USING CRYO-ELECTRON MICROSCOPY TO DETERMINE THE 3-DIMENSIONAL STRUCTURE OF FIBRINOGEN

by

Elizabeth G. Viverette

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Elizabeth G. Viverette

Greenville, NC

May, 2021

Approved by:

Dr. Nathan E. Hudson

Department of Physics of the Thomas Harriet College of Arts and Sciences

Contents

Table of Figures

1. Abstract

Fibrinogen is a large protein that is the soluble protein precursor to fibrin, which forms fibers that hold blood cells together to create blood clots. Research into the properties of fibrinogen is necessary because of its role in many thrombotic ailments, including strokes and heart attacks. The structure of fibrinogen has been studied through various methods, though certain regions of fibrinogen have not yet had their structures conclusively determined. These regions all have apparent flexibility and mobility of the region that is lost in sample preparation of other imaging methods such as the formation of crystal structures for x-ray crystallography.

Cryogenic Electron Microscopy (cryo-EM) is a technique for imaging proteins and other macromolecules at resolutions of 1-3 Å while preserving native structures such as flexibility. Cryo-EM involves freezing samples in a layer of amorphous ice to remove interference from frozen crystal structures. Because of this sample preparation procedure, fibrinogen can be imaged in its natural state. CryoSPARC and other Cryo-EM analysis software works by identifying isolated molecules and grouping them into relative angle orientations that are then condensed and averaged into 2-D structures. The initial process of 2-D angle classification and data refinement can be automated to some degree with neural networks, such as Topaz, developed and trained specifically to assist with Cryo-EM analysis. Once 2-D structures have been optimized, a complete three-dimensional render of the protein can be reconstructed using Gaussian best-fit algorithms.

A data set of over 500 GB has been collected using the Cryo-EM microscope at the National Institute of Environmental Health Sciences. This data set includes samples collected via negative staining. Multiple software packages for Cryo-EM analysis will be used to choose appropriate

particles for use in the final structure and render initial and final structures, including EMAN2, CryoSPARC, and a neural network specifically trained to process fibrinogen data. The particles to be used to image 2-D and 3-D maps of fibrinogen will be both manually and automatically selected and sorted into angle templates. This is complicated by the numerous possible flexibility states of fibrinogen. These templates will be refined and averaged over multiple iterations to create an optimal data set for the construction of a high-resolution structure map of fibrinogen in its natural, flexible form.

2. Introduction

2.1 Fibrinogen Structure and Function

Fibrinogen is a large soluble glycoprotein present in blood plasma that is essential for blood clotting (1). Fibrinogen has been observed to measure approximately 460 Å from end to end, with a notably flexible and elastic structure. Fibrinogen is bilaterally symmetrical across a central nodule and composed of two sets of three protein chains: Aα, formed from 610 amino acids; β, formed from 461 amino acids; and γ, formed from 411 amino acids (2). These symmetrical halves are referred to as the D region and are connected at the central E region by twenty-nine disulfide bonds. These three regions can be further subdivided into regions such as the coiled-coil region, the αc region, and the globular C-terminus of the β and γ polypeptide chains of the D regions. The central E region forms an asymmetrical globular nodule at the N-terminus of the three chains, with the exact structure of the α and β chains in the region being unresolved due to disorder or localized flexibility of the area. The coiled-coil regions of fibrinogen are present in both the E and D regions and are formed from the triple α helix coil of the α, β, and γ chains (3). The two coiled-coil regions terminate on either end of the fibrinogen

molecule with globular folding patterns of the C-terminus of the β and γ chains (4). The C-terminus of the α chain is connected to the main coiled-coil structure of fibrinogen and folds back upon itself briefly before separating from the main structure of fibrinogen. The α chain terminates in region known as the αC domain, which is connected to the D region via the αC connector.

When triggered by the enzyme thrombin, fibrinogen is converted into the insoluble protein fibrin, which is an important protein in the formation of blood clots. The transformation of fibrinogen to fibrin occurs when thrombin cleaves certain amino acids from the N-terminus ends of the α and β chains of fibrinogen, with the γ chain remaining intact (1). This cleaving leaves exposed knobs on the α and β chains in the E region of fibrin that correspond to holes on the C-terminus of the γ and β chains on the D region (3). This process allows fibrin to form half-staggered fibers that provide a scaffold-like structure for blood clot formation through a series of knob-hole binding sites, demonstrated in figure 1.

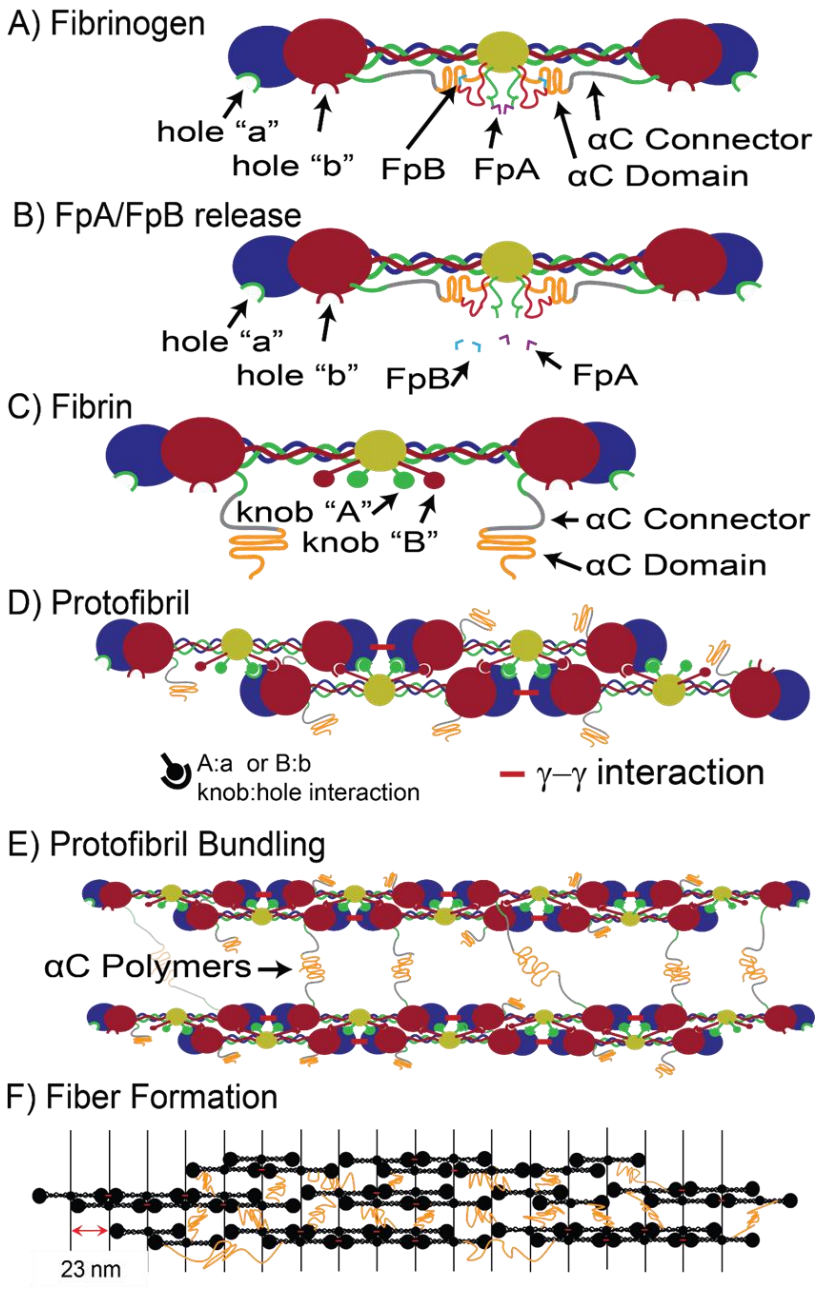

Fig. 1. Polymerization of fibrinogen to fibrin and fiber formation. In this diagram, the E region of fibrinogen is yellow and the D region is characterized by red and blue nodes, which are the β and γ regions, respectively.

The overall structure of fibrinogen has been studied through various methods, including through x-ray crystallography (4) and electron microscopy (5). Electron microscopy was first

used to confirm that fibrinogen takes the form of a three node protein with one central nodule in the E domain and two identical globular nodes on either end of the coiled-coil regions (6). Further analysis via x-ray crystallography has been used to determine the overall structure to resolutions of 3.3 Å in the coiled-coil regions with lower resolutions present in the globular regions of fibrinogen. The C-terminus of the α chain (the αC region) of fibrinogen folds back on itself in the D region of fibrinogen before it separates from the main structure. The globular αC region is connected to the main structure of fibrinogen by the αC connector region, which forms a structure that has been undetermined via x-ray crystallography. The difficulties in determining the structure of the αC region is compounded by the apparent flexibility and mobility of the region, likely as a result of the αC connector region. Similar difficulties caused by localized flexibility have occurred in the N-terminus of the α and β chains in the E region of fibrinogen.

There has been more success with determining the crystal structure of non-human fibrinogen, notably chicken fibrinogen, and existing crystallographic structures of human fibrinogen have been compared to the crystal structure for chicken fibrinogen, which includes more structural information about the αC region (4). Chicken fibrinogen has a more complete αC region because it lacks a repeated sequence of amino acids in the α chain that is present in the human α chain that forms the αC connector region (7, 8). Efforts to resolve more structural information about fibrinogen have been mostly focused on the α chain, specifically the αC region, which has not yet been fully resolved through x-ray crystallography. However, there has still never been any reliable structural information determined about the αC region, other than the confirmation of its elusive mobility (9).

In addition to the flexibility issues plaguing the structure mapping of the αC region, there are numerous other regions of fibrinogen that show a high degree of flexibility (4). This notably occurs in the coiled-coil regions of fibrinogen where the α, β, and γ chains contain sequences that are not compatible with the α helix structure of the region (10). This hinge site is located at residues 70-80 of the γ chain in the coiled-coil region, and has multiple possible orientations of its bend angle that have been observed through x-ray crystallography. Other sites with possible hinging include the region where the coiled-coil region meets the β nodule, where residues 148-160 of the α chain have been found to be accessible with fibrin and not fibrinogen (11). This suggests a certain amount of movement in the area that would allow for knob-hole binding. The β nodule may also hinge with respect to the γ nodule to expose binding sites along the γ chain residues 312-324 (12). These residues are located near the contact point between the two nodules.

This research project aims to use cryogenic electron microscopy to create a high-resolution structure of fibrinogen. This structural map of fibrinogen aims to include information about possible flexible hinge sites along the E region, coiled-coil regions, and D regions, such as exact locations and a range of motion at each site.

2.2 EM Theory

Electron microscopy is a technique first developed in the 1930s that uses the wave behavior of electrons to image a sample. The electron microscope was developed as a way of improving traditional optical microscopy as the wave behavior of electrons has a much shorter wavelength than that of light. The maximum resolution of a microscope image, or micrograph, is inversely dependent on the wavelength of the photons or electrons emitted by the

microscope; therefore, an electron microscope is capable of resolving images to a much higher resolution than optical microscopes (13). Transmission Electron Microscopy (TEM) is a specific method of electron microscopy where a focused electron beam is fired through a sample by a high voltage electron source. Diffraction patterns result from the electrons striking the sample, forming an electron density contrast pattern in the created image. This contrast pattern is based on both the amplitude and phase of the electrons when they are scattered by the sample. Amplitude contrast results from the effects of scattering removing some electrons from the density pattern, either by adsorption into the sample or by an angle of diffraction that results in the electron not being incident on an imaging device. Amplitude contrast in non-crystalline structures results from mass-thickness contrast, which is the relative difference in the angle of diffraction of areas of low mass compared to areas of high mass. Because of this, areas of relatively high mass will diffract electrons at a larger angle, resulting in a lower electron density in those areas.

