Towards A Systems Biology Understanding of Metabolic Syndrome

by

Clinton Mielke

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2013 by the
Graduate Supervisory Committee:

Lawrence Mandarino, Co-Chair
Joshua LaBaer, Co-Chair
D. Mitchell Magee
Valentin Dinu
Wayne Willis

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

This dissertation investigates the condition of skeletal muscle insulin resistance using bioinformatics and computational biology approaches. Drawing from several studies and numerous data sources, I have attempted to uncover molecular mechanisms at multiple levels. From the detailed atomistic simulations of a single protein, to datamining approaches applied at the systems biology level, I provide new targets to explore for the research community. Furthermore I present a new online web resource that unifies various bioinformatics databases to enable discovery of relevant features in 3D protein structures.

ACKNOWLEDGEMENTS

I would have never earned a PhD if it weren't for the support of my Advisor, Larry Mandarino. The first year of graduate school was challenging, and despite 5 different rotations in various labs in the Biodesign Institute, none of the available projects felt like a proper fit to my interests. I had just 2 weeks left to find a lab, and happened to run into Larry at a talk. I was not aware that any labs on campus were studying a disease that had such a profound personal connection to me and my family. Larry and I clicked because we both had entirely different backgrounds than biology, and despite having zero requisite skills needed for the task at hand, Larry gave me the opportunity to completely transform. This document is proof that I have. Thanks Larry, for putting up with me.

My co-chair, Josh LaBaer, similarly gave me a home within Biodesign to keep close ties to my program. I had rotated under Josh in his fantastic lab and learned a great deal of modern techniques. Although ultimately my PhD studies took me elsewhere, Josh allowed me to continue working in his lab and attend meetings to learn everything I know about cancer biology. The LaBaer lab, and its many friendly people, will always be family to me.

The heart of the LaBaer lab is indisputably Mitch Magee, who has been a warm friend throughout graduate school. He has always been there through the bad times, kept me sane, calmed me down, and been like a father. He is well known throughout the building of being the man with the biggest heart. He is a shining example of how effective academia and science could be if everyone was as friendly and professional.

Valentin Dinu has proven to be among the most loyal people I have ever had the privilege of serving under. When my first PhD project went south, Valentin gave me a

new lab home to play in with some of the kindest and most professional students I have ever corresponded with. Collaborating with Larry, Valentin provided me with the server infrastructure and support necessary to completely turn my PhD around in just a year. He has also been the strongest advocate of my abilities and has wished me success in my future career countless times. It is when conversing with Valentin that my dreams are kept alive, and he encourages me that I *can* go to better places and find my happiness.

Wayne Willis, you continue to put all to shame with your commanding *godlike* knowledge of metabolic pathways and mitochondrial physiology! You are the true notion of the genius professor, yet you are among the most kind and approachable faculty I've met. Never before have I met someone at your level that I could still routinely eat lunch or grab a beer with. I will always miss our conversations, where I learned all of my bioenergetics theory in grad school. You have been a core teacher, and if I ever do great things, you can thank yourself for my knowledge. Party on Wayne!

From the lab, I want to thank several fellow soldiers that trained me in the mission of fighting obesity: specifically Paul, Leon, April, and Jean for their incredibly patient mentorship. I needed a lot of help in the lab, and they taught me much. Without their guidance over the years, I could not be called a molecular biologist.

I thank my family for their love and support over the years. I also want to thank the Biological Design Graduate Program for the opportunity and its support. I thank all of my peers still finishing up, and I wish them luck in defending and finding bright futures. Many of my closest friends here and abroad have supported me through all of the rough times: Marc, Roeland, Brad, Joe, Brian, Karl, Andrew, Brianne, Rene, Kurt, Jack, and countless others. I adore you all! I also thank the programs administrative support staff

(Maria, JoAnn, and Jim) for keeping us all sane during the most stressful times. Lastly, to Tom Goodman, you have been like a father to me, and I will always cherish your wisdom of what the future of humanity holds.

DEDICATION

This dissertation is dedicated to people who have been punished entirely too much for their weight. I love them all.

Obesity is a serious disease with profound health consequences, but it also comes with it a great social stigma. It is the nature of our culture to ridicule these people, subject them to judgment and bias, ignore them in friendship or relational circumstances, alienate or otherwise abandon them, and torture them with accusations of laziness or stupidity. We cast blame entirely on their lifestyle whilst ignoring their genetic circumstances or previous medical history. Having been obese for much of my life, I can personally attest to the immense pain and sadness that obesity causes. It truly is a condition which dramatically reduces the quality of one's life, particularly in social or career circumstances. Obesity is the last remaining acceptable form of prejudice in our world, yet mounting molecular evidence indicates that it is a disease like any other.

I have been severely obese for my entire life, and I can personally attest to the immense pain this condition causes. I also have a complicated medical history of several endocrine problems, namely testosterone insufficiency and two bouts of thyroid cancer when I was very young. Nothing I tried got the weight off, even with intense dietary or exercise interventions, so I switched careers in grad school. As of the conclusion of this dissertation I've lost 120 lbs. The weight did not come off with traditionally defined diet and exercise, but with many episodes of forced starvation lasting several days. My story proves that lifestyle interventions *can* overcome genetic predispositions and prior medical conditions, but obese people have to try much harder! I believe the obese are stronger

people. They must deal with their own weight and the weight the world places upon them. There are few people in this world who lose so much weight, and even fewer who maintain the weight loss. These are the individuals who will stop at nothing to be healthy and be accepted by society, but their efforts require unreasonable sacrifice.

These victims are all my comrades, and I will continue to fight the enemy alongside them. The enemy is the hundreds of invisible molecular factors that make us ill, from genetic variability, to dietary toxins that are incompletely understood, to chemical compounds in our industrial environment that we are exposed to, and to pathogenic organisms that invade our gut and lead to inflammation. The enemy is not laziness or lack of will power. It never has been. It is the world fighting against our biology, at the molecular scale, in ways we have yet to comprehend.

Very recently, I have been genotyped by the consumer genetics testing company 23andMe. I have discovered that I have some of the worst genetic variants associated with metabolic syndrome, and many of these genes are discussed in this dissertation. This self-discovery has put into perspective an entire life of suffering. With this knowledge I have learned, I will carry forward my mission to heal. I will strive for a cure, so that the many people who suffer these medical problems can be saved, and can be happy.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

Metabolic syndrome is a growing health problem that places enormous weight on our society. It impacts the quality of life and longevity of the millions of victims suffering its effects, and the burden to our healthcare system is difficult to overestimate. Metabolic syndrome is a combination of several metabolic abnormalities that may have distinct causes but are commonly found together, including obesity, lipid abnormalities, and conditions of abnormal glucose homeostasis including insulin resistance and type 2 diabetes. Metabolic syndrome is also associated with cardiovascular disease, which remains the leading cause of death worldwide. It is a complex systemic condition that is influenced by a variety of factors, which include genetic heritability, environmental exposures, the burden of other illnesses, and lifestyle choices including diet and exercise. Evidence is accumulating that these pathologies arise from many distinct underlying causes. Elucidating the fundamental mechanisms of these abnormalities is crucial to save millions.

Metabolic syndrome is a complicated set of associated diseases in which the storage, transport, or utilization of energy is defective. These biological functions are precisely regulated at the cellular level by complex internal signaling mechanisms and metabolic pathways. Cells are *deciding* what to do based on their genetic makeup and the environment they are exposed to. Only by reverse engineering how cells *think*, we can come to an understanding of *why* they behave in specific ways and interact to form a complex system. This dissertation explores the molecular mechanisms underlying metabolic syndrome with a systems biology focus that promotes the viewpoint that the

entire system must be considered at all levels to understand the phenotype. New technologies are permitting us to scrutinize the inner workings of the cell at the genomic, transcriptomic, proteomic, and metabolomic levels. Unexpected connections can be found by “connecting the dots” between these various datasets. This underlying philosophy has resulted in several interesting discoveries that are discussed amongst these chapters. The stories explored in this dissertation range from the atomic details of single proteins, to entire signaling networks in cells, and indeed interactions between multiple organs. New strategies and tools are required to sort through the massive datasets and find the true points of control and failure within cellular information processing.

In this dissertation, I characterize the biophysical properties of a single protein critical for mitochondrial functioning, and provide a compelling argument as to how it can direct the energy status of the entire cell from the bottom up. Additionally, I develop a simple approach to find direct connections between genetic variation, expression profiles, and metabolic phenotypes of experimental subjects. Finally, I have developed a web-accessible database that unifies genomics and proteomics sequence data with protein structures, thereby enabling the visualization and discovery of novel regulators of protein function. These projects focus around the study of metabolic syndrome, and its associated disorder of glucose homeostasis known as diabetes.

Diabetes is a disease defined by elevated blood glucose, and additionally the inability to respond to an increase in blood glucose following a meal. In a healthy individual, blood glucose concentrations are held within a specific range to maintain homeostasis. To regulate this range, the pancreas contains specialized cells in the

pancreatic Islets of Langerhans to detect blood glucose concentrations and respond by secreting hormones. The hormone insulin is the best known, as it is a hormone secreted from beta cells in the islets when blood glucose levels rise. Insulin informs peripheral tissues in the body that glucose supply is in excess, and peripheral tissues such as muscle and fat respond to this signal by absorbing the excess glucose for storage and eventual utilization. The mechanisms of this process are complex. Insulin binds to insulin receptors on cells, which leads to signal propagation through hundreds of downstream associated proteins. At this stage, information processing takes place within the cell to *decide* if glucose is needed. If it is, the end result is insertion of the insulin-responsive glucose transporter protein (GLUT4) into the cell membrane. This final insertion permits glucose uptake into the cell.

Diabetes is a chronic state of elevated blood glucose, and is broadly classed in two types. In type 1 diabetes, autoimmune destruction of the pancreas leads to insufficient insulin secretion. The lack of the hormone insulin leads to increased blood glucose levels. The disease typically develops at an early age, and is treated by careful glucose monitoring and injection of insulin.

In type 2 diabetes, peripheral tissues ignore the insulin secreted by the pancreas. This state of the peripheral tissues is known as “insulin resistance”, which is thought to lead eventually to type 2 diabetes after prolonged stress leading to pancreatic burnout. The molecular mechanisms of insulin resistance, particularly in human muscle, have been a core focus of research. Despite binding of insulin to its receptor in target tissues, the glucose transporter is not translocated to and inserted properly in the plasma membrane. Somewhere between the insulin receptor and the glucose transporter, the pathway

becomes *switched off* for reasons that are not entirely clear. Over the last decade, numerous molecular sites of several *off switches* have been found, and they appear to correspond to different environmental conditions that the cell finds itself experiencing. Exposure to free fatty acids (FFAs) appears to be one instigator of this deactivation, leading many to conclude that high fat American diets are to blame.[1] The signaling pathways are complicated however, and more work needs to be undertaken to study where the true control points are located. Such work will provide insights into therapeutic targets.

Chapter 2 discusses my effort to uncover core signaling nodes within muscle that ultimately determine its insulin sensitivity. A study previously conducted by our lab recruited 14 human subjects to partake in an exercise bout. A hyperinsulinemic euglycemic clamp experiment was performed to measure insulin sensitivity. In this technique, insulin is injected into subjects to stimulate glucose uptake in peripheral tissues, and glucose is then infused at a rate that maintains blood glucose concentrations at a steady level. The glucose uptake rate in peripheral tissues can then be estimated based upon the rate of infusion. In this study, a muscle biopsy was taken from one leg prior to the clamp experiment. On a separate day, subjects underwent an exercise bout, followed by a second muscle biopsy 30 minutes after exercise. Gene expression microarrays were then used to determine the expression levels in both biopsies.

This dataset presented an excellent opportunity to attempt data mining into the gene expression profiles of muscle that correspond to disease. Typical gene expression microarray studies focus entirely on the comparison between two measured states. For example, two groups of people are compared, or comparisons are made before and after

some intervention such as exercise. It is thus the *fold change* that is commonly considered, and rarely the steady state expression values. In exercise studies, these types of analysis have taught us much about the exercise-induced adaptations to metabolic systems within muscle, and how these adaptations are less pronounced [2] in some subjects, which suggests a state of *exercise resistance*. Gene expression studies of this sort typically identify genes that change significantly (up or down) amongst the entire group of participants. This however misses genes that respond in opposite directions amongst a group.

In our exercise study, several metabolic parameters were collected from test participants, including their height and weight, insulin stimulated glucose uptake rate, parameters from a full lipid panel, as well as exercise performance characteristics such as VO2max and watts produced on the stationary cycle. These metabolic parameters are continuous values, which presented a unique opportunity to mine the gene expression data for *direct* associations in both single biopsies as well as gene expression fold changes. To this end, a simple technique was devised whereby every gene expression probe on the microarrays was correlated with these various metabolic phenotypes. Correlation coefficients were sorted, and the top and bottom most positive/negative correlated were investigated with deep literature review.

Remarkably, the findings were profound. We present a model whereby top correlates to glucose uptake are directly tied to the wnt signaling pathway which is deeply associated to type 2 diabetes susceptibility in genome wide association studies (GWAS). Furthermore, two top-correlating genes are found in GWAS studies for diabetes and BMI, and these connections validate our systems biology approach to unifying domain

knowledge from multiple sources. Many negative correlating genes correspond to a well-studied pathway of muscle atrophy, which is a phenomenon associated with metabolic syndrome [3]. Interestingly, muscle atrophy is initiated during times of fasting to provide amino acids to the liver to produce glucose in a process known as hepatic gluconeogenesis. When dietary glucose is scarce, the muscle must stop taking up glucose and start exporting amino acids, which keeps blood glucose available to the brain. These gene expression findings lead to a hypothesis that diabetic muscle *thinks* that the organism is starving, and leads us to propose a direct mechanistic connection between the muscle atrophy pathway and insulin resistance. This *off switch* is potentially mediated directly by proteolysis of proteins critical for glucose uptake, and I present evidence of this hypothesis.

The transcriptomics analysis leads to interesting findings which may unveil novel core regulatory pathways that regulate the phenotype of a cell, and indeed the phenotype of the whole organism. It also provided startling context to the nature of the disease. Gene expression values however only partially predict the abundance of proteins.[4] Furthermore, proteins undergo post-translational modification (PTMs), which can selectively change the biophysical properties of key residues. These modifications are not indicated in the genomic sequence but can rapidly and dynamically change the behavior of proteins. These processes are key to changing the reaction characteristics of metabolic enzymes and are also the primary mechanism by which signaling pathways in the cell process information. Thus the changing gene expression patterns in the cell only form half of the internal cellular *computer*. The remainder of the information processing is conducted within the proteome, and this layer must be additionally monitored to uncover

regulatory mechanisms that associate with disease.

It is biological mass spectrometry that has revolutionized the proteomics field in recent years. This amazing technology enables the routine detection and sequencing of intact proteins from biological samples, as well as the detection of post-translational modifications at specific sites in the protein sequence. This “protein sequencing” is now growing as a complement to the genome sequencing revolution that took biology in the previous decade. Briefly, proteins isolated from human samples are enzymatically digested into short peptides, separated with liquid chromatography, and then ionized into particle accelerators where their mass-to-charge ratios are measured. To sequence the peptides, they are literally blown apart at the peptide bonds via various fragmentation methods, and the sequence is determined from the pattern of peaks that emerges. This whole process is done routinely in the lab, and a typical tissue biopsy will unveil thousands of proteins and many associated modifications.

By scrutinizing these proteomes, many advances have been made in studying insulin resistance. The insulin signaling pathway has been mapped by looking for interacting partners of key signaling molecules. Furthermore, the signaling “state” of the pathway can be elucidated by detecting the characteristic phosphorylation patterns on central signaling proteins.[5] In mapping out these pathways, we gain insight into the signal propagation paths, from receptors to effectors, and we are slowly detangling the inputs and processing that determines metabolic phenotypes. These advancements are allowing us to directly read out the signal pathways from tissues and deduce what the cells were “thinking” prior to isolation.

There are several proteins in the insulin signaling pathway that have received considerable attention with mass spectrometric analysis. One such protein is known as the Insulin Receptor Substrate-1 (IRS-1) because it directly binds to the insulin receptor upon activation. IRS-1 is considered the primary signaling hub of the pathway, as it contains many [5-7] sites which become phosphorylated. These sites serve as switches that regulate binding of many downstream signaling proteins that connect with other signaling pathways. Some of these sites are activating, and some are deactivating of the signaling pathway, and many of these sites are modulated by feedback mechanisms within the cell that keep the signaling properly regulated. Understanding how these sites become activated due to various environmental conditions is a focus of current research.

Beyond signaling pathways, we are beginning to focus on metabolic enzymes in the cell. In characterizing them, we hope to find differences in abundance or modifications which regulate their activity and thus affect the metabolic properties of the cell. Chapter 3 discusses one such proteomic investigation that focuses on mitochondrial proteins and their associated modifications. Mitochondrial dysfunction has long been considered [2] to be physiologically associated with skeletal muscle insulin resistance. Recent proteomics analysis has revealed that post-translational modifications are abundant in mitochondria, specifically acetylation of lysines. Lysine acetylation is found in bacteria and considered [8] evolutionarily ancient, possibly developing before eukaryotic phosphorylation. Lysine acetylation has also been shown [9] to regulate the activities of many mitochondrial enzymes. This observation compels us to determine if mitochondrial acetylation is abundant in humans and whether it corresponds to metabolic abnormalities such as insulin resistance or mitochondrial dysfunction.

To this end, skeletal muscle biopsies were isolated from the vastus lateralis muscle in human subjects before and after an acute exercise bout. The mitochondria from these muscle biopsies were isolated, and HPLC-MS/MS analysis was performed to sequence the mitochondrial proteome. The MASCOT search engine was used to analyze the data and discover post-translational modifications to proteins. We searched for mass shifts that would correspond to lysine acetylation. Overall, we found the mitochondrial proteome to be abundantly acetylated before exercise, and we observed a dramatic reduction in protein acetylation one day after an exercise bout. This reduction of acetylation was more pronounced in insulin sensitive subjects.

Acetylation was discovered on a key mitochondrial protein: the Adenine Nucleotide Translocase-1 (ANT1). It is readily observed in human mitochondrial isolations as it is the most abundant protein found in the inner mitochondrial membrane. A member of the mitochondrial carrier family, ANT1 is the primary channel by which ADP and ATP exchange between the mitochondria and the cytosol. This protein is thus a key requirement for mitochondrial oxidative phosphorylation and delivery of ATP to be used by the cell. Several lysine residues were found to be acetylated, but the acetylation of a single lysine residue (K23) was found to change with exercise in a manner that correlated strongly with insulin sensitivity.

We next found crystal structures of the bovine ANT1 protein in the protein databank (PDB) and found that the homologous lysine (K23) was located at core of the protein. Next, we found prior molecular dynamics studies which simulated the mechanisms of ADP transport and binding into the channel. K23 was shown by several studies to be a key binding site which directly bound to the negatively charged

phosphates of ADP. We reasoned that since lysine acetylation resulted in a loss of positive charge, that this modification detected in humans would reduce binding affinity for ADP, and thus attenuate transport and overall oxidative phosphorylation "flux".

We constructed an acetyl-lysine residue *in silico*, and then conducted molecular dynamics simulations to characterize the effects that acetylation would have on protein structure and dynamics. We additionally simulated binding of ADP ligands and estimated the reduction of binding energies associated with acetylation. Overall, the modeling demonstrated that acetylation of this critical binding residue would have a pronounced impact on ADP binding, and would therefore affect transport. With transport compromised, we predict that there will be physiological consequences to mitochondrial function in the muscle *in vivo*.

Our modeling results are further strengthened by prior mutagenesis work in yeast, which systematically determined the key residues necessary for ADP/ATP transport.[10] The yeast protein was highly homologous, and we discovered that K23 in humans corresponded to a conserved lysine in yeast (K38) that was required for transport activity. With this critical lysine mutated, the yeast could not survive on a non-fermentable carbon source, which indicated a complete failure in mitochondrial oxidative phosphorylation. This gives us full confidence that this acetylation event in humans will have physiological consequences, but what remains to be determined is the fraction of damaged ANT1 proteins within the mitochondria of insulin resistant humans.

The ANT1 story, although incomplete at this stage, has provided a compelling insight into the profound effects that can be caused by a change in a single residue on a

protein within the cell. A loss of a single positive charge in ANT1 caused a dramatic reduction in binding. Indeed, at the molecular scale, as the basis of all cellular function, binding is *everything*. The ANT1 story emerged from a long and careful analysis of the protein structure, reading of numerous simulation studies, careful study of the protein sequences, and luckily finding the right mutagenesis study in another organism to base our central hypothesis. Connecting these "dots" took substantial manual effort to determine how a *single* modified residue on a *single* protein can modulate metabolic activity. Modern proteomics experiments however are finding thousands of modified residues on proteins. It was clear from this study that new systems were needed to find these connections faster.

New sequencing and genotyping techniques are rapidly contributing to our knowledge of genomic variants and their effects on protein function. With this, the field of biological mass spectrometry is growing quickly, and new databases are emerging which comprehensively map countless post-translational modifications. Additionally, databases such as Uniprot [11] offer detailed knowledge of manually and computationally discovered protein functional sites, such as catalytic active sites and binding motifs. Finding functional connections between these positions can be automated, potentially discovering new regulatory protein modifications or pathogenic gene mutations. To accomplish this, a system is needed which simply aggregates these multiple data sources together and finds co-localized residues of interest.

Chapter 4 discusses our solution to this problem. We have developed AMASS, which "amasses" together these databases into a unified platform for exploration and visualization. The AMASS web server is publically available (amass-db.org) and free to

use. The search interface pulls from Uniprot, and displays the protein residue-level annotations from that resource. Additionally, cancer mutation count data is pulled from the Catalogue of Somatic Mutations in Cancer (COSMIC)[12, 13] database, and phosphorylation sites from mass spectrometry datasets are displayed from the PhosphoSite [14] database. The core feature of our service however is direct integration with protein structures: when structures are found in the Protein Databank (PDB) that have at least 50% sequence identity, they are displayed alongside the annotations with an interactive Jmol [15] applet. Clicking anywhere in the protein sequence, or on interactive histogram plots, or on entries in the residue annotations table will show the user the exact location of that residue mapped to the 3D structure. This functionality allows investigators a quick and direct way of discovering novel biophysical interactions between mutations, modifications, and functional sites.

In summary, metabolic syndrome is an enormously complex condition. At the cellular level, metabolic networks regulate the net flux of countless metabolites, and the cell precisely controls the expression of these networks with intricate signaling pathways that process information and make decisions. System wide, it is ultimately the collection of countless cells in a cacophony of hormonal and neurological signaling that regulates the health of the organism. To understand complex diseases such as metabolic syndrome, we must monitor and decrypt this biological chatter. Doing this requires data from all levels, from disease associated variants in the genome, to transcriptional activity in the cell, and to proteomic alterations such as post-translational modifications that precisely regulate the biophysical interactions of signaling molecules and metabolic enzymes. Once we piece together all of this data, the true switches which regulate metabolism can

be found, giving us true targets for therapeutic intervention.

As the era of personalized medicine approaches, we will finally be able to understand metabolic syndrome for what it may be: a collection of numerous completely independent diseases that lead to the same phenotypes: obesity, diabetes, insulin resistance, dyslipidemia, and heart disease. New technologies are permitting personalized diagnostics by enabling individual profiling of genotypic variants. Soon this will be supplanted by inexpensive full genome sequencing. Perhaps someday, transcriptomics or proteomic profiling within target metabolic tissues will allow the scanning of cellular thought processes within patients to determine the bottom-up causality for each individual. For this dream to be realized however, we need to understand the whole system works. We need to map the signaling pathways where cellular decisions are made.

Chapter 2

TRANSCRIPTOMICS ANALYSIS REVEALS GENE EXPRESSIONS ASSOCIATED WITH METABOLIC PHENOTYPES

Introduction

Insulin resistance is a complicated phenomenon whereby a cell makes a "decision" to ignore an insulin signal and pass up on available glucose. When insulin resistance develops at a systemic level, blood glucose levels remain elevated, leading eventually to type 2 diabetes. The elucidation of the molecular mechanisms of this cellular thought process is crucial for the health of millions. In recent decades, the complexities of the insulin signaling pathway have been slowly unraveled to reveal hundreds of potentially involved signaling proteins.

Due to this seemingly unfathomable complexity, much research attention has been spent at the proximal and distal steps of the pathway. At the proximal end lies the insulin receptor and the associated proteins that are recruited upon stimulation, such as the insulin receptor substrate-1 (IRS-1). At this nodal site, substantial progress has been made in finding molecular alternations associated with insulin resistance, with many identified post translational modifications which correlate to disease in human populations.[5] At the distal end of the pathway, the insulin responsive glucose transporter (GLUT4) lies dormant in GLUT4 storage vesicles (GSVs), waiting to be transported and inserted into the cellular membrane. The molecular mechanisms of this process have been equally scrutinized by many labs.[16] Yet everything between these nodal points the pathway diverges along many confusing paths. Several other signaling

pathways in the cell exchange information via crosstalk and feedback, and it is simply too challenging to determine the true points of control without running in circles.

Mass spectrometry has leaded the field due to its capacity of finding interacting partners and mapping protein modifications. Information on both of these aspects of the pathway is needed to understand how the signal propagates from the receptor to GLUT4. Mass spectrometry is limited however in that it only sees the most abundant proteins in the cell, with a typical skeletal muscle dataset finding just several thousand proteins on modern instruments. Signaling molecules such as kinases and receptors are extremely low in abundance and rarely show up in whole proteomes. This limitation can be sidestepped with the use of immunoprecipitation, but this requires specific affinity reagents and insights from the investigator of where to look. Until new proteomics technologies arise that will "flatten" the proteome we will not have an unbiased view of the global proteome.

Although dated by today's standards, gene expression microarrays do give a global view of the entire transcriptome. On whole genome arrays, each gene is measured on independent spots, and the signal is more directly attributable to transcript concentration then is true of mass spectrometric measurement of protein abundance. Although gene expressions do not always predict true protein abundances and fail to account for post translational modifications, there is likely still information to be found that may fill in the gaps of our knowledge of the pathway. Gene expression measurements also give us profound insight into the activation of transcription factor circuits within a cell, and thus can reveal behavioral programs that are initiated during

different stimuli. Transcriptional activation of whole pathways can elucidate what the cell is *thinking* in pathological states.

Most studies using gene expression microarrays compare expression profiles between two discreet states. Subjects are either grouped into diseased/healthy dichotomies or compared pre/post some intervention such as exercise. Gene expressions are then assessed amongst the entire group, and transcripts that change significantly are filtered. This technique favors highly abundant transcripts that more easily pass statistical significance. Signaling pathways however can have changes that are far more subtle, with small changes in transcript or protein concentration yielding large modulations in signal. It is plausible that such small changes will never show statistical significance. Changes in gene expression are also typically considered as a group, and this can confound analysis in instances where a gene will increase in some subjects and decrease in others. Furthermore, grouping subjects into discreet categories of healthy/diseased or focusing just on pre/post comparisons removes much of the information content in the gene expression values. Subjects within any group will have a *continuum* of gene expression values. Furthermore, many of the phenotypes that we measure are also measured on a *continuum*. This information is lost when either gene expressions or subject phenotypes are divided into discreet categories.

In many of these studies continuous-valued phenotypes such as glucose disposal, BMI, and exercise performance measurements are experimentally determined alongside the microarray data, and these values can and should be directly compared with gene expressions. In this work, we present a simple technique whereby externally measured metabolic phenotypes are directly correlated with steady-state gene expression values in

the pre and post exercise conditions, as well as gene expression fold changes induced by exercise. Despite the simplicity of this technique, we find many gene targets that have striking connections with biological function.

Using correlation analysis against glucose uptake, we identify several genes that appear to connect with the wnt signaling pathway, well known to be associated with type 2 diabetes in genome wide association studies (GWAS). Two gene expressions are also found highly correlated with glucose uptake and body mass index (BMI) which have been previously identified as genome wide susceptibility loci for these phenotypes. These findings validate the GWAS data, and additionally unveil the target tissues where these genes exert pathological phenotypes. Furthermore, we demonstrate that many co-correlating genes have known associations with one another, either by forming direct complexes or partaking in the same signaling pathways. Overall, the results provide a birds-eye view of mostly-uncharted areas of insulin action and new mechanisms for insulin resistance and its modulation by exercise. This technique, if adopted in similar studies, may unveil new therapeutic targets to combat insulin resistance, diabetes, and obesity.

Exercise Study

Subjects. Fourteen normoglycemic volunteers took part in this study, which was approved by the Institutional Review Board of Arizona State University. All studies were conducted at the Clinical Research Unit at ASU. Informed, written consent was obtained from all subjects. All of the volunteers were sedentary, with none engaged in regular exercise, and no one reported having a change in body weight for at least 6 months before participating in this study. Subjects were instructed not to exercise for 48 hours before

study and to maintain their usual diet. A medical history, physical examination, 12-lead electrocardiogram, and a complete chemistry panel were obtained, and a 75-g oral glucose tolerance test was performed using American Diabetes Association criteria to exclude nondiabetic subjects with impaired glucose tolerance. No one was taking any medication known to affect glucose metabolism.

Peak aerobic capacity. Peak aerobic capacity (VO_{2peak}) was determined as previously described, with continuous heart rate monitoring . Briefly, exercise was started at a workload of 40 W and increased by 10 W/min until perceived exhaustion or a respiratory quotient of 1.10 was reached.

Muscle biopsy and hyperinsulinemic euglycemic clamp. Euglycemic clamps and muscle biopsies were performed as described . After a 10-h overnight fast, a percutaneous biopsy of the vastus lateralis muscle was obtained with a Bergstrom cannula under local anesthesia one hour before the start of insulin infusion. This biopsy served as the resting, non-exercised control. A stable isotope ([6,6-^{2}H] glucose) was used to trace glucose metabolism, and an insulin infusion rate of 80 mU m^{-2} min^{-1} was used.

Exercise bout with muscle biopsies. All subjects also underwent an exercise test consisting of a single bout of aerobic exercise. The exercise bout was conducted on a separate day after determination of VO_{2peak}, as described previously and at least 1 week either before or after the euglycemic hyperinsulinemic clamp. Design of the exercise bout is given in Figure 1. Subjects reported to the Clinical Research Unit at about 7 AM after fasting overnight and exercised on a stationary bicycle, alternating between 70% (8 minutes) and 90% (2 minutes) of heart rate corresponding to that at VO_{2max}, for a total of

4 sets of exercise, with two minutes rest between sets . Immediately after completing exercise, the subject was moved to a bed where a biopsy of the vastus lateralis muscle was performed within 30 min of the end of exercise.

mRNA expression assays: Muscle RNA was isolated as described. Total RNA (500 ng) was amplified and labeled using the Amino Allyl MessageAmp™ II aRNA Amplification Kit, as per manufacturer's instructions (Life Technologies, Carlsbad, CA). After labeling, antisense RNA (aRNA) was quantified using a NanoVue Spectrophotometer for cyanine 3 (Cy3) dye labeled aRNA concentration, RNA absorbance ratio (260 nm/280 nm), A260 values, ADye values, and dye/probe values. The Cy3-labeled aRNA was then fragmented using Agilent Gene Expression Hybridization Kit (Agilent Technologies, Palo Alto, CA), as per the instructions. The fragmented aRNA was hybridized to the 4x44K Whole Human Genome Microarray (Agilent Technologies, Palo Alto, CA) using a SureHyb DNA Microarray Hybridization Chamber at 65°C, for 17 hours in a rotating incubator. After hybridization, slides were washed in Gene Expression wash buffers 1 and 2 as per manufacturer's instructions, and then scanned with an Agilent DNA microarray scanner (Agilent Technologies, Palo Alto, CA). Feature Extraction Software version 10.5.1.1 (Agilent Technologies, Palo Alto, CA), was used for analysis of the array images.

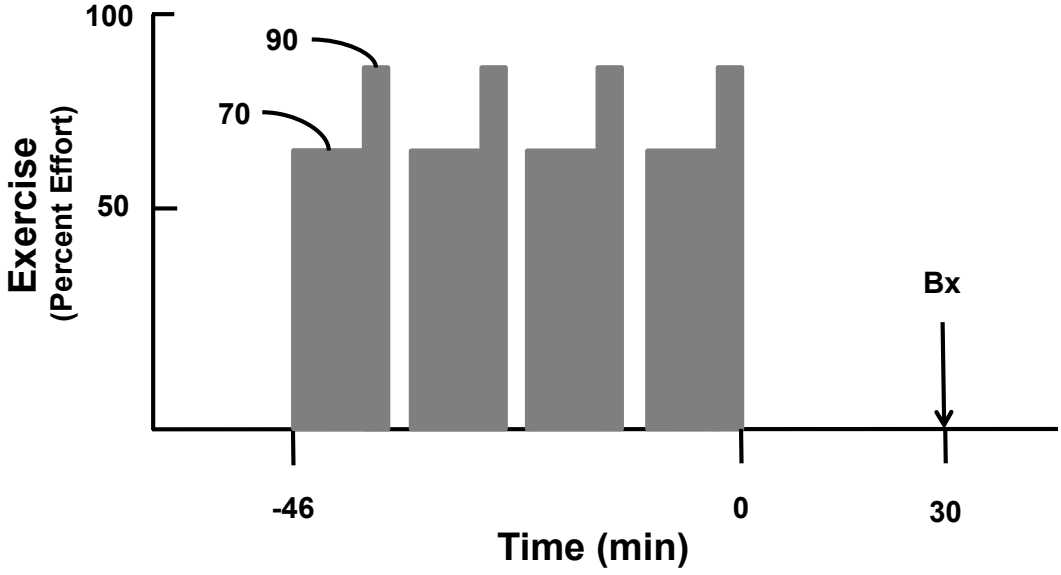


Figure 1 : Design of exercise bout

Data Analysis

Raw expression values for each individual probe were $\log_2$ transformed and quantile normalized. To determine fold change of gene expression between the post-exercise biopsy and the resting biopsy, the $\log_2$-transformed resting expression values were subtracted from the post-exercise values.

Correlations were computed using Pearson's product-moment method between gene expression values and externally measured metabolic phenotypes. Correlations were independently performed with the expression values in both the resting and post-exercise biopsies, as well as the gene expression fold changes induced by exercise. These three tests independently determine gene expressions that may influence glucose uptake (resting biopsy) and exercise performance (post-exercise biopsy), as well as exercise adaptations (fold changes) that are dependent on metabolic parameters. Correlations were calculated utilizing the Scientific Python (scipy) library[17] in a custom script which reads gene expression values from a comma-separated-values (csv) spreadsheet and corresponding subject metabolic phenotypes from a second csv file. The correlations are output into separate results spreadsheets sorted by correlation coefficient. We then performed literature searches on the top 20 most positively correlating probes and the bottom 20 most negatively correlating probes. We focused most of our attention on resting gene expression correlations with glucose uptake (M); however, additional scrutiny was given to all correlations with BMI and post-exercise gene expression correlations with exercise performance.

We observed several spurious correlations with non-biological control spots on the array. These so called "dark corners" are composed of sequences designed to form

hairpins, and otherwise not hybridize to sequences found within the human genome. These control spots are amongst the darkest spots on the array, and have signal levels considered to be background. It is possible that these dark control probes may be located near bright biological spots that have high correlations, and that cross-talk on the array is leading to the spurious correlations. Another possibility is that the spurious correlations are entirely random. Given that these spots have the lowest signal to noise on the array, and that the photon counting error follows Poisson statistics, it follows that the background-level probes will have the highest propensity toward spurious correlations. To compensate for this, we only considered correlating probes that had log-transformed intensities of 9 or higher. (Figure 2) These spot intensities are considered to be outside of the background signal range as determined by the distribution of dark corner intensities.

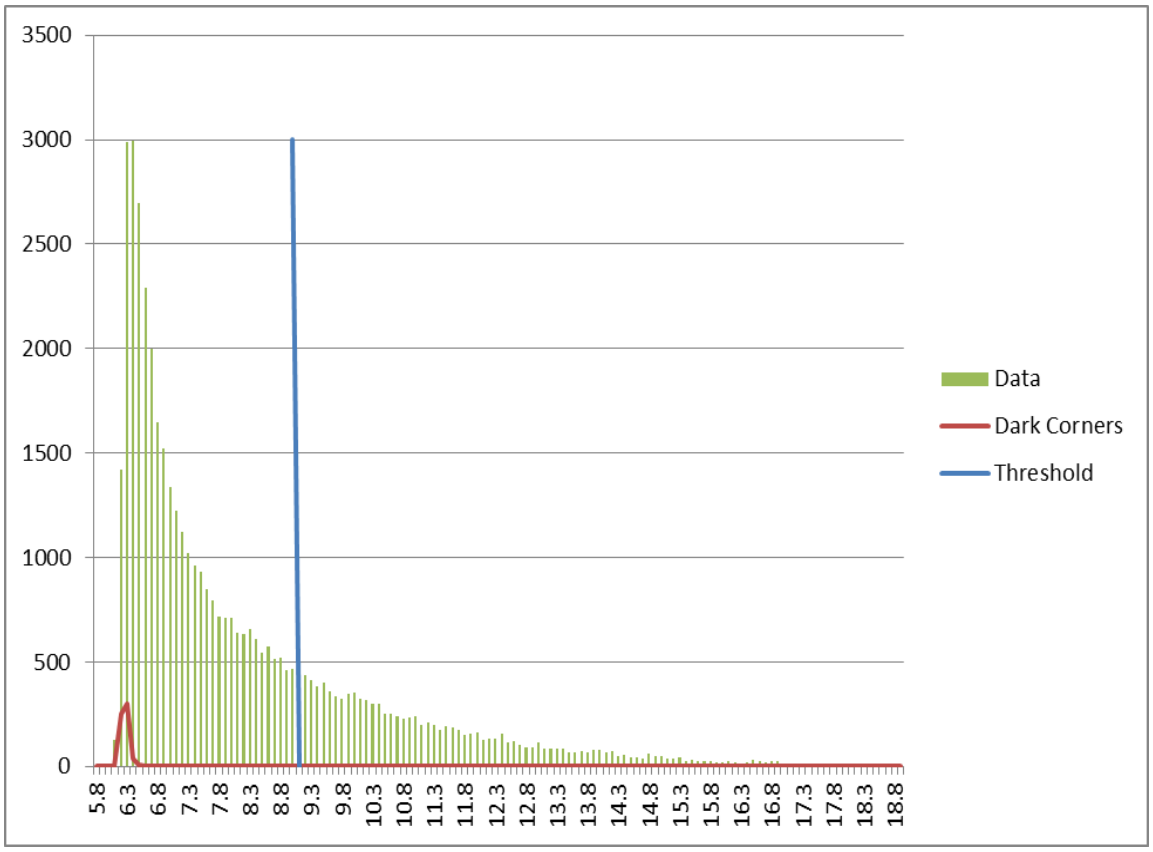


Figure 2 : Microarray signal intensities

Histograms showing distribution of signal intensities on microarray. Dark corner spots show the background level of signal intensities. For resting-state correlations, only the probes brighter than the threshold intensity of 9 (blue) were considered.

