A PROTEOMICS APPROACH TOWARDS UNDERSTANDING HYPERTROPHIC CARDIOMYOPATHY

By

Wenxuan Cai

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Molecular and Cellular Pharmacology)

At the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 7/25/2018

This dissertation is approved by the following members of the Final Oral Committee:

Ying Ge, PhD, Associate Professor, Department of Cell and Regenerative Biology

Youngsook Lee, PhD, Associate Professor, Department of Cell and Regenerative Biology

Lingjun Li, PhD, Professor, Pharmaceutical Sciences Division

Timothy J. Kamp, MD/PhD, Professor, Department of Cell and Regenerative Biology

J. Carter Ralphe, MD, Associate Professor, Department of Pediatrics

ACKNOWLEDGEMENTS

I find myself extremely fortunate to be able to conduct my thesis research under the guidance of Dr. Ying Ge, who introduced me to the fascinating field of proteomics and mass spectrometry. Dr. Ying Ge has been an extremely supportive and inspiring mentor. It is impossible to count all the ways that she has helped me in my research and professional career. I have always been inspired by her passion for science, her incredible work ethics and extreme kindness. She taught me to be a critical thinker, an effective communicator and a life optimist, which has benefited me during my Ph.D. training, and will continue to benefit me in my future endeavors. Second, I would like to thank Dr. Youngsook Lee, Dr. Lingjun Li, Dr. Timothy J. Kamp and Dr. J. Carter Ralphe for taking the time out of their busy schedule to serve on my thesis committee. Their expertise in cardiology, developmental biology, pharmacology and analytical science has offered me valuable insights in my research efforts. I also want to thank the Molecular and Cellular Pharmacology program. I have been very grateful to be part of this intelligent and friendly community. Third, I could not have come this far without the encouragement and love of my parents, who have been supporting me selflessly throughout my life. Lastly, I would like to thank member of the Ge research group who contributed to the research presented in this dissertation, as well as Madam Enigma, Sir Riddle and Sir Tomas for all the fun and joys they brought to my life.

ABSTRACT

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disorder estimated to affect 1 in 200 people in the population, and is the most common cause of sudden cardiac death in young adults. HCM is an autosomal dominant genetic disease caused mainly by mutations of proteins that constitute the sarcomeres. Although the genetic causes of HCM have been identified, the molecular mechanisms underlying contractile dysfunction, adverse cardiac remodeling and heart failure as a consequence of HCM remains incompletely understood. Post-translational modification (PTM) of the sarcomeric proteins represent important mechanisms in the regulation of contractile function and cardiac signaling, and therefore, contractile abnormalities in HCM may be contributed by alterations in the sarcomeric protein PTMs. In the studies detailed in this thesis, top-down mass spectrometry (MS)-based proteomics was employed to investigate changes in the sarcomeric protein PTMs in HCM using human tissues. The results revealed differential regulation of the sarcomeric protein phosphorylation in HCM. The phosphorylation levels of a number of sarcomeric proteins were down-regulated, whereas the phosphorylation levels of some other proteins were up-regulated in HCM tissues compared to donor tissues. In addition, the changes of phosphorylation were different in closely-related splicing sarcomeric protein isoforms, providing new insights into how different sarcomeric protein isoforms may contribute differently to cardiac function. This is also the first study to compare the HCM tissues from myectomy surgery and end-stage failing hearts. Interestingly, changes in the phosphorylation of a number of sarcomeric proteins in the myectomy tissues were reversed or partially reversed in the failing stage, suggesting important functional changes of the heart at different stages of the disease. Moreover, this study identified new sarcomeric protein PTMs that may represent novel mechanism of cardiac signaling.

Human pluripotent stem cell-derived cardiomyocytes (hPSC-CMs) are promising *in vitro* model for the study of HCM pathogenesis and disease progression; however, the immaturity of hPSC-CMs hinders application of this technology. By employing an integrated MS-based proteomics strategy, the research presented in this thesis provides a robust and reliable method for the benchmarking of hPSC-CM maturity, and identify potential novel markers for evaluating the maturation status of hPSC-CMs. The development of this integrated proteomics method will not only enable the study of HCM *in vitro*, but also facilitate mechanistic studies in the regulation of hPSC-CM development and maturation. Lastly, the development of novel serial size exclusion chromatography strategy for protein separation and comprehensive software tool for top-down MS data analysis have significantly advanced the field of top-down proteomics and pushed the boundary for the application of this technology in biomedical research.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1

INTRODUCTION

1.1 HCM EPIDEMIOLOGY AND CLINICAL PERSPECTIVES

Hypertrophic cardiomyopathy (HCM) is defined as abnormal non-dilated left ventricular (LV) hypertrophy in the absence of systemic diseases that can lead to pressure overload, such as hypertension and aortic valvular stenosis.[1-3] HCM is a common inherited heart disease previously estimated to affect at least 1 in 500 people[4] of both sexes and various ethnicities from all over the world. In contrast to its high prevalence, HCM is often under-recognized in clinical practice, partly due to a wide spectrum of disease presentations, with a large population of patients remaining mildly affected or asymptomatic. With recent advances in population genetic studies and contemporary diagnostic techniques, the prevalence of HCM was recently revised to account for the under-recognized population, and estimated to be at least 1 in 200 in the general population.[5] These data extrapolate to more than 1.5 million people in the U.S. either having HCM or at risk of developing HCM, which establishes HCM as the most common inherited heart disease.

With increasing clinical recognition and improvement in treatment options, the mortality rate of HCM is estimated to be 0.5% annually.[2, 3, 6] However, HCM remains a leading cause of sudden cardiac death in young adults, particularly competitive athletes.[7-9] Furthermore, HCM with or without outflow obstruction may progress to heart failure with preserved systolic function, and even evolve to decompensated ("burned-out phase") heart failure with systolic dysfunction.[2, 6, 10-12] Atrial fibrillation remains another complication of HCM and has been documented in more than 20% of HCM patients.[1, 13] HCM patients with atrial fibrillation have increased risk for progressive heart failure and embolic stroke.[13, 14]

HCM is usually suspected following the onset of symptoms, such as chest pain and dyspnea especially following exercise, a cardiac event or a heart murmur. Positive family history is of great

clinical value.[1, 3] Abnormal electrocardiogram can also be indicative of HCM.[1, 3] Clinical diagnosis of HCM is often established by imaging via 2-dimentional echocardiogram or cardiovascular magnetic resonance imaging (MRI).[1, 3] Imaging technique often reveals various degree of LV hypertrophy (wall thickening), which is a characteristic feature of HCM in many patients. The degree of LV wall thickness varies significantly from patient to patient, ranging from mild (13-15 mm) to massive (> 30 mm) [3, 15]. Importantly, the degree of LV hypertrophy is insufficient to predict the risk of sudden cardiac death.[15] The pattern of LV hypertrophy is also highly heterogeneous.[3] Many patients show diffusely distributed hypertrophy of the LV chamber with involvement of the inter-ventricular septum, whereas a significant subset of patients present with more localized and segmental midwall thickening,[16] such as the apical form. Some patients can remain asymptomatic until later in life at their 40-50s, while some patients experience cardiac insufficiency and arrhythmia very early in life.

At the tissue and cellular level, cardiomyocyte hypertrophy and disarray is a prominent feature of HCM. Normal myocardium is composed of cylindrical-shape (rod-shape) cardiomyocytes that aligned longitudinally, forming highly organized cardiac muscle bundles. In different areas of the HCM myocardium, adjacent cardiomyocytes or muscle bundles may align perpendicularly or obliquely to each other, forming a tangle mass.[17] In addition, multiple areas of the myocardium consist of longitudinally or transversely cut cardiomyocytes or bundles, giving the myocardium a "swirling" appearance.[17] Cardiomyocyte disorganization is extensive in both inter-ventricular septum and the LV free wall.[17] Intramural coronary arteries exhibit abnormal architecture with significantly narrowed lumen,[18, 19] and is thought to contribute to impaired circulation and greater risk of myocardial ischemia.[20] Myocardium injuries and repair in the form

of interstitial fibrosis with increased collagen matrix likely predispose to life-threatening electrical instability, leading to ventricular arrhythmia and cardiac arrest.[3, 11]

1.2. THE GENETIC BASIS OF HCM

HCM is transmitted in an autosomal dominant pattern, although *de novo* mutations do occur. The first reported HCM-causing variant (*MYH7* Arg403Gln encoding β-myosin heavy chain (β-MHC)) was described in 1990,[21] and since then different HCM-associated variants have been identified in multiple genes. Recent advances in genetic testing and high-throughput genome-sequencing technology have accelerated the identification of rare genetic variants, with over 1400 variants being identified in at least 11 genes associated with HCM (with strong evidence of pathogenicity **Table 1.1**).[1, 22] However, the evidence supporting the pathogenicity of each variant varies.

The most powerful and robust evidence for pathogenicity of a variant is often derived from genetic linkage analysis in large families.[22] Co-segregation of the genetic variant with afflicted individuals in the family, absence of the genetic variant in non-afflicted members, and absence of the genetic variant in unrelated and ethnic-matched normal controls, establish strong causality of such genetic variant with HCM.[22, 23] The most well-established disease-causing variants include the missense mutations on *MYH7* (encoding β-MHC) and frame-shifting mutations on *MYBPC3* (encoding cardiac myosin binding protein C (cMyBP-C)). However, due to the issues associated with incomplete penetrance and variable expressivity, genetic linkage studies are often difficult with variants that are relatively rare and less penetrant. Therefore, in addition to co-segregation of genetic variant with phenotype, a variant can be considered pathogenic (or likely pathogenic) if

identified or reported as a cause of HCM (in more than 1 patient); variation occurs in highly conserved residues across species; and/or protein structure and function is critically altered as a consequence of the variant.[22] Based on these criteria, a summary of the current knowledge of genetic causes of HCM is provided below (**Table 1.1**).[22]

Table 1.1 Genetic causes of HCM and evidence of pathogenicity

Gene	Protein	Prevalence[1]	Sarcomere Component or other function	Evidence of pathogenicity[22]
MYH7	β-myosin heavy chain (β-MHC)	35%	Thick filament	Strong
MYBPC3	Cardiac myosin binding protein C (c-MyBP-C)	35%	Thick filament	Strong
TPM1	α-tropomyosin (α-Tpm)	5%	Thin filament	Strong
TNNT2	Cardiac troponin T (cTnT)	5%	Thin filament	Strong
ACTC1	Cardiac α-actin (cα-actin)	1-5%	Thin filament	Strong
MYL2	Myosin light chain 2 (ventricular isoform) (MLC2)	1-5%	Thick filament	Strong
MYL3	Myosin light chain 1 (ventricular isoform) (MLC1)	1-5%	Thick filament	Strong
TNNI3	Cardiac troponin I (cTnI)	1-5%	Thin filament	Strong
TNNC1	Troponin C (TnC)	1-5%	Thin filament	Strong
ACTN2	α-actinin 2	1%	Z-disc	Strong
MYOZ2	Myozenin 2 (MYOZ2)	1%	Z-disc	Strong
MYH6	α-myosin heavy chain (α-MHC)	<1%	Thick filament	Lesser
TTN	Titin	<1%	Thick filament/Z-disc	Lesser
CSRP3	Muscle LIM protein (MLP)	<1 %	Z-disc	Lesser
TCAP	Telethonin	<1%	Z-disc	Lesser
VCL	Vinculin	<1%	Intercalated disc/Z-disc	Lesser
CASQ2	Calsequestrin (CASQ2)	<1%	Calcium handling	Lesser
JPH2	Junctophilin 2	<1%	Calcium handling	Lesser

Importantly, despite the rapid development and application of genetic testing in the clinic, there are still about 50%[1, 22] of the HCM patients having no mutations/variants found in the genes summarized in Table 1.1. Of the patients with identified genetic mutations, over 70% of the cases have mutations in two genes: *MYH7* and *MYBPC3*,[1, 22] encoding β-MHC and cMyBP-C, respectively. The remaining mutations with strong probability of pathogenicity were found in other genes (**Table 1.1**),[1, 22] majority of which encode the protein constituents of the sarcomeres.

Sarcomeres are the basic contractile apparatus in the myofibrils, composed of the actin-based thin filaments and the myosin-based thick filaments interconnected by complex protein structures known as the Z-discs (**Figure 1.1**).

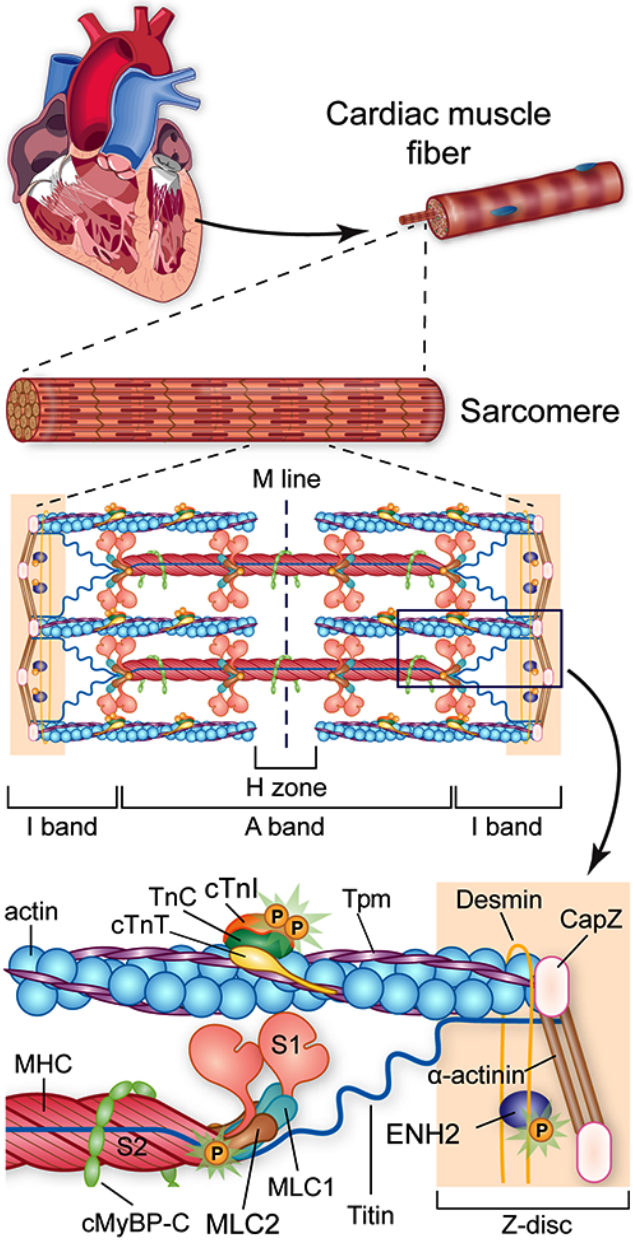


Figure 1.1: Schematic of the cardiac sarcomere composed of myofilaments (thin and thick filaments) and Z-discs. Only relevant proteins are shown in this diagram.[24] (adapted from previous work published by Peng et al.[24] with permission) (Copyright © 2014, The American Society for Biochemistry and Molecular Biology)

Diverse genetic causes partially explain the heterogeneous clinical expression of HCM. Some patients remain asymptomatic until later in life (at their 40-50s), while some experience cardiac insufficiency and arrhythmia at very young age.[25] In general, patients with mutations in *MYBPC3* (cMyBP-C) present with milder symptoms that do not develop until elderly age, while patients with mutations in *MYH7* (β-MHC) have earlier disease onset and worse symptoms.[25] Individuals with *TNNT2* (encoding cTnT) mutations typically present with milder hypertrophy; however, with an increased risk of sudden cardiac death.[26] However, over the past 30 years, it is increasingly recognized that mutation types are not reliable prognostic indicators for HCM.[27] Some mutations previously thought to be associated with early onset and sudden cardiac death, were later shown in individuals with very mild symptoms and normal life span.[27] Nevertheless, it is well recognized that patients carrying more than one HCM-causing mutations (compound heterozygous or homozygous) can present with severe hypertrophy with early onset and greater risk of sudden death, and the condition may rapidly progress to decompensated heart failure.[25, 26, 28, 29] Since patients with the same mutations may have drastic different age of disease onset, degree of hypertrophy and course of disease progression, there exist genetic and environmental modifiers which are yet to be defined, but encouraging for harnessing these "protective factors" for disease prevention and therapy.

HCM can be considered as a disease of the sarcomeres. However, the molecular mechanisms by which sarcomeric protein mutations lead to cardiac remodeling and hypertrophy, and eventually heart failure, remains incompletely understood.

1.3 DISEASE PATHWAYS IN HCM

1.3.1 Increased calcium sensitivity of force generation and/or hyper-contractility of myosin

Calcium-dependent force generation is a fundamental property of the cardiac sarcomeres. During systole, cell membrane depolarization upon electrical stimulation results in an increase in intracellular calcium though L-type calcium channel on plasma membrane, triggering subsequent calcium-dependent calcium release from sarcoplasmic reticulum (SR).[30-33] The cardiac troponin complex (TnC, cTnI and cTnT), together with α-Tpm are important regulators of the thin filaments. Upon rapid increase in intracellular calcium during systole, TnC binds to calcium and undergoes conformational change, leading to subsequent change of cTnI conformation and the release of the cTnI inhibitory region from the actin filament. In addition, weakening of the cTnT and α-Tpm interaction causes a change of α-Tpm positioning relative to the actin filament, and exposes the myosin-binding sites on the thin-filament.[30-33] In the presence of ATP, myosin binds to actin, resulting in sliding of the thick filament along the thin filament, which leads to muscle contraction. Crossbridge cycling continues until intracellular calcium level decreases in diastole, which lead to dissociation of calcium from TnC, subsequently blocking actin-myosin interaction and causing relaxation.[30-33] The power output and velocity of crossbridging is also regulated by cMyBP-C.[30-33]

Mutations in the troponin complex account for up to 15% of all the disease-causing mutations with a total of over 100 mutations found in the 3 components of the troponin complex.[34] Out of 6 common cTnC mutations, 4 of them (Ala8Val[35], Ala31Ser[36], Cys84Tyr[35], Asp145Glu[35]) have been consistently demonstrated to increase calcium sensitivity of force development.[34-37] In addition, Ala31Ser and Asp145Glu mutations on TnC were also shown to increase ATPase activity in the presence of calcium.[35, 36] TnC Leu29Gln mutation is perhaps the most extensively studied

mutation; however, the effects of Leu29Gln on the calcium sensitivity of force and ATPase activity remain conflicting.[37]

Of the many HCM-associated mutations on cTnI, at least five mutations: Arg145Gly, Arg145Gln, Arg162Trp, ΔLys183 and Lys206Gln, were shown to increase calcium sensitivity of myofibril ATPase activity in reconstituted skinned muscle fibers.[38] In addition, Arg145Gly, Arg145Gln and Arg162Trp were shown to increase the maximal ATPase activity in cardiac myofibrils.[38] Gly203Ser mutation on cTnI were shown to have a slight (but insignificant) tendency of increasing the calcium sensitivity of both ATPase activity and force generation.[38] However, Gly203Ser mutation on cTnI significantly increases relative minimal force in the absence of calcium (relaxing state),[38] similar to the case of Arg145Gly, Arg145Gln and Arg162Trp,[38] suggesting that Gly203Ser mutation of cTnI (in addition to Arg145Gly, Arg145Gln and Arg162Trp) impairs the ability of cTnI to inhibit force generation during relaxation. Extensive studies have shown that HCM-causing mutations of cTnT (Lys280Asn[39], Arg92Gln[39]/Trp[40]/Leu[40], Ser179Phe[39, 40], Arg94Leu[40], Glu163Arg[40]/Lys[41], Arg130Cys[40], Ala104Vla[40], Ile97Asn[41], Arg273Glu[39]) increased the calcium sensitivity of force development in reconstituted skinned cardiac muscle fibers with the exception of Glu244Asp[40]. However, Glu244Asp mutation lead to increased maximal force development.[40]

The studies of HCM-causing α-Tpm mutations showed that Ala63Vla and Lys70Thr both affect the flexibility of the α-Tpm N-terminal region, and subsequently increase calcium sensitivity of thin-filament sliding speed and calcium-dependent ATPase activity.[42, 43] This is consistent with another report suggesting that Ala63Vla, Lys70Thr and Glu180Gly increase calcium sensitivity of force generation.[44] One the other hand, Asp175Asn causes a slight increase in the calcium sensitivity of force generation, albeit not statistically significant. Moreover, the degree of calcium

sensitization due to α-Tpm mutation appears to correlate with the degree of clinical severity of the disease.[44]

cMyBP-C is a protein of the immunoglobulin superfamily and is associated with the thick-filament.[45-47] The C-terminal region of cMyBP-C interacts with the light meromyosin region of the myosin rod and titin, contributing to the thick-filament structure.[45-47] cMyBP-C also plays important roles in the regulation of contraction by acting as a "break" of actomyosin ATPase activity through the interaction between its N-terminal region and the S2 domain of the myosin molecule.[45-47] This molecular interaction reduces myosin ATPase activity and calcium sensitivity.[45-47] Interruption of cMyBP-C and S2 interaction lead to increase in calcium sensitivity, contractile force and time to half relaxation.[48, 49] Partial extraction of cMyBP-C from the muscle fibers lead to an increase in calcium sensitivity and shortening velocity.[50, 51] Phosphorylation of cMyBP-C at its N-terminus resulted in release of the S2 domain, and subsequently increased crossbridge cycling rate.[52] Transgenic mice with homozygous *MYBPC3* knock-out exhibit accelerated kinetics of contraction, increased calcium sensitivity of force generation and enhanced power output.[53] All these data supports that the loss of cMyBP-C can result in increase in calcium sensitivity, shortening velocity and myosin ATPase activity. Moreover, impaired relaxation is characteristic to loss-of-function mutations on cMyBP-C. Approximately two third of the HCM-causing mutations on cMyBP-C occur in the splicing donor/receptor of the *MYBPC3* gene locus, or introduce frameshift deletion/insertion in the coding sequence, resulting in a null-allele and haploinsufficiency.[46, 54] A number of mutations result in the direct introduction of stop codon.[46] The remaining HCM-causing mutations are missense mutations with single amino acid substitutions, which either leads to protein instability or mis-localization.[46, 55] Taking these

together, HCM-causing mutations in cMyBP-C predominantly causing the loss of cMyBP-C ("myosin ATPase break"), resulting in the "hyper-contractile" phenotypes underlying HCM.

Mutations on the β-MHC are the largely missense mutations with single amino acid substitution. Most of the published work supports that HCM-causing mutations on β-MHC increase contractility.[56] Arg403Gln mutation in mouse α-MHC (the predominant MHC isoform in mouse cardiac ventricles whereas β-MHC is the predominant isoform in human cardiac ventricles) increases sliding velocity, ATPase activity and crossbridge cycling kinetics.[57] Arg403Gln in human β-MHC was shown to increase sliding velocity by *in vitro* motility assay.[58] Arg719Trp mutation in human β-MHC increases ATPase activity and force generation.[59, 60] Arg723Gly leads to increased myosin sliding velocity and force generation. In addition, Asp778Gly[61], Asp778Val[62], Asp906Gly[62], Leu908Vla[58] have been shown to cause increased sliding velocity. Moreover, HCM-causing mutations in β-MHC generally increase the overall contractile force and velocity independent of calcium, which may explain in part the inability of β-blockers and calcium channel blockers to attenuate the progression of disease.[34] Each myosin molecule contains two MHC molecules with two essential and two regulatory light chain molecules, also known as myosin light chain 1 and 2, respectively (MLC1 and MLC2). Mutations in the MLC1/2 accounts for about 5% of all HCM-causing mutations. MLC2 mutation Arg58Gln was shown to increase calcium sensitivity of ATPase activity and steady-state force.[63] In addition, both Arg58Gln and Asn47Lys in MLC2 were shown to prolong calcium transient in intact mouse papillary muscle.[63] Consistently, MLC2 Asp166Vla mutation causes prolonged relaxation and increased resting force.[64]

In the past 20 years, many research studies have converged to the fact that increased calcium sensitivity and/or increased force generation (with impaired relaxation) is a nearly

universal characteristic of HCM, and the resulting hyper-contractility state is the underlying cause of subsequent cardiac remodeling and HCM phenotypic expression.

1.3.2 Impaired energetics of the HCM hearts.

As a result of increased calcium sensitivity and/or increased force generation, the crossbridge cycling rate increases and actin-activated ATPase activity is also enhanced.[65] Consequently, there is an increase of tension cost, which is the cost of energy to generate a given tension.[65-67] Therefore, the HCM hearts were expected to show reduced efficiency in energy utilization and increased energy demand.[66] Consistent with this rationale, using 31P NMR spectrometry technology, Jung WI et al. reported a reduction in the phosphocreatine-to-ATP ratio, a measurement of intracellular energy stock.[68] Furthermore, Grilley JG et al. reported consistent energy deficiency in HCM patients, regardless of the degree of hypertrophy.[69] Mitochondrial abnormalities manifest as inflated and swelling mitochondria with disorganized cristae were also observed in those HCM patients with abnormal force-frequency relationship.[70] Importantly, energy depletion is likely to affect many important cellular functions, particularly the maintenance of cardiomyocyte membrane potential, which is mediated by various ion transporter which are ATPases.

Impaired cellular energetics is certainly a consequence of hyper-contractility and increased calcium sensitivity due to the HCM-causing mutations. However, cellular energy deficiency can cause a serious of cellular dysfunction, and Ashrafian H et al. suggested that cellular energy depletion underlies the subsequent cardiac remodeling and hypertrophy.[66] Evidence supporting the central role of impaired energetics in HCM is derived from the observations that many diseases

caused by deficiency of myocardial energy production lead to asymmetric cardiac hypertrophy that can be clinically indistinguishable from HCM.[66] For instance, mutation in the mitochondrial tRNA causes cardiac phenocopy of HCM that is maternally inherited.[71] Mutations on the CD36 or very-long-chain acyl-CoA dehydrogenase, which affects fatty acid uptake and beta-oxidation, respectively, can also cause asymmetric cardiac hypertrophy.[72, 73] Moreover, mutations in AMP-activated protein kinase (AMPK), a critical sensor and master regulator of cellular energetics, result in Wolff–Parkinson–White syndrome and progressive conduction disease that resemble characteristic HCM phenotypes.[74, 75] Therefore, modification of myocardial energy metabolism presents an attractive target for attenuating disease progression and altering the natural course of HCM.

1.3.3. Abnormal calcium handling and activation of calcium-mediated hypertrophy pathways

Calcium handling is one of the most extensively studied aspects of HCM, and many studies have found abnormal calcium handling that links to HCM pathogenesis and arrhythmogenesis. In a mouse model of heterozygous Arg403Gln mutation on α-MHC (the analog of β-MHC in human), Fatkin D et al. observed abnormal calcium response,[76] which was likely a result of reduced SR calcium store. Further study in the same model confirmed reduction of SR calcium storage, and uncovered the down-regulation of Calsequestrin 2 (CASQ2),[77] which is a major SR calcium sequestering protein. In addition, there was significant reduction in the expression of Ryanodine receptor 2 (RyR2), Triadin and Junctin,[77] which interact with CASQ2 to regulate calcium release. Importantly, abnormality in calcium handling and alterations in the expression of the calcium handling proteins precede the development of cardiac hypertrophy,[77] indicating essential roles of

calcium dysregulation in HCM pathogenesis. In a transgenic mouse model of cTnT mutation Arg92Trp, Guinto PJ et al. reported altered SERCA2a-to-PLN (SR calcium ATPase 2a-to-phospholamban) ratio, and increased sodium-calcium exchanger expression, as well as decrease in baseline cTnI phosphorylation at 6-month of age.[78]

Consistently, in a model of HCM using human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-cardiomyocytes), Lan et al. reported that hiPSC-cardiomyocytes carrying the Arg663His mutation in β-MHC showed decreased SR calcium storage, increased cytosolic calcium concentration, and increased proportion of cells that exhibit abnormal calcium transient.[79] Furthermore, by subjecting the control hiPSC-cardiomyocytes to high calcium containing media, elevated intracellular calcium was observed and was sufficient to induce both hypertrophy and arrhythmia.[79] Therefore, the authors concluded that elevated cytosolic calcium concentration as a result of Arg663His mutation directly contributes to cardiac hypertrophy and arrhythmogenesis.[79]

Abnormal intracellular calcium concentration likely lead to dysregulation of a number of calcium-mediated signaling pathways, including the calcium/calmodulin-activated protein kinase II (CAMK II), calcium/calmodulin-calcineurin, and the protein kinase C pathways.[30] Calcineurin is also known as protein phosphatase 2B and contains two subunits, CnA and CnB. Calcineurin exhibit higher binding affinity towards calcium/calmodulin than CAMK II, and is believed to be more sensitive to small sustained increase in calcium concentration.[30, 80] Activation of calcineurin leads to dephosphorylation of nuclear factor of activated T-cells (NFAT) and the subsequent nuclear-translation of NFAT for activation of genes involved in hypertrophic response.[81] Extensive studies have consistently demonstrated the central role of the calcineurin-NFAT pathway in hypertrophy signaling.[81] Inhibition of calcineurin prevents development of hypertrophy induced by angiotensin II, pressure overload or isoproterenol infusion.[81-83] Moreover, several kinases such

as MAPK p38 and ERKs (p42/44) were shown to regulate cardiac hypertrophy response by phosphorylating NFAT.[84] Another important axis of calcium-mediated signaling is the CAMK II pathway, which is thought to respond to high-amplitude or frequent calcium oscillations.[30] Specifically, CAMK II mediates cardiac hypertrophic response through phosphorylation of class II histone deacetylases (HDAC), leading to consequential activation of MEF2.[85-87]

Interestingly, moderate exercise has been shown to prevent and reverse HCM in a mouse model with a β-MHC mutation.[88] The beneficial effects of exercise were partially due to a normalization of the phosphorylation level of glycogen synthase kinase-3β (GSK-3β) and CREB (a transcription factor).[88] A recent randomized clinical trial enrolling a total of 136 HCM patients also confirmed the benefits of moderate exercise for the improvement of exercise capacity.[89] The protective effects of exercise is throught to be mediated through PI3K-Akt-GSK3-mTOR signaling,[32, 90] and pharmacological activation of mTOR by rapamycin was shown to mitigate LV hypertrophy in rodents.[91] These data demonstrate the great promise of harnessing the power of exercise and exercise-induced signaling for the treatment of HCM.

1.4 PHARMACOLOGICAL MANAGEMENT OF HCM

The current medical intervention of HCM is limited to management of symptoms and supportive care, with the major goals of providing symptomatic relief, controlling anginal symptoms, limiting left ventricular outflow obstruction and heart failure, as well as preventing sudden cardiac death.[1, 92]

β-blockers (β-adrenergic receptor blockers) remain the first-line treatment agents for HCM since the first description of their beneficial effects in 1966.[93] At least 5 β-blockers (propanolol, sotalol, metipranolol, nadolol, bisoprolol) were studied in 12 trials with a total of 391 subjects.[92] β-blockers have been shown to be effective for reducing exertional angina or dyspnea, and are recommended to reduce the prevalence of non-sustained ventricular arrhythmias.[2, 93, 94] Sympathetic modulation by β-blockers reduces ventricular contractility and maximal contraction velocity, leading to subsequent reduction in outflow tract obstruction.[10, 92, 95-97] In addition, reduction in heart rate may increase time for diastolic filling, leading to improved coronary perfusion.[10, 92, 95-97] One retrospective study suggested the benefits of sotalol in reducing the risk of ventricular arrhythmia and sudden cardiac death.[98] Despite clear evidence of symptomatic relief with the use of β-blockers,[10, 92, 95-97] the benefits of long-term β-blocker treatments on HCM patients have not been convincingly established.[3]

Non-dihydropyridine calcium channel blockers, including verapamil and diltiazem, are another class of pharmacological agents used in the management of non-obstructive HCM. Compared to the dihydropyridine class, which is more vascular selective with less cardiac effects, non-dihydropyridine calcium channel blockers are more potent cardiac depressants with less vasodilating effects.[99] The negative inotropic and chronotropic effects of the non-dihydropyridine calcium channel blockers are thought to benefit the HCM patients by prolonging LV filling time.[10, 100, 101] Specifically, verapamil has been extensively studied in 12 trials with a collective enrollment of 367 patients and has been shown to improve exercise capacity.[92] Even though verapamil has been the standard treatment for non-obstructive HCM, there has been no convincing evidence suggesting functional improvement of the hearts.[2, 94, 97] Similar to verapamil, diltiazem causes strong depression in AV node conduction, leading to decreased heart rate, in addition to its

vasodilating effects that reduce peripheral resistance and afterload.[92, 102-104] Diltiazem has been shown to improve LV diastolic parameters,[102, 103] and one study demonstrated mitigation of myocardial ischemia with diltiazem[104]. However, only a small number of patients (a total of 55 patients in 4 trials) have been evaluated. Interestingly, Semsarian C et al. demonstrated prevention of cardiac hypertrophy in mice with diltiazem,[77] leading to an on-going trial to evaluate the use of diltiazem for HCM prevention in genotype-positive and phenotype-negative individuals (ClinicalTrials.gov Identifier: NCT00319982). Nevertheless, calcium channel blockers are not recommended for patients with significant LV outflow obstruction due to the potential of critical adverse hemodynamic effects and worsening of obstructive symptoms.[92]

Disopyramide is a Class IA antiarrhythmic that is usually used in combination with beta-blockers if the first-line treatment is not effective. Disopyramide has been demonstrated to effectively reduce pressure gradient and drag forces, and subsequently improve obstructive symptoms.[105, 106] However, side effects of disopyramide, including prolongation of QT intervals and anticholinergic effects, remain the major concerns.[107] Thus, disopyramide is often used short-term between surgical myectomy in symptomatic patients with severe outflow obstruction.[106] Nevertheless, studies have shown that cibenzoline may be used with improved diastolic function with less anticholinergic adverse effects,[108-110] meriting further investigation regarding its efficacy in the treatment of HCM. On the other hand, amiodarone, as the most effective and widely used treatment for atrial fibrillation, has been proven inadequate for preventing sudden cardiac death in HCM.[92] Occasionally, amiodarone was used to control ventricular arrhythmia in patients with an implantable defibrillator.[92]

Up-regulation of aldosterone has been shown to contribute to adverse cardiac and vascular remodeling in a rodent model of HCM,[111] indicating an important role of the renin-angiotensin-

aldosterone (RAAS) pathway that contributes to interstitial fibrosis, which is a characteristic feature of HCM. Pre-clinical studies using aldosterone inhibitor (spironolactone) or angiotensin II inhibitor showed promising results in the reduction of fibrosis and myocyte disarray with an improvement in the diastolic function.[111, 112] In a small group of 20 HCM patients, Kyriakidis M et al. reported the beneficial effects of combined enalaprilat and captopril, both of which are angiotensin-converting enzyme (ACE) inhibitors.[113] Valsartan is an angiotensin receptor blocker, and has been shown to decrease type I collagen synthesis in HCM patients in a small group of 23 HCM patients.[114] Losartan, another angiotensin receptor blocker, was shown to improve diastolic function in a group of 30 patients;[115] however, failed to improve myocardial function and exercise capacity in a trial with 133 patients.[116] Therefore, large prospective, double-blinded, placebo-controlled trial is required to draw definitive conclusions regarding the efficacy of RAAS inhibitors for the treatment of HCM.

Adverse microvascular remodeling results in impaired tissue perfusion and myocardial ischemia, and impaired myocardial blood flow is a strong independent predictor of cardiovascular mortality and adverse LV remodeling.[18, 20] Studies employing very-low-dose perindopril (ACE inhibitor) and indapamide (diuretic) have demonstrated improvement in coronary circulation in hypertension-induced cardiac hypertrophy,[117, 118] which shed light on the potential use of antihypertensive drugs to mitigate myocardial ischemia in HCM. Due to an apparent energy deficiency in the HCM hearts,[32] the use of perhexiline, a metabolic modulator that inhibits fatty acid metabolism, showed promising preliminary results in the improvement of diastolic function and exercise capacity.[119] Mavacamten (previously known as MYK-461), is a small molecular modulator that decreases myosin ATPase activity, and when given prophylactically, has been shown to prevent hypertrophy, cardiomyocyte disarray and fibrosis in a mouse model of HCM.[120]

Mavacamten is currently under-investigation in a Phase 3 clinical trial for symptomatic obstructive HCM (ClinicalTrials.gov Identifier: NCT03470545), and in a Phase 2 trial for symptomatic non-obstructive HCM (ClinicalTrials.gov Identifier: NCT03442764).

1.5 HUMAN PLURIPOTENT STEM CELL-DERIVED CARDIOMYOCYTES FOR HCM STUDIES

In addition to extensive *in vitro* biochemical and biophysical studies with HCM mutant proteins, many animal models, most notably rodent models, were generated for the investigation of HCM pathophysiology.[121] Many rodent models developed characteristic HCM features observed in the human patients, including ventricular hypertrophy, cardiomyocyte disarray, interstitial fibrosis and diastolic dysfunction.[121] These animal models not only provided strong evidence for the pathogenicity of the individual mutations, but also offered meaningful insights into the disease mechanisms described earlier in this chapter. Moreover, the rodent models of HCM also enabled drug screening and discoveries in a relatively cost-effective manner.[77, 112, 120]

Nevertheless, millions of years of divergent evolution resulted in significant difference between mice and humans in terms of contractile function and cardiac physiology. In particular, 1% of the human genome has no identifiable mouse homologs, and there exist obvious species-specific differences in gestation period, morphology, organ development and regulation of gene expression.[122] Consequently, even though mouse models of human heart diseases may recapitulate the disease phenotypes to some extent, not all aspects of the disease characteristics manifest in mouse models.[123] For instant, knocking-out phospholamban was shown to improve mouse cardiac contractility and rescue HCM phenotypes without affecting the life span of the phospholamban-

null mice.[124] However, null-mutations of phospholamban is detrimental to human, resulting in severe life-threatening dilated cardiomyopathy demanding for heart transplantation.[124] In addition, dystrophin null mutations lead to Duchenne Muscular Dystrophy that manifests with progressive muscle weakness soon after birth with 90% of the patients developing cardiac insufficiency.[125] Patients of Duchenne Muscular Dystrophy have a life expectancy of slightly over 20 years.[125] However, mice deficient of dystrophin exhibit mild cardiac and skeletal muscle dysfunction with a normal life span, due to the compensatory expression of utrophin.[125-127] Rodent hearts beat at a rate of 300-800 bpm, and therefore, rodent hearts have adapted to function with rapid systolic contraction and diastolic filling, which requires protein expression and regulation that are significantly different from human.[123] One of the most noted differences between mouse and human heart is that non-failing mouse ventricular cardiomyocytes express over 94% α-MHC,[123] while non-failing human ventricular cardiomyocytes expresses over 90% β-MHC.[128] Other small and large animal models were generated for the study of heart diseases, but significant challenges remains regarding transgenic manipulation of large animals and costs associated with maintenance of the animals.[123] Due to the differences between animal and human cardiac physiology, pharmacological studies in animal models remain difficult to translate to human therapy. For example, the statin class of compounds were shown to reduce cardiac hypertrophy and fibrosis in a rabbit model of HCM;[129, 130] yet clinical trial results in HCM patients were largely disappointing.[131]

Direct analysis of human cardiomyocytes is critical for the dissection the disease mechanisms and development of effective therapies, but human primary cardiomyocytes are a limited resource that cannot be maintained and expanded indefinitely in culture. Empowered by the recent advances in stem cell technologies, hiPSCs derived from somatic cells present highly

clinically relevant models for investigating the molecular mechanisms in HCM pathogenesis on an individual basis, as well as developing new patient-specific therapies. hiPSC-cardiomyocytes have been used to model a variety of cardiac disorders, including LEOPARD syndrome,[132] Duchenne Muscular D ystrophy,[133, 134] dilated cardiomyopathy,[135] long QT syndrome,[136] and HCM.[79] These hiPSC-cardiomyocyte models provided important molecular insights into the pathogenesis of a variety of cardiac diseases. Nevertheless, hiPSC-cardiomyocytes display the characteristics of fetal cardiomyocytes, limiting their applications, especially in the faithful recapitulation of disease phenotypes and pharmacological studies.[137, 138] Future efforts in promoting hiPSC-cardiomyocyte maturation to acquire the features of adult cardiomyocytes is critical to realize the full potential of this technology for the study and treatment of HCM and other cardiac disorders.

1.6 TOP-DOWN PROTEOMICS: TECHNOLOGY ADVANCEMENTS AND APPLICATIONS TO HEART DISEASES

Diseases of the heart are a leading cause of morbidity and mortality worldwide, and are associated with significant economic costs.[139-142] In 2011, over 600,000 deaths in the United States were caused by diseases of the heart, and the annual economic cost of these diseases in the United States is estimated to be roughly $215.6 billion.[139] Nevertheless, despite substantial effort over the past decades, the molecular mechanisms underlying diseases of the heart remain elusive. As a consequence, current therapeutic approaches for treating patients with heart disease focus primarily on symptom management and cannot alter disease progression. Thus, there is an urgent

need for new methods enabling delineation of the molecular mechanisms underlying diseases of the heart for the purposes of identifying novel disease biomarkers and therapeutic targets.

The field of systems biology aims to obtain a holistic, global, and integrated understanding of the molecular constituents of biological systems, including nucleic acids, proteins, and metabolites, as well as their interactions, and, thus, holds great promise for gaining mechanistic insights into a variety of diseases, including diseases of the heart. [143, 144] Following sequencing of the human genome, it became clear that genomic information alone was insufficient to understand biological functions. Therefore, in the post-genomic era, proteomics, which focuses on the analysis of proteins, the primary effectors of all cellular functions, has come to the forefront of systems biology.[145] In contrast to the genome, which is considered to be largely static, the proteome is highly complex and exhibits considerable plasticity owing to a myriad proteoforms—a term encompassing the plethora of protein species produced by a single gene via the alternative splicing of mRNA transcripts and protein post-translational modifications (PTMs)[146]—and highly homologous protein isoforms produced by multiple homologous genes (**Figure 1.2A**).[144, 147] In particular, protein PTMs play critical roles in the regulation of protein localization, activity, and degradation; and can be altered in disease.[148, 149] Moreover, the differential expression of protein isoforms is increasingly recognized as an important molecular mechanism contributing to disease.[150-152]

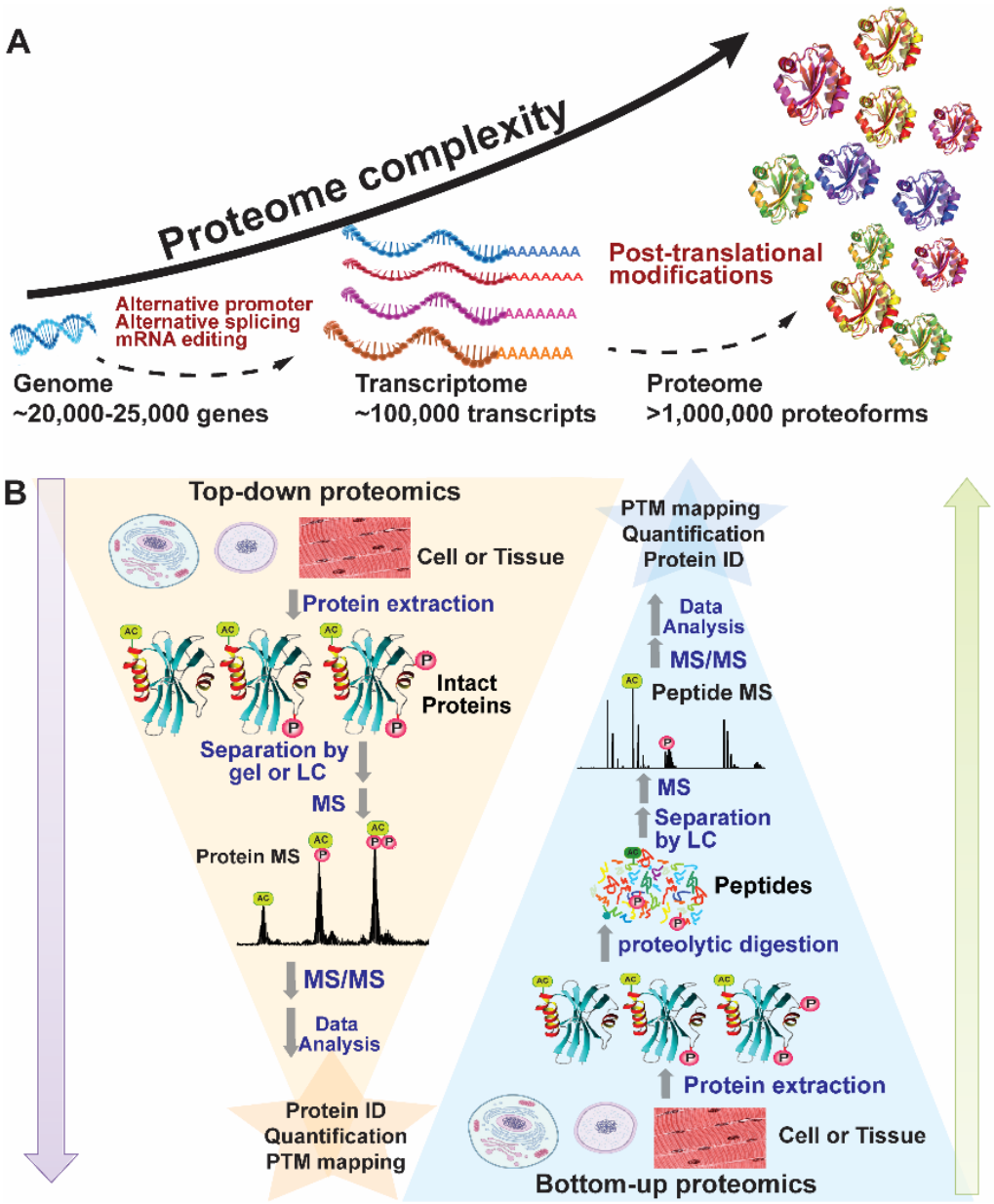


Figure 1.2. A. Exponential increase of proteome complexity owing to alternative splicing of mRNAs and post-translational modifications of proteins. B. Schematic depicting the basic workflow for top-down proteomics as compared with that of bottom-up proteomics. Adapted from ref[153] originally published in Pflügers Archiv European Journal of Physiology, with permission from Springer and authors. © Springer

Mass spectrometry (MS)-based proteomics has quickly become the method of choice for the quantitative analysis of proteoforms and cognate isoforms produced by different genes.[147] MS-based proteomics analysis has traditionally been carried out using the "bottom-up" approach,

wherein protein samples are digested into small peptides, separated using liquid chromatography (LC), and analyzed by MS (**Figure 1.2B**). While bottom-up proteomics is widely-used and enables the identification and quantification of thousands of proteins, this approach suffers from intrinsic limitations resulting from protein digestion, which makes it sub-optimal for the identification of protein isoforms (particularly those with high sequence homology), polymorphisms/mutations, and PTMs.[154] On the other hand, an emerging "top-down" proteomics approach, in which intact proteins are analyzed rather than peptides, provides a global or "bird's eye" view of all proteoforms and cognate isoforms produced by different genes and, thus, represents a superior approach for the comprehensive analysis of proteins and their post-translationally modified forms.[147, 154-158] Following intact protein MS analysis, specific proteoforms/isoforms of interest can be isolated in the gas phase and subjected to tandem MS (MS/MS) analysis to obtain sequence information that can be used to identify proteins and localize PTMs and sequence variations. Furthermore, top-down MS is semi-quantitative, permitting the relative quantification of proteoforms present within the same spectrum.[157, 159, 160]

Here, we will review recent advancements in the top-down proteomics methodology, including developments in sample preparation, protein separation, MS instrumentation, protein characterization, and quantification, as well as tools for top-down data analysis. Moreover, we will summarize recent applications of top-down proteomics to study the cardiac proteome towards a better understanding of the molecular mechanisms underlying diseases of the heart.

1.6.1 Top-down Proteomics Methodology

1.6.1a Sample preparation for top-down proteomics

Traditional sample preparation methods in biomedical research typically involve the use of high concentrations of salt (>100 mM) and/or surfactants, such as SDS or Triton X-100, for the efficient solubilization of proteins, particularly highly hydrophobic proteins.[147, 153, 155, 161, 162] However, most of the commonly used salts and surfactants are incompatible with protein analysis by MS as salt inhibits electrospray ionization (ESI) and surfactants can out-compete proteins for ionization and/or form multiply adducted species that are detrimental to the MS signal intensity of the proteins.[147, 153, 155, 163] Therefore, it is critical to remove MS-incompatible salts and surfactants, or replace them with MS-compatible ones, before MS analysis.

Protein samples containing high concentrations of MS-incompatible salts can be desalted using reverse phase chromatography (RPC).[156, 157, 164] Another option is to replace non-volatile MS-incompatible salts with volatile salts (e.g., ammonium acetate), which are compatible with MS analysis. The exchange of MS-incompatible for MS-compatible salts can be achieved either by substituting the compatible salt for the incompatible salts in the extraction buffer, or post-extraction using dialysis or ultrafiltration.[155, 165, 166] In addition to commonly employed ammonium salts, it was recently shown that ammonium tartrate is also compatible with MS. Xiu et al. demonstrated high-resolution separation of intact proteins using hydrophobic interaction chromatography (HIC) employing mobile phases containing ammonium tartrate.[167]

Surfactants can be generally categorized into ionic and non-ionic; with the former being extremely incompatible with MS due to severe suppression of protein ionization. Conversely, non-ionic surfactants are compatible with MS at low concentrations.[166] The seminal work by the

Robinson group identified a number of MS-compatible non-ionic surfactants that can be used for the analysis of intact membrane proteins and protein complexes.[166] Nevertheless, protein solubilization using non-ionic surfactants is limited in comparison to that achieved with ionic surfactants and, as a result, ionic surfactants are still routinely employed during sample preparation. These MS-incompatible surfactants are then frequently removed by protein precipitation in organic solvents.[168] However, this method suffers from low recovery, and, therefore, considerable effort has been invested in the development of degradable surfactants that can be broken down prior to MS analysis. To date, the majority of the degradable surfactants that have been developed, including ProteaseMAX (Promega), RapiGest (Waters), and MaSDeS,[163] are acid-labile and are degraded during LC separation due to the presence of acid in the mobile phase. Although these surfactants have shown significant efficacy in bottom-up proteomics, the successful use of degradable surfactants for intact protein analysis has not been demonstrated.

Hydrophobic proteins, such as membrane proteins, represent a particularly difficult group of proteins to analyze using top-down MS. In contrast to bottom-up proteomics, wherein some soluble peptides from membrane proteins can be recovered following protein digestion, permitting their identification and quantification, the analysis of intact membrane proteins is extremely challenging due to aggregation and precipitation of these proteins in the absence of surfactants. Despite the challenges, a number of research groups have successfully utilized a variety of systems to study intact membrane proteins and complexes *in vitro*, including detergent micelles, bicelles, amphipols, and nanodiscs.[169] As mentioned above, non-ionic surfactants were successfully used for the analysis of membrane proteins and protein complexes by the Robinson group. Membrane proteins or protein complexes can be prepared in buffer containing low concentrations of MS-compatible surfactants, which are subsequently removed in the gas phase by thermal activation.

Removal of the surfactants permits MS analysis of the ionized protein or protein assembly.[166] The solubilization of membrane proteins for MS analysis can also be achieved using detergent-free systems. Bicelles and nanodiscs employ small, discoidal arrangements of phospholipids to form lipid bilayers that mimic native membrane structure to deliver membrane proteins for MS analysis.[170] Amphipols are amphipathic polymers with alternating hydrophilic and hydrophobic side chains, which can also be used for the solubilization of hydrophobic proteins. Proteins or protein assemblies can be released in the gas phase from bicelles, nanodiscs, or amphipols by collisional activation similar to the removal of non-ionic surfactants.[170] Although collisional activation aids in the removal of surfactants and lipid-like molecules, over-activation can lead to local unfolding of protein structure, dissociation of protein complexes, and even protein fragmentation or the loss of labile PTMs such as phosphorylation or glycosylation. Alternatively, investigators can exploit the chemical properties of certain proteins to maintain protein solubility in the absence of MS-incompatible surfactants or lipid-like molecules. High concentration of formic acid have proved efficacious for the solubilization of membrane proteins for top-down MS analysis.[171-173] However, prolonged storage of protein samples in formic acid can result in protein degradation and artefactual modifications (formylation), and, therefore, immediate solvent transfer to a more inert solution is needed.[171] Similarly, high concentrations of organic solvents containing MS-compatible chaotropes have also been used for the extraction of membrane proteins from purified mitochondria for top-down proteomics analysis.[161] Despite recent advancements in protein solubilization, the solubility of intact proteins, particularly membrane proteins, remain a significant challenge that awaits further development of novel reagents and methods.

1.6.1b Protein purification and separation

The complexity and high dynamic range of the proteome necessitates the effective fractionation and separation of proteins prior to MS analysis.[174] Affinity purification (AP) remains a popular approach for the targeted analysis of intact proteins and protein complexes. In addition to AP for hypothesis-driven targeted proteomics, there exists a variety of gel-based and LC-based methods for large-scale protein separation for discovery-driven research.

Traditional biomedical research has heavily relied on AP for the selective purification of a single protein or protein complex of interest. AP is an effective and robust means for the targeted study of proteins with known biological relevance. The Ge group has extensively utilized AP for the purification of cardiac troponin I (cTnI) from animal and human cardiac tissue for the study of disease-associated cTnI proteoform alterations using the top-down approach.[156, 164, 175-177] Nevertheless, AP usually involves the use of MS-incompatible salts that need to be removed prior to MS analysis, making AP difficult to couple directly with MS. Additionally, as a targeted approach for the purification of specific proteins of interest, AP is limited by the availability and fidelity of the analyte-specific stationary phase (e.g., antibodies).[147] Furthermore, AP is relatively low throughput and provides a very limited view of complex biological systems such that many important/disease-associated changes in the system may be overlooked. Therefore, the development of high-throughput and high-resolution protein fractionation and separation methods for the interrogation of the entire proteome or large sub-proteomes remains an active area of research.

Gel-based methods often exploit size or isoelectric point for the separation of proteins, and are widely used in biomedical research. Proteins separated in gels can be digested and analyzed by

the bottom-up approach due to the relatively high recovery of small peptides from acrylamide gels. However, the recovery of intact proteins from the gel matrix remains extremely difficult. Notably, the development of Gel-Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE) has allowed for multiplexed, high-resolution size-based separation of intact proteins [168, 178]. Tran et al. were able to identify over 3000 proteoforms from a HeLa cell lysate by employing solution isoelectric focusing followed by GELFrEE and online nanocapillary RPC-MS.[179] This multi-dimensional platform demonstrated the power of separation based on a number of different physiochemical properties, including isoelectric point, size, and hydrophobicity, for the analysis of intact proteins from a complex cell lysate. Using this method, Catherman et al. identified 1220 proteins, approximately 25% of which were membrane proteins, and over 5000 proteoforms from the H1299 human cell line,[180] which represents the largest top-down proteomics analysis conducted to date. Nevertheless, despite the high-resolution protein separation that can be achieved using the GELFrEE system, this method requires the use of SDS, which, as mentioned above, is not compatible with MS and must be removed prior to MS analysis.

An appealing alternative to gel-based methods are LC-based methods, including RPC, HIC, ion-exchange chromatography (IEC), and size exclusion chromatography (SEC). A significant advantage of LC-based separation methods is that they can be coupled directly with MS.[181] Online LC-MS not only increases the overall throughput of sample analysis, but also reduces potential artefactual modifications associated with lengthy sample handling. RPC makes use of a hydrophobic stationary phase that proteins will interact with when low concentrations of organic solvents are included in the mobile phase. The analytes are subsequently eluted using increasing concentrations of organic solvents. Due to the high compatibility of RPC with MS, RPC remains a popular choice for intact protein separation coupled with MS. Using RPC alone with a

long capillary column (80 cm) and an extended gradient, Ansong et al. were able to identify 563 proteins and 1665 proteoforms from *Salmonella typhimurium.*[182] HIC and IEC utilize buffered salt concentration gradients for protein separation based on hydrophobicity and electrostatic interactions, respectively. In HIC, a gradient with a decreasing concentration of salt is used to separate proteins based on hydrophobic interactions. Commonly used salts for achieving high-resolution separation of proteins by HIC contains sulfate, phosphate, and citrate, which are incompatible with MS.[183] Ammonium acetate is an MS-compatible salt, but HIC separation using ammonium acetate suffers from low separation power due to inadequate protein retention when used with conventional HIC materials.[183] Chen et al. recently evaluated a series of HIC columns with heightened degrees of hydrophobicity to compensate for the inadequate retention of proteins by the stationary phase when using ammonium acetate in the mobile phase. They demonstrated improved protein retention using columns with stationary phases that were more hydrophobic than conventional HIC stationary phases and a mobile phase containing ammonium acetate, which opens the door for the use of this powerful high-resolution separation method for top-down proteomics analysis.[183] Importantly, HIC and RPC are orthogonal with different selectivity for intact protein separation,[184] despite the fact that both methods separate proteins based on hydrophobicity. Moreover, HIC is a non-denaturing separation method and holds great promises for the study of native protein confirmation, as well as protein-protein and protein-ligand interactions. In addition to HIC, recent work using online IEC coupled with MS also demonstrated great success in the study of native proteins and protein complexes.[185] Nevertheless, detection of high molecular weight (MW) proteins remains challenging due to an exponential decay in the signal-to-noise ratio (S/N) for proteins with increasing molecular weight,[186] and, therefore, large proteins, particularly those co-eluting with low MW proteins, are often difficult to detect. As a

consequence, size-based separation of proteins prior to MS analysis is necessary for the detection of high MW proteins. SEC is a highly versatile method with limited sample loss due to reduced interaction with stationary phase.[187] SEC can also utilize a variety of mobile phase solvents, both denaturing and non-denaturing for the analysis of proteins and protein interactions. Nevertheless, SEC traditionally suffers from low resolution and undesirable sample dilution due to the use of high volumes of solvent.[174] To address this, Chen et al. used ultra-high pressure SEC, and demonstrated high-resolution, high-efficiency separation of proteins in a wide mass range (6-669 kDa) in only 7 minutes.[188]

In the analysis of complex protein mixtures, great separation power can be achieved by multi-dimensional LC platforms combining two or more orthogonal LC methods. Due to the high compatibility of RPC with MS, RPC remains a popular choice for intact protein separation and is often used in combination with other orthogonal LC methods for multi-dimensional separation. Notably, the Paša-Tolić lab coupled weak anion exchange (a subtype of IEC) to RPC-MS. This study used RPC with a long C5 nanoLC column (80 cm) coupled directly to a 12 Tesla Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer, and identified 750 proteins from *Shewanella oneidensis* lysates.[189] Valeja et al. employed a three-dimensional LC platform coupling IEC, HIC, and RPC with an Orbitrap mass spectrometer, and were able to identify 640 proteins in a single IEC fraction (out of 35 fractions) from a HEK 293 cell lysate.[184] There has been significant progress for intact protein separation; however, the separation power needs to further improve to address the extreme complexity and dynamic range of the human proteome. The development of separation methods that can be coupled online with MS and allow for the orthogonal and high-resolution separation of intact proteins will be essential for the advancement of high-throughput top-down proteomics.

1.6.1c Top-down MS instrumentation

A typical mass spectrometer consists of a sample inlet, ion source, mass analyzer, and detector. MS analysis relies on the efficient ionization of molecules of interest, and the development of "soft-ionization" methods, including matrix-assisted laser desorption ionization (MALDI)[190] and electrospray ionization (ESI)[191], which do not induce fragmentation or dissociation of intact proteins or protein complexes, has significantly advanced the MS analysis of biomacromolecules. For the analysis of intact proteins, ESI is generally more favorable due to the resulting multiply charged ions produced using this ionization method.[186, 192] In contrast, MALDI produces predominantly singly or doubly charged ions. The multiply charged ions generated by ESI aid in the detection of proteins in mass analyzers with limited mass (*m/z*) range. In addition, multiply charged ions can be more effectively fragmented using electron-based fragmentation methods.

For the analysis of large intact proteins, high-resolution mass analyzers are desired because high-resolution instruments allow for highly accurate determination of the protein mass.[193] In addition, following MS/MS analysis, the resulting spectra are usually highly complex and contain a significant number of overlapping fragment ion peaks, which require high-resolution to differentiate. Therefore, early work on the MS analysis of intact proteins was almost exclusively conducted using FT-ICR mass spectrometers.[194-196] One difficulty of using FT-ICR mass spectrometers is that they require superconducting magnets, which can be difficult and costly to maintain, and, therefore, top-down technology was limited to only a few specialized labs. Advancements in instrumentation have now opened the field of top-down proteomics to other research groups. For example, time-of-flight (TOF) instruments have been used for the study of intact proteins and native protein complexes.[166, 170] In fact, TOF instruments are typically preferred

for the study of very large protein complexes due to the essentially unlimited mass (*m/z*) range of these instruments.[169] However, the old-generation TOF detectors, such as electron multipliers and microchannel plates, are suboptimal for the detection of high MW proteins due to the low kinetic energy of these molecules in the drift tube, which decreases the probability of generating secondary electrons and dramatically reduces the detection efficiency of high MW molecules.[197] With the nanomembrane TOF detector developed recently, the kinetic energy of the ions generates mechanical oscillations in the nanomembrane thereby significantly improving the detection of large proteins or protein assemblies.[198] In addition, the hybridization of TOF instruments with quadrupole mass analyzers (q-TOF), which can act as a mass filter or collision cell, has greatly improved the capabilities of these instruments.[169] Recent work by Mann and coworkers has demonstrated the high resolving power and mass accuracy of commercially available q-TOF instruments,[199] making them an affordable and robust option for top-down protein analysis. As an alternative to FT-ICR with relatively low cost and maintenance, the newest generation of the Orbitrap mass spectrometers[200, 201] can also offer good mass accuracy and high resolving power. The Kelleher group has demonstrated identification of thousands of intact proteins using Orbitrap mass spectrometers.[202, 203] Nevertheless, FT-ICR mass spectrometers still remain the preferred instrument for achieving isotopic resolution of proteins that are larger than 70 kDa.[194, 204] Importantly, the new dynamically harmonized FT-ICR cells developed by Nicolaev and coworkers provide significant improvements on the stability of the transient, which enhances resolution and makes it easier to achieve isotopic resolution of large proteins.[205] Significant advancements in MS instrumentation have been made in the past decades, but large-scale top-down proteomics analysis of proteins, especially high MW proteins, remains challenging. Further developments in

instrumentation, including enhancements in resolution and sensitivity over a broad mass range, will significantly advance the field of top-down proteomics.

1.6.1d Top-down protein characterization

Following MS analysis of intact proteins, specific protein species of interest can be isolated in the mass spectrometer and fragmented using various fragmentation methods. In general, these fragmentation methods can be classified as either energetic or electron-based dissociation methods.[206, 207] Energetic dissociation methods, such as collision induced dissociation (CID) and infrared multiphoton dissociation, dissociate bonds of the lowest activation energy. In the case of unmodified proteins/peptides, energetic dissociation results in cleavage of the CO-NH bonds in the protein backbone, yielding *b* and *y* fragment ions (**Figure 1.3A**). However, when labile PTMs (e.g., phosphorylation or glycosylation) are present, dissociation using energetic fragmentation methods preferentially cleaves the covalent bonds attaching these modifications to the protein, resulting in PTM loss and preventing localization of the modification (**Figure 1.3A**).[147] On the other hand, the electron-based fragmentation methods, including electron capture dissociation (ECD)[208] and electron transfer dissociation (ETD)[209], induce local fast cleavage of N-C_α bonds in the protein backbone, yielding *c* and $z^{\cdot}$ fragment ions, and preserving labile PTMs (**Figure 1.3B**). In addition, electron-based dissociation methods are non-selective and tend to produce cleavage that is randomly distributed along the protein backbone, which can increase sequence coverage and aid protein identification and the localization of PTMs.[147, 195]

Although ESI is generally preferred due to the production of multiply charged ions, the use of ESI results in dispersion of the protein signal over a large number of charge states and causes

detrimental reduction of the S/N, especially for high MW proteins.[186] The development of in-spray supercharging ESI following LC separation allows for preferential formation of ions with higher charge states,[210] which significantly increases S/N of the proteins, and holds great promise for the characterization of large proteins using electron-based dissociation methods. In addition to supercharging reagents for reducing charge state dispersion of protein signals, another technology, ion parking via ion/ion proton transfer reaction,[211] has proved to be particularly useful in reducing the number of charge states of high MW proteins, and has recently gained significant attention.[212] Ion parking allows for the concentration of protein signal over a small number of charge states, which allows for detection of low abundant and high MW proteins, as well as gas-phase "purification" of the proteoforms of interest for comprehensive sequence characterization.[212]

However, efficient fragmentation of large proteins (particularly those >50 kDa) remains challenging due to the fact that extensive intramolecular interactions promote the formation of highly compact gas phase structures that are refractory to fragmentation using conventional dissociation methods.[196] To enhance the fragmentation of large proteins, Han *et al*. utilized increased pre- and post-skimmer voltages and electrospray additives to reduce structural rearrangement of ions in the gas-phase and facilitate the fragmentation of large proteins.[196] Using their "pre-folding dissociation" strategies, the authors were able to achieve high coverage of fragmentation at the termini of large proteins up to 200 kDa.[196] More recently, the Brodbelt group has demonstrated enhanced sequence coverage of intact proteins using ultraviolet photodissociation (UVPD) in comparison to conventional fragmentation methods, including CID, high energy collision dissociation, and ETD.[213] UVPD with a 193 nm laser yields extensive fragmentation producing predominantly $a^{\cdot}$-x ions and a significant amount of b-y and c-$z^{\cdot}$ ions (**Figure 1.3C**), as well as other ions of side-chain cleavage.[213] Additionally, unlike ECD/ETD,

which exhibit significant charge state dependence of fragmentation, fragmentation efficiency with UVPD is relatively independent on the charge state of the precursor ions. High-resolution mass spectrometers implemented with UVPD presents a powerful tool for the analysis of intact proteins and protein complexes.

Nevertheless, fragmentation of large proteins produces very complex tandem mass spectra with a mixture of ions of various molecular weight and charge states that quite often overlap. Therefore, mass spectrometers with high resolution and sensitivity are needed for the analysis of complex product ions resulted from fragmentation of high molecular weight proteins. In addition, ion parking after MS/MS analysis allows for reduction of the ion species as well as concentration of the ion signals,[212] which makes it an appealing option for the characterization of high molecular weight proteins.

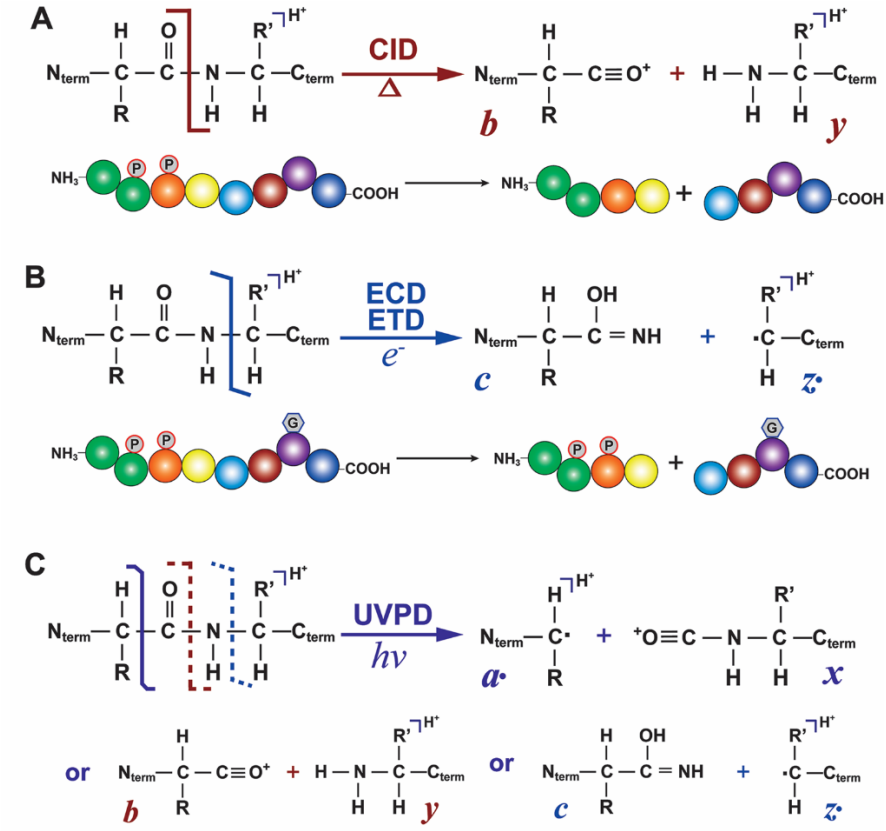


Figure 1.3. Schematics showing the fragment ions produced by CID (A), ECD/ETD (B), and UVPD (C). Noted that the dissociation pathways delineated are not exclusive. CID usually resulted in the loss of labile PTMs (A), whereas ECD/ETD typically preserves them (B).

1.6.1e Top-down protein quantification

Unlike in bottom-up proteomics, where the addition of modifications to peptides can significantly influence ionization, the physiochemical properties of intact proteins are relatively unaffected by the addition of small modifications or minor sequence variations,[214] which allows for relative quantification of modified and unmodified proteoforms present within the same spectrum. Relative quantification of protein PTMs and isoforms using top-down MS has been widely employed by our lab and others for uncovering disease-associated proteoform alterations and identifying biomarkers.[24, 156, 157, 160, 182, 214, 215] Using this quantification method, in a large-scale top-down proteomics analysis of pathogenic bacterium *Salmonella typhimurium,* Ansong et al. identified and quantified 1665 bacterial proteoforms, and revealed a unique protein PTM switch from predominantly protein S-cysteinylation to S-glutathionylation under infection-like conditions.[182] Similarly, Chamot-Rooke et al. employed top-down MS to investigate the PTMs of the intact pilin of *Neisseria meningitides* that causes cerebrospinal meningitis.[215] They discovered that enhanced glycerophosphorylation of pilin was responsible for the destabilization of type IV pili fiber bundles and enhanced bacterial detachment and migration across epithelial cells.[215] In addition to infectious diseases, top-down proteomics also revealed a decreased in the phosphorylation of cTnI in chronic heart failure.[157]

Although relative quantification of proteoform abundance changes aids in dissecting the molecular mechanisms underlying various diseases, including heart diseases and infectious

diseases, this method does not permit quantification of protein expression changes across samples. To quantify protein expression changes across samples, substantial efforts have been invested for developing methods to label intact proteins. Early work employing stable isotopic labeling by amino acids (SILAC) demonstrated reliable quantification of intact proteins[216, 217] in a small scale. Chemical labeling such as isobaric tags for relative and absolute quantitation (iTRAQ) allows for quantification of protein expression in samples other than cell culture or mouse models, and has been widely used in bottom-up proteomics. Adapted from iTRAQ, metal-coded affinity tags (MeCAT) were developed for the quantitative analysis of intact proteins.[218] Although SILAC and MeCAT have been successfully implemented for the quantitative analysis of protein expression changes in biological samples, these methods may suffer from detrimental deterioration of protein signals due to signal spreading with partial incorporation of the isotopic labels or metal-coded tags. On the other hand, Mazur et al. utilized label-free differential MS for the quantification of intact apolipoproteins in human high density lipoproteins (HDL). By carefully monitoring instrument response and sample load, Mazur et al. showed an average 4.7 (p-value 0.017) increase of O-glycosylated apolipoprotein C-III in the low versus high plasma HDL donors.[219] This study demonstrated the feasibility of using top-down MS for reliable quantification of protein abundance in a small scale. Taking this further, the Kelleher group employed label-free MS for the quantification of intact proteins less than 30 kDa in the growing and senescent fibroblasts, and confidently quantified 751 proteoforms, among which a proteoform of the enhancer of rudimentary homologue and Isoform 1 of FAM107B were down-regulated in the senescent cells.[203] Nevertheless, this study is limited due to the biased detection of low MW proteins. High-throughput large-scale quantification of protein expression level by top-down MS remains a significant challenge, and, thus, presents significant opportunities for future top-down research.

1.6.1f Top-down data analysis

Greatly aided by the well-developed computational tools for data analysis, bottom-up proteomics has gain considerable popularity for protein analysis. On the contrary, data analysis tools for top-down proteomics remain relatively under-developed. The use of high-resolution mass spectrometers and the resulting complex mass spectra necessitates sophisticated tools for the deconvolution of top-down mass and tandem mass spectra. Thorough High-Resolution Analysis of Spectra by Horn (THRASH)[220] was the first automatic computational tool developed for the deconvolution of high-resolution mass spectra, and is still widely used. THRASH compares the experimental isotopomer envelope with a theoretical envelope generated based on the averagine,[221] computes a figure of merit value for each envelope and reports those score higher than a user-defined fit score. MS-Deconv, which adopts a combinatorial approach for the deconvolution of overlapping isotopomer envelopes,[222] has also been applied for top-down data analysis. Recently, UniDec, which employs a Bayesian approach for the separation of mass and charge dimension,[223] can also be used for the deconvolution of high-resolution spectra. Although these algorithms significantly facilitate data processing and interpretation, the complexity of high-resolution top-down MS and MS/MS data often results in a considerable amount of mis-assigned peaks necessitating the implementation of visualization functions for the validation of the deconvolution results. Decon2LS uses a variation of THRASH for high-resolution spectral analysis and visualization.[224] The DataAnalysis (Bruker Daltonics) utilizes SNAP2, a THRASH-based algorithm, for spectral deconvolution, and allows for direct visual comparison of the experimental isotopomer envelopes with theoretical patterns (termed SNAP pattern). Although these tools allow for the visual comparison of the experimental isotopomer envelopes with the theoretical ones for

the determination of monoisotopic mass, manual adjustment of the computational output for the correction of charge state and mass error is not supported.

Spectral deconvolution of tandem mass spectra produces a mass list, which allows for protein identification. There are a number of software tools for protein identification in top-down proteomics, including ProSight PC,[225] Mascot Top Down (also known as Big-Mascot),[226] MS-TopDown,[227] and MS-Align+ .[228] Both ProSight PC and Mascot Top Down require specific PTMs to be defined prior to database search, which limits their utility in the identification of unexpected PTMs. MS-TopDown and MS-Align+ both employ the spectral alignment algorithm, which matches MS/MS data to proteins in the database, and allows for the identification of mass discrepancy between the experimental and theoretical fragment ions, which represents PTMs and/or sequence variations. In particular, MS-Align+ enables efficient filtering of candidate proteins when searches in a large protein database, and, therefore, allows for rapid identification of proteins and unexpected PTMs.[228] MS-Align+ also provides statistical evaluation of the identification results.[228] Recently, an upgraded version of MS-Align+, TOPPIC, was developed for faster protein identification and characterization with less computing requirements.

Recently, the Ge group has developed an integrated software tool, MASH Suite Pro,[229] which incorporates both THRASH and MS-Deconv for spectral deconvolution, MS-Align+ for protein identification, as well as a number of in-house developed functionality for protein quantification and characterization. Moreover, MASH Suite Pro provides easy access to the real spectra and visualization of each deconvoluted mass, with additional features allowing for deletion of false positives and correction of the mass values or charge states when mistakes occur. The visualization features and flexibility for correction of computational outputs significantly facilitate top-down data analysis with high accuracy and reliability. Importantly, MASH Suite Pro is capable

of processing both offline MS/MS and online LC-MS/MS data, which greatly simplify and speed up data analysis in top-down proteomics.

1.6.2 Application of top-down proteomics to heart diseases

Given the importance of protein isoform switching and PTM changes in cardiac contraction and relaxation, as well as in the development of heart failure,[151, 230] top-down proteomics has great potential to aid in the elucidation of the molecular mechanisms underlying physiological and pathophysiological processes in the heart.[230] In this section, we will summarize the applications of top-down proteomics to the advancement of our understanding of diseases of the heart. Moreover, we will highlight the potential of the technology to be translated into clinical setting for the identification of novel drug targets and the discovery and validation of biomarkers.

1.6.2a Identification of contractile protein isoforms and their associated proteoforms

Sarcomeres are the basic contractile units that are composed of myofilaments flanked by dense protein structure called Z-disc.[24, 160, 231] Myofilaments consist of the actin-based thin filaments and myosin-based thick filaments, the interaction of which mediates muscle contraction and relaxation in a Ca^{2+} dependent manner. The troponin complexes, cTnI, cardiac troponin T (cTnT), and cardiac troponin C are thin filament associated proteins, which induce conformational change of Tropomyosin (Tpm or Tm) with increased intracellular Ca^{2+}, and, subsequently, release the myosin binding site on the actin filament and initiate muscle contraction.[24, 160]

Unlike the bottom-up approach, top-down MS analyzes intact proteins, which preserves important sequence information of the proteins present in the samples. The ability of top-down MS for maintaining full sequence coverage allows for the analysis of highly homologous protein isoforms, as well as the comprehensive characterization of their associated proteoforms. Human Tpms are encoded by different genes (*Tpm1*, *Tpm2*, *Tpm3*, and *Tpm4*) and the alternative splicing event can generate a wide variety of proteoforms from a single gene. Using top-down MS, Peng et al. were able to unambiguously determine the presence and relative abundance of the Tm isoforms and their associated proteoforms in human and swine cardiac tissue, including α-Tpm, β-Tpm, and κ-Tpm, as well as the phosphorylated α-Tm and β-Tpm.[232, 233] α-Tpm and κ-Tpm are different isoforms encoded by *Tpm1* due to alternative splicing. α-Tm is well-known to be highly expressed in cardiac tissue, while κ-Tpm was only recently found to be expressed in the heart and confirmed using top-down MS.[233] Interestingly, the expression of κ-Tpm is higher in the left atrium as compared to the left ventricle,[233] suggesting important functional roles of κ-Tpm in regional heterogeneity.

In addition to Tpm, changes in the relative abundance of α-cardiac actin and α-skeletal actin in the heart have been associated with cardiac dysfunction.[234] However, precise quantification of these two isoforms is difficult because the two isoforms differ from each other by only 32 Da due to two juxtaposed amino acids.[235] Although gene expression levels are typically quantified by measuring the levels of the protein-specific mRNAs, studies have suggested that protein levels do not necessarily correlate with mRNA levels.[236] By employing high-resolution top-down MS, Chen et al. were able to distinguish between α-cardiac and α-skeletal actin in both human and swine heart tissue.[235] Moreover, top-down proteomics revealed that there is a marked

upregulation of α-skeletal actin relative to α-cardiac actin in the failing hearts as compared to the donor hearts, which can potentially serve as a biomarker for cardiac dysfunction.[235]

1.6.2b Mapping sequence variations and PTMs in contractile proteins

Top-down MS is the premier method for the unambiguous localization of sequence variations and PTMs without prior knowledge. By taking advantage of a variety of fragmentation methods, any protein sequence variations and PTMs can be localized to a single amino acid. Using high-resolution top-down MS, Zhang et al. identified a -28 Da mass discrepancy in cTnI from swine.[175] Top-down MS/MS analysis revealed that this mass discrepancy was the result of a naturally occurring polymorphism (Val to Ala switch at position 116) in swine cTnI.[175] Similarly, Peng et al. identified two amino acid polymorphisms (Pro to Gln switch at position 38, and Pro to Leu switch at position 64) in swine α-Tpm in the deep sequencing of swine Tpm isoforms.[233]

Protein PTMs regulate all aspects of protein biology, including protein localization, activity and degradation, and, consequently, play critical roles in cardiac physiology and pathophysiology.[230] While sequence variations/mutations may be readily detected by DNA sequencing, top-down MS is a powerful method for the comprehensive identification and localization of protein PTMs, as well as the determination of potential PTM cross-talk, which cannot be derived from DNA sequence. For examples, cTnI phosphorylation at Ser42/44 was identified *in vitro* and *in vitro* contractility assays showed that phosphorylation of these sites was detrimental for contractility.[237] However, it was not known whether these sites were phosphorylated *in vivo*. until Dong et al. employed a top-down proteomics approach and confirmed the phosphorylation of cTnI *in vivo* using rat models of spontaneously hypertensive heart failure.[156]

In addition to determining the biological relevance of PTMs identified *in vitro*, top-down MS is also capable of identifying unexpected PTMs for unraveling the regulatory pathways involved in cardiac dysfunction. As demonstrated in a swine model of acute myocardial infarction (AMI), commonly known as heart attack, top-down MS revealed phosphorylation in cTnI, myosin regulatory light chain (RLC) and a previously un-identified protein (UnK) (**Figure 1.4A-C**). Subsequent MS/MS analysis identified the unknown protein as enigma homolog 2 (ENH2), a Z-disc protein that was not previously found to be a phosphoprotein (**Figure 1.4C**). Importantly, MS/MS using ECD unambiguously localized the site of ENH2 phosphorylation to Ser118[24] (**Figure 1.4D**). As opposed to the traditional view of the Z-disc playing predominantly structural roles, this study provided definite evidence that Z-disc proteins can play critical regulatory roles post-myocardial infarction.[24]

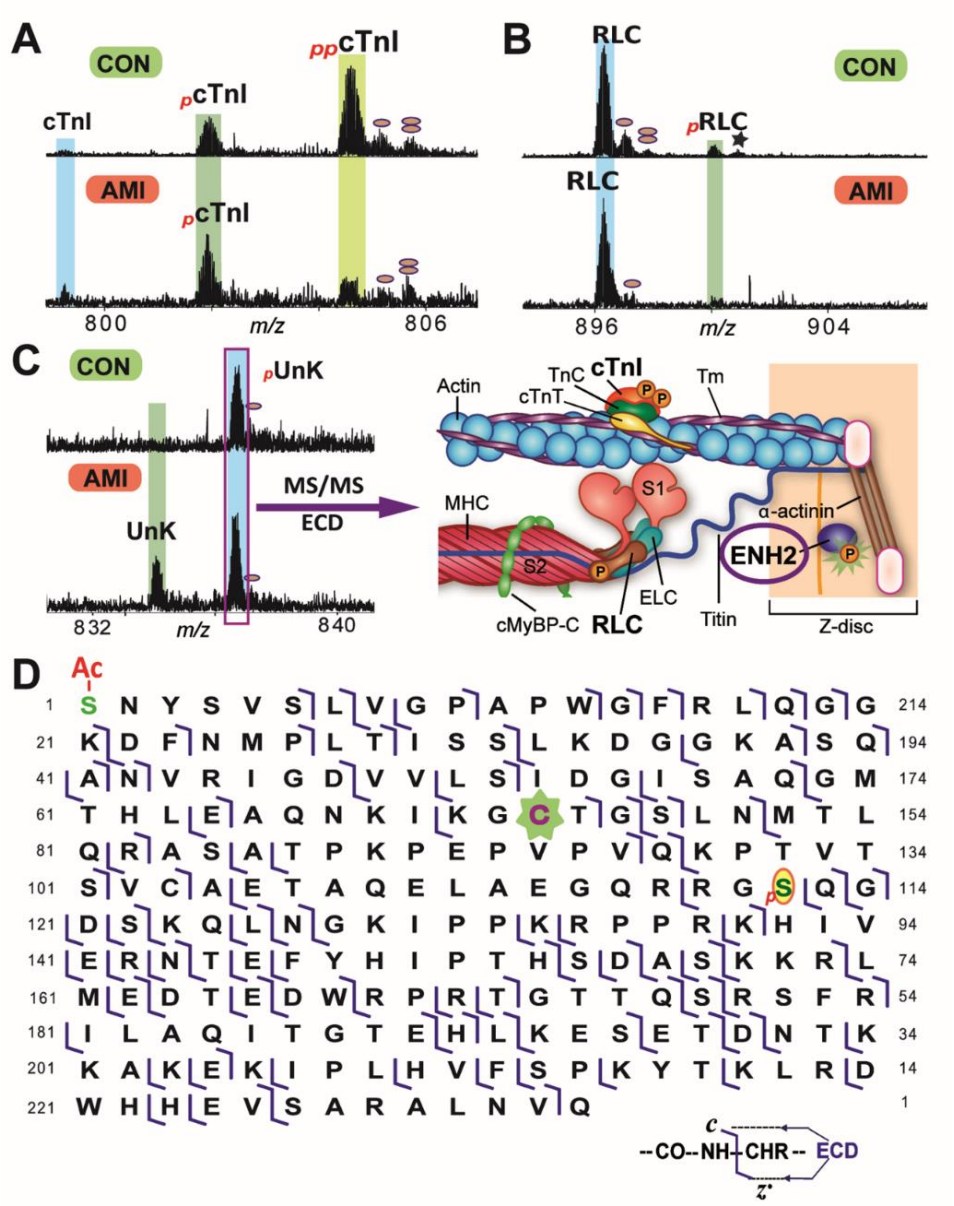


Figure 1.4. Top-down proteomics revealed a concerted reduction of protein phosphorylation in cTnI (A), RLC (B), and an unknown protein (UnK) (C), which was subsequently identified as ENH2. D. Top-down MS/MS analysis unambiguously localized the site of ENH2 phosphorylation to Ser118, and detected a previously unknown amino acid polymorphism at Cys72. This research was originally published in Mol Cell Proteomics.[24] Adapted from ref[24] with permission. © The American Society for Biochemistry and Molecular Biology.

