UNIVERSITY OF NAPLES FEDERICO II

DOCTORATE
MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXI CICLO

Functional assessment of new MYBPC3 variants associated with Hypertrophic Cardiomyopathy

Tutor
Prof. Giulia Frisso

Co-Tutor
Prof. Jorge Alegre-Cebollada

Candidate
Maria Rosaria Pricolo

COORDINATOR

Prof. Vittorio Enrico
Avvedimento

Academic Year 2017/2018

INDEX

A Mimmo

e ai miei nipoti,

Dario, Luna e Simona

Tra le tante canzoni che hanno accompagnato la scrittura di questa tesi c'è una in particolare che è rimasta al mio fianco dall'inizio alla fine, Thunder. Nella prima strofa cita:

"I was dreaming of bigger things
And wanna leave my own life behind
Not a yes sir, not a follower"

Ebbene si, forse questa frase racchiude alla perfezione il mio percorso di dottorato. Ho iniziato sognando, del resto sognatrice lo sono sempre stata. Scioccamente mi sono presa la libertà di immaginare un futuro. Ma da cattiva sognatrice so ritornare alla realtà e affrontarla. Ho voltato pagina. E' stata dura, ansie per il tempo che scorreva, esperimenti di un campo che all'inizio non era il mio, due anni in un altro paese lontano da casa. Ma sono qui a terminare questo percorso e carica per affrontare il mio nuovo futuro.

Lungo la strada tante persone sono state al mio fianco, e io ne ho molte che ho bisogno di ringraziare per per avermi incitato, sopportato e per aver condiviso con me questa passione per la scienza.

Prima tra tutte c'è Maria Pia, amica nonché "sorella". Ricordo ancora quando insieme aprimmo la graduatoria di ammissione al dottorato, e tutta la cioccolata che abbiamo mangiato per festeggiare! Abbiamo condiviso tutto,

anche il desiderio di cambiare e non posso che augurarle il meglio per il suo nuovo percorso spagnolo.
In secondo luogo la mia particolare gratitudine va al LabJAC. Mi avete accolto come una famiglia. In primis ringrazio Jorge per essere stato un capo ma anche un mentore. Elìas che mi ha insegnato la termodinamica e tutte le parole spagnole fondamentali! Ma un ringraziamento speciale va alla chicas. Carla e le nostre lunghe camminate parlando di scienza e vita, sperando che google maps non ci porti di nuovo nei meandri sperduti di Madrid! Diana, senza la quale non so come avrei potuto fare tutti gli esperimenti. Ringrazio anche la sua musica che accompagna sempre le nostre giornate in lab. Natalia e la sua grinta, la cui schiettezza ti riporta sempre con i piedi per terra. Carmen e le nostre corse di notte al CNIC. Grazie per avermi insegnato tutto sull'AFM e per aver ascoltato i miei lunghi monologhi!! Infine c'è Angel e le sue GIF. Ora ci starebbe bene la canzone di Snoop Dogg non credi?! Raquel e Antonio (detto Pio IX) che per me sono labJAC ugualmente.
Un forte ringraziamento va anche a Marianna, "mia compagna madrileña", e a tutte le volte che mi ha aiutata a staccare la spina.
In questi ringraziamenti non si possono non menzionare anche tutti gli amici di una vita. Un pensiero speciale va a te Gigi, fratello mio. Siamo entrambi lontani ma ugualmente tanto vicini. Sappi che sarai sempre il mio pensiero la domenica pomeriggio durante le mie camminate in solitaria!
Ma ringrazio anche Domenico che oltre Maria Pia ha dovuto sopportare anche me. Umberto e Antonella che sempre allegrano le mie giornate, anche se vediamo le cose e le persone differentemente!!
Ringrazio la famiglia di Mimmo, nonché la mia famiglia, i sui genitori, Tommaso e Rosaria, la sorella, Margherita. Vi sento al mio fianco sempre, in passato come ora. Devo a voi tanto, specialmente tutto l'amore che mi avete dato.
In questi ringraziamenti non posso non includere mio padre, per quanto diversi e talvolta lontani so che è orgoglioso di me. Ti voglio bene papà.
Infine ringrazio te Mimmo, compagno ma anche amico. Senza il tuo supporto non sarei mai arrivata "viva" alla fine. Sono contenta di ciò che abbiamo costruito insieme, e so che ovunque mi porti la vita, per quanto sia tortuosa la strada, io e te saremo sempre uno al fianco dell'altro.

ABSTRACT

Hypertrophic cardiomyopathy (**HCM**) is the most common genetic disease of the myocardium. In ~60% of the cases HCM is caused by mutations in sarcomeric proteins, such as cardiac Myosin Binding Protein C (cMyBPC), which are responsible for generating the molecular force of myocyte contraction. A cohort of HCM patients have been screened for mutations in sarcomeric genes, and some new variants of cMyBPC of uncertain significance (VUS) were found. These new variants include two intronic variants (MYBPC3-c.506-2 A>C and MYBPC3-c.2308+3 G>C) and one missense variant (cMyBPC I603M), which were selected for functional study to determine pathogenicity.

The MYBPC3-c.506-2 A>C mutation was analysed in mRNA extracted from peripheral blood of the patient. The analysis revealed the loss of the canonical splice site and the utilization of an alternative splicing site, causing the loss of the first 7 nucleotides of exon 5. For the other variant, minigene constructs were generated to transfect HEK-293 cells. The minigene assay showed that mutation MYBPC3-c.2308+3 G>C also produces altered pre-mRNA processing, resulting in the skipping of the exon 23.

The mutation I603M localizes to domain C4 of cMyBPC. Using bioinformatics sequence analyses, a deleterious effect for I603M was predicted, but mRNA studies do not show any alteration of the splicing mechanism. At the protein level, homology modelling of domain C4 shows I603 to be buried in the protein structure, suggesting a potential destabilizing role of the I603M mutant. Indeed, circular dichroism spectroscopy and differential scanning calorimetry show a ~15ºC lower melting temperature for the mutant C4 domain. Finally, results obtained by single-molecule atomic force microscopy do not show a mechanical fingerprint for C4 indicating a very low mechanical stability of this domain.

Taken our results together, we propose that mutations c.506-2 A>C, c.2308+3 G>C and I603M lead to haploinsufficiency and cMyBPC protein destabilization, respectively causing the development of HCM.

In conclusion, the study of the functional consequences of mutations leads to assignment of pathogenicity of variants of uncertain significance.

ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
ACTC1	cardiac α-actin gene
AFM	Atomic Force Microscopy
BPS	Branch site
CD	Circular Dichroism
DHPR	Dihydropyridine receptor
DSC	Differential scanning calorimetry
HBB	Human β-globin gene
HCM	Hypertrophic cardiomyopathy
HGMD	Human Gene Mutation Database
LMM	Light meromyosin
LVH	Left ventricle hypertrophy
MHC-α	α-myosin
MHC-β	ß-myosin
MyBPC	Myosin-binding protein C
MYBPC3	Cardiac myosin-binding protein C gene
MYH6	Cardiac α-myosin heavy chain gene
MYH7	Cardiac ß-myosin heavy chain gene
MYL2	Regulatory myosin light chain gene
MYL3	Essential myosin light chain gene
PMI	Precision Medicine Initiative
PPT	Polypyrimidine tract
PTC	Premature termination codons
RLC	Myosin regulatory light chain
RyR2	Ryanodine receptor
SCD	Sudden cardiac death
SR	Sarcoplasmic reticulum

SRX	Superrelaxed tate
T_m	Temperature at the midpoint of thermal transition
Tm	Tropomyosin
TnI	Troponin I
TNNI3	Cardiac troponin I gene
TNNT2	Cardiac troponin T gene
TnT	Troponin T
TPM1	Cardiac α tropomyosin gene
UsnRNP	Uridine-rich small ribonucleoproteins
VUS	Variants of uncertain significance
WLC	Worm-like chain model of polymer elasticity
ΔH_{cal}	Calorimetric enthalpy
ΔH_v	Van't Hoff enthalpy

INTRODUCTION

INTRODUCTION

In the 1950s, using high-resolution microscopy, two independent groups of scientists observed changes in sarcomere length during contraction (Huxley and Hanson 1954; Huxley and Niedergerke 1954). The sarcomere is the contractile unit of striated muscle, defined as the distance between two regions called Z-disks or Z-lines. These regions consist of a central A-band, made up of myosin, flanked by two I-bands, which are composed by actin filaments (*Figure 1*) (Sweeney and Hammers 2018). Huxley and colleagues observed that during sarcomere contraction, the length of the A-band remains constant while the I-band shortens (Krans 2010). Since x-ray diffraction studies proved inextensibility of both thick and thin filaments of striated muscle, they established the "sliding filament theory" as the interaction of myosin and actin that induces shortening of the sarcomere, and thus the muscle, through the sliding of thick filaments on thin filaments (Mijailovich et al. 1996; Lin et al. 2017). Other elements of myofilaments were subsequently discovered to be involved in this mechanism.

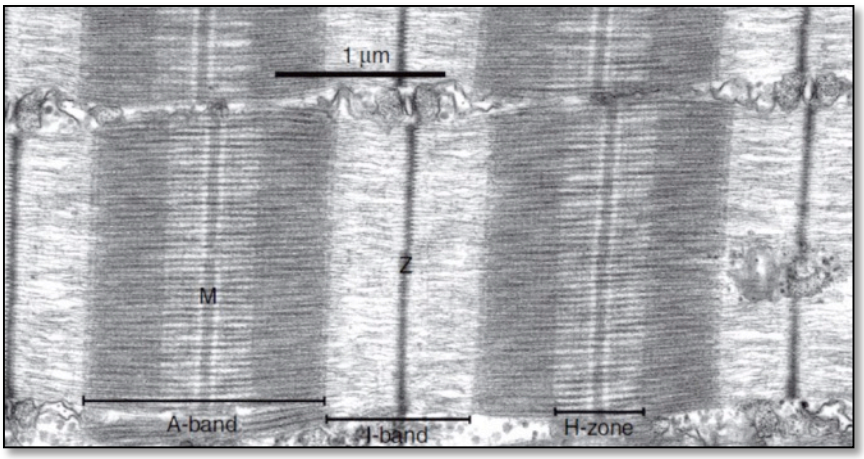

Figure 1: Electron micrograph of the sarcomere of skeletal muscle. A longitudinal thin section through a muscle fibre is shown. The major bands and lines are indicated as Z, Z-line, M, M-line, A-band, I-band and H-zone (Figure adapted from Sweeney and Hammers 2018)

Almost simultaneously to the Huxleys' research on sarcomere function during the 1950's, Robert Donald Teare, an English pathologist at St. George's Hospital in London, reported an autopsy of eight patients with asymmetrical hypertrophy of the heart and disorganized arrangement of muscle bundles associated with hypertrophy of individual muscle fibres and their nuclei (*Figure 2*). Seven of these caused sudden death in young adults (Teare 1958).

The disease that affected these patients was recognised as a distinct pathology, later named Hypertrophic Cardiomyopathy (HCM).

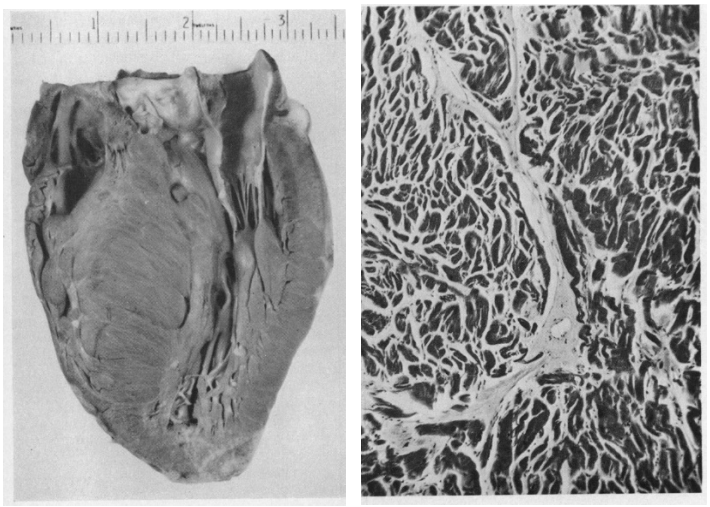

Figure 2: heart of the first HCM patient with localized hypertrophy of the interventricular septum (on the left). On the right is shown the disordered arrangement of muscle bundles of the same patient (Figure adapted from Teare 1958).

It was not until years later that the relationship between sarcomere defects and the development of hypertrophic cardiomyopathy was discovered.

1.1 The Sarcomere: the contractile unit of striated muscle

There are three types of muscle: skeletal, cardiac, and smooth. Skeletal and cardiac muscle are referred to as striated muscle because of the regular organization of sarcomeres. These contractile units consist of highly organized components that interact in a very precise manner to generate force and motion. Sarcomeres also integrate many signalling functions in muscle. Indeed, they are connected with other organelles, such as mitochondria or nucleus, by intermediate filaments to maintain cellular integrity and to contribute to mechanotransduction (Sweeney and Hammers 2018).

The sarcomere has three major filament systems that include thin filaments, thick filaments and the giant protein titin. The thin and thick filaments extend from the Z-disk and M-line, respectively, overlapping on each side of the M-line and creating distinctive regions within the sarcomere (*Figure 3*) (Henderson et al. 2017; Lin et al. 2017).

The primary component of thin filaments is **actin**, a helical rope-like structure

attached to the Z-disks. Actin monomers (G-actin) polymerize into long filaments (F-actin) at physiological ionic strength and in the presence of Mg^{2+} and K^{+}, forming a tightly wound helix (Roberto Dominguez and Kenneth C. Holmes 2011). G-actin consists of four subdomains (SD1- SD4) that form a binding pocket for ADP or ATP as well as the divalent ions Mg^{2+} and Ca^{2+}. The region SD1 and SD2 also interact with other myofilament proteins, such as myosin and myosin binding protein C (MyBPC). Specifically, actin SD1 binds to the myosin head, also known as myosin S1 via ionic interactions (Lin et al. 2017). Three homologous isoforms of actin (α, β and γ) are expressed in vertebrates. The α-actin, in turn, consist in skeletal, cardiac, and smooth variations. Skeletal and cardiac isoforms are differentially expressed: ACTC1 gene encodes α-cardiac-actin, whereas α-skeletal-actin is expressed by ACTA1 gene (Henderson et al. 2017).

Other thin filament proteins, **tropomyosin** (Tm) and **troponin** (Tn), regulate the contraction of striated muscle. Tropomyosin is an elongated coiled-coil molecule that binds to the surface of the thin actin filament (*Figure 3 and Figure 4*). Troponin comprises three subunits: troponin T (TnT), which binds to tropomyosin; troponin I (TnI), which binds to both actin and tropomyosin; and troponin C (TnC), which confers Ca^{2+} sensitivity to the system (Sweeney and Hammers 2018).

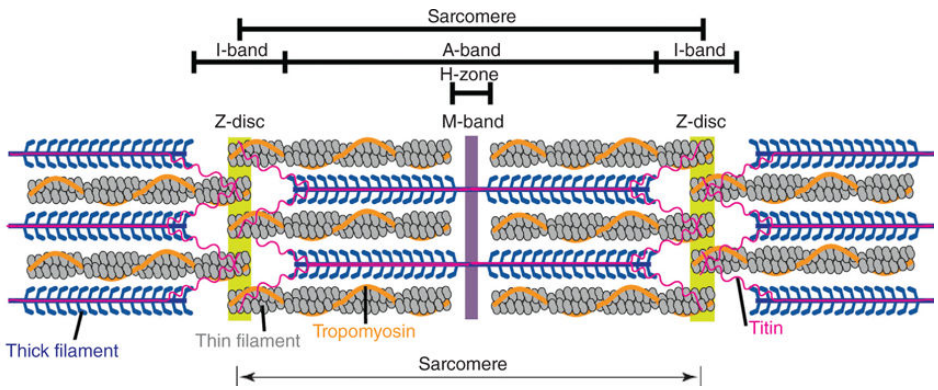

Figure 3: schematic diagram depicting the three major filament systems found in sarcomeres. Thin filament (gray), thick filament (blue), titin (pink) (Figure adapted from Henderson et al. 2017).

Myosin is the main component of the thick filament. Myosin is the molecular motor that drives muscle contraction trough interaction with thin filaments. Myosin is composed of heavy chain fragment (MHC) and light chain fragment (MLC). Myosin heavy chain consist of three regions: N-terminal head (S1),

neck (S2) and a C-terminal long tail region known as light meromyosin (LMM) (*Figure 4*). The S1 subfragment projects outward from the thick filament and contains a motor domain that binds actin filaments to form cross-bridges. This interaction drives actin-based movement via ATP hydrolysis. The neck region transduces force through its lever arm and the highly variable tail region mediates myosin interactions. Myosin light chains regulate MHC motor function by binding to the neck region (Henderson et al. 2017).

Two major isoforms of myosin are present in heart muscle: α-myosin (MHC-α) and ß-myosin (MHC-β) expressed by the MYH6 and MYH7 genes, respectively. In cardiac muscle of small mammals (e.g., mice), MHC-β is the predominant isoform during development and is replaced by MHC-α in the adult. In large mammals such as humans, MHC-α is expresses during development while MHC-β is the predominant isoform in the adult heart. In failing hearts (mouse and human), MHC can shift to the fetal isoform as a maladaptive response to severe cardiovascular stress (Lin et al. 2017).

Another component of the thick filament is the **myosin-binding protein** family (MyBP). This family consists of myosin-binding protein C (MyBPC) and myosin-binding protein H (MyBPH). Both MyBPs are located within the C-zone of A-band region and are arranged in 7 to 9 transverse stripes spaced at 43-nm intervals. There are three isoforms of MyBPC: a slow skeletal isoform, a fast skeletal isoform and a cardiac isoform, encoded by the distinct genes MYBPC1, MYBPC2, and MYBPC3, respectively. There is only one isoform of MyBPH (human gene MYBPH) expressed in the Purkinje fibers and fast twitch skeletal muscle fibers. MyBPC is thought to link the thick and thin filament systems and further regulate cross-bridge cycling by displacing tropomyosin and competing with myosin for actin binding. Specifically, the N-terminus of MyBPC binds actin filaments, regulating contraction by altering the actin-activated myosin ATPase activity. MyBPC's N-terminal M-motif also interacts with the S2 region of myosin, leading to regulation of the contraction. The C-terminus of cMyBPC binds titin and is necessary to localize MyBPC to the A-band (*Figure 4*). This interaction also allows MyBPC to possibly work in concert with all of the filament systems during contraction to impact force development, transmission, sensing, and signalling (Henderson et al. 2017).

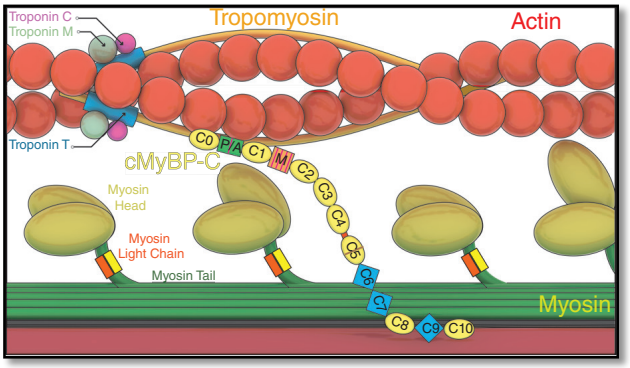

Figure 4: representation of structure and binding of thin and thick filaments proteins (Figure adapted from Lin et al. 2017).

Titin is the third myofilament system in the sarcomere spanning the length of the half sarcomere: the N-terminal region anchors to the Z-disk and extends across the I-band and A-band toward the M-line (*Figure 3*). The elastic I-band region of titin consists of a combination of tandem immunoglobulin domains (Ig), and random coil regions such as the PEVK (*Figure 5*). The PEVK region consists of approximately 70% proline (P), glutamic acid (E), valine (V), and lysine (K) residues, thus the name. The Ig domains and the random coil regions are crucial extensible sites conferring elasticity to the I-band region of titin. Titin within the A-band is composed of super repeats of seven fibronectin III domains (Fn3) and four Ig domains. This inextensible portion of titin interacts with MyBPC and the myosin tail domains, thus linking titin to the thick filaments. Differential splicing of the single titin gene (TTN) produces titin isoforms of different lengths. In adult cardiac muscle, two different splicing isoforms coexist (N2B and N2BA), which differ in their I-band regions. The N2B is the stiffest and shortest isoform (3.0 MDa), which is characterised by the absence of N2A element. N2BA is the longest isoform (3.3 to 3.5 MDa) and includes both N2A and N2B fragments.

Titin is thought of as a “molecular spring” and is key to the passive mechanical properties of the myofilaments contributing to passive tension during diastolic relaxation (Henderson et al. 2017; Lin et al. 2017).

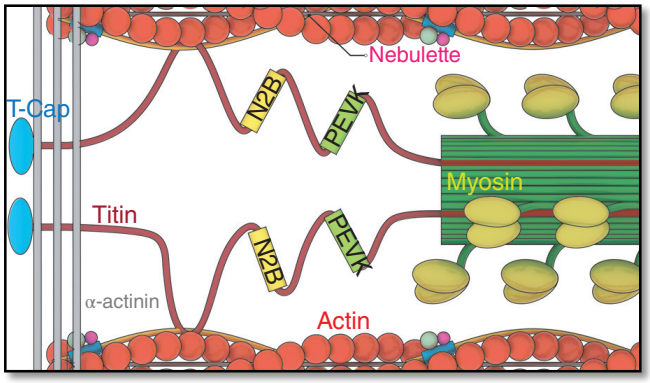

Figure 5: schematic structure of Titin containing N2B and PEVK regions. Interactions with thick and thin filaments are also represented (Figure adapted from Lin et al. 2017).

1.1.1 Sarcomere contraction and relaxation

The sarcomeres are responsible for contraction of muscle cells through a process known as excitation-contraction coupling. This process occurs via electrical stimulus linked to mechanical contraction movement. The process starts when an electrical signal, modulated in part by autonomic innervation, induces calcium release. The electrical signal results in an initial inward flux of Ca^{2+} from voltage-dependent Ca^{2+} channels, known as dihydropyridine receptors (DHPR). This initial ion flow is detected by ryanodine receptors (RyR2 in the heart), that release Ca^{2+} stores from the sarcoplasmic reticulum (SR) (Lin et al. 2017). The rising levels of Ca^{2+} initiate contraction by binding the amino-terminal Ca^{2+} binding-site of TnC within the Tn complex. The TnC consist of two globular domains (N- and C-terminal domains) connected via an α-helical bridge. The N-terminal region is a regulatory domain, whereas the C-terminal motif has a structural role by anchoring TnC to TnI. The binding of Ca^{2+} to the N-terminal globular domain partially opens a hydrophobic region of TnC and alters its interaction with TnI, which induces TnI movement. The shift of TnI is transmitted to Tm by TnT. When Ca^{2+} binds to TnC, TnT disengages from interaction with Tm, allowing Tm to move its position relative to actin, such that its myosin-binding site is exposed (*Figure 6*) (Dupuis et al. 2016; Lin et al. 2017). During this phase of contraction, myosin heads are in a pre-power stroke conformation in which the myosin binds ADP and P_i. Release of P_i hydrolytic product facilitates myosin head binding at a new position along the actin filament, and a myosin power stroke pulls Z-lines toward each other resulting in sarcomere shortening and muscle contraction. Finally, myosin releases ADP allowing new ATP binding and dissociation of the myosin head from actin. Myosin ATPase then hydrolyzes ATP, prompting the myosin head's converter domain to change shape. ATP hydrolysis enables the myosin

head to re-adopt a pre-power stroke conformation, so a new 'cross-bridge cycle' can be repeated (Garfinkel et al. 2018).

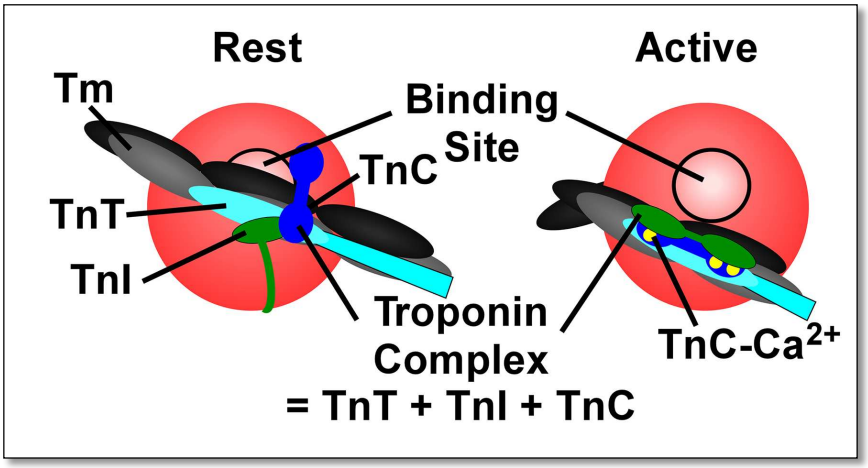

Figure 6: schematic representation of Troponin complex activation. (**Left**) At rest, the actin monomer (*in red*) is blocked by troponin complex. (**Right**) A conformational change in the Tn complex removes the Tm from the blocking position (Figure adapted from Dupuis et al. 2016).

