UNIVERSITY OF CALIFORNIA SAN DIEGO

Functional Characterization of Bioactive Lipids and Microproteins

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biology

by

Tina Chia-Tien Chang

Committee in charge:

Professor Alan Saghatelian, Chair
Professor Edward Dennis
Professor Andres Leschziner
Professor Gerald Shadel
Professor Dionicio Siegel

2022

The Dissertation of Tina Chia-Tien Chang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMA

LIST OF TABLES

LIST OF SUPPLEMENTARY FILES

Chang_01_Supplemental_Information1.doc. General information on 13-LAHLA synthesis and NMR spectra for Chapter 2.

Chang_02_ Supplemental_Information2.doc. General information on synthesis and general synthetic procedures for Chapter 3.

Chang_03_ Supplemental_Information3.doc. Supplementary figures for Chapter 4.

Chang_04_Supplementary Dataset1.xls. Mass spectrometry analysis of IgG and SLC35A4-MP eluted proteins for Chapter 4.

Chang_05_Supplementary Dataset2.xls. Conserved uORFs found in mouse and human transcriptomes that may encode novel functional microproteins for Chapter 4.

Chang_06_Supplementary Table1.xls. Antibodies and dilutions used for Chapter 4

LIST OF ABBREVIATIONS

3D	3-dimensional
3DEM	3-dimensional electron microscopy
ADTRP	androgen-dependent TFPI regulating protein
AG4OX	adipose-specific GLUT4 overexpressor
AIG1	androgen-induced gene 1
CJ	crista/cristae junction
CRISPR	clustered regularly interspaced short palindromic repeats
DHAHLA	docosahexaenoic acid ester of hydroxy linoleic acid
DNA	deoxyribonucleic acid
EM	electron microscopy
ER	endoplasmic reticulum
ERGIC	endoplasmic-reticulum-Golgi intermediate compartment
FAHFA	fatty acid ester of hydroxy fatty acid
FAHFA-TG	fatty acid hydroxy ester of hydroxy fatty acid-containing triglycerol
FCCP	Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
HEK	human embryonic kidney
HLA	hydroxy linoleic acid
HSA	hydroxy stearic acid
IBM	inner boundary mitochondrial membrane
IMM	inner mitochondrial membrane
IP	immunoprecipitation
IP-MS	immunoprecipitation-mass spectrometry
LA	linoleic acid
LAHLA	linoleic acid ester of hydroxy linoleic acid
LC-MS	liquid chromatography-mass spectrometry
MIC	MICOS, or mitochondrial contact site and cristae organizing system
MICOS	mitochondrial contact site and cristae organizing system
Mito	mitochondrial
MM	mitochondrial membrane
mORF	main open reading frame
uORF	upstream open reading frame
MP	microprotein
MPI	microprotein-protein interaction(s)
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
OA	oleic acid
OAHSA	oleic acid ester of hydroxy stearic acid
OCR	oxygen consumption rate
OMM	outer mitochondrial membrane
ORF	open reading frame
OXPHOS	oxidative phosphorylation

PA	palmitic acid
PAGE	polyacrylamide gel electrophoresis
PAHSA	palmitic acid ester of hydroxy stearic acid
PAHSA	palmitic acid ester of hydroxy stearic acid
PBS	phosphate buffered saline
PPI	protein-protein interaction(s)
PVDF	polyvinylidene difluoride
Ribo-Seq	ribosome profiling
RNA	ribonucleic acid
SAR	structure-activity relationship
SDS	sodium dodecyl sulfate
SEP	small open reading frame-encoded polypeptide
SLC35A4	solute carrier family 35 member A4
smORF	small open reading frame
TBST	tris-buffered saline with Tween-20
TCEP	tris(2-carboxyethyl)phosphine
TEAB	tetraethylammonium tetra hydroborate
UDP	uridine diphosphate glucose
UTR	untranslated region

ACKNOWLEDGEMENTS

During my academic training, it has been my deepest honor to work with Dr. Alan Saghatelian at the Salk Institute for Biological Studies. Not only is Alan a brilliant scientist, but he has also been an incredible mentor to me since the day I joined his lab as a technician. His kind nature has enabled me to learn science and challenge myself without restrictions. Over the years, the amount of guidance and support I received from Alan has made my graduate training an extraordinarily fulfilling experience, and I wouldn't trade it for the world.

I am also grateful to have worked closely with some of the brightest scientists in the lab, including Dr. Qian Chu and Dr. Thomas Martínez. Your mentorship has deeply and positively impacted the way I approach scientific problems. I cannot be more thankful for the time and patience you invested to help me grow as a scientist.

I would also like to thank my collaborators Dr. Dionicio Siegel and Dr. Mark Ellisman for playing an instrumental role in my doctoral training. Your expertise and advice are inspiring and have propelled my learning and research projects. Additionally, I am very fortunate to have Dr. Edward Dennis, Dr. Andres Leschziner, Dr. Gerald Shadel, and Dr. Dionicio Siegel serve on my doctoral committee. I truly appreciate your valuable time and research advice.

Last but definitely not least, I want to give my special thanks to senior scientists Joan Vaughan and Cindy Donaldson, my dearest family members and friends. All of you have been my best cheerleaders at difficult times, and the completion of my training would not have been possible without you.

Chapter 2, in full, is a reprint of the material with minor modifications as it appears in Linoleic acid esters of hydroxy linoleic acids are anti-inflammatory lipids found in plants and mammals. *The Journal of biological chemistry* vol. 294,27 (2019). The dissertation author was

the third investigator and author of this paper. Other authors include Matthew J. Kolar, Srihari Konduri, Huijing Wang, Clare McNerlin, Lena Ohlsson, Magnus Härröd, Dionicio Siegel, and Alan Saghatelian.

Chapter 3, in full, has been submitted for publication of the material with minor modifications as it may appear in Endogenous Fatty Acid Esters of Hydroxy Fatty Acids, Starting Points for New Chemical Regulators of Inflammation. *European Journal of Medicinal Chemistry* (2022). The dissertation author was the second investigator and author of this paper. Other authors include Huijing Wang, Srihari Konduri, Jeremiah D. Momper, Yuxin Lin, Arnold Garcia, José Rizo, Alan Saghatelian, and Dionicio Siegel.

Chapter 4, in full, is being prepared for publication of the material with minor modifications as it may appear in The *SLC35A4* upstream open reading frame encodes a microprotein regulator of mitochondrial structure and function. The dissertation author was the first investigator and author of this paper. Other authors include Guy Perkins, Andrea Rocha, Joan M. Vaughan, Jiao Ma, Peter C. Gray, Jolene K. Diedrich, Mark H. Ellisman, and Alan Saghatelian.

VITA

2015 Bachelor of Science in Biochemistry, Chemistry, University of California San Diego

2022 Doctor of Philosophy in Biology, University of California San Diego

PUBLICATIONS

1. **Chang, T.,** Perkins, G. A., Chu, Q., Rocha A., Vaughan, J. M., Ma, J., Gray, P. C., Diedrich, J. K., Ellisman, M. H., Saghatelian, A. The *SLC35A4* upstream open reading frame encodes a microprotein regulator of mitochondrial structure and function. [In preparation]

2. Wang, H., **Chang, T.**, Konduri, S., Momper, J., Lin, Y., Garcia, A., Rizo, J., Saghatelian, A., Siegel, D. (2021). Endogenous Fatty Acid Esters of Hydroxy Fatty Acids, Starting Points for New Chemical Regulators of Inflammation. European Journal of Medicinal Chemistry. [Under revision]

3. Wang, H., **Chang, T.**, Konduri, S., Huang, J., Saghatelian, A., & Siegel, D. (2019). Synthesis of chemically edited derivatives of the endogenous regulator of inflammation 9-PAHSA. *The Journal of antibiotics*, *72*(6), 498–506.

4. Kolar, M. J., Konduri, S., **Chang, T.**, Wang, H., McNerlin, C., Ohlsson, L., Härröd, M., Siegel, D., & Saghatelian, A. (2019). Linoleic acid esters of hydroxy linoleic acids are anti- inflammatory lipids found in plants and mammals. *The Journal of biological chemistry*, *294*(27), 10698–10707.

5. Chu, Q., **Chang, T.**, & Saghatelian, A. (2019). Substrate-Selective Enzyme Inhibitors. *Trends in pharmacological sciences*, *40*(10), 716–718.

6. Wang, H., Kolar, M. J., **Chang, T**., Rizo, J., Konduri, S., McNerlin, C., Saghatelian, A., & Siegel, D. (2019). Stereochemistry of Linoleic Acid Esters of Hydroxy Linoleic Acids. *Organic letters*, *21*(19), 8080–8084.

7. Chu, Q., Martinez, T. F., Novak, S. W., Donaldson, C. J., Tan, D., Vaughan, J. M., **Chang, T.**, Diedrich, J. K., Andrade, L., Kim, A., Zhang, T., Manor, U., & Saghatelian, A. (2019). Regulation of the ER stress response by a mitochondrial microprotein. *Nature communications*, *10*(1), 4883.

8. Kolar, M. J., Nelson, A. T., **Chang, T.**, Ertunc, M. E., Christy, M. P., Ohlsson, L., Härröd, M., Kahn, B. B., Siegel, D., & Saghatelian, A. (2018). Faster Protocol for Endogenous Fatty Acid Esters of Hydroxy Fatty Acid (FAHFA) Measurements. *Analytical chemistry*, *90*(8), 5358– 5365.

9. Gonçalves, J. T., Bloyd, C. W., Shtrahman, M., Johnston, S. T., Schafer, S. T., Parylak, S. L., Tran, T., **Chang, T.**, & Gage, F. H. (2016). In vivo imaging of dendritic pruning in dentate granule cells. Nature neuroscience, 19(6), 788–791.

ABSTRACT OF THE DISSERTATION

Functional Characterization of Bioactive Lipids and Microproteins

by

Tina Chia-Tien Chang

Doctor of Philosophy in Biology

University of California San Diego, 2022

Professor Alan Saghatelian, Chair

Lipids and proteins are essential natural small molecules with vital roles in regulating nearly all biological processes, including stress adaptation, development and growth, metabolism, and regeneration. An intricate interplay among these molecules facilitates proper functioning of cells and their machinery critical to human health. A disruption in their homeostasis would likely alter pathways that mediate physiology, leading to deleterious human diseases. For decades, a lack of available detection and experimental tools prevented our

understanding of natural small molecules and their biology. It was not until recent technologies that led to groundbreaking discoveries of beneficial lipids called FAHFAs and protein-coding small open reading frames (smORFs). The identifications of these unprecedented small natural molecules highlight that the current capacities of mammalian lipidome and transcriptome have been underestimated. Thus, this dissertation will focus on our efforts to functionally characterize these newly identified molecules and their biological roles.

Recent discovery of a new bioactive lipid class called branched fatty acid esters of hydroxy fatty acids (FAHFAs) demonstrated great structural diversity, with some, such as palmitic acid 9-hydroxy stearic acid (9-PAHSA), exhibiting inflammation-reducing effects. While over 80 distinct FAHFA isomers have been reported, the biological activities among these lipid species are poorly defined. Because the PAHSAs were previously demonstrated to mediate inflammation in murine and human disease models, we investigated if other bioactive FAHFAs exist in nature. We present that oat oil is rich of linoleic acid esters of hydroxy linoleic acids (LAHLAs), a newly identified FAHFA family. Comparison of 13-LAHLA and 9-PAHSA's ability to attenuate proinflammatory cytokine release in mouse macrophages under acute stress demonstrates that 13-LAHLA was more active than 9-PAHSA, implying a structural-activity relationship (SAR) exists within the lipid class.

Characterization of FAHFAs has been challenging due to their susceptibility to endogenous metabolism. FAHFAs are not only subjected to degradation by FAHFA-specific lipases, such as CEL, AIG1 and ADTRP, they can also be incorporated to form FAHFA-containing triglycerols (FAHFA-TGs). These pathways reflect that FAHFAs are dynamically regulated and metabolism of FAHFAs likely diminish their beneficial effects in physiology. Therefore, we hypothesized that synthetic FAHFAs with greater stability could protect these

derivates from endogenous metabolism, thus prolonging favorable effects systemically. Syntheses and activity screening of a library of synthetic FAHFA derivatives nominated a number of selective potent regulators of inflammation. Subsequent testing of the lead compound in a mouse colitis model confirms its anti-inflammatory activity and pharmacokinetics profile, potentiating a new non-steroidal therapeutic option against inflammation for this small molecule.

In addition to FAHFAs, smORFs belong to a group of underexplored molecules with translational potential. For decades, annotation of eukaryotic smORFs was hindered because scientists believed that translation would only occur once per transcript and only at the longest ORF. In particular, upstream ORFs (uORFs) are found in half of the human transcripts, representing the largest type of smORFs. However, little is known about the biochemical and cellular functions of uORF-encoded microproteins. Here, we report that the uORF-derived SLC35A4 microprotein (SLC35A4-MP) is a *bona fide* member of the mitochondrial contact site and cristae organizing system (MICOS) complex, a protein complex essential for mitochondrial crista formation and mitochondrial respiration. Loss of SLC35A4-MP disrupts mitochondrial crista and crista junction integrity and impaired the optimal oxidative capacities of cells. This finding demonstrates that new functional members of vital protein complexes can be extracted from the abundant reservoir of uncharacterized uORF-encoded microprotein.

As additional discoveries of natural small molecules are underway, we believe that our data provides an incentive for their further evaluation. Through development of systematic approaches to interrogate the fundamental biology of lipids and proteins, we aim to provide a template on how one can functionally characterize natural small molecules to counteract the molecular basis of human diseases. Our goal is to gain deeper knowledge of the mammalian lipidome and transcriptome, ultimately leading our path to designing better treatments.

CHAPTER 1

Introduction

1.1 Discovery of FAHFAs

Historically, fat or lipids are considered as a culprit for poor health as some are increasingly viewed as major risk factor for human diseases, such as cancer, obesity, insulin resistance, and type 2 diabetes (T2D). Elevated circulating fatty acids are commonly associated with insulin resistance (*1*) while soaring levels of cholesterol and triglycerides are one of the major risk factors for developing atherosclerotic diseases (*2*). Therefore, fats are generally considered deleterious to human health. Over the years, accumulating research evidence linking increased *de novo* lipogenesis in adipose tissue and improved metabolic health have suggested a beneficial role for these natural small molecules, indicating that not all lipids are harmful (*3*–*6*). Dietary supplementation of omega-3 fatty acids (*6*), for instance, can enhanced metabolic parameters of high fat diet (HFD)-fed mice (*7*) and lowered risk of T2D in human male subjects (*8*). Additionally, adipose-specific overexpression of glucose transporter Glut4 drives the production of circulating fatty acids and reverses defining hallmarks of diabetes by mediating fasting hyperglycemia and glucose intolerance in mice (*9*, *10*). These findings led to the hypothesis that increased *de novo* lipogenesis might lead to elevated levels of beneficial lipids, such as palmitoleate (*6*) and omega-3 fatty acids (*7*).

Using mice with adipose-selective overexpression of Glut4 (AG4OX), this hypothesis was tested using quantitative mass spectrometry (MS) lipidomics to profile the lipidome of adipose tissues from wildtype and AG4OX mice (*11*). Analysis of adipose tissues of AG4OX mice, which exhibited augmented lipogenesis, reveals a structurally novel class of mammalian lipids characterized by a branched ester linkage between a fatty acid and a hydroxy fatty acid chain. These lipids were collectively referred to as branched fatty acid esters of hydroxy fatty acids (FAHFAs).

FAHFAs are structurally diverse. In total, targeted MS analysis identified 16 unique FAHFA family members in the serum of AG4OX mice. Here, the term "family" is defined as the distinct combination of a fatty acid chain and a hydroxy fatty acid chain (**Figure 1.1a-b**) (*11*). Regioisomers of FAHFAs are also possible as the position of the branched ester can be found on different carbons of the hydroxy fatty acid chain (**Figure 1.1c**). Notably, recent biochemical analysis of 9-Palmitic Acid esters of Hydroxy Stearic Acid (9-PAHSA) biosynthesis and degradation demonstrated that production of this endogenous FAHFA is stereospecific (*12*). This finding implies that the beneficial functions of FAHFAs are likely modulated by their stereochemistry as the stereoisomer *R*-9-PAHSA was more predominantly present in the adipose tissues of AG4OX mice as opposed to S-9-PAHSA (*12*). Currently, at least 80 distinct FAHFA isomers including all the regioisomers have been identified (*13*). Within these lipids, over 20 different endogenous FAHFA families have been reported in mammalian tissues, including liver, heart, kidneys, and pancreas in mammals, and have been identified in plants and foods (*14–21*). The ubiquitous findings of FAHFAs confirms their endogenous abundance, indicating an important role for this lipid class in physiology.

Nonetheless, only three FAHFAs, the PAHSAs and OHASAs (*11*), and DHAHLA (*14*) have been further characterized in functional assays. The open questions in the field include how many more FAHFA isomers or FAHFA-containing lipid species exist and which one(s) are biologically active with beneficial effects in mammalian physiology. Additionally, how do their structural composition modify their biological activities and contribute to favorable parameters in mammalian physiology? Indeed, the discovery of FAHFAs and their beneficial metabolic effects upended our current perception that fats are deleterious in general, highlighting a gap in our knowledge of the mammalian lipidome. This limited understanding drove our curiosity and

provides a framework for this dissertation in which we will seek further evidence of their biological relevance in nature and scrutinize their biological activities by exploiting their chemical properties.

1.2 FAHFAs are anti-inflammatory novel lipids

The anti-inflammatory properties of FAHFAs were first described in HFD-fed AG4OX mice (*11*). Because 9-PAHSA was the most abundant FAHFA isomer in mouse adipose tissues and its levels were correlative with insulin sensitivity in human subjects, subsequent experiments were designed to interrogate whether an acute oral administration of 9-PAHSA could improve the metabolic parameters of diet-induced obese (DIO) mice. Animal models of DIO is commonplace in studying inflammation and T2D in biomedical research (*22*). Comparison of glucose levels between vehicle- and 9-PAHSA-treated HFD mice revealed that 9-PAHSA lowered basal glycemia and improved the glucose tolerance of the treated mice (*11*), suggesting that endogenous FAHFAs could be protective in inflammation models (**Figure 1.2a**). Direct analysis of adipose tissue macrophages prepared from mice gavaged with vehicle or 9-PAHSA confirms the anti-inflammatory role of 9-PAHSA as its treatment attenuated expressions of tumor necrosis factor (TNF) and interleukin 1β (IL-1β), markers of inflammation (**Figure 1.2b**) (*11*).

The clinical benefits and safety of using FAHFA as a therapeutic agent in reducing inflammation are continually being explored. In addition to the DIO models discussed, FAHFAs were also demonstrated to reduce acute inflammation in *in vitro* assays (*14*) and in animal models of chronic inflammation, such as colitis (*15*) and type 1 diabetes (T1D) (*2*). Docosahexaenoic acid-derived FAHFAs, such as DHAHLA is another novel member of the

FAFHA family discovered in human and mouse tissues after dietary n-3 PUFA supplementation (*14*). Using cell-based assays, Kuda et al. assessed the anti-inflammatory effects of DHAHLAs by quantifying its ability to abrogate secretion of inflammatory markers in murine macrophages and human peripheral blood mononuclear cells (PBMCs) under LPS-induced acute inflammation (*14*). Comparison of IL-6, TNF-α, and IL-1β inhibition between 9-PAHSA and 13-DHAHLA treated RAW264.7 macrophages under acute inflammation reveals that 13-DHAHLA exhibited striking anti-inflammatory activities (**Figure 1.2c**) (*14*). Similarly, FAHFA's anti-inflammatory effects were also tested in animal models of chronic inflammation, such as ulcerative colitis (UC) (*15*). UC is a chronic inflammatory bowel disease of poorly understood etiology that can affect any portion of the gastrointestinal tract from the mouth to the perianal area (*23*, *24*). Clinical assessment of DSS-induced colitis mice following an oral administration of 9-PAHSA showed an improvement in disease severity compared to mice treated with vehicle (**Figure 1.2d**) (*15*). This finding suggested that 9-PAHSA can mediate the innate and adaptive immune responses to protect against inflammation in the gut.

Currently, many acute and chronic inflammatory diseases are managed by glucocorticoids. While acute administration of glucocorticoids does not generally impose great harm, long term corticosteroid therapy is commonly associated with side effects, including dependency and relapses (*25*), drug resistance (*26*), and a comprehensive list of adverse effects on pulmonary, cardiovascular, and metabolic functions (*27*). Therefore, identification of additional potent non-steroidal compounds that can reduce inflammation is paramount. FAHFAs' inflammation-reducing effects in mammalian disease models make them promising therapeutic agents in clinical settings. Using 9-PAHSA as a template, we will create stable

variants of FAHFAs and assess whether an enhanced anti-inflammatory effects can arise from specific structural compositions of these derivatives.

1.3 Discovery of smORF-encoded microproteins

The existence of protein-coding small open reading frames (smORFs) was overlooked by traditional gene annotation algorithms as preset parameters intended to reduce false discovery inadvertently created a blind spot for detecting smORFs and functional smORF-encoded polypeptides (SEPs) (*28–30*), also referred to as microproteins. Setting a minimum length cut-off at 100 codons for open reading frames (ORF), scientists excluded shorter coding sequences from annotation with the assumption that most coding genes encode for larger proteins (*31*). Additionally, a monocistronic model of the transcriptome was commonly accepted as evidence supporting the coding potential of noncanonical ORFs was scarce (*32*).

Recent studies using ribosome profiling (Ribo-Seq), an innovative method used to locate the position of the ribosome on RNAs to identify sites of active translation, have transformed the understanding of eukaryotic transcriptomes by revealing that many transcripts are, in fact, polycistronic with the ability to encode for two or more proteins. It has also become clear that alternative ORFs of annotated transcripts previously thought as noncoding, such as those found on the 5' UTR (*33–35*), noncoding RNAs (*36–38*), and circular RNAs (*39*, *40*) are translated as regulatory elements or as unique bioactive peptides (**Table 1.1**) (*41*). Sequential discoveries of microproteins across multiple species, including bacteria (*42*), plants (*43*, *44*), yeast (*45*), nematode (*46*), flies (*47*, *48*), and humans (*49*, *50*) highlight the conservation of smORFs and their critical roles in biology. It also provided compelling evidence that the use of stringent criteria for prediction of protein-coding ORFs are no longer definitive requirements for accurate

genome annotation. Removal of these rulesets indicates that the coding capacity of the eukaryotic genome is underestimated and that more bioactive microproteins are present in the mammalian transcriptome where they await discovery.

1.4 Functions of microproteins and uORFs

The biological functions of smORFs are diverse. The functions of smORF translation include but are not limited to cell signaling, allosteric regulation, complex stabilization, translation regulation, and immune response (**Figure 1.3**) (*51*). While microproteins can be encoded by a variety of smORFs, upstream ORFs (uORFs) represents the largest class of smORFs in the mouse and human transcriptome. Analysis of mouse and human genomes revealed that at least 50% of all genes contained at least an uORF (*52*). More recently, Martínez *et al* discovered that over 75% of newly identified smORFs found in annotated genes within human cell lines HEK293T, HeLaS3, and K562 cells are uORFs (**Figure 1.4a**). While these evidence suggests a high abundance of uORFs in mammalian transcriptomes, the translational capacity of uORFs remains elusive because the conventional dogma in eukaryotic gene translation restricts translation to occur once per mRNA and only at the longest ORF found within the transcript (*32*).

The functional implication of uORFs has been extensively explored in literature, with most studies focusing on its roles in cis-regulation of the downstream ORF translation (*33–35*, *53*, *54*). uORFs are predominantly known as *cis*-acting elements as they can regulate the translation of the downstream ORF (main ORF or mORF) through engagement with the ribosome (**Figure 1.4b**). This action serves as a post-transcriptional mechanism to readily and dynamically control the expression of the canonical protein. For example, the *ATF4* (*34*) and

CHOP (*33*) uORFs are translated at homeostasis. These uORF repress the translation of the mORFs, but the induction of cellular stress disengages the uORF from the scanning ribosome, resulting in robust translation of canonical proteins ATF4 and CHOP. In addition to *ATF4* and *CHOP* uORFs, a couple other uORFs can also mediate translation of their respective mORFs (*35*, *54*). Because uORFs are pervasive in mammalian genomes and the mechanism of microprotein functions are widespread, one important question in the field remains if uORFs may also encode for microproteins with independent functions other than *cis*-regulation. Over the years, the identification of MIEF1-MP (*35*) and PKC-h (*55*) have shed light on this topic. Specifically, the uORF of the *MIEF1* gene, produces a 70-amino acid microprotein MIEF-MP that regulates mitochondrial fission and protein-protein interactions with the mitochondrial ribosome and MIEF1. Moreover, the uORF on the *PKC-h* gene, which encodes a microprotein inhibitor of the main PKC-h protein as well as other members of the novel PKC subfamily. These findings confirm that uORF-encoded microproteins can possess unique functions other than cis-regulation. To further understand the scope of microprotein biology and identify features associated with uORF-encoded microproteins, additional characterization of functional uORF-encoded microproteins is critical.

1.5 Scope of Work

The goal of this dissertation is to improve human health through discovery and functional characterization of natural small molecules, namely lipids and proteins, that are essential building blocks of life. We aim to bridge a gap in our current understanding of the eukaryotic lipidome and the proteome as a modality to tackle the molecular basis of human diseases and potentially design better treatments. This dissertation will dedicate Chapter 2 and Chapter 3 to

focus on lipid discovery and characterization, respectively. In Chapter 2, we identified oat oil as a natural source of LAHLA, an additional novel FAHFA with anti-inflammatory activities. In Chapter 3, using 9-PAHSA as a starting point, we designed a library of stable FAHFA variants and screened their anti-inflammatory activities to scrutinize how chemical modifications modulate their biological activities through structural stabilization. Last but not least, Chapter 4 will be solely dedicated to delineating our efforts in identifying and characterizing an uORF-encoded orphan microprotein SLC35A4-MP.

1.6 Conclusion

Lipids:

FAHFA's anti-inflammatory effects are well-demonstrated in literature, particularly in the case of PAHSAs (*11*). Though over 80 natural FAHFAs have been identified, analysis of their physiological effects has centrally focused on the PAHSA family, potentiating a gap in our understanding of the functional implications of other FAHFA species. Currently, a few members of different lipid families have demonstrated beneficial activities in metabolic health (*14*). Therefore, we are curious whether there exists additional bioactive FAHFAs in nature.

Interestingly, Parsons and Kolar *et al.* discovered specific FAHFA-degrading enzymes AIG1 and ADTRP (*56*). One of the prevailing difficulties in studying FAHFAs is that endogenous lipids like FAHFAs are readily metabolized into inactive species by these hydrolysis enzymes, limiting the therapeutic usage of FAHFAs. To overcome these limitations and capitalize FAHFA activity, we want to understand if FAHFAs can be structurally modified via chemical synthesis to achieve stability, thus improved metabolic potency. And if so, how do structural differences in these FAHFAs impact their biological activity in the context of

inflammation? To better compare the anti-inflammatory activities of FAHFA-derivatives, we will implement a cell-based assay to screen their activities attributed to different chemical functionalities. We will also use human and mouse models of inflammation to validate the effects of selective FAHFA-derivatives based on their potency in reducing inflammation. Collectively, this dissertation will provide further evidence demonstrating the prevalence and biological significance of FAHFAs in the mammalian lipidome. Additionally, this report will investigate if and how synthetic analogs of FAHFAs may achieve enhanced anti-inflammatory activities, thus clinical value.

Microproteins:

It is clear that smORFs are pervasive in the human genome (*57*). Public databases, including SmProt (*58*), sORFs.org (*59*), and Martínez *et al.* (*57*) are in place to provide accessible data for already-identified smORFs. Although 50-70% of them are uORFs, most functional microproteins already characterized are not derived from uORFs due to challenges of working with polycistronic genes. However, uORF-encoded microproteins like MIEF (*35*) confirms that uORF functions are more diverse than just cis-regulation and that more microproteins with independent functions can be derived from the sizable list of uORFs in the eukaryotic transcriptome. In this dissertation, we will reveal the functional diversity of uORFs by characterizing uORF found in the 5' leader of the *SLC35A4* gene encodes a functional mitochondrial SLC35A4 microprotein (SLC35A4-MP). Our findings will demonstrate that SLC35A4-MP is the newest member of the mitochondrial contact site and cristae organizing system (MICOS) complex, an important mitochondrial protein complex essential for mitochondrial cristae formation and oxidative function. Our findings exemplify that new

functional members of vital protein complexes exist in the abundant reservoir of uncharacterized uORF-encoded microproteins, providing further incentive for their continued evaluation. Using our experimental design, we aim to provide a template to accelerate the functional elucidation in future uORF-encoded microproteins discoveries. Our goal is to explore the functional potential of the eukaryotic smORFeome and to improve annotation as well as basic understanding of the mammalian genome.

1.7 Figures

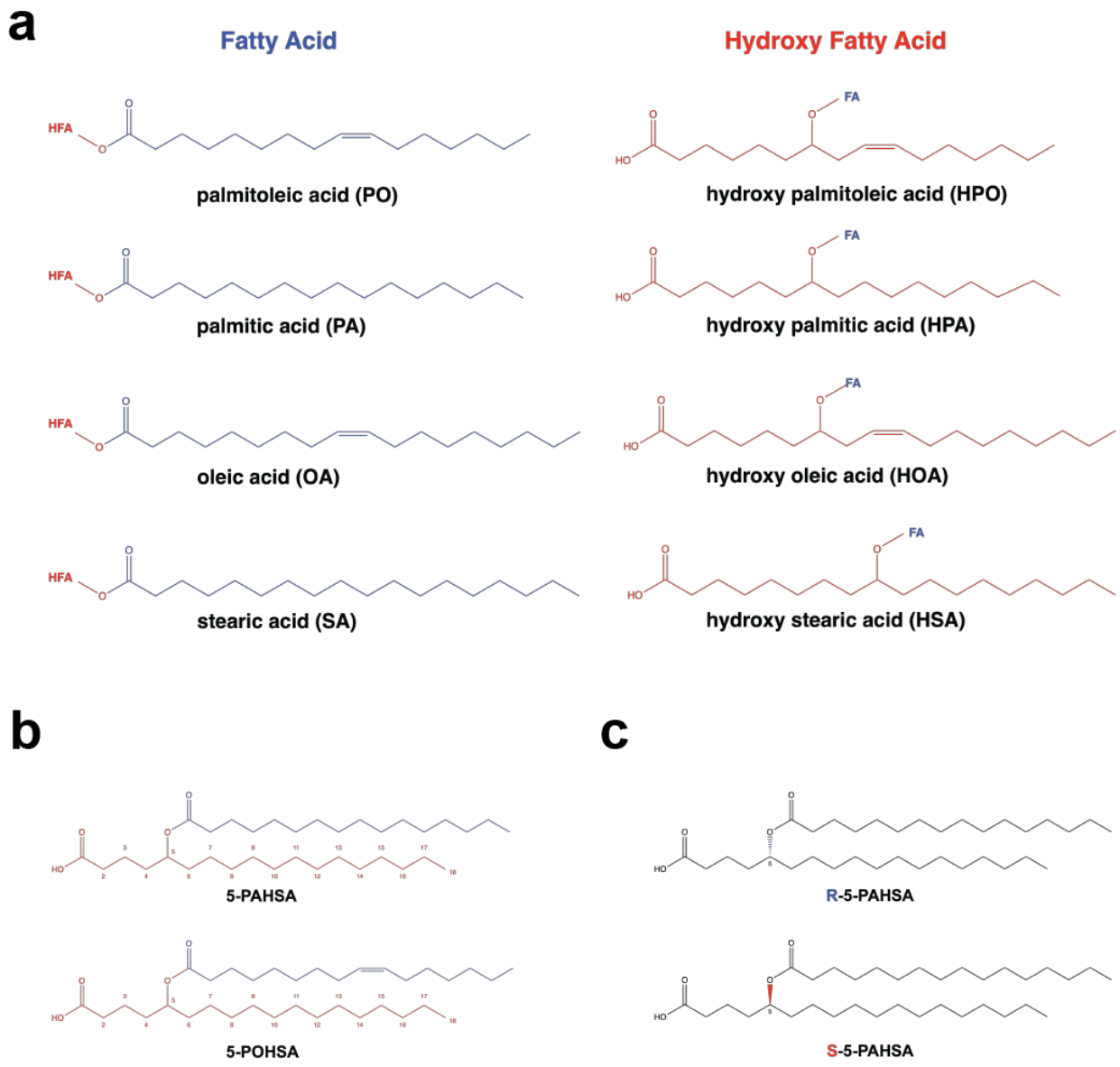


Figure 1.1. A FAHFA structure arises from a distinct combination of a fatty acid (FA) and a hydroxy fatty acid (HFA). (**a)** Examples of different FAs (PO, PA, OA, SA, indicated in blue), and HFAs (HPO, HPA, HOA, HSA indicated in red). (**b**) An example of two different FAHFA families (5-PAHSA and 5-POHSA), with blue indicating the FA and red indicating the HFA. Numbering on the HFA chains indicate the position of the ester linkage with respect to the carboxyl carbon. (**c**) R- and S- regioisomers of 5-PAHSA, with blue indicating R- and red indicating the S- configuration.

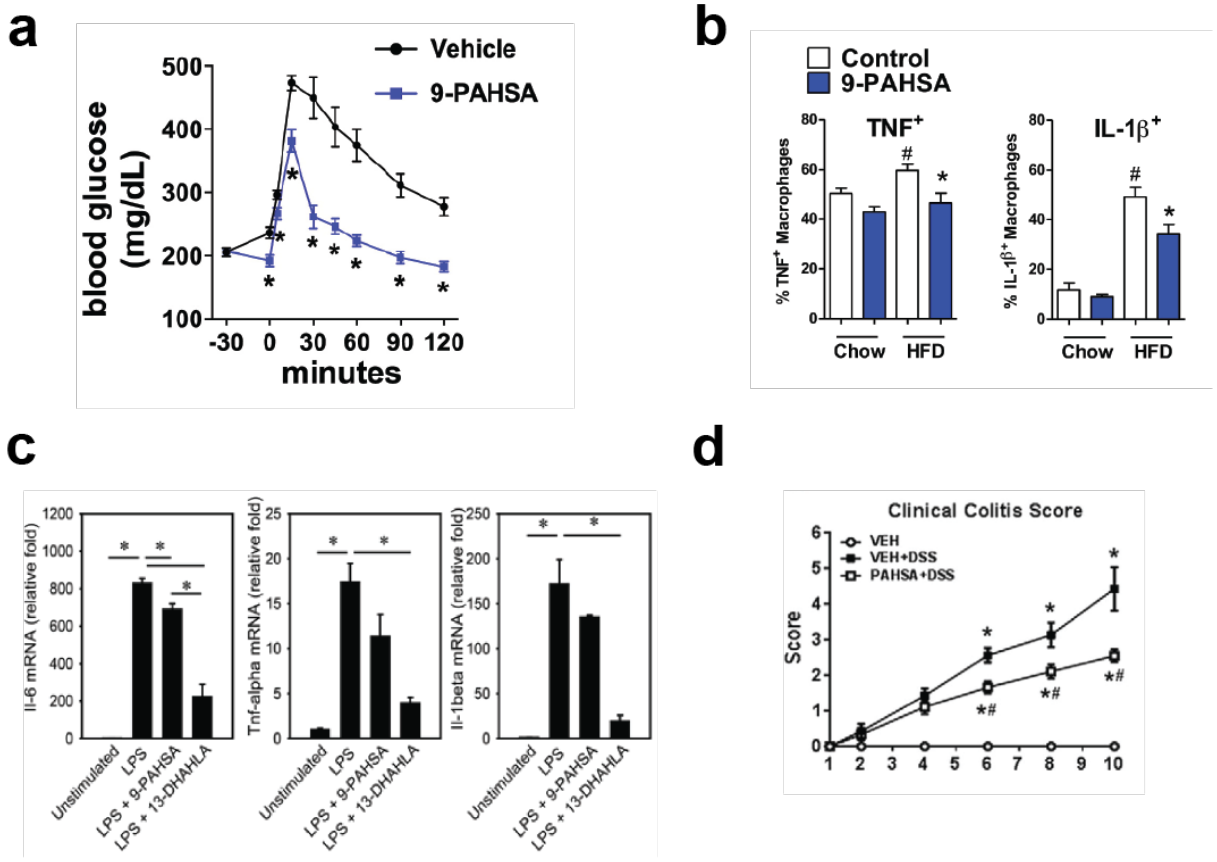


Figure 1.2. FAHFAs are anti-diabetic and anti-inflammatory. (**a**) A glucose tolerance test was used to measure the blood glucose of HFD-fed mice 30 minutes following an oral administration of vehicle or 9-PAHSA. 9-PAHSA treatment improves glucose tolerance in HFD mice. (**b**) Quantitation of adipose tissue (AT) macrophages expressing TNFα and IL-1β obtained from flow cytometry analysis. Chow- or HFD-fed mice were orally administered vehicle control or 9-PAHSA (30mg/kg for chow-fed mice and 45 mg/kg for HFD-fed mice) for 3 days. On day 4, AT macrophages derived from perigonadal white adipose tissue (PG-WAT) were stained using TNFα and IL-1β markers. 9-PAHSA treatment lowers markers of inflammation in PG-WAT. (**c**) Unstimulated or LPS-stimulated RAW264.7 macrophages were treated with 10 μmol/L of 9-PAHSA or 13-DHAHLA for 18 hours. mRNA quantitation of IL-6, TNFα and IL-1β was performed using treated cells. FAHFA, such as 9-PAHSA and 13-DHAHLA lowers mRNA levels of inflammation markers, with 13-DHAHLA showing a superior activity. (**d**) Clinical colitis scores were monitored in mice treated with vehicle (VEH) or 5- and 9-PAHSA (PAHSA) followed by colitis induction with vehicle (VEH) or 2% DSS water for 10 days. Treatment of PAHSA protects DSS-induced mice from colitis.

Table 1.1. Examples of human bioactive microproteins encoded by noncanonical transcripts.

Microprotein	Transcript	Length (aa)	Context	Physiological significance
ASAP	*LINC00467*	94	lncRNA	Promotes colorectal cancer proliferation
APPLE	Antisense to *ASH1L*	90	lncRNA	Overexpressed in acute myeloid leukemia
ASDURF	*ASNSD1*	96	Upstream	Encodes the 12th subunit of the PAQosome chaperone complex
ASRPS	*LINC00908*	60	lncRNA	Inhibits angiogenesis in triple-negative breast cancer
BiP uORF	*HSPA5*	9	Upstream	Generation of MHC class I antibodies during the integrated stress response
BRAWNIN	*C12orf73*	71	Transcript of unknown function	Essential for oxidative phosphorylation. Deletion causes mitochondrial disease
CASIMO1	*CASIMO1*	83	lncRNA	Aids in proliferation of breast cancer cells
Cdk2ap1ΔN	*Cdk2ap1*ΔN (isoform of *Cdk2ap1*)	88	Transcript isoform (alternative promoter)	N-terminally truncated protein variant driven by an alternative promoter. Mediates preimplantation embryo development
CHOP uORF or DDIT3 uORF/altDDIT3	*CHOP/DDIT3*	31 or 34	Upstream	Microprotein shown to downregulate CHOP expression. Binds with CHOP
CIP2A-BP	*LINC00665*	52	lncRNA	Inhibits migration and invasion of triple-negative breast cancer cells
circPPP1R12A-73aa	*circPPP1R12A*	73	circRNA	Promotes proliferation, migration, and invasion abilities of colon cancer cells
DWORF	*LOC100507537*	35	lncRNA	Facilitates Ca^{2+} cycling for muscle contractability in cardiomyocytes
ELABELA	*APELA*	54	lncRNA	Critical for cell survival in human embryonic stems cells. Activates PI3K/AKT/mTORC1 signaling
FORCP	*FORCP*	79	lncRNA	Inhibits proliferation and tumorgenicity in colorectal cancer cells
HJV uORF	*HJV*	18 or 27	Upstream	Microprotein shown to downregulate HJV expression which is involved in iron metabolism
HOXB-AS3	*HOXB-AS3*	53	lncRNA	Inhibits colorectal cancer tumorigenesis
Humanin and SHLP1-8	*MT-RNR2*	24-38	Mitochondrial transcript	Implicated in age-related disease progression. Cytoprotective and neuroprotective effects
KRASIM	*NCBP2-AS2*	99	lncRNA	Suppresses proliferation of hepatocarcinoma cells
MELOE-1 and MELOE-2	*Meloe*	46 and 39	lncRNA	Melanoma immunosurveillance and elicits melanoma-specific T cell response
MIAC	*RP11-469H8.6*	51	lncRNA	Suppresses tumor growth and metastasis in head and neck squamous cell carcinoma
MIEF1 uORF	*MIEF1*	70	Upstream	Binds to the mitochondrial ribosome regulating its translational activity. Binds with MIEF1
miPEP155	*MIR155HG*	17	lncRNA	Regulates antigen presentation on cells
Mitoregulin	*LINC00116*	56	lncRNA	Disruption perturbs lipid homeostasis
MOCCI	*C15ORF48*	83	Transcript of unknown function	Protective against excessive immune function within the mitochondria
MOTS-c	*MT-RNR1*	16	Mitochondrial transcript	Regulates metabolic homeostasis
MP31	*PTEN*	31	Upstream	Limits lactate–pyruvate conversion in the mitochondria by competing with mLDH for NAD+
MRI-2	*CYREN*	69	Transcript of unknown function	DNA repair
Myomixer	*MYMX*	84	Transcript of unknown function	Promotes myofiber formation during muscle development
Myoregulin	*LINC00948*	46	lncRNA	Regulates Ca^{2+} in skeletal muscle through interaction with SERCA
NoBody	*LINC01420*	68	lncRNA	Modulates P-body numbers in cells
Phospholamban	*PLN*	52	Transcript of unknown function	Regulates Ca^{2+} in cardiac muscle through interaction with SERCA
PIGBOS	Antisense to *PIGB*	54	lncRNA	Regulates the unfolded protein response in the endoplasmic reticulum
PINT87aa	*LINC-PINT* (*circPINT-exon2*)	87	circRNA	Inhibits transcriptional elongation in multiple oncogenes
PRL-1 and PRL-2 uORF	*PTP4A1* and *PTP4A2*	21	Upstream	Microprotein regulates magnesium homeostasis
RBRP	*LINC00266-1*	71	lncRNA	Enhances m6A recognition-dependent *c-myc* mRNA stability. Promotes colorectal cancer tumorigenesis
Sarcolipin	*SLN*	31	Transcript of unknown function	Regulates Ca^{2+} cycling and thermogenesis in skeletal muscle
SEHBP	*ZNF689*	46	Upstream	Interacts with chromatin-associated proteins and is likely a transcription factor
SMIM30	*LINC00998*	59	lncRNA	Promotes hepatocellular carcinoma tumorigenesis
SPAAR	*LINC00961*	90	lncRNA	Important for muscle regeneration. SPAR is proangiogenic
STORM	*LINC00689*	50	lncRNA	Competes with SRP19 to bind 7S RNA
YY1BM	*LINC00278*	21	lncRNA	Under nutrient deficiency, expression promotes esophageal squamous cell carcinoma apoptosis
ZFAS1	Antisense to *ZNFX1*	57	lncRNA	Promotes hepatocellular carcinoma cell migration

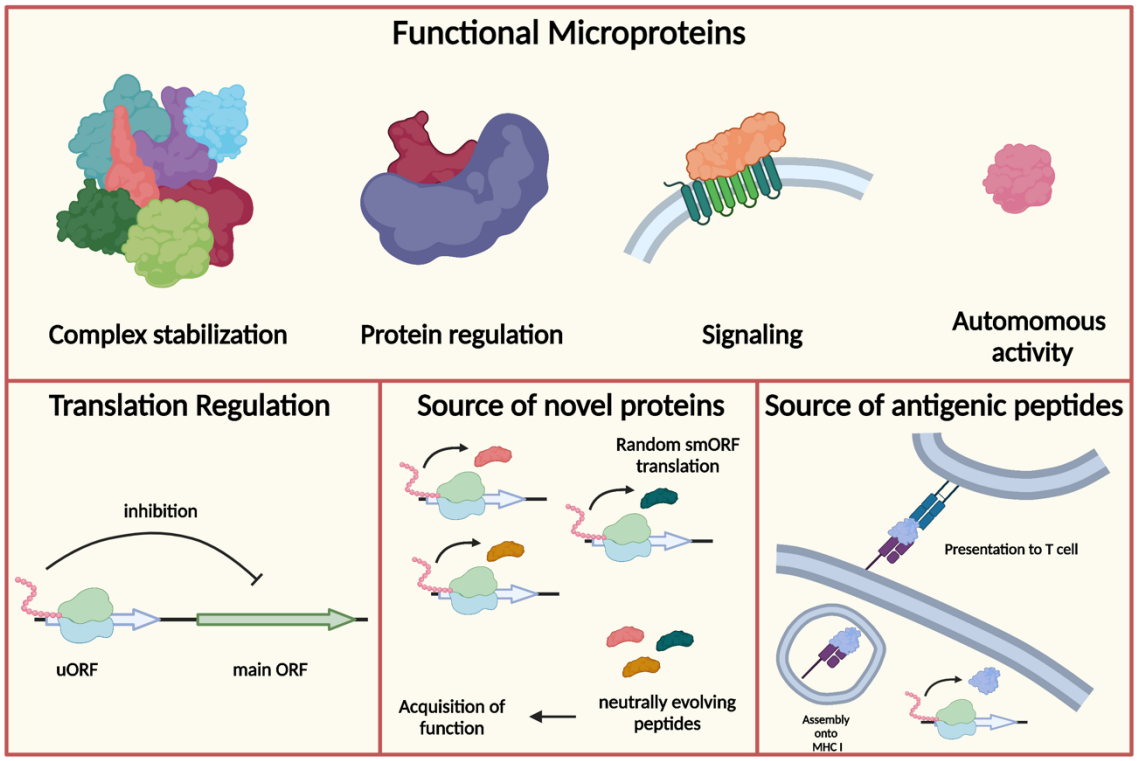


Figure 1.3. Functions of small ORF (smORF)-encoded microproteins are diverse in biology.

