

Making science tangible

Creating scientific animations from cryo-electron tomography
data of microbiological organisms

Vera C.J. Williams (s1390953)

Graduation thesis

Media Technology MSc program, Leiden University

July 2020

Thesis advisors: Prof.dr. A. Briegel and Prof.dr.ir. F.J. Verbeek

Introduction

The struggles of communicating scientific knowledge

Science fascinates many people, but to a lay person its subjects can easily become overwhelming due to its inherent complexities and abstractions. This specifically affects microbiological topics: here the investigated subjects are not visible to the naked eye and requires translational steps in order to transform data into a visualisation. While more accessible to the public, such visualisations can fall victim to oversimplifications that generate inaccurate mental models in the audience (Gibbons 2017). Another frequent undesired outcome is that visualisations may be accurate but fail to elicit interest in lay-observers.

Accurate scientific visualisations are of increasing importance within the scientific community as well. Oftentimes, researchers are highly specialised in their own research topic, but have only basic understanding of more distant research fields. In such cases, text alone often doesn't suffice to clearly describe the observations of complex microbiological systems (Iwasa 2016). Visualisations such as illustrations or animations greatly aid in making research results accessible to a wider audience. Scientists add these visualisations to their presentations at conferences to summarise and contextualise their research (Iwasa 2010). However, these animations are often more complex than those intended for the general public.

Despite their proven usefulness, the visualisations can be somewhat problematic within the scientific community. Burri (2013) found that for researchers, the effectiveness of visualisations is strongly dependent on the situation in which they are presented. Scientists regard visualisations as authoritative science in practical everyday use: MRI's for example, are described as "depicting reality 1:1" (Burri 2013). However, when the same scientist is asked to reflect on the visualisation's power, the images were described as constructed representations of reality (Burri 2013). This ambiguity originates from scepticism about the trustworthiness of the data manipulations when it is transformed into a visualisation (Burri 2013), which negatively effects the usefulness of visualisations within this community.

Lay-people demonstrate similar "seeing is believing" behaviour, as they have the tendency to accept visualisations as fact (Gibbons 2017). However, unlike scientists, it is more difficult for lay-people to reflect critically on a visualisation's power, due to the complexity of the underlying techniques. Few lay people are even aware of the abstractions and estimations that need to be made to get to these visualisations (Gibbons 2017). This becomes problematic when scientific images are taken from the realm of science and

are implemented in a popular context without reframing its original context. The resulting inaccurate mental models could negatively impact a persons' complete understanding of certain scientific subjects (Gibbons 2017).

Reframing scientific visualisations

It is therefore important to realise that visualisations cannot be experienced in a vacuum. Everyone is exposed to their specific cultural argument frames. These frames influence the expectations of a visualisation and even alter its focal point (Gibbons 2017). However, scientific visualisations exist within their own argument frames. The differences between these frames should be considered when communicating scientific findings from one frame to another.

A visualisation's framework is shaped by the research field from which it originates. Scientific visualisations are either directly or indirectly based on research results. These research results have to be interpreted in the context of the current understanding of the research topic. As such, visualisations based on research results should be interpreted in this context as well. Visualisations can then be regarded as *true* within its scientific context, but as powerful as these visualisations are, one should remain considerate of its relationship to the *truth*. Visualisations do not depict reality but rather represent an interpretation of the ground truth of a research topic.

It might therefore be argued that the underlying problem with scientific communication is more deeply rooted than expected. This highlights the importance of understanding the mindset of the observer with which they perceive a visualisation. The issue cannot be mended by solely adjusting the aesthetics or complexities of scientific visualisations. The observers need to understand that there are specific steps that lead to the end-product.

For scientists, transparency on this process could cement the visual power of visualisations more securely, and potentially increase their trust in its persuasive qualities. By providing transparency on the initial context of the observed subject, the intended goal of the visualisation, and the various levels of abstractions, a more appropriate frame in which to understand the visualisation's manipulations could be provided.

Additionally, to improve a lay-person's understanding of scientific visualisations, it might be beneficial to create visualisations from data that abstracts as little as possible from reality. In the biological sciences, cryo-

electron tomography (or cryo-ET) presents itself as an interesting candidate for this challenge due to this techniques qualities; cryo-ET produces high-resolution, 3D tomograms that allow us a glimpse of microbes in their near natural state.

In an attempt to recreate the frameworks in which visualisations can be understood on these two levels, this project strives to create scientifically credible visualisation from cryo-ET data and demonstrate a transparent process. To this end, cryo-electron tomography data will be transformed into an animation. As such, this project aims to answer the research question: *“How can we transform high-resolution cryo-electron tomography data into a widely accessible yet scientifically accurate visualisation?”*

Materials and methods

To answer the research question, this project designed a novel method for creating scientific visualisations. The answer to the posed “how” question is formulated as a “how-to” workflow. This workflow was designed through a case-study approach. The elements necessary to achieve this goal were known in principle, but needed to be converted into feasible steps. This was done in part through extensive brainstorm sessions on the types of software to be used, the different types of platforms to explore, the available sources to obtain the necessary practical skills, the variety of documentation methods, the range of product development and -testing methods, and the critical elements of the actual visualisation. These brainstorm sessions were often done with the help of experts from the fields of scientific visualisations, software technologies, hardware technologies, and education.

The abovementioned elements were investigated and discussed, and either explored and/or negated. Eventually, the animation format seemed the best fit for the data and the available/feasibly obtainable skillset. The animation process consisted of a trial and error approach since many of the skills necessary to create it were acquired through the process. Consequently, this project’s process is not at all streamlined; the goal of this project is to understand the visualisation process in all its individual steps so that a streamlined workflow can be described.

The overall goal of this project was to create scientific animations from cryo-electron tomography data for a specific audience. By reflecting on the process, distinguishing the critical moments, and determining the bottlenecks, the project aimed to design a *feasible, transparent, and reproducible* workflow for creating a scientific animation from cryo-ET data.

The key to designing the answer to the research question was to assess the methods' achievement in light of the two posed requirements in the research question: *accessibility* and *accuracy*. These parameters greatly affected the project as they influenced most important design choices. Accessibility was implemented by tailoring the animation to a specific audience. This quality was achieved by studying the target audience, their needs, background, available knowledge, and their specific use of jargon. Accuracy was maintained by evaluating the abstractions of all types of data used in the animation. This was achieved by selecting a fitting data type, and frequent feedback moments with experts on the story and the visual properties of the animation. These characteristics were balanced with the intended goal of the visualisation, with the aim of improving the reception of the final animation.

Background information

To better understand the complexities of this project, its relevant background information needs to be introduced. First, cryo-electron tomography is introduced to better grasp the data's properties and inherent abstractions. The choice for this data and its shortcomings will be discussed as well. Then, the scientific subject matters of the animations are described to provide a knowledge foundation for the final visualisations. Last, the target audiences are described in light of the accessibility features that shaped the design process.

A brief introduction to cryo-electron tomography

Through cryo-electron tomography we can observe microbial life at a very high resolution in its near-natural state. The samples are plunge frozen and imaged using an electron microscope (Oikonomou and Jensen 2016). This technique cools samples so quickly that water molecules stop rearranging before they can crystallise. As a result, the samples are embedded in a thin film of "vitreous ice" which preserves the native cellular structures of the microorganisms by preventing dehydration (Oikonomou and Jensen 2016).

The images are acquired with a transmission electron microscope (TEM). The TEM transmits a beam of electrons through the sample and an image is recorded. The sample is tilted in the microscope and its 3D information collected through the acquisition of a series of 2D projections. These projections have a regular orientation interval with respect to the electron beam. Typically, the samples are tilted in 2° increments from +/-60° (Fig. 1)¹. This "tilt series" is then used to computationally back-project in order to get the 3D image

1 https://en.wikipedia.org/wiki/Electron_cryotomography

(tomogram).

