

Understanding Biological Hydrogel Function Through Design of Simplified Peptides and Polymers

By

Wesley George Chen

B.S. Biological Engineering
California Institute of Technology (Caltech) 2012

Submitted to the Department of Biological Engineering in partial fulfillment of the requirements for the Degree of Philosophy at the Massachusetts Institute of Technology

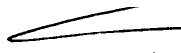
May 2017 [June 2017]

© Wesley George Chen. All Rights Reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.


Author

Signature redacted


Wesley George Chen
Department of Biological Engineering
May 12, 2017

Signature redacted

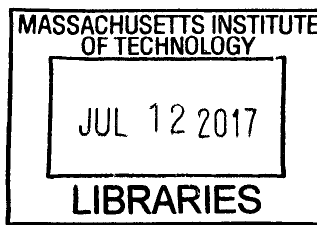
Certified By.


.....
Katharina Ribbeck
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted By.

Signature redacted

Mark Bathe
Graduate Academic Program Chair, Department of Biological Engineering



ARCHIVES

Understanding Biological Hydrogel Function Through Design of Simplified Peptides and Polymers

By
Wesley George Chen

Submitted to the Department of Biological Engineering on 12 May 2017, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Engineering

Abstract

Biological hydrogels exhibit complex properties that cannot be recapitulated by current synthetic materials. Examples include mucus, which acts as a barrier against toxins and pathogens while simultaneously hosting trillions of microbes within the gut; cartilage which resists repetitive compressive forces while maintaining highly lubricated layers for efficient movement; and nuclear pore matrices which act as selective barriers in the transport of proteins and nucleic acids. An underlying theme that gives biological hydrogels their unique mechanical and biological functions is the presence of long polymeric molecules. These polymers are typically comprised of repeating subunits that are essential for correct polymer function, such as the phenylalanine-glycine (FG) repeats in nucleoporin proteins of nuclear pore complexes (NPCs) and the proline-threonine-serine (PTS) domains in mucin polymers found in mucus. Although these polymeric subunits are well-identified, to date their structural complexity has limited our understanding of how they contribute to the overall hydrogel function.

In this thesis, we focus on two main biological hydrogels: the self-assembled matrix of the nuclear pore complex that controls the passage of molecules between the nucleus and the cytoplasm, and mucus, which protects against invading pathogens and toxins. As both hydrogels consist of functionally redundant polymers and associated factors, understanding the relationship between polymer sequence and hydrogel function is a significant technical challenge. To simplify the problem, we design structurally reduced peptides and polymers with targeted individual biological features such as amino acid identity, spatial localization of charge, and glycosylation identity. We then study the effect of one or a combination of these properties on the overall hydrogel function. Using this technique, we first demonstrate that peptide charge type and amino acid placement are important features for regulating selective transport through NPCs. For mucins, we identify single glycans that are sufficient to recapitulate the biofilm inhibition properties of mucin, and present novel evidence that mucins modulate horizontal gene transfer rates for opportunistic and commensal bacteria.

Thesis Supervisor: Katharina Ribbeck

Title: Associate Professor of Biological Engineering

Acknowledgements

First and foremost, I would like to thank my thesis mentor Katharina Ribbeck, who somehow managed to convince me in less than 20 minutes to study mucus and nuclear pore hydrogels for my Ph.D. She has always been wonderfully supportive of all my endeavors as a graduate student and provided the essential feedback to help me grow as a scientist.

I would also like to thank my committee members Alan Grodzinsky and Niels Holten-Andersen for their support these past five years. I consider myself very lucky to have found two committee members willing to have in depth discussions about my projects and provide the critical guidance from an external perspective.

I am immensely grateful to my parents Weizhen Gu and Sheng Chen and little sister Gina Chen for their continued support throughout my life. Without their dedication and love, I would not be where I am today.

I am also grateful to my undergraduate mentor Professor Grant Jensen at Caltech, who funded my research endeavors for four years. Without Professor Jensen's support, I would never have decided to apply for a Ph.D. program. I also owe everything to Dr. Morgan Beeby, who took the major risk in mentoring an extremely naïve freshman that knew nothing about biology. A big thank you is also needed for Dr. Elitza Tocheva, who spent two years training me in essential microbiology and biochemistry techniques, many of which I still use to this day.

To the rest of the Ribbeckers, thank you all for the wonderful time in lab. A special call out to Nicole Kavanaugh and Leon Li, my rotation mentors who convinced me to join the lab, and to Nicole Billings, Julia Co, Thomas Crouzier, and Erica Shapiro who supported me from day one. To the newer members Tahoura Samad, Brad Turner, Kate DuPont, Jacob Witten, Caroline Wagner, Emiko Zumbro, Caroline Werlang, Kelsey Wheeler, Ben Wang, Gerardo Oyarce-Carcamo, and Miri Krupkin, you all have been the best. It's thanks to you that made going into lab every day such an enjoyable experience.

I am also grateful to my climbing buddies Alex Turner, Ty Bubb, Justin Feng, Alix Chan, Nav Ranu, Mike Luong, Luis Gonzalez, and Lee Weinstein who keep me afloat both physically and mentally. They really should "take more" than their fair share of credit.

And lastly, to my fellow BE cohort. It's been an amazing time with all of you.

Table of Contents

| | |
|---|-----|
| Abstract..... | 2 |
| Acknowledgements..... | 3 |
| Table of Contents..... | 4 |
| Chapter 1 Introduction..... | 5 |
| Nuclear Pore Complex Self-assembly and Selective Transport | 6 |
| Mucus as a Microbial Phenotype Regulator..... | 8 |
| Chapter 2 Charge Influences Substrate Recognition and Self-assembly of Hydrophobic FG Sequences..... | 12 |
| Materials and Methods..... | 13 |
| Results | 18 |
| Discussion | 34 |
| Supplementary Figures..... | 37 |
| Chapter 3 Tryptophan Regulates Hydrophobic Selectivity of Nucleoporins | 43 |
| Materials and Methods..... | 43 |
| Results | 46 |
| Discussion | 56 |
| Chapter 4 Saliva and Salivary Mucin MUC5B Inhibit Transformation of Cavity-causing <i>Streptococcus mutans</i> | 58 |
| Materials and Methods..... | 59 |
| Results | 61 |
| Discussion | 69 |
| Supplementary Figures..... | 70 |
| Chapter 5 Gastric and Intestinal Mucins and Mucus Modulate Conjugation Rates for Commensal Bacteria..... | 72 |
| Materials and Methods..... | 74 |
| Results | 75 |
| Discussion | 80 |
| Chapter 6 Structure and function relationship of glycans on mucins | 82 |
| Chapter 7 Synthesis of Mucin Mimetics to Understand Function of Glycosylated Sugars | 88 |
| Materials and Methods..... | 89 |
| Results | 91 |
| Discussion | 96 |
| Chapter 8 Conclusion and Future Directions | 99 |
| References | 103 |

Chapter 1 Introduction

Biological hydrogels are ubiquitous in nature. At the human length scale, mucus hydrogels cover the entirety of our wet epithelia as a protective layer^{1,2}. At micron length scales, bacteria produce biofilms—polysaccharide hydrogels—that act as a protective coating against mechanical and environmental stresses³. At the nanometer lengths, the hydrogel within nuclear pore complexes (NPCs) act as selective regulators for transporting molecules between cytoplasmic and nuclear compartments in eukaryotic cells⁴⁻⁶. Depending on the context, biological hydrogels can serve as simply a physical barrier against offending stimuli, or act as more complex filters to selectively transport specific beneficial molecules. Mechanically, hydrogels such as cartilage can resist highly compressive forces in skeletal joints⁷⁻⁹ or provide a lubricating barrier to prevent mechanical stress¹⁰⁻¹⁴. Although these hydrogels exhibit dramatically different functions, a recurring theme in these gels is the presence of long polymeric molecules—either in the form of proteins, carbohydrates, or mixtures thereof—that give each hydrogel its unique properties. Further examination of the makeup of these polymers reveal that a significant fraction is comprised of repeating amino acid or carbohydrate subunits. For mucus, the major polymer is mucin, a megadalton glycoprotein that contains repeating regions of proline-threonine-series (PTS) domains, which are heavily glycosylated with O-linked sugars^{15,16}. For biofilms, a common polysaccharide is alginate^{17,18}, which is composed of repeating units of mannuronate and guluronate. For nuclear pores, the selective matrix is formed by proteins known as nucleoporins^{5,19,20}, which contain repeating FG-domains necessary for the unique selective properties of NPCs.

Although the essential polymers have been identified in many biological hydrogels, there is a lack of understanding as to how the biophysical and biochemical properties of the polymers give rise to the overall structure and function of the resultant self-assembled hydrogel. Here, in this thesis, we focus on two biological hydrogels: the self-assembled matrix found within NPCs and mucus. This thesis is separated into three major sections. Chapters 2 and 3 investigate the role of charged and hydrophobic amino acids in the molecular recognition of diffusing substrates that travel through NPCs. Chapters 4 and 5 examine how mucus in the oral cavity and gastrointestinal tract alter the rates of horizontal gene transfer between opportunistic and commensal bacteria. Chapters 6 and 7 seek to establish design principles as to how individual components of mucin glycoproteins contribute to the protective effects of mucosal hydrogels.

Nuclear Pore Complex Self-assembly and Selective Transport

The NPC is a major regulatory channel between the nucleus and cytoplasm within eukaryotic cells²¹⁻²³. The nucleoporins that create the self-assembled matrix within the NPC channel are intrinsically disordered proteins that contain stretches of hydrophobic phenylalanine-glycine (FG) domains^{5,21,24-29}. The hydrophobic FG domains self-assemble to form a saturated gel that excludes >99% of the proteins and nucleic acids synthesized by the cell. For efficient transport through the NPC, molecules require complexation with nuclear transport receptors (NTRs)^{23,27,30-35}. Efficient transport relies on transient hydrophobic interactions between hydrophobic patches on the NTR and the phenyl groups of FG domains.

From *in vivo* experiments, the minimal number of nucleoporins to maintain a selective NPC has been identified and several FG-domains have been discovered to be essential for NPC function³⁶.

However, despite in depth *in vivo* analyses, we still lack the understanding for establishing heuristic design principles that allow for development of synthetic gels that recapitulate the selective properties of the NPC. There are several reasons for the gap in knowledge. Primarily, the NPC contains many redundantly functional nucleoporins³⁶. Moreover, within each nucleoporin responsible for the selective function (termed FG nucleoporins), there are redundant FG domains that consist of multiple repeating sequences of FxFG, GLFG, or FG (general is FG sequences) interspersed by hydrophilic amino acids. The complexity of the NPCs and individual nucleoporin sequences has prevented systematic analysis of sequence to function associations. As a result, the focus within the nuclear pore field has primarily been on the easily identifiable phenylalanines and their role in maintaining a highly selective NPC. However, these FG sequences consist <10% of the >1000 amino acid long FG nucleoporins. A significant fraction of the FG nucleoporin sequence is disregarded and is speculated to only be essential for structural integrity of the NPC matrix and not selective function.

A significant challenge in dissecting the role of these highly polar and charged regions of the FG nucleoporins is the lack of conserved motifs. Unlike the conserved FG sequences, the regions interspersed between FG sequences are highly degenerate. Hence, systematic testing of how individual charged or polar amino acids contribute to FG domain function remains challenging *in vivo*. In Chapter 2, I describe a novel method using rationally designed short peptide sequences to circumvent this issue. With this technique, I determined that the charge type and identity of amino acids surrounding FG sequences impact the structure and selectivity of FG-based gels. Moreover, we showed that spatial localization of the charged amino acids with respect to the FG sequence determines the degree to which charge influences hydrophobic interactions. Taken

together, the data suggest that the amino sequence surrounding FG sequences represent a tunable mechanism to regulate the functions of FG domains to achieve a range of transport behaviors, from complete retention at the interface to free diffusion through the hydrogel dependent on the placement and charge type. In Chapter 3, I show that hydrophobic selectivity is not limited to phenylalanines in the FG domains. The presence of tryptophan in nucleoporin sequences may also act as a secondary regulatory of hydrophobic interactions and supports stringent selective transport in regions of the NPC devoid of phenylalanines.

Mucus as a Microbial Phenotype Regulator

Mucus is a highly hydrating hydrogel that coats the wet epithelia of the human body. It manifests in many different forms, from tears on the surface of eyes, saliva in the oral cavity, and the thick layer in the digestive tract. At these epithelial surfaces, mucus forms a protective layer that prevents microbial infections³⁷⁻⁴⁰. Mechanically, mucus hydrates the epithelia and lubricates surfaces to allow for processes such as blinking, talking, ingestion, and digestion. The polymer that gives mucus its mechanical properties is mucin^{15,41-44}, a long polymeric protein that is heavily glycosylated^{2,45}. Mucins come in two main forms: secreted and surface-attached. Here, we focus on the secreted mucins such as Muc5ac, Muc5b, and Muc2, which are able to self-assemble to form viscoelastic hydrogels. For a full review of the mucin glycoproteins and their structure, refer to Varki *et al*⁴⁶. Briefly, mucin glycoproteins are up to 2 MDa in molecular weight, with 50-80% of the molecular weight consisting of complex glycans. The glycans that protrude from the protein backbone are O-glycans linked to serines and threonines. The addition of complex carbohydrates is through stochastic processes governed by at least 20 different glycosyltransferases. The O-glycans are typically present in four core structures termed cores 1,

2, 3 and 4. The primary monomer attached to serines and threonines are N-acetyl galactosamines (GalNacs), which are then further modified with addition of N-acetyl glucosamine (GlcNac), galactose (Gal), sialic acid, or fucose (Fuc). Through these stochastic processes, over 100 various glycans have been identified on Muc2 alone⁴⁷, and are responsible for providing mucins their protective attributes.

Aberrant production and glycosylation of mucins are correlated with disease states such as cystic fibrosis^{48,49}, peptic ulcers⁵⁰⁻⁵², and other colorectal diseases⁵³. Inflammation and infections that arise from deregulated mucin production suggest that mucins play a particularly important role in regulating microbe function and physiology. In fact, mucins have a significant function in preventing surface attachment and biofilm formation of microbes^{54,55}. By adding mucins to a microbial culture, microbes such as *Candida albicans* and *Pseudomonas aeruginosa* downregulate various gene expression profiles related to virulence and toxicity⁵⁵. Studies examining how mucins affect multi-species dynamics also suggest that mucins may promote co-existence of antagonistic microbes. For instance, in the oral cavity, salivary mucins allow for prolonged co-existence of *Streptococcus mutans* and *Streptococcus sanguinis*⁵⁶ whereas gastric Muc5ac allows for co-cultures of *Staphylococcus aureus* and *P. aeruginosa* to thrive (unpublished results). These results provide insight as to why billions of microbes can live harmoniously on mucosal surfaces in the human body.

Although mucins have been identified as an essential glycoprotein in regulating specific virulence traits, gene expression studies now suggest that mucins may globally change the phenotype of microbes residing in the mucosal layer^{48,55}. Therefore, it is essential to begin

surveying on a broader scale what microbial functions mucins are able to influence. The particular virulence trait I will focus on in this thesis is the ability of microbes to acquire foreign DNA through a process termed horizontal gene transfer (HGT). Through HGT, microbes have been reported to obtain new traits such as antibiotic resistance⁵⁷⁻⁵⁹, toxin secretion⁶⁰, and novel metabolic pathways⁶¹. Of particular important from the global health perspective is the recent rise of antibiotic resistance in pathogens. In the US alone, over 2 million infections leading to 23,000 deaths annually are attributed to resistant bacteria. The annual economic impact is over \$20B in direct costs and \$35B in indirect costs stemming from 8 million additional days in hospitals. Globally up to 10 million deaths annually are attributed to antibiotic resistance, with costs expected to rise to \$100 trillion by 2050⁶².

The current paradigm suggests that HGT occurs readily in the human body such as in the oral cavity and gastrointestinal tract due to the shear density and high rate of interactions between microbes^{59,63-65}. The three main mechanisms of horizontal gene transfer are transformation, conjugation, and transduction⁶⁶. Briefly, transformation is the acquisition of exogenous DNA when the microbial cell enters a competent state. Once competent, microbes activate pathways that allow for internalization of DNA and incorporation into their chromosome and plasmids. Conjugation is the exchange of material between two bacteria through the formation of sex pili. Typically thought to occur at high frequencies in the gut, conjugation requires physical contact between donor and acceptor cells. Last, transduction is the spread of genetic material through bacteriophages, which infect their microbial hosts. In addition to antibiotic resistance genes, many virulence genes are recognized to spread through various mechanisms of HGT as well. Considering the high frequencies of HGT that can be achieved *in vitro* and the density of

microbes that inhabit our mucosal surfaces, microbes should be able to constantly exchange genetic material in the host environment. Despite that, antibiotic resistance and virulence genes are not present in all microbial genomes, suggesting that our body must have developed mechanisms to limit the spread of virulence and toxin-related genes as a protective measure. In this thesis, we explore how our body is able to modulate the exchange of genetic material for microbes that commonly inhabit our mucosal surfaces. Specifically, in Chapter 4, I examine how salivary mucin MUC5B and native saliva inhibit the natural transformation of the opportunistic pathogen *Streptococcus mutans*, which is known to cause dental caries. In Chapter 5, I show how gastrointestinal mucins and mucus inhibit conjugation between *E. coli* strains but upregulate conjugation events between *E. coli* and probiotics such as *Lactobacillus reuteri*. These results show that mucins and more generally, the mucosal environment, has a dramatic impact on the ability of microbes to undergo HGT and a deeper understanding is required as to what components of mucins regulate these virulence pathways. To investigate what components of mucins are able to regulate virulence traits of microbes, I take an engineering approach to identify essential glycans that may be of functional relevance and synthesize a small library of glycopolymers for *in vitro* assays. In Chapter 6, I describe a method in which to narrow down the potential glycans to test. In Chapter 7, I provide a synthesis protocol to show how to engineer simplified mucin-like glycopolymers. Here, I discovered that complex glycans are not necessary to inhibit virulence traits such as biofilm formation and gene transfer in *S. mutans*. Instead, simple monosaccharides grafted onto inert polymers are sufficient for strong inhibitory responses.

Chapter 2 Charge Influences Substrate Recognition and Self-assembly of Hydrophobic FG Sequences

The nuclear pore complex (NPC) is a megadalton structure that controls the exchange of material between the nucleus and cytoplasm²¹⁻²³ through a combination of passive and facilitated diffusion. Above a ~30 kDa cutoff, proteins require complexation with nuclear transport receptors (NTRs) to efficiently translocate at rates of nearly one thousand molecules per second⁶⁷⁻⁷⁰. This fast translocation rate relies on transient interactions between hydrophobic phenylalanine-glycine (FG) domains on intrinsically disordered FG-containing nucleoporins (FG nucleoporins) and hydrophobic patches on NTRs^{5,6,19,21,24-27,30,31,36,68,71-73}. Without this assistance from NTRs, proteins remain excluded from the NPC. Despite the necessity of FG domains (which contain repeating units of FG sequences such as FxFG or GLFG) for facilitated diffusion and self-assembly of the selective matrix, how hydrophobic FG domains exhibit such selectivity for specific hydrophobic domains on NTRs and what parameters tune the molecular recognition necessary for facilitated transport remain open questions. Sequence analysis and molecular dynamic simulations predict that the biochemistry and charge surrounding individual FG sequences should play an essential role in FG-mediated molecular recognition and therefore determine the organization and selectivity of the NPC^{67,69,74-77}. However, due to the complexity and redundancy of FG nucleoporins within NPCs, systematic biochemical dissection of the amino acid space surrounding individual FG sequences has remained an experimental challenge.

Here, we investigated whether and how electrostatic interactions surrounding FG sequences tailor both the self-assembly of FG sequences and the selective recognition of hydrophobic substrates. For systematic dissection of how charge type and localization influence FG-mediated

selectivity, full FG domains are too degenerate and complex to provide insight into how single amino acids contribute to selectivity. In peptide research, rationally designed peptides and polypeptides have previously been helpful to identify individual amino acids or short peptide sequences important to the function of much larger polymer-based biomaterials such as extracellular matrices⁷⁸, silk proteins^{79,80}, and elastin-like polypeptides⁸¹⁻⁸³. Here, we applied this approach and created systematically varied peptide sequences to identify the amino acids surrounding FG domains in native nucleoporins that may contribute to selectivity. While peptides may not recapitulate all properties of the original protein, they are a suitable model system for this study because they yield insight into the contributions of individual amino acids and their positioning to overall protein function, a task not feasible with intact proteins. Our data demonstrate that the identity and charge of amino acids surrounding FG sequences can impact the structure and stiffness of FG-based gels. Moreover, we determined that charged amino acids enable FG-containing peptide gels to discriminate substrates of varying charge and hydrophobicity, positioning us to create tunable filters that select substrates across a broad spectrum of complex biochemical properties. Last, we found that the distance at which a charged amino acid is localized with respect to the FG sequence determines the degree to which charge influences hydrophobic interactions. Together, our data suggest that FG domain function may be determined by the placement and type of amino acids surrounding FG sequences.

Materials and Methods

Sequence Analysis:

The logo of the consensus sequence was generated using Berkeley's WebLogo software⁸⁴ by aligning 15 repeats according to FSFG as reference. Conserved fraction was calculated for

positive charges (# cationic residues K or R)/(15 repeats) and similarly for negative charges (# anionic residue E or D)/(15 repeats) at each position indicated.

NTF2 Expression and Labeling:

pQE30-NTF2-6xHis was a gift from the Gorlich lab and transformed into DH5 α cells for cloning. A cysteine (C) was inserted for fluorescent labeling after the 6x His-tag using standard site-directed mutagenesis with the two primers (5' – ctcagctaattaagcttagcagtgatggtgatggtgatgagatctg - 3' and 5' - cagatctcatcaccatcaccatcactgctaagcttaattagctgag – 3') to form pQE30-NTF2-6xHis-C. The W7A-C mutant containing the terminal cysteine was created by using pQE30-NTF2-6xHis-C and applying standard site-directed mutagenesis with primers 5' – tgaggagccaatttgttccgcgatcggtttatcacccatg – 3' and 5' – catgggtgataaacgatcgcggaacaaattggctcctca – 3'. Expression of NTF2-C and W7A was completed in OverExpress C41(DE3) cells (Lucigen) and purified using standard nickel column purification and ion-exchange columns. Labeling of NTF2-C and W7A-C was completed using Fluorescein-5-maleimide (ThermoFischer Scientific, Catalog #F150) in accordance with the manufacturer's protocol; labeled product was gel purified. Labeling efficiency was approximately 50% (data not shown). The final concentrations of NTF2 and W7A for transport were adjusted to 10 μ M with 10% of the population labeled.

Peptide and Gel Preparation:

Unless specified otherwise, all chemicals were obtained from Sigma Aldrich (St. Louis, MO, USA). Peptides were prepared by MIT's Koch Institute Biopolymers and Proteomics Facility (Cambridge, MA, USA) and Boston Open Labs LLC (Cambridge, MA, USA). All peptides were

HPLC purified unless specified otherwise, desalted using reverse phase HPLC with 0.05% trifluoroacetic acid (TFA), and lyophilized after synthesis with >95% purity. For fluorescently labeled peptides, a 5-carboxyfluorescein (5-FAM, Anaspec; Fremont, CA, USA) fluorophore was added to the N-terminus, and the C-terminus was modified to be an amide. The dye labeling protocol was as follows: peptide-resin that contained an N-terminal free amine but was otherwise fully protected was washed six times with dimethylformamide (DMF) followed by six washes with dichloromethane (DCM) and dried. The material was subsequently reconstituted with dry DMF to re-swell the resin. The dye-labeling cocktail consisted of a 4X molar excess over peptide-resin of 5-FAM, N,N'-Diisopropylcarbodiimide (DIC), and hydroxybenzotriazole (HOBt) reconstituted in a minimal volume of dry DMF and stirred overnight at room temperature in the dark. The cocktail was allowed to pre-activate for 30 min before being added to the peptide-resin. Dye-labeled peptide resin was washed six times with DMF, then six times with DCM, and dried in preparation for standard fluorenylmethyloxycarbonyl chloride (Fmoc) cleavage and de-protection. All quality-control analyses for purity were provided by MIT's Proteomics facility or by Boston Open Labs (Supplementary Data S1). Neutral Hydrophilic (n) and Hydrophobic (n) fluorescent reporters could not be HPLC purified due to aggregation and were used as crude samples.

Fluorescent peptides were diluted into 200 mM NaCl with 20 mM HEPES [pH 7] at 10 μ M final concentration for diffusion experiments. Gel peptides were all dissolved in 20 mM NaCl, 20 mM HEPES [pH 7] at 2% (w/v). To facilitate solubilization and gel formation, peptides were vortexed for 30 s and briefly sonicated in a bath sonicator (Branson 2510) to reduce aggregation.

Capillary Diffusion Assay and Analysis:

Borosilicate square capillaries (1.5 inch) with 9 mm cross sectional width (Catalog # 8290, Vitrocom; Mountain Lakes, NJ, USA) were loaded by piercing pre-made hydrogels. Ten-micromolar solutions (200 mM NaCl, 20 mM HEPES [pH 7]) of fluorescent peptides were injected into the capillary and sealed with a 1:1:1 (by weight) mixture of Vaseline, lanolin, and paraffin. Time lapses of peptide diffusion were taken at 1-min intervals for up to 5 h on a Nikon Ti Eclipse inverted microscope using a Nikon CFI Plan UW 2X or on an AxioObserver D.1 with a EC Plan-Neofluar 1.25x/0.03 WD=3.9 and Hamamatsu C11440-22CU camera. All fluorescence profiles were obtained by averaging the fluorescence intensities across the width of the capillary in MatLab (MathWorks; Natick, MA, USA). Normalized concentration profiles were obtained by normalizing fluorescence intensities to the bath concentration of the capillary at the initial time point. The fluorescence signal was linear up to 50 μM (Supplementary Figure 2-1). To plot concentration profiles, the signal was not adjusted past the saturation point; data therefore represent the lower bound of the actual concentration of reporters accumulating in the gel. All data represent at least three independent replicates. Student's t test was applied to determine p-values between experimental conditions.

Effective diffusion rates were fit by minimizing the squared error of a simulated concentration timecourse in a region of the capillary on the gel side of the interface over a 100-min window. To achieve this fit, we numerically solved the diffusion equation for the concentration of probe c :

$$\frac{\partial c(x, t)}{\partial t} = D \frac{\partial^2 c(x, t)}{\partial x^2}$$

using MatLab's *pdepe* function (MathWorks; Natick, MA). The initial condition $c(x, 0)$ was set by the concentration profile at the first timepoint, and the boundary conditions $c(0, t)$ and $c(L, t)$

(for a fit over length L) were similarly determined by the concentration profiles at the edges of the region of interest. The D that minimized the squared difference between the simulated and actual concentration profiles was the value reported; minimization of error took place iteratively using a modified gradient descent algorithm. The window from 150-250 min was generally used for fitting, but for particularly fast-diffusing probes ($D > 1000 \mu\text{m}^2/\text{min}$), windows of 50-150 min or 30-130 min were used. The earlier time period allowed for the fitting to take place before the steady state or pseudo-steady state was reached, which was crucial for precise fitting. Examples of the diffusion-coefficient analysis across the time series as well as the general fit appear in Supplementary Figure 2-2.

Nile Red Fluorescence Assay:

Using the diffusion methods mentioned above, Nile Red dye (N1142, Thermo Fisher Scientific; Waltham, MA, USA) was loaded into capillaries at $10 \mu\text{g}/\text{mL}$ in 200 mM NaCl , 20 mM HEPES [pH 7].

Rheological Testing:

Rheological tests were performed on an Anton Paar MCR 302 Rheometer in a cone-plate geometry with a 25-mm diameter, 1° cone angle, and $51\text{-}\mu\text{m}$ truncation. The temperature was maintained at 25°C and evaporation was controlled with an H_2O -filled solvent trap. To identify the linear regime, amplitude sweeps were conducted at $\omega = 10 \text{ rad/s}$ from $\gamma_0 = 0.01\%$ to 100% strain. In the linear regime, frequency sweeps were conducted using the previously determined strain amplitude from $\omega = 100 \text{ rad/s}$ to 0.1 rad/s .

Transmission Electron Microscopy:

Images were taken with a JEOL-1200 transmission electron microscope. Gels were formed as described above and spotted onto glow-discharged carbon-coated copper Formvar grids (Ted Pella). Excess liquid and gel were removed with Whatman paper or parafilm. Grids were submerged in 1% uranyl acetate, blotted, and air dried for 15 min prior to imaging. Images of gels are representative of at least four images of each gel type.

Phenyl-sepharose chromatography:

Fluorescent reporters were dissolved in 200 mM NaCl, 20 mM HEPES [pH 7] and loaded into high-prep henyl FF 1 mL-capacity phenyl-sepharose columns (GE Healthcare Lifesciences; Pittsburgh, PA, USA) equilibrated with three volumes of 200 mM NaCl, 20 mM HEPES [pH 7]. For elution, flow rates were set to 1 mL/min with 0.5 mL fraction volumes. Fraction concentrations are representative traces of elution profiles.

