DOES HOST MUCIN - *SALMONELLA* ADHESION INTERACTION STEER DISEASE INVASIVENESS?: A COMPARATIVE STUDY USING COMMON CLINICALLY IMPORTANT SEROVARS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DENİZ TUNALIER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FOOD ENGINEERING

JULY 2023

Approval of the thesis:

DOES HOST MUCIN - *SALMONELLA* ADHESION INTERACTION STEER DISEASE INVASIVENESS?: A COMPARATIVE STUDY USING COMMON CLINICALLY IMPORTANT SEROVARS

submitted by **DENİZ TUNALIER** in partial fulfillment of the requirements for the degree of **Master of Science in Food Engineering, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Hami Alpas
Head of the Department, **Food Engineering**

Prof. Dr. Yeşim Soyer Küçükşenel
Supervisor, **Food Engineering, METU**

Assoc. Prof. Dr. Soraya Chaturongakul
Co-Supervisor, **Inst of Mol. Biosciences, Mahidol University**

Examining Committee Members:

Prof. Dr. Mecit Halil Öztop
Food Engineering, METU

Prof. Dr. Yeşim Soyer Küçükşenel
Food Engineering, METU

Assoc. Prof. Dr. Soraya Chaturongakul
Inst. of Mol. Biosciences, Mahidol University

Assoc. Prof. Dr. Çağdaş Devrim Son
Biological Sciences, METU

Assist. Prof. Dr. Emine Ertekin
Biological Sciences, METU

Date: 18.07.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : Deniz Tunalıer

Signature :

ABSTRACT

DOES HOST MUCIN - *SALMONELLA* ADHESION INTERACTION STEER DISEASE INVASIVENESS?: A COMPARATIVE STUDY USING COMMON CLINICALLY IMPORTANT SEROVARS

Tunalıer, Deniz
Master of Science, Food Engineering
Supervisor: Prof. Dr. Yeşim Soyer Küçükşenel
Co-Supervisor: Assoc. Prof. Dr. Soraya Chaturongakul

July 2023, 154 pages

Salmonella enterica subsp. *enterica* (*Salmonella*) which is a Gram-negative, rod-shaped, and motile bacterium, is one of the most prevalent foodborne pathogens. World Health Organization (WHO) highlighted that at least 100 million people suffer from salmonellosis annually. *Salmonella* has over 2600 serovars which are characterized in terms of host specificity or clinical manifestation. Host mucin 1 (MUC1) is a receptor for *Salmonella* giant adhesin *Salmonella* intestine infection (SiiE). The presence of *siiE* might affect the course of the disease. The severity of salmonellosis might be affected by MUC1 and SiiE interaction. In this study, it's hypothesized that the degree of invasiveness through this interaction varies regarding host specificity of *Salmonella.* Determining protein-protein interactions (PPIs) between SiiE adhesin of *Salmonella* and its host receptor MUC1 to observe binding characteristics of different *Salmonella* serovars is the aim of the study. The gene sequences of *siiE* of previously whole genome sequenced (WGS) local METU *Salmonella* isolates, as well as universal *Salmonella* isolates retrieved from National Center for Biotechnology Information (NCBI). METU isolates were selected by their original hosts, molecular subtypes, and frequency of making disease in humans in

Türkiye. Similar approach was used to select serovars world-wide from NCBI; serovars representing the most and the least observation rates for making disease in humans among the continents were used. By using sequence data of previously WGS METU *Salmonella* isolates (n=47), primer sets were designed to conduct multiplex PCR to detect presence of *siiE* gene. Overall, 78 representative *Salmonella* isolates were subjected to screening of the gene. According to the multiplex PCR results, all these isolates (n=78) have *siiE,* therefore it is concluded that *siiE* is an abundant gene among *Salmonella* serovars. For *in silico* analysis, genome data of important 37 *Salmonella* isolates were obtained from NCBI. Since SiiE is the biggest protein in *Salmonella* proteome, including 5559 amino acids, a representative region of this protein which is bacterial immunoglobulin-like domains 50^{th} to 52^{nd} (BIg50:52), was used for further analysis. A phylogenetic tree was constructed from SiiE BIg50:52 to observe amino acid substitutions in *Salmonella* isolates (n=84). Total of 20 clusters were observed in the phylogenetic tree and the most prevalent substitutions were detected on residues 5113 and 5336. Also, secondary structures of WGS METU *Salmonella* isolates (n=47) and selected isolates from NCBI (n=5) were determined by using threading approach to furtherly use in docking with rigid body method. From docking between SiiE and MUC1, PPIs showed that except disulfide bridges other interactions have a role on binding. Moreover, no correlation was observed for interaction of MUC1 and SiiE BIg50:52 of different *Salmonella* serovars in terms of host specificity. Finally, it was remarked that SiiE's residues 5174 and 5175 may have a vital role on host – receptor interaction between SiiE of *Salmonella* and MUC1 of *Homo sapiens*.

This study contributed to a better understanding of *Salmonella* – host interactions and observing the effect of *Salmonella* diversity on binding characteristics and invasiveness.

Keywords: *Salmonella*, multiplex polymerase chain reaction (PCR), secondary structure modelling, docking, protein – protein interactions

ÖZ

KONAK MÜSİN - *SALMONELLA* TUTUNMA ETKİLEŞİMİ HASTALIĞIN İNVAZYASYONUNU ETKİLİYOR MU?: KLİNİK OLARAK ÖNEMLİ YAYGIN SEROVARLARIN KULLANILDIĞI KARŞILAŞTIRMALI BİR ÇALIŞMA

Tunalıer, Deniz
Yüksek Lisans, Gıda Mühendisliği
Tez Yöneticisi: Prof. Dr. Yeşim Soyer Küçükşenel
Ortak Tez Yöneticisi: Doç. Dr. Soraya Chaturongakul