Mass-thickness contrast is utilized in multiple forms of TEM, including a technique called Negative Stain Electron Microscopy (negative staining). technique where samples are prepared in a layer of heavy metal salt to provide a high contrast for the imaging of small samples (14). It is simpler than other sample preparation methods and can be used as a quality control check for samples before imaging with more labor and material intensive techniques. Negative staining requires a uniform layer of heavy metal salt to be applied to the sample which diffracts the electrons of the source due to a high concentration of neutrons in the stain (15). This stain can be applied relatively quickly and cheaply, and does not require any specific imaging conditions beyond the normal vacuum environment of an electron microscope. Due to the size

of the metal salt used to stain a sample, the resolution of images that can be obtained is somewhat limited compared to other microscopy techniques. Despite this, negative staining remains a common technique, as it can be performed with any electron microscope and produce a sample image relatively quickly and cheaply. It is often an initial step in other, more involved imaging techniques, as a way of confirming the quality of the sample or as a method of providing rudimentary analysis.

Cryogenic Electron Microscopy (cryo-EM) is a powerful microscopy method that enables observation of samples in a condition similar their native environment (16). The eponymous sample preparation method reduces the chances that samples will be damaged through factors such as dehydration, interaction with the imaging film or stain, or even damage from the electrons of the microscope (17). Cryo-EM samples are prepared by forming a thin film of non-crystalline amorphous ice around the sample, eliminating any interference from a crystal lattice structure present in more common forms of ice. This thin ice layer also protects against the vacuum conditions of an electron microscope and suspends samples nearly instantly in contrast to the lengthy crystal-growing process needed for x-ray crystallography (18). Due to the flash-freezing technique of creating amorphous ice, the sample particles are also frozen in random orientations in the ice, allowing for analysis to be performed on multiple different angles of the particle without the need to tilt the sample itself (19). In addition, the sample will not degrade as long as it is kept at a suitable temperature, so the same sample can be imaged multiple times, allowing for a large data set to be constructed from a small sample. The random orientations of particles in the sample can be used to construct a 3-D model of a sample through various methods of angular reconstruction, (20) with model resolution for proteins and

other macromolecules reaching resolutions of up to 1 Å (21). Angular reconstruction has been automated via various computational algorithms that prioritize analysis for different forms of symmetry, though analysis for completely asymmetrical particles is also possible (22). This is because the algorithms attempt to find the best probable fit for different relative angles of the particles imaged in the ice layer through density mapping apparent particles into angular classes. These algorithms can be performed on the entire particle or on localized sections to increase the relative resolution of a region.

Cryo-EM methods can also be applied to particles that do not respond well to other, more traditional methods of microscopy (16). Notably, the random angles that the sample freezes in can reveal structural properties that are not discernable during sample preparation that relies on fundamental changes to the native structure of a particle, such as with dehydration due to the vacuum of an electron microscope or the growth of a crystal structure to facilitate x-ray crystallography (23). This allows for the observation of factors like flexibility and relative movement in particles (21). However, a high flexibility or variable position of a structure can negatively influence the overall resolution of a structure, due to the relative reduction in suitably congruent particles of various angles that can be imaged for a 3-D reconstruction (20).

2.3 Sample Preparation of Cryo-EM

There are many variables in the sample preparation phase that can affect the overall quality of a sample and imaging. Cryo-EM samples are prepared on TEM mesh grids made from a conductive metal, often copper or gold, and coated with a perforated carbon film layer (21). These perforations measure from 1-2 micrometers in diameter depending on the exact size of

grid, with standard sizes ranging from 200-400 holes per inch (24). There is a fair amount of interplay between the size and material of the mesh used, with a finer mesh providing more support for the sample but increasing the risk of interference from electrons colliding with the smaller grid of the mesh. A different sized mesh can also affect the thickness of the layer of ice around a sample (19). A thinner layer of ice is more desirable for higher resolution images of a particle, and acceptable ranges for ice thickness include from 100-800 Å, depending on the exact dimensions of the sample being imaged (25). An ideal ice layer is only slightly thicker than the size of the sample to reduce the loss of contrast from electrons being unable to penetrate the ice deeply. However, a thin layer of ice leaves the frozen particles more susceptible to radiation damage from the electron microscope compared to a thicker layer of ice (16). In addition, a layer of ice only slightly thicker than the particle's longest dimension increases the chance that the particle will settle into a preferred orientation and reduce the number of relative angles available for analysis. A thin ice layer also increases the chance that the specimen will be affected by the air-water interface, a hydrophobic layer that may induce denaturization of certain proteins (26). This interface is a smooth, deformable area with relatively high forces that are applied to any particle that comes in contact with it and can cause denaturization or even adsorption of proteins to the interface, meaning that a sample may become fixed at a certain angle relative to the interface. This angle-fixing phenomenon can complicate angular reconstruction methods, especially of samples with low symmetry, since it limits the number of unique angles that can be observed (20). Other factors that may impact the quality of a sample include the amorphous carbon-buffer interface along the grid itself as certain particles may have an affinity for the carbon-buffer interface and take a preferred

orientation along the grid itself, as well as the blotting process, which may damage a sample via the forces applied during blotting or through an interaction with the blotting paper itself. If the sample is particularly susceptible to challenges imposed by the air-water interface and tends to collect near the surface, the blotting paper may remove a majority of sample from the grid and harshly lower the apparent concentration.

Cryo-EM grids are loaded with relatively more protein than other electron microscopy methods. Protein concentration for cryo-EM is often over 1 mg/ml and grids are loaded with 3-5 microliters of sample (24). However, after blotting the grid with filter paper this concentration is often drastically reduced due to previously mentioned interaction with the air-water interface (16). A solution to this is to raise the conductivity and hydrophilia of the foil by adding a second layer of metal to the foil to increase the affinity of a sample to the grid and facilitate the settling of particles in the carbon holes (23). In addition, the sample may be reapplied to the grid and reblotted to maximize the number of particles that remain on the grid after blotting, since particles that settle into the grid will not be removed by the filter paper. Once the sample has been properly loaded onto a grid, it is plunged into a solution of either liquid nitrogen or hydrogen to freeze any water in the sample into an amorphous ice sheet without an interfering crystalline structure (18). The grid must be kept below the devitrification temperature of water, or approximately 140 K, once it has been frozen until after microscopic imaging is complete (17). Once grids have been frozen, they can be kept at appropriate temperatures indefinitely. This allows for subsequent imaging of the same grid, with focus on different areas. This is important since the surface tension of water forms a thicker layer of ice near the edge of the grid, and a thinner layer towards the center. As such, different orientations

and concentrations of the sample can be found at different locations of the grid. This also reduces the need to prepare large numbers of samples, since a single ideal grid can provide an ample number of particles and angles for subsequent analysis and formation of a structure model.

2.4 Software Packages for Analysis/Computational Analysis of Cryo-EM Images

A number of software packages exist to facilitate different methods and algorithms of angular reconstruction. The most common programs including EMAN/EMAN2, which is a program that focuses on analyzing post-acquisitional standard transmission electron data but is often applied to cryo-EM analysis, especially as a preliminary micrograph adjustment tool (27). Another popular program is RELION, a more specialized product that introduced a Bayesian approach to determine the most likely relative alignment of the angles seen in imaging (16). RELION and programs that employ similar algorithms such as CryoSPARC have been optimized for ease of use and streamlined analysis, enabling operators to quickly obtain and optimize a 3-D structure at high resolutions (28). In comparison, EMAN2 and comparable software packages have many specialized algorithms that are fully customizable via Python coding and the command-line interface as a tool to manually alter the algorithms used for analysis. EMAN2 also supports a wider array of file types than either RELION and CryoSPARC, and is supported by most operating systems, whereas RELION and CryoSPARC are only capable of running on Linux-based systems.

2.5 Data Preprocessing and Particle Picking

The first step in analysis of cryo-EM samples is to correct the micrographs obtained from imaging a sample for factors such as astigmatism of a particular microscope, contrast transfer

function (CTF), and defocus (16). In addition to a general defocus correction, micrographs must also be corrected for per-particle defocus, which is caused by particles in the sample being frozen at different thicknesses in the ice layer (22). Particles that are frozen deeper in the ice layer are subject to a lower signal-to-noise ratio which lowers the apparent resolution of the particle compared to others in the same micrograph. CTF is dependent on the interaction of the electrons from the microscope with the sample, meaning that the CTF factor is unique for each individual micrograph obtained (29). CTF distortion occurs as a Fourier transform of the virtual image of the electron beam as it passes through the microscope lens (12). This results in a micrograph image that is slightly distorted proportional to a point spread function, which can be applied due to the relatively small point source of the electron beam of a microscope. The point spread function is the diffraction pattern formed as a result of the differing distances the electrons must travel to strike different areas of the sample, with the diffraction pattern appearing as a series of concentric rings of constructive and destructive interference called Thon rings. Thon rings of cryo-EM micrographs are seen in figure 2.

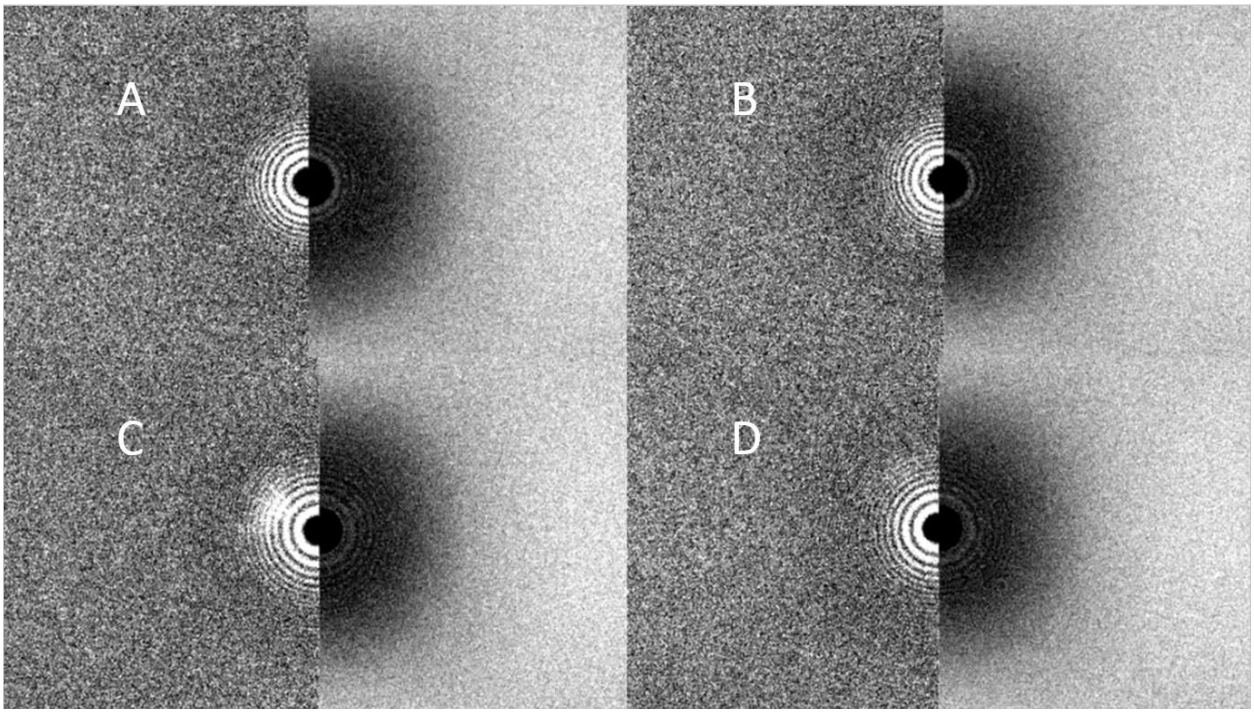

Fig. 2. Thon rings for four different cryo-EM micrographs, each with a different defocus and CTF correction value.