Cryo-electron tomography provides a strong basis for scientific visualisations as it preserves the 3D qualities of the microbes in their near natural state. The data needs fewer abstractions compared to other imaging techniques that require stained and dehydrated samples. Additionally, 3D data is not abstracted from reality like 2D images are, as it preserves our observed physical dimensions.

The shortcomings of cryo-ET. There are however

three important shortcomings of cryo-ET in light of this project. First, tomograms cannot capture the dimension of time. The samples are frozen and can therefore only capture a specific point in time. However, the variable of time is important in animations as they communicate dynamic events. The timing will have to be estimated from other sources of information to be incorporated in the animation.

Second, cryo-ET data is affected by an artefact called “missing wedge”. This artefact results in missing data in the z-axis parallel to the electron beam. It is caused by the increase in thickness with high angle tilts, which prevents complete a $\pm 90^\circ$ tilt series collection (Oikonomou and Jensen 2016). The “missing wedge” artefact results in elongated features in the z-axis that distort the image (Fig. 2). This generates an anisotropic image artefact that can affect the interpretability of the data, which will have to be compensated for.

Third, cryo-ET data is not shareable in its raw form; the file size can be too large, the file type inaccessible without proper software, and the data can be confusing to a lay-persons’ eye. Thus, even sharing raw files within the scientific community is cumbersome.

As a result of these constraints, abstractions of the data cannot be avoided in light of accessibility. However, it is pertinent to track the necessary

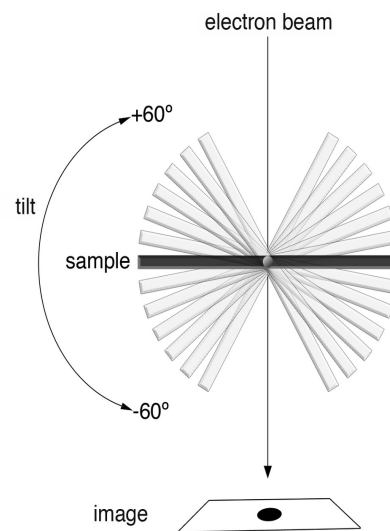


Figure 1. The sample is tilted with 2° increments ranging from $\pm 60^\circ$ in relation to the TEM.

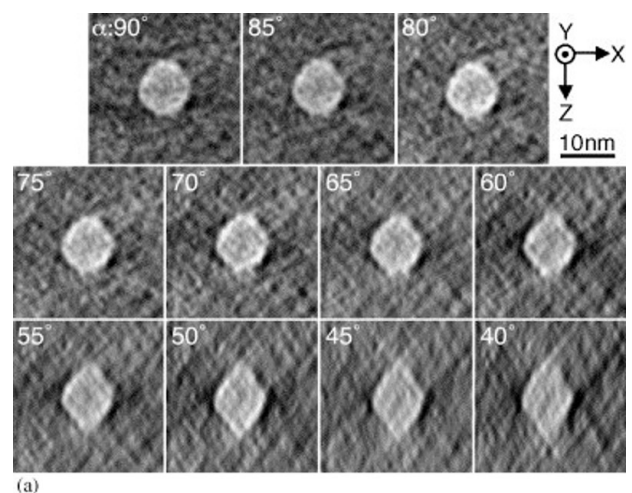


Figure 2. Demonstration of the “missing wedge” artefact as it elongates features in the z-axis (Kawase et al 2007)

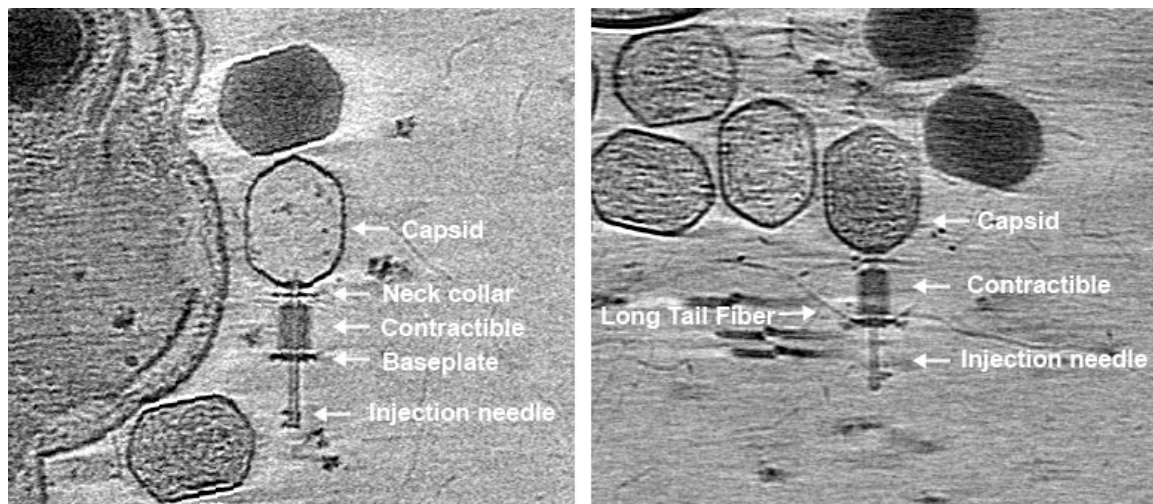


Figure 3. T4 bacteriophage subcomponents as described by (Hu et al. 2015b). From top to bottom: capsid, neck collar, contractible tail (STF's are hidden inside), baseplate, long tail fibres (LTF's), and injection needle.

abstractions as best as possible, to properly solidify the scientific credibility of the visualisation and shape the appropriate contextual framework.

A brief introduction to microbiology

The project covers the animation process of two visualisations: the first focusses on how T4 bacteriophage infects its host cell *Escherichia coli* in order to reproduce, the second visualisation is on the structural changes *E. coli* chemoreceptors undergo when the bacterium senses its surroundings. A short introduction to these interactions will be provided. Please note that neither short description tells the full story of these complex mechanisms. They do however encapsulate the essence of the stories in the animations of this project. The animations solely depict these processes, but keep in mind that it is not the full story.

Bacteriophage T4. The T4 bacteriophage is one of the more extensively studied model organisms in the field of Molecular Biology (Karam and Miller 2010). It is also one of the most complex viruses. The T4 bacteriophage consists of a capsid made up of more than 40 different proteins containing the 170-kb double stranded DNA genome, a neck collar, a contractible tail, a baseplate that harbours the short tail fibres (STF's), long tail fibres (LTF's), and an injection needle (Fig. 3) (Leiman et al. 2003; Hu et al. 2015).

The T4 bacteriophage propagates by infecting an *Escherichia coli* bacterium. In solution, most of the T4 phage's long tail fibres are folded against the tail (Fig. 4A). When in close proximity to its host, at least one of the T4 phage's LTF's interacts with the *E. coli*'s cell surface (Fig. 4B). This triggers a baseplate reconfiguration, revealing short tail fibres, and initiates reconfiguration of the neck-collar and contraction of

the tail (Fig. 4B-C). The contracted tail reveals the injection needle that allows the T4 phage to pierce the membrane of the *E. coli* (Fig. 4D). The injection needle cap protein (depicted in red) disassociates and the inner membrane of the *E. coli* bulges toward the injection needle (Fig. 4E), allowing the phage to inject its DNA into the host (Fig. 4E-F) (Williams 2019).