PTMs had typically been studied in isolation until, recently, emerging evidence suggested that proteins can be modified by multiple PTMs and that the presence or absence of specific PTMs

can help or hinder modifications at other sites.[238] Top-down proteomics is the premier method for the analysis of co-occurring PTMs and potential PTM crosstalk, as the study of intact proteins retains the combinatorial nature of PTMs. By the top-down approach, Zhang et al. was able to determine the sequence of cTnI phosphorylation using the cardiac tissue from patients with chronic heart failure. Ser22 was the only site of phosphorylation observed in mono-phosphorylated cTnI, while both Ser22/23 were present in bis-phosphorylated cTnI, suggesting that cTnI is first phosphorylated at Ser22 followed by phosphorylation at Ser23.[157] Cardiac myosin binding protein C (cMyBP-C) is an important component of the thick filament, and plays important roles in the regulation of actin-myosin cross-bridge and Ca^{2+} sensitivity of force generation.[239] Mutations in cMyBP-C have been shown to cause a variety of cardiac disorders, including dilated cardiomyopathy and hypertrophic cardiomyopathy.[239] A significant number of cMyBP-C mutations are caused by frame-shift or point mutations that lead to pre-mature termination of the coding sequence, resulting in truncated versions of cMyBP-C with various lengths.[46] In the study aiming to characterize the phosphorylation sites of full length and truncated cMyBP-C, Ge et al. discovered that truncation of cMyBP-C significantly altered its phosphorylation by protein kinase A (PKA).[158] While the PKA target sites in full-length cMyBP-C locate at Ser283, Ser292, and Ser312, with phosphorylation at Ser292 occurring prior to the other sites, truncation of cMyBP-C at the C-termini leads to altered specificity of the PKA-medicated phosphorylation. Ser292, Ser312, and Ser484 were found to be phosphorylated on the truncated cMyBP-C.[158] This study suggested that truncation can significantly alter specificity of modification events and provided important insights into the potential mechanisms underlying cardiac disorder caused by cMyBP-C truncation.

1.6.2c Quantification of cardiac contractile proteoforms

The ability of top-down proteomics to provide a "bird's eye" view of all proteoforms allows for robust and reliable quantification of proteoforms within a same spectrum. Top-down proteomics analysis of cTnI revealed a decline of cTnI phosphorylation with progressive cardiac hypertrophy, and a markedly abolishment of cTnI phosphorylation in chronic heart failure.[157] Subsequent MS/MS analysis of cTnI identified the sites decreased cTnI phosphorylation exclusively to Ser22/23, which are two well-established target sites of phosphorylation mediated by PKA.[237] This study represents the first application of top-down proteomics to study changes in contractile protein PTMs in clinically relevant samples, underscoring the potential of top-down proteomics for the quantification of biologically relevant PTM changes for biomarker discovery. Using a similar approach, Dong et al. quantified the phosphorylation of cTnI in a rat model of hypertensive heart failure, and uncovered an increase in cTnI phosphorylation. Interestingly, distinct from chronic heart failure in human, the sites of increases cTnI phosphorylation were localized to Ser22/23 and Ser42/44, the latter of which were established target sites of PKC *in vitro*, suggesting an important role of PKC in the development of hypertension-induced heart failure in rodents. Subsequently, Dong et al. evaluated the expression level of PKC, and found that PKCα was up-regulated in the rats with hypertensive heart failure.

By improving the strategies for contractile protein extraction and separation, Peng et al. were able to simultaneously quantify the proteoforms of multiple important contractile proteins, including cTnT, cTnI, RLC, α-Tpm, β-Tpm, and ENH2 in the swine model of AMI. While the phosphorylation of cTnT, α-Tpm and β-Tpm remained relatively unchanged following 90 min of ischemia, significant decreases of cTnI, RLC and ENH2 phosphorylation were observed (**Figure 1.4**).[24] This represents the first study investigating the PTM changes of contractile proteins

immediately following heart attack, and suggests that impaired PKA pathways are responsible for cardiac dysfunctions after heart attack.

In addition to the quantification of disease-associated proteoform changes in various animal models and human, top-down proteomics was used for the characterization of cardiac heterogeneity. Though heterogeneity of the heart is well-appreciated, transmural and regional difference of the cardiac contractile protein PTMs were not well-established, until the recent study performed by Gregorich et al. using top-down high-resolution MS.[160] Whereas minimal transmural heterogeneity in the phosphorylation of cTnT, cTnI, RLC and Tpm were observed, the study revealed lower basal phosphorylation of cTnI and, conversely, higher α-Tpm phosphorylation in the atria versus ventricles (**Figure 1.5**), which underscores the importance of sampling in the analysis of cardiac PTM changes.[160] Importantly, regional difference in the basal phosphorylation of cTnI and α-Tm likely contributes to the differences in the contractile function of the atrial and ventricular muscle preparations.[160]

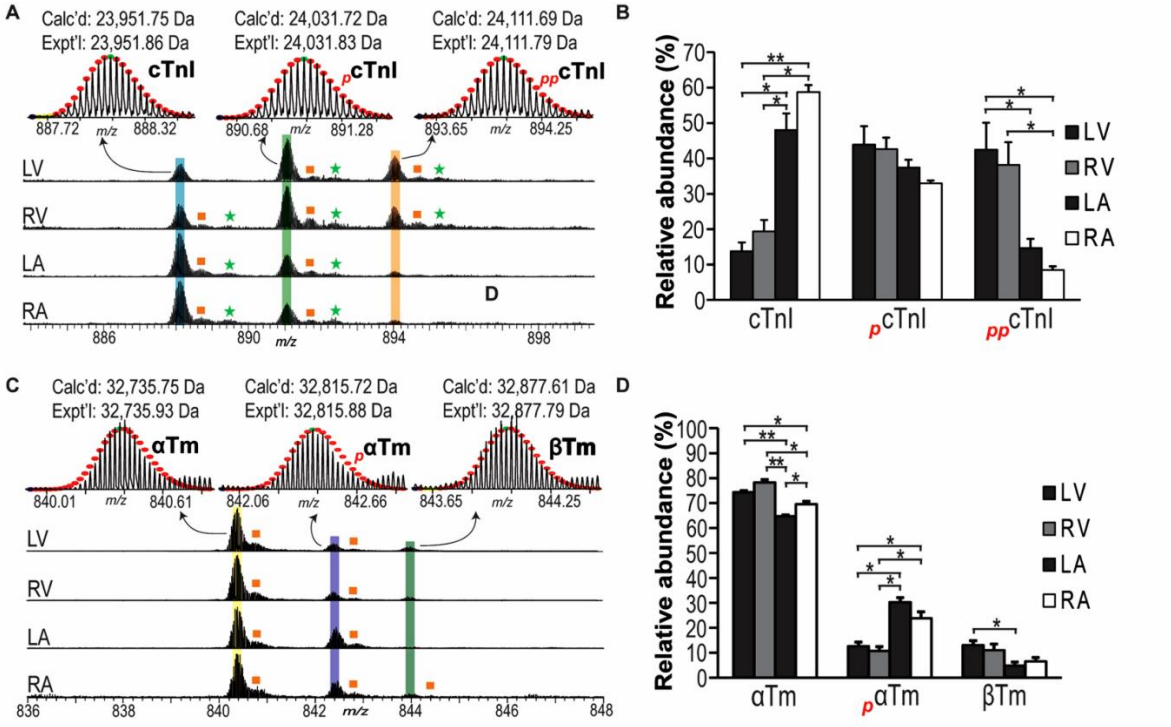


Figure 1.5. Top-down proteomics revealed regional heterogeneity of cTnI and α-Tm proteoforms. Basal phosphorylation of cTnI is higher in left and right ventricles versus atria (A-B), and basal phosphorylation of α-Tm is higher in the atria versus ventricles (D-E). αTm is equivalent to α-Tpm; βTm is equivalent to β-Tpm. This research was originally published in J Mol Cell Cardiol.[160] Adapted from ref[160] with permission from Elsevier and authors. © Elsevier

1.6.3 Expert Commentary

Top-down high-resolution MS-based proteomics is a premier method for the global and comprehensive study of protein isoforms and their associated proteforms, enabling the identification of novel protein isoforms and PTMs, characterization of un-identified sequence variations, and quantification of disease-associated proteoform alterations. Given the importance of isoform switching and PTM changes in the regulation of cardiac contraction and relaxation, and in the development of heart failure,[151, 231] top-down proteomics holds great promise for uncovering the molecular mechanisms underlying diseases of the heart towards the discovery of novel therapeutic targets/strategies to combat this devastating syndrome. In addition, the capability of top-down proteomics for a global view of the essential contractile proteins allows for the discovery of novel biomarkers for various heart diseases. Furthermore, owing to the drastic heterogeneity of patients and their clinical profiles, top-down proteomics employing rapid high-throughput LC-MS profiling of the cardiac proteins can be used for the comprehensive characterization of cardiac protein isoforms and the associated proteoforms, which allows for better understanding of the patients and proper selection of treatments.

1.6.4 Five-year view

Top-down MS has proved powerful in the research setting for the discovery of novel protein isoforms (produced by multiple homologous genes) and the associated proteoforms (produced by a single gene), and the quantification of biologically relevant proteoform alterations, which holds great promise for clinical applications. However, as an emerging technology, challenges remains in the development of top-down proteomics methodology and its application to heart diseases in the research and clinical settings. Improvements in the solubilization of hydrophobic proteins, detection of large and low abundant proteins, enhancement of overall throughput and automation, and quantification of protein expression are the focus areas of ongoing research. Heart diseases remain a devastating threat to the well-being of the general population. Top-down proteomics, in combination of cardiac physiology, metabolomics and functional analysis, will further advance our understanding of the mechanisms underlying various cardiac diseases; and uncover novel targets for the development of effective therapeutic approaches.

CHAPTER 2

TEMPERATURE-SENSITIVE SARCOMERIC PROTEIN POST-TRANSLATIONAL MODIFICATIONS REVEALED BY TOP-DOWN PROTEOMICS

ABSTRACT

Despite advancements in symptom management for heart failure (HF), this devastating clinical syndrome remains the leading cause of death in the developed world. Studies using animal models have greatly advanced our understanding of the molecular mechanisms underlying HF; however, differences in cardiac physiology and the manifestation of HF between animals, particularly rodents, and humans necessitates the direct interrogation of human heart tissue samples. Nevertheless, an ever-present concern when examining human heart tissue samples is the potential for artefactual changes related to temperature changes during tissue shipment or sample processing. Herein, we examined the effects of temperature on the post-translational modifications (PTMs) of sarcomeric proteins, the proteins responsible for muscle contraction, under conditions mimicking those that might occur during tissue shipment or sample processing. Using a powerful top-down proteomics method, we found that sarcomeric protein PTMs were differentially affected by temperature. Specifically, cardiac troponin I and enigma homolog isoform 2 showed robust increases in phosphorylation when tissue was incubated at either 4 ^{o}C or 22 ^{o}C. The observed increase is likely due to increased cyclic AMP levels and activation of protein kinase A in the tissue. On the contrary, cardiac troponin T and myosin regulatory light chain phosphorylation decreased when tissue was incubated at 4 ^{o}C or 22 ^{o}C. Furthermore, significant protein degradation was also observed after incubation at 4 ^{o}C or 22 ^{o}C. Overall, these results indicate that temperature exerts various effects on sarcomeric protein PTMs and careful tissue handling is critical for studies involving human heart samples. Moreover, these findings highlight the power of top-down proteomics for examining the integrity of cardiac tissue samples.

2.1 INTRODUCTION

Heart failure (HF) is a devastating condition that afflicts approximately 5 million Americans and accounted for 1 in 9 deaths in the year of 2013.[140, 141, 240] Despite recent advancements in the treatment strategies available to manage the symptoms of HF, the number of deaths attributable to this devastating clinical syndrome remained nearly as high in 2013 as it was in 1995.[240] Although animal models of HF have greatly advanced our understanding of the molecular mechanisms underlying this condition, differences in the manifestation of HF between animals, particularly rodents,[123, 241] and humans necessitates the direct interrogation of human heart tissue samples to fully elucidate the mechanisms of HF pathogenesis in humans. However, an ever-present concern when examining human heart tissue samples is the potential for artefactual changes related to temperature changes due to prolonged storage of the tissue samples, sharing and shipping of the samples between laboratories, and handling of the tissue samples. Flash-freezing cardiac tissue in liquid nitrogen remains the best method for preserving the integrity of the tissue;[242] yet, tissue warm-up post-procurement may occur during shipping and/or when processing the cardiac tissue samples, with unknown consequences at the protein level.

Sarcomeres are the basic contractile units of striated muscle, and consist of myofilaments flanked on either side by complex protein structures known as Z-discs.[33, 243, 244] Myofilaments are composed of the myosin-based thick filaments and actin-based thin filaments.[24, 33] The thick filaments contain myosin heavy chain and cardiac myosin binding protein C (cMyBP-C), as well as the myosin regulatory and essential light chains (RLC and ELC, respectively). The thin filaments consist of actin, tropomyosin (Tpm), and the cardiac troponin complex, which is made up of troponin C (TnC), cardiac troponin I (cTnI), and cardiac troponin T (cTnT). Recent research has convincingly demonstrated that alterations in myofilament post-translational modifications

(PTMs) can be causative in contractile dysfunction and HF.[230, 245-249] In addition, emerging evidence indicates that Z-discs represent critical signaling nodes within cardiomyocytes,[33, 250] and that the PTMs of Z-disc proteins can be altered in response to cardiac injury.[24, 251]

Herein, we employed high-resolution top-down proteomics[147, 206, 244, 252] to evaluate the impact of temperature on tissue quality and sarcomeric protein PTMs. In contrast to the conventional bottom-up proteomics approach wherein proteins are digested into peptides prior to mass spectrometry (MS) analysis, intact proteins are analyzed in the top-down proteomics workflow, which provides a "bird's eye" view of all proteoforms[146] (a term encompassing all protein variants of the same gene arising from genetic variation, alternative mRNA splicing and PTMs) and permits the comprehensive, and quantitative analysis of protein PTMs.[147, 244, 252, 253] Following MS analysis, a specific proteoform of interest can be "purified" in the gas phase and undergo fragmentation, known as tandem MS (MS/MS), for protein sequence characterization and PTM localization. Moreover, the top-down approach is the premier method for revealing small truncated species of proteins, which can be difficult to detect using bottom-up proteomics or conventional SDS-polyacrylamide gel electrophoresis. To gain insight into temperature-related changes in human cardiac tissue, we examined changes in sarcomeric protein PTMs at temperatures mimicking those that may occur during tissue storage, shipment, or sample processing, via a top-down proteomics.

2.2 MATERIALS AND METHODS

2.2.1 Reagents and chemicals

All reagents were purchased from Sigma-Aldrich, Inc. (St. Louis, MO, USA) unless otherwise noted. HPLC grade water, acetonitrile, and ethanol were purchased from Fischer Scientific (Fair Lawn, NJ, USA).

2.2.2 Human cardiac tissue collection

A donor heart characterized with normal cardiac function, but deemed unacceptable for transplantation, was obtained from the University of Wisconsin Hospital and Clinic. The procedure for the collection of human cardiac tissue has been approved by the Institutional Review Board of the University of Wisconsin-Madison. The donor heart was maintained in cold cardioplegic solution (4 °C) immediately after explanted and delivered to laboratory for dissection within 15 min as described previously.[157] Cardiac tissue dissection was carried out on a metal plate chilled with dry ice, and the dissection of the left ventricular tissue was carried out first and completed within 15 min. Individual pieces of tissue were wrapped with pre-labeled aluminum foil and snap-frozen in liquid nitrogen. The tissues were stored at -80 °C for subsequent analysis.

2.2.3 Time-course experiments and tissue storage condition test

One piece (~200 mg) of left ventricular tissue from the midwall region was used for all experiments. To assess time-related changes in sarcomeric protein PTMs, three portions of approximately 4-6 mg of tissue were excised on a metal plate pre-cooled with dry-ice and homogenized immediately to evaluate the baselines of sarcomeric protein PTMs (time 0). Additionally, 4-6 mg of tissue was excised and incubated at either 4 °C or 22 °C (room temperature) for 15 min, 30 min, 1 h, 2 h, or 4 h, followed by tissue homogenization and protein

extraction. Three small portions of the tissues were maintained in the HEPES buffer with protease and phosphatase inhibitors (25 mM HEPES pH 7.5, 50 mM NaF, 1 mM Na3VO4, 1 mM PMSF, 2.5 mM EDTA), and incubated at 4 °C for 30 min before protein extraction. The leftover tissue that remained frozen was rapidly cut into two parts (~80 mg each), and one portion was stored in liquid nitrogen while the other was stored at -80 °C. After approximately 18 months, about 4-6 mg of tissues from each portion were homogenized to extract sarcomeric proteins. The tissue that was stored in liquid nitrogen was then divided into 3 portions, and each were stored in liquid nitrogen, dry ice and wet ice, respectively, for 48 h to mimic shipment conditions. These tissues were then homogenized in the same way and the sarcomeric proteins were extracted.

2.2.4 Cardiac saromeric protein extraction

All sample processing steps were carried out in a cold room (4 °C), and the tissue portions were maintained completely frozen using dry ice prior to the temperature treatment. All homogenization steps were performed in the cold room, and homogenization of each samples took less than 20 s. Extraction of sarcomeric proteins was carried out as reported previously.[24, 160, 254] Briefly, tissue was first homogenized rapidly in 10 vol. (µl/mg tissue) of HEPES buffer containing protease and phosphatase inhibitors (25 mM HEPES pH 7.5, 50 mM NaF, 1 mM Na_3VO_4, 1 mM PMSF, 2.5 mM EDTA) using a Teflon pestle (1.5 mL tube rounded tip; Scienceware, Pequannock, NJ, USA). The homogenate was centrifuged at 17,000 rcf for 15 min, and the pellet was washed with 10 vol. of the HEPES buffer, and further homogenized in 10 vol. of trifluoroacetic acid (TFA) solution (1% TFA, 1 mM Tris-(2-carboxyethyl)phosphine). The homogenate, which contains primarily sarcomeric proteins, was centrifuged at 17,000 rcf for 15 min and the supernatant was collected and centrifuged again at 17,000 rcf for 60 min to remove residual particulate matter prior to liquid chromatography (LC)-MS analysis.

2.2.5 Reverse phase chromatography (RPC) and top-down proteomics analysis

LC-MS analysis was carried out using a NanoAcquity Ultra-high Pressure LC system (Waters, Milford, MA, USA) coupled to a high-resolution Impact II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker Daltonics, Bremen, Germany). To evaluate the reproducibility of protein extraction and the LC-MS method, three extraction replicates at the same time point (time 0) and three injection replicates of the same protein extract were tested. The sarcomeric proteins in each sample were eluted by a gradient of 5% to 95% mobile phase B (mobile phase A: 0.1% formic acid in water; mobile phase B: 0.1 % formic acid in 50:50 acetonitrile:ethanol) at a constant flow rate of 8 µL/min. Proteins eluted were delivered to the mass spectrometer via electrospray ionization. End plate offset and capillary voltage were 450 V and 4000 V, respectively. Nebulizer pressure was set to 0.5 Bar and dry gas flow rate was 4.0 L/min. In-source collisional energy was set to 10 V. Mass spectra were collected at a scan rate of 0.5 Hz over 500-2000 *m/z* range.

2.2.6 Offline fraction collection and top-down MS/MS analysis

The elution time of target proteins were identified and the eluents containing these proteins were collected and analyzed using a 12 Tesla solariX Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics) via an automated chip-based nanoESI source (Triversa NanoMate; Advion Bioscience, Ithaca, NY, USA). The proteoforms of interest were first purified in the gas phase using the quadrupole with an 8-12 *m/z* isolation window, and in-cell isolation was performed to remove peaks other than the peak corresponding to the proteoform of interest. The proteoform of interest was fragmented by electron capture dissociation (ECD). The excitation power for in-cell isolation was set to 1% with a pulse time of 0.02 ms. The electron energy and reaction time for ECD were determined on a case-by-case basis to achieve optimal

fragmentation. The transients from approximately 1000-4000 scans were summed to obtain high-quality spectra for the localization of protein PTMs.

2.2.7 Data analysis

All LC-MS data were processed and analyzed using the DataAnalysis software (version 3.2; Bruker Daltonics). Mass spectra were averaged over the retention time window wherein all proteoforms of the same protein eluted. The spectra were deconvoluted using the Maximum Entropy algorithm incorporated in the DataAnalysis software. The resolving power for Maximum Entropy deconvolution was set to 60,000 for proteins below 50 kDa, which were isotopically resolved by the mass spectrometer. Most-abundant mass was reported for all MS data, and monoisotopic mass was reported for all MS/MS data. The abundance of a particular proteoform is reported as the ratio of the peak heights of the proteoform to the summed peak heights of all proteoforms of the same protein. The percentages of the mono- ($\%P_{mono}$) and/or bis- ($\%P_{bis}$) phosphorylated proteoforms were defined as the summed abundances of mono- and/or bis-phosphorylated species over the summed abundances of the entire protein population, respectively. Based on these percentages, the total amount of phosphorylation (P_{total}; mol of Pi/mol of protein) of a single protein was calculated using the following equation:

$$P_{total}=\%P_{mono}+2\times\%P_{bis}$$

Tandem mass spectra were output from the DataAnalysis software and analyzed using MASH Suite Pro [229, 255] software developed in-house. The spectra were deconvoluted with a signal-to-noise ratio of 3 and a cutoff fit score of 60%. All the program-processed data were manually validated to obtain accurate sequence and PTM information.

2.2.8 Western blotting analysis

Temperature treatment of the tissues and homogenization with HEPES extraction buffer were described above. The remaining pellets were homogenized using 10 vol. extraction buffer (50 mM Tris-HCl, 50 mM NaCl, 1% Triton X-100, 0.5% SDS, 1 mM PMSF, 1 mM Na_3VO_4, 2.5 mM EDTA). Homogenization was performed in the cold room and took less than 30 s per sample. After centrifugation, the tissue extracts were analyzed by Western blotting to evaluate the phosphorylation of cTnI and cMyBP-C. The following primary antibodies were used for the analysis: rabbit polyclonal against phosphorylated Ser22/23 of cTnI (Cell Signaling Technology; cat. #4004), mouse monoclonal against cTnI (Abcam; cat. ab10231), rabbit polyclonal against PKA substrate (Cell Signaling Technology; cat. 9624) for probing cMyBP-C phosphorylation, and rabbit polyclonal against cMyBP-C (Thermo Fisher Scientific; cat. PA5-41993). Re-probing of the PVDF membrane with different antibodies was performed as described previously [256]. The detection of cTnI and phosphorylated cTnI was performed using a bi-color fluorescent detection method with the following secondary antibodies: IRDye 800CW anti-mouse IgG (LI-COR Biosciences; cat. 925-32210) and IRDye 680RD anti-rabbit IgG (LI-COR Biosciences; cat. 926-68070). The detection of cMyBP-C and phosphorylated cMyBP-C (PKA substrate) was by enhanced chemiluminescence. The images were taken using a LI-COR Odyssey imager (LI-COR Biosciences), and the quantification of signals were performed using the Image Studio Lite software (Version 5.2) (LI-COR Biosciences).

2.2.9 Enzyme-linked immunosorbent assay (ELISA) for quantitative analysis of cAMP (cyclic adenosine monophosphate)

Approximately 4-6 mg of tissue was excised on a pre-cooled metal plate and incubated at 4 °C for 1 min, 5 min, 10 min, or 30 min. The tissue was immediately homogenized in 10 vol. of

0.1 M HCl and centrifuged at 19,000 rcf for 5 min. The supernatants were diluted 5-fold using 0.1 M HCl for cAMP analysis, or 50-fold using water for Bradford protein assay. cAMP analysis was performed using a direct cAMP ELISA kit in accordance with the manufacturer's instructions (Enzo Life Sciences, Farmingdale, NY, USA). Briefly, cAMP standards of 0.78, 3.13, 12.5, 50, and 200 pmol/mL were prepared in 0.1 M HCl. The cAMP standards and the tissue extracts were added to wells coated with a GxR IgG antibody. The solution of cAMP conjugated to alkaline phosphatase (AP) was added and followed by addition of rabbit polyclonal antibody to cAMP. The mixtures were incubated at room temperature for 2 hrs to allow for the binding of the cAMP antibody to cAMP in the samples or cAMP-AP conjugates (competitors). After washing, the AP substrate, para-Nitrophenylphosphate (pNpp), was added to assay the amount of cAMP-AP conjugates, followed by addition of the stopping solution and photometric measurement at 405 nm. The lower amount of cAMP-AP conjugates indicated a higher amount of free cAMP in the sample. The Bradford protein assay was carried out to measure the total protein concentration, and the cAMP concentration was normalized to total protein concentration.

2.2.10 Statistical analysis

Three extraction replicates at time 0 were used as controls (control group), and two-way ANOVA was performed to evaluate the statistical significance of variation between groups. Differences between means were considered to be statistically significant at $p < 0.05$.

2.3 RESULTS

2.3.1 Increased phosphorylation of cTnI and Enigma Homolog Isoform 2 (ENH2) within minutes following increased temperature

We first assessed the reproducibility of RPC separation of proteins and the stability of the instrument response. Injection and extraction replicates were nearly identical with little variability in the detected relative abundances of sarcomeric protein proteoforms (**Figure 2.1, 2.2**), confirming the high reproducibility of the method. Next, to assess the impact of temperature on sarcomeric protein PTMs, we incubated pieces of tissue at temperatures mimicking those that might occur during tissue shipment (22 ^{o}C) or sample processing (4 ^{o}C). Following incubation at either 4 °C or 22 °C and extraction of the sarcomeric proteins, high-resolution top-down LC-MS analysis was carried out (**Figure 2.3**). These samples were compared to the samples extracted immediately before tissue defrosting (referred as the control group) to evaluate temperature-related changes in protein PTMs. Additionally, the effects of buffer and different storage conditions on the sarcomeric protein PTMs were investigated as outlined in **Figure 2.3a**.

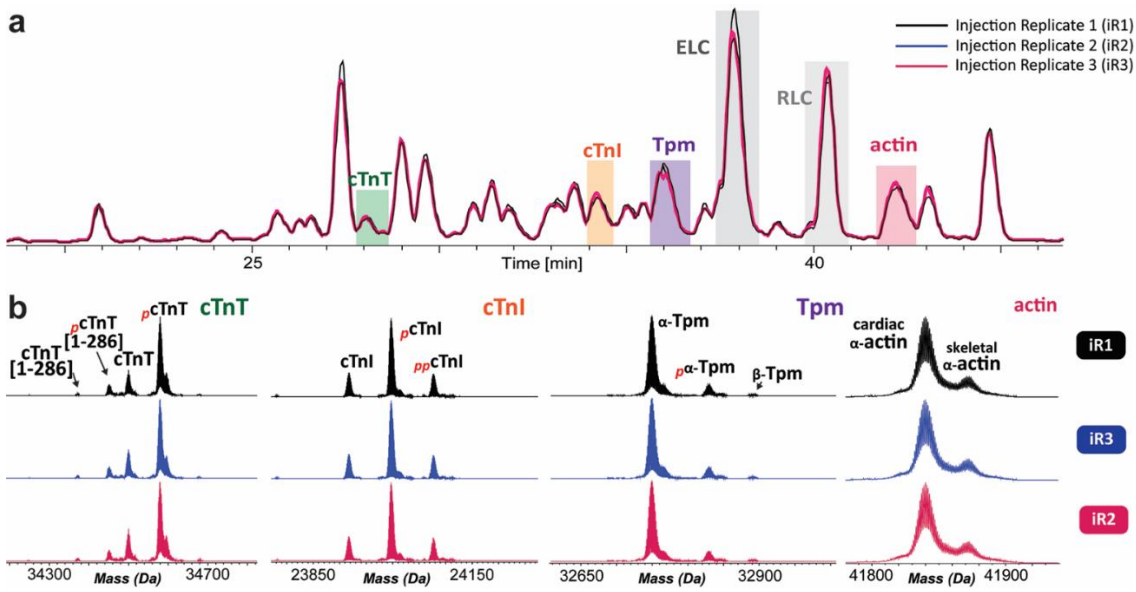


Figure 2.1. Three injection replicates demonstrate high reproducibility of protein separation and instrument performance. (a) Overlaid base peak chromatograms (BPCs) of three injection replicates (iR) were nearly identical. (b) Mass spectra showing relative abundances of the cTnT, cTnI, Tpm, and actin proteoforms demonstrated consistency of the instrument performance.

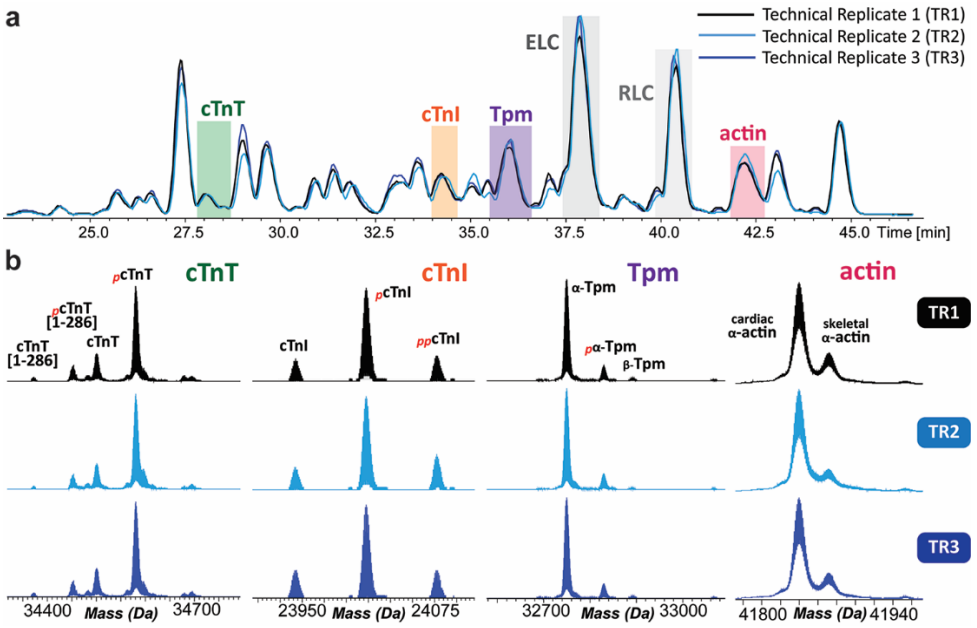


Figure 2.2. Reproducibility of protein extraction for quantification of protein PTMs. (a) Overlaid base peak chromatograms (BPCs) of three sarcomeric protein extracts showed high reproducibility. (b) Mass spectra showing relative abundances of the cTnT, cTnI, Tpm, and actin proteoforms demonstrated stable experimental performance.

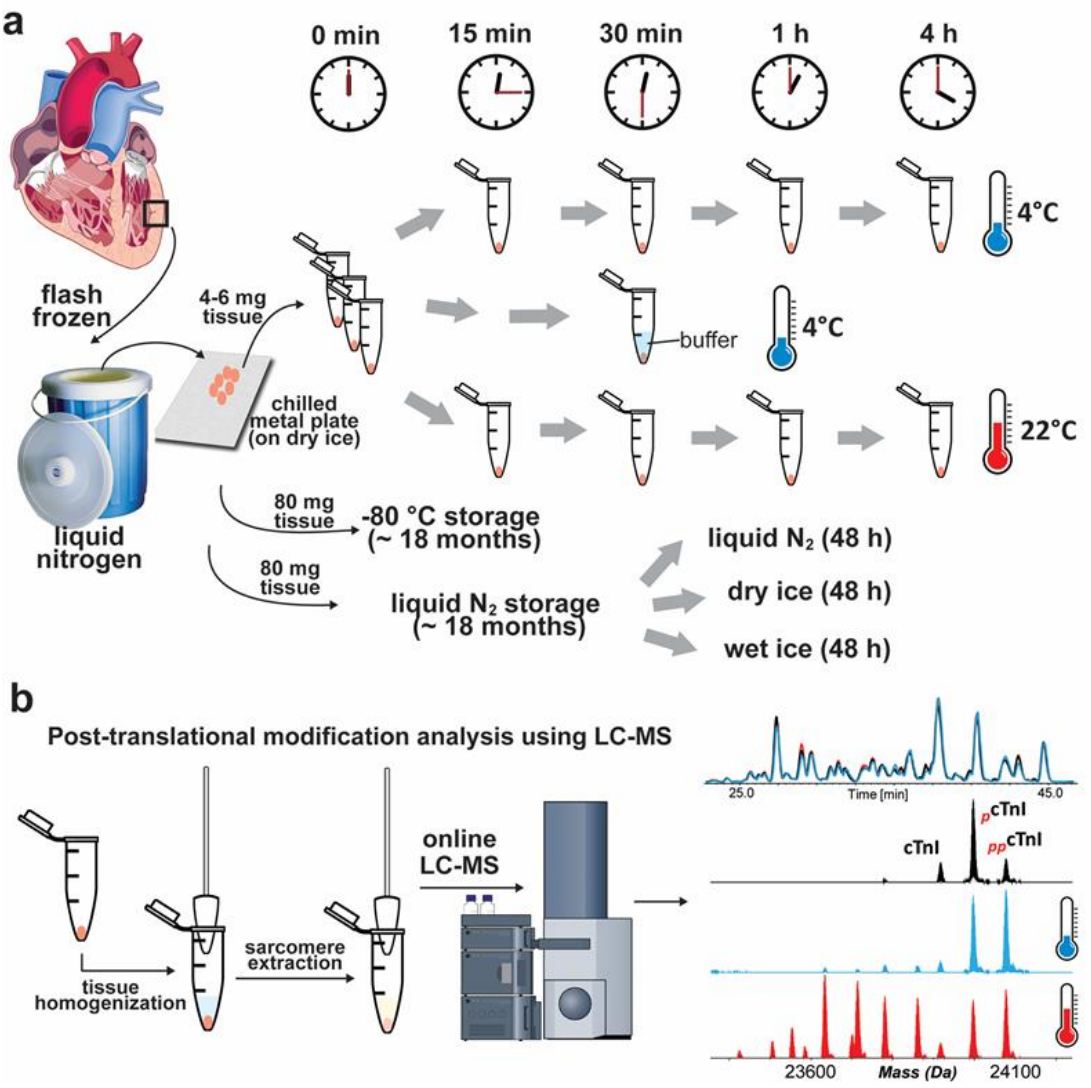


Figure 2.3. Schematic for assessing the effect of temperature on sarcomeric protein PTM changes. (a) Small pieces of cardiac tissue were kept at either 4 °C or 22 °C for various durations followed by tissue homogenization and protein extraction. Additionally, tissue was also stored in liquid nitrogen (N2) or at -80 °C for approximately 18 months to evaluate changes in protein PTMs under different storage conditions. Small pieces of tissues were also stored using liquid nitrogen (N2), dry ice and wet ice for 48 h to mimic the shipment and sample handling conditions. (b) Sarcomeric protein extracts were analyzed by high-resolution top-down LC-MS.

Following temperature treatment, prominent changes were observed in the phosphorylation of cTnI and ENH2, which is a Z-disc protein (**Figure 2.4**). Within 15 minutes of tissue incubation at 4 °C, the most abundant MS peak of cTnI changed from the mono-phosphorylated to bis-

phosphorylated proteoform. The relative abundance of bis-phosphorylated cTnI increased by approximately 150% within 15 min, and by about 230% after 1 h at 4 °C (**Figure 2.4a**). cTnI proteoforms resulting from removal of 2-4 amino acids from the C-terminus of the protein were detected within 30 min of incubation at 4 °C (**Figure 2.4a**). Similar to the tissue maintained at 4 °C, cardiac tissue thawed at 22 °C exhibited an increase in cTnI phosphorylation and degradation within 15 min (**Figure 2.4b**). Degraded proteoforms of cTnI with removal of 2 or 3 C-terminal amino acids became the dominant proteoforms after incubating the tissue at 22 °C for 1 h (**Figure 2.4b**). The high-resolution mass spectra of the cTnI proteoforms and their accurate molecular weights are shown in **Figure 2.5**.

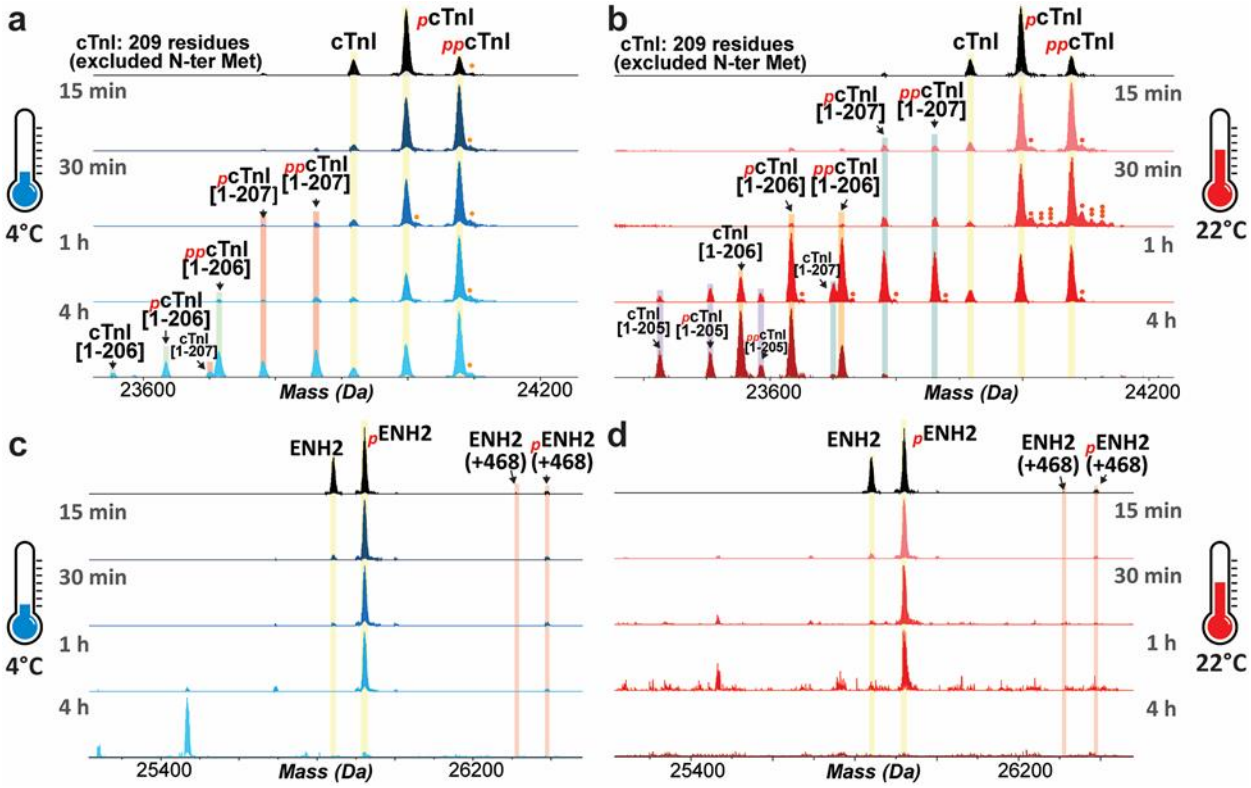


Figure 2.4. Rapid increase in the relative abundances of phosphorylated and degraded proteoforms of cTnI and ENH2 following tissue incubation at either 4 ºC or 22 ºC. Top-down mass spectra show that hyper-phosphorylation and truncation of cTnI occurred after tissue incubation at 4 °C (a) or 22 °C (b). Increased phosphorylation and degradation of ENH2 upon tissue defrosting at 4 °C (c) or 22 °C (d). Single,

double, and triple circles denote mono-oxidated, bis-oxidated and tri-oxidated protein proteoforms, respectively.

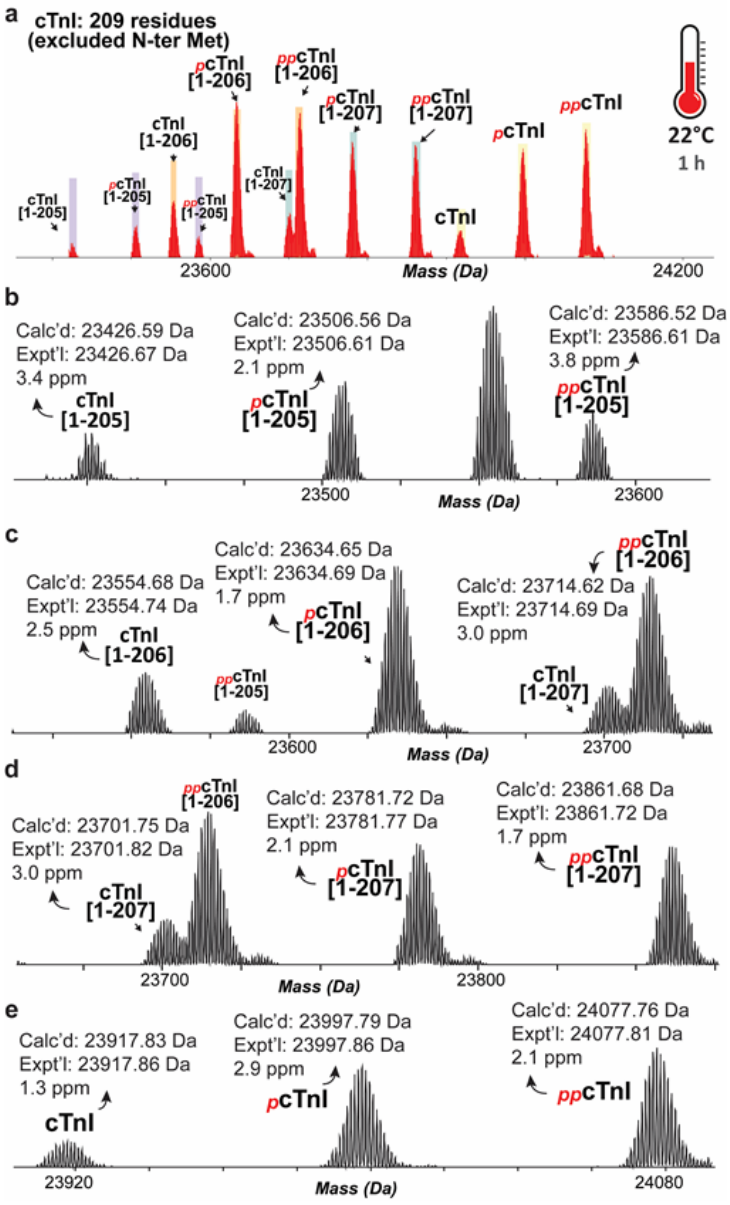


Figure 2.5. Representative mass spectra of the cTnI proteoforms detected by top-down proteomics. (a) Top-down mass spectra showing hyper-phosphorylation and truncation of cTnI upon tissue incubation at 22 °C for 1 hr. Representative mass spectra and accurate molecular weights of truncated cTnI proteoforms with removal of 3 (b), 2 (c), or 1 (d) C-terminal amino acids and intact cTnI proteoforms (e). Calc'd, calculated molecular weight based on cTnI sequence (with removal of N-terminal Met, and N-terminal acetylation). Expt'l, molecular weight obtained from the high-resolution mass spectrum. All molecular weights indicated are the most-abundant masses of the corresponding proteoforms. ppm, parts per million.

In addition to increased cTnI phosphorylation, ENH2 phosphorylation also increased upon upon temperature increase to 4 °C or 22 °C (**Figure 2.4c, d**). The high-resolution mass spectra of the ENH2 proteoforms and their accurate molecular weights are shown in **Figure 2.6**. Interestingly, even though degraded ENH2 proteoforms were not observed within the same elution time window (37.3-39.1 min), there was a progressive decline in the MS signal for ENH2 in the defrosted tissue after 1 h at 4 °C or 30 min at 22 °C. The intact ENH2 became nearly undetectable after 4 hrs at 4 °C or 1 h at 22 °C, indicating that ENH2 degradation occurred during the process of tissue defrosting. The degraded products of ENH2 may be small peptides that eluted at different retention time windows due to differences in hydrophobicity. We also observed two additional proteoforms of ENH2; one with an increase of 468 Da relative to the mass of the un-phosphorylated proteoform, and the second with a 468 Da mass increase relative to mono-phosphorylated ENH2 (**Figure 2.4c, d**). Due to the low abundance of these proteoforms, the sequences of these proteoforms were not further characterized.

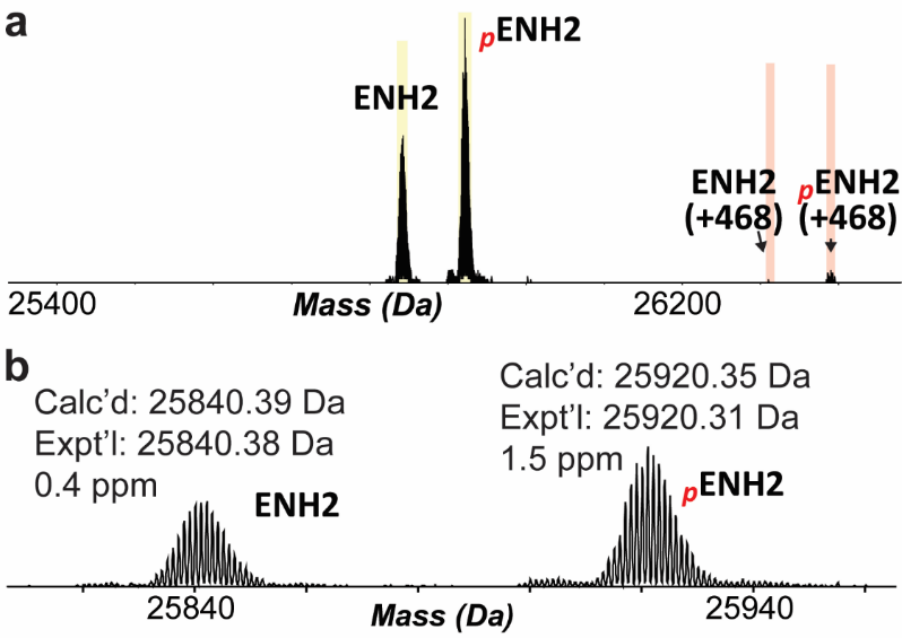


Figure 2.6. Representative mass spectra of the ENH2 proteoforms detected by top-down proteomics. (a) Top-down mass spectra showing the ENH2 proteoforms. (b) Representative mass spectra and accurate

molecular weights of ENH2 and $_p$ENH2. Calc'd, calculated molecular weight based on ENH2 sequence (with removal of N-terminal Met, and N-terminal acetylation). Expt'l, molecular weight obtained from the high-resolution mass spectrum. All molecular weights indicated are the most-abundant masses of the corresponding proteoforms. ppm, parts per million.

2.3.2 Decreased phosphorylation of cTnT and RLC following tissue incubation at 4 °C or 22 °C

In contrast to cTnI and ENH2 phosphorylation, which increased following incubation at either 4 °C or 22 °C, the phosphorylation of cTnT and RLC decreased in tissue kept at these temperatures (**Figure 2.7**). The decline in cTnT phosphorylation was mild when the tissue was maintained at 4 °C, but severe at 22 °C (**Figure 2.7a, b**). Particularly, cTnT phosphorylation in myocardial tissue decreased by approximately 70% when incubated at 22 °C for 4 h (**Figure 2.7b**). In addition to de-phosphorylation, degradation of cTnT was also observed in the cardiac tissue maintained at 22 °C for over 1 h (**Figure 2.7b, Figure 2.8**). The abundance of the C-terminally truncated cTnT proteoforms relative to the intact cTnT adult isoform was about 1:10 and 1:1 when the tissue was maintained at 22 °C for 1 h and 4 h, respectively (**Figure 2.7b**).

Even though RLC phosphorylation was low (less than 10%) in the control group, the decline in RLC phosphorylation was apparent after 30 min of tissue incubation at 4 °C, or 15 min incubation at 22 °C (**Figure 2.7c, d**). The accurate molecular weights of RLC proteoforms are shown in **Figure 2.9**. With prolonged temperature change (4 h), we also observed a slight increase in the phosphorylation of cardiac α-Tpm (formally known as Tpm1.1st^{257}); however, the phosphorylation of cardiac β-Tpm (formally known as Tpm2.2st^{257}) appeared to decrease to the extent that it was undetectable after tissue incubation for 4 h at either 4 °C or 22 °C (**Figure 2.10**).

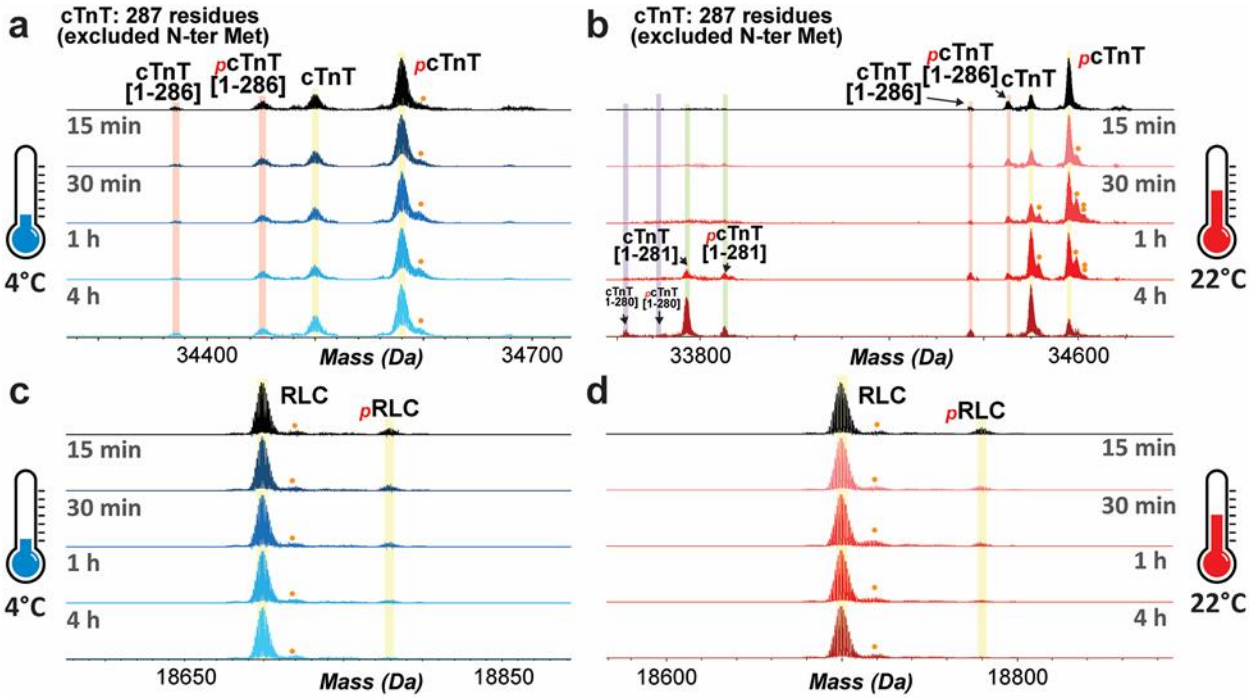


Figure 2.7. Decrease in the relative abundances of phosphorylated proteoforms of cTnT and RLC following tissue incubation at either 4 ºC or 22 ºC. Top-down mass spectra of cTnT proteoforms that show slightly decreased phosphorylation upon tissue incubation at 4 °C (a), and severe de-phosphorylation and degradation of cTnT occurred at 22 °C (b). Decreased phosphorylation of RLC was observed with tissue warm-up at either 4 °C (c) or 22 °C (d). Single and double circle denote mono- and bis-oxidized proteoforms, respectively.

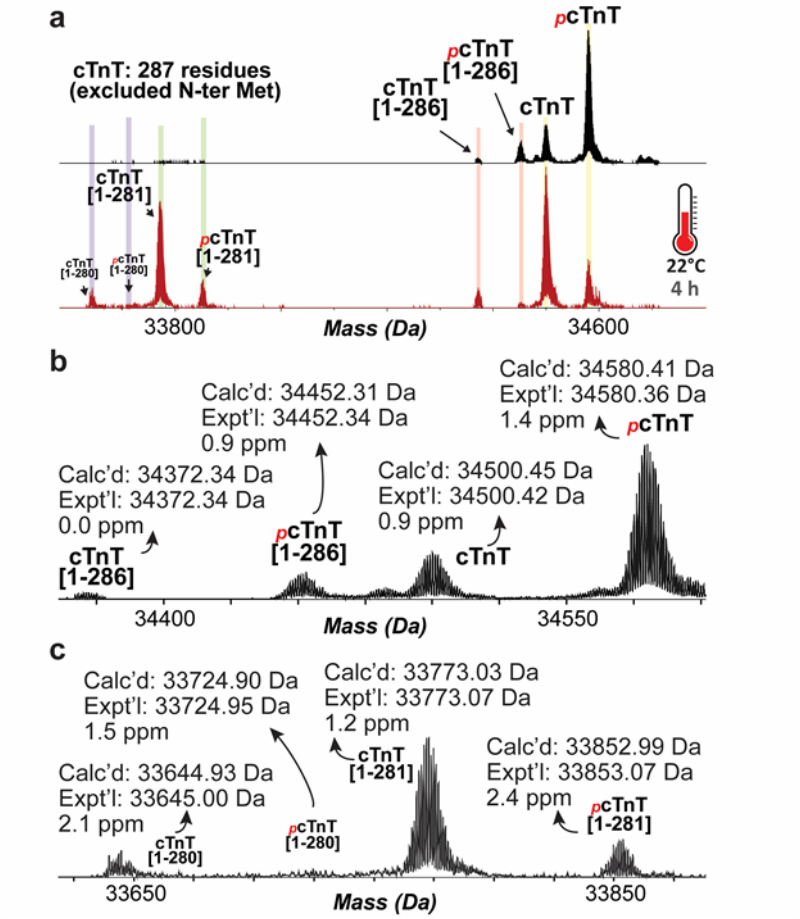


Figure 2.8. Representative mass spectra of the cTnT proteoforms detected by top-down proteomics. (a) Top-down mass spectra showing de-phosphorylation and truncation of cTnT upon tissue incubation at 22 °C for 4 h. Representative mass spectra and accurate molecular weights of intact cTnT and truncated cTnT proteoforms with removal of 1 C-terminal amino acid (b), and of truncated cTnT proteoforms with removal of 6 and 7 C-terminal amino acids (c). Calc'd, calculated molecular weight based on cTnT sequence (with removal of N-terminal Met, and N-terminal acetylation). Expt'l, molecular weight obtained from the high-resolution mass spectrum. All molecular weights indicated are the most-abundant masses of the corresponding proteoforms. ppm, parts per million.

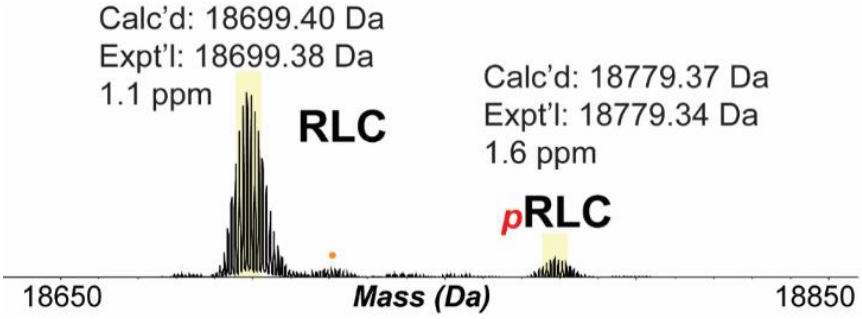


Figure 2.9. Representative mass spectra and accurate molecular weights of the RLC proteoforms. Calc'd, calculated molecular weight based on RLC sequence (with removal of N-terminal Met, and N-terminal acetylation). Expt'l, molecular weight obtained from the high-resolution mass spectrum. All

molecular weights indicated are the most-abundant masses of the corresponding proteoforms. ppm, parts per million.

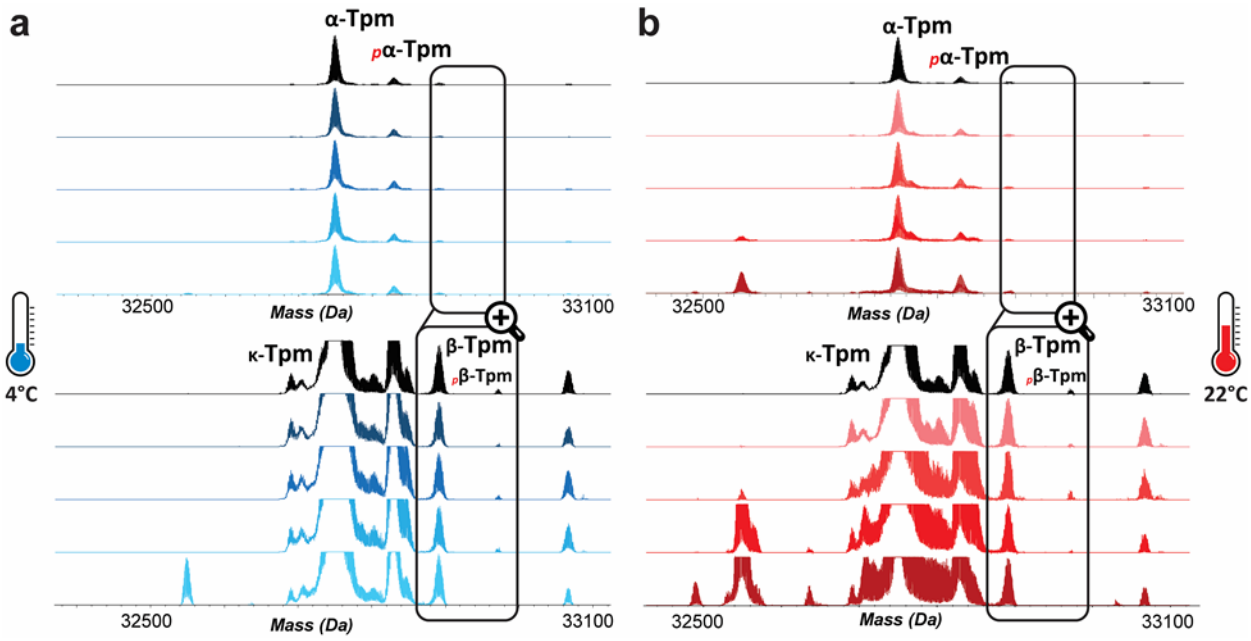


Figure 2.10. Slightly increased cardiac α-Tpm phosphorylation and decreased cardiac β-Tpm phosphorylation following tissue incubation at either 4 ºC or 22 ºC. Top-down mass spectra showing slight increase in the phosphorylation of cardiac α-Tpm and decreased phosphorylation of cardiac β-Tpm during tissue defrosting at 4 °C (a) and 22 °C (b).

As summarized in **Figure 2.11**, the phosphorylation of major myofilament proteins were differentially affected by tissue incubation at either 4 ºC or 22 ºC. While incubation at either 4 ºC or 22 ºC had a significant effect on all of the myofilament proteins analyzed, cTnI and ENH2 were affected to the greatest extent (**Figure 2.11a, b**). cTnT phosphorylation is relatively stable within 1 hr of temperature increase at 4 °C (**Figure 2.11c**), and RLC phosphorylation remains relatively unchanged within 15 min of tissue incubation at 4 °C (**Figure 2.11d**), but decreased significantly after 15 min at 4 °C or 22 °C.

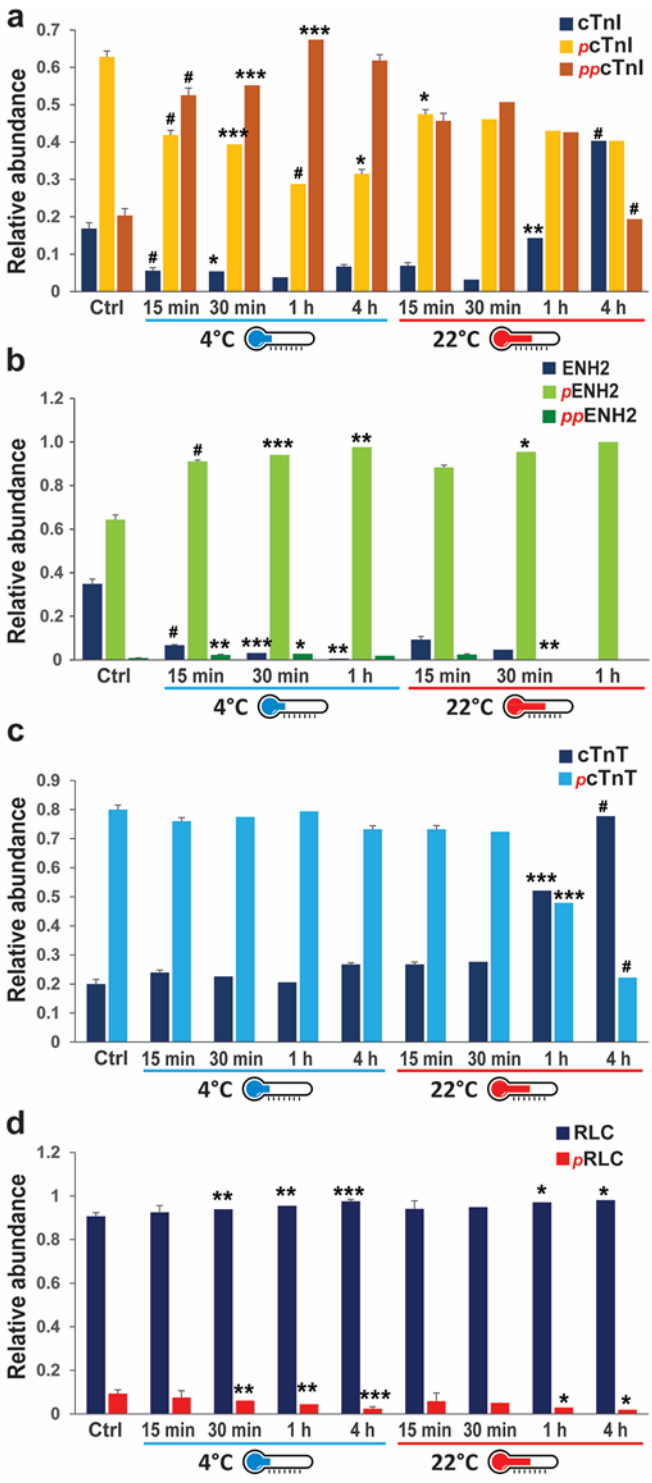


Figure 2.11. Temperature-related changes in the relative abundances of un-phosphorylated and phosphorylated sarcomeric protein proteoforms. Quantification of the relative abundances of the cTnI (a), ENH2 (b), cTnT (c), and RLC (d) proteoforms during tissue incubation at either 4 oC or 22 oC. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, # $p < 0.0001$ by two-way ANOVA.

2.3.3 Localization of the sites of phosphorylation in cTnI by ECD

While the observed decreases in the phosphorylation of cTnT and RLC likely occurred at the canonical sites, Ser1 and Ser14, respectively,[258, 259] increased phosphorylation of cTnI and ENH2 was intriguing. Particularly, cTnI has multiple sites of phosphorylation that can be targeted by various kinases, leading to dramatically different functional impacts on contractility.[245] In addition, ENH2 was only recently discovered as a phosphoprotein in a swine model of acute myocardial infarction,[24] and the site of ENH2 in human cardiac tissue has not been identified. To localize the sites of phosphorylation in cTnI and ENH2, we performed MS/MS experiments using an ultra-high-resolution 12 Tesla FT-ICR mass spectrometer. Mono- and bis-phosphorylated proteoforms of cTnI, as well as the mono-phosphorylated proteoform of ENH2, were fragmented by ECD, which preserves labile PTMs such as phosphorylation.[147, 244, 260]

In the case of $_{p}$cTnI fragmentation, after taking into account the removal of N-terminal Met, the masses of the N-terminal fragment ions (***c*** ions) prior to $\boldsymbol{c_{20}}$ exhibited an increase of 42.01 Da relative to the predicted masses for these fragment ions, indicating that the N-terminus is modified by acetylation (**Figure 2.12**), consistent with our previous reports.[157] The N-terminal fragments prior to $\boldsymbol{c_{20}}$ exclusively exhibit an increase of 42.01 Da without an additional 79.97 Da increase in mass (**Figure 2.12**), suggesting that the site of phosphorylation is localized to the C-terminus of amino acid 20. The N-terminal $\boldsymbol{c_{22}}$ ion has an increase of 79.97 Da in mass in addition to acetylation **(Figure 2.12)**, suggesting that Ser22 is phosphorylated. In addition, the fact that both un-phosphorylated and mono-phosphorylated $\boldsymbol{c_{22}}$ were observed (**Figure 2.12**) suggests Ser22 is partially phosphorylated, and that $_{p}$cTnI contains a mixed positional isomers with Ser22 being one of the phosphorylation sites. Given that $\boldsymbol{c_{20}}$ is exclusively un-phosphorylated, and $\boldsymbol{c_{24}}$ is exclusively mono-phosphorylated without detectable un-phosphorylated ion (**Figure 2.12b, d**), all

existing phosphorylation sites are between amino acid residues 20-24 (-Arg-Arg-Ser-Ser-Asn-), and therefore, the site of phosphorylation on $_{p}$cTnI is localized to either Ser22 or Ser23 (**Figure 2.12e**). A total of 59 ***c*** ions (N-terminal fragment ions) and 71 $z^{\cdot}$ ions (C-terminal fragment ions) from a single tandem mass spectrum were matched to the sequence of $_{p}$cTnI, which represents cleavage of 61% of all inter-residue bonds.

Similar to the localization of phosphorylation site on $_{p}$cTnI, the sites of phosphorylation on $_{pp}$cTnI was localized to Ser22 and Ser23 (**Figure 2.13**) with a total of 45 ***c*** and 70 $z^{\cdot}$ ions from a single tandem mass spectrum, representing cleavage of 56% intra-residue bonds. We further validated the increased cTnI phosphorylation by Western blotting using a specific antibody against phosphorylated (Ser22/23) cTnI (**Figure 2.13f**). We observed significant increase in the phosphorylation of cTnI at Ser22/23 following temperature treatment at 4 ^{o}C or 22 ^{o}C (**Figure 2.13f, g**), which was consistent with the MS results (**Figure 2.11**). However, the extent of cTnI phosphorylation change appeared different between the two quantification methods. This could result from different detection limits and linear range of the two methods. cTnI phosphorylation by Western blotting, when expressed in log scale, appeared more consistent with the quantification results by MS (**Figure 2.13h**).

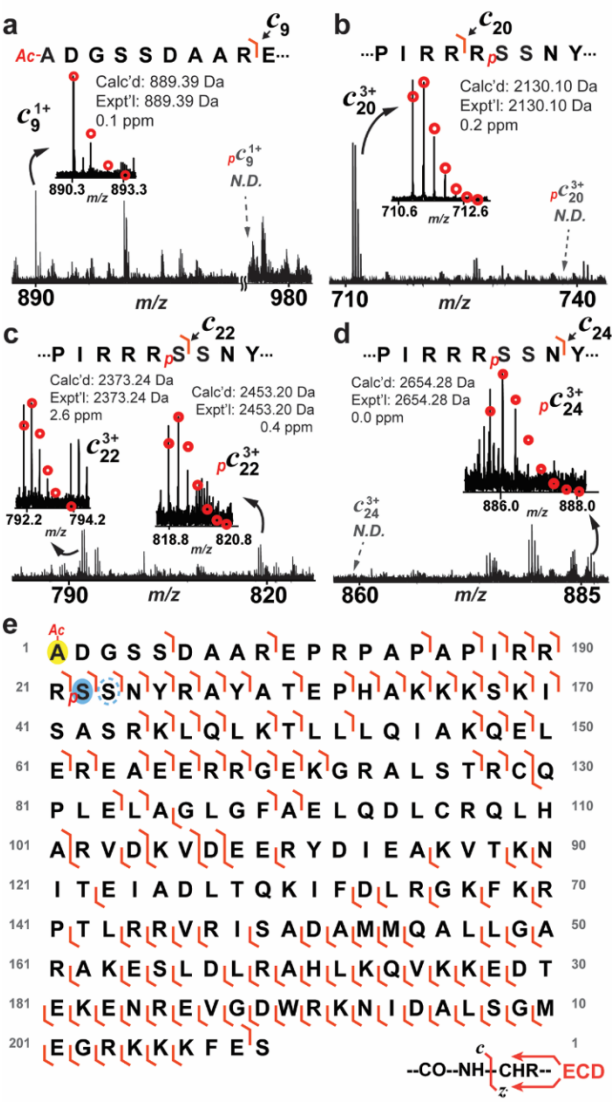


Figure 2.12. Localization of the phosphorylation site in $_p$cTnI by top-down MS/MS to either Ser22 or Ser23, representing positional isomers of $_p$cTnI . (a) and (b) Zoomed-in spectra for $\boldsymbol{c_9}$ and $\boldsymbol{c_{20}}$ ions, respectively, confirmed N-terminal acetylation of the $_p$cTnI. N-terminal fragments prior to $\boldsymbol{c_{20}}$ had no detectable mono-phosphorylated counterparts. (c) Zoomed-in spectrum for $\boldsymbol{c_{22}}$ ion showed detectable un-phosphorylated and mono-phosphorylated ions, suggesting that Ser22 is partially phosphorylated. (d) Zoomed-in spectrum for $\boldsymbol{c_{24}}$ showed that $\boldsymbol{c_{24}}$ ions were exclusively mono-phosphorylated without detectable un-phosphorylated ions. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned mass. N.D. not detected. (e) Fragmentation map shows the fragment ions that were matched to the sequences of $_p$cTnI (59 $\boldsymbol{c}$ and 71 $\boldsymbol{z}^{\cdot}$ ions total). Ac denotes acetylation. *p* denotes phosphorylation.

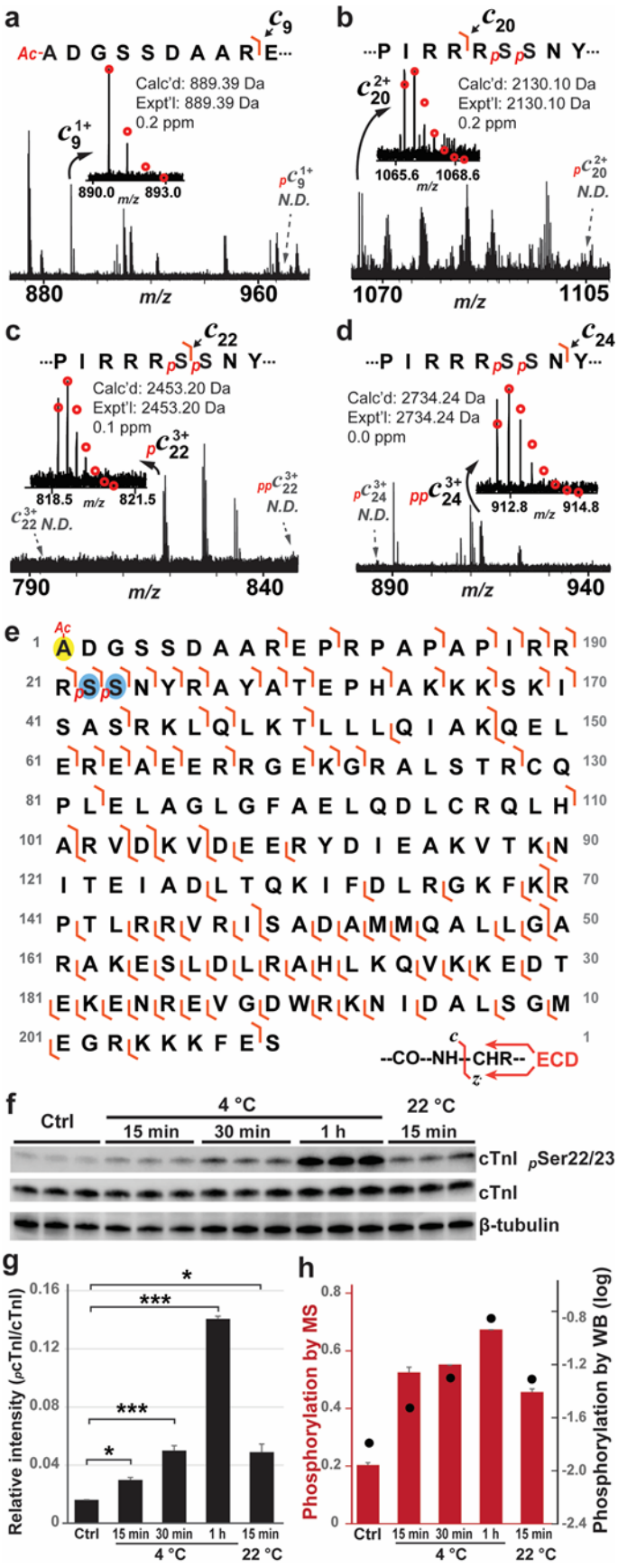

a
Ac-A D G S S D A A R E···
Calc'd: 889.39 Da
Expt'l: 889.39 Da
0.2 ppm
890.0 m/z 893.0
N.D.
880 m/z 960
b
···P I R R RpSpS N Y···
Calc'd: 2130.10 Da
Expt'l: 2130.10 Da
0.2 ppm
1065.6 1068.6 m/z
N.D.
1070 m/z 1105
c
···P I R R RpSpS N Y···
Calc'd: 2453.20 Da
Expt'l: 2453.20 Da
0.1 ppm
818.5 m/z 821.5
N.D.
N.D.
790 m/z 840
d
···P I R R RpSpS N Y···
Calc'd: 2734.24 Da
Expt'l: 2734.24 Da
0.0 ppm
N.D.
912.8 m/z 914.8
890 m/z 940
e
Ac A D G S S D A A R E P R P A P A P I R R
R pS pS N Y R A Y A T E P H A K K K S K I
S A S R K L Q L K T L L L Q I A K Q E L
E R E A E E R R G E K G R A L S T R C Q
P L E L A G L G F A E L Q D L C R Q L H
A R V D K V D E E R Y D I E A K V T K N
I T E I A D L T Q K I F D L R G K F K R
P T L R R V R I S A D A M M Q A L L G A
R A K E S L D L R A H L K Q V K K E D T
E K E N R E V G D W R K N I D A L S G M
E G R K K K F E S
--CO--NH--CHR-- ECD
f
Ctrl
4 °C
15 min
30 min
1 h
22 °C
15 min
cTnI pSer22/23
cTnI
β-tubulin
g
Relative intensity (pcTnI/cTnI)
Ctrl 15 min 30 min 1 h 15 min
4 °C 22 °C
h
Phosphorylation by MS
Phosphorylation by WB (log)
Ctrl 15 min 30 min 1 h 15 min
4 °C 22 °C

Figure 2.13. Localization of the phosphorylation sites in $_{pp}$cTnI by top-down MS/MS. (a) and (b) Zoomed-in spectra for $\boldsymbol{c_9}$ and $\boldsymbol{c_{20}}$ ions, respectively, confirmed N-terminal acetylation of the $_{pp}$cTnI. N-terminal fragments prior to $\boldsymbol{c_{20}}$ had no detectable mono-phosphorylated counterparts. (c) Zoomed-in spectrum for $\boldsymbol{c_{22}}$ ion showed only mono-phosphorylated ions, suggesting that amino acid sequence N-terminal to Ser22 is exclusively mono-phosphorylated. (d) Zoomed-in spectrum for $\boldsymbol{c_{24}}$ showed that $\boldsymbol{c_{24}}$ ions were exclusively bis-phosphorylated without detectable un-phosphorylated or mono-phosphorylated ion, suggesting that amino acid sequence prior to Ser24 contains both phosphorylation sites. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned mass. N.D. not detected. (e) Fragmentation map shows the fragment ions that were matched to the sequences of $_{pp}$cTnI (45 $\boldsymbol{c}$ and 70 $\boldsymbol{z^{\cdot}}$ ions total). Ac denotes acetylation. *p* denotes phosphorylation. (f) Western blotting analysis to confirm increased of cTnI phosphorylation at sites of Ser22/23. (g) Histogram showing quantification of phosphorylated cTnI by Western blotting. Two-way ANOVA test was performed to evaluate the statistical significance of difference. * $p < 0.05$, *** $p < 0.001$. (h) Comparison of relative abundance of bis-phosphorylated cTnI quantified by MS (red bars) and phosphorylated cTnI quantified by Western blotting in log scale (black dots).

2.3.4 Localization of the phosphorylation site in ENH2 by ECD

With the removal of the N-terminal Met, all $\boldsymbol{c}$ ions had a 42.01 Da increase in mass compared with the predicted fragment ions, indicating acetylation of the ENH2 N-terminus (**Figure 2.14a**), which is consistent with the finding in swine ENH2.[24] All N-terminal fragments prior to $\boldsymbol{c_{114}}$ were exclusively un-phosphorylated (**Figure 2.14a, b**), suggesting that the site of phosphorylation is localized after residue 114. The fact that the $\boldsymbol{c_{118}}$ and $\boldsymbol{c_{120}}$ ions are exclusively mono-phosphorylated without detectable un-phosphorylated counterparts (**Figure 2.14c, d**) suggests that the phosphorylation site is exclusively localized to the N-terminus of residual 118. Therefore, the phosphorylation site in $_{p}$ENH2 is localized between amino acid residues 114 and 118 (-Gln-Arg-Arg-Gly-Ser-). Since Ser118 is the only possible site of phosphorylation, we assigned Ser118 as the sole site of phosphorylation in human ENH2, which is consistent with the ENH2 phosphorylation site observed in swine.[24] In addition, $\boldsymbol{c_{118}}$ is exclusively mono-phosphorylated without detectable un-phosphorylated ions. This further proved that Ser118 is the only site of phosphorylation in $_{p}$ENH2. A total of 78 $\boldsymbol{c}$ ions and 52 $\boldsymbol{z^{\cdot}}$ ions from two tandem mass

spectra were matched to the sequence of $_p$ENH2, accounting for 52% of inter-residue bond cleavages.

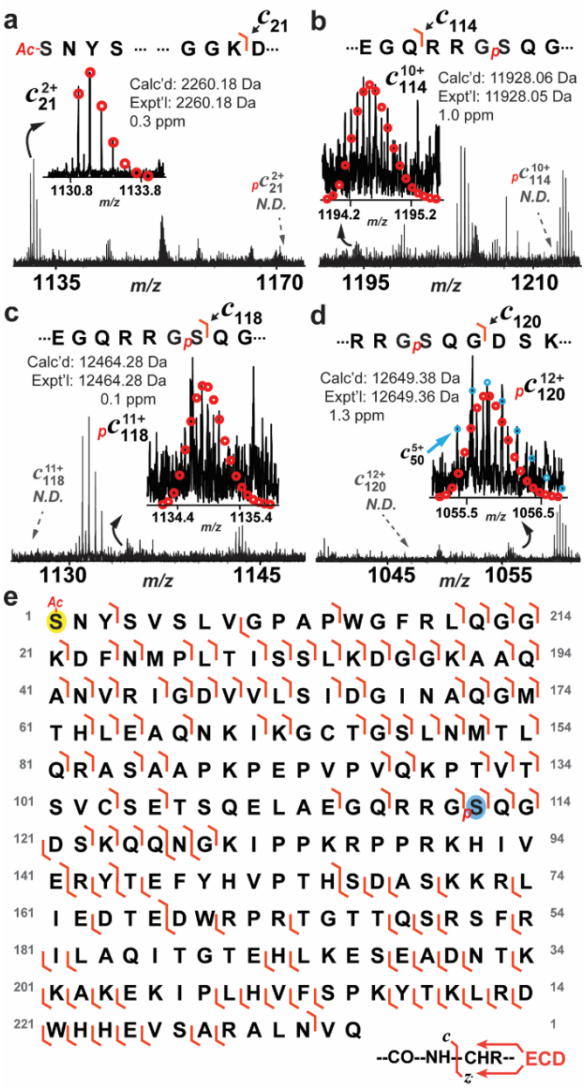


Figure 2.14. Localization of the phosphorylation site in $_p$ENH2 by top-down MS/MS. (a) Zoomed-in spectrum for $\boldsymbol{c_{21}}$ confirmed N-terminal acetylation of $_p$ENH2. (b) N-terminal fragments prior to $\boldsymbol{c_{114}}$ had no detectable mono-phosphorylated counterparts, suggesting that the phosphorylation site was located to the C-terminal of Gln114. (c) and (d) Zoomed-in spectra for $\boldsymbol{c_{118}}$ and $\boldsymbol{c_{120}}$ ion, respectively, showed only mono-phosphorylated ions, suggesting that amino acid sequence N-terminal to Ser118 is exclusively mono-phosphorylated. Circles represent the theoretical isotopic abundance distribution of the isotopomer peaks corresponding to the assigned mass. N.D. not detected. (e) Fragmentation map shows the fragment ions that were matched to the sequences of $_p$ENH2 (78 $\boldsymbol{c}$ and 52 $\boldsymbol{z}^{\cdot}$ ions total). Ac denotes acetylation. *p* denotes phosphorylation.

2.3.5 cAMP increased within 5 min of tissue incubation at 4 ^{o}C

Given that Ser22/23 of cTnI are well-known target sites of protein kinase A (PKA), increased phosphorylation at these sites following tissue incubation at 4 ^{o}C or 22 ^{o}C suggests that PKA might be activated in the warm-up myocardium. To determine if PKA was activated, we examined the concentration of cAMP in tissue incubated at 4 ^{o}C. The tissue extracts were incubated with primary antibodies against cAMP, as well as cAMP-akaline phosphatase conjugates (cAMP-AP) (**Figure 2.15a**). Free cAMP in the sample competed with cAMP-AP for antibody binding in a concentration-dependent manner and, therefore, the free cAMP concentration was inversely proportional to the cAMP-AP concentration, which is measured via a colorimetric reaction with an AP substrate, para-Nitrophenylphosphate (pNpp), (**Figure 2.15a**). The cAMP concentration was normalized to total protein concentration prior to comparison between different groups.

