Published in final edited form as:

Proteomics Clin Appl. 2014 August ; 8(0): 578–589. doi:10.1002/prca.201400039.

Identification of Cardiac Myofilament Protein Isoforms Using Multiple Mass Spectrometry Based Approaches

Viola Kooij^{#1}, Vidya Venkatraman^{#1,2}, Jonathan A. Kirk¹, Ceereena Ubaida-Mohien³, David R. Graham^{1,3}, Matthijs J. Faber⁴, and Jennifer E. Van Eyk^{1,2}

¹Department of medicine, Division of Cardiology, The Johns Hopkins University, Baltimore, USA.

²Advanced Clinical Biosystems Research Institute, Heart Institute and Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, USA. ³Department of Molecular and Comparative Pathobiology, The Johns Hopkins University, Baltimore, MD, USA ⁴Erasmus MC-Sophia, Department of Pediatrics, Division of Pediatric Cardiology, Rotterdam, The Netherlands.

These authors contributed equally to this work.

Abstract

Purpose—The identification of protein isoforms in complex biological samples is challenging. We, therefore, used a mass spectrometry (MS) approach to unambiguously identify cardiac myofilament protein isoforms based on the observation of a tryptic peptide consisting of a sequence unique to a particular isoform.

Experimental design—Three different workflows were used to isolate and fractionate rat cardiac myofilament subproteomes. All fractions were analyzed on an LTQ-Orbitrap MS, proteins were identified using various search engines (Mascot, X!Tandem, X!Tandem Kscore and OMSSA) with results combined via PepArML Meta-Search Engine, and a post-search analysis was performed by MASPECTRAS.

Results—The combination of multiple workflows and search engines resulted in a larger number of non-redundant proteins identified than with individual methods. A total of 102 myofilament annotated proteins were observed overlapping in two or three of the workflows. Literature search for myofilament presence with manual validation of the MS spectra was carried out for unambiguous identification: 10 cardiac myofilament and 17 cardiac myofilament-associated proteins were identified with 39 isoforms and sub-isoforms.

Conclusion and clinical relevance—We have identified multiple isoforms of myofilament proteins that are present in cardiac tissue using unique tryptic peptides. Changes in distribution of these protein isoforms under pathological conditions could ultimately allow for clinical diagnostics or as therapeutic targets.

Correspondence: Vidya Venkatraman Cedars-Sinai Medical Center Advanced Clinical Biosystems Research Institute Advanced Health Science Pavillion, Room A9307 8700 Beverly Blvd. Los Angeles, CA 90048 Vidya.Venkatraman@cshs.org.

Conflict of interest statement

The authors have declared no conflict of interest

Introduction

Protein isoforms can differ in their biological activity, regulatory properties, temporal and spatial expression and intracellular location [1]. Protein isoforms harbor variations in the amino acid sequence and can originate from separate genes that have evolved from a single ancestor gene generated by gene duplication or from a single gene by alternative splicing. Furthermore, there can be sub-isoforms when a particular protein isoform has a number of different possibilities of mRNA splicing, generating additional splice variants [1].

The myofilament proteins, arranged to form the sarcomere of a myofibril, are the main regulators in striated muscle (skeletal and cardiac) contractility. Several proteins comprising the contractile apparatus of the heart have been shown to have unique isoforms, with specific functions and localizations. Some of these isoforms are expressed only during cardiac development or in response to disease [2]. For example, it was thought that just three isoforms of Tm are expressed in striated muscle: α , β , γ , and that Tm- α is the predominant isoform in adult cardiac tissue [3, 4]. However, Denz and colleagues identified a novel cardiac-specific Tm isoform in the human heart, due to a splice variant, which they designated TPM1 κ [5]. Subsequently, quantification of human cardiac mRNA levels revealed that ~50% of the TPM1 mRNA is TPM1 κ with the remainder, TPM1 α [6]. At protein level, TPM1 κ is increased in patients with end-stage heart failure [6]. Another myofilament protein with multiple tissue specific isoforms is α -actinin. Alpha-actinin isoforms 2 and 3 are commonly associated with all striated muscle, with α -actinin 2 as the major isoform in cardiac tissue, and 1 and 4 proposed to be the non-muscle isoforms [7]. For the most part, sub-isoforms have not been investigated at the protein level and to date, none have been identified in the cardiac myofilament subproteome.

Even though efforts have been made to identify and characterize protein isoforms, a comprehensive analysis of protein isoforms and sub-isoforms of cardiac myofilament proteins has been difficult due to the challenges of identification [8]. This project aims to identify isoforms and sub-isoforms of the cardiac myofilament proteins in the myofilament-enriched subproteome of rat cardiac tissue. To increase confidence in the detection of protein isoforms in cardiac samples, we combined three workflows, combining two myofilament enrichment methods [9, 10] with two different protein separation strategies, mass spectrometry (MS) based analysis, and the use of multiple search engines. Unambiguous assignment of a particular protein isoform was designated based on the observation of a tryptic peptide comprising an amino acid sequence that is unique to the specific isoform or sub-isoform.

Materials and methods

Extraction methods of the myofilament subproteome

Whole hearts were collected from 7-8 weeks old Sprague-Dawley Rats (Pel Freezer Biologicals). A single heart was used for each workflow. The first method used to isolate cardiac muscle myofibrils was according to the original preparation by Solaro et al. using Triton X-100 [10]. Briefly, tissue from a rat heart was homogenized in 75 mM KCl₂, 2 mM MgCl₂, 2 mM EGTA, 1 mM NaN₃, 10 mM imidazole, 4 mM phosphocreatine, 1 mM ATP,

50 mM BDM, 1 mM DTT, 1 mM benzaamidine-HCl, pH 7.2 (commonly termed relax buffer [11]) with 1% Triton X-100, 0.1 M EDTA and 1/100 μ l phosphatase inhibitor II (Sigma) and protease inhibitor cocktail (Sigma) added. Homogenates were centrifuged at 3,000g for 5 min at 4 °C and the supernatant fraction removed. The pellet was resuspended in a 75 mM KCl₂, 2 mM MgCl₂, 2 mM EGTA, 1 mM NaN₃, 10 mM imidazole, pH 7.2 (also called rigor buffer) containing 1% Triton X-100 and the same protease and phosphatase inhibitors added to the buffer as with the relax buffer. The pellet after centrifugation at 3,000 g for 5 min at 4 °C was washed two times in rigor buffer without Triton X-100 added followed by one wash in relax buffer without Triton X-100. After the last washing step, the pellet was dissolved in Loading buffer (Invitrogen) containing 2% SDS and 0.1 M DTT.

