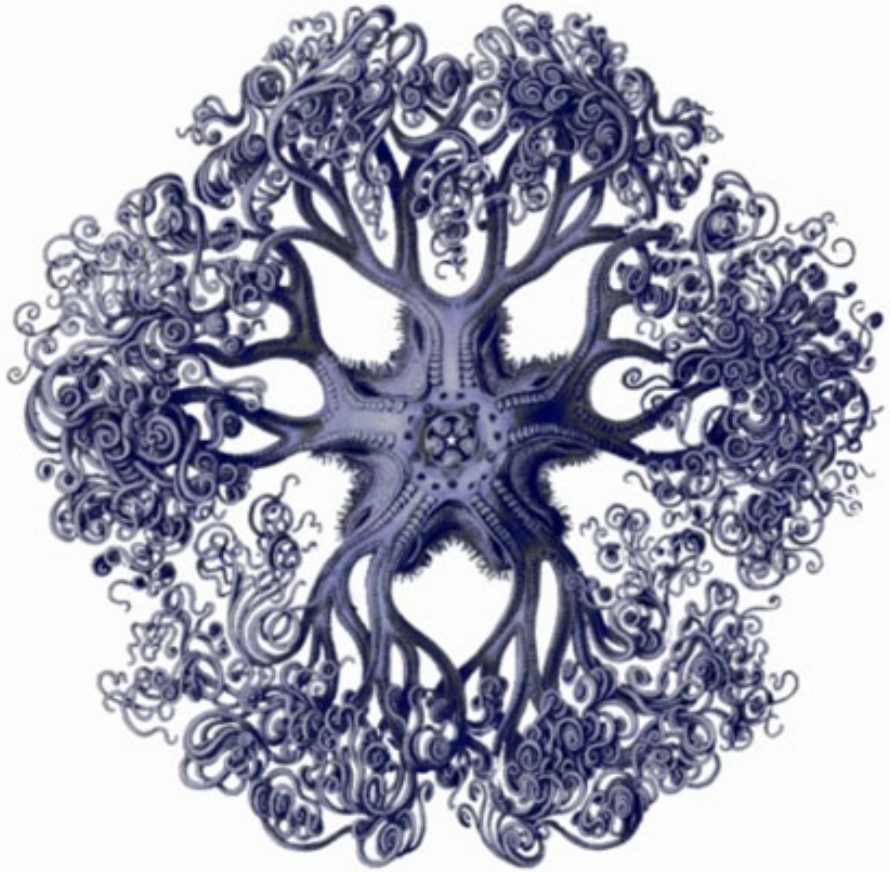
Scientific Visualizations of Microorganisms

Rashmi Kumari
186330007
IDC School of Design, IIT Bombay

Venkatesh Rajamanickam
IDC School of Design, IIT Bombay

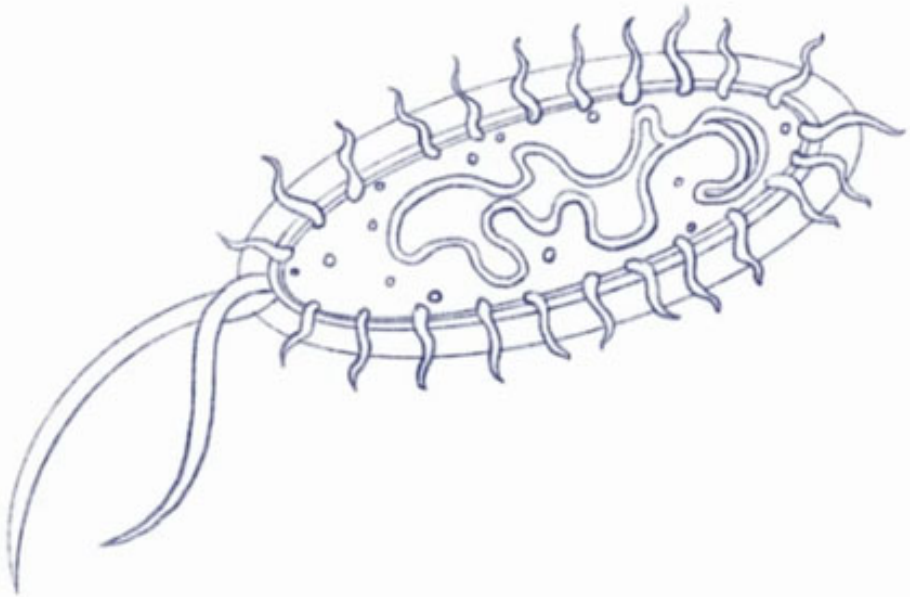


Scientific Visualizations

Figurative such as representative illustrations.

Image : Ernst Haeckel's Artforms in Nature, Artform: Astrophyton, Order: Ophiodea
https://en.wikipedia.org/wiki/Wikipedia:Reference_desk/Archives/Science/May_2006

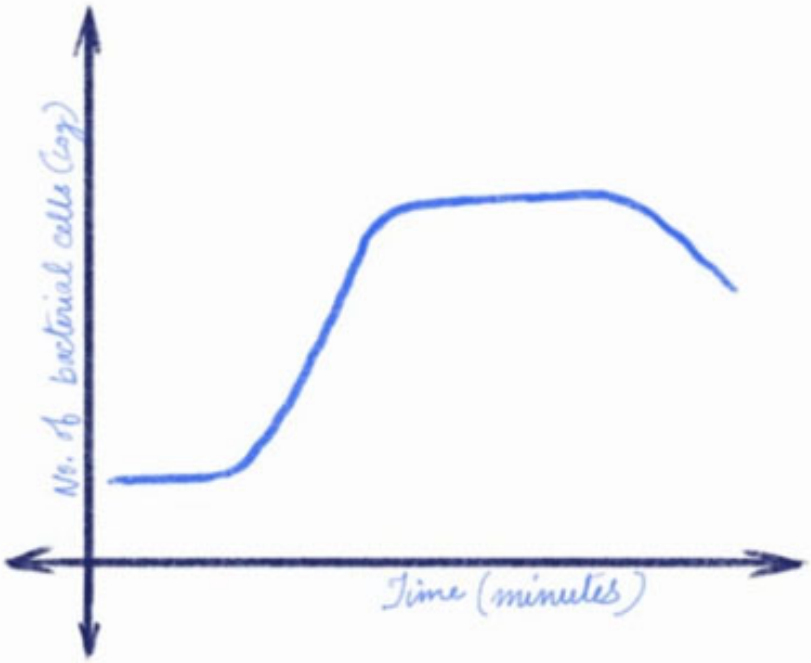
Scientific Visualizations



Figurative such as representative illustrations.

Diagrams to represent a concept.

Image : Diagram of a bacterial cell



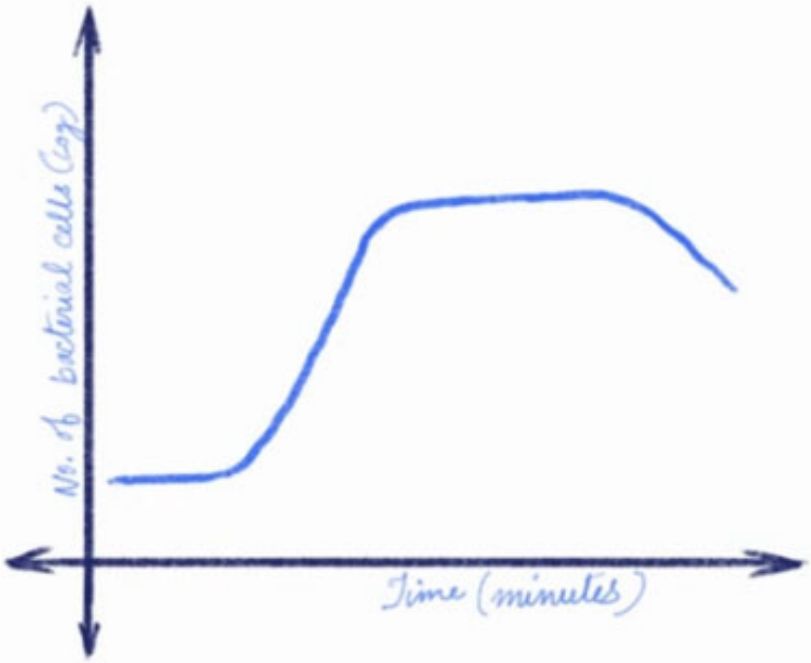
Scientific Visualizations

Figurative such as representative illustrations.

Diagrams to represent a concept.

Abstract representations such as plots and graphs to explain phenomenon or results.

Image : Typical bacterial growth curve
https://www.nature.com/articles/srep15159



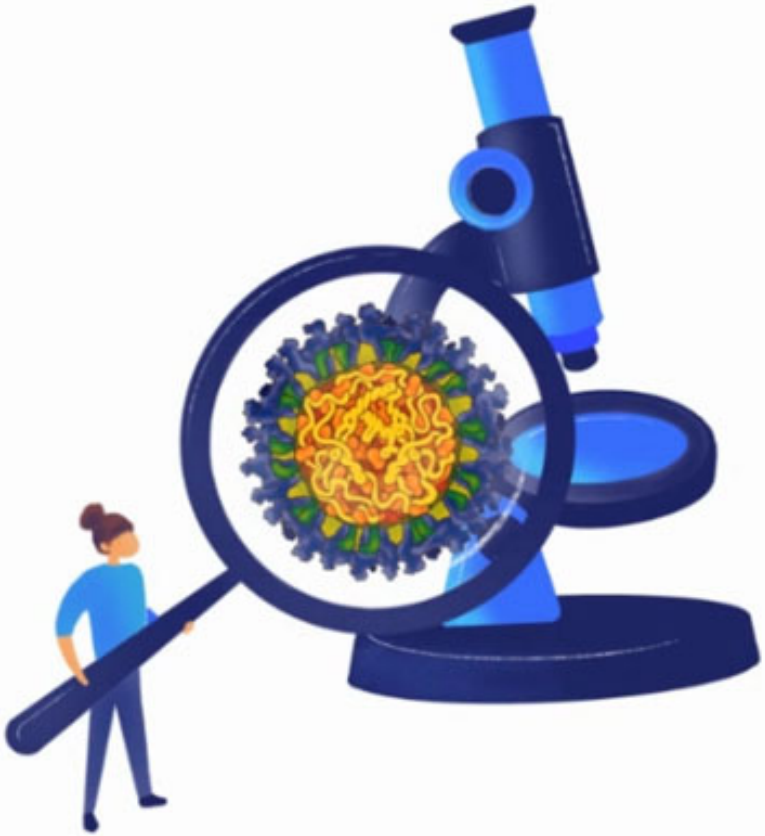
Data Visualizations

Figurative such as representative illustrations.

Diagrams to represent a concept.

Abstract representations such as plots and graphs to explain phenomenon or results.

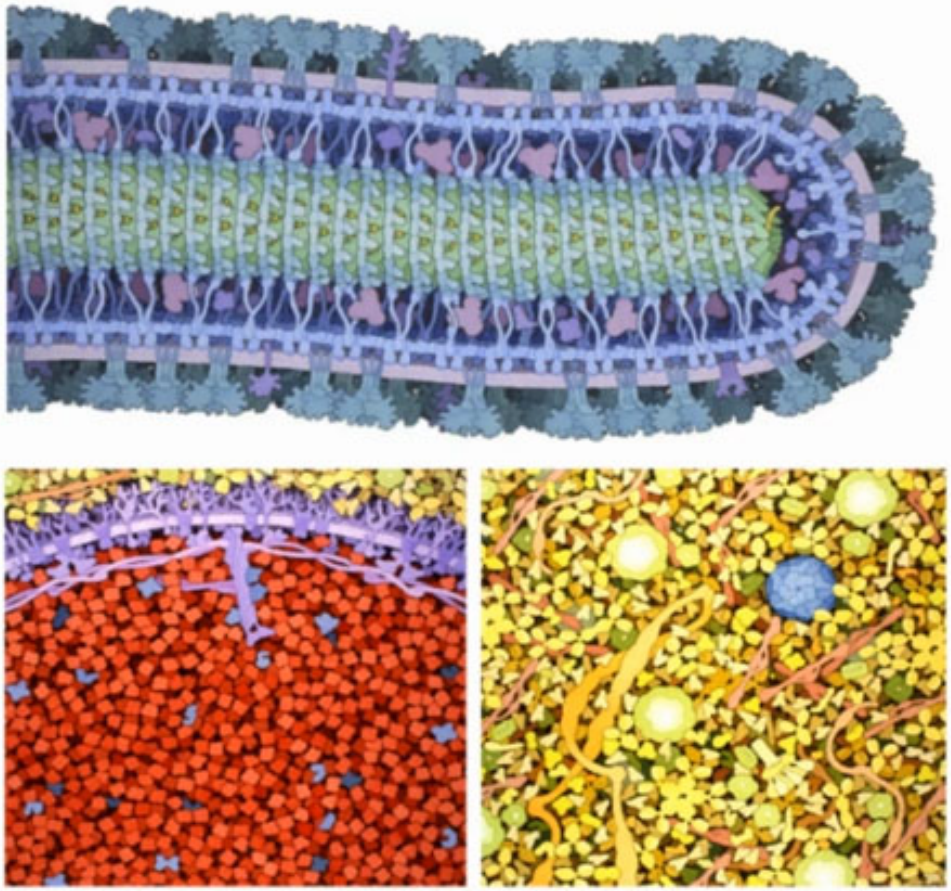
Image : Typical bacterial growth curve
https://www.nature.com/articles/srep15159



Goals

To present a near accurate view of the microorganism which addresses the size, scale, shape, position and quantity of its various components.

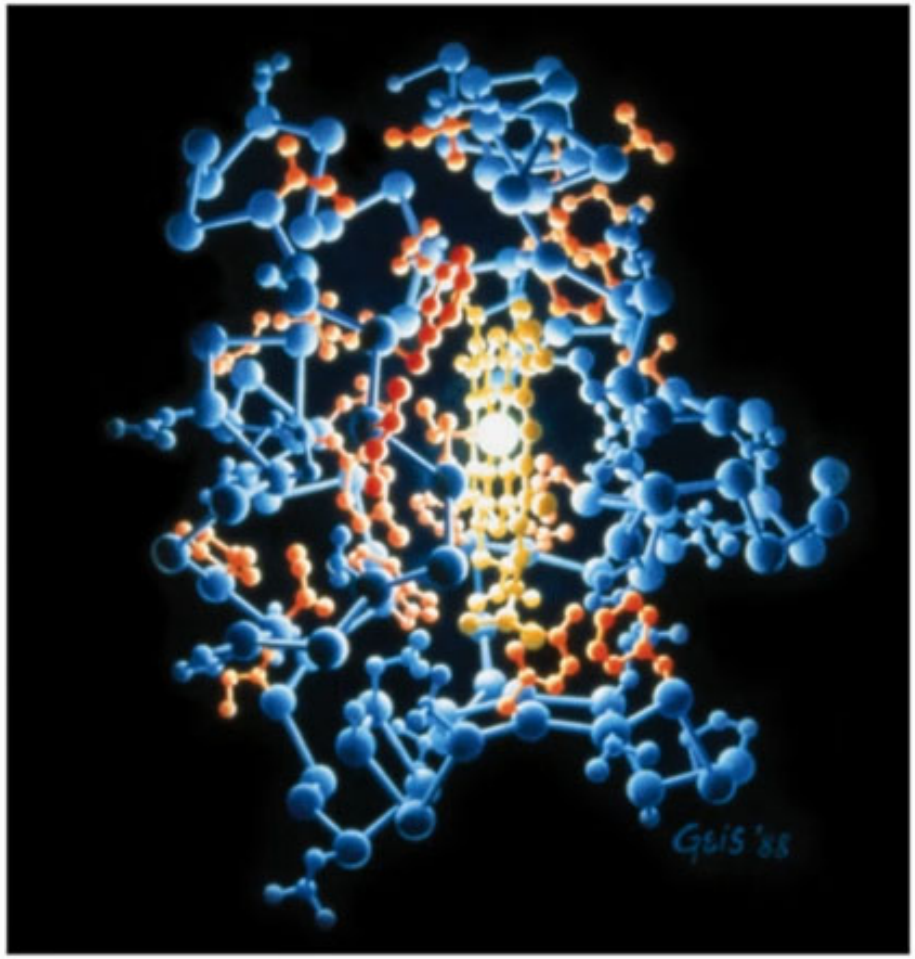
Research



David S. Goodsell

He has been making scientifically accurate paintings and illustrations of molecular structures of things related to cellular environments, proteins, scientific and medical processes etc., for more than 25 years.

Image (Clockwise from top) : Ebola virus, blood plasma and red blood cells.
https://pdb101.rcsb.org/sci-art/goodsell-gallery/

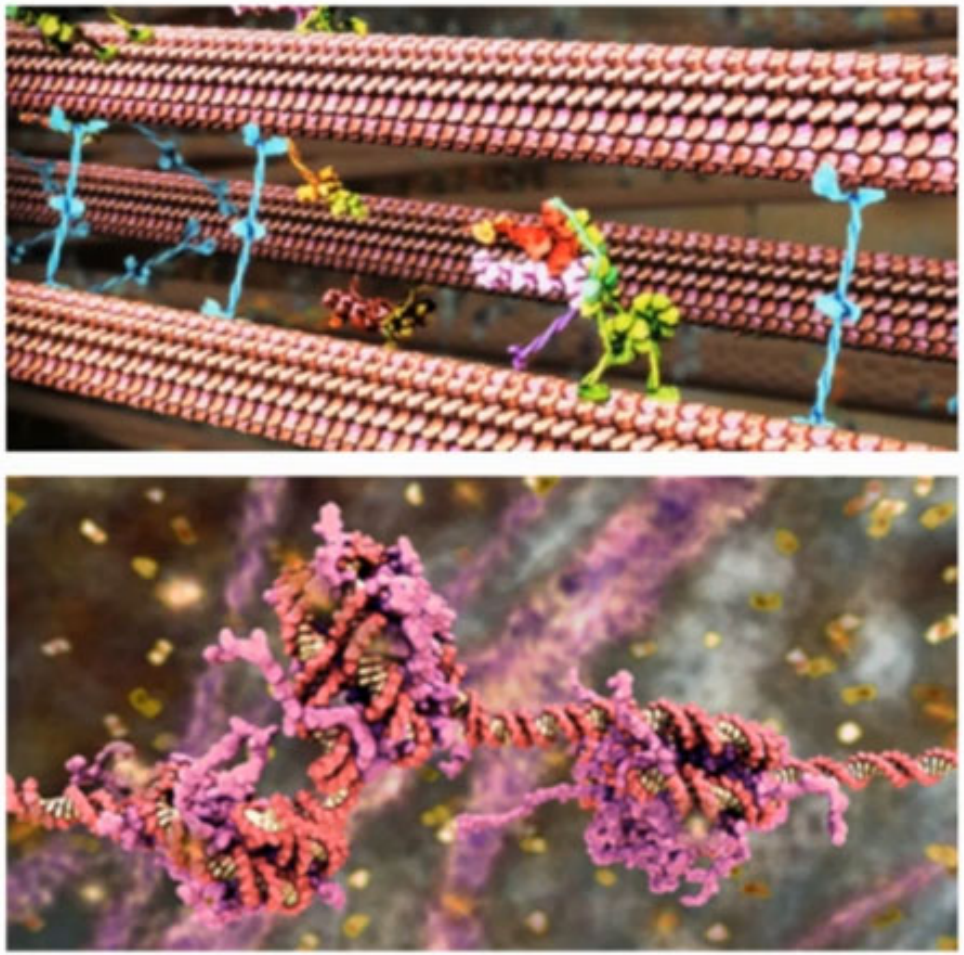


Irving Geis

" I have always been fascinated by making the invisible visible. "

Geis used a process he called "Selective Lying" to tweak the protein representation so that the end result is a protein structure and its molecular mechanism is easy to understand.

Image : Oil painting of 3-D protein structure of cytochrome C
https://pdb101.rcsb.org/sci-art/geis-archive/gallery/geis-1028-cytochrome-c



Drew Berry

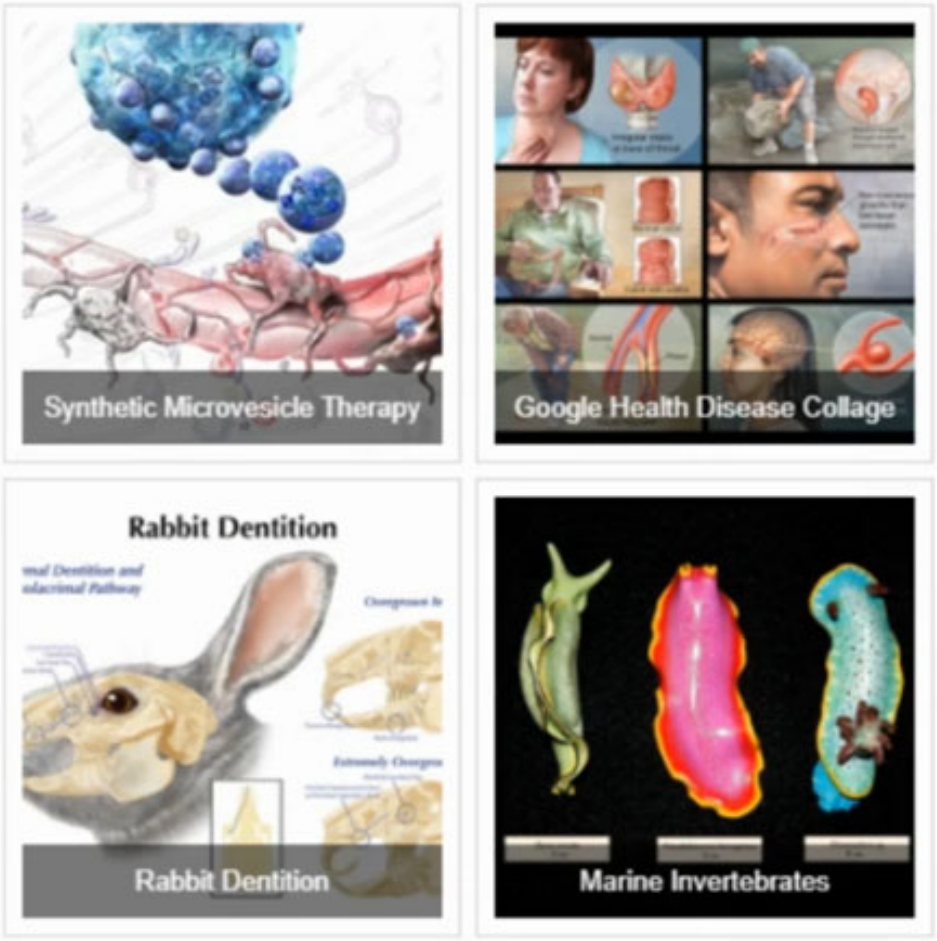
He creates stunning and scientifically accurate animations to illustrate how the molecules in our cell move and interact.
He employs many visual techniques to make the narrative clear, including consistent use of color, smooth transition to the level of detail, and design of the scene to include recognizable features that bridge scale levels.

Image : Stills from Animations of unseeable biology by Drew Berry
https://www.youtube.com/watch?v=WFCvkkDSfIU

Cells at Work!

Cells at Work! was first created by Akane Shimizu for magazine in March 2017.
It is much appreciated, major part of which is due to how scientifically accurate the series is despite being entertainment.

