Using Molecular Visualisation Techniques To Explain the Molecular Biology of SARS-CoV-2 Spike Protein Mutations to a General Audience

Sarah Iannucci, Dr William Harvey, Dr Joseph Hughes, Prof David Robertson, Dr Edward Hutchinson, Dr Matthieu Poyade

Abstract

Since the COVID-19 pandemic started in 2019, the virus responsible for the outbreak – SARS-CoV-2 – has continued to evolve. Mutations of the virus' spike protein, the main protein driving infectivity and transmissibility, are especially concerning as they may allow the virus to improve its infectivity, transmissibility, and ability to evade the immune system. Understanding how specific molecular changes can alter the behaviour of a virus is challenging for non-experts, but this information helps us to understand the pandemic we are living through and the public health measures and interventions needed to bring it under control. In response to communication challenges arising from the COVID-19 pandemic, we recently developed an online educational application to explain the molecular biology of SARS-CoV-2 spike protein mutations to the general public. We used visualisation techniques such as 3D modeling and animation, which have been shown to be highly effective teaching tools in molecular biology, allowing for the viewer to better understand protein structure, function, and dynamics. We also included interactive elements for users to learn actively by engaging with the digital content, and consequently improve information retention. This chapter presents the methodological and technological framework which we used to create this resource, the 'SARS-CoV-2 Spike Protein Mutation Explorer' (SSPME). It explains how molecular visualisation and 3D modeling software were used to develop accurate models of relevant proteins; how 3D animation software was used to accurately visualise the dynamic molecular processes of SARS-CoV-2 infection, transmission, and antibody evasion; and how game development software was used to compile the 3D models and animations into a comprehensive, informative interactive application on SARS-CoV-2 spike protein mutations.

This chapter indicates how cutting-edge visualisation techniques and technologies can be used to improve science communication about complex topics in molecular biology and infection biology to the general public, something that is critical to gaining control of the continuing COVID-19 pandemic.

Keywords

SARS-CoV-2, COVID-19, 3D modelling, 3D animation, interactive visualisation, molecular visualisation

1 Introduction

1.1 SARS-CoV-2 Spike Protein Mutations

Severe Acute Respiratory Syndrome Coronavirus 2 – SARS-CoV-2 – is the virus responsible for the COVID-19 pandemic, a world-changing disease outbreak which as of late January 2022 has caused over 5.5 million deaths (World Health Organization, 2021). Over the course of the pandemic, SARS-CoV-2 has mutated, resulting in changing viral properties. One of the four viral proteins found inside virus particles is the spike protein which decorates the outer viral particle like a crown or a solar corona (it is these distinctive spikes that give coronaviruses their name) (Figure 1.1). The spike (S) proteins assemble as trimers, with each monomer divided into S1 and S2 subunits. In its trimeric form, there is a distinct 'head' region of the protein sitting atop a 'stalk' which embeds within the viral membrane (Casalino, et al., 2020).. The S1 subunit of S contains two important regions of interest (ROI), which are determinants of infectivity and antigenicity (Casalino, et al., 2020).

The first of these ROIs is the receptor-binding domain (RBD), located at the tip of the spike protein (Figure 1.2). The RBD facilitates binding to Angiotensin-converting enzyme 2 (ACE-2), a protein found on the surface of the host cell (Peacock, et al., 2021; Letko, et al., 2020). The RBD can change its conformation to make the spike protein "open" or "closed", regulating it's binding capabilities (Ke, et al., 2020; Benton, et al., 2020). The second ROI is the N-terminal domain (NTD), located towards the upper sides of the spike protein (Figure 1.2). The NTD is one of the targets for the host's immune system (Casalino, et al., 2020). Additionally, The spike protein is chemically attached to sugar molecules called glycans, which are important for protein folding and may assist in 'shielding' the spike protein from searching host antibodies (Figure 1.2) (Rossen, et al., 1998; Xiong, et al., 2018).

Mutations of the spike protein can affect its properties and they are seen as particularly concerning if they could affect the key viral properties of infectivity and antigenicity, which could give strains of SARS-CoV-2 the potential to be more harmful to the human population (Letko, et al., 2020). As mutations accumulate in SARS-CoV-2, viral genomic surveillance can identify concerning viral variants and monitor their evolution (Harvey, et al., 2021). Variants with the potential to increase the impact of SARS-CoV-2 –

termed Variants of Concern (VOCs) – may require immediate action by public health authorities (Centers for Disease Control and Prevention, 2021).

Many groups around the globe are leading the charge on sequencing the SARS-CoV-2 genome and tracking these mutations. The COVID-19 Genomics UK Consortium (COG-UK) is one such organisation pioneering this genomic sequencing, helping to facilitate the understanding of SARS-CoV-2 transmission and evolution (COVID-19 Genomics UK Consortium, 2021). To share their analyses, COG-UK created the COG-UK Mutation Explorer (COG-UK/ME), an online interface allowing for the public to access SARS-CoV-2 mutation and variant data (Wright, et al., 2021). Given this resource's potential to impact the public's understanding of SARS-CoV-2 and COVID-19 public health regulations, this tool ideally should be accessible for the general public. However, when we began this project, the website was mainly used to present large amounts of detailed information (as data tables or protein structures) without explanatory context for non-specialists. Therefore, we decided that visualisation techniques and technologies may assist in making the data contained within the COG-UK/ME more accessible.

