55

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Léa Chuzel

Application of functional metagenomics to the field of glycobiology

Application of functional metagenomics to the field of glycobiology

Dissertation
zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

von Dipl. Biotech. Léa Chuzel
geboren am 15. November 1993 in Lyon (Frankreich)

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik der Otto-von-Guericke-Universität Magdeburg

Promotionskommission:

Prof. Dr. rer. nat. habil. Weiß (Vorsizt)
Prof. Dr.-Ing. Udo Reichl (Gutachter)
Dr. Erdmann Rapp (Gutachter)
Dr. Christopher H. Taron (Gutachter)

eingereicht am: 30. Juni 2021

Promotionskolloquium am: 1. November 2021

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Band 55

Léa Chuzel

Application of functional metagenomics to the field of glycobiology

Shaker Verlag
Düren 2022

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Magdeburg, Univ., Diss., 2021

Printed in Germany.

ISBN 978-3-8440-8437-5
ISSN 1439-4804

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Abstract

Bacteria account for ~15% of the earth's biomass and are the second largest biomass contributor after plants. Bacteria are ubiquitous and have adapted to live in all habitable ecosystems on our planet resulting in a vast diversity estimated to be 1 trillion (10^{12}) different species. The genomes of these microbes represent an extraordinary resource for novel enzyme discovery. The field of metagenomics strives to access these genomes, in particular, those from uncultivable species that have been otherwise out of reach.

Glycobiology is a field of research that addresses the structure, function, and biology of glycans and glycoconjugates. Glycans have wide-ranging importance in both basic biology and pharmaceutical science. Aberrant glycosylation has been implicated in many diseases. Additionally, glycosylation of therapeutic proteins affects their safety and efficacy, and is monitored during drug development and manufacturing. Enzymes acting on glycans are important and can represent therapeutic targets (*e.g.*, influenza A neuraminidase), therapeutic agents (*e.g.*, use of sialidases in cancer immunotherapy), or essential glycoanalytical tools (*e.g.*, PNGase F) that help deconvolute glycan structure. In this thesis, a functional metagenomic workflow was created for discovery of new enzymes that act upon glycans. This was used to answer fundamental and application-driven questions in the field of glycobiology.

The functional metagenomic workflow established in this thesis relies on large-insert metagenomic libraries created in *Escherichia coli*. A collection of almost 100,000 clones was created from diverse ecosystems including extreme environments, and is estimated to contain 3-4 million of environmental genes. Three activity-based screens were executed using libraries in this collection. The first led to identification of a novel exosialidase having a unique catalytic mechanism and new protein structure that defined a new glycoside hydrolase family (GH156). The second screen isolated two sialidases with a preference for a non-human form of sialic acid, a specificity not previously described. Finally, a third screen identified

two enzymatic activities: a sugar-specific sulfatase and a sulfate-dependent hexosaminidase that can act on sulfated glycans, an important chemical modification of *N*-glycans for which well-defined analytical tools (*e.g.,* enzymes acting specifically on sulfated sugars) have yet to be established.

To summarize, this work demonstrates the benefit of using functional metagenomics to identify precise enzyme specificities that can answer questions in the field of glycobiology and address technical challenges in glycoanalytics. It highlights that this method of discovery can yield novel protein families that act on glycans, unusual enzymatic specificities, and needed analytical tools. The workflow employed in this thesis is versatile and can easily be adapted to other enzyme discovery projects.

Zusammenfassung

Mit einer Biomasse von etwa 15% sind Bakterien, nach den Pflanzen, der zweitgrößte Produzent von Biomasse auf der Erde. Bakterien sind ubiquitär und haben sich an das Leben in allen Habitaten auf unserem Planeten angepasst, was zu einer enormen Vielfalt von geschätzten 1 Billion (10^{12}) verschiedenen Arten führt. Die Genome dieser Mikroben stellen eine außergewöhnliche Ressource für die Entdeckung (und Nutzung) neuer bakterieller Enzyme dar. Der Bereich der Metagenomik strebt nach der Erforschung dieser Genome insbesondere von Arten, die nicht kultiviert werden können und sonst unzugänglich wären.

Die Glykobiologie ist ein Forschungsgebiet, das sich mit der Struktur, Funktion und Biologie von Glykanen und Glykokonjugaten befasst. Glykane haben eine weitreichende Bedeutung sowohl in der Grundlagenbiologie als auch in der (bio)pharmazeutischen Forschung und Produktion.

Veränderungen der Glykosylierung werden mit einer Vielzahl an Erkrankungen in Verbindung gebracht. Darüber hinaus beeinflusst die Glykosylierung therapeutischer Proteine hinsichtlich deren Sicherheit und Wirksamkeit, und wird daher während der Arzneimittelentwicklung und -herstellung überwacht. Enzyme, die auf Glykane wirken, sog. Glyko-Enzyme können therapeutische Ziele (z. B. Influenza-A-Neuraminidase) oder therapeutische Wirkstoffe (z. B. Verwendung von Sialidasen in Krebsimmuntherapie) darstellen oder als essenzielle glykoanalytische Werkzeuge (z.B. PNGase F) fungieren, die zur Aufklärung der Glykanstruktur beitragen. In dieser Dissertation wurde ein funktionaler metagenomischer Workflow zur Entdeckung neuer Glyko-Enzyme entwickelt. Dieser Workflow wurde zur Beantwortung grundlegender und anwendungsorientierter Fragen im Bereich der Glykobiologie eingesetzt.

Der in dieser Arbeit etablierte funktionelle Metagenomik-Workflow basiert auf metagenomischen Bibliotheken mit großen Inserts, die in *Escherichia coli* erstellt wurden.

Eine Sammlung von fast 100.000 Klonen wurde aus verschiedenen Ökosystemen einschließlich extremer Umgebungen erstellt und enthält schätzungsweise 3-4 Millionen Umweltgene. Drei aktivitätsbasierte Screenings wurden unter Verwendung von Bibliotheken in dieser Sammlung ausgeführt. Das erste Screening führte zur Identifizierung einer neuen Exosialidase mit einem einzigartigen katalytischen Mechanismus und einer neuen Proteinstruktur, die eine neue Glykosid-Hydrolase Familie (GH156) definierte. Das zweite Screening isolierte zwei Sialidasen mit einer Präferenz für eine nicht-humane Form von Sialinsäure, eine Spezifität, die zuvor nicht beschrieben wurde. Schließlich identifizierte ein drittes Screening zwei enzymatische Aktivitäten: Eine zuckerabhängige Sulfatase und eine sulfatabhängige Hexosaminidase, die auf sulfatierte Glykane einwirken können, eine wichtige chemische Modifikation von N-Glykanen, für die noch genau definierte analytische Werkzeuge (z.B. Enzyme die spezifisch auf sulafatierte Zucker wirken) entwickelt werden müssen.

Zusammenfassend demonstriert diese Arbeit den Nutzen der funktionellen Metagenomik zur Identifizierung präziser Enzymspezifitäten, die Fragestellungen im Bereichen der Glykobiologie beantworten können und technische Herausforderungen der Glykoanalytik angehen. Es unterstreicht, dass diese Entdeckungsmethode neue, auf Glykane wirkende Proteinfamilien, ungewöhnliche enzymatische Spezifitäten und benötigte analytische Werkzeuge hervorbringen kann. Der in dieser Arbeit verwendete Workflow ist vielseitig und kann problemlos für Projekte zur Entdeckung anderer Enzyme angepasst werden.

Content