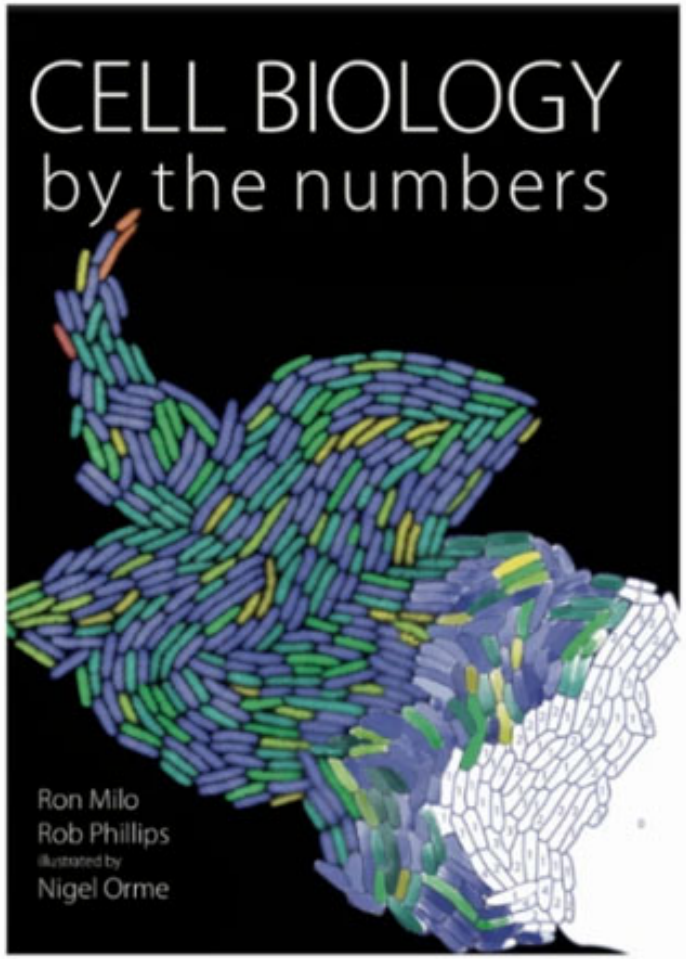
Image : Cells at Work! Characters representing red blood cell, white blood cell and platelets
https://www.netflix.com/in/title/81028791



Guild of Natural Science Illustration

GNSI is a global non-profit professional organization for all artists who work in the field of visual science communication.
It has been an important site for exploration and learning about scientific visualization.

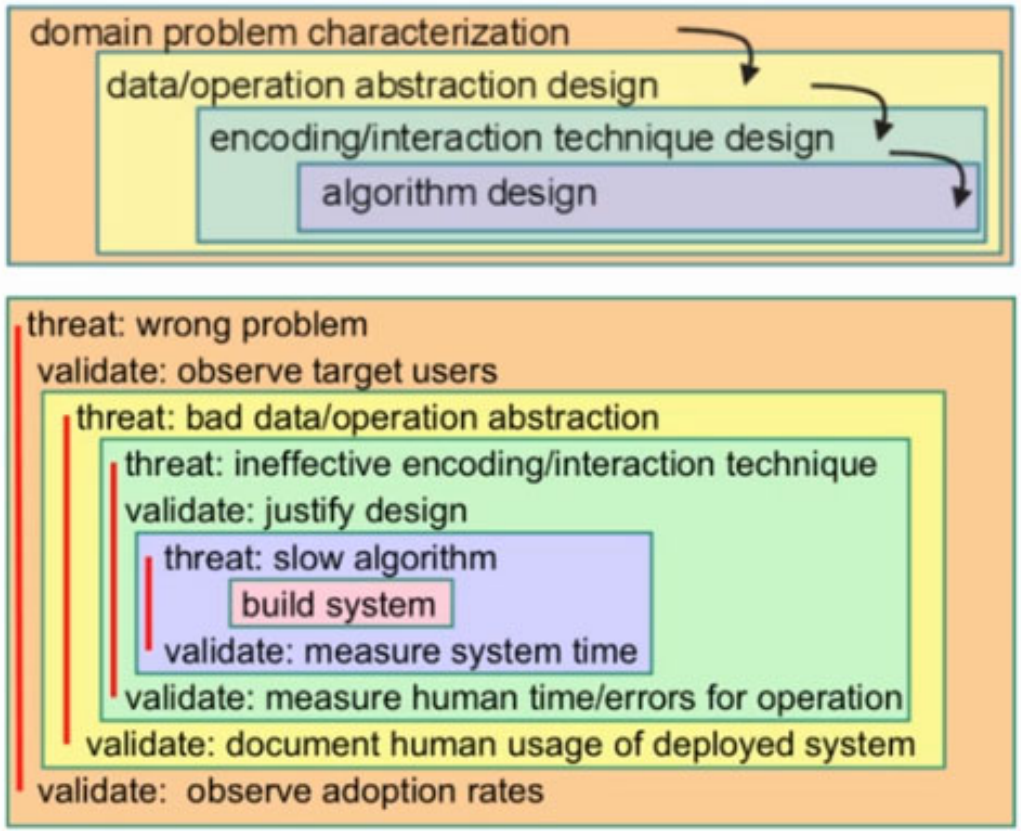
Image : Screenshot of an exhibit gallery on the GNSI website.
https://www.gnsi.org/image-gallery-2018-member-exhibit



Cell Biology by the Numbers

Intended for biology majors, the book addresses questions related to cell volume and cell component sizes along with absolute numbers for many biological objects.

Data Visualization Model



Munzner's Nested Model

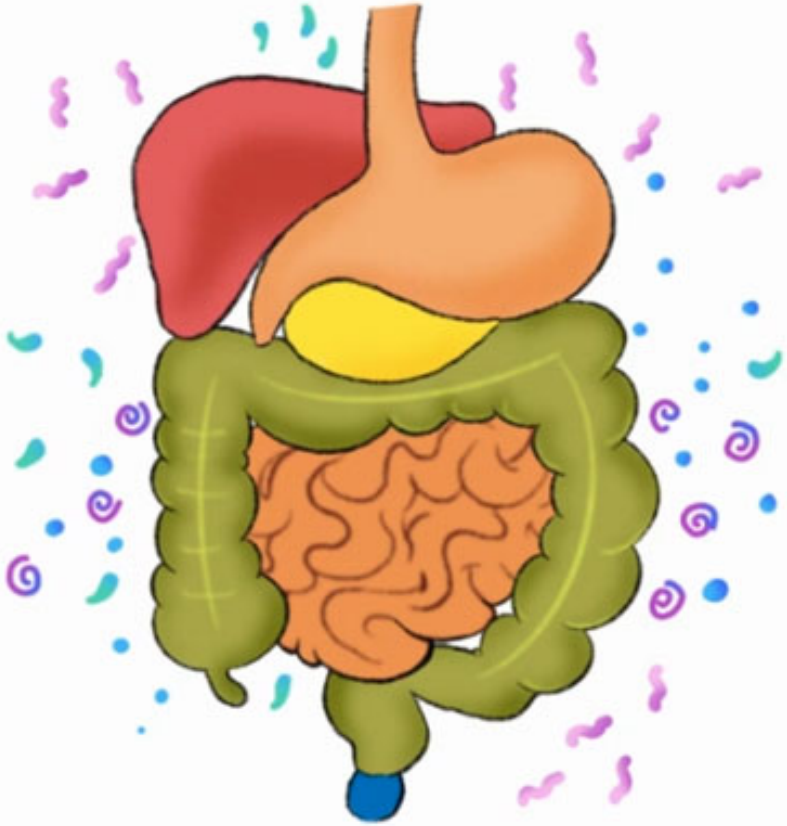
Tamara Munzner's Nested Model for visualization design and validation.
Identify threats and validation for each layer in association with this project.

Nested-layers of Tamara Munzner's model of visualization creation showing threats and validation for each layer.
https://www.cs.ubc.ca/labs/imager/tr/2009/NestedModel/NestedModel.pdf

Domain Characterization

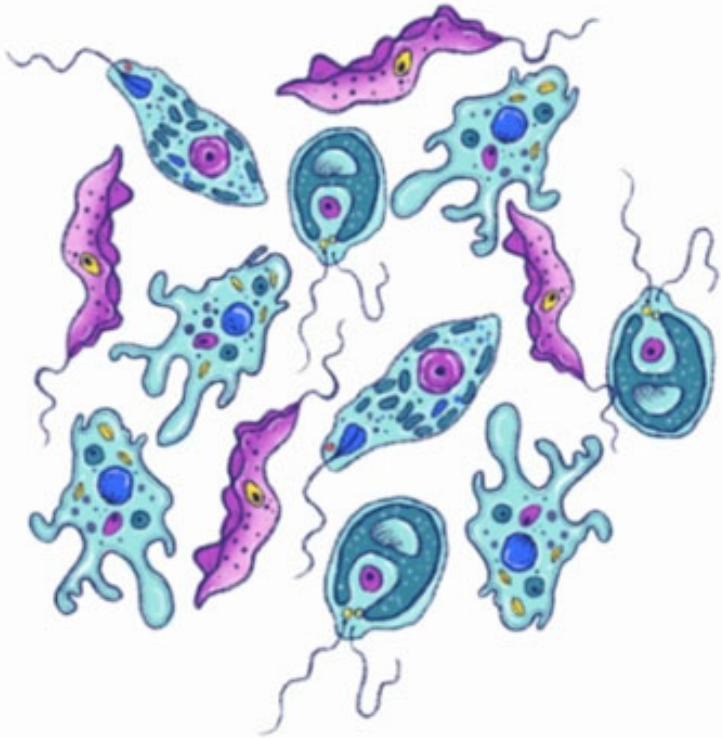
Domain Knowledge

Identification of various sub-fields and data sources for information gathering.
Research to identify and understand the data from each sub-field, the data formats and the processing of information.



Data Sources and Availability

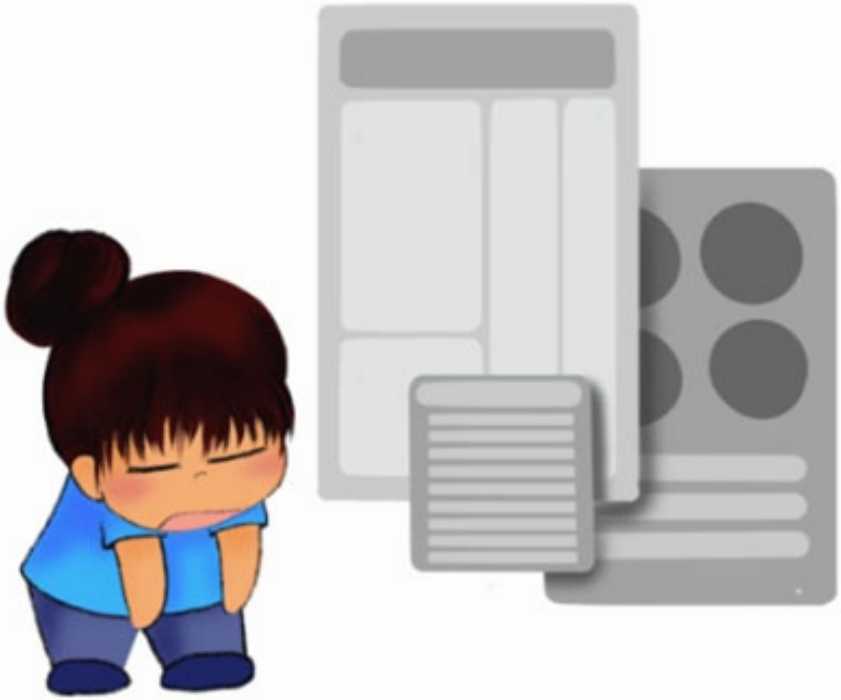
The human body contains trillions of microorganisms, outnumbering human cells by 10 to 1.
Human Microbiome Project was initiated in 2007 to identify and characterize the human microbial fauna.



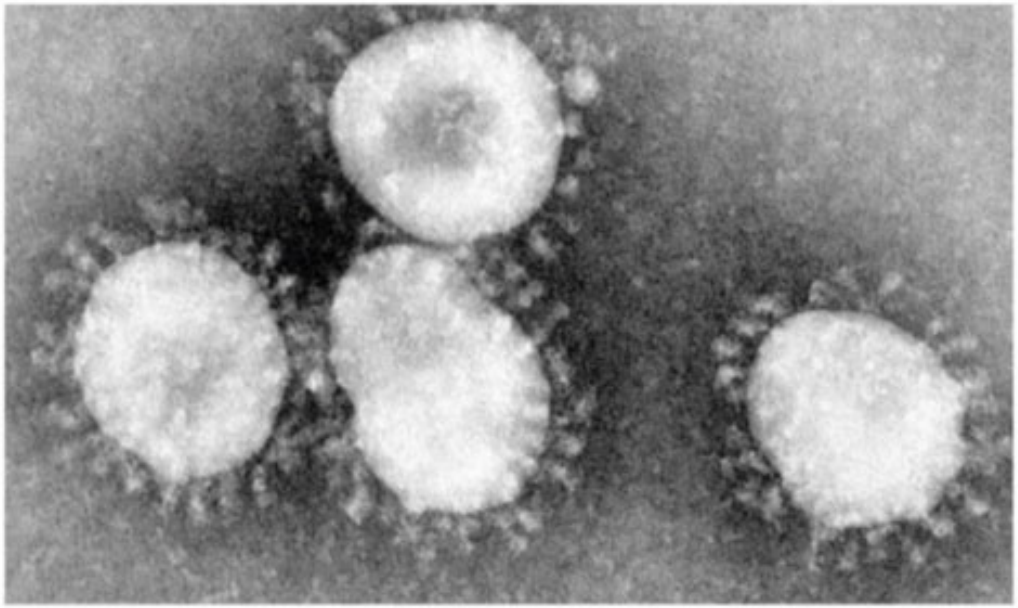
Data Sources and Availability

Extensive research has been done on microorganisms causing diseases worldwide, and a lot of information and research is available publically.

Data Type Abstraction

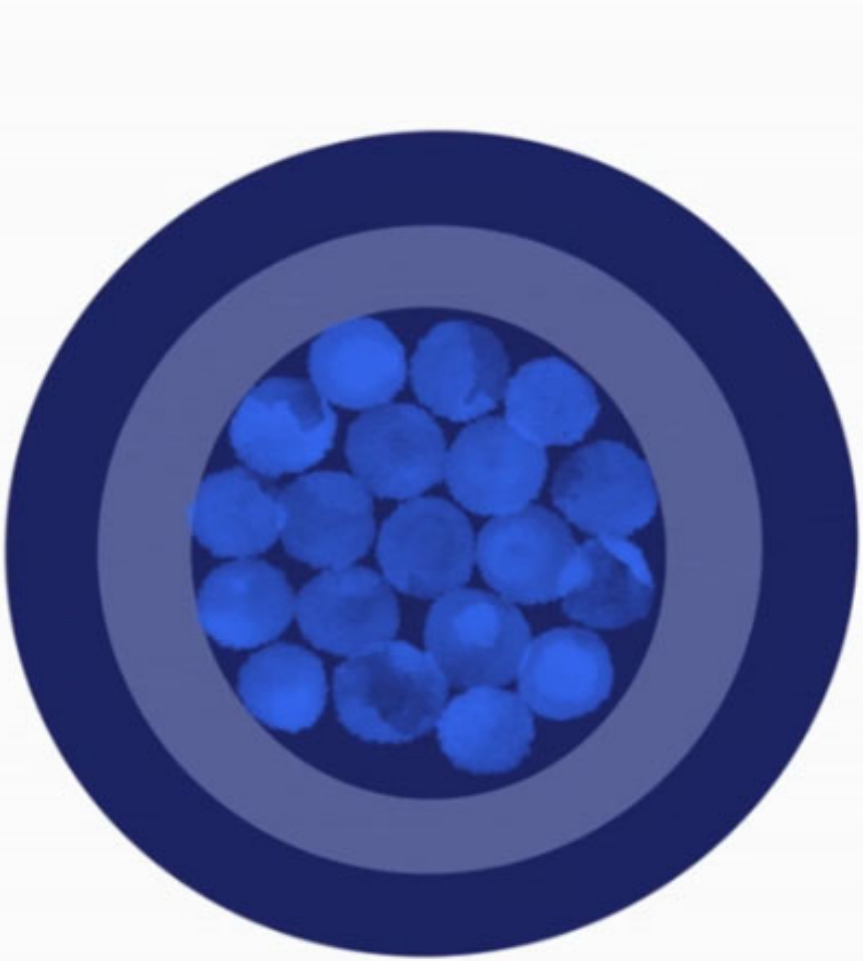


Cellular Ultrastructure



This information is most often obtained from electron microscopy, such as negative-stained thin sections for cross sectional mesoscale models and tomograms for 3D mesoscale models.

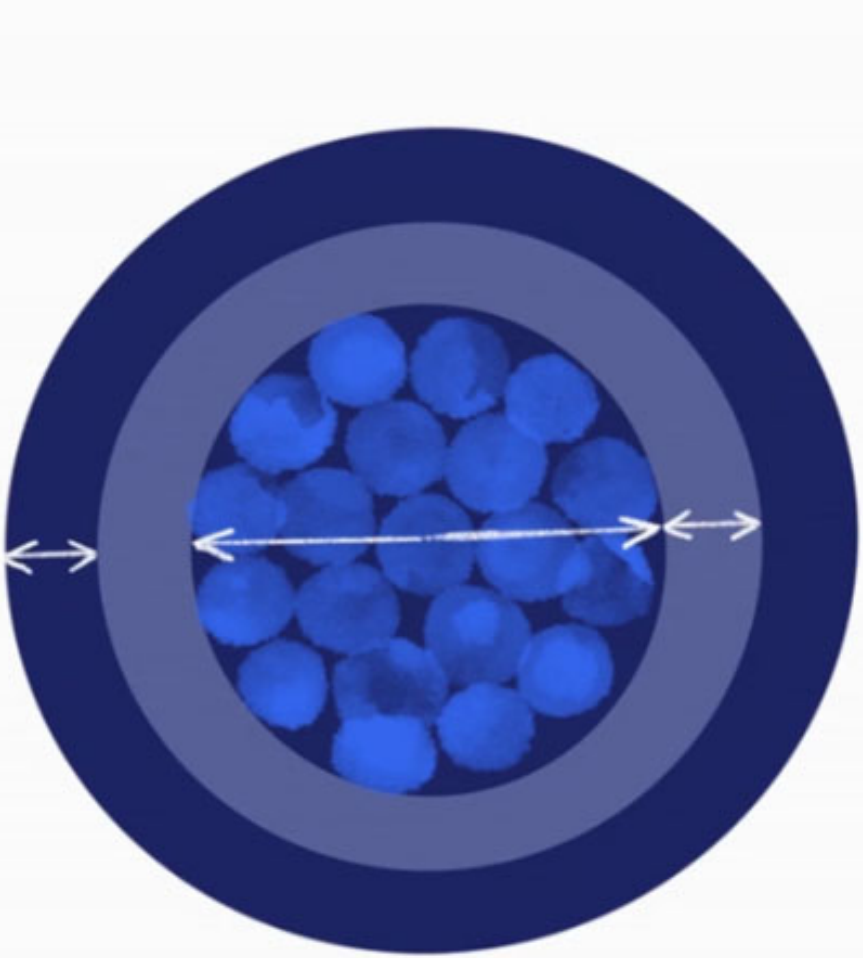
Image : Electron microscopy image of SARS Coronavirus
https://phil.cdc.gov/Details.aspx?pid=15523



Cellular Ultrastructure

This information is most often obtained from electron microscopy, such as negative-stained thin sections for cross sectional mesoscale models and tomograms for 3D mesoscale models.

These data reveal description of parts of the ultrastructure membranes and location of components of larger sizes.



Cellular Ultrastructure

This information is most often obtained from electron microscopy, such as negative-stained thin sections for cross sectional mesoscale models and tomograms for 3D mesoscale models.

These data reveal description of parts of the ultrastructure membranes and location of components of larger sizes.

	Entry	Entry name		Protein names	Gene names	Organism	Length	
☐	Q7TFA1	NS7B_CVHSA		**Protein non-structural 7b**	, 7b	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	44	
☐	P59595	NCAP_CVHSA		**Nucleoprotein**	**N**, 9a	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	422	
☐	P0C6X7	R1AB_CVHSA		**Replicase polyprotein 1ab**	**rep**, 1a-1b	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	7,073	
☐	P0C6U8	R1A_CVHSA		**Replicase polyprotein 1a**	, 1a	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	4,382	
☐	P59636	ORF9B_CVHSA		**Protein 9b**	, 9b	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	98	

Molecular Composition

A comprehensive list of proteins and nucleic acids that are synthesized can be provided by genomic information that is available on various online sources such as National Centre for Biotechnology Information (NCBI) and UniProt.

Image : Part of the table showing components of SARS Coronavirus available on UniProt
https://www.uniprot.org/proteomes/UP000000354

Functional category and most similar gene product	Spot identification	pI/M_r Observed	pI/M_r Expected	Accession number/PID	Most similar organism
Cell envelope					
HMW2	HA11	6.5/80	8.62/216.25	P47460	*M. genitalium* [a]
HMW2	HA12 (isoform)	6.5/80	8.62/216.25	P47460	*M. genitalium* [a]
HMW2	HA13 (isoform)	6.5/80	8.62/216.25	P47460	*M. genitalium* [a]
HMW3	A6	5.6/59	6.64/72.7	Q50369	*M. pneumoniae*
P200	A3	4.3/80	4.07/116.92	P75211	*M. pneumoniae* [c]
MgPa	A106 [c]	5.3/10	7.88/14.60	P20796	*M. genitalium* [a]
Central intermediary metabolism					
Inorganic pyrophosphatase	A55	6.2/25	5.62/21.63	P47593	*M. genitalium*
Glycerophosphoryl diester					

[a] Combined approach employed to add statistical weight;

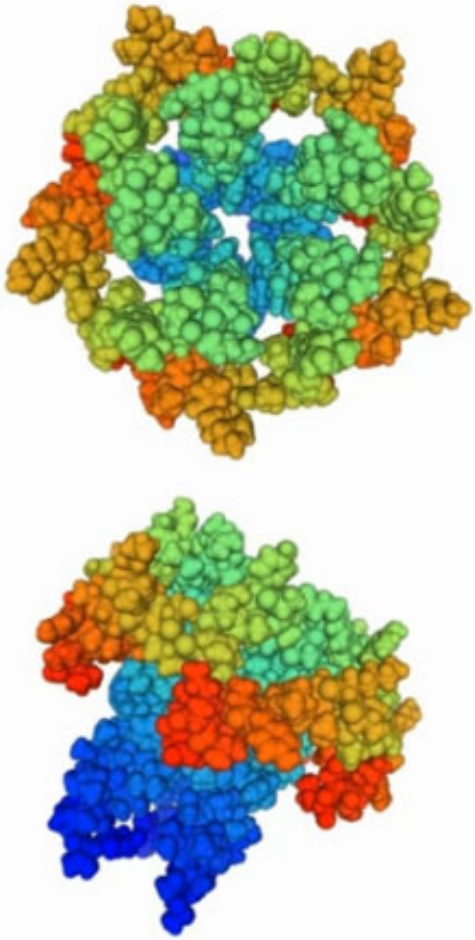
[b] protein spots found within >1 window of protein expression;

[c] proteins absent from late exponential phase gels.

