Integrative Modeling and Visualization of Exosomes

Julia Jiménez, Ludovic Autin, Inmaculada Ibáñez de Cáceres, David S. Goodsell

Information from proteomics, microscopy, and structural biology are integrated to create structural models of exosomes, small vesicles released from cells. Three visualization methods are employed and compared: 2D painting of a cross section using traditional media, manual creation of a cross section using the mesoscale 2.5D digital painting software cellPAINT, and generation of a 3D atomic model using the mesoscale modeling program cellPACK.

OPEN ACCESS

Keywords

Exosome, integrative structural biology, molecular visualization, non-photorealistic rendering, protein structure

Introduction

Modeling of the molecular structure of cells is rapidly moving from an art into a science. A vast amount of data is being collected and made available: genomics and proteomics are providing accurate knowledge on all of the molecular players, structural biology is determining the form, function, and assembly of these molecules, and techniques of microscopy are revealing smaller and more detailed views of intact cellular systems. When combined with the rapid development of information and computing technology, the day when we can generate and visualize a molecular model of an entire cell is within reach.

As a step towards this ambitious goal, we have been developing methods for integrating diverse experimental data into comprehensible visualizations. We began with traditional scientific illustration to create visualizations for dissemination, outreach, and education, focusing on defined portions of living cells and their interactions with viruses (Goodsell, Franzen, et al. 2018). We have recently extended this work by developing cellPAINT, a digital painting program that assists in the creation of these types of subcellular landscape illustrations (Gardner et al. 2018). Building on this work, we are currently developing tools to create and visualize accurate 3D models of the molecular structure of cellular compartments and entire cells (Johnson et al. 2015).

As part of this work, we have used small, well-defined systems to test and apply new methods. Because the work is highly integrative, a wide variety of data must be gathered (or estimated) in order to generate a convincing model. Early illustrative work focused on well-studied systems such as red blood cells and *Escherichia coli* (Goodsell 1992). For our 3D modeling work, we have used simpler systems, including HIV (Johnson et al. 2014) and mycoplasma bacteria (Goodsell, Autin, et al. 2018).

Exosomes provide an excellent test case for current mesoscale modeling methods. Sufficient proteomic, structural, and ultrastructural data are available to define a convincing model, however, there are still grey areas in this knowledge that may benefit from mesoscale modeling. Given their growing role in many disease states, many illustrations of exosomes have been created, ranging from simple schematics to highly rendered editorial and advertising imagery (an internet search of "exosome" will yield hundreds of examples). However, to our knowledge, no detailed integrative models of exosomes have been generated that are consistent with available data on scale and composition. In this report, we present the results of an internship project on the integrative modeling and visualization of exosomes using three mesoscale techniques.
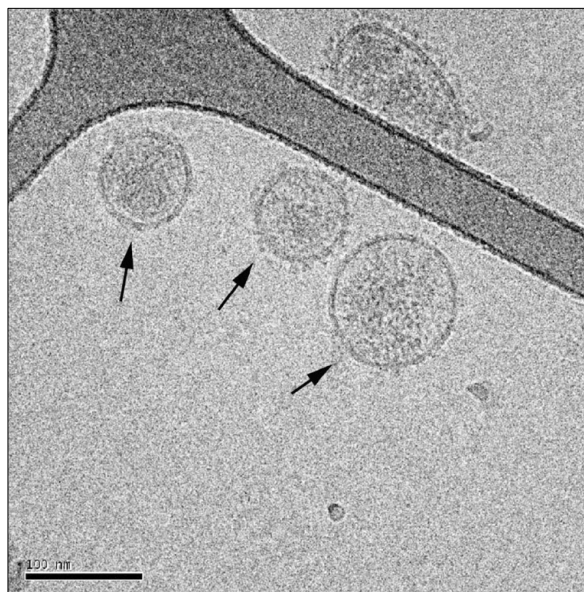

Figure 1. Cryo-electron micrograph of exosomes from prostate epithelial cells (prostasomes). Originally published in *Journal of Cell Biology* ©2013 (Raposo and Stoorvogel 2013).

Exosome Biology

Exosomes are nanovesicles that are released by most cell types into blood, urine, saliva, breast milk, and other bodily fluids (Raposo and Stoorvogel 2013). For much of the 30 years since their discovery, they have been described as garbage shuttles, but recent studies have revealed their role in cell communication and they are now seen as *information* shuttles (Huang-Doran et al. 2017). Exosomes are currently the subject of intense study as they are showing great promise for discovery of, for example, new approaches for cancer diagnosis and therapy (Barile and Vassalli 2017).

Exosomes have been reported to contain proteins, nucleic acids, lipids and metabolites, often enriched in specific molecules depending on the cell type and the environment surrounding the cells (Hessvik and Llorente 2018). For example, tumor-derived exosomes carry specific integrins to initiate the pre-metastatic niche (Hoshino et al. 2015) and induced pluripotent stem cells deliver ischemic-protective miRNAs to injured myocardium inside exosomes (Wang et al. 2015). Exosomes from different sources are often selectively loaded with a particular set of molecules, however, several components are shared between all of them. The most commonly detected proteins are used as exosome markers. These include CD9, CD81, CD63 and ALIX (PDCD6IP), as tabulated in the list of exosome markers at ExoCarta (http://exocarta.org/exosome_markers) (Keerthikumar et al. 2016).

Methodology

Exosome Structure and Components
Exosome Ultrastructure
Exosomes are surrounded by a single lipid bilayer membrane, with observed diameters ranging from 40 to 100 nanometers, and similar microvesicles may be up to ~10 times this large (Raposo and Stoorvogel 2013). In cryoelectron micrographs **(Figure 1)**, they show a roughly spherical shape, with a dense array of proteins extending from the surface and a clumpy, concentrated collection of soluble components in the interior. Early micrographic work observed a cup shape, but this has more recently been attributed to dehydration of the sample during preparation.