Results

To determine the relevance of charged amino acids in FG domains, we chose the yeast nucleoporin Nsp1 as a model system because it is essential in *Saccharomyces cerevisiae*⁸⁵ and contains a repeating subsequence (284-553) with a high density of charge (Figure 2-1A). Sequence analysis of 15 repeats of FG domains in Nsp1²⁸⁴⁻⁵⁵³ revealed a non-uniform distribution of charge in the sequence space separating FG sequences (Figure 2-1B). Cationic residues appear near the center and edges of the repeat, whereas anionic amino acids reside at least three positions away from FSFG sequences. Moreover, several highly conserved lysine (K) residues are situated 2-3 amino acids away from the FG sequence, at positions 6 and 17, and a

conserved glutamic acid (E) appears at position 9 (Figure 2-1A). The non-uniformity in conserved charge distributions and amino-acid identity suggests that the charge and biochemical properties of amino-acid sidechains may play a role in governing how neighboring FG sequences respond to environmental substrates; that function may be encoded in the non self-assembling domains of FG nucleoporins.

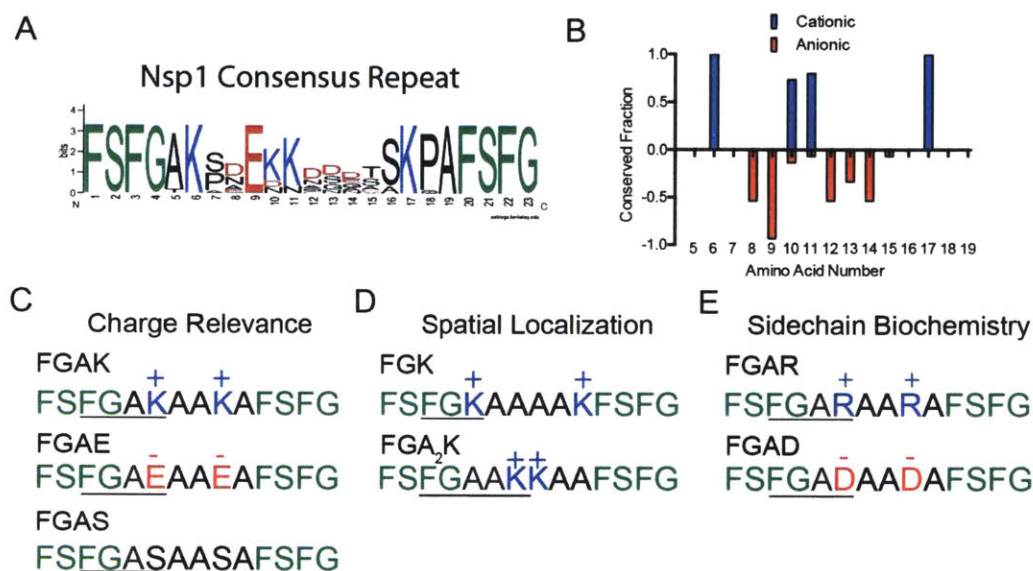


Figure 2-1: Identification of conserved repeat sequences and design of simplified FG self-assembling peptides

A) Conserved sequence identification of the C-terminal end of the essential yeast nucleoporin Nsp1. The letters in the top position indicate the most conserved residue at each location and represent the consensus sequence.

B) Conservation of charge between FG sequences in Nsp1 repeats. Positive or negative values refer to the presence of cationic or anionic residues, respectively. Amino acid numbers correspond to position in the consensus repeats, as in Figure 2-1A.

C) Simplified peptides consisting of 14 amino acids were designed with the sequence structure FSFGAXAAXAFSFG, where X represents the substituted amino acids K, E, or S. These peptides were used to determine how the presence of charge affects FG-mediated selectivity.

D) Designed peptides in which K is moved immediately adjacent to (FGK) or placed three amino acids away (FGA2K) from FSFG domains to test how the spatial localization of charge influences FG selectivity.

E) Class of peptides in which K is substituted with R or D to determine how biochemistry affects FG-mediated selectivity and self-assembly.

To test how the type and placement of charge, as well as sidechain chemistry, affect FG function, we synthesized the consensus Nsp1 peptide sequence and rationally designed 14-amino acid variations of it. All engineered peptides consisted of two terminal FSFG sequences with

neighboring charged or neutral amino acids. To establish how the presence of charge affects FG-mediated selectivity, we designed the peptide sequence FSFGAXAAXAFSFG, where X is K, E, or a neutral serine (S) (Figure 2-1C). Since K is the most conserved residue in this wildtype Nsp1 subsequence, we use the K-containing peptide (the FGAK peptide) as the reference for all other experimental comparisons. To determine whether the substrate-binding properties of FG domains depend upon the type of neighboring charge, we synthesized the anionic peptide FGAE, and as a control the neutral FGAS peptide in which each K was converted to S. To test how the positioning of K relative to the FG sequence affects FG function, we designed two variants of FGAK in which the K was placed directly adjacent to the FG domain or separated by two alanine (A) residues (peptides FGK and FGA₂K, respectively; Figure 2-1D). Last, since only certain charged amino acids such as K and E exhibit 100% conservation at positions 6, 9, and 17 (Figure 2-1A), we hypothesized that amino-acid biochemistries may also regulate FG-based molecular recognition. To test how the chemical structures of amino-acid sidechains affect the structure and function of FG domains, we designed a third class of peptides in which K was replaced with cationic arginine (R; FGAR) or aspartic acid (D; FGAD) (Figure 2-1E). With these 14-amino acid nucleoporin-based peptides, we evaluated how single amino-acid substitutions alter the selective binding and self-assembling capabilities of individual FG domains.

Characterization of the effect of charge on FG-mediated self-assembly

Before studying selective recognition by FG-based peptides, we first tested whether they form gels. We used a concentration of 2% (w/v) for each of the peptides, which corresponds to 28 mM FG sequence, a value that is well within the range estimated for densely packed NPCs^{20,36,86}. To

quantify gelation, we measured the stiffness of the resulting material using small amplitude oscillatory frequency sweeps. We report the storage (G') and loss (G'') moduli of the peptide solution (Figure 2-2). A gel forms when $G' > G''$, which indicates successful self-assembly of a stable network of peptides. The consensus peptide sequence of Nsp1 (Figure 2-1A) was unable to form a gel and flowed when inverted (data not shown), so it was not suitable for further analysis.

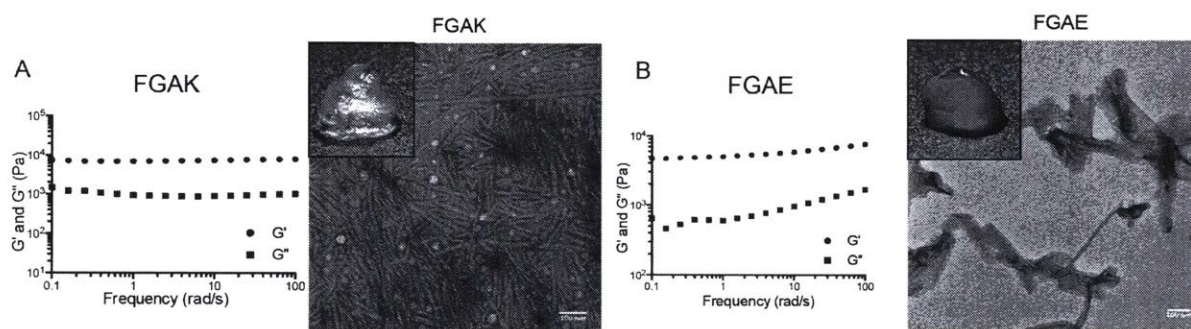


Figure 2-2: FGAK and FGAE self-assembly

Frequency sweeps of (A) FGAK gel and (B) FGAE gel, with G' (storage) and G'' (loss) moduli reported at 2% (w/v). Insets, transmission electron microscopy of self-assembled peptides and macroscopic gel. For both FGAK and FGAE peptides, $G' > G''$, indicating gel formation. The replacement of K with E alters the self-assembly properties of FG sequences to form different structures and gels with orders of magnitude varying stiffness.

For the engineered FG sequences, we established that the reference peptide FGAK forms a hydrogel with a stiffness of 10^4 Pa across the frequencies tested (Figure 2-2A). To ensure that the FG sequences were responsible for the self-assembly process, rather than the high density of A residues, which can promote stable self-assembly⁸⁷, we converted F to S (SGAK). With this substitution, the peptide remained in the aqueous phase and no longer exhibited a dominant storage modulus (Supplementary Figure 2-3A), suggesting that the F within the designed peptides provides the necessary hydrophobic interactions for gelation, as for intact FG nucleoporins^{5,25}. Reversing the charge via an E (FGAE peptides) revealed that the gel-forming properties are maintained with stiffness 5-9 KPa at the frequencies tested (Figure 2-2B). To

determine whether the charge is responsible for maintaining the hydration of the gel, we compared the material to the solution of neutral FGAS peptides. Without the charge, FGAS precipitated out of solution (Supplementary Figure 2-3B), indicating that the presence of charge is essential in maintaining a hydrated network of FG sequences. However, it appeared that too much charge could prevent gelation, presumably by increasing the solubility, as with the consensus sequence peptide. Without electrostatic repulsion in the gels, the network collapses and forms a precipitate. Taken together, our results suggest that the sequence adjacent to the FG sequence encodes information that impacts self-assembly and substrate binding by FG domains.

Binding of NTF2 is asymmetric in cationic and anionic FG-based gels

To determine whether FG-based peptide gels reconstitute the selective binding properties of native FG nucleoporins, we selected NTF2 as a model receptor because it contains an essential tryptophan (W7) that is required for binding FG domains. We replaced W with an A (NTF2^{W7A}) to ablate FG-binding capabilities²⁷, enabling us to test whether the FG domains are available in the gel for NTR binding. As cationic native and engineered nucleoporin gels are predicted to bind and facilitate the selective transport of native NTRs such as NTF2^{74,75}, we first determined whether the positively charged FGAK gel preferentially selected for NTF2 compared to the NTF2^{W7A} mutant. To prepare gels for selective transport assays, we dissolved FGAK peptides at 2% (w/v) as before and loaded the material into capillaries. Fluorescently labeled NTF2 or NTF2^{W7A} was then injected into the capillary and sealed to create a 1D diffusion chamber; the diffusion profile was monitored for up to 5 h at 1-min intervals (Figure 2-3A).

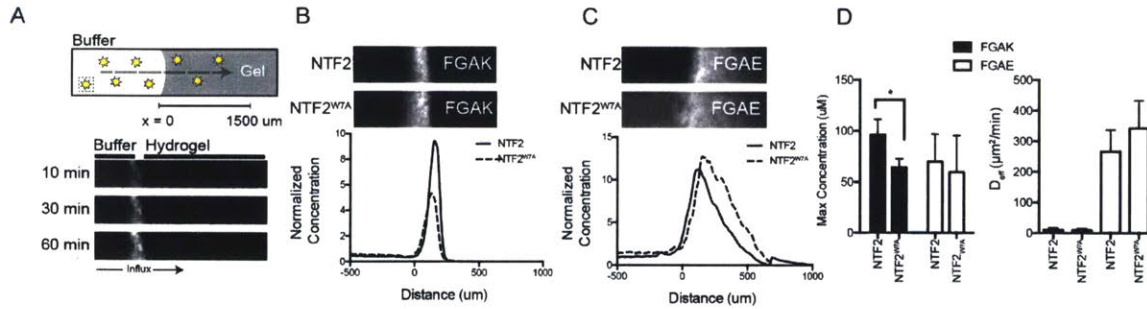


Figure 2-3: Molecular selectivity of the native transport receptor NTF2 in FGAK and FGAE gels

A) Top, schematic of slab 1D transport system within a sealed capillary, which provides no flux boundary conditions during the diffusion of fluorescent molecules (yellow stars). Bottom, the interface of the gel is reflected by the increase in fluorescence signal between the buffer and hydrogel.

B-C) Transport of representative fluorescently labeled NTF2 and NTF2W7A proteins into FGAK and FGAE gels at 5 h. FGAK is able to preferentially recognize NTF2 over NTF2W7A mutant whereas FGAE gels are unable to.

D) Left, maximum concentration of labeled proteins at the interface. Right, effective diffusion coefficients of labeled proteins. Error bars are standard deviations of at least three independent replicates. * $p < 0.05$, unpaired Student's t test.

To quantify the extent to which the gels differentiated between NTF2 and NTF2^{W7A}, we determined the maximum accumulation of the transport receptors at the gel interface and their effective diffusion coefficients inside the gel using time-evolving fluorescence profiles. Lower effective diffusivities and higher accumulation indicate stronger interactions with the gel. For a detailed description of diffusion coefficient analysis for these time series, refer to the Materials and Methods and Supplementary Figure 2-1. NTF2 accumulated at higher concentrations (mean \pm standard deviation: $96.2 \pm 15.2 \mu\text{M}$) at the gel interface than did the NTF2^{W7A} mutant ($63.9 \pm 8.8 \mu\text{M}$) (Figure 2-3B and D). Thus, the FGAK peptide has a binding preference for NTF2 over the NTF2^{W7A} mutant, and appears capable of recognizing a single hydrophobic amino-acid difference. While NTF2 displayed increased enrichment at the FGAK gel interface (Figure 2-3D), its effective diffusivity did not significantly differ from that of the NTF2^{W7A} mutant (NTF2, $10.6 \pm 6.5 \mu\text{m}^2/\text{min}$; NTF2^{W7A} $9.9 \pm 4.9 \mu\text{m}^2/\text{min}$). This experiment illustrates that the minimal *in vitro* system does not fully reconstitute the selective transport observed in the NPC *in vivo*, but it enables the characterization of the first step of selective transport in native NPCs,

namely, the requirements for the initial selective binding to FG-based gels, which is the focus of this work.

To test whether NTF2 binding is sensitive to charged residues neighboring the FG domain, we determined whether NTF2's interaction with FGAE recapitulated the selective properties of FGAK, despite the reversal of charge. Figure 2-3C and D show that FGAE selects for NTF2 ($69.9 \pm 27.1 \mu\text{M}$) and NTF2^{W7A} ($59.5 \pm 36.0 \mu\text{M}$) equally at the interface, with no significant difference in diffusion coefficients (NTF2 $266.5 \pm 71.1 \mu\text{M}$ and NTF2^{W7A} $342.7 \pm 91.0 \mu\text{M}$). These results indicate that the FGAE gel is less effective than the FGAK gel at differentiating between the native and mutant forms of NTF2. The FGAK and FGAE gels exhibit structural differences at the microscopic level (sheets vs. fibers; Figure 2-2A and B) and the macroscopic level (insets to Figure 2-2A and B). Therefore, it is possible that due to structural differences, the FG sequences are not exposed and are unavailable for binding by NTF2 in the FGAE gel. To test this hypothesis, we determined that Nile Red dye, which fluoresces within hydrophobic environments, was detected in both FGAK and FGAE gels (Supplementary Figure 2-4): the uncharged hydrophobic dye interacted with the FG domains within the peptide-based gels despite the structural variation. These results suggest that K and E may regulate the selective properties of the FG sequences from a charge interaction perspective, as opposed to simply altering the microstructure of the gel.

Presence of charge regulates selective recognition by FG domains

Although NTRs such as importin β (Imp β) and NTF2 have well-characterized specificity for particular FG-nup sequences, intact receptors contain multiple binding pockets with varying

affinities and charge distribution, or require dimerization for function. These factors complicate the systematic analysis of how charge contributes to FG function. Hence, as a replacement for complex NTRs in our minimal *in vitro* model, we designed fluorescent peptide reporters with defined spatial arrangements of charged and non-polar amino acids to systematically test the contribution of charge in hydrophobic selectivity (Figure 2-4A). The first two reporter peptides harbored three F residues, creating a hydrophobic tail for binding to FG sequences, but one contained an adjacent anionic sequence composed of E residues (termed Hydrophobic (-)), while the other contained cationic K residues (Hydrophobic (+)). As controls, we engineered two more reporters in which F residues were converted to hydrophilic asparagine (N) residues, termed Hydrophilic (- or +); these control peptides should not interact with FG domains. To confirm that the synthetic hydrophobic reporters interacted with the aromatic phenyl group in FG domains in the gels independently of their charged domains, we used phenyl-sepharose columns to assay hydrophobic interactions. This assay was previously employed to isolate intact NTRs from cell lysates⁶. Both hydrophobic reporters displayed longer retention times than their hydrophilic counterparts (Supplementary Figure 2-5), indicating that the hydrophobic reporters interact with the phenyl groups on the sepharose independent of the displayed charge.

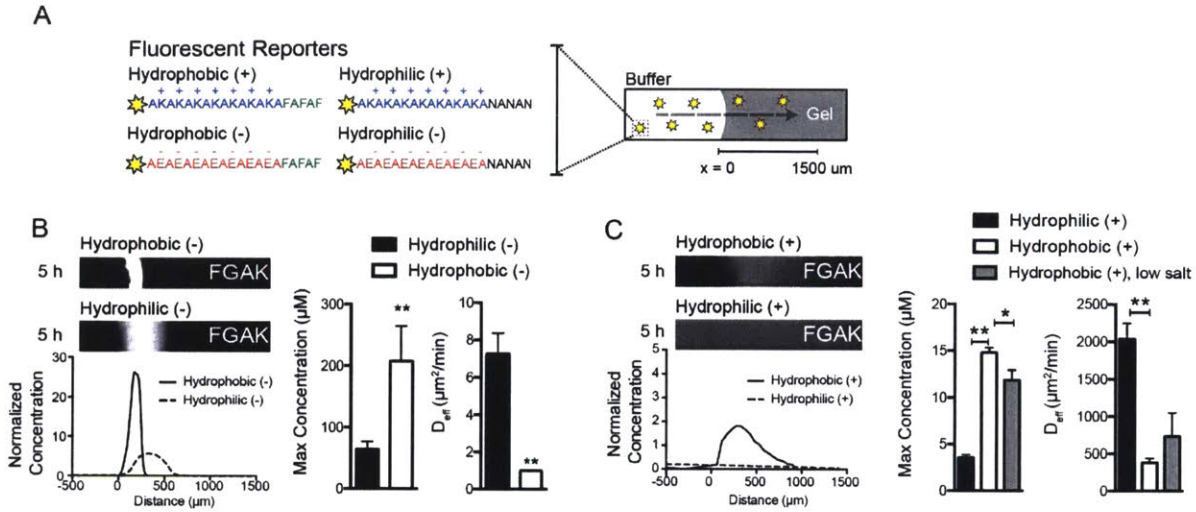


Figure 2-4: Charge affects hydrophobic molecular recognition in cationic FGAK gels

A) Schematic of the four fluorescent reporters with a charged domain and a hydrophilic (N) or hydrophobic (F) tail.

B) Representative fluorescence images of Hydrophobic (-) and Hydrophilic (-) reporters in the FGAK gel showing the increased selectivity for Hydrophobic (-) reporters over Hydrophilic (-) reporters

C) Representative fluorescence images of Hydrophobic (+) and Hydrophilic (+) reporters, which contain a cationic tail, into FGAK gel showing the increased selectivity for Hydrophobic (+) reporters over Hydrophilic (-) reporters. A low-salt condition (20 mM NaCl) was added for the Hydrophobic (+) reporters to show how electrostatic screening modulates hydrophobic interactions for Hydrophobic (+) reporters. For panels B and C, error bars are standard deviations of at least three independent replicates. * $p < 0.05$ and ** $p < 0.01$, unpaired Student's *t* test.

We used the FGAK gels to investigate whether the two hydrophobic fluorescent reporters underwent differential uptake into the gel. The cationic FGAK gel interacted with the Hydrophobic (-) reporters, and accumulated inside the gel. The maximum concentration of peptide at the interface was ~3-fold higher for the Hydrophobic (-) reporter ($64.6 \pm 12.1 \mu\text{M}$) than the Hydrophilic (-) reporter ($207.6 \pm 57.2 \mu\text{M}$) (Figure 2-4B), which does not contain a hydrophobic tail, suggesting that FG sequences in FGAK recognize reporter F residues when the reporter is negatively charged. The contribution of hydrophobic interactions is further corroborated by the strong reduction in the effective diffusion coefficient from $7 \mu\text{m}^2/\text{min}$ for Hydrophilic (-) to $<1 \mu\text{m}^2/\text{min}$ for Hydrophobic (-) (Figure 2-4B), indicating that the hydrophobic interactions induce tighter binding in the context of electrostatic attraction. We note that the interactions are so strong that the Hydrophobic (-) reporters essentially do not diffuse

over the 5-h period analyzed, suggesting that the binding is irreversible. In contrast, when the reporters contained a cationic tail (Figure 2-4C), the Hydrophobic (+) and Hydrophilic (+) reporters diffused into the gel with coefficients of $381.0 \pm 57 \text{ m}^2/\text{min}$ and $2038.0 \pm 211.8 \text{ }\mu\text{m}^2/\text{min}$, respectively. The Hydrophobic (+) reporters bound 1.5 times above the original bath concentration at the interface ($14.81 \pm 0.5 \text{ }\mu\text{M}$), whereas the Hydrophilic (+) reporters equilibrated ($3.6 \pm 0.3 \text{ }\mu\text{M}$) with no discernible partitioning (Figure 2-4C). Thus, hydrophobic interactions can occur in electrostatically repelling environments, but the interactions are much weaker than those detected with the Hydrophobic (-) reporters. Moreover, lowering the salt concentrations from 200 mM NaCl to 20 mM (to decrease electrostatic screening and to increase electrostatic repulsion) decreased the accumulation of Hydrophilic (+) reporter at the interface ($11.9 \pm 1.1 \text{ }\mu\text{M}$), while the diffusion coefficient ($732 \pm 317.7 \text{ }\mu\text{m}^2/\text{min}$) trended upward (Figure 2-4C), indicating that repulsion weakens the overall strength of the hydrophobic interactions with the FG domains. These results show that the K residues in FGAK gels distinguish between two substrates that contain the same hydrophobic domain but with different surrounding charge types. In particular, K allows for increased binding to hydrophobic domains with neighboring anionic residues.

Spatial localization of lysine affects FG-mediated self-assembly and binding selectivity

Our analysis of the Nsp1 repeat consensus sequence (Figure 2-1A and B) suggests that the conserved location of K within 2-3 amino acids of the FG sequence may be relevant for FG-mediated molecular recognition. To characterize this relationship, we tested whether placing a K immediately adjacent to the FG sequence (FGK) or moving it 3 amino acids away (FGA₂K) affected the sequence's selectivity compared to the original arrangement in the FGAK peptide.

While FGA₂K peptides readily formed a stiff material and could be tested for selective uptake, the FGK peptide gel was two orders of magnitude less stiff (Figure 2-5A and B), flowed when inverted, and dispersed when fluorescent reporters were loaded (data not shown). We therefore focused our comparison on diffusion and accumulation in FGAK and FGA₂K gels. The FGAK gel selectively enriched for the Hydrophobic (-) reporter but failed to uptake the Hydrophobic (+) reporter to the same degree (Figure 2-4B and C). The effective diffusivity of the Hydrophobic (-) ($0.5 \pm 0.14 \mu\text{m}^2/\text{min}$) reporter in the FGA₂K gel was an order of magnitude lower than that of the Hydrophilic (-) reporter ($7.9 \pm 1.7 \mu\text{m}^2/\text{min}$) (Figure 2-5C), showing that the combination of hydrophobic and electrostatic attraction synergizes for strong binding. However, the maximum accumulation at the interface of the two anionic reporters was similar ($153.6 \pm 29.6 \mu\text{M}$ for Hydrophobic (-) vs. $205.7 \pm 12.2 \mu\text{M}$ for Hydrophilic (-)) (Figure 2-5C), indicating that the FGA₂K gel is unable to differentiate between the two reporters to the same degree as the original FGAK gel. For cationic reporters, the Hydrophobic (+) peptide accumulated within the gel ($25.8 \pm 0.8 \mu\text{M}$) and displayed an order of magnitude lower effective diffusivity ($444.3 \pm 267.7 \mu\text{m}^2/\text{min}$) than the Hydrophilic (+) reporter ($6.5 \pm 1.3 \mu\text{M}$ and $2902 \pm 719.8 \mu\text{m}^2/\text{min}$, respectively) (Figure 2-5D). In addition, more Hydrophobic (+) reporter accumulates in FGA₂K gels than in FGAK gels (Figure 2-5D). These data show that placing K within 2 amino acids of an FG domain, approximately corresponding to the 1 nm Debye length at physiologically relevant salt concentrations, enables hydrophobic selection of substrates with opposite charge, while placing K three amino acids away reduces the contribution of electrostatic interactions to hydrophobic FG-mediated selectivity. As a result, by increasing the distance between charged residues and FG sequences, the FG sequences become less dependent on the electrostatic profile surrounding the hydrophobic substrate during selective transport.

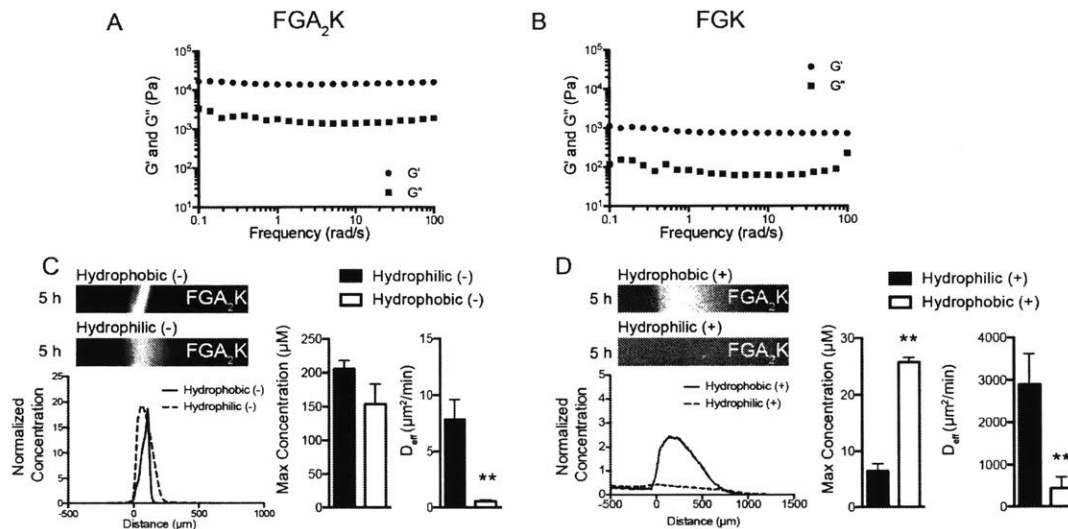


Figure 2-5: Effects of spatial localization of charge on FG-mediated self-assembly and selectivity

Frequency sweeps of (A) FGA2K and (B) FGK gels with G' (storage) and G'' (loss) moduli reported at 2% (w/v).

C) Representative fluorescence images of Hydrophobic (-) and Hydrophilic (-) reporters in a FGA2K gel showing the similarities between the selectivity of the two reporters.

D) Representative fluorescence images of Hydrophobic (+) and Hydrophilic (+) reporters, which contain a cationic, in a FGA₂K gel showing the ability of FGA₂K gels to select for Hydrophobic (+) reporters independent of electrostatic repulsion. Error bars are standard deviations of at least three independent replicates. * $p < 0.05$ and ** $p < 0.01$, unpaired Student's t test.

Glutamic acid reverses the selectivity of FG domains

Since the K residues in the FGAK gel help FG sequences differentiate between reporters containing anionic or cationic hydrophobic domains, we next asked whether FG-mediated recognition could be reversed by including anionic E residues. The selectivity of FGAE gels for the same class of fluorescent reporters was reversed from that of FGAK gels (Figure 2-6 A-D). Selectivity was a function of both hydrophobic and electrostatic interactions, as the Hydrophilic (+) reporters accumulated 3-fold less at the interface ($41.8 \pm 7.1 \mu\text{M}$) than did the Hydrophobic (+) reporters ($123.2 \pm 29.32 \mu\text{M}$) (Figure 2-6D). Moreover, the diffusion coefficient of the Hydrophobic (+) reporter ($6.6 \pm 2.6 \mu\text{m}^2/\text{min}$) was an order of magnitude lower than that of the Hydrophilic (+) reporter ($160.8 \pm 32.7 \mu\text{m}^2/\text{min}$) (Figure 2-6D). Conversely, neither the

Hydrophobic (-) nor the Hydrophilic (-) reporter interacted significantly with the gel; their similar effective diffusion coefficients ($1381.0 \pm 282.0 \mu\text{m}^2/\text{min}$ and $1451.0 \pm 223.0 \mu\text{m}^2/\text{min}$, respectively) (Figure 2-6C) show that electrostatic repulsion minimizes hydrophobic interactions with FG sequences. One general concern in synthesizing anionic peptides is residual TFA protonating anionic groups such as E. To ensure that the reporters were not neutralized in our buffering conditions, we synthesized neutral reporter peptides Hydrophilic (n) and Hydrophobic (n) (Supplementary Figure 2-6). In diffusion experiments with FGAK and FGAE gels, the neutrally charged reporters aggregated in solution and interacted minimally with the gels (Supplementary Figure 2-6), indicating that the anionic peptides are not neutralized by the residual TFA and indeed carry a net negative charge. Taken together, the data in Figure 2-6 and Supplementary Figure 2-6 suggest that charge proximal to FG sequences can tune hydrophobic selectivity, and that charge is essential in determining the hydrophobic moieties recognized by FG sequences.