Temmuz 2023, 154 sayfa

Gram-negatif, çubuksu ve hareketli *Salmonella* enterica subsp. *enterica* (*Salmonella*) en yaygın gıda kaynaklı patojenlerdendir. Dünya Sağlık Örgütü (DSÖ) yılda en az 100 milyon insanın salmonelloz hastalığına yakalandığını belirtmiştir. Konak özgüllüğü ya da klinik belirtilere göre sınıflandırılan *Salmonella*'nın 2600'den fazla suşu vardır. *Salmonella'*nın büyük adezini SiiE'nin konak reseptörü MUC1'dir. *siiE*'nin bulunması hastalığın seyrini etkileyebilmektedir. Salmonellozun şiddeti MUC1 ve SiiE etkileşiminden etkilenebilir. Bu çalışmada *Salmonella*'nın konak özgüllüğünün, bu etkileşim yoluyla nüfuz edebilmesinin derecesini etkilediği varsayılmaktadır. Bu çalışmanın amacı, farklı *Salmonella* suşlarının bağlanma özelliklerini gözlemlemek için *Salmonella*'nın SiiE adezini ve konak reseptörü MUC1 arasındaki protein–protein etkileşimlerini (PPE) incelemektir. Ulusal Biyoteknoloji Bilgi Merkezi'nden (UBBM), önceden tüm genomu dizilenmiş yerel ODTÜ *Salmonella* izolatlarının hem de evrensel *Salmonella* izolatlarının *siiE* gen dizileri alınmıştır. ODTÜ izolatları konakçıları, moleküler alt tipleri ve Türkiye'de insanlarda hastalık yapma sıklıklarına göre seçilmiştir. Benzer yaklaşım UBBM'den dünya çapındaki suşları seçerken de kullanılmıştır; kıtalar arasında insanlarda hastalık yapmada en sık ve en az gözlemlenen suşlar kullanılmıştır. Toplam gen dizilimi yapılmış ODTÜ *Salmonella* izolatlarının (n=47) dizi verileriyle, *siiE*'nin varlığını multipleks polymerize zincir reaksiyonuyla (PZR) belirlemek için primer setleri tasarlanmıştır. Toplamda 78 *Salmonella* izolatının geni taranmıştır. Multipleks PZR sonuçlarına göre, bütün izolatlarda (n=78) *siiE* bulunduğundan, *siiE*'nin *Salmonella* suşlarında yaygın olduğu sonucuna varılmıştır. Bilgisayarda analiz için, 37 önemli *Salmonella* suşunun toplam gen verileri UBBM'den alınmıştır. 5559 amino asit içeren SiiE, *Salmonella* proteomunun en büyük proteini olduğundan, bu proteinin temsili bölgesi, 50.'den 52.'ye bakteriyel immünglobulin benzeri domainleri (BIg50:52) ileriki analizlerde kullanılmıştır. *Salmonella* izolatlarındaki (n=84) amino asit farklılıklarını belirlemek için SiiE BIg50:52'den filogenetik ağaç oluşturulmuştur. Bu ağaçta 20 adet kümelenme belirlenmiştir ve en çok farklılığın 5113. ve 5336. amino asitlerde olduğu tespit edilmiştir. Ayrıca toplam gen dizilimi yapılmış olan ODTÜ *Salmonella* izolatları (n=47) ve UBBM'den seçilen *Salmonella* izolatlarının (n=5) ikincil yapı modellenmesi, sabit cisim yanaştırmasında kullanılmak için iplikleme yöntemiyle yapılmıştır. SiiE-MUC1 yanaştırmasından, PPE'ler disülfit köprüleri hariç diğer etkileşimlerin bağlanmada rol oynadığını göstermiştir. Ek olarak MUC1 ve farklı *Salmonella* suşlarının SiiE BIg50:52'sinin etkileşiminde, *Salmonella* konak özgüllüğü açısından ilişki gözlemlenmemiştir. Son olarak, SiiE'nin 5174. ve 5175. amino asitlerinin *Salmonella*'nın SiiE'si ve *Homo sapiens*'in MUC1'i arasındaki konak reseptör etkileşiminde önemli bir rolü olabileceğine dikkat çekilmiştir.

Bu çalışma *Salmonella* – konak etkileşimini ve *Salmonella* çeşitliliğinin konağa bağlanma özelliklerinde ve nüfuz edebilmesindeki etkisini daha iyi bir şekilde anlamaya katkıda bulunmuştur.

Anahtar Kelimeler: *Salmonella*, multipleks polimerize zincir reaksiyonu, ikincil yapı modelleme, yanaştırma, protein – protein etkileşimleri

To my beloved parents,

who made me the person I am today

ACKNOWLEDGMENTS

Primarily, I wish to thank my advisor Prof. Dr. Yeşim Soyer for her encouragements and guidance. She has been an inspiring role model to me since my undergraduate years. The reason why I decided to continue in academia is frankly her stance on life.

I would like to thank my co-advisor Assoc. Prof. Dr. Soraya Chaturongakul for her support.

I would like to express my gratitude to Assoc. Prof. Dr. Çağdaş Devrim Son for his guidance and support. He was always helpful with a smile whenever I needed.

I would also like to thank Emre Koç and Diala Konyali for introducing me to lab and teaching experimental procedures.

I also wanted to thank Uğurcan Yıldırım and Hayri Kerim İnce for introducing me to bioinformatics and encouragements.

I would like to express my appreciation to Muhammed Kürşat Yurt for his time to be my IT-guy.

I also would like to thank my lab mates Murat Dülger, Mustafa Güzel, İrem İskender, Hilal Samut, Ayşenur Yücefaydalı, Segâh Yetişkin and Filiz Yeniköylü. For their support and making this experience fun.

I would also express my gladness to Merve Dönmez, Nazlı Özkoca and Oğuzcan Dedeci. I am lucky to have you for more than a decade.

I also would like to thank Ege Can Kurter for his accompaniment in both good and bad times.

Last but most important, I wish to thank my beloved parents for their endless support and love on me. Children cannot choose their parents, but if I had a choice, I would still choose you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

SPI	*Salmonella* pathogenicity island
T1SS	Type 1 secretion system
BIg	Bacterial immunoglobulin
SEA	Sperm protein-enterokinase-agarin
NMR	Nuclear magnetic resonance
Cryo-EM	Cryo-electron microscopy
DNA	Deoxyribonucleic acid
MD	Molecular dynamics
PPI	Protein-protein interaction
METU	Middle East Technical University
NCBI	National Center for Biotechnology Information
FASTA	Fast-all
BLAST	Basic local alignment search tool
PFGE	Pulsed Field Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDB	Program database
AF	AlphaFold
OTU	Operational taxonomic unit
MSA	Multiple sequence alignment
AA	Amino acid
WGS	Whole genome sequence
RMSD	Root Mean Square Deviation

CHAPTER 1

INTRODUCTION

Foodborne illnesses play a crucial role in public health in terms of both mortality and socioeconomic effect. These diseases occur due to consumption of contaminated foods and can be encountered during any stage of farm to fork food supply chain (WHO, 2019). According to estimation of WHO, 600 million people get sick and 420,000 people die due to foodborne illnesses in the world annually (WHO, 2022). WHO also reported that among other pathogens that put foodborne burden on human health, *Salmonella* enterica subsp. *enterica* (*Salmonella*) is the number one contributor (WHO, 2015). Salmonellosis is the disease caused by *Salmonella* and has consequences of gastrointestinal effects and fever. Prime foods that are associated with salmonellosis are meat and poultry products, eggs, dairy products, nut butters, fresh vegetables, and raw sprouts (FDA, 2019). Although *Salmonella enterica* has more than 2600 different serovars, not all serovars cause the same disease or the same symptoms. Some of these serovars are known for causing invasive illnesses in humans such as *S*.Typhi and some of these are known for specific to swine such as *S*.Cholereasuis (Cheng et al., 2019). In other words, *Salmonella* diversity has an essential role on pathogenesis, disease severity and host specificity. When burdens on public health, various contamination routes and *Salmonella* diversity are taken into consideration, investigating the different *Salmonella* serovars' effect on invasiveness plays a significant role on revealing the impact of salmonellosis.