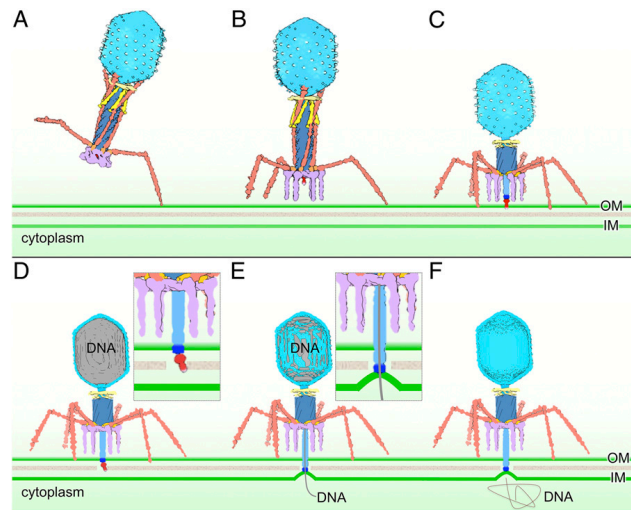


Figure 4. Bacteriophage T4 infection of *E. coli* as described by Hu et al (2015).

***Escherichia coli*.** Another well-studied organism in the field of biology is the *Escherichia coli* bacterium. *E. coli* is a gram-negative, rod-shaped bacterium that resides in the gastrointestinal tract of humans and animals (Lim, Yoon, and Hovde 2010). *E. coli*, like many other bacteria, is a motile bacterium that can sense its environment and respond to it. It can move to a more favourable location and retreat from toxic environments (Hazelbauer, Falke, and Parkinson 2008; Parkinson, Hazelbauer, and Falke 2015). This behaviour is called chemotaxis and this system is best understood in *E. coli*.

The chemotactic response in *E. coli* bacteria is moderated by structural proteins at the cell pole. The underlying sensory pathway of the chemotaxis system involves transmembrane receptors, called Methyl-accepting chemotaxis proteins or MCPs, that organize into membrane-bound arrays with the intracellular proteins CheW and CheA, which is made up of several domains (P1-P5) (Fig. 5) (Hazelbauer, Falke, and Parkinson 2008; Parkinson, Hazelbauer, and Falke 2015). CheA is a dimeric histidine kinase that is modulated by ligand binding events to the extracellular domains of MCPs. CheW is an adaptor protein that couples CheA to MCPs. The smallest functional unit of the array, referred to as the core unit, consists of six receptor dimers in a trimer-of-dimer oligomeric state, one dimeric CheA, and two CheW proteins (Li and Hazelbauer 2011). *In vivo*, these core units assemble into a large supramolecular complex with hexagonal (P6) symmetry that is

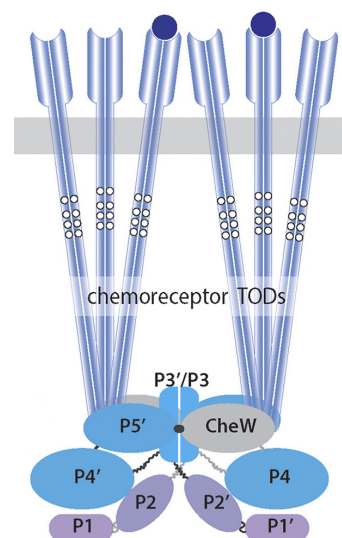


Figure 5. Structures of MCP receptors, CheW, and CheA kinase (P1-5) (Yang et. al. 2019).

relatively conserved in motile bacteria and archaea.

Multiple studies have found that components in the array undergo conformational changes when activated. In the kinase OFF-state, the receptor dimers in a trimer-of-dimers module, splay farther away from one another when compared to the kinase ON-state (Fig. 6). This effect is described as the dimer-zipping effect (Yang et al. 2019). In the kinase ON-state the P1 and P2 domains of CheA show an increased mobility. This suggests that the receptors control kinase activity at least in part by regulating the localization of the CheA P1 and P2 domains (Briegel et al. 2013).

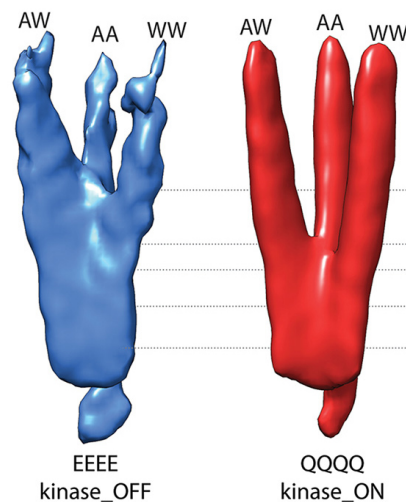


Figure 6. Density distribution of the receptor trimer of dimers in kinase-ON and kinase-OFF output states (Yang et al. 2019)

The target audiences

The animation on T4 bacteriophage infection of *E. coli* was designed specifically for the audience of the microbiology museum Micropia. The animation of chemotaxis in *E. coli* was designed to accompany a presentation at the American Society for Microbiology conference of June 2019. The target audiences' characteristics that influence the design process the most, will be highlighted next.

ARTIS-Micropia. Micropia is the world's first and only microbe museum to date and is a part of the Amsterdam Royal Zoo ARTIS². Both institutions strive to facilitate people from all background and ages, but also specifically tailor to families and young people ("Opzet Jaarverslag 2014 Natura Artis Magistra" 2014). Micropia aims to bring the otherwise invisible world of microbes to the public through a series of animations, interactive installations, and texts with general information of the microbe, their environment, and functionalities³. Its exhibitions are designed to provide an interactive and playful experience. These qualities are incorporated through the use of interactive tabletop screens for example. Due to the exhibit's overall playful nature, it is optimally designed to attract young people. However, the communicated information is relatively complex. Through these observations, and discussions with employee's at Micropia we estimate that visualisations fitting this setting, should be at least understandable for children in their

2 <https://www.micropia.nl/en/visit/what-is-micropia/>

3 <https://www.micropia.nl/en/footer/about-micropia/mission-vision/>

middle childhood stages up to the early adolescence (8 – 16 years). To achieve maximum effect in the most likely users, the visualisation on T4 bacteriophage infection will be designed to be understandable and enticing to this audience.

American Society for Microbiology, ASM. The animation on chemotaxis in *E. coli* was specifically designed to accompany a presentation at the ASM Microbe conference in June 2019. This conference is designed as a platform for microbiological experts to share their work with their peers. The goal of the conference is to share recent discoveries in the field of microbiology with other researchers and promote collaborations between peers⁴. As such, the target audience is expected to comprise researchers of various fields in microbiology, with advanced knowledge of their specific topics and a basic knowledge of more distant research fields. It is expected that most presentations at the conference will consist of talks and a supporting slideshow. Therefore, the goal of the animation is to aid the complex topics of the presentation with a visual demonstration. The animation can maintain a relatively high level of complexity since it is aimed at an expert audience. As such, it requires less abstractions from the original data.

The tangible products

These elements provide the basis and guiding principles for the visualisations created in this case-study. The design process of the most notable intermediary- and end-products will be described by highlighting the relevant sources, the feedback loops with important collaborators, the hardware specifications, and the software settings and most used functions. These products are created on a variety of platforms and can be accessed through the following link: <https://verawilliams.myportfolio.com/graduationproject2020> or by scanning the QR code in this paragraph. This portfolio website serves as a multi-platform appendix and is best viewed on a computer or laptop.