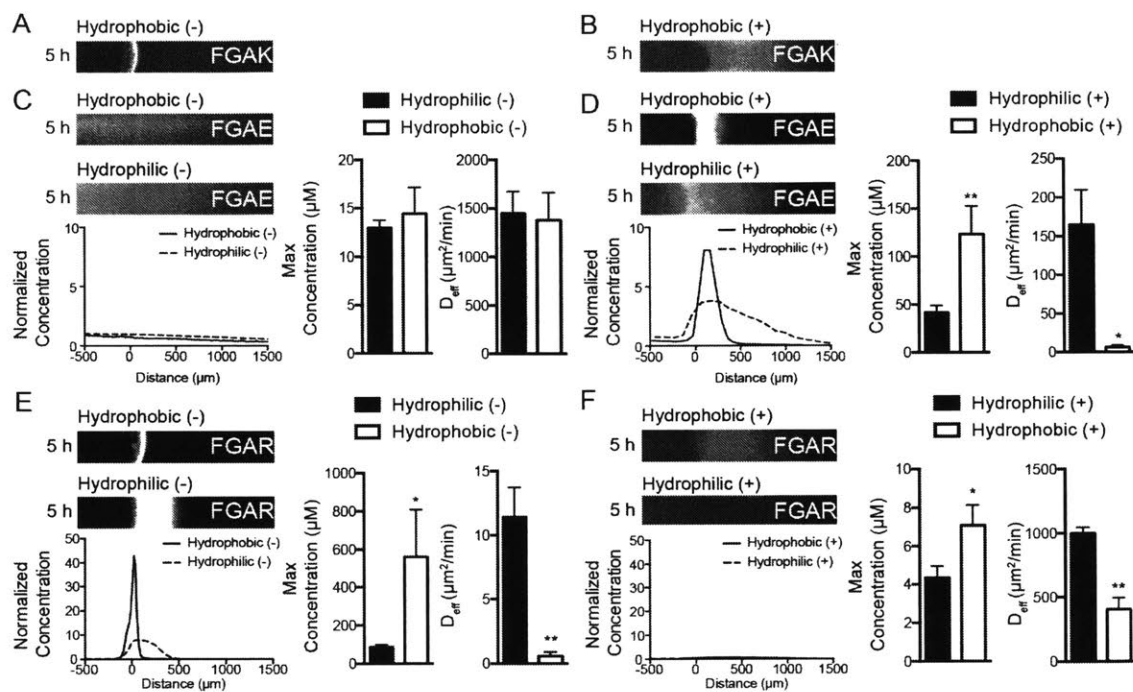


Figure 2-6: E residues reverse the selectivity of FG-based gels, while R-containing gels displays selectivity similar to that of K-containing gels

Representative (A) Hydrophobic (-) and (B) Hydrophobic (+) fluorescence images in FGAK gels.

C) Representative fluorescence images of Hydrophobic (-) and Hydrophilic (-) reporters in an FGAE gel with the corresponding concentration profiles showing the reversal of selective recognition when compared to FGAK gels with the same reporters. In FGAE gels, anionic reporters do not interact at the gel interface.

D) Representative fluorescence images of Hydrophobic (+) and Hydrophilic (+) reporters in an FGAE gel showing the reversal of selective recognition when compared to FGAK gels. In FGAE gels, the Hydrophobic (+) reporter binds more significantly at the interface compared to the Hydrophilic (+) reporter.

E) Representative fluorescence images of Hydrophobic (-) and Hydrophilic (-) reporters in an FGAR gel showing the similarity of selective recognition to that of FGAK gels.

F) Representative fluorescence images of Hydrophobic (+) and Hydrophilic (+) reporters in an FGAR gel showing the similarity of selective recognition to that of FGAK gels. Error bars are standard deviations of at least three independent replicates. * $p < 0.05$ and ** $p < 0.01$, unpaired Student's t test.

Amino acid sidechain chemistry is an additional regulator of FG function

Last, we tested how the sidechain chemistry of charged amino acids affects the self-assembly and molecular recognition of FG sequences. In Nsp1, the predominant cation is K rather than R (Figure 2-1A), suggesting that the two cations may not function equally. Similarly, E is highly conserved at position 9, whereas D is not as well conserved throughout the repeats (Figure

2-1A). To test the importance of sidechain chemistry, we determined that FGAR gels have uptake properties on the diffusion assay that are similar to those of FGAK gels (Figure 2-6A and 6E). The Hydrophobic (-) reporter accumulated ($561.1 \pm 247.7 \mu\text{M}$) and interacted with the interface of the FGAR gels ($0.6 \pm 0.3 \mu\text{m}^2/\text{min}$), while the Hydrophobic (+) reporter did not ($7.1 \pm 1.1 \mu\text{M}$ and $403.7 \pm 90.2 \mu\text{m}^2/\text{min}$, respectively) (Figure 2-6E and F). These data suggest that from an electrostatic standpoint, R is just as capable as K in helping an FG sequence differentiate between hydrophobic substrates. However, mechanically, FGAR forms gels with an approximate stiffness of 2000 Pa throughout the frequency sweep (Supplementary Figure 2-7A), 4-fold more compliant than FGAK gels (Figure 2-2A). The differences in the mechanical properties of the FGAK and FGAR gels suggest that the K and R residues may predominantly affect the structural self-assembly of FG domains.

Transmission electron microscopy revealed that FGAR peptides (Supplementary Figure 2-7B) form different structures than FGAK peptides (Figure 2-2). FGAD did not form a gel at the standard 2% (w/v), formed no repeating structures (Supplementary Figure 2-7C and S8D), and inverted when flowed (data not shown), suggesting that the small chemical differences between E and D transition the material from a selective gel to a viscous solution. Taken together, these data show that in addition to net charge, amino acid sidechain properties such as their inherent hydrophobicity, sidechain sterics, and charge distribution all may be important parameters that determine hydrophobic FG-mediated interactions.

Tryptophan (W) interactions are similarly modulated by electrostatics

An essential question in the selectivity of FG domains is how generalizable the tunability of hydrophobic selectivity is to natural aromatic amino acids other than F. In native NTRs such as NTF2 and other transport receptors, W is commonly used for the transient hydrophobic interactions required for transport. To test whether an electrostatic dependence persists when F is converted to W, we constructed hydrophobic reporters Hydrophobic (+W) and Hydrophobic (-W) from the original hydrophobic F reporters (termed +F and -F, respectively, for the remainder of the manuscript). The W-containing reporters were modulated by electrostatic interactions in a similar manner to F-containing reporters. The maximum accumulated fluorescence at the interface and diffusion did not significantly differ between the Hydrophobic -F and -W reporters (Figure 2-7A). The cationic reporters exhibited similar maximum interface properties, but the +W reporter showed a slight and non-significant trend toward slower diffusion than the +F reporter (+F: $479.9 \pm 217.2 \mu\text{m}^2/\text{min}$; +W: $210.5 \pm 24.6 \mu\text{m}^2/\text{min}$) (Figure 2-7B), indicating that although both aromatic residues can be modulated by electrostatic interactions, the identity of the hydrophobic residue has a small effect on transport in the gel. These results reveal that in the NPC, the phenomenon of electrostatic interactions regulating hydrophobic interactions may be generalizable to aromatic residues other than F; moreover, selective transport may also be tuned by the identity of the non-polar amino acid.

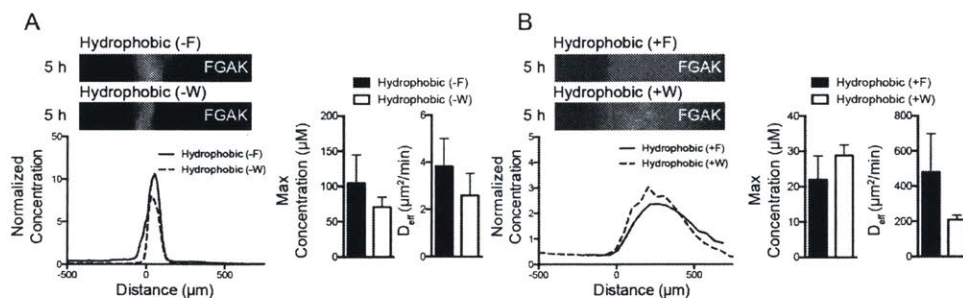


Figure 2-7: Electrostatic interactions modify hydrophobic recognition of W-containing reporters in FGAK gels

Comparison of (A) Hydrophobic (-F), (B) Hydrophobic (-W), (C) Hydrophobic (+F), and (D) Hydrophobic (+W) reporters in FGAK gels with representative fluorescence images and corresponding concentration profiles show that replacing F with W does

not alter hydrophobic molecular recognition for FGAK gels. Error bars are standard deviations of at least three independent replicates. No significant differences were detected via Student's t test.

Discussion

The current investigation demonstrates how the environment surrounding FG sequences tunes the hydrophobic interactions necessary for correct molecular recognition of hydrophobic substrates. Here, we have explicitly shown how the presence, placement, and type of charged amino acid are all essential parameters for FG function in gels. Translating concepts that are well established in the peptide field enabled us to establish how even simplified variants of the Nsp1 consensus sequence can encode complex molecular recognition. The data reported here support previous theoretical predictions that electrostatic interactions may be an important regulator of NPC selectivity⁷⁵ as well as sequence observations that the bias in the net charge of NTRs may be important in transport through the NPC⁷⁴. Moreover, recent reports on how charge influences interactions with hydrophobic surfaces^{71,88} emphasize that the interplay between electrostatics and hydrophobicity may be fundamental for various biological functions such as molecular recognition and DNA packing^{89,90}.

For the nuclear pore field, the current investigation shifts the spotlight from how hydrophobic effects determine NPC selectivity to how electrostatics contribute to this selectivity^{5,6,24,91}. The results described here provide insight into how individual hydrophobic FG sequences may distinguish a particular subset of hydrophobic substrates using a combination of electrostatic and hydrophobic domains that are in close proximity to each other (Figure 2-8). The closer the charged residues are to FG sequences, the greater the influence of charge on FG self-assembly (Figure 2-8A) and recognition of diffusing substrates (Figure 2-8B). As the charged residue is

situated greater than the theoretical Debye length at physiological conditions, the hydrophobic and electrostatic motifs are able to act independently for molecular recognition.

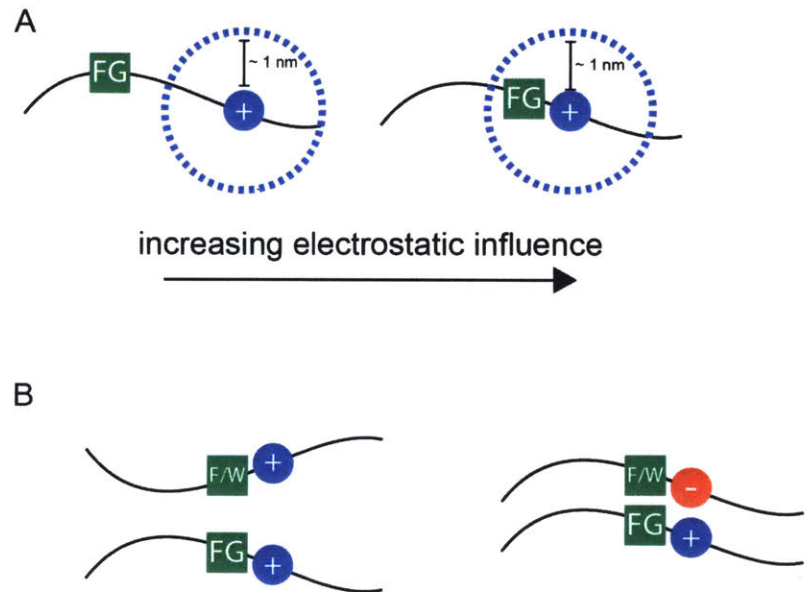


Figure 2-8: Schematic of the influence of electrostatics on hydrophobic interactions mediated by FG domains

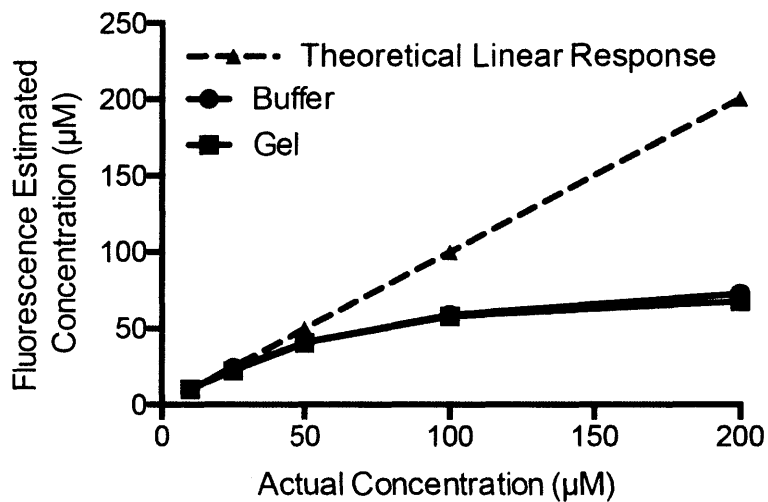
A) Complementary charge and hydrophobic interactions enable the strongest interactions (dotted lines) between a substrate and FG domains. Modulating charge types and the presence of hydrophobic domains can tune interactions from strong binding to free diffusion within the gel and in *in vivo* systems. B) At physiologically relevant salt concentrations, electrostatic interactions typically have a range of ~1 nm (Debye length), which determines how much a charged residue influences the FG sequence's recognition of and binding to hydrophobic substrates. When charged residues are moved further away from the FG sequence, electrostatics less significantly impact FG-mediated self-assembly and selective transport.

We recognize that although the study of engineered peptides can answer questions on how hydrophobic interactions are affected by the precise placement of charged amino acids, short-peptide gels do not recapitulate all of the unique facilitated diffusion properties of intact FG nucleoporins. This disconnect likely arises from structural differences, as short peptides form fibers and rods (Figure 2-2 and Supplementary Figure 2-7) that are not the dominant structures observed or predicted in the native nuclear pore^{70,92}. Moreover, the interactions between diffusing reporter peptides and the self-assembled gel may cause structural changes in the gel and also alter the diffusivity of the reporter, which has also been reported for Imp β in native

NPCs⁹³. Nevertheless, the current investigation of peptides provides strong evidence (Figure 2-4 through Figure 2-6) that charge is an essential regulator of transport through the NPC by tuning the essential first step of recognizing specific hydrophobic substrates. Moreover, engineered FG-containing peptides do capture the interplay between hydrophobic and electrostatic interactions at single amino-acid resolution (Figure 2-5), which is not possible with currently available *in vivo* techniques.

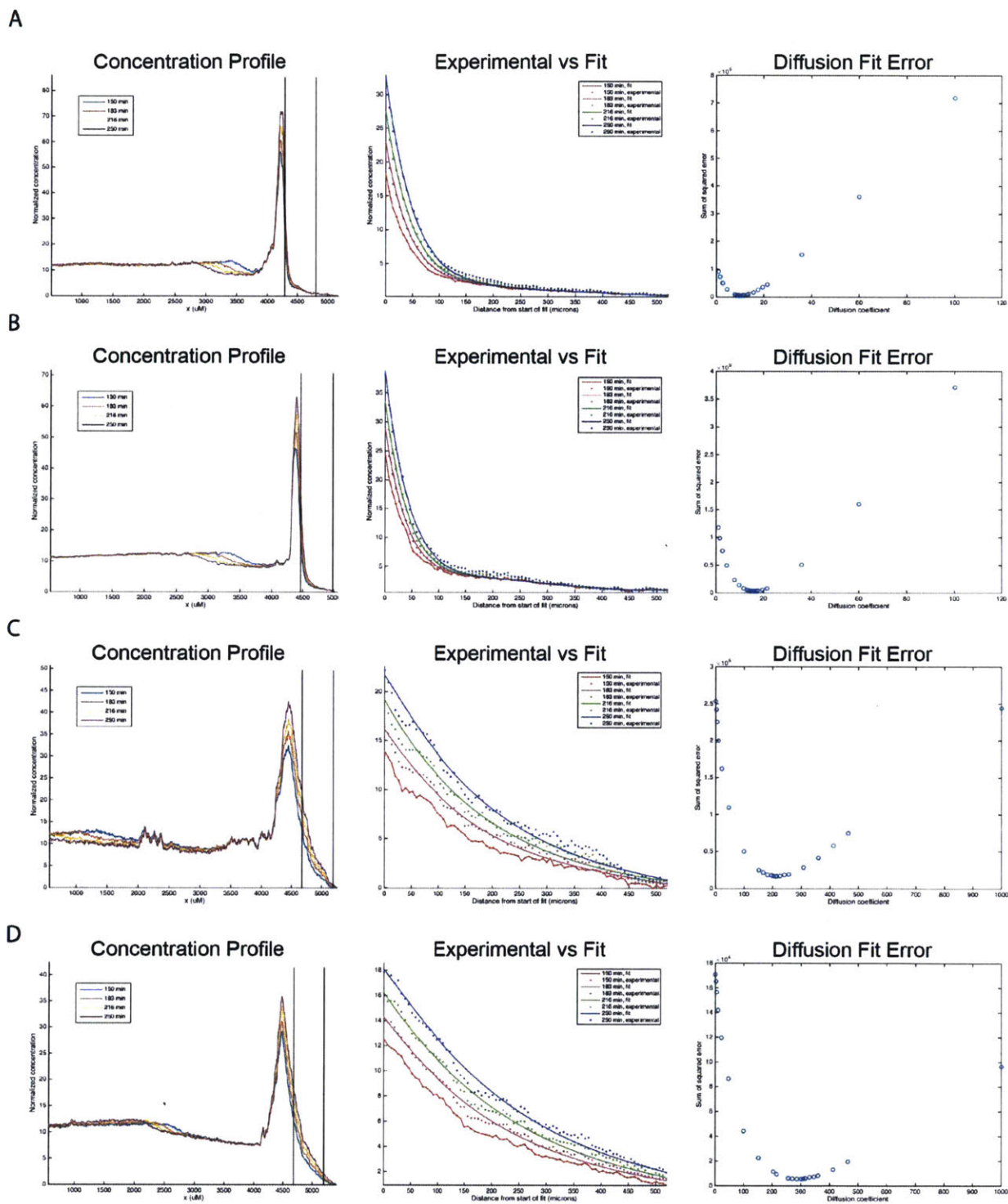
Our results build on previous studies of intact nucleoporins containing multiple repeats of FG domains where charge primarily plays a structural role in the cohesion of the self-assembled matrix⁹². Our results also support recent work in which Imp β was recognized to have varying binding capacities with synthetic FG-containing polyacrylamide gels depending on the charged state of the material⁹⁴. We believe that our strategy, which was inspired and informed by peptide engineering, is a useful approach that complements other well-established methods used to understand the selective transport mechanism of NPCs, such as the *in vivo* minimal NPC^{19,36}, gel and selectivity analysis of individual nucleoporins^{5,24,25,95}, and binding interactions with surface-grafted nucleoporin films^{92,96-98}. We expect that this rational approach from the peptide field, which focuses on conserved repeating sequences, can be extended to other biological gel systems such as mucus, byssal threads, and cartilage, where complex disordered proteins based on repeat units and reversible crosslinking constitute a significant proportion of the material and dictate its function^{2,11,99}.

Supplementary Figures



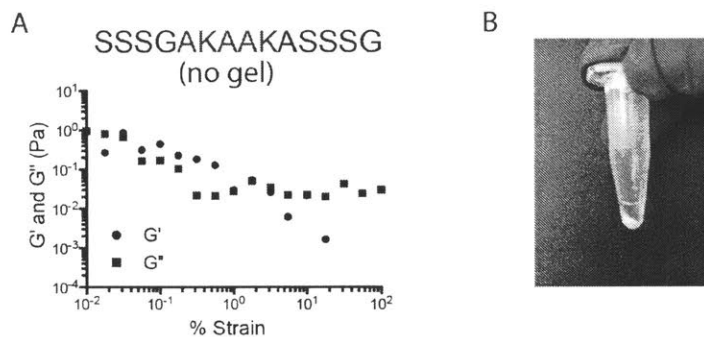
Supplementary Figure 2-1: Quantification of fluorescence signal as a function of fluorophore concentration

Fluorescence signal is approximately linear up to 50 µM and saturates by 100 µM in buffer and gel conditions. Dashed lines represent the theoretical linear response of fluorescence as a function of concentration. The gels and buffer calibration curves overlap in their associations. All concentrations are reported according to the experimental curve developed and represent lower values of the actual concentrations for values >100 µM.



Supplementary Figure 2-2: Analytical process for calculating effective diffusion coefficients

Examples are given for A) NTF2 diffusion into FGAK, B) W7A diffusion into FGAK, C) NTF2 diffusion into FGAE, and D) W7A diffusion into FGAE to show the reliability of the analytical process across multiple gels. The first column represents the region of the concentration profile where the fitting is implemented. The second column contains the actual data (circles) vs. fit (solid line) at four evenly spaced time points. The third column contains the error of the fit as a function of iterated effective diffusion coefficients.



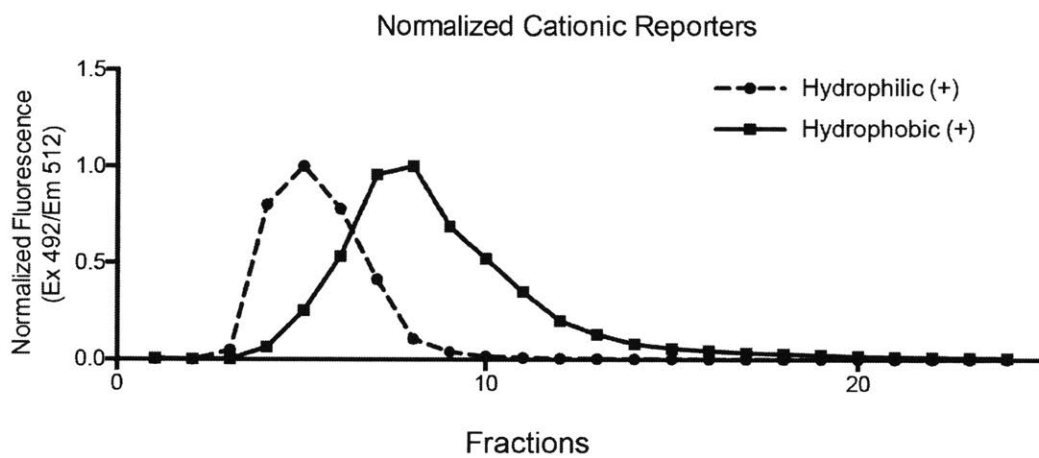
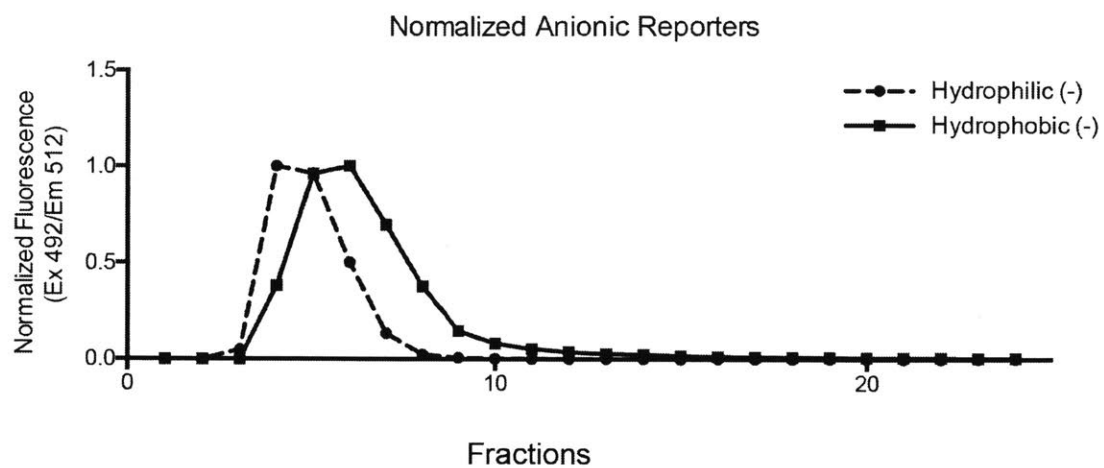
Supplementary Figure 2-3: Verification of F as essential amino acid for self-assembly in FGAK peptides

A) Frequency sweep of the F→S substitution (FGAK → SGAK) to determine the effect of F on the self-assembly of peptides. The elastic modulus (G') and loss modulus (G'') are reported. Note that the measured values are below the sensitivity of the rheometer using the specified cone-plate geometry due to the viscous nature of SGAK peptide solutions. B) Precipitated FGAS peptides in 20 mM NaCl, 20 mM HEPES [pH 7] after gentle centrifugation.



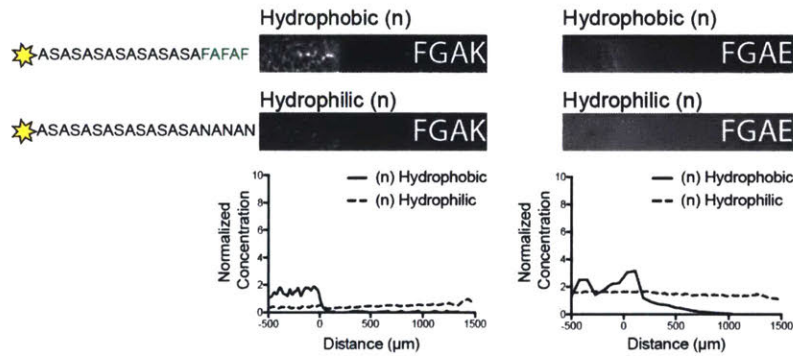
Supplementary Figure 2-4: Verification of hydrophobic domain availability in FG peptide gels

A) Transport of Nile Red into FGAK and FGAE gels after 0 h and 3 h of incubation. Fluorescence indicates that the dye is able to access hydrophobic environments created by FG domains within the gels. Images are of representative gels from three independent replicates.

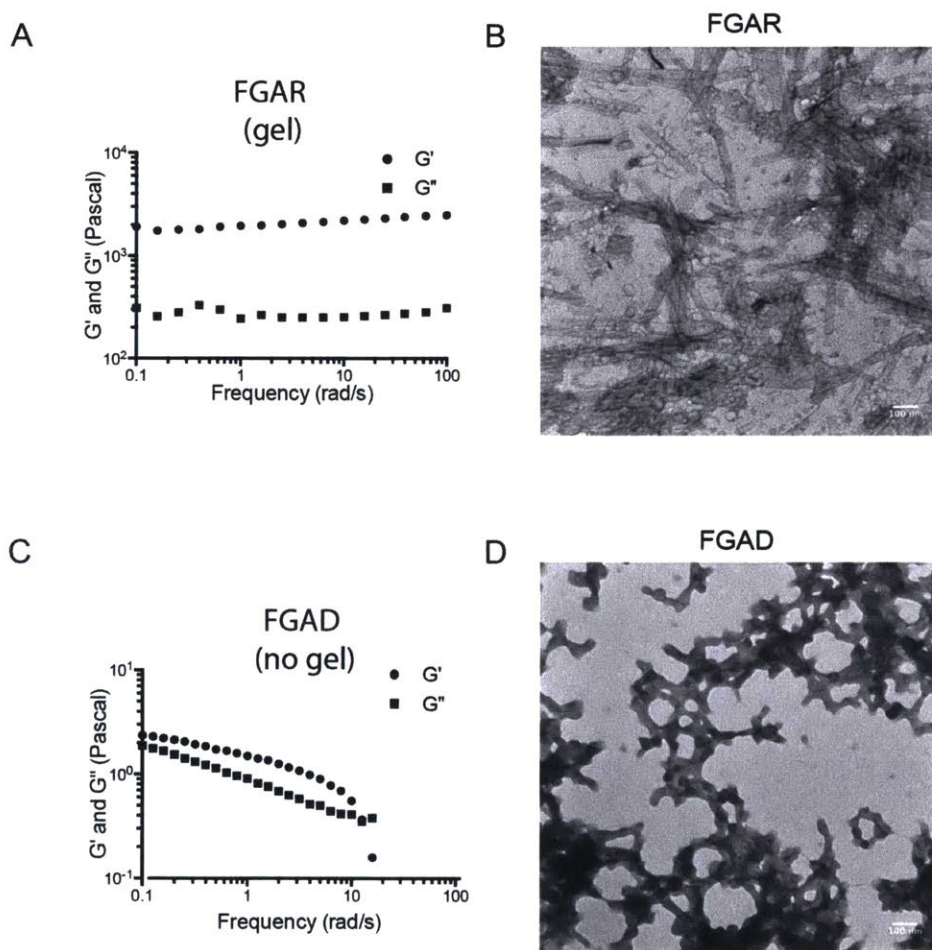


Supplementary Figure 2-5: Fractionation of hydrophilic reporters and their hydrophobic counterparts in phenyl-sepharose columns

Fluorescence signals from each fraction were collected and normalized to the signal with the highest intensity of emission. For both cationic and anionic reporters, the hydrophobic reporters eluted later. This increased retention time reflects stronger binding to phenyl-sepharose beads.