Exosome Proteins
Many proteomics studies have been performed for exosomes under different conditions and disease states. We chose to model a generic exosome, without specific components of any disease or environmental condition, with 30 proteins that are identified across multiple studies (Appendix 1). This list includes 20 out of the 25 top proteins tabulated in the marker list at ExoCarta. Five of the top 25 proteins were judged to be potential artifacts: cytoskeletal proteins actin and clathrin,

blood plasma proteins albumin and annexin A5, and annexin A2, which forms a complex with a protein not included in the list, S100A10/p11. Ten additional proteins were selected by manually compiling information from proteomic and review studies (Rashed et al. 2017, Kalra et al. 2016, Beach et al. 2014, Mathivanan et al. 2010, Thery et al. 2002).

For each protein, we chose the most complete structure from the list of Protein Data Bank (PDB, http://www.rcsb.org) entries at UniProt (https://www.uniprot.org) and manually curated to determine the biological state of oligomerization. In several cases, a complete structure was not available in the PDB archive, so we used structures of individual domains and published reports to build a complete structure of the protein using the Python Molecular Viewer (Sanner 1999), as follows. TSG101 has a ubiquitin E2 variant domain followed by a coiled-coil region, however the oligomerization of this protein remains unclear: some studies indicate that a single subunit participates in the ESCRT-I complex (Morita et al. 2007) while others point to a homotetramer (White et al. 2017). We chose to treat it as a monomer in the exosome. We manually connected available structures of the two domains of HSP70 based on (Stricher et al. 2013). The structure of CD81 was used to represent the three tetraspanins. Integrin was modeled from structures of the extracellular domain, single-pass transmembrane segment and short cytosolic tail based on (Lau et al. 2009). The transmembrane domains of HLA1 (human leukocyte antigen-1) and HLA2 were modeled based on the length of the chains annotated in UniProt. Glycosylation was added when indicated in UniProt using the Glycoprotein Builder (http://glycam.org) followed by manual repositioning to resolve clashes.

Exosome Nucleic Acids
Double-stranded DNA, mRNA, and miRNA have been detected in exosomes (Willms et al. 2018, Cai et al. 2016). The length of DNA fragments loaded into exosomes shows a wide range, from 35 to 17,000 base pairs, with an enrichment of the larger sizes from 6,000 to 17,000 base pairs. In addition, most studies find that these DNA fragments span all chromosomes. In the current model, we built a 10,000 base pair double-stranded DNA by repeating the structure of short oligonucleotide (PDB entry 6bna). Exosomal mRNA is typically small and heterogeneous, <200 nucleotides (Palanisamy et al. 2010), and was built using the single-stranded region of tmRNA (PDB entry 3iz4). Exosomes also contain miRNA, possibly in complex with argonaute proteins, which are currently the subject of intensive research because of their connections to cancer biology (Zhang et al. 2015). We used the structure of the pre-miR21 apical loop (PDB entry 5uzt) and the structure of argonaute2 bound to a guide RNA (PDB entry 4w5n).

Blood Plasma Proteins

Abundant information is available for blood plasma, presumably because it has historically been amenable for study. We included the 11 most abundant proteins (Appendix 2), with plasma concentration higher than 7 millimolar (Chan et al. 2004)—at this concentration, we would expect to have at least one copy present in the plasma volume represented in the model. In addition, we included miRNA in an argonaute2-miRNA complex, since there is some controversy about presence of miRNA in plasma (Gallo et al. 2012, Arroyo et al. 2011). Structures for blood plasma components were based on previous work with cellPAINT (Gardner et al. 2018) and low density lipoprotein structure was based on a cryo-electron study (entry EMD-5239 at http://www.emdatabank.org).

Concentrations

Information on absolute concentration of molecules in cells is typically the most difficult information to find in this type of integrative study. While a number of proteomics study have been done on exosomes, quantitative results are difficult to access and often contradictory. Furthermore, the quantification of exosomes is still an active field of research (Koritzinsky et al. 2017), and protein abundance is often variable among originating cell types (Yoshioka et al. 2013). This variation is currently the topic of intense study, due to its potential use for identifying cancer markers (Li et al. 2017).

As a result, we were unable to find comprehensive information on exosome protein concentration. At early stages of the project, when we were developing a conceptual model for use in a painting, we took a qualitative approach, estimating concentrations using entries from Exocarta. As the project matured, we used a more quantitative two-step approach for the 2.5D and 3D models: (1) we started with the concentration as in the donor cell, and (2) modified individual concentrations based on published reports of enrichment in exosomes. This is described in more detail below.

Platelets are the major source of exosomes in plasma (Aatonen et al. 2014), so we based our model on a quantitative proteomic study of platelet (Burkhart et al. 2012). This study reports a copy number per platelet cell. We assumed a volume of 9.7 femtoliter for the platelet, calculated molar concentrations for soluble components and area-based densities for membrane-bound components, and calculated initial copy numbers for an exosome of 100 nanometer diameter (Appendix 1). These initial values were doubled to bring the overall concentration to values similar to the typical range of concentrations in eukaryotic cytoplasm, justified by the observation in the proteomics study that "absolute quantification based on spectral counting can suffer from errors up to 200 percent in individual cases" (Burkhart et al. 2012). These platelet concentrations were used in our model for transcription factors (EEF1A1 and EEF2), enzymes (enolase, pyruvate kinase,

lactate dehydrogenase, GAPDH, etc.), two adapter proteins (YWHAZ and YWHAE), and G-protein.

We then applied an enrichment factor for proteins known to be enriched in exosomes: six-fold enrichment for HSPA8 (Faught et al. 2017), HSP90 at seven times more than HSPA8 (Clayton et al. 2005), 124-fold enrichment for tetraspanins CD9 and CD81 and seven-fold for CD63 (Escola et al. 1998). ESCRT-family proteins (PDCD6IP and TSG101) and rab7/rab5 proteins work with tetraspanins for intracellular trafficking and are commonly seen at high levels in proteomics (Tauro et al. 2012), so we applied an arbitrary enrichment factor of 20 for them. In addition, we collocated the tetraspanins and syntenin in the exosome models, since tetraspanins form microdomains that interact with syntenin (Latysheva et al. 2006). We were unable to find specific information about HLA2, so we set the concentration equal to that of HLA1, and enrichment values of eight for HLA2 and two for HLA1 (Escola et al. 1998). The concentration of argonaute2 was set equal to that of PDCD6IP, as both are implicated in miRNA packaging in exosomes (Iavello et al. 2016).

