University of Kent

Department of Biosciences

The Analysis of Myosin II Evolution in Mammals

Michael Coghlan

2015

Supervisors: Prof. Michael Geeves, Dr. Mark Wass and Dr. Anthony Baines

Table of Contents

Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Michael Coghlan

August 2015

Acknowledgements

Firstly, I would like to express my sincerest gratitude to my supervisors Mike, Mark and Anthony for their continued support and encouragement over the year. Balancing part-time work alongside this degree has not been easy and it has seen many frustrating moments, however never once has it been a problem. My thanks go out to members of the Geeves and Mulvihill lab for their constant support, encouragement and kind words; and all other members of the MaDCaP group for listening to my various presentations over the year and fully involving me in all aspects of life as a scientist. I say thanks to the Wass lab for providing continuous support throughout the year in the office and offering their time to listen to my quarrels over the year.

I would like to acknowledge the support of my family and friends over the course of the year and for always believing in me when I had doubts.

Lastly, I would like to say thank you to Sarah Jeanfavre, who truly paved the way for this amazing project, who it has been an absolute pleasure to work alongside and who has become a life long friend.

Abstract

Myosin-II is a family of myosin that contains fifteen different isoforms that vary in function, from maintaining a critical role in sarcomeric contractions to non-muscle movement. Myosin-IIs structure can be divided in to two key domains, the motor domain and the tail domain. It is the motor domain that contains the proteins catalytic activity. The observed phenomenon of an inverse relationship between organism size and heart rate tells us that heart rates do vary with size. The larger the animal mass, the slower the heart rate. This phenomenon allowed us to propose the idea that variations in the sequence of cardiac isoforms could be observed as mass of the species increases, due to the variance of heart rates. If so, any divergence seen in the sequence of the β-cardiac isoform would be reflected as species mass increases should not be seen in other isoforms that do not rely on heart rate variability. Through analysing the sequences of twelve mammalian species of varying mass, a strong divergence relationship was observed in the β-cardiac motor domain with a correlation coefficient of -0.945 that was no observed in other isoforms. Key results show that myosin sequence divergence does have a marked dependence on both evolutionary separation and size difference between species.

Abbreviations

ATP	Adenosine Triphosphate
ADP	Adenosine Diphosphate
Pi	Inorganic Phosphate
S1	Subfragment 1
S2	Subfragment 2
MYH1	Skeletal 2D/X myosin isoform
MYH2	Skeletal 2A myosin isoform
MYH3	Embryonic myosin isoform
MYH4	Skeletal 2B myosin isoform
MYH6	α-Cardiac myosin isoform
MYH7	β-Cardiac myosin isoform
MYH8	Perinatal myosin isoform
MYH9	Non-muscle A myosin isoform
MYH10	Non-muscle B myosin isoform
MYH11	Smooth muscle myosin isoform
MYH13	Extraocular myosin isoform
MYH14	Non-muscle C myosin isoform
MYH15	Extraocular-2 myosin isoform
MYH16	Jaw myosin isoform
ACT	Artemis Comparison Tool
SNAP	Synonymous and Non-synonymous Analysis Programme
R	Correlation coefficient

Introduction

The heart rate versus size phenomenon is one that is relatively accepted within the scientific community. That is, as the body mass of an organism increases, the heart rate decreases (Pellegrino, M., Canepari, M., Bottinelli, R. 2002). Knowing that there are many significant factors that contribute to the metabolism and heart rate of an organism (Savage, V., Allen, A., Brown, J. *et al*. 2007), relatively little is understood on the role myosin has in this relationship. Myosin is a key functional component of muscle and non-muscle movement in the body, with isoforms involved in processes from sarcomeric contractions to movements of cells. Mutations in these protein sequences have generated a large interest over recent years due to the increasing burden of cardiovascular diseases on economies both on a national and international scale. There have been hundreds of mutations identified that affect a number of different genes all affecting the proteins in cardiomyocytes (Moore, J., Leinwand, L., Warshaw, D. 2014). These mutations alter different kinetic properties of myosin that result in a number of different cardiomyopathies, namely hypertrophic cardiomyopathy and dilated cardiomyopathy.

At present, there are more than 300 mutations identified in the motor domain of the cardiac myosin protein with little understanding of how the mutations cause a disease phenotype. A possible way of understanding how mutations may affect kinetic properties of the protein would be through conducting an analysis of sequences and identifying points of variation that could possibly lead to either gain or loss of function. Any mutations found may lead to development of detection systems where there is a greater risk of cardiomyopathy in patients. As evolution occurs, different pressures are placed upon proteins in order to adapt and optimise processes for individual organisms, allowing for survival. Taking in to consideration the heart rate versus size phenomenon and the evolution of the protein, identifying residues that vary through a range of organisms may aid scientific understanding of this approach.

This introduction will introduce the muscle types found within the body, and their cellular organisations. The structures within these cells will be introduced, and the myosin motor protein's

structure and activity will be introduced to give an understanding of its importance in physiology. The varying different isoforms of myosin will be discussed, with focus on the cardiac isoforms. The importance of the heart rate versus size phenomenon will then be introduced to give context to the introduction, and the importance of DNA will be discussed to show how the proteins structure is affected from the genotype to the phenotype.

1.1 Muscle Types and Structure

The human body utilises two main types of muscle, smooth and striated muscle, each innervated in different ways. Smooth muscle is mainly involved in maintaining homeostasis of the body, and lines the vascular system and gastrointestinal tract. Smooth muscle varies in function so much it is difficult to classify, and is often categorised by: location, contraction pattern and the muscles communication with nearby cells (Silverthorn, D. 2012). This muscle type is not the focus of this project.

Striated muscle, aptly named due to the banding in the muscle as seen under the microscope, is involved more so with voluntary movements of the sympathetic nervous system. Of the striated muscle types, skeletal muscle differs from cardiac muscle in its appearance due to the cardiac isoforms presence of intercalated disks. These intercalated disks allow for transduction of electrical current across the cardiomyocytes allowing for more efficient, unanimous contractions (Silverthorn, D. 2012). Skeletal muscles are syncytia, meaning they are multinucleate from cell fusion and cardiac muscle is uninucleate (Alberts, B., Johnson, A., Lewis, J., *et al.* 2002).

A single muscle consists of many muscle fibres that are bunched together in what is known as a muscle fascicle. Each fibre contains a number of myofibrils that are surrounded by sarcoplasmic reticulum's, modified endoplasmic reticulums that act as calcium stores for muscle contraction (Silverthorn, D. 2012). A myofibril contains the sarcomere, as detailed in figure 1. This banding occurs from thick and thin filaments of the muscle unit. Thin filaments consist of actin, a long chain formed from individual 42 kDa globular actin units. Two of these chains twist together in skeletal muscle to create the thin filaments of the myofibril. Actin acts as the substrate from which the

active thick filament is able to pull on. The thick filament consists of conventional muscle myosin II, an ATP-dependent motor protein that creates movement. These two proteins make up the majority of each sarcomere, where 60%-70% of the cell is myosin and 20%-25% is actin (Devlin, T. 2011). These structures are shown in figure 1.2 A and 1.2 B.

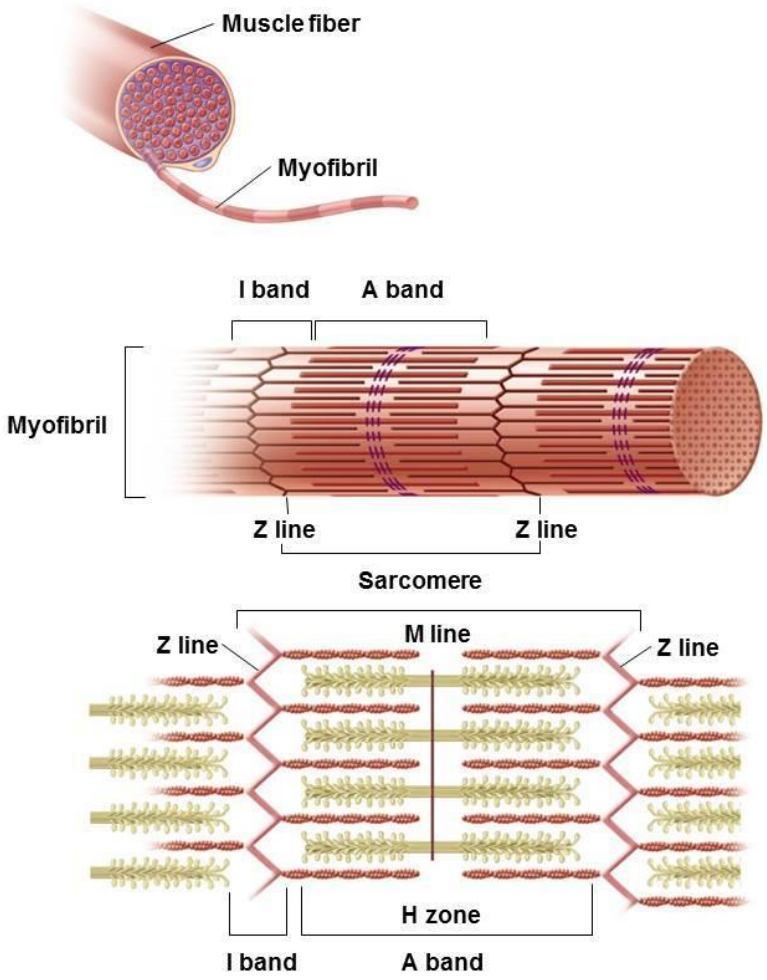


Figure 1.1 The muscle and its organisation. The muscle fibre is depicted, with a number of myofibrils within it. This myofibril is banded due to the structure of the muscle unit. An individual sarcomere is defined in between two Z lines, and are zigzag protein structures that form attachment sites for thin filaments. The I band spanning these Z lines are areas only occupied by thin filaments of the sarcomere, giving it a lighter appearance. The A band is the darkest of the bands and entails the thick filament, where overlapping occurs with the thin filament at the edges. The centre of the H zone is lighter in appearance than the rest due to it only being occupied by the heavy chain. The M line forms the attachment site for the thick filaments, and divides each A band in half. Adapted from *University of Wisconsin – Madison - Physiology*

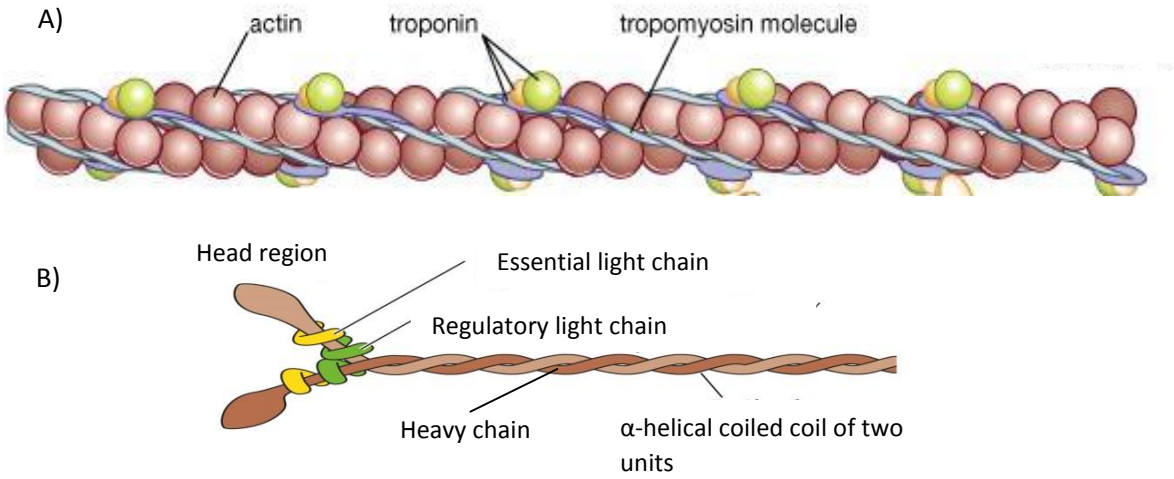


Figure 1.2 A The thin filament of the sarcomere. Single globular actin units, represented by the pink spheres, polymerise with each other in order to form filamentous actin. Each globular actin unit has a myosin binding site for the globular protein to bind to. Troponin and tropomyosin shown are regulatory units of the actin molecule, regulating the binding of the myosin heads.
1.2 B The thick filament of the sarcomere,myosin. The dimerised protein has a long tail region in an α-helical coiled coil formation leading to two globular heads. Each individual chain is about 230 kDa. The neck region of the protein contains the essential light chain and regulatory light chain, both involved in aiding the protein in its rotational movements (Borejdo, J., Ushakov, D., Akopova, I. 2002) and are around 20 kDa each (Devlin, T. 2011). Figures adapted from *Muscle: Actin and Myosin* and *What is Myosin?*.

When a sarcomere is activated, actin filaments are pulled over myosin filaments in a proposed sliding filament theory that results in the shortening of the I band and the H zone (Huxley, A., Niedergerke, R. 1954). Alone, the contraction of a sarcomere does not generate enough force to move a limb or organ, however the conjoined effect of all muscle sarcomeres in many different fascicles generates enough force to move. The three-dimensional organisation of the myofibrils in muscle fibres ensure that maximum potential is achieved upon contraction.

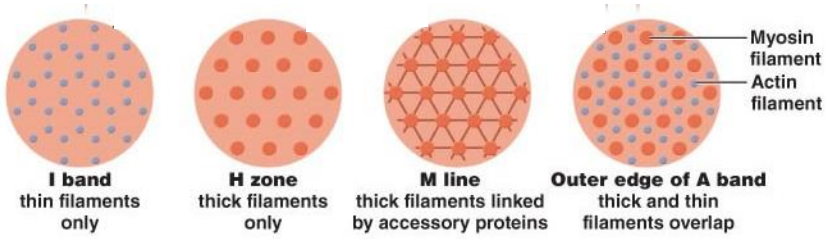


Figure 1.3 The 3D organisation of thick and thin filaments. Diagrams show cross sections of the sarcomere taken at different points, as shown. The arrangement of one thick filament for six thin filaments allows for efficient use of all myosin motor domains in the cylindrical thick filament. Adapted from *Human Physiology Slide Player*.

Actin is highly conserved between species (Devlin, T. 2011) suggesting that any differences in contraction rates are due to changes in the catalytic activity of the myosin protein. Different isoforms of conventional myosin have different contraction rates or shortening velocities due to their differing specialisations. Slow type myosins have been shown to be more common in larger species in order to adapt to differing pressures and physiologies (Pellegrino, M., Canepari, M., Bottinelli, R. 2003).

1.2 The cross-bridge cycle

The sliding filament theory introduces the kinetic properties of individual myosin and actin dynamics that the cross-bridge cycle goes on to define. This cycle has been conserved throughout all the different classes of myosin, each involved in individual cellular and physiological processes. The culmination of all of these individual interactions in a single movement generates the forces required for movement. At any single point, where the thick and thin filaments overlap in the H zone, only about 50% of the cross bridges are attached, therefore only generating 50% of potential force (Newsholme, E., Leech, T. 2009). The stages of the cycle are shown in figure 1.4.

Each individual myosin globular motor domain undergoes the cross bridge cycle independently of others, it is not a synchronous contraction. Due to the differing activation of each individual myosin protein, delayed dissociation of individual monomers can have an adverse effect. The amount of time the myosin is attached to the actin filament is known as the duty ratio. This duty ratio varies between myosins, where low-duty-ratio types will spend more time in the lower energy state Myosin.ATP and Myosin.ADP.Pi states. Higher duty-ratio types will spend more time occupying the Actin.Myosin and Actin.Myosin.ADP states (De La Cruz, E., Ostap, E. 2004). The conventional myosin class that is focussed on within this project and all of the isoforms within it generate the same force, even in their varied roles. It is the kinetic properties that differ (De La Cruz, E., Ostap, E. 2004).

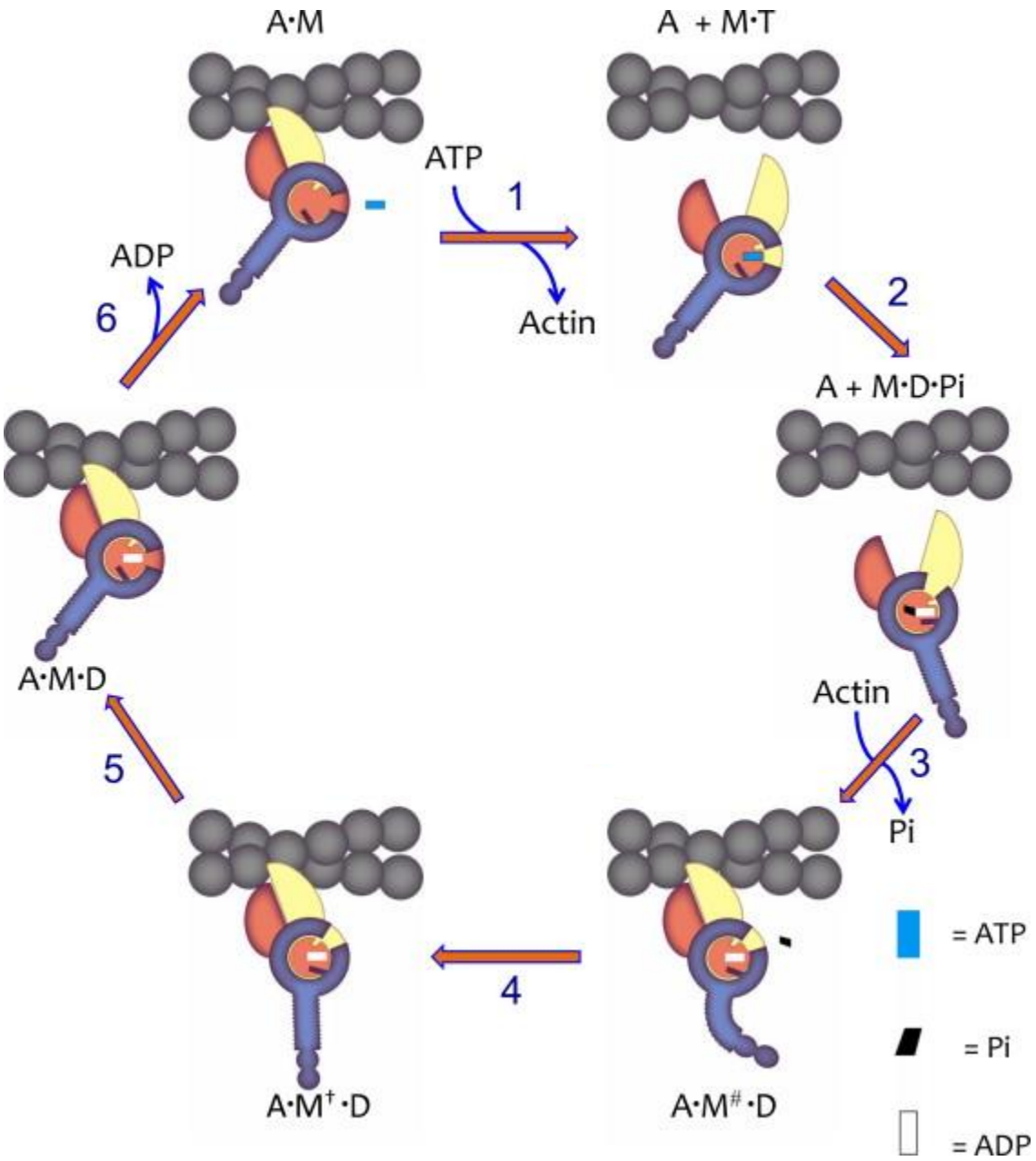


Figure 1.4 The cross-bridge cycle (adapted from *Geeves and Bloemink, 2011*). The actin filament is shown in black circles and the myosin protein is shown in three parts, red yellow and blue, with the central filled circle being the core of the myosin head that contains the ATPase activity. The upper 50K domain is shown in yellow with switch 1 in the centre. The blue unit represents the relay loop with switch 2 in the centre and the lever arm protruding outwards. The red unit is shown for a fixed reference point throughout the cycle. In the first step (1), ATP binds to myosin in the nucleotide pocket and it dissociates from the actin filament. The binding of ATP causes rotation of the upper 50K domain closes switch 1 onto the ATP, causing a conformational change that opens the actin binding site, leading to the dissociation of myosin from the filament. (2) The hydrolysis of ATP and the recovery stroke. The relay loop rotates to bring switch 2 in to contact with ATP, in turn rotating the converter domain to complete the process, cocking the myosin in position. Once both the switches are closed, ATP is hydrolysed to form the stable myosin.ADP.Pi complex. (3) On actin rebinding, the power stroke is initiated. After hydrolysis of ATP in to ADP the lower 50K domain rebinds actin, the cleft closes and both Upper and Lower 50K domains bind actin. This rotation of the subunits causes a conformational change in the ATP binding site, allowing release of Pi and furthermore, rotation of the lever arm, generating force. The exact ordering of this process remains under debate. (4) The force generated from (3) allows the protein to move the load 5-10 nm from the power stroke. (5) Once the force generated has been relieved through movement, the lever arm rotates further to open the ATP binding site and allows ADP to dissociate. The process is then free to continue once again (*Geeves and Bloemink, 2011*).

1.3 Myosin-II

The superfamily of myosin contains a number of different groups of proteins that are all essential for different processes in the body. They are an ATP-driven actin-based motor protein. Conventional muscle myosin, or myosin-II, will be the main group of proteins focussed on within this project.

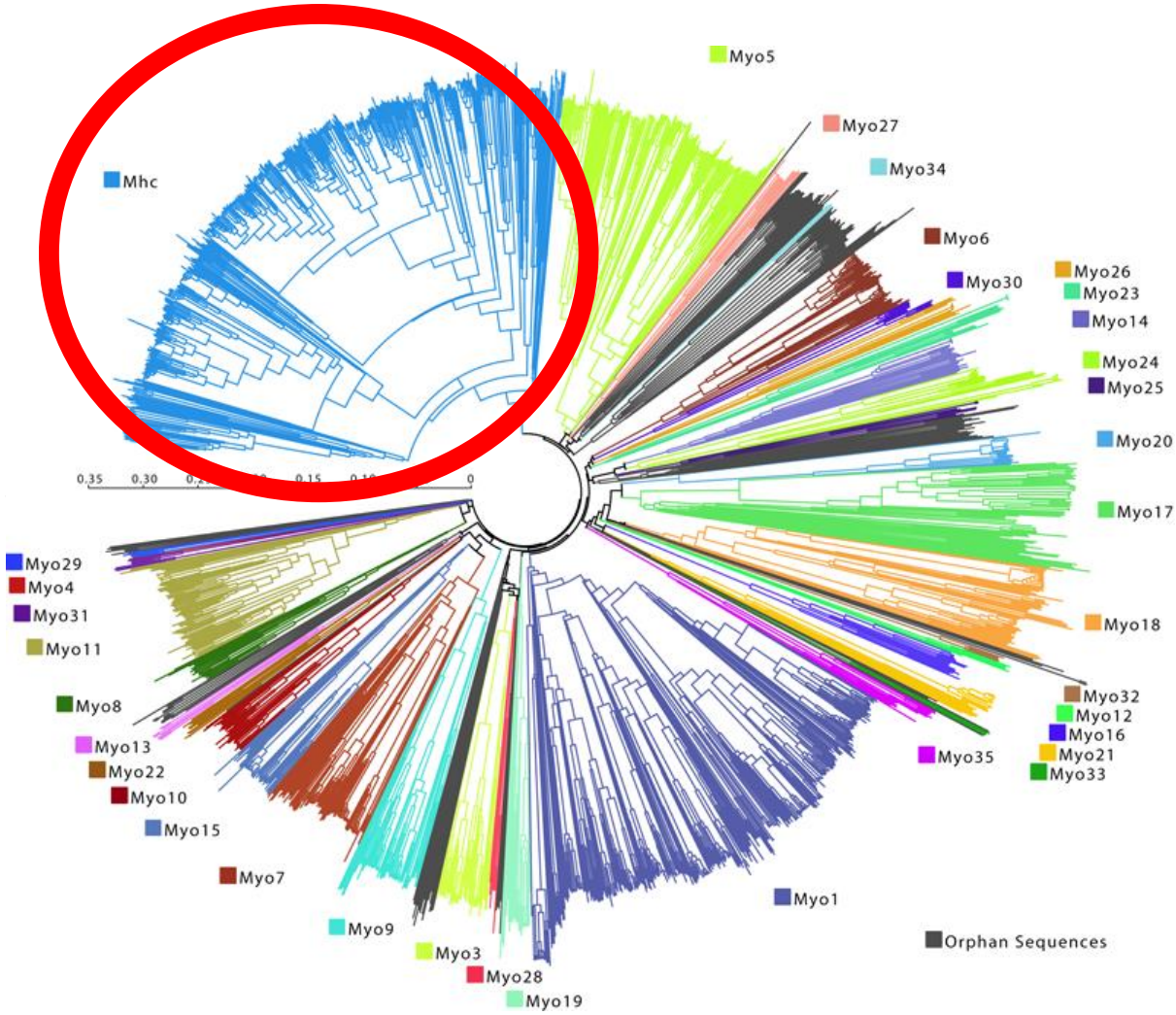


Figure 1.5 The 35 classes of myosin that have been identified. The group of conventional myosins studied in this project highlighted with a red circle. Adapted from *Myosin family analysis*.

The myosin-II protein is a dimer of two heavy chains. These chains consist of two, two-stranded α-helical coiled coil tail regions that lead to two globular motor domains adjacent from one another. The structure is shown in figure 1.6. It has previously been shown that these two domains have co-evolved (Korn, 2000), suggesting both domains are important for function of the protein.

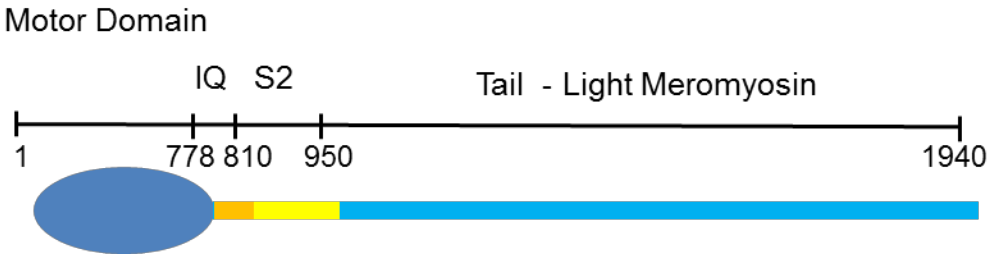


Figure 1.6 The structure of a single dimer of myosin. The domains are marked on the protein, with the blue circle indicating the motor domain from residues 1-778, the IQ domain indicated in orange from 778-810, the S2 hinge region indicated in yellow from 810-950, and the tail region in blue from 950-1940. Not to scale.

In sarcomeres, these tail regions aggregate together to form a symmetrical cylindrical thick filament that aid the typical striated filament seen under a microscope (Chew, M., Squire, J. 1995). The heptad repeat of the myosin protein, ubiquitous to all α-helices, contain the *a-b-c-d-e-f-g* conformation, where residues at the *a* and *d* positions are hydrophobic.

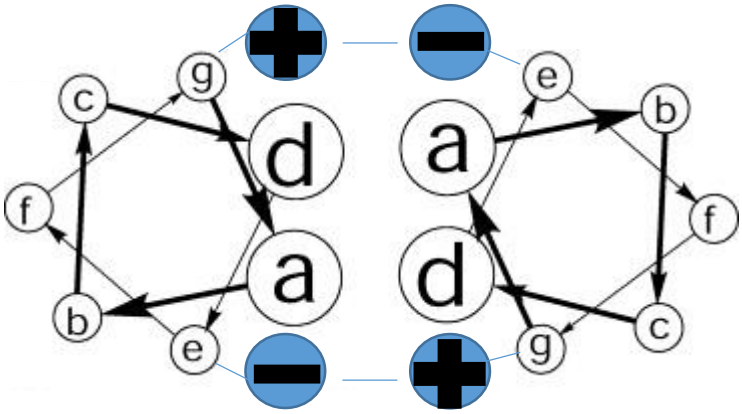


Figure 1.7 The heptad repeat of the myosin tail regions. Residues *a* and *d* are hydrophobic, allowing for dimersation of the two tail regions. The blue circles show how the charged *g* and *e* residues interact with eachother. Adapted from Hicks, M., Holberton, D., Kowalczyk, C., Woolfson, D. 1997

This hydrophobic stripe in the tail domains of the myosin allow for precipitation of the proteins at physiological salt concentrations (Hicks, M., Holberton, D., Kowalczyk, C., Woolfson, D. 1997). It is the tail domains that determine the cellular localisation and function of the myosin, i.e. whether the protein should form filaments or bind certain cargo (Coluccio, L. Holmes, K. (2007).

The motor domain of the protein contains a number of different structural and functional sites that are important for its activity. The motor domain is the functional domain of the protein where all catalytic activity occurs. In order to perform experiments at physiological salt concentrations, this

domain is often cleaved off the tail domain in to what is known as the S1 fragment, allowing the protein to dissolve and not precipitate due to the tail regions. The motor domain structure is shown in figure 1.8, indicating various sites and features of the protein.

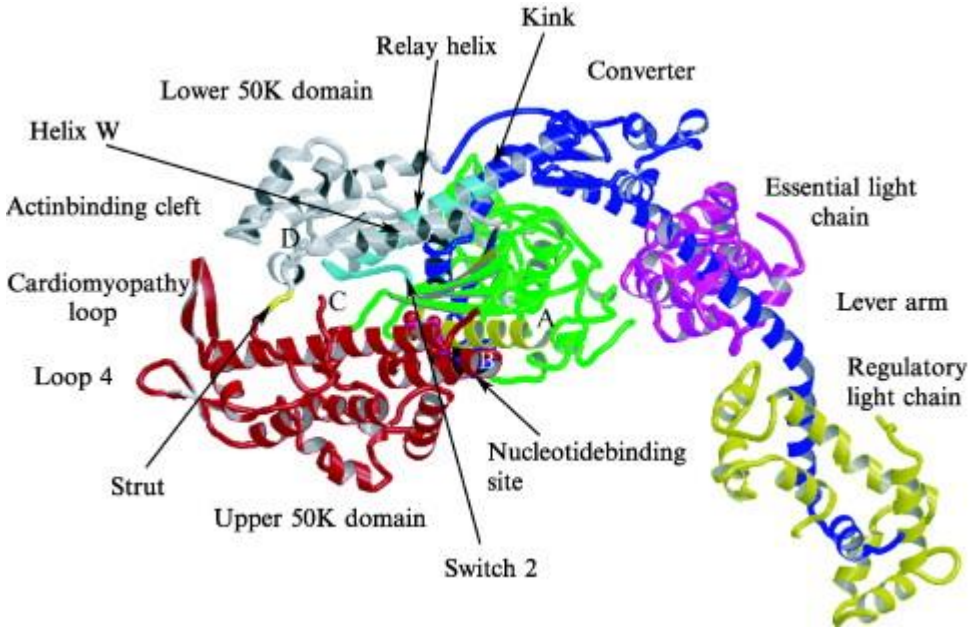


Figure 1.8 The ribbon diagram of the main sites in the myosin motor domain. Adapted from Holmes, K. and Geeves, M. 2005.

The ATPase activity of the S1 unit is found at the nucleotide binding site, known as the P-loop. The sequence is identified through the sequence motif GXXXXGK(T/S) where X indicates any amino acid (Ramakrishnan, C., Dani, V., Ramasarma, T. 2001). The important actin-binding site has a cleft about 13 Å by 13 Å (Devlin, T. 2011), which is able to open and close in response to the binding and dephosphorylation of ATP in to ADP. This dephosphorylation and release of energy also affects the converter domain that is able to rotate around a 60° angle to allow for the power stroke of the cross-bridge cycle as it binds the "lever arm" of the protein. It is this lever arm that binds the regulatory light chains of the protein. S1 can further be divided in to domains that are involved with the proteolytic activity of the protein. The 25K N-terminal, 50K and 20K C-terminal are the proteolytic fragments of S1, named after their molecular weights. The 50K domain is broken up in to the upper and lower domains, as shown in figure 1.9.

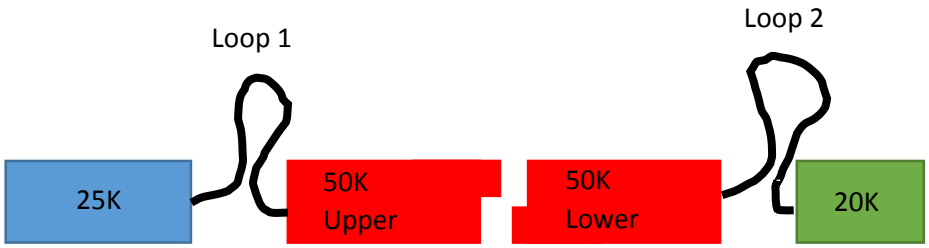


Figure 1.9 The 25K, 50K and 20K domains of the motor domain of myosin and the loop structures connecting them. The diagram also shows the cleft that distances the 50K upper and lower regions. Adapted from *Coluccio, L. Myosins, a superfamily of molecular motors*.

This cleft extends from the nucleotide binding site to the actin binding site, outlined in figure 1.8. The lower 50K subunit forms a critical part of the actin binding site, and the upper 50K subfragment forms part of the central β-sheet complex. The 20K domain contains the SH1 and SH2 helices that contain reactive thiol groups.