Muscle contraction is triggered and regulated by calcium levels and Ca^{2+}-sensor proteins. The calcium regulatory system acts in a length-dependent fashion by which elongated sarcomeres exhibit increased calcium sensitivity. Although the major gatekeeper of the calcium-based regulation is troponin-tropomyosin complex, cMyBPC and titin have been shown to participate in Ca^{2+}-mediated regulation of contraction. Essentially, cMyBPC phosphorylation increases myofilament calcium sensitivity and, at short sarcomere lengths, titin reduces the length-dependence of the calcium regulatory system (*Figure 7*) (Garfinkel et al. 2018).

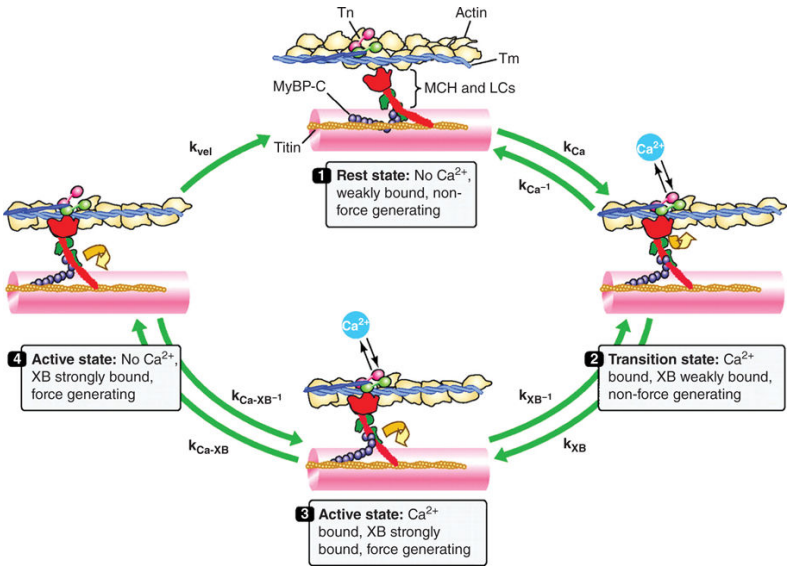

Figure 7: schematic drawing of the actomyosin cross-bridge cycle. Rest state (1), transition state (2), active state (3) and active state with loss of bound Ca2+ (4) are represented (Figure adapted from Henderson et al. 2017).

At low myofilament calcium concentrations, tropomyosin blocks the myosin-binding site on actin and results in sarcomere relaxation. Two distinct structures of paired and relaxed myosins have been identified, each associated with different rates of energy consumption (*Figure 8*). The disordered relaxed (DRX) state occurs when one myosin head interacts with the thick filament backbone, resulting in steric inhibition of its ATPase, whereas its partner myosin retains ATP hydrolysis activity and the potential for actomyosin interactions and force production. The superrelaxed (SRX) state occurs when both myosin heads dock onto the thick filament backbone, inhibiting ATP hydrolysis and withdrawing both myosins from thin filament interaction and force production (Garfinkel et al. 2018).

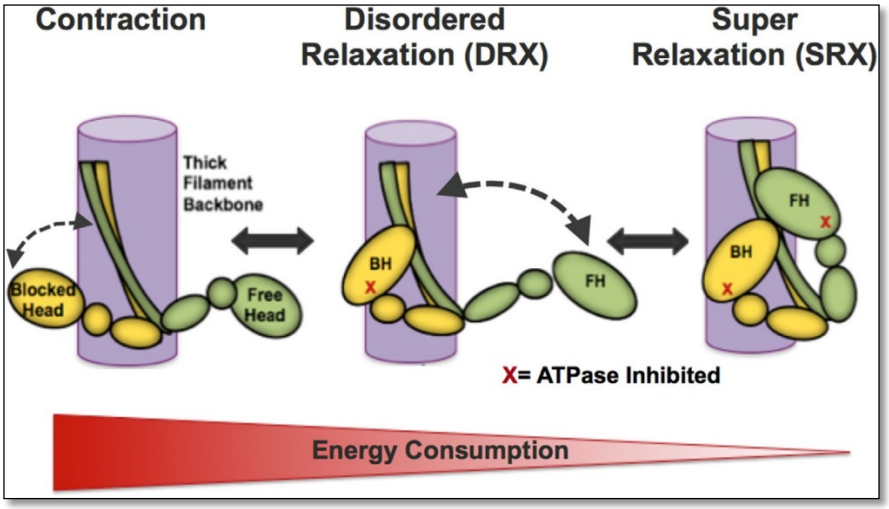

Figure 8: representation of sarcomere conformations and associated energy consumption throughout the cardiac cycle (Figure adapted from Garfinkel et al. 2018).

1.2 Hypertrophic Cardiomyopathy (HCM): definition and clinical manifestations

HCM is the most common inherited cardiovascular disease with an estimated prevalence ranging from 1:500 to 1:200 and the most frequent cause of sudden cardiac death (SCD). The risk of SCD is high in young patients, including athletes, and it may be the first manifestation of the disease (Wijnker et al. 2018).

The clinical manifestations of HCM are highly variable. Many patients are asymptomatic and are diagnosed incidentally, others may manifest shortness of breath, chest pain, palpitations or syncope, triggered by left ventricle hypertrophy (LVH). LVH typically manifests as asymmetric hypertrophy with greatest involvement of the basal interventricular septum subjacent to the aortic valve. At the cellular level, histopathologic study of heart muscle in HCM demonstrates that cardiac myocytes are hypertrophied, disorganized, and separated by areas of interstitial fibrosis (Geske et al. 2018).

Approximately, 5%-10% of patients with HCM progress to end-stage disease with impaired systolic function and dilatation. Atrial fibrillation is present in nearly 1 of 5 patients, accompanied by significant risk of stroke (Sabater-Molina et al. 2018).

1.2.1 Genetics bases of HCM

The first evidence of genetic involvement in Hypertrophic Cardiomyopathy was published in 1989, where a genetic-linkage analyses identified a locus on chromosome 14 that was inherited with hypertrophic cardiomyopathy in over 100 family members (Jarcho et al. 1989). Many genomic studies in the decades following have found that the majority of HCM cases are caused by mutations in sarcomere protein genes that cause HCM in an autosomal dominant pattern. For that reason, Hypertrophic Cardiomyopathy is recognized as a disease of the sarcomere (Ingles et al. 2018).

To date, most pathogenic variants detected in HCM patients affect by cardiac myosin-binding protein C (MYBPC3) and β-myosin heavy chain (MYH7) genes, together being responsible for approximately half of the familial HCM cases. Mutations in troponin T, troponin I, and Tm (TNNT2, TNNI3, TPM1) constitute less than 10% of diagnosed HCM cases. Other genes whose mutations cause HCM include MYL2 (myosin light chain 2), MYL3 (myosin light chain 3), and ACTC1 (cardiac α-actin), although with a much lower frequency (1%-5% each) (Table 1) (Liew et al. 2017).

Table 1: Mutations in sarcomeric genes that cause Hypertrophic Cardiomyopathy

Gene	Protein	Frequency (%)*
Cardiac myosin-binding protein C	MYBPC3	30-40%
β-myosin heavy chain	MYH7	20-30%
Cardiac troponin T	TNNT2	5-10%
Cardiac troponin I	TNNI3	4-8%
Regulatory myosin light chain	MYL2	2-4%
Essential myosin light chain	MYL3	1-2%
α tropomyosin	TPM1	<1%
α cardiac actin	ACTC1	<1%

*the frequencies refer to Liew et al. 2017

HCM is an inherited disorder with an autosomal dominant pattern characterised by incomplete penetrance and variable expressivity. The variability of the phenotype is due to the causal mutation acting in concert with many other genetic and non-genetic influences. Indeed, the primary defect is the mutation which affect the structure and function of proteins. Although, the phenotype includes also the structural changes and secondary molecular events in the myocardium, such as activation of the hypertrophic signalling pathways, which occur in response to changes in protein (*Figure 9*) (Marian and Braunwald 2017; Wijnker et al. 2018).

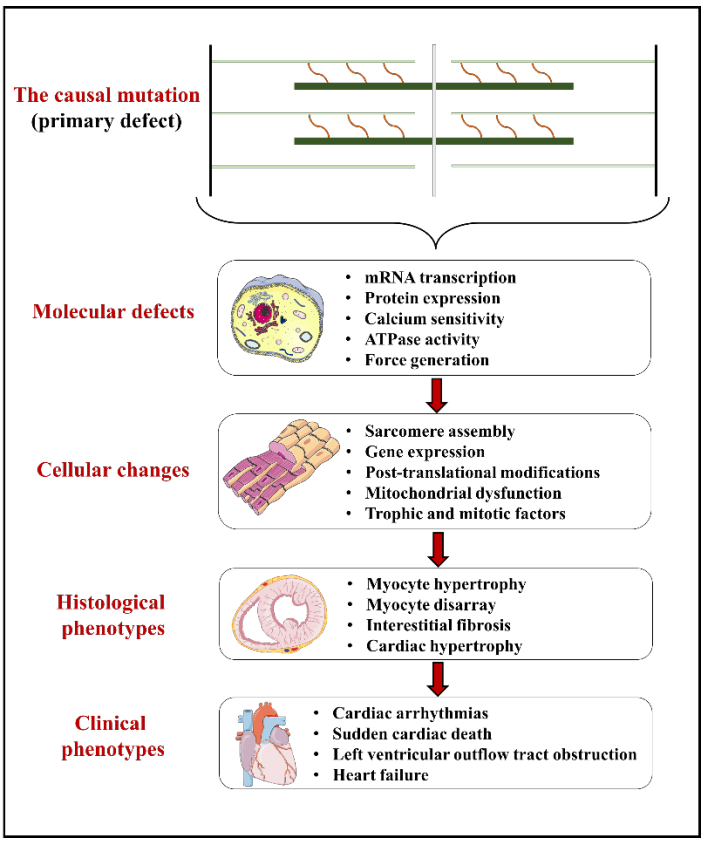

Figure 9: hypothetical pathogenesis of hypertrophic cardiomyopathy. The primary defect is the mutation in the sarcomere which causes a series of primordial defects. These initial defects activate expression of a series of molecular and cellular changes, which result in histological and morphological changes in the myocardium, such as myocyte hypertrophy and fibrosis. Finally, these changes lead to the clinical phenotypes of HCM, such as cardiac arrhythmias and heart failure.

The initial defects, in accord with the diversity of HCM mutations, are also diverse. The majority of the HCM mutations are missense mutations that alter proteins functions. Missense mutations that impart structural changes in the encoded protein may also reduce efficiency of sarcomere assembly. Another fraction of the HCM mutations result in premature truncation of the encoded proteins because of a gain of a stop codon or a frameshift. These kind of mutations commonly abolish protein expression (Marian and Braunwald 2017).

Such heterogeneity in the expression and incorporation of the mutant proteins at the myocyte level might in part explain the variable phenotypic expression of HCM.

1.2.2 Diagnosis of HCM and the role of genetic testing

The clinical diagnosis of HCM is based on the presence of left ventricular hypertrophy, typically defined by the end-diastolic ventricular septal thickness. Although, in the diagnostic criteria, the European Society of Cardiology guidelines recommend using a left ventricular wall thickness of ≥15 mm, HCM may be diagnosed with a value >13 mm, always occurring in the absence of abnormal loading conditions or other secondary causes. The cut point of 13 mm in adults offers high sensitivity in detecting HCM but has the risk of overdiagnosis (Marian and Braunwald 2017). However, in patient with lesser degrees of wall thickening (13–14 mm), the diagnosis of HCM requires evaluation of other features including family history, non-cardiac symptoms and signs, electrocardiogram abnormalities, multimodality cardiac imaging and genetic testing (Wijnker et al. 2018).

Up to 60% of patients who meet the HCM diagnostic criteria demonstrate a pathogenic sarcomeric mutation on genetic testing. However, the clinical role of genetic testing in HCM largely centers on family screening. If a causative genetic mutation is identified in a proband with HCM, testing for this mutation becomes the preferred method of family screening to facilitate presymptomatic diagnosis of family members, clinical surveillance and reproductive advice (Geske et al. 2018). It is very important to make a correct classification of variants and identification of their pathogenicity and functional relevance for an appropriate use of genetic testing in clinical practice.

Due to the increased complexity of analysis and interpretation of genetic tests, general guidelines for the interpretation of variants have been published (Hershberger et al. 2018). By applying the proposed score, variants can be classified into 5 main groups: pathogenic, likely pathogenic, uncertain significance (VUS), likely benign and benign. The American College of Medical Genetics and Genomics (ACMG) guidelines proposed a process to classify mutation composed by six items. The first one is frequency of variants in control population in international database (Exome Sequencing Project, 1000Genome or ExAc). Then, it. is important to know whether the variant has been reported previously as associated with disease by control disease databases such as Human Gene Mutation Database (HGMD) and ClinVar, or sequence databases like NCBI Genome and RefSeqGene. In silico pathogenicity scores is the third method proposed for the classification. This evaluation can be done using different software (eg, MutationTaster, Polyphen2, SIFT, Pmut, ESEfinder) to predict the potential impact of the variant on protein structure and/or function. The item number four is the evaluation of conservation to determining if the involved nucleotide and amino

acid residue are conserved among species and isoforms and if they are located in a strategic functional domain. However, this evidence needs careful evaluation in the absence of functional study or clinical cosegregation. For that reason, the most important items are the cosegregation studies and functional studies (Hershberger et al. 2018).

When the different criteria fail to produce assignment of pathogenicity, mutations are classifying as VUS. Clinical genetic testing often reports VUS with ambiguous and uncertain pathogenicity. VUS may cause difficult clinical interpretation and risk-assessment that limit counselling and treatment of individuals with unclear genetic predisposition to disease. Therefore, this can cause anxiety and stress in asymptomatic individuals and families who are found to carry a VUS (Ma et al. 2018).

Genetic test in HCM typically involves a minimum of 8 sarcomere genes (MYH7, MYBPC3, TNNT2, TNNI3, TPM1, MYL2, MYL3 and ACTC1), although approximately 80% who test positive have causative variants in MYBPC3 and MYH7. The mutations occurring in most sarcomeric genes are missense-type, with a dominant negative effect resulting in the formation of poison peptides, which can be incorporated into the sarcomere. Protein-truncating variants, which are not incorporated into the sarcomere cause HCM by haploinsufficiency. Most MYBPC3 mutations result in frame shifts and protein truncations (Sabater-Molina et al. 2018).

In this thesis, new mutations in the MYBPC3 gene were found and characterized.

1.3 Cardiac Myosin Binding Protein C (cMyBPC)

G. Offer et al. in 1973 discovered the Myosin Binding Protein C (MyBPC) as an impurity in skeletal muscle myosin preparations (Offer et al. 1973). Only 10 years later the MyBPC was determined and characterized (Yamamoto and Moos 1983).

The cardiac isoform (cMyBPC) is exclusively expressed in the heart resulting from the MYBPC3 gene composed of 34 coding exons (Carrier et al. 2015). cMyBPC is a multi-modular structural protein composed by eight immunoglobulin (Ig) and three fibronectin type-3 repeating domains (*Figure 10*). cMyBPC differs from its two skeletal isoforms by having a C0 domain, a proline-alanine-rich (Pro-Ala) region located between domains C0 and C1, a phosphorylatable motif (M-motif), which lies between the C1 and C2 domains, and a small insertion within the C5 domain (Sadayappan and De Tombe 2014).

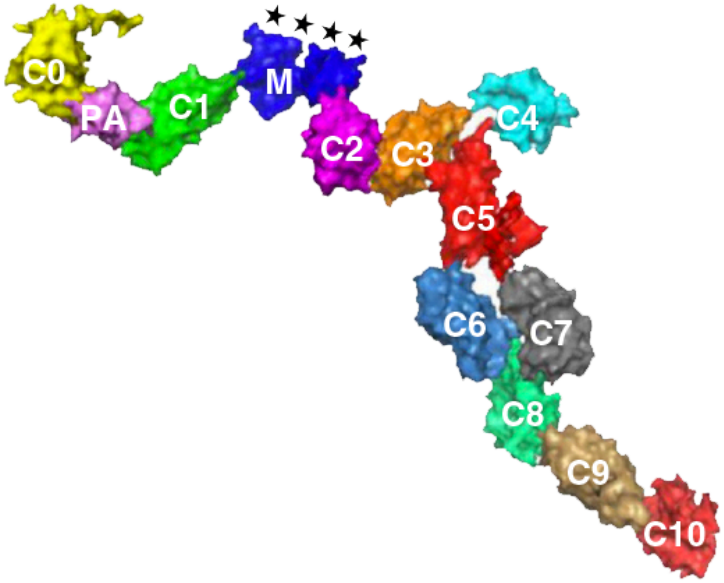

Figure 10: model of full-length human cardiac MyBP-C including all domains. The four serine phosphorylation sites on the M domain are indicated with asterisks (Nag et al. 2017).

cMyBPC is arranged in the C-zone of sarcomeres to run vertically through thick and thin filaments. The C-terminal domains of cMyBP-C bind to the light meromyosin (LMM) portion of β-myosin heavy chain and to titin. Residues in the N-terminus can interact with actin, the S1/S2 junction of myosin heavy chain, and the myosin regulatory light chain (RLC). The binding of both actin and myosin allows cMyBPC to perform the function as modulator of contraction and myofilament rigidity (McNamara et al. 2017). Indeed, the absence of cMyBP-C in the effective null mouse models leads to reduced rigor cross-bridge-dependent stiffness and thick filament stiffness (Palmer BM et al. 2011).

Mutations in the MYBPC3 gene represent the 40–50% of all HCM mutations, making it the most frequently mutated gene in this disease. Many MYBPC3 mutations are truncating mutations, including nonsense mutations, nucleotide insertions or deletions and RNA splicing mutations that result in frameshifts, premature termination codons (PTC) and C-terminal truncated polypeptides. The myocardium of patients harboring cMyBPC truncating mutations contains less cMyBPC levels, suggesting that haploinsufficiency may be a pathogenic mechanism in truncating mutations (Van Dijk et al. 2009; Marston et al. 2009).

Indeed, studies with engineered heart tissue provided that contractile function is affected when cMyBPC protein levels are below 73% (Wijnker et al. 2018). In contrast to truncating mutations, missense mutations lead, in most of the cases to stable mutant cMyBPCs that are, at least in part, incorporated into the sarcomere and could act as poison polypeptides on the structure and/or function of the sarcomere (Carrier et al. 2015).

HCM mutations are found in all domains of the protein, also in central domains that do not interact with myosin or actin. This observation has raised the hypothesis that thermodynamic as well as mechanical defects in these domains can result in altered function of cMyBPC. Indeed, missense mutations located in the central domains (C5-C6-C7) and clinically linked to HCM have been identified to interfere with domain folding (Nadvi et al. 2016). Furthermore, single-molecule Atomic Force Microscopy (AFM) data on C3 domains shows altered mechanical behaviour of mutant proteins (Suay-Corredera et al. 2018).

1.4 RNA Splicing mechanism

The removal of introns from pre-mRNA and the joining of exons is a critical aspect of gene expression. Introns are removed from primary transcripts by the process of RNA splicing, which links together the flanking exons to generate functional mature RNA (mRNA) (Papasaikas et al. 2016). The splicing reaction takes place in two catalytic steps involving two consecutive trans-esterification reactions (Figure 11).

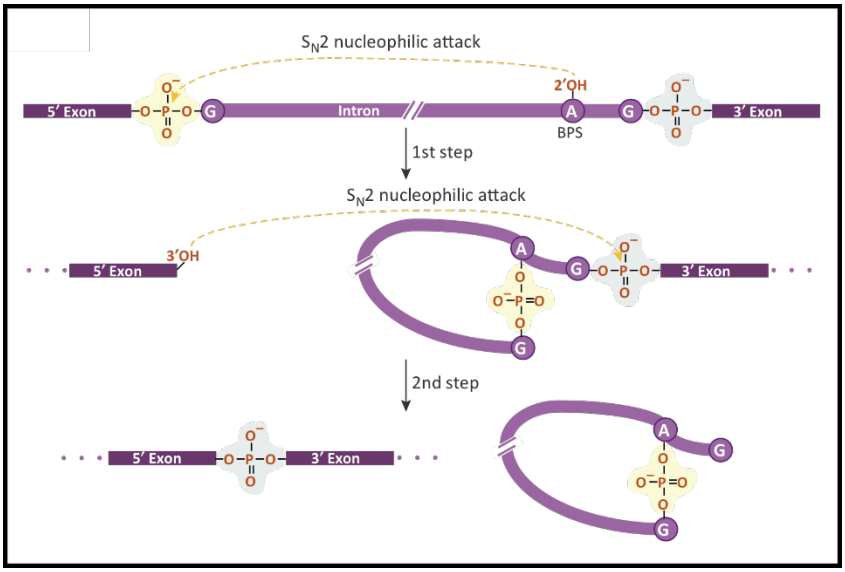

Figure 11: chemical mechanism of the RNA splicing reaction. RNA splicing entails

two transesterification steps. In the first, the hydroxyl group of an intronic adenosine residue of BPS carries out a nucleophilic attack on the phosphate group in the exon/intron junction of donor splice site, generating a lariat intermediate. In the second step, the free 3'-OH of the 5'-exon attacks the phosphate group in acceptor splice site, splicing the two exons together and releasing the intron lariat.

The exact sites for the trans-esterification reactions are defined by consensus sequences, called donor and acceptor sites. The human donor site (5'-splicing site) is MAG/GURAGU relative to the exon/intron junction (M is A or C; R is purine). In mammals only the first two bases of the intron (GU) are universally conserved. The acceptor consensus sequence at the 3'-splice site is defined by three separate elements: the branch site (BPS), the polypyrimidine tract (PPT), and the intron dinucleotide AG. Together these elements extend 100 nucleotides into the intron (Buratti et al. 2007).

The splicing reaction takes place in the spliceosome, which is formed by several uridine-rich small ribonucleoproteins (UsnRNP), and numerous non-snRNP splicing factors. These proteins excise each intron and join exons in the correct order through a multitude of RNA-RNA, RNA-protein, and protein-protein interactions (Baralle and Baralle 2005).

The most common consequences of splicing mutations are skipping of one or more exons and/or full intron retention (Tazi et al. 2009, Xiong et al. 2015). The mutations that cause splicing alterations involve disruption of native splicing sites and/or activation of aberrant 5'-donor and/or 3'-acceptor splice sites.

2. AIMS

Hypertrophic cardiomyopathy (HCM) is a genetic disorder, characterized by left ventricular hypertrophy in the absence of triggers such as hypertension. HCM is the most common inherited cardiovascular disease with an estimated prevalence ranging from 1:500 to 1:200. The genetic defects leading to HCM usually involve sarcomere disruption. The most frequently mutated HCM gene is MYBPC3, encoding cardiac myosin-binding protein C (cMyBPC).

In genetic screens of hereditary disorders, several different types of variants can be found: pathogenic mutations, benign polymorphisms and variants of uncertain significance (VUS). VUS are an ambiguous category of variants whose effects remain unclear, and therefore do not have a clinical value. According to the American College of Medical Genetics and Genomics (ACMG), in the absence of enough genetic support, functional studies are the most important criterion to establish causality of putative disease-causing mutations (Hershberger et al. 2018).

In order to understand the pathogenic basis of Hypertrophic Cardiomyopathy, to improve characterization of VUS and, finally, to improve molecular diagnosis, the evaluation of pathogenic effects by *in vitro* analysis can be helpful for interpretation of variants. In this context, the global aims of this project are:

1. Genotyping HCM patients to describe variants in proteins of the sarcomere that are associated with HCM.

2. Describing new variants in MYBPC3 that can cause HCM.

3. Integrating cosegregation analysis and functional assessment of MYBPC3 variants associated with HCM to refine assignment of pathogenicity.

MATERIALS AND METHODS

3. MATERIALS AND METHODS

3.1 Materials

3.1.1 Mammalian Cell culture

Human embryonic kidney 293 cell line (HEK-293) was used for Minigene transfection and RNA analyses. HEK-293 cells were grown as a monolayer in Dulbecco's modified eagle medium supplemented with 10% FBS, 2 mM L-glutamine and 1% penicillin/streptomycin in a humidified, 5% CO2 atmosphere at 37°C.

To transfect pMG plasmid, the cells were plated 1 day before transfection onto 60mm plates at a density of approximately $5x10^5$ cells/plate, in a 4ml of final volume of complete medium, in order to be 60-70% confluent the following day.