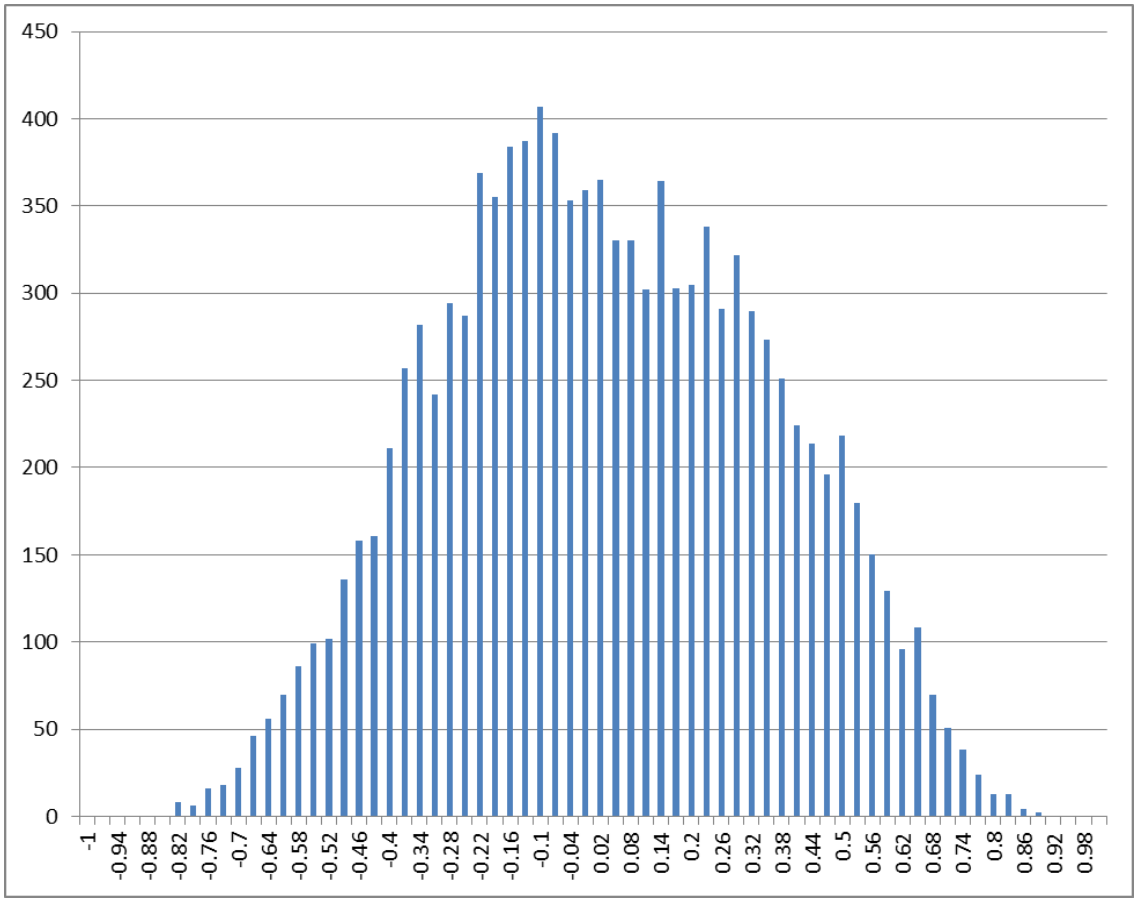


Figure 3 : Distribution of probe correlation coefficients

Results

We performed independent correlations of both the resting and post-exercise biopsy gene expressions, as well as the exercise-induced fold-changes of the gene expressions. Correlations were performed against insulin stimulated glucose uptake and body mass index (BMI) to look for interesting associations. Of the approximately 40,000 probes on the array, most showed little to no correlation, but a small handful of probes displayed significant positive or negative correlation. The distribution of correlations shows the rapid decline of probes with increasing correlation strength. (Figure 3) These top positive-correlating and bottom negative-correlating probes are scrutinized. For each of the genes represented by these probes, we conducted systematic literature review to find potential explanations. This analysis unveils both previously known and novel protein targets which may act as high-control nodes in affecting these various metabolic characteristics.

Resting state expressed genes that correlate with glucose uptake

In our study, the resting "pre exercise" muscle biopsy was obtained just one hour prior to a hyperinsulinemic euglycemic clamp. Given the close time proximity, we hypothesize that gene expression values in this resting state may be most-directly associated with the glucose disposal rate (M mg/kg/min) as measured in this experiment. Thus we focused most of our attention and developed the most comprehensive story based on correlations with this metabolic parameter. Excitingly, genes were uncovered that are genetically associated with diabetes and have direct connections to translocation of the glucose transporter. Furthermore, the results unveil a well-characterized muscle atrophy pathway that may provide a clue to the physiological purpose of insulin

resistance. Top correlating genes are presented for correlations with glucose disposal rate (Table 1) as well as the glucose disposal rate adjusted for fat-free mass. (Table 2) In both of these tables, genes with interesting connections are bold faced and discussed.

Correlation	ID	Description
0.86	**GOPC**	Homo sapiens golgi associated PDZ and coiled-coil motif containing
0.83	**ZBED3**	Homo sapiens zinc finger
0.83	MEA1	Homo sapiens male-enhanced antigen 1 (MEA1)
0.83	BC014384	Homo sapiens
0.83	GLUD1	Homo sapiens glutamate dehydrogenase 1 (GLUD1)
0.82	PDE4A	Homo sapiens phosphodiesterase 4A
0.81	EIF4ENIF1	Eukaryotic translation initiation factor 4E nuclear import factor 1
0.8	**PPP2R3A**	Homo sapiens protein phosphatase 2 (formerly 2A)
0.79	LOC401708	PREDICTED: Homo sapiens similar to Glutamine synthetase
0.79	**PPP2CB**	Homo sapiens protein phosphatase 2 (formerly 2A)
0.79	PUS7	Pseudouridylate synthase 7 homolog (S. cerevisiae) (PUS7)
0.79	MIER1	Homo sapiens mesoderm induction early response 1 homolog
0.78	BC037535	Homo sapiens cDNA clone IMAGE:5274219. [BC037535]
0.77	GIT2	Homo sapiens G protein-coupled receptor kinase interactor 2 (GIT2)
0.77	RAD23B	Homo sapiens RAD23 homolog B (S. cerevisiae) (RAD23B)
0.77	RBMX2	Homo sapiens RNA binding motif protein
0.77	TCEB2	Homo sapiens transcription elongation factor B (SIII)
0.77	FDPSL4	PREDICTED: Homo sapiens similar to farnesyl diphosphate synthase
Correlation	ID	Description
-0.78	RPLP0	Homo sapiens ribosomal protein
-0.78	RPLP0	Homo sapiens ribosomal protein
-0.78	**ADRM1**	Homo sapiens adhesion regulating molecule 1 (ADRM1)
-0.78	**NFKB1**	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
-0.78	**UBE2E3**	Homo sapiens ubiquitin-conjugating enzyme E2E 3
-0.78	PDXDC2	Homo sapiens mRNA; cDNA DKFZp686B0962
-0.78	RPLP0	Homo sapiens ribosomal protein
-0.79	**DYSF**	Homo sapiens dysferlin
-0.8	**EIF3S5**	Homo sapiens eukaryotic translation initiation factor 3
-0.8	EEF1B2	Homo sapiens eukaryotic translation elongation factor 1 beta 2
-0.8	RPS3A	Homo sapiens ribosomal protein S3A (RPS3A)
-0.8	RPLP0	Homo sapiens ribosomal protein
-0.81	A_24_P50489	Unknown
-0.83	**FBXO32**	Homo sapiens F-box protein 32 (FBXO32)
-0.85	LOC440350	Homo sapiens similar to nuclear pore complex interacting protein
-0.85	**NFKB1**	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
-0.87	**FBXO9**	Homo sapiens F-box protein 9 (FBXO9)

Table 1 : Top positive and bottom negative correlating gene expression probes with glucose uptake.

Correlation	ID	Description
0.87	EIF4ENIF1	Eukaryotic translation initiation factor 4E nuclear import factor 1
0.86	**GOPC**	golgi associated PDZ and coiled-coil motif containing
0.85	RAD23B	Homo sapiens RAD23 homolog B (S. cerevisiae) (RAD23B)
0.84	**PPP2CA**	Homo sapiens protein phosphatase 2 (formerly 2A)
0.84	LOC401708	Homo sapiens similar to Glutamine synthetase
0.83	MEA1	Homo sapiens male-enhanced antigen 1 (MEA1)
0.83	**PPP2R3A**	Homo sapiens protein phosphatase 2 (formerly 2A)
0.83	PDE4A	Homo sapiens phosphodiesterase 4A
0.83	THC2674354	Q36LB5_MARHY Nucleoside-diphosphate-sugar epimerases
0.82	ITGB1	Homo sapiens integrin
0.82	PSCD1	Homo sapiens pleckstrin homology
0.82	PUS7	Homo sapiens pseudouridylate synthase 7 homolog (S. cerevisiae)
0.82	ARS2	Homo sapiens ARS2 protein (ARS2)
0.82	SERBP1	Homo sapiens SERPINE1 mRNA binding protein 1 (SERBP1)
0.81	RMI1	Homo sapiens RMI1
0.81	MAPK14	Homo sapiens mitogen-activated protein kinase 14 (MAPK14)
0.81	**ZBED3**	Homo sapiens zinc finger
Correlation	ID	Description
-0.78	**NFKB1**	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
-0.78	RPLP0	Homo sapiens ribosomal protein
-0.79	SFRS8	Homo sapiens splicing factor
-0.79	A_24_P136011	Unknown
-0.79	PHKG1	Homo sapiens phosphorylase kinase
-0.80	**FBXO9**	Homo sapiens F-box protein 9 (FBXO9)
-0.81	CEP63	Homo sapiens centrosomal protein 63kDa (CEP63)
-0.82	RYR1	Homo sapiens ryanodine receptor 1 (skeletal) (RYR1)
-0.82	**EIF3S5**	Homo sapiens eukaryotic translation initiation factor 3
-0.82	RPL6	Homo sapiens ribosomal protein L6 (RPL6)
-0.82	LOC440350	Homo sapiens similar to nuclear pore complex interacting protein
-0.83	**NFKB1**	Nuclear factor of kappa light polypeptide gene enhancer in B-cells
-0.83	RPLP0	Homo sapiens ribosomal protein
-0.83	LSM3	Homo sapiens LSM3 homolog
-0.83	A_24_P50489	Unknown
-0.84	**FBXO32**	Homo sapiens F-box protein 32 (FBXO32)
-0.84	RPS3A	Homo sapiens ribosomal protein S3A (RPS3A)
-0.85	**UCP3**	Homo sapiens uncoupling protein 3 (mitochondrial

Table 2 : Top correlates of glucose uptake adjusted for fat free mass (mg/kg_ffm/min)

Positive Correlates : Diabetes Genetic Associations + Wnt Signaling

Beginning with investigation of positive correlating probes with glucose uptake, we find several genes that participate in the wnt/β-catenin signaling pathway. We will briefly summarize what is known about this pathway and its connections to type 2 diabetes. We direct the reader to an excellent reviews on the canonical wnt pathway[18] and its known role in regulating insulin sensitivity.[19, 20]

The downstream effector of the wnt pathway is the β-catenin protein. In the unstimulated cell, β-catenin concentrations are kept very low due to constant attack by a "destruction complex" consisting of the proteins Axin, Adenomatous Polyposis Coli (APC) , Glycogen Synthase Kinase-3 (GSK-3), and casein kinase 1α (CK-1α). When bound to this complex, β-catenin is phosphorylated, ubiquitinated, and degraded by the proteosome. When the wnt pathway is activated however, this destruction complex is inhibited. A variety of endogenous wnt protein ligands bind to cell-surface receptor complexes composed of Frizzled proteins and LDL receptor-related proteins (LPR). When activated, these receptor complexes recruit a variety of intracellular proteins. Dishevelled (Dsh) is recruited and activated, which inhibits GSK-3. Furthermore the destruction complex members Axin, APC, and GSK-3 are recruited and sequestered at the receptor, thereby rescuing β-catenin from degradation. This leads to an increase in cytosolic β-catenin concentration which then translocates into the nucleus and heterodimerizes with transcription factors of the T-Cell Factor (TCF) and Lymphoid Enhancer-binding Factor (LEF) family. These activated complexes bind to DNA and transcribe target genes.

There is profound genetic evidence that wnt/β-catenin signaling is central to diabetes pathogenesis and regulation of glucose homeostasis in a variety of tissues.[19] For many years, genome wide association studies (GWAS) have clearly shown that the TCF transcription factor TCF7L2 is the single most significant type 2 diabetes susceptibility locus with many pathogenic single nucleotide polymorphisms (SNPs).[20, 21] Despite this clear statistical evidence, the signaling mechanisms by which defects in TCF7L2 promote the development of type 2 diabetes are largely a mystery. Until recent years, most research on pathogenic TCF7L2 polymorphisms have focused on pancreatic development and function, however canonical wnt signaling is not active in adult β-cells after development.[22] This observation is directing research to other tissues of the body involved in glucose metabolism. TCF7L2 is additionally expressed in the brain, liver, muscle, and adipocytes.

A recent study by Abiola et. al. [23] has focused on wnt signaling in muscle, and promotes the hypothesis that wnt signaling modulates the balance between myogenic vs. adipogenic behavior. They established an inverse expression pattern of the endogenous wnt ligand wnt10b with the lipogenic transcription factor Sterol Response Element Binding Protein-1 (SREBP-1). During the development or post-injury regeneration of rat muscles, they observed that wnt10b is dominantly expressed alongside the transcription factor myogenin, whereas SREBP-1 expression was suppressed. After development, they observe that wnt10b becomes undetectable and that SREBP-1 expression is induced. This observational data in animals suggested an inverse relationship between these genes. To test this hypothesis, they performed numerous experiments with myoblasts and myotubes in culture.

They demonstrated that wnt10b knockdown in myoblasts led to a dramatic increase in the expression of SREBP-1, and wnt10b overexpression in myotubes lead to a suppression of SREBP-1. They additionally observed that PPARG was affected, mirroring the expression of SREBP-1. They next demonstrated that SREBP-1 knockdown was sufficient to induce the expression of wnt10b in contracting myotubes. From these combined results, they concluded that wnt10b and SREBP-1 show an inverse expression pattern, even after differentiation. To assess the effects of wnt signaling on insulin sensitivity, they performed 2-Deoxyglucose (2-DG) uptake experiments. They showed that culturing myotubes in high glucose conditions for 48 hours induced marked insulin resistance in which insulin would not stimulate any glucose uptake, but wnt10b overexpression completely restored insulin sensitivity in these cells.

They concluded that the primary signaling node which mediated this balance was Glycogen Synthase Kinase-3 (GSK-3), which they inhibited with the inhibitor BIO. They demonstrated with oil-red-o staining that BIO completely inhibited the formation of lipid droplets in myotubes cultured under high glucose conditions, and furthermore induced the formation of sarcomeres. From all of this evidence, they concluded that wnt signaling promotes a myogenic phenotype in muscle cells and that SREBP-1 inhibits the wnt signaling pathway and promotes an adipogenic phenotype in which intracellular lipids accumulate and insulin sensitivity is reduced. This exciting study suggests that the inhibition of adipogenic behavior by wnt signaling is as true in myocytes as it is in adipocytes.

In another recent study by Yoon et. al. [24] a high-throughput RNAi screen was used to determine genes that affect mitochondrial function. The wnt signaling pathway

was determined to be a direct activator of mitochondrial biogenesis. Furthermore, the study identified that Insulin Receptor Substrate-1 (IRS-1), the critical hub in the insulin signal transduction pathway, was a direct transcriptional target of wnt signaling. They concluded that wnt signaling directly potentiates insulin sensitivity in these cells.

Combined, these two recent studies give enthusiastic evidence that wnt signaling in muscle is directly related to insulin sensitivity. It may soon become evident that the ultimate downstream effector of this signaling is TCF7L2, which would provide validation of the GWAS results and evidence that muscle is a target tissue of its pathogenic effects. Even despite validation of this particular downstream wnt effector molecule, it is clear from these two studies that enhancement of wnt signaling in muscle will increase insulin-stimulated glucose uptake. Interestingly, several positive correlating genes have connections to wnt signaling, which may further hint at the importance of this pathway.

ZBED3: The second-ranking transcript showing high positive correlation (Figure 4) with glucose uptake is ZBED3. Little work has been done characterizing the function of this gene, however recent genomic evidence hints at a connection to type 2 diabetes. In a recent large scale genome wide association study (GWAS), a combined 42542 individuals with type 2 diabetes (T2D) and 60912 controls were studied to determine the genetic loci contributing to the disease.[21] 12 new susceptibility loci were identified that have not yet been reported in prior GWAS literature. One of these SNPs is located upstream of the ZBED3 gene on chromosome 5. Given that type 2 diabetes is a disease associated with insulin resistance, it is exciting that this gene expression correlate is now known to have associated polymorphisms with the disease. Aside from this GWAS study

however, there is little knowledge of the function of ZBED3 with only a single study characterizing its cellular function.[zbed3] It was determined to directly modulate the wnt/ β-catenin signaling pathway.

Chen et al.[25] were seeking novel interacting partners of Axin, which is the core scaffold protein of the β-catenin destruction complex. By using a yeast-two-hybrid assay, they identified zinc-finger BED domain-containing protein 3 (ZBED3). The authors directly confirmed interaction with in-vitro pulldown experiments. Their study also focused on a characteristic PPPPSPT motif found within ZBED3. This characteristic motif was shown to be the binding site for Axin, and that Glycogen Synthase Kinase 3 Beta (GSK3B) phosphorylates this motif at the serine and threonine residues. Furthermore, they demonstrate via pull-down experiments that interaction between Axin and ZBED3 is enhanced with the overexpression of wild-type GSK3B, but not with the overexpression of a kinase-dead version.

The authors went on to determine if ZBED3 overexpression modulated Wnt/ β-catenin signaling. Using a reporter gene in NIHT3T cells, they verified that expression of ZBED3 alone promoted transcription at LEF-1 binding sites. In a complementary experiment, they demonstrate that Wnt/ β-catenin signaling is decreased under conditions of a ZBED3 RNAi knockdown. This evidence illustrates that the concentration of ZBED3 alone may directly modulate wnt signaling and transcription of β-catenin target genes. The function of ZBED3 appears to be that of binding Axin and sequestering it in the same mechanism that Frizzled/LRP receptor complexes sequester it upon pathway activation, making ZBED3 an internal activator of the pathway. Given the prior in vitro

studies in muscle, it is reasonable that this gene would directly promote insulin sensitivity.

GOPC: The most positively correlating transcript to whole body glucose uptake in our exercise experiment (Figure 5) was GOPC, and this transcript is additionally the second most positively correlating gene for clamp glucose uptake adjusted for fat free mass. (Table 2) GOPC, also known as CAL and FIG, was originally described as *Protein Interacting Specifically with TC10* (PIST).[26] Very recently Bogan et. al. provided compelling evidence that GOPC/PIST is a direct activator of glucose transport in the cell.[27]

In the insulin unstimulated state, the insulin responsive glucose transporter (GLUT4) is stored within GLUT4 storage vesicles (GSVs) that are retained intracellularly until an insulin stimulus promotes mobilization, translocation, and ultimately docking and insertion of GLUT4 into the plasma membrane. Once inserted, GLUT4 permits glucose uptake into the cell. This concerted effort requires many proteins and several discreet biochemical stages, however the molecular mechanisms behind this process are incompletely understood.[28]

It has been elucidated that in the unstimulated state, GSVs are physically anchored to the Golgi apparatus by a tether protein known as TUG. It was proposed that insulin stimulation releases this tether and thus permits GSV translocation to the cellular membrane.[29] This hypothesis was confirmed recently by demonstrating that TUG proteins undergo endoprotylitic cleavage into separate N-terminal and C-terminal fragments.[27] The authors demonstrated that the C-terminal half of TUG anchors GSVs to the intracellular golgi apparatus, and that overexpression of a C-terminal fragment

resulted in translocation of GFP-tagged GLUT4 to the plasma membrane in the insulin unstimulated state. Additionally, glucose uptake assays were performed with 2-deoxyglucose, and it was demonstrated that glucose uptake is dramatically increased in both the resting and insulin-stimulated conditions when TUG was knocked down with shRNAs. While the fluorescence experiments confirmed that GLUT4 translocation can be directly activated via TUG knockdown, the direct glucose uptake assay confirms that releasing the TUG tether additionally mediates insertion of GLUT4 into the cell membrane.

Excitingly, GOPC/PIST was shown by Bogan et. al. [27] to be an interacting partner of TUG, and also shown to interact with the Golgi protein Golgin-160. The pull-down experiments imply a model in which GLUT4, TUG, GOPC/PIST, and Golgin-160 are bound in a linear arrangement. They propose a model whereby this complex of four proteins form the putative anchor which immobilizes stored GSVs to the golgi in the insulin unstimulated state. Given that GOPC/PIST is an effector of TC10a, and that TC10a has been shown to be required for insulin stimulated GLUT4 translocation in adipocytes,[30] Bogan et. al. hypothesized that GOPC/PIST may be the effector molecule that facilitates cleavage of TUG. Knockdown of TC10a with siRNAs inhibited TUG proteolytic processing, indicating that TC10a is required for cleavage of TUG and untethering of GLUT4 from the Golgi apparatus. This combined evidence hints that GOPC/PIST may be the final “on switch” for GLUT4 trafficking/insertion, and that the expression levels of this protein may indeed directly predict insulin sensitivity in muscle measured during a glucose clamp.

Intriguingly, beyond its role in regulating the glucose transporter, GOPC has been documented to facilitate the transport of several other proteins to the cell membrane. GOPC has been shown to be an interacting partner of frizzled receptors, which partake in wnt signaling. In a yeast-two-hybrid experiment, GOPC was demonstrated to bind to the c-terminal domain of Frizzled 8. GOPC has a PDZ domain, and was proven to interact with frizzled via a c-terminal PDZ binding motif (Ser/Thr–X–Val).[31] Upon co-transfection of GOPC and frizzled 5 into COS7 cells, the authors observed with immunofluorescent staining that GOPC and frizzled co-localized at the Golgi apparatus. At 20 hours after transfection, GOPC and frizzled were observed co-localized at the cell membrane. Thus GOPC is involved in a secretory pathway that translocates frizzled receptors to the cell membrane. From this evidence, we hypothesize that GOPC may have a secondary effect on insulin sensitivity by increasing wnt signaling, thereby increasing the activity of TCF7L2. This exciting connection warrants further investigation.

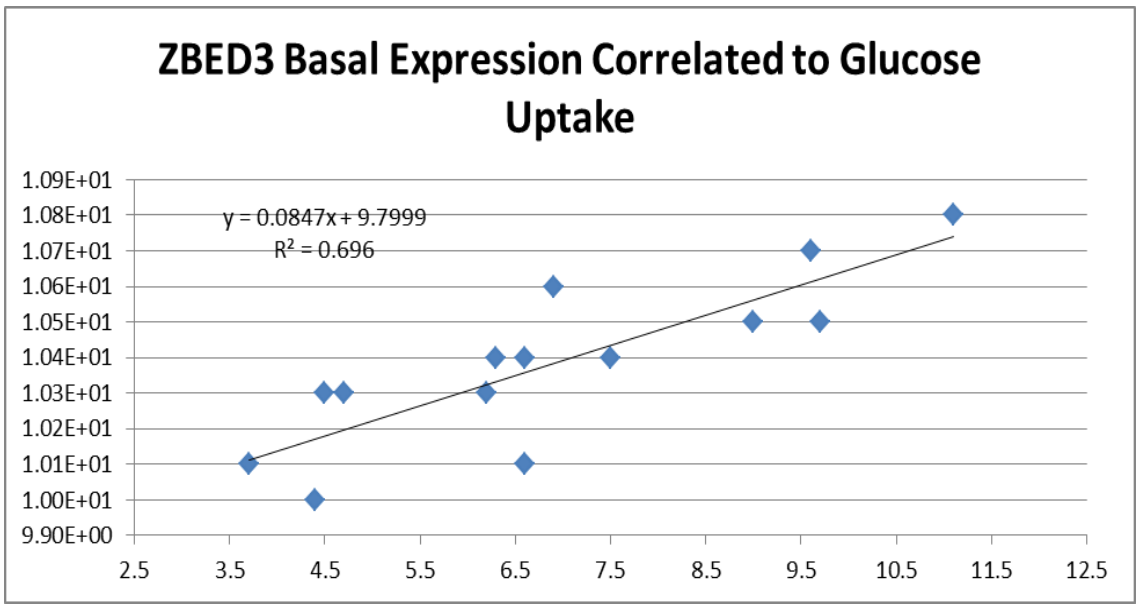


Figure 4 : ZBED3 correlation with glucose uptake

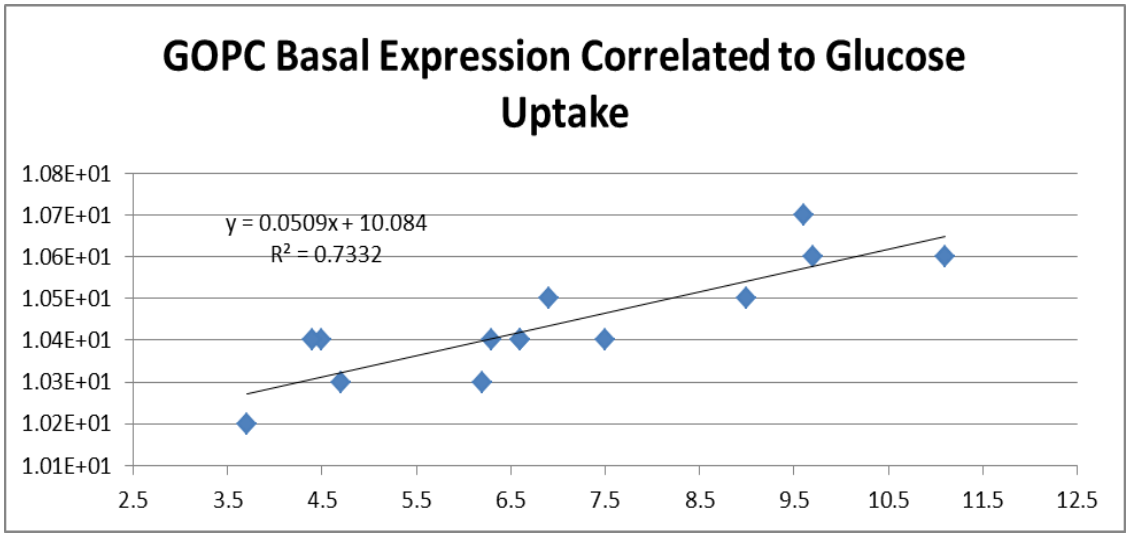


Figure 5 : GOPC correlation with glucose uptake

PP2: Protein Phosphatase 2, also known as PP2A, is a heterotrimeric serine/threonine protein phosphatase. It is composed of structural protein component, a catalytic protein component, and an optional regulatory subunit. The structural subunits are encoded by two isoforms (PPP2R1A and PPP2R1B) as are the catalytic subunits (PPP2CA and PPP2CB). There are many regulatory subunit genes encoding a variety of sizes with diverse functions. Interestingly, amongst the top positive correlates we find a catalytic gene (PPP2CB) and a regulatory subunit gene (PPP2R3A). The co-occurrence of these two probes in the top correlates list hints that PP2A activity may increase insulin sensitivity. Interestingly, PP2A [32] is a positive regulator of wnt signaling by specifically targeting β-catenin for dephosphorylation, although this was mediated with the pr55 regulatory subunit (PPP2R2). The regulatory subunit correlated in our study (PPP2R3A) however has conflicting effects on wnt signaling. The PPP2R3A gene encodes multiple transcript variants that encode both a 72kD (pr72) and a 130kD (pr130) isoform. A study has shown that the pr72 isoform down regulates[33] wnt signaling. A later study by the same lab discovered that the pr130 isoform up regulates[34] wnt signaling. It is the possible that alternative splicing of this gene may precisely fine-tune wnt signaling, but the microarray platform does not allow us to distinguish which variant is being transcribed.

Negative Correlates – The Muscle Atrophy/Wasting Pathway

We next turned our attention to the bottom most-negatively correlating genes. We have provided evidence that the top positive correlates may potentiate insulin signaling through wnt signaling, or even directly activate the GLUT4 translocation/insertion

machinery. These may be putative “on switches” of glucose metabolism. Likewise, gene expressions with strong negative correlation may provide evidence of the mechanisms of pathological insulin resistance. Indeed, if a glucose uptake “off switch” can be found, then novel therapeutic targets can be realized.

Interestingly, many of the strongest negative correlates appear to participate in the well-studied skeletal muscle atrophy pathway. Muscle atrophy and wasting is associated with many causes, including: starvation, disuse, inflammation, diabetes, cancer, and aging.[35] Despite several completely distinct causes for muscle atrophy, the phenomenon is mediated via a common cellular process: the ubiquitin proteosome system (UPS). The small protein tag known as ubiquitin is selectively ligated onto targeted proteins which are then recognized by the proteosome and subsequently degraded.

The ubiquitination of target proteins is accomplished by a cascade of reactions regulated by three classes of enzymes. The ubiquitin protein is first activated by a small set of activation enzymes (E1), and then transferred to one of about 40[36] mammalian ubiquitin conjugating enzymes (E2), and finally attached to a specific target protein by one of many hundreds of ubiquitin ligases (E3). Amongst the negative correlates to glucose uptake, we find a skeletal muscle specific E2 and two E3s, alongside a protein already demonstrated as a ubiquitin target in muscle atrophy. Alongside these genes, we find two probes for NFKB1, which is a transcription factor demonstrated as required for skeletal muscle atrophy via the UPS. This profound enrichment of genes in the most negative correlates suggests that skeletal muscle atrophy, mediated via the ubiquitin proteosome system, may be a direct “off switch” to glucose uptake in vivo.

UBE2E3 is classified as an E2 ubiquitin conjugating enzyme. The E2 proteins in the human genome show a marked tissue[36] specificity, and northern hybridization analysis[37] has demonstrated a strong and specific expression of UBE2E3 in skeletal muscle. Interestingly, in a recent microarray study of insulin sensitive and insulin resistant human subjects, UBE2E3 was the third highest ranked probe that was found significantly elevated in insulin resistant muscle.[38] This substantiates the strong negative correlation with glucose uptake in our analysis and suggests that this gene may be equally linked with both increased ubiquitin proteosome activity and insulin resistance.

E2 ubiquitin conjugating proteins are sometimes colloquially referred to as "ubiquitin carriers", as they carry ubiquitin tags until they eventually form complexes with the hundreds of E3 ubiquitin ligases that ultimately target downstream substrate proteins. A particularly large class of E3 ligases are the SCF (Skp1-Cullin-Fbox) complexes. Cullin forms a central scaffold which binds Skp1 and Roc1. Roc1 binds E2 conjugating enzymes that carry ubiquitin tags. Skp1 binds a large class of Fbox adaptor proteins that target specific downstream protein targets for ubiquitination. The full complex brings together E2 ubiquitin conjugates with target proteins and facilitates the transfer of the ubiquitin. With this modular system, the cell can regulate proteolysis by up regulating various Fbox proteins to handle degradation of different subsets of the proteome. Amongst the highest negative correlates of glucose uptake, we find two Fbox proteins associated with muscle proteolysis: FBXO32/Atrogin and FBXO9. Thus amongst the negative correlates, we see a strong association of both the E2 and E3 steps in the ubiquitination pathway with reduced glucose uptake.

FBXO32 is the fourth most-negatively correlating transcript with glucose uptake, and second when glucose uptake is adjusted for fat-free mass. Muscle atrophy can be readily induced with fasting, since the protein stores in muscle serve as a source of amino acids for hepatic gluconeogenesis in times when dietary glucose is restricted. To identify the specific molecular mechanisms behind this process, Gomes et. al.[3] used cDNA microarrays to identify genes up regulated during fasting. By comparing skeletal muscle samples of fed and fasted mice, they identified FBXO32 as a gene induced seven to nine fold after as little as 16 hours of fasting. Due to its clear role in inducing the atrophy process, the authors named this gene Atrogin-1.

Given that loss of muscle mass is a common characteristic[35] of metabolic disorder, it is compelling that this gene is amongst the most negative correlating genes with glucose uptake. It is furthermore compelling that this gene was originally identified in the skeletal muscle of fasted animals. Indeed the subjects in our study were instructed to fast for 10 hours overnight prior to the hyperinsulinemic euglycemic clamp. The measured M-values, adjusted for fat-free-mass, ranged from about 5 to 13 mg/kg/min. This molecular connection suggests that this range of insulin sensitivity in the human population may relate to an individuals propensity to muscle atrophy during times of fasting. Some individuals may be more prone to waste muscle on shorter time scales then others. This suggests that to understand insulin resistance, as measured by these glucose clamp experiments, we must understand short term muscle atrophy in these subjects and ascertain if it is higher in some subjects than in others.

FBXO9: Interestingly, another Fbox protein occupies the list of strong negative correlates to glucose uptake. FBXO9 is the most negative correlate for glucose uptake,

and the 12th most negative ranking correlate when glucose uptake is adjusted for fat free mass. FBXO9 is not as well-characterized in the literature as FBXO32/Atrogin-1, however in recent years a handful of studies have identified preliminary molecular targets of this protein.

FBXO9 was recently demonstrated to modulate mTOR signaling by mediating the ubiquitination and degradation of specific proteins within the mTOR complexes.[39] Telomere maintenance 2 (Tel2) and Tel2 interacting protein 1 (Tti1) are proteins that regulate the abundance of PI3K-related kinase (PIKK) family members such as mTOR. These proteins are integral complex members that are found in both the mTORC1 and mTORC2 complexes. The authors found that in certain cancers, FBXO9 overexpression selectively targets Tel2 and Tti1 in mTORC1 for ubiquitination and degradation. This causes reduction in signaling through mTORC1. It has long been accepted [40] that mTOR signaling is a central regulatory node for insulin sensitivity.

In a separate study, FBXO9 was demonstrated to be a required signaling molecule for adipogenesis.[41] The authors found that in the white fat tissue of mice, FBXO9 expression was elevated fivefold in mice fed a high fat diet as compared to a control diet. Furthermore, FBXO9 expression was shown to correlate highly with obesity. Expression was also significantly elevated when 3T3-L1 cells were stimulated with various inducers of differentiation, including dexamethasone and insulin. Lastly, the authors demonstrated that FBXO9 knockdown inhibited this adipogenesis. Blots were performed with this knockdown and with a scrambled siRNA control to assess the effects on various signaling molecules. Remarkably they show a marked reduction in β-catenin when FBXO9 is expressed. When knocked down, β-catenin stays elevated and adipogenesis is

inhibited. They also found that FBXO9 expression promoted PPARG expression, and FBXO9 knockdown lead to suppression of PPARG. They determined that this regulation was indirect and mediated via CEBPb.[41]

Altogether, the evidence provided by this study suggests that FBXO9 is a potent stimulator of adipogenic phenotypes in adipocytes. The question now becomes whether or not similar effects would impose an adipogenic phenotype in muscle. The degradation of β-catenin when FBXO9 is expressed is an interesting phenomenon worth investigating. It is unknown whether this is a direct target of FBXO9 ubiquitination or if another mechanism is leading to this suppression. This evidence does suggest however that the wnt signaling pathway may be directly suppressed by this gene. This is evidence of direct crosstalk between this negative-correlating atrophy pathway and the positive-correlating wnt pathway. Wnt signaling is anti-adipogenic, and FBXO9 is now established as pro-adipogenic.

EIF3S5 (eIF3-f) encodes the f subunit of the eukaryotic initiation translation factor 3, which is a major regulator of protein translation within the cell. Intriguingly, eIF3-f was recently shown [42] to be a direct interacting partner of FBXO32/Atrogin in a yeast-two-hybrid screen, and that atrogin ubiquitinates eIF3-f which results in proteosomal degradation. Atrogin knockdown in atrophying myotubes was sufficient to inhibit eIF3-f degradation. They also showed that eIF3-f may be a direct regulator of atrophy by showing that repression of eIF3-f alone induces atrophy and eIF3-f overexpression blocks atrophy in serum starved conditions. In the same study, eIF3-f was electroporated directly into the right leg tibialis anterior muscle of a live mouse, and an empty vector control into the left leg. They measured fiber diameters 14 days later, and

noted an 11% increase in fiber diameter with eIF3-f expression. They concluded that eIF3-f is a critical regulator of muscle hypertrophy by regulating the balance between protein translation and teardown.

Increased protein proteolysis is an essential aspect of muscle atrophy, and it is reasonable that decreased protein synthesis is an additional facet of this condition. This would especially be relevant in conditions of fasting or starvation in which amino acids must be conserved. Alongside its role in the proteolysis of sarcomeric proteins, the targeting of eIF3-f by Atrogin facilitates this dual strategy and fine tunes this balance between anabolism and catabolism.

It is however paradoxical that eIF3-f expression is *negatively* correlated with glucose uptake. One might expect that greater expression of the eIF3-f gene would result in higher quantities of the eIF3-f protein and a concomitant increase in protein translation in the cell. One possible explanation of this observation is that the cell may be upregulating the eIF3-f transcript to compensate for a dramatic reduction of the eIF3-f protein. Many proteins inhibit their own expression by binding to the promoter regions of their own genes. The concentration of the eIF3-f protein may be precisely regulated in the cell due to this process or some other indirect mechanism. To test this hypothesis, the eIF3-f protein itself must be assayed in muscle. Regardless of the seeming opposite direction of this correlation, the fact that eIF3-f interacts with Atrogin and appears as a strong negative correlate provides even more evidence that Atrogin is a core instigator of insulin resistance.

NFKB1: Intriguingly, two separate probes for NFKB1 appear in our list of strong negative correlates. NFKB1 encodes the p50 subunit of the nuclear factor kappa-light-chain-enhancer of activated b cells (NF-KB) transcription factor, which is a core activator of the inflammatory response in cells. There are 10 probes for this gene on the microarray, and they all show highly negative correlation coefficients ranging from -0.64 to -0.84. This is encouraging because several studies have indicated that NF-KB activity is a critical mediator of skeletal muscle atrophy. NFKB1 encodes the p50 subunit of the NF-KB transcription factor.

Very early work has demonstrated that KFKB activity directly activates FBXO32/Atrogin. Hunter et. al. [43] demonstrated that NFKB1 and BCL3 knockout mice are resistant to disuse atrophy. BCL3 heterodimerizes with p50, and activates transcription of target genes such as Atrogin, and thus both genes were shown to be required for muscle atrophy. A later study indicated that inhibition of the upstream IkappaB Kinase Alpha (IKKa) and Beta (IKKb) has a similar effect, along with inhibition of FOXO.[44]

A very recent publication provides far more detail into the transcriptional regulation of atrophy by performing chromatin immunoprecipitation (ChIP) assays to determine the subsets of transcribed genes targeted by NF-KB and Bcl3.[45] The results verify that p50, the protein encoded by NFKB1, does indeed bind to the promoter regions of many genes involved in skeletal muscle atrophy, including FBXO32/Atrogin, FBXO9, and FOXO3. Despite the fact that p50 does not increase in binding during muscle unloading, BCL-3 does increase in binding for several genes, including FBXO32/Atrogin. In the same study, quantitative real time PCR (qPCR) was used to

measure the expression of various atrophy related genes in both BCL3 and NFKB1 knockout mice. During hindlimb unloading, induction of FBXO32/atrogin expression was reduced to half with either gene knocked out.

These results suggest that despite the controversy concerning changes in p50 binding to the FBXO32/Atrogin promoter, binding of p50 to the promoter does stimulate atrophy due to heterodimerization with BCL3. Therefore it is reasonable to assume that increased NFKB1 gene expression and p50 protein content will promote muscle atrophy, possibly in part due to increased expression of FBXO32/Atrogin. The question remaining is why NFKB1 expression is increased in our insulin-resistant subjects as suggested by our correlation analysis. It has been shown[1] that expression of NF-KB proteins such at p65 and p50 (NFKB1) can be increased in L6 muscle cells with exposure to palmitate, and that this was associated with insulin resistance. From this we hypothesize that the accumulation of intracellular lipids may be a driving factor.