The CTF of a micrograph can be corrected for by considering the relative angle and thickness of the Thon rings and extrapolated to obtain and correct for the defocus of a micrograph as well (30). At higher resolutions, Thon rings become more frequent and lower in contrast, making it more difficult to resolve the defocus, blur, and CTF correction factors. CTF can be estimated using the equation

$$CTF(k) = \sin(\pi\lambda(\delta_0 + \delta_A \cos(2(\phi_k - \phi_A)))|k|^2 + \frac{\pi}{2} C_s \lambda^3 |k|^4 - \chi)$$

which is applied to produce a unique CTF factor for each micrograph, where λ is the wavelength of the electrons, δ_0 the average defocus, δ_A half the difference between the two effective defoci resulting from astigmatism, ϕ_k is the azimuthal angle of k, ϕ_A is the major astigmatism

axis, C_s is the spherical aberration constant of the microscope, and χ is the overall phase shift resulting from amplitude contrast and any phase plates used in imaging (22). The formation of a virtual blurred image over the standard micrograph is important for increasing resolution of a sample as it can be used as a pseudo-secondary image to remove noise from a micrograph sample (31). Applying CTF correction to a micrograph removes the effects of the point spread function by undistorting the virtual image and merging it with the actual micrograph. CTF correction is a mandatory step in single particle analysis since CTF distortion occurs for each micrograph and deforms the apparent shape of any particles present in an uncorrected micrograph image.

Each software package currently available for cryo-EM processing possesses an automated particle-picking functionality (22, 27, 28). Automated picking in programs such as RELION does not require a reference template created from user-picked particles and instead uses a Laplacian-of-Gaussian based filter to differentiate particles from noise interference in a micrograph. This filter takes into account the size of the targeted particle, multiplying the Fourier transform of the micrograph by

$$\frac{|k|^2}{\sigma^2} e^{\frac{|k|^2}{-2\sigma^2}}$$

where k is the image frequency, $\sigma = \frac{2}{d}$ where *d* is the diameter of the desired particle, which normalizes for a noise differentiation filter of σ^2. This method of identifying particles based on their diameter allows for upwards of hundreds of thousands of particles to be identified relatively quickly.

Another option for both particle picking and denoise correction is utilizing machine learning to train a neural network to optimize both processes for the unique parameters of each sample (32). This also corrects for more traditional denoise methods that may incorrectly identify small or globular proteins as noise interference rather than particles, or particle picking algorithms identifying noise as particles that must be manually removed or filtered out from the particle sample. Using a deep-trained neural network such as Topaz can speed preliminary refinement of new micrographs of a sample, as well as obtain a higher contrast denoised micrograph that one that would be resolved using a generalized denoising algorithm. Topaz also includes pre-trained denoising neural networks that have been optimized for denoising micrographs obtained from select microscopes, allowing for visually higher contrast between particles and ice without the need to dedicate time and data to training the neural network for a specific set of data.

2.6 Angular Reconstruction

The angular reconstruction of the structure of particles is performed using computational analysis, though initial data processing requires user input to begin identification and classification of particles (16). Under ideal conditions, a sample will be frozen in a random assortment of orientations, allowing for a 3-dimensional model to be reconstructed once the orientations are ordered relative to one another. Using software packages like CryoSPARC, extracted particles are ordered automatically into angle groups known as classes using Gaussian best-fit functions (28). For highly symmetric particles, automated sorting requires little user input beyond initiating the sorting process, but more irregular particles often require multiple phases of resorting. This stage of data processing, called 2-D classification, is also used to

identify and remove false-positives and low-quality particles that were automatically picked. Once a selection of 2-D classifications has been optimized, they can be ordered into a 3-dimensional volume using a combination of stochastic gradient descent and branch-and-bound maximum likelihood algorithms to create a 3-D volume map. These algorithms check for probable overlap between 2-D classifications using Bayesian statistical analysis to determine relative orientations of the classifications. Because of this, it is optimal to obtain the highest resolution possible during 2-D classification, often to the point where secondary structures are visible in each classification. However, due to the stochastic approximations present in the algorithms, highly symmetrical particles may be improperly reconstructed as the 3-D reconstruction combines two nearly identical but locally disparate 2-D classes into a single relative orientation, inadvertently removing the symmetry of the particle in favor of a single, nonsymmetric structure.

3-D reconstruction can be used as a secondary optimization technique when creating multiple simultaneous volume maps from the same data set. Since the data will be split into separately created volumes, each different random combination of classes and their respective particle sets will produce a different volume. In the case of an ideal set of data where the particle has taken no preferred orientation during imaging and has a suitably varied number of relative angles, these 3-D volumes will be very similar. 3-D reconstruction can also be performed with an initial volume map included as a template for the reconstruction to follow. This template is used to match the 2-D classifications to various orientations of the volume with gaussian best-fit functions. Because of the templating function of CryoSPARC and other software packages, a small, optimized set of data can be used to create a low resolution 3-D

volume, with that volume being used as a template for a higher resolution 3-D volume created using a much larger set of less optimal data. In addition, templates for 3-D reconstruction can be created from other data sets of the same particle, such as a relatively low-resolution volume obtained with negatively stained data being used as a template for the angular reconstruction of cryo-EM data. These volume maps can be rendered in 3-D using programs such as Chimera (33). Chimera allows not only for the visualization of 3-D datasets, but also for the analysis and masking of volume maps (34). The masking of a 3-dimensional area of the volume allows for post processing of the area to locally increase resolution in a number of ways (35), including by altering the fit of a particular structure within the particle itself, and analyzing the fit compatibility between substructures.

3. Materials and Methods

3.1 Sample Preparation

Samples of fibrinogen were obtained from Peak 1 fibrinogen produced at Enzyme Research Labs in South Bend, Indiana. These samples were also analyzed using western blot gel electrophoresis to confirm the presence of all three chains of fibrinogen, since their molecular weights are all distinct from one another and show up in distinct bands on a polyacrylamide gel.

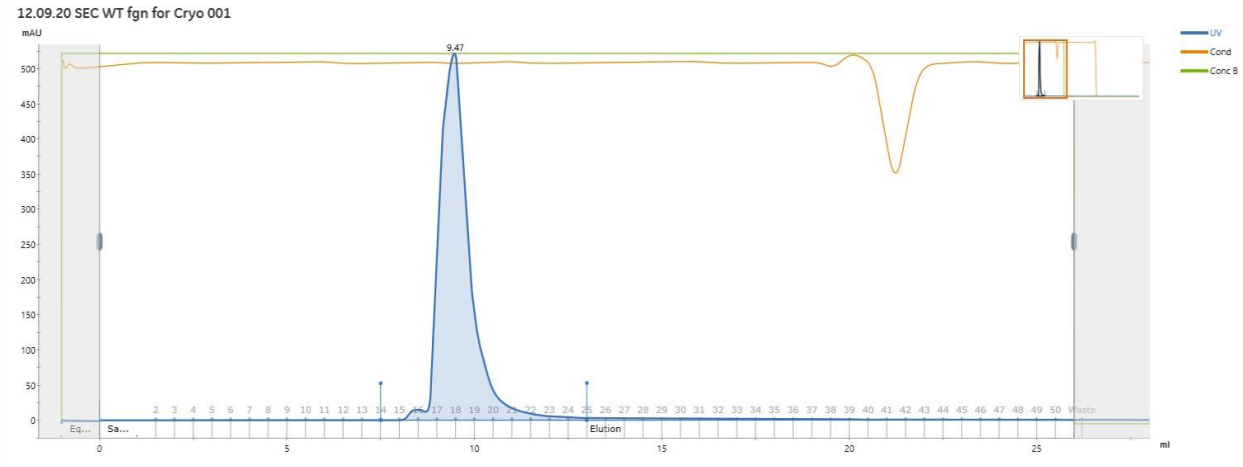

Fig. 3. FPLC chromatogram of Fibrinogen, showing a high concentration of fibrinogen in microcentrifuge tubes 18 and 19. Tube 18 was used for subsequent analysis.

Fast protein liquid chromatography (FPLC) was performed to remove any aggregate fibrinogen from the desired single fibrinogen molecules. The sample was exchanged into a low salt buffer of 10 mM NaCl + 20 mM HEPES that had been degassed and filtered in preparation for the procedure using a spin column. A Cytivia Äkta pure chromatography system and attached Superdex increase flow columns were used to fractionally elute fibrinogen into microcentrifuge tubes at a flow rate of 0.2 mL/min, with the resultant chromatogram seen in figure 3. The accepted curve was found to be strongest in fractional tube #18, and this sample was subsequently found to have a fibrinogen concentration of 0.895 mg/ML using a Thermo Scientific™ NanoDrop™ 2000, and aliquoted to concentrations of 0.8 mg/mL for both subsequent analysis and sample stock.