3.1.2 Bacteria: conservation and manipulation

Chemically Competent *Escherichia coli* cells were used for cloning, plasmid propagation and protein expression. Competent bacteria were prepared using calcium chloride protocol (J. F. Sambrook and D.W. Russell 2001). In short, a single colony of *E. coli* from fresh LB-plate without antibiotics, was pre-inoculated in 5ml of LB medium and incubated overnight, shaking at 37°C. Then, the starter culture of bacteria was inoculated in 100ml of LB medium and grown at 37°C. When the optical density at 600 nm (OD_{600}) reached 0.4, the cells were immediately put on ice. The bacteria culture was split into four parts, centrifuged at 4°C, resuspended in 20ml of ice-cold 50mM $CaCl_2$ each and left on ice for at least 40 minutes. Next, the cells were harvested by centrifugation at 2000g at 4°C. The pellet was resuspended in about 2 mL of ice cold 50mM $CaCl_2$ including 20% glycerol. The suspension was divided in 200µL aliquot into sterile 1.5 mL tubes and stored at -80°C.

Top10 (Invitrogen™) and BLR (DE3) (Novagen-Merck Millipore) strains of *E. coli* were used in this thesis.

TOP10 are provided at a transformation efficiency of 1×10^9 cfu/µg supercoiled DNA and are ideal for high efficiency cloning and plasmid propagation. They allow stable replication of high-copy number plasmids.
Genotype: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G].

BLR (DE3) is a *rec*A$^-$ derivative of BL21 that improves plasmid monomer yields and can help stabilize target plasmids containing repetitive sequences. DE3 indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lac*UV5 promoter. Such strains are suitable for production of protein from target genes cloned in vectors by induction with Isopropyl β-D-1-thiogalactopyranoside (IPTG).
Genotype: F$^-$ *omp*T *hsd*S$_B$(r_B^- m_B^-) *gal lac ile dcm* Δ(*srl-rec*A)306::T*n*10 (*tet*^{R})(DE3).

3.1.3 Plasmid constructs

To study the effect of genetic variants on splicing pattern, a minigene construct (pMG) was used. This plasmid results from introducing some modifications to pEGFP-C3 vector (Amato et al. 2012). The artificial minigene contains the human ß-globin gene fragment amplified from genomic DNA of the HEK293 cell line. A KpnI restriction site was inserted in the middle of the second intron of the ß-globin gene and was used to clone genomic DNA fragments for splicing studies. The graphic representation of minigene including exon 23 of MYBPC3 is shown in *Figure 11*.

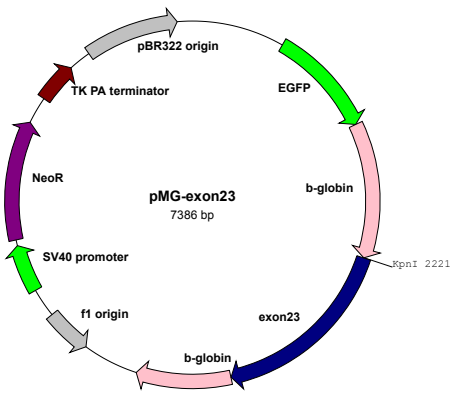

Figure 11: graphic representation of the minigene construct with MYBPC3 exon 23 insert (pMG-exon23). The sequences related to ß-globin include exon 1, exon2, exon 3 and related intron sequences; whereas for MYBPC3 the exon 23 together with its flanking regions were cloned between exon 2 and exon 3 of the ß-globin.

The proteins used for biophysical characterization were engineered and expressed using pT7 and pQE80L plasmids, respectively (*Figure 12 A, B*). Expression of proteins from pQE80L results in the addition of a His-tag at the N-terminus of the proteins, and a Cys-tag at the C-terminus. In order to avoid

C-terminal cysteines, pQE80L was modified for the cloning of monomer proteins.

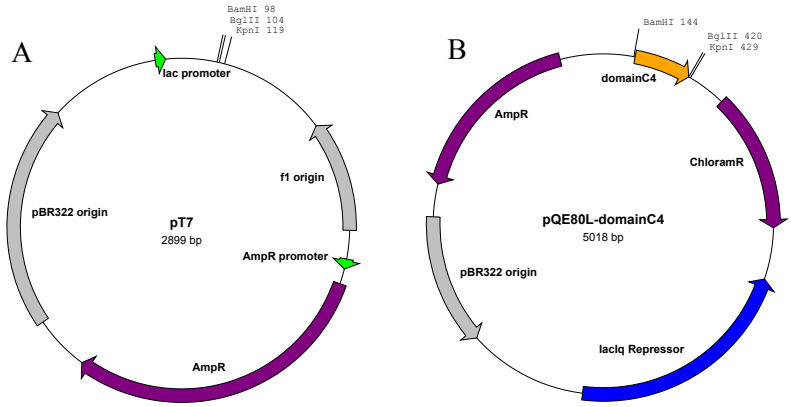

Figure 12: plasmids used for proteins engineerin and expression. Graphic representation of pT7 (A) and pQE80L (B) including a fragment codified for the C4 domain of cMyBPC.

3.1.4 Primers

The primers were designed using Primer3 tool (http://primer3.ut.ee/). The specification of primers for input PCR template was evaluated with Blat tool from UCSC Genome Browser, analysing multisequence alignments. The quality was evaluated with OligoAnalyzer Tool from IDT-Integrated DNA technologies.
In the Table 2 the primers used and their sequence are summarized.

Table 2: List of primers

Primer name	Tail Sequence	Primer Sequence (5'-3')
MYBPC3-ex3-4 Fw	----------	CAAGTCCCAAAGGGTCAAGCTC
MYBPC3-ex6 Rv	----------	GTGGACACCTCACAGCGGTA
MYBPC3-ex23 Fw	CGGGGTACC	CCTGGCTCCCTTCATCCTA
MYBPC3-ex23 Rv	CGGGGTACC	CACCCCCAGATCCAAAGAG
GFP Fwd	----------	ACGACGGCAACTACAAGACC
β-globin Rv	----------	CACACCAGCCACCACTTTC
MYBPC3-ex15 Fw	----------	CAAGCGTACCCTGACCATCA
MYBPC3-ex20-21 Rv	----------	GGATCTTGGGAGGTTCCTGC
MYBPC3-ex22-23 Rv	----------	CACACAGCAGCTTCTTGTCA
cMyBPC C4 Fw	CGCGGATCC	AAGCTGGAGGTGTACCAGAGCAT
cMyBPC C4 Rv	ATAGGTACCTTAGCAACAAGATCT	CTCCATGAAGTGGAGCTTGGCTG
cMyBPC C4 I603M Fw	----------	GTCCACAAACTGACCATGGACGA CGTCACACCT

The primers used for splicing analyses using mRNA from peripheral blood are designed to recognise exon sequences of MYBPC3. The forward primer (MYBPC3-ex3-4 Fw) anneals with the end region of exon 3 and initial sequence of exon 4. The MYBPC3-ex6 Rv recognises the exon 6 of MYBPC3. MYBPC3-ex23 Fw and MYBPC3-ex23 Rv were used to generate the insert of pMG vector. These primers carried a tail region that includes the sequence of KpnI restriction site (GGTACC) The oligonucleotides GFP Fwd and β-globin Rv were designed to amplify by RT-PCR and sequence the mRNA product in Minigene experiments, and anneal with the GFP sequence and the exon 3 of β-globin, respectively. The MYBPC3-ex15 Fw, MYBPC3-ex20-21 Rv and MYBPC3-ex22-23 Rv oligonucleotides anneal with exon 15 of MYBPC3, the region between exon 20 and exon 21, and the sequence straddling exons 22 and 23, respectively. To generate cDNA cloning for the C4 domain, the primers cMyBPC C4 Fw and cMyBPC C4 Rv were used. The Forward primer is complementary to exon 18 of MYBPC3 and carries a tail region with one BamHI restriction site (GGATCC). The reverse primer carries a tail region including KpnI and BglII (AGATCT) restriction sites, codons introducing two extra cysteines (GCAACA, the reverse and complentary sequence of TGT TGC) and a stop codon (TTA, reverse and complementary of TAA). The sequence of cMyBPC C4 Rv anneals with exon 19 of MYBPC3. Finally, the cMyBPC C4 I603M Fw was designed to introduce the mutation I603M in the domain C4.

3.2 Methods

3.2.1 General Molecular Biology

Standard molecular biology techniques (digestion, ligation, agarose electrophoresis, PCR, etc.) were performed according to standard protocols (J. F. Sambrook and D.W. Russell 2001).

3.2.2 Screening of HCM patients

Genomic DNA was isolated from peripheral whole blood with the Nucleon BACC2 kit (GE Healthcare, Life Sciences, Little Chalfont Buckinghamshire, UK). All coding exons, and 5′ and 3′ UTRs of genes involved in inherited cardiomyopathies associated to sudden death were amplified by PCR and analysed by automatic sequencing using previously reported protocols (F. Girolami et al. 2018). Thirty out of 249 patients were analysed using Next Generation Sequencing methodology for a large number of target genes. Informed consent to perform genetic analysis was obtained from patients according to the Helsinki Declaration. The nomenclatures are indicated

following HGVS recommendations for the descriptions of sequence variants (den Dunnen et al. 2016).

3.2.3 Bioinformatics assessment of mutations

Alamut software (Alamut® Visual, Interactive Biosoftware)was used for in silico prediction of splice-affecting nucleotide variants (Houdayer 2011). Genomic sequences (WT and mutant) were processed by this predictor software using five splicing prediction tools (SpliceSiteFinder-like, MaxEntScan, Neural Network Splice, GeneSplicer, and Human Splicing Finder). Each tool is based on different algorithms: **SpliceSiteFinder-like** uses position weight matrices computed from a set of human constitutive exon/intron junctions for donor and acceptor sites; **MaxEntScan** and **NNPLICE** are based on the Maximum Entropy principle and neural networks, respectively. **GeneSplicer** combines several splice site detection techniques, namely Markov models; and HSF is based on position weight matrices with some position-dependent logic.

Missense mutation at the protein level were evaluated with three independent bioinformatics tools which predict possible impact of amino acid substitutions on the structure and function of the protein (PolyPhen-2, SIFT, Provean). **PolyPhen-2** (Polymorphism Phenotyping v2) evaluates the impact of amino acid allelic variants via analysis of multiple sequence alignments and protein 3D-structures (Adzhubei et al. 2015). **SIFT** (Sorting Intolerant From Tolerant) presumes that important amino acids will be conserved in the protein family, so changes at well-conserved positions tend to be predicted as deleterious (Ng and Henikoff 2003). **PROVEAN** (Protein Variation Effect Analyzer) is a software tool which predicts whether an amino acid substitution has an impact on the biological function of a protein using pairwise sequence alignment scores (Choi and Chan 2015).

3.2.4 Homology modelling for protein structure

For modelling of cMyBPC's C4 domain, the protein sequence Q14896 from the UniProt database was used. The structure was modelled using I-TASSER tool using slow-MyBPC (2YUZ PDB) as template. I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction (Zhang 2008). It first identifies structural templates from the PDB by the multiple threading approach LOMETS. Then, full-length atomic models are constructed by iterative template fragment assembly simulations. PyMol software was used for molecular representation (DeLano 2002).

3.2.5 Splicing Analysis of mRNA from peripheral blood

Total RNA was extracted from 5-10 mL of patients' peripheral blood. Five volumes of 0.2% NaCl was added in order to lyse red blood cells and isolate a pellet that includes lymphocyte cells. Then, the RNA was isolated from this peleet using Trizol Reagent (Thermo Fischer Scientific, Waltham, MA, USA). RNA retro-transcription was performed by SuperScript VILO (Life Technologies) starting from 1μg of total RNA and using random primers. The cDNA obtained was amplified using specific consecutive-exon-spanning primers (Table 2). The same primers were used for sequencing of PCR fragment. Sequence was aligned with reference and analysed with CodonCode Aligner software (CodonCode Corporation, Dedham, MA, USA).

3.2.6 RNA Splicing Analysis by Minigene strategy

3.2.6.1 Insert Generation

Genomic DNA of patient carrying *MYBPC3-c.2308+3G>C* mutation was isolated from peripheral whole blood with the Nucleon BACC2 kit (GE Healthcare, Life Sciences, Little Chalfont Buckinghamshire, UK). The genomic regions affected by non-coding mutations were amplified by PCR using PCR Master Mix (Promega, Madison, WI, USA). The sizes of inserts obtained by PCR were approximately 1000 base pairs. The PCR amplicons contained the exon 23 and the 5′ and 3′ intronic flanking regions. The primers contained a KpnI restriction sequence at the 5′ terminus (Table 2). The PCR product was cloned in TOPO TA (Invitrogen™) amplification vector. The resulting *E. coli* colonies grown in LB plate with Ampicilin were used for mini-prep purification of plasmid DNA (QUIAGEN). The TOPO TA positive plasmids were Sanger-sequenced for insert verification. Since the patient was heterozygote both WT and 2308+3G>C fragments were obtained in the same cloning.

3.2.6.2. Minigene Plasmid Construction, Expression, and Transcript Analysis

The TOPO TA WT and mutated vectors and pMGgene vector were digested with the KpnI restriction enzyme (Gaildrat et al. 2010; Amato et al. 2012). The open pMGene vector was then dephosphorylated. The digested inserts were then cloned into the pMGene vector using the LigaFast Rapid DNA Ligation System (Promega). Top10 *E. coli* competent cells were transformed and grown in selective LB plate with Kanamicin. All clones were sequenced, and the WT and mutated forms were selected for expression experiments.

HEK-293 cells were transiently transfected with FuGene®HD (Promega) according to the manufacturer's instructions, using 2µg of WT or mutated pMGene and 3µl of FuGene®HD transfection reagent in 100µl absolute DMEM medium without FBS. After six hours the medium of cells was substituted with complete DMEM. Forty-eight hours after transfection cells were collected and RNA was extracted by Trizol Reagent (Life Technologies). RNA retrotranscription was performed by SuperScript VILO (Life Technologies) starting from 1µg of total RNA and using random primers. The cDNA obtained was amplified with GFP Fw and β-globin Rv primers. Since the forward primer annealed in the GFP-coding region, only located in the transcripts derived from the plasmid construct, the PCR did not select endogenous β-globin or MYBPC3 transcripts. The amplification products were analysed by Sanger sequencing.

3.2.7 Biophysical characterization of proteins

3.2.7.1 Protein expression and purification

The C4 domain sequence used to generate recombinant proteins and heteropolyproteins, corresponds to entry Q14896 (MYBPC3_HUMAN) of the UniProt databases.

The cDNA fragment including exons 18 and exon 19 of MYBPC3, was amplified from cardiac RNA with cMyBPC C4 Fw and cMyBPC C4 Rv oligonucleotides (see primers section). The primers were designed to clone just the coding sequence and to introduce BamHI, KpnI and BglII restriction sites. The PCR fragment was ligated into pT7 vector after digestion with BamHI and KpnI. Polymerase chain reaction-based site-directed mutagenesis was used with cMyBPC C4 I603M Fw internal primer to introduce the mutation I603M.

Production of monomeric proteins for thermodynamic analysis. The pT7-C4 WT and pT7-C4 I603M were digested using BamHI and BglII and the fragments were cloned into modified pQE80L expression vector. Expression plasmids were verified by Sanger sequencing. The resulting constructs, denoted as His_6-C4 WT and His_6-C4 I603M, encode the WT C4 domain and C4 I603M of cMyBP-C with an N- terminal hexahistidine affinity tag. Full protein sequences are reported in Table 3.

Table 3: amino acid sequences of C4 monomers

Protein	Sequence	# aa	Extinction coefficients
cMyBPC C4 WT	MRGSHHHHHHGSKLEVYQSIADLMVG AKDQAVFKCEVSDENVRGVWLKNGK ELVPDSRIKVSHIGRVHKLTIDDVTPAD EADYSFVPEGFACNLSAKLHFMERS	104	0.736 g/L
cMyBPC C4 I603M	MRGSHHHHHHGSKLEVYQSIADLMVG AKDQAVFKCEVSDENVRGVWLKNGK ELVPDSRIKVSHIGRVHKLTMDDVTPA DEADYSFVPEGFACNLSAKLHFMERS	104	0.724 g/L

Proteins were expressed in BLR (DE3) *E. coli* strain. Fresh cultures were grown at 37 °C to an OD_{600} of 0.6-1.0 and induced with 1mM IPTG. After 3 hours of induction at 37ºC and 250 rpm, C4 WT cultures were harvested. In the case of the C4 I603M, optimal expression was obtained by induction overnight at 14°C and 250 rpm. Cells were lysed by a combination of tip sonication and passes through a French Press. Proteins were purified from the soluble fraction using Ni-NTA agarose chromatography (Qiagen) with a column volume of 3 mL and pre-equilibrated with phosphate buffer (50 mM sodium phosphate pH 7 and 300 mM NaCl) supplemented with 10 mM DTT. Elution was performed in two steps with increasing imidazole concentration (from 20mM to 250mM). Further purification was achieved by size-exclusion chromatography in an AKTA Pure 25L system using a Superdex 200 Increase 10/300 GL column (GE Healthcare). The proteins were eluted in phosphate buffer, which is also the buffer used in circular dichroism (CD) and differential scanning calorimetry (DSC) experiments. The eluent was monitored by ultraviolet absorbance at 280nm. Concentrations of purified proteins were determined using theoretical extinction coefficients as estimated by the ProtParam tool (Table 3). Purity of samples was evaluated using SDS-PAGE.

Production of polyproteins for single-molecule experiments. To analyse the mechanical stability of C4 domain, heteropolyproteins were generated with two different markers, Protein-L and Sumo1 (Sadler et al. 2009; Kotamarthi et al. 2013)(Figure 13). Starting from pT7-C4 WT and pT-C4 I603M, the heteropolyproteins were generated with directional DNA concatemerization by self-ligation of the sticky ends of BamHI and BglI restriction sites (Alegre-Cebollada et al. 2010). Final cDNAs were inserted in the pQE80L expression plasmid using BamHI and KpnI and the resulting plasmids were verified by Sanger sequencing.

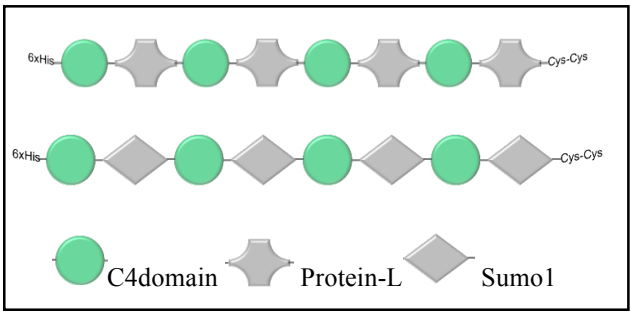

Figure 13: schematic representations of $(C4\text{-}L)_4$ and $(C4\text{-}Sumo1)_4$ heteropolyproteins.

The heteropolyproteins were expressed in BLR (DE3) *E. coli* cells. Fresh cultures (OD_{600} = 0.6-1.0) were induced with 1mM IPTG for 3 hours at 37ºC and at 250 rpm. Purification was achieved following the same procedure as for monomers (see above). The buffer used for the size-exclusion chromatography was 10 mM Hepes, pH 7.2, 150 mM NaCl, 1 mM EDTA, which is the same one used in AFM experiments. The eluent was monitored by the ultraviolet absorbance at 280nm. Then, the proteins fractions concentrations were determined using predicted extinction coefficients (ProtParam) and their purity was evaluated using SDS-PAGE gels.

3.2.7.2 Circular Dichroism (CD)

CD is defined as the unequal absorption of left-handed and right-handed circularly polarized light. The difference in absorbance of clockwise (E_R) and counterclockwise (E_L) polarized light is reported in units of ΔE, or in degrees ellipticity, defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse (Greenfield 2006). In particular, the CD data are reported as molar ellipticity $[\theta]$ in units of $deg \cdot cm^2 \cdot dmol^{-1}$.

CD spectra of cMyBPC domain C4 were collected for both the WT and I603M C4 samples with a Jasco-810 spectropolarimeter. The temperature was controlled using a Peltier thermoelectric system.

Purified proteins in phosphate buffer (20 mM sodium phosphate pH 6.5 and 50 mM NaCl) were loaded into a 0.1cm or 1cm path-length quartz cuvettes for data collection from far-UV (195-250nm) and near-UV (250-350nm) respectively. The proteins were analysed at concentration of 0.3 and 1 mg/mL for far-UV and near-UV respectively. Spectra were collected for the protein samples and buffer control at 25°C and 85°C with four accumulations of data. The parameters set for CD measurements are summarised in Table 4. The

buffer baseline spectrum was subtracted from each protein spectrum at each corresponding temperature to correct for the background signal.

Table 4: parameters used to measure the spectrum collection from far-UV (195-250nm) and near-UV (250-350nm)

Spectrum Measurement	
Start wavelength	250/350nm
End wavelength	195/250nm
Data pitch	0.2nm
Scanning mode	continuous
Scanning speed	50nm/min
response	1sec
Band width	1
Accumulation	4

Thermal denaturation analyses were carried out collecting the variations of ellipticity at 205nm and 230nm as a function of temperature (25°C-85°C). The thermal renaturation was also analysed with a temperature ramp-down (85°C-25°C). The spectrum measurement parameters are summarised in Table 5.

Table 5: parameters of thermal denaturation analysed at 215nm or 230nm

Variable Temperature measurement	
Wavelength	215/230nm
Start temperature	25ºC
End temperature	85ºC
Data pitch	0.5 ºC
Delay Time	0
Temperature Slope	30°C/h

Protein unfolding was plotted versus temperature and the data points were fit with a Boltzmann sigmoidal curve using Igor Pro software to obtain midpoint unfolding temperatures (T_m) for WT and I603M C4 domain. The changes in CD as a function of temperature, at characteristic wavelengths, can be used to determine the van't Hoff enthalpy (ΔH_v) of unfolding (Greenfield 2007).

To fit the change of CD at a single wavelength as a function of temperature, the Gibbs-Helmholtz equation was used considering a two-state process:

$$\Delta G = \Delta H\,(1-T/T_m) - \Delta C_p((T_m-T)+T\ln(T/T_m))$$

considering that

$$K=\exp(-\Delta G/(RT))$$

and

$$\alpha=K/(1+K)$$

$$\theta_t=\alpha(\theta_F-\theta_U)+\theta_U$$

Where α is the fraction unfolded protein, K is unfolding constant, T is the absolute temperature (Kelvin) and R is the Gas constant (1.98 cal/K/mol). ΔCp is the change in heat capacity going from the folded to the unfolded state. We considered ΔCp = 0.

3.2.7.3 Differential scanning calorimetry (DSC)

DSC is a powerful technique to characterize temperature-induced conformational changes in proteins and other biological macromolecules (Cooper and Johnson 1994). Calorimetric measurements were performed using a Microcal VP-DSC differential scanning calorimeter with 0.5 mL cells. Experiments were done increasing temperature from 25°C to 85°C at a rate of 0.5°C/min, using 0.085 mM (1 mg/mL) protein concentration in 20 mM sodium phosphate pH 6.5, 50 mM NaCl. The reversibility of the thermal transitions was assessed by reheating of the sample immediately after cooling from the previous scan, using the same rate of temperature change. The calorimetric traces were corrected for the instrumental background by subtracting a scan with buffer in both cells. The temperature dependence of the excess heat capacity was further analysed and plotted using Origin software (MicroCal, Northampton, MA). The thermal stability of the proteins was described by its T_m, and the calorimetric enthalpy (ΔH_{cal}), which was calculated as the area under the excess heat capacity function. The van't Hoff enthalpy (ΔH_v) associated with DSC thermograms was determined by a two-state fit of the thermograms.

$$C_p = C_{pN} + \left[\frac{K_A\,\Delta C_{pA}}{1+K_A} + \frac{K_A\,\Delta H_v^2}{(1+K_A)^2 RT^2}\right]$$

where C_{pN} is the molar heat capacity of the totally folded state and ΔC_{pA} is the change in heat capacity for unfolding the A domain (DSC Data Analysis in Origin®-Tutorial Guide).

The Gibbs free energy change (ΔG) was then calculated at any temperature using three experimental parameters: T_m, enthalpy change at T_M ($\Delta H(T_m)$) and

heat capacity change at T_m (ΔC_p). Using this parameters, the ΔG was derived from the above equation:

$$\frac{\Delta G}{T} = \Delta H(T_M)\left(\frac{1}{T} - \frac{1}{T_M}\right) + \Delta C_p\left(\ln\frac{T_M}{T} - \frac{T_M}{T} + 1\right)$$

The derivation assumes a reversible two-state transition and also assumes ΔC_p is independent of temperature (DSC Data Analysis in Origin®-Tutorial Guide). The ΔG was derivate for C4 WT and C4 I603M considering $\Delta C_p = 0$.

3.2.7.4 Single-molecule Atomic Force Microscopy (AFM)

Single-molecule AFM measurements were obtained in a commercial force-clamp AFS (Luigs & Neumann) following established protocols (Popa et al. 2013). In AFM experiments, we were able to stretch single polyprotein molecules tethered between a gold-coated surface and the tip of an AFM cantilever probe.