Based on the ELISA assays, cAMP concentration (nmol/ µg protein) increased dramatically upon tissue incubation at 4 ^{o}C (**Figure 2.15b**). The concentration of cAMP increased by nearly 50% within 5 min of tissue defrosting at 4 °C, and by about 100% after 30 min (**Figure 2.15b**). The increase in cAMP concentration in the myocardium supports activation of PKA in tissue incubated at 4 ^{o}C. This is highly consistent with the increased PKA-mediated phosphorylation of cTnI at Ser22/23 by both MS/MS and Western blotting analysis (**Figure 2.15**). To further validate increased PKA activity, we performed Western blotting analysis to evaluate the phosphorylation of cMyBP-C, a well-established substrate of PKA, using a PKA-substrate antibody. With 5 µg of total proteins being analyzed, this PKA-substrate antibody detected one major band between 130 and 250 kDa, and one weak band between 15 and 25 kDa (**Figure 2.16**). Moreover, using a specific antibody against cMyBP-C, we found that the band detected by the PKA substrate antibody overlapped perfectly with the band detected by the cMyBP-C antibody

(**Figure 2.16**). Therefore, we inferred that the band between 130 and 250 kDa detected by the PKA-substrate antibody was phosphorylated cMyBP-C. Consistently, PKA-mediated phosphorylation of cMyBP-C also increased following temperature treatment (**Figure 2.15c, d**).

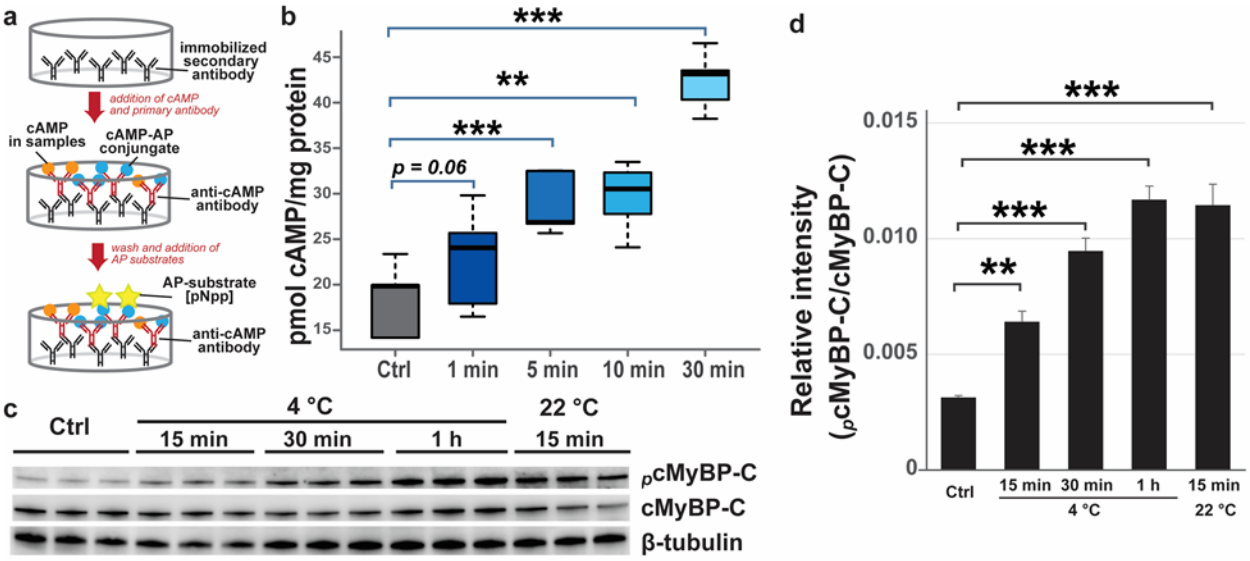


Figure 2.15. The concentration of cAMP increased in tissue incubated at 4 ºC. (a) Schematic of ELISA for quantitative analysis of cAMP concentration. AP, alkaline phosphatase. (b) Graph showing the concentration of cAMP during tissue defrosting at 4 °C. **p< 0.01, ***p< 0.001 by ANOVA. (c) Increased phosphorylation of cMyBP-C mediated by PKA after temperature treatment. Note that the the blot data are from the same blot as Figure 5 and, thus, the loading control is the same. (d) Histogram showing quantification of phosphorylated cMyBP-C by Western blotting. Two-way ANOVA test was performed to evaluate the statistical significance of difference. ** p < 0.01, *** p < 0.001.

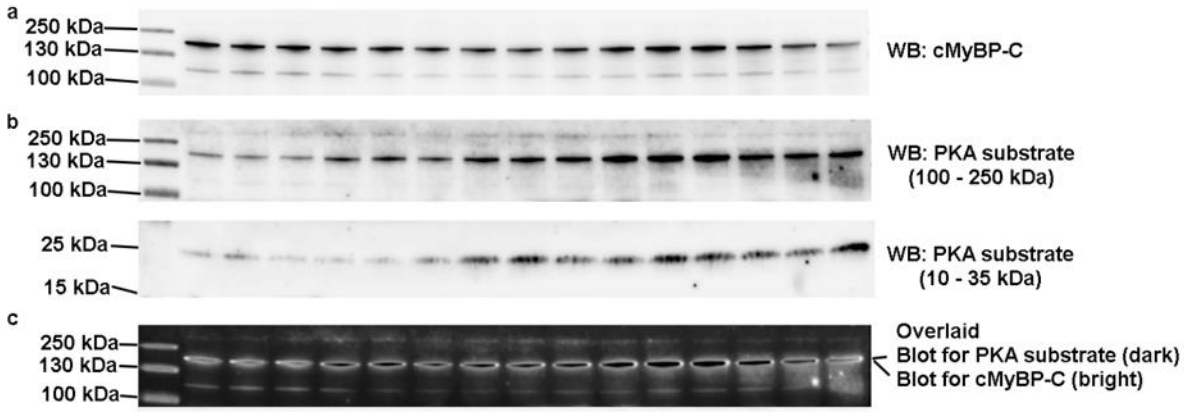


Figure 2.16. Western blotting results from specific cMyBP-C antibody and PKA-substrate antibody. (a) cMyBP-C specific antibody detected 1 band between 130 and 250 kDa. (b) PKA-substrate antibody

detected 1 band between 130 and 250 kDa, which was presumably cMyBP-C presumably, and the weak band between 15 and 25 kDa, which was presumably cTnI. (c) Overlaid image of (a) and (b).

In addition, the phosphorylation of ENH2 and cTnI both increased with increased temperature, suggesting that the phosphorylation of ENH2 may be co-regulated with cTnI phosphorylation. Interestingly, the increase of cTnI and ENH2 phosphorylation and degradation can be partially suppressed by immersing the tissue in a buffer containing protease and phosphatase inhibitors, as well as 2.5 mM EDTA (**Figure 2.17**). Nevertheless, the addition of a buffering solution containing protease and phosphatase inhibitors did not benefit the preservation of cTnT and RLC phosphorylation (**Figure 2.17**). Further investigation of the buffer composition that can prevent artefactual changes of protein PTMs is critical for the accurate interrogation of cardiac protein PTMs changes associated with heart diseases.

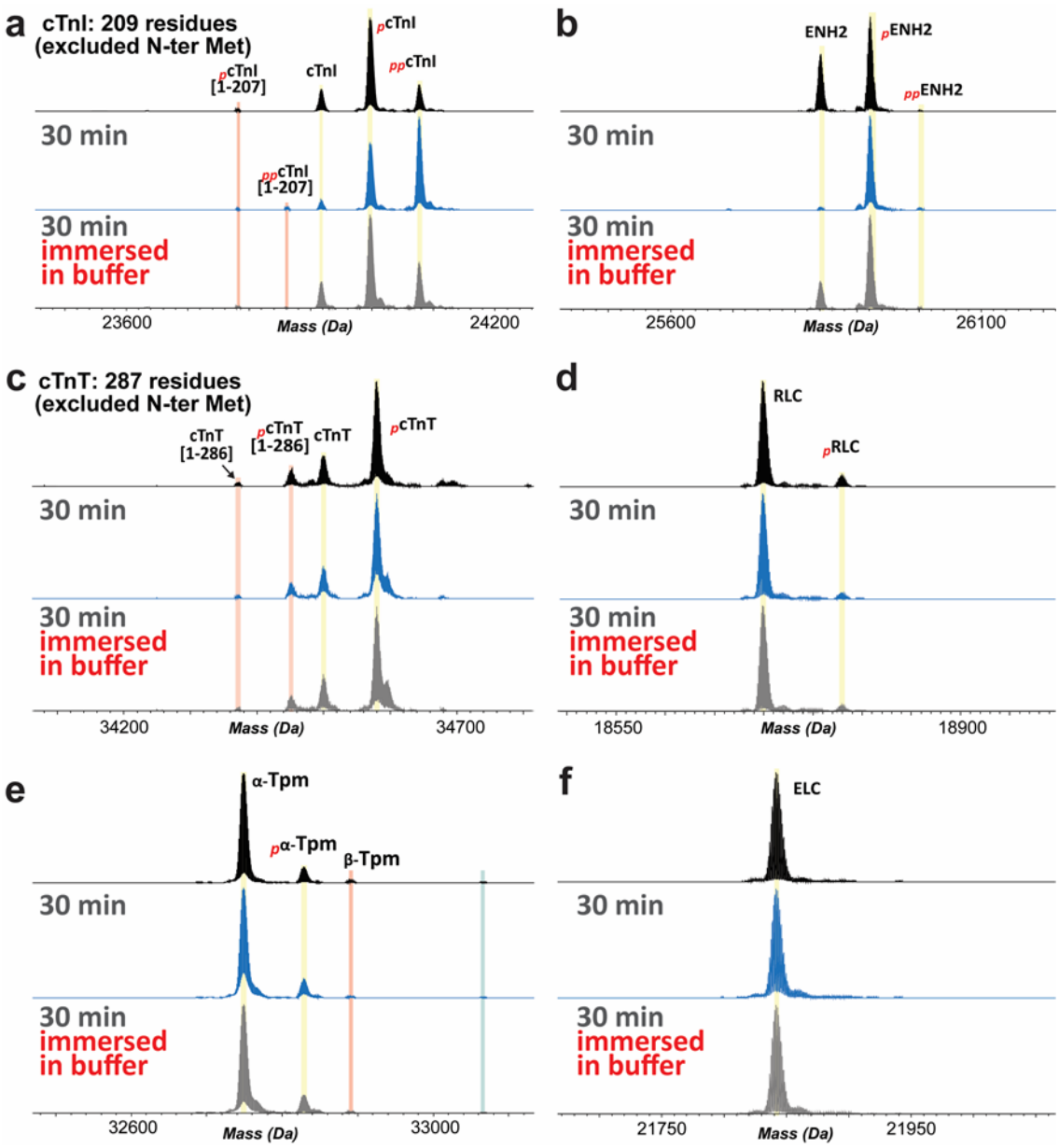


Figure 2.17. **Buffering solution containing protease/phosphatase inhibitors and EDTA can partially suppress temperature-dependent protein PTM alterations.** The increase in cTnI (a) and ENH2 (b) phosphorylation was partially suppressed. However, de-phosphorylation in cTnT (c) and RLC (d) persisted. There was no significant change in Tpm (e) and ELC (f).

We also found that frozen cardiac tissue can be stored in liquid nitrogen or -80 ^{o}C for up to 18 months with less than 10% changes in the phosphorylation of cTnI (**Figure 2.18a-c**), following the initial analysis (**Figure 2.4**). Even though the tissue stored at -80 ^{o}C appeared to

have lower cTnI phosphorylation that was statistically significant, this difference may not result from the difference in the temperature of storage, because higher temperature was shown to increase cTnI phosphorylation (**Figure 2.4**). This may be because tissue from the initial analysis (control) was partially exposed to increased temperature and, therefore, the phosphorylation level was slightly higher. Additionally, we tried to mimic shipment conditions by storing the samples in liquid nitrogen, dry ice, and wet ice for 48 h. There was no significant difference in the cTnI phosphorylation observed after 48 h "shipment" with liquid nitrogen or dry ice (**Figure 2.18c-f**). However, complete deterioration of the tissues occurred with wet ice storage after 48 h, and there was minimal amount of sarcomeric proteins recovered from the tissues, prohibiting proteomics analysis.

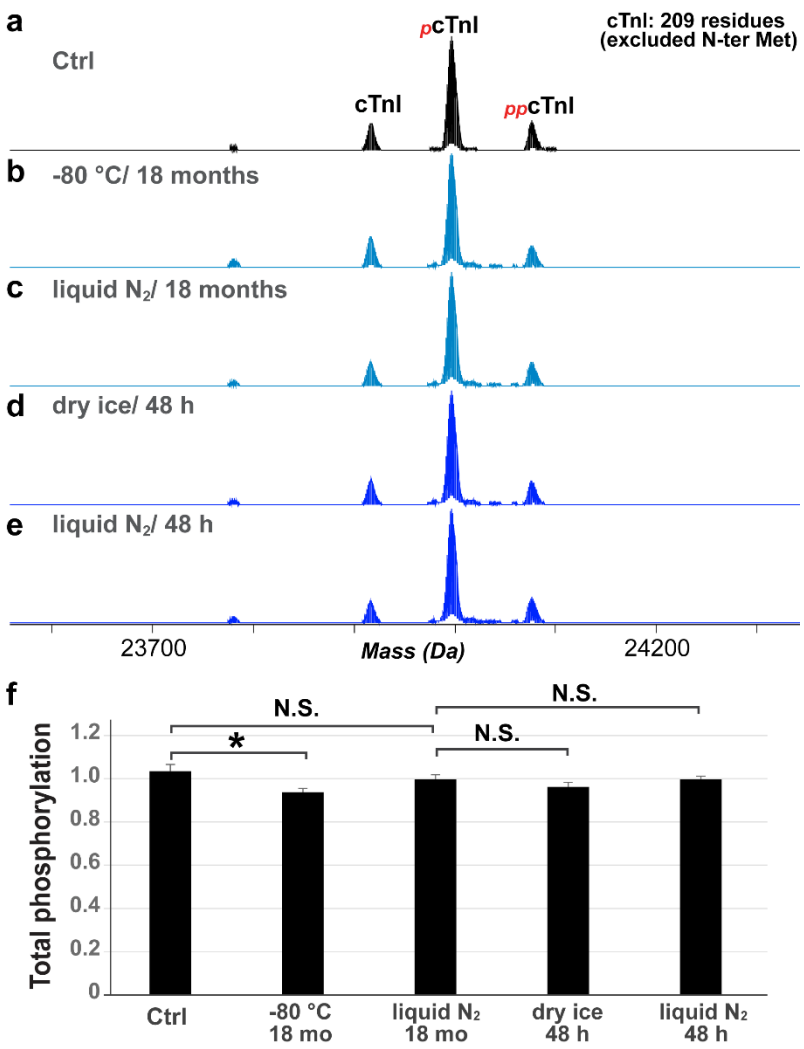


Figure 2.18. **Investigation of the effects of different storage conditions on the phosphorylation of cTnI.** (a). Phosphorylation of cTnI in Ctrl (same as in Figure 2) compared to the cTnI phosphorylation level after storage at in -80 °C (b) and liquid nitrogen (c) for 18 months. (d, e) Phosphorylation of cTnI with storage in -80 °C and liquid nitrogen for 48 hrs after analysis shown in (c). (f) Quantification of the relative abundance of cTnI, mono- and bis-phosphorylated cTnI by MS. Two-way ANOVA was performed to evalute the statistical significance of difference. * $p < 0.05$, N.S. not statistically significant.

2.4 DISCUSSION

Protease and phosphatase inhibitors have been routinely included in buffer solutions for the preparation of biological samples to suppress protein degradation and de-phosphorylation and, therefore, it was expected to observe truncation and de-phosphorylation of a number of

myofilament proteins upon incubation of myocardial tissue at either 4 ºC or 22 ºC. However, it was striking to find that cTnI and ENH2 phosphorylation actually increased upon tissue defrosting. cTnI is an important biomarker for myocardial infarction and a key regulator of muscle contraction and relaxation.[24, 245, 261, 262] cTnI forms a trimeric complex with cTnT and TnC for the positioning of Tpm on the actin filament in response to intracellular Ca^{2+}.[263] TnC binding of Ca^{2+} induces a conformational change in the troponin complex that leads to the repositioning of Tpm on the thin filament and relieves inhibition of actin-myosin interactions.[263] cTnI plays a central role in the regulation of contractility and the phosphorylation changes of cTnI at various sites have been implicated in animal models of heart disease and HF in humans.[24, 245, 264]

cTnI is a well-established substrate of PKA in response to β-adrenergic activation in normal physiological response and disease states.[261, 265] PKA-mediated phosphorylation of cTnI at Ser22/23 decreases the Ca^{2+} sensitivity of the myofilament.[261, 265] Although Ser22/23 are the well-established sites phosphorylated by PKA, it has been found that these sites can also be phosphorylated by protein kinase G (PKG) or protein kinase C (PKC).[266-269] Nevertheless, the finding that the concentration of cAMP, a direct PKA activator, increased rapidly in the cardiac tissue upon incubation of the tissue at 4 ºC strongly supports the activation of PKA, which is likely responsible for the increased phosphorylation of cTnI at Ser22/23 upon increased temperature. PKA-mediated phosphorylation of cTnI represents an important mechanism for the regulation of contraction and relaxation in health and diseases. Studies have shown that phosphorylation of cTnI at Ser22/23 is decreased in human HF;[247, 270-272] however, others reported that cTnI phosphorylation increased rodent HF models.[273] While many factors can account for the differences in these findings, our study has demonstrated the temperature effect on cTnI phosphorylation, which may in part explain the discrepancy in the literature. In addition to cTnI,

PKA also phosphorylates a number of proteins that are important regulators of contractility, including cMyBP-C and phospholamban (a regulator of the Ca^{2+} pump on the sarcoplasmic reticulum).[239, 274-277] Due to the large molecular weight of cMyBP-C, it remains challenging to directly study this protein using top-down proteomics. By employing Western blotting analysis, we also confirmed the increased PKA-mediated phosphorylated of cMyBP-C in the tissues following increased temperature. Recent advancements in multi-dimensional LC platform coupling serial size exclusion chromatography with reverse phase chromatography has demonstrated great promise in the top-down analysis of large proteins from the sarcomere.[278]

In addition to PKA-mediated protein phosphorylation, PKC can phosphorylate cTnI at Ser22/23, Ser42/44, and Thr143 (excluded N-terminal Met).[266] Depending on the PKC isoforms and the specific site, PKC-mediated phosphorylation of cTnI can exert various effects on the Ca^{2+} sensitivity, maximal force or cross-bridge cycling rate.[279-282] In addition, multiple studies have suggested that cTnI can be phosphorylated by PKG and p21-activated kinase 3 (PAK3) at Ser22/23 and Ser148 (excluded N-terminal Met), respectively.[268, 269, 283] However, Ser42/44, Ser148 and Thr143 were not identified in this study as the sites of phosphorylation, possibly due to low abundance of these PTMs. Previously, Zhang et al. reported an MS-based bottom-up method for identification of phosphorylation sites of cTnI, which were purified from the failing human hearts.[284] This study remains the current most sensitive cTnI phosphorylation assay, allowing for the identification of 14 phosphorylation sites.[284]

ENH2 is a Z-disc protein that was just recently identified as a phosphoprotein.[24] In a swine model of acute myocardial infarction, Peng et al. identified concerted de-phosphorylation of cTnI, ENH2, and RLC following 90-min of ischemic injury.[24] The site of ENH2 phosphorylation was localized to Ser118 in the swine LV tissue,[24] which is in agreement with the localization of ENH2

phosphorylation in human LV samples in the present study. In addition, Peng et al. identified a likely truncated proteoform of ENH2, the abundance of which increased following ischemic injury.[24] This suggests that ENH2 is unstable and prone to degradation, which is in high accordance with our finding that intact ENH2 became undetectable after 4 h at 4 °C.

We also found that addition of a buffered solution containing protease/phosphatase inhibitors and 2.5 mM EDTA can partially suppress the increase of cTnI and ENH2 phosphorylation and degradation during the process of tissue defrosting. While protease inhibitors are likely to play a role in the inhibition of protein degradation, since EDTA is an Mg^{2+} chelating agent and Mg^{2+} is essential for the catalytic activity of protein kinases, including PKA,[285, 286] the partial inhibition of PKA activity may be due to chelation of Mg^{2+} by EDTA.

In contrast to the phosphorylation of cTnI and ENH2, cTnT and RLC phosphorylation decreased in the myocardium with increased temperature. Previous studies have established Ser1 (excluding the N-terminal Met) as the site of phosphorylation on endogenous cTnT.[258, 287, 288] cTnT can also be phosphorylated at Thr203, Ser207, Thr212, Ser284, and Thr293 (excluded N-terminal Met).[258, 266, 289] However, since multiply phosphorylated cTnT was not observed in our mass spectra, it is possible that cTnT proteoforms with these putative phosphorylation sites other than Ser1 were too low in abundance to be detected. The protein kinase responsible for cTnT phosphorylation at Ser1 was isolated and identified as casein kinase 2 (CK2).[287, 290] It has been shown that cTnT phosphorylation at Ser1 has no effects on the Ca^{2+} sensitivity of actomyosin.[291] In skeletal muscle, troponin T (TnT) phosphorylation at Ser1 had no effects on the Ca^{2+} binding properties of the troponin complex,[292] or on the interaction between TnT and Tpm.[283] The physiological role of constitutive phosphorylation of cTnT Ser1 therefore remains unclear; however, it has been suggested that phosphorylation of cTnT Ser1 may be involved in the

stabilization of a contractile apparatus or its interaction with auxiliary proteins, rather than direct regulation of the contractile cycle.[258]

Unlike cTnT phosphorylation at Ser1, phosphorylation of RLC at Ser14 has direct and significant impact on the contractile properties.[293] Phosphorylation of RLC at Ser14 has been shown to increase Ca^{2+} sensitivity, maximal force, and the rate of cross-bridge cycling.[294] The decreased phosphorylation of RLC has been reported in human end-stage failing hearts.[295] Recently, cardiac-specific myosin light chain kinase (cMLK) has been identified as the primary kinase responsible for cardiac RLC phosphorylation at Ser14.[296] It remains unclear whether the decreased phosphorylation of RLC and cTnT is a result of decreased kinase expression/activity, or increased phosphatase expression/activity. While immersing the tissue in a buffer solution can partially suppress changes in cTnI and ENH2 PTMs, the decline of phosphorylation in cTnT and RLC persisted even though phosphatase inhibitors were included in the buffer. It is possible that 50 mM sodium fluoride, a Ser phosphatase inhibitor, is insufficient to suppress the artefactual decline in cTnT and RLC phosphorylation. Further studies on the composition of buffering solution for the preservation of protein PTMs are important for the accurate dissection of molecular changes in the failing and non-failing human hearts.

In addition to the evaluation of increased temperature, we also investigated the effects of tissue storage and shipment strategies on the phosphorylation of cTnI. We found that total phosphorylation of cTnI was not markedly affected with storage at -80 ^{o}C or liquid nitrogen for up to 18 months. There was also no significant changes in cTnI phosphorylation when stored in liquid nitrogen or dry ice for 48 h, mimicking most shipment conditions. As expected, storage of the tissue using wet ice led to complete tissue deterioration, prohibiting proteomics analysis. Even though -80 ^{o}C and liquid nitrogen storage did not have a profound effects on the PTM change, it

is important to recognize that small pieces of tissue (< 10 mg) are likely more susceptible to change of temperature, and therefore, liquid nitrogen remains the best option for shipment of small tissues and prolonged tissue storage.

Herein, we have demonstrated the impact of temperature on the preservation of protein PTMs in human cardiac tissue. Importantly, this study underscores the power of top-down MS-based proteomics for the assessment of tissue quality. Based on the fact that defrosted tissue exhibited a various degree of protein degradation on cTnI, ENH2, and cTnT, the truncated proteoforms of multiple myofilament and associated proteins can collectively be indicative of poor tissue quality. It is important to point out that many truncated proteoforms of cTnI and cTnT have similar molecular weight as the intact proteoforms, therefore making the detection of such truncated proteoforms difficult and ambiguous via the gel electrophoresis approach. Moreover, although the conventional bottom-up approach is capable of identifying and quantifying thousands of proteins from a complex mixture,[297, 298] due to limited sequence coverage as a result of proteolytic digestion, it remains challenging to distinguish highly homologous protein isoforms and proteolytic products with small truncation. On the contrary, top-down MS-based proteomics approach permits the analysis of intact proteins and unambiguously reveals the existing proteoforms including those with small truncation <1000 Da, as well as those with distinct sequence variations [299]. In addition, high-resolution top-down MS allows for the determination of accurate molecular weight with less than 10 ppm discrepancy. This permits the identification of the truncated proteoforms based on their accurate molecular weights and the mass differences from their intact counterparts.

In summary, our study highlights the importance of tissue handling in revealing accurate molecular alterations in the human failing versus non-failing hearts. To the best of our knowledge,

this is the first study to investigate the impact of temperature on the PTM changes of the major myofilament and associated proteins. By employing a high-resolution top-down proteomics approach, we have observed artefactual increases of cTnI and ENH2 phosphorylation, and artefactual decreases of cTnT and RLC phosphorylation, which demonstrated that sarcomeric protein PTMs were differentially affected by temperature. The sites of protein increased phosphorylation on cTnI were localized to Ser22/23, which in accordance with increased cAMP concentration in the warmed-up myocardium, suggests the activation of PKA upon increased temperature. Moreover, this study revealed degraded cTnI, ENH2, and cTnT upon tissue increased temperature, which can be indicative of poor tissue quality for future studies employing human cardiac tissue.

CHAPTER 3

TOP-DOWN PROTEOMICS REVEALTED DIFFERENTIAL REGULATION OF SARCOMERIC PROTEIN POST-TRANSLATIONAL MODIFICATIONS IN HUMAN HYPERTROPHIC CARDIOMYOPATHY

ABSTRACT

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disease and is estimated to affect approximately 1 in 200 people in the general population. It is the most common cause of sudden cardiac death in young adults, and frequently precipitates lethal arrhythmia and heart failure. HCM is caused by mutations in the genes that encode the protein constituents of the sarcomeres, which are the basic contractile apparatus. Although the genetic causes of HCM have been identified, the molecular mechanisms by which sarcomeric protein mutations lead to contractile dysfunction and adverse cardiac remodeling are not completely understood. Since protein post-translation modifications (PTMs) regulate many properties of the sarcomeric proteins, we hypothesize that contractile dysfunction in HCM is mediated by dysregulation in the sarcomeric protein PTMs. We undertook a high-throughput top-down proteomics approach to analyze sarcomeric protein PTMs from human HCM and donor tissues. Specifically, we analyzed HCM tissues from myectomy surgery and explanted hearts with HCM. We observed significant down-regulation in the phosphorylation of cardiac troponin I, enigma homolog isoform 2, α-Tropomyosin, myosin light chain 2 (MLC2) and Cypher isoform 6 in the HCM myectomy tissues compared to the donor tissues. In addition, we found that the phosphorylation levels of cardiac troponin T (cTnT) and Cypher isoform 5 were up-regulated in the HCM myectomy tissues compared to the donor tissues. Interestingly, we also found partial reversal of the cTnT, MLC2 and Cypher isoform 5 phosphorylation change at the failing stage of HCM to the level closer to the donor tissues. Furthermore, we identified cysteine rich protein 2 (CRIP2) to be thiol-sulfenylated, a new type of PTMs that were not previously reported for sarcomeric proteins, which may represent a novel mechanism of redox signaling in cardiac pathophysiology.

3.1 INTRODUCTION

Hypertrophic cardiomyopathy (HCM) is defined as abnormal non-dilated left ventricular hypertrophy in the absence of systemic diseases that lead to pressure overload of the heart such as hypertension and aortic valvular stenosis.[1-3] HCM is a common inherited heart disease previously estimated to affect at least 1 in 500 people[4] of both sexes and various ethnicities from all over the world. With recent advances in population genetic studies and contemporary diagnostic techniques, the prevalence of HCM was recently revised to account for the under-recognized population, and estimated to be at least 1 in 200 in the general population.[5] HCM is transmitted in an autosomal dominant pattern, although *de novo* mutations can occur. The first HCM-causing variant (*MYH7* Arg403Gln encoding β-myosin heavy chain (β-MHC)) was reported in 1990,[21] and since then over 1400 HCM-associated mutations were identified in at least 11 genes that encodes the protein constituents of the sarcomeres.[1, 22]

Sarcomeres are the repeating units that constitute the contractile apparatus in striated muscles. Sarcomeres consist of the myofilaments interconnected by complex protein structures known as Z-discs.[33, 243, 244] Myofilaments are composed of the myosin-based thick filaments and actin-based thin filaments.[24, 33] The thick filaments contain myosin heavy chain and cardiac myosin binding protein C (cMyBP-C), as well as the myosin light chain 1 and 2 (MLC1 and MLC2), also known as the myosin essential and regulatory light chain, respectively. The thin filaments consist of actin, tropomyosin (Tpm), and the cardiac troponin complex, which is made up of troponin C (TnC), cardiac troponin I (cTnI), and cardiac troponin T (cTnT). The Z-discs anchor the adjacent myofilaments, allowing for sliding of the thick filaments against the thin filaments for muscle contraction. In addition to the structural role, emerging evidence indicates that Z-discs represent critical signaling hotspots within the cardiomyocytes,[33, 250] and the post-translational modifications (PTMs) of Z-disc proteins can be altered in response to cardiac injury.[24, 251] Moreover, mutations

in the Z-disc proteins, such as muscle LIM protein (MLP), Titin and myozenin-2 (MYOZ2), were found to cause inherited cardiomyopathies.[300] Although the genetic causes of HCM have been identified and continue to evolve, the molecular mechanisms underlying HCM pathogenesis as a result of sarcomeric protein mutations are poorly understood. It is unclear how sarcomeric protein mutations lead to contractile dysfunction, adverse cardiac remodeling and progression to heart failure. Since sarcomeric protein PTMs are important mediators of cardiac signaling and contractile function, we hypothesize that sarcomeric protein PTM alterations lead to contractile dysfunction in HCM.

To address this question, we employed a high-resolution mass spectrometry (MS)-based top-down proteomics approach[147, 206, 244, 252] to evaluate the sarcomeric protein PTMs in human HCM and donor heart tissues. In contrast to the shotgun proteomics approach, which employs proteolytic digestion prior to MS analysis, the top-down proteomics workflow analyzes intact proteins, which can provide a "bird's eye" view of all proteoforms[146] (a term encompassing all protein variants of the same gene arising from genetic variation, alternative mRNA splicing and PTMs) and permits the comprehensive and quantitative analysis of protein PTMs.[147, 244, 252, 253] Following MS analysis, a specific proteoform of interest can be isolated in the gas phase for fragmentation, known as tandem MS (MS/MS), for protein sequence characterization and PTM localization.

3.2 MATERIALS AND METHODS

3.2.1 Reagents and chemicals

All reagents were purchased from Sigma-Aldrich, Inc. (St. Louis, MO, USA) unless otherwise noted. HPLC grade water, acetonitrile, and ethanol were purchased from Fischer Scientific (Fair Lawn, NJ, USA).

3.2.2 Human cardiac tissue collection and preservation

The donor hearts characterized with normal cardiac function, but deemed unacceptable for transplantation, were obtained from the University of Wisconsin Hospital and Clinic. The procedures for the collection of human cardiac tissue were approved by the Institutional Review Board of the University of Wisconsin-Madison. The donor hearts were maintained in cardioplegic solution before dissection. Following dissection, the tissues were immediately flash-frozen in liquid nitrogen. Myectomy tissues were acquired from the University of Wisconsin Hospital and Clinic during septal myectomy surgery and quickly flash-frozen in liquid nitrogen within 10 min of tissue removal. The procedures for myectomy tissue collection were approved by the Institutional Review Board of the University of Wisconsin-Madison. Alternatively, de-identified donor and HCM heart tissues were acquired from the Sydney Heart bank.[301]

3.2.3 Cardiac saromeric protein extraction

The extraction of sarcomeric proteins was carried out similarly as reported previously with minor modifications.[24, 160, 254] All homogenization steps were performed in the cold room (4 °C), and each homogenization step took less than 20 sec. The tissue (7-10 mg) was first homogenized rapidly in 50 µl HEPES buffer (30 mM HEPES pH 7.5, 120 mM NaF, 1.2 mM Na_3VO_4, 1.2 mM

PMSF, 3 mM EDTA, 1 mM DTT, 15 mM L-methionine, 1x protease inhibitor, 1x phosphatase inhibitor (catalog# ab65621, abcam) and 1x phosphatase inhibitor (catalog# 524629, Milipore)) using a Teflon pestle (1.5 mL tube rounded tip; Scienceware, Pequannock, NJ, USA). The homogenate was centrifuged at 21,000 rcf for 15 min, and the pellets were homogenized in another 50 µl HEPES buffer in the cold room. The samples were centrifuged at 21,000 rcf for 15 min and the HEPES fractions were combined. The remaining pellets were further homogenized in 50 µl trifluoroacetic acid (TFA) solution (1% TFA, 1 mM Tris-(2-carboxyethyl) phosphine). The homogenate, which contained primarily sarcomeric proteins, was centrifuged at 21,000 rcf for 20 min and the supernatant was collected. The sarcomeric protein enriched fractions were diluted 20-fold for Bradford protein assay, and all the samples were diluted to 100 ng/µl for liquid chromatography (LC)-MS analysis.

3.2.4 Reverse phase chromatography (RPC) and top-down proteomics analysis

LC-MS analysis was carried out using a NanoAcquity Ultra-high Pressure LC system (Waters, Milford, MA, USA) coupled to a high-resolution Impact II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker Daltonics, Bremen, Germany). The sarcomeric proteins in each sample were separated using a home-packed PLRP column (PLRP-S, 1000 Å pore size, 10 µm particle size, 500 µm of inner diameter) via a gradient of 5% to 95% mobile phase B (mobile phase A: 0.1% formic acid in water; mobile phase B: 0.1 % formic acid in 50:50 acetonitrile:ethanol) at a constant flow rate of 14 µL/min. The column pressure were maintained in the range of 800-1500 psi. The sarcomeric proteins eluted were ionized via electrospray ionization with the end plate offset and capillary voltage set at 450 V and 4000 V, respectively. Nebulizer pressure was set to 0.5 Bar and dry gas flow rate was 4.0 L/min. 10 V of in-source collisional energy was applied to

facilitate ionization. Mass spectra were collected at a scan rate of 0.5 Hz over the mass range of 500-2000 *m/z*.

3.2.5 Top-down MS/MS analysis

Individual protein fractions were collected from RPC separation and analyzed using a 12 Tesla solariX Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometer (Bruker Daltonics) or 7 Tesla LTQ/FT-ICR mass spectrometer (Thermo Scientific) via an automated chip-based nanoESI source (Triversa NanoMate; Advion Bioscience, Ithaca, NY, USA). Proteoforms of interest were first isolated in the gas phase and fragmented by electron capture dissociation (ECD). The electron energy and reaction time for ECD were determined on a case-by-case basis to achieve optimal fragmentation. The transients from approximately 1000-4000 scans were summed to obtain high-quality spectra for the characterization of protein sequence and localization of protein PTMs.

3.2.6 Data analysis

All LC-MS data were processed and analyzed using the DataAnalysis software (version 4.3; Bruker Daltonics). The chromatograms were smoothed using the Gauses algorithm with a smoothing width of 2.024. Mass spectra were averaged over the retention time window wherein all proteoforms of the same protein eluted. The spectra were deconvoluted using the Maximum Entropy algorithm embedded in the DataAnalysis software. The resolving power was set to 50,000 for proteins below 50 kDa, which were typically isotopically resolved by the mass spectrometer. The most-abundant mass was reported for all MS data, and monoisotopic mass was reported for all MS/MS data. The relative abundance (*A*) of a particular proteoform is reported as the ratio of the peak intensity of the proteoform to the summed peak intensities of all proteoforms of the same

protein. Based on the abundance of the individual phosphorylated species, the total amount of phosphorylation (P_{total}; mol of Pi/mol of protein) of a single protein was calculated using the following equation:

$$P_{total}=A_{\text{mono-phosphorylated}}+2\times A_{\text{bis-phosphorylated}}+3\times A_{\text{tri-phosphorylated}}+4\times A_{\text{tetra-phosphorylated}}$$

Tandem mass spectra were analyzed using the MASH Suite Pro[229, 255] software which was developed in-house. The spectra were deconvoluted with a signal-to-noise ratio of 3 and a cutoff fit score of 60%. All the program-processed data were manually validated to obtain accurate sequence and PTM information.

3.2.7 Statistical analysis

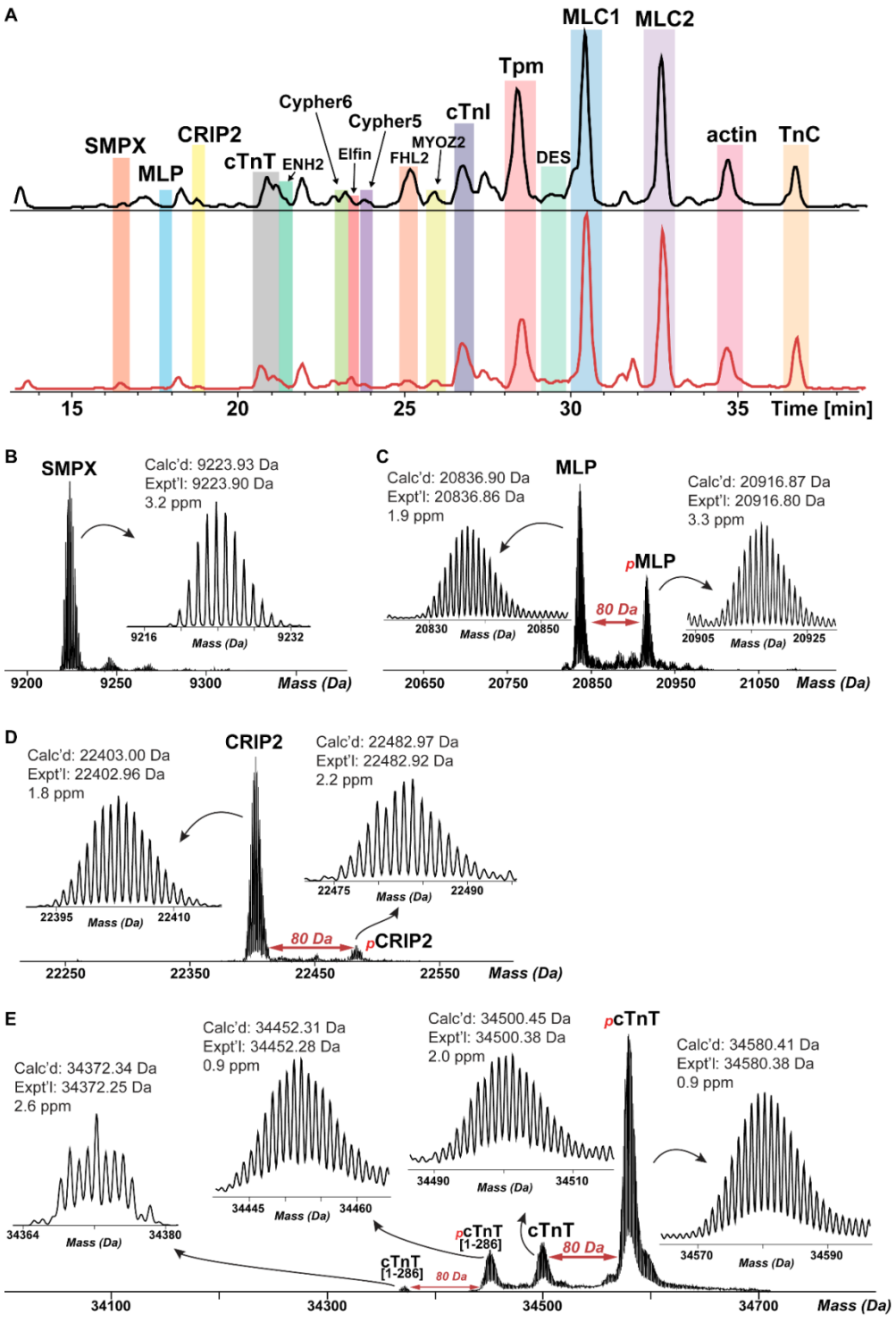
The ANOVA test was performed to evaluate the statistical significance of variation between groups. Differences between means were considered to be statistically significant at $p < 0.05$.

3.3 RESULTS

3.3.1 High-throughput and robust LC-MS method for the analysis sarcomeric proteins

We developed a robust liquid LC-MS method that allowed for high-throughput analysis of the cardiac sarcomeres. The method requires less than 10 mg tissue, and the entire data acquisition process including sample processing and LC-MS analysis took less than 3 hours. In a single LC-MS run, we were able to observe many sarcomeric proteins (< 60 kDa), including the major myofilament proteins: cTnT, cTnI, TnC, Tpm, actin, MLC1 and MLC2, and multiple Z-disc and sarcomere-associated proteins: MLP, small muscle protein X-linked (SMPX), cysteine rich protein

2 (CRIP2), Enigma homolog isoform 2 (ENH2), Cypher isoform 6 and 5 (Cypher6 and Cypher5), Elfin, four and half LIM protein 2 (FHL2), myozenin-2 (MYOZ2, also known as calsarcin-1) and desmin (DES) (**Figure 3.1**).



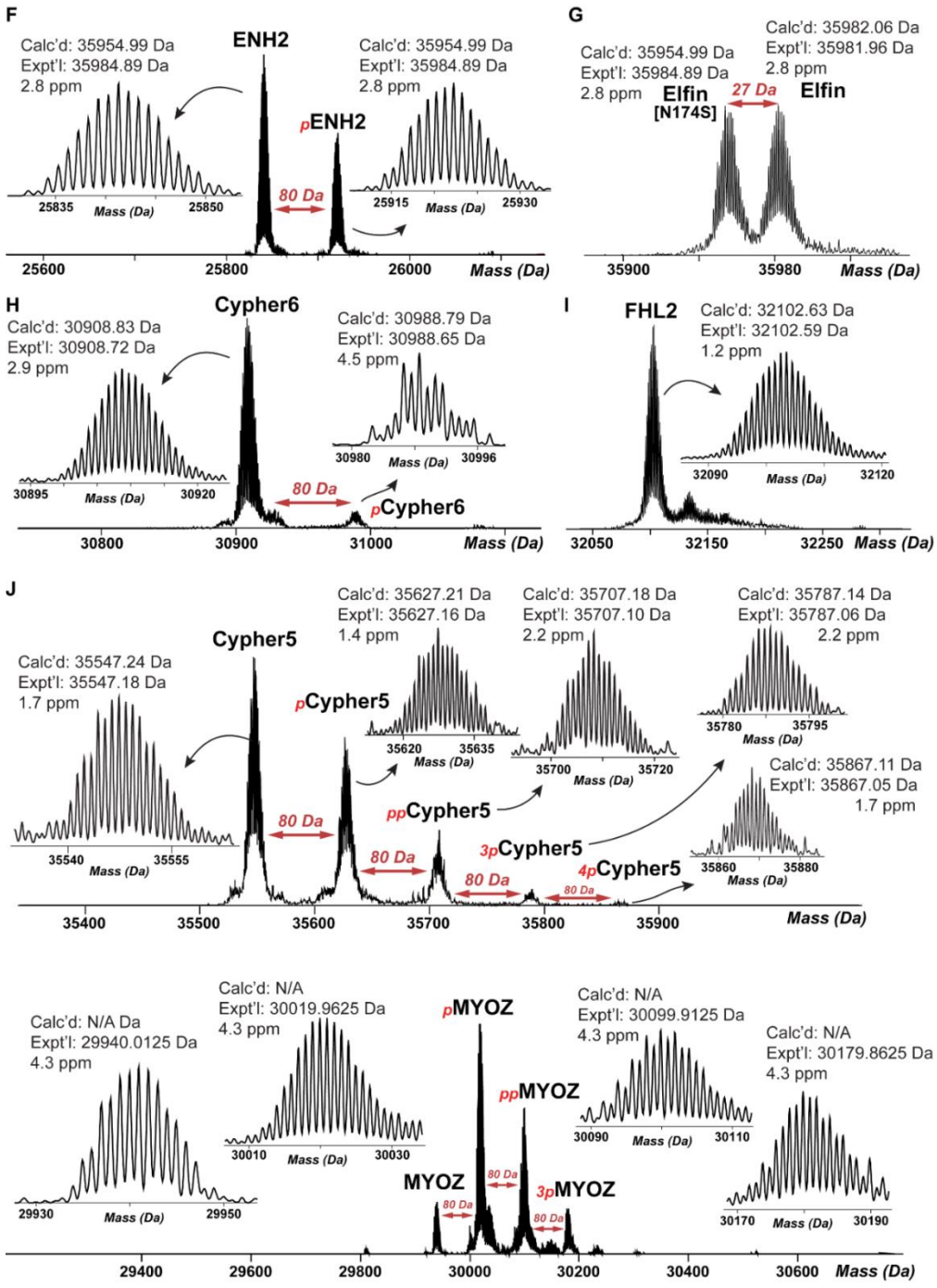


Figure 3.1. Top-down proteomics for the analysis of sarcomeric protein PTMs. (A) Overview of the chromatogram showing the elution time for individual sarcomeric proteins. (B-J) High-resolution mass spectra of SMPX (*SMPX gene*), MLP (*CSRP3 gene*), CRIP2 (*CRIP2 gene*), cTnT (*TNNT2 gene*), ENH2 (*PDLIM5 gene*), Elfin (*PDLIM1 gene*), Cypher6 (*LDB3 gene*), FHL2 (*FHL2 gene*), Cypher5 (*LDB3 gene*) and MYOZ2 (*MYOZ2 gene*).

3.3.2 Reduction in the phosphorylation of cTnI and ENH2 in HCM

The total phosphorylation of cTnI was between 0.6-2.0 mol Pi/ mol protein in the donor hearts, but remained consistently low (below 0.5 mol Pi/ mol protein) in the HCM tissues from myectomy surgery (**Figure 3.2A**). The explanted HCM tissues exhibit more variable distribution of cTnI phosphorylation, ranging between 0.1-1.8 mol Pi/ mol protein (**Figure 3.2A**). The phosphorylation of cTnI was down-regulated in both myectomy and explanted HCM heart tissues. However, low cTnI phosphorylation was more consistent among the myectomy tissues. We also found that phosphorylation of ENH2 was down-regulated in both the myectomy and explanted HCM heart tissues (**Figure 3.2B**). Consistent with the pattern observed in cTnI, ENH2 phosphorylation level showed less variation in the myetomy tissues (**Figure 3.2B**). Nevertheless, there was not statistically significant difference between the myectomy and explanted HCM tissues regarding the phosphorylation of cTnI or ENH2 (**Figure 3.2C**). Following quantification of total phosphorylation, we performed MS/MS analysis to localize the sites of phosphorylation in the donor hearts due to the high abundance of the phosphorylated proteoforms. The phosphorylation sites on the bis-phosphorylated cTnI were localized to Ser22/23 (**Figure 3.2D**), the canonical sites of protein kinase A (PKA)-mediated phosphorylation on cTnI. MS/MS analysis of ENH2 unambiguously localized the phosphorylation site on mono-phosphorylated ENH2 to Ser118, similarly as reported previously in a swine model of acute myocardial infarction.[24] Interestingly, the phosphorylation of cTnI and ENH2 exhibit a strong linear correlation with a R^2 value of 0.9159 (**Figure 3.2F**). This suggests that the phosphorylation of ENH2 and cTnI were co-regulated by the same enzymes/pathway or there was a cross-talk mechanism between the enzymes/pathways that regulate the phosphorylation of these two proteins. Given the fact that ENH2 phosphorylation is

localized to Ser118 with a consensus PKA substrate sequence (RRXS/T), it is likely that ENH2 is phosphorylated by PKA.

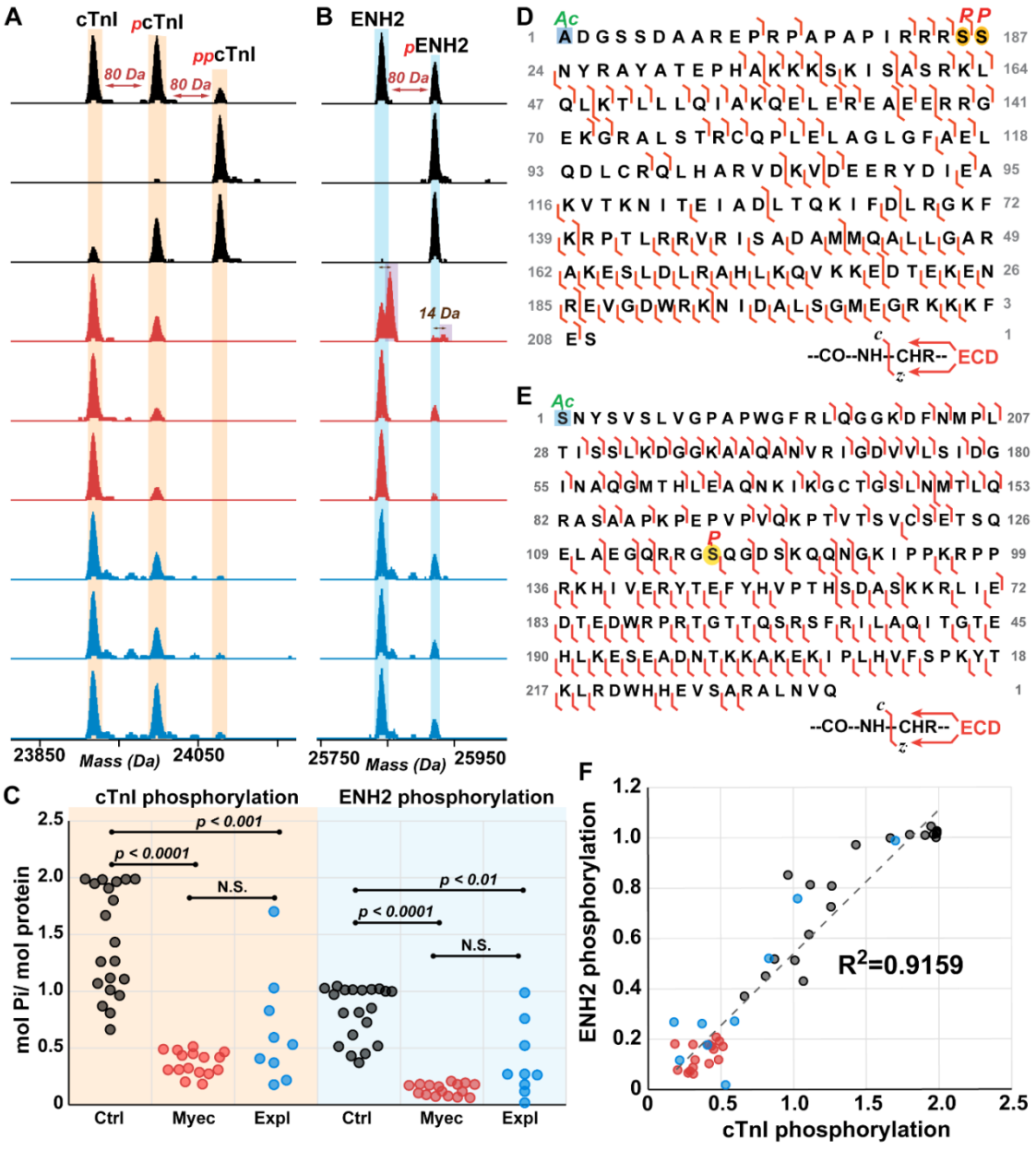


Figure 3.2. Decreased cTnI and ENH2 phosphorylation in myectomy and explanted HCM heart tissues. (A). Representative mass spectra for cTnI from donor hearts (black), myectomy HCM tissues (red), and explanted HCM hearts (blue). (B). Representative mass spectra for ENH2 from donor hearts (black), myectomy HCM tissues (red), and explanted HCM hearts (blue). (C). Quantification of total protein phosphorylation in cTnI and ENH2 in donor hearts (Ctrl), and myectomy (Myec) and explanted (Expl) HCM hearts. (D). Localization of cTnI phosphorylation to Ser22/23. (E). Localization of ENH2 phosphorylation to Ser118. (F). Linear correlation between cTnI phosphorylation and ENH2 phosphorylation.

Decreased cTnI phosphorylation at Ser22/23 is consistent with previously reported,[39, 302] Phosphorylation of cTnI at Ser22/23 mediated by PKA was shown to decrease the calcium sensitivity of the myofilament in the normal hearts.[261, 265] Nevertheless, in the case of HCM, despite a dramatic reduction in the phosphorylation of cTnI, the calcium sensitivity of the troponin remains almost the same,[302] indicating a reduction in the ability of cTnI phosphorylation to regulate myofilament calcium sensitivity in HCM. Furthermore, unlike the thin filaments of the donor hearts wherein increased cTnI phosphorylation decreased calcium sensitivity,[302] manipulation of the cTnI phosphorylation did not alter the thin filament calcium sensitivity in human HCM myofibrials.[302] These data suggest that cTnI phosphorylation was uncoupled from myofilament calcium sensitivity. Nevertheless, the exact molecular cause of this uncoupling event has not been defined. ENH2 is a Z-disc protein that was just recently identified as a phosphoprotein.[24] Interestingly, there was a concerted de-phosphorylation of cTnI, ENH2, and MLC2 following 90-min of ischemic injury in a swine model of acute myocardial infaction.[24] The site of ENH2 phosphorylation was localized to Ser118 in the swine LV tissue,[24] which is in agreement with the localization of ENH2 phosphorylation in the human heart tissue. The role of ENH2 in the function and regulation of contractility remains to be investigated.

3.3.3 Increased phosphorylation of cTnT and decreased phosphorylation of MLC2 and α-Tpm only in myectomy HCM tissues

The mono-phosphorylated cTnT was the major proteoform detected in all the samples and less than 1/3 of the protein was un-phosphorylated (**Figure 3.3A**). cTnT showed a slight propensity of increased phosphorylation in the myectomy HCM samples as compared to the donor tissues

(**Figure 3.3B**). Interestingly, there was no difference in the cTnT phosphorylation between the donor tissues and the explanted HCM samples (**Figure 3.3B**). Previous research identified Ser1 as the site the cTnT phosphorylation;[288] however, phosphorylation of cTnT at Ser1 did not seem to affect the calcium sensitivity of actomyosin.[291] cTnT was reported to be phosphorylated at Thr203, Ser207, Thr212, Ser284, and Thr293.[258, 266, 289] Since multiply-phosphorylated cTnT peaks were not observed in any human tissue sample, it is possible that the phosphorylation on these sites were very low abundant if exist. The exact site responsible for increased total cTnT phosphorylation in the HCM myectomy tissues remains to be investigated. Overall, the cTnT phosphorylation increased by ~12.4% in the HCM myectomy samples. Albeit being a small change, the significance of this change remains to be determined because a new site of cTnT phosphorylation is possible.

On the other hand, the phosphorylation levels of MLC2 and α-Tpm were found to decrease in the HCM myectomy samples compared to the donor samples, but there was no significant difference observed between the explanted HCM and the donor heart samples (**Figure 3.3C-F**). On average, MLC2 phosphorylation reduced by ~51% in the HCM myectomy tissues (**Figure 3.3D**). Previous study showed localization of human ventricular MLC2 phosphorylation to Ser14.[299] Unlike cTnT phosphorylation on Ser1, phosphorylation of MLC2 at Ser14 leads to increased calcium sensitivity, maximal force and the rate of cross-bridge cycling.[293, 294] Recently, cardiac-specific myosin light chain kinase (cMLK) was identified as the kinase responsible for cardiac MLC2 phosphorylation.[296] However, it remains unclear whether decreased MLC2 phosphorylation in HCM myectomy tissues was a result of decreased kinase activity or an increase in the activity of responsible phosphatase. α-Tpm phosphorylation was relatively low (less than 0.2 mol Pi/ mol protein) in all the tissues analyzed (**Figure 3.3E, F**). On average, α-Tpm phosphorylation decreased by ~20% in the HCM myectomy tissues compared to the donor tissues.

However, there was no statistically significant difference in the α-Tpm phosphorylation between the HCM explanted hearts and the donor hearts. Previous studies suggested that α-Tpm phosphorylation or pseudo-phosphorylation (Ser283Asp) did not altered steady-state calcium sensitivity,[303, 304] but pseudo-phosphorylation of α-Tpm at Ser283 lead to increased maximal ATPase activity, and slower relaxation without altering the rate of activation.[303] Intriguingly, prevention of α-Tpm phosphorylation (Ser283Ala) rescued HCM phenotypes in a mouse model of HCM harboring a mutation in α-Tpm (Glu180Gly).[305] Taken these together, α-Tpm phosphorylation reduction in the HCM myectomy tissues may represent a compensatory mechanism to offset the hyper-contractile[32, 34] phenotypes resulted from HCM-causing mutations.

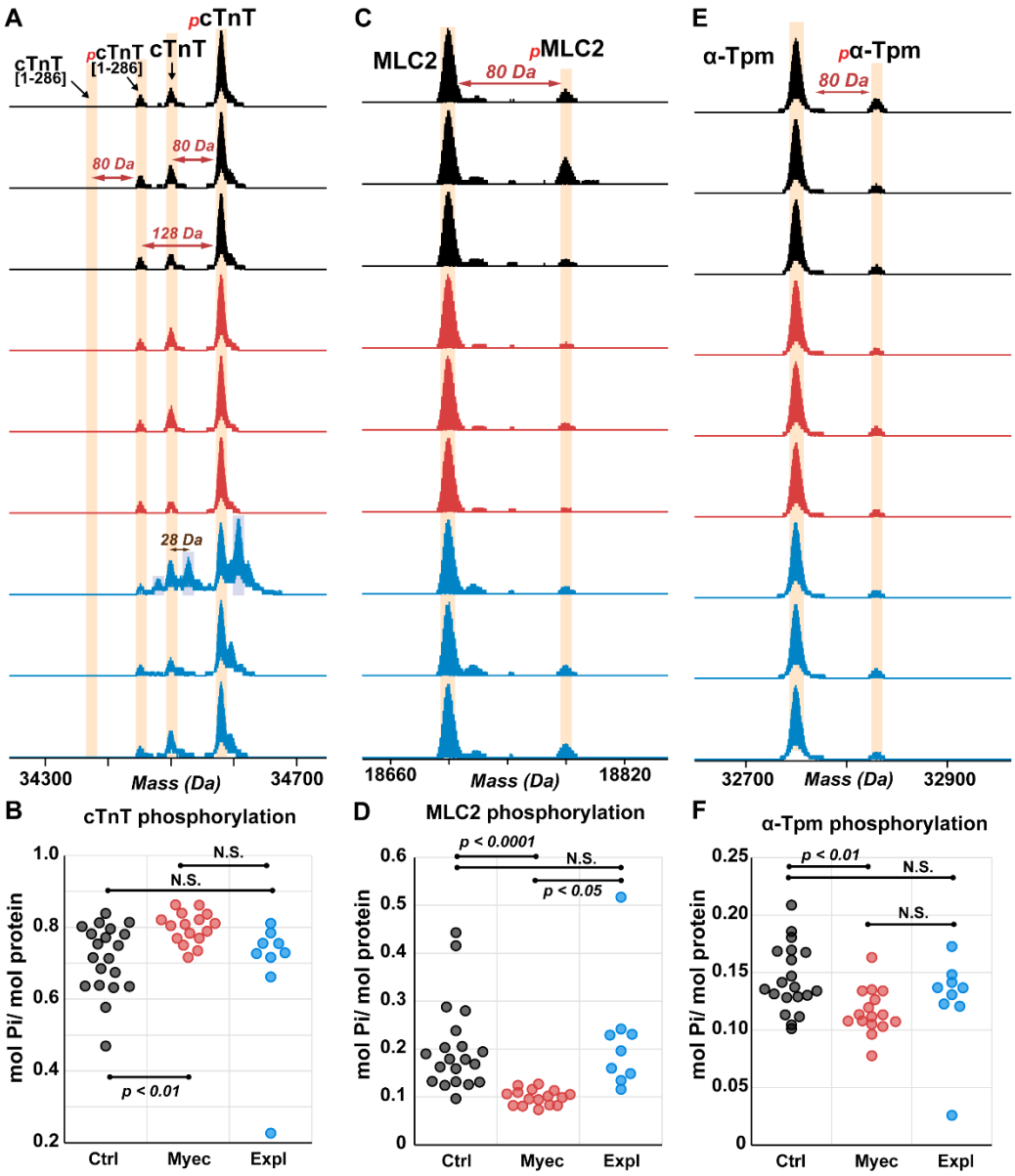


Figure 3.3. Changes of cTnT, MLC2 and α-Tpm phosphorylation in myectomy and explanted HCM tissues. Representative mass spectra for (A) cTnT, (C) MLC2, and (E) α-Tpm from donor hearts (black), myectomy HCM tissues (red), and explanted HCM heart tissues (blue). Quantification of total protein phosphorylation for (B) cTnT, (D) MLC2, and (F) α-Tpm in donor hearts (Ctrl), myectomy (Myec) and explanted (Expl) HCM hearts.

3.3.4 Increased phosphorylation of Cypher5 and decreased phosphorylation of Cypher6

In addition to ENH2, we identified another PDZ-LIM protein, Cypher, in the sarcomeric protein enriched fractions. Cypher is encoded by the *LDB3* gene, and expressed in the striated muscle tissue with many splicing isoforms. Using the top-down LC-MS method, we were able to identified two isoforms of Cypher, Cypher5 and Cypher6, both of which contain an N-terminal PDZ domain without the C-terminal LIM domains. We also detected phosphorylation of both isoforms. Specifically, Cypher5 is mono- and bis-phosphorylated in the donor heart tissues, whereas in the HCM tissues, tri- and tetra-phosphorylated Cypher5 were observed (**Figure 3.4A**). The total phosphorylation of Cypher5 increased in both the myectomy and explanted HCM heart tissues compared to the donor heart tissues (**Figure 3.4B**). However, there is a slight decrease in the phosphorylation of Cypher5 in the explanted HCM heart tissues compared to the myectomy tissues (**Figure 3.4B**). Cypher6 was found predominantly un-phosphorylated in all tissue samples, with about 0.02-0.18 mol Pi/ mol protein in the donor hearts. Interestingly, in contrast to the increased phosphorylation of Cypher5 in the HCM tissues, Cypher6 phosphorylation decreased in both myectomy and explanted HCM tissues (**Figure 3.4C and D**). Unlike Cypher5 phosphorylation, there is no statistically significant difference in the phosphorylation of Cypher6 between the myectomy and explanted HCM heart tissues (**Figure 3.4D**). These results indicate that Cypher5 and Cypher6, albeit being closely related, were regulated differently. The precise roles of each isoform in the regulation of contractile function and signaling remain to be investigated.

A previous proteomics study in mice identified 13 sites of phosphorylation on Cypher5,[306] of which 9 sites (Ser43, Ser97, Ser120, Ser122, Ser216, Ser222, Ser281, Ser307 and Ser329 (excluding N-terminal Met)) are conserved between human and mouse Cypher5 (**Figure 3.5A**). A

comparison between the sequence of Cypher5 and Cypher6 revealed 4 sites of phosphorylation (Ser120, Ser122, Ser216 and Ser222) that are unique to Cypher5 (**Figure 3.5B**). Next, we performed a prediction for the kinases possible for phosphorylate at least two of these sites, and interestingly, CDK1/5, GSK3 and CAMKII showed high probability for the phosphorylation of Ser120 and Ser122. Furthermore, GSK3 and CAMKII were also predicted to phosphorylate Ser216 and Ser222.

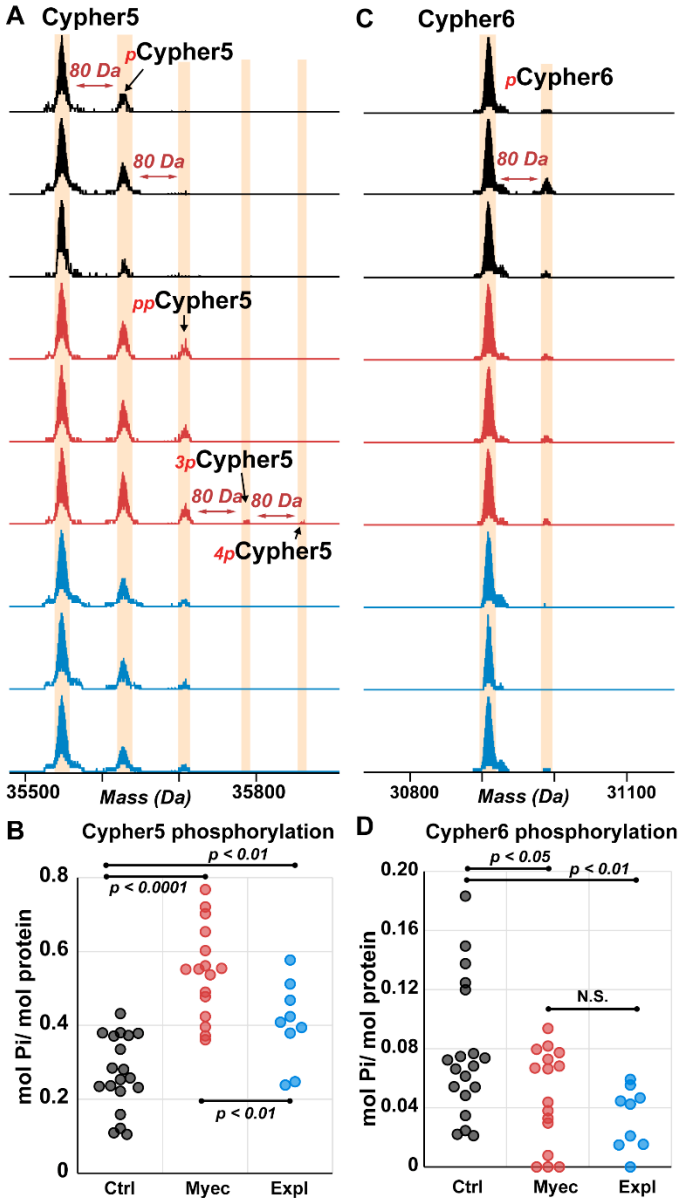


Figure 3.4. Changes of Cypher5 and Cypher6 phosphorylation in HCM. Representative mass spectra for (A) Cypher5 and (C) Cypher6 from donor hearts (black), myectomy HCM tissues (red), and explanted HCM heart tissues (blue). Quantification of total protein phosphorylation for (B) Cypher5 and (D) Cypher6 in donor hearts (Ctrl), and myectomy (Myec) and explanted (Expl) HCM hearts.

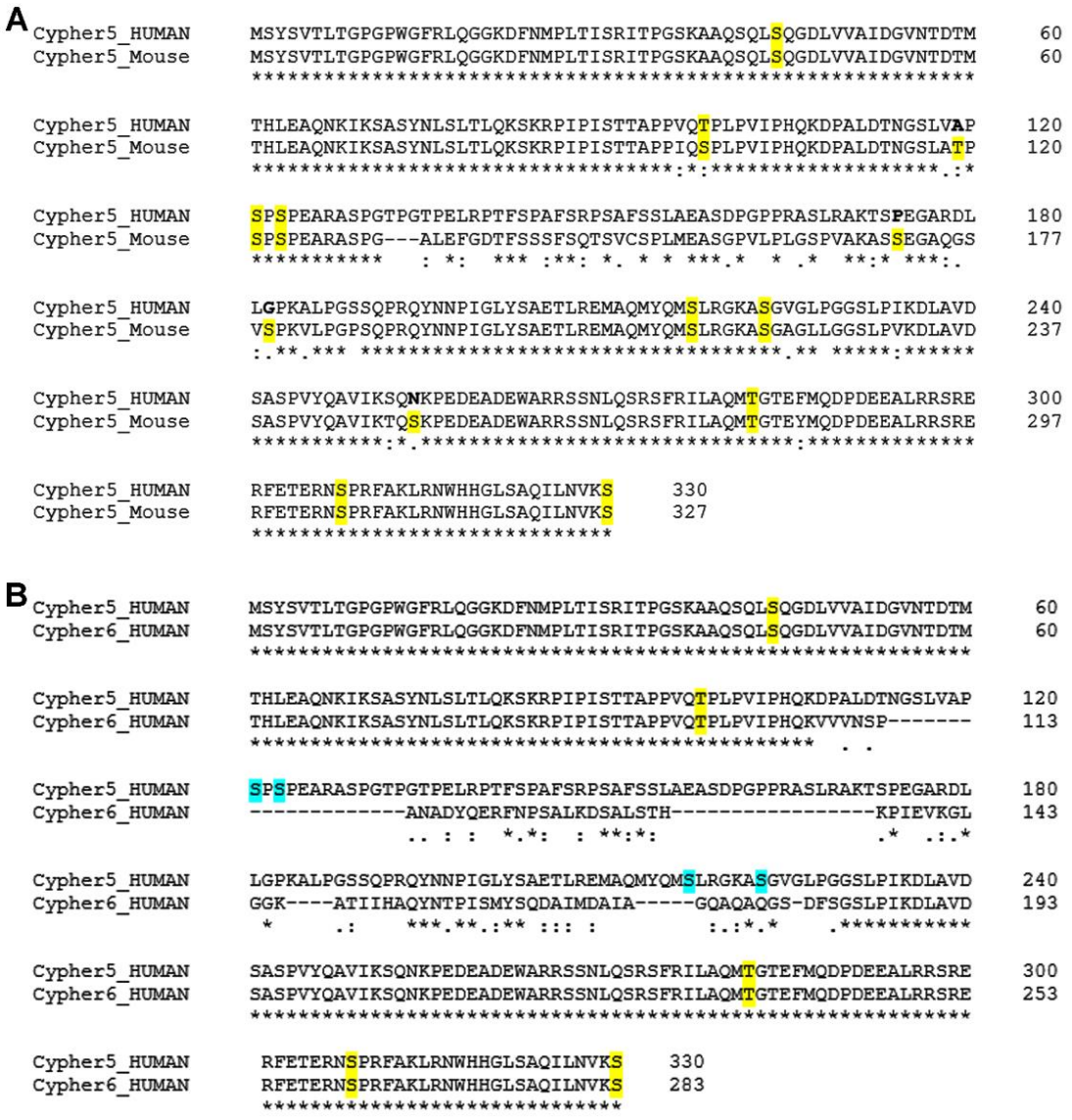


Figure 3.5. Protein sequence homology analysis. (A) Sequence alignment of human and mouse Cypher5. Yellow shade highlights the identified phosphorylation sites from a previous study.[306] (B) Sequence alignment of human Cypher5 and Cypher6. Yellow shade highlights the possible phosphorylation sites on both Cypher5 and Cypher6. Cyan shade highlights the sites unique to Cypher5.

3.3.5 CRIP2 phosphorylation remained unchanged but cysteine thiol-sulfenylation increased in explanted HCM heart tissues

CRIP2 was identified in the sarcomeric protein enriched fractions with variable level of phosphorylation (**Figure 3.6A, B**). There was no statistically significant difference in the total phosphorylation of CRIP2 between the donor and HCM heart tissues (**Figure 3.6B**). Interestingly, we identified various coeluting proteoforms that are different from the CRIP2 mass by one or multiple 48 Da (**Figure 3.6B**). Given that CRIP2 is a cysteine rich protein, the +48 Da likely resulted from thiol-sulfenylation of the cysteine residues.[307] More importantly, the level cysteine thiol-sulfenylation were significantly increased in the explanted HCM heart tissues by ~3-fold compared to the donor hearts (**Figure 3.6C**). Intriguingly, the level of cysteine thiol-sulfenylation were on average lower in the HCM myectomey tissues compared to the donor hearts (**Figure 3.6C**).

CRIP2 contains two LIM domains, and is highly expressed in the heart during development and in the adult hearts.[308, 309] A previous study showed that CRIP2 functions as a transcriptional suppressor by inhibiting the NF-κB-mediated pro-angiogenic signaling in cancer.[310] More recently, Zhou et al. showed that CRIP2 plays an important role in the inhibition of HIF-1α-mediated glycolysis in squamous cell carcinoma.[311] The function of CRIP2 in the heart has not been well defined. In this study, we identified two types of PTMs on CRIP2, including phosphorylation and thiol-sulfenylation. In contrast to protein phosphorylation, thiol-sulfenylation of proteins has not been extensively investigated. A recent study suggested protein tyrosine kinases such as EGFR were thiol-sulfenylated upon NADPH oxidase activation, suggesting a novel mechanism of redox signaling in the cells.[307] The discovery of CRIP2 thiol-sulfenylation and the fact that CRIP2 thiol-

sulfenylation increased in explanted HCM heart tissues suggested that CRIP2 may play an important role in cardiac redox signaling in health and diseases.

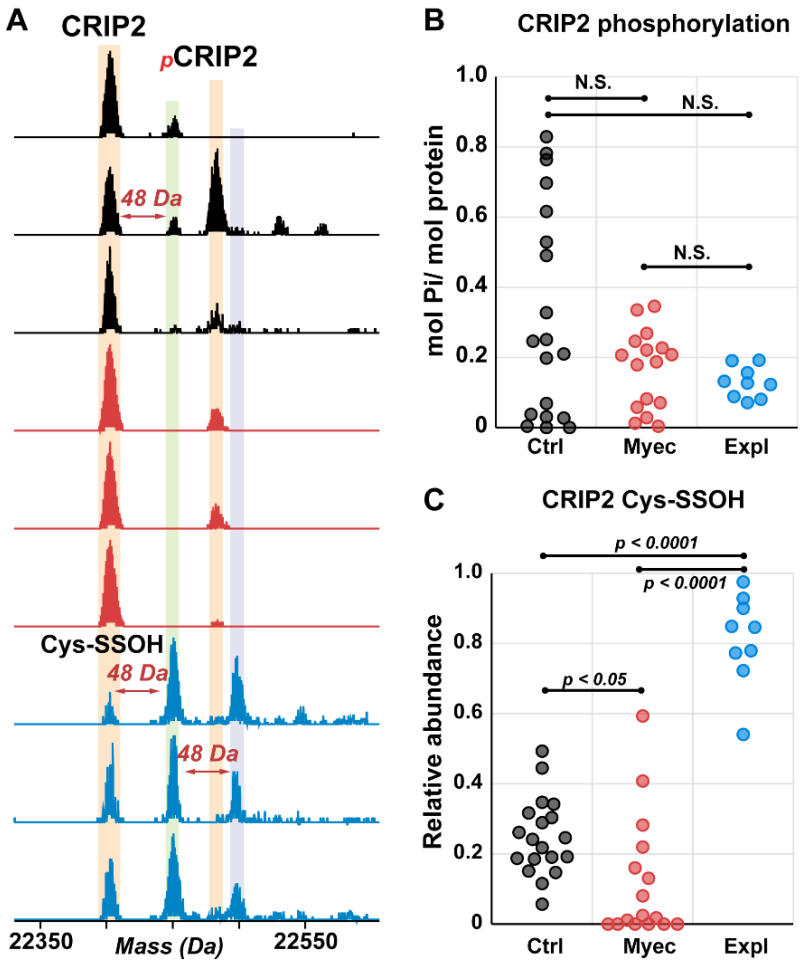


Figure 3.6. Identification of CRIP2 with phosphorylation and thiol-sulfenylation. (A) Representative mass spectra for CRIP2 from donor hearts (black), myectomy HCM tissues (red), and explanted HCM heart tissues (blue). Quantification of total CRIP2 (B) phosphorylation and (C) thiol-sulfenylation in donor hearts (Ctrl), and myectomy (Myec) and explanted (Expl) HCM hearts.

3.3.6 Total phosphorylation of MYOZ2 remained unchanged in HCM

MYOZ2 (also known as calsarcin-1) was identified in the sarcomeric protein enriched fractions with multiple phosphorylated proteoforms (**Figure 3.7A**). Mono-phosphorylated

MYOZ2 was the predominant MYOZ2 proteoform observed in all the tissue samples (**Figure 3.7A**). There was no statistically significant difference in the total phosphorylation of MYOZ2 between the donor and HCM heart tissues (**Figure 3.7B**). However, it is possible that phosphorylation level at individual residues are altered, which requires further investigation. A previous study identified 9 phosphorylation sites on mouse MYOZ2,[306] of which 8 residues are conserved between mouse and human MYOZ2 (Ser51, Ser95, Ser101, Thr107, Thr11, Ser116, Ser190 and Ser226) (**Figure 3.8**).

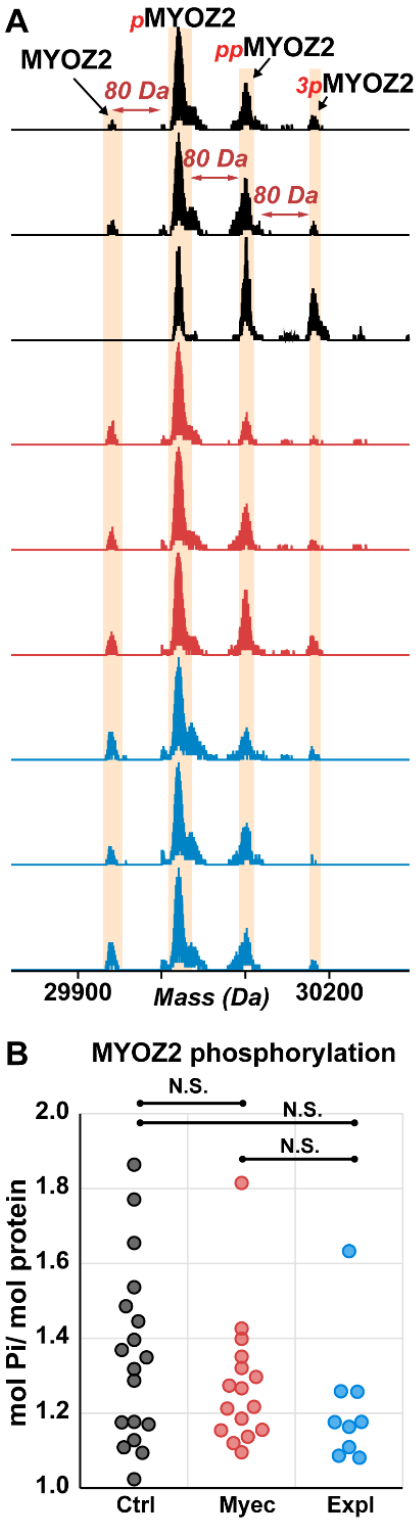


Figure 3.7. Identification of MYOZ2 with phosphorylation. (A) Representative mass spectra for MYOZ2 from donor hearts (black), myectomy HCM tissues (red), and explanted HCM heart tissues (blue). Quantification of total MYOZ2 (B) phosphorylation in donor hearts (Ctrl), and myectomy (Myec) and explanted (Expl) HCM hearts.

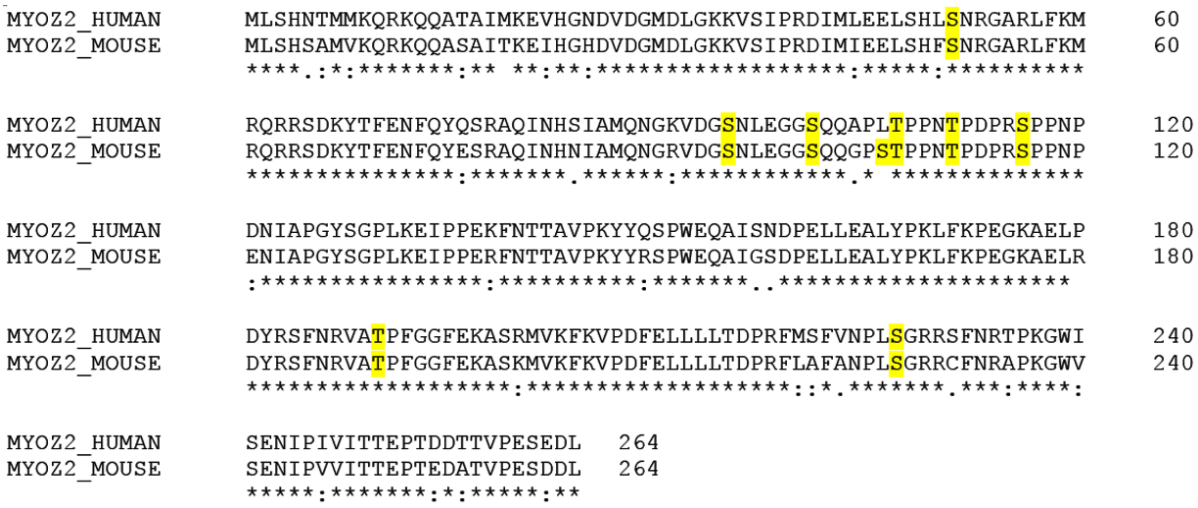


Figure 3.8. MYOZ2 sequence homology analysis. Sequence alignment of human and mouse MYOZ2. Yellow shade highlights the identified phosphorylation sites from a previous study.[306]

3.4 DISCUSSIONS

There are a number of confounding factors concerning the analysis of human heart tissues. The HCM patients, from whom cardiac tissues were collect, were using a variety of medications for symptomatic management, including β-blocker, angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers, and calcium or sodium channel inhibitors. In addition, due to the fact that surgical procedures required anesthesia, the observed blunt adrenergic response in the HCM groups, as demonstrated by decrease in PKA-medicated phosphorylation, could be a result of anesthesia. Furthermore, the explanted HCM hearts exhibited significant variation, which could be caused by co-morbidities of individual patients. Some patients experienced complete heart block, while some exhibited LV dilation with LV dysfunction. Moreover, one patient had rheumatoid arthritis and history of embolic transient ischemic attack, which may further complicate the cardiac phenotypes. Further analysis on a larger population of HCM with end-stage

heart failure is necessary to reveal the molecular alterations during the transition from hyperdynamic stage to contractile dysfunction at the failing stage.

Another limitation of human tissue analysis is that only the patients with severe symptoms underwent myectomy or transplantation. However, HCM is a highly heterogeneous disease with a wide spectrum of clinical presentations.[1] Some patients may remain asymptomatic or mildly affected without significant deterioration of life quality. The assessment of sarcomeric protein PTMs is nearly impossible among the HCM patients with mild disease presentations due to the invasive nature of the surgical procedures. Therefore, it remains unclear whether the sarcomeric protein alterations are features of the severely affected patients or the entire HCM patient population. The phenotypic heterogeneity of HCM is likely due to diverse genetic and environmental modifiers, and therefore, transgenic animal models are inadequate to address this question due to homogeneous genetic background. Recent advances in stem cell technology and human pluripotent stem cell-derived cardiomyocytes (hPSC-CMs) provide unprecedented opportunities to study genetic modifiers of HCM. hPSC-CMs can be derived from individual HCM patients with different phenotypic expression for the assessment of the key molecular alterations involved in HCM pathogenesis and progression. Furthermore, hPSC-CMs from different patients with the same HCM-causing mutations will allow for identification of key modifiers for the prevention and treatment of the disease.

Despite the limitations of human cardiac tissue analysis, this study revealed interesting HCM-associated alterations in the sarcomeric protein PTMs. In addition to decreased cTnI and ENH2 phosphorylation, α-Tpm and MLC2 phosphorylation were also down-regulated in HCM. In the contrary, phosphorylation of cTnT was found slightly increased in HCM and the potential of novel cTnT phosphorylation involved in hypertrophic signaling is exciting. Furthermore, we

identified two Cypher isoforms, the phosphorylations of which were differentially regulated in HCM. Cypher5 phosphorylation increased, whereas Cypher6 phosphorylation decreased in the HCM tissues, suggesting distinct roles of the Cypher isoforms in cardiac hypertrophy. While the full-length Cypher isoform containing both the N-terminal PDZ and C-terminal LIM domains was previously identified to be a PKA anchoring protein important for mediating phosphorylation of L-type calcium channel by PKA,[312] the roles of the the shorter Cypher isoforms (Cypher5 and Cypher6) with only the N-terminal PDZ domains remain elusive. Future studies using transgenic animals to restrict the expression of Cypher to certain isoform are important to reveal the role of Cypher in cardiac physiology and pathophysiology.

