The flexibility of myosin 7a

Thesis by

Glenn Stuart Peter Carrington

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

School of Molecular & Cellular Biology

September 2018

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements.

I am indebted to the Engineering and Physical Sciences Research Council for their funding, and the Faculty of Biological Sciences for making my PhD possible.

I would like to thank my supervisors Michelle Peckham, Sarah Harris, Daniel Read, and Oliver Harlen for their patience, tireless enthusiasm and advice throughout the ups and downs of the project. I am grateful to Ben Hanson, Albert Solernou, and Robert Welch for all things computer related and helping with my (inevitable) Seg-faults. I am especially thankful to Charlotte Scarff, Ruth Hughes and Peter Knight for their helpful advice and insightful knowledge on all things Biology, Mass-spec and EM related (in no particular order).

Thanks to Rachel George in the University of Leeds Mass Spectrometry Facility and Martin Fuller, Emma Hesketh, and Rebecca Thompson in the University of Leeds Bio-structural Facility for their support and use of equipment over the course of this PhD.

I am indebted to the past and present members of the Peckham lab for their invaluable support and well deserved coffee breaks. In particular Adriana, Alistair, Fran, Louise, Marcus, Marcin, Marta, and Matt B, it's been a long journey.

I should also probably thank Brendan, whom I have shared a constant back and forth of digs and jibes (Lonely Hearts Mondays) with over these last 4 years. I hope that you now see me than more than just a work colleague….. remember they don't call us Glendan for nothing.

I must also thank my Family, particularly my Mum and Dad for their support and encouragement during my time in Leeds. My final thanks goes to Jordan, whose continuous encouragement and faith in my ability is what kept me going all these years.

Abstract

Myosin 7a is a molecular motor found in hair cells of the ear and the photoreceptor cells of the eye. Myosin 7a is comprised of an actin-binding motor domain, a lever; which is composed of 5 IQ motifs that can potentially bind 5 light chains followed by a single alpha helical (SAH) domain, and a tail composed of 2 MyTH4-FERM domains. The lever is an essential mechanical element in myosin 7a function, but an understanding of its mechanical properties and how these derive from its substructure is lacking. It has been observed *in vitro* that myosin 7a is able to regulate its activity through a head-tail interaction.

How the flexibility of the sub-domains of the lever allows the molecule to fold up is not completely understood. To address this, the first aim of this study was to look for evidence of novel light chain binding partners in myosin 7a, which revealed calmodulin to be the preferred light chain. My second aim was to study the structure and flexibility of the lever of full-length myosin 7a using single-particle image processing of images from negative stain electron microscopy (EM). Image averaging revealed the lever to be much shorter than expected. Additionally, there was evidence of thermally-driven flexing at the motor-lever junction. A stiffness of 78 pN.nm.rad^{-2} for the flexing was inferred, which represents a significant compliance in the head. An investigation into lever bending analysis, by monitoring the decay of tangent-tangent correlations of the lever shapes, yielded a persistence length of 38 ± 3 nm.

Finally, long time molecular dynamics (MD) simulations were compared with a novel coarse-grained (CG) simulation technique called Fluctuating Finite Element Analysis (FFEA), which treats proteins as visco-elastic continua subject to thermal noise to probe the flexibility of myosin 7a. FFEA allows sufficiently long time simulations that are computationally less expensive than corresponding all-atom MD simulations to allow myosin 7a to explore its full range of configurations. Extraction of flexibility data from all-atom MD simulations calculated the bending stiffness of the SAH domain to be 60.5 pN.nm^2, with reasonable overlap of the major modes of motion between the all-atom and CG simulation types.

Table of Contents.

List of Tables.

List of Figures.

List of Abbreviations.

Units and constants

Å	10^{-10} metres
bp	Base pair(s)
˚C	Degrees Celsius
Da	Dalton(s)
g	Acceleration due to gravity = 9.8 ms-2
hr(s)	Hour(s)
kb(p)	10^3 base (pairs)
kDa	10^3 Daltons
K	Kelvin
min(s)	Minute(s)
mL	10^{-3} litre(s)
mM	10^{-3} molar
µL	10^{-6} litre(s)
µm	10^{-6} metre(s)
µM	10^{-6} molar
nm	10^{-9} metre(s)
nM	10^{-9} molar
pH	Potential of hydrogen
pN	10^{-12} newtons
MPa	Mega pascals
GPa	Giga pascals
s	Second(s)

Other abbreviations

2D	Two dimensional
3D	Three dimensional

4Knanobody	4K GFP nanobody
MD	Molecular dynamics
FFEA	Fluctuating Finite Element Analysis
ADP	Adenosine-5’-diphosphate
ATP	Adenosine-5’-triphosphate
ATPase	Adenosin 5’-triphosphatase
Ca^{2+}	Calcium ion
CaM	Calmodulin
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
LC	Light chain
ELC	Essential light chain
RLC	Regulatory light chain
EM	Electron Microscopy
F-actin	Fibrous/filamentous actin
GFP	Green Fluorescent Protein
HC	Heavy chain
IQ	IQ motif
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani
MW	Molecular weight
MWCO	Molecular weight cut off
N-	Amino-
C-	Carboxy-
OD	Optical Density (absorbance)
PBS	Phosphate buffered saline

PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
Pi	Inorganic phosphate
S1	Myosin subfragment 1
SAH	Single alpha helix
S.D.	Standard deviation
SDS	Sodium Dodecyl Sulphate
PAGE	Polyacrylamide Gel Electrophoresis
SFM	Serum Free Medium
SFX	Serum free media
Sf9	*spodoptera frugiperda*
SH3	Src-homology region type 3
S.O.C.	Super Optimal broth with Catabolite repression
TAE	Tris-Acetate-EDTA
TEM	Transmission electron microscopy
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
v/v	Volume / volume
w/v	Weight / volume

1. Introduction.

1.1 Molecular machines, motor proteins and movement.

The cell is a highly dynamic system that relies on multiple processes to maintain its function and enable its specific role. A key player in these processes are collectively known as molecular machines; assemblies of proteins, nucleic acids and other molecular components able to produce mechanical actions in response to specific external stimuli (Alberts, 1998; Ballardini et al., 2001). The cell has evolved many types of molecular machine to carry out its essential functions. These include cytoskeletal motor proteins (myosins, kinesins and dyneins) to generate directed movement along actin filaments and microtubules; rotary motors, pumps and translocation pores (e.g. the bacterial flagella motor, F- and V- ATPases) to generate rotational movement and move atoms across membranes to generate an energy potential; construction machines (e.g. the ribosome and polymerases) to build proteins and nucleic acid polymers; degradation machines (e.g. proteosomes) to break up proteins, as well as machines to separate, uncoil or pack nucleic acid strands (helicases, topoisomerases and condensins).

Many of the cell's molecular machines are also called molecular motors. A molecular motor or motor protein can be defined as a mechanochemical enzyme; a molecule that is able to convert different forms of chemical energy into mechanical work, through conformational changes and structural displacements (Bustamante et al., 2004).

The ability to generate and control movement is a defining characteristic of life at all structural levels, from the co-ordinated motions of many proteins that produce macroscopic scale movements such as muscle contraction or sperm propulsion, to the intracellular transport of cargoes along biopolymer tracks (Vale and Milligan, 2000). Motor proteins are the force-generating molecules that participate in all these processes. Therefore, understanding their structure and how their mechanical properties allow them to perform their function is an important research goal for biology and biophysics (Howard and Clark, 2002).

1.2 Cytoskeletal motor proteins.

The cytoskeletal motor proteins are mechanochemical-enzymes capable of converting the chemical energy from ATP hydrolysis into force for directed movement along cytoskeletal filaments. There are three distinct families: myosins, kinesins and dyneins. The cytoskeletal tracks to which these motors bind are actin filaments and microtubules. Both tracks have a polar structure, with faster- and slower-polymerising ends known as plus and minus ends respectively. Myosin binds to and make movements on actin filaments, and move towards the plus (or barbed) end, with the exception of myosin 6 which moves towards the minus end (Wells et al., 1999). Kinesins and dyneins bind and move on microtubules.

All three cytoskeletal motor families share the ability to move in a directed fashion along a filamentous track. The functional characteristic common to each class of motor is that there are coordinated cycles of ATP hydrolysis, track binding/unbinding and conformational changes. Each motor goes through a cycle of track attachment and detachment, which is coupled to ATP hydrolysis and product release. While attached to the track, the motor moves a mechanical element in the direction of travel and while detached from the track, the motor reverses this movement ready for the next cycle (Howard and Clark, 2002). In kinesins this mechanical element is a 13 amino acid neck linker region. When this linker docks onto the lead head of the kinesin dimer, it causes the detached trail head to swing forward and reach the next binding site (Rice et al., 1999). In cytoplasmic dynein, the mechanical element appears to be a much larger domain also called the linker. During dynein's mechanochemical cycle the linker swings across the AAA+ ring acting as a distance multiplying lever (Roberts et al., 2009; Imai et al., 2015). Its N-terminal tip moves ~17 nm during this swinging movement. In myosin, the mechanical element is called the lever and is generally an extended α-helix that emerges from the motor domain to which light chains are bound. Small nucleotide-dependent movements within the motor are amplified into large displacements of the lever, with the tip of the lever moving between 10 – 25 nm depending on the myosin.

1.3 Myosins.

Myosins form a diverse superfamily of motor proteins with many functions ranging

from muscle contraction to translocating vesicles and other cargoes along actin filaments (Sellers, 1999). In humans, the myosin superfamily is represented by 38 genes, which can be organised into 12 different classes using phylogenetic analysis (Berg et al., 2001). These can be divided into conventional and unconventional myosins. Myosin 2 was the first type of myosin to be discovered, is present in striated and smooth muscle, and is the most well studied member of the myosin family. Myosin 2 is an example of a conventional myosin. It is composed of two heavy chains, and four light chains. The two heavy chains are able to dimerise to form a coiled coil, and then further assemble to form filaments. This type of myosin generates muscle contraction. Non-muscle myosin 2 isoforms drive movements such as cell crawling and cell division (cytokinesis). All myosins that cannot form filaments are called 'unconventional' myosins.

Unconventional myosin can be further subdivided into a number of different classes, which either contain dimeric or monomeric isoforms. These myosins perform a multitude of different roles within the cell ranging from transport, anchoring/tethering, tension sensing, actin reorganization, and cell motility/adhesion (Hartman et al., 2011).

1.4 Basic structure of myosin.

All myosins are composed of 3 functional domains; a conserved motor domain, typically located at the N-terminus, which binds actin and hydrolyses ATP. The second functional domain is the lever, which is composed of a variable number of consensus sequences of IQXXXRGXXXR, termed IQ motifs (Cheney and Mooseker, 1992), which can bind light chains such as calmodulin and is responsible for converting the small conformational change in the motor into a large movement of the lever, termed the power stroke. The final functional domain is the tail, which is composed of many different domains that can bind cargo and ultimately determines the cellular role of the myosin.

1.5 Myosin motor.

The ~80 kDa N-terminal motor domain is highly conserved amongst all myosins and contains the actin- and nucleotide-binding sites. Hydrolysis of Mg.ATP leads to structural rearrangements within the motor. A number of crystal structures of motor domains from a range of myosins have been solved (Rayment et al., 1993;

Kollmar et al., 2002; Coureux et al., 2003; Coureux et al., 2004), and all of these structures show a similarly conserved structure (Figure 1).

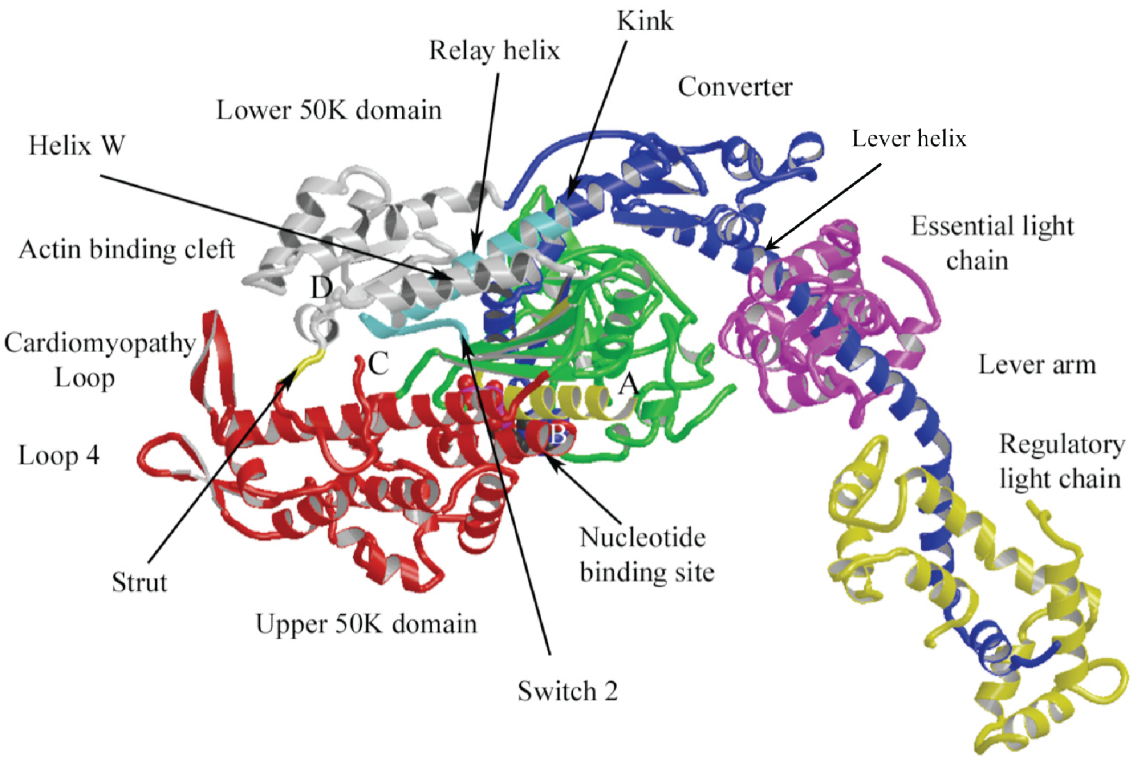


Figure 1 - Structure of the myosin head domain..

Ribbon diagram of chicken myosin 2 head crystal structure (Rayment, Rypniewski, et al., 1993), adapted from Geeves and Holmes (2005). The head appears in the orientation as if bound to actin (at the actin-binding cleft). Subdomains of the motor and the two LCs bound to the lever are shown. N-terminus region (green), nucleotide-binding P-loop and adjoining helix (yellow), upper 50K (red), lower 50K (grey) with relay helix (light blue). Converter and lever helix (blue) with Essential light chain (pink) and Regulatory light chain (yellow). In this conformation, the relay helix is straight and the converter and lever have rotated ~60° such that the lever is post-powerstroke in a ‘down’ position but the actin-binding cleft is open. The state is also termed ‘post-rigor’.

A 7-stranded β-sheet forms the core of the motor and is surrounded by α-helices. A deep cleft divides one side of the motor into the so-called upper and lower 50K domains, which are involved in actin-binding. The cleft opens and closes in response to nucleotide binding, which is coupled to weak (cleft open) or strong (cleft closed) actin-binding states. A small pocket near the apex of the 50K cleft forms the nucleotide-binding site and contains a phosphate-binding P-loop flanked by two catalytic loops named Switch 1 and 2, similar to those found in G-proteins (Smith and Rayment, 1996). Switches 1 and 2 are positioned to allow communication between the nucleotide and actin-binding sites. Switch 2 also connects to the relay helix, which is in turn positioned to make interactions with the converter domain, a compact structure at the C-terminal region of the motor. The converter domain is found close to the N-terminal region of the lever.

Structural rearrangement within the motor (twisting of the core β-sheet), triggered by the strong binding to actin, leads to a cascade of structural rearrangements within the motor, which in turn rotates the converter by ~60°, which then rotates the lever by the same amount (Geeves and Holmes, 2005). This sequence of events is known as the powerstroke (or working stroke), the mechanical action capable of transmitting force and generating directed movement of F-actin.

1.6 Myosin Lever.

1.6.1 IQ+CaM.

The myosin lever domain is able to convert the small conformational changes in the motor into a large displacement of the lever tip. It extends from the motor and is typically a long α-helix (the lever helix) containing one or more copies of a characteristic light chain (LC)-binding sequence, known as an IQ motif. This ~23-residue sequence was named an IQ motif, after the first two amino acids (IQ) in the consensus sequence of the conserved core (Cheney and Mooseker, 1992). Calmodulin or calmodulin-like LCs bind to the IQ motifs and provide rigidity to the lever. Loss of one or more of these light chains, through phosphorylation or altered calcium levels, leads to alterations within the lever domain which regulate the activity of the motor. For this reason, the lever is sometimes referred to as the regulatory domain (Houdusse and Cohen, 1996).

The IQ motifs appearing in the myosin lever, were originally identified as a ~23-25 amino acid long units, with a conserved 11-residue core fitting the consensus

sequence Ile-Gln-x-x-x-Arg-Gly-x-x-x-Arg (IQxxxRGxxxR) (where x is any amino acid) (Cheney and Mooseker, 1992). Most myosins contain between one and six IQ motifs (Sellers, 2000). They usually appear directly after the motor domain, arranged in tandem to form the lever. Bound LCs envelope the amphipathic IQ sequences, which is expected to prevent collapse of the lever, rigidifying this extended helical domain, so that it can transmit force or sustain load. Observations made by EM have revealed the shortening of the myosin heads (motor + lever) upon loss of one or more LCs (Walker and Trinick, 1989; Tominaga et al., 2012). This is consistent with the idea that loss of LCs results in collapse of the lever domain. It is therefore possible that LC dissociation (or conformational change) renders the lever pliant and thus non-functional and this is thought to be one of the mechanisms of myosin regulation (Trybus et al., 2007; Adamek et al., 2008). Many myosins, both conventional and unconventional, are sensitive to Ca^{2+} levels, and this sensitivity is conferred via the LCs. Conformational changes in LCs associated with Ca^{2+}- binding/dissociation (e.g. lobes opening, closing or detaching from the heavy chain), can give rise to altered interactions between LC and IQ motifs, LC and motor, as well as between adjacent LCs.

Sequence alignments of myosin IQ motifs reveal the conservation and properties of residues at different positions in the IQ sequence (Figure 2) (Houdusse et al., 1996; Rhoads and Friedberg, 1997). The first residue of the core is variable, but is commonly either Ile, Leu, or Val (with Met, Phe or Thr in some cases). Core position 11 varies between Arg, Lys or Gln. The Gly at position 7 is less well conserved. In addition, a hydrophobic residue (Φ) is frequently found at position 14, three residues beyond the core sequence (Figure 2).

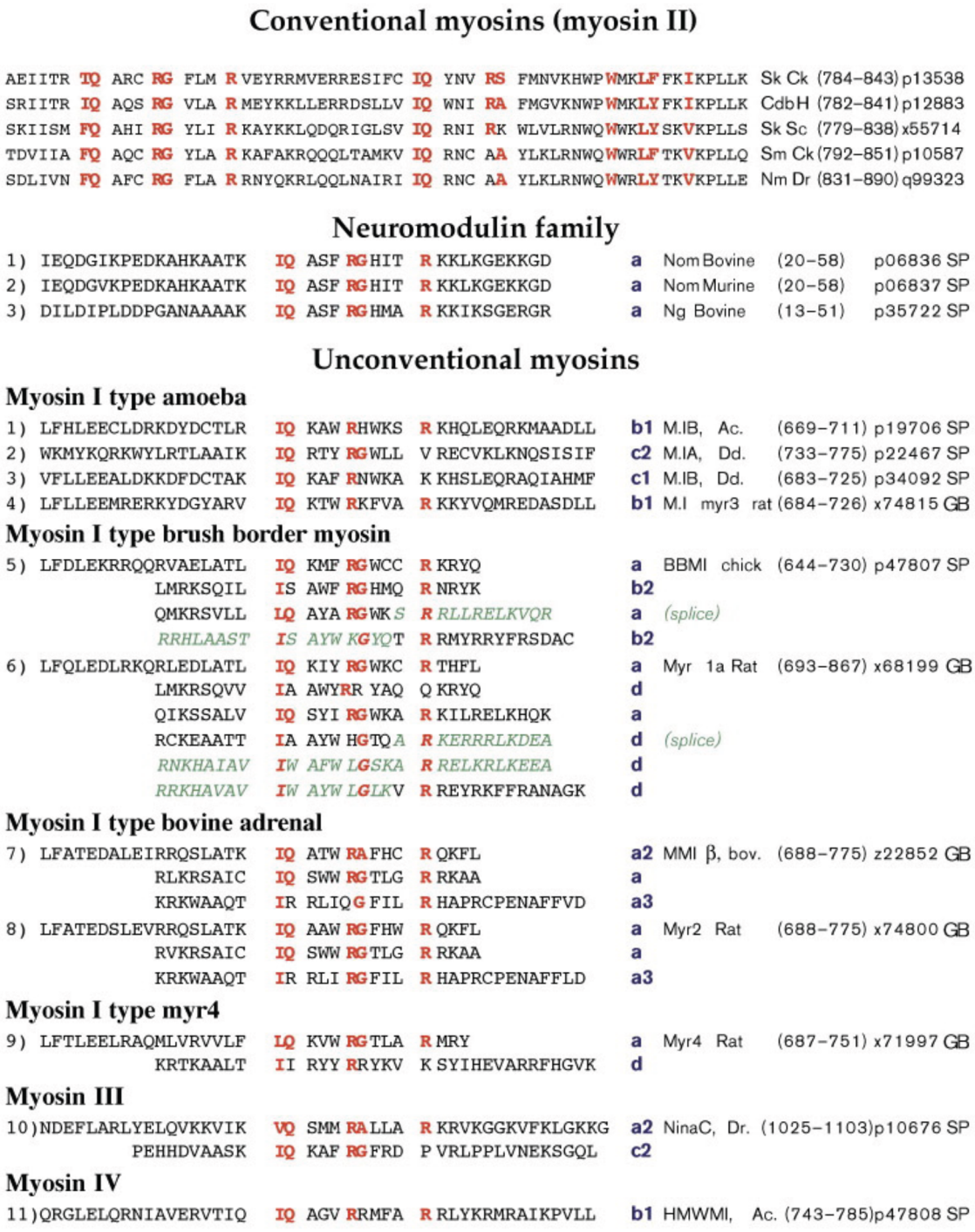


Figure 2 - Sequence alignments of IQ motifs in different myosin types.

Figure from Houdusse et al. (1996). NB: continues overleaf.

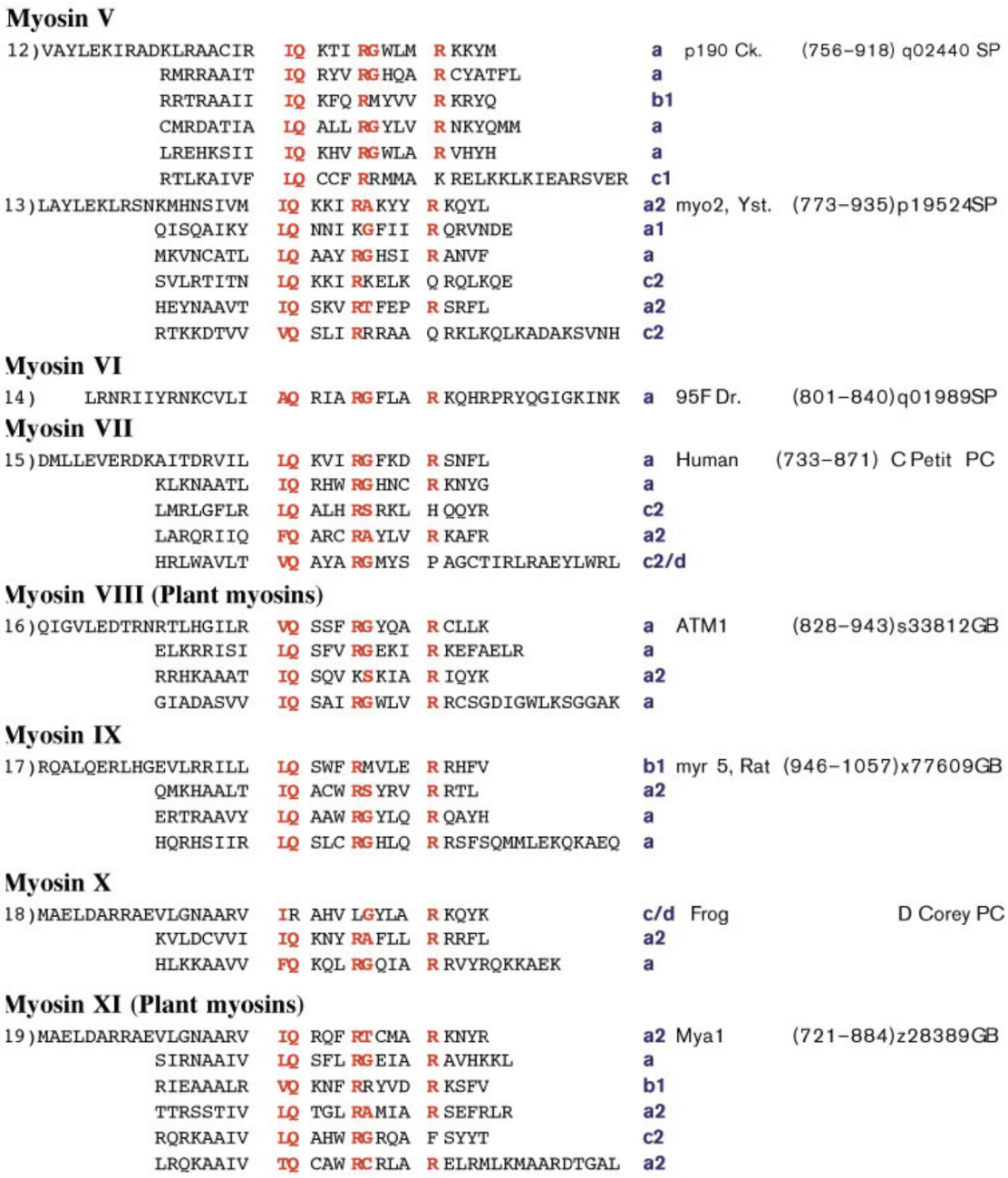


Figure 2 - Continued

The conserved residues in the core of the IQ motif are identified in red. Residue numbers for the sequences in each protein are given in parentheses. Accession numbers for GeneBank (GB) or Uniprot/SwissProt (SP) are given thereafter (PC implies personal communication to author). The letter code in purple after each sequence relates to the article text for the figure. Species abbreviations: Ac, Acanthamoeba; ATM1, Arabidopsis thaliana myosin 1; CdbH, human cardiac beta; Ck, chicken; Dct, Dictyostelium; Dr, Drosophila; HMWMI, high molecular weight myosin I; MM, mammalian myosin; Nm, non-muscle; Nom, neuromodulin; Ng, neurogranin; Sc, scallop; Sk, skeletal; Sm, smooth; and Yst, yeast.

A more general consensus sequence for the core, as seen in myosins, could therefore be written as:

[I,L,V]QXXXRGXXX[R,K,Q]XXΦ

A general pattern exhibited through the core of myosin IQ motifs is that two hydrophobic ‘patches’ occur. One is located in the N-terminal half of the helix, usually beginning four or five residues before the start of the core sequence, and the other (shorter) patch (~4-residues) occurs in the C-terminal half of the sequence, beginning at position 7. The hydrophobic residues are interspersed with positive residues, giving rise to a helix with hydrophobic and polar faces. The hydrophobic face is the target for CaM lobe binding.

Crystal structures of apo-CaM and other LCs bound to IQ motifs (Houdusse et al., 1996; Terrak et al., 2003; Terrak et al., 2005; Mukherjea et al., 2009), indicate that the two hydrophobic and polar faces in the heavy chain delineate the binding of the C- and N-lobes of the LC. The C-lobe always binds to the N-terminal half of the IQ-core (IQxxxR), while the N-lobe interacts with the C-terminal part (GxxxR). The C-lobe, which adopts the semi-open ‘gripping’ conformation, interacts most strongly, while the N-lobe is normally in a closed non-gripping conformation, interacting more weakly with the second half of the core, on the opposing face of the helix. The precise orientation of and interactions with the N-lobe is more variable, depending on the residues in the heavy chain sequence.

Inspection of a variety of IQ sequences (Figure 2) reveals that there is variation not only in the number of IQ motifs that different myosin types possess, but also in the amino-acid spacing between sequential IQ motifs. Counting inclusively from position 1 in the core sequence to the residue before the next core sequence starts, the two IQ motifs in myosin class 2 are always separated by 26 residues. In contrast, in the unconventional myosins are much more variable in the number of IQ motifs they possess, and the spacing varies from 22 to 36 residues (Table 1.1). The most common IQ-spacing is 23 residues, for example in the separation of the three IQ motifs in myosin 1c (human), IQs 1-5 of myosin 7a (human) and IQs 1-3 of myosin 10 (frog). 22-, 26- and 25-residue spacings also occur, though less often and are always flanked by 23-residue spacings. For example, myosin class 5 follow the spacing pattern 23-25-23-25-23, while myosin class 8 follows the pattern 23-26-23 (Table 1).

Myosin class/isoform	Number of IQ motifs	IQ spacing pattern
Myosin 1a	3	23-23
Myosin 1b	6	23-23-29-29-29
Myosin 1c	3	23-23
Myosin 1d	2	22
Myosin 1	1	N.A
Myosin 2	2	26
Myosin 3	1	N.A
Myosin 4	1	N.A
Myosin 5	6	23-25-23-25-23
Myosin 6	1	N.A
Myosin 7	5	23-23-23-23
Myosin 8	4	23-26-23
Myosin 9	1	N.A
Myosin 10	3	23-23
Myosin 11 (plants)	6	23-25-23-25-23

Table 1 – IQ motif spacing among various myosin classes.

Calmodulin (CaM) is a highly conserved calcium-binding protein expressed in all eukaryotic cells. It acts as a messenger protein; responding to changes in Ca^{2+} levels by binding or releasing Ca^{2+} cations and modifying its conformation and interaction with a wide variety of target proteins. It normally binds to myosin as a light chain in its apo- (Ca^{2+} free) form. CaM is formed from a 148 residue (16.7 kDa) polypeptide chain that folds into a dumbbell-shaped molecule with two

homologous globular domains, known as N- and C-lobes according to their position relative to the N-terminus (Figure 3). The lobes are joined by a flexible inter-lobe loop that allows the protein to ‘wrap around’ its target. CaM can bind up to four Ca^{2+} cations via four helix-loop-helix motifs known as EF-hands, two in each lobe (each separated by linkers).

While CaM is found to be the most common light chain in unconventional myosins (Heissler and Sellers, 2014), myosin can also bind essential light chain (ELC) and regulatory light chain (RLC). These and other myosin-binding LCs all share the same two-lobe dumbbell-like structure with EF-hands (Nelson and Chazin, 1998), though the ELC, RLC and other specific LCs lack most or all cation-binding capacity (Kawasaki et al., 1998). CaM and other LCs typically bind the IQ motif in Ca^{2+}-free conformation, known as apo-calmodulin (apo-CaM). At low Ca^{2+} concentrations, apo-CaM binds tightly to proteins such as neuromodulin, neurogranin, and to many of the unconventional myosins (Rhoads and Friedberg, 1997). An increase in the intracellular Ca^{2+} concentration can trigger a conformational change within CaM (Zhang et al., 1995), causing it to adopt an open conformation (Figure 3), which results in dissociation of the light chain from the IQ motif.

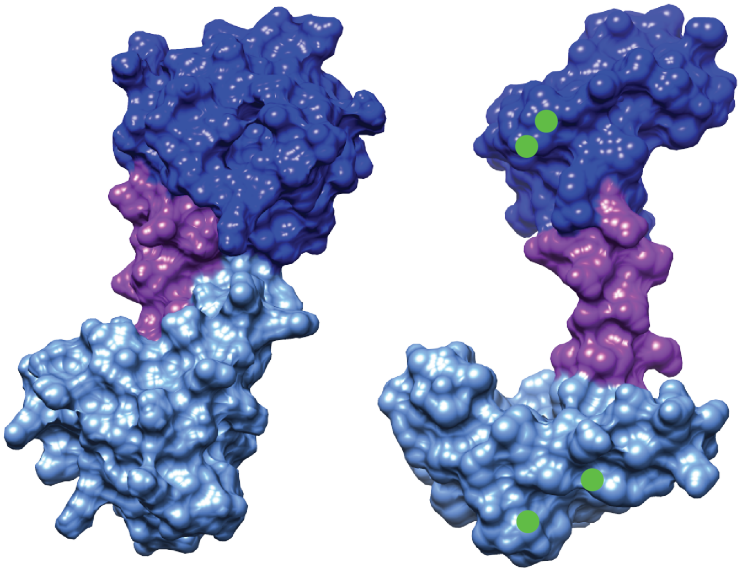


Figure 3 - Ca^{2+}-induced conformational change in calmodulin.
Ca^{2+}-free (left) and Ca^{2+}-bound (right) conformations of CaM. The structure on the left is from Xenopus laevis PDB: 1CFD, while the structure on the right is from Homo sapiens PDB: 1CLL. N- and C-terminal lobe regions are coloured light blue and dark blue, respectively, while the inter-lobe loop is coloured purple. On Ca^{2+} binding (binding depicted in green), the bundled lobes open out, exposing the non-polar amino acids in two grooves which are ready to grip the target protein.

Closer inspection of published crystal structures reveals IQ-CaM binding is not simply an on / off mechanism. Modelling work and analysis of apo-CaMs bound to part of the mouse myosin 5a lever, showed how apo-CaM can bind to a target by adopting a so-called ‘semi-open’ conformation (Figure 4a,b) (Houdusse et al., 1996; Houdusse et al., 2006). Here, despite the lack of Ca^{2+}, the C-lobe is partially open, forming a shallow groove with sufficient hydrophobic residues to accommodate residues in the first half of IQ motif core. The N-lobe remains closed and interacts less with the remainder of the IQ motif sequence. Crystal structures of apo-Mlc1p (a LC for the yeast class 5 myosin myo2p) bound to fragments of the myo2p IQ motif also exhibited the ‘semi-open’ C-lobe conformation (Terrak et al., 2003). However, the N-lobe is completely detached from the second half of the IQ motif (with no apparent interactions), with the LCs adopting a so-called ‘extended’ conformation, Figure 4c,d (Terrak et al., 2003).

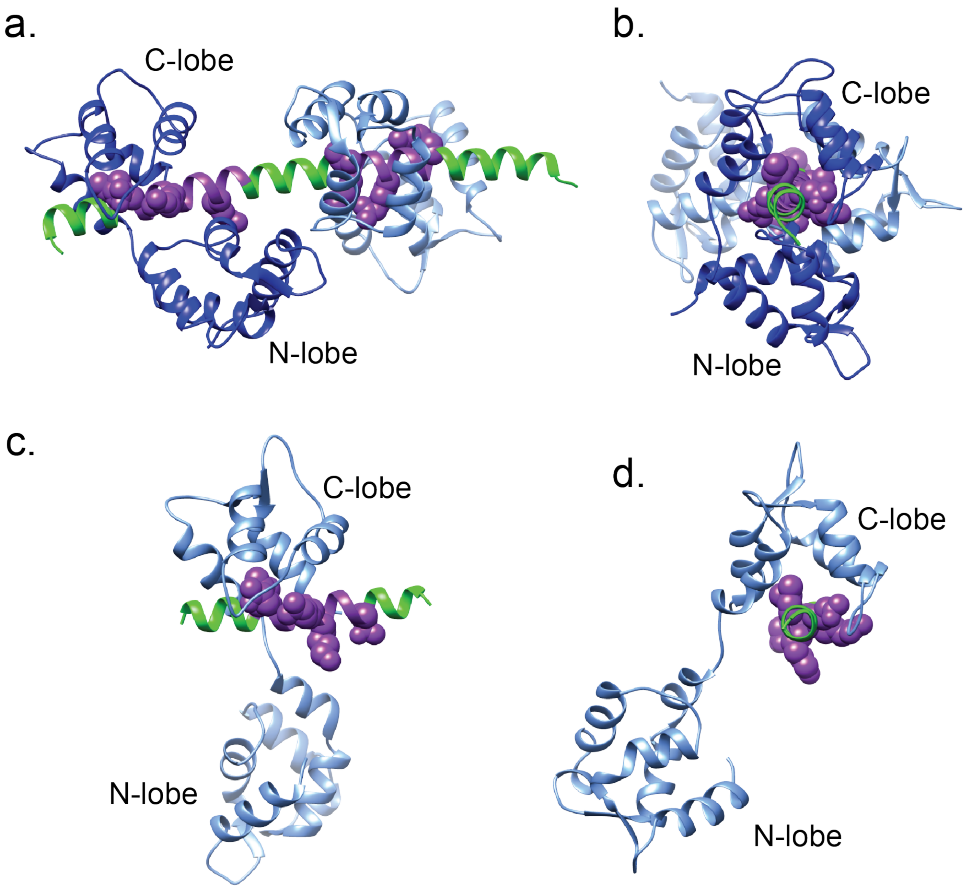


Figure 4 - apo-CaM bound to IQ motifs.

(a) ribbon depiction of apo-CaMs (Dark blue and light blue) bound to a fragment of mouse myosin 5a lever encompassing IQs 1-2. PDB: 2IX7. Heavy chain is shown in green with the IQ core sequence shown in purple and the conserved residues in space fill representation. (b) view of the same structure as in (a) looking end-on down the N-terminus of the lever helix. (c) Mlc1p light chain (Light blue) bound to the fourth IQ motif of the myo2p heavy chain (green). PDB: 1M46. IQ-core in purple with space fill representation of conserved residues. (d) same as (c) but viewed looking at the N-terminal end down the lever helix. Due to a bulky lysine at position 7 in the IQ-core and lack of a positively charged amino acid at position 11 (normally arginine) the N-lobe in the lower structures is detached from the C-terminus of the IQ motif. In the upper structures it makes interactions with the lever helix.

1.6.2 SAH domains.

Some myosins, notably myosins 6, 7a, and 10 from mammals, and MyoM from *Dictyostelium*, were originally suspected to contain coiled coil regions within their sequences. However, closer inspections of the sequences revealed these sequences to contain a large number of charged residues and lack of a hydrophobic seam (Knight et al., 2005). It has since been shown that these domains do not form coiled coils, but instead contain stable SAH (single α-helical) domains, and that all three of the mammalian myosins are monomeric (Lister et al., 2004; Yang et al., 2009; Umeki et al., 2011). Myosin 10 was first shown to contain an SAH domain instead of a predicted coiled coil (Knight et al., 2005). A peptide (37 amino acids long) from the N-terminal of the predicted coiled coil was shown to be almost 100% helical and monomeric in solution. In addition, a construct of myosin 10 containing the motor, lever and the entire predicted coiled-coil sequence was shown to be mostly monomeric *in vitro*. Importantly, even in the rare dimers, dimerisation involved only the extreme C-terminus of this construct. A SAH domain, approximately 15 nm long (equivalent to 100 residues at a rise of 0.15 nm per residue, as observed for α-helices) was observed in both monomers and dimers (Knight et al., 2005). It was also suggested in this report that the predicted coiled coil of myosins 6 and 7, and that for the Dictyostelium myosin MyoM, were also likely to contain a SAH domain from sequence homology between the predicted coiled-coil domains of these myosins. We now know that the predicted coiled-coil domains of all of these myosins forms a SAH domain (Spink et al., 2008), (Li et al., 2017), as predicted (Yang et al., 2009; Baboolal et al., 2009).

SAH domains are unusual in that these domains are highly stable in solution in isolation. In contrast, other α-helices are usually not stable in isolation but need to be stabilised by their interaction with other parts of the same protein or by binding to other proteins (as is the case for the α-helical lever in all myosins, which are stabilised by binding to light chains/CaM). SAH domains are rich in charged residues Glu, Arg, and Lys, and this allows the formation of an ionic bond network, allowing these structures to remain stable and helical over a wide range of pH and salt concentration when isolated in solution (Wolny et al., 2014). These salt bridge pairs both act as ‘staples’ to resist unfolding of the helix and to screen the hydrogen bonds between the polypeptide backbone amide

groups from attack by water (Garcia and Sanbonmatsu, 2002; Ghosh et al., 2003; Sivaramakrishnan et al., 2008). The side chains of arginine and lysine residues preferentially interact with glutamate side chains four residues downstream (*i*,*i*+4) or three residues upstream (*i*,*i*−3), whereas glutamate preferentially interacts with arginine and lysine residues at (*i*,*i*+3) and (*i*,*i*−4) (Batchelor et al., 2017).

Why these myosins contain a SAH domain is still a mystery. Experiments in which part of the IQ+CaM lever of myosin 5 is replaced with a similar length of SAH domain has shown the motor to retain its processivity (Baboolal et al., 2009), suggesting the SAH domain can replace the function of the canonical lever. It has also been shown the SAH domain unfolds under very low forces (<30 pN), which may allow myosins to perform its function in extremely crowded environments during its powerstroke (Wolny et al., 2014). As such, the bending stiffness of the SAH domain is expected to be much lower than that of the IQ+CaM lever domain (Baboolal et al., 2009). It has also been noted that the SAH domain can act as a constant force spring, whereby the force applied does not increase with extensibility (Wolny et al., 2014). However, how this behaviour affects the overall function and dynamics of the myosin has yet to be determined.

1.7 Myosin Tail.

The C-terminal tail follows the lever and is the most variable domain in terms of sequence, length, subdomain composition and organisation. The myosin tail determines its functional role and specificity, as it can target the myosin to a particular cellular location and determine its function e.g. through binding to different cargoes or binding partners or associating with other myosin tails to form filaments such as in myosin 2. The tail can also be involved in regulation, where the tail can fold back and contact the motor domain in absence of cargo, shutting the molecule down in a stable folded conformation as seen in myosins 2, 5 and 10 (Thirumurugan et al., 2006; Jung et al., 2008; Umeki et al., 2011).

Another key property of the tail is that it determines whether the myosin is monomeric or dimeric (Figure 5). Some myosins such as myosin 2 and myosin 5 contain long α-helical tail sequences featuring heptad repeats. These are 7-residue sequences labelled *abcdefg*, in which hydrophobic residues

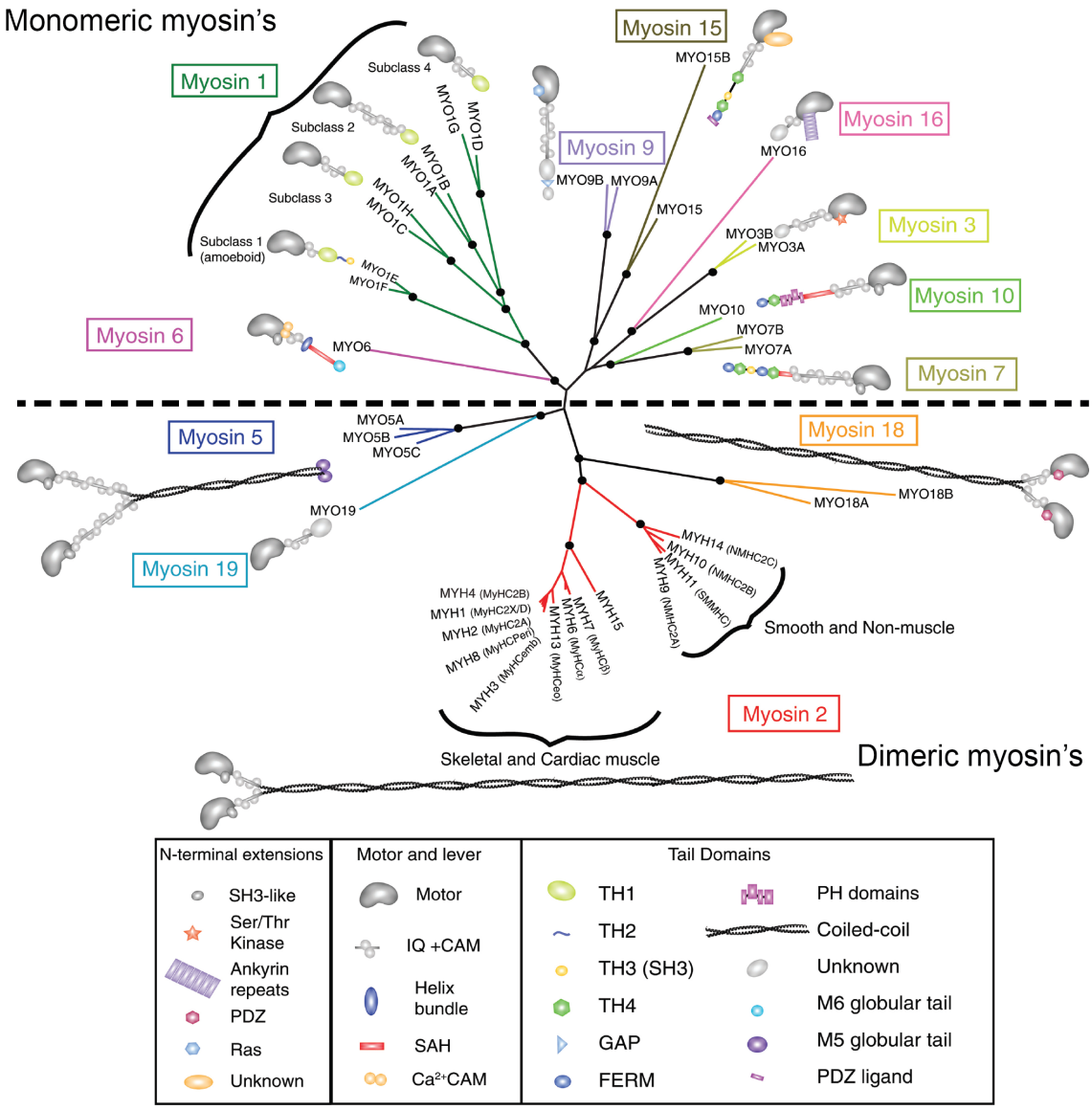


Figure 5 - Myosin tail domain diversity.

Image adapted from Peckham and Knight (2009). The myosin family tree in humans. The 39 genes encoding myosins in humans are organised into 12 different classes. Only 3 of the myosin classes contain genuine coiled-coil domains in their tails that dimerise them (myosins 2, 5 and 18). Monomeric myosins tend to contain a much more diverse C-terminal tail in terms of sequence, length, subdomain composition and organisation. The remainder of the myosins have a variety of domains in their tails as indicated by the key at the bottom of the figure. Myosins also have a variety of N-terminal extensions to their motor domains.

consistently occur at positions *a* and *d* (Lupas et al., 1991). The *a* and *d* residues form a hydrophobic core that slowly winds around the surface of the α-helix. When two such helices meet, the hydrophobic stripes pack against each other to form a coiled-coil (Parry, 1982; Lupas et al., 1991). This is the main form of dimerisation seen within myosins. Coiled-coil tail sequences that lead to dimerisation in this way are observed in myosins 2, 5 and 18. A small sequence immediately downstream of the SAH in Myosin 10 has been suggested to form an anti-parallel coiled coil (Lu et al., 2012), however deletion of this domain does not appear to affect formation and movement of filopodia, driven by this myosin (Baboolal et al., 2016).

1.8 Actomyosin interaction and swinging lever model.

As discussed earlier, myosins are able to translocate along tracks of actin. The binding of myosin to actin and the conformational changes associated with steps in the ATP hydrolysis cycle were first elucidated in the context of muscle contraction. Electron microscopy showed linkages, called cross-bridges, between the thick and thin filaments of the sarcomere, which were subsequently shown to originate from the thick filaments and identified as the heads (which correspond to the subfragment-1 (S1) region) of individual myosin molecules (Huxley, 1957; Huxley, 1963). Later it was proposed that the heads make cyclic interactions with the actin and undergo a conformational change that is fuelled by ATP hydrolysis (Bozler, 1953). This became known as the swinging-cross-bridge theory. The actomyosin cross-bridge cycle (Figure 6) is common to all myosins. It is now known that in myosin 2, and unconventional myosins, the movement of the lever, linked to the ATP cycle is responsible for force and movement, known as the powerstroke (Figure 6). The lever amplifies small (angstrom-sized movements) within the motor into larger (nm-sized) movements within the tail, as such, the lever needs to maintain its structure as it moves and when under load in order to propagate these movements (Oke et al., 2010).

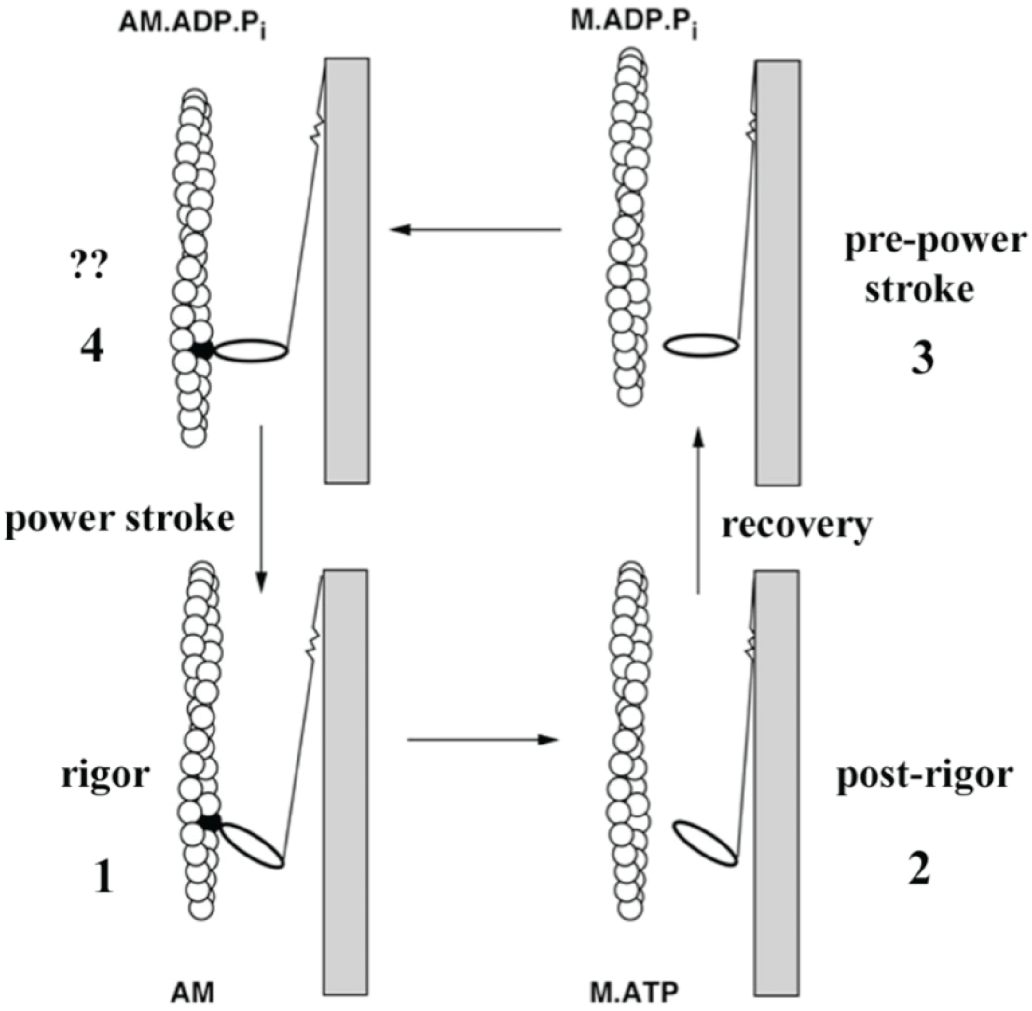


Figure 6 - The Lymn-Taylor cross-bridge cycle.

Figure taken from Geeves and Holmes (2005). A = actin, M = myosin, AM = actomyosin. The S1 domain (i.e. motor and lever) is depicted as a single lozenge-shaped domain, connected via its tail (a line) into the thick filament backbone (grey). The head cyclically interacts with the actin filament (spheres) and undergoes a powerstroke (a rotation about its contact point with actin) during its powerstroke. This is the force-generating step that causes the filaments to slide against each other within the sarcomere. ATP binding frees the head from actin and hydrolysis is associated with a recovery stroke, where the head rotates back so that it is primed to undergo another cycle. Details of the ATPase cycle are further explained in Figure 7.

In the case of a processive cargo-carrying myosins, such as myosin 5a, the movement of the lever is also coupled to the coiled-coil tail and its bound cargo. The resistance excerted by the cargo during the powerstroke (due to the viscosity or obstacles in the cytoplasm) means the lever must remain rigid enough to complete its powerstroke and find its next binding site. If the lever were pliant and buckled under load during its swing, then it would not be able to transmit the force generated in the motor domain to the tail domain (Nguyen and Higuchi, 2005). For this reason, the flexural rigidity (bending stiffness) and the specific mechanical properties of the lever are crucial to myosin function.

In some cases, it can be advantageous for the lever to bend or deform to some extent. In the general powerstroke model for myosin, the motor needs an elastic element (like a spring) which can store mechanical energy (Howard and Clark, 2002). When the motor undergoes its conformational change, strain is stored in this ‘internal spring’ and the restoring force is the motor force. Relief of the strain is the driving force for the motion, the powerstroke. In myosin the spring has not been definitively located in the structure of the protein. Indeed, elastic energy may be stored in more than one structural element. Flexing of the lever was suggested as one possible source of elasticity (Howard and Spudich, 1996). Alternatively, the elasticity may be due to deformation of the converter domain, or due to bending at a pliant region at the junction of the lever and converter. Evidence for the latter has been seen in crystal structures of scallop myosin 2 (Houdusse et al., 2000). In this case the internal spring would be more analogous to a torsion spring, which provides a restoring torque proportional to the angle it is rotated away from its equilibrium position.

Given the importance of the mechanical properties of myosin levers to function, it is interesting that lever structure is variable across different myosin types. This diversity suggests possible evolution of structure for different mechanical functions. One obvious difference is the difference in the length of the lever. Increasing the length of the lever, e.g. by addition of more LC-binding IQ motifs, increases the working distance. While class 2 myosins possess only 2 IQ motifs, myosin 5a and 5b have one of the longest levers with 6. The linear relationship between step size and lever length for myosin 5 has been shown in many studies (Köhler et al., 2003; Sakamoto et al., 2003; Moore et al., 2004). Interestingly, the myosin 1b lever is alternatively spliced in different tissues to give 3 isoforms with

4, 5 and 6 IQ motifs respectively. It has been suggested that the differently spliced levers confer different mechanical properties on the protein, important in different force-sensing roles (Laakso et al., 2010).

As well as the lever length, the spacing at which the IQ motifs appear along the lever also varies across myosins, ranging from 22 to 36 amino acid repeats and as described above (Houdusse et al., 1996). Precise IQ motif spacing is important as it dictates the relative orientation and so potential interactions between neighbouring LCs along the lever. More recent work has also shown that for some unconventional myosins the typical LC-IQ motif building block of the lever is varied by other structural components. For example, myosin 10 has a stable single α-helix (SAH) directly following its 3 IQ motifs which is thought to extend its lever and the working distance (Knight et al., 2005), though the flexural rigidity of the SAH is predicted to be lower than the LC-IQ motif section (Baboolal et al., 2009). Myosin 6 has perhaps the most unconventional lever, beginning with a unique bent insert sequence that binds a CaM with Ca^{2+} bound, followed by a conventional LC-IQ motif subunit followed by a 3-helix bundle (that may unfold as a lever extension (Mukherjea et al., 2009)) and then a SAH, which may further increase the length of the lever (Spink et al., 2008).

1.9 ATPase cycle.

All myosin isoforms studied, with the exception of myosin 18 (Guzik-Lendrum et al., 2013), are ATPases that catalyse the breakdown of ATP into ADP and Pi. In myosin, the substrate is MgATP and the ATPase activity is activated by binding to actin. Binding of myosin to actin catalyses the release of Pi, whereas the release of myosin from actin is catalysed by the binding of ATP. In this way, myosin's hydrolysis cycle (Figure 7) is tightly coupled to its cycle of attachment to and detachment from actin. The events in the actomyosin ATPase cycle appear to be common to all myosins known, though the rates between steps in the cycle vary widely depending on the specific myosin, which is linked to its function (De La Cruz and Michael Ostap, 2009).

The duty-ratio of a given myosin (defined as proportion of time a myosin spends strongly bound to actin during its hydrolysis cycle (Hackney, 1996) can vary greatly, and is an indicator of whether the myosin is processive (meaning it can make multiple interactions with the track per encounter, and thus move

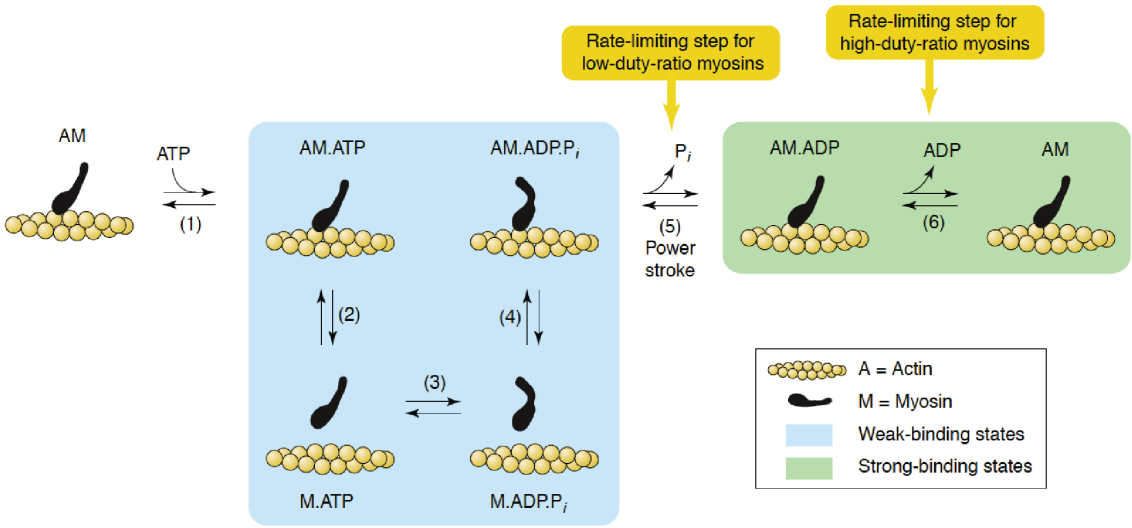


Figure 7 - The actomyosin ATPase cycle.

Figure taken from La Cruz and Ostap (2004). M = myosin, AM = actomyosin. In the absence of ATP, myosin binds actin tightly forming the 'strong' or 'rigor' complex. ATP binding to the myosin head (step 1) weakens the affinity and causes the head to dissociate from actin (step 2). ATP is then hydrolysed to form the stable products complex (M.ADP.Pi) and there is an accompanying conformational change known as the 'recovery stroke', where the lever swings and the head is primed ready to deliver its powerstroke. Myosin then re-binds actin with products bound (step 4). Pi is released from the binding pocket which is accompanied with the force-generating powerstroke (the lever swinging back again) (step 5). Myosin now occupies a strongly bound state again. ADP is released (step 6) and the cycle can repeat.

distances greater than a single ATP-driven step). For example, mammalian myosin 5a is a highly processive two-headed molecule, capable of taking multiple steps along actin before detaching (Mehta et al., 1999). The duty ratio of this myosin is very high ~0.7 (De La Cruz et al., 1999). In contrast, non-processive myosins undergo only a single cycle before detaching and cannot support movement over long distances. Skeletal muscle myosin 2 has a low duty-ratio, spending very little time during its cycle bound to actin (El-Mezgueldi and Bagshaw, 2008). This allows myosin 2 to function properly in large un-synchronised ensembles of motor domains along a thick filament. In general, for myosins with low duty-ratios, the release of Pi is the rate limiting step in the hydrolysis cycle, while for high duty-ratio myosins, the rate limiting step is ADP release (Figure 7).

1.10 Myosin function.

Myosins are known to have a variety of functions. Muscle and non-muscle myosin 2 form bipolar filaments that consist of dozens of myosin molecules. Myosin 5 is an unconventional myosin involved in the transport of organelles, membranous cargo, secretory vesicles, mRNA, lipids and proteins vesicles on actin tracks (Reck-Peterson et al., 2000). Myosin 6 is another cargo transporting myosin that contains a unique insert in its converter region that switches this myosin into reverse gear, enabling it to walk to the pointed (or minus), rather than the barbed end of actin filaments (Wells et al., 1999). Myosin 6 has been implicated in hearing (Avraham et al., 1995), in the development and maintenance of the stereocilia in the inner ear in mice (Self et al., 1999), in spermatogenesis in *C.elegans* (Kelleher et al., 2000), and in a wide variety of roles in *D. melanogaster*, including oogenesis, cell motility, and spermatogenesis. In mammalian cells, myosin 6 is involved in endocytosis (Buss et al., 2001; Aschenbrenner et al., 2003), in the maintenance of Golgi complex morphology and secretion (Warner et al., 2003), and in membrane ruffling (Buss et al., 1998).

1.11 Myosin Regulation.

The regulation of myosin is tightly controlled. Unregulated myosin activity would lead to the breakdown of many cellular pathways, and rapidly use ATP in cells. The regulation of myosins occur at three different levels. First, transcription

controls when and where myosin is expressed, together with switching between alternate splice isoforms in response to intra- and extracellular stimuli. The second and third levels of regulation includes substrate or effector binding and the regulation by posttranslational modifications (PTMs), respectively (Heissler and Sellers, 2016). Interestingly, no universal myosin regulation pattern exists: myosins from one class, which are highly conserved at the levels of primary sequence and tertiary structure between protozoa and vertebrates exhibit multiple and entirely different regulatory aspects. This paradox underlines that evolution does not create new enzymes within a superfamily but rather new allosteric modes of protein regulation (Gunasekaran et al., 2004). The various themes observed in myosin regulation are discussed below.

1.11.1 Light chain phosphorylation.

Many different proteins can be regulated by phosphorylation (Sellers, 1985). Myosin can be regulated by phosphorylation on both heavy and light chain. In class 2 myosins, phosphorylation of the regulatory light chain (RLC) in response to upstream signaling pathways is the major regulatory mechanism of vertebrate smooth and nonmuscle myosin isoforms, but not sarcomeric (skeletal and cardiac) muscle isoforms (Heissler and Manstein, 2013). Phosphorylation of Ser-19 of the smooth and nonmuscle myosin-2 RLC by kinases such as myosin light chain kinase (MLCK) is associated with increased (>1000 fold) actin-activated ATPase activity and allows myosin-2 to move actin in motility assays (Scholey et al., 1980). In contrast, RLC phosphorylation of mammalian sarcomeric myosins-2 has a minor impact on its enzymatic features but has modulatory effects on the mechanical performance of a muscle fiber (Kamm and Stull, 2011). In cardiac muscle, RLC phosphorylation is associated with increased myocardial performance and enhances myosin motility under loaded conditions *in vitro* (Karabina et al., 2015). Electron microscopy studies show that the motor domains of unphosphorylated skeletal and cardiac muscle myosin-2 filaments are held in a ‘closed’ position close to the filament backbone in a highly ordered manner. When the RLC is phosphorylated, the motor domains connected by the proximal coiled-coil regions move away from the surface of the filament and would be more likely to interact with actin (Levine et al., 1996).

Phosphorylation of non-muscle myosin-2 heavy chain in the coiled-coil tail and the non-helical tailpiece (NHT) have also been found to be regulatory targeting

mechanisms implicated in the inhibition of filament formation and the interaction with binding partners (Dulyaninova et al., 2005; Liu et al., 2012). Studies from amoeba myosin 2, have shown multiple serine and threonine residues in the tails are phosphorylation sites for myosin heavy chain kinases (Kuczmarski and Spudich, 1980; Collins et al., 1982). In *Acanthamoeba*, phosphorylation of serine residues within the NHT regulates filament assembly by inhibiting the dimerisation of myosin monomers and the assembly pathway of dimers into filaments (Liu et al., 2012).

1.11.2 Effect of Ca^{2+} binding to calmodulin: effects on the lever.

Changes in lever arm mechanics induced by Ca^{2+} have been described for myosin-5a and myosins 1a and c. In myosin 1c, binding of Ca^{2+} to the calmodulin on the 1st IQ motif dissociates calmodulin, introducing flexibility into the lever (Adamek et al., 2008). This is called the clutch model: when calmodulin is bound, the clutch is engaged, and when is dissociates, the clutch is disengaged, resulting in idling of the motor. Ca^{2+} binding to the calmodulin on the 1st IQ motif of myosin-5a also results in its dissociation, decreasing the mechanical stiffness of the lever, which impairs communication between the motor and the lever (Trybus et al., 2007). This loss of mechanical compliance reduces myosin motility in the presence of Ca^{2+} and this can be restored by adding excess calmodulin in *in vitro* assays (Trybus et al., 2007).

In myosin 7, Ca^{2+}-calmodulin interactions have been shown to induce conformational changes in the heavy chain at the distal IQ motif, which attenuates the interaction of the lever with the tail (Umeki et al., 2009), thus abolishing the tail-induced inhibition of myosin motor function (Umeki et al., 2009).

1.11.3 Regulation by autoinhibition.

Myosin activity is also commonly regulated by motor-tail interactions, as demonstrated for class 2, 5, 7 and 10 myosin isoforms. *S*mooth and non-muscle myosins-2 exist in a three-state equilibrium *in vitro* between a compact, autoinhibited conformation (10S), an extended conformation (6S), and filaments (Trybus et al., 1982; Craig et al., 1983). Electron micrographs of myosin-2 in the 10S conformation show that the two motor domains interact in an asymmetric manner and the tails are sharply bent at two locations to allow the tail to interact with and stabilise the motor-motor interaction (Craig et al., 1983; Wendt et al., 1999; Jung et al., 2011). The conformational restriction imposed by the folding of

the tail around the motor allosterically shifts the myosin into a kinetically inert and assembly-incompetent state. ATPase activity is restored upon phosphorylation of RLC at Ser-19, which promotes filament formation and an active molecule in the presence of actin.

Myosin-5a can either form an extended (14S) active conformation or an autoinhibited conformation (11S) (Wang et al., 2004; Li et al., 2004). Electron micrographs show that the autoinhibited myosin-5a has a triangular structure, with the two motor domains docked against the globular tail domains (GTD) (Liu et al., 2006; Thirumurugan et al., 2006). Binding of cargo, such as melanophilin to the myosin GTD shifts the equilibrium towards the extended, active conformer. The ability of myosin 5a to form the regulated conformation prevents futile ATP consumption under cargo-free conditions (Li et al., 2005).