Molecular Composition

Proteomic studies, which are a large-scale study of proteins produced in an organism, can define the proteins and estimate their abundance for different compartments in the cell.

Image : Part of table showing various components of Mycoplasma genitalium available as part of its proteomic study
https://febs.onlinelibrary.wiley.com/doi/full/10.1046/j.1432-1327.2000.01183.x

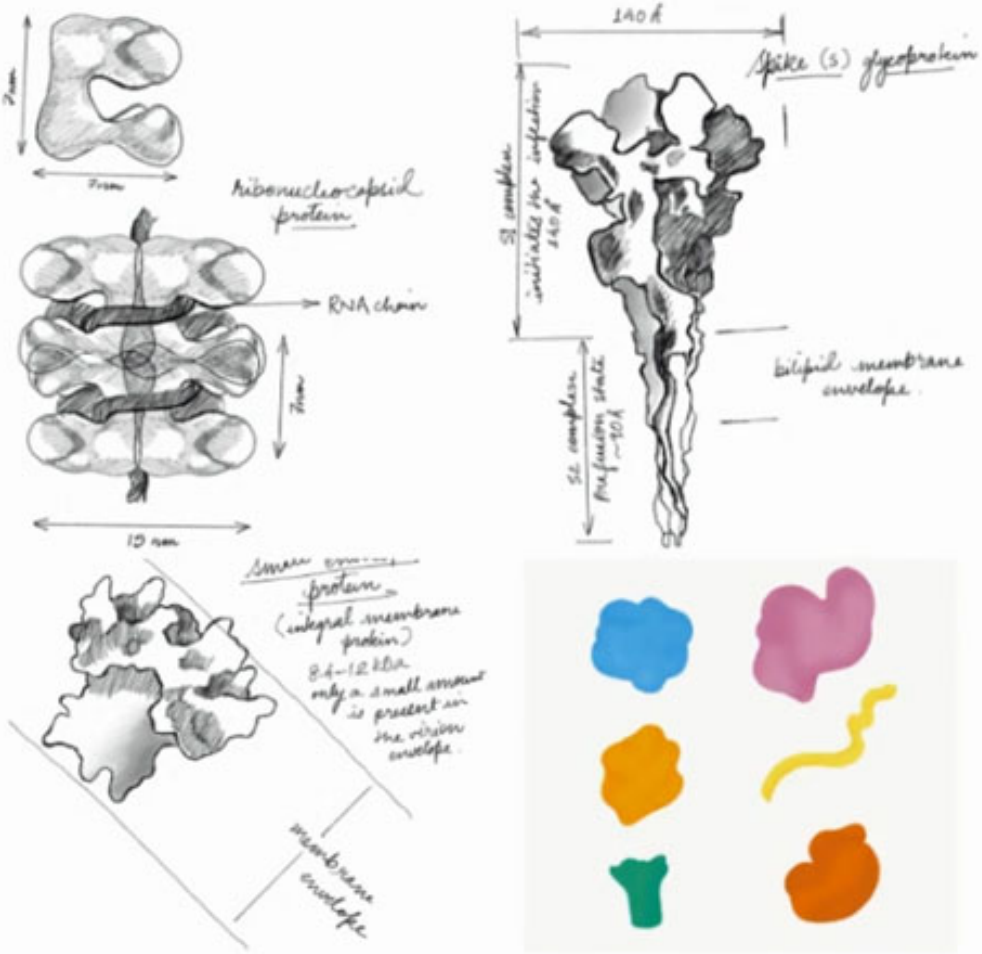


Molecular Structure

Structures are available for a large number of biomolecules through the Worldwide Protein Data Bank.

Another helpful resource is the archive of illustrations of biomolecules created by David Goodsell, which is also available on Protein Data Bank.

Image : Spacefill structure of Envelope small membrane protein, an integral membrane protein of SARS Coronavirus
https://www.uniprot.org/uniprot/P59637

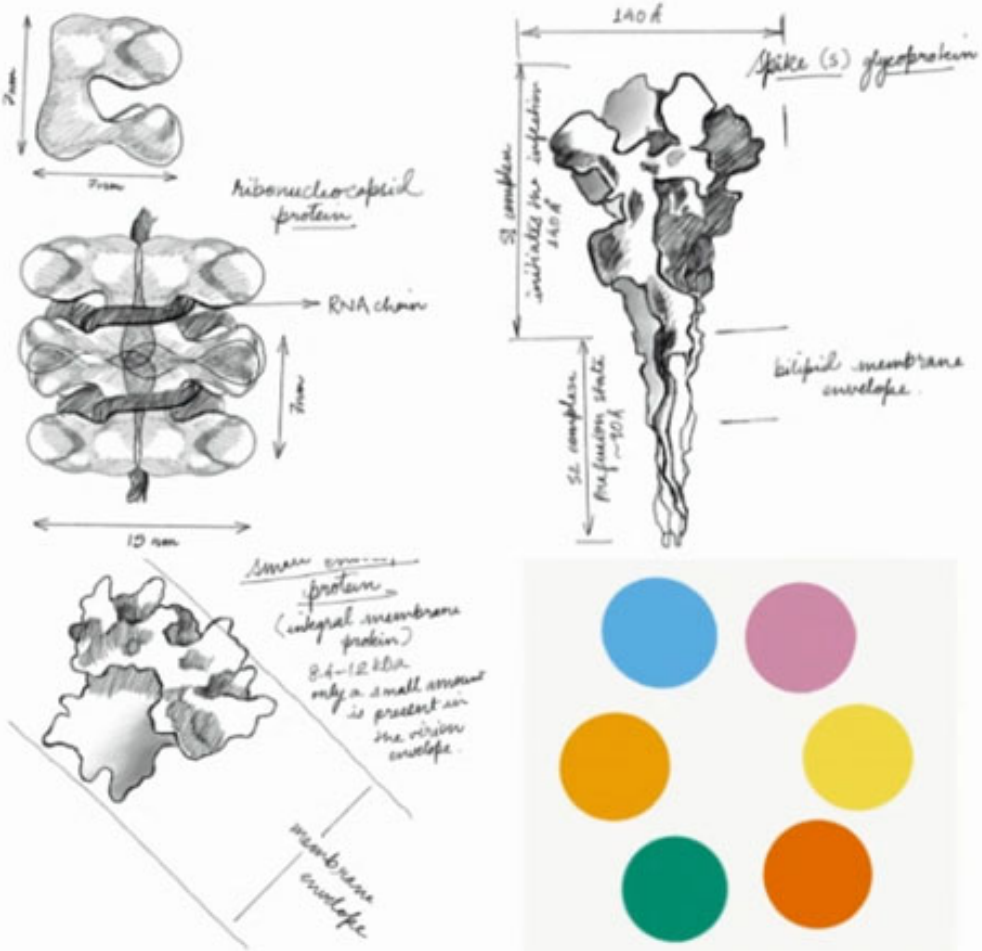


Molecular Structure

Structures are available for a large number of biomolecules through the Worldwide Protein Data Bank.

Another helpful resource is the archive of illustrations of biomolecules created by David Goodsell, which is also available on Protein Data Bank.

UniProt is another tool that is very helpful in gathering this information.

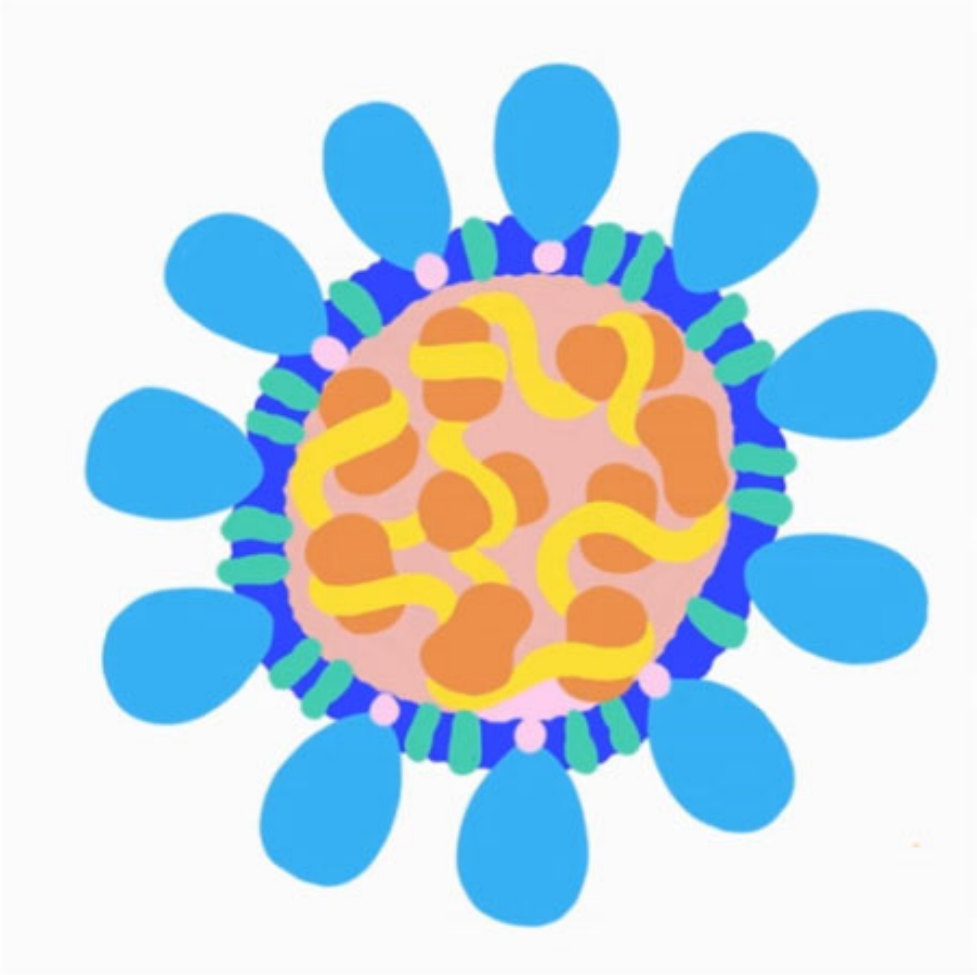


Molecular Structure

Structures are available for a large number of biomolecules through the Worldwide Protein Data Bank.

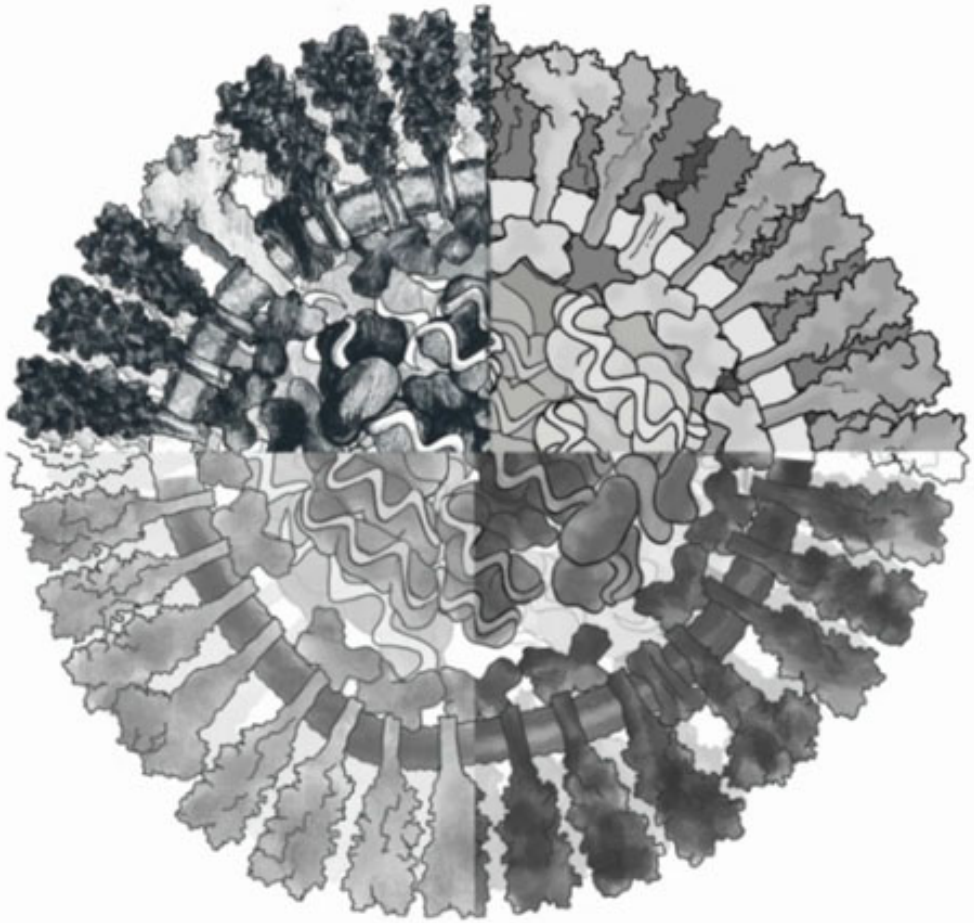
Another helpful resource is the archive of illustrations of biomolecules created by David Goodsell, which is also available on Protein Data Bank.

UniProt is another tool that is very helpful in gathering this information.

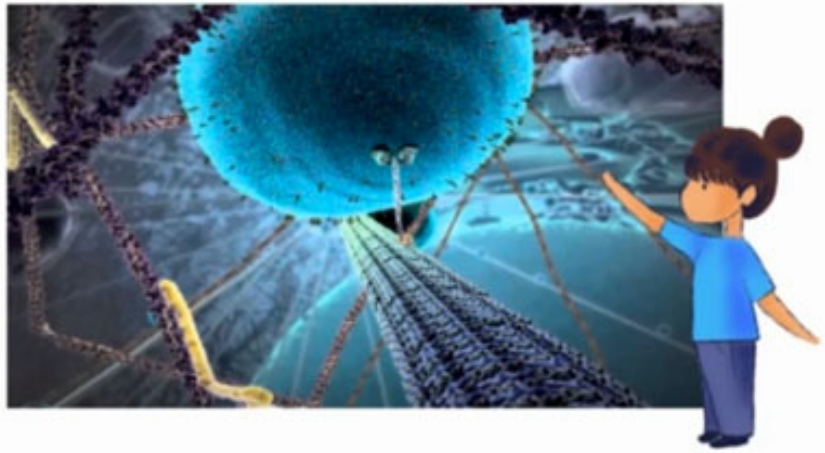


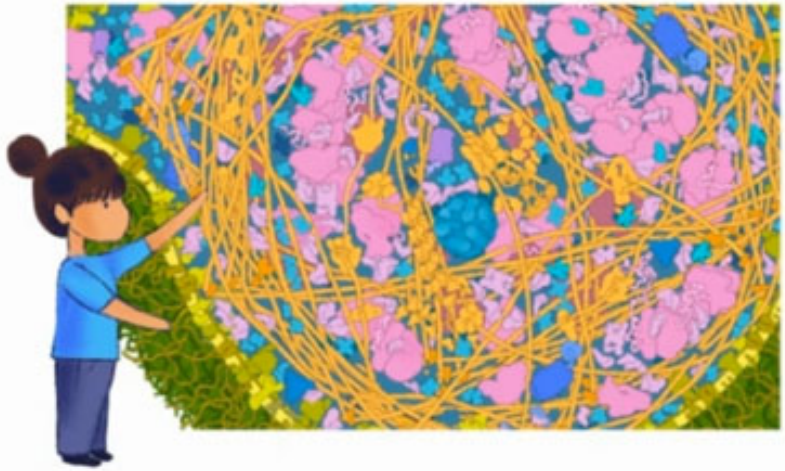
Molecular Interactions

Information on interactions between various biomolecules of a cell to determine or model the structure.



Data Encoding Design



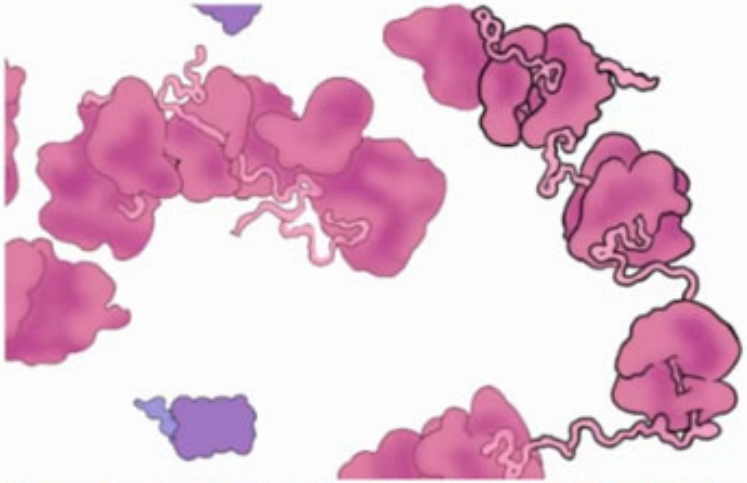


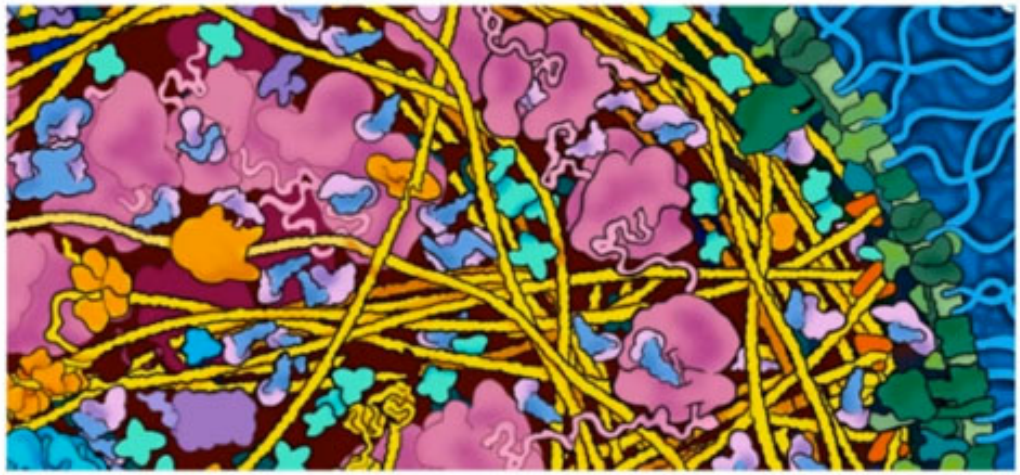
Projection

A cross-sectional approach has been employed which has little perspective, thus avoiding distortion of shape and size.

It enables the display of large sections of the cell, allowing all of the molecules to be drawn at the same scale, allowing easy comparison.

Image (Top) : Still from The Inner Life of the Cell
https://www.youtube.com/watch?v=wJyUtbn0O5Y





Visual Representation Style

A flat cartoony approach is employed.

The molecules are drawn with simple outlines and flat colors are used to highlight the packing and distribution of molecules instead of the details of each individual molecule.

Image : Structure of ribosomal complexes separately, and within the cellular environment. Part of the visualization attempted for Mycoplasma genitalium.

Color Encoding

Amino Acids	colour Name	Sample	RGB Values	Hexadecimal
ASP, GLU	Bright Red		[230,230, 10]	E60A0A
CYS, MET	Yellow		[230,230, 0]	E6E600
LYS, ARG	Blue		[20, 90,255]	145AFF
SER, THR	Orange		[250,150, 0]	FA9600
PHE, TYR	Mid Blue		[50, 50,170]	3232AA
ASN, GLN	Cyan		[0,220,220]	00DCDC
GLY	Light Grey		[235,235,235]	EBEBEB
LEU, VAL, ILE	Green		[15,130, 15]	0F820F
ALA	Dark Grey		[200,200,200]	C8C8C8
TRP	Purple		[180, 90,180]	B45AB4
HIS	Pale Blue		[130,130,210]	8282D2
PRO	Flesh		[220,150,130]	DC9682
Others	Tan		[190,160,110]	BEA06E

Color encoding for 21 amino acids.

Using amino acid composition of each of the molecules to generate their color.

Image: Chart showing color encoding for all the amino acids.
http://acces.ens-lyon.fr/biotic/rastop/help/colour.htm

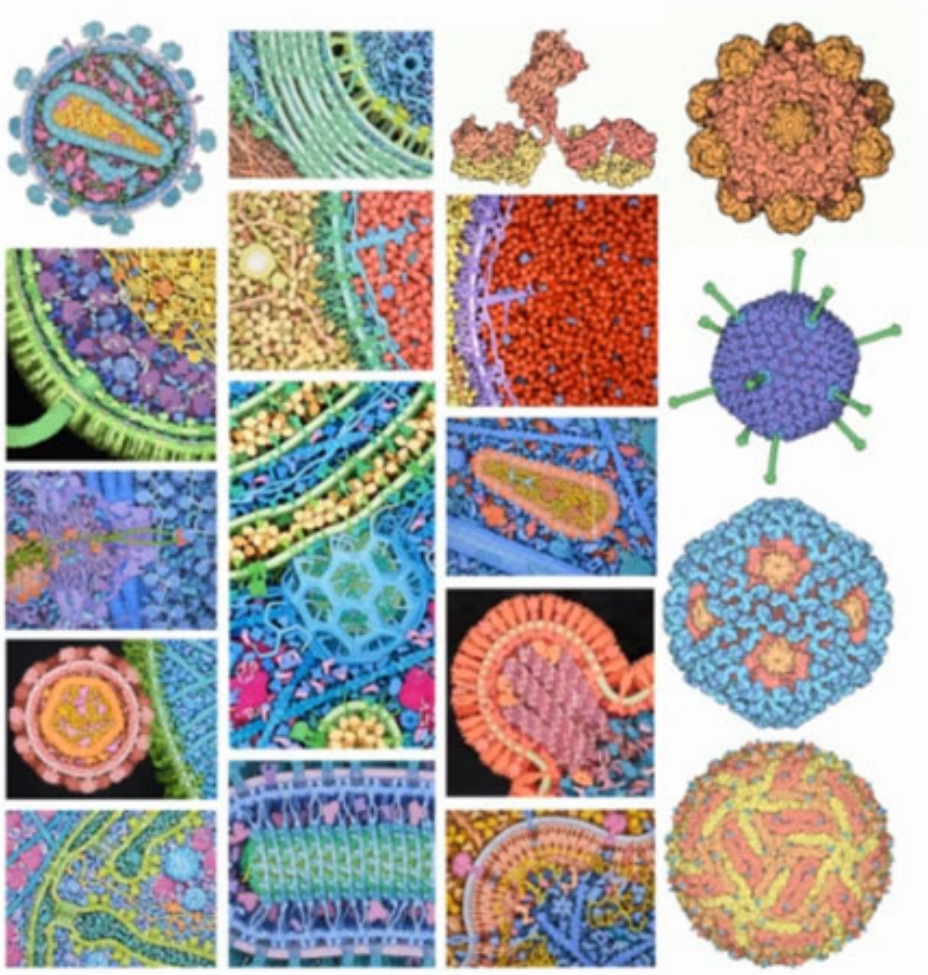
Color Encoding

Amino Acids	colour Name	Sample	RGB Values	Hexadecimal
ASP, GLU	Bright Red		[230,230, 10]	E60A0A
CYS, MET	Yellow		[230,230, 0]	E6E600
LYS, ARG	Blue		[20, 90,255]	145AFF
SER, THR	Orange		[250,150, 0]	FA9600
PHE, TYR	Mid Blue		[50, 50,170]	3232AA
ASN, GLN	Cyan		[0,220,220]	00DCDC
GLY	Light Grey		[235,235,235]	EBEBEB
LEU, VAL, ILE	Green		[15,130, 15]	0F820F
ALA	Dark Grey		[200,200,200]	C8C8C8
TRP	Purple		[180, 90,180]	B45AB4
HIS	Pale Blue		[130,130,210]	8282D2
PRO	Flesh		[220,150,130]	DC9682
Others	Tan		[190,160,110]	BEA06E

Difficult to group components based on the functions or subcellular locations across various visualisations.