The process of bacteria to attach other cells is defined as bacterial adhesion. Precondition of bacteria to grow, colonize and eventually cause infection on host, is highly dependent on this initial attachment (Schaeffer, 1993). Consequently, one of the first and foremost steps of bacteria to cause pathogenesis is adhesion to host cells. The process of bacterial attachment on host cells requires recognition between particular proteins on bacterial surface and corresponding receptors on the surface of host cells (Rehman et al., 2019). Adhesins of *Salmonella* are one of the virulence factors that *Salmonella* requires in adhesion. *Salmonella* intestine infection (SiiE) which is a non-fimbrial adhesin of *Salmonella* has a crucial role of binding glycostructures on cell surface since it is considered to be responsible of first contact to polarized host cells (Barlag & Hensel, 2015). Therefore, highlighting the structural differences of different *Salmonella* serovars' SiiE adhesin and investigating the binding properties between SiiE adhesin of *Salmonella* and its receptor on host cell, have a vital effect on enlighten the invasiveness and consequently course of salmonellosis. Hence, this study is based on the hypothesis that *siiE* region varies among different *Salmonella* serovars and these differences affect the binding to its host cell receptor, eventually degree of invasiveness.

In this study, sequence data of local *Salmonella* isolates, as well as universal *Salmonella* isolates retrieved from National Center for Biotechnology Information (NCBI). Serovars representing the most and the least observation rates for making disease in humans in Türkiye and among the continents were used to achieve a better understanding of *Salmonella* – host interaction, since different serovars affect the course of salmonellosis. Additional to in *vitro* methods, *in silico* methods were also applied to enlighten the relation between SiiE adhesin of *Salmonella* and its receptor on host cell by using various bioinformatic tools. In the light of bioinformatic methods, binding characteristics between *Salmonella* and host cell were investigated so that salmonellosis can be detected and even prevented from the earliest stage.

CHAPTER 2

LITERATURE REVIEW

2.1 *Salmonella*

Salmonella enterica subsp. *enterica* (*Salmonella*), which is one of the commonly encountered foodborne pathogen, leads to foodborne disease in humans (Hald, 2013). According to the annual estimation of WHO, at least 100 million people get sick due to *Salmonella* in the world (WHO, 2018). *Salmonella* is regarded as one of the most significant foodborne pathogens when prevalance, mortality and burden on economy are taken under consideration (WHO, 2022).

Salmonella is nonsporulating, rod-shaped and gram-negative prokaryotic bacterium. It belongs to *Enterobacteriaceae* family. Although it can be nonmotile in culture, it is mobile owing to having multiple flagella (Mumy, 2014). The name *Salmonella* was originated and given bringing honor to Daniel Elmer Salmon who was a veterinary pathologist from America, by virtue of his efforts (Monte & Sellera, 2020). For the purpose of classifying *Salmonella* in serovar level, White-Kauffmann-Le Minor (WKL) scheme is used. The scheme consists of antigens of outer membrane lipopolysaccharides (O), flagellar proteins (H1 and H2) and capsular polysaccharides (Vi) (Longo et al., 2019). To determine the group of *Salmonella* and to specify the serovars, O antigen and H antigen are used respectively. However, Vi antigen is observed in only 4 *Salmonella* serovars among more than 2600 serovars (Ryan et al., 2017).

Nomenclature of *Salmonella* was considered as elaborate since, dissimilar systems had been used among scientist, until WHO Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute based system is accepted universally and solely. A consistent and coherent system is essential for effective communication between people who work in farm-to-fork continuum, health sector and scientists (Brenner et al., 2000). According to present WHO Collaborating Centre based system, genus *Salmonella* has two species which are *S.enterica* and *S.bongori*. The species *S.enterica* includes 6 subspecies which are *S.enterica* subsp. *enterica* (I), *S.enterica* subsp. *salamae* (II), *S.enterica* subsp. *arizonae* (IIIa), *S.enterica* subsp. *diarizonae* (IIIb), *S.enterica* subsp. *houtenae* (IV) and *S.enterica* subsp. *indica* (VI) (Su & Chiu, 2007). In 2005, a third species which is *S.subterranea* of genus *Salmonella* is recognized, however it is not included to the current system (Shelobolina et al., 2004).

Table 2.1 Nomenclature of *Salmonella* (Su & Chiu, 2007)

Genus	**Species**	**Subpecies**
Salmonella	*enterica*	*enterica* (I)
		salamae (II)
		arizonae (IIIa)
		diarizonae (IIIb)
		houtenae (IV)
		indica (VI)
	bongori	V
	subterranea	

2.1.1 Diversity of *Salmonella* Serovars

Serovars which belong *Salmonella enterica* can be categorized into 3 groups in terms of host specificity. These groups are host-generalist, host-adapted and host-restricted. *S*. Enteritidis and *S*. Typhimurium are members of host-generalist, since they can affect broad range of hosts. Serovars under host-generalist can cause gastroentirical disease in both humans and animals (Huang et al., 2019). Host-adapted group consists of *S*. Dublin and *S*. Choleraesuis. Although the former mainly cause bacteremia in bowine and the latter mostly affect porcine, they may create infection in humans (Baumler & Fang, 2013). Host-restricted group includes *S*.

Typhi and *S*. Gallinarum. As the name implies, the members of this group can lead to fatal diseases to their related hosts. For example, *S*. Typhi only infect humans and cause lethal typhoid fever (Eswarappa et al., 2008).

Table 2.2 Examples of *Salmonella* serovars diversified by their abilities to cause diseases in different hosts (Singh, 2013)

Group	**Serovar**	**Host**
Host generalist	*S*. Typhimurium *S*. Enteritidis	Human, cattle, poultry, swine, mouse
Host adapted	*S*. Dublin *S*. Choleraesuis	Cattle, pig, mouse, chicken, human
Host restricted	*S*. Typhi *S*. Gallinarum *S*. Pullorum *S*. Abortusequi	Human, poultry, horse

Salmonella bongori are generally linked with illnesses in cold-bloded animals (Chan et al., 2003). *S. bongori* may also less frequently infect humans although, its reservoir is reptiles (Fookes et al., 2011).