Forces between the tip and the sample lead to a deflection of the cantilever according to Hooke's law:

$$F(pN) = \Delta x(nm) k_{sc}\left(\frac{pN}{nm}\right)$$

Where Δx is the displacement of cantilever and k_{sc} the spring constant.

The cantilever spring constant is described as follows:

$$k_{sc} = \frac{k_b T}{S^2\langle x^2\rangle}$$

Where k_b is Boltzmann's constant, T is the absolute temperature (K), S is the contact slope or deflection sensitivity (nm/V) and $\langle x^2\rangle$ is the mean squared displacement of cantilever (nm).

The cantilevers were calibrated before measurement according to the thermal fluctuations method in which the mean squared displacement of cantilever $\langle x^2\rangle$ was determined when the cantilever is positioned far from the gold-coated surface (Hutter and Bechhoefer 1993).

During an AFM experiment, Δx is calculated as changes in the A-B signal of

photodiode. Then, the signal in volts (V) coming from the photodiode, is converted in length units (nm) the using following equation:

$$x(nm) = S\left(\frac{nm}{V}\right)(A - B)(V)$$

Silicon nitride MLCT-C cantilevers were used with a 60-nm gold back side coating (Bruker). Typical spring constant values ranged between 15 and 20 pN/nm.

In the classical force-extension mode (FX) the traces are recorded at constant velocity of piezo retraction. In this mode, the cantilever measures the force response of the protein as its end-to-end length changes. Alternatively, in force-clamp mode the position of the piezo is adjusted such as the cantilever deflection is maintained at a given set-point force by a proportional–integral–derivative (PID) controller which adjusts the position of the piezo (Schlierf et al. 2004).

A small aliquot (2µL) of the purified protein is deposited directly into the buffer in the fluid chamber. The cantilever is brought in contact to the surface for 1s at 500-1500 pN to favor formation of single-molecule tethers. Then, the surface is retracted to achieve the set point force. If a single-molecule tether is formed, the force is increased linearly at 10 pN/s for 20 s while the length of the polyprotein is measured. Under this condition, protein unfolding events are detected as step increases in the length of the heteropolyproteins. Under force clamp, the length of protein-L unfolding steps is 16 ± 1 nm, whereas Sumo1 unfoldings produce 20 ± 1 nm unfolding steps (Pimenta-Lopes et al. 2018). To fingerprint the C4 domain, all traces that contain at least two events of the same size were analysed and represented as histograms.

In the data analysis, traces in FX mode were fitted to the worm-like chain (WLC) model of polymer elasticity (Bustamante et al. 1994):

$$F(x) = \frac{k_B T}{p}\left[\frac{1}{4}\left(1 - \frac{x}{L_c}\right)^{-2} - \frac{1}{4} + \frac{x}{L_c}\right]$$

where p and L_c are the persistence length and contour length, respectively, k_B is Boltzmann's constant and T is absolute temperature.

Data acquisition and analysis was performed using custom software written in Igor Pro (Wavemetrics).

To estimate the initial extension of the polyproteins only reproducible traces with four protein L unfolding events were analysed. The force-extension traces were fitted with the WLC model to calculate the contour lenght up to the first unfolding peak. The expected initial extension was calculated considering C4 was unfolded or folded. In the first case the expected initial extension results by the sum of the length of four folded L modules (4 x 3 nm), the handles and linkers included in polyproteins (28 amino acids x 0.4 nm/amino acid) and the length of the extended C4 domain (4 x 90 amino acids x 0.4 nm/amino acid) for a total length of 167.2 nm. In contrast, if C4 remained folded, the initial extension would correspond to the extension of the folded L domains, the handles and linkers and the folded C4 (4 x 4.1nm) for a total length of 39.6 nm. The length of each folded domain (L proteins and C4 domains) was obtained measuring the distance between the α-carbons of the first amino acid and the last amino acid in the high resolution structures of the domains.

The AFM tip picks the protein at random locations, so the number of peaks observed depend on the number of modules contained in the segment that is stretched by the AFM. Hence, the full extension of the heteropolyprotein has to be evaluated from traces with all four unfolding events of protein L. However, if the cantilever picked up the second domain of protein corresponding to L, the unfolding traces would not include one domain C4 and flanking linkers, a scenario that was taken into account to calculate a range of predicted initial extensions. Specifically, if the first C4 domain is not picked up by the cantilever, the initial extension would be 29.9 nm for C4 folded and 125.6 nm for C4 unfolded.

RESULTS

4. RESULTS

4.1 The spectrum of sarcomeric mutations in HCM patients

A large genetic screening of about 249 patients affected by hypertrophic cardiomyopathy was performed. No mutation was found in 38% of the patients, while the rest carried double o single mutations in genes coding for sarcomeric proteins. Most of them are localised in MYH7 (33%) and MYBPC3 (41%), in agreement with literature data (*Figure 14*) (Burke et al. 2016; Liew et al. 2017).

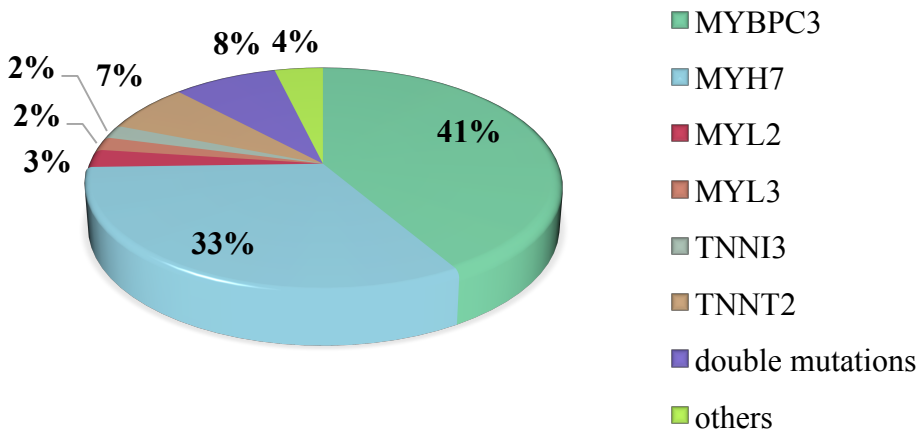

Figure 14: distribution of mutations in genes coding for sarcomeric proteins identified in genetic screening of HCM patients.

The focus of this thesis was MYBPC3, the most frequently mutated gene in HCM. The principal mutations of MYBPC3 related to Hypertrophic Cardiomyopathy include missense and truncating mutations (Marston et al. 2012; Mohamed et al. 2017). Among those patients harboring mutations in MYBPC in our cohort, 57% of them have truncating mutations compared to 43% of missense mutations. Protein truncation can result from frameshifts, premature termination codons (PTC) or mutations causing splicing alterations. In our HCM population, these possibilities occur with different frequencies: 25% are frameshifts, 7% are non-sense mutation, and 25% of intronic mutation that cause splicing alterations (Table 6). Most of the mutations found had already being described as pathogenic. The following mutations have not been reported before: p.T33RfsX15, p.R502Rfs, p.G897AfsX159, IVS4-2A>C and

IVS23+3G>C. Most of the identified missense mutations had been also already associated with HCM, and some of them were functionally characterised. However, mutations p.I603M and p.T705K had not been reported before (Table 7).

Table 6: MYBPC3 Truncated mutations identified in genetic screening

TRUNCATED MUTATIONS		
splicing	**non sense**	**frameshift**
IVS4-2 A>C	p.Q1012X	p.T33RfsX15
IVS9-1 G>C	p.Q1259X (3)	p.F305Pfs
IVS11-9 G>A (4)	p.W1214X	p.R502Rfs
IVS23-2 A>G (6)		p.M632Rfs (3)
IVS23+3 G>C		p.G897AfsX159
IVS32+2 T>A (6)		p.G1206Vfs
		p.K1065QfsX12 (11)

(Number in parenthesis are the number of patients with the corresponded mutation)

Table 7: MYBPC3 missense mutations identified in genetic screening

MISSENSE MUTATIONS	
p.P102L	p.G507R
p.G148R (2)	p.R574W
p.V178M	p.R597Q
p.V189I	p.R597W
p.S217G (3)	p.I603M
p.E258K (2)	p.E619K (2)
p.P371R (4)	p.T705K
p.E441K	p.R810H (2)
p.R495G	p.R820Q
p.R495Q	p.A848E
p.R502W	p.V906G
p.R502Q	p.Y1172C

(Numbers in parenthesis are the number of patients with the corresponded mutation)

In figure 15, the distribution of mutations along the different domains of cMyBPC protein is shown. The sixteen different truncating mutations were spread throughout all domains of protein. The terminal domains, which are interactors of actin, myosin and titin (Harris et al. 2011), were slightly enriched in truncating mutations. In contrast, the majority of missense mutations affect the central region of cMyBPC between domains C3 and C7. Indeed, among

twenty-five different missense mutations, ten map to the C3 and C4 domains, with no sarcomeric known interactor.

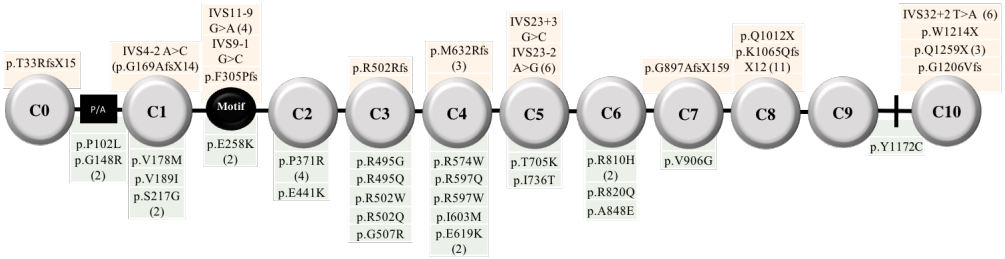

Figure 15: mutations of cMyBPC found in a genetic screen of 249 HCM patients. Missense (*bottom*) and truncating (*top*) mutations in cMyBPC are shown according to the affected domain. The number of patients with the corresponding mutation is indicated between parentheses.

4.2 Functional characterization of variants affecting splicing

The genetic screening highlighted several frameshift, and PTC variants in the MYBPC3 gene. The new variants p.T33RfsX15, p.R502Rfs, p.G897AfsX159 result in cMyBPC protein truncation. The premature truncation of the polypeptide probably results in lower protein levels, leading to cMyBPC haploinsufficiency and development of HCM in these variants (Van Dijk et al. 2009; Harris et al. 2011; Marston et al. 2012).

The pathogenicity of intronic variants is much more difficult to deduce without functional characterization. Therefore, a detailed functional study of new non-coding variants, IVS4-2 A>C (c.506-2 A>C) and IVS23+3G>C (c.2308+3 G>C), was developed.

The intronic variants were identified in two independent patients affected by HCM. Both variants were found in heterozygous pattern without mutations in other genes screened in these patients.

During this work, the *MYBPC3* c.506-2 A>C variant has been reported as disease-associated but no functional validation was provided. The variant co-segregates with disease, as showed in the family pedigree (*figure 16*). Although no HCM phenotype was clinically evident in two subjects who carry the mutation, all affected patient are positive in genetic screening for c.506-2 A>C. The MYBPC3 c.2308+3 G>C, is a new intronic variant not previously associated to HCM phenotype. Although no family screening was requested and the patient analysed presented borderline clinical evidence of the disease, the mutation was however selected for functional characterization.

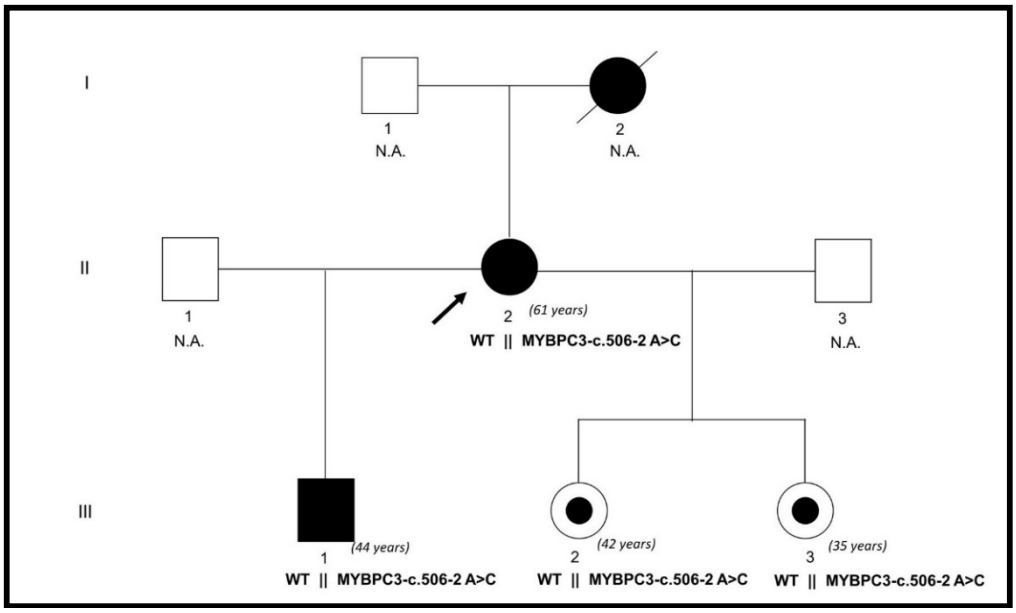

Figure 16: pedigree of HCM family with the c.506-2 A>C variant. *Open symbols* represent subjects with a negative phenotype. *Black symbols* represent clinically affected subjects. *Circles with solid centers* indicate unaffected female mutation carriers. The diagonal line indicates a deceased family member. The arrows indicate the proband. The ages of subjects are reported in brackets. N.A.: not analysed; WT: wild type

Both variants were located in a consensus region known as acceptor or donor splice sites. In particular, c.506-2A>C targets a consensus sequence of intron 4 acceptor splice site of the *MYBPC3* gene. c.2308+3 G>C involves the donor splice site of intron 23. Hence, the location of both mutations suggest that these variants may alter correct RNA splicing.

4.2.1 In silico studies to predict splicing changes

The pathomechanism of intronic mutations is, in most cases, due to defects in splicing process (Salman et al. 2018). The splicing is the editing of pre-mRNA transcript into a mature messenger RNA (mRNA) in which introns are removed and exons are ligated (see section 1.4 in the introduction). The mechanism involves the recognition of donor and acceptor splicing sites. Mutations in these specific sequences block canonical splicing, resulting in aberrant mRNAs (Baralle and Baralle 2005; Buratti et al. 2007).

In order to verify whether the two mutations result in aberrant alternative splicing of MYBPC3*, in silico* analysis by five different algorithms were performed using Alamut® Visual software.

The analyses of MYBPC3-c.506-2A>C mutation showed a severe impact on splicing process with completely loss of the natural acceptor splice site predicted by all five algorithms with a maximum score. Although with a low probability, three algorithms (SSF, MaxEnt and GeneSplicer) also identified a cryptic splice site at position c.513 of MYBPC3 exon 5, i.e. 7 nucleotide upstream of the native splicing site. The final result of Alamut elaboration is the high probability of exon 5 skipping (*figure 17A*).

For *MYBPC3*-c.2308+3G>C mutation, MaxEnt and NN Splice algorithms of the Alamut software predicted a consistent alteration of the natural donor site. The algorithms showed an close to 50% of reduction and a total site loss, respectively. The other three algorithms also predicted a donor site alteration, but the percentage of variation induced by the mutation was not as high (SSF $\geq -7.1\%$, GeneSplicer $\geq -32.8\%$; HSF $\geq -8\%$). No indications of formation of a new splicing site were predicted for this mutation (*figure 17B*).

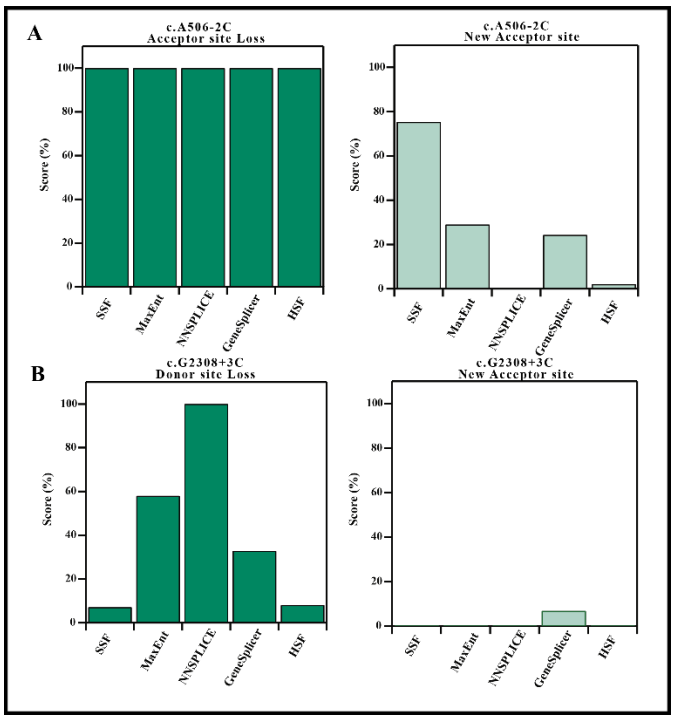

Figure 17: prediction of splicing alterations using Alamut software. The histograms represent the score given by the different prediction algorithms for a splicing site loss (*left*) and for the appearance of a new acceptor site (*right*). (**A**) Predictions for the c.506-2 A>C variant. (**B**) Prediction for c.2308+3 G>C variant. The evaluations were carried out with five algorithms of splicing prediction: SSF (SpliceSiteFinder-like), MaxEnt, NNSPLICE (Neural Network Splice), GeneSplicer and HSF (Human Splicing Finder).

4.2.2 Experimental validation of splicing pathomechanism

RNA splicing analyses for these variants were preferentially carried out using peripheral blood. In the absence of samples, splicing analyses using the minigene assay were performed.

The total RNA of the HCM patient carrying mutation c.506-2A>C was extracted from leukocyte fractions of peripheral blood and was retrotranscribed by RT-PCR using random primers. The region spanning from exon 3 to exon 6 of MYBPC3 was amplified by specific primers and the sequence was analysed. A healthy subject without MYBPC3 mutations was also analysed as a control. The RT-PCR amplified a band of the expected size (334bp) in both individuals. The sequence corresponding to the PCR product of the HCM patient showed both a normal and an aberrant splicing of mRNA accordingly with the heterozygous pattern. The normal mRNA carried the sequence of exon 4 followed by normal exon 5. Instead, the aberrant form showed a truncated exon 5 with a seven nucleotide loss (*figure 18*). This RNA alteration results in the substitution of glycine 169 with an alanine and a subsequent frame-shift and generation of a premature stop codon (MYBPC3-G169AfsX14).

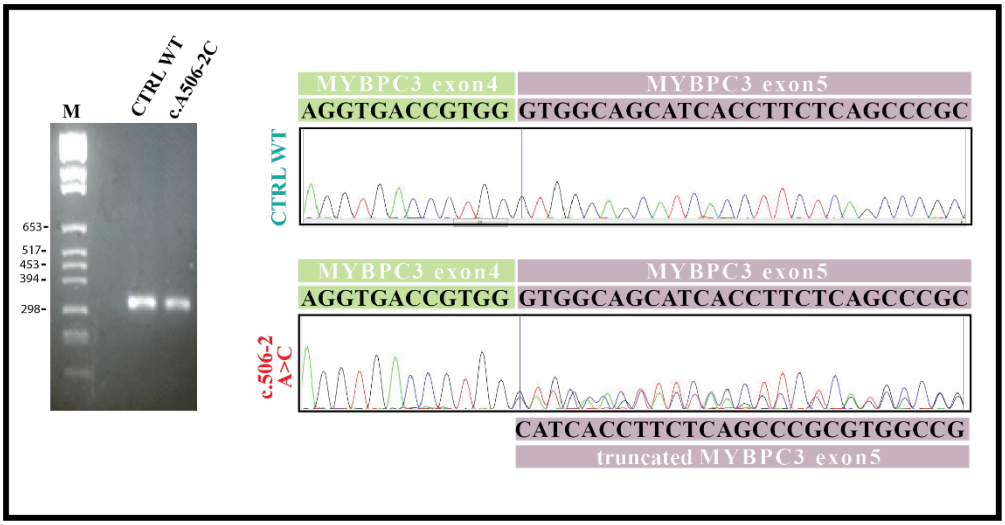

Figure 18: Effect of the c.506-2A>C mutation on the splicing of exon 5. (*Left*) Electrophoresis of the exon3-exon6 RT-PCR fragment of MYBPC3 from mRNA of healthy subject (WT) and patient. Lane 1: Marker IX (Roche Diagnostics); Lane 2: Control WT; Lane 3: c.506-2A>C mutation. (*Right*) Electropherograms obtained from Sanger sequencing of the fragments; black, red, green and blue peaks represent G, T, A, and C. The electropherograms capture the end of exon 4 (highlighted in green) and the start sequence of exon 5 of MYBPC3 cDNA (highlighted in pink), or a truncated mRNA in the case of the HCM patient carrying the mutation c.506-2A>C.

Since no cardiac biopsies or fresh peripheral blood sample of HCM patients carrying mutation c.2308+3G>C were available, a minigene system was used to verify the results obtained with the Alamut predictor software. A splicing minigene assay is based on an expression plasmid (pMG) containing the human β-globin gene (HBB), from the starting codon to the stop codon (including introns), and the GFP coding sequence fused with exon 1 of HBB. The genomic region containing the mutation (including the whole exon and the 5' and 3' intronic flanking regions) is inserted in pMG after amplification from genomic DNA. In particular, the insert for the analyses of c.2308+3G>C contained exon 23. The pMG-MYBPC3-WT and pMG-MYBPC3-2308+3G>C plasmids were transfected in HEK293 cells, the total RNA was extracted, retrotranscribed and sequenced. The analysis of the RNA obtained from the minigene system for c.2308+3G>C showed generation of two transcripts (normal and alternative) in pMG-MYBPC3-WT and only one (alternative) in pMG-MYBPC3-2308+3G>C (*figure 19*). Sanger sequencing shows that the alternative transcript arises from the skipping of MYBPC3's exon 23. It is unclear whether this non-canonical splicing site also occurs in wild-type human subjects. However, results in *figure 19* show that the acceptor site leading to inclusion of exon 23 in the mRNA of cMyPBC is weakened by the mutation 2308+3G>C, pointing towards a pathogenic status of the mutation. If exon 23 is skipped, the leucine 717 would be substituted by a threonine, causing a generation of a premature stop codon after 51 codons (MYBPC3-L717TfsX51).

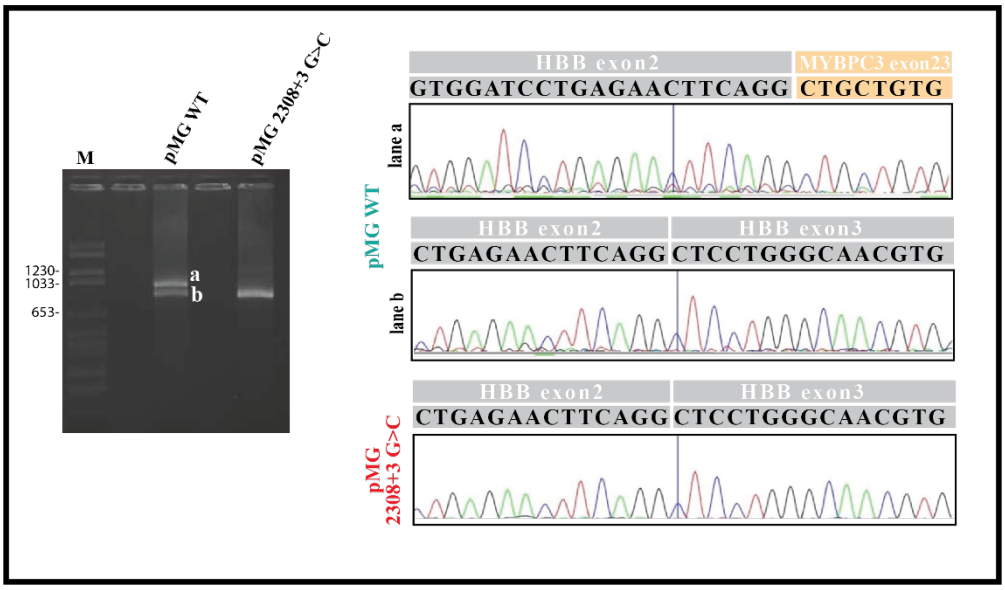

Figure 19: Effect of the MYBPC3-c.2308+3G>C mutation on the splicing of exon 23 of the MYBPC3 gene. (*left*) Electrophoresis of RT-PCR analyses of mRNA extracted from HEK293 cells transfected with pMG plasmids. Lane 1: Marker VI (Roche

Diagnostics); Lane 2: pMG WT; Lane 3: pMG 2308+3 G>C. (*right*) Electropherograms obtained from cDNA sequencing of the fragments; black, red, green and blue peaks represent the G, T, A, and C. Band a and band b sequences corresponded to the respectively band of pMG WT electrophoresis shown on the left. The electropherograms indicates the end of exon 2 and the start of exon 3 of HBB (highlighted in grey) and the start sequence of exon 23 of MYBPC3 cDNA (highlighted in yellow).