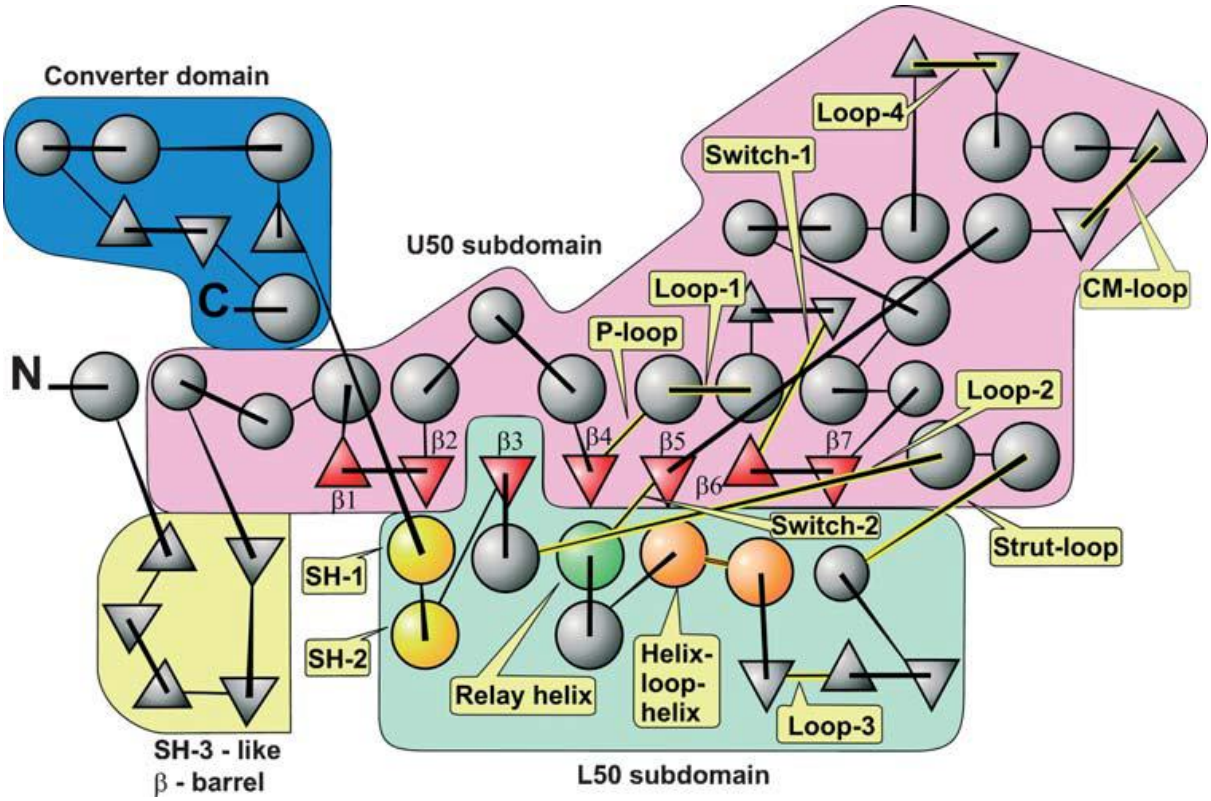


Figure 1.10 The roadkill map of myosin motor domain, adapted from Geeves, M. 2005. β-sheets are shown in triangles and α-helices are shown in circles. The upper 50K domain is highlighted in pink, the lower 50K domain shown in green, the converter domain in blue and the SH3-β-barrell complex is shown in yellow. The seven stranded β-sheet is shown in the red triangles in the middle.

The ATP-binding site is located within the large β-sheet complex in the middle of the protein, where switch 1 and 2 play critical roles in the catalytic activity. These two switches contact the highly

unstable ɣ-phosphate at the end of the ATP molecule in the nucleotide binding pocket and aid in its hydrolysis. The converter domain is the joint for the C-terminal neck domain that communicates with the active site. Loops 1 and 2 are vital for the function of myosin and are variable between the different myosin classes (Bobkov, A., *et al.* 1996). Variable loop 2 has been shown to affect a number of different product release rates, which may indicate it is a key structure to focus on when considering contraction velocity rates in consideration with organism mass. The cardiomyopathy loop is so called due to its important role for normal myosin activity. Mutations here are readily associated with cardiomyopathies, such as the well-studied R403C mutation in the β-cardiac sequence.

1.4 Isoforms

An isoform is loosely defined as a different form of the same protein (Brett, D., Pospisil, H., Bork, P. *et al*. 2002). These different isoforms of myosin-II have adapted specialised functions that are all independent from one another. Each one of these isoforms are associated with myosinopathies when they contain pathological mutations (Tajsharghi, H., Oldfors, A. 2013), further supporting how important each isoform is in its specialised function. The various isoforms of myosin-II are listed in table 1.1.

Gene name	Heavy Chain	Isoform	Short name
MYH1	MyHC-1	Skeletal 2B	2B
MYH2	MyHC-2	Skeletal 2A	2A
MYH3	MyHC-3	Embryonic	EMB
MYH4	MyHC-4	Skeletal 2D/X	2D
MYH6	MyHC-6	α cardiac	α
MYH7	MyHC-7	β cardiac	β
MYH8	MyHC-8	Perinatal	PERI
MYH9	MyHC-9	Non-muscle A	NMA
MYH10	MyHC-10	Non-muscle B	NMB
MYH11	MyHC-11	Smooth Muscle	SMTH
MYH13	MyHC-13	Extraocular	EXOC
MYH7b	MyHC-7b	Slow Tonic (β-Cardiac)	STβ
MYH14	MyHC-14	Non-muscle C*	NMC
MYH15	MyHC-15	Extraocular-2	ExtOC
MYH16	MyHC-16	Jaw	JAW

Table 1.1 The 15 myosin-II isoforms gene and heavy chain names, their conventional names and their short names. These will be referred to throughout this report (*Ozkan, E., Aceti, M., Rumbaugh, G. *et al* 2015)

In myosin-II, there are three main groups that each of these isoforms can be classified in to. These are: the fast movers, slow/efficient movers and strain sensors. These different groups are based on their different biochemical and kinetic properties (Bloemink and Geeves, 2011). For the fast movers (2A, 2B, 2D/X), they have a very weak ADP affinity and a low duty ratio. The slow/efficient movers such as β-cardiac and smooth muscle isoforms have a biphasic ADP-release step, and the strain sensors non-muscle A and B have a tight ADP affinity with a higher duty ratio than the fast movers (Bloemink and Geeves, 2011). The other isoforms can further be classified in to these three groups.

Skeletal muscle isoforms 2B, 2D/X and 2A are all adult myosins that are expressed throughout the developmental cycle. Although the skeletal isoform 2D/X is not expressed in humans at the protein level (Kurapati, R., McKenna, C., Blanco, G. 2011), 2A and 2B are the two isoforms that are involved in skeletal muscle contractions. 2A is the fast-twitch isoform and 2B is the slow-twitch isoform (Tajsharghi, H., Oldfors, A. 2013). Embryonic and perinatal myosin is expressed during early stage development of the foetus in the uterus, and is downregulated after birth. These isoforms are also expressed during muscle regeneration. Non-muscle A and non-muscle B isoforms are involved in cellular processes such as cytokinesis (Hindman, B., Goeckeler, Z., Sierros, K., Wysolmerski, R. 2015) and smooth muscle is alternatively spliced to produce different isoforms that are involved with smooth muscle contraction (Haase, H., Morano, I. 1996). The extraocular isoform is a fast twitch located around the eyes that assist with its rapid movement (Tajsharghi, H., Oldfors, A. 2013). Non-muscle C, slow tonic (β-cardiac), Extraocular (MYH15) and MYH16 are all ancient isoforms that have a varied expression pattern system (Rossi, A., Mammucari, C., Schiaffino, S. *et al*. 2010).

The cardiac isoforms α and β are predominantly found within the heart muscle throughout mammals. The slow-type β-cardiac myosin is found in slow twitch fibres too. The α-cardiac myosin is expressed in the atrium of all mammals and predominantly in the ventricle of small mammals such as mice; whereas the β-cardiac isoform is predominantly expressed in the ventricle of large mammals. The α-cardiac isoform contracts faster than the slow β cardiac isoform, which is important for the normal function of the heart as an organ, filling the ventricles with blood from

the atria before pumping the blood out (Lowey, S., Bretton, V., Trybus, K. 2013). The α isoform hydrolyses ATP ten-fold faster than β, has a five-fold weaker affinity for actin, and has a ten-fold faster ADP release.

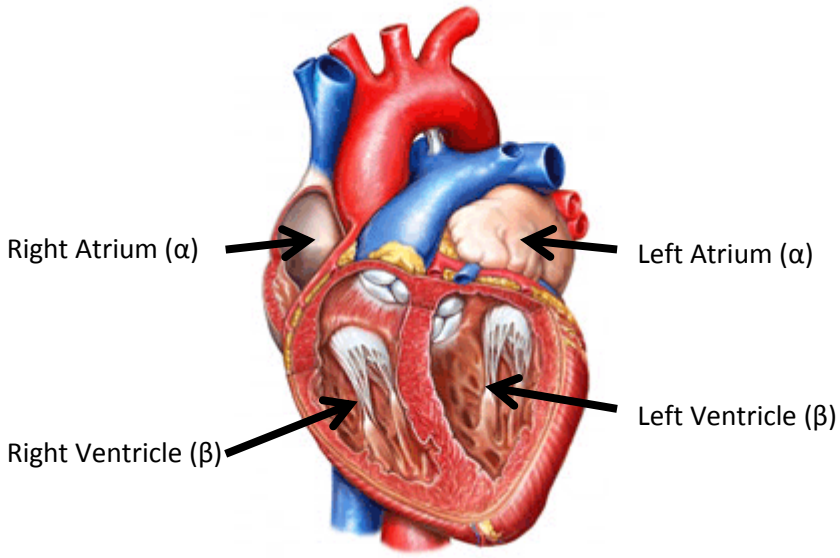


Figure 1.11 The key structures of the heart and locations of predominant α and β isoform presence in humans. Deoxygenated blood enters the heart from the body to the right atrium and subsequently to the right ventricle, where it is pumped in to the pulmonary artery for gaseous exchange in the lungs. Oxygenated blood enters the left atrium via the pulmonary vein and pumped in to the left ventricle where it is pumped out to the body via the aorta. Image adapted from *The Wellington Hospital: The Heart*.

The two cardiac isoforms share a 91% sequence similarity in their motor domains despite their differences in characteristics and kinetics (Deacon, J., Bloemink, M., Leinwand, L. 2012). This compares to the full length sequences of the other isoforms, where adult skeletal 2A, 2B and 2X/D share identities from 78.9% - 94.8%.

1.5 Heart Rate vs Size Phenomenon

Where isoforms so obviously differ in their kinetic properties due to their different functions, previous studies have established that external pressures such as mass can introduce evolutionary pressures (Pellegrino, M., Canepari, M., Bottinelli, R. 2002). As mammals vary hugely in terms of their body mass, for them to maintain similar locomotive speeds, an inverse shortening rate must be related to limb length and body size (Hill, A. 1950). In much the same way, the heart rate of a particular organism varies with their body size, where the larger the organism the slower the heartbeat (figure 1.12). To compensate for this slowed rate, the heart will increase in size to pump

out more blood across the body to sufficiently supply the extremities of the body (Savage, V., Allen, A., West, G. 2007). While muscles will vary in size, cell sizes stay similar throughout changes in mass (Pellegrino, M., Canepari, M., Bottinelli, R. 2002).

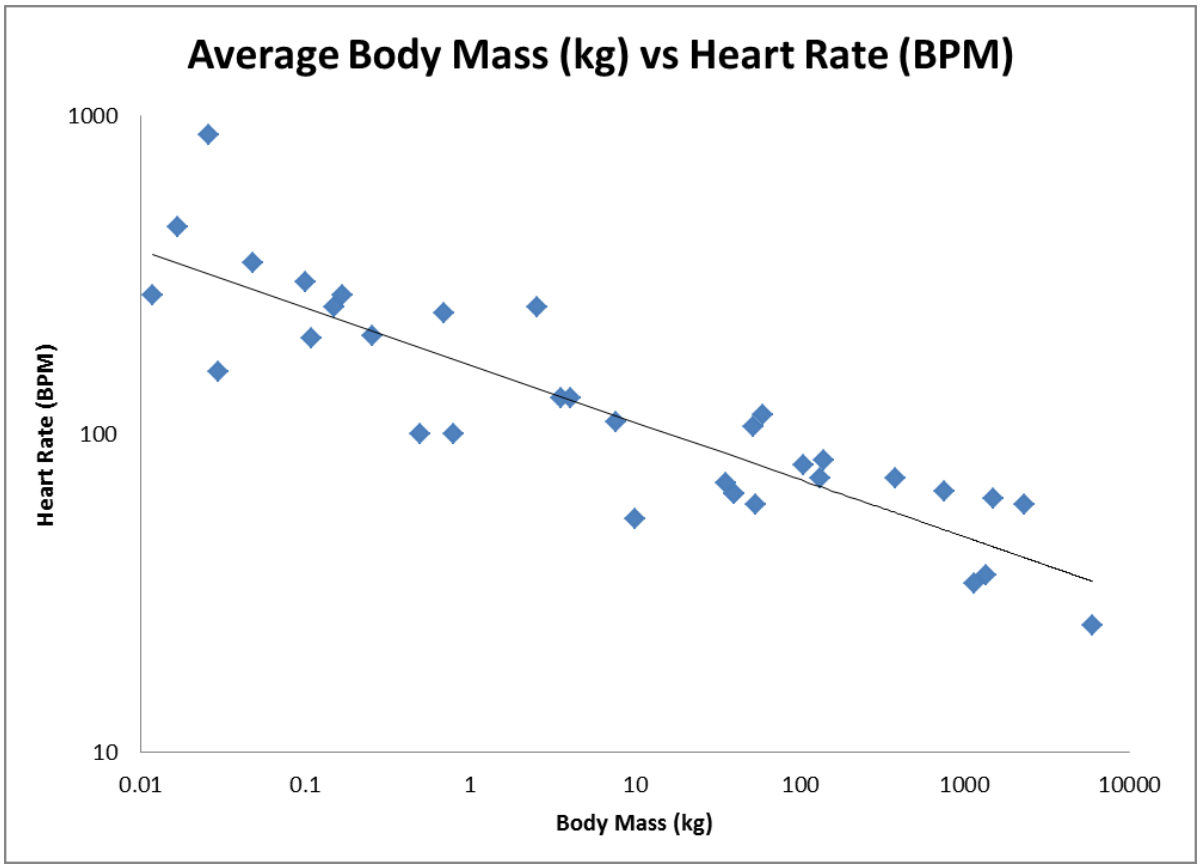


Figure 1.12 The heart rate versus size phenomenon. The average body mass has an inverse relationship with heart rate as it increases. Each animal had its heart rate and body mass plotted in order to show a correlation. A trend line highlights the negative correlation between the heart rate and body mass. Values for body mass and heart rate for each species used was found in literature and compiled by Jeanfavre, S. 2014. The graph is in a log scale. Image adapted from Jeanfavre, S. (2014).

The ability of a muscle to produce the necessary speed for movement is both dependent on the predominant isoform present and the ratio expressed of different isoforms in the muscle. As figure 1.12 highlights this heart rate versus size phenomenon, it reflects on the isoforms present in the heart and the rate of the cross-bridge interactions. It may therefore be stated that the myosin kinetics may play a role in dictating the heartbeat.

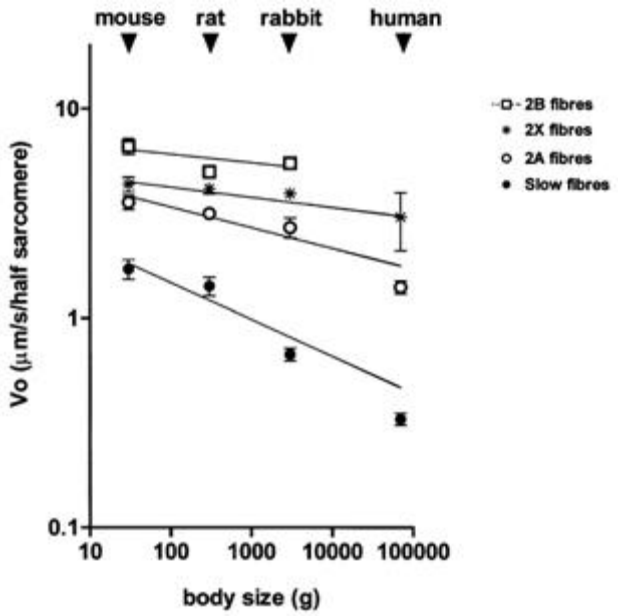


Figure 1.13 The shortening velocity of myosin decreases as body size increases. The mouse, rat, rabbit and human myosin skeletal and slow isoforms had their average shortening velocity (V_0) measured (Pellegrino, M., Canepari, M., Bottinelli, R. 2002).

Reflecting the importance of sequence changes on the difference of shortening velocity in similar isoforms across species, slower rates are observed in larger animals (Pellegrino, M., Canepari, M., Bottinelli, R. 2002)(Malmqist, U., Aronshtam, A., Lowey, S. 2004). Through comparing these different isoforms and observing point changes in the structure, it may be possible to determine functionally important residues that control rates of the cross-bridge cycle.

1.6 The DNA of Myosin-II

Point mutations observed in structures of proteins result from nucleotide mutations observed in DNA sequences of protein-coding genes. The genetic code maps 64 codon triplets to 20 amino acids and three stop codons. These codons are then translated and transcribed *in vivo* to synthesise a proteins structure. Looking at mutations at a nucleotide level gives an insight in to the evolutionary pressures of the protein through the presence of synonymous and non-synonymous mutations. Due to the degenerate nature of the genetic code, mutations in the nucleotide sequence may not result in changes of residues in the structure of the protein, as seen in figure 1.14.

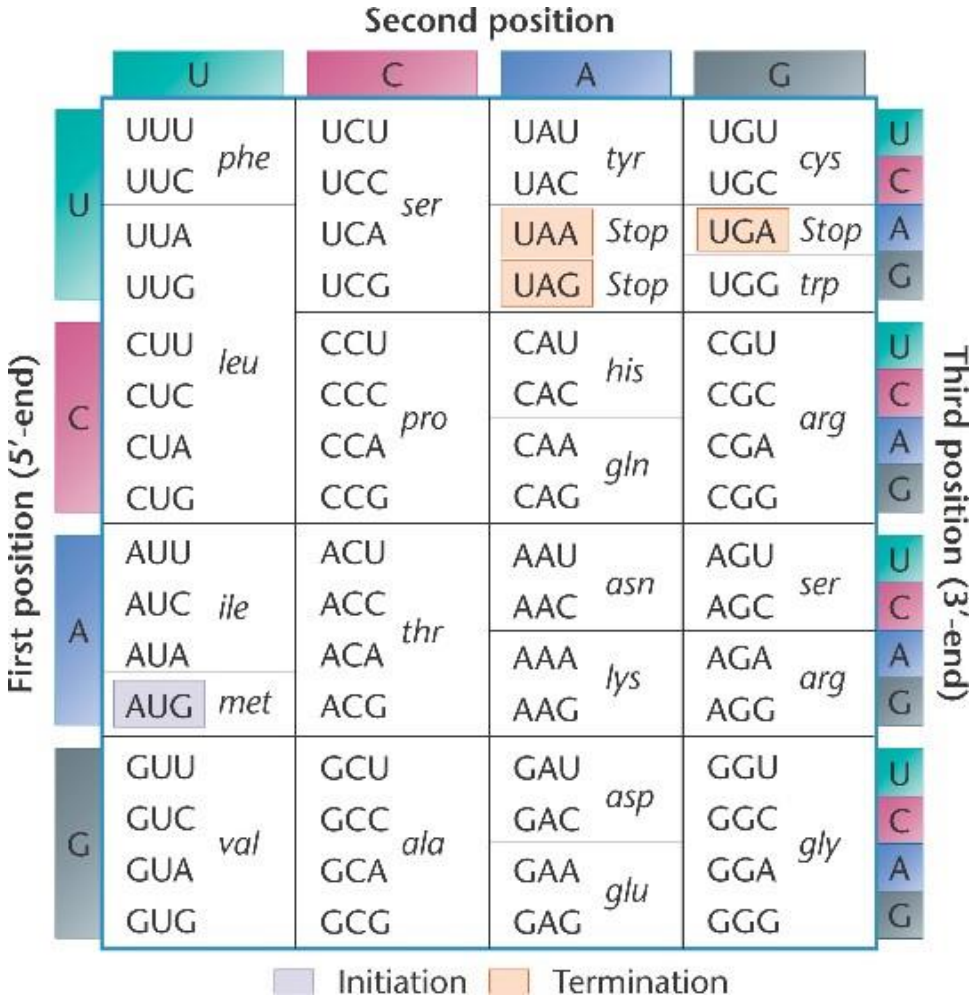


Figure 1.14 The genetic code. The degenerate nature of the code allows for different codons to code for the same amino acid. Stop and initiation codons are highlighted. Adapted from *Ground Rules for Gene Expression*.

A synonymous mutation is defined as a nucleotide substitution that results in the same amino acid being coded for, known as a silent mutation. A non-synonymous mutation is defined as a nucleotide substitution that results in a different amino acids being coded for (missense mutation) or a stop codon (nonsense mutation). Non-synonymous mutations affect a proteins structure, whereas synonymous mutations can give an insight in to the evolutionary pressures placed upon a particular gene (Yang, Z., Nielson, R. 2000).

The myosin genes for the many different isoforms are located on chromosomes 14 and 17 in humans (Saez, L., Gianola, K., Leinwand, L. 1987). Skeletal isoforms cluster in chromosome 17, whereas both the α and β genes cluster on chromosome 14. Analysis of these sequences may give a further insight in to the evolution of myosin-II in terms of both evolutionary distance and mass.

1.7 Aims

Determining whether β-Cardiac myosin has mutations that allow for different heart rates is the main premise for this project. In order to investigate this correlation, five main aims have been proposed to help evaluate the relationship β-cardiac myosin shares with other isoforms:

1: To investigate the rate of divergence in twelve myosin-II isoforms from twelve mammalian species. This will test whether the β-cardiac isoform diverges at a rate different to other sarcomeric and non-muscle isoforms. Divergence rates will be determined from both evolutionary distance and mass.

2: To investigate whether the rate of divergence in myosin isoforms seen in terms of evolutionary distance and mass in the twelve mammalian species is representative when taking in to consideration as many mammalian sequences as possible.

3: To investigate whether there are different evolutionary pressures placed on the motor domain of β-cardiac myosin that are not present in the tail domain. Comparing these rates to other myosin-II isoforms will allow for comparison when considering both evolutionary time and mass.

4: To investigate primary sequence divergence in β-cardiac myosin to verify residues of key interest that may impact the cross-bridge cycle.

5: To investigate the DNA sequences of myosin to further determine rates of divergence that may be seen in terms of synonymous mutations at the nucleotide level. These mutations may affect intron/exon boundaries, or may show large degrees of variation within in introns that would not be reflected in the protein structure.

2.0 Methods

2.1 Primary Sequence Database of Myosin Isoforms

Myosin sequences for the different isoforms used (MYH1, MYH4, MYH9, MYH11, MYH13, MYH14) that were compiled in addition to the data collated by a collaborator S. T. Jeanfavre (2014) for other isoforms (MYH2, MYH3, MYH6, MYH7, MYH8, MYH10), were extracted from Uniprot (Apweiler R., Bairoch A., Wu C.H, 2004), Ensembl (Hubbard, T., 2002) and NCBI (Wheeler, D., Chappey, C. *et al*. 2000). Primary sequences were downloaded in fasta format and accession numbers were noted (Appendix A). These extracted sequences formed the basis for comparisons between different species. Of the fifteen isoforms available, only twelve were studied due to their universal existence within mammals and their good characterisation. Isoforms that were not studied included were MYH14, MYH15 and MYH16.

Using the well characterised human myosin sequences as a reference for each individual isoform, the quality of sequences was assessed depending on their Basic Local Alignment Search Tool (BLAST) (Altschul, S., *et al*, 1990) score with the canonical human sequence for each isoform and the length of the sequence. Queries with around the same number of amino acids as the human sequences (varying from 1935-1976 amino acids with each isoform) and query coverage and identity of over 90% with confidence scores of <0.001 were selected for in order to ensure confidence and completeness of the sequence. Sequences with high BLAST scores but contained unidentified residues, represented by an X, were excluded. A database was constructed initially for twelve mammalian species in order to compare individual isoforms to one another. These twelve species were selected for their well-sequenced genome, the availability of the isoform sequences and gapless alignments. Not all genomes contained quality sequences or had sequences available, and analyses performed on these sequences have the missing sequences identified where relevant. A larger database was built in order to determine if the smaller sample size was representative of the taxa as a whole. The same processes were applied, however if the complete quality sequence

was present for that isoform it was included regardless of whether the same species sequence was present for other isoforms, in order to consider as many sequences as possible.

2.2 Comparing Evolutionary Divergence of Myosin Isoform Domains

In order to investigate divergence of the motor and tail domains of each isoform, each sequence was split in to their appropriate domains (figure 1.7) from sequence motifs as identified on Uniprot, excluding the IQ domain located between these domains as it is too short for comparison. These motifs are shown in table 2.1. Isoforms had differing sequence motifs due to changes in their primary sequence. Evolutionary distances were determined from TimeTree.org (Hedges, S., Dudley, J., Kumar, S. 2006) where each species distance was calculated in million years (Myr) from one another in order to create a matrix of evolutionary distance values. TimeTree.org uses a hierarchical tree-based system, compiling published data on molecular time estimates between the divergences of two taxa in to one database. TimeTree.org Identified that this data was difficult for the scientific community to find on their own, so this database introduces a user interface that generates data within seconds. Using an algorithm, query inputs are identified and the algorithm generates phylogenetic trees from common gene sequences found in NCBI. This allows for identification of a point at which the queries diverged through a common ancestor. Once this point has been identified, literature from published studies that involve Bayesien statistics, fossil-calibration and ribosomal RNA conservation on this time point is searched and an average time of divergence is conceived (Hedges, SB, Dudley, J, & Kumar, S., 2006)(Hedges, S., Marin, J., Kumar, S. *et al*. 2015).

Isoform	Head Motif Sequence	Tail Motif Sequence
2B	LEEMRD	LLKSAE
2A	LEEMRD	LLKSAE
EMB	LEEMRD	LLKSAE
2D/X	LEEMRD	LLKSAE
α	LEEMRD	LKSAET
β	LEEMRD	LLKSAE
PERI	LEEMRD	LLKSAE
NMA	LEEERD	LLQVSR
NMB	LEEERD	LQVTR
SMTH	LEEERD	LLQVTR
EXOC	LEEMRD	LLKSAE
STβ	LEEERD	LQVTR

Table 2.1 The sequence motifs for the domains of each isoform. Sequences were divided using these motifs.

2.3 Comparing Divergence of Myosin Isoform Domains with Mass

When considering mass as a parameter, the average mass of the adults of both male and females of the species was considered. Different databases were used in order to determine these weights, each of which are listed in table 2.2.

Species	Mass (kg)	Database
Human	68	Animal Diversity Web: animaldiversity.org
Bonobo	45.5	Animal Diversity Web: animaldiversity.org
Macaque	6.55	ARKive: Arkive.org
Tarsier	0.1315	Primatology.net
Rat	0.325	ARKive: Arkive.org
Mouse	0.017	ARKive: Arkive.org
Guinea Pig	0.9	Animal Diversity Web: animaldiversity.org
Hamster	0.1125	Animal Diversity Web: animaldiversity.org
Cow	755	Animal Diversity Web: animaldiversity.org
Brandt's bat	0.0069	ARKive: Arkive.org
Minke whale	9200	National Oceanic and Atmospheric Administration Fisheries: nmfs.noaa.gov
Opossum	0.048	ARKive: Arkive.org

Table 2.2 The average adult body mass of the twelve mammalian species used with the databases used to generate these values.

2.4 Multiple Sequence Alignments and Matrices

After splitting each sequence in to their appropriate motor and tail domains, Clustal Omega (Sievers, F., Wilm, A., Dineen, D., *et al*, 2011) was used in order to generate a multiple sequence alignment of the input sequences to determine areas of divergence and similarity. By ordering the sequences in order of either their evolutionary distance or mass size, and setting parameters on the server to keep the sequences in that order, patterns can be observed that highlight areas of

key interest. Each query that is put through the server gives a percent identity matrix that determines what percent each input sequence shares with another. After taking each of the scores and conditionally formatting them, colour-gradient scales can be generated that easily indicate areas of more or less divergence. Alignments were generated for the motor and tail domains of each isoform to determine rates of divergence in both these domains. Multiple sequence alignment outputs were annotated using ESPript (Robert, X., Gouet, P. 2014), that maps secondary structures elements on the alignment using known secondary structures. As crystal structures are limited for myosin-II isoforms, the β-cardiac multiple sequence alignment was annotated using the protein database structure 4DB1 (appendix C).

Once these were generated, further analysis involved the use of these percentage scores versus the evolutionary distance calculated or the masses obtained from literature. Using Origin (OriginLab, Northampton, MA), linear relationship fitting was used to calculate the divergence rates.

2.5 Protein Residue Conservation Prediction

After ensuring the completeness of sequences and alignments, a protein residue conservation test was performed on alignments by way of the Jensen-Shannon sequence divergence (Capra, J. Singh, M. 2007). This programme estimates the sequence conservation shared between proteins in positions of a multiple sequence alignment. From this, a perl script was written to determine protein divergence. The output file of the programme aligned the sequences in the order of their input in a way that this bioinformatics tool was able to be developed. This perl script counted the frequency of amino acids that share the same residue between species and the number of different amino acids present at each position in the alignment. Through the use of Origins, structures were marked on the motor domain of the protein that enabled further analysis to be done. This gave indications as to where in the multiple sequence alignments variations occurred.

2.6 DNA Analysis using Artemis, ACT and SNAP.

To further determine any evolutionary pressures placed on the myosin-II isoforms, databases of DNA sequences from the twelve species originally analysed were collated from NCBI (Wheeler, D., Chappey, C. *et al*. 2000) and Ensembl (Hubbard, T., 2002). The golden hamster (*Mesocricetus auratus*) sequence was excluded due to the presence of uncharacterised nucleotides (represented by N), concluding that the sequence was incomplete. The sequences were aligned using the Clustal Omega (Sievers, F., Wilm, A., Dineen, D., *et al*, 2011) DNA alignment software in order to analyse the synonymous and non-synonymous mutations. These sequences were then submitted to a Synonymous and Non-synonymous Analysis Programme (SNAP) that was able to determine the number of mutations observed within a given open reading frame. DNA sequences were also downloaded in .embl format to observe open reading frames within Artemis (Rutherford, K., Parkhill, J., Barrell, B. 2000) from the genes. These open reading frame sequences were extracted to better compare variations in the coding sequences. Sequences were also compared using the Artemis Comparison Tool (ACT) (Carver, T., Rutherford, K., Parkhill, J. 2005) to identify if exon/intron regions are conserved between species. Using the boxshade server written by Hoffman, K. and Baron, M., and hosted on ExPASy, these DNA exon sequences could be compared to better show intron/exon conservation throughout species sequences.

3.0 Results

3.1 Generating a myosin-II sequence dataset

The aims of this project were to analyse and identify any regions of variation in the myosin-II β-cardiac sequence and then examine any that would indicate a correlation between the mass of the organism sequence divergence and evolution over time. In order to do this, databases were constructed in order to complete analyses. Initially the sequence databases were searched to identify all available myosin-II sequences. These sequences are shown in table 3.1.

After obtaining these sequences, their domains were divided using sequence motifs as discussed in methods. Due to the importance of the motor (head) domain in myosin's activity, initial analysis was performed on these sequences to determine if any divergence patterns could be observed between species isoforms. The β-cardiac sequence was chosen as it is the isoform of interest in these experiments and shows a mass dependence. Non-muscle A was chosen as a direct comparison and a control, as it acts at the cellular level and would not be under the same stresses as the sarcomeric β-cardiac. Embryonic myosin was also chosen as it is expressed in foetal development and regeneration of muscle, and is downregulated after birth.

		Isoform Motor and Tail Domain Database											
Species	Average Adult Mass (kg)	2A	2B	2D/X	α	β	NMA	NMB	SM	ExOC	Slow	Peri	Emb
Mouse	0.017	×	×	×	×	×	×	×	×	×	×	×	×
Brandt's Bat	0.008	×	×	×	×	×	×	×		×	×	×	×
Opossum	0.048	×	×	×	×	×	×	×	×	×		×	×
Hamster	0.1125				×	×	×	×		×		×	×
Tarsier	0.1315		×	×	×	×	×	×		×	×	×	×
Rat	0.325	×	×	×	×	×	×	×	×	×	×	×	×
Guinea Pig	0.9	×	×	×	×	×	×	×	×	×	×	×	×
Macaque	6.55	×	×	×	×	×	×	×	×	×	×	×	×
Bonobo	45.5	×	×	×	×	×	×	×		×	×	×	×
Human	68	×	×	×	×	×	×	×	×	×	×	×	×
Cow	755	×	×	×	×	×	×	×	×	×	×	×	×
Minke Whale	9200	×				×	×	×	×		×	×	×

Table 3.1 The list of species used that had a number of fully sequenced isoforms available. The masses of each animal are listed and shown in kg. An × indicates that the sequence was present and a grey shaded box indicates that a sequence was missing. A total of 130 isoform sequences were analysed.

3.2 Evolutionary Divergence of the Motor Domain in Three Isoforms

Sequences were ordered in terms of their evolutionary distance from humans, as discussed in methods. Multiple sequence alignments were generated and percent identity matrices produced as shown in figure 3.1. Initial analysis was conducted on the β-cardiac sequence, non-muscle A and embryonic myosin, as they had a full set of sequences for each of the species listed and represented isoforms from different groups (figure 3.1).

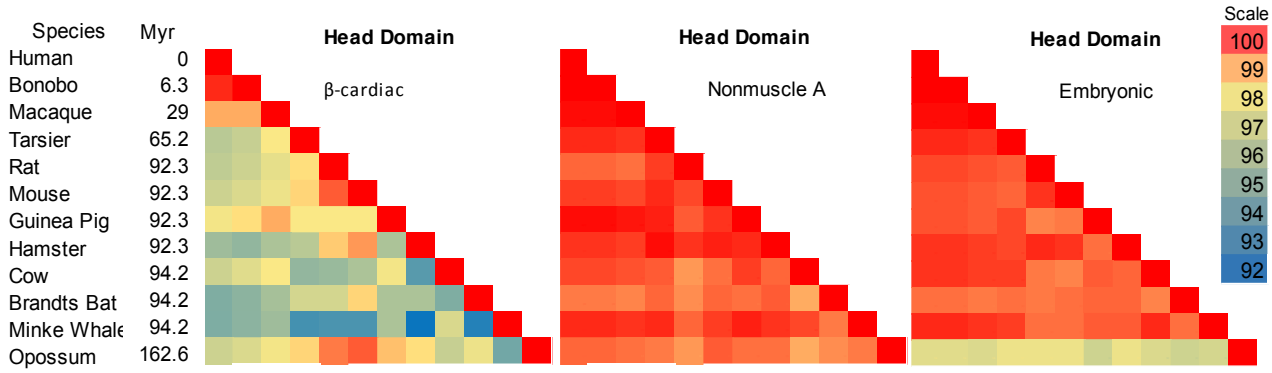


Figure 3.1 The percent identity matrices for β-cardiac, non-muscle A and embryonic motor domains. Each coloured square represents an individual comparison between two sequences. Red indicates a 100% sequence conservation, with a gradient spanning down to blue with a 92% sequence conservation. All species listed are in order of how closely they are related to humans in Myr.