T4 bacteriophage script. The story design for the T4 bacteriophage visualisation was based on available literature by Hinton (2010), Hu et al. (2015), and Leiman et al. (2003). The content of these papers was reformulated to fit the intended audience by adjusting its complexity and the use of jargon. To check if this had the desired effect, the script was evaluated by a representative of the target audience. This representative was a fourth year high-school student with an affinity for biology. The representative pointed out overly long sentences and complex words. The feedback was implemented to tailor the script to the audience

4 <https://fems-microbiology.org/opportunities/asm-microbe-2019/>

more fittingly. To check the scientific accuracy, the script was regularly evaluated by the supervisor and adjusted where necessary. Any final adjustments to the script were made during the recording phase. These adjustments were purely aesthetic and did not influence the content of the story.

T4 bacteriophage storyboard. The storyboard of the T4 bacteriophage animation was based on its script and the drawings from the paper by Hu et al. (2015). The storyboard was created in an incremental fashion, and therefore revised several times. The earliest versions contained rough sketches of the placement and dynamics of the elements. Later versions were based on 2D images of cryo-ET models and structural models from the papers by Hu et al. 2015a and Leiman et al. 2003, to increase scientific accuracy. For example, in the improved drawings the ratio of the phage to the *E. coli* was mimicked more accurately, and the molecular structures were more detailed.

T4 bacteriophage moving storyboard. The moving storyboard combined the detailed drawings of the storyboard and the narration from the script. The refined drawings were scanned and imported in Adobe Photoshop. The scans were cleaned up using the eraser function and increasing the contrast. The images were animated with Photoshop's built-in animation tools. The narration was recorded with an external microphone and edited in Adobe Audition. The final narration and background music were added to the animated drawing to improve accessibility to the animation. Both provided cues for the timing of the animation (e.g. the tail of the phage contracts at the moment that this is stated in the narration). The moving storyboard was presented to a representative of the target audience and supervisors for feedback on accessibility and scientific accuracy.

T4 Bacteriophage 3D volume from cryo-ET data. The initial 3D T4 bacteriophage model was created in during a 5ECTS project space (Williams 2019). The project aimed to obtain a 3D model of T4 bacteriophage. This was done by segmenting T4 bacteriophage data with AVIZO. The data was collected by Lasse Sprankel from the Briegel lab. First, the data was filtered using a Gaussian filter on all three axes ($x=1$, $y=1$, $z=1$). Then, the structures were extracted in each slice and segmented per type. Segmentation was done manually or with the “magic wand” auto-selection tool, using a mouse and a Wacom drawing tablet (model CTH-461). The data were exported as a wave file (.obj) after segmentation.

Next, the model was smoothed out and sculpted by hand in Blender, using a Wacom drawing tablet (CTL-6100WL) and the “clay strips”, “pinch”, “smooth”, and “scrape” functions. The capsid, neck, baseplate, and

injection needle were smoothed out. The missing wedge artefact was partially corrected for by filling in the missing data. The structures of the contractile were added with the “clay-strips” function. High resolution areas on the contractile were mimicked as closely as possible. Affected areas in the neck collar and baseplate were smoothed out and symmetrised by hand as accurately as possible. The long feet fibres were smoothed out and thickness was added where necessary.

Averaged, isolated baseplates from T4 Bacteriophage data. A more accurate depiction of the loose-lying baseplates in the T4 bacteriophage data was created through sub-tomogram averaging in DYNAMO. The baseplates were first identified as isolated, hexagonal components in the tomogram. To better depict the loose baseplate in 3D, sixteen loose baseplates were manually picked from nine tomograms. Two rounds of alignment were performed; the first to align the particles based on the tilting angles, the second by the in-plane rotation angles. Due to the limited number of loose baseplates available, a C6 symmetry was applied to the final average to enhance the signal-to-noise ratio.

Animation of the flexible domains of the chemotaxis array in *E. coli*. This animation demonstrates the zipping-effect of the MCPs when switching from a kinase-OFF to the kinase-ON state, and the mobility of the P1 and P2 domains upon activation. The visual design for the chemotaxis animation was based on the images in the papers by Yang and Briegel (2019), Yang et al. (2019), Briegel et al. (2013). The colours were based on their repeated use in the literature by several research groups. The box-modelled p' domains were based on a cartoon drawing from the paper by Yang and Briegel (2019) (Fig. 7). The top- and front view images were placed in the top- and front-view ports respectively and the shapes and sizes of the elements were mimicked in 3D. This was done by adding and moving vertices, and scaling, rotating, and transforming a cube to fit the drawing. Feedback on the animation's scientific accuracy, colour preferences, focal points and timing was collected and implemented through brief evaluations of scene overviews in various stages of the process.

Reflecting on the experience

Besides the practical steps of this experience,

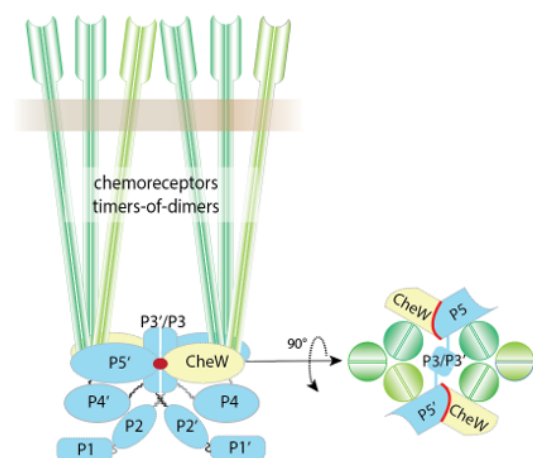


Figure 7. Front- and top-view images on which the box-modelled CheA and CheW units were based for the animation on chemotaxis in *E. coli* (Yang and Briegel 2019)

reflecting on the experience is potentially of greater importance in light of learning from it. The critical moments and bottlenecks are discussed in respect to the accessibility and accuracy features of the visualisations. Reflection on the observed bottlenecks with regard to the workflow proved of significant importance, as these obstacles influenced the feasibility and reproducibility of the final workflow significantly.

Reflection on the T4 bacteriophage project

The project on T4 bacteriophage infection of *E. coli* was created in close cooperation with a representative of the target audience. Feedback from a representative of the target audience proved invaluable during the writing and recording process, as it shaped the animation to become understandable for the target audience. This not only affected the content of the story, but also the narration's timing and emphasis, which were demonstrated to affect the efficacy of information transfer equally. The representative's feedback was implemented together with the input of biology professionals in an attempt to balance the accessibility features with proper accuracy.

Additionally, during this process the importance of close contact with experts on the subject matter became clear as well. Initially, T4 bacteriophage was selected as a subject for this project because it is very well understood. This project aimed to illustrate the available knowledge in a better and more accessible format. However, during the process we discovered unknown features, which we did not anticipate with such a thoroughly studied microorganism. We contacted phage experts at the TU Delft to discuss these findings. We arranged several meetings and had frequent digital correspondence. Unfortunately, this was not enough to maintain a feasible workflow. Feedback on the details of the data could not be collected as needed, which stagnated the project. This experience illustrated the importance of direct access to expert knowledge and its influence on the efficiency of the workflow.

Reflection on the chemotaxis in *E. coli* animation

The animation on chemotaxis of *E. coli* bacteria required fewer translatory steps because it was aimed at an expert audience. Accessibility was maintained by adhering to the colour palette used in the accompanying papers to avoid any miscommunication, and by basing 3D models of unavailable data on the illustrations in those papers. The movement of the transition from the kinase' OFF- to its ON-state was discussed frequently to accurately depict the current understanding of the conformational changes the molecules undergo.

The animation on chemotaxis of *E. coli* was presented during an opening lecture of the ASM Microbe conference which hosted 8000 attendees⁵. The animation was positively received, as it was remarked that the animation helped a researcher visualise the dynamics of the receptors and kinase domains more easily compared to just reading about the process in a scientific paper.