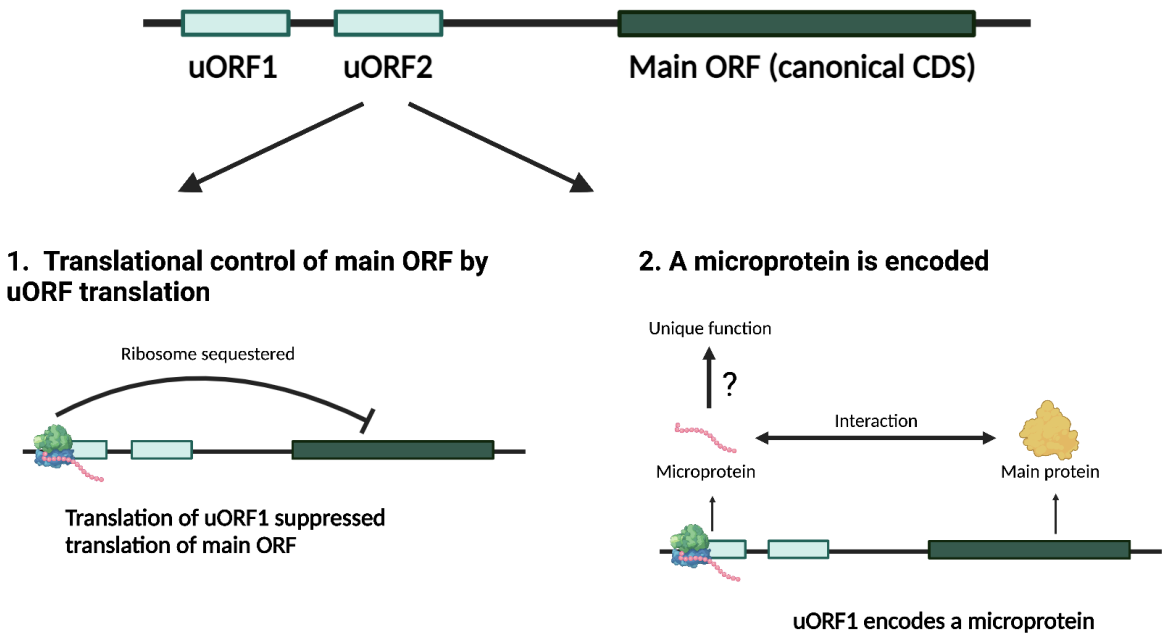


Figure 1.4. Upstream ORFs (uORFs) are conserved noncanonical transcripts with regulatory functions in the human genome. (**a**) The majority of human novel protein-coding smORFs in RefSeq transcripts are found in the 5'UTR region, excluding unannotated transcripts. Nonannotated (**b**) Current models of uORF functions.

1.8 References

1. G. Boden, G. I. Shulman, Free fatty acids in obesity and type 2 diabetes: defining their role in the development of insulin resistance and beta-cell dysfunction. *Eur. J. Clin. Invest.* **32 Suppl 3**, 14–23 (2002).

2. I. Syed, M. F. Rubin De Celis, J. F. Mohan, P. M. Moraes-Vieira, A. Vijayakumar, A. T. Nelson, D. Siegel, A. Saghatelian, D. Mathis, B. B. Kahn, PAHSAs attenuate immune responses and promote β cell survival in autoimmune diabetic mice. *J. Clin. Invest.* **129**, 3717–3731 (2019).

3. R. Roberts, L. Hodson, A. L. Dennis, M. J. Neville, S. M. Humphreys, K. E. Harnden, K. J. Micklem, K. N. Frayn, Markers of de novo lipogenesis in adipose tissue: associations with small adipocytes and insulin sensitivity in humans. *Diabetologia*. **52**, 882–890 (2009).

4. M. A. Herman, O. D. Peroni, J. Villoria, M. R. Schön, N. A. Abumrad, M. Blüher, S. Klein, B. B. Kahn, A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. *Nature*. **484**, 333–338 (2012).

5. M. D. Bruss, C. F. Khambatta, M. A. Ruby, I. Aggarwal, M. K. Hellerstein, Calorie restriction increases fatty acid synthesis and whole body fat oxidation rates. *Am. J. Physiol. Endocrinol. Metab.* **298**, E108-16 (2010).

6. H. Cao, K. Gerhold, J. R. Mayers, M. M. Wiest, S. M. Watkins, G. S. Hotamisligil, Identification of a Lipokine, a Lipid Hormone Linking Adipose Tissue to Systemic Metabolism. *Cell*. **134**, 933 (2008).

7. D. Y. Oh, S. Talukdar, E. J. Bae, T. Imamura, H. Morinaga, W. Q. Fan, P. Li, W. J. Lu, S. M. Watkins, J. M. Olefsky, GPR120 is an Omega-3 Fatty Acid Receptor Mediating Potent Anti-Inflammatory and Insulin Sensitizing Effects. *Cell*. **142**, 687 (2010).

8. J. K. Virtanen, J. Mursu, S. Voutilainen, M. Uusitupa, T. P. Tuomainen, Serum omega-3 polyunsaturated fatty acids and risk of incident type 2 diabetes in men: The kuopio ischemic heart disease risk factor study. *Diabetes Care*. **37**, 189–196 (2014).

9. P. R. Shepherd$, L. Gnudi, E. Tozzo, H. Yang, F. Leach, B. B. Kahn$, Adipose Cell Hyperplasia and Enhanced Glucose Disposal in Transgenic Mice Overexpressing GLUT4 Selectively in Adipose Tissue*. *J. Biol. Chem.* **268**, 22243–22246 (1993).

10. E. Carvalho, K. Kotani, O. D. Peroni, B. B. Kahn, Adipose-specific overexpression of GLUT4 reverses insulin resistance and diabetes in mice lacking GLUT4 selectively in muscle. *Am. J. Physiol. - Endocrinol. Metab.* **289**, 551–561 (2005).

11. M. M. Yore, I. Syed, P. M. Moraes-Vieira, T. Zhang, M. A. Herman, E. A. Homan, R. T. Patel, J. Lee, S. Chen, O. D. Peroni, A. S. Dhaneshwar, A. Hammarstedt, U. Smith, T. E. McGraw, A. Saghatelian, B. B. Kahn, Discovery of a class of endogenous mammalian

lipids with anti-diabetic and anti-inflammatory effects. *Cell*. **159**, 318–332 (2014).

12. A. T. Nelson, M. J. Kolar, Q. Chu, I. Syed, B. B. Kahn, A. Saghatelian, D. Siegel, Stereochemistry of Endogenous Palmitic Acid Ester of 9-Hydroxystearic Acid and Relevance of Absolute Configuration to Regulation. *J. Am. Chem. Soc.* **139**, 4943 (2017).

13. M. J. Kolar, S. Konduri, T. Chang, H. Wang, C. McNerlin, L. Ohlsson, M. Härröd, D. Siegel, A. Saghatelian, Linoleic acid esters of hydroxy linoleic acids are anti-inflammatory lipids found in plants and mammals. *J. Biol. Chem.* **294**, 10698–10707 (2019).

14. O. Kuda, M. Brezinova, M. Rombaldova, B. Slavikova, M. Posta, P. Beier, P. Janovska, J. Veleba, J. Kopecky, E. Kudova, T. Pelikanova, Docosahexaenoic acid-derived fatty acid esters of hydroxy fatty acids (FAHFAS) with anti-inflammatory properties. *Diabetes*. **65**, 2580–2590 (2016).

15. J. Lee, P. M. Moraes-Vieira, A. Castoldi, P. Aryal, E. U. Yee, C. Vickers, O. Parnas, C. J. Donaldson, A. Saghatelian, B. B. Kahn, Branched fatty acid esters of hydroxy fatty acids (FAHFAs) protect against colitis by regulating gut innate and adaptive immune responses. *J. Biol. Chem.* **291**, 22207–22217 (2016).

16. T. Zhang, S. Chen, I. Syed, M. Ståhlman, M. J. Kolar, E. A. Homan, Q. Chu, U. Smith, J. Borén, B. B. Kahn, A. Saghatelian, A LC-MS–based workflow for measurement of branched fatty acid esters of hydroxy fatty acids. *Nat. Protoc. 2016 114*. **11**, 747–763 (2016).

17. A. Vijayakumar, P. Aryal, J. Wen, I. Syed, R. P. Vazirani, P. M. Moraes-Vieira, J. P. Camporez, M. R. Gallop, R. J. Perry, O. D. Peroni, G. I. Shulman, A. Saghatelian, T. E. McGraw, B. B. Kahn, Absence of Carbohydrate Response Element Binding Protein in Adipocytes Causes Systemic Insulin Resistance and Impairs Glucose Transport. *Cell Rep.* **21**, 1021–1035 (2017).

18. M. Brezinova, O. Kuda, J. Hansikova, M. Rombaldova, L. Balas, K. Bardova, T. Durand, M. Rossmeisl, M. Cerna, Z. Stranak, J. Kopecky, Levels of palmitic acid ester of hydroxystearic acid (PAHSA) are reduced in the breast milk of obese mothers. *Biochim. Biophys. acta. Mol. cell Biol. lipids*. **1863**, 126–131 (2018).

19. E. Pflimlin, M. Bielohuby, M. Korn, K. Breitschopf, M. Löhn, P. Wohlfart, A. Konkar, M. Podeschwa, F. Bärenz, A. Pfenninger, U. Schwahn, T. Opatz, M. Reimann, S. Petry, N. Tennagels, Acute and Repeated Treatment with 5-PAHSA or 9-PAHSA Isomers Does Not Improve Glucose Control in Mice. *Cell Metab.* **28**, 217-227.e13 (2018).

20. I. Syed, J. Lee, P. M. Moraes-Vieira, C. J. Donaldson, A. Sontheimer, P. Aryal, K. Wellenstein, M. J. Kolar, A. T. Nelson, D. Siegel, J. Mokrosinski, I. S. Farooqi, J. J. Zhao, M. M. Yore, O. D. Peroni, A. Saghatelian, B. B. Kahn, Palmitic Acid Hydroxy Stearic Acids Activate GPR40 Which is Involved in Their Beneficial Effects on Glucose Homeostasis. *Cell Metab.* **27**, 419 (2018).

21. Q. F. Zhu, J. W. Yan, T. Y. Zhang, H. M. Xiao, Y. Q. Feng, Comprehensive Screening and Identification of Fatty Acid Esters of Hydroxy Fatty Acids in Plant Tissues by Chemical Isotope Labeling-Assisted Liquid Chromatography-Mass Spectrometry. *Anal. Chem.* **90**, 10056–10063 (2018).

22. D. Avtanski, V. A. Pavlov, K. J. Tracey, L. Poretsky, Characterization of inflammation and insulin resistance in high-fat diet-induced male C57BL/6J mouse model of obesity. *Anim. Model. Exp. Med.* **2**, 252–258 (2019).

23. T. Kobayashi, B. Siegmund, C. Le Berre, S. C. Wei, M. Ferrante, B. Shen, C. N. Bernstein, S. Danese, L. Peyrin-Biroulet, T. Hibi, Ulcerative colitis. *Nat. Rev. Dis. Prim. 2020 61*. **6**, 1–20 (2020).

24. J. R. Maxwell, J. L. Viney, *Curr. Protoc. Pharmacol.*, Chapter 5, Unit5.57 (2009)

25. P. J. Rutgeerts, Review article: the limitations of corticosteroid therapy in Crohn's disease. *Aliment. Pharmacol. Ther.* **15**, 1515–1525 (2001).

26. P. J. Barnes, Corticosteroid resistance in patients with asthma and chronic obstructive pulmonary disease. *J. Allergy Clin. Immunol.* **131**, 636–645 (2013).

27. M. Yasir, A. Goyal, S. Sonthalia, Corticosteroid Adverse Effects. *StatPearls* (2021) (available at https://www.ncbi.nlm.nih.gov/books/NBK531462/).

28. A. Saghatelian, J. P. Couso, Discovery and characterization of smORF-encoded bioactive polypeptides. *Nat. Chem. Biol.* (2015), pp. 909–16.

29. M. A. S. Mumtaz, J. P. Couso, Ribosomal profiling adds new coding sequences to the proteome. *Biochem. Soc. Trans.* **43**, 1271-6 (2015).

30. J. I. Pueyo, E. G. Magny, J. P. Couso, New Peptides Under the s (ORF)ace of the Genome. *Trends Biochem. Sci.* (2016), pp. 665–678.

31. B. Dujon, D. Alexandraki, B. André, W. Ansorge, V. Baladron, J. P. G. Ballesta, A. Banrevi, P. A. Bolle, M. Bolotin-Fukuhara, P. Bossier, G. Bou, J. Boyer, M. J. Buitrago, G. Cherét, L. Colleaux, B. Dalgnan-Fornier, F. Del Rey, C. Dion, H. Domdey, A. Düsterhöft, S. Düsterhus, K. D. Entian, H. Erfle, P. F. Esteban, H. Feldmann, L. Fernandes, G. M. Fobo, C. Fritz, H. Fukuhara, C. Gabel, L. Gaillon, J. M. Carcia-Cantalejo, J. J. Garcia-Ramirez, M. E. Gent, M. Ghazvini, A. Goffeau, A. Gonzaléz, D. Grothues, P. Guerreiro, J. Hegemann, N. Hewitt, F. Hilger, C. P. Hollenberg, O. Horaitis, K. J. Indge, A. Jacquier, C. M. James, J. C. Jauniaux, A. Jimenez, H. Keuchel, L. Kirchrath, K. Kleine, P. Kötter, P. Legrain, S. Liebl, E. J. Louis, A. Maia E Silva, C. Marck, A. L. Monnier, D. Möstl, S. Müller, B. Obermaier, S. G. Oliver, C. Pallier, S. Pascolo, F. Pfeiffer, P. Philippsen, R. J. Planta, F. M. Pohl, T. M. Pohl, R. Pöhlmann, D. Portetelle, B. Purnelle, V. Puzos, M. R. Rad, S. W. Rasmussen, M. Remacha, J. L. Revuelta, G. F. Richard, M. Rieger, C. Rodrigues-Pousada, M. Rose, T. Rupp, M. A. Santos, C. Schwager, C. Sensen, J. Skala, H. Soares, F. Sor, J. Stegemann, H. Tettelin, A.

Thierry, M. Tzermia, L. A. Urrestarazu, L. Van Dyck, J. C. Van Vliet-Reedijk, M. Valens, M. Vandenbo, C. Vilela, S. Vissers, D. Von Wettstein, H. Voss, S. Wiemann, G. Xu, J. Zimmermann, M. Haasemann, I. Becker, H. W. Mewes, Complete DNA sequence of yeast chromosome XI. *Nature*. **369**, 6479 (1994).

32. M. A. Brunet, S. A. Levesque, D. J. Hunting, A. A. Cohen, X. Roucou, Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* **28**, 609 (2018).

33. C. Jousse, Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR. *Nucleic Acids Res. **29**, 4341-4351* (2001).

34. E. Balsa, M. S. Soustek, A. Thomas, S. P. Gygi, A. Enriquez, ER and Nutrient Stress Promote Assembly of Respiratory Chain Supercomplexes through the PERK-eIF2α Axis. *Mol. Cell*. **74**, 877--890.e6 (2019).

35. A. Rathore, Q. Chu, D. Tan, T. F. Martinez, C. J. Donaldson, J. K. Diedrich, J. R. Yates, A. Saghatelian, MIEF1 Microprotein Regulates Mitochondrial Translation. *Biochemistry*. **57**, 5564–5575 (2018).

36. B. R. Nelson, C. A. Makarewich, D. M. Anderson, B. R. Winders, C. D. Troupes, F. Wu, A. L. Reese, J. R. McAnally, X. Chen, E. T. Kavalali, S. C. Cannon, S. R. Houser, R. Bassel-Duby, E. N. Olson, A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science*. **351**, 271–275 (2016).

37. A. Matsumoto, A. Pasut, M. Matsumoto, R. Yamashita, J. Fung, E. Monteleone, A. Saghatelian, K. I. Nakayama, J. G. Clohessy, P. P. Pandolfi, mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature*. **541**, 228–232 (2017).

38. K. Hanyu-Nakamura, H. Sonobe-Nojima, A. Tanigawa, P. Lasko, A. Nakamura, Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. *Nature*. **451**, 730–733 (2008).

39. Z. Yang, L. Xie, L. Han, X. Qu, Y. Yang, Y. Zhang, Z. He, Y. Wang, J. Li, Circular RNAs: Regulators of Cancer-Related Signaling Pathways and Potential Diagnostic Biomarkers for Human Cancers. *Theranostics*. **7**, 3106 (2017).

40. P. F. Zhang, C. Y. Wei, X. Y. Huang, R. Peng, X. Yang, J. C. Lu, C. Zhang, C. Gao, J. Bin Cai, P. T. Gao, D. M. Gao, G. M. Shi, A. W. Ke, J. Fan, Circular RNA circTRIM33-12 acts as the sponge of MicroRNA-191 to suppress hepatocellular carcinoma progression. *Mol. Cancer*. **18**, 1–15 (2019).

41. B. W. Wright, Z. Yi, J. S. Weissman, J. Chen, The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* **32**, 243–258 (2022).

42. C. S. Wadler, C. K. Vanderpool, A dual function for a bacterial small RNA: SgrS

performs base pairing-dependent regulation and encodes a functional polypeptide. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20454–9 (2007).

43. S. A. Casson, P. M. Chilley, J. F. Topping, I. M. Evans, M. A. Souter, K. Lindsey, The POLARIS gene of Arabidopsis encodes a predicted peptide required for correct root growth and leaf vascular patterning. *Plant Cell*. **14**, 1705–1721 (2002).

44. H. Röhrig, J. Schmidt, E. Miklashevichs, J. Schell, M. John, Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1915–1920 (2002).

45. J. P. Kastenmayer, L. Ni, A. Chu, L. E. Kitchen, W. C. Au, H. Yang, C. D. Carter, D. Wheeler, R. W. Davis, J. D. Boeke, M. A. Snyder, M. A. Basrai, Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. *Genome Res.* **16**, 365–373 (2006).

46. C. A. Gleason, Q. L. Liu, V. M. Williamson, Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. *Mol. Plant-Microbe Interact.* **21**, 576–585 (2008).

47. M. I. Galindo, J. I. Pueyo, S. Fouix, S. A. Bishop, J. P. Couso, Peptides encoded by short ORFs control development and define a new eukaryotic gene family. *PLoS Biol.* **5**, 1052–1062 (2007).

48. T. Kondo, Y. Hashimoto, K. Kato, S. Inagaki, S. Hayashi, Y. Kageyama, Small peptide regulators of actin-based cell morphogenesis encoded by a polycistronic mRNA. *Nat. Cell Biol.* **9**, 660–665 (2007).

49. Y. Hashimoto, T. Niikura, H. Tajima, T. Yasukawa, H. Sudo, Y. Ito, Y. Kita, M. Kawasumi, K. Kouyama, M. Doyu, G. Sobue, T. Koide, S. Tsuji, J. Lang, K. Kurokawa, I. Nishimoto, A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Aβ. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6336–41 (2001).

50. S. A. Slavoff, A. J. Mitchell, A. G. Schwaid, M. N. Cabili, J. Ma, J. Z. Levin, A. D. Karger, B. A. Budnik, J. L. Rinn, A. Saghatelian, Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **9**, 59–64 (2013).

51. D. Schlesinger, S. J. Elsässer, Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins. *FEBS J.* **289**, 53–74 (2022).

52. S. E. Calvo, D. J. Pagliarini, V. K. Mootha, Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–7512 (2009).

53. C. Hu, L. Shu, X. Huang, J. Yu, liuju Li, L. Gong, M. Yang, Z. Wu, Z. Gao, Y. Zhao, L. Chen, Z. Song, OPA1 and MICOS Regulate mitochondrial crista dynamics and formation. **11**, 1–17 (2020).

54. D. E. Andreev, P. B. O'connor, C. Fahey, E. M. Kenny, I. M. Terenin, S. E. Dmitriev, P. Cormican, D. W. Morris, I. N. Shatsky, P. V. Baranov, Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife*. **2015**, e03971 (2015).

55. H. Raveh-Amit, A. Maissel, J. Poller, L. Marom, O. Elroy-Stein, M. Shapira, E. Livneh, Translational Control of Protein Kinase Cη by Two Upstream Open Reading Frames. *Mol. Cell. Biol.* **29**, 6140 (2009).

56. W. H. Parsons, M. J. Kolar, S. S. Kamat, A. B. Cognetta, J. J. Hulce, E. Saez, B. B. Kahn, A. Saghatelian, B. F. Cravatt, AIG1 and ADTRP are atypical integral membrane hydrolases that degrade bioactive FAHFAs. *Nat. Chem. Biol.* **12**, 367–372 (2016).

57. T. F. Martinez, Q. Chu, C. Donaldson, D. Tan, M. N. Shokhirev, A. Saghatelian, Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).

58. Y. Hao, L. Zhang, Y. Niu, T. Cai, J. Luo, S. He, B. Zhang, D. Zhang, Y. Qin, F. Yang, R. Chen, SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief. Bioinform.* **19**, 636–643 (2018).

59. V. Olexiouk, J. Crappé, S. Verbruggen, K. Verhegen, L. Martens, G. Menschaert, SORFs.org: A repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* **44**, 324–329 (2016).

CHAPTER 2

Discovery of LAHLA as a natural lipid regulator of inflammation

Abstract

Fatty acid esters of hydroxy fatty acids (FAHFAs) are a recently discovered class of biologically active lipids. Here we identify the linoleic acid ester of 13-hydroxy linoleic acid (13-LAHLA) as an anti-inflammatory lipid. An oat oil fraction and FAHFA-enriched extract from this fraction showed anti-inflammatory activity in a lipopolysaccharide-induced cytokine secretion assay. Structural studies identified three LAHLA isomers (15-, 13-, and 9-LAHLA) as being the most abundant FAHFAs in the oat oil fraction. Of these LAHLAs, 13-LAHLA is the most abundant LAHLA isomer in human serum after ingestion of liposomes made of fractionated oat oil, and it is also the most abundant endogenous LAHLA in mouse and human adipose tissue. As a result, we chemically synthesized 13-LAHLA for biological assays. 13-LAHLA suppresses lipopolysaccharide-stimulated secretion of cytokines and expression of pro-inflammatory genes. These studies identify LAHLAs as an evolutionarily conserved lipid with anti-inflammatory activity in mammalian cells.

2.1 Introduction

Lipids are fundamental building blocks of life. They are the basic components of mammalian cells, in which they provide structural support of the plasma membrane to facilitate transport of molecules. Lipids are also signaling molecules and regulators of diverse cellular processes, including proliferation (*1*), growth and development (*2*), apoptosis (*3*), and regeneration (*4*). Moreover, metabolism of lipids creates sources of energy and bioactive lipid molecules that are essential for bioenergetics and homeostasis. Because lipids play a wide array of roles in biology, a disruption in lipid homeostasis are commonly associated with detrimental pathogeneses, such as insulin resistance (*5*), obesity and type 2 diabetes (T2D) (*6*), cardiovascular diseases (*7*), and neurodegenerative disorders (*8*). Therefore, a deeper knowledge of the human lipidome is critical to understanding the molecular basis of human diseases attributed to lipid imbalance.

Recently, fatty acid esters of hydroxy fatty acids (FAHFAs) were identified as a novel class of endogenous lipids with anti-diabetic and anti-inflammatory activities (*9*). Yore and colleagues discovered that overexpression of glucose transporter Glut4 in AG4OX mice resulted in de novo lipogenesis in adipose tissues and improved the metabolic profiles associated with HFD-induced inflammation (*10*, *11*). This finding potentiated the hypothesis that de novo lipogenesis could drive the production of beneficial lipids contributing to the favorable metabolic parameters observed, such as glucose tolerance, lower basal glycemia, and insulin sensitivity (*9*). Similar to a small group of other lipids with reported beneficial effects, such as omega-3 fatty acids (*12*) and C16:1n7-palmitoleate (*13*), the biological role of palmitic acid esters of hydroxy stearic acids (PAHSAs) in abrogating inflammation was explored and established *in vitro* and *in vivo* (*9*, *14*, *15*).

To date, more than 20 distinct FAHFA families have been identified in literature. However, only the PAHSAs (*9*, *14*, *15*) and docosahexaenoic acid esters of hydroxy linoleic acids (DHAHLA) (*16*) have been functionally elucidated. Interestingly, analysis of the biological activities of 13-DHAHLA revealed that the novel lipid was even more potent in reducing LPS-induced inflammation than 9-PAHSA (*16*). This indicates that different biological activities of FAHFAs may arise from distinct structural compositions of the lipid class. Therefore, to further understand and appreciate FAHFA's roles in the lipidome, we explored if additional bioactive FAHFA members exist in nature. We will investigate if natural sources with previously demonstrated anti-inflammatory activities contain biologically active lipids that would lead to bioactive FAHFAs.

Oat oil emerged as a potential source of bioactive FAHFAs because colloidal oatmeal has anti-inflammatory activity (*17*, *18*) and physiological studies revealed that ingestion of liposomes made of fractionated oat oil resulted in beneficial metabolic effects (*19*). Interestingly, a digalactosyldiacylglycerol (DGDG) estolide is the most abundant glycolipid found in oat kernels (*20*, *21*) and this glycolipid structure contains a distinct FAHFA structure, the linoleic acid ester of 15-hydroxy linoleic acid (15-LAHLA) (**Figure 2.1a**). This led us to reason that oat oil is likely a potential source of biologically active FAHFAs, which may contribute to the anti-inflammatory effects of colloidal oats. In this report, we identified LAHLA as an endogenous FAHFA family present in oats and mammalian tissues. Biological testing of synthetic 13-LAHLA reveals the lipid is another regulator of inflammation.

2.2 Results

Testing the anti-inflammatory activity of oat oil fractions

Previous reports have identified that colloidal oatmeal is anti-inflammatory and have beneficial metabolic effects upon consumption of liposomes made of fractionated oat oil rich in polar lipids (*17*). Crude oat oil, obtained by ethanol extraction of oats, can be fractionated using different concentrations of ethanol, water and sugar (*22*). The resulting fractions are rich in neutral lipids (called T1 fraction) or rich in polar lipids (called T2 fraction). We measured the anti-inflammatory effects of oat oil fraction T1 by treating lipopolysaccharide (LPS)-stimulated RAW264.7 cells with T1 and measuring interleukin 6 (IL-6) release following treatment. This assay was previously used to assess the anti-inflammatory activities of 9-PAHSA and 13-DHAHLA based on their ability to attenuate the release of proinflammatory cytokines, such as IL-6 (*16*). The T1 fraction exhibited a dose-dependent anti-inflammatory activity with significant suppression of LPS-stimulated IL-6 secretion starting at 10 ppm (~15% inhibition) (**Figure 2.1b**).

Solid-phase extraction (SPE) of the T1 fraction using a protocol for FAHFA enrichment (*23*), resulted in fraction (SPE-T1 fraction) with more potent anti-inflammatory activity than T1. For example, the SPE-T1 fraction resulted in approximately 50% IL-6 inhibition at 10 ppm and over 90% IL-6 inhibition at 100 ppm (**Figure 2.1c**). To show that these effects were due to activity and not cytotoxicity, we performed an MTT viability assay, and there was no observed cellular death at the tested concentrations of the T1 oat oil fraction and SPE-T1 fraction (**Figure 2.2**).

The SPE-T1 fraction contains Linoleic acid esters of hydroxy linoleic acid (LAHLAs)

We measured FAHFAs levels in the SPE-T1 fraction by liquid chromatography-mass spectrometry (LC-MS). These measurements were designed to detect different FAHFA families including PAHSAs, OAHSAs, and LAHLAs. LAHLAs are the most abundant FAHFAs in the SPE-T1 fraction (~300 pmol/mg), while PAHSAs and OAHSAs are detected at much lower concentrations (~1 pmol/mg) (**Figure 2.3a**). We detected three different LAHLA peaks in the LC-MS chromatogram (**Figure 2.3b** and **Figure 2.4**) indicative of three different LAHLA isomers. No other FAHFAs were as abundant as the LAHLAs, and, therefore, we focused on the structural and functional characterization of the LAHLAs.

LAHLA regioisomers

To be able to study these LAHLAs in greater detail, we needed to know which regioisomers were present and then synthesize pure version of these lipids for biological assays. To identify the structures of these LAHLA regioisomers, we applied a recently reported strategy for the structural elucidation of the 13-DHAHLA (*16*). This method utilizes multiple fragmentation steps (MS3) to identify the position of the hydroxy group (**Figure 2.5a**). We reasoned that because LAHLAs and DHAHLAs both contain hydroxy linoleic acid (HLA) the method should be directly transferable to identify LAHLA regioisomers.

We did not have access to an instrument capable of MS3, so we performed a pseudo-MS3 instead (*23*). The source voltage of the mass spectrometer was increased to induce in-source fragmentation of LAHLAs to linoleic acid and HLA, and the HLA was further fragmented in the mass spectrometer to reveal the position of the hydroxy group (**Figure 2.5a**). Analysis of the oat oil samples using this approach led to HLA fragments that co-elute with the different LAHLA

isomer peaks (**Figure 2.5b, c**). Using this approach, we identified the LAHLA isomers contain 15-, 13-, and 9-HLA (**Figure 2.5c**). The 13- and 9-LAHLA product ions are identical with reported fragments for 13- and 9-DHALHAs (*16*), providing strong evidence for these regioisomer assignments.

LAHLA levels after ingestion of liposomal oat oil

Liposomes made from fractionated oat oil rich in polar lipids (fraction T2) were used as a dietary supplement to look for beneficial postprandial metabolic effects (*19*). Ingestion of liposomal oat oil correlated with lower glucose levels and elevated levels of beneficial hormones, including peptide YY (PYY), glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2 (GLP-2) and cholecystokinin (CCK) in human subjects. Using serum acquired in a previous study (*19*), we measured FAHFA levels in these samples to determine the change in LAHLAs levels from control-fed or oat oil-fed human subjects.

LAHLAs levels in subjects ingesting liposomal oat oil were much higher than in control subjects, indicating that oat oil was the source of plasma LAHLA levels (**Figure 2.6a**). Furthermore, we see no change in PAHSAs or OAHSAs between the two groups, consistent with oat oil having high LAHLA levels (**Figure 2.6b, c**). In subjects that received liposomal oat oil, we detected the 15-, 13-, and 9-LAHLAs (**Figure 2.6d**). The levels of the LAHLA isomers are dynamic with 13- and 9-LAHLA increasing with time and 15-LAHLA decreasing (**Figure 2.7**). Overall, 13-LAHLA is the most abundant of all the LAHLAs (**Figure 2.6d**). The serum concentrations of LAHLAs reached approximately 40 nM, which is over 45 times higher than control subjects. These data identify 13-LAHLA as the primary serum LAHLA after ingestion of liposomal oat oil. Thus, there is a positive correlation between LAHLAs and the beneficial

metabolic and endocrine effects reported in these individuals (*19*). However, these experiments do not prove that the LAHLAs are the cause of these beneficial effects, since oat oil is comprised of many other molecules. Additional studies in animal models will be required to determine what, if any, effects dietary LAHLAs have in biology.

LAHLAs are endogenous mammalian lipids

PAHSAs are endogenous FAHFAs with the highest levels found in adipose tissue (*9*). There exist multiple PAHSA regioisomers with 9-PAHSA being the most abundant PAHSA. Although PAHSAs are present in some foods, they can be synthesized endogenously. We explored if LAHLAs were also present in adipose tissues. We examined subcutaneous white adipose tissue (SQWAT) of wild-type (WT) mice fed ad libitum for the presence of LAHLAs and observed that there were multiple LAHLA regioisomers present with 13-LAHLA being the most abundant regioisomer (**Figure 2.8a, b**). We also observed a similar finding with 13-LAHLA being the most abundant regioisomer in human white adipose tissue (**Figure 2.8c, d**). The discovery of 13-LAHLA as a naturally occurring lipid in rodents, humans, and oat oil pointed to 13-LAHLA as an ideal candidate for further investigation of LAHLA biology.

Synthesis of 13-LAHLA

To determine if 13-LAHLA has direct effects on inflammation, we first needed to synthesize 13-LAHLA (**Figure 2.9a**). Synthetic FAHFAs have exclusively contained saturated HFA fragments due to the early identification and relative ease of synthesis of this class of FAHFAs (*9*, *16*, *24*, *25*). For the laboratory synthesis of 13-LAHLA, the challenge was preparing large quantities of 13-hydroxylinoleic acid (13-HLA is synonymous with 13-HODE and coriolic

acid) with the correct olefinic configuration. While there have been several multistep syntheses of 13-HLA (*26–30*), we utilized a concise approach implementing a selective selenium dioxide oxidation of the methyl ester of linoleic acid (*31*) (**Figure 2.9a**). While the oxidative transformation proceeded in low yield the approach dramatically simplified the synthesis. Importantly, standard phase column chromatography could be applied to separating the multiple products, including olefin isomers, with the desired 13-HLA methyl ester produced as the major product. This reaction enabled access to ample quantities of 13-HLA methyl ester from readily available methyl linoleate. From this intermediate esterification of the secondary, allylic alcohol of 13-HLA methyl ester with the acid chloride of linoleic acid yielded the methyl ester of 13-LAHLA. Selective saponification of the methyl ester in preference to the fatty acid ester, due to steric effects, generated synthetic 13-LAHLA for the first time. Comprehensive NMR analysis confirmed the correct olefinic isomer was present and the placement of the ester linkage was at the 13^{th} carbon.

The characteristic precursor-to-product ion transitions and the retention times were the same for the synthetic and natural 13-LAHLA indicating a structural match (**Figure 2.9b**). Using the synthetic 13-LAHLA, we validated the structure of the LAHLA isomers in oat oil and human plasma as 13-LAHLA (**Figure 2.9c**).

13-LAHLA is an anti-inflammatory lipid

9-PAHSA and 13-DHAHLA have been shown to have anti-inflammatory effects. 9-PAHSA blocks LPS-stimulated dendritic cell (*9*) and macrophage (*16*) activation, and reduces inflammation in a mouse colitis model (*15*). We analyzed the effects of 9-PAHSA and 13-LAHLA on suppression of LPS-stimulated cytokine expression in RAW 264.7 macrophages.

Both treatment of 9-PAHSA and 13-LAHLA at 10 μM significantly suppressed the mRNA levels of IL-6 and IL-1β, with 13-LAHLA having a stronger effect (**Figure 2.10a, b**). Administration of 13-LAHLA to RAW 264.7 cells in the absence of LPS had no impact on control IL-6 levels (data not shown). In addition, 13-LAHLA (10 μM) also suppressed mRNA levels of iNOS and COX-2 in RAW 264.7 cells when stimulated with LPS (**Figure 2.10c, d**). The reduction in expression of pro-inflammatory cytokines as well as downstream regulators of inflammation iNOS and COX-2 indicate that 13-LAHLA has broad anti-inflammatory effects.

Three hydrolases—CEL, AIG1, and ADTRP—have been identified *in vitro* as FAHFA-specific lipases (*32*, *33*) and these enzymes led us to perform additional experiments to ensure that 13-LAHLA and not LAHLA hydrolysis products are responsible for the activity we observe. First, we measured the activity of 13-LAHLA and 13-HLA in this macrophage-stimulation assay and found that 13-LAHLA is significantly more active than its hydrolysis product 13-HLA (**Figure 2.11a**). Neither 13-LAHLA or 13-HLA was toxic towards cells. Comparison of linoleic acid and 13-HLA showed that they are equipotent in this assay (**Figure 2.11b**). These experiments demonstrate that 13-LAHLA is more active than either breakdown product, supporting that the LAHLA is the biologically active species in these studies.

2.3 Discussion

Lipids are fundamental regulators in biology with roles in membrane structure, sources of energy, and signaling molecules. Lipids in the latter category are often classified into families, such as prostaglandins (*34*) or oxysterols (*35*), that share a structural feature but with many distinct members. Individual members in a family can have distinct activities or different potencies for the same activity. For example, with FAHFAs, 9-PAHSA is anti-inflammatory but

5-PAHSA is not (*9*), and similar differences in structure-activity relationships are observed with other lipid classes (*34*). Analytical experiments suggest that there are many FAHFA families (*9*, *16*, *36*, *37*) but only a few of these have been functionally characterized (*9*, *16*). FAHFAs were detected in foods, including vegetables (*9*), and recent work has detected these lipids in rice and the plant model organism Arabidopsis thaliana (*38*). We hypothesized that there might be additional, uncharacterized FAHFAs in bioactive lipid extracts from plants.

The historical record reports the use of oats (Avena sativa) to treat skin conditions thousands of years of years ago. Today, oats, in particular colloidal oats, are present in many commercial personal care products (e.g., soaps, shampoos, and shaving creams (*39*, *40*). Identification of active molecules in oats identified avenanthramide alkaloids (*41*, *42*) which have anti-oxidant and anti-inflammatory activity *in vitro* (*43*) and *in vivo* (*44*). Our interest in oat oil was kindled by a physiological study in humans that found beneficial metabolic and hormonal effects of a liposomal oat oil dispersion (*19*).

We hypothesized that oat oil might contain biologically active FAHFAs because some of the observed effects of liposomal oat oil in humans mirrored that of FAHFAs in mice (*9*, *14*), and oats contained a complex lipid, a DGDG (*20*), which contains an esterified 15-LAHLA (**Figure 2.1a)**. In this work we have identified free FAHFAs in oat oil with LAHLAs being the most pronounced. Our hypothesis was bolstered by the observation that 13-LAHLA is anti-inflammatory.

We also detected LAHLAs as endogenous lipids and increased LAHLA concentrations upon ingestion of liposomal oat oil (fraction T2) (**Figure 2.6a, d)**. These observations do not prove that LAHLAs are the cause of the beneficial activity of liposomal oat oil in vivo, since the beneficial effects observed in these studies could be due to other molecules in the oat oil, but

only serve to identify the most abundant LAHLA regioisomer. Future studies that utilize pure LAHLAs pharmacologically or perturb LAHLA effects metabolically will be necessary to understand what role, if any, LAHLAs have in mammalian biology. This is the second observation of LAHLAs in vivo, as a study that developed a higher resolution method to measure FAHFA levels reported these lipids (*37*).

Recent work that characterized the function of 13-DHAHLA provided a guide (*16*) for the in vitro characterization of 13-LAHLA. We suspected that the 13-LAHLA and 13-DHALHA might have similar activities due to their similar structures of a 13-HLA core esterified with a polyunsaturated fatty acid. Testing of 13-LAHLA for its ability to inhibit LPS-stimulated macrophage activity revealed that this lipid is active with an IC_{50} of approximately 30 μM. While 9-PAHSA is active in this assay, we find that it is much less so than 13-LAHLA (**Figure 2.10**).

