and force-based constraints (d3js.org) shared in two different modes: view and edit mode. The tree structure encodes the topological nesting of compartments, which are visualized as circular fields in the interface (see figure 2). The D3 circle packing is used for the initial placement of all the circles upon loading. Individual biomolecular ingredients are represented as smaller circles that are constrained to lie within the appropriate compartment circle. Force-based constraints associated with the circles ensure that there are no overlaps, and that biomolecular ingredients stay in their compartment, producing an interpretable graph as the user makes changes. In addition, the Recipe View provides user-customization options to rename an ingredient or a compartment, as well as changing the size of the circle, the labeling scheme, and the coloring scheme, providing different visual feedback on the recipe (for example, changing radius or color according different properties). By default, the recipe's canvas is in view mode that lets users select and move ingredients while remaining in their defined compartment parent. Upon activating the edit mode, users can create/edit compartments and ingredients. These circles can be manipulated interactively to change their localization, creating concentric compartments or dragging ingredients into or out of compartments, or embedding them into membranes enclosing compartments. Molecular interactions are also represented and can be added, however, a general approach for defining such interactions will be an active topic for future work (as discussed below).

2.2.2. Recipe Table

Mesoscope also includes a detailed spreadsheet-like presentation of the recipe with the list of ingredients and a summary of the important information associated with each ingredient. The Recipe Table relies on the SlickGrid library (slickgrid.net), which offers features similar to a spreadsheet editor (cell edition, grouping, sorting, etc.) (see figure 3). Upon selection of individual entries in the Recipe View or the Recipe Table, the ingredient is highlighted in both views and all other windows are updated accordingly. The Recipe Table also includes basic options for searching the PDB and UniProt archives, placing results of the search in separate tabs on the Table.

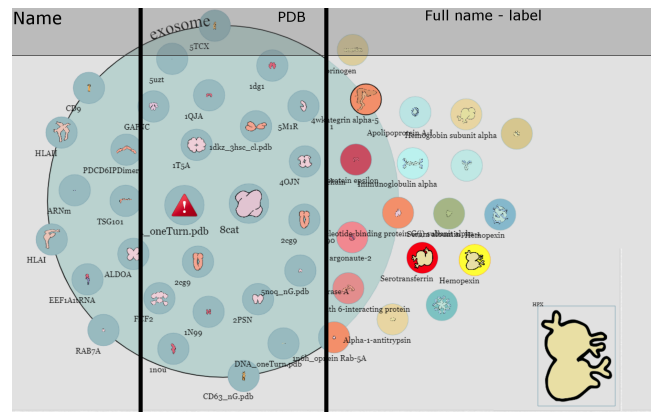


Figure 2: Recipe View

Recipe View with three options for labeling and coloring nodes: by name, PDB accession code, and with annotation information. This recipe includes components for exosomes, with hand-drawn sprites by intern Julia Jimenez. The red triangle indicates a component that requires curation, and icons may be dragged interactively between the inner, membrane, and exterior compartments.

Figure 3: Recipe Table

Detailed specifications for each molecule in the recipe is included in an editable spreadsheet, to allow curation of the data that will be exported for model building. Updates are immediately propagated to other Mesoscope components.

2.2.3. Inputting Recipes

For simple systems, users start from an empty canvas with the Recipe View in edit mode and interactively add compartments and ingredients, dragging compartments and ingredients from place to place to build the appropriate topology. This manual approach is ideal for small systems but becomes difficult for larger systems consisting of hundreds of ingredients. In these cases, the user inputs a simple comma-separated-value (csv) file or Excel spreadsheet. Upon loading a spreadsheet a modal interface appears that allows the user to define the mapping between their column information and the information required by Mesoscope. At a minimum, this information will include a name, concentration, membrane localisation (if relevant) and compartment parent for each ingredient. This information will populate the tree structure that is shared by the Recipe View and the Recipe Table.

2.3. Structure View

The atomic structure of each ingredient, if any, is specified using either a four character PDB code or a custom filename. In the custom filename case, the user needs to upload a data folder where Mesoscope will find the given file. The PDB code or the file is then loaded and represented using the NGL viewer [RH15]. NGL is a fully-fledged molecular viewer for the web, with easy access to all of the familiar molecular graphics representations, selection schemes, and rendering modes. Several specialized representations available in NGL are used during curation of structures, including display of a reduced sphere representations of the structures (e.g. coarse-grain bead which is used for multi-resolution representation and visualization of molecules and assemblies) and display of schematic representations of membrane orientation and boundaries (see figure 4).