However, seeing as the animation does not provide any context, it was not deemed suitable for a lay audience. In its current state, the animation shows the isolated receptor units but negates the surrounding structures, its relation to the bacterium, and its surroundings. As such, this animation could probably only be effectively used alongside an elaborate presentation on the topic.

This illustrates the importance of designing a visualisation with the target audience and context in mind. Consideration of the target audiences' background knowledge seems to impact the efficacy of information transfer significantly.

Reflection on the process

In addition to these findings, reflections on the experience as a whole provide insight for the design of the general workflow as well. Throughout the process, the workflow seemed overly chaotic as a result of the back-and-forth between the tangible products, as each product builds on another (e.g. the script influences the storyboard, the storyboard influences data collection, data collection influences the visual design etc.). The project also experienced time management issues. For example, colour suggestions were made after rendering the chemotaxis animation. As a result, the shaders had to be adjusted and the animation re-rendered. This was very time-consuming and cost many precious hours which made the process less efficient.

In an effort to create more structure, an attempt was made to create an overarching form of linearity. This was achieved by incorporating the experiences of this project into the workflow described by William Vaughan in his book "Digital modeling" (Vaughan 2012). In his book, Vaughan illustrates an industry-standard workflow and flowchart for creating 3D animations. Vaughan highlights four separate stages: pre-production, production, post-production, and parallel production, and the feedback loops within and between these stages. Most notably, Vaughan illustrates when the project is ready to move on to the next stage and stipulates the importance of the overarching linearity between these stages (Vaughan 2012). Vaughan's work puts most of the obstructions of this project in a broader perspective on the overall workflow.

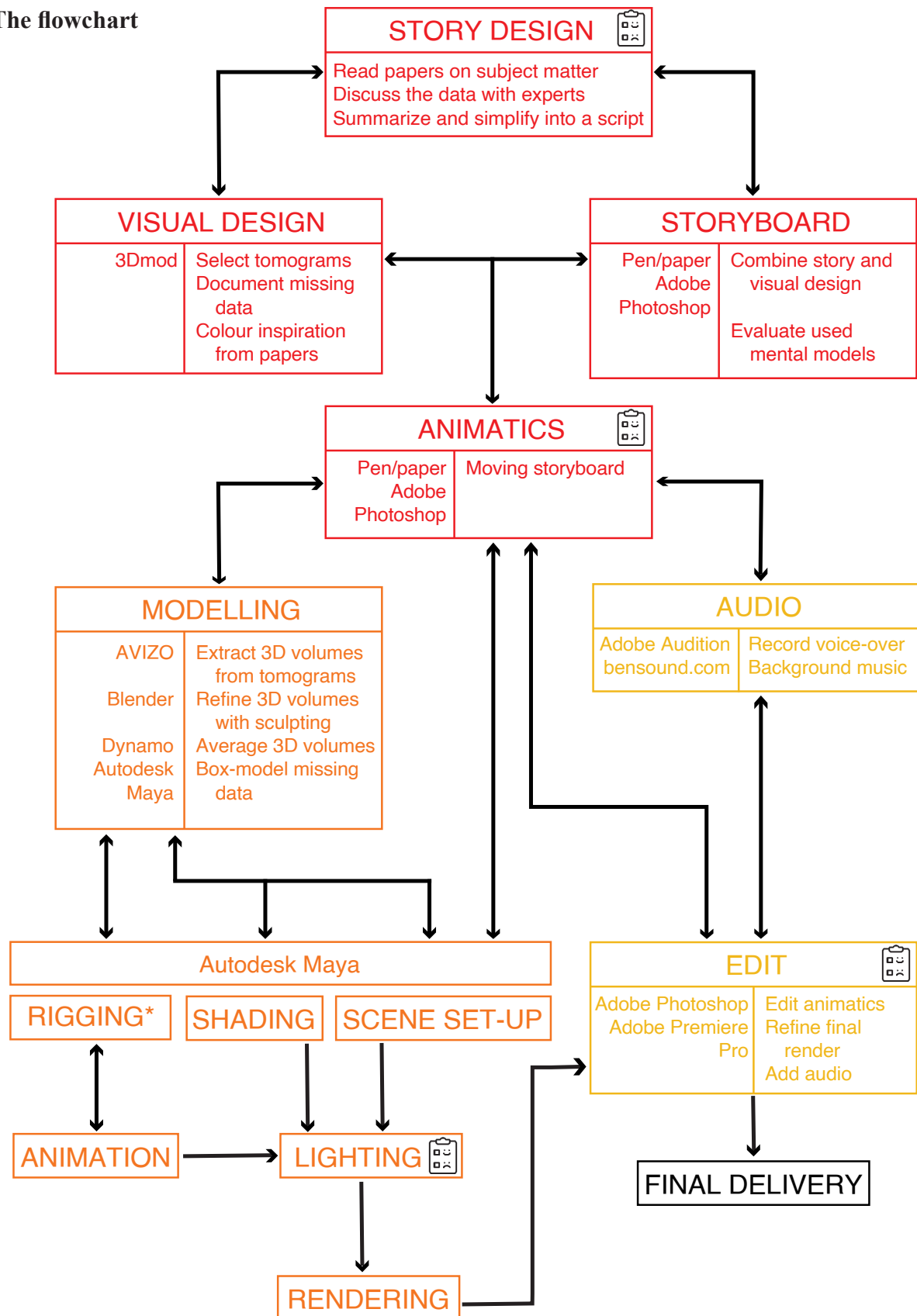
5 <https://www.youtube.com/watch?v=-YvGe36qyq4&feature=youtu.be>

By combining the experiences from this case-study with Vaughan's work, a novel workflow and flowchart could be created. The final workflow consists of three stages: Pre-production, Production, and Parallel production. A summary of this workflow will be provided by the accompanying flowchart. This flowchart will be introduced first as it provides a mental model for the workflow. This approach is necessary due to the complexity of the project; the project aims to combine several complex fields of expertise: microbiology, imaging technology, processing technology, and creative technology. Throughout the workflow, the intricacies of these fields of expertise are inextricably linked due to the non-linear nature of the process.

To avoid confusion, the interactions between each step are illustrated briefly in the flowchart before delving deeper into the specifics. The Pre-production stage is depicted in red, the Production stage is depicted in orange, and the Parallel production stage is depicted in yellow. The steps in each stage are depicted with a box. The colour of the box illustrates to which stage the step belongs. The box titles refer to the step's processes or tangible products. Beneath it, the individual tasks and accompanying software platforms are described on the right and left side, respectively. A detailed overview of the software platforms is provided in appendix A. The boxes are connected with one- or two-way arrows, indicating the workflow between the steps. The paragraphs following the flowchart describe the workflow and provide a more complete description of the three stages. The workflow will be colour-coded in accordance to the flowchart.

The flowchart is presented on the next page.

The flowchart



LEGEND

■ Pre-production
■ Production

■ Parallel production
↔ Two way process

→ One way process
📋 Feedback

Feedback icon made by www.flaticon.com/authors/those-icons

The workflow

Stage 1: Pre-production

During the pre-production stage, all the elements involved in production are prepared. The story, visual look, and the overall planning are decided on in this stage (Vaughan 2012). Any shortcuts taken in the Pre-production stage could directly affect the final outcome of the project (Vaughan 2012).

STORY DESIGN. First, decide on the initial concept of the animation. Break it down to the subject matter, the story and target audience. Contact experts on the subject and ask for guidance on appropriate papers. Read the selected papers to gain a good understanding of the subject matter. Formulate the story in a style that suits the target audience. The story can be written in a script format, but a screenplay is also an option (Vaughan 2012). Add any scene markers where necessary. This aids the creation of the storyboard.