We extended these measurements to include iNOS and COX-2 as additional inflammatory markers (**Figure 2.10c, d**). Nitric oxide (NO) signaling is a key trigger in the immune system to help rid infection, but in some cases overactive NO signaling is thought to damage healthy cells (*45*, *46*). There are three NO producing enzymes endothelial-, neuronal-, and inducible nitric oxide synthase (NOS) (eNOS, nNOS, and iNOS) (38). iNOS is the primary regulator of NO signaling in macrophages, and we measured changes in the expression of iNOS in the presence and absence of LPS. These experiments revealed modest inhibition of iNOS expression (50%) at 10 μM of 13-LAHLA. We also measured COX-2 levels to infer changes in the prostaglandin pathway (*47*). We find that administration 13-LAHLA (10 μM) inhibits COX-2 expression by over 50%. Together these data indicate that 13-LAHLA affects many branches

of the cellular inflammatory response including cytokine production and secretion, nitric oxide signaling, and prostaglandin production.

One challenge when testing endogenous lipids, such as the LAHLAs, is that these metabolites are susceptible to endogenous metabolism, and caution must be taken to conclude that a specific metabolite is biologically active. For instance, 13-DHAHLA was shown to be produced upon administration of DHA to cells and mice (*16*), suggesting that at least some of the biological activity of DHA might come from 13-DHAHLA (*16*). With 13-LAHLA, it is possible that the breakdown products 13-HLA or LA are active in these assays.

We measured the IC50s of 13-LAHLA and 13-HLA on IL-6 secretion and found that 13-HLA is less active than 13-LAHLA (approximately 30 μM for 13-LAHLA versus 274 μM for 13-HLA) (**Figure 2.11a**). In a separate experiment, we find that 13-HLA and LA have similar IC50s (**Figure 2.11b**) to indicate that both of these lipids are at least an order of magnitude less active than the intact 13-LAHLA. The data argues that even if 13-LAHLA is hydrolyzed, the products are less active and therefore less likely to contribute to the observed activity in our assays.

In summary, our findings demonstrate that FAHFAs are an evolutionary conserved class of lipids, from plants to humans, with anti-inflammatory activity. Moreover, the characterization of 13-LAHLA here along with the reported activity of 13-DHAHLA begin to reveal a trend that polyunsaturated FAHFAs are more active than the saturated FAHFAs tested so far (i.e. 9-PAHSA).

Previous work on the stereochemistry of hydroxy linoleic acids, specifically 15-HLA, was identified as possessing R stereochemistry (*20*). The existence of a stereocenter on this lipid indicates that the lipid is the product of an enzymatic, and therefore stereo specific pathway,

instead of a non-enzymatic pathway. Future efforts will look to synthesize LAHLAs enantiomers so that we can determine if these lipids from oat oil and tissues are enantiopure, and whether the different stereoisomers vary in their biological activity.

The identification of anti-inflammatory activity for a natural lipid that is non-toxic when consumed and the ability to modify this lipid while retaining activity indicate that we should continue to explore the biology and functions of FAHFAs, especially 13-LAHLA as we examine the functional roles in the body and explore the therapeutic potential of this pathway.

2.4 Conclusion

The identification of DGDG estolides in oats reveals that the structure contains a distinct FAHFA structure, 15-LAHLA. This drove our hypothesis to investigate if the biological activities of oat oil could be due to the presence of uncharacterized LAHLAs. Evaluation of synthetic 13-LAHLA indicate that the novel FAHFA is biologically active as it attenuates markers of inflammation in LPS-stimulated RAW264.7 macrophages. Simultaneous testing 13-LAHLA and 13-HLA confirms that the majority of anti-inflammatory effects, indeed, is attributed to 13-LAHLA and not its hydrolysis product. These findings corroborated the finding that FAHFAs are a conserved class of bioactive lipids found in plants and mammals at endogenous levels. Consistent with the functional characterization of DHAHLA, biological evaluation of 13-LAHLA highlights a trend that polyunsaturated FAHFAs may be more potent at reducing inflammation than the saturated FAHFAs characterized so far, such as 9-PAHSA. This suggests that a structure-activity relationship exist in FAHFAs and further elucidation of this relationship may improve understanding of the functional roles of FAHFAs and its potential therapeutic value.

2.5 Figures

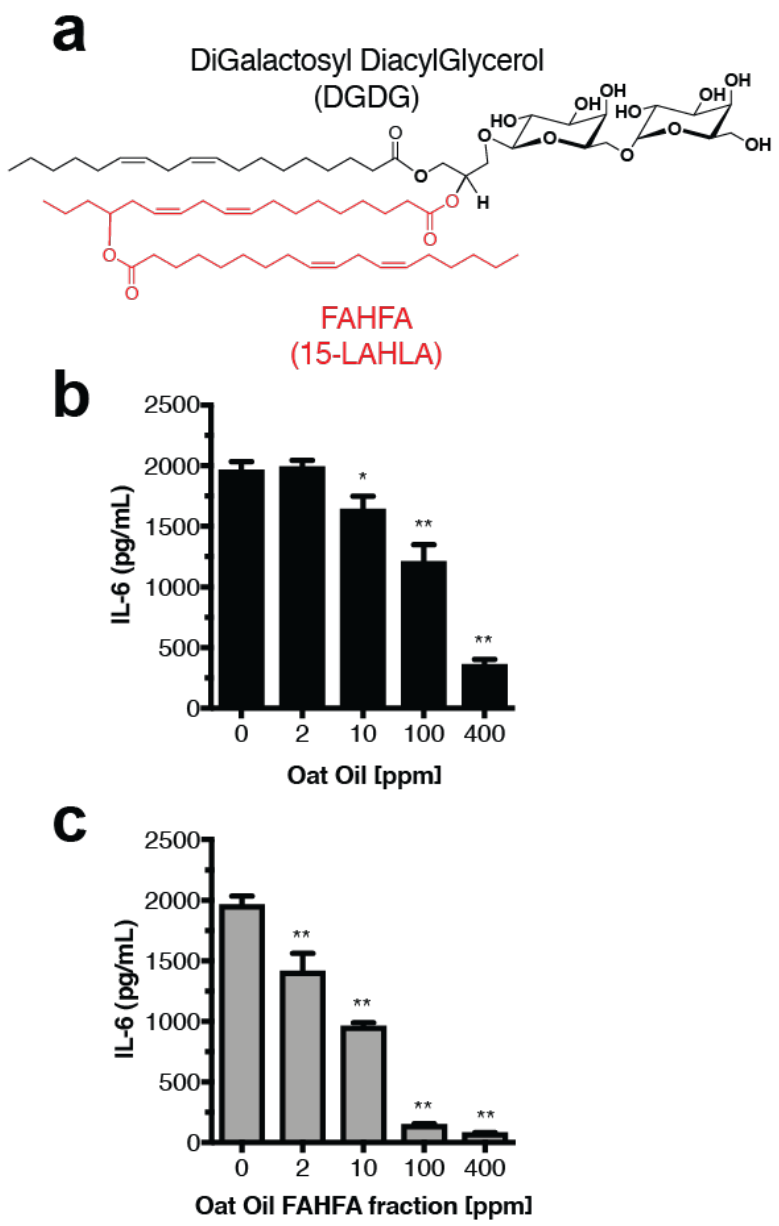


Figure 2.1. Anti-inflammatory activity of oat oil. (**a**) The reported structure of a DGDG from oat seed. (**b-c**) Activity of oat oil fraction T1 (**b**) and solid phase extracted sample of the T1 fraction (**c**) on IL-6 secretion from LPS-stimulated RAW 264.7 cells. Data represent mean ± S.E.M. for three biological replicates. **p, ≤ 0.0001; *p, ≤ 0.01; versus LPS by one-way ANOVA.

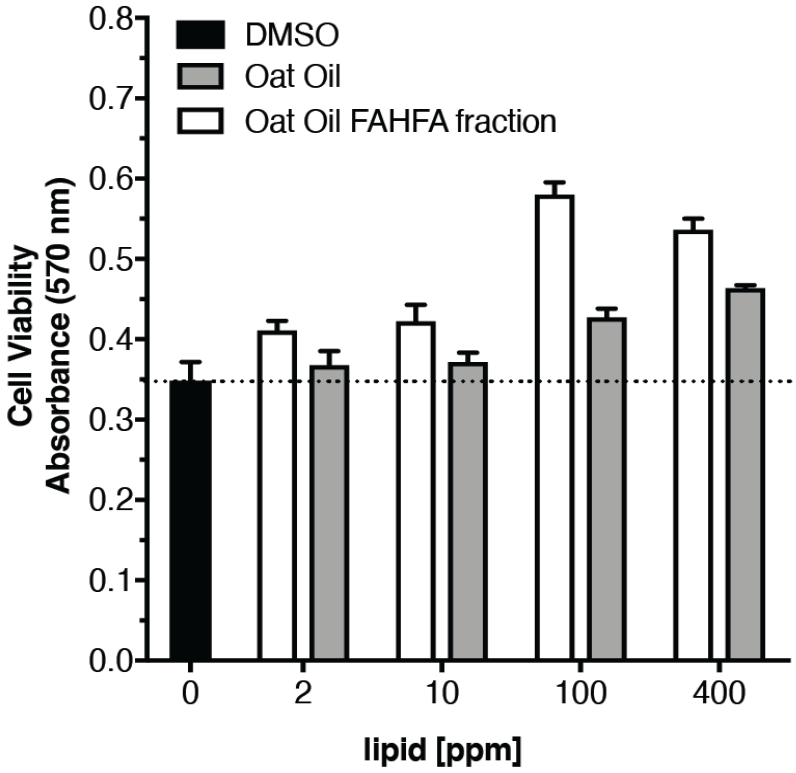


Figure 2.2. MTT cell viability assay for T1 oat oil fraction and SPE-T1 fraction corresponding to samples in **Figure 2.1b** and **Figure 2.1c**.

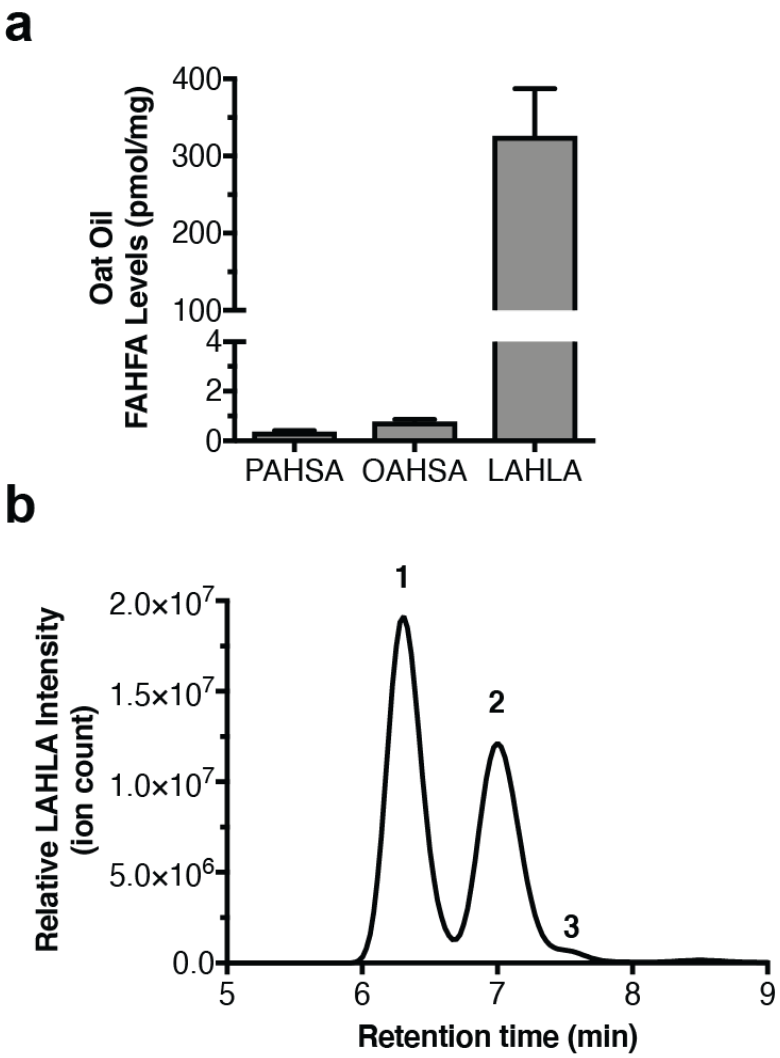


Figure 2.3. LAHLAs are primary oat oil FAHFAs. (**a**) LAHLAs are more abundant than PAHSAs or OAHSAs in oat oil. (**b**) LC-MS chromatogram showing three LAHLA regioisomers.

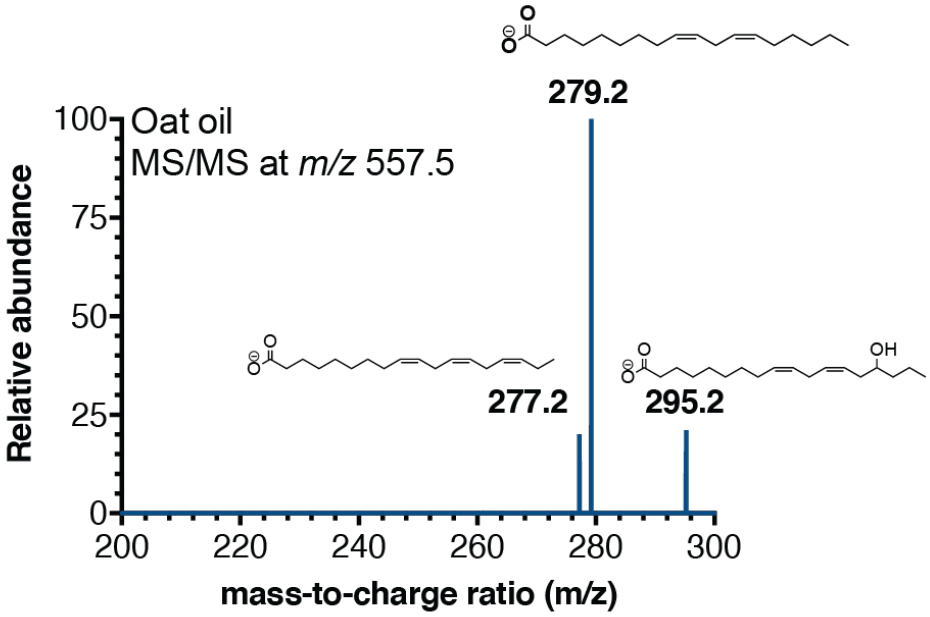


Figure 2.4. Representative MS2 of a LAHLA. All three LC-MS contained these fragment ions in their MS2 spectra to indicate that they are LAHLA isomers.

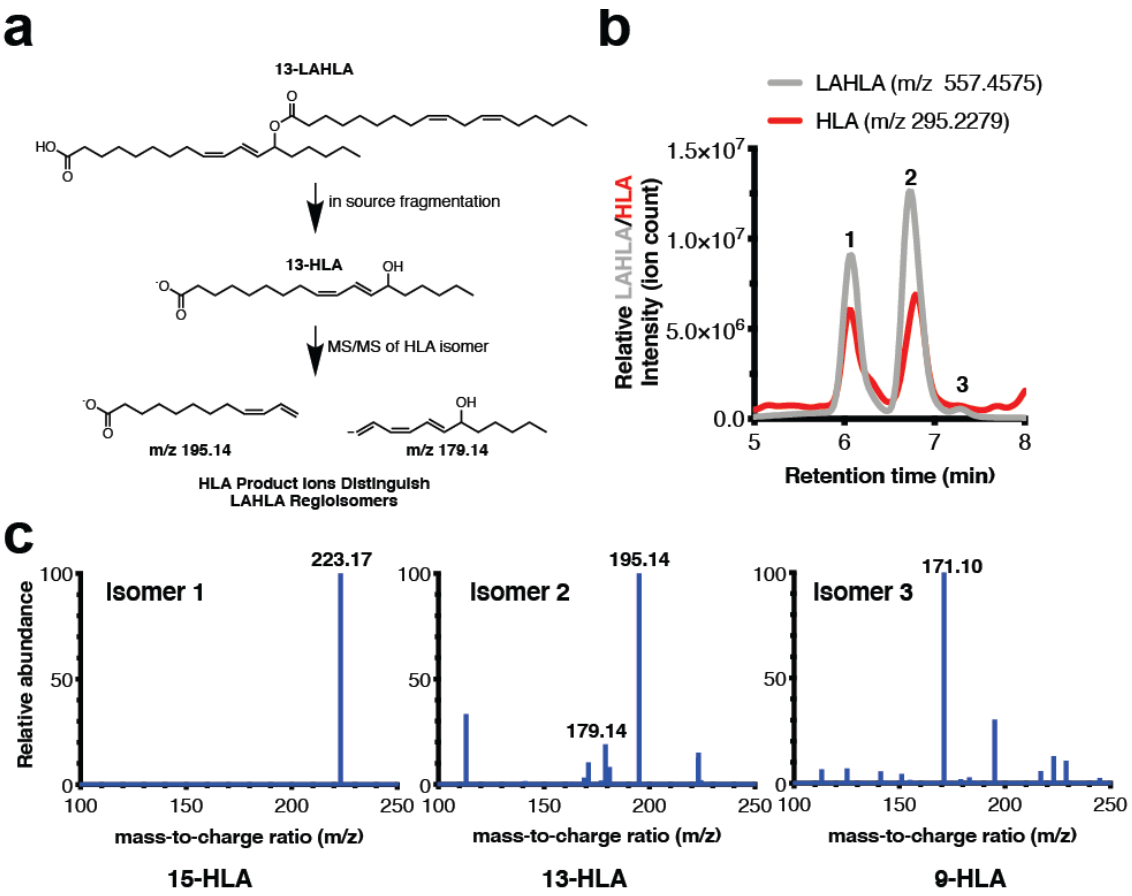


Figure 2.5. Identification of LAHLA regioisomers in oat oil. (**a**) Schematic of a pseudo-MS3 approach to identify the LAHLA regioisomers. (**b**) LC-MS chromatograms of FAHFA-enriched oat oil with a source energy of 80 eV showing the presence of LAHLAs (*m/z* 557.4575, grey) and HLA fragments from the LAHLAs (*m/z* 295.2279, red). (**c**) MS/MS of HLAs from regioisomers **1**-**3** identify the LAHLAs as 15-LAHLA (**1**), 13-LAHLA (**2**), and 9-LAHLA (**3**).

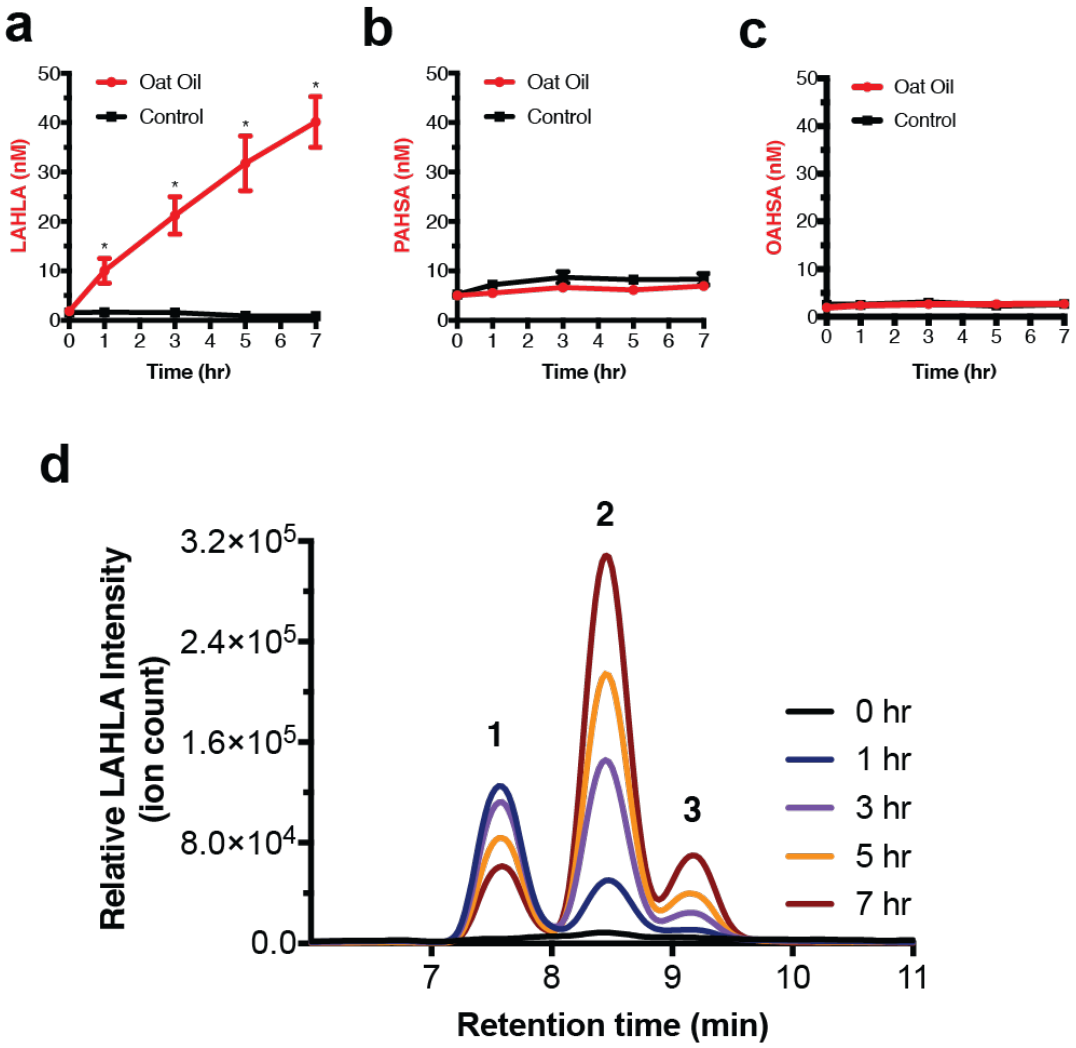


Figure 2.6. LAHLAs in plasma after ingestion of liposomal oat oil. (**a-c**) Levels of plasma LAHLAs (**a**), PAHSAs (**b**), and OAHSAs (**c**), in subjects fed liposomal oat-oil versus control fed subjects over 7 hours. Time 0 hours represents plasma FAHFAs after an overnight fast. Data represents mean ± S.E.M. (n = 10-12). (**d**) LC-MS chromatograms of LAHLA regioisomers in human plasma over a 7-hour time period after liposomal oat oil ingestion. Data represents mean ± S.E.M. (n = 10-12). *p, ≤ 0.05 by two-sided Student's t-test.

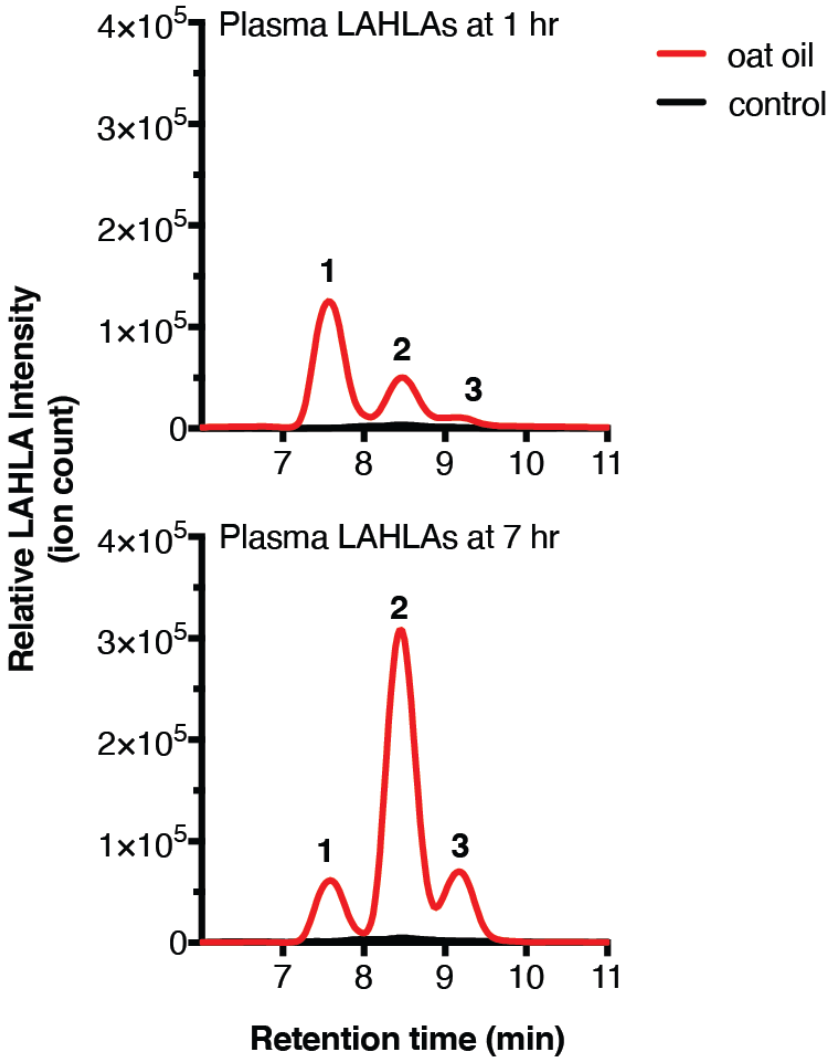


Figure 2.7. Plasma LAHLA chromatograms of liposomal oat oil treated and control treated subjects at 1 (top) and 7 hours (bottom).

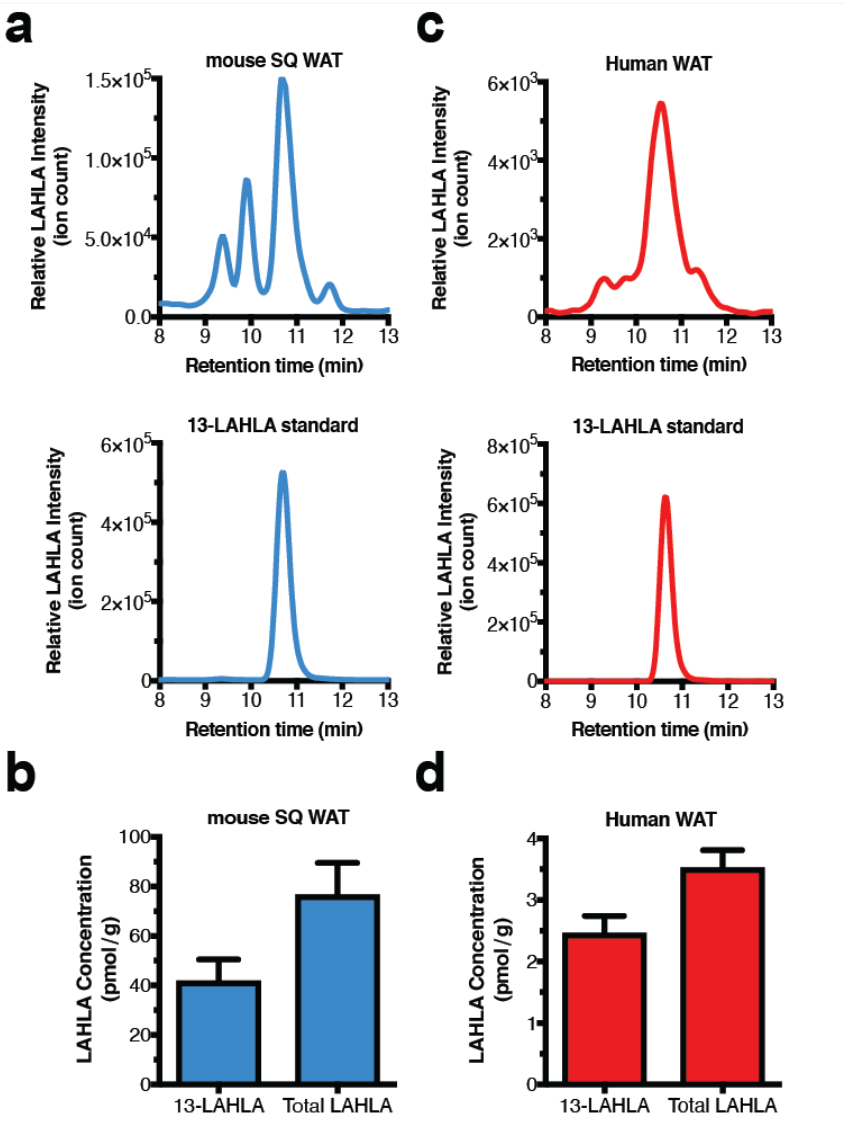


Figure 2.8. LAHLAs are present in mouse and human adipose tissue. (**a**) LC-MS chromatograms of mouse SQWAT LAHLAs (top) and the 13-LAHLA standard (bottom). (**b**) 13-LAHLA and total LAHLA levels in mouse SQWAT. (**c**) LC-MS chromatogram of human WAT LAHLAs (top) and the 13-LAHLA standard (bottom). (**d**) 13-LAHLA and total LAHLA in human WAT. Data represents mean $\pm$ S.E.M (n = 3).

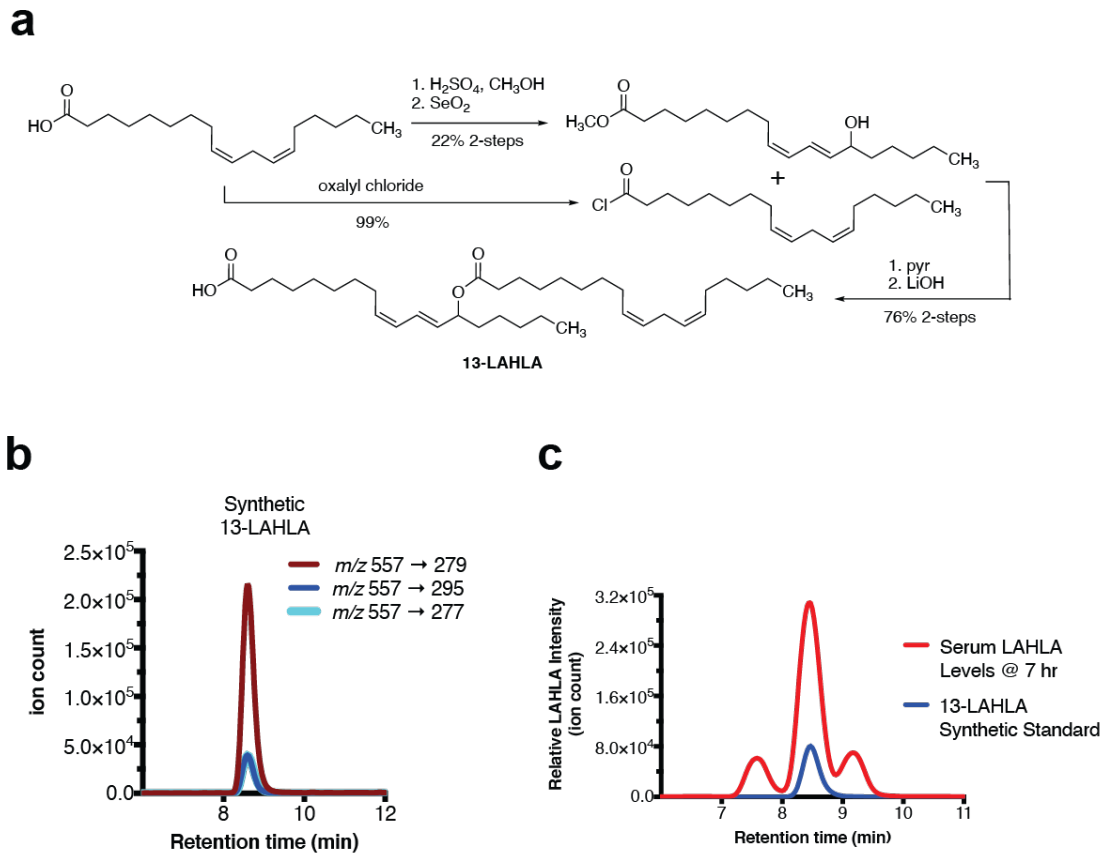


Figure 2.9. Synthesis of 13-LAHLA and confirmation that this is the major isomer in human serum post liposomal oat oil ingestion. (**a**) The synthesis of 13-LAHLA was initiated by conversion of linoleic acid to methyl linoleate under Fisher esterification conditions. Allylic oxidation of methyl linoleate with stoichiometric selenium dioxide afforded methyl 13-hydroxy linoleate. Linoleoyl chloride, prepared from linoleic acid and oxalyl chloride, was combined with methyl 13-hydroxy linoleate to generate the methyl ester of 13-LAHLA. Selective hydrolysis of the methyl ester yielded 13-LAHLA in 17% overall yield starting from linoleic acid. (**b**) MRM of 13-LAHLA (*m/z* 557) showing relative ratios of the three LAHLA-specific transitions. (**c**) Chromatogram of LAHLA isomers in human plasma seven hours after liposomal oat oil ingestion (red) compared to the 13-LAHLA synthetic standard (blue) provided confirmation of this regioisomer.

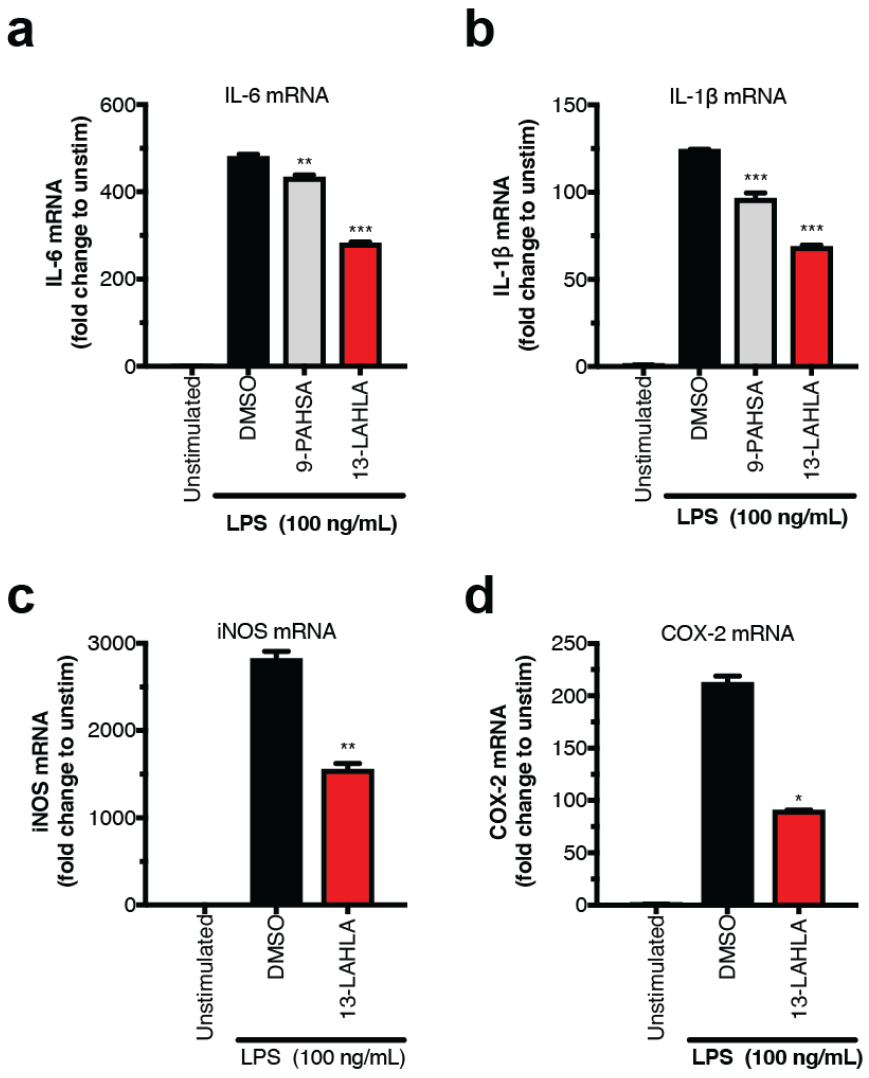


Figure 2.10. Anti-inflammatory effects of 13-LAHLA. (**a-b**) 9-PAHSA (10 μM) and 13-LAHLA (10 μM) reduce IL-6 (**a**) and IL-1β (**b**) mRNA levels in RAW 264.7 macrophages treated with LPS (100 ng/mL). (**c-d**) Effects of 13-LAHLA (10 μM) on iNOS (**c**) and COX-2 (**d**) mRNA levels in RAW 264.7 macrophages treated with LPS (100 ng/mL). Data represents mean ± S.E.M. (n = 3-4). *, $p \leq 0.05$; **, $p \leq 0.001$; ***, $p \leq 0.0001$ by two-sided Student's t-test for FAHFA treated versus LPS treated alone.

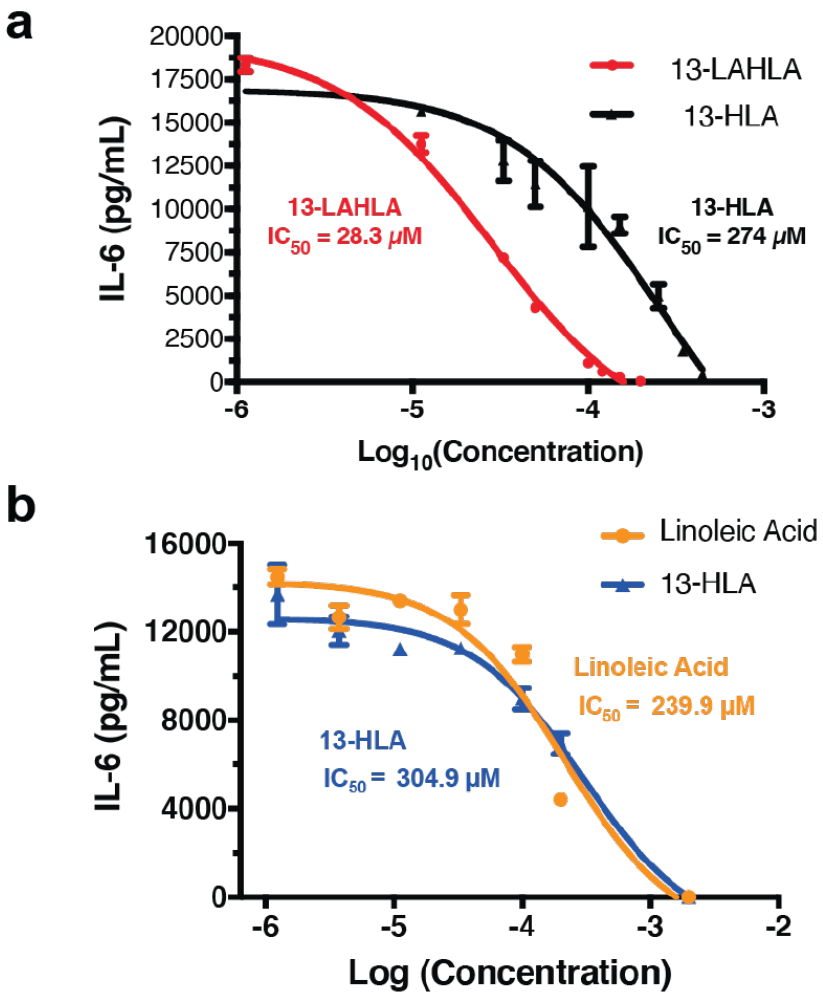


Figure 2.11. 13-LAHLA structure and activity. (**a**) Activity of 13-LAHLA and 13-HLA reveal greater activity for 13-LAHLA. Cells were viable over the entire dose range (data not shown). (**b**) IC_{50}s of linoleic acid and 13-HLA are approximately an order of magnitude worse than that of 13-LAHLA. Cells were viable over the entire dose range (data not shown).

2.6 Materials and Methods

General Information, Procedures on 13-LAHLA Synthesis, and NMR spectra

For general information, procedures on synthesis of LAHLA, and NMR spectra, please refer to "Chang_01_Supplemental_Information1.doc. General information on 13-LAHLA synthesis and NMR spectra for Chapter 2."

Chemicals

9-PAHSA was purchased from Cayman Chemical. 13-LAHLA was synthesized as described in the supporting information.

Oat oils

Ethanol extracted crude oat oil was produced in large scale at Swedish Oat Fiber, Bua, Sweden. This crude oat oil was fractionated using different mixtures of water, ethanol and sugar resulting in the neutral lipids rich T1 fraction, the polar lipid rich T2 fraction, and an oat syrup containing sugar and ethanol soluble oat protein (12). The main lipid classes in the T1 oil were triglycerides (90 wt.%); slightly polar lipids like free fatty acids, sterols, diglycerides and monoglycerides (6 wt.%); galactolipids (2 wt.%); and phospholipids (2 wt.%). The T2 fraction contains triglycerides (40 wt.%); slightly polar lipids like free fatty acids, sterols, diglycerides and monoglycerides (3 wt.%); galactolipids (29 wt.%); and phospholipids (29 wt.%). The SPE-T1 fraction was enriched using a previously reported solid-phase extraction method (40).

Mouse and human tissue for FAHFA measurements

Wild-type C57BL/6J mice were purchased from Jackson Laboratories. All animals were housed in groups on a 14-h light, 10-h dark schedule at the Salk. All animal care and experimental procedures were in accordance with the standing committee on the Use of Animals in Research and Teaching at Salk and Institutional Animal Care and Use Committee (IACUC), and the National Institutes of Health Guidelines for the Humane Treatment of Laboratory Animals. SQWAT was collected immediately after euthanasia and snap frozen using liquid nitrogen.

Human adipose tissue from a 46-year-old African American female was purchased (BioreclamationIVT, Hicksville, NY). Human plasma samples were obtained and prepared as previously described (8). All participants gave their written informed consent prior to the study. The studies were approved by the Regional Human Ethics Committee of Lund-Malmö, Sweden (Registration Number 2010/18, 2011/55). In that study initial blood samples were taken after the subjects fasted overnight. Subsequent blood samples were taken at 1, 3, 5, and 7 hours post-ingestion of a controlled breakfast supplemented with either liposomal oat oil or control diet. The breakfasts contained 35g lipids, i.e. the breakfast with oat oil contained 2mmol DGDG with FAHFA. An IRB letter of exemption for the deidentified human plasma samples used in this study is on file at the Salk Institute for Biological Sciences.