1.2 Molecular Visualisation in Education

Molecular biology is a highly complex subject to teach students, and even more challenging for a general audience without a scientific background to understand (Jenkinson, 2018). However, incorporating emerging technologies and visualisation techniques can be effective for facilitating the understanding of complex scientific topics (Jenkinson, 2018). Introducing interactivity into educational tools may lead to more active learning in users (Quintana, et al., 2004; Schrand, 2008). Active learning allows the user to be directly engaged with their learning experience, putting the user at the wheel of their own understanding. In addition to interactivity, visual resources are being used more regularly in science education. Static 2D and 3D illustrations, as well as animations, can aid the learner's understanding of complex scientific topics (Jenkinson, 2018). The use of three-dimensional (3D) graphics is particularly important in teaching molecular biology, as presenting molecular structures in 3D presents the user with depth cues which can be very important for understanding protein folding and binding sites, something which cannot be as easily achieved in 2D depictions (Höffler, 2010). In addition to explaining protein structure, 3D modelling of proteins can enhance the understanding of protein function. For example, learners may better understand how two proteins fit together when binding by visually

understanding the 3D structure of each protein's binding domains (Barak & Hussein-Farraj, 2013; Jenkinson, 2018).

One important feature of understanding molecular biology is the dynamic nature of proteins, as they are perpetually moving and interacting with structures in their environment; this can be difficult to grasp when using educational materials such as static images and textual descriptions (Robic, 2010; Queloz, et al., 2016; Jenkinson, 2018). Including 3D animations can help improve a learner's understanding of protein dynamics and key processes, better conveying the concept of molecular interactions as non-linear and highly fluid (Barak & Hussein-Farraj, 2013; Jenkinson, 2018). However, the creator of 3D molecular animations must also maintain a clear focus in the narrative and not depict an entirely chaotic molecular environment (Jenkinson, 2018). Considering the potential benefits of 3D molecular visualisation in science education, it is possible that interactive 3D models and animations of SARS-CoV-2 proteins and key viral processes may facilitate the understanding of SARS-CoV-2 mutations and VOCs for non-experts.

1.3 An Interactive Application for the COG-UK Mutation Explorer

This research set out to develop an online educational application about SARS-CoV-2 spike protein mutations in order to provide additional context for the COG-UK/ME and provide a visual aid for the key data presented in this resource. The application would incorporate interactive 3D models, 3D molecular animations, and information on SARS-CoV-2 spike protein mutations written in a style suitable for engaged but non-specialist audiences. The SARS-CoV-2 Spike Protein Mutation Explorer (SSPME) (https://sc2-application.itch.io/sars-cov-2-mutation-explorer) was developed as a collaboration between The Glasgow School of Art, The University of Glasgow, The MRC-University of Glasgow Centre for Virus Research, and COG-UK/ME.

2 Materials and Methods

2.1 Materials

A variety of software was used when designing and developing the SSPME (Table 2.1).

Table 2.1. The software used in the development of this research, including its purpose and developer.

Software	Purpose	Developer
Adobe After Effects	Digital visual effects, motion graphics, compositing	Adobe, Inc.
Adobe Illustrator	Vector graphics editor and design	Adobe, Inc.
Adobe Media Encoder	Video media transcoding	Adobe, Inc.
Adobe Photoshop	Raster graphics editor and design	Adobe, Inc.
Arnold Renderer	Advanced Monte Carlo ray tracing renderer	Solid Angle, Autodesk, Sony Pictures Imageworks
Autodesk Maya	3D modeling and animation	Autodesk, Inc.
Color Oracle	Colour blindness simulator	Bernie Jenny, Nathaniel Vaughn Kelso
Molecular Maya	Importing, modeling, rigging, and animating molecular structures	Digizyme, Inc.
PyMOL	Molecular visualisation	Shrödinger, Inc.
UCSF Chimera	Molecular visualisation	Resource for Biocomputing, Visualization, and Informatics (RBVI), UCSF
Unity	Application development engine	Unity Technologies
Visual Studio Code	Source-code editor	Microsoft

Additionally, existing visualisations and other resources were used for visual inspiration, information on protein structures, and as open-source assets for this development (Table 2.2).

Table 1.2. The resources used in the development of this research, including its use and reference.

Resource	Use	Reference
Annotation Lab: Coronavirus Animation	Inspiration on animation narrative and storytelling; visual inspiration on demonstrating viral properties	(Iwasa, et al., 2021)
A Visual Model for SARS-CoV-2 Membrane Fusion	Inspiration on animation narrative and storytelling; visual inspiration on demonstrating viral properties	(Khao, et al., 2020)
Abstract Texture	Texturing of the virus 3D model	(Paulo, 2021)
Click Sound	Application audio sounds	(Mixkit, n.d.)
Camera Snapshot Sound	Application audio sounds	(Soundjay, n.d.)
SARS-CoV-2 Paintings	Visual inspiration on colour scheme and composition; resource for details on relevant proteins to include in the animations	(Goodsell, 2020; Goodsell, 2020; Goodsell, et al., 2020)
Voiceover Generator	Generation of animation and application voiceover narratives	(Voicebooking, n.d.)