Another possibility is that NFKB1 expression is enhanced by extracellular inflammatory stimuli. The expression of NFKB1 can be induced[46] via exposure to tumor necrosis factor (TNF) and by NF-KB activation itself.[47] TNF-a has a longstanding reputation in obesity, diabetes, and insulin resistance. It is found elevated in the adipose tissue in obese human subjects.[48] Furthermore, administration of exogenous TNF-a can induce insulin resistance.[49] This plausible connection suggests that the inflammatory environment defined by elevated TNF-a levels could signal to the NF-KB pathway, and thereby upregulate muscle atrophy pathways.

ADRM1: Further connections to the ubiquitin proteosome system are found with ADRM1, a protein which binds to and activates [50] Uch37, a deubiquitinating enzyme located at the mouth of mammalian proteosomes. It has been shown [51] that knockdown of ADRM1 decreases the degradation of short-lived proteins in the cell. This evidence suggests that the expression levels of ADRM1 may directly accelerate the overall activity of the ubiquitin proteosome system, and thus accelerate skeletal muscle proteolysis.

UCP3: Remarkably, UCP3 shows as a strong negative correlate to glucose uptake rate, but only when normalized to fat free mass. (Table 2) UCP3 is an uncoupling protein found in the mitochondrial of skeletal muscle, and has long been a focus of research into metabolic syndrome. [52] Genetic variants of the UCP3 gene have been associated with insulin sensitivity, type 2 diabetes, and obesity, however the reports are highly controversial. [53] It has been shown that both the mRNA expression and protein content of UCP3 are reduced in diabetic muscle. [54] Recently, Senese et. al. [55] assessed the effects of UCP3 expression on insulin sensitivity by progressively reducing its expression in heterozygous and homozygous knockout mice, and concluded that greater UCP3 expression levels correlated with increased insulin sensitivity. These results are opposite to the correlations we find in this study, which indicate that increased expression correlates with reduced insulin stimulated glucose uptake. Despite the controversy, UCP3 has been a strong focus of research into skeletal muscle insulin sensitivity, and it is the most negative correlating gene for fat-free mass in this analysis. This suggests that it may have a central and complicated role in modulating insulin sensitivity in skeletal muscle that can only be resolved with further attention. It is possible that in our test subjects the UCP3 protein content is substantially reduced and that the gene expression is increased to

compensate for metabolic abnormalities such as increased ROS production. This relationship may behave in an opposing direction to less natural manipulations such as genetic knockout mice or overexpression.

DYSF: Dysferlin is a protein with strong genetic associations to muscular dystrophy, particularly Miyoshi myopathy and limb girdle muscular dystrophy.[56]The defining characteristic of these diseases is muscle weakness brought about by wasting. It is unclear how increased expression of this gene would contribute to this process, as it is typically the loss of function that is associated with disease. Dysferlin is responsible for repairing damaged cell membranes [57] in states of muscle inflammation, so it is possible that upregulation of the transcript is occurring in these conditions during insulin resistance.

Exercise-mediated gene expression changes correlating with glucose uptake

Given that the resting biopsy was performed immediately prior to the hyperinsulinemic euglycemic clamp, those expression values are the most relevant in deducing transcriptional control of glucose uptake. We did not consider correlations with the post-exercise gene expressions since no post-exercise glucose clamp was performed. We did however consider the hypothesis that the insulin sensitivity of a subject may relate to the changes in gene expression with exercise. Thus we performed correlations against the exercise-induced expression fold changes. (Table 3) Unlike correlations with the resting gene expression values, we did not observe any spurious correlations with dark corner control spots. We speculate that this is due to the fact that differencing the log-transformed values will negate any spurious correlations due to random noise. To this end, we considered all highly-correlating probes regardless of dim signal intensity.

Correlation	ID	Description
0.86	**KCNMB2**	potassium large conductance calcium-activated channel
0.86	**ACP5**	acid phosphatase 5
0.85	**LIN28B**	lin-28 homolog B (C. elegans) (LIN28B)
0.84	HAO2	hydroxyacid oxidase 2 (long chain) (HAO2)
0.83	AL157439	[AL157439]
0.81	**PRKAG2**	cDNA FLJ90194 fis
0.81	GPX4	glutathione peroxidase 4 (phospholipid hydroperoxidase)
0.80	**AF349445**	maxi-K channel HSLO mRNA
0.80	TJAP1	tight junction associated protein 1 (peripheral) (TJAP1)
0.79	TRIM43	tripartite motif-containing 43 (TRIM43)
0.79	AF116673	PRO1925 mRNA
0.79	**DGKZ**	diacylglycerol kinase
0.79	KIAA1772	KIAA1772 (KIAA1772)
0.79	SART3	squamous cell carcinoma antigen recognized by T cells 3
0.78	CBFA2T2	core-binding factor
0.78	**PFKFB3**	6-phosphofructo-2-kinase/fructose-2
0.78	**PFKFB3**	6-phosphofructo-2-kinase/fructose-2

Correlation	ID	Description
-0.76	BG115911	BG115911
-0.76	THUMPD1	THUMP domain containing 1
-0.77	RAB1A	RAB1A
-0.77	P2RY1	purinergic receptor P2Y
-0.77	AK026659	cDNA: FLJ23006 fis
-0.77	MAP3K7IP3	MAPKK 7 interacting protein 3
-0.77	BRCA1	breast cancer 1
-0.78	AA235942	cDNA clone IMAGE:687981 5'
-0.78	HAPLN1	hyaluronan and proteoglycan link protein 1
-0.79	C13orf27	chromosome 13 open reading frame 27
-0.79	THC2618142	Unknown
-0.79	ESR1	estrogen receptor 1 (ESR1)
-0.80	DDX52	DEAD (Asp-Glu-Ala-Asp) box polypeptide 52
-0.81	FLJ14959	cDNA FLJ14959 fis
-0.82	UEVLD	UEV and lactate/malate dehydrogenase domains
-0.83	ESR1	estrogen receptor 1
-0.87	A_24_P916453	Unknown

Table 3 : Top positive and negative gene expression fold change correlates with glucose uptake

BK Channels: Large conductance potassium channels, also known as Maxi-K, Big-K (BK), or slo1 channels, are calcium/voltage activated conductors of potassium ions. They function in smooth muscles to regulate vasodilation. Genetic polymorphisms of these channels are highly associated with hypertension in humans.[58-61] BK channels are composed of tetrameric alpha subunits which form the pore. Regulatory beta subunits encoded by four different genes can bind with the alpha tetramer to precisely regulate channel opening. Interestingly, amongst our top correlates we observe a probe encoding a beta subunit (KCNMB2) and a probe representing the alpha subunit (AF349445).

Interestingly, the probe detecting the alpha subunit does not sense the putative alpha subunit gene (KCNMA1), but instead it probes for a single 132bp exon which has been documented to differentially splice with the alpha subunits and modulate their activity.[62] It is well known that alpha subunits of BK channels are heavily spliced to precisely fine tune their calcium and voltage sensitivity in various tissues. This is evidence that exercise is possibly mediating this change, however the probe intensities for both of these genes are incredibly weak. This may be because the transcripts are likely only present in the blood vessels in the muscle biopsy samples. Confounding this issue further, there are two probes on the array detecting KCNMB2 expression. Alongside the high positive correlation probe we observe, the other probe has a strong negative correlation (-0.6) with glucose disposal rate. Both probes target the 3' UTR of the gene.

Given the low signal intensity and correlative ambiguity of this signal, we would not focus on this gene had we not uncovered evidence that it is relevant in diabetic vasculopathy. It has been demonstrated that the beta subunits of BK channels are downregulated substantially in diabetic blood vessels, and that this compromises BK

conductance, leading to reduced vasodilation.[63, 64] Remarkably, Zhang et al.[65] showed that the reduction of beta subunits was caused by targeted proteolysis mediated by FBXO32/atrogin and FBXO9. These E3 ligases are upregulated in the smooth muscle of diabetic blood vessels, and ultimately activated by FOXO3A under high glucose conditions. They recognized that FBXO32/Atrogin in particular recognized the beta subunits with a PDZ-binding motif (T/S-X-V) located in most BK-beta proteins. Mutation of the PDZ binding motif eliminated the proteolysis induced by FBXO32/Atrogin overexpression and rescued the BK channel conductance in cell culture models.

This surprising connection to diabetic vasculopathy and to the atrophy pathway hints that the regulation of vascular tone may underlie the insulin sensitivity gains in response to exercise. Unfortunately, the ambiguity of the two counter-correlating probes for KCNMB2 confounds this finding.

ACP5/TRAP: A strongly-correlating expression change is observed with the gene ACP5, known more commonly as tartrate resistant acid phosphatase (TRAP). Functionally TRAP is known to specifically dephosphorylate the protein osteopontin (OPN), which has roles in bone remodeling [66, 67] and in immune system function.

A recent study shed light on the potential metabolic relevance of TRAP by creating mice which overexpressed it in adipose tissue. The mice developed[68] severe hyperplastic obesity, but remarkably the mice remained completely insulin sensitive, in contrast to typical models of obesity. Elevated TRAP levels in the adipose tissue were shown to be secreted from resident macrophages. Since macrophage residence has long been thought [69] to directly mediate the connections between obesity and localized

insulin resistance, this hints that TRAP may partake in signaling pathways which regulate inflammatory response. Despite the increased fat mass of these animals, the investigators saw no evidence of global activation of inflammatory macrophages or the associated cytokines. Perplexingly, TNFa mRNA was elevated in adipose tissue, but the level of TNFa protein was not elevated in serum. Overall, these mice lacked the global inflammatory response typically observed in obesity.

Further evidence of this inflammatory connection comes from human genetics studies. Genetic deficiency of human TRAP causes [70] a disorder of chronic muscle inflammation known as inflammatory myositis, a condition which itself is characterized by marked insulin resistance of muscle and metabolic disorder.[71] In individuals with TRAP deficiency, phosphorylated osteopontin accumulates in the serum and is associated with autoimmunity.[72] This evidence suggests that the lack of TRAP results in chronic levels of inflammation, as opposed to an overexpression of TRAP in mice which abrogated inflammation.

The putative TRAP dephosphorylation substrate, osteopontin, has much stronger connections with metabolic disorder. Aside from its role in bone remodeling, osteopontin acutely attracts macrophages by binding to surface integrin receptors. Osteopontin is expressed and secreted by regenerating muscle and is thought to facilitate the invasion of inflammatory macrophages during repair.[73] It is also thought to mediate the inflammatory component of muscular dystrophy.[74] Given that macrophage infiltration into tissues is commonly associated with metabolic disorder, this protein may be a viable therapeutic target.

Osteopontin knockout mice are completely protected from diet induced insulin resistance.[75] Furthermore, injecting obese mice with antibodies which bind and neutralize osteopontin inhibits inflammation and its associated insulin resistance.[76] This effect was associated with attenuated infiltration of macrophages into the liver and adipose tissues in these animals. This evidence suggests that osteopontin is a dominant factor regulating the residence of inflammatory macrophages, and that it is these macrophages which impart the systemic insulin resistance associated with diet induced obesity.

The interactions between osteopontin and macrophages are mediated via binding to both b3-integrin receptors and CD44 receptors. Chemotaxis is enabled via the CD44 receptor, and macrophage activation is mediated through the b3-integrin receptor. Binding to the b3-integrin however requires phosphorylation of osteopontin.[77] Given that TRAP dephosphorylates osteopontin to mediate detachment from the avb3 integrin receptors of osteoclasts during bone resorption, it is possible that TRAP mediates a similar detachment in the context of macrophage infiltration.

This combined evidence leads us to hypothesize that TRAP expression at the site of regenerating muscle may deactivate the migration of macrophages into the muscle tissue. If this process reduces the level of resident inflammatory macrophages or reduces their level of activation, then the lower burden of inflammatory cytokines may lead to an improvement in insulin sensitivity. This model may explain the complete lack of inflammation in the TRAP overexpressing mouse study. In adipocytes, TRAP may downregulate the local inflammatory response, ultimately lowering cytokines which induce localized insulin resistance in the fat pads. In our system, we are unaware of

whether the TRAP is expressed by the muscle or resident leukocytes, or if the expression is localized to the muscle. These questions warrant further attention.

PRKAG2: AMP-activated protein kinase (AMP) is a critical energy-sensing protein in the cell which detects intracellular ratios of adenosine monophosphate and adenosine triphosphate. AMPK has long been thought to be a critical regulator of insulin sensitivity in the cell. The diabetic drug metformin is thought to target its activation.[78] Furthermore, there has been much focus studying the modulation of AMPK activity induced by exercise interventions.[79-83]

AMPK is composed of catalytic α, and regulatory β and γ subunits. The γ subunits contain multiple domains which bind AMP. Two isoforms for each of the α and β subunits, and three isoforms of the γ subunit are present in the human genome. The gene found correlated with glucose uptake in our dataset, PRKAG2, specifically encodes the γ-2 subunit.

A comprehensive study[84] has characterized the molecular functions of each of these γ subunits in humans. The three γ subunits of AMPK have markedly different tissue expression profiles, with the γ-1 subunit as the dominant isoform in nearly all tissues, except for the brain which has near equal partitioning of all three. In muscle the γ-1 subunit shows ~90% expression, with ~10% of γ-2 and negligible γ-3. The authors also characterized the enzymatic fold activation of AMPK with these various subunits. By measuring kinase activity without exposure to AMP, and with exposure to 200uM AMP, they denoted that the fold activation of various γ

subunits was markedly different. Regardless of the catalytic subunit used, the AMP-binding γ-2 subunit showed nearly twofold activity levels over the γ-1 or γ-3 subunits.

These results demonstrate that even though the γ-2 subunit has low expression in skeletal muscle, it is more sensitive to AMP levels and can activate the catalytic subunit more strongly. Thus small changes in the fractional expression of this particular subunit of AMPK during exercise may result in a marked increase in AMPK signaling, and thus mediate marked improvements in insulin sensitivity.

Lin28B: Micrornas (miRs) have a profound influence regulating the cellular transcriptome by hybridizing to select mRNA transcripts and promoting their degradation. In particular, the let-7 family of miRs have long been understood to inhibit cellular growth and proliferation, particularly in cancer models. Lin28 proteins suppress let-7 miRs by binding their precursors and targeting them for degradation. This process downregulates the pool of let-7 miRs, and thereby serves as a counterbalance against their inhibition of cellular growth. Most focus on the Lin28/let-7 axis has been in cancer research.

A recent paper connected the entire Lin28/let-7 axis to insulin metabolism and diabetes.[85] Transgenic mice were created overexpressing either Lin28a or LIN28B, and showed enhanced glucose tolerance and resistance to high-fat induced obesity. Opposingly, knocking out Lin28a in muscle or inducing the expression of let-7 produced mice with glucose intolerance. This data demonstrated that the Lin28/let-7 axis is a central regulator of glucose metabolism.

To determine the molecular mechanisms behind this adaptation, *in vitro* work was conducted in C2C12 myoblasts as a model of muscle. By overexpressing Lin28a they noted an increase in the mRNA levels of several genes in the insulin signaling pathway, including the insulin receptor, insulin receptor subtrate-1 (IRS-1), multiple PI3-Kinase

isoforms, and constituents of mTOR complexes such as Raptor and Rictor. They then used the TargetScan algorithm to find cononical let-7 miR binding sites in the 3' untranslated regions (UTRs) of several of these genes. Using a luciferase reporting system, they verified that overexpression of let-7f suppressed the insulin receptor, IRS2, AKT2, Rictor, and several other proteins relevant to insulin signal transduction. This overall evidence suggests that expression of Lin28 proteins may be a core regulatory node by which a cell can fine-tune insulin sensitivity. Our correlations analysis demonstrates that Lin28B may have high control strength on insulin sensitivity. Interestingly, the expression change of this transcript was zero amongst the group of subjects. Roughly half of the participants downregulated Lin28B, and the other half upregulated it, with a maximal expression change of just %15.

PFKFB3: 6-phosphofructo-2-kinase / fructose-2,6-bisphosphatase 3 is an enzyme that catalyzes the formation or degradation of fructose 2,6-bisphosphate. This molecule is a potent allosteric activator of 6-phosphofructo-1 kinase, which is the rate limiting enzyme in glycolysis. PFKFB3 is an inducible form of this enzyme, and its overexpression has been shown to markedly increase glucose uptake in adipocytes.[86] In a recent study, Guo et. al. [87]demonstrated that PFKFB3 is a transcriptional target of peroxisome proliferator-activated receptor gamma (PPARG) and is the enzyme responsible for improved insulin sensitivity obtained by treatment with thiazolidinediones, commonly used to treat diabetics. Disruption of PFKFB3 in knockout mice eliminated the effectiveness of these drugs in improving insulin sensitivity. Intriguingly, they also demonstrated that adipose tissue cytokine expression and inflammation was markedly increased in these animals. This combined evidence suggests

that this enzyme may similarly regulate glucose uptake and inflammation in skeletal muscle. Upregulation of this enzyme after exercise may be a critical regulatory node for increasing glucose uptake.

DGKZ: Diacylglycerol (DAG) is a cell membrane lipid that serves as an important second messenger in a variety of signaling pathways. It is formed by the cleavage of phosphatidylinositol 4,5-bisphosphate (PIP2) by the enzyme phospholipase C. After cleavage, the resulting DAG remains in the membrane and allosterically activates protein kinase C. It is widely accepted that DAG concentrations in the cell cause insulin resistance [88] by activating protein kinase C, which then inhibits insulin signaling via phosphorylation of insulin receptor substrate 1 (IRS-1).

Diacylglycerol kinases are enzymes which convert diacylglycerol (DAG) to phosphatidic acid. DGKZ is the zeta isoform of this class of enzymes. By reducing the concentration of DAGs, these enzymes attenuate protein kinase C signaling, and thus should improve insulin signaling. It has already been shown that downregulation of diacylglycerol kinases delta in skeletal muscle leads to insulin resistance. [89]Given that these enzymes catalyze the same reaction, it is likely that the increase in DGKZ expression after exercise would mediate similar effects.

Correlations with BMI:

In addition to correlations with glucose uptake measured in the hyperinsulinemic euglycemic clamp, we also performed correlations between the subjects' body mass index (BMI) against steady-state gene expression values and exercise-induced fold changes. (Table 4) Amongst the fold change correlations, the top correlating gene (LRP1B) has a genetic connection with BMI as determined by a recent large scale GWAS screen.

The Genetic Investigation of ANthropomorphic Traits (GIANT) consortium performed an association analysis against a sizeable population of 249,796 subjects. 18 new BMI susceptibility loci were uncovered [90] in this study. Among them was a BMI-associated polymorphic allele was found upstream the LRP1B gene with a minor allele frequency (MAF) of 18% in the population. Most of the previously-discovered BMI associated loci are associated with genes expressed predominantly in the central nervous system, particularly in the hypothalamus where energy expenditure and feeding behavior is regulated.[90] We have demonstrated however that the change in LRP1B expression due to exercise intervention is the most correlated fold change with the measured BMI. This evidence suggests that LRP1B may act in skeletal muscle or other peripheral tissues and affect systemic obesity. This connection clearly warrants further investigation.

The expression of this gene decreased with exercise intervention, and the decrease was greater in subjects with a lower BMI. (Figure 6) The immediate question that arises is whether or not the relationship is causal. Certainly pathogenic variants in LRP1B, if present in our subjects, would be causal as the BMI could have no effect on the genotype. These gene expression changes however could be influenced by BMI due to some

internal feedback mechanism. If the expression changes are causative of BMI however, then these results could suggest that test subjects respond to exercise in profoundly different ways. Exercise, as an intervention, may be less effective in subjects that do not appropriately downregulate the LRP1B gene. Future work must determine if these gene expression changes are associated with genotype at this pathogenic allele, and how these expression values are regulated due to the environmental circumstances of the cell. The polymorphism upstream the gene could affect resting expression levels or response to key transcription factors that are activated during exercise.

A recent study [91] describes what little is known on the function of LRP1B. It is primarily found expressed in the brain, thyroid gland, and skeletal muscle, and at low levels in other tissues. Interacting ligands have been identified with immunoprecipitation and mass spectrometric identification. LRP1B binds to many signaling ligands including fibrinogen, clusterin, and vitronectin. The study also demonstrated binding and uptake of ApoE containing very low densitiy lipoproteins (VLDLs).

We can infer additional function based on homology with other receptors in the LRP family. LRP1B is so named because it shares ~60% sequence homology [91] with LRP1, which is far more characterized in the literature. Interestingly, a proteomics analysis[92] of GLUT4 storage vesicles (GSVs) has shown that LRP1 is present abundantly and similarly trafficked to the cell membrane by insulin. The authors showed that LRP1 directly interacts with the luminal domains of the GLUT4 protein and the cytoplasmic tail of AS160, an important signaling node in the GLUT4 activation pathway. They showed that knockdown of LRP1 also reduced the expression levels of GLUT4, indicating that it is integral to the regulation of glucose uptake. If LRP1B has

similar or possibly competing functions of its close homologue LRP1, then it may affect dynamics of these GSVs.

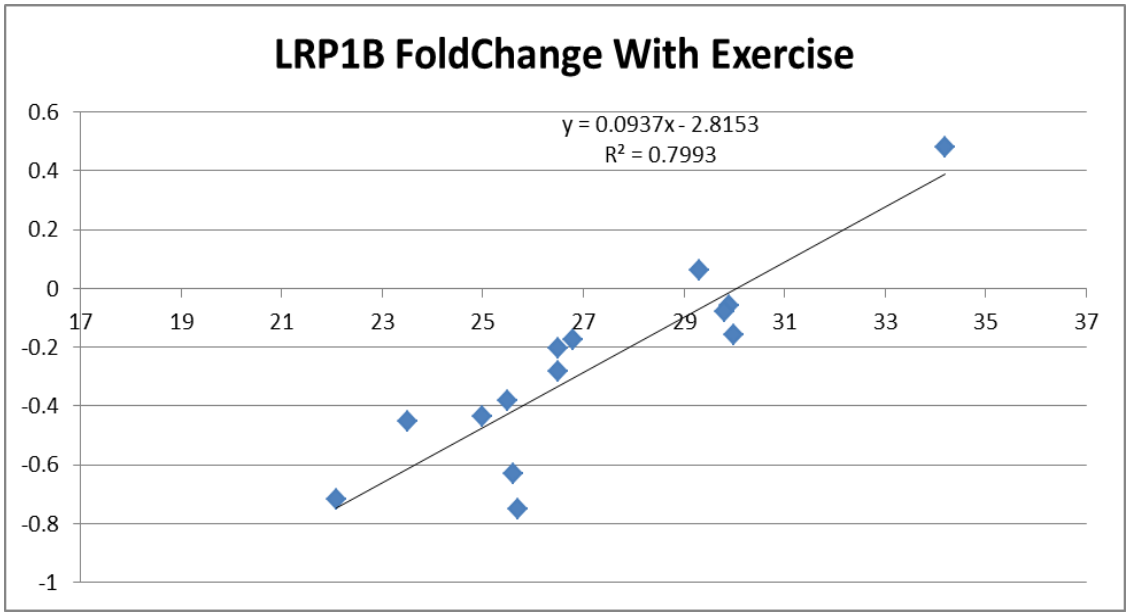


Figure 6 : BMI correlation with LRP1B fold change

Interestingly, LRP1 also modulates wnt signaling. Two related proteins from the same family, LRP5 and LRP6 bind to frizzled receptors to form functional receptor complexes at the cell membrane. These co-receptor proteins are required for frizzled receptor activation. LRP1 has been shown [93] to form similar dimers with frizzled-1 receptors in the membrane, however instead of activating the complex it sequesters the receptor and prevents activation by endogenous wnt3a ligands. Paradoxically, another study [94] explored the role of LRP1 knockout in mouse fibroblasts, and concluded that LRP1 is required for canonical wnt5a signaling. Without the wnt signaling pathway active, fibroblasts that are stimulated for adipocyte differentiation accumulate cholesterol and hypertrophy. These studies, although conflicting, point to a complicated role for LPR1 in regulating lipid transport and wnt pathway signaling.

If LPR1B behaves in a similar or opposing manner, then its exercise-mediated differential expression will likely have large impact on cellular metabolism by regulating wnt signaling, which as we have discussed appears to have a large influence on the lipid storage/utilization phenotype of muscle and fat cells.[23] The connection to cholesterol transport is also relevant as it possibly explains the wnt-surpressive effects of SREBP-1, a transcription factor which is activated when cellular cholesterol levels are low. Indeed, an inverse relationship is observed between the expression levels of LRP1 and various SREBP isoforms.[95]

Correlation	ID	Description
0.89	**LRP1B**	low density lipoprotein-related protein 1B (deleted in tumors)
0.88	AK024372	cDNA FLJ14310 fis
0.87	TPM3	tropomyosin 3 (TPM3)
0.86	TMEM126A	transmembrane protein 126A (TMEM126A)
0.85	LGI1	leucine-rich
0.84	C12orf60	chromosome 12 open reading frame 60 (C12orf60)
0.84	ZNF415	zinc finger protein 415 (ZNF415)
0.83	IKBKAP	inhibitor of kappa light polypeptide gene enhancer in B-cells
0.82	CRNKL1	crooked neck pre-mRNA splicing factor-like 1 (Drosophila)
0.82	GPAM	glycerol-3-phosphate acyltransferase
0.81	A_24_P341489	Unknown
0.81	THC2574008	Q475C5_RALEJ (Q475C5) Phosphate butyryltransferase
0.81	MYL2	myosin
0.81	HINT1	histidine triad nucleotide binding protein 1 (HINT1)
0.81	ALDH3A2	aldehyde dehydrogenase 3 family
0.81	SPIN1	spindlin 1 (SPIN1)
0.80	A_24_P49800	Unknown

Correlation	ID	Description
-0.80	POLD1	polymerase (DNA directed)
-0.80	BQ072652	AGENCOURT_6763016 NIH_MGC_118 cDNA clone
-0.80	TMEM48	transmembrane protein 48 (TMEM48)
-0.80	AA464246	Soares ovary tumor NbHOT cDNA clone
-0.81	BC017937	Homo sapiens
-0.82	THBS1	thrombospondin 1 (THBS1)
-0.82	GUSB	Glucuronidase
-0.83	GUSB	Glucuronidase
-0.83	HOXA11S	homeo box A11
-0.83	HEXIM2	hexamthylene bis-acetamide inducible 2 (HEXIM2)
-0.84	GUSB	Glucuronidase
-0.85	A_32_P80295	Unknown
-0.86	CR610211	full-length cDNA clone of Neuroblastoma
-0.86	BC089451	cDNA clone IMAGE:30554612. [BC089451]
-0.88	C12orf54	chromosome 12 open reading frame 54 (C12orf54)

Table 4: BMI correlations with exercise fold change

Discussion

We have performed a simple correlation analysis to determine which genes may activate or deactivate insulin-stimulated glucose uptake in the cell. Positive correlating genes appear to have connections to wnt signaling, of which the terminal effectors are transcription factor complexes of the protein β-catenin and proteins of the TCF or LEF families. The TCF family protein TCF7L2 is the single most associated gene with type 2 diabetes in GWAS studies. This gene has been shown to directly regulate transcription of the insulin receptor. Furthermore, signaling through this pathway increases insulin-stimulated glucose uptake and decreases the accumulation of intracellular lipids in muscle cells. Overall, these results suggest that wnt signaling is a primary determinant of the cells insulin sensitivity. Our top correlating gene, GOPC, has been established to directly traffic GLUT4 storage vesicles to the cell membrane, as well as Frizzled receptors which are the putative inputs to the wnt signaling pathway. Our second top correlate ZBED3 has been demonstrated to increase wnt signaling in the cell by binding and sequestering Axin. Overall, these results give new gene targets which regulate insulin sensitivity at the level of transcriptional regulation.

The most negatively correlating genes are associated with low insulin-stimulated glucose uptake, and are thus possibly the genes which bring about the state of insulin resistance. These genes have clear connections to the ubiquitin proteosome system, and thus the skeletal muscle atrophy pathway. Muscle atrophy and wasting is induced predominantly in states of fasting and starvation. With the reduction of dietary carbohydrates, the body must preserve blood glucose concentrations for proper functioning of the central nervous system. Hepatic gluconeogenesis can produce glucose

from free amino acids, and muscle provides the primary reserve of amino acids in the body. Thus, during starvation, putative atrophy activating genes such as Atrogin are dramatically upregulated to target muscle proteins for degradation by the proteosome. To further preserve blood glucose concentrations during starvation, peripheral tissues become insulin resistant, thereby decreasing their uptake and utilization of glucose. [96, 97] This connection between muscle atrophy and insulin resistance makes physiological sense. During starvation, muscles must cease the import of glucose and excrete free amino acids to support gluconeogenesis.

Muscle glucose uptake can be stimulated by both insulin signaling and exercise induced contraction. [98] This reflects the fact that glucose is needed for both anabolism at rest and as an energy source during exercise. These signaling pathways are disjoint with different sensory inputs, but nonetheless result in the common output of GLUT4 translocation and membrane insertion. We thus postulate that if glucose uptake must be blunted, that there are evolutionary advantages of inhibiting the signal at either the sensory inputs or at the terminal GLUT4 control points. If glucose uptake must be inhibited only for anabolic purposes, then it follows that inhibition of the insulin signaling pathway at the location of the insulin receptor or IRS1 is the most sensible point of control. If however, due to starvation, perceived starvation, or internal cellular stress, glucose uptake must be inhibited for any reasons, then it follows that inhibition at the level of GLUT4 translocation is the most advantageous.

Given this physiological connection, we hypothesize that proteolysis of a key unidentified protein may be the central *off switch* for insulin sensitivity. Intriguingly, our top positive correlating gene for insulin sensitivity selectively targets proteins for

membrane trafficking with a Ser/Thr-X-Val PDZ motif. GOPC transports the frizzled receptor [31] to the membrane, which activates the wnt signaling pathway. GOPC additionally inhibits the surface expression of the beta-adrenergic receptor by binding to a characteristic EKSV motif.[99] Curiously, in smooth muscle cells, Atrogin targets BK channels for degradation [65] by binding to the same PDZ binding motif (Ser/Thr-X-V). We thus hypothesize that Atrogin and GOPC share a common binding partner protein, and that this unknown protein may directly modulate insulin sensitivity. This protein could be a frizzled receptor, which are the activators of wnt signaling. Aside from Atrogin, FBXO9 was another strong negative correlate we scrutinized, and is also an Fbox ubiquitin ligase like Atrogin. It was demonstrated to be required for adipogenesis in adipocytes, and its expression was associated with the reduction of β-catenin. [41] It is unknown what molecular targets of FBXO9 induce the adipogenic program, but this program appears counter to the anti-adipogenic effects of wnt signaling. These connections could form a direct bridge between the insulin sensitivity *on switch* and the insulin resistance *off switch* in the cell.

If the skeletal muscle wasting pathway is indeed the cause of insulin resistance, we must now understand why it is activated in diabetic muscle. Much work has shown that muscle wasting is directly tied to fasting and starvation, and this connection between proteolysis and insulin resistance makes physiological sense. Other conditions however can induce muscle wasting, including disuse, sepsis, and inflammation. [35] NFKB signaling has been shown to be a direct upstream activator of muscle wasting [43, 45] due to either inflammation or disuse. Chronic inflammation is however already associated with insulin resistance in adipose tissue and muscle, and this chronic inflammation is

directly associated with resident activated pro-inflammatory macrophages [69] which secrete inflammatory cytokines and stimulate NFKB signaling. These macrophages are attracted and polarized by phosphorylated osteopontin [77] which can be dephosphorylated [66] via TRAP. Our correlations with exercise-induced expression fold changes demonstrate that insulin sensitive subjects overexpress TRAP to a larger degree with exercise intervention. This single protein may thus be a dominant upstream regulator of skeletal muscle insulin sensitivity. By dephosphorylating osteopontin, macrophage recruitment and activation is blunted, which reduces proinflammatory cytokine secretion, NFKB activation, and thus induction of skeletal muscle wasting. Under this model, it is systemic inflammation that may ultimately cause insulin resistance in non-fasting pathological states; however it is mediated through pathways that evolved to handle fasting conditions.

The polarization of proinflammatory macrophages in skeletal muscle may thus be the key upstream aspect of the disease manifestation. Understanding *why* these macrophages are activated is of paramount importance. In a recent work [100] it was demonstrated macrophages are resident in human skeletal muscle and correlate with insulin resistance. Furthermore, the study performed co-culture experiments with myotubes and macrophages, and showed that the presence of macrophages caused dramatic upregulation of proinflammatory cytokines such as TNFa and IL-1, as well as the gene Atrogin. This evidence provides a clear connection between macrophages and muscle wasting, likely due to TNFa-mediated NFKB signaling and induction of Atrogin. This evidence provides a plausible bridge between the immune system and the intracellular pathways we previously scrutinized.

Conclusion:

We have presented a technique whereby gene expression values are directly compared to externally-measured metabolic phenotypes. Despite the seemingly naïve simplicity of this technique, our analysis has elucidated several interesting findings that may be worth pursuing as novel therapeutic targets.

Future work using this technique could involve the simultaneous expression profiling of several distinct tissues, such as the brain, adipose tissue, and liver. This would permit a more cohesive understanding of how organs are communicating and coordinating global physiological responses such as starvation or inflammation. Additionally, a wide expansion of the characterized metabolic phenotypes may reveal more interesting patterns. Such phenotypes can include characterizations of fat metabolism and fuel selection, as well as the signaling of other metabolic hormones of interest. With these strategies, the adoption of new technologies such as RNAseq may enable more direct and reliable quantification of gene expressions values.

Two highly-correlating genes have been found that are associated with pathology at the level of genomic variance. A gene correlating with peripheral glucose disposal has pathogenic alleles associated with type 2 diabetes. Another gene, whose exercise-induced gene expression fold change correlates with BMI, has polymorphisms associated with BMI. These findings indicate that gene expression differences in skeletal muscle can correlate with disease status, and suggest that these expression differences may depend on the genotype of promoter-resident polymorphisms for these genes. Further validation of this hypothesis is needed. Genome wide association studies (GWAS) can survey hundreds of thousands of people for the underlying genetic contributors to phenotype,

however these studies fail to provide an organ-specific context for where molecular defect may manifest. This technique however suggests that these two susceptibility loci may exert influence in peripheral tissues such as skeletal muscle. Furthermore, these observations may point to differences in exercise tolerance and benefits to health, and can therefore guide lifestyle interventions for people with differing genotypes.

Alongside these two surprising hits, an additional glucose-uptake associated gene (GOPC) has been found which has a clearly documented connection with glucose transporter trafficking to the cell surface. This prior *in vitro* work investigating the molecular function of this gene gives enthusiastic validation that our correlation-based technique is a sensible approach to finding new targets to investigate. Overall, the targets found with this technique provide enthusiastic evidence that such gene expression correlations may unveil new insights into biological function if applied to the appropriate questions.

Chapter 3

ADENINE NUCLEOTIDE TRANSLOCASE ACETYLATION MAY ALTER ADP TRANSPORT IN MUSCLE

Clinton Mielke, Carrie G. Sharoff, Natalie Lefort, Jeanine M. Cordova, Paul R. Langlais, Andrew J. Bordner, Jerez A. Te, S. Banu Ozkan, Wayne T. Willis, and Lawrence J. Mandarino

ABSTRACT

Proteomics techniques reveal that lysine-acetylation is abundant in mitochondria. This study was undertaken 1) to determine how mitochondrial protein acetylation is regulated in human skeletal muscle, and 2) to use molecular modeling techniques to understand the consequences of acetylation of ANT1, the key protein required for ADP/ATP exchange across the inner mitochondrial membrane. Healthy subjects had euglycemic clamps and an exercise bout with muscle biopsies. A number of acetylated mitochondrial proteins were identified. Acetylation decreased after exercise ($P < 0.01$), and the acetylation at rest correlated with insulin sensitivity ($r = 0.603$, $P < 0.01$). One acetylation site in ANT1 that was deacetylated after exercise, K23, is in the ADP binding pocket. Its extent of deacetylation correlated with insulin sensitivity ($r = 0.90$, $P < 0.05$). Molecular dynamics methods were used to characterize changes in ANT1 structure/dynamics induced by acetylation. Ensemble docking predicted the ADP binding site at a pocket of critical positively-charged residues. We calculated binding affinities that appear physiologically relevant, and predicted substantial reductions in affinity upon acetylation. Given these

estimates, acetylation of ANT1 could have dramatic physiological effects on ADP/ATP exchange; the greater deacetylation of K23 in insulin sensitive individuals could enhance mitochondrial ADP/ATP exchange.

INTRODUCTION

Post-translational modifications of proteins provide a multidimensional mechanism for regulation of protein structure and function. For example, protein phosphorylation is an important means of regulating proteins, but other post-translational modifications also affect protein function. Protein lysine acetylation is a widespread, but poorly understood phenomenon. Once thought only to be important in regulation of chromatin structure, recent evidence shows lysine acetylation regulates function in other proteins. Lysine acetylation is an ancient, evolutionarily conserved regulatory mechanism, widespread across phyla, and more conserved across prokaryotes and eukaryotes than is the rest of the proteome [101]. In prokaryotes, acetylation coordinates metabolic flux [8, 102], so it is no surprise that mitochondrial proteins are over-represented in the acetylome of human cells [9, 101]. A published acetylome of human liver shows many metabolic enzymes to be acetylated [9].

Other more targeted studies show that reversible acetylation regulates mitochondrial function in mammals. Acetylation of enoyl-CoA hydratase, malate dehydrogenase, and long chain acyl CoA dehydrogenase regulates their activities [103]. Fasting alters acetylation of mitochondrial proteins, suggesting reversible acetylation may regulate function [104]. Sirtuins are evolutionarily conserved protein deacetylases, with a requirement for NAD+ linking them to metabolism [105]. Sirt3, 4, and 5 are present in mitochondria (42-44), but Sirt3 regulates mitochondrial fatty acid oxidation and appears

to be the primary mitochondrial sirtuin [103]. Fasting induces Sirt3 expression in mouse liver, and deletion of Sirt3 gene results in higher concentrations of acylcarnitines and triglycerides, due to defective fatty acid oxidation [103]. Sirt3 null mice also have reduced ATP production in liver and are cold-intolerant [106], and exhibit increased acetylation of mitochondrial proteins critical for energy homeostasis.

Little is known regarding regulation of the human mitochondrial acetylome or the effects of acetylation on function of specific human mitochondrial proteins. The unclear relationship between mitochondrial dysfunction and insulin resistance in humans provides a compelling rationale for understanding regulation of lysine acetylation in human tissues, particularly in mitochondria. In this study we combine clinical research, mass spectrometry, and molecular modeling approaches to show that the mitochondrial acetylome is related to insulin action in human muscle. We identify adenine nucleotide translocase (ANT)1 as a mitochondrial protein that is extensively acetylated. As the primary mitochondrial carrier for transport of ADP/ATP across the inner mitochondrial membrane, ANT1 is vital for oxidative metabolism. We show that, in humans, acetylation of mitochondrial proteins is related to insulin sensitivity, and that deacetylation of ANT1 is regulated by muscle contraction in a manner dependent on insulin sensitivity. Finally, we provide the results of molecular modeling simulations that indicate that acetylation of a key regulated lysine residue on human ANT1 may affect its ability to translocate adenine nucleotides across the inner mitochondrial membrane.

EXPERIMENTAL PROCEDURES

Subjects. A total of 21 normoglycemic volunteers took part in this study, which was approved by the Institutional Review Board of Arizona State University. All studies

were conducted at the Clinical Research Unit at ASU. Informed, written consent was obtained from all subjects. None of the volunteers engaged in regular exercise, nor did they report a change in body weight for at least 6 months before participating in this study. Subjects were instructed to not exercise for 48 hours before study and to maintain their usual diet. A medical history, physical examination, 12-lead electrocardiogram, and a complete chemistry panel were obtained, and a 75-g oral glucose tolerance test was performed using American Diabetes Association criteria to exclude nondiabetic subjects with impaired glucose tolerance. No one was taking any medication known to affect glucose metabolism. Other *in vivo* and mitochondrial function data and data for the subjects who had only a euglycemic clamp and not an exercise study has been reported elsewhere [107].