2.1.2 *Salmonella* Outbreaks

The unexpected increase in a disease occurrence which is the number of cases increases in a certain location for a certain time, defined as outbreak (Houlihan & Whitworth, 2019). When two or more people suffer from the same disease after consuming the same food, a foodborne disease outbreak is occured. To determine the source of the foodborne outbreak, epidemiologic investigation or laboratory conformation can be employed (MN Department of Health, 2022).

In Illinois, United States, 5295 confirmed cases of *Salmonella* infection and 9 deaths were reported in 1985. The outbreak originated due to *S*. Typhimurium contamination in a milk processing plant. This outbreak considered as the largest *Salmonella* outbreak in milk (CDC, 1985; The New York Times, 1985).

In United States, 224,000 people underwent gastroenteritis due to *S. enteritidis* in 1994. The food source was ice-cream. The outbreak was linked to cross contamination due to carrying ice-cream premix in a tank which had carried unpasteurized eggs previously (Hennessy et al., 1996).

In late 2008 and early 2009, the largest *Salmonella* outbreak in peanut butter happened due to *S*. Typhimurium contamination in United States. More than 500 people got infected and 9 people were dead. Peanut butter and peanut butter containing products were recalled and it is known as one of the greatest food recalls in US (CDC, 2009; Harris, 2009).

Besides gastroenteritis, a couple of *Salmonella* serovars can cause enteric fever. In a 5-year period between 2015 and 2020, 11 outbreaks were reported among the world. The source of these outbreaks was consuming contaminated food or drinking polluted water while, some of these outbreaks' food source still remains unknown (Popa & Popa, 2021).

The most recent *Salmonella* outbreak was reported in October 2022 in US. 33 people got infected with *Salmonella* and 13 people were hospitalized. The food source was raw salmon. The investigation of the outbreak is continuing (FDA, 2022).

2.1.3 Salmonellosis

The illness caused by nontyphoidal *Salmonella* is defined as salmonellosis (Department of Health, 2022). According to a research which was conducted by WHO, *Salmonella* puts the highest foodborne burden on human health among other pathogens (WHO, 2015). Annualy 180 million of diarrheal diseases occur due to *Salmonella* globally (Besser, 2018). Besides most common sources of *Salmonella* infection which are eggs, meat and dairy products, vegetables and water are also sources of the infection. In other words, food is the main source of salmonellosis especially in developed countries (Popa & Popa, 2021). Main symptoms of the diseases are stomach cramps, diarrhea and fever. These symptoms initiate 6 hours to 6 days and continues 4 to 7 days generally (CDC, 2022). Since salmonellosis is an infectious disease, multiplication of *Salmonella* initially in the small intestine followed by colonization and eventually results in invasion of intestinal tissue due to production of enterotoxins. Although most prevalent symptomps are mentioned previously, *Salmonella* can enter blood stream and even lymphatyic system. Put differently, *Salmonella* can also cause fatal illnesses. Also, the severity and the course of the disease depend to the host. Immunocompromised people such as infants, elderly, pregnants and people with chronical illnesses are more prone to salmonellosis (Bell & Kyriakides, 2009).

Table 2.3 Diseases which are caused by *Salmonella* (Bell & Kyriakides, 2009)

Disease	**Causative Serovars**	**Infective Load / Cause**
Gastroenteritis	*S.enterica* subsp. *enterica* and *S.enterica* subsp. *arizonae*	>10^4 cells.
Enteric fever	*S.* Typhi and *S.* Paratyphi	<10^3 cells.
Septicemia / bacteramia	*S.enterica* subsp. *enterica*	Presence of *Salmonella* in bloodstream
Sequelae	*S.enterica* subsp. *enterica*	Formerly underwent salmonellosis

2.1.4 Virulence Capacity of *Salmonella*

The ability of an organism to develop infection in the host and eventually to result with disease is defined as virulence. The molecules which aid the colonization of bacterium in the host are known as virulence factors. Virulence factors of a bacterium can be cytosolic, membrane associated or secretory (Sharma et al., 2016). The genes located on chromosomes and the genes located on virulence plasmid are the two main segments of *Salmonella*'s virulence factor encoding genes (Singh et al., 2018).

Salmonella virulence plasmid (Spv) are present in some serovars of *S. entrica* such as *S*. Typhimurium. Although the exact role of Spv is still unclear, a recent study pointed out that *Salmonella* virulence plasmid has a role on restraining innate immune responses of host (Wu et al., 2016). *Salmonella* pathogenicity islands (SPIs) are desciribed as certain regions which encode virulence genes and located on clusters of *Salmonella* chromosomes (Wang et al., 2020a). Up to date, there are 24 *Salmonella* pathogenicity islands exist (Cheng et al., 2019).

Table 2.4 *S.enterica* subsp. *enterica* pathogenicity islands (Cheng et al., 2019)

SPI	Size (kb)	Main genes / Gene products	Function
SPI-1	40	T3SS, effector proteins (e.g. *avrA, iacB, invB, sicA, sicP, sipA, sipB, sipC* and *sptP*)	Host cell invasion and enteropathy
SPI-2	40	T3SS, effector proteins (*ssaB, ssaE, sscA, sscB, sseA, sseF, sseG* and *ttr*)	Intracellular survival
SPI-3	17 – 36	Magnesium transport system (*mgtCB*), *misL*	Intracellular survival, intestinal colonization
SPI-4	25	T1SS (*siiABCDF*), non - fimbrial adhesin (*siiE*)	Adhesion and invasion of epithelial cells, virulence in mice and cows
SPI-5	11-44	Effectors of SPI-1 and SPI-2 (*pipABC, sopB, pipD* and *sigDE*)	Epithelial invasion, enteric salmonellosis, chicken colonization
SPI-6	47 – 59	T6SS, atypical fimbriae (*safABCD*), fimbriae (*tcfABCD*) and *sciS*	Invasion, intramacrophage survival, chicken colonization, virulence in mice
SPI-8	6.8	Bacteriocin fragment	Unkown
SPI-9	16	T1SS, adhesin	Transport, epithelial adherence
SPI-10	33	P4 – like prophage, *Sef* fimbriae	Virulence in mice and chickens, intramacrophage uptake / survival
SPI-11	6 – 10	*pagCD, envF* sRNA RaoN, typhoid toxin gene islet (*cdtB, pltA* and *pltB*)	Intramacrophage survival, serum resistance, typhoid fever pathology

Table 2.4 (cont'd)