2.2 Concept and Workflow

The concept of this project was to act as a visual educational tool, enabling users to better understand SARS-CoV-2 spike protein mutations and to better digest the information presented in the COG-UK/ME. Figure 2.1 shows a basic overview of the developmental workflow. After solidifying the concept, a Focus Group (FG) study was conducted within the COG-UK/ME in order to determine areas for improvement or refinement to be addressed in the SSPME. Once the FG study was completed, we conducted research into the structural information then available about relevant features of SARS-CoV-2, particularly the spike protein and its interaction partners. Following this, 3D modeling of protein structures and

model rigging was performed using the Molecular Maya plug-in for Autodesk Maya. After the molecular models were rigged, the 3D molecular animations were created. Graphic design was used to create additional visual elements for the animations in order to facilitate the visual narrative. Once the 3D models and animations were completed, the interactive application was built in Unity and hosted freely online.

2.3 Co-Design Workshop

The FG consisted of 9 participants, 5 female and 4 male, who were recruited from social media. Qualitative feedback received from the FG identified several common areas for further development with the current COG-UK/ME resource:

1. Include relevant background information on SARS-CoV-2 viral structure, specifically on the spike protein structure and function.
2. Include information on how mutations of ROI on the spike protein may impact SARS-CoV-2 viral properties.
3. Make the information more geared towards a non-expert audience.
4. Consider accessibility, such as colour blindness, when designing the application.
5. Include visualisations that facilitate the user's understanding of concepts.

This feedback was considered in the design and development process of the SSPME to ensure this resource met the expectations of the general public.

2.4 SARS-CoV-2 Structural Details

In order to create scientifically accurate, realistic 3D models of SARS-CoV-2 proteins, structural details were researched and documented for reference. The amino acid (aa) residues of spike protein ROIs were identified for creating individual 3D models of each region for the interactive application model (Table 2.3).

Table 2.3. The amino acid residues of the various regions of interest on the spike protein used in this development.

Region of Interest	Position in S protein (residue numbers)	Information Source
Spike protein "head"	1-1146	(Woo, et al., 2020)
Spike protein "stalk"	1147-1273	(Woo, et al., 2020)

Receptor-binding domain (RBD)	318-541	(Woo, et al., 2020)
N-Terminal domain (NTD)	13-305	(Woo, et al., 2020)
Furin cleavage site (FCS)	681-688	(Xia, et al., 2020)

The RCSB Protein Data Bank (PDB) was used to obtain experimentally solved protein structures for use within the 3D modelling software (Berman, et al., 2000). The PDB identifications (IDs) were obtained for all proteins to be used in the animations, as these IDs are necessary for importing protein structures directly through Molecular Maya. The Coronavirus Annotation animation resource was used as a guide to many of the structures listed in Table 2.4 (Iwasa, et al., 2021).

*Table 2.4. The RCSB Protein Data Bank identifications and file names for all proteins needed in this development. * indicates a protein structure that is not full-length and which was modelled by the source before use.*

Protein	**PDB ID/PDB File Name**	**Source**
Angiotensin-Converting Enzyme 2 (ACE-2)	6M17 chains B, E	(Yan, et al., 2020)
Full-length glycosylated spike (S) protein, 1 RBD open	6VSB_1_1_1*	(Woo, et al., 2020)
Full-length glycosylated spike (S) protein, all RBDs closed	6VXX_1_1_1*	(Woo, et al., 2020)
Membrane (M) protein	M_dimer_new_1	(Heo & Feig, 2020)
Envelope (E) protein	E_protein_1	(Heo & Feig, 2020)

Lysosome-associated Membrane Protein 1 (LAMP-1)	5GV0	(Terasawa, et al., 2016)
V-ATPase	5VOX	(Zhao, et al., 2017)
Transient Receptor Potential Mucolipin 1 (TRPML1)	5WJ5	(Schmiege, et al., 2017)
Solution structure of human secretory Immunoglobulin A 1 (IgA1)	3CHN	(Bonner, et al., 2009)
Crystal structure of Pembrolizumab, a full-length Immunoglobulin G 4 (IgG4) antibody	5DK3	(Scapin & Strickland, 2015)

The structural dimensions of SARS-CoV-2 and its structural proteins were researched and summarised in Table 2.5.

Table 2.5. The structural details used in the development of the 3D models for SARS-CoV-2 and its structural proteins.