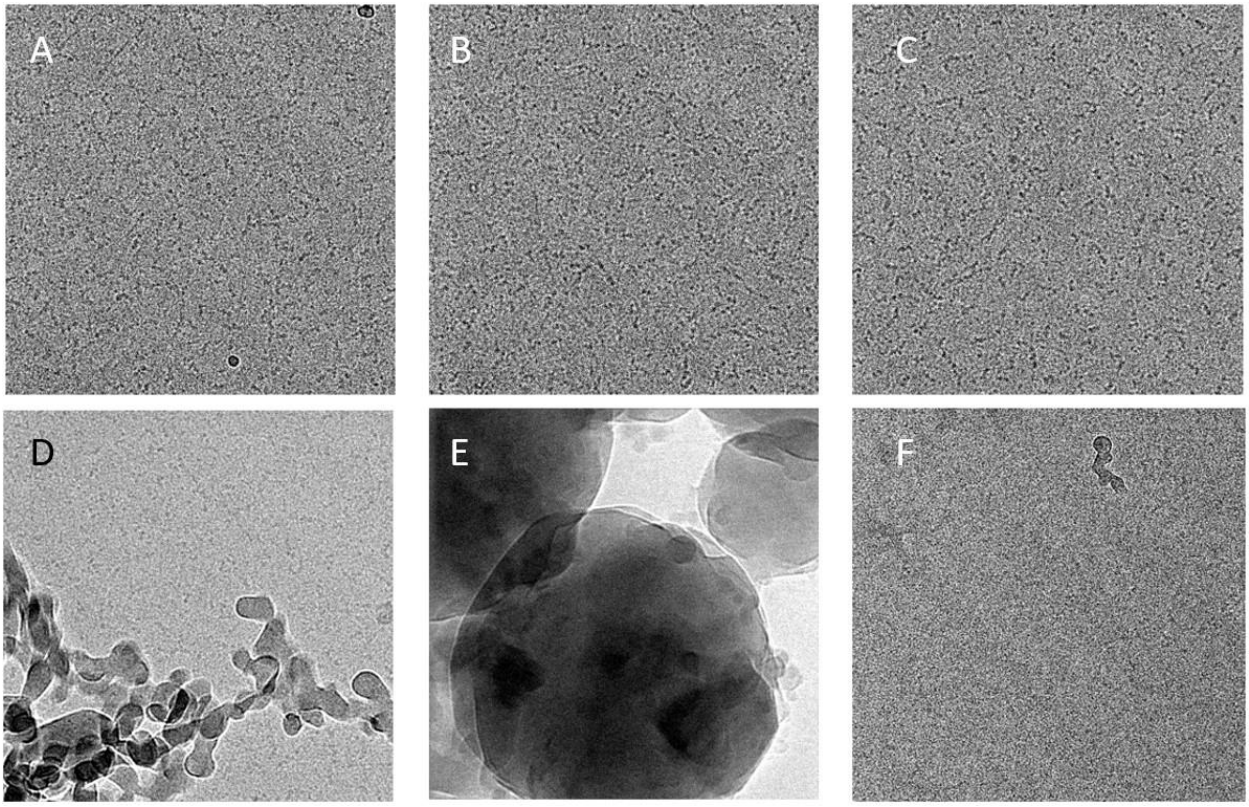

Fig. 4. Examples of good and bad micrographs obtained using cryo-EM. Note the particle separation and contrast on micrographs A, B, and C. Significant ice contamination is present in micrographs D and E, with low particle contrast present in micrographs D and F. All six micrographs show visible CTF lines. This CTF factor was corrected for later in data processing.

3.2 Cryo-EM Grid Preparation

Fibrinogen samples were prepared for cryo-EM using a size 200 gold mesh. 3 µL of sample was pipetted onto the mesh and blotted on filter paper to remove excess buffer. The sample was then flash frozen by plunging the grid into liquid nitrogen for two seconds and transported and stored in liquid nitrogen when not actively being imaged. These grids were imaged using the ThermoFischer Talos Artica Cryo-TEM microscope at the Molecular

Microscopy Consortium based in the National Institute of Environmental Health Sciences. A small initial set of micrographs were collected and visually analyzed to ensure that protein density was ideal for subsequent computational refinement and angular reconstruction, and that there was no crystallization of the ice layer or vertical overlap between fibrinogen particles. The samples were briefly analyzed at multiple locations to check for imperfections in the grid and the prevalence of any preferred orientations of fibrinogen. Sample concentrations ranging from 0.1 mg/mL – 4 mg/mL were tested. These initial micrograph images to confirm sample quality were not used for analysis. A final protein concentration of 0.1 mg/mL resulted in the most even distribution of well-behaved particles. Subsequent imaging was performed overnight and resulted in the collection of 1655 micrographs, a selection of which can be seen in figure 4.

Subsequent samples have been prepared to counteract the interaction between the air-water interface and fibrinogen, with negative stain samples prepared and imaged to test the quality of the improved preparation procedures. These samples of fibrinogen were prepared using the same techniques as listed above, with aliquots prepared with protein concentrations of 0.8 mg/mL for use as a stock for subsequent cryo-EM imaging.

3.3 Negative Stain Grid Preparation

A protein concentration of 20 µg/ml was also prepared for negative stain imaging to test the quality of the fibrinogen sample before overnight collection of cryo-EM data as well as for general negative stain analysis. The samples for negative stain microscopy were prepared using a size 200 carbon-gold mesh of standard thickness. After being cleaned by a glow-discharge device to create a hydrophilic surface on the grid mesh for the sample to adhere to, 6 µg of the

negative stain sample were pipetted onto the grid and blotted off using filter paper after waiting 30 seconds to allow the protein to settle into the grid holes. The blotted paper was washed with 1.5 % uranyl formate, blotted immediately, and washed a second time in 1.5 % uranyl formate. The grid was blotted again after 30 seconds and washed with buffer from the sample solution and immediately blotted once more on filter paper.

4. Results

4.1 Initial Cryo-EM Analysis of Fibrinogen

These micrographs were imported into CryoSPARC and preprocessed using included algorithms for motion correction and CTF estimation. Identification of fibrinogen particles was performed both manually and automatically. Manual picking was preformed by visually identifying distinct fibrinogen molecules, such as the one seen in figure 5. Automatic picking was performed both based on a manually created template based on the particles visually discernable with no micrograph filtering, as well as with a blob picker that differentiates particles from background noise present in the micrograph based on relative density values. In all cases, particles were picked with a circular box size with a diameter of 512 Å, which was chosen as it is both an optimal integer number for computational processing speed as well as being approximately 1.1 times the size of fibrinogen, large enough for an entire particle to fit inside the box diameter. A low-pass filter was manually applied to automatic particle picking algorithms to reduce the impact of a low signal-to-noise ratio, lowering the number of false positives based on noise in the micrograph.

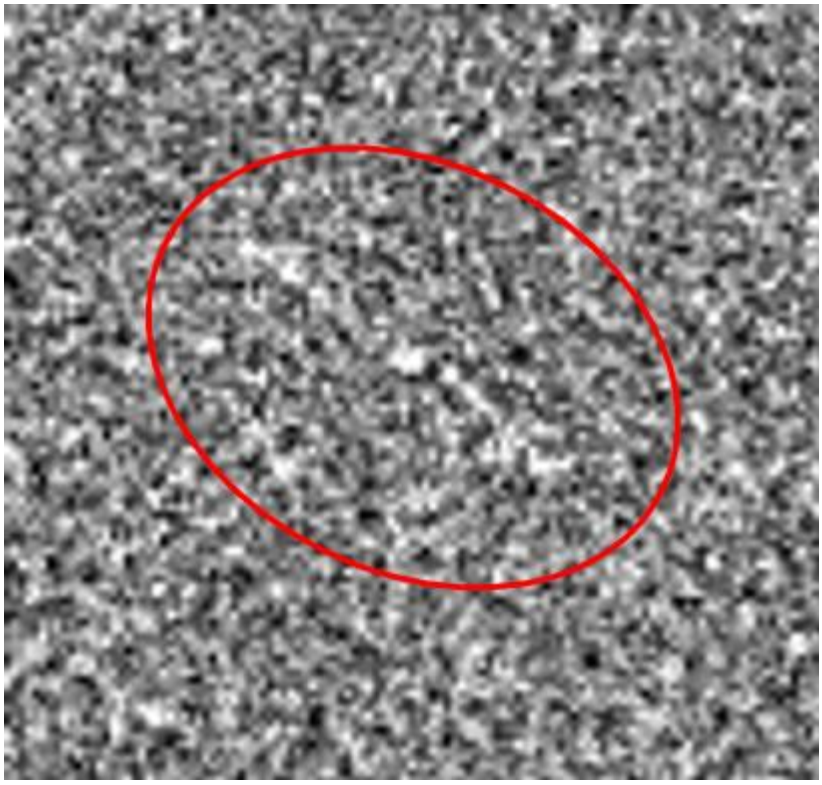

Fig. 5. An example of fibrinogen visually distinct in a micrograph. This particle was imaged using negative staining.

Approximately 65,000 fibrinogen particles were identified across the micrographs as appropriate for further analysis. These 65,000 particles were ordered into one hundred initial 2-D classes, seen in figure 6. These automatically produced classes were manually sorted for quality, including visually apparent structure and overall resolution of the classes. Classes that were comprised of false positives that were ignored by the applied low-pass filter, classes that contained two or more particles, and classes that demonstrated unusually unique flexibility were removed. The process of 2-D classification and manually inspecting classes was repeated until high-quality set of particles remained from the initial selection, with approximately 50 iterations of 2-D classifications being performed. It is notable that a majority of the particle

classes that were produced during 2-D classification showed a distinctly two node structure, in contrast to all existing publications which detail the known structure of fibrinogen as a three node structure. An initial render of a three node fibrinogen structure is shown in figure 7. In addition, reintroducing the two node particles in even small quantities of approximately 10% of the overall number of particles caused any volumes to tend towards a two node structure, shown in figure 8. The resolutions of 2-D classifications were found to range from approximately 10-20 Å for all classes.