Muscle biopsy and hyperinsulinemic, euglycemic clamp. Euglycemic clamps and muscle biopsies were performed in all participants, as described [7]. After a 10-h overnight fast, a percutaneous biopsy of the *vastus lateralis* muscle was obtained with a Bergstrom cannula under local anesthesia one hour before the start of insulin infusion. A stable isotope ([6,6-^{2}H] glucose) was used to trace glucose metabolism, and an insulin infusion rate of 80 mU m^{-2} min^{-1} was used.

Exercise bout with muscle biopsies. Five of the 21 subjects, besides having a euglycemic clamp, also underwent an exercise test consisting of a single bout of aerobic exercise. The exercise bout was conducted on a separate day after determination of VO2peak and at least 1 week either before or after the euglycemic hyperinsulinemic clamp. Design of the exercise bout is identical to that previously described [2]. Subjects reported to the Clinical Research Unit at about 7 AM after fasting overnight. Subjects

exercised on a stationary bicycle, alternating between 70% (8 minutes) and 90% (2 minutes) of heart rate corresponding to that at VO_{2max}, for a total of 4 sets of exercise, with two minutes rest between sets [2]. After completing exercise, the subject was discharged. They returned the following morning, again after an overnight fast, for a final biopsy of the *vastus lateralis* muscle, 24 hours after the end of exercise.

Peak aerobic capacity. Peak aerobic capacity (VO2peak) was determined on a separate day in these five subjects as previously described, with continuous heart rate monitoring [108]. Exercise was started at a workload of 40 W and increased by 10 W/min until perceived exhaustion or a respiratory quotient of 1.10 was reached.

Mass spectrometry. Whole muscle lysates were processed as previously described [4]. Mitochondria were isolated from freshly obtained muscle biopsies as described [109]. Muscle lysate or mitochondrial proteins were resolved by one-dimensional SDS-PAGE. The band corresponding to the molecular weight of ANT1 was excised, and samples were prepared and analyzed using a Thermo HPLC-nanospray LTQ-FTICR hybrid mass spectrometer, as described [4]. A data-dependent tandem mass spectrometry approach was utilized to identify peptides in which a full scan spectrum was acquired followed by collision-induced dissociation (CID) mass spectra of the 10 most abundant ions in the survey scan that was acquired using the FTICR mass analyzer in order to obtain high resolution, high mass accuracy data.

Data analysis and bioinformatics. Extracted tandem mass spectra were assigned charge states and searched against the IPI_HUMAN_v3.59 database (http://www.ebi.ac.uk/IPI/) as described [4]**.** Search parameters used were as described

[4], with the addition of acetylation as a modification and allowing for three miscleavages. The normalized spectral abundance factor (NSAF) method was used for label-free quantification [4, 109].

Creation of a *Homo sapiens* Homology Model. The crystal structure of bovine ANT1 is known (Protein Data Bank (PDB) entry: 1OKC) [110]. Protein sequence alignment showed that the bovine transporter is nearly identical to the human transporter with a sequence similarity of ~96%. We performed homology modeling using Modeller [111] to obtain a predicted human model. The protein was missing two residues at the N-terminus and four residues from the C-terminus. To add them, we used the molfacture plugin from VMD [112] to construct peptides of the first three (MGD) and last eight (YDEIKKYV) amino acids of the human sequence. Using the multiseq plugin [113] of VMD, we then performed structural alignments with the STAMP algorithm [114] to fit the glycine residue of the N-terminal extension and the first four residues (YDEI) of the C-terminal extension to the protein model. To visualize the final human homology model, we again used the VMD multiseq plugin to perform a whole-protein structural alignment with the original bovine crystal structure. Having aligned the structures, we then merged the original crystallographic water molecules from the bovine structure to ensure the protein would maintain proper hydration within tight pockets of the protein. The carboxyatractyloside (CATR) inhibitor, cardiolipins, and detergents in the original pdb were not included.

Preparation of Acetylysine Patch. To properly model the effects of acetylated lysine residues in MD simulations and docking experiments, we prepared an acetyllysine (KAC) patch. Following the same strategy described in Eichenbaum et. al. [115], we used the acetyl group from the existing CHARMM ACE patch. Linking the acetyl group to the lysine residue was accomplished by deleting the lysine amine-group hydrogens and forming an amide bond between the lysine NZ atom and the CY atom of the acetyl group. The atom type of NZ was changed to NH1, and the bond length was set to 1.345 Å, reflecting the equilibrium bond length defined in the CHARMM22 parameter file. Atom partial charges were assigned as performed in Eichenbaum et. al. [115].

Molecular Dynamics Simulations. All-atom simulations were performed with the NAMD [116] simulation package (Version 2.8). The VMD psfgen plugin was used to create an Xplor protein structure file (psf) of the human ANT1 structure using the CHARMM22 [117] topology for proteins and lipids with CMAP correction. Bulk water molecules were then added with the Solvate program using a minimum water shell thickness of 6 Å and 8 gaussians to define the solvent boundary.

To insert the protein into a lipid bilayer, water molecules that were within a 30 Å slice centered on the transmembrane axis of the protein were trimmed away. A 90x90 Å palmitoyl-oleyl-phosphatidyl-choline (POPC) bilayer was constructed with the VMD membrane builder plugin. The protein was then centered into the membrane by translating its center of mass, and lipid molecules within 3 Å of the protein were removed. Finally, the prepared membrane+protein system was fully solvated in a 90x90x90 Å water box with 18 chloride counterions to neutralize the net protein charge.

The solvated ANT1+Membrane system was equilibrated via a 4-stage process 1) The membrane was "melted" by performing a 0.5 ns simulation in which only the lipid tails were permitted to move. 2) With the protein constrained with a harmonic energy function, the entire system was subjected 1000 steps of minimization followed by 0.5 ns of equilibration at 310K. 3) An additional 0.5 ns of equilibration was performed without protein constraints to further equilibrate the entire system. 4) Finally, an additional 25 ns of equilibration was performed while holding the area of the lipid patch constant. Equilibration of the system over this lengthy trajectory was validated by plotting the RMSD of the system and observing a convergence.

We refer to the endpoint of this wild-type equilibration as the "equilibrated reference system", as it served as a common input structure to all of our computational experiments. Using the acetyllysine patch on this system, (see below) we prepared a series of three acetylated systems at lysines 10, 23, and 92. In each of these acetylated systems, the furthest chloride ion from the channel was deleted to re-neutralize the net charge. The wild-type (WT) and three modified systems (K10, K23, K92) were then subjected to re-minimization, 10 ns of independent equilibration, and finally 30 ns of full MD simulation in the isothermal-isobaric (NPT) ensemble. These four systems lacked the presence of ADP, and are thus hereafter referred to as the "Apo-ANT1" simulations.

Electrostatics Analysis. VMD PMEPOT [118] plugin was used to compute the electrostatic potentials of the assembled ANT1 systems and simulation results. To compare the functional effects of lysine acetylation, potentials were computed on the static protein after the application of an acetyllysine (KAC) patch to the K10, K23, and K92 residues. Isosurfaces of the potential were visualized in VMD. To produce 2D slices

and line-profile plots of the potential, we used the Python GridDataFormats library to import the OpenDX files exported from PMEPOT. Profiles were calculated by interpolating along a line drawn through the lumen of the channel. The origin of the system was taken as the geometric center of the alpha-carbon atoms belonging to the three proline-hinge residues (P28, P133, P230) which define the gate of the transporter. 2D slices were rendered with the matplotlib Python library.

Ensemble Docking. Autodock4 [119] was used to perform docking of ADP to an ensemble of protein conformations derived from 1500 equidistant snapshots from each of our four Apo-ANT1 systems of MD simulations. A custom TCL script was written to convert the NAMD dcd trajectory files. For each simulation snapshot, the protein was aligned to the equilibrated reference system and the protein structure was written to a separate file.

Instead of assigning partial charges via the Gasteiger method implemented in Autodock tools, we preserved the partial charge assignments previously defined from the CHARMM22 topology files used for all-atom simulations. To export these partial charges explicitly, PQR files were exported from the NAMD trajectories. The prepare_receptor4.py script from Autodock Tools (ADT) was then used to convert the PQR files to the native Autodock PDBQT format. We manually validated that protein snapshots from wild-type and acetylated systems each had +18 and +17 net charge respectively.

Autogrid4 was invoked with a default grid spacing of 0.375 Å. The bounding box was created for the lower pocket of the channel, positioned at the center of mass of the

protein. An XYZ grid size of 40x40x40 points was used. Docking was then performed on each ANT1 MD snapshot conformation with a Lamarckian genetic algorithm. A population of 256 was used, with 2500000 maximum energy evaluations and 27000 maximum generations. We adjusted these parameters extensively, but arrived at similar distributions for each experiment. The ADP ligand was exported from VMD into PQR format and prepared manually with Autodock Tools. Nonpolar hydrogens were merged, and 9 active torsions were defined by default. The net charge of ADP was reported at -3. After docking, binding energies and estimated binding constants were parsed from results files with the unix programs egrep, awk, and sort.

In addition to ensemble docking, we also performed ADP docking to the static equilibrated wild-type system. The docked pose was similar to the best poses found in ensemble docking to the Apo-ANT1 ensembles. The docked ligand was subsequently incorporated into the ANT1 structure, and two additional MD simulations were performed on both the wild-type (WT) and acetylated K23 systems. These bound "ANT1+ADP" simulations were subsequently analyzed and subjected to independent rounds of ensemble docking.

RESULTS

Acetylated proteins are widespread and abundant in human skeletal muscle. To assess the extent and sites of lysine acetylation in skeletal muscle proteins, muscle biopsies were lysed. Lysate proteins were resolved by one-dimensional SDS-PAGE. Samples were processed and mass spectrometry analysis performed as previously described [4, 109]. Acetylated proteins and specific sites of acetylation in whole muscle

are shown in (Table 5). Data shown in (Table 5) are a compilation of acetylation sites observed in 3 separate biopsies from different healthy individuals. Both mitochondrial and non-mitochondrial are represented, with myosin isoforms being heavily acetylated, as assessed by the large number of spectra for acetylated myosin proteins observed (not shown).

Mitochondrial proteins are heavily acetylated. To obtain better proteomic coverage of mitochondrial proteins, mitochondria were isolated from freshly obtained skeletal muscle biopsies (n=3) from healthy volunteers. Mitochondrial protein lysates were prepared and analyzed in the same manner as whole muscle lysates. Simplifying the complexity of whole muscle proteins revealed many more acetylation sites in a variety of mitochondrial proteins. The abundance of myosin spectra in whole muscle lysates masks other ions of lower abundance. Results of this analysis are shown in (Table 6). Some proteins, such as trifunctional enzyme, were found to be acetylated in whole muscle and mitochondrial lysates. Others were found mainly in the mitochondrial lysates. Proteins involved in several aspects of mitochondrial function were represented, including shuttle activity (aspartate aminotransferase), fatty acid metabolism (hydroxyacyl-coenzyme A dehydrogenase), citric acid cycle (malate dehydrogenase), NADP-dependent isocitrate dyhydrogenase, ATP synthesis (ATP synthase subunits b,d, and O), reactive oxygen species metabolism (Mn-superoxide dismutase), and adenine nucleotide translocase activity (ANT1).

Protein	Acetylated amino acid sequence
Myosin-2	$_{773}$AGLLGLLEEM(ox)RDDK(ac)$_{786}$
Myosin-7	$_{146}$K(ac)RSEAPPHIFSISDNAYQYM(ox)LTDR$_{169}$
Myosin light chain 3	$_{124}$NK(ac)DTGTYEDFVEGLR$_{138}$
Gamma-enolase	$_{202}$DATNVGDEGGFAPNILENSEALELVK(ac)$_{227}$
Histone H3.2	$_{19}$KQLATK(ac)AAR$_{27}$
Voltage-dependent anion-selective channel protein 1	$_{35}$SENGLEFTSSGSANTETTK(ac)VTGSLETK$_{61}$
Transitional endoplasmic reticulum ATPase	$_{2}$ASGADSK(ac)GDDLSTAILK$_{18}$
Cytochrome b-c1 complex subunit 2, mitochondrial	$_{353}$AAYNQVK(ac)TIAQGNLSNTDVQAAK$_{375}$
Tubulin beta-2C chain	$_{319}$GRMSMK(ac)EVDEQMLNVQNK$_{336}$
Myoglobin	$_{65}$HGATVLTALGGILK(ac)K$_{79}$
	$_{33}$LFK(ac)GHPETLEK$_{43}$
Isoform 1 of Myomesin-1	$_{372}$YK(ac)GEFDETRFHAGASTM(ac)PLSFGVTPYGYASR$_{402}$
Isoform 2 of L-lactate dehydrogenase A chain	$_{2}$ATLK(ac)DQLIYNLLK$_{14}$
Trifunctional enzyme subunit alpha, mitochondrial	$_{494}$K(ac)MGLVDQLVEPLGPGLK$_{516}$
	$_{213}$VIGM(ox)HYFSPVDK(ac)M(ox)QLLEIITTEK$_{230}$

Acetylation sites are indicated as K(ac); methionine oxidation M(ox)

Table 5 : Acetylation sites in whole muscle homogenates

Mitochondrial Protein	Acetylated amino acid sequence
ADP/ATP translocase 1	$_{2}$GDHAWSFLK(ac)DFLAGGVAAAVSK$_{23}$
	$_{8}$DFLAGGVAAAVSK(ac)TAVAPIER$_{31}$
	$_{81}$YFPTQALNFAFK(ac)DK$_{94}$
Trifunctional enzyme subunit beta	$_{292}$LK(ac)PAFIKPYGTVTAANSSFLTDGASAM(ox)LIM(ox)AEEK$_{325}$
Trifunctional enzyme subunit alpha	$_{214}$K(ac)M(ox)GLVDQLVEPLGPGLKPPEER$_{235}$
	$_{494}$VIGM(ox)HYFSPVDK(ac)M(ox)QLLEIITTEK$_{516}$
Aspartate aminotransferase	$_{82}$K(ac)AEAQIAAK$_{90}$
	$_{356}$TQLVSNLK(ac)K$_{364}$
	$_{394}$LIK(ac)EFSIYM(ox)TK$_{404}$
	$_{397}$EFSIYM(ox)TK(ac)DGR$_{407}$
	$_{410}$VAGVTSSNVGYLAHAIHQVTK(ac)$_{430}$
Hydroxyacyl-coA dehydrogenase	$_{166}$FAGLHFFNPVPVM(ox)K(ac)LVEVIK$_{179}$
Isoform 1	$_{254}$GDASK(ac)EDIDTAM(ox)K$_{266}$
Malate dehydrogenase	$_{53}$LTLYDIAHTPGVAADLSHIEK(ac)AAVK$_{78}$
	$_{177}$ANTFVAELK(ac)GLDPAR$_{191}$
	$_{298}$GIEK(ac)NLGIGK$_{307}$
	$_{302}$NLGIGK(ac)VSSFEEK$_{314}$
	$_{329}$K(ac)GEDFVK$_{355}$
Superoxide dismutase [Mn]	$_{54}$HHAAYVNNLNVTEEK(ac)YQEALAK$_{75}$
	$_{115}$GELLEAIK(ac)R$_{123}$
ATP synthase subunit d	$_{59}$ANVAK(ac)AGLVDDFEKK$_{73}$

	$_{112}$IVEYEK(ac)EM(ox)EK$_{121}$
Isocitrate dehydrogenase [NADP]	$_{46}$VAK(ac)PVVEM(ox)DGDEM(ox)TR$_{60}$
mitochondrial	$_{70}$LILPHVDIQLK(ac)YFDLGLPNR$_{89}$
	$_{81}$YFDLGLPNRDQTDDQVTIDSALATQK(ac)YSVAVK$_{112}$
	$_{173}$HAHGDQYK(ac)ATDFVADR$_{188}$
	$_{383}$GK(ac)LDGNQDLIR$_{393}$
ATP synthase subunit b	$_{155}$SQQALVQK(ac)R$_{163}$
ATP synthase subunit O	$_{58}$QNKLEQVEK(ac)ELLR$_{64}$
	$_{159}$TVLK(ac)SFLSQGQVLK$_{172}$
	$_{163}$SFLSQGQVLKLEAK(ac)$_{176}$
	$_{189}$IGEK(ac)YVDMSVK$_{199}$

Table 6 : Acetylation sites in isolated human muscle mitochondria

Adenine nucleotide translocase 1 (ANT1) contains multiple acetylation sites in human skeletal muscle mitochondria. Among the proteins deacetylated after exercise was the adenine nucleotide translocase, ANT1. ANT1 has significant control of mitochondrial respiration under physiological conditions [120], so we then focused on this protein. Using a combination of data-dependent and hypothesis-driven approaches[6], we identified a total of 5 acetylated lysine residues in ANT1. The primary amino acid sequence of ANT1, with verified acetylation sites highlighted, is shown in Figure 7. The crystal structure of ANT1, in the "c" conformation (open to the inner-mitochondrial space), complexed with the inhibitor carboxyatractyloside, is shown in Figure 8 and Figure 9. The human ANT1 homology model was prepared and inserted into a POPC membrane for simulation (Figure 10). Acetylation sites are highlighted. Amino acid sequences surrounding this acetylation site exhibit high evolutionary conservation (Figure 7).

Acetylation of mitochondrial proteins is related to insulin sensitivity. To determine the relationship between the mitochondrial acetylome and insulin resistance, total acetylation states of mitochondrial proteins (sum of all NSAF for acetylated spectra) were compared with insulin stimulation of glucose disposal during a euglycemic clamp. Characteristics of the subjects used for the analyses are given in (Table 7).

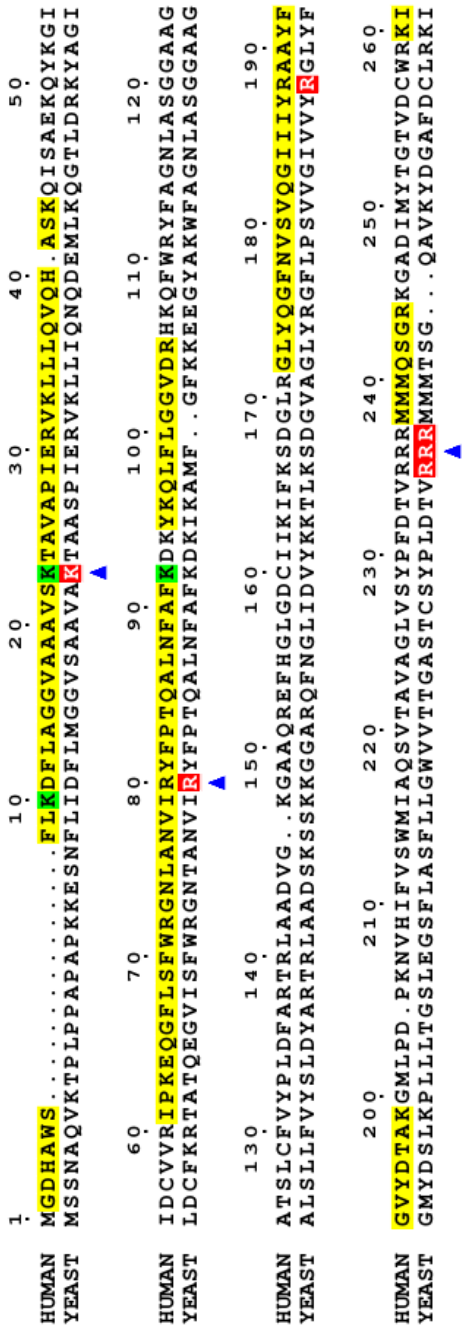


Figure 7 : Sequence alignment of human ANT1 and yeast AAC2

Yellow and green highlighting represents mass-spec coverage and observed lysine-acetylations respectively. For the yeast sequence, red highlights indicate residues required for ADP transport as demonstrated in a prior mutation study

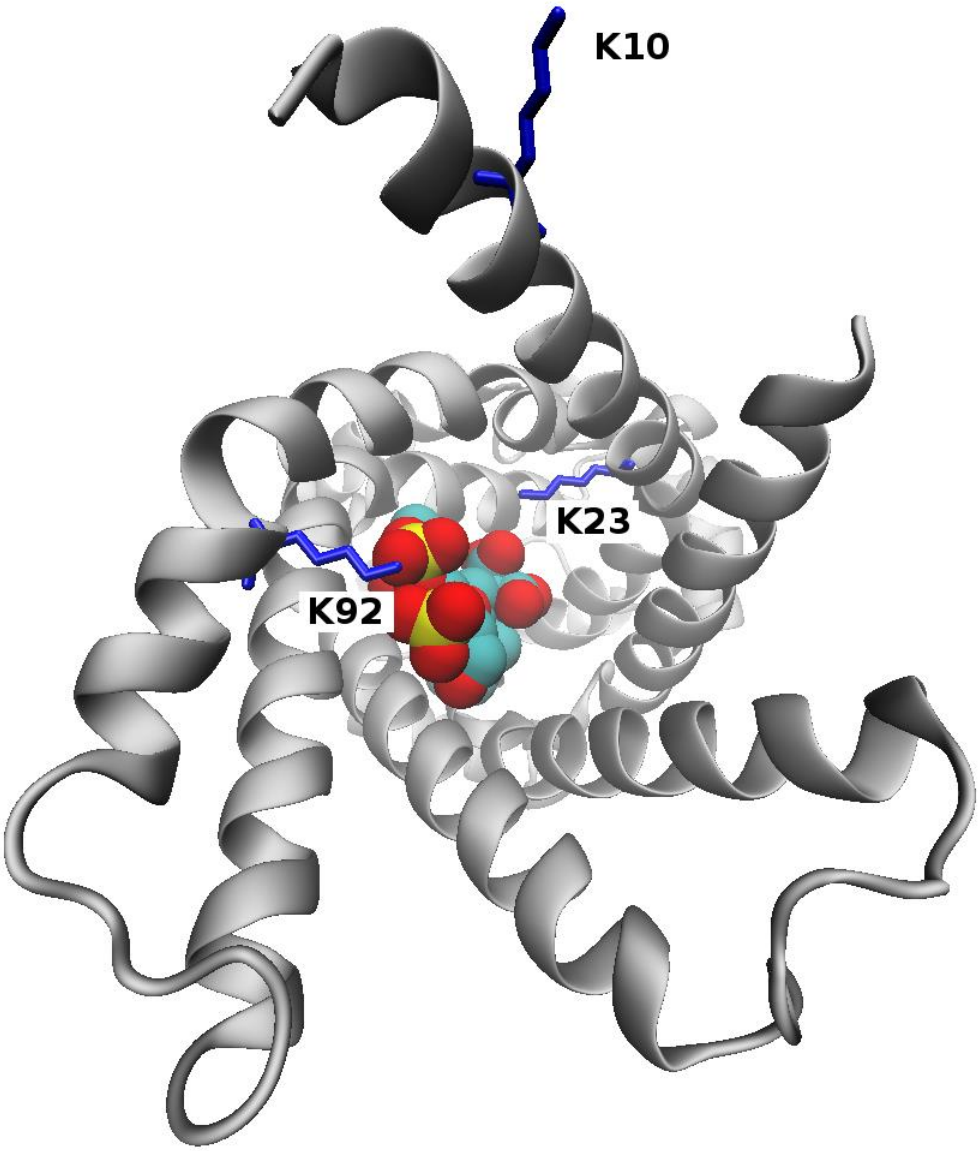


Figure 8 : Top-down view of bovine ANT1 complexed with the carboxyactractyloside inhibitor

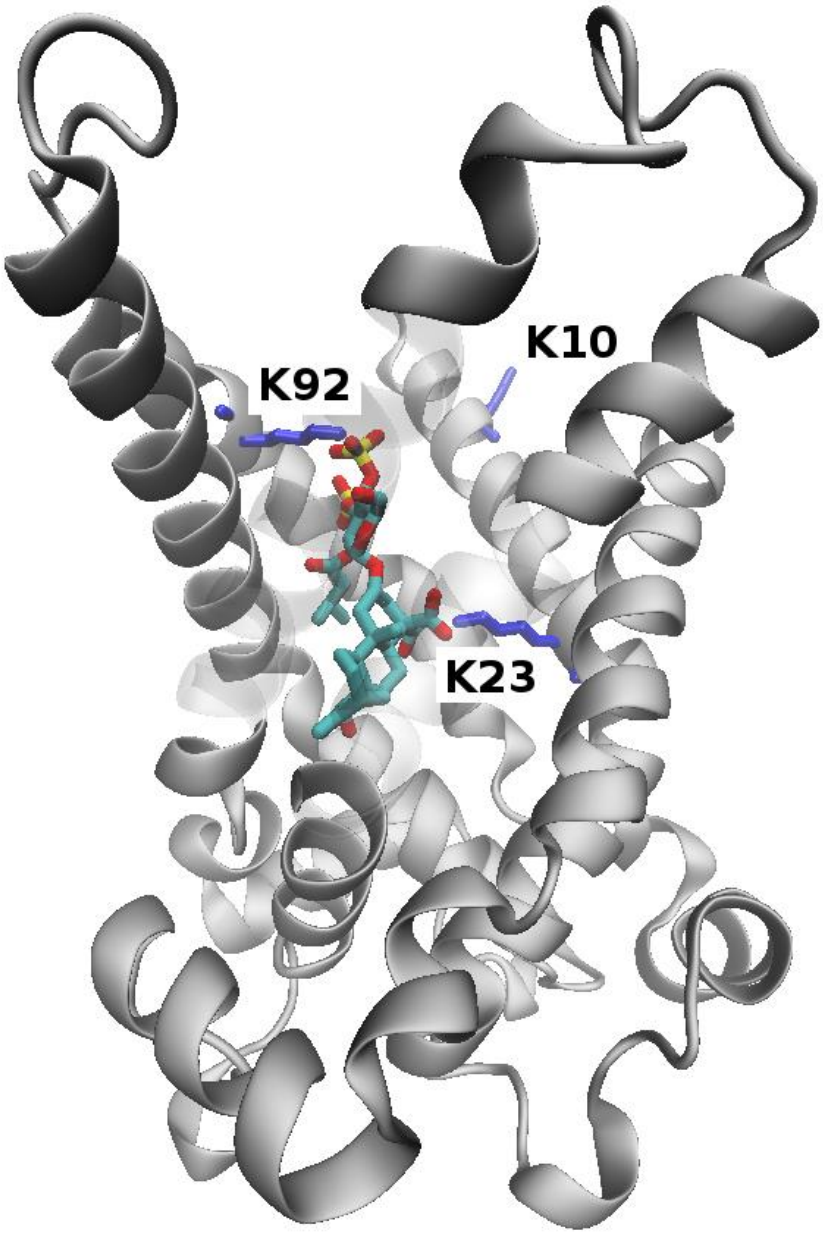


Figure 9 : Side view of bovine ANT1 complexed with the carboxyactractyloside inhibitor

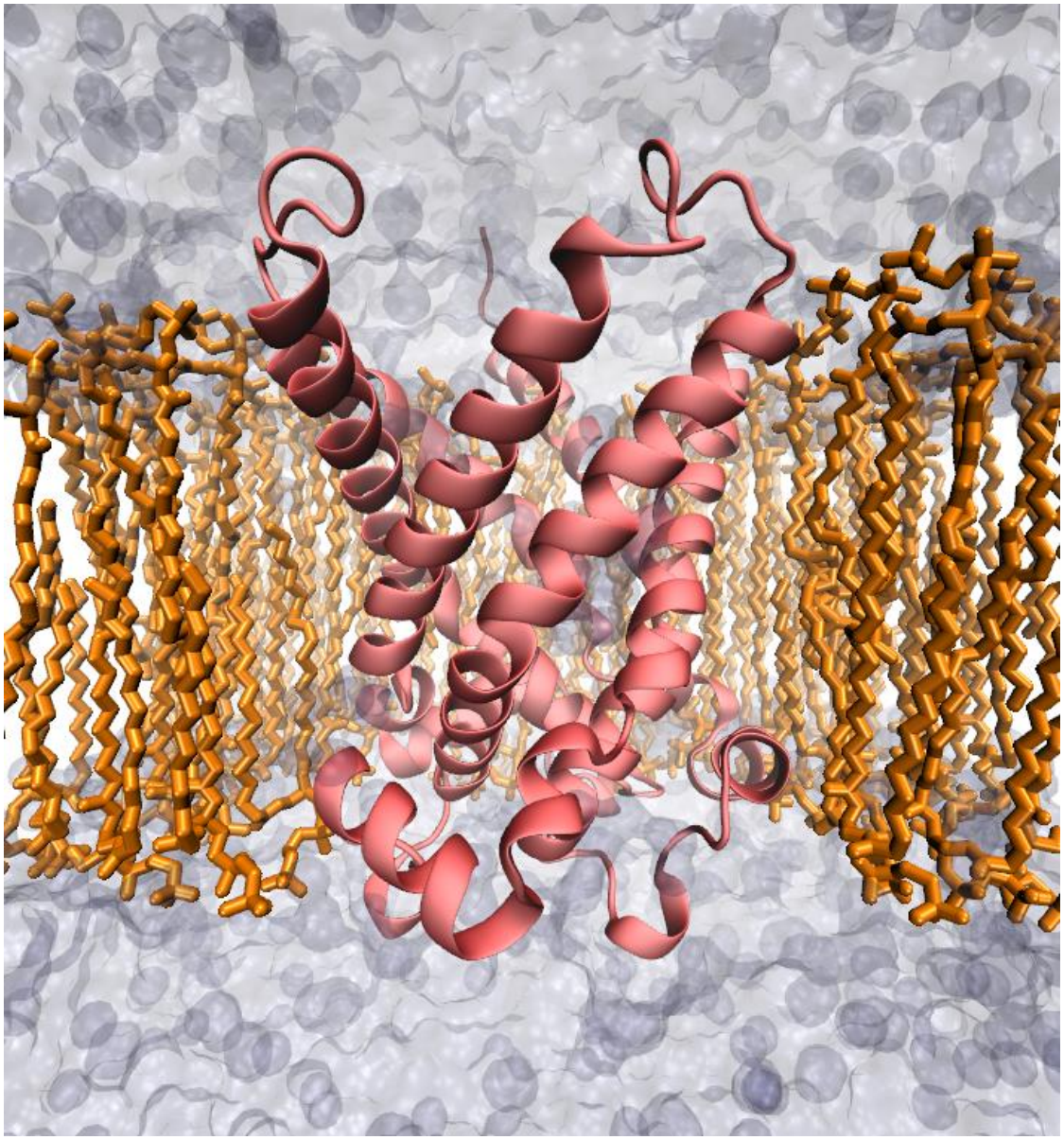


Figure 10 : Human ANT1 homology model inserted into a POPC membrane for modeling experiments.

		Mean / SEM	Range
Gender	(F/M)	8/8	
Age	years	39.9 / 3.0	24 – 60
BMI	Kg/m^2	29.1 / 1.8	19.1 – 39.1
Body Fat	percent	31.3 / 1.9	25.6 – 48.4
Basal Endogenous Glucose Production	$mg/kg^{-1}/min^{-1}$	2.07 / 0.09	1.4 – 2.5
Insulin-Stimulated Glucose Disposal	$mg/kg^{-1}/min^{-1}$	7.03 / 0.80	3.5 – 14.5
Hba1c	Percent	5.51 / 0.07	4.8 – 5.8
Total Cholesterol	mg/dl^{-1}	175 / 10	131 – 254
Plasma Triglycerides	mg/dl^{-1}	88 / 16	39 – 286
Data are given as Means □ SEM and ranges of values.			

Table 7 : Characteristics of subjects receiving only euglycemic clamps

		Mean / SEM	Range
Gender	(F/M)	2/4	
Age	years	33.5 / 4.3	23 – 48
BMI	Kg/m^2	27.8 / 1.7	22.1 – 29.9
Body Fat	percent	25.9 / 1.1	23.3 – 28.6
Insulin-Stimulated Glucose Disposal	$mg/kg^{-1}/min^{-1}$	9.3 / 1.4	6.1 – 13.2
Hba1c	percent	5.31 / 0.14	5.0 – 5.7
Total Cholesterol	mg/dl^{-1}	183 / 12	145 – 202
Plasma Triglycerides	mg/dl^{-1}	132 / 32	55 – 227
VO_{2max}	$mL/kg^{-1}/min^{-1}$	33.0 / 3.7	25.5 – 47.1
Maximum Heart Rate	BPM	187 / 4	174 – 197
Data are given as Means / SEM and ranges of values.			

Table 8 : Characteristics of subjects participating in exercise studies

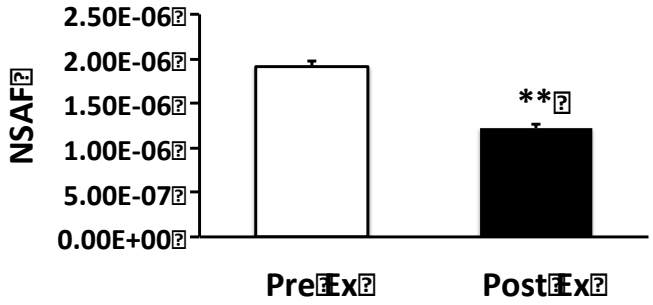


Figure 11 : Acetylation is correlated with insulin sensitivity

Acetylation is correlated with insulin-stimulated glucose disposal. Sixteen healthy volunteers underwent a two hour euglycemic, hyperinsulinemic clamp as described in the Methods. Muscle biopsies were taken at rest, after an overnight fast, approximately one hour before the insulin infusion was initiated. Mitochondria were isolated from the muscle biopsies and acetylation of mitochondrial proteins was quantified using the NSAF method. Data are given as rates of insulin stimulated glucose disposal in the last 30 minutes of the glucose clamp, expressed in units of mg glucose per minute per Kg fat-free mass. Data are displayed as Means / SEM. *P < 0.02.

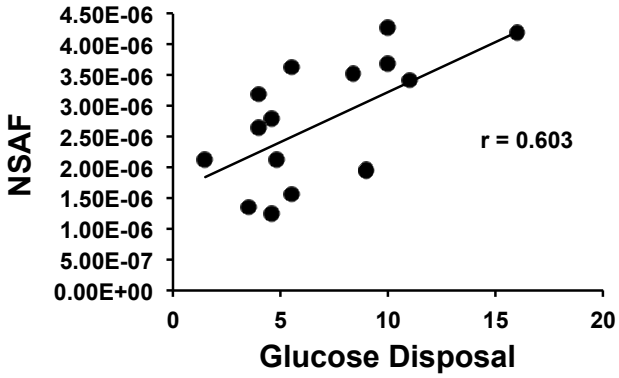


Figure 12 : Acetylation of mitochondrial proteins decreases with exercise

Acetylation of mitochondrial proteins decreases after exercise. Five healthy volunteers underwent a single bout of exercise as described in the Methods. Muscle biopsies were taken at rest, after an overnight fast, and 24 hours after the end of the exercise bout, again after an overnight fast. Mitochondria were isolated from the muscle biopsies and acetylation of mitochondrial proteins was quantified using the NSAF method. Data are shown as Means / SEM. **P < 0.001vs. pre-exercise values.

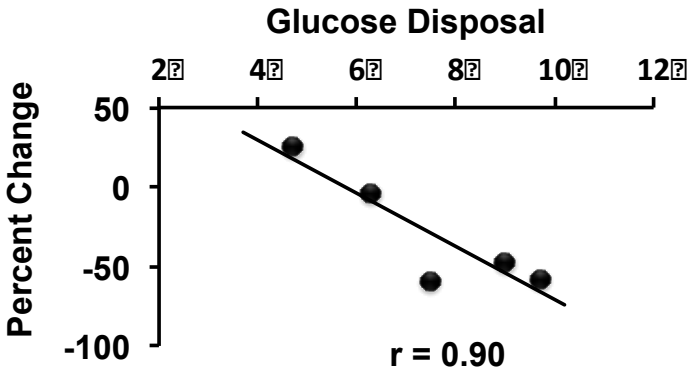


Figure 13 : Change of K23 acetylation with exercise

Ant1 acetylation at lysine 23 decreases after exercise. Five healthy volunteers underwent a single bout of exercise as described in the Methods. Muscle biopsies were taken at rest, after an overnight fast, and 24 hours after the end of the exercise bout, again after an overnight fast. Mitochondria were isolated from the muscle biopsies and acetylation of Ant1 was quantified using the NSAF method. Insulin stimulated glucose disposal was quantified during a two hour euglycemic, hyperinsulinemic clamp as described in the Methods Data are given as rates of insulin stimulated glucose disposal in the last 30 minutes of the glucose clamp, expressed in units of mg glucose per minute per Kg fat-free mass, and are shown as Means / SEM. *$P < 0.05$.

Acetylation of mitochondrial proteins is regulated by exercise. To determine whether a vigorous bout of exercise can result in deacetylation of mitochondrial proteins, five separate normoglycemic subjects (Table 8) underwent euglycemic clamps and on a separate day, vigorous bout of exercise. Overall, exercise resulted in a significant decrease in acetylation of mitochondrial proteins (Figure 11). Among the deacetylated proteins, ANT1 showed extensive deacetylation. Moreover, the extent of acetylation of mitochondrial proteins was significantly correlated with insulin action (Figure 12).

Regulation of deacetylation of specific ANT1 lysine residues after muscle contraction are related to insulin sensitivity. We reasoned that exercise/muscle contraction would increase the NAD^+/NADH ratio, activate sirtuins, and deacetylate critical mitochondrial proteins. Under resting conditions, there was no correlation between the sum of acetylation of ANT1 at all sites (total acetylation) and insulin sensitivity (r=0.42, NS). After exercise, there was a significant decrease in total ANT1 acetylation (Figure 13). When examined on a site-by-site basis, this trend was due to a significant ($P < 0.05$) decrease in acetylation of lysine 23 of ANT1, and this decrease was negatively correlated with insulin sensitivity (Figure 13).

Acetylation affects electrostatic properties of ANT1. Lysine acetylation results in a loss of positive charge. To assess the functional significance of these modifications, we analyzed the electrostatic properties of the wild-type (WT) and acetylated (K10, K23, K92) protein structures. After application of our acetylysine (KAC) patch, we produced 2D slices (Figure 14) and a 3D isosurface (Figure 15) of the electrostatic potential. ANT1 exhibited a strong electrostatic potential at the bottom of the cavity, which indeed serves

to direct the binding of negatively-charged ADP molecules. Interestingly, the observed potential is highest near the side chain of lysine 23 (Figure 16), emphasizing its importance in binding ADP. Upon acetylation, we noted only slight changes to the positions and shapes of the isopotentials, but the magnitude of the potential was reduced. To visualize the potential in a more straightforward way, we also interpolated the potentials along a line drawn through the axis of the channel. (Figure 15) As observed from the iso-electric surfaces the electrostatic potential of these 1-D plots near the binding pocket and experiences a rapid falloff with distance through the channel opening (Figure 16). However the potential at the core binding pocket was dramatically reduced with acetylation of lysine 23, which also suggests that acetylation of lysine 23 could affect ADP binding affinity.