Myosin-10 adopts an autoinhibited conformation in which the MyTH4/FERM and PH (pleckstrin homology) domain of the tail intramolecularly docks against the motor domain to form a tightly folded molecule with low actin-activated ATPase activity. This autoinhibited state is relieved upon the binding of PtdIns(3,4,5)P_3 to the PH domain, allowing the motor to form a catalytically active molecule (Umeki et al., 2011).

1.12 Myosin 7a.

Two class 7 myosin genes exist in humans. The myosin 7a and myosin 7b heavy chains are encoded by *MYO7A* and *MYO7B*, respectively, and display distinct kinetic and functional properties. The human myosin 7a heavy chain is composed of the N-terminal motor domain followed by a neck region with 5 IQ motifs, and then a SAH domain (Li et al., 2017). The tail region contains a myosin tail homology 4 (MyTh4) domain and band 4.1-ezrin-radixin-moesin (FERM) domain, followed by an SH3 domain, and then a second, C-terminal MyTH4-FERM tandem domain (Figure 8) (Hasson et al., 1995). Myosin 7b differs from myosin 7a by possessing a short insertion in its motor domain, which may promote its actin binding (Henn and De La Cruz, 2005).

Figure 8 - Myosin 7a domain structure.

Domain analysis of Human myosin 7a according to uniprot (Q13402). Number on the right indicates the number of residues within the protein.

Myosin 7b also apparently lacks a SAH domain distal to the IQ region. Myosin 7a associates with lysosomes and may be involved in lysosome trafficking (Soni et al., 2005). However, myosin 7a has been shown to play an essential role in both the development and physiological functions of the auditory and visual systems in mammals (Self et al., 1999; Wolfrum and Schmitt, 2000). It is also found in kidney, liver, testis, cochlea, lymphocytes (Gibbs et al., 2010).

In the visual system, myosin 7a is expressed in photoreceptor cells and retinal pigmented epithelial (RPE) cells and is essential for proper cilium formation as well as melanosome transportation in photoreceptors (Liu et al., 1997). Melanosomes in the RPE cells undergo light cycle–dependent movement. After light onset, there was a significant increase in the number of melanosomes in the apical processes, and it is thought that myosin 7a is required for the proper movement of melanosomes (Futter et al., 2004).

In the auditory system, myosin 7a, together with its associated proteins, is essential for the development and structural maintenance of stereocilia. These are actin-based protrusions at the apical surface of cochlear and vestibular hair cells in the inner ear, responsible for sound perception and spatial balancing respectively (Gillespie and Müller, 2009; Pan and Zhang, 2012). Within the hair cells, myosin 7a is found in the upper tip link density of the stereocilia (Grati and Kachar, 2011a), (Figure 9a – c). Hair cells are connected by tip links, which connect the top of a lower stereocilia, where mechano-electrical transduction (MET) channels are localised, to the side of an adjacent, upper tip link density; which are thought to contribute to resting tension and regulate adaptation. The upper and lower tip links are connected by cadherin 23 and protocadherin 15 (Yu et al., 2017). These links between the tip link densities are important for transmitting force to the mechano-electrical transduction (MET) channels that convert sound waves into an electrical signal (Hudspeth, 2014).

When the stereocilia are stimulated (i.e. deflected by sound waves), the hair cell bundles are deflected and the MET ion channels open, resulting in an influx of calcium causing a signal cascade that terminates with the perception of sound. Following this, the hair cells undergo a process called adaptation, which involves myosin 7a, to reduce the electrical response of the hair bundle, thus preventing saturation and ensuring that the bundle remains

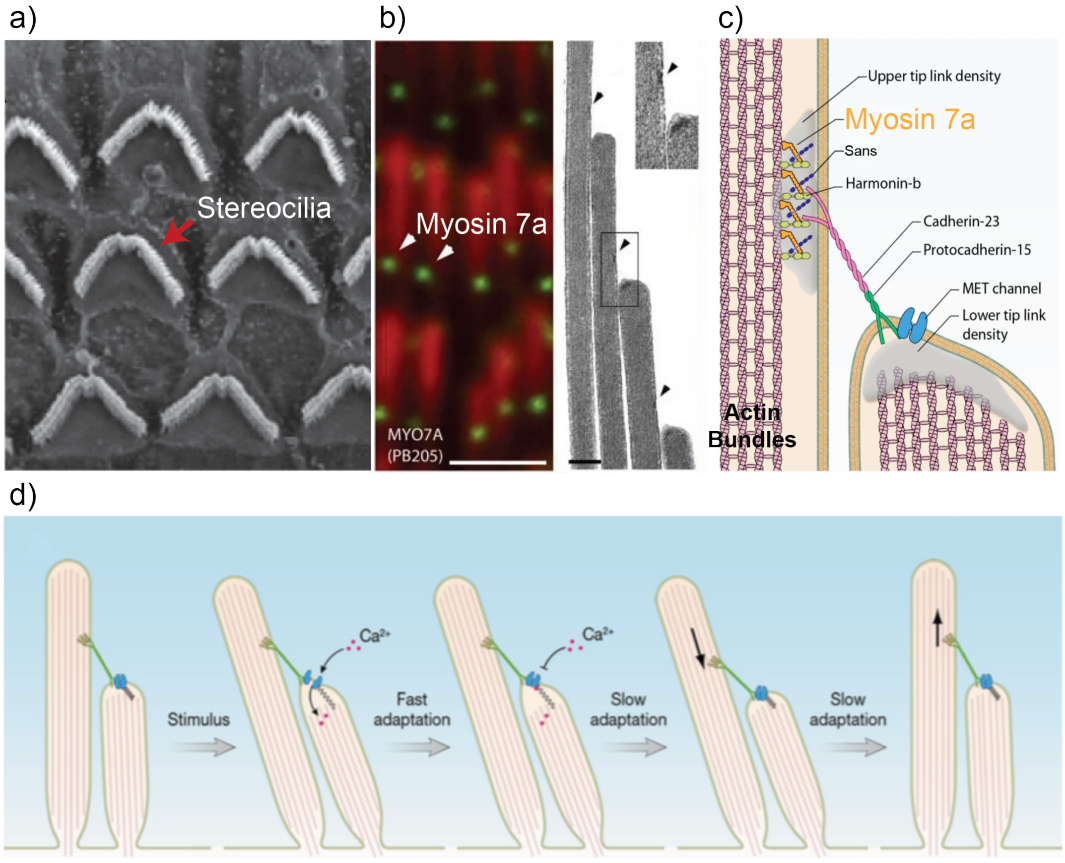


Figure 9 - Myosin 7a function.

Figure adapted from Grati and Kachar (2011). (a) Scanning electron microscopy image of hair cells from the inner ear. Myosin 7a is found in the stereocilia (red arrow). (b) Immunofluorescence confocal images of myosin 7a (green, arrowheads) in vestibular hair cells of guinea pig showing labelling. Stereocilia are counterstained in red with rhodamine phalloidin. (D) Myosin 7a is found in the Upper tip link density (arrowheads and Inset) in a thin section electron micrograph of a vestibular stereocilia bundle. (c) Schematic diagram of the stereocilia interaction network. Myosin 7a is bound to actin filaments at the Upper tip kink density. Myosin 7a is coupled to a number of proteins, including Sans, Harmonin-b, and Cadherin-23 which are coupled too ion channels at the Lowe tip link density. (d) Following deflection of the stereocilia by sound waves (stimuli), a signal cascade event is initiated known as the fast adaption response. Myosin 7a is suspected to restore the tension between the stereocilia to prepare for the next stimuli.

sensitive to further stimuli (Gillespie and Müller, 2009). The cadherin-based connections are essential for the formation of mechanically integrated bundles that must withstand shear stresses imposed on them. This is achieved by anchoring the cytoplasmic domains of cadherins to the actin cytoskeleton via a complex of myosin 7a and other adapter proteins (Yu et al., 2017). Additionally, it is suspected that myosin 7a can restore tension (by translocating along actin filaments) to the tip links (Grati and Kachar, 2011a)(Figure 9d).

Myosin 7a has been shown to have a high duty ratio (0.9), higher than that of myosin 5a (0.7) (Yang et al., 2005; Watanabe et al., 2006). A dimeric myosin 7a is thus expected to be processive which has been confirmed by dimerising two myosin 7a motor domains by a leucine zipper motif at the C-terminal end of the SAH domain, resulting in processive motion in the single-molecule TIRF motility assay (Yang et al., 2005). However, myosin 7a has been shown to be a monomer (Yang et al., 2009). It has been suggested that this motor could dimerize if brought into close proximity on actin, as apparently shown for myosin 6, or it could be dimerised by dimeric proteins that interact with the tail as also postulated for myosin 6 (Lister et al., 2004; Phichith et al., 2009). Many binding proteins have been described for mammalian myosin 7a (El-Amraoui, 2005). However, it is as yet unknown if any of these proteins dimerise myosin 7a.

The *in vitro* velocity of myosin 7a show that it moves slowly along actin (11 nm S^{-1}) (Sato et al., 2017) compared to myosin 5 (250 – 400 nm S^{-1}) (Mehta et al., 1999; Rief et al., 2000). As a monomer, it might best function to tether cargo to actin. Some myosins such as myosin 1 (Laakso et al., 2008) appear to act as molecular force sensors which stall when subjected to higher forces and remain attached to actin with little expenditure of ATP (Laakso et al., 2008). Given its high duty ratio and slow velocity, it is possible that myosin 7a may also work in this way. Alternatively, myosin 7a might function as an actin-bundling protein via the ATP-sensitive binding site in the motor and the ATP-insensitive site located in its tail or it might link actin via its motor domain to membranes via its FERM domain (Chishti et al., 1998). Clearly, there are many functional possibilities for myosin 7a.

1.13 Myosin 7 in disease.

Mutations in myosin 7a cause Usher syndrome type 1B (USH1B) (Well et al., 1995) and two forms of non-syndromic deafness, DFNB2 and DFNA11 (Hasson et al., 1995; Weil et al., 1997). Usher syndrome is responsible for the majority of deaf-blindness in humans and is inherited in an autosomal recessive pattern. Many of the mutations identified in patients are hypothesised to alter its structural stability or affect its interactions with binding partners (Wu et al., 2011; Pan and Zhang, 2012). The murine orthologue of *MYO7A* is causative for the *shaker-1* phenotype (Gibson et al., 1995). A large number of Usher syndrome missense mutations are found in the MyTH4-FERM domains, which bind various cargoes such as USH1 and accessory proteins (Grati and Kachar, 2011a). Mutations in myosin 7a can result in the improper localisation of USH1 proteins at the stereocilia in hair cells (Richardson et al., 2011). The inability to correctly localise and / or act as a force tether prevents these protein network scaffolds from functioning correctly, which ultimately leads to deaf-blindness.

1.14 Structure.

1.14.1 Motor.

Like all myosins, the myosin 7a motor domain contains a nucleotide binding site and an actin binding cleft. Motor activity is stimulated by F-actin with a K_{cat} of 4.3 s^{-1} and with 7 μM actin required for half-maximal activity (k_m) (Udovichenko et al., 2002). In a sliding filament motility assay, myosin 7a moved actin filaments with a velocity of 11 nm s^{-1} (Sato et al., 2017). While a crystal structure of the motor domain for myosin 7a is lacking, it does share ~40% sequence homology with the myosin 2 motor domain.

1.14.2 Lever.

Distal to the motor and proximal to the tail domain is the lever, which contains 5IQ motifs, suggesting it can potentially bind up to five LCs, and a SAH domain. Very little about the structure of this lever is known. However, the structure of the 5^{th} IQ motif bound to calmodulin, and the distal SAH domain (~78 residues) has recently been solved (Li et al., 2017).

Examining the sequence of each of the myosin 7a IQ motifs reveals that only the second IQ motif, fits the canonical IQ motif consensus sequence, while the others approximate as IQ motifs to varying degrees (Figure 10a). It is still not clear if all

5 IQ motifs are occupied in this myosin, or which specific light chains are bound. Immunoprecipitation experiments against endogenous myosin 7a obtained from bovine retina demonstrated that calmodulin binds the heavy chain in a Ca^{2+}-dependent manner (Udovichenko et al., 2002). Densitometry measurements suggested an average of 3.2 mols of calmodulin per heavy chain of myosin 7a. Other groups have also observed sub-stoichiometric binding of calmodulin to the myosin 7a heavy chain, with similar estimates of ~3 calmodulins binding each HC (Yang et al., 2009; Umeki et al., 2009).

The most intensive study examining light chain binding to Myosin 7a used a variety of GST-IQ fusion constructs (Figure 10 and (Sakai et al., 2015)). Pull down experiments in the presence of 1 mM EGTA (absence of Ca^{2+}) or 0.1 mM $CaCl_2$, (presence of Ca^{2+}) were performed. In the absence of calcium; IQ motifs 2 and 4 were occupied by calmodulin, while in the presence of calcium; IQ motifs 1, 2 and 5 were occupied. IQ motif 3 was never occupied. This study also tested binding of RLC and ELC, and found that while RLC did bind to the IQ motifs, CaM was the preferred LC. In contrast to this study, the crystal structure of IQ5 and the downstream SAH domain in the presence (5WSV) and absence (5WSU) of calcium (Figure 11) (Li et al., 2017) showed that calmodulin was bound in both cases, although the N-terminal region of calmodulin altered its conformation. It was shown that in the absence of calcium, the CaM adopted a canonical antiparallel orientation on the IQ motif, while in the presence of calcium, the N-lobe of the CaM dissociated from the C-terminal region of the 5^{th} IQ motif and adopted a parallel conformation on the IQ motif, with the N-lobe of CaM interacting with residues upstream of the 5^{th} IQ motif core sequence (Figure 11).

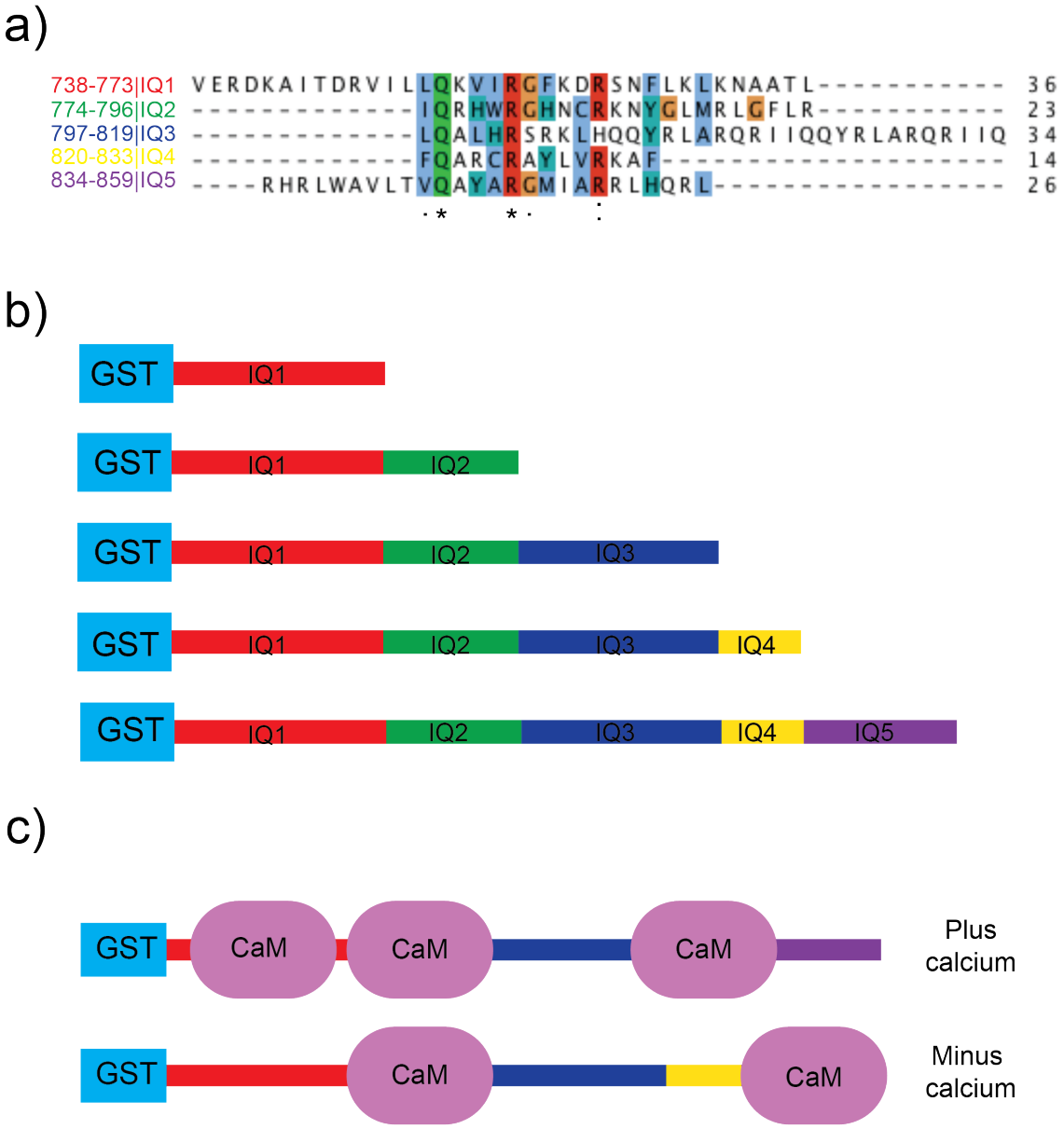


Figure 10 - Sakai et al (2015) Myosin 7a IQ constructs.

(a) sequence alignment of the IQ motifs added to each construct. The length of each construct is shown right. Residues are coloured according to their physiochemical properties which represent small hydrophobic residues (blue), basic residues (red), hydroxyl (green), ring (teal), and amine side chains (orange). (b) GST constructs generated by Sakai et al., (2015) with addition of IQ motif to previous construct. (c) summary of pull-down assays indicated IQ1, 2, and 4 are occupied by calmodulin in the presence of calcium, while IQ2 and 5 are occupied in the absence of calcium.

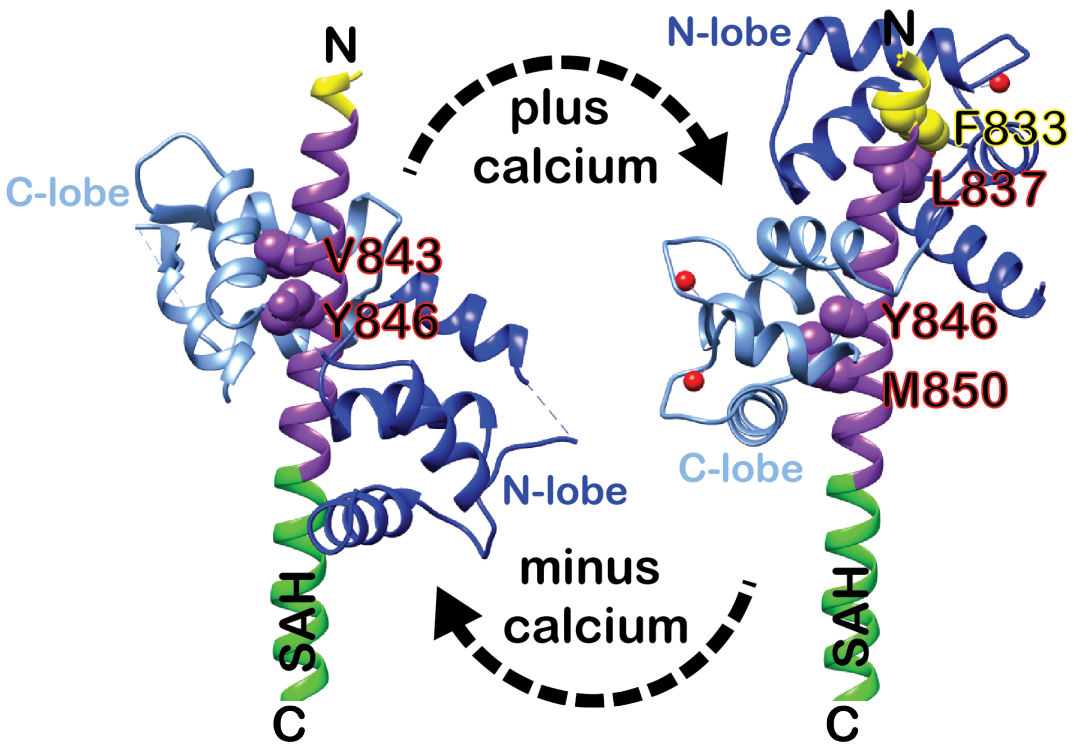


Figure 11 - IQ5 interaction with calmodulin.

Ribbon diagram showing the structure of myosin 7a IQ5 in complex with apo-CaM (left - 5WSU) or Ca^{2+}-CaM (right - 5WSV). The hydrophobic residues critical for binding are highlighted by space fill model. The SAH domain (green), IQ5 (purple), and IQ4 (yellow) are shown. Ca^{2+} ions are shown in red spheres. Recreated from Li et al., (2017).

These results demonstrate an unusual effect of Ca^{2+} on the binding of calmodulin to the IQ motifs, with only IQ2 able to bind calmodulin in both the presence and absence of Ca^{2+}. The lack of bound LC to IQ motif 3, would be expected to introduce compliance into the lever that would not be expected to allow the processive movement demonstrated in artificially dimerised myosin 7a (Yang et al., 2005). Whether the IQ motifs are really unoccupied is the topic of chapter 3 of this thesis and its effect on flexibility on the myosin 7a lever is further investigated in chapter 4.

Immediately following the IQ+CaM lever is a 78 residue SAH domain. Until recently, the SAH domain was only predicted to exist based on the sequence similarity to biochemically characterised structures from myosin 10 (Knight et al., 2005) and myosin 6 (Spink et al., 2008). Despite the lower potential numbers of possible salt bridges that can form between ERK residues when compared with myosin 6 and 10, the crystal structure of the myosin 7a SAH domain shows the presence of salt bridges between the side chains of some charged residues. Interestingly these salt bridges vary between the two structures in same unit cell (Li et al., 2017), hinting that these salt bridges are not fixed, but dynamic, consistent with our understanding of how SAH domains are stabilised by salt bridges (Wolny et al., 2017). The SAH domain adds a further 12 nm to the overall length of the lever, yet is very likely to have a much lower bending rigidity than the IQ+CaM domain, even assuming all the IQ motifs are occupied (Vilfan, 2005; Baboolal et al., 2009).

1.14.3 Tail.

Structural information about the myosin 7a tail does exist, with a 2.8 Å crystal structure (3PVL) of MyTH-FERM1 and the SH3 domain in complex with CEN1 domain of the sans cargo protein being present in the PDB (Wu et al., 2011). Additionally, a crystal structure (5MV7) of the MyTH4-FERM2 domain from myosin 7b in *Homo sapiens* has been obtained. The sequence identity between the MyTH4-FERM2 in myosin 7a and myosin 7b is ~61% (Yu et al., 2017).

In common with the structures of FERM domains from other proteins (Tepass, 2009), the myosin 7 FERM is composed of three lobes, F1, F2, and F3, which together form a cloverleaf configuration. The SH3 domain is coupled to the F3 lobe of MyTH-FERM1 by a short α-helix. This hydrophobic α-helix packs with the βB/βA/βE sheet of the SH3 domain and leaves the canonical target-recognition

pocket of the SH3 domain open. The MyTH4 domains adopt a 10-helix bundle architecture. The central six α helices (α2 and α5 to α9) are highly conserved among various MyTH4 domains (Wu et al., 2011) and assemble into a right-handed superhelical core. The residues forming the short α1, α3, and α4 helices of MyTH4 domains are more divergent. While both MyTH4-FERM1 and 2 show a very similar domain architecture, with an RMSD of 11.52Å between the two structures, the spatial-positioning of the two structures together is unknown.

1.15 Regulation.

The actin-activated Mg.ATPase activity of full-length myosin 7a has a V_{max} of approximately 1 s^{-1}, but K_{ATPase} is high, at approximately 40 μM (Yang et al., 2009). In contrast, while the V_{max} of an S1-like construct (motor domain and first IQ motif) *is* approximately 1 s^{-1}, the K_{ATPase} is only ~1 μM. This means that at low actin concentrations (i.e. 5 μM), the activity of the full-length myosin is very low, whereas that of the tail-less fragment is at V_{max}. This is explained by the fact that full length myosin 7a was mainly in a regulated, inactive state in these assays. Ca^{2+} has also been shown to regulate the activity of full-length myosin 7a (Umeki et al., 2009), possibly by altering the number of bound CaMs to the HC.

Electron microscopy analysis of full-length myosin 7a in the presence of ATP and low (<100 mM salt), showed that the molecule adopts a compact conformation in which the tail contacts the motor domain (Umeki et al., 2009). In the absence of ATP or in increased salt concentrations (>200 mM salt), the molecule adopts an extended conformation. Systematic truncations of C-terminal tail region demonstrated that deletion of the distal portion (F3) of the the final FERM domain (99 amino acids) was enough to activate the molecule, and that this molecule adopted an extended conformation (Yang et al., 2009). Mutation of two conserved positively charged amino acids to alanine in this region (R2140A & K2143A) also prevented the formation of the compact molecule, and reduced the actin-activated K_{ATPase} to ~1 μM (Yang et al., 2009). This suggests that these two acidic residues interact with basic residues within motor domain in the compact inactive molecule. This idea is further supported by the observation that increasing salt concentrations inhibits formation of the inhibited state.

Activation of myosin 7a at high actin concentrations may be due to the presence of the second low-affinity actin-binding site in the last FERM domain. This domain

was expressed and shown to be able to bind weakly to actin with a K_d of approximately 30 μM (Yang et al., 2009). This value is similar to the K_{ATPase} obtained for full length myosin 7a. Binding of myosin 7a to actin through this low-affinity second actin-binding site may causes the myosin to adopt the extended conformation when actin concentrations are high, and become active. This may also be the basis for targeting this molecule to actin bundles. It is also possible that myosin 7a can be regulated by binding to membranes or that proteins that bind to the tail region are able to interfere with the motor-tail interaction, enabling the molecule to adopt the extended conformation.

The region of the myosin which has enough flexibility for the full-length myosin 7a molecule to fold up and adopt a compact conformation is unclear. Potential regions of flexibility include the lever, both the IQ+CaM and SAH domains. This question will be explored further in this thesis.

1.16 Investigating Flexibility.

Central to this work is understanding the flexibility of a protein at different length and timescales. X-ray crystallography only offers a snapshot of a single conformation of (usually) a small part of the protein. Negative-stain electron microscopy allows us to visualise of whole molecules across a continuum of its conformational landscape (with some exceptions, see section 1.17.1). However, this is a relatively low-resolution technique that allows the visualisation of whole molecules and / or domains, as discussed in the next section.

To complement this, a number of modelling techniques now exist which can be used to support these experimental biochemical and imaging techniques (Solernou et al., 2018) These types of computational tools can explore many different lengths and timescales that are otherwise not visible by structural biology approaches and can provide insight into the molecular processes observed in experiments. This is further explored and discussed in section 1.18 and in chapter 5.

1.17 Transmission electron microscopy.

Transmission electron microscopy (TEM) has provided structural information on many biological samples for decades. However, a common problem encountered with TEM is obtaining sufficient contrast from the scattering of the electrons as it

passes through the sample. In TEM, electrons in the beam are scattered as they interact with the atoms in the sample. Contrast is formed by different amounts of electron scatter from different regions of the sample. Electrons scattered through high angles are excluded by a lens aperture. Phase-contrast (interference between scattered an unscattered waves) is the other contrast mechanism which can be enhanced by controlled defocus. To enhance contrast in TEM, heavy metal stains, such as uranyl acetate, can be applied that form a thin layer of crystallites across the grid upon drying, which is extremely electron dense. In areas of optimum stain depth, the biological samples protrude through and exclude stain, creating an envelope of stain around the biological sample, thereby creating regions which are much less electron dense than their surroundings. Contrast is created as the heavy metal atoms of the stain scatter electrons much more strongly than the light atoms of the biological sample, thereby the envelope around the molecule created is imaged with high contrast. The image of the molecule is thus formed from the absence of stain, hence the term 'negative stain'. In the case of uranyl acetate stain, which has one of the smallest grain sizes, the size of the stain crystallites limits the achievable resolution to ~20 Å (Ohi et al., 2004). Additionally, the stain also acts as a support that prevents the molecules from collapsing in the microscope vacuum, although distortions from the adsorption and grid drying processes are still a significant concern (see section 1.17.1).

Although the resolution of negative stain EM is lower than that obtainable in X-ray crystallography, it offers the advantage that crystals do not need to be formed, so different samples can be prepared and analysed relatively quickly. Additionally, whole populations of molecules in a range (sometimes a continuum) of conformations are imaged, in contrast to a single X-ray structure. In this way, EM (both negative stain and cryogenic techniques) can provide a better representation of the full dynamic range of a proteins' conformations.

Negative stain-EM has proved to be an invaluable tool for the analysis of biological specimens and offers particular advantages for the study of biological macromolecules that show discrete flexibility within its domains which would otherwise not be seen with other techniques in such detail (Burgess et al., 2003). It has also been particularly successful in imaging individual myosin molecules and their complexes with actin (Walker et al., 1985; Walker and Trinick, 1989;

Walker et al., 1991; Burgess et al., 1997; Walker et al., 2000; Yang et al., 2009).

1.17.1 Potential artefacts of negative stain EM.

While negative stain-EM is a very quick and easy method for preparing and visualising a population of heterogeneous molecules, it is important to be aware of both the technique's limitations and the processes involved that have the potential to produce artefacts. These processes, described below, can be both chemical and physical in nature.

Chemically induced processes concern the effects of the staining solution on the molecule as well as the changes to the molecules buffering environment. First of all, heavy metal stain solutions such uranyl acetate are acidic in nature. The 1% (w/v) aqueous uranyl acetate solution as prepared for negative staining in this PhD, is unbuffered and has a pH of 4.2 – 4.5. The acidic nature of the staining solutions can be harmful to protein samples that are unstable in acidic conditions, or can cause non-physiological changes in the structure and its appearance. However, some of these issues can be mitigated by the rapid (< 10 ms timescale) fixative properties of uranyl acetate, which has been shown to be able to rapidly stabilise ionic and hydrophobic interactions of myosin 2 filaments with actin before collapse or significant conformational change can occur within the structure (Zhao and Craig, 2003).

A second potential chemical artefact is the deposition of stain within cavities and around the protein. Though uranyl acetate in aqueous solution, in the pH range used in negative staining, exists mostly in un-ionised form (*i.e.* the uranyl ion is mainly associated with the acetate ion), with some charged ions species present (Hayat, 2002). These ions can exist as a series of complexes in both anionic and cationic forms and are thus capable of binding to both negatively and positively charged side chains of a protein. The binding of these different ion species is unpredictable and depends on many factors including stain concentration, pH, duration of staining and other components present in the solution (Hayat, 2002). Charged regions on the protein molecule can thus influence the deposition of stain, and, as a result, areas of a given molecule may be particularly stain-attracting or stain-excluding depending purely on factors described previously.

Another chemical artefact that must be considered concerns the behaviour of the sample as the stain and buffer solution dries on the grid. As the aqueous phase evaporates, the stain and salt concentrations dramatically increase, thereby

drastically altering the ionic strength of the protein sample environment. As described previously, the rapid fixative properties of uranyl acetate may help mitigate some of these issues by rapidly stabilising the structure before collapse can occur, but clearly the potential for this form of chemically-induced artefact should be noted.

In terms of physical mechanisms that can create artefacts in EM, these include the effects of radiation damage from the electron beam during visualisation, surface tension forces acting on molecules during the preparation of the grids, interactions between the substrate and adsorbing molecules and variability in imaging depending on the stain depth.

Radiation damage to a sample as it is exposed to a beam of accelerated electrons is the primary limiting factor in gaining high resolution images of biological macromolecules in TEM. Ionisation or excitation of specimen atoms caused is the main cause of radiation damage. The ions or radicals created lead to the breakage of bonds and other molecular damage, as well as the formation of new cross-links between fragments as they recombine. This is a particularly important consideration in cryo-EM, where the destructive interaction of electrons with frozen-hydrated specimens occurs in three stages: primary damage, as electrons ionize the sample, break bonds, and produce secondary electrons and free radicals; secondary damage, as the secondary electrons and free radicals migrate through the specimen and cause further chemical reactions; and tertiary damage, as hydrogen gas is evolved within the sample, causing gross morphological changes to the specimen (Baker and Rubinstein, 2010). However, the overall low resolution of negative stain EM, which is limited by the grain size of the stain, as well as the supporting nature of the stain may help to prevent gross morphological changes from occurring, perhaps by constraining the protein in its original conformation (Zhao and Craig, 2003).

Another physical mechanism to consider is the promotion of interactions between a protein and the grid substrate. Typically, the carbon coated grids are treated with UV radiation to increase hydrophilicity, and has been shown to be beneficial in stabilising certain proteins (particularly myosins – as individual molecules or filaments), against collapse or damage from the stain itself (Trinick and Elliott, 1979; Trinick, 1981; Knight and Trinick, 1984; Walker et al., 1985). However, interactions between the protein and carbon substrate can introduce artifacts,

which can lead to molecular distortions and certain orientations being preferred have been reported in the literature for elongated and flexible macromolecules, including myosins (Knight and Trinick, 1984; Walker et al., 1985; Burgess et al., 2002; Ohi et al., 2004). These favoured orientations are likely the result of specific interactions, be it charged and/or hydrophilic, on one side of the molecule which favorably interacts with the charged and/or hydrophilic regions of the substrate. In addition to this, it is also important to consider whether the molecule-substrate interaction is affecting the overall shape of the molecule. This is particularly evident in molecules such as titin and vimentin intermediate filaments, which form long string like structures (Tskhovrebova and Trinick, 1997; Mücke et al., 2004). For an interaction energy in the range of the thermal energy, the filaments will freely equilibrate in 2D on the support before they become adsorbed. This gives the overall molecule a more natural conformation across the substrate surface (Tskhovrebova and Trinick, 1997). Alternatively, for an interaction energy bigger than the thermal energy, the molecule may irreversibly bind to the substrate and become “trapped” by before having time to equilibrate. If the molecule binds in several places, the molecule may become trapped in this 2D ‘stressed’ conformation (Mücke et al., 2004). Ideally, the molecules un-equilibrated conformation onto the substrate can be modelled as a projection of the molecules 3D conformation onto the substrate surface (Mücke et al., 2009). The reality is that the 3D structure will re-equilibrate to an extent according to the 2D trapping principle described previously and this depends on several factors including substrate material, any preparatory treatments to increase hydrophilicity, and the type of sample being examined.

Distortion or collapse of macromolecules due to surface tension forces acting during grid preparation is another source for potential artefacts. The potential for macromolecules to appear flattened, elongated, or even collapsed compared to molecular mass expectations and crystal structures is well known in negative stain EM studies (Knight and Trinick, 1984; Radermacher et al., 1994; Cheng et al., 2006; Scarff et al., 2018). Different processes associated with the drying buffer/stain solution can cause this. The forces associated with the retreating meniscus by capillary action can be enough to cause molecules to unfold and collapse. This was found to be the case with C-protein, where the so called ‘side-blotting’ method, where excess sample / stain is removed by capillary action, is

enough to cause the molecule to alter its appearance. Other methods such as the ‘flicking’ method, where excess sample and stain is flicked off the grid, produces starkly different results where the elongated structure is observed (Scarff et al., 2018). Previous estimates of the forces associated with menisci formed around DNA and titin cross sections are relatively high, of the order of hundreds of pN (Bensimon et al., 1995; Tskhovrebova and Trinick, 1997; Tskhovrebova and Trinick, 2001), and so may well account for the kinds of flattening or structural collapse artefacts seen frequently in negative stain studies. Furthermore, during the drying process as the stain evaporates from the substrate surface, the adsorbed molecules break the liquid-air interface surface. Here, surface tension forces act both in parallel (shear) and perpendicularly towards the substrate, with shear forces and compression forces acting to distort the molecule. The magnitude of each force component will depend on the contact angle the liquid makes with the molecule (*i.e.* the ‘wettability’ of the molecule) and the surface tension at the air-liquid interface of the meniscus formed around the molecule’s perimeter.

One final consideration is the variation in the stain depth of negatively stained macromolecules. The depth of the stain can affect the appearance of the molecules on the grid. Ideally, the stain depth should be just deep enough for the molecule to protrude through the stain and accumulate around the perimeter. For example, if the stain is too deep, the background appears much darker and little or no molecule outlining is seen due to the molecule being buried in stain. Alternatively, if the stain is too shallow, little background contrast is seen and the molecules are easily damaged by irradiation in the electron beam (Burgess et al., 2004). A medium depth (usually found in areas exhibiting a gradation in stain depth) is optimal, since the accumulation of stain around the molecule perimeter is maximised, leading to highest contrast imaging. For this reason, it is important to be consistent in collecting images in areas of similar stain depth.

1.18 Modelling.

Biomolecular simulation allows scientists to study both equilibrium conformations and dynamics of biomolecules as a function of their structure. With much of the small length-scale and fast time-scale dynamical behaviour of biomolecules being out of scope for the experimental methods available today, simulation is a

necessary counterpart to experimentation, enabling us to understand how high resolution mechanisms driven by molecular structure lead to the lower resolution experimental observations.

The capabilities of simulation technology, particularly molecular dynamics (MD) (Alder and Wainwright, 1959) have followed directly from the exponential growth of computational power driven by Moore's Law (Moore, 1998). In 1967, Levitt was the first to develop generalised forcefields for atomic simulations (Levitt, 2001), with immediate applications including calculation of thermodynamic properties and vibrational spectra of n-alkane molecules (Lifson and Warshel, 1968) and structural refinement of macromolecular crystal structures using energy minimisation (Levitt and Lifson, 1969). These initial insights and applications paved the way for MD to be generalised to any protein with a known atomic structure. MD requires an initial structure to perform a simulation obtained via X-ray crystallography or NMR experiments with known structures being made available in the Protein Data Bank (PDB) (Bernstein et al., 1978).

Widely available MD packages such as AMBER (Salomon-Ferrer et al., 2013) and GROMACS (Van Der Spoel et al., 2005) use very general functional forms for their potential energies, with contributions from harmonic terms describing bond stretching, bending, and bond rotation, and finally a Lennard-Jones term for Van der Waals interactions and a Coulomb term for electrostatic interactions for each atom type (Figure 12). To distinguish between different types of atom, each term has a parameter specific to the pair of atoms in question, such as an effective bond spring constant (Leach, 2001). These parameters have been grouped to describe specific sets of bio-molecules such as amino acids,

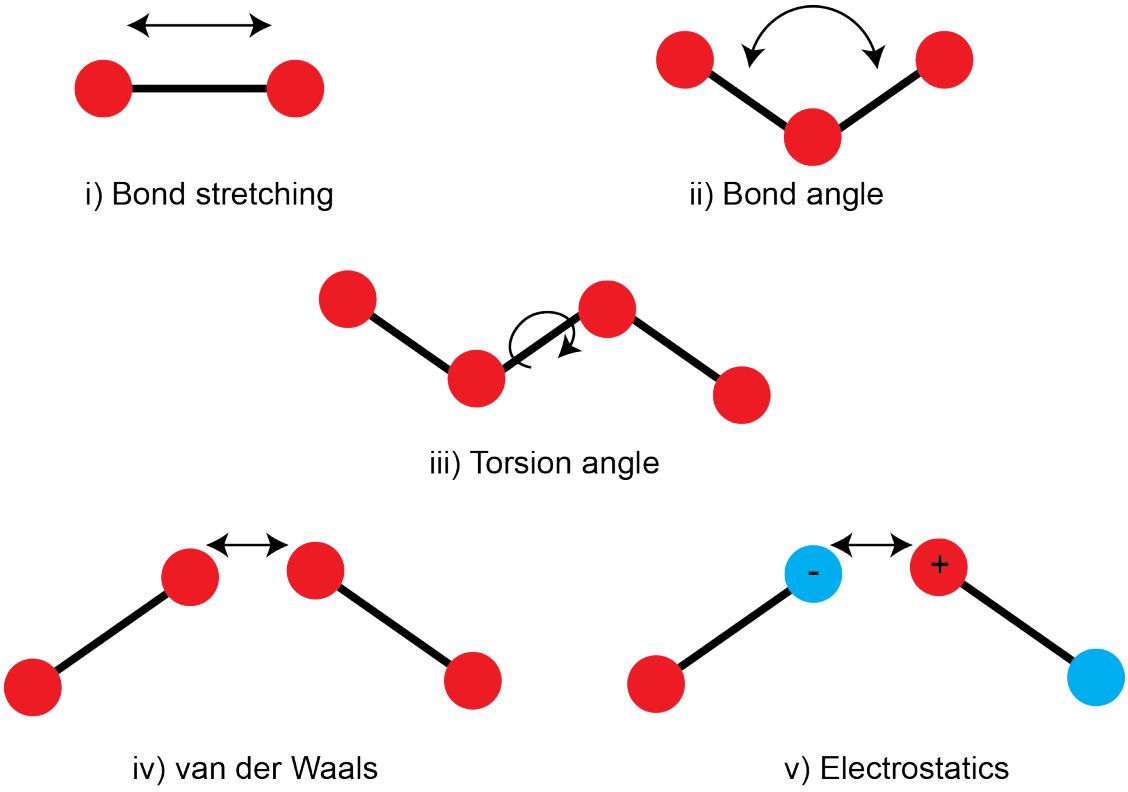


Figure 12 - Force fields describe interactions between atoms.

In molecular dynamics a molecule is described as a series of charged points (atoms) linked by springs (bonds). To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and electrostatic interactions between atoms, one uses a force field. The force field is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

and have been widely researched and validated through previous simulations (Van Der Spoel et al., 2005).

Also important for the dynamics of simulations are the external solvent applied to the atoms, as molecules found in many organisms have evolved to become heavily dependent on their natural environment (Hagen, 2010). A commonly studied process is protein folding, for which it has been shown that hydrodynamic interactions with the external solvent stabilise the protein as it progresses along its folding pathway and eventually help to stabilise the completely folded state, which is a necessity for its function in the organism (Vaiana et al., 2001). This increased stability is because hydrodynamic interactions with the surrounding solvent act both to dissipate energy from a system via inelastic collisions (drag) and also couple together the motions of all atoms in the system. To simulate hydrodynamics within an MD simulation we have the option of explicitly including all of the water molecules or applying an implicit force field.

Finally, the inclusion of a concentration of salt acts to screen the electrostatic interactions between particles by acting as a background negative charge. Although the potential terms within MD are approximations, the technique still shows remarkable accuracy with its predictions. Successful high-throughput docking campaigns using established force fields have been published for proteases, kinases, and bromodomains (Huanga and Caflischa, 2010; Unzue et al., 2016; Zhu and Caflisch, 2016). In a recent application a high-throughput docking program named SEED was used to screen for the CREBBP bromodomain a library of nearly 1500 fragments, which took less than one hour on a commodity computer, and resulted in a 50% success ratio (i.e. of 39 putative binders 20 were confirmed by ligand-observed NMR spectroscopy), and four crystal structures (Spiliotopoulos et al., 2017).

A major limitation on the use of MD is its computational expense, and as consequence software and hardware are constantly being developed to improve the performance of MD simulations. For example, the creation of ANTON (Shaw et al., 2008), a microprocessor developed and optimised solely for MD simulations, enabled simulations of much longer timescales to be performed. The first simulation of longer than a millisecond was published in 2009 and was performed using ANTON (Dror et al., 2012). However, despite these advances atomistic simulations of molecules of the size of full-length myosin 7a (~50000

atoms) for the microsecond timescales required to understand its function are unfeasible. As a consequence, coarse-grained (CG) simulation approaches are required for problems beyond the scale of MD. These techniques do not provide the same level of atomic detail as MD simulations, but enable the lower resolution features of the dynamics to be determined at a much lower computational cost. This is the subject of chapter 5, in which a novel CG simulation approach termed Fluctuating Finite Element Analysis (FFEA), which treats proteins as viscoelastic continua subject to thermal noise is used to study the larger scale motion of myosin 7a.

1.19 Thesis aims and structure.

A structure of the myosin 7a lever is lacking. How it is able to function as a processive motor while potentially having its third IQ motif unoccupied is a mystery. The aim of this research was to gain a better understanding of the myosin 7a lever by investigating the relationship between its structure and to characterise its flexibility. In the first instance, an investigation into light chain binding was performed in cells which endogenously express myosin 7a, which have not been performed in earlier studies (Sakai et al., 2015). It then required an assessment of the stoichiometry between the number of bound light chains and the heavy chain using biochemical and image processing techniques. Native mass spectrometry was used as an alternative technique to assess light chain binding. Finally, an investigation into the bending stiffness of the SAH domain at an all-atom and CG resolution was attempted using different simulation techniques. The bending stiffness results obtained from this work were then applied to full-length models of myosin 7a to investigate how the stiffness of the molecule allowed it to form the regulated state. In summary, the work undertaken sought to address the following research questions which, to date, have not been investigated elsewhere:

1. Does myosin 7a bind any unknown light chains *in vivo?*
2. Determine the stoichiometry of the between the myosin 7a HC and number of bound LC by gel densitometry and by native mass spectrometry.
3. Investigate the lever by negative stain EM to determine its length and flexibility in comparison to a myosin 5a lever construct.
4. Extract a bending stiffness value for the myosin 7a SAH domain from long-time all-atom molecular dynamics simulation.
5. Investigate whether the stiffness parameters obtained from this study are sufficient to allow myosin 7a to form its regulated state.

Following this introductory chapter, a materials and methods chapter is presented, before the three results-based chapters and a final conclusion. The materials and methods chapter contains only generic materials and laboratory methods that are common to the work in more than one of the results chapters. Where methods are specific to work in only one chapter, these are presented at the outset of the particular chapter, for convenience of reference.

2. General Materials and Methods.

General materials and methods chapter which includes information on materials and protocols common to more than one of the results chapters. Methods information for procedures specific to a particular chapter are located in the chapter concerned.

2.1 Chemicals and enzymes.

All chemicals, enzymes and reagents were supplied by Sigma, and restriction enzymes supplied by New England Biolabs (NEB), unless otherwise stated. For commonly used buffers and chemicals, see table 2.

Buffer name	Supplier	Buffer composition
P1	QIAGEN	50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNaseA
P2	QIAGEN	200 mM NaOH, 1% SDS
N3	QIAGEN	4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8
QBT	QIAGEN	750mM NaCl, 50mM MOPS, pH7.0, 15% isopropanol, 0.15% Triton X-100
QC	QIAGEN	1.0M NaCl, 50mM MOPS, pH 7.0, 15% isopropanol
QF	QIAGEN	1.25M NaCl, 50mM Tris-Cl, pH 8.5, 15% isopropanol
TE	QIAGEN	10 mM Tris-HCl pH 7.5, 1 mM EDTA
QG	QIAGEN	5.5 M GuSCN, 20 mM Tris HCl pH 6.6
PE	QIAGEN	10 mM Tris-HCl pH 7.5, 80% ethanol

Table 2 – list of NEB buffers and composition.

Strain	Genotype	Supplier
One Shot® TOP10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Thermo Fisher
NEB 5-alpha Competent *E. coli* (High Efficiency)	fhuA2 a(argF-lacZ)U169 phoA glnV44 a80a(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	NEB
Max Efficiency®DH10BAC competent cells	F-*mcr*A Δ(*mrr*-hsdRMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ-*rps*L *nup*G/pMON14272/ pMON7124	Invitrogen
BL21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int:: (lacI::PlacUV5::T7 gene1) i21 Δnin5*	Novagen
BL21(DE3) Rosetta 2	F- *ompT hsdS*B(rB- mB-) *gal dcm* (DE3) pRARE2 (CamR)	Novagen

Table 3 – Bacterial strains and genotypes.

2.2 Working with Escherichia coli (E. coli).

The following section details the growth and selection of *E. coli* bacteria used in this thesis. Bacterial strains used in this study are listed in table 3.

2.2.1 Growth media.

E. coli were cultured using Luria-Bertani (LB) broth (10 g SELECT Peptone 140, 5 g SELECT Yeast Extract, and 5 g NaCl per Litre (Invitrogen)) or 2YT (16 g/l tryptone, 10 g/l yeast extract, 100 mm NaCl, pH 7.4). Solid media was prepared using LB Broth with agar (Sigma Aldrich). Antibiotics were added to the media according to the concentrations listed in table 4 to select for bacteria containing antibiotic resistant plasmids.

Antibiotic	Working concentration (µg/ml)
Ampicillin	100
Kanamycin	50
Chloramphenicol	25
Gentamicin	7
Tetracycline	10

Table 4 – commonly used antibiotics and their working concentrations.

To obtain single colonies on agar plates, LB agar was prepared by dissolving 1 tablet LB Broth with agar (Sigma Aldrich) in 50 ml distilled water. LB agar was autoclaved and allowed to cool to 50˚C before adding the required antibiotic stock

solution to the final concentrations described above. Plates were poured into sterile petri dishes and allowed to set for 30 min. All sterile work was done on the bench by a lit bunsen flame.

2.2.2 Transformation of competent cells with plasmid DNA.

Chemical transformation of plasmid DNA (for list of plasmids used in this study see table 5) into NEB 5-α competent *E. coli* cells (NEB) or One Shot® TOP10 competent *E. coli* cells (Thermo Fisher) for DNA amplification was performed according to the manufacturer's instructions. Briefly, an aliquot of competent cells (50µl) was thawed on ice for 5 min before transferring 15 - 20 µl to a pre-cooled sterile 1.5 ml eppendorf tube on ice. 1 pg - 100 ng of plasmid DNA was added to the cells, gently mixed, and the cells incubated on ice for a further 30 min. Cells were heat shocked at 42˚C for 30 seconds in a water bath before returning to ice for 2 min. 950 µl Super Optimal broth with Catabolite repression (SOC) medium (NEB) was added to the cells and the cells were incubated at 37˚C with 220 rpm shaking for 1 hr. Between 10 – 50 µl of cells were then spread onto a prewarmed antibiotic selection plate and incubated at 37˚C overnight. Single colonies were picked from the plates for plasmid preparations.

Plasmid name	Source	Resistance	Genes present in vector	Use
pEGFP-C1	Addgene	Kanamycin	eGFP	Construction of eGFP tagged IQ motifs
pEGFP-myo7	Gift from Michelle Peckham	Kanamycin	eGFP, full length Mouse myosin 7a	Amplification of myosin 7a IQ motifs
pEGFP-6IQ	Clontech	Kanamycin	eGFP, myosin 5 6IQ motif	Generation of adeno- and baculo-viral constructs
pET28a_SUMO	Gift from Marcin Wolny		SUMO tag	Generation of bacterial 5IQ expression vector
pDC315	Microbix systems.	Ampicillin		Adenoviral shuttle vector
pBHGloxΔE1,3Cre	Microbix systems.	Ampicillin		Adenoviral rescue vector
pFAST-BAC1	Invitrogen	Ampicillin		Baculoviral shuttle vector
TOPO pCR2.1	Invitrogen	Ampicillin		Amplification of myosin 7a 5IQ

Table 5 – Plasmids used in this study.

2.2.3 Preparing bacterial cultures.

Single colonies from plates were used to inoculate 5ml LB cultures containing the appropriate antibiotic. The cultures were incubated overnight at 37˚C and 220 rpm. For larger cultures, 800 ml LB broth containing the appropriate antibiotic was inoculated with 5 ml starter culture and incubated overnight under the same conditions as described above.

2.2.4 Glycerol stocks.

Glycerol stocks were made for bacterial colonies that carry a validated plasmid of interest. 500 µl of an overnight culture was gently mixed with 700µl 50% sterile glycerol in a cryovial. Glycerol stocks were stored at -80˚C.

2.3 Working with DNA.

The following section details how DNA work was performed. All DNA stocks were stored at -20˚C until required and then thawed at room temperature (~25˚C).

2.3.1 "Miniprep" DNA preparations.

Small scale plasmid DNA was purified from *E. coli* using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. Briefly, 5ml of a 5ml overnight transformed NEB 5α culture was transferred to sterile Eppendorf tubes. The cell suspension was centrifuged at 3000 *xg* for 15 min at room temperature. The supernatant was discarded and the pellet thoroughly resuspended in 250 µl buffer P1 (resuspension buffer), containing RNaseA to remove RNA contamination and methylene blue which indicates alkaline cell lysis in cells by turning a blue colour and turns white upon neutralisation. 250 µl of buffer P2 (lysis buffer) was added and solution mixed by inversion 4-6 times. 350 µl buffer N3 (neutralisation buffer) was added and mixed as before. Lysed cells were centrifuged for 10 min at 17000 x g in a MICRO STAR 17R (VWR) centrifuge to pellet the cell debris. The supernatant was applied to a Qiagen spin column and centrifuged at 17000 x g for 1 min using a MICRO STAR 17R (VWR) centrifuge. The flow through was discarded and the column washed with 500 µl buffer PB. The flow through was again discarded and the column washed with 750 µl of buffer PE. The column was centrifuged at 17000 x g for 1 min using a MICRO STAR 17R (VWR) centrifuge and the flow-through discarded. Residual wash buffer was removed by centrifugation at 17000 x g for 1 min using a MICRO STAR 17R (VWR) centrifuge. The column was transferred to a sterile 1.5 ml Eppendorf and the DNA eluted in 30 µl of TE buffer.

2.3.2 "Maxiprep" DNA preparations.

In order to have sufficient DNA for subsequent cloning steps and for transfection of mammalian cells, larger DNA plasmid preparations were performed using the Qiagen Hi-speed Maxiprep kit, according to the manufacturer's instructions. Briefly, a 400 ml overnight bacterial culture was removed from the shaking

incubator, centrifuged at 3000 x g for 15 min at 5˚C and the pellet resuspended in 10 ml buffer P1. 10 ml buffer P2 (lysis buffer) was added and the tube was mixed by inversion six times and incubated at room temperature for 5 min. 10 ml buffer P3 (neutralisation buffer) was added and the solution mixed by inversion 6 times. The lysate was applied to a QIAfilter Maxi Cartridge, with the cap attached to the tip, and solution allowed to settle at room temperature for 10 min. The cleared cell lysate was then applied to a Hi Speed Maxi Tip, equilibrated with 10 ml buffer QBT. The lysate was allowed to enter the resin by gravity flow, enabling the DNA to bind. The QIAGEN Tip was washed with 60 ml buffer QC before the DNA was eluted with 15 ml buffer QF into a 50 ml falcon tube. The DNA was precipitated by adding 10.5 ml of room temperature isopropanol, gently mixing and incubating at room temperature for 5 min. The DNA/isopropanol mixture was passed through a QIAprecipitator Maxi module to bind the DNA. The DNA was washed with 2 ml 70% ethanol and the precipitator membrane dried by passing air forcefully over the membrane using a syringe. The DNA was eluted into a 1.5 ml centrifuge tube with 1 ml of TE buffer and the eluate was passed through the precipitator a second time to ensure full elution of the DNA.

2.3.3 Precipitation of nucleic acids.

DNA was concentrated by isopropanol precipitation when a higher concentration of DNA was required. To a volume of DNA, half the volume of 7.5 M ammonium acetate (NH_4OAc) and 2 volumes of isopropanol were added and incubated at room temperature for 10 min. The solution was centrifuged at 17000 x g for 10 min using a MICRO STAR 17R (VWR) centrifuge. The pellet was washed with 70% ethanol and the sample was briefly vortexed to break up the DNA pellet. DNA was pelleted by centrifugation at 17000 x g for 10 min. Ethanol was removed by a pipette and the pellet allowed to air dry for 10 min. The pellet was resuspended in a small volume of RNase- and DNase-free MilliQ pure water.

2.3.4 Spectrophotometric quantification of nucleic acids.

DNA concentrations were measured using a NanoPhotometer® NP80 spectrophotometer. Prior to measuring the DNA concentration, the sensor was washed two times with 5 µl distilled water. A blank measurement was then taken with 2 µl TE buffer before 2 µl sample DNA was loaded for measurement. The purity of the sample was determined by the A260/280 ratio, where a pure sample has a value of 1.8 or higher.

2.3.5 Restriction endonuclease digestions of plasmid DNA.

DNA constructs were digested for ligations and to confirm the presence of inserts within a vector using restriction endonucleases, according to the manufacturer's instructions (NEB). For general diagnostic digests, approximately 100 ng DNA was digested in a 20 µl volume containing 2 µl 10X buffer and less than 5 % (v/v) restriction enzyme (~10U/µl). The digestions were incubated at 37˚C for 1 hour to overnight. When larger amounts of DNA were digested the amount of enzyme, the volume of the digestion and the incubation time were increased accordingly. For digestions using multiple restriction enzymes, a buffer was chosen in which both enzymes had a greater than 75% activity, as indicated by the manufacturer. When this was not possible, DNA was digested with the first enzyme before an isopropanol precipitation was performed and the second digestion performed in its optimal buffer.

2.3.6 Separation of DNA fragments by agarose gel electrophoresis.

The digested DNA was separated in terms of its size by agarose gel electrophoresis. Agarose was melted into TAE buffer (0.04 M Tris-acetate, 1mM EDTA pH 8.0) at concentration between 2 and 0.5 % (w/v) depending on the size of the DNA fragments to be separated, and ethidium bromide (EtBr) (Thermo Fisher) was added at approximately 0.5 µg/ml. The gel was then poured into the required size gel casing and combs and allowed to set for 30 min. The gel was placed in a gel tank with TAE buffer and the DNA was mixed with 6X Gel loading dye Purple (NEB) (15% Ficoll® 400, 60 mM EDTA, 20 mM Tris-HCl (pH 8.0), 0.5 % SDS, 0.12 % Dye 1, and 0.006 % Dye 2) to produce a final 1x concentration. The DNA samples, along with 6 µl 1 kb DNA ladder or low MW ladder (NEB) were loaded into separate lanes onto the gel and the gel was run at 100V for ~45 min until the dye front had reached the end of the gel. The gel was imaged on the Molecular Imager® Gel Doc XR+ System (BioRad) and images processed using the Image Lab™ Software (BioRad).

2.3.7 Recovering DNA from agarose gels.

The separation of DNA on agarose gels was used to purify restriction digested fragments. The DNA band of interest was excised from the gel using the QIAquick® gel extraction kit (Qiagen) following the manufacturer's protocol. Briefly, the gel was viewed using a Gel Doc Imager (BioRad, Molecular Imager Gel Doc XR+) and bands carefully cut out on a transilluminator using a clean

scalpel. The band was transferred to a pre-weighed 2 ml Eppendorf tube and the mass of the gel band measured. Three gel volumes of buffer QG was added to the eppendorf, where 100mg of gel was equivalent to 100 µl of buffer. The bands were incubated in buffer QG at 50˚C until the gel slice had completely dissolved. One gel volume of isopropanol was added to the eppendorf tube and mixed to precipitate the DNA. The DNA was applied to a QIAquick spin column and centrifuged at 17000 x g for 1 min using a MICRO STAR 17R (VWR) centrifuge. All subsequent centrifugation steps were also performed at this speed. 500 µl buffer QG was applied to the column to further clean the DNA. This was centrifuged for 60 sec and the flow through was discarded. 750 µl buffer PE and centrifuged for 60 sec and the flow through discarded. The column was centrifuged for an additional 60 sec to remove any residual wash buffer. The QIAquick column was transferred to a sterile 1.5 ml eppendorf and the DNA was eluted with 30 µl TE buffer after 1 min incubation.

2.3.8 DNA Ligation of compatible “sticky” ends.

DNA fragments with compatible ends were ligated together using T4 ligase (NEB) with the supplied 10X T4 ligase buffer. Briefly, insert DNA (~60 pmol), vector DNA (~20 pmol), 10X ligation buffer and T4 DNA ligase were mixed together to a final volume of 20 µl. Molar ratios of between 3:1 and 6:1 between the insert and vector, respectively, were typically used, which could be calculated using the NEB ligation calculator (**http://nebiocalculator.neb.com/#!/ligation**). The reaction was incubated at room temperature for 10 min before the reaction was terminated at 65˚C for 10 mins.

2.3.9 Amplification and modification of DNA by the polymerase chain reaction.

The polymerase chain reaction (PCR) allowed for many manipulations of the DNA. Careful design of the primers allowed for the insertion of tags, novel restriction sites and stop codons which were incorporated into the DNA to be amplified. In addition, PCR was also used to verify if a cloning reaction had been successful. All primers were designed to have a %GC content of approximately 50%. Primers were synthesised by Integrated DNA Technologies (IDT). Lyophilised DNA primers were prepared as 100 µM stocks in TE buffer and stored at -20˚C. PCR reactions were performed using a T100 Thermocycler (BioRad). The PCR protocol was adapted for different PCR primers and products, based

on the length of the required product, the primer annealing temperature and the type of PCR reaction being performed.

2.3.10 Primer list.

Primers were designed for InFusion cloning following the manufacturer's protocol (Thermo Fisher) and using the InFusion primer design tool within SnapGene. Briefly, forward and reverse primers were designed with 15 bp homologous to the vector backbone and the remaining 18 - 20 bp covering the sequence to be amplified. In addition, some primers were also modified to include additional tags, restriction sites, and stop stop codons. Primers were therefore typically 30 - 50 bp long. The Tm was calculated without including the 15 bp overhang and was approximately 60˚C. Other primers are used the verification purposes. The list of primers used in this study are shown in Table 6.

Primer name	Sequence (5' – 3')	Additional features (Restriction sites, tags, STOP codons)	GC content (%)	Tm (˚C)	Use
M7a 5IQ FWD	GTC CGG ACT CAG ATC TGA CAT GTT GCT GGA GGT GGA G	BglII	57	59	Cloning M7a 5IQ
M7a 5IQ REV	GTC GAC TGC AGA ATT C**TT A**GT GAT GGT GAT GGT GAT GGA GGC GCC GCT GGT ACT CAA C	EcoRI, 6His, **STOP**	53	63	Cloning M7a 5IQ
pFAST_M7IQ FWD	CAT CGG GCG CGG ATC ATG GTG AGC AAG GGC GAG		67	61	Cloning M5IQ into pFASTBAC1
pFAST_M7IQ REV	TAC CGC ATG CCT CGA **TTA** GTG ATG GTG ATG GTG ATG GAG G	**STOP**	45	60	Cloning M5IQ into pFASTBAC1
pFAST_M5IQ FWD	CAT CGG GCG CGG ATC ATG GTG AGC AAG GGC GAG		67	60	Cloning M5IQ into pFASTBAC1
pFAST_M5IQ REV	TAC CGC ATG CCT CGA **TCA** CCT CTC CAC GGA CCG	**STOP**	64	59	Cloning M5IQ into pFASTBAC1
pSUMO_5IQ FWD	ACA GAT TGG TGG ATC CGA CAT GTT GCT GGA GGT GGA G	BamHI	54	59	Cloning M7a 5IQ into pET28a_SUMO
pSUMO_5IQ REV	CCG CAA GCT TGT CGA C**TC A**GA GGC GCC GCT GGT ACT CAA	SalI, **STOP**	62	63	Cloning M7a 5IQ into pET28a_SUMO
M13 FWD	CCC AGT CAC GAC GTT GTA AAA CG		52	63	Sequencing primer
M13 REV	AGC GGA TAA CAA TTT CAC ACA GG		43	61	Sequencing primer
FWD Polyhedrin	AAA TGA TAA CCA TCT CGC		39	51	Sequencing primer
REV Polyhedrin	GTC CAA GTT TCC CTG		53	49	Sequencing primer

Table 6 – primers used within this study.
Red characters indicate restriction sites, blue characters indicate a 6His tag, while capitals denote STOP codon

2.3.11 Colony PCR.

Standard PCR reactions for confirming positive colonies (colony PCR) were performed using Taq DNA Polymerase with ThermoPol® Buffer (NEB), following the manufacturer's instructions. A master mix containing all the reaction components for the colony PCR reaction, excluding the template, was prepared to an appropriate volume for the number of reactions to be carried out on ice. The master mix contained 1x ThermoPol® buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer and 0.625 U of Taq DNA polymerase made up to a total volume of 25 µl with nuclease-free MilliQ water. The master mix was then divided between the requisite number of PCR tubes before a single colony was dipped into the reaction before being used to inoculate 10 ml of LB broth with selective antibiotics for overnight culture.

For colony PCR, the bacterial cells were lysed and DNA denatured by heating the PCR reaction mix at 95˚C for 5 min prior to adding the Taq polymerase. The initial denaturation of DNA was at 95˚C for 5 min in order to lyse the bacterial cells and release the DNA. For subsequent steps the denaturation was performed at 95˚C for 30 sec before primer annealing at 3 - 5˚C below the calculated Tm for 15 – 60 sec. The DNA polymerisation step to extend the annealed primers was performed at 68˚C for 1 min/kb. The reaction was repeated for 30 cycles prior to a final 5 min extension at 68˚C. PCR products were resolved by agarose gel electrophoresis.

2.3.12 Phusion® High-Fidelity PCR

For InFusion cloning, Phusion® High-Fidelity DNA Polymerase (NEB) was used to reduce the error rate in PCR products. Phusion PCR reactions were prepared on ice in thin-walled PCR tubes, which contained 1X Phusion high-fidelity buffer, 200 µM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, < 250 ng template DNA, and 1U Phusion DNA polymerase made up to total volume of 25 µl with distilled water. The initial denaturation of DNA was at 98˚C for 30 sec, before denaturation at 98˚C for 5 - 10 sec for subsequent iterations. Following denaturation, the temperature was reduced to 3˚C below the Tm of the primers for 10 - 30 sec to allow the primers to anneal, before extension of the annealed primers was performed at 72˚C for 15 – 30 sec per kb. The reaction was repeated for 25 cycles prior to a final 5 - 10 min extension at 72˚C. Generally, less than 30 cycles were performed as the PCR can introduce random mutations, which

increased in likelihood with the number of cycles. PCR products were then resolved by agarose gel electrophoresis.

2.3.13 TOPO-TA cloning.

After amplification of the fragments by PCR, some fragments were cloned into the pCR2.1 vector using the “TOPO TA cloning kit” from Invitrogen. During PCR, Taq polymerase adds a single deoxyadenosine to the 3’ ends of the PCR products, which degrades away after time. The TOPO kit uses this overhang to ligate fresh PCR products into the pCR2.1 vector. This vector was provided linearized with 3’ deoxythymidine residues. 50 ng gel-purified PCR product was added to the ligation reaction containing 200 mM NaCl, 10 mM $MgCl_2$, 1 µl TOPO vector, and made up to 6 µl. The reaction was incubated at room temperature for 5 min. When ligated, 2 µl of this reaction was transformed into TOP10 competent cells. The pCR2.1 vector allowed blue white selection on Luria Broth plates containing 100 µg/mL ampicillin and 100 µg/mL X-gal.

