Deorphanizing Enzymes: Characterization of NXPE1 and Its Role in Mucin Glycosylation and Ulcerative Colitis Pathogenesis

HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of

have examined a dissertation entitled

Deorphanizing Enzymes: Characterization of NXPE1 and Its Role in Mucin Glycosylation and Ulcerative Colitis Pathogenesis

presented by Ranad Humeidi

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Ramnik Xavier

Signature _____

Typed name: Prof. Stuart Schreiber

Signature _____

Typed name: Prof. Philip Cole

Signature _____

Typed name: Prof. Benjamin Cravatt

Signature _____ Signed by:

Typed name: Prof. FFAF9/D5D/0341A

Date: January 9, 2025

Deorphanizing Enzymes: Characterization of NXPE1 and Its Role in Mucin Glycosylation and Ulcerative Colitis Pathogenesis

A dissertation presented

by

Ranad Humeidi

to

The Graduate School of Arts and Sciences in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

January 2025

Abstract

Orphan enzymes represent a largely unexplored frontier in molecular biology, with their unknown substrates and functions posing significant challenges and opportunities for understanding human health and disease. This thesis investigates the biochemical role of NXPE1, a gene implicated in ulcerative colitis (UC), to uncover its enzymatic function and physiological relevance. NXPE1 encodes an O-acetyltransferase that selectively modifies the sialic acid N-acetylneuraminic acid (Neu5Ac) at the 9-OH position, producing Neu5,9Ac2. This acetylation process, localized to the Golgi apparatus of colonic epithelial cells, is critical for modulating the biophysical properties of mucus glycoproteins. Using a multidisciplinary approach that integrates in silico modeling, activity-based protein profiling (ABPP), and organoid-based in vitro validation, this study demonstrates that the UC-protective variant NXPE1 G353R disrupts enzyme stability and acetylation activity.

Key findings reveal that NXPE1-mediated sialic acid acetylation is essential for maintaining the structural and functional integrity of the colonic mucosal barrier. The absence of NXPE1 or loss of function due to genetic variation correlates with increased mucus viscosity and altered mucosal sialoglycome, which may protect individuals from inflammation and dysbiosis. These findings not only elucidate a mechanistic basis for the protective effect of NXPE1 variants in UC but also highlight the broader significance of sialic acid modifications in health and disease. By deorphanizing NXPE1, this thesis

advances our understanding of enzyme function and its implications for therapeutic strategies targeting inflammatory bowel diseases.

Dedication

To my parents, whose love and sacrifices have shaped every step of my journey. Leaving behind your home, your careers, and your dreams to build a future for us was an act of immense courage. Though you could no longer practice medicine, you poured your love for healing and humanity into us. This doctorate is as much yours as it is mine—a reflection of your sacrifices, your dreams, and the values you instilled in me. To my mother, whose passion for medicine and science sparked my own, thank you for lighting the way.

To my siblings, my lifelong companions and cheerleaders, thank you for sharing every struggle and every triumph. To Raneem, your endless support through this degree, your belief in me, and your constant encouragement have been my anchor. I couldn't have done this without you.

To my people of Syria, whose courage and resilience shine even in the darkest of times. This work is for every soul who has endured the unthinkable, for those who have held onto hope amidst chaos, and for the dream of peace and renewal that we carry together.

To my friends, who have been my constant source of joy and strength. To Cannelle, my French sister from another mister, thank you for introducing me to the love of cycling and for your steadfast friendship. To Adhana, Zel, Celeste, and Dan, for being my closest friends and my greatest supporters, I am endlessly grateful for your love and

encouragement. To Dina and Aziza, thank you for showing me what Syrian excellence looks like and for inspiring me to always strive higher.

To Pipette, my scientifically curious and impossibly cute feline companion. You kept me company throughout the writing process, sitting on my computer as if guarding my words. Sometimes, I wonder if you hold the secrets of life but simply choose not to share them. Your presence brought both comfort and joy during this journey.

Finally, to Alfredo, my love, my anchor, and my partner in every sense. You have been my biggest champion, lifting me up in every moment of doubt and celebrating every small victory as if it were your own. Thank you for reading every email, practicing every presentation, and listening patiently. You deserve this PhD as much as I do—after all, you can probably recount every moment of it by heart.

Your kindness and belief in me have carried me through this journey. In a world that too often tries to silence voices like mine, you amplified it and made it stronger. Here's to you, my partner in life, my debugging companion, and the love who has turned every challenge into a shared triumph and every success into a celebration of us.

Acknowledgments

This work would not have been possible without the support, guidance, and encouragement of so many incredible individuals who shaped my journey.

First and foremost, I would like to thank my committee members and PhD advisors—Stuart, Ramnik, Ben, Phil, and Laura—for their valuable input and guidance throughout this process. Your expertise and mentorship have shaped this work and my growth as a scientist.

To Noriko, thank you for your hard work in developing organoids and immunostaining efforts. Your dedication to science and your friendship have been a source of strength and encouragement for me. I feel so fortunate to have had you by my side.

To Ben and Kayla, thank you for helping me onboard into the lab and teaching me the practical skills I needed to get started, including protein purification. Your assistance during those early days was invaluable, and I am grateful for your kindness and support.

To Mo, thank you for showing me the magic behind mass spectrometry and for helping me navigate its complexities. Your guidance was valuable during this journey, even amidst challenges.

To my fellow lab members: To my PhD buddy Thomas, thank you for always being there to listen, to help, and to offer words of wisdom. Your knack for knowing exactly what to say and your unwavering support have been invaluable to me.

To Claire, thank you for being a true role model. Your synthetic chemistry skills are nothing short of extraordinary, yet you always strive to make chemistry accessible to everyone through your patient teaching and generous spirit. Joining the lab during the

pandemic was daunting, but you created a sense of community that made it possible to thrive during such challenging times.

To Jon Ostrem and Will Gibson, thank you for being great and kind mentors. Your work ethic and scientific passion are deeply admirable, as is your commitment to leaving a positive impact on humanity. You have set an example I will always aspire to follow.

To my previous mentors: Dr. Michael Doyle, I met you when I was a fresh-off-the-boat immigrant, and you shaped me into the scientist and chemist I am today. I owe my career to you. To Dr. Delikatny, thank you for being the initial spark that made me fall in love with chemical biology and cancer biology. To Todd Golub and Steven Corsello, thank you for nurturing that spark and for teaching me about drug screening while giving me a community where I found my lifelong friends.

To Douglas Wheeler, thank you for being one of my biggest supporters. From reading my graduate school essays to reading this very thesis, your love of science is contagious and has been a source of encouragement throughout my academic journey.

To my cohort, thank you for the camaraderie and support throughout this journey. To Emma, thank you for being the glue that kept us together. To Bruce Birren and Gisselle Velez-Ruiz, thank you for showing me that coming from a diverse background can mean strength, not weakness, and for teaching me how to navigate the complex scientific environment.

To everyone who has walked this journey with me, your support means the world, and this accomplishment is as much yours as it is mine. Thank you.

Table of Contents

Chapter I.

1. Introduction to Inflammatory Bowel Disease and Genetic Factors

1.1. IBD, UC, and CD

Inflammatory bowel disease (IBD) encompasses two chronic inflammatory disorders of the gastrointestinal tract, Crohn's disease (CD) and ulcerative colitis (UC). While both conditions involve intestinal inflammation, they differ in location and depth of involvement. UC primarily affects the colon mucosal layer, whereas CD can affect any part of the gastrointestinal tract and involves transmural inflammation (**Figure 1**) (1). The exact cause of IBD remains unknown; nevertheless, various independent research efforts propose IBD could be attributed to a complex interplay between genetic susceptibility, environmental factors, and an abnormal immune response to the gut microbiota. (2) (3)

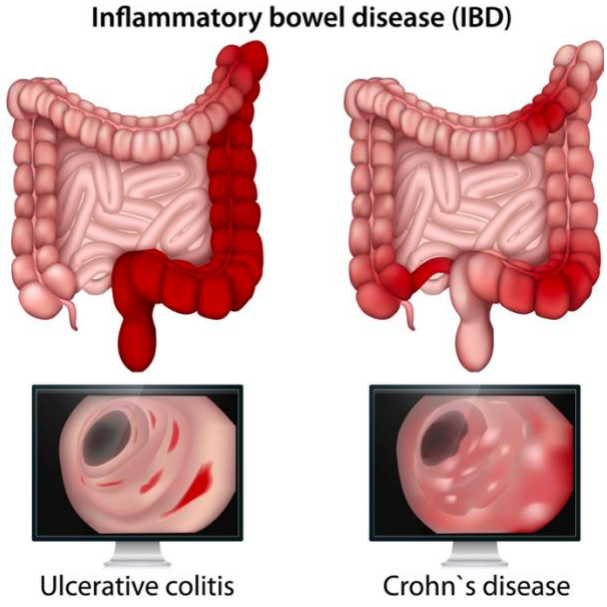


Figure 1 Visual representation of the different inflamed areas of the colon between UC and CD. UC affects the inner lining of the colon and rectum, while CD can affect any part of the GI tract. Image Source: Getty Images / medicalstocks

1.1.1 Introduction to GWAS and SNPS

Historically, research progress in the IBD field proved difficult given its suggested multifactorial nature and range and severity of presenting symptoms(4). However, 21st century improvements in sequencing technologies coupled with mass clinical efforts by hospitals, patient-advocacy, and academic research groups have resulted in robust patient samples banks to be used for sequencing and maintain clinical histories(5). Subsequent advances in algorithms and techniques for analyzing sequencing data such as genome-wide association studies (GWAS) have further increased the understanding of the genetic basis of complex diseases, including inflammatory bowel disease (IBD) (3).In brief, GWAS scans millions of single-nucleotide polymorphisms (SNPs) across the genome to identify genetic variants associated with disease risk(6). SNPs are single-base pair variations in DNA sequences that occur naturally throughout the genome, with an average frequency of approximately one per 1,000 nucleotides. GWAS compares the frequencies of SNPs between individuals with and without IBD, and identifies risk and protective variants(3). Risk variants, for example such as those found in NOD2, IL23R, and ATG16L1, occur more frequently in IBD patients and are associated with increased disease susceptibility(7). Conversely, protective variants such as NXPE1, G211R are more common in healthy individuals and may confer protection against IBD(7) **(Figure 2)**. The strength of these associations is typically expressed as an 'odds ratio', with values greater than 1.0 indicating increased risk and values less than 1.0 suggesting a protective effect.

In the past decade, genome-wide association studies have identified thousands of genetic variants associated with complex traits and diseases(8). These variants can be found dispersed throughout the various regions of the genome (e.g. Introns, exons, promoters,

etc..) with a more limited subsection of variants are found within protein-coding sequences(9). Although less common, these coding variants can have direct and often readily interpretable effects on protein function, allowing for additional exploration and confirmation in research settings(10,11). Despite their potential for functional interpretation, many coding variants identified through GWAS remain uncharacterized(11).

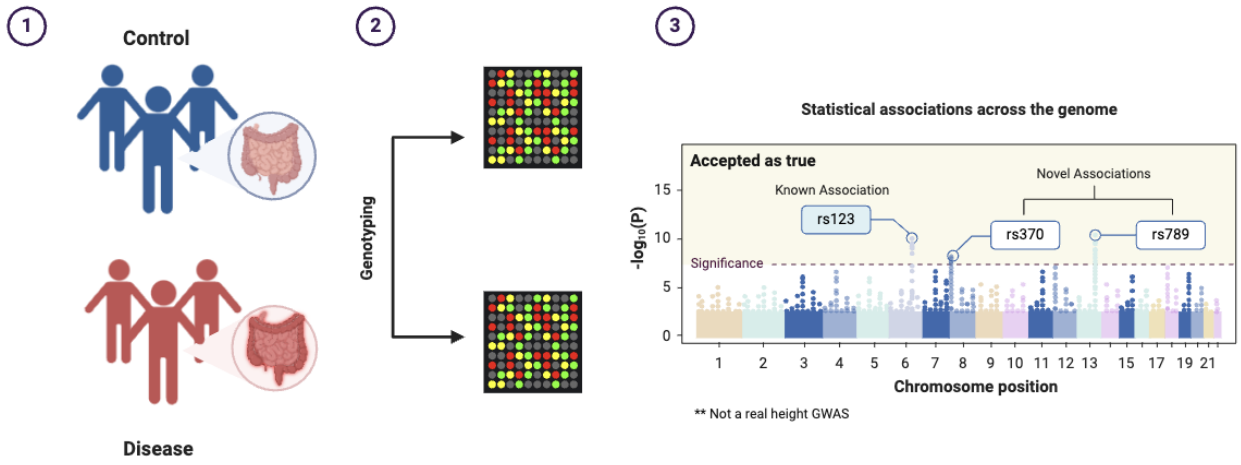


Figure 2 Overview of GWAS. Illustration depicting the key steps of GWAS including selection of feature population and control, genotyping and SNP calling from genomic data, and analysis to find SNPs statistically associated with traits from population of interest.

This lack of characterization represents a significant gap in our understanding of the biological mechanisms underlying complex traits and diseases (12). Coding variants can affect protein structure, stability, or function through various mechanisms, including affecting polarity in binding pockets, introducing premature stop codons, or affecting splicing(13). The characterization of these coding variants is crucial for several reasons. First, it provides direct insights into the molecular mechanisms of disease pathogenesis by

elucidating how alterations in protein function contribute to disease risk(14). Second, understanding the functional effects of these variants can facilitate the prioritization of potential therapeutic targets and inform drug development efforts, particularly in precision medicine approaches. Lastly, characterizing coding GWAS variants can enhance our ability to predict disease risk and develop personalized treatment strategies based on an individual's genetic profile (12).

Along with GWAS, accompanying advances in genomic screening and computational biology have further facilitated the characterization of coding variants. High-throughput experimental assays, such as deep mutational scanning and CRISPR-based screens, enable systematic assessment of variant effects on protein function (15). Additionally, improved computational methods for predicting the functional impact of coding variants have enhanced the capacity precision of variant prioritization for experimental validation (16). Continued efforts to functionally characterize coding GWAS variants are essential for translating genetic associations into actionable biological insights and therapeutic strategies(17).

1.2. Current Approaches in Characterization of orphan enzymes

Enzymes are ordered functional proteins that catalyze biochemical reactions by lowering activation energy and allowing metabolic processes to occur efficiently. As such, they are crucial in biological and disease contexts because they regulate critical pathways. In diseases like IBD, dysfunctional enzymes can lead to inflammation, impaired digestion, and altered gut microbiota. Sequencing in conjunction with functional modeling approaches have been used to delineate the subset of genes and proteins in the human

genome with potential enzymatic qualities. The characterization of such potential enzymes, commonly referred to as 'orphan enzymes', is nontrivial and challenging as their substrates, functional context, and biological roles in disease are unknown (18). Furthermore, the computational modeling used in projections is imprecise; and with the rapidly growing stores of genomic data, there are now far more suggested gene products than there are functions that we can ascribe (18,19). Even so, the characterization of orphan enzymes is crucial for advancing our understanding of physiological processes and has significant implications in multifactorial disease contexts such as that found in IBD (20). (18–20)

1.2.1 *In Silico* Approaches for Characterizing Orphan Enzymes

Computational approaches, or *in silico* methods, provide the most cost-effective and efficient starting points for characterizing orphan enzymes(18,21). In silico modeling leverages existing sequencing data, molecular data if available, and uses bioinformatics tools to predict the potential functions and substrates of orphan enzymes. Various approaches for such modeling exist.

Sequence-based Methods and Advanced Computational Approaches:

Sequence-based methods rely on comparing the nucleic acid and amino acid sequences of enzymes of unknown functions with those of already characterized enzymes (21). This approach is typically used as a first pass, or in conjunction with other approaches, but even alone it can provide valuable insights into the potential functions of uncharacterized proteins.

Homology Modeling:

Homology modeling generates three-dimensional structural models of orphan enzymes based on the known structures of related proteins. Although widely used, its accuracy depends on the availability of suitable templates and sequence similarity(21,22). It also relies on the annotation of specific regions such as binding pockets and terminal regions in these templates. In combination with sequence-based methods, this can suggest key regions of similarity or potential functional regions similar to already characterized enzymes.

Phylogenetic Analysis:

Phylogenetic analysis can infer potential enzymes and their functions based on evolutionary relationships (23). This approach provides insights into the functional diversification of enzyme families and helps to predict the activities of uncharacterized members. It utilizes cross-species data, including existing phylogenetic relationships within any given taxa, comparing the given protein across various species. This can provide key evidence by not only helping identify evolutionary enzymes, but also identify key regons or residues that most remain unchanged to maintain function. (23) (24) (25)

Advanced Computational Approaches: AlphaFold:

Advancements in artificial intelligence, particularly deep learning, have further built on sequence-based methods of enzyme characterization. AlphaFold, developed by DeepMind, represents a significant breakthrough in the field and such efforts (26). It combines both sequencing, structural, and experimental data in conjunction with deep

learning approach to predict protein structures with high accuracy (26). AlphaFold offers several advantages for enzyme characterization including:

1. High-accuracy structure prediction, even for proteins with low sequence similarity to known structures.
2. Insight into catalytic mechanisms by revealing the spatial arrangement of catalytic residues.
3. Substrate-binding site prediction, aiding in the prediction of enzyme function and specificity.
4. Protein-protein interaction prediction using AlphaFold-Multimer.
5. Large-scale structural genomic capability.

This sector of research is currently rapidly developing but will be by all indications essential and central to triaging the most functionally important enzymes in various disease contexts.

1.2.2 *In Vitro* Approaches for Characterizing Orphan Enzymes

While the advancements in AI and computational predications are great for hypothesis generation and to narrow down the function of the enzyme, validating these predictions is crucial and technically challenging. *In vitro* screening approaches are the logical next step and play a crucial role in elucidating the function of orphan enzymes such as the serine hydrolase family. These methods offer a controlled environment to study enzyme activity and substrate specificity, and with the right model can provide valuable real-world insights into their potential physiological roles.

Heterologous Expression Systems

Production of the target enzyme is a fundamental step in in vitro screening. Heterologous expression systems such as *Escherichia coli*, yeast, and insect cells are commonly employed to generate sufficient quantities of orphan enzymes for analysis. For many human enzymes, bacterial expression systems are often preferred because of their simplicity and high yield, although eukaryotic systems may be necessary for enzymes that require specific post-translational modifications.

High-Throughput Substrate Screening

One of the most powerful *in vitro* approaches is high-throughput screening with large libraries of potential substrates. This method involves:

1. **Compound libraries**: Diverse collections of esters, amides, and other hydrolyzable compounds are typically used to probe serine hydrolase activity.
2. **Automated screening**: Robotic systems enable rapid testing of thousands of compounds and reduce inherent variability in handling along with the potential introduction of human bias and error.
3. **Activity detection**: Various methods, including colorimetric assays (e.g., p-nitrophenyl ester hydrolysis), fluorescence-based assays, and mass spectrometry, can been employed to detect enzyme activity.

1.2.3 *In Vivo* Screening Approaches for Characterizing Orphan Enzymes

One powerful approach to identifying the function of enzymes involves the use of genetically altered organisms in which the gene of interest has been inactivated or deleted (knockout). This can be achieved by using transgenic mice and/or cultured cells. Some of the challenges in using this approach are: 1) the gene ortholog in mice may not have the same function as the protein of interest in humans, 2) related gene products may show compensatory activities, masking the functional loss of the gene of interest, and 3) the limitations of detection capabilities when using a complex system.

1.3 Serine Hydrolases

The largest and most diverse enzyme class in nature are serine hydrolases, comprising approximately 1% of all proteins in most proteomes (27). Serine hydrolases feature a conserved N-terminal Gly-Asp-Ser motif and a Ser-Gly-Asn-His active site and are capable of multiple types of enzymatic activity in diverse in metabolic processes. These enzymes play critical roles in numerous physiological processes including lipid metabolism, neurotransmitter signaling, blood coagulation, and inflammation (27, 28). Despite their importance, many hypothesized serine hydrolases are ‘orphans’ with unknown functions, substrates, and regulation (18, 29). The process of "deorphanization, " which assigns functions to these orphan enzymes, has proven to be a critical challenge in modern biochemistry and molecular biology (19). Subsequent sections explore various approaches to characterize orphan enzymes, with a focus on serine hydrolases as a paradigm.

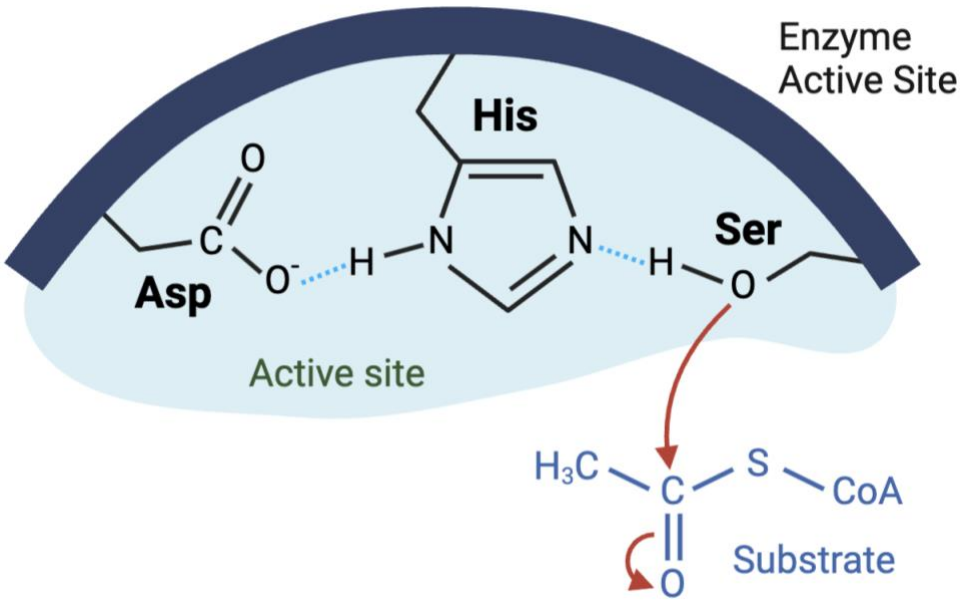


Figure 3 Serine Hydrolases Active Site. Serine hydrolase catalytic triad within an enzyme active site. The triad consists of aspartate (Asp), histidine (His), and serine (Ser) residues. These amino acids work together to facilitate substrate hydrolysis. Aspartate stabilizes the histidine, which acts as a base to deprotonate the serine. The activated serine then attacks the substrate's carbonyl carbon, forming a covalent acyl-enzyme intermediate, crucial for catalysis.