Table 3.2 introduces the evolutionary distances each species shares with one another. These values were used for evolutionary distance percentage identity plots.

Human	**0**										
Bonobo	6.3	0									
Macaque	29	29	0								
Tarsier	65.2	65.2	65.2	0							
Rat	92.3	92.3	92.3	92.3	0						
Mouse	92.3	92.3	92.3	92.3	25.4	0					
Guinea Pig	92.3	92.3	92.3	92.3	77.9	77.9	0				
Hamster	92.3	92.3	92.3	92.3	43.9	43.9	77.9	0			
Cow	94.2	94.2	94.2	94.2	94.2	94.2	94.2	94.2	0		
Brandts Bat	94.2	94.2	94.2	94.2	94.2	94.2	94.2	94.2	81.6	0	
Minke Whale	94.2	94.2	94.2	94.2	94.2	94.2	94.2	94.2	56	81.6	0
Opossum	162.6	162.6	162.6	162.6	162.6	162.6	162.6	162.6	162.6	162.6	162.6	0

Table 3.2 The evolutionary distance matrix for the percent identity matrices generated in figure 3.1. Values are in Myr. Each of these individual evolutionary distance values equate to the percent identity scores from figure 3.1.

Overall, divergence between sequences seen in the β-cardiac motor domain are not seen in the non-muscle A and embryonic myosin datasets. There is a higher rate of divergence in the β-cardiac head, as indicated by the presence of sequence identities in the 94-98% range. Motor domains of non-muscle A and embryonic isoforms have high sequence identity scores in the 99% region. In order to visualise these relationships better, the species sequence identities were compared to their evolutionary distance (table 3.2) and results are shown in figure 3.2.

In figure 3.2, the data for the β-cardiac sequence is seen to have a steeper trend in its sequence divergence than in both non-muscle A and embryonic myosin. The β-cardiac sequence shows a large scatter of data around the 94 Myr point, suggesting that the comparison of evolutionary distance with sequence divergence does not explain the correlation efficiently. The slope of -0.049 has a strong Pearsons correlation coefficient of 0.958, suggesting that there is a relationship between evolutionary distance and the divergence of the isoforms motor domain.

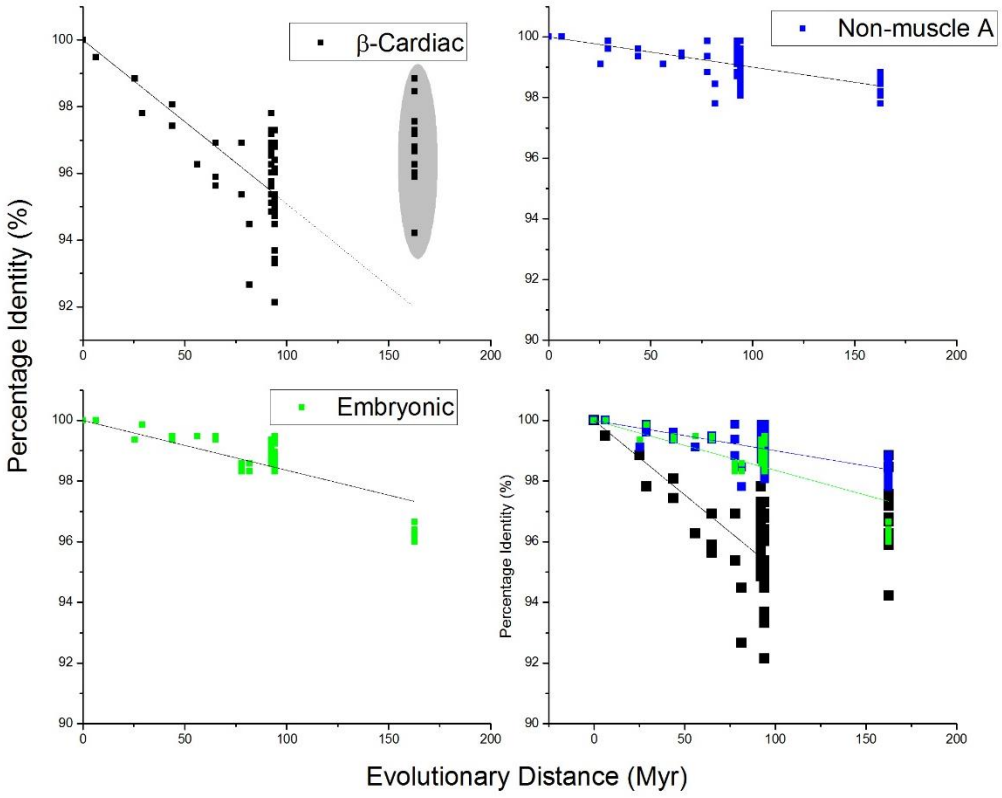


Figure 3.2 The evolutionary distance comparisons for β-cardiac, non-muscle A and embryonic isoforms. The percent identity scores each sequence shares with one another are plotted alongside each sequence comparisons evolutionary distance from one another. These plots allowed the generation of linear relationship to show any correlations. The β-cardiac sequence does excludes opossum data from the analysis as the data suggests the trend is shown in mammals with an evolutionary distance of up to 92 Myr. The opossum has a large distance of 162.6 Myr. β-cardiac sequence is represented by black squares, the non-muscle A sequence is represented by blue squares and the embryonic myosin is represented by green squares.

trend. The opossum data showed differing divergence over time due to a wide range of conservation and divergence between other species. As the opossum is equally distant to all species included in the sample set due to its being the only marsupial, it is difficult to determine whether this variance of divergence solely due to time.

Both non-muscle A and embryonic myosin have weaker rates of divergence with slopes of -0.01 and 0.016 respectively. Non-muscle A has a correlation coefficient of 0.931 and embryonic 0.928, suggesting that although not as strong, there is a relationship with sequence divergence as evolutionary distance increases.

3.3 Mass as a Parameter for Motor Domain Divergence in Three Isoforms

After considering the relationship between sequence divergence and evolutionary distance, next the relationship with average adult body mass was considered (figure 3.3). The Brandt's bat (*Myotis brandtii*) was found to be the smallest mammal in the dataset, and was used as a reference to compare other sequences. Due to the excelled divergence of the protein, data was skewed and reflected a divergence of physiology and use of the isoform in the bat as opposed to mass. A literature search indicated that *Myotis brandtii* is the longest lived mammal for its size and challenges a widely observed mass versus longevity relationship, where small mammals have shorter lifespans than larger animals (Seim, I., Fang, X., Lobanov, A. *et al*. 2012). This provided reasonable grounds to exclude it from being the reference sequence, and the well-studied mouse was used in lieu of this.

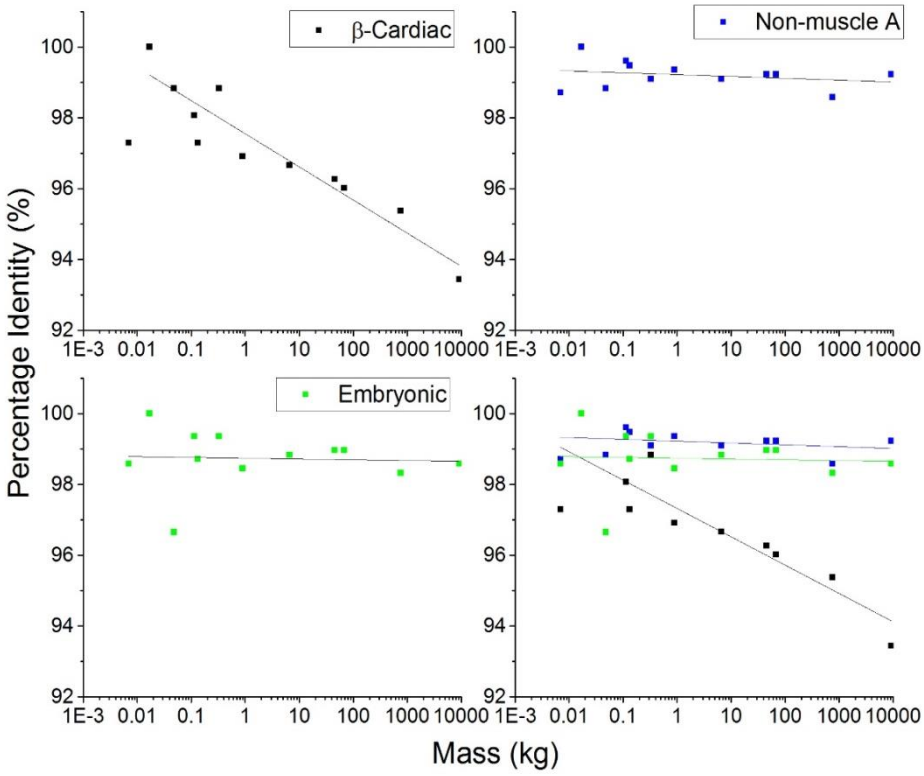


Figure 3.3 The percent identity scores each sequence shares with one another plotted alongside the average adult body mass for that species. This analysis was performed only on the identity scores from the mouse as body mass was used as a parameter instead of evolutionary distance. The plots are on a log scale and data points are labelled.

Figure 3.3 showed a strong correlation between sequence divergence and mass in the β-cardiac motor domain, with a slope of -0.936 and a correlation coefficient of -0.945. Shallower slopes for both non-muscle A (-0.052) and embryonic (-0.022) motor domains suggested that there is no relationship between mass and divergence. This is reflected by non-muscle A's weak correlation coefficient of -0.257 and the embryonic correlation coefficient of -0.053.

3.4 Evolutionary Divergence of the Tail Domain in Three Isoforms

The trends seen in the motor domain of β-cardiac, non-muscle A and embryonic myosin led to question as to whether similar relationships could be observed in the tail domains. The equivalent analysis was performed to determine whether the same sequences diverge at rates observed of the protein as a whole or only certain domains. Due to the catalytic activity of the motor domain, the majority of variations were expected to be observed here.

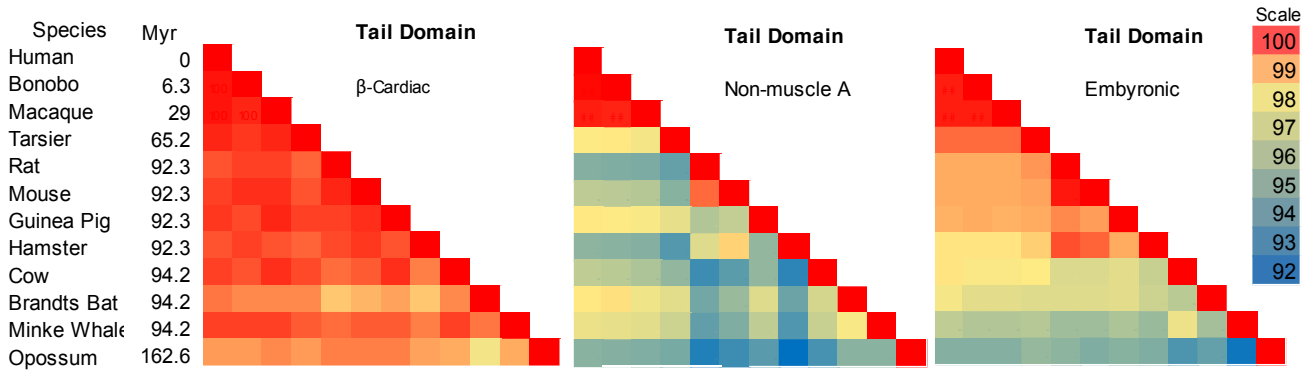


Figure 3.4 The percent identity matrix scores each sequence shares with one another for the three main isoforms studied. Red indicates total sequence conservation, spanning down to blue indicating there is a 92% sequence conservation.

This analysis provided a more detailed insight in to the evolution of the myosin-II protein. Trends seen in the motor domains for the protein appear to be the opposite in the tail domains. Where the β-cardiac motor domain sequence shows more divergence than both non-muscle A and embryonic myosin (figure 3.1), figure 3.4 shows that there is a greater divergence in both non-muscle A and Embryonic than there is in β-cardiac. Non-muscle A myosin tail domain sequences appear to show more sequences that show divergence in the 92-94% range than embryonic myosin. β-cardiac tail domain sequences are highly conserved, with most sequences showing conservation greater than 99%.

This prompted a more detailed investigation similar to that of the motor domains of the protein. Evolutionary distance plots were generated in the same way the motor domain sequence plots were, with the matrix scores plotted against the percentage identity scores.

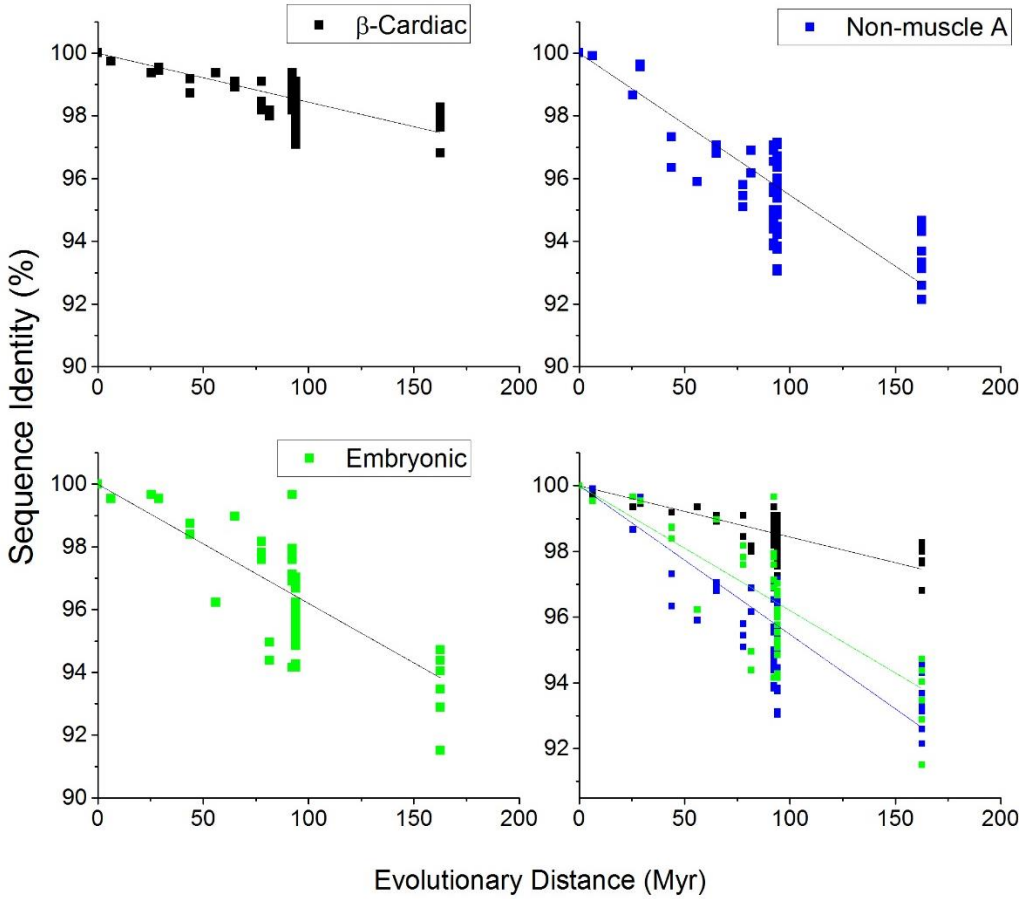


Figure 3.5 The evolutionary distance matrixes for the tail domains of β-cardiac myosin (black squares), non-muscle A (blue squares) and embryonic myosin (green squares).

This analysis highlighted the result of a reverse rate of divergence seen in the three isoforms between the domains. Where there is more divergence in the motor domain, there is more conservation in the tail domain, and vice versa. The β-cardiac sequence shows a slope of -0.017 and a correlation coefficient of 0.96, whereas non-muscle A has a slope of -0.045 and a correlation of 0.927 and embryonic has a slope of -0.038 and a correlation of 0.926.

3.5 Mass as a Parameter for Tail Domain Divergence in Three Isoforms

Similar investigations were performed on the sequence comparisons with increasing mass of the species to determine if the relationships observed for the motor domains held true in the tail domains.

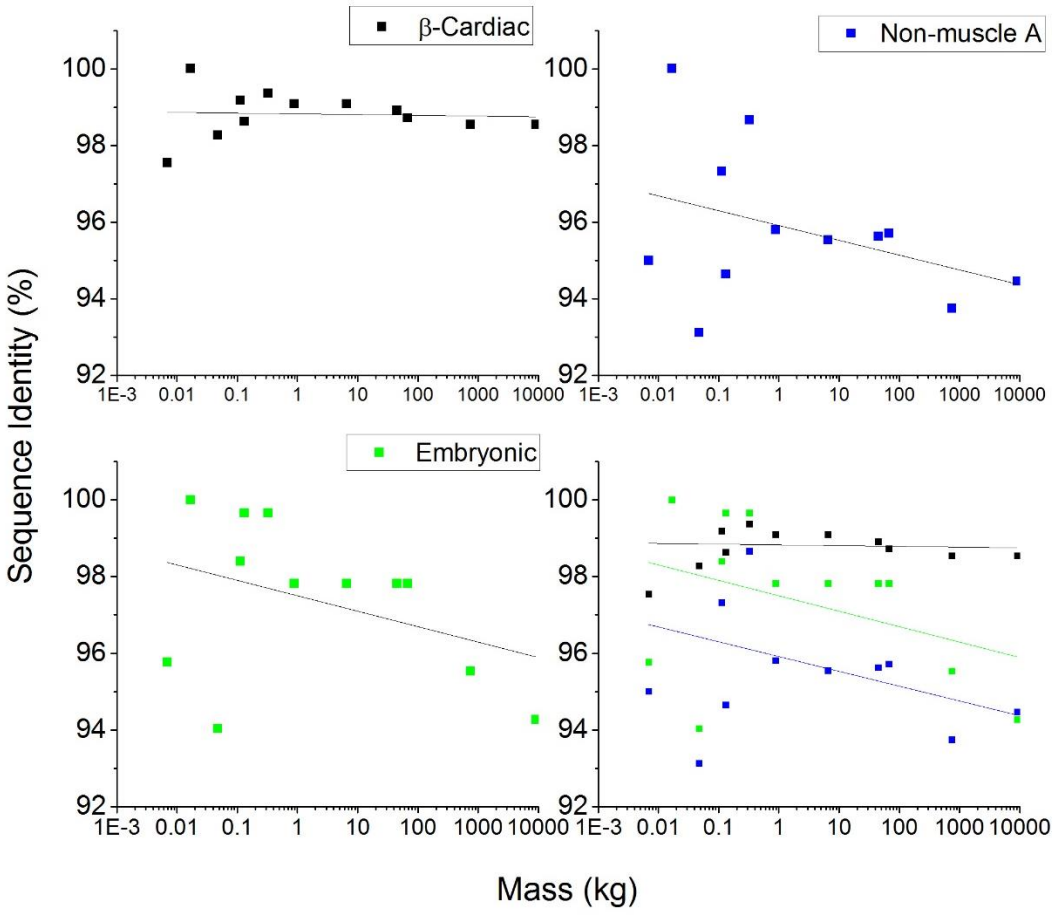


Figure 3.6 The mass divergence of the tail domains of β-cardiac, non-muscle A and embryonic myosin. Graphs are in a log scale.

Figure 3.6 comparisons showed that there is no mass dependence of the β-cardiac sequence tail domain, with a slope of -0.122 and a correlation of -0.471. There appears to be stronger slopes with both non-muscle A (-0.386) and embryonic (-0.403), however weak correlation coefficients of -0.376 and -0.381 respectively indicate that these trends do not hold true to the relationship.

3.6 Evolutionary Divergence of Motor Domains in other Isoforms

The three main isoforms were studied due to the presence of all twelve sequences. The other myosin-II sequences were studied, however most had fewer complete sequences, making it harder to draw conclusions.

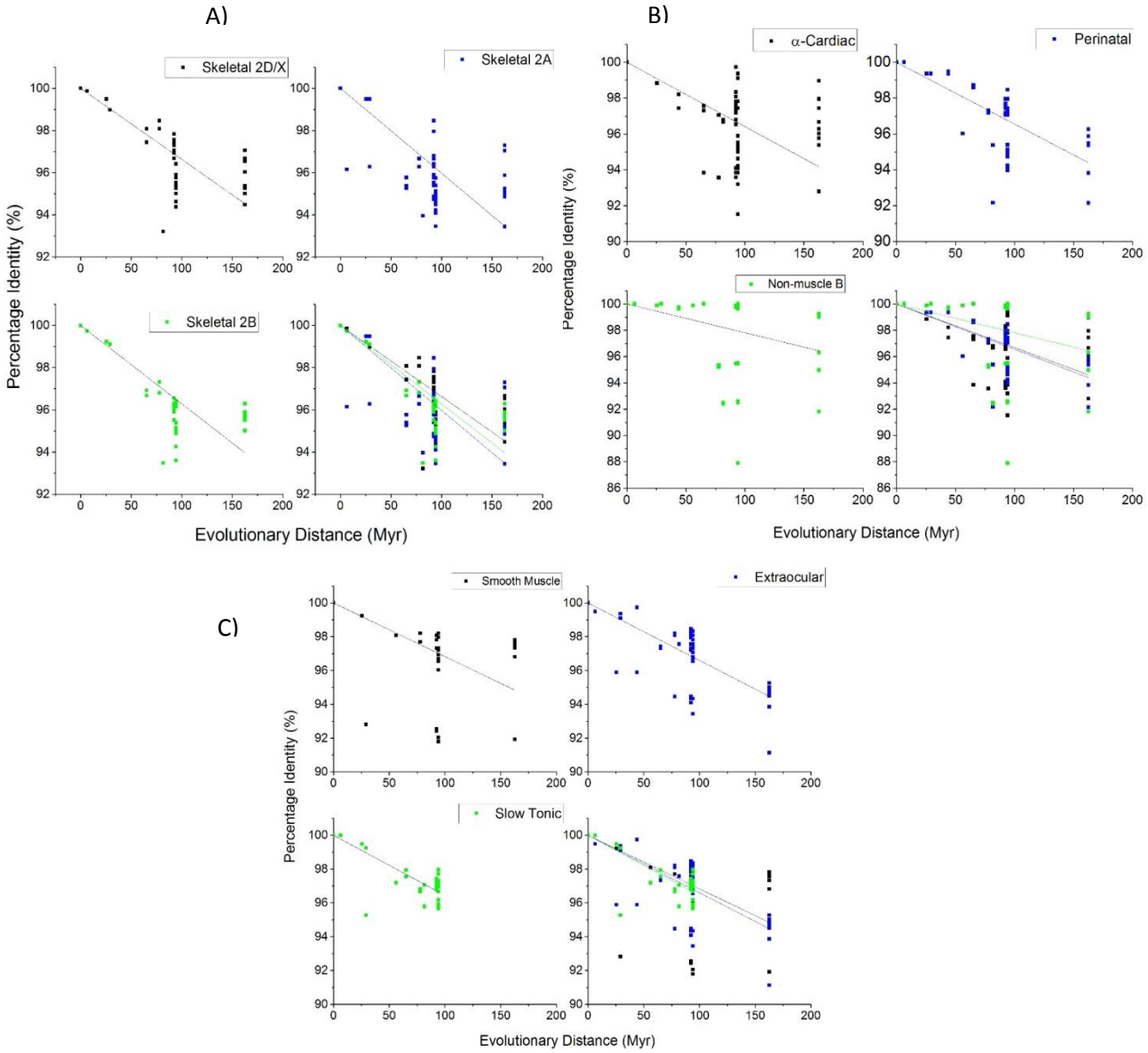


Figure 3.7 The evolutionary distance matrix plots versus the percent identity scores for motor domains. A) shows the skeletal muscle isoforms. All three isoforms have the hamster sequence missing. Skeletal 2B and D/X have the Minke whale missing and skeletal 2A has the Tarsier missing. B) shows α-cardiac, perinatal and non-muscle B isoforms. The α-cardiac sequence has the minke whale missing. C) shows smooth muscle, extraocular and slow tonic isoforms. Smooth muscle has the bonobo, tarsier, hamster and brandt's bat sequences missing, extraocular has the minke whale missing and slow tonic has the hamster and opossum data sequence missing.

Observations of the skeletal isoforms in figure 3.7 A indicate that there is a strong correlation between the evolutionary distance and sequence divergence of the protein. 2D/X shows a slope of -0.034 and a correlation coefficient of 0.918, 2A shows a slope of -0.041 and a correlation coefficient of 0.921 and 2B shows a slope of -0.037 and a correlation of 0.918.

The α-cardiac isoform in figure 3.7 B appears to show a large scattering of data points around both the 92 Myr and 162 Myr points. The line of best fit indicates a very similar slope to that of the perinatal isoform (-0.036 and –0.034 respectively). High correlation coefficients (0.921 and 0.928) indicate this is a strong relationship. Non-muscle B shows a large scattering of data points that are as low as 88%, reflected in the high standard error score of 0.004. The species sequence that causes these low conservation scores are from *Myotis brandtii,* suggesting that this isoform may be used differently in the physiology of this organism. Many data points are grouped at the 99% conservation rate, leading to a shallow slope of -0.022 with a correlation of 0.930.

Smooth muscle, extraocular and slow tonic all show similar slopes in their lines of best fit (-0.032, -0.034 and -0.036). These slopes are supported through strong correlation coefficients of 0.92, 0.92 and -0.956 respectively. The opossum sequence was not available and therefore the analysis is limited to a shorter evolutionary time frame.

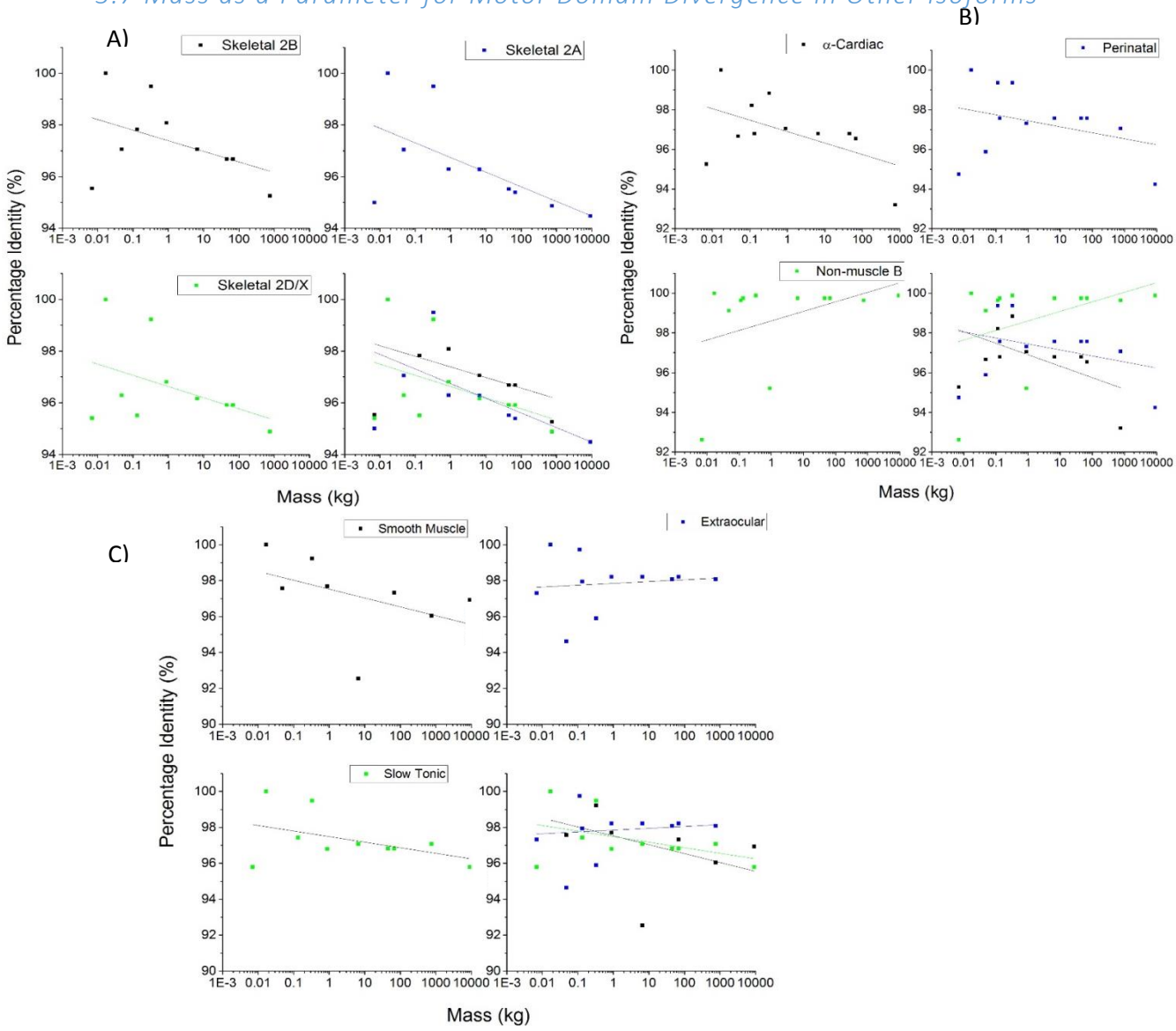


Figure 3.8 The percent identity matrix plots against the mass values of the twelve species studied.
A) shows Skeletal isoforms 2A 2B 2D/X. 2A is missing the Hamster and Tarsier sequence, 2B and 2D/X are missing the hamster and minke whale sequence.
B) shows the α-cardiac, perinatal and non-muscle B isoform sequences. The α-cardiac sequence has the minke whale missing.
C) shows smooth muscle, extraocular and slow tonic isoforms. Smooth muscle has the bonobo, tarsier, hamster and brandt's bat sequences missing, extraocular has the minke whale missing and slow tonic has the hamster and opossum data sequence missing. Plots are on a log scale.

Skeletal isoforms in 3.8 A appear to show a general trend, but the low correlation coefficients of -0.452, -0.607 and -0.435 for 2D/X, 2A and 2B respectively suggest that mass has a limited if any factor in determining the rate of divergence in the motor domains. Stronger coefficients when considering evolutionary distance suggest that this plays a greater role in the divergence of them.

Figure 3.8 B shows a slope in the α-cardiac myosin of -0.577 with a correlation coefficient of -0.528. Although weak, it appears to suggest that the isoform may have some mass dependence in

its evolution. The perinatal isoform shows a weaker slope (-0.301) and correlation (-0.332) and non-muscle B shows a weak positive relationship (-0.481 with a correlation of 0.398), due to the *Myotis brandtii* and *Mesocricetus auratus* having stronger divergence in the sequence. This may reflect the use of the isoform in their physiology. A large majority of the sequences show conservation in the 99% range.

Figure 3.8 C shows that for smooth, extraocular and slow tonic isoforms, mass dependence does not affect the divergence of the protein to the same extent that evolutionary distance does. The smooth muscle isoform shows a slope of -0.494 and a correlation of -0.446, extraocular shows a slope of 0.101 and a correlation of 0.107, and slow tonic shows a slope of -0.310 and a correlation of -0.446.

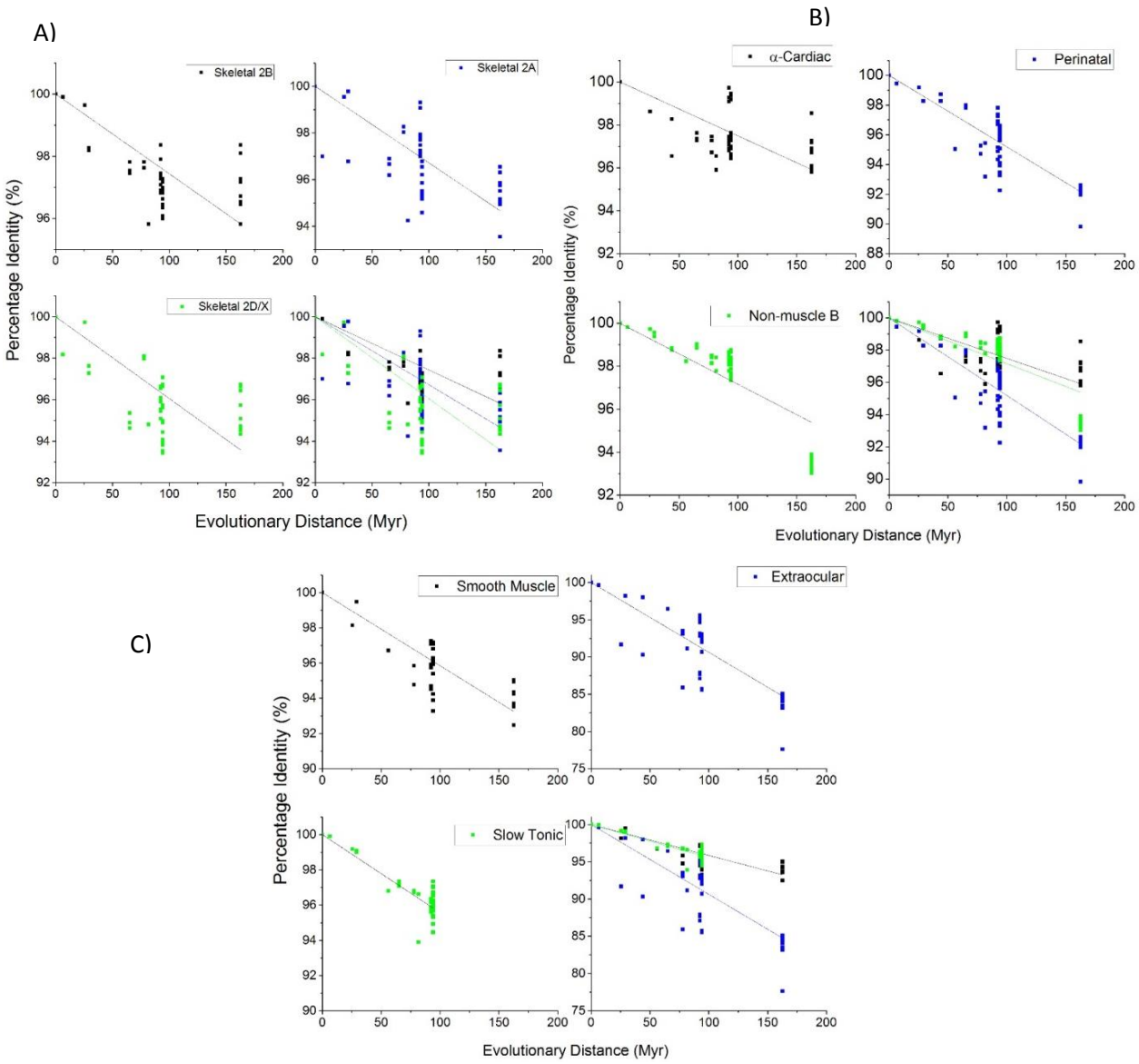


Figure 3.9 The percent identity matrix plots for the tail domains of the other isoforms when considering evolutionary distance. A) shows skeletal 2B, 2A and 2D/X isoforms where 2A is missing the Hamster and Tarsier sequence, 2B and 2D/X are missing the hamster and minke whale sequence. B) shows the α-cardiac, perinatal and non-muscle B isoform sequences. The α-cardiac sequence has the minke whale missing. C) shows smooth muscle, extraocular and slow tonic isoforms. Smooth muscle has the bonobo, tarsier, hamster and brandt's bat sequences missing, extraocular has the minke whale missing and slow tonic has the hamster and opossum data sequence missing.