Supplementary Figure 2-6: Diffusion of neutrally charged Hydrophilic (n) and Hydrophobic (n) reporters into cationic FGAK and anionic FGAE gels. Purely neutral reporters interact minimally with the FGAK and FGAE gels regardless of overall hydrophobicity.



Supplementary Figure 2-7: Affect of amino acid sidechain chemistry on self-assembly and mechanical properties of FG-containing peptides.

A) Frequency sweep of FGAR gel with G' (storage) and G'' (loss) moduli reported at 2% (w/v) showing the stable self-assembled matrix is maintained when converting from K to R.

B) Corresponding image from transmission electron microscopy showing the structural variation of FGAR peptide self-assembly when compared to that of FGAK peptides.

C) Frequency sweep of FGAD peptide solution with G' (storage) and G'' (loss) moduli reported at 2% (w/v) showing that FGAD does not form a gel.

D) Corresponding image from transmission electron microscopy showing the amorphous structure of FGAD peptide aggregates.

Chapter 3 Tryptophan Regulates Hydrophobic Selectivity of

Nucleoporins

In the NPC, phenylalanines are essential in regulating the selective transport of molecules between the nucleus and the cytoplasm²³. Within nucleoporin sequences, however, other hydrophobic aromatic amino acids are present such as tyrosine and tryptophan. For nuclear transport receptors, it is already well established that tryptophan is essential for efficient nuclear pore transport²⁷. Yet, despite the frequent presence of such amino acids in nucleoporins, whether tyrosine and tryptophan contribute to the selective properties of intact NPCs is unknown. One reason for this disparity in analysis is perhaps the easily identifiable repeating FSFG, FG, and GLFG sequences. For tryptophan and tyrosine, no such repeating sequence has been classified. Isolated experiments replacing phenylalanines with tyrosines in nucleoporin gels reveal that qualitatively, hydrophobic selectivity is maintained⁵. However, it is unclear why phenylalanines are the predominant hydrophobic amino acid in nucleoporins. Here, we take a systematic approach to understand how substitutions between phenylalanine, tryptophan, and tyrosine affect the self-assembly of nucleoporin-like proteins as well as the selective recognition of the resultant gel.

Materials and Methods

Nucleoporin Sequence Analysis:

All protein sequences annotated as nucleoporins were exported from National Center for Biotechnology Information (NCBI). From these sequences, 22860 were identified to contain FG sequences and are designated FG nucleoporins. A python script was written to analyze

the frequency and location of phenylalanine, tryptophan, and tyrosine for all FG nucleoporin sequences.

Peptide and Gel Preparation:

Unless specified otherwise, all chemicals were obtained from Sigma Aldrich (St. Louis, MO, USA). Peptides were prepared by MIT's Koch Institute Biopolymers and Proteomics Facility (Cambridge, MA, USA) and Boston Open Labs LLC (Cambridge, MA, USA). All peptides were HPLC purified unless specified otherwise, desalted using reverse phase HPLC with 0.05% trifluoroacetic acid (TFA), and lyophilized after synthesis with >95% purity. For fluorescent reporter peptides, Cy5 NHS ester (Lumiprobe; Hallandale Beach, FL, USA) was covalently linked to the N-terminal end.

Fluorescent peptides were diluted into 200 mM NaCl with 20 mM HEPES [pH 7] at 0.4 μ M final concentration for diffusion experiments. Gel peptides were all dissolved in 20 mM NaCl, 20 mM HEPES [pH 7] at 2% (w/v). To facilitate solubilization and gel formation, peptides were vortexed for 30 s and briefly sonicated in a bath sonicator (Branson 2510) to reduce aggregation.

Capillary Diffusion Assay and Analysis:

Borosilicate square capillaries (1.5 inch) with 9 mm cross sectional width (Catalog # 8290, Vitrocom; Mountain Lakes, NJ, USA) were loaded by piercing pre-made hydrogels. Ten-micromolar solutions (200 mM NaCl, 20 mM HEPES [pH 7]) of fluorescent peptides were injected into the capillary and sealed with a 1:1:1 (by weight) mixture of Vaseline, lanolin, and paraffin. Time lapses of peptide diffusion were taken at 1-min intervals for up to 3 h on an

AxioObserver D.1 with a EC Plan-Neofluar 1.25x/0.03 WD=3.9 and Hamamatsu C11440-22CU camera. All fluorescence profiles were obtained by averaging the fluorescence intensities across the width of the capillary in MatLab (MathWorks; Natick, MA, USA). Normalized concentration profiles were obtained by normalizing fluorescence intensities to the bath concentration of the capillary at the initial time point. All data represent at least three independent replicates.

Student's *t* test was applied to determine p-values between experimental conditions.

Effective diffusion rates were fit by minimizing the squared error of a simulated concentration timecourse in a region of the capillary on the gel side of the interface over a 100-min window.

To achieve this fit, we numerically solved the diffusion equation for the concentration of probe *c*:

$$\frac{\partial c(x, t)}{\partial t} = D \frac{\partial^2 c(x, t)}{\partial x^2}$$

using MatLab's *pdepe* function (MathWorks; Natick, MA). The initial condition $c(x, 0)$ was set by the concentration profile at the first timepoint, and the boundary conditions $c(0, t)$ and $c(L, t)$ (for a fit over length L) were similarly determined by the concentration profiles at the edges of the region of interest. The D that minimized the squared difference between the simulated and actual concentration profiles was the value reported; minimization of error took place iteratively using a modified gradient descent algorithm. The window from 150-250 min was generally used for fitting, but for particularly fast-diffusing probes ($D > 1000 \mu\text{m}^2/\text{min}$), windows of 50-150 min or 30-130 min were used. The earlier time period allowed for the fitting to take place before the steady state or pseudo-steady state was reached, which was crucial for precise fitting.

Rheological Testing:

Rheological tests were performed on an Anton Paar MCR 302 Rheometer in a cone-plate geometry with a 25-mm diameter, 1° cone angle, and 51- μm truncation. The temperature was maintained at 25 °C and evaporation was controlled with an H₂O-filled solvent trap. To identify the linear regime, amplitude sweeps were conducted at $\omega = 10$ rad/s from $\gamma_0 = 0.01\%$ to 100% strain. In the linear regime, frequency sweeps were conducted using the previously determined strain amplitude from $\omega = 100$ rad/s to 0.1 rad/s.

Results

A fundamental question that can be asked is if the tryptophans and tyrosines are present at the same frequency as phenylalanine. To approach this question, we identified all FG-containing nucleoporins that are recorded in the NIH protein database across all eukaryotic organisms. Using the protein sequences, we next identified for each FG nucleoporin, how many tryptophans, tyrosines, and phenylalanines are in each protein sequence. As seen in Figure 3-1, phenylalanines are present in the highest amount, with at least 50% of the FG nucleoporins containing 50 or more phenylalanines. In contrast, there are typically a maximum of 50 tyrosines or tryptophans, each, in the nucleoporin sequences. These findings reflect the bias in understanding the role of phenylalanines due to its high propensity in the nucleoporin sequences.

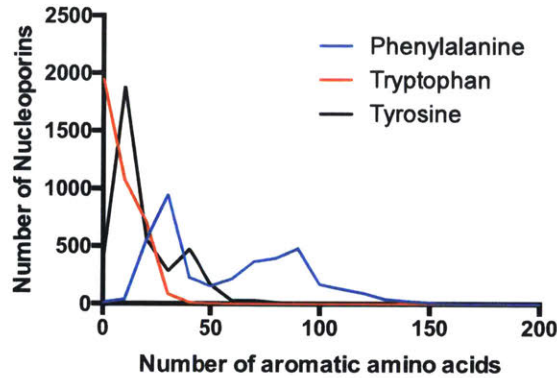


Figure 3-1: Distribution of FG nucleoporins based on the presence of phenylalanine, tryptophan, or tyrosine counts in the amino acid sequence. Phenylalanines are present in the highest amounts in FG nucleoporins, followed closely by tyrosine and lastly tryptophan.

To further characterize the contribution of these aromatic amino acids, we examined the spatial distribution along the protein sequence. We found that the spatial distribution of aromatic amino acids is non-uniform across nucleoporin sequences. Although overall, the number of phenylalanines outstrips tryptophan and tyrosine, phenylalanines are predominantly at the N-terminal region of nucleoporins, which are at the central regions of the nuclear pore complex. Tryptophans and tyrosine dominate the C-terminal regions, which are closer to the nuclear pore wall. Since the selectivity of NPCs is maintained throughout the entire channel (from center to periphery), it suggests that the presence of tryptophan and tyrosine should also act as regulators of hydrophobic selective transport. Previous structural analyses of intact NPCs revealed that the self-assembled matrix within the channel is not uniform^{100,101}. Typically, the central plug of the channel is denser than the matrix nearer to the nuclear pore wall. The structural disparities suggest that the passage of molecules through the NPC may take different routes^{102,103}. The variation in hydrophobic amino acid distribution suggests that in addition to structural changes, the various aromatic acids may have a significant role in molecular recognition of NTRs.

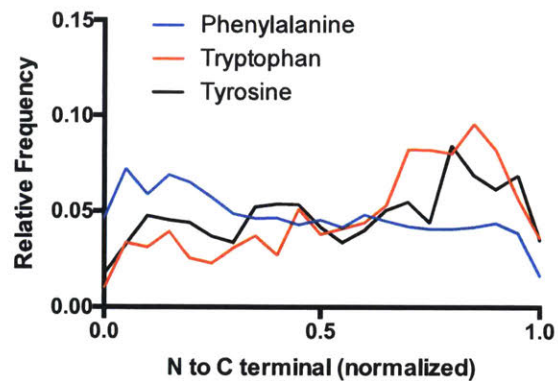


Figure 3-2: Spatial distribution of phenylalanine, tryptophan, and tyrosine residues across FG nucleoporin sequences (N to C terminal). Phenylalanines dominate at the N-terminal region at the center of NPCs whereas tryptophans and tyrosine are more prevalent at the C-terminal region along the NPC wall.

To systematically test the role of tryptophan and tyrosine in self-assembly and selective molecular recognition, we used short well-defined peptides that contain the essential components of FG nucleoporins as described in the previous chapter. We based the design of synthetic peptides on the consensus sequence identified in Nsp1, the essential FG nucleoporin found in *S. cerevisiae*¹⁰⁴. We synthesized FSFGAKAAKAFSFG as the reference peptide for all experimental comparisons and constructed two new gel-forming peptides WSWGAKAAKAWSWG and YSYGAKAAKAYSYG to test the role of tryptophan and tyrosine, respectively, in self-assembly and selective molecular recognition.

Phenylalanine and Tryptophan allow for self-assembly

For self-assembly assays, we dissolved the peptides at 2% (w/v) concentration in 20 mM NaCl, 20 mM HEPES [pH 7] and allowed for gels to form over 24 hours. FGAK and WGAK peptides formed gels with comparable stiffness (Figure 3-3A and B) whereas YGAK formed a gel that was at least three orders of magnitude less stiff in the frequency and strain regimes tested (Figure 3-3C). The tyrosine gels also flowed immediately when inverted, suggesting that the

self-assembled peptides had weak and unstable interactions. These experiments revealed that the choice of aromatic amino acid has a significant effect on the self-assembly properties of the hydrophobic domains. A small addition of a hydroxyl group to convert phenylalanine to tyrosine changes a stiff gel to a viscoelastic fluid that flows immediately when inverted. As tryptophan and tyrosine are the dominant aromatic residues at the C-terminal regions of FG nucleoporins (Figure 3-2), it implies that perhaps the self-assembled network near the wall of the nuclear pore is less cohesive and densely packed compared to the central region. Indeed, electron microscopy images have revealed previously that the central channel of the nuclear pore is more densely packed compared to the outer regions of the complex^{100,101}.

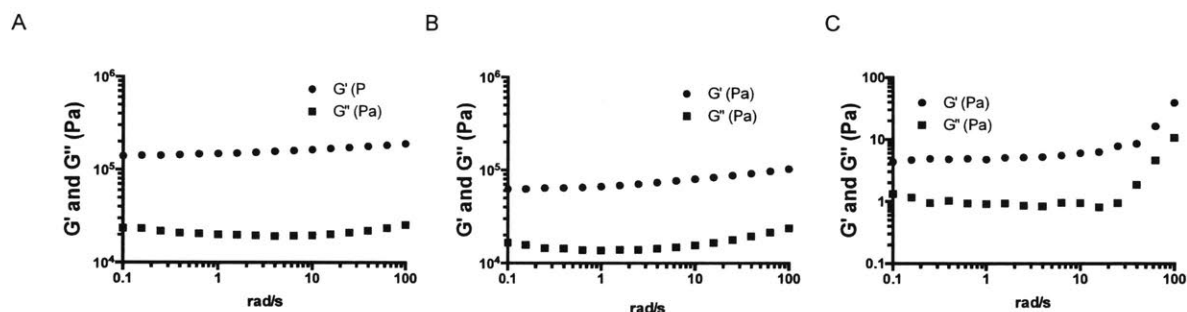


Figure 3-3: Frequency response of (A) FGAK, (B) WGAK, and (C) YGAK gels, revealing that FGAK and WGAK form gels that are four orders of magnitude stiffer than YGAK. These results suggest that phenylalanine and tryptophan are stronger regulators of hydrophobic self-assembly. Note, measurements for YGAK are below the sensitivity values for the rheometer.

As the choice of hydrophobic amino acid appears to regulate the molecular selectivity required for self-assembly, perhaps the choice between aromatic residues also determines the molecular selectivity for recognizing diffusing substrates. For the characterization of molecular recognition and its relation to transporting molecules, we only compared FGAK and WGAK gels as YGAK peptides did not form a stable self-assembled matrix. To determine how phenylalanine and tryptophan may defer in hydrophobic interactions, we analyzed how a hydrophobic fluorescent

dye diffuses into each of the FGAK and WGAK gels. As many fluorophores quench in the presence of tryptophan^{105,106}, we opted for far-red dyes that were previously experimentally verified to be more photostable in tryptophan-containing environments. For the series of transport experiments, we used Cy5 as it is neutrally charged. To test for interactions, we used the same diffusion chamber setup as described before in the previous chapter. Briefly, the FGAK and WGAK gels were formed and loaded into borosilicate capillaries. The fluorescent dye was injected into the capillary and the 1-D diffusion chamber was sealed on both ends by solidified wax. Diffusion was monitored for three hours at one-minute time intervals and concentration profiles were measured.

As seen in Figure 3-4, Cy5 interacted significantly with the FGAK and WGAK gels, with Cy5 binding at the interface and an effective diffusion coefficient of $43.7 \pm 8.4 \mu\text{m}^2/\text{min}$ in FGAK gels and $1.2 \pm 0.4 \mu\text{m}^2/\text{min}$ in WGAK gels (Table 3-1). Diffusion of Cy5 was slower in WGAK gels compared to FGAK gels, suggesting that the interactions between Cy5 and indole groups is stronger compared to Cy5 and phenyl groups. From the analysis of Cy5 diffusion coefficients, tryptophan may have different molecular selectivity properties compared to phenylalanine despite similar mechanical properties of the gel.



Figure 3-4: Cy5 reporters diffusing into FGAK (left) or WGAK (right) gels over five hours. The five-hour timepoints reveal that Cy5 is able to diffuse significantly further into FGAK than WGAK, suggesting that phenylalanine and tryptophan have different selectivities to the same hydrophobic substrates.

Table 3-1: Effective diffusion coefficient values of Cy5 into FGAK and WGAK gels when unmodified (Cy5 row) or modified by cationic residues (Hydrophilic (+)) or anionic residues (Hydrophilic (-)). The addition of cationic residues reduces interactions with the cationic FGAK and WGAK gels to increase the effective diffusivity of Cy5 despite the increase in molecular weight whereas addition of complementary anionic charges reduces the effective diffusivity.

| Reporter | FGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) | WGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) |
|-----------------|---|---|
| Cy5 | 43.7 ± 8.4 | 1.2 ± 0.4 |
| Hydrophilic (+) | 69.7 ± 28.4 | 24.2 ± 10.5 |
| Hydrophilic (-) | 1.9 ± 0.4 | 0.5 ± 0.4 |

As Cy5 is able to interact with the hydrophobic groups in the peptide gels, we next asked whether modifying the charge of Cy5 reporters affects hydrophobic selectivity. To determine whether the charge modulates binding effects of the hydrophobic Cy5 dye, Cy5 was appended to an N-terminal region of a net +7 or -7 charged peptide sequence (termed Hydrophilic (+) and Hydrophilic (-) reporters, respectively). Using the same 1-D diffusion chamber system as described in Chapter 2, we monitored the concentration profiles as they the reporters diffused into FGAK and WGAK gels. Here, we found that for the FGAK gels, the addition of cationic charges to Cy5 increased the diffusivity of the Cy5 reporter from $43.7 \pm 8.4 \mu\text{m}^2/\text{min}$ to $69.7 \pm 28.4 \mu\text{m}^2/\text{min}$ whereas the addition of anionic charges decreased the diffusivity down to $1.9 \pm 0.4 \mu\text{m}^2/\text{min}$. Similarly, for the WGAK gel, the addition of repulsive cationic charge increased the diffusivity of the Cy5 reporter from $1.2 \pm 0.4 \mu\text{m}^2/\text{min}$ up to $24.2 \pm 10.5 \mu\text{m}^2/\text{min}$ whereas

attractive anionic charges reduced the diffusion coefficient down to $0.5 \pm 0.4 \mu\text{m}^2/\text{min}$. These results showed that charge modulates the molecular selectivity of phenylalanine and tryptophan to a hydrophobic substrate such as Cy5. Moreover, reporters in the WGAK gels have a consistently lower diffusion coefficient when compared to that of FGAK gels, suggesting that the interactions between Cy5 and indole groups is once again stronger than interactions between Cy5 and phenyl groups. Taken together, these results with the Cy5-based reporters show that hydrophobic interactions between Cy5 and the tryptophan or phenylalanine containing gels are tunable by the presence of nearby charge residues. Although we show how hydrophobic interactions with Cy5 can be modulated by electrostatics, Cy5 is not a substrate that is typically selected for by native NPCs. Within *in vivo* systems, NPCs select for NTRs that typically contain tryptophan, phenylalanine, or isoleucine as the non-polar binding residues to FG nucleoporins.

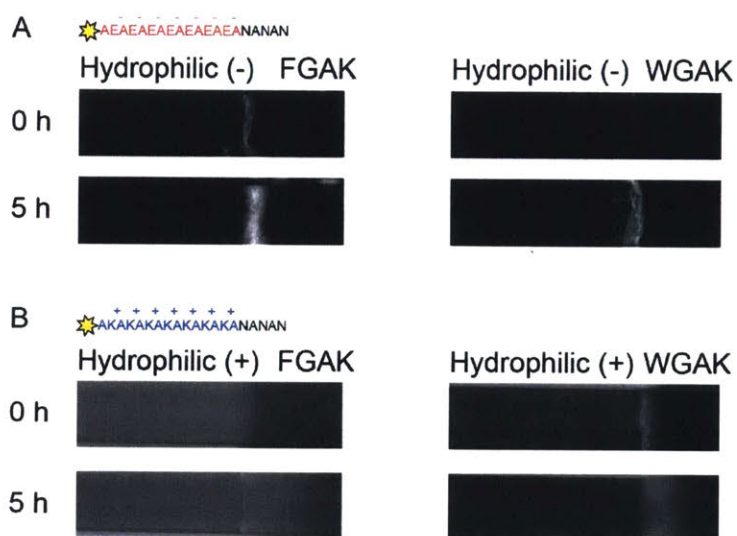


Figure 3-5: Charge regulates the hydrophobic interactions between Cy5 and phenyl groups (FGAK) or indole groups (WGAK). Complementary charge increases the binding of Cy5 at the interface of both gels (A) whereas charge repulsion minimizes the interactions and allows for diffusion of Cy5 into the gel (B). The sequence of the charged reporter is indicated in each subfigure. Star symbol refers to the fluorescent Cy5 dye.

To determine how well the FGAK and WGAK gels select for more physiologically relevant residues, fluorescent reporters containing a hydrophobic tail of phenylalanine were synthesized

as described in Chapter 2. Briefly, the asparagines (N) are converted to phenylalanines (F) to form Hydrophobic (+F) and Hydrophobic (-F) reporters (sequence structure in Figure 3-6). In Chapter 2, we showed that FGAK gels are able to select for hydrophobic F-containing reporters based on the charge profile of the substrate. We discovered that complementary charge between Hydrophobic (-F) and FGAK gels resulted in complete retention at the interface while electrostatic repulsion minimized hydrophobic interactions between Hydrophobic (+F) reporters and the gel. Electrostatic repulsion between Hydrophobic (+F) reporter and gel peptides was sufficiently strong to allow for diffusion of Hydrophobic (+F) throughout the gel matrix. In this section, we asked if WGAK gels could also discriminate the two phenylalanine-containing reporters based on the presence of charge interactions. In Figure 3-6 and Table 3-2, we found that FGAK and WGAK gels have similar selectivity for the Hydrophobic (+W) and Hydrophobic (-W) reporters. In both cases, when the reporter contains complementary charge to the gel, the diffusion coefficient is the smallest whereas electrostatic repulsion lessens the interaction with the gel matrix and allows for greater diffusivity. These results suggest that tryptophan has the same regulatory potential as phenylalanines in intact NPCs when the substrates contain phenyl groups.

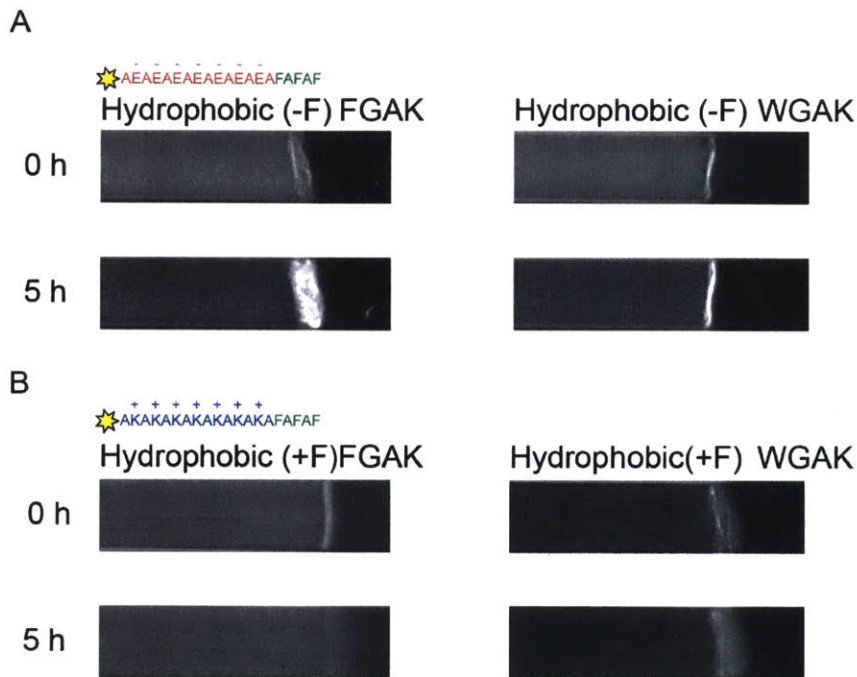


Figure 3-6: Representative images of Hydrophobic (-F) and Hydrophobic (+F) reporters into FGAK or WGAK over five hours. The images reveal that electrostatic attraction between the gel and the reporter allows for strong interactions at the interface (A) while electrostatic repulsion minimizes the hydrophobic interactions and allows for diffusion into the gel (B). These results suggest that hydrophobic selectivity is maintained even when phenylalanines are substituted with tryptophans in the gel-forming peptides.

Table 3-2: Diffusion coefficient values of Cy5 reporters containing physiologically relevant phenylalanines to facilitate hydrophobic interactions with FGAK or WGAK gels. The diffusion coefficient values show that the hydrophobic interactions with phenyl groups can be modulated by the presence of charge repulsion (Hydrophobic (+F)) or electrostatic attraction (Hydrophobic (-F)) to increase or decrease diffusivity, respectively.

| Reporter | FGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) | WGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) |
|------------------|---|---|
| Hydrophobic (+F) | 5.8 ± 1.7 | 3.5 ± 1.6 |
| Hydrophobic (-F) | 1.4 ± 0.4 | 0.3 ± 0.2 |

To broaden the perspective further, we next asked if FGAK and WGAK gels are also able to select for reporters containing other hydrophobic residues in a similar manner. In this section, we replaced phenylalanines in the Hydrophobic (+F) and Hydrophobic (-F) reporters with tryptophan (W; Figure 3-7) as W is a residue essential in NTR transport such as NTF2²⁷. As seen in Figure 3-7, all four reporters interact at the interface with minimal diffusion into either

FGAK or WGAK gels. The quantification of diffusion coefficients (Table 3-3) revealed that the interactions with tryptophan reporters are significantly strong enough that charge is no longer a major determinant in the molecular recognition capabilities of FGAK and WGAK gels. These results suggest that the choice of phenylalanine or tryptophan does not have a major impact on the selective properties on the self-assembled gel. However, from the transporting substrate perspective, the choice of tryptophan appears to minimize the impact that charge has on molecular recognition and suggests another design parameter in which to tailor the rate of selective transport across intact NPCs.

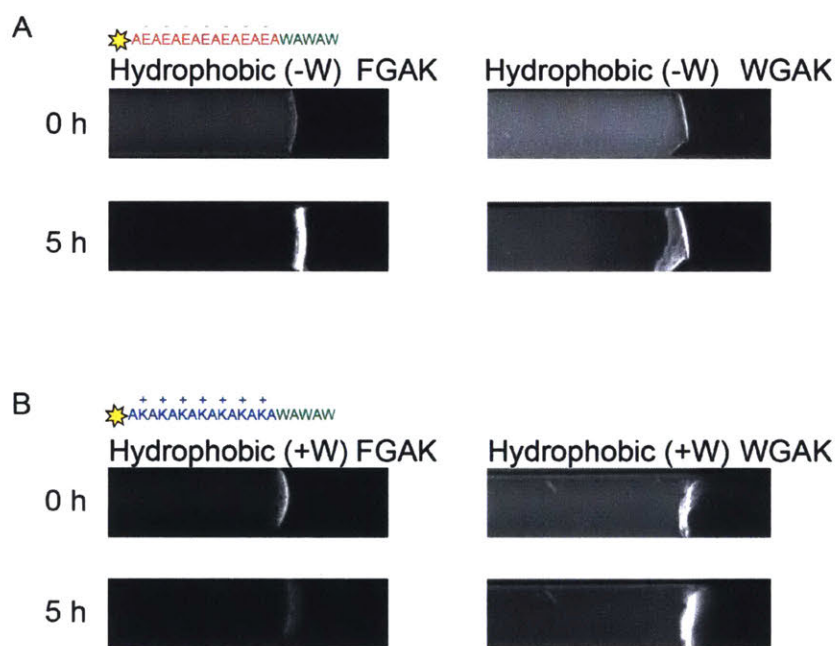


Figure 3-7: Representative time series of A) Hydrophobic (-W) and B) Hydrophobic (+W) reporters into FGAK and WGAK gels over the course of five hours. In all four scenarios, the tryptophan reporters interact strongly at the gel interface, suggesting that the tryptophan interactions are influenced less by the presence of charge.

Table 3-3: Diffusion coefficient values of Cy5 reporters containing physiologically relevant tryptophans to facilitate hydrophobic interactions with FGAK or WGAK gels. The diffusion coefficient values show that the hydrophobic interactions with indole groups is less sensitive to the presence of charge repulsion (Hydrophobic (+W)) or electrostatic attraction (Hydrophobic (-W)).

| Reporter | FGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) | WGAK: D_{eff} ($\mu\text{m}^2/\text{min}$) |
|------------------|--|--|
| Hydrophobic (+W) | 2.5 ± 1.6 | 1.6 ± 0.2 |

| | | |
|------------------|---------------|--|
| Hydrophobic (-W) | 0.5 ± 0.1 | Estimated <0.2. Below limit of analysis. |
|------------------|---------------|--|

Discussion

Hydrophobic exclusion mediated by FG domains is a well-established theory for how NPCs are able to control the selective passage of proteins and RNA across the nuclear membrane⁶. The selectivity is largely attributed to the presence of phenylalanines whereas other non-polar amino acids such as tryptophan and tyrosines have largely been ignored. Despite individual experiments showing that replacement of phenylalanines with tyrosine also allow for hydrophobic selectivity⁵, there has been no systematic dissection of the functional differences between phenylalanine, tryptophan, and tyrosine that would explain the strong bias towards phenylalanine. Here, we show how tryptophan can replace phenylalanine to produce self-assembled structure with similar mechanical and selective properties. Tyrosine, although close in molecular structure to phenylalanine, does not facilitate self-assembly as well and perhaps is not as stringent compared to phenylalanine and tryptophan. However, as previous experiments showed that tyrosine does have selective function⁵, it suggests that tyrosine may play a supporting role in maintaining selectivity in intact NPCs.