In this study, we also identified a novel PTM which is sulfenylation of cysteine on CRIP2. Cysteine sulfenylation was only recently reported to be a PTM on proteins such as EGFR.[307] Interesting, there are a number of cysteine-rich proteins, including CRIP2 and MLP, which are highly expressed in the cardiomyocytes. The regulatory functions of the cysteine residues on these proteins were not fully understood. Cysteine sulfenylation may represent a novel mechanism of cardiac redox-signaling. Future studies to identify enzymes that mediate cysteine sulfenylation and the reversal of the event are important to reveal the additional layer of regulation of the cysteine-rich proteins. Additionally, identification of the sulfenylated cysteine interactome is necessary to understand the role of this novel PTM.

CHAPTER 4

UNBIASED PROTEOMICS BENCHMARKING OF MATURATION OF HUMAN PLURIPOTENT STEM CELL-DERIVED CARDIOMYOCYTES

ABSTRACT

Human pluripotent stem cells-derived cardiomyocytes (hPSC-CMs) provide an unprecedented opportunity for patient-specific cardiac disease modeling, drug discovery, and cell-based therapy development. However, in vitro derived hPSC-CMs exhibit the properties of fetal CMs, which limit their applications. Various methods have been used to promote maturation of hPSC-CMs; however, there is a lack of consensus regarding the molecular markers for accurate and specific tracking of hPSC-CM maturity. Herein, we developed an unbiased proteomics method integrating high-throughput top-down targeted proteomics and bottom-up global proteomics for accurate and comprehensive assessment of hPSC-CM maturation. Utilizing hPSC-CMs from early- and late-stage two-dimensional monolayer culture and three-dimensional engineered cardiac tissue, we demonstrate that this method enables the detection of known maturation-associated contractile protein alterations with high reliability and reproducibility. More importantly, we identified and validated candidate maturation markers involved in sarcomere organization, cardiac excitability, and Ca^{2+} homeostasis in hPSC-CM maturation. We envision this unbiased proteomics method will provide accurate and specific benchmarking of hPSC-CM maturation.

4.1 INTRODUCTION

Since viable human cardiac tissue for experimental use is a highly scarce resource and adult human cardiomyocytes (CM) have limited proliferation potential, directed differentiation of human pluripotent stem cells (hPSCs) into CMs provides an unprecedented opportunity for patient-specific cardiac disease modeling, drug discovery, cardiotoxicity screening, and cell-based therapy development.[138, 313-315] However, hPSC-CMs exhibit the morphological, electrophysiological, metabolic, and contractile properties of fetal CMs, which limit their applications.[138, 316, 317] Promoting hPSC-CMs to acquire the key features of adult CMs is critical to realize the full promise of the hPSC-CM technology. Consequently, significant effort has been dedicated to promoting maturation of hPSC-CMs, including biochemical, mechanical, and electric interventions, as well as tissue engineering.[316-319] Traditionally, cell morphology, contractile function, electrophysiological properties, and changes in protein expression have been used as indicators of hPSC-CM maturation.[316] However, there remains no consensus regarding which markers are most suitable for accurate benchmarking of hPSC-CM maturity.[317] Unfortunately, some of the frequently used markers are not unique to hPSC-CMs, and the expression of others are highly variable and are not always indicative of maturation. Moreover, a number of standard markers are highly dependent on the measurement conditions and some markers could even provide a false assessment if considered as a stand-alone marker of maturation.[317] Hence, there is an urgent need to establish an unbiased method for the accurate and specific benchmarking of hPSC-CM maturity.

Discovery-based omics technologies have revolutionized research by enabling large-scale analysis of molecular constituents of biological systems without *a priori* knowledge and, thus, can provide an unbiased method to assess hPSC-CM maturity.[320] Although transcriptome analysis of

hPSC-CMs yields information regarding the differential regulation of genes at the mRNA level,[321] transcript levels by themselves are not sufficient to predict protein levels.[322] Moreover, transcript analysis offers no information regarding protein post-translational modifications (PTMs) which play essential roles in the regulation of protein activity and function.[244, 323] Therefore, direct interrogation of the proteome is necessary to provide definitive evidence of differential expression of protein isoforms and changes in their PTMs. The conventional mass spectrometry (MS)-based bottom-up proteomics (analysis of peptides digested from proteins) is a powerful method for the rapid and large-scale determination of differential regulation of thousands of proteins;[323] however, bottom-up proteomics is suboptimal for detecting the changes of protein isoforms especially those with high sequence homology and protein PTMs due to difficulties in achieving full sequence coverage.[153] In contrast, despite relatively limited proteome coverage, top-down MS-based proteomics (analysis of intact proteins without proteolytic digestion) is the most powerful technology for quantitatively assessing diverse protein PTMs and sequence variants simultaneously in one spectrum.[207, 244]

Here we present an unbiased MS-based proteomics method integrating top-down targeted proteomics and high-throughput bottom-up global proteomics for accurate and comprehensive assessment of hPSC-CM maturation. We have demonstrated high reproducibility and reliability of the targeted top-down method for concurrent quantification of sarcomeric (contractile) protein isoforms and PTMs. Subsequently, this top-down method was applied to the analysis of hPSC-CMs from early- and late-stage two-dimensional (2D) monolayer culture, as well as hPSC-CMs cultured in three-dimensional (3D) engineered cardiac tissue (ECT) constructs,[324, 325] for the assessment of hPSC-CM maturation. Moreover, we utilized bottom-up quantitative proteomics to assess changes in protein expression globally during hPSC-CM maturation. This integrated

proteomics platform not only enabled the detection of molecular changes known to occur during hPSC-CM maturation, but also facilitated the discovery of candidate markers of maturation including alteration in PTMs of key contractile proteins and other non-contractile molecules involved in sarcomere organization, cardiac excitability, and Ca^{2+} homeostasis. Subsequent validation of the identified changes in the ventricular myocardium of mice during development provided evidence that these alterations are *bona fide* markers of hPSC-CM maturation.

4.2 MATERIALS AND METHODS

4.2.1 Reagents and chemicals

All cell culture reagents were purchased from Invitrogen unless otherwise specified. All reagents used for preparation of samples for MS analysis were purchased from Sigma-Aldrich unless otherwise noted. High-performance LC reagents were purchased from Fisher Scientific.

4.2.2 hiPSC and hESC culture

One hiPSC line derived from foreskin fibroblasts without integration of vector and transgene, DF19-9-11T, and one hESC line, H1, were used in this study. hiPSCs and hESCs were maintained on Matrigel (GFR, BD Biosciences)-coated 6-well plates in mTeSR1 medium. Cells were passaged using Versene Solution every 4-5 days. Cell lines were between passage 30 and 75.

4.2.3 Human cardiac tissue collection and handling

A donor heart with normal cardiac function, but unsuitable for transplantation was collected from the University of Wisconsin Hospital and Clinic following procedures approved by

the Institutional Review Board of the University of Wisconsin-Madison. The donor (de-identified) was a 60-year-old man with no past heart disease history, and the echocardiogram was normal. The donor heart was maintained in cardioplegic solution and dissected on a metal plate chilled with dry ice. The tissue was cut into small pieces and flash frozen in liquid nitrogen and stored at -80 °C for later experiments. The left ventricular midwall tissue was used for top-down proteomics quantification of sarcomeric protein isoforms and their PTMs.

4.2.4 CM differentiation and culture

hiPSCs and hESCs were differentiated into CMs using a small molecule-directed protocol as described previously.[314, 315, 326] Briefly, hPSCs maintained on the mTeSR1/Matrigel system were dissociated into single cells and seeded onto Matrigel-coated 6-well plate at 1.5–2.0×10^6 cells/well in mTeSR1 medium supplemented with 10 μM Y-27632 (ROCK inhibitor, Tocris Bioscience, Bristol, UK). Cells were cultured for 5 days in mTeSR1 medium until reached 100% confluence. The differentiation (Day 0) started with changing mTeSR1 medium to 2.5 mL/well RPMI + B27 supplement without insulin and supplemented with 12 μM CHIR99021 (GSK-3β inhibitor, Tocris Bioscience), and cells were treated in the medium for 24 hrs (Day 1). At day 1, medium was changed to 3 mL/well RPMI+B27 without insulin and cells were cultured in this medium for 48 hrs (Day 3). 72 hrs (Day 3) after addition of CHIR99021, a combined medium was prepared by collecting 1.5 mL of medium from the wells and mixing with same volume of fresh RPMI+B27 without insulin medium and supplemented with 5 μM IWP2 (Tocris Bioscience), and cells were treated with IWP2 in the medium for 48 hrs (Day 5). The medium was changed to 3 mL RPMI+B27 without insulin on day 5. The medium was changed to RPMI+B27 complete supplement (with insulin) on day 7, and the differentiated cells were maintained in this medium

until day 15 with medium change every other day. The Day 15 differentiated CMs with a purity of 50-98% $cTnT^+$ cells were cryopreserved in the medium of 90% FBS+10%DMSO. The cryopreserved hPSC-CMs were thawed on the SyntheMax (Corning) coated 6-well plate in EB20 medium[12] and allow the cells to attach for 2 days, and then the hPSC-CMs were purified in the lactate medium of RPMI without glucose plus B27 supplement and 5 mM sodium DL-lactate (Sigma-Aldrich) for 10 days. hPSC-CMs were maintained in RPMI with B27 supplement until day 30 or day 60 for mass spectrometry analysis. hPSC-CMs were dissociated into single cells by incubation with 0.25% trypsin-EDTA for 5 min at 37 °C, 5 x 10^5 cells were pelleted as one sample.

4.2.5 Flow cytometry

The hPSC-CMs were fixed with 1% paraformaldehyde in 37 °C for 10 min in the dark, and permeabilized in ice-cold 90% methanol for 30 min on ice. The cells were washed once in FACS buffer (Mg^{2+}/Ca^{2+}-free DPBS, 0.5% BSA, 0.1% NaN_3) with 0.1% Triton X-100 (Sigma-Aldrich). After centrifugation, the supernatant was discarded leaving about 50 µL. Primary antibodies (anti-cTnT, cardiac isoform Ab-1, Clone 13-11, ThermoFisher Scientific, catalog#: MS-295-P) were diluted (1:200) in FACS buffer with 0.1% Triton X-100, and 50 µL diluted antibodies was added to each sample to make a total sample volume of 100 µL, followed by incubation at 4 °C for overnight. The cells were washed with 3 mL FACS buffer with 0.1% Triton X-100 and centrifuged. The supernatant was discarded leaving about 50 µL. Corresponding secondary antibodies were diluted in FACS buffer with 0.1% Triton X-100 and 50 µL was added to a final sample volume of 100 µL with the secondary antibody dilution being 1:1000. The samples were incubated for 30 min in dark at room temperature, washed with FACS buffer containing 0.1% Triton X-100. The supernatant was discarded and the samples were resuspended

in 300-500 µL FACS buffer with 0.1% Triton X-100 for flow cytometric analysis. The data were collected on a FACS Caliber flow cytometer (Becton Dickinson) and analyzed using FlowJo v8.5.2.

4.2.6 Immunocytochemistry

Primary antibodies were anti-MLC2v (ProteinTech Group, catalog#: 10906-1-AP, 1:200), and anti-MLC2a (Synaptic Systems, catalog#:311011, 1: 400). hPSC-CMs were fixed with 4% paraformaldehyde for 15 min at room temperature, and permeabilized in 0.2% Triton X-100 for 0.5-1 hr at room temperature. Cells were blocked with 5% non-fat dry milk (Bio-Rad) in 0.2% Triton X-100 solution and incubated for 1.5-2 hrs at room temperature on a rotator followed by two washes with DPBS. Primary antibodies were added in 1% BSA in DPBS solution with 0.1% Triton X-100, and incubated overnight at 4 °C. Cells were washed with 0.2% Tween 20 in DPBS twice. Secondary antibody specific to the primary IgG isotype were diluted (1:1000) in the same solution as the primary antibodies and incubated at room temperature for 1 hr in dark on a rotator. Cells were washed with 0.2% Tween 20 in DPBS twice. Nuclei were labeled with Hoechst. Imaging was carried out on EVOS epifluorescence microscope (Life technologies).

4.2.7 ECT preparation

CMs were differentiated from DF19-9-11T hiPSCs using the small molecule protocol as describes above, but without selection by lactate medium. At day 30 of differentiation, cells were dissociated with TrypLE from monolayers consisting of 70-90% beating hiPSC-CMs. The cells were subsequently suspended in 20 mL ECT medium (60.3% high-glucose DMEM; 20% F12 nutrient supplement; 1 mg/mL gentamicin; 8.75% FBS; 6.25% horse serum; 1% HEPES; 1x nonessential amino acid cocktail; 3 mM sodium pyruvate; 0.004% (wt/vol) $NaHCO_3$; 1 µg/mL

insulin; 400 µM tranexamic acid and 17.5 µg/mL aprotinin) and incubated for 1 hr on a rotating platform at 37 °C to form small and uniform clusters of viable cells. Approximately 2×10^6 hiPSC-CMs from the rotational culture were mixed with 1.25 mg/mL fibrinogen and 0.5 unit of thrombin in 200 µL ETC medium and loaded onto a 20x3 mm cylindrical mold of Flexcell Tissue Train silicone membrane culture plate, followed by incubation under pre-programmed vacuum condition for 60 min at 37 °C supplied with 5% CO_2 to allow for attachment of the ECT constructs to the nylon tabs at each end of the Tissue Train well. Following polymerization of the fibrin matrix, ECT were fed with ECT media and cultured for 14 to 44 days with media changes every 2 days. Assessment of ECT function and histology were performed as previously described.[325]

4.2.8 Sarcomeric protein enrichment from hPSC-CMs

Freshly isolated hPSC-CMs were washed with 1 mL Mg^{2+}/Ca^{2+}-free DPBS solution, and centrifuged at 1,100 g for 10 min at 4 °C. The procedures for enriching the cardiac sarcomeric proteins were described previously. Briefly, the wash solvent was removed and the pellets were homogenized in 50 µL HEPES buffer (25 mM HEPES pH 7.4, 60 mM NaF, 1 mM DTT, 1 mM L-methionine, 1 mM PMSF, 1 mM Na_3VO_4 containing protease and phosphatase inhibitors) at 4 °C. The HEPES extract containing predominantly cytosolic proteins were collected and the pellets enriched of sarcomeric proteins were frozen at – 80 °C for later experiments. The pellets depleted of cytosolic proteins were homogenized in 40 µL trifluroacetic acid (TFA) solution (1% TFA, 5 mM tris(2-carboxyethyl)phosphine), 5 mM L-methionine) at 4 °C, followed by centrifugation at 17,000 g for 20 min. Bradford protein assay was employed to determine the protein concentration and the samples were diluted to 100 ng/µL for top-down LC-MS analysis. The samples to be quantitatively compared were analyzed as continuous LC-MS runs to minimize instrumental variation.

4.2.9 Sarcomeric protein enrichment from ECT

The ECT construct was detached from the nylon tabs by cutting both ends of the construct. The ECT construct was then transferred to 1.5 mL microcentrifuge tubes, and washed with 1 mL Mg^{2+}/Ca^{2+}-free DPBS solution, followed by centrifugation at 1,100 g for 10 min at 4 °C. The washing solution was removed and 50 µL of the same HEPES buffer described above was added to each construct, and the construct was homogenized at 4 °C with a Teflon pestle. Followed by centrifugation at 17,000 g for 20 min, the supernatant was collected and the remaining pellet was further homogenized in 40 µL TFA solution. The TFA fraction containing predominantly sarcomeric proteins were collected for analysis by top-down LC-MS.

4.2.10 Reversed phase liquid chromatography and top-down MS analysis

LC-MS analysis was carried out using a NanoAcquity Ultra-high Pressure LC system (Waters) coupled to a high-resolution Impact II quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker Daltonics). To evaluate the reproducibility of the method and the linearity of instrument response, three injection replicates were performed at 100, 200, 300, 500, 750 ng of total proteins injected. The sarcomeric proteins in each sample were separated on a home-packed PLRP column (250 mm in length, 500 µm in inner diameter, 1000 Å pore size, 10 µm particle size) by a gradient of 5% to 95% mobile phase B (Mobile phase A: 0.1% formic acid in water, mobile phase B was 0.1% formic acid in 50:50 acetonitrile: ethanol) at a constant flow rate of 12 µL/min. Proteins eluted were delivered to the mass spectrometer via electrospray ionization. End plate offset and capillary voltage were 450 V and 4000 V, respectively. Nebulizer pressure was set to 0.5 Bar and dry gas flow rate was 4.0 L/min. In-source collisional energy was set to 10 eV. Mass spectra were collected at a scan rate of 0.5 Hz over the 500-2000 *m/z* range.

4.2.11 Online tandem MS (MS/MS) analysis and protein identification

Data-dependent online MS/MS analysis was performed by selecting the top 5 most abundant ions, followed by fragmentation via collision induced dissociation (CID) with the collisional energy between 15 eV and 30 eV. The scan rate for MS was set to 1 Hz and for MS/MS was set to 8 Hz for precursors with signal intensities higher than $1x10^5$, or 4 Hz for precursor ions with signal intensities between $1x10^3$ and $1x10^5$. The selected ions were excluded for 1 min. Alternatively, the ions between 700-800 *m/z* were collectively isolated and fragmented using collisional energy of either 20 or 25 eV in a single LC-MS/MS run after a full LC-MS run. The intact masses of the precursor ions were determined based on the first LC-MS run, and the fragment ions were determined by the following LC-MS/MS run. The mass spectra or tandem mass spectra were exported from the DataAnalysis v4.3 (Bruker Daltonics) as .ascii file and processed using MASH Suite Pro,[229] a software developed in-house. All ions were manually validated and protein identification was performed using MASH Suite Pro equipped with the MS-align+[228] algorithm.

4.2.12 Quantification of sarcomeric protein isoforms and PTMs

All top-down LC-MS data were processed and analyzed using the DataAnalysis v4.3 for protein quantification. All chromatograms were smoothened using the Gauss algorithm with a smooth window of 2.027. Briefly, for a specific protein isoform, the most abundant 3-5 charge state ions that do not overlap with other ions were selected for generating extracted ion chromatogram (EIC). The area under curve (AUC) of the EIC of a specific protein isoform represents the abundance of the selected protein isoform. The protein abundance was further adjusted to account for oxidation and other modifications of the same protein. The same ions were selected for the same protein isoform across all samples to be compared. Relative abundance of

the selected protein isoforms was measured by the ratio of their AUCs. Relative quantification of protein PTMs were performed as reported previously.[160, 254, 299] Briefly, the mass spectra were deconvoluted using the Maximum Entropy[327] algorithm, and the abundance of the modified protein form is calculated as the ratio of its peak height to the summed peak heights of all the protein forms of the same protein.

4.2.13 Statistical analysis for protein expression and PTMs by top-down proteomics

The student's *t*-test (two-tailed) was performed to evaluate the statistical significance of variation between groups for the top-down datasets. Differences between means were considered to be significant at $p < 0.05$.

4.2.14 Global label-free quantitative proteomics analysis

Approximately 0.4 million hPSC-CMs were washed with Mg^{2+}/Ca^{2+}-free DPBS solution and centrifuged at 1,100 g for 10 min, and the wash solvent was removed. The pellets were stored in -80 °C until time of analysis. 10 µL of lysis buffer (50 mM ammonium bicarbonate, pH 7.3, 0.4% anionic acid labile surfactant II (Protea), 0.2% non-ionic acid labile surfactant II (Protea), 10 mM L-methionine, 5 mM DTT, 1 mM PMSF, 1x protease and phosphatase inhibitors) were added to the cell pellets and the vials were vortex briefly, followed by sonication for 5 min at room temperature. The pellets were heated at 80 °C for 10 min and quickly frozen-thawed for 3 cycles. The vials were vortexed and 40 µL of 25 mM ammonium bicarbonate were added to each sample. The samples were sonicated for 10 min and incubated at 65 °C for 15 min, then centrifuged at room temperature for 20 min at 21,000 g. The supernatant was collected and Bradford protein assay was performed to determine the protein concentration. The samples were diluted to the same

concentration (0.4-0.8 mg/mL) and 100 µL of the protein extract was collected and reduced with 10 mM DTT at 37 °C for 1 hr. Freshly prepared iodoacetamine solution (0.25 M) was added to a final concentration of 25 mM and the samples were incubated in dark for 30 min. The protein extracts were digested with Trypsin Gold (Promega) at an enzyme-substrate mass ratio of 1:50 and the samples were incubated at 37 °C for overnight. The digested peptides were loaded onto a C18 trap for desalting for 10 min and then separated on a 250 mm PicoFrit capillary column packed with 1.9 µm, 175 Å Hypersil Gold aQ C18 particles (New Objective) at a flow rate of 300 nL/min by a gradient of 5-45% mobile phase B (0.1% formic acid in acetonitrile) (mobile phase A: 0.1% formic acid in water). The capillary column tip was customized to fit into the CaptiveSpray nanoESI source (Bruker Daltonics) for direct delivery of samples to the mass spectrometer. The most abundant 30 ions were selected for fragmentation by tandem MS (MS/MS) with an active (dynamic) exclusion time of 45 s. Protein identification (with at least 1 unique peptide) and quantification was performed using the MaxQuant v1.5.7. [328, 329] The protein expression represented by MS signal intensity was log2 transformed and normalized to the median of the dataset, and the Welch's modified *t*-test (two-tailed) was applied to evaluate the statistical significance of change. Proteins were considered significantly up- or down-regulated with a *p*-value smaller than 0.01, and a more than 1.4-fold change in the expression level. Gene ontology (GO) analysis was performed using the online PATHER[330] tool with a focus on the biological processes, and the GO terms for the differentially regulated proteins/genes were consolidated and validated manually for the functional classification of these proteins/genes. The differentially regulated proteins/genes were analyzed using the Reactome[331] plugin in Cytoscape[332] to include linker genes, and the group attribute layout was applied based on the functional classification of

the proteins/genes. The graphic view was output as .svg files and edited using Adobe Illustrator CS5.

4.2.15 Mouse cardiac tissue collection

The collection of mouse heart tissue was performed as described previously.[333] Wild type mice were anaesthetized through inhaled isoflurane (≥ postnatal day 21) or decapitation (≤ postnatal day 10), and the hearts were harvested. All hearts were rinsed in PBS solution, and the blood ejected using blunt forceps. Under a dissecting microscope, the great vessels were removed from the hearts. The atria were then separated from the ventricles, and each was flash frozen in liquid nitrogen. A post-mortem tail snip was taken from each mouse to allow subsequent sex determination by PCR. To generate embryonic pups, mice were bred for 24 hrs, after which the males were removed from the cages to establish the timing of pregnancy. Pregnant dams were given intraperitoneal injections of heparin (100 U) 5 min prior to anaesthetizing the dams by inhaled isoflurane on day 14-15 after the breeding date. The abdomens of the dams were opened and the uterine horns containing the embryos were placed in a petri dish with PBS solution. The embryos were removed from the uterine horn and embryonic sacs. Atria and ventricles were then harvested as described above. Postnatal day 0 is considered here as day of birth. This study was approved by the Animal Care and Use Committee of the School of Medicine and Public Health at the University of Wisconsin-Madison in accordance with the Guide for the Care and Use of Laboratory Animals (National Institutes of Health publication no. 85-23, revised 1985).

4.2.16. Protein extraction from mouse ventricular tissue and immunoblotting analysis

After sex determination by PCR, the female mouse hearts were used for protein extraction and immunoblotting. Approximately 10 mg of ventricular tissue was cut and homogenized for mice ≥ postnatal day 21, and 2 mg for mice at embryonic day 15. The tissue was homogenized first in 10 µL/mg of decellularization buffer 1 (0.5% SDS, 1 mM EDTA, 20 mM L-methionine, 20 mM L-glutamine in DPBS with protease and phosphatase inhibitors). The homogenate was incubated at room temperature for 2 hrs with continuous rotation and centrifuged at 21,000 g for 30 min, and the supernatant was collected. The pellet was washed twice with 5 µL/mg of DPBS with protease and phosphatase inhibitors, and the wash solution was collected and combined with the prior fraction. The pellet was then incubated with 10 µL/mg of decellularization buffer 2 (1% Triton X-100, 0.2% SDS, 1 mM EDTA, 20 mM L-methionine, 20 mM L-glutamine in DPBS with protease and phosphatase inhibitors) at room temperature for 5 hrs and at 4 °C for 10 hrs with continuous rotation until the tissue appeared pale. The homogenate was centrifuged at 21,000 g for 30 min and the supernatant was collected and combined with the prior fractions. Another 10 µL/mg of decellularization buffer 2 was added to the pellet and incubated at room temperature for 5 hrs with continuous agitation. The pellets were centrifuged at 21,000 g for 30 min and the supernatant was collected and combined with the prior fractions. Protein concentration was determined by Bradford protein assay (Bio-Rad) and the 15 µg protein from each sample was loaded to Criterion (Bio-Rad) TGX 4-20% polyacrylamide gel for separation. Proteins were transferred to methanol-activated PVDF membrane with a constant current of 60 mA for overnight and protein staining on blot was performed using the REVERT protein stain (LI-COR Biosciences) according to the manufacturer's instruction. The membrane was blocked with 2% BSA solution in TBS for 2 hrs and incubated with individual primary antibody for overnight at 4 °C. The membrane

was washed thoroughly and incubated with corresponding secondary antibodies at room temperature for 2 hrs, followed by thorough washing. Imaging was performed by fluorescent or chemiluminescent detection using the Odyssey imager (LI-COR Biosciences). The primary antibodies for immunoblotting were Anti-α-actinin 1 (Santa Cruz: sc-135819), Anti-dystrophin (H-5) (Santa Cruz: sc-365954), Anti- α-actinin 2 (ThermoFisher Scientific: 7H1L69), Anti-phosphlamban (F-7) (Santa Cruz: sc-393990), Anti-troponin I (Cell Signaling Technology: #4002), Anti- αB-crystallin (Abcam: ab13496). The dilution was determined on a case-by-case basis to have the final antibody concentration of 0.5-1 μg/mL. All images were analyzed and bands were quantified using the Image Studio Lite software (LI-COR Biosciences).

4.3 RESULTS

hPSCs, including both human induced pluripotent stem cells (hiPSCs) and embryonic stem cells (hESCs), were differentiated into CMs using a small molecule-directed protocol,[314, 326] which produced hPSC-CMs predominantly (91.5%) of the ventricular lineage based on electrophysiological properties.[326, 334] To establish a quantitative platform for assessment of hPSC-CM maturation, we first employed a simple maturation method with extended culture in 2D monolayer (**Figure 4.1a**). Following extended culture, hPSC-CMs exhibit increased cell size, greater myofibril density and organization, and improved contractile function.[335] hPSC-CMs were cultured and collected at early and late stage (approximately 30 and 60 days post-differentiation, respectively). Additionally, hPSC-CMs at Day 30 post-differentiation were dissociated for the preparation of ECTs, and the ECTs were collected at early and late stage accordingly.

The hPSC-CMs from 2D monolayer culture and 3D ECT were used for top-down targeted proteomics analysis with a focus on the analysis of sarcomeric proteins (the contractile proteins)

(**Figure 4.1b**). Studies have shown that multiple sarcomeric protein isoforms undergo a fetal to adult transition during cardiac development,[336-338] including switching of α- to β-myosin heavy chain (α-MHC, β-MHC), slow skeletal to cardiac troponin I (ssTnI, cTnI), and the atrial isoforms of myosin light chain 1 and 2 (MLC1a, MLC2a) to the corresponding ventricular isoforms (MLC1v, MLC2v) (**Figure 4.1b**).[336-338] The ability to detect these changes was used as a means of validating the proteomics method for tracking the maturity of hPSC-CMs. Additionally, this top-down proteomics method allowed for the identification of new candidate maturation markers and/or their PTMs in the contractile apparatus. Maturation of the hPSC-CMs is known to impact expression of genes related to metabolism, excitability, Ca^{2+} cycling, and cell cycle, which all change during development.[317] To further provide an unbiased analysis for the differential regulation of global protein expression beyond the sarcomeres, we employed a label-free bottom-up quantitative proteomics method to outline the molecular alterations between the hPSC-CMs from early- and late-stage cultures (**Figure 4.1c**).

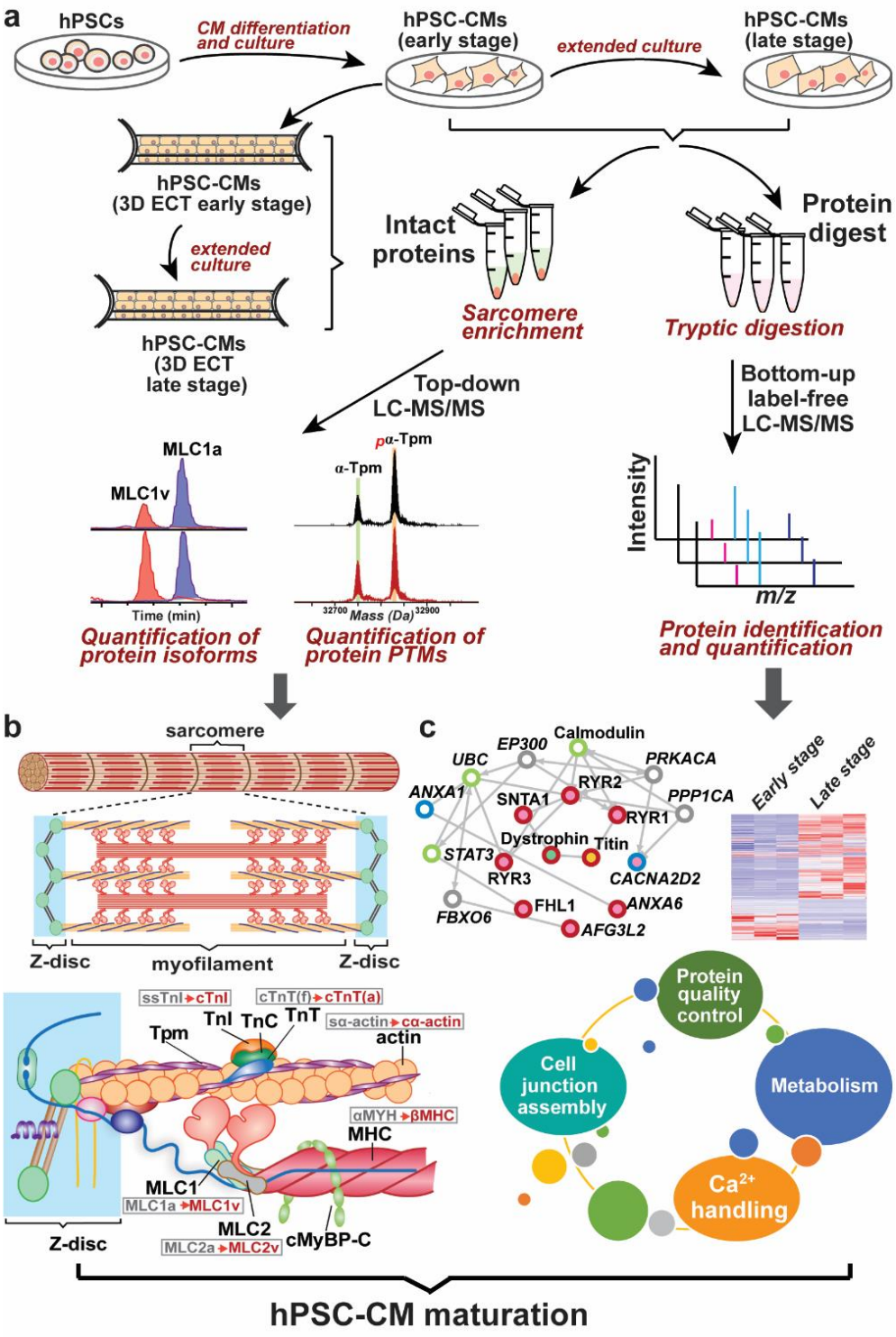


Figure 4.1. A comprehensive proteomics platform for the assessment of hPSC-CM maturation. (a) Experimental scheme for integrated top-down and bottom-up proteomics analyses of hPSC-CMs from early- and late-stage 2D monolayer culture and 3D engineered cardiac tissue (ECT). (b) Top-down LC-MS-based proteomics for concurrent quantification of contractile protein isoform expression and PTMs. cTnT(f)/(a), cardiac troponin T fetal/adult isoform; sα-actin/cα-actin, skeletal/cardiac α-actin; ssTnI/cTnI, slow skeletal/cardiac troponin I, α-MHC/β-MHC: α/β myosin heavy chain; MLC1a/MLC1v, myosin light

chain 1 atrial/ventricular isoform; MLC2a/MLC2v, myosin light chain 2 atrial/ventricular isoform. Grey and red font represents fetal and adult isoform, respectively, in the ventricular CMs. (c) Bottom-up label-free quantitative proteomics for determining global differential regulation of proteins.

4.3.1 Assessment of Reproducibility and Robustness of Top-down Proteomics for Simultaneous Quantification of Isoforms and their PTMs

To accurately quantify the protein isoform and PTM changes for the assessment of hPSC-CM maturation, we developed a robust liquid chromatography (LC)-MS-based top-down proteomics platform for targeted analysis of sarcomeric protein isoforms and their PTMs. We first evaluated the reproducibility of contractile protein enrichment from hPSC-CMs at Day 30. The cell pellets were first homogenized in a gentle neutral buffer (without detergent) to extract soluble (predominantly cytosolic) proteins and preserve the integrity of the sarcomere, followed by further homogenization in an acidic and MS-compatible solution for extracting the sarcomeric proteins.[24, 299] The extraction method has proven highly reproducible for hiPSC-CMs (DF19-9-11T line) from early- and late-stage cultures (**Figure 4.2a**). MHC was highly enriched in the sarcomeric protein fraction, and immunoblotting showed that ssTnI, the troponin I isoform expressed in immature/fetal CMs, was present exclusively in the sarcomeric protein fraction (**Figure 4.2a, b**).

After confirming reproducible protein separation and mass spectrometer response (**Figure 4.3**), we examined the linearity of the instrument response by injecting a defined amount of protein (100, 200, 300, 500, 750 ng) from the sarcomeric protein fraction (**Figure 4.2c**). Quantification of the sarcomeric protein was performed by generating extracted ion chromatogram (EIC) specific to each protein, and protein expression was represented as the area under curve (AUC) (**Figure 4.4**). From a single top-down LC-MS run, the EICs of the major sarcomeric proteins (< 50 kDa), including cardiac troponin T splicing isoform 6 (cTnT6, an adult isoform of cTnT, **Figure 4.5**),

ssTnI, α-tropomyosin (α-Tpm, formally known as Tpm2.2st[257]), MLC1v, MLC1a, MLC2v, MLC2a, and cardiac α-actin (cα-actin) were generated and the abundance of each protein is represented by the AUC (**Figure 4.2d**). cTnI remained undetected in the extract from hiPSC-CMs at Day 30, either because cTnI was not expressed or expressed in extremely low abundance (**Figure 4.2d**). A complete list of proteins identified by the top-down proteomics can be found in **Table 4.1** and **Figure 4.6**. We further confirmed the linearity of the instrument response as demonstrated by the linear correlation of the AUCs of MLC1v and MLC1a with varied total sarcomeric protein (100-750 ng) injection (**Figure 4.2e, f**), and the injection replicates proved highly consistent with less than 8% coefficient of variation (**Figure 4.2g**). The linear curves of the major sarcomeric proteins (< 50 kDa), including the above-mentioned proteins and cTnT splicing isoform 1 (cTnT1, a fetal isoform of cTnT, **Figure 4.5**), are shown in **Figure 4.2h**, and the mutual linear range of all major sarcomeric proteins is between 100 and 500 ng proteins from the sarcomeric protein extracts.

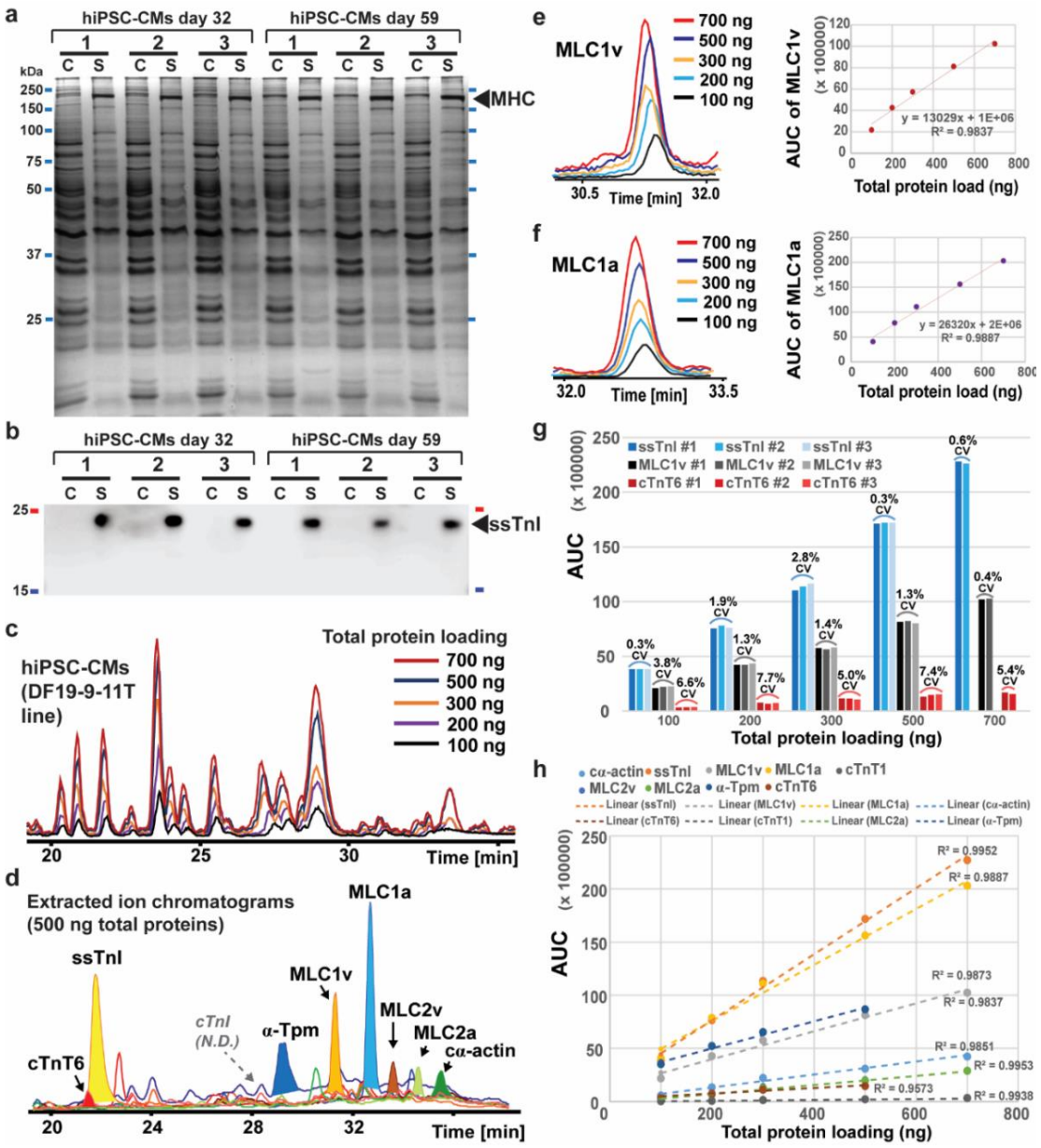


Figure 4.2. A robust and quantitative LC-MS-based top-down proteomics method for the quantification of sarcomeric protein isoforms from hPSC-CMs. (a) Assessment of the reproducibility of sample preparation method. 2 µg of proteins from the cytosolic and sarcomeric protein-enriched extracts were loaded and separated via SDS-PAGE, followed by silver staining. Cytosolic (C) and sarcomeric (S) protein fractions from three technical replicate extraction of hiPSC-CMs at day 32 or day 59 are highly consistent. (b) Western blotting showing the presence of ssTnI exclusively in the sarcomeric protein fraction. c) Base peak chromatograms (BPCs) of LC-MS analysis of hiPSC-CM sarcomeric protein fraction with various amount of total protein loading. (d) Extracted ion chromatograms (EICs) of the representative sarcomeric proteins at 500 ng of total protein loading of the sarcomeric protein-enriched fraction. The abundance of the protein is represented by the area under curve (AUC) of the corresponding EICs. N.D., not detected. cTnT6/cTnT1, cardiac troponin T splicing isoform 6 (adult)/1 (fetal). (e, f) EICs of MLC1v and MLC1a with various amount of total sarcomeric protein loading demonstrate linear response of the Q-TOF instrument. (g) Instrument reproducibility and stability assessment with three injection replicates of varied sarcomeric protein loading. The LC-MS-based quantification of three representative sarcomeric

proteins demonstrate high consistency with less than 8% coefficient of variation (CV). (h) Linear curves of the representative sarcomeric proteins.

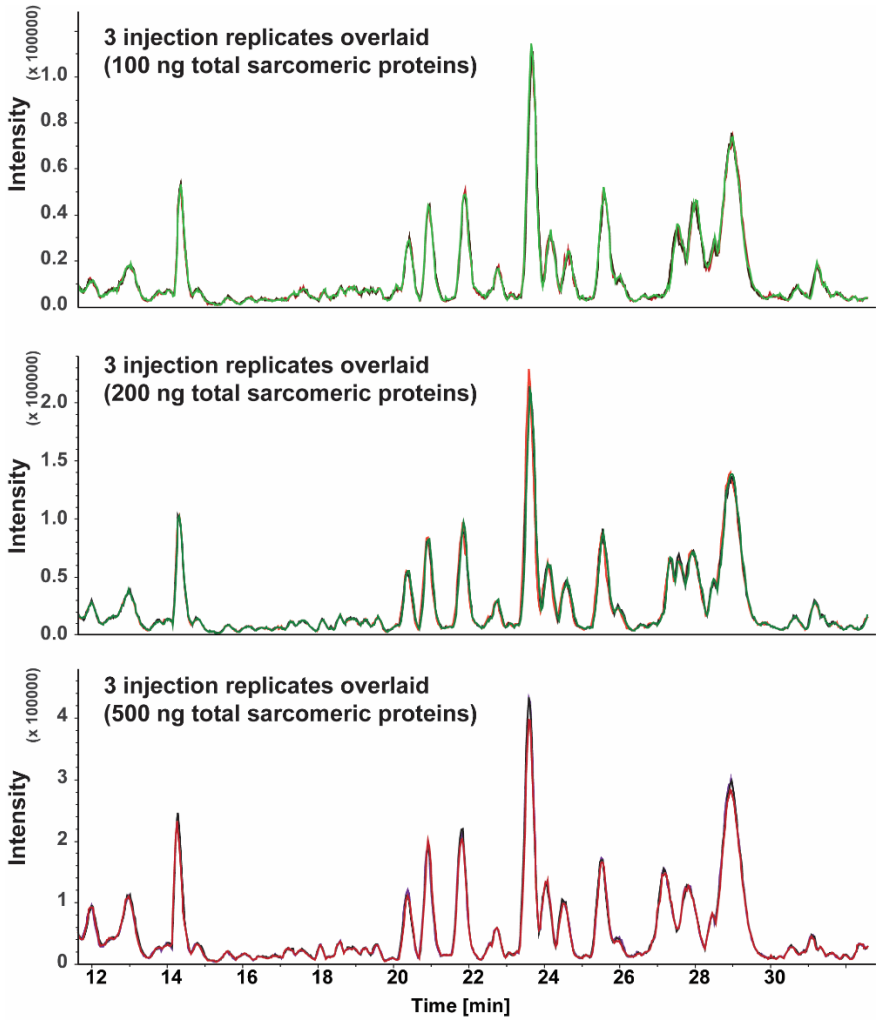


Figure 4.3. Assessment of the reproducibility of intact protein separation and the stability of the mass spectrometer. Top, middle, and bottom panels show the overlaid BPCs of three injection replicates from the injection of 100 ng, 200 ng, and 500 ng total proteins, respectively.

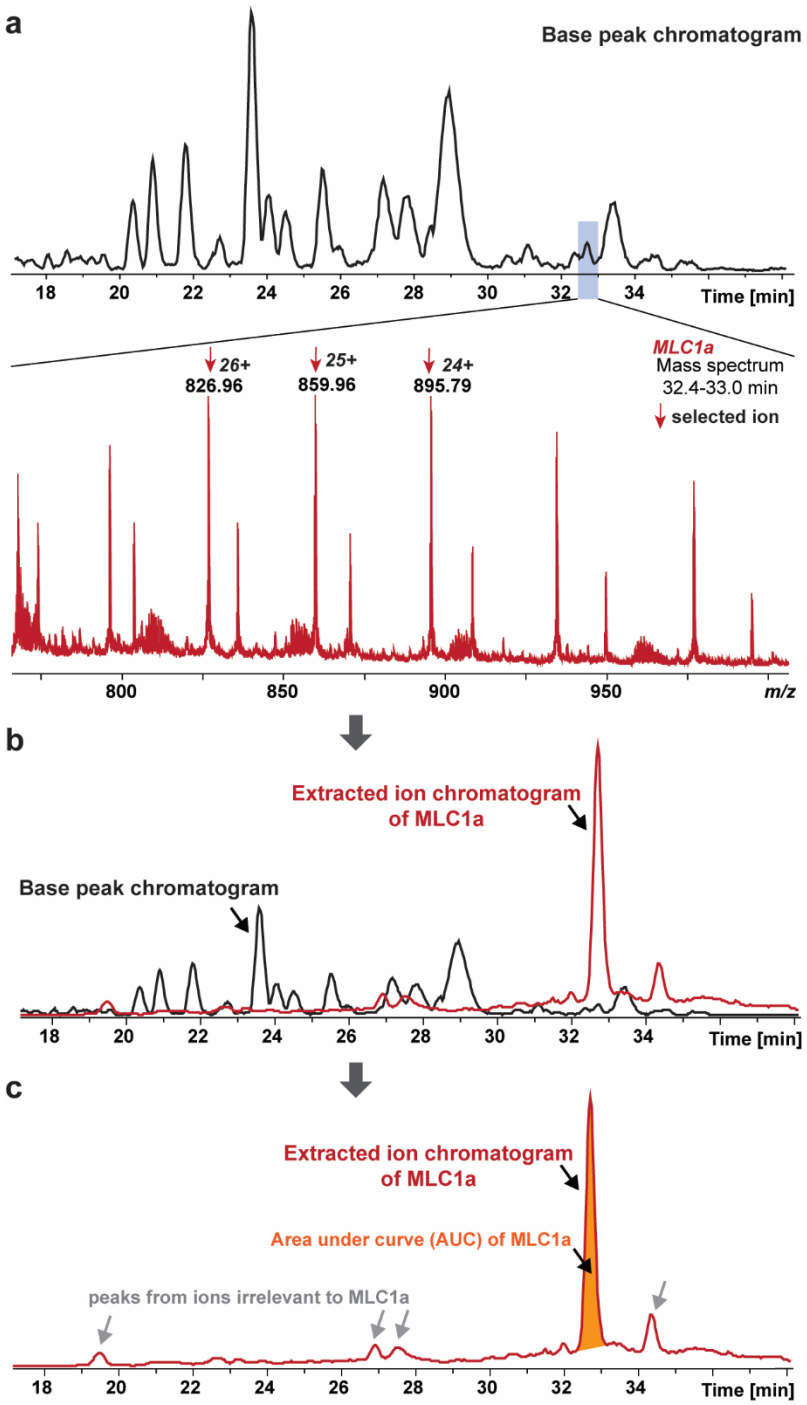


Figure 4.4. Workflow for top-down LC-MS-based protein quantification based on EICs. (a) The mass spectrum of the protein of interest (i.e. MLC1a) was surveyed, and the *m/z* values for the most abundant proteins forms from the top 3 charge states were selected to generate EICs. (b) EICs were generated based on the selected ion with a tolerance window of +/- 0.2 *m/z*. (c) The area was drawn over the elution time window where the protein of interest eluted, and the AUC represents the abundance of the protein of interest.

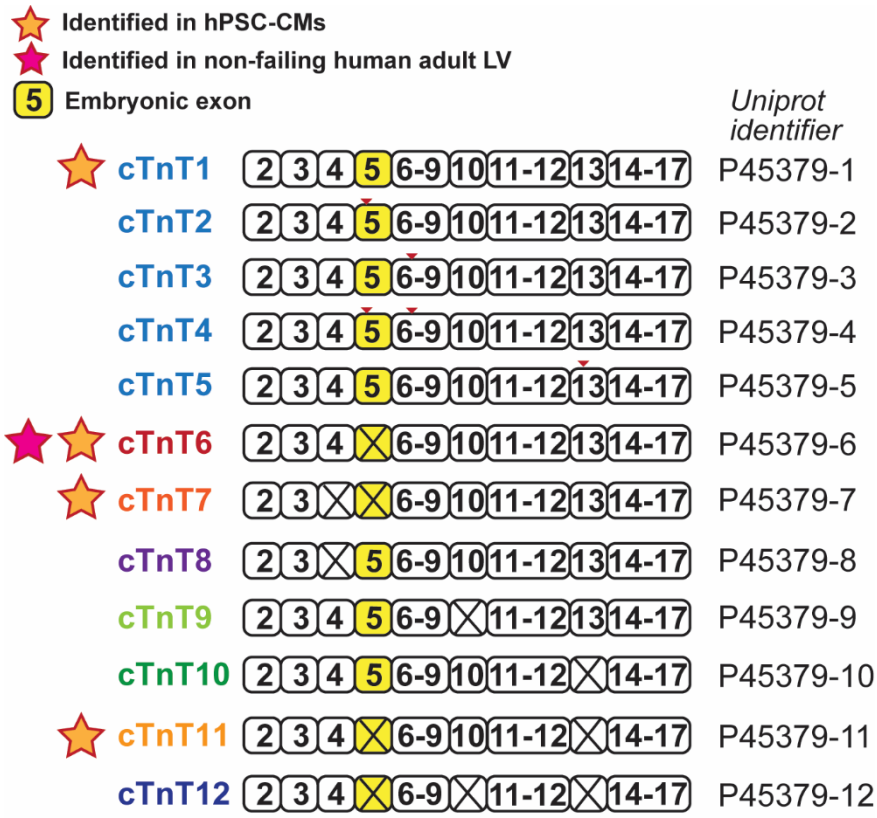


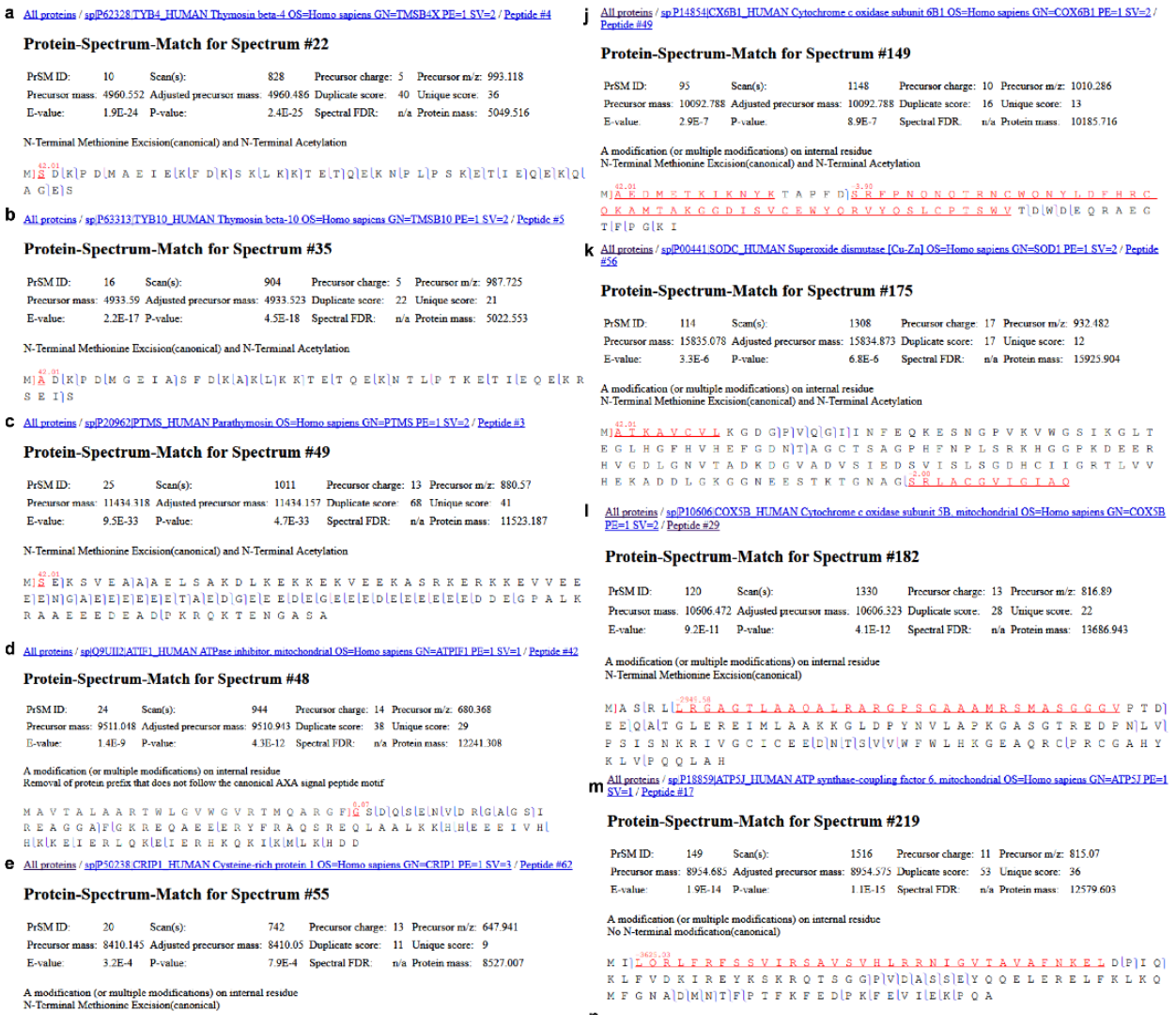
Figure 4.5. Splicing isoforms of cTnT (*TNNT2*) retrieved from Uniprot. cTnT1-5 contain all exons apart from exon 1. Minor sequence variations between cTnT1-5 are denoted by red triangles. Isoforms containing exon 5 are considered embryonic/fetal isoforms, whereas isoforms without exon 5 are considered adult forms. cTnT1, 6, 7, and 11 were identified in the hPSC-CMs in this study, in contrast, only cTnT6 was identified in non-failing human adult left ventricular (LV) tissue.

Table 4.1. **Intact proteins identified by top-down proteomics**. RT, retention time. Noted that specific retention time depends on the LC instrument, the length of the column and tubings. Specific LC-MS conditions have been included in the Methods section. M_r (Expt'l) and M_r (Calc'd), the observed most abundant mass from high-resolution MS data and the calculated most abundant mass from protein sequence (with modifications). Calculated most abundant mass was based on elemental formula simulation by Isopro3.0. aa, amino acid; N-term/C-term, N-terminal/C-terminal; † Information was derived from Uniprot protein database; ‡ Protein was identified by accurate intact mass and previous publications;[24, 232, 299] ∗ cTnI was identified only in hPSC-CMs from very late-stage (2D/3D) culture.

Peak #	RT (min)	M_r (Expt'l)	M_r (Calc'd)	Error (ppm)	Protein (*Gene*)	Sequence Modifications
1	~12.6	4962.52	4962.53	2.0	Thymosin β-4 (*TMSB4X*)	Met removal (N-term), Trimethylation (Ser1)
2	~12.7	4935.55	4935.56	2.0	Thymosin β-10 (*TMSB10*)	Met removal (N-term), Trimethylation (Ser1)
3	~12.9	11440.18	11440.17	0.9	Parathymosin (*PTMS*)	Met removal (N-term), Acetylation (Ser1)

4	~13.8	9515.88	9515.88	0.9	ATPase inhibitor, mitochondrial (*ATPIF1*)	aa1-25 removal (N-term)
5	~14.0	8415.01	8414.99	1.5	**CRIP1 (*CRIP1*)**	Met removal (N-term), Methylation (Arg67) †
6	~16.4	8085.72	8085.70	2.5	60S ribosomal protein L38 (*RPL38*)	Met removal (N-term)
7	~16.7	9223.92	9223.93	1.0	Small muscular protein (*SMPX*)	Met-Asn-Met removal (N-term), Trimethylation (Ser1)
8	~17.7	20836.88	20836.90	1.0	**MLP (*CSRP3*)** ‡	Met removal (N-term)
		20916.88	20916.87	0.5	**MLP (mono-phosphorylated)**	Met removal (N-term), Phosphorylation (site not mapped)
9	~18.5	15691.76	15675.70 +16.06	-	40S ribosomal protein S23 (*RPS23*)	Met removal (N-term), Hydroxyproline (Pro61) †
10	~18.6	22403.00	22403.00	0.0	**CRIP2(*CRIP2*)** ‡	Met removal (N-term), Acetylation (Ala1)
		22483.00	22482.97	1.3	**CRIP2 (mono-phosphorylated)**	Met removal (N-term), Acetylation (Ala1), Phosphorylation (site not mapped)
11	~19.2	10519.24	10519.22	1.9	Small ubiquitin-related modifier 2 (*SUMO2*)	Met removal (N-term), Acetylation (Ala1), Val-Tyr removal (C-term)
12	~19.8	10098.71	10102.70 -4.0	-	Cytochrome c oxidase subunit 6B1 (*COX6B1*)	Met removal (N-term), Acetylation (Ala1), (possibly 2 disulfide bonds) †
13	~20.9	34580.48	34580.41	2.0	**cTnT6 (adult isoform) (*TNNT2*)**‡	Met removal (N-term), Acetylation (Ser1), Phosphorylation (Ser1)
14	~20.9	34081.27	34081.18	2.6	**cTnT7 (adult isoform) (*TNNT2*)**‡	Met removal (N-term), Acetylation (Ser1), Phosphorylation (Ser1)
15	~20.9	34253.32	34253.25	2.3	**cTnT11 (adult isoform) (*TNNT2*)**‡	Met removal (N-term), Acetylation (Ser1), Phosphorylation (Ser1)
16	~20.9	35918.93	35918.88	1.4	**cTnT1 (fetal isoform) (*TNNT2*)** ‡	Met removal (N-term), Acetylation (Ser1), Phosphorylation (Ser1)
17	~21.0	21560.56	21560.55	0.5	**ssTnI (*TNNI1*)** ‡	Met removal (N-term)
18	~21.6	15845.92	15845.90	1.3	Superoxide dismutase [Cu-Zn] (*SOD1*)	Met removal (N-term), Acetylation (Ala1)
19	~22.1	10612.34	10612.32	1.9	Cytochrome c oxidase subunit 5B (*COX5B*)	aa1-31 removal (N-term)
20	~23.1	10841.88	10841.86	1.8	10 kDa heat shock protein, mitochondrial (*HSPE1*)	Met removal (N-term), Acetylation (Ala1)
21	~24.9	8959.56	8959.55	1.1	ATP synthase-coupling factor 6, mitochondrial (*ATP5J*)	aa1-32 removal (N-term)
22	~25.0	18412.69	18412.65	2.2	**Cofilin-1 (*CFL1*)**	Met removal (N-term), Acetylation (Ala1)
23	~25.7	13774.59	13774.57	1.5	Histone H2B type 1-C/E/F/G/I (*HIST1H2BC*)	Met removal (N-term)
24	~26.0	17880.90	17880.87	1.7	Peptidyl-prolyl cis-trans isomerase A (*PPIA*)	Met removal (N-term)
25	~26.5	23917.85	23917.83	0.8	**cTnI (*TNNI3*) *** ‡	Met removal (N-term), Acetylation (Ala1)
26	~26.6	13898.84	13898.83	0.7	Histone H2A type 2-C (*HIST2H2AC*)	Met removal (N-term), Acetylation (Ser1)
27	~28.5	32749.81	32749.74	2.1	**α-Tpm (*TPM1*)** ‡	Acetylation (Met1)
		32829.75	32829.71	1.2	**α-Tpm (mono-phosphorylated)**	Acetylation (Met1), Phosphorylation (Ser283)
28	~28.6	12344.19	12344.20	0.8	Macrophage migration inhibitory factor (*MIF*)	Met removal (N-term)
29	~30.6	21842.00	21841.96	1.8	**MLC1v (*MYL3*)** ‡	Met removal (N-term), Trimethylation (Ala1)

30	~32.0	21474.96	21474.86	3.7	**MLC1a (*MYL4*)** ‡	Met removal (N-term), Acetylation (Ala1)
31	~33.0	18699.44	18699.40	2.1	**MLC2v (*MYL2*)** ‡	Met removal (N-term), Trimethylation (Ala1)
		18779.39	18779.37	1.1	**MLC2v (mono-phosphorylated)**	Met removal (N-term), Trimethylation (Ala1), Phosphorylation
32	~33.5	16789.91	16747.83 +42.03/42.01	-	Calmodulin 1 (*CALM1*)	Met removal (N-term), Acetylation (Ala1), Acetylation/Trimethylation
33	~34.0	19358.63	19358.62	0.5	**MLC2a (*MYL7*)** ‡	Met removal (N-term), Acetylation (Ala1)
		19438.63	19438.59	2.1	**MLC2a (mono-phosphorylated)**	Met removal (N-term), Acetylation (Ala1), Phosphorylation
34	~34.8	41840.00	41839.92	1.9	**cα –actin (*ACTC1*)** ‡	Met-Cys removal (N-term), Methylation (His 73)
35	~37.0	18443.52	18443.57	2.7	**TnC (*TNNC1*)** ‡	Acetylation (Met1)



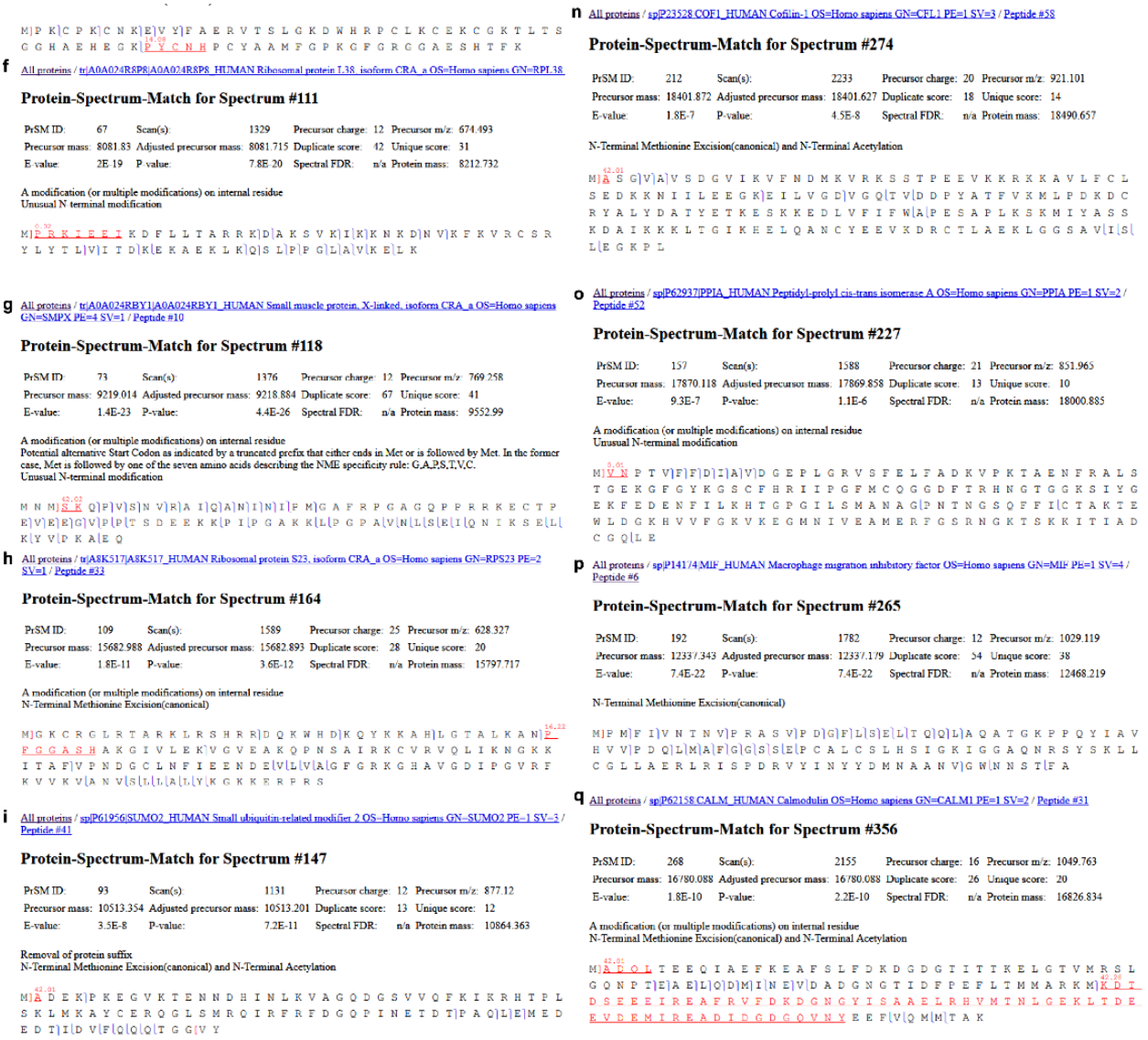


Figure 4.6. **Representative protein identification results output by MS-Align+.** Protein identification with an e-value less than 1.0E-04 was considered confident. Identification of (**a**) *TMSB4X*, (**b**) *TMSB10*, (**c**) *PTMS*, (**d**) *ATPIF1*, (**e**) *CRIP1*, (**f**) *RPL38*, (**g**) *SMPX*, (**h**) *RPS23*, (**i**) *SUMO2*, (**j**) *COX6B1*, (**k**) *SOD1*, (**l**) *COX5B*, (**m**) *ATP5J*, (**n**) *CFL1*, (**o**) *PPIA*, (**p**) *MIF*, and (**q**) *CALM1* by online LC/MS/MS. Protein identification was performed using MASH Suite Pro[229] incorporated with MS-Align+[228].

These results demonstrated reliability of the quantification method given equal amount of total proteins were injected, and quantification was performed over the linear range. To further demonstrate the reliability of this method for the quantification of sarcomeric protein isoforms as markers of hPSC-CM maturation, we tested CMs differentiated from a different pluripotent stem cell line, a human embryonic stem cell (hESC) H1 line. This method has also proven highly robust and quantitative for the assessment of sarcomeric isoform expression in hESC-CMs (**Figure 4.7**).

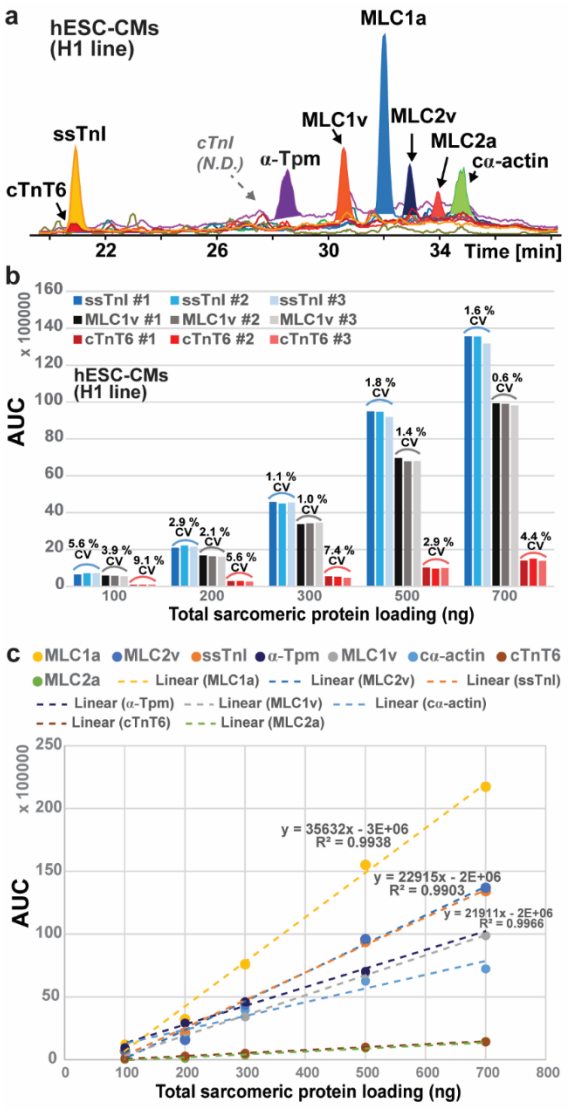


Figure 4.7. Reproducibility of protein quantification by top-down LC-MS for hESC-CMs. (a) EICs of the representative sarcomeric proteins. (b) Protein quantification from three injection replicates with varied total sarcomeric protein loading. (c) Linear curves of the major sarcomeric proteins.

4.3.2 Changes in Sarcomeric Protein Isoforms Following Prolonged Culture of hPSC-CMs Revealed by Top-down Proteomics

Following the validation of the reproducibility and quantitative properties of the top-down LC-MS-based proteomics method, we next evaluated whether this method permits reliable determination of hPSC-CM maturity. We first evaluated the maturity of the hESC-CMs (H1 line) from early- and late-stage culture by immunostaining. As expected, the expression of MLC2v increased whereas that of MLC2a decreased in hESC-CMs with prolonged culture (**Figure 4.8a**). After ensuring that the purity of hESC-CMs from the early- and late-stage cultures was greater than 90% (**Figure 4.8b**), we analyzed the hESC-CMs by top-down LC-MS to compare with the non-failing human adult left ventricular (LV) tissue (**Figure 4.8c, d**). We chose adult LV tissue as a reference because vast majority of hPSC-CMs generated with the small molecule protocol are fated to the ventricular lineage.[334]

Top-down proteomics revealed the major sarcomeric protein isoforms and their abundance represented by the AUCs (**Figure 4.8c, d**). Compared to the mature human LV tissue, wherein cTnI was expressed robustly with undetectable ssTnI and the atrial isoform of myosin light chains (MLC1a and MLC2a) (**Figure 4.8c**), hESC-CMs from both the early- and late-stage culture were considered fetal-like due to undetectable cTnI and robust expression of MLC1a and MLC2a (**Figure 4.8d**). Nevertheless, hESC-CMs at Day 60 exhibited increased maturity as measured by decreased cTnT1 (fetal cTnT isoform) (**Figure 4.8e**). cTnT6, the major adult cTnT isoform expressed in cardiac ventricular tissue[288, 339] remained unchanged. MLC2v, a common marker for evaluating the maturity of hPSC-CMs, increased significantly with extended culture, along with a

decrease in the MLC2a expression (**Figure 4.8e**). Intriguingly, the expression of both MLC1a and MLC1v decreased in the hESC-CMs with prolonged culture (**Figure 4.8e)**.

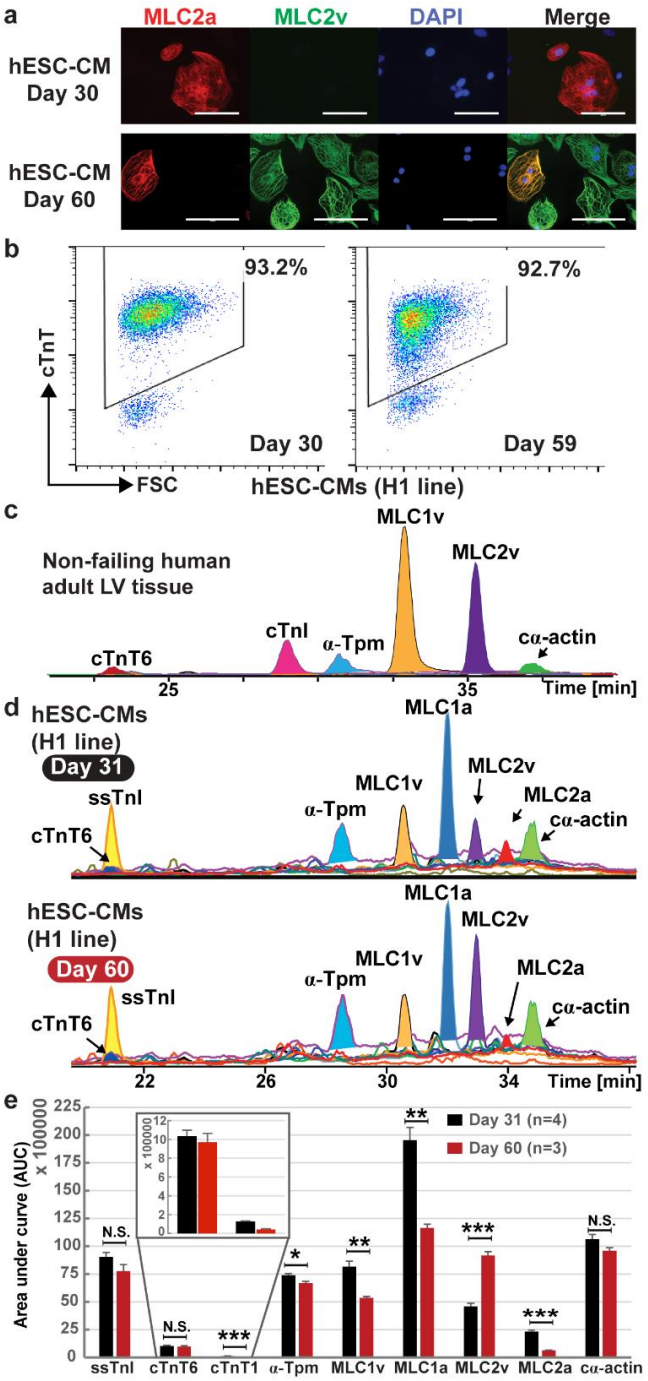


Figure 4.8. Assessment of sarcomeric protein isoform expression changes as markers of maturation in hESC-CMs with prolonged 2D monolayer culture. (a) Immunostaining showing increased MLC2v expression and decreased MLC2a expression in hESC-CMs (H1 line) at Day 60 compared to the cells at Day 30. Scale bar represents 100 µm. (b) Flow cytometric analysis of hESC-CMs at Day 30 and 59 showed

comparable CM purity between the early- and late-stage culture of hESC-CMs. (c) EICs of individual sarcomeric proteins in non-failing human adult LV tissue based on top-down LC-MS analysis (n=1). (d) EICs of individual sarcomeric proteins in hESC-CMs at Day 31 and 60 based on top-down LC-MS analysis. An equal amount (400 ng) of total protein was analyzed. N.D., not detected. e) Quantification of protein isoform expression in hESC-CMs at Day 31 and 60. ** $p < 0.01$ and *** $p < 0.001$ by student's two sample *t*-test (two-tailed). N.S. not statistically significant.

Similarly, differentiated hiPSC-CMs from the human iPSC line, DF19-9-11T, were studied after 32 and 59 days of culture. The hiPSC-CMs from early- and late-stage culture were immature relative to the non-failing human adult LV tissue. Nevertheless, hiPSC-CMs at Day 59 exhibited improved maturity compared to the cells at Day 32 (**Figure 4.9**), as demonstrated by increased MLC2v and decreased MLC2a expression in these cells. There was also a consistent decrease in the expression levels of cTnT1 (**Figure 4.9**). MLC1a remains highly abundant in both the hESC-CMs and hiPSC-CMs from late-stage culture (**Figure 4.8 and 4.9**). Over the time window analyzed, we did not observe consistent predictable patterns on the expression change of MLC1a and MLC1v in the hPSC-CMs analyzed.

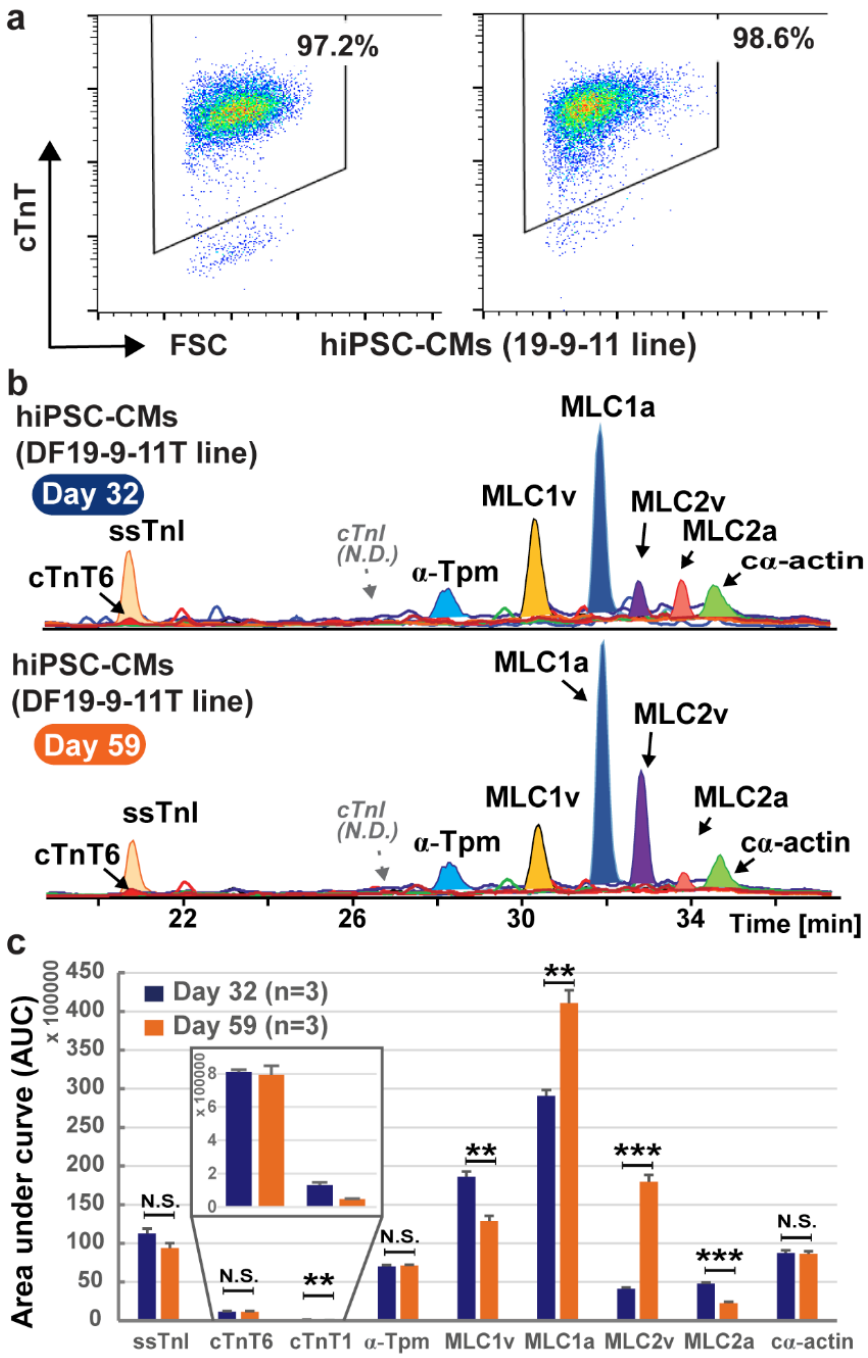


Figure 4.9. **Assessment of sarcomeric protein isoform expression changes in hiPSC-CMs with extended 2D monolayer culture.** (a) Flow cytometric analysis of hiPSC-CMs at Day 15 (left) and 59 (right) showed comparable CM purities between the early- and late-stage cultures. (b) EICs of individual sarcomeric proteins in hiPSC-CMs at Day 32 and 59. Equal amount (300 ng) of total sarcomeric protein was analyzed in three replicates. (c) Quantification of protein isoform expression in hiPSC-CMs at Day 32 and 59. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; N.S., not statistically significant.

4.3.3 Sarcomeric Protein PTMs as Potential Markers for hPSC-CM Maturation

The LC-MS-based top-down proteomics method we developed not only allows for the quantification of protein expression across multiple samples, but also permits relative quantification of protein PTMs simultaneously. From the same top-down LC-MS datasets wherein protein expression data were extracted, high-resolution mass spectra were examined for the quantification of protein PTMs. Phosphorylation of the cTnT isoforms was observed in the hESC-CMs similarly as in the non-failing human adult LV tissue (**Figure 4.10a**). ssTnI in the hESC-CMs was also found to be phosphorylated indicated by a species with a mass increase of 80 Da (**Figure 4.10b**). In addition, a potentially novel modification with an increase of 248 Da relative to the mass of ssTnI was detected (**Figure 4.10b**); however, ssTnI was not detected in the human LV tissue. In contrast, cTnI was found only in the mature LV tissue with two phosphorylated species detected (**Figure 4.10c**).

In the non-failing human adult LV tissue, mono-phosphorylated α-Tpm was low (approximately 8%) relative to the un-modified α-Tpm,[232] whereas the levels of α-Tpm phosphorylation were significantly higher in the hESC-CMs (**Figure 4.10d**). However, extended culture lead to decreased α-Tpm phosphorylation in the Day 60 hESC-CMs compared to the Day 31 cells, consistent with the Day 60 cells being relatively more mature (**Figure 4.10d**). Furthermore, previous investigation has indicated higher phosphorylation of α-Tpm in the atrial compared to ventricular tissue,[232] and confirmed a decline of α-Tpm phosphorylation during cardiac development in rat hearts.[340] Collectively, phosphorylation of α-Tpm can also serve as a validated marker of hESC-CM maturation. To the best of our knowledge, this is the first study suggesting protein PTMs can be used as markers of hPSC-CM maturation.