The skeletal isoform tail domains show weaker relationships when considering evolutionary distance as a parameter than the motor domains. The larger scattering of data points suggests that these relationships are poorly explained with increasing evolutionary distance.

For the α-cardiac isoform in 3.9 B, there is less of a spread of data points suggesting that increasing evolutionary distance explains the divergence of the tail domain more so than the motor domain. The linear fit is not as strong in the tail domain however it shows a correlation. Perinatal and non-

muscle B both show strong linear relationships in their tail regions with relatively little scatter, suggesting that this evolutionary distance parameter explains the divergence of the domain reasonably well.

In figure 3.9 C, the relationships seen for smooth muscle, extraocular and slow tonic isoforms all show a reasonable trend. Relatively little scatter around the trend line indicates that the divergence of each isoforms tail domain is explained well as evolutionary distance increases.

3.9 Mass as a Parameter for the Divergence of the Tail Domain in Other Isoforms

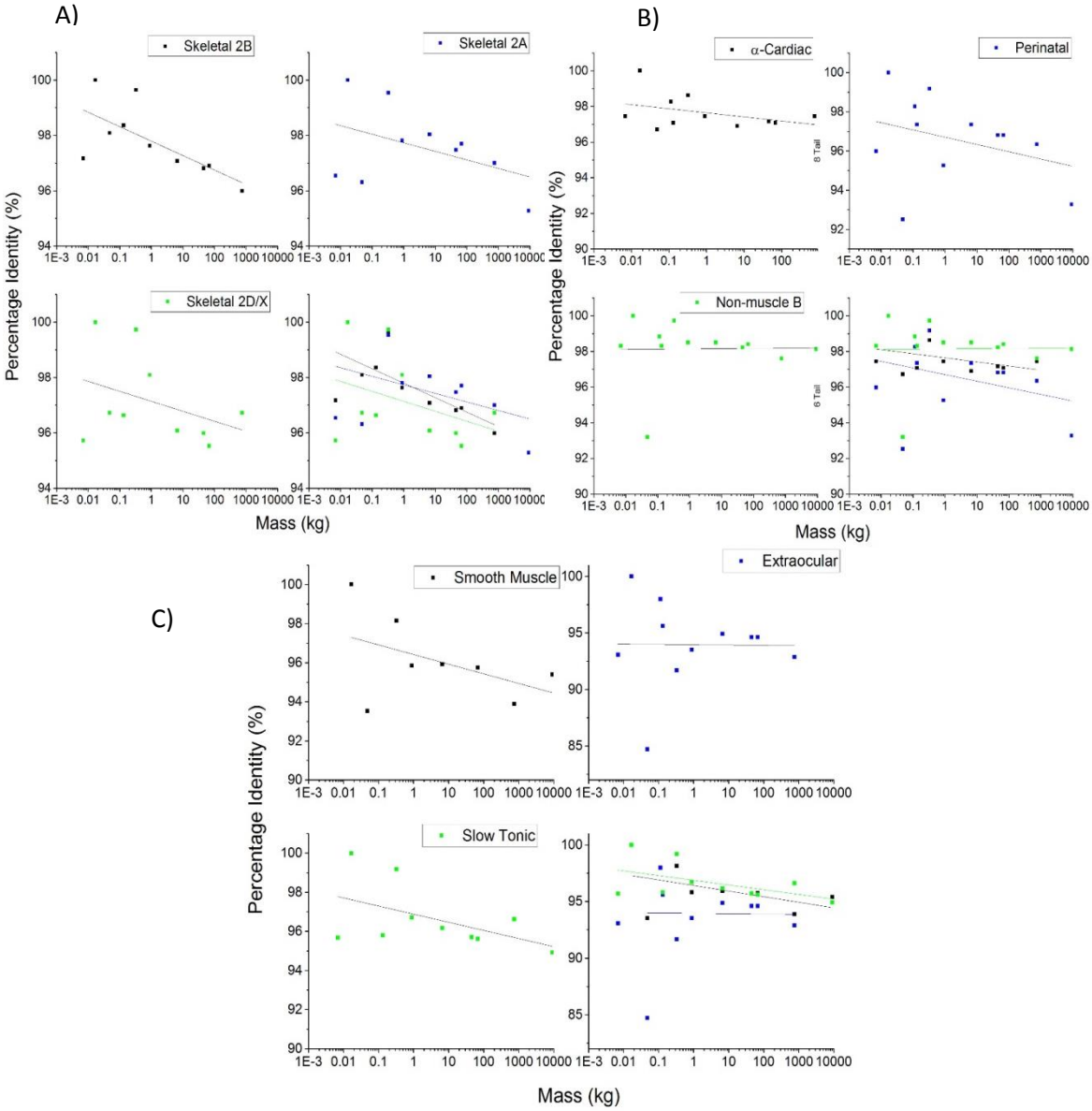


Figure 3.10 The percent identity matrix plots for the tail domains of the other isoforms when considering mass as a parameter. A) shows skeletal 2B, 2A and 2D/X isoforms where 2A is missing the Hamster and Tarsier sequence, 2B and 2D/X are missing the hamster and minke whale sequence. B) shows the α-cardiac, perinatal and non-muscle B isoform sequences. The α-cardiac sequence has the minke whale missing. C) shows smooth muscle, extraocular and slow tonic isoforms. Smooth muscle has the bonobo, tarsier, hamster and brandt's bat sequences missing, extraocular has the minke whale missing and slow tonic has the hamster and opossum data sequence missing. Plots are on a log scale.

For figure 3.10 A, using mass as a parameter to describe the divergence of the tail domains of the skeletal isoforms describes the relationships similarly to that of the evolutionary divergence plots. 2B has a slope of -0.36 and a correlation of -0.376, 2A has a correlation of -0.309 and a correlation of -0.443 and 2D/X has a slope of -0.522 and a correlation of -0.691. A weak correlation in 2D/X may indicate that mass has some dependence on the evolution of the tail domains, and weak correlations for 2A and 2B suggest no mass dependence.

Figure 3.10 B shows weak slopes for α-cardiac (-0.23), perinatal (-0.371) and non-muscle B (0.014) isoform tail domains. Correlation coefficients are -0.388, -0.330 and 0.016 respectively. These low coefficients suggest that mass has no role in determining the divergence of the tail domains of these isoforms.

Smooth muscle myosin shown in 3.10 C has a slope of -0.493 and a correlation coefficient of -0.471, extraocular has a slope of -0.028 and a correlation coefficient of -0.014, and slow tonic has a slope of -0.416 and a correlation of -0.506. Weak correlations seen in slow tonic and smooth muscle may indicate mass has a small role in determining the divergence of the tail domain of these isoforms, where no dependence is seen in the extraocular tail domain.

3.10 Comparing the heads and tails of the myosin isoform domains

By comparing all of the isoform motor and tail domains separately, it was clear to see that certain relationships were seen with either evolutionary distance or mass in one domain that were not seen in other domains. This lead to further question how these domains compared with the same isoform and to better analyse whether domains in an isoform as a whole diverge at faster or slower rates when considering evolutionary distance and mass.

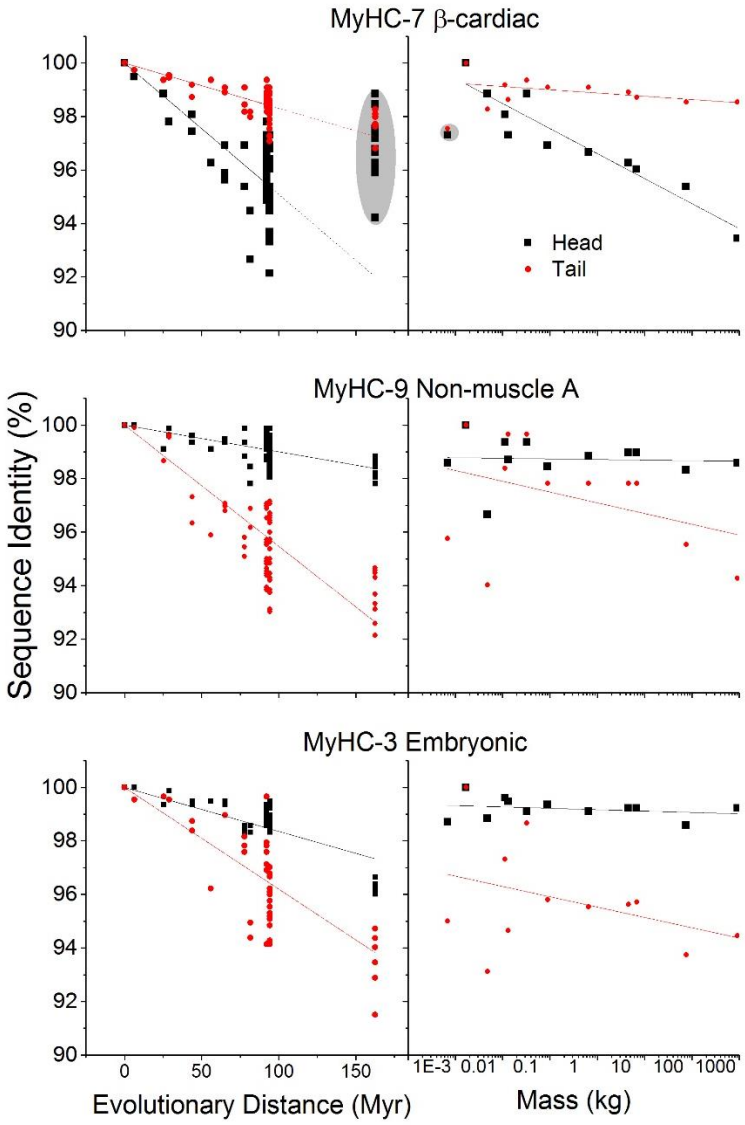


Figure 3.11 The motor domain and tail domain comparisons of each of the myosin isoforms. The β-cardiac, non-muscle A and embryonic myosin are all shown with the sequence identity and evolutionary distance on the left, and mass on the right. The mass plots are on a log scale. The head domains are marked in black squares and the tail domains are red circles, as indicated in the legend. The opossum data is shown but excluded from the analysis in the β-cardiac sequence due to odd data. The bat data are excluded from the β-cardiac mass plot due to differing physiology.

Figure 3.11 shows the motor and tail domains of β-cardiac myosin, non-muscle A and embryonic myosin. The evolutionary distance plot for β-cardiac myosin shows a better relationship in the tail domain than in the motor domain. The motor domain presents with a large spread of data suggesting that using the evolutionary distance of the species does not explain the protein divergence as well. The opposite is true when considering the mass data, the data points for the motor domain fit the trend well. When considering mass as a parameter for the tail domain, the data points fit the trend very well and show a high degree of conservation, suggesting that there is a high dependence on the protein to conserve its tail domain sequence as the motor domain diverges. The motor domain has a strong negative correlation in both the evolutionary distance plot (R = 0.958) and mass plot (R = -0.945) whereas the tail shows a strong correlation (R = 0.96) in terms of evolutionary distance, but not in mass (R = -0.471).

For non-muscle A, the tail is seen to diverge at a faster rate with a slope of -0.045 (R = 0.927) than the motor domain with a slope of -0.01 (R = 0.931) when considering evolutionary distance. There is scatter of data points around 94 Myr, however to a lesser extent as seen in β-cardiac. When considering the data points in mass order, the tail domain has a large spread of data that has no relationship (R = -0.376) observed. The motor domain of the protein has less scatter, but shows no relationship (R = -0.257), suggesting that mass does not have an effect on the conservation of the motor domain.

For the embryonic myosin in figure 3.11, the tail domain also diverges at a faster rate with a slope of -0.038 (R = 0.926) than the motor domain (slope of -0.016 R = 0.928) when considering evolutionary distance. Both domains show scattering of data points, suggesting the relationship is not explained using this parameter effectively. When considering mass, no relationships are seen through the presence of low correlation scores of -0.053 for the motor domain and -0.381 for the tail domain.

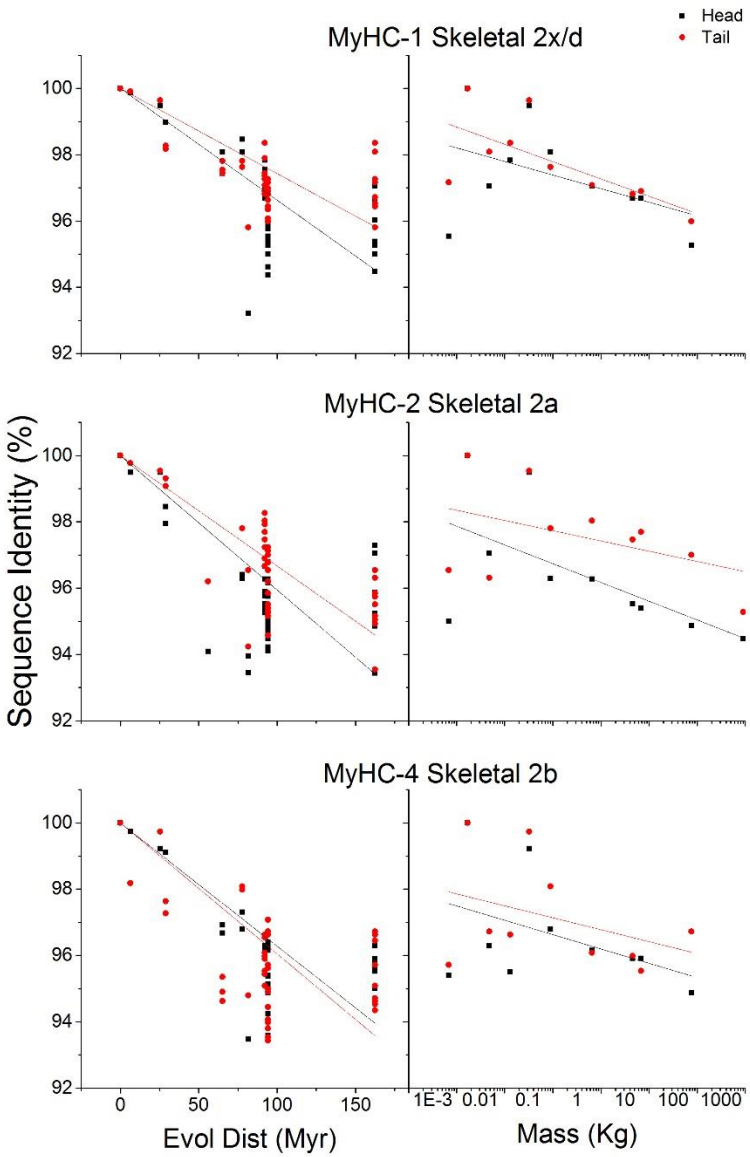


Figure 3.12 The motor domain and tail domain comparisons of each of the myosin isoforms. The skeletal 2D/X, 2A and 2B myosin are all shown with the sequence identity and evolutionary distance on the left, and mass on the right. The mass plots are on a log scale. The head domains are marked in black squares and the tail domains are red circles, as indicated in the legend.

For the skeletal isoforms, the evolutionary trends seen in both the motor domain and tail domains are similar, as can be seen in figure 3.12. For skeletal 2D/X, when considering evolutionary distance as a parameter, both the motor and tail domain show similar slopes of -0.034 (R = 0.918) and -0.026 (R = 0.92), suggesting that both domains co-evolved at a similar rate, with the motor domain diverging slightly faster. When considering mass, the motor and tail domains show a very weak relationship with correlation coefficients of -0.452 and -0.691 respectively.

The skeletal 2A isoform shows a strong evolutionary distance relationship in both the motor and tail domains, with slopes of -0.041 (R = 0.921) and -0.033 (R = 0.92) respectively. The motor domain appears to diverge at a slightly faster rate than the tail domain, however both show a strong linear regression. When considering mass, the strong relationship is lost, and it is the motor domain that appears to diverge at a faster rate than the tail domain. Weak correlation coefficients for the motor (R = -0.607) and tail (R = -0.443) domains suggest that there may be a weak relationship seen with the isoforms divergence as mass increases.

Skeletal isoform 2B in figure 3.12 shows similar divergence between the motor and tail domain when considering evolutionary distance, with slopes of -0.037 (R = 0.918) and -0.040 (R = 0.919). This suggests that both of the domains coevolved at similar rates. Scattering of data points suggest that there are other pressures on the evolution of this protein. When considering mass, the relationship is lost with poor correlation coefficients in the motor domain of -0.435 and -0.376 in the tail domain.

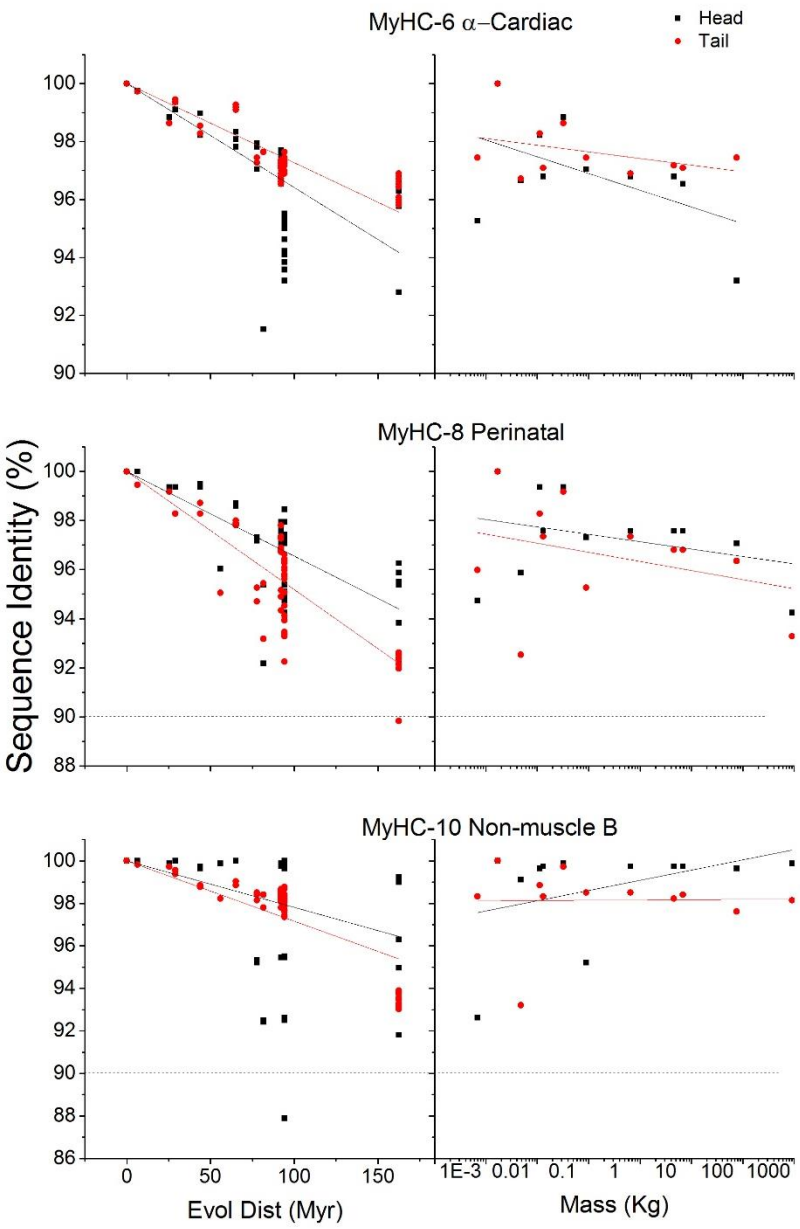


Figure 3.13 The motor domain and tail domain comparisons of each of the myosin isoforms. The α-cardiac, perinatal and non-muscle B myosin are all shown with the sequence identity and evolutionary distance on the left, and mass on the right. The mass plots are on a log scale. The head domains are marked in black squares and the tail domains are red circles, as indicated in the legend. The dotted lines on both perinatal and non-muscle B isoforms are to indicate a 90% sequence identity.

For the α-cardiac isoform, the tail domain diverges at a slower rate (-0.027) than the motor domain (-0.036) when considering evolutionary distance, similar to that of the β-cardiac isoform. The rate of divergence is not as great as the β-cardiac isoform. The trend appears to explain the motor domain divergence poorly, as there is a large spread of data at 92 Myr. The tail domains data has little scatter. When considering mass as a parameter, the same relationship holds true, where the motor domain diverges at a faster rate than the tail domain, as seen in the β-cardiac isoform,

however a weak correlation coefficient for the motor domain (-0.528) makes it difficult to draw any conclusions. The poor correlation coefficient for the tail domain of -0.388 suggests that the mass dependence on the divergence of this isoform is not as important.

The perinatal isoform appears to show a strong correlation between the divergence of the protein and the evolutionary distance of the species. Both the motor and tail domain appear to diverge at fast rates with slopes of -0.034 and -0.048 respectively, with the tail domain diverging more so. The large standard error of the slopes reflect the *Myotis brandtii* and *Mesocricetus auratus* sequences that diverge at rates not following the trend, possibly due to differing physiologies. A large majority of the sequences show conservation in the 99% range. When taking mass in to consideration, both the motor and tail domain lose their relationships with low correlation coefficients of -0.332 and -0.33 respectively. This suggests that the increasing mass of a species does not impact the evolutionary pressures of the isoform.

For non-muscle B, there is a large divergence of the motor domain around 92 Myr when considering evolutionary distance as a parameter. This indicates that the tail domain diverges at a faster rate than the motor domain through the slopes, but the large standard error is due to the divergence of *Myotis brandtii*'s motor domain. When considering mass as a parameter, the tail domain appears to show a degree of conservation, but a poor correlation coefficient of 0.016 suggests no relationship. The motor domain shows no relationship with mass, suggesting that it does not explain the divergence of the non-muscle B isoform in either of the domains.

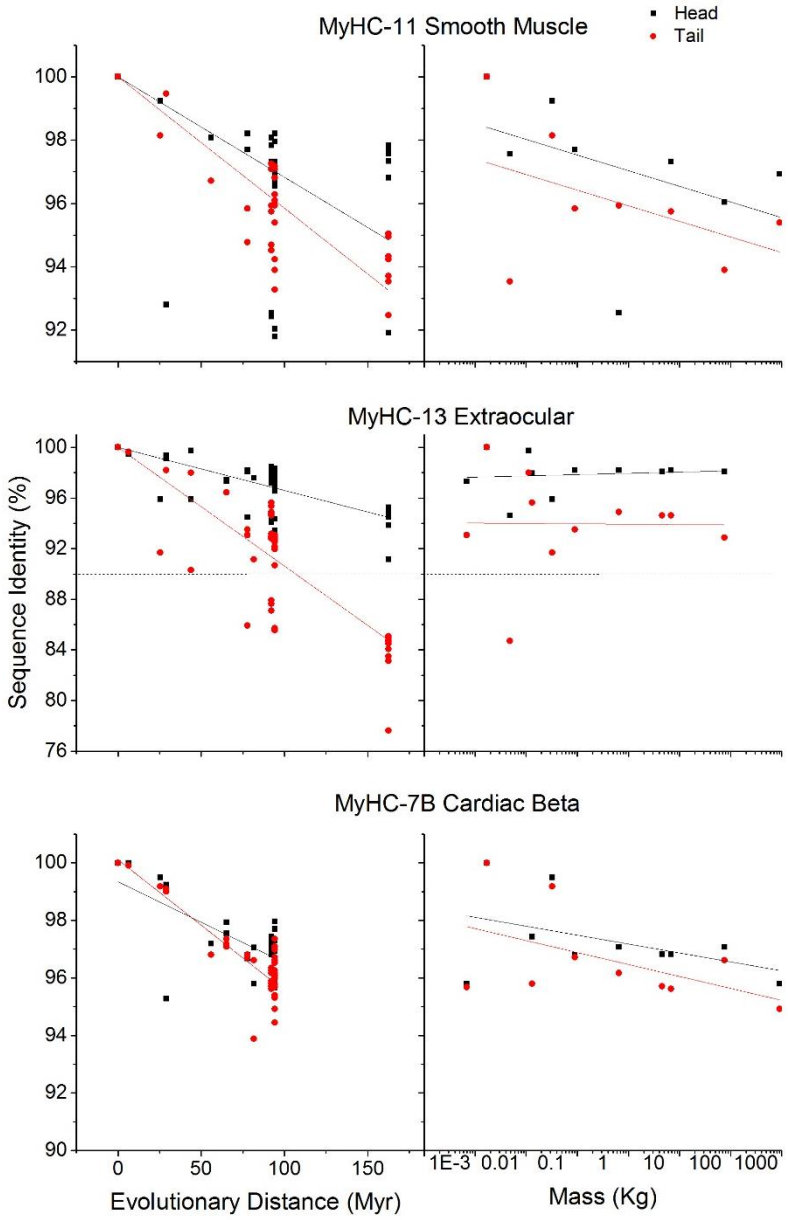


Figure 3.14 The motor domain and tail domain comparisons of each of the myosin isoforms. The smooth muscle, extraocular and7B Cardiac Beta myosin are all shown with the sequence identity and evolutionary distance on the left, and mass on the right. The mass plots are on a log scale. The head domains are marked in black squares and the tail domains are red circles, as indicated in the legend. The dotted lines in the extraocular graphs indicate a 90% sequence conservation.

In figure 3.14, for the smooth muscle isoform, a large spread of data is seen in the motor domain, which can possibly be explained by the isoforms innate process of being alternatively spliced (Haase, H., Morano, I. 1996). This may account for the grouping of motor domain data points around the 92% sequence identity, and reflects a high standard error of 0.005. When considering mass, both domains show large scattering and poor correlation coefficients of -0.446 for the motor

and -0.471 for the tail domain, suggesting that mass is not important in determining the divergence of the protein domains.

The extraocular isoform shows a strong divergence in the tail domain with a slope of-0.094 when considering evolutionary distance. With the tail domain diverging at a rate much faster than the motor domain (-0.034) and little scatter of data points, it suggests that evolutionary divergence is responsible for the divergence of this domain. The motor domain data points also have little scatter and a generous slope of -0.034, suggesting time also plays a role in the divergence of the domain. When considering mass, there is no relationship observed of the protein in either the motor or tail domain, suggesting that mass is not a factor for this isoform.

The slow tonic isoform shows a reasonable trend in both the motor and tail domains when considering evolutionary distance. The tail domain appears to diverge at a faster rate than the motor domain with slopes of -0.036 and -0.043 respectively, however there appears to be a spread of data around the 92 Myr point, suggesting the relationship is poorly defined. When considering mass, the relationships are weak and suggest mass has no/little influence on the divergence of the protein.

Table 3.4 shows the evolutionary divergence graph analysis for both the head and tail domains. This data includes the standard error and Pearson's correlation coefficient. Table 3.5 shows similarly derived data for the mass data plots discussed.

Correlation Coefficient Data for Evolutionary Divergence								**Number of Species (*n*)**
		Motor Domain			**Tail Domain**			
Isoform	**MyHC**	Slope	SE	R	Slope	SE	R	
β-Cardiac	7	-0.049	0.002	0.958	-0.017	0.001	0.960	12
Non-Muscle A	9	-0.010	0.001	0.931	-0.045	0.001	0.927	12
Embryonic	3	-0.016	0.001	0.928	-0.038	0.001	0.926	12
Skeletal 2D/X	1	-0.034	0.002	0.918	-0.026	0.002	0.920	10
Skeletal 2A	2	-0.041	0.002	0.921	-0.033	0.002	0.920	10
Skeletal 2B	4	-0.037	0.002	0.918	-0.040	0.002	0.919	10
Alpha Cardiac	6	-0.036	0.002	0.921	-0.027	0.001	0.921	11
Perinatal	8	-0.034	0.002	0.928	-0.048	0.002	0.925	12
Non-Muscle B	10	-0.022	0.004	0.930	-0.028	0.001	0.925	12
Smooth Muscle	11	-0.032	0.005	0.920	-0.042	0.002	0.916	8
Extraocular	13	-0.034	0.002	0.920	-0.094	0.004	0.907	11
Slow Tonic	7B	-0.036	0.001	-0.956	-0.043	0.001	-0.955	10

Table 3.4 The linear slopes of the lines of best fit for each of the isoforms when considering evolutionary distance. The calculated standard error (SE) and Pearson's correlation coefficient (R) are all included to three significant figures. These have been calculated for both the motor and tail domains, with the number of species for each isoform included.

Correlation coefficient data for Mass vs Sequence Divergence								
		Motor Domain			Tail Domain			Number of Species (*n*)
Isoform	**MyHC**	Slope	SE	R	Slope	SE	R	
β-Cardiac	7	-0.936	0.108	-0.945	-0.122	0.076	-0.471	12
Non-Muscle A	9	-0.052	0.062	-0.257	-0.386	0.301	-0.376	12
Embryonic	3	-0.022	0.132	-0.053	-0.403	0.309	-0.381	12
Skeletal 2D/X	1	-0.410	0.286	-0.452	-0.522	0.193	-0.691	10
Skeletal 2A	2	-0.566	0.262	-0.607	-0.309	0.221	-0.443	10
Skeletal 2B	4	-0.433	0.317	-0.435	-0.360	0.314	-0.376	10
Alpha Cardiac	6	-0.577	0.310	-0.528	-0.230	0.182	-0.388	11
Perinatal	8	-0.301	0.271	-0.332	-0.371	0.336	-0.330	12
Non-Muscle B	10	0.481	0.350	0.398	0.014	0.277	0.016	12
Smooth Muscle	11	-0.494	0.405	-0.446	-0.493	0.377	-0.471	8
Extraocular	13	0.101	0.310	0.107	-0.028	0.794	-0.012	11
Slow Tonic	7B	-0.310	0.220	-0.446	-0.416	0.251	-0.506	10

Table 3.5 The linear slopes of the lines of best fit for each of the isoforms when considering mass as a parameter. The calculated standard error (SE) and Pearson's correlation coefficient (R) are all included to three significant figures. These have been calculated for both the motor and tail domains, with the number of species for each isoform included.

3.11 Determining Whether Sample Sizes were Representative

Many trends observed were from a small sample set that may have biased any results seen. In order to test whether these sample sizes were representative, sequences for as many species as possible were collected in order to determine if the relationships seen still held true. Figure 3.15 represents the percent identity matrices that were generated for the β-cardiac sequences.

These matrices were generated in terms of evolutionary distance from humans. The overall trend of the datasets indicated a similar relationship to that of the smaller 12 mammals that were previously studied (figure 3.1 and 3.4), where the motor domain of β-cardiac myosin shows more divergence than the tail domain that is highly conserved. As the relationships are seen to be conserved in mammals, it was clear that it was not seen in other taxonomic groups. For reptiles and birds, there appears to be similar divergence rates in both the motor and tail domain. Fish showed their sequence divergence to be so great and their physiology so different, the protein was not comparable.

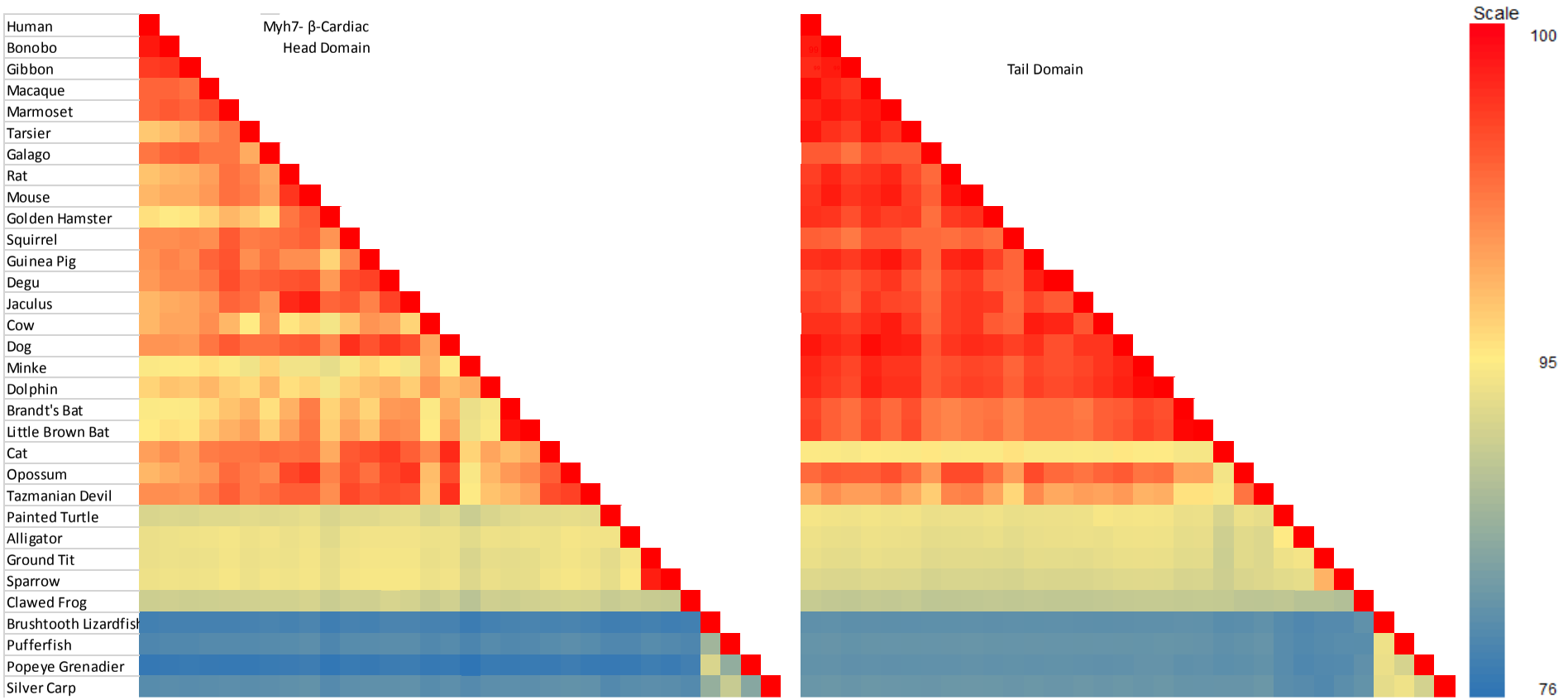


Figure 3.15 The percent identity matrices for 32 β-cardiac myosin motor and tail domains. As many sequences were collected as possible from Uniprot. Red indicates a 100% sequence conservation spanning down to blue with a 76% sequence conservation.