2.3.14 InFusion cloning.

The cloning of the lever constructs into a range of vectors was performed using the GeneArt® seamless cloning and assembly kit (Thermo Fisher), in which PCR primers are designed specific for the gene of interest but with an additional 15 bp homologous to the vector ends (Figure 13). The amplified insert can then be cloned into a digested vector by homologous recombination between the vector and overlapping portions added to the insert by high fidelity PCR amplification. PCR insert and linearized vector, made up to a total of 5 µl with deionized water were added to 5 µl GeneArt® 2X Enzyme Mix. The reaction was incubated at room temperature for 30 min before 5 µl of the reaction was immediately transformed into TOP10 Competent *E. coli* cells (Invitrogen).

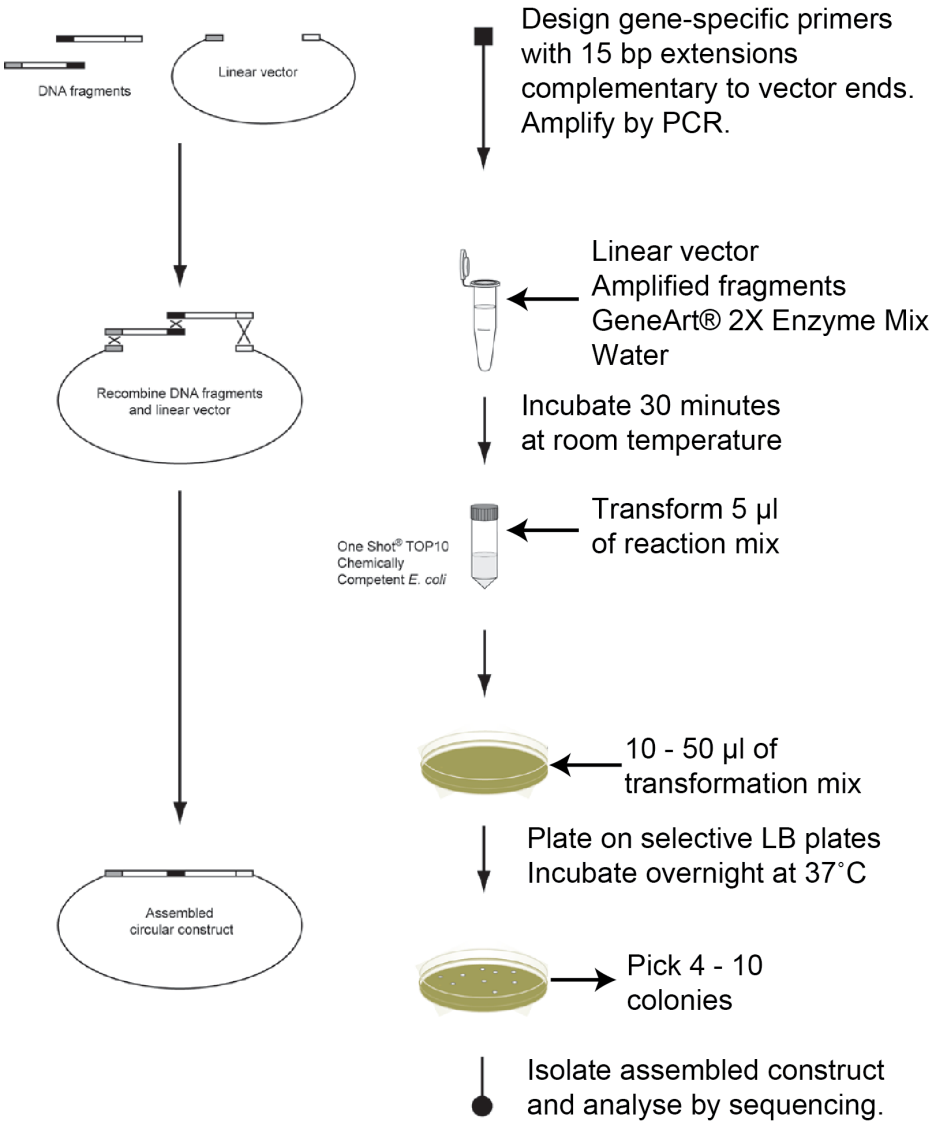


Figure 13 - An overview of the steps for InFusion cloning.

A diagram depicting the steps involved in InFusion cloning. A linearised plasmid and PCR amplified gene of interest are mixed in the InFusion cloning reaction for 30 min at room temperature. Reaction mixtures were transformed into OneShot TOP10 competent *E.coli* and positive clones selected. Adapted from GeneART seamless cloning User Manual (Thermo Fisher).

2.3.15 DNA sequencing.

Cloned constructs were sequenced by Source Bioscience or GATC Biotech Ltd. Available universal primers were used for common sites. Sequences were aligned and checked using SnapGene.

2.4 Mammalian Cell Culture.

All cell culture media and reagents were obtained from Gibco (Thermo Fisher Scientific). All reagents and materials used were sterile and all work was performed in a class II biological safety cabinet. All media was heated to 37˚C prior to use.

2.4.1 Growth of AD-293 cells.

AD-293 cells (ATCC)are derivatives of the HEK-293 cell-line (Graham et al., 1977). The cell line differs from the HEK-293 cell line in that it displays increased cell adhesion properties, which makes these cells less susceptible to disruption when grown as a monolayer in plasticware. The cell line is derived from human embryonic kidney cells transformed with fragments of human adenovirus type 5 DNA to maintain growth in culture (Graham et al., 1977). Additionally, the AD-293 cell line expresses the adenovirus E1 gene in *trans*, thus allowing for the propagation of infectious adenovirus particles when cells are transfected with complement E1-deficient adenovirus vectors, such as the pBHGlox vector (Graham et al., 1977; Parks et al., 1999). The AD-293 cells were cultured in growth medium consisting of Dulbecco's Modified Eagle Medium (DMEM) media (Gibco Thermo Fisher) supplemented with 10 % Fetal Bovine Serum (FBS, Gibco) and 1 % Pen/Strep (10,000 units/mL of penicillin, 10,000 µg/mL of streptomycin, Gibco) at 37˚C, 5% CO_2. Cells were typically passaged 2 - 3 times per week, once cells were ~70% confluent and media changed as required. Cells were maintained to passage 25.

2.4.2 Growth of B16 cells.

The B16 cell line (ATCC) is derived from *Mus musculus* melanoma cells (Fidler, 1975). B16 cells were cultured in growth medium consisting of Roswell Park Memorial Institute (RPMI) 1640 media (Gibco) with 10 % FBS and 1 % Pen/Strep (10,000 units/mL of penicillin, 10,000 µg/mL of streptomycin) at 37˚C, 5% CO_2. Cells were typically passaged 2 times per week, once cells were ~70% confluent and media changed as required. Cells were maintained to passage 25.

2.4.3 Recovery of mammalian cells from liquid nitrogen storage.

Vials of cells recovered from liquid nitrogen storage were rapidly thawed at 37˚C. 1 ml of thawed cells was diluted into 10 ml of pre-warmed media in a T75 tissue culture flask. The flask was incubated in a humidified incubator with 5% CO_2. After 24 hours of recovery, the old media was aspirated and changed to fresh media.

2.4.4 Passaging cells.

Once cells had reached a confluence of ~70% they were harvested and seeded at a lower density on a fresh tissue culture flask. Media was removed from the flask and the cells were gently washed with 5ml medium before being detached from the flask with 2 ml trypsin replacement TrypLE (Thermo Fisher) per 75 cm^3 flask. Cells were incubated with TrypLE for no longer than 5 min, and the trypsin was inactivated by the addition of 8 ml growth medium to give a final volume of 10 ml. AD-293 cells, were not trypsinised, but detached using a cell scraper in a 10 ml volume of medium. Cells were centrifuged at 1000 x g for 5 min at room temperature in a Harrier 18/80 (MSE) centrifuge to pellet before resuspending in 10 ml fresh growth medium. Cells were counted using a haemocytometer (Hawksley BS.748) and the cells seeded onto flasks at the required density (~2 x 10^6) in growth media. Cells were maintained to passage 25, at which point the cells were discarded.

2.4.5 Storage of mammalian cells.

It is important to keep stocks of cells at a low passage number in liquid nitrogen for future use when the passage of current cell stocks become too high. Cells were harvested with TrypLE, pelleted and resuspended at ~1x10^6 cells/ml in growth media supplemented with 10% Dimethyl sulfoxide (DMSO), which acts as a cryo-protectant. 1 ml of cell suspension was transferred to cryovials (Nunc) and frozen in a polystyrene box at -80˚C overnight. For long-term storage cryovials were transferred to a liquid nitrogen tank after 72 Hrs.

2.4.6 Lipid based transfection (FuGENE®) of mammalian cells.

To transiently transfect cells with a DNA construct, the lipid based transfection reagent, FuGENE 6 (Promega) was used. 16 Hours prior to transfection, cells were seeded on a 6 well plate at a density to ensure cells would reach 50% confluency on the day of transfection. FuGENE and DNA were mixed in a four to one ratio. FuGENE6 transfection reagent was warmed to room temperature and

10 µl was mixed with 100 µl 4˚C serum-free DMEM medium in a 1.5 ml eppendorf tube for 5 min. 2.5 µg DNA was added to the reaction, gently mixed and allowed to incubated at room temperature for 20 min. The growth media was removed from the cells to be transfected and 2 ml of fresh media was added. 100 µl of the transfection mixture was added dropwise to the centre of each well and incubated at required growth conditions for 24 – 48 hrs.

2.4.7 Electroporation.

Cells were electroporated using the Amaxa cell line Nucleofection kit R (Lonza). $1x10^6$ cells were resuspended at room temperature in 900µl Nucelofector solution and 2 µg of DNA to be transfected was added to the mixture. The cell/DNA suspension was transferred to a certified cuvette and placed within the Nucleofector cuvette holder (Lonza) and the cells were electroporated. Programs for specific cell lines are available at (https://bioscience.lonza.com/lonza_bs/CH/en/transfection-optimization-strategy-for-the-nucleofector-technology). After electroporation, 500 µl of growth medium was immediately added and the cells were then seeded in a 6 well plate in a total volume of 2 ml. The cells were incubated at the required growth conditions for 24 – 72 hrs.

2.5 Insect cell culture.

The following section details the procedures used to culture *Spodoptera frugiperda* (Sf9) insect cells. All cell culture media was obtained from Lonza and all plasticware was obtained from Corning (Sigma Aldrich). All reagents and materials used were sterile and all work was performed in a class II biological safety cabinet. All media was heated to 27˚C prior to use.

2.5.1 Growth and Passage of Sf9 cells.

Serum free adapted Sf9 cells (Thermo Fisher) were used as they can be grown to a higher density in suspension culture. Cells were cultured in Insect-Xpress medium with L-Glutamine (serum-free medium known as SFX media) (Lonza), in 125 ml Erlenmeyer flasks at 27˚C at 210 rpm in an incubator (New Brunswick Scientific Innova 44) to keep the cells aerated. Every 2 to 3 days, cells were counted and cell viability assessed by staining 50 µL cell-culture samples (diluted 5x in PBS) with 50 µL Trypan blue and examining under a light microscope in a haemocytometer (Hawksley BS.748). Once cells had reached a high density (4 – 6 x 10^6 cells/ml), the cells were split into a fresh 125 ml Erlenmeyer flask (Sigma

Aldrich) at a lower density of $0.5x10^6$ cells/ml in a total volume of 50ml with fresh medium. Cells were passaged to a maximum of 25 times.

2.5.2 Long-term Storage of Sf9 cells.

Low passage stocks of Sf9 cells were kept in liquid nitrogen for future use. Cells were grown to a density of 4 – $8x10^6$ cells/ml, and pelleted at 1000 x g at room temperature for 5 minutes in a centrifuge (Beckman Coulter Avanti J25). Half of the supernatant (conditioned media) was added to a fresh falcon tube and mixed to an equal volume of fresh SFX media and 7.5 % DMSO (v/v). The mixture was then filtered with a 0.22 micron filter (Millipore), to generate freezing solution. The remaining supernatant was removed from the cells and the cell pellet was resuspended to a concentration of $1x10^7$ cells/ml in the freezing solution. 1 ml aliquots of cells were transferred to cryovials (Nunc) and frozen in a polystyrene box at -80˚C overnight. For long-term storage, the cryovials were transferred after 72 Hrs to a liquid nitrogen tank.

2.5.3 Recovery of Sf9 cells from liquid nitrogen storage.

Frozen stocks were recovered by rapidly thawing the vials at 37˚C. The cells were immediately added to 24 ml of SFX media in a Erlenmeyer flask and returned to the incubator at 27˚C with 210 rpm shaking.

2.6 Adenovirus Production.

2.6.1 Transfection of AD-293 cells for viral production.

AD-293 cells were plated out onto a single well of a 6 well plate and incubated overnight. Cells were 50% confluent on the day of transfection. DNA plasmids were transfected into AD-293 cells using the FuGENE protocol (section 2.4.6). The two DNA plasmids required for the co-transfection (pBHGlox and pDC315_eGFP_5IQ_6His / pDC315_eGFP_6IQ constructs) were mixed at a ratio of 5:1. The pBHGlox vector is the E1 deficient packaging vector, containing all the viral proteins (with the exception of the E1 protein). The pDC315 vector is the shuttle vector, which recombines with the Packaging vector to insert the protein of interest into the viral sequence. The transfection mix was added in a dropwise manner directly to the media of a 6 well plate of AD-293 cells and incubated at 37˚C, 5% CO_2 overnight. Cells were inspected for GFP expression using an inverted fluorescent microscope (Nikon).

2.6.2 Adenoviral amplification.

24 hours post-transfection, the cells were divided between 2 wells of a 6 well plate. The cells were incubated at 37˚C, 5% CO_2 and inspected daily for viral production, indicated by the cells expressing GFP and becoming round and detaching from the surface. After 14 days the cells showed at least 40% cytopathic effect and were harvested using a cell scraper. Cells were pelleted at 1000 x g for 5 min at room temperature in a centrifuge (Harrier 18/80 (MSE) before the supernatant was aspirated and the pellet resuspended in 1.0ml Phosphate Buffered Saline (PBS) (Gibco) and transferred to an eppendorf tube.

Virus was released from the cells using a freeze-thaw method. Briefly, ethanol was added on top of dry ice in a small polystyrene box and the cells frozen by placing the eppendorf tube into the ethanol for 5 min. Cells were thawed in a beaker of room temperature water and once defrosted the cells were mixed by vortexing for 1 min. This process was repeated three times. The cells were centrifuged at 5000 x g for 10 min in a centrifuge (MICRO STAR 17R (VWR)) to pellet the cell debris, leaving the adenovirus in the supernatant. A single well of a 6 well plate containing 50 % confluent AD-293 cells were incubated with 500 µl of this viral supernatant together with 500 µl growth medium for 2 Hrs. 1ml of additional growth medium was then added and the cells incubated for 14 - 18 days until viral infection was visible. The virus was harvested as described above.

The virus was amplified four times in AD-293 cells to increase the adenoviral titre:

Amplification	Flasks of AD-293 Cells	Amount of Virus
1	1x6 well plate	500 µl of amplification per well
2	1xT75 flask (Thermo Fisher)	50 µl of amplification per flask
3	2x15cm^2 dish (Sigma Aldrich)	50 µl of amplification per dish
4	5x15cm^2 dish (Sigma Aldrich)	10 µl of amplification per dish

After 50% of the cells had begun to show signs of infection in the fourth amplification, the cells were centrifuged and the pellet frozen at -80˚C along with the media, which contains a significant amount of virus for subsequent purification.

2.6.3 Adenopack adenovirus purification.

Adenovirus purification was performed using the Vivapure Adenopack 100 kit (Sartorius) following the manufacturers' instructions. Briefly, the infected cells and media were defrosted and the cells freeze-thawed, as described above (2.6.2). The cells were centrifuged at 3500 x g for 15 min in a centrifuge (MegaFuge 16R, Thermo Scientific) to pellet the cell debris. The clarified viral supernatant (~ 100 ml) was transferred to two fresh 50 ml Falcon tube and 100 µl of Benzonase® nuclease (1 µl per 1 ml of culture) was added to the viral supernatant, mixed and the mixture was incubated at 37˚C for 30 min. For purification of the virus, a retort stand and clamp was prepared in a tissue culture hood and all subsequent steps were performed under sterile conditions. The tube set provided and a 50 ml syringe were set up in the clamp with the feed tube in the viral supernatant. The supernatant was drawn up into the syringe and the one-way valve used to ensure all air was expelled from the syringe and tubing. A Minisart plus filter unit was attached to the syringe assembly and the supernatant filtered into a fresh container, leaving a small volume of supernatant to prevent air entering the Minisart. Loading buffer at 1/9 volume of the total filtered supernatant was added and the Minisart discarded. The Adenopack, a filter containing an ion exchange membrane which binds viral particles, was prepared taking care to remove all air

to ensure efficient viral binding. A 10 ml syringe was filled with PBS, connected to the Adenopack and 5 – 6 ml of PBS gently passed through. The syringe plunger was pumped up and down a few times to remove air from the Adenopack before the remaining PBS was passed, through leaving 1ml in the syringe to prevent air entering, as this reduces the amount of virus that can bind.

The feed tube from the 50 ml syringe and tube set was placed into the prepared sample solution. The air was removed from the syringe and valve and connected to the Adenopack unit, ensuring no air entered the system. The prepared sample was slowly passed through the Adenopack unit at a rate of approximately 10 ml/min, to allow the viral particles to bind to the membrane. To prevent air entering the system 1-2 ml of sample was left in the syringe. The Adenopack was washed by slowly passing through 100 ml of washing buffer, which selectively removes non-specifically bound proteins from the membrane. As before, 1-2 ml buffer was left in the syringe to prevent air entering the system. To elute the virus, the Adenopack was removed from the tube set and attached to a 10 ml syringe filled with 5 ml elution buffer. The syringe was held over a sterile 15 ml tube and 1 ml of elution buffer was very slowly passed through the Adenopack over 1 - 2 min. The Adenopack was then incubated for 10 min at room temperature, to allow the viral particles to dissociate from the ion-exchange membrane, before slowly passing the remaining elution buffer through at approximately 1 ml/min.

2.6.4 Storage of adenovirus.

The virus was concentrated and the buffer exchanged to storage buffer (20 mM Tris-HCl, 25 mM NaCl, 2.5% glycerol (w/v), pH 8.0) using a Vivaspin 20 (100 kDa MWCO) spin column. 1 ml of the purified virus in storage buffer was dispensed into 10 x 50 µl and 1x 500 µl aliquots. Aliquots were frozen at -80^{o}C until required.

2.6.5 Viral Titre Assay.

The viral titre was calculated using the tissue culture infectious dose 50 ($TCID_{50}$) method (AdEasyVector System, Qbiogen, Inc.). A 96 well plate was prepared with $1x10^4$ AD-293 cells per well, and incubated at 37˚C, 5% CO_2 overnight. The purified virus was serially diluted in medium to give dilutions of 10^{-1} to 10^{-10} and added to each column of the 96 well plate, respectively, with the remaining two wells being left as a control without virus.

Cells were incubated for 10 days and the numbers of wells showing signs of viral infection (cytopathic effect; the presence of GFP cells) were counted. The relative number of wells showing cytopathic effect out of the 10 wells set up was calculated as a ratio for each viral dilution. The sum of these ratios was calculated and denoted as 's'. The $\log^{10}$ of each viral dilution is denoted as 'd'.

To calculate the viral titre:

$T = 10^{1+d(s-0.5)}$

$T = 10^{1+1(9.2-0.5)} = 10^{9.7}$ for 100 µl dilution. Therefore, $TCID_{50}$ is $10^{10.7}$/ml

Finally, the PFU/ml was calculated as:

$T = 1x10^{10.7-0.7} = 1x10^{10.0}$

PFU/ml = $1x10^{10}$ PFU/ml

2.7 Baculovirus Production using bacteria.

2.7.1 Donor plasmid transformation into DH10Bac™ cells.

Recombinant pFastBac1 donor plasmids were transformed into DH10Bac *E. coli* cells (Invitrogen) by heat shock as described previously (see section 2.2.2). These cells allow for the cloned insert to be transposed into the baculovirus genomic DNA. Briefly, 5 µL of plasmid DNA was gently mixed with a thawed aliquot of DH10Bac™ cells, which contain the baculoviral backbone vector and can recombine with the pFastBac1 shuttle vector. This mixture was incubated on ice for 30 mins. The mixture was heat shocked at 42˚C for 45 sec in a water bath, re-chilled on ice for 2 mins, before 900 µL SOC medium was added to the vial. The cells were then incubated in a shaking incubator for 4 hrs (37˚C, 270 rpm). 100 µL volumes of culture were prepared as serial dilutions (neat, 1/10, 1/100) and then spread on pre-prepared LB-agar plates containing antibiotics (50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracycline) plus 100 µg/mL Bluo-gal and 40 µg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG)). Plates were incubated for 48 hr at 37˚C.

2.7.2 Isolation of recombinant bacmid DNA.

DH10Bac colonies were screened by blue-white selection from the plates with white-coloured colonies likely containing the desired recombinant bacmid DNA. Ten white colonies were first streaked onto fresh Luria Broth-agar plates

containing antibiotics (50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracycline) plus 100 µg/mL Bluo-gal and 40 µg/mL isopropyl β-D-1-thiogalactopyranoside (IPTG)) and incubated overnight to confirm their colour. Positively-identified white colonies were picked and added to 5 mL of Luria Broth containing 50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracycline and cultured overnight (37˚C, 270 rpm shaking). Bacmid DNA was isolated from cultures using steps 1-3 of the microcentrifuge protocol from a QIAprep Spin Miniprep Kit (Qiagen) in section 2.3.1, before the DNA was precipitated by adding 800 µL ice-cold isopropanol to the 800 µL of centrifuged cell lysate and incubating on ice for 15 mins. This mixture was centrifuged for 15 min at 17000 × *g* for 15 min in a centrifuge (MICRO STAR 17R (VWR)), the supernatant was discarded and the pellet was washed twice with 70% ethanol. This mixture was centrifuged for 5 min at 17000 × *g* for 15 min in a centrifuge (MICRO STAR 17R (VWR)), the supernatant was then carefully removed with a pipette and the DNA pellet allowed to air-dry for 10 min. Finally, 30 µL TE buffer was added to re-dissolve each pellet and the isolated bacmid DNA samples were stored at 4˚C before use.

2.7.3 Verification of transposition to bacmid.

To confirm the transposition of the recombinant construct into the bacmid DNA, PCR analysis of the bacmid was performed as described in section 2.3.9 and Table 7. Here, M13 primers against regions flanking the gene insertion site in the bacmid were used. If transposition is successful, a PCR product of 2.3 kbp plus the clone insertion size of the recombinant DNA as seen by agarose gel electrophoresis.

Bacmid confirmation reaction:	**Reaction mix:** 1X Taq polymerase buffer, 200 µM dNTPs, 0.5 µM M13 forward primer, 0.5 µM M13 reverse primer, < 250 ng template DNA, and 1U Taq DNA polymerase made up to total of 50 µl with distilled water. **Heating cycle:** 1. 93˚C, 3 min 2. 94˚C, 45 s 3. 55˚C, 45 s 4. 72˚C, 7 min 5. repeat 2.- 4., ×30 times 6. 72 ˚C, 7 min 7. 4˚C, hold indefinitely

Table 7 – PCR reaction composition and conditions to confirm transposition into the bacmid.

2.7.4 Transfection of Sf9 cells with recombinant bacmid DNA.

Successfully-verified recombinant bacmid DNA was transfected into Sf9 insect cells following the Bac-to-Bac protocol (Thermo Fisher). 25 µg of bacmid DNA was suspended in 300 µl PBS and added to 300 µl polyethylenimine (1 mg/ml) (Sigma Aldrich). The mixture was incubated at room temperature for 10 min. The mixture was added to 100 mL of Insect-Xpress with L-glutamine serum free medium (SFX media) (Lonza) containing 0.5 x 10^6 cells/ml and then incubated for 5 days (27˚C, 210 rpm). To harvest the baculovirus, the cell suspension was centrifuged at 3000 x g for 15 min in a Beckman Coulter Avanti J25 centrifuge. The supernatant was then transferred to light-resistant falcon tubes and stored at 4˚C, this is the first baculovirus harvest (also known as P1).

2.7.5 Determining viral titre.

To determine viral titre, 125 µl purified virus, diluted 1/1000 in SFX medium, was added to the first column of a 96 well plate, so that each well contains 100 µl. The remaining 25 µl was diluted 1/5, and 100 µl added to each well for the next columns. This procedure was repeated for the next 8 columns, such that 10 out of the 12 columns contained serial dilutions of the virus. 100 µl of SF9 cells ($7.5x10^4$ cells per well) was then added to all the wells in all 12 columns and left to incubate for 5 days at 27˚C (Figure 14). The final two columns, which do not contain any added virus, act as a control. After 5 days, the cells were inspected for signs of infection, in this case the expression of the GFP-fusion protein can be visualised. The viral titre was then determined using the end-point dilution method (Hopkins and Esposito, 2009) and used to determine the number of plaque forming units per mL (PFU/mL).

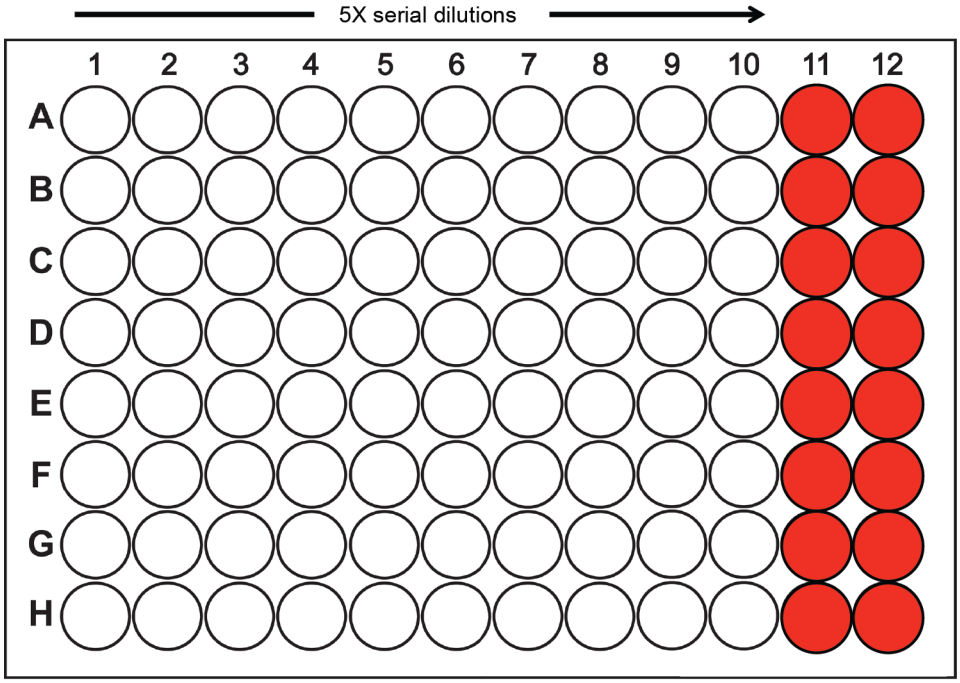


Figure 14 - Setting up baculoviral titer plates.

Lane 1 contains virus diluted at 10-3. The virus when then diluted 5 fold each column. Wells were scored for infection and then entered into an endpoint dilution assay spreadsheet. Red wells are control wells. The plates were incubated in a humified box at 27˚C for 5 days, and checked for signs of infection (expression of GFP).

2.7.6 Virus amplification.

Recombinant baculovirus was amplified from P1 stage through 1 more cycle to P2 stage, according to the Bac-to-Bac protocol (Thermo Fisher). The virus was harvested from infected cell culture supernatant (clarified by centrifugation at 15 min, 3000 × *g* in a Beckman Coulter Avanti J25 centrifuge) and used to re-infect fresh SF9 cells in SFX media. The initial P1 stock was used to infect cells with a MOI of between 0.05 – 0.1. A low multiplicity of infection (MOI); the ratio of viral particles to cells, is used during the amplification process to reduce the amount of defective virus particles formed. To obtain P2 stock, fresh SF9 cells (at 1×10^6 cells/mL) in 100 mL SFX media in a 250 ml Erlenmeyer flask was infected with the chosen virus with a multiplicity of infection (MOI) of between 0.05 - 0.1. The amount of virus to add is determined using the formula below.

$$\text{Inoculum (ml)} = \frac{\text{Desired MOI x cell density (cells/ml)x culture volume (ml)}}{\text{Viral titre (PFU/ml)}}$$

The viral amplification reaction was then incubated for 5 days at 27˚C with 210 rpm shaking. The supernatant was harvested and the titre measured as done previously. A table of viruses used in this study is found in table 8.

Baculovirus	Expression construct	Organism	Titer (10^6 PFU/ml)
M7-IQ	GFP + Myosin 7a 5IQ motif	*Mus musculus*	465
M5-IQ	GFP + Myosin 5a 6IQ motif	*Homo sapiens*	86
Calm2	Calmodulin	*Homo sapiens*	198
BLC	Regulatory + Essential light chains		227
UNC45B + HSP90	UNC45B + HSP90 co-chaperones		100

Table 8 – list of baculovirus used in this study.

2.8 Protein expression, purification, and quantification.

The following section details the methods used for expression, purification and quantification of the proteins used in this study.

2.8.1 Expression using *E. coli.*

To express protein in *E.coli*, the recombinant vectors (table 9) were transformed into either of the two strains of BL21(DE3) competent cells (Novagen) as described in section 2.2.2. BL21(DE3) Rosetta2 competent cells (Novagen) were used if the sequence contained a number of rare codons.

Plasmid name	Backbone vector	Inserted gene	Strain	Resistance
pSUMO_IQ	pET28a_SUMO	Myosin 7a 5IQ (745 – 849)	BL21(DE3) & BL21(DE3) Rosetta2	Ampicillin (and chlorophenicol)
pGEX_nanobody4K	pGEX-6P-1	Anti-GFP nanobody + precision protease site	BL21(DE3) Rosetta2	Ampicillin and chlorophenicol

Table 9 – Plasmids for recombinantly expressed proteins in E. coli used in this study.

Overnight cultures were obtained (see section 2.2.3) and used to inoculate 400 – 1000 ml of LB broth with appropriate selective antibiotics as described in table 4. The cultures were grown at 37˚C with 220 rpm. The cells were grown to OD_{600} = 0.6, with the optical density measured using the Carey 50 Bio UV-Visible spectrophotometer (Varian) set to a wavelength of 600 nm to ensure the cells were in the exponential growth phase for optimal protein expression. Protein expression was induced with the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM. The cells were returned to the incubator for between 4 – 16 Hr at 25 – 37˚C with 220 rpm shaking. Following expression, cells were pelleted at 3000 *x g* for 15 min in a centrifuge (MegaFuge 16R, Thermo Scientific). The supernatant was discarded and the pellet resuspended in 10 ml of appropriate lysis buffer with cOmplete, mini, EDTA-free, protease cocktail inhibitor (1 tablet/50mL) (Roche). At this stage, the resuspended pellet was either frozen at -80˚C or used directly for purification.

2.8.2 Expression using Sf9 cells.

Sf9 cells were grown in suspension to a density of between 1 – 1.5x10^6 cells/ml in 250 – 750 ml volumes in SFX media. The cultures were then infected with

baculovirus at an MOI of between 2 – 5 and returned to the incubator for 72 Hrs at 27˚C with 210 rpm shaking. Following expression, the cells were pelleted at 3000 *x g* for 15 min in a centrifuge (MegaFuge 16R, Thermo Scientific) and the supernatant discarded. At this stage, the pellet was either frozen at -80˚C for future use or resuspended in an appropriate lysis buffer with cOmplete, mini, EDTA-free, protease cocktail inhibitor (1 tablet/50mL) (Roche) and used directly for purification.

2.8.3 Expression using mammalian cells.

Adherent cells were grown to 80 % confluence in 5 x 15 cm^2 dishes (~80.0 x 10^6 cells) prior to infection with adenovirus. The medium was aspirated from the cells and the adenovirus was diluted to the required MOI (0.1 – 10) in 20 ml pre-warmed fresh growth media. The medium containing the virus was applied to the cells and returned to the incubator for 2 Hrs at 37˚C, 5 % CO_2. The medium was then topped up to the final volume with fresh growth medium and returned to the incubator for 24 – 72 Hrs at 37˚C, 5 % CO_2. The cells were harvested using a cell scraper and then pelleted at 1000 x g for 15 min. The media was aspirated and the pellet was either frozen at -80˚C for future use or resuspended in an appropriate lysis buffer with cOmplete, mini, EDTA-free, protease cocktail inhibitor (1 tablet/50mL) (Roche) and used directly for purification.

2.8.4 Cell Lysis prior to Protein Purification.

To prepare bacterial cells for protein purification, the fresh or freshly thawed pellets were resuspended fully in lysis buffer (Table 10) supplemented with cOmplete mini, EDTA-free, protease cocktail inhibitor (1 tablet/50mL) (Roche). The sample was sonicated for 10 seconds on / 10 seconds off for 8 cycles at 50 % amplitude with a sonicator (Soniprep 150 MSE (SANYO)). For mammalian and Sf9 cells, fresh or freshly thawed cells were mechanically homogenised in a chilled ground glass homogeniser on ice. A sample of the homogenate was inspected under the microscope to check for complete cell lysis.

The cellular lysate was clarified at 30,000 x g at 4˚C for 20 minutes in the Beckman Coulter Avanti J25. Samples of both the soluble and insoluble fractions were taken for analytical SDS-PAGE and western blot analysis.

Protein name	Lysis buffer composition
GFP nanobody	PBS (pH 7.4) + 2 mM DTT + 1 mM EDTA
pSUMO_5IQ	150 mM Tris-HCl (pH 7.5) + 150 mM NaCl + 2 mM DTT + 20 mM Imidazole
Adeno-M7-IQ / M5-IQ	150mM NaCl + 50mM Tris-HCl (pH 7.5) + 1mM EGTA + 1% Triton X100
Sf9 expressed-M5-IQ	PBS (pH 7.4) + 1 mM EGTA + 2 mM DTT
Sf9 expressed-M7-IQ	PBS (pH 7.4) + 500 mM KCl + 1 mM EGTA / 0.1 mM $CaCl_2$ + 2 mM DTT + 1 µM calmodulin (*Drosophila melanogaster)*

Table 10 – Lysis buffer composition for purification of the various proteins used in this study.

2.8.5 Protein purification using nickel affinity chromatography, GST or 4K GFP-nanobody.

The following section details several methods of protein purification. All follow a similar protocol and only differ in how the protein is eluted. During the process, all buffers used for purification as well as the collected samples were kept on ice.

Prior to sample loading, a 5 ml column was prepared by adding 1.5 ml resin suspension (50%) of either cOmplete His-tag purification resin (Roche), Glutathione Sepharose 4 Fast flow (GE Healthcare), or prepared 4K-GFP nanobody resin (see chapter 3.2.6 for details on generation of the anti-GFP-nanobody resin) suspension (~2.5 mg/ml) to generate a His-column, GST-column, or 4K nanobody column, respectively. Each column was washed with 10 column volumes (CV) of wash buffer to pre-equilibrate the column before the clarified cellular lysate was applied. Following addition of the lysate to the column, the column was rolled at 4˚C for 1 Hour to allow the protein to bind the resin. Following binding, the column was washed with 5 CV of wash buffer to remove contaminants. In case a His-tag column is used 15 – 50 mM imidazole is added to the buffer to remove non-specifically bound proteins. To elute the protein from their respective columns, the following procedures were followed:

Ni-NTA affinity chromatography for His-tagged proteins: 1 ml fractions of buffer containing 250 – 500mM imidazole was applied to the cOmplete his-tag purification resin (Roche) and incubated for 5 min at room temperature with the stopper in place. The stopper was then removed and the 1 ml elution fraction was

collected. An additional Five 1 ml elution fractions were collected before a final 1mL fraction of 1M imidazole was applied to clear any remaining bound protein.

Glutathione-conjugated agarose (GE Healthcare) for GST purification: To elute the recombinantly expressed GST-tagged protein which contains a precision protease cleavage site, 40µl of preScission protease (2000 units / ml) was dissolved in 1 ml of preScission protease cleavage buffer (50mM Tris (pH 7.5), 150mM NaCl, 1mM EDTA, 1mM DTT), applied to the column and left to incubate overnight at a roller at 4˚C. This overnight fraction was collected. A further four 1ml elution fractions (PBS, 1mM EDTA, 1mM DTT) were collected. Afterwards, the column was washed with 3 column volumes of resin regeneration buffer (50mM Tris, 10mM reduced gluthathione; pH 8.0), with the first wash also collected.

4K-nanobody: To elute the recombinantly expressed GFP-tagged protein from the resin, 1 ml 0.2M Glycine pH 2.7 was applied to the column with the stopper in place and incubated at room temperature for 5 mins. This was to disrupt the GFP: nanobody binding interface to release the GFP construct from the resin. Following incubation, the stopper was removed and a 1 ml elution was collected and immediately neutralised with the addition of 50 µl 1M Tris-HCl pH 8.0. A further four 1 ml elution fractions were collected in this way.

Once the columns had been stripped of any remaining bound protein, the columns were washed with three column volumes of wash buffer and two column volumes of 20% ethanol before being stored at 4˚C in PBS + 3mM NaN_3 for reuse. All fractions were collected for later analysis by SDS-PAGE. 40 µl protein was prepared in 2X Laemelli sample buffer (0.125 M Tris-HCl pH 6.8, 30% (w/v) sucrose, 0.04% (w/v) bromophenol blue, 2% (w/v) SDS, 5% (v/v) ß-mercaptoethanol).

2.8.6 Protein dialysis.

Protein containing fractions were pooled together and concentrated using a Vivaspin 6 (Sartorius) to the required volume according to the manufacturer's instructions. Dialysis was performed using the pooled fractions in dialysis tubing with a molecular weight cut off (MWCO) of 3.5 kDa (Gebaflex Maxi, Generon) placed in a 1000x volume of appropriate buffer. Dialysis was performed in a cold room at 4˚C, with continual stirring using a magnetic stir bar.

2.8.7 Measuring protein concentrations by absorbance at 280nm (A_{280}).

A rapid method of determining the concentration of protein solutions was to measure their absorbance at 280 nm using the NanoPhotometer® NP80 spectrophotometer. 2 µl of buffer was used as a blank before 2 µl of the sample (typically a stock dilution in appropriate buffer) was applied and the absorbance measured at 280 nm. The concentration was calculated based on the extinction coefficients predicted for each construct assuming all pairs of cysteine residues form cysteine bonds by Protparam (https://web.expasy.org/protparam/).

2.8.8 Freezing proteins.

Proteins were frozen drop-wise in liquid nitrogen and the frozen beads were stored in a cryovial in liquid nitrogen until later use.

2.8.9 Separating proteins by SDS-PAGE gel electrophoresis.

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed in order to separate the protein samples according to their size. Glass gel plates (BioRad) with 1mm spacers were cleaned with 70 % ethanol and dried. The gel casting assembly was placed in a casting stand and checked for leaks with water. A separating gel was prepared, using the percentage acrylamide best suited to the size of proteins to be resolved, with smaller proteins needing higher percentage acrylamide. Reagents were added in the order listed (Table 11), with ammonium persulfate (APS) and tetramethylethylenediamine (TEMED) being added simultaneously at the end immediately before the gel was poured. A layer of 70% ethanol was added on top to give the separating gel a level top and the gel was polymerized for 20 min at room temperature. Once the gel had set, a 4% stacking gel was prepared and poured on top of the separating gel (Table 12). The 1mm thick comb (10 or 15 well) was inserted and the gel allowed to set for 20 min. Alternatively, a pre-cast 4-20% polyacrylamide gel (BioRad) was transferred to a gel tank (BioRad).

1X running buffer was prepared from 10x stock (30 g/l Tris base, 144 g/l Glycine, 10% SDS) by diluting 100 ml into 1L of distilled water. Running buffer was added to the centre of the apparatus and the combs were removed. The protein samples were heated to 95˚C for 5 mins before loading onto the gel. 10 µl of PageRuler plus pre-stained protein ladder (Thermo Scientific) was loaded into the first well, before 10 µl of protein sample was loaded in subsequent wells. Remaining

running buffer was added to the tank, and the gel was run at a constant voltage of 100 V for 45 min. The gel was removed, placed in a staining box, ~20 ml of Instant Blue Coomassie stain (Expedeon) was added and the gel placed on a rocker for 30 mins to stain.

	% Acrylamide			
Volume of reagents required for 10 ml	**15**	**12**	**10**	**7.5**
Distilled Water (ml)	2.35	3.35	4.02	4.85
1.5 M Tris-HCl pH 8.8 (ml)	2.50	2.50	2.50	2.50
10% (w/v) SDS stock (µl)	100	100	100	100
Acrylamide/Bis (30% stock) (ml)	5.00	4.00	3.33	2.50
10% (w/v) APS (µl)	100	100	100	100
TEMED (µl)	10	10	10	10
Total (ml)	10	10	10	10

Table 11 – SDS-PAGE Separating gel composition.

Volumes of each reagent needed to produce a 4%, 0.125M Tris, pH6.8 SDS- PAGE stacking gel.	**4 %**
Distilled Water (ml)	6.10
0.5 M Tris-HCl pH 6.8 (ml)	2.50
10% (w/v) SDS stock (µl)	100
Acrylamide/Bis (30% stock) (ml)	1.30
10% (w/v) APS (µl)	100
TEMED (µl)	10
Total (ml)	10

Table 12 – 4 % Stacking gel composition.

2.8.10 Silver staining.

In order to perform silver stain protein gels, protein gels were first fixed with 60 ml acetone stock (50 % acetone (Sigma Aldrich)) + 1.5 ml TCA (50 % Trichloroacetic acid (Sigma Aldrich)) + 25 µl 37 % HCHO (Sigma Aldrich) for 30 mins with shaking. The gel was then washed 7X with distilled water before being incubated with 60 ml acetone stock for 5 mins. The gel was then pre-treated with 100 µl $Na_2S_2O_3$ stock (10 % w/v) in 60 ml distilled water for 1 min before being rinsed 3X with distilled water. The gel was then impregnated with 60 ml staining solution (0.8 ml $AgNO_3$ stock (20 % w/v), 0.6 ml 37 % HCHO, 60 ml distilled water) for 10 min before being rinsed 2X with distilled water. The gel was then developed with developing solution (1.2 g Na_2CO_3, 25 µl 37 % HCHO, 25 µl $Na_2S_2O_3$ stock, 60 ml distilled water) under constant agitation until the bands were sufficiently visible. This takes on the order of 2 – 5 mins. The reaction was stopped with the addition of 1% acetic acid.

2.8.11 Western blotting.

Standard immunoblotting techniques were used to probe for the protein of interest. To transfer proteins from SDS-PAGE gels to a Hybond-ECL nitrocellulose or PVDF membrane (Amersham), the gel was removed from the SDS-PAGE tank and the stacking gel removed using a scalpel. A transfer stack was prepared as shown (Figure 15). All the components were allowed to equilibriate in 1X transfer buffer (Pierce Western Blot Transfer Buffer) and the stack was carefully smoothed to ensure that no air bubbles were present. The cassette was transferred to the Bio-Rad transfer apparatus, the tank was filled with 1X transfer buffer and the transfer was run at 350 mA for 90 mins.

Once the transfer was complete, the stack was removed from the apparatus and the membrane checked to see if the coloured protein bands from the pre-stained ladder had properly transferred. The membrane was blocked in 5% non-fat milk in PBS + 0.05% Tween (AppliChem) for 1 Hr at room temperature on a roller. The primary antibody was diluted to the appropriate working concentration in 5 ml antibody buffer (5% milk in PBS + 0.05% Tween) and membrane was incubated with primary antibody overnight at 4˚C. After incubation, overnight at 4˚C, the membrane was washed 5 times with PBS + 0.05% Tween, for 5 minutes each at room temperature, on a roller. Next, horseradish peroxidase conjugated secondary antibody, diluted to 1:1000 in 5 ml in antibody buffer (5% milk in PBS

+ 0.05% Tween), was added and the membrane incubated with the secondary antibody at room temperature, on the roller for 1 hr. The membrane was then washed five times with PBS + 0.05% Tween as before. The membrane was then transferred to a sheet of acetate, 1 ml of chemiluminescent substrate (Super Signal West Pico) was applied to the membrane and incubated at room temperature for 5 min. The membrane was blotted dry on filter paper, returned replaced on the acetate sheet, covered with a second sheet of acetate and placed in an X-ray cassette. In a dark room the blot was exposed to Lumi-Film (chemiluminescent detection film; Roche) for times varying from 1 sec - 2 mins to obtain the optimal exposure. The film was then developed using an X-Ograph imaging system compact x4.

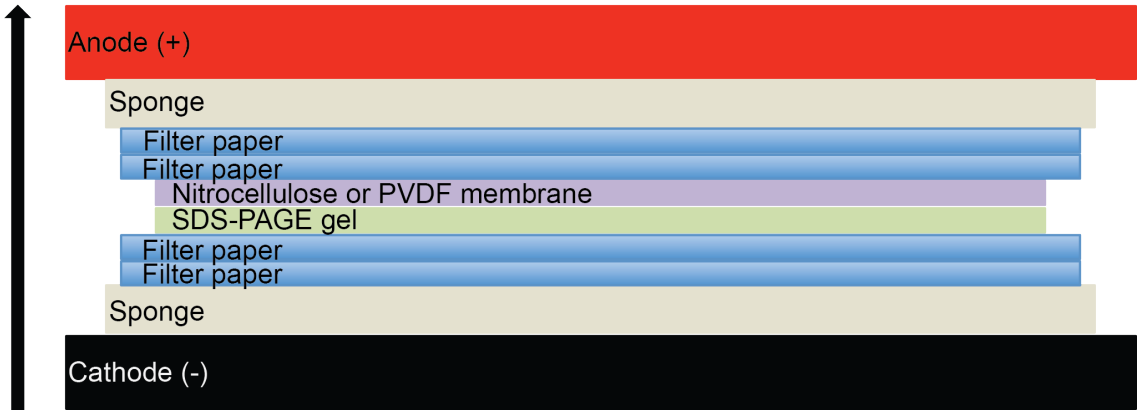


Figure 15 - The layout of a western blot sandwich.

The black band represents the side of the sandwich, which is present closest to the anode. Beige indicates the sponges and blue the filter papers at either side of the sandwich.

3. Investigating which light chains bind myosin 7a

3.1 Introduction.

The lever domain of myosin 7a consists of five predicted IQ motifs spaced equally 23 amino acids apart along the lever helix. However, expressed and purified *Drosophila* myosin 7a was found to bind only 3 mol of calmodulin per mole of heavy chain, and negative stain electron microscopy revealed that the neck length is much shorter than expected for 5 IQ motifs (Yang et al., 2009). This raised the possibility that calmodulin light chains may not bind all five IQ motifs of myosin 7a (Udovichenko et al., 2002).

Previous studies have suggested that unconventional myosins can bind other light chains such as regulatory (RLC) or essential (ELC) light chains (Guzik-Lendrum et al., 2013; Bird et al., 2014). Experiments using IQ motif constructs (Figure 10b) demonstrated that myosin 7a binds calmodulin, RLC, and ELC, with a preference for calmodulin, followed by RLC (Sakai et al., 2015). In addition, it was demonstrated that the presence/absence of calcium affected binding; in the presence of calcium, calmodulin associated with IQ motifs 1, 2, and 4, while in its absence, calmodulin only associated with IQ motifs 2 and 5 (Figure 10c). These data suggest that IQ3 is never occupied by calmodulin or other light chain. However, these experiments used IQ constructs expressed in *E. Coli*, which were tagged using glutathione-S- transferase (GST), which naturally forms a dimer. This may have introduced some steric interference between adjacent light chains which prevented them from binding light chains. In addition, the lengths of each IQ motif used was not consistent (Figure 10a). For example, the IQ4 segment is only 14 residues long, while the IQ1 segment is 36 residues long, and this may also have affected the ability of calmodulin to bind.

A recent crystal structure of the IQ5 together with the SAH domain of myosin 7a in the presence and absence of calcium (Li et al., 2017) revealed that calmodulin binding to the IQ motif did not depend on calcium, in contrast to the earlier study (Sakai et al., 2015). In the absence of calcium, calmodulin interacts with the 5th IQ motif (V843 – R857), where the C-terminal lobe of calmodulin is anchored at V843 and the N-terminal lobe of calmodulin interacts weakly with several residues in the C-terminal half of the IQ motif (Figure 11). Upon binding calcium, the N-terminal lobe of calmodulin shifts (~4nm) towards IQ4 and anchors at two

hydrophobic residues (Phe833 and Leu837), while the C-lobe moves towards the C-terminal end of the IQ5 helix, resulting in a parallel and more extended Calmodulin/IQ5 binding mode (Li et al., 2017). In this state, Ca^{2+}-Calmodulin covers part of the IQ4 sequence. It is possible that the contrasting results for calmodulin binding to IQ4 and 5 in the earlier study (6) could arise from the way in which the N-terminal lobe of calmodulin appears to interact with both IQ4 & 5 in the presence of calcium.

The finding that IQ motif 3 remains unoccupied in both the presence and absence of calcium is surprising, as the myosin lever is thought to be relatively stiff, and thus able to amplify small conformational changes within the motor into a large movement of the lever. An unoccupied IQ motif within the myosin 7a lever would introduce a region of compliance that would be expected to interfere with the mechanical properties of the lever of this myosin.

All of these previous studies used myosin constructs expressed and purified from either Sf9 of bacterial expression systems (Yang et al., 2009; Umeki et al., 2009; Sakai et al., 2015). Using this approach, only the light chains (calmodulin, RLC and/or ELC) co-expressed with the heavy chain are available to bind to the heavy chain. It is possible that there may be an as of yet unidentified light chain that may bind myosin 7a *in vivo,* allowing all 5 IQs to be occupied by a light chain.

To try to address the puzzling finding that not all IQ motifs in myosin 7a appear to be occupied by a light chain, I have explored the binding of light chains to myosin 7a using a novel approach and attempted to determine the stoichiometry of light chain to the heavy chain. First, details of expressing a myosin 7a 5IQ construct in endogenously expressing myosin 7a cells is described. Several expression protocols were considered for expressing the myosin 7a 5IQ construct within cells, described in the following sections. The second part of the chapter focuses on determining the stoichiometry between SF9 expressed Myosin 7a 5IQ constructs and light chains using both native- mass spectroscopy and gel densitometry. A detailed analysis on the myosin 7a 5IQ flexibility was also carried out (see chapter 4).

3.2 Materials & Methods.

3.2.1 Mammalian expression constructs.

An expression construct encoding all 5 myosin 7a IQ motifs (residues 733 - 867 from mouse myosin 7a heavy chain sequence (NCBI GenBank accession: NM_001256081.1)) was generated by PCR (table 6) using a GFP-myo7a construct (Belyantseva et al., 2005) and cloned into the *BglII* and *EcoRI* sites of pEGFP-C1 to yield pEGFP_Myo7-5IQ (Figure 16a).

A myosin 5a 6IQ (residues 755 – 917 from human myosin 5a heavy chain sequence (NCBI GenBank accession: NP_000250.3) containing an N-terminal eGFP was designed and synthesized (GenScript) to yield pEGFP_Myo5-6IQ (Figure 16b).

The open reading frame of both pEGFP_Myo7-5IQ and pEGFP_Myo5-6IQ was cloned into the *NheI* and *SalI* sites of pDC315 (Microbix) to generate the adenoviral plasmids pDC315_eGFP_Myo7-5IQ and pDC315_eGFP_Myo5-6IQ, respectively (Figure 17). These plasmids were then used to generate adenovirus (Adeno-M7-IQ and –M5-IQ, respectively) as described (Section 2.6).

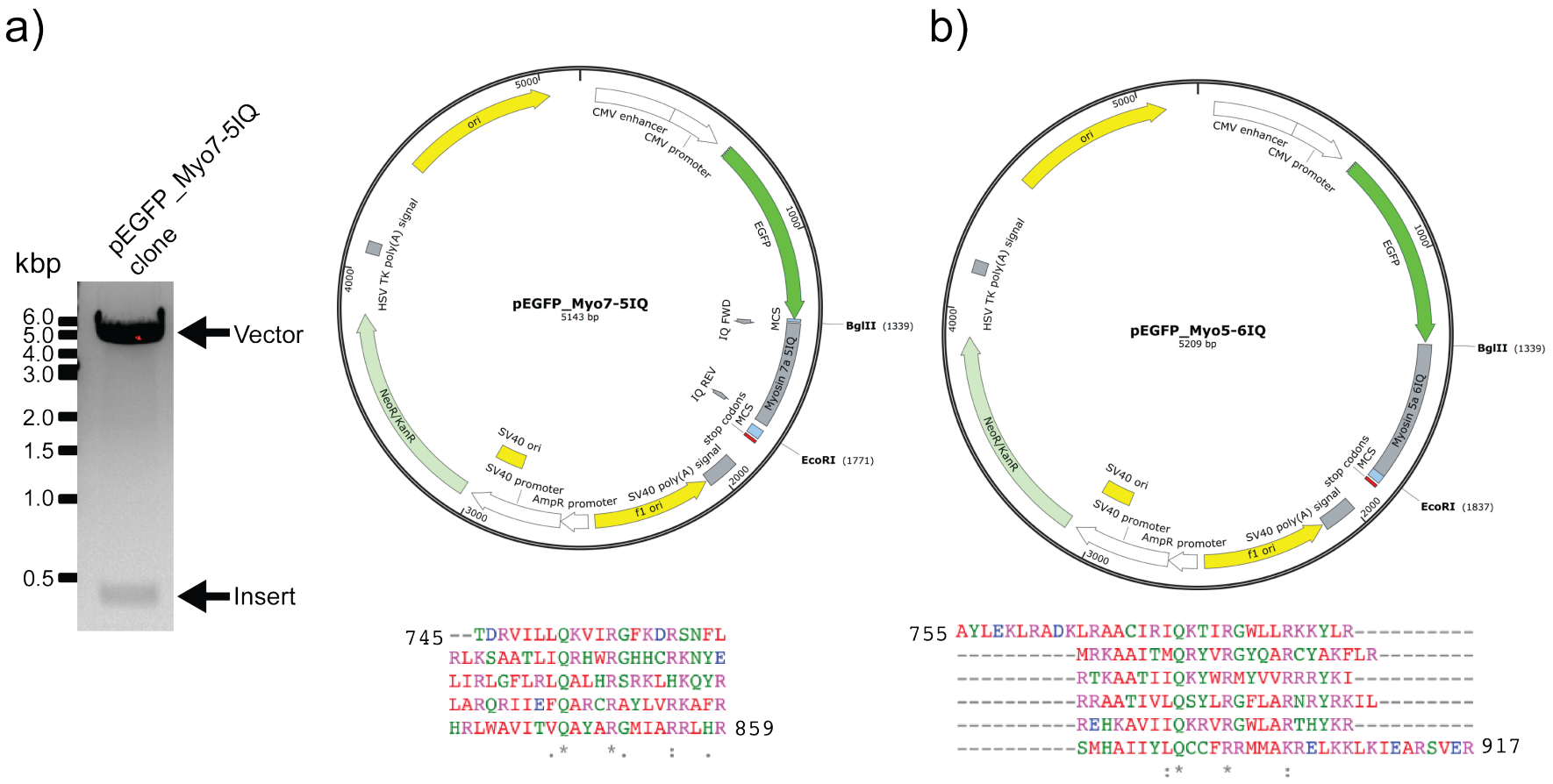


Figure 16 - pEGFP lever constructs.

(a) A 1.0% agarose gel of *BglII & EcoRI* restriction digest of pEGFP-C1 5IQ colonies 1 (Expected size: 5143 bp (vector 4697 bp + insert 446 bp) and complete pEGFP_Myo7-5IQ vector map sequence with IQ motif alignment and (b) Complete pEGFP_6IQ vector map sequence synthesised by GenScript with IQ motif alignment.

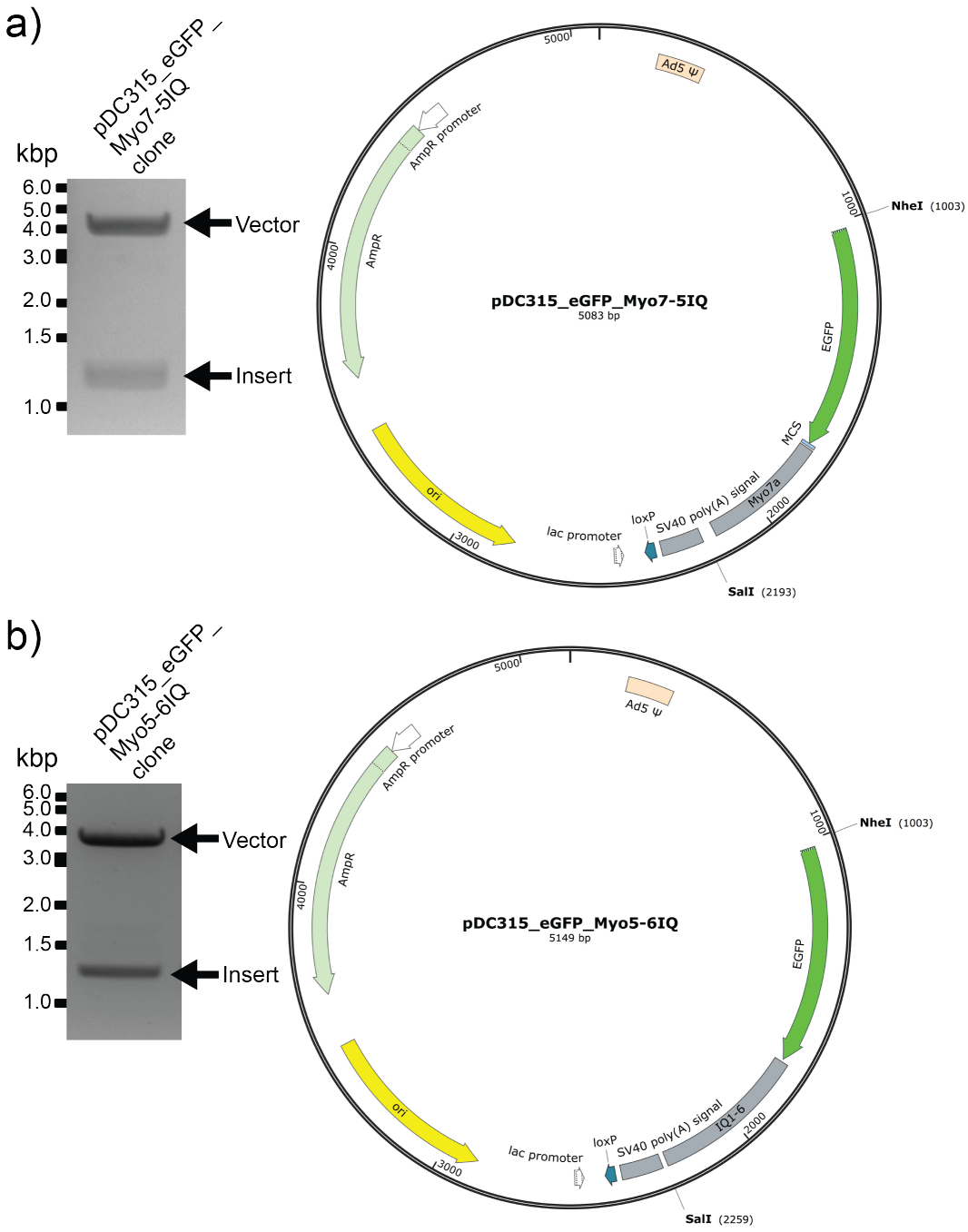


Figure 17 - Adenovirus lever constructs.

(a) NheI & SalI restriction digest of pDC315_eGFP_Myo7-5IQ (Expected size: 5083 bp (vector 3893 bp + insert 1190 bp) and vector map. (b) NheI & SalI restriction digest of pDC315_eGFP_Myo5-6IQ (Expected size: 5149 bp (vector 3893 bp + insert 1256 bp) and vector map.

3.2.2 Bacterial expression constructs.

The myosin 7a 5IQ motif (745 - 859 from mouse myosin 7a heavy chain sequence (NCBI GenBank accession: NM_001256081.1)) was cloned into the pET28a_SUMO vector by infusion cloning (Section 2.3.13 and table 6) into the *BamHI* and *SalI* sites to generate pSUMO_Myo7-5IQ (Figure 18). pSUMO_Myo7-5IQ was transformed into BL21(DE3) and BL21(DE3) Rosetta 2 cells for protein expression (section 2.8.1).

3.2.4 Sf9 expression constructs.

The open reading frames of both pEGFP_Myo7-5IQ and pEGFP_Myo5-6IQ were cloned into the *BamHI* and *XhoI* sites of pFASTBAC^{TM}1 by Infusion cloning (Section 2.3.13 and table 6) to yield pFAST_M7IQ and pFAST_M5IQ, respectively (Figure 19). These constructs were used to generate baculovirus as described (section 2.7). For expression, 250 ml suspension cultures ($1.5x10^6$ cells/ml) were infected with baculovirus at a range of MOIs and the length of time for expression was also varied (Table 13).

Expression Condition (MOI)	Expression duration (Hrs)
M7-IQ (2) + Calm2 (5)	72
M7-IQ (2) + Calm2 (5) + BLC (5)	72
M7-IQ (2) + Calm2 (5) + BLC (5) + HSP90 + UNC45B (2)	72
M5-IQ (2) + Calm2 (5)	72

Table 13 – Baculoviral expression conditions. The BLC virus contains the ELC and RLC ORF.

In addition, co-chaperones (HSP90 and UNC45B) were also added as these have been shown to help some myosins correctly fold (Bird et al., 2014). Light chain viruses were added to the M7-IQ or M5-IQ virus at a 2.5 to 1 MOI ratio, to ensure that sufficient light chain was present to bind the lever constructs.

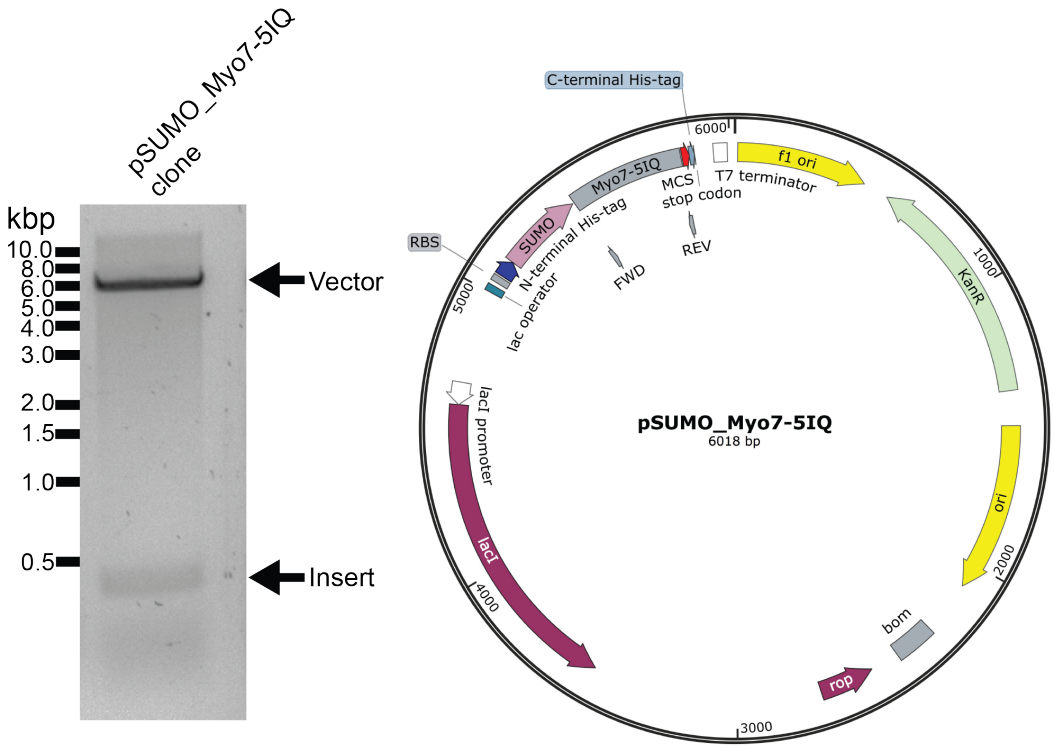


Figure 18 - Bacterial lever construct.

(a) SalI & BamHI restriction digest of pSUMO_Myo7-5IQ (Expected size: 6018 bp (vector 5604 bp + insert 414 bp) and vector map.