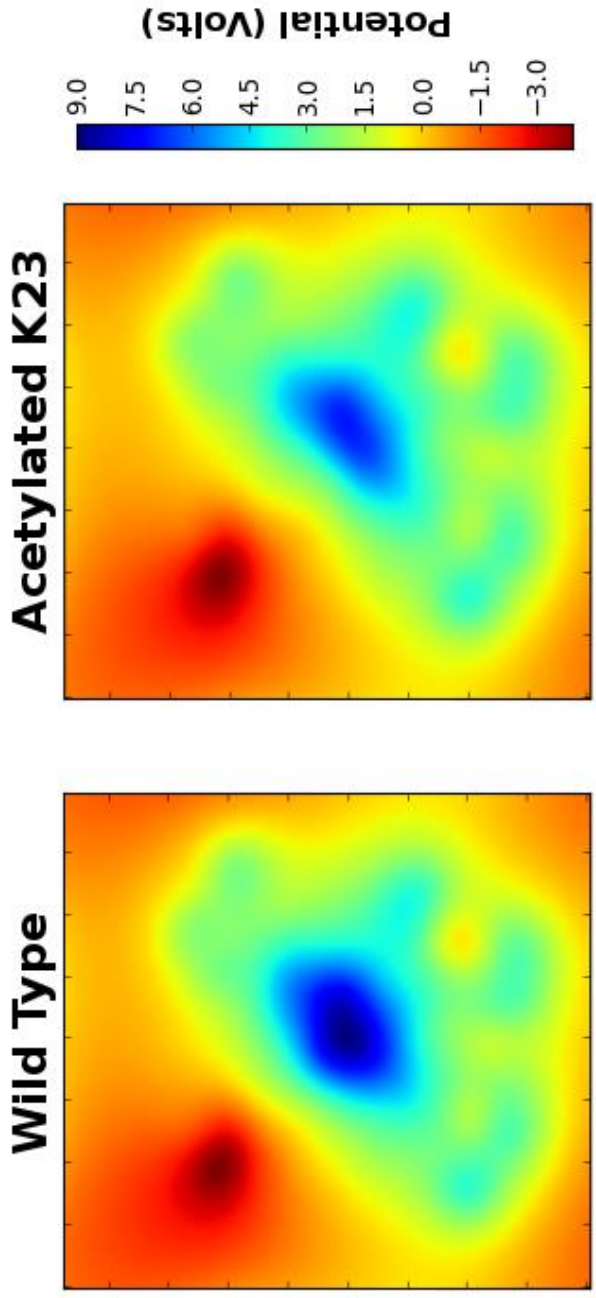


Figure 14 : 2D slices of the electrostatic potential

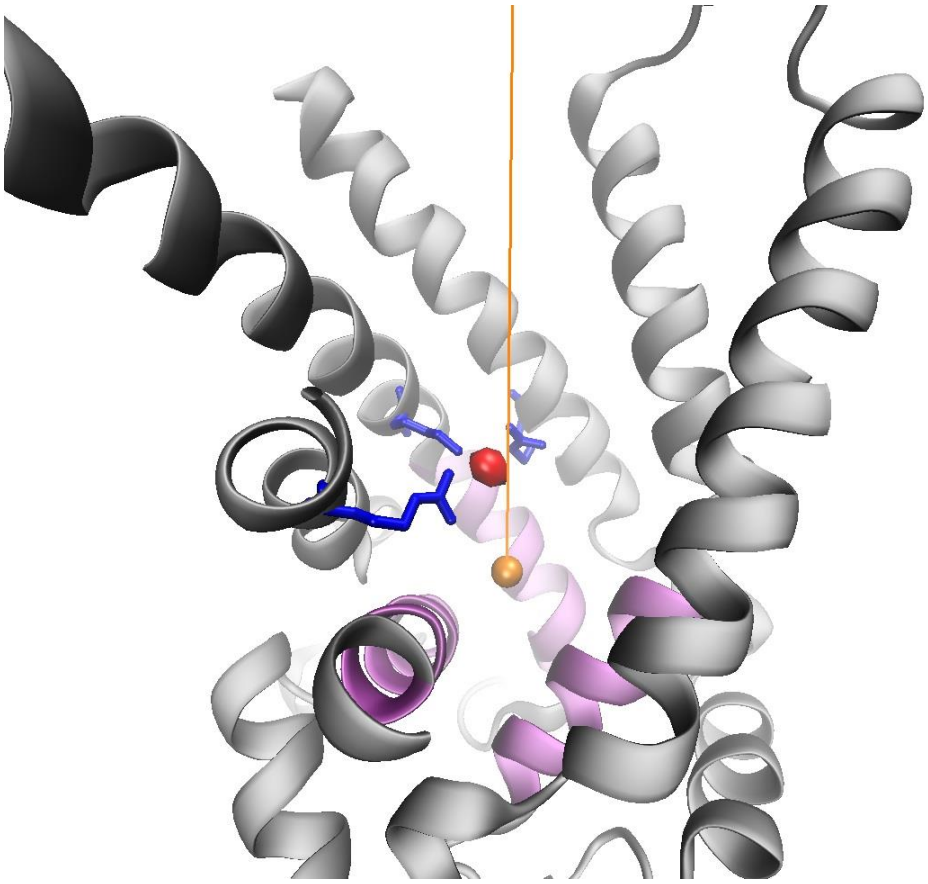


Figure 15 : Interpolation of electrostatic potential along channel

To visualize the electrostatic profile along a hypothetical reaction coordinate in which ADP would translocate, we performed interpolation of the potential along a 30 Angstrom line (orange) drawn in the vertical axis. The origin was defined as the geometric center of the proline hinge alpha carbon atoms. The three mitochondrial carrier motif gate-helices are highlighted in purple. The electrostatic potential is maximum at a site near the sidechain of K23 (red sphere).

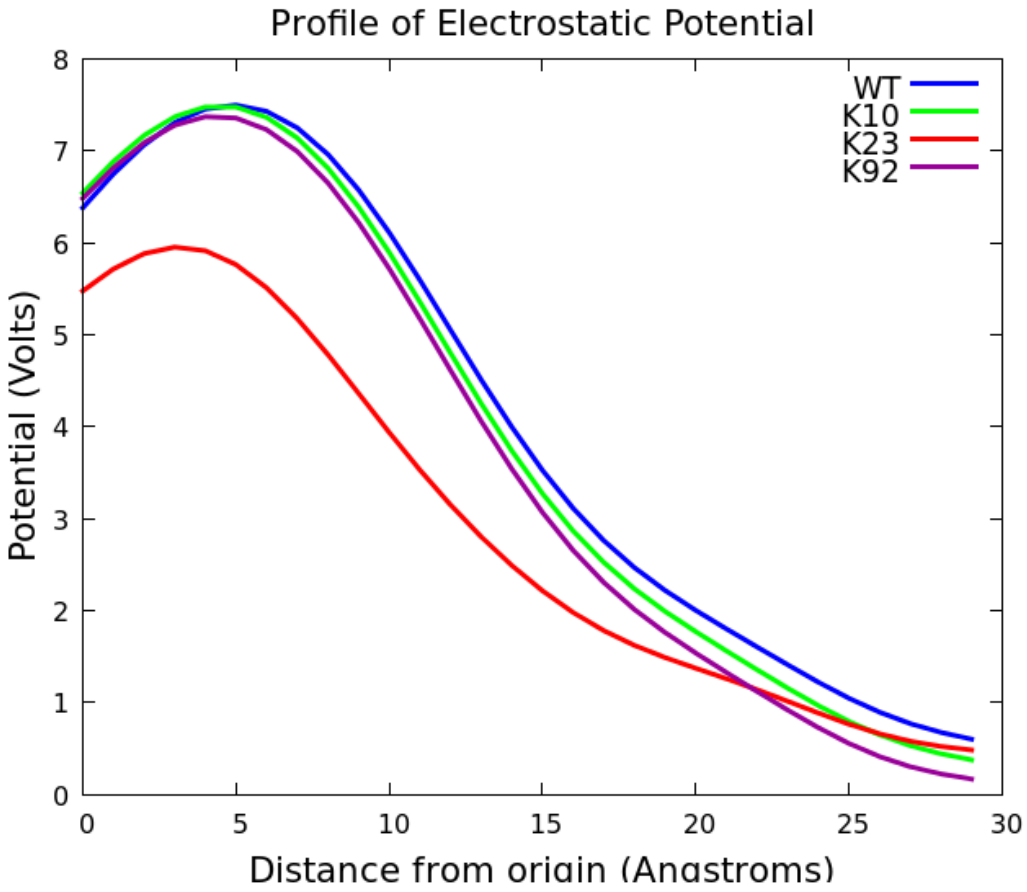


Figure 16 : Electric potential along reaction coordinate

Modeling of Apo ANT1. We prepared a wild-type (WT) and three lysine acetylated (K10, K23, K92) ANT1 systems and performed all-atom molecular dynamics (MD) simulations for 30 ns. These simulations were performed without the presence of ADP, and are thus referred to as "Apo-ANT1" simulations. To test the hypothesis that lysine acetylation may affect the structure or dynamics of Apo-ANT1, we performed root-mean-square-fluctuation (RMSF) analysis on the four simulation trajectories. The fluctuation values are plotted with respect to sequence position in Figure 17.

In each of the trajectories, there were small localized differences in fluctuations at single residues. More substantially however, a large and broad increase in fluctuations is seen in the vicinity of hydrophobic residues 76-87 localized to the midpoint of the second transmembrane helix of acetylated K23 simulation trajectory (Figure 18). Upon further investigation of the trajectory, we found that the sidechain of acetylated K23 drifts upwards away from the lower binding pocket. Rather than strong electrostatic interactions of K23 with E30 with the salt-bridge network observed in wildtype simulations, the hydrophobic part of the sidechain of acetylated K23 interacts with residues Y81 and N77 (Figure 18). These hydrophobic interactions appear to pull the second transmembrane helix slightly inward into the lumen of the channel. In the other three simulations, wild-type K23 remains localized to the binding pocket due to electrostatic interactions with E30 in the salt-bridge network.

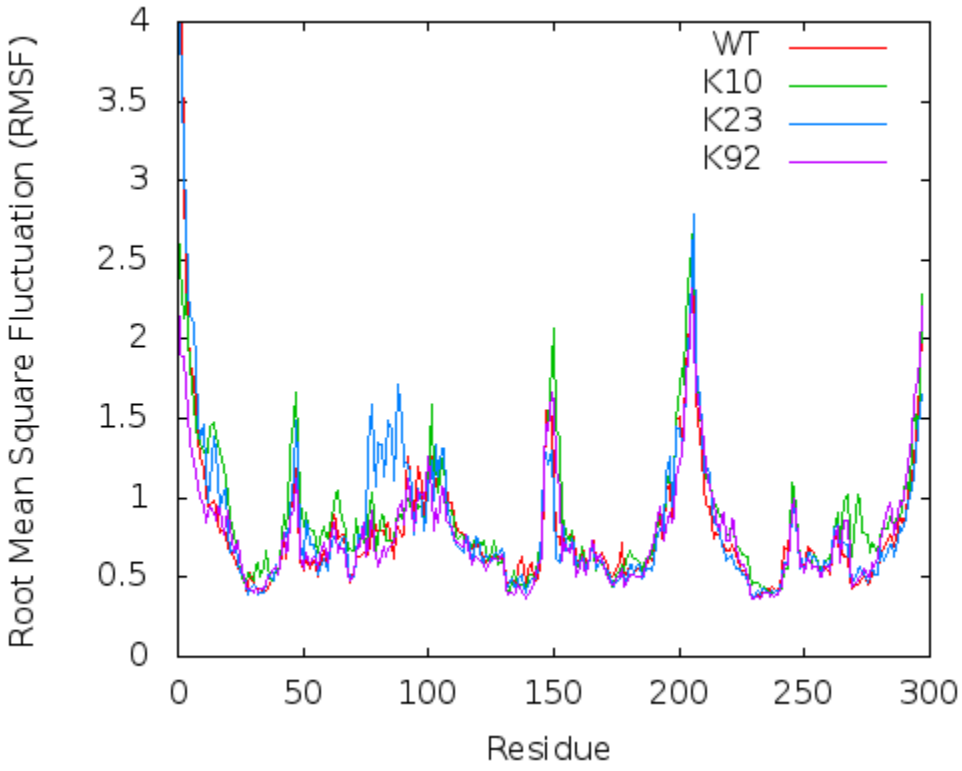


Figure 17 : RMSF profiles of the four Apo-ANT1 simulations.

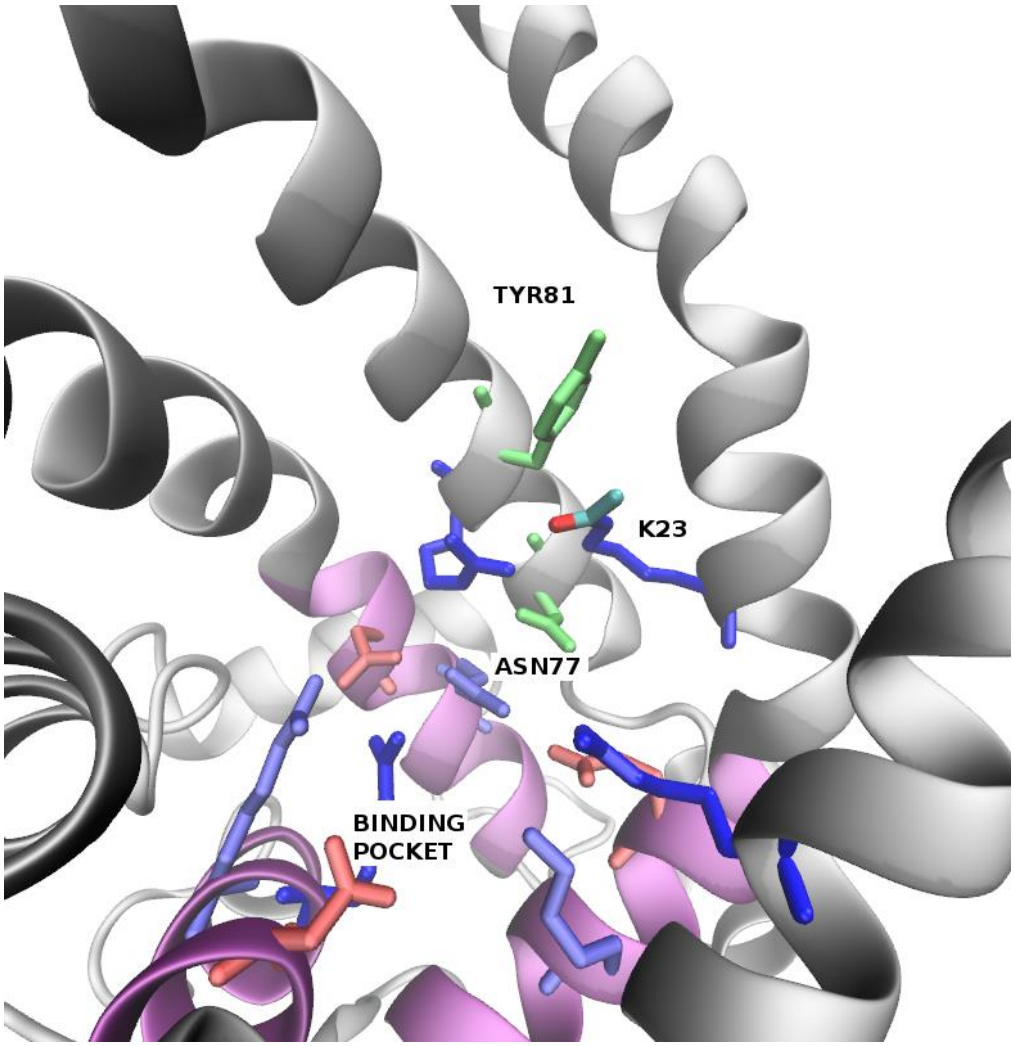


Figure 18 : Effect of K23 acetylation on ANT1 structure

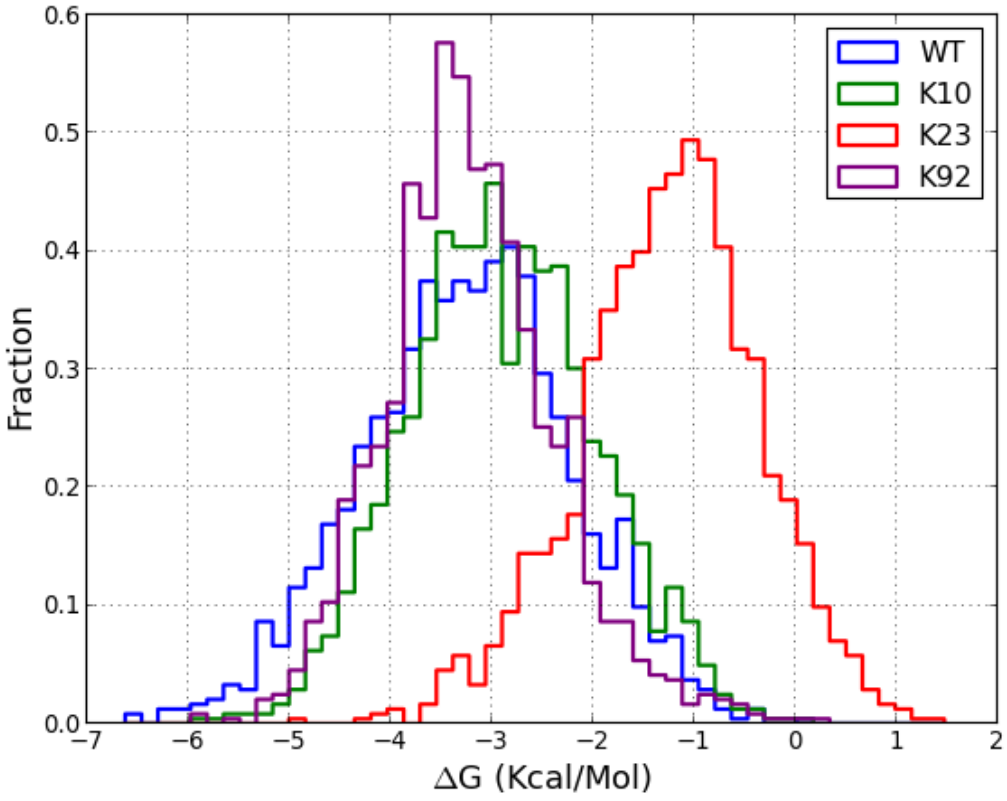


Figure 19 : Binding energy distributions obtained from ensemble docking of the four simulated systems.

To investigate the consequences of lysine acetylation on ADP binding, we used an ensemble docking approach. For each of the four Apo-ANT1 simulations, we performed independent docking runs to 1500 snapshots from the simulation trajectories using Autodock. Docking was focused to the core binding pocket of the channel, and thus docking experiments to the acetylated K10 and K92 systems served as replicates of the wild-type system. We found that ADP adopts highly similar docking poses among each of the four systems. The two phosphate groups were always arranged in a vertical orientation near the center of the pocket, preferentially interacting with four basic residues: Arginines 80, 280, 236, and Lysine 23. The adenine ring appears to not adopt any preferred orientations in the channel.

Despite no apparent differences in conformation of bound ADP, this analysis revealed substantial differences in binding affinity. We produced histograms (Figure 19) of the estimated binding energies of our ensemble docking reported by Autodock. These distributions were similar between the WT, K10, and K92 systems. However, acetylating K23 resulted in a large shift in the total binding energy of ADP. The peak of the distribution for the acetylated K23 system is located at approximately -1 kcal/mol. The other three systems, with an unmodified pocket, show binding energy peaks at approximately -3 kcal/mol. These binding energies are very weak, and would correspond to binding affinities in the millimolar range. Having arrived at these estimates, we hypothesized that docking experiments to static snapshots of Apo-ANT1 may yield weak binding affinities since the binding pocket is unable to rearrange to more favorably accommodate the ligand. To allow such rearrangements, we proceeded to conduct full molecular dynamics experiments with ADP docked to the binding pocket.

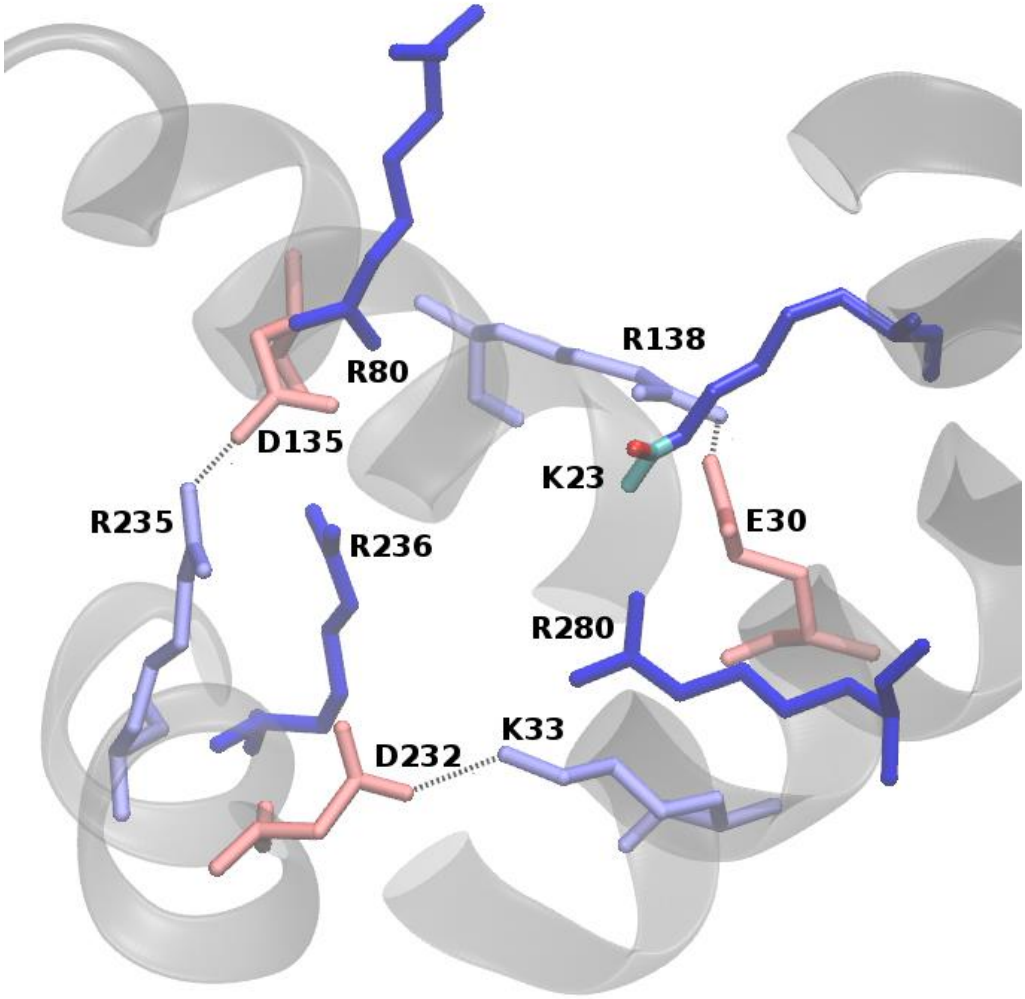


Figure 20 : Overlayed WT+K23 ANT1 systems before simulation

The overlayed binding pocket regions of both WT and Acetylated K23 systems before ADP is docked. The acetyl group atoms in the acetylated system are colored red for oxygen and teal for carbon

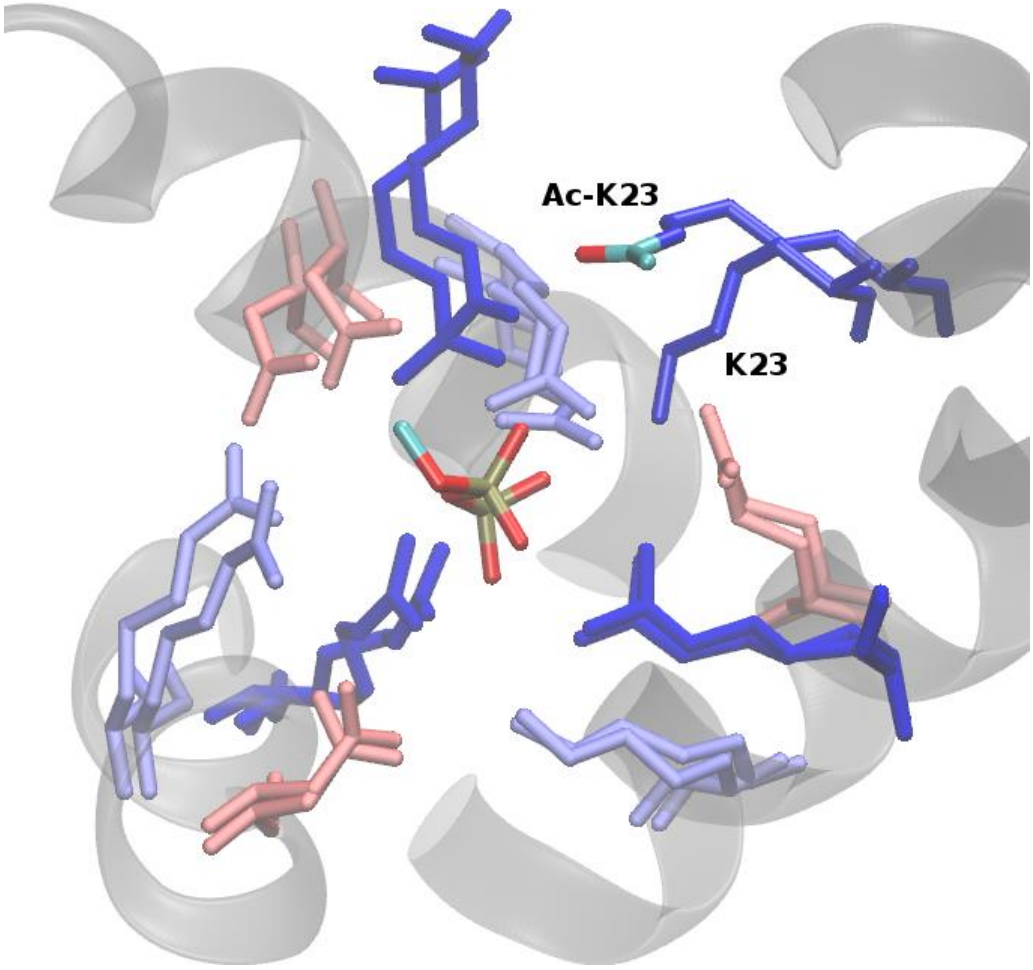


Figure 21 : Overlayed WT+K23 ANT1 systems after 30ns simulation

The overlayed binding pockets for both systems after ADP is docked at the systems are simulated for 30ns. From the overlay, we see that all charged residues have similar ending conformations except acetylated K23, which drifts away from ADP. Only the phosphates of ADP are shown for clarity. Dark blue sidechains interact the strongest with ADP and are homologues of critical yeast mutants. Lightly-colored residues form the underlying salt bridge network.

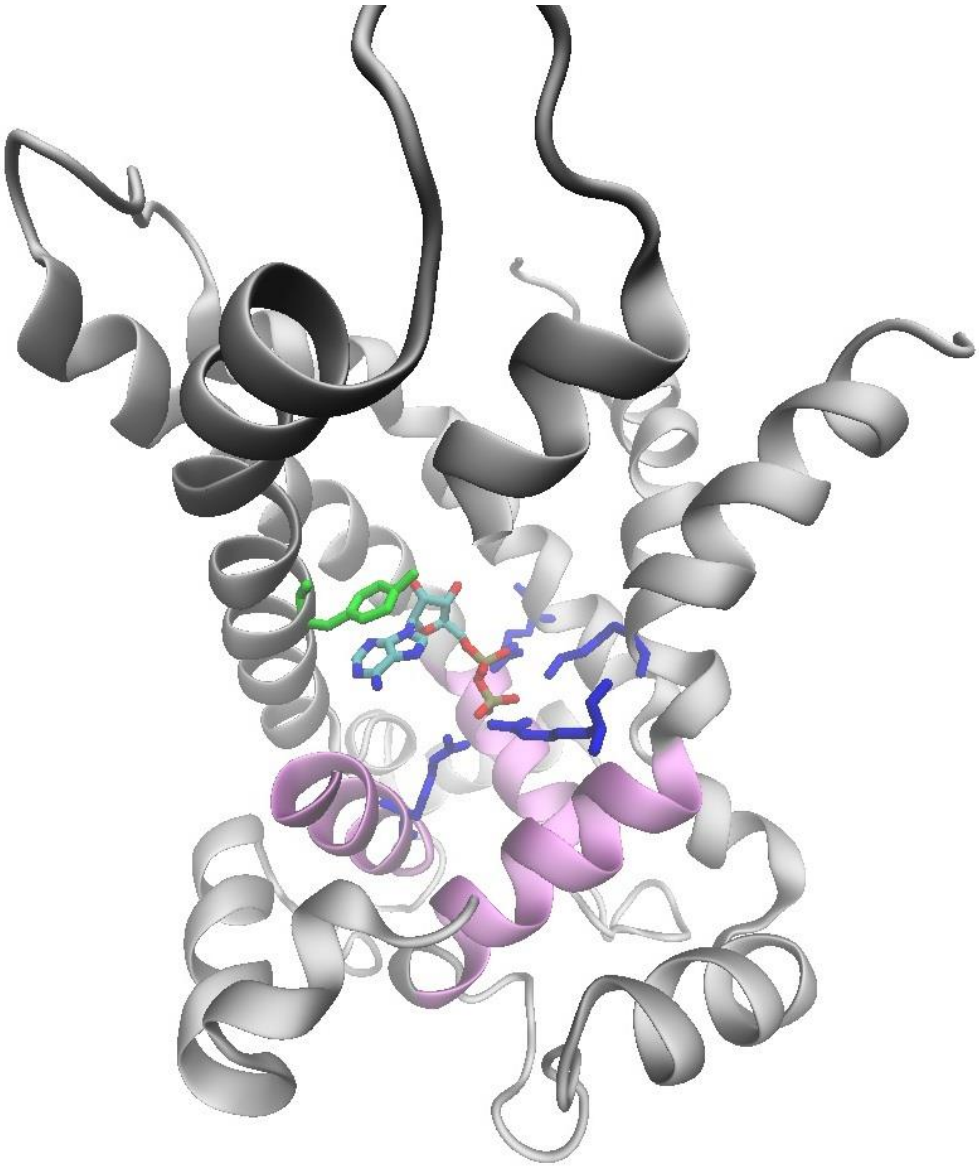


Figure 22 : Location of ADP binding pocket in ANT1. ADP is shown bound after docking and 30ns of MD simulation

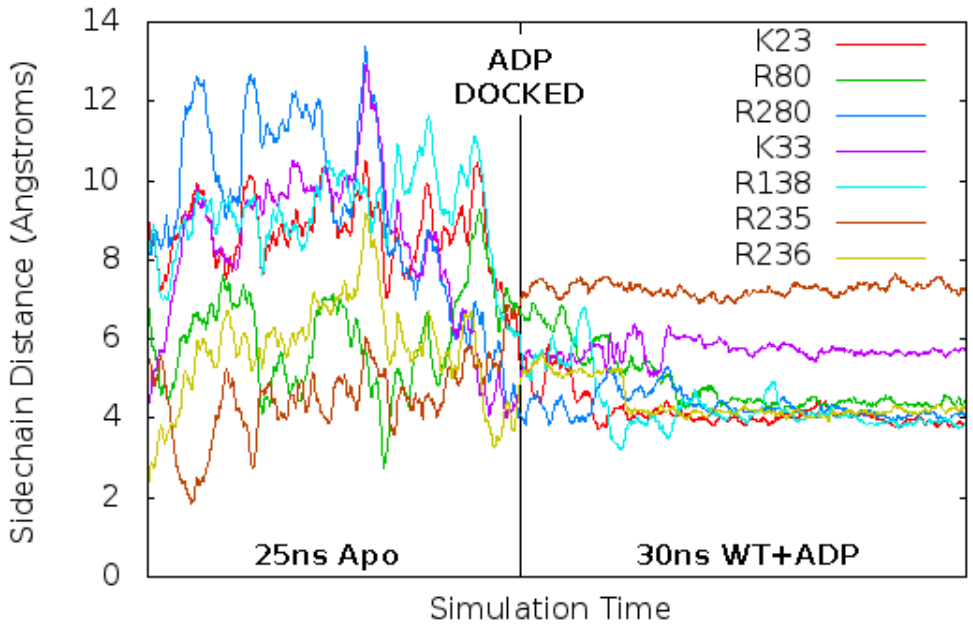


Figure 23 : Induced fit of binding pocket

We measured the radial distance of basic sidechains in the binding pocket in the Apo and Bound simulations. The dramatic stabilization after ADP insertion demonstrates an induced fit.

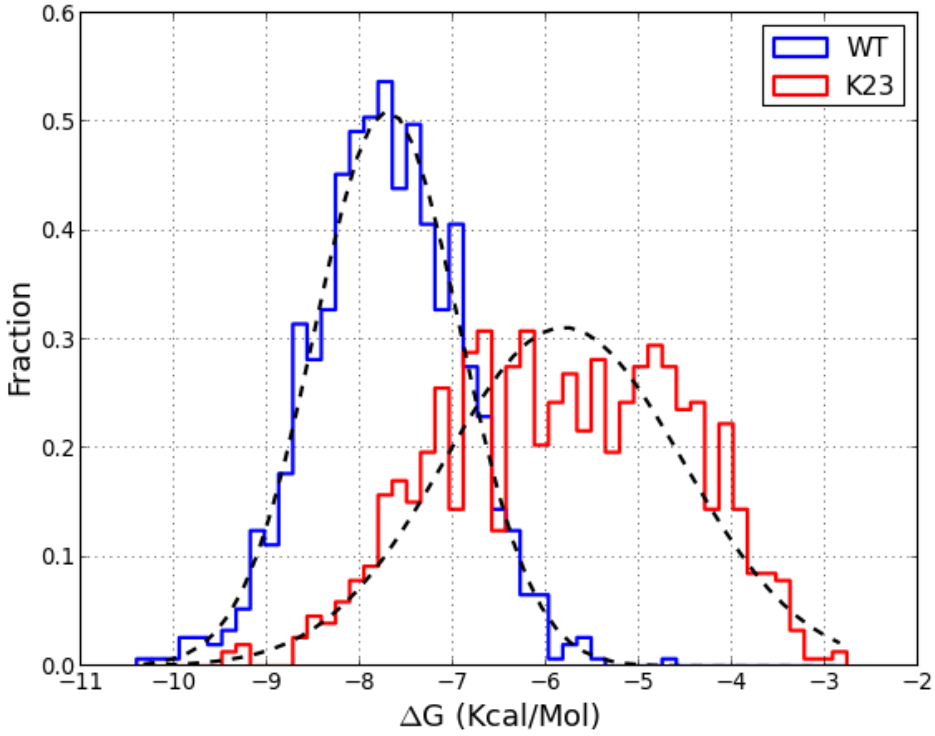


Figure 24 : Distribution of ADP binding energies

Distribution of binding energies obtained from ensemble docking experiments for the WT+ADP and K23+ADP simulations.

Modeling of ANT1 with ADP bound - We postulated that all-atom MD simulations of the ANT1 in the ADP-bound state might yield more insight into the dynamical consequences of lysine acetylation. We thus performed docking of ADP to the static equilibrated WT structure and used this system as an input for two further MD simulations. A wild-type bound (WT+ADP) and acetylated K23 bound (acetylated K23+ADP) system were prepared and simulated as before for 30 ns.

While visualizing the trajectories we observed a rapid reorganization of the charged residues within the binding pocket. Within the early stages of the wild-type simulation, R80, R280, R236, and K23 rearranged to form tight hydrogen bonds with the negatively-charged ADP phosphate groups. This observed effect suggests the concept of an induced fit model with regards to ADP binding to the ANT1 transporter. We present superimposed structures of the WT and K23 simulations with ADP bound. Both systems start in the same initial configuration (Figure 20) in which the critical residues of the binding pocket interact with charged residues at the periphery. After 30 ns of simulation however, the sidechains of these residues rearrange to a more centralized conformation to bind the ADP phosphates (Figure 21). In both systems, the final side chain conformations are highly similar for each charged residue, with the distinct exception of acetylated Lysine 23, which drifts away from the binding pocket (Figure 21). The location of the docked ADP within the WT ANT1 after 30ns of simulation is illustrated in Figure 22. The terminal phosphates of ADP interact primarily with 4 basic residues, one of which is K23.

To further characterize these structural rearrangements, we analyzed the radial positions of the conserved critical residues in the binding pocket (K23, R80, R280, K33,

R138, R235, R236) as a function of time. From the wild-type simulation, we computed the average position of the terminal phosphorus atom of ADP. Sidechain distances were then calculated from this position to the NZ atoms of K23 and K33, and the CZ atoms of R80, R138, R235, R236 and R 280 through the snapshots for both the 25 ns unbound (WT) equilibration and the 30 ns bound simulation (WT+ADP). Time evolutions of these distances which indirectly measure the interaction with ADP and these highly conserved residues critical for binding (Figure 23), shows a substantial change of binding pocket stability at the transition of the equilibration and the bound simulations. It appears that these residues drift in and out of the center binding pocket during the apo equilibration, but rapidly stabilize to a central configuration during simulations with ADP bound. Consequently, this suggests that the presence of ADP results in an induced-fit binding with the channel.

We hypothesized that this induced fit may strengthen ADP binding, and therefore we re-evaluate the binding affinities of ADP to wild type and acetylated K23 ANT in different ensemble docking results. We thus conducted an independent set of docking experiments on snapshots from the two bound ANT1+ADP simulations. We fit normal distributions to the binding energy histograms obtained from these ensemble docking experiments. (Figure 24) From the estimated mean binding energies, we observe a significant improvement of binding energies over the Apo-ANT1 docking experiments (Figure 19). ADP binding affinity of the wild-type ANT1 shows an increase from ~-3 kcal/mol in the Apo-ANT1 ensemble to -7.7 kcal/mol for the bound wild-type structure (WT+ADP). For the acetylated K23 system, ADP binds with an estimated binding energy of -5.8 kcal/mol to the bound (K23+ADP) ensemble. These stronger binding

energies further suggest that ADP binding to ANT1 is facilitated by an induced-fit of the binding pocket. These pocket rearrangements lead to a substantially stronger interaction with the ligand and may subsequently prepare the channel for opening and translocation.

From the histograms, it is clear that the ensemble docking technique has resulted in a large interval of estimated binding energies. To assess the physiological relevance of acetylation, we wish to associate these binding energies with equilibrium dissociation constants (Kd). Given the exponential dependence of K_d on binding energy, however, we must carefully consider the uncertainty in our estimates. To this end, we have compiled a table of K_d values (Table 9) calculated from the range of binding energies of 1 standard deviation to the left and right of the estimated means. From this table we see that small changes in binding energy can correspond to substantial ranges of K_d. Nonetheless, it is clear that acetylation of K23 may have significant physiological consequences on ADP binding given that concentrations of ADP in vivo are within the range of the wild-type K_d estimates.

	Wild Type	K23
Binding Energy (Mean)	-7.7 kcal/mol	-5.816 kcal/mol
K_d (of mean energy)	3.66 uM	79.39 uM
Binding Energy (Std Dev)	0.78 kcal/mol	1.28 kcal/mol
Binding Energy Range (+- 1 StdDev)	-8.49 to -6.928 kcal/mol	-7.09 to -4.53 kcal/mol
Corresponding K_d Range	1.02 to 13.042 uM	9.94 uM to 634 uM

Table 9: Estimated binding energies of ADP docked with either WT and K23-acetylated ANT1.

Means of binding energies were obtained through fitting the normal distribution to histograms of binding energies from ensemble docking experiments. The estimated standard deviations were used to consider a range of binding energies of 1 sigma to either side of the mean. Dissociation constants K_d were calculated from these estimated mean binding energies.

