

**SYSTEMATIC INVESTIGATION OF PROTEIN-METABOLITE
REGULATORY INTERACTIONS: METHODOLOGIES AND
CONTEXT**

A Thesis
Presented to
The Academic Faculty

By

McKenzie L. Smith

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Chemical & Biomolecular Engineering

Georgia Institute of Technology

December 2016

Copyright © McKenzie L. Smith 2016

SYSTEMATIC INVESTIGATION OF PROTEIN-METABOLITE REGULATORY INTERACTIONS: METHODOLOGIES AND CONTEXT

Approved by:

Dr. Mark P Styczynski, Advisor
School of Chemical & Biomedical
Engineering
Georgia Institute of Technology

Dr. Andreas S. Bommarius
School of Chemical & Biomedical
Engineering
Georgia Institute of Technology

Dr. Julie A. Champion
School of Chemical & Biomedical
Engineering
Georgia Institute of Technology

Dr. Facundo M. Fernandez
School of Chemistry
Georgia Institute of Technology

Dr. Wendy L. Kelly
School of Chemistry
Georgia Institute of Technology

Date Approved: November 1, 2016

ACKNOWLEDGEMENTS

There are many people I would like to thank for their support during my time at Georgia Institute of Technology.

Dr. Mark Styczynski for your support and mentorship since I entered the program.

My committee members, Dr. Andreas Bommarius, Dr. Julie Champion, Dr. Facundo Fernandez, and Dr. Wendy Kelly for your suggestions and assistance on my research over the last six years.

The Styczynski research group, past and present, for your companionship and frequent presence as a sounding-board, but especially Rob Dromms, Amy Su, Katie Vermeersch, and Daniel Watstein, for altogether more things than I can name.

My other friends from the program here at Georgia Tech and in the Atlanta area for your support, great science discussions, and not-so-science discussions through the years (and for making sure I left the lab and my apartment occasionally), especially Chris Butch and Andrew Sevrinsky.

My friends from farther afield who contributed just as much to helping me through the last several years, especially Matt Gen, Jennifer Howerter, John and Lindsey Morris, Betsy Muszynski, Chelsea Ray, and Austin Yoder.

Finally, my wonderful parents and brother for your love and consistent support throughout the years.

TABLE OF CONTENTS

Acknowledgements	iii
List of Tables	vii
List of Figures.....	viii
Abbreviations	ix
Summary.....	xii
Chapter 1 Introduction: Metabolic regulation is a systems-level phenomenon and a persistent methodological challenge.....	1
1.1 Systems biology and modeling	1
1.1.1 Metabolomics as a systems biology starting point	2
1.2 Allosteric regulation.....	3
1.2.1 Physical mechanism.....	3
1.2.2 Discovery of new regulatory interactions	5
1.3 <i>Saccharomyces cerevisiae</i>	7
1.4 Thesis overview	8
1.4.1 High-level insights from untargeted metabolomics	8
1.4.2 A pipeline framework for medium-throughput investigation of enzyme regulation via direct reaction assay	9
1.4.3 Challenges and potential of small-molecule microarrays for interaction discovery.....	10
1.5 References.....	11
Chapter 2 Untargeted metabolomics yields insights into high-level metabolic differences in two Atlantic salmon cohorts.....	15
2.1 Introduction.....	15
2.2 Results and discussion	18
2.2.1 Overall data set and reproducibility	18
2.2.2 Multivariate analysis of the metabolite profiles of the fasting and feeding fish	20
2.2.3 Significantly different analytes between fasting and feeding Atlantic salmon serum	24
2.2.4 Significantly accumulated analytes in inappetent Atlantic salmon serum.....	26
2.2.5 Spawning Atlantic salmon and whale sharks have different metabolic responses to inappetence	26
2.2.6 Fatty acid metabolism response to inappetence appears to be dependent on fatty acid structure	28
2.2.7 Polyol accumulation might be indicative of a currently unknown response to starvation	32
2.2.8 Amino acids supply additional energy sources to fasting.....	32

2.2.9	Limitations of this study	35
2.3	Conclusions and context	36
2.4	Materials and methods	38
2.4.1	Fish and experimental conditions	38
2.4.2	Sampling procedure and sample preparation	39
2.4.3	GCxGC-MS analysis	40
2.4.4	Data analysis	41
2.5	References	43
 Chapter 3 An <i>in vitro</i> reaction assay pipeline provides a framework for validation of putative regulatory interactions and a necessary increase in sensitivity range for <i>de novo</i> interaction discovery.....47		
3.1	Introduction	47
3.2	Results and discussion	50
3.2.1	Protein expression conditions	50
3.2.2	Reaction assay architecture	52
3.2.3	Screening putative regulator libraries for inhibitory/activating activity	55
3.2.4	GC-MS method optimization for screening individual enzymes	61
3.3	Conclusions	62
3.4	Materials and methods	62
3.4.1	Chemicals and media	62
3.4.2	Yeast culture and protein expression	62
3.4.3	Expression assays	63
3.4.4	Cell lysis and protein purification	64
3.4.5	Reaction assays	65
3.4.6	Deproteinization and GC-MS sample preparation	66
3.4.7	GC-MS	67
3.4.8	Reaction assay data analysis	67
3.5	References	68
 Chapter 4 Small-molecule microarrays (SMMs) present a potential platform for high-throughput binding interaction discovery, albeit with unique challenges.....71		
4.1	Introduction	71
4.2	Results and discussion	74
4.2.1	Initial validation and optimization of basic protocol	74
4.2.2	Sensitivity limitations	77
4.2.3	Photocrosslinking strategy for increasing primary hybridization sensitivity	77
4.2.4	Destruction of binding functionality in small printed metabolites	80
4.3	Conclusions	81
4.4	Materials and methods	81
4.4.1	Small-molecule microarray synthesis	81
4.4.2	Array hybridization and analysis	83
4.4.3	Addition and characterization of photocrosslinking step	83
4.4.4	Regulatory assay of chorismate mutase using <i>N</i> -acetylated regulators	83
4.5	References	84

Chapter 5	Conclusions and recommendations for further study	85
5.1	Conclusions	85
5.1.1	High-level insights from untargeted metabolomics	85
5.1.2	A pipeline framework for medium-throughput investigation of enzyme regulation via direct reaction assay	86
5.1.3	Challenges and potential of small-molecule microarrays for interaction discovery	87
5.2	Relevance	88
5.3	Recommendations for future study	91
5.3.1	Small-molecule microarray assay development	91
5.3.1.1	Additional reaction assays with substituted regulators	91
5.3.1.2	Validation of photocrosslinking strategy for sensitivity improvement	92
5.3.2	Integration of untargeted yeast metabolomics and regulator screening	92
5.3.2.1	Exploratory yeast metabolomics analysis	93
5.3.2.2	Guided regulator search via iterative library screening	93
5.3.3	Downstream mathematical modeling and bioinformatics analysis informed by regulator discovery	94
5.4	References	95
Appendix A:	Detailed GCxGC-MS methods	96
Appendix B:	Metabolite libraries	98
Appendix C:	Letter of permission for use of copyrighted material	100

LIST OF TABLES

Table 2.1: Significantly different metabolites between domesticated and sea-run fish.....	25
Table 2.2: Significantly enriched pathways from MetaboAnalyst pathway analysis.....	34
Table 3.1: Summary of enzymes and reactions.....	51
Table 3.2: Summary of enzyme behavior over differing reaction assay architecture, illustrating differences across enzymes.....	54
Table A.1: Modulation timing.....	97
Table A.2: Metabolite library lists.....	98

LIST OF FIGURES

Figure 2.1: Heatmap of all normalized metabolite levels.....	19
Figure 2.2: Principal components analysis shows strong separation between domesticated and sea-run fish.....	21
Figure 2.3: PCA visualization of batch effects.....	22
Figure 2.4: Unsupervised dimensional reduction yields separation in just two principal components.....	23
Figure 2.5: Partial least squares discriminant analysis shows strong separation between domesticated and sea-run fish.....	24
Figure 2.6: Fatty acids exhibit structure-specific differences between domesticated and sea-run fish.....	30
Figure 3.1: Relative quantification of expression of three enzymes under different expression protocols.....	51
Figure 3.2: Iterative regulator screening demonstrated in ARO7, narrowing down from a metabolite library to the known regulators.....	58
Figure 3.3: Extended library size comparison for ARO7 reactivity in lysate.....	60
Figure 4.1: Metabolite attachment chemistry.....	73
Figure 4.2: Simplified diagram of a double-hybridized SMM.....	73
Figure 4.3: Progressive validation of basic arraying and hybridization protocols...	75
Figure 4.4: Spot morphology comparison, showing FKBP12 bound to printed rapamycin.....	76
Figure 4.5: LC-SDA heterobifunctional crosslinker.....	78
Figure 4.6: Detectable non-specific capture of FKBP12 with 1 mM printed LC-SDA crosslinker when treated with UV (boxed slide segments on bottom row).....	79

ABBREVIATIONS

AAT2: aspartate aminotransferase

AMDIS: automated mass spectral deconvolution & identification system

ARO7: chorismate mutase

BCAA: branched-chain amino acid

BSA: bovine serum albumin

CoA: coenzyme A

DIPEA: *N,N*-diisopropylethylamine

DMF: *N,N*-dimethylformamide

DMSO: dimethylsulfoxide

DNA: deoxyribonucleic acid

EDTA: ethylenediaminetetraacetic acid

ESI: electrospray ionization

FAME: fatty acid methyl ester

FCY1: cytosine deaminase

FDR: false discovery rate

FK506: tacrolimus

FKBP12: FK506 binding protein 12

GAL: galactose

GC: gas chromatography

GFP: green fluorescent protein

His: histidine

HMDB: human metabolite database

IgG: immunoglobulin G

IMAC: immobilized metal ion affinity chromatography

K_D: dissociation constant

KNN: k-nearest neighbors

LC: liquid chromatography

LC-SDA: succinimidyl 6-(4,4'-azipentanamido)hexanoate

MALDI: matrix-assisted laser ionization desorption

MORF: moveable-open reading frame

MPEA: metabolite pathway enrichment analysis

MS: mass spectrometry

MSTFA: *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide

NHS: *N*-hydroxysuccinimide

NMR: nuclear magnetic resonance spectroscopy

OD: optical density

Oligo: oligonucleotide

ORF: open reading frame

PBS: phosphate-buffered saline

PBST: phosphate-buffered saline with Tween 20

PCA: principle components analysis

PEG: polyethylene glycol

PLS-DA: partial least squares discriminant analysis

PyBOP: (benzotriazol-1-yl-oxy)tripyrrolidinophosphonium hexafluorophosphate

RNA: ribonucleic acid

SC-ura: synthetic complete minus uracil

S. cerevisiae: *Saccharomyces cerevisiae*

SMM: small-molecule microarray

S/N: signal to noise ratio

SPROX: stability of proteins from rates of oxidation

S. salar: *Salmo salar*

TCA: tricarboxylic acid

THF: tetrahydrofuran

TMCS: trimethylchlorosilane

TOF: time-of-flight

tRNA: transfer ribonucleic acid

USFWS: U.S. fish and wildlife service

UV: ultraviolet

SUMMARY

Development of a systems-level understanding of metabolism is expected have far-reaching benefits from medicine to ecology to industry, as it will facilitate the comprehensive profiling and prediction of metabolic states in organisms of interest. Systematic characterization strategies have thus far been successfully applied at the genomic, transcriptomic, and proteomic levels, but downstream regulatory interactions have remained comparatively underexplored. We believe this is largely due to the wide sensitivity spectrum required to effect a similarly comprehensive study: the binding affinities of known protein-metabolite regulatory pairs span multiple orders of magnitude. The overarching aim of this work was therefore to develop and refine multiple complementary methodologies for discovery and characterization of these interactions, and to explore the potential mutual relevance such a strategy has with higher-level metabolomics studies.

An untargeted metabolomics study of inappetence in spawning salmon compared metabolic phenotypes between sea-run Alaskan salmon which do not eat for months at a time during the spawning season, and a domesticated cohort which were fed throughout their spawning period (Chapter 2). Two dimensional gas chromatography coupled to mass spectrometry was used to determine the metabolite profiles of serum samples. This analysis yielded insights into the high-level metabolic differences between inappetent and fed salmon. Structural class-specific changes were identified in metabolite levels in inappetent fish that suggest a potentially unique utilization of metabolites during spawning-induced inappetence. Also, in stark contrast to previous studies of inappetence in whale

sharks, the inappetent salmon's metabolite profiles were not indicative of a classical catabolic starvation phenotype. These results will provide direction and context for future protein- and pathway- level studies of salmonid metabolism, and the methodology can and should be applied to other organisms to inform systems-level analysis of metabolism, as is discussed in Chapter 5.

In order to directly pursue the goal of enabling broad-scale discovery and characterization of protein-metabolite regulatory interactions, an *in vitro* reaction assay-based pipeline was developed to provide a flexible framework for both the validation of putative regulatory interactions found via other methods, and the discovery of new regulatory interactions over a wide range of binding affinities (Chapter 3). Guidelines were developed not only for implementation of the reaction assay strategy itself, but also for the practical and logistical necessities of protein handling when many proteins are to be investigated in parallel. For *de novo* discovery of regulatory interactions, an iterative library screening strategy was developed and investigated using yeast chorismate mutase, which has two well-known allosteric regulators, as a proof-of-concept.

Small-molecule microarrays, originally developed for use in drug discovery, were explored as a potential platform for high-throughput discovery of endogenous binding interactions (Chapter 4). In the process of successfully reproducing the method validation results of previous work, we identified guidelines for reducing background in order to meet the needs of our intended use of the assay to screen for relatively weak interactions. Furthermore, a photocrosslinking strategy was proposed and preliminary investigations conducted in an

effort to increase the sensitivity of the assay and thereby expand the range of endogenous interactions that should be detectable via small-molecule microarray. We identified guidelines for further development of this strategy and future effective implementation of small-molecule microarrays as a discovery platform for endogenous binding interactions.

Taken together, the findings of this work provide a methodological framework to increase the efficiency and range of study for protein-metabolite regulatory interactions, as well as lay groundwork for further expansion of that range, ultimately in service of the future development of systems-level metabolic models. Recommendations for future development and pursuit of these investigative strategies are discussed in Chapter 5.

Chapter 1: Introduction: Metabolic regulation is a systems-level phenomenon and a persistent methodological challenge

1.1 Systems biology and modeling

Systems biology refers to the effort to understand the network behavior of biological systems, often involving mathematical modeling informed by experiment¹, and metabolism refers to all of the chemical reactions taking place inside a cell. In cells, “network behavior” encompasses comprehensive information about the genome, transcriptome, proteome, or metabolome, potentially including not only how these levels interact within themselves, but also with each other. On every level of metabolism, activity is regulated by transient ligand-binding events that present a much more finely-tuned and complex picture than simply looking at genes or proteins or metabolic pathways²⁻⁶).

Current state-of-the-art mathematical modeling methods, not discussed in detail here, have proven useful in optimizing metabolic pathways for a given objective parameter^{7,8}. One of the most common methods, flux balance analysis, is based on stoichiometric ratios and requires a quasi-steady-state assumption. Transcriptomic, proteomic, and metabolomic data have been used in such models to assist in predicting fluxes⁹, but the approach is paradigmatically the same. However, downstream metabolic regulation is transient and dynamic in nature, and is not adequately captured by these types of models (indeed, this is generally not attempted).

True dynamic modeling of metabolic behavior must be rooted in a mechanistic understanding of component interactions as well as an understanding of the basic pathway

architecture. This will involve incorporating multiple timescale of regulatory information and relationships, from the transient (e.g. allostery) to the longer-term (transcriptional regulation). The amount of data required to construct such a model is enormous. Systematic, high-throughput methods of characterizing component interactions have already been undertaken at various levels of metabolism (e.g. genomic, transcriptomic), but allosteric regulation is an area that has been comparatively underexplored to date². A similar, systematic analysis of metabolite-level regulation would provide a transformative contribution to systems biology because it would allow for a level and precision of dynamic modeling that has previously been unattainable.

1.1.1 Metabolomics as a systems biology starting point

Metabolomics is a high-level approach to studying metabolism focusing on comprehensive profiling of the metabolites in a biological sample, which results in a complex, phenotypic “fingerprint.” Such profiling has been previously used to great effect in the medical field for biomarker discovery, yielding diagnostic screens that may identify a disease state early, or differentiate between subtypes of a given disease, as in cancer¹⁰.

Differences in metabolic fingerprint between cohorts can point to areas of metabolism that will yield mechanistic insight with deeper study, but the fingerprint itself can only provide a suggestive snapshot of metabolism. Metabolomics methods are particularly useful in their ability to give quick access to some information about the metabolism of organisms that have not otherwise been widely studied, which may in turn present a starting point for more pointed experiments.

1.2 Allosteric regulation

1.2.1 *Physical mechanism*

Allostery refers to a regulatory event wherein an effector molecule binds to a protein in a location other than its active site and affects the activity of the protein. This effector is not a steric analogue of the substrate (hence, “allosteric”), and can influence the protein’s activity positively (in the case of an activator) or negatively (a repressor)^{11,12}. Early studies of allostery primarily defined and discussed it as a phenomenon exclusive to oligomeric proteins with identical subunits^{12,13}, and this understanding was later expanded and generalized to include monomeric proteins^{14,15}. Allosteric binding sites can be located at the interface of an oligomer’s subunits (its conformational coupling termed as “direct”) or fully-contained within the structure of a monomer or monomeric subunit (“indirect coupling”)¹⁵, often exhibiting cooperative binding in oligomeric proteins¹⁶⁻¹⁸. The term “allosteric” has also been used to describe conformational changes involved in protein-mediated active transport within a cell, wherein the allosteric effector is also considered the substrate¹⁹. This differs from the original definition in its emphasis on conformational change rather than on the distinction between substrate and effector as the most salient characteristic of the phenomenon. For the purposes of this work, the original definition will be used, with conformational change being considered a mechanism for the regulatory phenomenon rather than its defining characteristic.

Many major proteins, including transcription factors¹⁴, carrier proteins¹⁷, and enzymes²⁰⁻²², are explicitly known to be allosterically regulated. The broad spread of proteins with

known allosteric interactions, in itself, suggests the importance of this phenomenon's role in metabolism. Beyond the question of prevalence, allosteric interactions are fundamentally different from other forms of regulation – particularly transcriptional regulation, which is much more widely-studied. Allostery operates on a comparatively short timescale, allowing dynamic fine-tuning of metabolism as opposed to transcriptional regulation's often larger perturbations and significant lag time. It is also particularly notable in that it allows direct pathway cross-talk, as opposed to, for example, an indirect upstream linkage through shared substrates. The phenomenon's apparent prevalence² and unique function both support the idea that allosteric interactions are a part of metabolism worthy of discovery-based characterization efforts.

Perhaps even more intriguingly, energetic modeling of allosteric interactions¹⁶ has led to the prediction that significant changes to the magnitude or nature (activation or repression) of allosteric interactions could be brought about by relatively small alterations to the coupling enthalpy or entropy²³. This prediction has been borne out by studies in which allosteric interaction affinities were precisely tuned¹⁵ or even created *de novo*²⁴ by minor mutations. This mutability of allosteric interactions speaks to the phenomenon's importance from an engineering perspective. Broad characterization of these interactions – developing an understanding of the “allosterome” of entire organisms – may yield important targets for regulatory tuning in engineered pathways.

Furthermore, the point has been made that in the early days of protein research when effort to discover a new protein was on par with the effort required to characterize that protein's

regulatory behavior, most known proteins were known to exhibit allostery; in the past few decades, ease of protein discovery has skyrocketed while ease of allosteric characterization has not, and the rate of increase in known allosteric interactions is comparatively slow².

The likely-widespread importance and relative lack of research coverage of allostery in metabolism makes the phenomenon an attractive and potentially high-impact area of study for systems biology.

1.2.2 Discovery of new regulatory interactions

Characterization of allostery is, thanks to the generally weak (but exceptionally diverse in affinities), transient nature of the interactions involved², a nontrivially difficult task to do in a systematic manner. Historically, allosteric interactions have been discovered and characterized during intense study of a single protein or protein family^{20,22,25}, with specialized assays designed around the exact system being studied. This is untenable as a systems-scale analysis or discovery solution for obvious reasons.

Structure-based mathematical models have been created to describe allosteric interactions and attempt to predict them^{26,27}, but these models have not thus far been successfully implemented as tools of discovery. Machine learning has also been proposed as a technique to predict “allosteric hot spots”²⁸; the accuracy of the results did not, however, indicate that the approach as implemented would suffice as a reliable method of interaction discovery.

Several experimental strategies have been implemented in recent years to identify protein-metabolite binding interactions. A limited *in vivo* analysis of kinase regulation in yeast has shown promise as a methodology and identified ergosterol as a promiscuous binder and potentially important regulator⁵. Similarly, different mass spectrometry-based *in vitro* approaches have been used to study lipid-protein interactions^{3,4} and small metabolite-protein interactions⁶. The latter method used equilibrium dialysis to partition metabolite solution into different chambers containing or lacking a protein of interest and measured the enrichment of the metabolite in the protein-containing partition. This technique currently requires a large amount of protein (on the order of 1 mg per assay, approximately the total yield from a liter of culture with a “high expression” yeast ORF from the library that was used in this research²⁹), limiting its potential for extremely high-throughput usage. The former two methods made use of affinity interactions to immobilize a protein of interest on a solid matrix, then either eluted off metabolites bound *in vivo*⁵ or in an *in vitro* incubation step^{3,4,6}. The primary restriction that these methods have in common is that the protocols were optimized for detection of very specific classes of metabolites. The *in vivo* approach also, at present, requires an extremely expensive detection platform due to the small individual concentrations of the analytes.