As noted in a recent review (Koritzinsky et al. 2017), the nucleic acid content of exosomes is highly heterogeneous. Concentrations of specific miRNA differ by orders of magnitude, from about 0.00001 to 10 molecules per vesicle. In the current work, the abundance of mRNA and miRNA was arbitrarily set to the average value of the soluble proteins and one DNA strand was included per vesicle. Similarly to previous work (Gardner et al. 2018), concentrations of blood plasma components were based on proteomic studies (Chan et al. 2004).

These concentrations and membrane densities may be used directly for defining 3D models. In the 2.5D representations, the visualizations are typically partitioned into a foreground layer and two or more background layers. For these, we define a thickness of 10 nanometers for each layer, calculate the number of molecules that would fall within this volume, then distribute them within the layer.

Results

Structural Representation: Art and Science

One of our goals in this study is to explore and contrast different modes of creating mesoscale visualizations. Modeling approaches, which are often used in research applications, seek to incorporate as much of the available data as possible into the resultant model and visual representation, providing the ability to explore specific hypotheses about the relationship of structure with function. Illustrative approaches, on the other hand, are often used for dissemination, education and outreach, and may allow more room for artistic license, with the goal of

creating a visual representation that is intuitive and comprehensible to a wide audience.

As a first step, we explored the representation of the components of exosomes and blood plasma, to define the players that will be used to build a scene **(Figure 2)**. In each case, 3D models were built directly from atomic structures, and we generated images of full atomic models as well as smoothed models that are appropriate when depicting entire mesoscale scenes. Using these models, we searched for iconic views that reflect the shape and function of the molecules. These iconic views were manually chosen with several (occasionally conflicting) goals in mind. They may highlight the symmetry of a complex, making it easy to see how individual subunits interact (as in HSP90). They may highlight functional features of the molecule (as with binding clefts on the two domains of HSP70). The view may be constrained by how it interacts with other features of the environment (as in the placement of all membrane-bound proteins approximately parallel to the plane of the page).
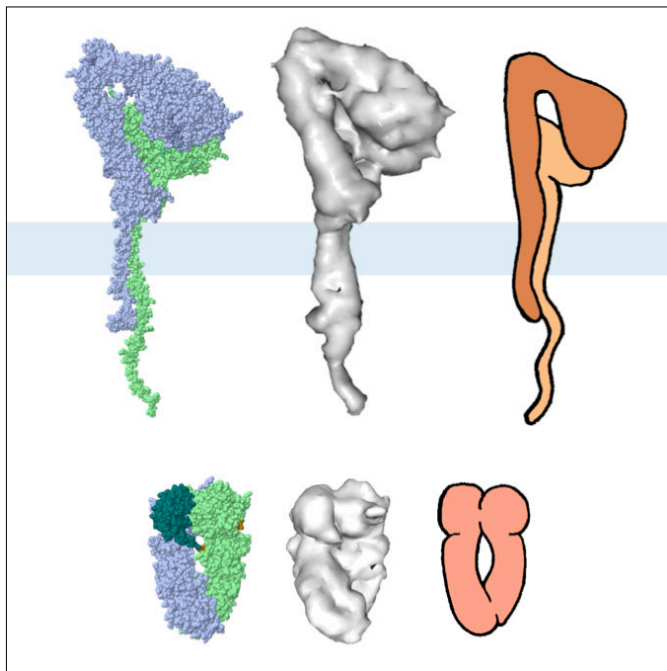

Figure 2. Development of protein representations. The atomic structure, surface representation, and hand-drawn image are shown for integrin (top) and HSP90 (bottom).

Once the viewing perspective was chosen, we based hand-drawn images on the surface representation of the structure and at proper scale **(Figure 3)**. The aim of the visual representation of the proteins is simplify the structure to make it more comprehensible and recognizable when embedded in a complex scene. To this end, we incorporated several conventions into these drawings. Flat colors and outlines provide a familiar style similar to cartoons we see throughout the popular media. In several cases, additional gesture lines are