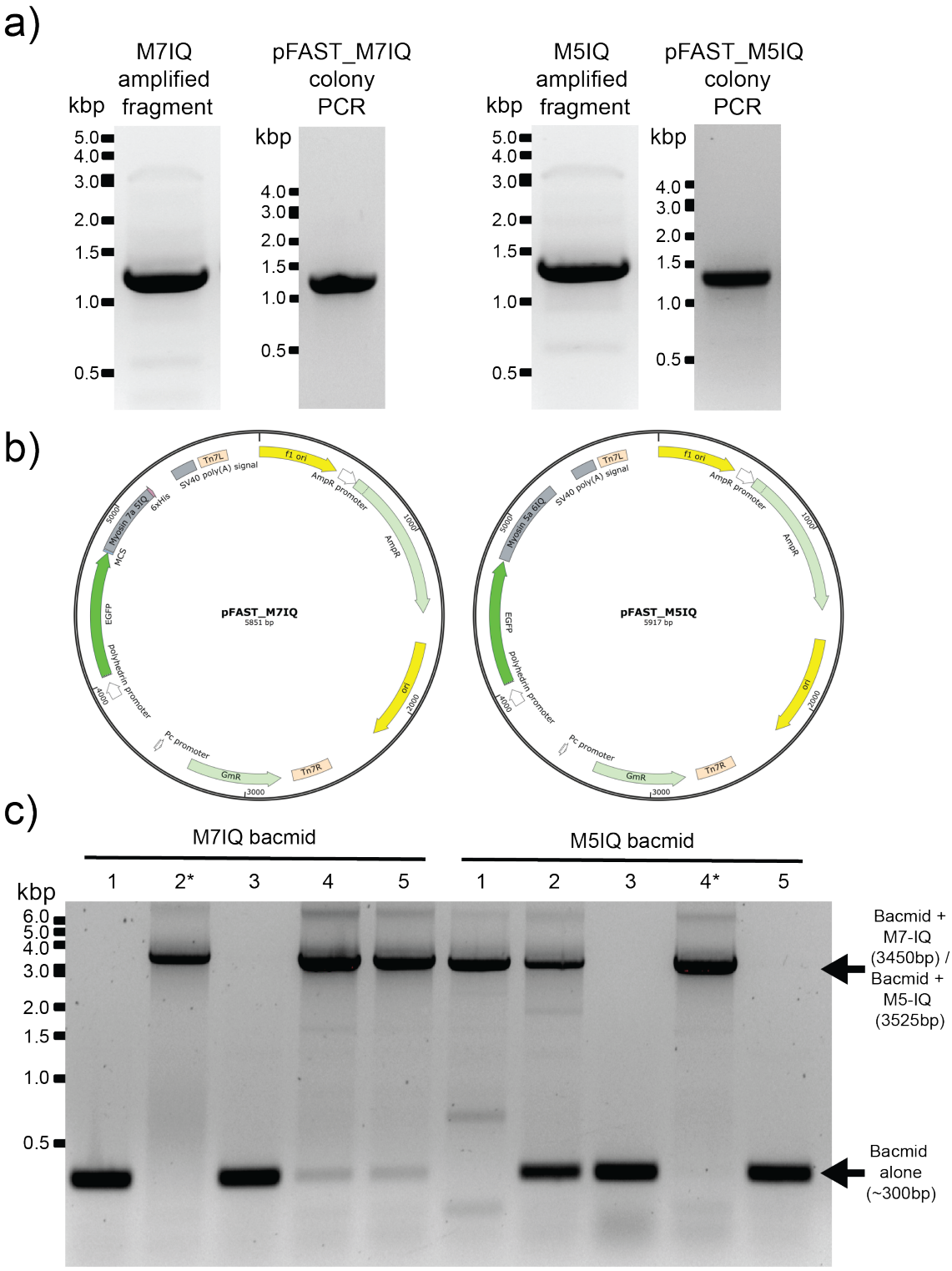


Figure 19 - Generation of lever construct bacmids.

(a) PCR amplification of M7-IQ and M5-IQ ORF (expected size 1188 bp & 1254 bp, respectively) and colony-PCR confirmation of the pFAST_M7IQ construct and pFAST_M5IQ construct. (b) pFAST_M7IQ and pFAST_M5IQ vector maps. (c) Colony-PCR confirmed the presence of the ORF within the bacmid. A successful insertion for the M7-IQ and M5-IQ would yield products of 3450 bp and 2525 bp, respectively. M7-IQ colony 2 and M5-IQ colony 4 (marked *) were used to generate infectious baculovirus.

3.2.5 Imaging of transfected cells.

Cells were imaged by deconvolution microscopy using the Delta Vision Widefield Deconvolution microscope, (Delta Vision, USA). The system uses an Olympus IX-70 epifluorescence microscope equipped with excitation/emission filters for DAPI (ex 360/40 nm, em 457/50 nm), FITC / GFP (ex 490/20 nm, em 528/38 nm), TRITC / rhodamine (ex 555/28 nm, em 617/73 nm) and Cy5 / AlexaFluor647 (ex 640/20 nm, em 685/40 nm). The camera, shutters and software are controlled by SoftWoRX software (Applied Precision Inc.). Cells were imaged to determine transfection efficiency and cell viability, in order to optimise and select the best transfection method.

3.2.6 Generation of anti-GFP-nanobody resin.

The GFP nanobody (a kind gift from Jan Faix, Hannover) was expressed in BL21(DE3) Rosetta2 cells overnight at 25˚C as described (Section 2.8.1). GFP nanobody was purified using its GST tag and cleaved on the column as described (Section 2.8.5). Samples collected from the column were then further purified by anion exchange chromatography with a 5 ml HiTrap SP Sepharose FF column (GE Healthcare), using a salt gradient (from 0.05 to 0.5 M NaCl) to elute the 4K nanobody. The GFP nanobody was concentrated using a Vivaspin 6 (3 kDa MWCO).

To generate the GFP nanobody column used for binding (IP), 2 mg of nanobody was bound to 1 ml activated CNBr Sepharose 4B (Sigma Aldrich) resin. The resin was activated by first hydrating the resin in ice cold 1mM HCl for 30 mins on a roller. The resin was then washed with 5 – 10 column volumes (CV) of distilled water before being activated with 1 CV of coupling buffer (0.1M $NaHCO_3$ + 0.5M NaCl pH 8.3). Following activation, 2 mg of 4K nanobody per ml of resin was incubated with the resin for 2 hours at room temperature on a roller. The resin was then washed with 3 CV of coupling buffer before being incubated with 0.2M glycine pH 2.7 for 2 hours at room temperature on a roller to block unreacted groups. The resin was then washed with 2 CV of coupling buffer before alternating washes between high and low pH buffers (0.1 M Tris-HCl, 0.5M NaCl pH 8.0 and 0.1M acetate buffer, 0.5M NaCl pH 4.0, respectively). The prepared resin was then stored in 1M NaCl + 3mM NaN_3.

3.2.7 Infecting AD-293 and B16 cells with Adeno-M7-IQ and –M5-IQ.

Low passage number AD-293 (ATCC) and B16 cells (ATCC) were used to seed 4 x 15cm^2 petri dishes. On the day of infection, the cells were approximately 75% confluent (a total of ~80.0 X 10^6 cells). Adeno-M7-IQ (1 x 10^8 PFU/ml) and Adeno-M5-IQ (1.25 x 10^8 PFU/ml) was used to infect AD-293 & B16 cells with an MOI of 0.05 and 10, respectively. Cells were incubated for 72 hours at 37˚C.

3.2.8 Anti-GFP-nanobody pulldowns.

Expressing cells were harvested by scraping into medium (80 ml). Loose cells were washed with PBS. Cells were pelleted by centrifugation at 1000 x g for 5 mins. Pelleted cells were resuspended in lysis buffer containing 150mM NaCl, 50mM Tris-HCl (pH 7.5), 1mM EGTA, 1% Triton X100, supplemented with a cOmplete mini protease cocktail inhibitor (1 tablet/50ml - Roche) and placed on ice. Cells were lysed on ice with vortexing every 5 mins for 1 min. Cellular debris was pelleted at 17000 x g for 20 min at 4˚C.

Clarified lysate was applied to 300 µl pre-equilibrated GFP-nanobody resin in a 2 ml eppendorf. Resin was then mixed on a roller for 1 hour at 4˚C. The resin mixture was centrifuged at 1000 x g for 1 min and a 100 µl sample of supernatant was taken for SDS-PAGE analysis before the supernatant was removed by aspiration. The resin was then washed 6 times with wash buffer (150mM NaCl, 50mM Tris-HCl (pH 7.5), 1mM EGTA). 50 µl of Laemelli sample buffer (section 2.8.5) was then added to the resin, and the samples heated to 95˚C for 5 mins before being analysed by SDS-PAGE gel electrophoresis.

3.2.9 Gel densitometry analysis.

Gels were stained using Instant Blue Coomassie stain (Expedeon) or silver stain (section 2.8.10), and then imaged on the Gel Doc Imager (BioRad, Molecular Imager Gel Doc XR+), before performing a densitometry analysis to determine the relative intensities of the band for the IQ motifs and the bands for the light chains, to calculate stoichiometry. The intensity of the bands was measured from the captured images using the gel macro tools in *ImageJ* software (v1.43) (NIH, USA). Images were first converted to 8-bit greyscale, a rectangular box used to select the first lane, and this box copied and pasted to select subsequent lanes. The band intensity profiles for each of the lanes were then plotted and the integrated area under the peaks corresponding to each gel band was measured. The relative intensities of the bands for the IQ motif and light chains was

calculated from these peak areas, and the stoichiometry between the IQ motifs and light chains was calculated as a ratio.

3.2.10 Mass spectroscopy.

Purified proteins were submitted as SDS-PAGE gel bands for protein identification by mass spectroscopy. Briefly, 20 µl of sample was loaded per well and an SDS-PAGE gel was run. Protein was stained with Instant Blue Coomassie stain (Expedeon) and the gel band removed using a clean scalpel. The gel band was stored in 30% ethanol and analysed by tandem mass spectrometry using the University of Leeds Mass Spectrometry Facility (run by James Ault). Native-mass spectroscopy and analysis was additionally performed (Rachel George, University of Leeds Mass Spectrometry Facility). To prepare the samples for this, ~100 µl of sample at 20 µM concentration was exchanged into 200 mM ammomium acetate using a vivaspin ultracentrifugation column.

3.3 Results.

3.3.1 Bacterial expression profiling of myosin 7a IQ motifs.

Expression tests for pSUMO_5IQ showed that a band of the correct size for SUMO-5IQ was only expressed when BL21 (DE3) Rosetta 2 cells were used (Figure 20a). This may be due to rare codons present within the IQ sequence. However, comparing relative amounts of SUMO_5IQ in the supernatant and pellet showed that the majority of the protein expressed was insoluble. This may be because the IQ motifs are unstable in the absence of light chains, and tend to aggregate.

To try to improve solubility, different combinations of temperatures and growth media were tested (Figure 20b). Switching between LB and 2YT medium had no effect, however, expression at 25˚C for 8 hours did improve solubility, although expression levels were reduced. Western blotting, using an anti-6His antibody, confirmed that the SUMO_5IQ protein was present in the soluble fraction. Two weaker bands were also positive for the anti-6His antibody. The ~20 kDa product is likely non-specific as it is also detected in the uninduced culture. The ~16 kDa product is likely to be His-SUMO (14 kDa), a degradation product or an incomplete expression product of the SUMO_5IQ construct.

Large scale expression and purification of SUMO_5IQ at 25˚C was unsuccessful. The major band purified was the 16 kDa fragment, likely to be His-SUMO, and very little intact SUMO_5IQ was purified (Figure 21). A significant amount of the protein was insoluble. Supplementing the lysis buffer with calmodulin when lysing the bacterial cells had no effect on SUMO_5IQ solubility. Due to the difficulty in obtaining soluble protein from bacteria, this expression method was not pursued any further.

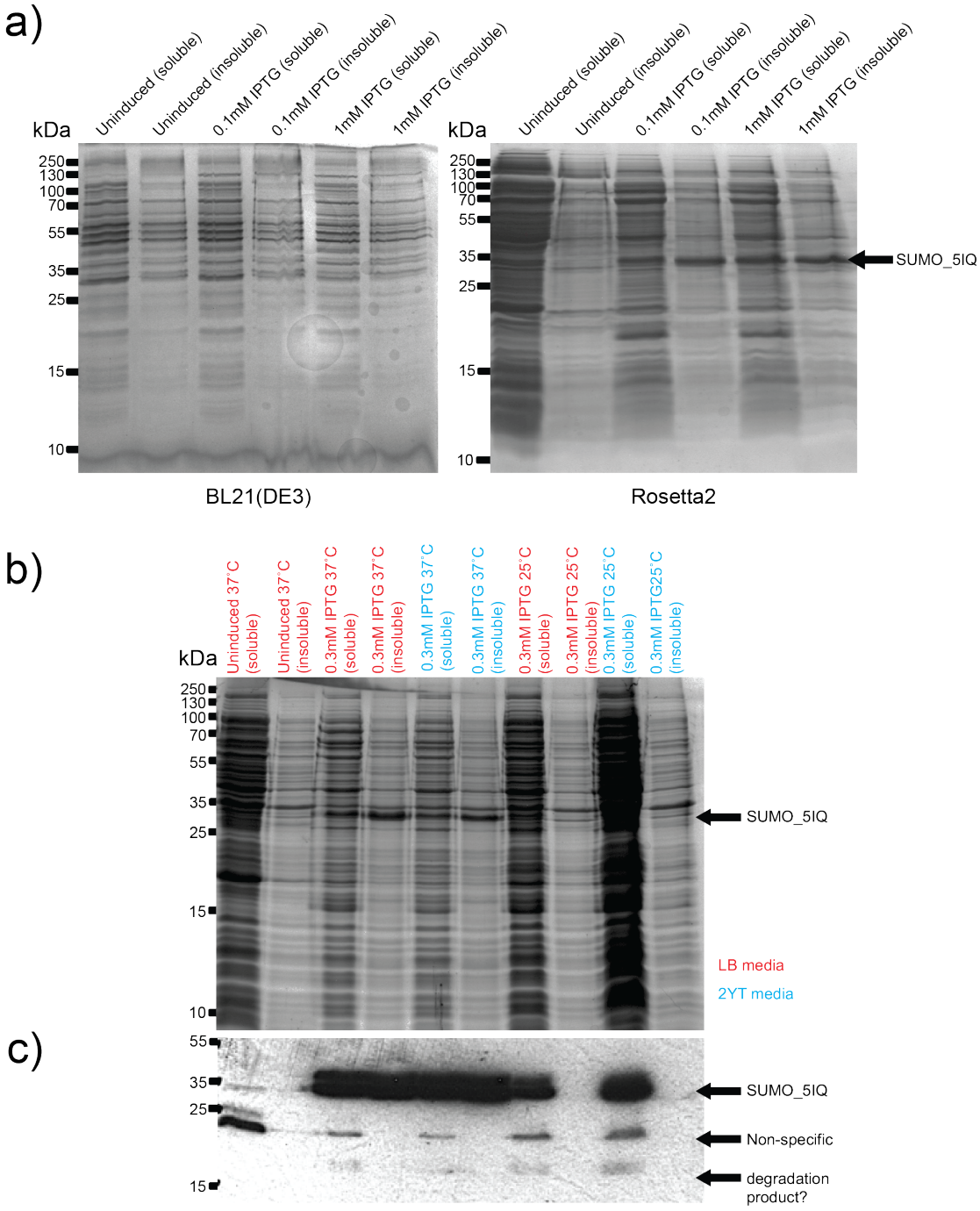


Figure 20 - SUMO_5IQ bacterial lever expression trials.

(a) expression testing in BL21(DE3) and Rosetta2 cells. Soluble and insoluble fractions of uninduced, induction with 0.1mM IPTG and 1mM IPTG. Expression is visualised in Rosetta2 cells induced with 0.1mM and 1mM IPTG. (b) Further expression trials in LB (red) and 2YT (blue) media at 37˚C and 25˚C. SUMO_5IQ appears to be expressed better at lower tempera tures and appears more soluble. (c) Anti-6His western of (b) shows at lower temperatures no pSUMO_5IQ (~33 kDa) is found in the insoluble fraction. Some non-specific epitote is detected (~19 kDa) as well as a degradation product at ~16 kDa.

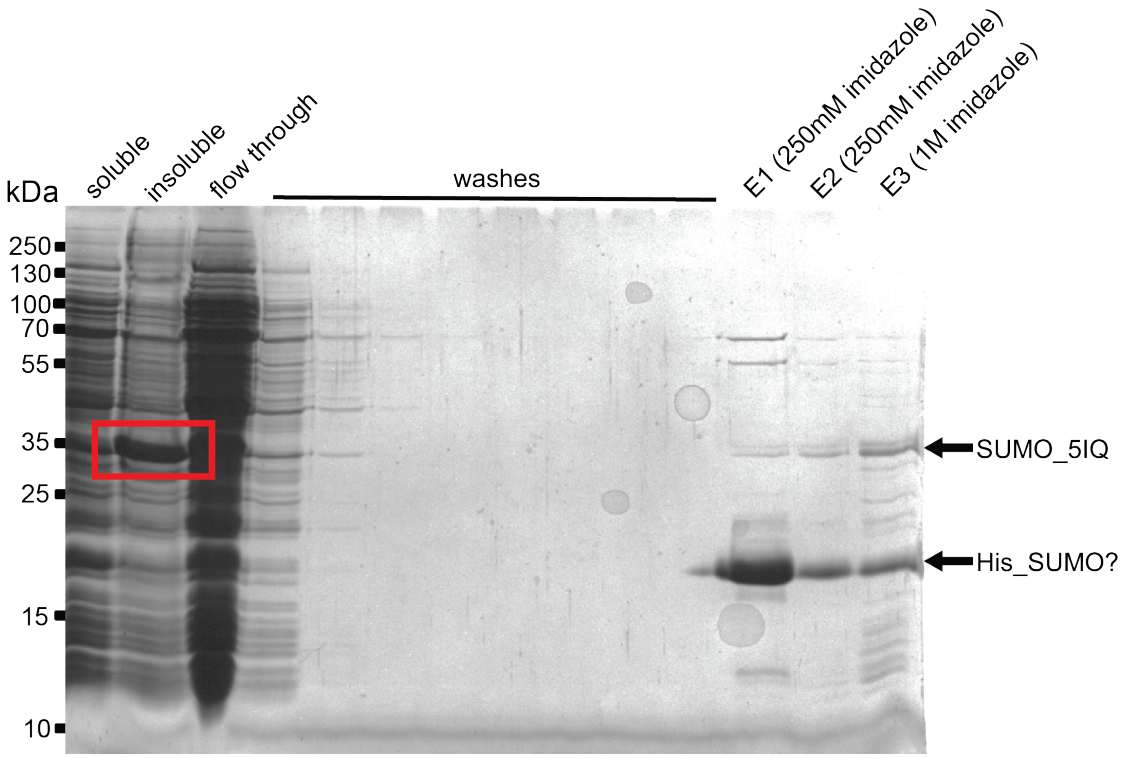


Figure 21 - SUMO_5IQ purification.

A 15% SDS-PAGE gel of SUMO_5IQ purification fractions. The SUMO_5IQ was not easily purified with most of the protein found again in the insoluble fraction (red box). A small amount of SUMO_5IQ was purified but the majority was a ~16 kDa fragment, which is likely a degradation product.

3.3.2 Testing for the presence of unidentified light chains that bind Myosin 7a IQ in mammalian cells.

To test for the presence of unidentified light chains, first 4 different cell lines were tested for expression of endogenous myosin 7a. If there is an unidentified light chain, it is most likely to be present in cells that express this myosin. Four cell lines (B16, AD-293, Cos-7, and HeLa) were tested for the presence of myosin 7a by western blotting using an anti-Myosin7a heavy chain antibody (Millipore, ab3481). Of the 4 cell lines tested, only B16 cells expressed myosin 7a, as shown by the band on the western of the expected size (~250kDa) (Figure 22). These cells were then used for the pull down experiments.

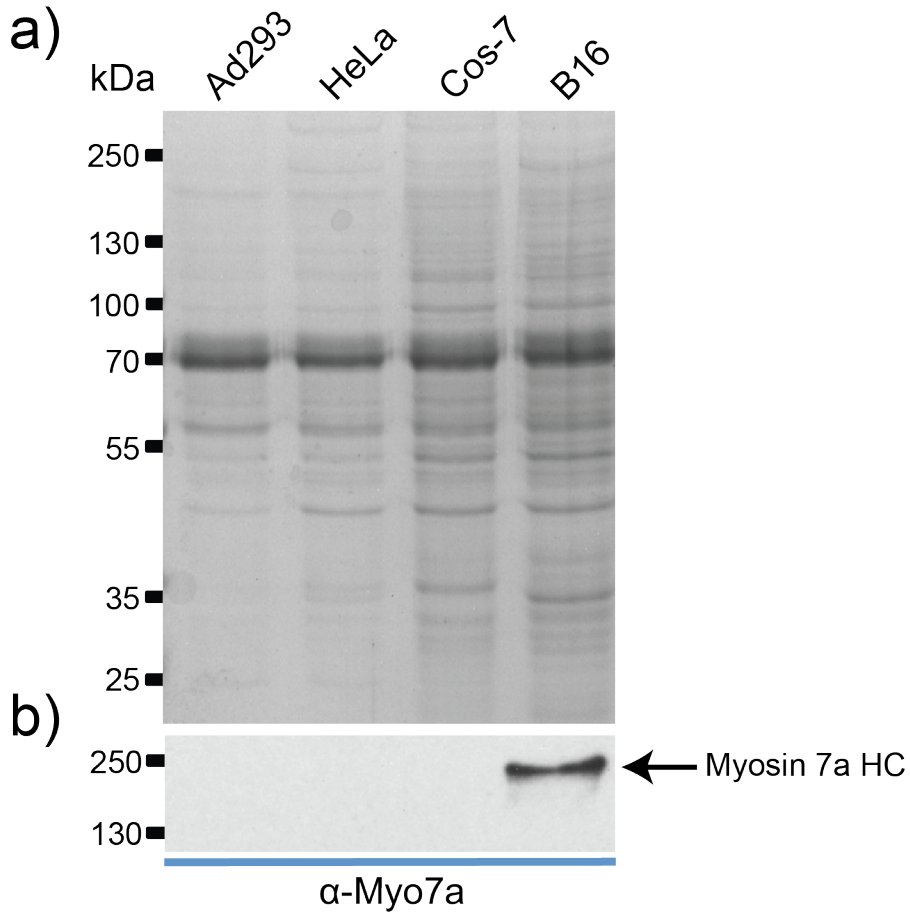


Figure 22 - B16 cells express myosin 7a.

a) 4 - 20% SDS-PAGE coomassie gel of whole cell extracts from Ad293, HeLa, Cos-7 and B16 cells. Molecular weights as shown. b) shows a western blot of the same samples probed with an anti-Myo7a antibody (Millipore). Arrow indicates the myosin 7a is present in the sample.

To express the myosin 7a IQ construct, three methods of transfection, electroporation, FuGENE, and adenoviral infection were tested to determine the best expression levels (Figure 23). The highest expression levels in B16 cells were obtained using adenoviral infection. Imaging the GFP construct shows expression was uniform throughout the cytoplasm the cells viable over a 72 Hr period. Both Fugene and electroporation resulted in low levels of expression, with only a few cells positive for GFP, and these cells tended to be small, and poorly spread (Figure 23a,b). The low levels of expression make these methods unsuitable for pull-down analysis.

The optimal MOI (from 0.1 to 10) and time (24 - 72 Hrs) for the highest levels of expression was determined to be an MOI of 10, and a time of 72 hours (Figure 23c – e). Lower MOIs of 1 and 0.1 resulted in lower levels of expression. Dot blots confirmed that maximum expression is observed at MOI 10 at 72 Hrs post infection (Figure 23f). GFP expression was observed in all MOIs at 72 hours post infection, agreeing well with the fluorescence images (Figure 23c – e). Therefore, adenoviral expression using an MOI of 10 was selected for further experiments.

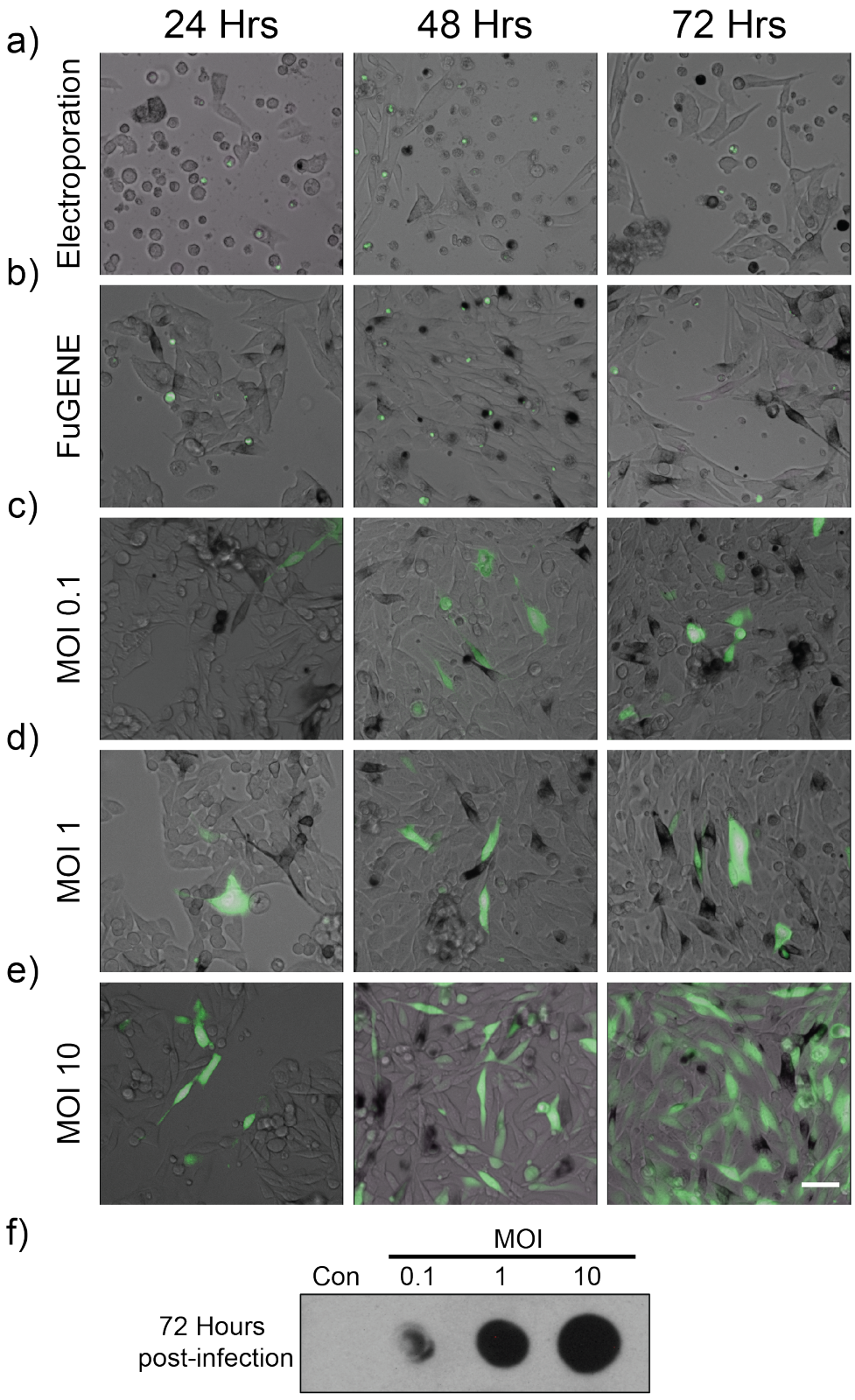
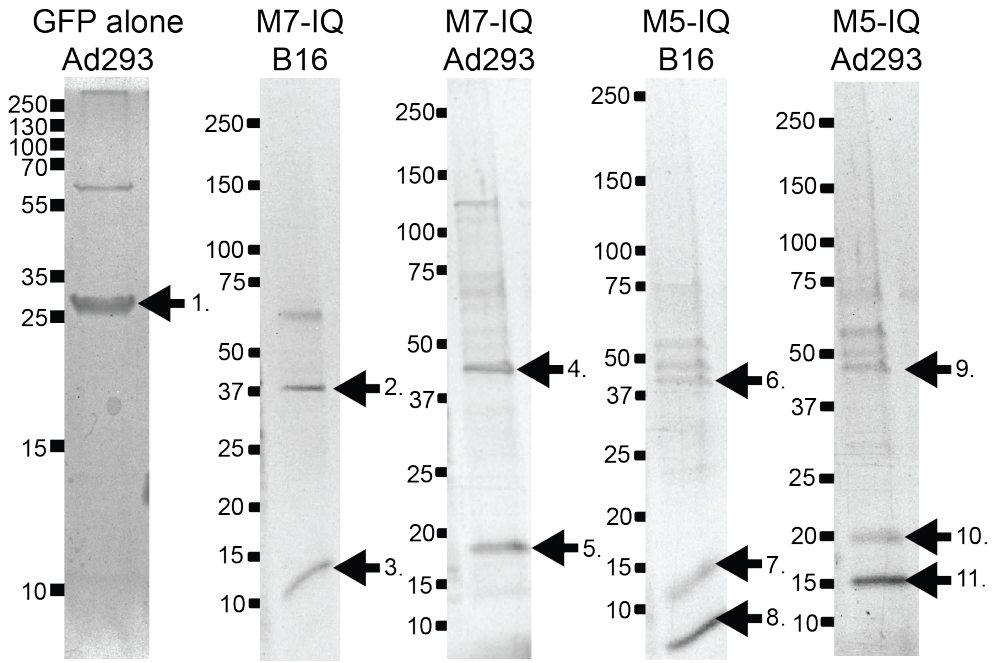


Figure 23 - A comparison of transfection strategies shows adenoviral based infection gave the highest expression.

(a - e) Images show expression of the eGFP-construct (in green) super-imposed on a bright field image (greyscale). An example image over a 72 Hr period for each condition are shown. (f) dot blot using anti-GFP antibody (Millipore) on whole cell lysate 72 Hrs post infection. Scale bar 50 microns.

Pull down experiments using the Adeno-M7-IQ lever construct expressed in B16 cells and AD-293 cells as a control (which do not express myosin 7a) show that the myosin 7a 5IQ motif only binds calmodulin in both B16 and AD-293 cells (Figure 24). No unidentified light chains were found to bind to this construct. In addition, neither regulatory or essential light chains were found to bind. As a further control pull down experiments using an Adeno-M5-IQ lever construct, known to only bind calmodulin, was also able to pull down calmodulin, and no other light chain was found to bind. Using GFP alone did not pull down light chains as expected.

a)



b)

Band ID	Identity	Coverage (%)	Unique peptides
1	GFP		
2	M7-IQ	50	26
3	Calmodulin	75	18
4	M7-IQ	45	23
5	Calmodulin	53	13
6	M5-IQ	43	8
7	Calmodulin	46	5
8	4K GFP nanobody	91	39
9	M5-IQ	52	11
10	Calmodulin	49	7
11	4K GFP nanobody	86	40

Figure 24 - Pulldown experiments in B16 cells show evidence for calmodulin binding only.

(a) shows SDS-PAGE gels of samples for control samples (GFP alone), the M7-IQ construct and the M5IQ construct in both B16 and Ad293 cells. b) shows the results from mass spectrometry for the bands indicated by arrows (band ID as numbered) in (a).

3.3.3 Evaluating light chain binding to M5-IQ and M7-IQ expressed in Sf9 cells.

In further experiments, both the M5-IQ lever construct and the M7-IQ construct were expressed in SF9 cells, and M7-IQ was co-expressed with calmodulin, RLC and ELC (table 13) as described previously to determine which light chains bind the M7-IQ construct. While the M5-IQ construct was easily purified using the 4K GFP nanobody, different expression conditions and purification methods had to be tested for M7-IQ (Figure 25b – d). Purification methods used either the GFP tag or the 6His tag, both of which were present in the M7-IQ construct. Using the GFP-tag for purification, the M7-IQ construct was purified, but in contrast to M5-IQ (Figure 25a), which readily refolded and remained soluble following elution in 0.2 M glycine pH 2.7 and neutralisation in 1M Tris-HCl pH 8.0, the M7-IQ construct was not soluble. The GFP was able to refold as evidenced by the return of green fluorescence following neutralisation. However, the solution remained turbid (Figure 25b), and clarifying the solution by centrifugation at 17000 x *g* resulted in pelleting of some protein. This suggests that the M7-IQ construct does not correctly refold and remain soluble. Additional purification attempts using a less acidic elution buffer (0.2 M Sodium citrate pH 4.0) failed to elute M7-IQ from the 4K resin (Figure 25c).

Purification of M7-IQ using the His-tag (section 2.8.5) was successful (Figure 25d) in purifying the M7-IQ construct with bands of the expected sizes for M7-IQ (~45 kDa) and CaM (~17 kDa) seen. An additional band of ~28kDa was determined to be GFP by mass spectroscopy suggesting GFP could be cleaved off the full length construct. A higher molecular weight band that additionally co-purified (65 kDa) was identified as chitinase by mass spectroscopy, a protein commonly upregulated in Sf9 expression (Thomas et al., 1998).

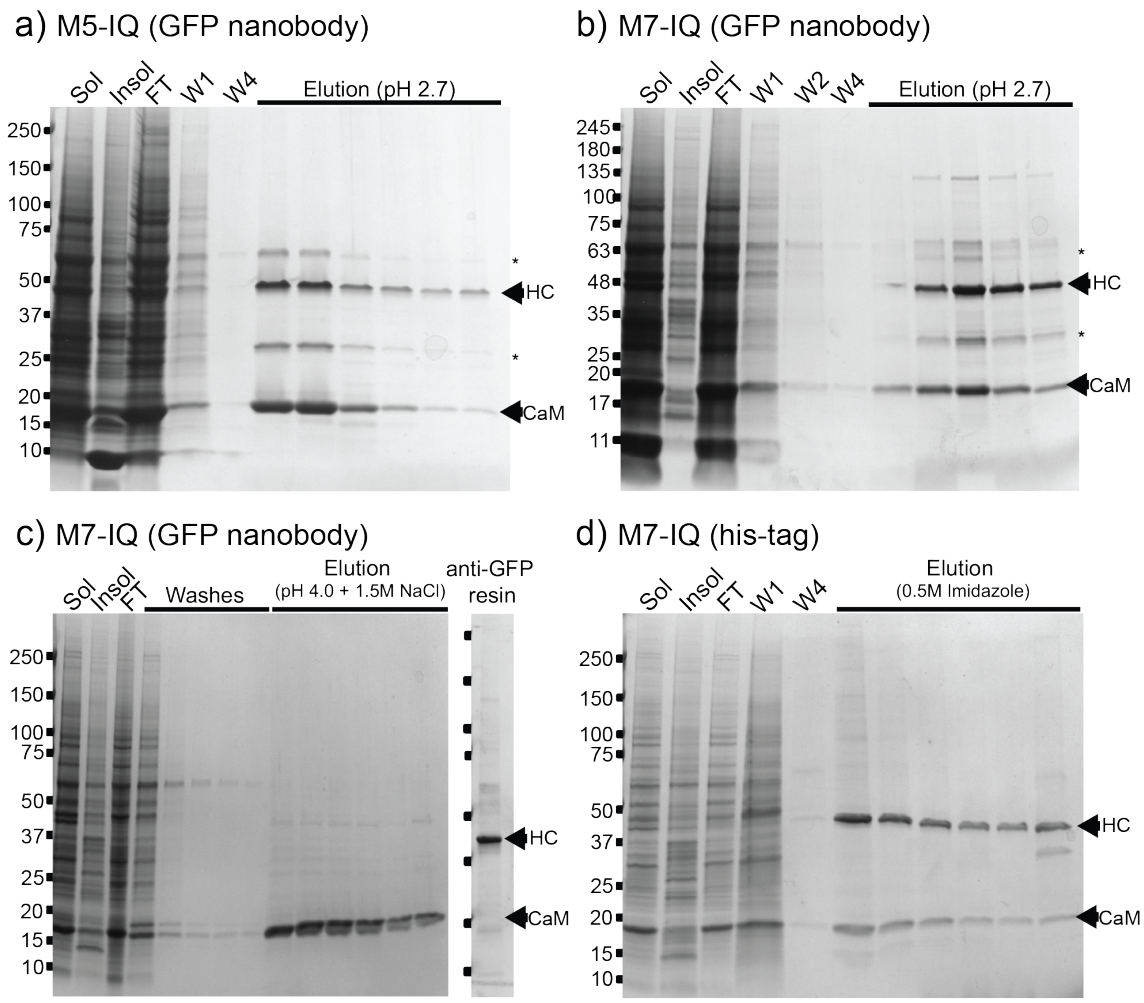


Figure 25 - Lever construct purification in Sf9 cells.

(a - c) SDS-PAGE of lever construct purification by GFP nanobody of (a) M5-IQ + CaM, (b - c) M7-IQ + CaM. Purification of M7-IQ in (b) and (c) was varied from using 0.2M glycine pH 2.7 (b) to using 0.2M sodium citrate + 1.5M NaCl pH 4.0. (d) M7-IQ+CaM purification by his-tag. Arrows show positions of bands corresponding to the lever construct HC (M7-IQ expected 45 kDa, M5-IQ expected 47 kDa) and CaM (expected 17 kDa) are evident. Additional bands at ~28 kDa and ~65 kDa (marked with *) representing contaminants or proteolytically-produced fragments.

Purification of M7-IQ co-expressed with different light chains showed that it can bind calmodulin, RLC, and ELC, when all three light chains are present (Figure 26e). Increasing the salt concentration to 500 mM KCl during purification, reduced the levels of additional co-purifying proteins in these experiments (Figure 26b – e). In addition, the M7-IQ complex only remained soluble if it was kept in 500 mM KCl. The additional band at ~23 kDa observed to co-purify with M5-IQ was identified as the 6IQ motif without GFP.

Recognition of contamination in both M7-IQ and M5-IQ lever preparations is very important, as the additional fragments will contribute to (and thus tend to overestimate) molecule concentration measurements by absorption spectroscopy. Furthermore, preparations containing additional fragments also needed to be acknowledged as they will be detected by native-mass spectroscopy and an explanation for these additional complexes needs to be determined.

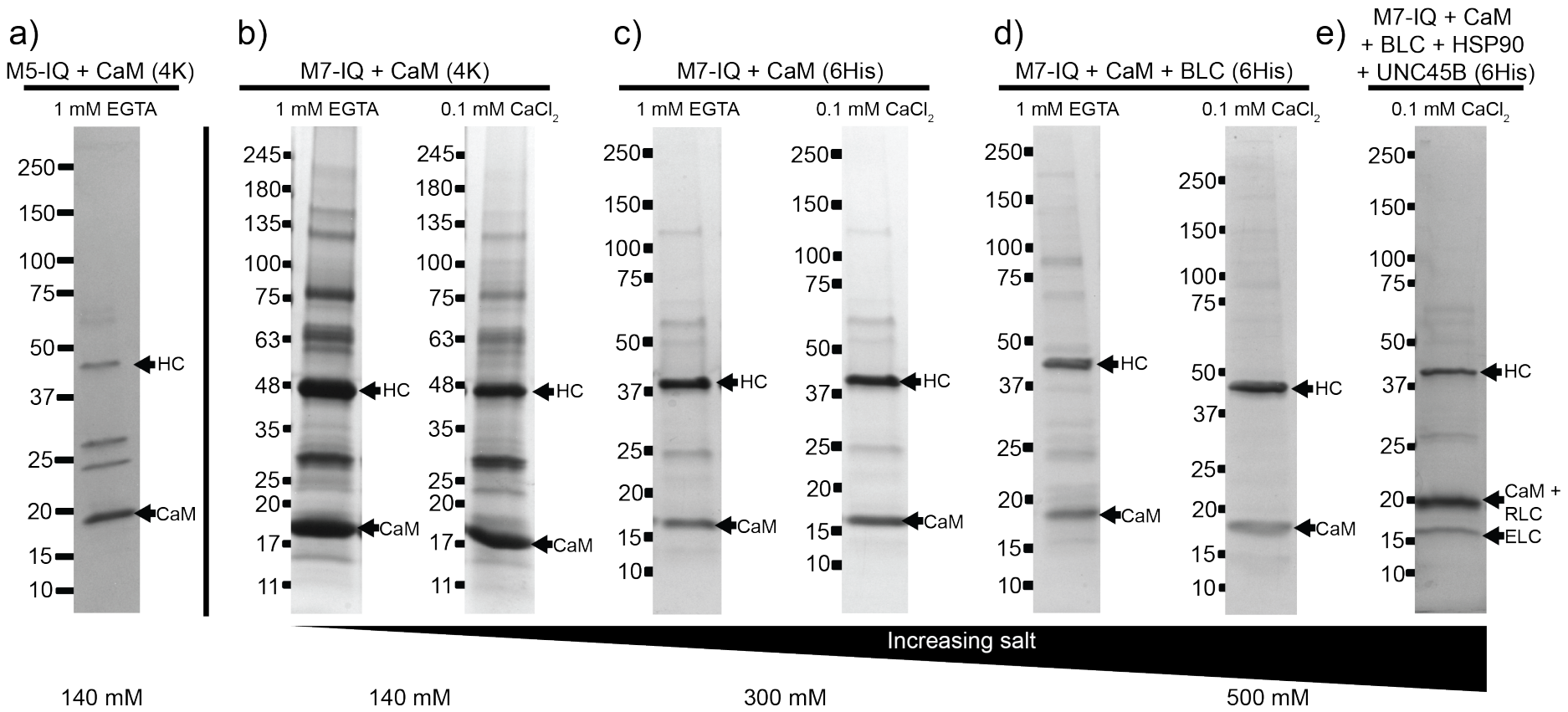


Figure 26 - Myosin 7 lever binds calmodulin, ELC and RLC.

SDS-PAGE of purified samples of M5-IQ and M7-IQ constructs expressed in Sf9 cells in the presence of different light chains. M7-IQ was purified in increasing concentrations of salt in the presence of 1mM EGTA or 0.1mM $CaCl_2$. (a) M5-IQ + CaM purified by 4K GFP nanobody in buffer containing 140mM NaCl and (b) M7-IQ + CaM purified under the same conditions. M7-IQ was purified by his-tag under increasing salt concentration of 300 mM KCl (c) and 500 mM KCl (d). Arrows corresponding to bands of the M5-IQ or M7-IQ lever construct (HC), CaM, RLC (expected size 19 kDa), and ELC (expected size 16 kDa) are shown.

Using gel densitometry analysis of SDS-PAGE (Figure 27), the stoichiometry of HC to LC was 4.1 for M5-IQ, and ranged from 1.0 to 2.9, HC:LC for M7-IQ (Figure 28). The results did depend on the method of staining with coomasie staining showing a more consistent staining between different samples. Results were not due to overloading of the gels, as calibration curves (Figure 27c) showed a linear relationship between protein concentration and band intensity below 5 – 10 μM calmodulin, and these experiments used concentrations of lever constructs of 1 μM. This densitometry approach does not appear to be highly reliable, as the ratio obtained for M5-IQ was 4.1, but we know that M5-IQ does bind 6 calmodulins. This could be explained by differences in staining for the heavy chain and the small calmodulin protein (Schleicher and Watterson, 1983).

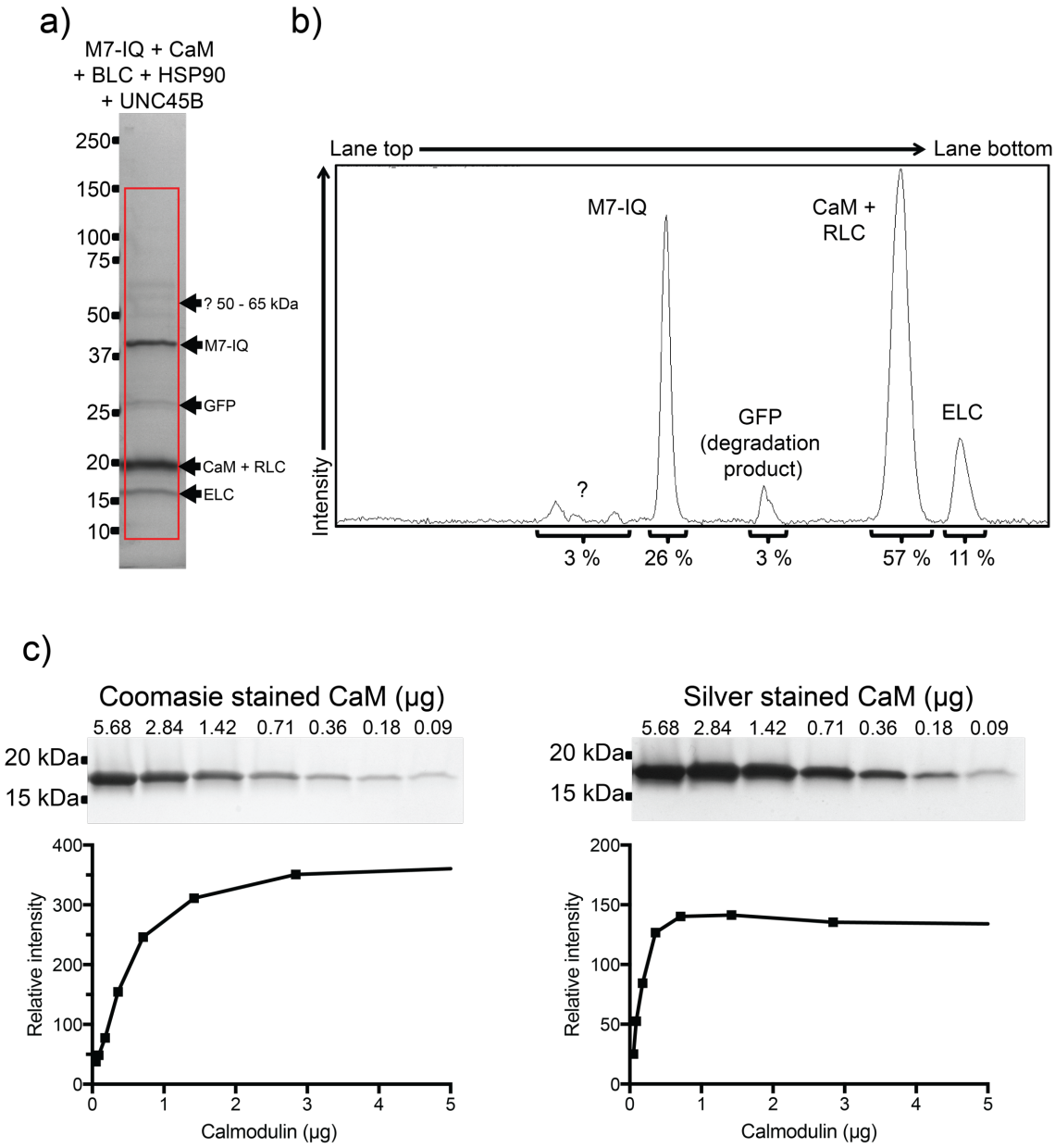
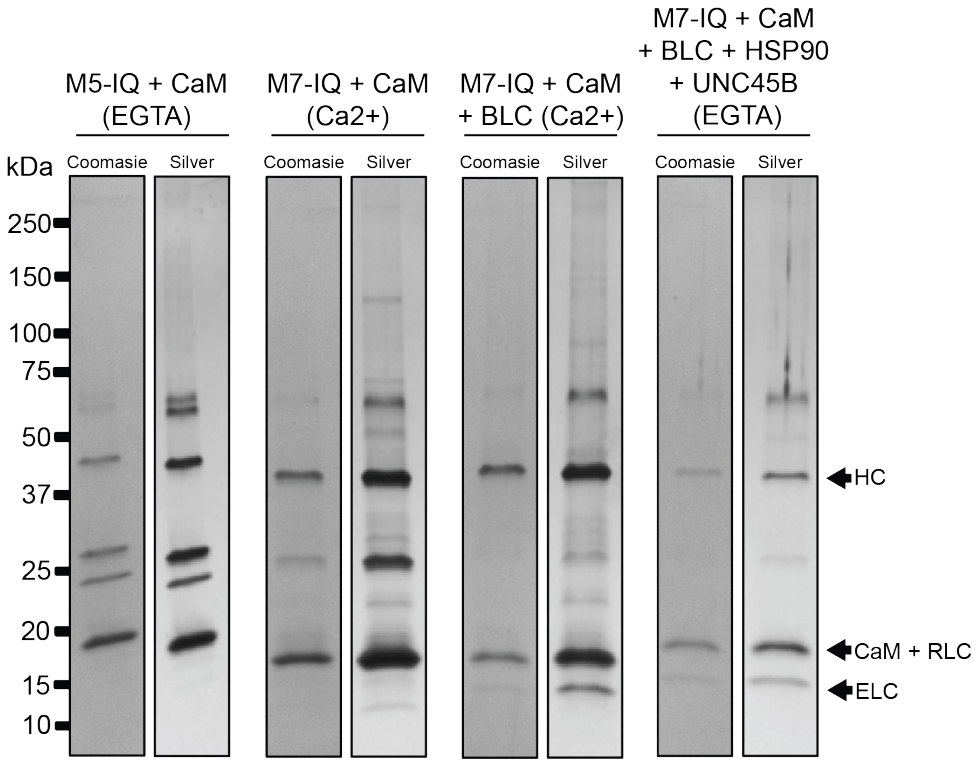


Figure 27 - Illustration of SDS-PAGE densitometry analysis for M7-IQ sample.

(a) a single lane is selected (red box) from an image of M7-IQ + CaM + BLC + HSP90 + UNC45B purified sample on a 4-20% SDS-polyacrylamide gel stained with coomasie. The M7-IQ (HC) band is identified along with CaM, RLC, and ELC and the various contaminating fragments such as the GFP. (b) using ImageJ software the image intensity profile of the bands from top to bottom (reading left to right on intensity plot horizontal axis) is shown. The percentages for the integrated peak areas calculated for this example are shown. (c) Decreasing amounts of calmodulin (µg) were analysed by SDS-PAGE gel stained with either coomasie (left) or silver stain (right). The relative peak intensities of each concentration was plotted to ensure staining was kept within the linear range.

a)



b)

	Ratio (mean + S.D) of LC to HC lever construct on coomasie stained gel		Ratio (mean + S.D) of LC to HC lever construct on silver stained gel	
	EGTA	Ca^{2+}	EGTA	Ca^{2+}
M5-IQ + CaM	4.08 ± 0.23	n.a	2.69	n.a
M7-IQ + CaM	1.02 ± 0.23	2.00 ± 0.14	n.a	2.48
M7-IQ + CaM + BLC	1.08 ± 0.09	0.74 ± 0.25	n.a	2.26
M7-IQ + CaM + BLC + HSP90 + UNC45B	2.94 ± 1.26	n.a	3.23	n.a

Figure 28 - Light chain binding to IQ lever constructs.

a) 4 - 20% SDS-PAGE coomasie gel and silver stained gels of lever constructs with various light chains. M5-IQ was purified by 4K GFP nanobody while M7-IQ was purified by his-tag. M7-IQ lever constructs were purified in the absence (1 mM EGTA) and presence (Ca^{2+}) of 0.1 mM calcium. (b) Values represent the ratio between light chain and lever construct (mean + S.D) from three independent experiments.

Native-mass spectroscopy was then used to estimate the numbers of light chains bound to the heavy chain for M5-IQ (Figure 29) and M7-IQ. For M5-IQ, the proteins identified corresponded to the calmodulin (16.7 kDa), GFP (27.2 kDa), 6 IQ motifs bound to six calmodulin light chains (120.7 kDa), and M5-IQ bound to six calmodulin (147.9 kDa), respectively. Detection of this full length construct bound to six calmodulin light chains confirms that the technique mass of the full length intact construct which remains intact following electrospray into the mass spectrometer.

Unfortunately, this approach did not work for M7-IQ. This protein could not be electrosprayed in the high salt buffer it is kept in to prevent aggregation, as this degrades the performance of native mass spectroscopy. When the M7-IQ construct was exchanged into 200 mM ammonium acetate with no salt, as required for this experiment, it is likely that the M7-IQ construct aggregated and failed to be electrosprayed. As a result, the masses of light chains were detected, but not those for heavy chains.

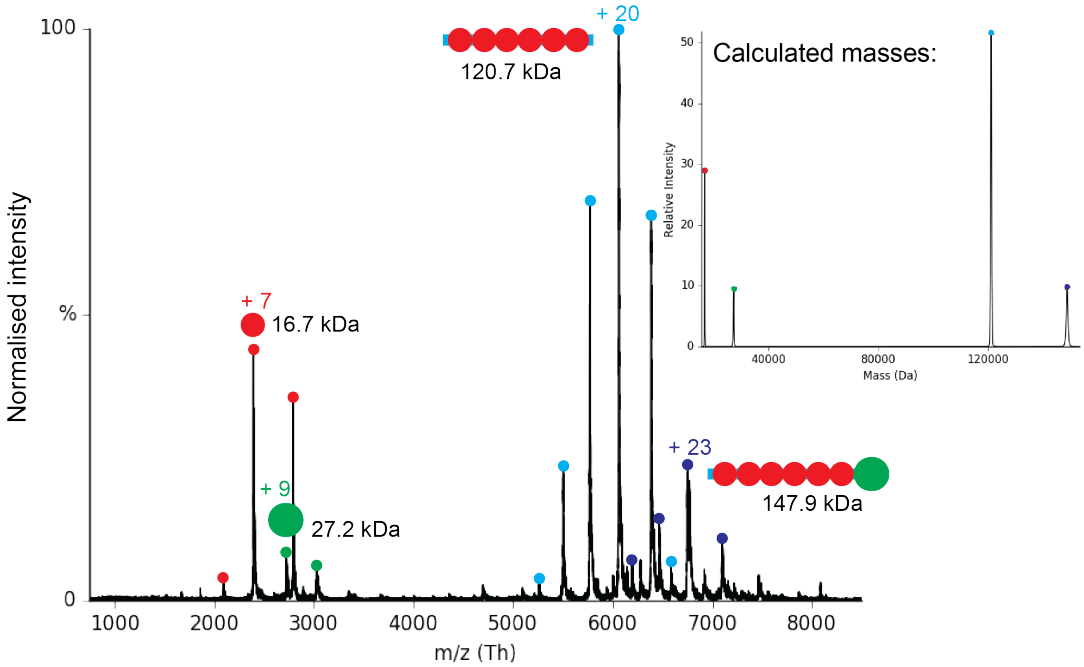


Figure 29 - Native mass spectra of M5-IQ.

Peaks showing the mass/charge (m/z) ratio against normalised peak intensity of the M5-IQ species detected by native mass spectrometry. Cartoon representations and expected masses of each species are shown next to the most abundant charge state. Species charge states (coloured) are calculated using (m/z)n-n=M, where n is the charge state and M is mass in daltons. Inset shows the calculated masses of the species present in the sample. Figure created using the Unidec software.

3.4 Discussion.

In this study, the myosin 7a 5IQ construct was found to only bind calmodulin in cells that endogenously express myosin 7a, and no unidentified light chain was discovered to bind to the 5IQ construct. The stoichiometry between the light chains and the 5IQ lever construct was determined to be ~ 3:1 by gel densitometry, suggesting some IQ motifs are unoccupied. Native mass-spectroscopy was unsuccessful in determining the intact complex mass of the myosin 7a 5IQ lever construct plus light chains due to issues with stability in low salt. This could be indicative of unoccupied IQ motifs, which tend to become insoluble in the absence of light chains as indicated by the bacterially expressed myosin 7a 5IQ (SUMO_5IQ) construct.

Pull downs experiments showed that both the M5-IQ and the M7-IQ constructs only bound calmodulin in both B16 and AD-293 cells. The result for M7-IQ supports earlier findings using bacterially expressed M7-IQ constructs, which also found that calmodulin was the preferred light chain (7). As B16 cells endogenously express myosin 7a, it is likely that if there was an unidentified light chain that does bind one or more of the IQ motifs in myosin 7a, it would have been present. However, no unidentified light chain was found to bind to M7-IQ. Myosin 7a is expressed in the stereocilia of hair cells, the pigmented epithelium and the photoreceptor cells of the retina (Hasson et al., 1995; Hasson et al., 1997), as well as kidney, liver, testis, cochlea, and lymphocytes (Gibbs et al., 2010; Lopes et al., 2011; Grati and Kachar, 2011b). However, B16 cells are derived from melanoma cells derived from the skin. It is possible that different light chains are expressed in other tissues such as hair cells, or the retina, which could be tested in future. Co-expression of M7-IQ with calmodulin, RLC and ELC does demonstrate that the IQ motifs can bind all three light chains, in contrast to the results from pull down experiments. However, this might be in part due to the high levels of expression of all three light chains in Sf9 cells, which does not represent the conditions *in vivo*.

Light chains other than CaM, RLC, and ELC have been identified to bind myosin. These include calmodulin like protein (CLP), which binds myosin 10 (Rogers and Strehler, 2001), and TgMLC1, a myosin light chain that co-precipitates with *Toxoplasma gondii* myosin A (TgMyoA) (Herm-Götz et al., 2002). Interesting, the TgMLC1 is larger than typical light chains (~15 – 18 kDa) at ~24 kDa in size. This

is in part due to an additional ~70 residue N-terminal extension which has been shown to form a complex with TgMyoA, TgGAP45, and TgGAP50 (Frénal et al., 2010). A crystal structure of TgMLC1 (5VT9) of the core calmodulin-like domain (73 – 211) shows its structure is highly similar to that of calmodulin. However, the N-terminal extension is not visible in the crystal structure. Interestingly, TgMLC1 is also predicted to contain at short coiled coil of about 70 – 90 residues by structural prediction programs.

The stoichiometry of light chain to heavy chain determined for myosin 7a also suggests that not all of the IQ motifs are occupied, but does not reveal which IQ motifs are unoccupied. The stoichiometry determined here is in agreement with previous studies (Yang et al., 2009; Sakai et al., 2015) using constructs expressed in bacteria and Sf9 cells respectively. The earlier finding that IQ3 is unoccupied, suggests that ~6.3 turns of the alpha helix in the central region of the lever would be solvent exposed. This would be expected to introduce compliance within the lever, which is hard to reconcile with the need of the lever to transmit force.

The densitometry experiments showed only 4.1 mols calmodulin per heavy chain for M5-IQ, which is lower than expected, and highlights the difficulty of non-linear staining of proteins for accurate determination of stoichiometry. Experiments with M7-IQ cannot therefore be reliably interpreted using this approach. Native mass spectrometry was able to demonstrate that M5-IQ binds 6 light chains. Unfortunately, due to the low solubility of M7-IQ in low salt, this approach did not work for this construct.

These results need to be further consolidated to answer some of the unresolved questions of this chapter. Perhaps of most interest from a biological perspective is that the myosin 7a lever only binds calmodulin *in vivo* and that only ~3 light chains apparently bind the lever. It would interesting to pursue another method for accurately determining the stoichiometry between the lever domain and light chains. Cross-linking may help, as this will prevent light chains from falling off the IQ motifs if the affinity is low. It would be interesting if a complete structure of the myosin 7a lever could be solved, either by cryo-EM or X-ray crystallography. Although if one or more of the IQ motifs are not occupied, the increased flexibility would likely make this extremely difficult.

4. Investigating the flexibility of myosin 7a by negative stain electron microscopy and image processing.

4.1 Introduction.

Negative stain EM can generate a lot of useful information about the structure and flexibility of macromolecules, as outlined in Section 1.17. With careful consideration of the potential flaws associated with the technique, structures generated by negative stain agree well with corresponding crystal structures, such as the structure of dynein (Roberts et al., 2009; Kon et al., 2011). This is especially the case when negative stain EM is allied with image alignment and classification methods (Ohi et al., 2004).

In EM, single particle analysis is fundamentally an averaging technique, conceptually similar to crystallography. While in crystallography, proteins orient themselves into a lattice structure, in EM this alignment and averaging must be performed *in silico*. It is this averaging process that allows the extraction of high resolution information from images with low contrast and signal to noise ratio. The key step is determining the orientation for each individual particle and calculating its angular relationship to the other particles in the dataset. Particles are then classified based on orientation prior to averaging. Several approaches have been used to implement this in alignment and classification in 2D including 'K means classification', 'equal distributed K means' and 'maximum likelihood' approaches (Sigworth, 1998).

First, a referencefree alignment is performed in either RELION (**RE**gularised **LI**k elihood **O**ptimisatio**N**) (Fernandez-Leiro and Scheres, 2017) or with Imagic (Van Heel et al., 1996). Alignment and classification with RELION and Imagic are in essence the same, however within RELION the alignment procedure is hidden behind the 2D classification graphical user interface (GUI), and not much customization can be done to improve the alignment. In Imagic the control and adaptability make it a much more powerful tool for 2D classification and having the ability to control the alignment and the pixel offset is much better for generating higher structural detail. In addition, masks can be applied to an area of particular interest to discover any subtle variations in this region, which may not be observed when considering the entire molecule. Regardless of this, the alignment procedure is first done in an unsupervised manner without the user

providing a reference structure. This removes some potential user bias as the particles cannot be forced into an incorrect alignment or class by the user. The particles are classified into a number of classes, typically on the order of 30 – 50 particles per class in early classifications. As the data became more homogeneous with subsequent iterations, the number of particles per class can then be increased to around 50 – 100 particles per class.

This methodology improves sample homogeneity by removing contaminants from the dataset for downstream applications. From here, this classification approach can take two paths; either continuing with 2D classification, to identify flexibility within the macromolecule using a series of focused masks in Imagic, or continuing onto 3D reconstruction.

My own and published work seems to suggest that not all of the IQ motifs in the neck of myosin 7a are occupied by calmodulin (or other light chain) (Udovichenko et al., 2002; Sakai et al., 2015)(Chapter 3). This is important to address as lack of light chain occupancy in the lever is likely to contribute to its flexibility. Therefore, it was important to try to determine the structure and flexibility of the neck region of this motor. To achieve this, I used a previously generated dataset of negatively stained myosin 7a (Yang et al., 2009), to allow me to compare the appearance of the lever with models of the 5IQ motif fully occupied by light chains, that were constructed using crystal structures of light chains on IQ motifs and to look for evidence of flexibility and incomplete decoration. This previous dataset was used, as the large size of the dataset and the high quality of staining were such that it would enable me to look for new structural details and conformational variety not previously described by single-particle image processing.

My first aim was to construct a 3D model of myosin 7a in the regulated state from negative stain EM images and assess where the different myosin domains may fit within this 3D structure. Secondly, I aimed to compare the head domain appearance obtained in the 3D structure with atomic models of the myosin 7a head, with the aim of fitting the atomic model into the 3D structure to characterise how the lever would fit into it. Finally, I aimed to use image classification to look for evidence of flexibility in different parts of the head. Two main types of flexibility were investigated: (i) motor tilting about the motor-lever junction, and (ii) lever

bending flexibility. I used measurements made on these molecules to predict the bending stiffness of the myosin 7a lever for modelling work (Chapter 5).

In parallel work, I also attempted to use negative-stain EM to further characterise the flexibility of the myosin 7a lever and compared this to the lever of myosin 5a, a well characterised domain within the literature, using baculoviral expressed lever constructs.

4.2 Materials and Methods.

4.2.1 Protein samples and negative stain EM procedures.

4.2.1.1 Full length myosin 7a.

A pre-existing negatively-stained dataset of full length recombinant myosin 7a molecules (Yang et al., 2009), was used. The sequence of the recombinant myosin 7a heavy chain was from *Drosophila Melanogaster* (NCBI GenBank accession: NM_010864) and contained amino acids 1-2167 with a FLAG tag epitope (DYKDDDDK) added at the C-terminus to facilitate purification. Myosin 7a constructs were cloned and co-expressed with CaM by Dr Yi Yang using a Baculovirus/Sf9 cell expression system (Yang et al., 2005) and purified as previously described (Yang et al., 2009).

Molecules were deposited on carbon- coated 300- or 400-mesh copper grids and stained with 1% (w/v) unbuffered uranyl acetate, and images were captured at x40,000 magnification using a JEOL 1200EX electron microscope (Yang et al., 2009). Micrographs of fields of negatively-stained myosin 7a molecules were recorded on Kodak S0-183 film (Eastman Kodak, Rochester, USA) using a 1 s exposure time. Films were developed and digitized as 16-bit grey scale TIFF images using an Imacon 848 scanner (Imacon / Hasselblad) at 1270 ppi. Object sampling in the image was calibrated by measurement of the 14.4 nm repeat of paramyosin filaments. This calculated as 0.52 nm/pixel for 40,000× nominal magnification. The TIFF files were subsequently converted to eman (.hdf) or RELION (.mrc) format using the *e2proc2d* programme (Tang et al., 2007).

4.2.1.2 Lever constructs.

Two lever constructs used here were the M7-IQ and M5-IQ constructs from myosin 7a and myosin 5a respectively, as described in chapter 3.2.4. EM was performed on fresh and previously drop-frozen samples, and no noticeable difference was observed. Purified M5-IQ and M7-IQ lever constructs (Chapter

3.3.3) at a concentration of ~100 nM, supplemented with 200 nM excess calmodulin were allowed to adsorb from solution to a thin carbon film supported by a copper mesh grid. All proteins and buffers were kept on ice at 4˚C prior to application to the carbon coated grid. To perform the staining, a grid was held in tweezers and 7 µL of sample applied to the carbon film side using a pipette. After 30 – 60 s the grid was flicked to remove the excess liquid. A drop of 1% (w/v) aqueous uranyl acetate (at ambient temperature) was applied using a Pasteur pipette. On contact the droplet was quickly flicked off the grid and then a further 5 – 7 drops were applied in the same way. Grids where stain had contacted the underside of the grid were discarded. Excess liquid was then wicked away from one side of the grid by touching the grid edge with torn filter paper (Whatman, Ashless grade #42 Filtration paper). The grid was then allowed to fully dry for a few minutes before viewing by EM.

The grids used for these experiments were carbon coated 400 mesh copper grids (Agar Scientific), and kindly provided by Martin Fuller (University of Leeds). Prior to adding the sample, the grids were glow discharged using the PELCO easiGlow™ Glow Discharge Cleaning System (Ted Pella), operating at 10 mA for 1 min. Alternatively, the grids were irradiated with ultraviolet (UV) light for 20 – 30 mins, using a low-pressure mercury vapour lamp (type R51, UV products Inc., Pasadena, CA, USA). The lamp has the low pass (black light) filter removed and emits over a broad range of wavelengths. Grids were placed ~5cm from the bulb, with their carbon side upwards and irradiated while enclosed in a cardboard box (in a fume cupboard) to retain ozone. Following UV treatment, grids were used immediately. Both UV treatment (which generates ozone), and glow discharge increases the hydrophilicity of the carbon surface, allowing the hydrophilic surfaces of proteins to readily adsorb. Generation of a hydrophilic surface also promotes a thin and even layer of stain on drying, and it may also promote interactions between the protein molecules and carbon that stabilise the protein against damage from staining (Walker et al., 1985).

Negatively-stained grids were examined using either the FEI Technai T12 transmission electron microscope equipped with a LaB_6 filament operating at 120 kV accelerating voltage, or a FEI Technai F20 transmission electron microscope equipped with a FEG, operating at 120 kV. Before use, the microscope apertures were aligned and a grid sample adjusted to the eucentric point using the

goniometer. Objective astigmatism was corrected.

To record images, grids were examined at ~10,000x magnification to locate optimal areas of negative staining. Once a suitable area was located images were collected at 30,000× magnification (3.74 Å/pixel) at a defocus of ~1.5 µm. Once the images had been collected, they were converted into .hdf format using the *e2proc2d.py* command in EMAN2 (Tang et al., 2007).

4.2.2 Image processing procedure.

The general work-flow for image processing of negative stain EM images was as follows. The steps in the process were: (i) particle 'picking' and windowing-out from micrographs, (ii) iterative reference-free rotational and translational alignment of images and (iii), classification and averaging of the images. A fourth optional step was 3D classification of particles.