Another family of strategies is based on the thermodynamic stabilizing effect on proteins of metabolite binding, tracking protein destabilization in lysate or with a chemical denaturant, such as in Stability of Proteins from Rates of Oxidation (SPROX)^{30,31}. In these assays, proteins are incubated with known metabolites and their stability is tracked relative to protein samples without ligand.

Other methods, such as small molecule-affinity chromatography combined with stable isotope labeling of amino acids in cell culture³², have used metabolites as the capture reagents, resulting in a search that runs in the opposite direction from the strategies described above (i.e., beginning with a metabolite that is desirable to investigate, rather than a protein). These methods have increasing popularity, but the disadvantage of requiring that the metabolites be modified. Also under the “metabolite-first” umbrella are chemoproteomic approaches, which allow detection of interactions in live cells by modifying the probe metabolite to be able to covalently modify its protein target³³.

The diverse array of experimental strategies for interaction discovery reflects the large degree of variability in both structural and practical specifics of these interactions.

1.3 *Saccharomyces cerevisiae*

The yeast *S. cerevisiae* was chosen as the primary organism of study for the direct protein investigations in this work because it is commonly used both as a model eukaryotic organism and a metabolic engineering workhorse. Its relative ease of culture, fast doubling time, and possession of eukaryotic chaperone proteins make it an attractive early-stage analogue for studying eukaryotic metabolism in general, and its approximately 6000-gene genome is comparatively small and simple versus that of higher eukaryotes. From an engineering perspective, these traits make *S. cerevisiae* a convenient host for many metabolic engineering efforts^{34,35}, and from a biological perspective, perhaps the easiest eukaryotic model to work with. Furthermore, much of the currently-existing genomic and

transcriptomic data are readily available in aggregate in internet databases^{36,37}. *S. cerevisiae* is therefore an attractive choice for development of methods for characterization and discovery of protein-metabolite interactions, as the information gained will be applicable both in theory and in direct practice.

1.4 Thesis overview

Systematic collection, assembly, and integration of protein-metabolite regulatory data would provide a much-needed level of detail in pursuit of effective, dynamic modeling of metabolism - a major goal across multiple subdisciplines. Many strategies have been developed and implemented over the past ten years, with varying degrees of ambition regarding throughput and scope, but no single methodology or pipeline has thus far provided a true framework for comprehensive study.

The purpose of this work was originally intended to be the development of such a single, comprehensive pipeline for characterization of regulatory interactions. However, it has become increasingly clear both in the course of this work and through the work of other researchers within the last handful of years that for the present, a mosaic of methodologies is necessary in order to account for the sheer variety represented in such interactions. This thesis, then, attempts an integrative view of studying regulation at the metabolite level, which takes into account that variety and addresses differing practical concerns that may emerge from it.

1.4.1 High-level insights from untargeted metabolomics

As described above, high-level metabolic differences between cohorts of a single organism can be suggestive of areas of interest for systems-level study. They indicate a metabolic response to perturbation, a significant part of which may take place at the protein-metabolite interaction level. The results of our early metabolomics study of Atlantic salmon were not further pursued in such a manner in this work, but nevertheless provide important context for a deeper study of protein-metabolite level metabolic regulation.

Wild Atlantic salmon exhibit inappetence during spawning. Studies with other organisms that display similar behavior have indicated a pathological metabolic state; this study was intended to see if Atlantic salmon had a similar metabolic response to inappetence, and to simply identify high-level changes in metabolism that correspond with a prolonged period of inappetence.

Control fish were hatchery-raised, and fed throughout their spawning period, and compared to sea-run salmon which undergo a period of inappetence as long as half a year. Serum samples were taken from each cohort and analyzed using two-dimensional gas chromatography coupled to time-of-flight mass spectrometry (GCxGC-TOFMS), yielding metabolite profiles which were then further investigated using statistical methods to identify the metabolite differences that represented a significant source of variation between the two cohorts.

1.4.2 A pipeline framework for medium-throughput investigation of enzyme regulation via direct reaction assay

Sensitivity restrictions and other practical limitations of screens for protein-metabolite binding interactions present a gap in currently-implemented systematic strategies for regulator discovery. Furthermore, all screening methods that identify binding interactions or putative binding interactions require downstream validation and characterization via reaction assay, and this step is too-often an afterthought in current literature. We developed an expression-to-characterization reaction assay pipeline to facilitate validation of potential regulators in future studies, and to simultaneously address the strategic gap in *de novo* discovery.

Three *S. cerevisiae* proteins of diverse reaction chemistry were chosen from a commercially-available yeast protein library for development case studies. Induction behavior was characterized over and beyond typically-recommended ranges of induction density. Then, the enzymes' relative reactivity was assayed over three different reaction assay architectures. Finally, a set of reaction assays using chorismate mutase and its two known allosteric regulators was used to provide proof-of-principle for our proposed library-based screening strategy for *de novo* regulator discovery.

1.4.3 Challenges and potential of small-molecule microarrays for interaction discovery

Development of an effective high-throughput screening strategy for protein-metabolite binding interactions is a highly desirable goal in the study of metabolic regulation. We explored the use of small-molecule microarrays, previously developed with an eye toward high-throughput drug discovery, as a potential platform for such a high-throughput screen for endogenous interactions.

Small-molecule microarrays were synthesized from amino-silane functionalized glass slides using isocyanate-based covalent attachment chemistry, producing an array of spotted metabolites tethered to the solid substrate via a short polyethylene glycol linker. Protocols for probing with a protein of interest were tested and adjusted with an eye to the challenges of screening for endogenous interactions (namely, since most allosteric interactions have binding strengths in a range that makes it particularly necessary to minimize fluorescence background on the hybridized slides). Finally, a novel strategy of photo-inducible crosslinking was explored as a potential means of increasing the sensitivity of the assay.

1.5 References

1. Ideker, T., Galitski, T. & Hood, L. A new approach to decoding life: Systems biology. *Annu Rev Genom Hum G* **2**, 343-372 (2001).
2. Lindsley, J.E. & Rutter, J. Whence cometh the allosterome? *Proc Natl Acad Sci U.S.A.* **103**, 10533-10535 (2006).
3. Kim, Y.G., Lou, A.C. & Saghatelian, A. A metabolomics strategy for detecting protein-metabolite interactions to identify natural nuclear receptor ligands. *Mol Biosyst* **7**, 1046-1049 (2011).
4. Tagore, R., Thomas, H.R., Homan, E.A., Munawar, A. & Saghatelian, A. A Global Metabolite Profiling Approach to Identify Protein-Metabolite Interactions. *Journal of the American Chemical Society* **130**, 14111-+ (2008).
5. Li, X.Y., Gianoulis, T.A., Yip, K.Y., Gerstein, M. & Snyder, M. Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **143**, 639-650 (2010).
6. Orsak, T. et al. Revealing the Allosterome: Systematic Identification of Metabolite-Protein Interactions. *Biochemistry* **51**, 225-232 (2012).
7. Lee, J.M., Gianchandani, E.P., Eddy, J.A. & Papin, J.A. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Computational Biology* **4** (2008).

8. Lee, J.M., Gianchandani, E.P. & Papin, J.A. Flux balance analysis in the era of metabolomics. *Brief Bioinform* **7**, 140-150 (2006).
9. Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. & Shlomi, T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics* **26**, i255-i260 (2010).
10. Vermeersch, K.A. & Styczynski, M.P. Applications of metabolomics in cancer research. *Journal of Carcinogenesis* **12**, 9 (2013).
11. Monod, J. & Jacob, F. General Conclusions - Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb Sym* **26**, 389-& (1961).
12. Monod, J., Wyman, J. & Changeux, J.P. On the Nature of Allosteric Transitions: A Plausible Model. *Journal of Molecular Biology* **12**, 88-118 (1965).
13. Koshland, D.E., Jr., Nemethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365-385 (1966).
14. Suzuki-Fujimoto, T. et al. Analysis of the galactose signal transduction pathway in *Saccharomyces cerevisiae*: interaction between Gal3p and Gal80p. *Molecular and Cellular Biology* **16**, 2504-2508 (1996).
15. Marvin, J.S. et al. The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors. *Proc Natl Acad Sci U S A* **94**, 4366-4371 (1997).
16. Robert, C.H., Decker, H., Richey, B., Gill, S.J. & Wyman, J. Nesting: hierarchies of allosteric interactions. *Proc Natl Acad Sci U S A* **84**, 1891-1895 (1987).
17. Daugherty, M.A. et al. Identification of the intermediate allosteric species in human hemoglobin reveals a molecular code for cooperative switching. *Proc Natl Acad Sci U S A* **88**, 1110-1114 (1991).
18. Zandany, N., Ovadia, M., Orr, I. & Yifrach, O. Direct analysis of cooperativity in multisubunit allosteric proteins. *Proc Natl Acad Sci U S A* **105**, 11697-11702 (2008).
19. Hill, T.L. A proposed common allosteric mechanism for active transport, muscle contraction, and ribosomal translocation. *Proc Natl Acad Sci U S A* **64**, 267-274 (1969).
20. Schnappauf, G., Lipscomb, W.N. & Braus, G.H. Separation of inhibition and activation of the allosteric yeast chorismate mutase. *Proc Natl Acad Sci U.S.A.* **95**, 2868-2873 (1998).

21. Collins, R.A., Kelly, S.M., Price, N.C., Fothergill-Gilmore, L.A. & Muirhead, H. Ligand-induced conformational changes in wild-type and mutant yeast pyruvate kinase. *Protein Eng* **9**, 1203-1210 (1996).
22. Schnappauf, G., Krappmann, S. & Braus, G.H. Tyrosine and tryptophan act through the same binding site at the dimer interface of yeast chorismate mutase. *Journal of Biological Chemistry* **273**, 17012-17017 (1998).
23. Reinhart, G.D., Hartleip, S.B. & Symcox, M.M. Role of Coupling Entropy in Establishing the Nature and Magnitude of Allosteric Response. *Proc Natl Acad Sci U.S.A.* **86**, 4032-4036 (1989).
24. Wu, S., Acevedo, J.P. & Reetz, M.T. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. *Proc Natl Acad Sci U.S.A.* **107**, 2775-2780 (2010).
25. Hardy, J.A., Lam, J., Nguyen, J.T., O'Brien, T. & Wells, J.A. Discovery of an allosteric site in the caspases. *Proc Natl Acad Sci U.S.A.* **101**, 12461-12466 (2004).
26. Freire, E. Can allosteric regulation be predicted from structure? *Proc Natl Acad Sci U.S.A.* **97**, 11680-11682 (2000).
27. Balabin, I.A., Yang, W.T. & Beratan, D.N. Coarse-grained modeling of allosteric regulation in protein receptors. *Proc Natl Acad Sci U.S.A.* **106**, 14253-14258 (2009).
28. Demerdash, O.N.A., Daily, M.D. & Mitchell, J.C. Structure-Based Predictive Models for Allosteric Hot Spots. *PLoS Computational Biology* **5** (2009).
29. Gelperin, D.M. et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development* **19**, 2816-2826 (2005).
30. DeArmond, P.D., Xu, Y., Strickland, E.C., Daniels, K.G. & Fitzgerald, M.C. Thermodynamic Analysis of Protein-Ligand Interactions in Complex Biological Mixtures using a Shotgun Proteomics Approach. *J Proteome Res* **10**, 4948-4958 (2011).
31. West, G.M., Tang, L. & Fitzgerald, M.C. Thermodynamic analysis of protein stability and ligand binding using a chemical modification- and mass-spectrometry based strategy. *Analytical Chemistry* **80**, 4175-4185 (2008).
32. Ong, S.E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
33. Manabe, Y., Mukai, M., Ito, S., Kato, N. & Ueda, M. FLAG tagging by CuAAC and nanogram-scale purification of the target protein for a bioactive metabolite

- involved in circadian rhythmic leaf movement in Leguminosae. *Chem Commun* **46**, 469-471 (2010).
34. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943 (2006).
 35. Atsumi, S. et al. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab Eng* **10**, 305-311 (2008).
 36. Okuda, S. et al. KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res* **36**, W423-W426 (2008).
 37. Hirschman, J.E. et al. Genome Snapshot: a new resource at the *Saccharomyces* Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res* **34**, D442-D445 (2006).

Chapter 2: Untargeted metabolomics yields insights into high-level metabolic differences in two Atlantic salmon cohorts

This chapter is largely taken from “Differential metabolite levels in response to spawning-induced inappetence in Atlantic salmon *Salmo salar*,”¹ which was published in *Comparative Biochemistry and Physiology, Part D*. All experimental design, handling of fish, and collection of serum samples was conducted by authors Rocco C. Cipriano and Alistair D.M. Dove. Sample work-up and analysis by gas chromatography-mass spectrometry was conducted by McKenzie L. Smith, and data processing/statistical analyses were conducted by McKenzie L. Smith and Kathleen A. Vermeersch, both under the advisement of Mark P. Styczynski.

2.1 Introduction

The salmonid family includes cool temperate anadromous fishes, including many species of important commercial and conservation concern such as Atlantic salmon *Salmo salar* (*S. salar*). All salmonid fish experience conditions of extreme physiological change over the course of their lives, including the shift from freshwater to seawater and subsequently back to freshwater for spawning. In addition, there are also extended periods with little to no active feeding during overwintering and during the spawning run. One might generally expect that extended periods of inappetence could have a deleterious impact on an animal, which could in turn be detected via circulating biomarkers; perhaps the most well-known example of this is ketosis and acidosis in humans, which can occur in response to starvation or prolonged fasting. It was recently found that inappetent whale sharks could be distinguished by circulating biomarkers, including markers such as keto-acids that are

similar to those expected in human starvation². However, it is not known if inappetence of during migration and spawning could even be attributed to or distinguished by some form of abnormal physiology. It is unclear whether long-term inappetence is an inherently stressful event (and thus potentially dangerous as it could compound the effect of anthropogenic stressors), or whether this is a natural condition that the fish have evolved or adapted to efficiently handle during their regular life cycle. Characterizing and understanding the response of salmonid fish to such an extreme condition may provide insight into the physiology of all fishes and the potential challenges faced by some important endangered populations during their complex life histories. Taken together, these insights may help enable more efficient monitoring of these species, with broader economic impact via the extensive efforts of federal and state wildlife management agencies.

Part of the challenge of exploring these potentially stressful physiological changes over the lives of these fish is in the overall lack of biochemical knowledge about the animals. A common approach in veterinary and husbandry practice is to measure serum or plasma levels of molecules with already-known clinical significance in the specific species under study^{3,4}, or molecules with known significance in other model species that are expected to be important in the species under study⁵⁻⁷. However, it is generally unknown what the serum levels of a given metabolite (e.g., glucose or free fatty acids) should be in normal versus different diseased or affected states for an arbitrarily-chosen species of fish. To this end, discovery-oriented approaches are particularly valuable to begin to understand the key determinants and identifiers of animal response to different conditions.

There has been increasing use of high-throughput, systems scale (“-omics”) technologies to assess the physiological state of salmonids under different conditions. Functional genomics approaches have studied the transcriptional changes in the liver and other tissues of salmonids⁸⁻¹⁰. There has also been some recent attention to exploration of salmonid physiology through the systematic measurement of small-molecule biochemical intermediates (metabolites) in the field known as “metabolomics”¹¹⁻¹⁵. The majority of the focus in such studies has been on the toxicological and environmental impacts on the fish, including petrochemicals¹² and pesticides¹⁵. While the study of such anthropogenic sources of disturbance to the salmonid life cycle are obviously of value, a better understanding of the salmonid’s response to natural life cycle events would help to provide a more detailed picture of salmonid physiology and biochemistry, as well as suggest whether these life cycle events might compound the effects of anthropogenic stressors.

Recently, there was one metabolomics study of a salmonid response to short-term fasting, in juvenile rainbow trout *Oncorhynchus mykiss*¹⁶. That study used proton nuclear magnetic resonance (NMR) spectroscopy to measure metabolite levels in plasma, liver, and muscle extracts, and found significant changes in lipids and energy metabolites in these samples. However, the authors noted that using hatchery-reared fish under laboratory conditions may limit generalization of the results, as wild fish are usually leaner and may have different responses to fasting. Additionally, the scope and number of analytes studied (due to sensitivity requirements and limitations on analyte identification with an NMR-based approach) was limited. Finally, that study involved a 28-day fasting experiment that did not include the impact of the months-long inappetence that salmonids experience during

spawning and overwintering. A subsequent study used mass spectrometry in place of NMR, but did not address the other limitations¹⁷.

The purpose of this work is to test whether there are indications that long-term inappetence of salmon is a pathological state as in other organisms, and to explore the changes in blood metabolite profiles caused by long-term inappetence. These goals are accomplished using a highly sensitive mass spectrometry-based approach to compare differences in blood metabolite levels between the same generation of hatchery-raised and -fed control fish as compared to wild, sea-run fish fasting for five to six months. Serum samples were taken from each of these corresponding cohorts and then analyzed to identify biochemical indicators of response to inappetence.

2.2 Results and discussion

2.2.1 Overall data set and reproducibility

The inappetence of returning spawning Atlantic salmon is a profound physiological change hypothesized to be reflected in their serum metabolite levels. Metabolite profiling using two-dimensional gas chromatography coupled to mass spectrometry supports this hypothesis. Overall, 139 analytes were retained after quality control filtering (Figure 2.1), including unknown and known analytes and with the potential of having multiple derivatization peaks for the same analyte. Known analytes were only identified as metabolites if they had entries in HMDB (used as a proxy for general biological validity, though potentially omitting some xenobiotics). Technical duplicate measurements showed

high reproducibility; the average pairwise (Pearson) correlation coefficient of duplicate measurements was 0.96. Of this set of reproducible analytes, 57 were resolved to unique, identifiable metabolites; this reduced set of analytes had an average pairwise correlation of 0.97.

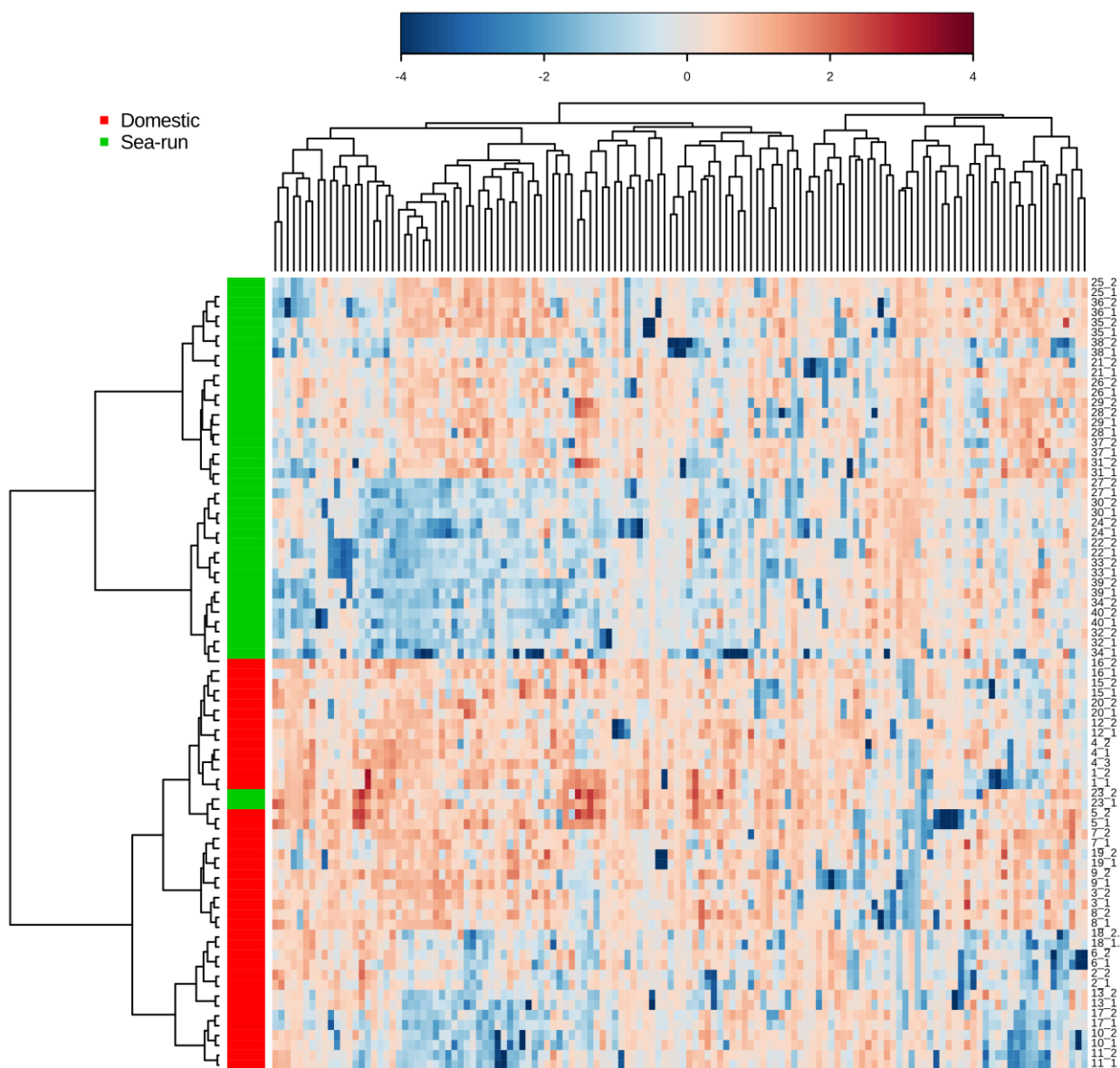


Figure 2.1: Heatmap of all normalized metabolite levels. Log-transformed and autoscaled metabolite peak areas presented in heatmap format for all measured analytes (including those not assigned identities during data analysis). Red entries indicate higher metabolite levels, blue indicates lower levels. Sample classes are indicated on the left by color code (red for domesticated, green for sea-run), and metabolites are organized using hierarchical clustering.