4.3 Investigations on pathogenic mechanism of missense mutation

The large genetic screening showed 43% mutations in MYBPC3 gene to be missense. As opposed to truncating mutations, the HCM-causing missense variants seem to localise more to the central domains of protein. Among them, a new missense variant, I603M, was found. The mutation is in the central domain C4.

I603M was identified in a proband with clinical features of HCM and family history. The family screening evidenced the presence of mutation in affected subjects in heterozygosis together with another MYBPC3 mutation, T33RfsX15. Only one patient, corresponding to III.5, is clinical affected but did not show the I603M mutation. Furthermore, two young subjects result positive to genetic screening without clinical evidences for HCM (*figure 20*).

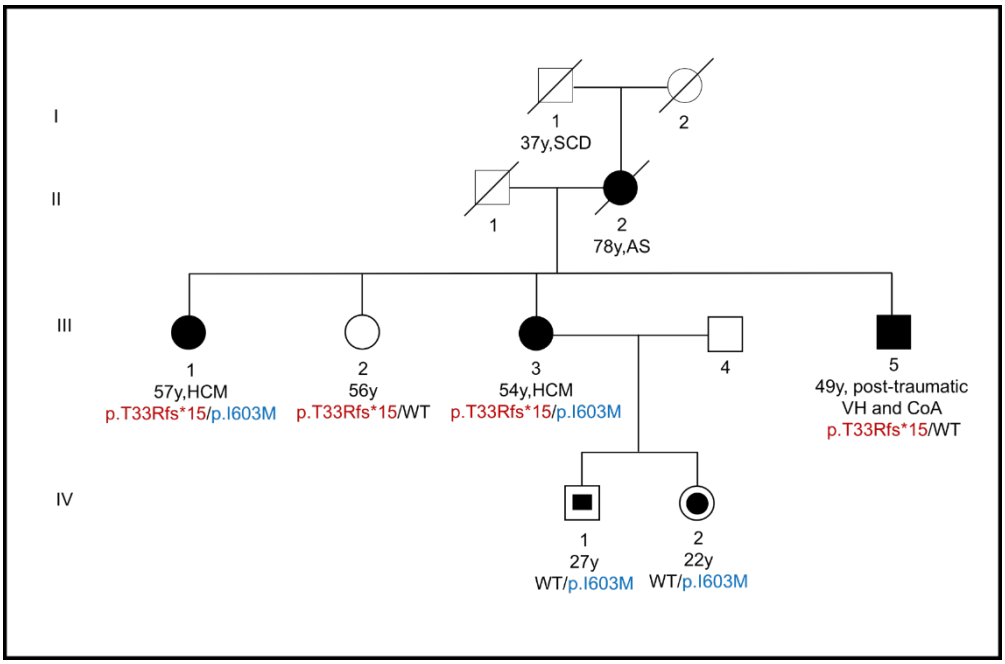

Figure 20: pedigree of four generations of HCM family with p.I603M variant. *Open symbols* represent subjects with a negative phenotype. *Black symbols* represent

clinically affected subjects. *Circles or square with solid centres* indicate unaffected mutation carriers. The symptoms of each patient are indicated: SCD, Sudden Cardiac Death; AS, Aortic Stenosis; HCM, Hypertrophic cardiomyopathy; VH, Ventricular Hypertrophy; CoA, coarctation of the Aorta. y:years.

Preliminary evaluation using three different bioinformatics tools showed strong evidence for deleterious effect of I603M variant (Table 8). Indeed, in comparison with another benign variant localised in the same domain, I603M showed a high score of pathogenicity.

Table 8: bioinformatics prediction of deleterious effect of p.I603M

cMyBPC variation	PolyPhen-2 Prediction	SIFT Prediction	PROVEAN Prediction
	(>0.6)	(<0.05)	(<-2.5)
p.I603M (c.1809 T>G)	1.0	0.001	-2.63
p.T607L (c.1820 C>T)	0.010	0.162	-1.71

Given the considerable evidence of deleterious effect, the variant I603M was selected for functional study to determine pathogenicity.

4.3.1 In silico and in vitro study of splicing alterations induced by missense mutations

Defects in RNA splicing is a mechanism by which missense mutations can result into truncated polypeptides, leading to protein haploinsufficiency that causes HCM. During the editing of pre-mRNA transcripts, the splicing machinery recognizes consensus sequences that also include the exons, particularly the regions close to the intron-exon boundaries (Yue et al. 2005; Lee and Rio 2015; Smelter et al. 2018). Several missense mutations of cMyBPC, whose pathogenicity status is well established, are known to cause RNA splicing disruptions (Carrier et al. 1997; Helms et al. 2014). We used available information to validate our *in silico* and *in vitro* analyses of RNA splicing alterations by mutations in exonic regions (missense mutations).

The disease-causing missense mutation E542Q (c.1624G>C), located at central domain C3 of cMyBPC, has been shown to cause alterations of splicing in human myocardium (Marston et al. 2012). The Alamut analyses of this mutation predicted the loss of a splice site, with maximum probability score determined by two algorithms (SSF and NNSplice) (*Figure 21*). We examined

the alteration of mRNA for the mutated transcript experimentally on blood samples. The region between exon 15 and exon 21 of cMyBP-C's mRNA was amplified using the leukocitary fraction of peripheral blood as starting material. In line with *in silico* prediction and the previously analyses on myocardial biopsies in literature (Carrier et al. 1997), the mutation E542Q induces the skipping of exon 17 (*Figures 21, 22*). This effect can be observed from the higher mobility band that is amplified (*Figure 22)*, and confirmed by the Sanger sequencing results, in which to exon-exon boundaries are detected, one corresponding to normal processing of mRNA (exon 17-exon 18, from wild type allele) and the other one to the skipping of exon 17 (*Figure 21*).

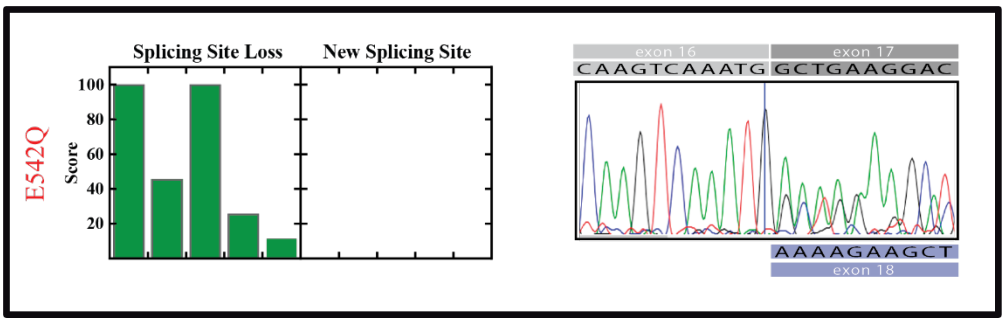

Figure 21: Splicing analyses of E542Q (c.1624G>C). (**on the left**) The histograms show the results of Alamut software. The canonic splice site loss has high score. No probability of new acceptor site creation was found. (**on the right**) Electropherogram obtained from cDNA sequencing of exon15-exon21 fragment, amplified from mRNA obtained from peripheral blood. The electropherogram indicates the end of exon 16 and the start of exon 17 (*top*) or exon 18 (*bottom*) of cMyBPC. Black, red, green and blue peaks represent the G, T, A, and C.

We also analysed splicing alterations in other disease-causing missense mutations affecting domain C3. Mutations R495Q (c.1484G>A), R502W (c.1504C>T) and R502Q (c.1505G>A) were predicted not to alter the canonical acceptor site. Only for R502Q, the Alamut software showed a potential new acceptor site (*Figure 22*). The effect of these mutations in the splicing of cMyBP-C's mRNA was analysed using mRNA from peripheral blood of carriers. As opposed to E542Q, mutations R495Q, R502W and R502Q gave rise to bands of the same size as the wild-type (*Figure 22*). Sanger sequencing confirmed no alteration on the transcripts (*Figure 22*). Importantly, our results using peripheral blood are coincident with results obtained usind myocardial samples from patients carrying mutations R495Q and R502W (Marston et al. 2012; Helms et al. 2014).

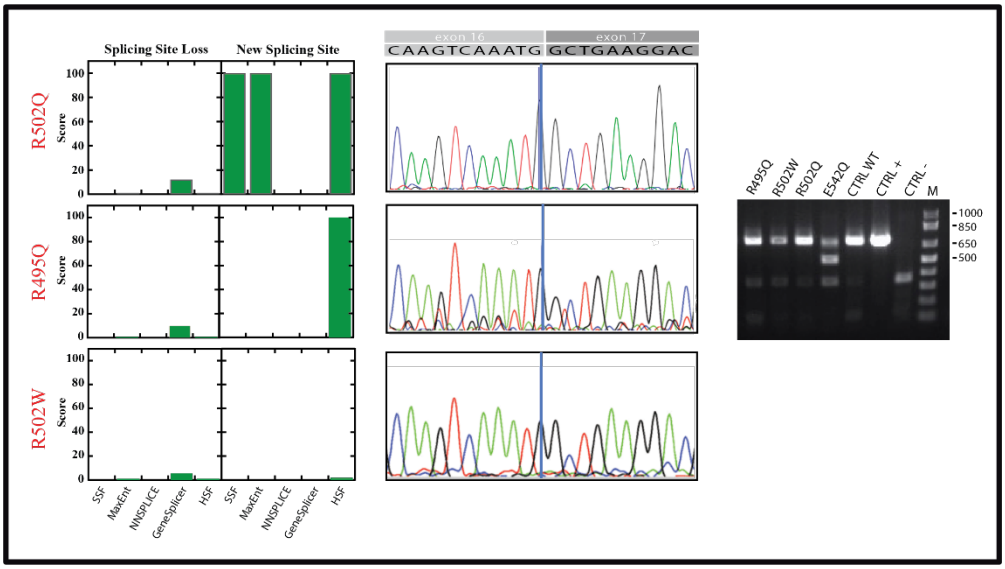

Figure 22: Splicing analyses of R502Q, R495Q and R502W. (**on the left**) The histograms represent the results of Alamut software. The evaluations were carried out with five algorithms of splicing prediction: SSF (SpliceSiteFinder-like), MaxEnt, NNSPLICE (Neural Network Splice), GeneSplicer and HSF (Human Splicing Finder). No splicing alterations were predicted. (**on the centre**) Electropherograms obtained from cDNA sequencing of exon15-exon21 fragment. The electropherograms indicates the end of exon 16 and the start of exon 17 of cMyBPC. Black, red, green and blue peaks represent the G, T, A, and C. (**on the left**) Electrophoresis of RT-PCR analyses of mRNA. M: 1Kb DNA ladder; CTRL WT: amplification product of mRNA extracted from healthy subject; CTRL+: amplification product of commercial cardiac mRNA; CTRL-: amplification product of mRNA extracted from Hela cells as negative control.

Overall, these analyses confirm that the pathological mechanism in missense mutations may occur as an mRNA defect rather than directly affecting the protein, such as in the case of E542Q. Defects in mRNA splicing for these mutations can be predicted with bioinformatics tools with high fidelity, and false positives can be avoided by analyzing readily available peripheral blood from patients.

In the quest to describe pathogenicity for mutation I603M (c.1809T>G), we applied the same analyses of RNA splicing alterations. The mutation occurs in the exon 19. The transcript fragment spanning from exon 15 to exon 21 was amplified from blood's mRNA. The electrophoresis gel showed an identical migration pattern of both mutant and WT fragments (*Figure 23*). The WT corresponded to a healthy subject. Sanger sequencing confirmed the correct splicing (*Figure 23*).

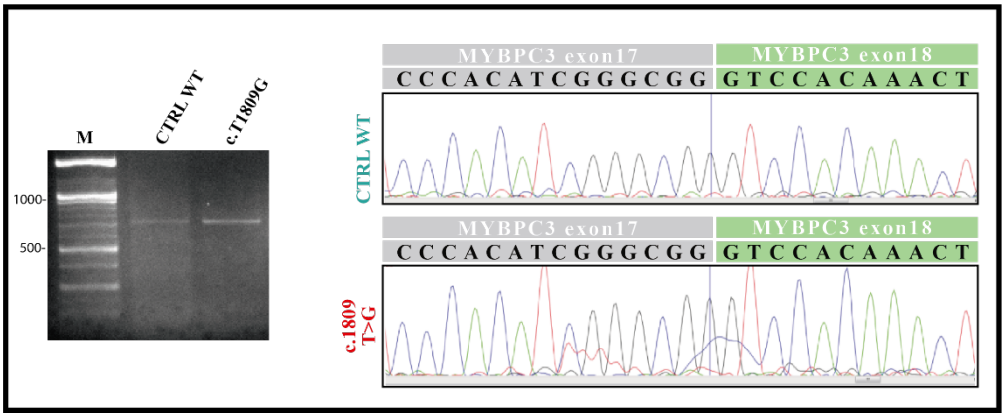

Figure 23: Effect of the MYBPC3-c.1809T>G (p.I603M) mutation on the splicing mechanism. (**on the left**) Electrophoresis of RT-PCR analyses of mRNA. Lane 1: Marker XIV (Roche Diagnostics); Lane 2: Control; Lane 3: c.1809T>G mutation. (**on the right**) Electropherograms obtained from cDNA sequencing of the fragments; black, red, green and blue peaks represent the G, T, A, and C. The electropherograms indicates the end of exon 17 (highlighted in grey) and the start sequence of exon 18 of MYBPC3 cDNA (highlighted in green).

4.3.2 Homology modelling of domain C4 structure

The position of I603M mutation is mapped on the structure of central domain C4 of cMyBPC. Since the structure of domain C4 is not known, to predict how the protein structure might be affected by I603M mutation, the homology modelling of the domain was carried out. The protein sequence related to C4 from the UniProt database was modelled using I-TASSER tool. The model showed that the C4 domain adopts an Ig-like folding (*Figure 24,A*) (Kufareva and Abagyan 2012). Instead, the prediction of mutant structure showed the same folding of WT (*Figure 24,B*). The RMSD of both predictions with respect to the template was 1.9±1.6Å for WT and 2.0±1.6Å for I603M.

In the mutation, an isoleucine is substituted by another non-polar amino acid, methionine, suggesting no effect on protein structure. Still, it is interesting to observe that Ile603 is buried in the core of the domain, so a mutation affecting this residue could not be easily accommodated by the domain leading to potential destabilization of the I603M mutant.

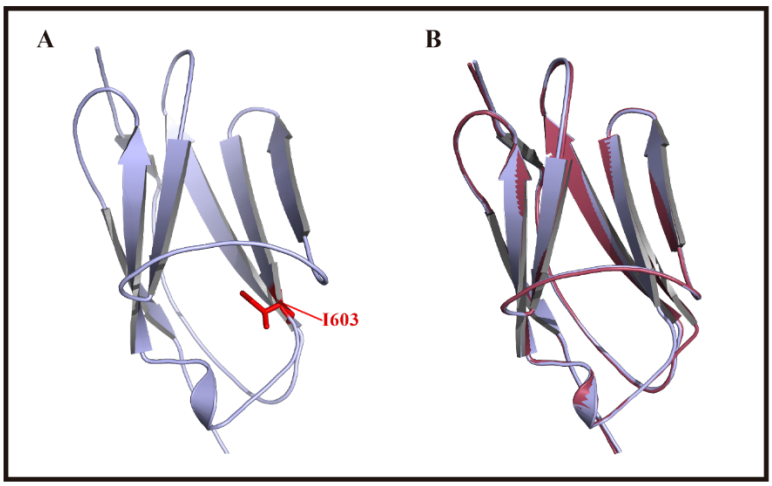

Figure 24: Homology modelling prediction of cMyBPC C4 domain. (**A**) Graphic representation of secondary structure of C4 WT domain, modelled with I-tasser using slow-MyBPC as a template (2YUZ PDB structure). The position of the isoleucine replaced in HCM patients is shown in red. (**B**) Overlapping of the secondary structure predictions of C4WT and C4 I603M. The mutant domain is indicating in magenta. The models of C4 domain have been prepared with PyMOL.

4.3.3 Structural and thermodynamic characterization of C4 and missense mutant by circular dichroism

Since RNA splicing seems not to be altered by mutation I603M, we investigate whether structural and biophysical properties of cMyBPC C4 domain may be affected in the mutant. To that end, the recombinant WT (aa.544-aa.633) as well as mutant I603M proteins were expressed and purified using *E.Coli* BLR21 (DE3) expression system. Site direct mutagenesis by PCR was used to generate the I603M mutant domain. All proteins were expressed in soluble form and purified by two steps: nickel-nitrilotriacetic acid (Ni-NTA) chromatography and size-exclusion chromatography using a fast protein liquid chromatography (FPLC) system (AKTA, GE Healthcare). All domains were at least 90–95% pure as judged by SDS-PAGE analysis (*Figure 25*).

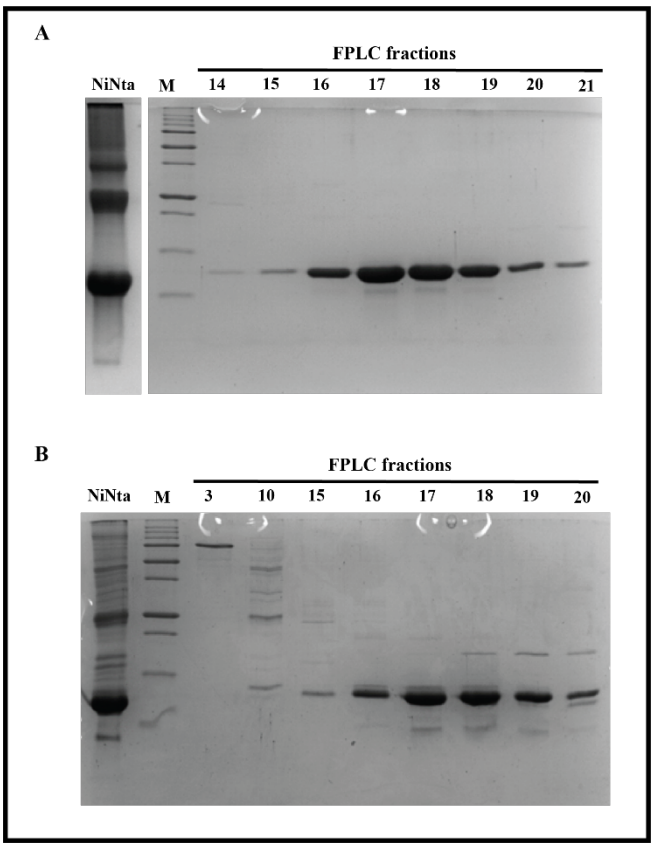

Figure 25: purification of recombinant C4 domain expressed in *E.Coli* BLR21 (DE3). Ni-Nta purification and gel filtration profile of the purified C4 WT (A) and C4 I603M (B). The nickel-nitrilotriacetic acid (Ni-NTA) pools were further purified with fast protein liquid chromatography (FPLC) using size-exclusion. All fractions were loaded in SDS-gels. The preparations were at least 90–95% pure.

Both proteins were analysed by Circular Dichroism (CD) to determine their secondary structure and folding properties. To investigate potential structural changes induced by the I603M mutation, far-UV (250-195nm) and near-UV (350-250nm) CD spectra were collected at 25°C for both C4 WT and C4 I603M (*Figure 26*).The spectra of C4 WT showed a minimum of ellipticity at 215 nm, characteristic of β-sheet structure (Greenfield 2006; Shen et al. 2018) in agreement with the homology modelling prediction. The comparison between C4 WT and C4 I603M spectra showed a similar native structure suggesting that the mutation did not have major impact in the fold of the protein (*Figure 26 C,D*). Change in the shape of the far-UV spectra was observed when the sample temperature was increased to 85°C, indicating the unfolding of protein and loss of secondary structure for both WT and I603M (*Figure 26 A, B*). The maximum differences between spectra at 25°C and 85°C were observed at 205 nm and at 230 nm. These wavelengths were chosen for

tracking the thermal denaturation of the proteins, to examine whether the mutation induces any thermodynamical destabilization.

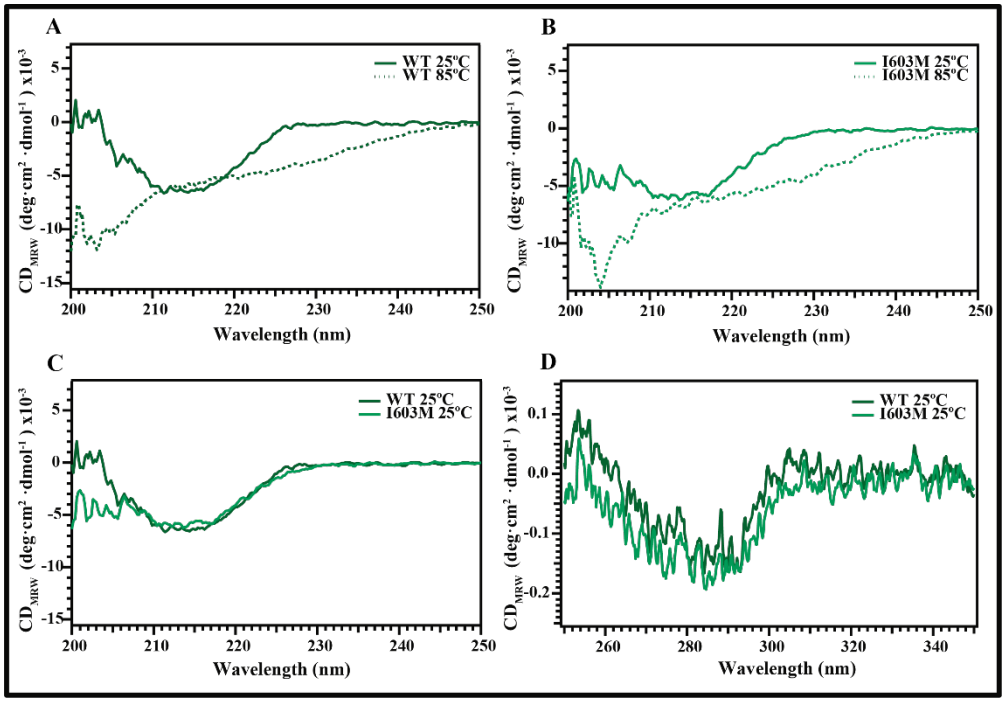

Figure 26: circular dichroism spectra of cMyBPC C4 WT and I603M. CD spectra of C4 WT (**A**) and C4 I603M (**B**) monitored in the far-UV at 25°C (solid line) and 85°C (dashed line). (**C**) Comparison of CD spectra in far-UV of native C4 WT and C4 I603M. (**D**) CD spectra of C4 WT and C4 I603M collected in near-UV.

Temperature-induced protein unfolding was investigated by monitoring the ellipticity as a function of temperature. The CD signals at 205nm and 230nm were tracked over a temperature range from 25°C to 85°C. Data of both WT and I603M mutant were fit to sigmoidal curves. The unfolding profiles measured of the two proteins at 205nm showed different midpoint unfolding temperatures, with a melting temperature (T_m) of 57°C for C4 WT and 42°C for mutant C4 (*Figure 27 A*). Similar T_m values were obtained from the unfolding transition monitored at 230 nm: 58°C for WT and 45°C for the mutant (*Figure 27 B*) These data show the destabilizing effect of mutation I603M on the C4 domain.

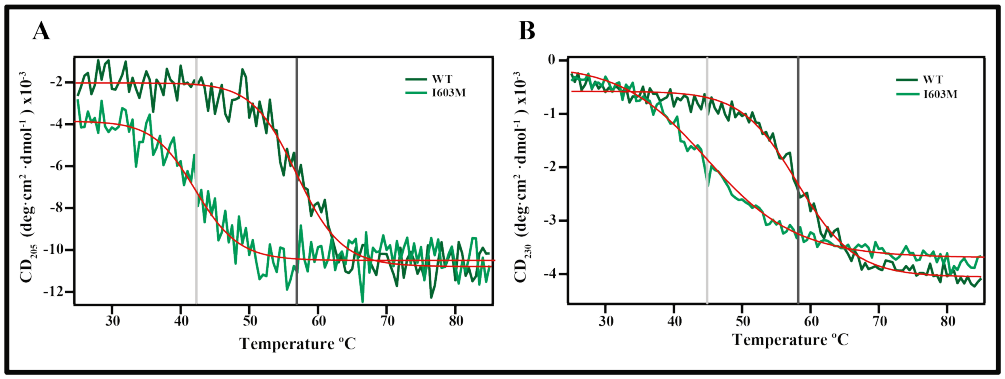

Figure 27: temperature-induced unfolding of the C4 domain monitored by CD. (A) CD signal at 205nm for the C4 WT (dark green) and C4 I603M (light green) collected as a function of temperature. The T_m of C4 I603M (42°C) is indicated with light grey line; T_m of C4 WT (57°C) is indicated with dark grey line. (B) Ellipticity at 230 nm as function of temperature for C4 WT (dark green) and C4 I603M (light green). The light grey line indicates the T_m of C4 I603M (45°C); dark grey line represents C4 WT T_m (58°C). The fit of the data to the sigmoidal curves is shown in red.