SPI-12	6 – 15	*sspH2*	Actin polymerization, virulence in mice
SPI-13	25	Putative lyase, hydrolase, oxidase and arylsulphatase regulator	Macrophage internalization, virulence in chickens and mice
SPI-14	9	*gpiAB*, putative acyl-CoA dehydrogenase	Chicken pathogenicity, epithelial invasion
SPI-16	4.5	Bactoprenol glucosyl transferase and translocase (*gtrAB*)	LPS modification, seroconversion
SPI-17	5	Bactoprenol glucosyl transferase and translocase (*gtrAB*)	LPS modification, seroconversion
SPI-18	2.3	*hlyE* hemolysin, *taiA* invasion – associated protein	Epithelial invasion
SPI-19	45	T6SS	Intramacrophage survival, chicken colonization
SPI-23	37	T3SS effectors (*sanA*, *chlR*, *shaU* and *dumE*)	Host cell adherence and invasion, invasion of pig epithelial cells
SPI-24	25	Outer membrane protein (*shdA*, *sivH*, *ratAB*, *sinl* and *potR*)	Fibronectin binding, murine intestinal colonization, intramacrophage survival

Virulence factors of *Salmonella* can have a role on colonization in host, adhesion, invasion and advance in survival of *Salmonella* as solely or in company with each other (Cui et al., 2021).

Salmonella Pathogenicity Island 4 (SPI-4) encodes 6 genes which are *siiABCDEF*. This pathogenicity island can be also observed in different subspecies of *Salmonella enterica*, such as *houtanae* and *indica*. SPI-4 can also be observed in *S. bongori* species (Cheng et al., 2019).

2.1.4.1 *Salmonella* Pathogenicity Island 4 (SPI-4)

Wong and his collegues were defined SPI-4 for the first time in 1998 (Main-Hester et al., 2008). 6 genes which are *siiABCDEF* are carried by *Salmonella* pathogenicity island – 4. Among these 6 genes, SiiC, SiiD and SiiF are responsible for constituting secretion apparatus for SiiE (*Salmonella* intestine infection) which is a 600 kDa adhesin of *Salmonella* and plays a role on adhesion to epithelial cells. (Kiss et al., 2007; Gerlach et al., 2007).

Figure 2.1 Representation of *Salmonella* pathogenicity island 4 adapted from Gerlach et al. (2007) and created with BioRender.com

2.1.4.2 Adhesins of *Salmonella*

Adhesins are one of the virulence factors which have a crucial role on attachment of pathogenic bacteria including *Salmonella* to host cells (Hallstrom & McCormick, 2015). For bacteria to generate pathogenesis, bacteria attach to mucosal surface and infect the mucosal tissue with the help of adhesins (Klemm & Schembri, 2000). In other words, host cell surface receptors can be recognized with adhesins. Adhesins can be cathegorized in three groups which are pili / fimbriae, non – pilus adhesins and curli which are multimeric molecules, monomeric or oligomeric proteins and thin fibers respectively (Ghosh et al., 2011).

Fimbrial adhesins are the most prevalent group of adhesins. Fimbriae / pili are proteins which composed of hundreds of major subunits with tipped or interspersed minor subunits. One or more of these subunits have the function of adhesion. Also, fimbrial adhesins' size in terms of length and diameter can be variable (Abraham et al., 2015). Moreover, by investigating the presence of fimbrial genes with various molecular methods, *Salmonella* and non – *Salmonella* can be differentiated and even *Salmonella* can be identified in serovar level (Rehman et al., 2019).

Non – fimbrial adhesins are known as heterogenous when compared to fimbrial ones which are known as homogenous. Non – fimbrial adhesins are secreted via type 1 secretion system (T1SS) or autotransported through bacterial membranes (Wagner & Hensel, 2011). For the former SiiE (*Salmonella* intestine infection) and BapA (Biofilm Association Protein) can be given as example. SiiE of *Salmonella* is responsible for attachment to epithelial cells, while BapA is responsible for producing cellulose composing biofilm and curli fimbriae (Azimi et al., 2020). ShdA, MisL and SadA are the examples of autotransported adhesins of *Salmonella*. Recognizing and particular binding to fibronectin are the function of ShdA, while MisL has a role on facilitating formation of biofilm and aggregation of bacteria (Urritia et al., 2014; Wang et al., 2018). Finally, Raghunathan and his collegues (2011) observed that SadA expression enhances adhesion of human epithelial cells, formation of biofilm and aggregation of cells.

The last group of adhesins are curli which are thin aggregative fibers (Römling et al., 1998). CsgA and CsgB can be given as examples of curli. These type of adhesins have an influence on biofilm stability and structural integrity (Nicastro et al., 2019).

Also developing vaccines and therapeutic drugs against salmonellosis can be achievable, since adhesins are present in the extracellular matrix (Rehman et al., 2019). For pathogenesis a single adhesin can have a vital role, although pathogens including *Salmonella* may express multiple adhesins (Klemm & Schembri, 2000). Adhesiome of *Salmonella enterica* consists of 20 adhesins which comprised of maximum 13 fimbrial and minimum 7 non-fimbrial adhesins contingent upon *Salmonella* serovar (Hansmeier et al., 2017). The function of SiiE which is a type of non-fimbrial adhesin of *Salmonella,* is proposed as responsible for the initial contact to host cell recently, and consequently SiiE has begun to attract attention in the literature (Barlag & Hensel, 2015; Li et al., 2019).

2.1.4.3 *Salmonella* Adhesin SiiE

SiiE which is a huge non – fimbrial adhesin of *Salmonella*, has a size of 595 kDa. It is known as the largest protein of *Salmonella* proteome (Barlag & Hensel, 2015). Also, SiiE is encoded by SPI – 4 and secreted by T1SS (Li et al., 2019). The giant adhesin of *Salmonella* consists of 5559 amino acid residues and has 53 repetetive bacterial immunoglobulin-like (BIg) domains. According to phylogenetic analysis of BIg domains of SiiE, domains 51 and 52 can be regarded as interpreters of other SiiE domains, while domain 50 is considered as a unique domain. Therefore, SiiE BIg50:52 have a special importance to gain an understanding on whole SiiE domains and research have been conducted for SiiE BIg50:52 in the literature (Griessl et al., 2013; Peters et al., 2017).