2.2.2 Multivariate analysis of the metabolite profiles of the fasting and feeding fish

Gross changes in metabolic physiology were easily detectable using principal components analysis of the entire set of measured analytes (Figure 2.2). The second principal component was the dominant axis of variation in determining the nutritional status of any given animal. The first principal component predominantly reflected a batch effect that was not able to be completely removed via data normalization (Figure 2.3). However, the combination of these two principal components (essentially allowing the second component to provide separation relative to the first component) provides almost perfect separation between the two sample classes. There are two notable outliers in the data: both technical replicates for sample 23, a fasted animal that clustered with the fed animals. This misclassification is consistent across all subsequent data analyses; it is unclear whether this is the result of a mislabeled tube, whether this animal had somehow eaten at some time pre- or post-spawning, or whether that animal just had such substantial biological variation from the rest of its cohort. Limiting analysis to only the identified metabolites, PCA showed similar separation based on the first and second principal components (Figure 2.4).

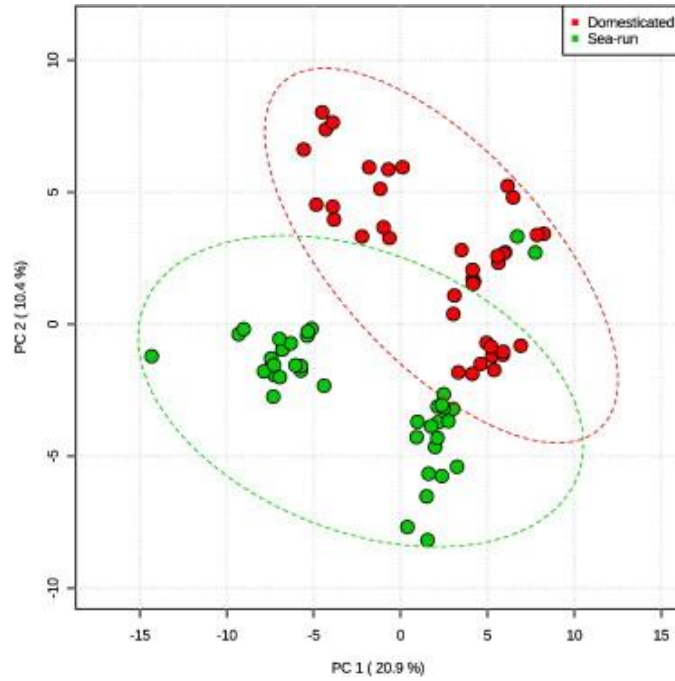


Figure 2.2: Principal components analysis shows strong separation between domesticated and sea-run fish. Unsupervised dimensional reduction using all measured analytes yields separation in just two principal components. The first principal component mostly captures a sample-ordering-based batch effect that could not be removed by data normalization approaches. The second principal component is the dominant axis of separation between the two classes, after accounting for batch-related differences. Red markers represent domesticated fish, green markers represent sea-run (inappetent) fish, and the dotted circles represent 95% confidence regions on the respective sample classes. Axes also indicate the fraction of overall variation capture by each principal component.

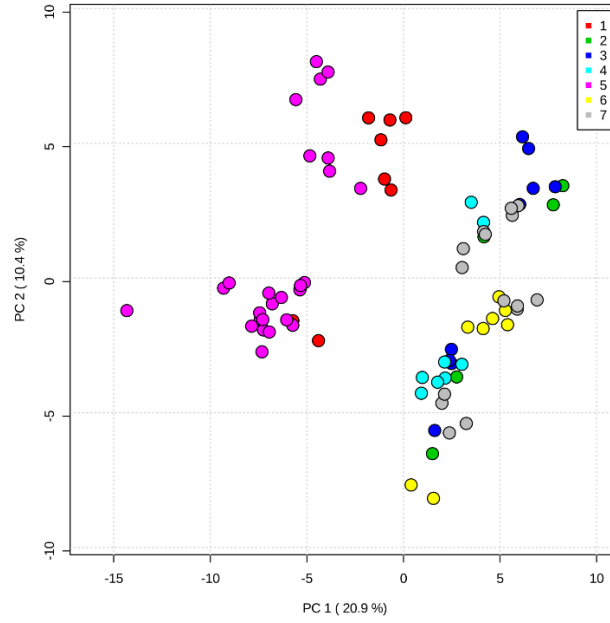


Figure 2.3: PCA visualization of batch effects. Due to an instrumentation issue during data acquisition, two batches of samples (1 and 5) exhibited batch effects evident during principal components analysis. Note that for each batch (indicated by different colors on the plot), principal component 1 predominantly represents the batch effect (though some fish sample class effect is also included in this principal component). Principal component two, as indicated in the main text, predominantly captures the differences between sample classes, which can be separated from the confounding batch issue due to our randomized batch design.

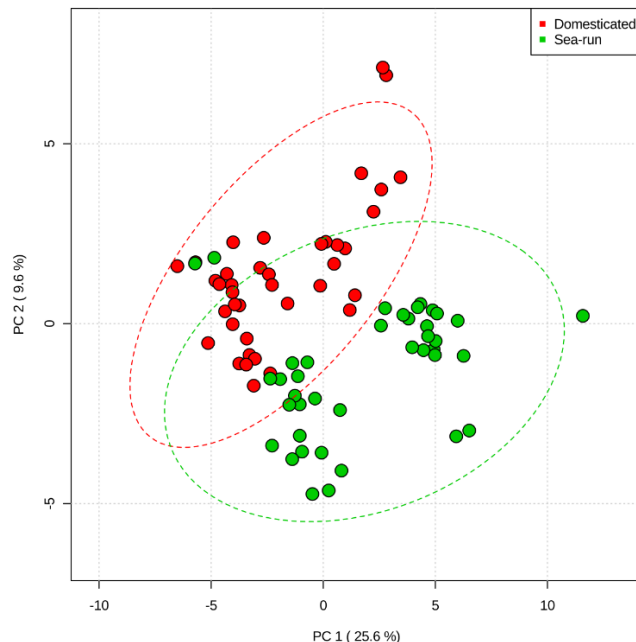


Figure 2.4: Unsupervised dimensional reduction yields separation in just two principal components. The first principal component mostly captures a sample ordering-based batch effect that could not be removed by data normalization and quality control. The second principal component is the dominant axis of separation between the two classes, after accounting for batch-related differences. Only identified metabolites (see Methods) are used for this analysis. Red markers represent domesticated fish, green markers represent sea-run (inappetent) fish, and the dotted circles represent 95% confidence regions on the respective sample classes. Axes also indicate the fraction of overall variation capture by each principal component.

This unsupervised principal components analysis indicates that once a sample-ordering batch effect was accounted for, one of the dominant modes of natural variation in the data was controlled by the fed vs. fasting condition of the animals. The fact that separation occurred naturally in this unsupervised approach in the first two principal components suggested that the differences that were being measured were real and biologically relevant. PLS-DA supported this finding (Figure 2.5): the two classes of animals, with the exception of the previously noted outliers, were clearly separated in only one component. 10-fold cross-validation yielded a Q^2 value of 0.82 in two components. The separation was significant with a $p < 0.0005$ using a label permutation test with two components. Again, performing the same analysis using only the putatively identified metabolites yielded

similar visualization results with similar statistical significance and cross-validation scores. The metabolites identified as most important in the PLS discriminating axis were largely the same as the metabolites that were individually statistically significant, and so those individually significant metabolites will be considered further below.

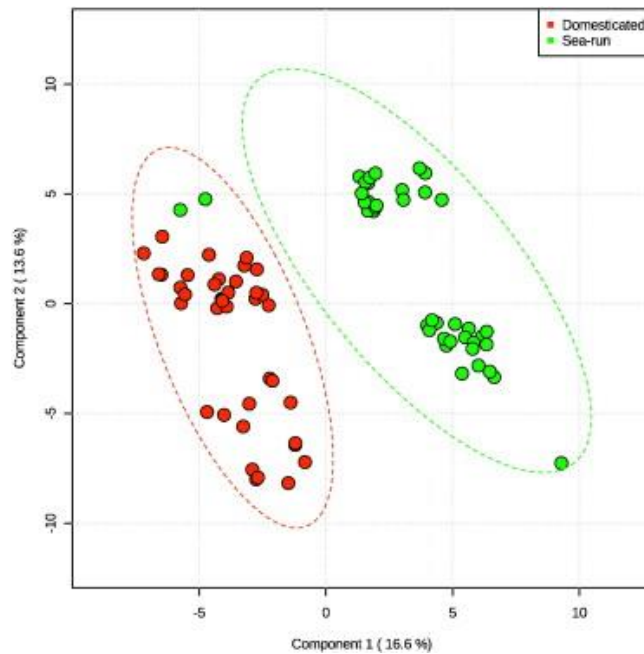


Figure 2.5: Partial least squares discriminant analysis shows strong separation between domesticated and sea-run fish. Strong separation between the two classes can be seen in only one dimension using a supervised dimensional reduction approach including all measured analytes. Red markers represent domesticated fish, green markers represent sea-run (inappetent) fish, and the dotted circles represent 95% confidence regions on the respective sample classes.

2.2.3 Significantly different analytes between fasting and feeding Atlantic salmon serum

Of the 137 overall consistently measured analytes, 58 were identified as significantly different between the sea-run inappetent and the domesticated feeding fish (t-tests with $p < 0.05$ after false discovery rate correction¹⁸. Of the 57 analytes resolved to unique, identifiable known metabolites, 27 of these were found to be significantly different between the sea-run inappetent and the domesticated feeding fish (Table 2.1). However, an inherent limitation of this study is the confounding effect of diet on serum levels. It is well-

known that having a high level of a specific metabolite in a food can cause elevated levels of that nutrient in the serum. For example, increasing levels of eicosapentaenoic acid in feed was shown to increase serum levels of eicosapentaenoic acid¹⁹. (However, this is not universally true; in that same study, increasing levels of alpha-linolenic and docosahexaenoic acid did not change serum levels of the respective molecules.) Nonetheless, to avoid any potential confounding factors, it is most prudent to focus on the analytes that increased in the inappetent fish; these are very unlikely to be artifacts of the feeding regimen.

Table 2.1: Significantly different metabolites between domesticated and sea-run fish. Only analytes with metabolite identities assigned during data processing (see Methods) are presented; when multiple derivatization products for the same metabolite were identified, the analyte with the fewest missing values across the samples and QCs was selected as representative for that metabolite. Fold changes are log₂, with negative values indicating depletion in sea-run fish and positive values indicating accumulation in sea-run fish. p-values are after false discovery rate multiple hypothesis correction; only metabolites with p-values < 0.05 are presented.

Metabolite	Log-fold change in sea-run	FDR-corrected p-value
Gondoic acid	1.98	1.92E – 10
Erucic acid	2.82	3.75E – 10
L-Alanine	– 0.72	3.21E – 07
Eicosapentaenoic acid	– 0.46	9.61E – 07
L-Valine	– 0.40	9.61E – 07
Putrescine	– 0.70	9.61E – 07
<i>myo</i> -Inositol	– 0.60	2.10E – 06
L-Leucine	– 0.50	8.88E – 06
γ-aminobutyric acid	– 1.61	8.02E – 05
Myristic acid	– 0.58	1.07E – 04
(S)-3-Hydroxyisobutyric acid	– 0.38	1.36E – 04
Docosahexaenoic acid	– 0.30	2.57E – 04
L-Tyrosine	– 0.35	4.70E – 04
2-Methyl-3-hydroxybutyric acid	– 0.63	0.0014
L-Isoleucine	– 0.38	0.0019
4-Hydroxyproline	0.47	0.0021
Palmitoleic acid	– 0.52	0.0021
Fumaric acid	– 0.48	0.0025
D-Glucose	– 0.57	0.0041
Scyllitol	0.53	0.0041

Table 2.1 (continued)		
<i>N</i> -Acetyl-D-glucosamine	– 0.49	0.0052
L-Methionine	– 0.99	0.0053
D-Fructose	– 0.53	0.0066
D-Tagatose	– 0.58	0.0087
α -Linolenic acid	– 0.23	0.028
Ribitol	0.16	0.029
2-Hydroxy-3-methylbutyric acid	– 0.44	0.04
Nervonic acid	0.85	0.043

2.2.4 Significantly accumulated analytes in inappetent Atlantic salmon serum

Of the 27 significantly different analytes between feeding and inappetent fish, only 6 were accumulated in the inappetent fish; the remainder were all depleted in the inappetent fish. This is perhaps not unexpected, as continuing inappetence to the point of starvation may deplete the blood of many nutrients, limited by the rate at which nutrients can be mobilized from storage and other tissues. If this is true, it suggests that the six metabolites that were accumulated may be particularly interesting. Of those six, three were ω -9 monounsaturated fatty acids (erucic acid [22:1 ω 9], gondoic acid [20:1 ω 9], and nervonic acid [24:1 ω 9]), and two were sugar alcohols (scyllitol and ribitol). Interestingly, no ω -9 monounsaturated fatty acids were found to be depleted. The two sugar alcohols represented half of the annotated sugar alcohols detected in this study. The other two sugar alcohols that were not increased in inappetent fish (*myo*-inositol and glycerol) have well-known signalling and lipid transport roles which may help to explain why they did not change in register with the rest of the sugar alcohols.

2.2.5 Spawning Atlantic salmon and whale sharks have different metabolic responses to inappetence

Long-term inappetence or starvation is a phenomenon that has been shown to have a significant effect on the physiology of an organism. It has been proposed that the starvation process goes through three phases, typically associated first with high metabolism of glycogen, followed by predominant use of lipids for energy generation, and finally ending in use of proteins for energy generation²⁰. This often results in changing levels of lipids, proteins, and the metabolites in blood. One common response during this second phase, catabolism of fatty acids, is the generation of ketone bodies that elevates circulating ketone levels. While this is a well-known and common response in endotherms, in ectotherms there is not evidence for an increase in circulating ketone bodies.

In their metabolomic study of whale sharks², Dove *et. al* detected seven keto-acids that were more prevalent among samples from pathologically inappetent whale sharks than from animals that were feeding normally. These acids were interpreted as representing a catabolic state wherein the inappetent whale sharks were breaking down their own tissue proteins and liberating different ketone species into the circulation. Interestingly, the current study did not detect a pronounced keto-acid pattern among inappetent despite known protein breakdown during salmon inappetence, for which three possible explanations have been identified. First, the analytical techniques differed between the two studies (direct MS for whale sharks, GCxGC-MS for Atlantic salmon) and may differ in sensitivity to keto-acids. However, keto-acids should generally be measurable by GC-MS and several keto-acids have previously been detected and tracked on the specific instrument used. Since the keto-acids were infrequently detectable in healthy samples from the whale sharks, this further suggests that the reason they were not observed in either set of fish may

be that they were not present. Second, there may be profound differences in response to pathological inappetence between sharks and salmon, where acidosis/aciduria is induced in one but not the other. Third, and probably most likely, there is a real metabolic difference between pathological inappetence in the whale sharks and adaptive or “normal” inappetence in the spawning run Atlantic salmon. This would imply that Atlantic salmon may have evolved tolerance to the prolonged anorexia during the spawning run, which would probably not come as a surprise.

2.2.6 Fatty acid metabolism response to inappetence appears to be dependent on fatty acid structure

The ω -9 monounsaturated fatty acids that were found to be accumulated in inappetent fish were erucic acid, gondoic acid, and nervonic acid; these are of particular interest due to the specificity of their behavior and the strength of their signal. Gondoic acid and erucic acid exhibited the two most significant and largest magnitude changes of all annotated metabolites (approximately 4-fold and 8-fold changes, respectively), suggesting a particularly important or unique role in the physiological response to starvation. While three measured ω -9 monounsaturated fatty acids were accumulated in inappetent fish, all three ω -3 polyunsaturated fatty acids were found to be depleted in inappetent fish (eicosapentaenoic and docosahexaenoic acids with $p < 0.001$, α -linolenic acid with $p < 0.05$ after FDR correction; Figure 2.6). Again, these measurements may be confounded by the feeding regimen of the control fish, which contains fish meal and thus is high in fatty acids. Nonetheless, previous studies showed changes of less than 50% in blood fatty acid levels for increased feeding content of docosahexaenoic and eicosapentaenoic acid from 0% to

2% of dry weight, which is consistent with the observed levels in this work¹⁹. (Unless indicated, this and all other referenced literature did not use exclusively female fish.) Importantly, these changes are substantially different from the 8-fold, 4-fold, and 2-fold increases observed in the ω -9 fatty acids. Thus, while the specific interpretation of ω -3 polyunsaturated acids cannot be distinguished as being diet-related or inappetence-related, it is clear that they are nowhere near the strongly accumulated status of the ω -9 monounsaturated acids. The mechanism for such specific regulation or utilization of certain unsaturated fatty acids remains unclear, but is noteworthy, particularly in light of previous knowledge of non-esterified or free fatty acids in response to starvation.

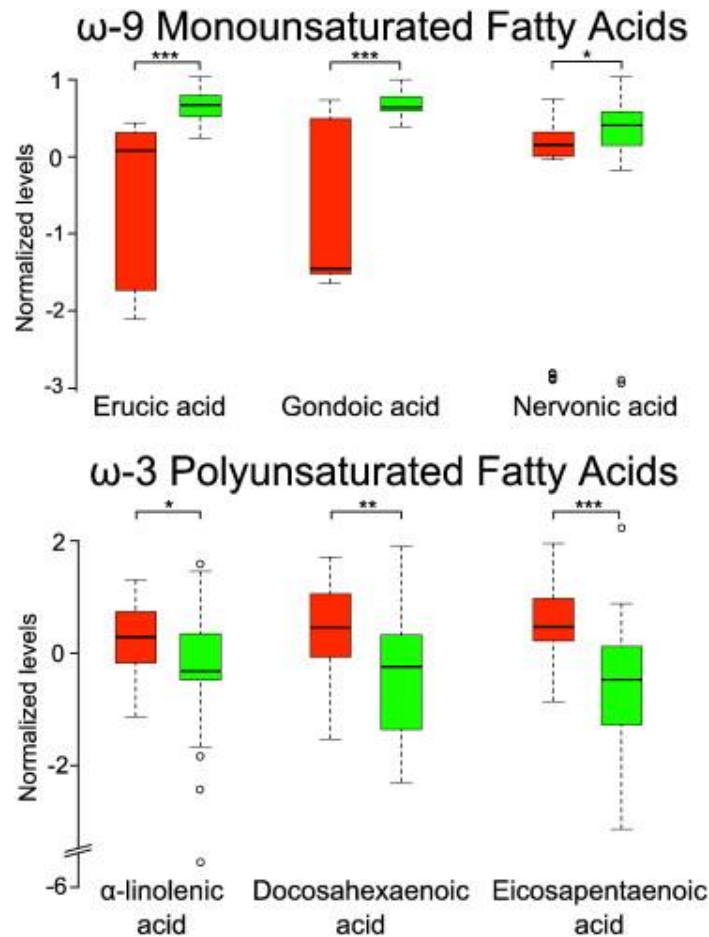


Figure 2.6: Fatty acids exhibit structure-specific differences between domesticated and sea-run fish. The ω -9 monounsaturated fatty acids measured (top row) exhibited significant changes between domesticated and sea-run fish that were in the opposite direction of the ω -3 polyunsaturated fatty acids (bottom row), suggesting some structure-specific utilization of fatty acids in response to prolonged inappetence. Red represents domesticated fish, green represents sea-run fish. Colored bar represents the span of the middle 50% of samples, with the median represented by a dark black line within the colored bar. Error bars show two standard deviations, with outliers represented as circles. Values are log transformed and autoscaled. Asterisks indicate FDR-corrected p-values: * $p < 0.05$, ** $p < 0.0005$, *** $p < 10^{-6}$.

It is well-known that fatty acids are mobilized from the liver and from muscle tissue in response to extended periods of starvation in fish²¹. Previous work studying fasting in a different salmonid species¹⁶ found clear tissue-specific mobilization patterns of lipids and some degree of structure-specific changes in lipid concentrations. Specifically, serum levels of unsaturated lipids increased after 28 days of fasting, while other measures of serum lipid levels either increased or decreased. A related study by Baumgarner and

Cooper found many saturated fatty acids to decrease after 28 days, but also saw docosahexaenoic acid decrease¹⁷, in contrast with the prior result. General analysis of free fatty acids, regardless of saturation, has shown decreases in plasma levels²². Other work by Pottinger and colleagues saw increases in plasma free fatty acids in the first two months of starvation, followed by a decrease in serum free fatty acids²³.

This work refines this analysis further, identifying not the general class of unsaturated lipids as increasing in fasted fish serum, but specifically ω -9 monounsaturated lipids as increasing while ω -3 polyunsaturated lipids decrease. Both 18:1 (ω -9) and 20:1 (ω -9) have been shown to be mobilized from liver and muscle stores during fasting²⁴, with 18:1 (ω -9) fatty acids being the most readily used energy source in teleosts²⁵. However, this study finds no significant change in 18:1 (ω -9) serum levels, while significant increases in 20:1 and 22:1 fatty acids are observed. Alternatively, after such an extended period of inappetence, the ω -3 fatty acids may be so depleted due to lack of intake (they are essential fatty acids) that the salmonid is producing more of the ω -9 fatty acids to make up for the deficiency for any functions that may be amenable to either type of molecule.