Altogether, these results suggest that tryptophan and tyrosine, although not as prevalent within NPCs, may play a significant role in governing the selective properties of the channel. Moreover, the discrepancies in molecular recognition between phenylalanine, tryptophan, and tyrosine are predicted to be an additional parameter in modulating the spatial transport routes of proteins through the complex. Future analysis into the binding kinetics and structure of phenylalanine-, tryptophan-, and tyrosine-based gels will provide

insight into why phenylalanine is the preferred aromatic amino acid within FG nucleoporin sequences.

Chapter 4 Saliva and Salivary Mucin MUC5B Inhibit

Transformation of Cavity-causing *Streptococcus mutans*

In the oral cavity, a key defensive mechanism against microbial infection is saliva¹⁰⁷⁻¹¹⁰. The human body naturally produces copious amounts of the viscoelastic fluid, which hydrates wet epithelial surfaces. A key constituent of saliva is the secreted glycoprotein known as mucins. These megadalton polymers, such as MUC5B, are highly glycosylated and give saliva its unique lubricating properties^{111,112}. Recent studies also show that MUC5B prevents biofilm formation of key cavity-causing bacteria, such as *Streptococcus mutans*^{16,113}.

Although mucins are recognized to be an essential regulator of biofilm formation, how they interfere in other virulence traits is still uncharacterized. In the oral cavity, *S. mutans* has the ability to form tenacious plaques and resist environmental stressors¹¹⁴⁻¹¹⁶. These attributes are considered *S. mutans*' major virulence traits and allow for persistent colonization. Investigations into these pathways revealed that *S. mutans* virulence is linked to competence pathways^{115,117,118}, which allow for uptake of foreign genetic elements. As MUC5B has been shown to inhibit biofilm formation of *S. mutans*¹¹³, we hypothesize that MUC5B should also influence natural transformation of *S. mutans*. Here, we show that MUC5B lowers rates of *S. mutans* transformation, whereas a commonly used mimetic, carboxymethylcellulose (CMC), is unable to recapitulate the properties. Moreover, we show that inhibition by MUC5B is robust to changes in extracellular concentrations of DNA and quorum sensing molecules that activate competence pathways. Last, we show that these inhibitory effects in MUC5B translate to native salivary conditions.

Materials and Methods

Strains and Reagents:

The bacterial strain *Streptococcus mutans* UA159 was kindly given as a gift by Dan Smith (Forsyth Institute). For general propagation and storage, UA159 was cultured and maintained in Todd-Hewitt (TH, BD 249240) at 37C with 5% CO₂. Plasmid DNA pVA838 (ATCC 37160) for transformation assays was purchased from ATCC and maintained in *E. coli* DH5 α (BD LB medium, 10 ug/ml chloramphenicol). All TH media was used within seven days of autoclaving.

Horizontal Gene Transfer Assay:

For all gene transfer assays, unless specified separately, UA159 cultures were grown overnight in 50% TH media at 37C, 5% CO₂ under static conditions. The next morning, cultures were diluted 1:20 into the 25% TH media in 100 uL volumes in 96-well plates and cultured statically for 2 hours. pVA838 (1.2 μ g/ml unless otherwise noted) was added and gently mixed into the solution and allowed to incubate for an additional two hours before serially diluting 1:10 in phosphate buffer saline (PBS). For transformants, dilutions were plated on selective TH agar plates (10 μ g/ml erythromycin). For total population counts, serial dilutions were plated on non-selective TH agar plates. All data presented are biological replicates with standard deviations reported for error bars.

For biofilm promoting conditions, media was supplemented with 1% (w/v) sucrose. During the serial dilution process, the planktonic phase was gently pipetted and separated. The attached biofilm was washed three times by 100 ul of sterile PBS. Biofilms were then detached by vigorous scraping with a pipette tip for 30 seconds and mixing prior to serial dilution.

For comparing media conditions, two alternate media conditions were used. For nutrient rich media conditions, UA159 was cultured in TH supplemented with 5% yeast extract (THYE) or TH alone for both overnight and gene transfer assays.

When testing the inhibitory effects of MUC5B, 0.5% (w/v) MUC5B was dissolved overnight in the appropriate THYE, TH, or 25% TH at 4° C, shaking, to disperse the polymer.

Saliva Collection:

Submandibular saliva was collected from 8 informed consenting volunteers using a custom vacuum pump setup as described previously¹¹³. To prevent suction of parotid gland secretions, cotton pads were placed between the molars and cheeks. To collect the submandibular secretions, the vacuum line was placed underneath the tongue. The collection vessel was placed on ice during this process to prevent degradation of proteins. For whole saliva testing, submandibular saliva was aliquoted into 1 mL volumes and snap frozen in liquid nitrogen and stored at -80C. Protocols involving the use of human subjects were approved by Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects.

MUC5B Purification:

After collection, saliva was diluted with 5.5 M NaCl, such that the final salt concentration was 0.16 M. Antibacterial agents and protease inhibitors were then added at the appropriate final concentrations: 0.04% (w/v) sodium azide, benzamidine HCl (5 mM), dibromoacetophenone (1 mM), phenylmethylsulfonyl fluoride (1 mM), and EDTA (5 mM, pH 7). For solubilization,

saliva solutions were gently stirred overnight at 4C. For storage, saliva was snap frozen in liquid nitrogen in 50 mL volumes before being placed at -80C. To remove cellular debris, saliva was thawed at 4C and centrifuged at 3,800g for 10 minutes in a swinging-bucket centrifuge. MUC5B was then purified by fast protein liquid chromatography on a Bio-Rad NGC equipped with an XK 50 column packed with 2 L of Sepharose CL-2B resin (GE HealthCare Biosciences).

Results

Prior to assaying the effect MUC5B has on transformation rates, we first established protocols that maximized the competent fraction of *S. mutans* populations. Previous reports suggested that modifying nutrient composition and yeast extract percentages regulate competency. Here, we used standard Todd-Hewitt yeast extract (THYE) as the baseline, and explored how removal of yeast extract (100%TH) and dilution of nutrients (25%TH) affected transformation rates. As readout of successful genetic uptake, we used the shuttle vector pVA838 that confers erythromycin resistance. As seen in Figure 4-1A, we found that the 25%TH condition promotes the highest transformation rates. Since 25%TH maximized the population of transformable *S. mutans*, all further transformation assays were performed in 25%TH unless otherwise stated.

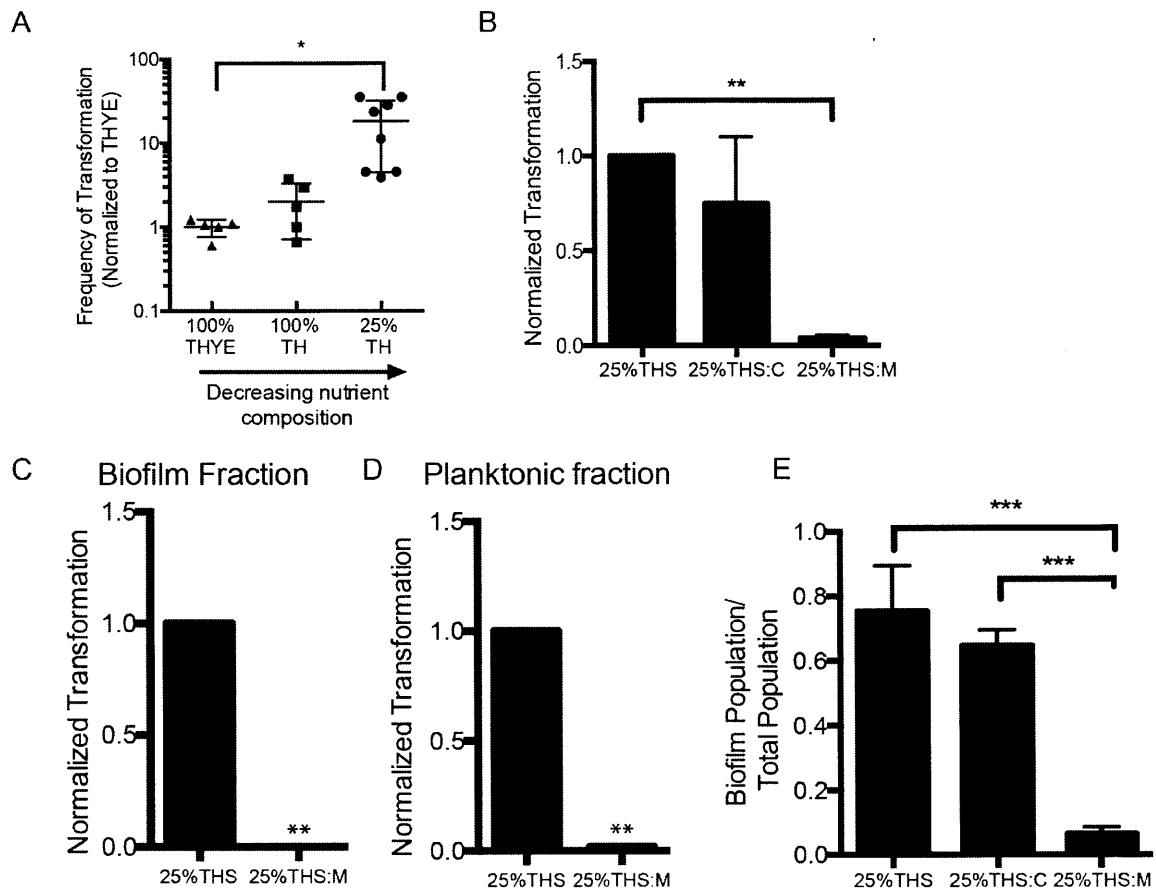


Figure 4-1: MUC5B suppresses *S. mutans* transformation in the biofilm state

- A) Identification of maximal transformation rates in rich 100%THYE media, 100%TH media, and nutrient-limited 25%TH showing that transformation is upregulated in lower nutrient conditions. N=3, * indicates $p < 0.05$, ** indicates $p < 0.01$.
- B) Normalized transformation rates of *S. mutans* 25%THS supplemented with either 0.5% (w/v) CMC (25%THS:C) or MUC5B (25%THS:M), showing inhibition of transformation by MUC5B only. Each independent biological experiment is normalized to the respective media-only condition. N = 3, * indicates $p < 0.05$, ** indicates $p < 0.01$.
- C) Normalized transformation rates of *S. mutans* for the biofilm population fraction showing that transformation is inhibited in the biofilm subpopulation.
- D) Normalized transformation rates of *S. mutans* for the planktonic population fraction showing that transformation is inhibited in the planktonic subpopulation.
- E) Fraction of surface-attached *S. mutans* populations in 25%THS when supplemented with 0.5% (w/v) CMC (25%THS:C) or 0.5% (w/v) MUC5B (25%THS:M) showing that biofilm formation is inhibited by MUC5B, which may lead to inhibition of transformation.

With the established media conditions, we first tested if MUC5B affected natural transformation rates of *S. mutans* in biofilms. Sugars, such as sucrose, are known to enhance *S. mutans* surface attachment and biofilm formation to dental enamel. To promote biofilm formation, we

supplemented 25% TH with 1% sucrose (25%THS) and compared transformation rates of *S. mutans* between MUC5B (25%THS:M) and media-only conditions. We found that MUC5B (0.5% w/v) suppressed transformation by 96% in the entire culture when compared to the untreated culture (Figure 4-1B). Further examination of the breakdown of transformed populations revealed that MUC5B inhibited transformation in both biofilm and planktonic subpopulations (Figure 4-1C and D).

To understand if inhibition was specific to MUC5B, we examined if another polymer recreated similar effects. We chose carboxymethylcellulose (CMC) for comparison, as it is an anionic polymer commonly used in salivary mimetics¹¹⁹⁻¹²¹. Using 0.5% (w/v) CMC (25%THS:C), we find that transformation rates of *S. mutans* are not inhibited (Figure 4-1C). The comparison with CMC suggested that physical polymer networks alone are not sufficient to inhibit transformation in biofilm-forming conditions. A possibility for the difference in transformation rates is due to MUC5B preventing biofilm formation. As biofilm formation is linked to increased competence of *S. mutans*^{117,122}, any inhibition of biofilm formation correlates with decreased transformation. We showed that in MUC5B conditions (Figure 4-1E), the biofilm population is only 10% of the total population, as opposed to in media and CMC conditions, where the biofilm fraction consists of over 60% of the total population. These results reveal that MUC5B inhibits transformation, whereas a salivary mimetic such as CMC is unable to.

To determine if MUC5B inhibits transformation independent of biofilm formation, we tested for transformation rates in conditions that decouple the effects of biofilm formation and competence. Therefore, we used 25%TH, which we previously showed to have the highest transformation

rates (Figure 4-1A) and does not contain sucrose for biofilm formation. With MUC5B (25%TH:M), we found that transformation rates were lowered by 96% compared to the wildtype rates (Figure 4-2A), without affecting total population counts (Figure 4-2B). To ensure that the suppression of transformation rates was due to MUC5B and not by DNA degradation from residual nucleases, we heat-treated the mucin solution to inactivate any residue nucleases during the purification process (Supplementary Figure 4-1). Again, we found the same inhibitory strength as before (Figure 4-2B and C). These experiments reveal that MUC5B was able to inhibit transformation of *S. mutans* through mechanisms other than disruption of biofilm formation.

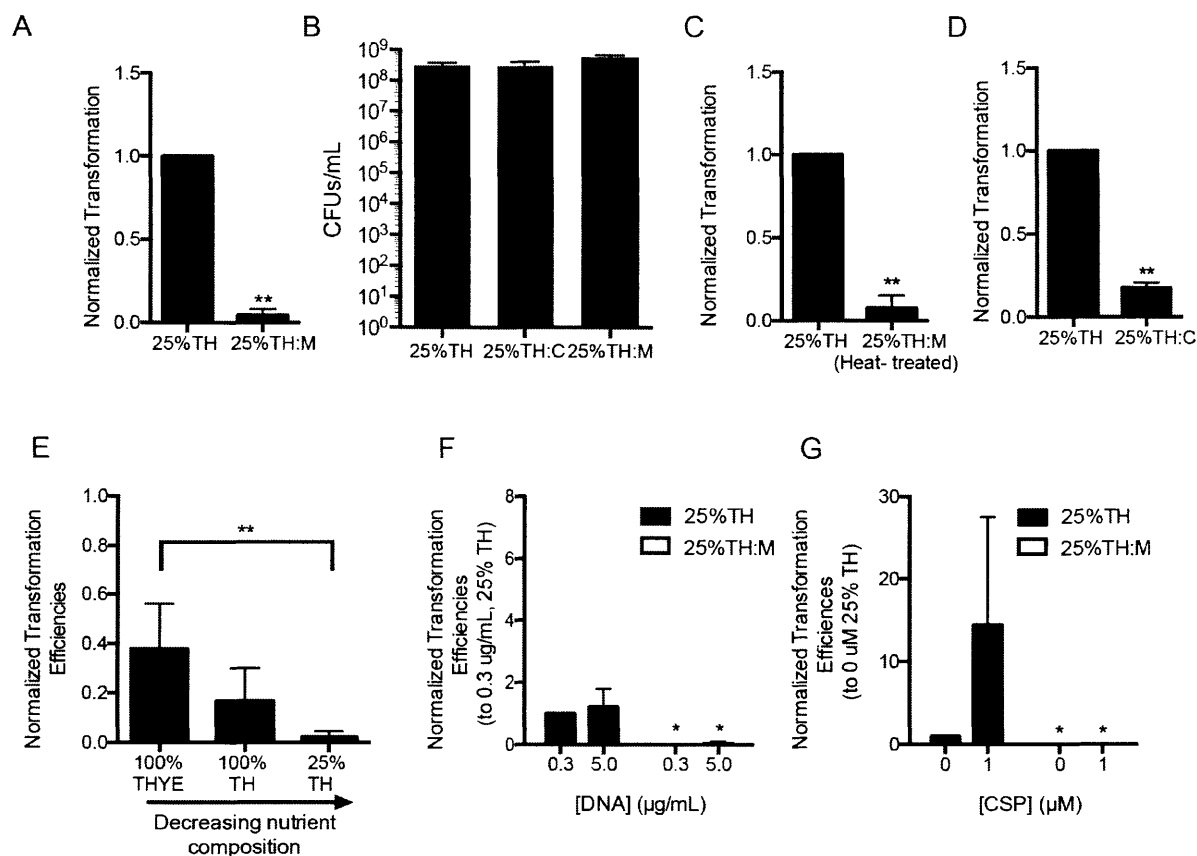


Figure 4-2: MUC5B inhibits *S. mutans* transformation in planktonic-only conditions and is robust to changes in extracellular DNA and CSP

- A) Normalized transformation rates of *S. mutans* in 25%TH and 25%TH:M, showing that transformation inhibition is not dependent on biofilm induction. Each independent biological experiment was normalized to its respective media-only condition. N = 3, * indicates $p < 0.05$, ** indicates $p < 0.01$.
- B) Total colony forming units (CFU) in 25%TH, 25%TH:C, and 25%TH:M conditions showing that CMC and MUC5B do not affect total population counts. N=3 independent experiments.
- C) Normalized transformation rates of *S. mutans* in 25%TH and 25%TH:M where MUC5B has been heat-treated. Inhibition of transformation is still maintained, showing that residual nuclease activity is not the inhibitory mechanism. Each independent biological experiment was normalized to its respective media-only condition. N=3, ** indicates $p < 0.01$.
- D) Normalized transformation rates of *S. mutans* in 25%TH and 25%TH:C, showing that CMC has some inhibitory effect against transformation. Each independent biological experiment was normalized to its respective media-only condition. N = 3, * indicates $p < 0.05$, ** indicates $p < 0.01$.
- E) Normalized transformation rates of *S. mutans* in rich 100%THYE, 100%TH, and 25%TH all supplemented with 0.5% MUC5B to show that inhibition of transformation by MUC5B is condition dependent.
- F) Normalized transformation rates of *S. mutans* in 25%TH and 25%TH:M supplemented with either 0.3 $\mu\text{g/ml}$ pVA838 or 5 $\mu\text{g/ml}$ pVA838 showing that inhibition by MUC5B is independent of DNA concentration. Each biological replicate is normalized to the 0.3 $\mu\text{g/ml}$ condition in 25%TH media only. N = 3, ** indicates $p < 0.01$ between comparisons of same DNA concentrations.

- G) Normalized transformation rates of *S. mutans* in 25%TH and 25%TH:M supplemented with either 0 μ M CSP or 1 μ M CSP showing that inhibition by MUC5B is independent of CSP concentration. Each biological replicate is normalized to the 0 μ M condition in 25%TH media only. N = 3, * indicates $p < 0.05$.

To investigate how MUC5B inhibits transformation of *S. mutans* in the planktonic state, we explored how polymer entanglement and biochemical signaling affects transformation rates. Previous research has shown that diffusion of plasmid DNA is restricted in mucin environments due to polymer entanglement¹²³. The physical hindrance may be sufficient to account for lack of DNA uptake by *S. mutans*. We confirmed via gel electrophoresis that MUC5B entangled with pVA838 to prevent DNA migration. pVA838 is identified and recovered only when MUC5B is digested enzymatically by pronase (Supplementary Figure 4-2). To determine if polymer entanglement is sufficient to prevent transformation, we tested if the anionic polymer CMC recapitulates the inhibitory response in planktonic conditions. As seen in Figure 4-2D, CMC was only able to inhibit transformation rates by 83%. This result suggests that in planktonic conditions, polymer entanglement affects transformation rates. However, in biofilm conditions with sucrose, addition of CMC is unable to inhibit transformation (Figure 4-2B), suggesting that polymer network entanglement is not sufficient to prevent transformation in every condition. This suggests that mucins may be inhibiting transformation by multiple mechanisms.

Biochemical cues from the environment are essential regulators of *S. mutans* competence. By modulating the nutrient composition of the media MUC5B is dissolved in, we determined how significant biochemical signaling by MUC5B influences the suppression of transformation. In rich Todd-Hewitt media with yeast extract added (THYE), we found that MUC5B is only able to suppress transformation rates by 60%. In Todd-Hewitt (TH) only, transformation rates were suppressed by 80%, and in nutrient-limited 25% Todd-Hewitt media (25% TH), we found the

greatest suppression with nearly 98% inhibition (Figure 4-2E). These results show that MUC5B inhibit transformation of *S. mutans* in a nutrient-dependent manner.

As environmental factors determine how effective MUC5B is at inhibiting transformation, we examined two parameters that are known to upregulate transformation rates: exogenous DNA and competence stimulating peptide (CSP) concentrations^{117,122,124}. To dissect how environmental concentrations of DNA affect transformation rates in the presence of MUC5B, we titrated in 0.3 or 5.0 $\mu\text{g/mL}$ of pVA838. Without MUC5B, we found that the 0.3 $\mu\text{g/mL}$ and 5.0 $\mu\text{g/mL}$ have comparable transformation rates, suggesting that 0.3 $\mu\text{g/mL}$ of DNA was sufficient to saturate the transformation rates of *S. mutans* in the media conditions tested. With MUC5B, the increase in DNA concentration was unable to recover the wildtype transformation rates (Figure 4-2F), indicating MUC5B inhibited transformation across varying concentrations of DNA. We next examined if transformation rates could be increased in the presence of MUC5B by supplying CSP exogenously. CSP is a peptide produced by the comCDE pathway, which is an essential inducer of competence in complex media^{122,124}. Exogenous addition of CSP is known to significantly promote competence of *S. mutans*. As seen in Figure 4-2G, we found that in the media-only conditions, the addition of 1 μM CSP was able to increase transformation rates by 15-fold. However, in the presence of MUC5B, the addition of CSP was unable to recover the wildtype transformation rates of the media-only conditions. These assays reveal that changes in DNA and CSP concentration do not affect MUC5B inhibition of *S. mutans* transformation.

With MUC5B alone significantly reducing the rate of *S. mutans* transformation in both planktonic and biofilm-inducing conditions, we next asked if MUC5B-based suppression

translates in the context of native saliva. To test the role of saliva on *S. mutans* transformation, we heat-treated the saliva to inactivate nucleases and sterilize the solution (Supplementary Figure 4-3A and B). We expected that MUC5B in saliva is stable in this treatment as Muc5b alone was still able to inhibit transformation after being heat-treated (Figure 4-2C). Again, we first tested whether saliva inhibited transformation in the presence sucrose, which promotes both biofilm formation and competence. As seen in Figure 4-3A, even when 1% sucrose was added, transformation rates were inhibited by 99%. Analyzing the breakdown of sessile and planktonic populations, we found that that whole saliva was able to inhibit gene transfer rates in both populations (Figure 4-3B and C). Moreover, the inhibition by saliva also appeared to be independent of biofilm formation, as *S. mutans* is able to form robust biofilms in the presence of saliva (Figure 4-3D). The biofilm formation in the presence of saliva can be attributed to constituents such as agglutinins, which have previously been shown to promote biofilm formation despite the presence of MUC5B¹¹⁴.

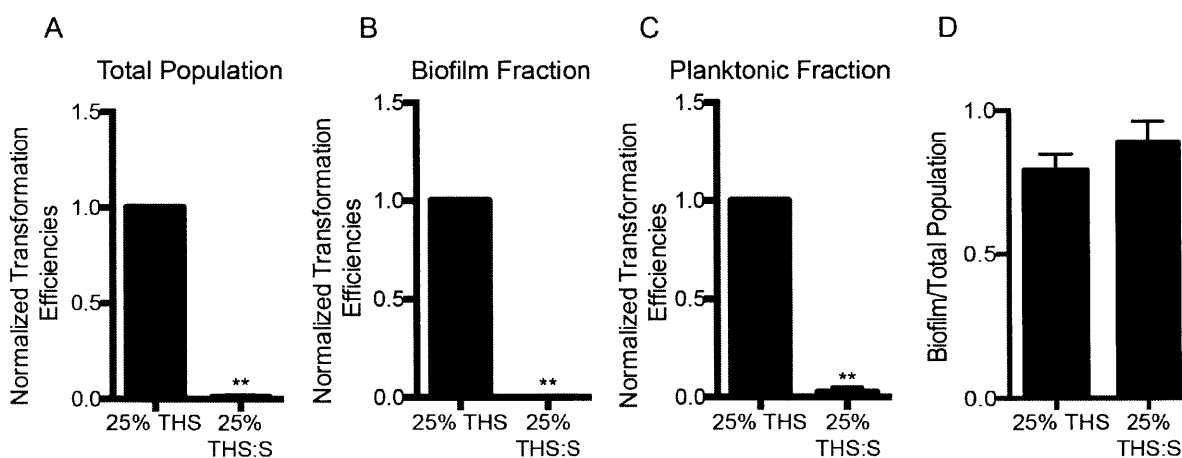


Figure 4-3: *S. mutans* natural transformation is inhibited by native saliva in biofilm-forming conditions. The inhibition is seen in the total population (A) as well as the biofilm (B) and planktonic (C) fractions. In saliva conditions, biofilm formation is not prevented, suggesting that transformation inhibitory effects by saliva are independent of biofilm pathways

To further characterize the inhibitory properties of saliva, we systematically tested whether saliva prevented gene transfer when biofilm formation and competence are decoupled in 25% TH. In these planktonic cultures of *S. mutans*, we found that saliva was able to inhibit transformation rates by 99% (Figure 4-4A). We confirmed that the inhibition was not due to loss of pVA838 from nuclease-mediated degradation as DNA could be recovered after 2 hours of incubation with heat-treated saliva (Supplementary Figure 4-3B). Moreover, the inhibition by saliva also appeared robust to exogenous addition of DNA and quorum sensing molecule CSP (Figure 4-4B-C), similar to what was observed before with MUC5B alone. These results indicate that for biofilm and planktonic phase *S. mutans*, the results from MUC5B can be translated to native saliva conditions as well.

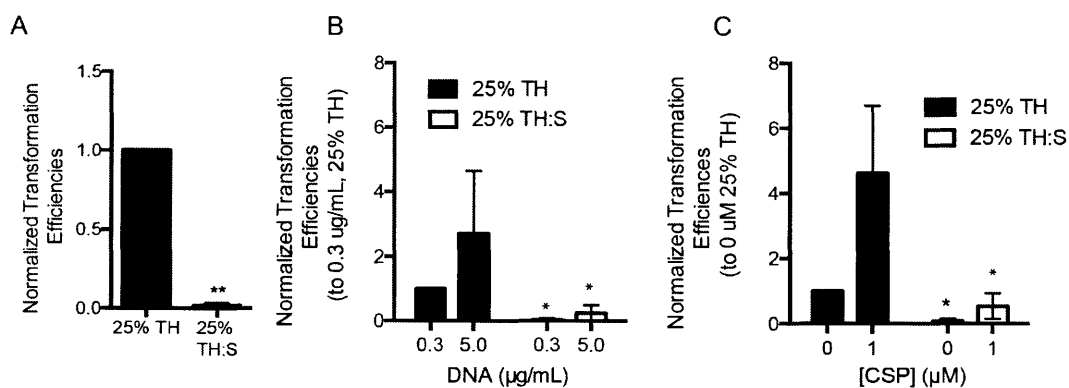


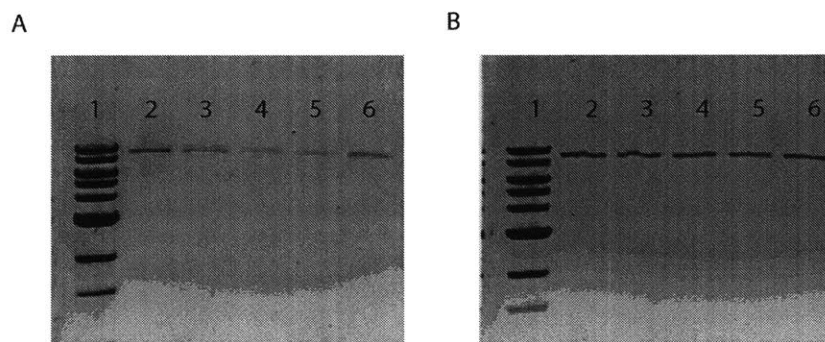
Figure 4-4: Inhibition of *S. mutans* transformation occurs in non-biofilm forming conditions (A) and is independent of DNA (B) and CSP (C) concentrations, suggesting that all the effects observed in MUC5B conditions translate to native saliva conditions.

Discussion

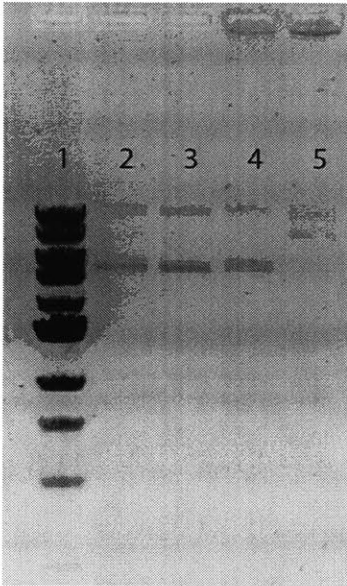
Oral health is primarily determined by host-microbial interactions. Here, we show that MUC5B inhibits rates of gene transfer for the cavity causing *S. mutans* without affecting overall microbial population counts. Moreover, we identify that the inhibition is robust to perturbations in DNA concentrations and exogenous addition of CSP. This work reveals that MUC5B may have a significant impact on the physiology of *S. mutans* and that such inhibitory processes remain

intact in the native salivary context. A possible mechanism for the robust inhibition by MUC5B is the glycan composition that is displayed on the protein backbone. Recent work has shown that *S. mutans* competence can be regulated by carbohydrate transport systems such as the phosphoenolpyruvate:sugar phosphotransferase system (PTS) system¹²⁵. Although the direct link between glycan metabolism and competence pathways has not been established, the glycans on MUC5B may modulate the energy-sensing and metabolic pathways to downregulate stress responses, such as biofilm formation and competence. Further studies to examine how individual mucin-based glycans affect the physiology and genetic response of *S. mutans* will be needed to fully characterize the mechanisms that govern the competence.