Lipid extraction and SPE

Lipid extraction of samples and solid-phase extraction was performed as previously described (40). Briefly, WAT (150 mg) was dounce homogenized on ice in a mixture of PBS, methanol, and chloroform (1.5 mL/1.5 mL/3 mL). Five pmol of $^{13}C_4$-9-PAHSA was added to the

chloroform prior to lipid extraction as an internal standard. The mixture was vortexed and then centrifuged at 2200 g for 5 min. The organic layer (bottom) was then transferred to a new vial, dried down, and then stored at -80°C for future use. The human plasma was prepared similarly, except no dounce homogenization was performed. SPE was performed at room temperature using a Strata SI-1 silica cartridge (500 mg silica, 3 mL, Phenomenex). The column was washed using 6 mL of ethyl acetate followed by column equilibration with 6 mL hexane. The lipid extract from the prior step was then added to the equilibrated column and neutral lipids were removed using 6 mL of 95:5 hexane:ethyl acetate followed by the elution of FAHFAs with 4 mL ethyl acetate. This eluate was dried down and then subjected to LC-MS analysis.

Cell culture methods

RAW264.7 cells were cultured in RPMI 1640, supplemented with 10% FBS at 37 °C and 5% CO_2.

IL-6 ELISA

RAW 264.7 cells were seeded using a 48-well plate (2.5 X 10^4 cells per well). The next day when cells were ~50% confluent, cells were co-treated with LPS (100 ng/mL) and compound (oat oil extract, FAHFA, 13-HLA, linoleic acid) in a total volume of 200 μL of media at 37°C. After 20 hours, the media was collected and secreted IL-6 was quantified using the mouse IL-6 ELISA MAX^{TM} Deluxe Kits following the manufacturer's protocol (BioLegend).

Cell Proliferation using MTT Assay

After the media was removed for IL-6 measurements, the remaining cells were assessed for their viability using the MTT assay according to the manufacturer's protocol (Calbiochem). Briefly, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was dissolved in sterile PBS (5 mg/mL) and filtered through a 0.2 μm Sterile Millex Filter to prepare a 500 μg/mL solution in RPMI. Prepared solution was added to adherent cells and incubated at 37°C for 4 hours. Sterile DMSO was then supplied upon removal of MTT solution. Relative cell viabilities were quantified using a plate reader at an absorbance of 570 nm.

qRT-PCR

RAW 264.7 cells (6-well plate, 60-70% confluent) were co-treated with FAHFA and LPS (100 ng/mL) or the equivalent amount of DMSO in cell media for 20 hours. Total RNA was isolated using a Purelink RNA Mini Kit (Thermo Fisher), according to the manufacturer's instructions. RNA from each replicate (2 μg) was reverse transcribed to double-stranded cDNA using the QuantiTect Reverse Transcription Kit (Qiagen). qRT-PCR was performed using a LightCycler 480 II (Roche), with SYBR Green qPCR Master Mix (Bimake). qRT-PCR reaction conditions were: 95 °C for 10 min; 35 x (95 °C for 15 s; 55 °C for 30 s; 72 °C for 30 s). The relative quantification of gene expression was calculated by using the $2^{-\Delta\Delta Ct}$ approximation method. The primers used in this study were GAPDH, iNOS1 (41), COX2 (41), IL-6 (6), and IL-1B (6) (IL-1B and IL-6 primer sequences were obtained through personal communication). GAPDH forward 5'-AGG TCG GTG TGA ACG GAT TTG-3', GAPDH reverse 5'-TGT AGA CCA TGT AGT TGA GGT CA-3'; COX2 forward 5'-GGA GAG ACT ATC AAG ATA GT-3', COX2 reverse: 5'-GGA GAG ACT ATC AAG ATA GT-3', iNOS forward 5'-AAT GGC AAC

ATC AGG TCG GCC ATC ACT-3', iNOS reverse 5'-GCT GTG TGT CAC AGA AGT CTC GAA CTC-3'; IL-6 forward 5'- AAC CAC GGC CTT CCC TAC TT-3', IL-6 reverse 5'- GCC ATT GCA CAA CTC TTT TCT C-3'; IL-1B forward 5'- ACC TGG GCT GTC CTG ATG AGA G-3', IL-1B reverse 5'-CCA CGG GAA AGA CAC AGG TAG C-3'.

Targeted LC-MS analysis of FAHFAs

FAHFAs were measured on a TSQ Quantiva LC-MS instrument using multiple reaction monitoring (MRM) in negative ionization mode as previously described (41). Resolution of FAHFAs was achieved using an Acquity UPLC BEH C18 column (1.7 μm, 2.1 mm X 100 mm, Waters) with a flow rate of 0.2 mL/min with a 93:7 (MeOH: H_2O) mobile phase with 5 mM ammonium acetate and 0.03% ammonium hydroxide. LAHLAs were analyzed by monitoring the following precursor-to-product ion transitions, m/z 557.5 → 279.2 and m/z 557.5 → 295.2, which correspond to the parent LAHLA to linoleic acid and LAHLA to hydroxy linoleic acid, respectively.

Pseudo-MS3 for analysis of LAHLA regioisomers

Oat oil (5 μL) was dissolved in 2:1 CHCl3:MeOH (400 μL) and then 5 μL was subjected to LC-MS analysis. LC separation was achieved using a Gemini 5U C-18 column (Phenomenex). Resolution of LAHLAs was achieved using an Acquity UPLC BEH C18 column (1.7 μm, 2.1 mm X 100 mm, Waters) with a flow rate of 0.2 mL/min with a 93:7 (MeOH: H_2O) mobile phase with 5 mM ammonium acetate and 0.03% ammonium hydroxide. MS analysis was performed using a Thermo Scientific Q Exactive Plus fitted with a heated electrospray ionization source. To facilitate in-source fragmentation, the in-source CID was set to 60.0 eV. To attain MS2 data of

HLA, the m/z of the negative HLA ion m/z 295.227 was added to the inclusion list to of masses to be fragmented.

2.7 Acknowledgments

Chapter 2, in full, is a reprint of the material with minor modifications as it appears in Linoleic acid esters of hydroxy linoleic acids are anti-inflammatory lipids found in plants and mammals. *The Journal of biological chemistry* vol. 294,27 (2019). The dissertation author was the third investigator and author of this paper. Other authors include Matthew J. Kolar, Srihari Konduri, Huijing Wang, Clare McNerlin, Lena Ohlsson, Magnus Härröd, Dionicio Siegel, and Alan Saghatelian.

2.8 References

1. T. E. Eling, W. C. Glasgow, Cellular proliferation and lipid metabolism: importance of lipoxygenases in modulating epidermal growth factor-dependent mitogenesis. *Cancer Metastasis Rev.* **13**, 397–410 (1994).

2. S. E. Carlson, Early determinants of development: a lipid perspective. *Am. J. Clin. Nutr.* **89**, 1523S (2009).

3. C. Huang, C. Freter, Lipid Metabolism, Apoptosis and Cancer Therapy. *Int. J. Mol. Sci.* **16**, 924 (2015).

4. D. A. Rudnick, N. O. Davidson, Functional Relationships between Lipid Metabolism and Liver Regeneration. *Int. J. Hepatol.* **2012**, 1–8 (2012).

5. D. B. Savage, K. F. Petersen, G. I. Shulman, No Title. *Physiol. Rev.* **87** (2007), doi:10.1152/PHYSREV.00024.2006.

6. E. E. Blaak, Fatty acid metabolism in obesity and type 2 diabetes mellitus. *Proc. Nutr. Soc.* **62**, 753–760 (2003).

7. J. Soppert, M. Lehrke, N. Marx, J. Jankowski, H. Noels, Lipoproteins and lipids in cardiovascular disease: from mechanistic insights to therapeutic targeting. *Adv. Drug Deliv. Rev.* **159**, 4–33 (2020).

8. Y. C. Kao, P. C. Ho, Y. K. Tu, I. M. Jou, K. J. Tsai, Lipids and Alzheimer's Disease. *Int. J. Mol. Sci.* **21**, 1505 (2020).

9. M. M. Yore, I. Syed, P. M. Moraes-Vieira, T. Zhang, M. A. Herman, E. A. Homan, R. T. Patel, J. Lee, S. Chen, O. D. Peroni, A. S. Dhaneshwar, A. Hammarstedt, U. Smith, T. E. McGraw, A. Saghatelian, B. B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell*. **159**, 318–332 (2014).

10. P. R. Shepherd$, L. Gnudi, E. Tozzo, H. Yang, F. Leach, B. B. Kahn$, Adipose Cell Hyperplasia and Enhanced Glucose Disposal in Transgenic Mice Overexpressing GLUT4 Selectively in Adipose Tissue*. *J. Biol. Chem.* **268**, 22243–22246 (1993).

11. E. Carvalho, K. Kotani, O. D. Peroni, B. B. Kahn, Adipose-specific overexpression of GLUT4 reverses insulin resistance and diabetes in mice lacking GLUT4 selectively in muscle. *Am. J. Physiol. - Endocrinol. Metab.* **289**, 551–561 (2005).

12. D. Y. Oh, S. Talukdar, E. J. Bae, T. Imamura, H. Morinaga, W. Q. Fan, P. Li, W. J. Lu, S. M. Watkins, J. M. Olefsky, GPR120 is an Omega-3 Fatty Acid Receptor Mediating Potent Anti-Inflammatory and Insulin Sensitizing Effects. *Cell*. **142**, 687 (2010).

13. H. Cao, K. Gerhold, J. R. Mayers, M. M. Wiest, S. M. Watkins, G. S. Hotamisligil, Identification of a Lipokine, a Lipid Hormone Linking Adipose Tissue to Systemic

Metabolism. *Cell*. **134**, 933 (2008).

14. I. Syed, J. Lee, P. M. Moraes-Vieira, C. J. Donaldson, A. Sontheimer, P. Aryal, K. Wellenstein, M. J. Kolar, A. T. Nelson, D. Siegel, J. Mokrosinski, I. S. Farooqi, J. J. Zhao, M. M. Yore, O. D. Peroni, A. Saghatelian, B. B. Kahn, Palmitic Acid Hydroxy Stearic Acids Activate GPR40 Which is Involved in Their Beneficial Effects on Glucose Homeostasis. *Cell Metab.* **27**, 419 (2018).

15. J. Lee, P. M. Moraes-Vieira, A. Castoldi, P. Aryal, E. U. Yee, C. Vickers, O. Parnas, C. J. Donaldson, A. Saghatelian, B. B. Kahn, Branched fatty acid esters of hydroxy fatty acids (FAHFAs) protect against colitis by regulating gut innate and adaptive immune responses. *J. Biol. Chem.* **291**, 22207–22217 (2016).

16. O. Kuda, M. Brezinova, M. Rombaldova, B. Slavikova, M. Posta, P. Beier, P. Janovska, J. Veleba, J. Kopecky, E. Kudova, T. Pelikanova, Docosahexaenoic acid-derived fatty acid esters of hydroxy fatty acids (FAHFAS) with anti-inflammatory properties. *Diabetes*. **65**, 2580–2590 (2016).

17. K. A. Reynertson, Garay. M., J. Nebus, S. Chon, S. Kaur, K. Mahmood, M. Kizoulis, M. D. Southall, Anti-Inflammatory Activities of Colloidal Oatmeal (Avena sativa) Contribute to the Effectiveness of Oats in Treatment of Itch Associated With Dry, Irritated Skin - JDDonline - Journal of Drugs in Dermatology. *J. drugs dermatology* (2015), pp. 43–8.

18. K. Capone, F. Kirchner, S. L. Klein, N. K. Tierney, Effects of Colloidal Oatmeal Topical Atopic Dermatitis Cream on Skin Microbiome and Skin Barrier Properties. *J. Drugs Dermatol.* **19**, 524–531 (2020).

19. L. Ohlsson, A. Rosenquist, J. F. Rehfeld, M. Härröd, Postprandial effects on plasma lipids and satiety hormones from intake of liposomes made from fractionated oat oil: Two randomized crossover studies. *Food Nutr. Res.* **58** (2014), doi:10.3402/fnr.v58.24465.

20. M. Hamberg, E. Liepinsh, G. Otting, W. Griffiths, Isolation and structure of a new galactolipid from oat seeds. *Lipids*. **33**, 355–363 (1998).

21. R. A. Moreau, D. C. Doehlert, R. Welti, G. Isaac, M. Roth, P. Tamura, A. Nuñez, The Identification of Mono-, Di-, Tri-, and Tetragalactosyldiacylglycerols and their Natural Estolides in Oat Kernels. *Lipids*. **43**, 533 (2008).

22. M. Härröd, US8865923B2 - Method for separating neutral and polar lipids and an oil rich in polar lipids - Google Patents, (available at https://patents.google.com/patent/US8865923B2/en).

23. T. Zhang, S. Chen, I. Syed, M. Ståhlman, M. J. Kolar, E. A. Homan, Q. Chu, U. Smith, J. Borén, B. B. Kahn, A. Saghatelian, A LC-MS–based workflow for measurement of branched fatty acid esters of hydroxy fatty acids. *Nat. Protoc. 2016 114*. **11**, 747–763 (2016).

24. A. T. Nelson, M. J. Kolar, Q. Chu, I. Syed, B. B. Kahn, A. Saghatelian, D. Siegel, Stereochemistry of Endogenous Palmitic Acid Ester of 9-Hydroxystearic Acid and Relevance of Absolute Configuration to Regulation. *J. Am. Chem. Soc.* **139**, 4943 (2017).

25. L. Balas, J. Bertrand-Michel, F. Viars, J. Faugere, C. Lefort, S. Caspar-Bauguil, D. Langin, T. Durand, Regiocontrolled syntheses of FAHFAs and LC-MS/MS differentiation of regioisomers. *Org. Biomol. Chem.* **14**, 9012–9020 (2016).

26. C. Chan, P. B. Cox, S. M. Roberts, Convergent stereocontrolled synthesis of 13-hydroxy-9Z, 11E-octadecadienoic acid (13-HODE). *J. Chem. Soc. Chem. Commun.*, 971–972 (1988).

27. A. V. R. Rao, E. R. Reddy, G. V. M. Sharma, P. Yadagiri, J. S. Yadav, A stereoselective synthesis of coriolic acid and dimorphecolic acid. *Tetrahedron Lett.* **26**, 465–468 (1985).

28. A. V. Rama Rao, S. P. Reddy, E. R. Reddy, Short and Efficient Syntheses of Coriolic Acid. *J. Org. Chem.* **51**, 4158–4159 (1986).

29. J. K. STILLE, M. P. SWEET, Palladium-Catalyzed Coupling Reactions of 1-(Tributylstannyl)-1-octen- 3-ol. *ChemInform*. **6**, 3189–3191 (1991).

30. J. S. Yadav, P. K. Deshpande, G. Sharma, Stereoselective synthesis of (S)-13-hydroxy octadeca- (9Z, 11E)-di- and (9Z, 11E, 15Z)-trienoic acids : selfdefensive substances against rice blast disease. *Tetrahedron*. **48**, 4465–4474 (1992).

31. Z. Li, V. H. Tran, R. K. Duke, M. C. H. Ng, D. Yang, C. C. Duke, Synthesis and biological activity of hydroxylated derivatives of linoleic acid and conjugated linoleic acids. *Chem. Phys. Lipids*. **158**, 39–45 (2009).

32. M. J. Kolar, S. S. Kamat, W. H. Parsons, E. A. Homan, T. Maher, O. D. Peroni, I. Syed, K. Fjeld, A. Molven, B. B. Kahn, B. F. Cravatt, A. Saghatelian, Branched Fatty Acid Esters of Hydroxy Fatty Acids Are Preferred Substrates of the MODY8 Protein Carboxyl Ester Lipase. *Biochemistry*. **55**, 4636–4641 (2016).

33. W. H. Parsons, M. J. Kolar, S. S. Kamat, A. B. Cognetta, J. J. Hulce, E. Saez, B. B. Kahn, A. Saghatelian, B. F. Cravatt, AIG1 and ADTRP are atypical integral membrane hydrolases that degrade bioactive FAHFAs. *Nat. Chem. Biol.* **12**, 367–372 (2016).

34. C. D. Funk, Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science*. **294**, 1871–1875 (2001).

35. J. M. Lehmann, S. A. Kliewer, L. B. Moore, T. A. Smith-Oliver, B. B. Oliver, J. L. Su, S. S. Sundseth, D. A. Winegar, D. E. Blanchard, T. A. Spencer, T. M. Willson, Activation of the nuclear receptor LXR by oxysterols defines a new hormone response pathway. *J. Biol. Chem.* **272**, 3137–3140 (1997).

36. Y. Ma, T. Kind, A. Vaniya, I. Gennity, J. F. Fahrmann, O. Fiehn, An in silico MS/MS

library for automatic annotation of novel FAHFA lipids. *J. Cheminform.* **7**, 1–5 (2015).

37. T. Hu, M. Lin, D. Zhang, M. Li, J. Zhang, A UPLC/MS/MS method for comprehensive profiling and quantification of fatty acid esters of hydroxy fatty acids in white adipose tissue. *Anal. Bioanal. Chem.* **410**, 7415–7428 (2018).

38. Q. F. Zhu, J. W. Yan, T. Y. Zhang, H. M. Xiao, Y. Q. Feng, Comprehensive Screening and Identification of Fatty Acid Esters of Hydroxy Fatty Acids in Plant Tissues by Chemical Isotope Labeling-Assisted Liquid Chromatography-Mass Spectrometry. *Anal. Chem.* **90**, 10056–10063 (2018).

39. M. Criquet, R. Roure, L. Dayan, V. Nollent, C. Bertin, Safety and efficacy of personal care products containing colloidal oatmeal. *Clin. Cosmet. Investig. Dermatol.* **5**, 183–193 (2012).

40. E. S. Kurtz, W. Wallo, Colloidal oatmeal: history, chemistry and clinical properties. *J. Drugs Dermatol.* **6**, 167–170 (2007).

41. F. W. Collins, Oat Phenolics: Avenanthramides, Novel Substituted N-Cinnamoylanthranilate Alkaloids from Oat Groats and Hulls. *J. Agric. Food Chem.* **37**, 60–66 (1989).

42. P. Mattila, J. M. Pihlava, J. Hellström, Contents of phenolic acids, alkyl- and alkenylresorcinols, and avenanthramides in commercial grain products. **53**, 8290–8295 (2005).

43. D. M. Peterson, M. J. Hahn, C. L. Emmons, Oat avenanthramides exhibit antioxidant activities in vitro. *Food Chem.* **79**, 473–478 (2002).

44. C. Y. Chen, P. E. Milbury, H. K. Kwak, F. W. Collins, P. Samuel, J. B. Blumberg, Avenanthramides and phenolic acids from oats are bioavailable and act synergistically with vitamin C to enhance hamster and human LDL resistance to oxidation. *J. Nutr.* **134**, 1459–1466 (2004).

45. J. MacMicking, Q. W. Xie, C. Nathan, Nitric oxide and macrophage function. *Annu. Rev. Immunol.* **15**, 323–350 (1997).

46. T. J. Guzik, R. Korbut, T. Adamek-Guzik, Nitric oxide and superoxide in inflammation and immune regulation. *J. Physiol. Pharmacol.* **54.4**, 469–87 (2003).

47. A. G. Eliopoulos, C. D. Dumitru, C. C. Wang, J. Cho, P. N. Tsichlis, Induction of COX-2 by LPS in macrophages is regulated by Tpl2-dependent CREB activation signals. *EMBO J.* **21**, 4831–4840 (2002).

CHAPTER 3

Endogenous Fatty Acid Esters of Hydroxy Fatty Acids, Starting Points for New Chemical Regulators of Inflammation

Abstract

Fatty Acid esters of Hydroxy Fatty Acids (FAHFAs) are a growing class of endogenous lipids possessing ester conjoined fatty acids. Members of these lipids lead to improvements in models of dysfunctional metabolism and inflammation. These endogenous lipids provide an ideal starting point for the discovery and development of new inflammation regulating small molecules. While similar to corticosteroids in that their discovery was based on endogenous compounds, FAHFAs and related analogues function through a different mechanism than corticosteroids providing an opportunity to overcome the limitations of existing anti-inflammatory agents. In line with developing natural compounds as therapeutics FAHFAs possess good pharmacodynamics but have several pharmacokinetic (PK) liabilities arising from their structures that hinder development. The biologically reactive functionality of FAHFAs present unique challenges with the carboxylic acid leading to the incorporation of FAHFAs into triglycerides and the ester undergoing hydrolysis reactions, both reactions inactivating FAHFAs' inflammatory regulating effects. Conversion of these functionality to stable variants through structural modifications has been found to have the added benefit of improved potency. With PK analysis and *in vivo* testing against inflammatory bowel disease these new analogues provide a path forward in creating new synthetic lipids to control inflammation.

3.1 Introduction

Studies of adipose-selective overexpression of glucose transporter 4 (Glut4) mice (AG4OX) revealed a class of endogenous lipids with anti-diabetic and inflammation regulating activities called Fatty Acid esters of Hydroxy Fatty Acids (FAHFAs). Paradoxically, AG4OX mice are obese but manifest favorable metabolic profiles (*1*). Lipidomic analysis of the adipose tissues from these mice identified FAHFAs as an upregulated lipid class, driving the hypothesis that endogenous FAHFAs are a class of beneficial lipids with physiological effects. Assessment of the biological activities of FAHFAs *in vitro* and *in vivo* have confirmed their activity against glucose resistance and inflammation (*1–5*). To date, at least 20 different FAHFA families have been identified in mammalian tissues endogenously, aside from AT, including liver, heart, kidneys, and pancreas in mammals, and have been found in many plants and foods (*2–12*).

While the FAHFA lipid class includes more than 80 distinct FAHFA members, the PAHSA family, specifically 5- and 9-PAHSA, has been the most extensively studied in a physiological context (*1–4*). Functional studies demonstrate that oral gavage of 9-PAHSA in high fat diet (HFD)-fed mice suppresses the mRNA levels of proinflammatory markers in AT macrophages including tumor necrosis factor alpha (TNFα) and interleukin-1β (IL-1β) compared to their counterparts gavaged with vehicle. Moreover, these effects were translated to protect mice from development of colitis, a disease arising from recurring inflammation of the inner lining of the colon and microbial dysbiosis (*13*). Oral dosing of a mixture of 5-PAHSA and 9-PAHSA in colitis-induced mice led to reduced T cell activation and abrogation of inflammatory chemokine and cytokine expression, protecting the mice from mucosal damage and development of colitis (*2*). In addition to 9-PAHSA, other FAHFA members, such as 13-DHAHLA (*5*) and 13-LAHLA (*6*) are consistently regulators of inflammation. Specifically, treatment of both lipids

individually inhibited the secretion of proinflammatory cytokines and attenuated the mRNA levels of inflammatory markers in LPS-stimulated RAW264.7 macrophages. As a result, FAHFAs are a novel class of bioactive lipids with diverse structural and beneficial roles in mammalian physiology.

Following the discovery of endogenous FAHFAs, further characterization of these lipids has been achieved through chemical synthesis. Simultaneous comparison of synthetic FAHFAs, such as 9-PAHSA with 13-DHAHLA (*5*, *14*, *15*) or with 13-LAHLA (*6*), reveals that both lipids showed superior inflammation-reducing activities than 9-PAHSA. This suggests a trend that unsaturated FAHFAs may be more active the saturated ones. Synthetic 9-PAHSA and 9-PAHSA analogs with modifications on the hydrocarbon lengths of either the 9-hydroxystearic acid or the palmitic acid hydrocarbon chain enabled a direct screen of their anti-inflammatory activities and demonstrated that an appropriate removal of hydrocarbons could maintain activity while reducing the molecular weight (*16*). Furthermore, another activity screen to investigate the anti-inflammatory activities within the LAHLA family was accomplished by the syntheses of LAHLA regioisomers and enantiopure stereoisomers (*17*). This assay determined that the position of the ester linkage influences the anti-inflammatory effects of 9-, 13, and 15-LAHLA, with 15-LAHLA being the most potent regioisomer (*6*). Meanwhile, the enantiopure S-15-LAHLA is slightly more active than the enantiopure R-15-LAHLA (*6*). Together, these results highlight the presence of a structure-activity relationship within the FAHFA family. Therefore, this report will underscore the value of chemical biology and assesses how chemical modifications of FAHFA structures can modulate their anti-inflammatory potencies.

One hindrance in studying FAHFAs or other endogenous lipids is lipid metabolism. FAHFAs are not only susceptible to endogenous degradation by hydrolysis enzymes (*6*, *18*, *19*),

FAHFAs are also established substrates for triglycerides biosynthesis (*20*). Specifically, FAHFAs can be incorporated to form endogenous FAHFA-containing triglycerols (FAHFA-TGs), and this reaction is likely due to the presence of carboxylic acid functional groups in a FAHFA structure (*20*). In fact, it was previously reported that carboxylic acid functional groups, though a critical constituent of a pharmacophore, is known to diminish metabolic stability (*21*). Therefore, designing stable FAHFA derivatives that are less or not susceptible to endogenous metabolism may convert natural FAHFAs into more potent regulators of inflammation and potentially enhance their favorable metabolic effects on mammalian physiology. To better understand the structure-activity relationship (SAR) in FAHFAs, we created a library of synthetic FAHFA analogs to explore the effects of the key functional groups in their structure on their activity. Due to the undesirable chemicophysical properties associated with carboxylic acid moieties, previous exploratory structural changes by replacing carboxylic acid groups with carboxylic acid bioisosteres were attempted (*21*, *22*). Because bioisosteres substitution, such as acyl-sulfonamides, demonstrated improved chemicophysical properties, we reason that acyl-sulfonamides could be used as a carboxylic acid surrogate in FAHFA structures to optimize the potency of FAHFAs. Thus, we report herein the SAR of FAHFAs relevant to the functionality, namely the ester linkage connecting the two acyl chains (**SAR1**) and acyl-sulfonamides (**SAR2**) as well as a cursory examination of the fatty acid chain was achieved (**SAR3**).

Screening of the anti-inflammatory activities of FAHFAs has been accomplished extensively using lipopolysaccharide (LPS)-stimulated murine RAW 264.7 macrophages (*5*, *6*, *16*, *17*). Proinflammatory cytokines, such as interleukin-6 (IL-6) and interleukin-1β (IL-1β) are mediators of stress following toll-like receptor 4 (TLR4) activation by LPS, an endotoxin derived from Gram-negative bacteria (*23–25*). In these cells, quantification of the suppression of LPS-

induced proinflammatory cytokines release following treatments of test compounds allows one to determine the efficiency of inflammation blockade by the compounds under acute stress. Therefore, using proinflammatory cytokine IL-6 as an indicator of inflammation in our *in-vitro* screen, synthetic FAHFA analog were screened for their ability to attenuate IL-6 production in LPS-stimulated RAW 264.7 macrophages. In support of this finding, we subsequently tested the most active FAHFA bioisosteres in human peripheral blood mononuclear cells (PBMCs) and a murine model of colitis to understand whether these anti-inflammatory effects are translatable. Together, an understanding of the SARs within FAHFAs is needed to help advance the compounds clinically as well as provide a better understanding of their biosynthesis, metabolism, and endogenous functions.

3.2 Results and Discussion

Chemistry:

The functionality was incorporated early on as focus on the metabolic liability of the ester was examined (SAR1). FAHFA degrading enzymes that hydrolyzed the ester bond linking the hydroxy fatty acid and fatty acid have been characterized[18] and esters are a well-established metabolic liability in drug development. The chemistry required to make diverse derivatives was extensive, however, nine variations replacing the ester were made enabling the identification of improved derivatives with different conjoining functionality.

Starting from 9-PAHSA, synthesized using the established protocol (*16*), the methane sulfonamide **2** analogue was prepared as a benchmark for comparison given the improved activity of the acylsulfonamides. Through 1-ethyl-3 (3-dimethylpropylamine) carbodiimide (EDCI) coupling the acyl-methanesulfonamide of 9-PAHSA (**2**) was generated in 84% yield

(**Scheme 3.1**). The syntheses of ester replacements were initiated by Dess Martin periodinane oxidation of methyl 9-hydroxyoctadecanoate **1** to yield ketone **3** in 88% yield, enabling reductive amination followed by EDCI coupling to form amide **4** in 75% yield over two steps. Treating amide **4** with $LiOH \cdot H_2O$ delivered carboxylic acid intermediate, which was reacted with a methane sulfonamide and EDCI to generate acylsulfonamide **5**. The acylsulfonamide/sulfonamide **6** and *N*-methyl amide derivative **8** were synthesized by similar sequences.

The synthesis of "reverse ester" of 9-PAHSA switching the connectivity of the ester was started by with treating undec-10-enoic acid **9** with LDA/DMPU/THF followed by alkylation of the enolate with 1-iodononane **10 (Scheme 3.2**). The successful alkylation afforded the acid **11** in 90% yield, which was esterified with 1-pentadecanol to provide ester **12**. Oxidative cleavage of the terminal alkene of **12** furnished carboxylic acid that was made to react with methanesulfonamide and EDCI to yield "reverse ester" analog **13**. Following a similar synthetic sequence, 1-pentadecanol was replaced by N-methylpentadecan-1-amine **14** to give "reverse *N*-methyl amide" analog **17** in 61% yield over three steps.

As shown in **Scheme 3.3**, the synthesis of analog **23** bearing an 1,2,4-oxadiazol in place of the ester began with mono-protection (*26*) of diol **18** to yield, after two additional steps, nitrile **19** in 73% yield over three steps. The α-alkylation of the anion of nitrile **19** with 1-bromononane formed **20** in 91% yield. The three-step sequence, silicon cleavage, oxidation, and methyl ester formation, afforded **21** in 66% yield. Nitrile **21** was transformed into acyl amidoxime, which underwent conversion to 1,2,4-oxadiazol **22**. Hydrolysis of the methyl ester followed by EDCI coupling with methane sulfonamide afforded analog **23** in 70% yield over two steps.

The synthesis of indole **27** began N-alkylation followed by phenolic demethylation of methyl ether **24** to provide phenol **25** in 52% yield over two steps (**Scheme 3.4**). Phenolic alkylation with 1-bromoheptane with NaH in DMF formed **26**. The targeted indole analogues **27** was obtained by hydrolysis of methyl ester followed by EDC coupling with methanesulfonamide.

The synthesis of analog **33** possessing an aryl core was outlined in **Scheme 3.5**, beginning with the alkylation of phenol **28** to yield aryl ether **29** in 83%. Ether **29** was reacted with N,O-dimethylhydroxylamine in the presence of trimethyl aluminum to provide the Weinreb amide derivative, which underwent mono Grignard addition to give ketone **30** in 62% yield over two steps. Subsequent Grignard addition into the ketone of **30** with non-8-en-1-ylmagnesium bromide yield tertiary alcohol **31**. The activated alcohol was reduced with TFA/Et_3SiH to yield **32**. Osmium tetroxide catalytic oxidative cleavage using Oxone® of **32** at 23 °C generated the carboxylic acid which was transformed to acylsulfonamide **33** in 43% yield over two steps.

Cursory inspection of the importance of unsaturation in the fatty acid region was achieved by incorporating one or two olefins in the hydrocarbon (**Scheme 3.6**). With carboxylic acid **11**, protection of carboxylic acid with a benzyl ester gave **34** in 93% yield. The alkene of **34** was transformed to acylsulfonamide **35** in 69% yield by oxidative cleavage and EDCI coupling. Reductive removal of the benzyl ester provided carboxylic acid **36**. Coupling of the methylamines **37** and **39** generated the desired products **38** and **40**, respectively.

To assess "reverse-amide" analogs with sulfonamides a large set was synthesized quickly to assess the importance of substitution on the sulfonamide (**Scheme 3.7**). A wide range of sulfonamides were selected to replace the carboxylic acid group of "reverse-amide" 9-PAHSA (**16**), delivering analog **41-68** via an EDCI coupling reactions.

Structure-Activity Relationships for 9-PAHSA:

To establish the SAR across analogues, we screened the compounds for their inflammation regulating properties in RAW264.7 macrophages. Upon LPS stimulation, the cells were simultaneously treated with DMSO or individual compounds at 25 μM for compounds in **Table 3.1** and **Table 3.3**, and at 10 μM for compounds in **Table 3.2**. Following a 20-hour treatment, cell media was subsequently collected to quantify the level of IL-6 (*2*, *5, 6, 16-17*). With 100% relative IL-6 indicating no inhibition of IL-6 at all, the anti-inflammatory properties of analogues were assessed by their ability to attenuate IL-6 upon LPS-induced acute stress in RAW264.7 macrophages (**Figure 3.3**). As shown in **Table 3.1**, replacement of the ester group at C9 position with an amide group (**5**) or a sulfonamide group (**6**) leads to less IL-6 attenuation at 86% and 59% when compared with the acylsulfonamide of 9-PAHSA (**2**) at 34%. However, the activity can be enhanced when the core structure is substituted with an N-methyl amide (**8**). Furthermore, compound (**13**), generated by reversing the ester group at C9 position, improved the relative IL-6 compared to the starting compound (**2**). The reverse-N-methyl amide at the core (**17**) delivered a derivative that was comparably in activity of reverse ester (**13**) with predicted improved stability. In addition, three cyclic core structures (**23**, **27**, and **33**) were also explored. The indole core (**27**) retained potency roughly equivalent to the starting ester (**2**) while aryl derivative (**33**) performed modestly. The 1,2,4-oxadiazole (**23**), however, was the most active cyclic core derivative tested.

While both **13** and **17** demonstrated outstanding inflammation reducing effects, cells treated with compound **17** retained a higher cell viability than those with **13**, indicating that some cytotoxicity could be attributed to **13** (data not shown). As our best knowledge, amidated FAHFA analogs are resistant to hydrolysis by the FAHFA lipase carboxyl ester lipase (CEL)

(*19*), which was demonstrated by an *in vitro* hydrolysis assay for linoleic acid esters of 13-hydroxy linoleic acid (13-LAHLA) and 13-aza-LAHLA (**Figure 3.4**). Thus, in addition to predicted improved stability N-methyl amide **17** was selected as the structure to use in examining modification on the sulfonamide (**Table 3.2**).

The synthesis of 29 analogues investigated the role of the sulfonamide group (**Scheme 3.7**, **SAR2**). Using an initial dose of 25 μM in testing **Table 3.2** analogues revealed that numerous compounds demonstrated potent effects in reducing IL-6 levels under acute inflammation (data not shown). To further examine the properties of individual analogues and to establish the SAR at the most optimal resolution, a lower dosage of 10 μM was used in our next screen (**Table 3.2**).

As shown in **Table 3.2**, replacing methyl group (**17**) with other groups can enhance IL-6 inhibition (**17**, **41-45**). However, treatment of sulfonamides (**61** and **68**) did not attenuate IL-6 as effectively as other sulfonamide analogues in **Table 3.2**. Interestingly, the incorporation of a phenyl group (**46**) showed an improved ability in hampering IL-6 relative to **17** and any alkyl substituted derivatives. Nonetheless, an addition of an ortho methyl group in a phenyl (**47**) reduced the potency when compared with phenyl alone (**46**). Compounds with halogens within the aromatic ring (**48-50**, **58**, and **59**) all demonstrated comparable activities with phenyl alone (**56**), with the ortho fluorine (**48**) outperforming the compared analogs. Other electron-withdraw groups (**51-57**) also exhibited similar activities with halogens within the aromatic rings (**48-50, 58 and 59**), with meta, ortho, and para substitutions being comparably active. Higher variations in IL-6 suppression were observed in heterocyclic substitution (**62-67**), with the 4-pyridyl derivative (**64**) retaining the highest ability in attenuating IL-6.

Unsaturated FAHFAs, such as linoleic acid ester of hydroxy linoleic acids (LAHLAs) and docosahexaenoic acid ester of 13-hydroxy linoleic acid (13-DHAHLA), exhibited superior anti-inflammatory activities than saturated FAHFAs, such as the 9-PAHSA (*5*, *6*, *17*). This trend continued in synthetic analogues with the two unsaturated analogs **38** and **40** containing 1 and 2 double bonds on the FA chain, respectively, relative to **17**. The synthesized unsaturated derivates demonstrated improved activities (**Table 3.3**, **SAR3**).

Consistently, our data suggests that analogue **17** carries potent inflammation regulating properties in reducing IL-6 in LPS-stimulated RAW264.7 macrophages. Next, we investigated if analogue **17** could have similar translational effects in human primary cells. Using human peripheral blood mononuclear cells (PBMCs) from healthy donors, we evaluated whether treatments of analogue **17** could also effectively abrogate the release of proinflammatory markers in these antigen-presenting cells under acute stress. In addition to IL-6, we included quantitation of another proinflammatory cytokine, interleukin-1β (IL-1β), as it has been demonstrated to play a role in initiating auto-immune responses and an indication for chronic inflammation such as, inflammatory bowel disease (IBD) (*27–29*). Human PBMCs were subjected to a 1-hour pretreatment of 17 at 1 μM and 25 μM prior to LPS stimulation and cytokine quantification. Our results indicate that **17**, at both doses, did not attenuate IL-6 release significantly, but selectively inhibited the secretion of IL-1β by more than 20% compared to cells that underwent vehicle treatment (**Figure 3**). Though IL-6 reduction was not as prominent as observed in RAW264.7 cells, this result was not surprising to us as human PBMCs were previously demonstrated to elicit a higher net production of IL-1β than IL-6 upon LPS stimulation (*30*).

As we confirmed that **17** retained translational effects in reducing inflammation in murine macrophages and human PBMCs, we performed an *in vivo* pharmacokinetic study to assess its

clinical value. Balb/c mice (*n*=6) were dosed with **17** via intraperitoneal injection at a dose of 20 mg/kg (**Figure 4**). Blood samples were obtained by retro-orbital sinus puncture at 0.5, 1, 2, 4, 8, and 24 hours post-dose. Quantitative determination of **17** in mouse plasma was achieved by liquid chromatography with tandem mass spectrometry (LC-MS-MS). The maximum plasma drug concentration (C_{max}), time to reach C_{max} (T_{max}), area under the plasma concentration-time curve from time zero to 24 hours (AUC_{0-24}), terminal elimination half-life ($t_{1/2}$), clearance (CL), steady state volume of distribution (Vss) were estimated using noncompartmental methods (Phoenix WinNonlin, version 6.3; Certara, Princeton, NJ, USA). **17** was absorbed from the peritoneum in all mice and detectable at the earliest time point (0.5 hours post-dose). The median (interquartile range; IQR) C_{max} was 1.45 μg/mL (1.03 – 2.05) at a T_{max} of 4 hours. Overall systemic exposure (AUC_{0-24}) ranged from 6.49 – 20.80 μg*hr/mL. The median (IQR) $t_{1/2,}$ CL, and Vss were 2.70 hours (1.48 – 4.09), 48.65 mL/hr (24.72 – 53.36), and 130.5 mL (105.6 – 230.6), respectively. These favorable pharmacokinetic results supported advancing **17** forward for oral dosing in a mouse dextran sodium sulfate (DSS) model of colitis, an inflammation model previously used evaluate 5- and 9-PAHSA's protective effects against colitis (*2*).

Using the same design as the PAHSA assays in mouse colitis models, we treated C57BL/6J mice with vehicle (50% PEG400/ 0.5% Tween 80/ 49.5% H_2O) or **17 (**15 mg/kg) via oral gavage for three days followed by another 10 days of DSS administration to induce colitis (*2*) (**Figure 5**). As mice were weighed daily, a clinical colitis score was assigned to each animal to monitor the severity of disease in the treated animals (**Figure 5a-b**). On days 12 and 13, DSS-treated mice that received **17** were more resistant to weight loss than vehicle-treated mice (**Figure 5a**). These results were consistent with a lower clinical colitis score in **17**-treated mice when compared to vehicle-treated mice from days 9-13 (**Figure 5b**). At our terminal timepoint

on day 13, we measured the colon weight of sacrificed mice and found that **17**-treated mice **17** had heavier colons (red bar) than vehicle-treated mice (black bar), indicating that treatment of **17** prevented DSS-induced colon reduction (**Figure 5c**). These preliminary studies verify our ability to identify orally active novel synthetic compounds that are effective in colitis mouse models.

3.3 Conclusion

Endogenous FAHFAs possessing inflammation regulating activities are promising natural small molecules with therapeutic potential. However, their physiological levels and favorable effects may be diminished due to endogenous metabolism or their involvement in biosynthetic pathways. Therefore, exploitation of synthetic FAHFA derivatives to intercept the same pathways involving FAHFAs establishes a new avenue to launch potent, non-steroidal drugs to meet the current limitations of anti-inflammatory agents, including dose limiting side effects and corticosteroids resistance. Assessment of the structure-activity relationships of FAHFA related to IL-6 inhibition has demonstrated that FAHFA analogues can improve the anti-inflammatory activities of natural FAHFAs. Our central compound of this study, *N*-methyl amide reverse 9-PAHSA methyl sulfonamide (**17**) exemplifies that bioisosteres can be synthetically incorporated into FAHFA structures to modify and enhance the biological activities of these natural lipids *in* vitro and *in vivo*. In this study, our design and characterization of synthetic FAHFA analogs provide additional starting points for the development of new agents to reduce inflammation.