Structure Parameter	**Value**	**Source**
Diameter of viral membrane	91 ± 11 nanometers (nm)	(Ke, et al., 2020)
Length of spike protein ectodomain	22.5 nm	(Yao, et al., 2020)
Number of spike proteins per SARS-CoV-2 virus particle	~26 spike protein trimers	(Yao, et al., 2020)

Number of spike proteins with RBD open versus closed	'Majority' in closed conformation	(Ke, et al., 2020)
Spike protein flexibility	Spike protein 'head' freely rotates around 'stalk' with a lean of up to 90 degrees relative to viral envelope normal axis; however, a lean over 50 degrees is less favourable	(Ke, et al., 2020)
Distribution of spike proteins on SARS-CoV-2 membrane	Random distribution (non-uniform); spike proteins have no apparent relationship to each other	(Ke, et al., 2020)

These structural details were referenced throughout the development of the SSPME, ensuring that the scaling and organisation of the protein 3D models were scientifically accurate.

2.4 3D Modeling

Two different sets of 3D models were required in the development of the SSPME: a lower-resolution 3D model of the spike protein with individual models of ROI for interactive use in the application and higher-resolution 3D models of the spike proteins and other relevant proteins for the 3D molecular animations. In the creation of all 3D models in this research, the colour blindness simulator Color Oracle was used to test and ensure all 3D model colours were accessible (Jenny & Kelso, 2018).

2.4.1 Interactive 3D Models

Using both the open and closed conformations of the spike protein (6VSB_1_1_1, 6VXX_1_1_1; see Table 2.4), selections were made in PyMOL of the aa residues correlating to the ROI identified in Table 2.3. Additionally, the glycosylation profile was selected using the PyMOL Generate function (Figure 2.2). These were all exported as PDB files for later importing into Autodesk Maya using Molecular Maya. In addition, the AA residues of the spike protein excluding the ROI were generated as PDB file in order to use as a base spike protein 3D model (AA residues 1-12, 306-317, 542-680, 689-1273). Due to issues importing the glycosylation PDB file in Molecular Maya, it was loaded into UCSF Chimera and, using

the Multiscale Model function, an OBJ file was generated and imported into the Autodesk Maya scene manually aligned to the spike protein model. The resulting 3D models were exported as FBX files and loaded into Unity for interactive application development. Additionally, a curved 3D plane was created to represent a portion of the viral membrane into which the spike protein could be inserted into (Figure 2.3).

2.4.2 Animation 3D Models

The proteins listed in Table 2.4 were imported into Autodesk Maya using Molecular Maya's Import PDB functionality at the lowest resolution preset settings possible. The majority of proteins had a geometry of 25,000 polygon (poly) faces or lower, however the glycans, V-ATPase, and IgA1 had higher poly counts due to more complex structures which might not show the level of detail required to communicate relevant concepts.

When defining the animation contents and narrative, it was decided that a back-drop of mucins would be necessary to depict the respiratory tract (these molecules are commonly found in this environment) where SARS-CoV-2 enacts its viral infection processes. Therefore, a digital painting of mucins was created for use as a backdrop to all animations; this painting was inspired by the portrayal of mucins by David Goodsell in his painting "Coronavirus" (Goodsell, 2020) (Figure 2.4). It was also decided that a full SARS-CoV-2 virus 3D model would be necessary in order to provide the viewer with an understanding of the full virus structure, in addition to individual 3D models of proteins for close-up scenes.

2.4.2.1 Full SARS-CoV-2 Virion

The structural dimensions detailed in Table 2.5 were used to scale the spike proteins appropriately to the virus particle, which was created as a basic sphere object. A 4:1 ratio of spike protein ectodomain length to SARS-CoV-2 outer membrane diameter was used. Since the majority of spike proteins should be in the closed conformation, with an average of 26 spikes per virus particle, 26 spike models were created with 20 in closed and 6 in open conformations. These were randomly dispersed by hand across the viral membrane at random angles generally not exceeding 50 degrees relative to the normal membrane axis, as described by Ke, et al., 2020 (Figure 2.5). Basic 3D model materials were used in colouring the model, and a texture was then applied to the viral membrane so that it resembled a lipid bilayer (Paulo, 2021).

2.4.2.2 Individual Protein 3D Models

All proteins were imported as described in Section 2.4.2 and the scene was set up with one spike in focus and two flanking spikes in the background, all sitting in the viral membrane decorated with E and M proteins (Figure 2.6). A human cell surface model was also created with associated cell surface receptor proteins (ACE-2, LAMP-1, V-ATPase, TRPML1) randomly distributed by hand around the membrane (Figure 2.7). Finally, 3D models of relevant host antibodies of isotypes that might be found in respiratory mucus (IgA1, IgG4) were arranged floating intermittently throughout the scene (Figure 2.8).

2.5 Animation

It was decided that two animations could be accomplished within project timelines, one showing the spike protein's RBD and the other the spike protein's NTD. Each animation would begin with a basic introduction to SARS-CoV-2, containing the full virus particle. Then, the animation would focus on one spike, demonstrating key concepts about spike functions that could be affected by spike protein mutations.

2.5.1 Molecular Rigging

In order to accurately convey protein dynamics within the animations, molecular 'rigs' were created using Molecular Maya's Rigging Kit. These act as skeletons underneath the molecular mesh, simulating key molecular dynamics (MD) throughout the animations. Rigs were created for each of the three protein chains in the spike, for both open and closed conformations. In order to prevent protein unfolding, Selections and Elastic Networks (ENs) were created for the alpha carbons (CA) of each full chain (AA 1-1273). After testing, EN strength values of 10 were used as they kept the proteins dynamic but bound together. A Field Magnitude (FM) of 100 was used as turbulence in MD, adding more dynamic movement to the entire structure.