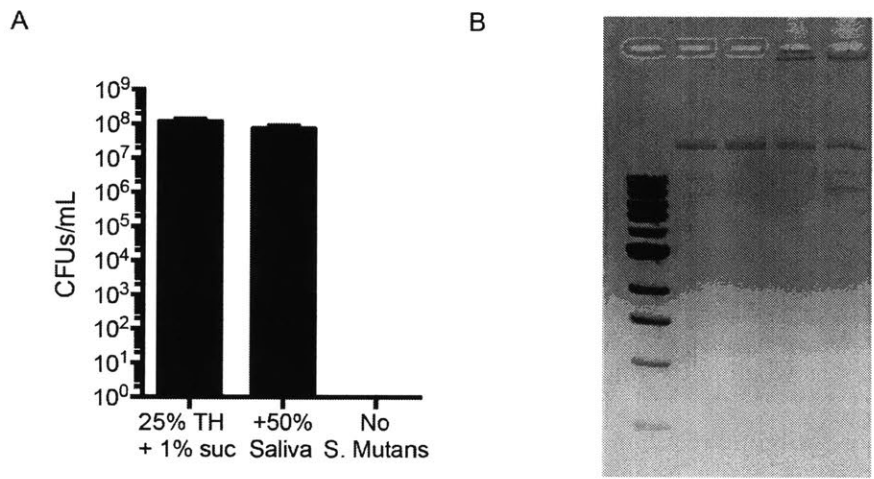
Supplementary Figures



Supplementary Figure 4-1: Gel images showing residual nuclease activity of purified MUC5B. A) From left to right, gel electrophoresis of NEB 1 kb ladder (1), pVA838 in 25%TH for 0 and 2 hour incubations (2 and 3), 25%TH + 0.5% MUC5B for 0 and 2 hour incubations (4 and 5), and a pVA838-only control (6). MUC5B solutions are digested with 1 mg/ml pronase to break down MUC5B polymers and release pVA838 for electrophoresis. B) Same description as before where lanes 4 and 5 contain heat-treated MUC5B (95°C for 5 minutes).



Supplementary Figure 4-2: Gel image showing how intact MUC5B prevents migration of DNA during gel electrophoresis. From left to right, gel electrophoresis of NEB 1 kb ladder, pVA838 in 25%TH for 0 and 2 hour incubations and 25%TH + 0.5% MUC5B for 0 and 2 hour incubations.



Supplementary Figure 4-3: Heat-treatment of saliva sterilizes the solution (A) and inactivates the nucleases (B). In the gel (B), the lanes from left to right are 1 kb NEB ladder, 0 hour incubation of pVA838 in 25%TH, 2 hour incubation of pVA838 in 25%TH, 0 hour incubation of pVA838 in 25%TH:S, and 2 hour incubation of pVA838 in 25%TH:S.

Chapter 5 Gastric and Intestinal Mucins and Mucus Modulate Conjugation Rates for Commensal Bacteria

The large intestine houses the majority of our gut microbiota. Recent investigations into the interactions of gut flora with host cells has revealed a complex dynamic that influences immune response¹²⁶⁻¹²⁹, brain function¹³⁰⁻¹³², and overall health^{110,133,134}. With the tremendous associations of gut flora to overall health and wellbeing, there is a significant push to use microbes as potential therapeutics to treat diseases with unmet clinical needs. Current examples include delivering fecal matter transplants (FMTs) to cure recurring *Clostridium difficile* infections, using microbes to modulate immune responses in diabetes, and also affecting neurological diseases such as autism, Parkinson's and Alzheimer's with bacteria. With the advent of using microbes as novel therapeutics, readouts have typically focused on the clinical efficacy of treatment and the ability of the microbes to engraft for long-term colonization. However, what is not well characterized is how these microbes influence the dynamics of the innate host microbiota and how long-term colonization affects the physiology of these artificially introduced microbes.

Of particular interest is whether the microbes used in treatments are able to undergo HGT and acquire virulence or antibiotic resistance genes. In the large intestine, which houses the majority of the gut microbiome, the high density of microbes and large surface area provides a unique environment where events such as conjugation are likely to happen¹³⁵. Conjugation between microbes can be divided into several categories based on the genetic mechanisms employed. To date, several classes of genetically mobilizable elements have been identified such as the F-plasmid, IncP plasmid groups, and IncQ as examples. For more in depth reviews on the types

and mechanisms, refer to the review by Lanka *et al*¹³⁶. From *in vitro* experiments using *E. coli*, conjugation typically acquires adherence to a surface (such as in biofilms) for stable formation of sex pili between two microbes¹³⁷. Based on these *in vitro* assays and genomic analysis of spread of virulence and antibiotic genes, it's postulated that that conjugation occurs at a significant rate within the gut and contributes to issues such as antibiotic resistance development.

Despite the significant results from *in vitro* assays and genome analysis, it is still unclear as to how well this translates to conjugation frequencies in the native context of mucosal linings. One drawback of the *in vitro* system is that conjugation occurs on a dehydrated agar surface¹³⁸ instead of the aqueous condition of a mucus hydrogel. This dehydrated surface allows for stable pili formation between acceptors and donors, which may not be possible in the mucus environment. Here, we first delve into asking fundamental questions as to how Muc5ac (gastric) and Muc2 (intestinal) are able to affect conjugation rates between two lab strains of *E. coli*. We identify that both Muc5ac and Muc2 are able to inhibit conjugation rates by >20-fold whereas a synthetic mimic such as CMC is unable to inhibit this process. Moreover, the inhibitory properties are also found to translate to mucus obtained from the gastric and intestinal lining. We then broadened the comparisons to test how bacteria such as *Lactobacillus reuteri*, a probiotic strain typically found in dairy products can conjugate with *E. coli* in a mucin environment. Here, we showed that conjugation rates are not inhibited in Muc2 and can be upregulated by >100 fold in the presence of intestinal mucus. Altogether, these results reveal how the mucus environment acts as potent regulator of microbial processes such as conjugation.

Materials and Methods

Strains and Reagents:

Unless otherwise specified, all reagents were obtained from Sigma Aldrich. *E. coli* strain OverExpressC41(DE3)pLysS was obtained from Lucigen Corporation (Wisconsin, USA). *E. coli* strain S17 λ carrying pSC189-KanR plasmid was a gift from Nicole Kavanaugh. Strain *Lactobacillus reuteri* 6475 was obtained from ATCC.

Conjugation assay:

Overnight cultures of *E. coli* were grown in 5 mL Luria-Bertani (LB) broth at 37° C, 225 rpm. For conjugation assays, S17 λ (donor strain) was mixed 1:20 ratio with OverExpressC41(DE3)pLysS (acceptor strain; Lucigen) in LB and diluted 1:100 into 100 μ L of fresh LB in 96-well plates. Conjugation was allowed to occur for five hours at 37° C in a static environment. For *E. coli* to *L. reuteri* conjugation assays, both strains were cultured overnight at 37° C in Brain Heart Infusion (BHI, Dibco). Conjugation assays were carried out as described above but mating was allowed to occur for 24 hours.

For estimating conjugation events, serial dilutions of the co-culture were plated onto LB agar plates containing 50 μ g/ml kanamycin and 30 μ g/ml chloramphenicol and incubated overnight at 37° C. For total colony forming unit population counts, the co-culture was plated on LB agar plates containing only 30 μ g/ml chloramphenicol to select for acceptor cells.

Muc2 and Muc5ac purification from porcine stomachs and small intestines:

Gastric mucus was obtained from porcine stomachs whereas intestinal mucus was obtained from porcine small intestines. Muc5ac and Muc2 were purified as described previously¹³⁹ minus the CsCl gradient. For conjugation assays, Muc5ac and Muc2 were dissolved overnight in LB at 0.5% (w/v) concentrations with constant shaking at 4° C.

Porcine gastric and intestinal mucus isolation:

Frozen porcine stomachs and small intestines were thawed for up to 48 hours. Mucus was scraped from stomach and small intestinal linings and immediately flash frozen in liquid nitrogen. For conjugation experiments, mucus was thawed at room temperature and diluted to 12.5% (v/v) in sterile LB. To sterilize, the LB-mucus solution was heat-treated at 95° C for 20 minutes.

Results

Within the gastrointestinal tract, many serotypes of *E. coli* can colonize the mucosal lining¹⁴⁰. Generally, *E. coli* is viewed as a benign inhabitant of the gut and may even produce beneficial effects¹⁴¹. However, certain serotypes such as the enteropathogenic *E. coli* (EPEC) and enterotoxigenic *E. coli* (ETEC) are pathogens that can cause illnesses such as diarrhea¹⁴². These are typically transmitted through fecal contact or contaminated water and food. *E. coli* are also a species of bacteria that are able to undergo high rates of conjugation to transfer both virulence and antibiotic resistance genes *in vitro*. For our model system, we use S17λ *E. coli* carrying the pSC189-KanR plasmid as the donor and C41-pLysS-CmR cells as the acceptor. As the donor conjugates with the acceptor, S17λ cells will transfer the kanamycin resistance cassette to C41 cells to produce a strain that is resistant to both kanamycin and chloramphenicol. This allowed for double selection during serial dilution and plating to analyze the frequency of conjugation

within a bacterial population. To replicate the three-dimensional environment of the mucosal lining, we allowed for conjugation to occur in liquid media (as opposed to on surfaces of agar plates) over 5 hours. To test how the presence of mucins affected conjugation rates, we supplemented the liquid culture with 0.5% (w/v) of either Muc5ac or Muc2. In the liquid culture, *E. coli* conjugated at an average frequency of 20 events per million cells (Figure 5-1). In the presence of Muc5ac and Muc2, the conjugation frequency was below the detection limit of the assay, showing that the Muc5ac and Muc2 inhibit conjugation by at least an order of magnitude. To determine if the inhibition was due to the presence of polymer-like structures, we also tested conjugation rates in the presence of CMC, a common mucin mimetic. In the presence of CMC, the conjugation rate doubled on average compared to wildtype rates alone (Figure 5-1). Moreover, the changes in the conjugation rates are due to conjugation events and not change in microbial populations, as shown in Figure 5-1. These changes suggest that the physical presence of polymers is unable to explain the inhibitory effects of mucin. Since CMC is unable to prevent the conjugation, perhaps mucins exhibit unique biochemical properties that inhibit the conjugation pathways.

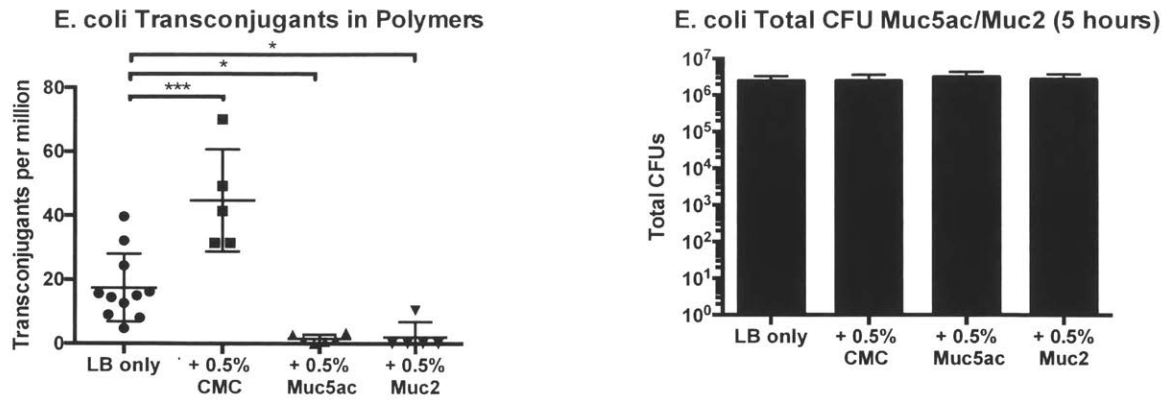


Figure 5-1: *E. coli* conjugation is inhibited in the presence of mucins Muc5ac and Muc2 but upregulated for a synthetic analog CMC. These changes in conjugation are not due to changes in the total population counts (right).

To determine if the inhibitory effects still hold in native mucus samples, we took fresh mucus scrapings from pig stomachs and intestines, sterilized through brief heat treatment, and diluted to 12.5% concentrations in media such that the resultant fluid was easily pipettable for serial dilution. Using the same five-hour conjugation period, we showed that PGM (Figure 5-2) and PIM (Figure 5-3) were able to inhibit *E. coli* to *E. coli* conjugation of antibiotic resistance genes by an order of magnitude without affecting the total population counts. These results revealed that at the concentrations of mucus tested, both pig gastric mucus (containing Muc5ac) and pig intestinal mucus (containing Muc2) were able to inhibit the conjugation events between various strains of *E. coli* without affecting the overall microbial populations.

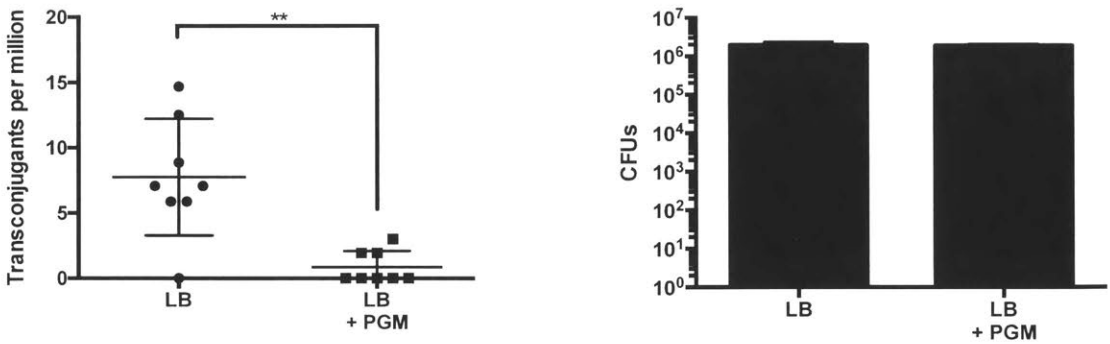


Figure 5-2: *E. coli* conjugation in pig gastric mucus (PGM) is inhibited by an order of magnitude without affecting total population counts.

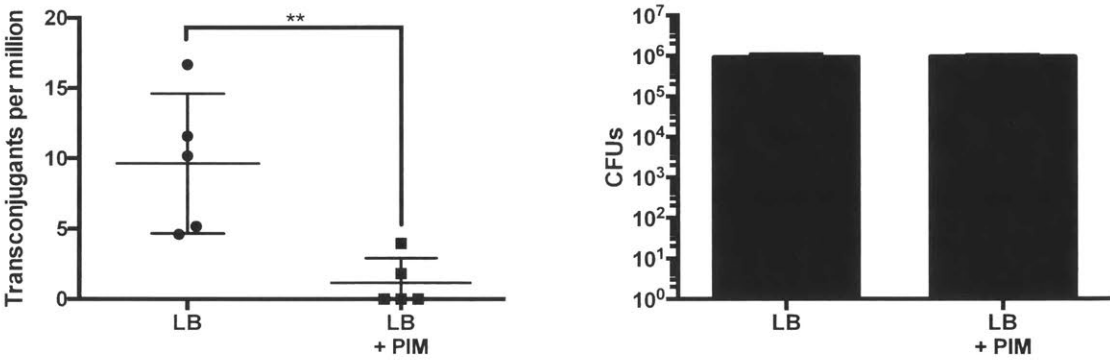


Figure 5-3: *E. coli* conjugation in pig intestinal mucus (PIM) is inhibited by an order of magnitude without affecting total population counts.

Changes in conjugations rates in mucus are dependent on microbial species

Although *E. coli* represents a useful model for understanding how microbes are able to conjugate in the presence of mucin polymers or native mucus, it is unclear how generalizable these results are. As microbes are being explored for probiotics and therapeutic purposes, it is essential that the dynamics of gene transfer in the LB intestine are well understood before such microbes are mass deployed for consumption. One well established probiotic is *Lactobacillus reuteri*, which is a microbe used to manage the immune system and maintain healthy gut¹⁴³⁻¹⁴⁵. Various strains are

used in probiotic formulations and yogurt cultures. Through constant ingestion, *L. reuteri* can maintain a stable population within the gut but does not engraft well long term. Although *L. reuteri* is used so widely commercially, there is limited data on the rates of genetic transfer between other microbial species present in the gut and its role in dissemination of antibiotic resistance genes. Here, we directly examine the rate a donor cell such as *E. coli* S17 λ is able to confer antibiotic resistance to *L. reuteri* within the mucus environment. To test the rate of conjugation between the two microbial species, we use a similar method compared to before. Here, to promote *L. reuteri* growth, we use BHI as the culturing medium and allow for conjugation to occur over 24 hours. As can be seen in Figure 5-4, we find that Muc2 slightly raises the conjugation rates between *E. coli* and *L. reuteri* (not significant). Although the upregulation is minor, the total population of *L. reuteri* is higher in the presence of Muc2, indicating that overall, the absolute number of antibiotic-resistant *L. reuteri* in the culture is higher. Using the same analysis for PIM, we again find that the conjugation frequency increases (Figure 5-5) without significantly changing total population. However, in this case, we find that the increase in conjugation is approximately an order of magnitude higher, indicating that some factors in mucus may be significantly increasing the propensity of *E. coli* to conjugate with *L. reuteri*. Although these tests are implemented in a diluted mucus condition, these assays reveal that mucins and whole mucus do exhibit the ability to both inhibit and promote conjugation between various microbial species.

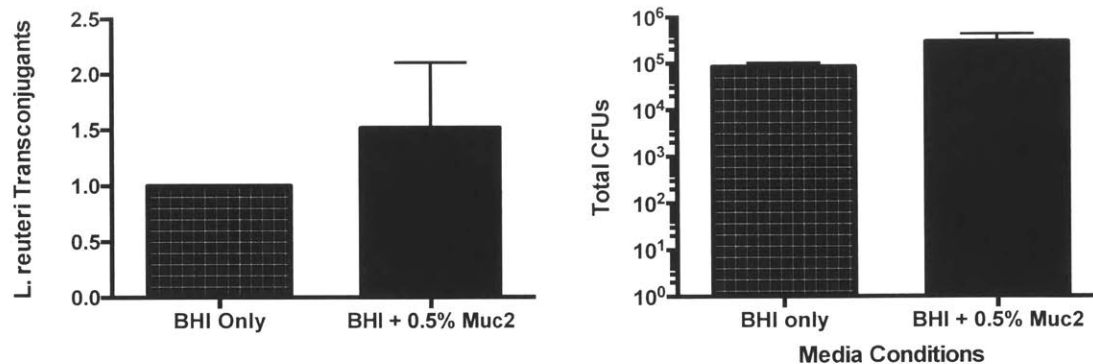


Figure 5-4: *E. coli* conjugation with *L. reuteri* in Muc2 shows a slight increase in conjugation in the presence of Muc2 over the BHI only conditions (left). Muc2 also appears to promote the growth of *L. reuteri* (right) compared to BHI only conditions, suggesting that the absolute conjugation events may be significantly higher in Muc2 conditions.

Discussion

These results show that the mucosal environment has dramatic impacts on rates of genetic exchange between microbial species. Previous research has revealed that the bacteriophages residing in the gut mucosa are a significant source of transduction¹⁴⁶. Here, we reveal that in various mucosal sites, transformation and conjugation mechanisms can be both inhibited and regulated based on the species involved. Within the oral cavity, *S. mutans* transformation is inhibited by the presence of salivary mucin MUC5B and native saliva. In the gastrointestinal tract, *E. coli* to *E. coli* conjugation is inhibited by both Muc5ac and Muc2 as well as by PGM and PIM samples. However, for *E. coli* to *L. reuteri* conjugation, Muc2 and PIM appear to promote conjugation. Although the mechanism is not known, we speculate that since *L. reuteri* displays surface agglutinins that help with anchorage to mucins and other proteins present in mucus^{147,148}, this may provide a stable environment for pili formation and bridging between two motile species of bacteria. Continuing with this logic, it stands to reason that conjugation rates between bacteria in a mucosal setting will be dramatically different for motile and immotile bacteria. Since the mechanisms for how mucus and mucins promote or inhibit conjugation is unknown,

further work is needed to understand multispecies dynamics and introduction of foreign microbes as part of therapies.

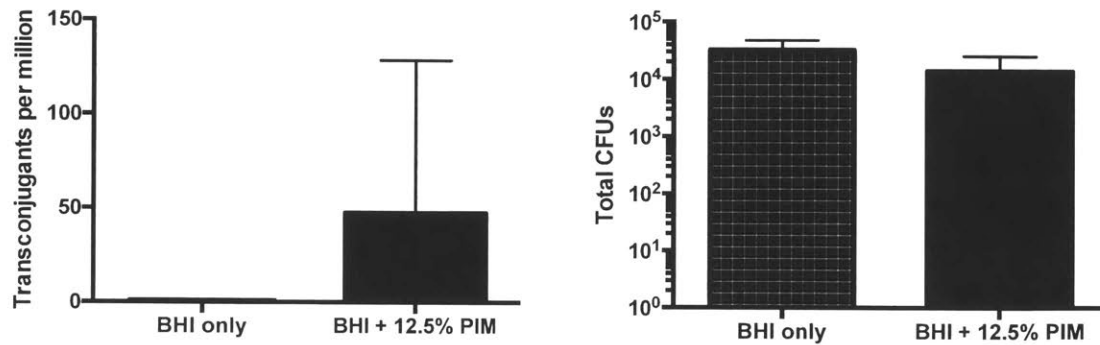


Figure 5-5: *E. coli* conjugation with *L. reuteri* in PIM revealed that mucus increases the rate of conjugation between the two strains of bacteria (left) without significant differences in the overall population of *L. reuteri* (right).

Chapter 6 Structure and function relationship of glycans on mucins

Mucins consist of two major components: the protein backbone and the dense glycosylation of sugars. The protein backbone can be divided into several general domains: the proline-threonine-serine (PTS) domain where glycans are conjugated to and the hydrophobic domains and cysteine knots that promote gelation¹⁵. The complex oligosaccharides added to the PTS domains contribute up to 90% of the molecular weight of mucin and are believed to provide the major functions of mucins^{2,16}. Without the contribution of glycans, naked mucin (apo-mucin) is insoluble in water and no longer has the anti-biofilm forming functions. Moreover, all lubricating and hydrating properties of mucins are no longer detected. Although these experiments reveal that glycan composition as a whole is essential for the overall function of a mucin glycoprotein, it is unclear how each individual glycan contributes to the function. For instance, how much redundancy is there in the glycan function covered by various carbohydrates? What structures of glycans confer the unique anti-virulence properties of mucins?

To understand the structure function relationship of glycan composition, attempts to catalogue the individual glycan types have revealed over 100 structures⁴⁷. As the synthesis of these glycans *in vivo* is highly stochastic, each mucin protein is a unique glycoform. The complex regulation of mucin production has so far made it intractable to understand what the structure-function relationship of each glycan and mucin monomer is. However, significant progress has been made for roles of sugar monomers such as fucose and sialic acid, especially in the context of gut microbiome. For instance, pathogens such as *Salmonella enterica* and *Vibrio cholera*, the presence and release of sialic acids from mucins provides a nutrient a source that gives these pathogens a competitive edge over the normal gut flora^{149,150}. Other microbes such as

Bacteroides thetaiotamicron, one of the primary colonizers of the gut, are pivotal in establishing a normal microbiome¹⁵¹⁻¹⁵³. *B. thetaiotamicron* display a wide repertoire of digestive enzymes that breakdown complex polysaccharides. The release of such sugars allows for utilization by beneficial bacteria such as *B. fragilis*, which helps promote a healthy gut microbiome¹⁵³. For infants especially, the presence of *B. thetaiotamicron* is essential for the establishment of a healthy microbiota to modulate immune response and protect against opportunistic pathogens. The concerted action between mucin glycans and microbial colonizers suggest that in depth investigations between the two components should be conducted.

With over 100 glycans identified, it is essential to understand what the functional capabilities are of the individual glycans. Iterative synthesis of each glycan for comprehensive analysis is technologically challenging, so a directed approach for systematic analysis needs to be implemented. Here, we describe a method that involves identifying what overlapping glycans exist in other parts of the body that may facilitate the same functions as mucin glycans. Preferably these are also secreted glycans or glycoproteins that are present in the gastrointestinal tract and have roles in microbiota and immune modulation. The primary candidate that fits these requirements is the class of sugars called human milk oligosaccharides (HMOs). HMOs are complex glycans found in human breast milk. By definition, these glycans are not metabolized by the infant and instead are processed by various microbes within the gastrointestinal tract. HMOs have garnered significant attention due to their multi-faceted functions. From controlling immune response in the blood stream to acting as decoy binding sites for pathogenic microbes, HMOs are key regulators of infant health and help establish the infant microbiota¹⁵⁴⁻¹⁵⁹.

Table 6-1: Table of the most prevalent HMOs that are similar in structure to identified mucin glycans.

| Glycan name | Milk Glycan Concentrations | Known epitope name |
|-------------------------------------|----------------------------|--------------------|
| Lacto-N-tetraose | 0.5-1.5 g/L | |
| Lacto-N-fucopentaose I (LNFP I) | 1.2-1.7 g/L | H-epitopes |
| Lacto-N-fecopentaose II (LNFP II) | 0.3-1.0 g/L | Le ^x |
| Lacto-N-fucopentaose III (LNFP III) | 0.01-0.2 g/L | Le ^x |
| Lacto-N-difucohexaose I | 0.1-0.2 g/L | Le ^x |
| NeuAc(alpha2-6)lactose | 0.3 - 0.5 g/L | |
| NeuAc(alpha2-3)lactose | 0.1-0.3 g/L | |
| NeuAc-lacto-N-tetraose a | 0.03 - 0.2 g/L | |
| NeuAc-lacto-N-tetraose b | 0.1 - 0.6 g/L | |
| NeuAc2-lacto-N-tetraose | 0.2 - 0.6 g/L | |

Structurally, HMOs also overlap significantly with mucins glycans. HMOs use lactose as the core sugar for creating complex glycans while mucins use GalNac attached to a serine or threonine. The terminal structures of the glycans are highly similar, with branched galactose and GlcNac chains terminated with either sialic acid or Fuc. Although the genetic and enzymatic regulation of HMO production is still not well-defined, the monosaccharides used in the synthesis pathways are the same for mucin glycans. Again, since these processes are stochastic, within a healthy mother’s breast milk, over 100 glycan varieties have also been identified¹⁵⁵. To narrow the scope of the proposed glycans to a manageable number, we rank sort the HMOs by similarity to previously identified mucin glycans and the yield in breast milk. Using this strategy, we find that the nine most commonly found HMOs have striking structural similarity to glycans previously identified on Muc2 (Table 6-1).

As the most common HMOs are available commercially, it is feasible to test how the addition of single types of HMOs are able to induce microbial responses. An assay commonly implemented is biofilm inhibition of various pathogenic bacteria. For HMO analysis, future experiments in *Pseudomonas aeruginosa* biofilm formation is suggested. *Pseudomonas aeruginosa* is a key player in chronic lung infections, particularly in cases for cystic fibrosis patients¹⁶⁰. In cases of chronic infection, *P. aeruginosa* forms robust biofilms that are robust to antibiotic treatment. Moreover, in long-term infections, *P. aeruginosa* is recognized to outcompete other microbes in the same environment through release of toxic molecules and its Type VI secretion systems⁴⁹. Previous results in the context of native mucins have shown that intact glycosylated mucins are able to significantly impair biofilm formation of *P. aeruginosa* as well as disperse established biofilms. Using similar experimental setups, the parallel experiments using HMOs can be implemented to determine how well the individual glycans contribute. By using the identified HMOs, the structure-function relationship between mucin glycan identity and microbial response can be elucidated.

Table 6-2: Estimation of effective concentrations of HMOs in comparisons to glycan composition found on intestinal mucins. For the same protective function, it is estimated that 3-8 mM of HMOs is required for any therapeutic or prophylactic effect whereas only 0.4 – 2 mM of mucin glycans (when linked to polymers) is needed.

| HMOs | | Mucins | |
|--|----------------|--|--------------|
| Average MW of top 9 HMOs | ~850 | Average MW of mucins (+glycans) | ~2 MDa |
| Estimated concentrations of HMOs (w/v) | 2.84 - 6.8 g/L | MW of mucin protein | 500,000 Da |
| Molar concentration | ~3 - 8 mM | Estimated concentrations of mucins (w/v) | 0.5 - 2 g/L |
| | | Putative glycosylation sites | 2000-3000 |
| | | Molar concentration of glycans | ~ 0.4 - 2 mM |

Once the candidate HMOs are identified, one question that can be asked is why does the human body expend the energy to create complex mucin glycoproteins when individual HMOs are sufficient for a protective effect. A major difference between the polymer structure and free floating glycans is the role of avidity. Avidity changes the binding kinetics of molecules by increasing the probability of substrate binding. One way to accomplish this is to locally increase the effective concentration of substrates by providing multiple binding sites (such as in the case of mucin glycans). By using orders of magnitude estimations for the production of HMOs and mucin glycans, we calculate 3.0-8.0 mM of HMOs is required for a protective effect. However, in the mucin context, only 0.4-2.0 mM of glycans is required for the same protective potential (Table 6-2). This suggests that avidity may play a major role in maximizing therapeutic effects of glycans even when overall concentrations are up to an order of magnitude lower. To test if avidity does minimize the concentration of HMOs required, I propose that individual HMOs can be conjugated to a base polymer (Figure 6-1). Moreover, this process will also determine whether or not covalently linked glycans have a different functional role compared to free-floating soluble glycans that can be metabolized by microbes.

