Coarse-grained Simulations of Myosin-VII Inhibition using Fluctuating Finite Element Analysis

Robert Welch

Abstract

The inhibition of the motor protein Myosin-VII occurs when the head domain binds to the tail domain. A coarsely-grained method in the continuum limit, that includes thermal fluctuations, was used to simulate the motion of Myosin-VII for the long ($> 1\mu s$) timescales this process requires. It has been suggested that Myosin-VII is able to bind only in the presence of the inhibitors ETGA and Ca^{2+}, which allow calmodulins to unbind from the lever domain, decreasing the rigidity of the molecule. A Python module was developed for analysing the end-to-end distance, angular distribution and twist. Once the molecule had been converted to a coarse-grained model and simulated under a variety of conditions, the results of the simulation confirmed that Myosin-VII is unable to bind in its uninhibited state, and established a relationship between the average end-to-end distance and the molecule's material parameters.

I. Introduction

I. Myosin-VII

Myosin 7 is a motor protein commonly found in the inner ear and retina. It performs functions relating to the development and maintenance of stereocilia, the sensory hair cells in the inner ear, and the transport of melanosomes (pigment sacs) in the pigmented layer of the retina.

Although the structure of Myosin-VII is well understood, the exact conformation of the folded state is still unknown.

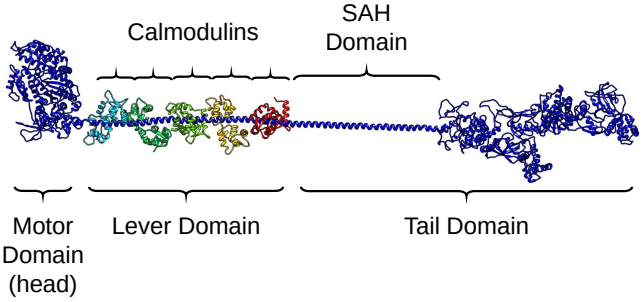

Figure 1: Atomistic data for Myosin-VII

A molecule of Myosin-VII consists of four main domains (regions) - a motor domain (the head) connected to a globular domain (the tail) by a long stable alpha-helix (SAH domain) and a lever, which consists of a number of IQ domains (binding sites), bound to calmodulins (messenger proteins).

Recent evidence from electron microscopy suggests that the molecule can 'curl up' in the presence of inhibitors such as Ca^{2+}, and that the head and tail bind to one another, allowing the molecule to remain in that 'folded' position [1], [2].

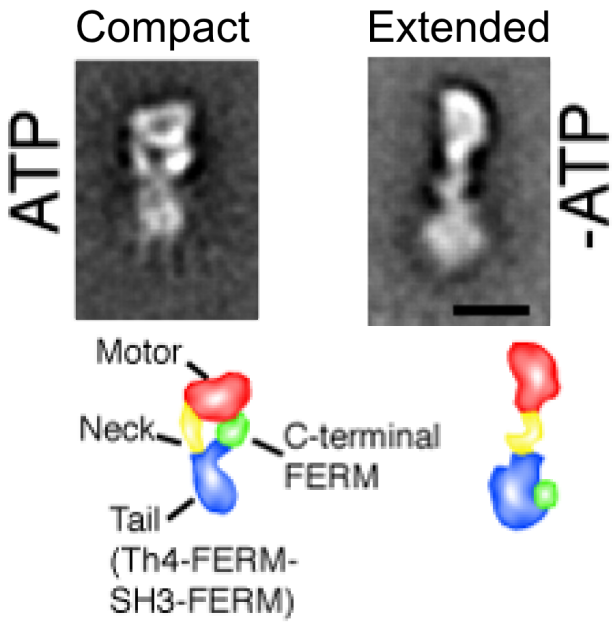

Figure 2: Electron microscopy image showing inhibited (left) and uninhibited (right) Myosin-VII [1]

Unfortunately, there are no atomistic structures available for the molecule's folded state. A simulation of Myosin-VII could provide a more precise idea of how the protein folds, and the material parameters needed for it to fold like this. Unfortunately, this folding could require timescales of order $1\mu s$ [3] or more. Although it is possible to simulate Myosin-VII using

atomistic molecular dynamics, the detail of second-order structures is unnecessary for this application, and a continuum model lends itself naturally to this type of motion.

II. FFEA

Biological simulations are often performed using atomistic molecular dynamics. In this type of simulation, molecules are defined as a number of atoms modelled as if they are connected by harmonic springs. The interactions between these atoms are modelled classically, by the electrostatic and Van Der Waals interactions. The potentials of these interactions are given by quantum mechanical calculations.

The naive implementation of this method would yield an algorithm of order $O(N^2)$, as every particle would need to exchange information with every other particle. In reality, the electrostatics are modeled with the particle mesh ewald (PME) method, and the range of the Van Der Waals force is hard-limited, resulting in an algorithm of order $O(N \log N)$ in the best-case scenario [4].

Nevertheless, this method requires supercomputers running for weeks in order to visualise larger, more complex molecules such as the Myosin family at timescales of less than 1 nanosecond.[1]

The unsuitability of MD simulations to proteins such as Myosin-VII has spurred the development of methods that are less computationally expensive, and are designed to operate on larger timescales, with larger, less-detailed systems. In the same way that atomistic simulations are parameterised by quantum-mechanical calculations, these methods can be parameterised by atomistic simulations and experimental data.

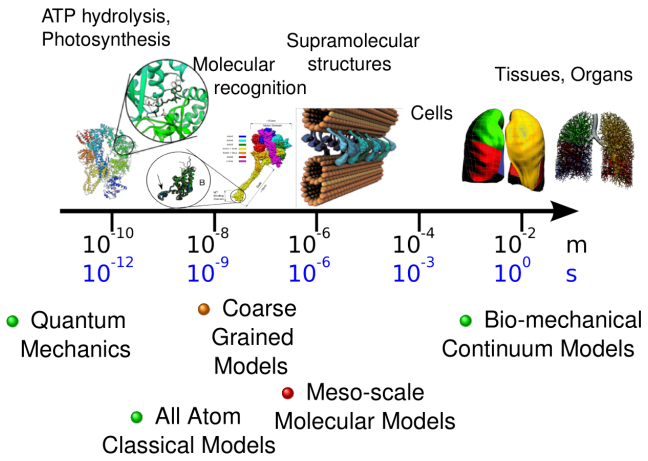

Figure 3: Relative scales of biological processes and the simulations used to model them. Atomistic simulations are designed with smaller structures in mind. There is some overlap between large atomistic structures and small coarse-grained structures (used by FFEA), where Myosin-VII lies. Although atomistic and continuum models have reached a high level of development, mesoscale simulations are relatively unexplored. Diagram by Albert Solernou\Sarah Harris.

The tool being developed at the University of Leeds uses a technique called 'Fluctuating Finite Element Analysis' (FFEA) to accomplish this [5] [6]. In FFEA, the atomistic data is converted into a much simpler 3-D volume comprised of tetrahedrons. The equations of motion are solved by the conjugate gradient method, which is of O(N) complexity in the best-case scenario[2].