A second method to obtain the myofilament subproteome was a modified version of the “IN Sequence” subproteomic extraction method, which uses only one buffer extraction to remove the cytoplasmic and soluble proteins [9]. Briefly, rat heart tissue was homogenized at 4 °C at the ratio of tissue weight:buffer volume = 1:4 (whole tissue homogenate) in 20 mM Tris (pH 6.8) and 2mM EDTA with 1/100 μ l protease and phosphatase inhibitors (Sigma). The homogenate was centrifuged at 3000g and the pellet homogenated in 0.05% trifluoroacetic acid (TFA) and 1mM Tris(2-carboxyethyl)phosphine hydrochloride (pH 2.3). After the pellet was homogenized, it was spun down and the supernatant collected. This soluble fraction is highly enriched for the myofilament proteins.

Protein concentrations for both the Triton X-100 and “IN Sequence” myofilament enriched samples were determined using the CB-X assay (G-Biosciences) in triplicate and the samples aliquoted and stored at –80 °C until further use.

Protein fractionation

High-performance liquid chromatography (HPLC): The “IN Sequence” separated myofilament fractions were mixed with solvent to a final concentration of 20% (v/v) acetonitrile (ACN), 1% (v/v) TFA, pH 2.3 and centrifuged at 18000 g for 30 min at 4 °C. 100 μ g of myofilament enriched protein was resolved on a reversed phase C8 column (4.6 mm id x 76 mm, 80 Å pore size, particle size 3 μ m, Beckman Coulter, Fullerton, CA, USA) with a flow rate of 0.5 mL/min and monitored at 214 nm. The optimal separation was achieved using a linear AB gradient from 20-70% in which solvent A consisted of 0.6% aqueous TFA and solvent B was 0.45 % TFA in ACN. The column was regenerated using a 7 minute 70-100% linear A/C gradient where solvent C was composed of 2% TFA in isopropanol:acetonitrile (1:4 v/v)) prior re-equilibration in starting buffer. One minute fractions were collected, concentrated using a SpeedVac concentrator (ThermoSavant) to 5-10 μ L, and neutralized by the addition of ammonium bicarbonate (1 M) raising the pH of the sample to 8.0. 1-DE: The myofilament samples obtained from either Triton X-100 or “IN Sequence” separation methods were diluted in 2x sample buffer (Invitrogen) with 2% SDS and 0.1 M DTT to a concentration of 1 μ g/ μ L. 20 μ g of myofilament enriched protein homogenate was separated on a 4-12% SDS gradient gel. The gels were fixated using 5% acetic acid and stained using Colloidal Coomassie Blue staining (Biorad). The gel bands were cut in 8 pieces for fractionation purposes.

Protein digestion

Neutralized HPLC fractions were digested using sequencing grade modified trypsin (1:20) (Promega, Madison, WI) and incubated overnight at 37 °C. In gel digestion was carried out as previously described [11, 12]. Eluted peptides were dried and then desalted using an Omix C18 tip column (Varian) following the manufacturer's protocol.

LC-MS/MS analysis

Each trypsin digested sample was analyzed on an Agilent 1200 nano-LC system (Agilent, Santa Clara, CA, USA) connected to an LTQ-Orbitrap mass spectrometer (Thermo, Waltham, MA, USA) equipped with a nanoelectrospray ion source. Peptides were separated on a C₁₈ RP-HPLC column (75 µm x 10 cm self-packed with 5 µm, 200 Å Magic C18; MichromBio Resources, Auburn, CA, USA) at a flow rate of 300 nL / min where mobile phase A was 0.1% v/v formic acid in water and mobile phase B was 90% acetonitrile, 0.1 % formic acid in water. Initial loading condition was 2% B for 8.5 minutes at a flow rate of 2 µL/min to allow for equilibration and for samples to reach the column, followed by a linear gradient of 10-45% B over 60 minutes, 45-65% B over 11 minutes and 65-95% B over 2 minutes. Each MS1 scan was followed by collision induced dissociation (CID, acquired in the LTQ part) of the 5 most abundant precursor ions. Dynamic exclusion was enabled with a repeat count of 2, repeat duration of 30 seconds and exclusion duration of 60 seconds. Only MS1 signals exceeding 1000 counts triggered the MS2 scans and +1 and unassigned charge states were not selected for MS2 analysis. For MS1, 2x10⁵ ions were accumulated in the Orbitrap over a maximum time of 500 ms and scanned at a resolution of 60,000 FWHM (from 375-2000 m/z). MS2 spectra were acquired in normal scan mode in the LTQ with a target setting of 10⁴ ions and accumulation time of 30 milliseconds. Normalized collision energy was set to 35% and one microscan was acquired for each spectra.