CD experiments to study refolding of domain C4 were carried out by ramping the temperature down while collected ellipticity. The refolding ability of the domains was determined by collecting a final far-UV CD spectrum at 25°C. Over 90% of the native protein refolded into original secondary structure after the cooling from 85°C to 25°C, for both WT and mutant I603M (*Figure 28*). Therefore, these data suggest that thermal denaturation of C4 WT as well as C4 I603M are reversible processes.

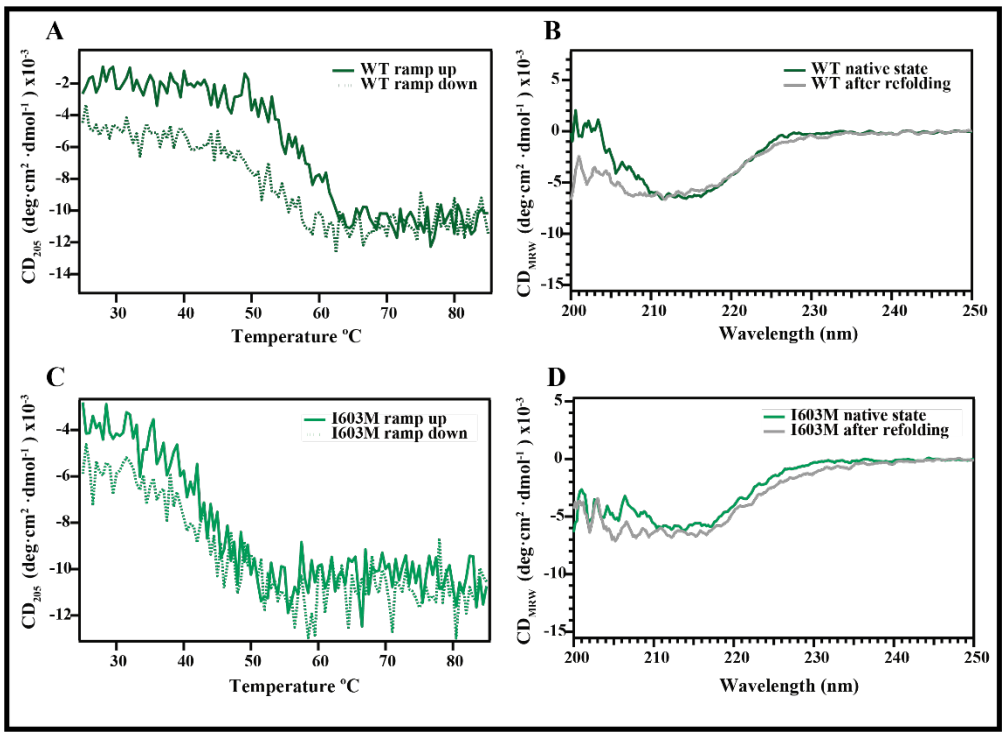

Figure 28: spectra characterization of the temperature-induced unfolding reversibility of C4 domain. Thermal denaturation (*solid line*) and renaturation (*dashed line*) of C4 WT (**A**) and C4 I603M (**C**) collected as change of ellipticity during thermal ramp up (from 25°C to 85°) and thermal ramp down (from 85°C to 25°C). The far-UV CD spectra of C4 WT (**B**) and C4 I603M (**D**) of protein under native state (*green traces*) and refolded state (*grey traces*) are shown.

When the unfolding of a protein is reversible and two-state, and there is only a single unfolding transition, the thermodynamic parameters evaluated by CD are almost identical with calorimetric methods (Greenfield 2007). The Gibbs-Helmholtz equation was used to fit the change of ellipticity at a single wavelength as a function of temperature. Using this fitting the van't Hoff enthalpy (ΔH_v) was calculated from CD data at 205 nm A small difference in enthalpy was observed in WT and mutant proteins with ΔH_v of 280 kJ/mol for C4WT and ΔH_v of 276 kJ/mol for C4 I603M. The same fitting was calculated also from CD data at 230 nm, with lower noise. The resulting van't Hoff enthalpy ΔH_v was different in WT and mutant, with ΔH_v of 220 kJ/mol for C4WT and ΔH_v of 130 kJ/mol for C4 I603M.

4.3.4 Thermodynamic stability of C4-WT and C4-I603M

The analysis of thermal stability by CD shows a marked decreased in T_m for the mutant, while the enthalpies of the unfolding transitions for the wild-type

and mutant I603M are very similar when monitored at 205 nm but different at 230 nm. To explore thermodynamic parameters by a more direct approach, differential scanning calorimetry (DSC) was carried out on domain C4-WT and C4 I603M. The thermograms of both WT and mutant were obtained in phosphate buffer and recorded at a scan rate of 30°C/h, which are the same conditions employed in CD experiments. The thermal peaks of both proteins indicate an exothermic transition with heat release (*Figure 29*). The transition could be seen also when the C4 WT and C4 I603M were subjected to a second heating scan after it were cooled from the first run (data not shown), showing that the unfolding of both polypeptides was almost completely reversible.

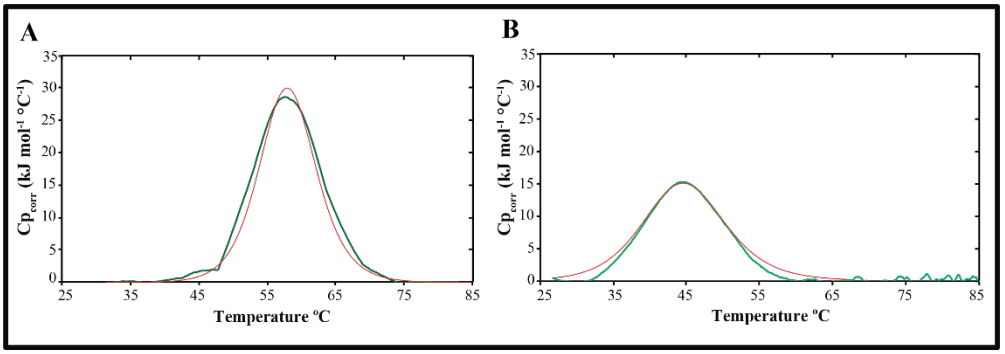

Figure 29: thermal unfolding monitored by DSC. Thermogram of C4 WT (**A**) and C4 I603M (**B**). Processed experimental data after baseline subtraction are shown in green and the fit of the data to the two-state model is shown in red.

The T_m values calculated in DSC are 58°C and 45°C for WT and I603M respectively, and were comparable to the values obtained in CD. The calorimetric enthalpy (ΔH_{cal}) was calculated by integration of the area under the thermogram peak and resulted in 364 kJ/mol for C4 WT and 196 kJ/mol for C4 I603M. The van't Hoff enthalpy (ΔH_v) was determined by the two-state fit of the thermogram. Again, the ΔH_v for C4 WT results higher than the value for C4 I603M (330 kJ/mol and 222 kJ/mol, respectively). The $\Delta H_{cal}/\Delta H_v$ ratio is 1.1 for WT and 0.89 for mutant. These values of ratio are close to the range of experimental error, and confirm that the two-state model is a good approximation to the unfolding process of C4 WT and I603M.

The Gibbs free energy change (ΔG) of the unfolding process was also calculated at 25°C considering $\Delta C_p=0$. The $\Delta\Delta G$ of C4 I603M resulted in 4.5 kcal/mol, suggesting that the mutant is less thermodynamically stable than wild-type.

4.3.5 Mechanical characterization of cMyBPC C4 domain by Atomic Force Microscopy (AFM)

Single molecule force spectroscopy has proven to be a powerful technique to study the mechanical properties of proteins as well as their conformational dynamics (Popa et al. 2013; Giganti et al. 2018). The information provided by force spectroscopy is particularly interesting for proteins that work under mechanical force, such as cMyBPC. Hence, we have studied the mechanical properties of C4 domain by single-molecule force-spectroscopy by AFM.

In single-molecule AFM experiments, it is necessary to fingerprint events of interest. A straightforward manner of achieving molecular fingerprinting is by the use of protein engineering to produce heteropolyproteins that contain marker proteins, of known mechanical properties. When pulling heteropolyproteins, mechanical unfolding of its constituent domains produce repetitive events of certain lengths. Those events corresponding to the marker proteins are accompanied, in the same single-molecule recording, by events corresponding to the domain of interest. In our case, we chose to engineer a heteropolyprotein of the C4 domain and Protein L, $(C4\text{-}L)_4$ (*Figure 30 A*). Protein L is naturally expressed in *P. magnus* as one of five homologous tandem domains. The protein is 63 amino acids in length and comprises a four-stranded β-sheet packed against a single a-helix. The Protein L is characterised by small and simple topology, but remarkable mechanical stability (Brockwell et al. 2005). The unfolding length of protein L has already been characterized (Glyakina et al. 2009; Sadler et al. 2009; Valle-Orero et al. 2017; Pimenta-Lopes et al. 2018).

The $(C4\text{-}L)_4$ heteropolyprotein was pulled at a constant velocity of 400nm/s to obtain the corresponding force *versus* extension curves. During force-extension measurements the separation between the cantilever and the surface changes linearly with time. These experiments measure the force applied to the protein as a function of its extension, giving rise to sawtooth pattern traces of unfolding, in which each "tooth" represents the unfolding and extension of an individual domain. To obtain the contour length increment for each protein, we used the worm-like chain (WLC) model of polymer elasticity. A typical force-extension curve for $(C4\text{-}L)_4$ is shown in *Figure 30B*. The last peak in the force-extension curve corresponds to the detachment of the polyprotein from the surface and/or the cantilever. In force-extension curves, the unfolding of Protein L is characterized by peaks around 100 pN with corresponding increments in contour length (Δc) of 19 nm (Sadler et al. 2009). Therefore, the unfolding events of Protein L provide a unique fingerprint that serves to identify successful traces. Surprisingly, no consistent unfolding peaks

assignable to the C4 domains were observed in these force-extension experiments, but only unfolding events corresponding to Protein L.

As an alternative approach more suited to identify unfolding steps, $(C4\text{-}L)_4$ was pulled in force-ramp mode applying a constant force thanks to control of piezo-electric element (Popa et al. 2013). The unfolding events were registered as stepwise changes in protein length in an extension *versus* time curve. The force was linearly varied at a rate of 10 pN/s for up to 20 s. The typical force-ramp trace showed four steps marking the unfolding of the Protein L, 17±1 nm long and occurring at around 50pN. As in force-extension recordings, no steps corresponding to the C4 domains were observed (*Figure 30 C*).

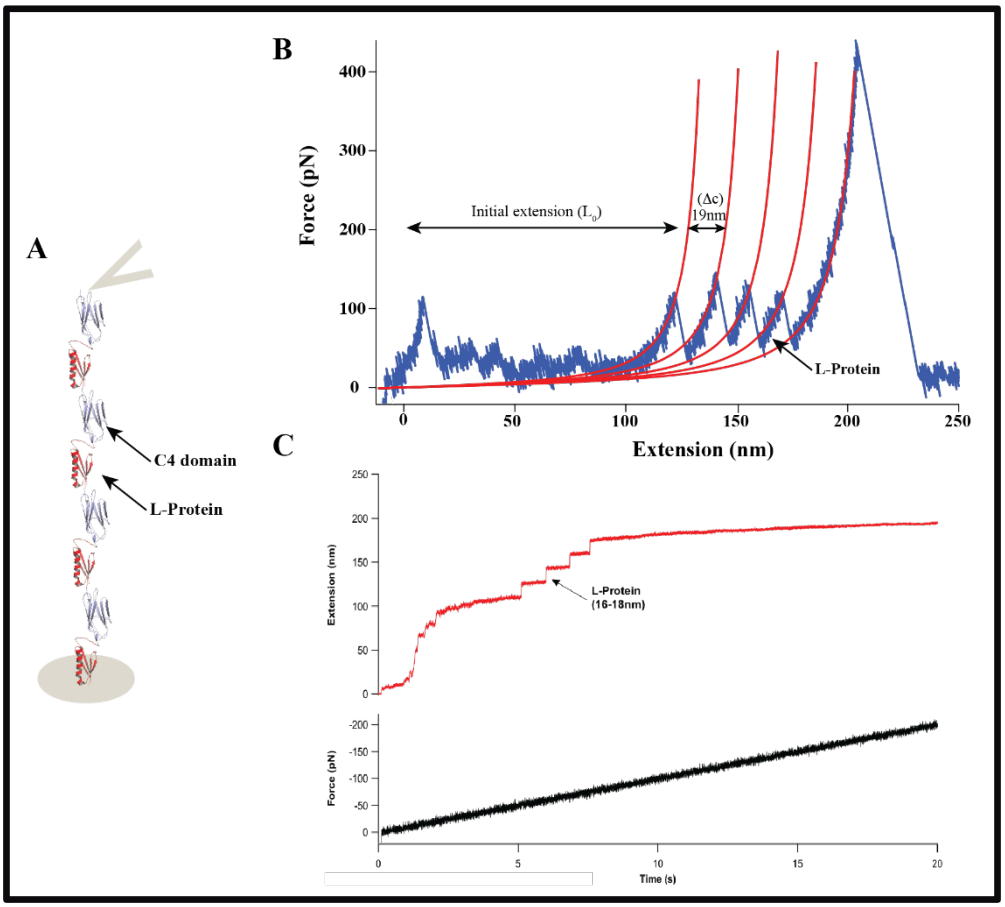

Figure 30: $(C4\text{-}L)_4$ heteropolyprotein investigated using single-molecule force-spectroscopy by AFM. (**A**) Schematic representation of the heteropolyprotein construct with a combination of four C4 (*grey structure*) and four fingerprinting protein-L (*red structure*) domains, used to probe the mechanical response of domain C4. (**B**) Typical force extension trace that captures four unfolding events of protein-L; no steps corresponding to the C4 domains were observed. The last event corresponds to the detachment of the polyprotien. WLC fit to the traces is shown in

red. The contour length (Δc) of 19 nm was measured for the four events, corresponding to L-Protein. The initial extension (L_0) was calculated from 0 to the first unfolding event. (**C**) Representative force-ramp trace showed the extension of $(C4\text{-}L)_4$ (*red* trace) and the force applied (*black trace*). Mechanical unfolding events of four Protein-L domains were detected, marked by unfolding lengths of 17±1 nm.

In order to define the molecular fingerprint of C4 domains in an unbiased manner, the step size and unfolding forces of all force-ramp traces containing at least 2 steps were analysed. The step lengths and the corresponding unfolding forces were represented in a bidimensional histogram (*Figure 31 A*; N = 2170 events). The histogram displays only one well-defined population corresponding to Protein L with 15±1 nm long and occurring around 50pN, as expected from previous AFM characterization of the domain (Sadler et al. 2009; Pimenta-Lopes et al. 2018). We conclude that there is no specific fingerprint associated with mechanical unfolding of C4 domain.

To rule out the possibility that Protein L was interfering with the mechanical response of domain C4, mechanical characterization of C4 was carried out using a different marker protein, Sumo1. Sumo1 is one of the four small ubiquitin-related modifiers (SUMOs) with known mechanical characteristics (Kotamarthi et al. 2013; Kotamarthi et al. 2015; Pimenta-Lopes et al. 2018). In the same manner as above, 390 traces of $(C4\text{-}Sumo1)_4$ under force-ramp, containing at least two events of the same size, were recorded and step size and unfolding forces were represented in a bidimensional histogram (*Figure 31 B*). Similar to results with $(C4\text{-}L)_4$, the data highlighted only one population, with 19±1 nm step size and around 130 pN unfolding force, which represent the Sumo1 protein.

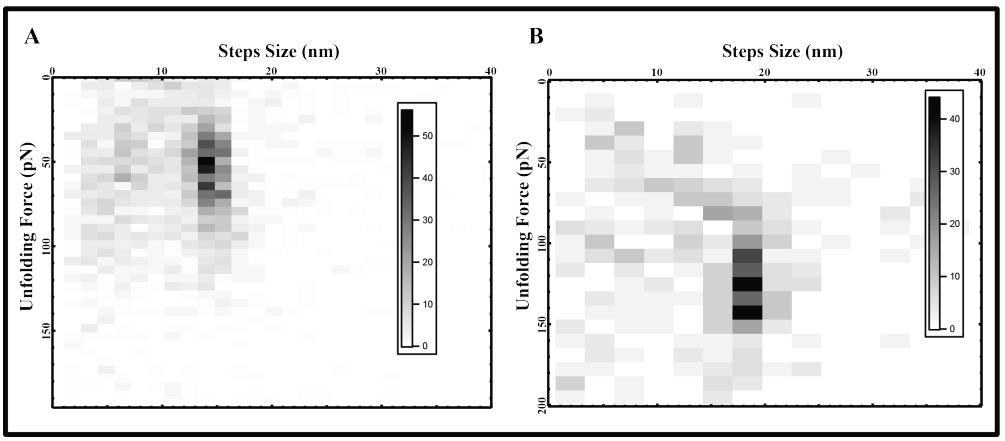

Figure 31: Fingerprinting of $(C4\text{-}L)_4$ and $(C4\text{-}Sumo1)_4$ heteropolyproteins. (**A**) Bidimensional histogram representing the step sizes and unfolding forces of 2170 events recorded in force-ramp mode for $(C4\text{-}L)_4$. A well-defined population

corresponds to unfolding of Protein-L (step size 15±1 nm at around 50pN) were identified. (**B**) Bidimensional histogram of the step sizes and the related unfolding forces from 390 AFM events of $(C4\text{-}Sumo1)_4$ in force ramp-mode. A well-defined populations correspond to unfolding of Sumo1 (step size 19±1 nm at around 130 pN) is observed.

The results of C4 fingerprinting are compatible with two different situations. The absence of homogeneous unfolding events coming from C4 might occur either if C4 extends before the first Protein L unfolding, or if it remains folded throughout the experiment. To determine which scenario is correct, the initial extension of the protein before the first Protein L unfolding event was estimated by fitting the worm-like chain model to the first unfolding peak of all traces that had the four Protein L events (n=10) (*Figure 30 B*). If the C4 domains do not unfold before Protein L, the initial extension would be in the range of 29.9-39.6nm. This value was calculated as the sum of length of three or four C4 modules (4 nm per domain), 4 folded Protein L domains (3 nm per domain), the handles and linkers included in the polyprotein. In contrast, if the C4 domains are unfolded before the first unfolding of protein-L, the initial extension should be 125.6-167.2nm. These values correspond to the extension of folded Protein L, the handles, linkers and three or four unfolded C4 domain.

An initial extension of 139 nm was obtained from a fit to the experimental histogram of initial extensions estimated using the WLC model (*Figure 30 B, Figure 32*). These data show that the C4 modules mechanically unfold before the Protein L modules.

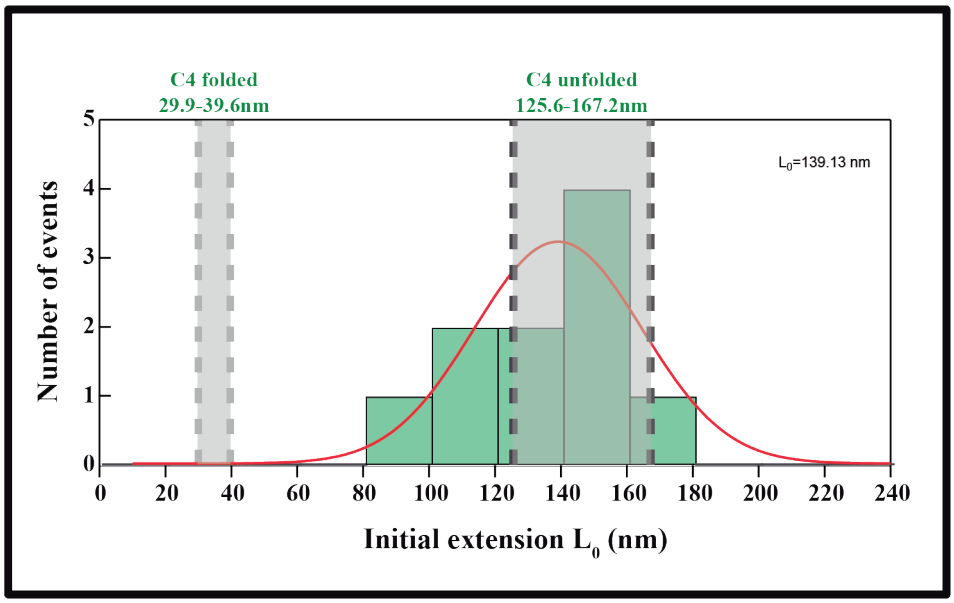

Figure 32: extensibility of C4 domain. Histogram of the initial extension for traces

with four L domains unfolding events (n=10). The expected initial extensions if the C4 behaved as random coils (*dotted line on the left*) or as mechanically stable domains (*dotted line on the right*) are shown. A Gaussian fit to the histogram is shown in red.

To investigate the possibility that the mutation I603M induced mechanical stabilization of the domain C4, characterization of (C4 I603M-Sumo1)$_4$ was carried out. The unfolding events were registered as changes in protein length in force-extension and force-ramp modes, as for the wild-type protein. Again, the typical force-extension trace showed four steps marking the unfolding of the Sumo1 and no steps corresponding to the C4 domains were observed (*Figure 33 A*). 890 traces under force ramp were recorded and analysed in order to define the mechanical fingerprinting of the mutant. Similarly, the bidimensional histogram highlighted only one population, with 19±1 nm step size and around 130 pN unfolding force, which represents the Sumo1 protein (*Figure 33 B)*. Hence, it is concluded that the mutation does not induce mechanical stabilization within the sensitivity of the AFM measurements.

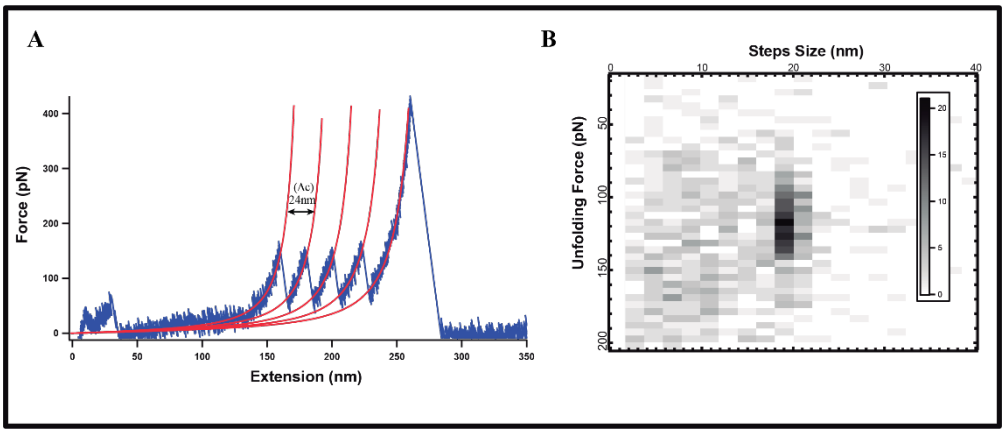

Figure 33: Fingerprinting of (C4 I603M-Sumo1)$_4$ heteropolyprotein. (**A**) Typical force extension trace represents four unfolding events of Sumo1; no peaks corresponding to the C4 domains were observed. The four peaks correspond to Sumo1 with a contour length (Δc) of 24 nm, measured with using the WLC model (in *red*) (**B**) Bidimensional histogram representing the step sizes and unfolding forces of 890 events recorded in force-ramp mode for (C4-Sumo1)$_4$. A well-defined population corresponds to unfolding of Sumo1 (step size 19±1 nm at around 130pN) was identified.

DISCUSSION

5.DISCUSSION

In the last few years, there has been an increasing interest in personalized medicine based on an individual's genes, environment, and lifestyle (Auffray et al. 2016; Nimmesgern et al. 2017; Scherr et al. 2018). The goal of Precision Medicine Initiative (PMI) is the identification of personalized interventions that not only include treatment, but also more targeted and cost-effective prevention and early-detection strategies. For those reasons, research has intensified to improve the understanding of disease mechanisms, diagnosis, and treatment outcomes mining the results of extensive genomic sequencing projects.

However, what do people want to know about their genes? According to a recent review, when asked about hypothetical genetic testing, most subjects would like to know all results except those that are uncertain, such as variants of uncertain significance (VUS) (Scherr et al. 2018). Indeed, in genetic screens of hereditary disorders, several different types of variants can be found and classified as pathogenic mutations, benign polymorphisms or VUS according to their known health impacts (Hershberger et al. 2018). The majority of clinical variants across all genes are currently classified as VUS. A recent study, indeed, show that in ClinVar, VUS constitute over 50% of entries for missense variants (*Figure 33*) (Weile and Roth 2018). The increase in relative and absolute number of VUS variants over the last few years is astonishing, and reflects the broad availability of Next Generation Sequencing (NGS) techniques (*Figure 33*).