The motion of a molecule in FFEA can be described by the Cauchy momentum equation for continuum mechanics:

$$\rho \frac{D\boldsymbol{u}}{Dt} = \nabla(\boldsymbol{\sigma}^v + \boldsymbol{\sigma}^e) + \boldsymbol{f} + \nabla\boldsymbol{\pi}$$

The molecule is subject to viscous stress (resistance to motion in the fluid medium) and elastic stress (a restoring force toward the molecule's original position). This is referred to as the Kelvin-Voight model of the solid.

Crucially, it is also subject to a thermal noise term (π). A good way to think of the motion of molecules under FFEA is in terms of the fluctuation-dissipation theorem. At small scales, dissipative processes, such as the conversion of kinetic energy to heat, work in reverse. The energy supplied by the thermal noise can push the head and tail close together and allow them to bind - if given enough time.

[1] The Adaptive Fast Multipole Method, which is currently under development for gravitational systems, promises these types of simulations in O(N) complexity, but still requires an N many orders of magnitude larger than is necessary for Myosin-VII

[2] For simulating single molecules that do not interact with themselves.

The Langevin equation:

$$\boldsymbol{M}\frac{d\boldsymbol{u}}{dt} = -\boldsymbol{K}\boldsymbol{u} - \boldsymbol{E} + \boldsymbol{N}$$

Is a weak formulation of the Cauchy momentum equation, based on the assumption that the solution is a linear differential equation. This equation is solved using finite element analysis.

The simulation software and the surrounding toolchain are still works-in-progress, but even at this early stage, they are mature enough to generate real, usable data.

To give an idea of the relative speeds of the two approaches: Myosin-VII is approximately 48,000 atoms in size, whereas a coarse-grained model may be as small as 1000 nodes. Using the AMBER MD package on 16 supercomputer cores for 2 weeks yields a simulation lasting 600 picoseconds. The same molecule on just four cores or a regular workstation PC using FFEA can yield a simulation lasting 600 nanoseconds, a 1000x speedup using a fraction of the computing resources.

II. Methodology

I. Model Generation and Parameterisation

The Protein Data Bank (PDB) is a free archive of structural information for biological molecules that can be readily downloaded and used in atomistic simulations. The data is acquired mostly from x-ray diffraction experiments, and is stored in the PDB format. Before information can be used, it must be converted into an FFEA-compatible 'coarse-grained' format. The conversion of atomistic data into these coarse-grained volumes is as much an art as it is a science, and the way they are converted can have a significant effect on the behaviour of the molecule being simulated.

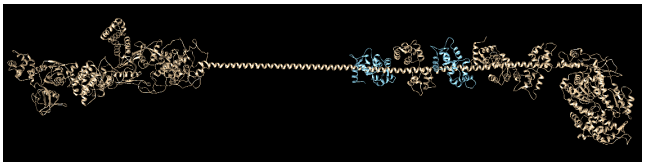

Figure 4: Atomistic form of Myosin-VII in UCSF Chimera visualisation package.

First, the structural information is opened in UCSF Chimera [7], a visualisation tool. UCSF Chimera is used to convert the PDB file (containing the positions, types and bonds of atoms) to an electron density map (.map) with ISO level 3.8. The electron density map is a visual representation of the probability of electrons being at a particular position. This gives the molecule a homogeneous blob-like appearance.

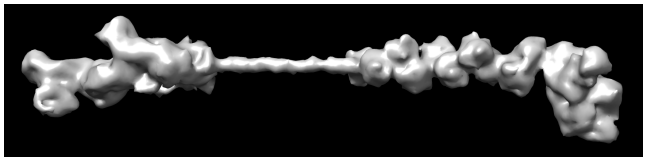

Figure 5: Electron density map of Myosin-VII in UCSF Chimera.

From there, the electron density map is smoothed using UCSF Chimera's built-in interpolation tools, with 2 smoothing iterations of factor 0.6, and converted to an .obj file. An .obj file is a 3D mesh representing the object's surface, so any details of internal structure are lost. This .obj file is then reduced in complexity, from many thousands of triangles to around 1000, by the FFEA surface coarse grainer tool. The granularity of the mesh can be varied, and areas less important to the motion are given less detail. The granularity of the meshes was generally between 6 and 21Å per node.

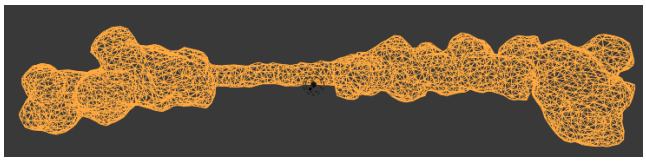

Figure 6: Myosin-VII Surface Mesh in Blender.

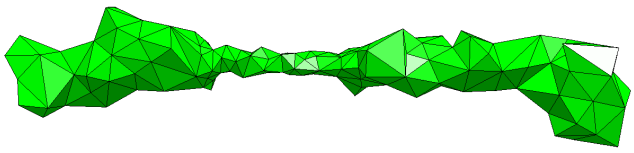

Figure 7: Coarse-grained model of Myosin-VII in Netgen. Note the smaller triangles in the SAH domain.

The internal structure of the new .obj file is restored using Netgen [8]. Netgen converts a triangular mesh to a 3-dimensional volume of tetrahedrons. This is the final model to be used by the FFEA simulation.

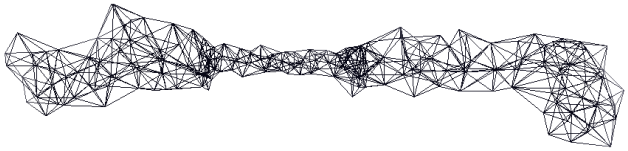

Figure 8: Myosin-VII molecule the in FFEA viewer with internal tetrahedrons shown.

Finally, the FFEA tools are used to generate an FFEA startup script, along with a number of other files, which store information associated with the material

and simulation parameters. These include the Stokes viscosity of the medium ($1 \times 10^{-3} Pa \cdot s$, later reduced to $1 \times 10^{-5} Pa \cdot s$ for inhibited simulations, see below), timestep ($1 \cdot 10^{-14} s$), and amount of thermal noise (a thermodynamic temperature of $kT = 4.11 \cdot 10^{-21} J$). The material parameters can be varied on a per-node basis, or for arbitrary volumes.

There is no definite method to parameterise any empirical model, and Myosin-VII is no exception. The parameterisation of Myosin-7 can be viewed as a balancing act between stability, speed and accuracy. The naive assumption is that the resolution of the mesh determines the accuracy of its simulation, and that a highly detailed mesh is best - but diminishing returns come from adding extra detail. Additionally, the smaller tetrahedrons are more likely to 'invert' (be turned inside-out). An inverted tetrahedron will cause the simulation to halt, as the Euler method is used to compute solutions to the Langevin equation, and a negative volume will result in undefined behaviour. For this reason, the Myosin-VII meshes were kept as coarse as possible, whilst still preserving the primary structure.

A systematic search of parameter space was undertaken in order to determine a set of parameters that give the best agreement with the experimental behaviour of the motor. These parameters are given in the appendix (fig. 30).