Database searching and processing

MS raw data were converted to mzXML format using msConvert 2.2.2814 from ProteoWizard [13]. Data were uploaded and searched using the PepArML Meta-Search engine [14] that includes Mascot [15], OMSSA [16] and X!Tandem [17] with native and K-Score pluggable scoring modules against the concatenated target/decoy rat Uniprot database release date 2011-10-05, version 2011_09 [18, 19]. The search was performed with the following parameters: Trypsin-based digestion with a maximum of one missed cleavage; parent-ion mass tolerance of 2.0 Da; fragment-ion mass tolerance of 0.5 Da; a fixed modification of Carbamidomethyl (C) and variable modifications of Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), Met Oxidation, Phosphorylation (ST), Phosphorylation (Y) and Pyro-carbamidomethyl (N-term C). The results from the search were subsequently processed through the PepArML result combiner that merges all the search results together and estimates the statistical significance of the end-result using the decoy searches. The PepArML combined search results as well as the individual X!Tandem, Mascot and OMSSA search results were parsed and analyzed using MASPECTRAS 2 at a peptide false discovery rate (FDR) of 5% with proteins identified by one or more unique peptides [20]. **Supplemental Figure 1** shows the receiver operating characteristic (ROC) curves used for the FDR calculation. The MS spectrum-peptide match FDR was set to 1%, while the

peptide-protein match FDR was set to 5%. Since the isoforms were to be manually validated, the decreased confidence in protein identification was justified in exchange for a broader mapping of isoforms. Furthermore, since several search engines were used to identify peptides, it was found that the majority of peptides identified at <5% FDR were identified by multiple search engines, increasing our confidence in this FDR level. Furthermore, the CD-HIT clustering tool was used to cluster highly similar proteins with sequence homology above 90% over the entire protein [21]. A Blastp search was additionally performed on uncharacterized protein hits to resolve multiple protein name assignments for the same peptide sequence [22]. The isoform validation process was performed by a preliminary sequence homology comparison of all the isoforms and splice variants of the primary protein of interest from Human, Mouse and Rat using Clustal Omega Program [23] followed by the manual spectral validation of all the identified isoform-specific peptides. The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium (<http://www.proteomexchange.org>) via the PRIDE partner repository [24] with the dataset identifier PXD000874.

Results

Identification of the cardiac myofilament subproteome

To comprehensively analyze the isoforms and sub-isoforms of the cardiac myofilament proteins and to increase the confidence in their detection, three different experimental workflows were used to enrich, separate, and analyze the myofilament-enriched subproteome of rat cardiac tissue. An overview of the three workflows is given in **Figure 1**. The workflow entitled 1DLC'InSeq' combines the "IN Sequence" enrichment protocol with RP-HPLC fractionation and in-solution digestion; 1DE'InSeq' combines the 'IN Sequence' method with SDS-PAGE fractionation and in-gel digestion; and 1DE'triton' combines Triton X-100 enrichment with SDS-PAGE fractionation and in-gel digestion. Eighteen fractions were collected using RP-HPLC (three reproducible 1DLC'InSeq' RP chromatograms are shown in **Supplemental Figure 2A**), which were analyzed by MS. The samples 1DE'InSeq' and 1DE'triton' were separated on a 4-12% SDS-PAGE gradient and fractionated (**Supplemental Figure 2B**).

Furthermore, multiple search engines were used to analyze the 3 different MS data sets in an attempt to increase peptide coverage within a protein. Proteins identified from each of the individual search engines, as well as the number of proteins after merging using PepArML combiner, were clustered using the CD-HIT clustering tool. The number of proteins (with 1 peptides, **Table 1**) after removal of the redundant proteins shows that the PepArML Meta-Search Engine identifies more non-redundant proteins than the individual search engines Mascot, X!Tandem, X!Tandem Kscore, and OMSSA alone. This underlines the importance and benefits of combining multiple search engines when searching a protein database. The total numbers of clustered and unclustered proteins are presented in **Supplemental Table 1**. 2570 total unique proteins were identified after combining the three workflows before clustering using CD-HIT (1DLC'InSeq', 1DE'InSeq' and 1DE'triton', proteins are listed in **Supplemental Table 2A**).

Redundant or uncharacterized proteins that could not be differentiated from each other on the basis of sequence homology are removed using the CD-HIT clustering tool at 90% amino acid sequence homology (**Supplemental Table 2B**). In total, 1665 unique proteins were identified after combining the three workflows and clustering using CD-HIT, and are shown in the Venn diagram of **Figure 2A** (proteins are listed in **Supplemental Table 2B**). Workflow 1DLC'InSeq' resulted in the identification of more proteins than the 1DE'InSeq' or the 1DE'triton' workflows (1198, 656 and 459 proteins, respectively). 461 proteins were found in common between any two or three workflows, and 187 proteins were common to all three workflows. The number of proteins unique to each workflow and number of proteins overlapping between the workflows are depicted in the Venn diagram of **Figure 2A**. **Supplemental Table 3** lists the 461 proteins with the 187 common proteins marked in red. Uniprot Subcellular Location annotation was used to identify the cellular localization of these 461 proteins (**Figure 2B**), which resulted in the identification of 102 proteins assigned as the "myofilament subproteome" (listed in **Supplemental Table 4**). These proteins can be assumed to be present in most myofilament or myofibril sample preparations. A subsequent literature search to verify cardiac myofilament localization along with manual validation of the MS spectra resulted in the identification 10 cardiac myofilament proteins (**Table 2**) and 17 myofilament-associated proteins (**Table 3**). These groups can be considered to be the core myofilament proteins. **Figure 3** is a schematic representation of a cardiac muscle sarcomere with the location of the core myofilament (associated) proteins (those with unique isoforms shown *).

Protein isoforms

Thirty nine specific isoforms or sub-isoforms were unambiguously assigned for a subset of the core set of myofilament proteins such as isoforms 1, 2 and 4 of α -actinin (**Figure 4A**), and tropomyosin (TM) isoform 1, known as α -TM1 (α -TM1). Two sub-isoforms for the α -TM1 were identified specifically, α -TM1-1 and α -TM1-2 (**Figure 4B**). In addition, TM isoform 3 β (also known as gamma TM) was identified. But, the isoforms TM- β , TM- γ , and the novel isoform TMP1 κ , were not. Moreover, the cardiac isoform of MyBP-C and isoforms 1, 2, and N2BA of titin were observed in the cardiac tissue samples (**Table 2**). The largest number of isoforms and sub-isoforms were identified for myosin light chain (MLC) and myosin heavy chain (MHC); MLC-2a, -2c, -3, -4, -6 (**Figure 5A**) and MHC-4, -6, -7, -8, -9, -13, -14 (**Figure 5B**). MHC-9 and -14 are known as non-muscle isoforms and MHC -4, -8 and -13 which are expressed in skeletal muscle but have not been previously shown to be present in cardiac muscle. MLC-6 is known as a smooth muscle/ non-muscle isoform and MLC-4 is the embryonic atrial isoform. In addition, we identified actin-binding LIM protein 1, the B chain of α -crystallin, ankyrin 2, F-actin-capping protein alpha 2, filamin C, four and a half LIM domain 2, cysteine-rich protein 2 and 3, myomesin 1, myozenin 2, nebulin, tropomodulin 1, spectrin beta chain 1, synaptopodin 2 and transglutaminase 2 as isoforms of the myofilament-associated proteins (**Table 3**).