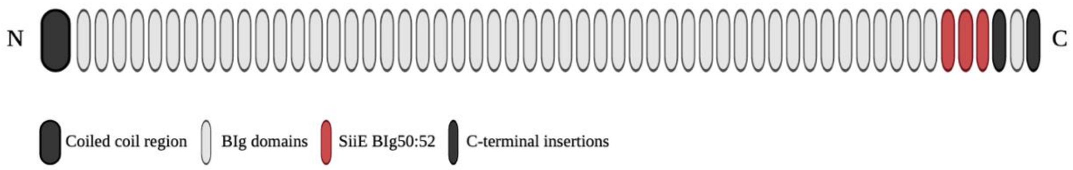


Figure 2.2 Bacterial immunoglobulin-like domain architecture of SiiE adapted from Griessl et al. (2013) and created with BioRender.com

The function of SiiE is known for facilitating the binding of *Salmonella* to polarized epithelial cells from the apical side (Gerlach et al., 2007). Further resarches of Gerlach and his collegues (2008) showed that a decrease in invasion of *Salmonella* to polarized cells when number of deleted BIg domains of SiiE is increased. Moreover, a more recent study revealed that SiiE facilitates the binding of polarized epithelial cells in a lectin – like fashion with the insights gained from lectin blockade experiments (Wagner et al., 2014). Finally, crystal structure of SiiE BIg50:52 is resolved with X – Ray diffraction experiments and 2 types of Ca^{2+} binding sites were determined. Also the researchers highlighted that Ca^{2+} binding leads SiiE to rigidify and aids stabilizing SiiE on cell surface (Griessl et al., 2013).

2.2 Human Defense Mechanism Against *Salmonella*

Immune system defends body against pathogenic bacteria including *Salmonella* with the coherent function of cells, proteins, chemicals, and organs by recognizing surface antigens of the pathogens. According to this recognition, the immune system is categorized into two lines of defense which are innate immunity and adaptive immunity (Thompson, 2015). The main difference between them is that the former is antigen-independent while the latter is antigen-dependent. In other words, innate immunity is non-specific, while adaptive immunity is specific against pathogens. Therefore, adaptive immunity has an immunologic memory against pathogens and a lag exposure time. On the contrary, innate immunity acts immediately after exposure and has no immunological memory (Marshall et al., 2018). T and B lymphocytes are the examples of adaptive immunity cells whereas, phagocytic cells, natural killer cells, antigen presenting cells and epithelia can be given as examples for innate immune system cells (Hato, T & Dagher, 2015).

Salmonella needs to pass through the mucus layer to get in contact with the epithelium during invasion (Broz et al., 2012). The epithelium which is a part of innate immune system is covered by mucus layer. Therefore, the mucus layer act as a physical barrier and form initial defense line against enteric pathogens including *Salmonella* (Dharmani et al., 2008).

2.2.1 Protection Properties of Mucus

Gastrointestinal (GI), urogenital and respiratory tracts which are not enveloped with skin and ar exposed surfaces of humans, are sheated with mucus (Xu et al., 2019). Water is the principal component of mucus and it corresponds to 95% of total mucus wet mass. Mucin glycoproteins and other molecules are other constituents of mucus which correspond to 5% and 2% of mucus respectively. Cellular debris, secretory IgA, lipids, peptides and nucleic acids are stand for the 2% of other molecules (Bakshani et al., 2018). Intestinal mucus layer does not only contact with resident microorganisms, but also is exposed to chemical irritants, toxins and intestinal pathogens including *Salmonella* (Dharmani et al., 2008). This mucus layer acts as an initial defense mechanism against these threats (Duangnumsawang et al., 2021). The defense mechanisms against intestinal pathogens including *Salmonella* can be listed as conserving resident flora, prevention of attachement of bacteria by simulating enterocytes' receptors, mucus glycoproteins' negative ionic charges, creating hydrophobic luminal barrier and presence of IgA (Braulio, 2005).

2.2.1.1 Mucins

The prime macromolecular constituents of mucus are mucins. There are more than 20 different mucins and each have specific sequence of repetative amino acids mostly threonine and / or serine (Voynow & Fischer, 2006). Mucins which are highly glycosylated glycoproteins can be cathegorized in two families which are secreted gel - forming mucins and membrane - tethered mucins (Brown et al., 2021; O'Sullivan & Montgomery, 2015). The latter which is also known as transmembrane mucins are components on all epithelial cells, particularly on small intestine's enterocytes (Dolan & Hansson, 2021). Secreted mucins make up physical barrier, while transmembrane mucins assits the barrier also and channel signals of survival and growth to internal parts of cells (Kufe, 2009).

Table 2.5 Transmembrane mucins of *Homo sapiens* (Dolan & Hansson, 2021)

Mucin gene	**Subtype***	**Main expression**
MUC1	SEA	Epithelial cells and non – epithelial cells
MUC3(A,B)	SEA	Intestinal epithelial cells
MUC4	NIDO-AMOP-VWD	Tonsils and other epithelial cells
MUC12	SEA	Intestinal epithelial cells
MUC13	SEA	Intestinal epithelial cells
MUC15	Other	Epididymis epithelium
MUC16	SEA	Uterus epithelium

Table 2.5 (cont’d)

MUC17	SEA	Intestinal epithelial cells
MUC21	Other	Esophagus epithelium
MUC22	Other	Esophagus epithelium

*: SEA : Sperm protein-enterokinase-agarin, NIDO: Nidogen-like, AMOP: Adhesion-associated, VWD: Von Willebrand factor

Interactions of receptor – receptor which stand for bacterial surface structures and host mucosal surface’s corresponding structures, are used when the topic is bacterial adhesion to mucosa of intestine. Pathogens including *Salmonella* can bind and / or degrade mucins, simultaneously mucins function as receptors for these pathogens. With the receptor – receptor binding between pathogens including *Salmonella* and mucins, pathogenic bacteria may be discharged with the flow of mucous or they can colonize in mucous or permeate the mucous layer and attach to epithelial tissues consequently (Braulio, 2005). Recently it was revealed that mucin 1 (MUC1), encoded by *MUC1* gene, is a host receptor for *Salmonella* adhesin SiiE (Li et al., 2019).

2.2.1.2 Human Mucin MUC1

MUC1 gene encodes MUC1 which has a molecular mass of higher than 300kDa. MUC1 which is also known as Mucin1 is a transmembrane glycoprotein. This type of mucin is present in all kind of epithelial cells and acts as a physical barrier (Lillehoj et al., 2013). MUC1 constitutes a tight mesh structure. Also, variable number tandem repeat region (VNTR), sperm protein-enterokinase-agarin (SEA) domain, transmembrane domain (TM) and cytoplasmic tail (CT) are the constituents of MUC1. The VNTR is heavily glycosylated and consists of a repeated amino acid sequence which is 20 amino acid in length and has the sequence of HGVTSAPDTRPAPGSTAPPA (Chen et al., 2021). The number of tandem repeats in this region varies between 20 to 125. Self-cleavage of MUC1 occurs in the SEA domain which have GSVVV motif. Through this cleavage, MUC1 extracellular domain is seperated from the cell surface, whereas TM and CT stay binded to surface of the cell. The TM region is consists of 28 amino acids and binds MUC1 to cell surface. Also, CT is located inside the cell and consists of 72 amino acids (Dhar & McAuley, 2019).