Ask the experts to provide feedback on your story, focusing mostly on scientific accuracy. Finetune your story based on the expert feedback. Contact your target audience to set-up and perform user tests, implement their feedback as well. Focus mostly on the jargon, pace, understandability, and efficacy. These processes could require several back-and-forth sessions.

Checkbox Story design

- ✓ *Check for accessibility and understandability with feedback from (a representative of) the target audience*
- ✓ *Check for scientific accuracy with feedback from experts*
- ✓ *Check for understandability and auditory flow during the recording stage*
- ✓ *Implement feedback on moving storyboard in an improved version of the script*

VISUAL DESIGN. Base the colour inspiration of the 3D-models on the images of selected papers to avoid confusion when communicating the work to the experts. But keep the primal effects on colours in mind to avoid any miscommunication to your audience (i.e. don't use red and green to depict opposite effects because the viewers could relate it to "right" and "wrong" which could lead to false interpretations of the story). Optionally, create a colour palette suitable for feedback in Adobe Photoshop and check its colour blindness qualities in Adobe Illustrator (view > proof setup > colour blindness > Protanopia-type / Deuteranopia-type).

Obtain the cryo-ET data in a workable format from the experts. Assess the available cryo-ET data

with AVIZO and 3Dmod. Use the script and the storyboard as a guideline to which models are needed and in which configuration. Keep a detailed log with the file names and -locations that contain proper data (e.g. in a logbook). Take note of any missing data. These can be modelled manually in the production stage or retrieved from online databases or other research groups.

Checkbox Visual design

- ✓ *Create the colour palette based on the selected papers*
- ✓ *Otherwise, design a colour palette to fit the overall feel of the visualisation*
- ✓ *Perform a colour blindness check*
- ✓ *Collect feedback from the target audience through the user tests on the moving storyboard*
- ✓ *Keep a detailed log on the available- and missing data, as well as averaging/box modelling plans*

STORYBOARD. A storyboard consists of a set of drawings depicting the key moments of the animation. A storyboard can simply be made using pen and paper. Divide the storyboard according to its scenes and subsequent acts as designed in the script (i.e. scene 1, act 1). Use a template to make this easier, examples can be found online. Briefly describe key moments in the allotted space in the template. Add camera directions (e.g. zoom-in and -out) and increases or decreases of pace. Feedback on the moving storyboard will circle back to the storyboard. Incorporate directions on scientific credibility, accessibility, and understandability to an improved storyboard.

Checkbox Storyboard

- ✓ *Divide the template into key scenes and acts*
- ✓ *Add camera directions*
- ✓ *Implement feedback on the moving storyboard into an improved storyboard. Focus on scientific credibility, accessibility, and understandability.*

ANIMATICS - creating a moving storyboard. Create a simple moving storyboard from the improved storyboard in Adobe Photoshop to get a better sense of the timing in the animation. Scan the refined drawings and import them into Adobe Photoshop. Use the eraser, magic wand, and contrast tools to clean up the drawings. Animate the drawings using the built-in animation functions and add the voice over and background music. Synchronise the critical moments in the animation with the voice over and background music to create a more harmonious product.

Collect feedback from the target audience, experts, and supervisors to evaluate the accuracy and efficacy of the product. Implement the feedback in the script, storyboard and moving storyboard where necessary.



Checkbox Moving storyboard

- ✓ *Ensure scientific credibility by collecting feedback from the experts*
- ✓ *Ensure accessibility by collecting feedback from the target audience*
- ✓ *Circle back to the script and storyboard to implement feedback*
- ✓ *Optionally, recreate a moving storyboard from the improved script and storyboard*


Checkbox Pre-production stage

Before moving on from stage 1:

- ✓ *Finalise the script*
- ✓ *Finalise the storyboard*
- ✓ *Finalise the visual design blueprint*
- ✓ *Collect feedback on the moving storyboard and implement it where necessary. (re)Create a final moving storyboard as the basis for the animation.*

Stage 2: Production

Do not move on to this phase before finalising the Pre-production stage. In the production stage, the final elements of the animation are created as designed in pre-production. Stay true to the script, storyboard and visual design that were created in the pre-production phase as much as possible (Vaughan 2012). This helps maintain a proper workflow throughout the process which saves time and costs. Before moving on to this phase, every element must be settled with any other party involved (e.g. supervisor or client). Any adjustments to the backbone of the project will obstruct the workflow significantly and possibly cause delays and obstructions. With certain modelling steps as the exception, this method's production stage is done entirely in Autodesk Maya 2018 software.



MODELLING. *Tomogram segmentation.* This is the first step to obtain 3D models from the cryo-ET data. Refer to the logbook from the “Visual design” step in the Pre-production stage to find the correct data files and -locations. Open the files in AVIZO. Apply a filter for cleaner segmentation if necessary. Assign a material to each subcomponent to maintain a good overview of the data. Segment the subcomponents with the magic-wand auto selection tool. This can best be done with a graphics tablet such as a Wacom drawing tablet, but a regular mouse works too.

Sculpting. Once segmentation is complete, export the data as a wave file (.obj). This allows the data to be imported into other 3D software. Refine the model in Blender if necessary. Correct for the layer-effect to smooth out the surfaces, connect loose-lying parts and add definition where needed. Use the “clay strips”, “pinch”, “smooth”, and “scrape” functions. Correct for the missing data in

the z-axis that results from the “missing wedge” artefact as well. Sculpt in any missing data but adhere to the existing data as much as possible. Refer to similar areas that are properly defined in the original data (For a visual example on how sculpting partially corrects for the missing data that results from the “missing wedge” artefact, refer to the “T4 Bacteriophage 3D volume from cryo-ET data” item on the portfolio website).

Sub-tomogram averaging. Refined models can also be obtained from the tomograms through sub-tomogram averaging in DYNAMO. Perform alignments based on the tilting angles and the in-plane rotation angles. Apply symmetry masks to enhance the signal-to-noise ratio. NOTE: due to the specificity and complexity of the software, help from an expert might be required.

Box-modelling. Models that cannot be obtained from the available data can be created through box-modelling in Autodesk Maya. Box-modelling is a common 3D modelling technique where objects are created by adding complexity to a simple cube (e.g. by adding edges and moving vertexes). Box-models can be based on “cartoon” figures of the data and replicated in 3D. Ideally, at least a top and front view figure should be available. Place these in the top- and front-view channels consecutively. Shape, scale and rotate the elements in the proper position.

Checkbox Modelling

- ✓ *Collect all the necessary models as described in your visual design log*
- ✓ *Design all models in accordance to the colour palette*
- ✓ *Sculpted models should adhere to the original data as much as possible*
- ✓ *Box-models should be based on available figures from papers*

***RIGGING.** Rigging can be described as giving the model the “hidden mechanics” that allow it to be animated in a correct way (Vaughan 2012). The workings of a model rig be compared to how our elbow allows us to bend our arm, for example. The rig allows the 3D model to deform in the most natural way. Although this step was not necessary for the animations in this project, rigging is of great importance in animations that require flexibility within 3D-models. As such, it could not be left out of workflow.

SHADING. Use Arnold shaders when planning to render with the Arnold renderer in Autodesk Maya. Base the shaders on the colour palette created in “Visual design” step of the the Pre-production stage. Adjust the specular highlights, Fresnel effects and specular transmission to fit the

overall look and feel of the animation.

SCENE SET-UP. When setting up the scene, mimic the storyboard as much as possible. Import the separate models into one scene in Maya. Focus mostly on the $t = 0$ moment: all elements need to be where they are needed at the beginning of the animation. Models can be put in place by using the translate tool in Autodesk Maya.