3.4 Figures, Schema, and Tables

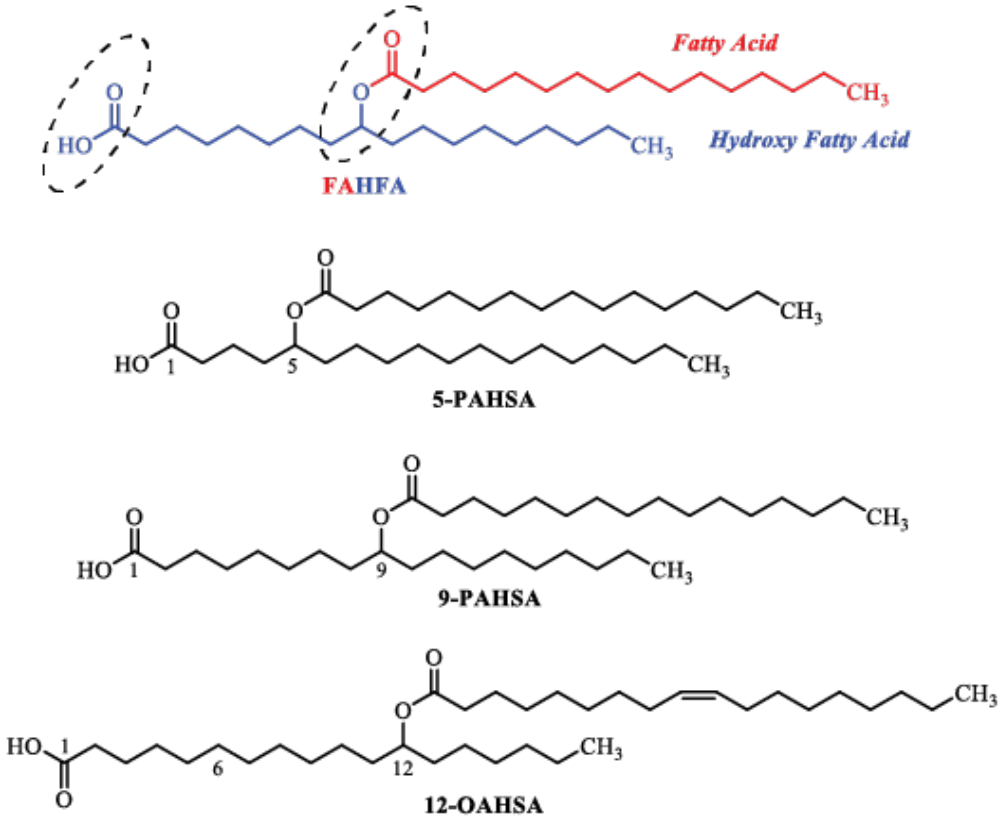


Figure 3.1. Endogenous FAHFA structures, with circles indicating the presence of carboxylic acid moieties.

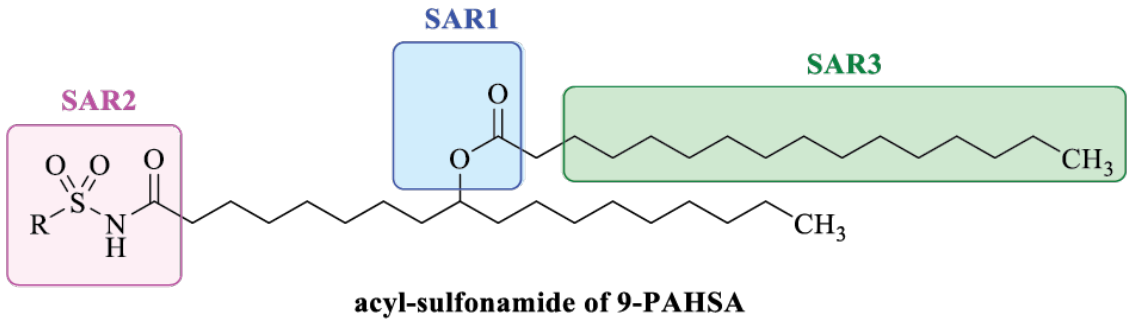
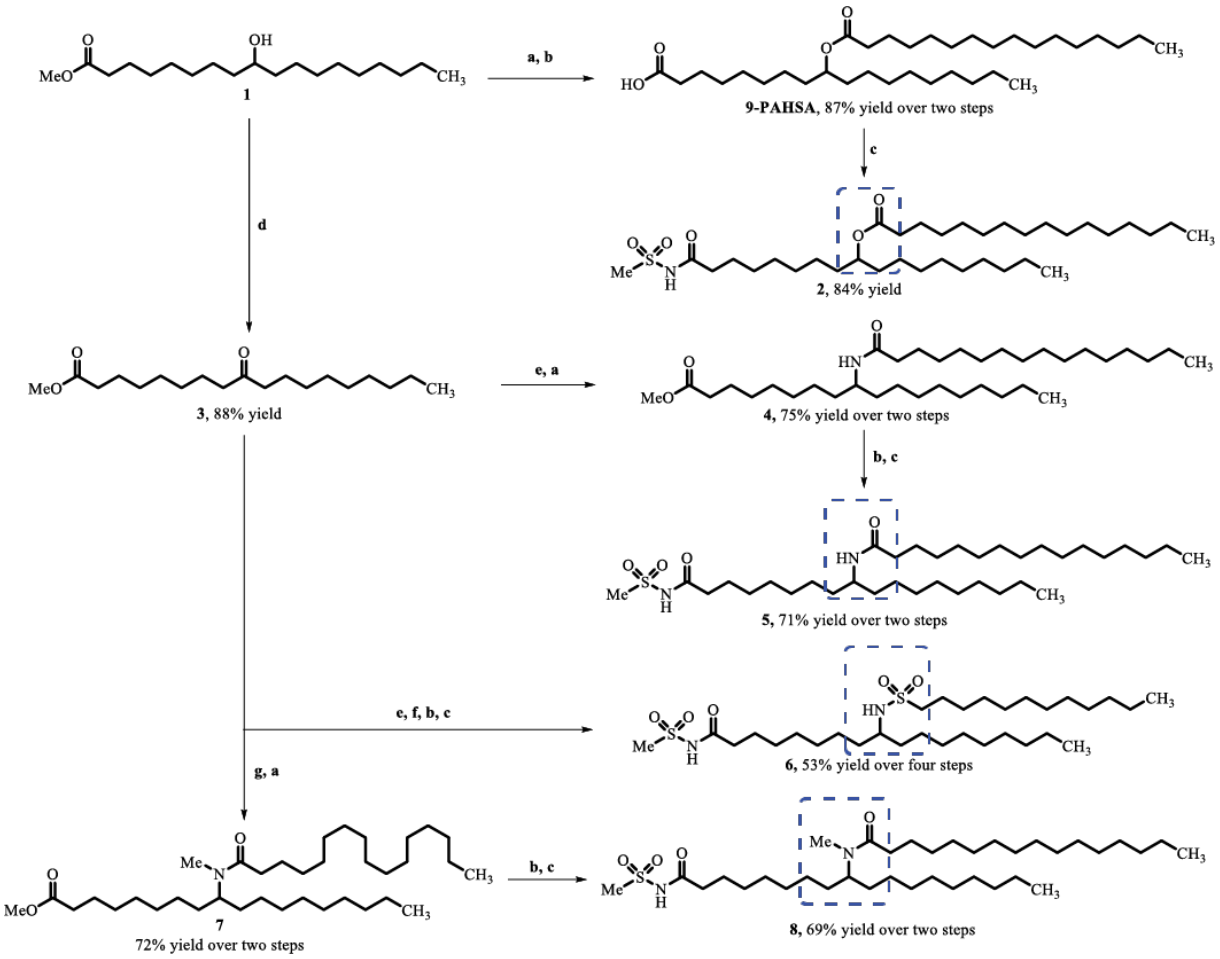
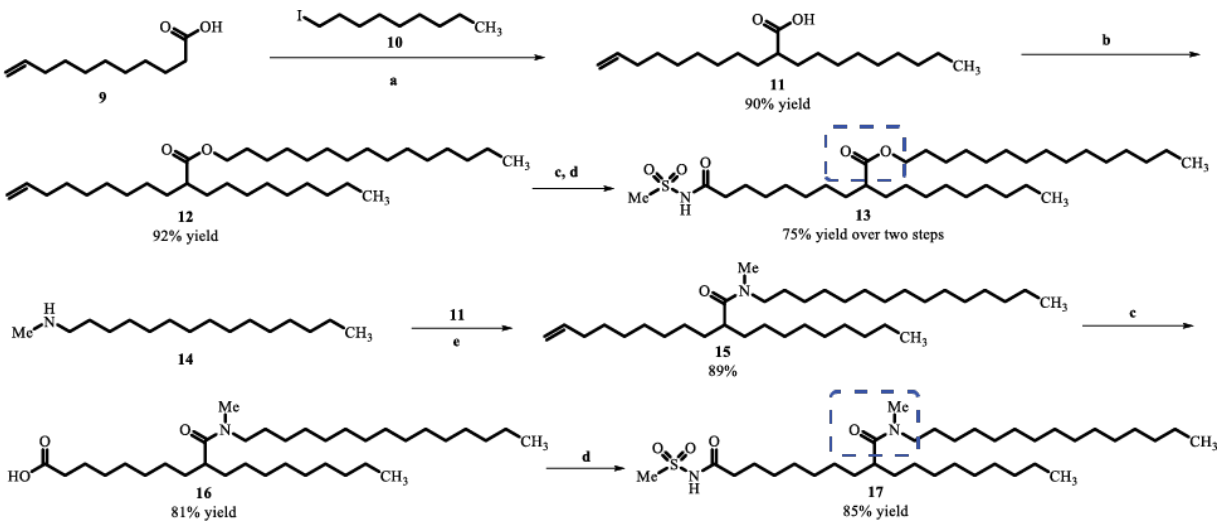


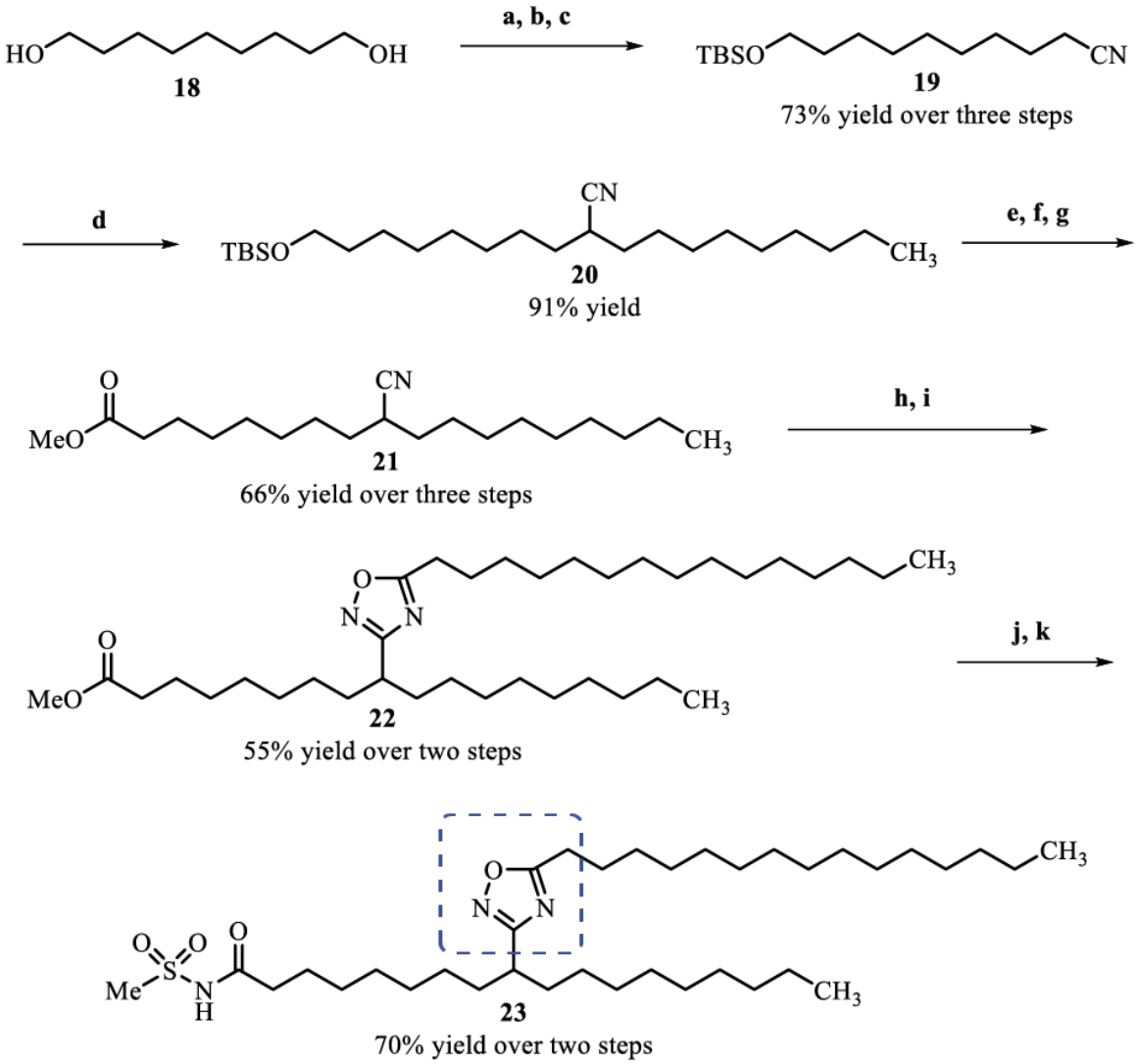
Figure 3.2. Metabolic liabilities in acyl-sulfonamide of 9-PAHSA.



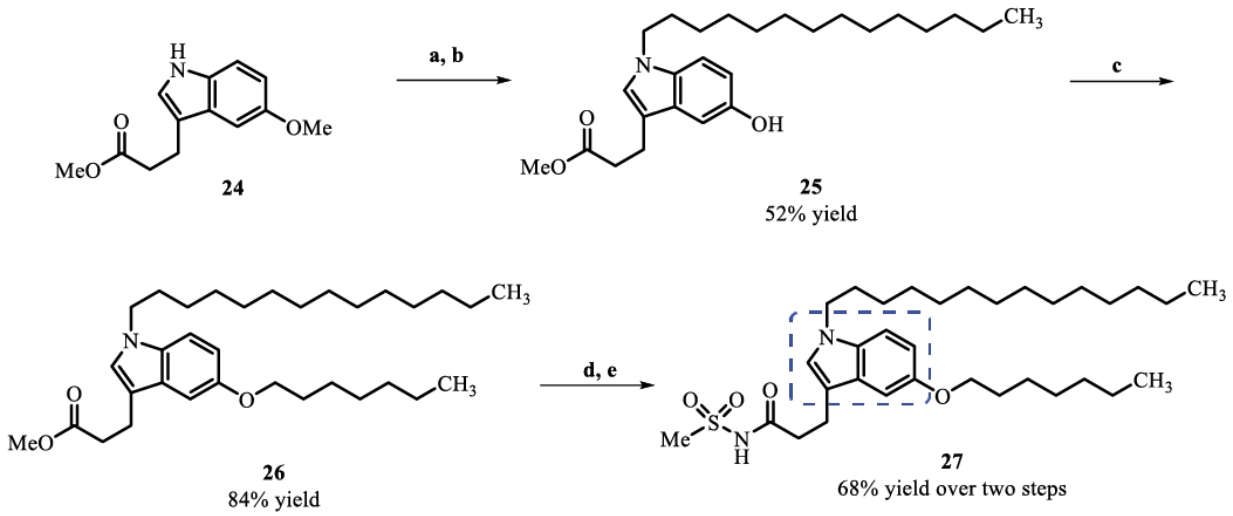
Scheme 3.1. Synthesis of 9-PAHSA analogs with different core structures. Reagents and conditions: a. palmitic acid, EDCI, DMAP, CH_2Cl_2, 23 °C; b. $LiOH \cdot H_2O$, THF:H_2O (1:1, v/v), 0 to 23 °C; c. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C; d. Dess-Martin periodinane, CH_2Cl_2, 0 to 23 °C; e. NH_4OAc, $NaBH_3CN$, MeOH, 23 °C; f. 1-dodecanesulfonyl chloride, Et_3N, CH_2Cl_2, 0 to 23 °C; g. HOAc, $MeNH_2$, $NaBH_3CN$, MeOH, 23 °C.



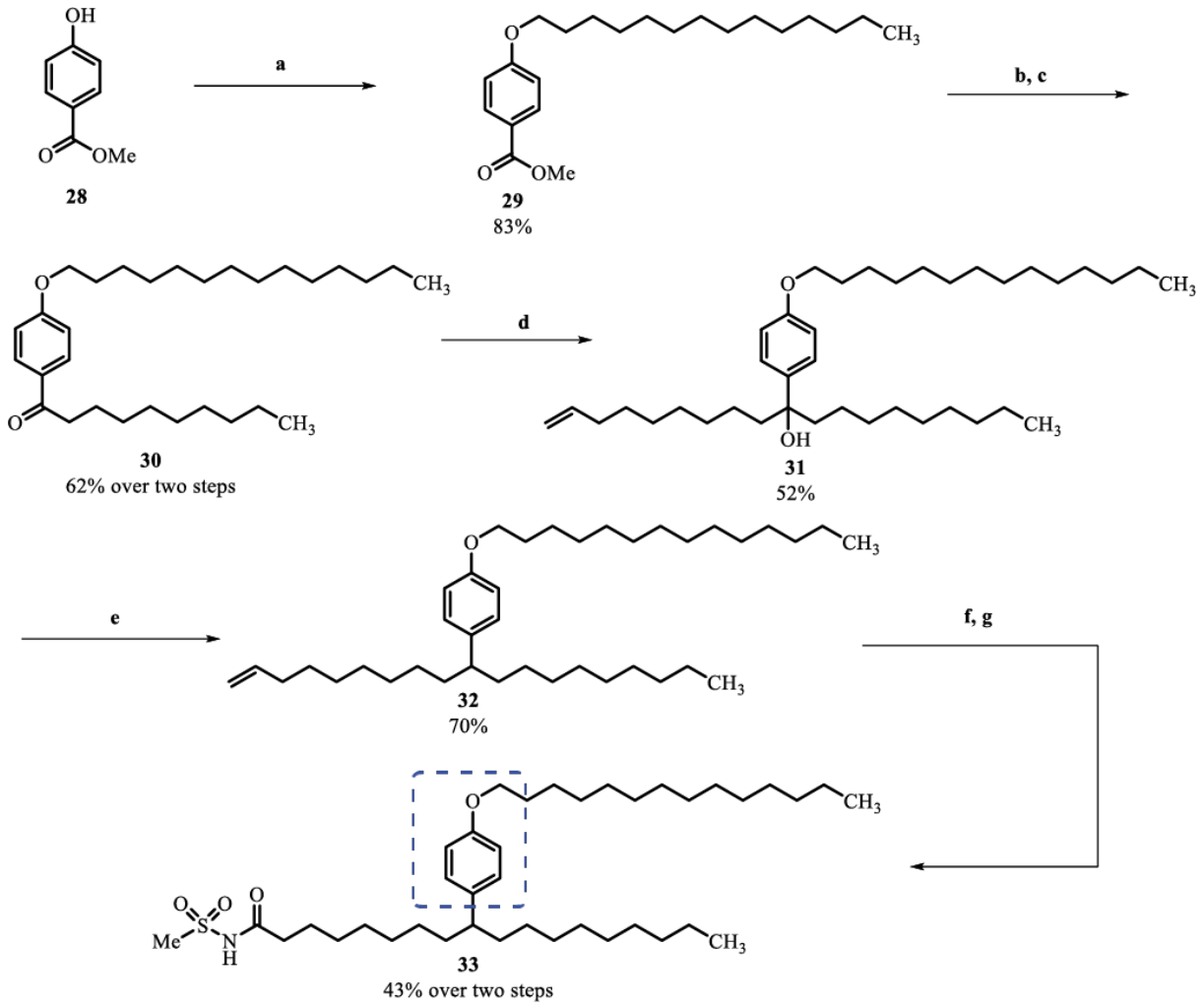
Scheme 3.2. Synthesis of "reverse ester" and "reverse amide" 9-PAHSA analogs. Reagents and conditions: a. LDA, DMPU, THF, 0 to 23 °C; b. 1-pentadecanol, EDCI, DMAP, CH_2Cl_2, 23 °C; c. $RuCl_3$, $NaIO_4$, CCl_4:MeCN:H_2O (2:2:3, v/v/v), 23 °C; d. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C; e. 11, EDCI, DMAP, CH_2Cl_2, 23 °C.



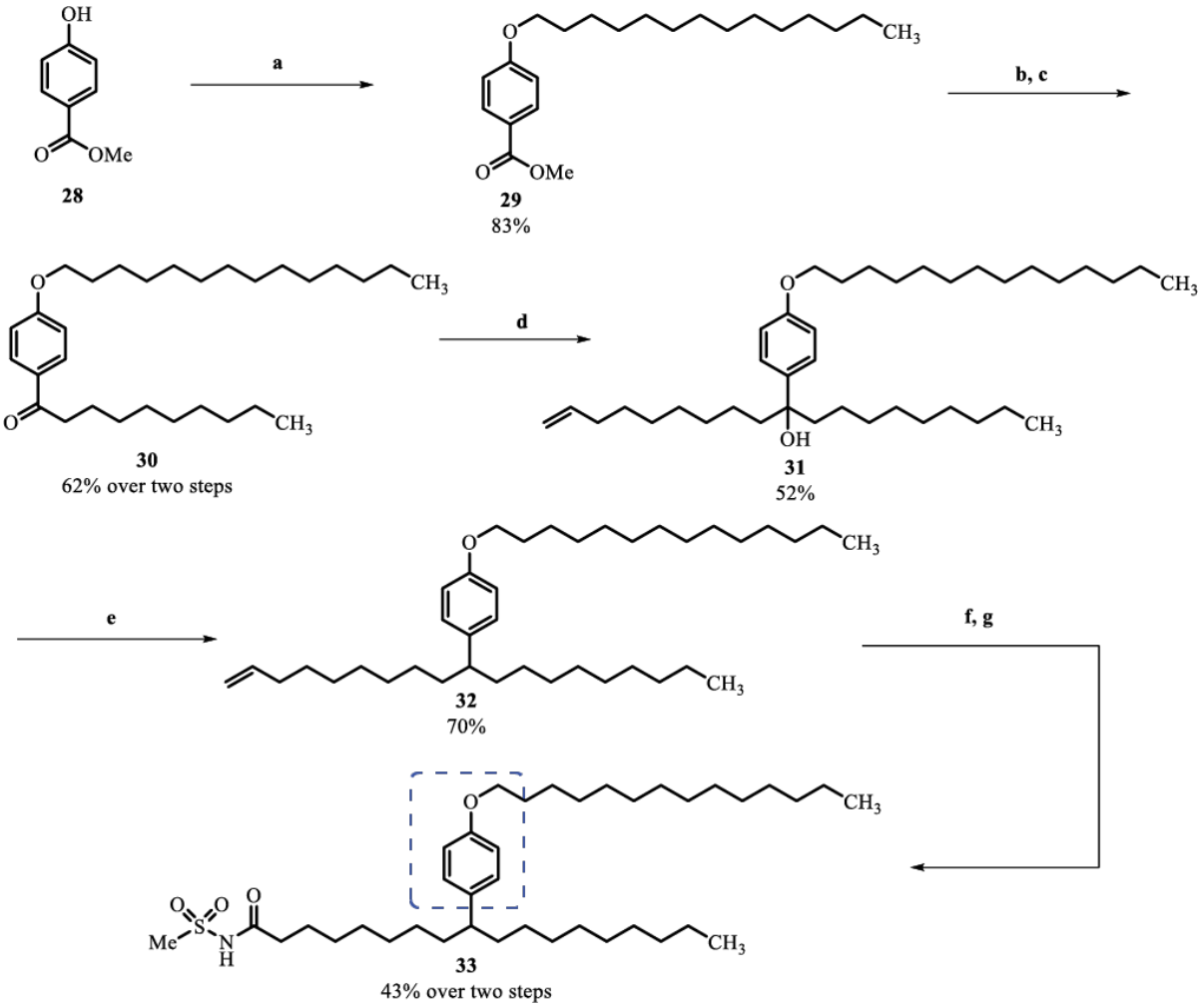
Scheme 3.3. Synthesis of analog 23 bearing an 1,2,4-oxadiazol ester replacement. Reagents and conditions: a. TBSCl, imidazole, DMAP, CH_2Cl_2, 0 to 23 °C; b. MsCl, Et_3N, DMAP, CH_2Cl_2, 0 to 23 °C; c. NaCN, DMSO, 50 °C; d. 1-bromononane, LDA, THF, –78 to 23 °C; e. TBAF, THF, 0 to 23 °C; f. CrO_3/H_2SO_4 (Jones' reagent), acetone, 0 to 23 °C; g. H_2SO_4, MeOH, 23 °C to reflux; h. $NH_2OH{\cdot}HCl$, Et_3N, EtOH, 23 °C to reflux; i. palmitic acid, EDCI, DMAP, CH_2Cl_2, 0 to 23 °C, then TBAF, THF, 0 to 23 °C; j. $LiOH{\cdot}H_2O$, THF:H_2O (1:1, v/v), 0 to 23 °C; k. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C.



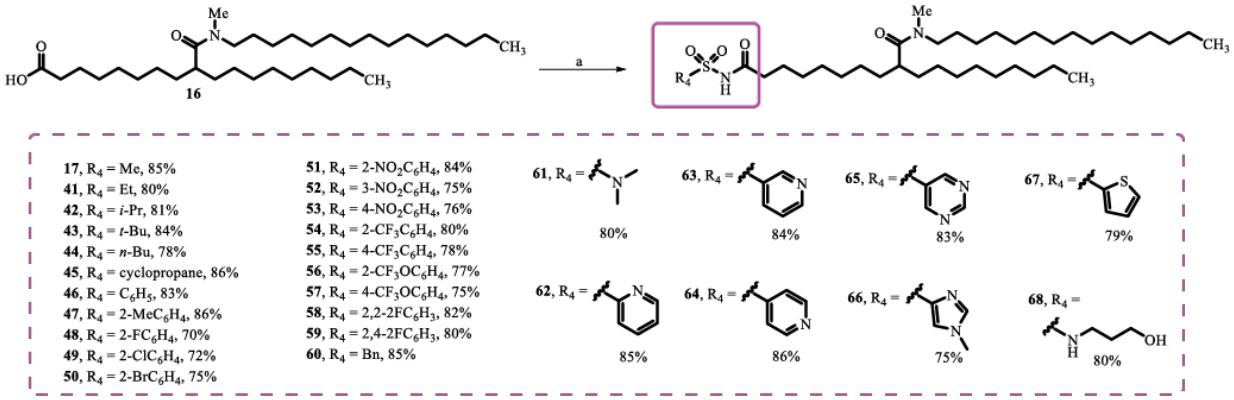
Scheme 3.4. Synthesis of analog 27 with an indole ester/hydrocarbon replacement. Reagents and conditions: a. 1-bromotetradecane, NaH, THF, 0 to 23 °C; b. BBr_3, CH_2Cl_2, –78 to 23 °C; c. 1-bromoheptane, NaH, DMF, 0 to 23 °C; d. $LiOH \cdot H_2O$, THF:H_2O (1:1, v/v), 0 to 23 °C; e. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C.



Scheme 3.5. Synthesis of analog 33 bearing a phenol framework as the core structure. Reagents and conditions: a. 1-bromotetradecane, K_2CO_3, DMF, 23 to 90 °C; b. N,O-dimethylhydroxylamine hydrochloride, Me_3Al, CH_2Cl_2, 23 °C; c. nonylmagnesium bromide, THF, 23 °C to reflux; d. Grignard reagent preparation: 9-bromonon-1-ene, Mg, DIBAL-H, THF, 35-50 °C; Grignard addition condition: Grignard reagent (non-8-en-1-ylmagnesium bromide in THF freshly prepared), THF, 10 to 23 °C; e. TFA, Et_3SiH, CH_2Cl_2, 23 °C; f. OsO_4, oxone, DMF, 23 °C; g. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C.

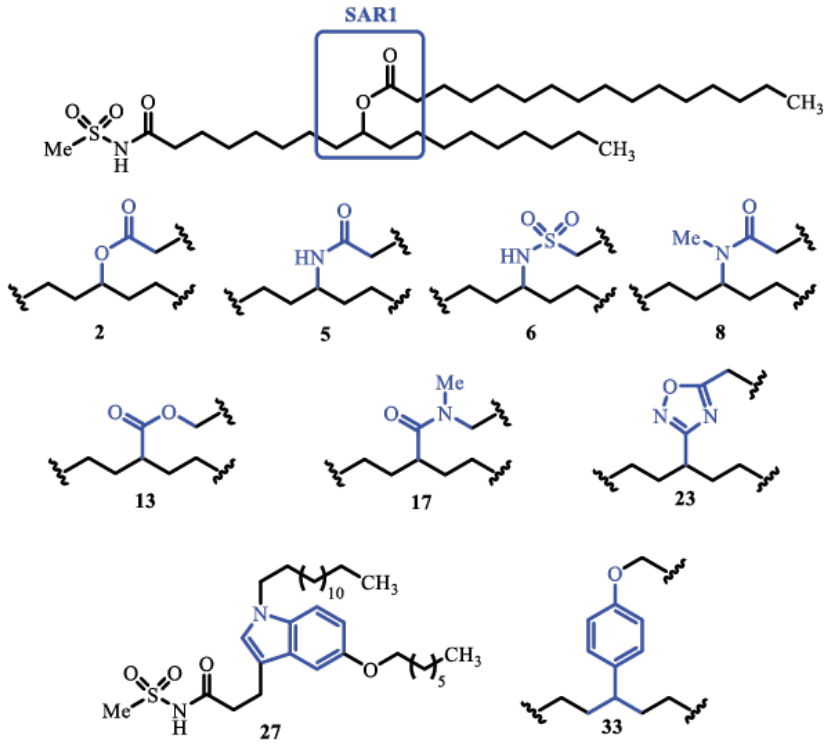


Scheme 3.6. Synthesis of analogs for SAR3. Reagents and conditions: a. BnOH, EDCI, DMAP, DMF, 23 °C; b. $RuCl_3$, $NaIO_4$, CCl_4:MeCN:H_2O (2:2:3, v/v/v), 23 °C; c. methanesulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C; d. Pd/C, H_2 (1 atm), MeOH, 23 °C; e. 37 or 39, EDCI, DMAP, CH_2Cl_2, 23 °C.



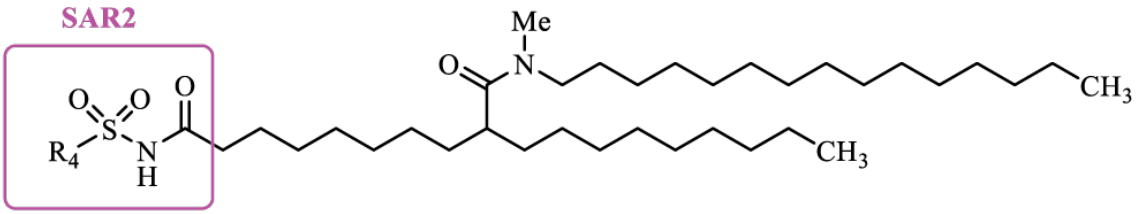
Scheme 3.7. Synthesis of sulfonamides for SAR2. Reagent and conditions: a. sulfonamide, EDCI, DMAP, CH_2Cl_2, 23 °C

Table 3.1. Analogues with ester replacements for the core connection and corresponding effects on IL-6 secretion in LPS-stimulated RAW264.7 cells at 25 µM.



Compound	Relative IL-6 (%)
2	34
5	86
6	59
8	21
13	17
17	18
23	20
27	30
33	46

Table 3.2. Analogues with sulfonamide modifications and corresponding effects on IL-6 secretion in LPS-stimulated RAW264.7 cells at 10 μM.



Compound	R_4	Relative IL-6 (%)
17	Me	64
41	Et	53
42	*i*-Pr	32
43	t-Bu	42
44	n-Bu	57
45	cyclopropane	37
46	C_6H_5	22
47	2-MeC_6H_4	64
48	2-FC_6H_4	20
49	2-ClC_6H_4	25
50	2-BrC_6H_4	22
51	2-$NO_2C_6H_4$	24
52	3-$NO_2C_6H_4$	21
53	4-$NO_2C_6H_4$	19
54	2-$CF_3C_6H_4$	24
55	4-$CF_3C_6H_4$	22
56	2-$CF_3OC_6H_4$	17
57	4-$CF_3OC_6H_4$	24
58	2,2-2FC_6H_3	25

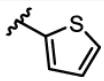
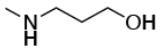
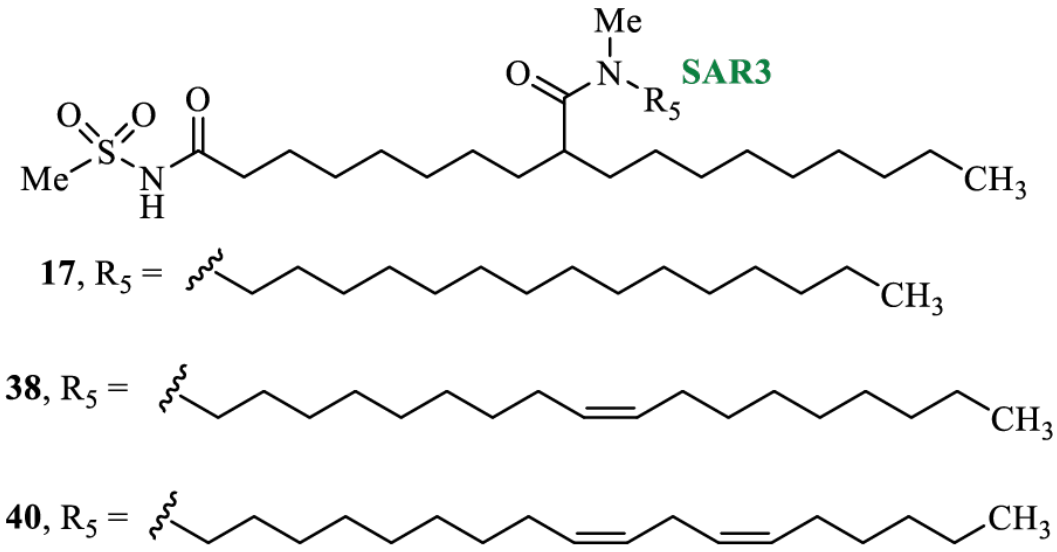
Compound	R_4	Relative IL-6 (%)
59	2,4-2FC_6H_3	24
60	Bn	31
61		100
62		37
63		61
64		17
65		48
66		29
67		29
68		94

Table 3.3. Analogues with changes to the fatty acid chain with increasing unsaturation and corresponding effects on IL-6 secretion in LPS-stimulated RAW264.7 cells at 25 µM.



Compound	Relative IL-6 (%)
17	11
38	9
40	5

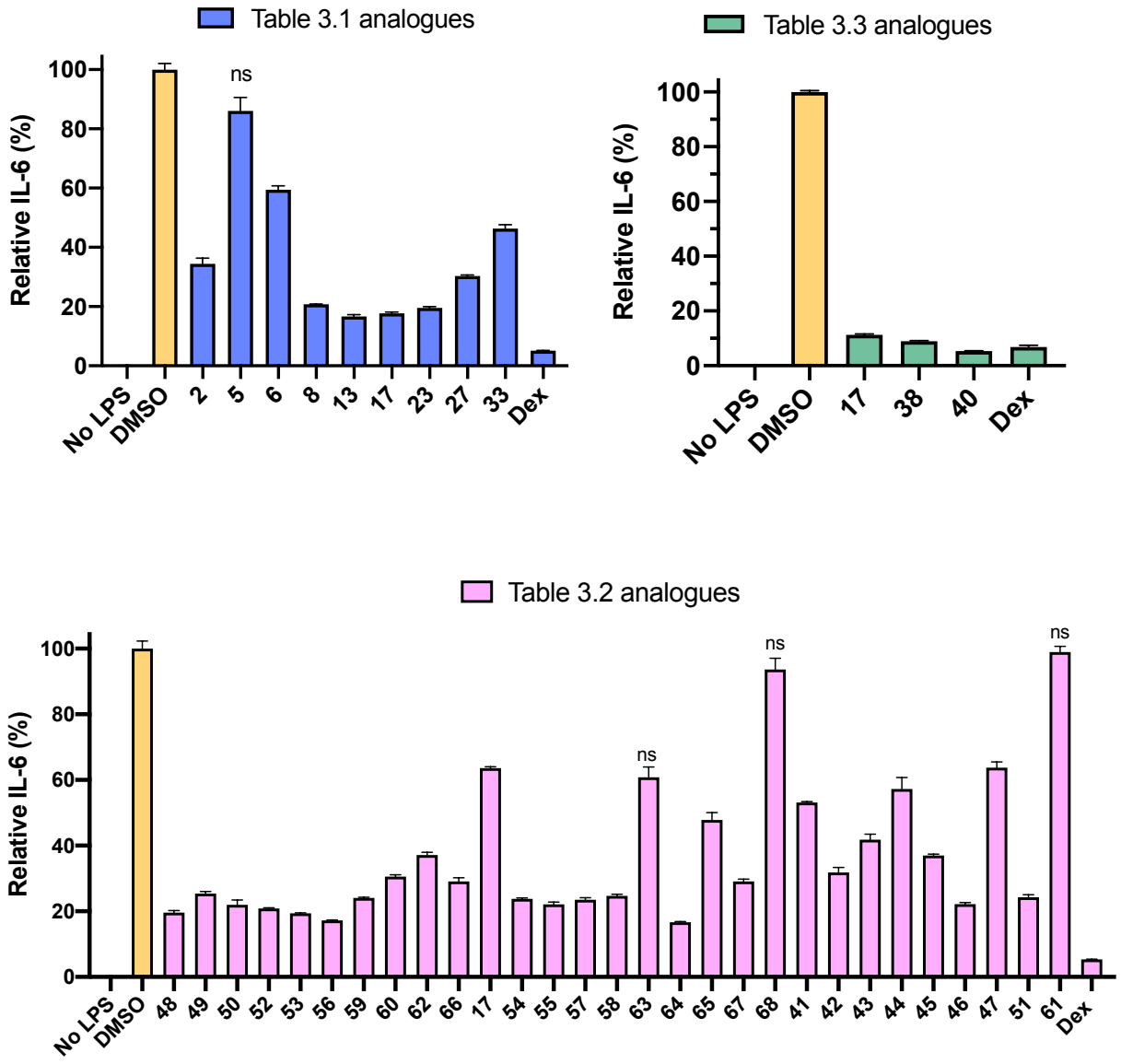


Figure 3.3. Quantification of IL-6 blockade following individual analogue treatments in LPS-stimulated RAW 264.7 cells. LPS-stimulated RAW264.7 cells were subjected to treatment of DMSO or individual analogues at 25 μM (Table 3.1 and Table 3.3) or at 10 μM (Table 3.2). Data are means ± S.E.M with $n = 3$ per treatment per dose. Statistical analysis: One-way ANOVA. All data shown obtained a P value ≤ 0.05 versus DMSO unless indicated otherwise.

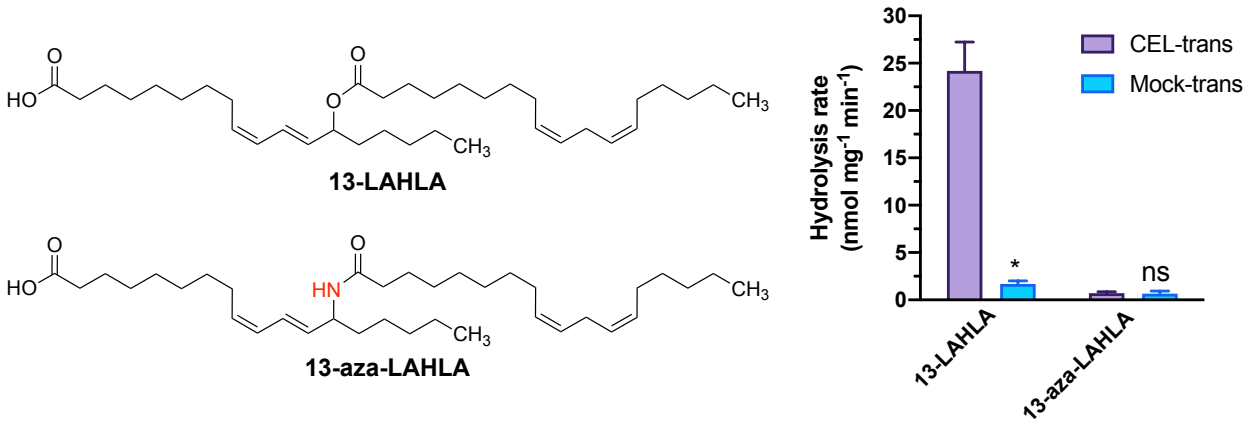


Figure 3.4. Amidated FAHFA analog 13-aza-LAHLA is resistant to hydrolysis by endogenous FAHFA lipase carboxyl ester lipase (CEL).

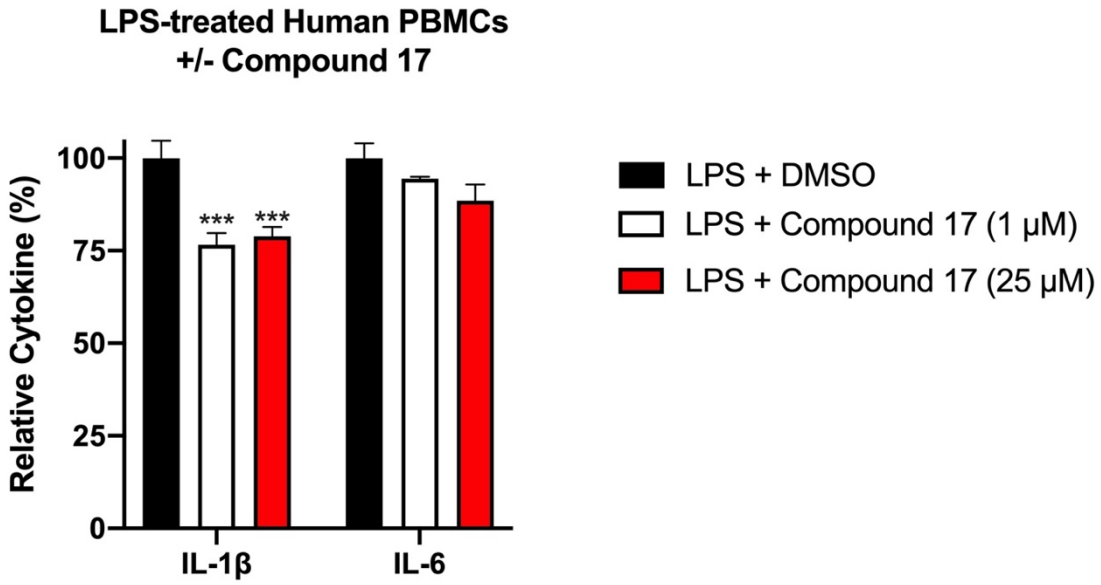


Figure 3.5. Analog **17** activity in primary human cells. **17** inhibits the secretion of IL-1β from LPS-stimulated human PBMCs. Statistical analysis: ANOVA, **P<0.01, ***P<0.001 vs control LPS + DMSO.

