

Integrative Structural Modeling and Visualization of a Cellular Organelle

Ludovic Autin (1), Brett Barbaro (1), Andrew I. Jewett (1), Axel Ekman (2), Shruti Verma (1),
Arthur J. Olson (1), David S. Goodsell (1,3,4)

1. The Scripps Research Institute, La Jolla, CA 92037, USA
2. Department of Anatomy, University of California, San Francisco, CA 94143, USA
3. Research Collaboratory for Structural Bioinformatics Protein Data Bank, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA
4. Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

corresponding author: David S. Goodsell, goodsell@scripps.edu

Abstract

Models of insulin secretory vesicles from pancreatic beta cells have been created using the CellPACK suite of tools to research, curate, construct, and visualize the current state of knowledge. The model integrates experimental information from proteomics, structural biology, cryoelectron microscopy, and X-ray tomography, and is used to generate models of mature and immature vesicles. A new method was developed to generate a confidence score that reconciles inconsistencies between three available proteomes using expert annotations of cellular localization. The models are used to simulate soft X-ray tomograms, allowing quantification of features that are observed in experimental tomograms, and in turn, allowing interpretation of X-ray tomograms at the molecular level.

Keywords:

insulin secretory granule, pancreatic beta cell, integrative mesoscale modeling, structural biology

INTRODUCTION

Structural modeling of entire cells is a new frontier 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 complex enveloped viruses, entire bacterial cells, and cellular organelles, as reviewed in: (Im *et al.* 2016; Feig and Sugita 2019; Goodsell *et al.* 2020).

A recent perspective (Singla *et al.* 2018) identified several characteristics that are needed for an effective integrative mesoscale structural model. The model needs to be complete over multiple levels of scale, from atomic details to overall ultrastructure. The model needs to couple a variety of representations in order to integrate different modalities of structural, biochemical, physiological, and bioinformatic knowledge. Also, as a key part of the process of integration, the uncertainty of each parameter defining the model needs to be quantified and made accessible in analysis and visualization of the model. Finally, the model must capture aspects of the heterogeneity of the system, allowing it to be both descriptive and predictive.

As part of the Pancreatic Beta-Cell Consortium (PBCC, pbccconsortium.org), we are developing methods to generate mesoscale models of functional regions of the pancreatic beta cell based on diverse experimental data from the PBCC and the larger research community. As our first proof of concept, we have chosen to model one of the defining characteristics of this cell, the insulin secretory granule (ISG). This is a particularly amenable initial target, given the abundance of available information, its manageable size and complexity, and its functional connection to disease states of the cell. With this report, we present an entire pipeline from data curation to model generation, and present potential applications that are facilitated by this

quantitative approach to mesoscale cellular biology. We also show preliminary work to simulate soft X-ray tomograms of vesicles. Soft X-ray tomography is an attractive experimental technique for imaging whole cells, since the experiment is performed in “near-native” conditions with no fixatives or freezing (McDermott *et al.* 2009) and is being actively applied to pancreatic beta cells in the PBCC (White *et al.* 2020; Loconte *et al.* 2022). We are exploring the use of integrative modeling to provide molecule-level interpretation of features in these tomograms, which typically have a resolution of 50-60 nm.

State of Knowledge for the Insulin Secretory Granule

Given its central role in the regulated delivery of insulin, there is abundant information available for ISG structure and function. We relied on several excellent reviews to provide general synthesis of current knowledge (Suckale and Solimena 2010; Germanos *et al.* 2021). In the generally-accepted view, immature ISGs have a single membrane and are filled with proinsulin and other proteins, which then mature under the action of several specific proteases. Insulin then crystallizes into a single crystal that fills most of the interior. A complex interaction with the cytoskeleton mediates storage and delivery of ISGs to the cell surface, and fusion results in dissolution of the insulin crystal and release into the blood. In the current work, we are limiting our modeling effort to just the ISG membrane and interior, in immature and mature forms.

However, there is still much uncertainty about basic information such as the ISG proteome components, concentrations, and interactions, as described in more detail below. Several proteomes are available that show discrepancies that must be reconciled. Even basic numbers such as the number of insulin molecules per vesicle show wide ranges, for example, from 200 thousand (Suckale and Solimena 2010) to one million (Eliasson *et al.* 2008). Much of the work presented in this paper is seeking to resolve this uncertainty to generate molecular models of the entire organelle that are consistent with the current state of knowledge (Figure 1A).

In this work, we gathered, curated, and integrated current information and used it to build idealized models of an ISG in cytoplasm. A manual literature search identified the 29 most widely-reported ISG protein components. These 29 were then used to train a method for scoring proteins from three proteomes, identifying 14 additional proteins that have evidence for association with the ISG but were missed in the manual search. In addition, 13 proteins not included in the proteomes, but with experimental data supporting localization in the ISG, were identified by a manual literature search. Structures for these 56 proteins were identified or modeled (Figure 1B), and were then used to build 3D models of idealized immature and mature ISGs. Finally, X-ray tomograms were simulated from the models and used to classify features in experimental tomograms of pancreatic beta cells.

METHODS

Confidence Scores for Proteomics Data

Results from three proteomes were reconciled using a simple scoring function that combines the proteomics statistics and location annotations into a confidence value. In brief, high scores are assigned to proteins found in multiple proteomes, that have annotations related to secretory vesicles, and that interact with proteins that also have annotations related to secretory vesicles. The set of 29 manually-curated proteins (Suppl. Table 1) were assigned as true positives, and remaining proteins from the three proteomes were assigned as trial negatives. Then, for each location annotation (loc = “nucleus”, “cytoplasm”, etc), a location score (LocScore(loc)) was evaluated:

$TPR(loc) = P(loc) / P_{tot}$	true positive rate
$P(loc)$	number of true positives with a particular location annotation
P_{tot}	total number of true positives

$FPR(loc) = N(loc) / N_{tot}$ false positive rate
 $N(loc)$ number of false positives with a particular location annotation
 N_{tot} total number of negatives

$$LocScore(loc) = [TPR(loc) - FPR(loc)] / [TPR(loc) + FPR(loc)]$$

A similar score, $IntLocScore(intloc)$, was calculated for the location of proteins that have been annotated as interacting with the protein of interest. This is based on the assumption that interacting proteins will be found in the same compartment. The confidence of a protein is calculated by:

$$confidence = N_{proteome}/3. + W_{loc} * \sum LocScore(loc) + W_{intloc} * \sum IntLocScore(intloc)$$

where the sum is performed over all of the location annotations for the particular protein and $N_{proteome}$ is the number of proteomes that include the protein.

A parameter sweep of the two weights was performed, yielding values of $W_{loc}=12.$ and $W_{intloc}=1.$ for the best ROC value over the entire data set. The entire set of proteins was then rescored based on these weights, and the scores were used for ranking the set. Given the small size of the set, cross-validation studies were not performed, so we expect that the resulting high scores of the true positives will be due to bias from use as the training set.

Location annotations were found by programmatically querying UniProt (uniprot.org) with gene names from each protein and rat taxid 10116 to get the UniProt entry with the highest annotation score, and extract location annotations listed under “Keywords - Cellular component”. We then programmatically queried the UniProt, StringDB, Biogrid, and Intact

databases to get a list of proteins that interact with the protein being evaluated, and extracted location annotations for these in the same way. This process was repeated for human taxid 9096 and mouse taxid 10090. Annotations and fractions of positives and negatives for the training set are included in Suppl. Table 2.

Protein Structures and Concentrations

To estimate relative concentrations within each proteomic study, “Mascot Protein Score” values were used for proteome 1 (Brunner *et al.* 2007), “score” values from proteome 2 (Hickey *et al.* 2009), and “Intensity HiC12mC13” values for proteome 3 (Schvartz, Brunner, *et al.* 2012). These values were divided by the length of the protein and then by the sum of all values to create a “Normalized Spectral Abundance Factor” (NSAF) for each entry, which indicates roughly the molar fraction for each protein detected in the sample. Each protein’s NSAF was then multiplied by its molecular weight and normalized again to determine its mass fraction. Proteins that were identified by manual curation but not observed in the proteomics studies were assigned arbitrarily low concentration values. Finally, the total mass fraction of non-insulin proteins was scaled to give a total mass fraction of 0.2, with insulin the crystal accounting for the remaining 0.8 mass fraction (Hutton 1989).