Overall, these results suggest an important or unique role for unsaturated fatty acids in the physiological response to prolonged starvation. While the unsaturated fatty acids split structure-specifically between accumulation and depletion in long-term inappetent fish, saturated fatty acids show no such strong signal. Of the 10 saturated fatty acids measured in the serum samples, only two are significantly changed, with palmitic acid increasing and myristic acid decreasing in inappetent fish. The ω -9 fatty acids may be playing a critical

role in mobilizing energy sources from liver and muscle; previous work has found this to be true for muscle levels of one ω -9 fatty acid¹⁶, which could cause higher serum levels of these acids if this is indeed occurring. However, it is also possible that unsaturated fatty acids are being non-specifically mobilized from tissue stores, with the ω -9 fatty acids not being utilized as heavily and thus essentially accumulating.

2.2.7 Polyol accumulation might be indicative of a currently unknown response to starvation

The strong changes in polyols are also of note. Polyols are generally known to serve as osmolytes with potential cryoprotectant capabilities²⁶. However, there is fairly little known about polyol levels in fish. The accumulated polyol scyllitol was first discovered in sharks and skates in 1858 by Staedler and Frerichs^{27,28}. It is known that skates have the highest levels of scyllitol in their kidneys and livers. Scyllitol is an osmolyte known to be present in the blood of teleosts and other deep-sea animals and which may function as a cryoprotectant^{29,30}. It has also received significant interest for its role in brain chemistry across species from fish to mammals, and in particular for its potential as a therapeutic for Alzheimer's disease, likely related to its stabilization of and interaction with proteins^{27,28}. The relevance of these increased sugar alcohol levels, particularly as related to prolonged starvation, warrants further study, as this may reflect an unknown metabolic or biochemical response to starvation that has not previously been characterized.

2.2.8 Amino acids supply additional energy sources to fasting

Branched-chain amino acids (BCAAs) are another noteworthy set of metabolites. Three of the fifteen most significant changes were in the BCAAs valine, leucine, and isoleucine, which were all depleted; this is consistent with previous results from short-term inappetence in domesticated rainbow trout¹⁷. Taken together, BCAAs have an important role in continued energy generation after depletion of muscle glycogen stores³¹. Amino acids, as breakdown products of muscle protein, are the major source of energy for fish during extended starvation^{32,33}. In fact, though, many different amino acids can be transformed into the anaplerotic biosynthetic intermediates needed to sustain oxidative phosphorylation and gluconeogenesis³⁴. A general utilization of proteolysis for energy production thus could potentially lead to a general decrease in serum amino acid levels, with the post-glycogen depletion role of BCAAs potentially leading to their even further and more specific uptake and use. This is particularly relevant here, as only 60% of the measured proteinogenic amino acids were found to be depleted in inappetent fish. As noted above, interpretation of these results may be limited based on the potential confounder of diet. However, a number of factors suggest that the effects observed may not be totally diet-related. First, the results are, as described above, consistent with multiple previous studies that did not have potential diet confounders^{17,35}. Second, the control fish feed was high in protein, yet only some of the amino acids were depleted in the inappetent fish, suggesting something more complex than just food-related elevation of serum amino acid levels in control fish.

Looking more broadly at amino acid metabolism, even more changes are seen in response to inappetence. Alanine was the third-most significantly altered metabolite (and the most

significant depleted metabolite), consistent with previous work identifying changes in alanine in response to short-term inappetence in rainbow trout¹⁷. However, neither glycine nor serine previously reported in domesticated, short-term inappetence were observed here, though the general trends of amino acid depletion were consistent. Pathway-level analysis using MetaboAnalyst identified a number of significantly altered pathways (Table 2.2), where among the most significant with the most support (in terms of numbers of analytes detected) were several specific amino acid metabolism pathways beyond the BCAAs (all FDR-corrected $p < 10^{-4}$). This supports the idea of significant alterations in amino acids and their biosynthetic intermediates in the blood of inappetent animals.

Table 2.2: Significantly enriched pathways from MetaboAnalyst pathway analysis. P-values are false discovery rate corrected.

Pathway	p-value
Arginine and proline metabolism	2.89E-09
Biosynthesis of unsaturated fatty acids	4.23E-09
Selenoamino acid metabolism	2.50E-07
Aminoacyl-tRNA biosynthesis	3.22E-07
Alanine, aspartate and glutamate metabolism	3.22E-07
Valine, leucine and isoleucine degradation	5.43E-07
Pantothenate and CoA biosynthesis	5.43E-07
Inositol phosphate metabolism	1.27E-06
Ascorbate and aldarate metabolism	1.27E-06
Valine, leucine and isoleucine biosynthesis	4.98E-06
Galactose metabolism	1.91E-05
Butanoate metabolism	0.000114
Tyrosine metabolism	0.000118
Glutathione metabolism	0.000151
Fatty acid biosynthesis	0.000289
Ubiquinone and other terpenoid-quinone biosynthesis	0.000297
Phenylalanine, tyrosine and tryptophan biosynthesis	0.000876
Phenylalanine metabolism	0.000876
Cysteine and methionine metabolism	0.00311
Starch and sucrose metabolism	0.00321
Amino sugar and nucleotide sugar metabolism	0.00403

2.2.9 *Limitations of this study*

While the approaches used to probe for changes in metabolic status in inappetent fish comprise a powerful framework for veterinary, husbandry, and scientific analyses, there are certainly limitations to this framework. The metabolite measurements were only semi-quantitative, enabling comparison between groups but not absolute quantitation of concentrations (highlighted by the inclusion in principal component one of batch effects). Future efforts should focus on including more spike-ins of labeled compounds along with the establishment of calibration curves to attempt to facilitate absolute quantitation. However, since each group had the same amount of serum assayed on the GCxGC-MS, direct comparison of the semi-quantitative results is still valid. In fact, the univariate results likely underestimate potential differences between the groups, as the observed slight batch effect only makes the measured differences less statistically significant. In addition, the fact that only serum was assayed – not the metabolite levels in liver, muscle, or any other potentially relevant tissue – limits the scope in interpreting measurements in a biochemical pathway context, as blood is essentially an integrative biomarker across the entire organism due its interaction with all tissues. Ongoing work measuring metabolite profiles in specific tissues is expected to help enable a more direct interpretation of the results shown here.

Finally, the fact that the control fish were domesticated rather than sea-run adds a potential confounder to the analysis that cannot be decoupled. Certain aspects of this issue are impossible to remedy through a control; for example, it is impossible to use a fed sea-run fish control because only domesticated fish can be made to feed leading up to spawning.

One would expect the long-term feeding patterns dictated by location, activity, and other factors to be different between a fed captive and fed sea-run group before any inappetence, which could contribute to some of the observed differences; however, since going over six months without food is such a significant perturbation, it is reasonable to expect that the inappetence and resulting mobilization of other energy resources would dominate the baseline differences between the animals. Additionally, the sea-run fish had already traveled approximately 75 miles upstream before being captured and held for six months while the domesticated fish had not, which could contribute more to the depletion of resources and metabolic changes in these animals. Again, based on the fact that non-domesticated spawning fish cannot be fed, this is another phenomenon that cannot be decoupled. However, this is an intrinsic and important part of spawning-induced inappetence, so the combination of these two factors is ultimately reasonable.

2.3 Conclusions and context

This study identified interesting, structural class-specific changes in serum metabolite levels in inappetent spawning fish. ω -9 fatty acids were specifically accumulated in inappetent fish, in stark contrast to the majority of significantly changed metabolites, including the ω -3 free fatty acids. This observation, combined with the significant depletion of branched chain amino acids, suggests a possibly uncharacterized unique and specific utilization of metabolites in the long-term stress of spawning-induced inappetence. The significant accumulation of sugar alcohols in the serum was another unexpected observation that is not explainable based on current knowledge of salmonid biochemistry, warranting further investigation.

The mass spectrometry-based strategy used here has shown to be powerful at identifying specific molecules and classes of molecules that change in response to a long-term natural potential stressor of health. Characterizing and understanding the physiological responses of salmonid fish to extreme conditions may provide insight into the natural physiological variability of and the physiology of all fishes, which could in turn serve as a starting point for the identification of markers that can be monitored to assess the health status of fish in a variety of environments, whether wild or bred. Due to the economic importance of and other fish, such endeavors may ultimately have a significant economic impact via the extensive efforts of federal and state wildlife management agencies. The metabolite class-specific findings identified here will provide the impetus for additional, targeted analyses of response to inappetence in Atlantic salmon, as well as justify the metabolomic investigation of other stresses that salmonids may experience over their lifespan.

Untargeted metabolomics techniques have been widely used to identify high-level differences between organismal or cellular cohorts, resulting in data and conclusions of the sort presented in this thesis. Such information can be directly actionable in the case of biomarker identification (particularly in a medical context)³⁶; however, understanding the mechanistic underpinnings of these high-level differences requires delving into analysis of individual pathways³⁷. In order for this information to in turn be applied at an engineering level, mathematical modeling is necessary, and the accuracy of such models can be greatly increased by a comprehensive understanding of regulation at the protein-metabolite level.

In this study of comparative metabolomics in Atlantic salmon, we identified metabolites that were accumulated in the inappetent cohort compared to the fed cohort. These results are tentatively suggestive of possibilities for the underlying biochemistry, but cannot provide definitive mechanistic information. Instead, they indicate areas of metabolism where deeper study and modeling may later yield key insights into Atlantic salmon metabolism, which efforts would be substantially aided by wider knowledge of metabolic regulation.

2.4 Materials and methods

2.4.1 Fish and experimental conditions

The control group was a set of 20 post-spawning adult female domestic *S. salar* housed at the White River National Fish Hatchery (Bethel, VT). These fish spent all 4-5 years at the hatchery, and were feeding spawn in October/November of 2010. The fish were fed a maintenance diet of 1.5% body weight one time daily with Corey AquaBrood (Fredericton, New Brunswick, Canada). Blood samples were collected post-spawn in December of 2010.

The inappetent group consisted of 20 feral sea run female *S. salar* that returned to the U.S. Fish and Wildlife Service (USFWS) Richard Cronin National Salmon Station (Sunderland, MA) in May of 2010. These fish originated at the USFWS White River National Fish Hatchery but were released as either non-feeding fry or one-year smolts into the Connecticut River. At age 5, the fish returned to the Connecticut River during the spring (April/May) to spawn but were captured at the Holyoke Dam fish lift (Holyoke, MA) and transported to the Richard Cronin National Salmon Station. The fish were maintained at

that facility in non-feeding status until they spawned in November of 2010. These fish stop active feeding at the onset of their spawning migrations and do not feed throughout the entire holding period at the Richard Cronin facility (approximately 5-6 months). Blood samples were collected post-spawn in December of 2010, the day after the control group samples were taken.

Sampling of fish began at 7:30 AM for each group of fish; the control fish had last been fed at 4:00 PM the night before. The average size of sampled fish were 78.3 cm and was similar between the two groups. The sea-run fish available for sampling were to be reconditioned for feeding and subsequent breeding after this study. Since the hatchery typically has enough sperm available for fertilization, only female kelt are maintained, and so the sea-run fish available for sampling were all female. Accordingly, the domesticated (control) fish selected for sampling were exclusively female to remove any possible gender effects from the study.

2.4.2 Sampling procedure and sample preparation

Blood was drawn from each fish from its hemal arch using heparinized syringes and stored on ice for one hour. A Pasteur pipette was inserted into the tube to ring the pre-formed clot and allow the red blood cells to collect at the bottom of the tube overnight at 4 °C; the resulting serum was then transferred to new tubes, and stored at -80 °C until metabolomics analysis. Aliquots of 50 µL of each of these samples were then further processed for mass spectrometry analysis by first using 150 µL of ice-cold acetonitrile to precipitate any protein that may have remained. The sample was then centrifuged at 21,100 xg for 7

minutes and the supernatant removed. The resulting serum was then evaporated to dryness using a CentriVap centrifugal concentrator at 40°C and derivatized following the protocol laid out by Kind, *et. al.*³⁸ Briefly, the samples were resuspended in 10 µL of 40 mg/mL *O*-methylhydroxylamine hydrochloride in pyridine and shaken at 1400 rpm for 90 minutes at 30°C. Then, 90 µL of *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) + 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Lafayette, CO) was added to the samples, which were then shaken at 1400 rpm for 30 minutes at 37 °C. Samples were centrifuged at 21,100 xg for 3 minutes, and 50 µL of the supernatant was added to an autosampler vial. Samples were spiked with 0.25 µL of a retention time standard solution consisting of fatty acid methyl esters (FAMES) and an internal standard of nonadecanoic acid methyl ester dissolved in dimethylformamide .

2.4.3 GCxGC-MS analysis

A LECO Pegasus 4D instrument with an Agilent 7890A gas chromatograph and 7683B autosampler was used to analyze the samples. Complete method details can be found in Appendix A. Briefly, 1 µL sample volume was injected into the inlet in splitless mode at 250 °C. A constant flowrate of helium was kept at 1 mL/min. The oven temperature program started at 50 °C and ramped to 295 °C over 28 minutes, with the secondary oven kept at an offset of 15 °C and the modulator temperature offset kept at 35°C from the main oven. The first column was an HP-5 30m x 0.320 mmID x 0.25 µm (Agilent, Santa Clara, CA) and the second was Rtx-200 2m x 0.180 mmID x 0.20 µm (Restek, Bellefonte, PA). Each experimental sample was analyzed with technical duplicate injections, with sample order randomized to limit batch effects. A pooled quality control sample was created from

a small aliquot of each sample. This pooled quality control sample was prepared in the same way as the two sets of experimental samples, and was injected after every three to six samples to facilitate later data correction.

2.4.4 Data Analysis

Sample runs were first analyzed in the instrument manufacturer's software, ChromaTOF (LECO, St. Joseph, MI), to determine baseline, peak area, and peak identification. Briefly, the following parameters were used: a baseline offset of 0.5, automatic smoothing, first dimension peak width of 24 seconds, second dimension peak width of 0.25 seconds, and a match of 700 required to combine peaks with a minimum S/N of 10 for all subpeaks. Peaks were required to have a signal to noise ratio (S/N) of 100 and have a minimum mass spectral similarity score of 800 (out of 1000) before assigning a name (both of these parameter choices are rather conservative). Unique mass was used for area and height calculation. To align the samples, MetPP³⁹ (<http://metaopen.sourceforge.net/metpp.html>) was used with all default parameters.

Peaks were filtered based on reproducibility in the quality control samples, as previously described⁴⁰. Only peaks with less than 50% missing values and less than 50% coefficient of variation in the quality control samples were deemed reproducible and retained for further analysis. An initial missing value imputation step was then performed based on injection replicates; if both replicates were missing a value for a given metabolite, they were replaced by a small value based on the smallest measured peak. If one replicate had a value and the other did not, the value was left as missing for later downstream imputation.

Finally, MetaboAnalyst (<http://metaboanalyst.ca/>) was used for statistical analysis⁴¹. The data matrix was log transformed and autoscaled before analysis, with variable filtering based on interquartile range and k-nearest neighbors (KNN) imputation of missing values. Analysis methods used include t-tests with false discovery rate correction (FDR) for multiple hypothesis testing¹⁸, principal components analysis as an unsupervised method for determining the axes of variation in the serum metabolite levels, and partial least squares discriminant analysis (PLS-DA) for supervised analysis, specifically targeting the combinations of metabolite levels that best separate the fed from the fasted animals. For enrichment analysis, metabolite pathway enrichment analysis (MPEA) was used. The data set was uploaded as discrete data with compound names. Metabolites were matched to their HMDB codes before processing the data. Data processing followed the same steps as listed above for missing value imputation and data normalization. The *Danio rerio* (zebrafish) pathway library was used for analysis and an in-house metabolite reference library based on detectable metabolites for the instrument used in this study was uploaded. Global test was used for pathway enrichment analysis, with relative-betweenness centrality as the pathway topology analysis. Pathways with an FDR < 0.05 were considered significantly enriched. Based on the lack of availability of a pathway library specifically for the species under study, analysis and interpretation of pathway-level results was generally limited and focused on putatively well-conserved pathways.

2.5 References

1. Cipriano, R.C., Smith, M.L., Vermeersch, K.A., Dove, A.D.M. & Styczynski, M.P. Differential metabolite levels in response to spawning-induced inappetence in Atlantic salmon *Salmo salar*. *Comp Biochem Phys D* **13**, 52-59 (2015).
2. Dove, A.D.M. et al. Biomarkers of Whale Shark Health: A Metabolomic Approach. *PLoS One* **7** (2012).
3. Jeffries, K.M. et al. Temporal changes in blood variables during final maturation and senescence in male sockeye salmon *Oncorhynchus nerka*: reduced osmoregulatory ability can predict mortality. *Journal of fish biology* **79**, 449-465 (2011).
4. Parry, G. Osmotic and Ionic Changes in Blood and Muscle of Migrating Salmonids. *J Exp Biol* **38**, 411-& (1961).
5. Dove, A.D.M., Arnold, J. & Clauss, T.M. Blood cells and serum chemistry in the world's largest fish: the whale shark *Rhincodon typus*. *Aquat Biol* **9**, 177-183 (2010).
6. Dove, A.D.M., Sokolowski, M.S., Bartlett, S.L. & Bowser, P.R. Spatio-temporal variation in serum chemistry of the lobster, *Homarus americanus* Milne-Edwards. *J Fish Dis* **28**, 663-675 (2005).
7. Hevroy, E.M. et al. Ghrelin is involved in voluntary anorexia in Atlantic salmon raised at elevated sea temperatures. *Gen Comp Endocr* **175**, 118-134 (2012).
8. Evans, T.G. et al. Transcriptomics of environmental acclimatization and survival in wild adult Pacific sockeye salmon (*Oncorhynchus nerka*) during spawning migration. *Mol Ecol* **20**, 4472-4489 (2011).
9. Martin, S.A., Douglas, A., Houlihan, D.F. & Secombes, C.J. Starvation alters the liver transcriptome of the innate immune response in Atlantic salmon (*Salmo salar*). *Bmc Genomics* **11** (2010).
10. Miller, K.M. et al. Salmon spawning migration: Metabolic shifts and environmental triggers. *Comp Biochem Phys D* **4**, 75-89 (2009).

11. Hines, A. et al. Discovery of Metabolic Signatures for Predicting Whole Organism Toxicology. *Toxicol Sci* **115**, 369-378 (2010).
12. Lin, C.Y. et al. Characterization of the metabolic actions of crude versus dispersed oil in salmon smolts via NMR-based metabolomics. *Aquat Toxicol* **95**, 230-238 (2009).
13. Solanky, K.S., Burton, I.W., MacKinnon, S.L., Walter, J.A. & Dacanay, A. Metabolic changes in Atlantic salmon exposed to *Aeromonas salmonicida* detected by H-1-nuclear magnetic resonance spectroscopy of plasma. *Dis Aquat Organ* **65**, 107-114 (2005).
14. Viant, M.R. Metabolomics of aquatic organisms: the new 'omics' on the block. *Mar Ecol Prog Ser* **332**, 301-306 (2007).
15. Viant, M.R., Pincetich, C.A. & Eerderna, R.S.T. Metabolic effects of dinoseb, diazinon and esfenvalerate in eyed eggs and alevins of Chinook salmon (*Oncorhynchus tshawytscha*) determined by H-1 NMR metabolomics. *Aquat Toxicol* **77**, 359-371 (2006).
16. Kullgren, A., Samuelsson, L.M., Larsson, D.G.J., Bjornsson, B.T. & Bergman, E.J. A metabolomics approach to elucidate effects of food deprivation in juvenile rainbow trout (*Oncorhynchus mykiss*). *Am J Physiol-Reg I* **299**, R1440-R1448 (2010).
17. Baumgarner, B.L. & Cooper, B.R. Evaluation of a tandem gas chromatography/time-of-flight mass spectrometry metabolomics platform as a single method to investigate the effect of starvation on whole-animal metabolism in rainbow trout (*Oncorhynchus mykiss*). *J Exp Biol* **215**, 1627-1632 (2012).
18. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U.S.A.* **100**, 9440-9445 (2003).
19. Ruyter, B., Rosjo, C., Einen, O. & Thomassen, M.S. Essential fatty acids in Atlantic salmon: effects of increasing dietary doses of n-6 and n-3 fatty acids on growth, survival and fatty acid composition of liver, blood and carcass. *Aquacult Nutr* **6**, 119-127 (2000).
20. Bar, N. Physiological and hormonal changes during prolonged starvation in fish. *Can J Fish Aquat Sci* **71**, 1447-1458 (2014).