Though MLC2a was not detected in the non-failing human LV tissue, the phosphorylation of MLC2a appeared to decrease in the Day 60 hESC-CMs compared to Day 31 cells (**Figure 4.10e, f**). Similarly, MLC2v phosphorylation was decreased and nearly abolished in the hESC-CMs following extended culture; however, in non-failing human LV tissue, MLC2v phosphorylation was detected at a low level (**Figure 4.10f**). In addition to thin and thick filament proteins, we identified phosphorylated forms of muscle LIM protein (MLP) and cysteine-rich protein 2 (CRIP2) in the hESC-CMs, similarly as seen in the human tissue (**Figure 4.10f, g, Figure 4.11**). MLP is an important Z-disc protein specifically expressed in cardiac and skeletal muscles, and mutations in MLP are known to be associated with dilated and hypertrophic cardiomyopathies.[341] Although a phosphorylated form of MLP was detected in hESC-CMs, the level of MLP phosphorylation appeared to be lower than that in non-failing human adult LV tissue (**Figure 4.10g**). Interestingly, cysteine-rich protein 2 (CRIP2) is highly expressed in the heart during development and in the adult heart,[308, 309] and was previously identified as a heart vascular marker,[309] but this study confirmed the expression of CRIP2 in the CMs (**Figure 4.10h**). We have also, for the first time, detected phosphorylation of CRIP2 in the hESC-CMs and human heart tissue, although the level of phosphorylation was lower than that observed in adult heart tissue (**Figure 4.10h).**

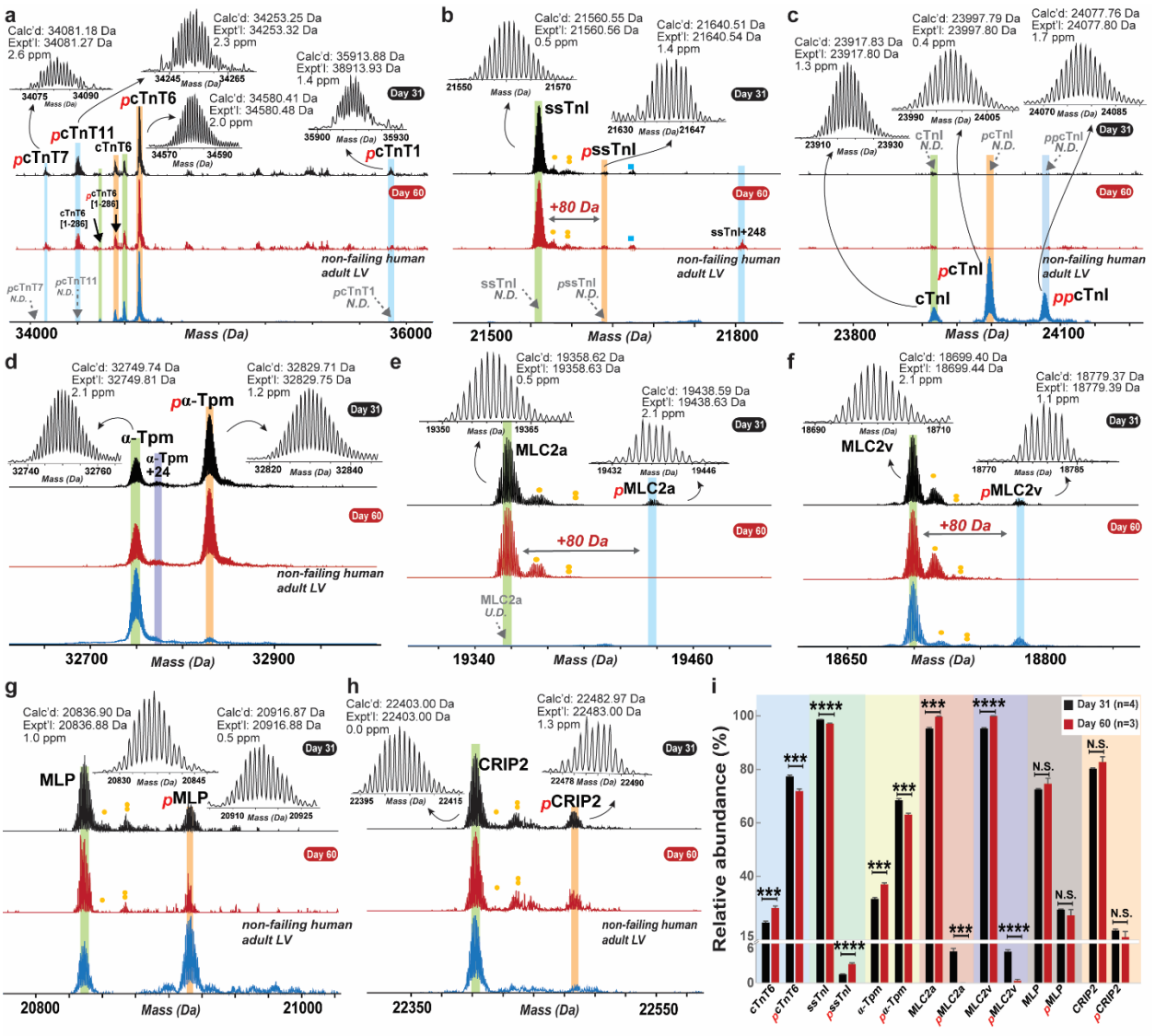


Figure 4.10. Relative quantification of the sarcomeric protein PTMs in hESC-CMs from 2D culture. (a-g) Comparison of the top-down mass spectra of cTnT, ssTnI, cTnI, muscle LIM protein (MLP), cysteine-rich protein 2 (CRIP2), α-Tpm, MLC2a and MLC2v, and their PTMs in hESC-CMs (H1 line) at Day 31 and 60, and non-failing human adult LV tissue. Yellow single and double circles represent singly and doubly oxidated species, respectively. Blue square represents non-covalent adductation from TFA. Italic red p represents phosphorylation. (i) Quantification of the relative abundance of the representative unmodified and modified sarcomeric proteins. N.D., not detected; *** $p < 0.001$ and **** $p < 0.0001$ by student's two sample *t*-test (two-tailed). N.S. not statistically significant.

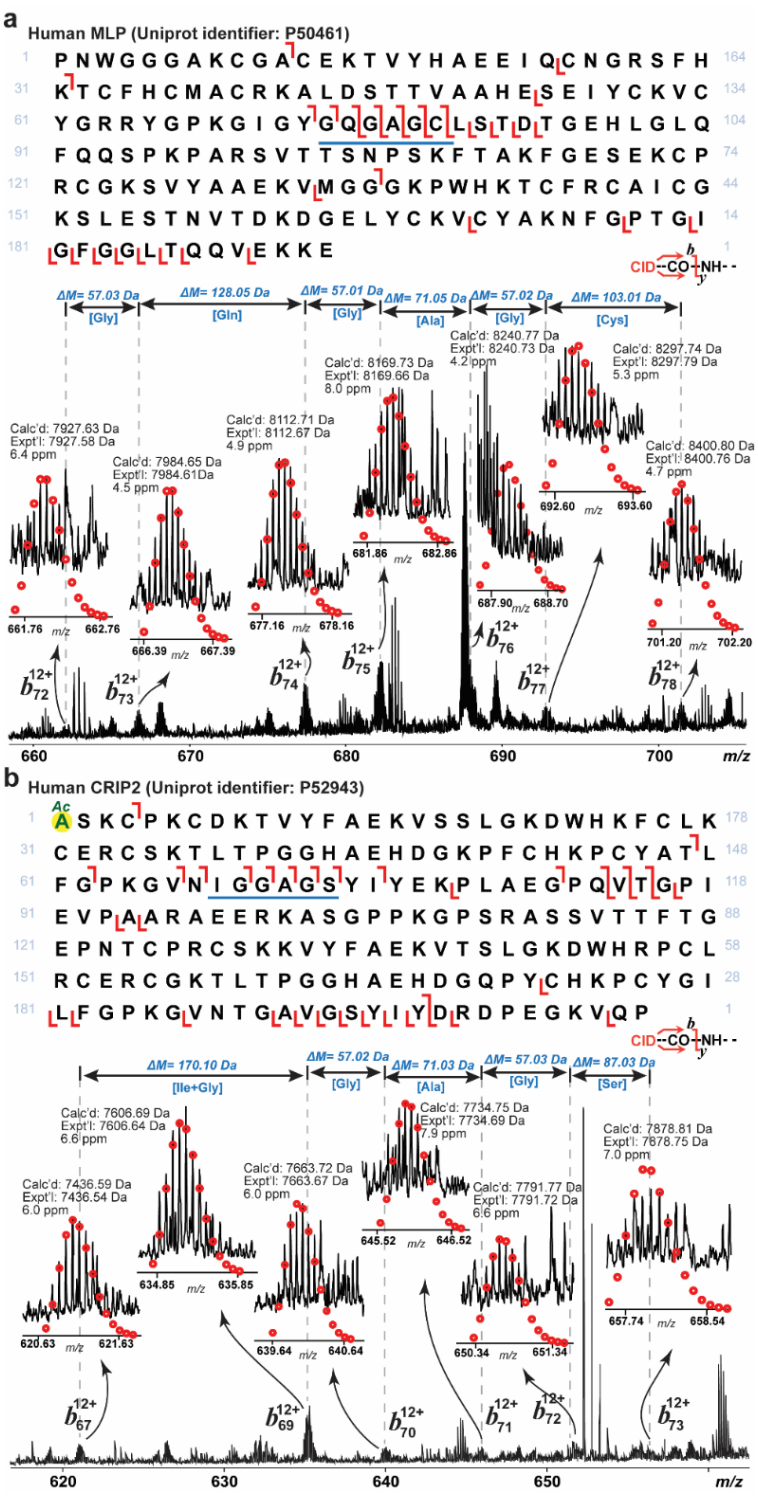


Figure 4.11. Identification of MLP (a) and CRIP2 (b) in hPSC-CMs by online top-down tandem MS (MS/MS). Top panels, protein sequence showing bond cleavages from a single MS/MS experiment with collision-induced dissociation (CID). Bottom panels, tandem mass spectra showing corresponding sequence underlined in blue in upper panels. *m/z*, mass-to-charge-ratio; Calc'd, calculated monoisotopic mass based on protein/fragment sequence; Expt'l, experimental monoisotopic mass measured by the mass spectrometer; ppm, part per million. ΔM, mass difference.

Quantification of the sarcomeric protein PTMs is shown in **Figure 4.10i**. Significant changes have been observed in the phosphorylation of ssTnI, cTnT, α-Tpm, MLC2a, and MLC2v in Day 60 hESC-CMs compared to Day 31 cells (**Figure 4.10i**). Since ssTnI and MLC2a were not detected in the non-failing adult LV tissue, the phosphorylation of ssTnI and MLC2a may not be good markers of hPSC-CM maturation at the later stage of development (beyond Day 60). On the other hand, since α-Tpm is expressed throughout hPSC-CM development and in adult hearts,[340] and a previous study has consistently demonstrated a decline of α-Tpm phosphorylation in developing rat hearts,[340] the declining phosphorylation of α-Tpm can be a novel marker of hPSC-CM maturation.

4.3.4 Global Differential Regulation of Proteins in hPSC-CM Maturation

Although the top-down LC-MS-based quantitative proteomics has proven to be a robust method for quantifying highly homologous protein isoforms concurrently with protein PTMs, detection and quantification of high molecular weight (MW) proteins and low abundance proteins from the hPSC-CMs remains challenging due to an exponential decay of signal-to-noise ratio (S/N) with increasing MW.[186, 244] Therefore, to investigate the differential regulation of hPSC-CM protein expression beyond the contractile proteins in early- and late-stage cultures, we undertook a label-free bottom-up proteomics approach for global quantitative profiling of protein expressions in the hPSC-CMs.

hiPSC-CMs from early- and late-stage culture with comparable CM purity were harvested and analyzed by an unbiased label-free quantitative proteomics approach, allowing for confident identification of 1898 protein groups at Day 30 and 59 (**Figure 4.12a**), among which 152 and 114

proteins were found to be significantly ($p < 0.01$, > 1.4-fold change) up- and down-regulated, respectively (**Figure 4.12b**). To first validate the label-free quantification method for assessing the maturation of hiPSC-CM, we surveyed a number of proteins that are known to be regulated developmentally, or are critical to contractile function and Ca^{2+} handling in relatively mature CMs. In a panel of 27 proteins, we have observed significant up-regulation of β-MHC and down-regulation of α-MHC (**Figure 4.12c**), which is consistent with a previous study that suggested a significant decrease in α-MHC expression in the ventricles of the human hearts between 7 and 12 weeks of gestation.[336] In addition, we identified a number of high MW sarcomeric proteins to be significantly up-regulated in Day 59 hiPSC-CMs compared to the Day 30 cells, including titin, α-actinin 2 (sarcomeric-specific isoform), cardiac myosin binding protein C (cMyBP-C) and myomesin 1 (MYOM1) (**Figure 4.12c**).

Subsequently, we performed protein interaction analysis and categorized the differentially regulated proteins based on their biological function. We found each category contains proteins from both up- and down-regulated protein groups, but the proteins involved in metabolic processes, cellular energy production, and oxidation-reduction processes were predominantly up-regulated (**Figure 4.12d**). This suggests that hiPSC-CMs were undergoing major metabolic changes between Day 30 to 59 of culture, consistent with the change from glycolytic predominant metabolism in the early heart to fatty acid β-oxidation in post-natal heart.[342] In particular, there was a decrease in enolase 1 (ENO1), a glycolytic enzyme known to be down-regulated during cardiac development,[343] and up-regulation of acyl-coA dehydrogenase family member 9 (ACAD9), acetyl-CoA acetyltransferase (ACAT1) and carnitine palmitoyltransferase 2 (CPT2) (**Figure 4.12d**,), which are important enzymes in β-oxidation of fatty acids. However, the changes in metabolic enzymes were complex with expression of some glycolytic enzymes increasing such as muscle

phosphofructokinase (PFKM) and at least one enzyme associated with β-oxidation of fatty acid, malonyl-CoA decarboxylase (MLYCD), decreasing in expression (**Figure 4.12d**).

A detailed analysis of the sub-interactome of proteins involved in or actin-myosin interaction revealed that early hiPSC-CMs expressed myosin light chain 6 (MYL6), myosin heavy chain 9 and 10 (MYH9, MYH10), the motor protein isoforms expressed in smooth muscle or non-muscle cells, which were down-regulated following prolonged culture (**Figure 4.12e**). We found robust expression of both α-actinin 1 and α-actinin 2 (**Figure 4.12c**), and the expression of α-actinin 1 decreased whereas the expression of α-actinin 2 increased in the hiPSC-CMs with prolonged culture (**Figure 4.12e**). α-actinin 2 is the predominant isoform expressed in the cardiac muscle, whereas α-actinin 1 is generally considered as the non-sarcomeric isoform expressed in non-muscle cells.[344, 345] These results indicated expression of α-actinin 1 in the early immature CMs, which switches to α-actinin 2 with maturation. The interactome analysis revealed a wide variety of proteins interacting with the α-actinin isoforms (**Figure 4.12e**), suggesting a key role for α-actinin in CM development and, possibly, maturation. Among the up-regulated proteins that interact with the sarcomeric proteins, many are involved in cell adhesion and cell junction assembly including dystrophin (**Figure 4.12e**), an important component of the sarcolemma membrane complex.[346] In addition, proteins involved in cardiac excitability including α1-syntrophin were found to be up-regulated (**Figure 4.12e**). α1-syntrophin has been shown to regulate CM membrane action potential via regulation of the cardiac sodium channels,[347, 348] and may contribute to the increase in voltage-gated sodium current observed with maturation. The maturation of hPSC-CMs requires orchestrated protein synthesis and folding, and consistently, we observed up-regulation of a number of molecular chaperons/co-chaperons including αB-crystallin (**Figure 4.12b, d**), a small heat shock protein recently shown to play key roles in the regulation of

cardiac hypertrophy[349] and maintenance of mitochondrial homeostasis in CMs.[350] Moreover, we found that phospholamban (PLB) was identified in all the Day 59 hiPSC-CM samples, and not identified in any Day 30 cell samples. PLB localizes to the sarcoplasmic reticulum, and plays an important role in Ca^{2+} homeostasis by regulating the cardiac sarcoplasmic reticulum Ca^{2+}-ATPase (SERCA).[351]

For the proteins identified to be up- or down-regulated in the hiPSC-CMs following prolonged culture, we further examined their expression in the developing mouse hearts. The mouse hearts were harvested at embryonic day 15 (E 15), post-natal day (PND) 1, 10, 21, and at adulthood (3-month old) (**Figure 4.12f, g**). The ventricles were separated from the atria, and the ventricular tissues were used for analysis because the vast majority of hPSC-CMs were of the ventricular lineage. Immunoblotting showed increased expression of dystrophin, α-actinin 2, cTnI, αB-crystallin and phospholamban, with a progressive decrease in the expression of α-actinin 1, in the developing mouse ventricles (**Figure 4.12f, g**), which were consistent with the observation in hPSC-CM maturation. These results demonstrated the great promise of these proteins as new markers of hPSC-CMs at various stages of development, and demonstrated the power of label-free quantitative proteomics for identifying novel markers of hPSC-CM maturation.

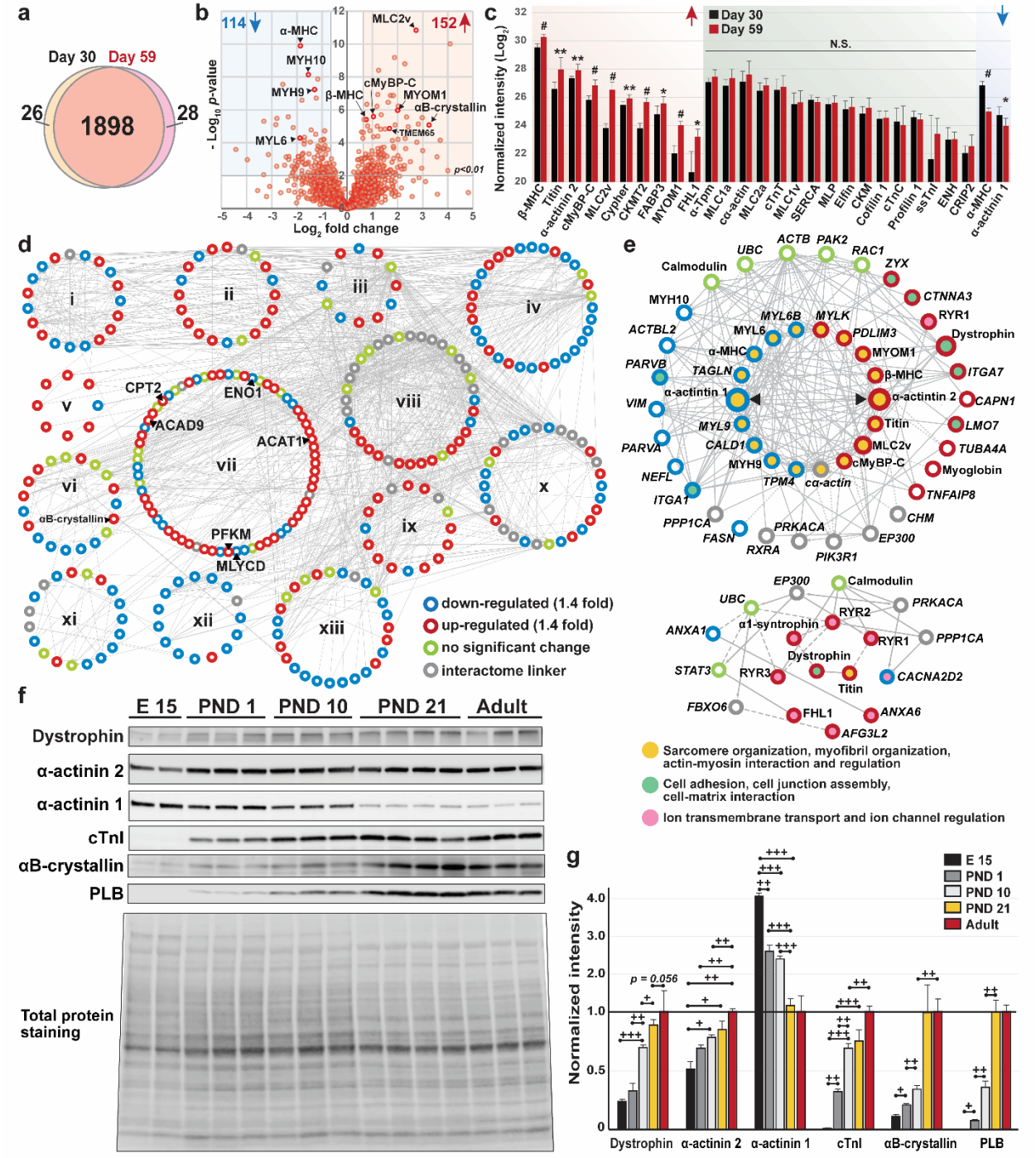


Figure 4.12. Large-scale bottom-up label-free quantification revealed global differential regulation of proteins in hiPSC-CMs with prolonged culutre. (a) Venn diagram showing the numbers of proteins identified in hiPSC-CMs (DF19-9-11T line) at Day 30 and Day 59. (b) Volcano plot showing differential regulation of proteins in Day 59 compared to Day 30 hiPSC-CMs. (c) Label-free quantification of the representative contractile proteins and proteins involved in Ca^{2+} handling in hiPSC-CMs. CKMT2, sarcomeric mitochondrial creatine kinase; FABP3, fatty acid binding protein found in the heart; MYOM1, myomesin 1; FHL1, four and half LIM domain protein; SERCA, sarcoplasmic reticulum Ca^{2+}-ATPase; CKM, creatine kinase M-type; ENH, enigma homolog. N.S, not statistically significant, * p<0.01, ** p< 0.001, *** p< 0.0001, # p< 0.00001 by the Welch's modified t-test. (d) Functional categorization and

interaction analyses of differentially regulated proteins in hiPSC-CMs in Day 59 compared to Day 30 cells. Protein grouping was based on (i), sarcomere organization, myofibril organization, or actin-myosin interaction and regulation; (ii), cell adhesion, cell junction assembly, or cell-matrix interaction; (iii), protein ubiquitination/deubiquination, proteasome-mediated protein catabolic process, or proteasome assembly; (iv), cytoskeleton organization, or cellular component movement and organization; (v), ion transmembrane transport and ion channel regulation; (vi), protein folding, unfolded protein response, or chaperon/co-chaperone-mediated protein assembly; (vii), metabolic processes, cellular energy production, or oxidation-reduction processes; (viii), response to stimuli and signal transduction; (ix), DNA damage response, DNA replication, cell cycle regulation, or cell division; (x), regulation of transcription and gene expression; (xi), ribosome assembly and protein translation, rRNA processing, or tRNA processing; (xii), mRNA processing, transport or metabolism; (xiii), protein transport and localization, vesicle-mediated transport, or endocytosis. (e) Extended interactome analyses of proteins involved in sarcomere organization, myofibril organization, or actin-myosin interation and regulation, and proteins involved in ion transmembrane transport and ion channel regulation. Italic font denotes gene names. RYR, ryanodine receptor. (f) Immunoblotting showing protein expression change in mouse ventricles during development. PLB, phospholamban. E 15, embryonic day 15; PND 1/10/21, post-natal day 1/10/21; adult, 3-month old. (g) Quantification of protein expression during mouse ventricular development. Biological replicates for E 15, PND 1, PND 10, PND 21 and adult animals were 2, 3, 3, 4, and 3, respectively. + $p< 0.05$, ++ $p< 0.01$, +++ $p< 0.001$ by the student's two-sample t-test (two-tailed).

4.3.5 Top-down LC-MS Analysis of CMs Cultured in 3D ECT

Recent development of tissue engineering technology has enabled the construction of ECTs, wherein CMs are cultured in a 3D environment with extracellular matrix to support the contacts between CMs and improve CM function *in vitro*.[317, 324, 352, 353] Previous studies have implicated improved hPSC-CM maturation with ECT culture compared to conventional 2D monolayer culture.[352] However, a detailed analysis of contractile protein isoform expression and their PTMs in hPSC-CMs cultured in 3D ECT has not been performed. Therefore, we employed the top-down LC-MS method to quantify the expression and PTMs of sarcomeric proteins in the ECT CMs as compared to the cells from 2D monolayer.

CMs were differentiated from DF19-9-11T hiPSCs in monolayer for 30 days, after which cells were dissociated and seeded into a fibrin-based matrix to form the 3D ECT constructs. ECTs

were cultured for an additional 14-44 days (**Figure 4.13a**). The hematoxylin and eosin stain showed a uniform distribution of hiPSC-CMs throughout the matrix (**Figure 4.13a**). Immunostaining of the ECT showed localization of α-actinin and cMyBP-C to the highly organized sarcomeres, as well as longitudinally orientated hiPSC-CMs within the ECT (**Figure 4.13b, c**). hiPSC-CMs grown in ECT revealed intact Frank-Starling relationship (**Figure 4.13d**). Furthermore, concurrent measurement of twitch force and Ca^{2+} transient suggested intact excitation-contraction coupling (**Figure 4.13e**). Sarcomeric proteins were reproducibly extracted and detected by top-down MS from 3-week ECTs (**Figure 4.14**).

Next, we performed parallel analysis on hiPSC-CMs from the same batch of differentiation cultured in either 2D or 3D cultures for the same length of culture (**Figure 4.15**). Due to interference from the matrix proteins in ECTs, quantification of sarcomeric protein isoforms cannot be performed by directly comparing the AUCs of the same protein across different samples and, therefore, the relative expression of the adult to fetal isoforms were compared when the hiPSC-CMs from 2D and 3D cultures were analyzed. At Day 76 post-differentiation, cTnI, the adult isoform, can be detected in the hiPSC-CMs with low abundance, whereas ssTnI, the fetal isoform, remains robustly expressed (**Figure 4.13f**). The expression of cTnI relative to ssTnI was higher in the hiPSC-CMs from the 3D culture compared to the cells maintained in the 2D culture (**Figure 4.13f**). The increased expression of cTnI relative to ssTnI in the 3D hiPSC-CMs compared to the 2D cells at the late-stage culture provided molecular evidence that 3D ECT can promote hiPSC-CM maturation.[317] We also observed a reduction in the phosphorylation of α-Tpm in the 3D-cultured cells as compared to those cultured in monolayer (**Figure 4.14**), which is consistent with improved maturation with prolonged culture. This further supports that 3D culturing may

promote maturation of hiPSC-CMs, and that α-Tpm phosphorylation might serve as a novel marker for hPSC-CM maturation.

For the myosin light chain isoforms, prolonged culture rather than the culture matrix (2D vs. 3D) had a greater impact on the expression of MLC1v relative to MLC1a (**Figure 4.13g**) and the expression of MLC2v relative to MLC2a (**Figure 4.13h**). In particular, the expression of MLC1v relative to MLC1a decreased in the 2D CMs following prolonged culture (**Figure 4.13g**), which is consistent with **Figure 4.8e**. This could be attributed to the cells still being in the early-stage of development wherein MLC1a expression was still increasing before it reached maximum and started to decrease. The expression of MLC2v relative to MLC2a was not significantly altered between the cells from the 3D and 2D culture (**Figure 4.13h**). Since the expression of MLC2v relative to MLC2a was already high (**Figure 4.13h**), minor changes due to the altering of the culturing method may not have a significantly detectable impact on the expression of MLC2v relative to MLC2a.

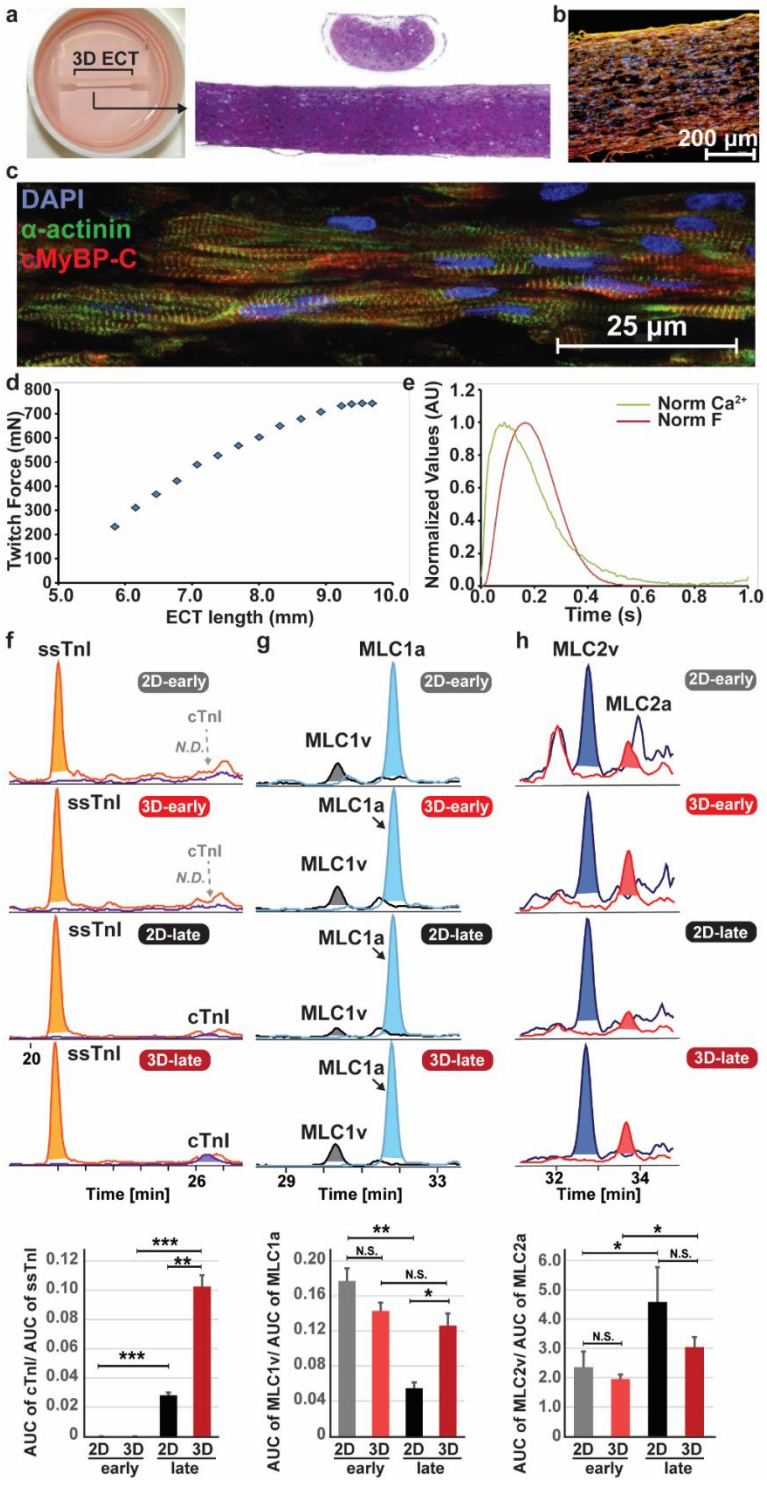


Figure 4.13. Quantification of sarcomeric protein isoforms in hiPSC-CMs from 2D monolayer and 3D ECT culture. (a) ECT cultured in 6-well plate and the hamatoxylin-and-eosin-stained cross section and longitudinal section showing the hiPSC-CMs (DF19-9-11T line) within the ECT construct. (b) Immunostaining showing the longitudinal orientation of the hiPSC-CMs and the sarcomere organization. (c) Human ECT demonstrated intact Frank-Starling stretch response. (d) Representative Ca^{2+} transient and twitch force tracings depicting normalized Ca^{2+} transient (green trace) and twitch force (red trace) over time

(s) from human ECT electrically stimulated at a frequency of 1 Hz. (e) Top-down proteomics for the analysis of 3D ECT with representative EICs of the major sarcomeric protein isoforms. EICs of (f) ssTnI and cTnI, (g) MLC1v and MLC1a, and (h) MLC2v and MLC2a from hiPSC-CMs cultured in 2D monolayer or 3D ECT at Day 44 (early) or Day 76 (late) post-differentiation. Noted that ECT cultures started at Day 30 post-differentiation. N.D., not detected; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ by student's two sample *t*-test (two-tailed). N.S. not statistically significant.

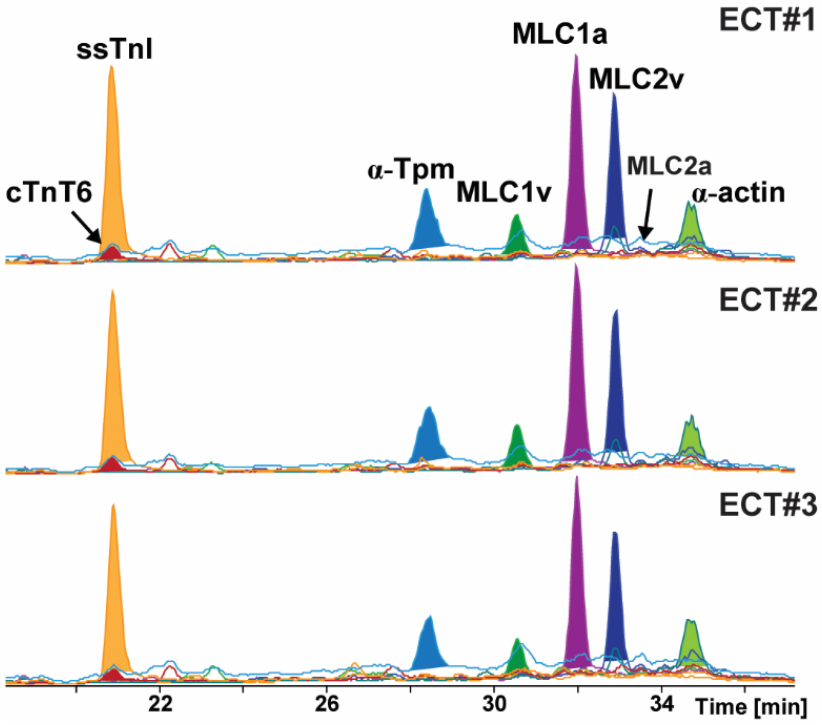


Figure 4.14. Reproducibility of the sample preparation and top-down LC-MS method for the analysis of 3D ECTs. Three extraction replicates from 2-week ECTs were performed and analyzed by top-down LC-MS. Extracted ion chromatograms (EICs) of the representative sarcomeric proteins were shown. The abundance of the protein is represented by the area under curve (AUC) of the corresponding EIC. The three extraction replicates show a high degree of agreement in the presence of the sarcomeric proteins. The elution time and relative abundance of the representative proteins were highly consistent across three replicates, demonstrating the reproducibility of the LC-MS method.

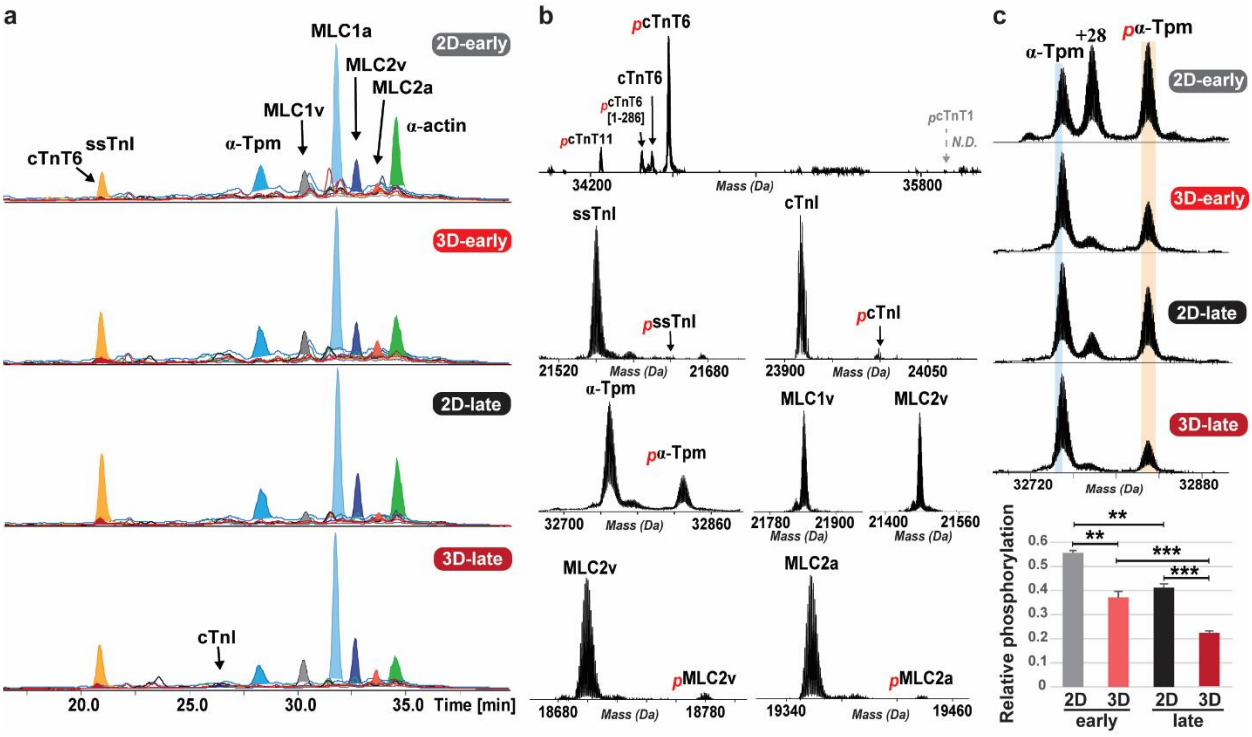


Figure 4.15. **Top-down MS analysis of hiPSC-CMs from 2D monolayer and 3D ECT culture.** (a) EICs of the sarcomeric proteins from in hiPSC-CMs from both 2D and 3D culture. The abundance of the protein is represented by the AUC of the corresponding EIC. (b) Mass spectra of cTnT6, ssTnI, cTnI, α-Tpm, MLC1v, MLC1a, MLC2v, and MLC2a from hiPSC-CMs cultured in 3D ECT at late stage (76 days post-differentiation). (c) Phosphorylation of α-Tpm decreased with prolonged culture, and also decreased in the hiPSC-CMs from 3D culture compared to those from 2D culture, indicating that 3D culture may improve hiPSC-CM maturation.

4.4 DISCUSSION

Despite substantial interest in hPSC-CM in basic and translational research, the immaturity of hPSC-CMs poses significant challenges, particularly for disease modeling, cell therapy development, and drug screening. Promoting hPSC-CM maturation requires accurate benchmarking at the molecular level. To facilitate benchmarking of the hPSC-CM maturity, we developed an unbiased proteomics strategy combining the strengths of both bottom-up and top-down proteomics to assess hPSC-CM maturity. With the bottom-up proteomics method providing

global identification and quantification of proteins in-breadth, and the top-down proteomics method offering in-depth analysis of protein isoform expression and their PTMs, we have demonstrated the power of an integrated proteomics strategy for the comprehensive characterization of hPSC-CM maturation.

For the first time, we developed a quantitative top-down proteomics method that can provide concurrent quantification of protein isoform expression and PTMs with high reproducibility and throughput (**Figure 4.2, 4.8, 4.10**). The entire process of top-down analysis, including sample preparation, takes less than 2 hrs. Compared to the traditional immunoblotting technique, with a single 40 min LC-MS run, we were able to detect and quantify the expression level of the major sarcomeric proteins including cTnT, cTnI/ssTnI, α-Tpm, MLC1v/MLC1a, MLC2v/MLC2a, cα-actin (**Figure 4.8, Table 4.1**) and their PTMs, most notably phosphorylation (**Figure 4.10**). Supported by the quantitative top-down proteomics method, we can perform a direct assessment of the hPSC-CM maturity compared to that of the adult CMs from non-failing human LV tissue (**Figure 4.8, 4.10**). Overall, hPSC-CMs from both early- and late-stage 2D and 3D culture were considered fetal-like due to robust expression of ssTnI and extremely low expression of cTnI. It is well recognized that ssTnI is expressed in the hearts at the fetal and neonatal stages, but cTnI becomes the only isoform expressed in adult CMs.[264, 337] Importantly, the expression of cTnI does not revert to the fetal isoform in pathological or hypertrophic conditions, making cTnI an ideal marker for assessing the maturity of hPSC-CMs. The molecular mechanisms governing the ssTnI-to-cTnI transition are largely unknown, but ssTnI-to-cTnI switch appears to be substantially later in hPSC-CM maturation. On the contrary, cTnT was expressed predominantly as the adult isoform at early-culture stage, with very minimal fetal isoform expression.

The expression of the atrial contractile isoforms in the ventricles is a recognized pattern of the fetal hearts. Consistently, the MLC1a and MLC2a were robustly expressed in the hPSC-CMs, which were predominantly of the ventricular lineage. Following prolonged culture, MLC2v was found to be up-regulated while MLC2a down-regulated. However, the expression of MLC1a persisted, and we did not observe consistent changes in the expression level of MLC1a in different hPSC lines. Whereas the MLC2a-to-MLC2v switch occurred early in culture, MLC1a-to-MLC1v switch was significantly later. The role of MLC1 isoforms for assessing hPSC-CM maturity is debatable; however, similar to cTnI, MLC2v was proposed to be a robust marker for maturation of ventricular hPSC-CMs due to their resistance to revert to the fetal gene expression under stress and disease conditions.[316] Our study suggests that MLC2 represents an relatively early maturation marker that predominantly expresses as the adult isoform by 60 days of culture, while cTnI expression is significantly later, representing a later marker of hPSC-CM maturation. These results demonstrated that the timing at which different contractile proteins switch from fetal to adult form is different. Whereas some markers may appear to be adult-like (e.g. cTnT, MLC2v), some other markers remain expressed as their fetal isoforms (e.g. ssTnI). Therefore, a single marker may provide a misleading assessment of hPSC-CM maturity. Our study highlights the importance of utilizing a panel of markers to accurately determine the maturation status the hPSC-CMs.

PTMs mediate protein activity and function, and are increasingly recognized as important regulators in numerous cellular processes.[149] Hence, protein PTMs may serve as important markers of hPSC-CM maturation. Notably, for the first time, we have identified phosphorylation of α-Tpm as a candidate maturation marker. There is a significant decrease of α-Tpm phosphorylation in hPSC-CMs following prolonged culture, and α-Tpm phosphorylation was found to be extremely low (less than 8%) in adult CMs (**Figure 4.10d, i**). α-Tpm is a regulatory protein of the thin

filament that blocks the myosin binding site on the actin filament, and upon increase of intracellular Ca^{2+} , α-Tpm positioning changes to allow for actin-myosin interaction.[232] α-Tpm is expressed throughout hPSC-CM differentiation and maturation and in the adult hearts, and a previous study has validated the declined phosphorylation of α-Tpm in developing rat hearts.[340] Hence, phosphorylation of α-Tpm can also serve as a validated marker of hPSC-CM maturation.

Nevertheless, analysis of low abundance high MW proteins larger than 50 kDa remains a significant challenge in the field of top-down proteomics.[186, 244, 278] Despite recent improvement in multi-dimensional separation methods for top-down proteomics,[180, 278] high sample consumption prohibits the practical use of these methods in hPSC-CMs, and the proteome coverage has yet matched that of the bottom-up proteomics approach. To address this, we employed a bottom-up global proteomics approach to complement the top-down analysis. The bottom-up proteomics approach outperformed the top-down method in the breadth of the proteome coverage, and permits the analysis of the high MW proteins, including the MHC isoforms, titin, and cMyBP-C (**Figure 4.13**), as well as low abundance proteins beyond the contractile apparatus, such as SERCA and RYRs (**Figure 4.13**). Empowered by the global proteomics method, we identified the expression changes of large proteins in hPSC-CMs following prolonged culture, such as up-regulation of β-MHC, cMyBP-C and titin, and down-regulation of α-MHC (**Figure 4.13b**). Previous studies suggest that human ventricular CMs switch from expressing the fast α-MHC isoform to the slow β-MHC during cardiac development[336] with approximately 90% β-MHC isoforms expressed in the non-failing human ventricles.[128] However, the expression of α- and β-MHC can be affected by pathological stress and heart failure, thus, compromising its reliability as a maturation marker in the setting of disease.[316]

Apart from proteins in the contractile apparatus, proteins involved in metabolic processes were predominantly up-regulated (**Figure 4.13d**). In particular, proteins responsible for fatty acid β-oxidation (e.g. ACAD9, ACAT1 and CPT2) were found to be up-regulated, with down-regulation of the glycolytic enzyme ENO1. This is consistent with a switch from glycolysis to fatty acid β-oxidation as major energy source during cardiac development. However, the overall changes of the metabolic proteins were complex with some glycolytic enzymes such as PFKM being up-regulated following prolonged culture.[354, 355] Nevertheless, the integrated changes in metabolism are not simply reflected by down-regulation of glycolytic enzymes and up-regulation of enzymes for fatty acid oxidation. For example, one key transcriptional activator for the metabolic switch in the developing heart is the PGC-1α and PPARβ/δ complex, which both increase the expression of key enzymes in fatty acid oxidation,[354], but at the same time increasing expression of some genes related to myocardial glucose metabolism such as PFKM.[355] In addition to proteins involved in cellular metabolism, we found up-regulation of α-actinin 2, dystrophin, αB-crystallin, and PLB in hPSC-CMs following prolonged culture. These proteins are important players in the assembly and maintenance of the contractile apparatus and sarcolemma, maintenance of mitochondrial integrity, and/or Ca^{2+} homeostasis, which are all critical for the survival and function of the mature CMs. Furthermore, we confirmed up-regulation of these proteins in the developing mouse ventricles, highlighting the reliability of the proteomics method for uncovering key players involved in cardiac maturation.

In summary, we have developed an integrated proteomics method integrating high-resolution top-down MS for in-depth quantification of contractile protein isoforms and PTMs, and high-throughput label-free bottom-up proteomics for in-breadth profiling of global protein expression for the assessment of hPSC-CM maturation. This method enables accurate

quantification of a number of verified markers of hPSC-CM maturation, which are constituents of the contractile apparatus. Furthermore, this integrated platform has identified new potential markers of hPSC-CM maturation, including the contractile protein PTMs and other proteins that were not previously known to be regulated developmentally. This method is unbiased, robust, and versatile, providing a direct and comprehensive assessment of hPSC-CM maturation with high efficiency and throughput. The application of this method to the analysis of hPSC-CMs at early- and late-stage 2D monolayer culture and 3D ECT further illustrated that prolonged culture and 3D tissue engineering promote the hPSC-CM maturation at the molecular level. Furthermore, this integrated proteomics strategy has laid a strong foundation for uncovering transcriptional and translational basis underlying human cardiac development, and understanding the complex disease-associated expression changes that can be modeled with patient-specific hiPSC-CMs. We envision this unbiased and direct method can be utilized for the accurate and specific tracking of the hPSC-CM maturity and facilitate the development of the optimal strategy for maturing hPSC-CMs *in vitro* to realize its full potential in regenerative therapy, drug discovery and disease modeling.

CHAPTER 5

MASH SUITE PRO: A COMPREHENSIVE SOFTWARE TOOL FOR TOP-DOWN PROTEOMICS

This chapter is reprinted from *Molecular and Cellular Proteomics*, Cai W, Guner H, Gregorich ZR, Chen AJ, Ayaz-Guner S, Peng Y, Valeja SG, Liu X, Ge Y. MASH Suite Pro: a comprehensive software tool for top-down proteomics. 2016; 15(2): 703-714,

ABSTRACT

Top-down mass spectrometry (MS)-based proteomics is arguably a disruptive technology for the comprehensive analysis of all proteoforms arising from genetic variation, alternative splicing, and post-translational modifications (PTMs). However, the complexity of top-down high-resolution mass spectra presents a significant challenge for data analysis. In contrast to the well-developed software packages available for data analysis in bottom-up proteomics, the data analysis tools in top-down proteomics remain under-developed. Moreover, despite recent efforts to develop algorithms and tools for the deconvolution of top-down high-resolution mass spectra and the identification of proteins from complex mixtures, a multi-functional software platform, which allows for the identification, quantitation, and characterization of proteoforms with visual validation, is still lacking. Herein, we have developed MASH Suite Pro, a comprehensive software tool for top-down proteomics with multifaceted functionality. MASH Suite Pro is capable of processing high-resolution MS and tandem MS (MS/MS) data using two deconvolution algorithms to optimize protein identification results. In addition, MASH Suite Pro allows for the characterization of PTMs and sequence variations, as well as the relative quantitation of multiple proteoforms in different experimental conditions. The program also provides visualization components for validation and correction of the computational outputs. Furthermore, MASH Suite Pro facilitates data reporting and presentation via direct output of the graphics. Thus, MASH Suite Pro significantly simplifies and speeds up the interpretation of high-resolution top-down proteomics data by integrating tools for protein identification, quantitation, characterization, and visual validation into a customizable and user-friendly interface. We envision that MASH Suite Pro will play an integral role in advancing the burgeoning field of top-down proteomics.

5.1 INTRODUCTION

With well-developed algorithms and computational tools for mass spectrometry (MS) data analysis, peptide-based bottom-up proteomics has gained considerable popularity in the field of systems biology.[356-364] Nevertheless, the bottom-up approach is suboptimal for the analysis of protein post-translational modifications (PTMs) and sequence variants as a result of protein digestion.[154] Alternatively, the protein-based top-down proteomics approach analyzes intact proteins, which provides a "bird's eye" view of all proteoforms,[146] including those arising from sequence variations, alternative splicing, and diverse post-translational modifications, making it a disruptive technology for the comprehensive analysis of proteoforms.[24, 147, 155, 172, 182, 202, 207, 215, 219, 253, 365-367] However, the complexity of top-down high-resolution mass spectra presents a significant challenge for data analysis. In contrast to the well-developed software packages available for processing data from bottom-up proteomics experiments, the data analysis tools in top-down proteomics remain under-developed.

The initial step in the analysis of top-down proteomics data is deconvolution of high-resolution mass and tandem mass spectra. Thorough High-Resolution Analysis of Spectra by Horn (THRASH), which was the first algorithm developed for the deconvolution of high-resolution mass spectra,[220] is still widely used. THRASH automatically detects and evaluates individual isotopomer envelopes by comparing the experimental isotopomer envelope with a theoretical envelope, and reporting those that score higher than a user-defined threshold. Another commonly used algorithm, MS-Deconv, utilizes a combinatorial approach to address the difficulty of grouping MS peaks from overlapping isotopomer envelopes.[222] Recently, UniDec, which employs a Bayesian approach to separate mass and charge dimensions,[223] can also be applied to the deconvolution of high-resolution spectra. Although these algorithms assist in data processing,

unfortunately, the deconvolution results often contain a considerable amount of mis-assigned peaks as a consequence of the complexity of the high-resolution MS and MS/MS data generated in top-down proteomics experiments. Errors such as these can undermine the accuracy of protein identification and PTM localization and, thus, necessitates the implementation of visual components that allow for the validation and manual correction of the computational outputs.

Following spectral deconvolution, a typical top-down proteomics workflow incorporates identification, quantitation, and characterization of proteoforms; however, most of the recently developed data analysis tools for top-down proteomics, including ProSightPC [225, 368], Mascot Top Down (also known as Big-Mascot),[226] MS-TopDown,[227] and MS-Align+,[228] focus almost exclusively on protein identification. ProSightPC was the first software tool specifically developed for top-down protein identification. This software utilizes "shotgun annotated" databases[369] that include all possible proteoforms containing user-defined modifications. Consequently, ProSightPC is not optimized for identifying PTMs that are not defined by the user(s). Additionally, the inclusion of all possible modified forms within the database dramatically increases the size of the database and, thus, limits the search speed.[228] Mascot Top Down[226] is based on standard Mascot, but enables database searching using a higher mass limit for the precursor ions (up to 110 kDa), which allows for the identification of intact proteins. Protein identification using Mascot Top Down is fundamentally similar to that used in bottom-up proteomics[370] and, therefore, it is somewhat limited in terms of identifying unexpected PTMs. MS-TopDown[227] employs the spectral alignment algorithm,[371] which matches the top-down tandem mass spectra to proteins in the database without prior knowledge of the PTMs. Nevertheless, MS-TopDown lacks statistical evaluation of the search results, and performs slowly when searching against large databases. MS-Align+ also utilizes spectral alignment for top-down protein identification.[228] It is capable of

identifying unexpected PTMs and allows for efficient filtering of candidate proteins when the top-down spectra are searched against a large protein database. MS-Align+ also provides statistical evaluation for the selection of proteoform spectrum match (PrSM) with high confidence. More recently, Top-down Mass Spectrometry based Proteoform Identification and Characterization (TopPIC) was developed (http://proteomics.informatics.iupui.edu/software/toppic/index.html). TopPIC is an updated version of MS-Align+ with increased spectral alignment speed and reduced computing requirements. In addition, MSPathFinder, developed by Kim et al., also allows for the rapid identification of proteins from top-down tandem mass spectra (http://omics.pnl.gov/software/mspathfinder) using spectral alignment. Although software tools employing spectral alignment, MS-Align+ and MSPathFinder, are particularly useful for top-down protein identification, these programs operate using command line, making them difficult to use for those with limited knowledge of command syntax.

Recently, new software tools have been developed for proteoform characterization.[255, 372] Our group previously developed MASH Suite, a user-friendly interface for the processing, visualization, and validation of high-resolution MS and MS/MS data.[255] Another software tool, ProSight Lite, developed recently by the Kelleher group,[372] also allows characterization of protein PTMs. However, both of these software tools require prior knowledge of the protein sequence for the effective localization of PTMs. In addition, both software tools cannot process data from liquid chromatography (LC)-MS and LC-tandem MS (MS/MS) experiments, which limits their usefulness in large-scale top-down proteomics. Thus, despite these recent efforts, a multi-functional software platform enabling identification, quantitation, and characterization of proteins from top-down spectra, as well as visual validation and data correction, is still lacking.

Herein, we report the development of MASH Suite Pro, an integrated software platform, designed to incorporate tools for protein identification, quantitation, and characterization into a single comprehensive package for the analysis of top-down proteomics data. This program contains a user-friendly customizable interface similar to the previously developed MASH Suite,[255] but also has a number of new capabilities, including the ability to handle complex proteomics datasets from LC-MS and LC-MS/MS experiments; as well as the ability to identify unknown proteins and PTMs using MS-Align+.[228] Importantly, MASH Suite Pro also provides visualization components for the validation and correction of the computational outputs, which ensures accurate and reliable deconvolution of the spectra and localization of PTMs and sequence variations.

5.2 MATERIALS AND METHODS

5.2.1 Software Design and Implementation

The default algorithm for spectral deconvolution in MASH Suite Pro is a modified version of THRASH [220] that we developed in house based on the Decon2LS open source code.[224] MS-Align+[228] has been integrated into the program and is used for top-down protein identification. The program is written on the Microsoft .NET framework. The main scientific algorithms are written in C++, and the visual development was written with C#. The windows were developed using Qios Devsuite, and the graphs and spectrum charts were developed using Microsoft Chart Controls. The graphical user interface of MASH Suite Pro was designed using tabbed document interface. The software architecture pattern was based on model view controller and the data tables were constructed based on Listview and Datagrid structures.[255]

5.2.2 Extraction of Myofilament Proteins

Approximately 20 mg of swine cardiac left ventricular tissue was homogenized thoroughly in 200 µl of HEPES extraction buffer (25 mM HEPES, pH 7.5, 50 mM NaF, 0.25 mM Na_3VO_4, 0.25 mM PMSF, 2.5 mM EDTA) at 4 °C using a Teflon pestle (1.5 ml tube flat tip, Scienceware, Pequannock, NJ, USA). The resulting homogenate was centrifuged at 17,000 relative centrifugal force (rcf) for 15 min at 4 °C, and the supernatant was removed.[24] The pellet was subsequently homogenized in 100 µl TFA extraction solution (1% TFA, 1 mM TCEP) to extract the myofilament proteins. The homogenate was centrifuged at 17,000 rcf for 15 min at 4 °C, and the supernatant, which is enriched in myofilament proteins, was transferred to a new 1.5 ml microfuge tube and centrifuged for an additional 60 min at 17,000 rcf and 4 °C. The resulting supernatant was subject to LC-MS analysis.

5.2.3 Online LC-MS Analysis and Manual Fraction Collection

LC separation of the myofilament proteins and low-resolution MS analysis were performed as previously described[24] with minor modifications. Briefly, 3 µl of the myofilament extract (equivalent to 600 µg of tissue per injection) was injected and the proteins in the mixture were eluted at a flow rate of 12.5 µl/min with a gradient going from 20% B to 95% B in 57 min (mobile phase A: 0.10 % formic acid in water; mobile phase B: 0.10 % formic acid in 1:1 acetonitrile: ethanol). The flow was split after LC separation with approximately 10% of the eluting sample being ionized via electrospray ionization through a 50 µm inner diameter tip, and analyzed directly by a linear ion trap (LTQ) mass spectrometer (Thermo Scientific, Bremen, Germany). The remaining ~90% of the sample was simultaneously collected as fractions on ice for subsequent high-resolution MS and MS/MS analyses.

5.2.4 Offline High-resolution Fourier Transform Ion Cyclotron Resonance (FT-ICR) MS and MS/MS Analyses

Fractions containing unknown proteins were analyzed using a 7 Tesla LTQ/FT-ICR mass spectrometer (LTQ/FT Ultra, Thermo Scientific, Bremen, Germany) equipped with an automated chip-based nano-electrospray ionization source (Triversa NanoMate, Advion Bioscience, Ithaca, NY, USA) as described previously.[24] The sample was introduced into the mass spectrometer using a spray voltage of 1.3 to 1.5 kV versus the inlet of the mass spectrometer. The resolving power of the FT-ICR was set at 200,000 at 400 *m/z*. The automatic gain control for a full scan in the LTQ, FT-ICR cell, MS^n FT-ICR cell, and electron capture dissociation (ECD) were 3E4, 5E5, 5E5, and 8E5, respectively. For MS/MS experiments, the protein molecular ions of the individual charge states were first isolated and then fragmented using 1.5% to 4.5% electron energy for ECD (corresponding to 0.6 to 3.5 eV) with a 70 millisecond duration without additional delay.

5.2.5 Multi-dimensional LC-MS/MS analyses

The detailed experimental procedures were described previously.[184] Human embryonic kidney (HEK) 293 cell lysate was subject to ion exchange chromatography followed by reverse phase chromatography coupled to a Q Exactive benchtop Orbitrap mass spectrometer (Thermo Scientific, Bremen, Germany). LC-MS/MS data were acquired with eight microscans at a resolving power of 70,000 (at 200 *m/z*) with automatic gain control set to 5E5 ions. A 10 V offset in the source was used for all of the experiments. In the top two data-dependent MS/MS scans, the intact protein ions were injected into the collision cell for higher energy collision dissociation (25 V) with a 10 second dynamic exclusion window.

5.2.6 Top-down MS Data Analysis

The tandem mass spectra collected for unknown proteins were used to test MASH Suite Pro in terms of protein identification. Deconvolution was performed using enhanced-THRASH with a signal-to-noise ratio (S/N) threshold of 3 and a minimum fit of 60%. All deconvoluted masses were manually validated prior to identification with the targeted database from NCBI (Sscrofa10.2, containing 24,476 protein sequences) using MS-Align+. Alternatively, the raw MS/MS data were converted to mzXML files in centroid mode, and deconvoluted using MS-Deconv[222] with a S/N of 1. The *m/z*, charge state, and mass for each of the precursor ions were manually validated; and the list of reported monoisotopic masses of the fragment ions was aligned against the targeted database (Sscrofa10.2, containing 24,476 protein sequences) using MS-Align+. To determine whether MASH Suite Pro can process LC-MS/MS data, the dataset reported in a previous publication was used.[184] Similarly, deconvolution was performed using either enhanced-THRASH with an S/N threshold of 3 and a minimum fit of 60%, or MS-Deconv with an S/N of 1. The precursor list was automatically retrieved by MASH Suite Pro and the dataset was searched against a human protein database (Uniprot-Swissprot database, released January 2013, containing 20,232 protein sequences).

5.3 RESULTS

MASH Suite Pro is an integrated software tool combining protein identification, quantitation, and characterization with visual displays for validation and correction of the computational outputs, as well as features enabling interface customization (**Figure 5.1**). All tasks are conducted in a user-friendly interface that guides the user through the process of data analysis

(**Figure 5.2**). The zoom-in and extended views of the individual windows of the software are shown in supplemental **Figures 5.3-5.5**. MASH Suite Pro can identify proteins with unexpected modifications, provide relative quantitation of proteoforms from different experimental/biological conditions, and localize PTMs and sequence variations. MASH Suite Pro also provides visualization components for validation and correction of the deconvoluted mass list and fragment ion assignments. It also allows users to output spectra and protein sequences directly. Moreover, MASH Suite Pro provides users with flexibility in terms of customizing the program interface, which can significantly improve the user experience.

Figure 5.1. Schematic summarizing the various functions and features of MASH Suite Pro. The major functions of MASH Suite Pro include protein identification, quantitation, and characterization of protein PTMs from top-down MS and MS/MS experiments. The program is equipped with various visualization components for the validation of the deconvolution results, identification results, and fragment ion assignments. Additional features include direct output of the graphics and customization of the program interface.

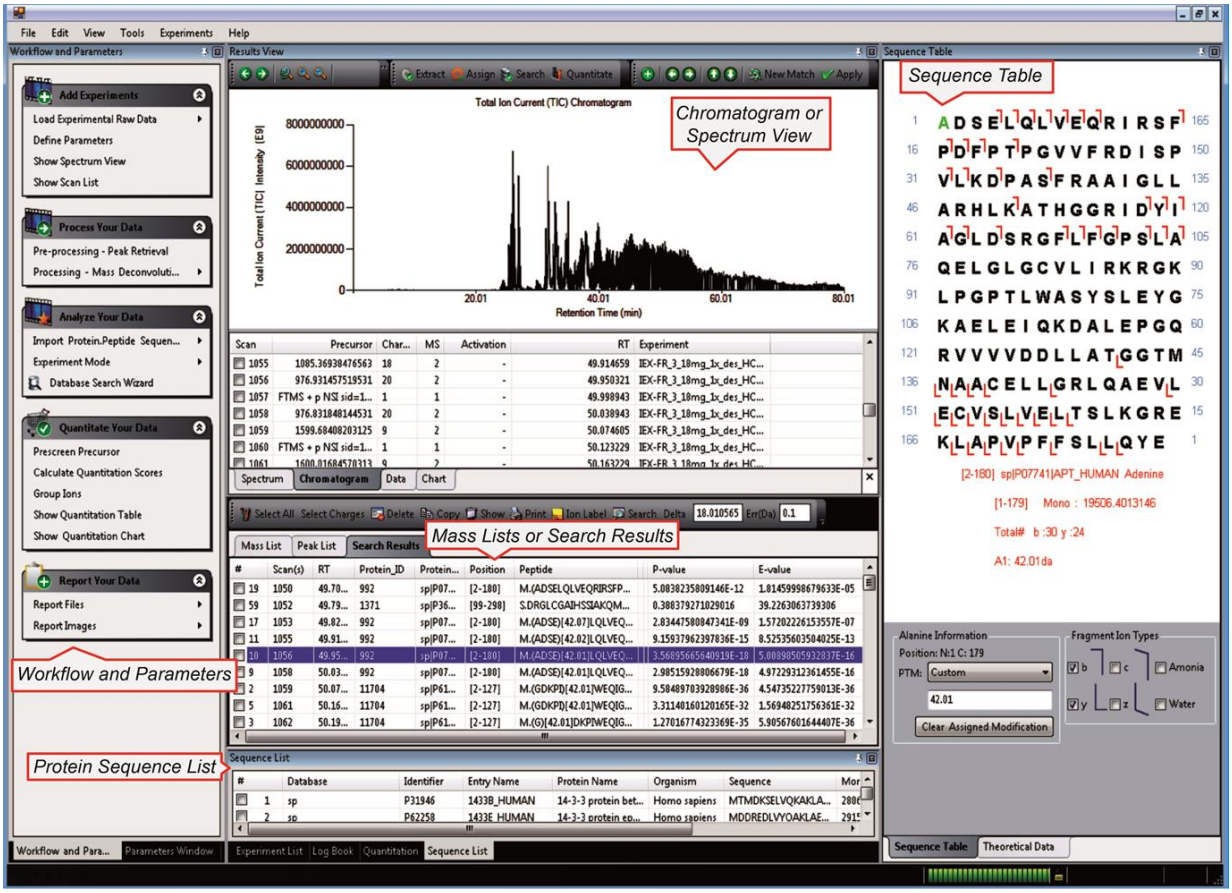


Figure 5.2. Overview of the MASH Suite Pro interface. The intuitive "*Workflow and Parameters*" window guides users through the various steps of high-resolution data analysis. Different windows allow for clear visualization of the data, including chromatograms and mass spectra, as well as the sequence table, which displays the fragmentation map of the selected protein with detailed bond cleavage information. The deconvoluted masses and search results are organized into lists, providing detailed information of each item. The complete and extended views of the "*Workflow and Parameters*", "*Chromatogram/Spectrum*" and "*Sequence Table*" windows are shown in **Figures 5.3-5.5**.

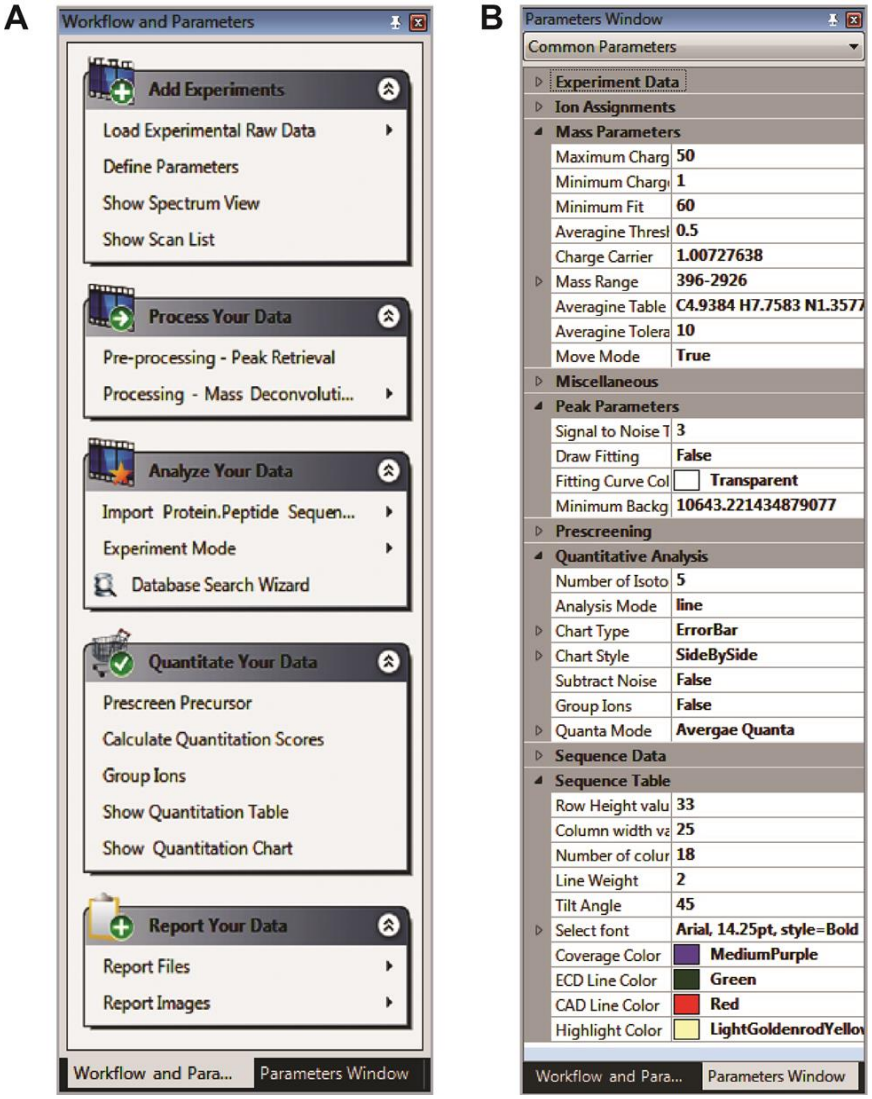


Figure 5.3. The "*Workflow and Parameters*" window and advanced parameter settings. (A). The intuitive "*Workflow and Parameters*" window guides users through the various steps of data processing and analysis. (B). Advanced parameters can also be defined in the "*Parameters Window*".

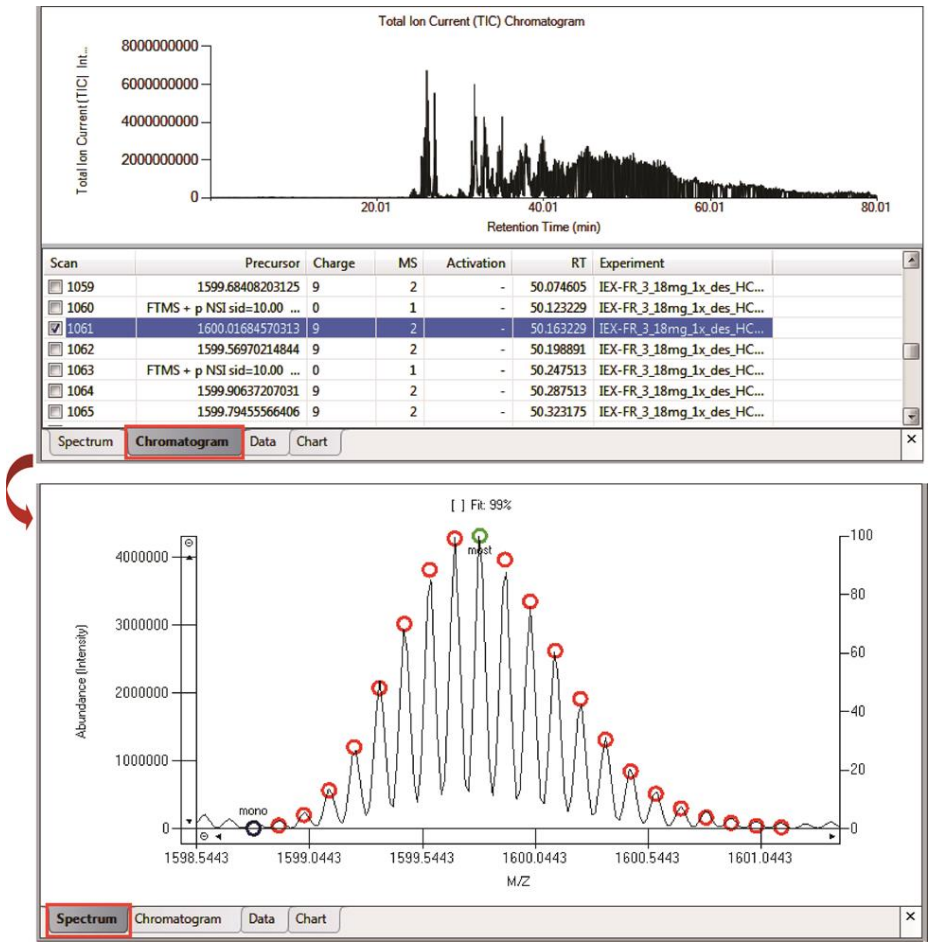


Figure 5.4. The chromatogram and spectrum view in MASH Suite Pro. MASH Suite Pro provides users with detailed information for each scan in an LC-MS/MS run. Each scan contains a mass spectrum that can be deconvoluted and visualized in the spectrum view tab.

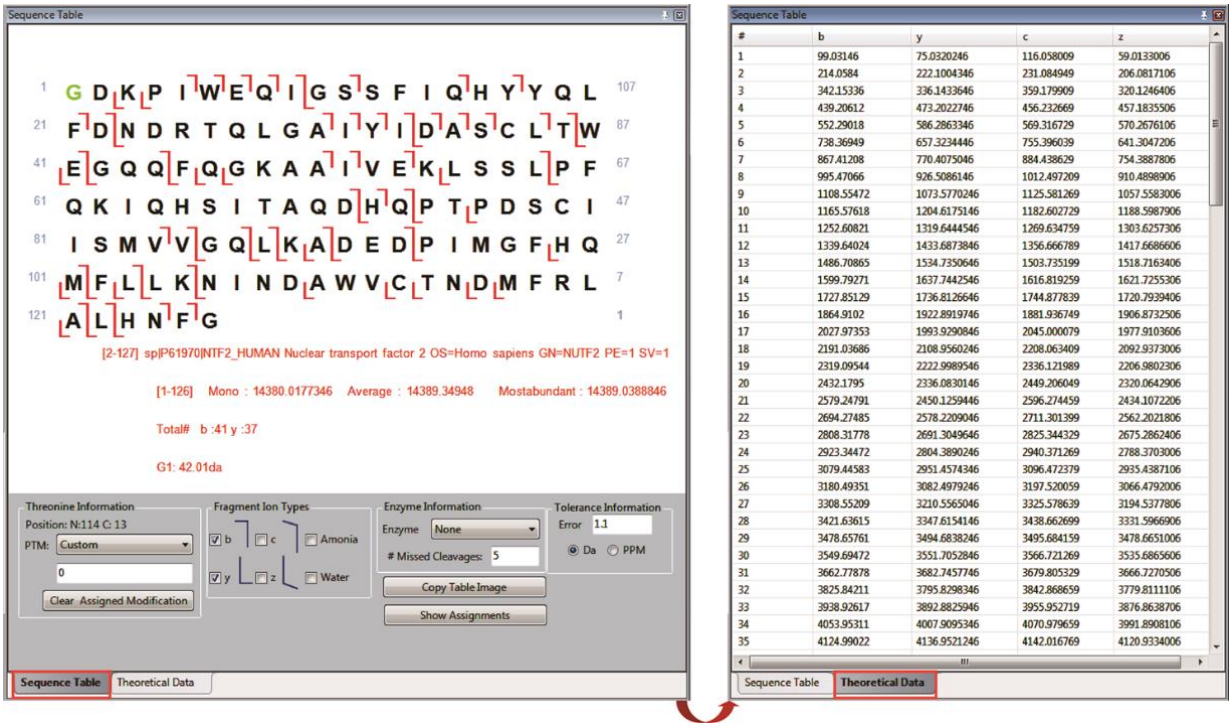


Figure 5.5. The "*Sequence Table*" in MASH Suite Pro. MASH Suite Pro displays the database sequence for identified proteins with bond cleavages representing fragment ions from the tandem mass spectrum. The "*Sequence Table*" allows users to modify the protein sequence with different PTMs for comprehensive protein characterization. Users can selectively display different fragment ions from the MS/MS data. The theoretical masses of all predicted ***b/y*** and ***c/z*** ions are also provided in the *"Theoretical Data"*.

5.3.1 MASH Suite Pro for Top-down Protein Identification

MASH Suite Pro utilizes the MS-Align+ algorithm to identify proteins from top-down LC-MS/MS and MS/MS data.[228] With MS-Align+, MASH Suite Pro can identify truncated proteoforms, as well as those with unexpected PTMs and sequence variations. The protein identification results are output as PrSMs with statistical evaluation (p-value and e-value),[228] allowing users to determine the confidence of the identification results. The workflow for protein identification in MASH Suite Pro is outlined in **Figure 5.6A**. The raw data (e.g., .raw files from Thermo Scientific instruments) can be directly read by the program and all the steps for protein identification can be performed in an integrated search wizard **(Figure 5.6B**). The "*Database*

Search Wizard" allows users to select different deconvolution methods (MS-Deconv or enhanced-THRASH) and fragmentation types. For LC-MS/MS data, MASH Suite Pro also allows users to select a certain scan range to minimize the duration of data processing and database search. The precursor *m/z*, charge states, and monoisotopic masses can also be automatically retrieved by the program. For MS/MS data lacking the precursor information, users are required to manually input the precursor information before searching the database. Additionally, as reported previously,[255] users are able to perform deconvolution using MASH Suite Pro and validate the program-determined precursor ion monoisotopic masses prior to the search.

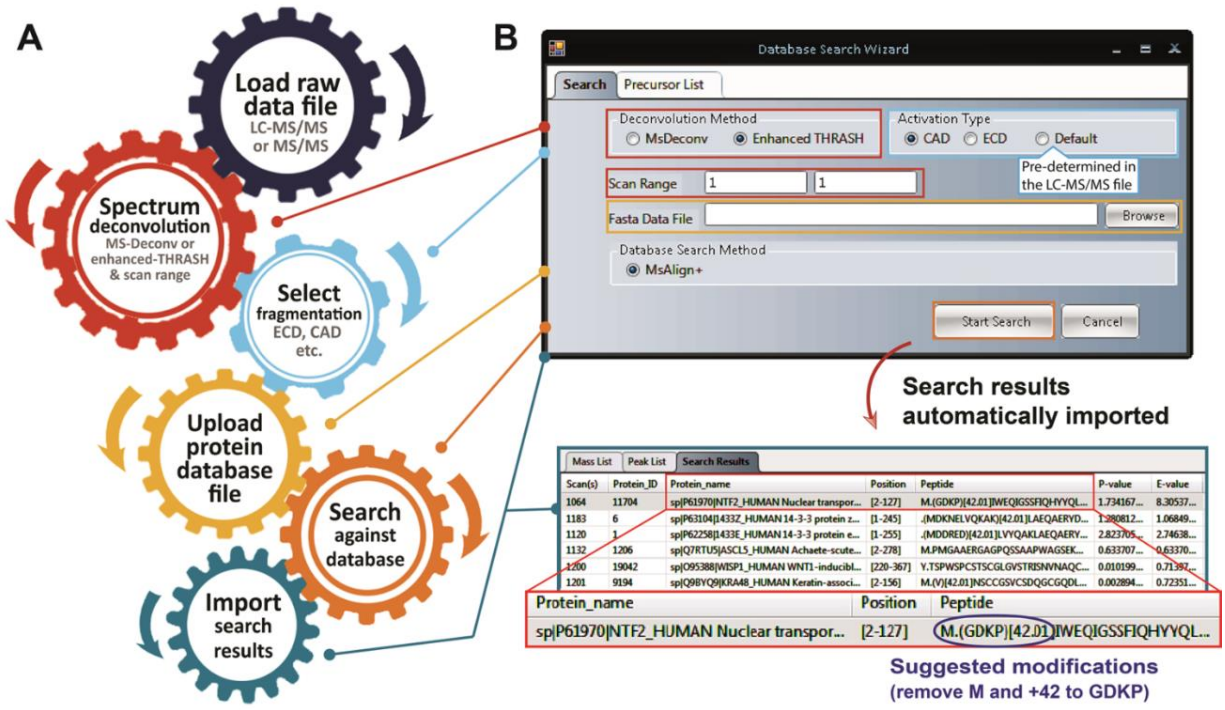


Figure 5.6. MASH Suite Pro for top-down protein identification. (A). The workflow for protein identification in MASH Suite Pro. The raw data (e.g. .raw from the instruments of Thermo Scientific) can be directly read by the program and all of the steps for protein identification can be performed in an integrated search wizard. (B). The highly integrated search wizard. MASH Suite Pro accommodates various options for deconvolution and database searching. The search results are automatically imported to the program for easy review and validation of the identification results.

MASH Suite Pro implements a "*Log Book*" (**Figure 5.7**), allowing for tracking of the search process. Once the database search is finished, the search results are automatically imported into the program (**Figure 5.2B**) with detailed information including the name and ID of the identified protein, the scan number, information pertaining to the identity and localization of any PTMs, p-value, e-value, among others (**Figure 5.8**). The identified proteins can be individually selected to view the protein sequence information, along with the localization of any mass discrepancies, in the "*Sequence Table*". In addition, bond cleavages within the sequence are also shown (**Figure 5.2 and 5.5**).

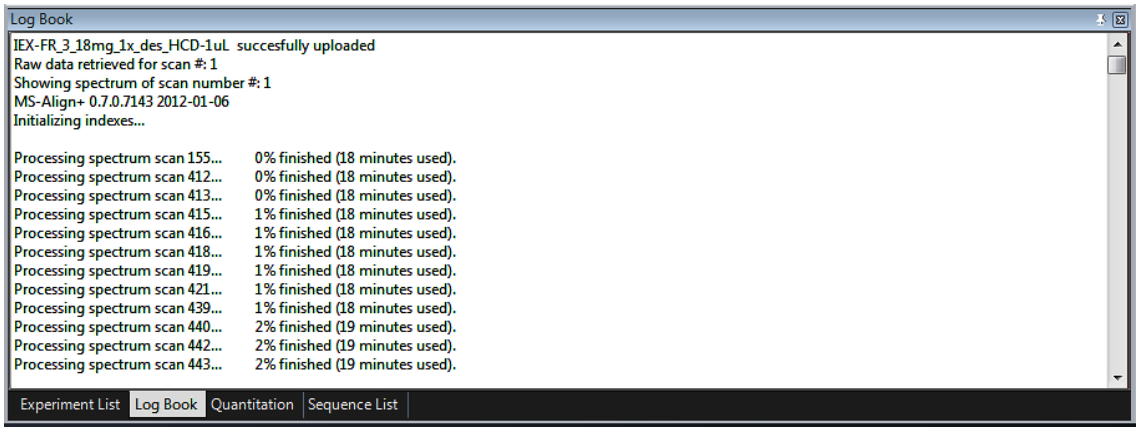
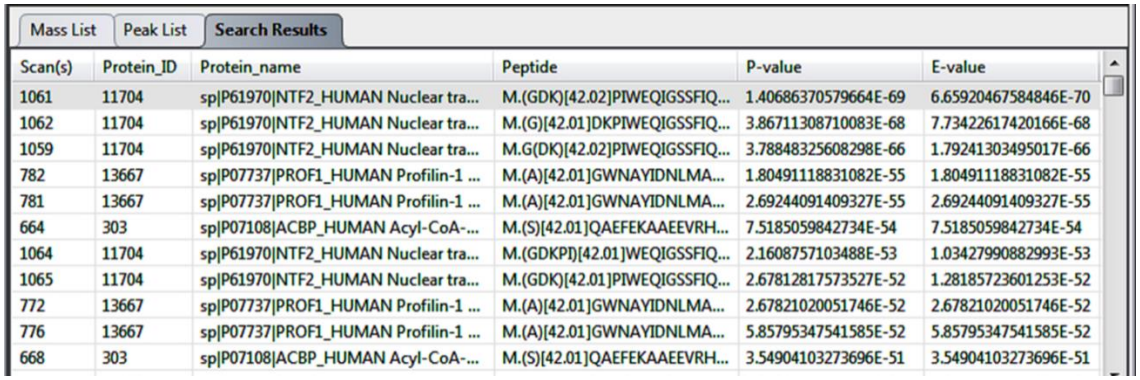


Figure 5.7. The *"Log Book"* in MASH Suite Pro. It provides detailed information regarding the progress of data processing and database search, allowing users to estimate the time needed for database search of large datasets.



Mass List | Peak List | Search Results

Scan(s)	Protein_ID	Protein_name	Peptide	P-value	E-value
1061	11704	sp|P61970|NTF2_HUMAN Nuclear tra...	M.(GDK)[42.02]PIWEQIGSSFIQ...	1.40686370579664E-69	6.65920467584846E-70
1062	11704	sp|P61970|NTF2_HUMAN Nuclear tra...	M.(G)[42.01]DKPIWEQIGSSFIQ...	3.86711308710083E-68	7.73422617420166E-68
1059	11704	sp|P61970|NTF2_HUMAN Nuclear tra...	M.G(DK)[42.02]PIWEQIGSSFIQ...	3.78848325608298E-66	1.79241303495017E-66
782	13667	sp|P07737|PROF1_HUMAN Profilin-1 ...	M.(A)[42.01]GWNAYIDNLMA...	1.80491118831082E-55	1.80491118831082E-55
781	13667	sp|P07737|PROF1_HUMAN Profilin-1 ...	M.(A)[42.01]GWNAYIDNLMA...	2.69244091409327E-55	2.69244091409327E-55
664	303	sp|P07108|ACBP_HUMAN Acyl-CoA-...	M.(S)[42.01]QAEFEKAAEEVRH...	7.5185059842734E-54	7.5185059842734E-54
1064	11704	sp|P61970|NTF2_HUMAN Nuclear tra...	M.(GDKPI)[42.01]WEQIGSSFIQ...	2.1608757103488E-53	1.03427990882993E-53
1065	11704	sp|P61970|NTF2_HUMAN Nuclear tra...	M.(GDK)[42.01]PIWEQIGSSFIQ...	2.67812817573527E-52	1.28185723601253E-52
772	13667	sp|P07737|PROF1_HUMAN Profilin-1 ...	M.(A)[42.01]GWNAYIDNLMA...	2.67821020051746E-52	2.67821020051746E-52
776	13667	sp|P07737|PROF1_HUMAN Profilin-1 ...	M.(A)[42.01]GWNAYIDNLMA...	5.85795347541585E-52	5.85795347541585E-52
668	303	sp|P07108|ACBP_HUMAN Acyl-CoA-...	M.(S)[42.01]QAEFEKAAEEVRH...	3.54904103273696E-51	3.54904103273696E-51

Figure 5.8. A representative list of the search results. The search results list provides information of each identified protein, including scan number, protein ID, protein name, modifications, p-value, e-value, and more.

It should be noted that the accuracy of the fragment ion monoisotopic masses resulting from deconvolution directly affects the identification results. To improve the accuracy of protein identification, MASH Suite Pro accommodates different algorithms for spectral deconvolution. Currently, MS-Deconv and enhanced-THRASH have been successfully implemented. Protein identification using enhanced-THRASH and MS-Deconv for deconvolution (**Figure 5.9**) reveals that the two methods yield similar results for the same dataset. For proteins that are identified with high confidence (e-value < 1E-30), MS-Deconv ususally results in smaller e-values (**Figure 5.9**), indicating higher confidence, compared with enhanced-THRASH. When the e-values are between 1E-10 and 1E-30, the reported e-values/p-values from MS-Deconv and enhanced-THRASH are similar (**Figure 5.9**). When the e-values are above 1E-10, the PrSMs are considered less confident. In these cases, enhanced-THRASH performs better than MS-Deconv (**Figure 5.9**). Nevertheless, both deconvolution methods can be optimized by adjusting parameters such as S/N, mass tolerance, and/or fit scores, depending on the datasets to be analyzed.