A rig specific to the spike protein 'head' was needed to convey the flexibility of this protein region on the 'stalk'. Selections and ENs were created for each protein chain's head, and additional ENs were created between heads of each of the three proteins in the trimeric spike to prevent them from disconnecting from one another. Then, a Target was created for each protein chain head, with a Handle target and strength value of 10. These Handles could be moved in the 3D scene and animated; the protein chain heads would then following this motion, creating the desired spike protein flexibility (Figure 2.9).

Additionally, a rig specific to the spike protein RBD was needed to simulate the movement from open to closed conformations. Since protein chain A in the 6VSB_1_1_1 PDB file was open, a rig was created for this specific protein chain. RBD-specific Selections and ENs were created. The Target was created as a Molecular Maya Chain, which was selected as the protein chain A in the 6VXX_1_1_1 PDB file, as this RBD is in closed conformation. Another Target was created as its own protein chain A. This would allow for the RBD to be animated as going back-and-forth between open and closed conformational states.

Finally, a basic rig was created for the ACE-2 PDB file in order to give this structure realistic MD.

2.5.2 Full SARS-CoV-2 Virion Animation

Using the translation and rotation tools only, the full virus models created in Section 2.4.2.1 were animated to appear as if they are floating in a fluid environment. These tools were animated to different positions and at different rotations in order to convey a sense of randomness to the movement and remove any potentially uniformity in the virus movements (Figure 2.10).

2.5.3 Individual Protein Animations

Considering the communication aims of the SSPME, there were several important viral processes to animate in the close-up animations of the SARS-CoV-2 spike protein:

1. Spike protein ‘head’ flexing back-and-forth on the protein ‘stalk’
2. Spike protein RBD changing between open and closed conformations
3. Spike protein RBD binding to ACE-2 on the host cell surface
4. Antibodies moving around the spike protein but not identifying it

First, the spike protein ‘head’ should demonstrate its flexibility by moving back-and-forth on the protein stalk. Using the translation tool, the Target Handle developed in Section 2.5.1 was animated to depict this movement. Next, the glycans on the spike protein should exhibit their characteristic wiggling movement (Grant, et al., 2020), as they are not stationary structures just sitting on the viral surface. In order to achieve this, the glycan models were duplicated, then manually stretched and moved on the spike model. Next, a Blend Deformer was used – this function connects the original mesh to the duplicated, deformed mesh. Fall-Off objects could then be used to alternate between the original and duplicated mesh. These

objects were animated back-and-forth, so that the glycan models deformed rapidly between their original, normal structure and position to the warped one. This created the wiggling illusion (Figure 2.11).

The next key process to animate was the spike's RBD changing from open to closed conformation. The strength value of the Target Selection described in Section 2.5.1 was animated between 0 to 10 for one protein chain, switching between open and closed conformations of that protein chain.

Next, in order to demonstrate how the spike protein's RBD binds to ACE-2 on the host cell, the ACE-2 model, along with the cell membrane and associated cell-surface receptors (Figure 2.7) were animated to move towards the spike protein. The background plane decorated with mucins was also animated the same distance at the same time, creating the illusion of the spike protein moving towards the ACE-2 receptor and binding (Figure 2.12).

Finally, for showing the host antibodies floating around the virus and not locating the spike protein, the antibody models were animated using the translation and rotation tools to create random movement.

For all animations, dim background lighting with a foggy environment was added to give visual appeal to the scene, allowing for a more mysterious cellular environment to be created. One additional light was used to highlight the main spike in the centre of the composition. A single camera was set up in front of the main spike protein to maintain the focus of this specific protein.

2.5.4 Animation Post-Processing

The post-processing of the animations was performed in Adobe After Effects, with added 2D graphical elements, voiceover narrative, and subtitles to facilitate and explain the visual narrative.

2.6 Interactive Application Development

The application was developed in Unity (2019.4.10f1) with an application structure outlined in Figure 2.13. The 3D models created in Section 2.4.1 were imported into Unity and the membrane model was given an image texture to give the appearance of a lipid bilayer (Paulo, 2021). Using PyMOL (as described in Section 2.4.1), images of the spike with different ROIs highlighted were created and added into the application in the sections

detailing ROI-specific information about function and the effects of mutations (Figure 2.14). Images highlighting specific mutations in the spike head domain that are characteristic of current VOCs - the Alpha, Beta, Gamma, Delta and Omicron variants - were obtained from the COG-UK/ME and implemented in the VOC-information section of the application (Figure 2.15). The Omicron VOC emerged after the initial release of the application, and due to its very rapid spread details about Omicron were rapidly added to the application to ensure that it remained relevant.

Using Unity and Visual Studio Code, a variety of functionality was developed to allow the user to explore the application (Table 2.6). Unity Documentation was referenced for all other basic application functioning (Unity Technologies, 2021).

Table 2.6. The purpose and sources of all scripts used in the development of the interactive application.