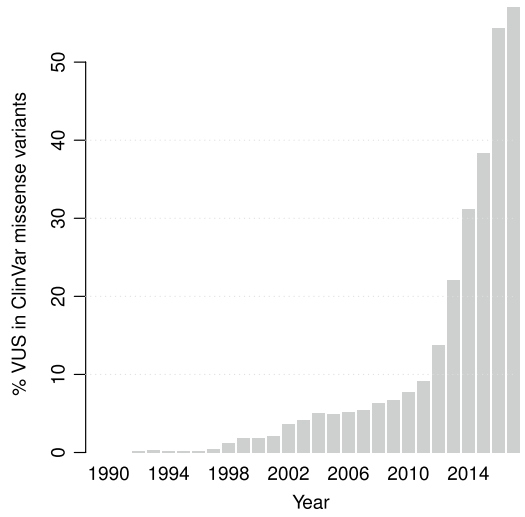

Figure 34: variants of uncertain significance (VUS) recorded in Clinvar database. The

histogram represents the percentage of missense VUS records over time from 1990 until 2017.(Adapted from Weile and Roth 2018)

Hypertrophic cardiomyopathy (HCM) is the most common familial heart disease with vast genetic heterogeneity (Burke et al. 2016; Liew et al. 2017; Sabater-Molina et al. 2018). The disease is caused by mutations in genes coding for sarcomeric proteins. Among them, cardiac myosin-binding protein C (MYBPC3) and β-myosin heavy chain (MYH7) genes are responsible for approximately half of the familial HCM cases.

The identification of HCM-causative mutations has revolutionized clinical management of patients and their families. Indeed, genetic testing is recommended after the initial identification of a pathogenic variant in order to classify asymptomatic relatives who might be at risk of disease-related complications, such as the risk of sudden cardiac death. However, the identification of VUS limits the reach of genetic testing in clinical practice due to the absence of pathogenicity assignment (Semsarian et al. 2015). Just think of MYBPC3, whose VUS constitute 44.6% of entries for missense variants shown in ClinVar database. Moreover, this percentage increases (66,5%) when all the mutations correlated to HCM are considered. Notably, the Association for Clinical Genetic Science does not recommend predictive testing on family members following the finding of a VUS in an index case (Ellard et al. 2017). Functional assessment of VUS can help define their pathogenic nature.

5.1. Selection of genetic variants for further functional screening

In this thesis the pathogenic effects of new genetic variants in MYBPC3 related to HCM have been investigated. MYBPC3 is the most frequently mutated gene in the cohort analysed. Indeed, 41% of the mutations found target MYBPC3. Similar numbers have been found in other genetic screens of HCM populations (Burke et al. 2016; Liew et al. 2017). Among the mutations targeting MYBPC3, seven variants had not been previously reported. Except for the null mutations, which by their nature have strong evidences of pathogenicity, c.506-2A>C, c.2308+3G>C and p.I603M variants can be classified as VUS and therefore were selected for functional studies.

Another missense mutation, p.T705K, had not been reported before. This mutation was excluded from functional study due to the absence of family genetic screening. Clinical cosegregation evidence could support further functional study of this variant in the future.

5.2 Pathogenicity study for c.506-2A>C and c.2308+3G>C variants

The c.506-2A>C and c.2308+3G>C are intronic variants. Intron variants play a crucial role in the aetiology of inherited cardiac conditions including cardiomyopathies. Indeed, approximately 10% of pathogenic variants causing cardiomyopathy are intronic mutations that lead to defects in alternative splicing mechanisms (Salman et al. 2018). Verification of a potential disease-related mutation was performed using *in silico* procedures and *in vitro* functional analysis of the process of RNA splicing.

The c.506-2A>C mutation disrupting the canonical AG/GT dinucleotides required for splicing resulted in a splice defect and the use of a cryptic acceptor site 7 nucleotides downstream (Figure 34). In the case of this variant, there was a 100% concordance between the *in silico* prediction, which highlighted the activation of a cryptic 3′ splice site, and the experimental results obtained using mRNA from the patient's blood (Figure 18).

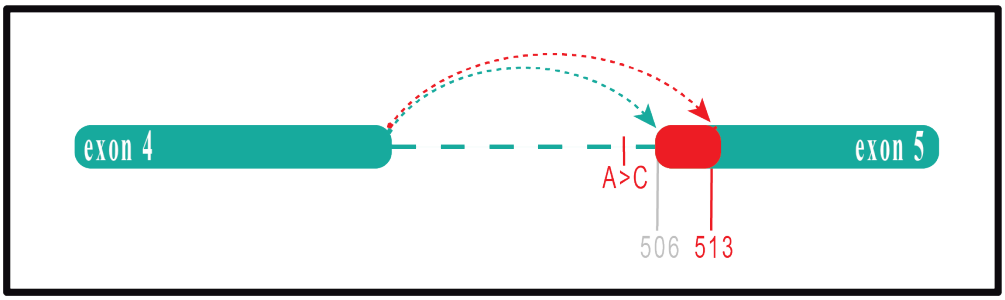

Figure 35: schematic representation of the splicing alteration induced by c.506-2A>C mutation. The arrows represent the intron removal during the splicing process (green arrow for canonical splicing and red arrow for mutant). The red box indicates the loss of seven nucleotides induced by the mutation. The A>C in red is the position of the mutation. The number in grey is the canonical acceptor site; the red number is the cryptic acceptor site.

Bioinformatics analysis of MYBPC3-c.2308+3G>C revealed that the score of the canonical donor splice site was decreased using all five prediction tools implemented in Alamut, and the decrease exceeded 50% for two of the prediction tools. Accordingly, minigene analysis of the mutated construct revealed skipping of MYBPC3 exon 23. In addition, HEK293 cells transfected with the WT minigene showed the coexistence of normal and skipped transcripts.

The aberrant splicing in WT sequence of MYBPC3 has also been described in a recent study (Ito, Parth N Patel, et al. 2017). The cell splicing assay in

HEK293 of MYBPC3's VUS, performed by Ito et al., show a high percentage of aberrant splice fragment in WT. In particular, the fragments including the WT donor site 2308G results in 25% of aberrant splicing, 75% without splice and only the 0,6% of fragments are processed canonically (Dataset S6 of Ito et al. 2017). We also implemented the minigene assay to study variants of C3 domain of MYBPC3, including exon 16 and 17. These assays also showed aberrant splicing in WT, including intron retention as well as exon 16 skipping (data not shown). Together, these data suggest that some canonical splice sites of MYBPC3 are weak and are not always recognized in HEK293 cells. Hence, minigene strategies have limited utility to examine the effect of mutations in RNA processing. Preferred methods involve the analysis of the patient's mRNA from myocardial or blood samples.

Finally, this study highlights the relevance of functional analysis to verify the pathogenicity of intronic variants, especially those that are located in exon-intron boundary. Hence, their importance in the diagnosis of HCM cannot be neglected, in agreement with recent reports using whole-genome sequencing (Bagnall et al. 2018).

5.3 Pathogenicity study for p.I603M variant

The p.I603M is a missense variant that affects a highly conserved residue in the central domain C4 of cMyBPC (*Figure 35*).

Figure 36: sequence alignment of Human cMyBPC C4 domain with other species. The red bar indicates the I603.

Results show that the variant c.1809T>G-p.I603M does not induce alterations in RNA splicing (*Figure 23*). At the protein level, the variant does not perturb much the structure of domain, according to the far-UV and near-UV CD spectra (*Figure 26*), in agreement with the homology models (*Figure 24*). Only small differences in the far-UV CD spectrum were observed in the range from

200nm to 205nm, which may be due to artefacts in a region of the spectra with higher noise. Although the mutant domain retains similar-to-WT structure, its thermodynamic stability is severely compromised, as shown by lower T_m in thermal denaturation experiments by CD and DSC, and by estimation of thermodynamic parameters by DSC.

The thermodynamic parameters obtained by CD at 205 nm provide very similar values of ΔH_v for WT and I603M (Table 9). Contrariwise, the van't Hoff enthalpy derived from CD at 230 nm result in a clear difference between WT and I603M. This contradiction probably reflects the fact that by using a single wavelength to monitor protein unfolding, we miss information about the global process of unfolding. In addition, the ΔH_v of CD estimated at 205 nm could be influenced by the higher noise of the experiment, and by the potential differences in the spectra at this region between wild-type and I603M (*Figure 27 A, 26 C*). In this regard, more accurate measurments of thermodynamic parameters can be obtained using DSC. The T_m values obtained by DSC resulted equal to CD and the thermodynamic parameters expressed as ΔH_v confirmed a strong difference between WT and I603M (Table 9).

Table 9: thermal stability parameters calculated by CD and DSC

	T_m CD	ΔH_v CD_{205}	ΔH_v CD_{230}	T_m DSC	ΔH_v DSC
C4 WT	57°C	280 kJ/mol	220 kJ/mol	58°C	330 kJ/mol
C4 I603M	42°C	276 kJ/mol	130 kJ/mol	45°C	222 kJ/mol

The Gibbs free energy change (ΔG) at 25ºC of the unfolding process was also calculated from the DSC thermograms. Since the thermograms of both WT and I603M did not allow accurate estimation of ΔC_p, the derivation of ΔG was carried out considering $\Delta C_p=0$. Using this approximation, the resulting $\Delta\Delta G$ of C4 I603M is 4.5 kcal/mol. Such degree of destabilization has been linked with pathogenicity in other domains of cMyBPC (Suay-Corredera et al. 2018). In that report, which analysed dozen of missense variants in cMyBPC, a specific feature of pathogenic mutations was strong thermodynamic destabilization over non-pathogenic variants. Values of $\Delta\Delta G$ higher than 3.0 kcal/mol were specifically associated with pathogenic mutations (*Figure 37*).

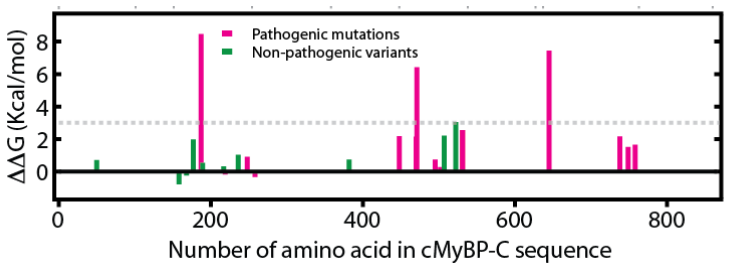

Figure 37: In silico examination of cMyBPC thermodynamic stability using the software FoldX. Each bar corresponds to ΔΔG calculated for a single variant. Pink bar correspond to pathogenic mutation, green bar correspond to non-pathogenic. The dotted line marks the highest change in ΔΔG detected for a non-pathogenic variant. (Figure adapted from Suay-Corredera et al. 2018).

Altogether, these data suggest that the mutant C4 I603M is much less thermodynamically stable than wild-type. Hence, it can be concluded that the I603M variant alters the stability of the central C4 domain of cMyBPC, which can lead to more frequent protein unfolding, degradation and protein haploinsufficiency, a hallmark of HCM (Van Dijk et al. 2009; Marston et al. 2009; Marston et al. 2012; Ito, Parth N. Patel, et al. 2017).

Finally, the mutation I603M was identified in clinical affected subjects in combination with another truncating variant of cMyBPC (T33RfsX15) (Figure 20). There is only one subject of the family that has cardiac hypertrophy in the absence of the mutation I603M, although the clinical manifestations probably are secondary to the coarctation of the aorta caused by an accident. In addition, there is one patient that carried only the truncated mutation and does not show HCM phenotype. The analysis of cosegregation of mutations in this family reflects the difficulties associated with genetic testing in a clinical setting, and the advantages offered by functional assessment of mutations. Overall, we can conclude that the mutation I603M is most likely pathogenic and that the presence of both mutated alleles can contribute synergistically to the clinical manifestation of HCM.

5.4 Pathogenic mechanisms of intronic and missense mutations in HCM

The functional study of intronic variants highlight the splicing defect affected by mutations. Usually, alternative splicing influences the structure of the mRNAs and their potential encoded proteins. The alternative splicing of mRNA is very important for the production of proteins with the right functionality, both in homeostasis and during development. The first study which showed the alternative splicing correlated with the contractile properties of embryonic myocardium regarded cardiac troponin T (cTnT) (Cooper and

Ordahl 1985; McAuliffe et al. 1990; Godt et al. 1993). In the embryonic heart, the inclusion of exon 5 of cTnT in mRNAs makes embryonic myofibrils more sensitive to calcium than adult and thereby influences the contractile properties of embryonic myocardium (McAuliffe et al. 1990; Godt et al. 1993). TnT is not the only protein influenced by alternative splicing in heart. Other genes, such as myomesin (Myom1), titin (ttn) and LIM domain-binding 3 (Ldb3) have been shown to have alternative splicing mRNA which leads to isoforms with distinct functions.(Huang et al. 2003; Lahmers et al. 2004; Schoenauer et al. 2011). However, any protein whose mRNA undergoes splicing is a potential target of mutations that disrupt this process, which can cause disease (Xiong et al. 2015). In cardiomyopathy, abnormal splicing of sarcomeric and ion channel genes has been reported in several studies. These changes can ultimately alter the normal internal architecture and homeostasis of the heart leading to heart failure (Lara-Pezzi et al. 2013; Van Den Hoogenhof et al. 2016; Zhu et al. 2017; Noyes et al. 2017).

The effect of mutations that alter the splicing process can be related to the activation of the nonsense-mediated decay (NMD) pathway. NMD is the process by which aberrant mRNAs containing premature termination codons (PTC) are targeted for accelerated degradation (Maquat 2004; Brogna and Wen 2009). This process is thought to protect cells from the potentially deleterious effects of inappropriately truncated proteins. The NMD has been shown to be the major pathway involved in clearance of MYBPC3 mRNA with PTC mutations in mice (Vignier et al. 2009). The transcript degradation via NMD pathway reduces mutant allele expression and the total amount of proteins is decreased. When the normal allele expression is insufficient, the haploinsufficiency model could explain the pathogenicity effects. However, previously studies have already shown that the reduced expressions of cMyBPC are correlated with HCM development and contractile deficits of affected patients (Van Dijk et al. 2009; Marston et al. 2009; Harris et al. 2011).

Regarding exonic variants, the most frequent mutations found in cMyBPC are truncating. Although protein truncation can result in normal levels of mutated mRNA, no detectable truncated polypeptide has been found in the myocardium of patients. The absence of truncated products is due to the degradation of aberrant polypeptides by cellular protein quality control systems, such as the ubiquitin proteasome system (UPS). The UPS represent the major eukaryotic proteolytic pathway (Ciechanover 1998). This degradation system, through an enzymatic complex (E1/E2/E3) marking proteins by polyubiquitination, which tags them for destruction by the 26S proteasome (Wilkinson 2000; Pickart and Cohen 2004; Xu et al. 2009). A number of E3 ligases have been shown to ubiquitinate misfolded proteins in an Hsp70-dependent manner (Arndt et al.

2007; Vembar and Brodsky 2008). Thus, the thermal instability, which leads the increasing of unfolding fractions of mutated proteins, can make them more sensitive to degradation. This mechanism could potentially contribute to the molecular pathogenesis of HCM.

Finally, we can conclude that both the alteration of RNA splicing and the thermodynamic destabilization of cMyBPC may lead to reduced protein levels, similar to the situation induced by truncating mutations, contributing with the HCM pathogenesis through haploinsufficiency. This model is supported by bioinformatics predictions and experimental data with domain C3 (Suay-Corredera et al. 2018).

However, mutations can also perturb protein-protein interactions. Indeed, mutations of sarcomeric components that affect the thermodynamic stability leads to alterations in protein binding and are correlated with HCM phenotype (Kremneva et al. 2004; Armel and Leinwand 2010; Da'as et al. 2018). Although the central domains of cMyBPC, including C4, are not described to interact with actin or myosin, we cannot entirely discount this possibility. Certainly, much remains to be learned about the specific function of central domains of cMyBPC.

5.5 Functional characterizations of VUS

Mutations in cMyBPC currently pose a major challenge to genetic diagnosis, since it is hard to move variants into the "likely pathogenic" category for any given individual. According to the American College of Medical Genetics and Genomics (ACMG), in the absence of enough genetic support, the most important criterion to establish causality of putative disease-causing mutations is a functional study to estimate the impact of mutations on a gene or protein function.

In summary, we propose a useful strategy to determine pathogenicity (*Figure 37*). When variants occur in introns as well as in the splice site of exon regions, the bioinformatics analyses with SSF, MaxEnt, NN Splice, GeneSplicer and HSF can be used to evaluate the splicing mechanism (*Figures 17, 21, 22*). Among the results, splice site loss is a strong prediction of alterations that can be confirmed with experimental mRNA analyses. According to our results, validation of predicted new splicing sites is a strict requirement to discriminate frequent false positives (*Figure 17, 19*). Ideally, for tissue-specific gene expression, mRNA should be extracted from the corresponding tissue, which in the case of the heart can be challenging. We have been confirmed that the MYBPC3 ectopic mRNA in blood sample gives equivalent results to cardiac tissue (*Figure 22*). In those cases where samples from HCM patients is

unavailable, the minigene system is a viable alternative to investigate splicing mechanism (*Figure 19*), although the analysis for some specific splicing reactions can be challenging in this system (Ito, Parth N Patel, et al. 2017).

Similarly, the first evaluations of pathogenicity of missense variants is bioinformatics. *In silico* profiling can be used to predict protein thermodynamic stability (C. Suay-Corredera et al. 2018). Mutations that target proteins for which high resolution structures are not available cannot be screened accurately for thermodynamic destabilization *in silico*, and therefore need to be studied experimentally. Both Circular Dichroism and Calorimetry have been demonstrated to highlight structural alterations as well as thermodynamic phenotypes.

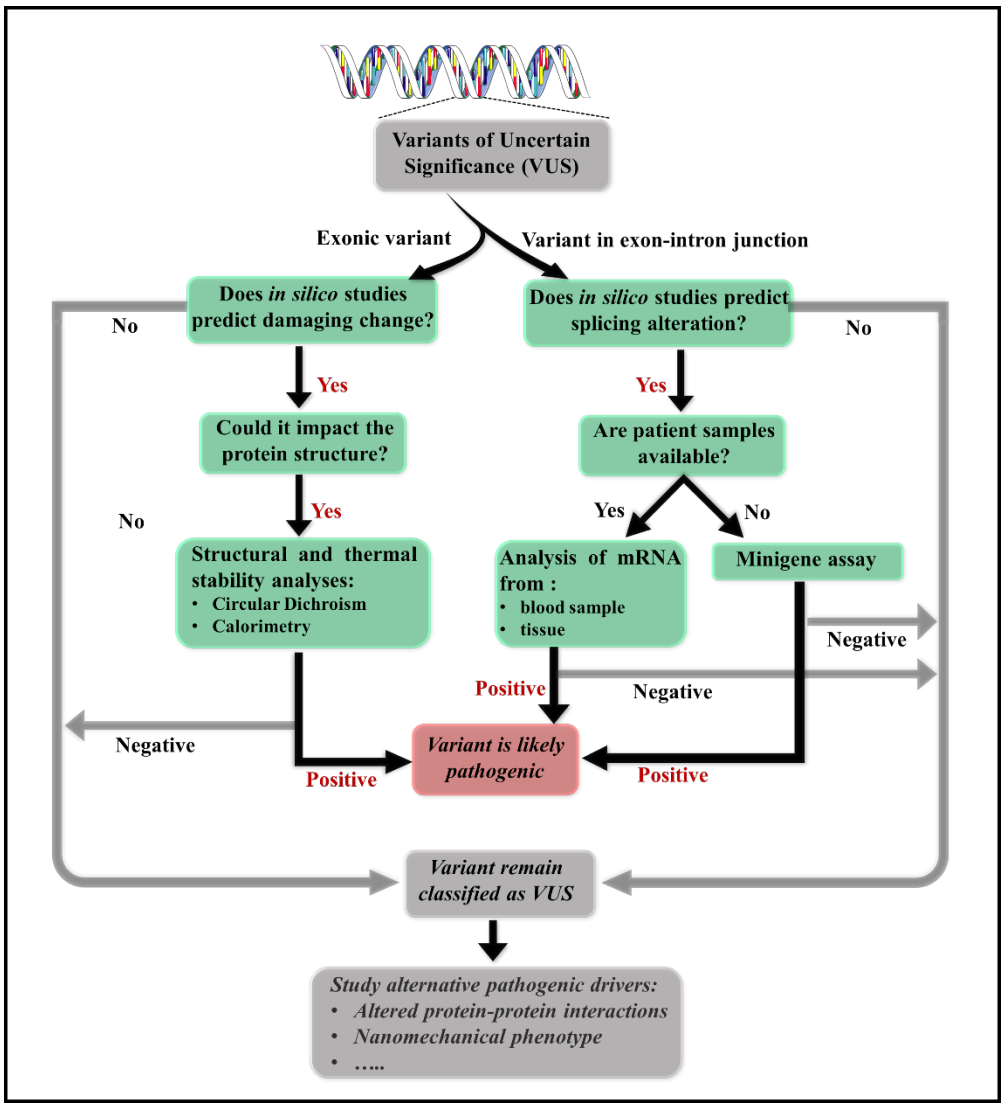

Figure 38: flowchart of functional determinations of VUS.

In summary, we propose that determination of RNA splicing and protein thermodynamic phenotypes can define HCM pathogenicity.

However, many missense pathogenic mutations in cMyBPC do not alter RNA splicing nor protein thermodynamic stability, resulting in stable proteins that are predicted to be readily incorporated in sarcomeres (Marston et al. 2012; Helms et al. 2014). In this case the study of pathogenicity should be focused on the other mechanisms. A plausible scenario is the alterations of binding to actin and/or myosin or other interactor who are involving in cardiomyocyte function as well as development (Vignier et al. 2009; Davis et al. 2016).

Recently, single-molecule AFM data suggested another alternative mechanism to explicate HCM pathogenicity. Indeed, any missense mutants, localized in central domain of protein, leads to nanomechanical destabilization of cMyBPC causing the alterations in the dynamics and strength of cMyBPC tethers in the sarcomere. This mechanical changes, resulting in altered regulatory properties of the mutated proteins, may be cause the pathogenicity of HCM.

Hence, it is highly recommended to study alternative pathogenic drivers for those variants that remain classified as VUS.

5.6 Mechanical characterization of cMyBPC C4 domain

The cMyBPC is a protein that works under mechanical force since it is tethered to the thick and thin filaments, which slide past each other during contraction. Hence, the study of mechanical properties of cMyBPC is of particular interest.

Single-molecule AFM experiments of cMyBPC show no consistent unfolding peaks assignable to the C4 domains (*Figure 31*). The absence of mechanical fingerprinting can be concluded because C4 extends before the first L unfolding, as demonstrated in the study of initial extensibility (*Figure 32*). The mutant I603M was also tested and revealed a same absence of mechanical stability then WT domain.

In light of these results, the central structure of cMyBPC should be examined more carefully. Interestingly, there is a 10-amino-acid segment between domains C4 and C5, which forms a flexible linker (*Figure 38*) (Idowu et al. 2003). The length of this linker is longer than the one found in slow skeletal isoform of myosin-binding protein. Indeed, typical linker lengths in poly-Ig proteins is shorter than 10 amino acids, suggesting that the extra length of this particular linker may be required to stabilize C4 and/or C5 domains. In agreement with this hypothesis, CD analyses on C5 domain shows the importance of a part of this linker in the stability of the C5 (Idowu et al. 2003). Consequently, we plan to examine the high-resolution crystallographic structure of C3-C5 multidomain construct, and to undertake single-molecule AFM experiments with the same construct.

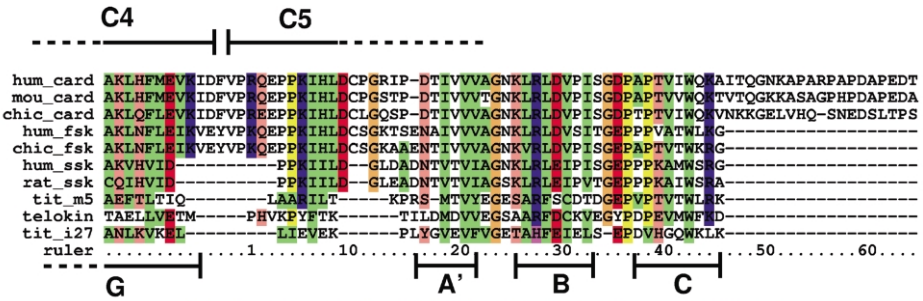

Figure 39: Sequence alignment of the C5 domain and part of the C4 domain with other Ig domains of known structure (from titin: tit_m5 PDB: 1nct; tit_i27 PDB 1tit and from telokin: 1tlk). The alignment is based on a comparison of the NMR structure of cC5 with the structure of the other Ig domains (Adapted from Idowu et al. 2003).