Activity-Based Protein Profiling (ABPP)

ABPP is a powerful chemical proteomic technique that has accelerated the study of enzyme functions and drug discovery, particularly in serine hydrolases. This approach allows researchers to directly assess the functional state of enzymes in complex biological systems, including living cells and organisms. ABPP relies on the use of chemical probes that react covalently with the active sites of specific classes of enzymes. Fluorophosphonate (FP) probes are commonly used for serine hydrolases. These probes typically consist of three key components (**Figure 4**)

1. A reactive group that binds to the enzyme's active site
2. A reporter tag (e.g., biotin or a fluorophore) for detection and enrichment
3. A linker region connecting the reactive group and the reporter tag

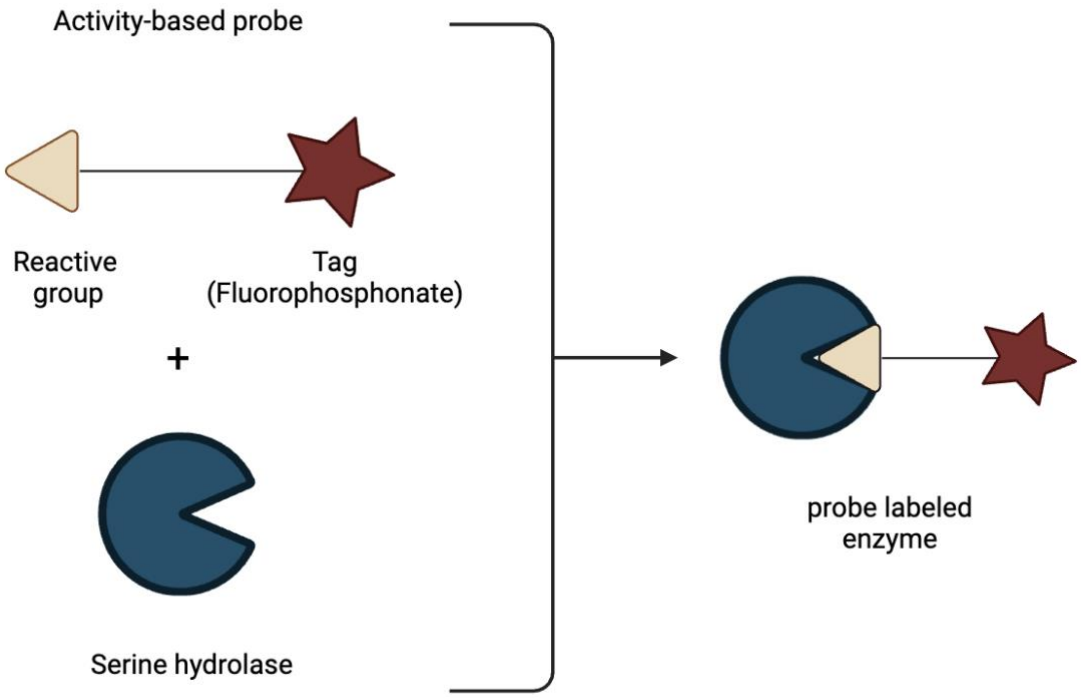


Figure 4 Activity-based protein profiling (ABPP). An activity-based probe, consisting of a reactive group and a fluorophosphonate tag, targets the serine hydrolase enzyme. The reactive group covalently binds to the active site of the enzyme, forming a stable probe-labeled enzyme complex. This labeling allows for the identification of serine hydrolases in a complex biological system.

Applications of ABPP

ABPP offers several advantages for characterizing orphan enzymes.

- **Enzyme activity profiling**: Simultaneous analysis of multiple enzyme activities within a complex proteome, allowing comparisons across different biological states or conditions.
- **Drug Discovery and Target Identification**: Assesses the selectivity of small-molecule inhibitors and identifies off-target interactions of drug candidates.
- **Functional Annotation**: Reveals active site reactivity and potential substrate preferences, aiding in the elucidation of enzyme roles in specific biological pathways.

In conclusion, *in vitro* screening approaches, particularly ABPP, offer a versatile, essential, and powerful toolkit for characterizing orphan serine hydrolases. By combining heterologous expression, high-throughput screening, activity-based profiling, and advanced analytical techniques, researchers can gain significant insight into the potential functions of these enigmatic proteins. The integration of these *in vitro* methods with genomic and proteomic data has greatly accelerated the process of deorphanizing enzymes such as serine hydrolases and has uncovered their roles in various biological processes. However, results from *in vitro* studies should be interpreted cautiously and ideally validated using complementary in vivo approaches to establish physiological relevance.

1.4 Sialic Acids, Sialic Acid Modifying Enzymes and Physiological Links.

N-acetylneuraminic acid (Neu5Ac), or sialic acid, is a monosaccharide with a distinctive nine-carbon backbone. Its nomenclature, sialic acid has an intriguing history: in 1941, German scientist Ernst Klenk coined "neuraminic acid" after discovering a sugar

cleavage product from brain lipids, while in 1952, Swedish biochemist Gunnar Blix introduced "sialic acid" (from Greek "sialos," meaning saliva) for a sugar he found in salivary mucins (30,31).

The stereochemistry of Neu5Ac's C2 chiral center (α or β) was determined by the relationship between the hemiacetal ring closure anomeric center C2 and the highest-numbered chiral center C7. An α-configuration results from the cis-orientation of oxygen atoms at these centers, whereas a trans orientation yields the β-configuration (31). Sialic acids exhibit specific configurational preferences in biological systems. As a component of glycoconjugates, it is exclusively found in the α-configuration, whereas in its high-energy donor form, CMP-sialic acid adopts the β-configuration(30). Sialic acids are predominantly α-linked at the terminal positions of N-glycans, O-glycans, and glycosphingolipids, contributing to the structural and functional diversity of glycoconjugates (30,32). The most common linkages at the C2 position include α2,3- and α2,6-linkages to galactose, α2,6-linkages to N-acetylgalactosamine, and α2,8-linkages to other sialic acid residues to form polysialic acids (33). This diversity in glycosidic linkages and the underlying carbohydrate structures further enhances the complexity of sialic acid-containing compounds (32).

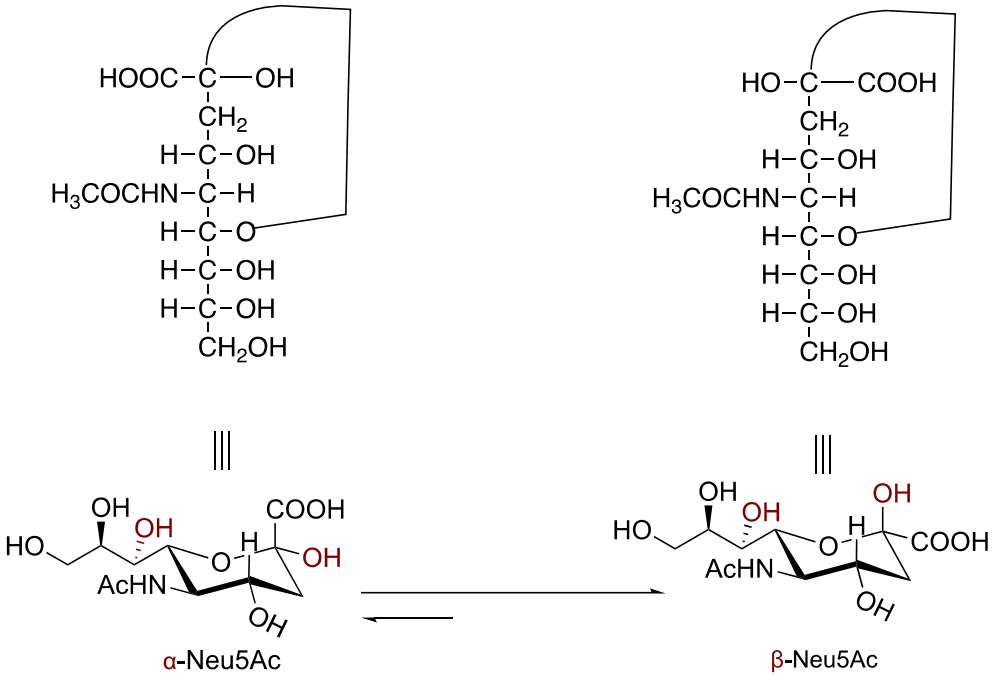


Figure 5 Structural conversion between α-Neu5Ac and β-Neu5Ac, highlighting the stereochemistry of the hydroxyl groups involved in the isomerization process.

1.4.1. Structural Diversity of Sialic Acids

Sialic acids are α-keto acids that undergo various modifications, resulting in a diverse family of over 50 known derivatives (34,35). Neuraminic acid (Neu) represents the unsubstituted core form of sialic acid, and its most common derivatives arise from modifications at the C-5 position, resulting in compounds such as N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc), and keto-deoxyneuraminic acid (Kdn)(34,36). Neu5Ac, Neu5Gc, and Kdn are widespread across different taxa, with Neu5Ac found in vertebrates, echinoderms, insects, protozoa, fungi, and bacteria (34). Neu5Gc is present in vertebrates, protozoa, and fungi, and Kdn is observed in vertebrates and bacteria (34). The structural diversity of sialic acids is further enhanced by the

modification of the hydroxyl groups at positions C4, C7, C8, and C9 through the addition of functional groups such as acetate, lactate, sulfate, phosphate esters, or methyl ethers (37). These modifications contribute to the functional versatility of sialic acids for various biological processes (38).

1.4.2 Sialic Acid Biosynthesis Pathway

The biosynthesis of sialic acid in mammals involves a multistep enzymatic pathway that begins in the cytosol. The initial step is catalyzed by UDP-GlcNAc 2-epimerase/ManNAc kinase (GNE), a bifunctional enzyme that converts UDP-N-acetylglucosamine (UDP-GlcNAc) to N-acetylmannosamine (ManNAc) and subsequently phosphorylates it to ManNAc-6-phosphate (39). This enzyme is allosterically regulated by CMP-Neu5Ac, providing feedback inhibition to control flux into the sialic acid pathway (40). The next step involves the condensation of ManNAc-6-phosphate with phosphoenolpyruvate (PEP), catalyzed by N-acetylneuraminic acid synthase (NANS), to form N-acetylneuraminic acid-9-phosphate (Neu5Ac-9P)(39,41). This was followed by dephosphorylation by N-acetylneuraminic acid phosphatase to yield free Neu5Ac. The final activation step occurs in the nucleus, where CMP-N-acetylneuraminic acid synthetase (CMAS) converts Neu5Ac to CMP-Neu5Ac, the activated form used by sialyltransferases in the Golgi apparatus for glycoconjugate synthesis (39,40). In contrast, bacterial sialic acid biosynthesis utilizes distinct enzymes such as NeuC (GlcNAc-6P 2-epimerase) and NeuB (Neu5Ac synthetase), presenting potential targets for antibiotic development.(42,43)

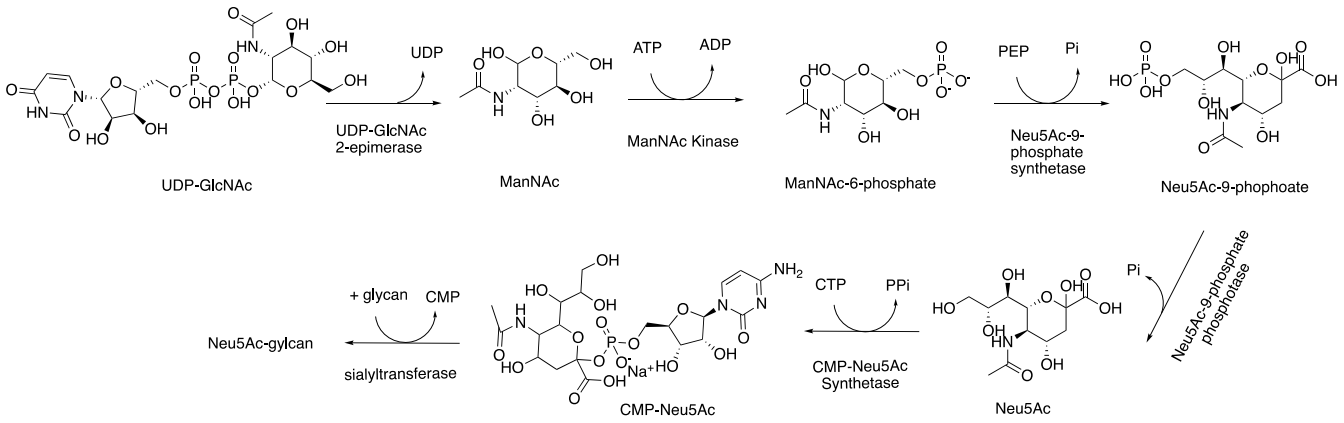


Figure 6 Biosynthetic pathway of Neu5Ac. The process begins with the conversion of UDP-GlcNAc to ManNAc by UDP-GlcNAc 2-epimerase. ManNAc is phosphorylated to ManNAc-6-phosphate by ManNAc Kinase using ATP. Neu5Ac-9-phosphate is then synthesized from ManNAc-6-phosphate by Neu5Ac-9-phosphate synthetase with PEP. Finally, Neu5Ac is formed from CMP-Neu5Ac through CMP-Neu5Ac Synthetase and CTP, followed by the action of Neu5Ac-phosphate phosphatase. This pathway is essential for cell signaling and pathogen recognition.

1.4.2.1 Sialyltransferases: Enzymes Catalyzing Sialic Acid Transfer

Sialyltransferases are a family of glycosyltransferases that catalyze the transfer of sialic acid from cytidine monophosphate-N-acetylneuraminic acid (CMP-Neu5Ac) to the terminal positions of glycoproteins and glycolipid carbohydrate groups (39,44). These enzymes exhibit high specificity, with each sialyltransferase catalyzing the transfer of sialic acid to sugar residues via distinct glycosidic linkages (α2,3-, α2,6-, or α2, 8-)(44,44). In humans, 20 different sialyltransferases have been identified, each of which plays a unique role in glycan structure formation (45).

These enzymes are classified into four main families based on their linkage specificity and acceptor substrates: ST3Gal (I-VI), ST6Gal (I-II), ST6GalNAc (I-VI), and ST8Sia (I-VI)(39,44,45). The ST3Gal family catalyzes α2,3-linkages to galactose residues; the

ST6Gal and ST6GalNAc families form α2,6-linkages to galactose and N-acetylgalactosamine, respectively, while the ST8Sia family creates α2,8-linkages to other sialic acid residues (39,44,45).

The specificity of sialyltransferases significantly influences the structural and biological activity of sialylated glycoconjugates (46). Alterations in sialyltransferase expression or activity have been implicated in various pathological conditions, particularly cancer, where changes in glycosylation patterns can contribute to tumor progression and immune evasion(46,47). Understanding the roles and regulation of these enzymes is crucial for elucidating the complex functions of sialic acids in biological systems and for developing potential therapeutic strategies targeting sialylation processes (48).

1.4.2.2 Sialidases (Neuraminidases): Enzymes for Sialic Acid Removal

Sialidases, also known as neuraminidases, are glycoside hydrolase enzymes that catalyze the removal of terminal sialic acid residues from glycoconjugates (49). In the colon, these enzymes play a crucial role in modulating the sialylation status of mucins and other glycoproteins, which can influence various biological processes, including cell recognition, signaling, and host-pathogen interactions (50,51). Sialidases exhibit varying specificities for different sialic acid modifications, such as O-acetylation, which is particularly prevalent in colonic mucins (49,52). The human genome encodes four distinct sialidase isoforms (NEU1-4), each with unique substrate preferences and subcellular localizations (49). In the context of colonic sialidases, their ability to remove modified sialic acids, especially O-acetylated forms, is of particular interest due to the developmental

regulation and altered expression of these modifications in colorectal cancer (52,53). The specificity of sialidases for O-acetylated sialic acids can vary, with some isoforms showing reduced activity against heavily O-acetylated substrates (51). This differential activity towards modified sialic acids contributes to the complex regulation of the colonic glycome and may have implications for both normal physiological processes and pathological conditions in the colon (52,53).

1.4.3 Sialic Acid Modifying Enzymes

Sialic acids exhibit remarkable structural diversity owing to various enzymatic modifications, including O-acetylation, N-glycolylation, and lactonization (54). O-Acetylation is a prevalent modification catalyzed by sialic acid O-acetyltransferases (SOATs), which significantly influence the biological properties and functions of sialic acid-containing glycoconjugates (54). These enzymes transfer acetyl groups to hydroxyl groups at the C4, C7, C8, and C9 positions of sialic acids (54). The most common form of O-acetylation in mammalian cells occurs at the C7, C8, and C9 positions and is catalyzed by CASD1, while 4-O-acetylation is less frequent (54).

As previously mentioned, sialic acids exhibit remarkable structural diversity owing to various enzymatic modifications, including O-acetylation, N-glycolylation, and lactonization (54). O-Acetylation is a prevalent modification catalyzed by sialic acid O-acetyltransferases (SOATs), which significantly influence the biological properties and functions of sialic acid-containing glycoconjugates (54). These enzymes transfer acetyl groups to hydroxyl groups at the C4, C7, C8, and C9 positions of sialic acids (54). The

most common form of O-acetylation in mammalian cells occurs at the C7, C8, and C9 positions and is catalyzed by CASD1, while 4-O-acetylation is less frequent (54). O-acetylated sialic acids exhibit increased resistance to sialidase cleavage, thereby providing protection to glycoconjugates (54). Abnormal levels of O-acetylation have been associated with various pathological conditions, including cancer and immune system dysfunctions(54).

Another significant modification is N-glycolylation, catalyzed by CMP-Neu5Ac hydroxylase (CMAH). This enzyme converts CMP-N-acetylneuraminic acid (CMP-Neu5Ac) into CMP-N-glycolylneuraminic acid (Neu5Gc)(54). Interestingly, humans lack functional CMAH owing to an inactivating mutation, resulting in the absence of endogenous Neu5Gc(54). However, the incorporation of Neu5Gc from dietary sources into human tissues has been linked to inflammation and cancer, highlighting the importance of this modification for health and disease(47,54). Understanding the enzymes responsible for sialic acid modification is crucial for elucidating their roles in various biological processes and developing potential therapeutic strategies targeting sialylation pathways (48).

1.4.4 Sialic Acid Acetylation and IBD

7(9)-O-Acetyltransferases have been extensively studied in rat liver and bovine submandibular glands. Despite their discovery approximately three decades ago, these enzymes have proven challenging to purify to homogeneity(52,54). Attempts to clone the sialate O-acetyltransferase gene(s) have previously been unsuccessful even though transfection of various transcripts has been shown to increase sialic acid O-acetylation in cultured cells (52). Human colonic mucins exhibit a high degree of sialic acid O-acetylation, with chemical analyses revealing that over 50% of colonic mucin sialic acids

are O-acetylated, including 30% containing oligo-O-acetyl forms with 2 or 3 acetyl esters per sialic acid residue(52). Histochemical data suggest that total sialate O-acetylation may

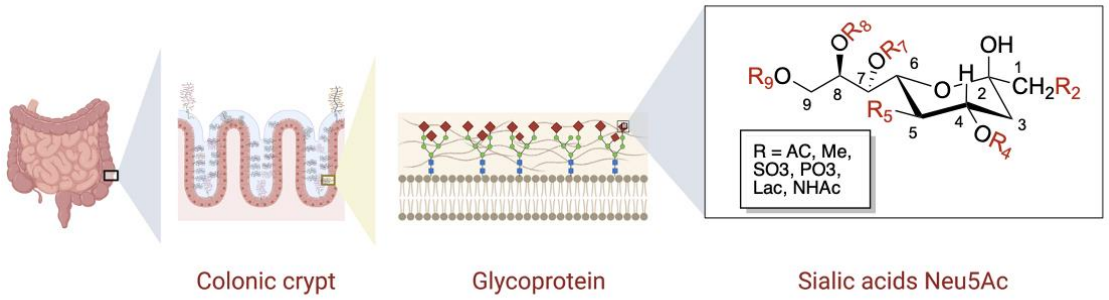


Figure 7 Sialic acid modifications in the colon. Colonic mucus is highly glycosylated with Sialic acid as the terminal sugar. Sialic Acids can be further modified by Acetylation, Sulfidation and methylation. These modifications play a crucial role in cellular interactions and protection of the mucosal surface.

reach at least 80%(52). This extensive O-acetylation is characteristic of the human colon and has not been observed in other human tissues, although comparable levels have been reported in bovine submandibular gland mucins (52).

The occurrence of O-acetylated sialic acids in rat and human colonic mucins is developmentally regulated, with sialate O-acetylation levels increasing significantly after birth (52). In contrast to certain cancer types, such as melanoma and basalioma, the degree of O-acetylation in colon cancer cells appears to be reduced compared to healthy colonic mucosa (55). Studies have additionally demonstrated that di- and tri-O-acetylated sialic acids are diminished in the early stages of malignant transformation, whereas a significant proportion of mono-O-acetyl sialic acid remains constant(53,55). These observations raise essential questions regarding the occurrence and enzymology of O-acetylated sialic acids in human colon. Although much of our knowledge concerning sialic acid O-acetylation has been derived from studies on bovine submandibular glands and rat liver, relatively little is

known about this process in human colonic tissue(52). This thesis aims to elucidate the substrate specificity, subcellular localization, and characteristics of 7(9)-OAT in human colonic mucosa(52).

References

1. McDowell C, Farooq U, Haseeb M. Inflammatory Bowel Disease. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 [cited 2024 Sep 17]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK470312/

2. Jarmakiewicz-Czaja S, Zielińska M, Sokal A, Filip R. Genetic and Epigenetic Etiology of Inflammatory Bowel Disease: An Update. Genes. 2022 Dec 16;13(12):2388.

3. De Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017 Feb;49(2):256–61.

4. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nat Rev Methods Primer. 2021 Aug 26;1(1):59.

5. Pizarro TT, Stappenbeck TS, Rieder F, Rosen MJ, Colombel JF, Donowitz M, et al. Challenges in IBD Research: Preclinical Human IBD Mechanisms. Inflamm Bowel Dis. 2019 May 16;25(Supplement_2):S5–12.