II. Model and Simulation Variations

In total, Myosin-VII was simulated 14 different times, with 7 unique model variations and 12 parameter variations. The first 7 simulations represent an attempt at testing the model generation toolchain and assessing the limitations of the FFEA simulation.

After that, the model was parameterised using physical data from the similar Myosin-V molecule as a guideline, as no data for Myosin-VII was available. The only material parameter available was bending stiffness [9], which was estimated from the stall force of the lever arm. The bending stiffness is related to the Young's modulus by:

$$B = EI$$

Where B is the stiffness, E is the Young's modulus, and I is the moment of inertia. I was approximated to be that of a rod of constant radius a:

$$I = \frac{\pi a^4}{4}$$

Finally, the value of Young's Modulus E is related to the actual material parameters used by FFEA, the shear and bulk moduli G and K, by:

$$G = \frac{E}{2(1+\nu)} \text{ and } K = \frac{E}{3(1-2\nu)}$$

Where ν is Poisson's ratio, approximately 0.35 in this case [6].

The experimental data suggested a Young's modulus of $382 MPa$ in the SAH domain and $827 MPa$ in the lever domain, and a simulation was run using these parameters.

Later, this was combined with experimental evidence to suggest that the SAH domain alone was too rigid to perform the power stroke, and that it may require a combination of the SAH and lever domain [10]. In the presence of Myosin-VII inhibitors, such as ETGA and Ca^{2+}, some of the calmodulins may unbind from the IQ domains, exposing the unstable alpha-helix within the lever domain, and significantly reducing the stiffness of the lever domain. In the presence of ETGA, the 1st, 3rd and 4th unbind, and in the presence of Ca^{2+}, the 3rd and 5th unbind. A test value was used for the Young's modulus of the lever itself, as the properties of the lever domain are not well understood in this scenario (see section 4 on page 10). However, the unstable alpha-helix is known to be extremely flexible and chaotic in its motion, so the value was set close to the lowest value that had previously been used without making the simulation unstable.

Figure 9: Myosin-VII molecule in the presence of Ca^{2+}. The 3rd and 5th calmodulins are unbound.

Figure 10: Myosin-VII molecule in the presence of ETGA. The 1st, 3rd and 4th calmodulins are unbound.

The final scenario to be tested was that of an alpha-helix collapse. When unbound from the calmodulins, it has also been suggested [10] that the alpha-helix is unstable enough for the secondary structure become tangled and collapse entirely. Again, this behaviour is not well understood, but as FFEA is not capable of representing this secondary structural information, the

collapse was represented by a decrease in the length of the regions of the lever with unbound calmodulins.

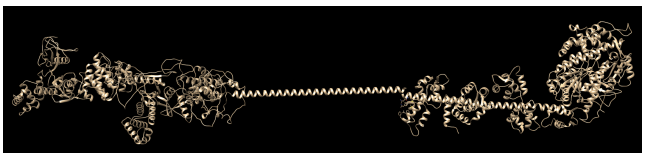

Figure 11: Myosin-VII molecule in the presence of Ca^{2+} with a collapsed alpha-helix.

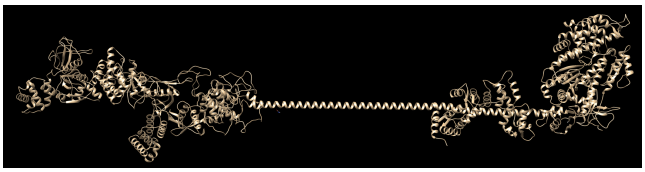

Figure 12: Myosin-VII molecule in the presence of ETGA with a collapsed alpha-helix.

An overview of the material parameters, granularities and other details of the models and simulation parameters can be found in the appendix of this report (fig. 30).

III. Data Analysis Tools

The motion of molecules in FFEA is stored in 'trajectory files', which represent the positions of a large number of indexed points ('nodes') across a number of frames. FFEA trajectories are structured hierarchically, with the first level down being comprised of objects internally referred to as 'blobs'. In the case of Myosin-VII, only one blob is needed.

The 'blob' is comprised of frames, with each frame containing an array with the x, y and z coordinates of every node in that frame. The row of the array that the node is being stored in is the index of that node. This allows the positions of individual nodes, or large groups of nodes, to be tracked between frames.

Blobs may also contain 'sub-blobs' which are defined by a list of indices of the nodes contained within the sub-blob. Sub-blobs can be created using 'pin files', which are in turn created by a pin file command-line interface (CLI) that allows the user to set a pin based on a central node and radius, in the same fashion that the user can set the material parameters. A sub-blob acts like a filter for the method which returns the frame data, returning only the positions of the nodes with a particular set of indices. A sub-blob can be addressed by either its generated ID or by a user-given name.
] All 'blobs' and 'sub-blobs' have a centroid associated with them that is calculated upon request, using the blob's 'get_centroid' method. This centroid is a trajectory containing the average position of every node in that sub-blob. The methods described in this section work by using sub-blob centroids rather than individual nodes, as sub-blobs are less affected by minor thermal fluctuations. It also allows for easier use of the CLI, as pin files are generated once for particular models and can then be loaded from a file, rather than having to specify points or radii manually each time the analysis scripts are run.

A number of modifications were made to the FFEA python libraries to allow useful information to be extracted from trajectory files. These include the distance between the head and tail, the angular distribution, and whether the molecule is even in the correct orientation for binding at all. Three metrics (figures 13, 14, and 15) were defined in order to describe this information:

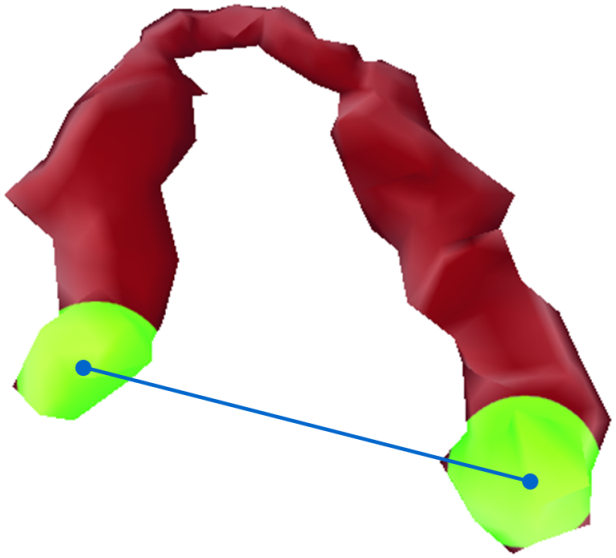

Figure 13: The end-to-end distance of the Myosin-VII molecule is defined as the distance between the centroids of two user-defined 'sub-blobs', containing all of the nodes in the head and tail, respectively. In this image, the centroids are shown as blue dots, and the sub-blobs that they represent are shown in green. The end-to-end distance is depicted as the blue line.