4.2.2.1 Particle picking.

Individual micrographs were opened in EMAN2 software. Particles were manually picked by setting the cursor on the center of the particle. The center of the particle is defined as the rotationally averaged centroid. The particles were then windowed in a box 1.5x the size of the particles longest axis (normally 120 x 120 pixels). To increase the throughput, the *SWARM* (Tang et al., 2007) semi-automatic picker was used. A number of reference particles were selected to provide a template for the *SWARM* autopicking software. *SWARM* then attempts to find any density in a micrograph that matches the templates provided, with additional user parameters, including approximate particle diameter, picking threshold and proximity, to minimise the selection of contamination, aggregation or particles lying on the grid boundary. Semi-automated picking is faster and less laborious than manual picking. To avoid introducing bias into the data, the template needs to be filtered to low (> 60Å) resolution (Henderson, 2013). Additionally, if the templates only represent a subset of the particle orientations, this may bias which particles are subsequently picked automatically, which can introduce artefacts.

4.2.2.2 Reference-free classification.

Selected particles were subjected to iterative rounds of reference free 2D alignment and classification in RELION (Figure 30) (Fernandez-Leiro and Scheres, 2017). For myosin 7a, this classification served two purposes, firstly, to identify poorly aligned particles or contaminants which were subsequently

removed and, secondly, to split the myosin 7a dataset into extended and regulated conformations. Regulated molecules form in low salt in the presence of ATP and analysis of the particles in these conditions revealed that only ~50 % of the molecules were in the regulated conformation (2215 out of 4180). Regulated (2215) and extended (4721) were subsequently taken forward. From here, extended particles were further processed by *IMAGIC* (described in the next section), while regulated particles were subjected to 3D-classification with RELION (described in section 4.2.4). For the lever constructs, a reference free 2D classification step was used to remove poorly aligned particles and contaminants.

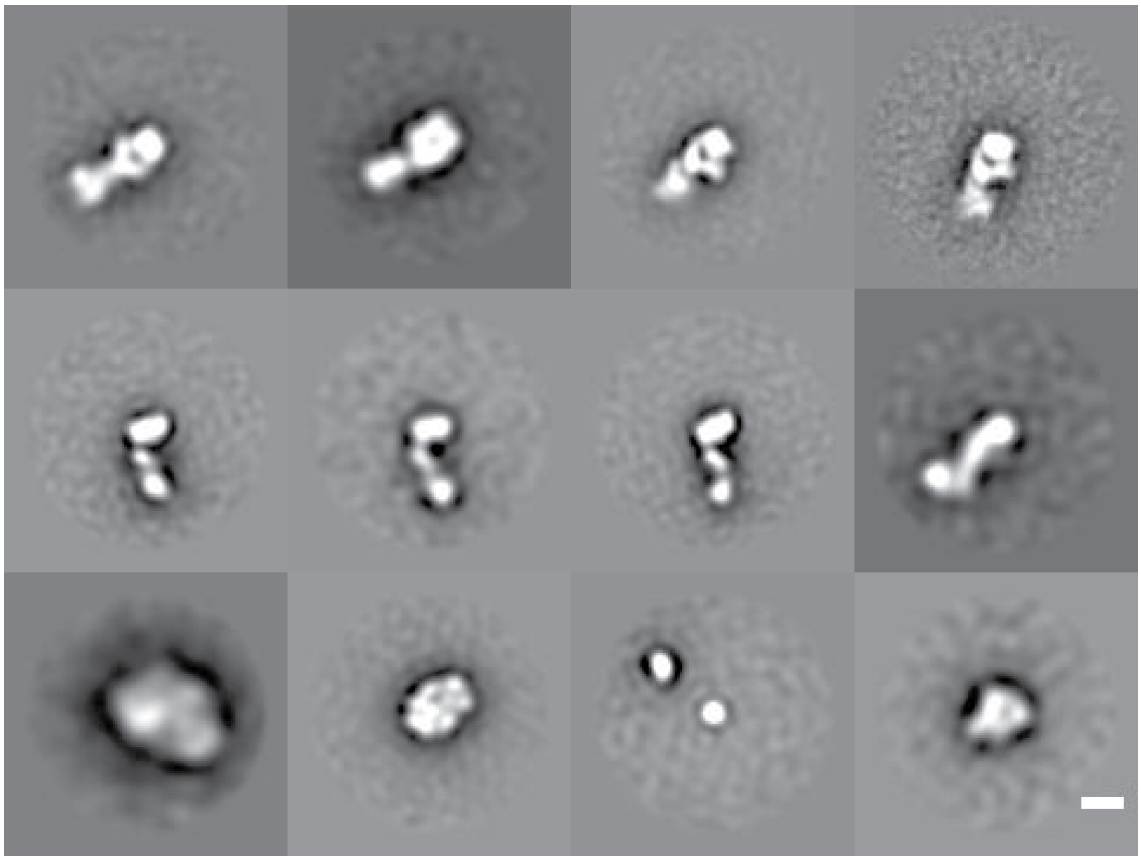


Figure 30 - Particle sorting with RELION.

The particles selected from the dataset were subjected to iterative rounds of 2D classification with relion, a subset of selected classes is shown. This split the particles into regulated particles (top row), extended particles (middle row), and contaminants (bottom row). The contaminants were removed following each round of classification. Three rounds of classification were performed to remove junk data. Scale bar 10 nm.

4.2.3 Analysis of extended molecules and lever constructs.

Extended particles were subjected to 2D classification using the IMAGIC 5 software package (Van Heel et al., 1996). The Imagic workflow (Figure 31), follows the general outline (i) reference-free rotational and translational alignment of images, (ii), selection of appropriate references to be used to perform multi-reference-alignment (m-r-a), and (iii) classification and averaging of the images. This is followed by (iv) iterative rounds of m-r-a and classification of the images. A total of 5521 extended full-length myosin 7a particles, 7882 (M5-IQ) and 351 (M7-IQ) lever construct particles, respectively, were processed and analysed using the IMAGIC software package. IMAGIC commands, denoted hereafter in `Courier font` in parentheses *e.g.* (`pretreat-image`), were issued at the command line or as a script of commonly used commands.

4.2.3.1 Image pre-processing.

Windowed particles were imported as an image stack into IMAGIC. The particles were first centred using (`center-image`) to center all the particles about a rotationally symmetrized version of the total sum of all particles. Images were then treated using (`pretreat-image`) to normalise the variance. This removes excessively bright or dark pixels (*e.g.* caused by film scratches or dust), and normalises the pixel intensity to a sigma value of 10. The particles were saved as an image stack.

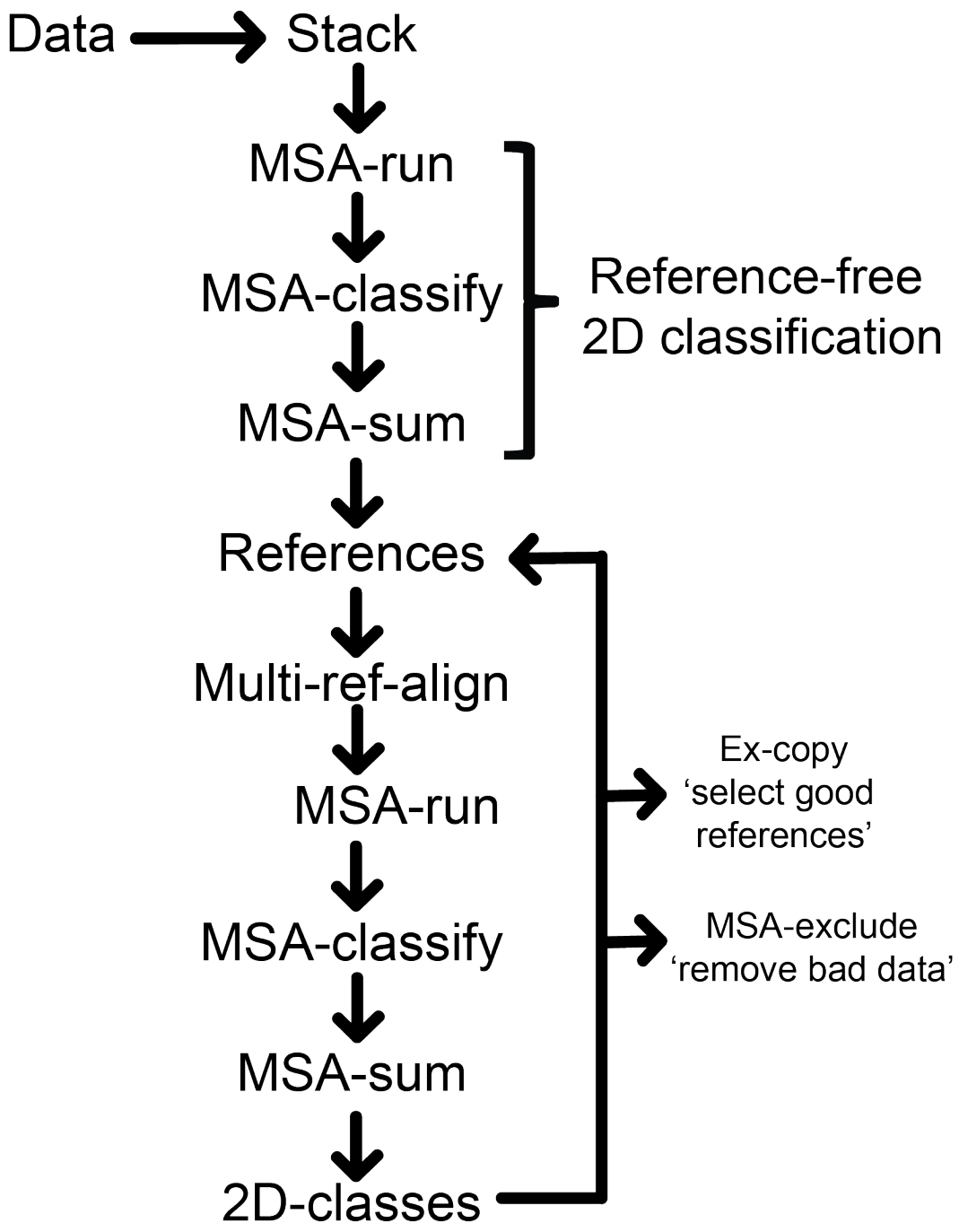


Figure 31 - Imagic workflow.

Data is imported into Imagic as an image stack. Reference-free alignment is performed from which good references are selected. Multi-reference alignment with the selected references is performed to align the particles to the references before classification is repeated. Iterative rounds of classification and selection of references is performed, with 'bad data' being removed.

4.2.3.2 Image alignment and classification.

Reference-free alignment using (`MSA-run, MSA-classify, & MSA-sum`) commands was performed to classify all particle sets (extended myosin 7a, M5-IQ, and M7-IQ lever constructs) into classes containing ~50 particles per class (for example 5521 extended myosin 7a particles into 110 classes). A circular mask, sufficient to encompass the whole particle was applied to the images. The class averages were visualised and a number of representative classes were selected (`extract-image`) for multi-reference alignment. The class averages were aligned (`prepare-mrarefs`), to maintain orientation and keep the alignment centered between iterations.

The prepared references were used to perform a fresh alignment of the pre-treated particle stack (`m-r-a`). This command performs a set of multi-reference alignments, one after the other, which includes translational and rotational alignment. Fresh multi-reference alignments were performed after each iteration to avoid additive interpolation errors caused by multiple transformations. The resulting aligned image stack was used as input to the next round of alignment (Figure 32). The alignment strategy adopted involved multiple rounds of multi-reference alignment and classification. Bad particles / junk were removed from the image stack using the (MSA-extract-classes) command during subsequent iterations. The final aligned images were oriented with the motor domain in the center of the window and the lever extending towards the top right corner. This procedure generated a final working image stack of 4721 aligned extended myosin 7a images, 5989 aligned M5-IQ, and 348 aligned M7-IQ lever construct images.

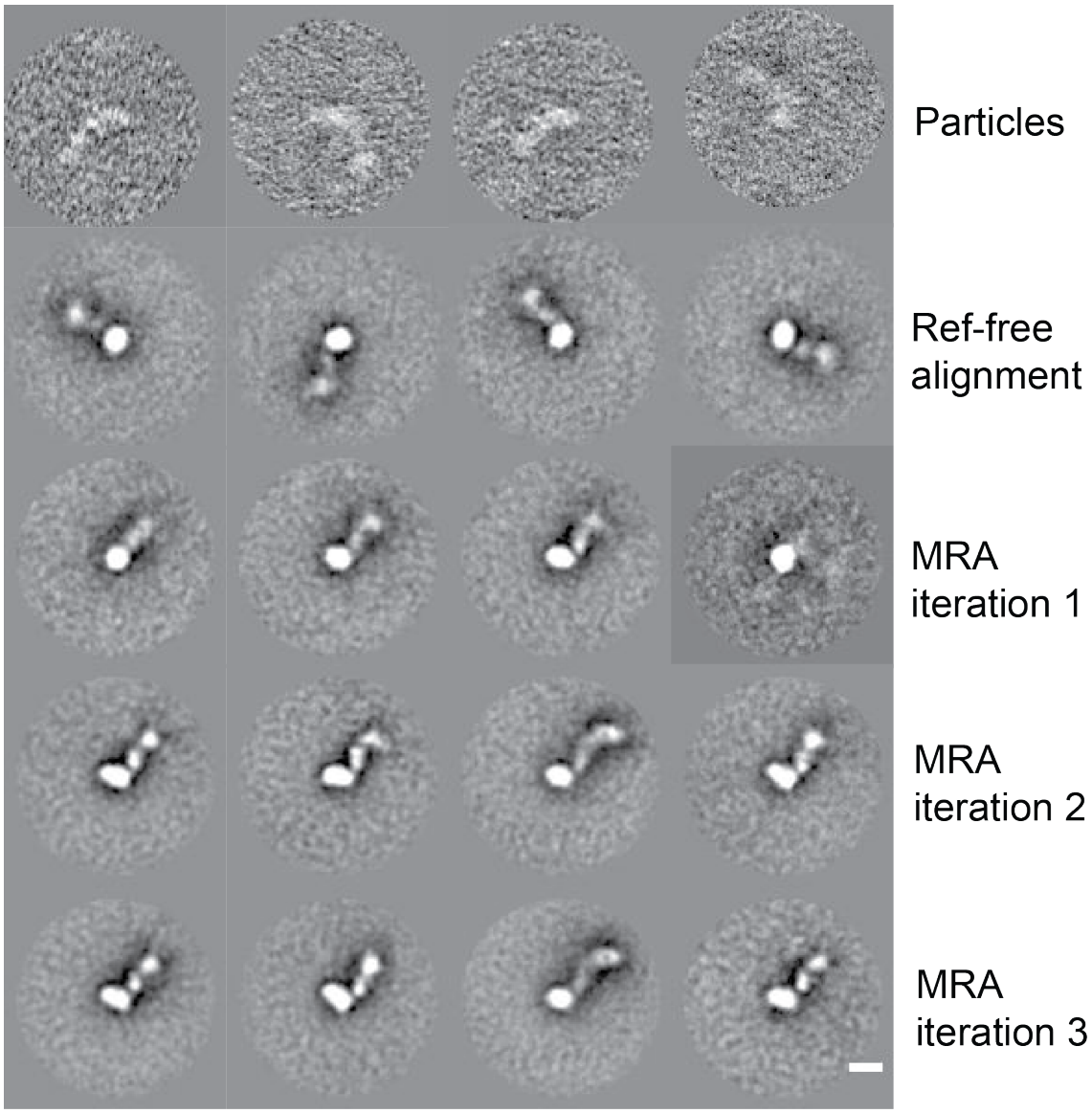


Figure 32 - IMAGIC image alignment and classification.

Individual centered particles (top row) are classified by reference-free alignment to produce class averages of similar particles (subset shown). A selection of references are chosen to perform multi-reference alignment to align the individual particles. This is performed iteratively (selection of class averages from each iteration shown) to align the particles and remove junk data. Scale bar 10nm.

4.2.4 3D classification of compact structure.

4.2.4.1 Generation of homology models.

Homology models of extended and regulated myosin 7a were constructed in order to facilitate generation of a 3D model from the EM dataset. A crystal structure for the motor domain of myosin 7a is lacking, therefore crystal structures of myosins with similar amino acid sequences to myosin 7a were used as templates to model the motor domain using ITASSER. The myosin 7a tail domain (amino acid residues 956 – 2215) was generated using two crystal structures of both the N- (PDB ID: 3PVL) and C- (PDB ID: 5MV7) terminal portions of the myosin 7 tail as a template. This homology model is referred to as M7tail from now onwards.

The 5IQ lever domain was constructed using the 2.5-Å resolution structure of apo-calmodulin bound to IQ motifs 1 and 2 of murine myosin 5 (PDB ID: 2IX7) that are 23 residues apart (section 1.6.1). Chimera was used to superimpose the backbone atoms of N-terminal CaM from a second 2IX7 structure onto those of the C-terminal CaM of the first 2IX7 structure, giving a 3IQ model. This process was repeated with a third and a fourth 2IX7 structure to give a 5IQ model with all IQ motifs having 23-residue spacing. Overlapping CaM and heavy chain sequences were deleted, and the segments of the backbone IQ helix were ligated and replaced with the correct residues of myosin 7a heavy chain.

To create pre-powerstroke head models (Motor - 5IQ+CaM), I-TASSER was used to model the head molecules using the scallop myosin pre-powerstroke structures (PDB ID: 1QVI) as a template. The myosin 7a head model was superposed onto the scallop myosin model, then the backbone atoms of the first IQ motif of the 5IQ lever model were superposed onto the first IQ motif of the scallop myosin. The scallop myosin model was then deleted and the head and 5IQ lever models ligated together. In addition, shorter head models consisting of 3 or 4 IQ motifs were also generated. Another domain within the lever, the SAH domain (858 – 956), was not included in these homology models as it is only ~1nm in diameter, which is much smaller than the grain size used in negative stain and is thus unlikely to be seen.

A model structure of myosin 7a in its compact form was constructed by aligning the head (Motor – IQ+CaM models) and M7tail homology models in CHIMERA to roughly approximate the structure of the compact form reported in the literature

(Yang et al., 2009; Sakai et al., 2015).

4.2.4.2 3D model construction.

The compact myosin 7a 5IQ homology model filtered to 60 Å was used as a starting model. The 2215 particles in the compact form were subjected to 25 rounds of 3D-refinement using RELION.

4.2.5 Analysis of motor-lever flexibility.

To analyse the flexibility in the motor-lever junction, the 4721 extended molecules, were realigned on the lever domain using the protocol described in section 4.2.3.2. A mask encompassing the motor domain and immediate surrounding area (excluding CaM-1) was drawn (Figure 39a), using the global average as a guide to determine an appropriate width. Image classification was then performed using *IMAGIC, as described in 4.2.3.2* using the motor-only mask to define the region of interest. Multiple classifications were performed (data not shown), splitting the images into differing numbers of classes as to obtain the largest number of classes while maintaining sufficient signal to noise. Division of the data into 90 classes (~50 images/class) was judged to be the best compromise between encompassing variability in the images and reinforcement of structural detail through class averaging.

ImageJ software (v1.43) (NIH, USA) was used to measure the angle between two straight line segments connecting three points positioned to define the motor-lever angle (Figure 39c). First, to ensure more precise hand-marking and to avoid artefacts arising from the constraints of the low-resolution pixel grid, the source images were first interpolated, by bicubic Fourier-based interpolation in *ImageJ* (v1.48v), to ×5 original size (from 120×120 to 600×600 pixels). For consistency in marking, points defining the tail and motor-lever junction were kept the same for each calculation as attempting to interpret the points that encompass the motor-lever junction in each class average would be variable. The variability in the appearance of the start of the lever makes interpretation of its precise angle of projection difficult and each lever average is, in any case, formed from a distribution images.

4.2.6 Analysis of lever constructs and spline fitting.

In order to analyse the lever shapes adopted by the lever constructs, the path of the underlying lever was marked out by hand for each average of the aligned lever construct image following classification of the images into its various shapes

using Easyworm (Lamour et al., 2014). Firstly, to enable more precise hand-marking and to avoid artefacts arising from the constraints of the low-resolution pixel grid, the source images were first interpolated, by bicubic Fourier-based interpolation in *ImageJ* (v1.48v), to ×10 original size (from 140×140 to 1400×1400 pixels) and then converted to 8-bit TIFF format. The class averages were then opened in Easyworm1 for marking. The following processing steps were then performed (Figure 47).

First, the scale of the image was set (0.374 Å/pixel) and the path of the lever was marked out, progressing from GFP-lever junction to the end of the lever, by clicking with a mouse along the implied lever helix path, using the 'Select Chain' tool. Care was taken to be as consistent as possible across the images marked. For example, the first mark was always positioned at a similar point at the centre of the motor lever junction, and, where possible, ending with a mark on the distal edge of the final CaM.

The number of coordinates marked per image was variable but ranged from ~ 4 – 7 along the lever. Images were only marked if the contour of the lever was clearly resolved and their position could be unambiguously determined, otherwise the image was rejected. Once the coordinates had been selected, Easyworm1 was used to fit a parametric spline using a least-square fit algorithm to connect points. Testing a range of trial spline fitting procedures by altering the 'fitting parameter' showed that a value of 125 was optimal to find the best fit shape (not shown). Of the 200 class averages produced, 174 classes were marked, with the remaining 26 omitted due to the lever's contour being too ambiguous or ill-defined. The fitted spline, which contains a much greater number of coordinates with much smaller separation, that typically contained ~ 30 points per curve. This means it had a resolution of ~5 times greater than the hand-marked points. The points were then saved and exported to Easyworm2, for further analysis (see section 4.2.7).

4.2.7 Calculation of persistence length.

The data generated by Easyworm1 was imported into Easyworm2 and the temperature was set to 25˚C (the approximate temperature at which the grids were prepared). The persistence length was calculated using the tangent-tangent correlations approach initially used by *Frontali et al., (1979)* to measure the persistence length of DNA (Frontali et al., 1979). This method monitors the

change in cosθ as ℓ increases according to Figure 33.

$$< cos\theta > = e^{-\frac{\ell}{sP}}$$

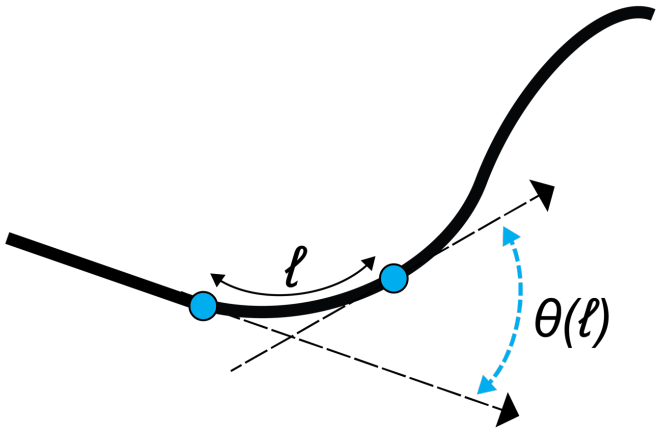


Figure 33 - Calculation of persistence length using the tangent-tangent correlations method.

The thermal bending of flexible polymers is commonly discussed in terms of the three-dimensional persistence length Lp = EI/kT, which is the arc length above which the angle θ(ℓ) becomes uncorrelated in three dimensional motion. Specifically, relating to the equation above, where θ is the angle between two segments (blue dots) of the spline separated by a distance ℓ along the chain contour. Since the chain is constrained to 2-dimensions, instead of 3-dimensions, a surface parameter of 2 is set. P is the persistence length.

In the above example, ***θ*** is the angle between two segments of the spline separated by a distance ***ℓ*** along the chain contour. ***s*** is a surface parameter that is set by the user to a value of 2 for chains that have equilibrated on the 2D surface and ***P*** is the persistence length. Trial binning procedures were attempted to find the best fit between show the full spread of the data while not introducing large numbers of outliers that could skew the fitting procedure done by Easyworm2 to find the persistence length (not shown). The data was subsequently fitted to 60 bins.

4.3 Results.

4.3.1 Global alignment of extended particles.

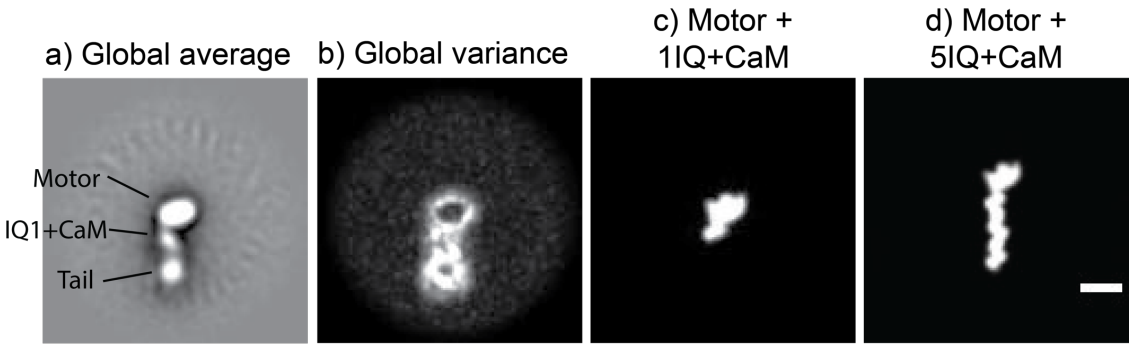
A global average image of 4721 extended particles (Figure 34a) shows the position of the motor domain at the top of the image, as well as the first calmodulin. The global variance image demonstrates that there is considerable variability in the amount of stain around the molecules (Figure 34b). Comparison with the projection models of the myosin 7a motor (ADP.Vi) with either the first IQ motif plus calmodulin (CaM) (Motor+1IQ+CaM, Figure 34c) or five IQ motifs plus CaM (Motor+5IQ+CaM, Figure 34d) helps to confirm the identity of the motor domain in the global average image (Figure 34a). In the global average, the motor was measured as ~10 nm along its long axis and ~6.5 nm across its width.

As reported in the previous analysis of this dataset (Yang et al., 2009), the expected five CaM substructure of the lever is not visible in the global average. Only a single CaM appears to be visible, which we assume is the first CaM, as an elliptically-shaped blob (Figure 34a). The dimensions of this CaM structure are ~5.8 nm along its long axis and ~3.5 nm across its width. This is consistent with the projection image in Figure 34c, where the CaM measures ~5.5 nm along its long axis and ~4 nm across its width. Downstream of the first CaM is a larger, less well-defined density, which is likely to encompass the rest of the lever and the tail. The less well-defined structural detail of this density could arise from the averaging procedure, and the large amount of variability in structure downstream of the first IQ motif.

The overall length of the molecule additionally appears much shorter than expected. Calculating the expected length of the lever, the rise per residue in an α-helix is 1.5 Å (0.15 nm), and each IQ motif in Myosin 7a is 23 residues long (Figure 34e), suggesting that the lever should be approximately 17nm long. However, in the global average the entire length of the molecule (including motor, lever, and part of the tail) is only ~22 nm.

The variance image (Figure 34b) indicates that there is significant variability in the appearance and staining pattern downstream of the first IQ-CaM, with some variance being seen around the first IQ-CaM and motor domain. Such variance can arise simply from variation in staining around the molecule, but it could also be an indication that the lever may be sufficiently flexible that it takes up a range

or different orientations relative to the motor.



e) Alignment of 5 IQ motifs in Drosophila myosin 7a

```
DLFLEQERDRVLTRKILI LQ RSI RG WVY R RRFL
          RLRAAAIT VQ RFW KG YAQ R KRYR
          NMRVGYMR LQ ALI RS RVL S HRFR
          HLRGHIVG LQ AHA RG YLV R REYG
          HKMWAVIK IQ SHV RR MIA M RRYRKLRLEHKQFA
```

Figure 34 - Global alignment of extended particles.

(a) Global average of the full length myosin 7a (M7) molecules generated from 4721 particles, aligned on the motor domain. (b) variance image of 4721 particles. Whiter pixels indicate more variance across the images. Global average and variance images were generated in IMAGIC. (c) 30 Å projection image generated by RELION of scallop myosin S1 in its pre-powerstroke state (1QVI), containing the motor with 1 IQ motif plus calmodulin and (d) projection image of the motor with 5 IQ motifs plus calmodulin. (e) alignment of the 5 IQ motifs in Drosophila myosin 7a. Consensus core residues of the IQ motifs are indicated in red. Note that each IQ motif is only 23 residues long. Scale bar 10nm.

A comparison of the lever in the global average with the lever in a projection image of scallop myosin S1 in the pre- powerstroke state reveals that the calmodulin appears to have a different tilt direction in the global average to that of the ELC bound to IQ1 in scallop myosin S1 (Figure 34). The left-down (LD) projection model (Figure 35b), whereby one lobe of the calmodulin (the left hand side lobe when inspecting the class average) appears lower than the other lobe (the right hand lobe), in which the first CaM is in its expected orientation on the 1st IQ of the lever, is not well matched to the appearance of CaM in the global average (Figure 35a), which appears in the right-down (RD) appearance. Rotating the CaM by 120˚ in the projection image (Figure 35c, right-down (RD) projection model) shows a better match with the appearance of the CaM in the global average. This suggests that the calmodulin in the global average is rotated ~ 120˚ compared to what would be expected.

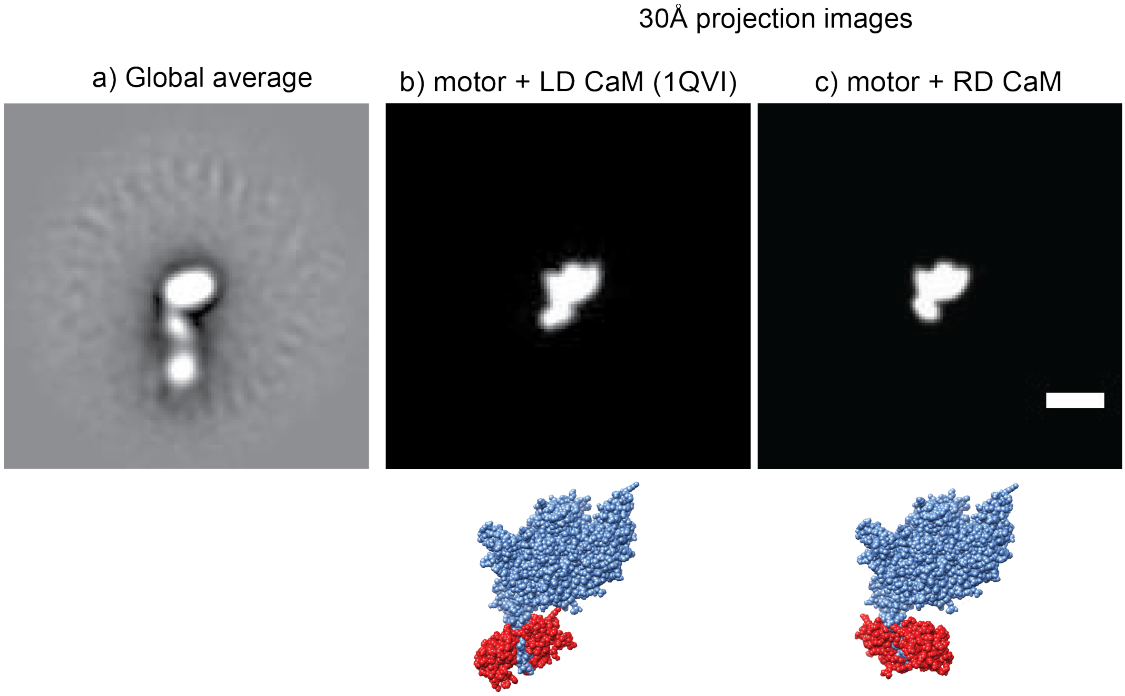


Figure 35 - Comparison of the global average image for myosin 7a, with projection images of scallop myosin S1 1in its pre-powerstroke state (1QV1).

(a) Global average image. (b) projection of 1QVI of the motor with the first CaM in a 'left down' (LD) conformation. (c) motor with the first CaM in a 'right down' (RD) conformation. Related atomic structures are shown beneath the projection images. Scale bar 10nm.

A mask was drawn around the 1st calmodulin (using the global variance as a guide) and applied to the aligned particles (Figure 36). Reclassification of the masked area should reveal if a large degree of compliance exists around the 1st IQ motif, and whether the calmodulin can fall in different orientations on the carbon film. The classifications reveal that 36% of the calmodulin seen in the class averages appear predominantly in the RD-orientation, while the remaining views fall into the straight appearance (19%), LD-orientation (15%), or undefined (30%). While the RD-orientation appears as the preferred orientation, the variation seen in the relative positioning of the calmodulin along the lever which suggests that the calmodulin can axially rotate independently of the surrounding parts of the head. This is exemplified by the number of particles (30%) that appear in an undefined view.

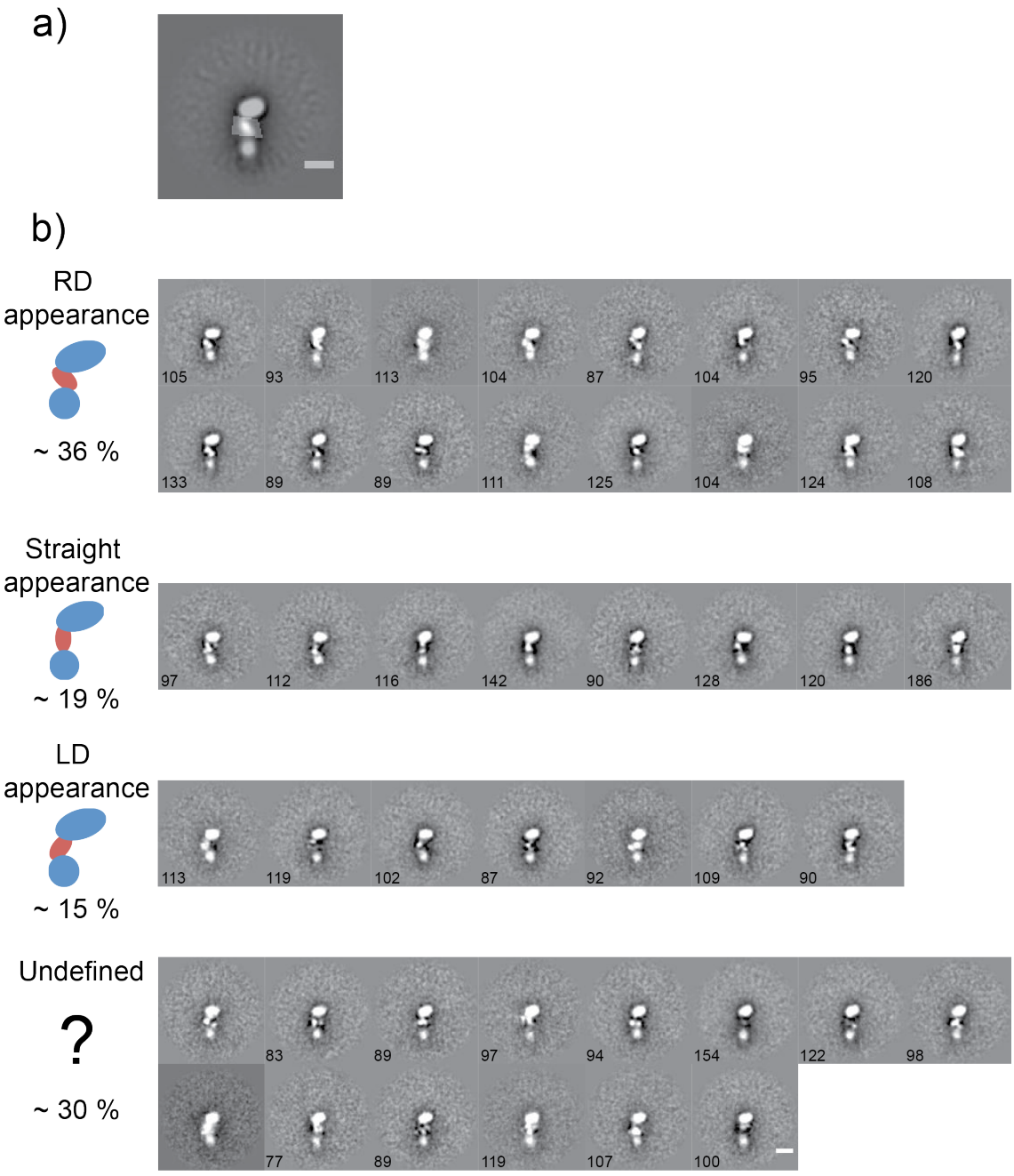


Figure 36 - IQ1+CaM principle views.

(a) the IQ1+CaM mask used is depicted superposed on the global average. (b) a montage of class averages showing the predominant appearance of the lever. Cartoons of the principal views seen in the class averages with approximate percentages of individual images exhibiting each appearance are shown to the left. Undefined images are where the CaM appearance was too ill-defined or ambiguous n the average to assign to a view. numbers in the lower left of each image represent the number of particles per class. Scale bar 10nm.

4.3.2 Missing lever and tail density.

Next, I investigated if the remaining part of the lever and tail are unable to be visualised due to averaging. A series of masks was applied to the motor aligned particles, using the global average as a guide, across the area where the lever is expected to lie, in an attempt to identify the remainder of the lever and the tail domains (Figure 37). While extra density was found in the expected region of the lever (Figure 37d-g) using this approach, it was not possible to unambiguously identify this density as belonging to either the lever or downstream tail domains.

Figure 37 - Locating missing tail density.

(a - j) masks superposed on the global average image were applied to the 4721 particles to locate missing lever and/or tail density lost due to averaging. The selection of image averages generated following classification (right) show classes exhibiting extra density. Extra density is typically seen in masks that are in-line with the lever (d - g), with classes showing more distinguishable features highlighted (*). Scale bar 10nm.

The overall density (in terms of particle size) occupied by the particles appears lower than expected, which may be due to a lack of occupancy of the IQ domains by calmodulin but could also be attributed to a lack of density from the tail domains. This lack of density could be accounted for if a part of the molecule is not lying flat on the carbon but is projecting out of plane. To assess this, the area occupied by regulated and extended particles was compared. Only molecules 'face in profile' and 'face away' were compared as to maintain the same angular distribution between the extended and regulated conditions. The total area of each particle was calculated for both extended and regulated particles (Figure 38).

On average, regulated particles occupied a higher average area (216 ± 32.8 nm^2) compared to extended particles (135 ± 30.1 nm^2), an increase of 81 nm^2 (Figure 38c). Myosin 7a is expected to occupy ~ 196.4 nm^2 (for a 3IQ model) or ~ 229.5 nm^2 (for a 5IQ model) for molecules in the regulated state and ~ 262.4 nm^2 (for a 3IQ model) or ~ 287.6 nm^2 (for a 5IQ model) for molecules in the extended state, based on homology models (Figure 38c). Since we would expect all the domains to lie in plane on the carbon substrate, we would expect both extended and regulated particles to occupy about the same area (with some variation based on the orientation of the domains on the carbon substrate). This confirms that not all of the molecule is visible on the carbon substrate, and it is extremely likely that a portion of the tail and/or lever is projecting out of plane, particularly in the extended particles where we see less than half the expected area occupied when compared with extended model projections. It should be noted that the gap in the density seen between lever and tail in the regulated class averages cannot be resolved in all of the single particle data. However, from the class averages we can see this gap in density accounts for ~ 3.5 nm^2. Even with this adjusted density area, we can still see that there is a large difference in the area between the regulated and extended particles.

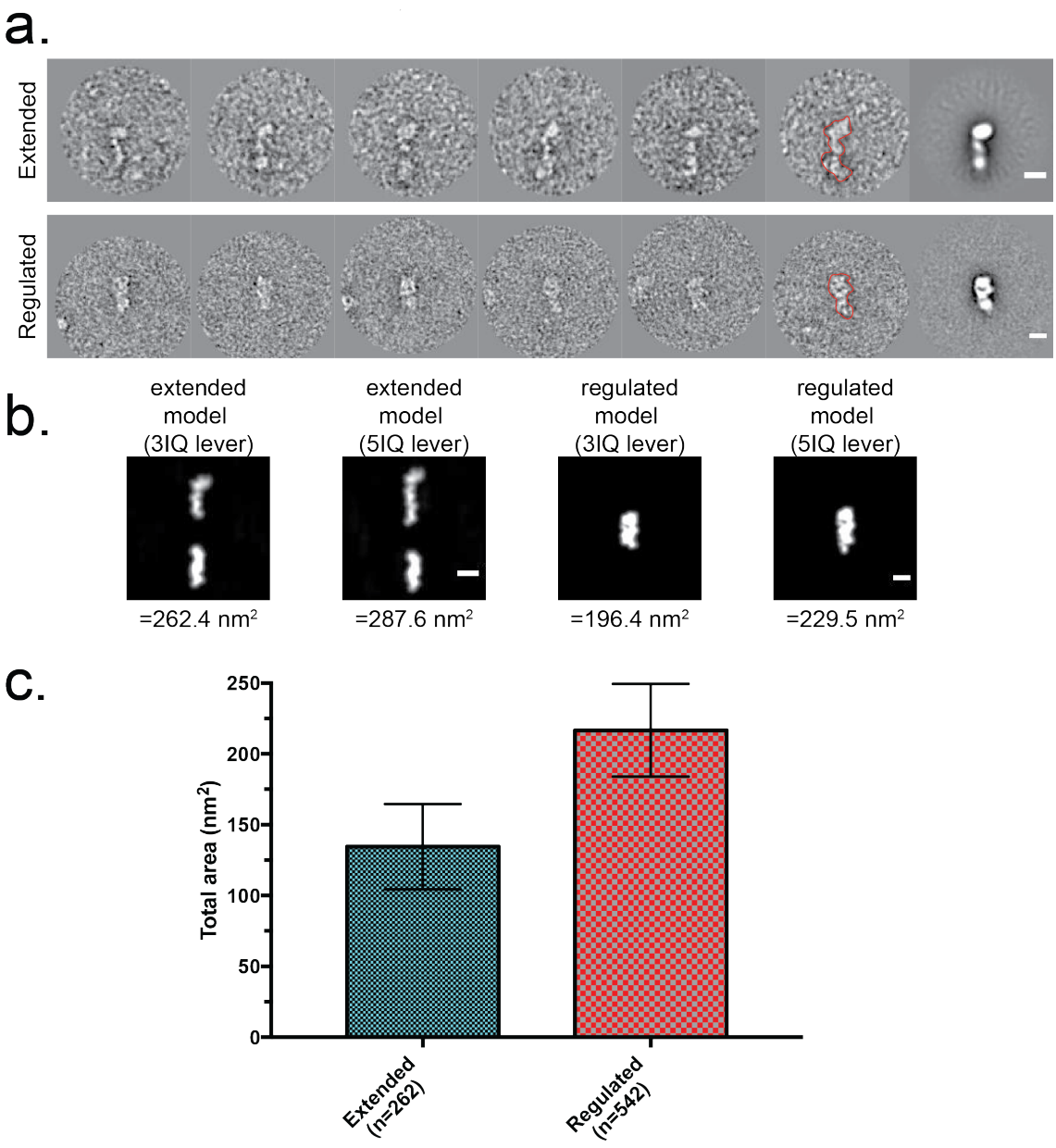


Figure 38 - Density projecting out of plane.

(a) The boundaries of individual particles were manually marked in ImageJ and the total area of each particle was recorded. (b) projections of model myosin 7a structure with either a 3IQ+CaM or 5IQ+CaM lever in both the extended and regulated form. Below each projection shows the total area occupied. (c) Total area of both extended and regulated particles ± S.D. Scale bar for all panels 10nm.

4.3.3 Motor tilting about the motor-lever junction.

One of the more noticeable features of the whole molecule class averages (Figure 32) was the apparent variability in the angle between the motor and lever about the motor-lever junction. Since ATP and no actin was present in the sample, all the myosin should be in the ADP.Pi date and the relative motor rotation is likely driven by thermal energy (fluctuations from Brownian motion), rather than conformational changes in the motor due to the effects of the powerstroke cycle. In this way, the motor-lever junction appears to act as a hinge, with the α-helix connecting the motor and IQ1+calmodulin bending somewhat. The hinge-like movement seen here is reminiscent of muscle myosin 2 S1 where a hinge point is observed between the converter and ELC at IQ1 (Billington et al., 2014).

To explore this further, the lever aligned extended molecules were further classified into 90 separate class averages using a motor domain mask (Figure 39a), with ~50 images in each class, which showed that the lever had a fairly consistent appearance. The first calmodulin was easily resolvable, though naturally not as well delimited as in the global average. These images also demonstrated the variability in the motor-lever angle (Figure 39b). These class averages were then used to analyse the distribution of angles formed between motor long axis and lever were measured in the 90-class classification. *ImageJ* software (v1.43) (NIH, USA) was used to measure the angle between two straight line segments connecting three points positioned to define the motor-lever angle. The 3 points consisted of the most distant resolved point of CaM1, the motor-CaM1 junction and the furthest point on motor circumference (Figure 39c). The angular distribution about the lever with motors tilting to the right, the ‘tilt angle’, was calculated (Figure 40a).

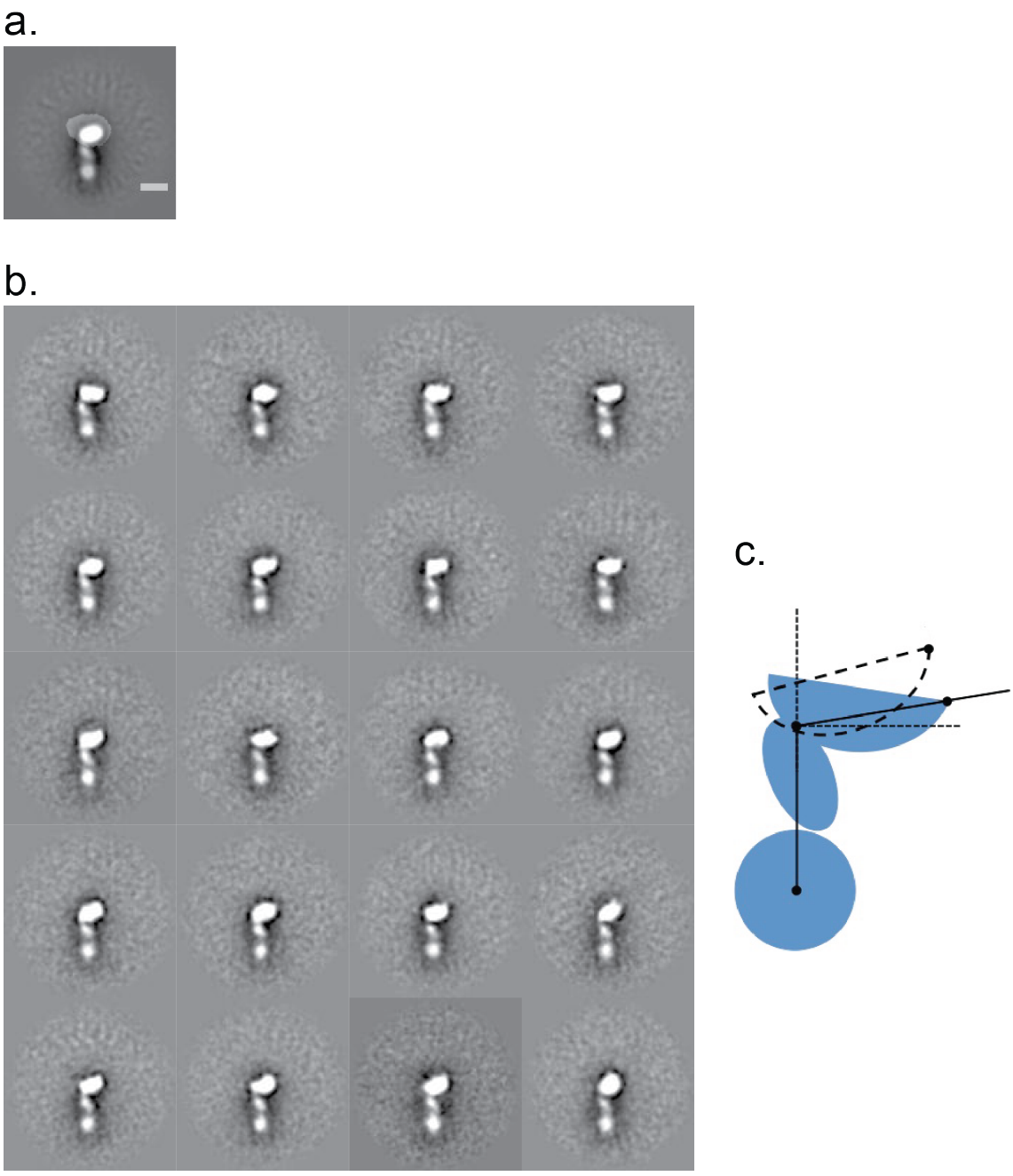


Figure 39 - Motor tilting about the lever junction.

(a) global average with motor-only classification mask superposed (white transparent region). (b) montage selection from 90 class averages from a classification of 4721 full-length extended myosin 7a images using the mask in (a). (c) motor-lever junctional variability identified in class averages about 3 points, with the motor tilting about the motor lever junction. Scale bar 10nm.

The spread of motor tip positions (Figure 40b) is concentrated in an arc that is densest near ~ 120˚. Since nucleotide is present in the sample, the rotation between the motor and the lever is expected to be thermally driven, with the motor-lever junction acting as a sort of hinge. A cumulative frequency plot of the motor-tilt angles shows a broad range of angles, ranging from 90° to 150° with a mean ± SD of 120.7 ± 13.1°. The motors tend to tilt more towards higher angles (> 120˚) than more shallow angles (< 120˚). A reason for this may be in part be due to steric hindrance. Looking at the atomic model and averages presented in Figure 36c, we can see that smaller angles (< 100˚) would bring the upper 50k domain of the motor into contact with the N-lobe of first calmodulin thereby blocking these conformations. This distribution suggests that thermal fluctuation about the mean is present, but with shallower angles being blocked.

For a thermally excited torsion spring, such as the hinge between the motor and the lever, an effective 'spring constant' can be approximated from the distributions through the Equipartition theorem as $k = k_B T/ \sigma^2$, where k is the torsion spring constant, k_B is Boltzmann's constant, T is absolute temperature, and σ^2 variance of the lever angle displacement (rad^2)(Reif and Scott, 1998). Applying the tilt angle data yields a torsion spring constant for the motor-lever joint of 78 $pN \cdot nm/rad^2$. These values are ~3 times larger than the equivalent measurements for apo-myosin 2 S1 fragments from rabbit, scallop and chicken (Billington et al., 2014), in which motor-lever torsion spring constants of ~23 $pN \cdot nm/rad^2$ on average for all S1 types were observed for motors in the nucleotide-free (apo) state.

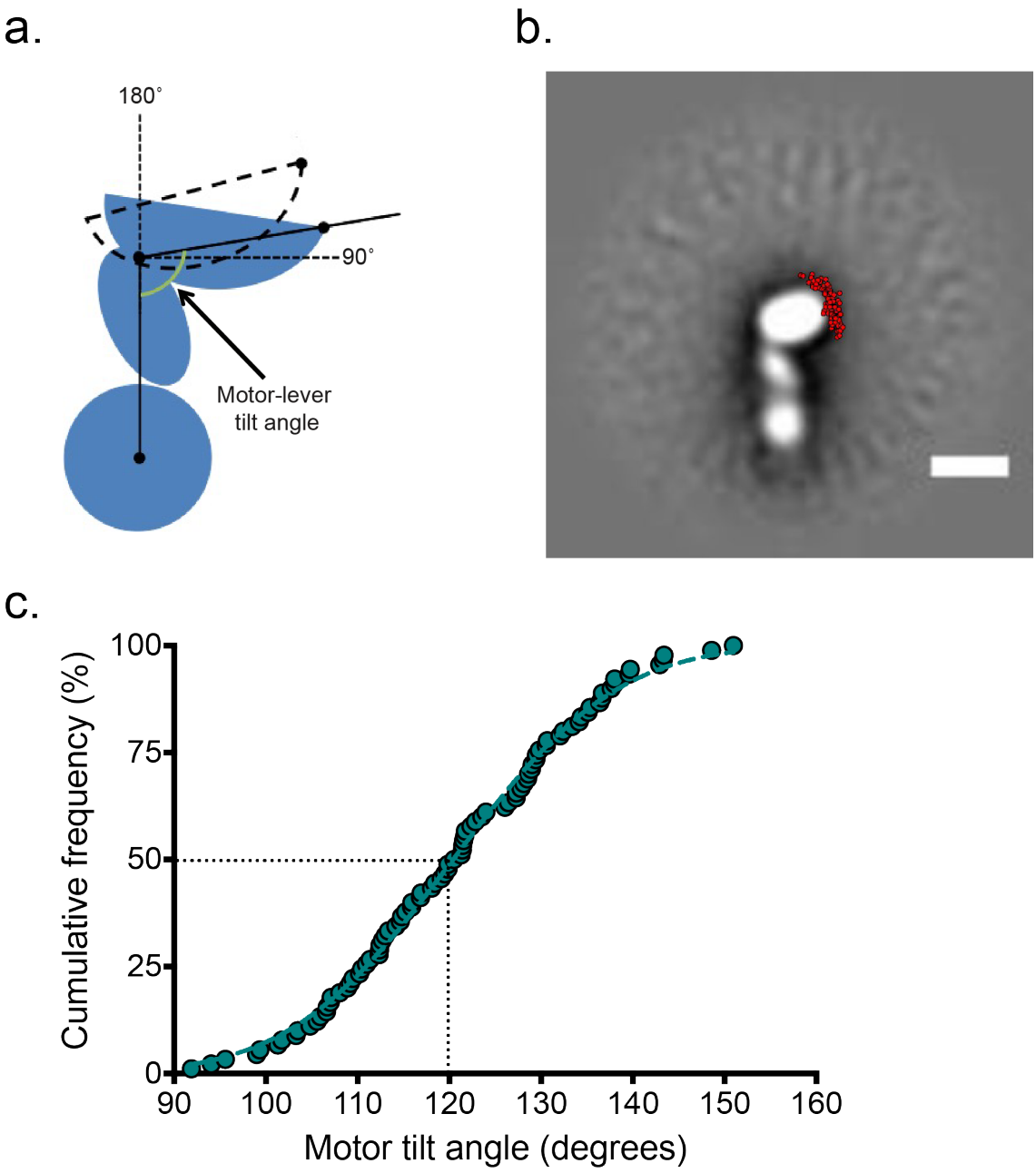


Figure 40 - Motor-lever tilt angle distribution.

(a) cartoon of molecule head showing motor tilt angles measured. Angles measured are between two line intersects at three points. (b) The marked positions of the tips of the motor after motor classification (red dots) superposed on the global average. (c) Cumulative frequency plot of tilt angles. Each point represents the measured tilt angle from each class average. Smooth dashed line is a Gaussian best fit of the data. Scale bar 10 nm.

4.3.4 3D classification of regulated structure.

The 2215 regulated molecules (see section 4.2.2.2) were subjected to 3D classification using a regulated myosin 7a 5IQ homology model filtered to 60 Å as a starting model. This generated a 34 Å resolution reconstruction, which displayed structural features distinct from the low pass filtered starting model (Figure 41a). The angular distribution (Figure 41b) shows that the majority of the myosin 7a particles tended to fall into two orientations, with views from the top and bottom being extremely rare. These orientations are unlikely to be observed upon the grid surface. Several factors limit the resolution of the map including the grain size of the heavy metal salt in negative stain, the overall particle number being relatively small, and the intramolecular flexibility exhibited by myosin 7a.

a.

b.

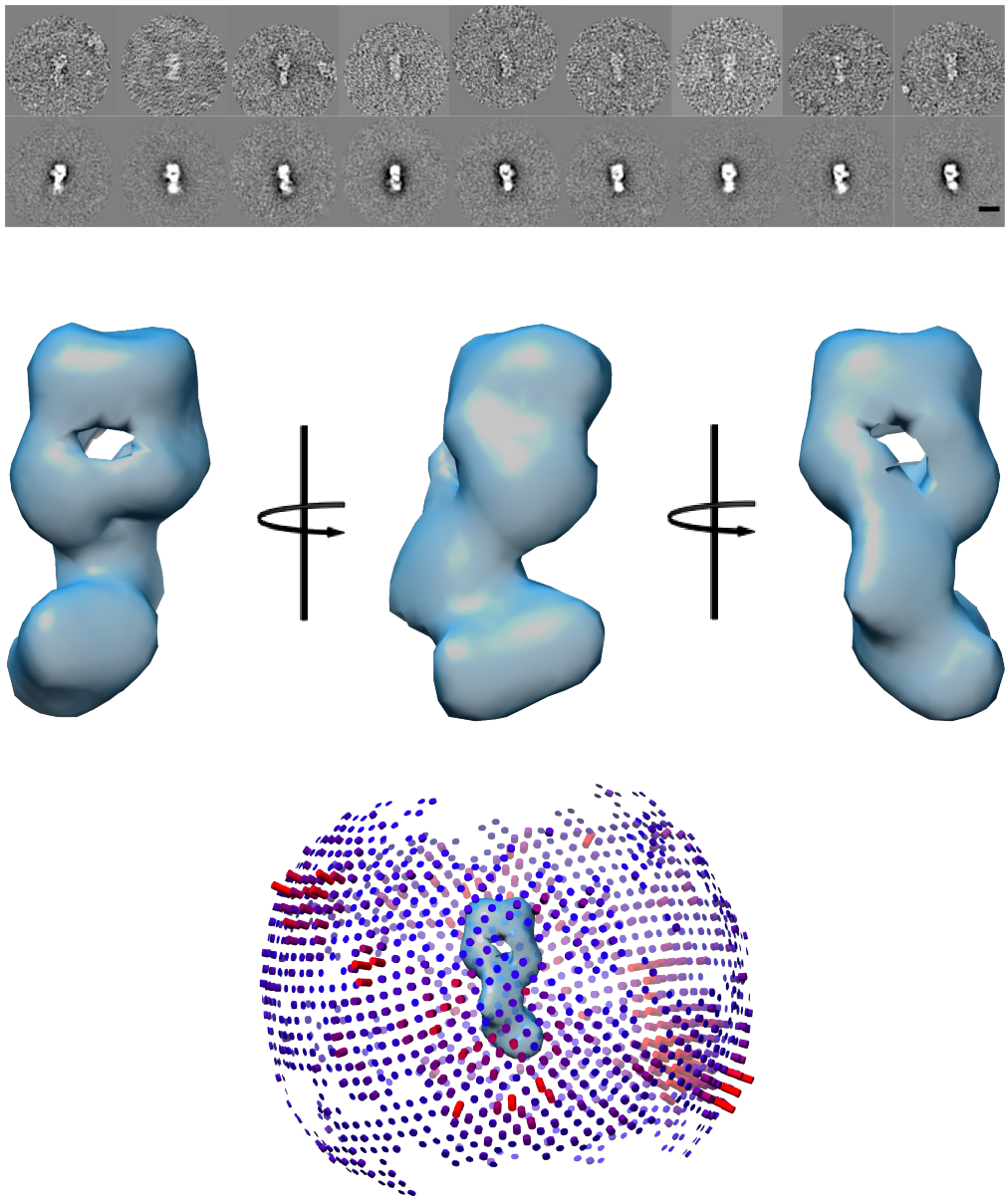


Figure 41 - 3D reconstruction of regulated myosin 7a.

(a) 2215 particles of myosin 7a in the regulated state were used to construct the model. The individual particles (top row) appear in a range of different views as shown by the class averages (bottom row). The 34Å density map generated by relion is shown (b) along with the angular distribution during three-dimensional refinement. The angular distribution represents a histogram of the projection direction (few in blue to many in red) of each boxed segment relative to the three-dimensional reconstruction (light blue). Scale bar 15nm.

To better understand the structural arrangement of the domains in the 3D density maps, various homology models of each domain were generated with ITASSER (Roy et al., 2010) and UCSF Chimera (v1.12) (see section 4.2.4.1), and compared to the 3D density map. Fitting the homology models into the 3D density map using a rough fitting, followed by the UCSF Chimera 'fit in map' function (which uses steepest-descent local optimisation to find the best fit) (Figure 42a,b) shows that the entire 5IQ+CaM lever cannot fit within the density (Figure 42c). Approximately 35 % of the atoms falling outside the map. The 4IQ+CaM lever was also too large to fit into the map, with 27 % of the atoms falling outside the map. The 3IQ+CaM lever on the other hand, was a better fit, with only 17.1 % of the atoms outside the density. It should be noted that the head domains fitted into the density map are rigid body structures, when in reality the structures will have a certain degree of flexibility to allow a better fit.

The M7tail domain (956 – 2215) model was also fitted to the density. The tail domain represents ~45 % of the molecule, and it is not surprising that the M7tail model fills a considerable amount of the density with 88.5% of its atoms inside density (Figure 42d). It should be noted that while crystal structures of both myosin 7 N- and C-terminal myTH4-FERM domains exist, the relative positions of each with respect to each other is not known. A region of additional density exists at the bottom of the map which could accommodate the missing parts of the lever, however, this would require the lever to bend sharply backwards. Taken together, it appears that the 3D model of the compact state can accommodate the tail and Motor + 3IQ-CaM lever (Figure 42e).

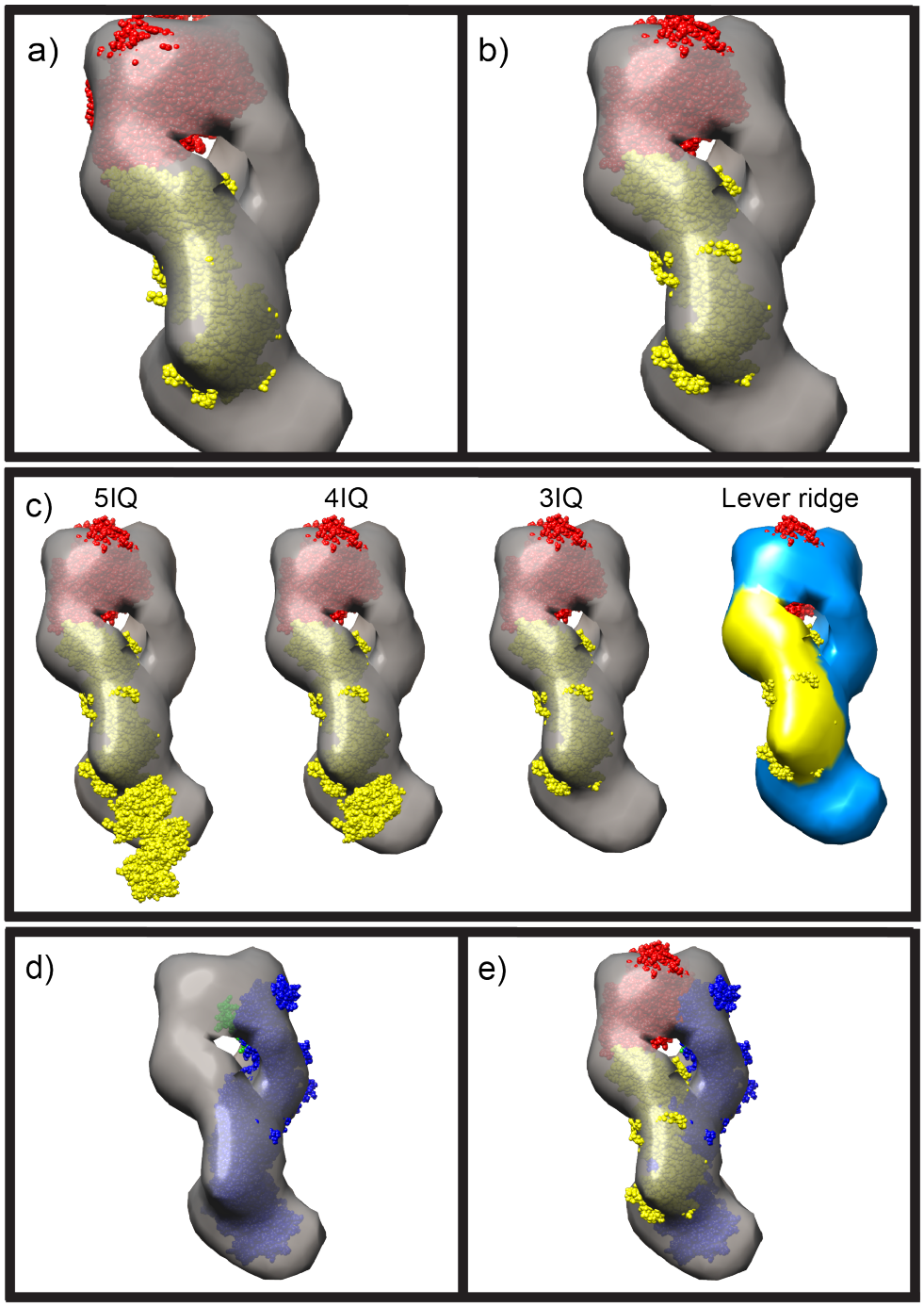


Figure 42 - Fitting domains within the regulated density.
(a) the motor domain structure is positioned, by-hand, in the density map (grey), in the region most likely accommodate the motor domain. (b) the motor domain in the density map following local optimisation fitting using UCSF Chimera. The motor has moved down and rotated a little from (a). (c) fitting different head models into the density map using UCSF Chimera 'fit in map' tool. Only the 3IQ model fits well within the density, with the lever following a 'ridge' in the density (yellow). (d) Fitting of the tail domain and (e) head and tail domains within the density.

4.3.5 E.M. Analysis of isolated lever domains.

The two main issues with examining full-length myosin 7a by electron microscopy is determining which density belongs to which domain (*e.g.* the tail or the lever). Moreover, this is made more complex, by some of the density protruding out of plane from the carbon substrate. To circumvent these issues, I used the GFP-lever constructs of myosin 7a and myosin 5a, termed M7-IQ and M5-IQ, respectively, as described in section 3.2.4. Both of these constructs were expressed and purified from Sf9 cells (Figure 43 and Chapter 3.3.3) and subsequently visualised by negative stain electron microscopy.