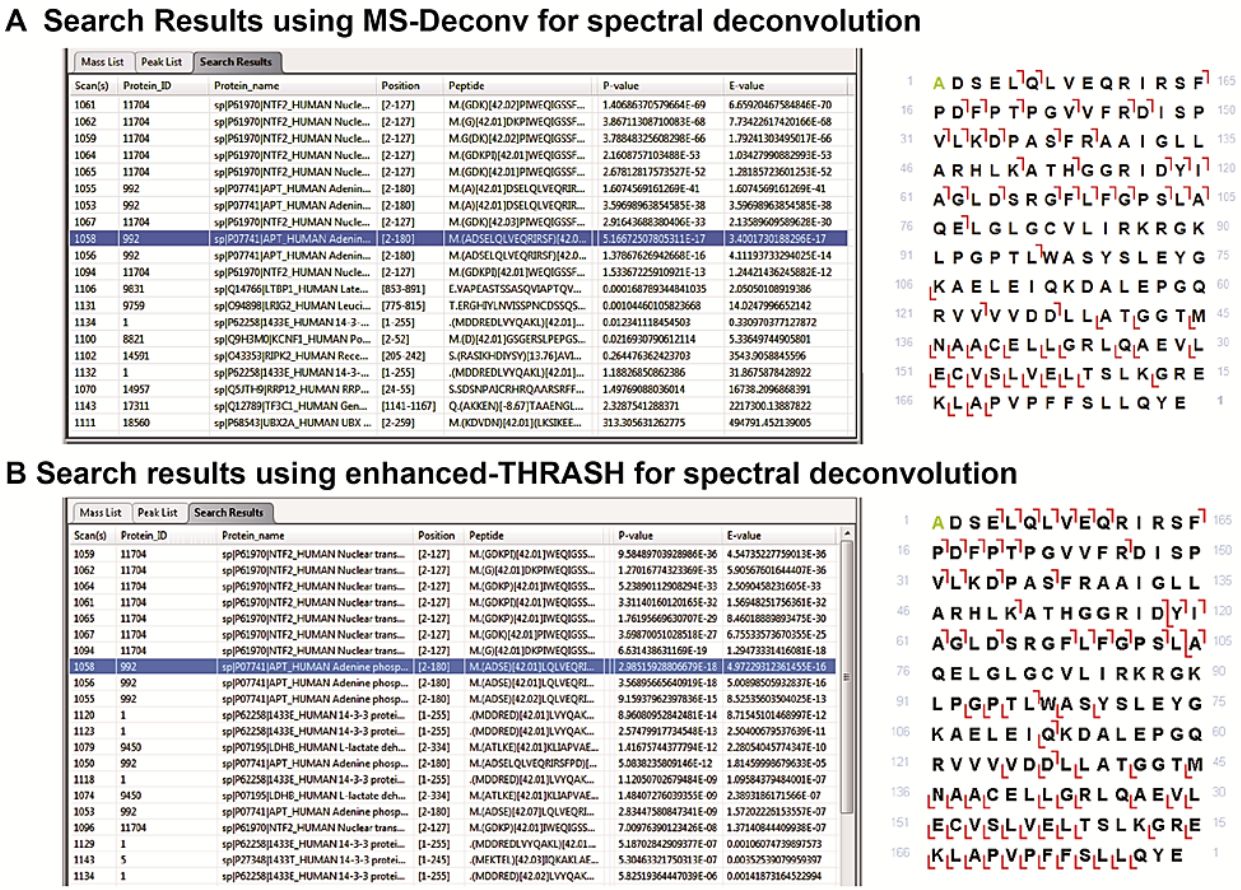


Figure 5.9. Protein identification using MS-Deconv and enhanced-THRASH for deconvolution. The same LC-MS/MS dataset was deconvoluted using (A) MS-Deconv or (B) enhanced-THRASH prior to protein identification using MS-Align+. The identification results from both deconvolution methods were similar. The fragmentation maps of the protein identified in scan 1058 are shown. According to the p-value, MS-Deconv outperformed enhanced-THRASH in some cases (e.g. scan 1061, 1155, 1067, etc.) while enhanced-THRASH performed better in other cases (e.g. scan 1056, 1094, 1134, etc). It is important to note that the quality of the deconvolution results depend on the dataset being analyzed and that deconvolution can be optimized by changing different parameters, including but not limited to S/N and mass tolerance.

5.3.2 MASH Suite Pro for Top-down Protein Quantitation

MASH Suite Pro significantly simplifies and accelerates the quantitation of various proteoforms from top-down mass spectra, and can present the quantitation results in both table and chart formats. Unlike the previous version,[255] MASH Suite Pro also allows for the quantitation of several proteoforms from mass spectra using multiple charge states (note that quantitation is based

on S/N values of the most abundant isotopomers from the mass spectra, rather than the tandem mass spectra) When multiple charge states are selected for quantitation, the program automatically normalize the proteoform abundance to the charge state to ensure accurate quantitation. In addition, it enables rapid quantitation and comparison of the selected proteoforms between different biological/experimental conditions (e.g. healthy *vs.* diseased). Furthermore, it provides the option for averaging of the quantitation results among replicates, allowing accurate quantitation with minimal manual intervention.

The workflow for quantitation in MASH Suite Pro is outlined in **Figure 5.10A**. Users can upload several data files and define the ions to be quantitated. For each raw data file, MASH Suite Pro calculates the relative abundances of the selected ions by summing the S/N values of the most abundant isotopomers. Users can also define the number of charge states to be calculated and the number of isotopomers of each charge state to be considered in the quantitation. Once the target ions are defined, the program will automatically search and carry out quantitation of the selected ions, generating a quantitation table and quantitation chart (**Figure 5.10**). Detailed procedures on the protein quantitation involving multiple charge states are outlined in supplemental **Figure 5.11**.

An example of quantitation in MASH Suite Pro is shown in **Figure 5.10B**. The relative quantities of troponin I (TnI) proteoforms, including un-phosphorylated, mono-phosphorylated TnI (*p*TnI), and bis-phosphorylated TnI (*pp*TnI), in the left ventricular tissue of control swine (Control) and swine with myocardial infarction (Disease) were compared. Two groups of samples from the diseased animals were analyzed (Disease 1 and Disease 2) (**Figure 5.10B**). For quantitation, the S/N ratios of the most abundant isotopomers within each isotopomer envelope were summed and the resulting values were defined as the quantities of the selected ions. Once the proteoforms and charge state(s) are defined, MASH Suite Pro automatically normalizes the

proteoform abundance to the corresponding charge states, finds the ions and calculates the relative percentage of each proteoform). MASH Suite Pro can also average the relative abundance of the same proteoform in different biological replicates and presents the quantitation chart with error bars (standard error of the mean). Most importantly, users can directly compare the relative abundance of various proteoforms in different biological conditions (e.g. healthy *vs.* diseased) (**Figure 5.10B**). MASH Suite Pro also provides users with a quantitation table showing the exact relative abundance of each proteoform in each experiment (**Figure 5.10B**, **Figure 5.12**). The quantitation feature in MASH Suite Pro significantly facilitates the analysis of molecular changes in different biological/experimental conditions.

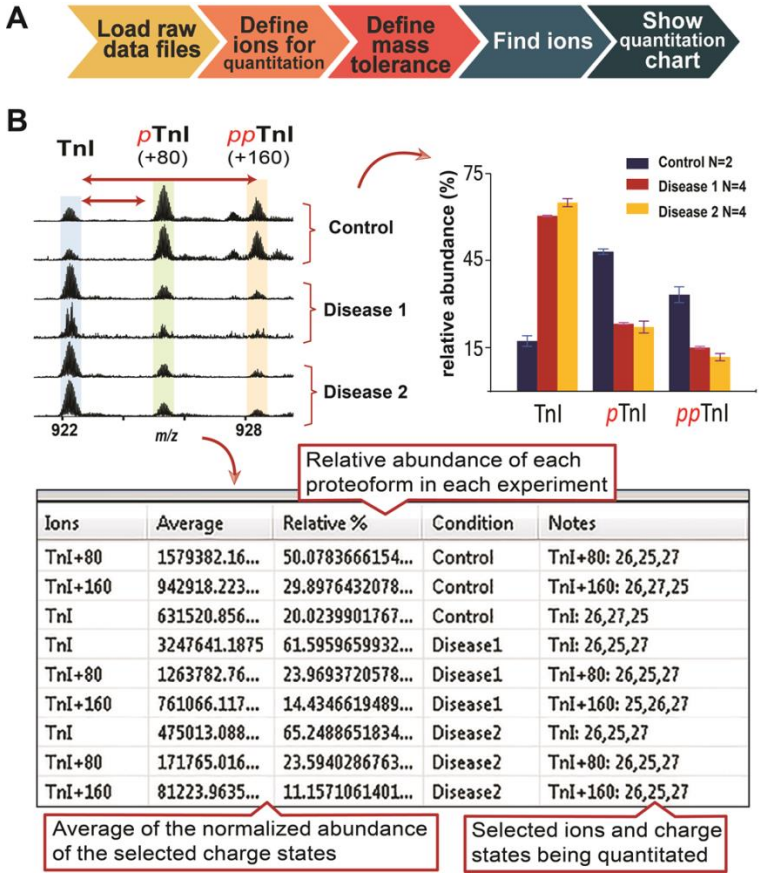
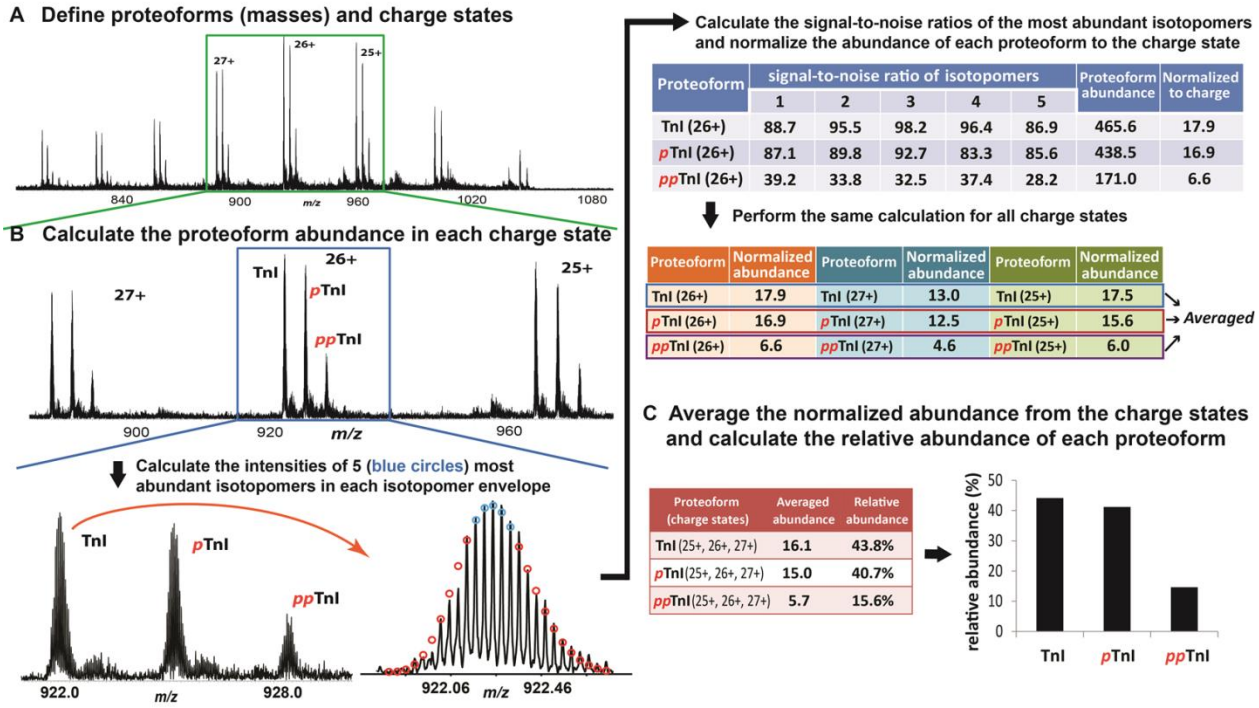


Figure 5.10. MASH Suite Pro for top-down protein quantitation. (A). The workflow for protein quantitation in MASH Suite Pro. Several raw data files corresponding to the same experimental/biological condition can be quantitated and averaged to ensure accurate quantitation. The relative abundances of proteoforms in different conditions can also be directly compared in MASH Suite Pro. (B). Quantitation of

Troponin I (TnI) proteoforms in cardiac left ventricular tissue from healthy (Control) and disease (Disease 1 and Disease 2) pigs. The Control, Disease 1, and Disease 2 conditions have 2, 4, and 4 biological replicates, respectively. The relative abundance of TnI, mono-phosphorylated TnI (*p*TnI), and bis-phosphorylated TnI (*pp*TnI) were determined using MASH Suite Pro. The quantitation chart provides a visual comparison of the relative percentage of each proteoform in different groups, and the quantitation table shows the exact relative percentage of the each proteoform in each experiment (only one replicate from each condition is shown). A complete quantitation table for all the experiments analyzed is shown in supplemental Figure S8.



Proteoform	signal-to-noise ratio of isotopomers					Proteoform abundance	Normalized to charge
	1	2	3	4	5		
TnI (26+)	88.7	95.5	98.2	96.4	86.9	465.6	17.9
*p*TnI (26+)	87.1	89.8	92.7	83.3	85.6	438.5	16.9
*pp*TnI (26+)	39.2	33.8	32.5	37.4	28.2	171.0	6.6

Proteoform	Normalized abundance	Proteoform	Normalized abundance	Proteoform	Normalized abundance
TnI (26+)	17.9	TnI (27+)	13.0	TnI (25+)	17.5
*p*TnI (26+)	16.9	*p*TnI (27+)	12.5	*p*TnI (25+)	15.6
*pp*TnI (26+)	6.6	*pp*TnI (27+)	4.6	*pp*TnI (25+)	6.0

Proteoform (charge states)	Averaged abundance	Relative abundance
TnI (25+, 26+, 27+)	16.1	43.8%
*p*TnI (25+, 26+, 27+)	15.0	40.7%
*pp*TnI (25+, 26+, 27+)	5.7	15.6%

Figure 5.11. Protein quantitation with multiple charge states. (A). Users define proteoforms of interest and charge states. (B). MASH Suite Pro calculates the abundance of each proteoform in each charge state. The proteoform abundance is represented by the sum of the S/N values of the most abundant isotopomers (blue circles) (top 5 is set as default). After calculating the abundance values for each of the proteoforms in each of the selected charge states, the abundance values for all of the proteoforms within a given charge state are normalized to the charge state. (C). MASH Suite Pro averages the normalized abundances of each of the proteoforms across the selected charge state, and subsequently calculates the relative abundances of each individual proteoform as a percentage of the entire proteoform population.

Quantitation

#	Replicate	Ions	Average	Relative %	Condition	Notes	File Name
1	Exp-1	TnI+80	14107.8077...	47.3990874105...	Control	TnI+80: 24,25,26	01092013_...
2	Exp-1	TnI+160	11103.4198...	37.3050140592...	Control	TnI+160: 24,25,26	01092013_...
3	Exp-1	TnI	4552.65297...	15.2958985302...	Control	TnI: 24,25,26	01092013_...
4	Exp-2	TnI+80	1579382.16...	50.0783666154...	Control	TnI+80: 26,25,27	01132013_...
5	Exp-2	TnI+160	942918.223...	29.8976432078...	Control	TnI+160: 26,27,25	01132013_...
6	Exp-2	TnI	631520.856...	20.0239901767...	Control	TnI: 26,27,25	01132013_...
7	Exp-3	TnI	2937134.86...	60.5518573924...	Disease1	TnI: 26,25,27	01212013_I...
8	Exp-3	TnI+80	1109767.09...	22.8789149623...	Disease1	TnI+80: 26,27	01212013_I...
9	Exp-3	TnI+160	803708.726...	16.5692276451...	Disease1	TnI+160: 25	01212013_I...
10	Exp-4	TnI	3247641.1875	61.5959659932...	Disease1	TnI: 26,25,27	01212013_I...
11	Exp-4	TnI+80	1263782.76...	23.9693720578...	Disease1	TnI+80: 26,25,27	01212013_I...
12	Exp-4	TnI+160	761066.117...	14.4346619489...	Disease1	TnI+160: 25,26,27	01212013_I...
13	Exp-5	TnI	2717905.98...	61.3963685413...	Disease1	TnI: 26,25,27	01212013_I...
14	Exp-5	TnI+80	1054706.57...	23.8253839616...	Disease1	TnI+80: 26,25,27	01212013_I...
15	Exp-5	TnI+160	654206.239...	14.7782474969...	Disease1	TnI+160: 25,26,27	01212013_I...
16	Exp-6	TnI	2285550.07...	61.3033750125...	Disease1	TnI: 26,25,27	01212013_I...
17	Exp-6	TnI+80	882936.625	23.6822617347...	Disease1	TnI+80: 26,25,27	01212013_I...
18	Exp-6	TnI+160	559774.710...	15.0143632527...	Disease1	TnI+160: 25,26,27	01212013_I...
19	Exp-7	TnI	609640.348...	69.9547393211...	Disease2	TnI: 25,24,26	01212013_...
20	Exp-7	TnI+80	180306.666...	20.6897491054...	Disease2	TnI+80: 25,24,26	01212013_...
21	Exp-7	TnI+160	81531.25	9.35551157337...	Disease2	TnI+160: 25,24,26	01212013_...
22	Exp-8	TnI	351327.302...	63.1980517305...	Disease2	TnI: 25,24,23	01212013_...
23	Exp-8	TnI+80	116811.421...	21.0124695324...	Disease2	TnI+80: 25,24,23	01212013_...
24	Exp-8	TnI+160	87776.0439...	15.7894787369...	Disease2	TnI+160: 26,25,24	01212013_...
25	Exp-9	TnI	523558.106...	64.8349212649...	Disease2	TnI: 26,25,27	01212013_...
26	Exp-9	TnI+80	193712.983...	23.9884854099...	Disease2	TnI+80: 26,25,27	01212013_...
27	Exp-9	TnI+160	90253.7695...	11.1765933250...	Disease2	TnI+160: 26,25,27	01212013_...
28	Exp-10	TnI	475013.088...	65.2488651834...	Disease2	TnI: 26,25,27	01212013_...
29	Exp-10	TnI+80	171765.016...	23.5940286763...	Disease2	TnI+80: 26,25,27	01212013_...
30	Exp-10	TnI+160	81223.9635...	11.1571061401...	Disease2	TnI+160: 26,25,27	01212013_...

Figure 5.12. An example of the complete quantitation table for the datasets presented in Figure 5.10. The quantitation table provides detailed information for the experimental datasets being analyzed, including the selected ions, the average abundance of the selected charge states, the relative abundance of each proteoform (relative to the entire proteoform population being quantitated), the experimental/biological conditions, and the charge states chosen for quantitation (Notes). Different experiments with the same condition (e.g. Exp-1 and Exp-2) are considered as replicates (technical or biological) and therefore the quantitation results are averaged and displayed with error bars (standard error of the mean) in the quantitation chart **(Figure 5.10**).

5.3.3 MASH Suite Pro for Comprehensive Protein Characterization

MASH Suite Pro can facilitate the characterization of PTMs and sequence variations by allowing assignment of modifications to individual amino acids in the protein sequence (**Figure 5.13**). Once the protein is identified, the amino acid sequence will be displayed in the "*Sequence Table*" window with bond cleavages (from the tandem mass spectrum) indicated. The identified PTMs are also automatically incorporated into the protein sequence to provide more accurate assignment of the fragment ions. For example, human nuclear transport factor 2 was identified in our LC-MS/MS experiment[184] using MS-Align+, and the identification results indicate that the N-terminal Met is removed and Gly1 is acetylated (**Figure 5.13**). Without taking into account acetylation at the N-terminus, only 5 ***b*** ions and 38 ***y*** ions could be assigned (**Figure 5.13**); however, when N-terminal acetylation was taken into account, an additional 36 ***b*** ions were assigned (41 ***b*** ions and 38 ***y*** ions total). MASH Suite Pro also allows users to manually change the modifications. Moreover, users can choose to display different types of fragment ions, including ***b/y*** and ***c/z***· pairs, as well as ***b/y*** ions with water loss ($-H_2O$) and neutral ammonium loss ($-NH_3$) (supplemental Figure **Figure 5.14**), and also visualize the fragment ions to validate and refine the fragment ion assignments (**Figure 5.13C**).

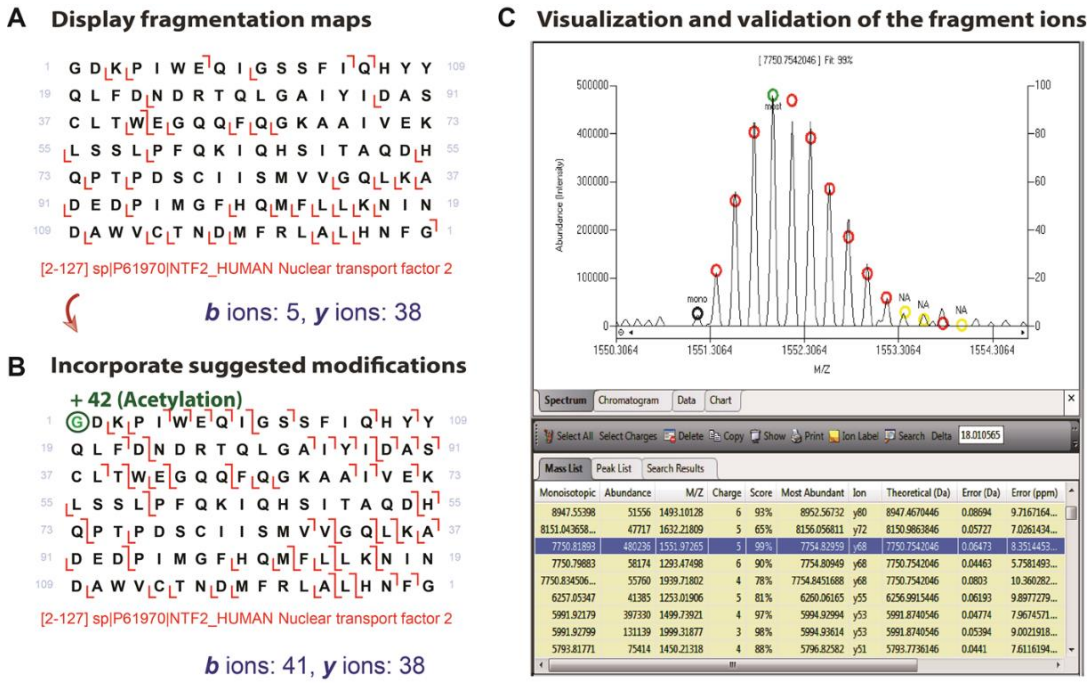


Figure 3.13. MASH Suite Pro for comprehensive characterization of protein PTMs and sequence variations. (A). Once the protein is identified (as shown in Figure 2), the protein sequence can be displayed in the *"Sequence Table"* showing bond cleavages. (B). Suggested PTMs from MS-Align+ can be incorporated into the protein sequence. (C). Users are also able to visualize and validate the fragment ions of each identified protein in the *"Mass List"*.

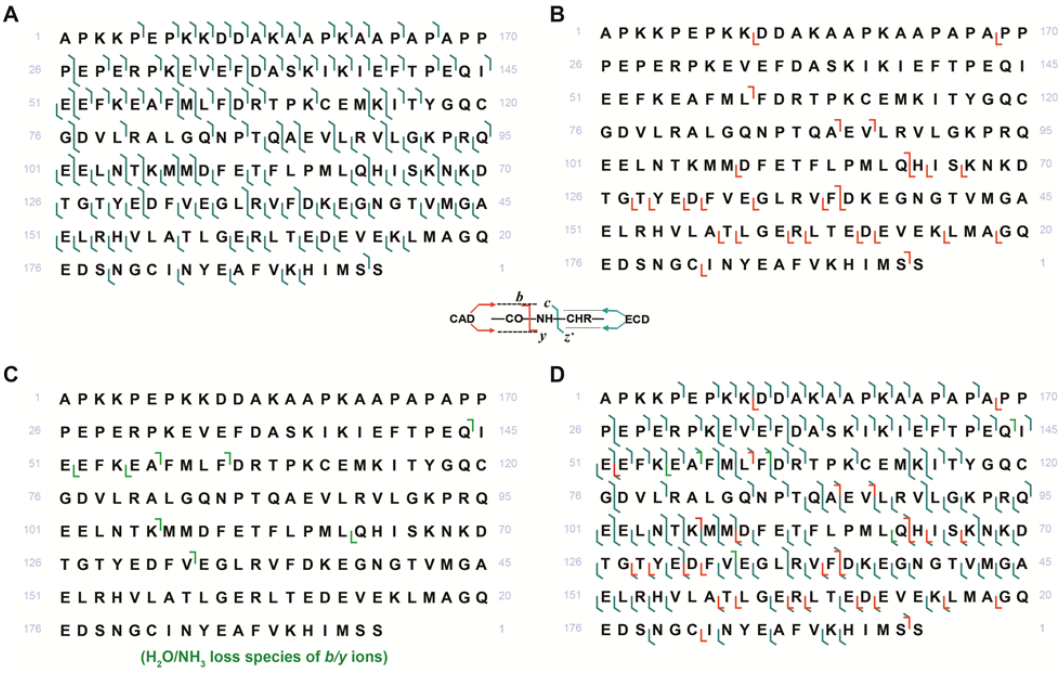


Figure 3.14. Direct output of different fragmentation maps for myosin light chain 1. Figures were produced based on (A) *c/z·* ions, (b) *b/y* ions, (C) *b/y* ions with water loss (-H_2O)/neutral ammonium loss (-

NH_3), and (D) combined all fragment ions. The fragmentation maps can be directly exported from MASH Suite Pro.

5.3.4 MASH Suite Pro for Identification, Characterization, and Quantitation of an Unknown Protein: An Example

To demonstrate the capabilities of MASH Suite Pro, we fraction collected an unknown protein that we observed during a LC-MS run. The fraction-collected protein was subsequently analyzed offline by high-resolution MS (**Figure 5.15A**). High-resolution MS analysis showed that the unknown protein has multiple proteoforms (P1, P2 and P3), and that the relative abundances of these proteoforms change under different experimental conditions (Condition 1 and Condition 2) (**Figure 5.15A**). The most abundant proteoform (P2) was selected for fragmentation using ECD and the tandem mass spectra were highly complex (**Figure 5.15B**), necessitating a sophisticated tool for data processing and analysis. Spectral deconvolution, protein identification, and characterization were performed in MASH Suite Pro, which identified the protein as a truncated form of ATPase inhibitor (**Figure 5.15C**). MASH Suite Pro can also identify unexpected PTMs and sequence variations, facilitating protein characterization. In this case, the ATPase inhibitor has one unexpected sequence variation at position 37 (Ala37Val). Once P2 was identified, according to the protein sequence, the other proteoforms (P1 and P3) were subsequently deduced (**Figure 5.15C**). The relative abundance of each proteoform under two experimental conditions was also determined by MASH Suite Pro and presented with a quantitation table for visual comparison (**Figure 5.15D**). This demonstrates the power of MASH Suite Pro for proteoform identification, comprehensive characterization, and quantitative analysis.

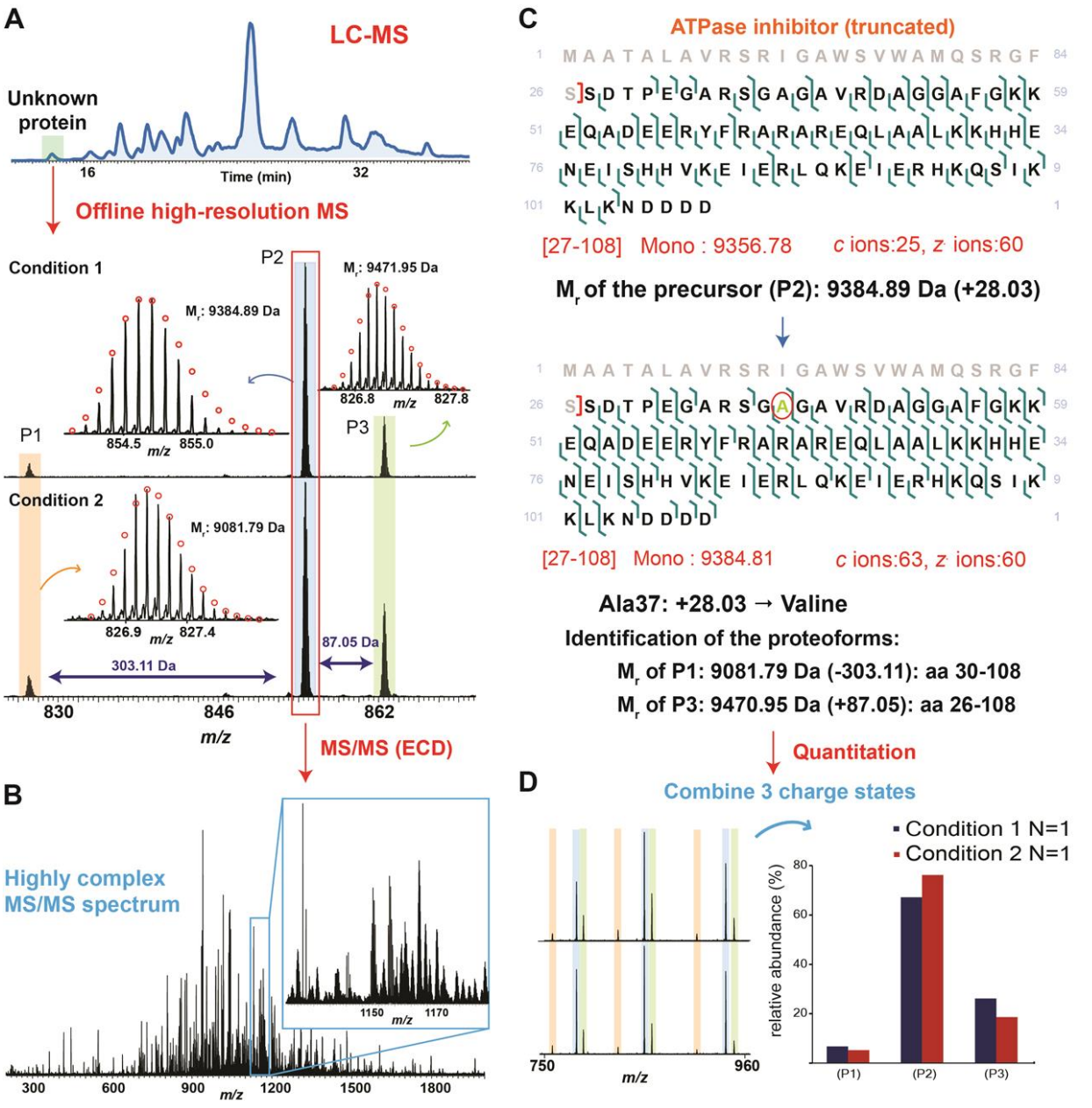


Figure 5.15. Identification and characterization of an unknown protein followed by quantitation of the proteoforms using MASH Suite Pro. (A). An unknown protein was detected in LC-MS and the fraction containing the protein was collected and analyzed by high-resolution MS. Top-down MS revealed 3 proteoforms (P1, P2 and P3). The relative abundance of each proteoform varied in two different experimental conditions. (B). The proteoform (P2) was selected for fragmentation using ECD, resulting in a highly complex MS/MS spectrum. (C). MASH Suite Pro was used for spectral deconvolution and protein identification. The proteoform P2 was identified to be a truncated form of the ATPase inhibitor containing amino acids (aa) 27-108. Characterization of the protein sequence using MASH Suite Pro identified a sequence variation (Ala37Val). Based on the protein sequence, the proteoforms (P1) and (P3) were deduced to be aa 30-108 and aa 26-108 of ATPase inhibitor. (D). MASH Suite Pro provided rapid quantitation to determine the relative abundances of the different proteoforms in different experimental conditions.

5.3.5 Data Visualization and Validation

For very complex spectra, spectral deconvolution methods sometimes fail to accurately deconvolute the isotopomer envelopes, resulting in miscalculation of the mass value or charge state (**Figure 5.16 and 5.17**). Charge state miscalculation and mass shifts are common primarily due to overlapping isotopomer envelopes (**Figure 5.16**). For high molecular weight ions and high charge state ions, even without the involvement of overlapping peaks, the current deconvolution methods can still result in charge state calculation errors and mass shifts. MASH Suite Pro provides users with visualization features for validation of the computational outputs. Users are able to visualize the spectrum of each ion assignment together with the automatically generated theoretical isotopomer envelope (based on averagine) in the "*Spectrum View*" (**Figure 5.18**). MASH Suite Pro allows users to correct the charge states (**Figure 5.16A**) and shift the theoretical isotopomer envelope incrementally by one Dalton to fit the real spectrum (**Figure 5.16B**) in order to obtain an accurate mass value for the selected ion (**Figure 3.16A**).[255]

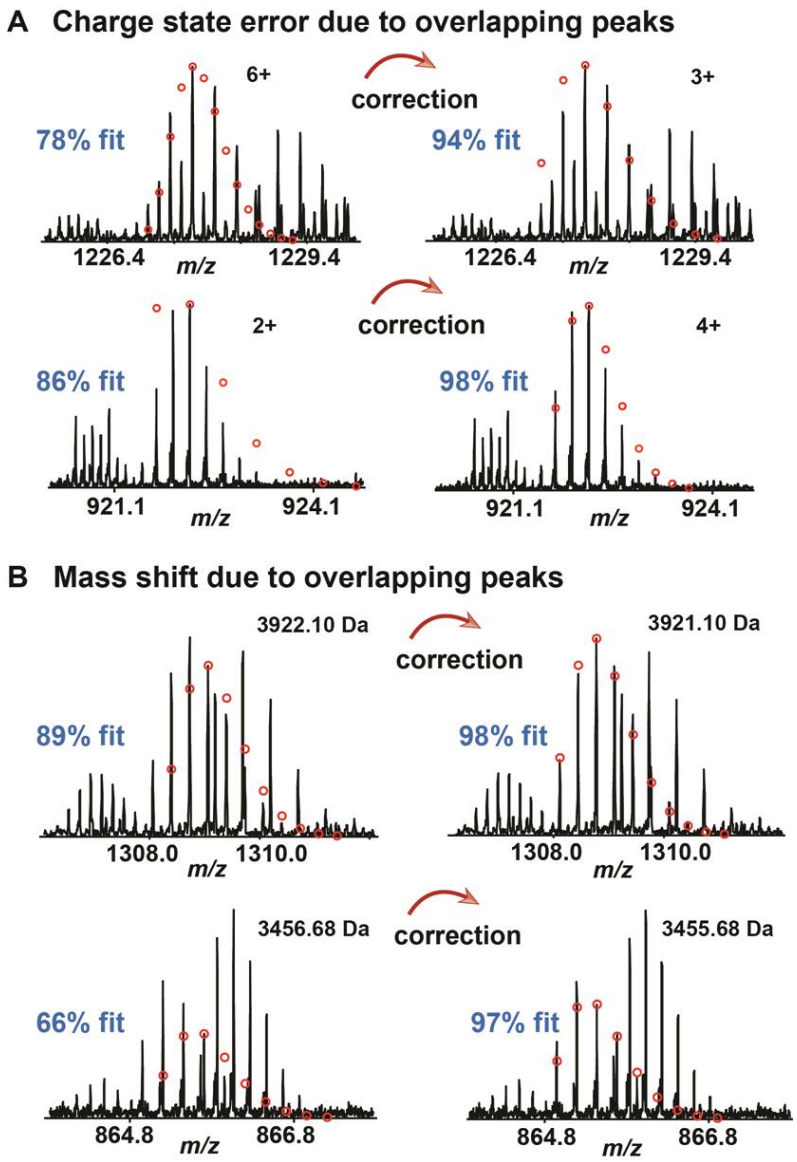


Figure 5.16. Visual validation of the computational outputs. (A). Charge state correction. Users can adjust/correct the charge state to better fit the data and acquire accurate mass values. (B). Mass shift correction. Users can shift the theoretical isotopomer envelope incrementally by one Dalton to fit the real spectrum for obtaining accurate mass values.

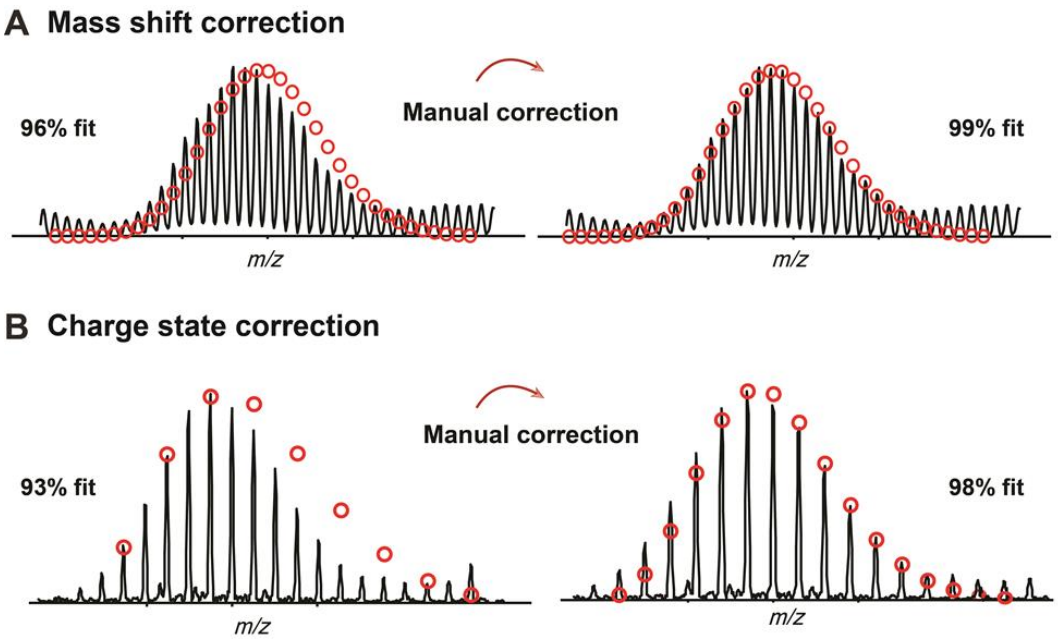


Figure 5.17. Examples of deconvolution errors (charge state and mass error). Automatic data processing commonly resulted in errors in mass accuracy (A) and charge state assignment (B). MASH Suite Pro allows users to visually validate the computational output for acquiring accurate mass values.

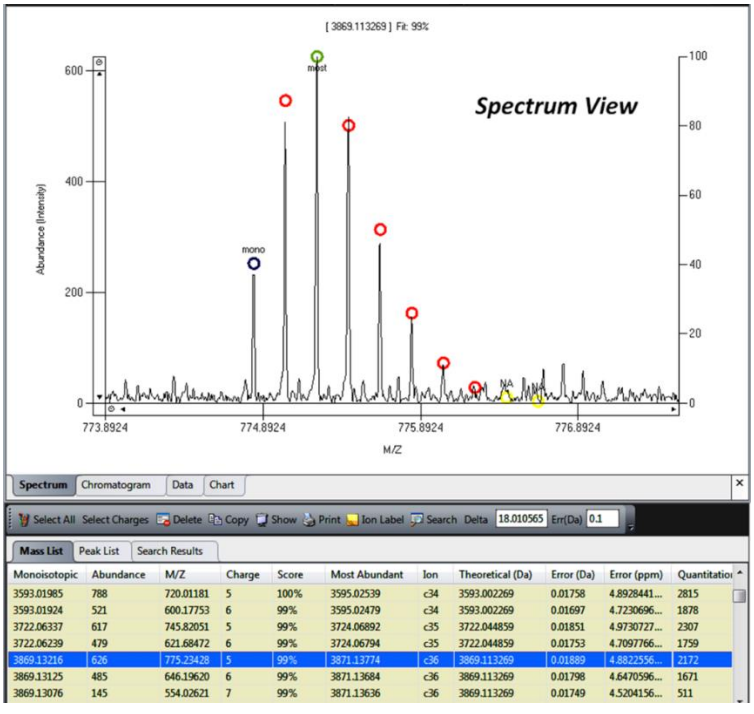


Figure 5.18. A representative spectrum view of a selected ion. Spectrum shows a fragment ion overlaid with the theoretical isotopic abundance distribution calculated using averagine and the calculated mass of the ion. The blue and green circles represent the monoisotopomer and most abundant isotopomer, respectively. The yellow circles represent isotopomers that are considered "not found" by the program because the isotopomers are under detection limit or outside the mass tolerance.

5.3.6 Interface Customization and Table/Graphic Output

Similar to the previous MASH Suite,[255] MASH Suite Pro provides users with flexibility in the arrangement of various windows and tabs in the program interface (**Figure 5.2 and 5.19**). Users can show or hide any window by pinning/unpinning it. The pinned windows will be shown in the program, and the unpinned windows will be hidden. When users hover over the hidden window, the window will toggle and appear temporarily, allowing the user to re-pin the window if desired. MASH Suite Pro arranges different views in the form of tabs, allowing users to switch between tabs to access different data or parameters (**Figure 5.19**). For example, as shown in **Figure 5.4**, users can switch between the "Spectrum" and "Chromatogram" to visualize different aspects of the data. Moreover, each window can be relocated by dragging the window to a different position, which allows for easy access to different features throughout the process of data analysis. To facilitate reporting of the analyzed data, MASH Suite Pro also supports direct output of the tables and graphics (**Figure 5.20**), enabling further analysis of the data and editing/publishing of the figures.

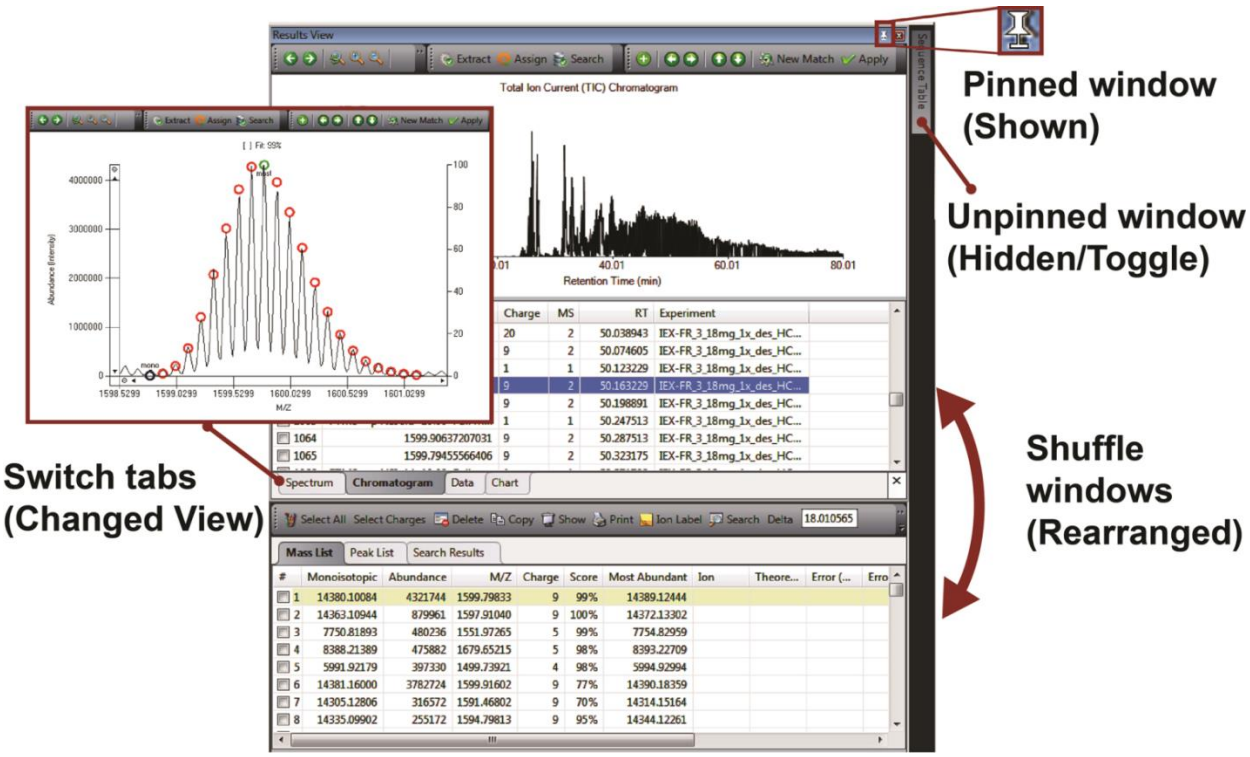


Figure 5.19. Customization of the program interface. Each window can be pinned or unpinned. Pinned windows are shown in the interface while the unpinned ones are hidden. Hovering over the unpinned window makes it toggle and the contents can be shown temporarily allowing the user to re-pin the window if desired. Users can switch tabs to change view and shuffle the windows to rearrange the entire interface.

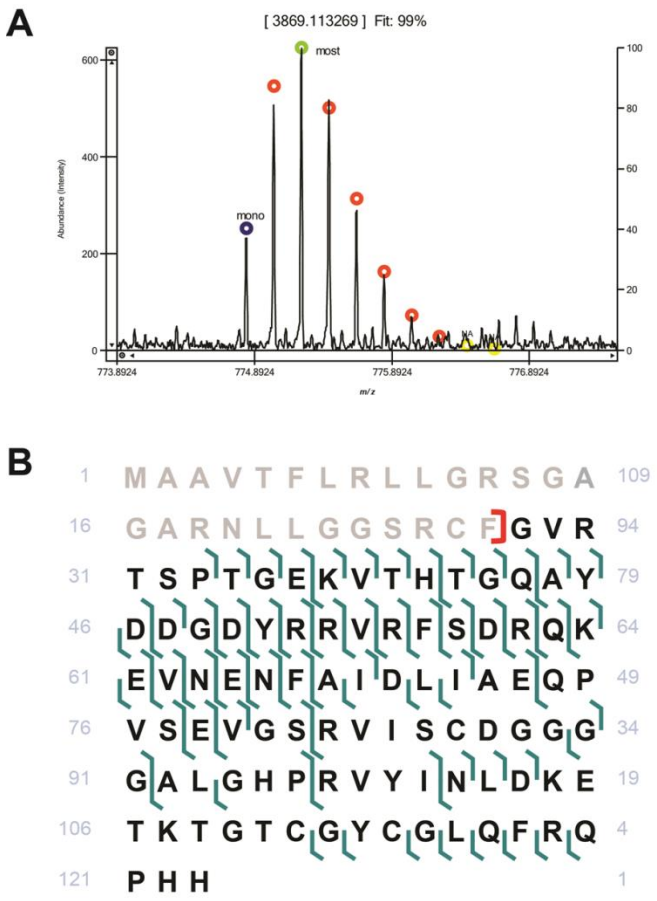


Figure 5.20. Direct graphical output of the spectrum view (A) and sequence table (B). The graphics directly exported from MASH Suite Pro are suitable for publication. Further editing of the graphics in other programs such as Microsoft PowerPoint or Adobe Illustrator is also supported.

5.4 DISCUSSION

Top-down proteomics is increasingly recognized as the premier method for the comprehensive analysis of proteoforms.[24, 147, 155-159, 172, 182, 202, 207, 215, 219, 235, 253, 365-367] Nevertheless, the implementation and practice of the top-down approach still face significant challenges, including solubilizing hydrophobic proteins, separating intact proteins in complex mixtures, detecting large (>50 kDa) and low-abundance proteins, and analyzing highly complex high-resolution mass and tandem mass spectra.[147] Recent advancements, such as the development of

MS-compatible surfactants for solubilizing hydrophobic proteins,[163] multi-dimensional LC separation methodologies for the separation of complex protein mixtures,[167, 184] and functionalized nanoparticles for phosphoprotein enrichment,[373] have begun to address these challenges. Nevertheless, challenges still remain in terms of analyzing highly complex top-down LC-MS and LC-MS/MS data. Herein, we have developed MASH Suite Pro to address the difficulties in analyzing complex spectra arising from LC-MS and LC-MS/MS experiments in large-scale high-resolution MS-based top-down proteomics studies. It is the first comprehensive software package that fully integrates spectral deconvolution, top-down protein identification, quantitation, and characterization, together with visual validation, into a user-friendly customizable interface.

Spectral deconvolution is a critical step in top-down data analysis because the accuracy of the deconvolution results directly affects the downstream data interpretation. MASH Suite Pro currently has accommodated two commonly used algorithms, MS-Deconv and THRASH, for the deconvolution of high-resolution mass and tandem mass spectra. These two algorithms use different methods to evaluate an experimental isotopomer envelope by matching it to the theoretical isotopomer envelope generated based on averagine;[221] THRASH computes a figure of merit value that measures the similarity of peak intensities while MS-Deconv uses a scoring function that involves intensities and *m/z* values of peaks. In addition, these two algorithms employ different approaches for envelop selection: a greedy approach by THRASH [220] and a dynamic programming approach by MS-Deconv.[222] Specifically, THRASH reports all identified masses with a reliability value above a user-specified threshold (referred to as a fit score in MASH Suite Pro), whereas MS-Deconv determines the number of masses reported from a spectrum by estimating the length of the target protein using the precursor mass. Therefore, MS-Deconv is relatively conservative in selecting and reporting isotopomer envelopes. As a consequence, MS-

Deconv usually reports fewer overall masses than THRASH, but also fewer false positives. It is important to note that the statistical evaluation for PrSMs takes into account both the matched and unmatched masses, which offers an accurate estimation of the confidence of protein identification. For large-scale top-down proteomics studies, the majority of identified proteins have e-values between 1E-30 and 1E-10.[167, 184] In this confidence range, MS-Deconv and THRASH yield very similar e-values (usually within 3 orders of magnitude) for the same PrSMs. For tandem mass spectra with relatively high complexity (more product ions), low noise level, and high signal intensity, high confidence in protein identification (e-values < 1E-30) can be achieved using either MS-Deconv or THRASH (**Figure 5.9**). However, under these conditions, MS-Deconv usually results in better (lower) e-values than THRASH due to the fact that more stringent criteria are used for envelope reporting in MS-Deconv (i.e., less false positives will be produced leading to less unmatched masses and an overall increase in the confidence of identification). On the other hand, for tandem mass spectra with relatively low complexity (less product ions), high noise level, and low signal intensity, the identified PrSMs are generally less confident (e-values > 1E-10) when deconvolution is performed using either MS-Deconv or THRASH. Yet, in these cases, THRASH tends to yield better e-values than MS-Deconv due to the fact that more masses will be matched to the sequence (**Figure 5.9**) as a consequence of the lower stringency for envelope reporting. Therefore, as different deconvolution algorithms may perform better for different datasets, the choice of deconvolution method should be decided on a case-by-case basis. It should also be noted that the use of appropriate parameter settings (e.g., S/N), regardless of the algorithm used, is also important for accurate and reliable spectral deconvolution. Besides MS-Deconv and THRASH, additional deconvolution algorithms can also be added to MASH Suite Pro with the permission of the developers. For example, the recently developed UniDec[223] can perform mass and charge

separation of complex spectra, which can also be used for the deconvolution of high-resolution MS data.

Due to the complexity of high-resolution top-down mass and tandem mass spectra, which contain many overlapping isotopomer envelopes, currently available deconvolution methods tend to generate a considerable amount of mis-assigned peaks (**Figure 5.16 and 5.17**). Peak assignment errors can prevent confident identification of the proteins, as well as accurate localization of PTMs and sequence variations. Therefore, visual validation and manual correction of the automatically processed top-down MS data is necessary to reduce false positives and correct mass shift and charge state errors if they occur. There are software packages that provide visual components for comparing the experimental isotopomer envelopes with theoretical isotopomer envelopes generated using averagine.[221] For example, Decon2LS uses a variation of THRASH for high-resolution spectral analysis and visualization.[224] DataAnalysis (Bruker Daltonics) utilizes SNAP2, a THRASH-based algorithm, for spectral deconvolution, and allows for direct visual comparison of the experimental isotopomer envelopes with theoretical envelope patterns (termed SNAP pattern). Nevertheless, these software packages do not allow for manual correction of the computational outputs. In contrast, MASH Suite Pro provides easy access to the real spectra and visualization of each deconvoluted mass, with additional features allowing users to delete false positives and correct the mass values or charge states when mistakes occur. The visualization features and flexibility for data correction make MASH Suite Pro advantageous in analyzing top-down spectra with high accuracy and reliability.

MASH Suite Pro utilizes MS-Align+ algorithm[228] for protein identification. As demonstrated previously,[24, 182, 374, 375] MS-Align+ is highly effective for the identification of proteins and unexpected PTMs from top-down tandem mass spectra. A previous study has shown

that MS-Align+ completes matching 1000 spectra in 18 min versus 22 min using the biomarker search mode of ProSightPC.[228] This corresponds to an improvement of approximately 20% in the speed of protein identification. In addition, the time needed for ProSightPC to complete the search in the advanced search mode (searching against annotated top-down database) was an order of magnitude longer [228]. Moreover, MS-Align+ can identify unexpected PTMs whereas ProsightPC cannot. This demostrated the advantages of MS-Align+ in protein identification. By incorporating MS-Align+, we have developed a user-friendly interface for protein identification in top-down proteomics with high simplicity and speed. Instead of operating protein identification via command lines in the previous MS-Align+, users are able to finish all the necessary steps for protein identification in a single intuitive platform. In addition to MS-Align+, MASH Suite Pro can also accommodate other search algorithms in the future, such as the very recently developed TopPIC (http://proteomics.informatics.iupui.edu/software/toppic/index.html) and MSPathFinder (http://omics.pnl.gov/software/mspathfinder). Similar to MS-Align+, TopPIC identifies proteoforms with unexpected sequence variations and PTMs, and estimates statistical confidence of the identified PrSMs. TopPIC improves on MS-Align+ in terms of search efficiency and computing requirements. MSPathFinder also employs spectral alignment algorithm for top-down protein identification. It requires the users to input the spectrum file, a protein sequence file (.fasta), and a list of potential modifications. MSPathFinder matches the spectra against the user-specified protein sequences containing the defined modifications, and reports the search results as PrSMs with their scores, which indicates the degree of confidence in the identification. When considering only one modification (acetylation), MSPathFinder is faster than MS-Align+. However, MSPathFinder only includes the user-defined modifications, so it is somewhat limited for the

identification of unexpected PTMs. Therefore, MS-Align+ is favored for large-scale proteomics studies and PTM discovery.

MASH Suite Pro enables rapid and accurate relative quantitation of proteoforms in different biological/experimental conditions and, thus, has great potential in the dissection of disease-associated proteoform alterations.[24, 147, 156, 157, 159, 215, 219, 367, 376, 377] MASH Suite Pro allows for the quantitation of different proteoforms with multiple charge states from multiple top-down spectrum data files. The program also automatically generates a quantitation table and chart for a direct comparison of the relative abundances of various proteoforms in different conditions. When disease-related changes in the abundance of proteoforms are observed, MASH Suite Pro can also facilitate the characterization of the proteoforms, including the localization of sequence alterations and PTMs (**Figure 5.15**). This highlights the potential of MASH Suite Pro in unraveling the proteoform-associated disease mechanisms.

Regarding the data format, MASH Suite Pro can directly process and analyze the raw data format (.raw) generated by the instruments of Thermo Scientific without conversion. Furthermore, MASH Suite Pro can also process mzXML files [378], which are an open data format for the storage and exchange of MS data. Proprietary file formats from most vendors can be converted to the open mzXML format, allowing the use of MASH Suite Pro for processing and analyzing data files generated from the instruments of most vendors.

In summary, MASH Suite Pro is a comprehensive, user-friendly and freely available program tailored for large-scale top-down proteomics data analysis, including spectral deconvolution, protein identification, quantitation, and characterization. Distinguished from the previous version for single protein characterization, MASH Suite Pro has implemented important

new features for large-scale proteomic data analysis, including the ability to process LC-MS and LC-MS/MS data, identification of unknown proteins and PTMs, as well as multiplex proteoform quantitation involving various biological/experimental conditions. MASH Suite Pro also allows for visual validation and correction of the computational outputs to ensure the accuracy of data interpretation. With a user-friendly and customizable interface, MASH Suite Pro greatly simplifies and speeds up the interpretation of high-resolution data, and, therefore, will play an integral role in advancing the field of top-down proteomics.

CHAPTER 6

TOP–DOWN PROTEOMICS OF LARGE PROTEINS UP TO 223 KDA ENABLED BY SERIAL SIZE EXCLUSION CHROMATOGRAPHY STRATEGY

This chapter is reprinted from *Analytical Chemistry*, Cai W, Tucholski TM, Chen B, Alphert AJ, McIlwain S, Kohmoto T, Jin S, Ge Y. Top-Down proteomics of large proteins up to 223 kDa enabled by serial size exclusion chromatography strategy, 2017; 89(10): 5467-75,

ABSTRACT

Mass spectrometry (MS)-based top-down proteomics is a powerful method for the comprehensive analysis of proteoforms that arise from genetic variations and post-translational modifications (PTMs). However, top-down MS analysis of high molecular weight (MW) proteins remains challenging mainly due to the exponential decay of signal-to-noise ratio with increasing MW. Size exclusion chromatography (SEC) is a favored method for size-based separation of biomacromolecules, but typically suffers from low resolution. Herein we developed a serial size exclusion chromatography (sSEC) strategy to enable high-resolution size-based fractionation of intact proteins (10-223 kDa) from complex protein mixtures. The sSEC fractions could be further separated by reverse phase chromatography (RPC) coupled online with high-resolution MS. We have shown that 2D sSEC-RPC allowed for the identification of 4044 more unique proteoforms and a 15-fold increase in the detection of proteins above 60 kDa, compared to 1D RPC. Notably, effective sSEC-RPC separation of proteins significantly enhanced the detection of high MW proteins up to 223 kDa, and also revealed low abundance proteoforms that are post-translationally modified. This sSEC method is MS-friendly, robust and reproducible, and thus, can be applied to both high-efficiency protein purification and large-scale proteomics analysis of cell or tissue lysate for enhanced proteome coverage, especially for low abundance and high MW proteoforms.

6.1 INTRODUCTION

Top-down mass spectrometry (MS)-based proteomics is the method of choice for the comprehensive analysis of intact proteins to facilitate the structural and functional characterization of myriad proteoforms[379] (a term encompassing all protein products of a single gene arising from genetic variation, alternative mRNA splicing, and post-translational modifications (PTMs)), and holds great promise for providing novel insights into cellular signal transduction and regulation, as well as disease mechanisms.[147, 154, 155, 207, 244, 360, 380, 381] However, top-down MS analysis of high molecular weight (MW) proteins remains challenging due to the high complexity and wide dynamic range of the human proteome and an exponential decay in the signal-to-noise ratio (S/N) of proteins with increasing MW,[186] especially for those that coelute with low MW proteins. Thus, size-based protein separation is necessary for the detection and characterization of high MW proteins by top-down MS.

The development of multiplexed gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)[168, 202, 382, 383] allowed for the separation of intact proteins based on size, and enabled a deep coverage of the human proteome when coupled with isoelectric focusing and reverse phase chromatography (RPC). In addition, significant progress has been made in the analysis of proteins in the range of 30 – 80 kDa with the use of GELFrEE coupled with capillary zone electrophoresis.[384] Even though the GELFrEE system can achieve high-resolution separation, the use of SDS, an MS-incompatible surfactant, during the size-sorting step necessitates detergent removal procedures such as protein precipitation, which inevitably results in detrimental sample loss, particularly for low abundance and high MW proteins.

Size exclusion chromatography (SEC) is an appealing alternative to the gel-based methods for the fractionation and separation of proteins by size or hydrodynamic volume.[174, 186, 188] SEC has advantages over many other liquid chromatography (LC) methods due to its high compatibility with various solvent systems and reduced sample loss due to minimal interaction between the analytes and the stationary phase. Therefore, SEC is a favored method and has been extensively employed for the analysis of antibody-drug conjugates and the assessment of drug purity.[385, 386] However, SEC has conventionally been considered a low-resolution chromatographic method accompanied by sample dilution [387], and therefore, has not been widely used for the fractionation of highly complex protein mixtures. Although a previous method coupling SEC with RPC for the analysis of cell lysate allowed for the identification of over 370 proteins with low MW (<40 kDa),[388] top-down analysis of high MW proteins (>60 kDa) remains challenging due to the low resolution and separation power of the conventional SEC methods.

Herein, we introduce serial size exclusion chromatography (sSEC) to enable high-resolution size-based separation of intact proteins over a broad MW range. We employed sSEC for the fractionation of a complex protein mixture extracted from the cardiac sarcomeric subproteome consisting of proteins ranging from 10 to 223 kDa, and demonstrated high-resolution separation via the combination of different pore size in series and an increase in effective separation length. A two-dimensional (2D) platform combining sSEC with reverse phase chromatography (RPC) surpasses 1D RPC for the analysis of the sarcomeric protein mixture with 4044 more unique proteoforms identified. Notably, there was a 15-fold increase in the number of high MW proteoforms (> 60 kDa) detected by top-down MS using the 2D method. This sSEC method is MS-friendly, robust and versatile,

and therefore, can be applied to large-scale proteomics analysis of cell or tissue lysate and purification of high MW proteins.

6.2 MATERIALS AND METHODS

6.2.1 Chemicals and reagents

All reagents were purchased from Sigma Aldrich Inc. (St. Louis, MO, USA) unless otherwise noted. HPLC grade H_2O, acetonitrile, ethanol, and ultracentrifugal 10 kDa molecular weight cut-off (MWCO) filters (0.5 mL) were purchased from Fischer Scientific (Fair Lawn, NJ, USA). 10-20% precast Criterion Tris-HCl gels for SDS-PAGE were purchased from Bio-Rad Laboratories (Hercules, CA, USA).

6.2.2 Preparation of Cardiac Protein Extract

Donor heart tissue was obtained from University of Wisconsin Hospital and Clinics according to protocols approved by the Institutional Review Board of University of Wisconsin-Madison as described previously.[389] The donor heart was maintained in cardioplegic solution before dissection. Dissected tissue was snap-frozen immediately in liquid nitrogen and stored at -80 °C for subsequent analysis. All procedures involving sample handling were performed at or below 4 °C at all times to reduce oxidation and degradation of samples. Extraction of the sarcomeric proteins was performed as follows. Cardiac tissue was homogenized using a Polytron homogenizer in 10 vol (mL/g tissue) of wash buffer (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 5 mM $MgCl_2$, 0.5 mM EGTA, 0.1 M NaCl, 1% Triton X-100, 5 mM DTT, 1 mM PMSF, and 1 mM Na_3VO_4 containing protease inhibitor cocktail). Homogenate was centrifuged at 17,000 rcf for 3 min at 4 °C and the supernatant was discarded. The washing step was repeated twice and the

resulting pellet was homogenized in 5 vol (mL/g tissue) of LiCl extraction buffer (5 mM EGTA, 0.1 mM $CaCl_2$, 0.7 M LiCl, 25 mM Tris-HCl (pH 7.5), 5 mM DTT, 1 mM PMSF, 1 mM Na_3VO_4 and protease inhibitor cocktail) and incubated at 4 °C for 10 min to extract sarcomeric subproteome. The homogenate was centrifuged at 17,000 rcf and the supernatant containing the sarcomeric proteome was collected and further centrifuged at 17,000 rcf for 30 min to remove residual pellet. MS-SAFE protease and phosphatase inhibitor cocktail was used in the extraction of sarcomeric proteome to be analyzed by MS and tandem MS (MS/MS).

Prior to 1D RPC and 2D sSEC-RPC analysis, the sarcomeric protein extract was desalted by washing 200 µL of the aliquot through 10 kDa MWCO filters using 0.1% formic acid (FA) in H_2O twice, 1% FA in H_2O four times, followed by 3 final washes with 0.1 % FA in H_2O. The volume of the protein extract recovered after washing was normalized to the starting volume (200 µL) and the sample was collected. Protein concentration after the desalting procedure was approximately 2 mg/mL. Hexafluoroisopropanol was added to the desalted sample to a final concentration of 100 mM and the sample was centrifuged at 16,100 rcf for 30 min at 4 °C prior to 1D RPC-MS and 2D sSEC-RPC-MS analysis.

6.2.3 Serial Size Exclusion Chromatography

SEC and sSEC experiments were performed using an ACQUITY H-class UPLC system (Waters, Milford, MA, USA) equipped with a UV detector and an automatic fraction collector. PolyHYDROXYETHYL A (PolyHEA) columns (single or in series) from PolyLC Inc. (Columbia, MD, USA) were used for all experiments. For SEC with a single column, columns (200 mm x 9.4 mm, 3 µm) with pore sizes 1000 Å, 500 Å, and 300 Å were evaluated individually. For sSEC with two columns (2sSEC), one 1000 Å column

was connected in series with one 500 Å column using capillary tubing (Valco PEEK tubing, 100 um i.d., 2 cm). Similarly, sSEC with three columns (3sSEC) combined one 1000 Å and two 500 Å columns in series (1000 Å - 500 Å - 500 Å). Sarcomeric proteins were separated isocratically at a flow rate of 0.5 mL/min using 1% FA in H_2O. Time-based automatic fraction collection was performed using the Waters Fraction Manager System. Fractions were collected in 1 minute intervals (0.5 mL per fraction) and subsequently concentrated to 100 µL using 10 kDa MWCO centrifugal filters for SDS-PAGE and LC-MS/MS analysis. Tris(2-carboxyethyl)phosphine (TCEP) was added to the sSEC fractions to be analyzed by MS and MS/MS to a final concentration of 1 mM to minimize protein oxidation.

6.2.4 Reverse Phase Chromatography and Top-down MS analysis

5 µL of the desalted LiCl extract or each sSEC fraction was injected into a PLRP column (1000 Å, 250 mm x 0.5 mm, 10 µm) and separated by RPC using an ACQUITY M-class 2D UPLC system (Waters, Milford, MA, USA) configured to bypass the first LC dimension. Sarcomeric proteins were eluted by a gradient of 20% to 95% mobile phase B (Mobile phase A: 0.1% FA in H_2O, mobile phase B was 0.1% FA in 50:50 acetonitrile: ethanol) at a constant flow rate of 8 µL/min.

Proteins eluted were analyzed by a MaXis II quadrupole-time of flight (Q-TOF) mass spectrometer (Bruker Daltonics, Bremen, Germany) via electrospray ionization. The end plate offset and capillary voltage were 500 V and 4500 V, respectively. Nebulizer pressure was set to 0.5 Bar and dry gas flow rate was 4.0 L/min. Mass spectra were taken at a scan rate of 0.5 Hz over 500-2000 *m/z* range.

Targeted MS/MS was performed for Fraction 5 and 7 from the 3sSEC experiment. These representative fractions were chosen based on the prevalence of high MW proteins. The elution time and most abundant charge states of the target proteins were determined from the previous MS runs. Precursor ions of the protein of interest were isolated in the quadrupole and fragmented in the collision cell over the time segment where the protein of interest eluted. MS/MS was performed using collisionally activated dissociation (CAD, 15-20 eV). The same MS/MS runs were repeated with adjusted collisional energy to ensure optimal fragmentation for individual proteins.

6.2.5 Data Analysis, Protein Identification, and Proteoform Profiling

All total ion chromatograms (TIC) were smoothed by the Gauss algorithm with a 3.45 smoothing width. Mass spectra were deconvoluted using the Maximum Entropy algorithm[1] incorporated in the DataAnalysis software. The resolving power was set to 80,000 for proteins that were isotopically resolved online, and 10,000 for proteins that were not isotopically resolved. Tandem mass spectra were exported as .ascii files from the DataAnalysis software and analyzed by the MASH Suite Pro software[2] developed in-house using an S/N threshold of 3 and minimum fit score of 60%. All fragment ions were manually validated and the mass lists were output for protein identification using the MS-Align+ search algorithm (version 0.7.1).[3] The search was performed against the Uniprot human database (updated June 16, 2015). Protein identifications were considered good candidates if p-values were below 0.01. All protein identifications were manually validated following database output using MASH Suite Pro software. Mass tolerance for both precursor ions and fragment ions was set to 10 ppm.

Mass spectra were averaged every 0.5 min from 11 min to 45.5 min in the chromatogram (11-11.5min, 11.5-12, etc.). The resulting averaged mass spectra were treated as "compound mass spectra", and were deconvoluted using the Maximum Entropy algorithm.[1] High-resolution

deconvolution using a resolving power of 80,000 was performed for the proteoforms that were (potentially) isotopically resolved (3-50 kDa) in the 1D analysis and 2D analysis (sSEC Fractions 7-12). The SNAP algorithm was used to generate the mass lists containing the mono-isotopic masses of the proteoforms for the compound mass spectra that were deconvoluted using high-resolution setting. Retention time window for each proteoform was determined based on the time segments in which the species was found. Parameters for the SNAP algorithm intensity threshold set to 0.01 % with an absolute intensity of 500, and the Quality factor and S/N was set to 0.6 and 3, respectively. Low-resolution deconvolution with a resolving power of 10,000 was performed for the proteoforms that were not well isotopically resolved by the instrument (40,000-250,000 Da) for both the 1D analysis and 2D analysis (sSEC Fraction 1-7). For all low-resolution deconvoluted spectra, the Sum Peak algorithm was used to generate the mass lists. The intensity threshold for 1D and 2D analysis using the Sum Peak algorithm was set to 1000 and 800, respectively, based on relative signal intensity. A 0.01% intensity threshold and S/N of 3 were used for the Sum Peak parameters for both 1D and 2D data processing. It is to be noted that the mass range for the high-resolution and low-resolution deconvolution was overlapped for the inclusive counting of the proteoforms.

A mass list of proteoforms found in 1D analysis in mass range 3-50 kDa by high-resolution deconvolution was exported to Microsoft Excel. Similarly, mass lists from 2D (sSEC Fractions 7-12) of proteoforms in mass range 3-50 kDa (high-resolution deconvolution method) were generated for 2D analysis. Molecular mass ($M + H^+$) for these lists were rounded to the nearest integer to account for cases in which proteoforms were double counted due to the difference in the mass determination using the SNAP algorithm. Following rounding, the mass lists of 2D Fractions 7–12 were combined for a list of total proteoforms in the mass range of 3-50 kDa.

The 1D and 2D mass lists described above were further processed in Microsoft Excel to remove redundant proteoforms. There were two cases for having redundant proteoforms within the mass lists. In the first case, a proteoform eluted over a minute and was counted in two adjacent 0.5 min retention windows. To account for these cases, "Remove Duplicates" function was used for the mass lists generated from 1D and 2D data processing. In the second case, a proteoform could elute in two or more sSEC fractions in the same retention time window (only applicable to the 2D analysis). In order to account for these redundant proteoforms, the mass lists from 2D analysis were combined and "Remove Duplicates" function was applied to remove redundancies between different sSEC fractions. Absolute intensity cutoffs of 1,000 and 500 were applied for 1D and 2D mass lists, respectively, due to a relative low intensity (2-fold reduction) of the 2D data compared to 1D.

Manual processing of the 1D and 2D compound mass spectra for Fractions 1 – 7 was performed to generate a list of proteoforms greater than 50 kDa. First, the mass spectra of the 2D analysis (sSEC Fraction 1-7) were manually analyzed to identify proteoforms > 50 kDa, and determine their retention time. Based on the proteoforms > 50 kDa found in the 2D analysis and their retention time, mass spectra of the 1D analysis were analyzed for visible charge states of the proteoforms detected in 2D analysis. Cutoffs for including the potential high MW proteins found in 1D were determined based on having higher than 0.01% intensity relative to the m/z peak of the highest intensity, within the same compound mass spectra.

For a comparison of proteoforms detected in 1D and 2D analysis, non-redundant mass lists containing proteins of 3–50 kDa were combined for a total count of proteoforms detected between the two. Duplicates were removed from this combined list to reveal the number of species detected by both 1D and 2D analysis. The proteoforms > 50 kDa that were found in both 1D and 2D analysis

was determined manually. To determine the total proteoform count for 1D and 2D analysis, the mass lists containing proteoforms in the mass range 3–50 kDa were combined with the lists of those > 50 kDa that were generated manually.

6.3 RESULTS AND DISCUSSION

6.3.1 High-resolution Size-based Separation of Intact Proteins by sSEC

Size-based protein separation is critical for the detection and characterization of high MW proteins by top-down MS. Conventional SEC methods suffer from low resolution due to significant band broadening primarily as a consequence of slow mass transfer in the packing pores, which limits the number of baseline-separated compounds.[390, 391] Consequently, SEC has been primarily used to analyze simple mixtures of biomacromolecules. Therefore, we sought to develop an SEC method for the high-resolution separation of proteins from a complex mixture. The separation efficiency was evaluated by SDS-PAGE analysis according to the number of proteins that coeluted within a defined chromatographic interval (1 min), as well as the MW range of all the proteins that coeluted in the same fraction. Specifically, a smaller number of coeluting proteins within the same chromatographic interval, and a narrower MW range of all the proteins that coeluted in a same fraction, were indicative of better separation.

To enable size-based separation of individual proteins rather than protein complexes, we employed a denaturing condition by using 1% FA (pH 1.8) as the mobile phase to facilitate protein unfolding and complex dissociation. We first evaluated the

fractionation range of individual SEC columns with different pore size (1000 Å, 500 Å, or 300 Å) under the denaturing condition (**Figure 6.1**). The SEC column with 500 Å pore size provided better separation than 300 Å, with less proteins coeluted in each fraction within every 1 min chromatographic interval, and a narrower MW range for the proteins eluted in each fraction (**Figure 6.1A, B**). Moreover, an SEC column with 1000 Å pore size provided significantly better separation of the high MW proteins (> 60 kDa) from the intermediate (35-60 kDa) MW ones, with predominantly high MW proteins present in the earliest four fractions (**Figure 6.1B, C**).

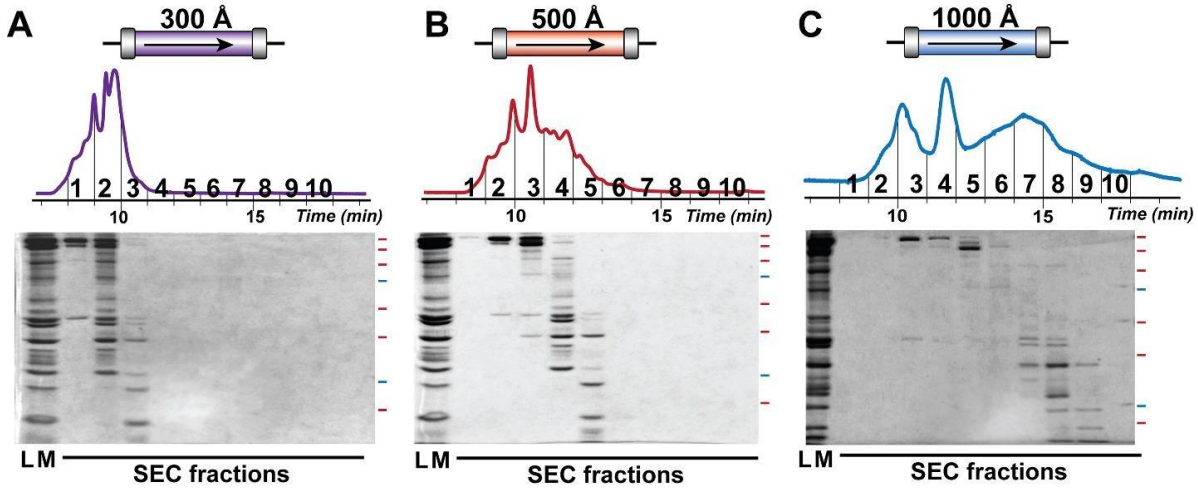


Figure 6.1: Comparison of SEC with pore sizes (A) 300 Å, (B) 500 Å, and (C) 1000 Å for the fractionation of sarcomeric protein extract. Top panel shows representative UV chromatograms and bottom panel shows corresponding SDS-PAGE analysis of fractions collected and pooled from two technical replicates (12.5 %).

To achieve efficient size-based separation over a broad MW range, we explored serial SEC (sSEC) by combining two SEC columns with 1000 Å and 500 Å pore size (termed 2sSEC). This would allow for better separation of high MW protein from intermediate MW ones, as well as the intermediate from low MW proteins. In comparison

with SEC with a single column of 500 Å pore size, the addition of a column with 1000 Å pore size resulted in significant improvement in the separation of the high MW proteins (> 60 kDa) from the intermediate MW ones (35 - 60 kDa) (**Figure 6.2A, B**). In addition, the overall separation improved with 2sSEC compared to SEC, as demonstrated by the smaller number of coeluting proteins present in the Fraction 6-8 of 2sSEC versus Fraction 2-5 of SEC (**Figure 6.2A, B**). This likely resulted from the increased path length with 2 columns ($R \propto \sqrt{L}$), which offers improved resolution. The fractionation method proved to be highly reproducible, with little variation in each fraction collected in three technical replicates (**Figure 6.3**).

To further exploit the benefit provided by increased path length, we combined an additional column of 500 Å pore size with the 1000 Å-500 Å series (3sSEC) to test the separation of complex protein mixtures. Proteins eluted later in 3sSEC with majority of the proteins eluting between 29-41 min versus 18-28 min in 2sSEC. However, the interval time required for most of the proteins to elute from the column was similar (11 min in 3sSEC vs. 9 min in 2sSEC) (**Figure 6.2B, C**). Furthermore, 3sSEC demonstrated significantly improved resolution over 2sSEC in the separation of proteins in the intermediate to low MW range (< 60 kDa) (**Figure 6.2B, C**).

Our study agrees with previous research, which demonstrated that MW range and selectivity of an SEC column are highly dependent on the pore size of the column packing materials.[392] Importantly, we have achieved high-resolution separation of a complex protein mixture with a large MW range (10-220 kDa) by combining columns with different pore size and increasing the effective separation length (**Figure 6.2**). This method is robust,

highly reproducible (**Figure 6.3**), and compatible with a wide variety of solvents, and therefore, can be generally applied for the separation and fractionation of protein mixtures.

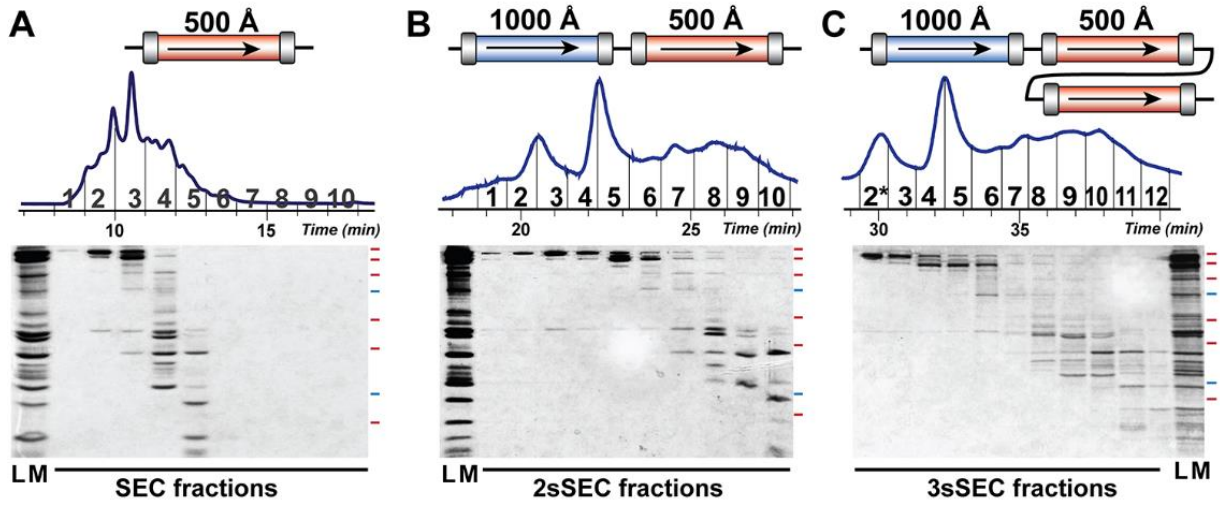


Figure 2.2: **High-resolution sSEC separation of complex protein mixture**. Comparison of (A) SEC (500 Å), (B) sSEC 2sSEC (1000 Å – 500 Å), and (C) 3sSEC (1000 Å – 500 Å – 500 Å) for the fractionation of sarcomeric protein extract. Top panel: representative UV chromatogram of each experiment with the collected fractions (1-10 for SEC and 2sSEC, 2*-12 for 3sSEC) annotated. Bottom panel: SDS-PAGE analysis of the corresponding SEC, 2sSEC or 3sSEC fractions collected and pooled from two technical replicates. Red and blue marks to the right of each gel represent molecular weight markers (250, 150, 100, 75, 50, 37, 25 kDa from top to bottom). LM, loading mixture prior to SEC/sSEC fractionation. The lane corresponding to Fraction 2* in (C) represents visualization of equal volumes of both Fraction 1 and 2, which were combined for RPC-MS analysis due to their similarity in the protein contents.

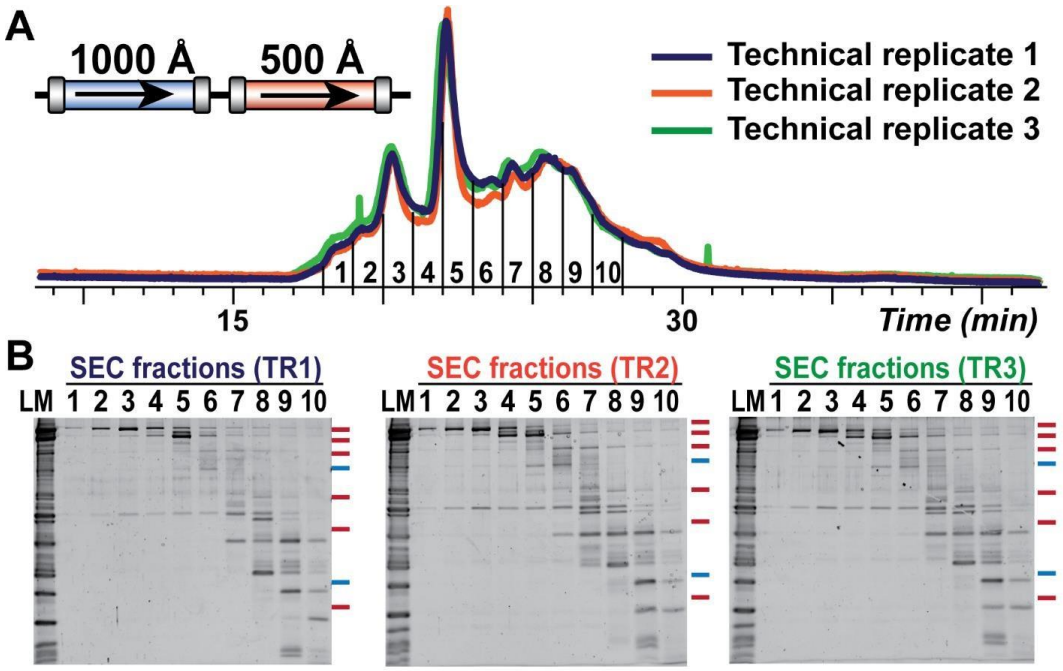


Figure 6.3: Three technical replicates of sSEC experiment demonstrated highly reproducible separation. (A). UV chromatograms for three 100 µL (200 µg total protein) injection of the sarcomeric protein extract for sSEC 1000 – 500 Å column series. (B). SDS-PAGE analysis of sSEC fractions collected from individual runs. TR: technical replicate. Red and blue marks to the right of each gel represent molecular weight markers.