Nevertheless, it cannot be excluded that the force range of single-molecule AFM is high for detect mechanical unfolding transitions of the C4 domain. To test this hypothesis, we plan to do experiments by magnetic tweezers, another single-molecule manipulation techniques that has better resolution at low forces (Neuman and Nagy 2008; Popa et al. 2016).

6.CONCLUSIONS

The conclusions of this thesis can be summarized as follows:

1. three new mutations in Myosin Binding Protein C have been described;
2. the study of the RNA splicing process show that *in silico* evaluations are efficient to predict canonical splice site loss, but can be unreliable to determine cryptic splice site acquisition;
3. the analyses of mRNA of MYBPC3 in peripheral blood gives equivalent results to mRNA obtained from cardiac tissue, and hence, it is useful to study the effects of mutations in the splicing process;
4. mutations c.506-2A>C and c.2308+3G>C lead to altered splicing and are predicted to be pathogenic;
5. the C4 domain of cMyBP-C is thermodynamically stable but appears to be mechanically weak;
6. the mutation I603M induces strong mechanical destabilization of the domain C4, pointing towards its pathogenic nature;
7. the study of the functional consequences of mutations leads to assignment of pathogenicity of variants of uncertain significance.

REFERENCES

Adzhubei I, Jordan DM, Sunyaev SR. 2015. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2.

Alegre-Cebollada J, Badilla CL, Fernández JM. 2010. Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. J Biol Chem. doi:10.1074/jbc.M110.102962.

Amato F, Bellia C, Cardillo G, Castaldo G, Ciaccio M, Elce A, Lembo F, Tomaiuolo R. 2012. Extensive molecular analysis of patients bearing CFTR-related disorders. J Mol Diagnostics. doi:10.1016/j.jmoldx.2011.09.001.

Armel TZ, Leinwand LA. 2010. A mutation in the β-myosin rod associated with hypertrophic cardiomyopathy has an unexpected molecular phenotype. Biochem Biophys Res Commun. doi:10.1016/j.bbrc.2009.11.062.

Arndt V, Rogon C, Höhfeld J. 2007. To be, or not to be - Molecular chaperones in protein degradation. Cell Mol Life Sci. doi:10.1007/s00018-007-7188-6.

Auffray C, Caulfield T, Griffin JL, Khoury MJ, Lupski JR, Schwab M. 2016. From genomic medicine to precision medicine: Highlights of 2015. Genome Med. doi:10.1186/s13073-016-0265-4.

Bagnall RD, Ingles J, Dinger ME, Cowley MJ, Ross SB, Minoche AE, Lal S, Turner C, Colley A, Rajagopalan S, et al. 2018. Whole Genome Sequencing Improves Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. J Am Coll Cardiol. doi:10.1016/j.jacc.2018.04.078.

Baralle D, Baralle M. 2005. Splicing in action: Assessing disease causing sequence changes. J Med Genet. doi:10.1136/jmg.2004.029538.

Brockwell DJ, Beddard GS, Paci E, West DK, Olmsted PD, Smith DA, Radford SE. 2005. Mechanically unfolding the small, topologically simple protein L. Biophys J. doi:10.1529/biophysj.105.061465.

Brogna S, Wen J. 2009. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol. doi:10.1038/nsmb.1550.

Buratti E, Chivers M, Královičová J, Romano M, Baralle M, Krainer AR, Vořechovský I. 2007. Aberrant 5′ splice sites in human disease genes: Mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. Nucleic Acids Res. doi:10.1093/nar/gkm402.

Burke MA, Cook SA, Seidman JG, Seidman CE. 2016. Clinical and Mechanistic Insights Into the Genetics of Cardiomyopathy. J Am Coll Cardiol. doi:10.1016/j.jacc.2016.08.079.

Bustamante C, Marko J, Siggia E, Smith S. 1994. Entropic elasticity of lambda-phage DNA. Science (80-). doi:10.1126/science.8079175.

Carrier L, Bonne G, Bährend E, Yu B, Richard P, Niel F, Hainque B, Cruaud C, Gary F, Labeit S, et al. 1997. Organization and sequence of human cardiac myosin binding protein C gene (MYBPC3) and identification of mutations predicted to produce truncated proteins in familial hypertrophic cardiomyopathy. Circ Res.

Carrier L, Mearini G, Stathopoulou K, Cuello F. 2015. Cardiac myosin-binding protein C (MYBPC3) in cardiac pathophysiology. Gene. doi:10.1016/j.gene.2015.09.008.

Choi Y, Chan AP. 2015. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. doi:10.1093/bioinformatics/btv195.

Ciechanover A. 1998. The ubiquitin-proteasome pathway: On protein death and cell life. EMBO J. doi:10.1093/emboj/17.24.7151.

Cooper A, Johnson C. 1994. Differential Scanning Calorimetry. In: Microscopy, Optical Spectroscopy, and Macroscopic Techniques SE - 10.

Cooper TA, Ordahl CP. 1985. A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternate splicing. J Biol Chem. doi:10.1186/1472-6947-10-6.

Da'as SI, Fakhro K, Thanassoulas A, Krishnamoorthy N, Saleh A, Calver BL, Safieh-Garabedian B, Toft E, Nounesis G, Lai FA, et al. 2018. Hypertrophic cardiomyopathy-linked variants of cardiac myosin binding protein C3 display altered molecular properties and actin interaction. Biochem J.:BCJ20180685. doi:10.1042/BCJ20180685.

Davis J, Davis LC, Correll RN, Makarewich CA, Schwanekamp JA, Moussavi-Harami F, Wang D, York AJ, Wu H, Houser SR, et al. 2016. A Tension-Based Model Distinguishes Hypertrophic versus Dilated Cardiomyopathy. Cell. doi:10.1016/j.cell.2016.04.002.

DeLano WL. 2002. The PyMOL Molecular Graphics System. Schrödinger LLC wwwpymolorg. doi:citeulike-article-id:240061.

Van Dijk SJ, Dooijes D, Remedios C Dos, Michels M, Lamers JMJ, Winegrad S, Schlossarek S, Carrier L, Cate FJT, Stienen GJM, et al. 2009. Cardiac myosin-binding protein C mutations and hypertrophic ardiomyopathy haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction. Circulation. doi:10.1161/CIRCULATIONAHA.108.838672.

den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, Mcgowan-Jordan J, Roux AF, Smith T, Antonarakis SE, Taschner PEM. 2016. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Hum Mutat. doi:10.1002/humu.22981.

Dupuis LJ, Lumens J, Arts T, Delhaas T. 2016. Mechano-chemical Interactions in Cardiac Sarcomere Contraction: A Computational Modeling Study. PLoS Comput Biol. 12(10):1–20. doi:10.1371/journal.pcbi.1005126.

Ellard S, Baple EL, Owens M, Eccles DM, Abbs S, Zandra C. 2017. ACGS Best Practice Guidelines for Variant Classification 2017. Assoc Clin Genet Sci.:1–12.

Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. 2010. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol. 653:249–257. doi:10.1007/978-1-60761-759-4.

Garfinkel AC, Seidman JG, Seidman CE. 2018. Genetic Pathogenesis of Hypertrophic and Dilated Cardiomyopathy. Heart Fail Clin. 14(2):139–146. doi:10.1016/j.hfc.2017.12.004.

Geske JB, Ommen SR, Gersh BJ. 2018. Hypertrophic Cardiomyopathy: Clinical Update. JACC Hear Fail. doi:10.1016/j.jchf.2018.02.010.

Giganti D, Yan K, Badilla CL, Fernandez JM, Alegre-Cebollada J. 2018. Disulfide isomerization reactions in titin immunoglobulin domains enable a mode of protein elasticity. Nat Commun. doi:10.1038/s41467-017-02528-7.

Glyakina A V., Balabaev NK, Galzitskaya O V. 2009. Mechanical unfolding of proteins L and G with constant force: Similarities and differences. J Chem Phys. doi:10.1063/1.3183974.

Godt RE, Fogaça RTH, Silva IK, Nosek TM. 1993. Contraction of developing avian heart muscle. Comp Biochem Physiol -- Part A Physiol. doi:10.1016/0300-9629(93)90197-C.

Greenfield NJ. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc. doi:10.1038/nprot.2006.202.

Greenfield NJ. 2007. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc. doi:10.1038/nprot.2006.204.

Harris SP, Lyons RG, Bezold KL. 2011. In the thick of it: HCM-causing mutations in myosin binding proteins of the thick filament. Circ Res. doi:10.1161/CIRCRESAHA.110.231670.

Helms AS, Davis FM, Coleman D, Bartolone SN, Glazier AA, Pagani F, Yob

JM, Sadayappan S, Pedersen E, Lyons R, et al. 2014. Sarcomere mutation-specific expression patterns in human hypertrophic cardiomyopathy. Circ Cardiovasc Genet. doi:10.1161/CIRCGENETICS.113.000448.

Henderson CA, Gomez CG, Novak SM, Mi-Mi L, Gregorio CC. 2017. Overview of the muscle cytoskeleton. Compr Physiol. 7(3):891–944. doi:10.1002/cphy.c160033.

Hershberger RE, Givertz MM, Ho CY, Judge DP, Kantor PF, McBride KL, Morales A, Taylor MRG, Vatta M, Ware SM. 2018. Genetic Evaluation of Cardiomyopathy—A Heart Failure Society of America Practice Guideline. J Card Fail. doi:10.1016/j.cardfail.2018.03.004.

Van Den Hoogenhof MMG, Pinto YM, Creemers EE. 2016. RNA Splicing regulation and dysregulation in the heart. Circ Res. doi:10.1161/CIRCRESAHA.115.307872.

Houdayer C. 2011. In silico prediction of splice-affecting nucleotide variants. Methods Mol Biol. doi:10.1007/978-1-61779-176-5_17.

Huang C, Zhou Q, Liang P, Hollander MS, Sheikh F, Li X, Greaser M, Shelton GD, Evans S, Chen J. 2003. Characterization and in vivo functional analysis of splice variants of cypher. J Biol Chem. doi:10.1074/jbc.M211875200.

Hutter JL, Bechhoefer J. 1993. Calibration of atomic-force microscope tips. Rev Sci Instrum. doi:10.1063/1.1143970.

Huxley AF, Niedergerke R. 1954. Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres. Nature. 173:971.

Huxley H, Hanson J. 1954. Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation. Nature. 173:973.

Idowu SM, Gautel M, Perkins SJ, Pfuhl M. 2003. Structure, stability and dynamics of the central domain of cardiac myosin binding protein C (MyBP-C): Implications for multidomain assembly and causes for cardiomyopathy. J Mol Biol. doi:10.1016/S0022-2836(03)00425-X.

Ingles J, Bagnall RD, Semsarian C. 2018. Genetic Testing for Cardiomyopathies in Clinical Practice. Heart Fail Clin. doi:10.1016/j.hfc.2017.12.001.

Ito K, Patel PN, Gorham JM, McDonough B, DePalma SR, Adler EE, Lam L, MacRae CA, Mohiuddin SM, Fatkin D, et al. 2017. Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing. Proc Natl Acad Sci U S A. doi:10.1073/pnas.1707741114.

Ito K, Patel PN, Gorham JM, McDonough B, DePalma SR, Adler EE, Lam L, MacRae CA, Mohiuddin SM, Fatkin D, et al. 2017. Identification of pathogenic gene mutations in *LMNA* and *MYBPC3* that alter RNA splicing. Proc Natl Acad Sci. doi:10.1073/pnas.1707741114.

J. F. Sambrook and D.W. Russell. 2001. Molecular cloning: A Laboratory Manual.3rd Ed. Vol.1. Cold Spring Harb Lab Press New York, USA.

Jarcho JA, McKenna W, Pare JA, Solomon SD, Holcombe RF, Dickie S, Levi T, Donis-Keller H, Seidman JG, Seidman CE. 1989. Mapping a gene for familial hypertrophic cardiomyopathy to chromosome 14q1. N Engl J Med. doi:10.1056/NEJM198911163212005.

Kotamarthi HC, Sharma R, Koti Ainavarapu SR. 2013. Single-molecule studies on polySUMO proteins reveal their mechanical flexibility. Biophys J. doi:10.1016/j.bpj.2013.04.008.

Kotamarthi HC, Yadav A, Ainavarapu SRK. 2015. Small peptide binding stiffens the ubiquitin-like protein SUMO1. Biophys J. doi:10.1016/j.bpj.2014.11.3474.

Krans JL. 2010. The Sliding Filament Theory of Muscle Contraction. Nat Educ. 3(9):66.

Kremneva E, Boussouf S, Nikolaeva O, Maytum R, Geeves MA, Levitsky DI. 2004. Effects of two familial hypertrophic cardiomyopathy mutations in α-tropomyosin, Asp175Asn and Glut180Gly, on the thermal unfolding of actin-bound tropomyosin. Biophys J. doi:10.1529/biophysj.104.048793.

Kufareva I, Abagyan R. 2012. Methods of protein structure comparison. Methods Mol Biol. doi:10.1007/978-1-61779-588-6_10.

Lahmers S, Wu Y, Call DR, Labeit S, Granzier H. 2004. Developmental Control of Titin Isoform Expression and Passive Stiffness in Fetal and Neonatal Myocardium. Circ Res. doi:10.1161/01.RES.0000115522.52554.86.

Lara-Pezzi E, Gómez-Salinero J, Gatto A, García-Pavía P. 2013. The alternative heart: Impact of alternative splicing in heart disease. J Cardiovasc Transl Res. doi:10.1007/s12265-013-9482-z.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem. doi:10.1146/annurev-biochem-060614-034316.

Liew A, Vassiliou V, Cooper R, Raphael C. 2017. Hypertrophic Cardiomyopathy—Past, Present and Future. J Clin Med. doi:10.3390/jcm6120118.

Lin BL, Song T, Sadayappan S. 2017. Myofilaments: Movers and rulers of

the sarcomere. Compr Physiol. 7(2):675–692. doi:10.1002/cphy.c160026.

Ma N, Zhang J, Itzhaki I, Zhang SL, Chen H, Haddad F, Kitani T, Wilson KD, Tian L, Shrestha R, et al. 2018. Determining the Pathogenicity of a Genomic Variant of Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells. Circulation. doi:10.1161/CIRCULATIONAHA.117.032273.

Maquat LE. 2004. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol. doi:10.1038/nrm1310.

Marian AJ, Braunwald E. 2017. Hypertrophic cardiomyopathy: Genetics, pathogenesis, clinical manifestations, diagnosis, and therapy. Circ Res. 121(7):749–770. doi:10.1161/CIRCRESAHA.117.311059.

Marston S, Copeland O, Gehmlich K, Schlossarek S, Carrrier L. 2012. How do MYBPC3 mutations cause hypertrophic cardiomyopathy? J Muscle Res Cell Motil. doi:10.1007/s10974-011-9268-3.

Marston S, Copeland O, Jacques A, Livesey K, Tsang V, McKenna WJ, Jalilzadeh S, Carballo S, Redwood C, Watkins H. 2009. Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency*. Circ Res. doi:10.1161/CIRCRESAHA.109.202440.

McAuliffe JJ, Gao L, Solaro RJ. 1990. Changes in myofibrillar activation and troponin C Ca2+binding associated with troponin T isoform switching in developing rabbit heart. Circ Res. doi:10.1161/01.RES.66.5.1204.

McNamara JW, Li A, Lal S, Bos JM, Harris SP, Van Der Velden J, Ackerman MJ, Cooke R, Dos Remedios CG. 2017. MYBPC3 mutations are associated with a reduced super-relaxed state in patients with hypertrophic cardiomyopathy. PLoS One. doi:10.1371/journal.pone.0180064.

Mijailovich SM, Fredberg JJ, Butler JP. 1996. On the theory of muscle contraction: Filament extensibility and the development of isometric force and stiffness. Biophys J. doi:10.1016/S0006-3495(96)79348-7.

Mohamed IA, Krishnamoorthy NT, Nasrallah GK, Da'as SI. 2017. The Role of Cardiac Myosin Binding Protein C3 in Hypertrophic Cardiomyopathy-Progress and Novel Therapeutic Opportunities. J Cell Physiol. doi:10.1002/jcp.25639.

Nadvi NA, Michie KA, Kwan AH, Guss JM, Trewhella J. 2016. Clinically Linked Mutations in the Central Domains of Cardiac Myosin-Binding Protein C with Distinct Phenotypes Show Differential Structural Effects. Structure. doi:10.1016/j.str.2015.11.001.

Nag S, Trivedi D V., Sarkar SS, Adhikari AS, Sunitha MS, Sutton S, Ruppel KM, Spudich JA. 2017. The myosin mesa and the basis of hypercontractility caused by hypertrophic cardiomyopathy mutations. Nat Struct Mol Biol. doi:10.1038/nsmb.3408.

Neuman KC, Nagy A. 2008. Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy. Nat Methods. doi:10.1038/nmeth.1218.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. doi:10.1093/nar/gkg509.

Nimmesgern E, Benediktsson I, Norstedt I. 2017. Personalized Medicine in Europe. Clin Transl Sci. doi:10.1111/cts.12446.

Noyes AM, Zhou A, Gao G, Gu L, Day S, Andrew Wasserstrom J, Dudley SC. 2017. Abnormal sodium channel mRNA splicing in hypertrophic cardiomyopathy. Int J Cardiol. doi:10.1016/j.ijcard.2017.08.071.

Offer G, Moos C, Starr R. 1973. A new protein of the thick filaments of vertebrate skeletal myofibrils. Extraction, purification and characterization. J Mol Biol. doi:10.1016/0022-2836(73)90055-7.

Pickart CM, Cohen RE. 2004. Proteasomes and their kin: Proteases in the machine age. Nat Rev Mol Cell Biol. doi:10.1038/nrm1336.

Pimenta-Lopes C, Suay-Corredera C, Velázquez-Carreras D, Sánchez-Ortiz D, Alegre-Cebollada J. 2018. Orthogonal fingerprinting for accurate and fast single- molecule mechanical profiling of proteins. Prepr biorxiv.:1–19.

Popa I, Kosuri P, Alegre-Cebollada J, Garcia-Manyes S, Fernandez JM. 2013. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. Nat Protoc. doi:10.1038/nprot.2013.056.

Roberto Dominguez and Kenneth C. Holmes. 2011. Actin structure and function. Annu Rev Biophys. doi:10.1146/annurev-biophys-042910-155359.

Sabater-Molina M, Pérez-Sánchez I, Hernández del Rincón JP, Gimeno JR. 2018. Genetics of hypertrophic cardiomyopathy: A review of current state. Clin Genet. 93(1):3–14. doi:10.1111/cge.13027.

Sadayappan S, De Tombe PP. 2014. Cardiac myosin binding protein-C as a central target of cardiac sarcomere signaling: A special mini review series. Pflugers Arch Eur J Physiol. doi:10.1007/s00424-013-1396-8.

Sadler DP, Petrik E, Taniguchi Y, Pullen JR, Kawakami M, Radford SE, Brockwell DJ. 2009. Identification of a Mechanical Rheostat in the Hydrophobic Core of Protein L. J Mol Biol. doi:10.1016/j.jmb.2009.08.015.

Salman OF, El-Rayess HM, Abi Khalil C, Nemer G, Refaat MM. 2018. Inherited Cardiomyopathies and the Role of Mutations in Non-coding Regions of the Genome. Front Cardiovasc Med. doi:10.3389/fcvm.2018.00077.

Scherr C, Aufox S, Ross A, Ramesh S, Wicklund C, Smith M. 2018. What People Want to Know About Their Genes: A Critical Review of the Literature on Large-Scale Genome Sequencing Studies. Healthcare. doi:10.3390/healthcare6030096.

Schlierf M, Li H, Fernandez JM. 2004. The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. Proc Natl Acad Sci. doi:10.1073/pnas.0400033101.

Schoenauer R, Emmert MY, Felley A, Ehler E, Brokopp C, Weber B, Nemir M, Faggian GG, Pedrazzini T, Falk V, et al. 2011. EH-myomesin splice isoform is a novel marker for dilated cardiomyopathy. Basic Res Cardiol. doi:10.1007/s00395-010-0131-2.

Semsarian C, Ingles J, Maron MS, Maron BJ. 2015. New perspectives on the prevalence of hypertrophic cardiomyopathy. J Am Coll Cardiol. doi:10.1016/j.jacc.2015.01.019.

Shen C, Du Y, Qiao F, Kong T, Yuan L, Zhang D, Wu X, Li D, Wu Y-D. 2018. Biophysical and structural characterization of the thermostable WD40 domain of a prokaryotic protein, Thermomonospora curvata PkwA. Sci Rep. 8(1):12965. doi:10.1038/s41598-018-31140-y.

Smelter DF, de Lange WJ, Cai W, Ge Y, Ralphe JC. 2018. The HCM-linked W792R mutation in cardiac myosin-binding protein C reduces C6 FnIII domain stability. Am J Physiol Circ Physiol. doi:10.1152/ajpheart.00686.2017.

Suay-Corredera C, Pricolo MR, Herrero-Galán E, Velázquez-Carreras D, Sánchez-Ortiz D, Urrutia-Irazabal I, García-Giustiniani D, Delgado J, Vilches S, Dominguez F, Barriales-Villa R, Frisso G, Serrano L, García-Pavía P, Monserrat L, Alegre-Cebollada J. 2018. Molecular phenotyping defines pathogenicity of cMyBP-C mutants that cause hypertrophic cardiomyopathy. Sci. Transl. Med. Submitted.

Sweeney HL, Hammers DW. 2018. Muscle contraction. Cold Spring Harb Perspect Biol. 10(2). doi:10.1101/cshperspect.a023200.

Teare D. 1958. Asymmetrical hypertrophy of the heart in young adults. Br Heart J. 20(1):1–8. doi:10.1136/hrt.20.1.1.

Valle-Orero J, Rivas-Pardo JA, Popa I. 2017. Multidomain proteins under force. Nanotechnology. doi:10.1088/1361-6528/aa655e.

Vembar SS, Brodsky JL. 2008. One step at a time: Endoplasmic reticulum-associated degradation. Nat Rev Mol Cell Biol. doi:10.1038/nrm2546.

Vignier N, Schlossarek S, Fraysse B, Mearini G, Krämer E, Pointu H, Mougenot N, Guiard J, Reimer R, Hohenberg H, et al. 2009. Nonsense-mediated mrna decay and ubiquitin-proteasome system regulate cardiac myosin-binding protein c mutant levels in cardiomyopathic mice. Circ Res. doi:10.1161/CIRCRESAHA.109.201251.

Weile J, Roth FP. 2018. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. Hum Genet. doi:10.1007/s00439-018-1916-x.

Wijnker P, Sequeira V, Kuster D, Velden J van der. 2018. Hypertrophic cardiomyopathy: a vicious cycle triggered by sarcomere mutations and secondary disease hits. Antioxid Redox Signal.:ars.2017.7236. doi:10.1089/ars.2017.7236.

Wilkinson KD. 2000. Ubiquitination and deubiquitination: Targeting of proteins for degradation by the proteasome. Semin Cell Dev Biol. doi:10.1006/scdb.2000.0164.

Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, Hua Y, Gueroussov S, Najafabadi HS, Hughes TR, et al. 2015. The human splicing code reveals new insights into the genetic determinants of disease. Science (80-). doi:10.1126/science.1254806.

Xu P, Duong DM, Seyfried NT, Cheng D, Xie Y, Robert J, Rush J, Hochstrasser M, Finley D, Peng J. 2009. Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation. Cell. doi:10.1016/j.cell.2009.01.041.

Yue P, Li Z, Moult J. 2005. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol. doi:10.1016/j.jmb.2005.08.020.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. doi:10.1186/1471-2105-9-40.

Zhu C, Chen Z, Guo W. 2017. Pre-mRNA mis-splicing of sarcomeric genes in heart failure. Biochim Biophys Acta - Mol Basis Dis. doi:10.1016/j.bbadis.2016.11.008.

LIST OF PUBLICATIONS

Frisso G, Detta N, Coppola P, Mazzaccara C, **Pricolo MR**, D'Onofrio A, Limongelli G, Calabrò R, Salvatore F.
Functional Studies and In Silico Analyses to Evaluate Non-Coding Variants in Inherited Cardiomyopathies.
Int J Mol Sci. 2016 Nov 10;17(11). pii: E1883.

Suay-Corredera C, **Pricolo MR**, Herrero-Galán E, Velázquez-Carreras D, Sánchez-Ortiz D, Urrutia-Irazabal I, García-Giustiniani D, Delgado J, Vilches S, Dominguez F, Barriales-Villa R, Frisso G, Serrano L, García-Pavía P, Monserrat L, Alegre-Cebollada J.
Molecular phenotyping defines pathogenicity of cMyBP-C mutants that cause hypertrophic cardiomyopathy.
Sci. Transl. Med. 2018. Submitted.

MR Pricolo, E Herrero-Galán, C Mazzaccara, J Alegre-Cebollada, G Frisso.
Thermodynamic phenotypes guide the pathogenicity ascertainment of a variant of uncertain significance in cardiac myosin binding protein C.
Manuscript in preparation.