May lead to visualizations with indistinguishable components.

Image: Chart showing color encoding for all the amino acids.
http://acces.ens-lyon.fr/biotic/rastop/help/colour.htm

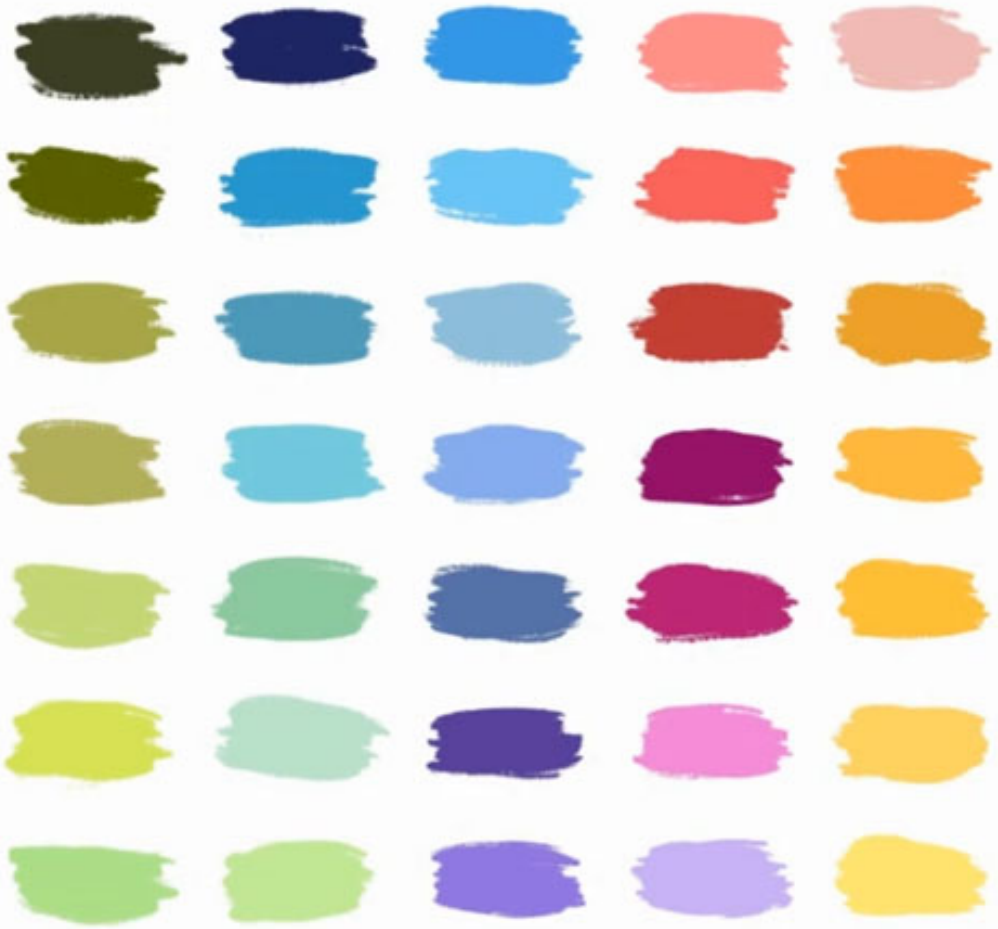


Color Encoding

A color palette was derived from Goodsell's extensive body of work.

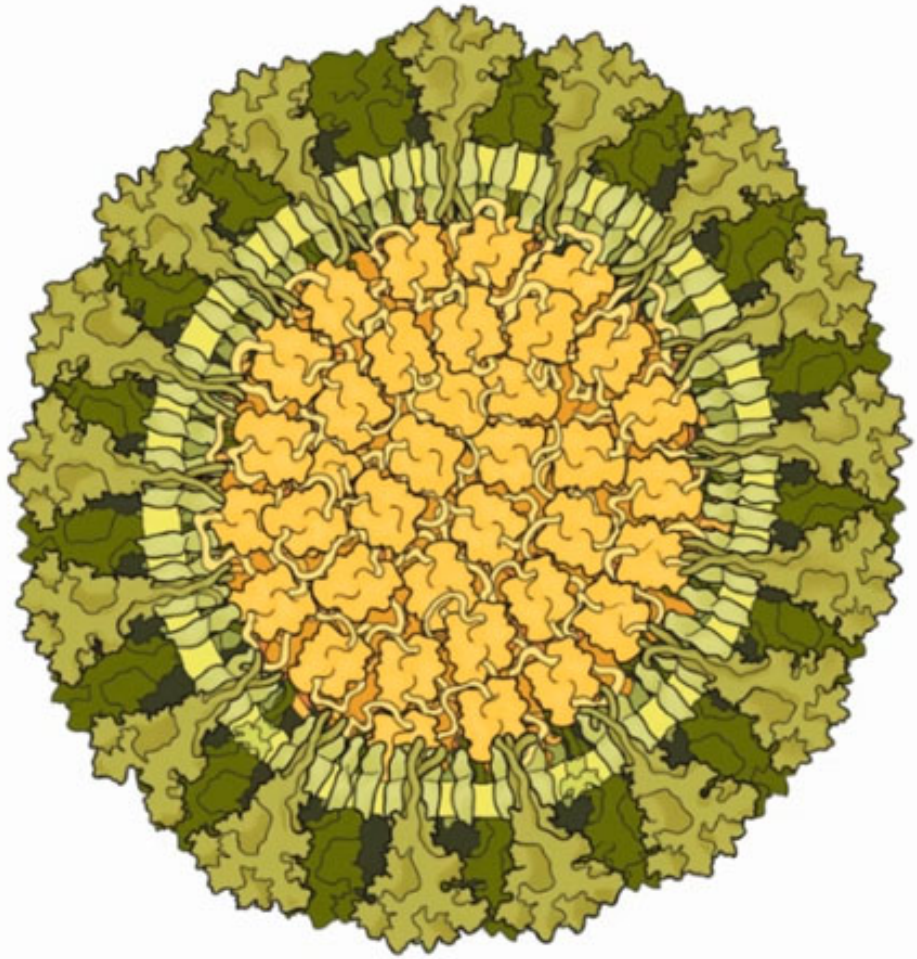
The biomolecules are color encoded on the basis of their subcellular location and then function.

Image : Works of David Goodsell
https://pdb101.rcsb.org/sci-art/goodsell-gallery/



Color Encoding

The colors are chosen to highlight the ultrastructural arrangement with the cell cytoplasm in blues, the membrane and membrane proteins in green. The molecules of central dogma, that is the molecules responsible for replication, transcription and translation are presented in shades of yellow.



Color Encoding

Difficulty in distinguishing between closely placed heterogeneous components due to similar coloring.

Non-microbiology related associations of the visualizations.

Image : Scientifically accurate, hand drawn, 2d illustration of SARS Coronavirus in the described color encoding.

Color Encoding

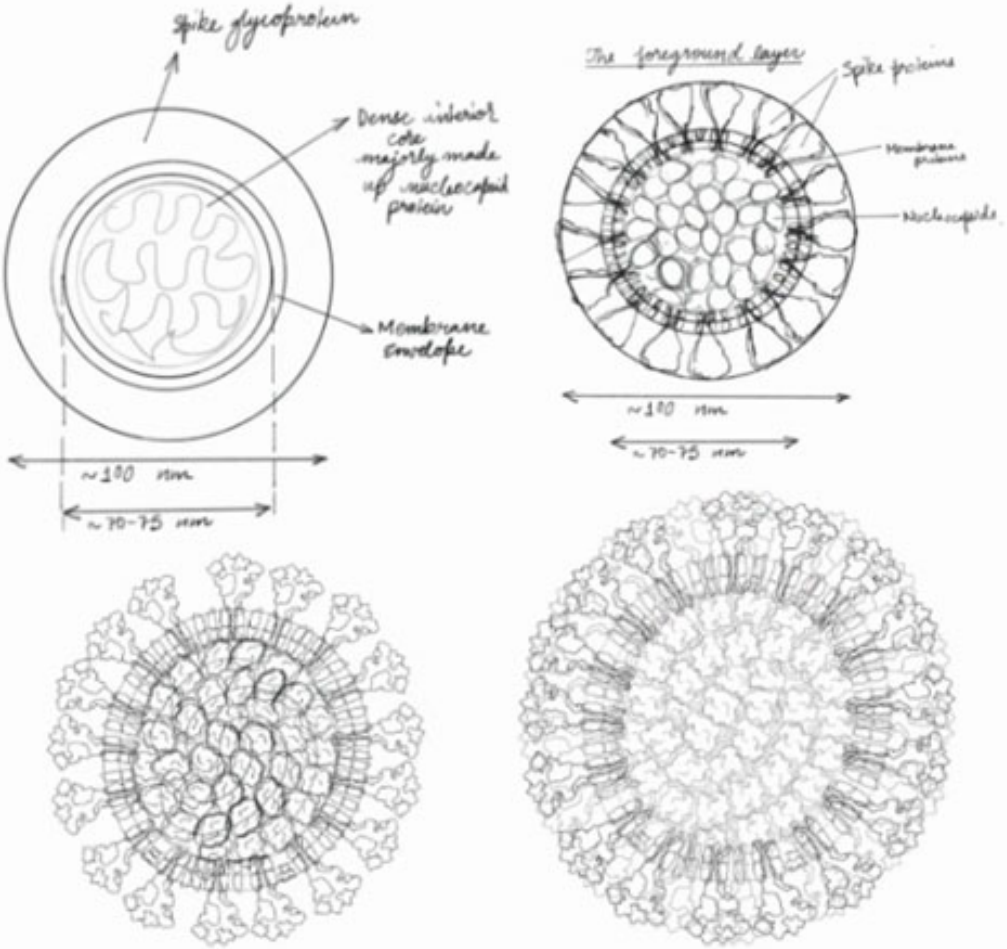
Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

Palette created to make the visualizations more accessible by using colors optimized for color blind individuals.

Previous characterization of biomolecules on the basis of their biological functions and sub-cellular locations were used.

Image: Colors optimised for color-blind individuals. P and D indicate simulated colors as seen by individuals with protanopia and deuteranopia.
https://www.nature.com/articles/nmeth.1618.pdf

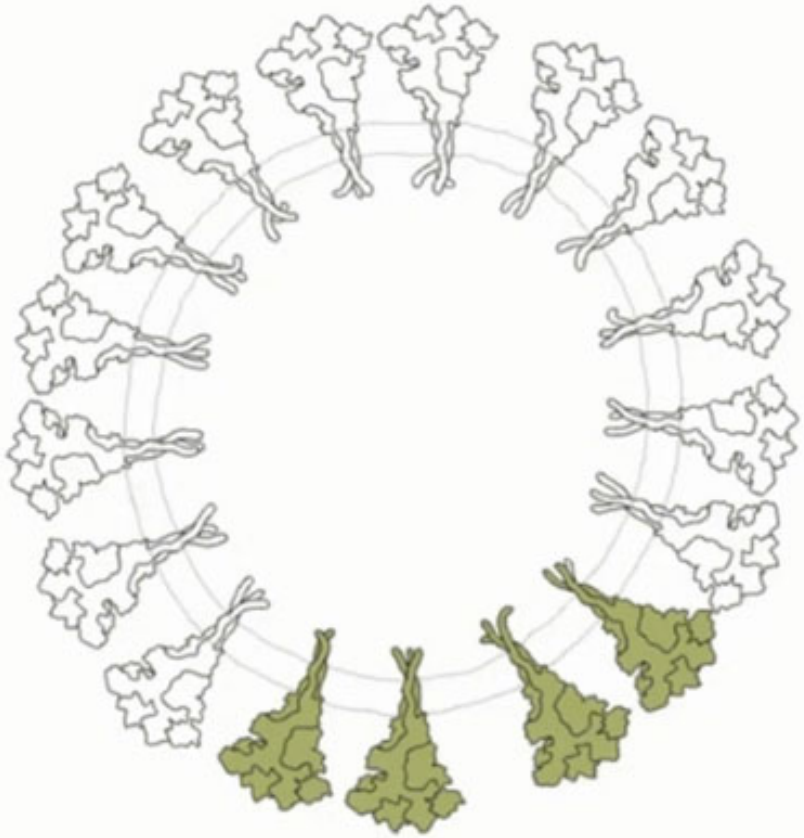
Visualization Process



Creating the Illustration

We work from the ultrastructure to the molecular structure. We start with the largest and the most important players, later adding the smaller components, building up the proper overall concentration.

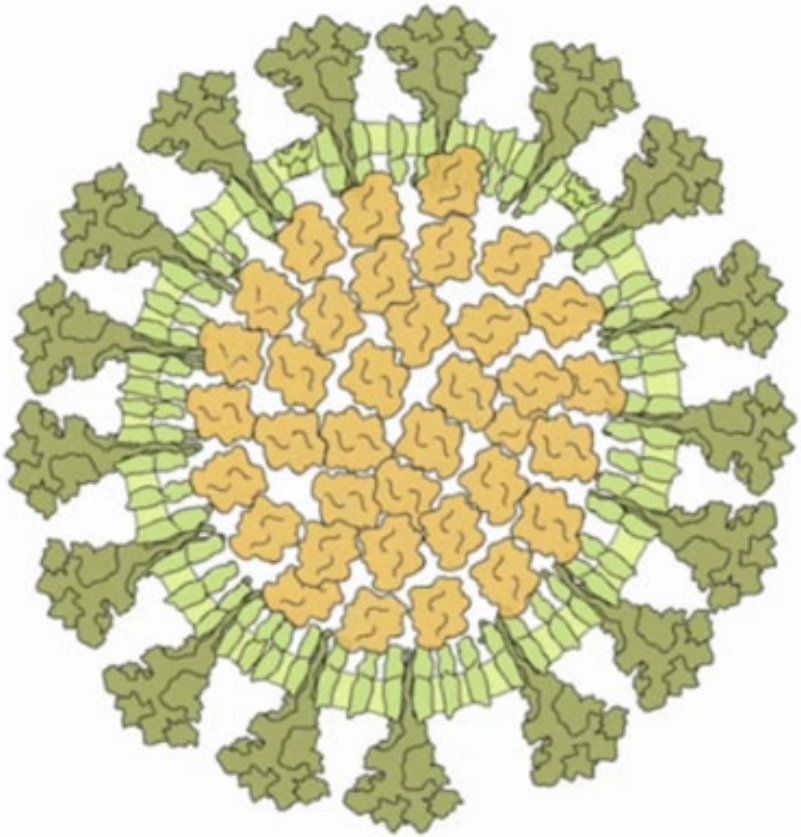
Image : Steps followed to create the scientifically accurate sketch of SARS Coronavirus. A basic scaled layout, followed by a storyboard with all the major structural components. Using storyboard as reference, various components are added layer by layer to create a single view of all the molecules lying on the same plane, which is then iterated.



Rendering the Illustration

Similar approach of creating the major components in the beginning and then moving on to the smaller ones followed while creating the sketch was followed while rendering.

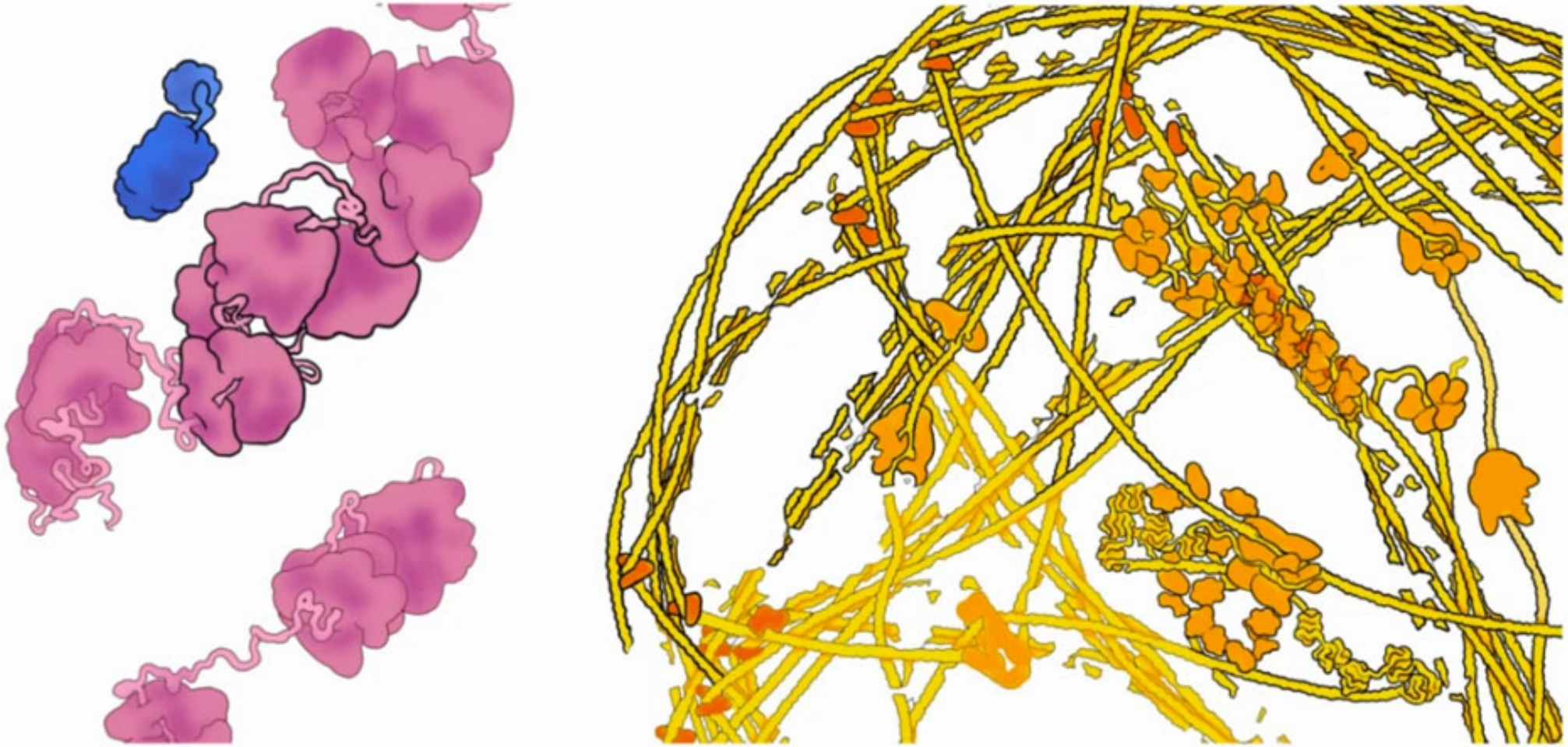
A single layer of multiple components lying on the same plane is created. Multiple layers of the same in darkened tones were added to create a sense of depth and highly packed view of the microorganism.



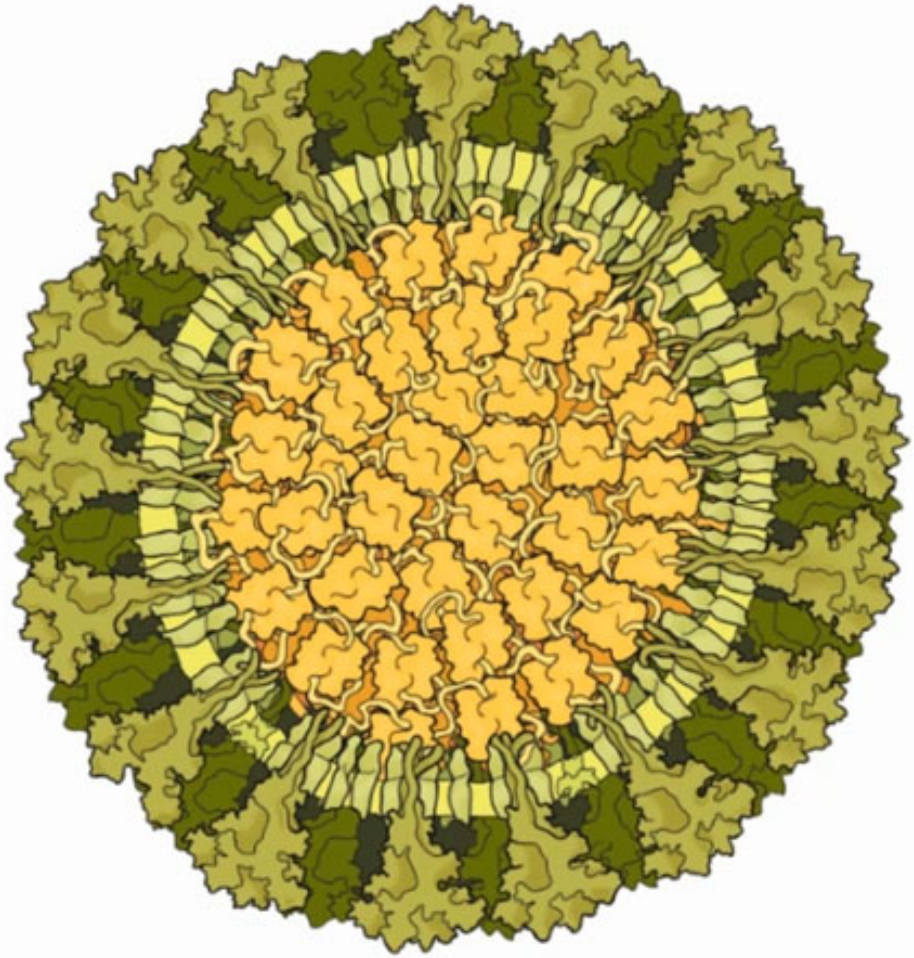
Rendering the Illustration

Similar approach of creating the major components in the beginning and then moving on to the smaller ones followed while creating the sketch was followed while rendering.

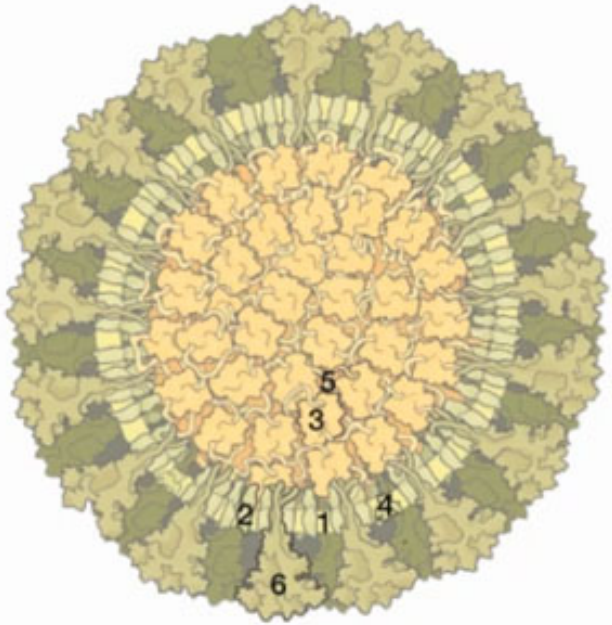
A single layer of multiple components lying on the same plane is created. Multiple layers of the same in darkened tones were added to create a sense of depth and highly packed view of the microorganism.