In order to determine if the relationships seen in these domains were reflected when considering mass as a parameter, the sequence identities were generated and plotted, shown in figure 3.16. Mammals see the relationship holds true, where the motor domain diverges at a faster rate than the tail domain and the data points have little scatter. Through separating out the reptiles, birds and fish sequences, it was clear to see that these sequences did not fit in with the mass relationship observed in mammals and reinforced the fact this phenomenon was exclusive to mammals.

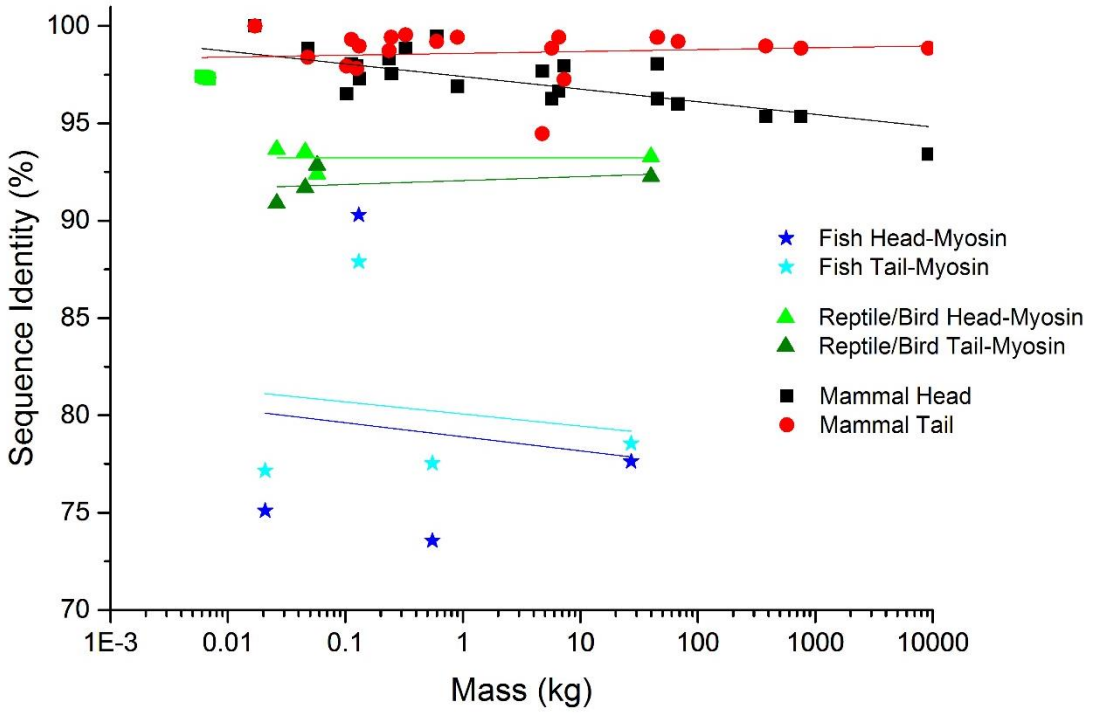


Figure 3.16 The percent identity scores each β-cardiac sequence shares with one another plotted using mass as a parameter. Red circles and black squares indicate the mammalian sequences, green triangles represent reptile and bird sequences, blue stars indicate fish sequences. The two bat mammalian sequences are shown but excluded from the analysis, and are marked in green circle and squares.

Determining if the relationship was exclusive to mammals led to further investigations to see whether the data points used in the original analysis were representative of mammals as a whole. By combining the data used with the new sequences obtained, the relationships could be determined, as seen in figure 3.17.

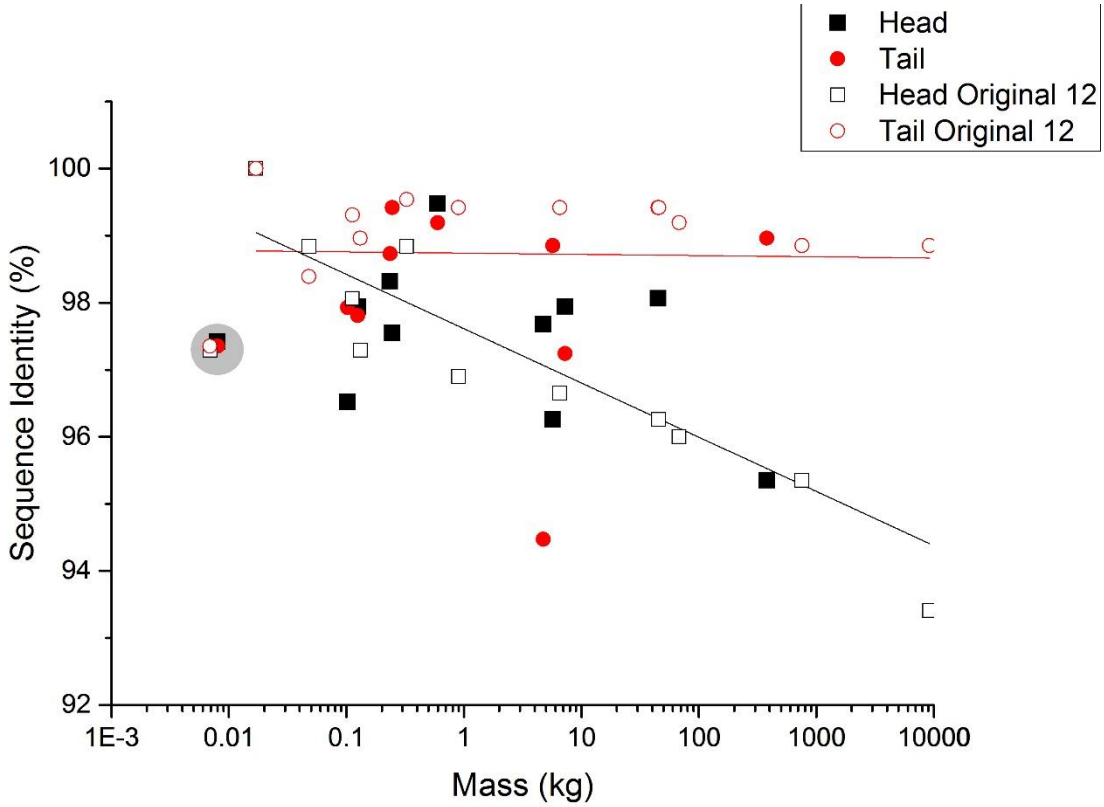


Figure 3.17 The percent identity scores for 23 mammalian β-cardiac sequence shares with one another plotted using mass as a parameter. These plots allowed the generation of linear relationship to show any correlations. Original data points from the 12 species used are indicated with clear squares and circles, and the additional 11 sequences are indicated in black squares and circles for motor and tail domains respectively. The brandt's bat and little brown bat sequences are shown but excluded from the dataset and are surrounded with a grey circle.

The strong trend seen indicates that the 12 species chosen for previous analysis are representative of a larger set of mammals. With the extra sequences fitting the trend very well and complementing the already established data points for both the motor and tail domains, confidence was established in the previous sample size. The motor domain showed a strong correlation and divergence with increasing mass, and the tail domain maintains a higher degree of conservation as the motor domain diverges, as seen in table 3.6.

Correlation coefficient data for Mass vs Sequence Divergence in the β-cardiac isoform with 23 mammalian sequences								
Isoform	**MyHC**	Slope	SEM	R	Slope	SEM	R	**Number of Species (*n*)**
β-Cardiac	7	-0.809	0.136	-0.808	-0.018	0.175	-0.024	23

Table 3.6 The linear slopes of the lines of best fit for the 23 mammalian isoforms when considering mass as a parameter.

Percent identity matrices were generated for both non-muscle A and embryonic isoform mammalian sequences to further determine if similar relationships were seen. Figures 3.18 and 3.19 show similar trends to that observed previously in the twelve species set, including 19 species sequences for each isoform. In the motor domain of non-muscle A in figure 3.18, sequences are more conserved than in the tail domain where there is a higher degree of divergence. In figure 3.19, both the motor and tail domains of the embryonic myosin show similar degrees of conservation, as previously observed (figure 3.4).

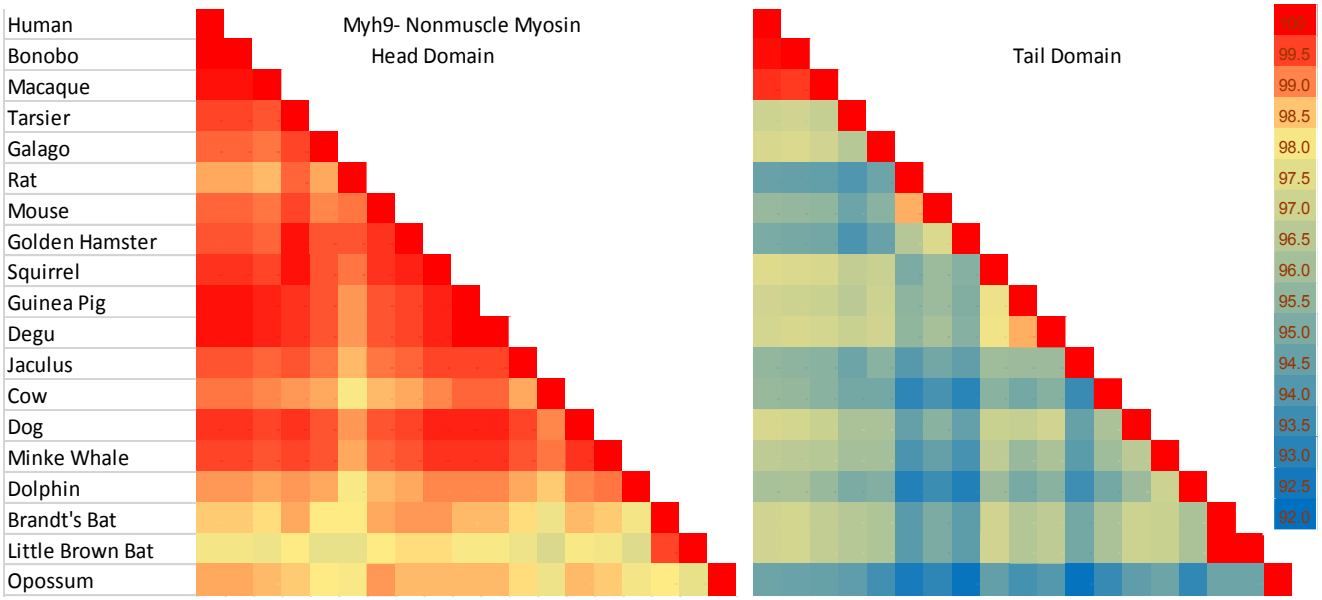


Figure 3.18 The percent identity matrices for 19 mammalian non-muscle A myosin motor and tail domains. As many sequences were collected as possible from Uniprot. Red indicates a 100% sequence conservation spanning down to blue with a 92% sequence conservation.

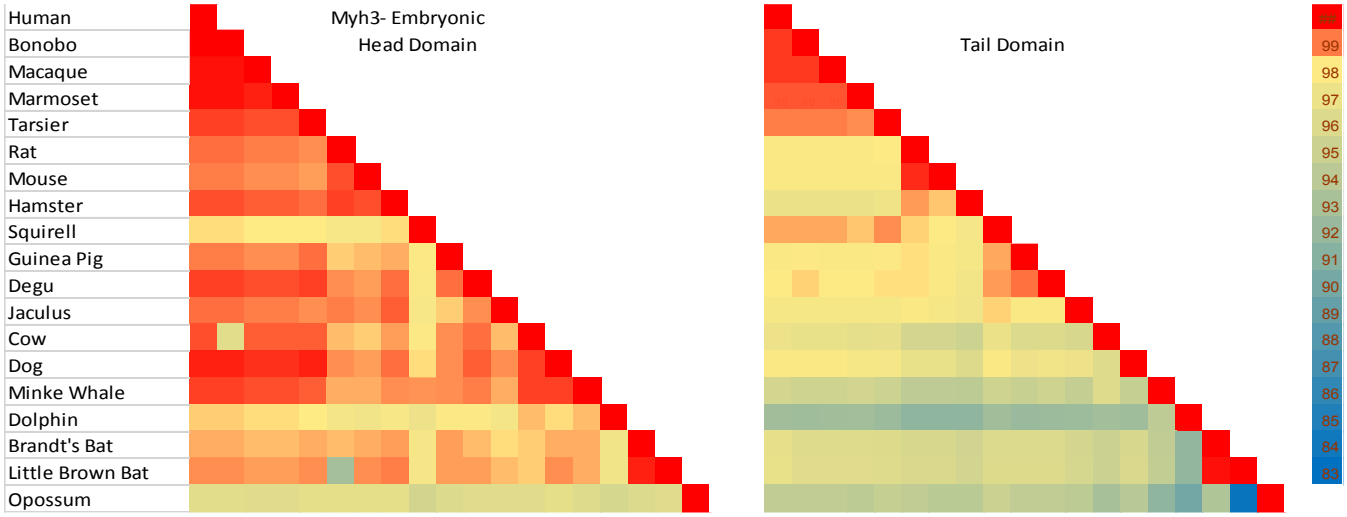


Figure 3.19 The percent identity matrices for 20 embryonic myosin motor and tail domains. As many sequences were collected as possible from Uniprot. Red indicates a 100% sequence conservation spanning down to blue with an 88% sequence conservation.

After analysing the relationships between isoforms and their domains, determining where the variations occurred within the protein structures was next. After our datasets proved representative, a perl script was written to calculate for each position both the most frequent amino acid and the number of amino acids present at each residue position. This was done for β-cardiac, non-muscle A and embryonic sequences (*n*=12).

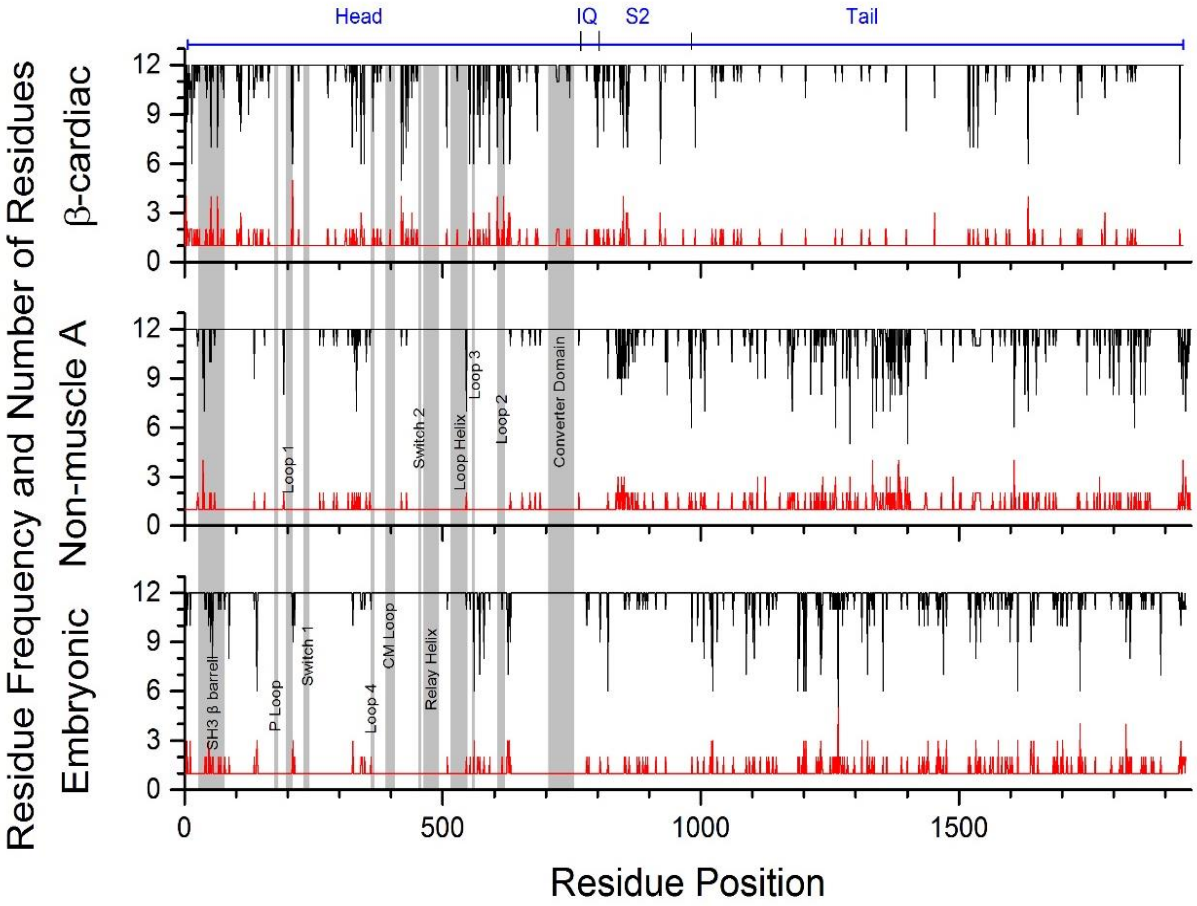


Figure 3.20 The variation analysis perl script output. The number of residues at a specific site and the number of sequences sharing the same amino acid were counted. The whole protein sequence was analysed and is divided in to its various domains via the blue line. The black line indicates the number of sequences that share the same residue at that point and the red line indicates how many amino acids are present at that site. Key structures of interest were marked on and are shaded over to highlight and see if any variations are seen at these sites. These analyses were done for β-cardiac, non-muscle A and embryonic myosin as indicated.

The whole protein was analysed in order to determine where variations occurred in the structures. Important structural sites of the motor domain were marked on to highlight any regions of importance that could give an insight in to the impact these variations would have on the protein.

Where the number of sequences sharing the same amino acid (figure 3.20 black line) vary, clusters of variation can be seen in the motor domain of the β-cardiac domain that are not seen in non-muscle A and embryonic isoforms. The inverse is true of these two isoforms, where there is greater variation in the tail domain than the motor domain. This further highlights the variation seen from the percent identity matrices.

Through marking on the key structures of the motor protein, relatively little conclusions were determined from this. Clustering of variation outside of these regions were more prominent than variations inside them. This may indicate that residues affecting near-by sites of key structural elements of the protein *in-vivo* have a more important role in determining its speed than mutations inside them.

For the 12 species sample size, counting the number of amino acids at each residue provided further insight in to any biochemical properties of the variations. Where there appears to be a sequence variation, the number of amino acids are shown at that site (red line, figure 3.20). This quantitative analysis allowed for the identification of sites that had multiple residues at a particular site, or for the identification of sites that has a variation in its sequence that was also seen in other species.

Identified Residues from Myh7 Sequence Divergence Structure Plot That Have 4+ Residue Difference

Residue Position	Within a Structure?	Amino Acids	Number of Amino Acids at Position			
51	SH3 β-Barrell	V L M I	**7 V**	3 L	1 M	1 I
64	SH3 β-Barrell	Y H N F	1 Y	3 H	**7 N**	1 F
209	Loop 1	S G N T A	1 S	1 G	3 N	**6 T** 1 A
420	CM Loop	I A S V	2 I	**5 A**	**4 S**	1 V
606	No	G A D E	**7 G**	3 A	1 D	1 E
618	Before Loop 2	T N S I	4 T	**6 N**	1 S	1 I
850	No	S T N L	2 S	**7 T**	N 1	L 2

Table 3.7 The identified residues from the protein conservation prediction plot that have 4+ residues at a particular site for the β-cardiac sequence. The number of sequences that contain the amino acid are listed next to the single letter identifier of the amino acid.

For the β-cardiac sequence, where residue locations had 4 or more variations, these were noted and analysed in order to determine their biochemical properties, as shown in table 3.7. Four or more variations at a single point appeared to be high, warranting investigations as to what these residues were. At residue 51, all amino acids present are hydrophobic, suggesting that there would be little impact on the proteins structure. Residue 64 has mutations that vary from hydrophobic residues to positively charged residues and polar residues. This site may play a role in affecting the proteins structure. Position 209 sees the presence of five amino acids at that point, which is high. It is only two sequences that have serine and glycine present, suggesting that these residues are not conserved between species.

As this small sample size generated data of interest and larger datasets were available, the script was also done for the larger sample sizes that were collected. The plot with larger sample sizes (figure 3.21) did not generate results that were significantly different from the data for the 12

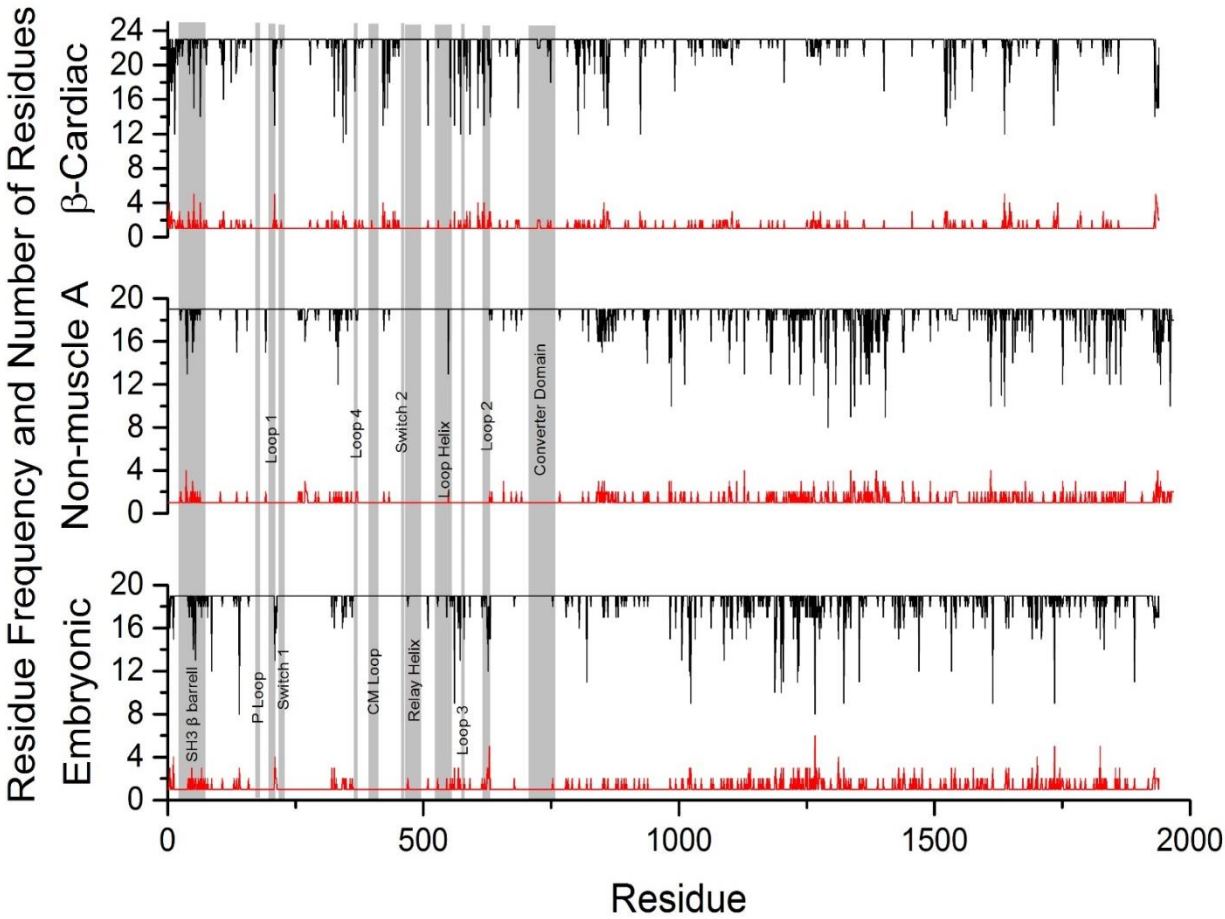


Figure 3.21 The variation analysis perl script output for the large sets of mammalian sequences. The number of residues at a specific site and the number of sequences sharing the same amino acid were counted for β-cardiac (*n*=23), non-muscle A (*n*=19) and embryonic (*n*=19) isoforms. The black line indicates the number of sequences that share the same residue at that point and the red line indicates how many amino acids are present at that site.

species, further suggesting that the sample size was representative of the larger set of mammals. Similar patterns are seen, but are amplified due to the presence of more sequences. The motor domain in the β-cardiac sequence appears to show more variation than the tail domain, and the inverse is true for non-muscle A and embryonic isoforms.

3.13 Residues of Interest

Through identifying areas of variation that lead to significant interest, analysis was done on the multiple sequence alignments of species ordered in mass size. Residue locations that contained two or more amino acids that appeared to show a pattern of conserved residues with smaller/larger mammals that varied as mass increased/decreased.

These 35 identified sites were all chosen due to them showing a trend with mass, some more clear than others (table 3.8). Residues 77, 434, 610, 616, 627 and 629 are all sites that have amino acids conserved in smaller mammals. Residues 20, 164, 366, 569, and 580 are sites that have amino acids conserved in larger mammals. These variable patterns between smaller and larger mammals may give an insight in to why contraction speeds of the β-cardiac isoform varies as mass increases.

Residue 77 represents a residue site that shows high conservation rates in mammals of smaller sizes, and a variation in the sequence as average adult body mass increases, in the dolphin, cow and minke whale. This residue variation from methionine to leucine is a shift from a methionine to leucine, which are both hydrophobic side chains. This variation can therefore be predicted to have less of an effect on the structure of the protein, but may interact with surrounding residues which would alter its activity. This site is located in the SH3 β –barrell that may affect the proteins activity.

Residue 125 represents a residue site that shows conservation of valine in smaller mammals apart from the brandt's bat and the little brown bat, and isoleucine in the larger dolphin, cow and minke whale sequences. Here, where mass may be indicative of the presence of an isoleucine in the larger mammals, the importance of the physiology of the mammal is also highlighted. This residue may play a role in affecting the activity of the protein due to the presence of isoleucine in the bats that have shown to have a rate of divergence in the β-cardiac sequence that does not follow the trend

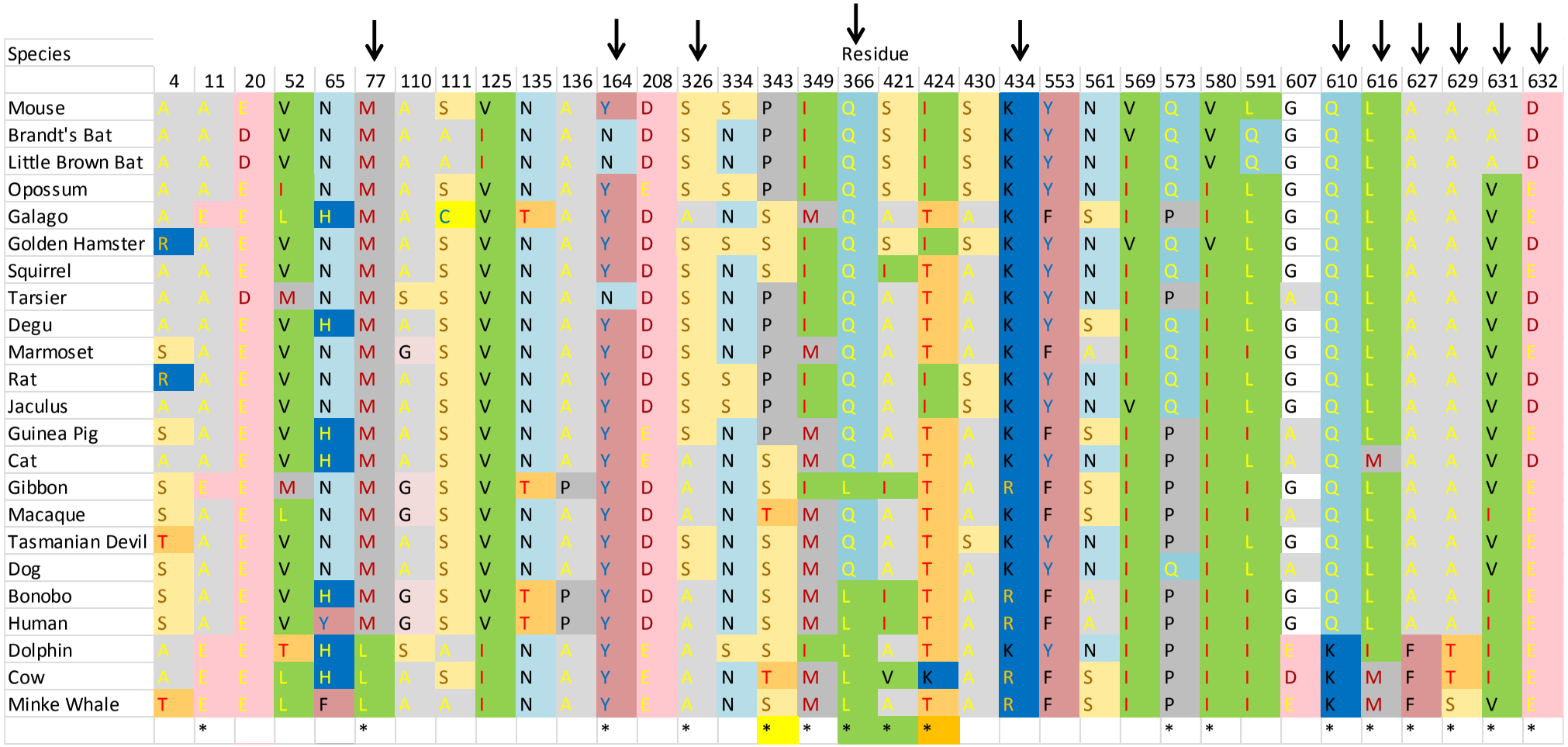
seen, as discussed. Valine and isoleucine are both hydrophobic residues, meaning the variation may not affect the structure of the protein, but may affect interactions with surrounding residues.

Residue 627 highlights a residue that is conserved throughout smaller mammals and varies in the larger dolphin, cow and minke whale. This residue variation from alanine to phenylalanine introduces a bulky residue to the sequence that may affect the proteins structure. Although both negatively charged residues, alanine does not contain an aromatic ring whereas phenylalanine does. This may interact with surrounding residues at the site, or, due to its location within loop 2, may restrict its movement.

Residue 610 highlights a residue that is conserved throughout smaller species and varies in the larger dolphin, cow and minke whale sequences. The change from glutamine, a polar uncharged residue, to the positively charged lysine residue may impact both the structure of the protein and the interactions with other residues surround the site in the protein. This site is located near to loop 2, however has the possibility form electrostatic interactions with other residues due to the tertiary structure of the protein.

Residue 20 demonstrates a residue that is conserved throughout larger mammals and varies in the smaller tarsier, brandt's bat and little brown bat. Here, the change from glutamic acid to aspartic acid would have little impact on the interactions of the site with surrounding residues due to them both being negatively charged. This site is located before the SH3 β-barrell, early in the motor domain sequence, which may have little effect on the interactions of the protein.

Residue 135 reflects a relationship of evolutionary divergence as opposed to mass. Here, the primates of the dataset conserve a threonine in their sequence, whereas other species conserve an asparagine. Both of these residues have polar uncharged side chains, in which the variation would not cause large disruption to the proteins structure surrounding the site. The variation may interact with residues when in tertiary conformation.