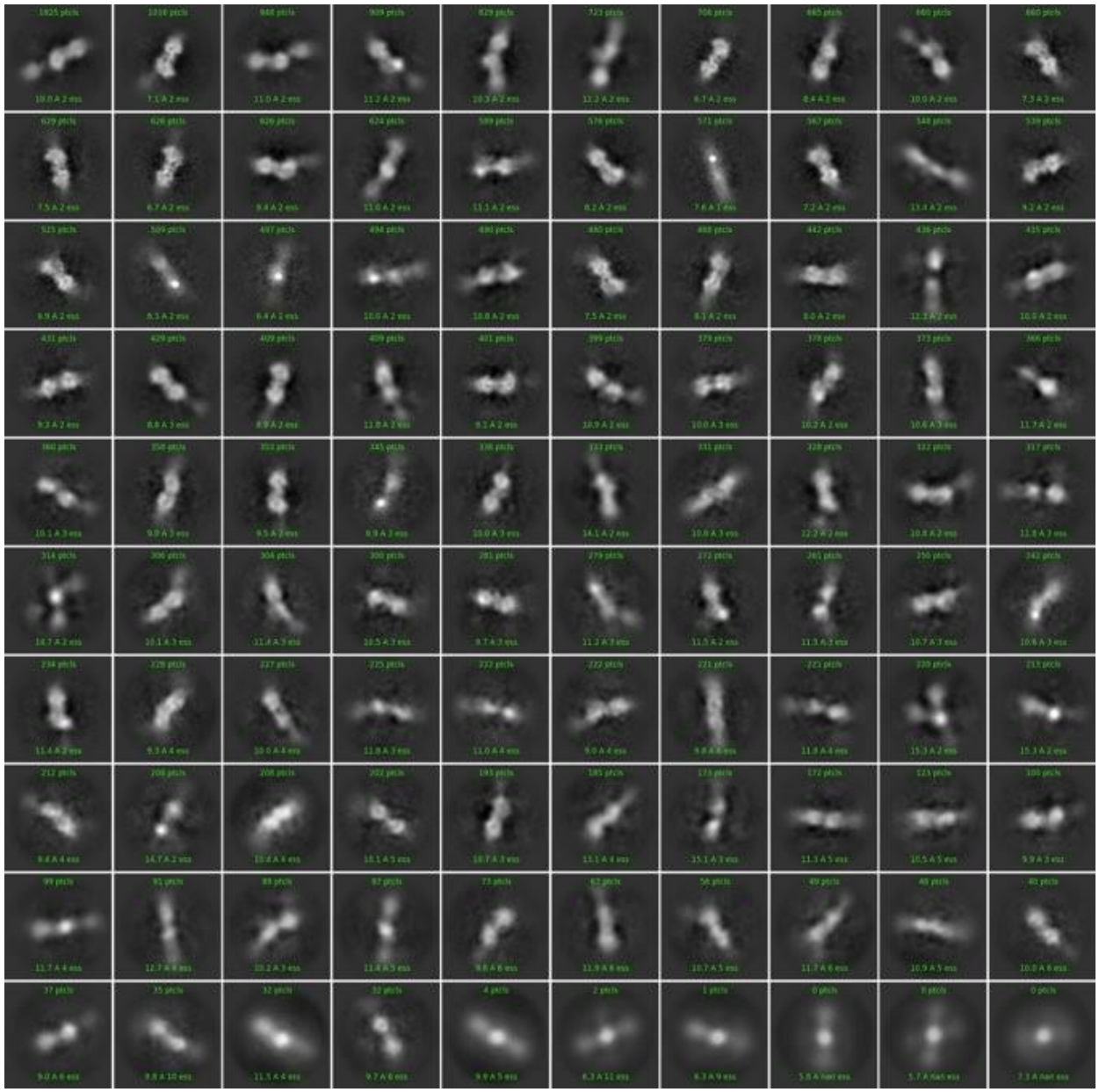

Fig. 6. Initial 2-D classes created with CryoSPARC using a selection of approximately 35,000 particles. Most classes appear to have two nodes, rather than the three node structure expected of fibrinogen. Secondary structures of the particles become apparent when the class resolution approaches approximately 10 Å, with the overall resolution of the classes ranging from 7-16 Å. These classes are organized in order of decreasing particles per class. Note that the number of particles in each class does not directly correlate with the class resolution.

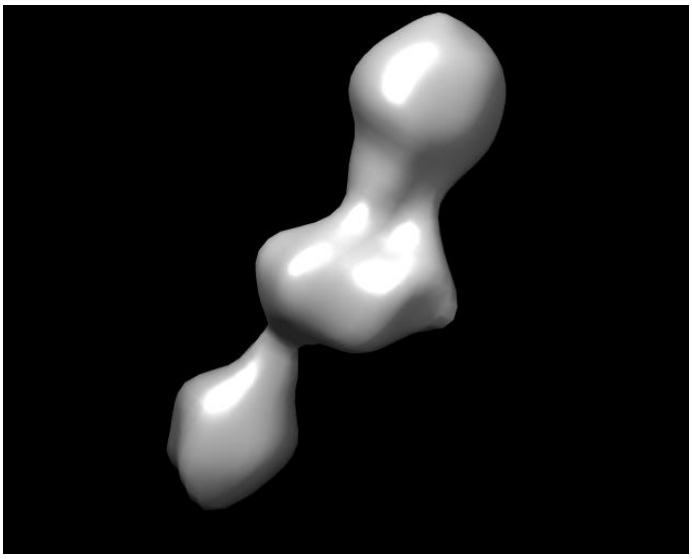

Fig. 7. An initial volume render created using the classes in Fig. 6. This volume was created using CryoSPARC and cryo-EM data.

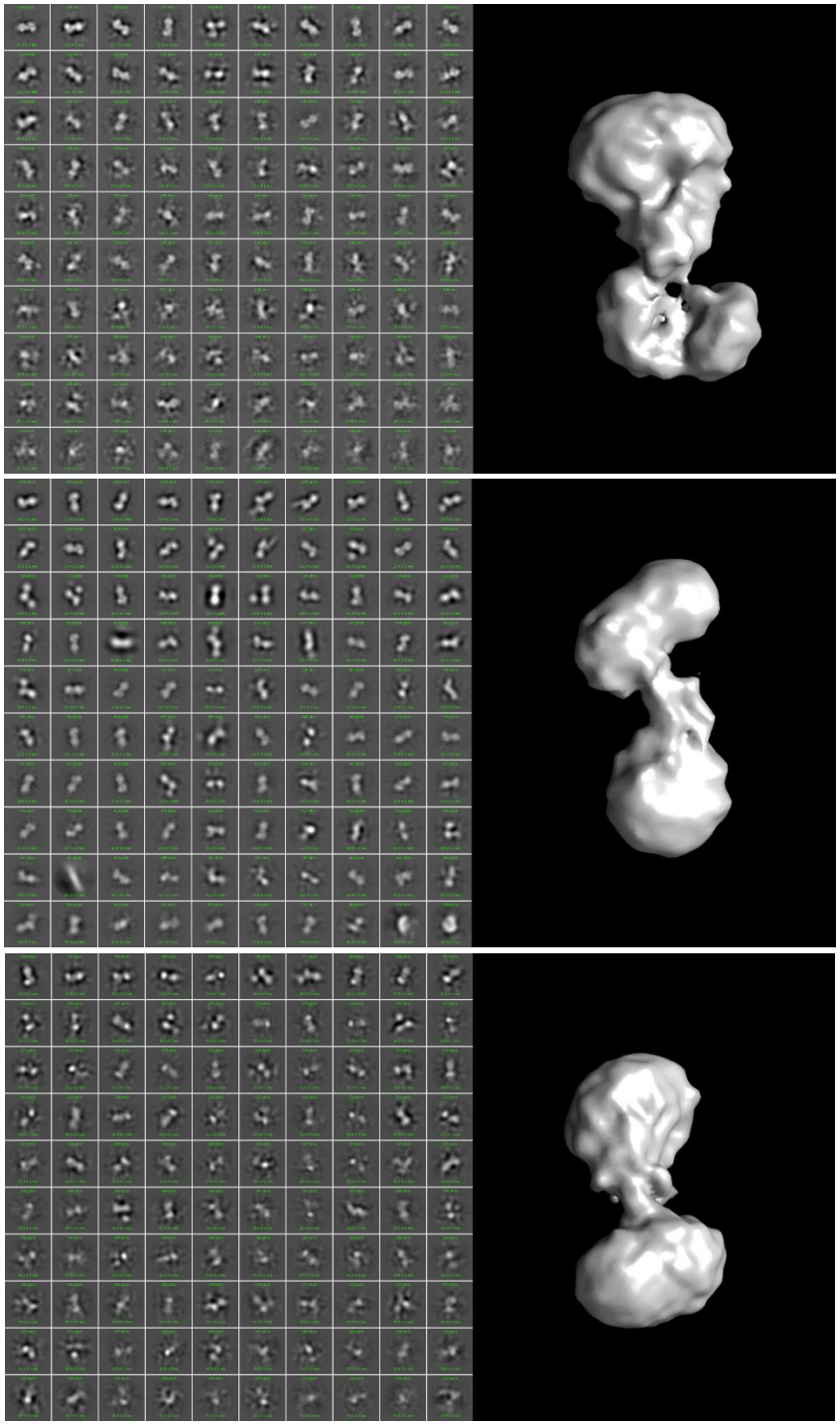

Fig. 8. 2-D classes and resulting 3-D volume maps of fibrinogen. All three volumes tend towards a two node structure. Both the 2-D classes and 3-D volume maps were created using CryoSPARC. The particles in these maps were randomly split into three volumes from the classes in fig. 6.

Initial results, including 2-D class averages of various angles of the entire fibrinogen molecule were inconclusive. When initial volumes were formed from angular reconstruction and visualized using the software package Chimera (33), the structure of fibrinogen tended towards the two-node structures visible in figure 8 rather than the three-node structure expected from previous methods of imaging fibrinogen. In addition, complications arise from the flexibility of the fibrinogen molecule, which can cause it to appear in multiple configurations that are not compatible for angular reconstruction. This is because a bend at any possible hinge site has the chance of appearing to a 2-D classification algorithm as a separate orientation, rather than slightly hinged.

The flexible behavior of fibrinogen can artificially lower the sample size as multiple hinge angles can cause secondary structures or even entire regions to average out during 2-D classification. An attempted solution for this issue is to only use particles that show little visible flexibility to create a standard linear structure template for subsequent local refinement, with particles of this type used to create the classes seen in figure 9. However, the number of fibrinogen particles appropriate for this type of analysis was too low to produce any meaningful results. 2-D classification was performed to select a data set of approximately 3,000 visibly three node fibrinogen molecules from which a template 3-D volume was created for refinement of a larger data set using the nonlinear fibrinogen. However, due to the relatively small percentage of three node fibrinogen compared to visually disparate fibrinogen present in the sample, the template volume was unsuccessful as an algorithmic guide to create a higher-resolution render of fibrinogen structure.

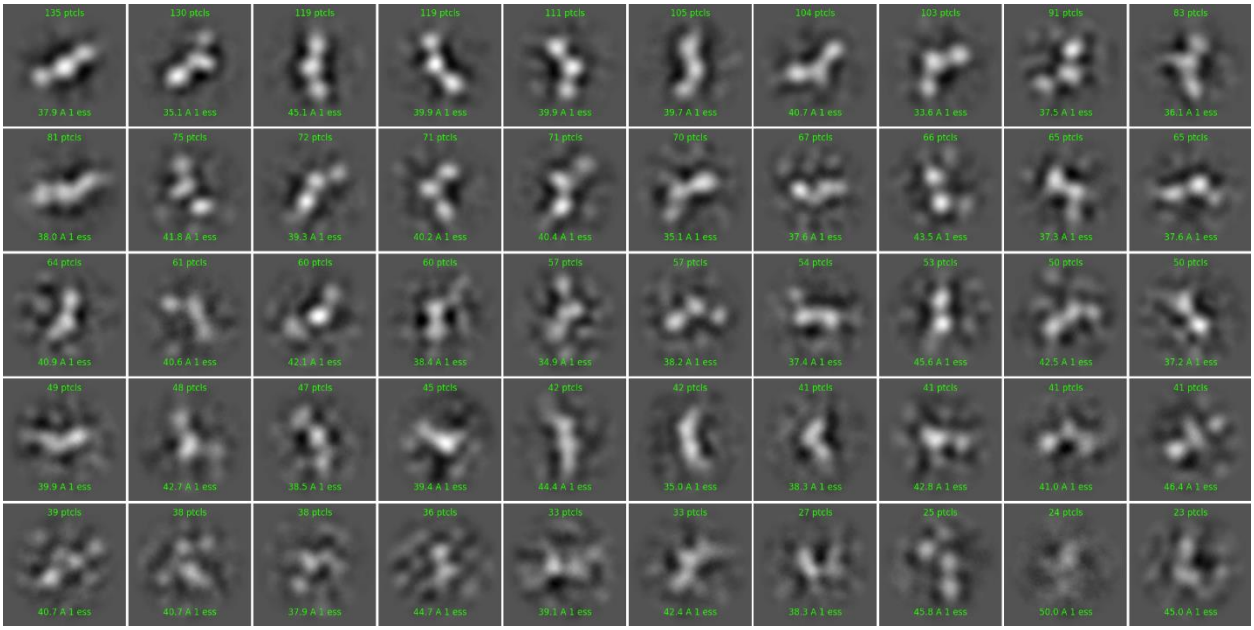

Fig. 9. 2-D classes created in CryoSPARC by selecting for only particles that are visually similar to the known structure of fibrinogen. These classes were created using a subset of slightly more than 3,000 particles. The classes are organized in order of decreasing particles per class and have a resolution range of 35-50 Å as a result of the small dataset.