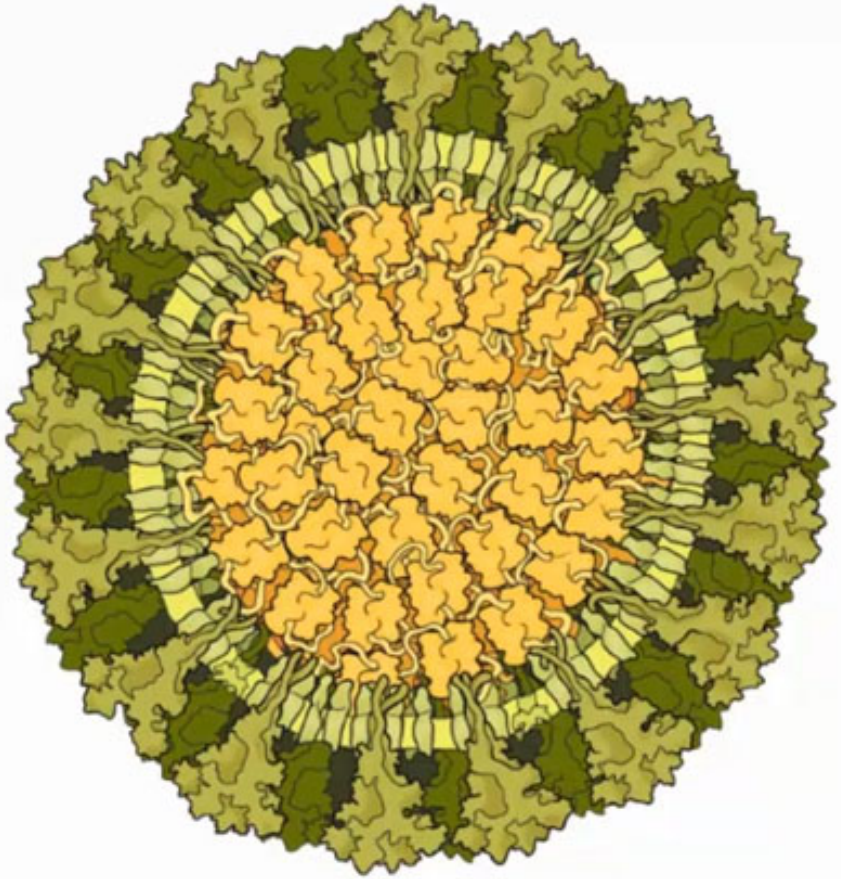
Validation



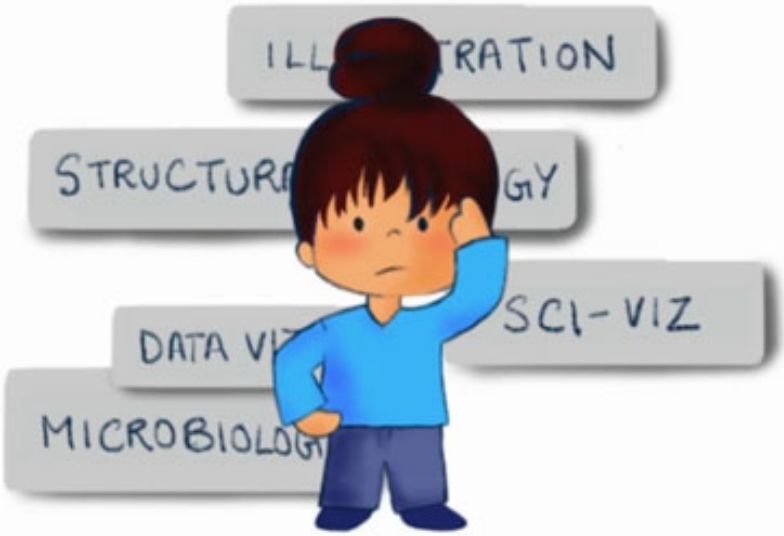
Process Validation



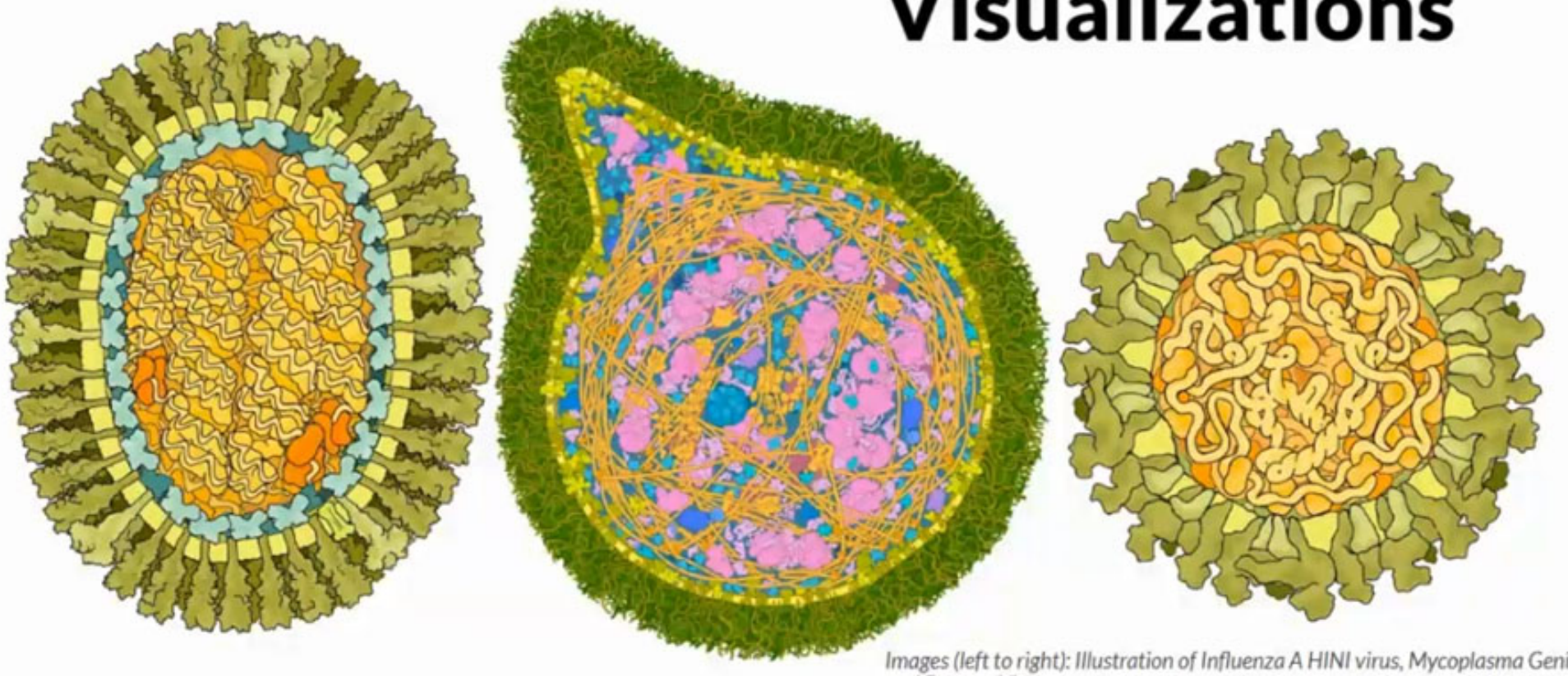
Scientifically accurate, hand drawn, 2d illustration of SARS Coronavirus. (1) Lipid bilayer. (2) Membrane protein. (3) Nucleoprotein. (4) Envelope small membrane protein. (5) RNA. (6) Spike glycoprotein



Expert Review



Visualizations



Images (left to right): Illustration of Influenza A H1N1 virus, Mycoplasma Genitalium and Dengue Virus

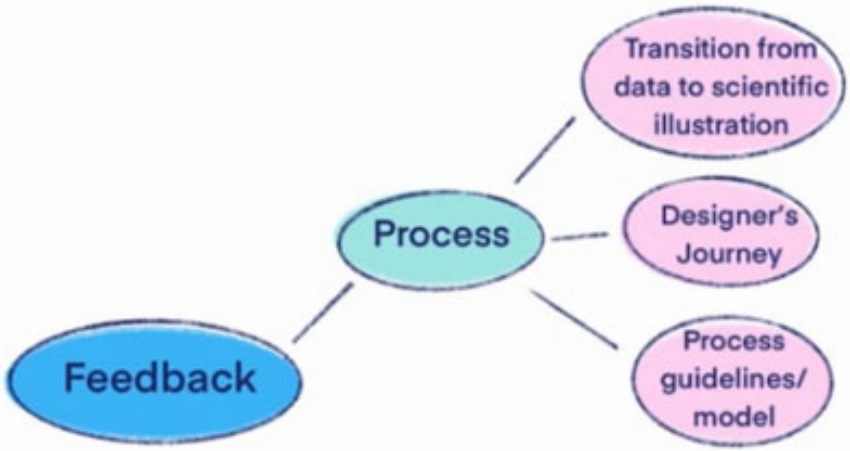
Experts

Leslee Lazar, Cognitive Neuroscientist and Science communicator, IIT Gandhinagar
Aditya Nayak, Co-founder, Plantik Biosciences
Nirupa Rao, Botanical Illustrator, National Geographic
Rasagy Sharma, Principal Information Designer, Gramener
Kuhu Gupta, General Director (Early Career), Data Visualization Society

Expert Feedback

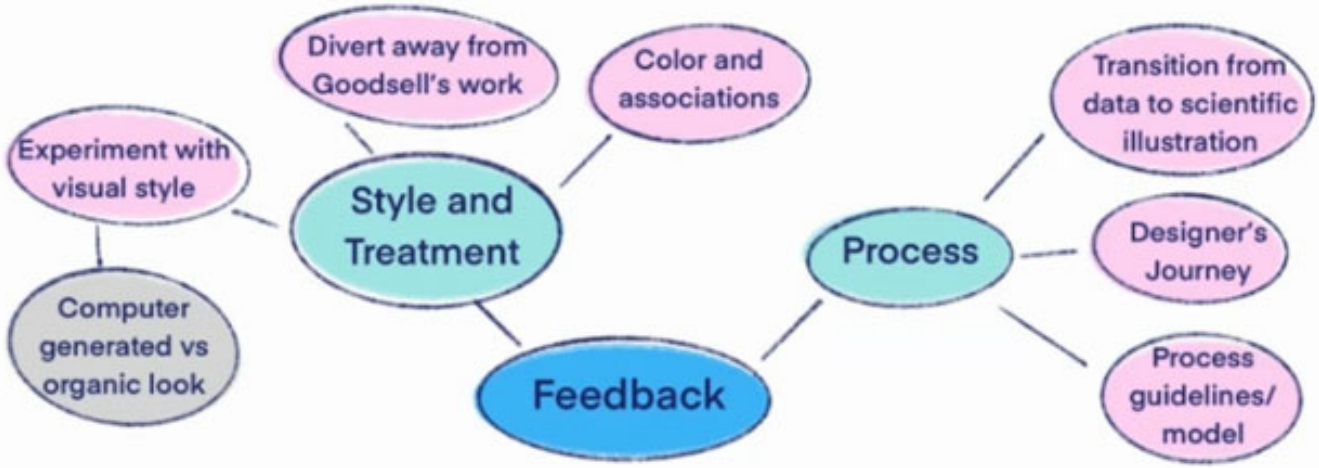
The project was presented to the mentioned panel of experts over three zoom sessions.

Process was also shared with David Goodsell.



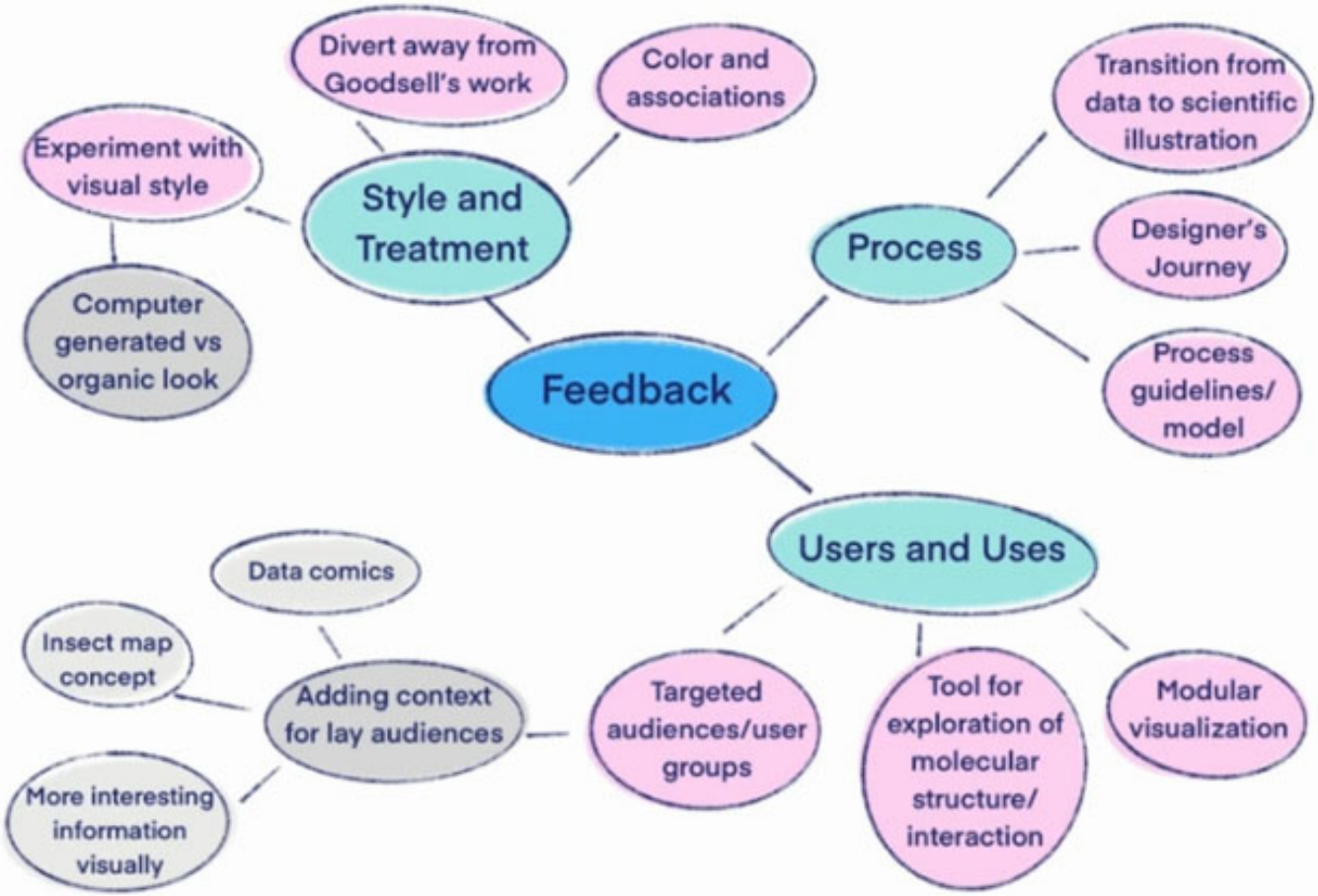
"I really admire the effort you put in to understand the science behind the structures, going to PDB and getting the illustrations done"

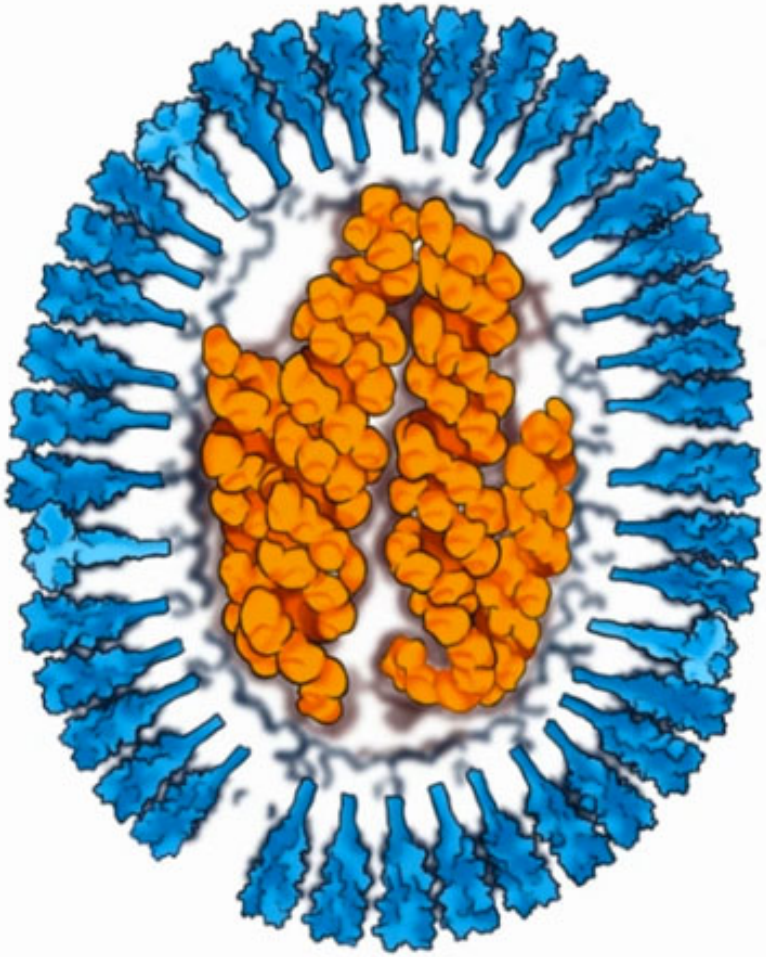
- Aditya Nayak



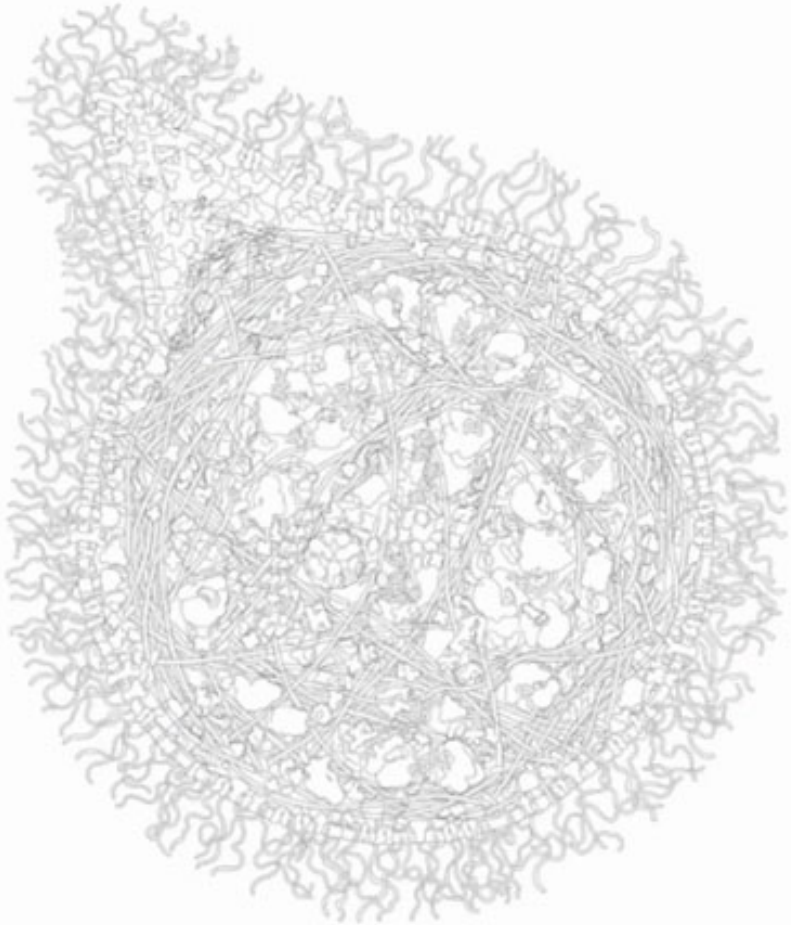
"Looks like an extension of David Goodsell's work, although you have individually worked with a process from scratch to arrive at the results. Might be a good idea to break away from Goodsell's style."