added to relate the structures to features often used in diagrams; for example, lines were added to tetraspanins to show the multiple crossings of the chain across the membrane.
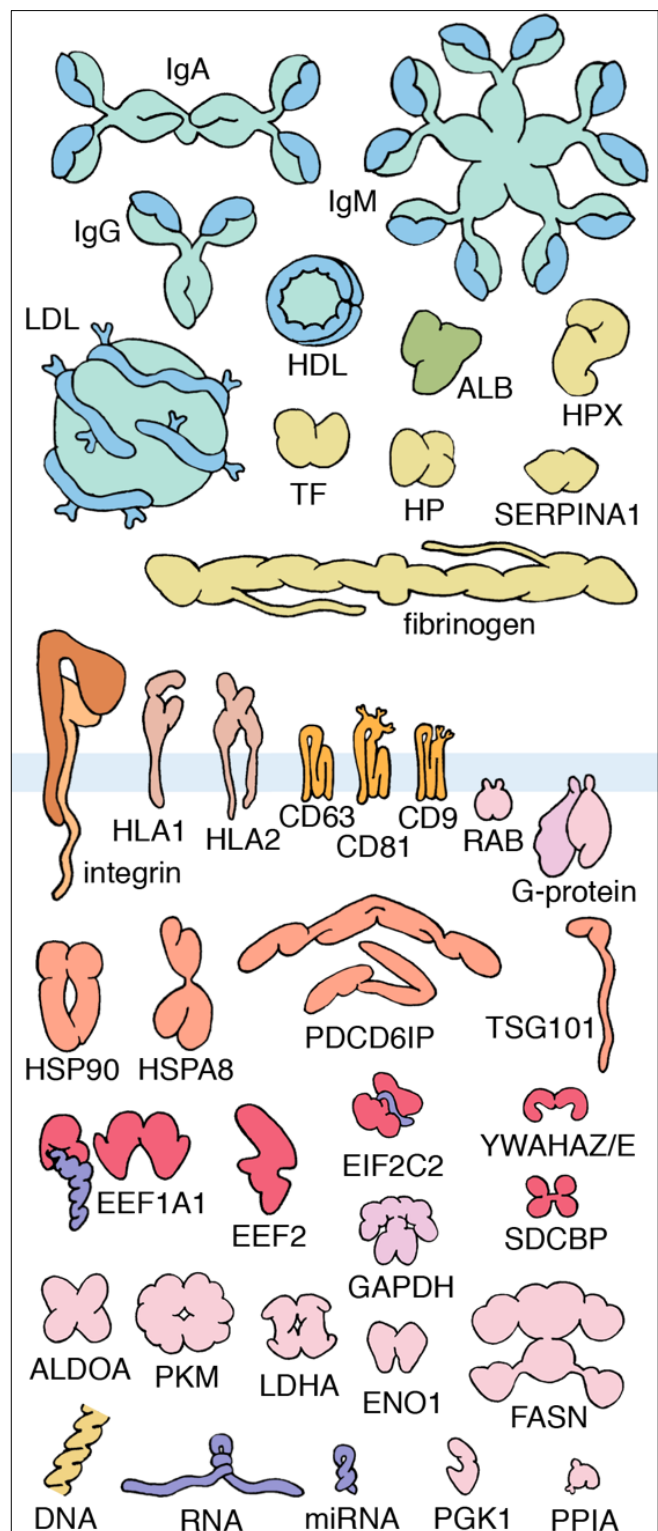

Figure 3. Proteins included in the visualizations. Blood plasma proteins are at the top, exosome membrane proteins at the center, and exosome lumen proteins at the bottom.

Hand-Rendered Illustration

We created a hand-rendered illustration **(Figure 4)** using our previously reported watercolor technique (Goodsell 2016). As in previous work, this illustration incorporates many modes of artistic license to improve the readability of the scene. For example, we use only standard views for molecule, so, for instance, all antibodies in the blood plasma are seen with their iconic Y-shape. The entire cross-sectional scene shows only a shallow slice through the environment, with sharp depth cueing to enhance spatial relationships and choosing a clipping plane that intersects the vesicles at their center, allowing intuitive understanding of the shape of the vesicle.

For the overall layout of the model, we chose to represent three circulating exosomes with diameters of 50, 100 and 140 nanometers. The 100 nanometer exosome, which is the size of a "typical" exosome, was placed in the most visually prominent position. The coloring scheme was chosen to enhance the different compartments, with warm colors for the exosome lumen to highlight the main characters in the scene.

Figure 5. CellPAINT illustration of the scene depicted in Figure 4.

Figure 4. Watercolor illustration of three circulating exosomes (2018, Julia Jiménez). The exosome on the right has a diameter of 100 nanometers.

CellPAINT 2.5D Digital Illustration

CellPAINT allows users to create similar illustrations using an interface akin to a paint program, providing a menu of molecules and membranes that may be added into a cross-sectional view of a cellular scene (Gardner et al. 2018). As with the painting, only a thin slice of the scene is visible, with a bright foreground layer and several background layers progressively obscured with distance-dependent depth cueing to black. In the current work, cellPAINT was modified to incorporate the molecular images from Figure 3 into the painting menu, allowing creation of scenes similar to the hand-drawn illustration.
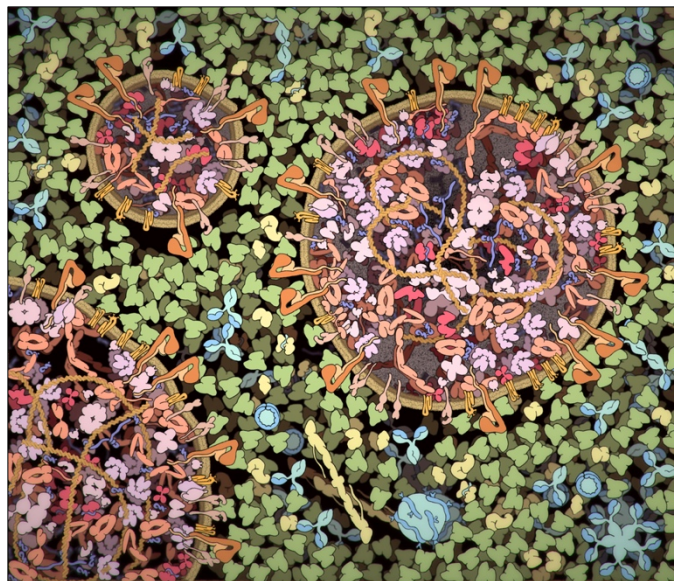
The great advantage of cellPAINT is that these digital scenes may be produced rapidly, sketching a variety of hypotheses. For example, Figure 5 shows a scene based on the watercolor painting, but generated in an hour or so. The two illustrations in Figure 6 show the results of a recent study using engineered exosomes for treatment of Parkinson's disease (Haney et al. 2015). In the study, native exosomes from macrophages, which show a diameter of about 100 nanometers, were mixed with a solution of catalase and extruded several times through small pores in an Avanti Lipids Extruder. This causes breakage and reformation of slightly larger exosomes (155 nanometers diameter) loaded with catalase, and presumably also losing most of the native exosome interior contents in the process. These loaded "exoCAT" exosomes then provide an effective mode of delivery of the therapeutic catalase across the blood-brain barrier to target neural and vascular cells. These illustrations required 5 to 10 minutes to produce and thus may be easily updated as new research results are obtained.
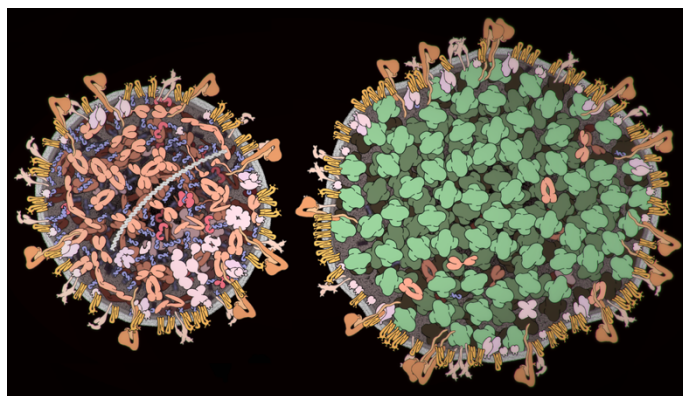