For actin, a major component of the thin filaments in the cardiac sarcomere, only the beta-actin like 2 isoform was unambiguously identified based on detection of a peptide exclusively found in the beta form. While this might appear surprising since the vast majority of actin present in the cardiomyocyte is cardiac alpha actin, with ~20% skeletal

alpha actin and minimal amount of beta actin [25]. However, this does not mean that alpha actin is not present, but rather that we were unable to differentiate at the peptide level between the various actin isoforms. This is, most likely due to the high amino acid sequence homology of the actin isoforms, with the majority of difference residing in the extreme N-terminus. Skeletal and cardiac alpha actin only differs in 4 amino acids in the N-terminus along the whole protein, and without detection of the tryptic N-terminal peptide these isoforms will not be unambiguously observed. As well, we could not identify isoforms for proteins troponin T and obscurin, which means that there were no unique peptides observed that could be assigned for a specific isoform of these proteins. Importantly, the failure to detect any unique tryptic peptide to an isoform suggests that they are either not present or are present in an abundance too low to be detected [26]. The comprehensive list of myofilament (associated) isoform-specific peptides are listed in **Table 2** and **Table 3**. The manually validated spectra are shown in **Supplemental Figure 3**.

Discussion

This study aimed to comprehensively analyze the isoforms of the cardiac myofilament proteins using two commonly used enrichment methods, two different protein separation methods, and with MS data analyzed by using multiple search engines for maximum assignment. 102 ‘high probability’ myofilament proteins were identified based on Uniprot Subcellular Location. Furthermore, we determined a final core cardiac myofilament (10 proteins) and myofilament-associated (17 proteins) subproteome that was based on further literature evaluation, and found 39 myofilament protein isoforms and sub-isoforms.

Myofilament proteins isoforms

The vast majority of the isoforms were observed for myosin, the thick filament enzyme involved in ATP hydrolysis required to produce the energy needed for muscle contraction. Myosin is comprised of two heavy and four light chains, and we unambiguously identified 7 MHC and 5 MLC isoforms. Comparison between the various MLC isoforms showed that MLC-2c was the most abundant isoform (242 spectral counts, which is an indication for abundance [24]) observed in the samples. This isoform is known as the ventricular/ cardiac isoform. MHC-7, the cardiac muscle β -isoform had the highest spectral count of the 7 different MHC isoforms based on spectral counts. Beta-MHC is predominantly expressed in the rat heart during development, and is replaced with α -MHC (MHC-6) after birth. With hypertrophy, the expression of β -MHC is re-induced. The rat hearts analyzed in this study were adult non-diseased hearts, where the alpha isoform should have been dominant. This demonstrates the limitation of the peptide-based MS approach. Since the identification of an isoform is solely dependent on the observation of a specific peptide that is unique to a specific isoform, the characteristic of this peptide in the MS instrument will dictate its sensitivity. Thus, even at the same concentration, two different peptides can have considerably different spectral counts.

Protein isoforms are different forms of the same protein that originated from separate genes or from one gene by alternative mRNA splicing. TM is expressed from 4 genes, TM1, TM2, TM3 and TM4, with isoforms generated by alternative exon expression, by the use of

alternate promoters. The alternate promoters result in two TM classes: high molecular weight (HMW) tropomyosins of ~284 residues which are expressed in both muscle and non-muscle cells, and low molecular weight (LMW) tropomyosins of ~248 residues found in the cytoskeleton. LMW TM have been shown to have less capacity to form head-to-tail polymers and have lower affinity for actin compared to HMW TM [27]. Interestingly, we identified the LMW isoform α -TM3, which is also known as gamma TM, and TM5, which is also known as alpha-TM, the predominant isoform expressed in cardiac adult tissue [28, 29]. For the first time, we identified two additional sub-isoforms for the alpha-TM isoforms: α -TM1-1, α -TM1-2, which differs in cell type (skeletal versus smooth muscle, respectively). Thus, it is not surprising that both sub-isoforms were detected since the heart tissues being analyzed have arteries embedded within the myocardium.

Alpha-actinin isoforms 1, 2 and 4 were identified in this study. Isoforms 2 and 3 are commonly associated with skeletal, cardiac and smooth muscle with α -actinin 2 being the major isoform present in cardiac tissue. Although isoform 1 is known as a non-muscle isoform [7], it is also present in the Z-line of striated muscle. Alpha-actinin isoform 4 is also known as a non-muscle isoform [7], but is present in epithelial cells and fibroblasts, cell types that are present in myocardium. We did not observe any unique tryptic peptide for α -actinin isoform 3, which might have been due to the high amino acid sequence homology of this isoform compared to the other isoforms.

A number of proteins known to associate directly to the cardiac myofilament, but whose roles in the myofilament are less characterized, were also included in the myofilament subproteome. This includes tissue type transglutaminase, or TG2. TG2 has the enzymatic activity of transamidation and the formation of a covalent bond between a free amine group (e.g., protein- or peptide-bound lysine) and the gamma-carboxamide group of protein- or peptide-bound glutamine. Bonds formed by transglutaminase exhibit high resistance to proteolysis [30]. The transglutaminases have direct myocardial actions such as regulating myocardial growth, fibrosis, and wound healing after myocardial infarction [30]. Human cardiac troponin T, a core myofilament protein, forms a substrate for transglutaminase in the presence of calcium overload [31], indicating direct involvement at myofilament level in the cardiomyocyte. Another protein is synaptopodin 2, which belongs to the synaptopodin family whose members have been shown to stimulate actin polymerization and bundling of actin filaments. Synaptopodin 2, also known as myopodin or fesselin is known to localize in the Z-bands of striated and heart muscle cells [32]. It is known to be an F-actin binding protein, and interacts with the myofilament-associated proteins filamin C and α -actinin [33, 34].