- Leslee Lazar


Divert away from Goodsell's work
Color and associations
Experiment with visual style
Style and Treatment
Computer generated vs organic look
Transition from data to scientific illustration
Process
Designer's Journey
Process guidelines/ model
Feedback
Users and Uses
Data comics
Insect map concept
Adding context for lay audiences
More interesting information visually
Targeted audiences/user groups
Tool for exploration of molecular structure/ interaction
Modular visualization



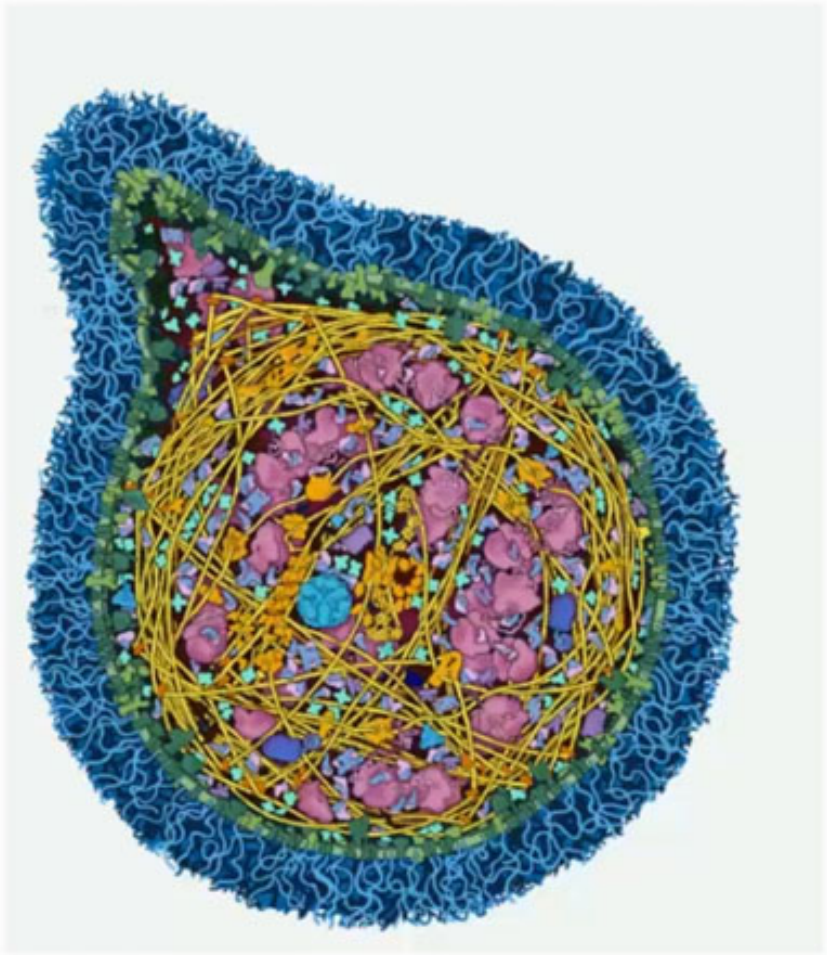
Final Visualizations

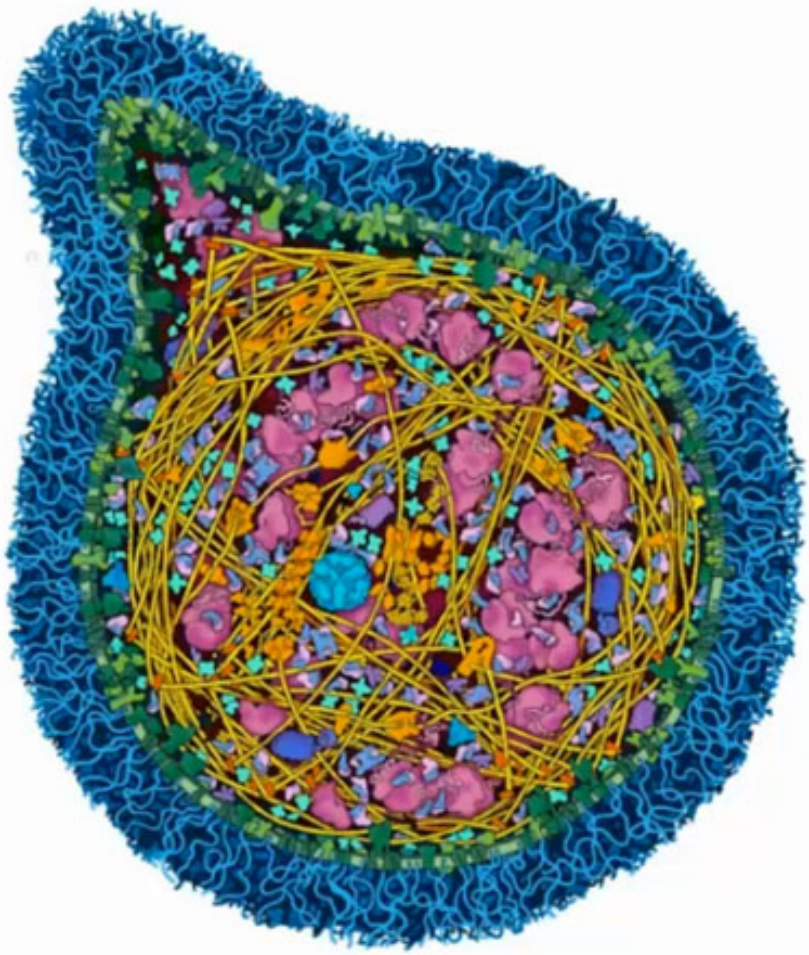


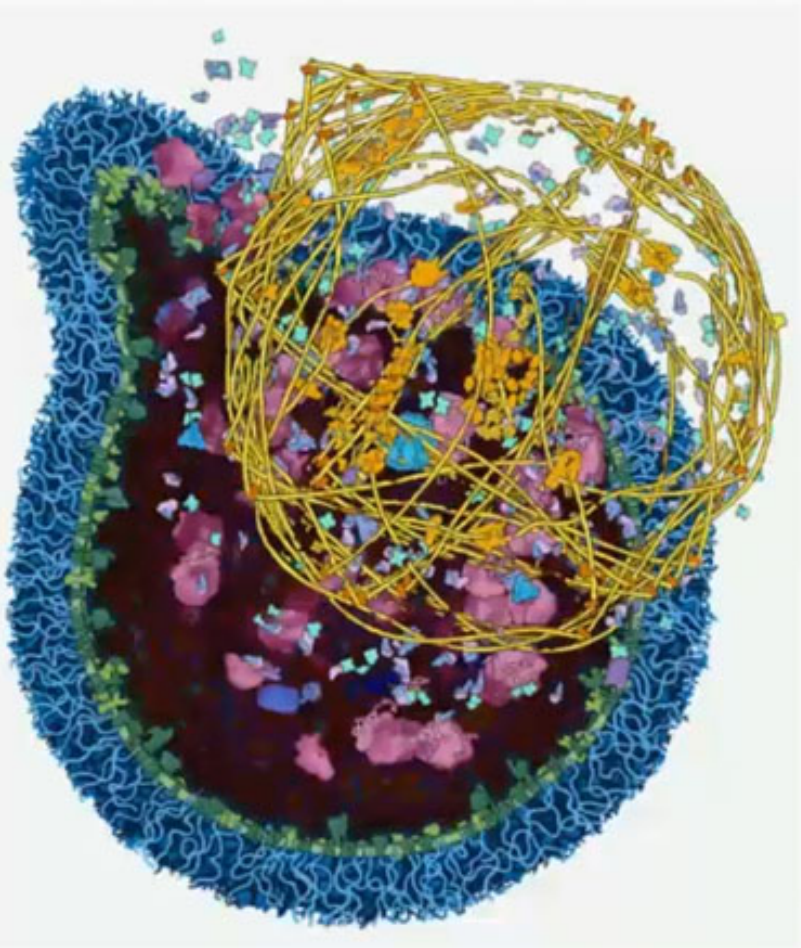
Mycoplasma Genitalium

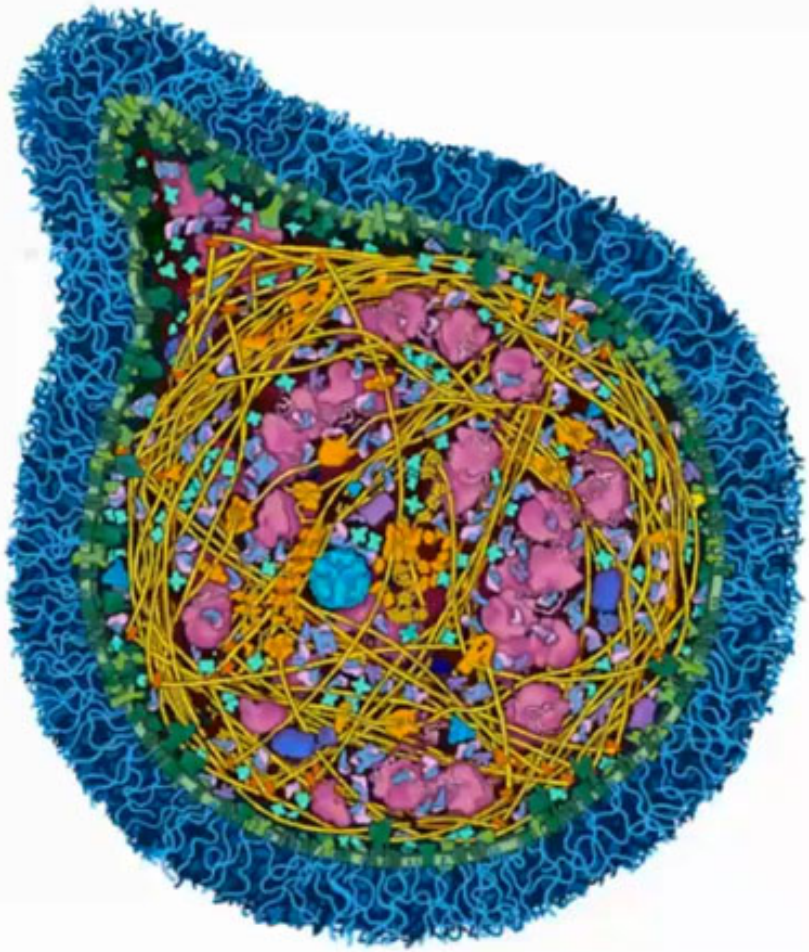
Status: Completed
No. of Iterations: 6
Time: 6 Weeks

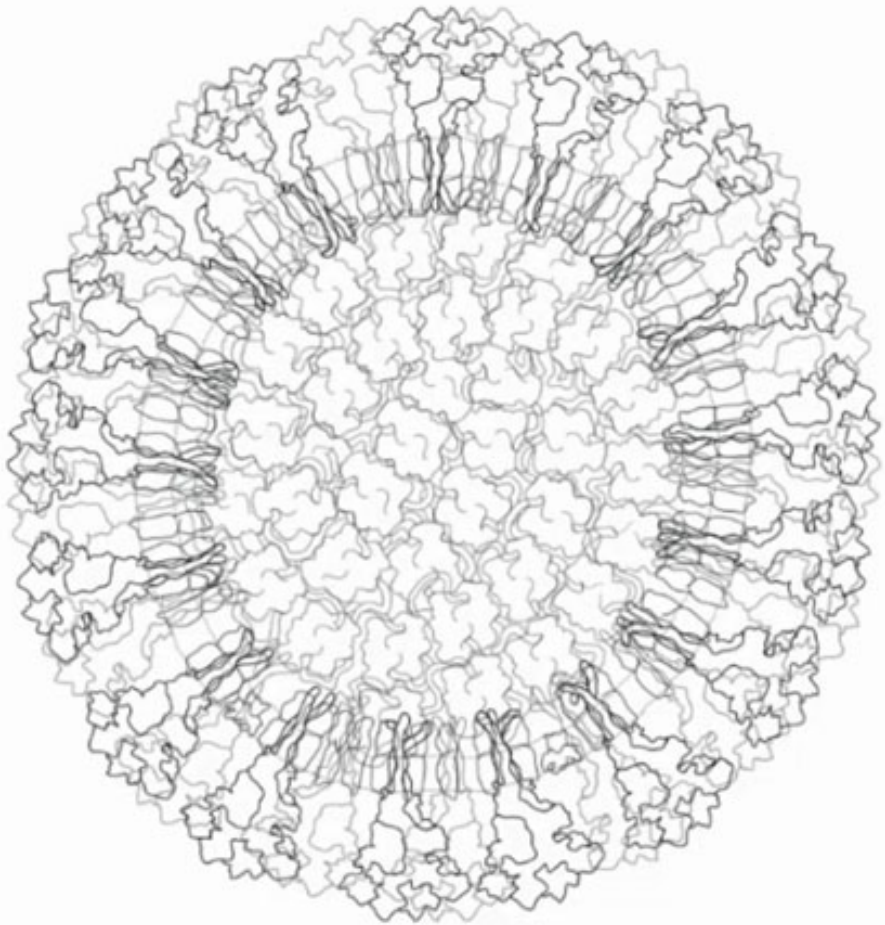
Mycoplasma genitalium is the smallest bacterial pathogen found in the human body.
It is also the most complex organism visualized and completed as part of this project.







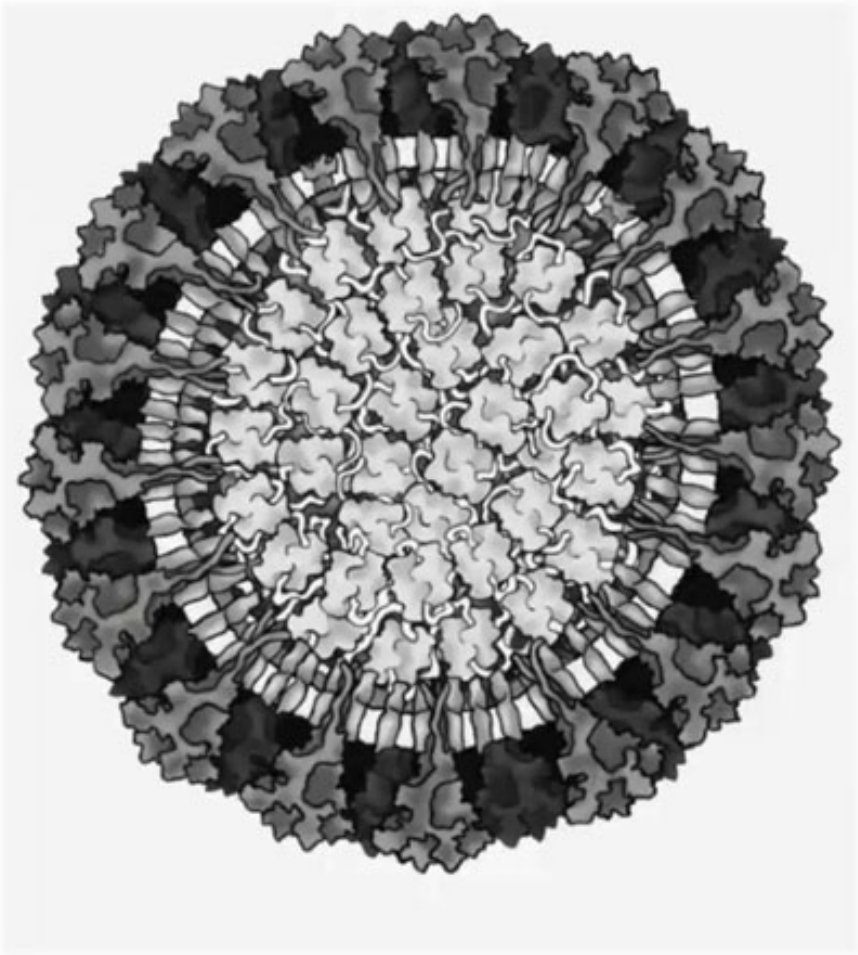


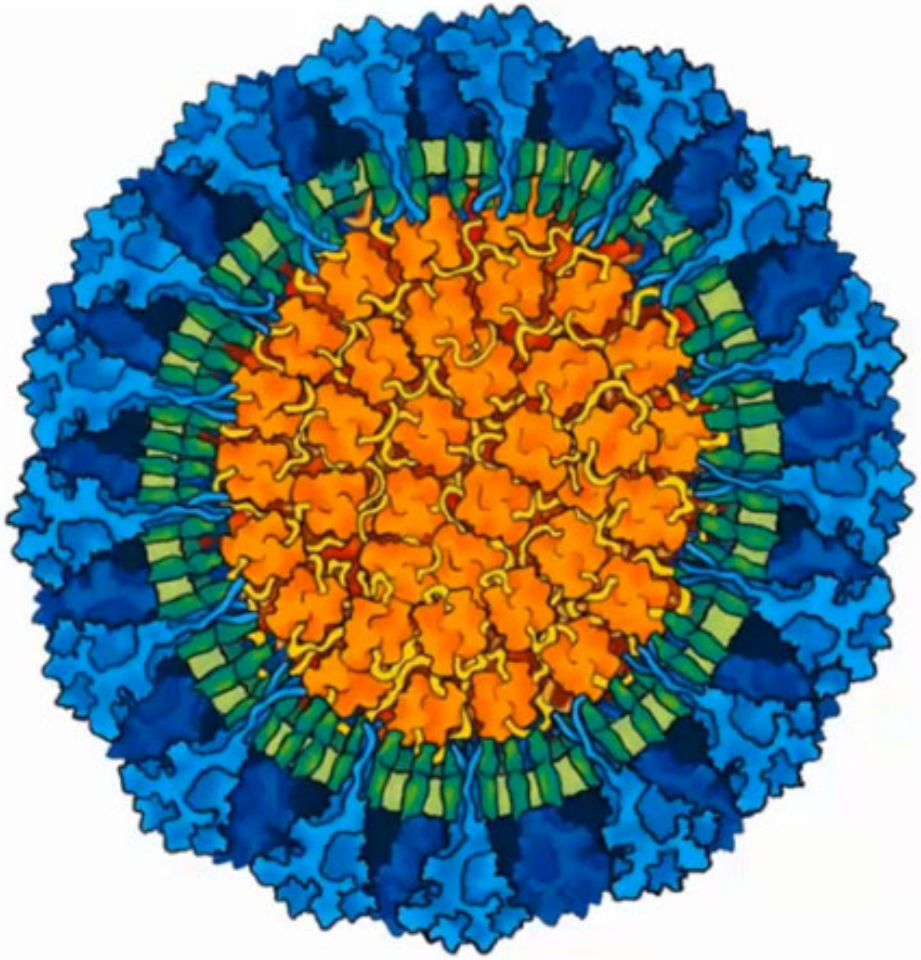


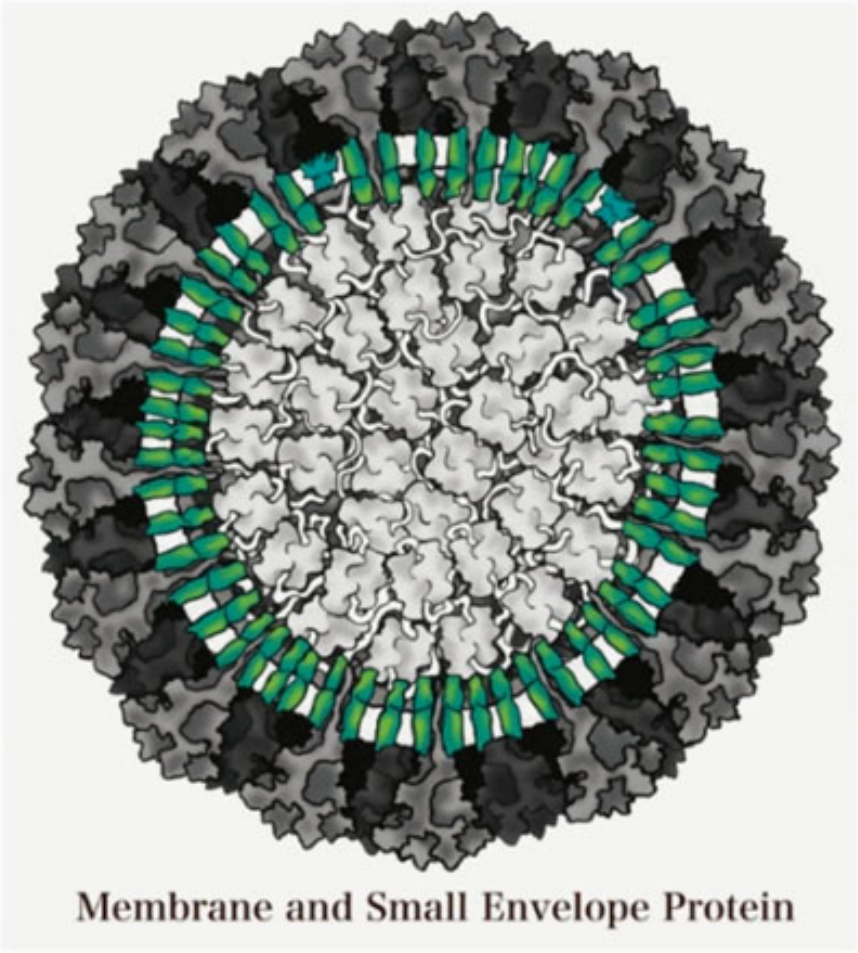
SARS Coronavirus

Status: Completed and Validated
No. of Iterations: 4
Time: 4 Weeks

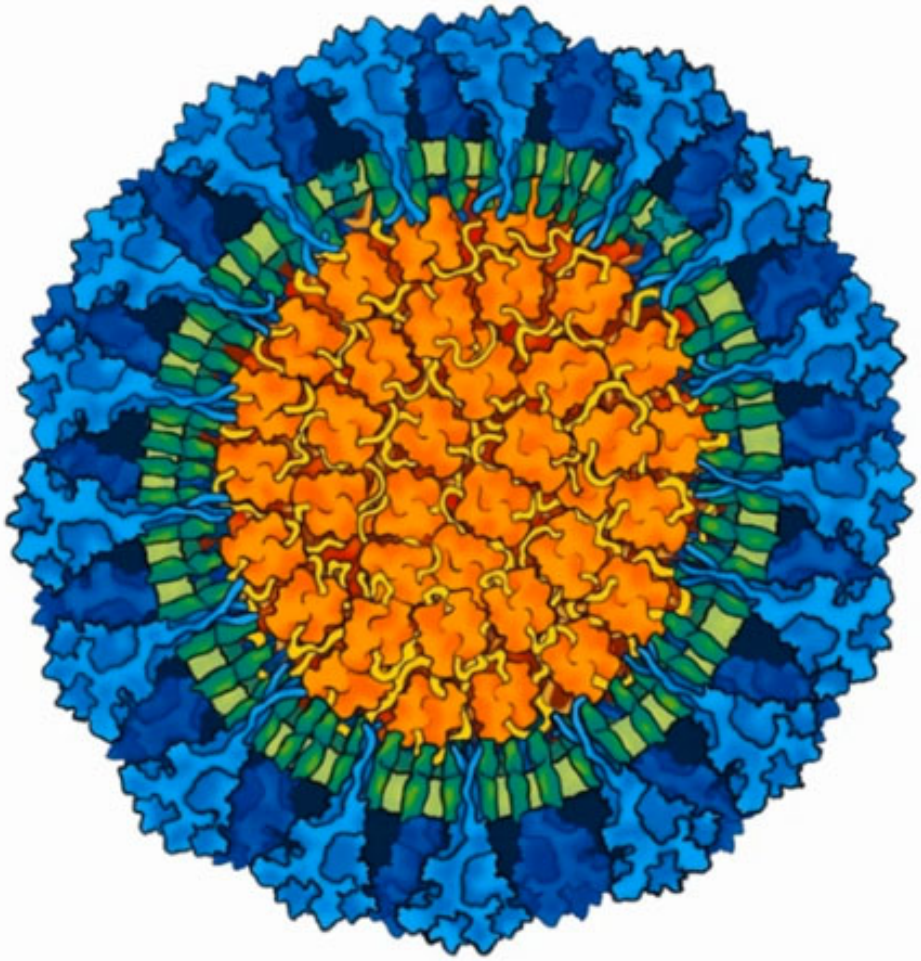
Information gathering for this particular organism was easiest as compared to other microorganisms due to the current situation.