Absolute copy numbers for proteins were calculated for a generic ISG with a crystal diameter of 200 nm (Zhang *et al.* 2020). The insulin crystal is based on entry 1trz from the RCSB Protein Data Bank (rcsb.org). This crystal form has insulin hexamers packed into a rhombohedral H3 lattice, the most common space group observed for hexameric insulin structure determinations. A value of 1.22 g/cm³ for proteins was used (Andersson and Hovmöller 1998), with the asymmetric unit molecular weight obtained from the RCSB PDB structure summary page, yielding a volume occupancy of 0.686. The second most prevalent space group is a P2₁ lattice as exemplified with PDB ID 1ev6, which yields a volume occupancy of 0.584. Manual

exceptions were made for several proteins: IAPP abundance was assigned 1% of insulin abundance (Caillon *et al.* 2016); the copy number of vATPase was calculated by averaging copy numbers of the subunits with appropriate stoichiometry; and the experimental copy number for carboxypeptidase E from (Schvartz, Brunner, *et al.* 2012) was omitted from the copy number calculation because it is ~100 times greater than the other two proteomes.

An initial set of structures was gathered automatically from online databases using csvcomplete10.0.9

(<https://github.com/brettbarbaro/csvcomplete/blob/master/csvcomplete10.0.9.py>). The UniProt RESTful API was used to retrieve sequences, and N-terminal signal sequences annotated in UniProt were automatically removed. The resulting sequences were then used to query the RCSB PDB BLAST-based RESTful API and return the top 10 structure matches in the PDB database. Subsequent manual curation was required in several cases. Multiple files were needed to model cytoplasmic and lumen domains of Syt5, Pam, Atp6ap, and EphA, and transmembrane segments were taken from integrin (PDB ID 2k1a). Similarly, files representing cytoplasmic and lumen domains of Ptpn and Ptpn2 were combined with transmembrane segments from receptor tyrosine kinase ErbB1 (PDB ID 2m0b). All of the Vamps were modeled with Vamp-2, as PDB ID 2kog includes the entire protein in a lipid-bound environment. Granins were treated as coiled-coils based on biophysical studies of chromogranin-A (Mosley *et al.* 2007). Information was difficult to find for several proteins, resulting in poor or fragmentary structures for Nucb1, Nucb2, Dnajc2, Stc1, and Nptx1.

As we were completing this project and drafting this manuscript, structures generated by AlphaFold2 (Jumper *et al.* 2021) became readily available through UniProt. We reevaluated the entire structural proteome to determine if these predicted structures added value. In the most challenging cases, such as the granins, AlphaFold2 provided structures that are largely

unfolded and of low confidence. Predicted structures of Nucb1, Nucb2, Dnajc2, Stc1, and Nptx1 all showed a well-folded core similar to the existing homologous structures found by the methods above, flanked by long disordered regions of low confidence. These low confidence regions could reflect intrinsic disorder in these structures or deficiencies in the AlphaFold2 method due to, for example, unmodeled oligomerization. Ultimately, we chose to use one structure from AlphaFold2, for CD36, which showed a well-folded domain and two membrane-spanning helices, all consistent with UniProt annotations and existing partial structures. We envision that AlphaFold2 will become a more integral part of the structural pipeline in future, as methods are developed to create credible models that incorporate oligomerization, definition of membrane-spanning regions, and intrinsic disorder.

The cytoplasm was modeled using 50 cytoplasmic proteins with highest abundance in a recent proteomic study (Beck *et al.* 2011) (Suppl. Table 3). Relative abundances of these proteins were normalized to a total concentration of 0.2g/ml (Luby-Phelps 1999).

Model Generation

Models were generated using a modified version of our instant distribution software (Klein *et al.* 2018), based on cellPACK (Johnson *et al.* 2015), similarly to the method used for building a full model of *Mycoplasma genitalium* cell (Maritan *et al.* 2022). During the packing process, a reduced representation of each protein is used, comprised of a list of representative beads or spheres selected using K-means clustering manually tuned to give a good ratio of number of beads to coverage of the entire protein. This manual tuning is particularly important for proteins with high aspect ratios, and is facilitated by the use of the web-based curation tool Mesoscope (Autin *et al.* 2020). The beads are used to (i) estimate the space occupied by each protein in a master grid and (ii) relax overlapping proteins.

Molecules are distributed by a parallel algorithm that partitions the available space on a grid, places molecules, then performs a local relaxation to resolve conflicts. However, the insulin granule can consist of up to 2 million beads, so local relaxation couldn't be applied due to limitations in the NVIDIA Flex library to ~1 million beads. Thus, we exported and relaxed the model using LAMMPS (Thompson *et al.* 2022) through Langevin rigid body dynamics. A custom approach based on cellPAINT (Gardner *et al.* 2018, 2021) is used for the membrane proteins. Each membrane protein is augmented with two triads of beads on either side of the membrane that constrain the protein within the membrane during the simulation. All bead interactions used a soft potential (https://docs.lammps.org/pair_soft.html).

The ultrastructure of the vesicle and boundary of the insulin crystal are defined as primitive signed distance fields (e.g. soft Boolean operation of spheres) or computed signed distance fields from a user-defined polygonal mesh (e.g. obtained from segmentation). In the current work, we used a spherical boundary for the crystal, and insulin was placed procedurally using the rhombohedral H3 lattice of PDB ID 1trz.

The instant packing method also includes a method for distributing lipids, using an approach similar to LipidWrapper (Durrant and Amaro 2014). The vesicle membrane is represented by a spherical signed distance field, polygonized with a dual contour algorithm. The resulting triangulated surface is then tiled using small patches of lipids cookie-cut from equilibrated flat lipid bilayers. The operation is run in parallel on the gpu for each triangle giving interactive performance.

Segmentation of X-ray Tomograms

X-ray tomograms were obtained from the PBCC website (pbccconsortium.isrd.isi.edu). These datasets have a voxel width of 37.42 nm. Segmentation files were also obtained from the PBCC

that define the location of the cell boundary, nucleus, endoplasmic reticulum, and ISG locations. Blob detection was performed with VISFD (github.com/jewettaij/visfd/releases/tag/v0.19.10) in several steps. First, the position and orientation of the glass tube was identified automatically, and a mask created to omit areas within 7-10 voxels of the tube from blob detection and visualization. A mask was created to identify cytoplasmic regions inside the cell, using the PBC manual segmentation of the cell membrane and nucleus. Additional manual curation removed small features that were disconnected from the bulk of the cell. Blob detection was then performed on the maps, and blobs that overlap and blobs with weak scores were discarded. The threshold for discarding weak blobs was obtained by tuning the threshold for each map individually, choosing threshold values that correctly classify a test set of 30 manually-chosen positives and 30 manually-chosen negative decoys. Finally, analysis of blobs was performed, including a radial density profile centered on the brightest pixel in the blob, distance from the cell surface, and classification by brightness.

Simulation of Absorption of X-rays

From (Ekman *et al.* 2018), image formation in soft X-ray tomography can be approximated as:

$$-\ln \frac{I_{\text{im}}}{I_{\text{im}}^0} \approx \mathbf{A}_h \boldsymbol{\mu},$$

where $\mathbf{A}_h$ is a linear projection matrix incorporating the 3D point spread function (PSF) of the system and $\boldsymbol{\mu}$ is a discretized vector representation of the Linear Attenuation Coefficient (LAC) that is calculated from the atomistic model of the vesicle. Values of the Mass Attenuation Coefficient (MAC, μ/ρ where ρ is the density) at a given photon energy (here 517 eV) are typically available for different materials (Henke *et al.* 1993) (henke.lbl.gov/optical_constants/atten2.html), and may be combined using the mixture rule, which is a mass-fraction-weighted average of the MAC of each component:

$$(\mu/\rho)_{\text{voxel}} = W_{\text{protein}} (\mu/\rho)_{\text{protein}} + W_{\text{lipid}} (\mu/\rho)_{\text{lipid}} + W_{\text{water}} (\mu/\rho)_{\text{water}}$$

where W_{protein} , W_{lipid} , and W_{water} are the mass fraction of the component materials. The voxel protein LAC μ_{voxel} is then calculated as the mass attenuation coefficient in the voxel $(\mu/\rho)_{\text{voxel}}$ times the protein density in the voxel ρ_{voxel} (protein weight / voxel volume):

$$\mu_{\text{voxel}} = (\mu/\rho)_{\text{voxel}} * \rho_{\text{voxel}}$$

MAC values were calculated explicitly using the atomic composition of proteins in each voxel of the model, again using the mixture rule and the weight fraction of MAC values for each of the atom types. Lipids were based on DOPC ($\text{C}_{40}\text{H}_{80}\text{NO}_8\text{P}$), with MAC value of 9264.835/cm and a volume of 1150.0 Å³ per lipid (Greenwood *et al.* 2006), and a value of 1114.279/cm was used for water.