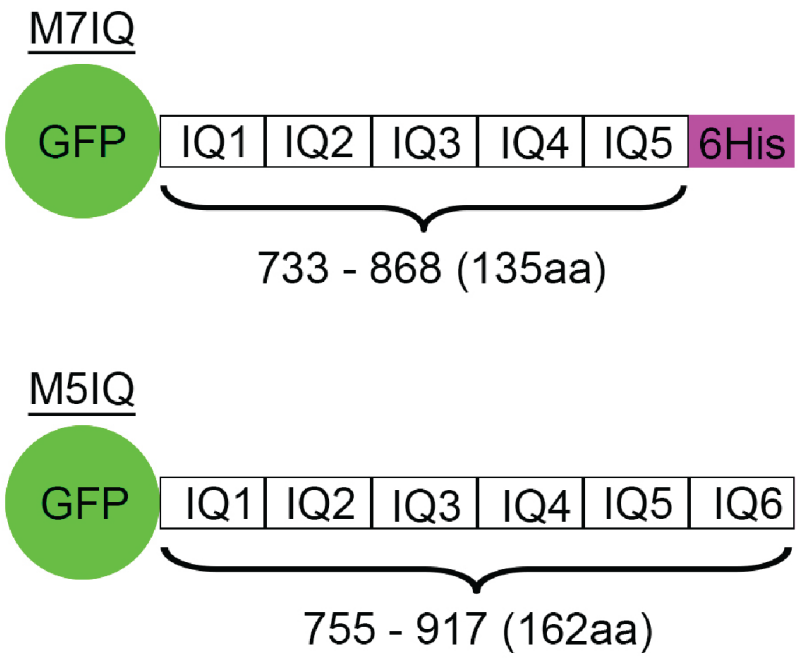


Figure 43 - Myosin 7a and myosin 5a lever constructs.

Constructs containing an N-terminal GFP followed by either 5 or 6 IQ motifs from mouse myosin 7a or myosin 5a, respectively, with the M7-IQ construct containing a C-terminal his-tag, were designed for expression in Sf9 cells.

4.3.6 Characterisation of M5-IQ lever.

EM images of negatively-stained M5-IQ molecules (Figure 44a) shows the appearance of the molecules is consistent with a GFP lever construct, showing an N-terminal GFP 'head' followed by a C-terminal tail consisting of 6IQ motifs + CaM. The appearance of the levers suggests that each of the IQ motifs is occupied by CaM. The levers vary from being relatively straight to containing sharp kinks (Figure 44b). The lever appears to be ~ 19 nm long, consistent with observations of the myosin 5a lever in previous studies (Walker et al., 2000). The inclusion of 2 mM EGTA in the dilution buffer in these experiments should ensure that the levers are fully occupied with calmodulin as discussed in the introduction. Additionally, the size of the GFP 'head', is somewhat variable. Although the reason for this is not clear, it may be due to a slightly deeper stain around some particles. Some molecules appear to lack GFP, which may result from proteolytic cleavage of the GFP from the rest of the molecule (discussed in Chapter 3).

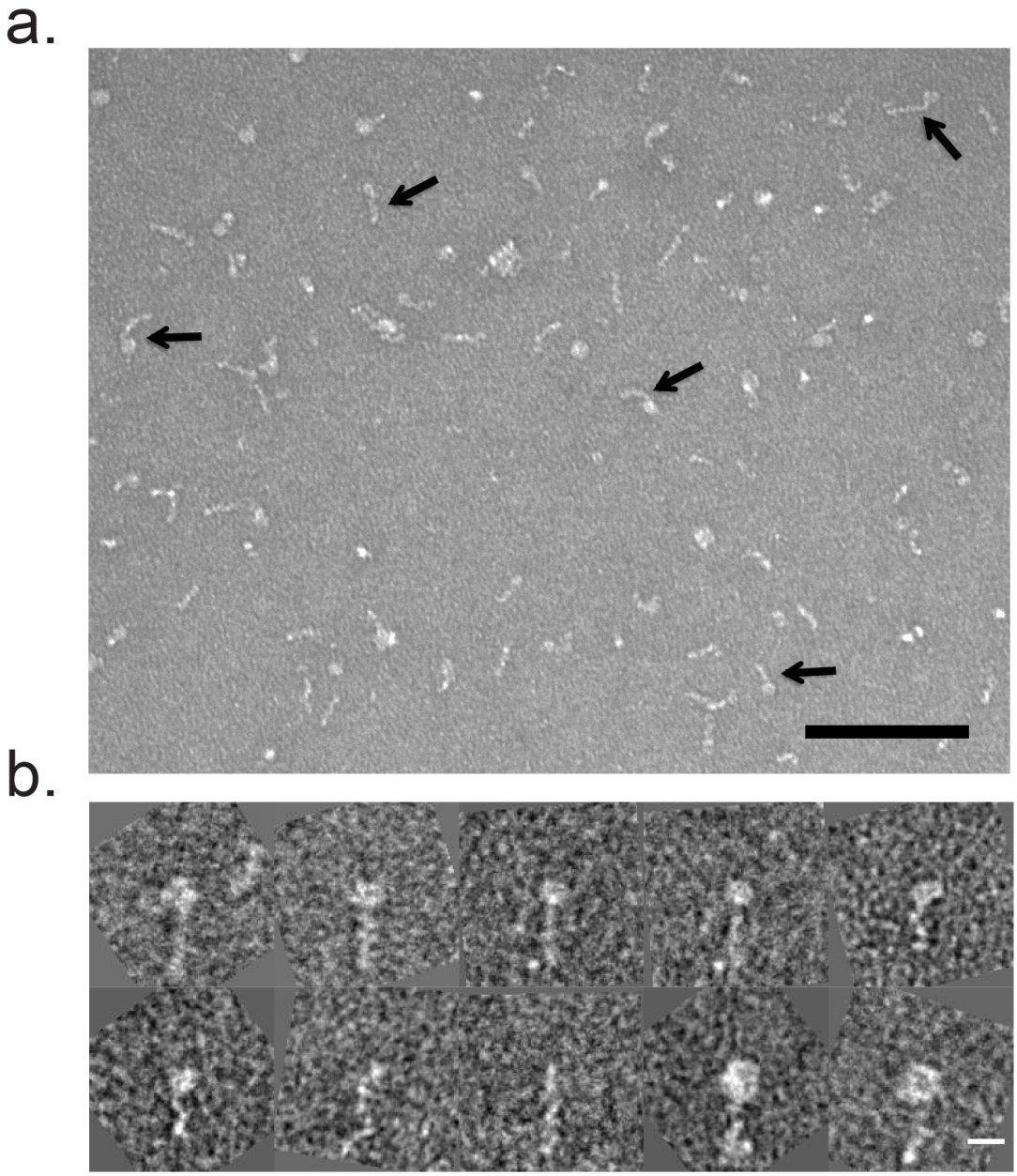


Figure 44 - Field of negatively-stained M5-IQ molecules.

(a) Examples of ‘tadpole’ shaped molecules are highlighted with arrows. The bulbous tips on the ends of the levers is the GFP domain at the N-terminus. Scale bar 100 nm. Grid conditions prior to staining: phosphate-buffered saline (PBS), 2 mM EGTA, pH 7.3. 1% uranyl acetate stain. (b) selection of individual M5IQ molecules. Scale bar 10 nm.

5989 particles were picked, aligned with their levers pointing down and then classified using *IMAGIC* to perform a K-means classification of the lever region appearance. An oval-shaped mask was drawn (using the global average as a guide) (Figure 45) to define the region of interest and encompass the full range of expected lever shapes. Based on trial classifications, the 5989 particles were divided into 200 classes, such that there were ~ 30 images per class – to ensure a sufficient amount of signal to noise and obtain the full spectrum of lever shapes (Figure 45).

In the classification averages, the N-terminal region (GFP) of the construct is easily discernable as the ‘head’. It is important to note that the junction formed between the GFP and lever does not form a continuous unbroken helix as it would in the head of myosin, and significant flexibility exists at the GFP-lever junction. In most cases, averaging is sufficient that an entire lever is seen, with a few exceptions where the lever appears truncated or blurs towards its C-terminus end (caused by heterogeneity in the images averaged together in that class). The details of the individual CaM lobes outlining the lever are not well resolved. This is likely the result of axial rotation (looking down the lever helix), as the molecules probably do not have a preferential orientation on the carbon substrate and thus larger scale features (such as lever shape) are averaged together with smaller features (such as individual CaM orientation. With this in mind, preferences for bending in a certain direction cannot be determined, as the lever has essentially been rotationally averaged (along its length).

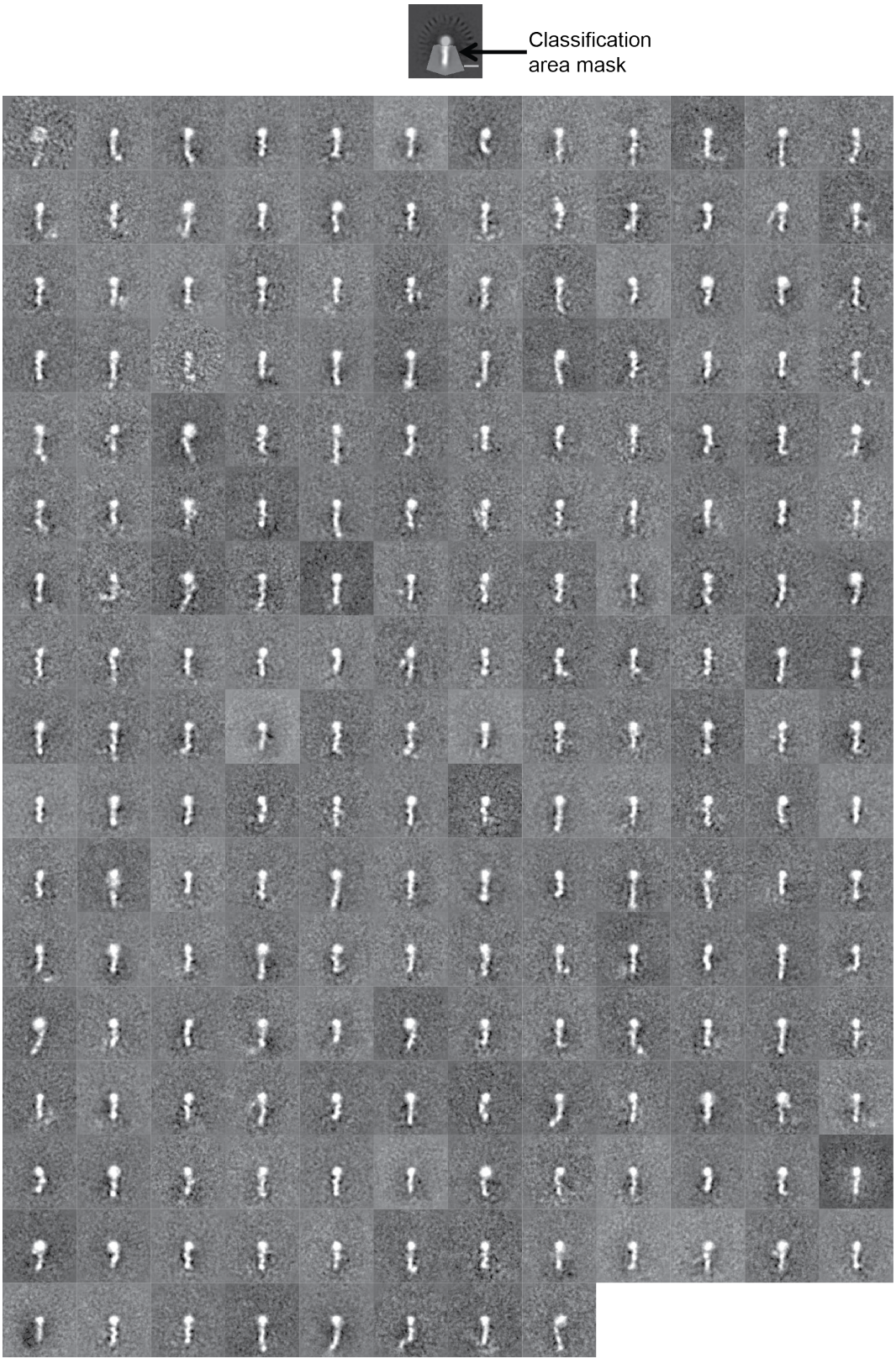


Figure 45 - Lever region classification of M5-IQ molecules.

The global average image (all molecules) is shown above with the classification mask used to focus on the lever region superimposed. The 200 class averages produced by K- means classification by IMAGIC are shown above. Scale bar 10nm.

The wide range of lever conformations observed demonstrates the significant bending flexibility in the lever. The levers may appear roughly straight, smoothly curved to the left and right of the vertical axis, display S-shaped bending or sharply kink (see Figure 46 for examples). This range of curved shapes is reminiscent of modes of oscillation. The degree of curvature appears to vary smoothly across the averages, from straight to a pronounced curve, indicative of a continuum of possible shapes rather than a discrete set of lever conformations. The evenly populated classes are also consistent with a continually varying shape. Pronounced curvature examples are less frequent than the straighter ones. Overall, the distribution of lever conformations suggests the lever is driven by thermal fluctuations about a straight (or near straight) minimum energy conformation.

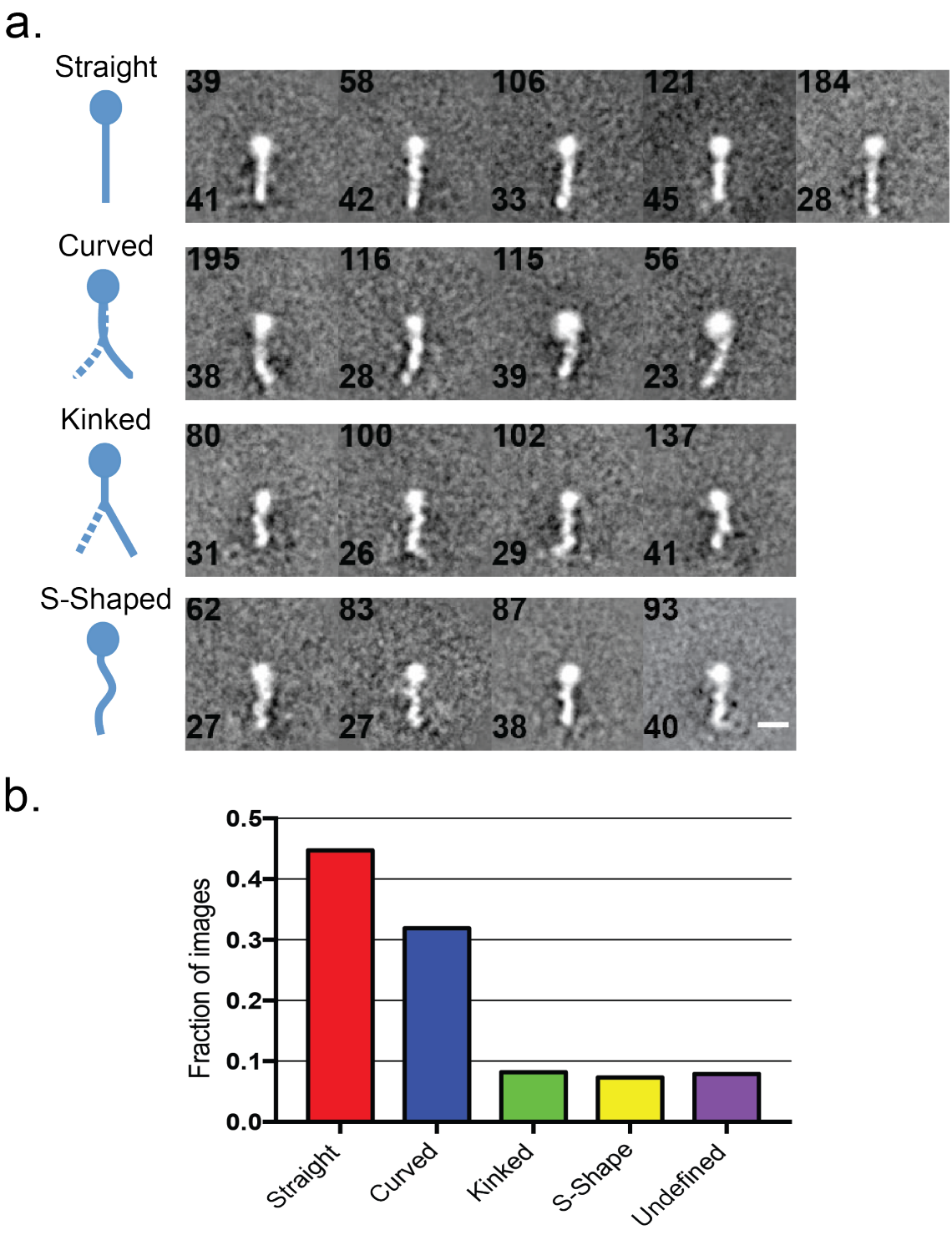


Figure 46 - Identifying predominant lever shapes from classification of M5-IQ.

(a) Selected examples of class averages exhibiting the principal types of lever shape found are shown. Cartoons of each lever shape category are depicted in blue left. Class numbers are indicated in the top left of images, number of images forming the average bottom right. (b) The fraction of images (n=5989) falling into classes of different lever shapes are shown. Scale bar 10nm.

4.3.7 Measuring the curvature and persistence length of the lever.

The shape of the lever was determined from the 174 class averages using fitted splines (Figure 47), allowing the contour length of the lever to be determined. Figure 48 shows the distribution of lengths obtained from the class average data. The data yielded a mean ± S.D of 22.4 ± 2.9 nm, for the length of the lever beyond the GFP. This agrees reasonably well with the value of 21.4 nm predicted from measuring the lever in the atomic model of myosin 5a (Liu et al., 2006) and 21.6 nm expected from a rise per residue of 0.15 nm. The value measured here equates to a length of 3.68 nm per CaM-IQ, which is close to that predicted. It should be noted that the slightly higher than expected value could be in part due to the difficulty in marking the start of the lever at the GFP-lever junction, and the end of the lever, which tends to appear less well defined.

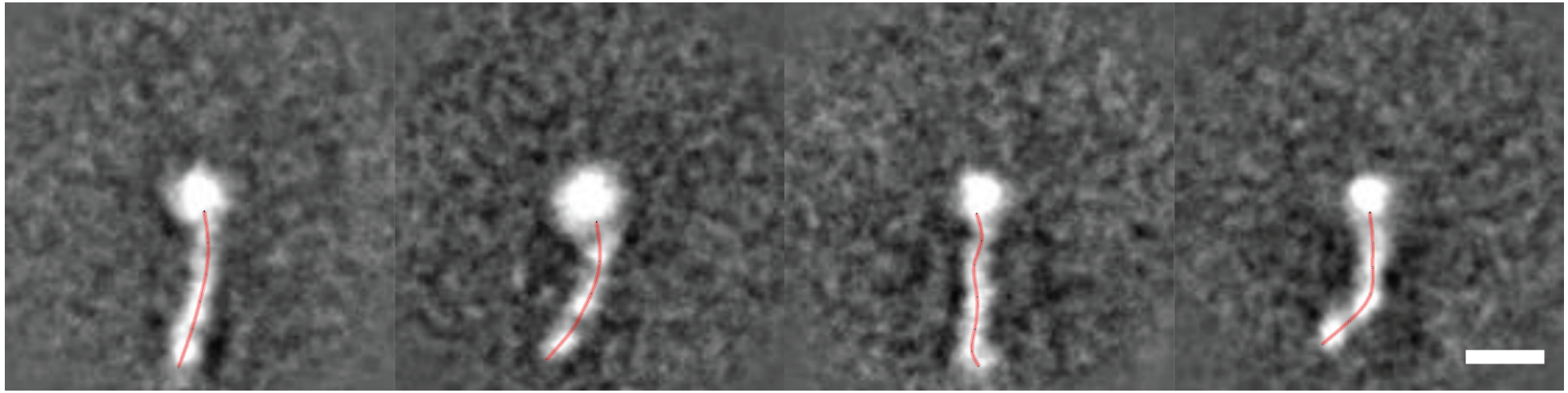


Figure 47 - Example lever shapes marked by hand with fitted spline curves.

Examples of class averages with differently curved lever shapes. Original hand-marked points are shown (black dots) over the spline fit generated by Easyworm1 through these points (red line). Marking the lever is an interpretive process and assigning the end of the lever can be difficult. Scale bar 10 nm.

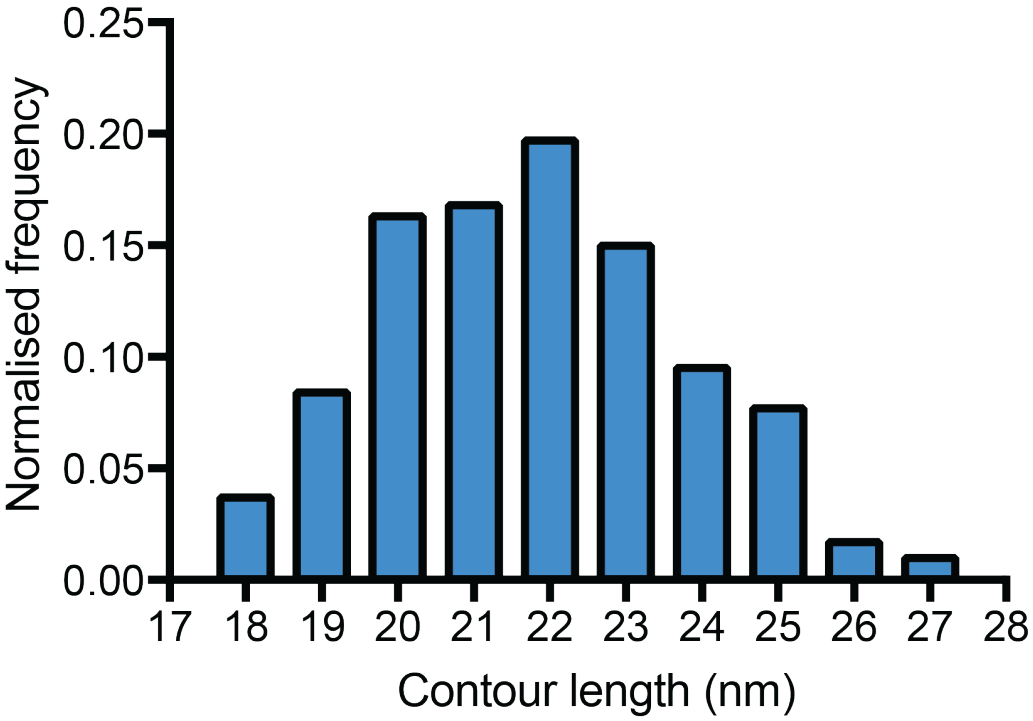


Figure 48 - Histogram of lever lengths.

Bin size 1 nm, n=5444 image measurements. Arc length labels are the mid-point values of each bin. The data has been weighted according to class sizes and the frequencies are normalised (summing to 1).

The persistence length was measured as 28 ± 3 nm or a bending stiffness of 115 ± 12 pN.nm^2 (Figure 49a). A requirement for the use of this statistically-based method is that the measured angle distributions be Gaussian (Frontali et al., 1979). The plot in Figure 49a deviates from the linear relationship expected for a Gaussian distribution in $\boldsymbol{\theta(\ell)}$, an indication that the statistical polymer chain model does not fit the lever shapes particularly well. The negative gradient of the data tends to increase with s (implying greater flexibility/shorter $\boldsymbol{P}$ with increasing $\boldsymbol{\ell}$). As discussed there is some discrepancy in deciding exactly where the lever begins and ends. If we exclude data at the beginning of the implied lever (< 5 nm) and data at the end of the lever (< 22 nm). Considering data only in $\boldsymbol{\ell}$ = 5 – 22 nm data range we get a persistence length of 38 ± 3 nm or a bending stiffness of 156 ± 12 pN.nm^2 (Figure 49b).

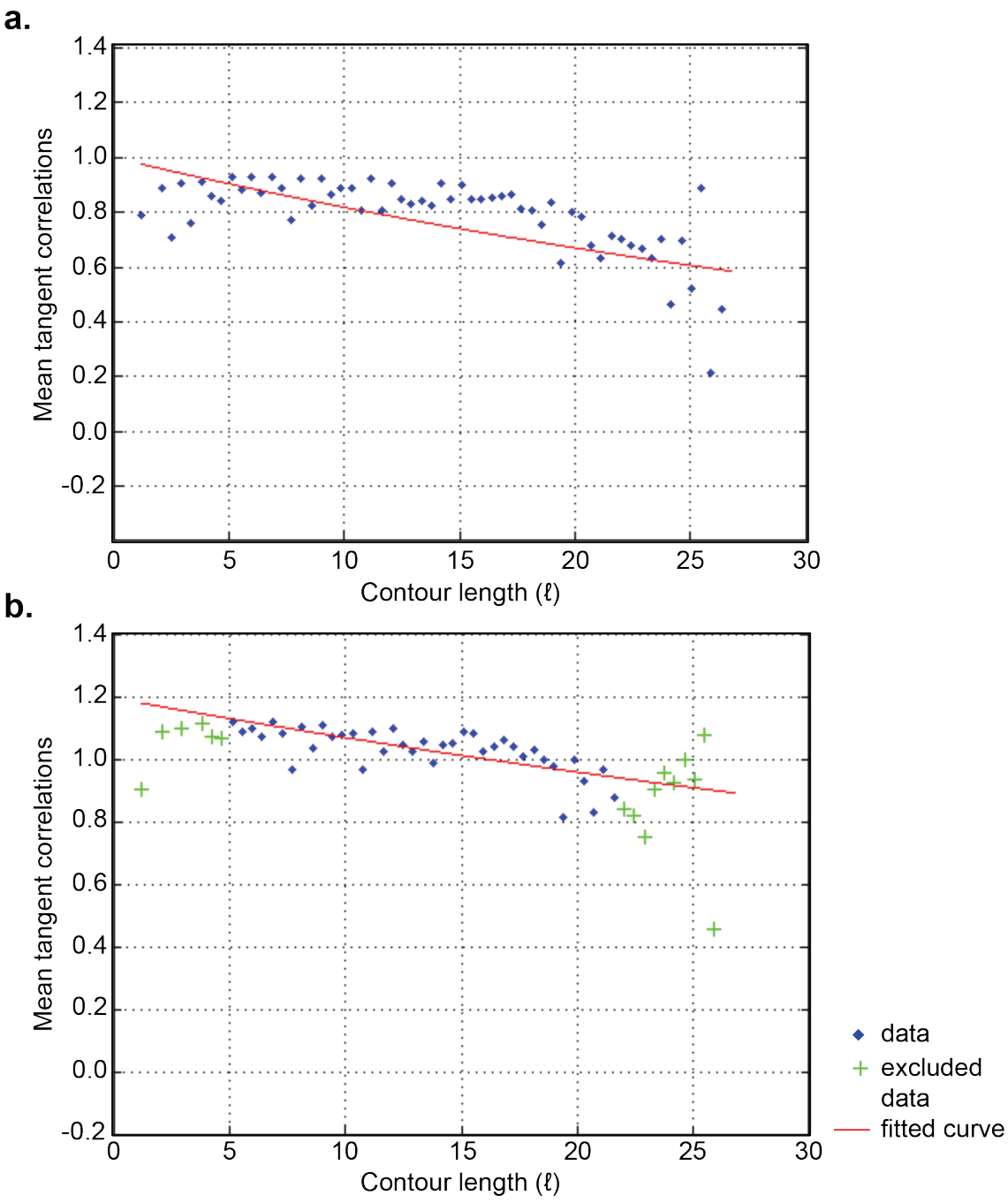


Figure 49 - Tangential angle correlation analysis of lever curve data.

<cosθ> values for data from the 174 class averages of lever curves are plotted. A linear trend would be expected if the angular data for θ(ℓ) was Gaussian. Values appear to deviate from a single linear trend when all data is used in (a). A second linear fit for data (5 - 22 nm) is shown which gives a better linear fit (b).

4.3.8 Characterisation of M7-IQ.

Despite numerous attempts, in which concentrations, buffer conditions and staining methods were varied, EM images of negatively-stained M7-IQ molecules were difficult to obtain. Molecules could only be observed if x10 higher concentrations of proteins were used and an additional wash step using distilled water to wash away excess salt was also required.

Molecules imaged do not contain recognisable features of a GFP-lever construct (*i.e.* a GFP head followed by an IQ+CaM tail) (Figure 50a,b). Instead, only similarly sized blobs can be identified. It is possible that the M7-IQ construct is not particularly stable (discussed at length in chapter 3.3.3) and given that protein will precipitate in low salt conditions (< 500 mM KCl), it is possible that the calmodulin has dissociated from the IQ motifs and collapsed. A number of particles (n=351) were selected and aligned using *IMAGIC5.* The appearance of the class averages and global average reveals no distinctive features (Figure 50c). There is quite a lot of variability in terms of size. The dimensions of the blob are ~ 8.5 nm in diameter giving an area of 57 nm^2. In contrast, the dimension of the global average of the GFP within the M5-IQ construct are ~ 8 nm in diameter giving an area of ~ 50 nm^2. The size of the blob is consistent with the expected dimensions of a GFP molecule, with the variability seen in the global average possibly being attributed to the collapsed IQ motifs on the surface of the GFP molecule.

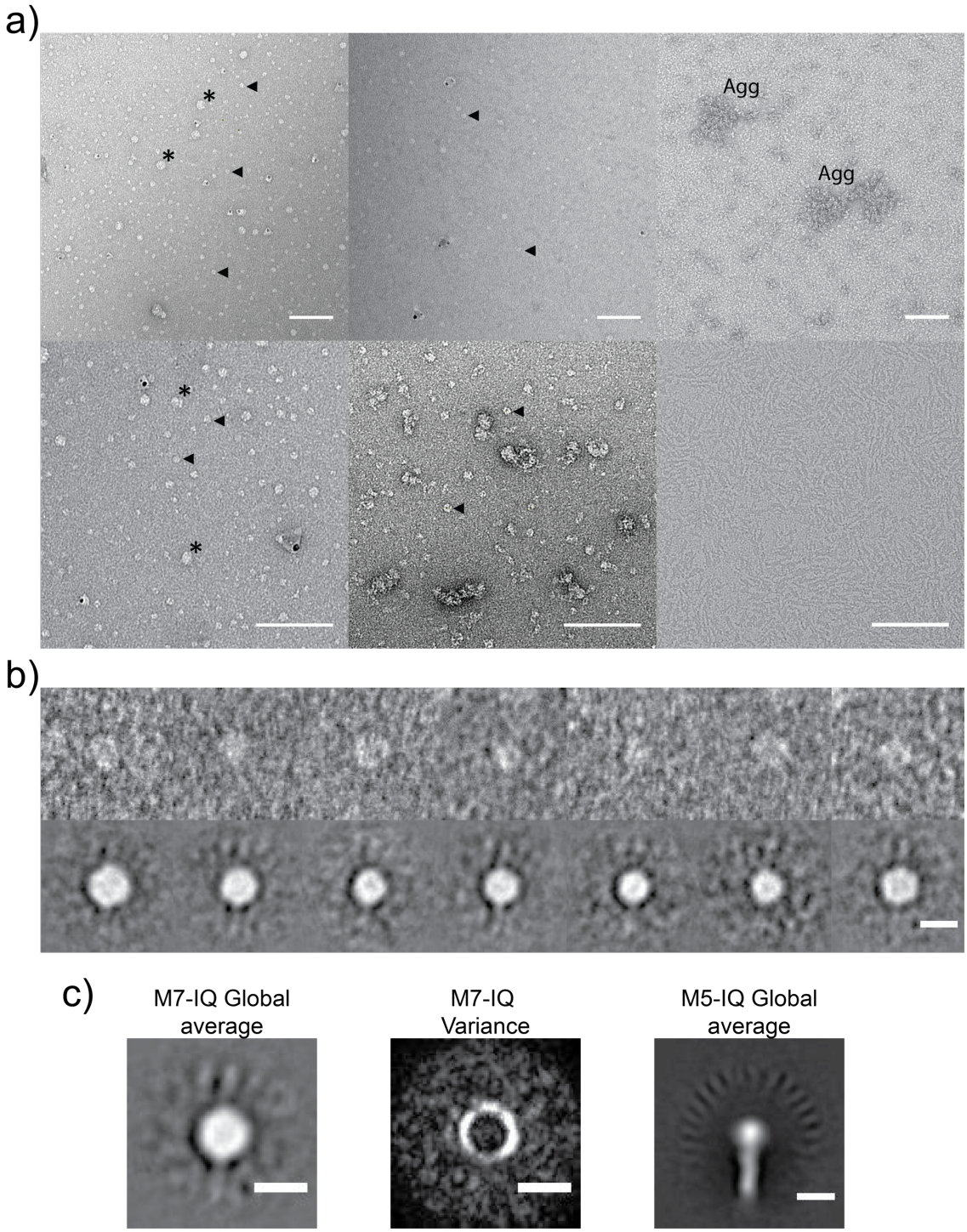


Figure 50 - M7-IQ field of view and classification.

(a) Different concentrations, buffer conditions, and staining techniques were trailed to prepare M7-IQ grids. Recognisable lever constructs could not be identified except for similar sized blobs (arrows) or slightly larger blobs (*) which are potentially unfolded lever constructs. In low salt conditions, M7-IQ is seen to aggregate (Agg). (b) selection of individual blob-like densities (top row) and resulting class averages (bottom row). (c) comparison of the global average and variance produced by aligning selected particles, against the global average of the M5-IQ. The aligned particle shows similar dimensions to the GFP molecule in the M5-IQ construct. Scale bar 100 nm in (a) and 10 nm in (b + c).

4.4 Discussion and future prospects.

The data presented in this chapter has demonstrated that the lever of myosin 7a is shorter than expected both in 2D and in 3D structures, and is unlikely to accommodate an IQ+CaM lever longer than the equivalent of 3IQ motifs. The data additionally appears to suggest that IQ1 may not be occupied by CaM, as the first CaM domain appears to be rotated by 120˚ compared to the expected orientation. Experiments using isolated levers fused to GFP worked well for myosin 5a, but poorly for myosin 7a, further suggesting that not all of the IQ motifs in myosin 7a are occupied by CaM. In addition, analysis of the motor-lever junction showed some flexibility at this point.

Our observation that the first IQ may not be occupied by CaM is consistent with earlier biochemical studies (Sakai et al., 2015) that demonstrated that IQ1 is not occupied in absence of calcium (EGTA) for human myosin 7a. That study showed that IQ2, but not IQ1 was occupied under these conditions. This leads us to suspect that the IQ + CaM we observed in our EM structures is in fact the second IQ motif. It was suggested in that study that the lack of binding of CaM to IQ1 in the absence of Ca^{2+} allowed the motor-lever to be flexible, enabling the motor to fold up into the regulated conformation. This was observed in myosin 7a from Drosophila melanogaster which is able to form the regulated conformation in the absence of Ca^{2+} (Yang et al., 2009). Comparison of the sequence of the first IQ motifs from these studies reveals the core of the IQ motif is well conserved and we should expect CaM binding to behave similarly between the two species.

VIL **LQ**K VI**R** **G**FK D**R**S NFL (Human)

ILI **LQ**R SI**R** **G**WV Y**R**R RFL (Drosophila)

The lever of myosin 7a is also unusual compared to that of myosin 5a, as each IQ motif only contains 23 residues, whereas that for myosin 5s contains a repeating pattern of 23,25. This is important, because previous modelling studies have shown that the α-helix would need to be nearly straight, in order to avoid steric clashes between two adjacent CaMs where the IQ motifs are only separated by 23 residues (Houdusse, Silver and Cohen, Structure, 1996). Therefore, it might very well be difficult for all 5 IQ motifs in myosin 7a to be occupied by CaM as this would require the lever to remain straight, consistent with our EM images.

A second source of flexibility within the head was characterised at the motor-lever junction, with the motor being observed to rotate by ~60° about the lever. The distribution of angles found suggests thermally driven tilting fluctuations about an equilibrium position, with the motor-lever junction acting something like a torsion spring, with the motor able to tilt like the bar in a torsion balance. By application of the equipartition theorem to the tilt angle distribution, an effective stiffness for the motor-lever joint of 78 $pN \cdot nm \cdot rad^{-2}$ was estimated. This is consistent with values from recent single-molecule fluorescence microscopy assays (Shiroguchi et al., 2011), and indicates that inherent motor-lever hinge-like flexibility is a significant source of compliance in the head, and one that is often overlooked in myosins (Billington et al., 2014), especially considering its importance in the actomyosin crossbridge cycle. If the 1st IQ motif were unoccupied, this would be expected to generate a high amount of variance in the global average around the start of the lever due to Brownian motion about the motor-lever junction. Our data is inconsistent with an unoccupied motif as a consistent angle between motor and lever is seen, ableit with a degree of torsional flexibility. If the first IQ motif were unoccupied, a gaussian distribution about the mean would not be observed. In support of this, myosin 7b artificially dimerised by a leucine zipper, is a processive motor, which moves at 72 nm s^{-1}, with an apparent step size (in TIRFM based motility assays) of 30nm, in the absence of Ca^{2+}. Moreover, it was shown to be a high duty ratio motor (0.88) with a rate limiting ADP release, and fast Pi release (Yang et al., 2005). If the first IQ does not bind calmodulin, it should uncouple the motor from the lever, and this is unlikely given its ability to work as a processive, high duty ratio motor. This tells us that the myosin 7a lever can transmit force and the torsional flexibility about the motor lever junction suggests that the motor and lever are not uncoupled. It is possible that the first IQ motif may be unoccupied but can still form kind of stable structure in the absence of CaM in order to transmit force.

Finally, lever bending analysis (*i.e.* curving of the lever helix) and an estimation of the persistence length was also made. Clearly, bending is another significant type of flexibility in the lever. Analysis of intact M5-IQ lever fragments showed a lever length of 22.4 ± 2.9 nm, consistent with expected results (Oke et al., 2010), and a persistence length of 38 ± 3 nm was obtained. Previous modelling study estimates, which put the persistence length at 120 nm (Lan and Sun, 2005; Skau

et al., 2006) to 370 nm (Vilfan, 2005), are slightly higher to an order of magnitude higher than our estimates.

A reason for this lower estimate could be in part due to the experimental protocol. This concerns the process of attachment of the molecules to the carbon film substrate of the EM grid, a process which, for negative stain EM, is not well understood. The assumption for performing tangent-tangent correlations analysis approach assumes to molecules are able re-equilibrate in 2D during attachment to the substrate, and that interaction with the adsorbing substrate does not alter the local bending rigidity of the molecules (Frontali et al., 1979). Such a process would generally require a weak interaction strength between molecule and substrate, in the range of thermal energy (Mücke et al., 2009). The contrasting scenario is that the molecule-substrate interaction is much greater than thermal energy, and that the molecule binds irreversibly to the surface, beginning at a few contact points from where adsorption is quickly propagated. In this process, the molecule is 'trapped' in a frustrated conformation before 2D equilibration can occur. In 'ideal' trapping, the un-equilibrated conformation on the surface can be modelled as a projection of the molecule's 3D conformation onto the plane of the substrate (Frontali et al., 1979; Mücke et al., 2009). Such 'trapping' and 3D-projection has the effect of condensing the molecule shapes and gives rise to a shorter apparent *Lp* (Mücke et al., 2009) so this process might be an explanation for the low lever *Lp* value obtained here.

Additionally, the inherent uncertainty and interpretative nature of manually tracing the lever curves from the position of the CaMs has been discussed in section 4.3.9. Given the difficulty in precisely determining the GFP-lever junction, which is likely connected by a coil, and the ends of the lever which become 'blurred out' due to averaging make estimation difficult. To mitigate the effects of uneven coordinate spacing and short length-scale roughness, the spline-fitting and coordinate re-sampling process was devised. Nevertheless, this process still requires an interpretative judgement as to best sampling frequency that satisfactorily captures the lever's shape but avoids artefacts.

With these considerations in mind, the persistence length of our myosin 5a lever can be compared to that of other biological macromolecules. It is longer (*i.e.* stiffer) than that of titin (L_p ~ 13.5 nm) (Tskhovrebova and Trinick, 2001) or of a typical stable single α-helix (L_p ~ 15 nm) (Sivaramakrishnan et al., 2009), but

interestingly of similar length (stiffness) to that of DNA (L_p ~ 50 nm) (Frontali et al., 1979). Could the levers of myosin 5a molecule be replaced by equivalent lengths of DNA and the molecule still be mechanically competent and able to move processively on F-actin?

That the M7-IQ lever could not be viewed by EM is disappointing, although this result in itself may shed some light on the occupancy of the number of CaMs on the HC. As we saw with full length extended myosin 7a (section 4.3.1), a fully occupied 5IQ construct could not be visualised but only a single CaM shaped structured following the motor domain. The fact that we see the first CaM in a RD-appearance, as opposed to a LD-appearance as expected for a LC docked onto the first IQ motif of other myosins suggest that the first IQ motif isn't occupied. Yet we see that a consistent angle between the motor and lever (see section 4.3.3), which tells us that the motor and lever are not uncoupled. It could be that the 1st IQ motif is able to dock against the converter domain to form a stable structure that allows this torsional spring behaviour. Future work could involve solving crystal structures of each IQ motif (or pairs of IQ motifs) with bound LC along with the myosin 7a motor + 1st IQ motif + LC to help solve this mystery.

5. Modelling the flexibility of Myosin 7a.

5.1 Introduction.

Bio-molecular simulation is becoming an important tool in understanding biological processes. Together with advances within the simulation software and computing power itself (Bond and Leimkuhler, 2007), bio-molecular simulation is an incredibly useful tool in solving important problems ranging from the study of protein folding mechanics to designing novel types of drugs (Borhani and Shaw, 2012).

In describing the range of different simulation tools available it is important to first consider the different length and time scales present in biology, which range from the macroscale, which encompasses large structures such as bones and whole organs, down to the nanoscale, which encompasses small structures such as individual proteins.

5.1.1 The nanoscale.

The nanoscale refers to the regime in which the discrete nature of atomic structures becomes apparent in the dynamics of systems. At these length-scales, we cannot neglect the stochastic motion caused by thermal fluctuations as it is of the same order of magnitude as the size of the structures of interest.

Structures generated by techniques such as X-ray crystallography fall within this regime, with the structures being described by a set of atomic positions connected by covalent bonds in 3D space. However, the structures generated by these techniques cannot be considered to be rigid structures but are actually highly dynamic systems whose motions are important for their functional role.

Many bio-simulation packages have been developed to study biological processes using the structural information generated by these techniques. The most well-known of these is molecular dynamics (MD), which is used to simulate molecules at the atomic level. In this technique, forces are assigned to each individual atom, along with accounting for interactions between all relevant pairs of atoms, as described in section 1.18 and Figure 12. These terms are divided into non-bonded interactions (electrostatics and van der Waals) and bonding potentials representing co-valent bonds. This allows a dynamic trajectory of the motion of atoms in the simulated molecule to be calculated by numerically solving Newtons laws of motion for each atom. The result of this collective motion allows

us to make predictions about the small-scale conformation changes and binding events that occur within the molecule.

MD is a time-dependent approach, which simulates the time evolution of the molecular system to provide a trajectory. The evolution of the molecular system is given as a series of snapshots taken at close time intervals, usually on the order of femtoseconds. Typical simulations of small proteins including surrounding solvent cover the range of several hundred nanoseconds, i.e., they incorporate millions of elementary time steps. The deterministic capabilities of MD make it a very attractive simulation technique.

The initial set of atomic positions (i.e., geometry) is usually obtained from sources such as X-ray crystallography or NMR data. Alternatively, large online structure repositories such as the Protein Data Bank (PDB) are available which contain the structures of many molecules in the accessible .pdb format.

Once the starting structure has been obtained and prepared, the general workflow is to “minimise” and “equilibrate” the system prior to the “production” run. To produce a trajectory of n time steps each of length Δt:

1. Calculate forces on each of the atoms using an appropriate form for the potential energy,

$$\overrightarrow{F_i} = -\nabla V_i$$

2. Use Newton’s Second Law to obtain equations of motion for each atom, i,

$$\frac{d^2\overrightarrow{r_i}}{dt^2} = \frac{\vec{F}_i}{m_i}$$

3. Numerically integrate over Δt to update positions and velocities of atoms,

$$\overrightarrow{r_i}(t + \Delta t) = \vec{r_i}(t) + \overrightarrow{v_i}(t)\Delta t + \frac{1}{2}\overrightarrow{a_i}(t)\Delta t^2$$

4. Repeat steps 1 – 3 (n) times.

By knowing acceleration, we can calculate the velocity of an atom in the next time step. From the atom’s position, velocity, and acceleration at any moment in time,

we can calculate atom positions and velocities at the next time step. Integrating these infinitesimal steps yields the trajectory of the system for any desired time range.

Another important consideration when performing an MD simulation are external forces applied to the molecule (i.e. the solvent), as many biological molecules have evolved to effectively function in their native environments. This includes how the external aqueous environment is described, be it implicitly, in which the solvent is represented as a continuous medium, or explicitly, in which individual water molecule are represented. Both models have their own advantages and disadvantages, and it is up to the user to decide whether or not explicitly representing the hydrodynamics and specific interaction of the water molecules with the system is necessary.

Although MD is a powerful tool it does have its limitations. In principle, MD could answer biological questions relating to the systems of any size, however, due to the sheer computational cost of simulating large systems this is unfeasible. As MD has been shown to be accurate and biologically representative (Piana et al., 2014), advances in both software and hardware continue to be developed to improve the effective performance of MD. Nevertheless, extending MD simulations beyond the nanoscale time and length scale is still beyond the feasibility of local computing power and even generic supercomputing power at present. As studying problems beyond the nanoscale frameset is now becoming a common requirement, coarse-grained approaches are often being turned to as an alternative to MD.

5.1.2 The mesoscale.

Our definition of the nanoscale (< 5 nm on the order of hundreds of nanoseconds) lies at the small-scale extreme of the biological length scale of interest, and it is, to some extent, intuitively clear how we should physically view the systems. At the nanoscale we require discrete atomic description. However, the function of many biological systems occurs at significantly longer length and time-scales, where the overall dynamics of the system tends to rely less on their discrete atomic detail and more on their overall shape. We define this realm as the mesoscale, where molecules on the order of tens of nanometers to a couple microns in size perform processes on the order of microseconds in length.

Furthermore, obtaining a known structure for a system under study is not always possible. This is often the case when studying larger, less ordered structures, which are often harder to crystallise or solve by NMR methods.

We can therefore begin to remove certain degrees of freedom by coarse-graining the structure to group together sets of atoms. In general, coarse-grained (CG) models group heavy atoms together into a single large particle, thus reducing the number of discrete sites. This allows CG simulations to sample phase space at up five to ten faster than a corresponding all atom simulation (Marrink et al., 2004).

The MARTINI model is an example of a CG simulation approach that is similar to MD, but where the fundamental units are groups of atoms (Marrink et al., 2004). On average, every 4 heavy atoms are represented as a single CG particle, with specific charged, polar, non-polar and apolar particle `types', each with additional hydrogen bonding capabilities. The bonded force parameters are determined from the underlying atomic structure, and non-bonded interactions are still treated using a Lennard-Jones potential for Van der Waals and Coulomb electrostatics for charged particles (Marrink and Tieleman, 2013). Once parameterised, the CG simulation progresses in a manner identical to MD. Another example of CG simulation technique is the Gaussian Network model (Rader et al., 2006), which represents biological molecules as an elastic mass-and-spring network whereby the Cα of the backbone are connected by Hookean springs with a defined spring constant.

5.1.3 Fluctuating Finite Element Analysis (FFEA).

FFEA differs from other coarse-grained simulation techniques as it operates in the continuum limit rather than as groups of particles. The finite element method, or finite element analysis (FEA) is a modelling strategy often utilised at the macroscale, which represents objects as a continuum. At this length scale, the discrete atomic detail has very little effect on the overall structural properties of the object. Instead, the material properties and shape of the object can be described as smooth and continuous functions that describe the macroscopic variables of the material. By describing an object as a continuum, the overall number of degrees of freedom is reduced and hence the computation time required to calculate a solution is similarly reduced. To solve via the finite element method, it is first necessary to replace the individual complex shape with an

equivalent network of simple elements (referred to as the finite element mesh), for which numerical solutions to the partial differential equations (PDEs) for each element can be obtained.

Fluctuating finite element analysis (FFEA) extends the principles of the general FEA technique to include the effects of thermal fluctuations (Oliver et al., 2012). Objects at the mesoscale are sufficiently small that the energy provided by thermal fluctuations drives the molecule to continuously explore conformational space (Hanson et al., 2015). FFEA represents biomolecules as soft, inhomogeneous, continuum materials that are subject to deformation from both thermal fluctuations and from interactions with other biomolecules. This continuum model is described by continuous material properties, such as its density, Young's modulus and viscosity of the protein and its surrounding environment (Oliver et al., 2012). These parameters should be sufficient to describe the biomolecules at the continuum level. FFEA can be generalised to include any intermolecular forces relevant to bio-molecular dynamics, such as van der Waals and electrostatic interactions, provided the necessary force-field parameters can be obtained (Hanson et al., 2015). Once these parameters have been defined, the trajectory describing the changing shape of the protein as a result of the thermal fluctuations can be calculated by iteratively integrating Newton's equations of motion over short time steps and moving each node of the mesh accordingly (Richardson et al., 2014; Gray et al., 2015).

FFEA studies molecules at the mesoscale, and is appropriate for length scales on the order of hundreds of nanometers to micrometers. Since FFEA operates at the continuum limit, there is theoretically no upper length limit. However, an important consideration is determining the lower length scale, below which it becomes important to resolve the atomistic detail. This lower length scale is estimated to be on the order of 5 Å – 10 Å.

An important advantage of FFEA is that it is able to use low-resolution volumetric data arising from experiential techniques such as cryo-EM / cryo-ET or data repositories containing such as the EMDB as inputs for these calculations. This is particularly useful when the discrete atomistic detail of the structure is unknown. In FFEA, a network of volumetric elements represents the complex three-dimensional shape of the protein, with the most convenient element shape being the tetrahedron. The size of these elements need to be large enough to be

approximated as a continua and to avoid introducing instabilities in the FFEA model (which necessitates introducing a small time step) while also being sufficiently small to maintain the level of detail of the shape of the molecule in question. This meshing procedure is non-trivial but software is available to automate this process (Schöberl, 1997).

FFEA calculates trajectories by calculating the evolution of velocity field **u** from the solution of the momentum equation (equivalent to Newtons 2nd law for a continuum) for a continuous material subject to thermal noise (4):

$$\rho \frac{D\mathbf{u}}{Dt} = \nabla \cdot \boldsymbol{\sigma} + \mathbf{f} + \nabla \cdot \boldsymbol{\pi}$$

Here ρ is the density, $\boldsymbol{\sigma}$ is the continuum stress resulting from short-range atomic forces, **f** are additional forces arising from any electrostatic or van der Waals interactions present in the system; and $\boldsymbol{\pi}$ is the stress due to thermal noise. The local deformation of the material is quantified by the strain (the fractional change in shape from a reference conformation). Forces within the material are described by the stress, $\boldsymbol{\sigma}$, which is the force per unit area acting across an embedded surface within the material. The stiffness of the object is characterised by the material modulus, which gives the elastic stress per unit of strain. Typically, the modulus associated with changes in volume (the 'bulk modulus') is significantly larger than the modulus associated with shearing deformations (the 'shear modulus').

Since objects in FFEA are coarse-grained, it is possible to simulate very large protein structures for timescales up to a microsecond. While interactions at the atomic level are essential for bio-molecular specificity, as we approach length scales within the mesoscale, the overall dynamics become less dependent of discrete atomic detail. Objects on the order of length scales between 5nm to 500nm are treated as thermally excited molecules which can deform in response to their environment, which in the crowded environment is a hugely important factor with regards to the potential protein function (Zimmerman and Trach, 1991; McGuffee and Elcock, 2010). This, coupled with the need to simulate bio-molecules over longer length and time scales makes FFEA an interesting simulation tool to study these complex questions.

5.2 Methods.

5.2.1 Designing an all-atom myosin 7a model.

Atomistic models of the myosin 7a molecule are generated in pieces that must be assembled into a complete structure. The only atomic resolution structure available for human myosin 7a is the N-terminal portion of the tail (PDB ID: 3PVL; (Wu et al., 2011)), which includes the first myTH4, FERM and SH3 domain. However, we can use X-ray crystal structures of other myosins with similar amino acid sequences to myosin 7a as templates to model the various domains of myosin 7a. The 5IQ lever region of myosin 7a was constructed using the 2.5-Å resolution structure of apo-calmodulin bound to IQ motifs 1 and 2 of murine myosin 5 (2IX7) (Houdusse et al., 2006). Chimera was used to superimpose the backbone atoms of calmodulin motif 2 on those of calmodulin motif 1 from a second 2IX7 structure, giving a 3IQ model. This process was repeated with a third and a fourth 2IX7 structure to give a 5IQ model with all IQ motifs having 23-residue spacing. Overlapping CaM sequences were deleted, and the backbone IQ helix was joined and replaced with the correct residues of myosin 7a heavy chain.

To create the pre-powerstroke head model, I-TASSER (Roy et al., 2010) was used to model the head molecules using the scallop myosin pre-powerstroke structures (PDB:1QVI) as template. The myosin 7a head models were superposed onto their respective scallop myosin models, then the backbone atoms of the first IQ motif of the 5IQ lever model were superposed onto the first IQ motif of the scallop myosin. The scallop myosin models were then deleted and the head and 5IQ lever models connected. The tail domain was modeled using I-TASSER using the atomic-resolution tail structure (PDB ID 3PVL) as a template. PSIPRED was used to model the proximal region of the tail (935 to 1016). Residues 858-935 are speculated to form a single alpha helix (SAH) domain (Knight et al., 2005). Since no structure for the SAH domain exists, a perfectly α-helical conformer was generated. The models generated were joined together by removing overlapping structures, renaming the peptide chain identifiers, renumbering the residues and then writing the PDB files for the complete structure.

Molecular dynamics simulations were performed on both atomistic models to check if they were physically viable. The simulations were used to correct errors

that may have been introduced during modelling (e.g. overlapping atoms and, distorted bond lengths and angles) and check for secondary structure stability.

My own and previous studies appear to suggest that not all of the IQ motifs are occupied by light chain. *Sakai et al., (2015)* is the only study to have investigated which IQ motifs bound which light chains in the presence and absence of calcium (Figure 10). Additional models taking these results into consideration were generated. Models with unoccupied 'folded' IQ motifs were created, while a folded alpha-helix is unlikely to exist, the bending stiffness of these unoccupied IQ motifs will be lower as a result of the cross-sectional area. Additionally, models with 'collapsed' IQ motifs were created, whereby Motor plus 2IQ+CaM / 3IQ+CaM models were connected to the SAH domain as described earlier.

5.2.2 Molecular dynamics simulations.

5.2.2.1 Molecular dynamics packages used.

Implicit molecular dynamics (MD) simulations were run with Amber14 molecular dynamics package using the ff99SB force field (Salomon-Ferrer et al., 2013), while explicit MD simulations were performed with GROMACS (v 5.1.2) also using the Amber ff99SB force field and SPC216 water model (Lindorff-Larsen et al., 2010; Abraham et al., 2015). Before the MD runs can take place, the co-ordinate files also need accompanying parameters to describe the system. The workflow in Figure 51 shows an overview of the different steps taken for MD file preparation and these different steps are described in more detail below.

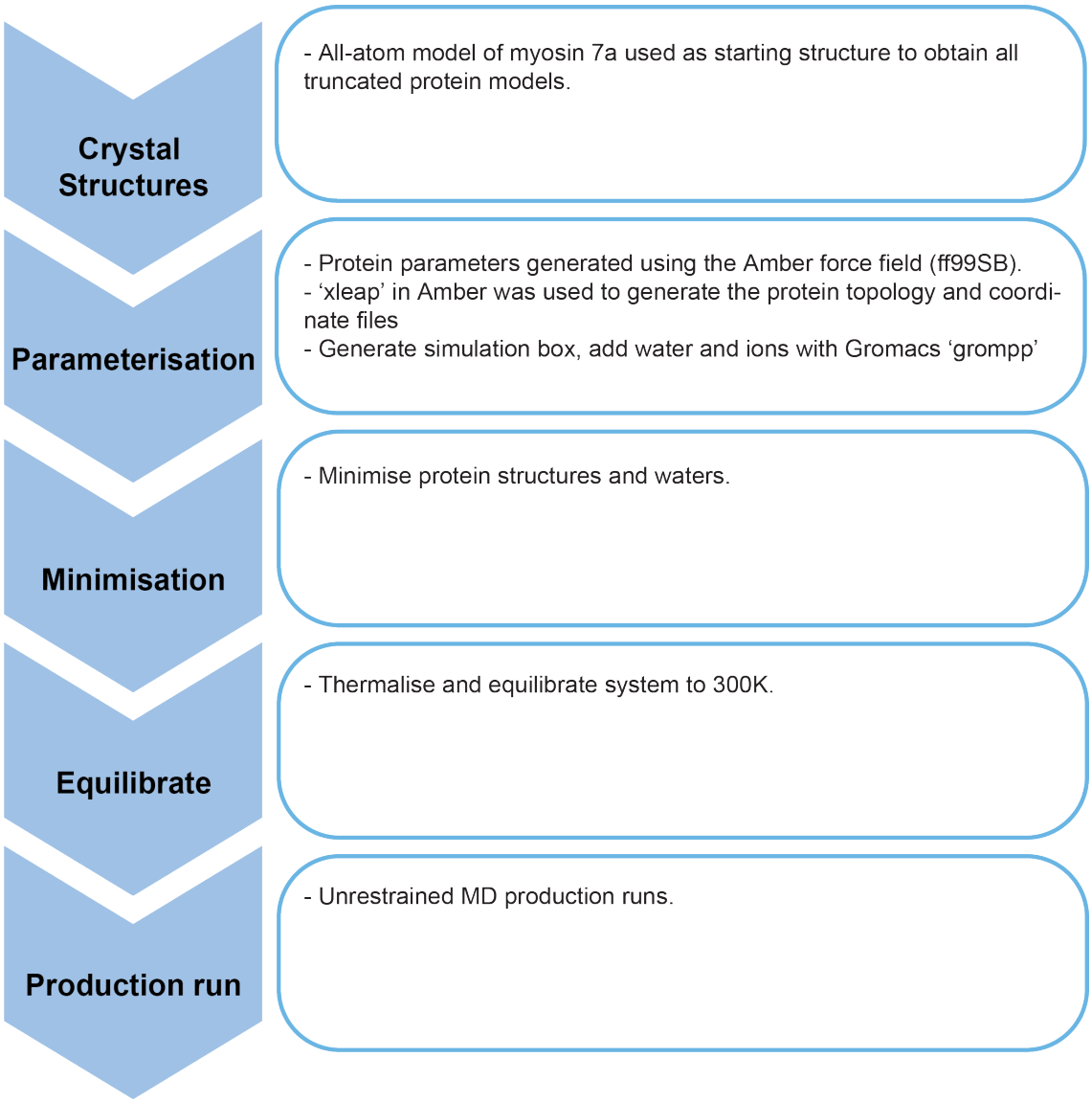


Figure 51 - Workflow for the production of the MD simulation.

The all-atom model of myosin 7a was the starting point for all simulations, with certain domains being extracted from the model. The files were then parameterised, minimised, equilibrated, and then production runs were performed.

5.2.2.2 Computing resources.

The simulations were run either locally or on local supercomputers including ARC1, ARC2, or ARC3 in the Leeds High Performance Computing facility.

5.2.2.3 Preparing protein structure files.

Using the PDB model generated in 5.2.1, the xleap tool in Amber or the pdb2gmx tool in GROMACS, both of which used the ff99SB force field, were used to generate the necessary files for the MD runs. The GROMACS system was then neutralised by addition of sodium counter ions and merged into periodic box surrounded by 39475 water molecules. The system was then minimised using a steepest descent minimization followed by a conjugate gradient algorithm, and then thermalized where the temperature is increased in a step-wise fashion with the constraining forces slowly moved over number of steps [126]. Unrestrained MD simulations with periodic boundary conditions at constant pressure (1 atm) and temperature (300 K). “Shake” was used to constrain all bonds involving hydrogen atoms with time step of 2 fs for integration of Newton’s equations and a 10.0 Å cut off was used for non-bonded interactions. For replica SAH simulations, the same starting structures were used but the values for initial velocity were assigned using the random seed generator in Amber. The input files for the energy minimisation and the equilibration steps as well as the production MD input are shown in Appendix.

5.2.2.4 Simulation data analysis.

After the simulation runs had completed, the raw trajectory files were processed with the Amber CPPTRAJ tool. This was used to strip certain components from the trajectory, such as the solute (water and ions) from the files and re-image the protein (i.e. centre of mass aligned to the centre of the simulation box) to remove the diffusion of the protein around the simulation box and to extract Cα backbone positions for future analysis.

For the simulation data analysis, end-to-end distance and secondary structure were analysed using the Wordom trajectory analysis tool (Seeber et al., 2007) and GROMACS do_dssp tool (Kabsch and Sander, 1983), respectively. Principal component analysis (PCA) was performed using the PCAzip compression and analysis toolkit (Meyer et al., 2006). PCA is used to gain insight into the main modes of motion of simulated systems. The PCAzip code calculates the averaged correlations between atom co-ordinates from the MD trajectory to get

a co-variance matrix that is diagonalised to get a set of eigenvectors with corresponding eigenvalues to get the direction and amplitudes for the main motions in the simulation. These files can also be converted to short animations to visualise the largest motions and extreme structures of the simulated system.

5.2.3 FFEA simulations.

5.2.3.1 Building an FFEA model of myosin 7a.

Using the full-length atomic models of myosin 7a (or the various structural components) generated in section 5.2.1, FFEA models were generated using the following protocol.

First the PDB was converted to an electron density map (EDM) using [`pdbtoemmap`], which overlays a cuboidal voxel mesh over the atomistic structure and creates a sphere around each atom of the given radius, from which a density is assigned to each voxel based on where these spheres overlap. The EDM generated was inspected in Chimera (v1.12) and contoured to give a representation of the outer surface being careful to leave no internal cavities or separated densities. The electron density maps were subsequently processed with [`commands`] found in ffeatools (Solernou et al., 2018). A surface profile of the structure composed of triangulated faces was generated using [`emmaptosurf`], which was further coarsened using [`surftocgsurf`]. The level of coarseness is determined by the user, with the overall value being high enough to describe the overall shape of the structure but low enough to increase computational efficiency. Surface profiles were typically coarsened to values of 8 – 18 Å. Finally, we populated the triangulated surface with tetrahedra using the Tetgen software package (Si, 2015), completing the FFEA modelling procedure. The procedure is illustrated in Figure 52.

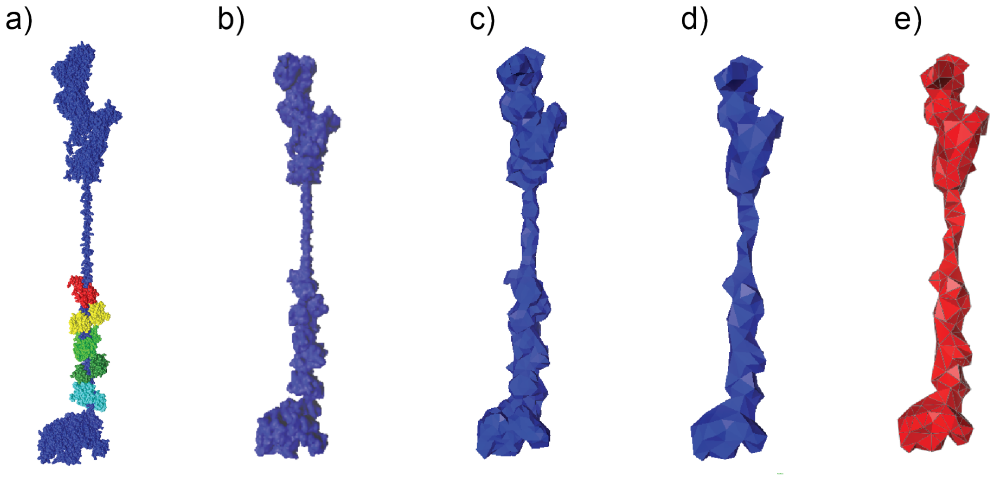


Figure 52 - FFEA conversion procedure for the atomistic full length myosin 7a model.

(a) The atomistic structure of myosin 7a. (b) The electrostatic surface calculated using VMD. (c) first coarsened triangulated surface. (d) Final coarsened surface with a shortest distance of 14Å. (e) The final FFEA model visualised in PyMOL.

5.2.3.2 Parameterisation of FFEA models.

In addition to the overall shape of the molecule the FFEA method also requires material parameters to be assigned. Generalised force-fields cannot necessarily be developed for FFEA models due to the differences in elastic and viscous properties for structures that appear identical at the mesoscale. For example, the coiled-coil structures present in myosin, kinesin and dynein, although structurally similar at the mesoscale, do not have the same flexibilities (Wolgemuth and Sun, 2006). We can however use experimentally obtained values of the bending stiffness of similarly related macromolecules from the literature to parameterise our models.

Alternatively, if an all-atom MD structure is available however, then simulations can be performed on the entire structure, or smaller segments of it, and the flexibility data extracted. These can be converted to FFEA material parameters to reproduce the same flexibility seen in the higher resolution MD, and used in subsequent FFEA simulations to probe emergent properties that are often beyond the capabilities of MD due to time restrictions.

The volumetric maps generated were parameterised with the classical material parameters, such as density, viscosity, and elasticity which describe the continuum properties of the molecule, which are shown in table 14. The external viscosity of the system was set to that of water.

Density	1500 kg m^{-3}
Shear Viscosity	1 mPa.s
Bulk Viscosity	1 mPa.s
External Viscosity	1 mPa.s

Table 14 – Material parameters used in FFEA.

The density of biomolecules is easy to calculate as the masses of atoms are known and there are many complete crystal structures in the PDB (Bernstein et al., 1978). On average the density of biomolecules is around 1500 kg m^{-3} or 1.5 times that of water (Quillin and Matthews, 2000). The internal viscosity of a protein can be determined experimentally by at least two methods. These are solvent based techniques (Cellmer et al., 2008) and AFM (Wang and Zocchi, 2011). The solvent based techniques work by measuring the unfolding / refolding

rates for different solvent viscosities and by extrapolation determines the contribution of the internal viscosity. AFM techniques work by measuring the relaxation time of a protein in response to either oscillating the AFM tip across a frequency range (Wang and Zocchi, 2011) or by deforming a protein and then retracting the AFM cantilever while monitoring the deflection of the tip (Radmacher et al., 1994). In our case, we use water viscosity to describe protein viscosity.