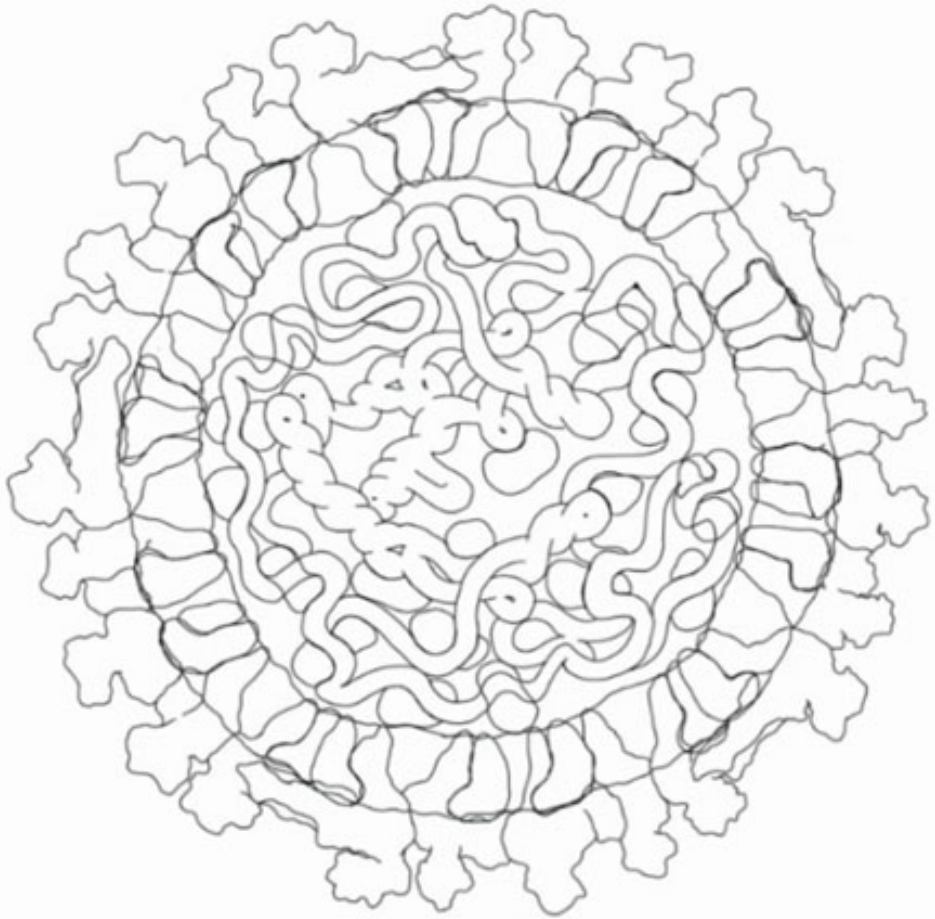






Membrane and Small Envelope Protein

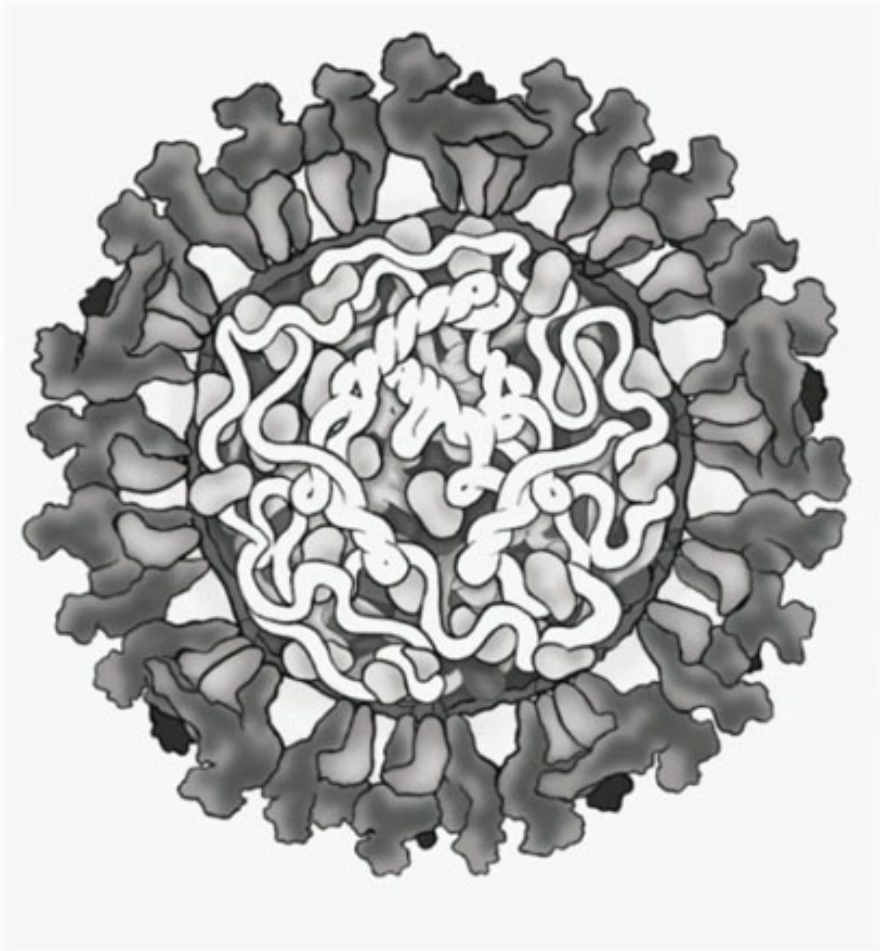


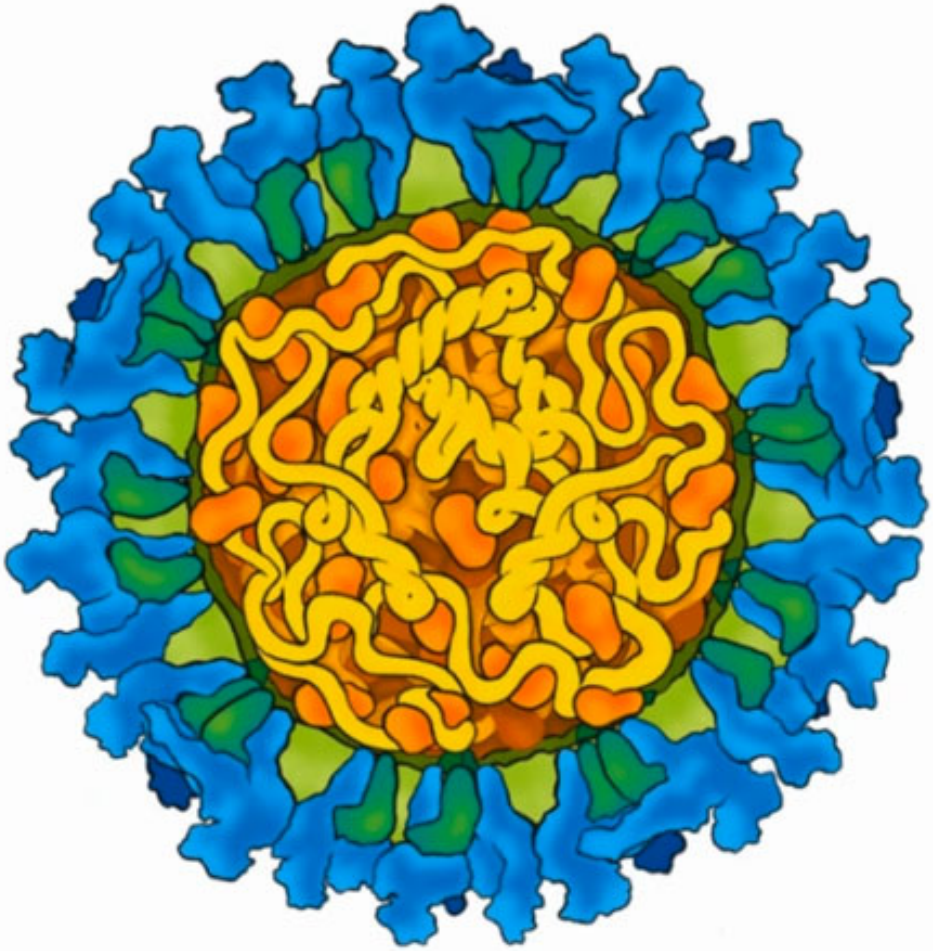


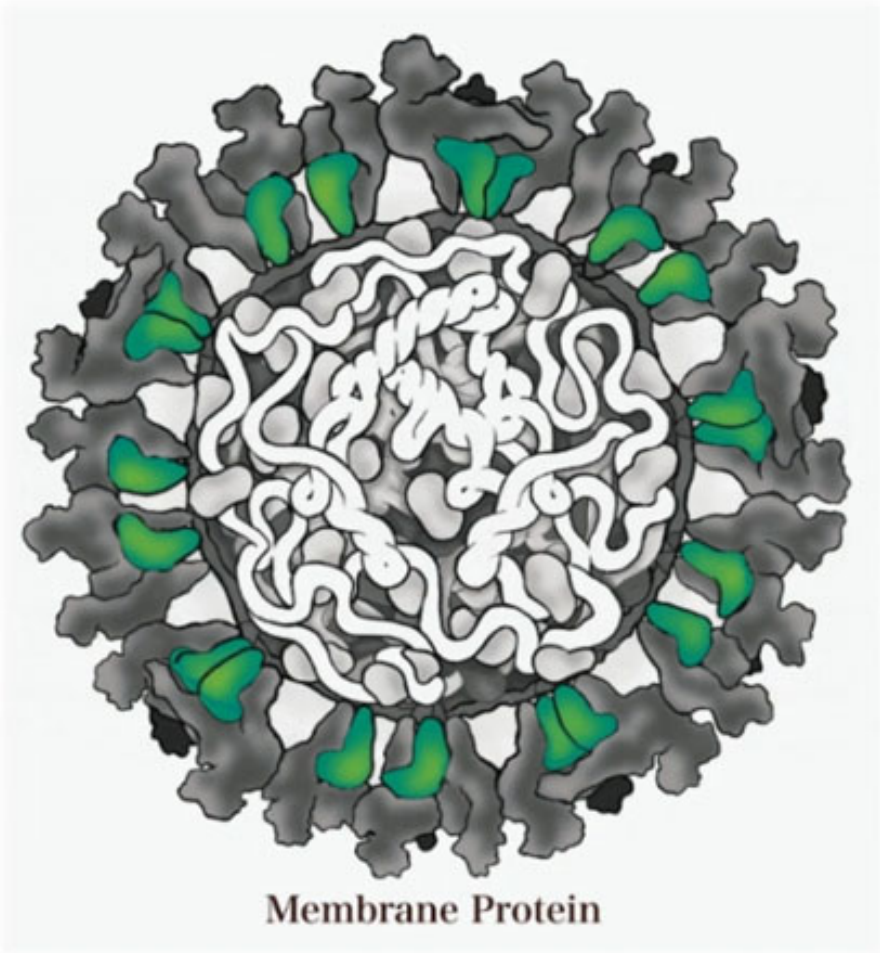
Dengue Virus

Status: Completed
No. of Iterations: 3
Time: 2 Weeks

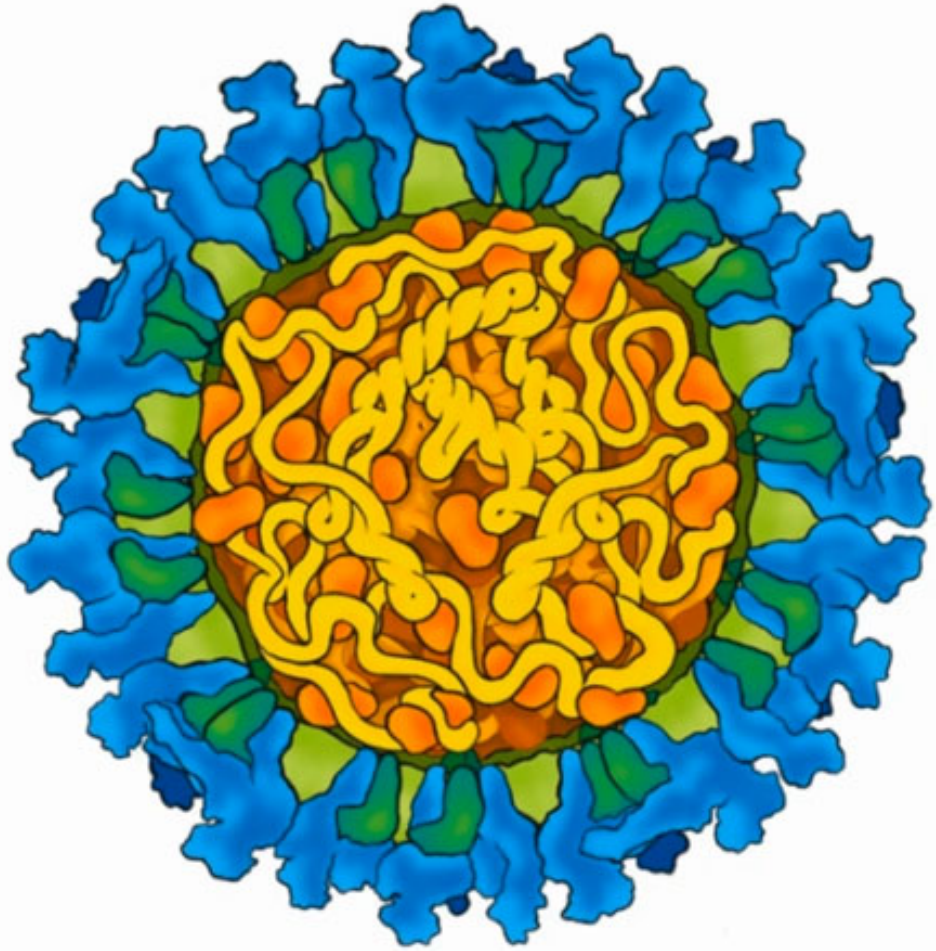
Dengue virus belongs to the genus Flavivirus which are relatively smaller viruses.
This is the simplest organism attempted in this project.