To generate the final images, the initial calculated volume is embedded in a larger volume the size of the experimental data. The simulated reconstructions were obtained similar to the experimental ones: the projection images were distorted by Poisson noise corresponding to the shot noise of the experiment, and random translations were added to the tomograms to mimic sub-pixel alignment errors of the image registration (Chen *et al.* 2022). The reconstruction was repeated 10 times for each volume and averaged for calculation of the radial profiles.

RESULTS AND DISCUSSION

Combining Proteomic and Annotation Information

We manually curated a proteome based on the current state of knowledge for the system. This proteome is based largely on a comprehensive review (Suckale and Solimena 2010), followed

by literature searching to find supporting reports on each molecule. This includes 29 familiar proteins, including two forms of insulin, granins, enzymes involved in maturation, regulatory molecules, and a variety of membrane-bound transporters and fusion-related proteins. These 29 proteins are included in Suppl. Table 1 with citations for the studies localizing them to the ISG.

Manual curation, however, is intrinsically limited by the vagaries of literature search methods and user bias, so we developed a method to identify bona fide ISG proteins from proteomics data, removing false positives from ISG isolation impurities. Three proteomes were used in the current study, with a total of 270 proteins, of which eight are observed in all three (two forms of insulin; converting enzymes Pcsk2 and Cpe; granins ChgA, ChgB and Scg2; nucleobindin-2). The studies used different protocols for isolation of ISG: proteome 1 used a gradient (Brunner *et al.* 2007), proteome 2 added an affinity purification step (Hickey *et al.* 2009), and proteome 3 used a 3-step gradient purification along with Stable Isotope Labeling with Amino acids in Cell culture (SILAC) (Schvartz, Brunner, *et al.* 2012). Proteome 3 identified 668 proteins, but we included the 140 that were considered “specific to ISGs”, as done in a meta-analysis of all three studies (Crèvecoeur *et al.* 2015).

The simplest approach to reconcile these differences would be to use the number of proteomic observations as a weighting factor. However, this has potential problems. For example, the converting factor Pcsk1, a necessary component of the ISG, showed up in only one study, but several subunits of mitochondrial ATP synthase, which is most likely a mitochondrial impurity, showed up in two. We developed a ROC-type scoring method that combines the number of proteomic observations and expert annotations of the cellular location. The study starts with the full set of proteins from the three proteomes, and assigns our manually-curated set as true positives and the rest as trial negatives. We then evaluate each location annotation for its ability to distinguish between the two. Annotations like “cytoplasmic vesicle” have power to

discriminate true positives, and “mitochondrion” and “nucleus” are found for trial negatives (see Suppl. Table 2). A simple scoring function was developed to quantify this discrimination. Finally, an overall confidence score was developed that combines these location annotation scores with the number of times each protein is observed in a proteome, as described in Methods.

A manual literature search was then performed for all proteins with confidence > 0.333 , as well as all proteins present in two or more proteomes. As described below, in some cases evidence for presence in the ISG was found, in other cases evidence was found for presence in other compartments of the cell, and a few were ambiguous.

VAMPs are necessary for the fusion of vesicles with the membrane. Vamp2, Vamp3 and Vamp8 were included in our curated set. Vamp2 and Vamp3 scored well, but Vamp8, which has been immunolocalized to the ISG (Zhu *et al.* 2012), showed a mid-range score of 0.21, due to being found in only one proteome and having lysosome and endosome localizations. Lamp2 also showed an intermediate score (0.76), but we could find no evidence for its presence in the ISG.

Rabs and similar membrane-associated proteins play an essential role in regulation of the vesicle life cycle. Rab37, Rap1a, and Rab3a were all included in the curated list, and showed high to intermediate scores. Immunolocalization studies were found for Rph3a1 (Matsunaga *et al.* 2016), Rab3c, and Rab3d (Iezzi *et al.* 1999). Rab27a is connected to secretory vesicles in general (Suckale and Solimena 2010), but no evidence was found for several Rab5 members, Rab35, or Rhog. Rab1a, Rab2a, and several G-protein subunits were found in two proteomes but received low scores because of localization, and we could not find evidence for their presence in the ISG. Two additional proteins, Rac1 and Cdc42, were included based on the literature search (Wang and Thurmond 2009), although they were not included in any of the proteomes.

Several proteins were not included in the original curated list, but strong evidence for their inclusion was found in the subsequent literature search. Peptide-amidating enzyme PAM (Garmendia *et al.* 2002), neurosecretory protein Vgf (Stephens *et al.* 2017), neuronal pentraxin-1 (Schvartz, Couté, *et al.* 2012), and stanniocalcin-1 (Zaidi *et al.* 2012) all received high to intermediate scores. Conversely, several proteins were included in the original list, but scored poorly. Two nucleobindins, which have been placed in the ISG by immunolocalization (Ramesh *et al.* 2015), scored low due to nuclear and ER annotations, in spite of nucleobindin-2 being found in all three proteomic studies. vATPase has been localized to the ISG (Sun-Wada *et al.* 2006), but most of the subunits scored poorly, due to their presence in a single proteome and a variety of conflicting localizations. Macrophage migration inhibitory factor was found in only one proteome, but has been immunolocalized to the ISG (Waeber *et al.* 1997).

Several proteins showed high scores, and evocative but non-conclusive evidence was found in the literature search. Tmem163, a putative zinc transporter (Sharma *et al.* 2017), and Enpp2 (Gorelik *et al.* 2017) have connections to type 2 diabetes. Depalmitase Abhd17 (Won *et al.* 2018), Dnajc5 (Gorenberg and Chandra 2017), and Wnt ligands like Wif1 are involved in the secretory process (Schinner *et al.* 2007). These may be candidates for immunolocalization studies.

A variety of proteins were found in two of the proteomic studies, but appear to be impurities from other compartments in the cells. These include lysosomal proteins alpha-glucosidase, septin-11, and cathepsin D; cytoplasmic enzymes ATP-citrate synthase, fructose biphosphate aldolase A, and glyceraldehyde-3-phosphate dehydrogenase; a subunit of mitochondrial ATP synthase; and tubulin and fibronectin. All showed low scores due to the localization annotations and were culled from the list used for modeling.

Models of the Insulin Secretory Granule

Idealized models of mature and immature vesicles were created with the final proteome of 56 proteins (Suppl. Table 1), an insulin crystal with diameter 200 nm (Zhang *et al.* 2020), and vesicle diameter of 320 nm (Suckale and Solimena 2010). The initial model created with the interactive method of CellPACKgpu roughly places molecules in the proper compartments, but has a number of steric contacts, particularly with highly asymmetric molecules such as the granins. Subsequent optimization with LAMMPS resolves these contacts to create the final model. Relaxation required between 1 and 2 hours using 30 threads, followed by around ~30 minutes to recompute the position/rotation of the protein instance from the bead coordinates for each frame of the simulation (CPU AMD Ryzen Threadripper 1950X 16-Core Processor, GPU NVIDIA Quadro R8000). The final models reveal the ISG as being densely packed with soluble proteins and with a protein-rich bounding membrane. The full model (idealized mature and immature granules surrounded by cytoplasmic proteins) consists of 424384 individual proteins represented by 2343538 beads of radius 17.0 Å.