Figure 6. CellPAINT illustrations of a typical exosome (left) and a therapeutic exoCAT, with catalase molecules in green (right).

CellPACK 3D Model

Detailed 3D models were created using cellPACK **(Figure 7)** (Klein et al. 2018). CellPACK builds stochastic 3D models based on a recipe that defines structures, locations, and concentrations of molecular components randomly distributed within a scene composed of membrane-bounded compartments. The exosome recipe includes a spherical exosome membrane with diameter of 100 nanometers, and molecular parameters for the blood plasma, exosome membrane and exosome lumen based on the information included in Appendices 1 and 2.
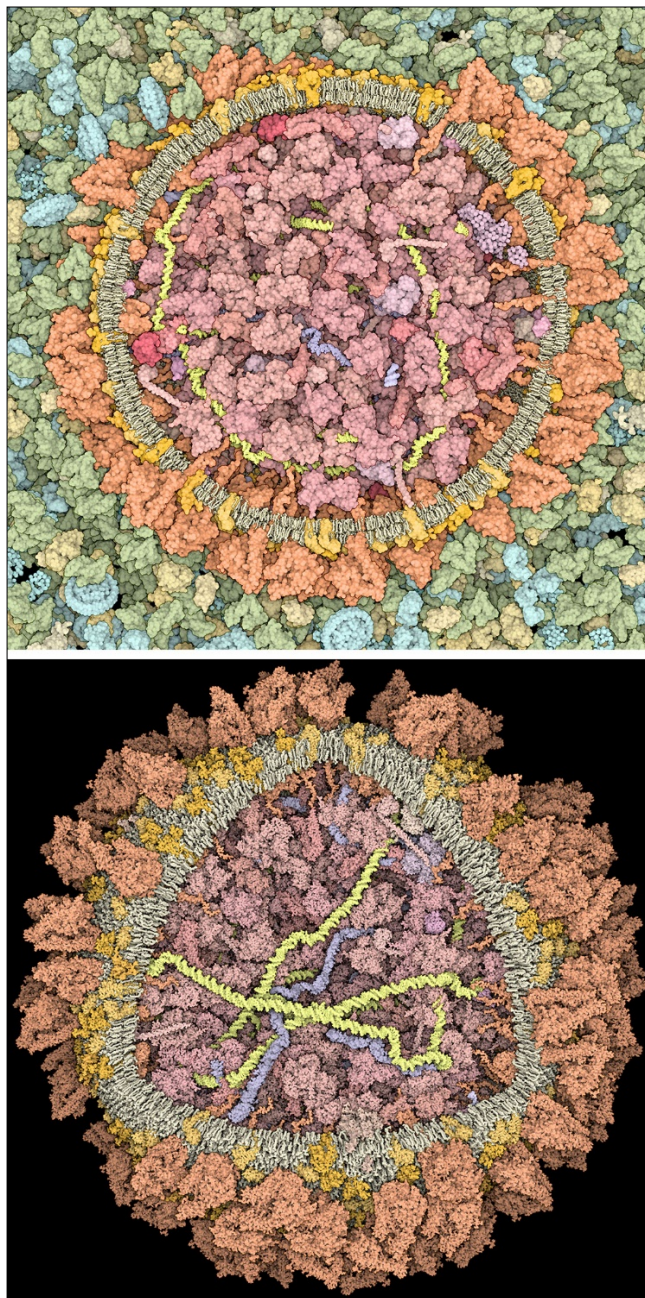

Figure 7. 3D model of an exosome. (Top) Cross-section of the exosome surrounded by blood plasma, with proteins in smooth surfaces. (Bottom) Isolated exosome with one quadrant of membrane removed to reveal the cargo molecules inside.

Discussion

Art/Science Dialog

One of our goals of this work is to explore the dialog between art and science in our ongoing mesoscale modeling work. We are currently using three approaches: traditional illustrations of cross-sectional views, illustrations created digitally with cellPAINT, and visualization of full 3D models from cellPACK. Our insights from this study are best understood in the context of different communities of end users of the results, in research and in education.

In a research setting, this study followed a course that we have used in past studies on HIV and mycoplasmas. The artistic approaches, using watercolor and cellPAINT, are useful for gathering data, to help crystallize an integrated view and identify potential problems and gray areas. An artistic approach allows freedom to explore possibilities, sketching in approximations when needed and defining the areas that need exploration when moving to the rigor of detailed 3D modeling. The artistic approach also provides many insights that may be employed to aid interpretability when moving on to 3D visualizations. In turn, 3D models bring added levels of rigor to the artistic approaches. Methods like cellPAINT bridge the two approaches, allowing facile creation of new illustrations based on insights from the research with models.

In dissemination, outreach and education settings, we also have experienced a robust dialog between illustrative and modeling approaches. In general, the 2D and 2.5D illustrations are more comprehensible than visualizations of the 3D models, which may be due to the extensive use of artistic license to enhance both shape recognition of individual components and their spatial relationships in the scene. We fully expect that 2D and 2.5D illustrations will be essential in the dissemination of our work on exosomes, for use in introductory material to give viewers the context of results. Then, once the general context is understood, more complex visualizations of 3D models may be presented to delve deeply into the science.