6. Ntunzwenimana JC, Boucher G, Paquette J, Gosselin H, Alikashani A, Morin N, et al. Functional screen of inflammatory bowel disease genes reveals key epithelial functions. Genome Med. 2021 Nov 11;13(1):181.

7. Crohn's disease: MedlinePlus Genetics [Internet]. [cited 2024 Sep 17]. Available from: https://medlineplus.gov/genetics/condition/crohns-disease/

8. Abdellaoui A, Yengo L, Verweij KJH, Visscher PM. 15 years of GWAS discovery: Realizing the promise. Am J Hum Genet. 2023 Feb 2;110(2):179–94.

9. Lichou F, Trynka G. Functional studies of GWAS variants are gaining momentum. Nat Commun. 2020 Dec 8;11(1):6283.

10. Sun BB, Kurki MI, Foley CN, Mechakra A, Chen CY, Marshall E, et al. Genetic associations of protein-coding variants in human disease. Nature. 2022 Mar;603(7899):95–102.

11. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun. 2018 Jun 21;9(1):2427.

12. Alzu'bi AA, Zhou L, Watzlaf VJM. Genetic Variations and Precision Medicine. Perspect Health Inf Manag. 2019 Apr 1;16(Spring):1a.

13. Field MA. Bioinformatic Challenges Detecting Genetic Variation in Precision Medicine Programs. Front Med. 2022 Apr 8;9:806696.

14. Uddin M, Wang Y, Woodbury-Smith M. Artificial intelligence for precision medicine in neurodevelopmental disorders. Npj Digit Med. 2019 Nov 21;2(1):1–10.

15. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014 Aug;11(8):801–7.

16. Roca I, Fernández-Marmiesse A, Gouveia S, Segovia M, Couce ML. Prioritization of Variants Detected by Next Generation Sequencing According to the Mutation Tolerance and Mutational Architecture of the Corresponding Genes. Int J Mol Sci. 2018 May 27;19(6):1584.

17. Flynn ED, Lappalainen T. Functional Characterization of Genetic Variant Effects on Expression. Annu Rev Biomed Data Sci. 2022 Aug 10;5:119–39.

18. Sorokina M, Stam M, Médigue C, Lespinet O, Vallenet D. Profiling the orphan enzymes. Biol Direct. 2014 Jun 6;9:10.

19. Pouliot Y, Karp PD. A survey of orphan enzyme activities. BMC Bioinformatics. 2007 Jul 10;8(1):244.

20. Liu Y, Patricelli MP, Cravatt BF. Activity-based protein profiling: The serine hydrolases. Proc Natl Acad Sci U S A. 1999 Dec 21;96(26):14694–9.

21. Hadadi N, MohammadiPeyhani H, Miskovic L, Seijo M, Hatzimanikatis V. Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites. Proc Natl Acad Sci U S A. 2019 Apr 9;116(15):7298–307.

22. Kumar S. Computational identification and binding analysis of orphan human cytochrome P450 4X1 enzyme with substrates. BMC Res Notes. 2015 Jan 17;8(1):9.

23. Lai J, Jin J, Kubelka J, Liberles DA. A phylogenetic analysis of normal modes evolution in enzymes and its relationship to enzyme function. J Mol Biol. 2012 Sep 21;422(3):442–59.

24. Rauwerdink A, Kazlauskas RJ. How the Same Core Catalytic Machinery Catalyzes 17 Different Reactions: the Serine-Histidine-Aspartate Catalytic Triad of α/β-Hydrolase Fold Enzymes. ACS Catal. 2015 Oct 2;5(10):6153–76.

25. Anderson AC, Stangherlin S, Pimentel KN, Weadge JT, Clarke AJ. The SGNH hydrolase family: a template for carbohydrate diversity. Glycobiology. 2022 Jul 23;32(10):826–48.

26. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

27. Long JZ, Cravatt BF. The Metabolic Serine Hydrolases and Their Functions in Mammalian Physiology and Disease. Chem Rev. 2011 Oct 12;111(10):6022–63.

28. Bachovchin DA, Cravatt BF. The pharmacological landscape and therapeutic potential of serine hydrolases. Nat Rev Drug Discov. 2012 Jan;11(1):52–68.

29. Navarrete M, Ho J, Krokhin O, Ezzati P, Rigatto C, Reslerova M, et al. Proteomic characterization of serine hydrolase activity and composition in normal urine. Clin Proteomics. 2013 Nov 15;10(1):17.

30. Frontiers | Diversity of sialic acids and sialoglycoproteins in gametes and at fertilization [Internet]. [cited 2024 Sep 30]. Available from: https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2022.982931/full

31. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J. Sialic Acids. In: Essentials of Glycobiology [Internet]. Cold Spring Harbor Laboratory Press; 1999 [cited 2024 Sep 29]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK20724/

32. Suzuki A. Sialic Acids and Human Evolution. Glycoforum. 2002 Mar 18;6:A2.

33. Büll C, Stoel MA, Den Brok MH, Adema GJ. Sialic Acids Sweeten a Tumor's Life. Cancer Res. 2014 Jun 15;74(12):3199–204.

34. Ghosh S. Sialic acid and biology of life: An introduction. Sialic Acids Sialoglycoconjugates Biol Life Health Dis. 2020;1–61.

35. Chen X, Varki A. Advances in the Biology and Chemistry of Sialic Acids. ACS Chem Biol. 2010 Feb 19;5(2):163–76.

36. Schauer R. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol. 2009 Oct;19(5):507–14.

37. Frontiers | Biological functions of sialic acid as a component of bacterial endotoxin [Internet]. [cited 2024 Sep 30]. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.1028796/full

38. Zhu W, Zhou Y, Guo L, Feng S. Biological function of sialic acid and sialylation in human health and disease. Cell Death Discov. 2024 Sep 30;10(1):1–15.

39. Li Y, Chen X. Sialic acid metabolism and sialyltransferases: natural functions and applications. Appl Microbiol Biotechnol. 2012 May;94(4):887–905.

40. Du J, Meledeo MA, Wang Z, Khanna HS, Paruchuri VDP, Yarema KJ. Metabolic glycoengineering: Sialic acid and beyond. Glycobiology. 2009 Dec 1;19(12):1382–401.

41. Real-time monitoring of the sialic acid biosynthesis pathway by NMR - Chemical Science (RSC Publishing) DOI:10.1039/D2SC06986E [Internet]. [cited 2024 Sep 27]. Available from: https://pubs.rsc.org/en/content/articlehtml/2023/sc/d2sc06986e

42. Chen SC, Huang CH, Lai SJ, Yang CS, Hsiao TH, Lin CH, et al. Mechanism and inhibition of human UDP-GlcNAc 2-epimerase, the key enzyme in sialic acid biosynthesis. Sci Rep. 2016 Mar 16;6(1):23274.

43. Tanner ME. The enzymes of sialic acid biosynthesis. Bioorganic Chem. 2005 Jun;33(3):216–28.

44. Gu J, Isaji T. Specific sialylation of N-glycans and its novel regulatory mechanism. Glycoconj J. 2024 Jun 1;41(3):175–83.

45. Hugonnet M, Singh P, Haas Q, von Gunten S. The Distinct Roles of Sialyltransferases in Cancer Biology and Onco-Immunology. Front Immunol. 2021 Dec 17;12:799861.

46. Pietrobono S, Stecca B. Aberrant Sialylation in Cancer: Biomarker and Potential Target for Therapeutic Intervention? Cancers. 2021 Apr 22;13(9):2014.

47. Sialic acids in cancer biology and immunity | Glycobiology | Oxford Academic [Internet]. [cited 2024 Sep 30]. Available from: https://academic.oup.com/glycob/article/26/2/111/2355440?login=false

48. Dobie C, Skropeta D. Insights into the role of sialylation in cancer progression and metastasis. Br J Cancer. 2021 Jan;124(1):76–90.

49. Sialidases as regulators of bioengineered cellular surfaces | Glycobiology | Oxford Academic [Internet]. [cited 2024 Oct 1]. Available from: https://academic.oup.com/glycob/article/25/7/784/1988824?login=true

50. Barnard KN, Alford-Lawrence BK, Buchholz DW, Wasik BR, LaClair JR, Yu H, et al. Modified Sialic Acids on Mucus and Erythrocytes Inhibit Influenza A Virus Hemagglutinin and Neuraminidase Functions. J Virol. 2020 Apr 16;94(9):e01567-19.

51. Mammalian sialidases: Physiological and pathological roles in cellular functions | Glycobiology | Oxford Academic [Internet]. [cited 2024 Oct 1]. Available from: https://academic.oup.com/glycob/article/22/7/880/1987986?login=false

52. Shen Y, Tiralongo J, Kohla G, Schauer R. Regulation of sialic acid O-acetylation in human colon mucosa. Biol Chem. 2004 Feb;385(2):145–52.

53. Hutchins JT, Reading CL, Giavazzi R, Hoaglund J, Jessup JM. Distribution of Mono-, Di-, and Tri-O-Acetylated Sialic Acids in Normal and Neoplastic Colon1. Cancer Res. 1988 Jan 1;48(2):483–9.

54. Visser EA, Moons SJ, Timmermans SBPE, de Jong H, Boltje TJ, Büll C. Sialic acid O-acetylation: From biosynthesis to roles in health and disease. J Biol Chem. 2021 Jun 19;297(2):100906.

55. Reduction of sialic acid O-acetylation in human colonic mucins in the adenoma-carcinoma sequence | Glycoconjugate Journal [Internet]. [cited 2024 Oct 1]. Available from: https://link.springer.com/article/10.1023/A:1007026314792

Chapter II.

The Ulcerative Colitis-Associated Gene NXPE1 Catalyzes Glycan Modifications on Colonic Mucin

Ranad Humeidi[1,2,3], Noriko Rapley[2,3], Xiebin Gu[2,3], Joon Soo An[4], Ashwin N. Ananthakrishnan[5,6], Elizabeth A. Creasey[3], Mark J. Daly[2,7], Stuart L. Schreiber[2,8], Daniel B. Graham[2,3,9,*], Mohammad R. Seyedsayamdost[4,10,*], Ramnik J. Xavier[2,3,9,*]

[1]Program for Chemistry & Chemical Biology, Harvard Medical School, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[4]Department of Chemistry, Princeton University, Princeton, NJ, USA

[5]Department of Medicine, Harvard Medical School, Boston, MA, USA

[6]Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA

[7]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[8]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA, USA

[9]Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[10]Department of Molecular Biology, Princeton University, Princeton, NJ, USA

*Correspondence: dgraham@broadinstitute.org; mrseyed@princeton.edu; xavier@molbio.mgh.harvard.edu

ABSTRACT

Colonic mucus forms a first line of defense against bacterial invasion while providing nutrition to support coinhabiting microbes in the gut. Mucus is composed of polymeric networks of mucin proteins, which are heavily modified post-translationally. The full compendium of enzymes responsible for these modifications and their roles in health and disease remain incompletely understood. Herein we determine the biochemical function of NXPE1, a gene implicated in ulcerative colitis (UC), and demonstrate that it encodes an acetyltransferase that modifies mucin glycans. Specifically, NXPE1 utilizes acetyl-CoA to regioselectively modify the mucus sialic acid, 5-*N*-acetylneuraminic acid (Neu5Ac), at the 9-OH group to generate 9-*O*-acetylated Neu5Ac (Neu5,9Ac$_2$). We further demonstrate that colonic organoids derived from donors harboring the missense variant NXPE1 G353R, which is protective against UC, exhibit severely impaired acetylation of Neu5Ac on mucins. Together, our findings support a model in which NXPE1 masks the alcohols of mucus sialoglycans via acetylation, which is important for modulating mucus barrier properties that limit interactions with commensal microbes.

INTRODUCTION

Inflammatory bowel disease (IBD) is a group of disorders comprised of Crohn's disease (CD) and ulcerative colitis (UC), which involve chronic inflammation of the digestive tract. In 2020, there were ~2.5 million IBD patients in the United States, a number that is estimated to rise to ~3.5 million by 2030 (1). IBD can be debilitating and can lead to life-

threatening complications. Studies have linked IBD to specific environmental conditions, genetic risk factors, and immune dysregulation, which suggests a complex pathophysiologic process (2–4). Genome-wide association studies (GWAS) have been particularly successful in identifying genetic loci associated with CD and UC, some of which can be fine-mapped to implicate causal genes and variants (5–9). One such gene, *NXPE1* (formerly FAM55A), encodes for a previously uncharacterized protein that has a missense variant protective against UC through yet unknown mechanisms (6). Characterization of these variants, therefore, can provide insights into IBD as well as new molecular targets and possible treatment options (10).

Based on its protein sequence and homology to known enzymes, NXPE1 is classified as a putative member of the GDSL/SGNH esterase superfamily and, within it, the PC-esterase family. Members of these families are serine hydrolases, one of the largest functional enzyme classes in mammals comprising 1-2% of the total proteome (11). Serine hydrolases feature a conserved Gly-Asp-Ser-Leu motif and a Ser-Gly-Asn-His active site, after which the superfamily is named, and are capable of catalyzing multiple reactions in diverse metabolic processes (12). PC-esterases have been purified and characterized from plants and bacteria, where they play roles in mediating surface glycan acylation levels, a feature key to defining chemical properties of glycans such as tensile strength, metabolite recycling, hydrophobicity, and susceptibility to degradation (11, 13, 14). In animals however, only one family member has been characterized thus far, CASD1, an acetyltransferase that specifically catalyzes the modification of sialic acids, which are essential carbohydrates that cap glycoproteins and glycolipids (15). This process is linked to cell recognition and, importantly, pathogen resistance mechanisms.

Other human proteins that may be involved in this process have yet to be identified (16, 17).

The primary barrier between epithelial cells of the intestinal lining and the microbial flora in the gut is the mucosal layer, which consists of highly glycosylated proteins (18–20). Defects in this layer and the subsequent loss of barrier integrity result in dysregulation of intestinal homeostasis and are known to trigger an inflammatory response (21–24). Intestinal mucus is primarily composed of networks of mucin glycoproteins containing numerous oxygen-linked glycans that may be further modified (25–27). Due to the dynamic nature of gut mucus, these glycoproteins are heavily and reversibly modified post-translationally (23, 28). However, the enzymes responsible for these modifications are incompletely described (29, 30). Given the putative enzymatic function of NXPE1 and its association with UC, we hypothesized that NXPE1 may regulate the production and maintenance of the mucosal barrier, thus playing a crucial role in controlling the gut immune response. We, therefore, initiated a multi-pronged study to better understand its role in UC.

Here, we report that NXPE1 is an *O*-acetyltransferase of mucosal sialic acids and is involved in the production and maintenance of the mucosal barrier. NXPE1, is thus an important factor in mediating the gut immune response. These findings are validated by organoid studies, which revealed significant impairment of sialic acid acetylation in organoids carrying the UC-protective variant or in NXPE1 knockouts. Our results thus provide a mechanistic basis for the protective effect of NXPE1 variants. Moreover, they illustrate the power of deorphanizing enzymes encoded in the human genome to gain mechanistic insights into human biology and disease.

RESULTS

NXPE1 is highly expressed in colonic epithelial cells

To begin, we sought to determine the expression levels and localization of NXPE1 in epithelial cells and lineages thereof, which are the main source of mucus in the intestinal lining (31). Analysis of previously available, spatially defined single-cell RNA sequencing (scRNA-seq) data from patient-derived biopsies (32) revealed the highest expression of NXPE1 in goblet cells (GCs), transit amplifying (TA) cells, intestinal epithelial stem cells, and absorptive enterocytes (**Figure 1A**). The expression of other NXPE paralogs, such as NXPE2 and NXPE4, tracked with that of NXPE1 (**Figure 1A**).

As further evidence, we examined the localization of NXPE1 in situ by staining GC-polarized intestinal organoids from human donor-derived biopsies. Using CRISPR/Cas9, $NXPE1^{-/-}$ organoids were generated as a control. We observed increased expression of NXPE1 in GC-differentiated organoids compared to undifferentiated stem cells. To analyze the subcellular localization of NXPE1, we performed immunofluorescence staining and imaging studies. First, we generated a monoclonal NXPE1 antibody for immunofluorescence imaging and showed that it specifically detected NXPE1 in organoids, while the signal was ablated after CRISPR-mediated knock-out of NXPE1 (**Figure 1B**). Next, we focused on locating cellular compartments known to control protein glycosylation and co-stained with the ER marker ERp72 and the Golgi markers GM130 and CCDC186. Accordingly, we found that NXPE1 localizes to the Golgi apparatus rather than the ER (**Figure 1C and D**). Together, these results show that

NXPE1 is highly expressed in colonic epithelial cells and localizes to the Golgi compartment.

NXPE1 is an *O*-acetyltransferase of the serine hydrolase family

Based on its protein sequence and homology to known enzymes, NXPE1 has been classified as a serine hydrolase and a member of the GDSL/SGNH esterase superfamily (**Figure 2A**) (33). The UC-protective variant carries a modification in this sequence, converting the essential glycine in the GDSL sequence (G353) to arginine (G353R). Moreover, sequence analysis shows that serine-355 is conserved and likely acts as the active site covalent catalyst. To determine whether NXPE1 is indeed a serine hydrolase, we used an established activity-based protein profiling (ABPP) assay, wherein the putative active site serine is specifically and covalently labeled with a fluorophosphonate-based probe that is then detected through a biotin tag (**Figure 2B**) (34). A control experiment with a validated serine hydrolase, the bovine coronavirus esterase (BCovHe), showed expected labeling and the double-band pattern previously described, representing the monomeric and dimeric forms of the protein (35–37). Likewise, experiments with NXPE1, but not the S355A variant, revealed labeling with the ABPP probe (**Figure 2B**). These studies were not possible with the NXPE1 G353R variant, as it was insoluble in both mammalian and bacterial expression systems (**Figure 2C**). In fact, other substitutions at G353, such as a tryptophan, lysine, asparagine, and even alanine, resulted in insoluble protein. These results point to the crucial role of G353 in maintaining an intact and folded NXPE1 structure.

Serine hydrolases are one of the largest enzyme superfamilies and are capable of diverse enzymatic activities in multiple metabolic processes. To characterize the function of NXPE1, we investigated acetyltransferase activity by testing whether it could form a covalent acetyl-enzyme intermediate upon incubation with acetyl-coenzyme A (CoA) in the absence of an acceptor substrate. Upon incubation with acetyl-CoA, followed by proteolytic digestion, and analysis of the peptide fragments via liquid chromatography-coupled electrospray ionization-mass spectrometry (LC-ESI-MS), we found a peptide containing the active site residue S355 in its modified, acetylated form, as indicated by a 42 Da mass shift (**Figure 2D**). The mutant NXPE1 S355A, however, was not acetylated upon incubation with acetyl-CoA. These results indicate that NXPE1 is a novel acetyltransferase and that S355 serves as the active site covalent catalyst.

NXPE1 mediates 9-*O*-acetylation of Neu5Ac with high substrate specificity

Acetyltransferases transfer an acetyl group from a common donor, such as acetyl-CoA, to an acceptor substrate (38, 39). To identify the acceptor substrate of NXPE1, we conducted a screen with various monosaccharides, given the Golgi localization of the enzyme. We tested 18 different sugars as potential acetyl group acceptors, including 11 nucleotidyl sugars and seven free monosaccharides (**Figure 3A**). After incubating NXPE1 with acetyl-CoA and each acceptor molecule, the reaction was acid-quenched to preclude nonenzymatic acetyl group transfer, precipitated NXPE1 was removed by centrifugation, and the sugars were modified with 1,2-diamino-4,5-methylenedioxy-benzene (DMB) to facilitate their detection. The substrate/products were then analyzed using high-performance liquid chromatography-coupled quadrupole time-of-flight mass

spectrometry (HPLC-qTOF-MS). The results showed acetyltransferase activity with several sugars, the strongest occurring with *N*-acetylneuraminic acid (Neu5Ac) and Cytidine-5′-monophosphate-*N*-acetylneuraminic acid (CMP-Neu5Ac). Neu5Ac, the predominant sialic acid in human and many mammalian cells, is essential for various cellular functions, while CMP-Neu5Ac serves as a critical intermediate in the biosynthesis of sialoglycoconjugates (40, 41). Additionally, approximately 10-fold weaker acetyltransferase activity was observed with UDP-Xylose, a nucleotide sugar involved in the biosynthesis of glycosaminoglycans. NXPE1 also showed ~100-fold weaker, but detectable turnover with several other sugars, such as *N*-acetylgalactosamine (GalNac) and UPD-galactose (**Figure 3A**). Importantly, no reaction was detected at all with the NXPE1 S355A mutant that disrupts the catalytic serine.

To test the specificity of NXPE1, we incubated it with equimolar concentrations of the most active acceptors, leveraging isotopically labeled Neu5Ac (Neu5Ac-$^{13}C,d_3$) to distinguish it from CMP-Neu5Ac. Strikingly, NXPE1 showed high selectivity towards Neu5Ac, generating 20-fold higher levels of acetylated product than with any of the other substrates tested (**Figure 3B**). The activity was enzyme concentration-dependent with product accumulating in a time-dependent manner. No acetylated product was observed with NXPE1 S355A. Extended incubation of NXPE1 with Neu5Ac led to the appearance of small amounts of bisacetylated product (0.2–2% relative to monoacetylated Neu5Ac, see below).

Neu5Ac contains several alcohol functionalities that could be acetylated by NXPE1. To determine the precise location of acetylation, we conducted two experiments. First, we co-injected the product of the reaction with a standard consisting of Neu5,9Ac$_2$,

which was generated synthetically, and found the two compounds to co-elute, indicating they are identical (**Figure 3C**). Moreover, we carried out a large-scale reaction of NXPE1 with Neu5Ac, purified the product and determined its structure using 1D/2D NMR spectral analysis. The data were consistent with the analysis above and showed Neu5,9Ac$_2$ as the reaction product. Specifically, a new acetyl group was detected in the product with δ_H/δ_C of 2.05/20.2 for the methyl group and δ_C of 174.2 for the carbonyl-carbon. HMBC correlations from the methylene-protons at C-9 to the carbonyl carbon (δ_C 174.2) established the location of the modification (**Figure 3D**). Together, these findings show that NXPE1 is an acetyltransferase that acts preferentially on Neu5Ac and acetylates the C-9 hydroxy group to generate Neu5,9Ac$_2$.