Interpreting Experimental Soft X-ray Tomograms

Figure 2 demonstrates use of this methodology to interpret experimental soft X-ray tomograms. Tomograms of whole pancreatic beta cells show many easily-distinguishable features with high LAC values. The ones with the highest LAC values are presumed to be lipid droplets, and the remainder are insulin secretory vesicles. We address two questions that have been posed with the experimental work.

First, there has been some question about the contribution of the membrane to the observed features. We created models and simulated tomograms for vesicles with and without the membrane (Figure 2) and calculated a 2D radial profile for all four volumes. These show that the

membrane for these idealized vesicles accounts for 10-20% of the observed X-ray absorbance in these features. For the full model with lipids, the projected mature blob peak value is $0.538/\mu\text{m}$ for mature and $0.437/\mu\text{m}$ for immature, and the surrounding cytoplasm has an average value of $0.35/\mu\text{m}$.

Second, it has been noted that immature vesicles, which do not include the dense crystal of insulin, may not show enough contrast to be observed in experimental soft X-ray tomograms. As seen in Figure 2, immature vesicles are less visible in simulated tomograms. When we applied our VISFD segmentation protocol to these simulated tomograms, only the mature vesicles were detected with a contour diameter of 184 nm, while the immature vesicles were undetected. This result indicates that a pool of immature vesicles may not be detected with the VISFD segmentation protocol in the analysis described below. We also evaluated the need for the relaxation step in model generation, when these models are used to simulate soft X-ray tomograms, and as seen in the graph in Figure 2, unrelaxed the full models and unrelaxed models show very similar radial profiles. In the following larger study, a quick optimization was performed to ensure that all lumen proteins are inside the membrane, but a full relaxation was not performed.

We then estimated the size and insulin content of all visible vesicles in an experimental whole-cell tomogram (cell 766_5, Figure 3A). We segmented the experimental volume with VISFD, resulting in 1136 features interpreted to be vesicles (Figure 3E). We generated a library of 3456 models to sample potential variations in vesicle and cytoplasmic properties (four hours of computation). Sixteen vesicle diameters were sampled from 130 nm to 473 nm. For each vesicle size, six different insulin crystal sizes were sampled and the number of soluble and membrane proteins was scaled based on the number of insulin molecules in the crystal. Six

steps of transition between mature and immature vesicles were generated by partitioning this amount of insulin as mature insulin in a crystal, proinsulin in immature vesicles, and cleaved insulin monomer in the lumen in transitional forms. Six different concentrations of cytosolic proteins were also scanned from 0.06-0.2 g/ml. A quick relaxation of 150 iterations was applied on the gpu to force proteins (in particular, the highly extended granins) inside the vesicles. Finally, model LAC values were computed for each of these sample vesicles in a 636 nm bounding box (17*17*17 voxels with voxel size of 37.42 nm), then embedded in a surrounding volume of size 247*17*247 voxels (Figure 3BC). Each volume was projected and reconstructed ten times (37 hours of computation), used to calculate an average 2D radial profile, and an R^2 score between the experimental profile and the simulated profile was calculated, yielding a “confidence level” in the assignment. R^2 scores are calculated for six radial distance points to focus on the vesicle profile, and for 15 points to include information on the surrounding cytoplasmic LAC. The simulated profile with the maximum R^2 score was selected for each of the 1136 features in the experimental tomogram.

We obtained an average R^2 score of 0.94 ± 0.1 across all 1136 experimental blobs, with 83% with a score > 0.93 and 68% with a score > 0.96 . Looking at the steps of transition, 128 blobs are assigned as immature vesicles, 747 in transitional states, and 261 assigned as mature vesicles (Figure 3F). This predominance of immature and transitional forms is consistent with recent cryoelectron tomography studies, which found that the percentage of mature vesicles ranged from 12% to 32% depending on the location in the cell (Zhang *et al.* 2020). In addition, note that this analysis may be missing a pool of immature vesicles that are not visible in the soft X-ray tomographic experiment, as described above. Several examples are shown in detail in Figure 4.

OUTLOOK

The convergence of experimental methodology and computational capability is putting whole-cell structural modeling within reach. Using this methodology, it is becoming possible to interpret entire cellular tomograms with integrative molecular models. In Figure 3, an entire X-ray tomogram has been segmented to identify mature secretory vesicles, and models have been generated with crystal sizes that match the observed absorption of the vesicle. This model is a first step toward a quantitative mesoscale interpretation of this cell. For example, in cell 766_5 from the PBCC, the observed features include a total of ~606,000,000 insulin monomers across the cell, with an average of 533,000 +/- 253,000 per vesicle, and with ~125,000,000 out of the total in crystal form.

One of the goals of our work on the CellPACK suite is to create tools that will be widely usable, however these models are straining the current capabilities of consumer-level computational hardware and software. The CellPACK suite heavily leverages advances from the gaming community for use of gpu hardware, which underlies the instant packing algorithm used to generate initial models (Klein *et al.* 2018). For the relaxation/optimization steps, the size of these models required a move from the interactive methods of Flex to the traditional batch mode of LAMMPS. We thus developed an initial simplistic input for LAMMPS that will serve as the foundation for more advanced simulation of similar large systems. Finally, visualization before and after simulation is provided by cellVIEW (Muzic *et al.* 2015) within our cellPACK application built within Unity. Simulation can be cached and played back directly in our application. Final models are saved/exported in different resolution and file format (.bin, .pdb, .cif) to enable visualization with other molecular graphics software (e.g. OVITO, VMD, UCSF Chimera, Mol*). These files typically include coordinates for each type of protein along with transformation information for placing all of the instances of the protein into the overall model. However, as of today only Mol* (Sehna *et al.* 2018) is capable of reading and visualizing the full all-atom model

(Figure 5) and this at a low frame rate using a coarse Gaussian surface representation of the molecules (for example, visit the GitHub site included below to see Mol* visualization of the mature and immature vesicles). Other molecular graphics tools were able to load coordinate files for each type of protein but currently failed to build the many instances of each protein to visualize the entire model.

Of course, many challenges remain as we move forward in this mesoscale structural view of a cell. As is always the case with biology, there are numerous unique features for each of the compartments of the cell, and new methodologies will be needed, for example, to incorporate complex structural features like intrinsic disorder in nuclear pores, the cytoskeleton and its interactions with organelles, the dynamic endomembrane system and the proteins that manage it, and the many structural and functional states of chromatin. With all of this, structural mesoscale modeling will proceed with a combination of automated methods when possible, and manual attention when necessary.

Availability

Models and volumes for the mature and immature vesicles are available at <https://github.com/ccsb-scripps/ISG>. Links to Mol* visualizations are also available for mature and immature models at the GitHub site.

Acknowledgements and Financial Support

We thank Kate L. White, Xianjun Zhang, Carolyn Larabell, Yu Gao, and Raymond C. Stevens for helpful discussions. This work was supported by NIH R01 GM120604 (DSG) and the Pancreatic Beta Cell Consortium (BB). The National Center for X-ray Tomography was supported by NIH NIGMS (P30GM138441) and DOE Office of Biological and Environmental Research (DE-AC02-5CH11231), and is located at the Advanced Light Source, a U.S. DOE

Office of Science User Facility under contract no. DE-AC02-05CH11231. RCSB PDB (DSG) is funded by the National Science Foundation (DBI-1832184), US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health (R01GM133198).

Author Contributions

Investigation: LA, BB, AIJ, SV, AJO, DSG

Methodology: LA, BB, AIJ, AE, SV, DSG

Writing: LA, BB, AIJ, AE, AJO, DSG

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Information

Supplementary Table 1. Insulin Secretory Granule Proteins

Supplementary Table 2. UniProt Localization Annotations for Proteome Confidence Scoring Function

Supplementary Table 3. Cytoplasmic Proteins

References

Andersson KM, Hovmöller S (Eds) (1998) The average atomic volume and density of proteins.

Zeitschrift Für Kristallographie - Crystalline Materials **213**, 369–373.

doi:10.1524/zkri.1998.213.7-8.369.