Limitations

Enrichment methods do not provide pure sub-proteomes, but only enrich for certain proteins. Therefore, this study combines three different myofilament enrichment workflows on rat cardiac tissue to increase the chance and confidence of myofilament protein isoform identification. Moreover, our myofilament (associated) protein lists, which were used to identify the presence of protein isoforms, was searched through the literature to ensure 'true' myofilament localization (**Table 2** and **Table 3**). Also, cardiac tissue contains

cardiomyocytes, fibroblasts and smooth muscle cells (e.g. endothelial cell), therefore, we cannot rule out that some of the cardiac myofilament protein isoforms of **Table 2**, or the myofilament-associated protein isoforms of **Table 3** are not cardiomyocyte specific.

CDHIT was used to remove redundant proteins, which has the potential to remove some unique protein isoforms. To minimize this, the CDHIT clustered protein groups were re-evaluated manually to ensure no minor isoforms were eliminated during removal of redundant proteins. In cases where they were removed, we added them back into our final list (e.g. some of the myosin isoforms). However, there is a probability that some low abundant isoforms are not included in this study due to i) lack of peptide evidence that allows us to distinguish between the isoforms, ii) lack of high-confidence spectrum during the validation process, iii) detection of an isoform in only one workflow instead of multiple workflows, which was our threshold.

Future clinical applications will likely involve only a single workflow, instead of the complex multi-workflow approach used here. As an alternative to a broad isoform interrogation, targeted clinical assays to quantify individual or small groups of isoforms, such as ELISA or multiple reaction monitoring, will allow for simplified and sensitive clinical workflows. The results presented here will help to develop these assays.

Conclusion

The aim of this project was to identify cardiac myofilament (associated) proteins isoforms present in adult rat heart. It is known that protein isoforms have a high impact on cardiac function and their activity, temporal and spatial expression as well as cellular location can change under pathological conditions. It is therefore important to identify, characterize function, and document changes during cardiac disease, which would ultimately allow the use of protein isoforms as therapeutic targets for clinical diagnostics. More research is necessary to identify the function, and potential change during disease, of the protein isoforms identified in this study. Moreover, as new technologies mature, such as SWATH mass spectrometry [35], these tools could help further elucidate the protein isoforms present in myofilament. The finding presented here could act as a benchmark for assessing new methods and technologies as they emerge.

A peptide from the unique amino acid sequence of a protein isoform is required for unambiguous identification, which hampers the identification of high sequence homology isoforms. This will have resulted in an under representation of various protein isoforms and subisoforms in our myofilament (associated) protein tables (**Table 2** and **Table 3**). However, isoform information can be used for the development of specific quantitative methods including antibodies or for the targeted MS-based method, multiple reaction monitoring.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the following funding source for their work: P01-HL077180 (to J.E. Van Eyk), Proteomic Initiative contracts NHLBI-HV-10-05(2) and HHSN268201000032C (to J.E. Van Eyk) and American Heart Association post-doctoral fellowship 12POST11520006 (to V. Kooij). Also, the authors would like to thank Dr. N. J. Edwards from Georgetown University for additional bioinformatics and computing support.

Abbreviations

1DE	one dimensional electrophoresis
1D-LC	one dimensional liquid chromatography
ATP	adenosine triphosphate
BDM	2,3- butanedione monoxime
FDR	false discovery rate
FWHM	full-width half-maximum
HCl	hydrochloric acid
HMW	high molecular weight
LMW	low molecular weight
MHC	myosin heavy chain
MLC	myosin light chain
MgCl₂	magnesium chloride
MS	mass spectrometry
NaN₃	sodium azide
SC	spectral count
TM	tropomyosin

References

1. Gunning PW. Protein Isoforms and Isozymes. eLS. 2006
2. Marston SB, Redwood CS. Modulation of thin filament activation by breakdown or isoform switching of thin filament proteins: physiological and pathological implications. *Circ.Res.* 2003; 93:1170–1178. [PubMed: 14670832]
3. Peng Y, Chen X, Zhang H, Xu Q, Hacker TA, et al. Top-down targeted proteomics for deep sequencing of tropomyosin isoforms. *J.Proteome Res.* 2013; 12:187–198. [PubMed: 23256820]
4. Wolska BM, Wieczorek DM. The role of tropomyosin in the regulation of myocardial contraction and relaxation. *Pflugers Arch.* 2003; 446:1–8. [PubMed: 12690456]
5. Denz CR, Narshi A, Zajdel RW, Dube DK. Expression of a novel cardiac-specific tropomyosin isoform in humans. *Biochem.Biophys.Res.Comm.* 2004; 320:1291–1297. [PubMed: 15249230]
6. Rajan S, Jagatheesan G, Karam CN, Alves ML, Bodi I, et al. Molecular and functional characterization of a novel cardiac-specific human tropomyosin isoform. *Circulation.* 2010; 121:410–418. [PubMed: 20065163]
7. Sjoblom B, Salmazo A, Djinoic-Carugo K. Alpha-actinin structure and regulation. *Cell Mol.Life Sci.* 2008; 65:2688–2701. [PubMed: 18488141]
8. Stastna M, Van Eyk JE. Analysis of protein isoforms: can we do it better? *Proteomics.* 2012