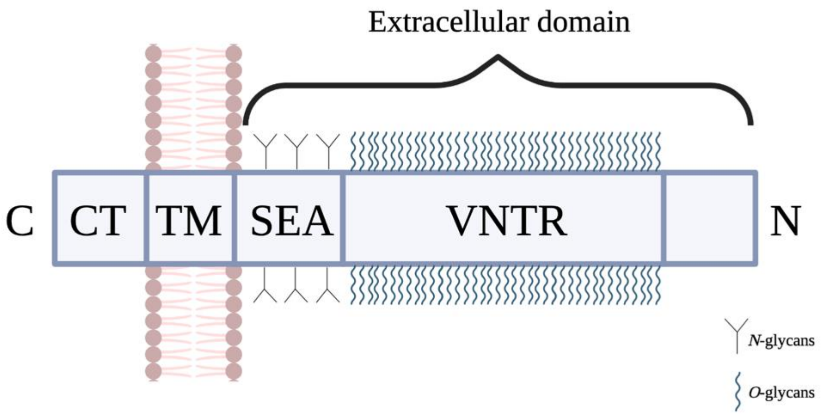


Figure 2.3 Schematic representation of MUC1 adapted from Dhar & McAuley (2019) and created with BioRender.com

Through the function of an extracellular sperm protein-enterokinase-agarin (SEA) domain, this barrier operates. The SEA domain of MUC1 plays a role on preventing epithelial cells from rupturing, harsh environmental conditions and immune attacks. Actioning as a decoy receptor against pathogens is another vital function of the SEA domain of MUC1 (Chen et al., 2021). Explicitly, with presence of extracellular oligosaccharides, MUC1 binds to pathogenic bacteria and prevents attainment of pathogens to cell surface. Hence, MUC1 has a crucial importance on protecting cells from infection which are caused by pathogens (Ke & Liu, 2014). Protective activity of MUC1 against foodborne pathogens is shown for several microorganisms (*H. pylori*, *C.jejuni*) previously (Lindén et al., 2009; McAuley et al., 2007). On the other hand, proinflammatory role of MUC1 for different foodborne pathogens (enteroaggregative *E. coli*, *S.* Enteritidis, *S.* Typhimurium) is also demonstrated (Boll et al., 2017; Li et al., 2019). Therefore, MUC1 is considered as having a dual function for pathogens (Bose & Mukherjee, 2020).

The importance of MUC1 arise from its activity against wide range of pathogens including *Salmonella* and its presence in different types of host tissues (Yuste, 2017). Therefore, MUC1 is the most comprehensively studied one among other mucins. However, most of the studies give information on role of MUC1 in cancer biology instead it's function in defense mechanism (Linden et al., 2008).

2.3 Host – *Salmonella* Interaction

Salmonella can enter the gut via three routes which are uptake by M cells, uptake by dendritic cells and penetration of epithelial cells of intestine (Fattinger et al., 2021). *Salmonella* enters the intestinal epithelial cells via its type 3 secretion system (T3SS) which is encoded by *Salmonella* Pathogenicity Island 1 (SPI-1) (Broz et al., 2012). After its entry to entrocytes, *Salmonella* uses T3SS which is encoded by *Salmonella* Pathogenicity Island 2 (SPI-2) for its survival intracellularly (Birhanu et al., 2018). The prerequisite of these mechanisms, which *Salmonella* uses to enter the intestinal epithelial cells and to persist in the host, is the initial adhesion of *Salmonella* to epithelial cells (Birhanu et al., 2018). *Salmonella* uses different fimbriae and adhesins during its adherence to epithelial cells (Fattinger et al., 2021). According to a recent research, the role of *Salmonella* adhesin SiiE was revealed as initial contact to polarized epithelial cells from the apical side (Barlag & Hensel, 2015). Therefore, *Salmonella* adhesin SiiE has a special importance during invasion of *Salmonella*, since *Salmonella* to grow, colonize and eventually cause infection is highly dependent to its initial attachment to host cells (Schaeffer, 1993).

2.3.1 SiiE – MUC1 Interaction

In 2015, a model which describes role of SPI – 4 of *S. enterica* on polarized cell invasion was proposed. According to the model, the function of SiiE was described as initial contact of polarized epithelial cells from their apical sides. The researchers also presented that the giant adhesin of *Salmonella* follows a zipper – like binding pattern due to highly repetetive domain structure of SiiE (Barlag & Hensel, 2015).

A more recent study which was published in 2019 revealed that for the first time, mucin MUC1 is a receptor for *Salmonella* adhesin SiiE. The researchers observed that mutant *Salmonella* strain which *siiE* was knocked out, did not have the ability of invading intestinal epithelial cells via mucin MUC1. Hence, this study highlighted that during SiiE regulated invasion of enterocytes from apical sides, MUC1 is compulsory and it is the receptor for giant adhesin SiiE of *Salmonella* (Li et al., 2019).

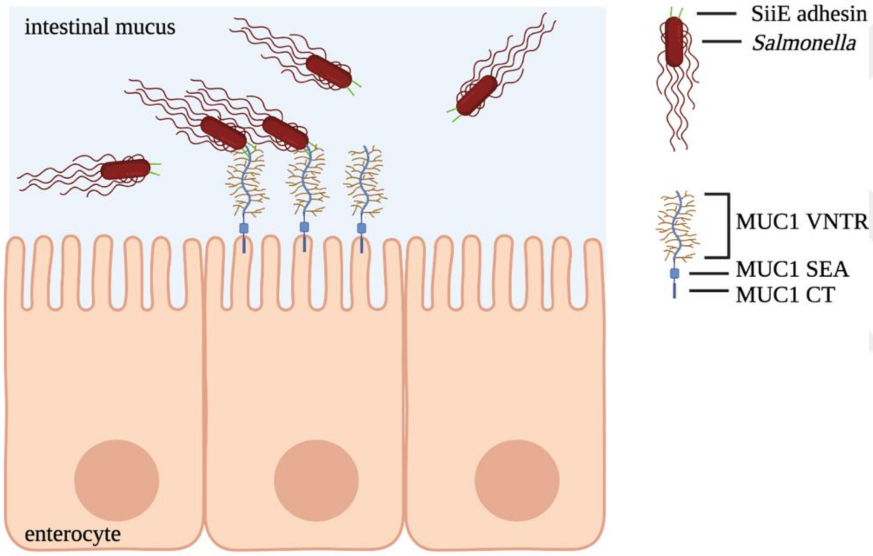


Figure 2.4 Proposed interaction of SiiE - MUC1 during intestinal epithelial cell invasion of *Salmonella* adapted from Li et al. (2019) and created with Biorender.com

2.4 Protein Modelling

Determination of protein's structure via experiments and modelling proteins computationally are intertwined disciplines. Both of them supply 3D information of proteins in atomic level, since proteins are polymers of amino acid residues (Mullins, 2012). X – Ray crystallography, nuclear magnetic resonance (NMR) and cryo –

electron microscopy (cryo – EM) are widely used methods when proteins are modelled experimentally (Dokholyan, 2020).