Script	Purpose	Source
Animation Start	Allows both animations to load on app start-up in Web-GL build	Custom Script
Camera Manager	Allows the camera to zoom and rotate around the 3D spike protein model	Custom Script
Mouse Manager	Allows the users mouse to interact with ROI on the spike protein; interactive ROI highlight when hovered over; if clicked, the scene changes to that ROI's information scene	(Khan, 2020; Katus, 2015)
Quit Active	Allows all pop-up panels to block out background mouse events, if opened, until the pop-up panel is closed	Brian Loranger
Screenshot Handler	Allows the screenshot functionality to work; when pressed, a screenshot of the app scene is automatically saved to the user's Downloads folder	(Unity Forum, 2016)

UI Manager	Controls all UI button functionality within the app	(Ouweland, 2017)

All UI elements were designed in Adobe Illustrator with reference to the colours presented in the COG-UK/ME in order to be visually cohesive with the website. Audio clicks were added to all application buttons in order to provide the user with audio feedback when using the application.

Finally, the text describing relevant scientific concepts was assembled (as discussed in 1 Introduction, above). Key statements were checked against information published by the CDC and the COG-UK/ME (Centers for Disease Control and Prevention, 2021; Wright, et al., 2021). Additionally, virologists with experience in public engagement and communication (Dr Ed Hutchinson and Dr William Harvey) checked the content for accuracy and suitability for the intended non-expert audience. Key terminology was provided in a glossary panel which could be accessed throughout the application.

3 Developmental Results

3.1 Animations

Two 3D animations were developed in this research, one focusing on the SARS-CoV-2 spike protein's RBD and the other on the NTD. Both were rendered at the lower-resolution size of 640 x 480 pixels due to time and equipment restraints. Both animations begin with an introductory scene depicting several full SARS-CoV-2 virions floating in the respiratory mucus (Figure 3.1). The purpose of this first scene is to introduce the viewer to the virus particle and its general structure. For both animations, this is followed by an animation showing two key properties of the spike:

- The spike protein exhibiting it's flexible 'head' and stationary 'stalk' (Figure 3.2a)
- The glycans covering the spike protein (Figure 3.2b)

Following this, the RBD and NTD animations diverge in narrative but use many of the same rendered stills to go into further detail on viral properties relevant to the ROI. The RBD animation visualises the following three processes:

- The RBD moving from open to closed conformation (Figure 3.3a)
- The RBD binding to the ACE-2 cell surface receptor (Figure 3.3b)
- The RBD moving to closed conformation to avoid antibody detection (Figure 3.3c)

The NTD animation instead visualises the following process, using the same frames as the animation described in Figure 3.3c but with a slightly altered narrative and different 2D graphics:

- The spike protein avoiding antibody detection (Figure 3.4)

In all animations, overlaid 2D graphics, narration, and subtitles were used to facilitate a more accessible and comprehensible story for the viewer to follow.

3.2 SARS-CoV-2 Spike Protein Mutation Explorer

The SSPME was embedded online in a Web-GL format in order to be accessible to users anywhere in the world with internet access. The design of the application interface was directly inspired by the COG-UK/ME website at the time of development; the SSPME was therefore created with simplistic, clean scenes in a similar colour scheme to that of the website. The SSPME application flow follows Figure 2.13, beginning with an Opening Scene (OS) to present the SSPME's purpose to the user and provide appropriate credits where applicable (Figure 3.5). From the OS, the user can move to the Instructions Scene (IS) to understand better how to use their mouse buttons to navigate the application (Figure 3.6). From the IS, the user enters the Main Scene (MS) containing the interactive 3D spike protein model (Figure 3.7). Here, the user can zoom in on and rotate around the 3D model, highlight the ROI on the spike, change the RBD conformation, take a snapshot, and access the VOC and glossary panels (Figure 3.8).

It is from the MS that users can go more in-depth with the four interactive ROI on the spike protein: the RBD, NTD, FCS, and glycan profile. Each of these ROI Scenes (RS) provides detailed written information about the ROI and the potential impact of mutations in this region. To improve accessibility for all users, options were given to display the text in 'big' and 'small' (default) options, and an auditory voiceover function was available for the text (Walker, 2014) (Figure 3.9). From the RBD and NTD RS, the user can access the 3D animations mentioned in Section 3.1.

4 Discussion and Conclusion

4.1 Critical Review of Methodology

There were many challenges confronted during the development of the 3D animations and interactive application in this research. When developing the animations, In order to fully comprehend the functionality and potential of mMaya, the online course *Introduction to*

Molecular Maya's Rigging Kit (McGill, 2018) was completed in order to fully comprehend the functionality and potential of mMaya. This facilitated the processes involved in creating 3D models of the many proteins used within this research, as well as rigging and animating these 3D models. Additionally, due to the complex nature of the multiple proteins needed in the animations, the 3D scenes in Maya were highly costly in terms of polygon count, which was not optimal for the equipment used in this research. This negatively impacted the animation development process as mMaya calculates molecular dynamics in real-time, causing significant delays when any animation process required review. However, this was deemed necessary in order to preserve the realistic appearance of the proteins. For the SSPME, several scripting issues were encountered early on when creating the interactive camera and screenshot feature within the application. Additionally, the VOC menu was added after the main application was already designed, therefore it was not considered in the design plan and may have negatively impacted the flow of this section. Ideally, this menu would be adapted in the future to be more interactive and in better harmony with the interactive 3D spike model. Overall, the SSPME is functional and works on popular web browsers such as Google Chrome, although preliminary internal testing showed that the animations will not play correctly in Safari.