The use of a background plane is optional but recommended. Use the storyboard and visual design as guidelines. Pick a background that fits with the overall look and feel of your animation.

ANIMATION. Animate the models in Autodesk Maya with the built-in animation functions.

Base the animation on the moving storyboard; this provides instruction on the timing and pace in accordance to the selected background music and voice-over.

Set the keyframes for each individual object with the right timing, the software automatically interpolates the movements inbetween. For example: the chemoreceptor complex in the chemotaxis animation is rotated by setting the first key frame on rotation $x = 0$ and $t = 0$, and the next on $x = 360$ and $t = 6$. This allows that the model to turn 360° in three seconds from the start of the animation.

LIGHTING. Keep the lighting simple. For example: use an Arnold dome light to light the entire scene and an Arnold spotlight to draw more attention to the focal point. Arnold lights are necessary when rendering with the Arnold renderer.

RENDERING. Rendering is done in the foreground of the computer and therefore takes up most of the computer's power. For efficiency sake, render the animation overnight. For larger projects with multiple scenes, render the priority scene's out first (Vaughan 2012). Renders output a frame sequence of .png image files. Import the image sequence in Adobe Premiere Pro to make final edits.

Checkbox Production stage

- ✓ Stay true to the visual design blueprint when creating the models, and assigning their shaders
- ✓ Stay true to the storyboard and moving storyboard for the scene set-up and animation processes
- ✓ Prioritise important scenes when rendering multiple scenes
- ✓ Render overnight for efficiency

Parallel production

Parallel production consists of audio design and editing. In contrast to the edits that are done throughout the design process of the individual steps, the step “Edit” refers to adjustments made to the overall visualisation. Editing in this sense focuses on the moving storyboard and the final animation. Editing can also be described as finetuning the visualisation. Much of this is done during the creation of the animatic, but the animation can be tweaked in other stages, as well (Vaughan 2012). Like editing, audio design runs parallel to the other stages because it is an important part of the animatics. The narration is first implemented and finetuned in the Pre-production stage, and is eventually implemented in the final animation during the Production stage.

AUDIO. Narration. Record the narration from the script designed in the Pre-production stage with Adobe Audition and a microphone. The built-in microphone of a laptop provides adequate quality, but an external microphone would be better. Recorded voice-over in a quiet space to reduce background noise. Replay the recordings to assess the intonation, clarity and focus of the story. Soften the “t” and reduce hissing on the “s”. Avoid breathing into the microphone and articulate as clearly as possible. Speak at a slower pace to make the narration more comfortable to listen to.

Once the recordings are satisfactory, edit them in Adobe Audition. For example, reduce the background noise with “Noise reduction”, “Hiss reduction”, “De-hummer”, and “Automatic click-remover”. Cut the silences before and after the actual narration and edit any mistakes. Optionally, add fade-in and -out effects to smooth the transitions between recordings. Export the recording as .m4a file.

Background music. Select background music from a royalty free website such as www.bensound.com. Pick background music that fits with the overall style of the story and the target audience. This fills up the silence and helps ease the audience into the story. The movements of the animation can be timed with the background music for more emphasis on certain events. But do not stray too far from the original design.

EDIT. The animatics can be edited by adjusting its components: the storyboard, the script, the visual design, and the audio. Use the feedback on the moving storyboard to guide the edits. Implement the feedback as accurately as possible. Each product can be edited individually and then brought back together in a new rendition of the moving storyboard.

Final edits of the animation can be done in Adobe Premiere pro. Import the render sequence obtained in the Production stage into Adobe Premiere Pro. Adjustment the timing if necessary and add the narration and background music. Invest in finetuning the final rendition of the visualisation for the best possible result.

Checkbox Parallel production stage

- ✓ *Ensure the quality of the narration recordings*
- ✓ *Find suitable background music*
- ✓ *Use the narration and music to emphasise certain events, but don't stray too far from the blueprint*
- ✓ *Edit the feedback into the work to tailor the final product to your audience as precisely as possible*
- ✓ *Invest in finetuning the final visualisation*

Discussion

This project aimed to demonstrate a novel method for creating scientifically credible visualisations. It posed the question: “*How can we transform high-resolution cryo-electron tomography data into a widely accessible yet scientifically accurate visualisation?*” Through a case-study approach, this project attempted to answer this question by developing a “how-to” workflow. The critical characteristics for its product design were the two requirements posed in the research question: *accuracy* and *accessibility*. These qualities were incorporated by studying the background and needs of the target audience, monitoring the necessary abstractions, and balancing these findings with the animations’ goal. Reflection on the critical moments and bottlenecks of this process, combined with an industry standard workflow for 3D animation, resulted in the proposed workflow and flowchart.

The proposed method aimed to demonstrate a *feasible, transparent, and reproducible* workflow in an attempt to resolve some of the core issues concerning the reception of scientific visualisations. By providing transparency on the data transformations, the project hoped to inspire trust in the visualisations, and create a more accurate framework in which to interpret them. The method shows promising feasibility and illustrates transparency where it could be provided, which implies the potential of reproducibility of this method.

However, the method’s feasibility is heavily dependent on the available resources. The quality of the animations’ accuracy and accessibility features are dictated by the level of access to expert knowledge and the target audience, respectively. To achieve the optimal level of cooperation between such far-removed groups could be a rather challenging feat. However, these close collaborations could be invaluable in the endeavour of making scientific knowledge more accessible beyond the world of science, and therefore worth the effort.

Additionally, to provide full transparency, the abstractions made in each step should be communicated more precisely. Although this method provides transparency on the abstractions of the cryo-ET data, it could be beneficial to provide transparency on the other data-types as well. However, this would require detailed knowledge of each involved field of expertise: microbiology, imaging technology, processing technology, and creative technology. The trade-offs would be too severe if such a broad spectrum of knowledge were to be expected of a single person. If this level of transparency is to be achieved, it should be attempted through a group effort. However, it would be interesting to investigate in how far total transparency actually improves trust in visualisations and provides an appropriate framework, to discern if this is even necessary.

Lastly, the reproducible qualities of this method can in so far only be speculated about. The proposed method would probably be reproducible if a certain level of proficiency in microbiology, data collection, 3D modelling, storytelling, design, and animation is available in advance. However, such restrictive requirements do not lend themselves to a convincingly reproducible method. As such, the method probably does not provide sufficient instruction for the average scientists that would like to transform their data into a visualisation. It would therefore be interesting to first compare this method with similar projects from the field of scientific visualisations, to ascertain its reproducible qualities on this niche level.

Future work

The development of this method could greatly benefit from much more: the quality of the products could be improved with structured user-tests. This could specifically illuminate the elements that influence the accessibility features of a visualisation most. In light of scientific accuracy, it would be ideal if all the used 3D-models were based on averaged cryo-ET datasets. This would minimize the amount of man-made data and illustrate a more stereotypical example. The method could also profit from a deeper understanding in design, modelling, and animation principles, which would improve the overall quality of its products. Additionally, it might be interesting to investigate if this method is applicable in other scientific fields as well. It might be useful for research that works with other complex datasets such as physics, industrial design, and astronomy. A plethora of alternative software could also be explored even further. The proposed software is somewhat expensive. In light of reproducibility it might therefore be interesting to experiment with alternative software. Additionally, modelling and animation technologies are developing very rapidly. It would be fascinating to use such cutting-edge techniques to make scientific visualisations even more appealing. In light of this, there is still much to discover in the field scientific visualisations, and the

possibilities for innovation seem boundless and exciting.