Species	4	11	20	52	65	77	110	111	125	135	136	164	208	326	334	343	349	366	421	424	430	434	553	561	569	573	580	591	607	610	616	627	629	631	632
Mouse	A	A	E	V	N	M	A	S	V	N	A	Y	D	S	S	P	I	Q	S	I	S	K	Y	N	V	Q	V	L	G	Q	L	A	A	A	D
Brandt's Bat	A	A	D	V	N	M	A	A	I	N	A	N	D	S	N	P	I	Q	S	I	S	K	Y	N	V	Q	V	Q	G	Q	L	A	A	A	D
Little Brown Bat	A	A	D	V	N	M	A	A	I	N	A	N	D	S	N	P	I	Q	S	I	S	K	Y	N	I	Q	V	Q	G	Q	L	A	A	A	D
Opossum	A	A	E	I	N	M	A	S	V	N	A	Y	E	S	S	P	I	Q	S	I	S	K	Y	N	I	Q	I	L	G	Q	L	A	A	V	E
Galago	A	E	E	L	H	M	A	C	V	T	A	Y	D	A	N	S	M	Q	A	T	A	K	F	S	I	P	I	L	G	Q	L	A	A	V	E
Golden Hamster	R	A	E	V	N	M	A	S	V	N	A	Y	D	S	S	S	I	Q	S	I	S	K	Y	N	V	Q	V	L	G	Q	L	A	A	V	D
Squirrel	A	A	E	V	N	M	A	S	V	N	A	Y	D	S	N	S	I	Q	I	T	A	K	Y	N	I	Q	I	L	G	Q	L	A	A	V	E
Tarsier	A	A	D	M	N	M	S	S	V	N	A	N	D	S	N	P	I	Q	A	T	A	K	Y	N	I	P	I	L	A	Q	L	A	A	V	D
Degu	A	A	E	V	H	M	A	S	V	N	A	Y	D	S	N	P	I	Q	A	T	A	K	Y	S	I	Q	I	L	G	Q	L	A	A	V	D
Marmoset	S	A	E	V	N	M	G	S	V	N	A	Y	D	S	N	P	M	Q	A	T	A	K	F	A	I	Q	I	I	G	Q	L	A	A	V	E
Rat	R	A	E	V	N	M	A	S	V	N	A	Y	D	S	S	P	I	Q	A	I	S	K	Y	N	I	Q	I	L	G	Q	L	A	A	V	D
Jaculus	A	A	E	V	N	M	A	S	V	N	A	Y	D	S	S	P	I	Q	A	I	S	K	Y	N	V	Q	I	L	G	Q	L	A	A	V	D
Guinea Pig	S	A	E	V	H	M	A	S	V	N	A	Y	E	S	N	P	M	Q	A	T	A	K	F	S	I	P	I	I	A	Q	L	A	A	V	E
Cat	A	A	E	V	H	M	A	S	V	N	A	Y	E	A	N	S	M	Q	A	T	A	K	Y	N	I	P	I	L	A	Q	M	A	A	V	D
Gibbon	S	E	E	M	N	M	G	S	V	T	P	Y	D	A	N	S	I	L	I	T	A	R	F	S	I	P	I	I	G	Q	L	A	A	V	E
Macaque	S	A	E	L	N	M	G	S	V	N	A	Y	D	A	N	T	M	Q	A	T	A	K	F	S	I	P	I	I	A	Q	L	A	A	I	E
Tasmanian Devil	T	A	E	V	N	M	A	S	V	N	A	Y	D	S	N	S	M	Q	A	T	S	K	Y	N	I	P	I	L	G	Q	L	A	A	V	E
Dog	S	A	E	V	N	M	A	S	V	N	A	Y	D	S	N	S	M	Q	A	T	A	K	Y	N	I	Q	I	L	A	Q	L	A	A	V	E
Bonobo	S	A	E	V	H	M	G	S	V	T	P	Y	D	A	N	S	M	L	I	T	A	R	F	A	I	P	I	I	G	Q	L	A	A	I	E
Human	S	A	E	V	Y	M	G	S	V	T	P	Y	D	A	N	S	M	L	I	T	A	R	F	A	I	P	I	I	G	Q	L	A	A	I	E
Dolphin	A	E	E	T	H	L	S	A	I	N	A	Y	E	A	S	S	I	L	A	T	A	K	Y	N	I	P	I	I	E	K	I	F	T	I	E
Cow	A	E	E	L	H	L	A	S	I	N	A	Y	E	A	N	T	M	L	V	K	A	R	F	S	I	P	I	I	D	K	M	F	T	I	E
Minke Whale	T	E	E	L	F	L	A	A	I	N	A	Y	E	A	N	S	M	L	A	T	A	R	F	S	I	P	I	I	E	K	M	F	S	V	E
		*				*						*		*		*	*	*	*	*						*	*			*	*	*	*	*	*

Table 3.8 showing residues of interest that were identified for the large β-cardiac set of sequences. The species and residue numbers are indicated. The colour co-ordination of these sequences indicate residues that share similar properties. Pink shaded boxes indicate negatively charged residues. Green shaded boxes indicate hydrophobic residues. Yellow shaded boxes indicate polar uncharged side chains. Red shaded boxes indicate residues with aromatic rings. White boxes indicate small residue side chains. Dark grey boxes indicate proline and methionine. Blue shaded boxes represent positively charged amino acids. Asterisks' indicate residue type shifts. Arrows indicate residues that show strong patters of mass dependence.

3.14 DNA Synonymous and Non-synonymous Variation Analysis

Through detailing possible mutations that may lead to the differences in contraction rates of β-cardiac myosin in mammals, it was questioned whether mutation rates seen in the DNA of the β-cardiac isoform reflected these changes. Due to the degenerate nature of the genetic code, mutations seen in the nucleotide sequences of proteins may not result in the change of an amino acid. Where this change is not disadvantageous to an organism, natural selection over evolution will retain it. Where the primary sequence of the protein reflects non-synonymous, residue changing mutations, analysing the DNA sequence allows for synonymous mutations to be observed. Investigating these nucleotide sequences allows for the analysis of selection pressure within a given system. Through comparing the 12 mammalian species DNA sequences, these levels of synonymous and non-synonymous mutations could be observed, as seen in table 3.9.

Species sequence comparisons		Myr	Synonymous (dS)	Non-synonymous (dN)	Ratio of dS/dN
Mouse	Brandt's Bat	94.2	441.0	53.0	39.8
Mouse	Opossum	162.6	704.8	63.2	74.1
Mouse	Tarsier	65.2	424.5	41.5	48.5
Mouse	Rat	25.4	252.0	7.0	149.4
Mouse	Guinea Pig	77.9	447.5	48.5	44.5
Mouse	Macaque	92.3	405.2	42.8	43.9
Mouse	Bonobo	92.3	407.7	51.3	36.8
Mouse	Human	92.3	409.2	56.8	33.4
Mouse	Cow	94.2	437.5	121.5	16.9
Mouse	Minke whale	94.2	414.0	78.0	24.6
Brandt's Bat	Opossum	162.6	680.5	94.5	46.0
Brandt's Bat	Tarsier	94.2	366.0	60.0	27.5
Brandt's Bat	Rat	94.2	431.0	56.0	36.4
Brandt's Bat	Guinea Pig	94.2	422.7	75.3	26.4
Brandt's Bat	Macaque	94.2	348.7	65.3	23.7
Brandt's Bat	Bonobo	94.2	346.7	75.3	20.3
Brandt's Bat	Human	94.2	347.7	75.3	20.4
Brandt's Bat	Cow	81.6	375.2	134.8	12.4
Brandt's Bat	Minke whale	81.6	329.2	89.8	15.9
Opossum	Tarsier	162.6	713.3	84.7	57.0

Opossum	Rat	162.6	711.8	57.2	83.5
Opossum	Guinea Pig	162.6	701.3	79.7	58.2
Opossum	Macaque	162.6	681.8	81.2	53.7
Opossum	Bonobo	162.6	690.8	94.2	47.2
Opossum	Human	162.6	696.8	96.2	47.2
Opossum	Cow	162.6	690.8	152.2	29.0
Opossum	Minke whale	162.6	686.8	108.2	40.6
Tarsier	Rat	92.3	419.0	37.0	53.3
Tarsier	Guinea Pig	92.3	393.2	53.8	33.7
Tarsier	Macaque	65.2	334.5	39.5	37.4
Tarsier	Bonobo	65.2	303.7	52.3	24.9
Tarsier	Human	65.2	311.7	52.3	25.8
Tarsier	Cow	94.2	347.3	125.7	12.1
Tarsier	Minke whale	94.2	318.5	79.5	17.3
Rat	Guinea Pig	77.9	442.5	42.5	49.9
Rat	Macaque	92.3	424.0	37.0	54.0
Rat	Bonobo	92.3	427.0	47.0	42.7
Rat	Human	92.3	424.0	51.0	39.0
Rat	Cow	94.2	424.3	115.7	17.0
Rat	Minke whale	94.2	390.5	72.5	24.5
Guinea Pig	Macaque	92.3	394.5	36.5	49.9
Guinea Pig	Bonobo	92.3	384.0	50.0	35.0
Guinea Pig	Human	92.3	382.0	51.0	34.1
Guinea Pig	Cow	94.2	415.2	111.8	17.2
Guinea Pig	Minke whale	94.2	375.0	72.0	23.5
Macaque	Bonobo	29	148.0	27.0	21.2
Macaque	Human	29	155.0	25.0	24.2
Macaque	Cow	94.2	360.8	103.2	15.5
Macaque	Minke whale	94.2	344.0	62.0	24.4
Bonobo	Human	6.3	36.0	7.0	18.7
Bonobo	Cow	94.2	346.8	112.2	13.5
Bonobo	Minke whale	94.2	312.5	65.5	20.4
Human	Cow	94.2	350.8	115.2	13.3
Human	Minke whale	94.2	318.5	66.5	20.6
Cow	Minke whale	56	222.0	100.0	8.9

Table 3.9 The nucleotide sequence comparisons between species. The number of synonymous (dS) and non-synonymous (dN) mutations observed in the sequence as a whole are shown to 1 significant figure. The ratio of the observed synonymous and non-synonymous mutations indicate selective pressures on the gene.

Overall, it is clear to see that the number of synonymous mutations in all sequence comparisons are higher than the observed non-synonymous mutations seen. Rates of synonymous mutations

reach a peak of up to 713.3, however these high values are seen with the opossum where evolutionary distance is at its greatest. With correlations previously shown from the domain divergences over time, higher rates of synonymous and non-synonymous mutations are expected to be seen with increased evolutionary distance. The highest rate of non-synonymous mutations seen is 152.2, between the opossum and the cow. The dS/dN ratio varies throughout these comparisons, and has the highest ratio of 149.9 between the mouse and rat sequences, showing that there is a high positive selection for mutation between these two sequences. This high ratio is due to the presence of a lot of synonymous mutations and very little non-synonymous mutations, which may be explained through their close evolutionary distance of 25.4 Myr.

3.15 Intron/Exon Boundary Conservation

Through investigating the synonymous and non-synonymous mutation rates found in the DNA sequences of isoforms, observing whether intron/exon boundaries were conserved throughout species would give an insight in to the evolution of sequences.

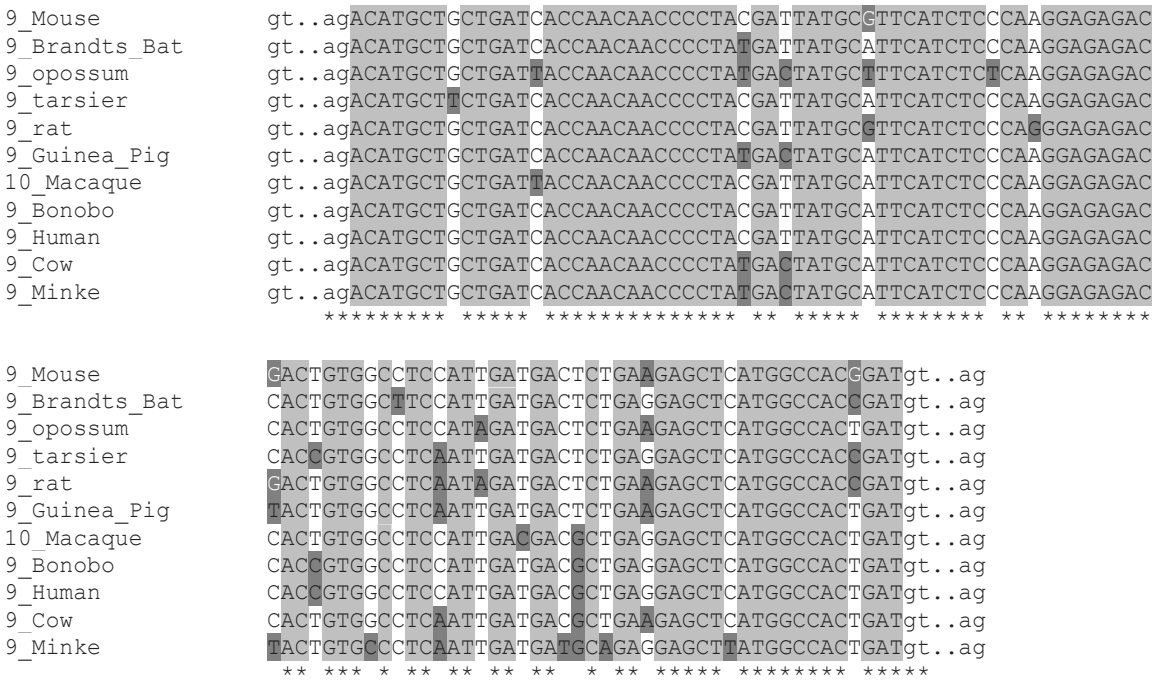


Figure 3.22 Exon 9 (10 in macaque) of the β-cardiac sequence DNA alignment showing conservation of the ag-gt boundaries Areas of conservation are highlighted grey, areas of divergence are highlighted in white with the nucleotides that vary from the consensus highlighted in dark grey. The golden hamster was not included in this sequence due to the presence of unidentified nucleotides in the sequence. Ends of exon 8 (9 in macaque) boundaries are shown and beginning of exon 9 (11 in macaque) boundaries are shown. The intron/exon boundaries are important for their involvement in splicing. Along with other important features conserved within introns (nucleating adenine and pyrimidine rich region), the GU (GT in DNA) region is recognised by the U1, U2 and U4-U6 snRNPs of the spliceosome, and is cleaved at the 5' site where it forms a lariat with the nucleating adenine. The 3' AG region is cleaved and the two resulting exons are ligated together (Lewin, B., Goldstein, E., Kilpatrick, S. 2013).

Use of the Artemis Comparison Tool allowed for analysis showing that intron/exon boundaries are conserved throughout species coding DNA. This indicates that despite the considerable synonymous and non-synonymous mutation rates observed at the DNA level, there is no tolerance for the loss of these boundaries. Figure 3.22 highlights this, with exon 9 (10 in the macaque) having both the AG and GT boundaries that are important in splicing of the introns between coding DNA. This exon also reflects the mutation rates that are observed in table 3.8, with a considerable amount of variation between species. Similar patterns were observed for all the introns in the β-cardiac sequence across the sample set.

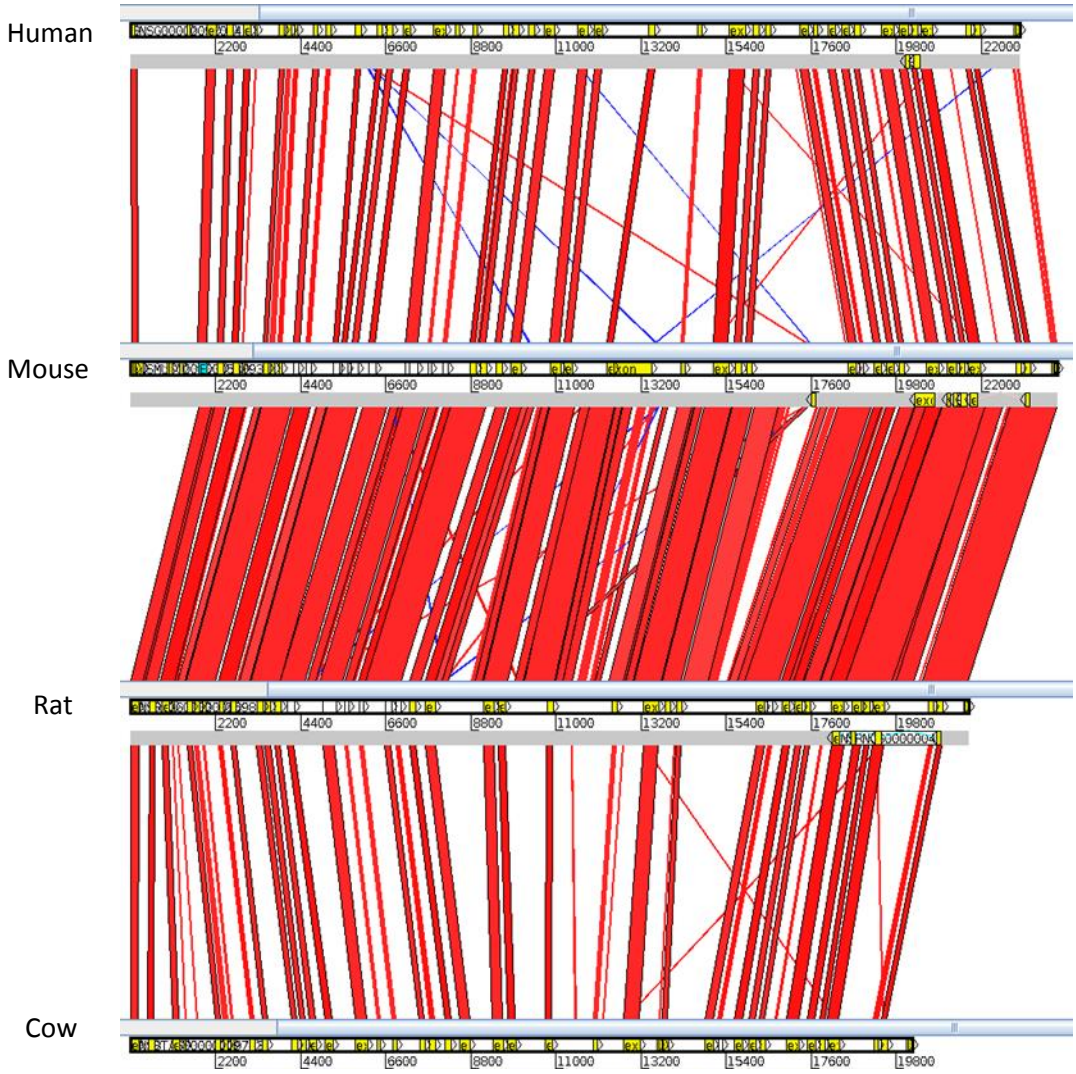


Figure 3.23 The Artemis Comparison Tool showing regions of DNA conserved between the human, mouse, rat and cow β-cardiac sequences. Regions of high similarity are indicated by red boxes, blue lines indicate areas where sequences of similarity appear inverted. Exons are indicated in yellow squares.

Figure 3.23 further highlights that there is considerable variation within the sequences but also areas of conservation between species. Using the human, mouse, rat and cow exon coding sequences, the figure shows regions of DNA that are similar between the species. The human and mouse DNA comparisons show that there are a number of sequences, often located in exons, that are conserved. This reflects that conservation in these areas are important for the function of the protein, and mutations that are seen in introns do not affect the protein. The blue lines indicate areas of similarity that are seen in the protein, however are inverted from one another. The mouse and rat comparison shows a vast amount of conservation between the two sequences, both within exons and introns. There are small regions of variance, however the high rates of conservation reflect the close evolutionary distance these two species share from one another (25.4 Myr). There is less conservation between the rat and cow than the rat and mouse, which can be expected as the evolutionary distance is greater. The conservation regions are often located within exons, reflecting that the conservation of the exon sequences in the β-cardiac isoform is important, however intergenic regions are able to undergo greater divergence without affecting the proteins expression.

4.0 Discussion

Analysis of the 130 sequences available throughout this project provided reasonable grounds to determine the evolution between myosin-II isoforms as opposed to the evolution between different myosin classes. This gave this project novelty and has allowed for the potential to progress the analyses performed further. The main aim of this investigation was to determine whether divergence rates seen in β-cardiac myosin were reflective of the changes seen in heart rate, and whether other isoforms showed similar divergences. Through analysing the divergence rates in terms of both evolutionary distance and mass, the following aims were achieved:

1. To determine the rate of divergence between myosin isoforms and whether similar rates were seen between β-cardiac myosin and other isoforms.
2. To determine where evolutionary pressures lie within the protein for each isoform and how this pressure relates to the isoforms function.
3. To identify and investigate any patterns of divergence in the proteins sequence that may be due to the increasing mass and therefore slower heart rate in mammals.
4. To determine whether the rates of DNA evolution seen in the β-cardiac sequence give an insight in to the proteins evolution through synonymous and non-synonymous mutations.

4.1 Rates of Divergence between Isoforms

The comparisons of the isoforms lead to a number of conclusions. Of the three main isoforms compared, β-cardiac, non-muscle A and embryonic, the percentage identity matrices (figures 3.1 and 3.4) were indicative that patterns of divergence observed in β-cardiac were not seen in the other 9 isoforms. Through dividing these sequences in to their motor and tail domains, we were able to more clearly see where any relationships lied. Comparing the β-cardiac motor domain to the non-muscle A and embryonic motor domain sequences (figure 3.1), it was clear that there was a greater divergence in β-cardiac sequence that wasn't observed in other isoforms through the presence of sequence identities in the 94-98% range. As the β-cardiac function is different from both non-muscle A and embryonic myosin, it is indicative that this isoform has had a greater pressure to diverge in order to compensate for different needs of the organisms over time. The

high conservation of the sequences motor domain in non-muscle A shows that the isoform has maintained a steady state in the samples over time, suggesting that there is little tolerance for changes in the catalytic activity of the protein. The embryonic myosin shows total conservation throughout all of the species listed in terms of their evolutionary distance from humans apart from the Opossum, which shows a slight increase in its variance compared to other species. This may be due to the differing physiology of this mammal, or the large evolutionary distance between other mammals. Being a marsupial, it gives birth to a neonate that then feeds on a teat in a pouch of the mother as opposed to undergoing full foetal development, like the rest of the mammals included. As embryonic myosin is expressed during foetal development, this small increase in divergence may be reflected by this differing reproductive cycle.

The Brandt's Bat also raised the importance of taking in to consideration the physiology of each mammal when considering the isoforms sequence identity. As the Brandt's bat was the smallest mammal in the set, it was originally the reference sequence to compare all other mammals to. However, the sequence proved to be more divergent from other mammals at a rate that was not indicative of mass, causing it to become an outlier where other sequences would more closely match in similarity. This led to an investigation as to question whether the Brandt's bat used myosin-II isoforms in the way that other mammals did. Literature revealed that the bat already tested the longevity versus size relationship, and is the longest lived mammal for its size. This provided reasonable grounds to not use it as the reference sequence but to use the mouse as it is more widely studied.

In terms of evolutionary distance, the β-cardiac motor domain appears to diverge at a faster rate over time than both non-muscle A and embryonic myosin. When taking in to consideration standard protein divergence from the highly conserved Histone H2 and Haemoglobin β, Histone H2 diverges at a rate around 0.25 point mutations per 100 residues per 100 million years (Mut/100Myr) and Haemoglobin β diverges at a rate around 30 Mut/100Myr. Taking these proteins divergence rates as guidelines for whether proteins are more or less divergent, the rates of myosin-II divergence fall within this range. With the slope of the β-cardiac motor domain being -0.049 with

a Pearson's correlation coefficient (R) of 0.958, this strong slope fits a linear relationship very well. Non-muscle A and embryonic myosin's slopes of -0.010 and -0.016 with R values 0.931 and 0.928 respectively, show β-cardiac motor domain diverges at a faster rate. Where there appears to be correlations in both non-muscle A and embryonic myosin motor domains with divergence over time, it is at a rate slower than β-cardiac. Compared to all the other myosin motor domains, β-cardiac has the strongest negative slope in terms of evolutionary distance (table 3.3).

The different sample sizes make it difficult to draw conclusions from, which was why β-cardiac, non-muscle A and embryonic myosin were primarily used as comparators and controls. However, for all of the isoforms, evolutionary distance explained the relationship between divergence quite well. Where there is a large scattering around a certain evolutionary distance (e.g. figure 3.2 β-cardiac 94 Myr), this indicates that there may be other factors involved. β and α –cardiac, skeletal 2D/X and 2A motor domains all show a large scattering around the 94 Myr point when considering evolutionary distance, suggesting the relationship is not clearly explained by evolutionary divergence alone.

When considering mass as a parameter for protein evolution, the β-cardiacs motor domain maintains a strong correlation (R = -0.945) as seen in the evolutionary distance plots (R = 0.958), suggesting that the protein does diverge over time as expected, but also appears to diverge when average adult body mass increases in the mammal. Any trends from all other isoforms observed in the evolutionary distance plots are lost when considering mass.

Overall, all isoforms show that the motor and tail domains do show sequence divergence is correlated with evolutionary divergence, through the presence of strong correlations (table 3.3). It is the rates of divergence in the domains that differ, with some diverging at a faster rate (β-cardiac, slope -0.049) and some at a slower rate (Non-muscle A, slope -0.01). What can be noted is that it is the β-cardiac motor domain that shows the fastest rate of divergence out of all the isoforms when considering evolutionary divergence.

The mass relationships seen for all isoforms are much weaker in both motor and tail domains. However, the β-cardiac motor domain is the only sequence that shows a strong correlation in this analysis, reflecting the variations seen in the proteins structure as mass increases.

Confidence in these values can be taken from the larger analysis considering 23 mammals. Small sample sizes proved difficult to draw conclusions from, as it may have not reflected relationships shown in the taxonomy as a whole. Figure 3.6 highlighted that this relationship with mass and divergence is only applicable to mammals, due to other taxonomic groups having different physiologies. Figure 3.17 shows that the original species reflect the relationship well when combined with other sequences that were available. This increased sample size generated a strong negative slope in the motor domain (-0.809) with a strong correlation coefficient of -0.808, reflecting the strong relationships seen in the motor domain of the isoform in the smaller sample set. The tail domain goes on to show no relationship with divergence when considering mass also. Similar patterns of divergence can be seen in larger sample sizes for both non-muscle A and embryonic myosin also, as shown in figures 3.18 and 3.19. Non-muscle A shows a greater rate of conservation in its motor domain than its tail domain where there are greater rates of divergence, possibly reflecting its function and adaptation. Embryonic myosin also shows similar trends (3.19) as seen in figure 3.1, where both domains show similar conservation rates.

4.2 Variations in the Proteins Structure

The rates of variation observed indicated that there are relationships seen when considering evolutionary distance and mass of species. This led to investigating what these variations were, and where they lied in the sequences of these isoforms. Whole protein sequences were analysed and the frequency of each amino acid was counted alongside the number of different amino acids at each position. Key structures of the motor domain were highlighted to detail if these points of variation occurred in areas they would have been predicted to be seen. Figure 3.20 reflects patterns seen in the matrices from figure 3.1, where there appears to be more variations in the motor domain of the β-cardiac as opposed to non-muscle A and embryonic isoforms.

Structures of the domains were marked on, and showed that areas of more divergence in the proteins are not actually found within the structures themselves. This observation may indicate that the residues involved in these structures are functionally important and there is little tolerance for variations at these points. In the β-cardiac sequence, high numbers of variations outside of these structures may show that it is surrounding residues of these structures that play an important role in the variance of the proteins activity as body mass increases and heart rate decreases.

For non-muscle A, the high variation observed in the tail domain may reflect its role *in vivo* where it acts at the cellular level. The catalytic domain may not have a significant need for difference in its kinetics, whereas the tail domain may have different pressures to bind different structures in the cell, causing a pressure and increasing the variation.

The SH3 β-barrell shows consistent variation between all three isoforms, highlighted in figure 3.21. The β-barrells function remains relatively unknown, however evidence suggesting that the structure is involved in the binding of the extension of the essential light chain may reflect this variation (Lowey, S., Saraswat, L., Hanein, D., *et al*. 2007). Whether different affinities for actin of isoforms in different species alters remains unknown, however variation in this β-barrell may indicate slight variations that are significant in determining the kinetics of the interactions.

4.3 Residues of Interest

By identifying locations of residues that contain variations, further detail in to these locations in the β-cardiac was done to identify residues that may impact the structure of the protein when considering mass (table 3.8). A list of 35 residues were identified from a multiple sequence alignment that reflected a trend with mass, with either presence of residues in small mammals that varied in larger animals and vice versa. Some residues reflected relationships more clearly, such as residue 77 where there is a conservation of methionine in species up to the dolphin, cow and minke whale where the variation of a leucine is introduced. The opposite pattern can be seen in residue 580, where valine is present in the mouse, brandt's bat, little brown bat and golden hamster, however isoleucine is present in all other species at this point. These varying residues may play a

role in affecting the contraction rates of the protein to compensate for the varying heart rates observed as mass changes.

The physiology of the bats and their use of the β-cardiac isoform is also reflected in these analyses. An example is at residue 591 where the presence of leucine and isoleucine in the majority of mammals is present, however the bats both express glutamine, suggesting that they have permitted a variation that other species have not, which may be due to their use of the isoform as opposed to a mass dependence.

These analyses would have been better supported through their mapping on to the structure of the myosin protein, allowing for mutations to be visualised within the structure to see how they interact with surrounding residues.

4.4 DNA Analysis

High rates of synonymous mutations in the sequence indicate that silent mutations are constantly playing a part in these isoform sequences over their divergence that are not reflected at the protein level. These evolutionary pressures identified may indicate that the proteins are under evolutionary pressure, however little tolerance is accepted for missense mutations, as shown in lower rates of non-synonymous mutations. The ratios indicated in the sequence suggest that the sequence comparisons are under positive selection.

Although there are high synonymous mutation rates in the coding sequence of the proteins, their intron/exon boundaries are highly conserved throughout the evolution of the β-cardiac sequence. This shows that the mutations observed in the coding sequences do not affect the splicing activity of introns and exons throughout the species comparisons and any variations seen are not due to the introduction of an intron.

4.5 Future Work

Many comparisons were done in this project that highlighted the variations seen in the β-cardiac sequence. Although the sample sizes included proved to be representative, there also appeared to be exceptions to the rule through the variations observed in *Myotis brandtii*. The lack of sequences available for all isoforms also limited conclusions that were able to be drawn. Future work would involve investigating more mammalian sequences for the isoforms to better determine any evolutionary rates of proteins, as strong correlations were seen in all isoforms, justifying further investigations. For the β-cardiac sequence, evolutionary distance explained the divergence of the motor domain up until 94 Myr, where the opossum did not follow the same divergence. Whether this was down to the physiology of the organism or evolutionary distance is not clear, so further investigations using mammals that diverged further than 94.2 Myr from humans could be performed to determine the lack of relationship seen with the opossum sequence. The 12 species sample size contain species from the Boroeutheria classification, which may not reflect the relationship seen throughout mammals as a whole. Further investigations using sequences from mammals that lie outside of this classification, such as Eutheria, Metatheria and Prototheria may indicate more accurate relationships of isoform divergences across species. However, it should be taken in to consideration that any relationships seen may reflect what the isoform is used for in the organisms due to differing physiology.

Problems of finding completed sequences that were of good quality were also regularly encountered throughout the project. Although these sequences may have contained correct exon sequences, misalignment of the sequences justified removal for this project. Future work could be performed on these exon sequences that contain key structures to determine any variation within them, which would expand the number of sequences included for comparison drastically.

Biochemical work could be carried out on the myosin proteins in order to better determine the kinetics between the isoforms from different species. Bioinformatic analysis revealed variations that do occur as mass increases in the β-cardiac isoform, justifying the investigation of how variations affect the kinetics of the protein as heart rate decreases. *In vitro* motility assays of the

whole protein could be performed to visualise the movement of fluorescently labelled actin across a cover slip. This would give an insight in to the variation of kinetics the proteins have.

Following on from biochemical analysis, further bioinformatics work could be done to determine how mutations outside of key structures are predicted to affect the proteins activity, as found through the variation analysis. Mapping the variations on to the proteins structure through the use of PyMol (The PyMOL Molecular Graphics System, Schrödinger) and other programmes would provide insights in to any interactions wild type and mutated residues would have with surrounding structures. However, limited crystal structures of myosin-II throughout all stages of the cross-bridge cycle may prove this difficult to predict, due to the conformational changes that are seen in the protein. Molecular dynamic studies could be done to better determine how the variations observed in the β-cardiac sequence from small to large mammals affect its activity. Molecular dynamics simulates the movements of atoms within a given system, which would indicate at the atomic level how variations affect the protein.

5.0 Conclusions

Overall, it has been shown that the β-cardiac sequence has a correlation between its divergence and mass that is not seen in other isoforms. This project highlighted that all other isoforms are shown to have a correlation with divergence over evolutionary distance, but this relationship is lost when considering mass as a parameter. By further characterising the mass dependence on divergence of the β-cardiac isoform, future investigations can be carried out on the variations observed in the sequence. Where this relationship has been shown in the β-cardiac sequence, residues that may play a role in affecting the activity of the protein have been identified through analysing where in the sequence the variations are. Characteristics for other isoforms have also been defined, for instance in non-muscle A, where there is pressure on the tail domain of the protein that may be due to its function. While there still maintains to be a wide scope of further work to be completed on each of the aims discussed for this project, conclusions from the analyses performed have justified further work.

References

Apweiler, R., Bairoch, A., Wu, C. (2004) "Protein Sequence Databases". *Current Opinion in Chemical Biology* 8(1): 76-80.

Alberts, B., Johnson, A., Roberts, K. (2002) *Molecular Biology of the Cell.* 4th Ed. New York: Garland Science.

Bobkov, A., Bobkova, E., Reisler, E. *et al* (1996) "The role of surface loops (residues 204-216 and 627-646) in the motor function of the myosin head" *Proc Natl Acad Sci* 93(6): 2285-2289.

Borejdo, J., Ushakov, D., Akopova, I. (2002) "Regulatory and Essential Light Chains of Myosin Rotate Equally during Contraction of Skeletal Muscle" *Biophysical Journal* 82: 3150-3159.

Brett, D., Pospisil, H., Bork, P. *et al* (2002) "Alternative splicing and genome complexity". *Nat Genet.* 30(1): 29-30.

Capitanio, M., Canepari, M., Cacciafesta, P. *et al* (2005) "Two independent mechanical events in the interaction cycle of skeletal muscle myosin with actin" *Proc Natl Acad Sci* 103(1): 87-92.

Capra JA and Singh M. (2007) "Predicting functionally important residues from sequence conservation". *Bioinformatics* 23(15):1875-82.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG and Parkhill J. *ACT: the Artemis Comparison Tool*. Bioinformatics (Oxford, England) 2005;21;16;3422-3

Chew, M., Squire, J. (1995) "Packing of α-Helical Coiled-Coil Myosin Rods in Vertebrate Muscle Thick Filaments". *Journal of Structural Biology* 115: 233-249.

Coluccio, L. Holmes, K. (2007). Myosins: A Superfamily of Molecular Motors. 7th ed. Netherlands: Springer. 35-52.

De La Cruz, E., Ostap, E. (2004) "Relating biochemistry and function in the myosin superfamily" *Current Opinion in Cell Biology* 16: 61-67.

Deacon, J., Bloemink, M., Leinwand, L. *et al* (2012) "Identification of functional differences between recombinant human α and β cardiac myosin motors". *Cell Mol Life Sci* 69(13): 2261-2277.

Devlin, T. (2011) *Textbook of Biochemistry: With Clinical Correlations*. Hoboken, NJ: John Wiley & Sons.

Geeves, M., Holmes, K. (2005) "The Molecular Mechanism of Muscle Contraction". *Advances in Protein Chemistry* 71: 161-193.

Ground Rules for Gene Expression (2014) Available from: http://biomoocnews.blogspot.co.uk/ [12/06/2015]

Haase, H., Morano, I. (1996) "Alternative splicing of smooth muscle myosin heavy chains and its functional consequences". *J Cell Biochem* 60(4):521-8.

Hedges SB, Dudley J, & Kumar S (2006). "TimeTree: A public knowledge-base of divergence times among organisms". *Bioinformatics* 22: 2971-2972

Hedges, S., Marin, J., Kumar, S., *et al*. (2015) "Tree of Life Reveals Clock-Like Speciation and Diversification". *Mol. Biol. Evol.* 32(4): 835-845.

Hicks, M., Holberton, D., Kowalczyk, C., Woolfson, D. (1997) "Coiled-coil assembly by peptides with non-heptad sequence motifs" *Folding and Design* 2(3): 149-158.

Hindman, B., Goeckeler, Z., Wysolmerski, R. *et al* (2015) "Non-Muscle Myosin II Isoforms Have Different Functions in Matrix Rearrangement by MDA-MB-231 Cells". *PLoS One* 10(7): e0131920.