4.2 Limitations & Future Research

There were several limitations faced during this research. Firstly, this research was conducted during a Master's dissertation and the research was adapting in real time to communicating about a pandemic. Therefore, the restrictive timeline limited the objectives of the animations and application. This prohibited the creation of a third and fourth full animation for the FCS and glycosylation profile of the spike protein. This also did not allow for the VOC menu of the application to be reimagined and enhanced. In addition to time constraints, the pace of development was limited by the equipment available. For example, the 3D molecular modeling and animation processes were very slow due to relatively low computer processing speeds on the Macbook Pro (2.2 GHz Quad-Core Intel Core i7, 16 GB 1600 MHz DDR3, Intel Iris Pro 1536 MB) which was available for use during remote work. In order to counteract this, low-resolution settings were used where possible.

Despite these limitations, this project demonstrates the potential for using three-dimensionality and interactivity in scientific communication and education. The SSPME successfully incorporates a SARS-CoV-2 spike protein 3D model into an interactive

application in order to educate users on its various ROIs and about how mutations of these ROIs may impact viral properties, bringing about the VOCs that have had major impacts on the way we have had to live our lives over the course of the pandemic. One example of this is the newly-emerging Omicron variant, which has a high number of mutations in the spike protein. We hope that the SSPME may make information about new VOCs presented in the media, as well as the data contained within the COG-UK/ME, more understandable to a general audience without a molecular virology background. This may help to boost the public's understanding of COVID-19, public health control measures, and the need for the restrictions that were imposed on daily life during the pandemic. Considering how the COVID-19 pandemic has progressed over the last two years and how misinformation surrounding infection and disease has spread globally, further research into tools that communicate the science of these topics are vital to combatting further negative development of the pandemic.

4.3 Conclusion

Though it has been two years since SARS-CoV-2 began its spread across the globe, it is showing no signs of stopping, and may continue to without the dedication and collaboration of us all. Promoting scientifically sound information to the general public can only help promote scientific communication and fight the spread of misinformation around SARS-CoV-2 and COVID-19. The SSPME (https://sc2-application.itch.io/sars-cov-2-mutation-explorer) developed in this research is hopefully one of many resources being developed for this very purpose. Between its first release in July 2021 and its current status in January 2022, the SSPME has been viewed over 1,700 times by users worldwide and a variety of positive and constructive feedback has been received from user-testing, press releases, and social media responses. We intend to continue to update the application as new VOCs emerge so that the SSME can be as accessible and far-reaching as possible.

Author Contributions

S.I.: Conceptualization, Visualisation, Methodology, Writing – Review and Editing; W.H.: Supervision, Visualisation; J.H.: Supervision; D.R.: Supervision; M.P.: Supervision, Resources, Writing - review and editing; E.H.: Conceptualization, Supervision, Writing - review and editing, Funding acquisition

Acknowledgements

The work described in this article originated as a project in The University of Glasgow / The Glasgow School of Art MSc in Medical Visualisation and Human Anatomy. We wish to thank Faye Watson and Lois Mason (MRC-University of Glasgow Centre for Virus Research) for support and advice. Work in the Hutchinson group is funded by the UKRI Medical Research Council [MR/N008618/1].

Figure Captions

Bibliography

Barak, M. & Hussein-Farraj, R., 2013. Integrating Model-Based Learning and Animationsfor Enhancing Students'Understandingof Proteins Structure and Function. *Research in Science Education,* 43(2), pp. 619-636.

Benton, D. J. et al., 2020. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature,* 588(7837), pp. 327-330.

Berman, H. M. et al., 2000. *RCSB Protein Data Bank.* [Online]
Available at: https://www.rcsb.org/
[Accessed 23 12 2021].

Casalino, L. et al., 2020. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science,* 6(10), pp. 1722-1734.

Centers for Disease Control and Prevention, 2021. *SARS-CoV-2 Variant Classifications and Definitions.* [Online]
Available at: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html
[Accessed 23 12 2021].

COVID-19 Genomics UK Consortium, 2021. *COVID-19 Genomics UK Consortium.* [Online]
Available at: https://www.cogconsortium.uk/
[Accessed 23 12 2021].

Goodsell, D., 2020. *Coronavirus.* [Art] (RCSB Protein Data Bank).

Goodsell, D., 2020. *RCSB PDB-101: Molecule of the Month.* [Online]
Available at: https://pdb101.rcsb.org/motm/246
[Accessed 23 12 2021].

Grant, O. C., Montgomery, D., Ito, K. & Woods, R. J., 2020. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports,* 10(1), pp. 1-11.

Höffler, T. N., 2010. Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review. *Educational Psychology Review,* 22(3), pp. 245-269.

