

Diss. ETH No. 23442

Nontargeted Metabolomics Approaches for Enzyme Discovery and High-Resolution Metabolome Phenotyping

Thesis submitted to attain the degree of
DOCTOR OF SCIENCES OF ETH ZÜRICH
(Dr. sc. ETH Zürich)

presented by
Daniel Charles Sévin
M.Sc. Ing. (FH), Hochschule Mannheim
born the 19th of February 1987 in Strasbourg
citizen of Germany and France

Accepted on the recommendation of
Prof. Dr. Uwe Sauer
Prof. Dr. Jörg Stelling
Dr. Nicola Zamboni
PD Dr. Stefan Pelzer

2016

Abstract

Metabolism is fundamental to all known organisms in transforming nutrients into building blocks and energy required to sustain life. Cellular metabolism consists of a network of thousands mostly enzyme-catalyzed reactions interconverting small chemical molecules, so-called metabolites. Decades of biochemistry research and, more recently, the development of genome sequencing technology, have led to the characterization of the metabolic networks of many organisms. Yet, in many organisms more than half of the genes have no known function assigned, and it is estimated that up to one third of these functionally uncharacterized genes may encode metabolic enzymes. Consequently, although the topology of metabolism is relatively well-understood, many missing enzymes may exist that remain unaccounted for in current models. These missing enzymes, together with an incomplete understanding of regulatory mechanisms, limit the ability of metabolic models to analyze and contextualize experimental data and reduce the accuracy of their predictions.

During this thesis, I developed a novel method for the systematic discovery of metabolic enzymes. It is based on the incubation of purified or overexpressed proteins in complex mixtures of metabolites, followed by the detection of catalytic events by nontargeted mass spectrometry. In **Chapter 3**, I describe the development and validation of this method and report its application to all functionally uncharacterized proteins of the model bacterium *Escherichia coli*, identifying over two hundred putative novel enzymes, twelve of which were followed up and functionally validated. During the method development stage, I noticed that many known enzymes seemed to catalyze additional reactions. In **Chapter 4**, I therefore systematically investigated all known metabolic enzymes of *E. coli* for additional catalytic capabilities using the developed technology. Indeed, for over four hundred enzymes novel substrates or products were observed, suggesting these enzymes to have so-called promiscuous activities. The subsequent successful functional validation of selected promiscuous enzymes corroborated the accuracy of these findings. The effect of such wide-spread promiscuity on metabolism was then tested using a constraint-based modeling approach, predicting that enzyme promiscuity may facilitate the evolution of *E. coli* to utilize novel nutrients and hence conquer new habitats.

Despite missing enzymes, the topology of the metabolic network is reasonably well understood, whereas much less is known about metabolic responses to genetic or environmental perturbations. Nontargeted metabolomics methods are instrumental to measure the concentrations of a large number of metabolites, thereby enabling a comprehensive and largely unbiased assessment of the metabolic phenotype. In **Chapters 5 and 6** of this thesis, I used nontargeted metabolomics to study the metabolic responses of fifteen diverse organisms to hyperosmotic salt stress as a model for complex environmental perturbations. Beyond unexpectedly global metabolic responses involving half of the covered metabolome, nontargeted metabolomics led to the identification of a novel osmoprotection mechanism in *E. coli*, namely the stabilization of its cytoplasmic membrane by accumulation of the isoprenoid and respiratory electron carrier ubiquinone-8.

Overall, this thesis demonstrates the potential of nontargeted metabolomics for large-scale and data-driven investigations in diverse fields of biological research.