9. Neverova I, Van Eyk JE. Application of reversed phase high performance liquid chromatography for subproteomic analysis of cardiac muscle. *Proteomics*. 2002; 2:22–31. [PubMed: 11788988]
10. Solaro RJ, Pang DC, Briggs FN. Purification of Cardiac Myofibrils with Triton X-100. *Biochimica et Biophysica Acta*. 1971; 245:259–&. [PubMed: 4332100]
11. Kooij V, Holewinski RJ, Murphy AM, Van Eyk JE. Characterization of the cardiac myosin binding protein-C phosphoproteome in healthy and failing human hearts. *J.Mol.Cell Cardiol*. 2013; 60:116–120. [PubMed: 23619294]
12. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr.Protoc.Mol.Biol*. 2009 Chapter 10, Unit10.
13. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*. 2008; 24:2534–2536. [PubMed: 18606607]
14. Edwards N, Wu X, Tseng CW. An unsupervised, model-free, machine-learning combiner for peptide identifications from tandem mass spectra. *Clinical Proteomics*. 2009; 5.1:23–36.
15. Helsens K, Martens L, Vandekerckhove J, Gevaert K. MascotDatfile: An open-source library to fully parse and analyse MASCOT MS/MS search results. *Proteomics*. 2007; 7:364–366. [PubMed: 17203510]
16. Barsnes H, Huber S, Sickmann A, Eidhammer I, Martens L. OMSSA Parser: An open-source library to parse and extract data from OMSSA MS/MS search results. *Proteomics*. 2009; 9:3772–3774. [PubMed: 19639591]
17. Muth T, Vaudel M, Barsnes H, Martens L, Sickmann A. XTandem Parser: An open-source library to parse and analyse X!Tandem MS/MS search results. *Proteomics*. 2010; 10:1522–1524. [PubMed: 20140905]
18. MacLean B, Eng JK, Beavis RC, McIntosh M. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics*. 2006; 22:2830–2832. [PubMed: 16877754]
19. The Uniprot C. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2012; 40:D71–D75. [PubMed: 22102590]
20. Mohien CU, Hartler J, Rix U, Rix LR, Winter GE, et al. MASPECTRAS 2: an integration and analysis platform for proteomic data. *Proteomics*. 2010; 10:2719–2722. [PubMed: 20455215]
21. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010; 26:680–682. [PubMed: 20053844]
22. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389–3402. [PubMed: 9254694]
23. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011; 7:539. [PubMed: 21988835]
24. Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, et al. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res*. 2013; 41:D1063–1069. [PubMed: 23203882]
25. Copeland O, Nowak KJ, Laing NG, Ravenscroft G, Messer AE, et al. Investigation of changes in skeletal muscle alpha-actin expression in normal and pathological human and mouse hearts. *J.Muscle Res.Cell Motil*. 2010; 31:207–214. [PubMed: 20706863]
26. Tang HY, Beer LA, Tanyi JL, Zhang R, Liu Q, et al. Protein isoform-specific validation defines multiple chloride intracellular channel and tropomyosin isoforms as serological biomarkers of ovarian cancer. *J Proteomics*. 2013; 89:165–178. [PubMed: 23792823]
27. Matsumura F, Yamashiro-Matsumura S. Purification and characterization of multiple isoforms of tropomyosin from rat cultured cells. *J.Biol.Chem*. 1985; 260:13851–13859. [PubMed: 4055761]
28. Kobayashi T, Jin L, de Tombe PP. Cardiac thin filament regulation. *Pflugers Arch*. 2008; 457:37–46. [PubMed: 18421471]
29. Kobayashi T, Solaro RJ. Calcium, thin filaments, and the integrative biology of cardiac contractility. *Annu.Rev.Physiol*. 2005; 67:39–67. [PubMed: 15709952]

30. Sane DC, Kontos JL, Greenberg CS. Roles of transglutaminases in cardiac and vascular diseases. *Front Biosci.* 2007; 12:2530–2545. [PubMed: 17127261]
31. Gorza L, Menabo R, Di Lisa F, Vitadello M. Troponin T cross-linking in human apoptotic cardiomyocytes. *Am.J.Pathol.* 1997; 150:2087–2097. [PubMed: 9176400]
32. Schroeter MM, Orlova A, Egelman EH, Beall B, Chalovich JM. Organization of F-Actin by Fesselin (avian smooth muscle synaptopodin 2). *Biochemistry.* 2013
33. Linnemann A, Vakeel P, Bezerra E, Orfanos Z, Djinovic-Carugo K, et al. Myopodin is an F-actin bundling protein with multiple independent actin-binding regions. *J.Muscle Res.Cell Motil.* 2013; 34:61–69. [PubMed: 23225103]
34. Linnemann A, van der Ven PF, Vakeel P, Albinus B, Simonis D, et al. The sarcomeric Z-disc component myopodin is a multiadapter protein that interacts with filamin and alpha-actinin. *Eur.J.Cell Biol.* 2010; 89:681–692. [PubMed: 20554076]
35. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics.* 2012; 11:O111 016717. [PubMed: 22261725]
36. Pyle WG, Solaro RJ. At the crossroads of myocardial signaling - The role of Z-discs in intracellular signaling and cardiac function. *Circulation Research.* 2004; 94:296–305. [PubMed: 14976140]
37. Barrientos T, Frank D, Kuwahara K, Bezprozvannaya S, Pipes GC, et al. Two novel members of the ABLIM protein family, ABLIM-2 and -3, associate with STARS and directly bind F-actin. *J Biol Chem.* 2007; 282:8393–8403. [PubMed: 17194709]
38. Golenhofen N, Htun P, Ness W, Koob R, Schaper W, et al. Binding of the stress protein alpha B-crystallin to cardiac myofibrils correlates with the degree of myocardial damage during ischemia/reperfusion in vivo. *J Mol Cell Cardiol.* 1999; 31:569–580. [PubMed: 10198188]
39. Chopra A, Patel A, Shieh AC, Janmey PA, Kresh JY. alpha-Catenin localization and sarcomere self-organization on N-cadherin adhesive patterns are myocyte contractility driven. *PLoS One.* 2012; 7:e47592. [PubMed: 23077648]
40. Lazarides E, Hubbard BD. Immunological characterization of the subunit of the 100 A filaments from muscle cells. *Proc Natl Acad Sci U S A.* 1976; 73:4344–4348. [PubMed: 1069986]
41. Caldwell JE, Heiss SG, Mermall V, Cooper JA. Effects of CapZ, an actin capping protein of muscle, on the polymerization of actin. *Biochemistry.* 1989; 28:8506–8514. [PubMed: 2557904]
42. Koteliansky VE, Glukhova MA, Gneushev GN, Samuel JL, Rappaport L. Isolation and localization of filamin in heart muscle. *Eur J Biochem.* 1986; 156:619–623. [PubMed: 3516696]
43. Sheikh F, Raskin A, Chu PH, Lange S, Domenighetti AA, et al. An FHL1-containing complex within the cardiomyocyte sarcomere mediates hypertrophic biomechanical stress responses in mice. *Journal of Clinical Investigation.* 2008; 118:3870–3880. [PubMed: 19033658]
44. Wei TC, Lin HY, Lu CC, Chen CM, You LR. Expression of Crip2, a LIM-domain-only protein, in the mouse cardiovascular system under physiological and pathological conditions. *Gene Expr Patterns.* 2011; 11:384–394. [PubMed: 21601656]
45. Grove BK, Kurer V, Lehner C, Doetschman TC, Perriard JC, et al. A new 185,000-dalton skeletal muscle protein detected by monoclonal antibodies. *J Cell Biol.* 1984; 98:518–524. [PubMed: 6537951]
46. Takada F, Vander Woude DL, Tong HQ, Thompson TG, Watkins SC, et al. Myozenin: an alpha-actinin- and gamma-filamin-binding protein of skeletal muscle Z lines. *Proc Natl Acad Sci U S A.* 2001; 98:1595–1600. [PubMed: 11171996]
47. Fowler VM. Identification and purification of a novel Mr 43,000 tropomyosin-binding protein from human erythrocyte membranes. *J Biol Chem.* 1987; 262:12792–12800. [PubMed: 3624279]
48. Hayes NV, Scott C, Heerkens E, Ohanian V, Maggs AM, et al. Identification of a novel C-terminal variant of beta II spectrin: two isoforms of beta II spectrin have distinct intracellular locations and activities. *J Cell Sci.* 2000; 113(Pt 11):2023–2034. [PubMed: 10806113]