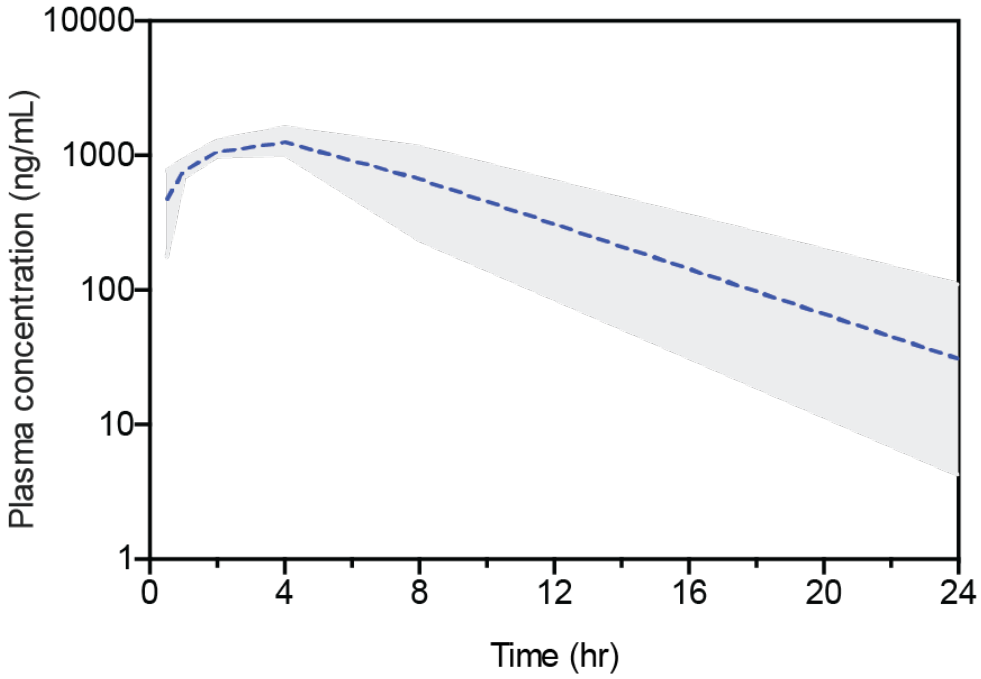


Figure 3.6. Median plasma concentration versus time profile of 17 following intraperitoneal injection in mice. The shaded area displays the interquartile range. (n=6)

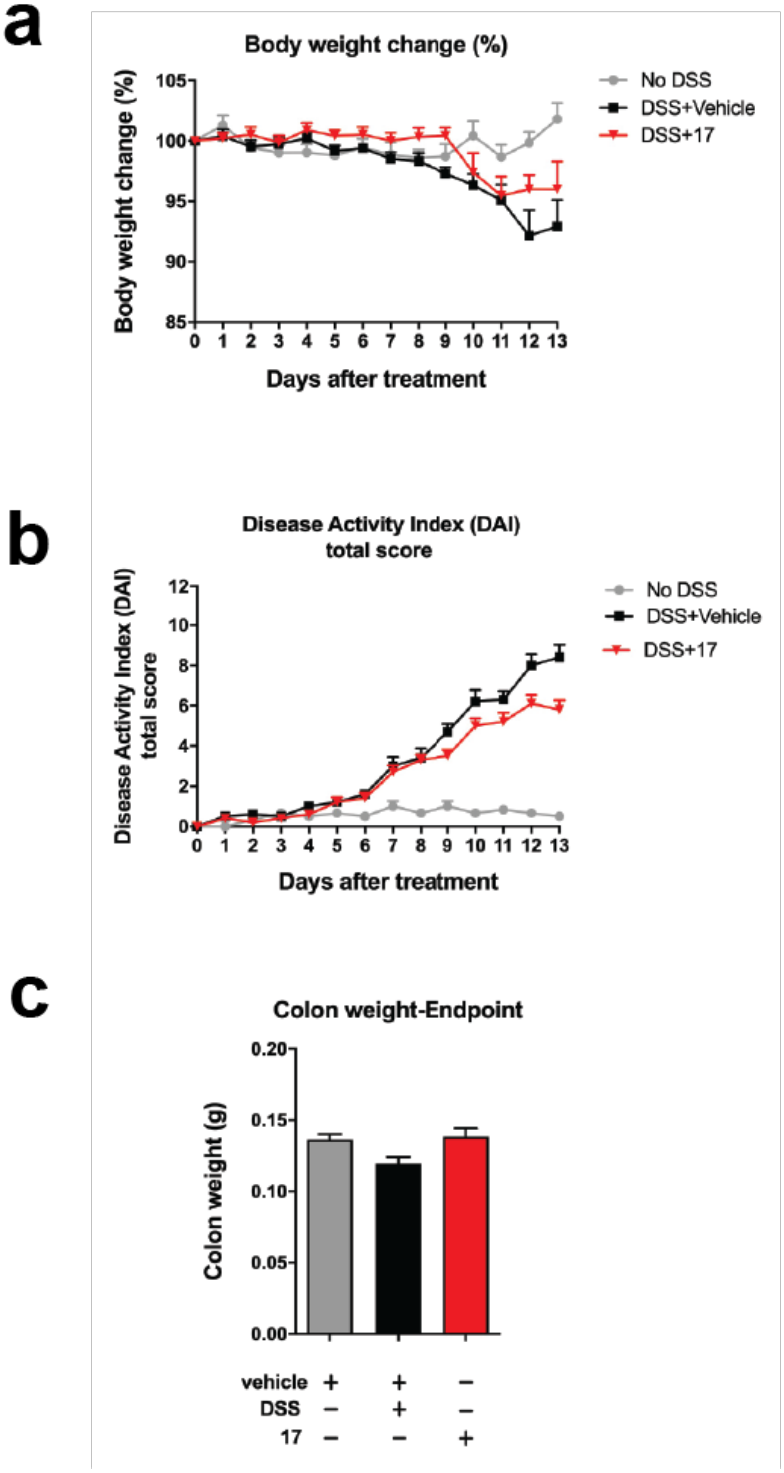


Figure 3.7. Oral administration of **17** inhibits DSS-induced colitis. Mice were treated with vehicle or **17** by oral gavage and 2% DSS (w/v). (**a**) Bodyweight and (**b**) disease activity index measured daily, and post measurements of (**c**) colon weight. Statistical analysis: ANOVA. *P<0.05, **P<0.01, ****P<0.0001 versus vehicle.

3.5 Materials and Methods

General Information and Procedures on Synthesis

For general information and procedures on synthesis of FAHFA analogs, please refer to "Chang_02_ Supplemental_Information2.doc. General information on synthesis and general synthetic procedures for Chapter 3."

Culturing

RAW264.7 cells (Sigma) and human irradiated PBMCs (iQ Biosciences) were cultured in RPMI 1640 supplemented with L-glutamine (Coring, 10-040-CV) and 10% fetal bovine serum (FBS) at 37 °C and 5% CO_2. All experiments were performed under passage 10 for RAW264.7 cells and immediately after thawing for human PBMCs without expansion.

Preparation of test compounds

All test compounds were solubilized in sterile DMSO at 20 mM. Stocks of test compounds are stored in -80 °C until testing day. Upon thawing, test compounds were diluted in culture media at desired concentrations. Final concentrations of DMSO controls were calculated to match that of the test compounds. Final DMSO concentrations were kept at 0.2% or less.

Treatment of analogues in RAW264.7 cells

RAW264.7 cells were seeded onto a 48-well plate at 25,000 cells per well the day prior to treatments. Individual treatments of analogues were prepared in culture media in the presence or absence of 100 ng/mL LPS (Sigma, L4391) at final testing concentrations. Cell culture medium were aspirated and replaced with individual treatments at 200 μL per well, and compounds were

tested in triplicates. Following a 20-hour incubation at 37 °C, cell media were collected for subsequent quantifications of cytokines.

Treatment of analogues in human PBMCs

Immediately upon thawing, human PBMCs were spun down at 200 g x 3 minutes, and the cell pellet was resuspended in fresh culture medium. Cells were seeded at 200,000 cells in 150 μL media per well onto a 48-well plate. Individual analogues were prepared into 20X stock solutions in culture media, and the individual analogues stocks were added onto each well in volumes of 10 μL followed by a 1-hour incubation at 37 °C. Subsequently, 40 μL of culture media with or without 5X LPS stock solutions were added in each well to achieve a final concentration of 100 ng/mL of LPS in a total volume of 200 μL per well. Cell media were collected after 20-hours post treatment for subsequent quantifications of cytokines.
Post-Treatment Analysis of Cell Viability using MTT Assay. 3- (4,5-dimethylthiazol-2-yl)- 2,5-diphenyltetrazolium bromide (MTT) was dissolved in sterile PBS (5 mg/mL) and filtered through a 0.22 μm Sterile Millex Filter to prepare a 500 μg/mL solution in culture medium. Prepared MTT solution was added to adherent cells and incubated at 37 °C for 4 hours. Sterile DMSO was added upon removal of the MTT solution. Plate absorbance was read at 570 nm.

Quantification of IL-6 or IL-1β suppression upon LPS-Stimulation Assay

Collected cell media were subjected to quantifications of mouse and human cytokines for RAW264.7 cells and human PBMCs, respectively using mouse IL-6, human IL-6, and human IL-1β ELISA MAX^{TM} Deluxe kits (BioLegend). Unknown concentrations of cytokines in treated wells were determined and analyzed using the RIA regression feature in GraphPad PRISM.

Relative IL-6 inhibition indicates the percentage of IL-6 attenuated by the test compound relative to 0% IL-6 inhibition achieved by DMSO controls under LPS stimulation.

DSS-induced Colitis Model

Animal experiments to test compound **17** in the DSS colitis model were carried out at BioDuro as a research contract. Male C57BL/6J wild-type mice were obtained from Shanghai Jihui Laboratory Animal Care Co., Ltd. at seven weeks of age. Mice were randomized to the treatment group based on body weight at eight weeks of age. Mice had *ad libitum* access to chow (irradiated, Shanghai SLAC Laboratory Animal Co. Ltd., China) and water. Mice were housed singly at BioDuro with 12-hour light and 12-hour dark cycles. Mice were sacrificed with CO_2, and tissues were harvested, snap-frozen in liquid nitrogen, and stored at -80 °C for further processing. All aspects of animal care were in accordance with federal guidelines and approved by the Institutional Animal Care and Use Committee of the BioDuro.

Eight-week-old mice were treated once daily by oral gavage with vehicle (50% PEG-400, 0.5% Tween 80, and 49.5% distilled water) or compound **17** (15 mg/kg) starting three days before dextran sodium sulfate (DSS, MP biomedicals) treatment. DSS was dissolved in sterile water to a concentration of 2%. Vehicle or **17** treatments continued concurrently with either sterile water or 2% DSS water for ten days. Body weight and clinical colitis score (ratings of 0 – 4 based on stool consistency, rectal bleeding, and mouse appearance) were measured daily throughout the experimental treatment period and used as an indirect measure of colitis severity. All mice were sacrificed by CO_2 by day ten, when body weight loss in the control DSS group was equal to or greater than 20% of the starting body weight or when the maximum clinical

colitis score of seven was reached. Fresh colon was harvested for weight measurement and stored at -80 °C.

3.6 Acknowledgment

Chapter 3, in full, has been submitted for publication of the material with minor modifications as it may appear in Endogenous Fatty Acid Esters of Hydroxy Fatty Acids, Starting Points for New Chemical Regulators of Inflammation. *European Journal of Medicinal Chemistry* (2022). The dissertation author was the second investigator and author of this paper. Other authors include Huijing Wang, Srihari Konduri, Jeremiah D. Momper, Yuxin Lin, Arnold Garcia, José Rizo, Alan Saghatelian, and Dionicio Siegel.

3.7 References

1. M. M. Yore, I. Syed, P. M. Moraes-Vieira, T. Zhang, M. A. Herman, E. A. Homan, R. T. Patel, J. Lee, S. Chen, O. D. Peroni, A. S. Dhaneshwar, A. Hammarstedt, U. Smith, T. E. McGraw, A. Saghatelian, B. B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell*. **159**, 318–332 (2014).

2. J. Lee, P. M. Moraes-Vieira, A. Castoldi, P. Aryal, E. U. Yee, C. Vickers, O. Parnas, C. J. Donaldson, A. Saghatelian, B. B. Kahn, Branched fatty acid esters of hydroxy fatty acids (FAHFAs) protect against colitis by regulating gut innate and adaptive immune responses. *J. Biol. Chem.* **291**, 22207–22217 (2016).

3. I. Syed, J. Lee, P. M. Moraes-Vieira, C. J. Donaldson, A. Sontheimer, P. Aryal, K. Wellenstein, M. J. Kolar, A. T. Nelson, D. Siegel, J. Mokrosinski, I. S. Farooqi, J. J. Zhao, M. M. Yore, O. D. Peroni, A. Saghatelian, B. B. Kahn, Palmitic Acid Hydroxy Stearic Acids Activate GPR40 Which is Involved in Their Beneficial Effects on Glucose Homeostasis. *Cell Metab.* **27**, 419 (2018).

4. I. Syed, M. F. Rubin De Celis, J. F. Mohan, P. M. Moraes-Vieira, A. Vijayakumar, A. T. Nelson, D. Siegel, A. Saghatelian, D. Mathis, B. B. Kahn, PAHSAs attenuate immune responses and promote β cell survival in autoimmune diabetic mice. *J. Clin. Invest.* **129**, 3717–3731 (2019).

5. O. Kuda, M. Brezinova, M. Rombaldova, B. Slavikova, M. Posta, P. Beier, P. Janovska, J. Veleba, J. Kopecky, E. Kudova, T. Pelikanova, Docosahexaenoic acid-derived fatty acid esters of hydroxy fatty acids (FAHFAS) with anti-inflammatory properties. *Diabetes*. **65**, 2580–2590 (2016).

6. M. J. Kolar, S. Konduri, T. Chang, H. Wang, C. McNerlin, L. Ohlsson, M. Härröd, D. Siegel, A. Saghatelian, Linoleic acid esters of hydroxy linoleic acids are anti-inflammatory lipids found in plants and mammals. *J. Biol. Chem.* **294**, 10698–10707 (2019).

7. J. Ding, T. Kind, Q. F. Zhu, Y. Wang, J. W. Yan, O. Fiehn, Y. Q. Feng, In-Silico-Generated Library for Sensitive Detection of 2-Dimethylaminoethylamine Derivatized FAHFA Lipids Using High-Resolution Tandem Mass Spectrometry. *Anal. Chem.* **92**, 5960–5968 (2020).

8. A. Vijayakumar, P. Aryal, J. Wen, I. Syed, R. P. Vazirani, P. M. Moraes-Vieira, J. P. Camporez, M. R. Gallop, R. J. Perry, O. D. Peroni, G. I. Shulman, A. Saghatelian, T. E. McGraw, B. B. Kahn, Absence of Carbohydrate Response Element Binding Protein in Adipocytes Causes Systemic Insulin Resistance and Impairs Glucose Transport. *Cell Rep.* **21**, 1021–1035 (2017).

9. T. Zhang, S. Chen, I. Syed, M. Ståhlman, M. J. Kolar, E. A. Homan, Q. Chu, U. Smith, J. Borén, B. B. Kahn, A. Saghatelian, A LC-MS–based workflow for measurement of

branched fatty acid esters of hydroxy fatty acids. *Nat. Protoc. 2016 114*. **11**, 747–763 (2016).

10. Q. F. Zhu, J. W. Yan, T. Y. Zhang, H. M. Xiao, Y. Q. Feng, Comprehensive Screening and Identification of Fatty Acid Esters of Hydroxy Fatty Acids in Plant Tissues by Chemical Isotope Labeling-Assisted Liquid Chromatography-Mass Spectrometry. *Anal. Chem.* **90**, 10056–10063 (2018).

11. M. Brezinova, O. Kuda, J. Hansikova, M. Rombaldova, L. Balas, K. Bardova, T. Durand, M. Rossmeisl, M. Cerna, Z. Stranak, J. Kopecky, Levels of palmitic acid ester of hydroxystearic acid (PAHSA) are reduced in the breast milk of obese mothers. *Biochim. Biophys. acta. Mol. cell Biol. lipids*. **1863**, 126–131 (2018).

12. M. Brezinova, T. Cajka, M. Oseeva, M. Stepan, K. Dadova, L. Rossmeislova, M. Matous, M. Siklova, M. Rossmeisl, O. Kuda, Exercise training induces insulin-sensitizing PAHSAs in adipose tissue of elderly women. *Biochim. Biophys. acta. Mol. cell Biol. lipids*. **1865**, 158576 (2020).

13. F. Fava, S. Danese, Intestinal microbiota in inflammatory bowel disease: Friend of foe? *World J. Gastroenterol.* **17**, 557 (2011).

14. A. H. Morgan, V. J. Hammond, L. Morgan, C. P. Thomas, K. A. Tallman, Y. R. Garcia-Diaz, C. McGuigan, M. Serpi, N. A. Porter, R. C. Murphy, V. B. O'Donnell, Quantitative assays for esterified oxylipins generated by immune cells. *Nat. Protoc. 2010 512*. **5**, 1919–1931 (2010).

15. B. Neises, W. Steglich, Simple Method for the Esterification of Carboxylic Acids. *Angew. Chemie Int. Ed. English*. **17**, 522–524 (1978).

16. H. Wang, T. Chang, S. Konduri, J. Huang, A. Saghatelian, D. Siegel, Synthesis of chemically edited derivatives of the endogenous regulator of inflammation 9-PAHSA. *J. Antibiot. (Tokyo).* **72**, 498–506 (2019).

17. H. Wang, M. J. Kolar, T. Chang, J. Rizo, S. Konduri, C. McNerlin, A. Saghatelian, D. Siegel, Stereochemistry of Linoleic Acid Esters of Hydroxy Linoleic Acids. *Org. Lett.* **21**, 8080–8084 (2019).

18. W. H. Parsons, M. J. Kolar, S. S. Kamat, A. B. Cognetta, J. J. Hulce, E. Saez, B. B. Kahn, A. Saghatelian, B. F. Cravatt, AIG1 and ADTRP are atypical integral membrane hydrolases that degrade bioactive FAHFAs. *Nat. Chem. Biol.* **12**, 367–372 (2016).

19. M. J. Kolar, S. S. Kamat, W. H. Parsons, E. A. Homan, T. Maher, O. D. Peroni, I. Syed, K. Fjeld, A. Molven, B. B. Kahn, B. F. Cravatt, A. Saghatelian, Branched Fatty Acid Esters of Hydroxy Fatty Acids Are Preferred Substrates of the MODY8 Protein Carboxyl Ester Lipase. *Biochemistry*. **55**, 4636–4641 (2016).

20. D. Tan, M. E. Ertunc, S. Konduri, J. Zhang, A. M. Pinto, Q. Chu, B. B. Kahn, D. Siegel,

A. Saghatelian, Discovery of FAHFA-Containing Triacylglycerols and Their Metabolic Regulation. *J. Am. Chem. Soc.* **141**, 8798–8806 (2019).

21. C. Ballatore, D. M. Huryn, A. B. Smith, Carboxylic acid (bio)isosteres in drug design. *ChemMedChem*. **8**, 385–395 (2013).

22. P. Lassalas, B. Gay, C. Lasfargeas, M. J. James, V. Tran, K. G. Vijayendran, K. R. Brunden, M. C. Kozlowski, C. J. Thomas, A. B. Smith, D. M. Huryn, C. Ballatore, Structure Property Relationships of Carboxylic Acid Isosteres. *J. Med. Chem.* **59**, 3183–3203 (2016).

23. K. K. Nyati, K. Masuda, M. Mahabub-Uz Zaman, P. K. Dubey, D. Millrine, J. P. Chalise, M. Higa, S. Li, D. M. Standley, K. Saito, H. Hanieh, T. Kishimoto, TLR4-induced NF-κB and MAPK signaling regulate the IL-6 mRNA stabilizing protein Arid5a. *Nucleic Acids Res.* **45**, 2687 (2017).

24. A. Ngkelo, K. Meja, M. Yeadon, I. Adcock, P. A. Kirkham, LPS induced inflammatory responses in human peripheral blood mononuclear cells is mediated through NOX4 and G iα dependent PI-3kinase signalling. *J. Inflamm.* **9**, 1–7 (2012).

25. C. R. H. Raetz, C. Whitfield, Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* **71**, 635 (2002).

26. P. G. McDougal, J. G. Rico, Y. I. Oh, B. D. Condon, A convenient procedure for the monosilylation of symmetric 1,n-diols. *J. Org. Chem.* **51**, 3388–3390 (2002).

27. L. Mao, A. Kitani, W. Strober, I. J. Fuss, *The role of NLRP3 and IL-1β in the pathogenesis of inflammatory bowel disease* (Frontiers Media S.A., 2018), vol. 9.

28. L. M, S. PL, K. F, R. D, M. Ligumsky, Role of interleukin 1 in inflammatory bowel disease-enhanced production during active disease. *Gut*. **31**, 686–689 (1990).

29. E. Hadadi, B. Zhang, K. Baidzajevas, N. Yusof, K. J. Puan, S. M. Ong, W. H. Yeap, O. Rotzschke, E. Kiss-Toth, H. Wilson, S. C. Wong, Differential IL-1β secretion by monocyte subsets is regulated by Hsp27 through modulating mRNA stability. *Sci. Rep.* **6** (2016), doi:10.1038/srep39035.

30. L. Janský, P. Reymanová, J. Kopecký, Dynamics of Cytokine Production in Human Peripheral Blood Mononuclear Cells Stimulated by LPS or Infected by Borrelia. *Physiol. Res.* **52**, 593–598 (2003).

CHAPTER 4

The SLC35A4 upstream open reading frame encodes a microprotein regulator of mitochondrial structure and function

Abstract

Upstream open reading frames (uORFs) are cis-acting elements within the 5'-untranslated regions (5'-UTRs) of eukaryotic RNAs that post-transcriptionally regulate the expression of downstream main ORFs (mORFs). Eukaryotic translation was thought to occur once per transcript, usually at the longest ORF. Because uORFs are predominantly small ORFs (smORFs), most have gone unannotated and uncharacterized. The eukaryotic ORF landscape is dynamically changing, however, as newer technologies have recently revealed thousands of translated mouse and human uORFs upending the current dogma in the field.

Eukaryotic uORFs encode conserved, functional small peptides, including SLC35A4-MP, a 103-amino acid microprotein from the *SLC35A4* uORF. Lack of homology to any characterized proteins required the *de novo* characterization of SLC35A4-MP to decipher whether the microprotein is functional. SLC35A4-MP is a *bona fide* member of the mammalian proteome with robust expression in mammalian cells and tissues, including peripheral tissues and the central nervous system. Analysis of the SLC35A4-MP sequence revealed a single-pass transmembrane domain that localizes the microprotein to the inner mitochondrial membrane. As a class, microproteins primarily function as components of protein complexes. Protein interaction studies identified SLC35A4-MP as an integral member of the mitochondrial contact site and cristae organizing system (MICOS) complex. The MICOS complex facilitates the formation of crista architecture and is essential for mitochondrial functions. Knockdown or knockout of SLC35A4-MP results in smaller cristae and shortened crista junctions (CJs) and reduced metabolic output, revealing a fundamental role for this microprotein. Furthermore, these results support that uORFs can encode microproteins that serve vital cellular roles.

4.1 Introduction

The accepted model of eukaryotic genes had been that each mRNA transcript produces a single protein via ribosome translation of the largest open reading frame (ORF) (*1*). However, some transcripts are known to contain one or more upstream ORFs (uORFs), which are typically small ORFs (smORFs) less than 100-150 codons (*2*, *3*). uORFs serve to post-transcriptionally regulate the expression of the longer downstream ORF, or main ORF (mORF), through interactions with the scanning ribosome. Well-studied uORFs found in *ATF4* and *CHOP* transcripts, for instance, demonstrate dynamic regulation of downstream mORF translation. Under homeostatic conditions, the *ATF4* and *CHOP* uORFs are translated and repress the translation of the downstream ORFs, but the induction of cell stress leads to ribosome phosphorylation which allows for uORF read through and robust translation of ATF4 and CHOP. This post-transcriptional mechanism provides an elegant way for the cell to rapidly control protein expression without having to generate a new transcript.

Genome sequences provide the raw data for the computational prediction of protein-coding ORFs, but the length cutoffs and the monocistronic model of eukaryotic translation impeded annotation of most uORFs. The number of uORFs in eukaryotic genomes has expanded over the last decade, however, through advances on several fronts. Computational strategies that targeted uORFs suggested that half of all human and mouse transcripts had at least one uORF (*4*). And the advent and application of proteogenomics and ribosome profiling (Ribo-Seq) have resulted in the annotation of thousands of uORFs by direct detection of the peptides or translated uORFs, respectively. These empirically detected uORFs have rewritten the rules of eukaryotic translation by showing that polycistronic transcripts exist and are prevalent within genomes.

Functional studies indicate that the deletion of uORFs can increase the translation of the downstream mORFs by 30-80%, consistent with uORFs role as a cis-regulator of mORF translation. uORFs may also encode functional peptides or small proteins, collectively referred to as microproteins. The *MIEF1* uORF, for example, produces a 70-amino acid microprotein (MIEF-MP) that regulates mitochondrial fission and protein-protein interactions with the mitochondrial ribosome or its downstream mORF protein (*5*). Moreover, an unbiased protein-interaction analysis of ten uORF-encoded microproteins showed that about half are functional and interact with the proteins from the downstream mORFs (*6*). With only a few uORF microproteins characterized, there remains thousands of uORF microproteins expressed in cells and tissues of unknown function (*6*, *7*) and the characterization of these genes represents a frontier in molecular biology.

Here, we characterize SLC35A4 microprotein (SLC35A4-MP), a functional microprotein derived from the *SLC35A4* mRNA uORF. This uORF is an established translational regulator of the *SLC35A4* mORF (*8*). We hypothesized that this uORF also encodes a functional microprotein because it is conserved between distant organisms, well expressed, and contains a single pass transmembrane domain, which has been observed in several functional microproteins. While some uORF microproteins are reported to interact with their downstream mORF proteins (*6*), we find that this is not the case for SLC35A4-MP, which is an inner mitochondrial membrane microprotein, and, therefore, cannot interact with the Golgi-localized SLC35A4 mORF-encoded protein (*9*). Since biologically active microproteins operate via interactions with larger protein complexes, we carried out SLC35A4-MP immunoprecipitation experiments and discovered that SLC35A4-MP is a component of the mitochondrial contact site and cristae organizing system (MICOS) complex. The MICOS complex drives the formation of mitochondrial cristae and crista

junctions (CJs), producing the characteristic inner mitochondrial membrane folding observed in mitochondrial images. Cristae house the machinery for ATP production and are required for proper mitochondrial function. We find that SLC35A4-MP has an active role in the mitochondria as a reduction or loss of SLC35A4-MP results in pronounced changes to crista structure, which impacts mitochondrial respiration and ATP production. Thus, SLC35A4-MP is an essential component of an optimally functioning cell, and the characterization of this microprotein highlights new biology associated with uORFs.

4.2 Results and Discussion

Current model of uORF functions

Informatic studies suggested that uORFs are highly prevalent. Found in at least 50% of the human and mouse transcripts, uORFs represent the largest class of additional ORFs in the genome, where they serve as cis-regulators of downstream translation (*7*). Functional studies indicate that removal of uORFs can amplify translation of the downstream coding sequence by 30-80% (*10*). Moreover, the regulation of downstream translation by uORFs are dynamic under global translational changes (*8*). For example, under homeostatic conditions, the uORFs on ATF4 and CHOP transcripts are engaged to repress the translation of the downstream ORFs. However, upon the induction of stress and eventual phosphorylation of the ribosome, the scanning ribosome can bypass these uORFs to initiate annotation of the downstream ORFs, resulting in "leaky" translation (*11*, *12*) (**Figure 4.1**). Thus, it is currently understood that uORFs predominantly serve as a mechanism to mediate and finetune translation of downstream mORFs.

While the regulation of translation might be the primary function of uORFs, it is not their only role. uORFs can also encode for peptides and small proteins with unique biological

functions, collectively referred to as microproteins. The uORF on the MIEF1 gene, for example, produces a 70-amino acid microprotein (MIEF-MP) that has been shown to regulate mitochondrial fission and involve in protein-protein interactions with the mitochondrial ribosome and MIEF1 (*5*). In addition, a search of mRNAs identified several non-coding RNAs whose transcripts contained smORFs that can encode for microproteins with important roles in the muscle. Notably, myoregulin is a 46-amino acid polypeptide that folds into a single-pass transmembrane helix (*13*). This structure allows myoregulin to interact with the transmembrane domain of sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) to inhibit calcium dynamics and reduce skeletal muscle function. Further studies identified a 34-amino acid microprotein called DWORF, which also serves as a potent activator of SERCA to enhance muscle function in the heart (*14*). Together, the importance of uORFs and transmembrane microproteins drove our interest to investigate if the uORF located in the 5'-UTR of the SLC35A4 mRNA transcript encodes a functional microprotein with unique functions (**Figure 4.1**).

Genomic locus and pre-existing knowledge of the function of human *SLC35A4*

Human SLC35A4 (Solute Carrier Family 35 Member A4; UniProt entry Q96G79) is a 324-amino acid protein belonging to the SLC superfamily of solute carriers, the second largest family of membrane proteins after G-protein coupled receptors (*15*). Located on chromosome 5, the human *SLC35A4* mRNA is spliced together from three exons, forming two predominant isoforms: a 312 base-pair (bp) SLC35A4 uORF and a 975-bp mORF (**Figure 4.2a**). While the *SLC35A4* mORF is encoded entirely within exon 3, the uORF sequence extends across exon 1, exon 2, and exon 3 with a 407-bp gap in between the two ORFs.

Biochemical findings demonstrated that the *SLC35A4* mORF encodes a 324-a.a. SLC35A4 canonical protein localized near or within the Golgi apparatus (*9*). Based on its sequence homology with other members in the SLC35 family, the SLC35A4 canonical protein is predicted to be a UDP-sugar transporter, though its exact biological role has yet to be experimentally confirmed (*9*, *16*). Nonetheless, the annotation score of the canonical protein is low. Moreover, biochemical data supporting a potential biological role in the *SLC35A4* uORF is extremely limited. We were unaware of any functional studies of the *SLC35A4* uORF other than its ability to *cis*-regulate the translation of the mORF (*8*). Similar to other uORFs, such as those in *ATF4* and *CHOP*, the *SLC35A4* uORF regulates transcription of the mORF in *cis* and controls the translation of the *SLC35A4* mORF during arsenite-induced stress (*8*). While the cis-regulatory functions of the *SLC35A4* uORF in translational control had been explored, we were curious whether the *SLC35A4* uORF could produce a microprotein with unique functions.

Prediction and validation of SLC35A4-MP translation

The evolutionary conservation of protein sequences implicate fitness, hence genomic function, under the test of natural selection (*17*, *18*). Rising usage of Basic Local Alignment Search Tool (BLAST) and PhyloCSF scores (*19*), which analyzes nucleotide sequence alignment across species, help scientists compare sequence conservation, predict protein coding regions, and more importantly, convey the translational potential of a gene or transcript. Using BLAST, we searched for the human 103-a.a. SLC35A4-MP protein and obtained its microprotein sequence, if translated. Sequence alignment of human SLC35A4-MP returned remarkable conservation across many other vertebrate species, including but not limited to non-human

primates, felines, rodents, and zebrafish (**Figure 4.2b***)*. A high degree of conservation suggests that this microprotein is likely functional as it has been retained throughout evolution.

Next, ribosome footprints as a proxy for protein synthesis, we investigated the translational potential of the human SLC35A4 transcript in HEK293T cells. Ribosome profiling resulted in robust read counts across the uORF region spanning exons 1-3 and there are considerably fewer read counts across the mORF (**Figure 4.2c**) (*7*), suggesting that the SLC35A4 uORF possesses coding potential in HEK293T cells. This is also consistent with published reports that the major protein product of the *SLC35A4* mRNA is likely derived from the uORF and not the canonical mORF in HEK293T cells under homeostasis (*8*).

To allow downstream investigation of SLC35A4-MP, we generated a rabbit polyclonal antiserum raised against the N-terminal 34 amino acids of human SLC35A4-MP and demonstrate that endogenous SLC35A4-MP is abundant in HEK293T cells as we robustly detected SLC35A4-MP peptide fragments with our antiserum by proteomics (**Figure 4.2d**). In addition, the endogenous expression of SLC35A4-MP is also found in other human cells lines, including A549 and HeLaS3 cells (**Chang_03_ Supplemental_Information3.doc, Figure S1**). The SLC35A4-MP antiserum also detects the homologous mouse uORF, enabling us to determine the tissue distribution of this microprotein in C57BL/6 mice (**Figure 4.2e**). We found that the expression of SLC35A4-MP is ubiquitous and abundant across most peripheral tissues and the central nervous system at endogenous levels. From these data, we conclude that SLC35A4-MP is a stable, long-lived microprotein with robust expression in mammalian cells and tissues.

Sequence analysis and model prediction of SLC35A4-MP

Analysis of the SLC35A4-MP sequence using the structure prediction tool TMHMM Server v.2.0 identified a single-pass transmembrane domain (TMD) between residues 62 and 84 (**Figure 4.3a***)*. Next, searching the microprotein sequence against the human non-redundant protein sequence database on BLASTP (sequence ID: **NP_001124401.1**) reveals that SLC35A4-MP exhibit sequence similarity with short transmembrane mitochondrial protein 1 (STMP1), a certified mitochondrial transmembrane microprotein (*20*). We found that while STMP1 is only 47 a.a. in length, it shares 16 identical residues with SLC35A4-MP, particularly in the TMD of SLC35A4-MP (**Chang_03_ Supplemental_Information3.doc, Figure S2**). This overlapped region is identified as a domain of unknown functions (DUF 4535, Accession: PF15054) by the PFAM database. Although this shared region currently lacks a well-defined function for us to gain further functional insights of the microprotein, TMDs are not generic but reflective features of the physical properties of the bilayer in which the protein resides (*21*). Therefore, we reason that SLC35A4-MP may be potentially involved in organelle-specific activities, similar to STMP1. Using DeepLoc, an algorithm that relies on deep neural networks to confer protein localization (*22*), we analyzed the protein sequence of the SLC35A4-MP to predict where the microprotein could be found within cells. This prediction provided a possible subcellular localization of SLC35A4-MP near or within the mitochondria (**Figure 4.3b**).

SLC35A4-MP is a mitochondrial-localized microprotein

To test the veracity of these model predictions, we performed immunofluorescence experiments and biochemical fractionation of organelles to pinpoint the true subcellular location of SLC35A4-MP. First, we compared the expression levels of SLC35A4-MP across a panel of

subcellular fractions of HEK293T cells. Immunoblotting analysis demonstrates that SLC35A4-MP is expressed in total lysate and exclusively enriched in the mitochondrial fraction (**Figure 4.3c**). Conceivably, SLC35A4-MP's true localization may be altered by the mitochondrial isolation method used (*23*, *24*). Therefore, we sought additional evidence to corroborate its localization to mitochondria. To avoid isolation artifacts, we performed confocal microscopy to observe endogenous SLC35A4-MP expression in intact HeLa cells using our in-house generated rabbit polyclonal SLC35A4-MP antiserum (**Figure 4.3d**). The results show that the fluorescence signal for SLC35A4-MP (**Figure 4.3d**, green) overlaps with the mitochondrial label MitoTracker (**Figure 4.3d**, red), confirming the microprotein's localization onto the mitochondria.

Because mitochondria are encapsulated by two membranes, an inner mitochondrial membrane (IMM) and an outer mitochondrial membrane (OMM), we next determined whether SLC35A4-MP resides in the IMM or the OMM. By using a proteinase K protection assay, we compared the effects of treating isolated mitochondria from HEK293T cells with proteinase K in the absence or presence of swelling caused by osmotic pressure. In the absence of swelling, proteinase K treatment causes degradation of OMM proteins but leaves IMM proteins intact, whereas in the presence of swelling the OMM is compromised allowing proteinase K to access and degrade IMM proteins (*25*). As controls for this assay, we measured the effects of proteinase K treatment on degradation of the OMM protein Tom20 and the IMM protein TIMM50 (*26*, *27*). Immunoblotting confirmed that proteinase K treatment in the absence of swelling caused degradation of the OMM protein Tom20 but left the IMM protein TIMM50 intact, whereas proteinase K treatment in the presence of swelling caused the degradation of both proteins (**Figure 4.3e**). SLC35A4-MP showed a similar pattern to what was observed for TIMM50

(**Figure 4.3e**), indicating that SLC35A4-MP is localized to the IMM. Overall, our findings indicated that SLC35A4-MP is a single transmembrane IMM microprotein.

Identification of SLC35A4-MP putative binders by IP/MS (3A-C)

Since functional microproteins are typically components of complexes containing proteins with known functions, elucidation of microprotein-protein interactions (MPI) provides an efficient method to accelerate understanding of microprotein functions (*5*, *28–30*). Therefore, we explored the interactome of SLC35A4-MP using a proteome-wide immunoprecipitation-mass spectrometry (IP-MS) approach. Initial attempts at these experiments using overexpressed epitope tagged-SLC35A4-MP were unsuccessful and highlighted a general challenge when characterizing small, mostly unstructured microproteins. Overexpression of SLC35A4-MP-FLAG in HEK293T cells resulted in excessive and unphysiological levels of the microprotein. This led to irreproducible datasets of enriched proteins including many predicted to be localized to non-mitochondrial cellular compartments. We realized that these inconsistent results may be attributed to the physical characteristics of microproteins in general. In particular, the small size of microproteins and their lack of rigid domain structures found in larger proteins enable them to interact with many proteins simultaneously when expressed at high concentrations, leading to the apparent false positives we observed in IP-MS experiments following SLC35A4-MP overexpression.

To overcome this challenge, we optimized immunoprecipitation conditions for endogenous SLC35A4-MP and utilized isolated mitochondria from HEK293T cells to reduce background and increase chances of identifying the most relevant SLC35A4-MP interaction partners. Immunoprecipitation of HEK293T mitochondrial lysates with rabbit immunoglobulin G

(IgG) or SLC35A4-MP antiserum followed by proteomics identified potential microprotein interacting proteins (**Chang_04_Supplementary Dataset1.xls., tab 1**). To refine our search, we curated the list of potential interactors by only retaining enriched proteins with at least 5 spectral counts and a minimum of 3 enrichment folds. This method narrowed down our search to 45 putative binding partners with SLC35A4-MP (**Chang_04_Supplementary Dataset1.xls., tab 2**). Using MitoCarta3.0 as a reference inventory of annotated mitochondrial proteins, we further excluded ubiquitous contaminants and non-mitochondrial proteins (*31*, *32*), restricting our search to a smaller list of 15 proteins (**Figure 4.4a**). The remaining proteins were then inputted onto the protein-protein interaction (PPI) database STRING to identify known protein complexes within this list of mitochondrial-localized putative binders (*33*) (**Figure 4.4a-c**).

The analysis revealed that the largest complex SLC35A4-MP interacts with contains CHCHD3 (MIC19), APOO (MIC26), IMMT (MIC60), ACO2, and IDH3B (**Figure 4.4b**). Of these, MIC19, MIC26, MIC60 are subunits of the mitochondrial contact site and cristae organizing system (MICOS) complex (*34–36*). Enriched at the crista junctions, the MICOS complex is a conserved, major regulator of crista architecture and biogenesis (*34–36*). Additionally, SLC35A4-MP pulled down Krebs cycle enzymes isocitrate dehydrogenase [NAD] subunit B (IDH3B) and aconitate hydratase (ACO2) (*37*, *38*).

Immunoblotting analysis of SLC35A4-MP putative binders

Indeed, our IP/MS workflow was optimized to retain protein hits with the highest likelihood of interacting with SLC35A4-MP. However, we found that not all proteins nominated by IP/MS were true interactors with SLC35A4-MP as our subsequent immunoprecipitation experiments indicated that ACO2 and IDB3B were false positives (**Chang_03_**

Supplemental_Information3.doc, Figure S3a-d). Therefore, a thorough biochemical validation using antibodies with high affinity and specificity is imperative for accurate determination of genuine MPIs.

Next, we validated the putative interactions between SLC35A4-MP and the remaining MICOS complex proteins MIC19, MIC26, and MIC60. While our workflow only retained 3 subunits of the human MICOS complex, the cristae-modulating complex consists of a total of seven subunits, including the additional proteins MIC10, MIC13, MIC25, and MIC27 (*39*). Although MIC10 was not initially identified in our IP-MS list (**Chang_04_Supplementary Dataset1.xls***)*, MIC13, MIC25, and MIC27 were enriched but conditionally removed by the filter in the raw data due to lower enrichment folds (**Chang_03_ Supplemental_Information3.doc, Figure S3e**). Because these proteins are known members of the MICOS complex and interact with other MIC proteins enriched by SLC35A4-MP, they were included in our subsequent validation experiments. In addition, we also included the dynamin-like GTPase OPA1 in our subsequent validation as it is epistatic to the core component of MICOS complex MIC60 and a regulator of crista structure (*40*, *41*).

Immunoblotting analysis demonstrates that all seven MICOS subunits are co-immunoprecipitated by pulldown of SLC35A4-MP from HEK293T lysates (**Figure 4.4d**). OPA1 was also enriched by SLC35A4-MP immunoprecipitation, but only from the mitochondrial fraction (**Chang_03_ Supplemental_Information3.doc, Figure S3f-g**). In support of these results, we further validated the interaction between SLC35A4-MP and the MICOS complex using reciprocal immunoprecipitations (**Figure 4.4e***,* **Chang_03_ Supplemental_Information3.doc, Figure S3h**). Pulldowns of HEK293T whole cell lysate with

MIC19, MIC25, MIC26, and MIC60 antibodies co-immunoprecipitated SLC35A4-MP, confirming SLC35A4-MP as a *bona fide* interactor with MICOS complex (**Figure 4.4e**).

Validation of *bona fide* interactions using SLC35A4-MP KO HEK293T cells

Next, we generated an SLC35A4-MP knockout (KO) HEK293T cell line via CRISPR/Cas9 (**Figure 4.4f**) and used these cells to ensure that enrichment with the SLC35A4-MP antiserum requires the presence of the microprotein. SLC35A4-MP was undetectable in KO cells by immunoblotting of whole cell lysates and immunoprecipitated samples (**Figure 4.4g**), and deletion of SLC35A4-MP did not affect the expression levels of MIC19, MIC26, and MIC60 in WT and SLC35A4-MP KO HEK293T cells (**Figure 4.4h**). Pulldown of SLC35A4-MP co-immunoprecipitated MIC19, MIC26, and MIC60 as expected in the WT lysate, but very weak or no signal was detected from SLC35A4-MP KO HEK293T cell lysates (**Figure 4.4h**). Together, these results provide robust evidence that SLC35A4-MP forms multiple contacts sites with MICOS complex, which suggests a role for SLC35A4-MP in crista structure and function.