21. Doucett, R.R., Booth, R.K., Power, G. & McKinley, R.S. Effects of the spawning migration on the nutritional status of anadromous Atlantic salmon (*Salmo salar*): insights from stable-isotope analysis. *Can J Fish Aquat Sci* **56**, 2172-2180 (1999).
22. Ballantyne, J.S. et al. Plasma nonesterified fatty acid profiles in male and female sockeye salmon, *Oncorhynchus nerka*, during the spawning migration. *Can J Fish Aquat Sci* **53**, 1418-1426 (1996).
23. Pottinger, T.G., Rand-Weaver, M. & Sumpter, J.P. Overwinter fasting and re-feeding in rainbow trout: plasma growth hormone and cortisol levels in relation to energy mobilisation. *Comp Biochem Phys B* **136**, 403-417 (2003).
24. Jezierska, B., Hazel, J.R. & Gerking, S.D. Lipid Mobilization during Starvation in the Rainbow-Trout, *Salmo-Gairdneri* Richardson, with Attention to Fatty-Acids. *Journal of fish biology* **21**, 681-692 (1982).
25. Tocher, D.R. Metabolism and functions of lipids and fatty acids in teleost fish. *Rev Fish Sci* **11**, 107-184 (2003).
26. Jamal, S. et al. Relationship between functional activity and protein stability in the presence of all classes of stabilizing osmolytes. *Febs J* **276**, 6024-6032 (2009).
27. Fenili, D., Ma, K. & McLaurin, J. scyllo-Inositol: A Potential Therapeutic for Alzheimer's Disease. *Rsc Drug Discov*, 94-116 (2010).
28. Ma, K.R., Thomason, L.A.M. & McLaurin, J. scyllo-Inositol, Preclinical, and Clinical Data for Alzheimer's Disease. *Adv Pharmacol* **64**, 177-212 (2012).
29. Yancey, P.H., Blake, W.R. & Conley, J. Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants. *Comp Biochem Phys A* **133**, 667-676 (2002).
30. Yancey, P.H., Rhea, M.D., Kemp, K.M. & Bailey, D.M. Trimethylamine oxide, betaine and other osmolytes in deep-sea animals: Depth trends and effects on enzymes under hydrostatic pressure. *Cell Mol Biol* **50**, 371-376 (2004).
31. Gualano, A.B. et al. Branched-chain amino acids supplementation enhances exercise capacity and lipid oxidation during endurance exercise after muscle glycogen depletion. *J Sport Med Phys Fit* **51**, 82-88 (2011).

32. Love, R.M. The Chemical Biology of Fishes. *Academic Press, New York* (1980).
33. Walton, M.J. & Cowey, C.B. Aspects of Intermediary Metabolism in Salmonid Fish. *Comp Biochem Phys B* **73**, 59-79 (1982).
34. Mommsen, T.P. Salmon spawning migration and muscle protein metabolism: the August Krogh principle at work. *Comp Biochem Phys B* **139**, 383-400 (2004).
35. Mommsen, T.P., French, C.J. & Hochachka, P.W. Sites and Patterns of Protein and Amino-Acid Utilization during the Spawning Migration of Salmon. *Can J Zool* **58**, 1785-1799 (1980).
36. Vermeersch, K.A. & Styczynski, M.P. Applications of metabolomics in cancer research. *Journal of Carcinogenesis* **12**, 9 (2013).
37. Treutler, H. et al. Discovering Regulated Metabolite Families in Untargeted Metabolomics Studies. *Analytical Chemistry* **88**, 8082-8090 (2016).
38. Kind, T. et al. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Analytical Chemistry* **81**, 10038-10048 (2009).
39. Wei, X.L. et al. MetPP: a computational platform for comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics. *Bioinformatics* **29**, 1786-1792 (2013).
40. Dunn, W.B. et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* **6**, 1060-1083 (2011).
41. Xia, J.G., Mandal, R., Sinelnikov, I.V., Broadhurst, D. & Wishart, D.S. MetaboAnalyst 2.0-a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* **40**, W127-W133 (2012).

Chapter 3: An *in vitro* reaction assay pipeline provides a framework for validation of putative regulatory interactions and a necessary increase in sensitivity range for *de novo* interaction discovery

The majority of this chapter is taken from “A Gas Chromatography/Mass Spectrometry-based enzyme activity assay pipeline for discovery and validation of putative regulatory interactions,” which has been submitted to *Biotechniques*.

3.1 Introduction

In the push toward more accurate and high-power mathematical models of metabolism, there is an increasingly evident need for the characterization of reversible regulatory interaction between metabolites and proteins; while technologies for collecting genomic, transcriptomic (and to a lesser extent, proteomic) data are mature and routine, regulatory data lags behind due to the transient, often weakly-binding nature of these interactions¹. Ongoing efforts to systematize discovery and characterization of protein-metabolite regulatory interactions have resulted in many iterations on assays to detect protein-metabolite binding, including pull-down assays^{2,3}, dialysis partitioning⁴, affinity-selection mass spectrometry^{5,6}, desorption electrospray ionization mass spectrometry⁷, and various computational methods for predicting such interactions⁸⁻¹¹. As these methods provide only binding pairs (or putative binding pairs, in the case of computational methods), all of these methods share the downstream requirement to validate and characterize the regulatory function of the (putative) binders via enzyme activity assays. While this step is rarely focused on in regulator discovery methods, it can actually be quite complex due to protein

diversity and varying considerations based on how the proteins are to be expressed (and usually purified) for use in the assays.

Enzymatic activity screening has previously been carried out through a number of mass spectrometry-based assays¹², which represents an improvement in generalizability (and thus, efficiency) over individualized, enzyme-specific¹³ assays. The increasing emphasis on developing methods for discovery and prediction of binding interactions indicates a growing need for an efficient, flexible framework for activity-based screens. The specialized assays with product-specific reporters classically used for intensive study of an individual enzyme are not practical for screening beyond one or two enzymes. Mass spectrometry provides a generalized platform for analysis of these reactivity screens, with analytical techniques utilizing various types of mass spectrometry often (though not always) coupled to chromatographic separation. MALDI-MS¹⁴⁻¹⁶ and ESI-MS¹⁷⁻²⁰ have both been previously used for inhibitor screening; these methods have the advantage of throughput and minimal sample alteration before analysis. Adding a chromatography step (most often liquid chromatography) to the analysis adds to the time needed to analyze each sample, but in exchange allows more complex mixtures to be accurately analyzed.

Interestingly, gas chromatography-mass spectrometry (GC-MS) has not often to our knowledge been used for such assays, even though it offers excellent chromatographic separation and “hard” (electron bombardment) ionization, which allows some otherwise-difficult conversions (e.g. various isomerizations) to be distinguished more readily than with other mass spectrometry methods. GC-MS also typically provides a higher level of

sensitivity per instrument expense than LC-MS, making it an accessible platform for many labs. We believe that this, in combination with its other analytical advantages, makes GC-MS an appropriate and particularly attractive platform for systematic analysis of enzymatic activity, despite requiring a more complex sample work-up procedure than various other methods.

In addition to validation and characterization of previously-predicted regulators, such versatile approaches to screen for regulatory activity may also be used for *de novo* regulator discovery, by testing not just putative regulators, but a wide swathe of potential regulators. This is important because current *in vitro* screening for protein-metabolite binding interaction is hampered as a discovery method by the sensitivity limits of these methods; where limits of detection have been reported or estimated in previous studies, they typically land in the tens of micromolar range for the dissociation constant^{3,21}. As known allosteric interactions range into K_{DS} as high as millimolar-scale, other methods must be employed to reliably screen for weaker-binding regulators.

In this work, we developed a medium-throughput, GC-MS-based *in vitro* assay pipeline to be used both for rapid validation of observed or predicted binding interactions, and for the potential direct discovery of new regulatory pairs in a manner that bypasses the difficulty of conducting binding assays with weak-affinity partners. Protein expression is discussed, as recommendations for how this should be done tend to vary widely (in both content and level of strictness) between labs. As it is common practice for proteins under *in vitro* study (particularly with regard to screening for binding interactions) to be expressed with

attached affinity tags which may be used for purification, immobilization, and detection, we discuss three possible formats for the reaction assay step and demonstrate that the most appropriate format can differ by protein, which should be taken into account in systematic studies. Finally, we present a proof-of-concept for a suggested iterative library screening via reaction assays. These recommendations as well as the overall scheme may easily be generalized to use with other organisms and expression systems.

3.2 Results and Discussion

3.2.1 Protein expression conditions

As it is impossible in most cases to purchase purified metabolic enzymes for a systems-scale study and what few might be commercially available are expensive, most studies of multiple enzymes must begin with in-house expression. Protein expression recommendations typically provide ranges of values for induction, including induction time and optical densities at 600 nm (OD_{600}) at which to induce, yet efficiency of expression (both per unit time and unit mass) become important as experiments are multiplexed. We investigated a range of induction densities spanning a typical mid-log recommendation (OD_{600} 0.7-1.2), as well as near saturation (OD_{600} ~3.5), and periods of 4, 9, and 16 hours.

Three *S. cerevisiae* enzymes of diverse chemistry were chosen for the purposes of pipeline development – chorismate mutase (ARO7), cytosine deaminase (FCY1), and aspartate aminotransferase (AAT2). Their reactions are summarized in Table 3.1.

Table 3.1: Summary of enzymes and reactions

Gene abbreviation	Enzyme	Reaction
ARO7	Chorismate mutase	chorismate \rightarrow prephenate
FCY1	Cytosine deaminase	cytosine + water \rightarrow uracil + ammonia
AAT2	Aspartate aminotransferase	L-aspartate + 2-oxoglutarate = oxaloacetate + L-glutamate

As seen in Figure 3.1, the expression patterns differ for the various proteins. Longer induction periods typically produced significantly more protein per mass than shorter ones, whereas induction density was generally less important over the range surveyed. The differences between proteins suggest that some degree of optimization for each different protein to be studied could provide modest improvements, but for most practical purposes this is not necessary.

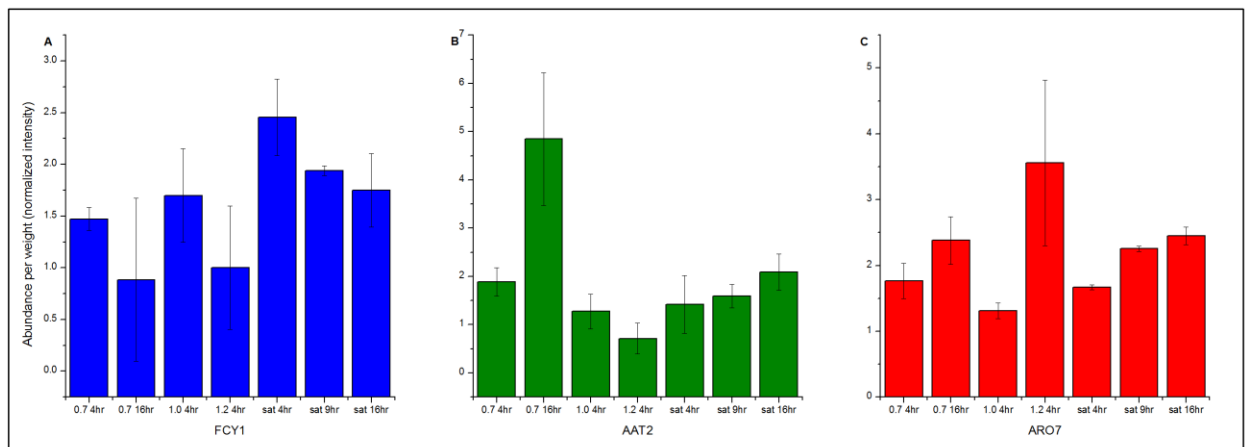


Figure 3.1: Relative quantification of expression of three enzymes under different expression protocols. Protein expression is most efficient when induction is carried out during log-phase growth, but precise induction density is minimally important. Values calculated per wet cell pellet mass; error bars represent the standard error of the mean, n=3.

While inducing at very high cell density is hardly optimal in terms of expression efficiency per cell mass and time, with an overnight induction period, the disadvantages of doing so are reasonably ameliorated by the logistical advantages, as growth timing can sometimes be a tricky proposition on a practical/logistical level. Of note, we show that “overgrown” cultures may still be induced to express an acceptable amount of the target protein with likely greater convenience.

3.2.2 Reaction assay architecture

In most research documentation dealing with regulator discovery, the reaction assay itself is frequently trivialized because it has previously been developed or is commercially available, but the in-lab reality is often one of tedious troubleshooting of individual proteins, which can drastically reduce the efficiency of the research and is directly counterproductive to attempts to systematize these discovery efforts. For broad studies of many enzymes, a middle ground is needed between intensive focus on individual enzymes and a single, “catch-all” method that may not in reality behave acceptably for many proteins. We therefore sought to establish a reaction assay format framework that can be varied according to specific needs of the research to be conducted as well as the respective sensitivities of the proteins involved, while still maintaining some degree of standardization: in some cases it may be reasonable to sacrifice protein purity or baseline activity for speed, while in others, it may be desirable or necessary to conduct reaction assays as soon after lysis as possible. In others still, large affinity tags on the proteins may interfere with folding or ligand binding, and thus require cleavage in order for the enzyme to become active. (When using expression systems other than the MORF library, which

includes a C-terminal ZZ-tag of ~17 kDa, any large affinity tag will pose the same potential issues.)

The three proteins used in this work reflect the diversity of requirements and permissible environments for protein function. ARO7 was robust; it retained activity following multiple different purification methods, allowed easy detection of activity even from crude lysates, and remained stable at 4 °C for several days. AAT2 was also robust in terms of activity, but the lysate matrix made measurement of activity unduly difficult in a lysate-based assay. FCY1, by contrast, was active in lysate, but lost all activity by the time it was purified and dialyzed. We identified that this loss in activity was due to FCY1's relative lack of stability in solution once extracted from the cell; within a few hours with or without dialysis, FCY1 lost almost all activity.

To address the limitations in FCY1's stability in anticipation of similar problems with other enzymes, we developed and validated an "on-bead" assay for enzyme activity wherein the reaction is carried out while the enzyme is still non-covalently attached to an affinity-based purification medium. This assay format allows a compromise between speed and protein purity; while more logistically involved than a lysate assay, it lacks the lengthy downstream steps of dialysis out of an elution buffer and/or cleavage of an affinity tag, providing essentially pure enzyme without a long incubation. It would be inappropriate for enzymes that require the removal of an attached affinity tag to avoid steric hindrances of function, and the agarose matrix may itself slow activity in some cases due to steric hindrance.

Nonetheless, for cases where a protein has poor stability and catalyzes a fast reaction but where a lysate assay is undesirable due to matrix issues, it is an appropriate choice.

Comparative results for all three test-case proteins are summarized in Table 3.2. ARO7 demonstrates the highest relative conversion in an on-bead assay. For FCY1, conversion is complete in both the lysate and on-bead conditions, but the on-bead assays showed high variability and much lower overall uracil abundances than the lysate condition, likely due to incomplete removal of the reaction mixture from the agarose bead matrix. No-substrate control samples of FCY1 lysate show a significant amount of natively-present uracil, but those amounts are orders of magnitude lower than in the substrate-added condition, rendering a lysate assay a reasonable and convenient option for that protein. Finally, AAT2's relatively complex lysate background (substantial native quantities of both glutamate and aspartate) and bi-molecular chemistry make the fully-purified structure decidedly preferable.

Table 3.2: Summary of enzyme behavior over differing reaction assay architecture, illustrating differences across enzymes

Protein	Lysate	On-bead	Fully purified
ARO7	Clear activity with no appreciable background	2x more conversion/time relative to lysate	Low but detectable activity
FCY1	Fast, complete conversion with simple background	Complete conversion but inconsistent detection	Minimal activity
AAT2	Activity but with complex, obfuscating background	No activity detected	Activity comparable to lysate, no background

Given these observations, we suggest the following for methodology selection:

1. Full purification is the default preferred option, given that it offers the fewest complications that may impede the protein's unregulated activity or make data analysis difficult. If the ZZ-tag interferes with protein folding, full purification and cleavage are strictly necessary, and ongoing reactions other than the one of interest can present an interpretation problem in lysate.
2. Lysate assays offer speed and convenience with potential sacrifice of data clarity, as their utility is limited to those reactions whose substrate(s) and product(s) are both easily detected against a cell lysate background and are not major participants in other post-lysis reactions.
3. On-bead assays offer a compromise between speed and clarity, but the large tags that remain on the enzymes may cause some hindrance of metabolite binding and/or protein subunit assembly, and metabolite absorption by the agarose matrix may also factor into reduced reactivity and detected compounds.

3.2.3 *Screening putative regulator libraries for inhibitory/activating activity*

While there are many potential applications for enzymatic activity assays, our main motivation behind the study and characterizations above was to implement an efficient pipeline for validating predicted (via *in vitro* or *in silico* techniques) regulatory interactions, and for screening libraries of metabolites for *de novo* regulator discovery.

For regulator validation, assuming a small number of hypotheses generated either via observation of binding interactions *in vitro*^{2-4,22} or *in silico* prediction^{8,9} this is a simple matter: one putative regulator is added to one reaction mixture, and the enzyme's activity in the presence of the compound is compared to its unregulated activity, subject to appropriate controls.

Using this methodology for *de novo* discovery or validation screening of a larger number of predicted potential regulators (as may be produced by some *in silico* modeling) requires iterative screening of a library followed by smaller subsets of that library to narrow down potential sources of regulatory activity. We used *S. cerevisiae* chorismate mutase (ARO7) as the test case to demonstrate this approach, as it has two well-studied^{23,24} allosteric regulators - it is inhibited by tyrosine and activated by tryptophan - that allow for positive controls in a hypothetical process of regulator library screening.

An initial screening library of ten compounds was selected for purposes of illustration, including tyrosine and tryptophan. Other compounds in the library were chosen from central carbon metabolism, nucleotide biosynthesis, and elsewhere in amino acid metabolism. The reasoning behind these selections was that amino acid metabolism and nucleotide biosynthesis pathways have a high degree of overlap, and various metabolites in glycolysis and the tricarboxylic acid (TCA) cycle are also used as feedstocks for secondary metabolism pathways; altogether, the library presents a logical assortment of metabolites that could be regulators based on their potential to indicate metabolic state.

The general scheme for this library screening is to conduct reaction assays with the full library, then divide the library repeatedly according to activity changes observed, as diagrammed in Figure 3.2A. Groups D, E, and F are representative possible sub-groupings from B and C for purposes of demonstration. In an actual discovery-oriented screening study, B and C would simply be arbitrarily divided in half/near-half without recombination, but in this case we included Group D to represent a case where two as-yet unknown regulators were allocated to the same sub-library. Figure 3.2B shows changes in activity due to library and sub-library aliquots added to the reaction mixtures. This data clearly narrows down the library of “putative” regulators to tryptophan as an activator; however the influence of tyrosine in this data is indistinct due to low unregulated activity. Depending on the kinetics of a given enzyme and amounts being used, it may be desirable to lengthen the reaction time or shorten it – here, tyrosine inhibition of ARO7 is indistinct due to a likely overly-short choice of reaction time, whereas with, for instance, FCY1, an activator would be difficult to distinguish *except* over a very short timeframe, as the unregulated reaction goes to completion in our standard protocol. While this adds another point of assay diversification/tailoring to individual enzymes, it is a minor and easily-implemented one.

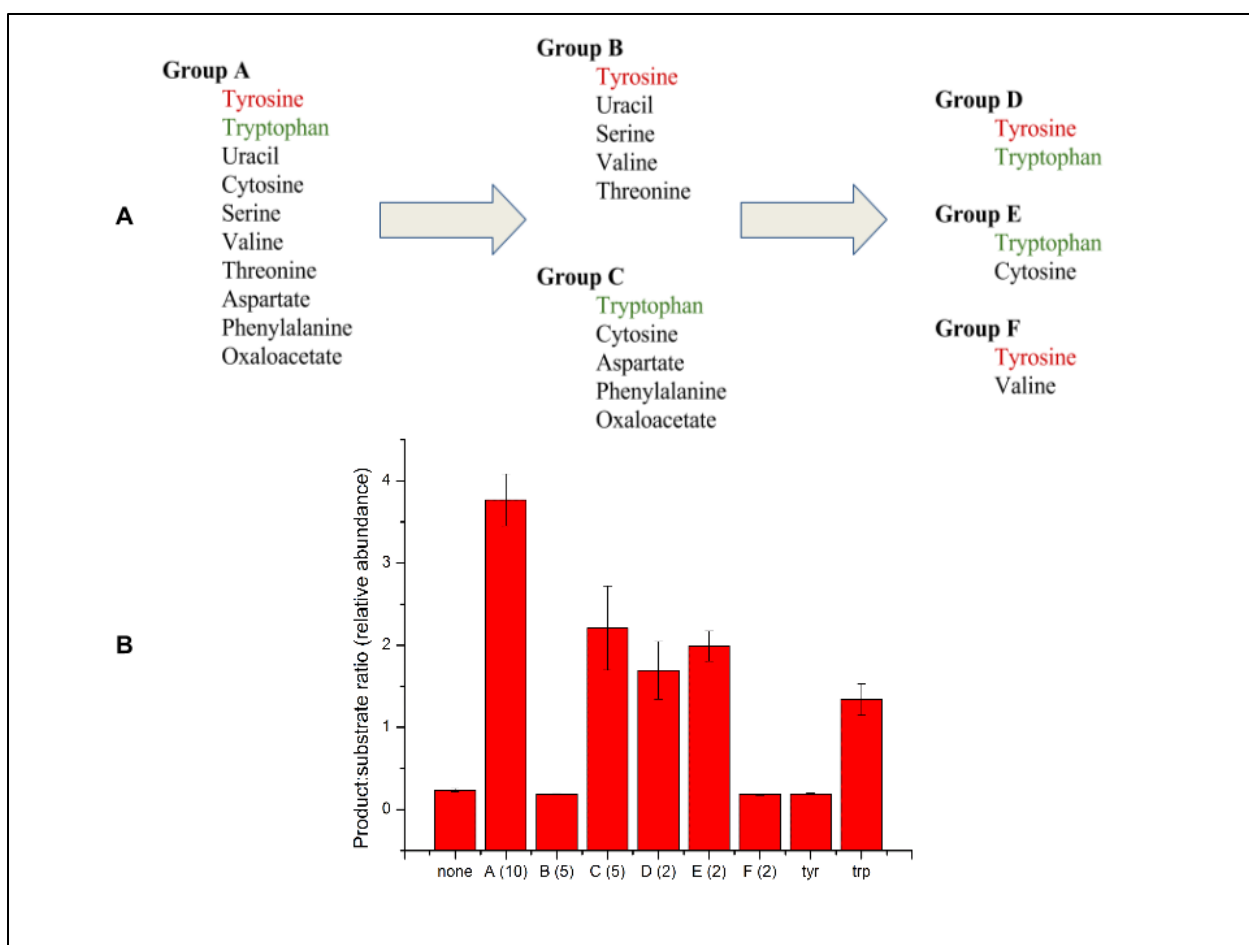


Figure 3.2: Iterative regulator screening demonstrated in ARO7, narrowing down from a metabolite library to the known regulators. Here tryptophan shows clear activation; inhibition by tyrosine is not significant, indicating a longer reaction time is appropriate. Activation in Group D demonstrates dominance of one allosteric regulator over another. A) Libraries used. Known activators and inhibitors written in green and red, respectively. B) Comparative activity of ARO7 exposed to different candidate libraries. Error bars represent the standard error of the mean; n=4.