Harvey, W. T. et al., 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology,* Volume 19, pp. 409-424.

https://www.cogconsortium.uk/, 2021. *COVID-19 Genomics UK Consortium.* [Online]
Available at: https://www.cogconsortium.uk/
[Accessed 23 12 2021].

Iwasa, J. et al., 2021. *Coronavirus Annotation.* [Online]
Available at: https://coronavirus-annotation-3.sci.utah.edu/
[Accessed 07 2021].

Jenkinson, J., 2018. Molecular Biology Meets the Learning Sciences: Visualizations in Education and Outreach. *Journal of Molecular Biology,* 430(21), pp. 4013-4027.

Jenkinson, J., 2018. Molecular Biology Meets the Learning Sciences: Visualizations in Education and Outreach. *Journal of Molecular Biology,* 430(21), pp. 4013-4027.

Jenny, B. & Kelso, N. V., 2018. *Color Oracle,* Melbourne, Australia: Monash University.

Katus, 2015. *Change Color on mouse over in Unity3D.* s.l.:Youtube.

Ke, Z. et al., 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature,* 588(7838), pp. 498-502.

Khan, M. F., 2020. *Object Glow Unity - Glow GameObject on Mouse Click C#.* s.l.:Youtube.

Letko, M., Marzi, A. & Munster, V., 2020. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology,* 5(4), pp. 562-569.

McGill, G., 2018. *Introduction to Molecular Maya Rigging Kit.* [Online]
Available at: https://clarafi.com/training/introduction-to-molecular-maya-rigging-kit/
[Accessed 21 05 2021].

Mixkit, n.d. *Free Click Sound Effects.* [Online]
Available at: https://mixkit.co/free-sound-effects/click/
[Accessed 2 7 2021].

Ouweland, L. v. d., 2017. *Using a Unity button to switch between scenes..* [Online]
Available at: https://www.loekvandenouweland.com/content/use-unity-button-to-change-between-scenes.html
[Accessed 2 7 2021].

Paulo, J., 2021. *Organic Abstract 001.* [Art] (https://3dtextures.me/).

Peacock, T. P., Penrice-Randal, R., Hiscox, J. A. & Barclay, W. S., 2021. SARS- CoV-2 one year on: evidence for ongoing viral adaptation. *Journal of General Virology,* 102(4), p. 001584.

Queloz, A. K. M., Stern, E., Hafen, E. & Köhler, K., 2016. Debunking Key and Lock Biology: Exploring the prevalence and persistence of students' misconceptions on the nature and flexibility of molecular interactions. *Matters Select,* 2(8), p. e201606000010.

Quintana, C. et al., 2004. A Scaffolding Design Framework for Software to Support Science Inquiry. *The Journal of the Learning Sciences,* 13(3), pp. 337-386.

Robic, S., 2010. Mathematics, Thermodynamics, and Modeling to Address Ten Common Misconceptions about Protein Structure, Folding, and Stability. *CBE-Life Sciences Education,* 9(3), pp. 189-195.

Rossen, J. et al., 1998. The Viral Spike Protein Is Not Involved in the Polarized Sorting of Coronaviruses in Epithelial Cells. *Journal of Virology,* 72(1), pp. 497-503.

Schrand, T., 2008. Tapping into Active Learning and Multiple Intelligences with Interactive Multimedia: A Low-Threshold Classroom Approach. *College Teaching,* 56(2), pp. 78-84.

Soundjay, n.d. *Camera Sound Effects.* [Online]
Available at: https://www.soundjay.com/camera-sound-effect.html
[Accessed 2 7 2021].

Unity Forum, 2016. *HOW TO: screenshot in a new tab of the browser (without pop-up blocking)..* [Online]
Available at: https://forum.unity.com/threads/how-to-screenshot-in-a-new-tab-of-the-browser-without-pop-up-blocking.440051/
[Accessed 2 7 2021].

Unity Technologies, 2021. *Unity Documentation.* [Online]
Available at: https://docs.unity3d.com/Manual/index.html
[Accessed 2 7 2021].

Voicebooking, n.d. *Free Voice Over Generator.* [Online]
Available at: https://www.voicebooking.com/en/free-voice-over-generator
[Accessed 15 7 2021].

Walker, M., 2014. *Top Tips for Accessible Apps.* [Online]
Available at: https://abilitynet.org.uk/news-blogs/top-tips-accessible-apps
[Accessed 12 07 2021].

Woo, H. et al., 2020. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *The Journal of Physical Chemistry B,* 124(33), pp. 7128-7137.

World Health Organization, 2021. *WHO Coronavirus (COVID-19) Dashboard.* [Online]
Available at: https://covid19.who.int/
[Accessed 23 12 2021].

Wright, D. W. et al., 2021. *COG-UK Mutation Explorer.* [Online]
Available at: https://sars2.cvr.gla.ac.uk/cog-uk/
[Accessed 23 12 2021].

Xiong, X. et al., 2018. Glycan Shield and Fusion Activation of a Deltacoronavirus Spike Glycoprotein Fine-Tuned for Enteric Infections. *Journal of Virology,* 92(4), pp. e01628-17.