Accessibility of Mesoscale Methodology

This report presents the results of a proof-of-principle study, where methods were applied and evaluated by a single student (JJ) during a two-month summer internship at the Scripps Research Institute. Building on an informal foundation of artistic practice and based on her graduate research project focusing on experimental study of exosomes in cancer, she was able to research, plan, and render the watercolor illustration in this time, and to use the integrated information to create visualizations within the existing cellPAINT and cellPACK software suite. This underscores the growing accessibility of mesoscale techniques, both for creation of explanatory imagery and for detailed mesoscale modeling in research. We are currently expanding the functionality of both

cellPACK and cellPAINT to make them more general and more accessible, and plan to do a more comprehensive user study to identify current needs of the research and education community, address these needs in our software development, and assess the utility of the resulting methods. The exosome version of cellPAINT, as well as a version tailored for creation of scenes with HIV-1, blood plasma, and a cell surface, is
available at http://cellpaint.scripps.edu. More information on the entire cellPACK suite of programs is available at http://ccsb.scripps.edu/cellpack.

Acknowledgements

References

Aatonen, M. T., Ohman, T., Nyman, T. A., Laitinen, S., Gronholm, M., and Siljander, P. R. 2014. Isolation and characterization of platelet-derived extracellular vesicles. *J Extracell Vesicles* 3:1.

Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., Mitchell, P. S., Bennett, C. F., Pogosova-Agadjanyan, E. L., Stirewalt, D. L., Tait, J. F., and Tewari, M. 2011. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A* 108:5003-8.

Barile, L., and Vassalli, G. 2017. Exosomes: Therapy delivery tools and biomarkers of diseases. *Pharmacol Ther* 174:63-78.

Beach, A., Zhang, H. G., Ratajczak, M. Z., and Kakar, S. S. 2014. Exosomes: an overview of biogenesis, composition and role in ovarian cancer. *J Ovarian Res* 7:14.

Burkhart, J. M., Vaudel, M., Gambaryan, S., Radau, S., Walter, U., Martens, L., Geiger, J., Sickmann, A., and Zahedi, R. P. 2012. The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood* 120:e73-82.

Cai, J., Wu, G., Jose, P. A., and Zeng, C. 2016. Functional transferred DNA within extracellular vesicles. *Exp Cell Res* 349:179-183.

Chan, K. C., Lucas, D. A., Hise, D., Schaefer, C. F., Xiao, Z., Janini, G. M., Buetow, K. H., Issaq, H. J., Veenstra, T. D., and Conrads, T. P. 2004. Analysis of the human serum proteome. *Clinical Proteomics Journal* 1:101-225.

Clayton, A., Turkes, A., Navabi, H., Mason, M. D., and Tabi, Z. 2005. Induction of heat shock proteins in B-cell exosomes. *J Cell Sci* 118:3631-8.

Escola, J. M., Kleijmeer, M. J., Stoorvogel, W., Griffith, J. M., Yoshie, O., and Geuze, H. J. 1998. Selective enrichment of tetraspan proteins on the internal vesicles of multivesicular endosomes and on exosomes secreted by human B-lymphocytes. *J Biol Chem* 273:20121-7.

Faught, E., Henrickson, L., and Vijayan, M. M. 2017. Plasma exosomes are enriched in Hsp70 and modulated by stress and cortisol in rainbow trout. *J Endocrinol* 232:237-246.

Gallo, A., Tandon, M., Alevizos, I., and Illei, G. G. 2012. The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. *PLoS One* 7:e30679.

Gardner, A., Autin, L., Barbaro, B., Olson, A. J., and Goodsell, D. S. 2018. CellPAINT: Interactive illustraiton of dynamic mesoscale cellular environments. *IEEE Computer Graphics Appl.* 38:51-66.

Goodsell, D. S. 1992. A Look Inside the Living Cell. *American Scientist* 80:457-465.

Goodsell, D. S. 2016. Cellular landscapes in watercolor. *J Biocommun* 40:22-26.

Goodsell, D. S., Autin, L., and Olson, A. J. 2018. Lattice models of bacterial nucleoids. *J Phys Chem B* 122:5441-5447.

Goodsell, D. S., Franzen, M. A., and Herman, T. 2018. From Atoms to Cells: Using Mesoscale Landscapes to Construct Visual Narratives. *J Mol Biol* 430:3954-3968.

Haney, M. J., Klyachko, N. L., Zhao, Y., Gupta, R., Plotnikova, E. G., He, Z., Patel, T., Piroyan, A., Sokolsky, M., Kabanov, A. V., and Batrakova, E. V. 2015. Exosomes as drug delivery vehicles for Parkinson's disease therapy. *J Control Release* 207:18-30.

Hessvik, N. P., and Llorente, A. 2018. Current knowledge on exosome biogenesis and release. *Cell Mol Life Sci* 75:193-208.

Hoshino, A., Costa-Silva, B., Shen, T. L., Rodrigues, G., Hashimoto, A., Tesic Mark, M., Molina, H., Kohsaka, S., Di Giannatale, A., Ceder, S., Singh, S., Williams, C., Soplop, N., Uryu, K., Pharmer, L., King, T., Bojmar, L., Davies, A. E., Ararso, Y., Zhang, T., Zhang, H., Hernandez, J., Weiss, J. M., Dumont-Cole, V. D., Kramer, K., Wexler, L. H., Narendran, A., Schwartz, G. K., Healey, J. H., Sandstrom, P., Labori, K. J., Kure, E. H., Grandgenett, P. M., Hollingsworth, M. A., de Sousa, M., Kaur, S., Jain, M., Mallya, K., Batra, S. K., Jarnagin, W. R., Brady, M. S., Fodstad, O., Muller, V., Pantel, K., Minn, A. J., Bissell, M. J., Garcia, B. A., Kang, Y., Rajasekhar, V. K., Ghajar, C. M., Matei, I., Peinado, H., Bromberg, J., and Lyden, D. 2015. Tumour exosome integrins determine organotropic metastasis. *Nature* 527:329-35.