The initial library (Group A)’s significant activity boost over its tryptophan-including subdivisions was investigated and determined to be due to A’s higher acidity, which is a known factor in chorismate mutase activity. As pH sensitivity of enzymes is not uncommon, we recommend measuring the pH of the libraries to be used and adjusting as appropriate, either for the specific protein if a sensitivity is known, or simply to a standard value.

In order to investigate the suitability of larger initial screening libraries, further reaction assays were carried out using ARO7 lysate and metabolite libraries of 15, 25, and 50 compounds at 1 mM. The components of these libraries were chosen based largely on availability, and are listed in Appendix B. The working hypothesis for this experiment was that at larger library sizes, the effects of ARO7's known regulator might be made indistinguishable due to matrix effects and/or competing presence of multiple other regulators.

Substrate and product peaks were easily distinguished over all conditions, and the results were surprising (Figure 3.3). Rather than cause a loss of observable regulation by tyrosine and tryptophan with larger libraries, use of the 25-compound and 50-compound libraries resulted in a dramatic decrease in chorismate mutase activity (and these conditions were also distinct from one another), while the results for the 15-metabolite libraries were statistically indistinguishable from the unregulated and tyrosine- or tryptophan-only conditions (though the tyrosine-only and tryptophan-only conditions were distinct from each other). The obfuscation of tyrosine and tryptophan's effects in the 15-compound libraries is tentatively explainable by the fact that the assays were conducted in lysate, and both of those amino acids participate in many other parts of metabolism. Tyrosine, in particular, was only barely detectable in the samples where it had been added as part of the library. The fact that the tyrosine- and tryptophan-containing versions of the 25- and 50-compound libraries did not produce results distinct from each other indicates that the regulatory effects of those two metabolites were indeed being obfuscated, but likely by the presence of a far-stronger inhibitor of ARO7 (possibly more than one, given that the 25-

and 50-compound conditions were distinct from one another as well as from the other conditions).

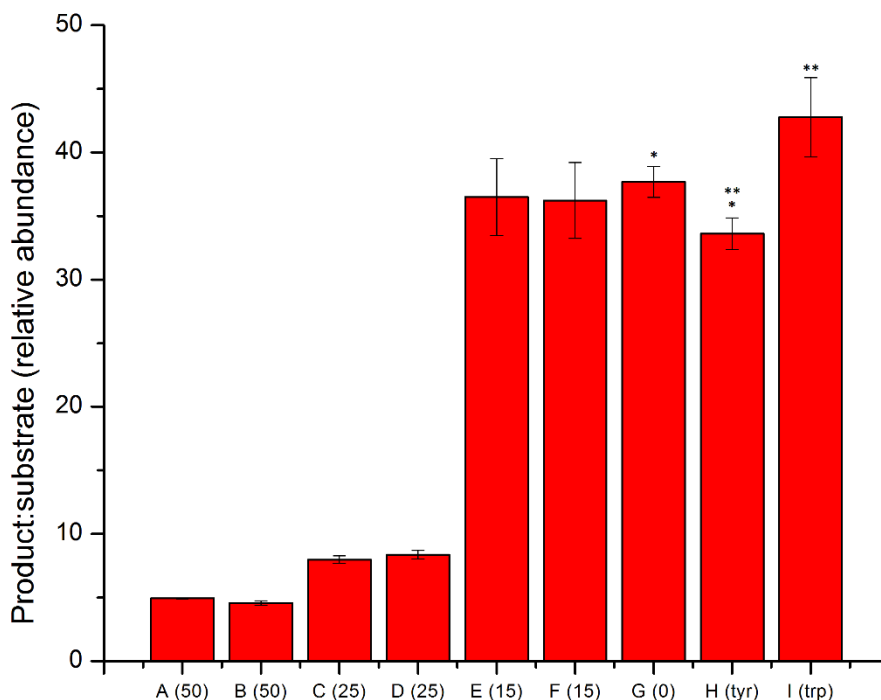


Figure 3.3: Extended library size comparison for ARO7 reactivity in lysate. Numbers in parentheses are the number of compounds in the library added to the reaction mixture for each condition. Every library size pair contains one of tyrosine or tryptophan (A, C, and E contain tyrosine as their final compound; B, D, and F contain tryptophan). Error bars represent the standard error of the mean. Asterisks indicate p-values via t-test: * $p < 0.04$; ** $p < 0.03$.

Taken together, these results do not suggest the presence of a generalizable upper limit to initial library size. Practical/logistical considerations seem likely to be a reasonable guide for initial library construction lacking other basis for decision. However, rather than construct a single initial screening library containing all the endogenous metabolites available to hand, it may be prudent to begin instead with two or three different libraries of intermediate (relative to the practical maximum determined by available resources) size, if a very large number of compounds are available. It of course remains a possibility with any given protein that competing regulators of similar strength might offset one another's

effects; as with any broadly-applied search strategy for putative regulators, some false negatives may be expected to occur.

Finally, the results summarized in Figure 3.3 indicate a clear avenue of future investigation regarding the apparent existence of one or more strong inhibitors of yeast chorismate mutase.

3.2.4 GC-MS method optimization for screening individual enzymes

Since this validation and discovery methodology is based on observed differences in enzymatic activity, good chromatographic resolution is only necessary around the product and substrate peaks. This allows for GC methods to be tailored to a given enzyme to reduce the time needed for analysis of a large set of screening reactions. For example, the substrate chorismate and product prephenate of chorismate mutase (ARO7) elute closely to each other, but do not co-elute. Thus we created a shortened screening method for chorismate mutase activity by only designating a slow temperature ramp (10°C/min) during the elution temperature window of chorismate and prephenate, and using a faster one outside of it. This approach decreased sample-to-sample time by ~40% relative to typical sample runs in our lab, which are designed for complex matrices. Runtime could be reduced even further from this initial optimization with additional effort.

3.3 Conclusions

Activity-based screening is both a necessary component for validation of previously-predicted regulatory interactions and a flexible platform to broaden the range of affinity strengths of regulatory interactions for which we can assay in medium-throughput. In this work, we have developed a set of guidelines and considerations for adapting a general mass spectrometry-based activity assay to more specific needs, and an iterative scheme for efficient screening for *de novo* regulator discovery. Our methodology provides a flexible schema for validating hypothesized regulators and pursuing *de novo* regulator discovery. As both a part of the ongoing screening studies for protein-metabolite binding pairs and a higher-sensitivity strategy orthogonal to them, we believe it can significantly speed regulator discovery efforts and thereby improve our understanding of metabolic regulation.

3.4 Materials and methods

3.4.1 Chemicals and media

Specialty yeast media ingredients (yeast nitrogen base, Synthetic Complete-uracil dropout amino acid mix) were purchased from Sunrise Scientific. Other media and buffer components were purchased from VWR Scientific. Chemical reagents for the reaction assays were purchased from Sigma-Aldrich.

3.4.2 Yeast culture and protein expression

The Moveable-ORF yeast protein library was purchased in yeast glycerol stock format from OpenBiosystems. Individual strains over-expressing ARO7, FCY1, and AAT2

respectively were streaked from stocks onto uracil-dropout SC-ura + 2% dextrose plates and incubated at 30 °C for two days. Single colonies were inoculated into 5 mL liquid SC-ura/2% dextrose and incubated in a rotating drum at 30 °C for 1-2 days, and the cultures subsequently used to make working glycerol stocks.

For preparative protein expression, appropriate MORF clones were grown in 5 mL liquid SC-ura/2% dextrose for 1-2 days to visual turbidity. From these cultures, 500 µL was used to inoculate 300 mL SC-ura/2% dextrose cultures in shake flasks, which were grown 16 hours in a 30 °C water bath, with shaking at 170 rpm.

Expression was induced at OD₆₀₀ 0.7-1.2 by media exchange (as the GAL promotor is inhibited by glucose): under sterile conditions, the 300 mL cultures were split into six 50 mL conical tubes and centrifuged at 3000 rpm for 3.5 minutes. The media was decanted from the pellets, which were then each resuspended by vortexing in 10 mL SC-ura/2% galactose/3% glycerol/2% ethanol expression media and recombined in their flasks, with media then added to 450 mL final volume.

The cultures were then returned to the shaker for 10-16 hours to allow protein expression, and harvested into two 50 mL conical tubes by centrifugation (3500 rpm, 3 minutes). Cell pellets were then either lysed immediately or stored at -80 °C.

3.4.3 *Expression assays*

To compare expression conditions for each protein of study, pre-expression cultures were grown as described above in 300 mL SC-ura/dextrose media. At OD₆₀₀ 0.7, 1.0, 1.2, and near-saturation, 5 mL samples were withdrawn and induced, at which point they were incubated at 30°C with rotation for either 4, 9, or 16 hours prior to harvesting by centrifugation for 3 minutes at 2739 xg and storage at -80 °C.

The cell pellets for each protein were lysed using the method described in the section below, scaled down to 200 µL each lysis buffer and beads.

Comparative expression was assayed via dot blot for each protein of interest. Briefly, 2 µL of each sample was spotted onto a nitrocellulose membrane and allowed to dry. The blots were blocked for at least 1 hour on a rocker plate in 5% bovine serum albumin (BSA) in PBST. The blocking solution was removed, and the blots hybridized for 1 hour on a rocker plate using anti-5xHis-Alexa647 (Qiagen), diluted 1:10,000 in PBST/1% BSA. The blots were then washed 3 times for 5 minutes each with PBST, followed by a 1-minute wash in PBS, and scanned for fluorescence at 647 nm.

Each individual condition was assayed in triplicate for each protein.

3.4.4 Cell lysis and protein purification

Preparative culture cell pellets (~0.8-1.2 g wet mass) were lysed in 700 µL PBST with 10 µL no-EDTA protease inhibitor cocktail (G Biosciences) and 1.5 mL 0.5 mm soda lime glass beads, by vortexing (3x 5-minute cycles, with a 1-minute quick-chill at -80 °C after

each cycle). The liquid was transferred to microcentrifuge tubes, and the beads washed twice with 700 μ L of the lysis buffer (also added to the tubes). The lysate was centrifuged for 3 minutes at 10,000 rpm to remove cell debris, and then either stored at -80 °C or used immediately.

The MORF library proteins have two tags which may be used for purification²⁵ – a large Protein A tag and a 6x-histidine epitope tag. For preparative purifications, we used immobilized metal ion affinity chromatography (IMAC) via HisPur Co²⁺ resin (Thermo Scientific). Crude lysates prepared as described above were diluted 1:1 with PBS + 10 mM imidazole, then added to 500 μ L cobalt resin (pre-washed with PBS + 10 mM imidazole) and incubated for 20 minutes. After 700 xg centrifugation and removal of the supernatant, the beads were washed 3x with 1 mL PBS + 10 mM imidazole. Purified proteins were eluted with 3 washes of 500 μ L PBS + 150 mM imidazole, then clarified via 0.22 μ m cellulose acetate spin filter. To remove the ZZ-tag (which is cleavable via a 3C protease cut site) and imidazole, half the pooled eluates were placed in mini-dialysis cartridges (GE Healthcare) with 4 μ g Turbo GST-3C protease (Accelagen), mixed, and dialyzed into PBS overnight at 4 °C. The unused half was saved for diagnostic analysis if necessary. The next day, the protease was removed by 30-minute incubation with 20 μ L glutathione-agarose resin (Thermo Scientific). The protein samples were then filtered at 0.22 μ m before use.

3.4.5 *Reaction assays*

Reaction assays were set up to assess relative enzyme activity as follows: 10 μ L of each substrate solution (10 mM in water) was added to a protein sample (in the form of crude lysate, bead-bound, and fully purified and tag-cleaved protein dialyzed into PBS – volumes 20 μ L, 40 μ L, and 15 μ L respectively such that each sample would be the same fraction of the overall volume at that stage) in a microcentrifuge tube, and incubated for one hour at 30 °C.

In the bead-bound condition, 40 μ L of 1:1 diluted lysate was added to 20 μ L cobalt resin and incubated on a rocker for 30 minutes. The samples were centrifuged for 1 minute at 700g and the supernatant removed, with two subsequent washes of 60 μ L PBS/10 mM imidazole, followed by one wash of only PBS. The beads were resuspended in 10 μ L PBS, with the substrate added and incubated as above. To separate the reaction mixture and the beads, the reactions were transferred to a 0.22 μ m cellulose acetate spin filter, and the clarified reaction sample removed for GC-MS analysis.

Reaction assays for library screens were carried out using 10 μ L ARO7 protein sample (in purified or lysate form), 10 μ L chorismate (10 or 20 mM in water), and 10 μ L metabolite library or single known regulator (all constituents at 1 mM in PBS) or PBS alone, mixed together in a microcentrifuge tube and incubated for 1 hour at 30 °C.

3.4.6 Deproteinization and GC-MS sample preparation

The reaction assay samples were prepared for GC-MS analysis by acetonitrile extraction and trimethylsilyl derivatization. Briefly, 3 sample volumes of ice-cold acetonitrile were

added to the samples, which were then mixed by vortexing and centrifuged at 20,000 $\times g$ for 3.5 minutes. The supernatant was pipetted off any solid material and transferred to clean tubes, then dried completely at 40 °C under vacuum in a Labconco CentriVap.

The dried samples were derivatized per the procedures of Fiehn, et al²⁶ (90 minutes in 10 μ L 40 mg/mL methoxyamine in pyridine, with agitation, at 30°, followed by addition of 90 μ L MSTFA + 1%TMCS (Thermo Scientific) and further incubation for 30 minutes at 37° with agitation) prior to GC-MS analysis.

3.4.7 GC-MS

All samples were analyzed on an Agilent 7890B gas chromatograph with 5977A quadrupole MSD, using 1 μ L splitless injection onto a HP-5MS (phenyl methyl silox) column, with a temperature ramp from 45°C to 310°C.

3.4.8 Reaction assay data analysis

Peak deconvolution and integration was performed in AMDIS 2.71²⁷, with subsequent alignment using SpectConnect²⁸, both with default parameters. Protein activity was assessed by calculating the ratio of a characteristic product peak to a characteristic substrate peak. For lysate-based samples, additional samples of lysate with no substrate were prepared, to account for any detectable amount of substrates and products already present in the lysate.

3.5 References

1. Lindsley, J.E. & Rutter, J. Whence cometh the allosterome? *Proc Natl Acad Sci U.S.A.* **103**, 10533-10535 (2006).
2. Tagore, R., Thomas, H.R., Homan, E.A., Munawar, A. & Saghatelian, A. A Global Metabolite Profiling Approach to Identify Protein-Metabolite Interactions. *Journal of the American Chemical Society* **130**, 14111-+ (2008).
3. Li, X.Y., Gianoulis, T.A., Yip, K.Y., Gerstein, M. & Snyder, M. Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **143**, 639-650 (2010).
4. Orsak, T. et al. Revealing the Allosterome: Systematic Identification of Metabolite-Protein Interactions. *Biochemistry* **51**, 225-232 (2012).
5. Annis, D.A., Nazef, N., Chuang, C.C., Scott, M.P. & Nash, H.M. A general technique to rank protein-ligand binding affinities and determine allosteric versus direct binding site competition in compound mixtures. *Journal of the American Chemical Society* **126**, 15495-15503 (2004).
6. Whitehurst, C.E. & Annis, D.A. Affinity selection-mass spectrometry and its emerging application to the high throughput screening of G protein-coupled receptors. *Combinatorial Chemistry & High Throughput Screening* **11**, 427-438 (2008).
7. Liu, P.Y., Zhang, J., Ferguson, C.N., Chen, H. & Loo, J.A. Measuring Protein-Ligand Interactions Using Liquid Sample Desorption Electrospray Ionization Mass Spectrometry. *Analytical Chemistry* **85**, 11966-11972 (2013).
8. Demerdash, O.N.A., Daily, M.D. & Mitchell, J.C. Structure-Based Predictive Models for Allosteric Hot Spots. *PLoS Computational Biology* **5** (2009).
9. Kruger, F.A. & Overington, J.P. Global Analysis of Small Molecule Binding to Related Protein Targets. *PLoS Computational Biology* **8** (2012).

10. Collier, G. & Ortiz, V. Emerging computational approaches for the study of protein allostery. *Archives of Biochemistry and Biophysics* **538**, 6-15 (2013).
11. Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nature Biotechnology* **31**, 357-+ (2013).
12. Liesener, A. & Karst, U. Monitoring enzymatic conversions by mass spectrometry: a critical review. *Analytical and Bioanalytical Chemistry* **382**, 1451-1464 (2005).
13. Bonowski, F. et al. Computer Controlled Automated Assay for Comprehensive Studies of Enzyme Kinetic Parameters. *PLoS One* **5** (2010).
14. Bothner, B. et al. Monitoring enzyme catalysis with mass spectrometry. *Journal of Biological Chemistry* **275**, 13455-13459 (2000).
15. Greis, K.D. et al. MALDI-TOF MS as a label-free approach to rapid inhibitor screening. *Journal of the American Society for Mass Spectrometry* **17**, 815-822 (2006).
16. Smith, A.M.E., Awuah, E., Capretta, A. & Brennan, J.D. A matrix-assisted laser desorption/ionization tandem mass spectrometry method for direct screening of small molecule mixtures against an aminoglycoside kinase. *Analytica Chimica Acta* **786**, 103-110 (2013).
17. Takayama, S. et al. Chemoenzymatic preparation of novel cyclic imine sugars and rapid biological activity evaluation using electrospray mass spectrometry and kinetic analysis. *Journal of the American Chemical Society* **119**, 8146-8151 (1997).
18. Wu, J.Y., Takayama, S., Wong, C.H. & Siuzdak, G. Quantitative electrospray mass spectrometry for the rapid assay of enzyme inhibitors. *Chemistry & Biology* **4**, 653-657 (1997).
19. Norris, A.J., Whitelegge, J.P., Faull, K.F. & Toyokuni, T. Kinetic characterization of enzyme inhibitors using electrospray-ionization mass spectrometry coupled with multiple reaction monitoring. *Analytical Chemistry* **73**, 6024-6029 (2001).
20. Hu, F.L., Zhang, H.Y., Lin, H.Q., Deng, C.H. & Zhang, X.M. Enzyme inhibitor screening by electrospray mass spectrometry with immobilized enzyme on

- magnetic silica microspheres. *Journal of the American Society for Mass Spectrometry* **19**, 865-873 (2008).
21. Bradner, J.E. et al. A robust small-molecule microarray platform for screening cell lysates. *Chemistry & Biology* **13**, 493-504 (2006).
 22. Gao, H. & Leary, J.A. Multiplex inhibitor screening and kinetic constant determinations for yeast hexokinase using mass spectrometry based assays. *Journal of the American Society for Mass Spectrometry* **14**, 173-181 (2003).
 23. Schmidheini, T., Sperisen, P., Paravicini, G., Hutter, R. & Braus, G. A Single Point Mutation Results in a Constitutively Activated and Feedback-Resistant Chorismate Mutase of *Saccharomyces cerevisiae*. *Journal of Bacteriology* **171**, 1245-1253 (1989).
 24. Schnappauf, G., Krappmann, S. & Braus, G.H. Tyrosine and tryptophan act through the same binding site at the dimer interface of yeast chorismate mutase. *Journal of Biological Chemistry* **273**, 17012-17017 (1998).
 25. Gelperin, D.M. et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development* **19**, 2816-2826 (2005).
 26. Kind, T. et al. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Analytical Chemistry* **81**, 10038-10048 (2009).
 27. Stein, S.E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry* **10**, 770-781 (1999).
 28. Styczynski, M.P. et al. Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Analytical Chemistry* **79**, 966-973 (2007).

Chapter 4: Small-molecule microarrays (SMMs) present a potential platform for high-throughput binding interaction discovery, albeit with unique challenges

4.1 Introduction

In order to construct systems-level mathematical models of metabolism, broad-scale characterization of regulatory interactions is necessary. As discussed previously, high-throughput discovery strategies for protein-metabolite interactions have not yet been effectively implemented - time and labor requirements for regulatory interaction discovery and characterization have remained high compared to other areas of investigation for which the analytical technology is mature.

Microarrays are a massively-parallel platform for probing binding interactions, with a genesis in probing for gene expression levels^{1,2}. While the microarray principle has since been applied to proteins³, RNA⁴, and small molecules⁵⁻⁷ in addition to DNA oligos, the basic concept remains the same: solutions of molecules to be tested are printed in micron-scale spots with a known configuration on glass slides. The slides may be functionalized in some manner, which may be covalent or noncovalent⁶, to improve the binding behavior of the printed molecules. A prepared array may then be probed with a solution of the putative binding partner(s). Binding detection is typically accomplished by fluorophore labeling.