6.3.2 Comparison of 2D sSEC-RPC with 1D RPC for MS Analysis of Intact Proteins in a Complex Mixture

Next, we analyzed the sarcomeric protein extract before and after sSEC fractionation by high-resolution top-down MS. 1D RPC-MS is adequate for detecting the most abundant sarcomeric proteins and their proteoforms, nearly all of which are below 50 kDa (**Figure 6.4A, B**). Cardiac troponin T (cTnT), cardiac troponin I (cTnI), myosin light chain 1 and 2 (MLC1 and MLC2), actin, and cardiac troponin C (cTnC) are the major constituents of the cardiac sarcomere, and can be detected by 1D RPC-MS (**Figure 6.4A**) as reported previously.[393] Many of these proteins are post-translationally modified as shown in **Figure 6.4B**.

Using the abundant sarcomeric proteins (cTnT, cTnI, MLC1, MLC2, actin and cTnC) as references, we evaluated the efficiency of sSEC fractionation. As shown in **Figure 6.4C**, the most abundant sarcomeric proteins were segregated by size into different sSEC fractions. Actin (41.8 kDa) and cTnT (34.6 kDa) were found primarily in sSEC Fractions 7-9, while the cTnI (24 kDa) was found in Fractions 9-10, and the smaller sarcomeric proteins (MLC1, MLC2 and cTnC) were found in the last fractions (Fractions 11-12). Importantly, sSEC fractionation did not alter the abundance of the post-translationally modified proteoforms relative to the unmodified ones (**Figure 6.5**), indicating that relative quantification of proteoforms is reliable with sSEC fractionation prior to RPC-MS analysis.

While 1D RPC is sufficient for MS analysis of the most abundant sarcomeric proteins and their corresponding proteoforms, it failed to detect many proteins larger than 50 kDa that were present in the mixture (SDS-PAGE analysis in **Figure 6.2**). The 2225 proteoforms found in the 1D RPC-MS analysis were primarily low MW proteins ranging between 10 and 25 kDa in size. 2D sSEC-RPC-MS analysis outperformed the 1D method, and detected a total of 5360 proteoforms (with 1316 proteoforms in common). More importantly, 47 proteoforms above 60 kDa were detected in the 2D platform compared to only 3 proteoforms in the 1D analysis, representing a 15-fold increase in the number of high MW proteoforms detected (**Figure 2.4D**). Overall, the 2D method surpassed 1D analysis by detecting 4044 more unique proteoforms that were not found in the 1D method. (**Figure 2.4D**).

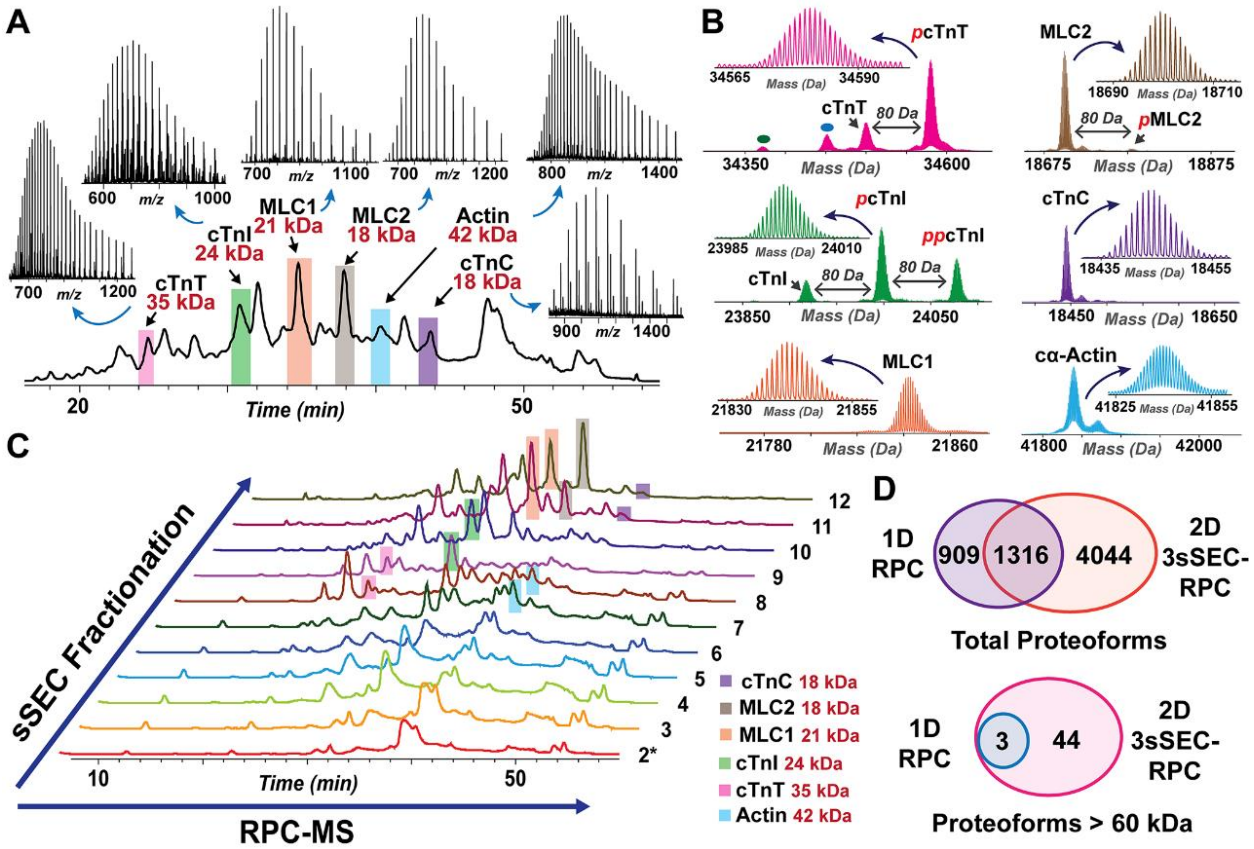


Figure 2.4: Comparison of 1D RPC-MS and 2D 3sSEC-RPC-MS for the top-down analysis of a complex protein mixture. (A) Total ion chromatogram (TIC) of 1D RPC-MS analysis for 10 µg sarcomeric protein extract. The most abundant sarcomeric proteins are highlighted (pink, cTnT; green, cTnI; orange, MLC1; brown, MLC2; blue, actin; purple, cTnC) and their raw mass spectra are shown. The MW of each species is indicated in red font. (B) High-resolution deconvoluted mass spectra of the sarcomeric proteins highlighted in 1D RPC-MS TIC. The major proteoforms of each protein species are annotated. Red italic p and pp represents mono-phosphorylation and bis-phosphorylation, respectively. (C) TICs of 2D sSEC-RPC-MS analysis coupling sSEC fractionation with RPC-MS. Highlighted peaks correspond to the peaks of the abundant sarcomeric proteins highlighted in the 1D TIC. (Note: The highlights indicate the fractions in which the proteins are present most abundantly, but do not indicate that the species is exclusively present in the highlighted fractions.) (D) Venn diagrams showing the numbers of total proteoforms and proteoforms with MW > 60 kDa detected in 1D versus 2D analysis.

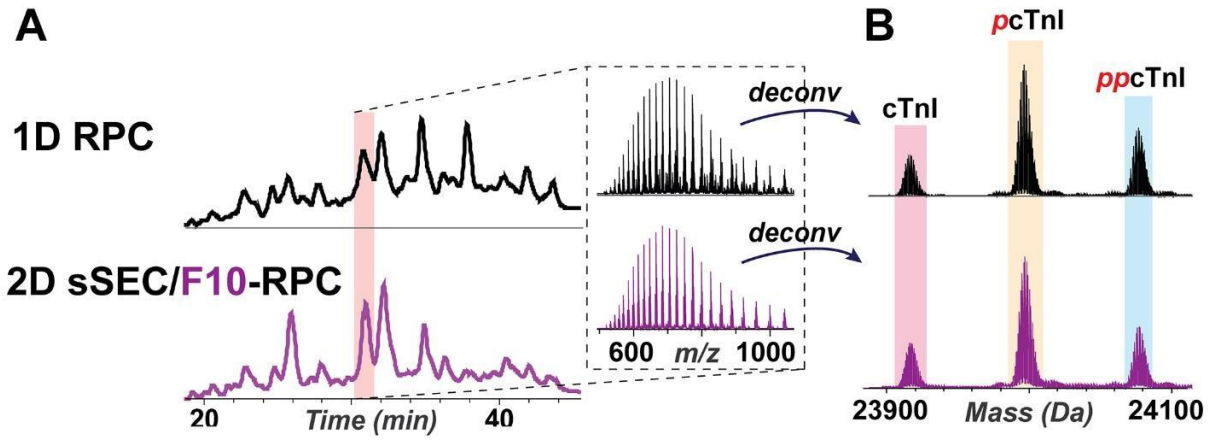


Figure 6.5: sSEC prior to RPC-MS analysis did not alter relative abundance of proteoforms of cTnI: (A). TIC of 1D RPC-MS analysis (black trace) of the whole sarcomeric protein extract aligned with TIC of RPC-MS (purple trace) analysis of sSEC Fraction 10. Highlighted region (30.5 – 31 min) corresponds to the elution of cTnI. Top-down MS of cTnI for 1D and 2D are shown. (B). Deconvoluted mass spectra of cTnI proteoforms showing similar relative abundance between 1D and 2D analysis.

These results are consistent with previous studies, wherein multiple orthogonal LC separation methods can increase the number of proteins and proteoforms identified.[386, 388, 394-398] Multi-dimensional protein separation methods have been reported to expand the coverage of the complex proteomes.[381, 399, 400] 2D LC strategies coupling ion-exchange chromatography (IEC), hydrophilic interaction chromatography, or hydrophobic interaction chromatography with RPC have allowed for top-down MS analysis of complex cell lysates and detection of hundreds of proteins.[394, 396, 401] Furthermore, using a three-dimensional (3D) LC strategy, our group has reported the identification of 640 proteins in a single IEC fraction (out of 35), demonstrating the capability of multi-dimensional LC separation for top-down MS analysis of complex protein mixtures.[397] Kelleher and co-workers utilized isoelectric focusing and GELFrEE separation, followed by RPC-MS analysis in a large-scale top-down study of human proteome, and were able to identify 1220 proteins and over 5000 proteoforms, representing the largest top-down study of the human proteome to date.[180] While these studies demonstrate the power of multi-dimensional

separation platforms for the deep profiling of proteomes by top-down MS, none of these works achieved top-down MS analysis of proteins larger than 80 kDa.[388, 395, 398] Zhang et al. employed isoelectric focusing and superficially porous silica LC for the separation of mouse cardiac tissue extract, and were able to detect myosin heavy chain variants.[402] However, the coverage of the cardiac proteome was limited. The sSEC method developed in this study enabled robust and high-resolution size-based separation of proteins over a broad MW range. sSEC coupled with RPC-MS using a Maxis II Q-TOF mass spectrometer significantly facilitated high-resolution top-down MS analysis of complex protein mixtures.

6.3.3 sSEC-RPC Enabled Top-down MS Analysis of High MW Proteins (up to 223 kDa)

With an exponential decay of S/N for proteins with an increase in MW,[186] signal interference from coeluting low MW impedes the detection of large proteins. Hence, many LC-MS platforms are insufficient for detecting high MW proteins (especially those above 60 kDa) and the corresponding proteoforms in a complex protein mixture. Moreover, isolation and MS/MS fragmentation of ions from high MW proteins is challenging due to the extremely low S/N. Size-based fractionation of complex protein mixtures prior to RPC-MS analysis not only reduces the complexity of the sample, but also decreases ion suppression against high MW species, enabling top-down MS analysis of high MW proteins and their proteoforms.

As shown in **Figure 6.6**, we were able to detect high MW proteins up to 223.1 kDa in the early sSEC fractions (Fraction 1-6). sSEC Fraction 1 and 2 primarily contain a high

MW protein (~ 220 kDa) as shown in the SDS polyacrylamide gel (**Figure 6.2C**), and this protein was detected in the subsequent top-down MS analysis (**Figure 6.6**). In addition, other high MW proteins, including a 140.8 kDa, 116.4 kDa, 93.3 kDa, and 80.9 kDa proteins among others, were also found in the sSEC fractions (**Figure 6.6**).

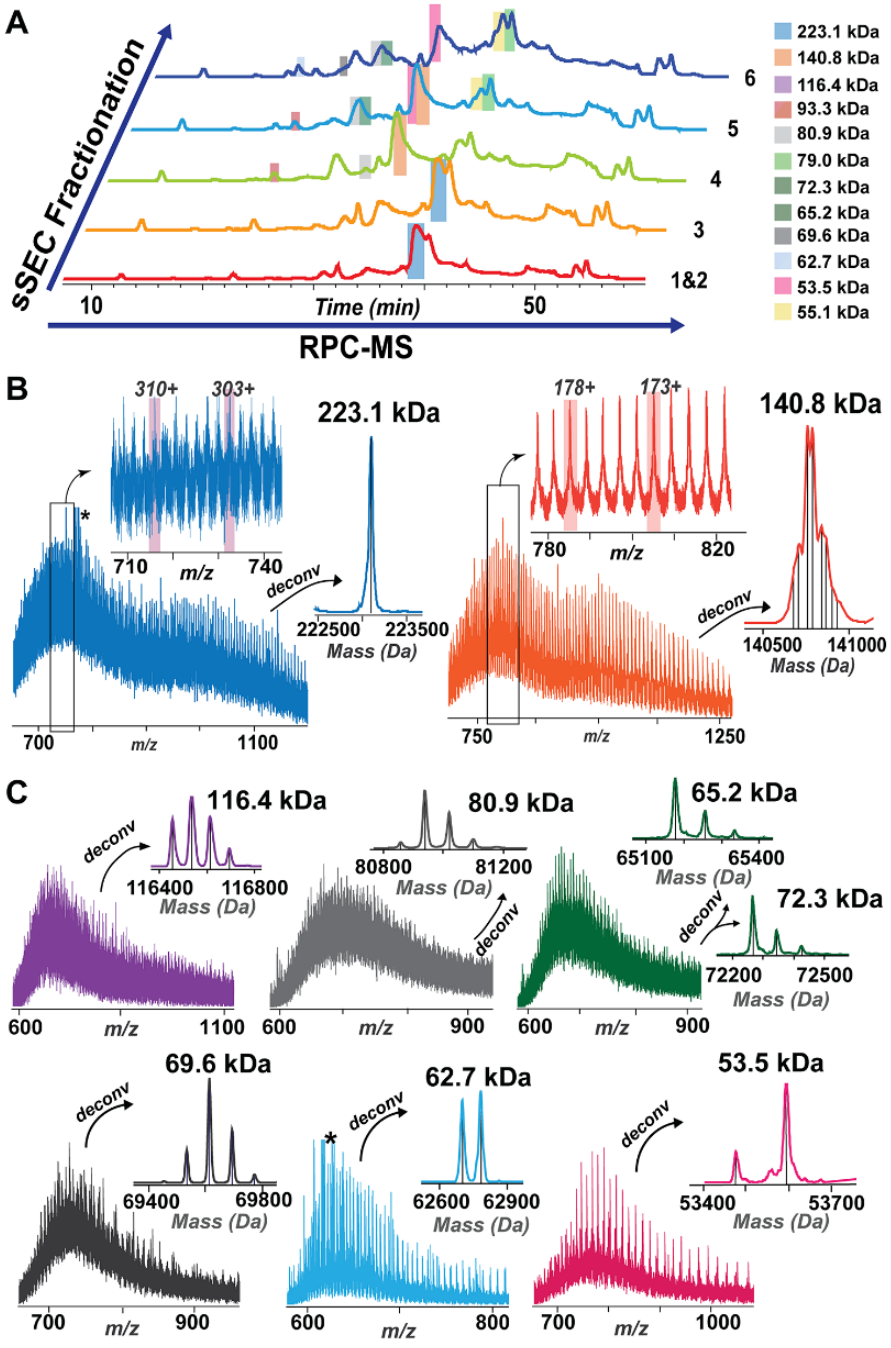


Figure 6.6: **sSEC fractionation enabled detection of high MW proteins**. (A) TICs of RPC-MS for sSEC Fractions 1-6. Highlighted regions of the TICs represent retention windows for the corresponding high MW

proteins. (B) Top-down mass spectra for a 223.1 kDa and a 140.8 kDa with zoom-in views of the charge states and the corresponding deconvoluted spectra. The deconvoluted spectrum of the 140.8 kDa protein shows multiple proteoforms. (C) Top-down mass spectra and the deconvoluted spectra of proteins with MW 116.4 kDa, 80.9 kDa, 65.2 kDa, 72.3 kDa, 69.6 kDa, 62.7 kDa and 53.5 kDa.

Coeluting low MW proteins were segregated into the later sSEC fractions, permitting the detection of large proteins in the earlier fractions. For example, in the 1D RPC-MS analysis, a 116.4 kDa and a 47.6 kDa protein coeluted with very low MW proteins (e.g. 13.4 kDa and 7.8 kDa), which dominated the mass spectrum (**Figure 6.7**). By separating these by size, sSEC permitted the detection of the 116.4 and 47.6 kDa proteins in the sSEC Fraction 5 and 6, respectively (**Figure 6.7**). Deconvolution of the mass spectrum revealed 4 proteoforms of the 116.4 kDa protein, each with an 80 Da mass shift from the adjacent proteoform, suggesting that the 116.4 kDa protein is multiply-phosphorylated (**Figure 6.7B**). In addition, though the 47.6 kDa protein could be detected in the 1D RPC-MS analysis, the S/N was too low to efficiently isolate the precursor ions for MS/MS experiments. Reduction of the sample complexity enabled enrichment of the 47.6 kDa protein in Fraction 6 (**Figure 6.7B**), permitting the characterization of this proteins by MS/MS. It is worth noting that the 116.4 kDa and the 47.6 kDa protein were found in adjacent sSEC fractions, confirming that sSEC separation is necessary and effective for the separation of high MW proteins from intermediate MW ones. In addition to the separation of high MW proteins from intermediate and low MW species (**Figure 6.7**), intermediate MW proteins were also fractionated from the smaller ones (**Figure 6.8**). For example, a 47.5 kDa and 50.9 kDa proteins normally coeluted with the abundant sarcomeric protein (MLC1) in the 1D RPC-MS analysis (**Figure 6.8**). sSEC allowed for

the separation of the 47.5 kDa and 50.9 kDa proteins from MLC1 to facilitate top-down MS and MS/MS analysis of these intermediate MW proteins.

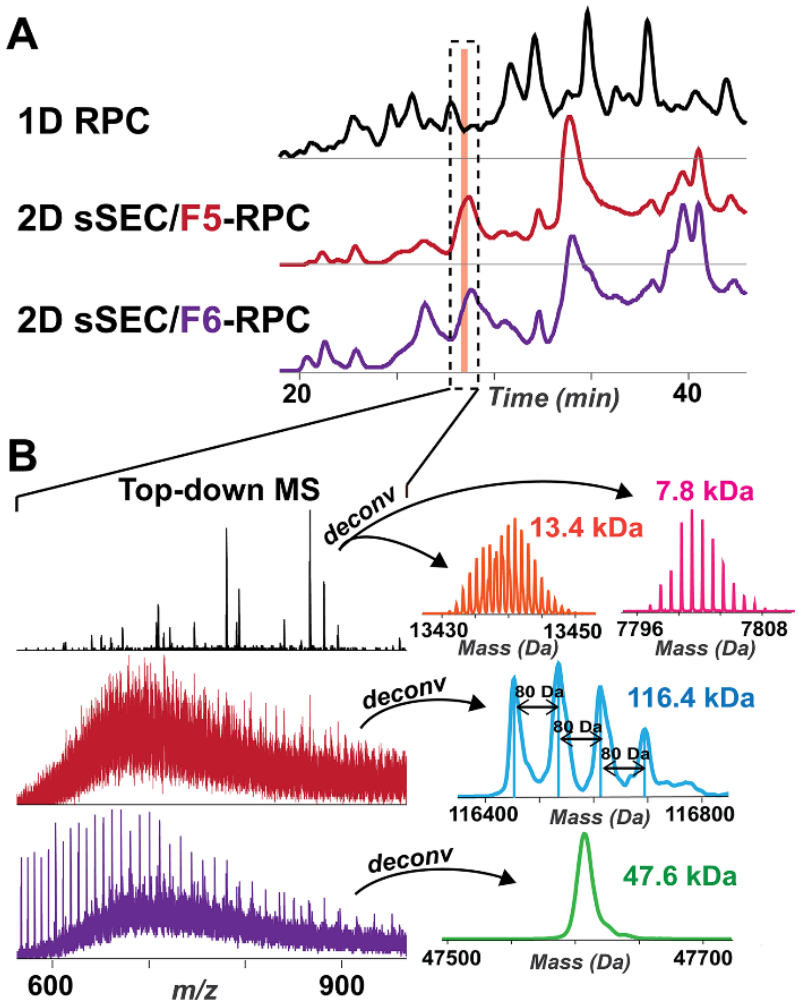


Figure 6.7: **Coeluted high MW and low MW proteins were separated by sSEC, permitting top-down MS analysis**. (A) TIC of 1D RPC-MS (black trace) analysis for the whole sarcomeric protein mixture aligned with TICs of RPC-MS analysis for sSEC Fractions 5 and 6 (red and purple traces, respectively) of the same protein extract. (B) The corresponding top-down mass spectra of the proteins eluted between 28.0 and 28.5 min. Protein species with MW 116.4 kDa and 47.6 kDa were revealed in sSEC Fractions 5 and 6, respectively, which regularly coeluted with smaller protein species (13.4 kDa, 7.8 kDa) in 1D RPC-MS analysis and remained undetected. To the right of the top-down mass spectra shows the corresponding deconvoluted spectra of the major proteins detected in 1D RPC-MS analysis (13.4 kDa and 7.8 kDa proteins), and RPC-MS analysis of sSEC Fractions 5 and 6 (116.4 kDa and 47.6 kDa protein, respectively). Deconvoluted mass spectrum of the 116.4 kDa proteins (resolving power 10000) revealed complex proteoforms each with 80 Da mass shifts.

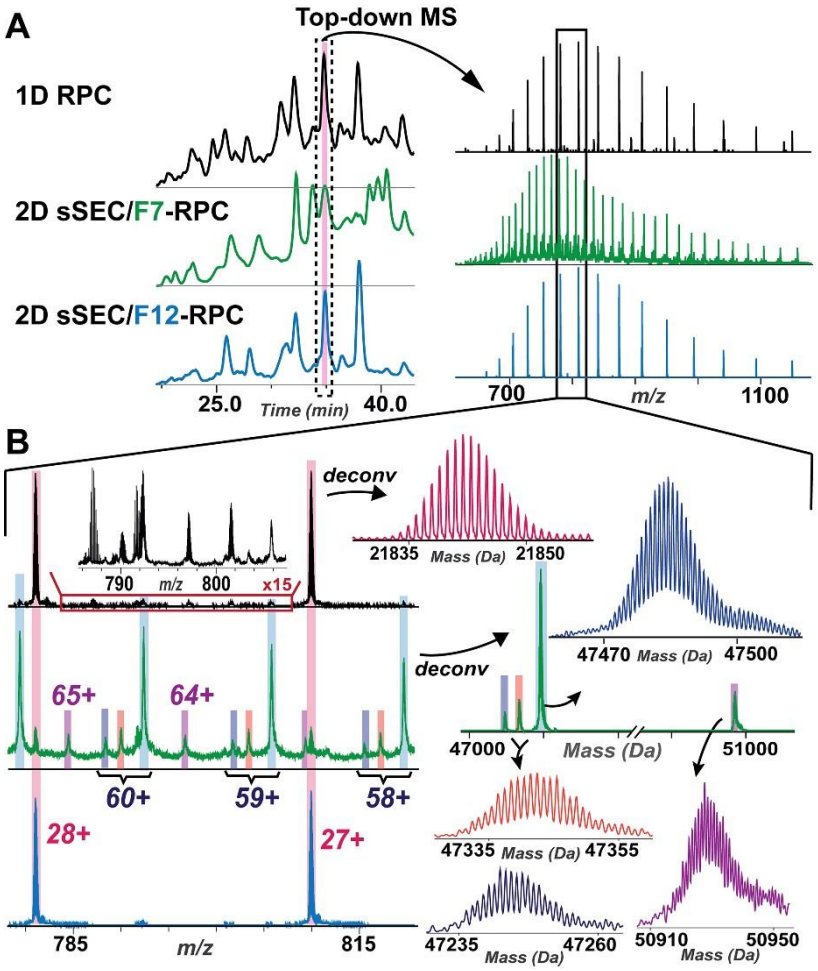


Figure 6.8: **3sSEC fractionation separates intermediate MW proteins, which were otherwise suppressed by coeluted low MW proteins**. (A) TIC of 1D RPC-MS (black trace) analysis for the whole sarcomeric protein mixture aligned with TICs of RPC-MS analysis for sSEC Fractions 7 and 12 (green and blue traces, respectively) of the sarcomeric protein extract. Proteins with MW 47.4, 47.3, 47.2, and 50.9 kDa were revealed in Fraction 7, which coeluted with abundant sarcomeric protein MLC1 (21.8 kDa) within retention time window 34.5 – 35.2 min in the 1D analysis. (B) Zoom-in views of the mass spectra of the proteins found in 1D or 2D analysis showing ions from the MLC1 (pink highlight), 47.4 kDa (indigo highlight), 47.3 kDa (orange highlight), 47.2 kDa (purple highlight), and 50.9 kDa (violet highlight). Deconvoluted mass spectrum of the each proteoform is shown to the right of the raw spectrum.

6.3.4 sSEC-RPC Enabled Detection of Low Abundance Protein PTMs

sSEC fractionation prior to RPC-MS analysis facilitated the detection of low abundance protein PTMs (**Figure 6.9 and Figure 6.10**). In the 1D RPC-MS analysis, a 65.2 kDa protein was found coeluted with other low MW proteins (**Figure 6.9**). Although

certain charge states of the 65.2 kDa protein can be found in the zoom-in view of the mass spectrum, the S/N of this protein was extremely low (**Figure 6.9B**), and therefore, the post-translationally modified forms of the 65.2 kDa protein remained undetected in the 1D RPC-MS analysis due to their low abundance. Detection of the post-translationally modified forms of the 65.2 kDa protein was made possible with sSEC fractionation. The charge states of the 65.2 kDa proteoforms were easily recognizable in the RPC-MS analysis of sSEC Fraction 5 (**Figure 6.9B**). Protein deconvolution showed two additional proteoforms of the 65.2 kDa protein, each with 80 Da mass shifts. In addition, a 72.3 kDa protein, previously undetected in the 1D analysis, coeluted with the 65.2 kDa protein in the sSEC-RPC-MS analysis (**Figure 6.9**). Importantly, sSEC allowed for not only the detection of the 72.3 kDa protein by top-down MS, but also its post-translationally modified forms (**Figure 6.9B**). Similarly, the post-translationally modified form of a 79.0 kDa protein, which remained undetected in the 1D analysis, was revealed after sSEC fractionation (**Figure 6.10A**).

Protein PTMs represent important mechanisms in the regulation of protein activity and function, and play critical roles in signal transduction in health and diseases.[147-149, 244] Despite the prevalence of protein PTMs, the abundance of modified proteins relative to unmodified ones is usually low, hindering the detection and comprehensive characterization of post-translationally modified proteoforms. Many groups sought to develop effective methods for the enrichment of proteoforms with a particular PTM, such as phosphorylation,[373, 403, 404] to facilitate the structural and functional characterization of post-translationally modified proteoforms. Here we have shown that high-resolution sSEC fractionation aids in reducing sample complexity, permitting the analysis of low abundance protein PTMs by top-down MS. More importantly, this sSEC separation method does not

alter the relative abundance of the related proteoforms (**Figure 6.5**), unlike the enrichment approach, and therefore, enables reliable quantification.

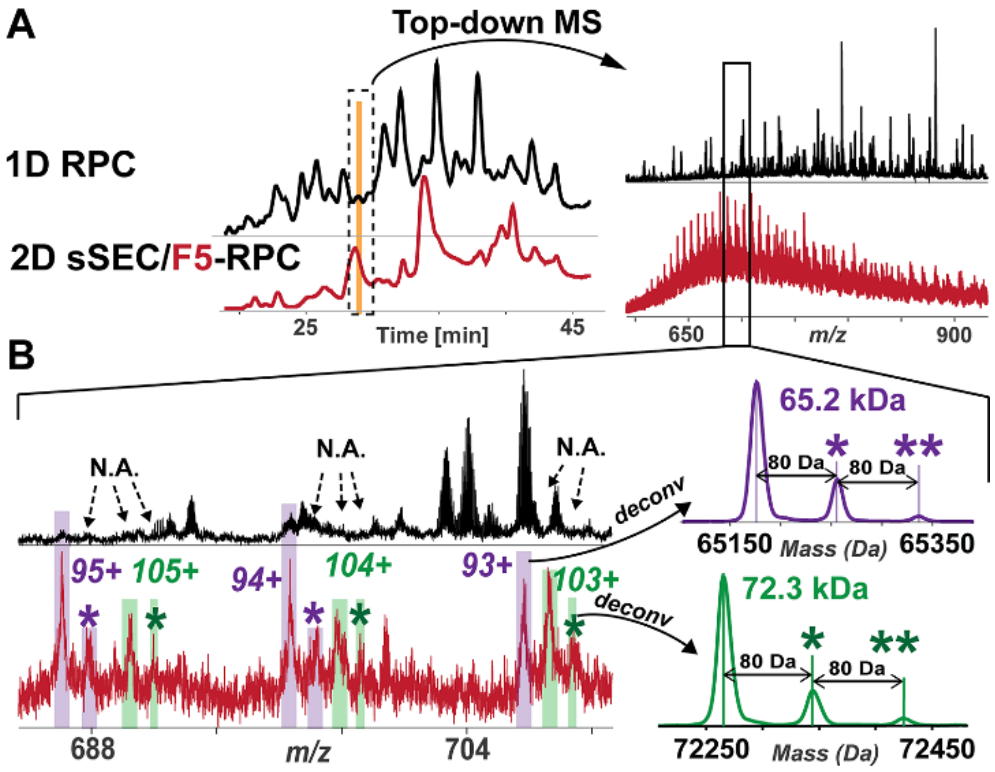


Figure 6.9: 3sSEC fractionation enabled top-down MS analysis of low abundance protein PTMs. (A) TIC of 1D RPC-MS (black trace) analysis for the whole sarcomeric protein mixture aligned with TICs of RPC-MS analysis for sSEC Fraction 5 of the same protein extract. The corresponding top-down mass spectra of the proteins eluted between 28.5 and 29 min are shown at the right panel. (B) Zoom-in views of the top-down mass spectra and the corresponding deconvoluted spectra of the high MW proteins detected in sSEC Fraction 5. The charge states of the 65.2 kDa protein were detected in the 1D analysis. However, the low abundance PTMs of the 65.2 kDa was not revealed. sSEC allowed for the detection of the low abundance post-translationally modified form of the 65.2 kDa, as well as another protein that is 72.3 kDa. Deconvoluted mass spectrum of the 65.2 kDa and 72.3 kDa proteins (resolving power 10000) revealed low-abundant PTMs each with 80 Da shift. * denotes mono-phosphorylated proteoform; ** denotes bis-phosphorylated proteoform.

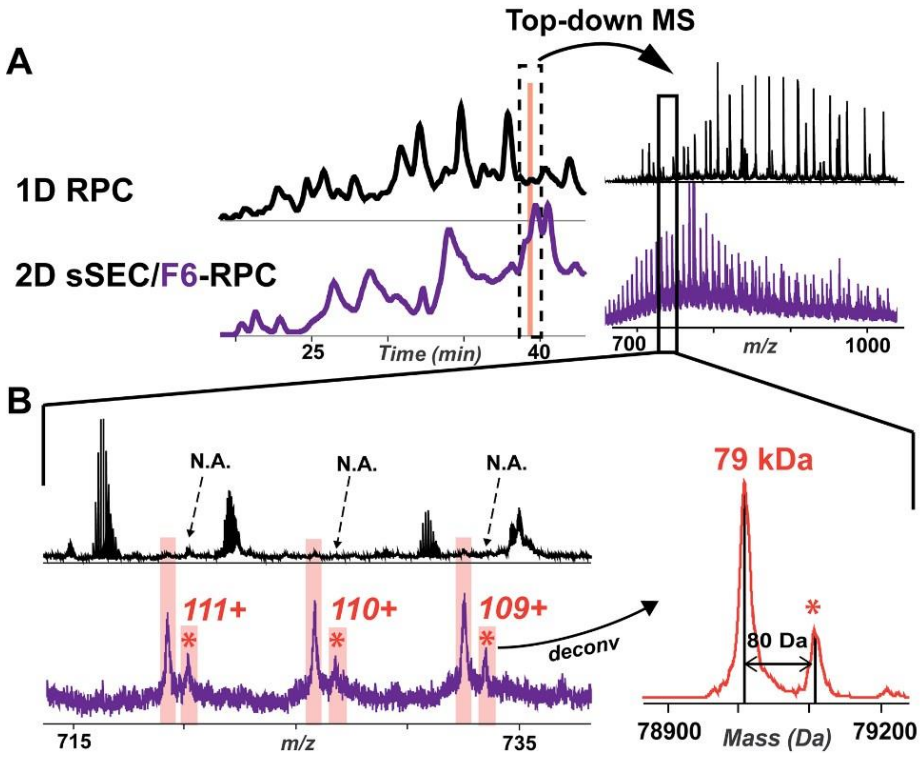


Figure 6.10: **sSEC-RPC-MS reveals low abundant protein PTMs**. (A) TIC of 1D RPC-MS (black trace) analysis for the whole sarcomeric protein mixture aligned with TICs of RPC-MS analysis for sSEC Fraction 6 (purple trace) of the sarcomeric protein extract. The corresponding top-down mass spectra of the proteins eluted are shown at the right panel. A protein with MW 79 kDa was revealed in sSEC Fraction 6, which regularly coeluted with smaller proteins the in 1D RPC-MS analysis and remained undetected. (B) Zoom-in views of the top-down mass spectra and the corresponding deconvoluted spectra of the 79 kDa protein detected in RPC-MS analysis of sSEC Fraction 6 (2D sSEC/F6-RPC). Deconvoluted mass spectrum (resolving power 10000) revealed a proteoform with 80 Da mass shift from the major peak, indicating phosphorylation of the 79 kDa protein.

6.3.5 Online Targeted High-resolution MS/MS for Protein Identification

Detection of charge state envelopes of large proteins was made possible with effective sSEC fractionation prior to RPC-MS. This enabled efficient isolation of precursor ions of the high MW proteins for MS/MS fragmentation. Targeted MS/MS with online CAD fragmentation was performed on high and intermediate MW proteins in sSEC Fraction 5 and 7 (**Figure 6.11, 6.12 and 6.13**). Through the analysis of the first LC-MS run (Run 1), we were able to determine the charge state ions, molecular weight, and retention time of each target protein (**Figure 6.11A, B**).

To identify target proteins, we performed a second LC-MS run (Run 2) with pre-set time segments to isolate precursor ions of the proteins of interest for fragmentation (**Figure 6.11A**). Online CAD was performed in multiple time segments, with corresponding TIC of the tandem mass spectra shown in **Figure 6.11A** (purple trace). In the remaining time segments where no proteins of interest eluted, MS data were collected (orange trace, **Figure 6.11A**).

A 42.9 kDa protein was detected in the first LC-MS run (Run 1) in the 33.5-34.5 min elution window. The absence of coeluting proteins allowed multiple charge states of the 42.9 kDa protein from 700-800 *m/z* to be isolated for CAD fragmentation in the second LC-MS/MS run (Run 2) within the time segment set between 33.5 and 34.5 min. High-resolution MS/MS allowed for accurate determination of the fragment ion charge states in the complex tandem mass spectrum (**Figure 6.11B**). As shown in **Figure 6.11B**, ions with charge state 35+ and 29+ were resolved close to the baseline. The high resolution provided by the Maxis II Q-TOF mass spectrometer enabled determination of the monoisotopic masses of the fragment ions with high accuracy. MASH Suite Pro[405] equipped with the MS-Align+ algorithm [228] was employed for the manual validation of the mass list, and the identification of the target protein. We achieved confident identification of the 42.9 kDa protein as creatine kinase, M-type (CKM, Uniprot Human P06732), with a *p*-value of 2.4 E-24, which indicates high confidence in the identification result (**Figure 6.11C**). Importantly, in the tandem mass spectrum of only 20 averaged scans, 23 *b* ions and 36 *y* ions of CKM were identified with high resolution and high mass accuracy (**Figure 6.11C, D**). Using this method, we were able to achieve identification of proteins up to 47.4 kDa. In particular, the 47.4 protein is a novel proteoform of the trifunctional enzyme, β-subunit (Uniprot Human P55084) (**Figure 6.14**) with removal of the N-terminal 33 amino acids. This 33-amino acid peptide is likely a signaling peptide to aid the transportation of this protein to the mitochondria and is

cleaved after transport is completed. Importantly, this proteoform is not present in the protein database, which showcased the power of top-down MS and MS/MS for the identification and characterization of novel proteoforms.

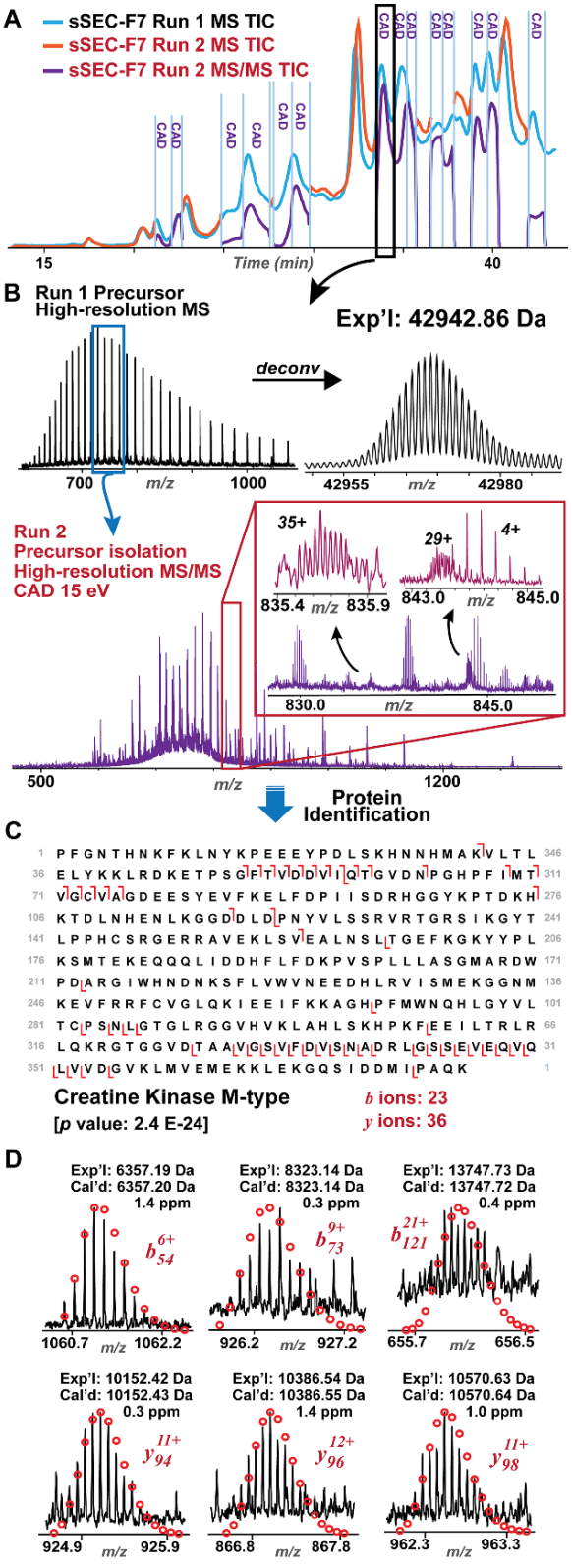


Figure 6.11: **Top-down targeted MS/MS for protein identification in the sSEC fraction.** (A) TIC of the first MS experiment (Run 1, blue trace) aligned with TICs of the second experiment (Run 2, purple and orange trace) for targeted MS/MS analysis (purple trace) of the proteins of interest in defined time segments.

For the remaining time segments wherein no proteins of interest were found, MS data were collected (orange trace). (B) High-resolution mass spectrum and the deconvoluted spectrum (80,000 resolving power) for the 42.9 kDa protein eluted between 33.5 and 34.5 min from the first MS experiment. Precursor ions of the 42.9 kDa protein were fragmented by CAD in Run 2, yielding complex high-resolution tandem mass spectrum. Insets show zoom-in view of the isotopically resolved fragment ions. (C) The 42.9 kDa protein was identified as creatine kinase M-type (CKM) with high confidence (p-value: 2.4 E-24). An average of 20 tandem mass spectrum yielded 23 b ions and 36 y ions from CKM. 55 bonds of 379 were cleaved, yielding 15% bond cleavage. (D) Representative b and y fragment ions of CKM with high mass accuracy.

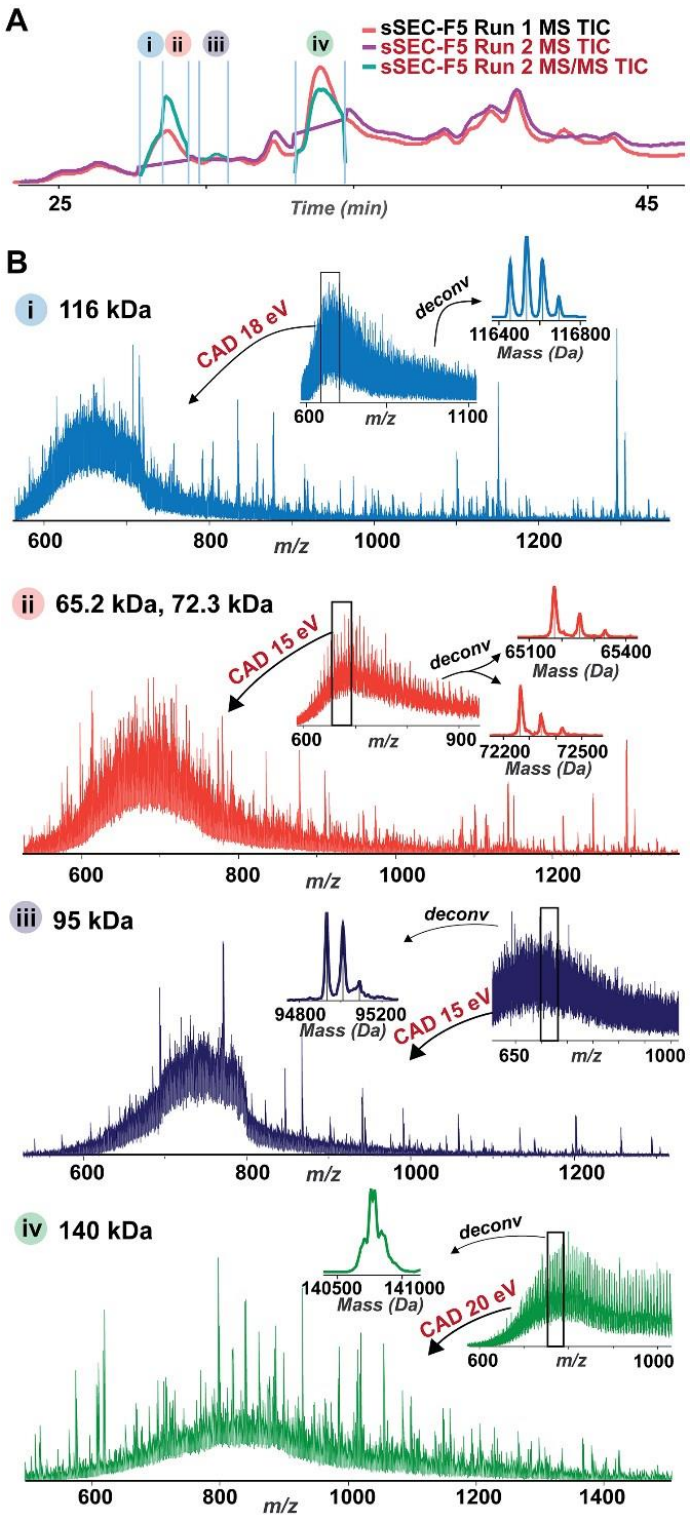


Figure 6.12: **2D sSEC-RPC enabled targeted MS/MS analysis of high MW proteins in 3sSEC Fraction 5**. (A) TIC of the first MS experiment (Run 1, orange trace) aligned with TICs of the second experiment (Run 2, purple and teal trace) for targeted MS/MS analysis (teal trace) on the proteins of interest in defined time segments. For the remaining time segments wherein no proteins of interest were found, MS data were collected (purple trace). (B) Run 1 allowed for the determination of retention windows for

individual proteins of interest, and the resulting top-down mass spectra and the deconvoluted spectra allowed for the determination of precursor MW. Online MS/MS analysis by CID yielded complex high-resolution tandem mass spectra for individual proteins of interest. Collisional energy for each protein of interest is highlighted in red font.

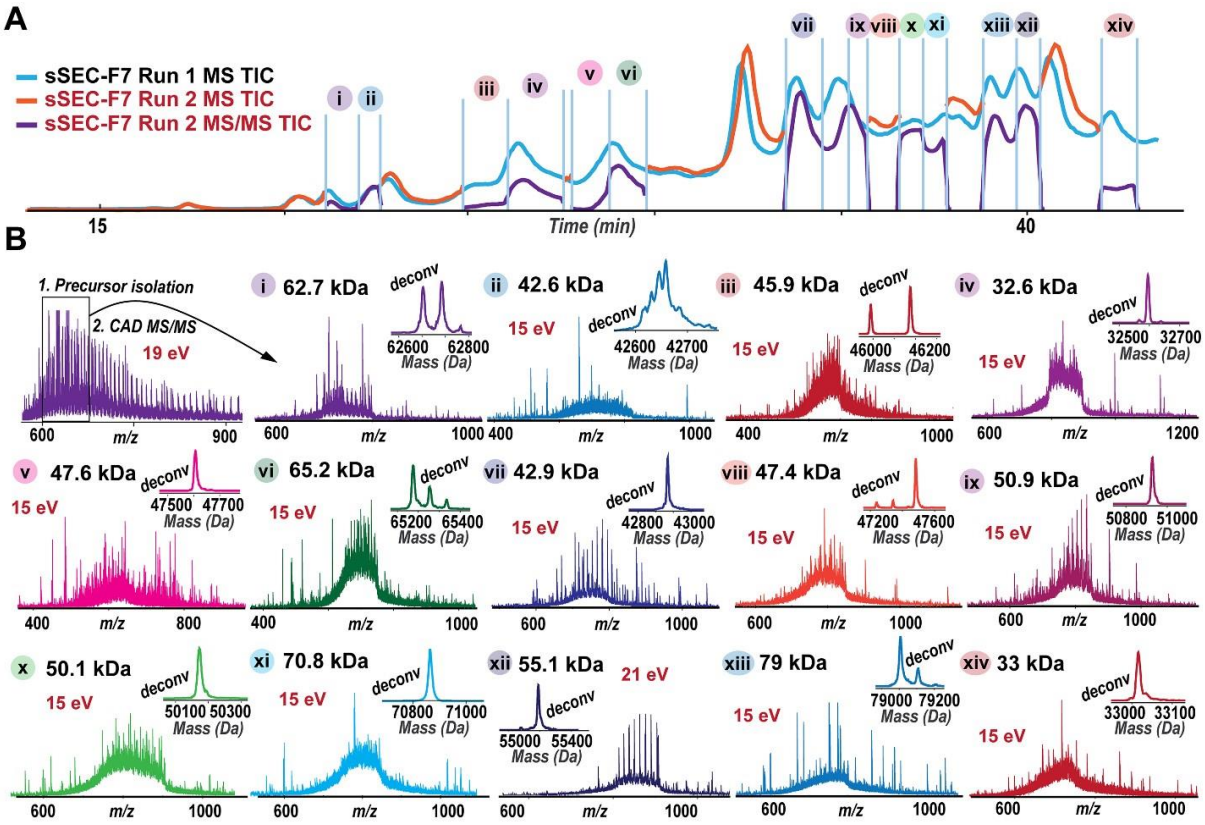


Figure 6.13: **2D sSEC-RPC enabled targeted MS/MS of high/intermediate MW proteins in 3sSEC Fraction 7**. (A) TIC of the first MS experiment (Run 1, blue trace) aligned with TICs of the second experiment (Run 2, purple and orange trace) for targeted MS/MS analysis (purple trace) on the proteins of interest in defined time segments. For the remaining time segments wherein no proteins of interest were found, MS data were collected (orange trace). (B) Run 1 allowed for the determination of retention windows for individual proteins of interest, and the resulting top-down mass spectra and the deconvoluted spectra allowed for the determination of precursor MW. Online MS/MS analysis by CID yielded complex high-resolution tandem mass spectra for individual proteins of interest. Collisional energy for each protein of interest is highlighted in red font.

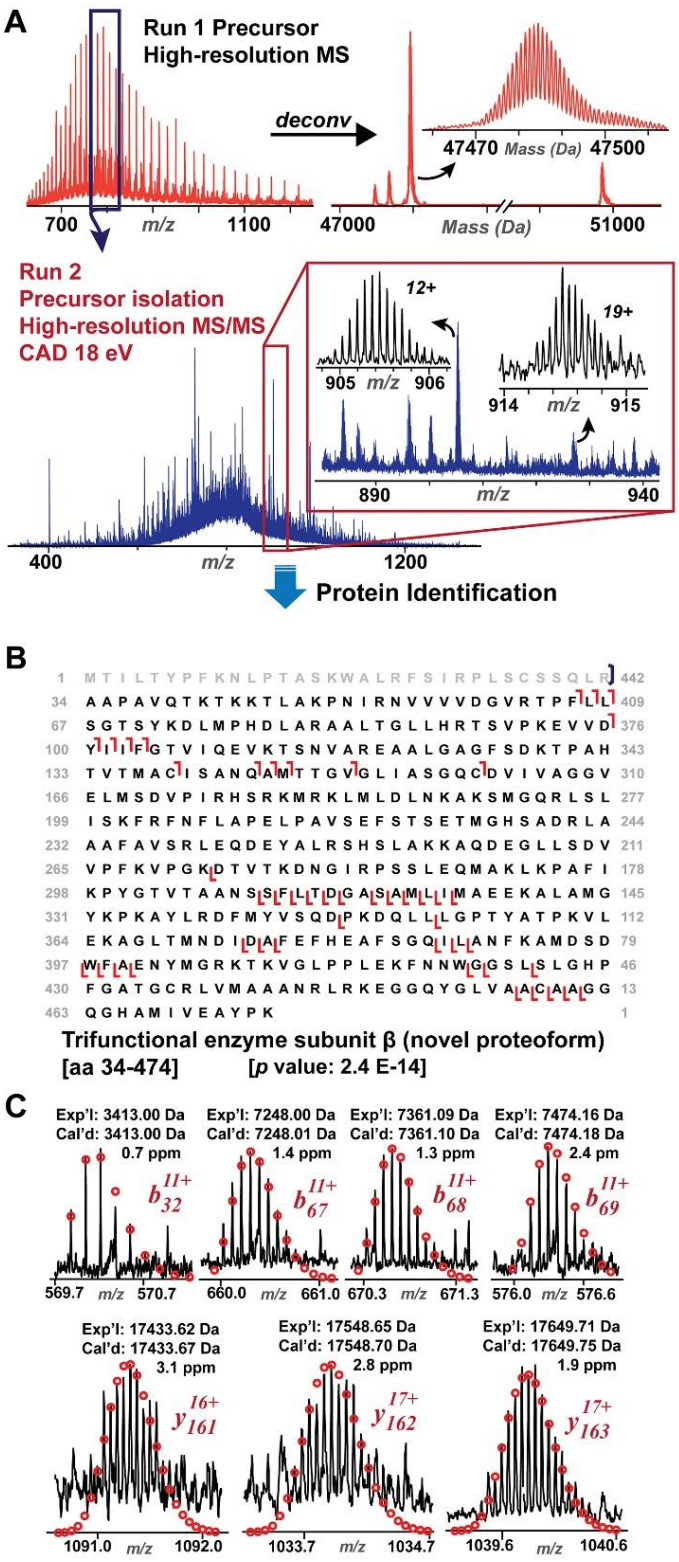


Figure 6.14: (A) Top-down mass spectrum and the deconvoluted spectrum for the 47.4 kDa protein (sSEC fraction 7, RT 34.5 – 35.2 min) determined by first MS run (Run 1). Targeted MS/MS experiment was performed by isolating precursor ions isolated across 750 – 850 *m/z* in Run 2, yielding complex high-resolution tandem mass spectrum. Insets show zoom-in views of the isotopically resolved fragments. 44

bonds of 440 were cleaved, yielding 10 % bond cleavage. (B) A novel proteoform of trifunctional enzyme subunit β was identified with high confidence (2.4 E-14). (C) Representative *b* and *y* fragment ions with high mass accuracy validated using MASH Suite Pro.

Nevertheless, we were unable to achieve confident identification of proteins with MW higher than 50 kDa, primarily because these proteins were not isotopically resolved by the mass spectrometer. This presented difficulties in the determination of accurate monoisotopic mass of the precursor ion, which is important for protein identification using MS-Align+. Online CAD of the 116.4 kDa protein (**Figure 6.7**) yielded a complex tandem mass spectrum with over 400 fragment ions, which were all manually validated using the MASH Suite Pro. Manual interpretation of the tandem mass spectrum revealed a portion of the amino acid sequence (**Figure 6.15**). However, the search algorithm did not match the tandem mass spectrum to any protein in the database with high confidence, suggesting that more advanced algorithms for top-down protein identification, independent of the precursor mass, is needed for the confident identification of high MW proteins. Furthermore, other fragmentation methods, such as electron-based dissociation and ultraviolet photodissociation method,[406, 407] can be utilized in the 2D sSEC-RPC-MS pipeline to improve protein identification and characterization.

Considerable sample dilution also reduced S/N of the intermediate to high MW proteins, and resulted in difficulties in protein identification. Sample dilution was a consequence of fraction collection using columns with large inner diameter at high flow rate. In addition, individual proteins/proteoforms were spread across two or more fractions, which required more samples to enable top-down MS and MS/MS analysis. To this end, this strategy could be improved with the implementation of online sSEC-RPC-MS. The

development of automated 2D LC systems and MS-compatible chromatographic strategies[183] have opened the door for the online versions of this 2D LC platform.

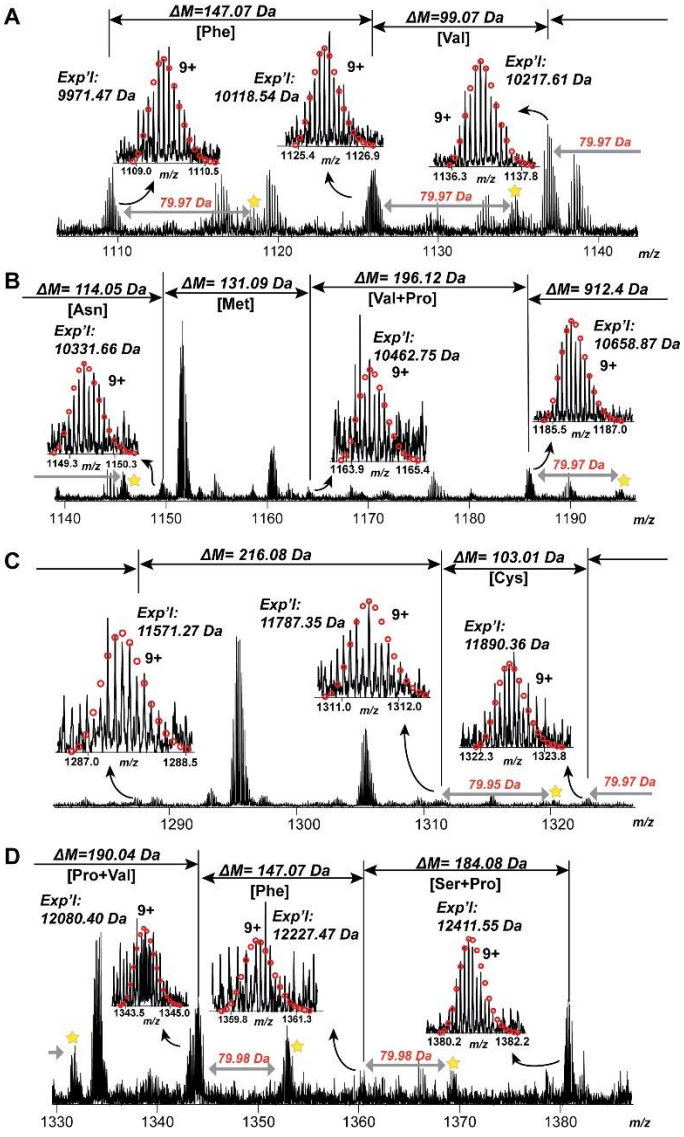


Figure 6.15: **Partial sequence obtained by manual interpretation of the tandem mass spectrum of the 116.4 kDa protein**. Fragment ions with 9+ charge were used to deduce amino acid sequence. (A), (B) shows confident sequence determination of Phe-Val-Asn-Met-(Val+Pro), and (C), (D) shows Cys-(Pro+Val)-Phe-(Ser+Pro). Ions with 79.97 Da mass shift were also identified by manual interpretation of the spectra, indicating the presence of phosphorylation, which is consistent with the multiply-phosphorylated proteoforms seen in the mass spectrum of the 116.4 kDa protein.

In this study, we have shown that combining SEC columns with different pore sizes in series can achieve high-resolution separation of proteins in a complex mixture over a broad MW range (10-223 kDa). Implementing sSEC with MS-compatible solvent coupled to RPC-MS analysis significantly enhanced the detection high MW proteins above 60 kDa. This 2D system (sSEC-RPC) coupled to a high-resolution Q-TOF mass spectrometer greatly surpasses the 1D RPC-MS method by enabling the detection of 4044 more unique proteoforms and high MW proteins up to 223 kDa. Effective protein fractionation and separation allowed for the detection of low abundance protein PTMs and online targeted MS/MS analysis of intermediate to high MW proteins for protein identification and characterization. This versatile method opens up new opportunities for the large-scale top-down proteomics analysis of high MW proteins from complex protein mixtures. Moreover, sSEC can also be used in non-denaturing conditions coupled with other non-denaturing chromatography methods, such as hydrophobic interaction chromatography or IEC,[185] holding great promise for the analysis of native protein complexes.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardiac disease estimated to affect 1.6 million people in the U.S.,[1, 5] and is the most common cause of sudden cardiac death in young adults.[8, 9] Over the past couple decades, the genetic causes of HCM were identified to be mutations in the protein constituents of the sarcomeres, and our knowledge on the HCM-associated genetic variants continues to evolve.[2, 25, 26] However, it remains unclear how sarcomeric protein mutations result in adverse cardiac remodeling and contractile dysfunction. While muscle contraction is ultimately triggered by elevated intracellular calcium, substantial evidence points to the importance of post-translation modifications (PTMs) for the regulation of contractile function,[24, 230, 250, 408] and the roles of sarcomeric protein PTMs in cardiac signal transduction have been increasingly recognized.[250, 409] The research presented in this dissertation aimed to uncover alterations in the sarcomeric protein PTMs in human HCM samples.

In chapter 2, the impacts of temperature and sample handling on the quality of human tissue were explored using a top-down proteomics strategy. Due to the limited availability of human tissue samples, sharing of samples between laboratories is common, but an ever-present concern when examining human heart tissue samples is the potential for artefactual changes related to temperature changes in the process of sample shipping and handling. Flash-freezing cardiac tissue in liquid nitrogen remains the best method for preserving the integrity of the tissue; yet, tissue warm-up post-procurement may occur during shipping and/or when processing the cardiac tissue samples, with unknown consequences at the protein level. In this study, we revealed that

sarcomeric protein PTMs were differentially affected by increased temperature. Specifically, while cardiac troponin T (cTnT) and MLC2 phosphorylation decreased following tissue warm-up, there was a drastic increase in the phosphorylation of cardiac troponin I (cTnI) and enigma homolog isoform 2 (ENH2) within 15 min of temperature increase at 4 °C or 22 °C. Tandem MS (MS/MS) revealed increased phosphorylation of cTnI to Ser22/23, the canonical sites for protein kinase A (PKA)-mediated phosphorylation,[245, 261, 272] suggesting activation of PKA upon temperature increase. Consistently, cyclic AMP (cAMP), the direct activator of PKA, increased rapidly within 5 min of temperature increase. Additionally, minor degradation products of cTnI and cTnT were observed with temperature increase, which can be indicative of poor tissue quality for the quality control of the human tissue samples.

In chapter 3, quantitative top-down proteomics was employed to study changes in the sarcomeric protein PTMs in human donor and HCM heart tissues. Significant down-regulation in the phosphorylation of cTnI and ENH2 was observed in the myectomy and explanted HCM heart tissues compared to the donor tissues. There was a strong linear correlation between the level of cTnI and ENH2 phosphorylation in the individual heart tissue sample, indicating co-regulation of these two proteins, or a cross-talk mechanism between the signaling pathways that mediate the phosphorylation of these two proteins. Interestingly, the phosphorylation levels of cTnT, MLC and α-Tpm were altered in the myectomy HCM tissues compared to the donor tissues, but not in the explanted HCM tissues. In addition to ENH2, at least 6 more Z-disc and sarcomere-associated proteins were identified, including muscle LIM protein (MLP), cysteine-rich protein 2 (CRIP2), Elfin, four and half LIM protein 2 (FHL2), Cypher and myozenin-2 (MYOZ2). Similar to ENH2, Elfin and Cypher belongs to the family of the PDZ-LIM domain containing proteins.[409, 410] Elfin,

encoded by *PDLIM1*, had no detectable PTMs from top-down proteomics analysis. On the other hand, top-down proteomics revealed two splicing isoforms of Cypher, encoded by *LDB3*, which are the short isoforms Cypher5 and Cypher6 that contain only the N-terminal PDZ domain, without the C-terminal LIM domains seen in the full length protein. Intriguingly, while the overall phosphorylation of Cypher6 decreased in HCM, the overall phosphorylation of Cypher5 increased in HCM compared to the donor tissues, suggesting that these closely-related protein isoforms were regulated differently in cardiac hypertrophic remodeling. Cypher was previously identified to be a PKA anchoring protein that plays important role in the localization of PKA for mediating the phosphorylation of L-type calcium channel.[312] However, the impacts of phosphorylation on Cypher function have not been convincingly demonstrated. Moreover, the roles of different Cypher isoforms in cardiac signaling and contractile function regulation remain to be elucidated. Phosphorylation of CRIP2 was detected in the human samples, but the degree of phosphorylation varied widely. Interestingly, CRIP2 was observed with multiply thiol-sulfenylated proteoforms, and the degree of thiol-sulfenylation increased significantly in the explanted HCM heart tissues, but not in the myectomy tissues, compared to the donor tissues. CRIP2 is a LIM domain containing protein closely related to MLP. In contrast to MLP, which has been more extensively studied in the context of cardiac physiology and diseases, the roles of CRIP2 in cardiac function remain unexplored, albeit being highly expressed in the heart during development and at adult stage.[309] Interestingly, CRIP2 was previously identified to be a tumor suppressor via suppression of angiogenesis mediated by the NF-κB pathway, and it also regulates the HIF-1α signaling in metabolism.[310, 311] CRIP2 is a cysteine-rich protein, and for the first time, was shown to be thiol-sulfenylated in the explanted HCM heart tissues. Cysteine thiol-sulfenylation is a new type of

protein PTM that was only reported very recently on tyrosine kinases such as EGFR.[307] This new PTM may play a unique role in cellular redox signaling.

Recent advances in the stem cell technology and directed differentiation of human pluripotent stem cells into cardiomyocytes (CMs) provide unprecedented opportunities for the study of heart diseases.[79, 135, 137, 313] Unlike the study of human tissues, which is subject to complications from a mixture of cell types, the use of medications and other co-morbidities, human pluripotent stem cell-derived cardiomyocyte (hPSC-CMs) permits the study of alterations in protein PTMs, intracellular signaling and contractile properties in relatively pure CM population, independent of complications from medication uses and co-morbidities. Therefore, hPSC-CMs provide a good system for modeling human HCM *in vitro* to uncover the molecular alterations underlying the disease. Moreover, this model provides a platform to study disease onset and progression, which is usually not feasible with human patients. Nevertheless, HCM typically develops in the adulthood; yet, hPSC-CMs exhibit the properties of fetal CMs. Promoting hPSC-CM maturation to acquire the properties of adult CMs is important to realize the full capacity of this new model for the study and treatment of heart diseases, but is difficult in part due to a lack of comprehensive methods for benchmarking the maturation of hPSC-CMs.

In chapter 4, an unbiased proteomics method was developed to combine the strengths of top-down and bottom-up proteomics in order to gain a comprehensive view of the molecular alterations during hPSC-CM maturation. A robust and quantitative top-down LC-MS method was developed to quantify protein isoform expression and PTMs simultaneously. The method was proven reliable for quantifying protein isoform expression, and was applied to evaluate the

maturity of hPSC-CMs from early- and late-stage 2D monolayer culture, as well as 3D engineered cardiac tissue (ECT). Over 90% of the hPSC-CMs produced were ventricular CMs; however, the atrial and ventricular isoforms of the myosin light chains (MLC1 and MLC2) were found co-expressed in all the samples analyzed in this study. With prolonged culturing in either 2D or 3D cultures, MLC2 ventricular isoform expression increased while the atrial isoform expression decreased, which was consistent with improved ventricular hPSC-CM maturation. However, there was no consistent changes observed in terms of the expression of MLC1 isoforms, and the molecular switch that regulates MLC1 isoform change remains to be identified. Slow skeletal isoform of troponin I (ssTnI) is the isoform expressed in the fetal heart, and during development, ssTnI expression decreases while cTnI increases and become the only isoform expressed in the heart. However, hPSC-CMs showed robust expression of ssTnI with virtually undetectable level of cTnI. The expression of cTnI relative to ssTnI was significantly increased in the hPSC-CMs from 3D ECT compared to the 2D culture, supporting that 3D ECT can promote hPSC-CM maturation. Furthermore, the phosphorylation of many sarcomeric proteins was found altered in the hPSC-CMs during maturation. In particular, α-Tpm phosphorylation decreased in the more mature hPSC-CMs, which is consistent with low level of α-Tpm phosphorylation in the non-failing adult ventricular tissues. This was also confirmed in the developing rat hearts.[340]

Top-down proteomics is powerful for the quantification of protein isoforms and PTMs, whereas bottom-up proteomics permits the study of large proteins and low abundant proteins beyond the sarcomeres. By employing a bottom-up proteomics approach, we identified changes in proteins responsible for metabolism, calcium homeostasis and cardiac excitability in hPSC-CMs during maturation. Specifically, a number of metabolic proteins involved in fatty acid oxidation

were found up-regulated and some glycolytic enzymes were found down-regulated in more mature hPSC-CMs. In addition, a number of large sarcomeric proteins, including cardiac myosin binding protein C (cMyBP-C), α-actinin 2 and myomesin 1 were found up-regulated in more mature hPSC-CMs. Proteins that link the sarcomeres to the plasma membrane, such as dystrophin, were also found up-regulated. Importantly, proteins involved in calcium regulation and cardiomyocyte excitability, including phospholamban and calsequestrin 2, were also up-regulated in the more mature hPSC-CMs. The alterations of these proteins were further confirmed in the developing mouse cardiac ventricles, indicating that these proteins can be novel markers to evaluate the maturation of hPSC-CMs.

Top-down proteomics provides unique advantages for the analysis of intact proteins to reveal alterations in protein isoforms and PTMs in a given biological system. Nevertheless, as an emerging field, there are significant challenges remaining to be tackled before this technology can realize its full potential.[147, 244] Specifically, in contrast to the bottom-up proteomics wherein data analysis tools are well-developed and widely available, there was a lack of comprehensive software tool for the analysis of top-down proteomics data. In addition, due to an exponential decay of signal-to-noise ratio with increasing molecular weight,[186] the analysis of proteins with molecular weight higher than 60 kDa remains challenging.

Chapter 5 and 6 presents the methodological improvement in the top-down proteomics workflow. A comprehensive software tool, MASH Suite Pro, was developed to facilitate the data analysis process in top-down proteomics. MASH Suite Pro is user-friendly and versatile, and equipped with multiple functions for protein identification and quantification, as well as protein

sequence and PTM characterization. The software allows for visual validation of the computational outputs according to the real mass spectra, which ensures accurate data interpretation. In chapter 6, a high-resolution serial size-exclusion chromatography (sSEC) method was developed, and when coupled to the conventional reverse phase chromatography, this 2D separation platform not only extended the coverage of proteoforms, but also improved the detection of high molecular weight proteins, permitting the detection of cMyBP-C and β-MHC. cMyBP-C and β-MHC are two major components of the thick filaments, with a molecular weight of ~140 and 220 kDa, respectively. Mutations on cMyBP-C and β-MHC collectively account for over 70% of all the known HCM-causing mutations;[1] yet analysis of these two proteins by top-down proteomics remains extremely difficult due to their high molecular weights. Future improvement on the sSEC workflow with reduced sample consumption will contribute to the top-down proteomics analysis of cMyBP-C and β-MHC in the context of HCM.

In summary, the results presented in this dissertation revealed alterations in the sarcomeric protein PTMs in response to temperature alterations, and presents an effective method for evaluating human tissue quality which is essential for the study of molecular alterations in human cardiac diseases. Empowered by the top-down proteomics method, the study of sarcomeric protein PTM alterations was conducted using human tissues, and revealed alterations of sarcomeric protein phosphorylation in HCM that likely contributes to contractile dysfunction and cardiac remodeling. Moreover, this is the first study to compare the HCM tissues from myectomy surgery and end-stage failing hearts with HCM. Interestingly, changes in a number of sarcomeric protein phosphorylation in the myectomy tissues were reversed or partially reversed in the failing stages, suggesting important functional changes of the heart, consistent with changes from preserved

ejection fraction at the hyperdynamic stage to reduced ejection fraction at the failing stage. Moreover, this study identified new sarcomeric protein PTMs, including multiply phosphorylation of Cypher5 and cysteine thiol-sulfenylation of CRIP2. The study also revealed differential regulation of PTMs on closely-related protein isoforms such as Cypher5 and Cypher6. To facilitate the use of hPSC-CMs as *in vitro* model of cardiac diseases including HCM, an integrated proteomics method combining the strengths of top-down and bottom-up proteomics were developed to assess the maturation of hPSC-CMs. This study revealed important changes of sarcomeric protein isoforms and PTMs and differential regulation of other proteins during hPSC-CM maturation. Importantly, the study identified a number of proteins that were up-regulated during hPSC-CM maturation, which can serve as potential new markers for benchmarking hPSC-CM maturity. To further address the challenges with employing top-down proteomics in biomedical research, a comprehensive software tool was developed to facilitate data analysis in top-down proteomics. Furthermore, a novel protein separation strategy was developed, which enables future investigation on the large sarcomeric proteins and their PTMs in HCM.

The research presented in this dissertation opens up new opportunities for future studies of human cardiac diseases using the cutting-edge high-throughput proteomics technology. In this study, top-down proteomics revealed alterations in the phosphorylation of many sarcomeric proteins in human HCM tissues compared to donor tissues. Nevertheless, it remains unclear whether alterations in the sarcomeric protein phosphorylation are causes of cardiac hypertrophy or consequence of hypertrophic remodeling. Furthermore, the roles of the many sarcomeric protein phosphorylations remain elusive. For example, Cypher6 was found mono-phosphorylated, and Cypher5 was found mono- and bis-phosphorylated in the donor heart tissues. However, at least

two more phosphorylated proteoforms of Cypher5 were observed in the HCM tissues. The roles of Cypher5 phosphorylation in cardiac physiology and pathophysiology are intriguing. Future studies to investigate the function of Cypher phosphorylation by mutagenesis and high-throughput proteomics are promising to unravel additional mechanisms of contractile function regulation and cardiac remodeling. In addition, Cypher isoforms appeared differentially regulated in cardiac pathophysiological response, and the roles of individual Cypher isoform in cardiac development, hypertrophic response and adaptation are interesting areas of research. In this study, human cardiac tissues were obtained from patients with relatively severe clinical presentations. However, HCM exhibits a wide spectrum of phenotypes with many patients remaining asymptomatic or mildly affected. The assessment of sarcomeric protein PTMs among mildly affected HCM patients are important for revealing the significance of individual PTMs in HCM phenotypic expression. However, this is challenging due to the invasive nature of the myectomy and transplantation surgery. Future studies utilizing hPSC-CMs for modeling HCM will provide more insights into the association of individual PTMs with severity of the disease, and add to our knowledge regarding HCM onset and progression. Moreover, the facts that individuals with the same HCM mutation have different clinical presentations suggest that genetic background and environment are important modifiers of the disease. Investigations of the same HCM mutations under different genetic backgrounds using hPSC-CMs from patients with known clinical course of disease progression are extremely meaningful for uncovering the protective factors/signaling pathways to aid prevention and treatment of HCM.

BIBLIOGRAPHY

1. Maron BJ and Maron MS. Hypertrophic cardiomyopathy. *Lancet*. 2013;381:242-55.

2. Maron BJ, McKenna WJ, Danielson GK, Kappenberger LJ, Kuhn HJ, Seidman CE, Shah PM, Spencer WH, Spirito P, Ten Cate FJ, Wigle ED, Documents ACoCFTFoCEC and Guidelines ESoCCfP. American College of Cardiology/European Society of Cardiology Clinical Expert Consensus Document on Hypertrophic Cardiomyopathy. A report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents and the European Society of Cardiology Committee for Practice Guidelines. *Eur Heart J*. 2003;24:1965-91.

3. Maron BJ. Hypertrophic cardiomyopathy: a systematic review. *JAMA*. 2002;287:1308-20.

4. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT and Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. Echocardiographic analysis of 4111 subjects in the CARDIA Study. Coronary Artery Risk Development in (Young) Adults. *Circulation*. 1995;92:785-9.

5. Semsarian C, Ingles J, Maron MS and Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2015;65:1249-54.

6. Maron BJ, Ommen SR, Semsarian C, Spirito P, Olivotto I and Maron MS. Hypertrophic cardiomyopathy: present and future, with translation into contemporary cardiovascular medicine. *J Am Coll Cardiol*. 2014;64:83-99.

7. Maron BJ, Roberts WC, McAllister HA, Rosing DR and Epstein SE. Sudden death in young athletes. *Circulation*. 1980;62:218-29.

8. Maron BJ, Epstein SE and Roberts WC. Causes of sudden death in competitive athletes. *J Am Coll Cardiol*. 1986;7:204-14.

9. Maron BJ, Doerer JJ, Haas TS, Tierney DM and Mueller FO. Sudden deaths in young competitive athletes: analysis of 1866 deaths in the United States, 1980-2006. *Circulation*. 2009;119:1085-92.

10. Spirito P, Seidman CE, McKenna WJ and Maron BJ. The management of hypertrophic cardiomyopathy. *N Engl J Med*. 1997;336:775-85.

11. Olivotto I, Maron BJ, Appelbaum E, Harrigan CJ, Salton C, Gibson CM, Udelson JE, O'Donnell C, Lesser JR, Manning WJ and Maron MS. Spectrum and clinical significance of systolic function and myocardial fibrosis assessed by cardiovascular magnetic resonance in hypertrophic cardiomyopathy. *Am J Cardiol*. 2010;106:261-7.

12. Harris KM, Spirito P, Maron MS, Zenovich AG, Formisano F, Lesser JR, Mackey-Bojack S, Manning WJ, Udelson JE and Maron BJ. Prevalence, clinical profile, and significance of left ventricular remodeling in the end-stage phase of hypertrophic cardiomyopathy. *Circulation*. 2006;114:216-25.

13. Tian T, Wang Y, Sun K, Wang J, Zou Y, Zhang W, Bao J, Zhu L, Shen H, Hui R, Zhou X and Song L. Clinical profile and prognostic significance of atrial fibrillation in hypertrophic cardiomyopathy. *Cardiology*. 2013;126:258-64.

14. Maron BJ, Olivotto I, Bellone P, Conte MR, Cecchi F, Flygenring BP, Casey SA, Gohman TE, Bongioanni S and Spirito P. Clinical profile of stroke in 900 patients with hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2002;39:301-7.

15. Elliott PM, Gimeno Blanes JR, Mahon NG, Poloniecki JD and McKenna WJ. Relation between severity of left-ventricular hypertrophy and prognosis in patients with hypertrophic cardiomyopathy. *Lancet*. 2001;357:420-4.

16. Klues HG, Schiffers A and Maron BJ. Phenotypic spectrum and patterns of left ventricular hypertrophy in hypertrophic cardiomyopathy: morphologic observations and significance as assessed by two-dimensional echocardiography in 600 patients. *J Am Coll Cardiol*. 1995;26:1699-708.

17. Maron BJ, Anan TJ and Roberts WC. Quantitative analysis of the distribution of cardiac muscle cell disorganization in the left ventricular wall of patients with hypertrophic cardiomyopathy. *Circulation*. 1981;63:882-94.

18. Maron BJ, Wolfson JK, Epstein SE and Roberts WC. Intramural ("small vessel") coronary artery disease in hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 1986;8:545-57.

19. Tanaka M, Fujiwara H, Onodera T, Wu DJ, Matsuda M, Hamashima Y and Kawai C. Quantitative analysis of narrowings of intramyocardial small arteries in normal hearts, hypertensive hearts, and hearts with hypertrophic cardiomyopathy. *Circulation*. 1987;75:1130-9.

20. Maron MS, Olivotto I, Maron BJ, Prasad SK, Cecchi F, Udelson JE and Camici PG. The case for myocardial ischemia in hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2009;54:866-75.

21. Geisterfer-Lowrance AA, Kass S, Tanigawa G, Vosberg HP, McKenna W, Seidman CE and Seidman JG. A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell*. 1990;62:999-1006.

22. Maron BJ, Maron MS and Semsarian C. Genetics of hypertrophic cardiomyopathy after 20 years: clinical perspectives. *J Am Coll Cardiol*. 2012;60:705-15.

23. Marian AJ and Belmont J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circ Res*. 2011;108:1252-69.

24. Peng Y, Gregorich ZR, Valeja SG, Zhang H, Cai W, Chen Y-C, Guner H, Chen AJ, Schwahn DJ, Hacker TA, Liu X and Ge Y. Top-down Proteomics Reveals Concerted Reductions in Myofilament and Z-disc Protein Phosphorylation after Acute Myocardial Infarction. *Molecular & Cellular Proteomics*. 2014;13:2752-2764.

25. Alcalai R, Seidman JG and Seidman CE. Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics. *J Cardiovasc Electrophysiol*. 2008;19:104-10.

26. Konno T, Chang S, Seidman JG and Seidman CE. Genetics of hypertrophic cardiomyopathy. *Curr Opin Cardiol*. 2010;25:205-9.

27. Landstrom AP and Ackerman MJ. Mutation type is not clinically useful in predicting prognosis in hypertrophic cardiomyopathy. *Circulation*. 2010;122:2441-9; discussion 2450.

28. Maron BJ, Maron MS and Semsarian C. Double or compound sarcomere mutations in hypertrophic cardiomyopathy: a potential link to sudden death in the absence of conventional risk factors. *Heart Rhythm*. 2012;9:57-63.

29. Hodatsu A, Konno T, Hayashi K, Funada A, Fujita T, Nagata Y, Fujino N, Kawashiri MA and Yamagishi M. Compound heterozygosity deteriorates phenotypes of hypertrophic cardiomyopathy with founder MYBPC3 mutation: evidence from patients and zebrafish models. *Am J Physiol Heart Circ Physiol*. 2014;307:H1594-604.

30. Bers DM. Calcium cycling and signaling in cardiac myocytes. *Annu Rev Physiol*. 2008;70:23-49.

31. Chung JH, Biesiadecki BJ, Ziolo MT, Davis JP and Janssen PM. Myofilament Calcium Sensitivity: Role in Regulation of. *Front Physiol*. 2016;7:562.

32. Ashrafian H, McKenna WJ and Watkins H. Disease pathways and novel therapeutic targets in hypertrophic cardiomyopathy. *Circ Res*. 2011;109:86-96.

33. Hwang PM and Sykes BD. Targeting the sarcomere to correct muscle function. *Nat Rev Drug Discov*. 2015;14:313-28.

34. Ren X, Hensley N, Brady MB and Gao WD. The Genetic and Molecular Bases for Hypertrophic Cardiomyopathy: The Role for Calcium Sensitization. *J Cardiothorac Vasc Anesth*. 2018;32:478-487.

35. Landstrom AP, Parvatiyar MS, Pinto JR, Marquardt ML, Bos JM, Tester DJ, Ommen SR, Potter JD and Ackerman MJ. Molecular and functional characterization of novel hypertrophic cardiomyopathy susceptibility mutations in TNNC1-encoded troponin C. *J Mol Cell Cardiol*. 2008;45:281-8.

36. Parvatiyar MS, Landstrom AP, Figueiredo-Freitas C, Potter JD, Ackerman MJ and Pinto JR. A mutation in TNNC1-encoded cardiac troponin C, TNNC1-A31S, predisposes to hypertrophic cardiomyopathy and ventricular fibrillation. *J Biol Chem*. 2012;287:31845-55.

37. Kalyva A, Parthenakis FI, Marketou ME, Kontaraki JE and Vardas PE. Biochemical characterisation of Troponin C mutations causing hypertrophic and dilated cardiomyopathies. *J Muscle Res Cell Motil*. 2014;35:161-78.

38. Takahashi-Yanaga F, Morimoto S, Harada K, Minakami R, Shiraishi F, Ohta M, Lu QW, Sasaguri T and Ohtsuki I. Functional consequences of the mutations in human cardiac troponin I gene found in familial hypertrophic cardiomyopathy. *J Mol Cell Cardiol*. 2001;33:2095-107.

39. Messer AE, Bayliss CR, El-Mezgueldi M, Redwood CS, Ward DG, Leung MC, Papadaki M, Dos Remedios C and Marston SB. Mutations in troponin T associated with Hypertrophic Cardiomyopathy increase Ca(2+)-sensitivity and suppress the modulation of Ca(2+)-sensitivity by troponin I phosphorylation. *Arch Biochem Biophys*. 2016;601:113-20.

40. Harada K and Potter JD. Familial hypertrophic cardiomyopathy mutations from different functional regions of troponin T result in different effects on the pH and Ca2+ sensitivity of cardiac muscle contraction. *J Biol Chem*. 2004;279:14488-95.

41. Sommese RF, Nag S, Sutton S, Miller SM, Spudich JA and Ruppel KM. Effects of troponin T cardiomyopathy mutations on the calcium sensitivity of the regulated thin filament and the actomyosin cross-bridge kinetics of human β-cardiac myosin. *PLoS One*. 2013;8:e83403.

42. Heller MJ, Nili M, Homsher E and Tobacman LS. Cardiomyopathic tropomyosin mutations that increase thin filament Ca2+ sensitivity and tropomyosin N-domain flexibility. *J Biol Chem*. 2003;278:41742-8.

43. Redwood C and Robinson P. Alpha-tropomyosin mutations in inherited cardiomyopathies. *J Muscle Res Cell Motil*. 2013;34:285-94.

44. Michele DE, Albayya FP and Metzger JM. Direct, convergent hypersensitivity of calcium-activated force generation produced by hypertrophic cardiomyopathy mutant alpha-tropomyosins in adult cardiac myocytes. *Nat Med*. 1999;5:1413-7.

45. Harris SP, Lyons RG and Bezold KL. In the thick of it: HCM-causing mutations in myosin binding proteins of the thick filament. *Circ Res*. 2011;108:751-64.

46. Flashman E, Redwood C, Moolman-Smook J and Watkins H. Cardiac myosin binding protein C: its role in physiology and disease. *Circ Res*. 2004;94:1279-89.