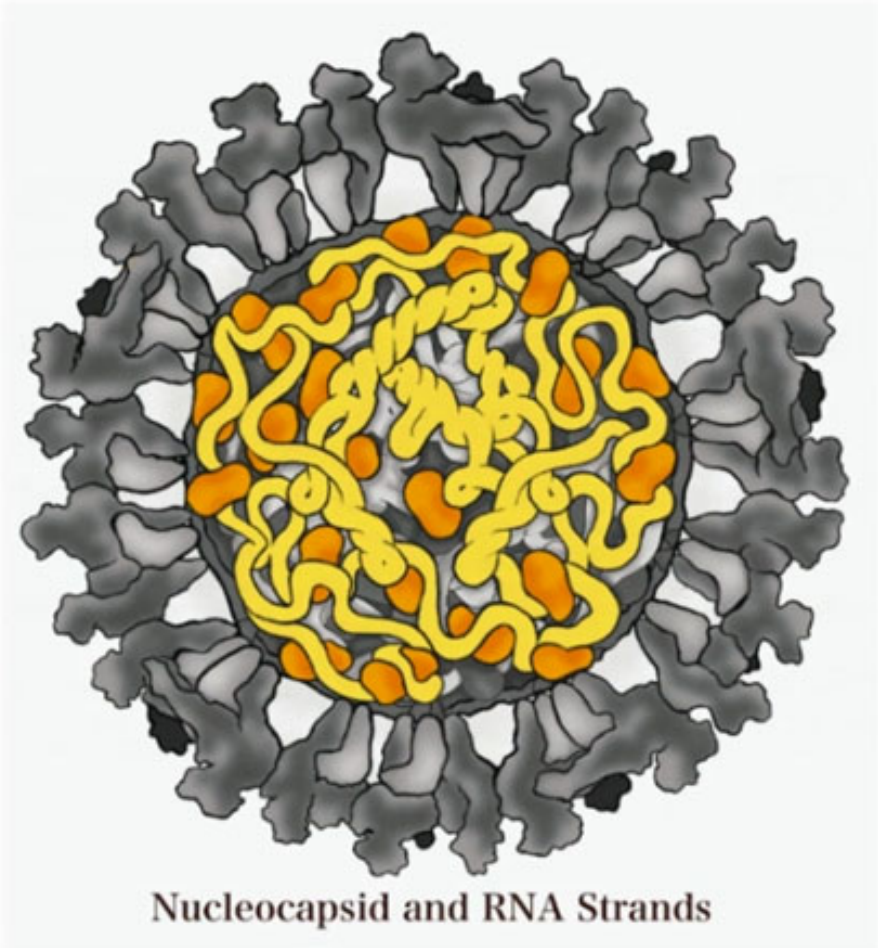




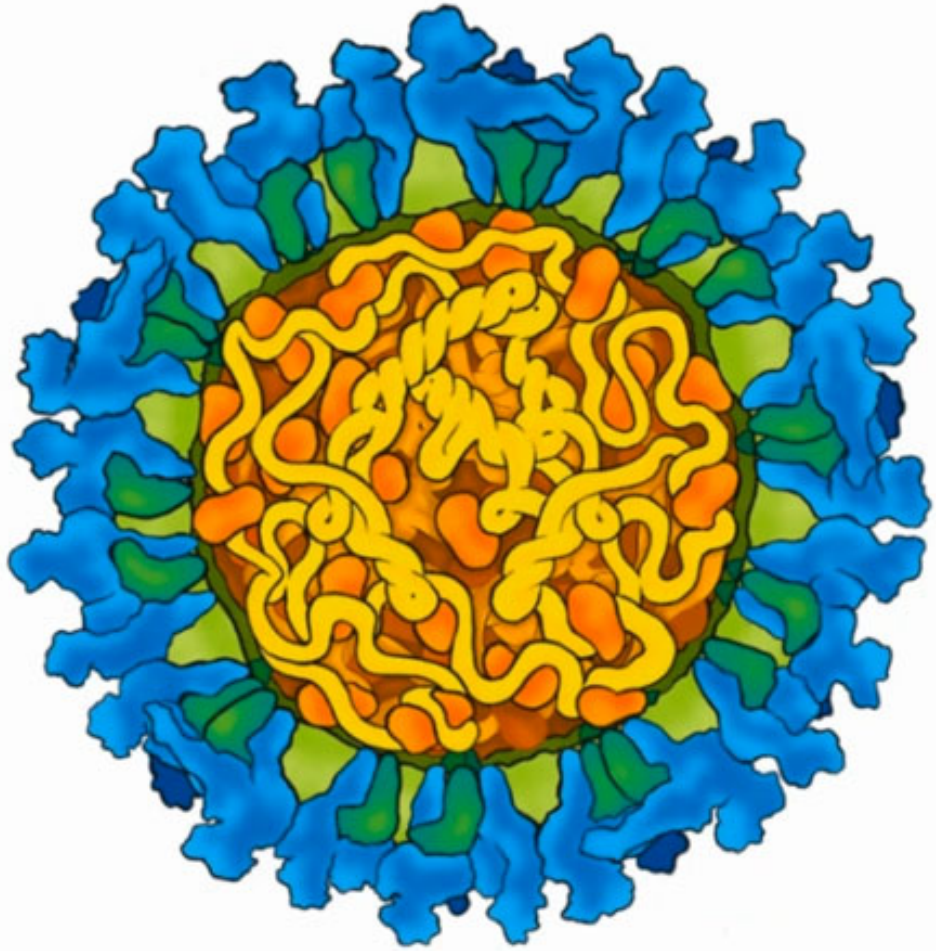


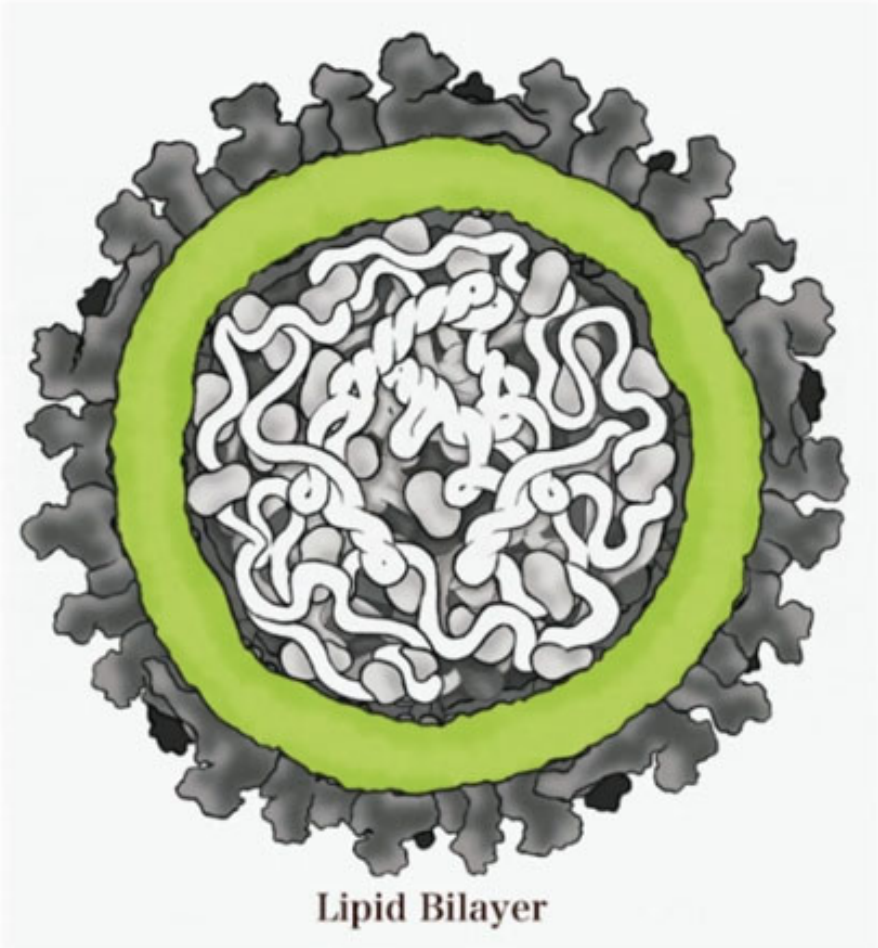
Membrane Protein

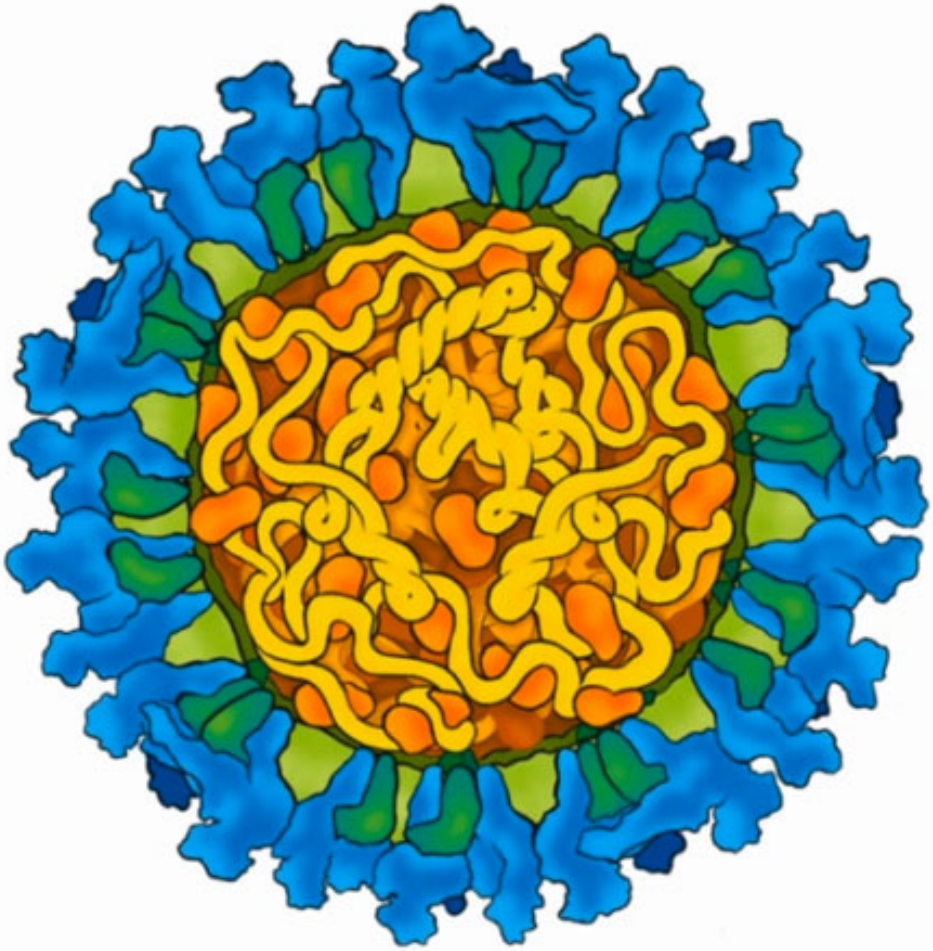


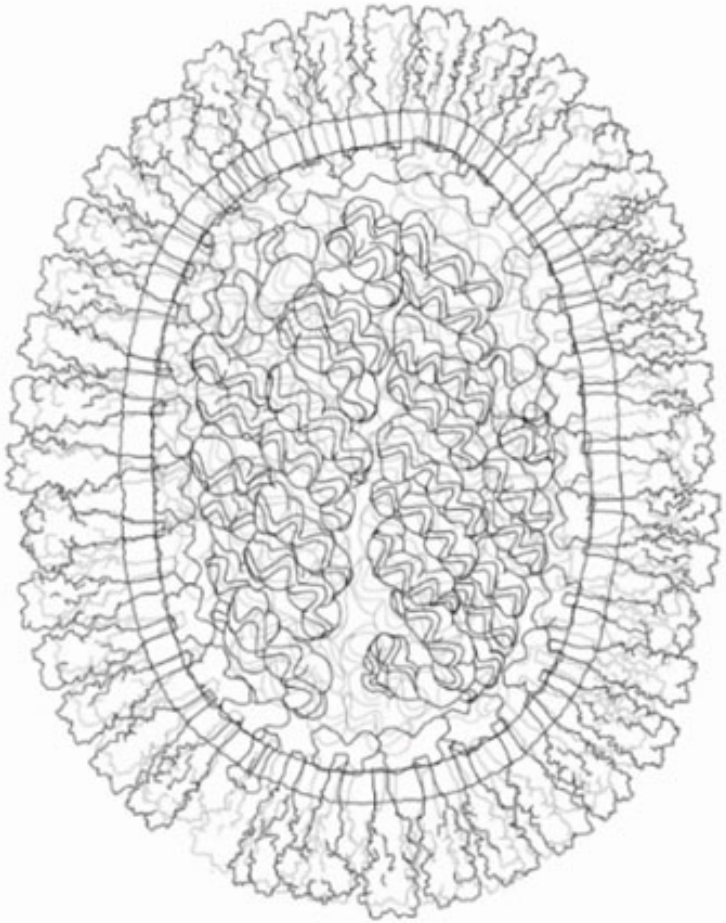


Nucleocapsid and RNA Strands




Lipid Bilayer

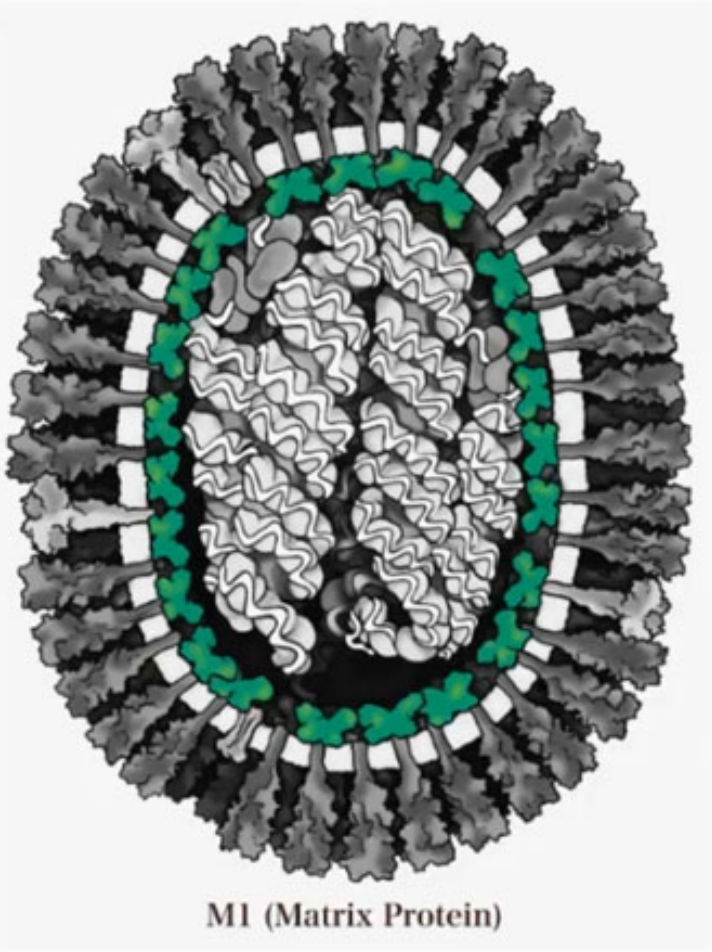




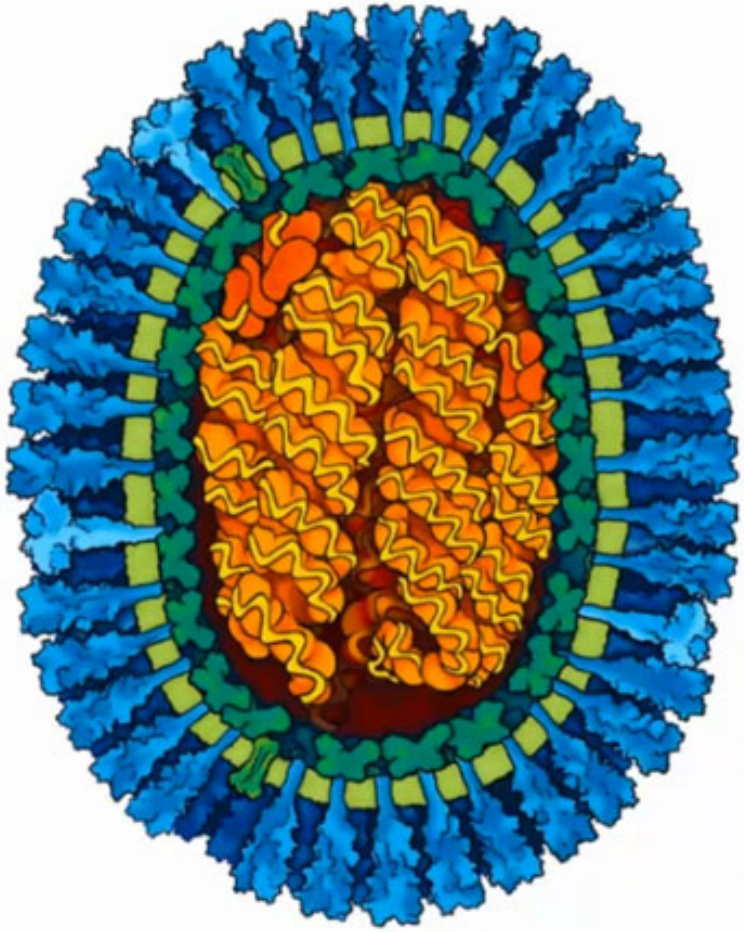
Influenza A (H1N1)

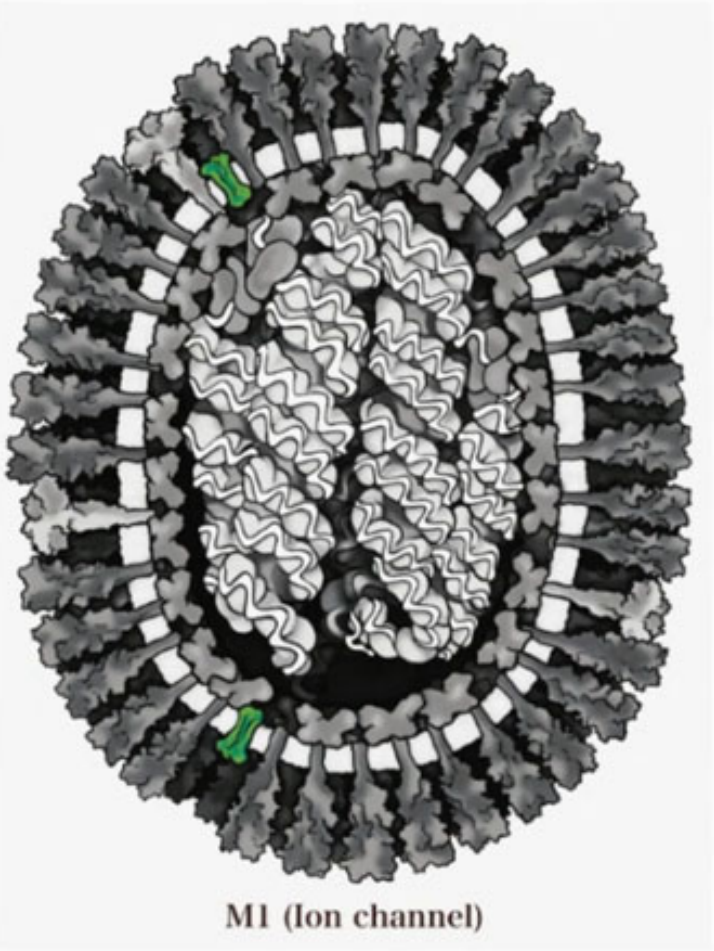
Status: Completed and Validated
No. of Iterations: 3
Time: 3 Weeks

Influenza A is the most complex of the 3 of the viruses attempted in this project. The most difficult part was to depict the crowdedness of the cell correctly.

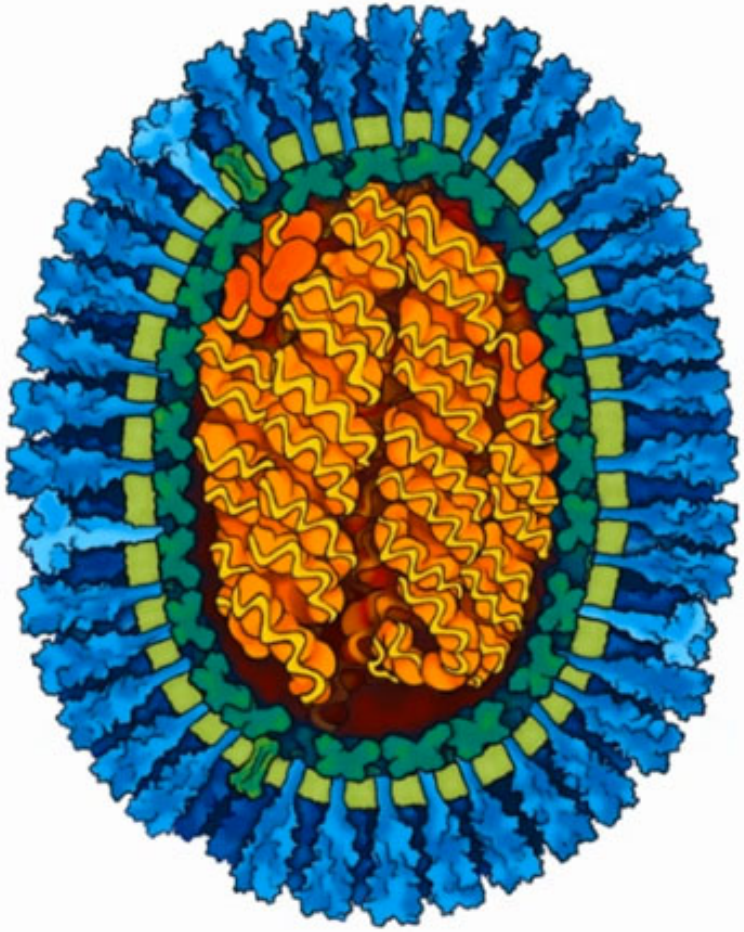


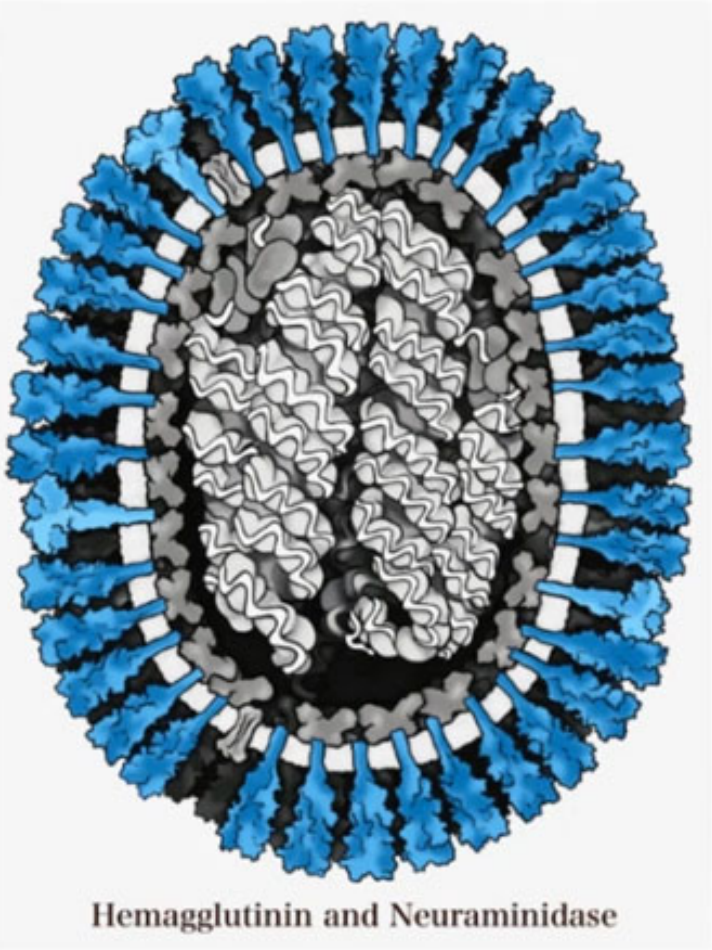
M1 (Matrix Protein)



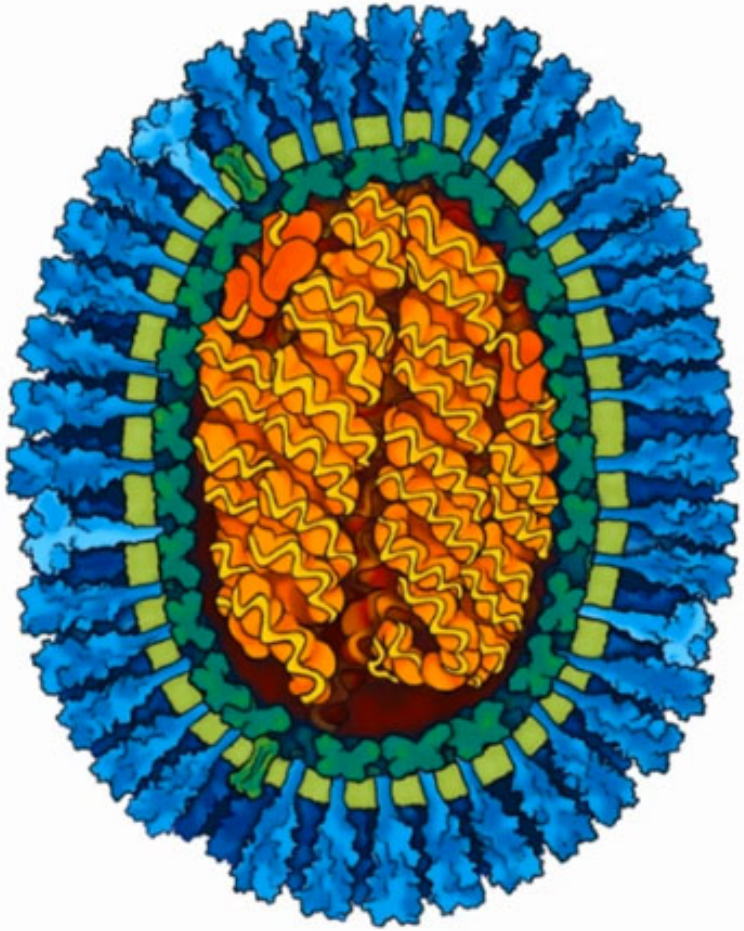


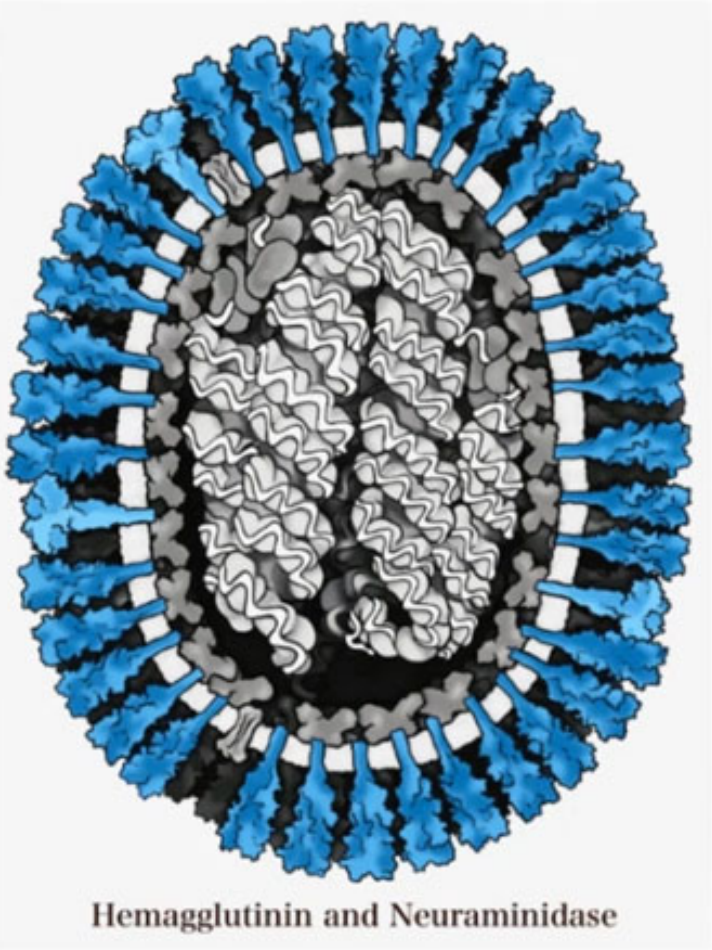
M1 (Ion channel)



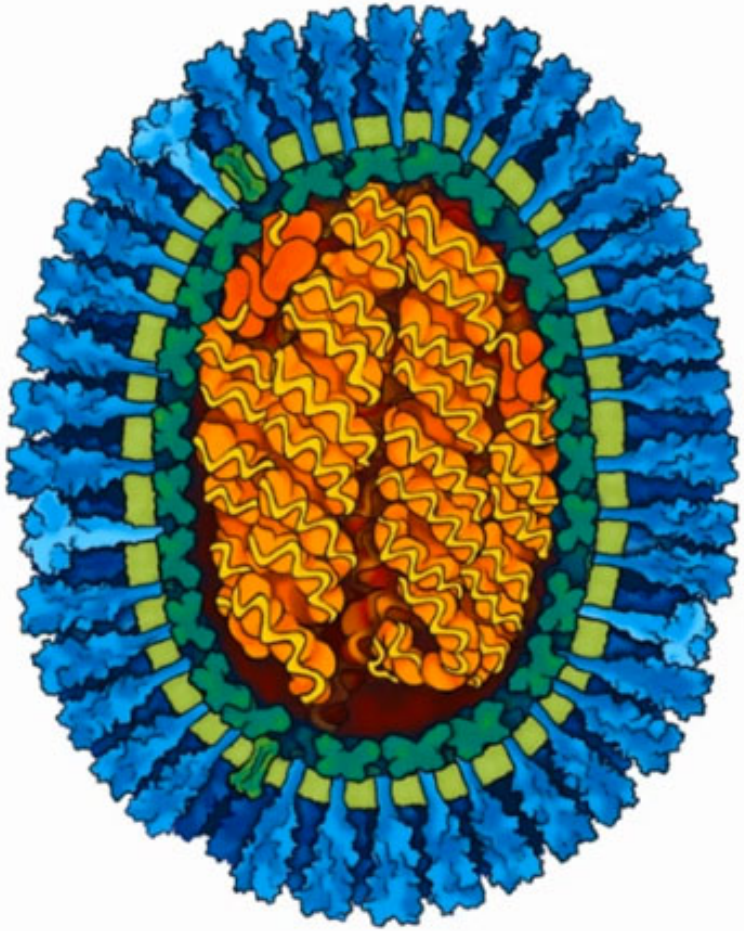


Hemagglutinin and Neuraminidase

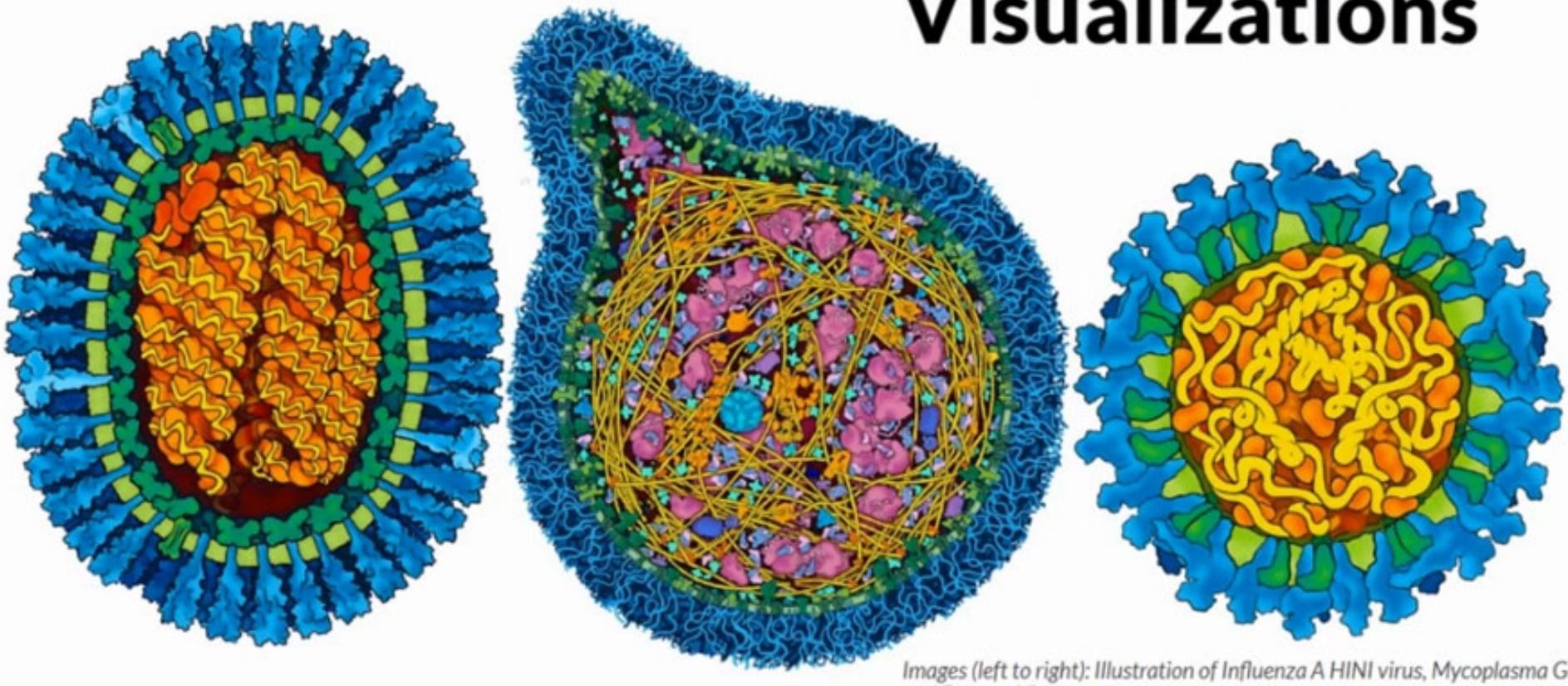




Hemagglutinin and Neuraminidase



Visualizations



Images (left to right): Illustration of Influenza A H1N1 virus, Mycoplasma Genitalium and Dengue Virus


Thank You!