2.4. Sequence View

If the ingredient is a PDB entry, several resources from the PDB Component Library (www.ebi.ac.uk/pdbe/pdb-component-library) are used to visualise the sequence features and the topology. If a UniProt id is specified then the full sequence feature is visualised with ProtVista [WGPM17], and the full UniProt-to-PDB mapping [DGT*18] can be used to highlight region in NGL. The Sequence Features, ProtVista, and Topology tabs provide third-party information on the structure being displayed in NGL, to assist with curation tasks. Furthermore, these windows are synchronized directly with the structure visualized in the NGL view. This is especially useful for membrane proteins, allowing the user to highlight the transmembrane region on the 3D structure based on information from UniProt (see figure 4).

2.5. Assignment and Curation of Component Properties

The major impetus for the development of Mesoscope was the difficulty in assigning and curating all of the properties that are needed during model generation. Biomolecules are highly diverse, and there are numerous exceptions and oddities that need to be addressed when sorting through a list of ingredients. Mesoscope is designed to assign these properties automatically using a set of widely-accepted defaults, and then allow rapid manual curation and update to correct any errors in this automated assignment.

The NGL viewer window (see figure 4) provides much of the functionality for this process. NGL option windows allow users to specify advanced information for the selected proteins, including:

- Selection of an appropriate biological unit from the information included in the coordinate file, including generation of symmetry mates, chain selection, choice of model number, etc.
- Building a representative mesh geometry that is used for rendering and collision in cellPACK [JAAA*15].
- Building a coarse-grain bead representation for use in packing and simulation. The user provides the number of desired beads, and Mesoscope will do a K-means clustering on the atomic structure and automatically assign the bead radius or used a user provide unique radius.
- Rendering a sprite illustration of the ingredient with the current

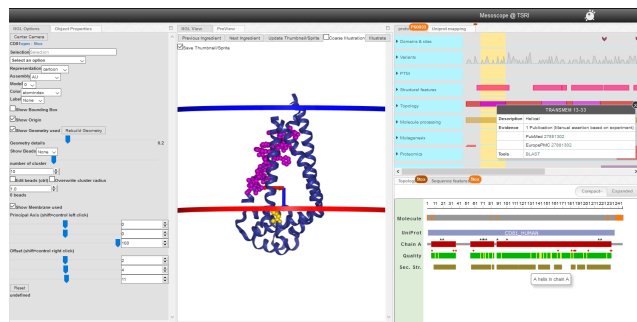


Figure 4: NGL Molecular Viewer

NGL is used to display and manipulate atomic structures of components. Most standard NGL visual representations are supported. A membrane-bound protein is shown here, showing the location of the membrane-spanning portion, which may be tuned using interactive sliders. NGL viewport is synchronized with sequence information from the Sequence features and the ProtVista widget. The mouse click and over position on the sequence displays the corresponding 3D residues (purple and yellow ball and stick) helping refining membrane orientation.

camera orientation using a screenshot of the NGL viewport or using our Illustrate cgi-server [GAO19]. By default the sprite is taken directly from the RCSB PDB (www.rcsb.org).

- If the protein is a membrane protein, Mesoscope will check if it's present in the "Orientations of Proteins in Membranes" (OPM) server [LPJ*12]. The user can also manually set the orientation and portion buried in the membrane, using an axis and an offset.
- If the protein is defined as a fiber, a repetition of one structural unit is input and the user can specify the main axis along which the unit is repeated.

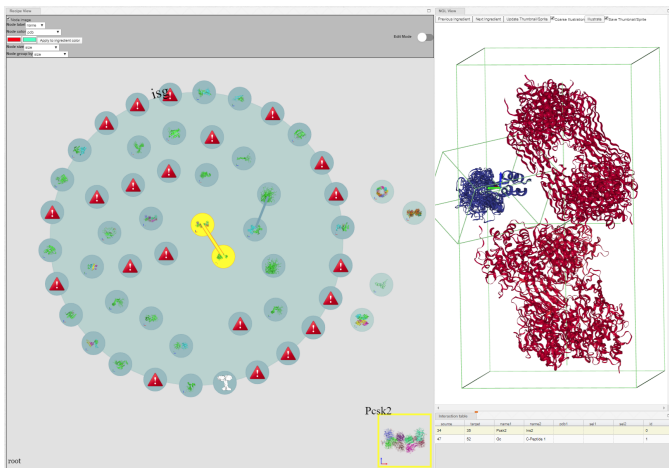
Several other windows manage additional information. The Object Properties window will let the user have a summary of all the information related to the selected ingredient. The user can also define here the type of compartment (sphere, primitive, mesh).