Zusammenfassung

Stoffwechsel ist ein grundlegender Prozess in allen bekannten Lebewesen, in dessen Verlauf Nährstoffe in lebensnotwendige Bausteine und Energie umgewandelt werden. Der zelluläre Stoffwechsel besteht aus einem Netzwerk tausender meist enzymkatalysierter Reaktionen kleiner chemischer Moleküle, sogenannter Metaboliten. Jahrzehnte biochemischer Forschung und, in jüngerer Zeit, die Entwicklung von Genomsequenzierungstechnologien, haben zur Aufklärung der Stoffwechselnetzwerke vieler Lebewesen geführt. Jedoch ist in vielen Lebewesen die Funktion mehr als der Hälfte der Gene weiterhin unbekannt, und es ist anzunehmen, dass bis zu ein Drittel davon für Stoffwechselenzyme kodieren könnte. Obwohl die Struktur von Stoffwechselnetzwerken im Allgemeinen gut untersucht ist existiert daher möglicherweise eine Vielzahl bislang unbekannter Enzyme, denen derzeitige Modelle nicht Rechnung tragen. Diese fehlenden Enzyme, zusammen mit einem nur begrenzten Verständnis von Regulationsmechanismen, schränken die Fähigkeit solcher Modelle ein, experimentelle Daten zu interpretieren und verringern ihre Vorhersagegenauigkeit.

Im Rahmen dieser Dissertation wurde ein neuartiger Ansatz zur systematischen Entdeckung von Stoffwechselenzymen entwickelt. Dieser beruht auf der Inkubation aufgereinigter oder überexprimierter Proteine in einer Metabolitenmixtur, gefolgt von der Detektion katalytischer Ereignisse durch nicht-zielgerichtete Massenspektrometrie. In **Kapitel 3** wird die Entwicklung und Validierung dieser Methode beschrieben, sowie deren Anwendung auf alle funktionell nicht charakterisierten Proteine des Modellbakteriums *Escherichia coli* dargelegt. In dieser Studie wurden mehr als zweihundert mögliche neue Enzyme identifiziert, von denen zwölf näher untersucht und validiert wurden. Während der Methodenentwicklung war auffällig, dass viele bereits bekannte Enzyme scheinbar weitere Reaktionen katalysierten. In **Kapitel 4** wurden daher alle in *E. coli* bekannten Stoffwechselenzyme systematisch auf zusätzliche katalytische Aktivitäten hin untersucht. In der Tat wurden bei mehr als vierhundert Enzymen bislang unbekannte Substrate oder Produkte detektiert, welches auf deren sogenannte Promiskuität hindeutet. Die Zuverlässigkeit dieser Ergebnisse wurde durch erfolgreiche Validierung ausgewählter promiskuitiver Enzyme bestätigt. Schlussendlich wurde der Einfluss promiskuitiver Enzyme auf den gesamten Stoffwechsel mittels eines beschränkungsbasierten Modellierungsansatzes untersucht, welcher einen Beitrag von Enzympromiskuität zur evolutionären Weiterentwicklungsfähigkeit des Stoffwechsels hinsichtlich der Verwendung neuer Nährstoffe einhergehend mit der Kolonialisierung neuer Lebensräume durch *E. coli*/vorhersagte.

Trotz fehlender Enzyme ist die Struktur des Stoffwechselnetzwerkes üblicherweise hinreichend bekannt, wohingegen die Stoffwechselantwort auf genetische oder umweltbedingte Veränderungen wesentlich weniger gut verstanden ist. Nicht-zielgerichtete Metabolomikmethoden können eingesetzt werden um die Konzentration einer Vielzahl von Metaboliten zu messen, wodurch die umfassende und weitgehend unvoreingenommene Einschätzung des metabolischen Phänotyps ermöglicht wird. In **Kapiteln 5 und 6** dieser Dissertation wurden nicht-zielgerichtete Metabolomikmethoden genutzt um die Stoffwechselantworten fünfzehn unterschiedlicher Lebewesen auf hyperosmotischen Salzstress, ein Modellsystem für komplexe Umweltstörungen, zu untersuchen. Über unerwartet

weitläufige, die Hälfte des abgedeckten Metaboloms umfassende Stoffwechselantworten hinaus, führten diese Untersuchungen zur Entdeckung eines neuen Schutzmechanismus gegenüber osmotischem Stress in *E. coli*, nämlich die Stabilisierung dessen zytoplasmatischer Membran durch die Akkumulierung des isoprenoiden Elektronentransporters Ubiquinon-8.