The stiffness of a molecule is defined through its elastic moduli. On the assumption that the material properties are homogenous and isotropic, the elastic properties can be described the bulk (*K*) and shear (*G*) moduli. As noted by Oliver et al. (2013) these can alternatively be defined through the Young's modulus (*E*) and Poisson ration (υ). The bulk and shear moduli can then be calculated from the following, where E is the Young's Modulus and υ the Poisson ratio:

$$G = \frac{E}{2(1 + \upsilon)}$$

$$K = \frac{E}{3(1 - 2\upsilon)}$$

Only the Young's modulus for all simulations was varied, with the remaining parameters remaining constant. The value of the Poisson ratio was taken to be υ = 0.35 following Oliver et al. (2013).

5.2.3.3 Parameterisation of SAH domain.

Although the bending stiffness of SAH domains has not yet been directly measured, the SAH domain (~ 10 nm) in myosin 6 is predicted to have a persistence length of ~ 12 nm (Spink et al., 2008), from which a bending stiffness (*Bs*) of model SAH domains was calculated as 154 pN.nm^2, in part from MD simulations (Sivaramakrishnan et al., 2008). From the bending stiffness of a SAH domain, we can calculate the Young's modulus using the following relationship from beam theory

$$\boldsymbol{Bs = EI}$$

where *I* is the moment of inertia of the beam cross-section. In the case of a uniform cylindrical rod *I* can be expressed as

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

where *a* is the radius. For a bending stiffness of 154 pN.nm^2, this gives a Young's moduli of 392 MPa. Given the differences in sequences between SAH domains from different organisms and techniques used to estimate the bending stiffness, a range of different Young's moduli (1523 – 150 MPa) were used in the FFEA models.

5.2.3.4 Parameterisation of IQ+CaM lever.

As discussed in Chapters 3 and 4, my own and published results indicate that the 5IQ lever is not fully occupied with calmodulin and only 2 light chains bind the lever in the presence of calcium and three in the absence of calcium, echoing the results of *Sakai et al., (2015)*. As the visualisation and determination of bending stiffness of the M7IQ constructs in chapter 4.3.10 was not possible due to the lever not been able to be identified, the bending stiffness of the M5-IQ construct (156 pN.nm^2) was used instead to parameterise our model. Predictions of the bending stiffness of the lever provide an estimate of the bending stiffness as being 1500 pN.nm^2 (Vilfan, 2005). Three systems were employed (Figure 53) i) where a fully occupied 5IQ+CaM lever was employed ii) where the IQ motifs are unoccupied according to the *Sakai et al., (2015)* study, but the underlying IQ motif is still structured iii) the underlying IQ motif has collapsed.

While the 5IQ+CaM lever structure is not a uniformly thick cylinder (Figure 54b), we took the average diameter of the structure to be 3nm. Using the radius of the 5IQ+CaM lever and the bending stiffness's described earlier, we estimate the Young's moduli as being 129 MPa.

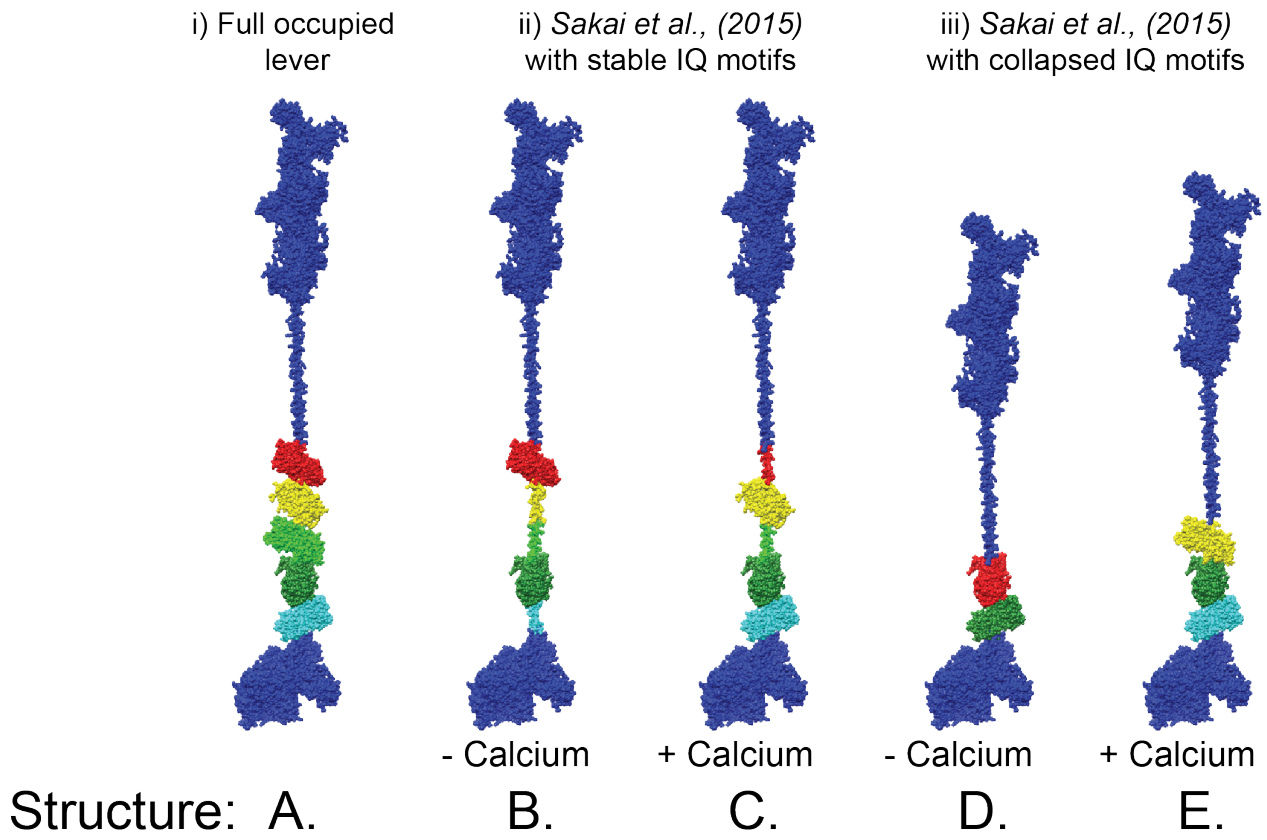


Figure 53 - FFEA lever models.

The following 5 levels models (A - E) were used during FFEA simulations i) a fully occupied 5IQ+CaM lever. Four models in which the 5IQ lever with bound CaM according to the results of Sakai et al., (2015) in the presence and absence of calcium in which the underlying IQ motifs ii) remain stable or iii) have collapsed.

5.2.3.5 Converting FFEA trajectories into pseudo-atomistic trajectories.

In order to compare FFEA trajectories with corresponding atomistic trajectories, we used the [`atomicMapper`] tool (within ffeatools) (Solernou et al., 2018), which converts the continuum FFEA trajectory into an estimated trajectory for the corresponding atomistic structure, using the atomic structural information used to generate the initial structure. Geometrically, a biomolecule within FFEA is simply a set of nodes connected topologically as a set of tessellating tetrahedra. During the simulation the coordinates of the nodes are continuously changing, while its topology remains constant. This enables us to relate the positions of the atoms comprising the biomolecule to the coordinates of the nodes via a linear interpolation scheme (Solernou et al., 2018).

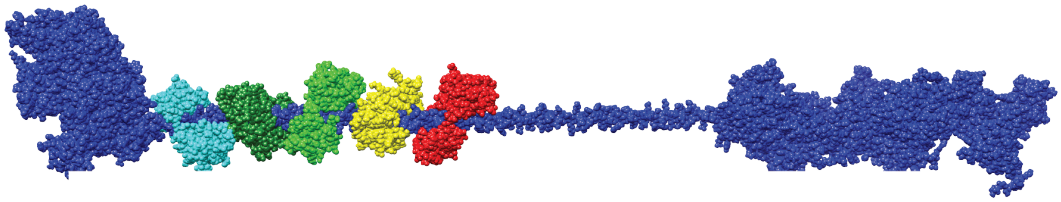
5.3 Results.

5.3.1 Building an atomic model.

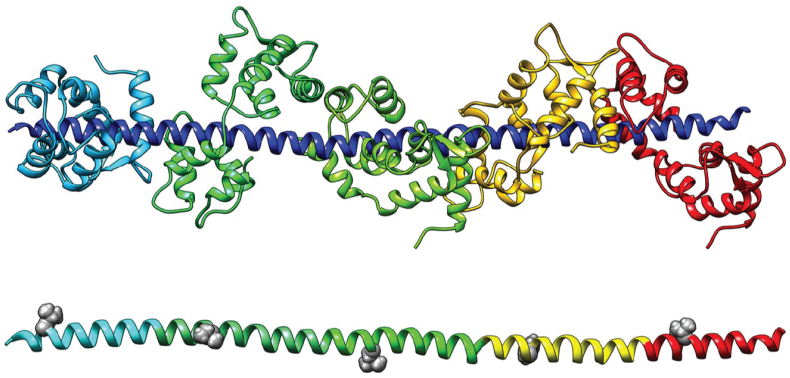
A model of myosin 7a with 5 bound calmodulin in the pre-powerstroke state was generated from homology models and crystal structures (Figure 54a). The molecule in the extended state is 53.4nm long. The molecule contained a total of 2940 residues, which includes the myosin 7a heavy chain (2215 resides) and 5 bound calmodulins (145 residues each).

The positioning and orientation of bound calmodulin as a result of iteratively superimposing calmodulin on top of one another created a thick, anisotropic structure along the lever. The relative spacing between adjacent IQ domains was 23 residues in length, with 12 residues separating adjacent domains, as is found in myosin 7a (Figure 54b). As discussed in the introduction, it should also be noted that the 5 IQ domains do not all follow the canonical IQ motif, IQxxxRGxxxR (Figure 54c).

a)



b)



c)

```
Canonical_IQ      XXXXXXIQXXXRGXXXRXXXX 21
745-765|IQ1       TDRVILLQKVIRGFKDRSNFL 21
768-788|IQ2       KNAATLIQRHWRGHNCRKNYG 21
791-811|IQ3       RLGFLRLQALHRSRKLHQQYR 21
814-834|IQ4       RQRIIQFQARCRAYLVRKAFR 21
837-857|IQ5       LWAVLTVQAYARGMIARRLHQ 21
                        .*   *.   :
```

Figure 54 - All-atom myosin 7a model generation.

(a) The full length all-atom myosin 7a model. (b) the heavy chain (dark blue) is bound to 5 calmodulin light chains. The IQ motifs are 23 residues long between the isoleucine (spacefill) of IQ motif. (c) Not all IQ motifs follow the canonical IQ motif sequence (IQXXXRGXXXR).

The SAH domain (858-935) was modelled as a perfect helix by setting internal dihedral angles to $\phi = -47°$ and $\varphi = -57°$. The crystal structure of the SAH domain (862-932) from *Mus musculus* (PDB:5WST) has since been published. The backbone RMSD between the two structures was 1.75 Å, however, it should be noted that within the unit cell 2 structures of the SAH domain exist which each contain different pairs of ER/K interactions between the two structures (Figure 55). This suggests that the structure of the SAH domain is dynamic and that our representation as a perfect helical conformer is justified.

The proximal region of the tail, which connects the SAH domain to the first MyTH4 domain of the tail (935 to 1016) was not resolved in the template crystal structure and poorly modelled by I-TASSER. PSIPRED (Buchan et al., 2013), an online secondary structure prediction tool suggests that this region forms a mostly unstructured coil with small regions of helicity. As such, the atomic coordinates for this missing segment were modelled as an unstructured loop region using the MODELLER tool within chimera (Pettersen et al., 2004).

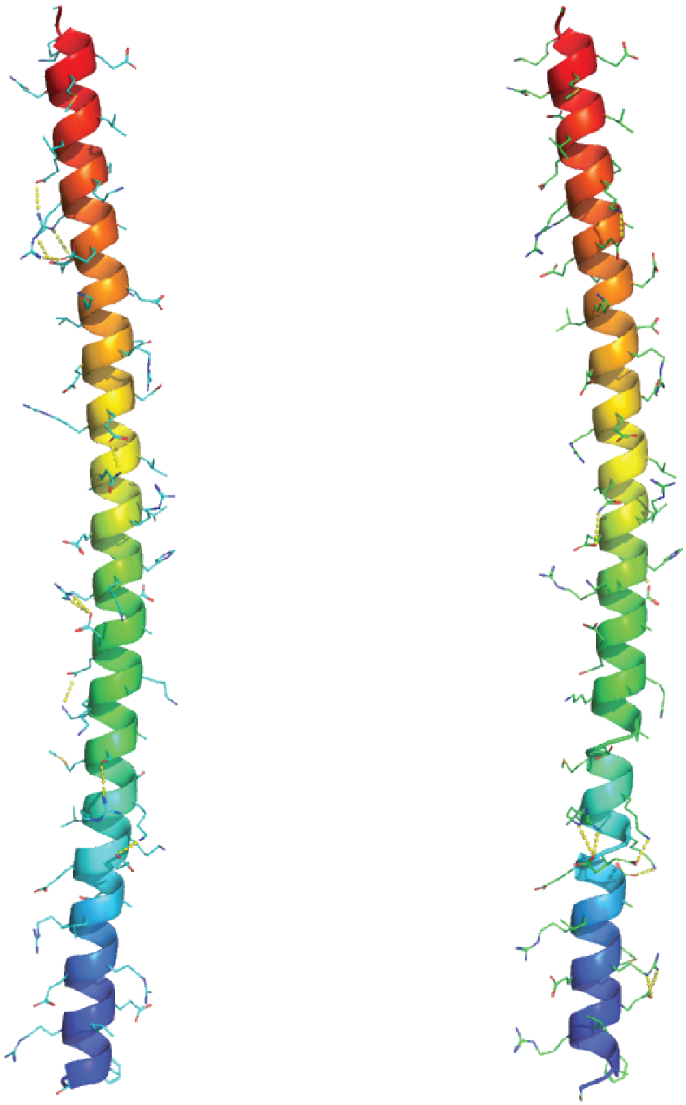


Figure 55 - SAH crystal structure salt bridge interactions.

The two SAH crystal structures found within the unit cell from *Mus musculus* (5WST) showing differing pairs of ionic interactions (yellow dashed lines) between the side chains (stick representation).

5.3.2 Comparison of all-atomistic and FFEA SAH simulations.

The distributions in the changes in length were calculated for both atomistic and FFEA simulations (Figure 56). The FFEA simulations show a well-defined Gaussian distribution which becomes wider as the stiffness is decreased, as expected for an isotropic rod. Comparison of the mean length (mean + S.D) for the 1523 MPa (11.66 ± 0.15 nm) and 150 MPa (11.21 ± 0.64 nm) SAH models reveals the increase in flexibility within the models.

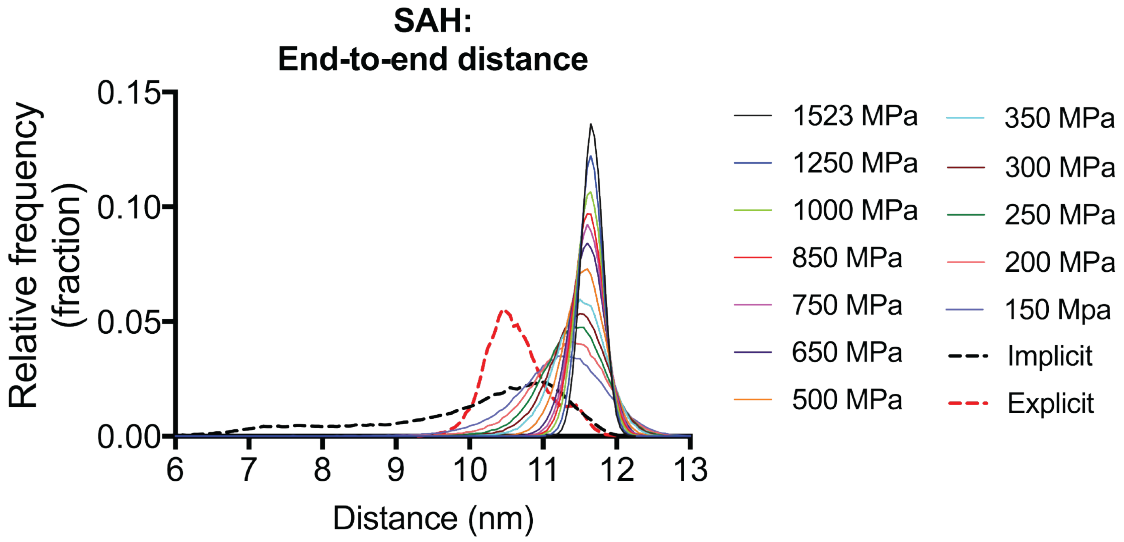


Figure 56 - SAH end-to-end distance.

The end-to-end distance of the SAH domain from both all-atom (implicit and explicit - dashed lines) and FFEA (parameterised with different stiffnesses - dashed lines) simulations. bin size = 0.5Å.

The corresponding measurements for the atomistic models give average end-to-end distances significantly lower than FFEA models with the explicit (10.64 ± 0.42 nm) and implicit (9.97 ± 1.33 nm), with the shape of each distribution also not being symmetric. The primary reason for this discrepancy is that the residues selected for measuring the end-to-end distance tend to unfold and lose their helicity at both the N- and C- termini (Figure 57), with unfolded proteins more likely to adopt a globular form (Haran, 2012). Secondly, it is possible that localised breaks within the alpha helix occur which allows the helix to introduce compliance at specific points. An analysis of the secondary structure per residue (Figure 58) reveals that for the most part, the explicitly simulated SAH domain remains mostly helical during the simulation, with the exception of the N- and C-termini as described in Figure 57. However, the implicitly simulated SAH domain reveals consistent breaks in the alpha helix which tends to appear at R873, E883, and K916 along with other sporadic breaks throughout the simulation (Figure 58). These two factors would account for the smaller mean end-to-end distance observed.

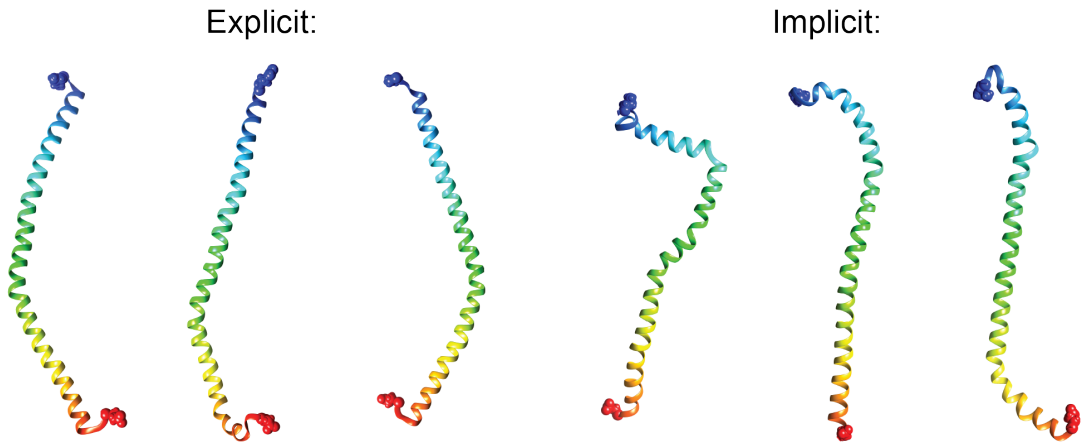


Figure 57 - SAH secondary structure snapshots.

Snapshots of the SAH domain from explicit (left) and implicit (right) simulations. The N- (blue) and C- (red) terminus tends to unravel and not remain helical. Shown in space fill are residues used to measure end-to-end distance.

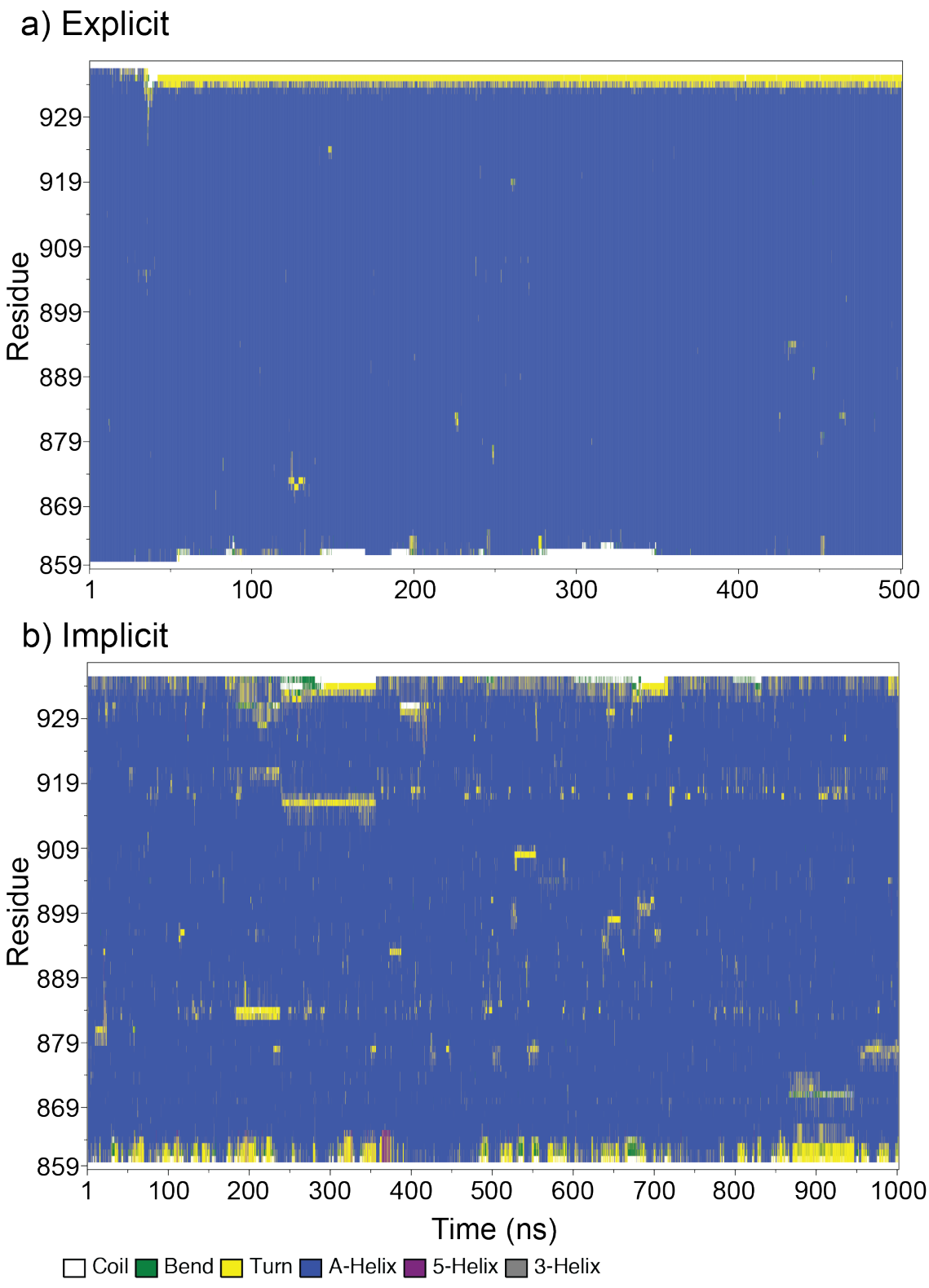


Figure 58 - SAH secondary structure per residue.

The secondary structure of the SAH domain over time in (a) explicit and (b) implicit simulations.

To compare further the dynamics of the SAH domain between atomistic and FFEA simulations, PCA was performed on both sets of the simulation trajectories to determine the major modes of motion. Using the pyPcazip interface, we were able to determine the 5 most flexible modes of motion from the trajectories, the eigenvalues of which are shown in Figure 59. The factored difference in the Young's moduli between the FFEA simulations gives the same factor difference between the eigenvalues, which represent the flexibility of the molecule. Closer inspection of the first 5 Eigenmode animations (Figure 60) obtained from the FFEA simulations and explicit atomistic simulations reveal that the modes are similar between the two types of simulation method, with the first major mode being a bend, which is to be expected for an elongated beam structure. The second and third components represent a compress & twist and stretch mode, respectively, which correspond to an over twist and under twist, respectively, when compared to the atomistic simulations. This becomes far more apparent when the FFEA trajectories are converted to pseudo-atomistic trajectories described later. The fourth and fifth modes represent 2nd order bends which forms an S-shape, with compression and stretch aspect to the motion, similar to modes two and three, respectively.

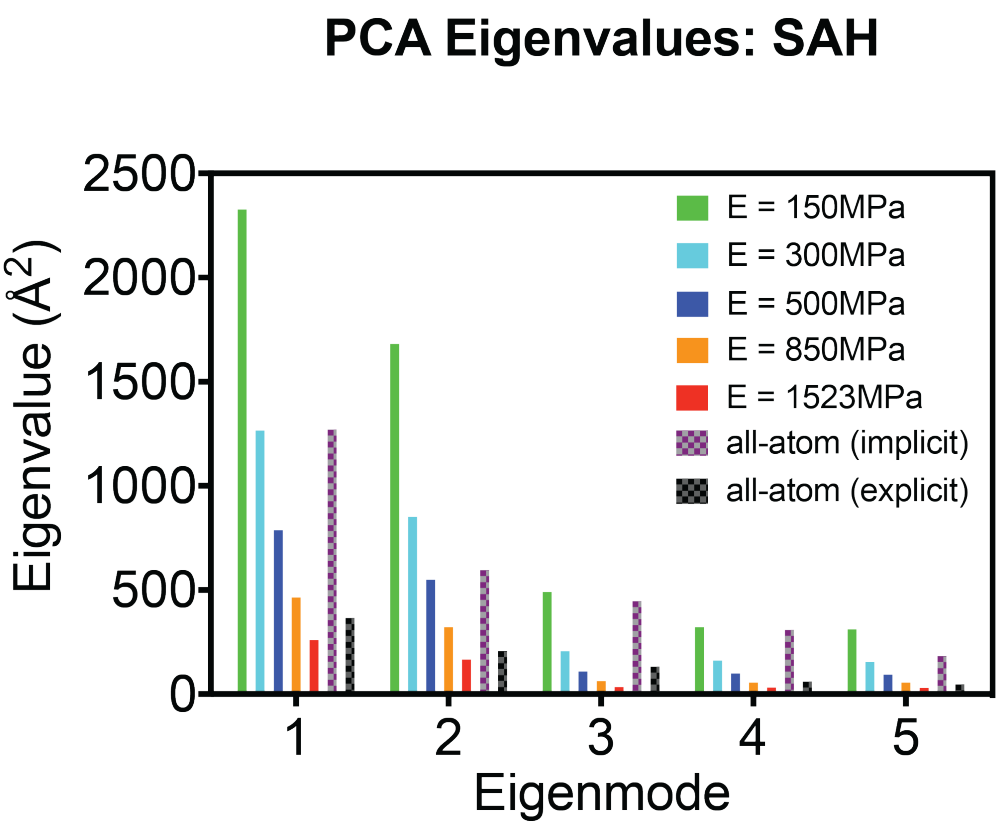


Figure 59 - Top 5 eigenmodes of FFEA and all-atom simulation.

The eigenvalues, which correspond to positional variance, of the 5 most flexible modes found from PCA analysis of the simulations from FFEA (solid colour) and all-atom simulations (checked colour) of the SAH domain.

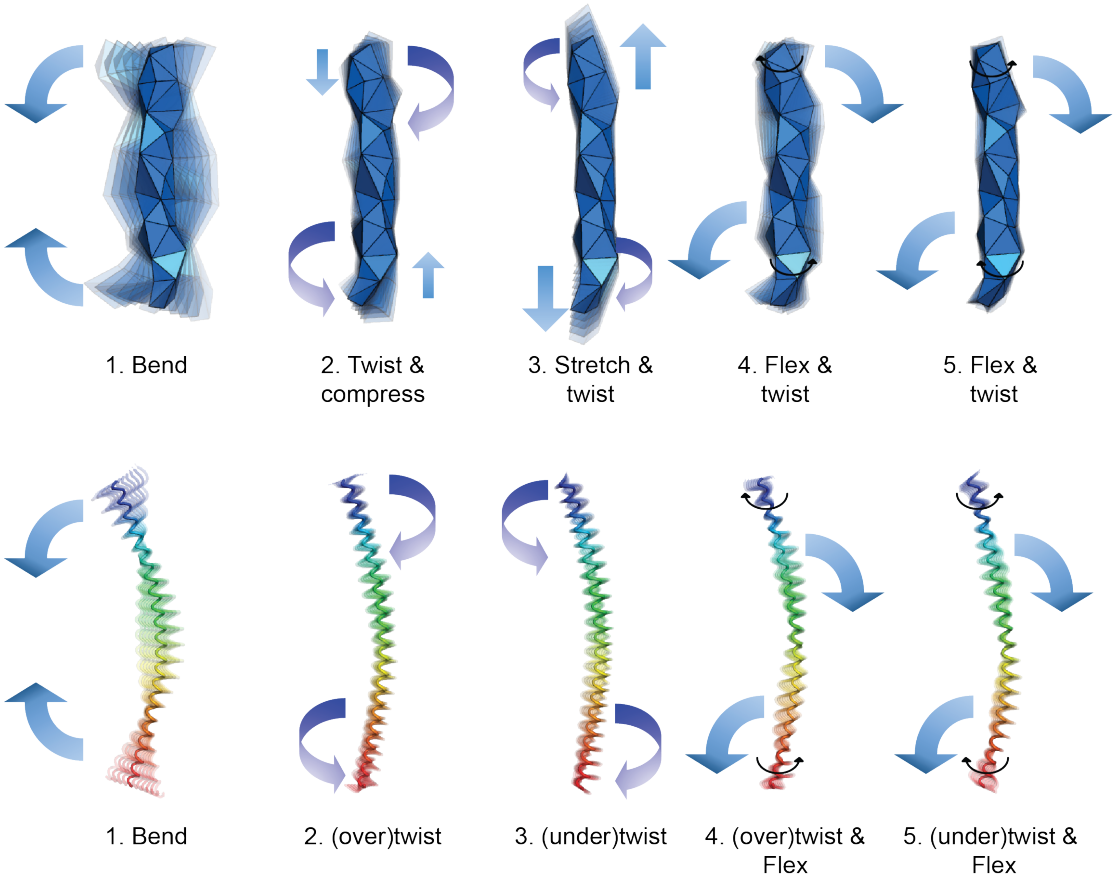


Figure 60 - PCA animations of top five modes.

Structures showing the superimposed frames from PCA animations of the top five eigenvector movements from FFEA (150MPa) (top) and explicitly simulated (bottom) SAH domains. It can be seen that mode 1 shows the largest movements, and they progressively get smaller.

For our purposes, component 1 is the most important for representing global bending of the myosin 7a molecule, while components 2 and 3 may be important for the dynamics at the local resolution, they become less important for the global dynamics. Comparing the eigenvalues for component 1 (bending), reveals that our stiffest FFEA SAH domain (1523 MPa), explores a smaller conformational space than our explicitly simulated SAH domain, which indicates the 1523 MPa SAH model is too stiff even compared with the explicitly simulated SAH domain. Modes 2 and 3 of the atomistically simulated SAH domain are less apparent in the FFEA model, which correspond to an “over” and “under” twist of the alpha helix. It is possible that the isotropic elasticity used in FFEA is a poor approximation to the twisting seen in the atomistic models. A twist in the FFEA model would involve some shearing of the structure perpendicular to the length. Mode 2 is seen as a slight compression but predominantly as a bend during the compression. Mode 3 is seen as a stretch which corresponds to an undertwist. We also note that in each individual simulation, many of the eigenvalues are similar in magnitude, indicating a near degeneracy of several of the eigenmodes. This is to be expected given the substantial symmetry of the SAH structure.

Using the mapping procedure described in chapter 5.2.3.3, we transformed each of the FFEA SAH trajectories into pseudo-atomistic trajectories before performing principal component analysis. Performing this transformation from the FFEA model onto the atomistic structure allows us to compare directly the dynamics of the two types of simulation trajectory on a common coordinate system. For our remaining 4 models (850, 500, 300 and 150 MPa), we compared the atomistic and FFEA pseudo-atomistic PCA eigensystems by calculating the inner product matrices of the 10 most flexible eigenmodes from each data set. We also performed an eigenvector comparison between the first and second halves of the atomistic trajectories, for which the these matrices are shown in Figure 61. In general, the explicit simulations tend to match the FFEA simulations much more closely than the implicit simulations, with the 150 MPa model showing high correlation with the explicit simulation (eigenspace overlap of 0.830) (Figure 61a). The lower correlation shown with the implicit simulations is likely due to the mismatch in the order of the eigenmodes described earlier (Figure 61).

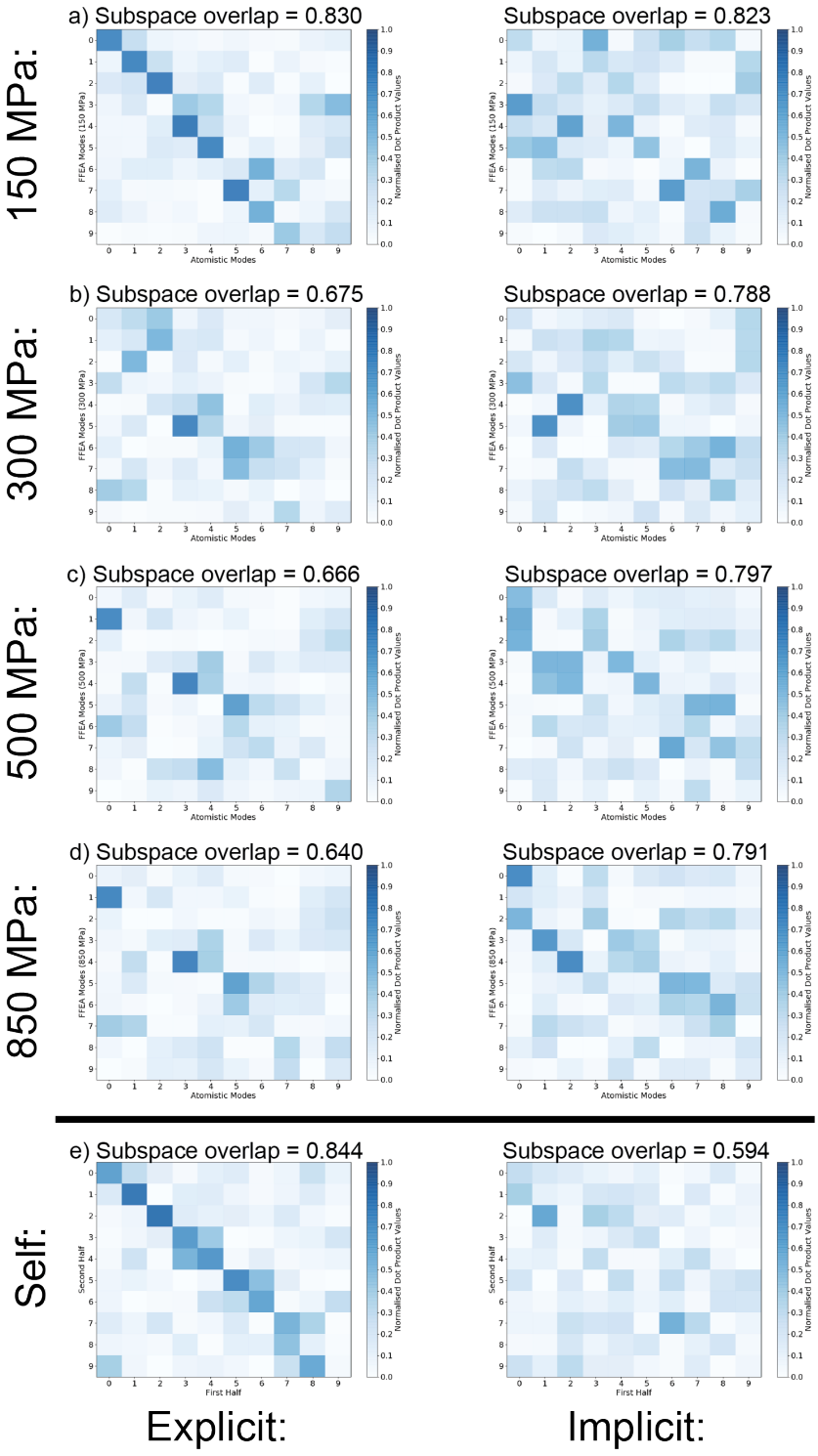


Figure 61 - Eigen-system inner product matrices for the dominant 10 modes of FFEA and all-atom SAH models.

a), b), c), and d) represent 150; 300; 500; and 850 MPa FFEA SAH models compared with explicitly (left) and implicitly (right) simulated SAH domain. The higher the subspace overlap indicates more similarity between the mode of motions between the systems. e) is the comparison between the first and second halves of the atomistic datasets.

5.3.3 Full molecule analysis.

We performed 8 FFEA simulations with bending stiffnesses for the SAH domain and IQ+CaM lever ranging from 12.5 - 1500pN.nm^2 (see table in Figure 62a), to examine the effect of these stiffness values on the ability of myosin 7a to achieve a configuration in which regulation is possible. The residues required for regulation are known to be located within the tip of the tail (R2140 & K2143), and we can estimate the binding interface on the motor from negative stain images of the compact state (Yang et al., 2009) (Figure 62b).

a)

Model #	Model structure ID	IQ+CaM lever Bending stiffness (pN.nm^2)	SAH Bending stiffness (pN.nm^2)	IQ+CaM lever Young's Modulus (MPa)	SAH Young's Modulus (MPa)	Description
1	A	1500	154	827	382	Values from literature *(Vilfan., 2005)* *(Sivaramakrishnan et al., 2008)*
2	A	270	135	320	50	On the order of this study
3	A	135	67.5	160	25	On the order of this study
4	A	67.5	33.75	80	12.5	Lowest possible
5	B	420	26	355	67.5	On the order of this study
6	C	420	26	355	67.5	On the order of this study
7	D	420	26	355	67.5	On the order of this study
8	E	420	26	355	67.5	On the order of this study

b)

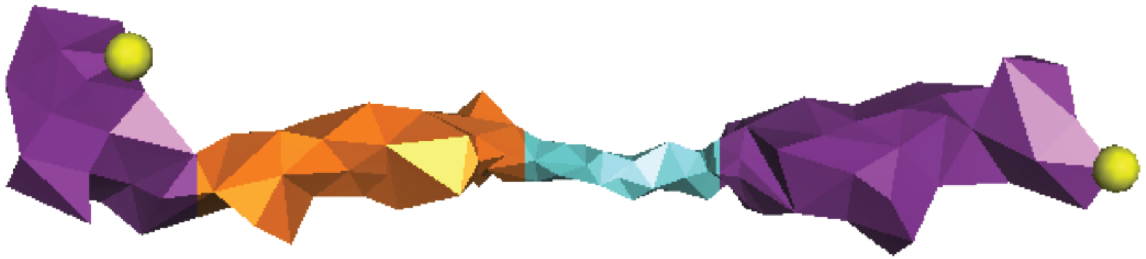


Figure 62 - Full-length myosin 7a model domain stiffness.

(a) The 8 FFEA models used in this section using model structures (refer to Figure 53) and the bending stiffness values for the (b) various structural domain of the IQ+CaM (orange) and SAH (blue) were parameterised with corresponding Young's Moduli. The end-to-end distance was measured between two nodes (yellow) for each of the simulation.

The end-to-end distance between these two points over the course of the simulation is plotted in Figure 63. Model 1 (Figure 63a), which has bending stiffness values obtained from the literature for the SAH domain and the IQ+CaM lever remains extended during the simulation (50.72 ± 0.69nm). As these values were far too stiff to enable regulation the Youngs moduli of both the IQ+CaM lever and SAH domain were incrementally reduced (Figure 63b – d) until the simulations were no longer numerically stable. We can see that model 4 (Figure 63) displays the shortest end-to-end distance yet model 3 shows a greater variance (Figure 63c).

Models created according to the findings reported by *Sakai et al., (2015)* in which the IQ motifs remain folded (Figure 63e – f) and where they have collapsed (Figure 63g – h) also show that the molecules never get into configurations where regulation is possible. These simulations reveal that obtaining these configurations through a stochastic process takes a long time as the molecule needs to explore its full range of configurations and these simulation timescales haven’t been long enough to probe this.

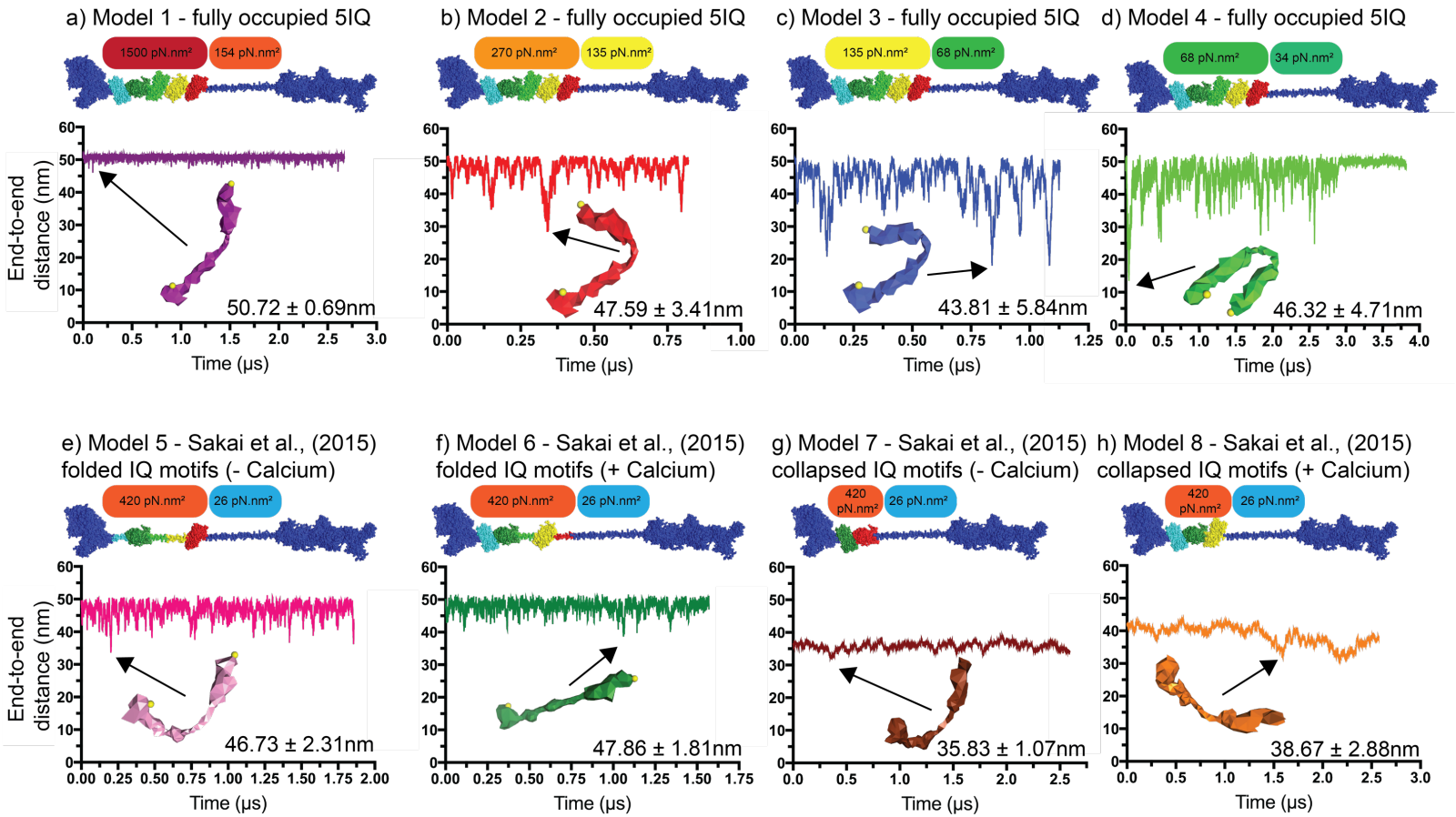


Figure 63 - End-to-end distance of full-length myosin 7a models with various stiffness parameters.

The end-to-end distance of each model over the course of the simulation was monitored between the points identified in Figure 62b and yellow spheres inset. The myosin 7a model (relating to Figure 53) and the bending stiffness values of the IQ+CaM lever and SAH domain (relating to Figure 62a) used in each simulation is shown above the corresponding graph. The models correspond to a fully occupied 5IQ (a - d), folded IQ motifs partially occupied by CaM according to *Sakai et al* (e,f), and collapsed IQ motifs partially occupied by CaM according to *Sakai et al* (g,h). Shown inset within each graph is a model corresponding to the closest distance achieved during the simulation. The average distance (mean ± S.D) is also included.

5.3.4 Intra-molecular bending.

Two independent all-atom simulation of myosin 7a revealed a tendency for the molecule to bend at SAH – 5IQ+CaM junction (Figure 64a). To investigate whether this interdomain bending is a real feature of the system, simulations of the 5IQ+CaM – SAH domain were performed to see if this phenomenon is repeated. To our surprise, the intramolecular bending was seen again with a what appeared like the SAH domain sticking against the distal calmodulin (Figure 64b).

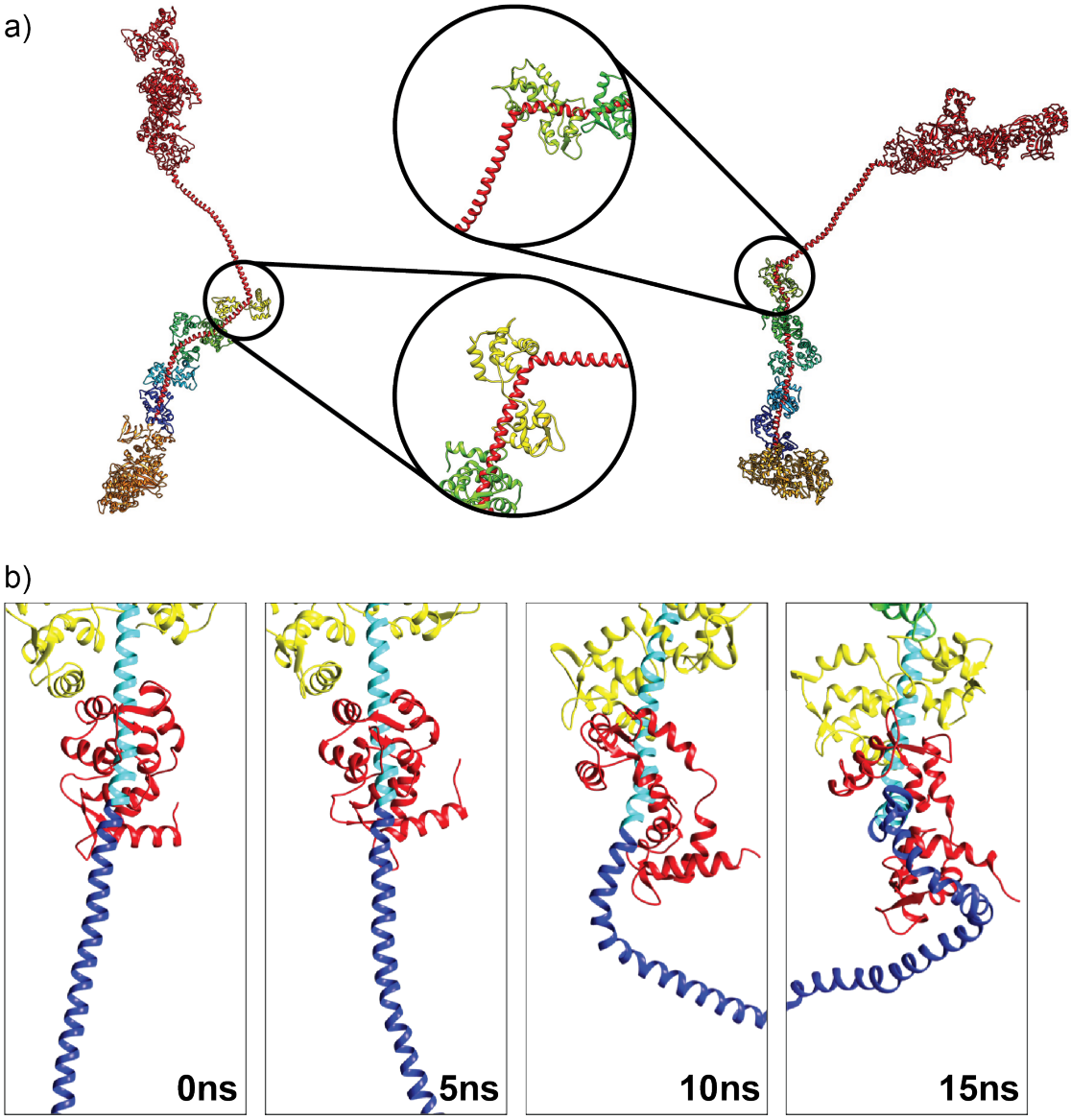


Figure 64 - SAH domain bending against the distal calmodulin.

(a) independent short implicit simulations of full length myosin 7a revealed the SAH domain would bend against the distal calmodulin. (b) Snapshots at different time points during a representative simulation reveals the SAH (dark blue) domain bending back against the distal calmodulin (red).

To examine the effect that the distal calmodulin was having on the secondary structure of the heavy chain, separate simulations with and without calmodulin bound to the 5th IQ were performed and the secondary structure of the heavy chain (V840 – E868) was monitored (Figure 66). The secondary structure of the heavy chain in the presence of the distal calmodulin was shown to form breaks much more prevalently in the alpha-helical structure when compared to the structure in which the distal calmodulin has been removed. This region (I851 – L863) is full of a number of charged residues, most notably arginine. Do these interactions between acidic residues of the SAH-junction and basic residues within the CaM sufficiently weaken the helical structure to allow an intramolecular hinge to form or is this merely a force field observation. It should be noted that this intramolecular bending has been reported by *Li et al., (2017)*. Comparison of the crystal structures with calmodulin bound to the 5th IQ motif of *Mus musculus* in the presence and absence of calcium reveals our structure resembles that of crystal structure of IQ5+CaM in the absence of calcium with the CaM straddling the IQ motif in anti-parallel fashion as expected. In this sense, our simulations should not have exposed the ~ 1 nm of IQ helix due to the calmodulin switching orientations in the IQ motif seen by *Li et al., (2017),* thus the melting of the lever helix at positions 855 – 862.

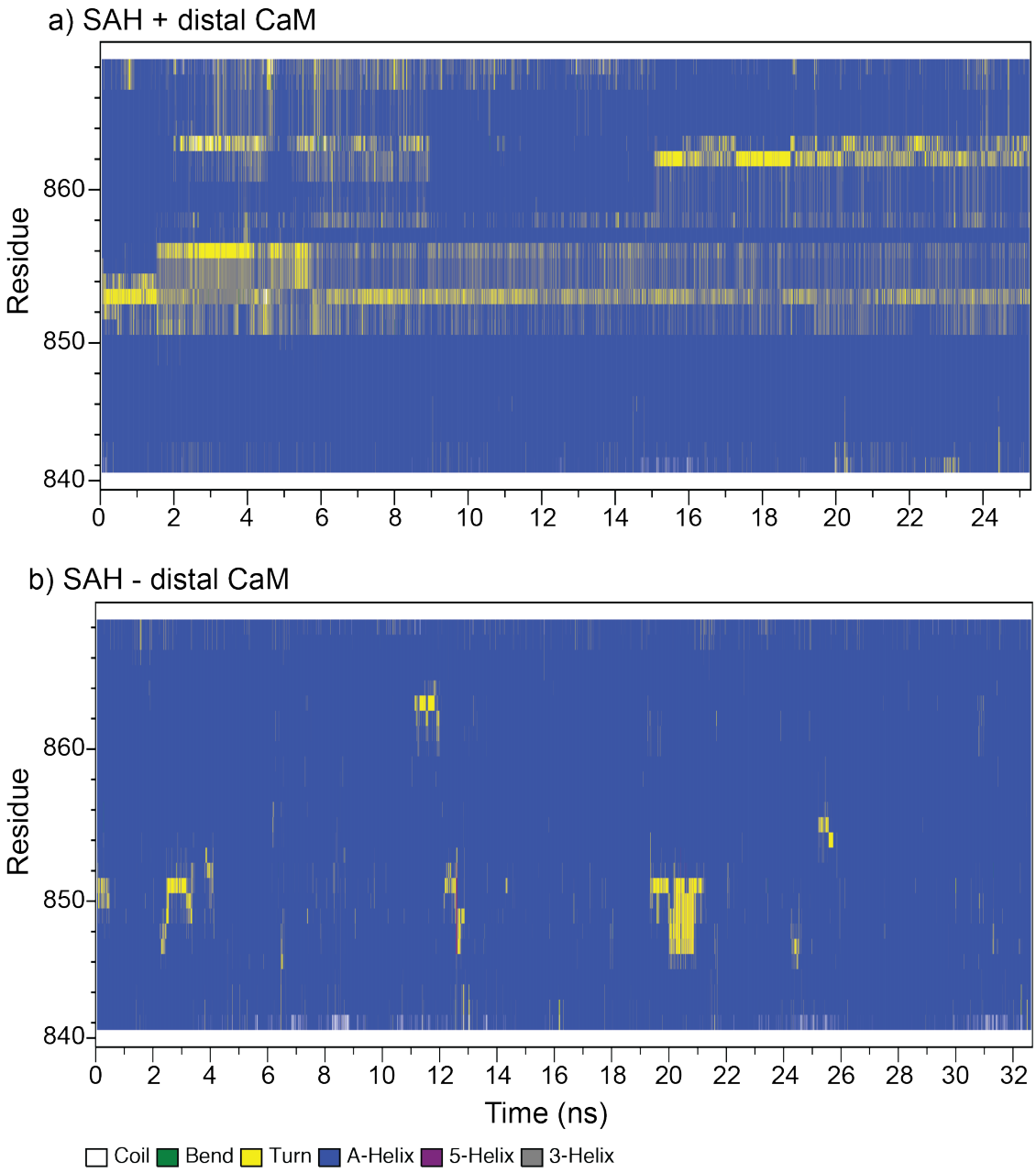


Figure 65 - Distal CaM effect on SAH secondary structure.

The secondary structure of the SAH domain (V840 - E868) was monitored during the course of the simulation in the (a) presence and (b) absence of calmodulin bound to IQ5.

5.4 Discussion.

With regards to the SAH domain, we have seen that we can replicate the overall dynamics of all-atom molecular dynamic simulations using our FFEA model. Moreover, the relative ordering of the eigenmodes is also closely replicated. From our simulations, an FFEA SAH domain model with a Young's moduli of 150 MPa, which corresponds to a bending stiffness of 60.5 $pN.nm^2$, most closely replicated the results of our all-atom SAH domain simulations in terms of eigenspace overlap. Due in part to the simple structure of the SAH domain, the relative dynamics of the model are easy to visualise in regards to its function within the myosin 7a. However, due to the nature of this continuum methodology, atomistic breaks in the alpha helical backbone are not replicated within FFEA due to its CG nature (Figure 58). This raises the question of whether these backbone breaks are important on the timescales of interest? This is unlikely due to the stochastic nature in which these breaks occur. However, more periodic breaks in the sequence that have been visualised can be replicated in FFEA by reducing the effective Young's modulus of the elements at this atomistic positions in order to mimic the effects seen in atomistic simulations.

A similar example was observed when looking at the effect of the distal calmodulin on the secondary structure of the SAH domain. This raises an interesting caveat about studying biomolecules. Studying the SAH domain by itself does not reveal this induced 'bending' observed when simulating the SAH domain with the distal calmodulin. This raises an interesting question, when studying large protein systems should we ever truncate the proteins being studied? A similar result has also been observed in which the crystal structure of the myosin 7a SAH domain and IQ5 from *Mus musulus* (834 – 935) in the presence of apo-CaM and Ca^{2+}-CaM has recently been solved (Li et al., 2017). As previously described (Section 1.15.2), the authors found that the distal calmodulin was able to change positions on the 5^{th} IQ motif to expose the SAH-IQ5 junction. Trypsin digestion and MD simulations revealed a Ca^{2+} induced flexibility change within the molecule akin to observations we had made. However, our lever model more closely resembles a CaM bound to IQ5 in the canonical antiparallel arrangement (absence of calcium) where *Li et al., (2017)* do not see this intramolecular bending. In our simulations, arginine is seen to

form bonds between the heavy chain and the CaM docked onto IQ5. Why we see this intramolecular bending still occurs needs to be further investigated.

We have successfully used FFEA to look at the dynamics of full-length myosin 7a which would otherwise not be within scope of other modelling techniques such as MD. Our simulations, qualitatively at least, have shown that in order to explore their full configuration space, simulations of the order of hundreds of microseconds need to be performed (Hanson et al., 2015). These lengths of simulations are far beyond the current computation capabilities for a molecule of this size. Myosins are known to be highly dynamic proteins in which thermal diffusion drives the dynamics of the molecule to perform its function. Take for example myosin 5 walking along actin filaments, following detachment of the trailing head, the free head must undergo a thermally-driven diffusive search, following the powerstroke distance in order to find the next binding site. iSCAT experiments examining the stepping dynamics have shown it takes on the order of hundreds of milliseconds for the trailing motor to move to the next binding site (Andrecka et al., 2015). These experiments examine the entire powerstroke cycle, not just the diffusive search following the powerstroke, yet it illustrates that the timescale of these events can be far in excess of the atomistic regime. FFEA is already able to explore the microsecond timescale regime, and with further optimisation of timestep length and the detail in which the system is described this can be increased.

Generally speaking, proteins have bending stiffnesses on the order of GPa to a few hundred MPa (Guthold et al., 2007). In our simulations, we found that far lower values were required in order to get into configurations in which the regulation could feasibly occur. This tells us a couple of things; first, that the simulations are not long enough and due to the stochastic nature of this problem this is unlikely to be seen; second, something structurally is explicitly missing. This can include structural deformations within secondary structures, as evidenced within the SAH domain.

One thing lacking from these simulations is the is the effect of unoccupied IQ motifs on the atomistic structure. In our simulations of full-length myosin 7a, the unoccupied IQ motifs either remain helical; which is unlikely to occur, or collapse; in which we simply remove the collapsed IQ motif sequence, which does not take into account the compliance that unstructured coil would introduce within the

lever. The introduction of calcium to myosin 5 within a flow cell was observed to inhibit the actin–based motility (Homma et al., 2000). Such inhibition was associated with dissociation and loss of one or more CaM molecules, since motility was restored in the absence of Ca^{2+} only by adding exogenous CaM. Similar inhibition/restoration was observed for chicken brush-border myosin-I (Collins et al., 1990). In both cases, the interpretation was that binding of Ca^{2+} to the CaM light chain weakened the interaction with an IQ sequence, thus causing dissociation and introducing compliance within the lever. This translates into a region which cannot be described as having a bending stiffness; as there should be no resistance against a bending force, but instead using a spring constant; which describes the overall extensibility of the IQ motif in its collapsed configuration. This would allow adjacent CaMs to pivot about the collapsed IQ motif. However, this requires a spring constant and equilibrium length in order to describe the collapsed IQ motif within FFEA (Solernou et al., 2018). These values could be obtained from MD simulations of collapsed IQ motifs, which would also tell us about the propensity of the hydrophobic residues to ‘stick’ to the interfaces between adjacent CaM molecules, which would also need to be parameterised into FFEA.

One of the major benefits demonstrated in this chapter is the advantage of using FFEA over all-atom MD simulate protein dynamics. Typically, MD needs to be run on a supercomputer for a long time using a small timestep, in contrast, FFEA can be run using a small number of processors with a much longer timestep to get very long time simulations. For mesoscopic processes that occur over time-scales of tens of microseconds to milliseconds length- scales of between 10 nm and 1 μm, such as the walking of cytoplasmic dynein along microtubules (Roberts et al., 2013) and the aggregation of fibrin molecules during blood clotting (Weisel and Litvinov, 2013), the development of a technique to explores these much larger length and timescales is of paramount importance.

Taken together, our simulations of myosin 7a has generated several interesting results but is still only the beginning of a much larger project. Many structural aspects contribute to the flexibility of myosin 7a across different length and timescales. These range from the very small, such as breaks in the alpha helical backbone, to very large, such as large structural dependencies such as an unoccupied IQ motif. To parameterisation of these features onto our FFEA model

will help us understand how flexibility arises in the myosin 7a molecule to allow it to adopt the compact state.

6. Summary and concluding remarks.

6.1 Summary.

This thesis has provided new information on the structure and flexibility within the myosin 7a lever. Perhaps the most important finding of this research is that the myosin 7a lever only binds calmodulin *in vivo*, something that many previous studies have neglected to acknowledge in their expression systems. In the second part of this thesis, an investigation into the occupancy of the myosin 7a lever by calmodulin was undertaken. Analysis of the old negative stain images of full-length myosin 7a from *drosophila melanogaster* revealed the lever to be much shorter than expected and that the orientation of the first calmodulin following the motor domain is different than that of published crystal structures from other myosins. The final chapter in this thesis tried to bridge the gap between computational simulation approaches of proteins with that of 'low resolution' experimental techniques such as negative stain EM and biochemical data. While only initial characterisations into the structural elements could be performed, they provide useful insight into ways the field can advance as a whole.

6.2 Summary of specific findings.

The principal findings from the different parts of this thesis are summarised below with possible directions for future work given thereafter.

6.2.1 Myosin 7a binds only calmodulin *in vivo.*

In chapter 3, a myosin 7a 5IQ lever construct (M7-IQ) was expressed in cells that endogenously express myosin 7a. Pull down experiments revealed that only calmodulin was co-precipitated with the lever construct suggesting that calmodulin is the preferred binding partner. This finding is important as alternative expression systems, such as Sf9 and *E. coli,* need to co-express light chain with the myosin 7a construct. In these cases, it was unknown whether calmodulin (or another light chain) was the preferred binding partner as they are simply provided in these cases with no information that about the possibility of other light chains binding to the myosin 7a IQ motifs. The adenoviral system used to express the M7-IQ and M5-IQ lever constructs could be expressed in different cell types to look for evidence of different light binding in different cell types which may allow the flexibility of the lever to be tailored for different functions in different cell types.

6.2.2 Truncated myosin 7a lever is very unstable.

Expression of the M7-IQ lever construct in Sf9 cells revealed the stability of the construct to be very low. Soluble protein could only be maintained in high (500 mM) salt conditions in comparison to a M5-IQ lever construct which was stable in ~140 mM salt. This made characterisation of the M7-IQ lever construct by negative stain EM and native mass-spectrometry unfeasible. Numerous expression attempts under increasing salt conditions with the addition of extra light chain and co-chaperone viruses did aid obtaining cleaner expression preparations. Given that calmodulin is the preferred light chain, it was surprising to see that RLC and ELC were co-purified with the M7-IQ construct. This may elucidate to some of the stability issues that we saw, with the extremely high amounts of light chain present during expression not being 'physiological', it may be that the IQ motifs may not be able to correctly fold to bind the correct light chain. Also stated previously, some myosins need to be co-expressed with chaperones in order to properly fold and function (Bird et al., 2014). Further work needs to be undertaken to find the optimal expression strategy to express the myosin 7a IQ motifs, perhaps by decreasing the MOI during expression, to obtain more folded protein.

6.2.3 Determination of light chain stoichiometry.

One of the main aims of chapter 3 was to determine the stoichiometry of the number of bound light chains to the myosin 7a lever. Gel densitometry measurements, using either coomasie or silver stained gels, tended to under estimate the number of bound light chains as evidenced using our M5-IQ lever construct which was shown to bind ~4 mol light chain per heavy chain, while we know it should bind 6 (Figure 28 and 29). Native mass-spectrometry was used as a quantitative methodology to determine the intact complex mass of the M7-IQ and M5-IQ constructs. Native mass-spec was successful in determining to correct complex mass of the M5-IQ lever construct, it also had the added benefit of telling us about other species present within the sample, namely that M5-IQ construct could be cleaved into a GFP and 6IQ+6CaM sub species which both remained stable (Figure 29). Unfortunately, calculating the intact mass of the M7-IQ was not possible, due to issues with stability as described in section 6.2.2. Due to high salt buffer that the M7-IQ construct was kept in to prevent aggregation, the sample needed to be exchanged into 200 mM ammonium acetate as high salt

conditions degrade the performance of native mass spectroscopy. Unfortunately, the M7-IQ construct aggregated following buffer exchange and failed to be electrosprayed. As a result, the masses of light chains were detected, but not those for heavy chains.

Native mass spectroscopy is a far superior technique to densitometry due to its precision and quantitative nature. However, not all proteins are amenable to native mass spectroscopy due to issues with storage in non-volatile buffers and other reagents nessacery to maintain stability of the protein, such as detergents used to stabilise membrane proteins (Leney and Heck, 2017). As discussed previously, if the M7-IQ expression and purification protocol can be improved, this would be a very useful experiment to repeat.

6.2.4 Negative stain-EM shows a shorter than expected lever.

Reanalysis of full-length myosin 7a in its extended and regulated forms in chapter 4 revealed the lever to be shorter than expected, with only a single CaM light chain being reliably resolved in the extended molecule (section 4.3.1). Masking did not reliably identify the lever which may have been lost from averaging and comparison of the area occupied by the regulated (216 ± 32.8 nm^2) and extended (135 ± 30.1 nm^2) particles revealed that some of the molecule may be projecting out of plane in the extended state. An interesting finding from this study revealed the orientation of the first CaM was in a RD-orientation, which is opposite to the orientation of the ELC docked onto the first IQ motif of scallop myosin S1 (1QVI), which is representative of myosins with the first IQ motif occupied. To us, this suggests that first IQ motif is unoccupied, as described by *Sakai et al., (2015)* where the first IQ motif is unoccupied in the presence of 1mM EGTA as these images were collected under. It should be noted that a common angle between the motor-lever junction was determined, which suggests that the motor and lever are not decoupled (discussed in 6.2.5).

A 34 Å resolution structure of the regulated state of myosin 7a could only reliably accommodate 3 CaM light chains plus the motor and tail into the density map. This finding reinforces the result that some of the molecule is projecting out of plane. Looking at myosin 7a by cryo-EM is the next logical step, as this would circumvent issues with the molecule projecting out of plane, as well the issues with preferred orientations on the carbon substrate (section 4.3.4) as all views should be possible in ice.