DISCUSSION

Although a significant body of evidence exists in favor of the concept that mitochondrial dysfunction is associated with insulin resistant skeletal muscle [121], this finding is not universal [122]. Some investigators report subtle, but potentially important, changes in mitochondrial function [123]. Given the wide variety of methods and measurements used by many investigators, and because many measurements reflect overall pathway activity rather than individual steps, it may not be surprising that there are disparate results. The present study was undertaken to understand how one potential regulator of mitochondrial proteins, acetylation, might affect the function of mitochondria, particularly with respect to insulin sensitivity.

We used mass spectrometry-based proteomics to identify a number of lysine acetylation sites in skeletal muscle, including mitochondrial and other proteins. Because we did not use any enrichment methods, these proteins are likely to represent the most abundantly acetylated proteins in human skeletal muscle and a minimum acetylation. To our knowledge, this is the first reported acetylome of human skeletal muscle. There undoubtedly are many more, lower abundance acetylation sites that can be revealed using enrichment methods. Nevertheless, many of the proteins we observed to be acetylated are involved in metabolism, which is consistent with the idea that acetylation is a widespread regulator of metabolism[9].

Using a label-free mass spectrometry quantification method that we previously validated [4], we observed that the overall extent of acetylation of mitochondrial proteins was positively correlated with insulin-stimulated glucose disposal during a euglycemic

clamp, which quantifies insulin sensitivity in skeletal muscle [124]. We also showed that exercise is associated with an overall decrease in acetylation of mitochondrial proteins. In particular, acetylation of lysine 23 in the adenine nucleotide translocase (ANT1), gene name SLC25A4 decreased after exercise in a manner dependent on insulin sensitivity, that is, to a greater degree in insulin sensitive individuals. Therefore, it appears that in general insulin sensitivity is associated with a greater range of protein acetylation, that is, higher values under basal conditions and greater physiological changes. This phenomenon is reminiscent of the concept of "metabolic inflexibility" of insulin resistant skeletal muscle [125], and may provide a clue to understanding the underlying molecular mechanism.

Under physiological conditions, ANT1 displays significant control strength over the rate of mitochondrial respiration [120]. Because abnormalities in mitochondrial function have been linked to insulin resistance in muscle [123], we sought to understand how acetylation of ANT1 might alter its function. To address this, we chose a molecular modeling approach, since the crystal structure of the highly homologous bovine ANT1 is known [110]. Abnormalities in the structure of ADP/ATP carriers have long been investigated in metabolic defects and disease. In a recent work both biochemical assays and all-atom MD simulations were used to characterize six pathological mutants observed in humans [126]. Several of these mutations have been linked to fatigue, weakness, and exercise intolerance in individuals, as well as extreme ptosis [126]. These associations demonstrate a clear point of control at the ADP/ATP carrier for proper metabolism in muscle.

The functional importance of charged residues in ANT1 has already been demonstrated in mutation studies. Nelson and colleagues performed a screen for site-directed mutants of the yeast AAC2 transporter, a close homologue of human ANT1 [10]. By growing yeast on a non-fermentable carbon source such as glycerol, they were able to screen for mutants of AAC2 that affected ADP/ATP transport. They demonstrate that six arginines and one lysine residue are absolutely required for growth. Sequence alignment of the yeast AAC2 carrier with human ANT1 (Figure 7) demonstrates that the two proteins are highly homologous. Of the six arginines identified as critical for transport, three are part of the RRRMMM motif, which is a characteristic feature of all ADP/ATP transporters [127]. The second arginine in this motif is R236, which interacts closely with the ADP phosphate groups in our simulations and docking. Another two mutant arginines in yeast correspond to R80 and R280 in the human transporter, which bind ADP and flank K23 on both sides. Finally, the single most critical lysine in yeast, K38, is homologous to K23 in human ANT1. As our mass spectrometry results indicate that K23 is acetylated in humans, regulated physiologically, and related to insulin resistance, we feel there is strong evidence that channel transport and metabolic activity may be tightly controlled by this critical modification.

Given the strong evidence in the literature that ANT1 is a central point of control for oxidative metabolism, and that defects of ANT1 at specific sites can cause profound dysfunction, we focused our efforts to characterize the biophysical mechanisms that ANT1 transport might be inhibited. Acetylation of critical residues could affect the electrostatic properties of ANT1 and thereby inhibit either the initial capture or tight binding of ADP. Acetylation could also perturb the structure or dynamics of ANT1,

possibly changing the shape of its binding pocket or altering molecular dynamics that may facilitate ligand recognition or channel opening. We thus explored all of these possibilities in our modeling efforts.

To capture and transport ADP, ANT1 relies upon a substantial +18 charge to attract negatively-charged ADP molecules into the mitochondrial matrix. In prior work, Wang et. al.[128] demonstrated using bioinformatics analysis that this large charge is extraordinary in comparison to most of the mitochondrial proteome. It is thus critical to understand the effects that lysine acetylation has on the electrostatic properties of the channel. Of the three lysines addressed in our study, K23 is uniquely positioned at the binding pocket where the electrostatic potential is maximum (Figure 15). Furthermore, a dramatic loss of electrostatic potential is observed when this single lysine is acetylated. (Figure 16) We believe that this electrostatic loss will profoundly affect ADP capture in vivo.

We additionally hypothesized that lysine acetylation may perturb either the structure or dynamics of ANT1 and thus affect its transporter activity. After the initial ADP capture and binding to the lower pocket, ANT1 is thought to undergo a conformational change that leads to channel opening and passage of ADP into the mitochondrial matrix. ANT1 is a member of the mitochondrial carrier family (MCF) of proteins, all of which are characterized by three "mitochondrial carrier motifs" PX(D/E)XX(K/R). The three motifs co-localize in the structure of bovine ANT1 (PDB: 1OKC) and form a triple-helix “gate” at the matrix side of the channel. The proline residues form kinks in the alpha-helices, presumably allowing a hinge motion to occur during opening and closing of the channel. The two charged (D/E) and (K/R) residues of

each helix form salt-bridges that bind the neighboring helices together in a closed conformation. Channel opening has long been thought to require the coordinated disruption of this salt-bridge network, presumably by the binding of ADP to the three basic residues located on these helices. ANT1 has long been postulated to undergo a conformational switch between a cytosol-facing C state and a matrix-facing M state, in which either ADP or ATP is bound and translocated in alternating order. We therefore postulated that a perturbation of these dynamics via lysine acetylation may influence the function of the channel.

We conducted fluctuation analysis to determine if the gate helices would be perturbed by acetylation but found no such changes in the Apo simulations. We instead, however, discovered a large structural difference in the channel brought on specifically by acetylation of lysine 23. The second transmembrane helix from residues 76-87 is dramatically perturbed due to specific interactions with acetylated K23. With the charge of K23 completely neutralized by acetylation, hydrophobic interactions dominate and induce local packing with the hydrophobic residues on the sidewall of the channel. These results suggest that channel narrowing may be a structural consequence of K23 acetylation that may conceivably effect ADP transport.

Alongside our investigations into the Apo-ANT1 structure, we sought more direct computational evidence that capture or binding of ADP would be reduced. Two prominent computational studies [128, 129] have simulated the early capture and translocation events of ADP using all-atom MD techniques with very long simulations. Both studies focused on the strong electrostatic properties of ANT1 acting as a "funnel" to capture ADP molecules from the inner-membrane space and transport them to the

lower binding pocket. These simulations required a diverse range of translocation times due to the stochastic nature of transport. In these simulations, ADP became trapped in various locations and orientations for extended durations. We felt that these stochastic and chaotic processes precluded the use of all-atom MD to simulate capture events for purposes of comparing the effects of lysine acetylation.

We thus focused on the docking of ADP directly into the lower binding pocket where K23 is located. In a recent work [130], Autodock was successfully employed to perform a screen of possible ATP binding sites on the closed matrix-facing surface of the ANT1 protein, and so we adopted similar methodologies for ADP binding in the cytosol-facing pocket. Our experiments resulted in highly-reproducible bound conformations of ADP that agrees with previous all-atom MD results [128]. Furthermore, the Autodock software provided us with estimates of binding energies that give us perspective of how a single acetylation event can have a significant effect at such a critical core residue in the pocket.

Ensemble docking to our Apo-ANT1 molecular dynamics results was used to successfully assess the binding energies in four variants of the ANT1 channel. These results suggested there could be a dramatic inhibition of ADP binding that resulting from a single acetylation at K23. We then extended the docking technique to produce an ANT1+ADP complex for further all-atom simulations. These simulations and subsequent ensemble docking experiments further elucidated the structural rearrangements associated with ADP binding and provided even more compelling evidence that lysine 23 is critical for proper function.

In our docking experiments and our all-atom MD simulations, we specifically observed R80, R280, R236 and K23 to be the primary binding partners of the ADP ligand. These critical residues account for the majority of the critical mutation sites previously identified [10] in yeast (Figure 7). Over longer timeframes, we also observed R138 and R235 moving towards the ligand. It is apparent from our simulations that the strong negative electrostatic field of the ADP ligand is beginning to disrupt two of the salt-bridges that keep the channel closed. Without the presence of ADP, the basic side chains are highly mobile and have highly varying distances from the channel center. In the bound ANT1 simulations (ANT1+ADP) however, the basic residues quickly rearrange to a stable minimal distance from the average phosphate position. Intriguingly, two of the three basic salt-bridge residues show a strong rearrangement. This is direct evidence that the salt-bridge network holding the channel closed is indeed perturbed by the binding of ADP.

Other studies have found acetylation of mitochondrial proteins to regulate function. For example, fasting alters acetylation of mitochondrial proteins, suggesting reversible acetylation may regulate function [104]. Sirtuins are important protein deacetylases that depend on NAD^+ for activity. Sirt3, 4, and 5 are present in mitochondria and may regulate aspects of mitochondrial function [131-133]. For example, fasting induces Sirt3 expression in mouse liver, elevating fatty acid oxidation, and deletion of Sirt3 gene results in higher concentrations of acylcarnitines and triglycerides, due to defective fatty acid oxidation (29). Deacetylation of LCAD by Sirt3 is critical for fatty acid oxidation [103]. Sirt3 null mice also have reduced ATP production in liver and are cold-intolerant [106], and exhibit increased acetylation of

mitochondrial proteins critical for energy homeostasis. Sirt3 expression is increased by exercise and fasting in mouse muscle, where it is highly expressed in oxidative muscle [80].

The results of this study provide important new information about the role of reversible acetylation in mitochondrial function. Our data provide strong evidence that acetylation of lysine 23, through its electrostatic effects, profoundly reduces the affinity of ADP binding to ANT1. ADP dissociation constants (K_d; the concentration of ADP ligand required for half-occupancy of the ANT1 channel) calculated from the mean binding energies from ensemble docking experiments give insight into the performance of WT and acetylated ANT1 *in vivo*.

The estimated K_d values were 3.66 □M in wild type ANT1, but rose to 79.39 □M with the acetylation of K23. These K_d values would correspond roughly to K_MADP for respiration values if 1) ANT1 was the single locus of mitochondrial respiratory flux control and 2) ANT1 flux and cytosolic [ADP] conformed to a simple Michaelis-Menten (M-M) relationship. In fact, in mixed human skeletal muscle contracting in the mild-to-moderate aerobic metabolic range, ANT1 is the major locus of aerobic flux control [120] and ^{31}P-nmr non-invasive measurement of energy phosphate concentrations have shown that simple M-M kinetics can, under such conditions, provide a reasonable fit of the [ADP]:VO_2 relationship [134], however higher order models have been demonstrated as well [135]. Thus, it is of interest to compare these K_d values to free [ADP] and apparent K_MADP values measured *in vivo*. These studies demonstrate that *in vivo*, resting human skeletal muscle free [ADP] is in the 10-20 □M range and the apparent K_MADP has been reported to be roughly 30-50 □M. These values reported in the literature fall between our

estimates of the binding affinities of the WT and K23 acetylated systems. Rearrangement of the Michaelis-Menten equation yields an estimate of the fractional velocity of ANT1:

$$v/V_{max} = [ADP]/([ADP] + K_MADP)$$

This calculation predicts, for example, that a resting [ADP] of 10 □M, along with a K_MADP of 3.66 □M, yields an ANT1 fractional velocity of 0.73, while the same estimate with K_MADP of 79.39 □M yields 0.11. Since human skeletal muscle rests at a VO_2 roughly 1 or 2% of its aerobic capacity [125, 136, 137], it seems the former, wild type scenario greatly overestimates mitochondrial ATP output, while the acetylated condition yields a reasonable match.

Important caveats obviously attend such hypothetical scenarios. For one example, the scenario assumes ANT1 V_{max} equals the maximum rate of mitochondrial ATP production, when it may be threefold higher, based on isolated mitochondria in which maximal respiration is titrated out by carboxyatractyloside (CAT) titration, an ANT1 inhibitor that binds with very high, nearly 1:1, affinity [138]. But the analysis nevertheless affords useful insight into the role of ANT1 in the control of mitochondrial respiration. The observation that the in vivo apparent K_MADP for respiration is intermediate to the widely disparate K_d values of wild type vs. acetylated ANT1 modeled here may reflect a mechanism that adjusts mitochondrial sensitivity to the ADP respiratory signal in a manner geared to metabolic demand. In the context of basic biology, such a mechanism might partly explain the observation that the kinetics by which ADP controls respiration appear to be higher order, rather than simple Michaelis-Menten [135].

In summary, acetylation of mitochondrial proteins, and in particular the adenine nucleotide translocase ANT1, is abundant in human skeletal muscle mitochondria, is related to insulin sensitivity, and physiologically regulated by an acute exercise bout in an insulin sensitive manner, especially at lysine 23. The use of simulations and docking experiments revealed that lysine 23 is a critical residue for controlling the affinity of ADP binding to ANT1, with the acetylated system having a much lower binding affinity for ADP. Therefore, the brisk deacetylation of K23 by exercise in insulin sensitive individuals could have important implications in regulating mitochondrial function.

Software Credits

- Sequence alignment in Figure 7 was drawn with ESPript[139]

ACKNOWLEDGEMENTS

The authors wish to thank the participants who volunteered for this study and gratefully acknowledge the expert editorial assistance of Irene Beauvais. The study was supported by NIH grants R01 DK047936 (LJM) and DK066483 (LJM), and an allocation of supercomputer time from the ASU Advanced Computing Center. The authors also thank Dianne DeNardo, RN, for expert nursing assistance and Mark Mattern, MD, for medical oversight of the euglycemic clamp experiments.

Chapter 4

AMASS: A NEW DATABASE FOR INVESTIGATING PROTEIN MODIFICATIONS

Clinton Mielke, Lawrence J. Mandarino, Valentin Dinu

Abstract:

Motivation: Bioinformatics is traditionally sequence-driven, however new technologies have created an explosion of protein structural data. Sequence annotations must be unified to these structures to enable visualization and understanding of protein function and molecular abnormalities potentially involved in disease.

Results: We have created the AMASS Database, which seeks to bridge the gap between 1D sequence annotation databases and 3D protein structures. We additionally provide a service whereby users can query arbitrary sites in protein sequences to find possible functions. The system is built upon the premise that active sites, coding mutations, and post-translational modifications within proteins may co-localize and share common functional mechanisms.

Availability and Implementation: AMASS DB is implemented with Biopython and Apache as a freely-available web server at www.amass-db.org.

Contact: For inquiries or technical support, please contact amass.db@gmail.com

Introduction:

Due to the rapid evolution of genome sequencing techniques, the understanding of the molecular basis of life has experienced a great revolution. Our ability to read the "source code" of life has produced large databases of biological sequences, and the bioinformatics field has emerged to find patterns in the code. Bioinformatics has originated as a sequence focused enterprise, and these techniques have yielded much insight into how genes and proteins work. Sequence conservation, at the nucleotide or residue level, provides tantalizing hints into where the underlying molecular mechanisms of a gene or protein reside. Thus, just from sequence data, the field has deciphered many biological puzzles. The result is a growing collection of functional site annotations in genes and proteins. Active sites in enzymes have been elucidated, conserved binding motifs have been found, and specific cancer mutations have been detected that are abundant in tumors.

Complementing the advances in genomic sequencing, new technologies have ushered in the era of proteomics. Mass spectrometry has enabled the routine and rapid "sequencing" of thousands of proteins found within a given biological sample. This technique enables the high-throughput discovery of protein post-translational modifications (PTMs) such as phosphorylation and acetylation, which can dramatically alter biological function in ways not described by the genome sequence. These post-translational modifications provide another layer of complexity that describes molecular mechanisms at single sites within protein sequences. These residue-specific annotations will only continue to grow in a variety of bioinformatics databases.

Structural proteomics has also experienced rapid growth with projects such as the Protein Structure Initiative[140] aiming to determine the atomic structure of every human protein. Bioinformatics must adapt to this change, and transition from its one-dimensional (1D) sequence-oriented origins to a unified three-dimensional (3D) sequence/structure level of understanding. Connecting sequence annotations to protein structures may enable the discovery of co-localized functional sites of interest, and may allow investigators to find novel mechanisms of regulation.

As the number of crystallized proteins has increased, multiple projects and web portals have been designed and published that offer various analysis and visualization capabilities. Most of these projects focus on the analysis of single-nucleotide-polymorphisms (SNPs) that are studied in genome wide association studies. The primary objective of most of these portals is to offer automated predictions of functional effects of SNPs based on either sequence analysis or predicted biophysical changes to protein structure. Some of these resources permit interactive visualization of mutated sites, but with limited support for exploration of other annotated functional sites. Furthermore, most of these resources only permit query of single sites, or do not permit any kind of queries as they index fixed SNP databases such as dbSNP.

SNPs3D[141] provides a pipeline that performs analysis of SNPs within protein structures, and assesses the biophysical consequences of SNPs based on changes such as surface accessibility and electrostatic interactions. Upon these predicted effects, a support vector machine classifier is used to predict whether or not a given SNP will have deleterious consequences. The location of each SNP can be visualized using Jmol on separate pages within the web interface. The SNPeffect[142] website offers a similar

analysis pipeline for predicting the effects of SNPs, and focuses on binding/aggregation predictions and backbone stability analysis.

The PolyPhen[143] server offers a sophisticated analysis service wherein users can query arbitrary SNPs of a given protein. The site uses eight sequence-based and 3 structure-based algorithms to guess the potential functional consequences of an amino acid substitution. Among the structural features the authors investigated, they scrutinized the crystallographic B-factor of the residue, the hydrophobicity, and the accessible surface area. The authors constructed a probabilistic classifier to predict whether a given SNP will have a benign or damaging effect on protein function. The authors reported favorable prediction performance when compared with other SNP prediction platforms.

MSV3D[144] offers similar features for analyzing SNPs, and provides a similar resource, and has mapped SNPs for 20199 human proteins by offering either direct mappings or homology models. The resource offers detailed tabular descriptions of many physical and chemical aspects of residue substitutions, however only offers limited visualization capabilities with Jmol.

MutDB[145-147] offers a platform whereby variants from dbSNP and SwissProt are unified to protein structures. For visualization capabilities the site relies on embedded Jmol applets, but also offers plugins for UCSF Chimera and PyMOL for even more advanced display options. The interface allows the locations of variants to be easily displayed.

While these tools are excellent resources for molecular biologists studying novel SNPs, these portals focus primarily on automated classification and not interactive exploration and visualization. Some of these tools offer Jmol visualizations of the

location of SNPs, but rarely provide other annotated sites that are mapped to the structure. This makes the visualizations of limited utility to end-users who may wish to see the big picture of how a protein functions. Finally, these tools focus on SNPs, and thus provide no direct facility to investigate post-translational modifications from mass spectrometry data.

Recently, a handful of sites have emerged to focus on post-translational modifications. Phospho3D curates phosphosites from the Phospho.ELM[148] database and permits interactive visualization of these sites in separate Jmol pages. Additionally, Phospho3D mines 3D structures to determine *zones* surrounding each phosphosite, defined as the neighborhood of residues within a 12 Angstrom distance. Phospho3D additionally computes several structural properties for each site, including secondary structure, solvent accessibility, B-factors, depth and protrusion indices, and disorder probabilities. While Phospho3D is a comprehensive resource, its visualization capabilities are of limited utility as they do not associate textual sequence annotations to protein structures. Furthermore, the site offers no ability to query arbitrary sites on proteins or upload peptide modification datasets for analysis. Instead a service is offered whereby users can upload PDB structure files to find similar 3D zones surrounding phosphosites.

Another recent project focusing on post-translational modifications is ptmfunc.[149] This system and website is a comprehensive analysis pipeline that extracted nearly 200,000 phosphorylation, acetylation, and ubiquitinylation sites from several data sources. The authors developed algorithms to prioritize the functional importance of modifications to determine interesting sites for future study. Their analysis

methods included determining phosphorylation hotspots within domain families and finding modifications that occur at interfaces between two bound proteins in structures. This mass analysis effort was conducted and the results were assembled into an online website. Despite the impressive analysis, the website only provides minimal information on the scrutinized sites in a tabular format. No capabilities are provided to visualize the proteins and modifications. Furthermore, no capability exists whereby users can submit datasets for analysis.

While there is clearly a plethora of structure mapping web resources, they offer a heterogeneous set of features. Most focus has been on analyzing variants, yet more attention needs to be placed on the growing wealth of modification data. Even more importantly, unification of variants, modifications, and functional annotated sites may lead to additional insights into protein function. Finally, we feel that less focus on automated analysis/classification and greater emphasis on exploration/visualization will allow researchers to better complement their research.

We have developed AMASS, a web-accessible system that brings together multiple bioinformatics databases, associates 1D sequence features to 3D coordinates in protein structures, and then enables the discovery and visualization of co-localized protein features. AMASS aims to bridge the gap between the wealth of 1D sequence feature annotations and the rapidly growing collection of 3D protein structures. Our system aggregates sequence-level annotations from multiple databases: The Uniprot KnowledgeBase (UniprotKB)[11], the Catalogue of Somatic Mutations in Cancer (COSMIC)[12], and the PhosphoSite database[14] of post-translational modifications. Alongside this aggregation of sequence level annotations, our system matches searched

proteins with structures from the Protein Databank (PDB). Structures that have a similar sequence are automatically aligned to the query sequence, and sequence-level annotations are subsequently mapped to the 3D structure. Structures are visualized with the Jmol[15] browser based PDB viewer, and sequence annotations can be interactively explored by clicking links within the interface.

With this system, users can explore co-localized functional residues, which may include important cancer mutations, post-translational modifications, or annotated sequence features such as binding motifs and active sites. The system also permits queries of arbitrary residues within proteins, and upload of sequence data (such as mass spectrometry data) to find possible functional roles of thousands of queried sites. The inspiration for the AMASS system is rooted in mass-spectrometry based discovery of protein modifications that correlate with metabolic disorder. The system is also designed from the ground up to bring together multiple bioinformatics databases to provide a coherent picture of how a protein works. To this end, we have named our resource "AMASS", and it can be accessed at amass-db.org.

System and Methods:

The AMASS server allows visitors to both 1) *browse* for their favorite proteins to investigate sequence-structure relationships, and 2) to *analyze* peptide sequences and modification data derived from mass-spectrometry based experiments. Both usage modes provide very similar interfaces. When browsing for proteins by accession or symbol, a table of possible protein matches is presented. Alternatively when a user uploads a dataset for analysis, the input is a list of peptide or protein sequences with queried sites denoted as lower-case letters. This set of sequences is matched against sequences from

the Uniprot database. Matching proteins are presented in a similar table, with added columns indicating features in the protein structure that are proximal to the queried modification sites. These features include ligands in the structure and functional sequence elements mapped to the structure.

For each protein viewed on the site, we present a table of residue-level textual annotations derived from Uniprot and PhosphoSite. The system preferentially focuses on any annotation records that indicate active sites, clinically relevant variants, ligand binding residues, or sites that have been studied in prior mutagenesis experiments. Additionally, we present an interactive bar chart which displays single-residue count data derived from several sources. Somatic mutation frequencies in tumor samples are displayed from the COSMIC database. Additionally, we present post-translational modification count data derived from the PhosphoSite database, which aggregates modification count data based on prior mass-spec datasets and literature publications. By presenting such count data from multiple sources, users can visualize the functional importance of various residues. This count data serves to complement the textual annotations that we aggregate and present in tabular form. The counts plot and sequence annotations table are both linked, so that exploration of either with the mouse causes automatic highlighting to occur in the other. This feature gives users the ability to explore the one-dimensional sequence features in a unified way.

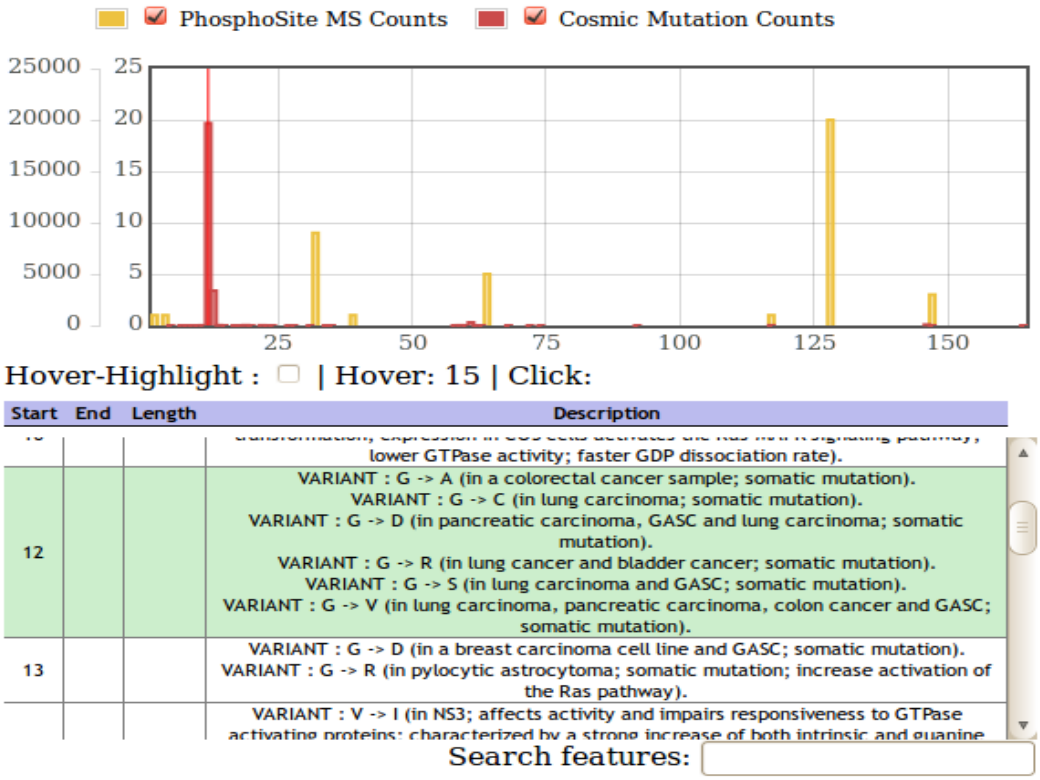


Figure 25 : Sequence counts and annotations display

The 1D counts plot in the top panel shows modifications from the PhosphoSite database (yellow) and cancer mutations (red) from the COSMIC database. The table in the bottom panel shows annotations from Uniprot

A major advantage of AMASS is that it has a mechanism for unifying this sequence annotation data with protein structures from the protein databank (PDB). If a queried protein has PDB structures, or is at least 50% similar to homologous proteins with structures, then these structures are presented in a table sorted by sequence similarity. Users can select structures to visualize with an embedded Jmol Java applet. Once a structure is loaded, users can interactively explore the structure by clicking on textual annotation records or on peaks in the counts plot. These actions also highlight the sequence alignment, which shows the user how the mappings are made between the queried protein and loaded structure. The alignments themselves can be clicked, allowing zoom-in of any residue on any specific chain. The structure viewing panel also contains several tools to change the color and transparency options of the model, as well as a contacts widget which uses Jmol to automatically find and highlight neighboring ligands which form Van-Der-Waals contacts with the most recently selected residue.

As an example, we use AMASS to search for human KRAS, a well-studied GTPase mutated in human cancers. Figure 25 illustrates how AMASS represents 1D sequence annotations, and Figure 26 shows a paired protein structure, complete with a visualization using Jmol and a displayed sequence alignment. A prominent somatic mutation (G12) and commonly-observed phosphotyrosine (Y32) are observed near the GTP binding site.

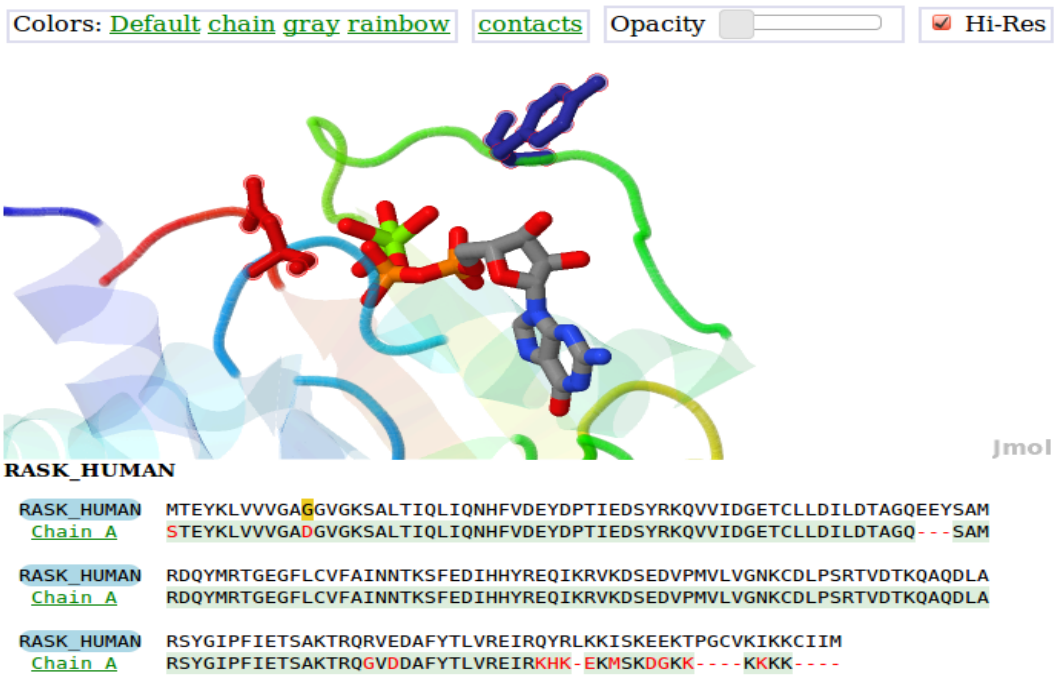


Figure 26 : Human KRAS and sequence alignment

To attain structural insights into the mutations, modifications, and sequence features of human KRAS, a structure from the protein databank (4dsu) is loaded and aligned to the KRAS sequence. Residues of interest in the counts plot (from Figure 25) were clicked to highlight them in the structure. The prominent mutation peak at G12 is associated to a mutated aspartic acid in the structure and alignment (red stick model, left). The prominent phosphorylation peak at Y32 corresponds to a tyrosine residue (Blue stick model, top) that is near the GDP binding site.

Development of AMASS required substantial database preparation and alignment computations. Most of this effort was performed ahead of time to allow rapid query and analysis for end-users. The development of the full pipeline involved writing a collection of Python scripts. On a modern server, downloading the relevant flat-file databases and populating our unified database takes approximately one week.

We used the Uniprot database as the core foundation of our platform. Given the comprehensive coverage of this database across many proteins and species, as well as its central role in mapping to many other external databases, we selected Uniprot to serve as the primary indexed identifier for each protein in our platform. Uniprot is composed of Trembl, a large collection of protein sequences prepared by translating genome sequences, and SwissProt, a much smaller but manually validated set of protein sequences. Swissprot is also commonly used as the primary sequence database for protein identification in mass-spectrometry experiments. Our pipeline downloads the compressed flat-file releases of these databases and parses them with the Biopython library, which enables efficient handling of the sequence data and feature annotations that Uniprot collects from many sources.

Our primary goal with AMASS is to enable the investigation of sequence-structure relationships. The protein data bank has seen exponential growth over the last decade, but despite this growth there has been little concerted effort to correlate deposited structures with protein sequence identifiers. Several projects, such as PDBSWS [150] have been published over the last several years to close this divide by extracting protein sequences from PDB files and performing blast searches against protein sequence databases such as Uniprot. This strategy requires great care to ensure that the proper

pairings are performed between sequences and structures. Fortunately, the SIFTS consortium [151] has recently established a unified database of Uniprot sequence mappings to individual chains within PDB files. While this dataset does not provide alignments, it provided a reliable source of sequence identities to which we could perform our own alignments.

We mirrored the entire contents of the World Wide Protein DataBank (wwPDB) to our server, and used the Prody [152] PDB library to parse and analyze each PDB file in a large-scale effort. Header information, such as the title and crystallized ligands were stored in a relational datastore. We then used Prody to enumerate each chain parsed in the PDB structures. For polypeptide chains, a sequence of residue names and residue IDs was extracted. Single-letter amino acid sequences were interpreted from the 3-letter residue names, and these chain sequences were associated to Uniprot sequences with the SIFTS database. Each chain sequence was then individually aligned using Clustal Omega [153] to its associated Uniprot sequence to determine the residue-level mappings. These independent alignments were merged together into single "chain alignment" FASTA files which contain a parent Uniprot sequence and all associated chain sequences for that given PDB structure. We have observed that the vast majority of alignments between sequences and structures are identical as described by the SIFTS database. Minor discrepancies exist however, such as the presence of affinity tags within crystallized proteins or occasional site-directed mutants at relevant functional sites.

Aside from the chain sequence alignments, the residue IDs for each PDB chain were stored into a Python array and serialized to an object-oriented database for rapid retrieval. Given the computed "chain alignments" and these residue ID mappings, we

were then able to associate any Uniprot residue to residues within PDB chains, and subsequently to the atomic coordinates of particular residue IDs within a PDB file. This process formulates the core of our system, and allows mapping of 1D sequence annotations to 3D locations within protein structures.

In addition to extracting and aligning sequences from every PDB, we also extracted topological data from each structure. For every non-water residue, the Prody selection engine was used to find other non-water residues within a 7 angstrom distance. These searches were performed on every residue in every PDB file. These protein *zones* were stored in an indexed object-oriented database to allow rapid lookup for later analysis efforts. This approach was largely inspired by the approach taken by Phospho3D in which 3D *zones* were calculated for all phosphosites. Our system performs this analysis for all residues, modified or not.

Despite the large number of structures deposited in the Protein DataBank, its coverage of the proteome for any organism is far from complete. Many human proteins have not been crystallized. Usually however, highly similar homologous proteins can be found in the PDB, either as human paralogues or orthologues in other species. To expand the exploratory horizon of our platform, we designed it from the start to enable pairing of queried proteins with homologous proteins in the protein databank. Similar web resources that have been developed, such as MSV3D, instead produce homology models of uncrystallized proteins, which is more computationally demanding. Our approach is to simply find similar structures and present sequence alignments to end users. This is far simpler, and also more transparent, as it allows users to directly inspect the similarity between their queried protein and the paired structure. Inferences made at the single

residue level can be sanity checked if the paired structure is not conserved in other species.

To accomplish this sequence-structure pairing, we used all structure sequences documented in the SIFTS database as a query in an all-vs-all BLAST [154] search against the SwissProt database. Sequences that had at minimum of a 50% match were stored in a relational database. Subsequently, we pre-computed all sequence alignments between the SIFTS sequences and matched SwissProt sequences using Clustal Omega. [153] These "homology alignments" were separately stored as independent files to disk. These alignments are later merged with the PDB specific chain alignments to permit the mapping of sequence feature annotations from any protein sequence to 3D structures of homologous proteins. This alignment merging, and subsequent residue mapping, is the core methodology employed by the AMASS server.

Algorithm:

During either live exploration of proteins on the web interface or analysis of uploaded datasets, the AMASS server pairs each protein sequence of interest with relevant structures from the protein databank. To find structures in the PDB with similar sequences, the system first consults a MySQL database which stores precomputed BLAST search result. In search mode, these structures are presented to the user, whereas in analysis mode these structures are each systematically investigated.

For each structure, the system loads the pre-computed "homology alignment" between the queried protein sequence and the Uniprot sequence assigned to that given structure. The system then loads the separate "chain alignment" files that map the Uniprot sequence to individual peptide chains within the PDB. The system merges these two

alignments based on the common PDB Uniprot sequence defined by the SIFTS database. This alignment merging is implemented in Python. Sequences in both alignments are first converted to linked-list data structures for efficient manipulation, and then gaps are inserted into each alignment while scanning the joined Uniprot sequence in each. After this, the sequences are converted back to strings and then packed back into a Biopython compatible sequence alignment object.

Once this merged alignment is produced between a queried protein and a homologous protein structure, we can map sequence positions on the query protein to the protein sequence in the PDB structure, and then onto single residues in each chain. To perform this mapping rapidly, we implemented a simple data structure in Python that represents the alignment as a series of Python integer arrays for efficient random access. Each protein and chain sequence within the alignment has an associated “forward map”, which is an integer array with length equal to the sequence. The integer values within the array map the sequence index to the position within the alignment. Each sequence also has an associated “reverse map”, which is an integer array with length equal to the sequence alignment. The integer values within this array map the alignment positions back to indexes within the sequences. Thus, to map any position within the query sequence, the forward map is used to find the column in the merged alignment. Then, to map to a single residue within a chain, this chain's reverse map is used to find the corresponding position in the chain sequence. Given this sequence position, we can finally determine the residue ID within the PDB file by consulting the residue ID mapping arrays that were stored to disk during the database creation. This sidesteps the

fact that many PDB files use residue numbering schemes that do not agree with sequence indexes in the native protein.

While our site allows visualization of any protein of interest, AMASS also provides an analysis backend to systematically scrutinize in "batch" mode single modifications on many peptide sequences, e.g., derived from mass-spectrometry experiments. The input format follows a standard in the mass-spectrometry field, in which peptide sequences are capitalized, while detected modifications are denoted with lower-case letters. The user provides an email address for notification when the job has completed, and indicates the species that the sample is derived from. After submitting the data, the system checks for common errors such as non-amino acid characters, and then de-duplicates non-unique peptide sequences while keeping an internal count of coverage and modifications. We encourage users to upload raw peptide data, wherein the same peptide sequence is listed many times. This "spectral count" data is reported back to the user for each and every residue in unmodified and modified counts.

Given the list of unique peptide sequences, our system first discovers the set of Uniprot sequences that potentially match the sample, given the species of origin that the user indicated during submission. This search is efficiently implemented with the UNIX fgrep program. After this set of target proteins is found, peptide sequences are assigned to whole protein sequences using the string containment algorithm implemented in Python. Site coverage and modification count data is accumulated in integer arrays, and then later stored to a MySQL database for long-term aggregation of the most commonly occurring protein modifications.

After submission and validation, the job enters a queue. When executed, the backend analysis pipeline finds similar PDB structures and merges alignments as previously described. Modifications are then mapped to 3D coordinates, and a neighborhood of structural features within 7 Angstroms is found. Within the neighborhood residue set, our algorithm tests membership of residues that have important sequence features such as active sites, mutagenesis sites, or binding residues as annotated within the Uniprot or PhosphoSite databases. Additionally, our system determines if the site is located at the binding interface between two distinct proteins in a complex. Finally, the system checks nearby to look for interesting ligands bound to the protein, which can include enzymatic metabolites or drug compounds that bind the active sites of kinases.

Implementation:

The AMASS database is hosted on a dedicated server with dual Xeon Westmere processors, 32GB of RAM, and 8TB of RAID storage. The server runs the Ubuntu GNU/Linux Operating System. The backend analysis engine and web frontend are both implemented entirely in the Python [17] programming language, utilizing the Biopython [155] bioinformatics library for sequence processing and the Prody PDB parser for structural analysis.