47. Oakley CE, Hambly BD, Curmi PM and Brown LJ. Myosin binding protein C: structural abnormalities in familial hypertrophic cardiomyopathy. *Cell Res*. 2004;14:95-110.

48. Kunst G, Kress KR, Gruen M, Uttenweiler D, Gautel M and Fink RH. Myosin binding protein C, a phosphorylation-dependent force regulator in muscle that controls the attachment of myosin heads by its interaction with myosin S2. *Circ Res*. 2000;86:51-8.

49. Calaghan SC, Trinick J, Knight PJ and White E. A role for C-protein in the regulation of contraction and intracellular Ca2+ in intact rat ventricular myocytes. *J Physiol*. 2000;528 Pt 1:151-6.

50. Hofmann PA, Hartzell HC and Moss RL. Alterations in Ca2+ sensitive tension due to partial extraction of C-protein from rat skinned cardiac myocytes and rabbit skeletal muscle fibers. *J Gen Physiol*. 1991;97:1141-63.

51. Hofmann PA, Greaser ML and Moss RL. C-protein limits shortening velocity of rabbit skeletal muscle fibres at low levels of Ca2+ activation. *J Physiol*. 1991;439:701-15.

52. Weisberg A and Winegrad S. Alteration of myosin cross bridges by phosphorylation of myosin-binding protein C in cardiac muscle. *Proc Natl Acad Sci U S A*. 1996;93:8999-9003.

53. Harris SP, Bartley CR, Hacker TA, McDonald KS, Douglas PS, Greaser ML, Powers PA and Moss RL. Hypertrophic cardiomyopathy in cardiac myosin binding protein-C knockout mice. *Circ Res*. 2002;90:594-601.

54. van Dijk SJ, Dooijes D, dos Remedios C, Michels M, Lamers JM, Winegrad S, Schlossarek S, Carrier L, ten Cate FJ, Stienen GJ and van der Velden J. Cardiac myosin-binding protein C mutations and hypertrophic cardiomyopathy: haploinsufficiency, deranged phosphorylation, and cardiomyocyte dysfunction. *Circulation*. 2009;119:1473-83.

55. Smelter DF, de Lange WJ, Cai W, Ge Y and Ralphe JC. The HCM-linked W792R mutation in cardiac myosin-binding protein C reduces C6 FnIII domain stability. *Am J Physiol Heart Circ Physiol*. 2018;314:H1179-H1191.

56. Moore JR, Leinwand L and Warshaw DM. Understanding cardiomyopathy phenotypes based on the functional impact of mutations in the myosin motor. *Circ Res*. 2012;111:375-85.

57. Palmer BM, Fishbaugher DE, Schmitt JP, Wang Y, Alpert NR, Seidman CE, Seidman JG, VanBuren P and Maughan DW. Differential cross-bridge kinetics of FHC myosin mutations R403Q and R453C in heterozygous mouse myocardium. *Am J Physiol Heart Circ Physiol*. 2004;287:H91-9.

58. Palmiter KA, Tyska MJ, Haeberle JR, Alpert NR, Fananapazir L and Warshaw DM. R403Q and L908V mutant beta-cardiac myosin from patients with familial hypertrophic cardiomyopathy exhibit enhanced mechanical performance at the single molecule level. *J Muscle Res Cell Motil*. 2000;21:609-20.

59. Seebohm B, Matinmehr F, Köhler J, Francino A, Navarro-Lopéz F, Perrot A, Ozcelik C, McKenna WJ, Brenner B and Kraft T. Cardiomyopathy mutations reveal variable region of myosin converter as major element of cross-bridge compliance. *Biophys J*. 2009;97:806-24.

60. Kirschner SE, Becker E, Antognozzi M, Kubis HP, Francino A, Navarro-López F, Bit-Avragim N, Perrot A, Mirrakhimov MM, Osterziel KJ, McKenna WJ, Brenner B and Kraft T. Hypertrophic cardiomyopathy-related beta-myosin mutations cause highly variable calcium sensitivity with functional imbalances among individual muscle cells. *Am J Physiol Heart Circ Physiol*. 2005;288:H1242-51.

61. Miller G, Maycock J, White E, Peckham M and Calaghan S. Heterologous expression of wild-type and mutant beta-cardiac myosin changes the contractile kinetics of cultured mouse myotubes. *J Physiol*. 2003;548:167-74.

62. Alpert NR, Mohiddin SA, Tripodi D, Jacobson-Hatzell J, Vaughn-Whitley K, Brosseau C, Warshaw DM and Fananapazir L. Molecular and phenotypic effects of heterozygous,

homozygous, and compound heterozygote myosin heavy-chain mutations. *Am J Physiol Heart Circ Physiol*. 2005;288:H1097-102.

63. Wang Y, Xu Y, Kerrick WG, Guzman G, Diaz-Perez Z and Szczesna-Cordary D. Prolonged Ca2+ and force transients in myosin RLC transgenic mouse fibers expressing malignant and benign FHC mutations. *J Mol Biol*. 2006;361:286-99.

64. Kerrick WG, Kazmierczak K, Xu Y, Wang Y and Szczesna-Cordary D. Malignant familial hypertrophic cardiomyopathy D166V mutation in the ventricular myosin regulatory light chain causes profound effects in skinned and intact papillary muscle fibers from transgenic mice. *FASEB J*. 2009;23:855-65.

65. Belus A, Piroddi N, Scellini B, Tesi C, D'Amati G, Girolami F, Yacoub M, Cecchi F, Olivotto I and Poggesi C. The familial hypertrophic cardiomyopathy-associated myosin mutation R403Q accelerates tension generation and relaxation of human cardiac myofibrils. *J Physiol*. 2008;586:3639-44.

66. Ashrafian H, Redwood C, Blair E and Watkins H. Hypertrophic cardiomyopathy:a paradigm for myocardial energy depletion. *Trends Genet*. 2003;19:263-8.

67. Frey N, Brixius K, Schwinger RH, Benis T, Karpowski A, Lorenzen HP, Luedde M, Katus HA and Franz WM. Alterations of tension-dependent ATP utilization in a transgenic rat model of hypertrophic cardiomyopathy. *J Biol Chem*. 2006;281:29575-82.

68. Jung WI, Sieverding L, Breuer J, Hoess T, Widmaier S, Schmidt O, Bunse M, van Erckelens F, Apitz J, Lutz O and Dietze GJ. 31P NMR spectroscopy detects metabolic abnormalities in asymptomatic patients with hypertrophic cardiomyopathy. *Circulation*. 1998;97:2536-42.

69. Crilley JG, Boehm EA, Blair E, Rajagopalan B, Blamire AM, Styles P, McKenna WJ, Ostman-Smith I, Clarke K and Watkins H. Hypertrophic cardiomyopathy due to sarcomeric gene mutations is characterized by impaired energy metabolism irrespective of the degree of hypertrophy. *J Am Coll Cardiol*. 2003;41:1776-82.

70. Unno K, Isobe S, Izawa H, Cheng XW, Kobayashi M, Hirashiki A, Yamada T, Harada K, Ohshima S, Noda A, Nagata K, Kato K, Yokota M and Murohara T. Relation of functional and morphological changes in mitochondria to myocardial contractile and relaxation reserves in asymptomatic to mildly symptomatic patients with hypertrophic cardiomyopathy. *Eur Heart J*. 2009;30:1853-62.

71. Merante F, Tein I, Benson L and Robinson BH. Maternally inherited hypertrophic cardiomyopathy due to a novel T-to-C transition at nucleotide 9997 in the mitochondrial tRNA(glycine) gene. *Am J Hum Genet*. 1994;55:437-46.

72. Tanaka T, Sohmiya K and Kawamura K. Is CD36 deficiency an etiology of hereditary hypertrophic cardiomyopathy? *J Mol Cell Cardiol*. 1997;29:121-7.

73. Aoyama T, Souri M, Ushikubo S, Kamijo T, Yamaguchi S, Kelley RI, Rhead WJ, Uetake K, Tanaka K and Hashimoto T. Purification of human very-long-chain acyl-coenzyme A dehydrogenase and characterization of its deficiency in seven patients. *J Clin Invest*. 1995;95:2465-73.

74. Blair E, Redwood C, Ashrafian H, Oliveira M, Broxholme J, Kerr B, Salmon A, Ostman-Smith I and Watkins H. Mutations in the gamma(2) subunit of AMP-activated protein kinase cause familial hypertrophic cardiomyopathy: evidence for the central role of energy compromise in disease pathogenesis. *Hum Mol Genet*. 2001;10:1215-20.

75. Gollob MH, Green MS, Tang AS, Gollob T, Karibe A, Ali Hassan AS, Ahmad F, Lozado R, Shah G, Fananapazir L, Bachinski LL, Roberts R and Hassan AS. Identification of a gene responsible for familial Wolff-Parkinson-White syndrome. *N Engl J Med*. 2001;344:1823-31.

76. Fatkin D, McConnell BK, Mudd JO, Semsarian C, Moskowitz IG, Schoen FJ, Giewat M, Seidman CE and Seidman JG. An abnormal Ca(2+) response in mutant sarcomere protein-mediated familial hypertrophic cardiomyopathy. *J Clin Invest*. 2000;106:1351-9.

77. Semsarian C, Ahmad I, Giewat M, Georgakopoulos D, Schmitt JP, McConnell BK, Reiken S, Mende U, Marks AR, Kass DA, Seidman CE and Seidman JG. The L-type calcium channel inhibitor diltiazem prevents cardiomyopathy in a mouse model. *J Clin Invest*. 2002;109:1013-20.

78. Guinto PJ, Haim TE, Dowell-Martino CC, Sibinga N and Tardiff JC. Temporal and mutation-specific alterations in Ca2+ homeostasis differentially determine the progression of cTnT-related cardiomyopathies in murine models. *Am J Physiol Heart Circ Physiol*. 2009;297:H614-26.

79. Lan F, Lee AS, Liang P, Sanchez-Freire V, Nguyen PK, Wang L, Han L, Yen M, Wang Y, Sun N, Abilez OJ, Hu S, Ebert AD, Navarrete EG, Simmons CS, Wheeler M, Pruitt B, Lewis R, Yamaguchi Y, Ashley EA, Bers DM, Robbins RC, Longaker MT and Wu JC. Abnormal calcium handling properties underlie familial hypertrophic cardiomyopathy pathology in patient-specific induced pluripotent stem cells. *Cell Stem Cell*. 2013;12:101-13.

80. Dolmetsch RE, Lewis RS, Goodnow CC and Healy JI. Differential activation of transcription factors induced by Ca2+ response amplitude and duration. *Nature*. 1997;386:855-8.

81. Heineke J and Molkentin JD. Regulation of cardiac hypertrophy by intracellular signalling pathways. *Nat Rev Mol Cell Biol*. 2006;7:589-600.

82. Zou Y, Hiroi Y, Uozumi H, Takimoto E, Toko H, Zhu W, Kudoh S, Mizukami M, Shimoyama M, Shibasaki F, Nagai R, Yazaki Y and Komuro I. Calcineurin plays a critical role in the development of pressure overload-induced cardiac hypertrophy. *Circulation*. 2001;104:97-101.

83. Molkentin JD, Lu JR, Antos CL, Markham B, Richardson J, Robbins J, Grant SR and Olson EN. A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell*. 1998;93:215-28.

84. Molkentin JD. Calcineurin-NFAT signaling regulates the cardiac hypertrophic response in coordination with the MAPKs. *Cardiovasc Res*. 2004;63:467-75.

85. Passier R, Zeng H, Frey N, Naya FJ, Nicol RL, McKinsey TA, Overbeek P, Richardson JA, Grant SR and Olson EN. CaM kinase signaling induces cardiac hypertrophy and activates the MEF2 transcription factor in vivo. *J Clin Invest*. 2000;105:1395-406.

86. Backs J and Olson EN. Control of cardiac growth by histone acetylation/deacetylation. *Circ Res*. 2006;98:15-24.

87. Zhang CL, McKinsey TA, Chang S, Antos CL, Hill JA and Olson EN. Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell*. 2002;110:479-88.

88. Konhilas JP, Watson PA, Maass A, Boucek DM, Horn T, Stauffer BL, Luckey SW, Rosenberg P and Leinwand LA. Exercise can prevent and reverse the severity of hypertrophic cardiomyopathy. *Circ Res*. 2006;98:540-8.

89. Saberi S, Wheeler M, Bragg-Gresham J, Hornsby W, Agarwal PP, Attili A, Concannon M, Dries AM, Shmargad Y, Salisbury H, Kumar S, Herrera JJ, Myers J, Helms AS, Ashley EA and Day SM. Effect of Moderate-Intensity Exercise Training on Peak Oxygen Consumption in Patients With Hypertrophic Cardiomyopathy: A Randomized Clinical Trial. *JAMA*. 2017;317:1349-1357.

90. McMullen JR, Amirahmadi F, Woodcock EA, Schinke-Braun M, Bouwman RD, Hewitt KA, Mollica JP, Zhang L, Zhang Y, Shioi T, Buerger A, Izumo S, Jay PY and Jennings GL. Protective effects of exercise and phosphoinositide 3-kinase(p110alpha) signaling in dilated and hypertrophic cardiomyopathy. *Proc Natl Acad Sci U S A*. 2007;104:612-7.

91. Marin TM, Keith K, Davies B, Conner DA, Guha P, Kalaitzidis D, Wu X, Lauriol J, Wang B, Bauer M, Bronson R, Franchini KG, Neel BG and Kontaridis MI. Rapamycin reverses hypertrophic cardiomyopathy in a mouse model of LEOPARD syndrome-associated PTPN11 mutation. *J Clin Invest*. 2011;121:1026-43.

92. Spoladore R, Maron MS, D'Amato R, Camici PG and Olivotto I. Pharmacological treatment options for hypertrophic cardiomyopathy: high time for evidence. *Eur Heart J*. 2012;33:1724-33.

93. Cherian G, Brockington IF, Shah PM, Oakley CM and Goodwin JF. Beta-adrenergic blockade in hypertrophic obstructive cardiomyopathy. *Br Med J*. 1966;1:895-8.

94. Gersh BJ, Maron BJ, Bonow RO, Dearani JA, Fifer MA, Link MS, Naidu SS, Nishimura RA, Ommen SR, Rakowski H, Seidman CE, Towbin JA, Udelson JE, Yancy CW, Guidelines ACoCFAHATFoP, Surgery AAfT, Echocardiography ASo, Cardiology ASoN, America HFSo, Society HR, Interventions SfCAa and Surgeons SoT. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2011;124:2761-96.

95. Cohen LS and Braunwald E. Amelioration of angina pectoris in idiopathic hypertrophic subaortic stenosis with beta-adrenergic blockade. *Circulation*. 1967;35:847-51.

96. Thompson DS, Naqvi N, Juul SM, Swanton RH, Coltart DJ, Jenkins BS and Webb-Peploe MM. Effects of propranolol on myocardial oxygen consumption, substrate extraction, and haemodynamics in hypertrophic obstructive cardiomyopathy. *Br Heart J*. 1980;44:488-98.

97. Gilligan DM, Chan WL, Joshi J, Clarke P, Fletcher A, Krikler S and Oakley CM. A double-blind, placebo-controlled crossover trial of nadolol and verapamil in mild and moderately symptomatic hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 1993;21:1672-9.

98. Tendera M, Wycisk A, Schneeweiss A, Poloński L and Wodniecki J. Effect of sotalol on arrhythmias and exercise tolerance in patients with hypertrophic cardiomyopathy. *Cardiology*. 1993;82:335-42.

99. Frishman WH. Calcium channel blockers: differences between subclasses. *Am J Cardiovasc Drugs*. 2007;7 Suppl 1:17-23.

100. Bonow RO, Ostrow HG, Rosing DR, Cannon RO, Lipson LC, Maron BJ, Kent KM, Bacharach SL and Green MV. Effects of verapamil on left ventricular systolic and diastolic function in patients with hypertrophic cardiomyopathy: pressure-volume analysis with a nonimaging scintillation probe. *Circulation*. 1983;68:1062-73.

101. Bonow RO, Frederick TM, Bacharach SL, Green MV, Goose PW, Maron BJ and Rosing DR. Atrial systole and left ventricular filling in hypertrophic cardiomyopathy: effect of verapamil. *Am J Cardiol*. 1983;51:1386-91.

102. Suwa M, Hirota Y and Kawamura K. Improvement in left ventricular diastolic function during intravenous and oral diltiazem therapy in patients with hypertrophic cardiomyopathy: an echocardiographic study. *Am J Cardiol*. 1984;54:1047-53.

103. Betocchi S, Piscione F, Losi M A, Pace L, Boccalatte M, Perrone-Filardi P, Cappelli-Bigazzi, Briguori C, Manganelli F, Ciampi Q, Salvatore M and Chiariello M. Effects of diltiazem on left ventricular systolic and diastolic function in hypertrophic cardiomyopathy. *Am J Cardiol*. 1996;78:451-7.

104. Sugihara H, Taniguchi Y, Ito K, Terada K, Matsumoto K, Kinoshita N, Azuma A, Ushijima Y, Maeda T and Nakagawa M. Effects of diltiazem on myocardial perfusion abnormalities during exercise in patients with hypertrophic cardiomyopathy. *Ann Nucl Med*. 1998;12:349-54.

105. Matsubara H, Nakatani S, Nagata S, Ishikura F, Katagiri Y, Ohe T and Miyatake K. Salutary effect of disopyramide on left ventricular diastolic function in hypertrophic obstructive cardiomyopathy. *J Am Coll Cardiol*. 1995;26:768-75.

106. Sherrid MV, Barac I, McKenna WJ, Elliott PM, Dickie S, Chojnowska L, Casey S and Maron BJ. Multicenter study of the efficacy and safety of disopyramide in obstructive hypertrophic cardiomyopathy. *J Am Coll Cardiol*. 2005;45:1251-8.

107. Bexton RS, Hellestrand KJ, Cory-Pearce R, Spurrell RA, English TA and Camm AJ. The direct electrophysiologic effects of disopyramide phosphate in the transplanted human heart. *Circulation*. 1983;67:38-45.

108. Hamada M, Shigematsu Y, Ikeda S, Hara Y, Okayama H, Kodama K, Ochi T and Hiwada K. Class Ia antiarrhythmic drug cibenzoline: a new approach to the medical treatment of hypertrophic obstructive cardiomyopathy. *Circulation*. 1997;96:1520-4.

109. Kondo I, Mizushige K, Nozaki S, Iwado Y, Masugata H, Kohno M and Matsuo H. Effect of cibenzoline, a class ia antiarrhythmic agent, on left ventricular diastolic function in hypertrophic cardiomyopathy. *Cardiovasc Drugs Ther*. 2001;15:459-65.

110. Hamada M, Ikeda S, Ohshima K, Nakamura M, Kubota N, Ogimoto A and Shigematsu Y. Impact of chronic use of cibenzoline on left ventricular pressure gradient and left ventricular remodeling in patients with hypertrophic obstructive cardiomyopathy. *J Cardiol*. 2016;67:279-86.

111. Tsybouleva N, Zhang L, Chen S, Patel R, Lutucuta S, Nemoto S, DeFreitas G, Entman M, Carabello BA, Roberts R and Marian AJ. Aldosterone, through novel signaling proteins, is a fundamental molecular bridge between the genetic defect and the cardiac phenotype of hypertrophic cardiomyopathy. *Circulation*. 2004;109:1284-91.

112. Lim DS, Lutucuta S, Bachireddy P, Youker K, Evans A, Entman M, Roberts R and Marian AJ. Angiotensin II blockade reverses myocardial fibrosis in a transgenic mouse model of human hypertrophic cardiomyopathy. *Circulation*. 2001;103:789-91.

113. Kyriakidis M, Triposkiadis F, Dernellis J, Androulakis AE, Mellas P, Kelepeshis GA and Gialafos JE. Effects of cardiac versus circulatory angiotensin-converting enzyme inhibition on left ventricular diastolic function and coronary blood flow in hypertrophic obstructive cardiomyopathy. *Circulation*. 1998;97:1342-7.

114. Kawano H, Toda G, Nakamizo R, Koide Y, Seto S and Yano K. Valsartan decreases type I collagen synthesis in patients with hypertrophic cardiomyopathy. *Circ J*. 2005;69:1244-8.

115. Araujo AQ, Arteaga E, Ianni BM, Buck PC, Rabello R and Mady C. Effect of Losartan on left ventricular diastolic function in patients with nonobstructive hypertrophic cardiomyopathy. *Am J Cardiol*. 2005;96:1563-7.

116. Axelsson A, Iversen K, Vejlstrup N, Ho CY, Havndrup O, Kofoed KF, Norsk J, Jensen M and Bundgaard H. Functional effects of losartan in hypertrophic cardiomyopathy-a randomised clinical trial. *Heart*. 2016;102:285-91.

117. Mourad JJ, Hanon O, Deverre JR, Camici PG, Sellier P, Duboc D and Safar ME. Improvement of impaired coronary vasodilator reserve in hypertensive patients by low-dose ACE inhibitor/diuretic therapy: a pilot PET study. *J Renin Angiotensin Aldosterone Syst*. 2003;4:94-5.

118. Neglia D, Fommei E, Varela-Carver A, Mancini M, Ghione S, Lombardi M, Pisani P, Parker H, D'amati G, Donato L and Camici PG. Perindopril and indapamide reverse coronary

microvascular remodelling and improve flow in arterial hypertension. *J Hypertens*. 2011;29:364-72.

119. Abozguia K, Elliott P, McKenna W, Phan TT, Nallur-Shivu G, Ahmed I, Maher AR, Kaur K, Taylor J, Henning A, Ashrafian H, Watkins H and Frenneaux M. Metabolic modulator perhexiline corrects energy deficiency and improves exercise capacity in symptomatic hypertrophic cardiomyopathy. *Circulation*. 2010;122:1562-9.

120. Green EM, Wakimoto H, Anderson RL, Evanchik MJ, Gorham JM, Harrison BC, Henze M, Kawas R, Oslob JD, Rodriguez HM, Song Y, Wan W, Leinwand LA, Spudich JA, McDowell RS, Seidman JG and Seidman CE. A small-molecule inhibitor of sarcomere contractility suppresses hypertrophic cardiomyopathy in mice. *Science*. 2016;351:617-21.

121. Maass A and Leinwand LA. Animal models of hypertrophic cardiomyopathy. *Curr Opin Cardiol*. 2000;15:189-96.

122. Zhu Z and Huangfu D. Human pluripotent stem cells: an emerging model in developmental biology. *Development*. 2013;140:705-17.

123. Milani-Nejad N and Janssen PM. Small and large animal models in cardiac contraction research: advantages and disadvantages. *Pharmacol Ther*. 2014;141:235-49.

124. Haghighi K, Kolokathis F, Pater L, Lynch RA, Asahi M, Gramolini AO, Fan GC, Tsiapras D, Hahn HS, Adamopoulos S, Liggett SB, Dorn GW, MacLennan DH, Kremastinos DT and Kranias EG. Human phospholamban null results in lethal dilated cardiomyopathy revealing a critical difference between mouse and human. *J Clin Invest*. 2003;111:869-76.

125. Wilson K, Faelan C, Patterson-Kane JC, Rudmann DG, Moore SA, Frank D, Charleston J, Tinsley J, Young GD and Milici AJ. Duchenne and Becker Muscular Dystrophies: A Review of Animal Models, Clinical End Points, and Biomarker Quantification. *Toxicol Pathol*. 2017;45:961-976.

126. Delfín DA, Zang KE, Schill KE, Patel NT, Janssen PM, Raman SV and Rafael-Fortney JA. Cardiomyopathy in the dystrophin/utrophin-deficient mouse model of severe muscular dystrophy is characterized by dysregulation of matrix metalloproteinases. *Neuromuscul Disord*. 2012;22:1006-14.

127. Bulfield G, Siller WG, Wight PA and Moore KJ. X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc Natl Acad Sci U S A*. 1984;81:1189-92.

128. Miyata S, Minobe W, Bristow MR and Leinwand LA. Myosin heavy chain isoform expression in the failing and nonfailing human heart. *Circ Res*. 2000;86:386-90.

129. Patel R, Nagueh SF, Tsybouleva N, Abdellatif M, Lutucuta S, Kopelen HA, Quinones MA, Zoghbi WA, Entman ML, Roberts R and Marian AJ. Simvastatin induces regression of cardiac hypertrophy and fibrosis and improves cardiac function in a transgenic rabbit model of human hypertrophic cardiomyopathy. *Circulation*. 2001;104:317-24.

130. Senthil V, Chen SN, Tsybouleva N, Halder T, Nagueh SF, Willerson JT, Roberts R and Marian AJ. Prevention of cardiac hypertrophy by atorvastatin in a transgenic rabbit model of human hypertrophic cardiomyopathy. *Circ Res*. 2005;97:285-92.

131. Nagueh SF, Lombardi R, Tan Y, Wang J, Willerson JT and Marian AJ. Atorvastatin and cardiac hypertrophy and function in hypertrophic cardiomyopathy: a pilot study. *Eur J Clin Invest*. 2010;40:976-83.

132. Carvajal-Vergara X, Sevilla A, D'Souza SL, Ang YS, Schaniel C, Lee DF, Yang L, Kaplan AD, Adler ED, Rozov R, Ge Y, Cohen N, Edelmann LJ, Chang B, Waghray A, Su J, Pardo S, Lichtenbelt KD, Tartaglia M, Gelb BD and Lemischka IR. Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature*. 2010;465:808-12.

133. Lin B, Li Y, Han L, Kaplan AD, Ao Y, Kalra S, Bett GC, Rasmusson RL, Denning C and Yang L. Modeling and study of the mechanism of dilated cardiomyopathy using induced pluripotent stem cells derived from individuals with Duchenne muscular dystrophy. *Dis Model Mech*. 2015;8:457-66.

134. Dick E, Kalra S, Anderson D, George V, Ritso M, Laval SH, Barresi R, Aartsma-Rus A, Lochmüller H and Denning C. Exon skipping and gene transfer restore dystrophin expression in human induced pluripotent stem cells-cardiomyocytes harboring DMD mutations. *Stem Cells Dev*. 2013;22:2714-24.

135. Sun N, Yazawa M, Liu J, Han L, Sanchez-Freire V, Abilez OJ, Navarrete EG, Hu S, Wang L, Lee A, Pavlovic A, Lin S, Chen R, Hajjar RJ, Snyder MP, Dolmetsch RE, Butte MJ, Ashley EA, Longaker MT, Robbins RC and Wu JC. Patient-specific induced pluripotent stem cells as a model for familial dilated cardiomyopathy. *Sci Transl Med*. 2012;4:130ra47.

136. Itzhaki I, Maizels L, Huber I, Zwi-Dantsis L, Caspi O, Winterstern A, Feldman O, Gepstein A, Arbel G, Hammerman H, Boulos M and Gepstein L. Modelling the long QT syndrome with induced pluripotent stem cells. *Nature*. 2011;471:225-9.

137. Matsa E, Burridge PW and Wu JC. Human stem cells for modeling heart disease and for drug discovery. *Sci Transl Med*. 2014;6:239ps6.

138. Karakikes I, Ameen M, Termglinchan V and Wu JC. Human induced pluripotent stem cell-derived cardiomyocytes: insights into molecular, cellular, and functional phenotypes. *Circ Res*. 2015;117:80-8.

139. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB and Subcommittee AHASCaSS. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation*. 2015;131:e29-322.

140. Bui AL, Horwich TB and Fonarow GC. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol*. 2011;8:30-41.

141. McMurray JJ and Pfeffer MA. Heart failure. *Lancet*. 2005;365:1877-89.

142. Mudd JO and Kass DA. Tackling heart failure in the twenty-first century. *Nature*. 2008;451:919-28.

143. Drake TA and Ping P. Thematic review series: systems biology approaches to metabolic and cardiovascular disorders. Proteomics approaches to the systems biology of cardiovascular diseases. *J Lipid Res*. 2007;48:1-8.

144. Sabidó E, Selevsek N and Aebersold R. Mass spectrometry-based proteomics for systems biology. *Curr Opin Biotechnol*. 2012;23:591-7.

145. Pandey A and Mann M. Proteomics to study genes and genomes. *Nature*. 2000;405:837-46.

146. Smith LM, Kelleher NL and Proteomics CfTD. Proteoform: a single term describing protein complexity. *Nat Methods*. 2013;10:186-7.

147. Gregorich ZR and Ge Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*. 2014;14:1195-210.

148. Mann M and Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003;21:255-61.

149. Karve TM and Cheema AK. Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease. *J Amino Acids*. 2011;2011:207691.

150. Geraldes P and King GL. Activation of protein kinase C isoforms and its impact on diabetic complications. *Circ Res*. 2010;106:1319-31.

151. Marston SB and Redwood CS. Modulation of thin filament activation by breakdown or isoform switching of thin filament proteins: physiological and pathological implications. *Circ Res*. 2003;93:1170-8.

152. Risco A and Cuenda A. New Insights into the p38γ and p38δ MAPK Pathways. *J Signal Transduct*. 2012;2012:520289.

153. Gregorich ZR, Chang YH and Ge Y. Proteomics in heart failure: top-down or bottom-up? *Pflugers Arch*. 2014;466:1199-209.

154. Chait BT. Chemistry. Mass spectrometry: bottom-up or top-down? *Science*. 2006;314:65-6.

155. Zhang H and Ge Y. Comprehensive analysis of protein modifications by top-down mass spectrometry. *Circ Cardiovasc Genet*. 2011;4:711.

156. Dong X, Sumandea CA, Chen YC, Garcia-Cazarin ML, Zhang J, Balke CW, Sumandea MP and Ge Y. Augmented phosphorylation of cardiac troponin I in hypertensive heart failure. *J Biol Chem*. 2012;287:848-57.

157. Zhang J, Guy MJ, Norman HS, Chen YC, Xu Q, Dong X, Guner H, Wang S, Kohmoto T, Young KH, Moss RL and Ge Y. Top-down quantitative proteomics identified phosphorylation of cardiac troponin I as a candidate biomarker for chronic heart failure. *J Proteome Res*. 2011;10:4054-65.

158. Ge Y, Rybakova IN, Xu Q and Moss RL. Top-down high-resolution mass spectrometry of cardiac myosin binding protein C revealed that truncation alters protein phosphorylation state. *Proc Natl Acad Sci U S A*. 2009;106:12658-63.

159. Peng Y, Ayaz-Guner S, Yu D and Ge Y. Top-down mass spectrometry of cardiac myofilament proteins in health and disease. *Proteomics Clin Appl*. 2014;8:554-68.

160. Gregorich ZR, Peng Y, Lane NM, Wolff JJ, Wang S, Guo W, Guner H, Doop J, Hacker TA and Ge Y. Comprehensive assessment of chamber-specific and transmural heterogeneity in myofilament protein phosphorylation by top-down mass spectrometry. *J Mol Cell Cardiol*. 2015;87:102-112.

161. Carroll J, Fearnley IM and Walker JE. Definition of the mitochondrial proteome by measurement of molecular masses of membrane proteins. *Proc Natl Acad Sci U S A*. 2006;103:16170-5.

162. Catherman AD, Li M, Tran JC, Durbin KR, Compton PD, Early BP, Thomas PM and Kelleher NL. Top down proteomics of human membrane proteins from enriched mitochondrial fractions. *Anal Chem*. 2013;85:1880-8.

163. Chang YH, Gregorich ZR, Chen AJ, Hwang L, Guner H, Yu D, Zhang J and Ge Y. New mass-spectrometry-compatible degradable surfactant for tissue proteomics. *J Proteome Res*. 2015;14:1587-99.

164. Ayaz-Guner S, Zhang J, Li L, Walker JW and Ge Y. In vivo phosphorylation site mapping in mouse cardiac troponin I by high resolution top-down electron capture dissociation mass spectrometry: Ser22/23 are the only sites basally phosphorylated. *Biochemistry*. 2009;48:8161-70.

165. Sterling HJ, Batchelor JD, Wemmer DE and Williams ER. Effects of buffer loading for electrospray ionization mass spectrometry of a noncovalent protein complex that requires high concentrations of essential salts. *J Am Soc Mass Spectrom*. 2010;21:1045-9.

166. Laganowsky A, Reading E, Hopper JT and Robinson CV. Mass spectrometry of intact membrane protein complexes. *Nat Protoc*. 2013;8:639-51.

167. Xiu L, Valeja SG, Alpert AJ, Jin S and Ge Y. Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. *Anal Chem*. 2014;86:7899-906.

168. Tran JC and Doucette AA. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal Chem*. 2008;80:1568-73.

169. Mehmood S, Allison TM and Robinson CV. Mass spectrometry of protein complexes: from origins to applications. *Annu Rev Phys Chem*. 2015;66:453-74.

170. Hopper JT, Yu YT, Li D, Raymond A, Bostock M, Liko I, Mikhailov V, Laganowsky A, Benesch JL, Caffrey M, Nietlispach D and Robinson CV. Detergent-free mass spectrometry of membrane protein complexes. *Nat Methods*. 2013;10:1206-8.

171. Whitelegge JP. HPLC and mass spectrometry of intrinsic membrane proteins. *Methods Mol Biol*. 2004;251:323-40.

172. Whitelegge J. Intact protein mass spectrometry and top-down proteomics. *Expert Rev Proteomics*. 2013;10:127-9.

173. Souda P, Ryan CM, Cramer WA and Whitelegge J. Profiling of integral membrane proteins and their post translational modifications using high-resolution mass spectrometry. *Methods*. 2011;55:330-6.

174. Doucette AA, Tran JC, Wall MJ and Fitzsimmons S. Intact proteome fractionation strategies compatible with mass spectrometry. *Expert Rev Proteomics*. 2011;8:787-800.

175. Zhang J, Dong X, Hacker TA and Ge Y. Deciphering modifications in swine cardiac troponin I by top-down high-resolution tandem mass spectrometry. *J Am Soc Mass Spectrom*. 2010;21:940-8.

176. Wang Y, Pinto JR, Solis RS, Dweck D, Liang J, Diaz-Perez Z, Ge Y, Walker JW and Potter JD. Generation and functional characterization of knock-in mice harboring the cardiac troponin I-R21C mutation associated with hypertrophic cardiomyopathy. *J Biol Chem*. 2012;287:2156-67.

177. Sancho Solis R, Ge Y and Walker JW. Single amino acid sequence polymorphisms in rat cardiac troponin revealed by top-down tandem mass spectrometry. *J Muscle Res Cell Motil*. 2008;29:203-12.

178. Tran JC and Doucette AA. Multiplexed size separation of intact proteins in solution phase for mass spectrometry. *Anal Chem*. 2009;81:6201-9.

179. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM and Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011;480:254-8.

180. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM and Kelleher NL. Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol Cell Proteomics*. 2013;12:3465-73.

181. Neverova I and Van Eyk JE. Role of chromatographic techniques in proteomic analysis. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2005;815:51-63.

182. Ansong C, Wu S, Meng D, Liu X, Brewer HM, Deatherage Kaiser BL, Nakayasu ES, Cort JR, Pevzner P, Smith RD, Heffron F, Adkins JN and Pasa-Tolic L. Top-down proteomics reveals a unique protein S-thiolation switch in Salmonella Typhimurium in response to infection-like conditions. *Proc Natl Acad Sci U S A*. 2013;110:10153-8.

183. Chen B, Peng Y, Valeja SG, Xiu L, Alpert AJ and Ge Y. Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-Down Proteomics. *Anal Chem*. 2016;88:1885-91.

184. Valeja SG, Xiu L, Gregorich ZR, Guner H, Jin S and Ge Y. Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. *Anal Chem*. 2015;87:5363-71.

185. Muneeruddin K, Nazzaro M and Kaltashov IA. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. *Anal Chem*. 2015;87:10138-45.

186. Compton PD, Zamdborg L, Thomas PM and Kelleher NL. On the scalability and requirements of whole protein mass spectrometry. *Anal Chem*. 2011;83:6868-74.

187. Alpert AJ. Size Exclusion High-Performance Liquid

Chromatography of Small Solutes *Column Handbook for*

Size Exclusion Chromatography San Diego: Academic Press; 1999: 249-66.

188. Chen X and Ge Y. Ultrahigh pressure fast size exclusion chromatography for top-down proteomics. *Proteomics*. 2013;13:2563-6.

189. Sharma S, Simpson DC, Tolić N, Jaitly N, Mayampurath AM, Smith RD and Pasa-Tolić L. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J Proteome Res*. 2007;6:602-10.

190. Karas M and Krüger R. Ion formation in MALDI: the cluster ionization mechanism. *Chem Rev*. 2003;103:427-40.

191. Fenn JB, Mann M, Meng CK, Wong SF and Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989;246:64-71.

192. Scherperel G and Reid GE. Emerging methods in proteomics: top-down protein characterization by multistage tandem mass spectrometry. *Analyst*. 2007;132:500-6.

193. Mann M and Kelleher NL. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A*. 2008;105:18132-8.

194. Kelleher NL, Senko MW, Siegel MM and McLafferty FW. Unit resolution mass spectra of 112 kDa molecules with 3 Da accuracy. *Journal of the American Society for Mass Spectrometry*. 1997;8:380-383.

195. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP and McLafferty FW. Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry. *J Am Chem Soc*. 2002;124:672-8.

196. Han X, Jin M, Breuker K and McLafferty FW. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science*. 2006;314:109-12.

197. Chen X, Westphall MS and Smith LM. Mass spectrometric analysis of DNA mixtures: instrumental effects responsible for decreased sensitivity with increasing mass. *Anal Chem*. 2003;75:5944-52.

198. Park J, Qin H, Scalf M, Hilger RT, Westphall MS, Smith LM and Blick RH. A mechanical nanomembrane detector for time-of-flight mass spectrometry. *Nano Lett*. 2011;11:3681-4.

199. Beck S, Michalski A, Raether O, Lubeck M, Kaspar S, Goedecke N, Baessmann C, Hornburg D, Meier F, Paron I, Kulak NA, Cox J and Mann M. The Impact II, a Very High-Resolution Quadrupole Time-of-Flight Instrument (QTOF) for Deep Shotgun Proteomics. *Mol Cell Proteomics*. 2015;14:2014-29.

200. Makarov A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal Chem*. 2000;72:1156-62.

201. Perry RH, Cooks RG and Noll RJ. Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom Rev*. 2008;27:661-99.

202. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM and Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*. 2011;480:254-8.

203. Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M and Kelleher NL. Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. *J Proteome Res*. 2016;15:976-82.

204. Valeja SG, Kaiser NK, Xian F, Hendrickson CL, Rouse JC and Marshall AG. Unit mass baseline resolution for an intact 148 kDa therapeutic monoclonal antibody by Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem*. 2011;83:8391-5.

205. Kostyukevich YI, Vladimirov GN and Nikolaev EN. Dynamically harmonized FT-ICR cell with specially shaped electrodes for compensation of inhomogeneity of the magnetic field. Computer simulations of the electric field and ion motion dynamics. *J Am Soc Mass Spectrom*. 2012;23:2198-207.

206. McLafferty FW, Breuker K, Jin M, Han X, Infusini G, Jiang H, Kong X and Begley TP. Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS J*. 2007;274:6256-68.

207. Siuti N and Kelleher NL. Decoding protein modifications using top-down mass spectrometry. *Nat Methods*. 2007;4:817-21.

208. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK and McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*. 2000;72:563-73.

209. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J and Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 2004;101:9528-33.

210. Miladinović SM, Fornelli L, Lu Y, Piech KM, Girault HH and Tsybin YO. In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Anal Chem*. 2012;84:4647-51.

211. McLuckey SA, Reid GE and Wells JM. Ion parking during ion/ion reactions in electrodynamic ion traps. *Anal Chem*. 2002;74:336-46.

212. Huang TY and McLuckey SA. Top-down protein characterization facilitated by ion/ion reactions on a quadrupole/time of flight platform. *Proteomics*. 2010;10:3577-88.

213. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL and Brodbelt JS. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. *J Am Chem Soc*. 2013;135:12646-51.

214. Pesavento JJ, Mizzen CA and Kelleher NL. Quantitative analysis of modified proteins and their positional isomers by tandem mass spectrometry: human histone H4. *Anal Chem*. 2006;78:4271-80.

215. Chamot-Rooke J, Mikaty G, Malosse C, Soyer M, Dumont A, Gault J, Imhaus AF, Martin P, Trellet M, Clary G, Chafey P, Camoin L, Nilges M, Nassif X and Duménil G. Posttranslational modification of pili upon cell contact triggers N. meningitidis dissemination. *Science*. 2011;331:778-82.

216. Du Y, Parks BA, Sohn S, Kwast KE and Kelleher NL. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Anal Chem*. 2006;78:686-94.

217. Waanders LF, Hanke S and Mann M. Top-down quantitation and characterization of SILAC-labeled proteins. *J Am Soc Mass Spectrom*. 2007;18:2058-64.

218. Bergmann U, Ahrends R, Neumann B, Scheler C and Linscheid MW. Application of metal-coded affinity tags (MeCAT): absolute protein quantification with top-down and bottom-up workflows by metal-coded tagging. *Anal Chem*. 2012;84:5268-75.

219. Mazur MT, Cardasis HL, Spellman DS, Liaw A, Yates NA and Hendrickson RC. Quantitative analysis of intact apolipoproteins in human HDL by top-down differential mass spectrometry. *Proc Natl Acad Sci U S A*. 2010;107:7728-33.

220. Horn DM, Zubarev RA and McLafferty FW. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom*. 2000;11:320-32.

221. Senko MW, Beu SC and McLaffertycor FW. Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions. *J Am Soc Mass Spectrom*. 1995;6:229-33.

222. Liu X, Inbar Y, Dorrestein PC, Wynne C, Edwards N, Souda P, Whitelegge JP, Bafna V and Pevzner PA. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol Cell Proteomics*. 2010;9:2772-82.

223. Marty MT, Baldwin AJ, Marklund EG, Hochberg GK, Benesch JL and Robinson CV. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem*. 2015;87:4370-6.

224. Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA and Smith RD. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. *BMC Bioinformatics*. 2009;10:87.

225. Zamdborg L, LeDuc RD, Glowacz KJ, Kim YB, Viswanathan V, Spaulding IT, Early BP, Bluhm EJ, Babai S and Kelleher NL. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res*. 2007;35:W701-6.

226. Karabacak NM, Li L, Tiwari A, Hayward LJ, Hong P, Easterling ML and Agar JN. Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry. *Mol Cell Proteomics*. 2009;8:846-56.

227. Frank AM, Pesavento JJ, Mizzen CA, Kelleher NL and Pevzner PA. Interpreting top-down mass spectra using spectral alignment. *Anal Chem*. 2008;80:2499-505.

228. Liu X, Sirotkin Y, Shen Y, Anderson G, Tsai YS, Ting YS, Goodlett DR, Smith RD, Bafna V and Pevzner PA. Protein identification using top-down. *Mol Cell Proteomics*. 2012;11:M111.008524.

229. Cai W, Guner H, Gregorich ZR, Chen AJ, Ayaz-Guner S, Peng Y, Valeja SG, Liu X and Ge Y. MASH Suite Pro: A Comprehensive Software Tool for Top-down Proteomics. *Mol Cell Proteomics*. 2015.

230. de Tombe PP and Solaro RJ. Integration of cardiac myofilament activity and regulation with pathways signaling hypertrophy and failure. *Ann Biomed Eng*. 2000;28:991-1001.

231. Yin Z, Ren J and Guo W. Sarcomeric protein isoform transitions in cardiac muscle: a journey to heart failure. *Biochim Biophys Acta*. 2015;1852:47-52.

232. Peng Y, Yu D, Gregorich Z, Chen X, Beyer AM, Gutterman DD and Ge Y. In-depth proteomic analysis of human tropomyosin by top-down mass spectrometry. *J Muscle Res Cell Motil*. 2013;34:199-210.

233. Peng Y, Chen X, Zhang H, Xu Q, Hacker TA and Ge Y. Top-down targeted proteomics for deep sequencing of tropomyosin isoforms. *J Proteome Res*. 2013;12:187-98.

234. Copeland O, Nowak KJ, Laing NG, Ravenscroft G, Messer AE, Bayliss CR and Marston SB. Investigation of changes in skeletal muscle alpha-actin expression in normal and pathological human and mouse hearts. *J Muscle Res Cell Motil*. 2010;31:207-14.

235. Chen YC, Ayaz-Guner S, Peng Y, Lane NM, Locher MR, Kohmoto T, Larsson L, Moss RL and Ge Y. Effective Top-Down LC/MS+ Method for Assessing Actin Isoforms as a Potential Cardiac Disease Marker. *Anal Chem*. 2015;87:8399-406.

236. Gry M, Rimini R, Strömberg S, Asplund A, Pontén F, Uhlén M and Nilsson P. Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics*. 2009;10:365.

237. Solaro RJ and van der Velden J. Why does troponin I have so many phosphorylation sites? Fact and fancy. *J Mol Cell Cardiol*. 2010;48:810-6.

238. Hunter T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell*. 2007;28:730-8.

239. de Tombe PP. Myosin binding protein C in the heart. *Circ Res*. 2006;98:1234-6.

240. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Isasi CR, Jiménez MC, Judd SE, Kissela BM, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Magid DJ, McGuire DK, Mohler ER, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Rosamond W, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Woo D, Yeh RW, Turner MB and Subcommittee AHASCaSS. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. *Circulation*. 2016;133:e38-e360.

241. Houser SR, Margulies KB, Murphy AM, Spinale FG, Francis GS, Prabhu SD, Rockman HA, Kass DA, Molkentin JD, Sussman MA, Koch WJ, Koch W and American Heart Association Council on Basic Cardiovascular Sciences CuoCC, and Council on Functional Genomics and

Translational Biology. Animal models of heart failure: a scientific statement from the American Heart Association. *Circ Res*. 2012;111:131-50.

242. Walker LA, Medway AM, Walker JS, Cleveland JC and Buttrick PM. Tissue procurement strategies affect the protein biochemistry of human heart samples. *J Muscle Res Cell Motil*. 2011;31:309-14.

243. Yuan C and Solaro RJ. Myofilament proteins: From cardiac disorders to proteomic changes. *Proteomics Clin Appl*. 2008;2:788-99.

244. Cai W, Tucholski TM, Gregorich ZR and Ge Y. Top-down Proteomics: Technology Advancements and Applications to Heart Diseases. *Expert Rev Proteomics*. 2016;13:717-30.

245. Layland J, Solaro RJ and Shah AM. Regulation of cardiac contractile function by troponin I phosphorylation. *Cardiovasc Res*. 2005;66:12-21.

246. Yar S, Monasky MM and Solaro RJ. Maladaptive modifications in myofilament proteins and triggers in the progression to heart failure and sudden death. *Pflugers Arch*. 2014;466:1189-97.

247. Messer AE, Jacques AM and Marston SB. Troponin phosphorylation and regulatory function in human heart muscle: dephosphorylation of Ser23/24 on troponin I could account for the contractile defect in end-stage heart failure. *J Mol Cell Cardiol*. 2007;42:247-59.

248. Sadayappan S, Gulick J, Osinska H, Martin LA, Hahn HS, Dorn GW, Klevitsky R, Seidman CE, Seidman JG and Robbins J. Cardiac myosin-binding protein-C phosphorylation and cardiac function. *Circ Res*. 2005;97:1156-63.

249. Sheikh F, Ouyang K, Campbell SG, Lyon RC, Chuang J, Fitzsimons D, Tangney J, Hidalgo CG, Chung CS, Cheng H, Dalton ND, Gu Y, Kasahara H, Ghassemian M, Omens JH, Peterson KL, Granzier HL, Moss RL, McCulloch AD and Chen J. Mouse and computational models link Mlc2v dephosphorylation to altered myosin kinetics in early cardiac disease. *J Clin Invest*. 2012;122:1209-21.

250. Pyle WG and Solaro RJ. At the crossroads of myocardial signaling: the role of Z-discs in intracellular signaling and cardiac function. *Circ Res*. 2004;94:296-305.

251. Kirk JA, Holewinski RJ, Kooij V, Agnetti G, Tunin RS, Witayavanitkul N, de Tombe PP, Gao WD, Van Eyk J and Kass DA. Cardiac resynchronization sensitizes the sarcomere to calcium by reactivating GSK-3β. *J Clin Invest*. 2014;124:129-38.

252. Catherman AD, Skinner OS and Kelleher NL. Top Down proteomics: facts and perspectives. *Biochem Biophys Res Commun*. 2014;445:683-93.

253. Savaryn JP, Catherman AD, Thomas PM, Abecassis MM and Kelleher NL. The emergence of top-down proteomics in clinical research. *Genome Med*. 2013;5:53.

254. Gregorich ZR, Peng Y, Cai W, Jin Y, Wei L, Chen AJ, McKiernan SH, Aiken JM, Moss RL, Diffee GM and Ge Y. Top-Down Targeted Proteomics Reveals Decrease in Myosin Regulatory Light-Chain Phosphorylation That Contributes to Sarcopenic Muscle Dysfunction. *J Proteome Res*. 2016.

255. Guner H, Close PL, Cai W, Zhang H, Peng Y, Gregorich ZR and Ge Y. MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J Am Soc Mass Spectrom*. 2014;25:464-70.

256. Yeung YG and Stanley ER. A solution for stripping antibodies from polyvinylidene fluoride immunoblots for multiple reprobing. *Anal Biochem*. 2009;389:89-91.

257. Geeves MA, Hitchcock-DeGregori SE and Gunning PW. A systematic nomenclature for mammalian tropomyosin isoforms. *J Muscle Res Cell Motil*. 2015;36:147-53.

258. Katrukha IA and Gusev NB. Enigmas of cardiac troponin T phosphorylation. *J Mol Cell Cardiol*. 2013;65:156-8.

259. Scruggs SB, Reisdorph R, Armstrong ML, Warren CM, Reisdorph N, Solaro RJ and Buttrick PM. A novel, in-solution separation of endogenous cardiac sarcomeric proteins and identification of distinct charged variants of regulatory light chain. *Mol Cell Proteomics*. 2010;9:1804-18.

260. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*. 2004;15:12-6.

261. Solaro RJ, Moir AJ and Perry SV. Phosphorylation of troponin I and the inotropic effect of adrenaline in the perfused rabbit heart. *Nature*. 1976;262:615-7.

262. Jaffe AS, Ravkilde J, Roberts R, Naslund U, Apple FS, Galvani M and Katus H. It's time for a change to a troponin standard. *Circulation*. 2000;102:1216-20.

263. Solaro RJ and Rarick HM. Troponin and tropomyosin: proteins that switch on and tune in the activity of cardiac myofilaments. *Circ Res*. 1998;83:471-80.

264. Sasse S, Brand NJ, Kyprianou P, Dhoot GK, Wade R, Arai M, Periasamy M, Yacoub MH and Barton PJ. Troponin I gene expression during human cardiac development and in end-stage heart failure. *Circ Res*. 1993;72:932-8.

265. Zhang R, Zhao J, Mandveno A and Potter JD. Cardiac troponin I phosphorylation increases the rate of cardiac muscle relaxation. *Circ Res*. 1995;76:1028-35.

266. Noland TA, Raynor RL and Kuo JF. Identification of sites phosphorylated in bovine cardiac troponin I and troponin T by protein kinase C and comparative substrate activity of synthetic peptides containing the phosphorylation sites. *J Biol Chem*. 1989;264:20778-85.

267. Noland TA, Guo X, Raynor RL, Jideama NM, Averyhart-Fullard V, Solaro RJ and Kuo JF. Cardiac troponin I mutants. Phosphorylation by protein kinases C and A and regulation of Ca(2+)-stimulated MgATPase of reconstituted actomyosin S-1. *J Biol Chem*. 1995;270:25445-54.

268. Blumenthal DK, Stull JT and Gill GN. Phosphorylation of cardiac troponin by guanosine 3':5'-monophosphate-dependent protein kinase. *J Biol Chem*. 1978;253:324-6.

269. Yuasa K, Michibata H, Omori K and Yanaka N. A novel interaction of cGMP-dependent protein kinase I with troponin T. *J Biol Chem*. 1999;274:37429-34.

270. Purcell IF, Bing W and Marston SB. Functional analysis of human cardiac troponin by the in vitro motility assay: comparison of adult, foetal and failing hearts. *Cardiovasc Res*. 1999;43:884-91.

271. van der Velden J, Papp Z, Zaremba R, Boontje NM, de Jong JW, Owen VJ, Burton PB, Goldmann P, Jaquet K and Stienen GJ. Increased Ca2+-sensitivity of the contractile apparatus in end-stage human heart failure results from altered phosphorylation of contractile proteins. *Cardiovasc Res*. 2003;57:37-47.

272. Bodor GS, Oakeley AE, Allen PD, Crimmins DL, Ladenson JH and Anderson PA. Troponin I phosphorylation in the normal and failing adult human heart. *Circulation*. 1997;96:1495-500.

273. Belin RJ, Sumandea MP, Allen EJ, Schoenfelt K, Wang H, Solaro RJ and de Tombe PP. Augmented protein kinase C-alpha-induced myofilament protein phosphorylation contributes to myofilament dysfunction in experimental congestive heart failure. *Circ Res*. 2007;101:195-204.

274. Kranias EG, Garvey JL, Srivastava RD and Solaro RJ. Phosphorylation and functional modifications of sarcoplasmic reticulum and myofibrils in isolated rabbit hearts stimulated with isoprenaline. *Biochem J*. 1985;226:113-21.

275. James P, Inui M, Tada M, Chiesi M and Carafoli E. Nature and site of phospholamban regulation of the Ca2+ pump of sarcoplasmic reticulum. *Nature*. 1989;342:90-2.

276. Gautel M, Zuffardi O, Freiburg A and Labeit S. Phosphorylation switches specific for the cardiac isoform of myosin binding protein-C: a modulator of cardiac contraction? *EMBO J*. 1995;14:1952-60.

277. Colson BA, Bekyarova T, Locher MR, Fitzsimons DP, Irving TC and Moss RL. Protein kinase A-mediated phosphorylation of cMyBP-C increases proximity of myosin heads to actin in resting myocardium. *Circ Res*. 2008;103:244-51.

278. Cai W, Tucholski T, Chen B, Alpert AJ, McIlwain S, Kohmoto T, Jin S and Ge Y. Top-Down Proteomics of Large Proteins up to 223 kDa Enabled by Serial Size Exclusion Chromatography Strategy. *Anal Chem*. 2017;89:5467-5475.

279. Kentish JC, McCloskey DT, Layland J, Palmer S, Leiden JM, Martin AF and Solaro RJ. Phosphorylation of troponin I by protein kinase A accelerates relaxation and crossbridge cycle kinetics in mouse ventricular muscle. *Circ Res*. 2001;88:1059-65.

280. Takeishi Y, Chu G, Kirkpatrick DM, Li Z, Wakasaki H, Kranias EG, King GL and Walsh RA. In vivo phosphorylation of cardiac troponin I by protein kinase Cbeta2 decreases cardiomyocyte calcium responsiveness and contractility in transgenic mouse hearts. *J Clin Invest*. 1998;102:72-8.

281. Westfall MV and Borton AR. Role of troponin I phosphorylation in protein kinase C-mediated enhanced contractile performance of rat myocytes. *J Biol Chem*. 2003;278:33694-700.

282. Bowling N, Walsh RA, Song G, Estridge T, Sandusky GE, Fouts RL, Mintze K, Pickard T, Roden R, Bristow MR, Sabbah HN, Mizrahi JL, Gromo G, King GL and Vlahos CJ. Increased protein kinase C activity and expression of Ca2+-sensitive isoforms in the failing human heart. *Circulation*. 1999;99:384-91.

283. Buscemi N, Foster DB, Neverova I and Van Eyk JE. p21-activated kinase increases the calcium sensitivity of rat triton-skinned cardiac muscle fiber bundles via a mechanism potentially involving novel phosphorylation of troponin I. *Circ Res*. 2002;91:509-16.

284. Zhang P, Kirk JA, Ji W, dos Remedios CG, Kass DA, Van Eyk JE and Murphy AM. Multiple reaction monitoring to identify site-specific troponin I phosphorylated residues in the failing human heart. *Circulation*. 2012;126:1828-37.

285. Adams JA. Kinetic and catalytic mechanisms of protein kinases. *Chem Rev*. 2001;101:2271-90.

286. Knape MJ, Ahuja LG, Bertinetti D, Burghardt NC, Zimmermann B, Taylor SS and Herberg FW. Divalent Metal Ions Mg^{2+} and Ca^{2+} Have Distinct Effects on Protein Kinase A Activity and Regulation. *ACS Chem Biol*. 2015;10:2303-15.

287. Villar-Palasi C and Kumon A. Purification and properties of dog cardiac troponin T kinase. *J Biol Chem*. 1981;256:7409-15.

288. Zhang J, Zhang H, Ayaz-Guner S, Chen YC, Dong X, Xu Q and Ge Y. Phosphorylation, but not alternative splicing or proteolytic degradation, is conserved in human and mouse cardiac troponin T. *Biochemistry*. 2011;50:6081-92.

289. Sumandea MP, Pyle WG, Kobayashi T, de Tombe PP and Solaro RJ. Identification of a functionally critical protein kinase C phosphorylation residue of cardiac troponin T. *J Biol Chem*. 2003;278:35135-44.

290. Dobrovol'skiĭ AB, Risnik VV and Gusev NB. [Troponin T kinase: possible relationship to casein kinases of the G type]. *Biokhimiia*. 1981;46:1006-14.

291. Tobacman LS and Lee R. Isolation and functional comparison of bovine cardiac troponin T isoforms. *J Biol Chem*. 1987;262:4059-64.

292. Gusev NB. [Troponin from the myocardium and skeletal muscles: structure and properties]. *Biokhimiia*. 1986;51:1993-2009.

293. Scruggs SB and Solaro RJ. The significance of regulatory light chain phosphorylation in cardiac physiology. *Arch Biochem Biophys*. 2011;510:129-34.

294. Olsson MC, Patel JR, Fitzsimons DP, Walker JW and Moss RL. Basal myosin light chain phosphorylation is a determinant of Ca2+ sensitivity of force and activation dependence of the kinetics of myocardial force development. *Am J Physiol Heart Circ Physiol*. 2004;287:H2712-8.

295. van der Velden J, Papp Z, Boontje NM, Zaremba R, de Jong JW, Janssen PM, Hasenfuss G and Stienen GJ. The effect of myosin light chain 2 dephosphorylation on Ca2+ -sensitivity of force is enhanced in failing human hearts. *Cardiovasc Res*. 2003;57:505-14.

296. Chan JY, Takeda M, Briggs LE, Graham ML, Lu JT, Horikoshi N, Weinberg EO, Aoki H, Sato N, Chien KR and Kasahara H. Identification of cardiac-specific myosin light chain kinase. *Circ Res*. 2008;102:571-80.

297. Chang YH, Ye L, Cai W, Lee Y, Guner H, Kamp TJ, Zhang J and Ge Y. Quantitative proteomics reveals differential regulation of protein expression in recipient myocardium after trilineage cardiovascular cell transplantation. *Proteomics*. 2015;15:2560-7.

298. Yang L, Gregorich ZR, Cai W, Zhang P, Young B, Gu Y, Zhang J and Ge Y. Quantitative Proteomics and Immunohistochemistry Reveal Insights into Cellular and Molecular Processes in the Infarct Border Zone One Month after Myocardial Infarction. *J Proteome Res*. 2017;16:2101-2112.

299. Gregorich ZR, Cai W, Lin Z, Chen AJ, Peng Y, Kohmoto T and Ge Y. Distinct sequences and post-translational modifications in cardiac atrial and ventricular myosin light chains revealed by top-down mass spectrometry. *J Mol Cell Cardiol*. 2017;107:13-21.

300. Teekakirikul P, Kelly MA, Rehm HL, Lakdawala NK and Funke BH. Inherited cardiomyopathies: molecular genetics and clinical genetic testing in the postgenomic era. *J Mol Diagn*. 2013;15:158-70.

301. Lal S, Li A, Allen D, Allen PD, Bannon P, Cartmill T, Cooke R, Farnsworth A, Keogh A and Dos Remedios C. Best Practice BioBanking of Human Heart Tissue. *Biophys Rev*. 2015;7:399-406.

302. Bayliss CR, Jacques AM, Leung MC, Ward DG, Redwood CS, Gallon CE, Copeland O, McKenna WJ, Dos Remedios C, Marston SB and Messer AE. Myofibrillar Ca(2+) sensitivity is uncoupled from troponin I phosphorylation in hypertrophic obstructive cardiomyopathy due to abnormal troponin T. *Cardiovasc Res*. 2013;97:500-8.

303. Nixon BR, Liu B, Scellini B, Tesi C, Piroddi N, Ogut O, Solaro RJ, Ziolo MT, Janssen PM, Davis JP, Poggesi C and Biesiadecki BJ. Tropomyosin Ser-283 pseudo-phosphorylation slows myofibril relaxation. *Arch Biochem Biophys*. 2013;535:30-8.

304. Heeley DH. Investigation of the effects of phosphorylation of rabbit striated muscle alpha alpha-tropomyosin and rabbit skeletal muscle troponin-T. *Eur J Biochem*. 1994;221:129-37.

305. Schulz EM, Wilder T, Chowdhury SA, Sheikh HN, Wolska BM, Solaro RJ and Wieczorek DF. Decreasing tropomyosin phosphorylation rescues tropomyosin-induced familial hypertrophic cardiomyopathy. *J Biol Chem*. 2013;288:28925-35.

306. Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME and Gygi SP. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell*. 2010;143:1174-89.

307. Heppner DE, Hristova M, Ida T, Mijuskovic A, Dustin CM, Bogdándi V, Fukuto JM, Dick TP, Nagy P, Li J, Akaike T and van der Vliet A. Cysteine perthiosulfenic acid (Cys-SSOH): A novel intermediate in thiol-based redox signaling? *Redox Biol*. 2018;14:379-385.

308. Yu TS, Moctezuma-Anaya M, Kubo A, Keller G and Robertson S. The heart LIM protein gene (Hlp), expressed in the developing and adult heart, defines a new tissue-specific LIM-only protein family. *Mech Dev*. 2002;116:187-92.

309. Wei TC, Lin HY, Lu CC, Chen CM and You LR. Expression of Crip2, a LIM-domain-only protein, in the mouse cardiovascular system under physiological and pathological conditions. *Gene Expr Patterns*. 2011;11:384-94.

310. Cheung AK, Ko JM, Lung HL, Chan KW, Stanbridge EJ, Zabarovsky E, Tokino T, Kashima L, Suzuki T, Kwong DL, Chua D, Tsao SW and Lung ML. Cysteine-rich intestinal protein 2 (CRIP2) acts as a repressor of NF-kappaB-mediated proangiogenic cytokine transcription to suppress tumorigenesis and angiogenesis. *Proc Natl Acad Sci U S A*. 2011;108:8390-5.

311. Zhou L, Wang Y, Zhou M, Zhang Y, Wang P, Li X, Yang J, Wang H and Ding Z. HOXA9 inhibits HIF-1α-mediated glycolysis through interacting with CRIP2 to repress cutaneous squamous cell carcinoma development. *Nat Commun*. 2018;9:1480.

312. Lin C, Guo X, Lange S, Liu J, Ouyang K, Yin X, Jiang L, Cai Y, Mu Y, Sheikh F, Ye S, Chen J, Ke Y and Cheng H. Cypher/ZASP is a novel A-kinase anchoring protein. *J Biol Chem*. 2013;288:29403-13.

313. Burridge PW, Keller G, Gold JD and Wu JC. Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell*. 2012;10:16-28.

314. Lian X, Zhang J, Azarin SM, Zhu K, Hazeltine LB, Bao X, Hsiao C, Kamp TJ and Palecek SP. Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat Protoc*. 2013;8:162-75.

315. Zhang J, Klos M, Wilson GF, Herman AM, Lian X, Raval KK, Barron MR, Hou L, Soerens AG, Yu J, Palecek SP, Lyons GE, Thomson JA, Herron TJ, Jalife J and Kamp TJ. Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circ Res*. 2012;111:1125-36.

316. Bedada FB, Wheelwright M and Metzger JM. Maturation status of sarcomere structure and function in human iPSC-derived cardiac myocytes. *Biochim Biophys Acta*. 2016;1863:1829-1838.

317. Feric NT and Radisic M. Maturing human pluripotent stem cell-derived cardiomyocytes in human engineered cardiac tissues. *Adv Drug Deliv Rev*. 2016;96:110-34.

318. Nunes SS, Miklas JW, Liu J, Aschar-Sobbi R, Xiao Y, Zhang B, Jiang J, Massé S, Gagliardi M, Hsieh A, Thavandiran N, Laflamme MA, Nanthakumar K, Gross GJ, Backx PH, Keller G and Radisic M. Biowire: a platform for maturation of human pluripotent stem cell-derived cardiomyocytes. *Nat Methods*. 2013;10:781-7.

319. Shadrin IY, Allen BW, Qian Y, Jackman CP, Carlson AL, Juhas ME and Bursac N. Cardiopatch platform enables maturation and scale-up of human pluripotent stem cell-derived engineered heart tissues. *Nat Commun*. 2017;8:1825.

320. Breker M and Schuldiner M. The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nat Rev Mol Cell Biol*. 2014;15:453-64.

321. van den Berg CW, Okawa S, Chuva de Sousa Lopes SM, van Iperen L, Passier R, Braam SR, Tertoolen LG, del Sol A, Davis RP and Mummery CL. Transcriptome of human foetal heart compared with cardiomyocytes from pluripotent stem cells. *Development*. 2015;142:3231-8.

322. Liu Y, Beyer A and Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*. 2016;165:535-50.

323. Aebersold R and Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016;537:347-55.

324. Ralphe JC and de Lange WJ. 3D engineered cardiac tissue models of human heart disease: learning more from our mice. *Trends Cardiovasc Med*. 2013;23:27-32.

325. Raval KK, Tao R, White BE, De Lange WJ, Koonce CH, Yu J, Kishnani PS, Thomson JA, Mosher DF, Ralphe JC and Kamp TJ. Pompe disease results in a Golgi-based glycosylation deficit in human induced pluripotent stem cell-derived cardiomyocytes. *J Biol Chem*. 2015;290:3121-36.

326. Lian X, Hsiao C, Wilson G, Zhu K, Hazeltine LB, Azarin SM, Raval KK, Zhang J, Kamp TJ and Palecek SP. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proc Natl Acad Sci U S A*. 2012;109:E1848-57.

327. Ferrige AG, Seddon MJ, Jarvis S, Skilling J and Aplin R. Maximum entropy deconvolution in electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry*. 1991;5:374-377.

328. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N and Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*. 2014;13:2513-26.

329. Cox J and Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26:1367-72.

330. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D and Thomas PD. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017;45:D183-D189.

331. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, Milacic M, Roca CD, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Viteri G, Weiser J, Wu G, Stein L, Hermjakob H and D'Eustachio P. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*. 2018;46:D649-D655.

332. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13:2498-504.

333. Farrell ET, Grimes AC, de Lange WJ, Armstrong AE and Ralphe JC. Increased Postnatal Cardiac Hyperplasia Precedes Cardiomyocyte Hypertrophy in a Model of Hypertrophic Cardiomyopathy. *Front Physiol*. 2017;8:414.

334. Mummery CL, Zhang J, Ng ES, Elliott DA, Elefanty AG and Kamp TJ. Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: a methods overview. *Circ Res*. 2012;111:344-58.

335. Lundy SD, Zhu WZ, Regnier M and Laflamme MA. Structural and functional maturation of cardiomyocytes derived from human pluripotent stem cells. *Stem Cells Dev*. 2013;22:1991-2002.

336. Reiser PJ, Portman MA, Ning XH and Schomisch Moravec C. Human cardiac myosin heavy chain isoforms in fetal and failing adult atria and ventricles. *Am J Physiol Heart Circ Physiol*. 2001;280:H1814-20.

337. Bhavsar PK, Dhoot GK, Cumming DV, Butler-Browne GS, Yacoub MH and Barton PJ. Developmental expression of troponin I isoforms in fetal human heart. *FEBS Lett*. 1991;292:5-8.

338. Suurmeijer AJ, Clément S, Francesconi A, Bocchi L, Angelini A, Van Veldhuisen DJ, Spagnoli LG, Gabbiani G and Orlandi A. Alpha-actin isoform distribution in normal and failing human heart: a morphological, morphometric, and biochemical study. *J Pathol*. 2003;199:387-97.

339. Gomes AV, Venkatraman G, Davis JP, Tikunova SB, Engel P, Solaro RJ and Potter JD. Cardiac troponin T isoforms affect the Ca(2+) sensitivity of force development in the presence of slow skeletal troponin I: insights into the role of troponin T isoforms in the fetal heart. *J Biol Chem*. 2004;279:49579-87.

340. Heeley DH, Heeley DA, Moir AJ and Perry SV. Phosphorylation of tropomyosin during development in mammalian striated muscle. *FEBS Lett*. 1982;146:115-8.

341. Buyandelger B, Ng KE, Miocic S, Piotrowska I, Gunkel S, Ku CH and Knöll R. MLP (muscle LIM protein) as a stress sensor in the heart. *Pflugers Arch*. 2011;462:135-42.

342. Lopaschuk GD and Jaswal JS. Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation. *J Cardiovasc Pharmacol*. 2010;56:130-40.

343. Keller A, Rouzeau JD, Farhadian F, Wisnewsky C, Marotte F, Lamandé N, Samuel JL, Schwartz K, Lazar M and Lucas M. Differential expression of alpha- and beta-enolase genes during rat heart development and hypertrophy. *Am J Physiol*. 1995;269:H1843-51.

344. Gupta V, Discenza M, Guyon JR, Kunkel LM and Beggs AH. α-Actinin-2 deficiency results in sarcomeric defects in zebrafish that cannot be rescued by α-actinin-3 revealing functional differences between sarcomeric isoforms. *FASEB J*. 2012;26:1892-908.

345. Sjöblom B, Salmazo A and Djinović-Carugo K. Alpha-actinin structure and regulation. *Cell Mol Life Sci*. 2008;65:2688-701.

346. Lapidos KA, Kakkar R and McNally EM. The dystrophin glycoprotein complex: signaling strength and integrity for the sarcolemma. *Circ Res*. 2004;94:1023-31.

347. Wu G, Ai T, Kim JJ, Mohapatra B, Xi Y, Li Z, Abbasi S, Purevjav E, Samani K, Ackerman MJ, Qi M, Moss AJ, Shimizu W, Towbin JA, Cheng J and Vatta M. alpha-1-syntrophin mutation and the long-QT syndrome: a disease of sodium channel disruption. *Circ Arrhythm Electrophysiol*. 2008;1:193-201.

348. Cheng J, Van Norstrand DW, Medeiros-Domingo A, Valdivia C, Tan BH, Ye B, Kroboth S, Vatta M, Tester DJ, January CT, Makielski JC and Ackerman MJ. Alpha1-syntrophin mutations identified in sudden infant death syndrome cause an increase in late cardiac sodium current. *Circ Arrhythm Electrophysiol*. 2009;2:667-76.

349. Kumarapeli AR, Su H, Huang W, Tang M, Zheng H, Horak KM, Li M and Wang X. Alpha B-crystallin suppresses pressure overload cardiac hypertrophy. *Circ Res*. 2008;103:1473-82.

350. Diokmetzidou A, Soumaka E, Kloukina I, Tsikitis M, Makridakis M, Varela A, Davos CH, Georgopoulos S, Anesti V, Vlahou A and Capetanaki Y. Desmin and αB-crystallin interplay in the maintenance of mitochondrial homeostasis and cardiomyocyte survival. *J Cell Sci*. 2016;129:3705-3720.

351. Stammers AN, Susser SE, Hamm NC, Hlynsky MW, Kimber DE, Kehler DS and Duhamel TA. The regulation of sarco(endo)plasmic reticulum calcium-ATPases (SERCA). *Can J Physiol Pharmacol*. 2015;93:843-54.

352. Hirt MN, Hansen A and Eschenhagen T. Cardiac tissue engineering: state of the art. *Circ Res*. 2014;114:354-67.

353. de Lange WJ, Hegge LF, Grimes AC, Tong CW, Brost TM, Moss RL and Ralphe JC. Neonatal mouse-derived engineered cardiac tissue: a novel model system for studying genetic heart disease. *Circ Res*. 2011;109:8-19.

354. Cheng L, Ding G, Qin Q, Huang Y, Lewis W, He N, Evans RM, Schneider MD, Brako FA, Xiao Y, Chen YE and Yang Q. Cardiomyocyte-restricted peroxisome proliferator-activated receptor-delta deletion perturbs myocardial fatty acid oxidation and leads to cardiomyopathy. *Nat Med*. 2004;10:1245-50.

355. Burkart EM, Sambandam N, Han X, Gross RW, Courtois M, Gierasch CM, Shoghi K, Welch MJ and Kelly DP. Nuclear receptors PPARbeta/delta and PPARalpha direct distinct metabolic regulatory programs in the mouse heart. *J Clin Invest*. 2007;117:3930-9.

356. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D and Gavin AC. Visualization of omics data for systems biology. *Nat Methods*. 2010;7:S56-68.

357. Cox J and Mann M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem*. 2011;80:273-99.

358. Larance M and Lamond AI. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol*. 2015;16:269-80.

359. Bensimon A, Heck AJ and Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem*. 2012;81:379-405.

360. Yates JR, Ruse CI and Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng*. 2009;11:49-79.

361. Nesvizhskii AI, Vitek O and Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods*. 2007;4:787-797.

362. Cravatt BF, Simon GM and Yates JR. The biological impact of mass-spectrometry-based proteomics. *Nature*. 2007;450:991-1000.

363. Huang PH and White FM. Phosphoproteomics: unraveling the signaling web. *Mol Cell*. 2008;31:777-81.

364. Miteva YV and Cristea IM. A proteomic perspective of Sirtuin 6 (SIRT6) phosphorylation and interactions and their dependence on its catalytic activity. *Mol Cell Proteomics*. 2014;13:168-83.

365. Zhang Z, Wu S, Stenoien DL and Paša-Tolić L. High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)*. 2014;7:427-54.

366. Ryan CM, Souda P, Bassilian S, Ujwal R, Zhang J, Abramson J, Ping P, Durazo A, Bowie JU, Hasan SS, Baniulis D, Cramer WA, Faull KF and Whitelegge JP. Post-translational modifications of integral membrane proteins resolved by top-down Fourier transform mass spectrometry with collisionally activated dissociation. *Mol Cell Proteomics*. 2010;9:791-803.

367. Dang X, Scotcher J, Wu S, Chu RK, Tolić N, Ntai I, Thomas PM, Fellers RT, Early BP, Zheng Y, Durbin KR, Leduc RD, Wolff JJ, Thompson CJ, Pan J, Han J, Shaw JB, Salisbury JP,

Easterling M, Borchers CH, Brodbelt JS, Agar JN, Paša-Tolić L, Kelleher NL and Young NL. The first pilot project of the consortium for top-down proteomics: a status report. *Proteomics*. 2014;14:1130-40.

368. LeDuc RD, Taylor GK, Kim YB, Januszyk TE, Bynum LH, Sola JV, Garavelli JS and Kelleher NL. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res*. 2004;32:W340-5.

369. Pesavento JJ, Kim YB, Taylor GK and Kelleher NL. Shotgun annotation of histone modifications: a new approach for streamlined characterization of proteins by top down mass spectrometry. *J Am Chem Soc*. 2004;126:3386-7.

370. Mazur MT and Fyhr R. An algorithm for identifying multiply modified endogenous proteins using both full-scan and high-resolution tandem mass spectrometric data. *Rapid Commun Mass Spectrom*. 2011;25:3617-26.

371. Pevzner PA, Dancík V and Tang CL. Mutation-tolerant protein identification by mass spectrometry. *J Comput Biol*. 2000;7:777-87.

372. Fellers RT, Greer JB, Early BP, Yu X, LeDuc RD, Kelleher NL and Thomas PM. ProSight Lite: Graphical software to analyze top-down mass spectrometry data. *Proteomics*. 2014.

373. Hwang L, Ayaz-Guner S, Gregorich ZR, Cai W, Valeja SG, Jin S and Ge Y. Specific enrichment of phosphoproteins using functionalized multivalent nanoparticles. *J Am Chem Soc*. 2015;137:2432-5.

374. Wu S, Brown RN, Payne SH, Meng D, Zhao R, Tolić N, Cao L, Shukla A, Monroe ME, Moore RJ, Lipton MS and Paša-Tolić L. Top-Down Characterization of the Post-Translationally Modified Intact Periplasmic Proteome from the Bacterium Novosphingobium aromaticivorans. *Int J Proteomics*. 2013;2013:279590.

375. Wu S, Brown JN, Tolić N, Meng D, Liu X, Zhang H, Zhao R, Moore RJ, Pevzner P, Smith RD and Paša-Tolić L. Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. *Proteomics*. 2014;14:1211-22.

376. LaVoie MJ, Ostaszewski BL, Weihofen A, Schlossmacher MG and Selkoe DJ. Dopamine covalently modifies and functionally inactivates parkin. *Nat Med*. 2005;11:1214-21.

377. Rakhit R, Crow JP, Lepock JR, Kondejewski LH, Cashman NR and Chakrabartty A. Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J Biol Chem*. 2004;279:15499-504.

378. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W and Aebersold R. A common open representation of mass spectrometry data and its application to proteomics research. *Nat Biotechnol*. 2004;22:1459-66.

379. Smith LM and Kelleher NL. Proteoform: a single term describing protein complexity. 2013;10.

380. Sabido E, Selevsek N and Aebersold R. Mass spectrometry-based proteomics for systems biology. *Curr Opin Biotechnol*. 2012;23:591-7.

381. Zhang Z, Wu S, Stenoien DL and Pasa-Tolic L. High-throughput proteomics. *Annu Rev Anal Chem (Palo Alto Calif)*. 2014;7:427-54.

382. Vellaichamy A, Tran JC, Catherman AD, Lee JE, Kellie JF, Sweet SM, Zamdborg L, Thomas PM, Ahlf DR, Durbin KR, Valaskovic GA and Kelleher NL. Size-sorting combined with improved nanocapillary liquid chromatography-mass spectrometry for identification of intact proteins up to 80 kDa. *Anal Chem*. 2010;82:1234-44.

383. Kellie JF, Catherman AD, Durbin KR, Tran JC, Tipton JD, Norris JL, Witkowski CE, 2nd, Thomas PM and Kelleher NL. Robust analysis of the yeast proteome under 50 kDa by molecular-mass-based fractionation and top-down mass spectrometry. *Anal Chem*. 2012;84:209-15.

384. Li Y, Compton PD, Tran JC, Ntai I and Kelleher NL. Optimizing capillary electrophoresis for top-down proteomics of 30-80 kDa proteins. *Proteomics*. 2014;14:1158-64.

385. Hong P, Koza S and Bouvier ES. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *J Liq Chromatogr Relat Technol*. 2012;35:2923-2950.

386. Li Y, Gu C, Gruenhagen J, Zhang K, Yehl P, Chetwyn NP and Medley CD. A size exclusion-reversed phase two dimensional-liquid chromatography methodology for stability and small molecule related species in antibody drug conjugates. *J Chromatogr A*. 2015;1393:81-8.

387. Alpert AJ. 8 - Size Exclusion High-Performance Liquid Chromatography of Small Solutes A2 - Wu, Chi-san *Column Handbook for Size Exclusion Chromatography* San Diego: Academic Press; 1999: 249-266.

388. Simpson DC, Ahn S, Pasa-Tolic L, Bogdanov B, Mottaz HM, Vilkov AN, Anderson GA, Lipton MS and Smith RD. Using size exclusion chromatography-RPLC and RPLC-CIEF as two-dimensional separation strategies for protein profiling. *Electrophoresis*. 2006;27:2722-33.

389. Yi-Chen C, Ayaz-Guner S, Peng Y, Lane NM, Locher MR, Kohmoto T, Larsson L, Moss RL and Ge Y. Effective top-down LC/MS+ method for assessing actin isoforms as a potential cardiac disease marker. 2015.

390. Berek D. Size exclusion chromatography – A blessing and a curse of science and technology of synthetic polymers.

391. Van Kreveld ME and Van den Hoed N. Mass transfer phenomena in gel permeation chromatography. *Journal of Chromatography A*. 1978;149:71-91.

392. Kostanski LK, Keller DM and Hamielec AE. Size-exclusion chromatography-a review of calibration methodologies. *J Biochem Biophys Methods*. 2004;58:159-86.

393. Peng Y, Gregorich ZR, Valeja SG, Zhang H, Cai W, Chen Y-C, Guner H, Chen AJ, Schwahn DJ, Hacker TA, Liu X and Ge Y. Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. 2014.

394. Sharma S, Simpson DC, Tolic N, Jaitly N, Mayampurath AM, Smith RD and Pasa-Tolic L. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J Proteome Res*. 2007;6:602-10.

395. Stobaugh JT, Fague KM and Jorgenson JW. Prefractionation of intact proteins by reversed-phase and anion-exchange chromatography for the differential proteomic analysis of Saccharomyces cerevisiae. *J Proteome Res*. 2013;12:626-36.

396. Xiu L, Valeja SG, Alpert AJ, Jin S and Ge Y. Effective protein separation by coupling hydrophobic interaction and reverse phase chromatography for top-down proteomics. 2014.

397. Valeja SG, Xiu L, Gregorich ZR, Guner H, Jin S and Ge Y. Three dimensional liquid chromatography coupling ion exchange chromatography/hydrophobic interaction chromatography/reverse phase chromatography for effective protein separation in top-down proteomics. 2015;87.

398. Pesavento JJ, Bullock CR, LeDuc RD, Mizzen CA and Kelleher NL. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J Biol Chem*. 2008;283:14927-37.

399. Zhang X, Fang A, Riley CP, Wang M, Regnier FE and Buck C. Multi-dimensional liquid chromatography in proteomics--a review. *Anal Chim Acta*. 2010;664:101-13.

400. Tian Z, Zhao R, Tolic N, Moore RJ, Stenoien DL, Robinson EW, Smith RD and Pasa-Tolic L. Two-dimensional liquid chromatography system for online top-down mass spectrometry. *Proteomics*. 2010;10:3610-20.

401. Wilson SR, Jankowski M, Pepaj M, Mihailova A, Boix F, Vivo Truyols G, Lundanes E and Greibrokk T. 2D LC Separation and Determination of Bradykinin in Rat Muscle Tissue Dialysate with On-Line SPE-HILIC-SPE-RP-MS. *Chromatographia*. 2007;66:469-474.

402. Zhang J, Roth MJ, Chang AN, Plymire DA, Corbett JR, Greenberg BM and Patrie SM. Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography. 2013;85.

403. Porath J, Carlsson J, Olsson I and Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*. 1975;258:598-9.

404. Schmidt SR, Schweikart F and Andersson ME. Current methods for phosphoprotein isolation and enrichment. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007;849:154-62.

405. Cai W, Guner H, Gregorich ZR, Chen AJ, Ayaz-Guner S, Peng Y, Valeja SG, Liu X and Ge Y. MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol Cell Proteomics*. 2016;15:703-14.

406. Cammarata MB, Thyer R, Rosenberg J, Ellington A and Brodbelt JS. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *J Am Chem Soc*. 2015;137:9128-35.

407. Robotham SA and Brodbelt JS. Comparison of Ultraviolet Photodissociation and Collision Induced Dissociation of Adrenocorticotropic Hormone Peptides. *J Am Soc Mass Spectrom*. 2015;26:1570-9.

408. Metzger JM and Westfall MV. Covalent and noncovalent modification of thin filament action: the essential role of troponin in cardiac muscle regulation. *Circ Res*. 2004;94:146-58.

409. Hoshijima M. Mechanical stress-strain sensors embedded in cardiac cytoskeleton: Z disk, titin, and associated structures. *Am J Physiol Heart Circ Physiol*. 2006;290:H1313-25.

410. Krcmery J, Camarata T, Kulisz A and Simon HG. Nucleocytoplasmic functions of the PDZ-LIM protein family: new insights into organ development. *Bioessays*. 2010;32:100-8.