2.6. Under development

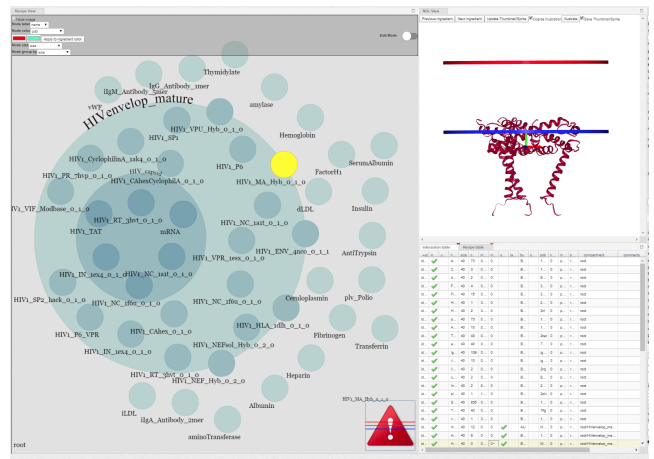
Working on a web platform is highly dynamic and flexible, also several features are currently under development, and are included as placeholders or minimally-functional methods in the beta version. These include an interaction table as part of the Recipe Table, which will be expanded with more comprehensive representations of interactions between ingredients (figure 5(a)), and a preview window, which will act as a mesoscale model viewer. Model will be generate in the fly from a recipe or loaded from more sophisticated modeling methods (figure 5(b)).

2.7. Export

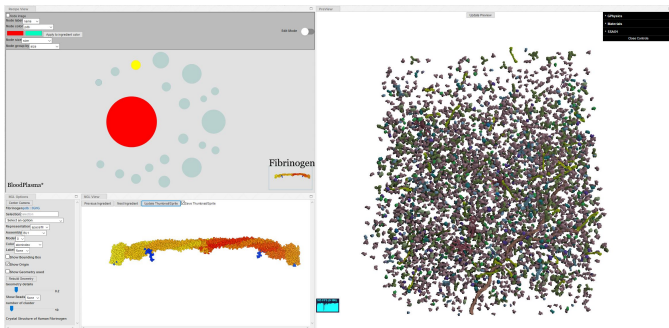
All the widgets are windows that are opened on the same underlying data represented in the tree structure. This ensures that all windows are synchronized, and, for example, changing the name of a circle in the Recipe View will change the underlying node name and query an update of Recipe Table View. Similarly, changes in the



(a) Interaction



(a) Recipe



(b) Preview

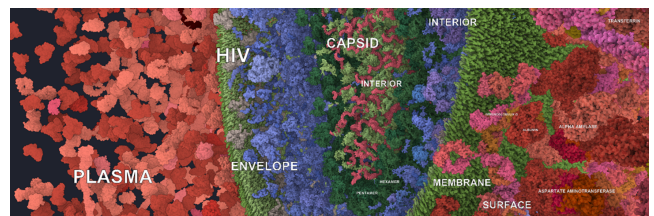
Figure 5: Mesoscope Features Currently Under Development

a) Interaction - A prototype capability to define specific interactions between molecules is implemented by connecting icons in the Recipe View, and providing a structure of the complex of the two ingredients. b) Preview - A prototype tool to preview interactive instances of the model showing the number of each ingredient in the current state of the recipe.

Object properties such as the compartment type or ingredient type will be reflected in the Recipe Table and vice versa. In this way, curation may be performed in any of the Views, the information will be updated throughout Mesoscope, and the user can use the Recipe View (coloring nodes by viewed state) to highlight molecules that have not yet been curated.