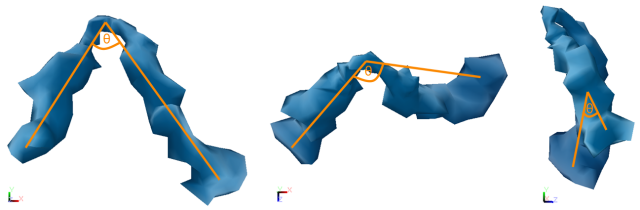

Figure 14: The angular distribution of the Myosin-VII molecule is defined by an angle θ, which is found from lines constructed from the positions of the head, tail and active region. This angle is computed individually for projections of the molecule in the xy, yz and xz planes.

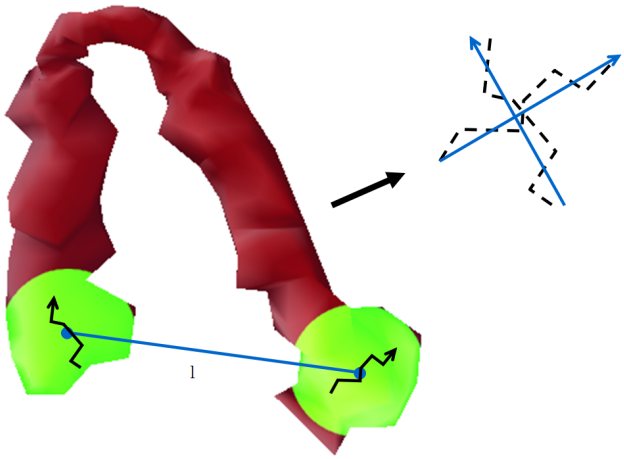

Figure 15: The twist experienced by the molecule is defined as the angle between two vectors embedded in the head and tail, which are orthogonal to the long axis of the molecule. These vectors are projected into a plane in the direction of the current orientation of the molecule, and the angle between them is found using the dot product. This is divided by the end-to-end distance (see fig. 13) to get a value of twist in radians/m.

This analysis required extensive modifications to the existing FFEA libraries and the development of new command-line interfaces, as well as a number of bugfixes and debugging improvements. The exact 'diffs' for these changes are shown in the appendix of this document.

The motion of the molecule can also be analysed using principal component analysis, which breaks down the motion into a series of vibrational modes, with the 'inverse stiffnesses' of the vibrational modes being represented by the eigenvalues of the covariance matrix. As the simulation goes on, these eigenvalues will converge around a particular value - if they have converged, it means that the simulation has fully explored conformational space. A package called PyPcaZip, (originally designed to compress large trajectory files) was used for this purpose. A PCA analysis script runs the PCA algorithm many times (n times) per trajectory, and gives the eigenvalues in intervals of $\frac{1}{n}$ frames (e.g. for a 10,000 frame trajectory at $n = 10$, it would give the eigenvalues for the first 1,000 frames of the trajectory, then the first 2,000, then the first 3,000, etc.).

III. Results

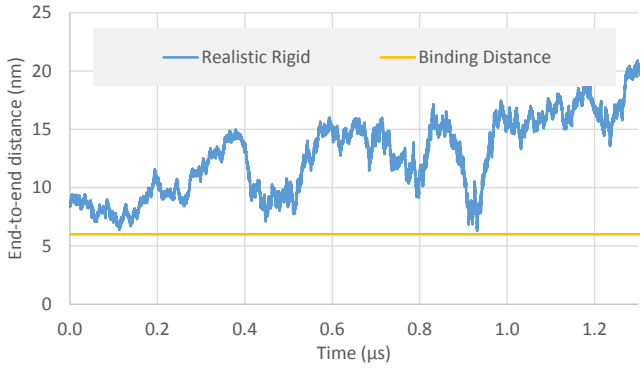

Figure 16: The end-to-end distance (see fig. 13) of the Myosin-VII molecule (fig. 1) with 'realistic' material parameters, with all of the IQ domains in the lever bound. In this case, the most flexible region was the alpha-helix, with a Young's modulus of 382MPa. The approximate end-to-end distance required for binding is indicated with a blue line. This value incorporates the distance from the centroid to the binding site of both the head and the tail. The molecule is not able to bind in this configuration, which supports the idea that some of the IQ domains must be unbound for inhibition to occur.

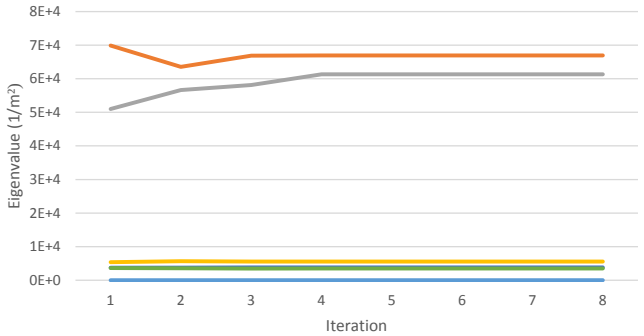

Figure 17: Convergence of associated with the normal modes of vibration. The trajectory was divided into 12 chunks (iterations) with each iteration representing around 1000 frames. The y-axis represents the eigenvalues of the normal modes, which represent the 'inverse stiffness' of the material. The higher eigenvalues (associated with the larger motions) appeared to converge very quickly, suggesting that the molecule has fully explored conformational space.

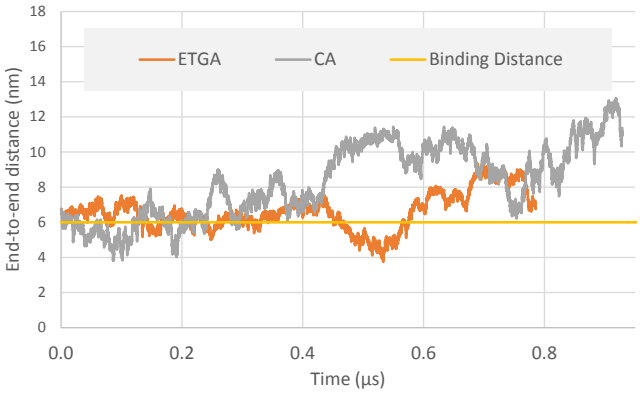

Figure 18: In the presence of Ca^{2+} and ETGA (fig. 9 and 10), some of the IQ domains unbind from calmodulins, dramatically reducing the Young's modulus of that region. A trial value of 80MPa was used, as there is no experimental data for Myosin-VII in this configuration. The rest of the molecule is parameterised through experimental data, as in figure 17. In both cases, the end-to-end distance (fig. 13) passes the binding threshold. In the presence of ETGA, the molecule experiences a lower mean end-to-end distance and a lower minimum distance. This may be a consequence of the fact that ETGA unbinds an additional calmodulin (2 of 5 are bound, as opposed to 3 of 5 in the presence of Ca^{2+}).