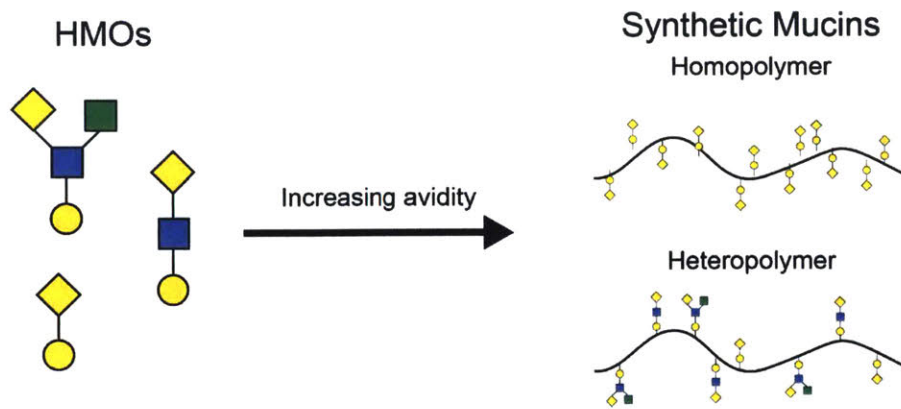


Figure 6-1: For testing the efficacy of HMOs as a protective substance, HMOs can be linked to inert polymer backbones to increase the avidity of interactions. These synthetic mucin polymers can be designed to contain a single HMO (homopolymer) to target single microbial pathways or mixtures of HMOs (heteropolymers) to broaden the functional specificity of the synthetic mucin.

These proposed experiments should link two fields that currently operate independently. A few reviews tangentially comment on the structural similarity between HMOs and mucin glycans, but again, no systematic analysis of the glycans has been implemented. As glycan synthesis technologies continue to advance, it becomes much more feasible for high throughput iterative testing of these glycans to determine their therapeutic and protective potential against invading microbial species. The experiments provided here provide the foundation for screening 10-20 glycans that have a strong potential to regulate microbial function, whether it be biofilm inhibition, quorum sensing inhibition, or tailored growth of beneficial bacteria.

Chapter 7 Synthesis of Mucin Mimetics to Understand Function of Glycosylated Sugars

In the previous chapter, we discussed an approach to understand the role of complex HMO and mucin glycans in protecting against specific microbes such as *P. aeruginosa*. To accomplish these tasks, it is necessary to establish methods to synthesize and test synthetic mucins, which will be discussed in the current chapter. For synthetic mucins, we identified several criteria that need to be satisfied for an ideal polymer base:

- 1) The polymer backbone must be sufficiently long to mimic the protein backbone length of mucins.
- 2) The glycosylation method must be covalent and not susceptible to enzymatic degradation.
- 3) The glycosylation must be uniform and relatively dense.
- 4) The final glycopolymer should exhibit a net negative charge to mimic the charge of mucins.

From the glycan perspective, one possible library of glycans that can be used for synthesis and testing are the HMOs previously identified in Chapter 6. However, there are several issues with this approach. Primarily, HMOs, although available commercially, are prohibitively expensive for routine synthesis testing of prototypes. Moreover, even if individual HMOs are identified for specific protective effects, it is unclear if a subset of the glycan is sufficient or if the entirety of the HMOs is required for correct function.

An approach to circumvent these challenges is based on the assumption that the entire complex glycan is not necessary for the functional properties of the sugars. Since the glycans are densely packed on the protein backbone of mucins^{2,15}, it stands to reason that there are geometric

constraints to what is available for binding by microbial cell surface proteins or free floating enzymes. Because of steric limitations, the argument can be made that only the terminal sugars are freely available for binding. The different structural conformations merely change the kinetics and affinity for binding, but at sufficiently high densities of glycosylation, low affinity effects can be overcome. Using this perspective, we can attach monosaccharides to the polymer backbone to understand how individual sugars are able to change the physiological response of microbes. Here, we screen four monosaccharides glucose, galactose, fucose, and mannose to test for inhibition of *S. mutans* biofilm inhibition. We identify glucose and galactose as the strongest inhibitors of biofilm formation and that this inhibition is independent on polymer backbone length. Moreover, we discover the glucose-containing polymers are able to inhibit additional virulence traits such as natural transformation, a property previously identified to be specific to MUC5B in Chapter 4.

Materials and Methods

Strains and Reagents:

Unless otherwise stated, all reagents are obtained from Sigma Aldrich. The bacterial strain *Streptococcus mutans* UA159 was kindly given as a gift by Dan Smith (Forsyth Institute). For general propagation and storage, UA159 was cultured and maintained in Todd-Hewitt (TH, BD 249240) at 37C with 5% CO₂. Plasmid DNA pVA838 (ATCC 37160) for transformation assays was purchased from ATCC and maintained in *E. coli* DH5 α (BD LB medium, 10 ug/ml chloramphenicol). All TH media was used within seven days of autoclaving.

Synthesis of glycosylamines:

The aldehyde group of reducing sugars is converted to a primary amine via the standard Kochetkov reaction. Briefly, reducing sugars are incubated with ammonia bicarbonate at a 1:40 w/w ratio for four days at room temperature (Kochetkov amination). Liquid containing solubilized sugars is aspirated and collected. Excess ammonia is removed via repeated lyophilization until weight stabilizes. Glycosylamines are used fresh for conjugation to CMC backbones.

Glycosylamine conjugation to CMC:

Prior to conjugation, 90 kDa or 250 kDa CMC is dissolved overnight at 10 mg/ml concentrations. The effective molar concentrations of carboxyl groups are estimated based on the reported degrees of substitution (D.S. = 1.2). HoBt and EDC are added at a 1:1:1 molar ratio to the carboxyl groups and solubilized for up to 30 minutes, stirring. Once fully dissolved, glycosylamines are added at a 2:1 molar ratio to estimated carboxyl groups and stirred for 30 minutes at room temperature. Once reaction is completed, solution is diluted 10-fold with milliQ water to stop reactions. Residual glycosylamines and reactants are removed by dialyzation using a 10 kDa cutoff membrane. Dialyzed glycosylated CMC is lyophilized and stored at room temperature in a desiccation chamber.

Horizontal Gene Transfer Assay:

For all gene transfer assays, unless specified separately, UA159 cultures were grown overnight in 50% TH media at 37C, 5% CO₂ under static conditions. The next morning, cultures were diluted 1:20 into the 25% TH media in 100 uL volumes in 96-well plates and cultured statically for 2 hours. pVA838 (1.2 ug/ml unless otherwise noted) was added and gently mixed into the

solution and allowed to incubate for an additional two hours before serially diluting 1:10 in phosphate buffer saline (PBS). For transformants, dilutions were plated on selective TH agar plates (10 ug/ml erythromycin). For total population counts, serial dilutions were plated on non-selective TH agar plates. All data presented are biological replicates with standard deviations reported for error bars.

Biofilm inhibition assay:

For all biofilm inhibition assays, unless specified separately, UA159 cultures were grown overnight in 50% TH media at 37C, 5% CO₂ under static conditions. The next morning, cultures were diluted 1:20 into the 25% TH media containing 1% (w/v) sterile sucrose in 100 uL volumes in 96-well plates and cultured statically for 4 hours. During the serial dilution process, the planktonic phase was gently pipetted and separated. The attached biofilm was washed three times by 100 ul of sterile PBS. Biofilms were then detached by vigorous scraping with a pipette tip for 30 seconds and mixing prior to serial dilution.

Results

To satisfy the above criteria, the choice of the polymer backbone is essential for the glycopolymer synthesis process. Of the various polymers considered, carboxymethylcellulose (CMC) was designated the ideal choice. The carboxylation provides a net negative charge that is reminiscent of intact mucins, the polymer can be produced at various lengths, has tunable density of glycosylation sites, and functionally, is relatively inert when tested against microbes. Specifically, the carboxyl groups on CMC polymers also provide glycosylation sites through EDC-HObT reactions with aminated substrates. Reducing sugars such as glucose, galactose, fucose, and mannose can have their aldehyde groups converted to a free amine by the well-

known Kochetkov reaction. Moreover, CMC comes in multiple molecular weights, ranging from 30 kDa to >700 kDa depending on the cellulose source. For each molecular weight category, generally the degree of carboxylation ranges from 0.7 to 1.2 per glucose monomer, which allows for tunability in both the hydrophilicity as well as degree of glycosylation. For the complete synthesis procedure, refer to Material and Methods.

For the first library of glycopolymers, we used the following monosaccharides: glucose, galactose, mannose, and fucose (termed Glc-CMC, Gal-CMC, Man-CMC, and Fuc-CMC, respectively, for the glycopolymer notation). Using glucose as the proxy for all testing purposes, we used glucose oxidation assays to detect the extent of labeling. With this absorbance assay, we determine that per milligram of glucose-CMC, the concentration of labeled glucose is about 0.1 mg, which is 10% of the molecular weight (data not shown). These results reveal that not all possible carboxyl groups are glycosylated with the monosaccharide of choice. Assuming that all the glycosylation densities are consistent for each of the tested sugars, we then tested the functionality of the glycopolymers.

To demonstrate the idea that the terminal sugar monomers are sufficient in regulating microbial function, we used biofilm formation of *S. mutans* as a functional readout. With the previously designed assays from Chapter 4, we tested for biofilm inhibition of these glycopolymers compared to the original CMC polymer alone. As seen in Figure 7-1, CMC alone did not have an inhibitory effect whereas each of the glycosylated polymers significantly prevents biofilm formation. Of the glycopolymers tested, Glc-CMC had the strongest inhibitory effects against biofilm formation, followed by Gal-CMC, Fuc-CMC, and finally Man-CMC with the weakest

effect. This assay shows that the glycosylation of monosaccharides to CMC significantly changes the function of the polymer and inhibits biofilm formation of *S. mutans* by at least 4-fold difference.

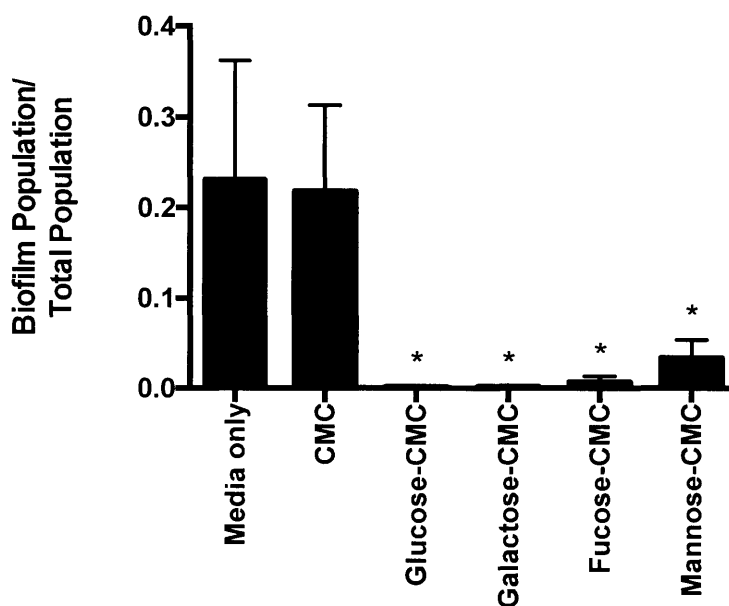


Figure 7-1: Covalent linkage of monosaccharides to CMC produces novel inhibitory functions for CMC. Attachment of glucose and galactose produces the strongest inhibitory effects against biofilm formation of *S. mutans* whereas addition of mannose has the weakest effect, suggesting that the choice of monosaccharide is essential in determining the function of the glycopolymer.

To further identify that the grafted glycans are an essential regulator of function, we compared Glc-CMC to unglycosylated CMC with free-floating glucose in the environment. This test determines if glucose alone is sufficient for the inhibitory response or if glucose must be in its grafted form on the CMC. The results, as seen in Figure 7-2, revealed that the grafted form of glucose on CMC induced a dramatically different physiological response compared to the free-floating form. Glc-CMC inhibited biofilm formation as before but when CMC was physically separate from glucose, the inhibitory effects were not maintained, even when glucose was added at 1% (w/v) concentrations. These results suggest that glycans linked to mucins may exhibit different biological functions in their grafted form compared to their free-floating form in solution. A possible explanation for the difference in function is the ability of microbes to

metabolize monosaccharides in its soluble form. When covalently linked to CMC, glucose is unable to be metabolized by *S. mutans*. This is evidenced by the fact that the addition of Glc-CMC to the media did not enhance growth of *S. mutans*, which suggests that Glc-CMC is not a nutrient source (Figure 7-3). Although Glc-CMC cannot be metabolized by *S. mutans*, glucose may still act as a biochemical signal to alter *S. mutans* physiology. As an example to support this hypothesis that covalently linked forms of glycans and free glycans exhibit disparate functions, there are reported bacteria blooms of opportunistic pathogens upon cleavage of sialic acid and fucose from mucin glycans^{149,150}.

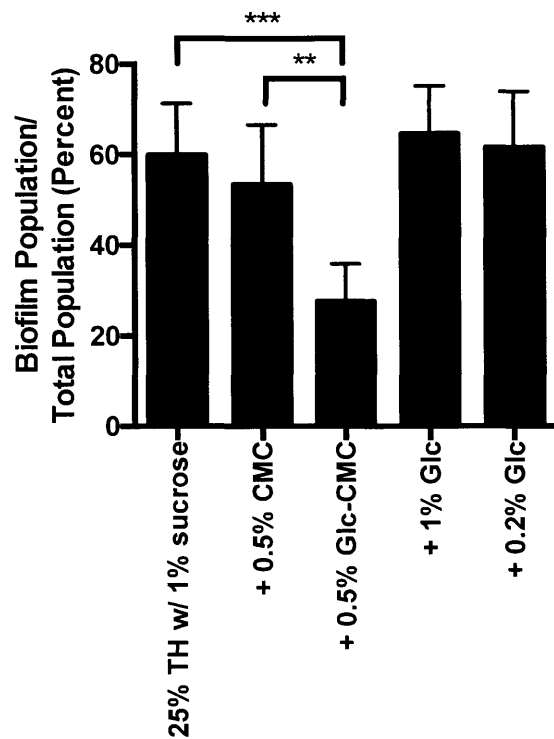


Figure 7-2: Inhibitory effects of Glc-CMC rely on the covalently linked glucose to CMC. When glucose is added to unaltered CMC, no inhibitory effects are observed even when glucose is added to 1% (w/v). These results suggest that the covalently linked form of glucose induces different physiological responses in *S. mutans* compared to unconjugated glucose.

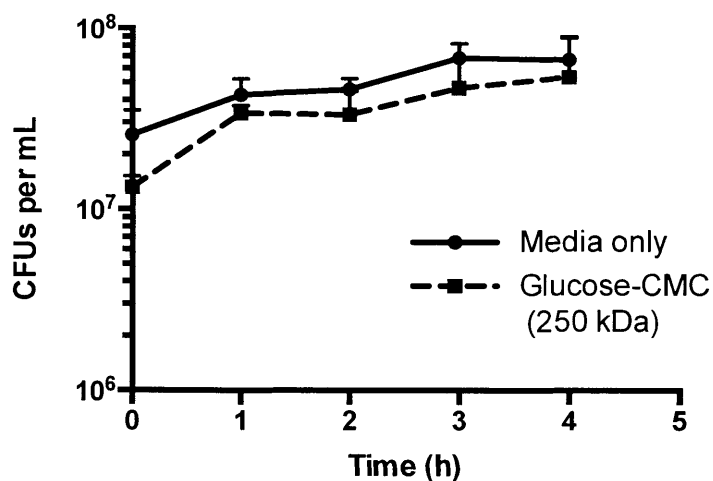


Figure 7-3: Growth rate of *S. mutans* in media only conditions and media supplemented with Glc-CMC are not significantly different, suggesting that Glc-CMC is not acting as an additional nutrient source for *S. mutans*.

The last parameter investigated in these studies is to understand if polymer length also affects the function of the resultant glycopolymer. Mucin polymers are typically thousands of amino acids long with many repeating subsequences where glycosylation occurs. A major question in the field is to understand what the relevance of having such long polymers is. One suggested reasoning is that the long polymer formation allows for structural stability of the mucus hydrogel by allowing for processes such as entanglement to occur¹⁵. To investigate the role of polymer length on function, we chose two polymer lengths: 90 kDa CMC and 250 kDa CMC. Using the same process as before and same molar ratios for glycan grafting, glucose was glycosylated onto the CMC backbone. In comparisons for testing the biofilm inhibitory properties of CMC, it was found that both 90 kDa Glc-CMC and 250 kDa Glc-CMC have similar inhibitory effects (Figure 7-2 and Figure 7-4). From a functional perspective, these two polymers are not significantly different. However, the two different lengths of CMC are known to have varying degrees of viscosity at the same weight percent formulations. These results suggest that the functional component of glycan-grafted CMC can be independent from the mechanical properties. With

these two independent design parameters, materials can be developed that have the same biological effects, but dramatically different viscoelastic profiles.

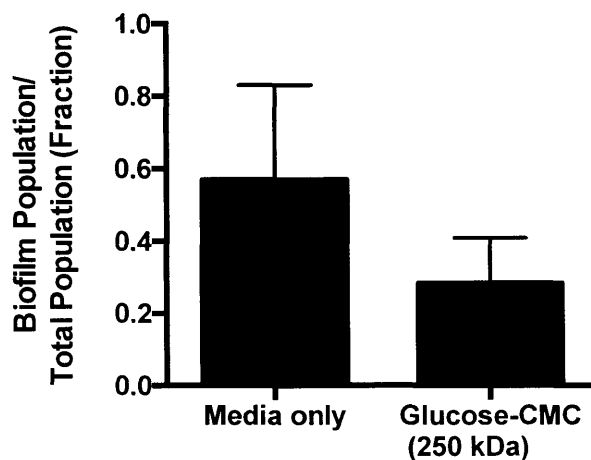


Figure 7-4: Longer polymer lengths (250 kDa) of CMC have the same inhibitory effects as 70 kDa CMC (Figure 7-2) against *S. mutans* biofilm formation when glycosylated with glucose.

Discussion

With these particular polymers, there are several future areas that can be explored. An immediate application is that these polymers can be developed as prebiotics or therapeutics for the prevention and treatment of chronic infections. We have *in vitro* proof of concept results for inhibiting plaque formation of *S. mutans*, but these polymers can be applied for other opportunistic biofilm-formers such as *P. aeruginosa*, *S. aureus*, and *C. albicans*. Moreover, in assaying how *S. mutans* responds to the presence of these glycosylated polymers, we found that not only is biofilm formation altered, other virulence related pathways such as competence in *S. mutans* is strongly inhibited (Figure 7-5). Altogether, these results suggest that the glycosylation of simple monosaccharides to a polymer backbone can impact multiple metabolic pathways of microbes and act as a broad-spectrum inhibitor of virulence traits without affecting cell populations.

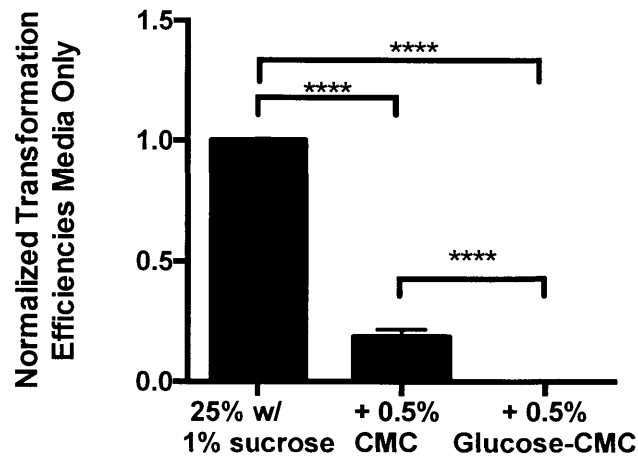


Figure 7-5: Inhibition of *S. mutans* natural transformation in the presence of Glc-CMC is significantly stronger compared to CMC alone, suggesting that the covalently bound glucose to CMC provides new biological function for the glycopolymer.

Alternatively, these synthetic glycosylated polymers allow for systematic analysis of which grafted sugars are functionally relevant. These polymers represent a simple toolkit in which biochemical pathways can be investigated to understand which microbial cell surface receptors are activated or inhibited by mucin-like glycans. A possible master regulator is the phosphotransferase system (PTS), which is specific for the translocation of sugars across bacterial membranes¹⁶¹. These two-component systems are essential for not only metabolizing nutrients, but are also implicated in the control and expression of virulence systems in bacteria¹⁶²⁻¹⁶⁴. The PTS system revolves around using phosphoenolpyruvate (PEP) as an energy source. First discovered in *E. coli*, the PTS system has been identified in almost all bacterial species to date. In general, the PTS consists of two components: the intracellular Enzyme I (EI) and HPr, which are the main drivers in phosphorylation and downstream signaling, and Enzyme II (EII) as the cell surface transporter that is responsible for carbon source recognition¹⁶³. In *E. coli*, at least fifteen EII enzymes have been identified and are responsible for a wide range of carbohydrates¹⁶³. The EII cell surface transporters and sensors can be categorized into four major families depending on their specificities to saccharides. Phosphorylation of EIIA subunits dictates

what carbohydrates the bacteria preferentially react to and govern the overall genetic expression profile and physiology of the bacteria. The phosphorylation pathways of carbohydrate metabolism follow a tiered approach, with glucose generally being the preferred catabolite^{161,165}. As such, we expect that the response of opportunistic pathogens is highly dependent on the type of saccharides exposed on the polymer backbone. For instance, one hypothesis is that the glucose-conjugated CMC polymers may induce responses from the broadest range of bacterial species due to glucose being a major catabolite-repressor.

As carbohydrate sensing dictates the metabolic response of bacteria, the presence and absence of sugars determines the virulence profile of pathogenic bacteria¹⁶³. In several bacterial species, expression of virulence genes is controlled by carbohydrate catabolite repression. For *Clostridium difficile*, which secretes colitis- and diarrhea-inducing exotoxins, the presence of monosaccharides inhibits expression of *toxA* and *toxB* genes¹⁶³. In the cavity causing bacteria, *Streptococcus mutans*, mutants without the EIIABC transporter exhibit higher levels of the virulence gene fructose hydrolase¹⁶⁶. These observations suggest that by providing bacteria with the correct carbohydrate signals, the expression of virulence genes may be modulated or inhibited. Taken together, our approach of using simplified glycopolymers to test the structure-function relationship of monosaccharides is a powerful tool to understanding how mucin glycans are able to provide the broad spectrum protective effects. This work provides the basis for systematic analysis of both simple glycans as well as complex oligosaccharides such as HMOs (Chapter 6) for understanding how grafted glycans affect microbial physiology and metabolism.

Chapter 8 Conclusion and Future Directions

This thesis details novel methods to understand the heuristic design principles for self-assembling polymers and how they relate to functional properties of biological hydrogels. In Chapters 2 and 3, I developed a peptide-based approach to analyze two fundamental questions in nuclear pore biology. In Chapter 2, I determined whether the presence of charge affects the molecular selectivity of essential hydrophobic FG domains and how localization of charge affects selective function. In Chapter 3, I analyzed the preferential bias for phenylalanines in the hydrophobic domains of nucleoporin sequences and determine how substitution with other natural aromatic amino acids such as tryptophan and tyrosine affect nuclear transport function. In both cases, I reduced the complexity of native nucleoporin sequences from >1000 amino acids to 14-amino acid peptides. Although I identified how charge, spatial localization of charge, and choice of aromatic residues all affect the selective properties of FG-based sequences, there are several limitations with this approach. Primarily, these short peptide-based gels do not recapitulate the selective properties of intact nucleoporin gels^{5,24,95}. A major reason for this discrepancy is that the short peptides formed into fibers and rods that are not typically seen in long polymer self-assemblies. To circumvent these challenges, a possible approach is to utilize the 14-amino acid sequences identified in Chapter 2 and create longer peptides that contain multiple repeats of those sequences. Since the original Nsp1 sequence contains approximately 16 repeats, we can create a library of FGAK peptide variants containing the original single repeat, 2 repeats, 4 repeats, 8 repeats, and 16 repeats. As the repeat numbers increase, we expect that polymer entanglement will play a more important role in self-assembly. With these structural rearrangements, the longer polymers will allow for saturation of gel formation and exhibit hydrophobic exclusion typically seen in native nucleoporin gels.

In Chapters 4 and 5, I identified two new functions of mucins and correlated their effect to native mucus samples. In Chapter 4, I described how salivary mucin MUC5B inhibits natural transformation of *S. mutans* and showed this effect translates to native saliva. In Chapter 5, I showed how Muc5ac and Muc2 can be potent modulators of conjugation between commensal bacteria. Between two *E. coli* strains, both Muc5ac and Muc2 inhibited HGT and prevented the spread of antibiotic resistance genes. Moreover, these results were translatable to native porcine gastric mucus as well as porcine intestinal mucin, suggesting that Muc5ac and Muc2 are critical in the inhibitory response. However, for other probiotics such as *L. reuteri*, Muc2 and intestinal mucus appeared to upregulate conjugation rates. The discrepancies between the various species suggest that there are strain-specific effects that need to be elucidated. For instance, *L. reuteri* is canonically considered a non-motile bacteria and also displays surface agglutinins that allow for binding to mucins and self-aggregation¹⁴⁸. *E. coli*, on the other hand, is highly motile in mucins and mucus and is not known to aggregate using the same mechanisms. The disparities in the phenotypic response between *L. reuteri* and *E. coli* suggest that the ability of the bacteria to stably colonize in a mucosal environment determines the rate at which the species can conjugate. Future experiments that investigate the mechanisms of horizontal gene transfer in the context of mucins is required. For instance, for *S. mutans*, it is not clear whether mucin glycans are the causative agent, and if so, which bacterial pathways are modulated by the glycans. For conjugation between *E. coli* and other gut microbes, it is essential to understand if motility is a major factor in altering the conjugation rates in a mucosal environment, or cell-surface adhesins to mucin polymers is the dominant effect.

In Chapters 6 and 7, I developed strategies to simplify complex mucin glycoproteins down to glycopolymers that contain a single type of glycan as opposed to the hundreds typically identified on native mucins. In Chapter 6, I described an approach inspired by HMO research, a field that historically has little crosstalk with mucin analysis. By analyzing the overlapping glycan structures, I identified approximately 10-20 glycans that overlap structurally between HMO and mucin glycans. These overlapping glycans suggest that certain complex saccharides have conserved protective functions and should be investigated further. In Chapter 7, I developed methods to generate glycopolymers containing single monosaccharides. Based on assumptions regarding steric hindrance and monosaccharide identities in mucin glycans, I created a library of glycopolymers that have strong inhibitory effects against biofilm formation of cavity-causing *S. mutans* bacteria. I showed that the inhibition is independent on polymer length and is dependent on sugar identity and the covalent linkage to the polymer. Although these glycopolymers showed that grafting of glycans onto the polymer backbone are essential for controlling virulence of *S. mutans*, there are several limitations to this approach. Primarily, quality control of the synthesis process is challenging. CMC is a polymer derived from plant-based material and is highly dispersed in polymer length and carboxylation uniformity. Since the starting material is heterogeneous, controlling glycosylation density and purity remains a challenge. For future work, identification of a base polymer that has a low polydispersity index and well-characterized chemistries that allow for uniform glycosylation is necessary. Once the library of second-generation glycopolymers is synthesized, systematic analysis of microbial response using RNA-seq should provide directions as to how covalently bound sugars are able to induce microbial responses. As stated in Chapter 7, a likely universal pathway that mucins are activating is the PTS pathways. By binding to the EIIA cell surface receptors, mucin and HMO glycans may be

upregulating the catabolic repression pathways to inhibit bacterial virulence traits. However, PTS pathways are only present on bacteria. Eukaryotic microbes such as *C. albicans* also exhibit lower virulence profiles in mucins, suggesting that mucins may be activating multiple pathways that affect both prokaryotic and eukaryotic organisms.

Altogether, I lay the groundwork for establishing what parameters are essential in governing the function of polymers in biological hydrogels. The focus here has been on polymers found in NPCs and mucus, but this approach can be generalized to other biological hydrogels. Another biological hydrogel that contains proteins with multiple repeating amino acid sequences is the cuticle of mussel byssal threads. The proteins within cuticles contain subunits enriched in tyrosines that can be chemically modified to form covalent linkages. These linkages help byssal thread resist large cyclic strains and adhere to wet surfaces in the ocean¹⁶⁷⁻¹⁷⁰. Although the repeat peptides have been identified, it's unclear exactly how the exact sequence affects the function of the resultant hydrogel. Other hydrogels that can be examined in such a way are the polysaccharide-based biofilm gels. Tuning polymer length, charge distribution, and compositions of the polysaccharides should provide heuristic principles for understanding how each polymer contributes to mechanical stability and selective protection against environmental stressors.