Scalability is the primary incentive for microarray analysis; microarrays of all types share the advantage of extremely low resource use per interaction probed. Whereas an in vitro

binding assay requires amounts of solution on the order of one milliliter per sample, an entire batch of microarrays requires only microliter amounts. Furthermore, in the case of SMMs, the protein needed to probe a single array is $< 10\ \mu\text{g}$, as opposed to much higher amounts (on the order of hundreds of micrograms) used in a milliliter-scale in vitro experiment, or the milligram amounts needed for some previously-developed assays⁸. Arrays on a glass substrate also present the advantage of comparatively low nonspecific binding tendency for hybridization materials versus porous blotting substrates¹.

In the context of protein-metabolite interaction discovery, SMMs provide the capability to test a given protein against hundreds of metabolites simultaneously, with all steps of the procedure including the final scan and data analysis conducted in parallel. If successfully implemented, this would therefore provide an orders-of-magnitude scale increase of assays per time over mass spectrometry-based methods.

Small-molecule microarrays as used in this research make use of reactive isocyanate-functionalized glass slides. Isocyanate reacts with most functional groups in biological molecules (e.g. thiols, alcohols, amines, indoles, carboxylic acids, phenols⁶), resulting in covalent attachment of the metabolites to the glass substrate. A flexible polyethylene glycol linker sits between the isocyanate and the glass, allowing for increased conformational freedom for the printed metabolites (Figure 4.1). Once prepared, the arrays are probed with protein displaying a histidine affinity tag, and a secondary hybridization with fluorophore-labeled anti-His antibody provides the means to detect the bound protein⁵ (Figure 4.2).

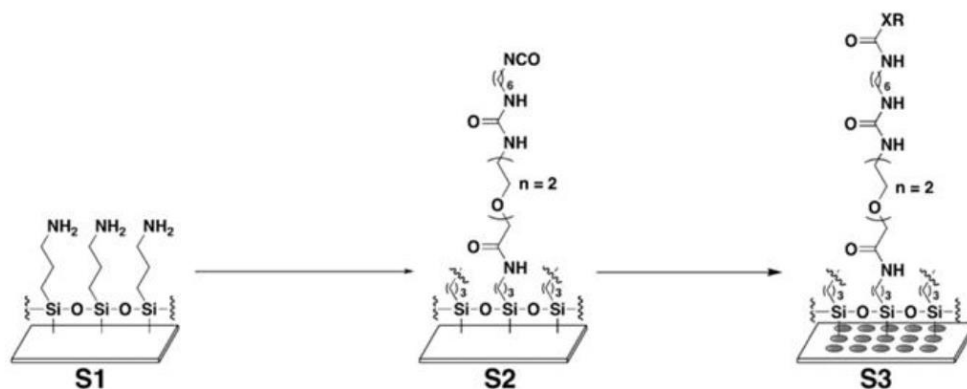


Figure 4.1: Metabolite attachment chemistry. A flexible PEG linker and reactive isocyanate group are attached to amine-functionalized glass slides. The metabolites are then printed onto the slides, where certain functional groups can react with the isocyanate moiety to form covalent bond. Figure adapted from Bradner, et al, 2006.⁶

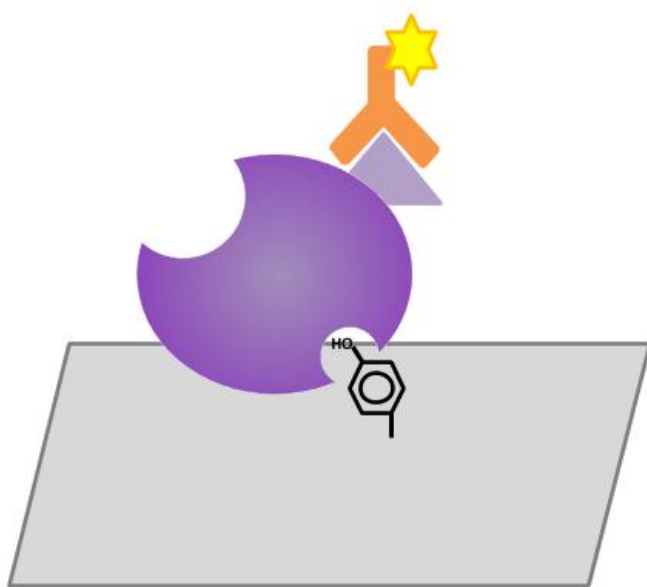


Figure 4.2: Simplified diagram of a double-hybridized SMM. The protein binds to the printed molecule in the primary hybridization. The secondary hybridization step attaches a fluorescently-labeled antibody to the 6xHis tag, allowing detection of the bound protein.

This SMM methodology was originally developed with high-throughput screening of compounds for drug discovery in mind. We sought to adapt this method to create a *Saccharomyces cerevisiae* “endogenous metabolome microarray” to screen for endogenous protein-metabolite interactions. This chapter will discuss early protocol

optimization efforts using high-affinity positive protein-metabolite pairs and explore the additional challenges posed by applying SMMs to endogenous interaction discovery.

4.2 Results and Discussion

4.2.1 Initial validation and optimization of basic protocol

As the construction and hybridization of SMMs is a complex procedure with several major points of failure, we carried out sequential validation of the following: array synthesis and covalent printing, primary hybridization, secondary hybridization.

Array synthesis with subsequent covalent attachment of a small molecule was validated by printing the fluorescent dye rhodamine B onto an active array and scanning for fluorescence after downstream workup. Primary hybridization was demonstrated on the streptavidin-biotin binding interaction by printing biotin and hybridizing with fluor-labeled streptavidin. Secondary hybridization was initially demonstrated using 6x-histidine-tagged streptavidin and fluorophore-labeled anti-5xHis antibody, and further shown in the same manner with the weaker-binding pairs FKBP12-rapamycin and FKBP12-FK506. We were thus able to replicate detection of positive control interactions suggested in the original protocol (Figure 4.3).

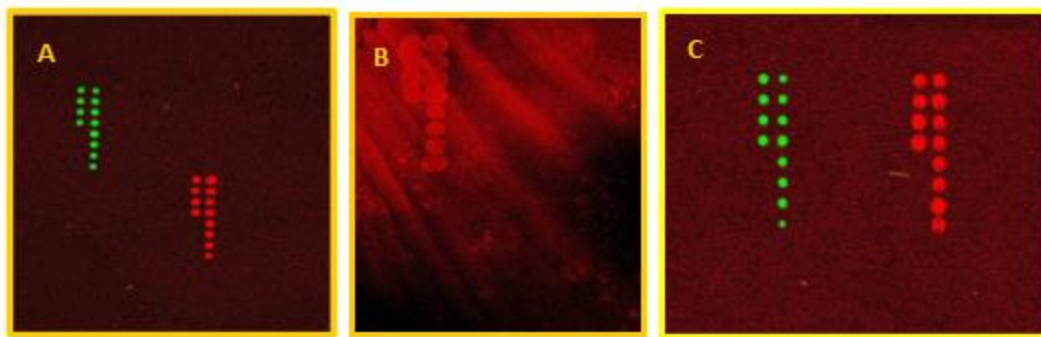


Figure 4.3: Progressive validation of basic arraying and hybridization protocols. A. Rhodamine B (green) and primary hybridization of fluor-tagged streptavidin to printed biotin (red). B. First successful secondary hybridization; 6xHis-tagged streptavidin bound to printed biotin, with detection via secondary hybridization using anti-5xHis-Alexa647. C. Rhodamine B (green) and clean primary-secondary hybridization of 6xHis-FKBP12 bound to printed rapamycin and detected via anti-5xHis-Alexa647 (red).

Print quality (namely, metabolite spot size and consistency) was an issue encountered and explored at various points throughout the SMM work (Figure 4.4). The specific needs of a lab making use of SMMs will vary with instrumentation; however, it is worth noting on a general level that spot quality was heavily dependent on the following factors:

1. Solvent used for the metabolite to be printed. This is a particularly relevant concern for a broad-spectrum application to endogenous metabolites, as their solubility profiles vary widely. With our instrument and pins, water produced the smallest and most regular spots, followed by dimethyl sulfoxide (DMSO), followed by dimethylformamide (DMF). Neither ethanol nor 50:50 ethanol/water could be printed at an acceptable spot density.
2. Microarraying pin design. All work presented here was completed using channeled silicon wafer microarraying pins; a different, “partially-etched channel” pin model from the same manufacturer resulted in smaller spot size (which, on a practical level, increases flexibility in solvents and printing density).

3. Microarraying pin condition and cleanliness. Most microarray printers are equipped with a wash station, but designing a protocol appropriate to the pins and solutions being used is still nontrivial. We found it necessary to remove from the printhead and manually inspect the pins (using a microscope) on a regular basis to ensure that they remained in good condition and spot quality, consistent.

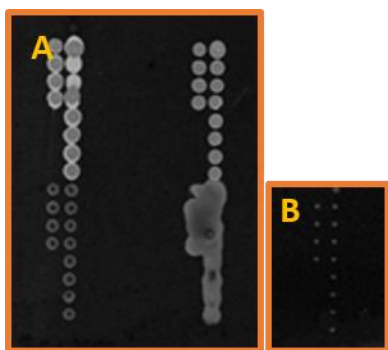


Figure 4.4: Spot morphology comparison, showing FKBP12 bound to printed rapamycin. A. Variation of spot morphology within a single print run. Each cluster of twelve spots was made by one pin; solvent was DMSO. A pin being slightly dirty or damaged results in the spots running together during printing. B. Rapamycin printed in 50/50 DMSO/water and spotted with partially-etched channel pins.

Unsurprisingly, specific details of the hybridization and wash procedures had a strong impact on the quality of the final array scan. Gentle agitation on a rocker plate during all incubation steps was preferable to the protein-conserving “Parafilm method” for primary hybridization (which involved pipetting a ten-fold smaller volume than used for a dish incubation onto a piece of Parafilm and placing the array face-down on the droplet such that it covered the array surface), but the latter gave detectable spots even if high or inconsistent background. However, the weaker the binding pair to be detected, the higher the hybridization quality generally required; such a higher-background method is therefore unlikely to be suitable in screens for endogenous protein-metabolite interactions, as these affinities may be many orders of magnitude weaker than the positive control cases

presented. (In the original target application of the SMM methodology adopted for this work, which was a chemical biology search for drug candidates, detecting low-affinity interactions was generally undesirable anyway, allowing a higher tolerance in parts of the protocol that affect final array quality.)

4.2.2 Sensitivity limitations

In the originally-published protocol, the reported detectability limit is a dissociation constant of 50 μM ⁵. Many, perhaps the majority, of known allosteric interactions are near (within an order of magnitude) that limit, or even higher. As previously discussed, *S. cerevisiae* chorismate mutase (ARO7) presented a useful positive control for endogenous interaction detection, as its two known regulators, tyrosine and tryptophan, have affinities of 50 μM and 1.5 μM , respectively.

4.2.3 Photocrosslinking strategy for increasing primary hybridization sensitivity

Repeated lack of success in detecting either of these interactions on an array in addition to our understanding of other allosteric interaction affinities led us to pursue a means of improving the sensitivity of the assay. Our hypothesis was that the bound metabolite was being washed off of its binding partner during the subsequent antibody hybridization and multiple washes, such that the protein must remain bound through those steps in order for the interaction to be detected. If dissociation of a protein-metabolite binding pair could be prevented after the initial capture, there would be potential for dramatic improvement in the assay's practical limit of detection, albeit at the likely cost of an increased frequency of background signal or false positive results.

We therefore explored the use of a photo-inducible crosslinker as a means of “fixing” the primary hybridization in place prior to proceeding with the antibody hybridization. Succinimidyl 6-(4,4'-azipentanamido)hexanoate (NHS-LC-Diazirine, or LC-SDA; Figure 4.5) is a heterobifunctional crosslinker with an amine-reactive moiety (NHS ester) and a nonspecifically-reactive, UV-inducible moiety (diazirine), which are connected by a spacer arm. The NHS ester hydrolyzes readily in solution, leaving behind a carboxyl group that is compatible with the isocyanate attachment chemistry in SMM production, allowing it to be mixed into the metabolite solutions to be printed and resulting in mixed metabolite-crosslinker spots. After primary hybridization of the array, exposure to UV light prior to removing the primary hybridization solution activates the diazirine moiety, forming a carbene intermediate which then reacts nonspecifically with an amino acid side-chain or peptide backbone of the nearby protein. This fixes nearby protein covalently to the spot, which prevents that protein from being washed off in subsequent hybridization and wash steps in the array workup.

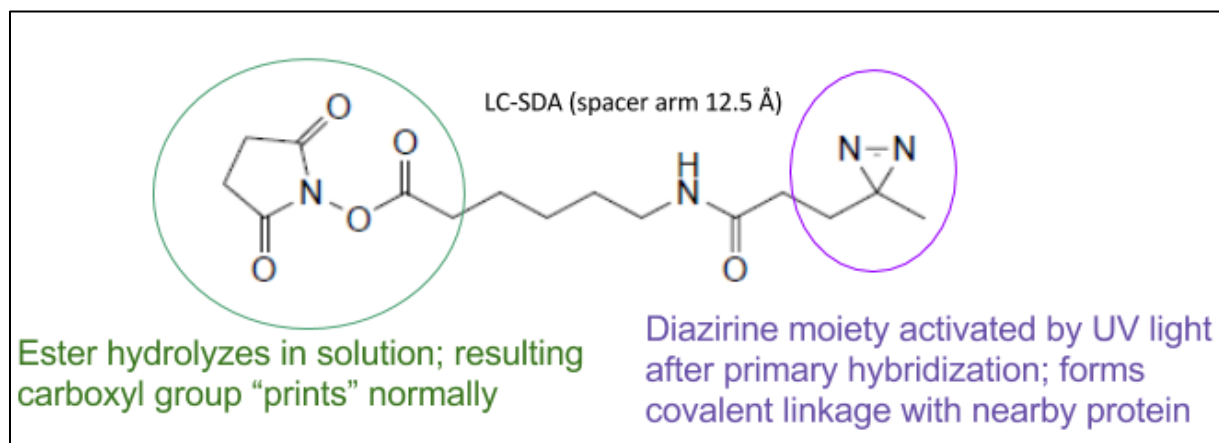


Figure 4.5: LC-SDA heterobifunctional crosslinker.

We conducted initial validation of this concept's chemistry by crosslinking rhodamine B dye or green fluorescent protein (GFP) to IgG-agarose beads, and by printing LC-SDA crosslinker only (not mixed with a metabolite) onto arrays at various concentrations followed by a typical two-hybridization array workup with an arbitrary tagged protein (FKBP12-6xHis). The crosslinker behaved as expected in the IgG bead/fluorophore crosslinking tests. When printed onto a microarray at a concentration of 1 mM, a detectable amount of protein was nonspecifically captured during the UV exposure, indicating that this capture scheme can work on an array (Figure 4.6). At LC-SDA print concentration of 0.1 mM or less, nonspecific capture was not detectable, indicating that mixing small amounts of it with metabolites to be printed should mitigate the increase in false positives. (For comparison, FKBP12-rapamycin was easily detected when rapamycin was printed at a concentration of 2 μ M.)

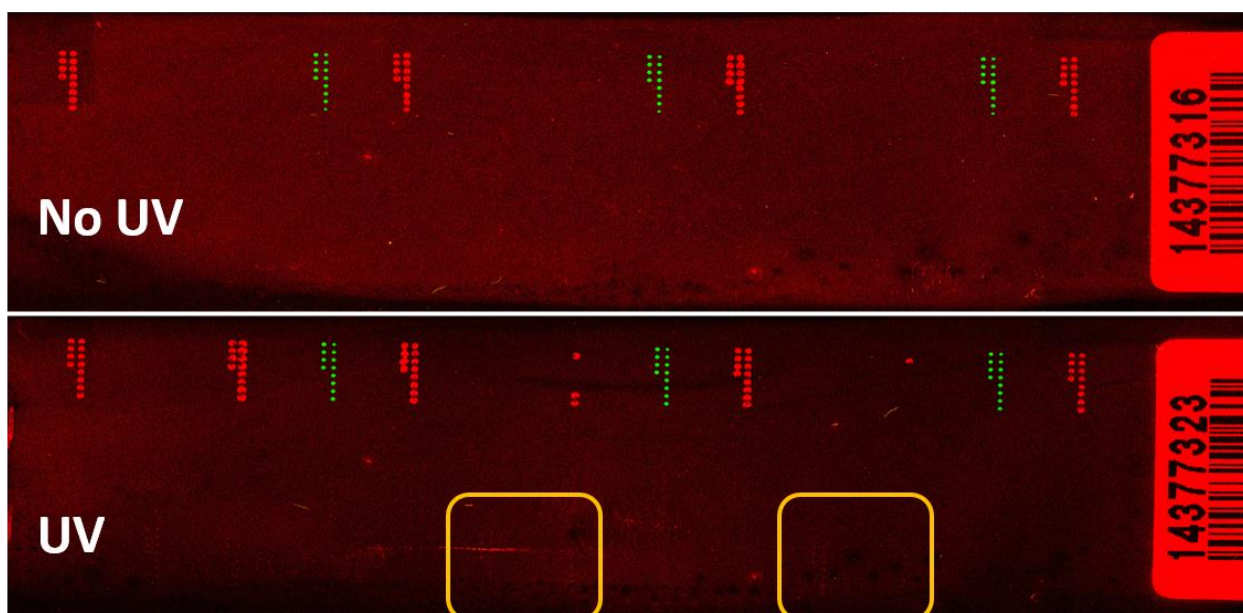


Figure 4.6: Detectable non-specific capture of FKBP12 with 1 mM printed LC-SDA crosslinker when treated with UV (boxed slide segments on bottom row). The top row of each slide was printed with rhodamine B and rapamycin. (Differences in the top row spots were determined to be due to pin malfunction.)

Attempts to demonstrate the intended sensitivity increase, however, were unsuccessful, albeit due to the lack of an appropriate test case. The interaction between FKBP12 and rapamycin is strong ($K_D = 0.5$ nM) enough that there was no appreciable attenuation of signal with increasing length and number of post-hybridization washes. Incubation with a solution of 8 M urea + 100 mM beta-mercaptoethanol resulted in an overall reduction of signal, but did not succeed in completely stripping slides of bound FKBP12 with or without crosslinking; we therefore concluded that any sensitivity advantage conferred by the crosslinking step would not be observable using FKBP12-rapamycin. Attempts to apply the method to ARO7 still did not allow detection of its known interactions.

4.2.4 Destruction of binding functionality in small printed metabolites

Repeated validation of ARO7's correct folding and functionality (see Chapter 3) via reaction assay led to the hypothesis that the attachment of its regulators to a microarray destroyed the moieties required for proper binding to the protein. Tyrosine and tryptophan (181.2 g/mol and 204.2 g/mol, respectively) are quite small in comparison to rapamycin (914.2 g/mol), and have fewer functional groups available to react with the isocyanate to form a bond to the array, resulting in a higher likelihood that their ability to bind to ARO7 may be destroyed by array attachment.

Evidence supporting this hypothesis was gathered by conducting a reaction/regulation assay of ARO7 with tyrosine, tryptophan, *N*-acetyl-tyrosine, and *N*-acetyl-tryptophan. Unsubstituted tyrosine and tryptophan demonstrated the expected regulatory behavior,

while the *N*-acetylated amino acids did not, indicating a strong likelihood that ARO7's interactions with tyrosine and tryptophan are unable to be shown using this array methodology.

4.3 Conclusions

Given general knowledge of structural trends for metabolites in central carbon and amino acid metabolism (small) and binding affinity trends for known interactions (weak), the results described above indicate that SMMs may not be appropriate tools for screening for binding interactions of small metabolites if they are uniformly attached to the substrate via a moiety that is likely to be involved in protein binding. While some false negatives are acceptable and expected in such an assay, since the small endogenous metabolites we are interested in studying have few potential array linkage points for a moiety-specific attachment chemistry (even one with multiple potential attachment moieties such as the isocyanate chemistry), the false negative rate will be increased to a level we expect to be impractical. Photocrosslinking, however, may still provide a reasonable avenue for increasing the assay's sensitivity: attaching metabolites in a non-specific manner to the array in a way that allows for some fraction of the attachments to be non-destructive to their native binding interactions could circumvent these issues.

4.4 Materials and Methods

4.4.1 Small-molecule microarray synthesis

SMMs were synthesized following the protocol described by Bradner, *et al.*⁵ Briefly, amino-functionalized glass slides (Corning GAPS II) were completely submerged in a PEG

solution (1 mM fmoc-8-amino-3,6-dioxaoctanoic acid, 2 mM PyBOP, 0.5 mM DIPEA in DMF) and incubated at room temperature with stirring overnight (or at least 4 hours). The slides were briefly rinsed with DMF, then submerged in 1% vol/vol piperidine in DMF for 10 minutes, before being washed in DMF for 1 minute with stirring, then submerged in 1% 1,6-diisocyanatohexane in DMF with stirring for 30 minutes to install the reactive isocyanate group.

The slides were then washed twice with DMF, 3 minutes with stirring, then a third time with THF for 2 minutes before being allowed to fully air dry and transferred to the print bed of a Genetix QArray2 microarray printer equipped with a Parallel Synthesis silicon wafer print head and silicon wafer arraying pins. Metabolite solutions were printed at various concentrations in water, DMSO, or DMF and the spots allowed to completely air dry. Rhodamine B was additionally printed as a positional reference point and an indicator that the array synthesis was successful. The printed slides were placed in a vacuum desiccator chamber and exposed to pyridine vapor overnight.