MySQL [156] is used as a relational database to store several datasets, including Uniprot protein accessions, IDs, description fields, and identifier mappings to external databases. Additionally, metadata extracted from PDB files was stored in relevant tables. MySQL enabled quick development of search interfaces on our web frontend with its full text search indexes.

In addition to a conventional SQL datastore, we also made extensive use of object-oriented storage of Python objects in serialized binary files with the cPickle and Shelve modules. For certain datatypes, this allowed for increased performance of the backend analysis engine and web frontend. Python objects describing PDB chain alignments, structural neighborhood mappings, and Biopython Uniprot records were all stored as binary Python objects directly to disk. This strategy improved analysis speeds and simplified implementation as no object to relational mapping was required. This strategy also enabled easy portability of the database and distributed computation to be performed with shared network filesystems.

All PDB file parsing and analysis was performed using the excellent Prody library. At the beginning of our efforts, we surveyed several PDB parsing libraries and found Prody to be the fastest, and also had the most convenient object-oriented interface. Python scripts were written to pre-process all PDB files. This extensive process involved chain sequence enumeration and extraction using Prody, followed by sequence alignments with Biopython. Additionally, Prody was used to calculate structural neighborhoods for all non-water residues in each structure. By building this large database ahead of time, we greatly improved performance of our analysis engine when presented with novel datasets.

For the web frontend, the Apache2 [157] suite is used to serve the content. Dynamic pages are implemented as Python CGI scripts. The Jquery JavaScript library was used extensively to create interactive content on the pages. The DataTables [158] Jquery plugin was used to allow client-side searching and sorting in several HTML

tables. The Flot [159] plotting library was used to create interactive plots of site-specific count data.

Protein structure visualizations are powered by the open source Jmol [15] PDB viewer. The Jmol javascript API was used to implement several interactive interfaces which allow end-users to explore sequence features and analysis results.

Case Studies:

To illustrate the utility of AMASS for uncovering and exploring interesting post-translational modifications, we have analyzed a prior mass spectrometry dataset collected on isolated mitochondria from human skeletal muscle. [107] Within this data, AMASS found several proteins with potentially interesting modifications near active sites.

Pyruvate Dehydrogenase (PDH):

The pyruvate dehydrogenase complex is a central metabolic enzyme that connects glycolysis with the citric acid cycle by converting pyruvate to acetyl-coa. Our mass-spectrometry data from human muscle demonstrates that serines 293 and 295 of E1 subunit alpha (Uniprot: ODPA_HUMAN) are phosphorylated *in vivo*. Both of these serines are documented in the PhosphoSite database, with S293 in particular having 625 references to external mass-spectrometric datasets. Furthermore, S293 is the most studied mutagenesis site, with 13 references documented in PhosphoSite. Most interestingly, these two phosphorylations are located in a cluster of 6 phosphosites present in a short span of the protein sequence. (Figure 27) Two of these sites (S293 and S300) are annotated in Uniprot as targets of pyruvate dehydrogenase kinase (PDK4), which is a kinase which regulates the activity of PDH. [160]

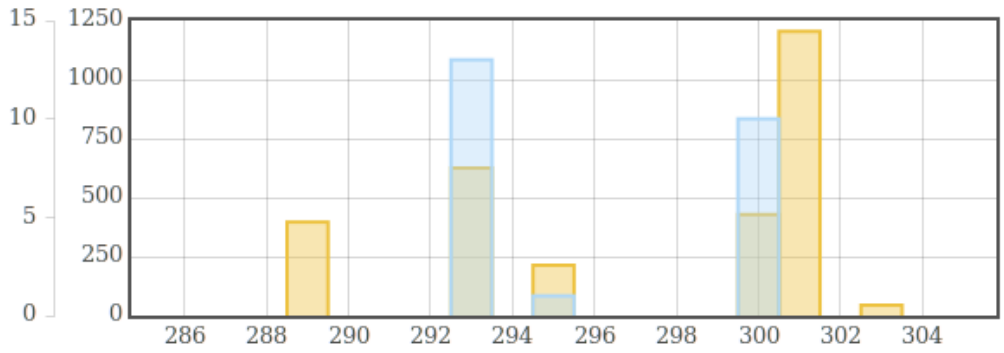


Figure 27: Phosphorylation counts of PDH

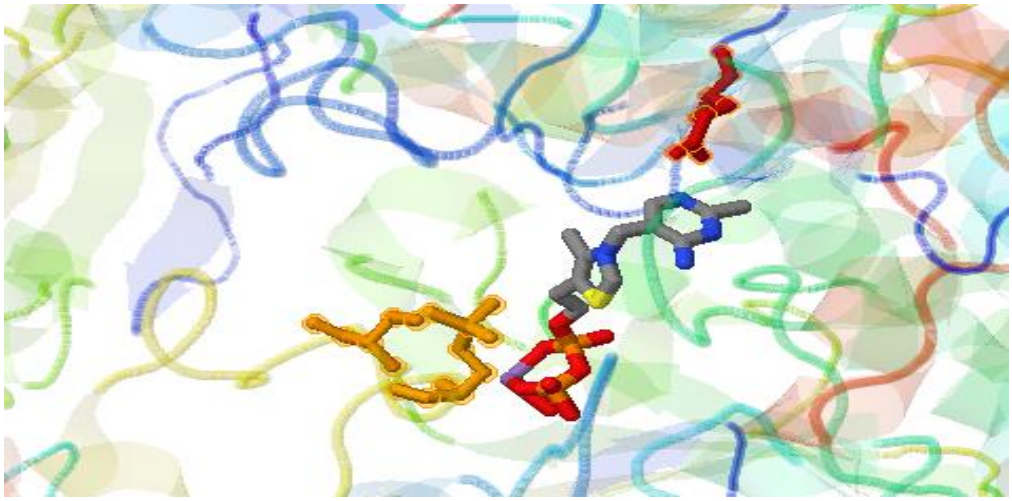


Figure 28: Phosphosites in PDH structure

A serine and phosphoserine residue (orange) are located on a loop region in the crystal structure. This loop is proximal to TPP, a cofactor of PDH.

The protein has 7 crystal structures in the PDB, and in one of these structures (3exh) AMASS found that S293 is proximal to an interesting ligand. Upon loading the structure and visualizing the alignments, we note that the crystal structure was prepared from the mature enzyme that has a 29-residue mitochondrial transit peptide removed. The phosphosite S293, identified in the full-length Uniprot sequence, aligns with S264 in the mature enzyme. Despite this discrepancy, the sequence alignment procedure of AMASS ensured that the correct full-length phosphosites were correctly associated to the chains in the structure. The crystal structure has four chains of this protein, and this residue is actually found phosphorylated in one of these chains. This structure was prepared by phosphorylating PDH with PDK4, and the heterogenous phosphorylation of the 4 chains permits comparison of secondary structure.

AMASS found that in chain C in particular, this phosphoserine is located very close to the ligand Thiamine PyroPhosphate (TPP), a derivative of Thiamine (Vitamin B1), and a required cofactor of PDH. Furthermore, we readily observe that the other phosphorylated serine S295 (S266 in the mature form) is also repositioned in the chains where S264 is phosporylated. It is discussed in the corresponding publication of this crystal model [160] that phosphorylation of the S293 residue (via PDK4) causes derangements of this phosphorylation loop region and prevents substrate channeling to the TPP cofactor, thus inactivating the entire PDH complex. In this case, S293 is well studied, yet the annotations within Uniprot only discuss PDK4 as the responsible kinase but do not assign a strong functional role of S293 in regulating metabolic activity. Nonetheless, AMASS found the proximity to TPP, which enabled us to uncover the literature citation associated with this PDB structure and ultimately provided us with

insight to our data. This emphasizes the potential of AMASS to serve as a tool to integrate data and help the research community improve functional annotations of sites.

ATP Synthase Beta (ATPB):

In human skeletal muscle, peptides from human ATP synthase are particularly abundant. On the beta subunit of this complex, AMASS found a phosphothreonine modification (T213) that is located near interesting ligands in several structures in the protein databank. Human ATP synthase beta has not been crystallized, but several models exist of the bovine protein (97% sequence identity) and the yeast protein (78% identity). In many of these structures, AMASS finds that phosphorylated T213 is located nearby magnesium ions and either ATP or ADP molecules. This site was previously investigated in relation to insulin resistant skeletal muscle, and deduced to likely be located near the ATP binding domain based on sequence analysis. [161]

Using AMASS to directly visualize the ATP synthase beta protein (PDB 4asu), we observe that three beta subunits and three alpha subunits complex to form the F1 subunit. Each of the three beta subunits has an ADP molecule present at the active site. Additionally, magnesium cofactor ions are seen bound to the terminal phosphates of the ADP molecules. The T213 sidechain is positioned such that direct contact is made with the magnesium ions (Figure 29). This finding is interesting, as this phosphorylation of T213 could thus be an autocatalytic mechanism due to the proximity of the ADP phosphates, or perhaps an unknown reaction intermediate. We postulate that phosphorylation of this threonine may effect catalytic function by inhibiting binding of either the ADP substrate or the magnesium cofactor.

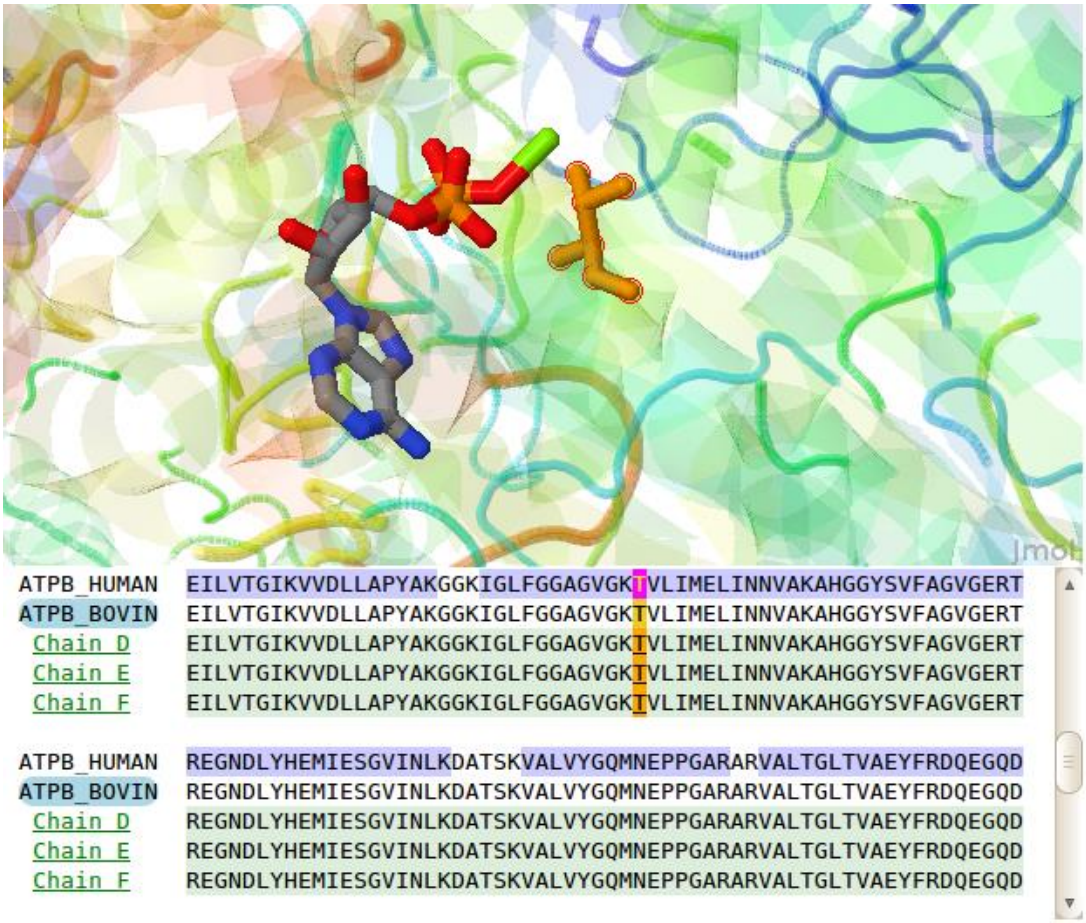


Figure 29: Bovine ATP synthase and sequence alignment

ADP and T213 (orange) which binds the magnesium cofactor (green). Below: Multiple sequence alignment showing mass spec coverage (purple) and alignment between human and bovine ATPB, and alignments to chains in PDB.

Enolase Beta (ENOB):

Our muscle dataset identifies the three isozyme subunits of enolase. Alpha (24 unique peptides), beta (27 unique peptides), and gamma (5 unique peptides). Two phosphorylations (S37 and T41) were detected that map to all three sequence forms, whereas a third phosphorylation (S176) only maps to the beta sequence. The beta form is characteristic of striated muscle, and so we focus our analysis solely on this isozyme subunit (ENOB_HUMAN). All three phosphorylations are documented in the PhosphoSite database. S37 and S176 only have a few mass-spectrometric counts in prior datasets, whereas T41 has 44 documented ocurrances in human samples. Investigating the PhosphoSite count data in AMASS shows that S37 and T41 are proximal to Y44, which is phosphorylated in 1623 distinct human samples annotated in PhosphoSite.

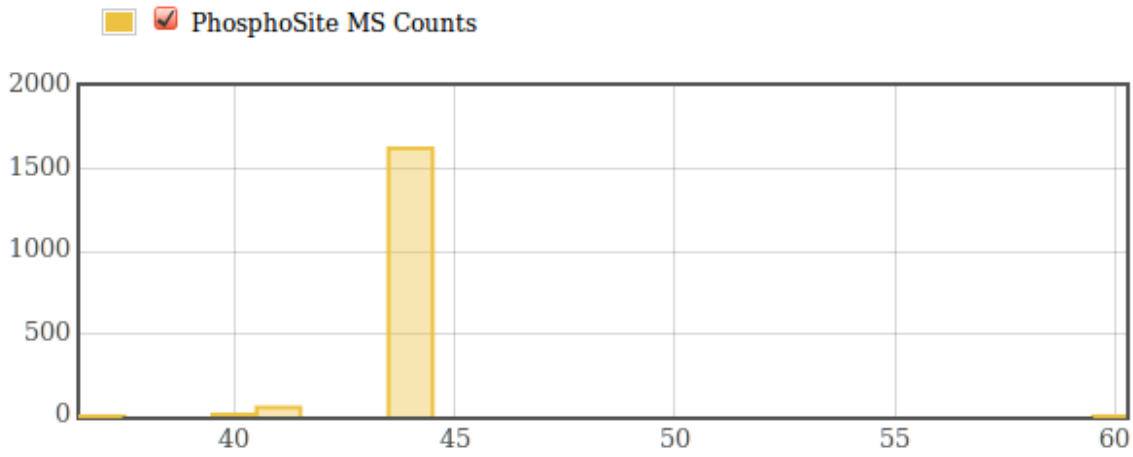
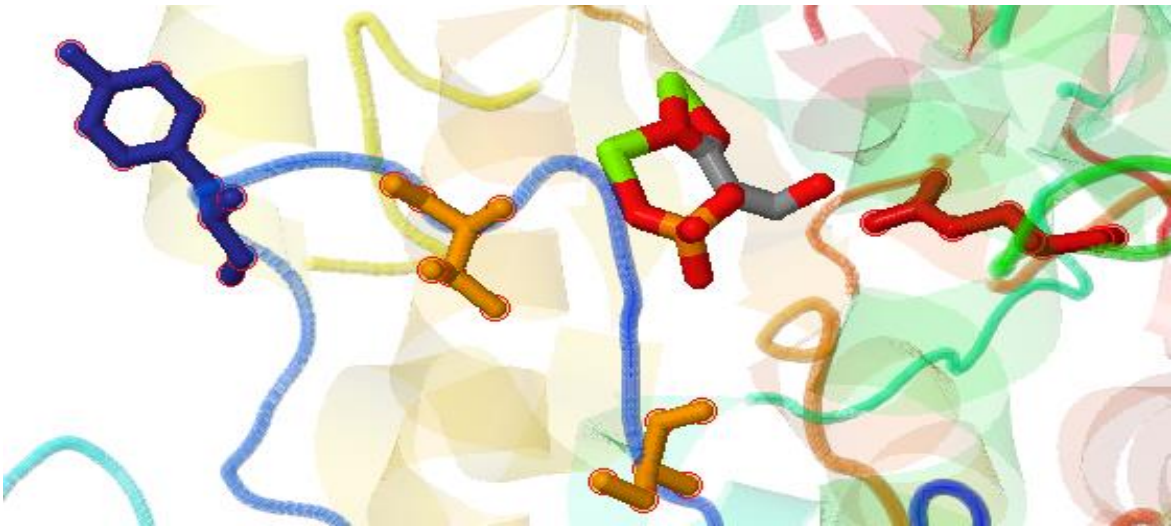


Figure 30: Phosphorylation counts of Beta Enolase



Two serines that are phosphorylated in our dataset (orange) are proximal to the bound substrate (2-PG) at the active site of the enzyme. An additional serine S40 (not depicted yet) serves as a direct binding site of 2-PG. A tyrosine (blue) is located on the same loop but distal to the active site, and is highly phosphorylated in many datasets documented by the PhosphoSite database, but unobserved in our data. The active site residue of the enzyme is depicted in red.

Only a single structure for ENOB_HUMAN exists in the PDB, however AMASS failed to find any interesting proximal features to these three sites. The gamma isozyme (ENOG_HUMAN) however has nine crystal structures in the PDB, and several of these structures were prepared with 2-phosphoglyceric acid and phosphoenolpyruvate, the substrates of this enzyme. Given that gamma enolase has 84% sequence identity with the beta form, AMASS aligned the queried modifications to those structures for similar analysis. S37 and T41 are both found proximal to the substrate binding sites.

We chose the structure 3ucd for visualization as it contained two polypeptide chains, one with 2-phosphoglyceric acid (2-PG) and the other with phosphoenolpyruvate (PEP) occupying the active site. In either chain, we find that S37 and T41 sit on a loop that overlaps the substrate binding pocket. Both serines are within a nanometer of either 2-PG or PEP. The heavily-phopshorylated tyrosine Y44 is located on the same loop, but further from the binding pocket. Interestingly Y44 forms a salt bridge with a nearby D298, and might stabilize this loop region. Given that these phosphosites all cluster on this loop which contacts the bound substrate, we postulate that phosphorylations may effect catalytic function through electrostatic interactions with bound substrates.

Argininosuccinate Synthase:

In human liver, we identify argininosuccinate synthase (ASSY_HUMAN) with a serine phosphorylation at S180 and lysine acetylations at K228 and K355. The enzyme catalyzes the formation of argininosuccinate from ATP, citrulline, and aspartate, and is involved in the hepatic urea cycle. Mutations to this protein can cause citrullinemia, a disease characterized by elevated citrulline levels in serum and urine as well as ammonia intoxication in infants.[162]

A single structure in the PDB exists for human argininosuccinate synthase (2nz2), and is co-crystallized with citrulline. AMASS found that the S180 residue is proximal to the citrulline ligand in the structure, and also reported that S180 is an annotated citrullin binding residue in the Uniprot database. Interestingly, the sequence annotations table from Uniprot also describes S180 as a mutation site. In a patient suffering from neonatal citrullinemia, S180 was found mutated to an asparagine.[163] This is compelling as asparagine has only a marginally larger sidechain, yet its substitution was sufficient to clearly lead to enzyme deactivation in this particular case. This hints at the functional importance of S180.

A homologous structure (1j1z) from thermus thermophilus is also present in the PDB. The sequence is only 53% identical, however the structure appears similar, and more importantly the model was prepared with citrulline, aspartate, and ATP all bound at the active site. When aligned, human S180 corresponds to S173 in this organism. When the location of this serine is visualized using the AMASS contacts widget, we observe that it forms Van-Der-Waals contacts with citrulline and the gamma phosphate of ATP. It is inferred from this arrangement that the introduction of a negative charge at this serine through phosphorylation may inhibit binding of negatively-charged ATP. Serine phosphorylation may be an auto-catalytic property of this enzyme or an external regulatory mechanism that requires further investigation. Phosphorylation of S180 in humans has only been observed in one dataset annotated in the PhosphoSite database.

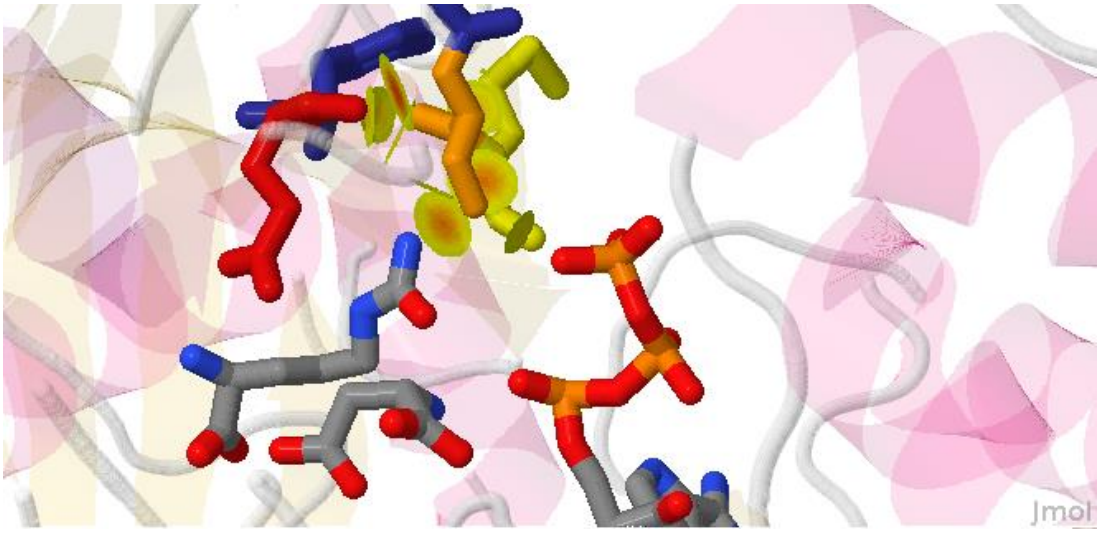


Figure 31 : Active site of argininosuccinate synthase

S180 (top, orange) shown interacting via VDW contacts with citrulline (left) and the gamma phosphate of ATP (right). Aspartate is depicted bottom-left.

Fructose Bisphosphate Aldolase:

Fructose bisphosphate aldolase is an important enzyme in glycolysis, and is found in three isoforms (A,B and C). Isoform A (ALDOA_HUMAN) is most prevalent in muscle, and we thus observe the most unique peptides for this form. For this protein, we detect 5 distinct phosphosites and one lysine acetylation. Intriguingly, three phosphosites are clustered in the sequence (S36, T37, and S39), and previously documented in a handful of samples in PhosphoSite.

This protein is crystallized in 7 structures in the PDB, and one of these structures (4ald) has the fructose bisphosphate (2FP) ligand present. We observe that these three clustered phosphosites are located proximal to a terminal phosphate on the 2FP molecule. The two serines make direct VDW contact with the 2FP phosphate, whereas the threonine sidechain is oriented in an opposing direction. By using the contact widget, we additionally observed that the sidechain of K180 also binds this phosphate, and appears to be a primary electrostatic interactor with the ligand. We hypothesize that phosphorylation of serines 36 and 39 may effect 2FP binding through electrostatic repulsion of the terminal phopshate of 2FP. These serines could be phosphorylated by the bound ligand itself, and may inactivate the enzyme.

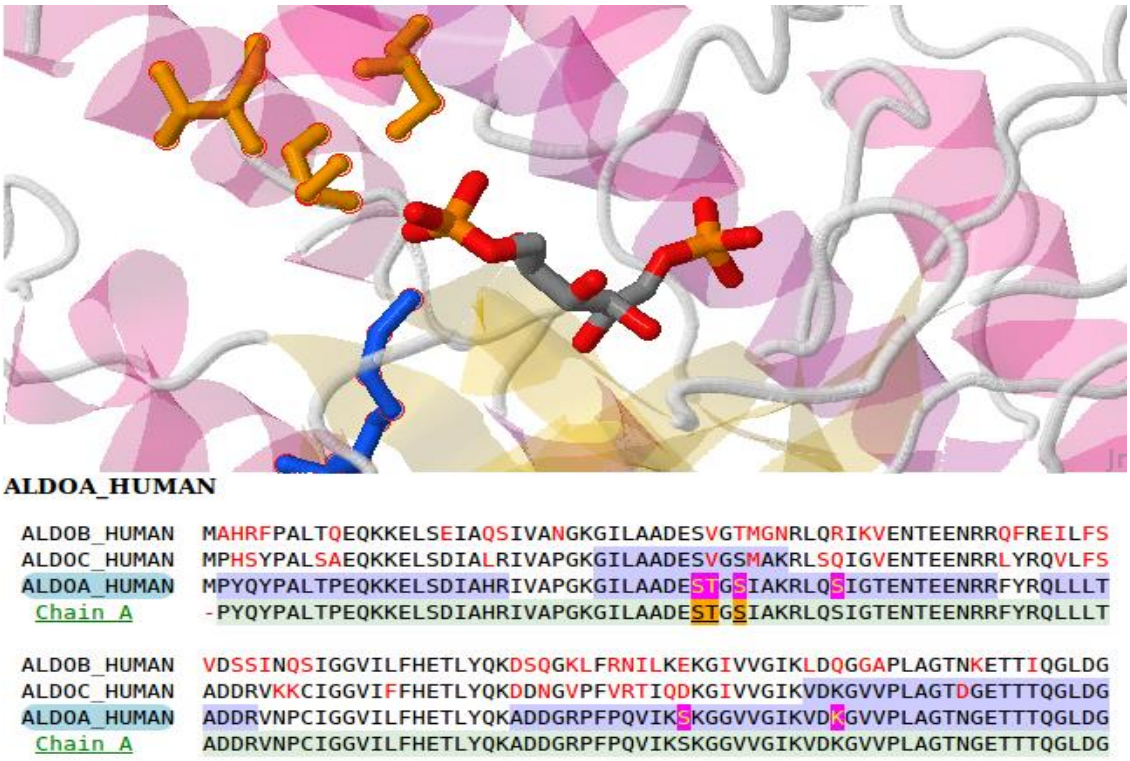


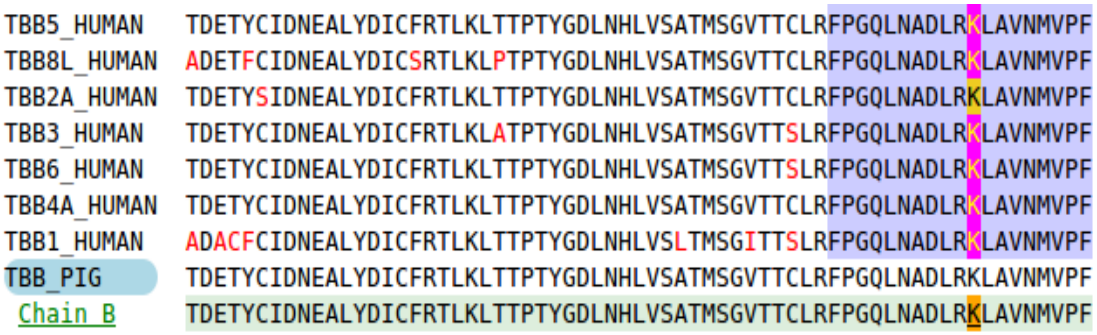
Figure 32 : Structure and alignment of Aldolase A

(Top) the structure of ALDOA_HUMAN showing three clustered phopshosites. Two of these sites are serines, and are forming direct VDW contact with a phosphate group of fructose bisphosphate. To the left is K108 (blue), a lysine which electrostatically binds this phosphate. (Bottom) Sequence alignments showing coverage of peptides mapped to all possible proteins determined from the dataset. Mass spectrometer coverage is indicated with blue highlighting, and modifications are shown in purple in the sequence. Orange highlighting indicates chain residues with interesting findings.

Tubulin:

In human liver, we detect many peptides that are assigned to a variety of tubulin isoforms. Altogether, 11 isoforms and tubulin-like proteins from SwissProt are identified with peptides from liver. The isoform with the most representation and coverage is Tubulin beta-2A (TBB2A_HUMAN) with 27 unique peptides. Tubulin proteins are the constituents of microtubules, and polymerize by forming heterodimers between alpha and beta subunits. A GTP molecule is bound at the dimerization interface between chain A and chain B. An acetylaton is detected at K252, which corresponds to the same sequence position in the other tubulin isoforms. PhosphoSite reports that for Tubulin beta-2A, this lysine is acetylated in 75 distinct samples.

AMASS mapped this modification to many distinct structures in the PDB of tubulin proteins from cow, sheep, and pig. In several of these structures, GTP is detected as a proximal ligand to this lysine. In visualizing a pig structure (2wbe) we note that K252 is present on the B chain and directly binds the alpha phosphate of GTP located at the binding pocket of tubulin chain A. Given that lysine acetylation results in a loss of positive charge, we postulate that acetylation of this crucial residue may inhibit GTP phosphate binding and thus regulate tubulin polymerization. Excitingly, tubulin acetylation has been documented in the literature, and has been demonstrated to affect polymerization function.[164]



Sequence	Mods			
TBB1_HUMAN	E123	K252	K324	N379
TBB2A_HUMAN	E123	K252	K324	K379
TBB2B_HUMAN	E123	K252	K324	K379
TBB3_HUMAN	E123	K252	K324	K379
TBB4A_HUMAN	E123	K252	K324	K379
TBB4B_HUMAN	E123	K252	K324	K379
TBB5_HUMAN	E123	K252	K324	K379
TBB6_HUMAN	E123	K252	K324	K379
TBB8L_HUMAN	E123	K252	R324	T379
TBB8_HUMAN	E123	K252	R324	K379
YI016_HUMAN	E51	K180	R252	T307
TBB_PIG	E123	K252	K324	K379
Chain B	E	K	K	K

Figure 33: Sequence alignment (top) and modification mapping table (bottom) from the AMASS web interface showing how detected modifications map to the various possible isoforms of proteins. The peptides that mass spectrometers identify can be associated with several protein isoforms. Tubulin is an excellent example of how some peptides and modifications may have many possible mappings.

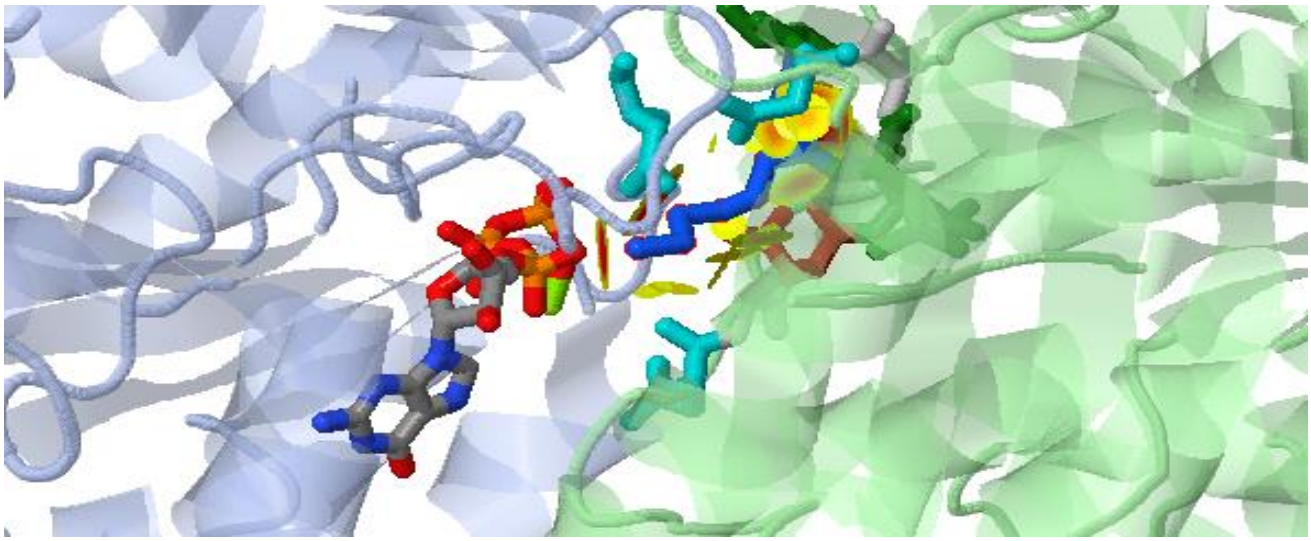


Figure 34 : Tubulin dimer interface

Tubulin chains A (blue) and B (green) are shown co-crystallized with a bound GTP ligand located at the tubulin A GTP binding pocket. Lysine 252 (blue, center) is shown binding the alpha phosphate of GTP.

Isocitrate Dehydrogenase (IDH):

In proteomics experiments conducted on human muscle tissue, we identify several peptides that match the cytosolic (IDHC_HUMAN) and mitochondrial (IDHP_HUMAN) forms of isocitrate dehydrogenase. In muscle tissue, we detect two lysine acetylations which can be attributed to the mitochondrial isoform: K48 and K180. K48 is not a reported modification in the Phosphosite database, however PhosphoSite reports that K180 is modified in human IDH in 28 datasets. The COSMIC database reports that two residues are frequent cancer mutation hostpots: R140 and R172. We display the sequence coverage of IDHP from our experimental data, with the two acetylated lysines highlighted in purple. Also shown is a plot of PhosphoSite modification count data (yellow) and COSMIC mutation frequencies (red).

Human mitochondrial IDH has not been crystallized, however AMASS associated it with the sequence from pigs which has 97% sequence identity. Pig mitochondrial IDH has a single structure in the PDB (1lwd), which was crystallized with a bound isocitrate ligand at the active site.[] AMASS determined that K180 acetylation may have a functional effect to the activity of this enzyme. When aligned with the pig sequence, K180 (human) corresponds to K149 (pig), which is a single residue from Y148 (Y179 in humans). The Uniprot record for IDHP_PIG has two sequence annotations for Y148. It is documented as a binding residue for isocitrate, and was also shown in a prior mutagenesis study[] to be critical for catalysis. Mutations to phenylalanine (F) lead to a significant decrease in Vmax. Alongside these annotations, Uniprot also refers to the two cancer mutation hotspot residues as binding sites for the isocitrate ligand.

Sequence

MAGYLRVVRSLCRASGSRPAWAPAALTAPTSQEQPRRHYADKRIKVAKPVVEMDGDEMTRII
WQFIKEKLILPHVDIQLKYFDLGLPNRDQTDDQVTIDSALATQKYSVAVKCATITPDEARVEEF
KLKKMWKSPNGTIRNILGGTVFREPIICKNIPRLVPGWTKPITIGRHAHGDQYKATDFVADRA
GTFKMVFTPKDGSGVKEWEVYNFPAGGVGMGMYNTDESISGFAHSCFQYAIQKKWPLYMST
KNTILKAYDGRFKDIFQEIFDKHYKTDFDKNKIWYEHRLIDDMVAQVLKSSGGFVWACKNYD
GDVQSDILAQGFGSLGLMTSVLVCPDGKTIEAEAAHGTVTRHYREHQKGRPTSTNPIASIFAW
TRGLEHRGKLDGNQDLIRFAQMLEKVCVETVESGAMTKDLAGCIHGLSNVKLNEHFLNTTD
FLDTIKSNLDRALGRQ

IDHP_HUMAN - Annotations

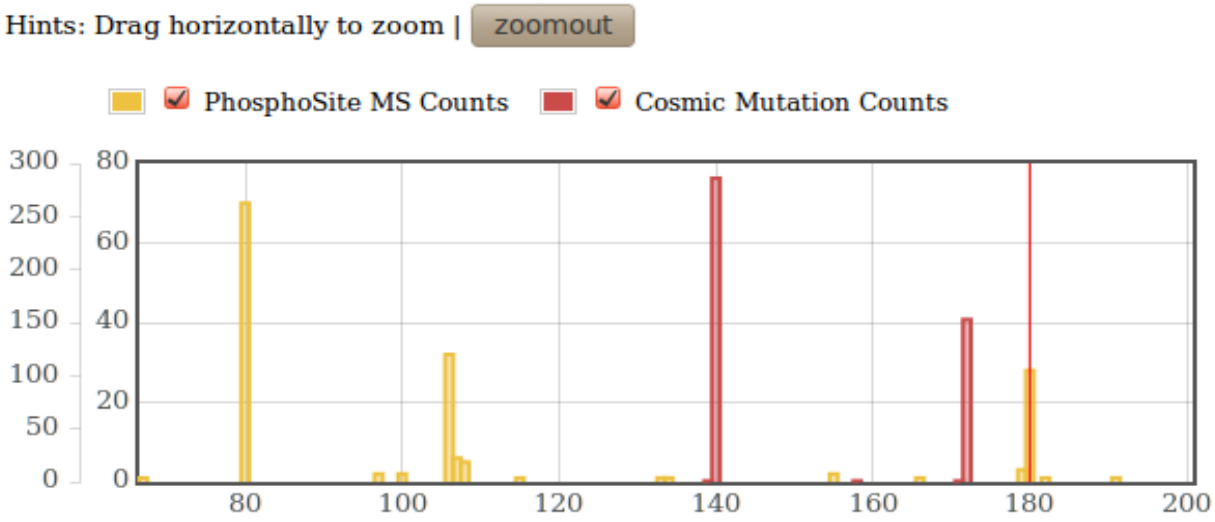


Figure 35 : IDH sequence coverage, phospho and mutation counts

(A) IDH mass spectral sequence coverage in blue with modified residues highlighted in purple. (B) Count data showing cancer mutation frequencies from the COSMIC database (red) and post-translational modification count data from PhosphoSite (yellow). K180, which is acetylated in human muscle, is abundantly acetylated in datasets aggregated in PhosphoSite.

By interacting with the page, we are able to easily determine the structural locations of these critical residues. As we depict in Figure 36, Y148 is indeed a binding residue of isocitrate, as are the two arginine residues commonly mutated in cancers. The acetylated lysine, K180 (K149 in pig) is a single residue away from the binding tyrosine Y148, however the sidechain oriented in an orthogonal direction. The lysine appears to be positioned at the top of a cleft in the protein formed between two seperate halves of the active site. By using the contact finder widget on the page, we find that this lysine contacts a nearby asparagine residue located on the opposing wall of this cleft. From this visualization, we hypothesize that acetylation of this lysine may affect the structural stability of the binding pocket. K149 acetylation may also effect the orientation of Y148, thereby effecting catalysis. Mutagenesis studies and/or computational modeling is required to investigate this hypothesis further, however AMASS has proven useful to find this possible connection.