Finally, when fully curated, users export the table in formats appropriate for different modeling applications:

- Reduced csv format includes basic information but omits information such as mesh geometry or bead coordinates.
- CellPACK recipe includes mesh geometry and bead representations, for use in our original CellPACK software [JAAA*15], and a sprite/image for each ingredient if used in CellPAINT [GAB*18].
- CellPACK-gpu recipe supports the combined model-building and visualization engine cellPACK-gpu [KAK*18], based on a



(b) CellVIEW

Figure 6: HIV in Blood Plasma

a) The recipe includes two concentric bounding elements, the inner capsid and the outer envelope membrane, and soluble/fibrous components inside the capsid, in the viral lumen, and outside the viral envelope. The membrane location of the viral matrix protein (highlighted in yellow in the recipe view) is being curated in the NGL window. b) CellVIEW visualization of a cross-section through the HIV final model, using context-specific labeling methods to identify the major features of the current view.

parallel instant distribution method. This recipe doesn't require a mesh geometry or bead coordinates, but curation of a bead representation is recommended if simulation studies will follow the modeling.

- Color palettes or concentration/copy number palettes may also be exported, to allow users to have one recipe with different conditions and colors.

3. Applications

The development of Mesoscope has been driven by challenges we have faced in several collaborative applications. These include modeling an HIV virion surrounded by blood plasma as part of a multidisciplinary research center, building a structural models of insulin secretory vesicles that reconcile experimental data from soft x-ray tomography, and generating bacterial models from systems biology data. These biological drivers prompted the generation, evaluation and addition of new features in the context of modeling by biomolecular researchers.

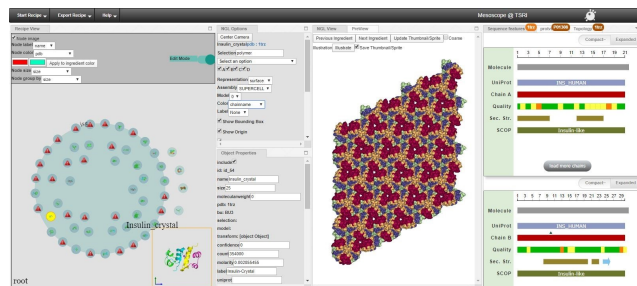
The initial development of CellPACK was performed using a

small but deceptively tricky system: an HIV virion surrounded by blood plasma [JGA*14], in collaboration with researchers at the HIVE Center (hive.scripps.edu). Manual generation and curation of this system underscored the need for a tool like Mesoscope, and has provided an excellent use-case throughout development. A few of the challenges implicit in this model include: the need to define a compartment bounded by a membrane and also a compartment bounded by a protein capsid, and the incorporation of a flexible fiber composed of RNA and associated proteins (see figure 6(a)). We have also enjoyed extensive collaboration with Ivan Viola group (www.cg.tuwien.ac.at/staff/IvanViola.html) on modeling methods and visualization based on this recipe (see, for example [KCK*18]), which has spurred improvement of the export formats (see figure 6(b)).

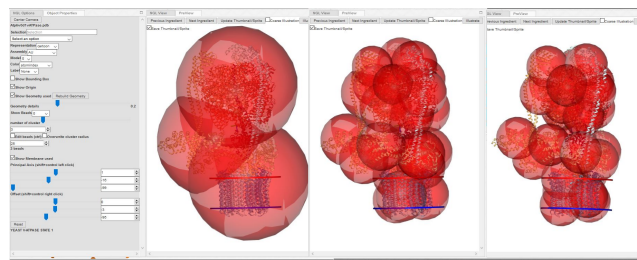
The basic functionality was implemented and tested during work with the Pancreatic Beta-cell Consortium [SMW*18]. The goal of the Consortium is to build a predictive model of an entire eukaryotic cell, and we have chosen the characteristic organelle of these cells, the insulin secretory granule (ISG), as a first step (figure 7). An iterative dialog between development and application in this system led to the refinement of the coarse-grain bead interface, letting user overwrite the radius of the beads, but also letting user change the position of individual beads through the implementation of an edit mode relying on NGL event framework (figure 7b). Multiple third-party tools that showed utility during the process of curation have been incorporated directly into Mesoscope as well. We feel that having nimble access to these curation tools is one of the major value-added features of Mesoscope, greatly streamlining the process of validating the many ingredients that are incorporated into these recipes.

Extensive feature development and testing of Mesoscope was then performed during curation of data for the simple bacterium *Mycoplasma genitalium* (figure 1 and 8). The goal of this project is to generate snapshots of the bacterium at defined time points in the life cycle, integrating data parameters from whole-cell systems biology simulations from the Covert laboratory [KSM*12]. The outcome represents the first 3D mesoscale model of an entire bacterial cell with atomic detail, and is the most complex recipe that we have attempted, with 683 ingredients including complex fibrous elements such as DNA and RNA.