Hubbard, T., *et al*. (2002) "The Ensembl genome database project". *Nucleic Acids Res* 30(1): 38-41.

Human Physiology Slide Player (2009) Available at: http://slideplayer.com/slide/4452656/

Huxley, A., Niedergerke, R. (1954) "Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres". *Nature* 173: 971-973.

Iorga, B., Adamek, N., Geeves, M. (2007) "The slow skeletal muscle isoform of myosin shows kinetic features common to smooth and non-muscle myosins". *J Biol Chem* 282(6): 3559-70.

Jeanfavre, S. (2014) *Correlation of Organism Size, Myosin Structure, and ADP-Release in Beta-Cardiac Myosin Isoform.* Unpublished Masters Thesis. University of Kent.

Korn, E., 2000. "Coevolution of head, neck, and tail domains of myosin heavy chains". *Proc Natl Acad Sci*, 97(23): 12559–64

Krebs, J., Lewin, B., Goldstein, E. *et al* (2013) *Lewin's essential genes*. 3rd Ed. Burlington, MA: Jones and Bartlett Publishers.

Krenz, M., Sanbe, A., Robbins, J. *et al* (2003) "Analysis of Myosin Heavy Chain Functionality in the Heart". *JBC* 278: 17466-17474.

Kurapati, R., McKenna, C., Blanco, G. *et al* (2011) "Myofibrillar myopathy caused by a mutation in the motor domain of mouse MyHC IIb" *Hum. Mol. Genet.* 1: 1-19.

Lowey, S., Bretton, V., Trybus, K. *et al* (2013) "Transgenic Mouse α- and β-Cardiac Myosins Containing the R403Q Mutation Show Isoform-dependent Transient Kinetic Differences". *J Biol Chem* 288(21): 14780-14787.

Lowey, S., Saraswat, L., Liu, H. *et al* (2007) "Evidence for an interaction between the SH3 domain and the N-terminal extension of the essential light chain in class-II myosins". *J Mol Biol* 371(4): 902-13.

Malmqvist, U., Aronshtam, A., Lowey, S. (2004) "Cardiac Myosin Isoforms from Different Species Have Unique Enzymatic and Mechanical Properties" *Biochemistry* 43: 15058-15065.

Moore, J., Leinwand, L., Warshaw, D. (2012) "Invited Review: The Myosins: Exploration of the Development of Our Current Understanding of These Mutations in the Motor". *Circ Res* 111(3): 375-385.

Muscle: Actin and Myosin (2011) Available at: http://www.britannica.com/science/actin-filament/images-videos/The-structure-of-actin-and-myosin-filaments/46641

Myosin family analysis (2012) Available at: http://www.motorprotein.de/myosin.html

NCBI and NIH (2013) *The New BLAST Results Page, enhanced graphical presentation and added functionality* [PDF] [accessed 8/06/2015]. Available at: ftp://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_NewBLAST.pdf

Newsholme, E., Leech, T. (2010) *Functional Biochemistry in Health and Disease.* Chichester, UK: Wiley-Blackwell.

Ozkan, E., Aceti, M., Rumbaugh, G., *et al* (2015) "Input-specific regulation of hippocampal circuit matruation by non-muscle myosin IIB" *Journal of Neurochemistry* 134(3): 429-444.

Pellegrino, M., Canepari, M., Bottinelli, R. *et al* (2003) “Orthologous myosin isoforms and scaling of shortening velocity with body size in mouse, rat, rabbit and human muscles” *J Physiol.* 435(3): 677-689.

Ramakrishnan, C., Dani, V., Ramasarma, T. (2002) “A conformational analysis of Walker motif A [GXXXXGKT(S)] in nucleotide-binding and other proteins”. *Protein Eng.* 15(10): 783-789.

Robert, X. and Gouet, P. (2014) "Deciphering key features in protein structures with the new ENDscript server". *Nucl. Acids Res*. 42(1): 320-324.

Rossi, A., Mammucari, C., Schiaffino, S. *et al* (2010) Two novel/ancient myosins in mammalian skeletal muscles: MYH14/7b and MYH15 are expressed in extraocular muscles and muscle spindles”. *J Physiol.* 588(2): 353-64.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA and Barrell B *Artemis: sequence visualization and annotation*. Bioinformatics (Oxford, England) 2000;16;10;944-5

Saez, L., Gianola, K., Leinwand, L. *et al* (1987) “Human cardiac myosin heavy chain genes and their linkage in the genome” *Nucleic Acids Research* 15(13): 5443 – 5459.

Savage, V., Allen, A., West, G. *et al* (2007) “Scaling of number, size, and metabolic rate of cells with body size in mammals” *Proc Natl Acad Sci* 104(11)L 4718-4723.

Schiaffino, S., Reggiani, C. (2011) “Fiber types in mammalian skeletal muscles” *Physiol Rev* 91: 1447-1531.

Seim, I., Fang, X., Lobanov, A. *et al* (2012) “Genome analysis reveals insights into physiology and longevity of the Brandt’s bat *Myotis brandtii*” *Nature Communications* 4: 2212.

Sievers, F., Wilm, A., Dineen, D., *et al* (2011) “Fast, scalable generation of high-quality protein multiple sequence alignment using Clustal Omega” *Molecular Systems Biology* 7(1): 539.

Silverthorn, D. (2013) *Human Physiology: An Integrated Approach*. 6th Ed. IL: Pearson Education.

The Wellington Hospital: The Heart (2007) Available from: http://www.thewellingtoncardiacservices.com/the-heart.asp

University of Wisconsin – Madison – Physiology. (2012) Available from: https://www.studyblue.com/notes/note/n/exam-2/deck/5980386 [14/06/2015].

Weiss, A., Schiaffino, S., Leinwand, L. (1999) “Comparative sequence analysis of the complete human sarcomeric myosin heavy chain family: implications for functional diversity”. *Journal of Molecular Biology* 290 (1): 61-75.

What is Myosin? (2009) Available from: http://oregonstate.edu/instruction/bi314/summer09/cytoskel.html

Wheeler, D., *et al* (2000) “Database resources of the National Center for Biotechnology Information”. *Nucleic Acids Res* 28(1): 4-10.

Yang, Z., Nielsen, R. (2000) “Estimating Synonymous and Non-synonymous Substitution Rates Under Realistic Evolutionary Models”. *Mol Biol Evol* 17(1): 32-43.

Appendix A)

Species	Isoform												Avg Mass (kg)
	Skeletal 2d/x	Skeletal 2a	Embryonic	Skeletal 2b	α-cardiac	β-cardiac	Perinatal	Non-muscle A	Non-muscle B	Smooth Muscle	Extraocular	Slow Tonic	
Human	P12882	Q9UKX2	251757455	Q9Y623	P13533	P12883	P13535	P35579	219841954	13432177	110624781	599045671	68[a]
Bonobo	675746236	397494570	675746242	675746226	397473260	397473262	675746209	675764569	675746138		675746206	675798456	45.5[a]
Macaque	544497116	544497114	544497126	109113269	544446347	544446351	544497122	383408157	384940798	387541766	544497107	544465262	6.55[b]
Tarsier	640786419		640786435	640786417	640818214	640818212	640786413	640796733	640805785		640786411	640822915	0.1315[c]
Rat	205830438	G3V6E1	6981234	Q29RW1	P02563	P02564	F1M8F6	Q62812	Q9JLT0	282158051	F1M789	211938461	0.325[b]
Mouse	Q5SX40	G3UW82	153792649	Q5SX39	Q02566	Q91Z83	P13542	Q8VDD5	Q61879	20070691	124486959	145864471	0.017[b]
Guinea pig	348560933	D6N3E0	524926861		348577524	348577522	348560930	348569442	514454472	H0VCX1	348560928	514456949	0.9[a]
Hamster			348560937	514454246	P13539	P13540	524926859	524925525	524926815		524926855		0.1125[a]
Cow	Q9BE40	Q9BE41	156120319	E1BP87	O97496	Q9BE39	F1N775	F1MQ37	554569298	151554905	741963173	741948777	755[a]
Brandt's Bat	554569234		554569240	554569232	554581513	554581511	554569226	G1PWB8	Q27991		554569314	554572455	0.0069[b]
Minke Whale		594653830	511880634			594665031	594653837	594622148	594623548	594681427		594692255	9200[d]
Opossum	126308741	126308745	126308751	126308737	611969297	126277437	126308731	126339824	611994225	126334598	611994046	F7GF97	0.048[b]
	1	2	3	4	6	7	8	9	10	11	13	7b	
Myosin Heavy Chain - MyHC													

Appendix table 1 showing the Uniprot and NCBI GI accession numbers of the isoform sequences used. Identifiers starting with upper case letters are from Uniprot, identifiers consisting only of numbers are secondary access GI codes for Refseq Protein database. Databases used to collect average adult body mass have been included.

[a] = Animal Diversity Web www.animaldiversity.org

[b] = Arkive.org www.arkive.org

[c] = Primatology.net www.primatology.net

[d]= National Oceanic and Atmospheric Administration Fisheries www.nmfs.noaa.gov

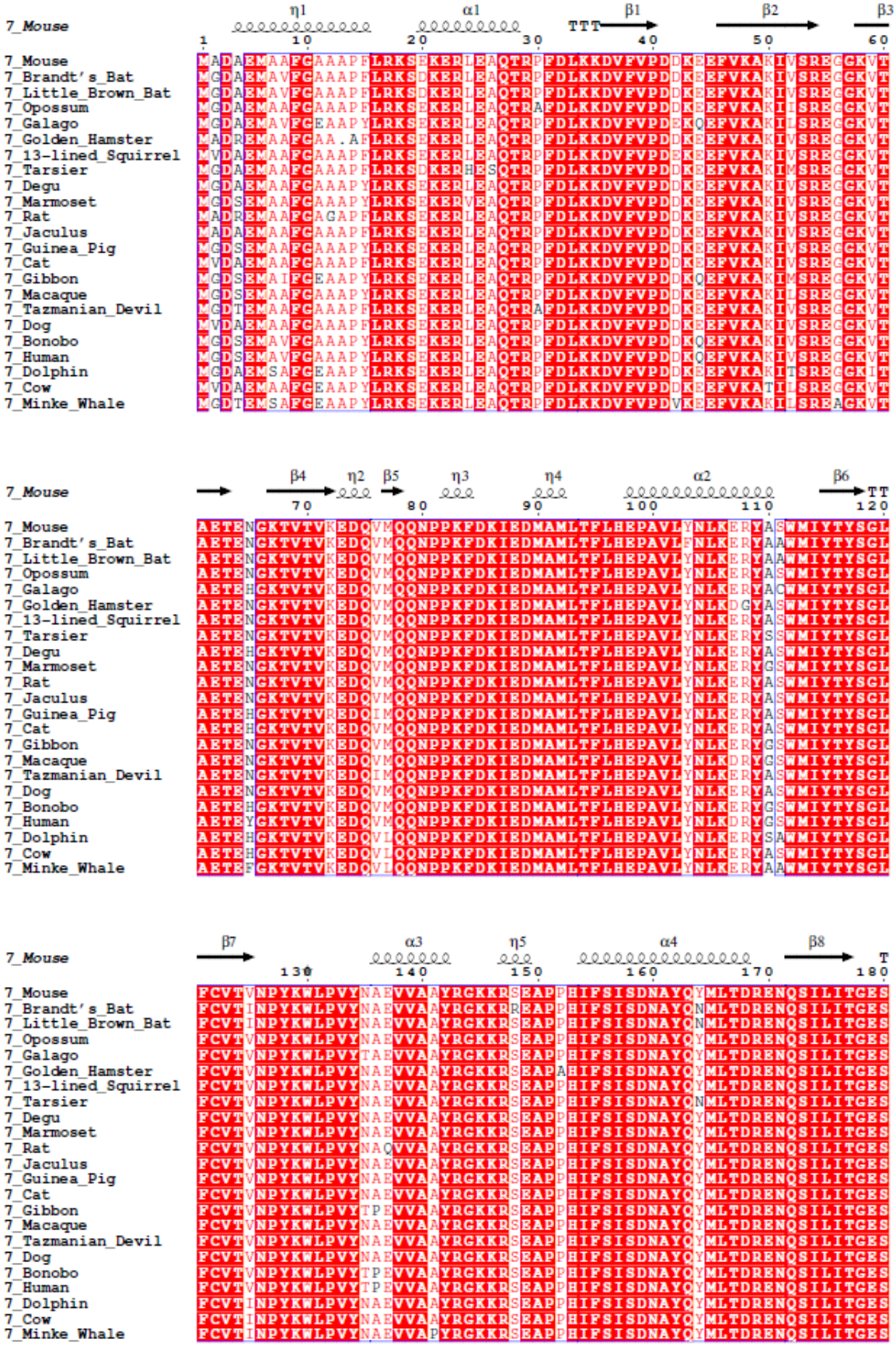
Appendix B)

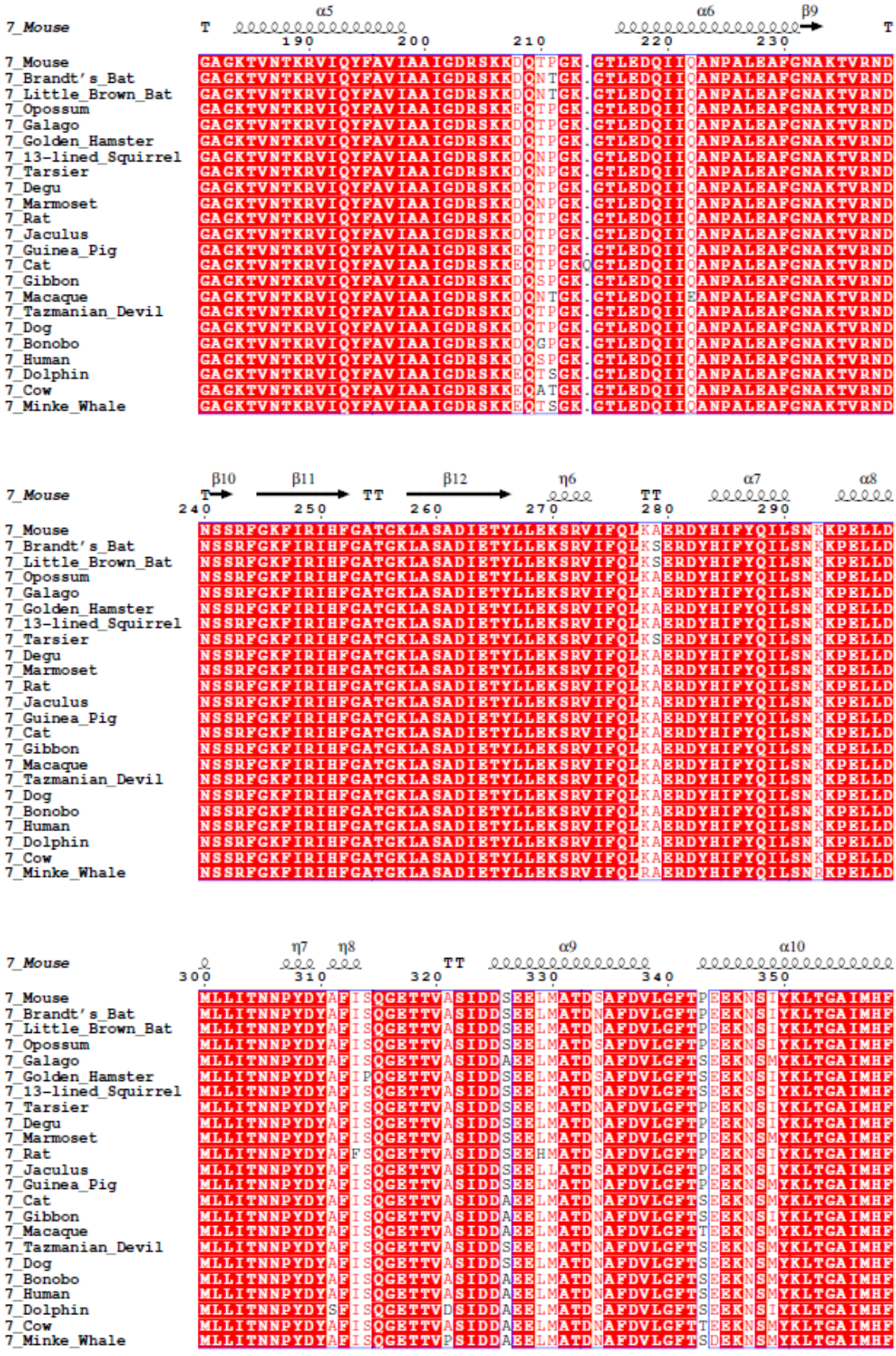
Additional Mammalian Species Gene Identifiers for β-cardiac (7) Sequence			
Species	Identifier (Uniprot or gi)	Average Adult Mass (kg)	Mass Source Website Name
Little Brown Bat	G1PWD9	0.008	Animal Diversity Web[a]
Galago	H0WZ33	0.1025	Primate Info Net[b]
13-Lined Ground Squirrel	I3M0R9	0.125	Smithsonian National Museum of Natural History[c]
Degu	507709647	0.235	Animal Diversity Web[a]
Marmoset	F7I4V9	0.246	Primate Info Net[b]
Jaculus	507556199	0.6	Arkive[d]
Cat	M3WAS7	4.75	Animal Diversity Web[a]
Gibbon	441667084	5.7	Animal Diversity Web[a]
Tasmanian Devil	395503058	7.259	Parks and Wildlife Service-Tasmania[e]
Dog	P49824	45	Animal Diversity Web[a]
Dolphin	602728381	380	Animal Diversity Web[a]

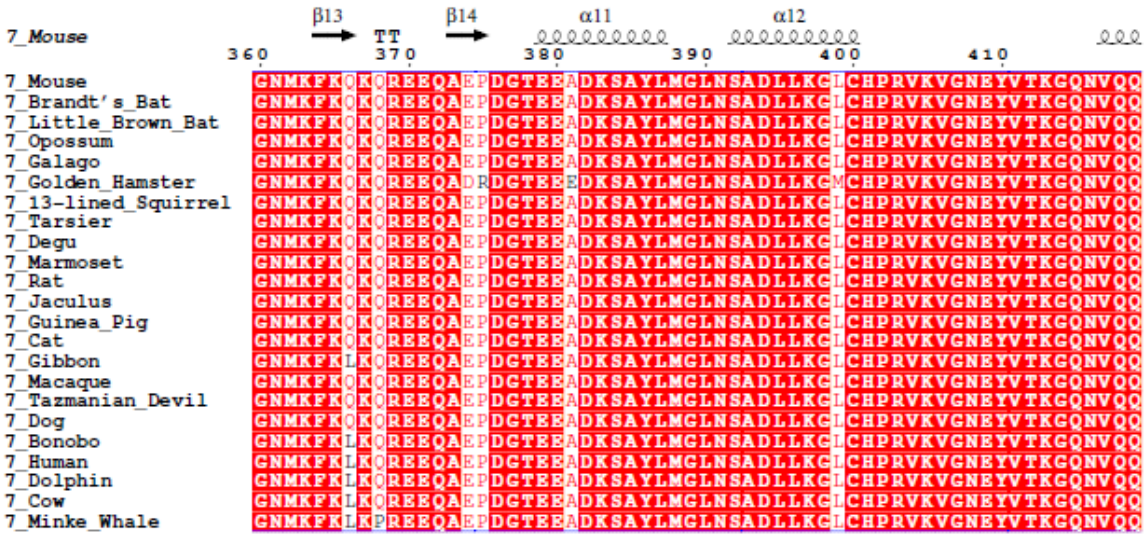
Appendix table 2 showing the additional mammalian species used for the β-cardiac larger data set in figure 3.17. The sequence identifiers are included where Uniprot accession numbers begin with an upper case letter and GI accession codes consist solely of numbers. The average adult masses for each species is included along with the database the data was extracted from.

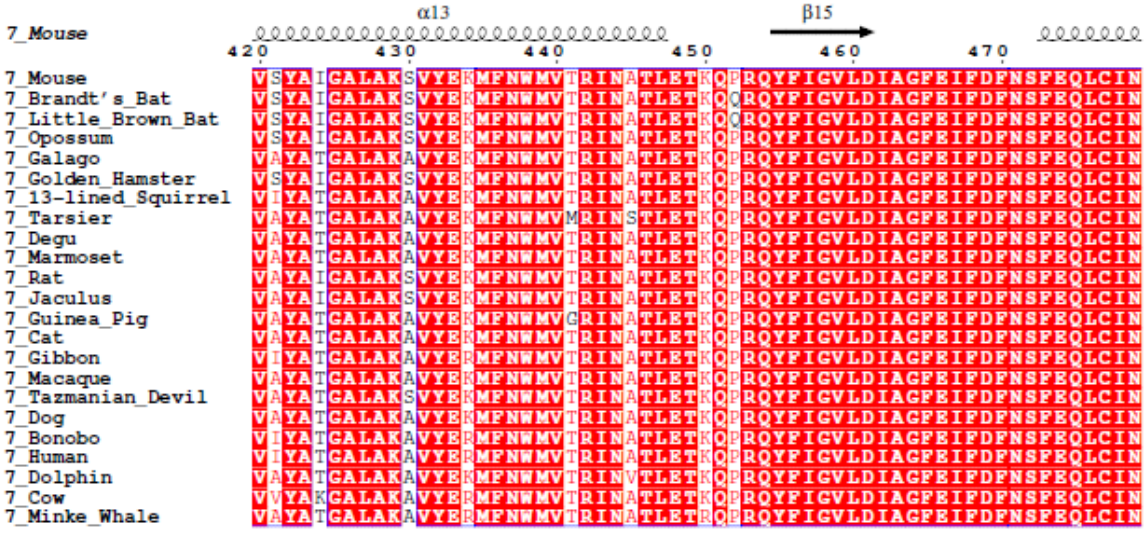
a= www.animaldiversity.org b = pin.primate.wisc.edu c = www.mnh.si.edu d = www.arkive.org e = www.parks.tas.gov.au

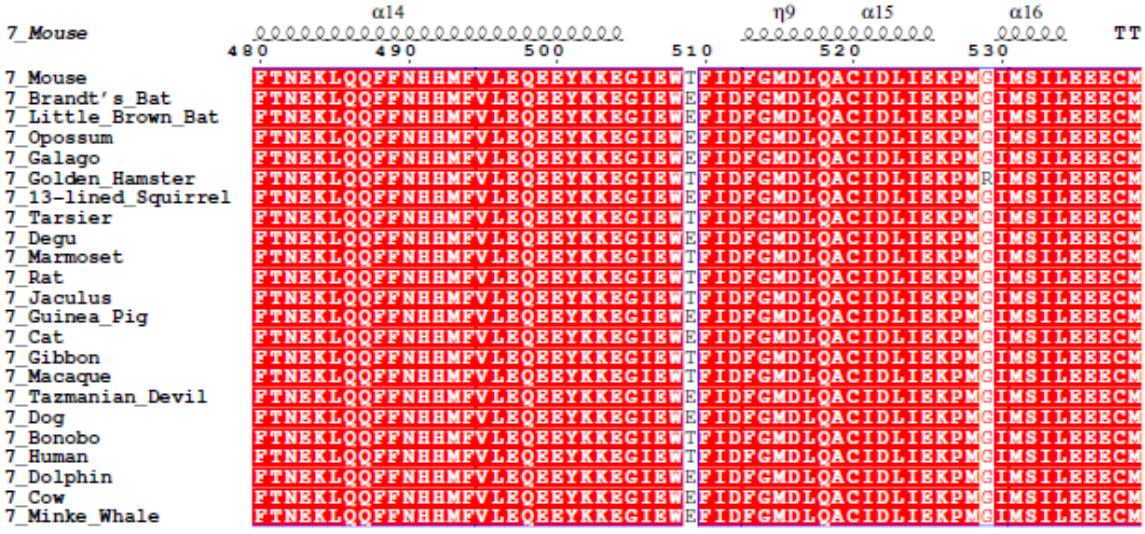
Appendix C) Multiple Sequence Alignment for β-cardiac myosin with annotated structures.

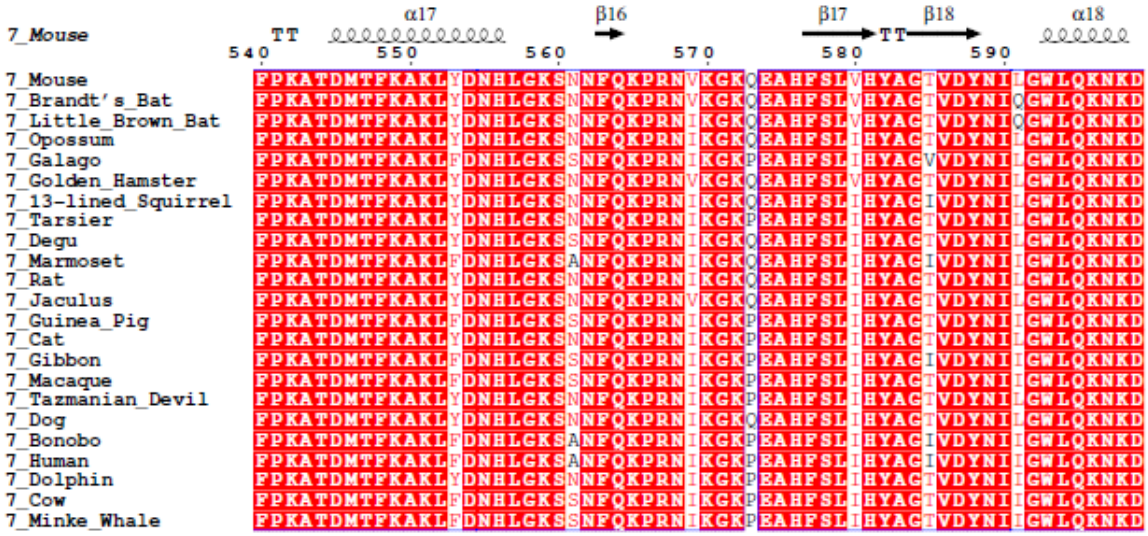


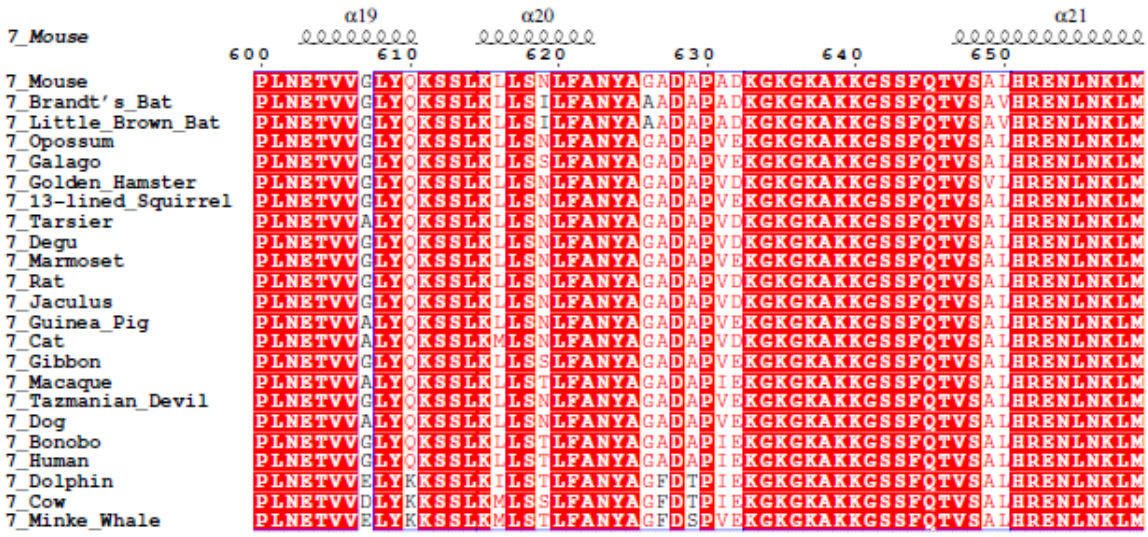

7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

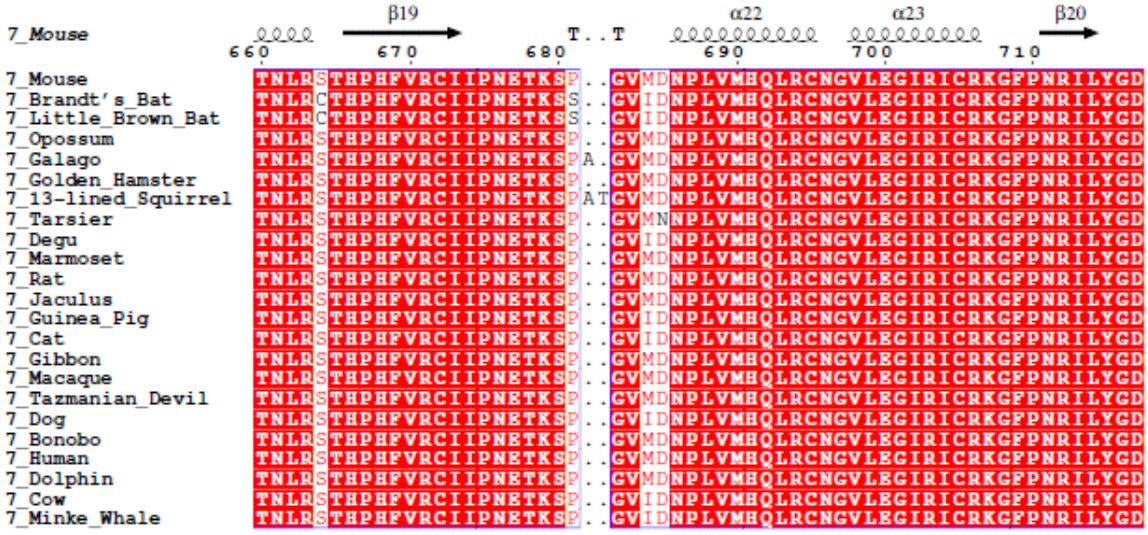

β13
β14
α11
α12
7_Mouse
TT
360
370
380
390
400
410
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

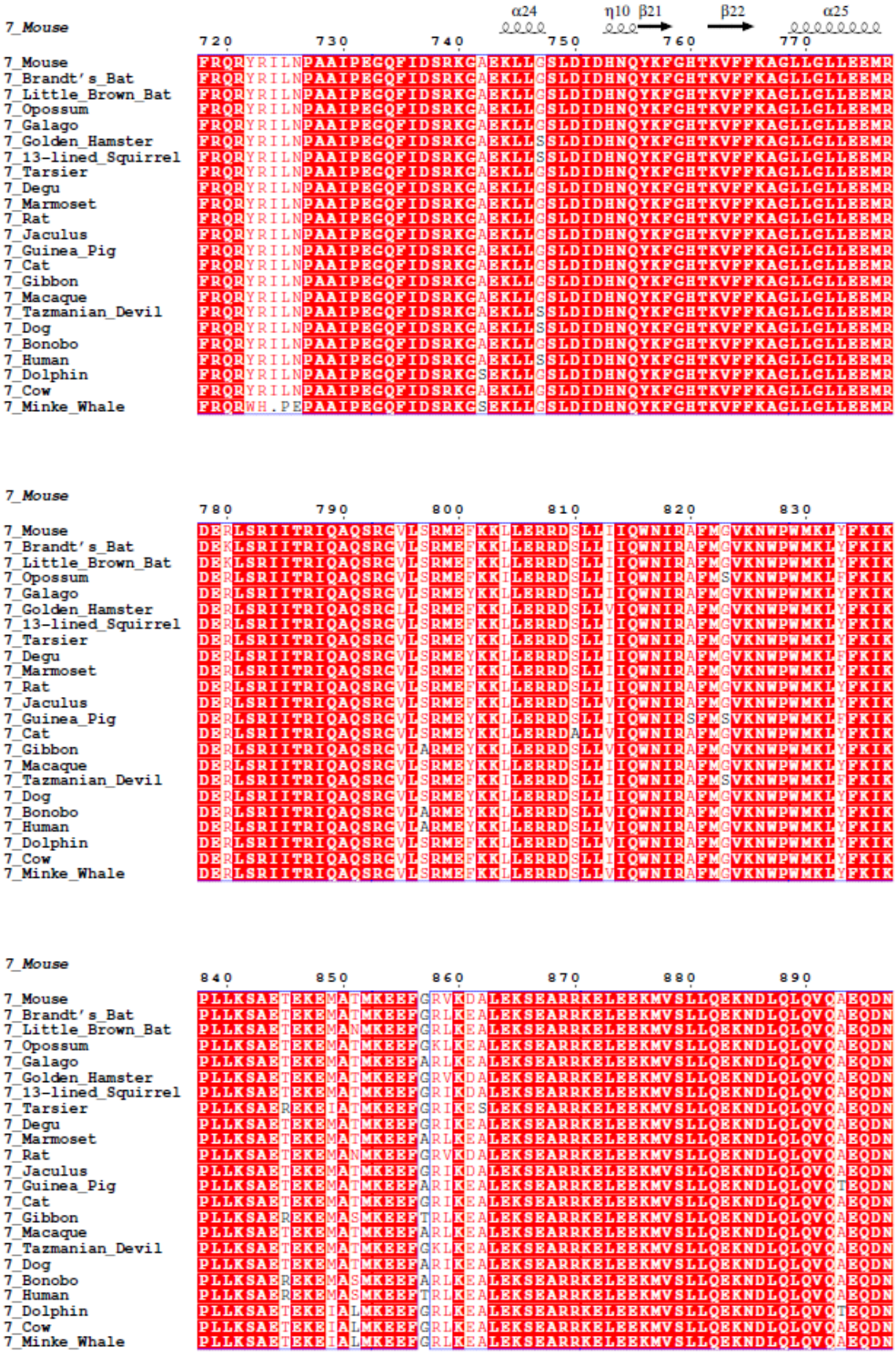

α13
β15
7_Mouse
420
430
440
450
460
470
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