Clinical relevance

Protein isoforms, the same protein varying in amino acid sequence, originate from separate genes or alternate mRNA splicing of a single gene. Protein isoforms can have an impact on biological function and differ in activity, expression, and/or cellular location, and can be alternately expressed with development or under pathological conditions. Myofilament proteins, arranged to form the sarcomere of a myofibril, are the main regulators of cardiac contractility. Several myofilament proteins are known to have isoforms, of which some are expressed only in response to cardiac disease; however, a comprehensive analysis of the myofilament protein isoforms is lacking. This is for a large part due to the challenges of unambiguous identification of isoforms at protein level, which is hampered by high amino acid homology and often low protein expression levels. This study uses a mass spectrometry approach to identify cardiac myofilament protein isoforms based on the observation of a tryptic peptide composed of an amino acid sequence that is unique to a particular isoform. We have identified cardiac myofilament and myofilament-associated proteins with one or more isoforms, and some of these isoforms have not been previously observed. Changes of distribution of these protein isoforms under pathological conditions could ultimately allow for clinical diagnostics or as therapeutic targets.

Figure 1. Experimental workflows

Three workflows are combined to identify the myofilament protein isoforms in cardiac samples from rat. 1DLC'InSeq': myofilament proteins were enriched using the "IN Sequence" method, fractionated using HPLC and in-solution digested. 1DE'InSeq': enrichment using the "IN Sequence" method combined with SDS-PAGE fractionation and in-gel digestion. 1DE'triton': enriched using the Triton X-100, fractionated using SDS-PAGE and in-gel digested. All samples were run on an LTQ-Orbitrap, proteins identified using Mascot, X!Tandem, X!Tandem Kscore, and OMSSA with results merged using PepArML Meta-Search engine, and post-search analysis performed using MASPECTRAS.

Figure 2. Identified cardiac myofilament proteins using the meta-search engine PepArML Clustered proteins identified using PepArML Meta-Search engine are portrayed in a Venn diagram shown in **Figure 2A**. The 1DLC-InSeq workflow resulted in 1198 proteins, 1DE-InSeq in 656 proteins and 1DE-triton 459 proteins (with 1 peptides) (**Supplemental Table 1B**). 461 proteins were identified using two or three workflows (**Supplemental Table 3**). **Figure 3B** illustrates the cellular localization of these proteins using Uniprot Subcellular Location. Uniprot annotated 102 proteins as myofilament or myofilament-associated proteins (**Supplemental Table 4**). Some proteins have multiple location annotations and where therefore counted double. All proteins identified in this Uniprot annotated myofilament list were validated in the literature for direct association with the myofilament. This resulted in 27 “core” myofilament proteins and the identification of 39 isoforms and sub-isoforms (**Table 2**).

Figure 3. Schematic sarcomere overview with myofilament protein localization
The localization of the 27 “core” myofilament proteins listed in **Table 2** are depicted in the sarcomere schematic. Proteins identified with one or more isoforms have been marked (*).

Figure 4. Schematic representation of the isoforms of α -actinin and tropomyosin

A. Isoforms 1, 2 and 4 of α -actinin with the known binding sites to actin, dendrin, Ca^{2+} and PIP_2 marked. **B.** Isoforms 1-1, 1-2 and 3 of tropomyosin with the overlapping sites and the actin binding regions marked. Unique peptide denotes the region where the unique peptide of the isoform was located.

Figure 5. Schematic representation of the isoforms of MLC and MHC

A. Isoforms 2a, 2c, 3, 4 and 6 of MLC with marked EF-hand regions. **B.** Isoforms 4, 6, 7, 8, 9, 13, and 14 of MHC with the ATP-binding sites, IQ-motifs and actin binding sites marked. Unique peptide denotes the region where the isoforms unique peptide was located.