The next day, the slides were removed from the chamber and immersed in 5% ethylene glycol and 0.1% pyridine in DMF for 30 minutes with stirring, followed by a rinse in DMF and a 1 hour, stirred wash in DMF and two subsequent 3 minute washes in THF. They were allowed to air dry, then either used immediately or packaged into five-slide boxes sealed with Parafilm and stored at -20 °C in a secondary container with desiccant for future use.

4.4.2 Array hybridization and analysis

6xHis-tagged proteins of interest were diluted to a concentration of 0.1-5.0 µg/mL in phosphate-buffered saline plus 0.1% Tween 20 (PBST) at 4 °C. Microarrays were covered in protein solution and incubated for 1 hour on a rocking platform. The protein solution was then removed and the slides briefly rinsed in chilled PBST, followed by 1 hour of incubation on a rocking platform with anti-5xHis-Alexa647 antibody (diluted 1:1000 in PBST + 0.1% BSA). The antibody solution was removed and the slides rinsed with chilled PBST, followed by 3 washes in PBST each 2 minutes with agitation. The arrays were washed once more in chilled PBS for 2 minutes, then dried by centrifugation and scanned for fluorescence on a GSI Lumonix microarray scanner.

4.4.3 Addition and characterization of photocrosslinking step

LC-SDA (Thermo Scientific) was prepared as a 10 mM stock solution in DMSO immediately prior to use, then diluted either into water or pre-made metabolite solutions to the desired concentrations. When photocrosslinking was desired during a hybridization, before removing the primary hybridization solution, the arrays were exposed to long-wave UV light (365 nm from an 8-watt lamp, at a distance of ~1 cm) for 20 minutes to allow the diazirine reaction to proceed. Secondary hybridization and washes were then carried out as described above.

4.4.4 Regulatory assay of chorismate mutase using N-acetylated regulators

ARO7 was expressed from the MORF library⁹ as discussed in Chapter 2 and purified via immobilized metal ion chromatography, followed by cleavage of its ZZ-tag using 3C

protease and dialysis into PBS. 30 μ L ARO7 at \sim 30 μ g/mL was added to 15 μ L of 10 mM chorismate and 15 μ L of 1 mM of the appropriate amino acid, gently mixed, then incubated for 1 hour at 30 °C. The samples were deproteinized using acetonitrile and analyzed via GC-MS as described in Chapter 2.

4.5 References

1. Shalon, D., Smith, S.J. & Brown, P.O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639-645 (1996).
2. Brown, P.O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* **21**, 33-37 (1999).
3. Zhu, H. et al. Global analysis of protein activities using proteome chips. *Science* **293**, 2101-2105 (2001).
4. Collett, J.R. et al. Functional RNA microarrays for high-throughput screening of antiprotein aptamers. *Anal Biochem* **338**, 113-123 (2005).
5. Bradner, J.E., McPherson, O.M. & Koehler, A.N. A method for the covalent capture and screening of diverse small molecules in a microarray format. *Nat Protoc* **1**, 2344-2352 (2006).
6. Bradner, J.E. et al. A robust small-molecule microarray platform for screening cell lysates. *Chemistry & Biology* **13**, 493-504 (2006).
7. Vegas, A.J., Fuller, J.H. & Koehler, A.N. Small-molecule microarrays as tools in ligand discovery. *Chem Soc Rev* **37**, 1385-1394 (2008).
8. Orsak, T. et al. Revealing the Allosterome: Systematic Identification of Metabolite-Protein Interactions. *Biochemistry* **51**, 225-232 (2012).
9. Gelperin, D.M. et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development* **19**, 2816-2826 (2005).

Chapter 5: Conclusions and recommendations for further study

5.1 Conclusions

The work presented here explored the methodological requirements and realities of characterizing protein-metabolite regulatory interactions and of conducting other systems-biological studies. Ultimately, we contextualize previous work with new data and present a flexible-frame approach that addresses an aspect of regulatory characterization which is often inappropriately treated as an afterthought or presumed to already exist, as well as offering a discovery strategy that does not suffer from sensitivity or practical constraints in the manner seen for some other popular approaches.

5.1.1 *High-level insights from untargeted metabolomics¹*

Metabolomics analysis of serum from inappetent, wild salmon versus fed, domesticated salmon during spawning did not seem to indicate that the inappetence of the wild fish resulted in a pathological state. ω -9 free fatty acids were found to be accumulated in the inappetent group, contrasting the behavior of ω -3 fatty acids, which is a surprisingly specific structural point of divergence. Accumulation of sugar alcohols was also observed in the inappetent group, which nothing currently known about salmonid metabolism can suggest an explanation. These differences suggest potential value in the pursuit of related studies, which may ultimately aid in developing a firmer understanding of salmonid physiology and response to differing types of stress, and offer context and guidance for studying regulation at the protein-metabolite level.

5.1.2 *A pipeline framework for medium-throughput investigation of enzyme regulation via direct reaction assay*

Working with a published *Saccharomyces cerevisiae* protein library², we developed a pipeline framework for use of *in vitro* reaction assays in validation/characterization of hypothesized regulatory interactions as well as *de novo* discovery of such interactions. Three enzymes with differing chemistries were selected for analysis in order to represent a part of protein diversity and assist in development of an experimental framework with the flexibility to handle that diversity.

Given the large variation in both content and strictness of adherence to protocols for protein expression in literature, we conducted induction and expression time course assays for the three proteins of study. While “efficiency” of expression (protein quantity versus overall biomass) was reduced in cultures that were allowed to proceed out of log-phase growth prior to induction, in general the differences observed across the range of the experiment indicated that the primary factor in the expression stage of such work can be convenience, rather than adhering to a strict window for induction.

Different enzymes frequently have different requirements and sensitivities with regard to handling, and these differences can be problematic for studying many proteins in parallel. Enzyme reactivity was compared for the three proteins using a lysate preparation, fully-purified protein, and an “on-bead” approach developed to provide a compromise between the benefits and drawbacks of the other two methods. Each of the proteins investigated

had a different ideal assay structure due to differences in stability and the nature of the reactions themselves.

Finally, we presented an iterative library-screening approach for regulator validation or discovery, demonstrating how one would efficiently narrow down from an arbitrary metabolite library to a regulator of the enzyme using ARO7 and its known regulators.

5.1.3 Challenges and potential of small-molecule microarrays for interaction discovery

Small-molecule microarrays³⁻⁵ are a “protein-first” method for identifying binding interactions between a protein and a potentially very large library of metabolites simultaneously, which we worked to adapt to the context of endogenous interaction discovery. By using known (exogenous) binding pairs as positive controls, we were able to demonstrate sequential validation of the assay’s components, and identified guidelines for reducing background to suit the needs of using the assay primarily to screen for interactions on the weaker end of its detection range.

As sensitivity was a substantial concern with regard to using the assay to screen for endogenous interactions, we attempted to implement a photocrosslinking strategy for “fixing” captured protein in place after the primary hybridization. Final validation of the strategy was problematic due to a lack of appropriate positive controls, but it remains a potentially viable strategy for extending the sensitivity range of a small-molecule microarray assay.

To test the hypothesis that the particularly small size of the metabolites we intended to begin working with (including tyrosine and tryptophan, whose interactions with ARO7 we were attempting to use as a positive control), we compared regulatory activity of N-acetylated tyrosine and tryptophan to the unmodified amino acids via direct reaction assay. The lack of regulation by the substituted amino acids is evidence that SMMs may present more difficulties in searching for interactions in central metabolism than initially expected.

5.2 Relevance

Systems biology is fundamentally integrative in nature, taking information from multiple techniques aimed at various levels of metabolism. At the highest level, the work presented in this thesis makes a practical argument in favor of a consciously multi-angled approach to studying metabolite-level regulation within one research group or project, as opposed to a search for the One True Platform.

This work produced novel findings regarding salmonid metabolism, and how it may be affected by and/or tied to environmental factors. The conclusion that inappetence likely does *not* involve a pathological state provides contrast to previous studies with other organisms, in turn highlighting future potential avenues of investigation that may uncover more physio-biochemical differences and increase understanding of all species involved. Furthermore, the results of the salmon inappetence study lay long-term groundwork for the potential pursuit of identifying metabolic biomarkers in one or more animals that can speak to the health of the ecosystem as a whole.

The methodologically-focused investigation of protein-metabolite regulatory interactions in this work via reaction assays and small-molecule microarrays directly contributes to the available toolbox of methods in the field of systems biology. It also presents a critical evaluation of the practical realities of conducting these discovery efforts in a systematic manner, and strategic recommendations for the field moving forward.

In order to conduct an effective, broad-range investigation of protein-metabolite interactions in an organism, it is important to have not only an effective set of assays, but a framework for handling large numbers of diverse proteins in the context of these assays. In previously-existing work on discovery methodology for protein-metabolite interactions, the protein-handling and validation steps are downplayed, with the focus placed squarely on assay development. This work directly addresses those practical concerns using a widely-available protein library for a widely-studied, important model eukaryote, ultimately presenting a flexible pipeline that will be applicable to many future discovery efforts.

As previously discussed, the wide variety of assays previously developed to identify binding interactions each have unique strengths and weaknesses that in turn are affected by the high level of variation observed in strengths and steric requirements of binding interactions.⁶ Very weak interactions present a particularly challenging problem for most of these strategies. In some cases these strategies' effective scopes are also limited by the need to chemically alter the binding partners in some way. Both of these limitations are explored further in this work. This highlights a gap in the available methodologies in the

context of medium-to-high-throughput studies, to which this work offers a compromise solution in the form of a medium-throughput direct reaction assay strategy. This iterative library-screening method requires no alteration of either the proteins or the putative regulators, and its sensitivity is not limited by the need to directly detect metabolite binding, as no matter how weakly a regulator binds, if it does in fact regulate enzymatic activity, that is reflected in the assay result.

Furthermore, this work takes steps to adapt a previously-established, potentially time- and resource-efficient, high-throughput detection strategy⁴ for protein-small molecule interactions in chemical biology to screen instead for endogenous binding interactions. Small-molecule microarrays have been previously shown to be an effective platform for identification of potential drug targets and leads. In this work, we successfully replicated the technique and identified elements of the protocol that are particularly critical for the detection of interactions near the sensitivity limit of the assay. We also developed a novel potential means of increasing the effective sensitivity of the assay, which, with further testing, may greatly facilitate the practical implementation of this technique for endogenous interaction discovery in the future. Finally, we investigated the hypothesis that SMMs may not be an appropriate screening tool for a significant proportion of metabolism due to the likelihood of interfering with protein-metabolite binding when the metabolite in question is particularly small and has few functional groups that can attach it to the array. The failure of chorismate mutase to be regulated by acetylated derivatives of its known regulators is evidence in favor of this hypothesis, which should further inform future development efforts with regard to this type of assay.

5.3 Recommendations for future study

5.3.1 *Small-molecule microarray assay development*

Despite the limitations identified in this work, small-molecule microarrays still offer a high degree of potential utility in regulator screening efforts if applied to an appropriate area of metabolism, and if sensitivity is improved. Successful implementation of the technique across even a few classes of metabolites would allow for efficient identification of putative regulators in those sectors of metabolism for any species as long as correctly-folded protein can somehow be isolated.

5.3.1.1 Additional reaction assays with substituted regulators

The experiment presented in Chapter 4 wherein yeast chorismate mutase was shown not to be regulated by *N*-acetylated versions of its two known regulators offers evidence that the problem of hindrance due to array attachment might be of significant detriment to the assay's effectiveness for very small metabolites. Further investigation of SMM potential for endogenous interaction discovery should include repeating similar versions of that experiment with as many other known enzyme-regulator pairs as is practical. (N-acetylated derivatives are available for many amino acids, and present a reasonable model for an array-attached amino acid, in particular, due to the fact that primary amines are among the most reactive functional groups with the isocyanate moiety used in the array attachment chemistry.) These results will give much greater insight into the assay's limits of applicability.

5.3.1.2 Validation of photocrosslinking strategy for sensitivity improvement

The roadblock reached in this work that caused us to shift focus away from development of the photocrosslinking technique was the lack of a conveniently-available, appropriately weakly-binding test case: no sensitivity increase could be observed when photocrosslinking capture was implemented with FKBP12-rapamycin binding, but this was ultimately unsurprising given the difficulty in stripping bound FKBP12 even under relatively extreme conditions. Identification of an appropriate test case(s) will allow validation of the technique. Proteins with known allosteric regulators in the micromolar affinity range (as the estimated sensitivity limit for these SMMs at present is on the order of $10\ \mu\text{M}$)³ should be tested with those regulators on an array; those whose binding is successfully detected to any degree present potentially appropriate candidates for photocrosslinking validation.

When these candidate pairs are established, a series of hybridizations with different levels of wash stringency should be carried out using those pairs. Any cases where signal can be easily attenuated by increasing the number or stringency of post-hybridization washes should then be used as test case pairs. Hybridizations should then be carried out with photocrosslinking as described in section 4.4.3 and the fluorescence signal measured for various wash stringencies to characterize the effect of photocrosslinking on assay sensitivity.

5.3.2 Integration of untargeted yeast metabolomics and regulator screening

Chapter 2 discusses the results of an untargeted metabolomics study in Atlantic salmon and points out the potential contextual relevance of such a study to regulator discovery and characterization. Given that the remainder of the experimental work in this thesis was completed using yeast proteins and that the lab expects to continue studies with yeast, it would be highly appropriate to conduct yeast metabolomics studies, subjecting *S. cerevisiae* to environmental perturbations and using observed differences in metabolite abundances to guide a search for regulators making use of the other methods described in this thesis.

5.3.2.1 Exploratory yeast metabolomics analysis

S. cerevisiae cultures may be easily grown side-by-side, then subjected to differing environmental conditions at a given point in their growth and sampled for metabolomics analysis (via cold methanol quenching)⁷ both before and after the perturbation. Specific perturbations that would be an appropriate starting point for an integrated study of regulation would be removal of key nutrients from the growth medium (via exchange into a dropout medium) or a change of carbon source. Since protein-metabolite level regulation occurs on a fast timescale relative to other levels of metabolic regulation, samples should be collected immediately after the perturbation is applied, and once or twice more at short time intervals. After all sampling is complete, quenched samples would be subjected to hot ethanol extraction⁷ followed by GC-MS analysis in the manner described in Chapter 2.

5.3.2.2 Guided regulator search via iterative library screening

Our work with Atlantic salmon resulted in the identification of a small number of metabolites of different classes that were accumulated in inappetent fish. With yeast in a nutrient deprivation experiment, the pathways involving accumulated metabolites would be similarly of interest. Enzymes in such pathways would then be initial targets for library-based reaction assay screening for regulator discovery. Also, specifically, the results described in section 3.2.3 should be further pursued to investigate the possible existence of an as-yet unknown ARO7 regulator.

5.3.3 Downstream mathematical modeling and bioinformatics analysis informed by regulator discovery

Both the regulation data and the high-level metabolomics could subsequently be integrated into ongoing efforts within the lab to create dynamic mathematical models of yeast metabolism. This would likely not be the work of one student, but a collaborative effort among a computational student pursuing yeast metabolic modeling strategies, an experimentalist investigating yeast metabolomics, and another with a focus on protein-metabolite level regulatory interactions. The latter researcher need not be working with yeast metabolism alone, as the discovery strategies discussed in this work are directly generalizable to any enzyme that can be practically overexpressed and/or purified, and work in phylogenetically-near homologues would potentially confer an additional, very exciting benefit: identification and sequence comparison of very close homologues with differing regulatory profiles could allow researchers to derive structural insight on protein-metabolite level regulatory interactions.

5.4 References

1. Cipriano, R.C., Smith, M.L., Vermeersch, K.A., Dove, A.D.M. & Styczynski, M.P. Differential metabolite levels in response to spawning-induced inappetence in Atlantic salmon *Salmo salar*. *Comp Biochem Phys D* **13**, 52-59 (2015).
2. Gelperin, D.M. et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes & Development* **19**, 2816-2826 (2005).
3. Bradner, J.E. et al. A robust small-molecule microarray platform for screening cell lysates. *Chemistry & Biology* **13**, 493-504 (2006).
4. Bradner, J.E., McPherson, O.M. & Koehler, A.N. A method for the covalent capture and screening of diverse small molecules in a microarray format. *Nat Protoc* **1**, 2344-2352 (2006).
5. Vegas, A.J., Fuller, J.H. & Koehler, A.N. Small-molecule microarrays as tools in ligand discovery. *Chem Soc Rev* **37**, 1385-1394 (2008).
6. McFedries, A., Schwaid, A. & Saghatelian, A. Methods for the Elucidation of Protein-Small Molecule Interactions. *Chemistry & Biology* **20**, 667-673 (2013).
7. Loret, M.O., Pedersen, L. & Francois, J. Revised procedures for yeast metabolites extraction: application to a glucose pulse to carbon-limited yeast cultures, which reveals a transient activation of the purine salvage pathway. *Yeast* **24**, 47-60 (2007).

APPENDIX A

Detailed GCxGC-MS Methods

AS Method

An Agilent 7683 autosampler was used. Three pre-washes with pyridine were performed before each injection. The sample was then pumped into the syringe 4 times. The syringe size was 10 μ L with a sample volume of 1 μ L injected into the inlet. Three post-washes with pyridine were performed after injection.

GC Method

An Agilent 7890 gas chromatograph adapted to GCxGC analysis was used. The first column was an HP-5 30m x 0.320 mmID x 0.25 μ m and the second was Rtx-200 2m x 0.180 mmID x 0.20 μ m. The excluded masses in auto mass defect mode option was chosen. Helium was used as the carrier gas with a corrected constant flowrate of 1.00 mL/min. An inlet septum purge flow of 3mL/min was chosen. The inlet was operated in splitless mode with a purge flow of 100 mL/min set to start 30 seconds after injection, yielding a total flow of 101 mL/min. Gas saver mode was used, with a flow of 20 mL/min set to start a minute after injection.

The oven temperature program started at 50 °C for one minute and ramped to 295 °C over 28 minutes followed by a 4.4 minute hold at 295 °C, with the secondary oven kept at an offset of 15 °C and the modulator temperature offset kept at 35 °C from the main oven.

Modulator times are presented in Table A.1. An equilibration time of 60 seconds was set for the ovens. The transfer line was set to 320 °C for the entire run.

Table A.1: Modulation Timing

#	Start (s)	End (s)	Modulation period (s)	Hot pulse time (s)	Cold pulse time (s)
1	Start	1128	4.00	0.50	1.50
2	1128	1259.2	3.20	0.75	0.85
3	1259.2	1315.2	3.50	0.75	1.00

MS Method

A Leco Pegasus IVD time of flight mass spectrum (TOF-MS) was used. The total MS method time was based on the GC method time. The acquisition delay was set to 400 seconds, with the filaments being turned off until then. The collection mass range was from 50 to 500 u. The acquisition rate was set to 200 spectra/second. The detector voltage was set to 100 V above the optimized voltage with the electron energy set to -70 V. The mass defect mode was set to manual with the mass defect 0 mu/ 100 u. The ion source temperature was 220 °C and the run had to wait for the ion source temperatures to reach the set point before starting acquisition.

APPENDIX B

Metabolite Libraries

Table A.2: Metabolite library lists

15-compound	25-compound	50-compound
(Tyrosine or tryptophan)	(Tyrosine or tryptophan)	(Tyrosine or tryptophan)
		1-Methyl-L-histidine
		2'-Deoxyuridine
		2-Deoxyribose 5-phosphate
	4-hydroxybenzoate	4-Hydroxybenzoate
		4-Hydroxyphenylpyruvate
		5-Phospho-D-ribose 1-diphosphate
Adenine	Adenine	Adenine
		Adenosine-5-monophosphate
		Alanine
	Anthranilic acid	Anthranilic acid
Arginine	Arginine	Arginine
Asparagine	Asparagine	Asparagine
Aspartic acid	Aspartic acid	Aspartic acid
		Cysteine
Cytosine	Cytosine	Cytosine
		D-Glucuronic acid
	D-Mannose	D-Mannose
GABA	GABA	GABA
		Glutamine
Glycine	Glycine	Glycine
Guanine	Guanine	Guanine
		Guanosine
		Histidine
		Homoserine
		Isoleucine
	Kynurenic acid	Kynurenic acid
		Kynurenine
Methionine	Methionine	Methionine
	Myo-inositol	Myo-Inositol
	Ornithine HCl	Ornithine HCl
	Orotic acid	Orotic acid
	Oxalic acid	Oxalic acid
		Oxaloacetic acid

Table A.2 (continued)		
Phenylalanine	Phenylalanine	Phenylalanine
		Phenylpyruvate
		Phosphoenolpyruvate
		Proline
Serine	Serine	Serine
		Shikimic acid
		Threonine
		Thymidine
Thymine	Thymine	Thymine
Uracil	Uracil	Uracil
		Uridine
Valine	Valine	Valine
		Xanthosine dehydrate
		α -Ketoglutaric acid
	β -alanine	β -Alanine

APPENDIX C

Letter of permission for use of copyrighted material

Marlene Kraml <cbpjrnl.mkraml@gmail.com>

Tue 11/8, 9:12 AM

Dear McKenzie,

There is need for extra permission from the journal. As you have already done, we do ask that you clearly state that part(s) of the thesis were published in CBP and that you include the citation(s) in the appropriate chapters of your thesis (article title, journal title, date of publication and page numbers).

We wish you all the best,

Kind regards,

Marlene

Marlene Kraml <cbpjrnl.mkraml@gmail.com>

Yesterday, 8:07 AM

Dear McKenzie,

Sorry that wasn't very clear. My email constitutes permission. We only ask that you include the citations.

Kind regards,

Marlene