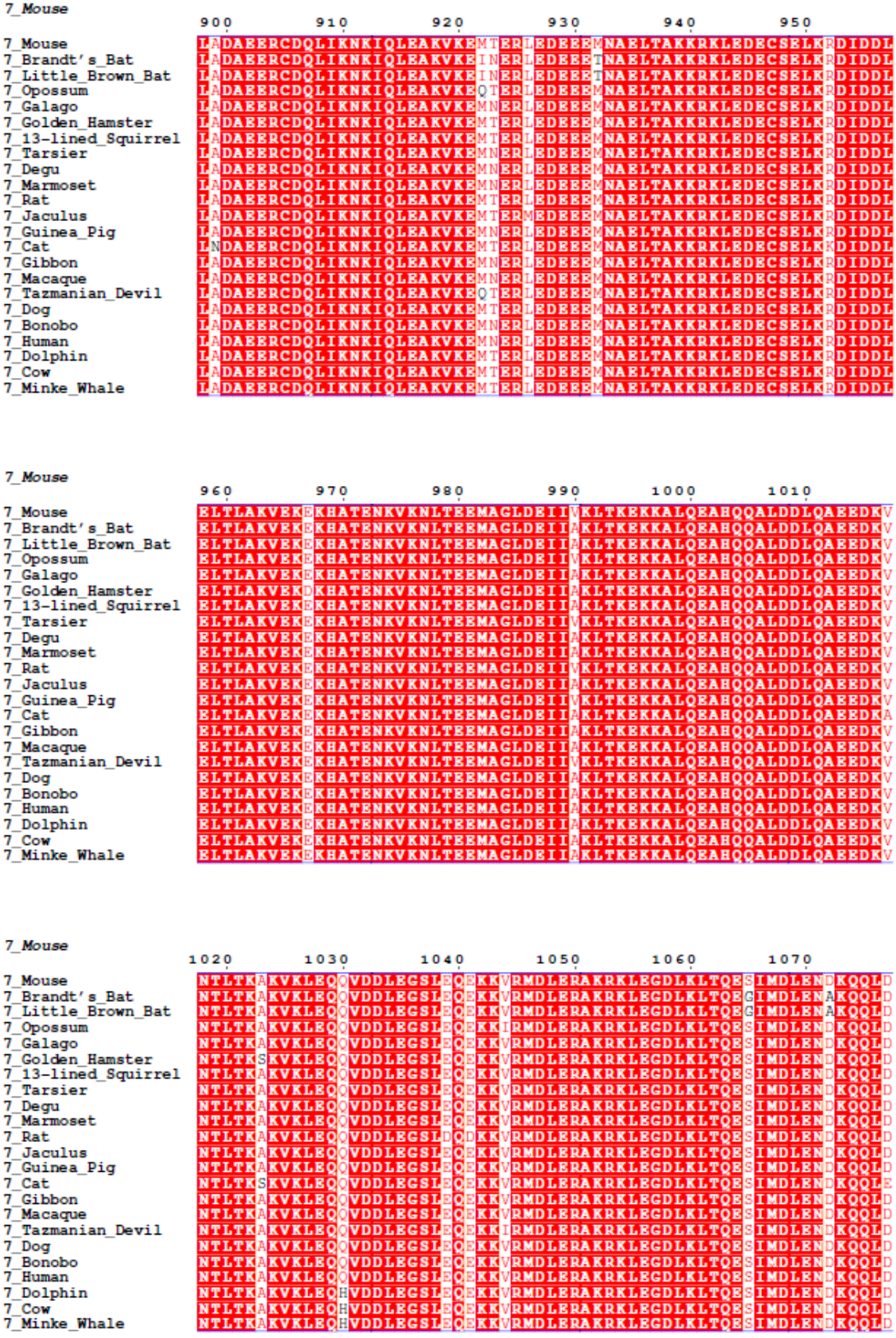

α14
η9
α15
α16
TT
7_Mouse
480
490
500
510
520
530
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

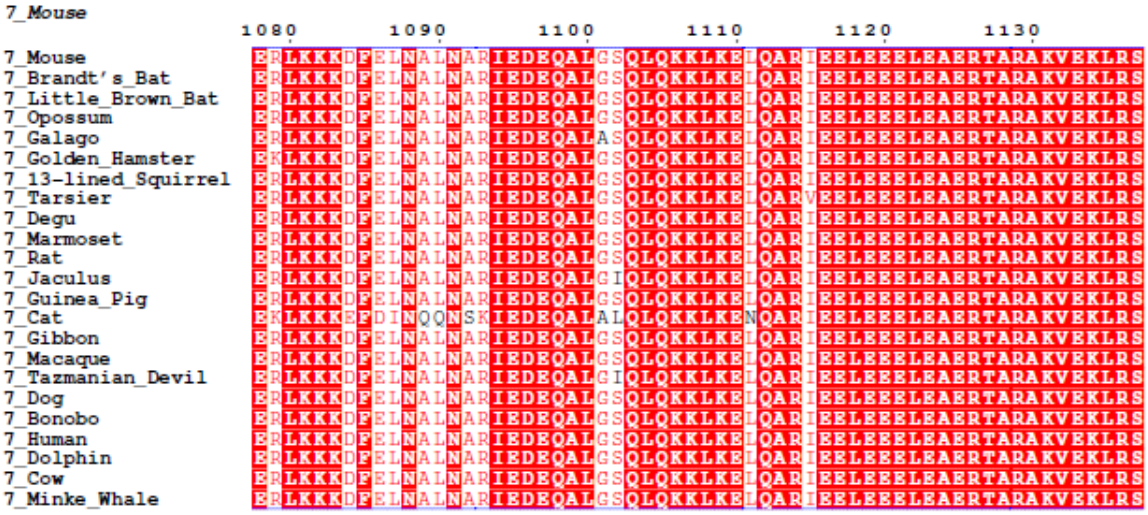

7_Mouse
α17
β16
β17
β18
α18
540
550
560
570
580
590
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

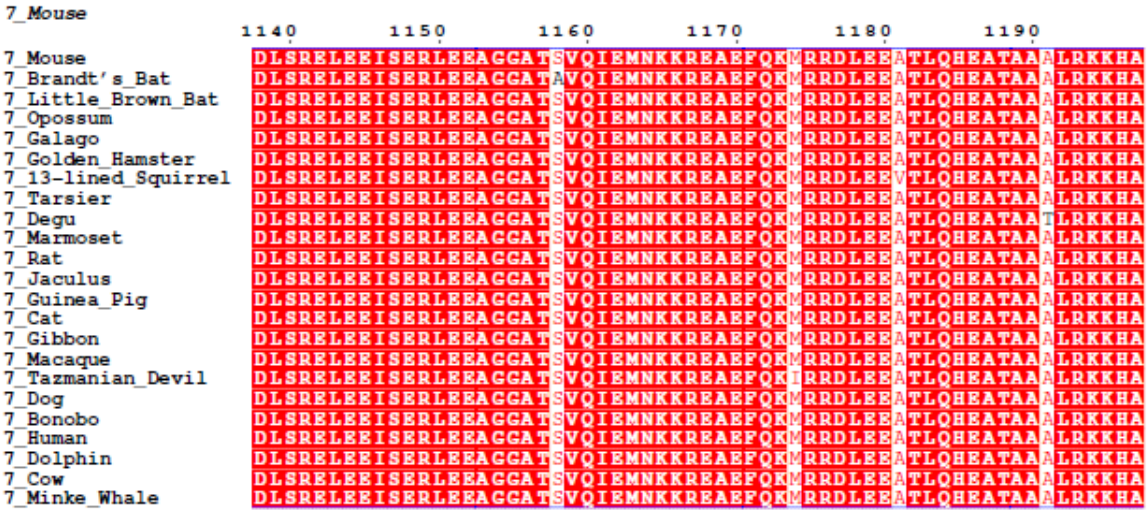

7_Mouse
α19
α20
α21
600
610
620
630
640
650
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

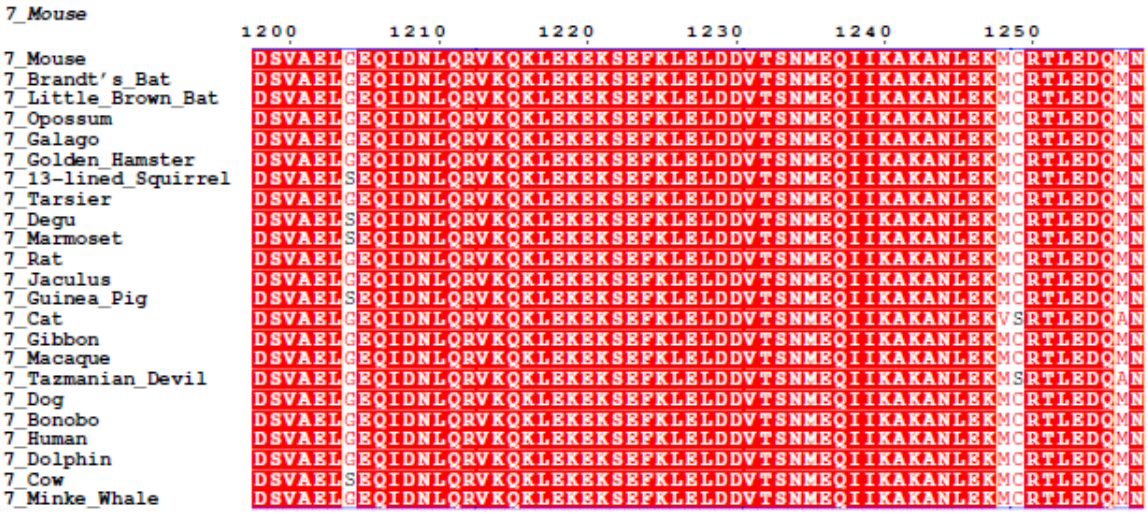

7_Mouse
β19
α22
α23
β20
660
670
680
690
700
710
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

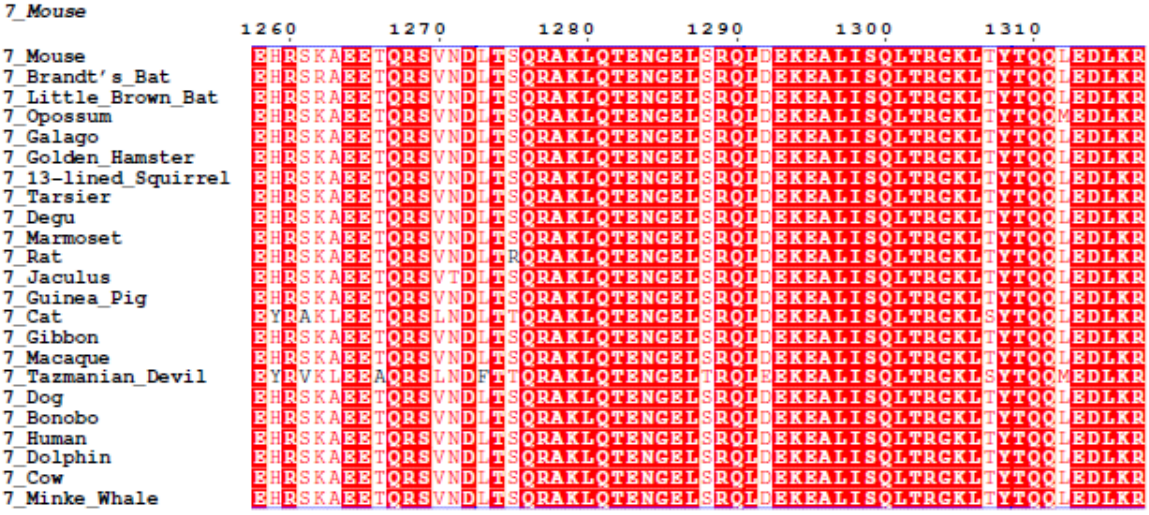

α24
η10 β21
β22
α25
7_Mouse
720
730
740
750
760
770
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale
780
790
800
810
820
830
840
850
860
870
880
890

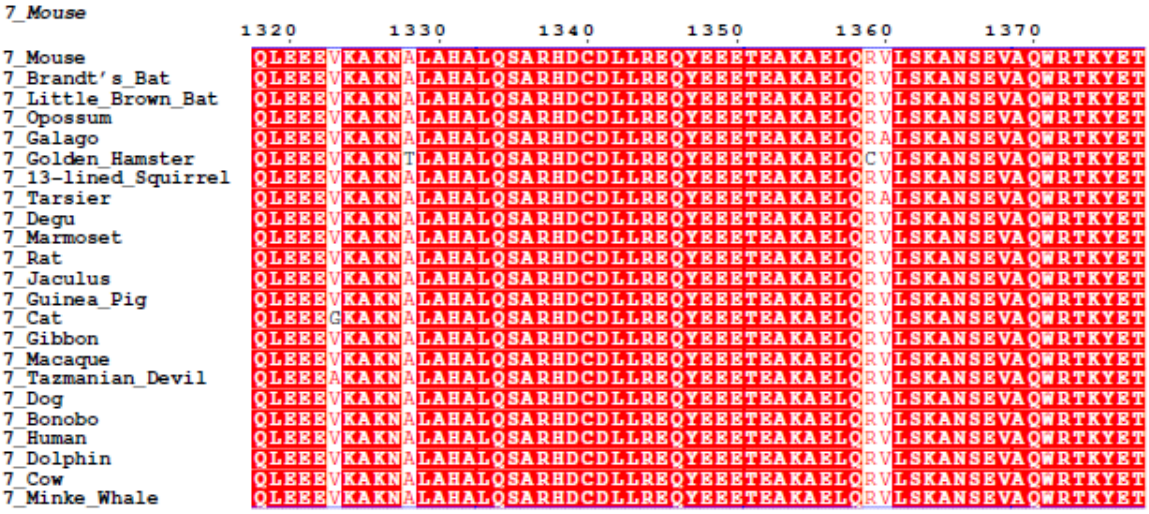

7_Mouse
900 910 920 930 940 950
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale
7_Mouse
960 970 980 990 1000 1010
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale
7_Mouse
1020 1030 1040 1050 1060 1070
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

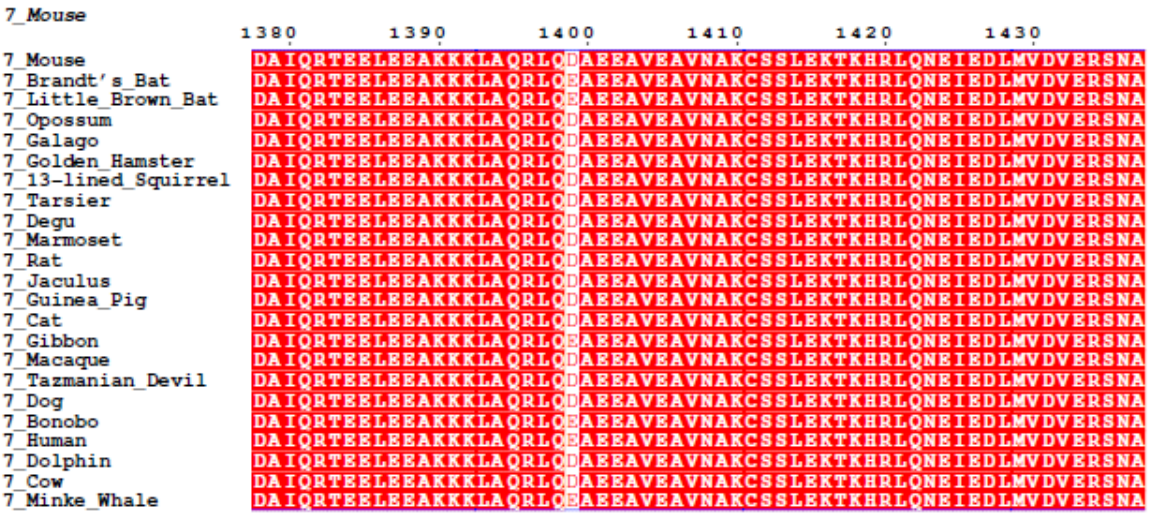

7_Mouse
1080
1090
1100
1110
1120
1130
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

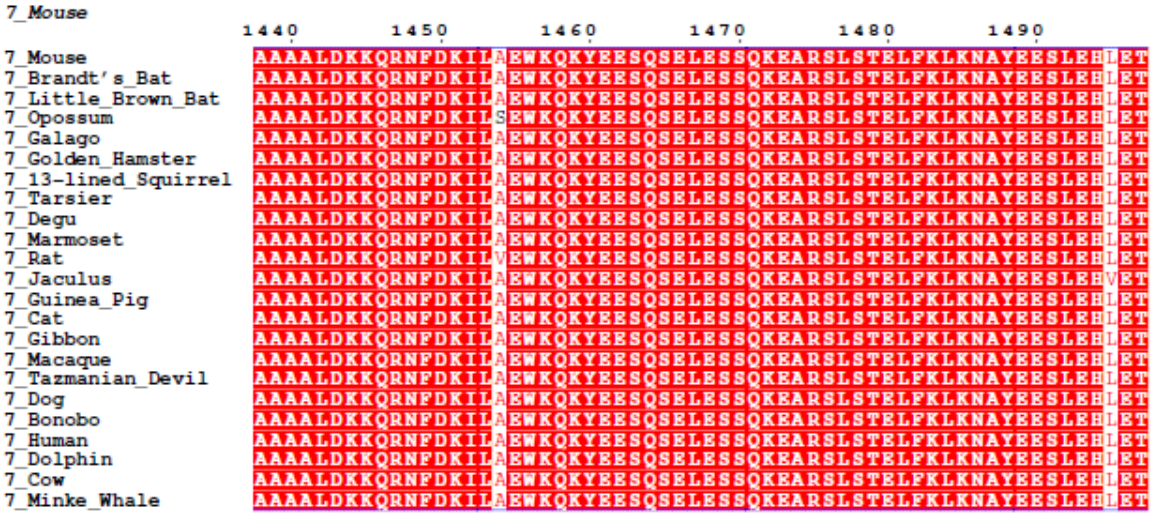

7_Mouse
1140
1150
1160
1170
1180
1190
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

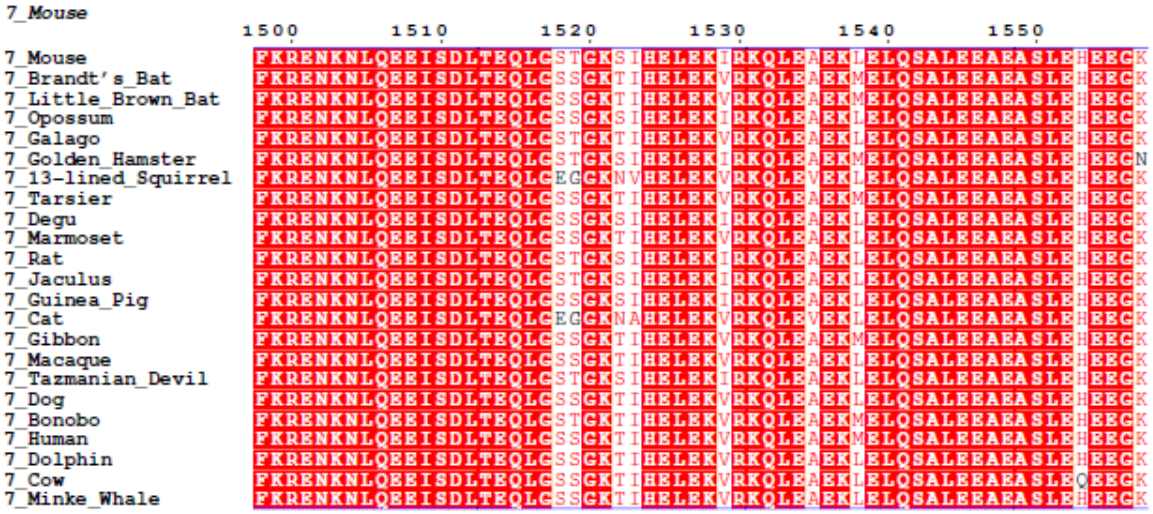

7_Mouse
1200
1210
1220
1230
1240
1250
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale


7_Mouse
1260
1270
1280
1290
1300
1310
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

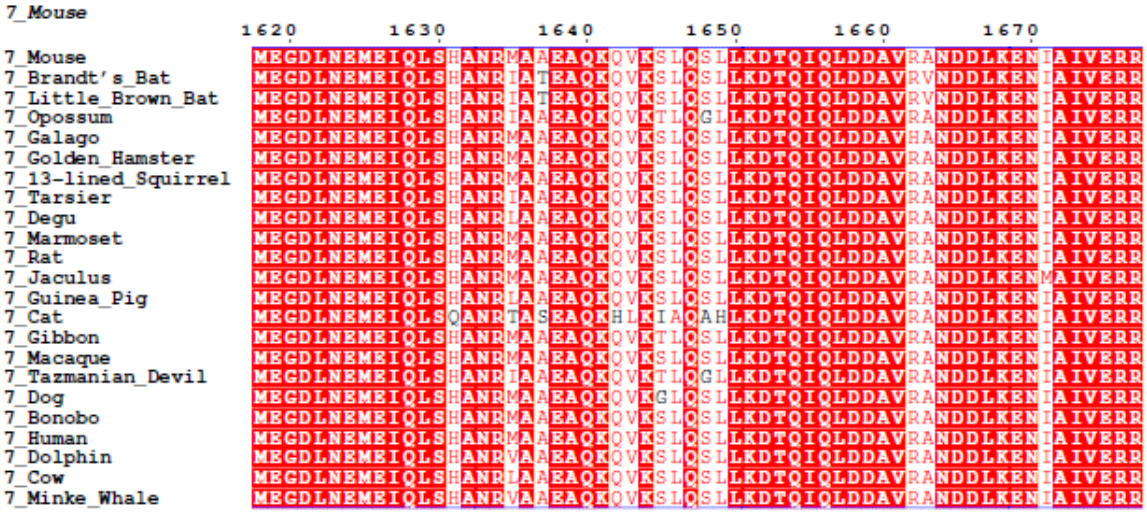

7_Mouse
1320
1330
1340
1350
1360
1370
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

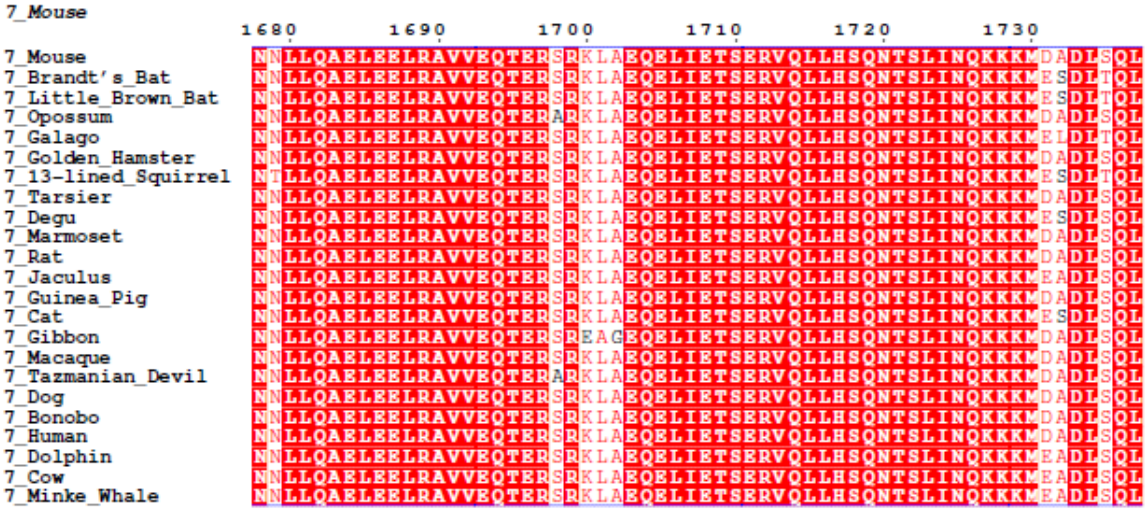

7_Mouse
1380
1390
1400
1410
1420
1430
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

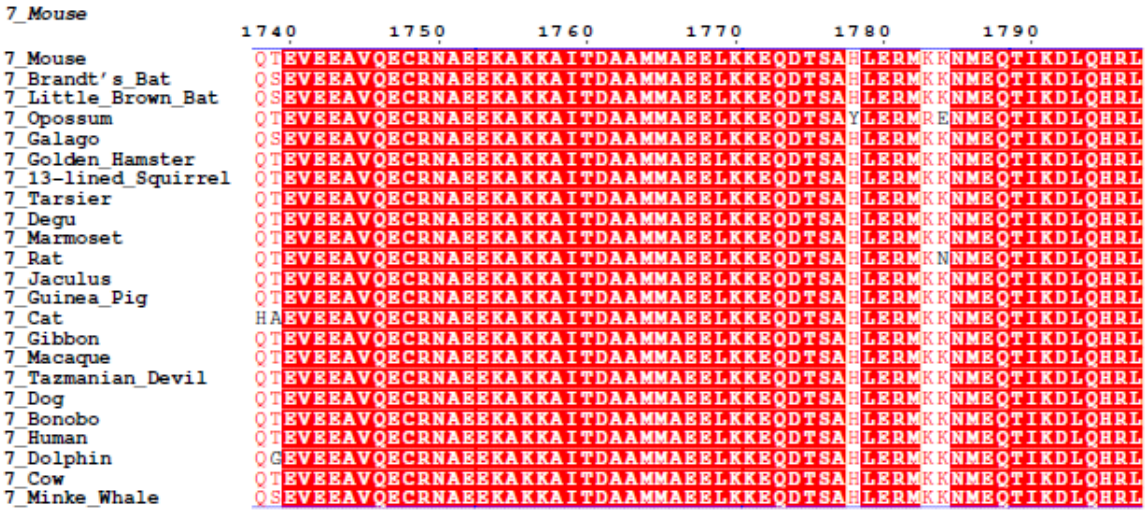

7_Mouse
1440
1450
1460
1470
1480
1490
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

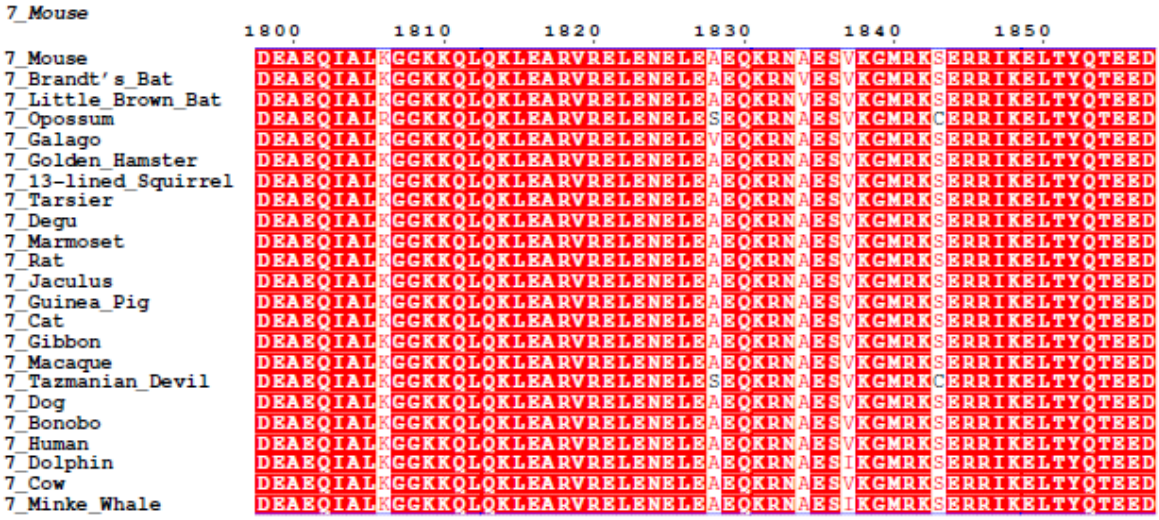

7_Mouse
1500
1510
1520
1530
1540
1550
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

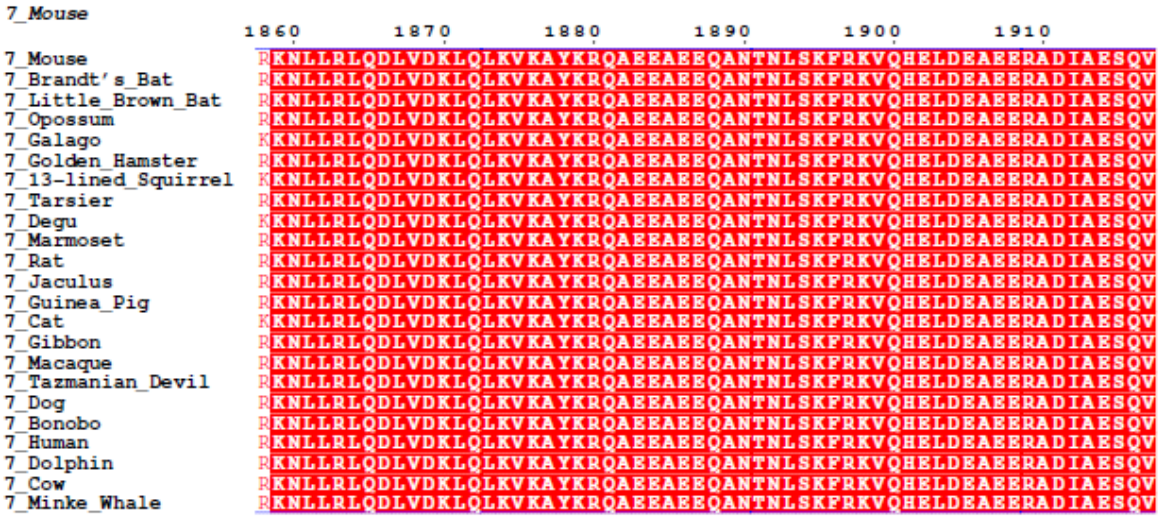

7_Mouse
1560
1570
1580
1590
1600
1610
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale

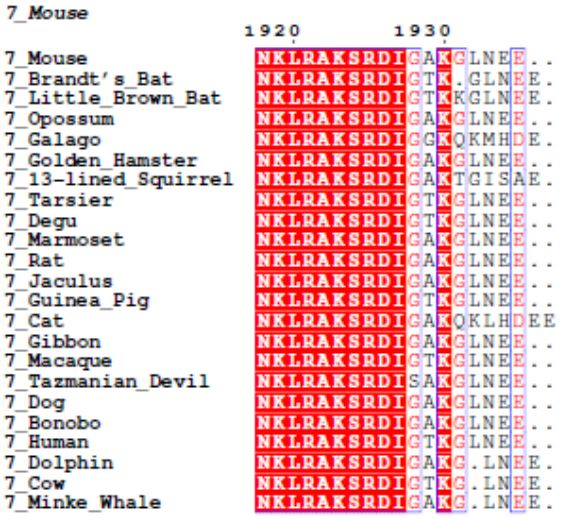

7_Mouse
1620
1630
1640
1650
1660
1670
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale


7_Mouse
1680
1690
1700
1710
1720
1730
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale


7_Mouse
1740
1750
1760
1770
1780
1790
7_Mouse
7_Brandt's_Bat
7_Little_Brown_Bat
7_Opossum
7_Galago
7_Golden_Hamster
7_13-lined_Squirrel
7_Tarsier
7_Degu
7_Marmoset
7_Rat
7_Jaculus
7_Guinea_Pig
7_Cat
7_Gibbon
7_Macaque
7_Tazmanian_Devil
7_Dog
7_Bonobo
7_Human
7_Dolphin
7_Cow
7_Minke_Whale


7_Mouse
1800
1810
1820
1830
1840
1850
7_Mouse DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Brandt's_Bat DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNVESVKGMRKSERRIKELTYQTEED
7_Little_Brown_Bat DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNVESVKGMRKSERRIKELTYQTEED
7_Opossum DEAEQIALRGGKKQLQKLEARVRELENELESEQKRNAESVKGMRKCERRIKELTYQTEED
7_Galago DEAEQIALKGGKKQLQKLEARVRELENELEVEQKRNAESVKGMRKSERRIKELTYQTEED
7_Golden_Hamster DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_13-lined_Squirrel DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Tarsier DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Degu DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Marmoset DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Rat DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Jaculus DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Guinea_Pig DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Cat DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Gibbon DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Macaque DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Tazmanian_Devil DEAEQIALKGGKKQLQKLEARVRELENELESEQKRNAESVKGMRKCERRIKELTYQTEED
7_Dog DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Bonobo DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Human DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Dolphin DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESIKGMRKSERRIKELTYQTEED
7_Cow DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESVKGMRKSERRIKELTYQTEED
7_Minke_Whale DEAEQIALKGGKKQLQKLEARVRELENELEAEQKRNAESIKGMRKSERRIKELTYQTEED


7_Mouse
1860
1870
1880
1890
1900
1910
7_Mouse RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Brandt's_Bat RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Little_Brown_Bat RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Opossum RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Galago KKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Golden_Hamster RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_13-lined_Squirrel KKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Tarsier RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Degu KKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Marmoset RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Rat RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Jaculus RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Guinea_Pig RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Cat KKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Gibbon RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Macaque RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Tazmanian_Devil RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Dog RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Bonobo RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Human RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Dolphin RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Cow RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV
7_Minke_Whale RKNLLRLQDLVDKLQLKVKAYKRQAEEAEEQANTNLSKFRKVQHELDEAEERADIAESQV


7_Mouse
1920
1930
7_Mouse NKLRAKSRDIGAKGLNEE..
7_Brandt's_Bat NKLRAKSRDIGTK.GLNEE.
7_Little_Brown_Bat NKLRAKSRDIGTKKGLNEE.
7_Opossum NKLRAKSRDIGAKGLNEE..
7_Galago NKLRAKSRDIGGKQKMHDE.
7_Golden_Hamster NKLRAKSRDIGAKGLNEE..
7_13-lined_Squirrel NKLRAKSRDIGAKTGISAE.
7_Tarsier NKLRAKSRDIGTKGLNEE..
7_Degu NKLRAKSRDIGTKGLNEE..
7_Marmoset NKLRAKSRDIGAKGLNEE..
7_Rat NKLRAKSRDIGAKGLNEE..
7_Jaculus NKLRAKSRDIGAKGLNEE..
7_Guinea_Pig NKLRAKSRDIGTKGLNEE..
7_Cat NKLRAKSRDIGAKQKLHDEE
7_Gibbon NKLRAKSRDIGAKGLNEE..
7_Macaque NKLRAKSRDIGTKGLNEE..
7_Tazmanian_Devil NKLRAKSRDISAKGLNEE..
7_Dog NKLRAKSRDIGAKGLNEE..
7_Bonobo NKLRAKSRDIGAKGLNEE..
7_Human NKLRAKSRDIGTKGLNEE..
7_Dolphin NKLRAKSRDIGAKG.LNEE.
7_Cow NKLRAKSRDIGTKG.LNEE.
7_Minke_Whale NKLRAKSRDIGAKG.LNEE.