- Autin L, Maritan M, Barbaro BA, Gardner A, Olson AJ, Sanner M, Goodsell DS (2020) Mesoscope: A web-based tool for mesoscale data integration and curation. In Workshop on Molecular Graphics and Visual Analysis of Molecular Data, (ed. J. Byska, M. Krone, B. Sommer), pp. 23-31. Geneva: The Eurographics Association.
doi:10.2312/MOLVA.20201098.
- Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (2011) The quantitative proteome of a human cell line. *Molecular Systems Biology* **7**, 549. doi:10.1038/msb.2011.82.
- Brunner Y, Couté Y, Iezzi M, Foti M, Fukuda M, Hochstrasser DF, Wollheim CB, Sanchez J-C (2007) Proteomics analysis of insulin secretory granules. *Molecular & Cellular Proteomics* **6**, 1007–1017. doi:10.1074/mcp.M600443-MCP200.
- Caillon L, Hoffmann ARF, Botz A, Khemtemourian L (2016) Molecular structure, membrane interactions, and toxicity of the islet amyloid polypeptide in type 2 diabetes mellitus. *Journal of Diabetes Research* **2016**, 1–13. doi:10.1155/2016/5639875.
- Chen J-H, Vanslembrouck B, Loconte V, Ekman A, Cortese M, Bartenschlager R, McDermott G, Larabell CA, Le Gros MA, Weinhardt V (2022) A protocol for full-rotation soft X-ray tomography of single cells. *STAR Protocols* **3**, 101176. doi:10.1016/j.xpro.2022.101176.
- Crèvecoeur I, Rondas D, Mathieu C, Overbergh L (2015) The beta-cell in type 1 diabetes: What have we learned from proteomic studies? *PROTEOMICS - Clinical Applications* **9**, 755–766. doi:10.1002/prca.201400135.
- Durrant JD, Amaro RE (2014) LipidWrapper: An algorithm for generating large-scale membrane models of arbitrary geometry (RF Murphy, Ed.). *PLoS Computational Biology* **10**, e1003720. doi:10.1371/journal.pcbi.1003720.

- Ekman A, Weinhardt V, Chen J-H, McDermott G, Le Gros MA, Larabell C (2018) PSF correction in soft X-ray tomography. *Journal of Structural Biology* **204**, 9–18.
doi:10.1016/j.jsb.2018.06.003.
- Eliasson L, Abdulkader F, Braun M, Galvanovskis J, Hoppa MB, Rorsman P (2008) Novel aspects of the molecular mechanisms controlling insulin secretion: Exocytosis in pancreatic β -cells. *The Journal of Physiology* **586**, 3313–3324.
doi:10.1113/jphysiol.2008.155317.
- Feig M, Sugita Y (2019) Whole-cell models and simulations in molecular detail. *Annual Review of Cell and Developmental Biology* **35**, 191–211. doi:10.1146/annurev-cellbio-100617-062542.
- Gardner A, Autin L, Barbaro B, Olson AJ, Goodsell DS (2018) CellPAINT: Interactive illustration of dynamic mesoscale cellular environments. *IEEE Computer Graphics and Applications* **38**, 51–66. doi:10.1109/MCG.2018.2877076.
- Gardner A, Autin L, Fuentes D, Maritan M, Barad BA, Medina M, Olson AJ, Grotjahn DA, Goodsell DS (2021) CellPAINT: Turnkey illustration of molecular cell biology. *Frontiers in Bioinformatics* **1**, 660936. doi:10.3389/fbinf.2021.660936.
- Garmendia O, Rodríguez MP, Burrell MA, Villaro AC (2002) Immunocytochemical finding of the amidating enzymes in mouse pancreatic A-, B-, and D-cells: A comparison with human and rat. *Journal of Histochemistry & Cytochemistry* **50**, 1401–1415.
doi:10.1177/002215540205001013.
- Germanos M, Gao A, Taper M, Yau B, Kebede MA (2021) Inside the insulin secretory granule. *Metabolites* **11**, 515. doi:10.3390/metabo11080515.

- Goodsell DS, Olson AJ, Forli S (2020) Art and science of the cellular mesoscale. *Trends in Biochemical Sciences* **45**, 472–483. doi:10.1016/j.tibs.2020.02.010.
- Gorelik A, Randriamihaja A, Illes K, Nagar B (2017) A key tyrosine substitution restricts nucleotide hydrolysis by the ectoenzyme NPP 5. *The FEBS Journal* **284**, 3718–3726. doi:10.1111/febs.14266.
- Gorenberg EL, Chandra SS (2017) The role of co-chaperones in synaptic proteostasis and neurodegenerative disease. *Frontiers in Neuroscience* **11**, 248. doi:10.3389/fnins.2017.00248.
- Greenwood AI, Tristram-Nagle S, Nagle JF (2006) Partial molecular volumes of lipids and cholesterol. *Chemistry and Physics of Lipids* **143**, 1–10. doi:10.1016/j.chemphyslip.2006.04.002.
- Henke BL, Gullikson EM, Davis JC (1993) X-ray interactions: Photoabsorption, scattering, transmission, and reflection at E = 50–30,000 eV, Z = 1–92. *Atomic Data and Nuclear Data Tables* **54**, 181–342. doi:10.1006/adnd.1993.1013.
- Hickey AJR, Bradley JWI, Skea GL, Middleditch MJ, Buchanan CM, Phillips ARJ, Cooper GJS (2009) Proteins associated with immunopurified granules from a model pancreatic islet β -cell system: Proteomic snapshot of an endocrine secretory granule. *Journal of Proteome Research* **8**, 178–186. doi:10.1021/pr800675k.
- Hutton JC (1989) The insulin secretory granule. *Diabetologia* **32**, 271–281. doi:10.1007/BF00265542.

- Iezzi M, Escher G, Meda P, Charollais A, Baldini G, Darchen F, Wollheim CB, Regazzi R (1999) Subcellular distribution and function of Rab3A, B, C, and D isoforms in insulin-secreting cells. *Molecular Endocrinology* **13**, 202–212. doi:10.1210/mend.13.2.0228.
- Im W, Liang J, Olson A, Zhou H-X, Vajda S, Vakser IA (2016) Challenges in structural approaches to cell modeling. *Journal of Molecular Biology* **428**, 2943–2964. doi:10.1016/j.jmb.2016.05.024.
- Johnson GT, Autin L, Al-Alusi M, Goodsell DS, Sanner MF, Olson AJ (2015) cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nature Methods* **12**, 85–91. doi:10.1038/nmeth.3204.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. doi:10.1038/s41586-021-03819-2.
- Klein T, Autin L, Kozlikova B, Goodsell DS, Olson A, Groller ME, Viola I (2018) Instant construction and visualization of crowded biological environments. *IEEE Transactions on Visualization and Computer Graphics* **24**, 862–872. doi:10.1109/TVCG.2017.2744258.
- Loconte V, Singla J, Li A, Chen J-H, Ekman A, McDermott G, Sali A, Le Gros M, White KL, Larabell CA (2022) Soft X-ray tomography to map and quantify organelle interactions at the mesoscale. *Structure* **30**, 510–521.e3. doi:10.1016/j.str.2022.01.006.

- Luby-Phelps K (1999) Cytoarchitecture and physical properties of cytoplasm: Volume, viscosity, diffusion, intracellular surface area. *International Review of Cytology* **192**, 189–221. doi:10.1016/S0074-7696(08)60527-6.
- Maritan M, Autin L, Karr J, Covert MW, Olson AJ, Goodsell DS (2022) Building structural models of a whole mycoplasma cell. *Journal of Molecular Biology* **434**, 167351. doi:10.1016/j.jmb.2021.167351.
- Matsunaga K, Taoka M, Isobe T, Izumi T (2017) Rab2a and Rab27a cooperatively regulate transition from granule maturation to exocytosis through the dual effector Noc2. *Journal of Cell Science* **130**, 541–550. doi:10.1242/jcs.195479.
- McDermott G, Le Gros MA, Knoechel CG, Uchida M, Larabell CA (2009) Soft X-ray tomography and cryogenic light microscopy: the cool combination in cellular imaging. *Trends in Cell Biology* **19**, 587–595. doi:10.1016/j.tcb.2009.08.005.
- Mosley CA, Taupenot L, Biswas N, Taulane JP, Olson NH, Vaingankar SM, Wen G, Schork NJ, Ziegler MG, Mahata SK, O'Connor DT (2007) Biogenesis of the secretory granule: Chromogranin A coiled-coil structure results in unusual physical properties and suggests a mechanism for granule core condensation. *Biochemistry* **46**, 10999–11012. doi:10.1021/bi700704r.
- Muzic ML, Autin L, Parulek J, Viola I (2015) cellVIEW: a tool for illustrative and multi-scale rendering of large biomolecular datasets. In Eurographics Workshop on Visual Computing for Biology and Medicine, (ed. K. Buhler, L. Linsen, M.W. John), pp. 61–70. Geneva: The Eurographics Association. doi:10.2312/VCBM.20151209.