Insgesamt veranschaulicht diese Dissertation das Potential von nicht-zielgerichteten Metabolomikmethoden für großangelegte und datengesteuerte Untersuchungen in verschiedenen biologischen Forschungsfeldern.

Résumé

Le métabolisme est fondamental pour tous les organismes connus en transformant les substances nutritives en composants essentiels et l'énergie nécessaire pour soutenir la vie. Le métabolisme cellulaire consiste d'un réseau de milliers de réactions chimiques entre des petites molécules appelées métabolites, la plupart catalysée par des enzymes. Des décennies de recherche biochimique et, récemment, le développement de méthodes de séquençage de génome, ont conduit à la caractérisation des réseaux métaboliques de nombreux organismes. Cependant, dans de nombreux organismes la fonction de plus de la moitié des gènes reste inconnue, et on estime que jusqu'à un tiers de ces gènes pourraient coder pour des enzymes métaboliques. En conséquence, même si la topologie des réseaux métaboliques est généralement bien connue, il est concevable qu'il existe un nombre important d'enzymes inconnues qui ne figurent pas dans les modèles actuels. Ces enzymes manquantes, ainsi que les mécanismes de régulation peu compris, limitent la capacité des modèles métaboliques d'analyser et contextualiser les données expérimentales et réduisent la précision de leurs prédictions.

Dans cette thèse, une nouvelle approche a été développée pour la découverte systématique d'enzymes métaboliques. Elle se base sur l'incubation d'enzymes purifiées ou surexprimées dans un mélange complexe de métabolites, suivi par la détection d'événements catalytiques par la spectrométrie de masse non-ciblée. Dans le **3^{ème} Chapitre**, le développement et la validation de cette méthode et son application à toutes les protéines de fonction inconnue dans la bactérie modèle *Escherichia coli* sont présentés. Cette étude menait à la découverte de plus de deux cents enzymes putatives dont douze étant caractérisées en plus de détail. En développant la méthode, nous nous sommes aperçus que plusieurs enzymes connues semblaient catalyser des réactions supplémentaires. Dans le **4^{ème} Chapitre**, nous avons étudié toutes enzymes métaboliques connues d'*E. coli* sur leur capacité de catalyser davantage de réactions. En effet, pour plus de quatre cents enzymes des nouveaux substrats ou produits ont été observés, indiquant que ces enzymes ont des activités promiscuitives. L'exactitude de ces résultats est soutenue par la confirmation fonctionnelle de certaines enzymes choisies. Finalement, l'effet des enzymes multifonctionnelles sur le metabolisme a été étudié en utilisant une approche de modélisation numérique par contrainte, prédisant que la promiscuité enzymatique pourrait faciliter l'évolution d'*E. coli* envers l'utilisation de nouvelles substances nutritives en colonisant de nouveaux habitats.

Malgré les enzymes manquantes, la topologie du réseau métabolique est suffisamment bien étudiée, tandis que la réponse métabolique aux perturbations génétiques ou environnementales est moins bien connue. Les méthodes de métabolomique non-ciblées sont instrumentales pour mesurer les concentrations d'un nombre important de métabolites, ainsi permettant un constat compréhensif et pour la plupart impartial du phénotype métabolique. Dans le **5^{ème} et 6^{ème} Chapitre**, nous utilisons la métabolomique non-ciblée pour étudier les réponses de quinze organismes divers au stress hyper-osmotique comme modèle pour les perturbations environnementales complexes. Au-delà des réponses métaboliques globales inattendues impliquant la moitié du métabolome couvert, la métabolomique non-ciblée a mené à la

découverte d'un nouveau mécanisme osmoprotectif dans *E. coli* : La stabilisation de sa membrane cytoplasmique par l'accumulation de la molécule isoprénioïde et porteur d'électrons ubiquinone-8.

Dans l'ensemble, cette thèse démontre le potentiel de la métabolomique non-ciblée pour des analyses à grande échelle et fondé sur les données dans des divers domaines de recherche biologique.