NXPE1 mediates sialic acetylation of colonic mucus

Having established the biochemical activity of NXPE1, we next sought to assess its role in cell models using colonic organoids. Specifically, we developed organoids from five donors expressing the common allele of NXPE1 and seven donors homozygous for the UC-protective variant G353R. We also generated NXPE1$^{-/-}$ organoids using CRISPR/Cas9 as a negative control. Since human cells also encode the paralogous NXPE4, which we reasoned may have redundant functions that could potentially compensate for NXPE1 KO, we also generated NXPE1$^{-/-}$/NXPE4$^{-/-}$ double KO organoids. The organoids were then assessed by western blot and/or MS analysis to determine the role of NXPE1. NXPE1 was only detected in organoids expressing the common allele but not those expressing the G353R variant, presumably because it is a hypopmorphic allele encoding an unstable protein (**Figure 4A**).

To interrogate the sialic acid content of the organoids, we grew them as monolayers in air-liquid interface (ALI) cultures with vasoactive intestinal peptide (VIP), which promotes production of a hydrated apical mucus barrier. Accordingly, mucus was harvested and processed to release sialic acids, and we labeled these with DMB and assessed them by HPLC-qTOF-MS. Neu5Ac was detected in mucus from cells expressing the common allele, G353R variant, as well as the NXPE1$^{-/-}$ and NXPE1$^{-/-}$/NXPE4$^{-/-}$ (**Figure 4B**). However, the product (Neu5,9Ac$_2$) was produced at high levels only in mucus from cells expressing the common allele. It was undetectable in mucus from homozygous G353R variants and observed minimally in mucus from the NXPE1$^{-/-}$ organoids. Importantly, mucus from NXPE1$^{-/-}$/NXPE4$^{-/-}$ cultures revealed a similar decrease in Neu5,9Ac$_2$ as NXPE1$^{-/-}$, although to a slightly lesser extent due to less efficient knockout of NXPE1 in double knockout organoids. Together these findings suggest that NXPE4 does not compensate for the loss of NXPE1, because they do not have redundant functions with respect to acetylation of sialic acid on mucus (**Figure 4B**). Furthermore, acetylation of sialic acids from NXPE1$^{-/-}$ organoids could be rescued by incubating mucus from the knockout with recombinant NXPE1, further indicating that it is required and sufficient for this activity (data not shown). Detailed analysis of the HPLC-qTOF-MS data also revealed a second minor product, that we identified as a tris-acetylated neuraminic acid (Neu5,7,9Ac$_3$); this product was also observed in biochemical assays of NXPE1 upon extended incubation with Neu5Ac and Ac-CoA (data not shown). Together, these results show that NXPE1 generates Neu5,9Ac$_2$ and smaller amounts of Neu5,7,9Ac$_3$, thus modulating the sialoglycome of colonic mucus. Given the recognized importance of mucus glycan modification in regulating the integrity of the intestinal

barrier (42, 43), our findings offer new insights by identifying the UC-associated gene NXPE1 as a novel enzyme that is required for acetylation of sialic acids on mucins.

DISCUSSION

We have discovered the reaction catalyzed by NXPE1, a protein with previously unknown function that was implicated by GWAS with a protective effect in UC patients. We find that NXPE1 is highly expressed in intestinal epithelial cells within the Golgi apparatus and that it is a serine hydrolase that can be captured via ABPP. NXPE1 catalyzes self-acetylation upon incubation with acetyl-CoA, a reaction that is diagnostic of an acetyltransferase. A substrate screen showed that NXPE1 transfers the acetyl group to the prevalent sialic acid Neu5Ac to generate the 9-*O*-acetyl product and, to a lesser extent, the 7,9-*O*-bisacetyl product. This reaction can modulate mucus viscosity in colonic organoids, a function that is abrogated in cells carrying the G353R variant, which we report is a loss-of-function allele.

The reaction catalyzed by NXPE1 and its loss-of-function variant are key to explaining its possible role in UC, a disorder that is associated with alterations of intestinal barrier function and dysbiosis of commensal flora. Mucus is produced in various body parts, and cells modulate mucus properties using several strategies. Our results herein suggest that acetylation of alcohol functionalities on sialic acids, which are subsequently appended onto mucins, is one such strategy that may modulate viscosity by altering mucus hydration and physical transitions between liquid and gel-like states. We propose that the UC-protective variant loses this ability and could thus produce a constitutively viscous mucus. In the context of UC, this can be beneficial as it produces a

more rigid barrier and keeps bacteria that contribute to inflammation at bay. To gain further insights into this process and further assess this model, understanding the functions of other NXPE paralogs will be critical.

The corollary of the proposed model is that inhibition of NXPE1 may be beneficial in UC, and the enzyme therefore provides a new, attractive target that directly addresses intestinal barrier disruption. Therapeutic interventions to restore barrier integrity represent a significant unmet need and offer opportunities for the development of first-in-class drugs. Following our learnings from genetics and the mechanistic basis for the protective effects of NXPE1 G353R, we hypothesize that small molecule inhibitors of NXPE1 may mimic this protective effect on barrier function. The spatially and temporally well-defined expression profiles of NXPE1 suggest that inhibition would not have broad adverse on-target effects, though this would need to be tested. Aside from pointing to a possible target for UC treatment, our results more broadly highlight the power of deorphanizing human enzymes to uncover new biological processes that may be targeted therapeutically.

METHODS

General Experimental Procedures. An Agilent Technologies 1200 series HPLC instrument coupled 6120 quadrupole MS equipped with an electrospray ionization (ESI) source was used to obtain low-resolution HPLC-ESI-MS data. NMR spectra to determine the structure of the NXPE1 product were collected at the Princeton University Department of Chemistry NMR Core facilities. 1D/2D NMR spectra were recorded on a Bruker Avance 800 MHz NMR spectrometer carrying a multipurpose cryoprobe. Semi-

preparative HPLC purifications for NMR experiments were performed on an Agilent 1290 Infinity Series HPLC system which was equipped with an automatic liquid sampler, a diode array detector, and an automated fraction collector.

Homology modelling. The three-dimensional structure of NXPE1 was predicted using Alpha Fold. To improve its stability, we strategically truncated a sequence at T67 at the N-terminal region with low folding confidence. Subsequently, we introduced a secretion signal from CrypA to enhance the folding and secretion properties of this NXPE1.

Expression and purification of NXPE1. Plasmids enabling the secreted expression of NXPE1 in Expi293 mammalian cells were based on a modified pRUSHmam vector (Invitrogen) carrying sequence stretches encoding an *N*-terminal signal peptide (CrypA) and a *C*-terminal TEV-TSFLAG tag. The sequence encoding residues E68–C547 of human NXPE1 was amplified by PCR. The resulting PCR products were ligated into the restriction sites of the modified pRUSHmam vector, resulting in the plasmids pRUSHmam-CrypA-T67-NXPE1-TEV-TS-FLAG, pRUSHmam-CrypA-T67-S355A-NXPE1-TEV-TS-FLAG, and pRUSHmam-Crypa-T67-G353RNXPE1-TEV-TS-FLAG. Expi293F human cells were grown at 37ºC in shaking cultures at 125 RPM in Expi293™ Expression Medium (Thermo, A1435101) and maintained at a density of 0.8–4 x 10^6 viable cells per ml.

Transient expression of WT NXPE1 and S355A-NXPE1 was performed using the ExpiFectamine™ 293 Transfection Kit system (Thermo Fisher), as described above. Conditioned medium was collected 96 h after infection, concentrated by centrifugation (5000g, 10 min, 4ºC). The supernatant was then supplemented with BioLock and

incubated for 30 min. It was then dialyzed against 50 mM sodium phosphate, 200 mM NaCl (pH 7.5) and passed over a Strep-Trap HP column (IBA). After washing with 20 column volumes of the same sodium phosphate buffer supplemented with 5% glycerol, bound protein was eluted by a desthiobiotin gradient. The desthiobiotin was subsequently removed by gel filtration using a Superdex 200 10/300 GL column equilibrated with 20 mM Tris-HCl, 200 mM NaCl, 5% glycerol, and 1 mM TCEP (pH 7.5).

Differential Scanning Calorimetry (DSF). All proteins (BCovHE, CASD1, WT NXPE1, and S355A-NXPE1) were used at a final concentration of 3 μM for this assay. SYPRO Orange (Invitrogen S6651) was used at a final concentration of 10X. Experiments were carried out in 20 mM Tris-HCl, 200 mM NaCl, 5% glycerol, 1 mM TCEP (pH 7.4). All DSF experiments were performed using an Eppendorf Realplex2 Mastercycler. Each sample was divided into four 10 μL replicates. Sample solutions were dispensed into 96-well optical reaction plates (Thermo Fisher Scientific 4306737), which were sealed with an optical PCR plate sheet (Thermo Fisher Scientific AB-1170). Fluorescence intensity was measured via the JOE emission filter (550 nm) and "PTS clear plate" was set as the background for the calibration. Temperature was continuously increased at a rate of 1°C/min. Melting curves were directly exported from the instrument and analyzed using Prism 6 (GraphPad Software Inc.).

In vitro serine hydrolase assay. Serine hydrolase activity was measured in a total volume of 50 μl containing in 20 mM Tris-HCl, 200 mM NaCl, 5% glycerol, 1 mM TCEP (pH 7.5). The reaction contained 10 μM protein (BCoV-HE, WT NXPE1 or S355A-NXPE1) and 1 mM ActiveX Desthiobiotin serine hydrolase probe (SHP, Sigma #88317). The samples were incubated overnight at 37ºC.

In vitro formation of the acetyl-enzyme intermediate. Purified WT NXPE1 or S355A-NXPE1-S355A (5 µg) was incubated for 15 min at 37°C with or without 2.5 mM acetyl-CoA (Sigma Aldrich) in a 50 µl reaction in Tris buffer (20 mM Tris-HCl, 200 mM NaCl, 5% glycerol, 1 mM TCEP, pH 7.5). After storage at 20°C for at least 18 h to stabilize the acetyl-enzyme intermediate, the reaction mixture was separated by SDS–PAGE, the protein band was excised, digested with trypsin, and analyzed by HPLC-ESI-MS.

In vitro sugar screen. Enzyme assays were performed for 3 h at 37°C in 60 µl containing 50 mM MES, 10 mM $MnCl_2$, 0.5 mM acetyl-CoA (pH 6.5) with or without 3.8 µM WT NXPE1 or S355A-NXPE1. The following sugars were tested as acceptors at a concentration of 0.5 mM: CMP-Neu5Ac, GDP-L-fucose, GDP-D-mannose, GDP-xylose, UDP-galactose, UDP-glucoronic acid, UDP-galuronic acid, UDP-*N*-acetylglucosamine, UDP-*N*-acetylgalactosamine, UDP-*N*-acetylazidoglucosamine, UDP-*N*-acetylazidogalactosamine, CMP-*N*-acetylazidoneuraminic acid, 9-azido-Neu5Ac, CMP-*N*-glycolylneuraminic acid, Neu5Ac, galactose, *N*-Acetyl-D-glucosamine, fucose, mannose, N-glycolylneuraminic acid (Chemily Glycoscience). After 3 h, the reaction was quenched by addition of 12 µl of 2 M propionic acid and incubation at 80ºC for 1 h. Propionic acid was used to reduce *O*-acetyl group loss/migration during this step. After 1 h, the quenched reaction was cooled to RT for 10 min, spun down to remove precipitated protein, and the sugars then labeled with DMB by addition of 9 µl of an 80 mM DMB stock (prepared in acetone), 5 µl of β-mercaptoethanol, and 5 µl of sodium dithionite (0.3 M stock prepared in water). The mixture was incubated for 2.5 h at 50ºC. It was then spun down, filtered, and analyzed by HPLC-qTOF-MS using an analytical Phenomenex

Polar C18 column (4.6 x 150 mm, 1.6 μm, 100Å) and a mobile phase consisting of water/acetonitrile supplemented with 0.1% (v/v) formic acid. Elution was carried out isocratically at 16% acetonitrile in water for 12 min at a flow rate of 0.15 ml/min. The amount of acylated product was determined by relative quantification via extraction of the desired product ion and integration of the area underneath the peak. The amount of product observed in the reaction of WT NXPE1 with Neu5Ac was set to 100%. The average of at least three independent experiments are reported for the substrate screen and other enzymatic assays.

Substrate specificity experiment. To characterize the preferred substrate of NXPE1, a competition experiment was carried out where the exact same reaction conditions were used as described above, except that NXPE1 was incubated with equimolar concentrations of three substrates that gave the highest turnover in the substrate screen. These were Neu5Ac, CMP-Neu5Ac, and UDP-xylose. Because DMB treatment results in the same product from Neu5Ac and CMP-Neu5Ac, isotopically labeled Neu5Ac (Neu5Ac-$^{13}C,d_3$) was used to differentiate it from CMP-Neu5Ac. The reaction was worked up and the amount of product determined as described above in the substrate screen.

Purification of DMB-Neu5,9Ac$_2$. To determine the position of acetylation by NXPE1, large scale (8 ml) experiments were carried using the same conditions as describe above in the substrate screen. After formation of the DMB adduct, the sample was lyophilized to dryness, resuspended in 400 μl of 50% MeOH/Water and then subjected to semi-preparative HPLC purification using a Kinetex C18 column (Phenomenex, 5 μm, 10 x 250 mm). DMB-Neu5,9Ac$_2$ eluted at a retention time of 26

min using an 80 min gradient at 2 ml/min starting with 5% acetonitrile in water and ending at 45% acetonitrile in water. The product was lyophilized to dryness and then subjected to 1D/2D NMR analysis.

Human colonic organoid culture. Endoscopic biopsy samples were obtained from healthy donors enrolled in the Prospective Registry in IBD Study at MGH (PRISM). Written informed consent was obtained from study participants and the study protocols were reviewed and approved by the Mass General Brigham Human Research Committee (#2004-P-001067). Human colon spheroids were established and maintained as previously described (44) with a slight modification. Briefly, biopsy samples were washed with PBS containing penicillin (100 units/ml), streptomycin (0.1 mg/ml) and gentamicin (50 ug/ml) for 5 min at room temperature. After three washings with PBS, tissues were incubated with 5 mM EDTA in PBS for 45 mins at 4°C. After wash with PBS, several pipetting yielded supernatants enriched in colonic crypts. Crypts were pelleted by centrifugation at 200 g for 5 mins and then plated in 15 ul of Matrigel (Corning) and maintained in 500 ul of 50% L-WRN conditioned media supplemented with 10 uM Y-27632 (R&D systems, cat# 1254) and 10 uM SB431542 (Tocris Bioscience, 1614). Medium was changed every 2-3days and organoids were passaged to 1: 3 using TrypLE every 7 days.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R01 AI109725, P30 DK043351, and RC2 DK135492 to R.J.X. and R35 GM152049 to M.R.S.), the Crohn's and Colitis Foundation, the Center for Microbiome Informatics and Therapeutics, and the

Klarman Cell Observatory. This work was supported by a GSK-Broad Institute research collaboration (D.B.G., R.J.X.). We thank members of the GSK team, including Nadia Eckert, Sunil Nagpal, Aaron Cheng, Karen Simpson, Tharini Selvakumar and Silvia Gonzalez-Del Valle for their scientific discussions regarding this work. We thank Alok Jaiswal and Lingjia Kong for analysis of NXPE1 expression and Theresa Reimels for editorial assistance.

REFERENCES

1. M. H. Weisman, *et al.*, Inflammatory Bowel Disease prevalence: Surveillance data from the U.s. national Health and Nutrition Examination Survey. *Prev. Med. Rep.* **33**, 102173 (2023).
2. D. B. Graham, R. J. Xavier, Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* **578**, 527–539 (2020).
3. D. R. Plichta, D. B. Graham, S. Subramanian, R. J. Xavier, Therapeutic opportunities in inflammatory bowel disease: Mechanistic dissection of host-microbiome relationships. *Cell* **178**, 1041–1056 (2019).
4. H. H. Uhlig, F. Powrie, Translating immunology into therapeutic concepts for inflammatory bowel disease. *Annu. Rev. Immunol.* **36**, 755–781 (2018).
5. L. Jostins, *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
6. H. Huang, *et al.*, Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).

7. K. M. de Lange, *et al.*, Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
8. J. Z. Liu, *et al.*, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
9. Z. Liu, *et al.*, Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat. Genet.* **55**, 796–806 (2023).
10. B. Khor, A. Gardet, R. J. Xavier, Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).
11. A. C. Anderson, S. Stangherlin, K. N. Pimentel, J. T. Weadge, A. J. Clarke, The SGNH hydrolase family: a template for carbohydrate diversity. *Glycobiology* **32**, 826–848 (2022).
12. J. Z. Long, B. F. Cravatt, The metabolic serine hydrolases and their functions in mammalian physiology and disease. *Chem. Rev.* **111**, 6022–6063 (2011).
13. H. Chepyshko, C.-P. Lai, L.-M. Huang, J.-H. Liu, J.-F. Shaw, Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (Oryza sativa L. japonica) genome: new insights from bioinformatics analysis. *BMC Genomics* **13**, 309 (2012).
14. C. C. Akoh, G.-C. Lee, Y.-C. Liaw, T.-H. Huang, J.-F. Shaw, GDSL family of serine esterases/lipases. *Prog. Lipid Res.* **43**, 534–552 (2004).
15. A.-M. T. Baumann, *et al.*, 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. *Nat. Commun.* **6**, 7673 (2015).

16. R. Nason, *et al.*, Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* **12**, 4070 (2021).

17. I. Surolia, *et al.*, Functionally defective germline variants of sialic acid acetylesterase in autoimmunity. *Nature* **466**, 243–247 (2010).

18. J. M. Pickard, M. Y. Zeng, R. Caruso, G. Núñez, Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. *Immunol. Rev.* **279**, 70–89 (2017).

19. E. Theodoratou, *et al.*, The role of glycosylation in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **11**, 588–600 (2014).

20. A. Di Sabatino, G. Santacroce, C. M. Rossi, G. Broglio, M. V. Lenti, Role of mucosal immunity and epithelial-vascular barrier in modulating gut homeostasis. *Intern. Emerg. Med.* **18**, 1635–1646 (2023).

21. Y. H. Sheng, S. Z. Hasnain, T. H. J. Florin, M. A. McGuckin, Mucins in inflammatory bowel diseases and colorectal cancer. *J. Gastroenterol. Hepatol.* **27**, 28–38 (2012).

22. Y. H. Jeong, *et al.*, Rare exonic minisatellite alleles in MUC2 influence susceptibility to gastric carcinoma. *PLoS One* **2**, e1163 (2007).

23. J. A. Grondin, Y. H. Kwon, P. M. Far, S. Haq, W. I. Khan, Mucins in intestinal mucosal defense and inflammation: Learning from clinical and experimental studies. *Front. Immunol.* **11**, 2054 (2020).

24. M. C. Visschedijk, *et al.*, Pooled resequencing of 122 ulcerative colitis genes in a large Dutch cohort suggests population-specific associations of rare variants in MUC2. *PLoS One* **11**, e0159609 (2016).

25. K. Bergstrom, *et al.*, Proximal colon-derived O-glycosylated mucus encapsulates and modulates the microbiota. *Science* **370**, 467–472 (2020).

26. W. X. Shi, R. Chammas, A. Varki, Induction of sialic acid 9-O-acetylation by diverse gene products: implications for the expression cloning of sialic acid O-acetyltransferases. *Glycobiology* **8**, 199–205 (1998).

27. A. Guzman-Aranguez, P. Argüeso, Structure and biological roles of mucin-type O-glycans at the ocular surface. *Ocul. Surf.* **8**, 8–17 (2010).

28. J.-F. Sicard, G. Le Bihan, P. Vogeleer, M. Jacques, J. Harel, Interactions of intestinal bacteria with components of the intestinal mucus. *Front. Cell. Infect. Microbiol.* **7** (2017).

29. S. Jafaripour, *et al.*, A21 role of sialic acid o-acetylation in maintaining mucus integrity and homeostasis of the colonic mucosa. *J. Can. Assoc. Gastroenterol.* **6**, 11–12 (2023).

30. Z. Wu, *et al.*, Characterization of O-acetylation in sialoglycans by MALDI-MS using a combination of methylamidation and permethylation. *Sci. Rep.* **7** (2017).

31. M. R. Kudelka, S. R. Stowell, R. D. Cummings, A. S. Neish, Intestinal epithelial glycosylation in homeostasis and gut microbiota interactions in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 597–617 (2020).

32. C. S. Smillie, *et al.*, Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).

33. UniProt. Available at: https://www.uniprot.org/uniprotkb/Q8N323/entry [Accessed 11 November 2024].

34. A. E. Speers, B. F. Cravatt, Activity-based protein profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT mass spectrometry. *Curr. Protoc. Chem. Biol.* **1**, 29–41 (2009).

35. P. B. Rosenthal, *et al.*, Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* **396**, 92–96 (1998).

36. Q. Zeng, M. A. Langereis, A. L. W. van Vliet, E. G. Huizinga, R. J. de Groot, Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9065–9069 (2008).

37. E. S. Rangarajan, *et al.*, Structural and enzymatic characterization of NanS (YjhS), a 9-O-Acetyl N-acetylneuraminic acid esterase from Escherichia coli O157:H7. *Protein Sci.* **20**, 1208–1219 (2011).

38. R. Zhong, D. Cui, Z.-H. Ye, Evolutionary origin of O-acetyltransferases responsible for glucomannan acetylation in land plants. *New Phytol.* **224**, 466–479 (2019).

39. S. Sharma, M. Ahmed, Y. Akhter, Fungal acetyltransferases structures, mechanisms and inhibitors: A review. *Int. J. Biol. Macromol.* **157**, 626–640 (2020).

40. J. Song, *et al.*, Reassembled biosynthetic pathway for a large-scale synthesis of CMP-Neu5Ac. *Mar. Drugs* **1**, 34–45 (2003).

41. A. K. Münster, *et al.*, Mammalian cytidine 5'-monophosphate N-acetylneuraminic acid synthetase: a nuclear protein with evolutionarily conserved structural motifs. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9140–9145 (1998).

42. A. S. Luis, G. C. Hansson, Intestinal mucus and their glycans: A habitat for thriving microbiota. *Cell Host Microbe* **31**, 1087–1100 (2023).