- Ramesh N, Mohan H, Unniappan S (2015) Nucleobindin-1 encodes a nesfatin-1-like peptide that stimulates insulin secretion. *General and Comparative Endocrinology* **216**, 182–189. doi:10.1016/j.ygcen.2015.04.011.
- Schinner S, Ülgen F, Papewalis C, Schott M, Woelk A, Vidal-Puig A, Scherbaum WA (2007) Regulation of insulin secretion, glucokinase gene transcription and beta cell proliferation by adipocyte-derived Wnt signalling molecules. *Diabetologia* **51**, 147–154. doi:10.1007/s00125-007-0848-0.
- Schvartz D, Brunner Y, Couté Y, Foti M, Wollheim CB, Sanchez J-C (2012) Improved characterization of the insulin secretory granule proteomes. *Journal of Proteomics* **75**, 4620–4631. doi:10.1016/j.jprot.2012.04.023.
- Schvartz D, Couté Y, Brunner Y, Wollheim CB, Sanchez J-C (2012) Modulation of neuronal pentraxin 1 expression in rat pancreatic β -cells submitted to chronic glucotoxic stress. *Molecular & Cellular Proteomics* **11**, 244–254. doi:10.1074/mcp.M112.018051.
- Sehnal D, Rose A, Koca J, Burley S, Velankar S (2018) Mol*: Towards a common library and tools for web molecular graphics. In Workshop on Molecular Graphics and Visual Analysis of Molecular Data, (ed. J. Byska, M. Krone, B. Sommer), pp. 29-33. Geneva: The Eurographics Association. doi:10.2312/MOLVA.20181103.
- Sharma V, Sharma I, Sethi I, Mahajan A, Singh G, Angural A, Bhanwer AJS, Dhar MK, Singh V, Rai E, Sharma S (2017) Replication of newly identified type 2 diabetes susceptible loci in Northwest Indian population. *Diabetes Research and Clinical Practice* **126**, 160–163. doi:10.1016/j.diabres.2017.02.013.

- Singla J, McClary KM, White KL, Alber F, Sali A, Stevens RC (2018) Opportunities and challenges in building a spatiotemporal multi-scale model of the human pancreatic β cell. *Cell* **173**, 11–19. doi:10.1016/j.cell.2018.03.014.
- Stephens SB, Edwards RJ, Sadahiro M, Lin W-J, Jiang C, Salton SR, Newgard CB (2017) The prohormone VGF regulates β cell function via insulin secretory granule biogenesis. *Cell Reports* **20**, 2480–2489. doi:10.1016/j.celrep.2017.08.050.
- Suckale J, Solimena M (2010) The insulin secretory granule as a signaling hub. *Trends in Endocrinology & Metabolism* **21**, 599–609. doi:10.1016/j.tem.2010.06.003.
- Sun-Wada G-H, Toyomura T, Murata Y, Yamamoto A, Futai M, Wada Y (2006) The $\alpha 3$ isoform of V-ATPase regulates insulin secretion from pancreatic β -cells. *Journal of Cell Science* **119**, 4531–4540. doi:10.1242/jcs.03234.
- Thompson AP, Aktulga HM, Berger R, Bolintineanu DS, Brown WM, Crozier PS, in 't Veld PJ, Kohlmeyer A, Moore SG, Nguyen TD, Shan R, Stevens MJ, Tranchida J, Trott C, Plimpton SJ (2022) LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **271**, 108171. doi:10.1016/j.cpc.2021.108171.
- Waeber G, Calandra T, Roduit R, Haefliger J-A, Bonny C, Thompson N, Thorens B, Temler E, Meinhardt A, Bacher M, Metz CN, Nicod P, Bucala R (1997) Insulin secretion is regulated by the glucose-dependent production of islet β cell macrophage migration inhibitory factor. *Proceedings of the National Academy of Sciences* **94**, 4782–4787. doi:10.1073/pnas.94.9.4782.

- Wang Z, Thurmond DC (2009) Mechanisms of biphasic insulin-granule exocytosis – roles of the cytoskeleton, small GTPases and SNARE proteins. *Journal of Cell Science* **122**, 893–903. doi:10.1242/jcs.034355.
- White KL, Singla J, Loconte V, Chen J-H, Ekman A, Sun L, Zhang X, Francis JP, Li A, Lin W, Tseng K, McDermott G, Alber F, Sali A, Larabell C, Stevens RC (2020) Visualizing subcellular rearrangements in intact β cells using soft x-ray tomography. *Science Advances* **6**, eabc8262. doi:10.1126/sciadv.abc8262.
- Won SJ, Cheung See Kit M, Martin BR (2018) Protein depalmitoylases. *Critical Reviews in Biochemistry and Molecular Biology* **53**, 83–98. doi:10.1080/10409238.2017.1409191.
- Zaidi D, Turner JK, Durst MA, Wagner GF (2012) Stanniocalcin-1 co-localizes with insulin in the pancreatic islets. *ISRN Endocrinology* **2012**, 1–6. doi:10.5402/2012/834359.
- Zhang X, Carter SD, Singla J, White KL, Butler PC, Stevens RC, Jensen GJ (2020) Visualizing insulin vesicle neighborhoods in β cells by cryo–electron tomography. *Science Advances* **6**, eabc8258. doi:10.1126/sciadv.abc8258.
- Zhu D, Zhang Y, Lam PPL, Dolai S, Liu Y, Cai EP, Choi D, Schroer SA, Kang Y, Allister EM, Qin T, Wheeler MB, Wang C-C, Hong W-J, Woo M, Gaisano HY (2012) Dual role of VAMP8 in regulating insulin exocytosis and islet β cell growth. *Cell Metabolism* **16**, 238–249. doi:10.1016/j.cmet.2012.07.001.

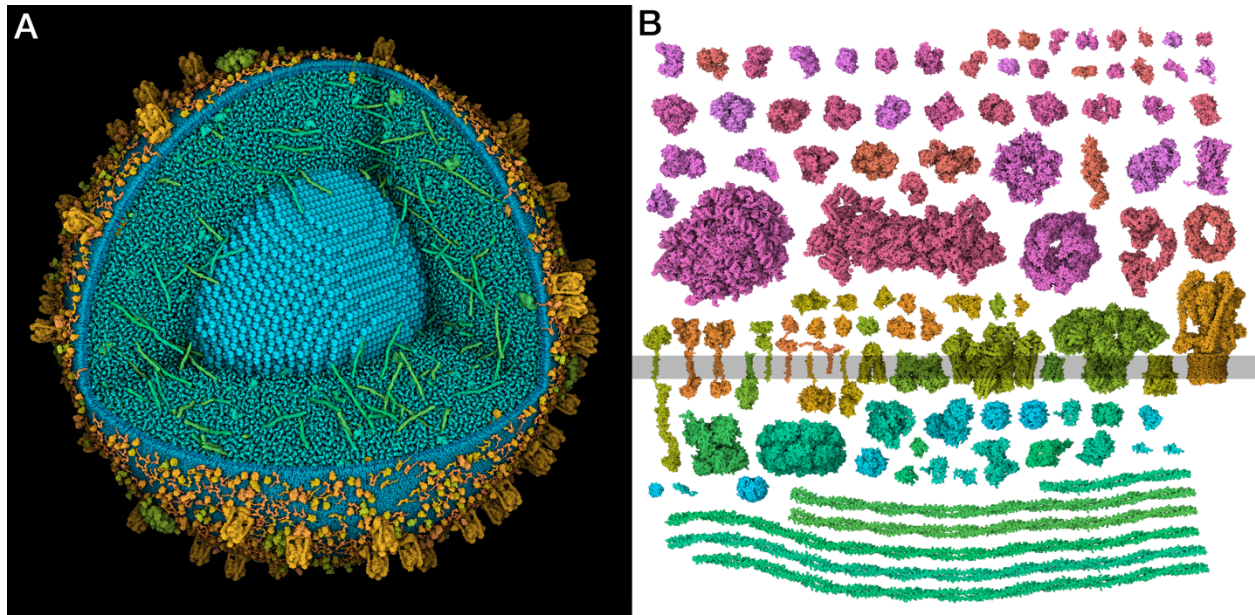


Figure 1. A) Integrative 3D model of a mature insulin secretory granule. One quadrant is clipped away to show the insulin crystal (turquoise). The long coiled-coil proteins in green are granins, and the lumen is dominated by many copies of the small beta-peptide left over from the maturation of insulin. (B) Structural proteomes used to build the models. Cytoplasmic proteins are at top in red and magenta, vesicle membrane-spanning and membrane-associated proteins are at center in orange and yellow-green, and vesicle lumen proteins are at bottom in blue-green.