References

1. Fahy, J. V. & Dickey, B. F. Airway Mucus Function and Dysfunction. *N. Engl. J. Med.* **363**, 2233–2247 (2010).
2. Rose, M. C. & Voynow, J. A. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol. Rev.* **86**, 245–278 (2006).
3. Hall-Stoodley, L., Costerton, J. W. & Stoodley, P. Bacterial biofilms: from the Natural environment to infectious diseases. *Nat. Rev. Microbiol.* **2**, 95–108 (2004).
4. Raices, M. & D'Angelo, M. A. Nuclear pore complex composition: a new regulator of tissue-specific and developmental functions. *Nature Reviews Molecular Cell Biology* **13**, 687–699 (2012).
5. Frey, S. & Görlich, D. A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *CELL* **130**, 512–523 (2007).
6. Ribbeck, K. & Görlich, D. The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. *EMBO J* **21**, 2664–2671 (2002).
7. Newell, N. *et al.* Biomechanics of the human intervertebral disc: A review of testing techniques and results. *J Mech Behav Biomed Mater* **69**, 420–434 (2017).
8. Laasanen, M. S. *et al.* Biomechanical properties of knee articular cartilage. *Biorheology* **40**, 133–140 (2003).
9. Tabeian, H. *et al.* Cyclic Tensile Strain Reduces TNF- α Induced Expression of MMP-13 by Condylar Temporomandibular Joint Cells. *J. Cell. Physiol.* **232**, 1287–1294 (2017).
10. Greene, G. W. *et al.* Lubricin: A versatile, biological anti-adhesive with properties comparable to polyethylene glycol. *Biomaterials* **53**, 127–136 (2015).
11. Waller, K. A. *et al.* Role of lubricin and boundary lubrication in the prevention of chondrocyte apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* – (2013). doi:10.1073/pnas.1219289110
12. Chang, D. P. *et al.* Friction Force Microscopy of Lubricin and Hyaluronic Acid between Hydrophobic and Hydrophilic Surfaces. *Soft Matter* **5**, 3438–3445 (2009).
13. Jay, G. D. & Waller, K. A. The biology of Lubricin: Near frictionless joint motion. *Matrix Biology* (2014). doi:10.1016/j.matbio.2014.08.008
14. Ali, L., Jin, C. & Karlsson, N. G. Glycoproteomics of Lubricin-Implication of Important Biological Glyco-and Peptide-Epitopes in Synovial Fluid.
15. Bansil, R. & Turner, B. S. Mucin structure, aggregation, physiological functions and biomedical applications. *Current Opinion in Colloid & Interface Science* **11**, 164–170 (2006).
16. Frenkel, E. S. & Ribbeck, K. Salivary mucins in host defense and disease prevention. *J Oral Microbiol* **7**, 29759 (2015).
17. Hentzer, M. *et al.* Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *Journal of Bacteriology* **183**, 5395–5401 (2001).
18. Mann, E. E. & Wozniak, D. J. *Pseudomonas* biofilm matrix composition and niche biology. *FEMS Microbiol. Rev.* **36**, 893–916 (2012).
19. Patel, S. S., Belmont, B. J., Sante, J. M. & Rexach, M. F. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *CELL* **129**, 83–96 (2007).
20. Denning, D. P. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proceedings of the National Academy of Sciences*

- 100**, 2450–2455 (2003).
21. Stewart, M. Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology* **8**, 195–208 (2007).
 22. Stewart, M. *et al.* Molecular mechanism of translocation through nuclear pore complexes during nuclear protein import. *FEBS Lett.* **498**, 145–149 (2001).
 23. Görlich, D. & Kutay, U. Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* **15**, 607–660 (1999).
 24. Frey, S. & Görlich, D. FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J* **28**, 2554–2567 (2009).
 25. Frey, S., Richter, R. P. & Görlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006).
 26. Bayliss, R., Littlewood, T., Strawn, L. A., Wenthe, S. R. & Stewart, M. GLFG and FxFG nucleoporins bind to overlapping sites on importin-beta. *J. Biol. Chem.* **277**, 50597–50606 (2002).
 27. Bayliss, R. *et al.* Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J. Mol. Biol.* **293**, 579–593 (1999).
 28. Bayliss, R., Littlewood, T. & Stewart, M. Structural Basis for the Interaction between FxFG Nucleoporin Repeats and Importin- β in Nuclear Trafficking. *CELL* **102**, 99–108 (2000).
 29. Bayliss, R. Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *EMBO J.* **21**, 2843–2853 (2002).
 30. Ribbeck, K., Kutay, U., Paraskeva, E. & Görlich, D. The translocation of transport-cargo complexes through nuclear pores is independent of both Ran and energy. *Current biology* (1999).
 31. Bednenko, J., Cingolani, G. & Gerace, L. Importin beta contains a COOH-terminal nucleoporin binding region important for nuclear transport. *J Cell Biol* **162**, 391–401 (2003).
 32. Jakel, S. Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.* **17**, 4491–4502 (1998).
 33. Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M. & Görlich, D. NTF2 mediates nuclear import of Ran. *EMBO J* **17**, 6587–6598 (1998).
 34. Siebrasse, J. P. Rapid translocation of NTF2 through the nuclear pore of isolated nuclei and nuclear envelopes. *EMBO Rep.* **3**, 887–892 (2002).
 35. Clarkson, W. D., Kent, H. M. & Stewart, M. Separate binding sites on nuclear transport factor 2 (NTF2) for GDP-Ran and the phenylalanine-rich repeat regions of nucleoporins p62 and Nsp1p. *J. Mol. Biol.* **263**, 517–524 (1996).
 36. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S. & Wenthe, S. R. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol.* **6**, 10–206 (2004).
 37. Hansson, G. C. Role of mucus layers in gut infection and inflammation. *Current Opinion in Microbiology* **15**, 57–62 (2012).
 38. Roy, M. G. *et al.* Muc5b is required for airway defence. *Nature* (2013). doi:10.1038/nature12807
 39. Lieleg, O., Lieleg, C., Bloom, J., Buck, C. B. & Ribbeck, K. Mucin biopolymers as broad-spectrum antiviral agents. *Biomacromolecules* **13**, 1724–1732 (2012).

40. Cone, R. A. Barrier properties of mucus. *Adv. Drug Deliv. Rev.* **61**, 75–85 (2009).
41. Lai, S. K., Wang, Y.-Y., Wirtz, D. & Hanes, J. Micro- and macrorheology of mucus. *Adv. Drug Deliv. Rev.* **61**, 86–100 (2009).
42. Chaudhury, N. M. A., Proctor, G. B., Karlsson, N. G., Carpenter, G. H. & Flowers, S. A. Reduced MUC7 mucin sialylation and altered saliva rheology in Sjogren's syndrome associated oral dryness. *Mol. Cell Proteomics* (2015). doi:10.1074/mcp.M115.052993
43. Sheehan, J. K., Howard, M., Richardson, P. S., Longwill, T. & Thornton, D. J. Physical characterization of a low-charge glycoform of the MUC5B mucin comprising the gel-phase of an asthmatic respiratory mucous plug. *Biochem. J.* **338 (Pt 2)**, 507–513 (1999).
44. Thornton, D. J. & Sheehan, J. K. From mucins to mucus: toward a more coherent understanding of this essential barrier. *Proc Am Thorac Soc* **1**, 54–61 (2004).
45. Larsson, J. M. H. N., Karlsson, H., vall, H. S. & Hansson, G. C. A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* **19**, 756–766 (2009).
46. Varki, A. *Essentials of Glycobiology*. (CSHL Press, 1999).
47. Thomsson, K. A. *et al.* Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* **22**, 1128–1139 (2012).
48. Fung, C. *et al.* Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. *J. Med. Microbiol.* **59**, 1089–1100 (2010).
49. Jeffries, J. L. *et al.* *Pseudomonas aeruginosa* pyocyanin modulates mucin glycosylation with sialyl-Lewis(x) to increase binding to airway epithelial cells. *Mucosal Immunol* (2015). doi:10.1038/mi.2015.119
50. Montecucco, C. & Rappuoli, R. Living dangerously: how *Helicobacter pylori* survives in the human stomach. *Nature Reviews Molecular Cell Biology* **2**, 457–466 (2001).
51. Juge, N. Microbial adhesins to gastrointestinal mucus. *Trends in Microbiology* **20**, 30–39 (2012).
52. Corfield, A. P. *et al.* Mucins and mucosal protection in the gastrointestinal tract: new prospects for mucins in the pathology of gastrointestinal disease. *gut.bmj.com*
53. Bartman, A. E. *et al.* Aberrant expression of MUC5AC and MUC6 gastric mucin genes in colorectal polyps. *Int. J. Cancer* **80**, 210–218 (1999).
54. Caldara, M. *et al.* Mucin biopolymers prevent bacterial aggregation by retaining cells in the free-swimming state. *Curr. Biol.* **22**, 2325–2330 (2012).
55. Kavanaugh, N. L., Zhang, A. Q., Nobile, C. J., Johnson, A. D. & Ribbeck, K. Mucins Suppress Virulence Traits of *Candida albicans*. *mbio.asm.org*
56. Frenkel, E. S. & Ribbeck, K. Salivary mucins promote the coexistence of competing oral bacterial species. *ISME J* **11**, 1286–1290 (2017).
57. Kazimierzak, K. A. & Scott, K. P. Antibiotics and resistance genes: influencing the microbial ecosystem in the gut. *Adv. Appl. Microbiol.* **62**, 269–292 (2007).
58. Blake, D. P., Hillman, K., Fenlon, D. R. & Low, J. C. Transfer of antibiotic resistance between commensal and pathogenic members of the Enterobacteriaceae under ileal conditions. *J Appl Microbiol* **95**, 428–436 (2003).
59. Wolska, K. I. Horizontal DNA transfer between bacteria in the environment. *Acta Microbiologica Polonica* (2003).
60. Mallon, C. A., Elsas, J. D. V. & Salles, J. F. Microbial Invasions: The Process, Patterns,

- and Mechanisms. *Trends in Microbiology* **23**, 719–729 (2015).
61. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation : Article : Nature. *Nature* **405**, 299–304 (2000).
 62. Gandra, S., Barter, D. M. & Laxminarayan, R. Economic burden of antibiotic resistance: how much do we really know? *Clin. Microbiol. Infect.* **20**, 973–980 (2014).
 63. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
 64. Stecher, B. *et al.* Gut inflammation can boost horizontal gene transfer between pathogenic and commensal Enterobacteriaceae. *Proceedings of the ...* (2012).
 65. van Reenen, C. A. & Dicks, L. Horizontal gene transfer amongst probiotic lactic acid bacteria and other intestinal microbiota: what are the possibilities? A review - Springer. *Archives of microbiology* (2011).
 66. Huddleston, J. R. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect Drug Resist* **7**, 167–176 (2014).
 67. Ando, D. *et al.* Nuclear pore complex protein sequences determine overall copolymer brush structure and function. *Biophys. J.* **106**, 1997–2007 (2014).
 68. Ribbeck, K. & Görlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J* **20**, 1320–1330 (2001).
 69. Ghavami, A., Veenhoff, L. M., van der Giessen, E. & Onck, P. R. Probing the Disordered Domain of the Nuclear Pore Complex through Coarse-Grained Molecular Dynamics Simulations. *Biophysj* **107**, 1393–1402 (2014).
 70. Gamini, R., Han, W., Stone, J. E. & Schulten, K. Assembly of Nsp1 nucleoporins provides insight into nuclear pore complex gating. *PLoS Comput. Biol.* **10**, e1003488 (2014).
 71. Ma, C. D., Wang, C., Acevedo-Vélez, C., Gellman, S. H. & Abbott, N. L. Modulation of hydrophobic interactions by proximally immobilized ions. *Nature* **517**, 347–350 (2015).
 72. Strawn, L. A., Shen, T. & Wentz, S. R. The GLFG regions of Nup116p and Nup100p serve as binding sites for both Kap95p and Mex67p at the nuclear pore complex. *J. Biol. Chem.* **276**, 6445–6452 (2001).
 73. Grant, R. P., Neuhaus, D. & Stewart, M. Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1 Å resolution. *J. Mol. Biol.* **326**, 849–858 (2003).
 74. Colwell, L. J., Brenner, M. P. & Ribbeck, K. Charge as a selection criterion for translocation through the nuclear pore complex. *PLoS Comput. Biol.* **6**, e1000747 (2010).
 75. Tagliazucchi, M., Peleg, O., Kröger, M., Rabin, Y. & Szleifer, I. Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proceedings of the National Academy of Sciences* **110**, 3363–3368 (2013).
 76. Ando, D., Colvin, M., Rexach, M. & Gopinathan, A. Physical Motif Clustering within Intrinsically Disordered Nucleoporin Sequences Reveals Universal Functional Features. *PLoS ONE* **8**, e73831 (2013).
 77. Yamada, J. *et al.* A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell Proteomics* **9**, 2205–2224 (2010).
 78. Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nature*

- Biotechnology* **21**, 1171–1178 (2003).
79. Hinman, M. B., Jones, J. A. & Lewis, R. V. Synthetic spider silk: a modular fiber. *Trends Biotechnol.* **18**, 374–379 (2000).
 80. Xu, M. & Lewis, R. V. Structure of a protein superfiber: spider dragline silk. *Proceedings of the National Academy of Sciences* **87**, 7120–7124 (1990).
 81. Meyer, D. E. & Chilkoti, A. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules* **5**, 846–851 (2004).
 82. Nettles, D. L., Chilkoti, A. & Setton, L. A. Applications of elastin-like polypeptides in tissue engineering. *Adv. Drug Deliv. Rev.* **62**, 1479–1485 (2010).
 83. Wright, E. R. & Conticello, V. P. Self-assembly of block copolymers derived from elastin-mimetic polypeptide sequences. *Adv. Drug Deliv. Rev.* **54**, 1057–1073 (2002).
 84. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
 85. Hurt, E. C. A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. *EMBO J* **7**, 4323–4334 (1988).
 86. Denning, D. P., Uversky, V., Patel, S. S. & Fink, A. L. The *Saccharomyces cerevisiae* nucleoporin Nup2p is a natively unfolded protein. ... *of Biological Chemistry* (2002).
 87. Rathore, O. & Sogah, D. Y. Self-assembly of beta-sheets into nanostructures by poly(alanine) segments incorporated in multiblock copolymers inspired by spider silk. *J. Am. Chem. Soc.* **123**, 5231–5239 (2001).
 88. Chen, S. *et al.* Ionic interactions. Subnanoscale hydrophobic modulation of salt bridges in aqueous media. *Science* **348**, 555–559 (2015).
 89. Migliori, A. D., Smith, D. E. & Arya, G. Molecular interactions and residues involved in force generation in the T4 viral DNA packaging motor. *J. Mol. Biol.* **426**, 4002–4017 (2014).
 90. Migliori, A. D. *et al.* Evidence for an electrostatic mechanism of force generation by the bacteriophage T4 DNA packaging motor. *Nat Commun* **5**, 4173 (2014).
 91. Hülsmann, B. B., Labokha, A. A. & Görlich, D. The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *CELL* **150**, 738–751 (2012).
 92. Ader, C. *et al.* Amyloid-like interactions within nucleoporin FG hydrogels. *Proceedings of the National Academy of Sciences* **107**, 6281–6285 (2010).
 93. Lowe, A. R. *et al.* Importin- β modulates the permeability of the nuclear pore complex in a Ran-dependent manner. *Elife* **4**, (2015).
 94. Friedman, A. K. & Baker, L. A. Synthetic hydrogel mimics of the nuclear pore complex display selectivity dependent on FG-repeat concentration and electrostatics. - PubMed - NCBI. *Soft Matter* (2016).
 95. Labokha, A. A. *et al.* Systematic analysis of barrier-forming FG hydrogels from *Xenopus* nuclear pore complexes. *EMBO J.* **32**, 204–218 (2013).
 96. Eisele, N. B., Labokha, A. A., Frey, S., Görlich, D. & Richter, R. P. Cohesiveness tunes assembly and morphology of FG nucleoporin domain meshworks - Implications for nuclear pore permeability. *Biophys. J.* **105**, 1860–1870 (2013).
 97. Eisele, N. B., Frey, S., Piehler, J., Görlich, D. & Richter, R. P. Ultrathin nucleoporin phenylalanine-glycine repeat films and their interaction with nuclear transport receptors. *EMBO Rep.* **11**, 366–372 (2010).

98. Zahn, R. *et al.* A physical model describing the interaction of nuclear transport receptors with FG nucleoporin domain assemblies. *Elife* **5**, e14119 (2016).
99. Holten-Andersen, N., Zhao, H. & Waite, J. H. Stiff coatings on compliant biofibers: the cuticle of *Mytilus californianus* byssal threads. *Biochemistry* **48**, 2752–2759 (2009).
100. Reichelt, R. *et al.* Correlation between structure and mass distribution of the nuclear pore complex and of distinct pore complex components. *The Journal of cell ...* (1990).
101. Bestembayeva, A., Kramer, A. & Labokha, A. A. Nanoscale stiffness topography reveals structure and mechanics of the transport barrier in intact nuclear pore complexes. *Nature* (2014).
102. Ma, J., Kelich, J. M. & Yang, W. SPEED Microscopy and Its Application in Nucleocytoplasmic Transport. *Methods Mol. Biol.* **1411**, 503–518 (2016).
103. Yang, W. Distinct, but not completely separate spatial transport routes in the nuclear pore complex. *Nucleus* **4**, 166–175 (2013).
104. Nehrbass, U. *et al.* NSP1: A yeast nuclear envelope protein localized at the nuclear pores exerts its essential function by its carboxy-terminal domain. *CELL* **61**, 979–989 (1990).
105. Togashi, D. M., Szczupak, B., Ryder, A. G., Calvet, A. & O'Loughlin, M. Investigating tryptophan quenching of fluorescein fluorescence under protolytic equilibrium. *J Phys Chem A* **113**, 2757–2767 (2009).
106. Marmé, N., Knemeyer, J.-P., Sauer, M. & Wolfrum, J. Inter- and intramolecular fluorescence quenching of organic dyes by tryptophan. *Bioconjug. Chem.* **14**, 1133–1139 (2003).
107. Derrien, M. & van Passel, M. Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut* (2010).
108. Dewhirst, F. E. *et al.* The Human Oral Microbiome. *Journal of Bacteriology* **192**, 5002–5017 (2010).
109. Adler, C. J. *et al.* Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat. Genet.* **45**, 450–5–455e1 (2013).
110. Wade, W. G. The oral microbiome in health and disease. *Pharmacological research* (2013).
111. Tabak, L. A. In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annu. Rev. Physiol.* **57**, 547–564 (1995).
112. Levine, M. J. *et al.* Structural aspects of salivary glycoproteins. *J. Dent. Res.* **66**, 436–441 (1987).
113. Frenkel, E. S. & Ribbeck, K. Salivary mucins protect surfaces from colonization by cariogenic bacteria. *Applied and Environmental Microbiology* **81**, 332–338 (2015).
114. Ahn, S.-J., Ahn, S.-J., Wen, Z. T., Brady, L. J. & Burne, R. A. Characteristics of biofilm formation by *Streptococcus mutans* in the presence of saliva. *Infect. Immun.* **76**, 4259–4268 (2008).
115. Ahn, S. J., Wen, Z. T. & Burne, R. A. Multilevel Control of Competence Development and Stress Tolerance in *Streptococcus mutans* UA159. *Infect. Immun.* **74**, 1631–1642 (2006).
116. Duarte, S. *et al.* Influences of starch and sucrose on *Streptococcus mutans* biofilms. *Oral Microbiol. Immunol.* **23**, 206–212 (2008).
117. Li, Y.-H. *et al.* A quorum-sensing signaling system essential for genetic competence in *Streptococcus mutans* is involved in biofilm formation. *Journal of Bacteriology* **184**,

- 2699–2708 (2002).
118. Liao, S. *et al.* Streptococcus mutans extracellular DNA is upregulated during growth in biofilms, actively released via membrane vesicles, and influenced by components of the protein secretion machinery. *Journal of Bacteriology* **196**, 2355–2366 (2014).
 119. Christersson, C. E., Lindh, L. & Arnebrant, T. Film-forming properties and viscosities of saliva substitutes and human whole saliva. *Eur. J. Oral Sci.* **108**, 418–425 (2000).
 120. Aguirre, A. A. *et al.* Lubrication of selected salivary molecules and artificial salivas. *Dysphagia* **4**, 95–100 (1989).
 121. Pinna, R., Campus, G., Cumbo, E., Mura, I. & Milia, E. Xerostomia induced by radiotherapy: an overview of the physiopathology, clinical evidence, and management of the oral damage. *Ther Clin Risk Manag* **11**, 171–188 (2015).
 122. Li, Y. H., Lau, P. C. Y., Lee, J. H., Ellen, R. P. & Cvitkovitch, D. G. Natural Genetic Transformation of Streptococcus mutans Growing in Biofilms. *Journal of Bacteriology* **183**, 897–908 (2001).
 123. Shen, H., Hu, Y. & Saltzman, W. M. DNA diffusion in mucus: effect of size, topology of DNAs, and transfection reagents. *Biophysj* **91**, 639–644 (2006).
 124. Reck, M., Tomasch, J. & Wagner-Döbler, I. The Alternative Sigma Factor SigX Controls Bacteriocin Synthesis and Competence, the Two Quorum Sensing Regulated Traits in Streptococcus mutans. *PLoS Genet.* **11**, e1005353 (2015).
 125. Moye, Z. D. *et al.* Effects of Carbohydrate Source on Genetic Competence in Streptococcus mutans. *Applied and Environmental Microbiology* **82**, 4821–4834 (2016).
 126. Strober, W. Impact of the gut microbiome on mucosal inflammation. *Trends in immunology* (2013).
 127. Devkota, S. & Chang, E. B. Nutrition, microbiomes, and intestinal inflammation. *Current opinion in gastroenterology* (2013).
 128. Dickerson, F., Severance, E. & Yolken, R. The microbiome, immunity, and schizophrenia and bipolar disorder. *Brain Behav. Immun.* **62**, 46–52 (2017).
 129. Celiberto, L. S., Bedani, R., Rossi, E. A. & Cavallini, D. C. U. Probiotics: The scientific evidence in the context of inflammatory bowel disease. *Crit Rev Food Sci Nutr* **57**, 1759–1768 (2017).
 130. O'Mahony, S. M., Clarke, G., Borre, Y. E. & Dinan, T. G. Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. *Behavioural brain ...* (2015).
 131. Galland, L. The Gut Microbiome and the Brain. *Journal of Medicinal Food* **17**, 1261–1272 (2014).
 132. Foster, J. A. & McVey Neufeld, K.-A. Gut–brain axis: how the microbiome influences anxiety and depression. *Trends in Neurosciences* **36**, 305–312 (2013).
 133. Pei, R., Martin, D. A., DiMarco, D. M. & Bolling, B. W. Evidence for the effects of yogurt on gut health and obesity. *Crit Rev Food Sci Nutr* **57**, 1569–1583 (2017).
 134. Vaiserman, A. M., Koliada, A. K. & Marotta, F. Gut microbiota: A player in aging and a target for anti-aging intervention. *Ageing Res. Rev.* **35**, 36–45 (2017).
 135. Schjørring, S. & Krogfelt, K. A. Assessment of Bacterial Antibiotic Resistance Transfer in the Gut. *International Journal of Microbiology* **2011**, 1–10 (2011).
 136. Lanka, E. & Wilkins, B. M. DNA Processing Reactions in Bacterial Conjugation. *Annu. Rev. Biochem.* **64**, 141–169 (1995).
 137. Hausner, M. & Wuertz, S. High Rates of Conjugation in Bacterial Biofilms as Determined by Quantitative In Situ Analysis. *Applied and Environmental Microbiology*

- (1999).
138. Phornphisutthimas, S., Thamchaipenet, A. & Panijpan, B. Conjugation in *Escherichia coli*: A laboratory exercise. *Biochem Mol Biol Educ* **35**, 440–445 (2007).
 139. Celli, J. *et al.* Viscoelastic Properties and Dynamics of Porcine Gastric Mucin. *Biomacromolecules* **6**, 1329–1333 (2005).
 140. Tenailon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
 141. Schultz, M. Clinical use of *E. coli* Nissle 1917 in inflammatory bowel disease. *Inflamm. Bowel Dis.* **14**, 1012–1018 (2008).
 142. Katouli, M. Population structure of gut *Escherichia coli* and its role in development of extra-intestinal infections. *Iran J Microbiol* **2**, 59–72 (2010).
 143. Shornikova, A. V., Casas, I. A., Isolauri, E., Mykkänen, H. & Vesikari, T. *Lactobacillus reuteri* as a therapeutic agent in acute diarrhea in young children. *J. Pediatr. Gastroenterol. Nutr.* **24**, 399–404 (1997).
 144. Valeur, N., Engel, P., Carbajal, N., Connolly, E. & Ladefoged, K. Colonization and immunomodulation by *Lactobacillus reuteri* ATCC 55730 in the human gastrointestinal tract. *Applied and Environmental Microbiology* **70**, 1176–1181 (2004).
 145. Jones, S. E. & Versalovic, J. Probiotic *Lactobacillus reuteri* biofilms produce antimicrobial and anti-inflammatory factors. *BMC Microbiol.* **9**, 35 (2009).
 146. Barr. Bacteriophage adhering to mucus provide a non-host-derived immunity. *pnas.org* Available at: <http://www.pnas.org/content/early/2013/05/16/1305923110.full.pdf?withds=yes>. (Accessed: 20 May 2013)
 147. Van Tassell, M. L. & Miller, M. J. *Lactobacillus* adhesion to mucus. *Nutrients* **3**, 613–636 (2011).
 148. Reniero, R., Cocconcelli, P., Bottazzi, V. & Morelli, L. High frequency of conjugation in *Lactobacillus* mediated by an aggregation-promoting factor. *Journal of General Microbiology* **138**, 763–768 (1992).
 149. Pacheco, A. R., Curtis, M. M., Ritchie, J. M. & Munera, D. Fucose sensing regulates bacterial intestinal colonization. *Nature* (2012).
 150. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96–99 (2013).
 151. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
 152. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
 153. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–335 (2012).
 154. Hunt, K. M. *et al.* Human Milk Oligosaccharides Promote the Growth of Staphylococci. *aem.asm.org*
 155. Kunz, C., Rudloff, S., Baier, W., Klein, N. & Strobel, S. Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu. Rev. Nutr.* **20**, 699–722 (2000).
 156. Marcobal, A. *et al.* Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10**, 507–514 (2011).
 157. Bode, L. Recent advances on structure, metabolism, and function of human milk

- oligosaccharides. *J. Nutr.* **136**, 2127–2130 (2006).
158. Rudloff, S. & Kunz, C. Milk oligosaccharides and metabolism in infants. *Adv Nutr* **3**, 398S–405S (2012).
 159. De Leoz, M. L. A. *et al.* Lacto-N-tetraose, fucosylation, and secretor status are highly variable in human milk oligosaccharides from women delivering preterm. *J. Proteome Res.* **11**, 4662–4672 (2012).
 160. Lambiase, A., Catania, M. R. & Rossano, F. Anaerobic bacteria infection in cystic fibrosis airway disease. *New Microbiol.* **33**, 185–194 (2010).
 161. Lengeler, J. W. & Jahreis, K. Bacterial PEP-dependent carbohydrate: phosphotransferase systems couple sensing and global control mechanisms. *Contrib Microbiol* **16**, 65–87 (2009).
 162. Deutscher, J., Francke, C. & Postma, P. W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).
 163. Postma, P. W., Lengeler, J. W. & Jacobson, G. R. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. (1993).
 164. Kok, M. Effect of enzyme I of the bacterial phosphoenolpyruvate : sugar phosphotransferase system (PTS) on virulence in a murine model. *Microbiology* **149**, 2645–2652 (2003).
 165. Schaeffer, P., Millet, J. & Aubert, J. P. Catabolic repression of bacterial sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **54**, 704–711 (2005).
 166. Abranches, J., Chen, Y. Y. M. & Burne, R. A. Characterization of *Streptococcus mutans* Strains Deficient in EIIABMan of the Sugar Phosphotransferase System. *Applied and Environmental Microbiology* **69**, 4760–4769 (2003).
 167. Deming, T. J. Mussel byssus and biomolecular materials. *Curr Opin Chem Biol* **3**, 100–105 (1999).
 168. Waite, J. H. Evidence for a repeating 3, 4-dihydroxyphenylalanine-and hydroxyproline-containing decapeptide in the adhesive protein of the mussel, *Mytilus edulis* L. *J. Biol. Chem.* **258**, 2911–2915 (1983).
 169. Waite, J. H., Housley, T. J. & Tanzer, M. L. Peptide repeats in a mussel glue protein: theme and variations. *Biochemistry* **24**, 5010–5014 (1985).
 170. Lin, Q. *et al.* Adhesion mechanisms of the mussel foot proteins mfp-1 and mfp-3.