43. Y. Yao, *et al.*, Mucus sialylation determines intestinal host-commensal homeostasis. *Cell* **185**, 1172-1188.e28 (2022).

44. T. Sato, *et al.*, Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

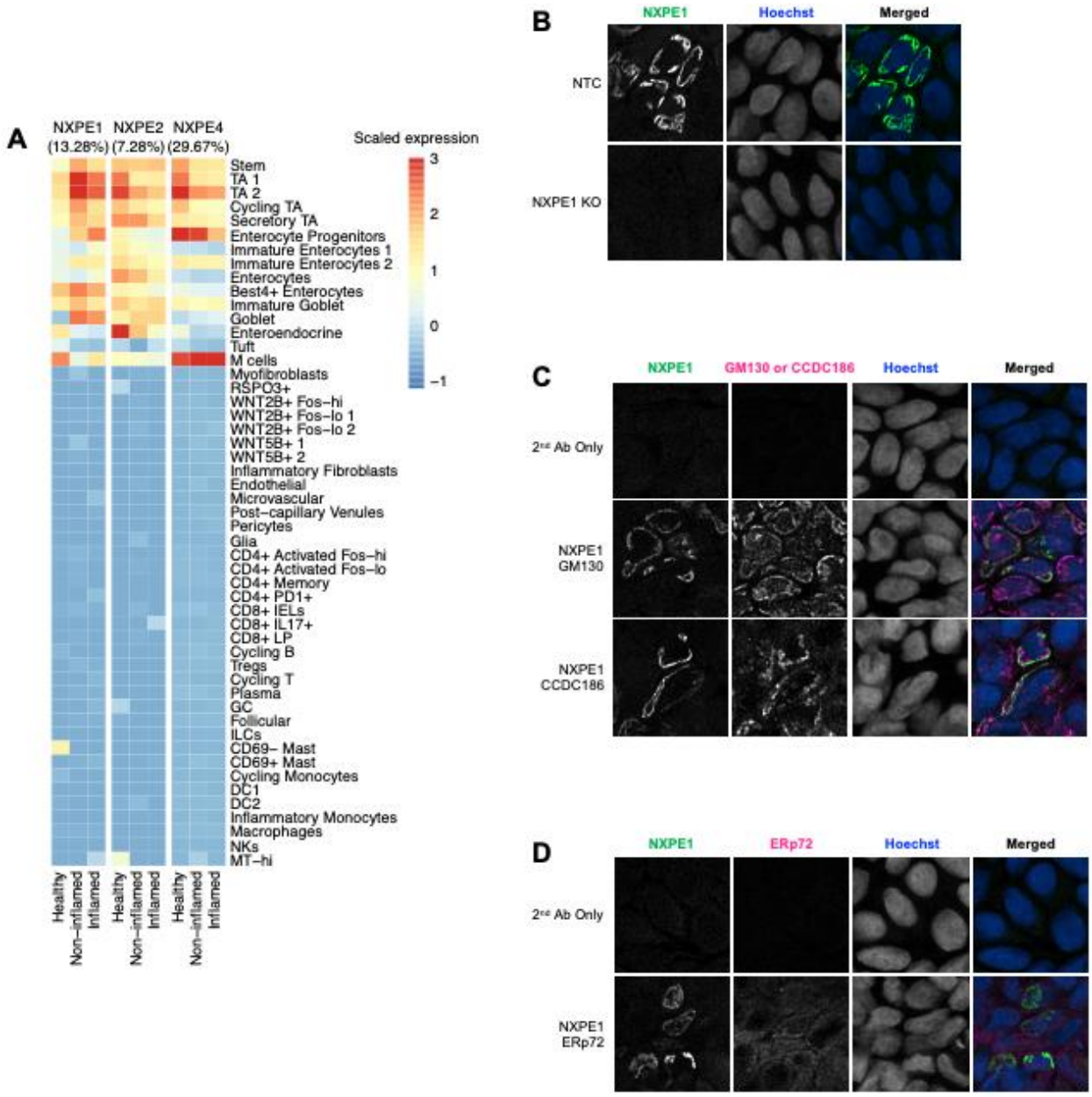


Figure 1. NXPE1 is expressed in intestinal epithelial cells within the Golgi apparatus. (**A**) Heatmap showing expression levels of NXPE1, NXPE2, and NXPE4 from human colon biopsies (32). Percentages in parentheses indicate the mean percent of epithelial cells in which the transcript is detected. (**B**) Detection of NXPE1 with monoclonal antibodies in human colonic organoids grown in monolayers. Staining conducted for DNA (Hoechst) and NXPE1 is shown in organoids after CRISPR/Cas9

modification with nontargeting control (NTC) or single guide RNA (sgRNA) targeting NXPE1 (KO). (**C**) NXPE1 co-localization with Golgi markers GM130 and CCDC186 in human colonic organoids. (**D**) NXPE1 co-localization stain with the ER marker ERp72 in human colonic organoids.

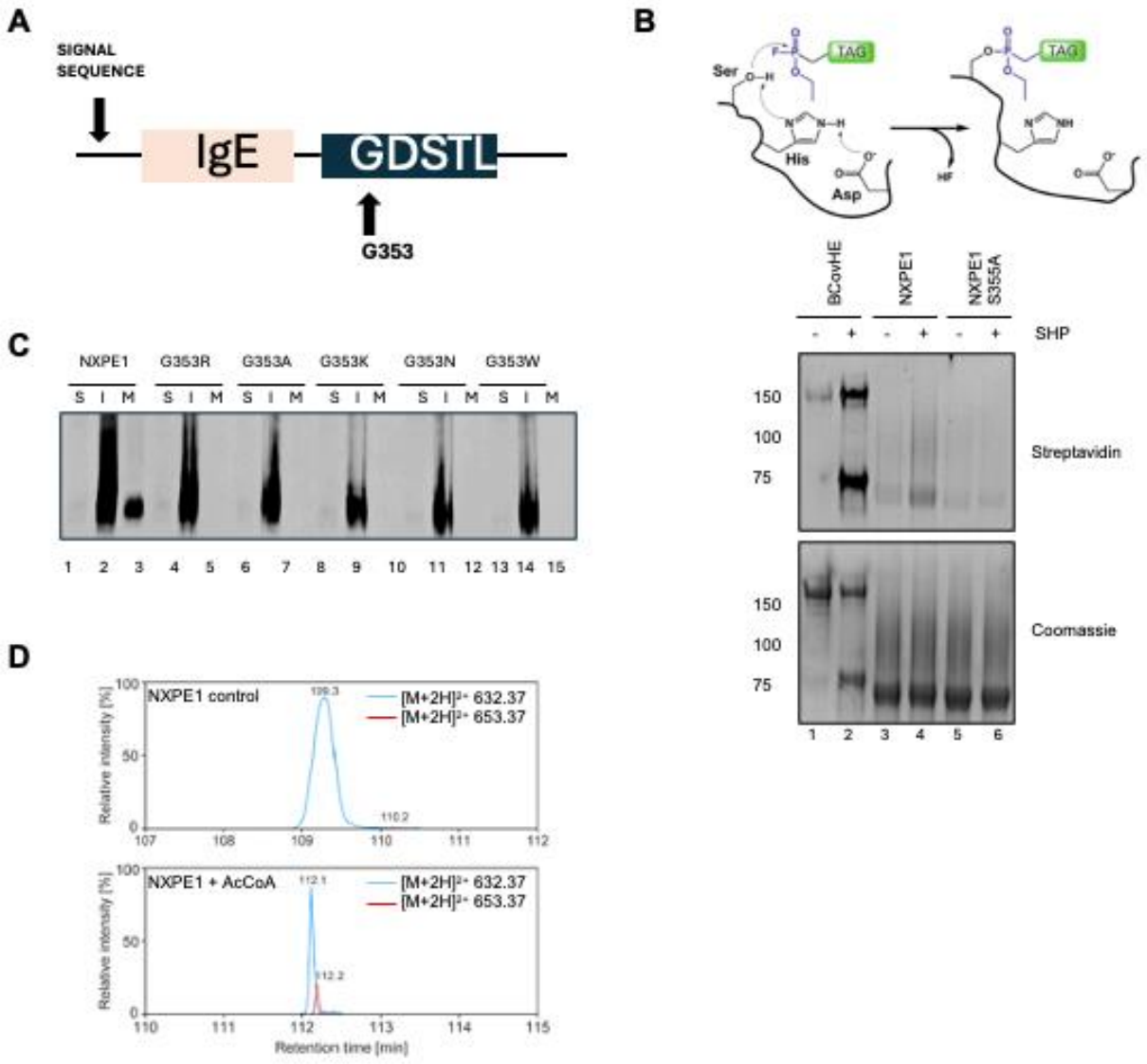


Figure 2. NXPE1 is an *O*-acetyltransferase and member of the serine hydrolase family. (**A**) Domain structure of NXPE1. Sequence homology suggests that NXPE1 is a serine hydrolase of the GDSL/SGNH family with a characteristic GDSTL active site sequence. The UC-protective variant (G353R) is highlighted. S355 is the active site covalent catalyst. (**B**) (Top) Schematic of the fluorophosphonate ABPP probe and reaction with NXPE1. (Bottom) The well-known serine hydrolase BCovHE (bovine coronavirus esterase, positive control), NXPE1, and NXPE1 S355A were incubated with

and without the serine hydrolase probe (SHP). Active-site labeling was then monitored using a streptavidin-based Western blot. Native NXPE1, but not the S355A mutant, can be captured with the SHP. (**C**) Assessing soluble protein expression with the G353 variants. Only the endogenous protein is soluble and secreted into the medium ('M'), whereas the G353R/A/K/N/W variants are insoluble ('I') and not observed in the soluble crude extract ('S') nor secreted into the medium. (**D**) Formation of a covalent acetyl-enzyme intermediate captured by MS. Purified NXPE1 and S355A mutant were incubated with or without acetyl-CoA, subjected to SDS-PAGE, trypsin treatment, and analysis by HPLC-ESI-MS. Acetyl-adduct is only observed with native NXPE1 in the presence of Ac-CoA.

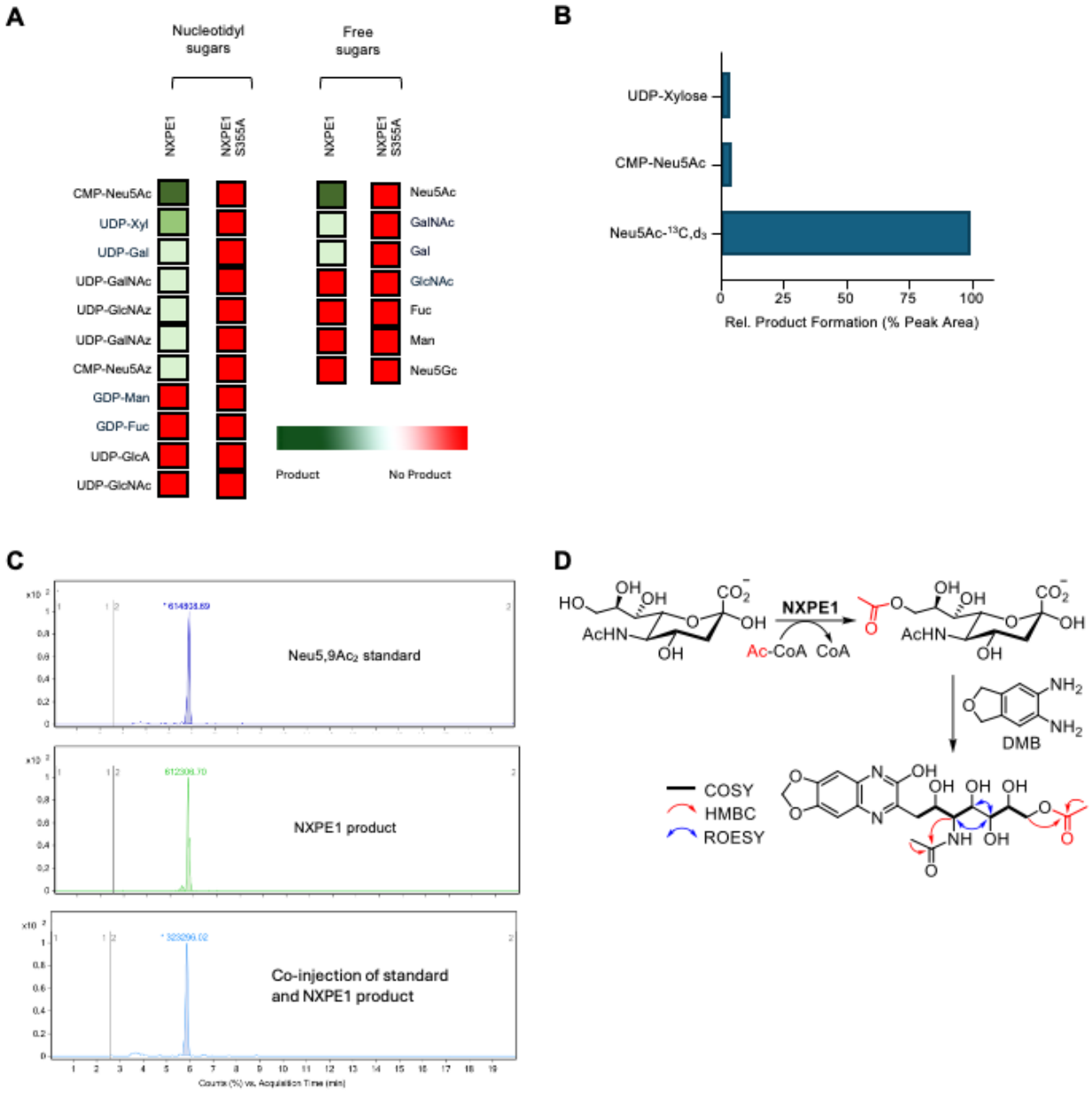


Figure 3. NXPE1 catalyzes 9-*O*-acetylation of Neu5Ac. (**A**) NXPE1 and S355A mutant were screened against 18 acceptor substrates as indicated. The amount of acetylated product was then quantified using HPLC-qTOF-MS analysis, as indicated by the color bar. The data are normalized to the reaction of NXPE1 with Neu5Ac, which showed the highest turnover. (**B**) NXPE1 substrate specificity was determined by incubating it with equimolar concentrations of UDP-xylose, CMP-Neu5Ac, and Neu5Ac-^{13}C,d$_3$. The

isotopically labeled form of Neu5Ac was used to distinguish it from CMP-Neu5Ac. Product formation was quantified using HPLC-qTOF-MS, which is shown in the bar graph and is normalized to Neu5Ac-$^{13}C,d_3$ (set to 100%). The average of three independent measurements are shown. (**C**) HPLC traces of DMB-labeled authentic Neu5,9Ac_2 (top), the NXPE1 product (middle), and co-injection of the two samples. Coelution indicates the compounds are similar. (**D**) Reaction catalyzed by NXPE1 and analysis by DMB. NMR-based structure elucidation of the reaction product is shown with key NMR correlations highlighted.

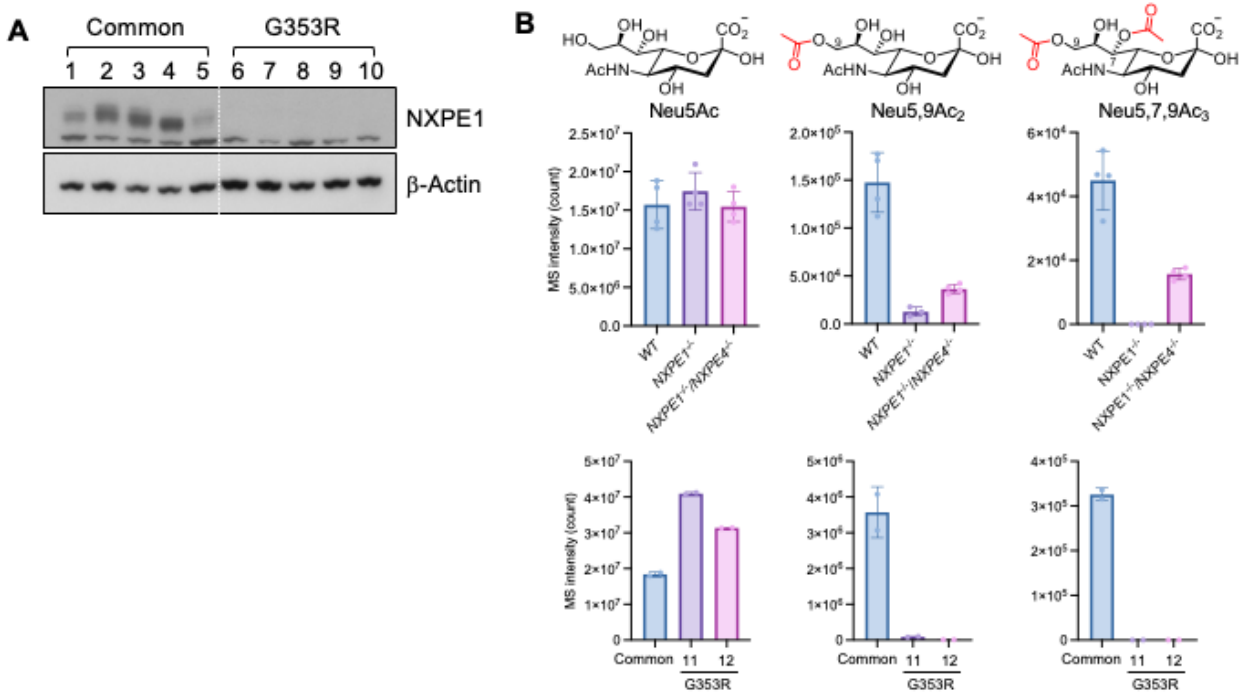


Figure 4. NXPE1 acetylates Neu5Ac in mucus from human colonic organoids, and G353R variant is a loss-of-function allele. (**A**) Immunoblot of NXPE1 using differentiated human intestine organoid lysates obtained from patients with the common allele of NXPE1 or UC-protective variant G353R. (**B**) Sialic acid content of organoids expressing endogenous NXPE1 or the G353R variant. Sialic acids were released from the outer layer secreted mucus and their distribution determined by HPLC-qTOF-MS. The sialic acid monitored is shown above each column. The average of five independent experiments is reported; error bars represent S.D.

SI APPENDIX

The Ulcerative Colitis-Associated Gene NXPE1 Catalyzes Glycan Modifications on Colonic Mucin

Ranad Humeidi[1,2,3], Noriko Rapley[2,3], Xiebin Gu[2,3], Joon Soo An[4], Ashwin N. Ananthakrishnan[5,6], Elizabeth A. Creasey[3], Mark J. Daly[2,7], Stuart L. Schreiber[2,8], Daniel B. Graham[2,3,9,*], Mohammad R. Seyedsayamdost[4,10,*], Ramnik J. Xavier[2,3,9,*]

[1]Program for Chemistry & Chemical Biology, Harvard Medical School, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Department of Molecular Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[4]Department of Chemistry, Princeton University, Princeton, NJ, USA

[5]Department of Medicine, Harvard Medical School, Boston, MA, USA

[6]Division of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA

[7]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[8]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA, USA

[9]Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[10]Department of Molecular Biology, Princeton University, Princeton, NJ, USA

*Correspondence: dgraham@broadinstitute.org; mrseyed@princeton.edu; xavier@molbio.mgh.harvard.edu

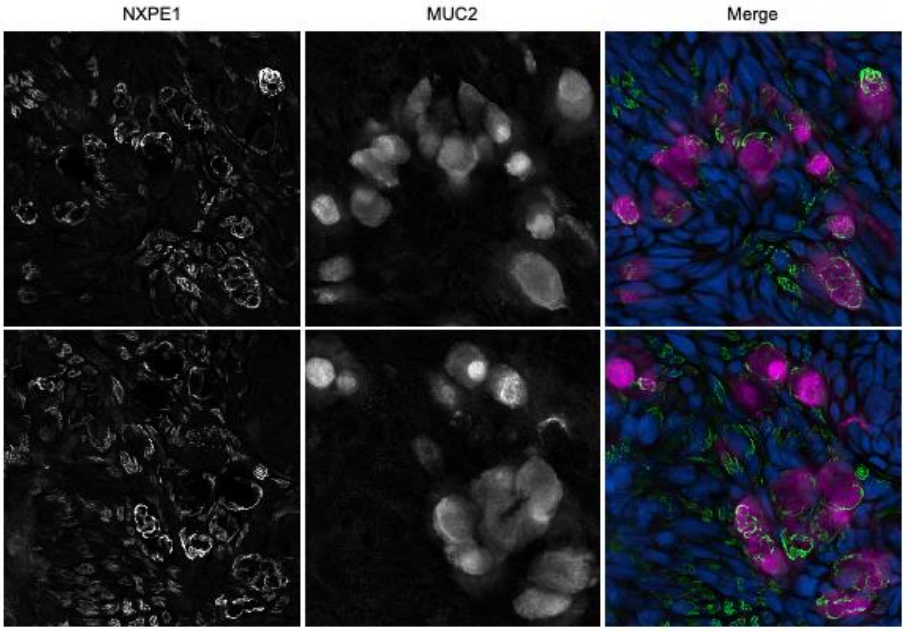


Figure S1. Characterization of NXPE1 and MUC2 expression in human colonic organoids. Immunocytochemistry of NXPE1 (Green), MUC2 (Magenta), and DNA (Blue) in human colonic organoids. NXPE1 is expressed in Goblet cells and other epithelial cells.