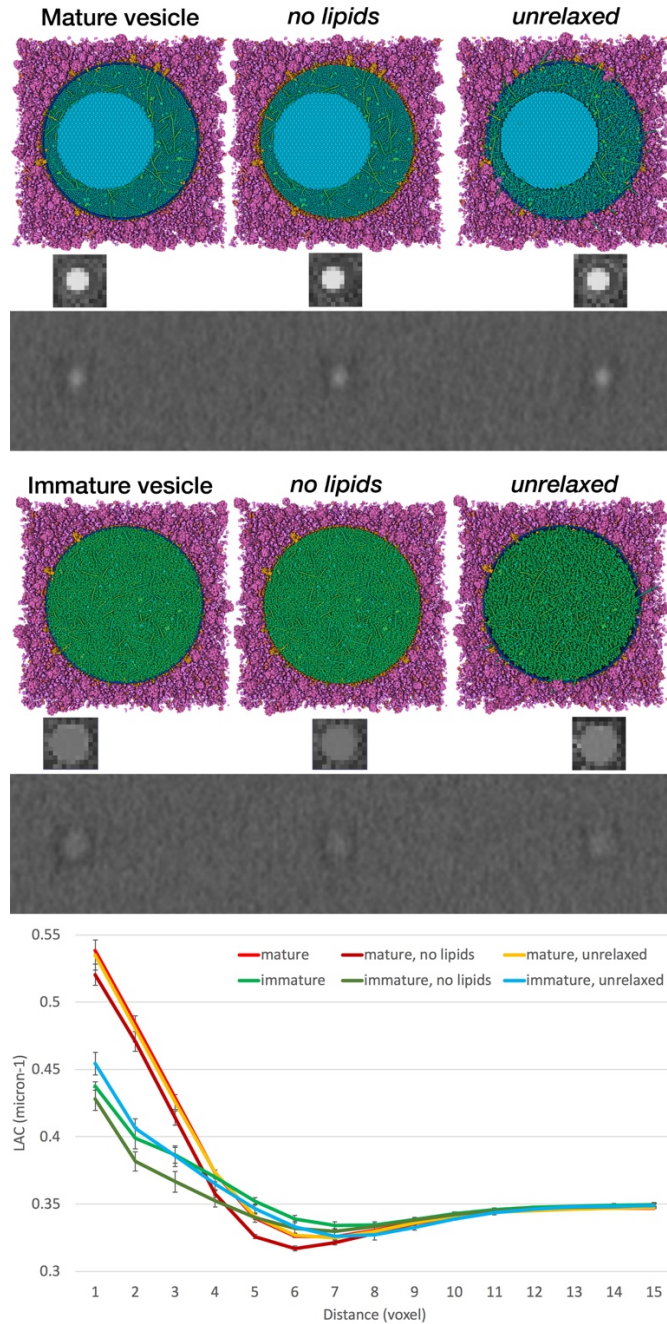


Figure 2. Simulation of X-ray tomograms. Mature and immature vesicles are shown in three states: the full model, a model without lipids, and a rough model generated without the relaxation step. The small insets are “phantoms”: voxelized representations of the simulated LAC for each vesicle. These phantoms are embedded in a large volume of cytoplasm and used to simulate tomograms that reflect the experimental imaging and processing, as shown below the models and phantoms. Radial density profiles for the six models are shown at bottom.

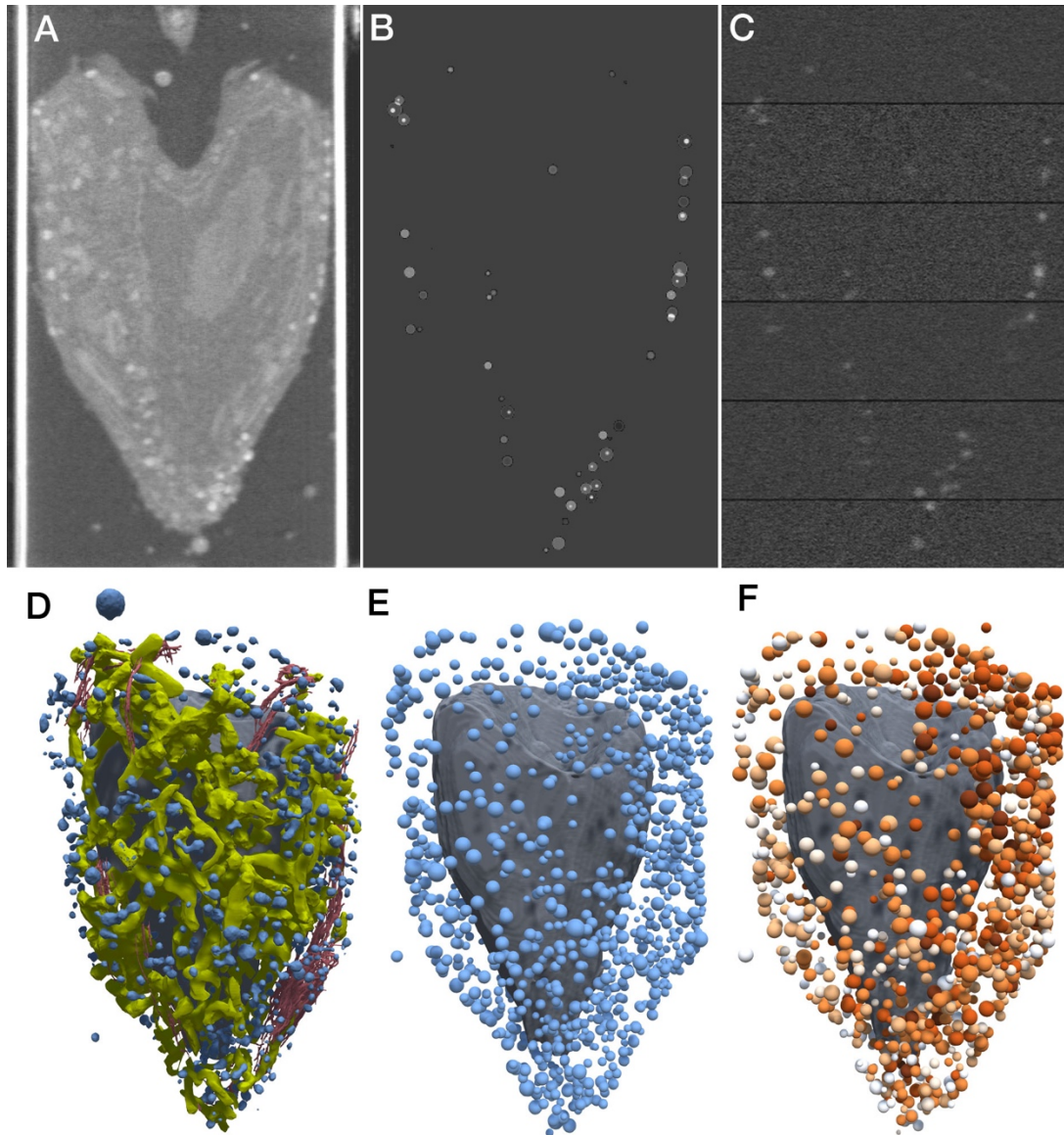


Figure 3. Interpretation of an experimental X-ray tomogram. (A) Volume rendering of a slice through the X-ray tomogram of cell 766_5 from the PBCC. Bright white vertical bands at the edges are the capillary used to hold the cell. (B) Idealized simulated absorption for this slice from vesicle models placed at features in the tomogram. (C) Simulated absorption of this slice mimicking the experimental imaging and processing. Horizontal bands are due to calculation of the volume in sections. (D) Manual segmentation from the PBCC showing mitochondria (yellow), nucleus (gray), endoplasmic reticulum (red), and vesicles (blue). (E) VISFD automatically segments blob-like features as spheres (blue). The manually-segmented nucleus

is included in gray for context. (F) Interpretation of the automated segmentation with idealized spherical vesicles, showing the predicted vesicle membrane radius and colored with mature vesicles in dark red, immature vesicles in white, and transitional forms in pastel shades.