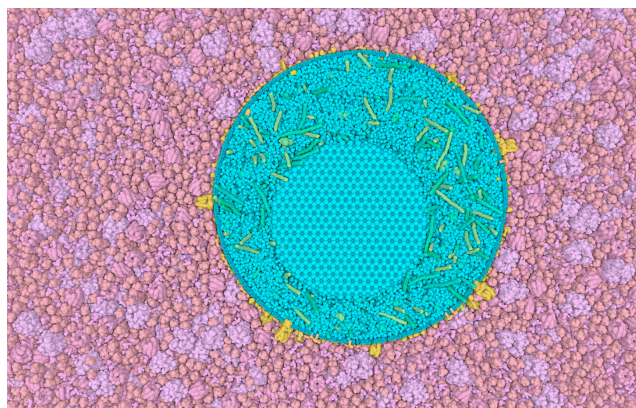
Key information about the ingredients (protein identifications, functions, molecular weights, locations, UniProt codes) were extracted from WholeCellKB (www.wholecellkb.org [KSM*13]) as json files, while molarities were obtained from processing simulations as hdf5 files downloaded from WholeCell-SimDB (www.wholecellsimdb.org [KPC14]). The initial draft recipe was generated as a csv file with a Python script that incorporates all the data extracted from the simulations and the structural models assigned to each ingredient. Mesoscope was used to visually inspect every structure in the model, allowing ready curation of specific biological assemblies and assignment of protein chains. Also, Mesoscope was instrumental in the proper orientation for membrane proteins that are not deposited in the OPM. The Protvista panel proved particularly useful in this task, as it retrieves the transmembrane domain location from the UniProt entry, and was used as a guide to define the membrane orientation.



(a) ISG Recipe



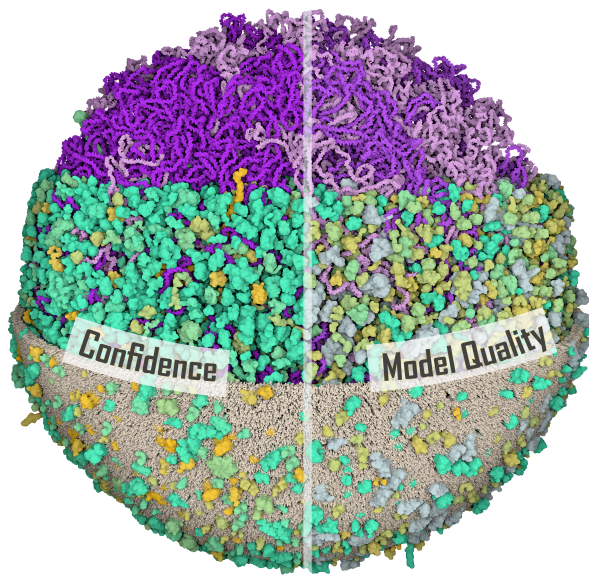
(b) ISG Beads Setup



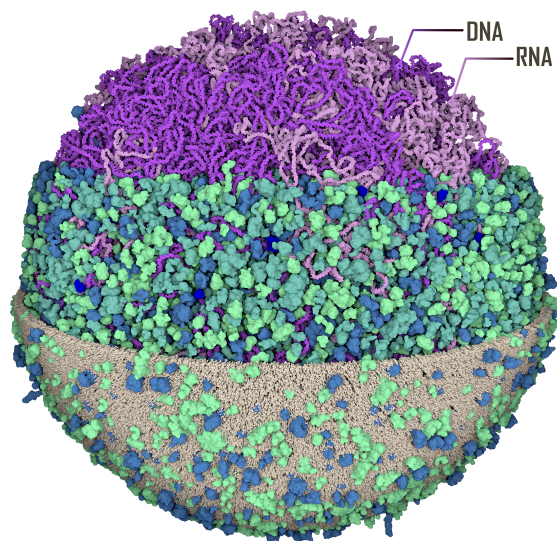
(c) ISG Model

Figure 7: Insulin Secretory Vesicle

a) The recipe includes membrane-bound and soluble interior ingredients for the vesicle, and a simplified recipe for generic cytoplasm. The recipe also includes specifications for a small crystal of insulin, being previewed in the NGL viewer, which is often found in mature forms of the vesicle. b) Through K-means clustering of atomic position, Mesoscope generates n -beads representing the current structure viewed in NGL. Radii of the beads may be calculated from the cluster boundary (left and center, 3 and 25 beads auto radius) or defined by the user (right, 25 beads with radius of 25 Angstroms). c) cellPACK-gpu visualization of a cross-section through the model, with the vesicle in turquoise and the surrounding cytoplasm in pink. The large circular feature inside the vesicle is the insulin crystalline inclusion.