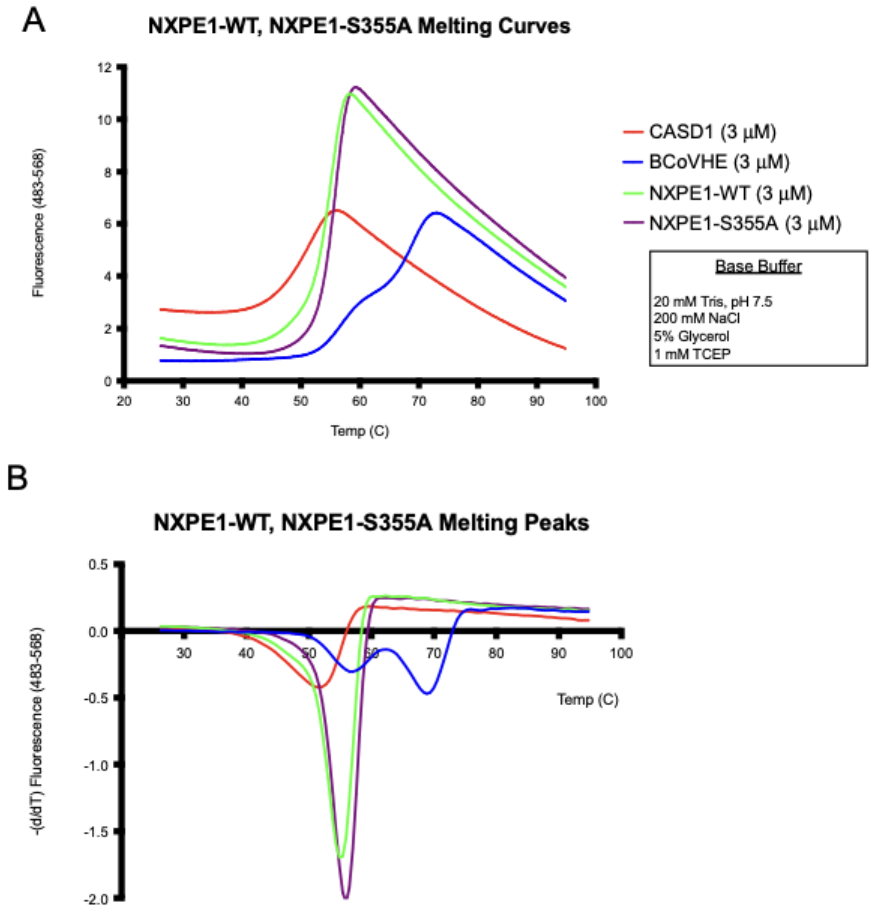


Figure S2. Melting curve assays of NXPE1-wt and NXPE1-S355A. Melting curve analysis (A) and melting peaks (B) for NXPE1-WT and NXPE1-S355A as determined by differential scanning fluorimetry. CASD1 and the bovine coronavirus esterase (BCovHe) are shown as controls. All constructs are stable at ambient temperatures.

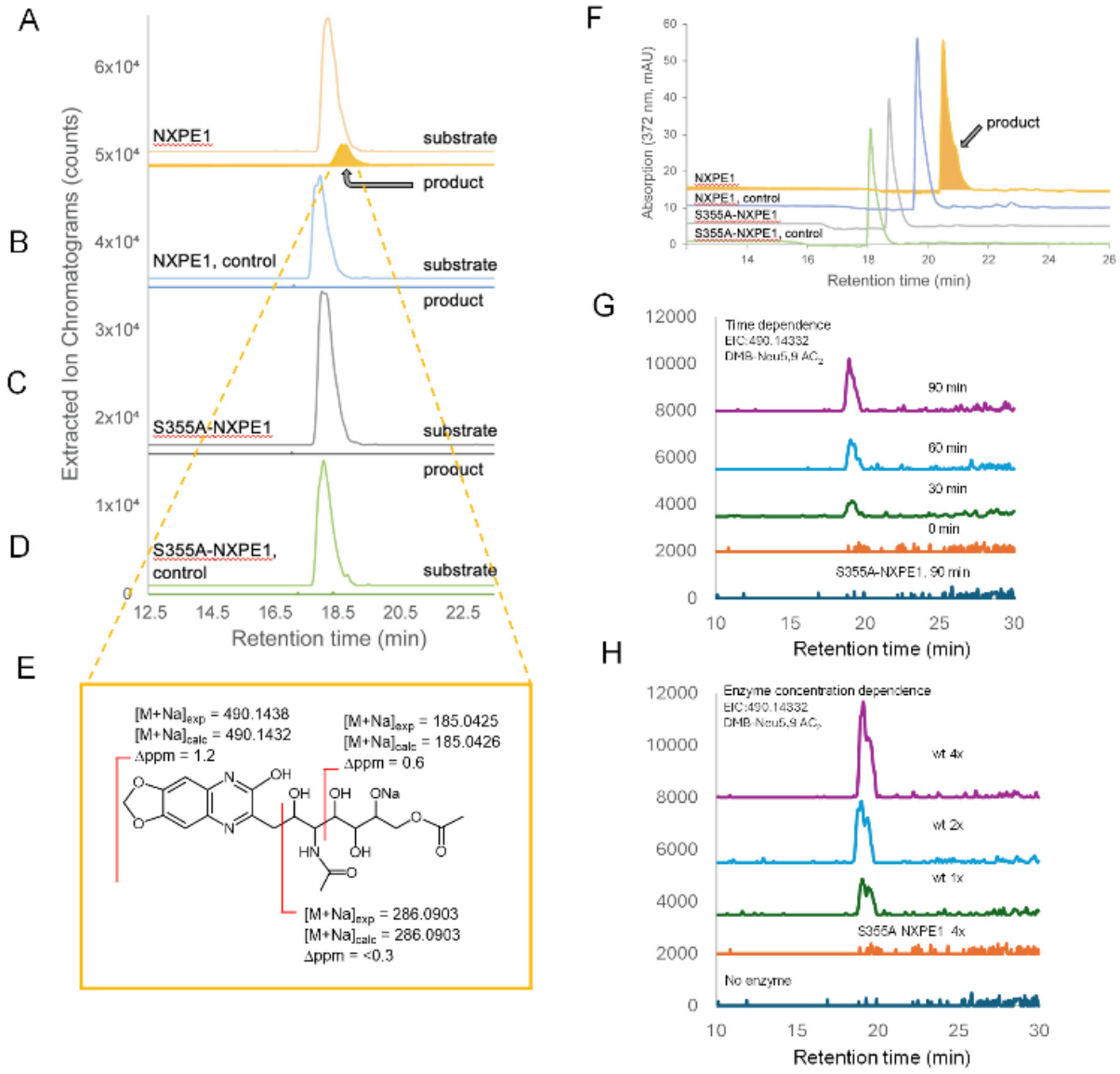


Figure S3. Enzymatic assays with NXPE1. (A-E) HPLC-qTOF-MS elution profiles of DMB-derivatized samples obtained after incubation of Neu5Ac and acetyl-CoA with wt NXPE1 or S355A-NXPE1. Shown are extracted ion chromatograms for substrate and product in the reaction with wt NXPE1 (A), the associated no-enzyme control (B), the reaction with S355A-NXPE1 (C), and the associated no-enzyme control (D). The inset (E) shows results from HR-MS/MS analysis of the product and the detected fragments in the reaction with wt NXPE1. Product is only observed in panel (A). (F) Reactions in panel (A-D) monitored by the UV-vis absorption using the 372 nm feature of the DMB-derivatized product. As in (A-D), product is only observed in the reaction with wt NXPE1. (G) Time-dependent DMB-Neu5,9Ac$_2$ (product) formation, as determined by extracted ion chromatography (EIC), with wt NXPE1 monitored after 0, 30, 60, and 90 min of reaction time. The blue profile shows the EIC trace for product with NXPE1-S355A after 90 min. (H) Enzyme-dependent DMB-Neu5,9Ac$_2$ (product) formation, as determined by EIC at various enzyme concentrations. No product is observed in the no-enzyme control nor in the reaction with 4x S355A-NXPE1.

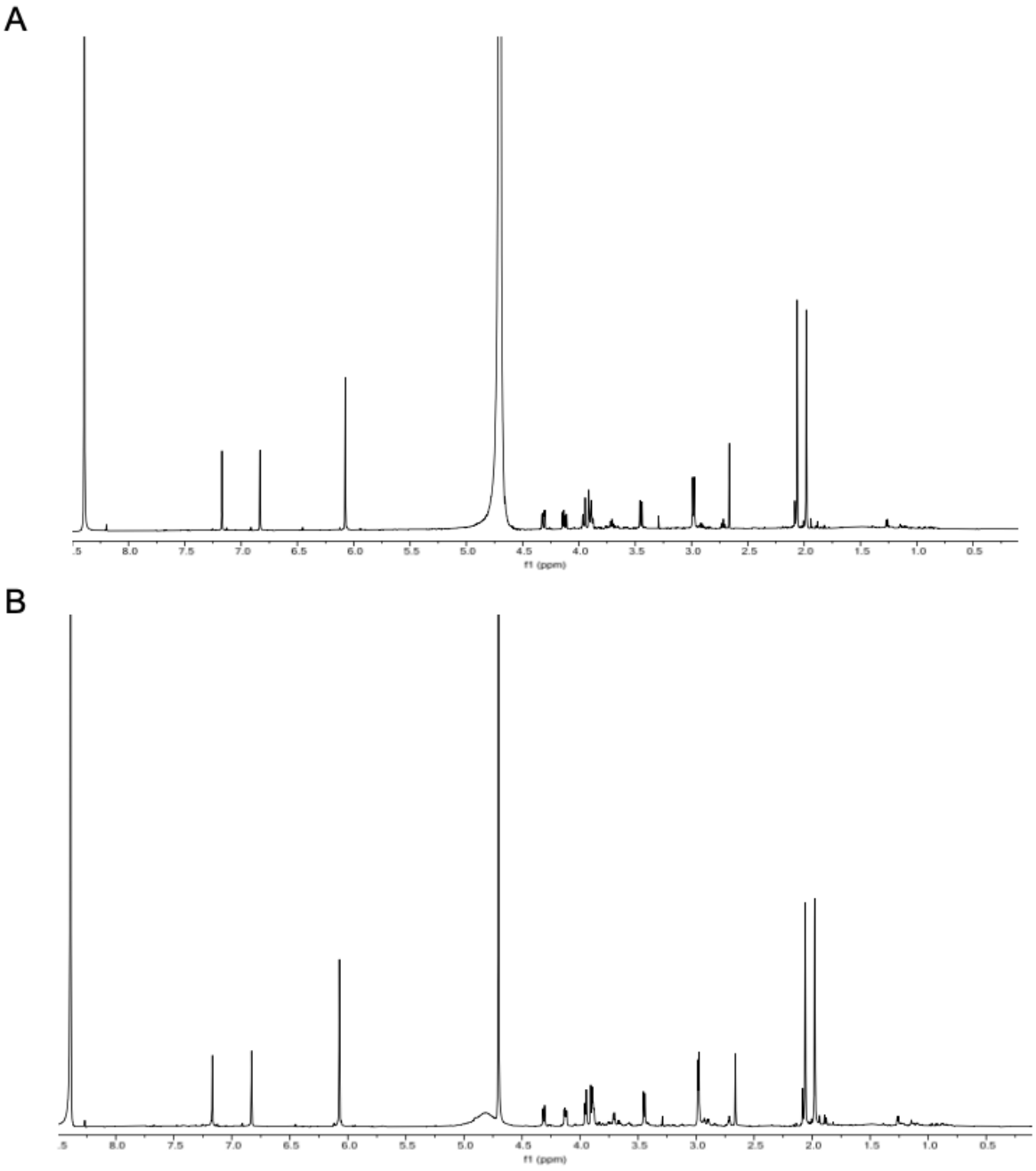


Figure S4. ^{1}H NMR spectra of the NXPE1 product. (A) ^{1}H NMR spectrum of NXPE1 product at 500 MHz in D_2O immediately after purification. (B) ^{1}H NMR spectrum of NXPE1 product at 800 MHz in D_2O after two days. No degradation or acetyl group migration is observed.

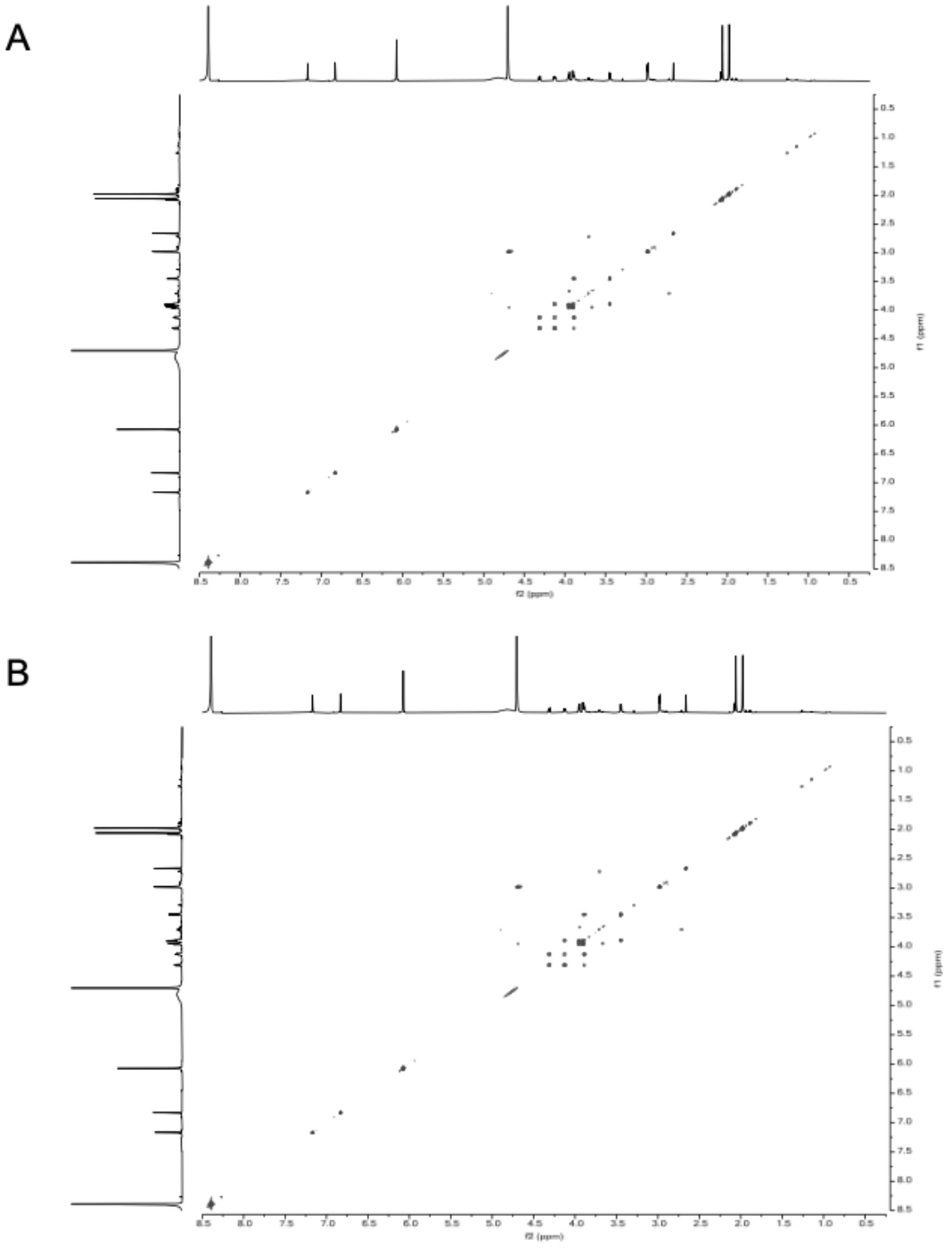


Figure S5. COSY and HSQC NMR spectra of the NXPE1 product. (A) COSY NMR spectrum of NXPE1 product at 800 MHz in D_2O. (B) HSQC NMR spectrum of NXPE1 product at 800 MHz in D_2O.

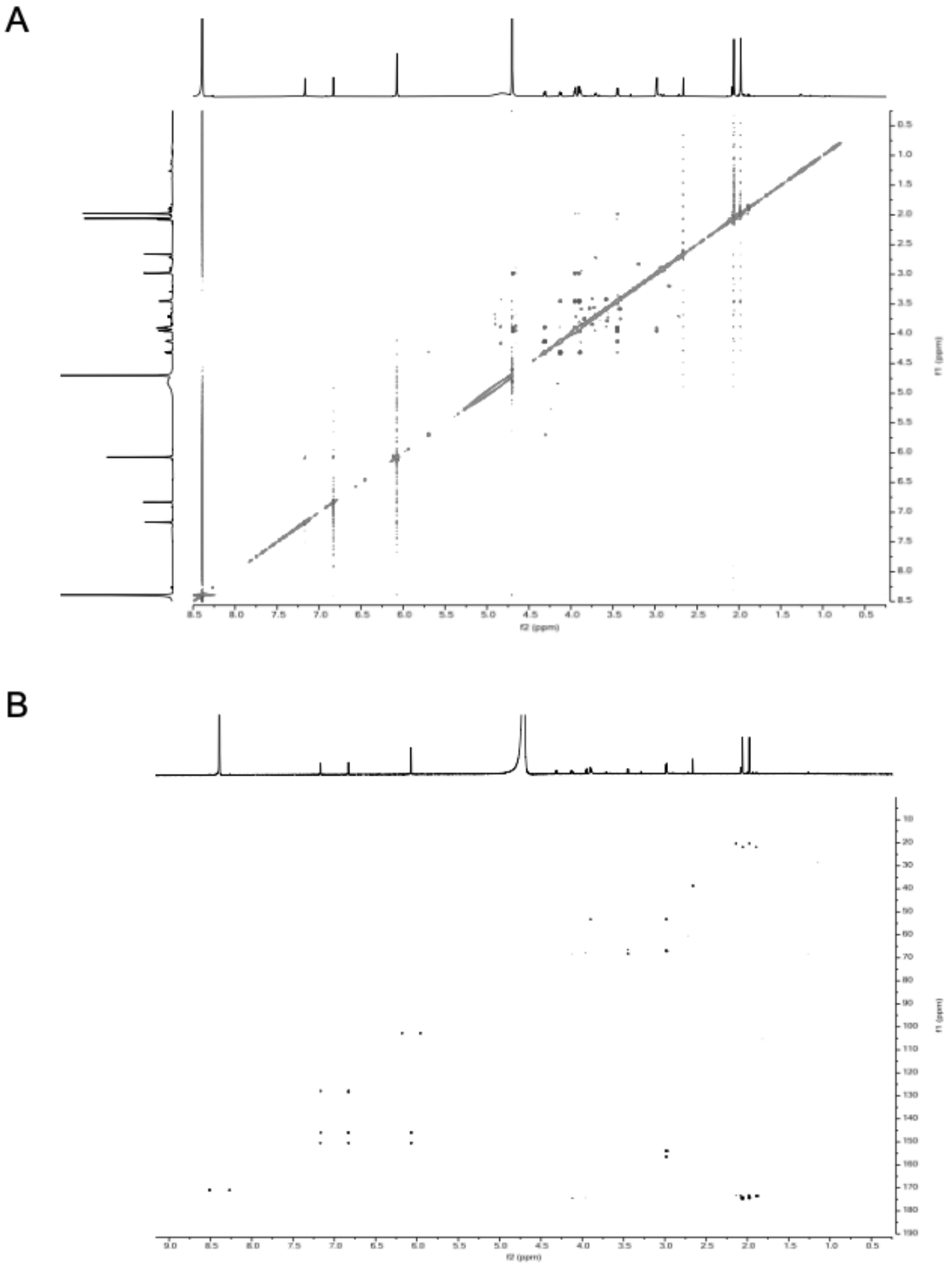
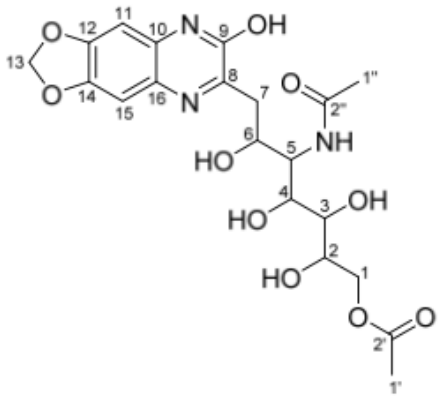


Figure S6. ROSY and HMBC NMR spectra of the NXPE1 product. (A) ROESY NMR spectrum of NXPE1 product at 800 MHz in D_2O. (B) HMBC NMR spectrum of NXPE1 product at 800 MHz in D_2O.

Table S1. NMR ^{1}H and ^{13}C NMR data of DMB-Neu5,9Ac_2 in D_2O.[a] The structure and numbering scheme for DMB-Neu5,9Ac_2 are shown below the table.

position	δ_C, type	δ_H, mult (J in Hz)
1	66.3, CH_2	4.31, dd (11.5, 1.7)
		4.12, dd (11.5, 5.7)
2	68.2, CH	3.88, m
3	69.0, CH	3.44, d (9.5)
4	67.6, CH	3.90, m
5	53.3, CH	3.95, m
6	66.9, CH	4.67, m
7	37.6, CH_2	2.98, d (6.5)
8	154.0, C	
9	156.4, C	
10	128.3, C	
11	105.4, CH	7.17, s
12	150.5, C	
13	102.6, CH_2	6.07, s
14	146.0, C	
15	95.1, CH	6.83, s
16	127.7, C	
1'	20.2, CH_3	2.05, s
2'	174.2, C	
1''	21.7, CH_3	1.97, s
2''	174.5, C	

[a] ^{1}H NMR spectrum was recorded at 800 MHz and the ^{13}C chemical shifts were determined by comprehensive analysis of ^{1}H, HSQC, and HMBC spectra, which were recorded at 800 MHz.