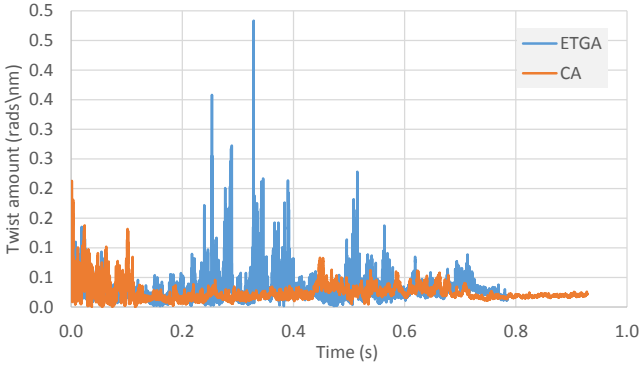

Figure 20: Twist (see fig. 15) comparison between collapsed (fig. 11 and 12 and uncollapsed (fig. 9 and 10) lever results: any twist represents a deviation from the ideal bending trajectory.

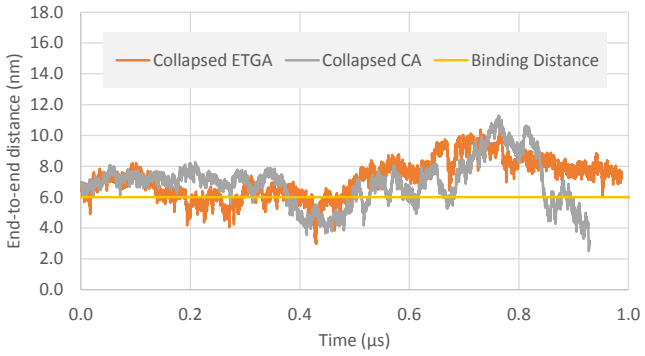

Figure 19: Collapse of the lever alpha-helix when calmodulins unbind in the presence of ETGA (fig. 12) and Ca^{2+} (fig. 11): Sakai et al [10] suggest that the SAH domain may collapse when the calmodulins unbind. This has been represented by the shortening of the unbound regions in the underlying alpha-helix. Both versions experiece a lower minimum end-to-end distance. Although Ca^{2+} experiences a higher mean end-to-end distance, this should not be taken as evidence that this configuration is less optimal; it may be a random sampling issue.

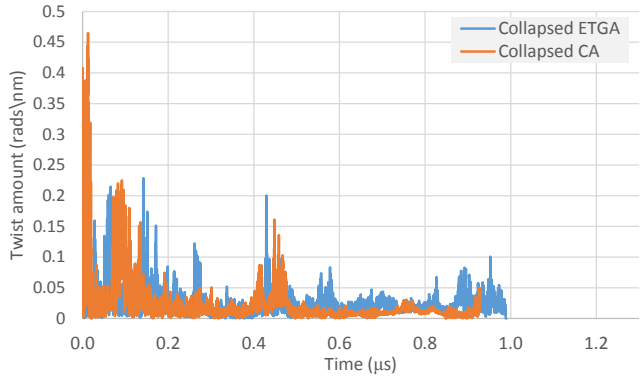

Figure 21: The molecule experienced a similar degree of twist (fig. 15) in its collapsed (fig. 11 and 12) and uncollapsed alpha-helix state (fig. 9 and 10). As the twist is normalized to the length of the molecule, the length is not a factor in this test.

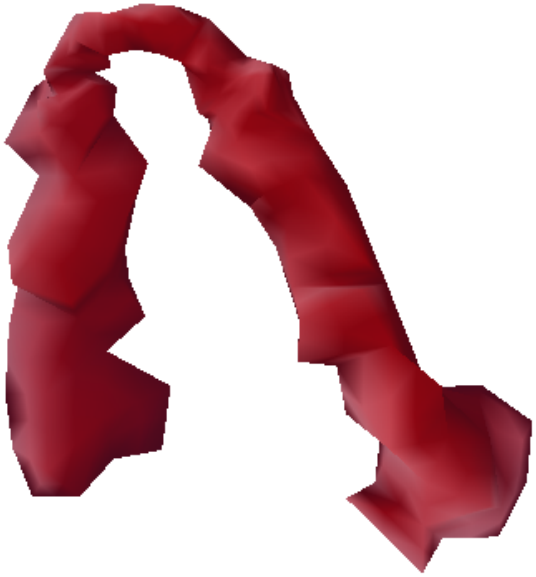

Figure 22: A Myosin-VII molecule in a highly-twisted state. The head and tail are misaligned and are thus unable to bind.

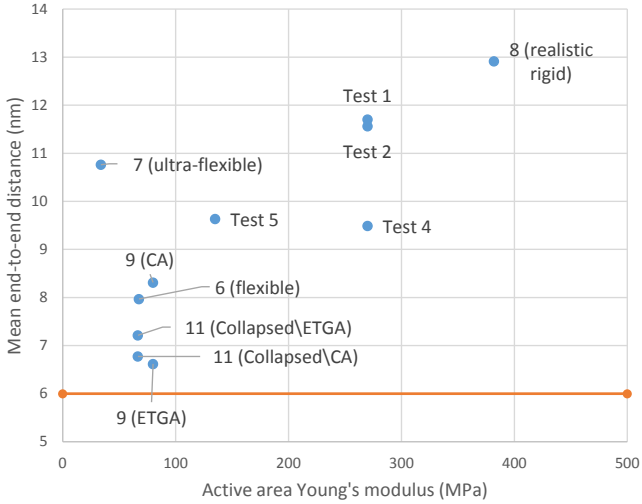

Figure 24: Active area Young's modulus compared to mean end-to-end distance. The maximum end-to-end distance required for binding to occur is indicated with a line. A lower Young's modulus in the active area (defined as the lever in the presence of Ca^{2+} and ETGA, and the alpha-helix otherwise) generally results in a lower mean end-to-end distance. More information on the labels can be found in a summary table in the appendix (fig.30).

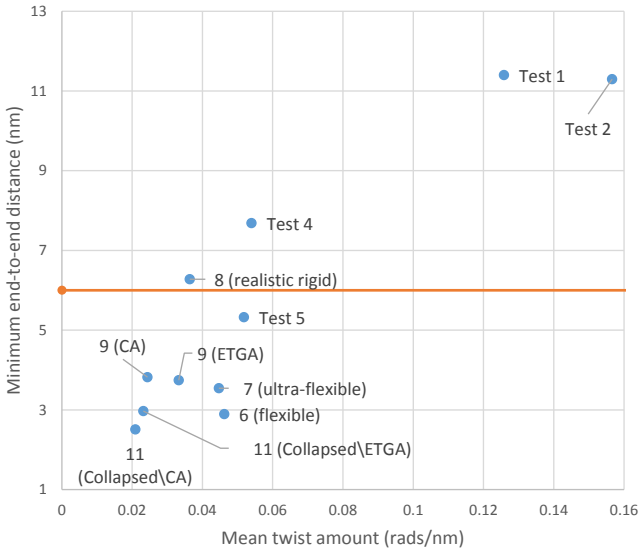

Figure 23: Active (bendiest) area Young's modulus compared to minimum end-to-end distance: the simulations below the line indicate that the molecule entered a configuration in which it was able to bind successfully. The binding site forces were not enabled for any simulations. Disabling binding allowed the molecule to explore conformational space for the entire duration of the simulation. As expected, a reduction in the Young's modulus results in a reduction in minimum end-to-end distance, with all of the inhibited Myosin-VII simulations able to bind successfully. More information on the labels can be found in a summary table in the appendix (fig.30).