Functional characterization of SLC35A4-MP

The biological functions of most eukaryotic microproteins have been discovered through MPIs, as the function of the binding partner can be leveraged to accelerate understanding of the microprotein (*5*, *28–30*). Several studies have shown that MICOS complexes play critical roles in regulating crista structure and respiratory functions (*34*, *35*, *42–47*) and the identification of MICOS proteins as SLC35A4-MP interacting partners suggests a similar role for SLC35A4-MP. For example, MIC10 and MIC60 are considered the core subunits of the MICOS complex as their depletion show the most pronounced phenotypes in CJ loss, making them the most extensively studied components within the complex (*42*, *48*). Additionally, individual cell lines

bearing MIC10 KO or MIC60 KO both demonstrated impaired oxidative capacity (*48*). To examine whether SLC35A4-MP shares similar functions with MICOS complex, we next assessed if deletion of the microprotein in SLC35A4-MP KO HEK293T cells also led to loss-of-function phenotypes in cristae morphology and cellular respiration.

Analysis of SLC35A4-MP's impact on cristae structure using EM

We used the three-dimensional electron microscopy (3DEM) technique of EM tomography to generate volumes of mitochondria to assess the structure of the cristae in WT and SLC35A4-MP KO HEK293T cells. Deletion of SLC35A4-MP did not alter the number of mitochondria between WT and KO cells (**Chang_03_ Supplemental_Information3.doc, Figure S4a**). However, morphological differences in the appearance of the crista structure were striking (**Figure 4.5a**). Comparison of tomographic volumes revealed that the absence of SLC35A4-MP led to reduction in crista density, defined as the crista membrane surface area divided by the mitochondrial volume (**Figure 4.5a, j**). To visualize the cristae structures in 3D, surface rendering were created by segmenting the OMM and each crista (**Figure 4.5b**). We found that the WT cristae were mostly arranged in stacks and had a lamellar shape with little branching, whereas the KO cristae were much smaller in size. Comparison between the mitochondrial volumes between WT and KO mitochondria also showed a prominent decrease in the number of CJs per mitochondria (**Figure 4.5c**), possibly due to less elongated CJs, or fewer CJs or fewer cristae. To distinguish between these possibilities, we first generated graphical models of the CJs by using illustrative rods to represent the extent of the CJ openings (**Figure 4.5d**), and the diameter of these rods were set at 16 nm as this number was determined by the mean CJ widths in both WT and KO mitochondria (**Figure 4.5e**). Examining the models from a

side and 30-degree tilt perspectives shows that the mitochondria from KO cells exhibited decreased CJ elongation (height) compared with its WT counterparts (**Figure 4.5d, f**). We next found that the CJ density (**Figure 4.5g**) and the number of CJs per crista (**Figure 4.5h**) were no different between WT and KO cell mitochondria, with slightly higher variance in the KO cell mitochondria. Finally, we interrogated whether there were fewer cristae in KO cell mitochondria and found that even though there was no difference in the number of cristae per mitochondrion between WT and KO cells (**Figure 4.5i**), there was around a 2-fold decrease in the crista density in KO cell mitochondria (**Figure 4.5j**), consistent with our observations of a subpopulation of smaller cristae in the KO cell mitochondria (**Figure 4.5b**). These results demonstrate that SLC35A4-MP not only interacts with the MICOS complex but also regulates the structure of cristae since genetic removal of the microprotein results in major effects on crista density, size, and morphology. We also conclude that even though SLC35A4-MP does not appear to be involved with the formation of CJs or CJ density, its presence is critical for maintaining the elongation of the CJ openings. These structural alterations led us to reason that deletion of SLC35A4-MP may have an impact on crista function and the generation of ATP.

Investigation of SLC35A4-MP's impact on oxidative capacity

Due to the altered structure and reduced number of cristae in SLC35A4-MP KO HEK293T cells, we hypothesized that the mitochondria of these cells may also have impaired oxidative capacity. Using a Seahorse XFe96 Analyzer, we interrogated how genetic ablation of SLC35A4-MP would impact the cells' oxidative capacity by sequentially introducing mitochondria-perturbing compounds, such as oligomycin, carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP), and a mixture of rotenone and antimycin A and

comparing the oxygen consumption rates (OCRs) between WT and SLC35A4-MP KO cells. Comparison of the maximal respiration rates stimulated by FCCP-induced uncoupling of OXPHOS indicates a mild but significant reduction of OCR in SLC35A4-MP KO HEK293T cells compared to WT cells (**Figure 4.6a-b***)*. We observed that even though loss of SLC35A4-MP resulted in pronounced reduction in crista morphology, deletion of the microprotein only modestly diminished respiration in SLC35A4-MP KO cells (**Figure 4.6b***)*. This observation is consistent with previous functional studies such that pronounced cristae defects caused by the deletion of a MICOS subunit do not always provoke strong OCR impairment phenotypes. For example, deletion of MIC10 in human HAP1 cells abolishes CJ densities, yet the OCR level was only modestly reduced (*48*). The OCR changes we observed between WT and SLC35A4-MP HEK293T cells appear to be the direct result of the deletion of SLC35A4-MP rather than loss of expression of OXPHOS proteins since the levels of these proteins were the same in both groups (**Chang_03_ Supplemental_Information3.doc, Figure S4b**). Additionally, we validated the role of SLC35A4-MP in regulating mitochondrial respiration by showing that knockdown of the microprotein in HeLaS3 cells using siRNA also causes a reduction in OCR (**Chang_03_ Supplemental_Information3.doc, Figure S4c-e**).

Moreover, we compared citrate synthase (CS) activity in WT and SLC35A4-MP KO HEK293T cells (**Chang_03_ Supplemental_Information3.doc, Figure S4f-g**). CS activity serves as a marker of mitochondrial fitness since low CS activity has been associated with compromised metabolic activity (*49*). Comparison of CS activities in mitochondrial fractions of WT and SLC35A4-MP KO cells show that CS activity was approximately 16% lower in KO cells compared to WT cells (**Chang_03_ Supplemental_Information3.doc**, **Fig. S4*G***). Because loss of SLC35A4-MP attenuated maximal respiration rates and CS activity, we next hypothesize

that the microprotein may also be a regulator of mitochondrial biogenesis. Using cellular ATP levels as a surrogate for energy production, we next quantify cellular ATP levels from viable HEK293T WT and SLC35A4-MP KO cells and found that deletion of SLC35A4-MP attenuated the cells ability to generate ATP (**Figure 4.6c**). Our data indicates that SLC35A4-MP is a regulator of oxidative metabolism as deletion of the microprotein results in lower respiratory input and biogenesis output. This result further supports an important role for SLC35A4-MP in regulating aerobic metabolism.

4.3 Conclusion

uORFs in the 5'-UTR of mRNAs are primarily known as post-transcriptional regulators of mORF translation. Here we show that the highly conserved *SLC35A4* uORF functions independently of mORF translational regulation and encodes a unique functional microprotein SLC35A4-MP. SLC35A4-MP is an inner mitochondrial membrane protein and a member of the mitochondrial contact site and cristae organizing system (MICOS) complex, essential for cristae shape formation and maintenance. Targeted deletion of SLC35A4-MP in human cells diminishes crista integrity and inhibits mitochondrial oxidative capacity. Identification of SLC35A4-MP as a component of the MICOS complex provides a potential molecular modality in elucidating the pathological phenotypes associated with this vital mitochondrial complex. Functional characterization of SLC35A4-MP is significant because it expands the conventional role of uORF-encoded microproteins.

4.4 Figures

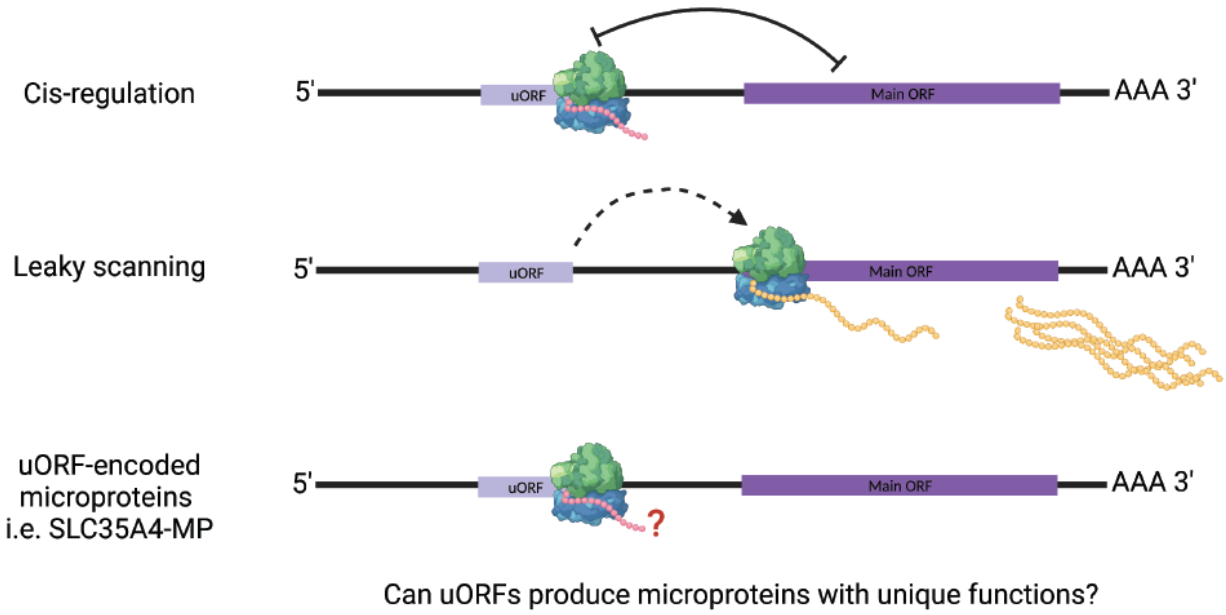


Figure 4.1. Schematic showing models of uORF translation regulation in functional bicistronic mRNAs. **Top**: uORFs serve as *cis*-acting elements as their translation can inhibit translation of the downstream mORF. **Middle**: Alternatively, in "leaky scanning" the ribosome can bypass the uORF translational start site to initiate translation of the downstream main ORF. **Bottom**: In some cases, translation of uORFs produces microproteins with novel biological functions.

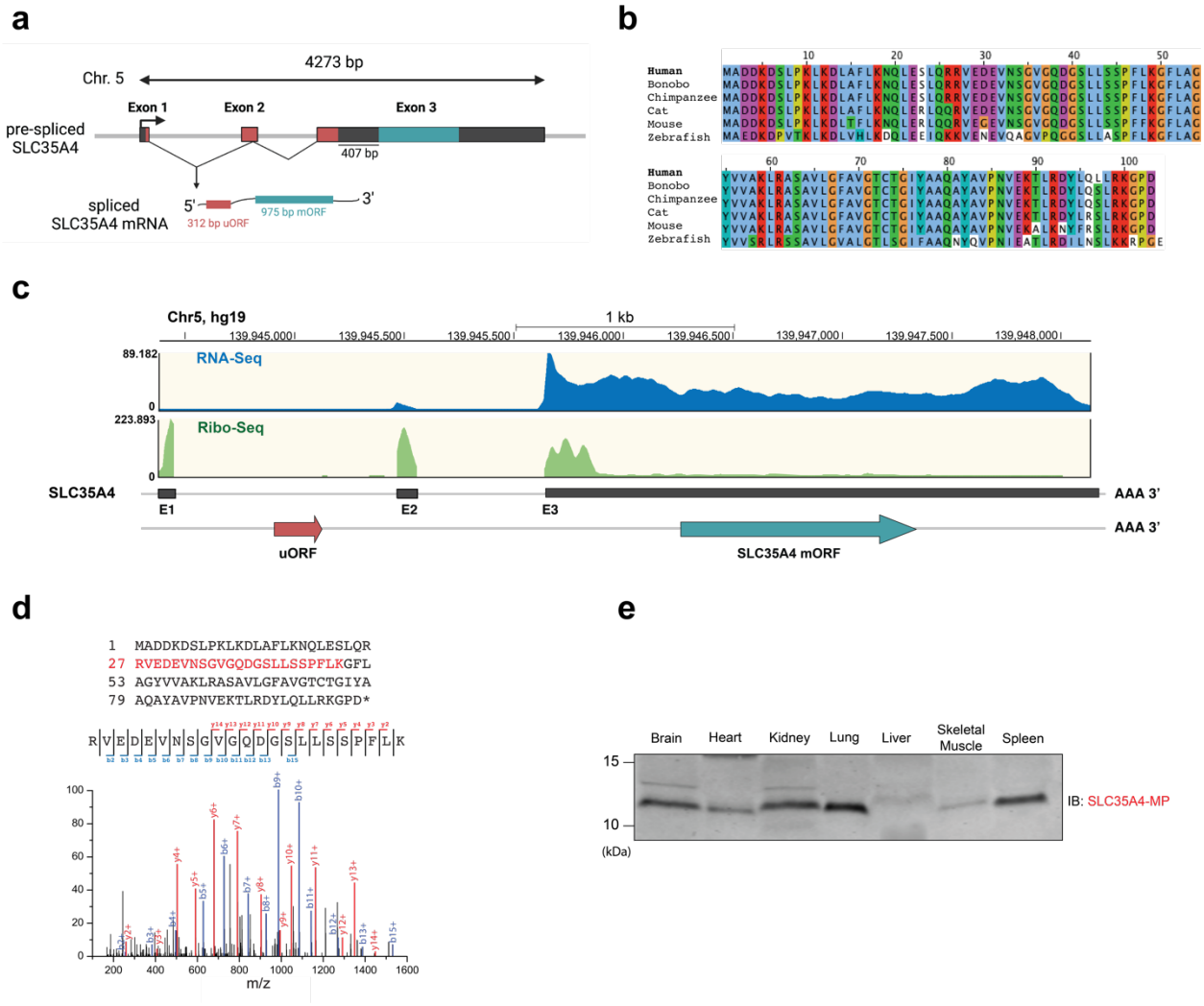


Figure 4.2. SLC35A4-MP is a conserved microprotein that is abundantly expressed in mammalian cells and tissues. (**a**) Schematic of the human SLC35A4 genomic locus. Located on chromosome 5, the SLC35A4 uORF (red) consists of 3 exons whereas the SLC35A4 mORF (turquoise) consists of 1 exon. (**b**) Amino acid sequence alignment of SLC35A4-MP shows that it is highly conserved across vertebrate species. The colors indicate the properties of individual amino acid residues (blue, hydrophobic; purple, negative charge; red, positive charge; and green, polar). (**c**) Ribo-Seq and RNA-Seq data indicate that the SLC35A4 mRNA contains an uORF comprising exons 1, 2 and part of exon 3 encoding the 103 amino acid SLC35A4-MP. (**d**) MS/MS spectrum of a unique tryptic peptide fragment corresponding to the region of SLC35A4-MP highlighted in red in (**c**) identified by mass spectrometry of human K562 and HEK293T cells. (**e**) Immunoblotting analysis of C8-extracted lysates prepared from the indicated tissues from C57BL6 mice was performed using an in house-generated rabbit polyclonal SLC35A4-MP antiserum. Protein loading was normalized using total protein amount of 100 µg as the extraction method used excluded actin, the standard loading control used for normalization.

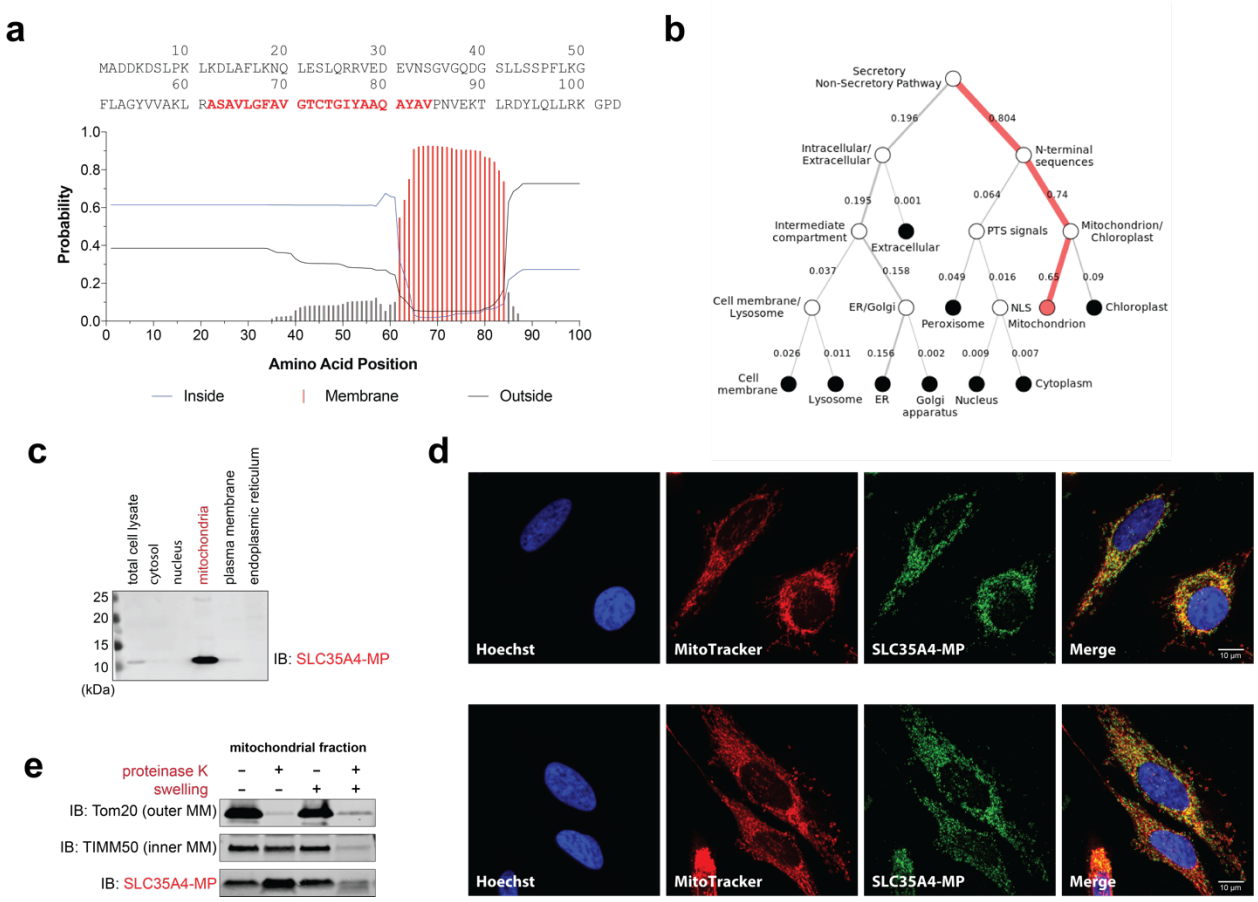


Figure 4.3. SLC35A4-MP is localized to the inner mitochondrial membrane. (**a**) Analysis of the SLC35A4-MP amino acid sequence using TMHMM Server v.2 identifies a high probability single-pass transmembrane domain (TMD) consisting of amino acid residues 62 through 84. The predicted transmembrane sequence is highlighted in red. (**b**) Algorithmic prediction of subcellular localization of SLC35A4-MP through deep learning using DeepLoc model. (**c**) Immunoblotting analysis of total lysate and lysates of subcellular fractions prepared from HEK293T cells using a rabbit polyclonal SLC35A4-MP antiserum (in-house). (**d**) Immunofluorescence of fixed HeLa cells shows co-localization of SLC35A4-MP and mitochondria. SLC35A4-MP immunolabeling is indicated in green, mitochondrial-specific MitoTracker labeling is indicated in red and nuclear Hoechst staining is in blue. Scale bar: 10 µm. (**e**) The mitochondria fraction of HEK293T cells was treated with or without osmotic pressure and proteinase K to cause loss of proteins associated with both inner and outer mitochondrial membranes or only outer mitochondrial proteins, respectively. Lysates were analyzed by immunoblotting using antibodies against Tom20 (outer mitochondrial membrane marker) and TIMM50 (inner mitochondrial membrane marker), and a rabbit polyclonal antiserum against SLC35A4-MP. MM, mitochondrial membrane.

Figure 4.4. SLC35A4-MP interacts with members of the MICOS complex. (**a**) Schematic IP/MS workflow for identifying putative binders of SLC35A4-MP in the mitochondria. (**b**) STRING analysis identifies SLC35A4-MP interacting partners from HEK293T mitochondrial enriched samples using workflow described in (**a**). The presented network illustrates a full STRING network types, excluding disconnected nodes. Network edges are based on evidence, and the active interaction sources include experiments and databases. Minimum interaction score: 0.4. (**c**) Bars representing fold enrichment for the proteins identified in (**b**) indicate the average spectral counts from pull downs with SLC35A4-MP antiserum versus those from pull downs using IgG control (MIC, MICOS). (**d**) Immunoblotting analysis of HEK293T whole cell lysates (WCL; input) and eluted proteins following immunoprecipitation with either IgG control or SLC35A4-MP antiserum. Proteins were detected using the indicated MICOS protein antibodies. (**e**) Immunoblotting analysis of HEK293T whole cell lysates (WCL; input) and eluted proteins following reciprocal immunoprecipitation with either IgG control or individual antibodies against MIC19, MIC25, MIC26, and MIC60. Proteins were detected using SLC35A4-MP antiserum. (**f**) Schematic illustrating the strategy used to knockout the SLC35A4-MP using CRISPR/Cas9-mediated genome editing. Two guide RNAs were designed to target the first and second exons of the SLC35A4 gene. **(g)** Immunoblotting analysis of HEK293T wild type (WT) and SLC35A4-MP knockout (KO) whole cell lysates (WCL; input) with SLC35A4-MP antiserum. **(h)** Immunoblotting analysis of HEK293T wild type (WT) and SLC35A4-MP knockout (KO) whole cell lysates (WCL; input) and eluted proteins following immunoprecipitation with SLC35A4-MP antiserum. Proteins were detected using the indicated MICOS protein antibodies or SLC35A4-MP antiserum (MIC, MICOS).

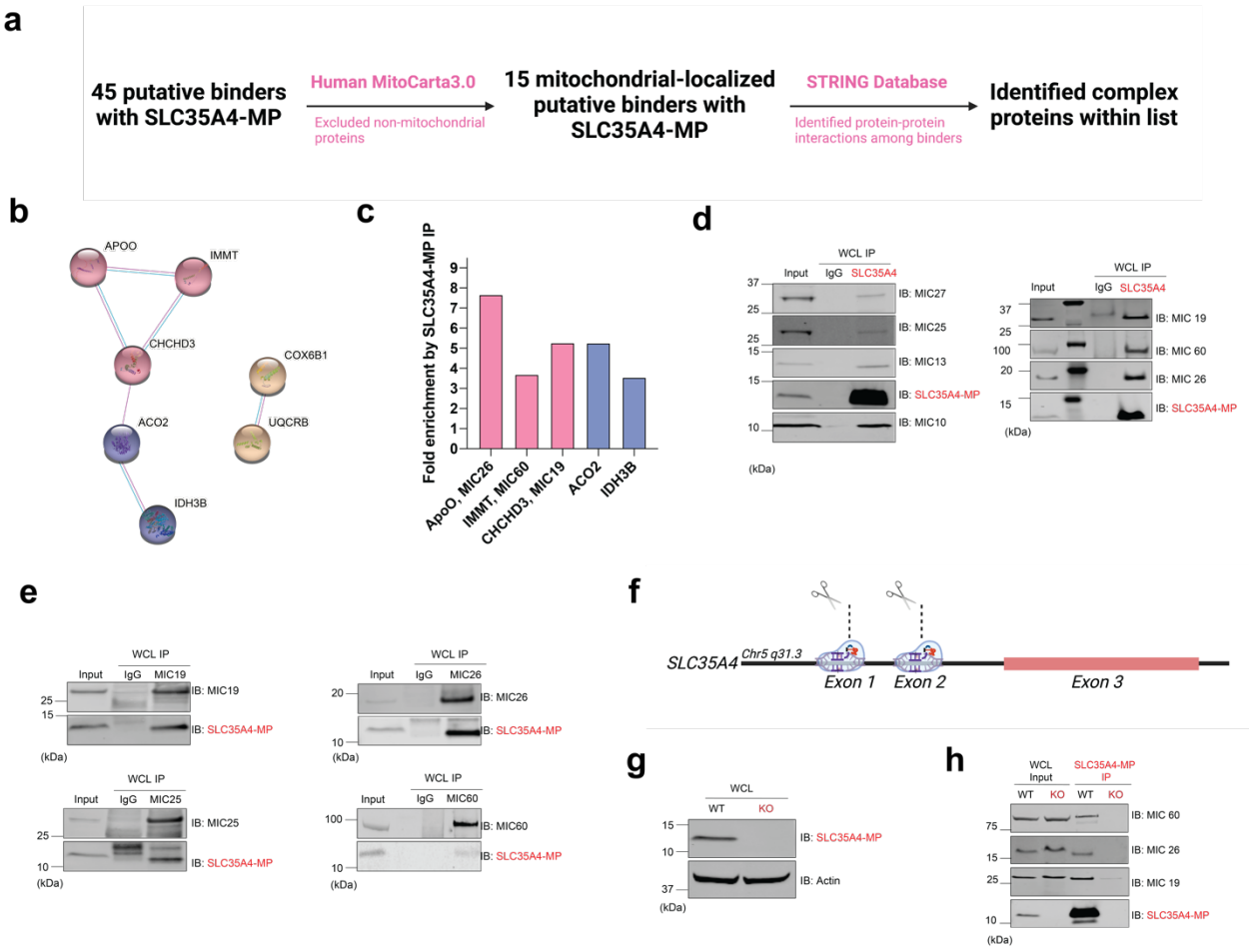
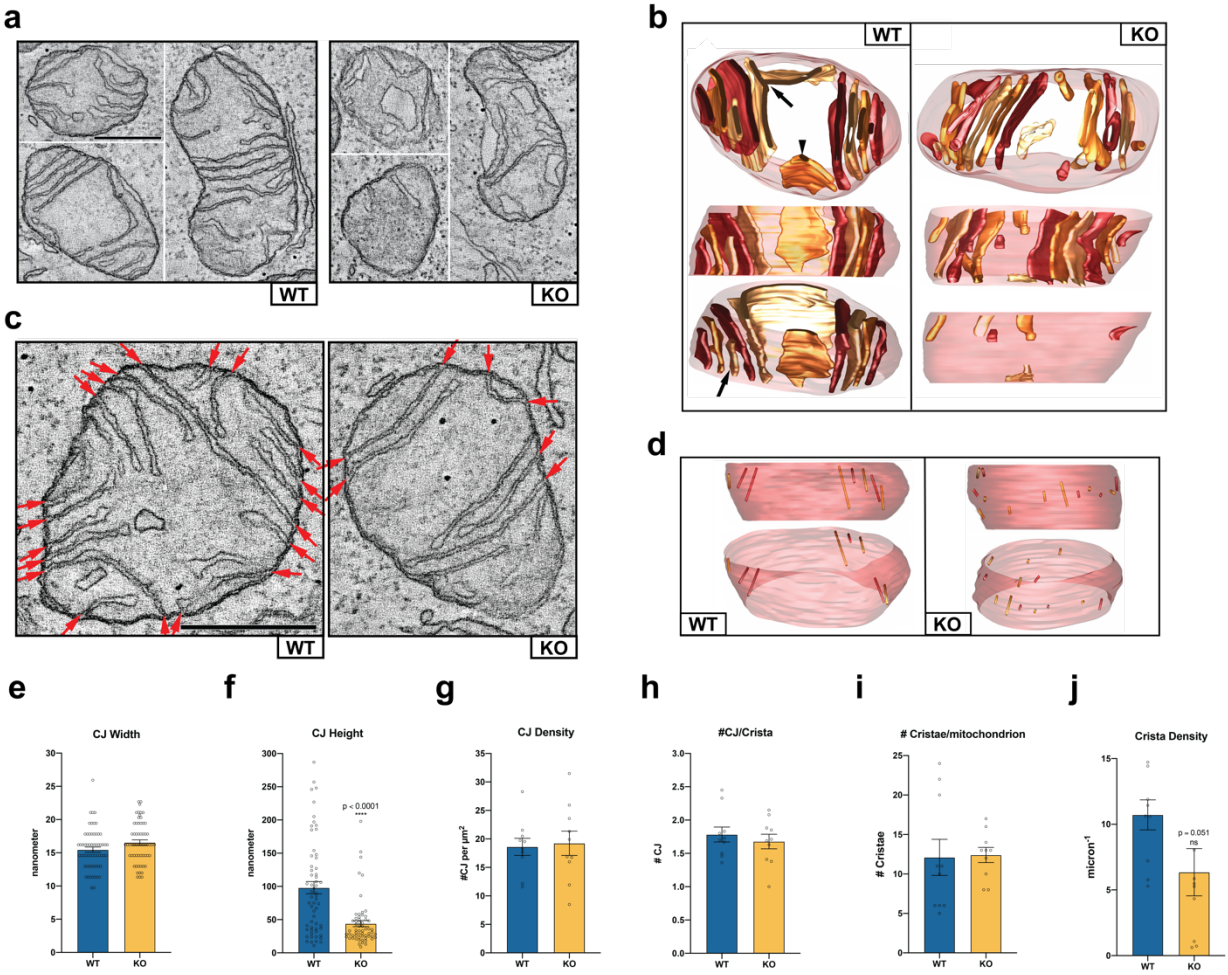

a
45 putative binders with SLC35A4-MP
Human MitoCarta3.0
Excluded non-mitochondrial proteins
15 mitochondrial-localized putative binders with SLC35A4-MP
STRING Database
Identified protein-protein interactions among binders
Identified complex proteins within list
b
APOO
IMMT
CHCHD3
COX6B1
ACO2
UQCRB
IDH3B
c
Fold enrichment by SLC35A4-MP IP
ApoO, MIC26
IMMT, MIC60
CHCHD3, MIC19
ACO2
IDH3B
d
WCL IP
Input
IgG
SLC35A4
IB: MIC27
IB: MIC25
IB: MIC13
IB: SLC35A4-MP
IB: MIC10
(kDa)
IB: MIC 19
IB: MIC 60
IB: MIC 26
e
IB: MIC19
IB: MIC26
IB: MIC25
IB: MIC60
f
SLC35A4
Chr5 q31.3
Exon 1
Exon 2
Exon 3
g
WCL
WT
KO
IB: Actin
h
SLC35A4-MP IP

Figure 4.5. Loss of SLC35A4-MP alters mitochondrial ultrastructure. (**a**) Slices of thickness 1.6 nm through the middle of EM tomography volumes of mitochondria from WT (left panels) with typical crista density and SLC35A4-MP KO (right panels) HEK293T cells with decreased crista density. Scale = 0.5 microns and applies to all panels. (**b**) Surface rendering of the outer mitochondrial membrane (translucent maroon) and all the cristae (various shades of brown) showing that all 10 cristae have a lamellar shape and little branching. *WT top*: Top view showing the mostly stack-like arrangement of the cristae and that the middle of the mitochondrion is devoid of cristae, which are common features of HEK293T cells. The arrow points to an unusual crista that both is part of the crista stack and also bends to follow the outer mitochondrial membrane. There is only 1 crista (arrowhead) that is not part of the stack and lines the outer mitochondrial membrane such that the empty middle is framed by almost a "box" of cristae. *WT middle*: Side view emphasizing the lamellar shape of the cristae and that they extend throughout the volume. *WT bottom*: View rotate 40 degrees to better visualize the curvature of the large crista (marked by an arrow at top) and to show the only crista (arrow) that does not extend throughout the volume. *KO top*: Top view showing all 15 cristae, a number of which are smaller. *KO middle*: Side view showing that the smaller cristae do not extend through the volume. *KO bottom*: Only the 6 smaller cristae are shown for comparison. **(c)** Example slices of thickness 1.6 nm through the middle of tomographic volumes of mitochondria from WT (left panel) and SLC35A4-MP KO (right panel) HEK293T cells suggesting either fewer crista junctions (CJ or less elongated CJs as indicated by the seemingly lower density of CJs in the KO. CJs are indicated by red arrows; 20 in the WT and 7 in the KO. Scale = 0.5 microns and applies to all panels. **(d)** Modeling of CJ dimensions in WT and SLC35A4-MP KO HEK293T cells with respect to the mitochondrial outer membrane. The outer membrane is shown in translucent maroon and the CJs as rods. (e-j) Quantitation of crista junctions (CJ) and cristae comparing WT and SLC35A4-MP KO HEK293T cells. Data represent mean ± S.E.M. Statistical significance was performed using two-sided Student's t test. **(e)** CJ width; N = 60 junctions; NS, $p = 0.0592$, NS. **(f)** CJ height (elongation); N = 60 junctions; ****, $p < 0.0001$. For CJ width and height, the measurement tool was accurate to 1 pixel (1.6 nm) accounting for the clustering of values. **(g)** CJ density defined as the total number of crista junctions in a mitochondrion divided by the outer membrane surface area of that mitochondrion; N = 10; NS, $p = 0.8175$. **(h)** The number of CJs per crista defined as the total number of junctions divided by the total number of cristae in a mitochondrion; N = 10; NS, $p = 0.5114$. **(i)** The number of cristae per mitochondrion; N = 10, $p = 0.9050$. **(j)** Crista density defined as the total number of cristae in a mitochondrion divided by the outer membrane surface area of that mitochondrion; N = 10, $p = 0.0551$.


a
WT
KO
b
WT
KO
c
WT
KO
d
WT
KO
e
CJ Width
nanometer
WT
KO
f
CJ Height
nanometer
p < 0.0001
****
WT
KO
g
CJ Density
#CJ per µm²
WT
KO
h
#CJ/Crista
CJ
WT
KO
i
Cristae/mitochondrion
Cristae
WT
KO
j
Crista Density
micron⁻¹
p = 0.051
ns
WT
KO

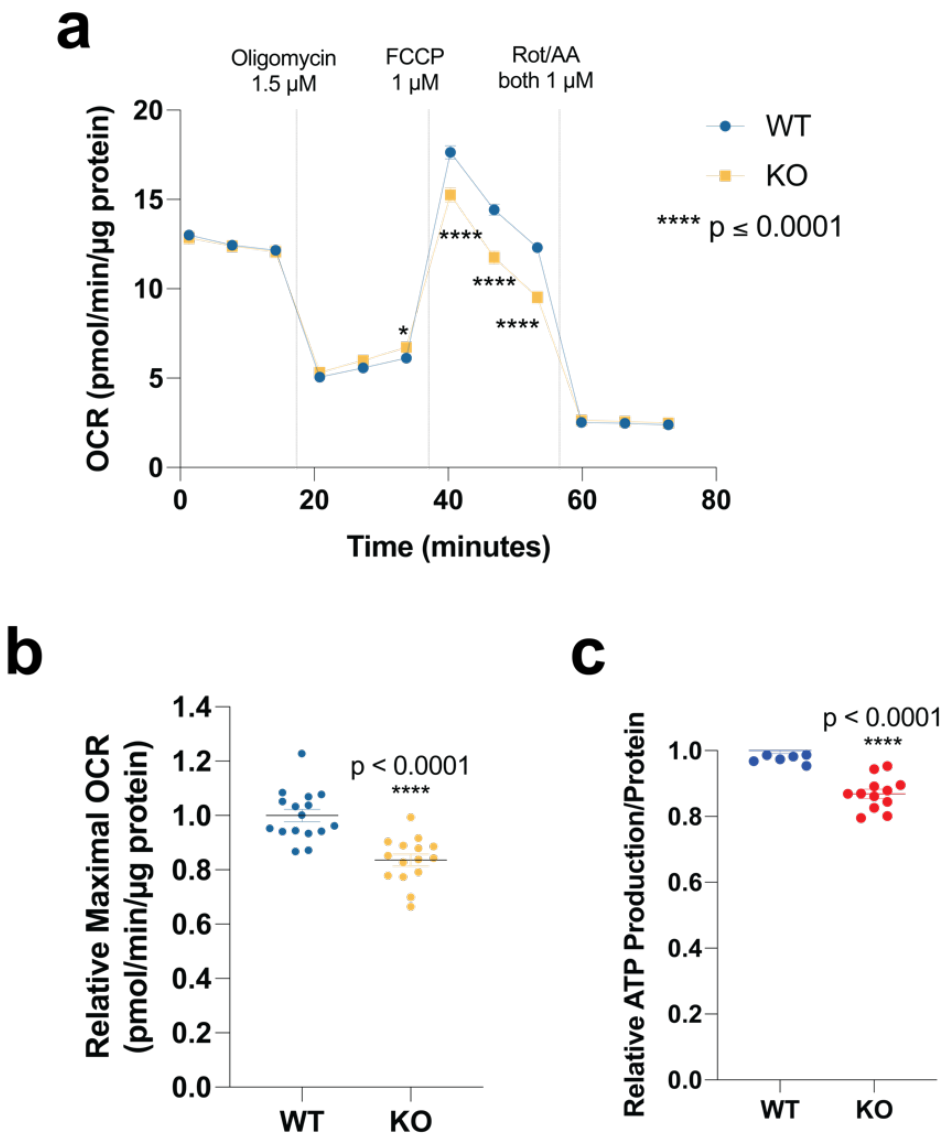


Figure 4.6. Genetic ablation of SLC35A4-MP attenuates mitochondrial function. (**a**) Oxygen consumption rates (OCR) of WT and SLC35A4-MP KO HEK293T cells was measured using a Seahorse XF96 extracellular flux analyzer, with sequential additions of the metabolic inhibitors/activators oligomycin rotenone/antimycin (Rot/AA) (OCR, oxygen consumption rate; Data represent mean ± S.E.M (n = 15-16). ****, $p < 0.0001$ by two-sided Student's t test.). (**b**) Quantification of maximal OCR from (**a**) representing the maximal cellular respiration following injection of FCCP, a mitochondrial uncoupler. Data represent mean ± S.E.M (n = 15-16). Statistical significance was performed using two-sided Student's t test. (**c**) Quantitation of cellular ATP levels using a CellTiter Glo assay in viable HEK293T WT and SLC35A4-MP KO HEK293T cells. Data was normalized to protein concentration using a BCA protein assay. Data represent mean ± S.E.M (n = 12). Statistical significance was performed using two-sided Student's t test.

4.5 Materials and Methods

Cell lines and culturing

HeLa S3 (CCL-2) and HEK293T (CRL-11268) cell lines were purchased from American Type Culture Collection (ATCC). Both cell lines were cultured in DMEM (Corning, 10-013-CV) supplemented with 10% fetal bovine serum (FBS) at 37 °C and 5% CO_2.

Generation of SLC35A4-MP knockout cell line

SLC35A4-MP KO HEK293T cells were generated using two designed guide RNAs targeting exon 1 and exon 2 of the SLC35A4 genomic region (gRNAs: GGGGAAGATGGCGGATGACA and GCAGCGGCGTGTAGAAGACG, respectively). Guide RNAs were purchased from GenScript and cloned into a pSpCas9 BB-2A-Puro (PX459) vector before transfecting into HEK293T cells using Lipofectamine 2000 (Thermo Fisher Scientific, 11668-019) and Opti-MEM (Gibco, 31985-070). Following a 24-h transfection, media was replaced with fresh DMEM supplemented with 1.25 μg/mL puromycin (SCBT, sc-108071) for selection. Media containing puromycin was replaced daily until removal of non-transfected cells was complete. The surviving cells were validated for knockout efficiency by western blot using the in-house rabbit polyclonal SLC35A4-MP antiserum prior to propagation.

Transfection of siRNA

Transfection of plasmid constructs was done using RNAiMAX (Thermo Fisher Scientific, 13778-075) in Opti-MEM (Gibco, 31985-070) according to instructions provided by the manufacturers. Briefly, 500,000 cells/well were seeded onto a 6-well plate. Upon cell

adherence, 25 pmol of siRNA (med-GC or si-SLC35A4-MP) was transfected into each well using RNAiMAX and Opti-MEM and incubated for 48 h.

Preparation of whole cell lysate for immunoprecipitation and western blot

Trypsinized cells were collected, spun down at 600g x 3 min, and washed with warm PBS. The cell pellet was then lysed with Lysis Buffer (50 mM Tris, 150 mM NaCl, 1% Triton X-100, and protease inhibitor cocktail) on ice with occasional vortexing. Following a 15-min incubation, the cells were centrifuged at 13,000 g x 10 min and the supernatant was collected as whole cell lysate.

Subcellular fractionation

Isolation of individual subcellular fractions from HEK293T cells was performed as described previously with slight modifications (*29*, *50*). HEK293T cells were homogenized in ice cold Isolation Buffer (225 mM *d*-mannitol, 75 mM sucrose, 0.1 mM EGTA, and 30 mM Tris-HCl pH 7.4) until most cells were lysed. Then, the homogenate was centrifuged at 600g for 5 min to collect pellet **P1** and clarify the supernatant **S1**. Pellet **P1** containing the nuclear fraction was washed three additional times with Isolation Buffer and the final pellet was resuspended in RIPA Buffer (Thermo Fisher Scientific, 89901). Supernatant **S1** was then centrifuged at 7,000g for 15 min to isolate the crude mitochondrial fraction in pellet **P2** from the supernatant **S2**. Pellet **P2** was then washed with Isolation Buffer twice and the final pellet was resuspended in RIPA buffer (crude mitochondrial fraction). The remaining supernatant **S2** was further centrifuged at 20,000g for 30 min to separate the plasma membrane in pellet **P3** from the supernatant **S3**, which contains the ER and cytosolic fractions. Following two additional Isolation Buffer washes, the

final pellet of **P3** was resuspended in RIPA Buffer (plasma membrane fraction). To isolate the ER and cytosolic fractions, **S3** was further centrifuged at 100,000g for 1 h. The supernatant (cytosolic fractions) was then saved while the pellet **P4** was washed three additional times with Isolation Buffer. The final pellet of **P4** was then resuspended in RIPA Buffer (ER fraction).