6.2.5 Motor-lever junction flexibility.

Classification of aligned head images provided evidence of particular flexibility at the motor-lever junction. A rotation of ~60° about the lever axis was identified, consistent with the findings in striated muscle myosin 2 S1 described by Billington et al. (2014). The distribution of angles suggested thermally-driven fluctuations about a mean, with a torsional spring constant of 78 pN·nm/rad^2 for movement of the motor relative to the lever about the motor-lever junction. The value presented here is higher than that reported by Billington et al. (2014), although this is likely due to the lever acting as a back stop in our experiments, whereas myosin 2 in the nucleotide free state was used in the Billington et al. (2014) study to allow a greater range of angle to be explored. The finding here suggests that the motor and lever are not decoupled from one another, which supports the results that artificially dimerised myosin 7b is processive (Yang et al., 2005).

However, this is in contrast to our previous finding that the first IQ motif is unoccupied (which would otherwise decouple the motor and lever) and that the 5IQ lever is not fully occupied by light chain. Perhaps the first IQ motif is able to form a stable structure against the motor / converter domain or some form of bridging between a single light chain and adjacent IQ motifs is occurring as described by Bayley et al. (2005). Many of these unexplained results could be resolved with information from an atomic level structure of the Myosin 7a S1 + bound light chain although due to the sheer size of this proteins this may not be possible. However, smaller structural domains may be more readily resolvable by crystallography.

Identifying these intramolecular ‘hinge’ regions within myosins is likely essential for their function, and paramertising the torsional spring constants which define these hinge regions is something that will need to be added to our FFEA models in the future.

6.2.6 Lever bending analysis

Lever bending of the M5-IQ construct was studied by image classification and by a method of numerically approximating the lever’s shape and mapping curvature with arc length. The purpose of the work was to determine the rigidity parameters to describe the bending stiffness. Assuming a simple isotropic rod model and assuming the lever as an unconstrained rod that can equilibrate in 2D a value for the bending stiffness of the lever of 156 ± 12 pN.nm^2 which is equivalent to a

persistence length of 38 ± 3 nm was determined. This stiffness number is surprisingly low, roughly equivalent to the bending stiffness of DNA. Could the myosin 5a levers be replaced with lengths of DNA and still function?

It is disappointing that the same type of analysis could not be performed for the M7-IQ construct. Not only would this allow us to compare the flexibility against the myosin 5 lever, but also to look for anisotropies in the lever, due to unoccupied IQ motifs, which introduce compliance within the lever. This methodology will need to be revisited to assess the bending stiffness of the M7-IQ lever construct when the stability issues are corrected for.

6.2.7 Capturing flexibility from simulations

The topic of chapter 5 was to examine the flexibility myosin 7a using FFEA. Unlike well-defined all-atom modelling regimes which have defined force fields to define atomic interactions, FFEA lacks a generalised force field to describe the stiffness parameters of various structural elements within proteins. The stiffness parameters of the SAH domain were elucidated from long-time molecular dynamics simulations by comparison with FFEA simulations parameterised with different stiffness values. A bending stiffness of 60.5 $pN.nm^2$ was found to most closely approximate the stiffness of the SAH in terms of major modes of motion explored and end-to-end distance. FFEA is not good for replicating effects caused by discrete atomic effects, namely breaks in the alpha helical backbone due to the coarse-grained nature of the simulation approach. While these discrete atomic details are unlikely to affect the major modes of motion of the SAH domain as an individual structural element, it is important to consider them. For example, the observation of an intra molecular hinge region between the distal calmodulin and SAH domain observed by us (section 5.3.4) and described by Li et al. (2017) would not be seen by our FFEA simulations as the atomic detail is lost. This is finding will need to be incorporated into future FFEA simulations of myosin 7a along with more accurate description of the bending stiffnesses of the lever domain obtained from results in this thesis.

The development of FFEA has allowed us to sample microsecond time scales with relative ease (compared to all-atom MD), taking us firmly into the mesoscale. However, with any stochastic process involving large molecules such as motor proteins, that occur across many tens of microseconds, bio simulation will always be difficult. FFEA offers a step in the right direction to bridge the gap between

occurring processes occurring at the experimental timescale regime (> μs) and processes at the nanoscale timescale regime (< ns).

6.3 Final conclusions.

This study into flexibility of myosin 7a has given some new insights into where flexibility exists in the myosin 7a lever. However, there are still a number of questions remaining for future study, some of which have been discussed here. Myosin 7a is a very large protein with very little structural data about these domains available. How it the stiffness of these domains both individually and together affect function of myosin 7a as a whole need to be fully explored.

7. Bibliography.

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindah, E. 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1**–**2**, pp.19–25.

Adamek, N., Coluccio, L.M. and Geeves, M.A. 2008. Calcium sensitivity of the cross-bridge cycle of Myo1c, the adaptation motor in the inner ear. *Proceedings of the National Academy of Sciences*. **105**(15), pp.5710–5715.

Alberts, B. 1998. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. *Cell*. **92**(3), pp.291–294.

Alder, B.J. and Wainwright, T.E. 1959. Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*. **31**(2), pp.459–466.

Andrecka, J., Arroyo, J.O., Takagi, Y., de Wit, G., Fineberg, A., MacKinnon, L., Young, G., Sellers, J.R. and Kukura, P. 2015. Structural dynamics of myosin 5 during processive motion revealed by interferometric scattering microscopy. *eLife*. **2015**(4).

Aschenbrenner, L., Lee, T. and Hasson, T. 2003. Myo6 facilitates the translocation of endocytic vesicles from cell peripheries. *Molecular biology of the cell*. **14**(7), pp.2728–43.

Avraham, K.B., Hasson, T., Steel, K.P., Kingsley, D.M., Russell, L.B., Mooseker, M.S., Copeland, N.G. and Jenkins, N.A. 1995. The mouse snell's waltzer deafness gene encodes an unconventional myosin required for structural integrity of inner ear hair cells. *Nature Genetics*. **11**(4), pp.369–375.

Baboolal, T.G., Mashanov, G.I., Nenasheva, T.A., Peckham, M. and Molloy, J.E. 2016. A combination of diffusion and active translocation localizes myosin 10 to the filopodial tip. *Journal of Biological Chemistry*. **291**(43), pp.22373–22385.

Baboolal, T.G., Sakamoto, T., Forgacs, E., White, H.D., Jackson, S.M., Takagi, Y., Farrow, R.E., Molloy, J.E., Knight, P.J., Sellers, J.R. and Peckham, M. 2009. The SAH domain extends the functional length of the myosin lever.

Proceedings of the National Academy of Sciences. **106**(52), pp.22193–22198.

Baker, L.A. and Rubinstein, J.L. 2010. Radiation damage in electron cryomicroscopy. *Methods in Enzymology*. **481**(C), pp.371–388.

Ballardini, R., Balzani, V., Credi, A., Gandolfi, M.T. and Venturi, M. 2001. Artificial molecular-level machines: Which energy to make them work? *Accounts of Chemical Research*. **34**(6), pp.445–455.

Batchelor, M., Wolny, M., Knight, P.J., Paci, E., Kalverda, A.P. and Peckham, M. 2017. Structural Dynamics in the Myosin 7A Single α-Helix Domain. *Biophysical Journal*. **112**(3), p.345a.

Belyantseva, I.A., Boger, E.T., Naz, S., Frolenkov, G.I., Sellers, J.R., Ahmed, Z.M., Griffith, A.J. and Friedman, T.B. 2005. Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia. *Nature Cell Biology*. **7**(2), pp.148–156.

Bensimon, D., Simon, A.J., Croquette, V. and Bensimon, A. 1995. Stretching DNA with a receding meniscus: Experiments and models. *Physical Review Letters*. **74**(23), pp.4754–4757.

Berg, J.S., Powell, B.C. and Cheney, R.E. 2001. A millennial myosin census. T. D. Pollard, ed. *Molecular biology of the cell*. **12**(4), pp.780–94.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. 1978. The protein data bank: A computer-based archival file for macromolecular structures. *Archives of Biochemistry and Biophysics*. **185**(2), pp.584–591.

Billington, N., Revill, D.J., Burgess, S.A., Chantler, P.D. and Knight, P.J. 2014. Flexibility within the heads of muscle myosin-2 molecules. *Journal of Molecular Biology*. **426**(4), pp.894–907.

Bird, J.E., Takagi, Y., Billington, N., Strub, M.-P., Sellers, J.R. and Friedman, T.B. 2014. Chaperone-enhanced purification of unconventional myosin 15, a molecular motor specialized for stereocilia protein trafficking. *Proceedings of the National Academy of Sciences*. **111**(34), pp.12390–12395.

Bond, S.D. and Leimkuhler, B.J. 2007. Molecular dynamics and the accuracy of

numerically computed averages. *Acta Numerica*. **16**, pp.1–65.

Borhani, D.W. and Shaw, D.E. 2012. The future of molecular dynamics simulations in drug discovery. *Journal of Computer-Aided Molecular Design*. **26**(1), pp.15–26.

Bozler, E. 1953. The mechanism of muscular relaxation. *Experientia*. **9**(1), pp.1–6.

Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K. and Jones, D.T. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research*. **41**(Web Server issue), pp.W349–W357.

Burgess, S., Walker, M., Wang, F., Sellers, J.P., White, H.D., Knight, P.J. and Trinick, J. 2002. The prepower stroke conformation of myosin V. *Journal of Cell Biology*. **159**(6), pp.983–991.

Burgess, S.A., Walker, M.L., Sakakibara, H., Knight, P.J. and Oiwa, K. 2003. Dynein structure and power stroke. *Nature*. **421**(6924), pp.715–718.

Burgess, S.A., Walker, M.L., Thirumurugan, K., Trinick, J. and Knight, P.J. 2004. Use of negative stain and single-particle image processing to explore dynamic properties of flexible macromolecules. *Journal of Structural Biology*. **147**(3), pp.247–258.

Burgess, S.A., Walker, M.L., White, H.D. and Trinick, J. 1997. Flexibility within myosin heads revealed by negative stain and single- particle analysis. *Journal of Cell Biology*. **139**(3), pp.675–681.

Buss, F., Arden, S.D., Lindsay, M., Luzio, J.P. and Kendrick-Jones, J. 2001. Myosin VI isoform localized to clathrin-coated vesicles with a role in clathrin-mediated endocytosis. *EMBO Journal*. **20**(14), pp.3676–3684.

Buss, F., Kendrick-Jones, J., Lionne, C., Knight, A.E., Côté, G.P. and Luzio, J.P. 1998. The localization of myosin VI at the Golgi complex and leading edge of fibroblasts and its phosphorylation and recruitment into membrane ruffles of A431 cells after growth factor stimulation. *Journal of Cell Biology*. **143**(6), pp.1535–1545.

Bustamante, C., Chemla, Y.R., Forde, N.R. and Izhaky, D. 2004. Mechanical Processes in Biochemistry. *Annual Review of Biochemistry*. **73**(1), pp.705–748.

Cellmer, T., Henry, E.R., Hofrichter, J. and Eaton, W.A. 2008. Measuring internal friction of an ultrafast-folding protein. *Proceedings of the National Academy of Sciences*. **105**(47), pp.18320–18325.

Cheney, R.E. and Mooseker, M.S. 1992. Unconventional myosins. *Current Opinion in Cell Biology*. **4**(1), pp.27–35.

Cheng, Y., Wolf, E., Larvie, M., Zak, O., Aisen, P., Grigorieff, N., Harrison, S.C. and Walz, T. 2006. Single particle reconstructions of the transferrin-transferrin receptor complex obtained with different specimen preparation techniques. *Journal of Molecular Biology*. **355**(5), pp.1048–1065.

Chishti, A.H., Kim, A.C., Marfatia, S.M., Lutchman, M., Hanspal, M., Jindal, H., Liu, S.C., Low, P.S., Rouleau, G.A., Mohandas, N., Chasis, J.A., Conboy, J.G., Gascard, P., Takakuwa, Y., Huang, S.C., Benz, E.J., Bretscher, A., Fehon, R.G., Gusella, J.F., Ramesh, V., Solomon, F., Marchesi, V.T., Tsukita, S., Tsukita, S., Arpin, M., Louvard, D., Tonks, N.K., Anderson, J.M., Fanning, A.S., Bryant, P.J., Woods, D.F. and Hoover, K.B. 1998. The FERM domain: A unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends in Biochemical Sciences*. **23**(8), pp.281–282.

Collins, J.H., Kuznicki, J., Bowers, B. and Korn, E.D. 1982. Comparison of the Actin Binding and Filament Formation Properties of Phosphorylated and Dephosphorylated Acanthamoeba Myosin II. *Biochemistry*. **21**(26), pp.6910–6915.

Collins, K., Sellers, J.R. and Matsudaira, P. 1990. Calmodulin dissociation regulates brush border myosin I (110-kD-calmodulin) mechanochemical activity in vitro. *Journal of Cell Biology*. **110**(4), pp.1137–1147.

Coureux, P.D., Sweeney, H.L. and Houdusse, A. 2004. Three myosin V structures delineate essential features of chemo-mechanical transduction. *EMBO Journal*. **23**(23), pp.4527–4537.

Coureux, P.D., Wells, A.L., Ménétrey, J., Yengo, C.M., Morris, C.A., Sweeney, H.L. and Houdusse, A. 2003. A structural state of the myosin V motor without bound nucleotide. *Nature*. **425**(6956), pp.419–423.

Craig, R., Smith, R. and Kendrick-Jones, J. 1983. Light-chain phosphorylation

controls the conformation of vertebrate non-muscle and smooth muscle myosin molecules. *Nature*. **302**(5907), pp.436–439.

Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H. and Shaw, D.E. 2012. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics*. **41**(1), pp.429–452.

Dulyaninova, N.G., Malashkevich, V.N., Almo, S.C. and Bresnick, A.R. 2005. Regulation of myosin-IIA assembly and Mts1 binding by heavy chain phosphorylation. *Biochemistry*. **44**(18), pp.6867–6876.

El-Amraoui, A. 2005. Usher I syndrome: unravelling the mechanisms that underlie the cohesion of the growing hair bundle in inner ear sensory cells. *Journal of Cell Science*. **118**(20), pp.4593–4603.

El-Mezgueldi, M. and Bagshaw, C.R. 2008. The Myosin Family: Biochemical And Kinetic Properties *In*: L. M. Coluccio, ed. *Myosins*. Dordrecht: Springer Netherlands, pp.55–93.

Fernandez-Leiro, R. and Scheres, S.H.W. 2017. A pipeline approach to single-particle processing in RELION *In*: *Acta Crystallographica Section D: Structural Biology*., pp.496–502.

Fidler, I.J. 1975. Biological Behavior of Malignant Melanoma Cells Correlated to Their Survival in Vivo. *Cancer Research*. **35**(1), pp.218–224.

Frénal, K., Polonais, V., Marq, J.B., Stratmann, R., Limenitakis, J. and Soldati-Favre, D. 2010. Functional dissection of the apicomplexan glideosome molecular architecture. *Cell Host and Microbe*. **8**(4), pp.343–357.

Frontali, C., Dore, E., Ferrauto, A., Gratton, E., Bettini, A., Pozzan, M.R. and Valdevit, E. 1979. An absolute method for the determination of the persistence length of native DNA from electron micrographs. *Biopolymers*. **18**(6), pp.1353–1373.

Futter, C.E., Ramalho, J.S., Jaissle, G.B., Seeliger, M.W. and Seabra, M.C. 2004. The role of Rab27a in the regulation of melanosome distribution within retinal pigment epithelial cells. *Molecular biology of the cell*. **15**(5), pp.2264–75.

Garcia, A.E. and Sanbonmatsu, K.Y. 2002. -Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proceedings of the National*

Academy of Sciences. **99**(5), pp.2782–2787.

Geeves, M.A. and Holmes, K.C. 2005. The molecular mechanism of muscle contraction. *Advances in Protein Chemistry*. **71**, pp.161–193.

Ghosh, T., Garde, S. and García, A.E. 2003. Role of Backbone Hydration and Salt-Bridge Formation in Stability of α-Helix in Solution. *Biophysical Journal*. **85**(5), pp.3187–3193.

Gibbs, D., Diemer, T., Khanobdee, K., Hu, J., Bok, D. and Williams, D.S. 2010. Function of MYO7A in the human RPE and the validity of shaker1 mice as a model for Usher syndrome 1B. *Investigative Ophthalmology and Visual Science*. **51**(2), pp.1130–1135.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K.A., Antonio, M., Beisel, K.W., Steel, K.P. and Brown, S.D.M. 1995. A type VII myosin encoded by the mouse deafness gene shaker-1. *Nature*. **374**(6517), pp.62–64.

Gillespie, P.G. and Müller, U. 2009. Mechanotransduction by Hair Cells: Models, Molecules, and Mechanisms. *Cell*. **139**(1), pp.33–44.

Graham, F.L., Smiley, J., Russell, W.C. and Nairn, R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *Journal of General Virology*. **36**(1), pp.59–72.

Grati, M. and Kachar, B. 2011a. Myosin VIIa and sans localization at stereocilia upper tip-link density implicates these Usher syndrome proteins in mechanotransduction. *Proceedings of the National Academy of Sciences*. **108**(28), pp.11476–11481.

Grati, M. and Kachar, B. 2011b. Myosin VIIa and sans localization at stereocilia upper tip-link density implicates these Usher syndrome proteins in mechanotransduction. *Proceedings of the National Academy of Sciences*. **108**(28), pp.11476–11481.

Gray, A., Harlen, O.G., Harris, S.A., Khalid, S., Leung, Y.M., Lonsdale, R., Mulholland, A.J., Pearson, A.R., Read, D.J. and Richardson, R.A. 2015. 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**(Pt 1), pp.162–172.

Gunasekaran, K., Ma, B. and Nussinov, R. 2004. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function and Genetics*. **57**(3), pp.433–443.

Guthold, M., Liu, W., Sparks, E.A., Jawerth, L.M., Peng, L., Falvo, M., Superfine, R., Hantgan, R.R. and Lord, S.T. 2007. A comparison of the mechanical and structural properties of fibrin fibers with other protein fibers. *Cell Biochemistry and Biophysics*. **49**(3), pp.165–181.

Guzik-Lendrum, S., Heissler, S.M., Billington, N., Takagi, Y., Yang, Y., Knight, P.J., Homsher, E. and Sellers, J.R. 2013. Mammalian myosin-18A, a highly divergent myosin. *Journal of Biological Chemistry*. **288**(13), pp.9532–9548.

Hackney, D.D. 1996. The kinetic cycles of myosin, kinesin, and dynein. *Annual review of physiology*. **58**, pp.731–50.

Hagen, S.J. 2010. Solvent viscosity and friction in protein folding dynamics. *Current Protein & Peptide Science*. **11**(5), pp.385–395.

Hanson, B., Richardson, R., Oliver, R., Read, D.J., Harlen, O. and Harris, S. 2015. Modelling biomacromolecular assemblies with continuum mechanics. *Biochemical Society Transactions*. **43**(2), pp.186–192.

Haran, G. 2012. How, when and why proteins collapse: The relation to folding. *Current Opinion in Structural Biology*. **22**(1), pp.14–20.

Hartman, M.A., Finan, D., Sivaramakrishnan, S. and Spudich, J.A. 2011. Principles of Unconventional Myosin Function and Targeting. *Annual Review of Cell and Developmental Biology*. **27**(1), pp.133–155.

Hasson, T., Gillespie, P.G., Garcia, J.A., MacDonald, R.B., Zhao, Y.D., Yee, A.G., Mooseker, M.S. and Corey, D.P. 1997. Unconventional myosins in inner-ear sensory epithelia. *Journal of Cell Biology*. **137**(6), pp.1287–1307.

Hasson, T., Heintzelman, M.B., Santos-Sacchi, J., Corey, D.P. and Mooseker, M.S. 1995. Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type 1B. *Proceedings of the National Academy of Sciences of the United States of America*. **92**(21), pp.9815–9819.

Hayat, M.A. 2002. *Principles and techniques of electron microscopy: biological applications*.

Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. and Schatz, M. 1996. A new generation of the IMAGIC image processing system. *Journal of Structural Biology*. **116**(1), pp.17–24.

Heissler, S.M. and Manstein, D.J. 2013. Nonmuscle myosin-2: Mix and match. *Cellular and Molecular Life Sciences*. **70**(1), pp.1–21.

Heissler, S.M. and Sellers, J.R. 2014. Myosin light chains: Teaching old dogs new tricks. *Bioarchitecture*. **4**(6), pp.169–188.

Heissler, S.M. and Sellers, J.R. 2016. Various Themes of Myosin Regulation. *Journal of Molecular Biology*. **428**(9), pp.1927–1946.

Henderson, R. 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*. **110**(45), pp.18037–18041.

Henn, A. and De La Cruz, E.M. 2005. Vertebrate myosin VIIb is a high duty ratio motor adapted for generating and maintaining tension. *Journal of Biological Chemistry*. **280**(47), pp.39665–39676.

Herm-Götz, A., Weiss, S., Stratmann, R., Fujita-Becker, S., Ruff, C., Meyhöfer, E., Soldati, T., Manstein, D.J., Geeves, M.A. and Soldati, D. 2002. Toxoplasma gondii myosin A and its light chain: A fast, single-headed, plus-end-directed motor. *EMBO Journal*. **21**(9), pp.2149–2158.

Homma, K., Saito, J., Ikebe, R. and Ikebe, M. 2000. Ca2+-dependent regulation of the motor activity of myosin V. *Journal of Biological Chemistry*. **275**(44), pp.34766–34771.

Hopkins, R.F. and Esposito, D. 2009. A rapid method for titrating baculovirus stocks using the Sf-9 Easy Titer cell line. *BioTechniques*. **47**(3), pp.785–788.

Houdusse, A. and Cohen, C. 1996. Structure of the regulatory domain of scallop myosin at 2 Å resolution: Implications for regulation. *Structure*. **4**(1), pp.21–32.

Houdusse, A., Gaucher, J.-F., Krementsova, E., Mui, S., Trybus, K.M. and Cohen, C. 2006. Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. *Proceedings of the National Academy of Sciences*. **103**(51), pp.19326–19331.

Houdusse, A., Silver, M. and Cohen, C. 1996. A model of Ca2+-free calmodulin binding to unconventional myosins reveals how calmodulin acts as a regulatory switch. *Structure*. **4**(12), pp.1475–1490.

Houdusse, A., Szent-Gyorgyi, A.G. and Cohen, C. 2000. Three conformational states of scallop myosin S1. *Proceedings of the National Academy of Sciences*. **97**(21), pp.11238–11243.

Howard, J. and Clark, R. 2002. Mechanics of Motor Proteins and the Cytoskeleton. *Applied Mechanics Reviews*. **55**(2), p.B39.

Howard, J. and Spudich, J.A. 1996. Is the lever arm of myosin a molecular elastic element? *Proceedings of the National Academy of Sciences of the United States of America*. **93**(9), pp.4462–4464.

Huanga, D. and Caflischa, A. 2010. Library screening by fragment-based docking. *Journal of Molecular Recognition*. **23**(2), pp.183–193.

Hudspeth, A.J. 2014. Integrating the active process of hair cells with cochlear function. *Nature Reviews Neuroscience*. **15**(9), pp.600–614.

Huxley, H.E. 1963. Electron microscope studies on the structure of natural and synthetic protein filaments from striated muscle. *Journal of Molecular Biology*. **7**(3), pp.IN1-IN30.

Huxley, H.E. 1957. the Double Array of Filaments in Cross-Striated Muscle. *The Journal of Cell Biology*. **3**(5), pp.631–648.

Imai, H., Shima, T., Sutoh, K., Walker, M.L., Knight, P.J., Kon, T. and Burgess, S.A. 2015. Direct observation shows superposition and large scale flexibility within cytoplasmic dynein motors moving along microtubules. *Nature Communications*. **6**.

Jung, H.S., Billington, N., Thirumurugan, K., Salzameda, B., Cremo, C.R., Chalovich, J.M., Chantler, P.D. and Knight, P.J. 2011. Role of the tail in the regulated state of myosin 2. *Journal of Molecular Biology*. **408**(5), pp.863–878.

Jung, H.S., Burgess, S. a, Billington, N., Colegrave, M., Patel, H., Chalovich, J.M., Chantler, P.D. and Knight, P.J. 2008. Conservation of the regulated structure of folded myosin 2 in species separated by at least 600 million years of independent evolution. *Proc Natl Acad Sci U S A*. **105**(16),

pp.6022–6026.

Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. **22**(12), pp.2577–2637.

Kamm, K.E. and Stull, J.T. 2011. Signaling to myosin regulatory light chain in sarcomeres. *Journal of Biological Chemistry*. **286**(12), pp.9941–9947.

Karabina, A., Kazmierczak, K., Szczesna-Cordary, D. and Moore, J.R. 2015. Myosin regulatory light chain phosphorylation enhances cardiac β-myosin in vitro motility under load. *Archives of Biochemistry and Biophysics*. **580**, pp.14–21.

Kawasaki, H., Nakayama, S. and Kretsinger, R.H. 1998. Classification and evolution of EF-hand proteins. *BioMetals*. **11**(4), pp.277–295.

Kelleher, J.F., Mandell, M.A., Moulder, G., Hill, K.L., L'Hernault, S.W., Barstead, R. and Titus, M.A. 2000. Myosin VI is required for asymmetric segregation of cellular components during C. elegans spermatogenesis. *Current Biology*. **10**(23), pp.1489–1494.

Knight, P. and Trinick, J. 1984. Structure of the myosin projections on native thick filaments from vertebrate skeletal muscle. *Journal of Molecular Biology*. **177**(3), pp.461–482.

Knight, P.J., Thirumurugan, K., Xu, Y., Wang, F., Kalverda, A.P., Stafford, W.F., Sellers, J.R. and Peckham, M. 2005. The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *Journal of Biological Chemistry*. **280**(41), pp.34702–34708.

Köhler, D., Ruff, C., Meyhöfer, E. and Bähler, M. 2003. Different degrees of lever arm rotation control myosin step size. *Journal of Cell Biology*. **161**(2), pp.237–241.

Kollmar, M., D??rrwang, U., Kliche, W., Manstein, D.J. and Kull, F.J. 2002. Crystal structure of the motor domain of a class-I myosin. *EMBO Journal*. **21**(11), pp.2517–2525.

Kon, T., Sutoh, K. and Kurisu, G. 2011. X-ray structure of a functional full-length dynein motor domain. *Nature Structural and Molecular Biology*. **18**(6), pp.638–642.

Kuczmarski, E.R. and Spudich, J.A. 1980. Regulation of myosin self-assembly: Phosphorylation of Dictyostelium heavy chain inhibits formation of thick filaments (cell movement/nonmuscle contraction/myosin ATPase/protein kinase). *Cell Biology*. **77**(12), pp.7292–4296.

De La Cruz, E.M. and Michael Ostap, E. 2009. Chapter 6 Kinetic and Equilibrium Analysis of the Myosin ATPase. *Methods in Enzymology*. **455**(A), pp.157–192.

De La Cruz, E.M., Wells, A.L., Rosenfeld, S.S., Ostap, E.M. and Sweeney, H.L. 1999. The kinetic mechanism of myosin V. *Proceedings of the National Academy of Sciences*. **96**(24), pp.13726–13731.

Laakso, J.M., Lewis, J.H., Shuman, H. and Ostap, E.M. 2010. Control of myosin-I force sensing by alternative splicing. *Proceedings of the National Academy of Sciences*. **107**(2), pp.698–702.

Laakso, J.M., Lewis, J.H., Shuman, H. and Ostap, E.M. 2008. Myosin I can act as a molecular force sensor. *Science*. **321**(5885), pp.133–136.

Lamour, G., Kirkegaard, J.B., Li, H., Knowles, T.P.J. and Gsponer, J. 2014. Easyworm: An open-source software tool to determine the mechanical properties of worm-like chains. *Source Code for Biology and Medicine*. **9**(1).

Lan, G. and Sun, S.X. 2005. Dynamics of myosin-V processivity. *Biophysical Journal*. **88**(2), pp.999–1008.

Leach, A.R. 2001. *Molecular modelling : principles and applications*. Pearson Prentice Hall.

Leney, A.C. and Heck, A.J.R. 2017. Native Mass Spectrometry: What is in the Name? *Journal of the American Society for Mass Spectrometry*. **28**(1), pp.5–13.

Levine, R.J.C., Kensler, R.W., Yang, Z., Stull, J.T. and Sweeney, H.L. 1996. Myosin light chain phosphorylation affects the structure of rabbit skeletal muscle thick filaments. *Biophysical Journal*. **71**(2), pp.898–907.

Levitt, M. 2001. The birth of computational structural biology. *Nature Structural Biology*. **8**(5), pp.392–393.

Levitt, M. and Lifson, S. 1969. Refinement of protein conformations using a

macromolecular energy minimization procedure. *Journal of Molecular Biology*. **46**(2), pp.269–279.

Li, J., Chen, Y., Deng, Y., Unarta, I.C., Lu, Q., Huang, X. and Zhang, M. 2017. Ca2+-Induced Rigidity Change of the Myosin VIIa IQ Motif-Single α Helix Lever Arm Extension. *Structure*. **25**(4), p.579–591.e4.

Li, X.D., Ikebe, R. and Ikebei, M. 2005. Activation of myosin Va function by melanophilin, a specific docking partner of myosin Va. *Journal of Biological Chemistry*. **280**(18), pp.17815–17822.

Li, X.D., Mabuchi, K., Ikebe, R. and Ikebe, M. 2004. Ca2+-induced activation of ATPase activity of myosin Va is accompanied with a large conformational change. *Biochemical and Biophysical Research Communications*. **315**(3), pp.538–545.

Lifson, S. and Warshel, A. 1968. Consistent Force Field for Calculations of Conformations, Vibrational Spectra, and Enthalpies of Cycloalkane and *n* -Alkane Molecules. *The Journal of Chemical Physics*. **49**(11), pp.5116–5129.

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics*. **78**(8), pp.1950–1958.

Lister, I., Schmitz, S., Walker, M., Trinick, J., Buss, F., Veigel, C. and Kendrick-Jones, J. 2004. A monomeric myosin VI with a large working stroke. *EMBO Journal*. **23**(8), pp.1729–1738.

Liu, J., Taylor, D.W., Krementsova, E.B., Trybus, K.M. and Taylor, K.A. 2006. Three-dimensional structure of the myosin V inhibited state by cryoelectron tomography. *Nature*. **442**(7099), pp.208–211.

Liu, X., Hong, M.-S., Shu, S., Yu, S. and Korn, E.D. 2012. Regulation of the filament structure and assembly of Acanthamoeba myosin II by phosphorylation of serines in the heavy-chain nonhelical tailpiece. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, pp.E33-40.

Liu, X., Vansant, G., Udovichenko, I.P., Wolfrum, U. and Williams, D.S. 1997.

Myosin VIIa, the product of the Usher 1B syndrome gene, is concentrated in the connecting cilia of photoreceptor cells. *Cell Motility and the Cytoskeleton*. **37**(3), pp.240–252.

Lopes, V.S., Gibbs, D., Libby, R.T., Aleman, T.S., Welch, D.L., Lillo, C., Jacobson, S.G., Radu, R.A., Steel, K.P. and Williams, D.S. 2011. The Usher 1B protein, MYO7A, is required for normal localization and function of the visual retinoid cycle enzyme, RPE65. *Human Molecular Genetics*. **20**(13), pp.2560–2570.

Lu, Q., Ye, F., Wei, Z., Wen, Z. and Zhang, M. 2012. Antiparallel coiled-coil - mediated dimerization of myosin X. *Pnas*. **109**, pp.17388–17393.

Lupas, A., Van Dyke, M. and Stock, J. 1991. Predicting coiled coils from protein sequences. *Science*. **252**(5009), pp.1162–1164.

Marrink, S.J. and Tieleman, D.P. 2013. Perspective on the martini model. *Chemical Society Reviews*. **42**(16), pp.6801–6822.

Marrink, S.J., de Vries, A.H. and Mark, A.E. 2004. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B*. **108**(2), pp.750–760.

McGuffee, S.R. and Elcock, A.H. 2010. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLOS Comptational Biology*. **6**(3), p.e1000694.

Mehta, A.D., Rock, R.S., Rief, M., Spudich, J.A., Mooseker, M.S. and Cheney, R.E. 1999. Myosin-V is a processive actin-based motor. *Nature*. **400**(6744), pp.590–593.

Meyer, T., Ferrer-Costa, C., Pérez, A., Rueda, M., Bidon-Chanal, A., Luque, F.J., Laughton, C.A. and Orozco, M. 2006. Essential dynamics: A tool for efficient trajectory compression and management. *Journal of Chemical Theory and Computation*. **2**(2), pp.251–258.

Moore, G.E. 1998. Cramming more components onto integrated circuits. *Proceedings of the IEEE*. **86**(1), pp.82–85.

Moore, J.R., Krementsova, E.B., Trybus, K.M. and Warshaw, D.M. 2004. Does the myosin V neck region act as a lever? *Journal of Muscle Research and Cell Motility*. **25**(1), pp.29–35.

Mücke, N., Klenin, K., Kirmse, R., Bussiek, M., Herrmann, H., Hafner, M. and Langowski, J. 2009. Filamentous biopolymers on surfaces: Atomic force microscopy images compared with brownian dynamics simulation of filament deposition. *PLoS ONE*. **4**(11).

Mücke, N., Kreplak, L., Kirmse, R., Wedig, T., Herrmann, H., Aebi, U. and Langowski, J. 2004. Assessing the Flexibility of Intermediate Filaments by Atomic Force Microscopy. *Journal of Molecular Biology*. **335**(5), pp.1241–1250.

Mukherjea, M., Llinas, P., Kim, H.J., Travaglia, M., Safer, D., Ménétrey, J., Franzini-Armstrong, C., Selvin, P.R., Houdusse, A. and Sweeney, H.L. 2009. Myosin VI Dimerization Triggers an Unfolding of a Three-Helix Bundle in Order to Extend Its Reach. *Molecular Cell*. **35**(3), pp.305–315.

Nelson, M.R. and Chazin, W.J. 1998. Structures of EF-hand Ca2+-binding proteins: Diversity in the organization, packing and response to Ca2+ binding. *BioMetals*. **11**(4), pp.297–318.

Nguyen, H.A. and Higuchi, H. 2005. Motility of myosin V regulated by the dissociation of single calmodulin. *Nature Structural and Molecular Biology*. **12**(2), pp.127–132.

Ohi, M., Li, Y., Cheng, Y. and Walz, T. 2004. Negative staining and image classification - Powerful tools in modern electron microscopy. *Biological Procedures Online*. **6**(1), pp.23–34.

Oke, O.A., Burgess, S.A., Forgacs, E., Knight, P.J., Sakamoto, T., Sellers, J.R., White, H. and Trinick, J. 2010. Influence of lever structure on myosin 5a walking. *Proceedings of the National Academy of Sciences*. **107**(6), pp.2509–2514.

Oliver, R., Read, D.J., Harlen, O.G. and Harris, S.A. 2012. A Stochastic Finite Element Model for the Dynamics of Globular Macromolecules. *Journal of Computational Physics*. **239**(0), pp.147–165.

Pan, L. and Zhang, M. 2012. Structures of Usher Syndrome 1 Proteins and Their Complexes. *Physiology*. **27**(1), pp.25–42.

Parks, R.J., Cummings, D.T., Evelegh, C.M., Sankar, U. and Graham, F.L. 1999. A High-Efficiency Cre/loxP-Based System for Construction of

Adenoviral Vectors. *Human Gene Therapy*. **10**(16), pp.2667–2672.

Parry, D.A.D. 1982. Coiled-coils in α-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. *Bioscience Reports*. **2**(12), pp.1017–1024.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. **25**(13), pp.1605–1612.

Phichith, D., Travaglia, M., Yang, Z., Liu, X., Zong, A.B., Safer, D. and Sweeney, H.L. 2009. Cargo binding induces dimerization of myosin VI. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(41), pp.17320–17324.

Piana, S., Klepeis, J.L. and Shaw, D.E. 2014. Assessing the accuracy of physical models used in protein-folding simulations: Quantitative evidence from long molecular dynamics simulations. *Current Opinion in Structural Biology*. **24**(1), pp.98–105.

Quillin, M.L. and Matthews, B.W. 2000. Accurate calculation of the density of proteins. *Acta Crystallographica Section D: Biological Crystallography*. **56**(7), pp.791–794.

Rader, A.J., Chennubhotla, C., Yang, L.-W. and Bahar, I. 2006. The Gaussian Network Model: theory and applications. *Normal Mode Analysis - theory and applications to biological and chemical systems*. **10**(20), pp.41–64.

Radermacher, M., Rao, V., Grassucci, R., Frank, J., Timerman, A.P., Fleischer, S. and Wagenknecht, T. 1994. Cryoelectron Microscopy And 3-Dimensional Reconstruction Of The Calcium-Release Channel Ryanodine Receptor From Skeletal-Muscle. *Journal of Cell Biology*. **127**(2), pp.411–423.

Radmacher, M., Fritz, M., Cleveland, J.P., Walters, D.A. and Hansma, P.K. 1994. Imaging Adhesion Forces and Elasticity of Lysozyme Adsorbed on Mica with the Atomic Force Microscope. *Langmuir*. **10**(10), pp.3809–3814.

Rayment, I., Rypniewski, W.R., Schmidt-Bäse, K., Smith, R., Tomchick, D.R.,

Benning, M.M., Winkelmann, D.A., Wesenberg, G. and Holden, H.M. 1993. Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science*. **261**(5117), pp.50–58.

Reck-Peterson, S.L., Provance, D.W., Mooseker, M.S. and Mercer, J.A. 2000. Class V myosins. *Biochimica et Biophysica Acta - Molecular Cell Research*. **1496**(1), pp.36–51.

Reif, F. and Scott, H.L. 1998. *Fundamentals of Statistical and Thermal Physics* (N. Y. McGraw-Hill, ed.). New York.

Rhoads, A.R. and Friedberg, F. 1997. Sequence motifs for calmodulin recognition. *The FASEB journal*. **11**(5), pp.331–340.

Rice, S., Lin, A.W., Safer, D., Hart, C.L., Naber, N., Carragher, B.O., Cain, S.M., Pechatnikova, E., Wilson-Kubalek, E.M., Whittaker, M., Pate, E., Cooke, R., Taylor, E.W., Milligan, R.A. and Vale, R.D. 1999. A structural change in the kinesin motor protein that drives motility. *Nature*. **402**(6763), pp.778–784.

Richardson, G.P., de Monvel, J.B. and Petit, C. 2011. How the Genetics of Deafness Illuminates Auditory Physiology. *Annual Review of Physiology*. **73**(1), pp.311–334.

Richardson, R.A., Papachristos, K., Read, D.J., Harlen, O.G., Harrison, M., Paci, E., Muench, S.P. and Harris, S.A. 2014. Understanding the apparent stator-rotor connections in the rotary ATPase family using coarse-grained computer modeling. *Proteins: Structure, Function and Bioinformatics*. **82**(12), pp.3298–3311.

Rief, M., Rock, R.S., Mehta, A.D., Mooseker, M.S., Cheney, R.E. and Spudich, J.A. 2000. Myosin-V stepping kinetics: A molecular model for processivity. *Proceedings of the National Academy of Sciences*. **97**(17), pp.9482–9486.

Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K. and Burgess, S.A. 2013. Functions and mechanics of dynein motor proteins. *Nature Reviews Molecular Cell Biology*. **14**(11), pp.713–726.

Roberts, A.J., Numata, N., Walker, M.L., Kato, Y.S., Malkova, B., Kon, T., Ohkura, R., Arisaka, F., Knight, P.J., Sutoh, K. and Burgess, S.A. 2009. AAA+ Ring and Linker Swing Mechanism in the Dynein Motor. *Cell*. **136**(3),

pp.485–495.

Rogers, M.S. and Strehler, E.E. 2001. The Tumor-sensitive Calmodulin-like Protein Is a Specific Light Chain of Human Unconventional Myosin X. *Journal of Biological Chemistry*. **276**(15), pp.12182–12189.

Roy, A., Kucukural, A. and Zhang, Y. 2010. I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*. **5**(4), pp.725–738.

Sakai, T., Jung, H.S., Sato, O., Yamada, M.D., You, D.J., Ikebe, R. and Ikebe, M. 2015. Structure and regulation of the movement of human myosin VIIA. *Journal of Biological Chemistry*. **290**(28), pp.17587–17598.

Sakamoto, T., Wang, F., Schmitz, S., Xu, Y., Xu, Q., Molloy, J.E., Veigel, C. and Sellers, J.R. 2003. Neck length and processivity of myosin V. *Journal of Biological Chemistry*. **278**(31), pp.29201–29207.

Salomon-Ferrer, R., Case, D.A. and Walker, R.C. 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. **3**(2), pp.198–210.

Sato, O., Komatsu, S., Sakai, T., Tsukasaki, Y., Tanaka, R., Mizutani, T., Watanabe, T.M., Ikebe, R. and Ikebe, M. 2017. Human myosin VIIa is a very slow processive motor protein on various cellular actin structures. *Journal of Biological Chemistry*. **292**(26), pp.10950–10960.

Scarff, C.A., Fuller, M.J.G., Thompson, R.F. and Iadaza, M.G. 2018. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. *Journal of Visualized Experiments*. (132).

Schleicher, M. and Watterson, D.M. 1983. Analysis of differences between Coomassie blue stain and silver stain procedures in polyacrylamide gels: Conditions for the detection of calmodulin and troponin C. *Analytical Biochemistry*. **131**(2), pp.312–317.

Schöberl, J. 1997. An advancing front 2D/3D-mesh generator based on abstract rules. *Computing and Visualization in Science*. **1**(1), pp.41–52.

Scholey, J.M., Taylor, K.A. and Kendrick-jones, J. 1980. Regulation of non-muscle myosin assembly by calmodulin-dependent light chain kinase. *Nature*. **287**(5779), pp.233–235.

Seeber, M., Cecchini, M., Rao, F., Settanni, G. and Caflisch, A. 2007. Wordom: A program for efficient analysis of molecular dynamics simulations. *Bioinformatics*. **23**(19), pp.2625–2627.

Self, T., Sobe, T., Copeland, N.G., Jenkins, N.A., Avraham, K.B. and Steel, K.P. 1999. Role of myosin VI in the differentiation of cochlear hair cells. *Developmental Biology*. **214**(2), pp.331–341.

Sellers, J.R. 1985. Mechanism of the phosphorylation-dependent regulation of smooth muscle heavy meromyosin. *Journal of Biological Chemistry*. **260**(29), pp.15815–15819.

Sellers, J.R. 2000. Myosins: A diverse superfamily. *Biochimica et Biophysica Acta - Molecular Cell Research*. **1496**(1), pp.3–22.

Sellers, J.R. 1999. *Myosins*. Oxford; New York: Oxford University Press.

Shaw, D.E., Chao, J.C., Eastwood, M.P., Gagliardo, J., Grossman, J.P., Ho, C.R., Lerardi, D.J., Kolossváry, I., Klepeis, J.L., Layman, T., McLeavey, C., Deneroff, M.M., Moraes, M.A., Mueller, R., Priest, E.C., Shan, Y., Spengler, J., Theobald, M., Towles, B., Wang, S.C., Dror, R.O., Kuskin, J.S., Larson, R.H., Salmon, J.K., Young, C., Batson, B. and Bowers, K.J. 2008. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM*. **51**(7), p.91.

Shiroguchi, K., Chin, H.F., Hannemann, D.E., Muneyuki, E., de la Cruz, E.M. and Kinosita, K. 2011. Direct observation of the myosin va recovery stroke that contributes to unidirectional stepping along actin. *PLoS Biology*. **9**(4).

Si, H. 2015. TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator. *ACM Transactions on Mathematical Software*. **41**(2), pp.1–36.

Sigworth, F.J. 1998. A maximum-likelihood approach to single-particle image refinement. *Journal of Structural Biology*. **122**(3), pp.328–339.

Sivaramakrishnan, S., Spink, B.J., Sim, A.Y.L., Doniach, S. and Spudich, J.A. 2008. Dynamic charge interactions create surprising rigidity in the ER/K alpha-helical protein motif. *Proceedings of the National Academy of Sciences of the United States of America*. **105**(36), pp.13356–61.

Sivaramakrishnan, S., Sung, J., Ali, M., Doniach, S., Flyvbjerg, H. and Spudich, J.A. 2009. Combining Single-Molecule Optical Trapping and Small-Angle

X-Ray Scattering Measurements to Compute the Persistence Length of a Protein ER / K a - Helix. *Biophysj*. **97**(11), pp.2993–2999.

Skau, K.I., Hoyle, R.B. and Turner, M.S. 2006. A kinetic model describing the processivity of myosin-V. *Biophysical Journal*. **91**(7), pp.2475–2489.

Smith, C.A. and Rayment, I. 1996. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. *Biophysical Journal*. **70**(4), pp.1590–1602.

Solernou, A., Hanson, B.S., Richardson, R.A., Welch, R., Read, D.J., Harlen, O.G. and Harris, S.A. 2018. Fluctuating Finite Element Analysis (FFEA): A continuum mechanics software tool for mesoscale simulation of biomolecules. *PLoS Computational Biology*. **14**(3), p.e1005897.

Soni, L.E., Warren, C.M., Bucci, C., Orten, D.J. and Hasson, T. 2005. The unconventional myosin-VIIa associates with lysosomes. *Cell Motility and the Cytoskeleton*. **62**(1), pp.13–26.

Spiliotopoulos, D., Zhu, J., Wamhoff, E.C., Deerain, N., Marchand, J.R., Aretz, J., Rademacher, C. and Caflisch, A. 2017. Virtual screen to NMR (VS2NMR): Discovery of fragment hits for the CBP bromodomain. *Bioorganic and Medicinal Chemistry Letters*. **27**(11), pp.2472–2478.

Spink, B.J., Sivaramakrishnan, S., Lipfert, J., Doniach, S. and Spudich, J.A. 2008. Long single α-helical tail domains bridge the gap between structure and function of myosin VI. *Nature Structural and Molecular Biology*. **15**(6), pp.591–597.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*. **26**(16), pp.1701–1718.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I. and Ludtke, S.J. 2007. EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology*. **157**(1), pp.38–46.

Tepass, U. 2009. FERM proteins in animal morphogenesis. *Current Opinion in Genetics and Development*. **19**(4), pp.357–367.

Terrak, M., Rebowski, G., Lu, R.C., Grabarek, Z. and Dominguez, R. 2005. Structure of the light chain-binding domain of myosin V. *Proceedings of the*

National Academy of Sciences of the United States of America. **102**(36), pp.12718–23.

Terrak, M., Wu, G., Stafford, W.F., Lu, R.C. and Dominguez, R. 2003. Two distinct myosin light chain structures are induced by specific variations within the bound IQ motifs - Functional implications. *EMBO Journal*. **22**(3), pp.362–371.

Thirumurugan, K., Sakamoto, T., Hammer, J.A., Sellers, J.R. and Knight, P.J. 2006. The cargo-binding domain regulates structure and activity of myosin 5. *Nature*. **442**(7099), pp.212–215.

Thomas, C.J., Brown, H.L., Hawes, C.R., Lee, B.Y., Min, M.K., King, L.A. and Possee, R.D. 1998. Localization of a baculovirus-induced chitinase in the insect cell endoplasmic reticulum. *Journal of virology*. **72**(12), pp.10207–12.

Tominaga, M., Kojima, H., Yokota, E., Nakamori, R., Anson, M., Shimmen, T. and Oiwa, K. 2012. Calcium-induced mechanical change in the neck domain alters the activity of plant myosin XI. *Journal of Biological Chemistry*. **287**(36), pp.30711–30718.

Trinick, J. and Elliott, A. 1979. Electron microscope studies of thick filaments from vertebrate skeletal muscle. *Journal of Molecular Biology*. **131**(1), pp.133–136.

Trinick, J.A. 1981. End-filaments: A new structural element of vertebrate skeletal muscle thick filaments. *Journal of Molecular Biology*. **151**(2), pp.309–314.

Trybus, K.M., Gushchin, M.I., Lui, H.J., Hazelwood, L., Krementsova, E.B., Volkmann, N. and Hanein, D. 2007. Effect of calcium on calmodulin bound to the IQ motifs of myosin V. *Journal of Biological Chemistry*. **282**(32), pp.23316–23325.

Trybus, K.M., Huiatt, T.W. and Lowey, S. 1982. A bent monomeric conformation of myosin from smooth muscle. *Proceedings of the National Academy of Sciences of the United States of America*. **79**(20), pp.6151–5.

Tskhovrebova, L. and Trinick, J. 1997. Direct visualization of extensibility in isolated titin molecules. *Journal of Molecular Biology*. **265**(2), pp.100–106.

Tskhovrebova, L. and Trinick, J. 2001. Flexibility and extensibility in the titin

molecule: Analysis of electron microscope data. *Journal of Molecular Biology*. **310**(4), pp.755–771.

Udovichenko, I.P., Gibbs, D. and Williams, D.S. 2002. Actin-based motor properties of native myosin VIIa. *Journal of cell science*. **115**(Pt 2), pp.445–450.

Umeki, N., Jung, H.S., Sakai, T., Sato, O., Ikebe, R. and Ikebe, M. 2011. Phospholipid-dependent regulation of the motor activity of myosin X. *Nature Structural and Molecular Biology*. **18**(7), pp.783–788.

Umeki, N., Jung, H.S., Watanabe, S., Sakai, T., Li, X., Ikebe, R., Craig, R. and Ikebe, M. 2009. The tail binds to the head-neck domain, inhibiting ATPase activity of myosin VIIA. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(21), pp.8483–8488.

Unzue, A., Xu, M., Dong, J., Wiedmer, L., Spiliotopoulos, D., Caflisch, A. and Nevado, C. 2016. Fragment-Based Design of Selective Nanomolar Ligands of the CREBBP Bromodomain. *Journal of Medicinal Chemistry*. **59**(4), pp.1350–1356.

Vaiana, S.M., Manno, M., Emanuele, A., Palma-Vittorelli, M.B. and Palma, M.U. 2001. The role of solvent in protein folding and in aggregation. *Journal of Biological Physics*. **27**(2–3), pp.133–145.

Vale, R.D. and Milligan, R.A. 2000. The way things move: Looking under the hood of molecular motor proteins. *Science*. **288**(5463), p.88.

Vilfan, A. 2005. Elastic lever-arm model for myosin V. *Biophysical Journal*. **88**(6), pp.3792–3805.

Walker, M., Knight, P. and Trinick, J. 1985. Negative staining of myosin molecules. *Journal of Molecular Biology*. **184**(3), pp.535–542.

Walker, M., Knight, P. and Trinick, J. 1991. Properties of the myosin molecule revealed by negative staining. *Micron And Microscopica Acta*. **22**(4), pp.413–422.

Walker, M. and Trinick, J. 1989. Electron microscopy of negatively stained scallop myosin molecules. Effect of regulatory light chain removal on head structure. *Journal of Molecular Biology*. **208**(3), pp.469–475.

Walker, M.L., Burgess, S.A., Sellers, J.R., Wang, F., Hammer, J.A., Trinick, J.

and Knight, P.J. 2000. Two-headed binding of a processive myosin to F-actin. *Nature*. **405**(6788), pp.804–807.

Wang, F., Thirumurugan, K., Stafford, W.F., Hammer, J.A., Knight, P.J. and Sellers, J.R. 2004. Regulated Conformation of Myosin V. *Journal of Biological Chemistry*. **279**(4), pp.2333–2336.

Wang, Y. and Zocchi, G. 2011. The folded protein as a viscoelastic solid. *Epl*. **96**(1).

Warner, C.L., Stewart, A., Luzio, J.P., Steel, K.P., Libby, R.T., Kendrick-Jones, J. and Buss, F. 2003. Loss of myosin VI reduces secretion and the size of the Golgi in fibroblasts from Snell's waltzer mice. *EMBO Journal*. **22**(3), pp.569–579.

Watanabe, S., Ikebe, R. and Ikebe, M. 2006. Drosophila myosin VIIA is a high duty ratio motor with a unique kinetic mechanism. *Journal of Biological Chemistry*. **281**(11), pp.7151–7160.

Weil, D., Kussel, P., Blanchard, S., Levy, G., Levi-Acobas, F., Drira, M., Ayadi, H. and Petit, C. 1997. The autosomal recessive isolated deafness, DFNB2, and the Usher 1B syndrome are allelic defects of the myosin-VIIA gene. *Nature Genetics*. **16**(2), pp.191–193.

Weisel, J.W. and Litvinov, R.I. 2013. Mechanisms of fibrin polymerization and clinical implications. *Blood*. **121**(10), pp.1712–1719.

Well, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levilliers, J., Weston, M.D., Kelley, P.M., Kimberling, W.J., Wagenaar, M., Levi-Acobas, F., Larget-Piet, D., Munnich, A., Steel, K.P., Brown, S.D.M. and Petit, C. 1995. Defective myosin VIIA gene responsible for Usher syndrome type IB. *Nature*. **374**(6517), pp.60–61.

Wells, A.L., Lin, A.W., Chen, L.Q., Safer, D., Cain, S.M., Hasson, T., Carragher, B.O., Milligan, R.A. and Sweeney, H.L. 1999. Myosin VI is an actin-based motor that moves backwards. *Nature*. **401**(6752), pp.505–508.

Wendt, T., Taylor, D., Messier, T., Trybus, K.M. and Taylor, K.A. 1999. Visualization of head-head interactions in the inhibited state of smooth muscle myosin. *Journal of Cell Biology*. **147**(7), pp.1385–1389.

Wolfrum, U. and Schmitt, A. 2000. Rhodopsin transport in the membrane of the

connecting cilium of mammalian photoreceptor cells. *Cell Motility and the Cytoskeleton*. **46**(2), pp.95–107.

Wolgemuth, C.W. and Sun, S.X. 2006. Elasticity of α-helical coiled coils. *Physical Review Letters*. **97**(24).

Wolny, M., Batchelor, M., Bartlett, G.J., Baker, E.G., Kurzawa, M., Knight, P.J., Dougan, L., Woolfson, D.N., Paci, E. and Peckham, M. 2017. Characterization of long and stable de novo single alpha-helix domains provides novel insight into their stability. *Scientific Reports*. **7**.

Wolny, M., Batchelor, M., Knight, P.J., Paci, E., Dougan, L. and Peckham, M. 2014. Stable single ??-Helices are constant force springs in proteins. *Journal of Biological Chemistry*. **289**(40), pp.27825–27835.

Wu, L., Pan, L., Wei, Z. and Zhang, M. 2011. Structure of myTH4-FERM domains in myosin VIIa tail bound to cargo. *Science*. **331**(6018), pp.757–760.

Yang, Y., Baboolal, T.G., Siththanandan, V., Chen, M., Walker, M.L., Knight, P.J., Peckham, M. and Sellers, J.R. 2009. A FERM domain autoregulates Drosophila myosin 7a activity. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(11), pp.4189–4194.

Yang, Y., Kovács, M., Xu, Q., Anderson, J.B. and Sellers, J.R. 2005. Myosin VIIB from Drosophila is a high duty ratio motor. *Journal of Biological Chemistry*. **280**(37), pp.32061–32068.

Yu, I.M., Planelles-Herrero, V.J., Sourigues, Y., Moussaoui, D., Sirkia, H., Kikuti, C., Stroebel, D., Titus, M.A. and Houdusse, A. 2017. Myosin 7 and its adaptors link cadherins to actin. *Nature Communications*. **8**, p.15864.

Zhang, M., Tanaka, T. and Ikura, M. 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nature Structural Biology*. **2**(9), pp.758–767.

Zhao, F.Q. and Craig, R. 2003. Capturing time-resolved changes in molecular structure by negative staining. *Journal of Structural Biology*. **141**(1), pp.43–52.

Zhu, J. and Caflisch, A. 2016. Twenty Crystal Structures of Bromodomain and PHD Finger Containing Protein 1 (BRPF1)/Ligand Complexes Reveal

Conserved Binding Motifs and Rare Interactions. *Journal of Medicinal Chemistry*. **59**(11), pp.5555–5561.

Zimmerman, S.B. and Trach, S.O. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of Molecular Biology*. **222**(3), pp.599–620.

8. Appendix.

MD parameter files

Amber – scripts used to run implicitly simulated molecules

Minimisation

```
#$ -cwd -V
#$ -l h_rt=03:00:00
#$ -l nodes=2
module add amber

mpirun sander.MPI -O -i IS_min.in -o IS_min.out -c ../path/to/*.crd -p ../path/to/*.top -r *_min.rst
```

IS_min.in

```
Minimization to relax initial bad contacts, implicit solvent
 &cntrl
   imin=1,
   ncyc=2500,
   maxcyc=5000,
   ntpr=100,
   ntb=0,
   cut=10.0,
   igb=5,
   ntr=0,
 /
```

Equilibration

```
#$ -cwd -V
#$ -l h_rt=24:00:00
#$ -l nodes=4
module add amber

mpirun sander.MPI -O -i IS_equi.in -o IS_equi.out -p ../path/to/*.top -c ../path/to/*_min.rst -r *_equi.rst -x
equi.mdcrd -ref ../path/to/min.rst
```

IS_equi.in

```
Implicit solvent initial heating mdin
 &cntrl
   imin=0, irest=0, ntx=1,
   ntpr=500, ntwx=500, nstlim=n,
   dt=0.002, ntt=3, tempi=0,
   temp0=300, gamma_ln=3.0, ig=-1,                    #temp0 done in stages
   ntp=0, ntc=2, ntf=2, cut=10.0,
   ntb=0, igb=5, ioutfm=1,
 /
```

Production

```
#$ -cwd -V
#$ -l h_rt=48:00:00
#$ -l nodes=n
module add amber

mpirun sander.MPI -O -i IS_MD.in -o IS_MD.out -p ../path/to/*.top -c ../path/to/final/*_equi.rst -r *_prod.rst -
x *_prod.mdcrd
```

IS_MD.in

```
Implicit solvent molecular dynamics
 &cntrl
       ntf=2, ntc=2, ntb=0, cut=10.0,
       nstlim=n, dt=0.002,
       tempi=300.0, temp0=300.0, ntt=3,
```

```
    gamma_ln=5, ig=-1,
    ntwr=1000,
    imin=0, irest=1, ntx=5,
    igb=5, gbsa=1, saltcon=n,
    ntpr=1000, ntwx=1000, ioutfm=1,
 &end
```

Gromacs – scripts used to run explicitly simulated molecules

Minimisation

```
#!/bin/sh
#$ -cwd -V
#$ -l h_rt=24:00:00
#$ -l nodes=4
module add gromacs


gmx grompp -f Min.mdp -c ../Add_ions/ions.tpr -p ../path/to/SAH.top -o em_min.tpr

gmx mdrun -v -deffnm em_min.tpr
```

Min.mdp

```
# minim.mdp - used as input into grompp to generate em.tpr
integrator = steep              # Algorithm (steep = steepest descent minimization)
emtol = 1000.0                  # Stop minimization when the maximum force < 1000.0 kJ/mol/nm
emstep = 0.01                   # Energy step sizeimization when the maximum force < 100
nsteps = 0.= 50000              # Energy; Maximum number of (minimization) steps to perform
nsteps = 0.= 50000              # Energy; Maximum number of (minimization) steps to perform
# Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlisteters describ= 1         # Frequency to update the neighbor list and long range forces
cutoff-scheme = Verlet          # Frequency to update the neighbor list and long range forces
ns_typeeters describ= gridw     # Method to determine neighbor list (simple, grid)
coulombtypeme  = Ve= PME        #Treatment of long range electrostatic interactions
rcoulombcheme  = Ve= 10.0d      # Fr; Short-range electrostatic cut-off (simple, grid)
rvdwombtypeforces  = 10.0d      # Short-range Van der Waals cut-offatic interactions
pbculombypeme  = Ve= 1.= xyz    # Periodic Boundary Conditions (yes/no) interactions
```

Equilibration

```
#!/bin/sh
#$ -cwd -V
#$ -l h_rt=48:00:00
#$ -l nodes=16
module add gromacs

gmx grompp -f NVT.mdp -c ../path/to/em_min.gro -p ../path/to/SAH.top -o NVT.tpr
gmx mdrun -v -deffnm NVT.tpr

gmx grompp -f NPT.mdp -c NVT.gro -t NVT.cpt -p ../path/to/SAH.top -o NPT.tpr
gmx mdrun -v -deffnm NPT.tpr
```

NVT.mdp

```
title = explicit SAH domain NVT equilibration
define = -DPOSRES               # position restrain the protein
# Run parameters
integrator = md                 # leap-frog integrator
nsteps = 50000                  # 2 * 50000 = 100 ps
dt = 0.002                      # 2 fs
# Output control
nstxout = 500                   # save coordinates every 1.0 ps
nstvout = 500                   # save velocities every 1.0 ps
nstenergy = 500                 # save energies every 1.0 ps
nstlog = 500                    # update log file every 1.0 ps
# Bond parameters
Continuation = no               # first dynamics run
constraint_algorithm = lincs    # holonomic constraints
constraints = all-bonds         # all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1                  # accuracy of LINCS
```

```
lincs_order = 4                 # also related to accuracy
# Neighborsearching
cutoff-scheme = Verlet
ns_type = grid                  # search neighboring grid cells
nstlist = 10                    # 20 fs, largely irrelevant with Verlet
rcoulomb = 1.0                  # short-range electrostatic cutoff (in nm)
# Electrostatics
coulombtype = PME               # Particle Mesh Ewald for long-range electrostatics
pme_order = 4                   # cubic interpolation
fourierspacing = 0.16           # grid spacing for FFT
# Temperature coupling is on
tcoupl = V-rescale              # modified Berendsen thermostat
tc-grps = Protein Non-Protein   # two coupling groups - more accurate
tau_t = 0.1     0.1             # time constant, in ps
ref_t = 300     300             # reference temperature, one for each group, in K
# Pressure coupling is off
pcoupl = no                     # no pressure coupling in NVT
# Periodic boundary conditions
pbc = xyz                       # 3-D PBC
# Dispersion correction
DispCorr = EnerPres             # account for cut-off vdW scheme
# Velocity generation
gen_vel = yes                   # assign velocities from Maxwell distribution

gen_temp = 300                  # temperature for Maxwell distribution

gen_seed = -1                   # generate a random seed
```

NPT.mdp

```
title = explicit SAH domain NPT equilibration
define = -DPOSRES               # position restrain the protein
# Run parameters
integrator = md                 # leap-frog integrator
nsteps = 50000                  # 2 * 50000 = 100 ps
dt = 0.002                      # 2 fs
# Output control
nstxout = 500                   # save coordinates every 1.0 ps
nstvout = 500                   # save velocities every 1.0 ps
nstenergy = 500                 # save energies every 1.0 ps
nstlog = 500                    # update log file every 1.0 ps
# Bond parameters
continuation = yes              # Restarting after NVT
constraint_algorithm = lincs    # holonomic constraints
constraints = all-bonds         # all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1                  # accuracy of LINCS
lincs_order = 4                 # also related to accuracy
# Neighborsearching
cutoff-scheme = Verlet
ns_type = grid                  # search neighboring grid cells
nstlist = 10                    # 20 fs, largely irrelevant with Verlet scheme
rcoulomb = 1.0                  # short-range electrostatic cutoff (in nm)
rvdw = 1.0                      # short-range van der Waals cutoff (in nm)
# Electrostatics
coulombtype = PME               # Particle Mesh Ewald for long-range electrostatics
pme_order = 4                   # cubic interpolation
fourierspacing = 0.16           # grid spacing for FFT
# Temperature coupling is on
tcoupl = V-rescale              # modified Berendsen thermostat
tc-grps = Protein Non-Protein   # two coupling groups - more accurate
tau_t = 0.1     0.1             # time constant, in ps
ref_t = 300     300             # reference temperature, one for each group, in K
# Pressure coupling is on
pcoupl = Parrinello-Rahman      # Pressure coupling on in NPT
pcoupltype = isotropic          # uniform scaling of box vectors
tau_p = 2.0                     # time constant, in ps
ref_p = 1.0                     # reference pressure, in bar
compressibility = 4.5e-5        # isothermal compressibility of water, bar^-1
refcoord_scaling = com
3 Periodic boundary conditions
```

```
pbc = xyz                       # 3-D PBC
# Dispersion correction
DispCorr = EnerPres             # account for cut-off vdW scheme
# Velocity generation
gen_vel = no                    # Velocity generation is off
```

Production

```
#!/bin/sh
#$ -cwd -V
#$ -l h_rt=48:00:00
#$ -l nodes=16
module add gromacs

gmx grompp -f MD.mdp -c ../path/to/NPT.gro -p ../path/to/*.top -o PROD.tpr
gmx mdrun -v -deffnm PROD.tpr
```

MD.mdp

```
title = explicit SAH domain MD simulation
# Run parameters
integrator = md                 # leap-frog integrator
nsteps = 250000000              # 2 * 250000000 = 500000 ps (500 ns)
dt = 0.002                      # 2 fs
# Output control
nstxout = 5000                  # save coordinates every 10.0 ps
nstvout = 5000                  # save velocities every 10.0 ps
nstenergy = 5000                # save energies every 10.0 ps
nstlog = 5000                   # update log file every 10.0 ps
nstxout-compressed = 5000       # save compressed coordinates every 10.0 ps
                                # nstxout-compressed replaces nstxtcout
compressed-x-grps = System      # replaces xtc-grps
# Bond parameters
continuation = yes              # Restarting after NPT
constraint_algorithm = lincs    # holonomic constraints
constraints = all-bonds         # all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1                  # accuracy of LINCS
lincs_order = 4                 # also related to accuracy
# Neighborsearching
cutoff-scheme = Verlet
ns_type = grid                  # search neighboring grid cells
nstlist = 10                    # 20 fs, largely irrelevant with Verlet scheme
rcoulomb = 1.0                  # short-range electrostatic cutoff (in nm)
rvdw = 1.0                      # short-range van der Waals cutoff (in nm)
# Electrostatics
coulombtype = PME               # Particle Mesh Ewald for long-range electrostatics
pme_order = 4                   # cubic interpolation
fourierspacing = 0.16           # grid spacing for FFT
# Temperature coupling is on
tcoupl = V-rescale              # modified Berendsen thermostat
tc-grps = Protein Non-Protein   # two coupling groups - more accurate
tau_t = 0.1     0.1             # time constant, in ps
ref_t = 300     300             # reference temperature, one for each group, in K
# Pressure coupling is on
pcoupl = Parrinello-Rahman      # Pressure coupling on in NPT
pcoupltype = isotropic          # uniform scaling of box vectors
tau_p = 2.0                     # time constant, in ps
ref_p = 1.0                     # reference pressure, in bar
compressibility = 4.5e-5        # isothermal compressibility of water, bar^-1
# Periodic boundary conditions
pbc = xyz                       # 3-D PBC
# Dispersion correction
DispCorr = EnerPres             # account for cut-off vdW scheme
# Velocity generation
gen_vel = no                    # Velocity generation is off
```