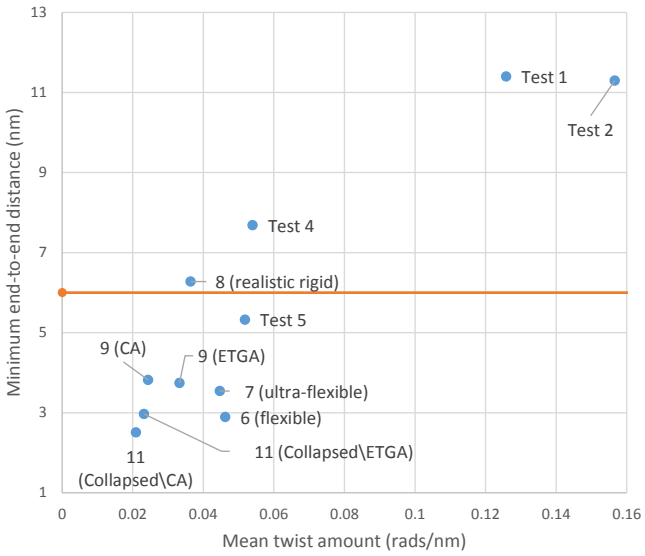

Figure 25: Mean twist amount compared to minimum binding site separation: this comparison demonstrates the negative effect twist can have on the behaviour of the molecule. Molecules with larger twist amounts generally experienced a higher minimum end-to-end distance, owing to effects such as those demonstrated in fig 22. This is also evidence that a lower Young's modulus does not necessarily mean a greater twist amount: structural differences and the position of the 'bendy region' are also important. More information on the labels can be found in a summary table in the appendix (fig.30).

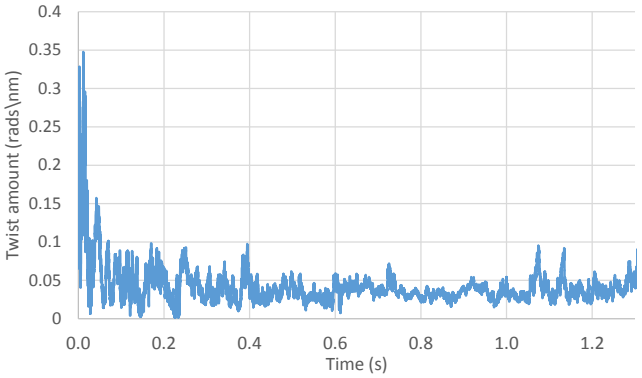

Figure 26: Overall reduction in motion of molecule for extremely long trajectory: the twist amount of a Myosin-VII molecule parameterised with dummy values. The simulation was started close to the beginning of the investigation, and was the most long-lived. By the end of the trajectory, the molecule has almost entirely stopped twisting. The fluctuations in end-to-end distance follow a similar pattern. Around 2/3rds of the way into the trajectory, the molecule undergoes an extremely sudden reduction in motion in every degree of freedom. The reason for this is not yet understood - although unlikely, it could be a conversion of vibrational energy to translational energy, which is not represented in our metrics or visualisations. It could also be a bug in the simulation software or automatic restart script. No change was found in the potential energy of the molecule at this point.

IV. Discussion

The investigation has produced valuable new insights into the behaviour of Myosin-VII. First, it was found that the Myosin-VII molecule was unable to bind to itself in its uninhibited state. A relationship was established between the Young's modulus of the molecule and the end-to-end distance, and the twist of the molecule. Observations of Myosin-7's trajectory were made in large quantities, and at extremely long timescales that allowed for an extensive sampling of conformational space. The molecule was found to be able to bind to itself in the presence of either Ca^2+ or ETGA, as predicted by Yang and Umeki et al [1], [2], in both its collapsed and uncollapsed state. The timescale of the collapse and the structure that results are not known, but it has been shown to be possible for the binding to occur at both extremes - with no collapse, and with an extreme collapse.

The 'twist' measurement provides an entirely new way to visualise the conformation of Myosin-VII, and has already provided an interesting result - the gradual reduction in twist of the Myosin-VII molecule as seen in fig. 26. There is no explanation for this behaviour yet, but future analysis is recommended, possibly in the form of more FFEA simulations, followed by a comparison with long-lived atomistic MD data. The 'twist' measurement also demonstrated the negative effect of twist on end-to-end distance, as shown in figure 25.

Some opportunities for improvements to future methods were also identified.

At the parameterisation stage, the stiffness data used was extracted from Myosin-V, not Myosin-VII, the former containing one extra IQ domain. The bound IQ domains contribute to the overall stiffness of the lever, which means that the stiffness is likely to be a slight over-estimate. The assumption that the molecule is a rod may have increased the stiffness estimate further.

Luckily, there are already experiments looking at more accurate ways of measuring the rigidity of Myosin molecules using large, time-averaged samples of the molecule's conformation [11].

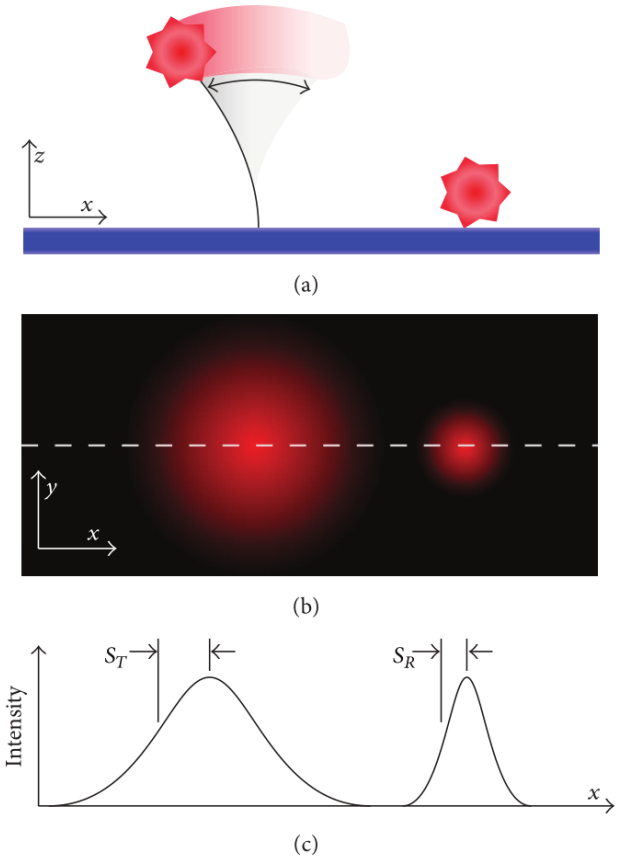

Figure 27: Motion of quantum dot attached to tethered Myosin molecule. The distribution of the quantum dot's position under Brownian motion is Gaussian. This distribution is wider than that of the quantum dot simply bound to the surface. The difference in widths is used to estimate the molecular stiffness. This method has already been used to calculate the stiffness of Myosin-V.