Mitochondrial isolation IP-MS analysis

Isolation of pure mitochondrial fraction from HEK293T cells was performed as described previously with slight modifications (*29*, *50*). Briefly, for each replicate 15 x 15 cm dishes of HEK293T cells were grown to confluence. Cells were trypsinized, washed with ice cold PBS and centrifuged at 600g for 5 min to collect the cell pellet. Using a dounce homogenizer, a pellet of pooled cells was homogenized in 20 mL of ice-cold Isolation Buffer (225 mM mannitol, 75 mM sucrose, 50 mM HEPES pH 7.5, 1 mM PMSF (Sigma, 78830) supplemented with protease inhibitor cocktail (Roche, 05892791001) until most cells were lysed. The homogenate was then centrifuged at 600g for 5 min to remove cell debris and clarify the supernatant. The supernatant was then centrifuged at 7,000g for 20 min to pellet the crude mitochondria. The pellet was resuspended in 10 mL of Isolation Buffer and further centrifuged at 10,000g for 10 min to remove contaminations by plasma membrane, cytosolic fractions, lysosomes, and microsomes. The final pellet was resuspended in Resuspension Buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 10% glycerol, 1 mM PMSF, and 1% digitonin (Biosynth Carbosynth, D-3200) for downstream biochemical analysis or MS preparation.

Proteinase K protection assay

Proteinase K protection assay was performed as described (*29*, *51*). Briefly, pure mitochondrial samples were prepared from HEK293T cells using the method described as above. The pellets were then resuspended in Isolation Buffer (225 mM mannitol, 75 mM sucrose, 50 mM HEPES pH 7.5) or in hypotonic buffer (2 mM HEPES pH 7.5) and incubated on ice for 30 min. Each buffer condition was then equally divided into two samples for a total of 4 samples; proteinase K (New England Biolabs P8107S) was added to one sample per condition at a final concentration of 50 μg/mL. Following a 30-min incubation on ice, 1 mM PMSF was added to the samples to inactivate proteinase K activity. All resulting samples were then precipitated with 12% TCA and separated by SDS-PAGE followed by immunoblotting.

Measurements of oxygen consumption rate (OCR)

The analysis was performed using a Seahorse XFe96 Analyzer (Agilent) as described previously with slight modifications (*52*, *53*). Briefly, cells were detached using trypsin (Gibco, 25200-056), resuspended in DMEM with 10% FBS, and equally divided into two collection tubes per siRNA condition or per genotype before spinning down the pellets at 600g x 3 min. One tube was used for western blot while the other was used for reseeding the cells onto the Seahorse plate. For western blot, a whole cell lysate was prepared and the protein concentration was determined using BCA protein assay (Thermo Fisher Scientific, 23225) prior to SDS-PAGE. For reseeding, the pellet was resuspended in DMEM with 10% FBS and then equally seeded onto a 0.01% (w/v) poly-L-lysine precoated 96-well Seahorse plate at 1 x 10^4 cells/100μL/well. Following a 16-h incubation, the media was replaced with prewarmed XF DMEM medium pH 7.4 (Agilent Technologies, 103575-100) supplemented with 10 mM D-glucose, 1 mM sodium

pyruvate (Gibco, 11360-070), and 2 mM glutamax (Gibco, 35050-061). To perform a mito stress assay, 1.5 μM oligomycin (Sigma, O4876), 1 μM carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (CCCP, Sigma, 2920), and 1 μM rotenone (Sigma, R8875) and antimycin A (Sigma, A2754) were sequentially added to the assay. Assay cycle included 3 min of mixing and 3 min of measurements. Upon completion of an assay run, the cell plate was retrieved and the cells were lysed with RIPA buffer. The protein concentrations of replicates were determined using a BCA protein assay.

Measurements of citrate synthase (CS) activity

Determination of CS activity of isolated mitochondria prepared from HEK293T and SLC35A4-MP KO cells was performed as previously described (*54*). Briefly, mitochondrial fractions were isolated from the cells using the MitoCheck® Mitochondrial isolation kit (Cayman, 701010). Protein concentrations were subsequently determined using a BCA assay and the more concentrated sample was adjusted to match the protein concentration of the lower concentrated sample. Next, using the MitoCheck® Citrate Synthase Activity Assay Kit (Cayman, 701040), reactions were initiated by the addition of the diluted samples or positive control and absorbance values were measured every 30 seconds for 20 min at 412 nm using a BioTek Synergy H5 microplate reader. Plotting the absorbance values over time generated a curve, which monitored the turnover rate of oxaloacetate using a readable signal generated by its byproduct thiol-coenzyme A at 412 nm. The reaction rates were calculated by determining the slope of the reaction curve. The reaction rate was then converted to specific activity and normalized based on protein concentration.

Quantitation of cellular ATP

ATP cell viability assay was performed using CellTiter-Glo® (Promega, G9241) according to manufacturer's instructions. Briefly, 30,000 HEK293T WT or SLC35A4-MP KO cells per well were seeded onto a 96-well plate. After adherence overnight, cells were subjected to CellTiter-Glo reagent treatments for 2 min on an orbital shaker. Following a 10-min incubation at room temperature, the luminescence signal was measured using a BioTek Synergy H5 microplate reader.

Confocal imaging

HeLa S3 cells were seeded onto coverslips precoated with 50 µg/mL poly-L-lysine (Sigma, P1399). Grown to approximately 70% confluence, cells were stained with MitoTracker Deep Red FM (Life Technologies, M22426) to label the mitochondria. Cells were then fixed with 4% paraformaldehyde (Polysciences, Inc., #18814) followed by permeabilization with 0.2% Triton X-100. Next, cells were incubated with antigen affinity purified SLC35A4-MP antiserum at 1:500 dilution overnight at 4°C. After washing with PBS + 0.1% Triton X-100, cells were stained with Alexa-488 goat anti-rabbit secondary antibody for 1 h at room temperature. Nuclei were counterstained with Hoechst 33258 (Sigma, #94403, 1:2000 in PBS) and then imaged using a Zeiss LSM 880 Airyscan confocal microscope.

Fixation of HEK293T for EM

Cells were cultured in 10-cm dishes until 70% confluence. Cells were fixed with 2.5% glutaraldehyde (Electron Microscopy Sciences, 16019) and 2.5 mM calcium chloride (Thermo Fisher Scientific, BP510-500) in 0.1 M sodium cacodylate, pH 7.4 (Electron Microscopy

Sciences, 12300) for 5 min followed by another 60-min on ice. Following 0.1 M sodium cacodylate buffer pH 7.4 rinses, cells were blocked in 10 mM glycine (VWR International, 470301-176) and 10 mM potassium cyanide (Sigma-Aldrich, 60178) for 15 min on ice and washed an additional 3 times with 0.1 M sodium cacodylate buffer. A solution of 3% potassium ferrocyanide (Acros Organics, cat. no. 424130050) in 0.3 M cacodylate buffer and 4 mM calcium chloride was combined with an equal volume of 4% (wt/wt) aqueous osmium tetroxide (Electron Microscopy Sciences, 19150) then added to the samples and incubated for 30 min on ice. After rinsing the cells in cold double deionized water (DDW), the cells were placed in 2% aqueous uranyl acetate for 1 h. The cells were then rinsed in cold DDW before dehydration series using 20%, 30%, 50%, 70%, 90%, 100% ethanol. A Durcupan ACM resin was prepared as follows: 11 g Part A (Sigma-Aldrich, cat. no. 44611), 10 g Part B (Sigma-Aldrich, 44612), 0.3 g Part C (Sigma-Aldrich, 44613), and 0.1 g Part D (Sigma-Aldrich, 44614). The samples were placed into 50% solution of Durcupan in ethanol for 30 min at room temperature followed by 100% Durcupan for 30 min. The resulting sample was then placed at 60°C for 48 h before analysis.

Transmission electron microscopy

Thin sections about 70 nm thick were cut from the cell blocks with a Leica ultramicrotome and placed on 200-mesh uncoated thin-bar copper grids. A Tecnai Spirit (FEI; Hillsboro, OR) electron microscope operated at 120 kV was used to record images with a Gatan 2Kx2K CCD camera at 6.0 nm/pixel. ImageJ was used to measure the number of mitochondria and cytoplasmic areas.

EM tomography

Semi-thick sections of thickness about 400 nm were cut from the cell blocks with a Leica ultramicrotome and placed on 200-mesh uncoated thin-bar copper grids. 20-nm colloidal gold particles were deposited on each side of the grid to serve as fiducial cues. The grids were irradiated for about 15 min before initiating a double-tilt series to limit anisotropic specimen thinning during image collection. During data collection, the illumination was held to near parallel beam conditions. Tilt series were captured using SerialEM (University of Colorado, Boulder, CO) software on a Tecnai HiBase Titan (FEI; Hillsboro, OR) electron microscope operated at 300 kV and 0.81 nm/pixel. Images were recorded with a Gatan 4Kx4K CCD camera. Each double-tilt series consisted of first collecting 121 images taken at 1 degree increment over a range of -60 to +60 degrees followed by rotating the grid 90 degrees and collecting another 121 images with the same tilt increment. To improve the signal-to-noise ratio, 2x binning was performed on each image by averaging a 2x2 x-y pixel box into 1 pixel.

The IMOD package (https://en.wikipedia.org/wiki/IMOD_(software)) was used for alignment, reconstruction, and volume segmentation. R-weighted back projection was used to generate the reconstructions. Volume segmentation of the mitochondrial outer membrane was performed using IMOD by tracing in each of the 1.62 nm-thick *x-y* planes that the object appeared that then created stacks of contours with the Drawing Tools plug-in in IMOD. The traced contours were then surface-rendered by turning contours into meshes to generate a 3D model. The surface-rendered volumes were visualized using 3DMOD. Measurements of crista junction width, height, and number, crista number, and mitochondrial outer membrane surface area were made within tomographic volumes following volume segmentation using the Drawing Tools plug-in for IMOD and reported using IMODinfo. Crista density, defined as the total

membrane surface area of the cristae divided by the volume of the mitochondrion, was measured using the ImageJ stereology plug-in on slices from tomographic volumes.

Antisera production-animal care

All animal procedures were approved by the Institutional Animal Care and Use Committee of the Salk Institute and were conducted in accordance with the PHS Policy on Humane Care and Use of Laboratory Animals (PHS Policy, 2015), the U.S. Government Principles for Utilization and Care of Vertebrate Animals Used in Testing, Research and Training, the NRC Guide for Care and Use of Laboratory Animals (8th edition) and the USDA Animal Welfare Act and Regulations. All animals were housed in an AAALAC accredited facility in a climate-controlled environment (65-72 degrees Fahrenheit, 30-70% humidity) under 12 light/12 dark cycles. Upon arrival, animals were physically examined by veterinary staff for good health and acclimated for at least two weeks prior to initiation of antisera production. Each animal was monitored daily by the veterinary staff for signs of complications and weighed every two weeks. Routine physical exams were also performed by the veterinarian quarterly on all rabbits. For production of antisera against human SLC35A4 MP, three 10- to 12-week-old, female New Zealand white rabbits, weighing 3.0 to 3.2 kg at beginning of the study, were procured from Irish Farms (I.F.P.S. Inc., Norco, California, USA). Rabbits were provided with ad libitum feed (5326 Lab Diet High Fiber), micro-filtered water and weekly fruits and vegetables and alfalfa hay for enrichment.

Generation of SLC35A4 MP antisera

Antisera against SLC35A4 was raised in rabbits against a synthetic peptide fragment encoding human Cys^{34} SLC35A4(2-34) coupled to maleimide activated keyhole limpet hemocyanin (Thermo Fisher Scientific) per manufacturer's instructions. The SLC35A4 peptide, ADDKDSLPKLKDLAFLKNQLESLQRRVEDEVNC, was synthesized, C18 HPLC purified to 99%, and amino acid sequence verified using mass spectrometry by RS Synthesis (Louisville, KY). The immunogen was prepared by emulsification of Freund's complete adjuvant-modified *Mycobacterium butyricum* (EMD Millipore) with an equal volume of phosphate buffered saline (PBS) containing 1.0 mg conjugate/ml for the first two injections. For booster injections, incomplete Freund's adjuvant was mixed with an equal volume of PBS containing 0.5 mg conjugate/ml. For each immunization, an animal received a total of 1 ml of emulsion in 20 intradermal sites in the lumbar region, 0.5 mg total protein conjugate for the first two injections and 0.25 mg total protein conjugate for all subsequent booster injections. Three individual rabbits were injected every three weeks and were bled one week following booster injections, <10% total blood volume. Rabbits were administered 1-2 mg/kg Acepromazine IM prior to injections of antigen or blood withdrawal. At the termination of study, rabbits were exsanguinated under anesthesia (ketamine 50 mg/kg and acepromazine 1 mg/kg, IM) and euthanized with an overdose of pentobarbital sodium and phenytoin sodium (1 ml/4.5 kg of body weight IC to effect). After blood was collected the death of animals was confirmed. All animal procedures were conducted by experienced veterinary technicians, under the supervision of Salk Institute veterinarians.

Characterization and purification of SLC35A4 MP antisera

Each bleed from each animal was tested at multiple doses for the ability to recognize the synthetic peptide antigen; bleeds with highest titers were further analyzed by western immunoblot for the ability to recognize the full-length endogenous protein and to check for cross-reactivity to other proteins. Antisera with the best characteristics of titer against the synthetic peptide antigen, ability to recognize the endogenous protein, and specificity were antigen affinity purified and used for all studies. Rabbit PBL#7383 was purified using human Cys^{34} SLC35A4(2-34) coupled to SulfoLink coupling resin (Thermo Fisher Scientific Cat #20401), 2 mg peptide on 2 ml bed volume resin, per manufacturer's instructions. To ensure that the same batch of purified antibody could be used for this study, a large volume, ~ 20 ml sera pooled from bleeds with similar profiles, was purified. The pooled serum was mixed with an equal volume of PBS, filtered through a 5 mm syringe filter, and then rotated overnight with SLC35A4(2-34)-SulfoLink resin. Unbound material was collected using a column, the resin was washed four times, using 2 column volumes each time, with 50 mM NaHepes, 100 mM NaCl, 0.1% Triton X-100, pH 7.5 followed by washing three times, using 2 column volumes each time, with 100 mM NaCitrate pH 4.5. Finally, the column was eluted with 1 N acetic acid and fractions of 1 column volume each were collected. The fractions were monitored for protein and the immunoglobulins containing fractions were pooled and dialyzed against phosphate buffered saline at 4C using a Slide-A-Lyzer cassette MWCO 7K (Thermo Fisher Scientific).

Preparation of antiserum linked resin for immunoprecipitations

Antigen affinity purified SLC35A4 MP antiserum was covalently coupled to Affi-Gel10 N-hydroxysuccinimide activated resin per manufacturer's instructions (Bio-Rad Cat #1536099),

1.2 mg antiserum to 1.5 ml bed volume resin. For control resin, a similar ratio of rabbit IGG whole molecule, Rockland Cat # 011-0102, was covalently linked to Affi-Gel 10. Following coupling of immunoglobulins to resins by rotating overnight at 4C, the resins were washed with 10 column volumes of 1 N acetic acid to remove any unbound antiserum and then equilibrated and stored at 4C in PBS with 0.05% sodium azide as 25% gel slurries.

Extraction of mouse tissues for western blot analysis

Tissues were extracted using a mixture of hot (90°C) 1 N acetic acid/0.1 N HCl, homogenized with a polytron blender, centrifuged at 30K g for 30 min at 4°C, and supernatants removed and filtered through 5 mm syringe filters. Supernatants were enriched for microproteins using BondElut C8 silica cartridges (Agilent), ~10 mg extracted protein per 200 mg sorbent. Cartridges were prepared with one column volume methanol and equilibrated with two column volumes triethylammonium formate (TEAF) pH 3.0 before the samples were applied. The cartridges were washed with two column volumes TEAF and the enriched microprotein fraction eluted with acetonitrile:TEAF pH 3.0 (3:1) and lyophilized to dryness using a Savant Speed-Vac concentrator.

SDS-PAGE and immunoblotting

Samples were loaded onto 4-12% Bis-Tris gels (Thermo Fisher Scientific, NW04120BOX) and run in 1X MES-SDS buffer (Thermo Fisher Scientific, B0002) at 200 V for 27 min. Gels were transferred onto PVDF membranes (Thermo Fisher Scientific, IB24001) using the iBlot 2 Dry Gel Transfer System (Thermo Fisher Scientific, IB21001) under the following steps: 20 V for 1 min, 23 V for 4 min, and 25 V for 2 min. Membranes were blocked with

Intercept® (TBS) Blocking Buffer (LI-COR, 927-60001) at room temperature for 1 h prior to the addition of corresponding primary antibodies at desired dilutions (**Chang_06_Supplementary Table1.xls.**). Membranes were incubated in primary antibodies overnight on a shaker at 4°C. Upon removal of membranes from primary antibodies, membranes were washed with TBS-T (0.1% Tween-20) and incubated in Alexa-fluor labelled secondary antibodies at desired dilutions (**Chang_06_Supplementary Table1.xls**) for 1 h. Following TBS-T washes, images of membranes were captured using LiCor Odyssey CLx at IR700 and IR800. Subsequent analysis and quantitation were performed using ImageStudio.

Forward immunoprecipitations

Lysates were added to washed IgG- or purified SLC35A4-MP antiserum coupled to Affi-Gel 10 beads and rotated at 4 °C overnight. Beads were washed and rotated in TBS-T (0.1% Tween-20) 5 times, with a minimum of 20 min each time. 2X sample buffer (Thermo Fisher Scientific, 84788) or SDS loading buffer with 2-mercaptoethanol (Sigma Aldrich, M6250) were added to the beads for western blot and MS analysis, respectively. Bound proteins were eluted by heating at 95 °C for 10 min. Following centrifugation at 5000 rpm for 2 min, samples were collected without disrupting the beads. For western blot, 2-mercaptoethanol was added to samples prior to loading.

Reciprocal immunoprecipitations

A whole cell lysate was prepared from at least 8 confluent 15cm dishes of HEK293T cells. Normal rabbit (in-house), normal mouse IgG (SCBT, sc-2025) or corresponding primary antibody was added to lysates and rotated at 4 °C overnight. Antibody-incubated samples were

added to pre-washed protein A/G PLUS-agarose beads (SCBT, sc-2003) and rotated at 4 °C overnight. Beads were washed and rotated in TBS-T (0.1% Tween-20), a minimum of 20 min each time. Bound proteins were eluted from beads by adding 2X sample buffer (Thermo Fisher Scientific, 84788) and heating at 95 °C for 10 min. Following centrifugation at 5000 rpm for 2 min, samples were then collected without disrupting the beads. 2-mercaptoethanol was added to samples prior to loading.

Preparation of MS samples

Trypsin digestion and sample preparation for MS/MS was performed as previously (*55*). Briefly, samples were precipitated with 6.1 N trichloroacetic acid (TCA) (MP Biomedicals, 196057) to reach a final TCA concentration of 23% (v/v) and incubated on ice overnight. Protein precipitate was washed with ice cold acetone twice before dissolving in 8 M urea (Alfa Aesar, 36428) followed by reduction with tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Thermo Fisher Scientific, 20491, 5 mM final concentration) then alkylation with 2-chloroacetamide (TCI, C2536, 10 mM final concentration). Upon addition of triethylammonium bicarbonate buffer (TEAB, Sigma, T7408, final urea concentration is 2 M), samples were digested overnight at 37 °C in the dark by trypsin (Promega, V5111). Reaction was quenched with formic acid to achieve a final concentration of 5% (v/v). The digested sample was then centrifuged at 15,000g for 15 min and the supernatant was collected for further analysis on a Q Exactive mass spectrometer (Thermo Fisher Scientific).

4.6 Acknowledgments

Chapter 4, in full, is being prepared for publication of the material with minor modifications as it may appear in The *SLC35A4* upstream open reading frame encodes a microprotein regulator of mitochondrial structure and function. The dissertation author was the first investigator and author of this paper. Other authors include Guy Perkins, Andrea Rocha, Joan M. Vaughan, Jiao Ma, Peter C. Gray, Jolene K. Diedrich, Mark H. Ellisman, and Alan Saghatelian.

4.7 References

1. M. A. Brunet, S. A. Levesque, D. J. Hunting, A. A. Cohen, X. Roucou, Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. *Genome Res.* **28**, 609 (2018).

2. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (80-.).* **324**, 218–23 (2009).

3. A. Rathore, T. F. Martinez, Q. Chu, A. Saghatelian, Small, but mighty? Searching for human microproteins and their potential for understanding health and disease. *Expert Rev. Proteomics*. **15**, 963 (2018).

4. S. E. Calvo, D. J. Pagliarini, V. K. Mootha, Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–7512 (2009).

5. A. Rathore, Q. Chu, D. Tan, T. F. Martinez, C. J. Donaldson, J. K. Diedrich, J. R. Yates, A. Saghatelian, MIEF1 Microprotein Regulates Mitochondrial Translation. *Biochemistry*. **57**, 5564–5575 (2018).

6. J. Chen, A. Brunner, J. Z. Cogan, J. K. Nuñez, P. Fields, B. Adamson, D. N. Itzhak, J. Y. Li, M. Mann, D. Leonetti, J. S. Weissman, HHS Public Access. **367**, 1140–1146 (2020).

7. T. F. Martinez, Q. Chu, C. Donaldson, D. Tan, M. N. Shokhirev, A. Saghatelian, Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).

8. D. E. Andreev, P. B. O'connor, C. Fahey, E. M. Kenny, I. M. Terenin, S. E. Dmitriev, P. Cormican, D. W. Morris, I. N. Shatsky, P. V. Baranov, Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife*. **2015**, e03971 (2015).

9. P. Sosicka, D. Maszczak-Seneczko, B. Bazan, Y. Shauchuk, B. Kaczmarek, M. Olczak, An insight into the orphan nucleotide sugar transporter SLC35A4. *Biochim. Biophys. Acta - Mol. Cell Res.* **1864**, 825–838 (2017).

10. C. Barbosa, I. Peixeiro, L. Romão, Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLOS Genet.* **9**, e1003529 (2013).

11. E. Balsa, M. S. Soustek, A. Thomas, S. P. Gygi, A. Enriquez, ER and Nutrient Stress Promote Assembly of Respiratory Chain Supercomplexes through the PERK-eIF2α Axis. *Mol. Cell*. **74**, 877--890.e6 (2019).

12. C. Jousse, Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR. *Nucleic Acids Res.* **29**, 4341–4351 (2001).

13. D. M. Anderson, K. M. Anderson, C. L. Chang, C. A. Makarewich, B. R. Nelson, J. R. McAnally, P. Kasaragod, J. M. Shelton, J. Liou, R. Bassel-Duby, E. N. Olson, A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell*. **160**, 595–606 (2015).

14. C. A. Makarewich, A. Z. Munir, G. G. Schiattarella, S. Bezprozvannaya, O. N. Raguimova, E. E. Cho, A. H. Vidal, S. L. Robia, R. Bassel-Duby, E. N. Olson, The DWORF micropeptide enhances contractility and prevents heart failure in a mouse model of dilated cardiomyopathy. *Elife*. **7**, e38319 (2018).

15. R. Fredriksson, K. J. V. Nordström, O. Stephansson, M. G. A. Hägglund, H. B. Schiöth, The solute carrier (SLC) complement of the human genome: Phylogenetic classification reveals four major families. *FEBS Lett.* **582**, 3811–6 (2008).

16. M. Wiktor, W. Wiertelak, D. Maszczak-Seneczko, P. J. Balwierz, B. Szulc, M. Olczak, Identification of novel potential interaction partners of UDP-galactose (SLC35A2), UDP-N-acetylglucosamine (SLC35A3) and an orphan (SLC35A4) nucleotide sugar transporters. *J. Proteomics*. **249**, 104321 (2021).

17. S. Asthana, M. Roytberg, J. Stamatoyannopoulos, S. Sunyaev, Analysis of sequence conservation at nucleotide resolution. *PLoS Comput. Biol.* **3**, e254 (2007).

18. S. Omer, T. J. Harlow, J. P. Gogarten, Does Sequence Conservation Provide Evidence for Biological Function? *Trends Microbiol.* (2017), pp. 11–18.

19. M. F. Lin, I. Jungreis, M. Kellis, PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. **27**, i275-82 (2011).

20. D. Zhang, Y. Xi, M. L. Coccimiglio, J. A. Mennigen, M. G. Jonz, M. Ekker, V. L. Trudeau, Functional prediction and physiological characterization of a novel short trans-membrane protein 1 as a subunit of mitochondrial respiratory complexes. *Physiol. Genomics*. **44**, 1133–1140 (2012).

21. H. J. Sharpe, T. J. Stevens, S. Munro, A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. *Cell*. **142**, 158 (2010).

22. J. J. Almagro Armenteros, C. K. Sønderby, S. K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*. **33**, 3387–3395 (2017).

23. R. Acin-Perez, I. Y. Benador, A. Petcherski, M. Veliova, G. A. Benavides, S. Lagarrigue, A. Caudal, L. Vergnes, A. N. Murphy, G. Karamanlidis, R. Tian, K. Reue, J. Wanagat, H. Sacks, F. Amati, V. M. Darley-Usmar, M. Liesa, A. S. Divakaruni, L. Stiles, O. S. Shirihai, A novel approach to measure mitochondrial respiration in frozen biological samples. *EMBO J.* **39**, e104073 (2020).

24. M. Picard, T. Taivassalo, D. Ritchie, K. J. Wright, M. M. Thomas, C. Romestaing, R. T.

Hepple, Mitochondrial Structure and Function Are Disrupted by Standard Isolation Methods. *PLoS One*. **6**, e18317 (2011).

25. D. F. Parsons, G. R. Williams, B. Chance, Characteristics of isolated and purified preparations of the outer and inner membranes of mitochondria. *Ann. N. Y. Acad. Sci.* **137**, 643–666 (1966).

26. H. Schneider, T. Söllner, K. Dietmeier, C. Eckerskorn, F. Lottspeich, B. Trülzsch, W. Neupert, N. Pfanner, Targeting of the master receptor MOM19 to mitochondria. *Science (80-.).* **254**, 1659–1662 (1991).

27. H. Yamamoto, M. Esaki, T. Kanamori, Y. Tamura, S. ichi Nishikawa, T. Endo, Tim50 is a subunit of the TIM23 complex that links protein translocation across the outer and inner mitochondrial membranes. *Cell*. **111**, 519–528 (2002).

28. Q. Chu, A. Rathore, J. K. Diedrich, C. J. Donaldson, J. R. Yates, A. Saghatelian, Identification of Microprotein-Protein Interactions via APEX Tagging. *Biochemistry*. **56**, 3299–3306 (2017).

29. Q. Chu, T. F. Martinez, S. W. Novak, C. J. Donaldson, D. Tan, J. M. Vaughan, T. Chang, J. K. Diedrich, L. Andrade, A. Kim, T. Zhang, U. Manor, A. Saghatelian, Regulation of the ER stress response by a mitochondrial microprotein. *Nat. Commun.* **10**, 4883 (2019).

30. N. G. D'Lima, J. Ma, L. Winkler, Q. Chu, K. H. Loh, E. O. Corpuz, B. A. Budnik, J. Lykke-Andersen, A. Saghatelian, S. A. Slavoff, A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* **13**, 174–180 (2017).

31. D. Mellacheruvu, Z. Wright, A. L. Couzens, J. P. Lambert, N. A. St-Denis, T. Li, Y. V. Miteva, S. Hauri, M. E. Sardiu, T. Y. Low, V. A. Halim, R. D. Bagshaw, N. C. Hubner, A. Al-Hakim, A. Bouchard, D. Faubert, D. Fermin, W. H. Dunham, M. Goudreault, Z. Y. Lin, B. G. Badillo, T. Pawson, D. Durocher, B. Coulombe, R. Aebersold, G. Superti-Furga, J. Colinge, A. J. R. Heck, H. Choi, M. Gstaiger, S. Mohammed, I. M. Cristea, K. L. Bennett, M. P. Washburn, B. Raught, R. M. Ewing, A. C. Gingras, A. I. Nesvizhskii, The CRAPome: a Contaminant Repository for Affinity Purification Mass Spectrometry Data. *Nat. Methods*. **10**, 730 (2013).

32. S. Rath, R. Sharma, R. Gupta, T. Ast, C. Chan, T. J. Durham, R. P. Goodman, Z. Grabarek, M. E. Haas, W. H. W. Hung, P. R. Joshi, A. A. Jourdain, S. H. Kim, A. V. Kotrys, S. S. Lam, J. G. McCoy, J. D. Meisel, M. Miranda, A. Panda, A. Patgiri, R. Rogers, S. Sadre, H. Shah, O. S. Skinner, T. L. To, M. A. Walker, H. Wang, P. S. Ward, J. Wengrod, C. C. Yuan, S. E. Calvo, V. K. Mootha, MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541 (2021).

33. C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).

34. M. Darshi, V. L. Mendiola, M. R. Mackey, A. N. Murphy, A. Koller, G. A. Perkins, M. H. Ellisman, S. S. Taylor, ChChd3, an inner mitochondrial membrane protein, is essential for maintaining Crista integrity and mitochondrial function. *J. Biol. Chem.* **286**, 2918–2932 (2011).

35. S. Koob, M. Barrera, R. Anand, A. S. Reichert, The non-glycosylated isoform of MIC26 is a constituent of the mammalian MICOS complex and promotes formation of crista junctions. *Biochim. Biophys. Acta - Mol. Cell Res.* **1853**, 1551–1563 (2015).

36. G. B. John, Y. Shang, L. Li, C. Renken, C. A. Mannella, J. M. L. Selker, L. Rangell, M. J. Bennett, J. Zha, The mitochondrial inner membrane protein mitofilin controls cristae morphology. *Mol. Biol. Cell*. **16**, 1543–1554 (2005).

37. Y. O. Kim, H. J. Koh, S. H. Kim, S. H. Jo, J. W. Huh, K. S. Jeong, I. J. Lee, B. J. Song, T. L. Huh, Identification and Functional Characterization of a Novel, Tissue-specific NAD+-dependent Isocitrate Dehydrogenase β Subunit Isoform. *J. Biol. Chem.* **274**, 36866–36875 (1999).

38. D. B. Mirel, K. Marder, J. Graziano, G. Freyer, Q. Zhao, R. Mayeux, K. C. Wilhelmsen, Characterization of the human mitochondrial aconitase gene (ACO2). *Gene*. **213**, 205–218 (1998).

39. R. Anand, A. S. Reichert, A. K. Kondadi, Emerging Roles of the MICOS Complex in Cristae Dynamics and Biogenesis. *Biology (Basel).* **10**, 600 (2021).

40. C. Glytsou, E. Calvo, S. Cogliati, A. Mehrotra, I. Anastasia, G. Rigoni, A. Raimondi, N. Shintani, M. Loureiro, J. Vazquez, L. Pellegrini, J. A. Enriquez, L. Scorrano, M. E. Soriano, Optic Atrophy 1 Is Epistatic to the Core MICOS Component MIC60 in Mitochondrial Cristae Shape Control. *Cell Rep.* **17**, 3024 (2016).

41. C. Hu, L. Shu, X. Huang, J. Yu, liuju Li, L. Gong, M. Yang, Z. Wu, Z. Gao, Y. Zhao, L. Chen, Z. Song, OPA1 and MICOS Regulate mitochondrial crista dynamics and formation. **11**, 1–17 (2020).

42. R. Anand, V. Strecker, J. Urbach, I. Wittig, A. S. Reichert, Mic13 is essential for formation of crista junctions in mammalian cells. *PLoS One*. **11**, e0160258 (2016).

43. H. Li, Y. Ruan, K. Zhang, F. Jian, C. Hu, L. Miao, L. Gong, L. Sun, X. Zhang, S. Chen, H. Chen, D. Liu, Z. Song, Mic60/Mitofilin determines MICOS assembly essential for mitochondrial dynamics and mtDNA nucleoid organization. *Cell Death Differ. 2016 233*. **23**, 380–392 (2015).

44. T. A. Weber, S. Koob, H. Heide, I. Wittig, B. Head, A. van der Bliek, U. Brandt, M. Mittelbronn, A. S. Reichert, APOOL Is a Cardiolipin-Binding Constituent of the Mitofilin/MINOS Protein Complex Determining Cristae Morphology in Mammalian Mitochondria. *PLoS One*. **8**, 63683 (2013).

45. V. Guarani, E. M. McNeill, J. A. Paulo, E. L. Huttlin, F. Fröhlich, S. P. Gygi, D. Van Vactor, J. Wade Harper, QIL1 is a novel mitochondrial protein required for MICOS complex stability and cristae morphology. *Elife*. **4**, 1–23 (2015).

46. H. Rampelt, R. M. Zerbes, M. van der Laan, N. Pfanner, Role of the mitochondrial contact site and cristae organizing system in membrane architecture and dynamics. *Biochim. Biophys. Acta - Mol. Cell Res.* (2017), pp. 737–746.

47. J. An, J. Shi, Q. He, K. Lui, Y. Liu, Y. Huang, M. S. Sheikh, CHCM1/CHCHD6, novel mitochondrial protein linked to regulation of mitofilin and mitochondrial cristae morphology. *J. Biol. Chem.* **287**, 7411–7426 (2012).

48. A. K. Kondadi, R. Anand, S. Hänsch, J. Urbach, T. Zobel, D. M. Wolf, M. Segawa, M. Liesa, O. S. Shirihai, S. Weidtkamp-Peters, A. S. Reichert, S. Weidtkamp-Peters, A. S. Reichert, Cristae undergo continuous cycles of membrane remodelling in a MICOS-dependent manner. *EMBO Rep.* **21**, e49776 (2020).

49. Y. Alhindi, L. M. Vaanholt, M. Al-Tarrah, S. R. Gray, J. R. Speakman, C. Hambly, B. S. Alanazi, B. M. Gabriel, A. Lionikas, A. Ratkevicius, Low Citrate Synthase Activity Is Associated with Glucose Intolerance and Lipotoxicity. *J. Nutr. Metab.* **2019**, 8594825 (2019).

50. M. R. M. R. Wieckowski, C. Giorgi, M. Lebiedzinska, J. Duszynski, P. Pinton, Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells. *Nat. Protoc. 2009 411*. **4**, 1582–1590 (2009).

51. L. Aerts, K. Craessaerts, B. De Strooper, V. A. Morais, PINK1 Kinase Catalytic Activity Is Regulated by Phosphorylation on Serines 228 and 402. *J. Biol. Chem.* **290**, 2798 (2015).

52. X. Gu, Y. Ma, Y. Liu, Q. Wan, Measurement of mitochondrial respiration in adherent cells by Seahorse XF96 Cell Mito Stress Test. *STAR Protoc.* **2**, 100245 (2021).

53. B. Y. Mok, M. H. de Moraes, J. Zeng, D. E. Bosch, A. V. Kotrys, A. Raguram, F. S. Hsu, M. C. Radey, S. B. Peterson, V. K. Mootha, J. D. Mougous, D. R. Liu, A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nat. 2020 5837817*. **583**, 631–637 (2020).

54. T. L. Marin, B. Gongol, F. Zhang, M. Martin, D. A. Johnson, H. Xiao, Y. Wang, S. Subramaniam, S. Chien, J. Y. J. Shyy, AMPK promotes mitochondrial biogenesis and function by phosphorylating the epigenetic factors DNMT1, RBBP7, and HAT1. *Sci. Signal.* **10**, eaaf7478 (2017).

55. J. Ma, J. K. Diedrich, I. Jungreis, C. Donaldson, J. Vaughan, M. Kellis, J. R. Yates, A. Saghatelian, Improved Identification and Analysis of Small Open Reading Frame Encoded Polypeptides. *Anal. Chem.* **88**, 3967 (2016).

CHAPTER 5

APPENDIX

5.1 Conclusion and Future Directions

Our work in functional elucidation of FAHFAs and microproteins expands the current knowledge of the lipidome and micropeptidome capacity. However, there is still a lot to be learned about these functional natural molecules and their translatable activities in the context of human diseases. More mechanism of action studies involving FAHFAs are underway, but it is currently understood that some of the biological activities of 9-PAHSA function through activation of G protein-coupled receptors (GPCRs), such as GPR40 (*1*) and GPR120 (*2*). Though FAHFA's mechanism of action is not fully elucidated, it remains plausible that FAHFAs are transported into the cells via the classical endogenous lipoprotein pathway. Endogenous lipids are often cargos of lipoproteins, such as VLDL, in circulation. Lipoproteins play critical roles in shuttling lipids like triglycerides to tissues requiring energetic demands for further degradation (*3*). And because FAHFAs share similar chemical and physical properties with triglycerides, their biological functions may act on the same transport mechanism. It was previously demonstrated that FAHFAs can be incorporated to form FAHFA-TGs (*4*), yet it is still unclear whether they are essential for phospholipid or sphingolipid anabolism. To date, the challenge of characterizing additional bioactive FAHFAs is largely due to their susceptibility to endogenous metabolism. We predict that synthetic FAHFAs, however, will provide a new utility and avenue for exploring these open-ended questions and benefit future functional studies involving FAHFAs as some are demonstrated to be more stable with enhanced activities. The demonstrated activities of synthetic FAHFAs in our studies have potentiated these molecules as novel non-steroidal reducers of inflammation. Further assessment of their efficacy and safety in anti-inflammatory indications would be pertinent clinically.

Similarly, our choice of SLC35A4-MP for characterization is dependent on the current limited understanding of uORF-encoded microprotein functions, despite their high abundance in the human genome(*5*). uORFs are predominantly known for its cis-regulating activities in translational control (*6–8*). Function elucidation of SLC35A4-MP, however, underscores the diversity of uORF functions. Genetic ablation of SLC35A4-MP demonstrates that the microprotein is required for optimal formation of CJ ultrastructure, presumably through stabilization of its *bona fide* interactor MICOS complex. We found that removal of SLC35A4-MP in SLC35A4-MP KO HEK293T cells did not significantly influence the protein expressions of other known MICOS proteins. This is contrary to the phenotypes generated in MIC10 or MIC60 KO HAP1 cells (*9*), suggesting that SLC35A4-MP may not be a scaffolding subunit of MICOS. Moving forward, topological studies shall be a priority to further assess the effects of SLC35A4-MP on MICOS complex assembly and coordination to mediate cristae or CJ morphologically. In addition, the transmembrane domain of SLC35A4-MP may also contribute to its localization within the MICOS complex. Truncation of the amino acid residues within its transmembrane domain may also help interrogate the importance of its transmembrane domain on complex assembly.

Similarly, while MICOS complex and SLC35A4-MP have crucial roles, the making of cristae is a highly complex mechanism that is not well-understood. Previously, it was suggested that cristae formation is dependent on the mitochondrial activity of the tissue. Tissues requiring high energetic demands, such as skeletal muscle and neurons, exhibit lamellar-shaped cristae, while tubular cristae are representative of steroidogenic tissues (*10*). This may also indicate that different tissues require different sets of molecules to construct the cristae as their mitochondrial activities vary. Future examination of SLC35A4-MP's physiological consequences using disease

models could also reinforce our understanding of SLC35A4-MP's role on physiology and human health. Though SLC35A4-MP's endogenous expression is abundant in mammalian cells and most C57BL/6 mouse tissues, its levels across peripheral tissues in mice varies, with the liver having a very low expression. A possible explanation for this observation is that SLC35A4-MP's activity may be tissue-specific. Overall, we found that the oxidative capacity of SLC35A4-MP KO mitochondria is reduced. Aerobic exercise, such as long distance training output, may be used as a function to measure the impact of SLC35A4-MP on mammalian physiology, thus fitness.

5.2 References

1. I. Syed, J. Lee, P. M. Moraes-Vieira, C. J. Donaldson, A. Sontheimer, P. Aryal, K. Wellenstein, M. J. Kolar, A. T. Nelson, D. Siegel, J. Mokrosinski, I. S. Farooqi, J. J. Zhao, M. M. Yore, O. D. Peroni, A. Saghatelian, B. B. Kahn, Palmitic Acid Hydroxy Stearic Acids Activate GPR40 Which is Involved in Their Beneficial Effects on Glucose Homeostasis. *Cell Metab.* **27**, 419 (2018).

2. M. M. Yore, I. Syed, P. M. Moraes-Vieira, T. Zhang, M. A. Herman, E. A. Homan, R. T. Patel, J. Lee, S. Chen, O. D. Peroni, A. S. Dhaneshwar, A. Hammarstedt, U. Smith, T. E. McGraw, A. Saghatelian, B. B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects. *Cell*. **159**, 318–332 (2014).

3. K. R. Feingold, C. Grunfeld, Introduction to Lipids and Lipoproteins. *Endotext* (2021) (available at https://www.ncbi.nlm.nih.gov/books/NBK305896/).

4. D. Tan, M. E. Ertunc, S. Konduri, J. Zhang, A. M. Pinto, Q. Chu, B. B. Kahn, D. Siegel, A. Saghatelian, Discovery of FAHFA-Containing Triacylglycerols and Their Metabolic Regulation. *J. Am. Chem. Soc.* **141**, 8798–8806 (2019).

5. S. E. Calvo, D. J. Pagliarini, V. K. Mootha, Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7507–7512 (2009).

6. T. Zhang, A. Wu, Y. Yue, Y. Zhao, uORFs: Important Cis-Regulatory Elements in Plants. *Int. J. Mol. Sci.* **21**, 1–14 (2020).

7. C. Barbosa, I. Peixeiro, L. Romão, Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PLOS Genet.* **9**, e1003529 (2013).

8. Y. Chu, J. Huang, G. Ma, T. Cui, X. Yan, H. Li, N. Wang, An Upstream Open Reading Frame Represses Translation of Chicken PPARγ Transcript Variant 1. *Front. Genet.* **11**, 165 (2020).

9. A. K. Kondadi, R. Anand, S. Hänsch, J. Urbach, T. Zobel, D. M. Wolf, M. Segawa, M. Liesa, O. S. Shirihai, S. Weidtkamp-Peters, A. S. Reichert, S. Weidtkamp-Peters, A. S. Reichert, Cristae undergo continuous cycles of membrane remodelling in a MICOS-dependent manner. *EMBO Rep.* **21**, e49776 (2020).

10. M. Harner, C. Körner, D. Walther, D. Mokranjac, J. Kaesmacher, U. Welsch, J. Griffith, M. Mann, F. Reggiori, W. Neupert, The mitochondrial contact site complex, a determinant of mitochondrial architecture. **30**, 4356–4370 (2011).