(a) Structure



(b) Score

Figure 8: Mycoplasma Cell

Colors are used to explore confidence levels and provenance of ingredients in a model of an entire mycoplasma cell. These values are curated in the Mesoscope spreadsheet and exported to CellView for display. (a) Color coding of confidence levels for the protein ingredients of the recipe. (b) Color coding of protein ingredients by the source of the molecular structure.

**Figure 9: Recipe Exported to CellPAINT**

A recipe composed of hand-drawn ingredients was designed and curated in Mesoscope, then exported to CellPAINT, a prototype paint program designed for interactive creation of 2.5D mesoscale images.

Two new functionalities were implemented during the curation of this mycoplasma model: the possibility to apply user defined colors to ingredients and more flexible options to update specific parts of existing draft recipes or merge two draft recipes. When dealing with a considerable number of ingredients, the coloring system on the 'Recipe View' panel now allows users to instantly highlight desired ingredient properties (i.e. molecular weight, copy number, confidence on the structural model assigned) and visually spot the ingredients that require more attention. The 'merging' options enable simultaneous updates of all ingredient copy numbers extracted at different time frames of the simulations, allowing a curated recipe for one time point to be applied to all time points in a simulation. Aspects of file export were also streamlined during this study, exporting of the color palettes and protein molarities as single files readable from cellPACK-gpu, greatly facilitating the interpretation and the navigation of MG-WholeCell model.

Our laboratory is also exploring alternative interfaces for modeling and visualizing the cellular mesoscale, which rely on Mesoscope recipes to define the systems under study. For example, we have created an application similar to a paint program, CellPAINT, that allows users to draw membranes interactively and spray in collections of proteins (figure 9). To support CellPAINT, an option was added to Mesoscope to capture still images of ingredients, using our illustrative rendering techniques and views chosen in the Mesoscope NGL window. These images are then used as sprites in CellPAINT for interactive creation of mesoscale illustrations. We are currently exploring an extension of CellPAINT that takes a similar painting metaphor into three dimensions within a virtual reality environment (figure 10).

4. Future Plans

Mesoscope, like much of the field of mesoscale modeling, is actively under development. The beta version should be treated as a proof-of-concept for the basic functionality needed for simple cellular systems that include membranes, soluble biomolecules, and a

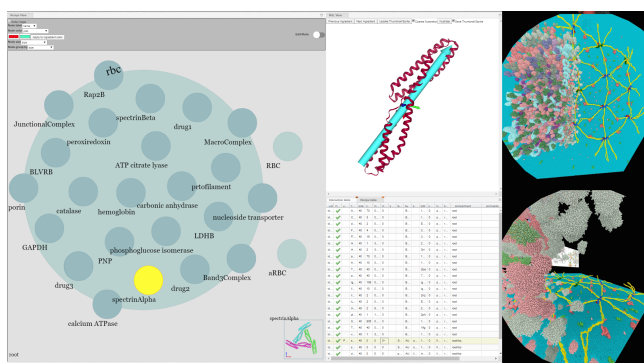


Figure 10: Recipe Exported to CellPAINT-VR

A recipe for the erythrocyte cytoskeleton was designed and curated in Mesoscope. Two domains from spectrin are being curated in the NGL viewer at center, to specify how they will be connected to form a larger fibrous molecule. The recipe was exported to CellPAINT-VR, to allow construction of the network-shaped cytoskeleton using virtual reality viewing and interactive manipulation.

few specialized higher-order structures. Future development will be needed to address the many additional levels of hierarchical complexity that are implicit in biological systems. Our current goal is to define a useful and usable representation of the network that defines interactions between molecules, which will include both information on pairings that occur and the structural details of the interactions. We are also actively developing methods for incorporating ultrastructural information from cryoelectron tomography for detailed definition of compartments.