Thank you I would like to thank everyone who helped make this project possible. Thank you, Ariane Briegel and Fons Verbeek for supervising this project and supporting my personal exploration of the topic. Thank you, Janet Iwasa for video chatting with me when I would get stuck with the project, and for being a peer during my presentation. Thank you, Jasper Buikx, for your input on the animations and for being a peer during my presentation, as well. Thank you, Briegel Lab colleagues for sharing your expertise of imaging software and chemotaxis topics and letting me tag-along your cryo-ET data collections at NeCEN. Thank you, fellow Media Technology students and Media Technology teachers for brainstorming with me. Thank you, Vera van de Seyp for improving the flowchart's lay-out. Thank you, Lael* for being an amazing representative of the target audience for the T4 bacteriophage story. And, of course, a massive thank you to my partner, parents, and sister, for supporting me when things were tough. And thank you to all my friends, coaches, and councillors who enabled me to continue with this project. I would not have been able to do it without you!

Bibliography

- Briegel, Ariane, Peter Ames, James C Gumbart, Catherine M Oikonomou, John S Parkinson, Grant J Jensen, and Salt Lake. 2013. "The Mobility of Two Kinase Domains in the Escherichia Coli Chemoreceptor Array Varies with Signalling State," 1–11. <https://doi.org/10.1111/mmi.12309>.
- Burri, Regula Valérie. 2013. "Visual Power in Action : Digital Images and the Shaping of Medical Practices Visual Power in Action : Digital Images and the Shaping of Medical Practices" 5431. <https://doi.org/10.1080/09505431.2013.768223>.
- Gibbons, Michelle G. 2017. "Seeing the Mind in the Matter: Functional Brain Imaging as Framed Visual Argument." *Argumentation and Advocacy* 43 (3–4): 175–88. <https://doi.org/10.1080/00028533.2007.11821673>.
- Hazelbauer, Gerald L., Joseph J. Falke, and John S. Parkinson. 2008. "Bacterial Chemoreceptors: High-Performance Signaling in Networked Arrays." *Trends in Biochemical Sciences* 33 (1): 9–19. <https://doi.org/10.1016/j.tibs.2007.09.014>.
- Hinton, Deborah M. 2010. "Transcriptional Control in the Prereplicative Phase of T4 Development." <https://doi.org/10.1186/1743-422X-7-289>.
- Hu, Bo, William Margolin, Ian J Molineux, and Jun Liu. 2015a. "Structural Remodeling of Bacteriophage T4 and Host Membranes during Infection Initiation," no. 28. <https://doi.org/10.1073/pnas.1501064112>.
- . 2015b. "Structural Remodeling of Bacteriophage T4 and Host Membranes during Infection Initiation." *Proceedings of the National Academy of Sciences of the United States of America* 112 (35): E4919–28. <https://doi.org/10.1073/pnas.1501064112>.
- Iwasa, Janet H. 2010. "Animating the Model Figure." *Trends in Cell Biology* 20 (12): 699–704. <https://doi.org/10.1016/j.tcb.2010.08.005>.
- . 2016. "Scienti Fi c Life The Scientist as Illustrator." *Trends in Immunology* 37 (4): 247–50. <https://doi.org/10.1016/j.it.2016.02.002>.
- Leiman, P G, S Kanamaru, V V Mesyanzhinov, F Arisaka, and M G Rossmann. 2003. "Review Structure and Morphogenesis of Bacteriophage T4." *CMLS, Cell. Mol. Life Sci* 60: 2356–70. <https://doi.org/10.1007/s00018->

- Li, Mingshan, and Gerald L. Hazelbauer. 2011. "Core Unit of Chemotaxis Signaling Complexes." *Proceedings of the National Academy of Sciences of the United States of America* 108 (23): 9390–95. <https://doi.org/10.1073/pnas.1104824108>.
- Lim, Ji Youn, Jang W. Yoon, and Carolyn J. Hovde. 2010. "NIH Public Access." *J Microbiol Biotechnol* 20 (1): 5–14.
- Oikonomou, Catherine M., and Grant J. Jensen. 2016. "A New View into Prokaryotic Cell Biology from Electron Cryotomography." *Nature Reviews Microbiology* 14 (4): 205–20. <https://doi.org/10.1038/nrmicro.2016.7>.
- "Opzet Jaarverslag 2014." *Natura Artis Magistra*, 2014. <https://www.artis.nl/nl/footer/over-natura-artis-magistra/jaarverslagen/>
- Parkinson, John S., Gerald L. Hazelbauer, and Joseph J. Falke. 2015. "Signaling and Sensory Adaptation in Escherichia Coli Chemoreceptors: 2015 Update." *Trends in Microbiology* 23 (5): 257–66. <https://doi.org/10.1016/j.tim.2015.03.003>.
- Vaughan, William. 2012. "Digital modeling." *Berkeley: Pearson Education*.
- Williams, Vera. 2019. "[Revisiting T4Phage."
- Yang, Wen, and Ariane Briegel. 2019. "Diversity of Bacterial Chemosensory Arrays." *Trends in Microbiology*, 1–13. <https://doi.org/10.1016/j.tim.2019.08.002>.
- Yang, Wen, C Keith Cassidy, Peter Ames, Christoph A Diebolder, Klaus Schulten, and Zaida Luthey-schulten. 2019. "In Situ Conformational Changes of the Escherichia Coli Serine" 10 (4): 1–14.

Appendix

A. List of software

3Dmod (v 4.7.4) by the University of Colorado is a software that allows you to view the tomograms¹. It was used to view the available data and select the files containing promising 3D volumes.

Adobe Photoshop (v. 21.0.3) from the Adobe creative suite is an artistic software used to create graphics or edit photos. In this project it was used to create the moving storyboard. I scanned in pen and paper drawings and imported them. Using Photoshop's basic animation functions I created a simple animation that highlights the timing and focal points of the storyboard in an easily editable sketch format.

AVIZO/AMIRA (v. 6.7) by Thermo Fisher is high performance 3D visualisation and analysis tool for scientific data². The selected tomograms were imported into AVIZO and 3D volumes obtained with the tracing tools. This was done on a Linux operating system.

DYNAMO (v. 1.1.398) is an open source software package for processing tomographic data. In this project it was used to create sub-tomogram averages of specific subcomponents of the data. This was done on a Linux operating system.

Blender (v. 2.79b), an open source 3D software³, was used to refine the 3D volumes created in AVIZO. Through digital sculpting the models were smoothed out using mainly the “clay strips”, “pinch”, “smooth”, and “scrape” functions.

Maya 2018 (student version) by Autodesk is the industry standard for animation, 3D-modeling, simulation and rendering⁴. During this project it was used to box-model any missing 3D volumes, create the animation, shading, scene set-up, lighting and rendering.

UCSF Chimera (v. 1.13.1) is a program for interactive visualisation and analysis of molecular structures⁵. Any missing models could also be replaced by existing crystal structures obtained from the PDB database. These were viewed and converted in 3D files (.obj) through Chimera.

Adobe Audition (v. 12.1.5) from the Adobe creative suite is an audio recording and editing software. It used to record and edit the voice-over and background music. Voice recordings were done with the built-in microphone of my laptop (Realtech® Audio codec with Dolby APO).

bensound.com provided the royalty free background music used in the animations.

Adobe Premiere Pro (v. 13.1.5) from the Adobe creative suite is a movie editing software. It was used to edit the rendered animation. Any edits with regards to the timing, the background music and voice-over recordings are done in this stage.

1 <https://bio3d.colorado.edu/imod/doc/3dmodguide.html>

2 <https://www.thermofisher.com/nl/en/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/3d-visualization-analysis-software.html>

3 <https://www.blender.org/>

4 <https://www.autodesk.com/products/maya/overview>

5 <https://www.cgl.ucsf.edu/chimera/>