3. Early finding of NXPE4 Characterization

3.1 Introduction

To investigate the biological processes driving IBD, our lab uses genomic screening and molecular biology approaches to identify specific single nucleotide polymorphisms (SNPs) that are associated with altered risk of IBD and investigate the molecular mechanisms and functions of these mutations. In the last years, genome-wide association studies (GWAS) identified thousands of single-nucleotide polymorphisms (SNPs) associated with complex diseases and have been particularly successful in CD and UC. In Chapter 2, I go into great details of how we characterized one of these variants. Another coding variant identified in the same family of genes is NXPE4. The Manhattan plot in (**Figure 1**) displays a significant genetic association signal in the chromosomal region containing NXPE1, NXPE2, and NXPE4. NXPE1 coding variant has a clear peak of association that surpasses the genome-wide significance threshold (indicated by the horizontal green line). The x-axis represents the chromosomal position within the 11q region, while the y-axis shows the -log10(P-

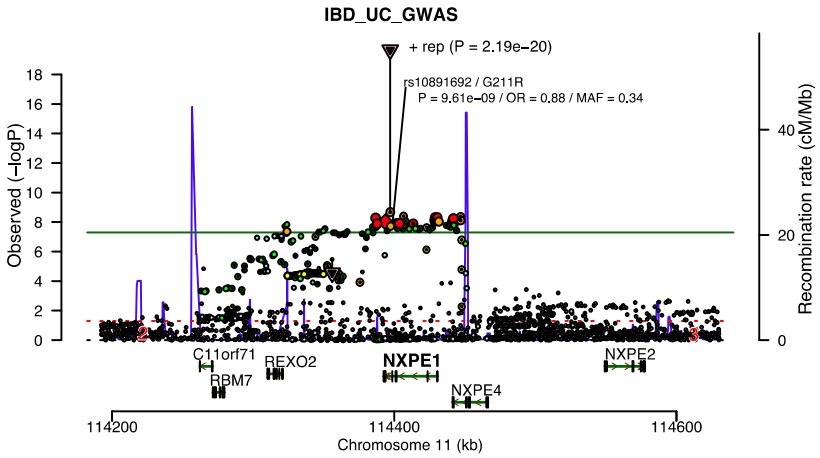


Figure 1. Exome sequencing in GWAS studies of NXPE1, NXPE4 and NXPE2. Each dot represents a single-nucleotide polymorphism (SNP). Y axis represents strength of their association measured as $-\log_{10}$ transformed P values. Green line marks genome-wide significance threshold of $P < 5 \times 10^{-8}$, X axis indicates chromosome position along human chromosome 11.

value) of the association statistics. don’t see the same statistical associations with the homologs NXPE4 and NXPE2.

3.2 Characterizing the effect of this pathway in animal models

A common method to investigate the function of specific enzymes involves the use of genetically modified mice, in which the target gene is either partially inactivated or completely deleted, a process known as generating a knock-out (1). By observing the phenotypic and molecular changes that occur in these mice, the role of the disrupted gene can often be inferred. Such models allow for controlled exploration of gene function in a living system, providing insights into complex biological processes (1,2).

However, the use of mice as a model system comes with inherent limitations. One major challenge lies in the differences between human genes and their mouse orthologs. The gene ortholog in mice may not fully replicate the function of its human counterpart, which can limit the relevance of the findings (3,4). Additionally, compensatory mechanisms in mice can obscure the functional consequences of knocking out a gene. For instance, related genes may upregulate their expression or assume the function of the disrupted gene, masking the true phenotypic impact of its loss(2,5). This redundancy can make it difficult to identify the direct contributions of the targeted gene to specific biological processes.

Another significant obstacle is determining how and where to assess the effects of gene disruption in the mouse. Choosing which tissues to analyze is a critical decision, as phenotypic changes may only be observable in specific cell types or tissues(1).

In the specific case of NXPE1, the gene poses unique challenges when using mice as a model system. NXPE1 exists as a pseudogene in mice, meaning it is non-functional and cannot be directly studied in vivo through conventional knock-out models. This lack of functional expression in mice complicates efforts to investigate its biological role. However, mice express a closely related gene, NXPE4, which belongs to the same family of enzymes and shares significant sequence similarities with NXPE1. NXPE4 may offer an alternative for studying related enzymatic functions in mice, to infer the potential role of NXPE1 in a different context.

Despite these challenges, genetically modified mice remain a powerful tool for exploring gene function. While certain limitations are unavoidable, careful experimental design, such as leveraging related genes like NXPE4, can help overcome these obstacles and provide valuable insights into the functional roles of genes such as NXPE4. This chapter will cover early findings of NXPE4 characterization *in vivo* and *in vitro*.

3.2.1 Restricted spatial expression of NXPE4

To investigate the tissue-specific expression of NXPE4, we leveraged the comprehensive spatial transcriptomic dataset generated in our group, as detailed in Mayassi et al. (Nature, 2024). This dataset provides a high-resolution spatial and transcriptional map

of the murine intestine, constructed using the Swiss roll technique to maintain the native spatial orientation of the tissue. Specifically, the dataset includes over 138,000 spatially resolved spots from 46 Swiss roll preparations, spanning both the small intestine and colon, enabling detailed analysis of gene expression under steady-state and perturbed conditions.

Using this dataset, we analyzed NXPE4 expression across the murine colon. The Swiss roll technique provided a unified coordinate system for data integration and visualization, ensuring spatially accurate reconstruction of gene expression patterns. Our analysis (**Figure 2**) revealed a distinct spatial distribution of NXPE4 expression across three biological replicates. Expression levels were visualized as normalized gradients, with high expression regions highlighted in yellow and low expression regions in purple. Notably, the expression pattern localized NXPE4 predominantly in the ileum showing

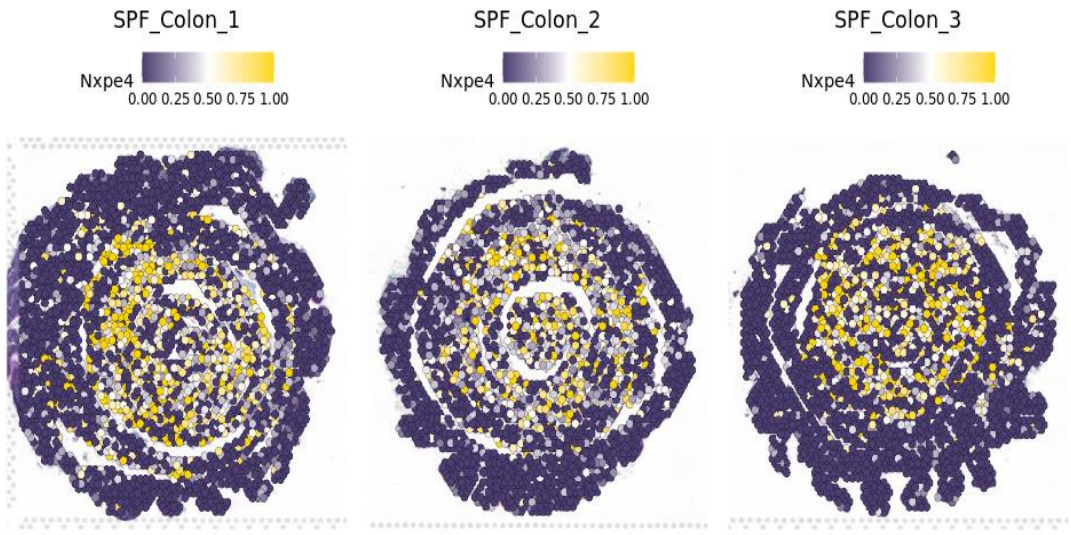


Figure 2. Heatmap showing spatial expression patterns of NXPE4 across different regions of a Swiss rolled mouse colon. Columns represent individual biological replicates.

selective spatially restricted expression in the colonic architecture.

Building on the findings of NXPE4's spatially restricted expression in the murine colon, we next examined its expression in human tissues using the GTEx dataset. In humans, NXPE4 exhibits similarly selective expression within the gastrointestinal tract with its highest levels observed in the colon (**Figure 3**).

Within the colonic epithelium, NXPE4 is enriched in specialized epithelial cell populations, Goblet cells, primary producers of mucus. Similarly to NXPE, the conserved, organ-specific expression pattern suggests a critical role for NXPE4 in maintaining mucosal barrier integrity through the regulation of glycosylation and mediating host-microbiota interactions. These findings show the importance of characterizing the function of NXPE4 to better understand how it's function can impact colonic homeostasis and

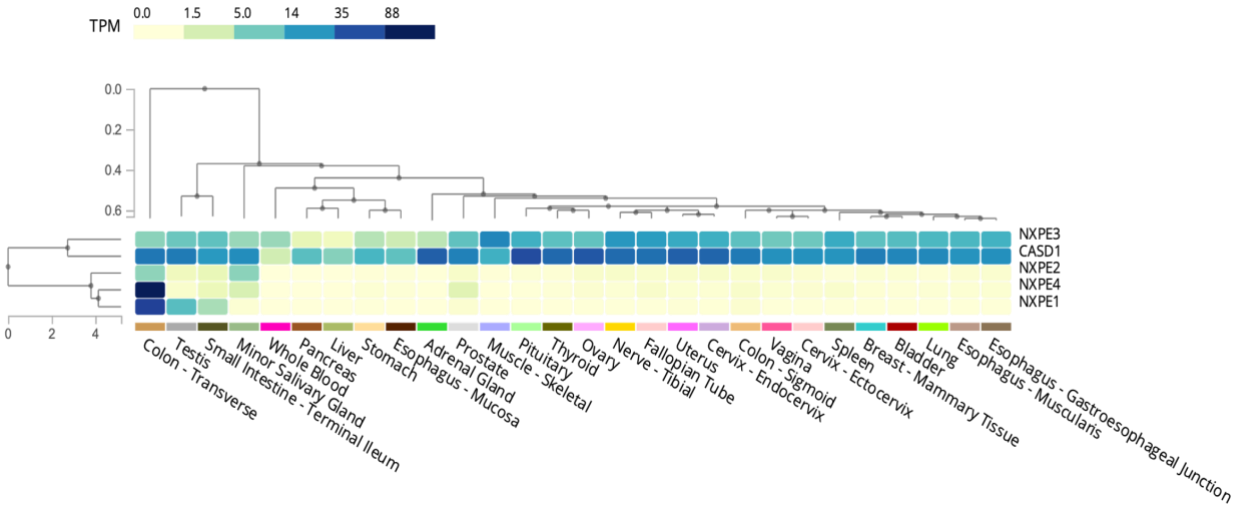


Figure 3. NXPE4 is specifically expressed in the GI tract in humans. RNA expression of NXPE genes in the indicated human tissues. Data processing and normalization by GTEx portal. TPM, transcripts per million.

barrier function.

3.2.2 Creating knockout mice

The CRISPR/Cas9 system has revolutionized the creation of knockout mice (6). This method uses the Cas9 endonuclease and a guide RNA (gRNA) to introduce precise double-strand breaks at the target gene locus(6). These breaks are typically repaired by non-homologous end joining (NHEJ), often resulting in small insertions or deletions (indels) that disrupt the gene's coding sequence. Key advantages of the CRISPR/Cas9 method include:

1. Direct delivery into mouse embryos via microinjection or electroporation, bypassing the need for ES cell intermediates
2. Significantly faster production of knockout mice, often in a single generation
3. Ability to target multiple genes simultaneously through multiplexing with multiple gRNAs

It is then crucial to confirm the intended gene modification. Once validated, the mice are bred to establish stable lines for phenotypic characterization and experimentation. Gene knockout mouse models continue to play a central role in functional genomics and translational medicine, providing invaluable insights into gene function and disease mechanisms.

3.2.3 Methods of isolating colonic crypts

NXPE4 shows high expression profile in the epithelial layer of colon tissue; to characterize the endogenous enzyme product of NXPE4 in vivo here we extracted the sialic acids from colonic crypts and compared the acetylation levels of the sugars in WT versus KO mice.

To isolate colonic crypts from a mouse colon, the procedure begins with careful preparation of the tissue. First, the colon is removed and opened longitudinally to expose the inner surface. It is then thoroughly washed with phosphate-buffered saline (PBS) to remove debris and contaminants. The tissue is cut into small fragments, typically around 5 mm in size, and rinsed in cold PBS multiple times to ensure cleanliness. Next, the tissue fragments are incubated in PBS containing 2 mM ethylenediaminetetraacetic acid (EDTA) for 45 to 60 minutes at 4°C. EDTA helps to chelate calcium ions, which disrupts cell-cell adhesion and facilitates the detachment of the epithelial layer from the underlying tissue. Following incubation, the tissue is vigorously pipetted using a 10 mL pipette to mechanically dislodge the crypts from the tissue.

The suspension containing the crypts is then passed through a 70 µm cell strainer to separate the crypts from larger tissue fragments. The isolated crypts are collected by centrifugation at 900 × g for 5 minutes, resulting in a pellet of crypt cells. These crypts can be further processed for applications such as RNA isolation, quantitative RT-PCR.

3.2.4 Methods of isolating Sialic acids from colonic crypts

The obtained crypt sample was then incubated with neuraminidase for 4 hours to ensure specific release of Sialic acids. The released sugars were then labeled with DMB

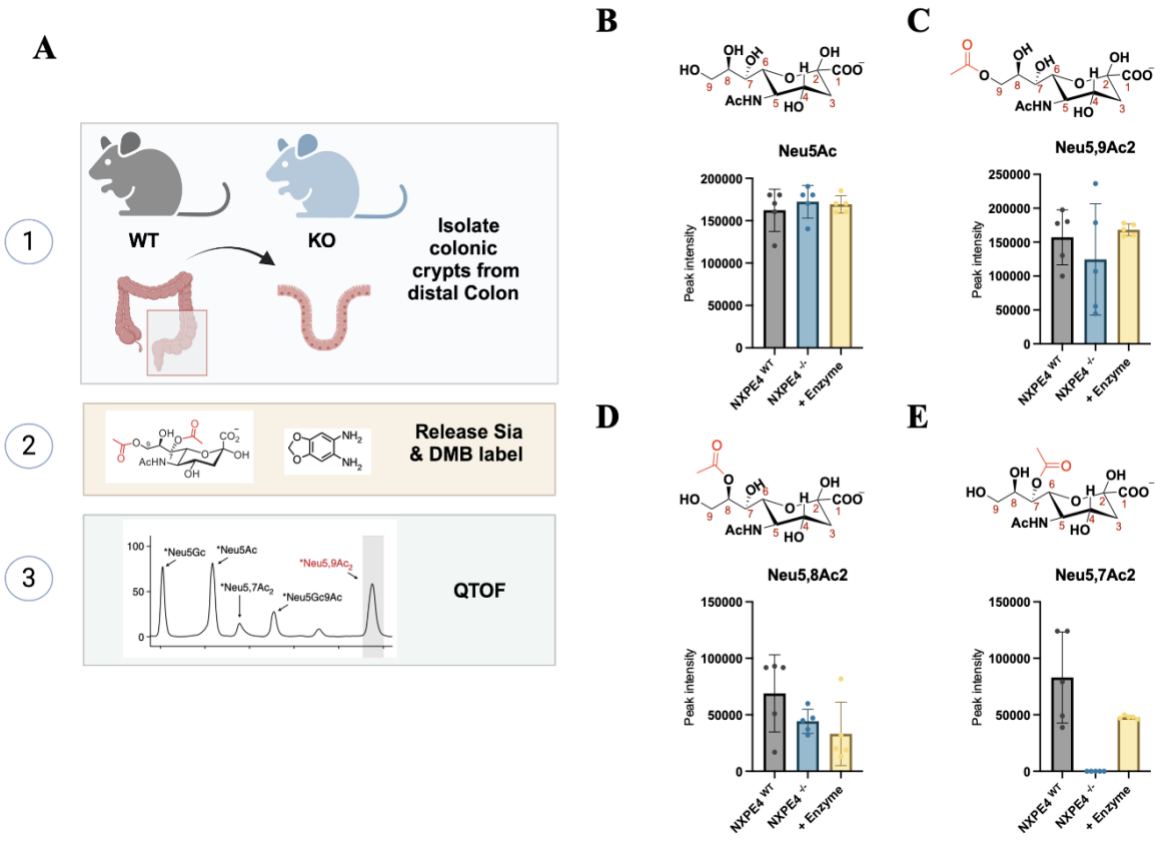


Figure 4. Quantification of Sialic Acid Acylation in Colonic Crypts of NXPE4 KO and WT Mice. A) Experimental workflow: colonic crypts were isolated from the distal colon of wild-type (WT) and NXPE4 knockout (KO) mice. Released sialic acids (Sia) were derivatized with 1,2-diamino-4,5-methylenedioxybenzene (DMB) for fluorescent labeling and quantified using liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC-QTOF). B-E) Peak intensities of DMB-derivatized Neu5Ac, Neu5,9Ac2, Neu5,8Ac2, and Neu5,7Ac2, measured in WT, NXPE4 KO. NXPE4 KO mice. Error bars represent standard deviation (SD).

and analyzed on Quadruple time-of-flight (QTOF) MS. We were able to isolate and detect singly and doubly acetylated forms of Sialic acids from the in vivo sample and compare

the mass and retention time of both were compared to the sample obtained from in vitro reaction. Further experiments are required to identify the exact O-acetylation position.

3.2.5 Preliminary Evidence Suggests NXPE4's Role In Vivo is Production of 7-O-Acetyl-Neu5Ac in Colonic Crypts.

To investigate the role of NXPE4 in sialic acid acetylation, we analyzed the sialic acid profiles of colonic crypts isolated from 10 NXPE4 knockout (KO) mice and 10 wildtype (WT) controls. Using acid release followed by high-resolution Q-TOF mass spectrometry, we identified significant differences in the acetylation patterns between the two groups (**Figure 4**). In WT mice, a distinct peak corresponding to 7-O-acetyl-Neu5Ac was consistently detected, along with peaks for 8- and 9-O-acetyl-Neu5Ac. In contrast, the 7-O-acetyl-Neu5Ac peak was completely absent in NXPE4 KO mice, while other acetylated species, such as 8- and 9-O-acetyl-Neu5Ac, remained present.

To confirm the enzymatic activity of NXPE4, we added purified NXPE4 and acetyl-CoA to sialic acids extracted from KO mice. This restored the production of 7-O-acetyl-Neu5Ac, demonstrating that NXPE4 catalyzes the formation of this specific acetylated sialic acid species in vivo. These findings reveal that NXPE4 is required for the generation of 7-O-acetyl-Neu5Ac, highlighting its potential role in modulating mucosal barrier functions and host-microbiota interactions.

3.3 Microbiota-Driven Regional Gene Expression and Sialic Acid Modifications in the Murine Colon

Gene expression within specific regions of the murine colon is profoundly influenced by the presence of gut microbiota. As demonstrated in Mayassi et al. (Nature, 2024), the spatial transcriptional landscape of the intestine adapts to microbial colonization in a region-specific manner, with certain genes and cellular populations showing distinct microbiota-driven patterns.

Previous studies have established a link between gut microbiota and sialic acid catabolism, often mediated through microbial sialidases that cleave sialic acids from host glycoproteins as part of a foraging mechanism(7,8). These sialic acids, key components of colonic mucus, serve as a nutrient source for the microbial community and influence microbial composition. In response, the host regulates sialic acid-modifying enzymes to preserve the integrity of the mucus layer and minimize nutrient loss. To explore how sialic acid modifications are influenced by gut microbiota, we isolated sialic acids from the colonic crypts of germ-free (GF) and specific pathogen-free (SPF) mice and compared their acetylation levels. Our analysis revealed that GF mice exhibited increased sialic acid acetylation compared to SPF mice (**Figure 5**), suggesting that certain microbial species in SPF conditions actively cleave acetyl group modifications. This observation points to a potential host adaptation to microbial colonization aimed at preserving sialic acid bioavailability and regulating host-microbe interactions.

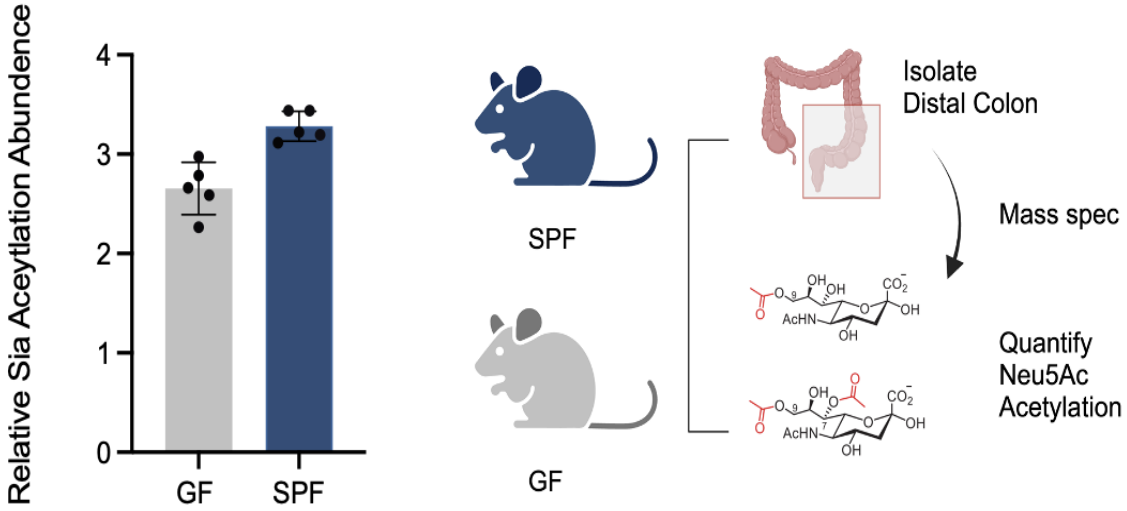


Figure 5. Neu5Ac modification abundance in germ-free (GF) and specific pathogen-free (SPF) mice. Sialic acid acetylation was assessed by DMB-HPLC analysis, amount of DMB-Neu5,9Ac2 (product) and DMB-Neu5Ac (substrate) quantified by integrating the corresponding peak areas. The area under the peaks was calculated from the extracted ion chromatograms (EICs) for $[H^{+}+Na^{2+}]$, and data are presented as mean ± SEM for biological replicates.

3.4 Microbiota-Dependent Spatial Expression of NXPE4 and Its Role in Mucosal Integrity

NXPE4 is predicted to play a role in these protective mechanisms by acting on sialic acids, potentially modifying them in ways that make them less accessible to microbial sialidases. This hypothesis is supported by the observed microbiota-dependent regulation of NXPE4 expression, which mirrors trends seen in other sialic acid-modifying enzymes. Its robust expression in SPF mice, compared to the minimal expression in GF mice, suggests a functional adaptation to microbial colonization, aimed at maintaining mucosal integrity and regulating host-microbe interactions. To test this this hypothesis, we examined the expression of NXPE4, under germ-free (GF) and specific pathogen-free (SPF) conditions using the data generated by Mayassi et al. (Nature, 2024). NXPE4 displayed robust and spatially localized expression in SPF mice, predominantly in the middle and distal colon (**Figure 6**) suggesting a microbiota-dependent regulation.

These findings align with previous observations that the microbiota drives adaptations in the colonic mucosa, including enhanced mucus secretion and structural cell specialization this data underscores NXPE4's potential role in maintaining mucosal integrity and supporting the colonic environment in the presence of commensal microbes.