In addition, the higher-than-average number of two-node structures can be contributed to the possibility of a preferred vertical or partially vertical orientation of fibrinogen in the samples, with only one D-region visible at the surface of the ice sheet and the other being too deep within the ice layer to properly image. The exact thickness of the amorphous ice layer surrounding the sample is unknown, but the range of possible ice thicknesses ranges from 100-800 Å (25); fibrinogen is approximately 460 Å from end to end. To reduce the chances of fibrinogen taking this possible preferred orientation, subsequent cryo-EM grids will be prepared to minimize the thickness of the ice layer around the samples. The effect of the air-water interface during sample preparation for cryo-EM may also cause fibrinogen to denature along

the central E region into two symmetrical half-molecules, leaving few intact fibrinogen particles with the majority of 2-D classes showing a two-node structure of the central E region and the distal D region, rather than the expected three-node structure of two D regions on either end of the central E region. When measured in EMAN2, both the 2-D classes and the volume renders were between 100 and 200 Å in legnth, less than half of the expected legnth of a fibrinogen molecule. In addition, the visually intact fibrinogen particles were often arranged with different hinge angles along the central E region. Because of this, 2-D classifications from different relative angles could not be used to reconstruct a 3-D volume as the angles of flexibility were incompatible.

4.2 Negative Stain EM Analysis of Fibrinogen

The grids prepared for negative stain microscopy were imaged, producing 150 micrographs. These negatively stained micrographs were preprocessed for CTF alignment corrections using CryoSPARC, along with other preprocessing procedures specific to analysis of negatively stained micrographs in CryoSPARC, including an inversion of selection processes, as particles in negatively stained data appear as unpigmented regions, rather than as pigmented particles as occurs with cryo-EM data due to the high molecular weight and density of the uranyl formate stain compared to fibrinogen. This is due to the mass-thickness contrast effect of electron beam scattering present in electron microscopy. Initial automatic particle picking was performed in CryoSPARC to determine the approximate number of fibrinogen molecules per micrograph.

Subsequent picking used for data analysis was performed with manual and automatic particle picking occurred in EMAN2, which does not require the inversion of negative stained

data. EMAN2 algorithms are only able to determine light particles on a dark background, which is the type of contrast that negative staining produces; cryo-EM and other types of imaging require data to be visually inverted before accurate analysis can be completed. To account for the high number of possible arrangements fibrinogen can be oriented in when factoring in possible flexible regions, particle picking focused only on the D region of fibrinogen, specifically the β and γ C-terminus regions. This region appears in the micrographs as a distinct pair set of nodes at a constant separation regardless of the relative angle or flexibility of the respective particle. The β and γ C-terminus regions in the D region of fibrinogen are also nearly identical in known structure, with any averaging between the two regions having little to no effect on the final created volume or classes due to this similarity. All subsequent analysis using both negative stain and cryo-EM data focused on reconstructing a high-resolution volume map of the D region of fibrinogen. This is because of the clarity of the β and γ nodes seen in the micrographs and classes, as well as its relatively distinct structure compared to areas that appeared to have been averaged out during computational analysis, such as hinging in the E region. The high resolution D region volume will also be used as a reference for future structural analysis such as the orientation of the coiled-coil region.

A sample of approximately one hundred visually ideal particles were manually picked from micrographs with a variety of defocus and CTF correction values and identified as optimal templates for the auto picking algorithm included in EMAN2. These manually picked particles prioritized distinct β and γ nodes in the D region, and had a box size of 96 Å. The auto picker picked for the remainder of the micrographs, as well as for particles that were not visually discernable on the manually picked micrographs, producing a sample size of over 89,000

particles. Due to the large number of visibly discernable false positives identified by the auto picker, the auto picker was subsequently run again with a smaller box size due to a large number of false positive particles. This second auto picking pass used the same manually identified set of particles as the first pass with a smaller box size of 64 Å, and over 39,000 particles were identified for single particle analysis. The effects on the algorithm caused by these two different box sizes is shown in figure 10. These particles were averaged into initial 2-D classes, seen in figure 11, which were refined over multiple class averaging iterations to remove false positives and low-quality particles that were not optimal for analysis.

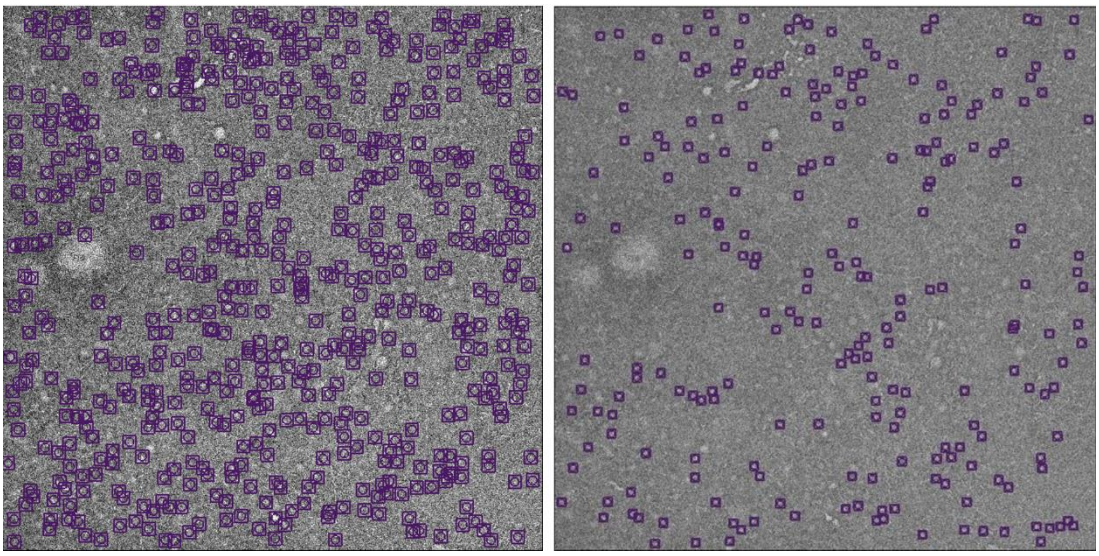

Fig 10. A micrograph that has been automatically picked with two different box sizes. The left image shows a box size of 96 Å, with approximately 600 particles picked per micrograph. The right image has been picked with a box size of 64 Å, with approximately 260 particles picked per micrograph. Notably, both micrographs show pairing of particle locations; there are two of the targeted D regions on each fibrinogen molecule.

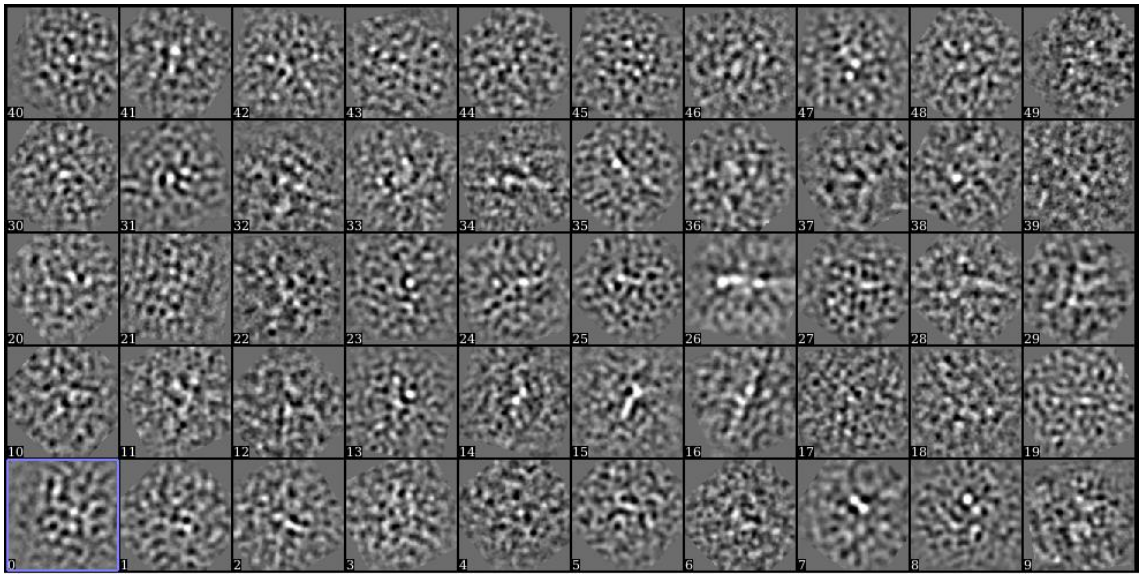

Fig 11. Initial classes found using approximately 39000 particles. These classes are inconclusive, with a high signal-to-noise ratio. The size of the D region (60 Å) is near the limit of negative stain imaging (19).

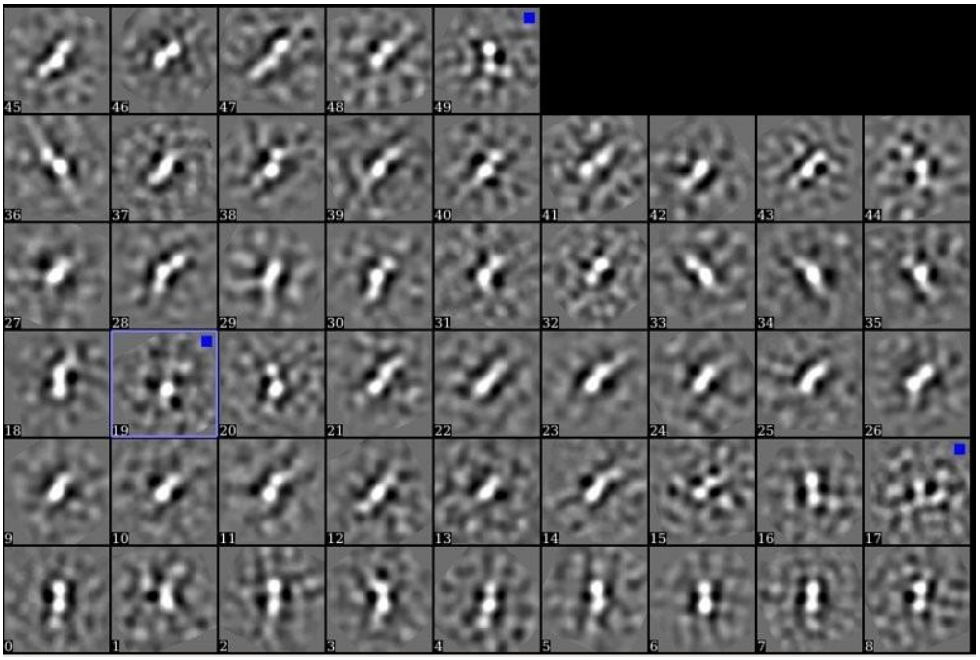

Fig. 12. 2-D classes after multiple rounds of set refinement. A two node structure is visible in multiple classes, which is expected for the nodular β and γ C-terminus regions. These 2-D classes were created in EMAN2 using negative stain data, and subsequently used as templates for cryo-EM particle picking.