We are also very interested in improving the user experience. Our immediate goal will be to incorporate a user-friendly preview method, allowing users to identify potential problems before more resource-heavy modeling efforts. The size of these systems has proven to be our major stumbling block in initial efforts, which only show effective performance with hardware with a gpu. We are currently collaborating with the developers of Mol*(molstar.org), which provides many effective methods for approaching systems of this complexity. The Mol* viewer work in progress with example of cellPACK model can be seen at molstar.org/viewer.

5. Availability

- Mesoscope is freely available at mesoscope.scripps.edu/beta and has been tested on Firefox, Chrome and other modern browsers.
- A tutorial for Mesoscope use with cellPACK is available at www.autopack.org/documentation/tutorials/nbcr-summer-training-2018
- A tutorial for Mesoscope use with cellPAINT is available at ccsb.scripps.edu/cellpaint/cellpaintexo-tutorial

6. Acknowledgements

This work was supported by grant GM120604 from the US National Institutes of Health. This is manuscript 29944 from the Scripps Research Institute.

References

- [AM18] AMARO R. E., MULHOLLAND A. J.: Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat Rev Chem* 2, 4 (Apr 2018). [PubMed Central:PMC6445369] [DOI:10.1038/s41570-018-0148] [PubMed:30949587]. 2
- [DGT*18] DANA J. M., GUTMANAS A., TYAGI N., QI G., O'DONOVAN C., MARTIN M., VELANKAR S.: SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Research* 47, D1 (11 2018), D482–D489. URL: <https://doi.org/10.1093/nar/gky1114>, arXiv:<https://academic.oup.com/nar/article-pdf/47/D1/D482/27437256/gky1114.pdf>, doi:10.1093/nar/gky1114. 4
- [FS19] FEIG M., SUGITA Y.: Whole-Cell Models and Simulations in Molecular Detail. *Annu. Rev. Cell Dev. Biol.* 35 (10 2019), 191–211. [PubMed Central:PMC6783340] [DOI:10.1146/annurev-cellbio-100617-062542] [PubMed:27957539]. 2
- [Ful82] FULTON A. B.: How crowded is the cytoplasm? *Cell* 30, 2 (Sep 1982), 345–347. [DOI:10.1016/0092-8674(82)90231-8] [PubMed:6754085]. 2
- [GAB*18] GARDNER A., AUTIN L., BARBARO B., OLSON A. J., GOODSSELL D. S.: CellPAINT: Interactive Illustration of Dynamic Mesoscale Cellular Environments. *IEEE Comput Graph Appl* 38, 6 (2018), 51–66. [PubMed Central:PMC6456043] [DOI:10.1109/MCG.2018.2877076] [PubMed:30668455]. 2, 5
- [GAO18] GOODSSELL D. S., AUTIN L., OLSON A. J.: Lattice Models of Bacterial Nucleoids. *J Phys Chem B* 122, 21 (05 2018), 5441–5447. [PubMed Central:PMC5980677] [DOI:10.1021/acs.jpcc.7b11770] [PubMed:12897855]. 2
- [GAO19] GOODSSELL D. S., AUTIN L., OLSON A. J.: Illustrate: Software for Biomolecular Illustration. *Structure* 27, 11 (11 2019), 1716–1720. [PubMed Central:PMC6834899] [DOI:10.1016/j.str.2019.08.011] [PubMed:30668455]. 4
- [HLE17] HACKER W. C., LI S., ELCOCK A. H.: Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. *Nucleic Acids Res.* 45, 13 (Jul 2017), 7541–7554. [PubMed Central:PMC5570083] [DOI:10.1093/nar/gkx541] [PubMed:27705758]. 2
- [JAAA*15] JOHNSON G. T., AUTIN L., AL-ALUSI M., GOODSSELL D. S., SANNER M. F., OLSON A. J.: cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nat. Methods* 12, 1 (Jan 2015), 85–91. [PubMed Central:PMC4281296] [DOI:10.1038/nmeth.3204] [PubMed:8742726]. 2, 4, 5
- [JGA*14] JOHNSON G. T., GOODSSELL D. S., AUTIN L., FORLI S., SANNER M. F., OLSON A. J.: 3D molecular models of whole HIV-1 virions generated with cellPACK. *Faraday Discuss.* 169 (2014), 23–44. [PubMed Central:PMC4569901] [DOI:10.1039/c4fd00017j] [PubMed:9759480]. 2, 6
- [KAK*18] KLEIN T., AUTIN L., KOZLIKOVA B., GOODSSELL D. S., OLSON A., GRÖLLER M. E., VIOLA I.: Instant Construction and Visualization of Crowded Biological Environments. *IEEE Trans Vis Comput Graph* 24, 1 (01 2018), 862–872. [PubMed Central:PMC5746312] [DOI:10.1109/TVCG.2017.2744258] [PubMed:20221255]. 5
- [KCK*18] KOURIL D., CMOLIK L., KOZLIKOVA B., WU H. Y., JOHNSON G., GOODSSELL D. S., OLSON A., GRÖLLER M. E., VIOLA I.: Labels on Levels: Labeling of Multi-Scale Multi-Instance and Crowded 3D Biological Environments. *IEEE Trans Vis Comput Graph* (Aug 2018). [DOI:10.1109/TVCG.2018.2864491] [PubMed:30130195]. 6