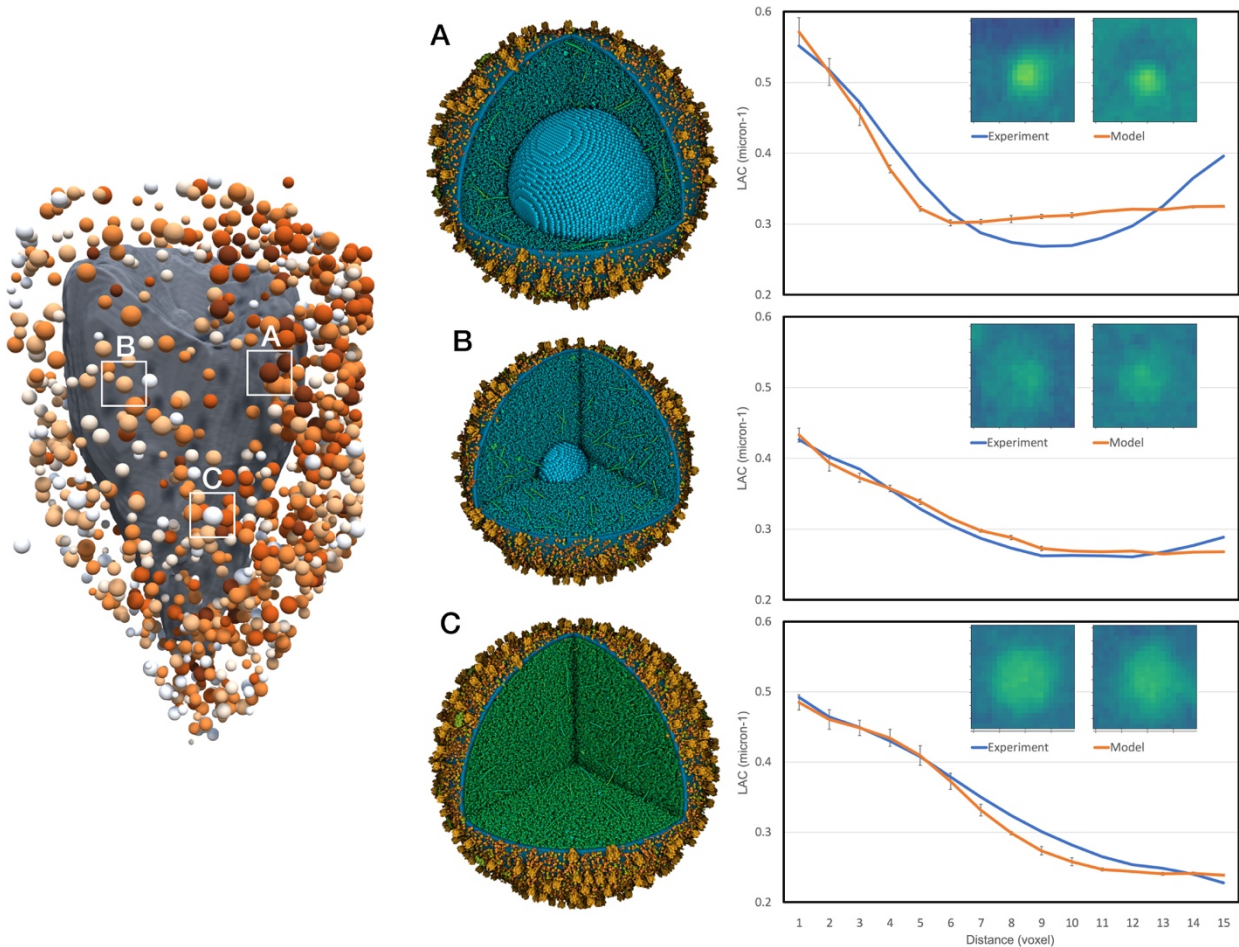


Figure 4. Interpretation of selected features in the experimental tomogram. Three vesicles are shown: (A) mature vesicle with radius 236 nm and crystal of 124 nm, and cytoplasmic LAC of 0.34/μm, (B) transitional vesicle with radius 202 nm and crystal of 50 nm, and cytoplasmic LAC of 0.28/μm, and (C) immature vesicle with radius 236 nm and cytoplasmic concentration to give an average LAC of 0.25/μm. The vesicle models are shown at center and the match of experimental (blue curve) and simulated (orange curve) radial density profiles are shown at right with a central slice through the tomograms.

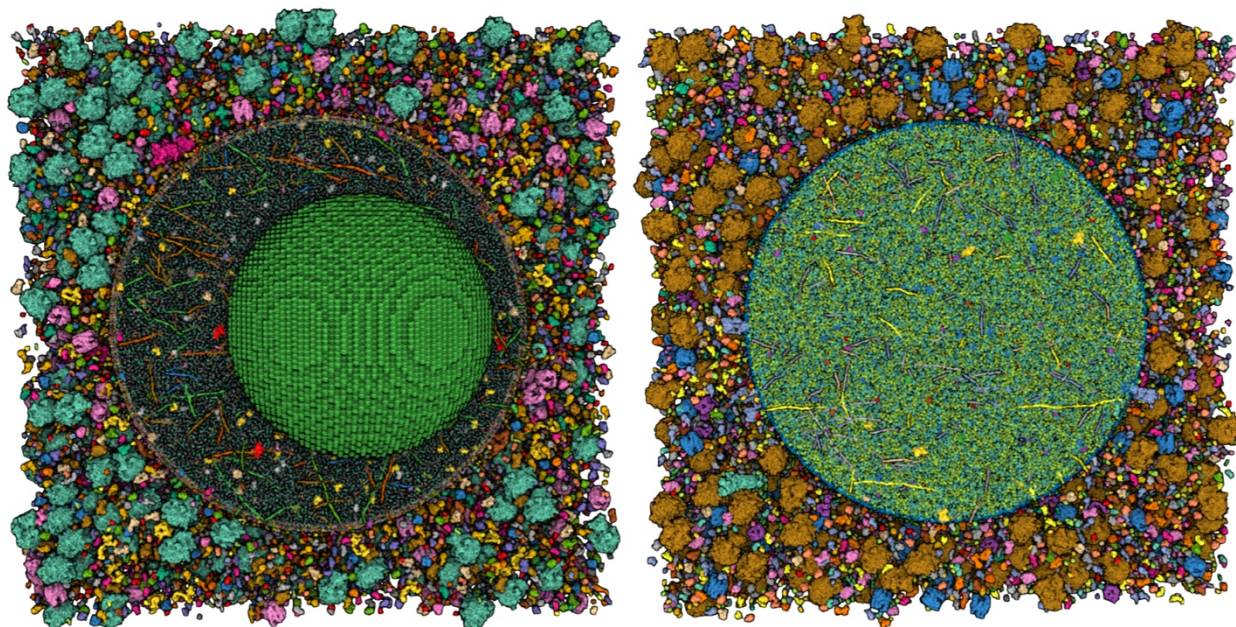


Figure 5. Idealized models of (left) mature and (right) immature vesicles viewed interactively in Mol* using a coarse Gaussian surface and colored by chain.