Huang-Doran, I., Zhang, C. Y., and Vidal-Puig, A. 2017. Extracellular Vesicles: Novel Mediators of Cell Communication In Metabolic Disease. *Trends Endocrinol Metab* 28:3-18.

Iavello, A., Frech, V. S., Gai, C., Deregibus, M. C., Quesenberry, P. J., and Camussi, G. 2016. Role of Alix in miRNA packaging during extracellular vesicle biogenesis. *Int J Mol Med* 37:958-66.

Johnson, G. T., Autin, L., Al-Alusi, M., Goodsell, D. S., Sanner, M. F., and Olson, A. J. 2015. cellPACK: a virtual mesoscope to model and visualize structural systems biology. *Nat Methods* 12:85-91.

Johnson, G. T., Goodsell, D. S., Autin, L., Forli, S., Sanner, M. F., and Olson, A. J. 2014. 3D molecular models of whole HIV-1 virions generated with cellPACK. *Faraday Discuss* 169:23-44.

Kalra, H., Drummen, G. P. C., and Mathivanan, S. 2016. Focus on Extracellular Vesicles: Introducing the Next Small Big Thing. *International Journal of Molecular Sciences* 17:170.

Keerthikumar, S., Chisanga, D., Ariyaratne, D., Al Saffar, H., Anand, S., Zhao, K., Samuel, M., Pathan, M., Jois, M., Chilamkurti, N., Gangoda, L., and Mathivanan, S. 2016. ExoCarta: A Web-Based Compendium of Exosomal Cargo. *J Mol Biol* 428:688-692.

Klein, T., Autin, L., Kozlikova, B., Goodsell, D. S., Olson, A., Groller, M. E., and Viola, I. 2018. Instant Construction and Visualization of Crowded Biological Environments. *Ieee Transactions on Visualization and Computer Graphics* 24:862-872.

Koritzinsky, E. H., Street, J. M., Star, R. A., and Yuen, P. S. 2017. Quantification of Exosomes. *J Cell Physiol* 232:1587-1590.

Latysheva, N., Muratov, G., Rajesh, S., Padgett, M., Hotchin, N. A., Overduin, M., and Berditchevski, F. 2006. Syntenin-1 is a new component of tetraspanin-enriched microdomains: mechanisms and consequences of the interaction of syntenin-1 with CD63. *Mol Cell Biol* 26:7707-18.

Lau, T. L., Kim, C., Ginsberg, M. H., and Ulmer, T. S. 2009. The structure of the integrin alphaIIbbeta3 transmembrane complex explains integrin transmembrane signalling. *EMBO J* 28:1351-61.

Li, A., Zhang, T., Zheng, M., Liu, Y., and Chen, Z. 2017. Exosomal proteins as potential markers of tumor diagnosis. *J Hematol Oncol* 10:175.

Mathivanan, S., Ji, H., and Simpson, R. J. 2010. Exosomes: Extracellular organelles important in intercellular communication. *Journal of Proteomics* 73:1907-1920.

Morita, E., Sandrin, V., Alam, S. L., Eckert, D. M., Gygi, S. P., and Sundquist, W. I. 2007. Identification of human MVB12 proteins as ESCRT-I subunits that function in HIV budding. *Cell Host Microbe* 2:41-53.

Palanisamy, V., Sharma, S., Deshpande, A., Zhou, H., Gimzewski, J., and Wong, D. T. 2010. Nanostructural and transcriptomic analyses of human saliva derived exosomes. *PLoS One* 5:e8577.

Raposo, G., and Stoorvogel, W. 2013. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol* 200:373-83.

Rashed, M. H., Bayraktar, E., Helal, G. K., Abd-Ellah, M. F., Amero, P., Chavez-Reyes, A., and Rodriguez-Aguayo, C. 2017. Exosomes: From Garbage Bins to Promising Therapeutic Targets. *International Journal of Molecular Sciences* 18:E538.

Sanner, Michel F. 1999. Python: A Programming Language for Software Integration and Development. *J. Mol. Graphics Mod.* 17:57-61.

Stricher, F., Macri, C., Ruff, M., and Muller, S. 2013. HSPA8/HSC70 chaperone protein: structure, function, and chemical targeting. *Autophagy* 9:1937-54.

Tauro, B. J., Greening, D. W., Mathias, R. A., Ji, H., Mathivanan, S., Scott, A. M., and Simpson, R. J. 2012. Comparison of ultracentrifugation, density gradient separation, and immunoaffinity capture methods for isolating human colon cancer cell line LIM1863-derived exosomes. *Methods* 56:293-304.

Thery, C., Zitvogel, L., and Amigorena, S. 2002. Exosomes: Composition, biogenesis and function. *Nature Reviews Immunology* 2:569-579.

Wang, Y., Zhang, L., Li, Y., Chen, L., Wang, X., Guo, W., Zhang, X., Qin, G., He, S. H., Zimmerman, A., Liu, Y., Kim, I. M., Weintraub, N. L., and Tang, Y. 2015. Exosomes/microvesicles from induced pluripotent stem cells deliver cardioprotective miRNAs and prevent cardiomyocyte apoptosis in the ischemic myocardium. *Int J Cardiol* 192:61-9.

White, J. T., Toptygin, D., Cohen, R., Murphy, N., and Hilser, V. J. 2017. Structural Stability of the Coiled-Coil Domain of Tumor Susceptibility Gene (TSG)-101. *Biochemistry* 56:4646-4655.

Willms, E., Cabanas, C., Mager, I., Wood, M. J. A., and Vader, P. 2018. Extracellular Vesicle Heterogeneity: Subpopulations, Isolation Techniques, and Diverse Functions in Cancer Progression. *Front Immunol* 9:738.

Yoshioka, Y., Konishi, Y., Kosaka, N., Katsuda, T., Kato, T., and Ochiya, T. 2013. Comparative marker analysis of extracellular vesicles in different human cancer types. *J Extracell Vesicles* 2:10.3402/jev.v2i0.20424.