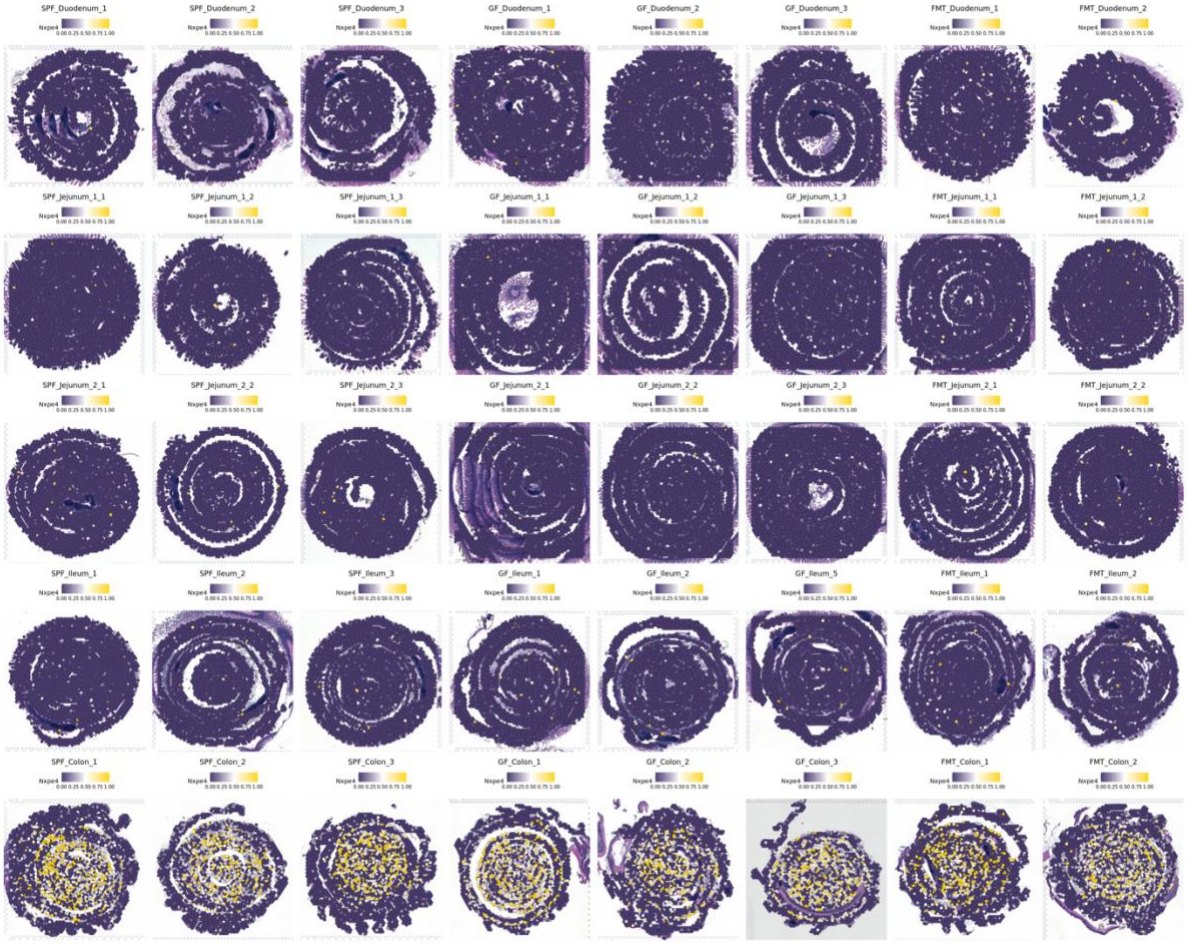


Figure 6. Heatmap showing spatial expression patterns of NXPE4 across regions of Swiss-rolled mouse colons from specific pathogen-free (SPF) and germ-free (GF) mice. Columns represent individual biological replicates, and rows represent distinct colonic regions, from proximal to distal. Color intensity indicates transcript expression levels, with darker shades representing higher expression. Spatial distribution highlights region-specific gene expression in SPF mice.

.

3.5 Distinction between NXPE4 expression profile to its related genes

To investigate the functional specialization of NXPE family genes and their potential roles in maintaining colonic mucosal integrity, we compared the expression profiles of NXPE2, NXPE3, NXPE4, and CASD1 across various cell types in the murine colon. NXPE4, in particular has been implicated in sialic acid modification and host-microbiota interactions, making it essential to assess how its expression compares to other NXPE family members and CASD1, a known acetyltransferase involved in sialic acid biology.

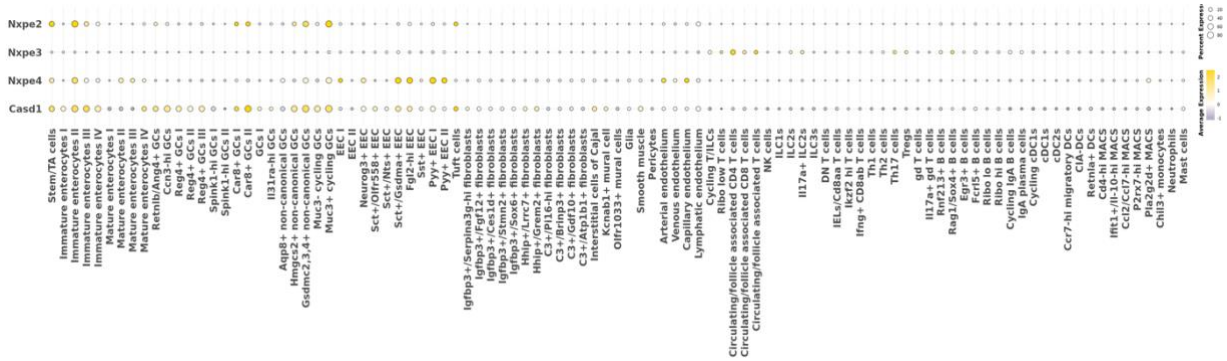


Figure 7. Dot plot showing average expression level and percent expression of NXPE gene family in individual cell types of mouse colon.

The analysis (**Figure 7**) revealed that NXPE4 exhibits a highly selective expression profile, predominantly localized to structural cells, including epithelial cell populations. This is in stark contrast to NXPE2 and NXPE3, which display broader expression patterns

across multiple immune and stromal cell types. CASD1, as expected, demonstrates widespread expression, spanning both epithelial and immune compartments.

The selective enrichment of NXPE4 in epithelial cells supports its hypothesized role in maintaining epithelial integrity and regulating host-microbiota interactions, possibly through sialic acid modification. In contrast, the broader distribution of NXPE2, NXPE3, and CASD1 suggests their involvement in diverse cellular processes within the colonic microenvironment. This comparative analysis underscores the unique functional specialization of NXPE4 relative to other NXPE family members and highlights its potential as a key regulator of colonic mucosal function.

To investigate the functional specialization of NXPE family genes and their potential roles in maintaining colonic mucosal integrity, we compared the expression profiles of NXPE2, NXPE3, NXPE4, and CASD1 across various cell types in the murine colon. NXPE4 has been implicated in sialic acid modification and host-microbiota interactions, making it essential to assess how its expression compares to other NXPE family members and CASD1, a known acetyltransferase involved in sialic acid biology (**Figure 8)**.

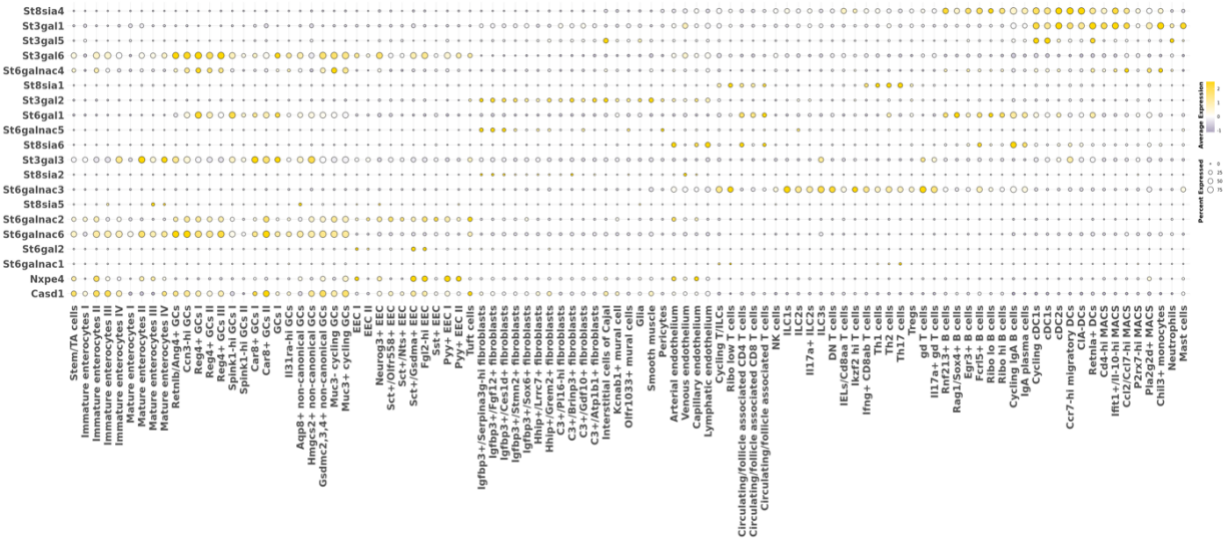


Figure 8. Dot plot showing average expression level and percent expression of NXPE4, CASD1 and Sialyltransferases across distinct goblet cell subtypes, as labeled on the x-axis. Heat map represents expression level, dot size represents percent expressed.

Given the critical role of goblet cells in producing and maintaining the colonic mucus layer, we specifically analyzed the expression of NXPE4, CASD1, and sialyltransferases across various goblet cell subsets in the murine colon. Goblet cells are essential for mucosal barrier integrity and are central to sialic acid-mediated interactions, making them a key focus for understanding the functional relevance of these enzymes.

This analysis revealed that NXPE4 shows selective expression in specific goblet cell subsets, particularly cycling goblet cells (e.g., Muc3+ and Muc2+ cycling goblet cells), suggesting a role in regulating mucus integrity and function. In contrast, CASD1, a known acetyltransferase, displays broader expression across multiple goblet cell types, indicating its general involvement in sialic acid acetylation. Sialyltransferases, such as ST6Gal1 and

ST3Gal1, exhibit cell-type-specific patterns, with high expression in certain cycling and non-canonical goblet cells. Additionally, specialized sialyltransferases, like ST8Sia1, are enriched in distinct goblet cell subsets, pointing to their roles in subset-specific functions (**Figure 9**).

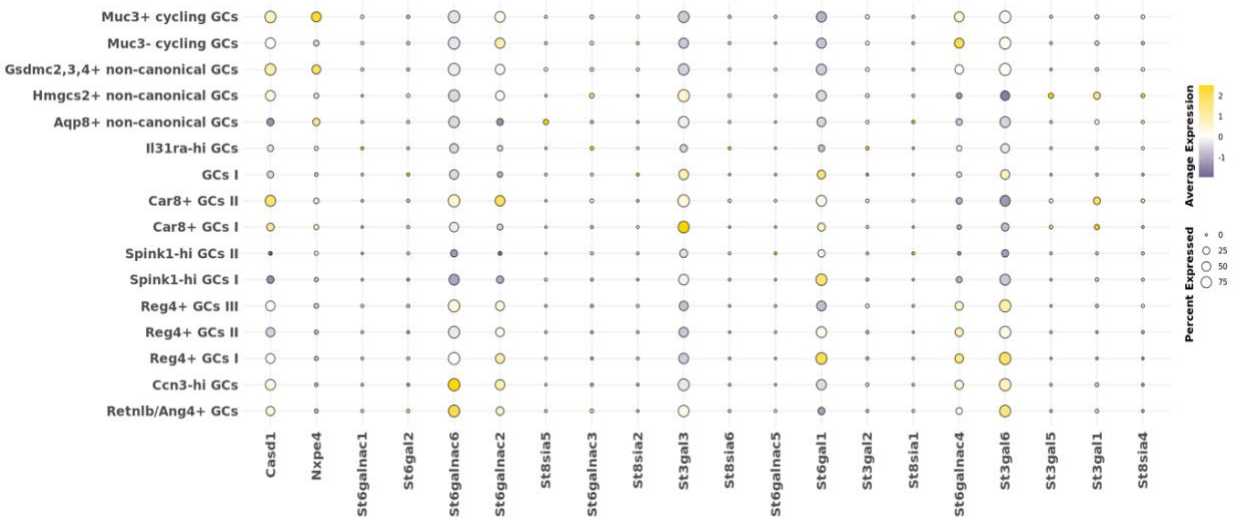


Figure 9. NXPE4 Expression levels in Goblet cell subtypes. Dot plot showing average expression level and percent expression of NXPE4, CASD1 and Sialyltransferases across distinct goblet cell subtypes, as labeled on the x-axis. Heat map represents expression level, dot size represents percent expressed.

These findings emphasize the heterogeneous expression of sialic acid-modifying enzymes within goblet cells, highlighting a division of labor among subsets in mucus production, sialic acid modification, and adaptation to host-microbiota interactions. The selective enrichment of NXPE4 in goblet cells underscores its specialized role in supporting epithelial and mucosal homeostasis.

References

1. Zuo E, Cai YJ, Li K, Wei Y, Wang BA, Sun Y, et al. One-step generation of complete gene knockout mice and monkeys by CRISPR/Cas9-mediated gene editing with multiple sgRNAs. Cell Res. 2017 Jul;27(7):933–45.

2. Appearances can be deceiving: phenotypes of knockout mice | Briefings in Functional Genomics | Oxford Academic [Internet]. [cited 2024 Dec 5]. Available from: https://academic.oup.com/bfg/article/6/2/91/240212?login=false

3. Uhl EW, Warner NJ. Mouse Models as Predictors of Human Responses: Evolutionary Medicine. Curr Pathobiol Rep. 2015;3(3):219–23.

4. Ha D, Kim D, Kim I, Oh Y, Kong J, Han SK, et al. Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. Nucleic Acids Research. 2022 Feb 28;50(4):1849–63.

5. El-Brolosy MA, Stainier DYR. Genetic compensation: A phenomenon in search of mechanisms. PLoS Genet. 2017 Jul 13;13(7):e1006780.

6. Hall B, Limaye A, Kulkarni AB. Overview: Generation of Gene Knockout Mice. Curr Protoc Cell Biol. 2009 Sep;CHAPTER:Unit-19.1217.

7. Bell A, Severi E, Owen CD, Latousakis D, Juge N. Biochemical and structural basis of sialic acid utilization by gut microbes. J Biol Chem. 2023 Feb 8;299(3):102989.

8. Nishiyama K, Mukai T. Sialic acid impact on the gut microbiome and function. Glycoforum. 2022 Apr 1;25:A3.

Chapter IV.

4. Conclusion, Perspective and Future Work

4.1. Recap of previous work

In Chapter 1 of this thesis, I introduced the complex interplay between inflammatory bowel disease (IBD) and genetic factors, emphasizing the significance of genome-wide association studies (GWAS) in identifying risk and protective variants. This chapter highlighted the challenges in characterizing genetic variations and their functional consequences, particularly in the context of IBD-associated genes. The chapter set the foundation for understanding how these genetic insights, coupled with modern approaches like in silico modeling, in vitro experimentation, and in vivo validation, pave the way for deorphanizing enzymes implicated in disease pathogenesis.

In Chapter 2, I presented an in-depth investigation into the ulcerative colitis-associated gene NXPE1, which encodes a previously uncharacterized enzyme. This chapter demonstrated that NXPE1 is a novel O-acetyltransferase that modifies sialic acid glycans in colonic mucins. Through biochemical assays, organoid-based models, and structural analyses, I established that NXPE1-mediated acetylation is crucial for maintaining the mucosal barrier and modulating host-microbe interactions. The findings elucidate a mechanistic basis for the UC-protective effects of the NXPE1 G353R variant and emphasize the role of sialic acid modifications in gut health.

In Chapter 3, I explored the characterization of NXPE4, focusing on its spatial expression and functional roles in mouse models. This chapter began with an investigation into the tissue-specific expression patterns of NXPE4, revealing its localized presence in

colonic epithelial cells. Building on this, I established a connection between NXPE4 expression and the gut microbiota, demonstrating that microbial colonization modulates NXPE4 levels and activity. Through knockout models and microbiota-altering interventions, I uncovered a potential feedback mechanism where NXPE4 influences mucosal glycan modifications, which in turn shapes the microbial composition. These findings underscore the interplay between host genetic factors and microbiota in maintaining intestinal homeostasis.

In Chapter 4, I summarized the overarching findings of this thesis and provided a perspective on the future directions for research. This chapter emphasized the importance of deorphanizing enzymes such as NXPE1 in uncovering novel molecular pathways and highlighted the potential to expand this work to other sialic acid-modifying enzymes. Future efforts will aim to further characterize the NXPE family and explore the therapeutic implications of modulating sialic acid acetylation in IBD and other mucosal disorders.

4.2 Perspective on Therapeutic Targeting of the Mucosal Barrier and Its Implications for Ulcerative Colitis

In this thesis, I describe our efforts to uncover the enzymatic and physiological role of NXPE1, a previously uncharacterized gene implicated in ulcerative colitis (UC). While the central question was how NXPE1 contributes to mucosal barrier integrity and immune regulation, this required a systematic approach to first deorphanize the enzyme and identify its substrate. By integrating biochemical assays, structural modeling, and organoid-based

functional studies, we demonstrated that NXPE1 is an O-acetyltransferase responsible for modifying colonic mucin sialic acids. Our work also revealed how genetic variants in NXPE1, particularly the UC-protective G353R variant, impair enzymatic activity and alter the mucosal glycome. These findings not only provide mechanistic insights into UC pathogenesis but also highlight NXPE1 as a potential therapeutic target for modulating gut barrier function in inflammatory bowel diseases.

The current therapeutic landscape for UC is limited, relying heavily on broad immunosuppressive drugs such as corticosteroids, biologics, or JAK inhibitors. While effective for some, these treatments often come with significant systemic side effects and fail to address the root cause of mucosal barrier dysfunction. The work presented here opens a novel therapeutic avenue by targeting the mucosal layer directly through the modulation of NXPE1 activity. By developing inhibitors or modulators of NXPE1, it may be possible to fine-tune the mucosal barrier in a tissue-specific manner, preserving its protective functions while minimizing systemic effects. This precision-based approach could offer a safer, more effective treatment option for UC patients who are unresponsive to current therapies.

Furthermore, this work has implications for preventative interventions in individuals at high risk for UC. Genetic screening for protective and risk variants of NXPE1 could identify susceptible populations who might benefit from early therapeutic modulation of the mucosal layer. By strengthening the intestinal barrier before the onset of inflammation, such interventions could serve as a prophylactic strategy, reducing disease incidence and severity. This tissue-specific, mechanism-based approach represents a shift

from systemic immunosuppression to targeted barrier modulation, offering a new paradigm in the prevention and treatment of UC.

The therapeutic potential of this work is promising, and has the potential to positively affect millions of patients. Identifying small molecules or biologics capable of modulating NXPE1 activity and understanding how they interact with the broader gut environment, are the logical next steps. Ultimately, the precision and tissue specificity of this approach provide a compelling framework for improving outcomes in UC and other diseases involving mucosal barrier dysfunction.

4.3 Future Directions: Screening Lundbeck's Serine Hydrolase Inhibitor Library for NXPE1 and NXPE4 Modulation

Future research should focus on leveraging Lundbeck's comprehensive serine hydrolase-directed inhibitor library to identify modulators of NXPE1 and potentially NXPE4. While NXPE1 has been shown to play a critical role in modifying colonic mucins through sialic acid acetylation, preliminary findings suggest that NXPE4, a paralog of NXPE1, might have overlapping or complementary functions. It remains unclear whether inhibiting NXPE1 alone is sufficient to achieve a physiological effect on the mucosal barrier or if dual inhibition of NXPE1 and NXPE4 is necessary. Addressing this question is pivotal for the design of effective therapeutic strategies targeting mucosal barrier dysfunction in ulcerative colitis (UC).

To identify specific inhibitors, a gel-based activity-based protein profiling (ABPP) assay using FP-TAMRA (fluorophosphonate-TAMRA) probes would be highly advantageous. Compared to FP-biotin probes, FP-TAMRA enables direct fluorescence detection, simplifying workflow and allowing for high-throughput screening. The first step would involve validating the FP-TAMRA probe with NXPE1 and NXPE4 to confirm its reactivity with the active site serine residues of these enzymes. This step is critical to ensure that the probe can reliably label both targets for competitive inhibition assays.

Once the assay is optimized, the Lundbeck library can be screened to identify inhibitors with specificity for NXPE1, NXPE4, or both. For compounds that demonstrate promising inhibition, dose-response experiments will be conducted to determine IC50 values, selectivity, and potential off-target effects. Structural modeling and mutagenesis studies could further elucidate the molecular basis of inhibitor binding, offering insights into the active site dynamics and potential for enhancing specificity.

Furthermore, parallel functional combinational studies will be essential to assess whether inhibition of NXPE1 in combination with NXPE4 result in significantly increased mucosal barrier integrity resulting in even greater measurable physiological effect. These studies could involve the use of organoid models and NXPE1/4 knockout systems to compare phenotypes under single and dual inhibition conditions. For example, experiments could evaluate changes in sialic acid acetylation patterns, mucosal glycome composition, and barrier properties under different inhibitor regimens. Additionally, mouse models with targeted deletions of NXPE1, NXPE4, or both could provide valuable insights into the in vivo relevance of these enzymes and their potential redundancy or synergistic roles.

Ultimately, these investigations will inform whether therapeutic strategies should focus on NXPE1 alone or include NXPE4 as a co-target. If dual inhibition proves even more efficacious, screening for inhibitors with dual specificity or developing combination therapies will become a priority. Conversely, if NXPE1 alone is sufficient to drive critical modifications, inhibitors targeting it could be fine-tuned for maximal efficacy and minimal off-target effects.

This integrated approach—combining advanced chemical biology tools, high-quality inhibitor libraries, and robust functional assays—has the potential to transform our understanding of NXPE1 and NXPE4's roles in mucosal biology. By addressing these questions, we can refine therapeutic strategies aimed at modulating the mucosal barrier in UC and related diseases, paving the way for more effective, tissue-specific interventions with reduced systemic side effects.