These final classes shown in figure 12 were used to angularly reconstruct an initial 3-D volume. This initial volume was used as a template for subsequent refinements of the volume using angular reconstruction techniques, ultimately resulting in a final volume with a resolution of 16.8 Å. Both volumes are shown in figure 13. This final negative stain volume was compared with the known crystallographic structure of fibrinogen as seen in figure 14, and its respective 2-D classes were imported into CryoSPARC to use in analysis of cryo-EM data. This high correlation between the angularly reconstructed volume and the known crystal structure confirms that the D region was intact and able to be accurately reconstructed using negative stain analysis. The 2-D classes collected using the negative stain data were used as templates for automatic particle picking of the D region using existing cryo-EM data in order to increase the resolution of the volume. A sample size of approximately 50,000 D region particles was collected from micrographs imaged using cryo-EM and used for 2-D classification and volume reconstruction.

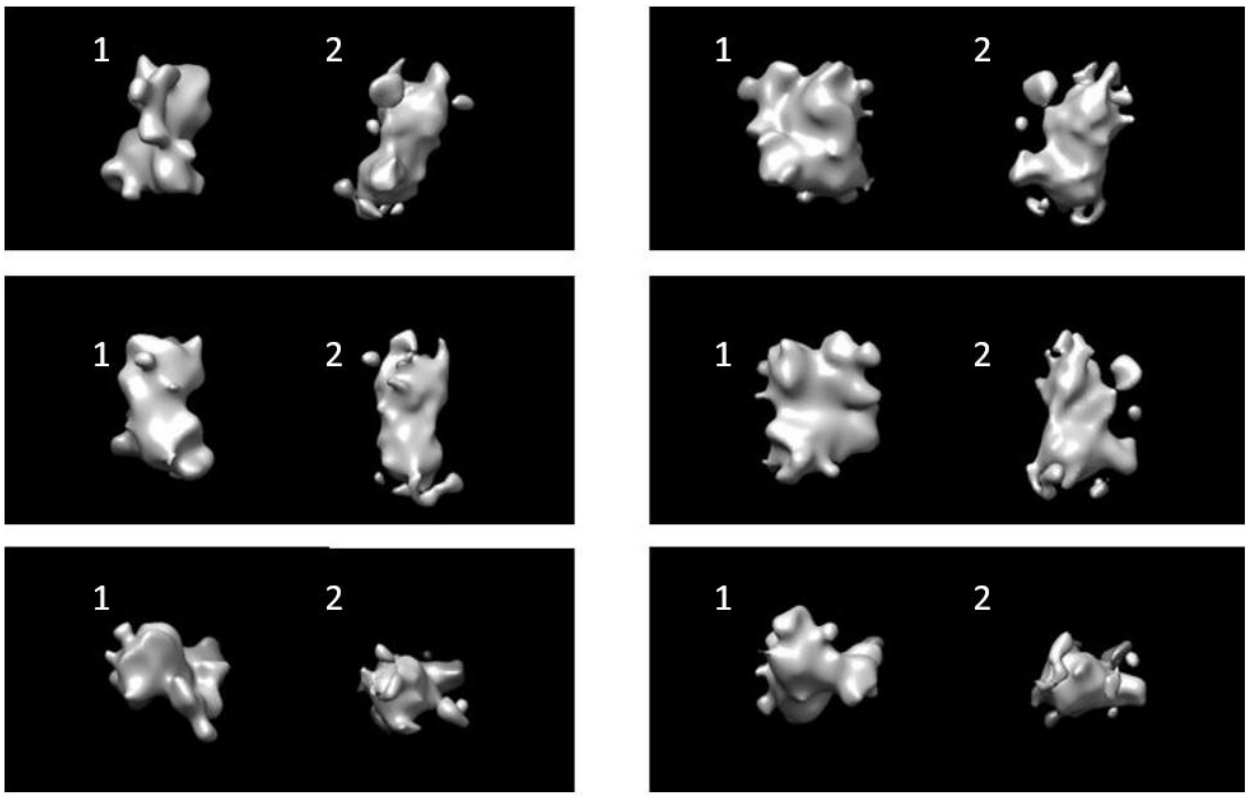

Fig. 13. Initial 3-D volume (volume 1) of the D region of fibrinogen compared to a refined 3-D volume (volume 2). In all comparisons, the volumes are at the same relative angle. The initial volume has a resolution of 21.7 Å compared to 16.8 Å of the refined volume. The refined volume shows greater definition between globular β and γ regions, as well as a defined coiled-coil structure protruding from the volume. These volumes were created using negative stain data.

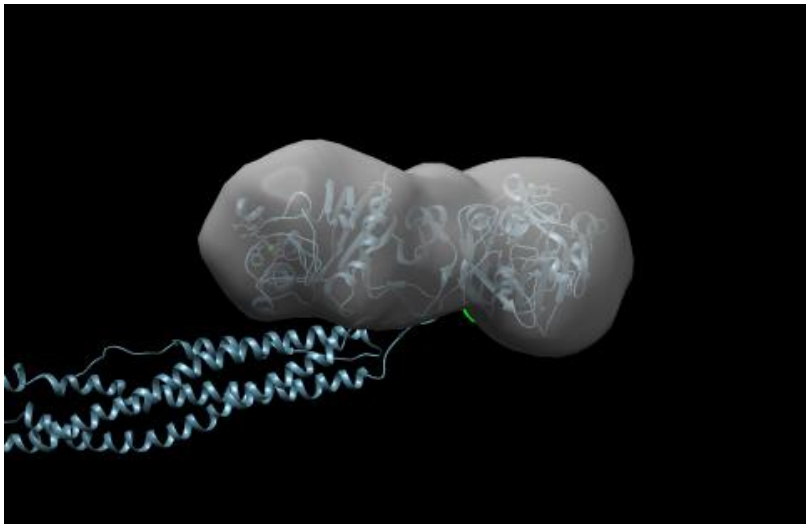

Fig. 14. D region volume created using negative stain data and EMAN2 overlaid with crystal structure. This volume has an estimated resolution of approximately 17 Å.

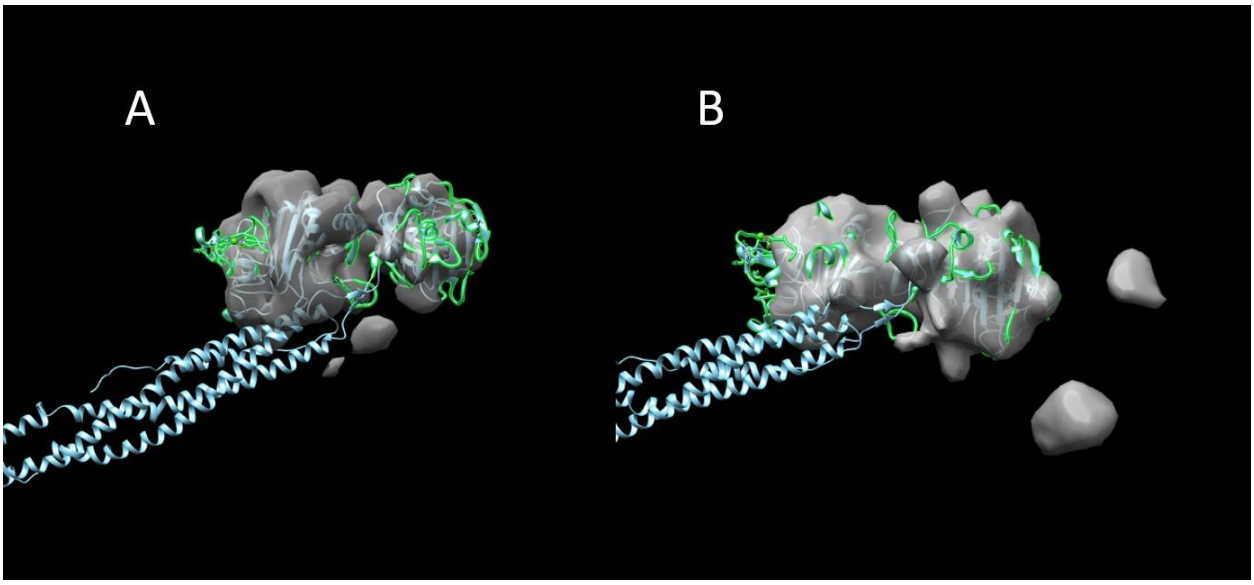

Fig. 15. An initial (A) and homogenous refined (B) volume of the D region, overlaid with the known crystal structure of fibrinogen. Both volumes were created with cryo-EM data in CryoSPARC using the same randomly assigned 2-D class subset containing approximately 12,000 particles.

4.2 Resuming Cryo-EM Analysis

Multiple initial volumes were created using CryoSPARC using ab initio reconstruction, where the best fit algorithms attempt to recreate a most likely structure based on the fit overlap between 2-D classes. By default, CryoSPARC will create one volume using every particle available, but when more than one volume is created in the same ab initio reconstruction job using pre-job input parameters, CryoSPARC will automatically and randomly separate the 2-D classes into each resultant volume. This random assignment of 2-D classes can lead to tendencies for certain types of classes to form various structures based on the percentage of 2-D classes that seem to average to those features. Because of this, 3-D volume construction can act as another particle refinement step where 3-D features reveal false positives or poor-quality particles that were not visible in 2-D classification. Volumes were then visually examined and compared to the existing known crystallographic structure of the D region of fibrinogen. The volumes with the highest fit correlation are then homogenously refined using the particle sets they were randomly assigned during ab initio reconstruction; no new particles are introduced to volume construction at this stage. This initial refinement results in the construction of a volume with a higher resolution since the best fit algorithms now have a 3-D volume template to use as a reference for angular reconstruction, rather than fitting 2-D classes in a relatively random order, with the results of this refinement algorithm on the d region shown in figure 15. This allows for more precision in the placements of 2-D classes and a higher resolution in the created volume. Homogenous refinement does not create new structures in the volume, instead it attempts to create a better orientation match between the particles and existing structures.

These refined volumes were compared to the existing crystal structure of fibrinogen, and subsequently masked for local refinement in order to increase the resolution of the volume. This masking primarily involved removing the floating globules on the periphery of the main D region structure, as can be seen in figure 16. These extant regions of the volume were removed using built-in tools included in Chimera, specifically the segger and volume eraser tools. The resulting volume map was imported into CryoSPARC as a masked template for localized refinement of the D region. Localized refinement was performed with the full set of approximately 49,000 picked particles, producing a volume with higher resolution and an improved crystal-structure fit, which is shown in figure 17.