A possible 'sampling bias' is introduced by the FFEA restart script. When an element inverts, and causes the simulation to crash, the simulation is restarted by the restart script. The frame that caused the crash is removed from the trajectory. This restart script works on the assumption that crashes will come as a purely random result of thermal noise. In some cases, it was found that the more flexible models induced more crashes from elements in the more flexible regions (such as the lever and SAH domain) than the other regions. If these elements are inverting due to extreme bends, then this could introduce a sampling bias that favours trajectories with less extreme bends. To counteract this, the granularity of the active region was kept relatively high in later simulations, such as the inhibited Myosin-VII simulations. In the future, more sophisticated logging techniques from the restart script could be used. This would be combined with a new metric to asses how the crash logs differ from the expectation value due to random noise.

It could also be argued that the simulation of the collapsed alpha-helix is a misuse of the FFEA simulation technique, as the collapsed alpha-helix, unstable and lacking hydrogen bonds, undergoes hydrophobic interactions with the surrounding medium. The change in secondary structure that results from this collapse is not reflected in the FFEA simulation. However, this represents a good opportunity to compare FFEA to MD simulations, and the properties of the unstable alpha-helix may become more predictable at longer timescales. Indeed, internal investigations into the comparability of FFEA data and MD simulations are already underway.

One more assumption is the use of centroids in the head and tail domain to track the motion of the binding sites. Using the binding site nodes alone makes the data too noisy to understand[3] - but it does not reflect the conformation of the molecule. This is evident in the graphs relating to Myosin-VII in the collapsed alpha-helix state (see figures 19 and 21), as the overall length of the molecule is very short, and this makes the binding sites appear quite close, even though they are not in the correct configuration to bind. Future implementations of the FFEA analysis tools will improve this method, likely by increasing the granularity locally at the binding site, allowing for a larger sampling of nodes in a small area. Although this error is difficult to quantify, the twist measurement can be used as an indicator, as it represents the misalignment of the head and tail.

Overall, FFEA is a technique with potentially profound implications on the field of molecular dynamics. Even these early FFEA simulations are capable of producing results that are in agreement with experimental evidence, such as the electron microscopy data from Yang and Umeki et al [1], [2]. The relationship of the Young's modulus to end-to-end distance, the relationship of twist to end-to-end distance, the effect of introducing the inhibitors Ca^{2+} and ETGA, the timescales required for the molecule to explore conformational space, and many others have provided a large and valuable body of data that does not contradict existing literature.

FFEA development itself is extremely fast-paced and new features are implemented on a regular basis. With more documentation, the project will also be ready for an open-source release.

The ultimate goal of FFEA is not to simulate individual molecules, but rather to look at mesoscale phenomena that MD simulations are not capable of simulating - supramolecular structures consisting of tens or potentially hundreds of molecules, and complex processes that are outside of our understanding, such as fibrin clot formation.

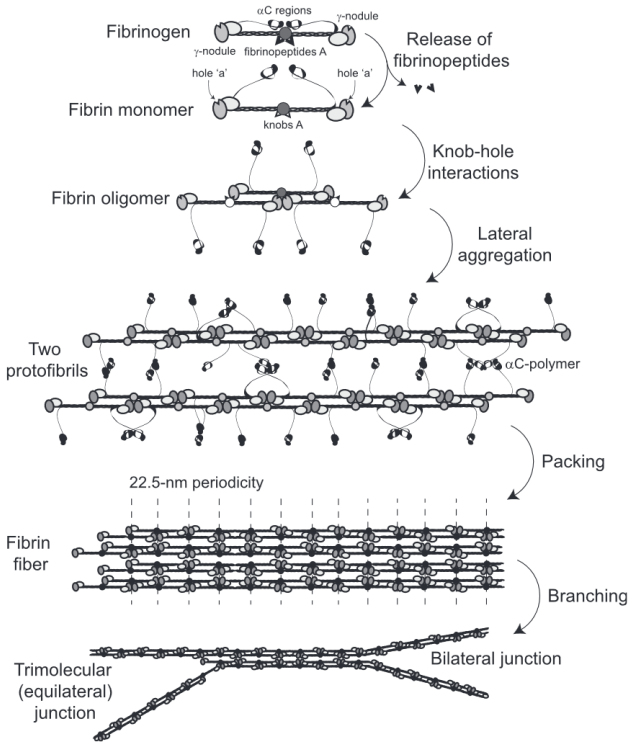

Figure 28: Fibrin is a protein responsible for blood clotting. Fibrin polymerisation [12] is an example of a complex biological process that may one day be modeled by coarse-grained simulation techniques such as FFEA.

[3]There is an FFEA_trajectory method to get the distance between two arbitrary nodes, but it outputs unusabe data. The tradeoff between getting closer to the binding site and adding more noise is not a fair one. Ultimately, the method was rejected entirely.

This is also FFEA's true strength - huge motor proteins such as Dynein, or large supramolecular structures do not need the fine level of detail used in atomistic simulations. Molecular dynamics simulations are often perceived as niche techniques that are not broadly applicable to biological systems, but newer simulation techniques designed with larger systems in mind could help to change this perception.

V. Acknowledgements

Thanks to the FFEA team at the University of Leeds: Glenn Carrington, Benjamin Hanson, Oliver Harlen, Sarah Harris, Daniel Read, Albert Solernou and James Taylor-Ims.

VI. References

[1] Yi Yang, Thomas G. Baboolal, Verl Siththanandan, Michael Chen, Matthew L. Walker, Peter J. Knight, Michelle Peckham, and James R. Sellers. A ferm domain autoregulates drosophila myosin 7a activity. *Proceedings of the National Academy of Sciences*, 106(11):4189–4194, 2009.

[2] Nobuhisa Umeki, Hyun Suk Jung, Shinya Watanabe, Tsuyoshi Sakai, Xiang-dong Li, Reiko Ikebe, Roger Craig, and Mitsuo Ikebe. The tail binds to the head-neck domain, inhibiting atpase activity of myosin viia. *Proceedings of the National Academy of Sciences*, 106(21):8483–8488, 2009.

[3] Alan Gray, Oliver G Harlen, Sarah A Harris, Syma Khalid, Yuk Ming Leung, Richard Lonsdale, Adrian J Mulholland, Arwen R Pearson, Daniel J Read, and Robin A Richardson. In pursuit of an accurate spatial and temporal model of biomolecules at the atomistic level: a perspective on computer simulation. *Acta Crystallographica Section D: Biological Crystallography*, 71(1):162–172, 2015.

[4] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[5] Robin Archibald Richardson. *A mesoscale model for coarse-grained protein dynamics*. PhD thesis, University of Leeds, 2014.

[6] Robin Oliver. *A stochastic finite element model for the dynamics of globular proteins*. University of Leeds, 2013.

[7] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera - a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[8] Joachim Schöberl. Netgen an advancing front 2d/3d-mesh generator based on abstract rules. *Computing and visualization in science*, 1(1):41–52, 1997.

[9] Andrej Vilfan. Elastic lever-arm model for myosin v. *Biophysical journal*, 88(6):3792–3805, 2005.