- [KPC14] KARR J. R., PHILLIPS N. C., COVERT M. W.: WholeCell-SimDB: a hybrid relational/HDF database for whole-cell model predictions. *Database (Oxford) 2014* (2014). [PubMed Central:PMCID4165886] [DOI:10.1093/database/bau095] [PubMed:15129285]. 6
- [KSM*12] KARR J. R., SANGHVI J. C., MACKLIN D. N., GUTSCHOW M. V., JACOBS J. M., BOLIVAL B., ASSAD-GARCIA N., GLASS J. I., COVERT M. W.: A whole-cell computational model predicts phenotype from genotype. *Cell 150*, 2 (Jul 2012), 389–401. [PubMed Central:PMCID3413483] [DOI:10.1016/j.cell.2012.05.044] [PubMed:20671182]. 6
- [KSM*13] KARR J. R., SANGHVI J. C., MACKLIN D. N., ARORA A., COVERT M. W.: WholeCellKB: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res. 41*, Database issue (Jan 2013), D787–792. [PubMed Central:PMCID3531061] [DOI:10.1093/nar/gks1108] [PubMed:17962296]. 6
- [LPJ*12] LOMIZE M. A., POGOZHEVA I. D., JOO H., MOSBERG H. I., LOMIZE A. L.: OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res. 40*, Database issue (Jan 2012), D370–376. [PubMed Central:PMCID3245162] [DOI:10.1093/nar/gkr703] [PubMed:11804593]. 4
- [RH15] ROSE A. S., HILDEBRAND P. W.: NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res. 43*, W1 (Jul 2015), W576–579. [PubMed Central:PMCID4489237] [DOI:10.1093/nar/gkv402] [PubMed:12825620]. 2, 4
- [SMW*18] SINGLA J., MCCLARY K. M., WHITE K. L., ALBER F., SALI A., STEVENS R. C.: Opportunities and Challenges in Building a Spatiotemporal Multi-scale Model of the Human Pancreatic β Cell. *Cell 173*, 1 (03 2018), 11–19. [PubMed Central:PMCID6014618] [DOI:10.1016/j.cell.2018.03.014] [PubMed:28768879]. 2, 6
- [THS*06] TAKAMORI S., HOLT M., STENIUS K., LEMKE E., GRØNBORG M., RIEDEL D., URLAUB H., SCHENCK S., BRÜGGER B., RINGLER P., MÜLLER S., RAMMNER B., GRÄTER F., HUB J., DE GROOT B., MIESKES G., MORIYAMA Y., KLINGAUF J., GRUBMÜLLER H., HEUSER J., WIELAND F., JAHN R.: Molecular anatomy of a trafficking organelle. *Cell 127*, 4 (Nov 2006), 831–846. [DOI:10.1016/j.cell.2006.10.030] [PubMed:17110340]. 2
- [WGP17] WATKINS X., GARCIA L. J., PUNDIR S., MARTIN M. J.: ProtVista: visualization of protein sequence annotations. *Bioinformatics 33*, 13 (Jul 2017), 2040–2041. [PubMed Central:PMCID5963392] [DOI:10.1093/bioinformatics/btx120] [PubMed:25348405]. 4
- [WMT*14] WILHELM B. G., MANDAD S., TRUCKENBRODT S., KRÖHNERT K., SCHÄFER C., RAMMNER B., KOO S. J., CLASSEN G. A., KRAUSS M., HAUCKE V., URLAUB H., RIZZOLI S. O.: Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science 344*, 6187 (May 2014), 1023–1028. [DOI:10.1126/science.1252884] [PubMed:24876496]. 2