Zhang, J., Li, S., Li, L., Li, M., Guo, C., Yao, J., and Mi, S. 2015. Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics Proteomics Bioinformatics* 13:17-24.

Authors

Julia Jiménez is a predoctoral researcher in the Experimental Therapies and Biomarkers group at The Sanitary Research Institution IdiPAZ in Madrid and the Cancer Epigenetics group at University Hospital La Paz, Madrid. She has participated in the publication of 4 papers and 2 patents related to her research projects. She is performing her Ph.D. in molecular biosciences at the Autonomous University of Madrid. Contact her at jul.jimenez@estudiante.uam.es.

Ludovic Autin is a Staff Scientist with the Scripps Research Institute. He is Lead Developer of methods for mesoscale modeling, visualization, and augmented reality. He received his Ph.D. in molecular modeling from Paris 5 University. Contact him at autin@scripps.edu.

Inmaculada Ibáñez de Cáceres is the coordinator of the Experimental Therapies and Biomarkers group at The Sanitary Research Institution IdiPAZ in Madrid, and is the Director of the Cancer Epigenetics Group at University Hospital La Paz, Madrid. She has published more than 30 papers in the epigenetics field, is the main author of three patents based on biomarkers of clinical use, one of them already licensed and under exploitation, and has been serving as a Full Member representing her institution at the biomarkers platform form the European Infrastructure for Translational Medicine (EATRIS). Contact her at inma.ibanezca@salud.madrid.org.

David S. Goodsell (corresponding author) is an Associate Professor with the Scripps Research Institute and a Research Professor with Rutgers University. He currently divides his time between research in computational biology and science outreach. He received his Ph.D. in biochemistry from University of California, Los Angeles. Contact him at goodsell@scripps.edu.

Licensing

Appendix 1. Exosome Proteins

Name	Description	UniProt ID	PDB ID	Number / Exosome (1)	Oligomeric State
Lumen Proteins					
HSP90	Heat Shock Protein	P07900 P08238	2cg9	86	Heterotetramer
HSPA8	Heat Shock Cognate 71 kD Protein	P11142	1dkz, 3hsc	30	Monomer
ENO1	Enolase	P06733	2psn	5	Homodimer
PKM	Pyruvate Kinase	P14618	1t5a	2	Homotetramer
LDHA	L-lactate Dehydrogenase	P00338	4ojn	2	Homotetramer
ALDOA	Fructose-bisphosphate Aldolase A	P04075	5ky6	3	Homotetramer
PGK1	Phosphoglycerate Kinase 1	P00558	5m1r	4	Monomer
FASN	Fatty Acid Synthase	P49327	2cf2	1	Homodimer
GAPDH	Glyceraldehyde-3-phosphate Dehydrogenase	P04406	4wnc	3	Homotetramer
PPIA	Peptidyl-prolyl cis-trans Isomerase A	P62937	5noq	11	Monomer
PDCD6IP	ALIX	Q8WUM4	2oev	23	Monomer or Homodimer
TSG101	Tumor Susceptibility Gene 101 Protein	Q99816	3obs, 3iv1	8	Monomer or Homotetramer
EIF2C2	Protein Argonaute-2	Q9UKV8	4w5n	23	Monomer
SDCBP	Syntenin-1	O00560	1n99	1	Monomer or Homodimer
YWHAZ	14-3-3 Protein zeta/delta	P63104	1qja	10	Homodimer or Heterodimer
YWHAE	14-3-3 Protein epsilon	P62258	2br9	7	Homodimer or Heterodimer
EEF1A1	Elongation Factor 1-alpha 1	P68104	1dg1	1	Homodimer
EEF2	Elongation Factor 2	P13639	1n0u	1	Monomer
Membrane-associated Proteins					
CD63	CD63 Tetraspanin	P08962	(5tcx)	59	Monomer
CD81	CD81 Tetraspanin	P60033	5tcx	355	Monomer
CD9	CD9 Tetraspanin	P21926	(5tcx)	281	Monomer
ITG	Integrin	P08648 P05556	4wk4, 2k9j, 1m8o	201	Heterodimer

RAB5A	Ras-related Protein Rab-5A	P20339	1n6h	18	Monomer
RAB7A	Ras-related Protein Rab-7A	P51149	1vg8	54	Monomer
G-protein	Heterotrimeric G-protein	P10824 P62871 P63212	1gg2	72	Heterotrimer
HLAI	Human Leukocyte Antigen Class I	P03989 P61769	1hsa	7	Heterodimer
HLA2	Human Leukocyte Antigen Class II	P01903 P04229	1dlh	28	Heterodimer

(1) Values are those used for the 3D cellPACK model, estimated for an exosome with diameter of 100 nm.

Appendix 2. Blood Plasma Proteins

Name	Description	UniProt ID	PDB ID	Estimated Concentration (uM)	Oligomeric State
ALB	Serum Albumin	P02768	1e7i	640.	Monomer
TF	Serotransferrin	P02787	3v83,chain A	42.	Monomer
IgG	Immunoglobulin	P01857 P0DOX7/8	2ig2, 1fc2	53.-120.	Heterooligomer
IgA	Immunoglobulin	P01876 P0DOX7/8	(2ig2, 1fc2)	2.8-14.1	Heterooligomer
IgM	Immunoglobulin	P01871 P0DOX7/8	2rcj	0.60-2.6	Heterooligomer
Fibrinogen	Fibrinogen	P02671 P02675	1m1j	7.2	Heterotetramer
SERPINA1	Alpha-1-antitrypsin	P01009	1psi	39.	Monomer
HP	Haptoglobin	P00738	4x0l	3.8	Heterotetramer
HPX	Hemopexin	P02790	1qjs	65.	Monomer
HDL	High Density Lipoprotein	P02647	2msc, chains A,C	23.2	Heterooligomer
LDL	Low Density Lipoprotein	P04114	(EMD-5239)	1.-3.	Heterooligomer
EIF2C2	Protein Argonaute-2	Q9UKV8	4w5n	trace	Monomer