[10] Tsuyoshi Sakai, Hyun Suk Jung, Osamu Sato, Masafumi D. Yamada, Dong-Ju You, Reiko Ikebe, and Mitsuo Ikebe. Structure and regulation of the movement of human myosin viia. *Journal of Biological Chemistry*, 290(28):17587–17598, 2015.

[11] Arthur J Michalek, Guy G Kennedy, David M Warshaw, and M Yusuf Ali. Flexural stiffness of myosin va subdomains as measured from tethered particle motion. *Journal of Biophysics*, 2015, 2015.

[12] John W. Weisel and Rustem I. Litvinov. Mechanisms of fibrin polymerization and clinical implications. *Blood*, 121(10):1712–1719, 2013.

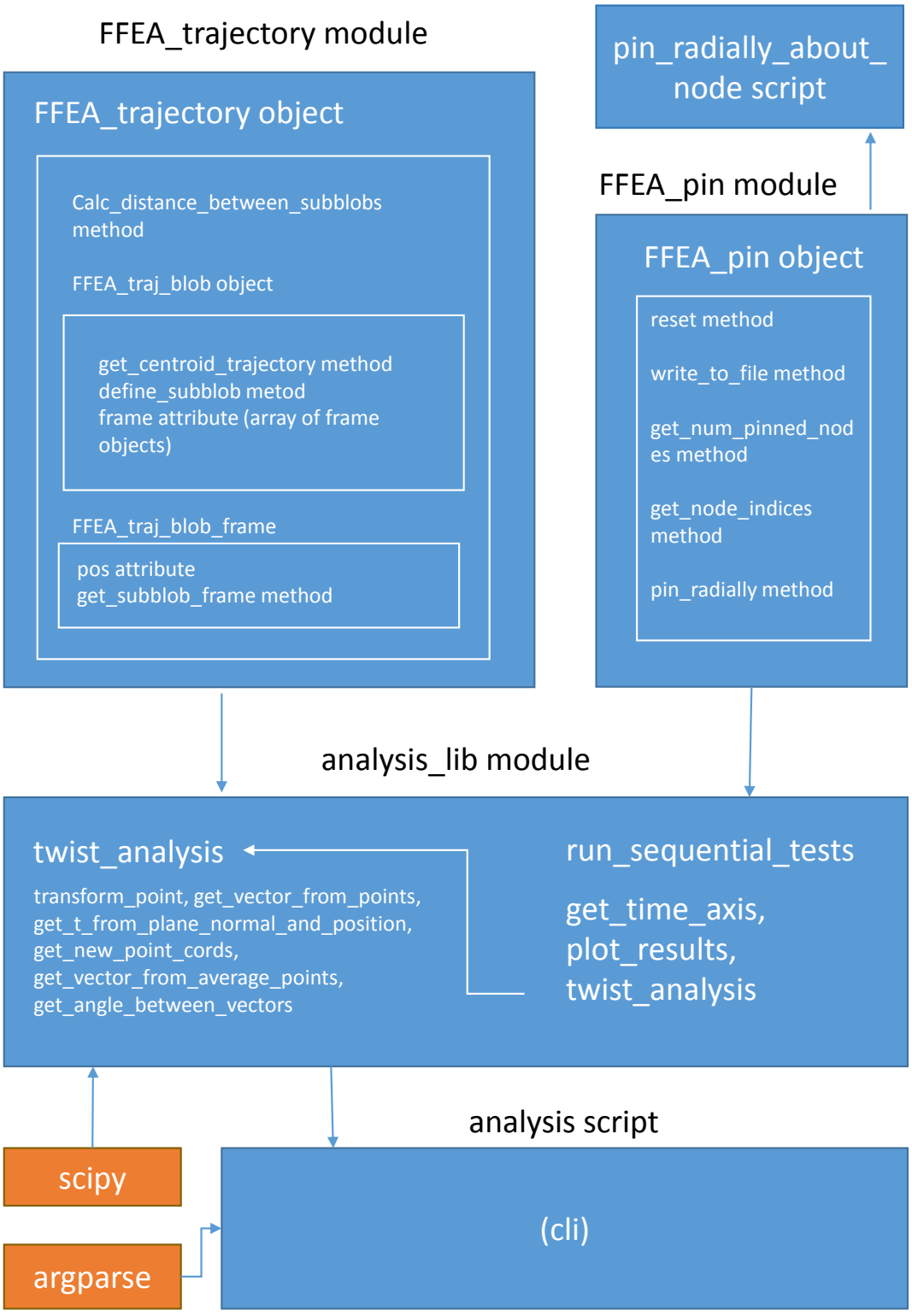

Figure 29: Overview of FFEA analysis scripts. The analysis script is a frontend for the analysis_lib module, powered by the argparse module. The analysis_lib module contains thre vector math functions needed to compute the twist, end-to-end distance and angular distribution, and the plotting functions to visualise the results. It depends upon the pin module, to extract the node indices from the pin files, and the data access methods in the FFEA_trajectory module.

Sim Number	1	2	4	5	6	7	8	9	9	11	11
Description	Early model test	Early model test	Variable graining test	Variable graining test	Unrealistically low Young's modulus	Extremely low Young's modulus	Realistic Young's modulus	1, 3 and 4 calmodulins unbound (ETGA)	3 and 5 calmodulins unbound (CA2+)	Collapsed alpha-helix (ETGA)	Collapsed alpha-helix (CA2+)
Min end-to-end distance (nm)	11.4	11.3	7.7	5.3	2.9	3.5	6.3	3.7	3.8	3.0	2.5
Max end-to-end distance (nm)	12.0	12.0	11.3	14.0	12.5	19.6	21.6	9.3	13.0	10.4	11.3
Avg end-to-end distance (nm)	11.7	11.6	9.5	9.6	8.0	10.8	12.9	6.6	8.3	7.2	6.8
Min twist (rads/nm)	0.07	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Max twist (rads/nm)	0.17	0.20	0.24	0.33	0.29	0.35	0.38	0.48	0.21	0.23	0.47
Avg twist (rads/nm)	0.13	0.16	0.05	0.05	0.05	0.04	0.04	0.03	0.02	0.02	0.02
E (active area) (Mpa)	270	270	270	135	67.5	33.75	382	80	80	66.59	66.59
E (lever) (Mpa)	270	270	270	270	135	135	827	80	80	66.59	66.59
E (SAH) (Mpa)	270	270	270	135	67.5	33.75	382	355	355	355	355
Distance Apart (approx) (nm)	5.4	5.3	1.7	-0.7	-3.1	-2.5	0.3	-2.3	-2.2	-3.0	-3.5
Active area granularity (nm)	1	1	0.8	0.8	0.8	0.8	0.8	1	1	0.9	0.9
Max granularity (nm)	1	1	2.2	2.2	2.2	2.2	2.2	2.1	2.1	2.1	2.1
Binding condition fulfilled?	✕	✕	✕	✓	✓	✓	✕	✓	✓	✓	✓

Figure 30: Summary table containing model generation parameters for each simulation.