Table 1

Number of clustered proteins (with 1 peptides) identified using multiple search engines.

Search Engine	IDLC'InSeq' 1 peptides	IDE'InSeq' 1 peptides	IDE'triton' 1 peptides
PepArml Combiner	1198	656	459
Xtandem! Kscore	956	496	448
Xtandem!	933	489	424
Mascot	1028	513	428
OMSSA	1084	410	391

Total numbers of unclustered and clustered proteins are listed in Supplemental Table 1. All 1665 unique proteins that were identified across all workflows are listed in Supplemental Table 2B.

Table 2

Cardiac myofilament protein isoforms

Protein Name	Isoform	Protein ID	Unique Isoform Peptide	SC
Actin[28]	Beta-actin-like protein 2	D3ZRN3 (Q562R1)	IWYHTFYNELR	3
Alpha actinin[36]	1	Q9Z1P2 (P12814)	AGTQIENIEEDFRDGLK	1
	2	D3ZCV0 (P35609)	TINEVETQILTR	11
	4	Q9QXQ0 (O43707)	HRPELIEYDK	1
Myosin-binding protein[36]	C (cardiac type)	P56741 (Q14896)	SVFTVEGAKEDEGVYTVTVK	24
Myosin Heavy Chain[36]	4 (2b)	Q29RW1 (Q9Y623)	SELQASLEEAESLEHEEGK	4
	6 (alpha)	P02563 (P13533)	SLNDFTTQR	17
	7	P02564 (P12883)	SEFKLELDDVTSNM[15.99]EQHIK	119
	8	F1M8F6 (P13535)	KHADS[79.96]M[15.99]AELGEQIDNLQR	6
	9	Q62812 (P35579)	AS[79.96]IAALEAK	7
	13	F1M789 (Q9UKX3)	SLSTEIFK	9
	14	Q6URW6 (Q7Z406)	AEELLIQLGR	1
Myosin Light Chain[36]	2a	F1M7K3 (C9JEG4)	LNGTDPEEAILSAFR	13
	2c	P08733 (P10916)	FSKEEIDQM[15.99]FAAFPPDVTGN LDYK	242
	3	P16409 (P08590)	AAPAPAAAPAAPEPERPK	78
	4	P17209 (P12829)	IDFSADQIEEFKEAFSLFDR	3
	6	Q64119 (P60660)	VFDKEGNGTVMGAEIR	7
Titin[36]	1	D4A8Y1 (Q8WZ42-1)	ITQETLRQETEIAASVVVATAK	117
	2	F1M7S9 (Q8WZ42-2)	ATSTATJYVEAR	4
	N2BA	F1MAA0	VKKPEEPPPPPK	19
Tropomyosin[28]	Alpha 1-1	P04692 (P09493)	SIDDLEDELYAQK	194
	Alpha 1-2	P04692-2 (P09493-2)	VLEELHKAEDSLLAADETAAK	1
	Alpha 3	Q63610 (P06753)	MELQEIQLK	14
Troponin C[28]	NE	E9PTA1 (P63316)	NE	29
Troponin I[28]	NE	P23693 (P19429)	NE	28
Troponin T[28]	NI	P50753 (P45379)	NI	193

Accession numbers and protein names have been retrieved from the Uniprot database. The human protein accession number is provided between brackets. Unique peptides for each isoform and sub-isoform have been determined using the Clustal Omega Program [23] across rat, mouse and human Uniprot database along with spectral validation. SC represents the spectra counts associated with the isoform specific unique peptide. NE: different isoforms not expected. NI: different isoforms not identified.

Table 3

Isoforms of cardiac myofilament-associated proteins

Protein Name	Isoform	Protein ID	Unique Isoform Peptide	SC
Actin-binding LIM protein[37]	1	D3ZWB3 (O14639)	AIYDIERPDLITYEPFYTSGYEDK	4
Alpha-crystallin[38]	B chain	P23928 (P02511)	IPADVDPLTITSSSLSDGVLTVNGPR	117
Ankyrin[36]	2	F1LM42 (Q01484)	LIEVGPSGAQFLGPVIVEIPHFAALR	5
Catenin[39]	alpha 1	Q5U302 (P35221)	VLTDAVDDITSIDDFLAVSENHILEDVNK	4
Desmin[40]	NE	P48675 (P17661)	NE	585
F-actin-capping protein(CapZ)[41]	alpha 2	Q3T1K5 (P47755)	FIHAPPGEFNEVFNDVR	8
Filamin[42]	C	D3ZHA0 (Q14315)	VQYTAYEEGVHLVEVLYDEVAVPK	5
Four and a half LIM domains protein (FHL)[43]	2	O35115 (Q14192)	YISFEER	3
Cysteine-rich protein[44]	2	P36201 (P52943)	GVNTGAVGSYIYDKDPEGTVQP	6
	3	P50463 (Q6Q6R5)	NFGPTGIGFGGLTHQVEK	25
Myomesin[45]	1	F1M7T8 (P52179)	FALFDLVEGK	11
Myozenin[46]	2	D4A724 (Q9NPC6)	DIMIEELSHFGNR	10
Nebulin[36]	Nebulette	D4A4G1 (O76041)	RPGIIVDLK	2
Obscurin[36]	NI	F1M8C9 (Q5VST9)	NI	10
Tropomodulin[47]	1	P70567 P28289)	TLENELDELDPDNALLPAGLR	9
Spectrin beta chain [48]	1	F1LRX5 (P11277)	LTTLELLEVR	5
Synaptopodin[32]	2	D4A702 (Q9UMS6)	GKKPLNTLDVMK	9
Transglutaminase[31]	2	Q9WVJ6 (P21980)	LTLYFEGR	13

Accession numbers and protein names have been retrieved from the Uniprot database, and the human protein accession number is provided between brackets. Unique peptides for each isoform and sub-isoform have been determined using the Clustal Omega Program [23] across rat, mouse and human Uniprot database along with spectral validation. SC represents the spectra counts associated with the isoform specific unique peptide. NE: different isoforms not expected. NI: different isoforms not identified.