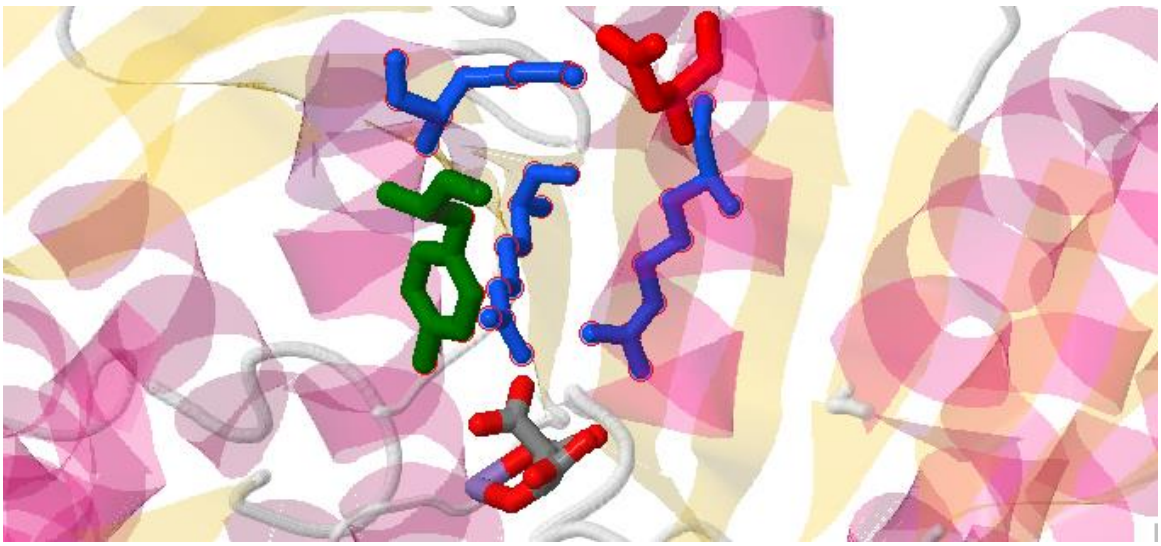


Figure 36 : IDH structure

Acetylated lysine (top blue) is positioned directly above the critical binding residue Y148 (left green), and interacts with negatively charged asparagine (top red).

Discussion:

We hope that AMASS will grow into a valuable resource for the biomedical research community, and we thus have a substantial roadmap of desired features planned for future releases. We plan to integrate many more biological databases to bring together as many functional annotations as possible. For example, the COSMIC database has given us an excellent jumpstart in providing mutational frequencies of single amino acids, however the recent release of full tumor sequencing data from The Cancer Genome Atlas (TCGA) project [165] hints that much more cancer mutation data can be integrated in the coming years as sequencing becomes cheaper.

Our analysis system is incredibly simplistic, as it only considers proximity of co-localized features to infer significance. Our primary focus for this system was to create an interactive user-friendly interface. We feel that visualization and appropriate aggregation of data will allow users to generate their own hypotheses about functional sites. Nonetheless, several other tools available perform far more sophisticated analysis of the structural consequences of modified sites, typically for SNP databases. For example, PolyPhen and SNPs3D both offer computerized classification of whether or not amino acid changes may affect function based on changing biophysical properties such as sidechain hydrophobicity, charge, or packing. Adding this deeper level of analysis in the future may improve the utility of our tool. The challenge for post-translational modifications will be to find appropriate algorithms to assess functional consequences of these novel sidechains.

We have observed that many proteins in the Uniprot database lack comprehensive functional annotations, especially when proteins from less-studied species are queried.

Furthermore, highly-valuable feature records such as mutagenesis sites are species-specific, since those annotations arise from academic publications that focused in a single model organism. We plan to improve this discrepancy in the future by integrating a homology clustering system into AMASS, whereby all homologous proteins (and their associated site annotations) can be visualized in alignments and simultaneously projected onto 3D structures. This improvement poses considerable user interface design challenges, but may vastly increase the exploratory horizon of our system.

Many protein structures within the PDB have multiple proteins in complex, however several distinct structures can be structurally aligned on common shared protein sequences to elucidate other possible complexes that may exist in vivo. For example, a structure containing proteins A and B can be structurally aligned with a separate structure consisting of proteins B and C. Such an alignment may hint how proteins A and C can interact in vivo. Performing such alignments across all structures may expand the set of possible neighboring features that can be discovered. Furthermore, adopting a technique similar to ptmfunc [149], whereby conserved domains amongst many distinct proteins are aligned, may allow greater generalization to be made for the functions of single sites.

With the development of AMASS, we have a system that is capable of quickly finding possible functional explanations of any queried modifications. This analysis however is only performed upon user submission of data, or when the user interactively explores a structure through the web interface.

In the future, we plan to perform a massive automated screen which will find co-localized features in all protein structures. Aggregating these findings and clustering relevant functional sites, cancer mutations, and high-abundance modifications remains a

challenge, as does presenting the results in a medium that permits the ranking of identified clusters. To enable such ranking, one could devise simple heuristics based on the number and density of functional sites within the neighborhood of active site residues or ligand binding pockets. Additionally, representation of clusters in 3D, using the Jmol applet, would allow for the easy visualization of functionally relevant surfaces of proteins. We will additionally investigate the idea of rendering colored heat-maps onto 3D structures that can represent variant statistics amongst haplotypes, cancer mutation rates, or modification count data for each residue.

Given our developed system and planned features, we envision that AMASS will serve as a valuable tool to the research community at large, by both enabling the understanding of datasets, as well as the discovery and annotation of unrecognized functional sites of proteins.

Chapter 5

THE MISSION AHEAD

Metabolic syndrome is a complex phenomenon, and is thusfar documented to have many associated causes and pathological associations with no clear direct cures in sight. Despite the simplistic explanation that metabolic syndrome can be avoided by affecting energy balance between dietary caloric intake and energy expenditure, the obesity epidemic continues to grow. Furthermore, the few individuals who are successful in combating the disease rarely maintain their achievements over the long term.

In this dissertation, I have investigated molecular mechanisms of this condition at multiple levels of complexity. I have simulated the atomic biophysical properties of a single residue in a key protein at the center of mitochondrial metabolism, attempted to reverse engineer the central developmental pathways that regulate metabolism of energy in muscle and fat cells, and even attempted to understand the whole-body systemic regulation of energy utilization during times of fast and feast. These studies have taught me the true depth and breadth of how metabolic disorder manifests. It is fundamentally a condition that is influenced both “top down” and “bottom up”.

Analysis of the transcriptome of human skeletal muscle yielded profound insights into the nature of skeletal muscle insulin resistance. I developed a novel methodology of analyzing gene expression data, wherein expression values are directly correlated with metabolic phenotypes. This analysis revealed several pathways that have coherent underlying themes. In particular, genes correlating positively with glucose uptake are associated with the wnt signaling pathway. This pathway is now demonstrated to centrally regulate lipogenesis and glucose homeostasis in the cell. Genetic evidence from

type 2 diabetes GWAS studies confirm this association with insulin resistance, as core signaling elements in the wnt pathway are susceptibility loci.

In insulin resistance, several genes from the muscle atrophy pathway were found as strong correlates. In conditions of muscle atrophy, enhanced proteolysis from the ubiquitin proteosome system tears down cellular proteins. In the case of starvation, this mechanism provides amino acids which can be used for hepatic gluconeogenesis in times of restricted dietary glucose. During starvation, it is thus beneficial to induce a state of insulin resistance in peripheral tissues, as in these conditions glucose is scarce and must be reserved for the brain. When starving, the muscle must simultaneously stop taking in glucose, and it must start exporting amino acids. Thus, proteolysis and insulin resistance may be directly mechanistically connected. I propose that the primary insulin resistance *off switch* is the degradation of an unidentified central protein involved in glucose transporter trafficking. A clue to this is that the top correlating gene GOPC binds to proteins using the same PDZ motif which is targeted by Atrogin-1 for proteolysis. This could indicate that either GOPC, or an interacting partner of GOPC, is the target. Frizzled receptors share the common PDZ motif, and have been shown to be trafficked to the cell membrane by GOPC. Thus, Atrogin-mediated proteolysis of frizzled receptors may induce insulin resistance by eliminating wnt signaling.

During muscle atrophy, protein synthesis is also dramatically inhibited. This is partially mediated via targeted ubiquitination of a key translation initiation factor by the atrophy-inducing gene atrogin-1. This implies that protein turnover is dramatically reduced in insulin resistant subjects, and I propose that this phenomenon may explain the changes in acetylation that we observed in mitochondrial proteomes after exercise.

Insulin resistant individuals can less efficiently replace acetylated and defective enzymes with newly-synthesized proteins. This hypothesis can be tested in human subjects by unifying proteomic and transcriptomic data to ascertain whether or not the reductions in mitochondrial acetylation are associated with the expression levels of atrophy related genes.

The wnt pathway appears to be the primary activating signal for insulin stimulated glucose uptake in the cell, and this pathway also suppresses fat accumulation. The muscle atrophy pathway is activated during fasting, and appears to be the primary instigator of insulin resistance. This connection makes sense at the level of the whole organism. The atrophy pathway can be activated by other confounding factors however, namely inflammation of the tissue. Exposure to free fatty acids and residence of proinflammatory macrophages can lead to activation of the NFKB pathway via TNFa signaling, and this directly activates muscle proteolysis. Thus, in the diabetic muscle, enhanced proteolysis may be due to inflammation as opposed to starvation as the primary causal factor. The downstream effector proteins are the same, however. Thus if a mechanistic insulin *off switch* can be found that is mediated via proteolysis, then the state of insulin resistance can be inadvertently activated by inflammation.

The importance of proinflammatory macrophages in inducing skeletal muscle insulin resistance is hinted by a strong increase of the TRAP gene with exercise. Insulin sensitive subjects express TRAP to a larger degree after exercise. Prior literature indicates that TRAP has a role in modulating inflammation. TRAP dephosphorylates osteopontin, which is a critical adhesion molecule that attracts and polarizes macrophages into an M1 pro-inflammatory state. Thus, it could be that the increased expression of TRAP in

insulin sensitive subjects reduces macrophage-induced inflammation, and that this reduces NFKB signaling and subsequent activation of muscle proteolysis pathways. This connection could be a direct "top down" pathway whereby inflammation is viewed as the upstream cause of the molecular mechanisms of insulin resistance. The work ahead involves coming to a deeper understanding of why macrophage recruitment and activation is higher in insulin resistant muscle. Why does the inflammation start?

During this journey, I have also built new tools and skills. I designed and developed a web-available informatics resource that attempts to "Amass" sequence annotations from multiple sources and map it onto 3D protein structures. This system will enable researchers to discover and visualize connections between genetic variants, post-translational modifications, and annotated functional sites in proteins and complexes. This resource is developed around the common theme of this dissertation: unifying biological evidence from all available sources to find unexpected connections.

With the studies I have performed, it is now clear to me more than ever that metabolic syndrome is a complicated condition with many distinct causes. Thinking in a systemic fashion, and learning everything I can about the relevant signaling and metabolic pathways, I have formulated ideas of how metabolic syndrome can slowly develop. Although this short list of findings only inches us closer to a cure, the overall theme of systemic thinking will guide my future research career. I will continue to find connections, both by manual investigation of the literature and by developing new software and databases to aggregate data and "connect the dots". Overall, my goal is to find the underlying causes of these metabolic conditions.

With personalized medicine rapidly approaching on the horizon, different interventions will be needed for each individual, as I believe it is likely that everyone develops these illnesses for uniquely distinct reasons. No longer can we expect to help people by treating the symptoms. We must find each distinct cause of the disease. With this systemic knowledge, we can more effectively drop the weight of the obesity epidemic by treating patients with personalized therapies tailored to their genotype.

REFERENCES

1. Barma, P., et al., *Lipid induced overexpression of NF-kappaB in skeletal muscle cells is linked to insulin resistance.* Biochim Biophys Acta, 2009. **1792**(3): p. 190-200.

2. De Filippis, E., et al., *Insulin-resistant muscle is exercise resistant: evidence for reduced response of nuclear-encoded mitochondrial genes to exercise.* Am J Physiol Endocrinol Metab, 2008. **294**(3): p. E607-14.

3. Gomes, M.D., et al., *Atrogin-1, a muscle-specific F-box protein highly expressed during muscle atrophy.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14440-5.

4. Hwang, H., et al., *Proteomics analysis of human skeletal muscle reveals novel abnormalities in obesity and type 2 diabetes.* Diabetes, 2010. **59**(1): p. 33-42.

5. Langlais, P., et al., *Global IRS-1 phosphorylation analysis in insulin resistance.* Diabetologia, 2011. **54**(11): p. 2878-89.

6. Yi, Z., et al., *Global assessment of regulation of phosphorylation of insulin receptor substrate-1 by insulin in vivo in human muscle.* Diabetes, 2007. **56**(6): p. 1508-16.

7. Cusi, K., et al., *Insulin resistance differentially affects the PI 3-kinase- and MAP kinase-mediated signaling in human muscle.* J Clin Invest, 2000. **105**(3): p. 311-20.

8. Zhang, J., et al., *Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli.* Mol Cell Proteomics, 2009. **8**(2): p. 215-25.

9. Zhao, S., et al., *Regulation of cellular metabolism by protein lysine acetylation.* Science, 2010. **327**(5968): p. 1000-4.

10. Nelson, D.R., et al., *Site-directed mutagenesis of the yeast mitochondrial ADP/ATP translocator. Six arginines and one lysine are essential.* J Mol Biol, 1993. **230**(4): p. 1159-70.

11. *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Res, 2012. **40**(Database issue): p. D71-5.

12. Forbes, S.A., et al., *COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer.* Nucleic Acids Res, 2011. **39**(Database issue): p. D945-50.

13. Bamford, S., et al., *The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website.* Br J Cancer, 2004. **91**(2): p. 355-8.

14. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.

15. *Jmol: an open-source Java viewer for chemical structures in 3D.* *http://www.jmol.org/*.

16. Thong, F.S., C.B. Dugani, and A. Klip, *Turning signals on and off: GLUT4 traffic in the insulin-signaling highway.* Physiology (Bethesda), 2005. **20**: p. 271-84.

17. *Python Reference Manual, PythonLabs, Virginia, USA, 2001. Available at* *http://www.python.org*.

18. Logan, C.Y. and R. Nusse, *The Wnt signaling pathway in development and disease.* Annu Rev Cell Dev Biol, 2004. **20**: p. 781-810.

19. Jin, T., *The WNT signalling pathway and diabetes mellitus.* Diabetologia, 2008. **51**(10): p. 1771-80.

20. Jin, T. and L. Liu, *The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus.* Mol Endocrinol, 2008. **22**(11): p. 2383-92.

21. Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.* Nat Genet, 2010. **42**(7): p. 579-89.

22. Nobrega, M.A., *TCF7L2 and glucose metabolism: time to look beyond the pancreas.* Diabetes, 2013. **62**(3): p. 706-8.

23. Abiola, M., et al., *Activation of Wnt/beta-catenin signaling increases insulin sensitivity through a reciprocal regulation of Wnt10b and SREBP-1c in skeletal muscle cells.* PLoS One, 2009. **4**(12): p. e8509.

24. Yoon, J.C., et al., *Wnt signaling regulates mitochondrial physiology and insulin sensitivity.* Genes Dev, 2010. **24**(14): p. 1507-18.

25. Chen, T., et al., *Identification of zinc-finger BED domain-containing 3 (Zbed3) as a novel Axin-interacting protein that activates Wnt/beta-catenin signaling.* J Biol Chem, 2009. **284**(11): p. 6683-9.

26. Neudauer, C.L., G. Joberty, and I.G. Macara, *PIST: a novel PDZ/coiled-coil domain binding partner for the rho-family GTPase TC10.* Biochem Biophys Res Commun, 2001. **280**(2): p. 541-7.

27. Bogan, J.S., et al., *Endoproteolytic cleavage of TUG protein regulates GLUT4 glucose transporter translocation.* J Biol Chem, 2012. **287**(28): p. 23932-47.

28. Bogan, J.S., *Regulation of glucose transporter translocation in health and diabetes.* Annu Rev Biochem, 2012. **81**: p. 507-32.

29. Yu, C., et al., *The glucose transporter 4-regulating protein TUG is essential for highly insulin-responsive glucose uptake in 3T3-L1 adipocytes.* J Biol Chem, 2007. **282**(10): p. 7710-22.

30. Chang, L., S.H. Chiang, and A.R. Saltiel, *TC10alpha is required for insulin-stimulated glucose uptake in adipocytes.* Endocrinology, 2007. **148**(1): p. 27-33.

31. Yao, R., et al., *Identification of a PDZ domain containing Golgi protein, GOPC, as an interaction partner of frizzled.* Biochem Biophys Res Commun, 2001. **286**(4): p. 771-8.

32. Zhang, W., et al., *PR55 alpha, a regulatory subunit of PP2A, specifically regulates PP2A-mediated beta-catenin dephosphorylation.* J Biol Chem, 2009. **284**(34): p. 22649-56.

33. Creyghton, M.P., et al., *PR72, a novel regulator of Wnt signaling required for Naked cuticle function.* Genes Dev, 2005. **19**(3): p. 376-86.

34. Creyghton, M.P., et al., *PR130 is a modulator of the Wnt-signaling cascade that counters repression of the antagonist Naked cuticle.* Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5397-402.

35. Guasconi, V. and P.L. Puri, *Epigenetic drugs in the treatment of skeletal muscle atrophy.* Curr Opin Clin Nutr Metab Care, 2008. **11**(3): p. 233-41.

36. van Wijk, S.J. and H.T. Timmers, *The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins.* FASEB J, 2010. **24**(4): p. 981-93.

37. Ito, K., et al., *cDNA cloning, characterization, and chromosome mapping of UBE2E3 (alias UbcH9), encoding an N-terminally extended human ubiquitin-conjugating enzyme.* Cytogenet Cell Genet, 1999. **84**(1-2): p. 99-104.

38. Yang, X., et al., *Microarray profiling of skeletal muscle tissues from equally obese, non-diabetic insulin-sensitive and insulin-resistant Pima Indians.* Diabetologia, 2002. **45**(11): p. 1584-93.

39. Fernandez-Saiz, V., et al., *SCFFbxo9 and CK2 direct the cellular response to growth factor withdrawal via Tel2/Tti1 degradation and promote survival in multiple myeloma.* Nat Cell Biol, 2013. **15**(1): p. 72-81.

40. Saltiel, A.R. and C.R. Kahn, *Insulin signalling and the regulation of glucose and lipid metabolism.* Nature, 2001. **414**(6865): p. 799-806.

41. Lee, K.W., et al., *F-box only protein 9 is required for adipocyte differentiation.* Biochem Biophys Res Commun, 2013. **435**(2): p. 239-43.

42. Lagirand-Cantaloube, J., et al., *The initiation factor eIF3-f is a major target for atrogin1/MAFbx function in skeletal muscle atrophy.* EMBO J, 2008. **27**(8): p. 1266-76.

43. Hunter, R.B. and S.C. Kandarian, *Disruption of either the Nfkb1 or the Bcl3 gene inhibits skeletal muscle atrophy.* J Clin Invest, 2004. **114**(10): p. 1504-11.

44. Reed, S.A., et al., *Inhibition of IkappaB kinase alpha (IKKalpha) or IKKbeta (IKKbeta) plus forkhead box O (Foxo) abolishes skeletal muscle atrophy.* Biochem Biophys Res Commun, 2011. **405**(3): p. 491-6.

45. Wu, C.L., S.C. Kandarian, and R.W. Jackman, *Identification of genes that elicit disuse muscle atrophy via the transcription factors p50 and Bcl-3.* PLoS One, 2011. **6**(1): p. e16171.

46. Tan, X., et al., *PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: quantitative analysis by competitive PCR.* Biochim Biophys Acta, 1994. **1215**(1-2): p. 157-62.

47. Ten, R.M., et al., *The characterization of the promoter of the gene encoding the p50 subunit of NF-kappa B indicates that it participates in its own regulation.* EMBO J, 1992. **11**(1): p. 195-203.

48. Hotamisligil, G.S. and B.M. Spiegelman, *Tumor necrosis factor alpha: a key component of the obesity-diabetes link.* Diabetes, 1994. **43**(11): p. 1271-8.

49. Moller, D.E., *Potential role of TNF-alpha in the pathogenesis of insulin resistance and type 2 diabetes.* Trends Endocrinol Metab, 2000. **11**(6): p. 212-7.

50. Yao, T., et al., *Proteasome recruitment and activation of the Uch37 deubiquitinating enzyme by Adrm1.* Nat Cell Biol, 2006. **8**(9): p. 994-1002.

51. Qiu, X.B., et al., *hRpn13/ADRM1/GP110 is a novel proteasome subunit that binds the deubiquitinating enzyme, UCH37.* EMBO J, 2006. **25**(24): p. 5742-53.

52. Brand, M.D. and T.C. Esteves, *Physiological functions of the mitochondrial uncoupling proteins UCP2 and UCP3.* Cell Metab, 2005. **2**(2): p. 85-93.

53. Jia, J.J., et al., *The polymorphisms of UCP2 and UCP3 genes associated with fat metabolism, obesity and diabetes.* Obes Rev, 2009. **10**(5): p. 519-26.

54. Liu, J., et al., *The role of uncoupling proteins in diabetes mellitus.* J Diabetes Res, 2013. **2013**: p. 585897.

55. Senese, R., et al., *Uncoupling protein 3 expression levels influence insulin sensitivity, fatty acid oxidation, and related signaling pathways.* Pflugers Arch, 2011. **461**(1): p. 153-64.

56. Liu, J., et al., *Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy.* Nat Genet, 1998. **20**(1): p. 31-6.

57. Confalonieri, P., et al., *Muscle inflammation and MHC class I up-regulation in muscular dystrophy with lack of dysferlin: an immunopathological study.* J Neuroimmunol, 2003. **142**(1-2): p. 130-6.

58. Tomas, M., et al., *Genetic variation in the KCNMA1 potassium channel alpha subunit as risk factor for severe essential hypertension and myocardial infarction.* J Hypertens, 2008. **26**(11): p. 2147-53.

59. Senti, M., et al., *Protective effect of the KCNMB1 E65K genetic polymorphism against diastolic hypertension in aging women and its relevance to cardiovascular risk.* Circ Res, 2005. **97**(12): p. 1360-5.

60. Gollasch, M., et al., *The BK channel beta1 subunit gene is associated with human baroreflex and blood pressure regulation.* J Hypertens, 2002. **20**(5): p. 927-33.

61. Fernandez-Fernandez, J.M., et al., *Gain-of-function mutation in the KCNMB1 potassium channel subunit is associated with low prevalence of diastolic hypertension.* J Clin Invest, 2004. **113**(7): p. 1032-9.

62. Korovkina, V.P., et al., *Characterization of a novel 132-bp exon of the human maxi-K channel.* Am J Physiol Cell Physiol, 2001. **281**(1): p. C361-7.

63. Lu, T., et al., *Impaired arachidonic acid-mediated activation of large-conductance Ca2+-activated K+ channels in coronary arterial smooth muscle cells in Zucker Diabetic Fatty rats.* Diabetes, 2005. **54**(7): p. 2155-63.

64. Lu, T., et al., *Regulation of coronary arterial BK channels by caveolae-mediated angiotensin II signaling in diabetes mellitus.* Circ Res, 2010. **106**(6): p. 1164-73.

65. Zhang, D.M., et al., *Muscle-specific f-box only proteins facilitate bk channel beta(1) subunit downregulation in vascular smooth muscle cells of diabetes mellitus.* Circ Res, 2010. **107**(12): p. 1454-9.

66. Ek-Rylander, B., et al., *Dephosphorylation of osteopontin and bone sialoprotein by osteoclastic tartrate-resistant acid phosphatase. Modulation of osteoclast adhesion in vitro.* J Biol Chem, 1994. **269**(21): p. 14853-6.

67. Choi, S.T., et al., *Osteopontin might be involved in bone remodelling rather than in inflammation in ankylosing spondylitis.* Rheumatology (Oxford), 2008. **47**(12): p. 1775-9.

68. Lang, P., et al., *Monomeric tartrate resistant acid phosphatase induces insulin sensitive obesity.* PLoS One, 2008. **3**(3): p. e1713.

69. Olefsky, J.M. and C.K. Glass, *Macrophages, inflammation, and insulin resistance.* Annu Rev Physiol, 2010. **72**: p. 219-46.

70. Briggs, T.A., et al., *Tartrate-resistant acid phosphatase deficiency causes a bone dysplasia with autoimmunity and a type I interferon expression signature.* Nat Genet, 2011. **43**(2): p. 127-31.

71. Coyle, K., et al., *Metabolic abnormalities and cardiovascular risk factors in children with myositis.* J Pediatr, 2009. **155**(6): p. 882-7.

72. Lausch, E., et al., *Genetic deficiency of tartrate-resistant acid phosphatase associated with skeletal dysplasia, cerebral calcifications and autoimmunity.* Nat Genet, 2011. **43**(2): p. 132-7.

73. Uaesoontrachoon, K., et al., *Osteopontin and skeletal muscle myoblasts: association with muscle regeneration and regulation of myoblast function in vitro.* Int J Biochem Cell Biol, 2008. **40**(10): p. 2303-14.

74. Porter, J.D., et al., *A chronic inflammatory response dominates the skeletal muscle molecular signature in dystrophin-deficient mdx mice.* Hum Mol Genet, 2002. **11**(3): p. 263-72.

75. Chapman, J., et al., *Osteopontin is required for the early onset of high fat diet-induced insulin resistance in mice.* PLoS One, 2010. **5**(11): p. e13959.

76. Kiefer, F.W., et al., *Neutralization of osteopontin inhibits obesity-induced inflammation and insulin resistance.* Diabetes, 2010. **59**(4): p. 935-46.

77. Weber, G.F., et al., *Phosphorylation-dependent interaction of osteopontin with its receptors regulates macrophage migration and activation.* J Leukoc Biol, 2002. **72**(4): p. 752-61.

78. Zhou, G., et al., *Role of AMP-activated protein kinase in mechanism of metformin action.* J Clin Invest, 2001. **108**(8): p. 1167-74.

79. Winder, W.W., *Energy-sensing and signaling by AMP-activated protein kinase in skeletal muscle.* J Appl Physiol, 2001. **91**(3): p. 1017-28.

80. Palacios, O.M., et al., *Diet and exercise signals regulate SIRT3 and activate AMPK and PGC-1alpha in skeletal muscle.* Aging (Albany NY), 2009. **1**(9): p. 771-83.

81. Ojuka, E.O., et al., *Regulation of GLUT4 biogenesis in muscle: evidence for involvement of AMPK and Ca(2+).* Am J Physiol Endocrinol Metab, 2002. **282**(5): p. E1008-13.

82. Lee, W.J., et al., *AMPK activation increases fatty acid oxidation in skeletal muscle by activating PPARalpha and PGC-1.* Biochem Biophys Res Commun, 2006. **340**(1): p. 291-5.

83. Holmes, B.F., E.J. Kurth-Kraczek, and W.W. Winder, *Chronic activation of 5'-AMP-activated protein kinase increases GLUT-4, hexokinase, and glycogen in muscle.* J Appl Physiol, 1999. **87**(5): p. 1990-5.

84. Cheung, P.C., et al., *Characterization of AMP-activated protein kinase gamma-subunit isoforms and their role in AMP binding.* Biochem J, 2000. **346 Pt 3**: p. 659-69.

85. Zhu, H., et al., *The Lin28/let-7 axis regulates glucose metabolism.* Cell, 2011. **147**(1): p. 81-94.

86. Atsumi, T., et al., *Expression of inducible 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase/PFKFB3 isoforms in adipocytes and their potential role in glycolytic regulation.* Diabetes, 2005. **54**(12): p. 3349-57.

87. Guo, X., et al., *Involvement of inducible 6-phosphofructo-2-kinase in the anti-diabetic effect of peroxisome proliferator-activated receptor gamma activation in mice.* J Biol Chem, 2010. **285**(31): p. 23711-20.

88. Kahn, S.E., R.L. Hull, and K.M. Utzschneider, *Mechanisms linking obesity to insulin resistance and type 2 diabetes.* Nature, 2006. **444**(7121): p. 840-6.

89. Chibalin, A.V., et al., *Downregulation of diacylglycerol kinase delta contributes to hyperglycemia-induced insulin resistance.* Cell, 2008. **132**(3): p. 375-86.

90. Speliotes, E.K., et al., *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index.* Nat Genet, 2010. **42**(11): p. 937-48.

91. Haas, J., et al., *LRP1b shows restricted expression in human tissues and binds to several extracellular ligands, including fibrinogen and apoE-carrying lipoproteins.* Atherosclerosis, 2011. **216**(2): p. 342-7.

92. Jedrychowski, M.P., et al., *Proteomic analysis of GLUT4 storage vesicles reveals LRP1 to be an important vesicle component and target of insulin signaling.* J Biol Chem, 2010. **285**(1): p. 104-14.

93. Zilberberg, A., A. Yaniv, and A. Gazit, *The low density lipoprotein receptor-1, LRP1, interacts with the human frizzled-1 (HFz1) and down-regulates the canonical Wnt signaling pathway.* J Biol Chem, 2004. **279**(17): p. 17535-42.

94. Terrand, J., et al., *LRP1 controls intracellular cholesterol storage and fatty acid synthesis through modulation of Wnt signaling.* J Biol Chem, 2009. **284**(1): p. 381-8.

95. Costales, P., et al., *Selective role of sterol regulatory element binding protein isoforms in aggregated LDL-induced vascular low density lipoprotein receptor-related protein-1 expression.* Atherosclerosis, 2010. **213**(2): p. 458-68.

96. Stannard, S.R. and N.A. Johnson, *Insulin resistance and elevated triglyceride in muscle: more important for survival than "thrifty" genes?* J Physiol, 2004. **554**(Pt 3): p. 595-607.

97. Boden, G., *Free fatty acids (FFA), a link between obesity and insulin resistance.* Front Biosci, 1998. **3**: p. d169-75.

98. Goodyear, L.J. and B.B. Kahn, *Exercise, glucose transport, and insulin sensitivity.* Annu Rev Med, 1998. **49**: p. 235-61.

99. He, J., et al., *Interaction with cystic fibrosis transmembrane conductance regulator-associated ligand (CAL) inhibits beta1-adrenergic receptor surface expression.* J Biol Chem, 2004. **279**(48): p. 50190-6.

100. Varma, V., et al., *Muscle inflammatory response and insulin resistance: synergistic interaction between macrophages and fatty acids leads to impaired insulin action.* Am J Physiol Endocrinol Metab, 2009. **296**(6): p. E1300-10.

101. Choudhary, C., et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, 2009. **325**(5942): p. 834-40.

102. Wang, Q., et al., *Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux.* Science, 2010. **327**(5968): p. 1004-7.

103. Hirschey, M.D., et al., *SIRT3 regulates mitochondrial fatty-acid oxidation by reversible enzyme deacetylation.* Nature, 2010. **464**(7285): p. 121-5.

104. Gerhart-Hines, Z., et al., *Metabolic control of muscle mitochondrial function and fatty acid oxidation through SIRT1/PGC-1alpha.* EMBO J, 2007. **26**(7): p. 1913-23.

105. Lomb, D.J., G. Laurent, and M.C. Haigis, *Sirtuins regulate key aspects of lipid metabolism.* Biochim Biophys Acta, 2010. **1804**(8): p. 1652-7.

106. Verdin, E., et al., *Sirtuin regulation of mitochondria: energy production, apoptosis, and signaling.* Trends Biochem Sci, 2010.

107. Lefort, N., et al., *Proteome profile of functional mitochondria from human skeletal muscle using one-dimensional gel electrophoresis and HPLC-ESI-MS/MS.* J Proteomics, 2009. **72**(6): p. 1046-60.

108. Christ-Roberts, C.Y. and L.J. Mandarino, *Glycogen synthase: key effect of exercise on insulin action.* Exerc Sport Sci Rev, 2004. **32**(3): p. 90-4.

109. Lefort, N., et al., *Increased reactive oxygen species production and lower abundance of complex I subunits and carnitine palmitoyltransferase 1B protein despite normal mitochondrial respiration in insulin-resistant human skeletal muscle.* Diabetes, 2010. **59**(10): p. 2444-52.

110. Pebay-Peyroula, E., et al., *Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside.* Nature, 2003. **426**(6962): p. 39-44.

111. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models, and associated resources.* Nucleic Acids Res, 2011. **39**(Database issue): p. D465-74.

112. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

113. Roberts, E., et al., *MultiSeq: unifying sequence and structure data for evolutionary analysis.* BMC Bioinformatics, 2006. **7**: p. 382.

114. Russell, R.B. and G.J. Barton, *Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels.* Proteins, 1992. **14**(2): p. 309-23.

115. Eichenbaum, K.D., et al., *The energetics of the acetylation switch in p53-mediated transcriptional activation.* Proteins, 2010. **78**(2): p. 447-56.

116. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

117. MacKerell, A.D., Jr., N. Banavali, and N. Foloppe, *Development and current status of the CHARMM force field for nucleic acids.* Biopolymers, 2000. **56**(4): p. 257-65.

118. Aksimentiev, A. and K. Schulten, *Imaging alpha-hemolysin with molecular dynamics: ionic conductance, osmotic permeability, and the electrostatic potential map.* Biophys J, 2005. **88**(6): p. 3745-61.

119. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-91.

120. Jeneson, J.A., et al., *Magnitude and control of mitochondrial sensitivity to ADP.* American journal of physiology. Endocrinology and metabolism, 2009. **297**(3): p. E774-84.

121. Petersen, K.F., et al., *Mitochondrial dysfunction in the elderly: possible role in insulin resistance.* Science, 2003. **300**(5622): p. 1140-2.

122. Holloszy, J.O., *Skeletal muscle "mitochondrial deficiency" does not mediate insulin resistance.* Am J Clin Nutr, 2009. **89**(1): p. 463S-6S.

123. Lowell, B.B. and G.I. Shulman, *Mitochondrial dysfunction and type 2 diabetes.* Science, 2005. **307**(5708): p. 384-7.

124. DeFronzo, R.A., J.D. Tobin, and R. Andres, *Glucose clamp technique: a method for quantifying insulin secretion and resistance.* Am J Physiol, 1979. **237**(3): p. E214-23.

125. Kelley, D.E. and L.J. Mandarino, *Fuel selection in human skeletal muscle in insulin resistance: a reexamination.* Diabetes, 2000. **49**(5): p. 677-83.

126. Ravaud, S., et al., *Impaired transport of nucleotides in a mitochondrial carrier explains severe human genetic diseases.* ACS Chem Biol, 2012. **7**(7): p. 1164-9.

127. Clemencon, B., et al., *Yeast ADP/ATP carrier isoform 2: conformational dynamics and role of the RRRMMM signature sequence methionines.* J Biol Chem, 2011. **286**(41): p. 36119-31.

128. Wang, Y. and E. Tajkhorshid, *Electrostatic funneling of substrate in mitochondrial inner membrane carriers.* Proc Natl Acad Sci U S A, 2008. **105**(28): p. 9598-603.

129. Dehez, F., E. Pebay-Peyroula, and C. Chipot, *Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel.* J Am Chem Soc, 2008. **130**(38): p. 12725-33.

130. Di Marino, D., et al., *Mapping multiple potential ATP binding sites on the matrix side of the bovine ADP/ATP carrier by the combined use of MD simulation and docking.* J Mol Model, 2012. **18**(6): p. 2377-86.

131. Cooper, H.M. and J.N. Spelbrink, *The human SIRT3 protein deacetylase is exclusively mitochondrial.* Biochem J, 2008. **411**(2): p. 279-85.

132. Hallows, W.C., B.N. Albaugh, and J.M. Denu, *Where in the cell is SIRT3?--functional localization of an NAD+-dependent protein deacetylase.* Biochem J, 2008. **411**(2): p. e11-3.

133. Nakagawa, T., et al., *SIRT5 Deacetylates carbamoyl phosphate synthetase 1 and regulates the urea cycle.* Cell, 2009. **137**(3): p. 560-70.

134. Chance, B., et al., *Control of oxidative metabolism and oxygen delivery in human skeletal muscle: a steady-state analysis of the work/energy cost transfer function.* Proc Natl Acad Sci U S A, 1985. **82**(24): p. 8384-8.

135. Jeneson, J.A., et al., *The signal transduction function for oxidative phosphorylation is at least second order in ADP.* J Biol Chem, 1996. **271**(45): p. 27995-8.

136. Andersen, P. and B. Saltin, *Maximal perfusion of skeletal muscle in man.* J Physiol, 1985. **366**: p. 233-49.

137. Zurlo, F., et al., *Skeletal muscle metabolism is a major determinant of resting energy expenditure.* J Clin Invest, 1990. **86**(5): p. 1423-7.

138. Willis, W.T. and P.R. Dallman, *Impaired control of respiration in iron-deficient muscle mitochondria.* Am J Physiol, 1989. **257**(6 Pt 1): p. C1080-5.

139. Gouet, P., et al., *ESPript: analysis of multiple sequence alignments in PostScript.* Bioinformatics, 1999. **15**(4): p. 305-8.

140. Burley, S.K., et al., *Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers.* Structure, 2008. **16**(1): p. 5-11.

141. Yue, P., E. Melamud, and J. Moult, *SNPs3D: candidate gene and SNP selection for association studies.* BMC Bioinformatics, 2006. **7**: p. 166.

142. De Baets, G., et al., *SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants.* Nucleic Acids Res, 2012. **40**(Database issue): p. D935-9.

143. Adzhubei, I., D.M. Jordan, and S.R. Sunyaev, *Predicting functional effect of human missense mutations using PolyPhen-2.* Curr Protoc Hum Genet, 2013. **Chapter 7**: p. Unit7 20.

144. Luu, T.D., et al., *MSV3d: database of human MisSense Variants mapped to 3D protein structure.* Database (Oxford), 2012. **2012**: p. bas018.

145. Dantzer, J., et al., *MutDB services: interactive structural analysis of mutation data.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W311-4.

146. Mooney, S.D. and R.B. Altman, *MutDB: annotating human variation with functionally relevant data.* Bioinformatics, 2003. **19**(14): p. 1858-60.

147. Singh, A., et al., *MutDB: update on development of tools for the biochemical analysis of genetic variation.* Nucleic Acids Res, 2008. **36**(Database issue): p. D815-9.

148. Diella, F., et al., *Phospho.ELM: a database of phosphorylation sites--update 2008.* Nucleic Acids Res, 2008. **36**(Database issue): p. D240-4.

149. Beltrao, P., et al., *Systematic functional prioritization of protein posttranslational modifications.* Cell, 2012. **150**(2): p. 413-25.

150. Martin, A.C., *Mapping PDB chains to UniProtKB entries.* Bioinformatics, 2005. **21**(23): p. 4297-301.

151. Velankar, S., et al., *SIFTS: Structure Integration with Function, Taxonomy and Sequences resource.* Nucleic Acids Res, 2013. **41**(Database issue): p. D483-9.

152. Bakan, A., L.M. Meireles, and I. Bahar, *ProDy: protein dynamics inferred from theory and experiments.* Bioinformatics, 2011. **27**(11): p. 1575-7.

153. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.* Mol Syst Biol, 2011. **7**: p. 539.

154. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

155. Cock, P.J., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics.* Bioinformatics, 2009. **25**(11): p. 1422-3.

156. *MySQL* - *www.mysql.com*.

157. Fielding, R.T. and G. Kaiser, *The Apache HTTP Server Project.* Internet Computing, IEEE, 1997. **1**(4): p. 88-90.

158. *www.datatables.net*.

159. *www.flotcharts.org*.

160. Kato, M., et al., *Structural basis for inactivation of the human pyruvate dehydrogenase complex by phosphorylation: role of disordered phosphorylation loops.* Structure, 2008. **16**(12): p. 1849-59.

161. Hojlund, K., et al., *Proteome analysis reveals phosphorylation of ATP synthase beta -subunit in human skeletal muscle and proteins with potential roles in type 2 diabetes.* J Biol Chem, 2003. **278**(12): p. 10436-42.

162. Vilaseca, M.A., et al., *Phenotype and genotype heterogeneity in Mediterranean citrullinemia.* Mol Genet Metab, 2001. **74**(3): p. 396-8.

163. Kobayashi, K., et al., *Heterogeneity of mutations in argininosuccinate synthetase causing human citrullinemia.* J Biol Chem, 1990. **265**(19): p. 11361-7.

164. Maruta, H., K. Greer, and J.L. Rosenbaum, *The acetylation of alpha-tubulin and its relationship to the assembly and disassembly of microtubules.* J Cell Biol, 1986. **103**(2): p. 571-9.

165. Collins, F.S. and A.D. Barker, *Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies.* Sci Am, 2007. **296**(3): p. 50-7.