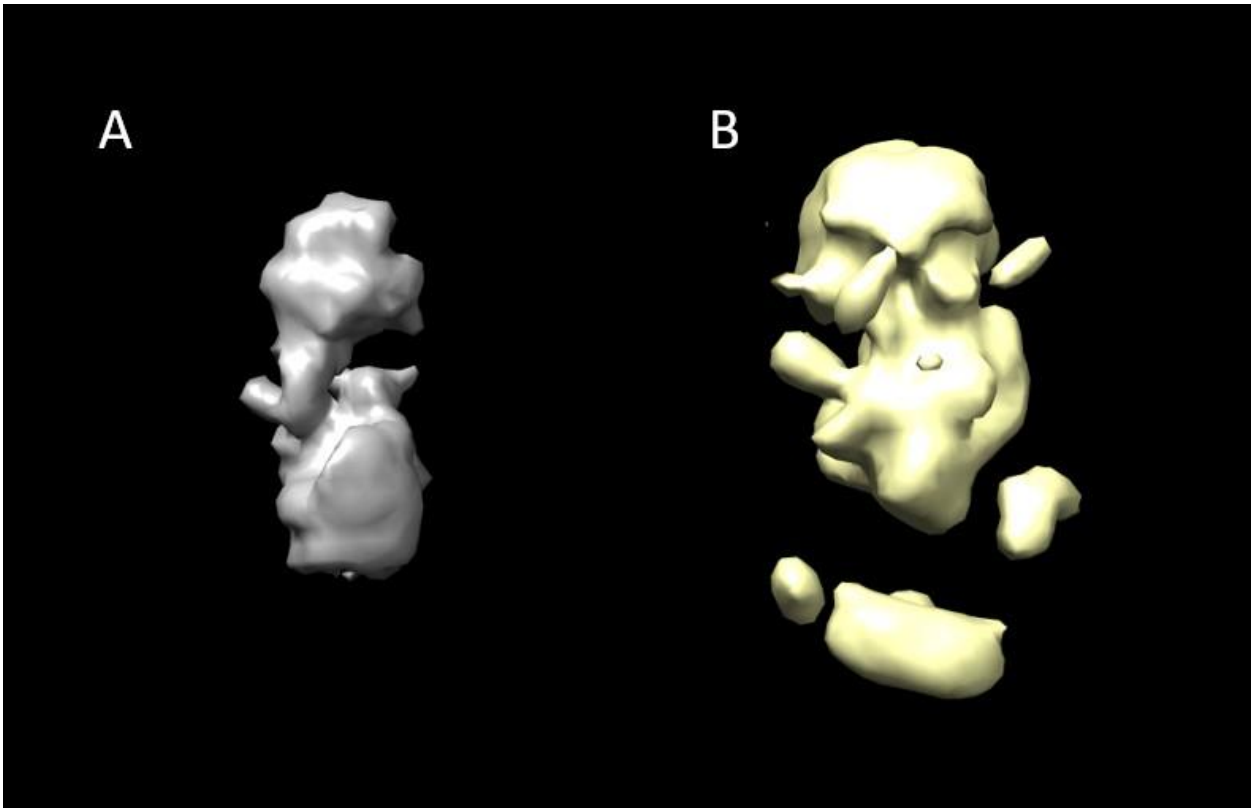

Fig. 16. 3-D volume of the D region after masking (left) and after initial refinement (right). The initial volume is an alternate angle of the refined volume seen in fig. 15. The two volumes are at the same relative angle.

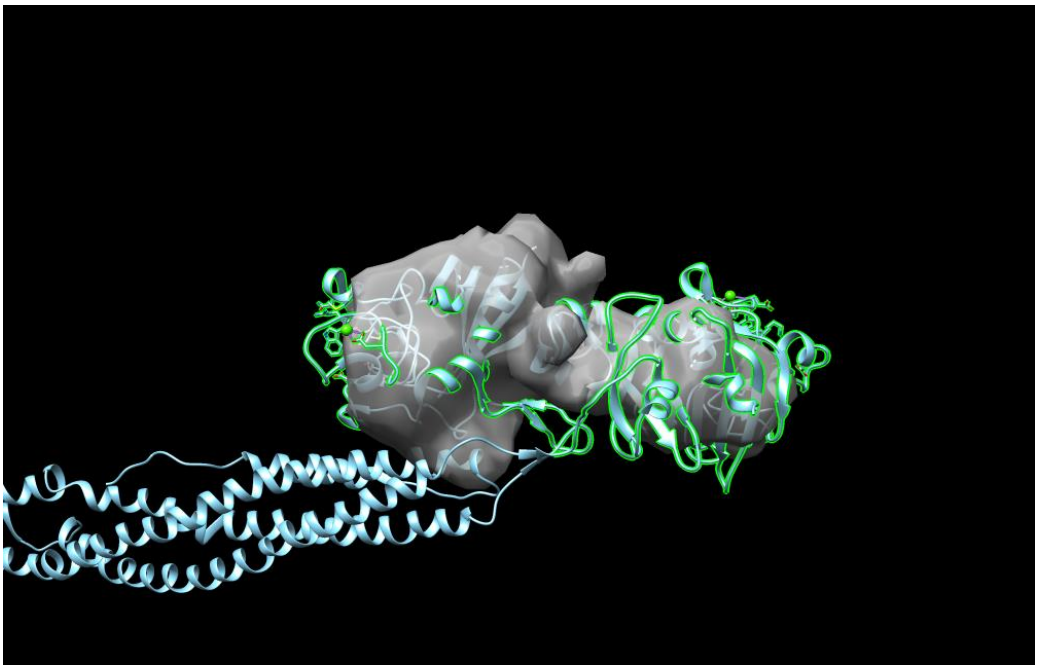

Fig. 17. Mask refined volume of D region of fibrinogen overlaid with the known crystal structure.

5. Conclusion and Future Directions

Fibrinogen is a large protein that shows flexibility at multiple sites along its structure, a property that challenges structural analysis by both traditional methods like crystallography and novel methods like cryo-EM. In cryo-EM analysis of fibrinogen, initial micrographs of the protein show multiple angles of flexibility, lowering the effective sample size of particles available for angular reconstruction using cryo-EM software packages. Currently, single particle analysis of a whole fibrinogen molecule is pending the collection of more cryo-EM data, with initial results being inconclusive. The D region of fibrinogen is a globular region that distinctly appears on both negative stain and cryo-EM micrographs as a two node structure with relatively consistent distance between the β- and γ- C-terminus regions. This stable appearance as well as the proposed hinge sites between the D region and the coiled-coil region and

between the β- and γ- chains make it a strong candidate for structural analysis using both negative stain and cryo-EM techniques. Negative stain analysis was performed on the D region, resulting in 2-D classes and initial volume structures of the region. These classes and volumes were used as templates for subsequent cryo-EM analysis, which has resulted in the volumes seen above in figure 17. Future data collection will occur to increase the number of particles available for structural analysis of both the D region and the entire fibrinogen molecule.

The D region volumes are currently undergoing further refinement with an aim of creating a high-resolution volume map structure of the region. This volume map will be used to determine possible hinge sites in the D region, including between the β and γ C-terminus regions and between the D region and the coiled-coil region. Other areas of the fibrinogen molecule that are candidates for a similar cryo-EM analysis approach of targeting an area of interest include the central E region of fibrinogen. In addition, improvement of sample preparation can lead to a higher percentage of intact fibrinogen available for angular reconstruction of a full and complete cryo-EM structure of fibrinogen.

6. Acknowledgements

I would like to thank Dr. Nathan Hudson for his advice and guidance throughout this project, which would never have begun without him. I would also like to acknowledge the NIEHS Cryo-EM facility, Dr. Mario Borgnia and Dr. Venkata Dandey. This work would not have been possible without funding from NIH R15HL150666 and the ECU Undergraduate Research and Creative Activity (URCA) award. Finally, I thank East Carolina University and the Honors College for continued support.

7. References

[1] J. W. Weisel, FIBRINOGEN AND FIBRIN.

[2] S. Herrick, O. Blanc-Brude, A. Gray, and G. Laurent, The international journal of biochemistry & cell biology 31, 741 (1999).

[3] J. W. Weisel and R. I. Litvinov, Fibrin Formation, Structure and Properties (Springer International Publishing, Cham, 2017), 82, p. 405.

[4] J. M. Kollman, L. Pandi, M. R. Sawaya, M. Riley, and R. F. Doolittle, Biochemistry (Easton) 48, 3877 (2009).

[5] S. Kattula, J. Byrnes, and A. Wolberg, Arteriosclerosis, thrombosis, and vascular biology 37, e13 (2017).

[6] C. Cohen, G. N. Phillips, and J. W. Weisel, Nature (London) 289, 263 (1981).

[7] Lawrence Weissbach and Gerd Grieninger, Proceedings of the National Academy of Sciences - PNAS 87, 5198 (1990).

[8] B. A. Cottrell, M. Riley, K. W. K. Watt, R. F. Doolittle, and D. D. Strong, Nature (London) 280, 464 (1979).

[9] Z. Yang, J. M. Kollman, L. Pandi, and R. F. Doolittle, Biochemistry (Easton) 40, 12515 (2001).

[10] R. F. Doolittle, D. M. Goldbaum, and L. R. Doolittle, Journal of molecular biology 120, 311 (1978).

[11] W. J. G. Schielen, H. P. H. M. Adams, K. v. Leuven, M. Voskuilen, G. I. Tesser, and W. Nieuwenhuizen, Blood 77, 2169 (1991).

[12] Wim J. G. Schielen, Marijke Voskuilen, Godefridus I. Tesser, and Willem Nieuwenhuizen, Proceedings of the National Academy of Sciences - PNAS 86, 8951 (1989).

[13] B. Fultz and J. M. Howe, Transmission Electron Microscopy and Diffractometry of Materials (Springer Berlin / Heidelberg, Berlin, Heidelberg, 2013).

[14] R. H. Wade, Ultramicroscopy 46, 145 (1992).

[15] S. De Carlo and J. R. Harris, Micron (Oxford, England : 1993) 42, 117 (2011).

[16] Y. Cheng, N. Grigorieff, P. A. Penczek, and T. Walz, Cell (Cambridge) 161, 438 (2015).

[17] J. Lepault, A. W. McDowall, M. Adrian, and J. Dubochet, Nature (London) 308, 32 (1984).

[18] J. Dubochet and E. Knapek, PLoS biology 16 (2018).

[19] M. Ohi, Y. Li, Y. Cheng, and T. Walz, Biol. Proced. Online 6, 23 (2004).

[20] M. Van Heel, Ultramicroscopy 21, 111 (1987).

[21] D. Bhella, Biophys Rev 11, 515 (2019).

[22] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, Nature methods 14, 290 (2017).

[23] J. R. Meyerson, P. Rao, J. Kumar, S. Chittori, S. Banerjee, J. Pierson, M. L. Mayer, and S. Subramaniam, Scientific reports 4, 7084 (2014).

[24] R. F. Thompson, M. Walker, C. A. Siebert, S. P. Muench, and N. A. Ranson, Methods (San Diego, Calif.) **100**, 3 (2016).

[25] L. A. Passmore and C. J. Russo, Methods in enzymology 579, 51 (2016).

[26] E. A. Vogler, Biomaterials 33, 1201 (2011).

[27] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, Journal of structural biology 157, 38 (2007).

[28] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, Nature methods 14, 290 (2017).

[29] J. A. Mindell and N. Grigorieff, Journal of structural biology 142, 334 (2003).

[30] R. H. Wade, Ultramicroscopy 46, 145 (1992).

[31] K. Ahi and M. Anwar, in (SPIE, Apr 29, 2016), p. 98560N.

[32] T. Bepler, K. Kelley, A. J. Noble, and B. Berger, Nature communications 11, 5208 (2020).

[33] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, Journal of computational chemistry 25, 1605 (2004).

[34] T. D. Goddard, C. C. Huang, and T. E. Ferrin, Journal of structural biology 157, 281 (2007).

[35] Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, T. D. Goddard, E. C. Meng, A. Sali, and T. E. Ferrin, Journal of structural biology 179, 269 (2012).