Table 2.6 Pros and cons of experimental protein structure modelling methods (Dokholyan, 2020)

Method	**Advantages**	**Disadvantages**
X – Ray crystallography	High resolution structures and large proteins	Need of high protein concentration, some crystallographic artifacts, only for stable proteins,
NMR	Multiple techniques for structure determination, ability to quantify dynamics and witness conformational dynamics of proteins	Need of high protein concentration, atomic assignment may be challenging
Cryo - EM	Large structures, ability to witness conformational states, faster workflow	Small proteins are a challenge

Although these methods have some noticeable superiorities, handicaps of these methods cannot be underestimated. Experimental methods of protein modelling are time consuming, remarkably expensive and sometimes unpracticable (Mullins, 2012). With the advance of DNA sequencing, 'omics' methods have been drawing exceptional attention. Computational modeling of proteins can be divided in two classes which are homology modeling and *ab initio* modelling. The former uses known structures of homologous proteins with the protein to be modelled which its structure is unknown, while the latter only uses primary structure of a protein and builds the model from scratch (Leach & Thomas, 2017). The significance of protein modelling can be explained with the prime dependence of proteins' function with their 3D structures which are contingent upon their amino acid residues. Also, proteins' sequences are less conserved than their structures (Agnihotry et al., 2022).

2.4.1 Protein Docking

Protein docking is regarded as molecular modelling. With the assistance of algorithms and methods of computer science, position and orientation of two molecules which form complex together is predicted. One of the constituents of the complex can be a small molecule like nucleic acid or it can be a protein, while the other one is a protein (Tradigo et al., 2019). Rigid body approach formes the baseline of many docking methods. Although there exists various methodologies, molecular dynamics (MD) and Monte Carlo which are based on energy minimization principle are examples of protein docking methodologies (Silakari & Singh, 2021). The importance of protein docking is that this discipline helps to specify protein – protein interface residues and to predict protein – protein interactions (PPIs) (Stanzione et al., 2021).

2.4.2 Protein - Protein Interactions

Secondary structures which are α – helix and β – sheet of a protein, are mainteined by hydrogen bonding. H - bond is formed between two amino acids when carboxyl group' oxygen atom is bonded to amino group' nitrogen. This hydrogen bonding play a role in stabilizing proteins' secondary structures. Although only covalent bonds (disulfide bridges) have an influence on stabilizing tetiary structures of proteins, ionic bonds and hydrophobic interactions are help to determine proteins' tertiary structure. Disulfide linkages occur due to interaction between cysteine residues of amino acids. As well, ionic bonds form between amino acids which have opposite electrical charges, while hydrophobic interactions arise due to nonpolar side chains of amino acid residues where lie interior core of a protein (Libretexts, 2022).

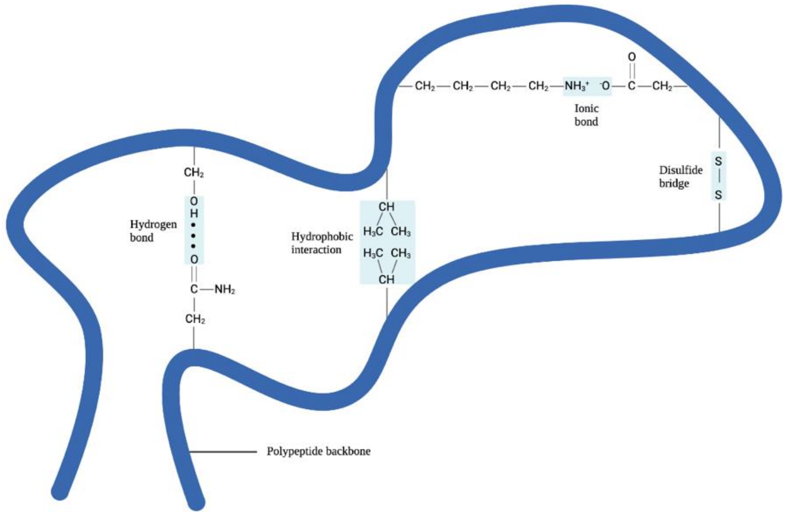


Figure 2.5 Protein-protein interactions (PPIs) adapted from Libretexts (2022) and created with BioRender.com

The importance of PPIs is that a protein will lose its function when the protein loses its 3D shape which is determined by the interactions explained above in detail (Libretexts, 2022).

Additionally, by investigating binding of two molecules, the interaction of the two molecules with each other can be interpreted. These two molecules can be a small molecule like nucleic acids or can be a bigger molecule such as proteins (Pollard, 2010). PPIs are responsible for the physical contact between two proteins (Farooq et al., 2021). Non-covalent interactions, except disulfide bridges which are covalent in characteristics, mediate the interaction between two proteins mostly (Soleymani et al., 2022).

CHAPTER 3

MATERIALS AND METHODS

3.1 Chemicals and Materials

A list of chemicals and materials which are meticulously selected for the experimental procedure can be seen in the Appendix (Table B.1).

3.2 Databases and Bioinformatic Tools

A list of databases and bioinformatic tools that are used in this study are presented in the Appendix (Table C.1).

3.3 *Salmonella* Isolates

3.3.1 Whole Genome Sequenced *Salmonella* Isolates

Salmonella enterica subsp. *enterica* (*Salmonella*) isolates which had already been isolated from Türkiye and whole genome sequenced by using Illumina MiSeq sequencing method can be seen in Table 3.1 with detailed information. These isolates were previously frozen with glycerol, stored in METU Food Safety Laboratory of Food Engineering Department at -80℃ and given ID codes. In total, 47 whole genome sequenced *Salmonella* isolates were used in this study.

Table 3.1 Whole genome sequenced *Salmonella* isolates

Isolate ID Code	BioProject Accession	BioSample Accession	Genus	Serovar	Source Type	Isolation Year
MET S1-24	PRJNA266622	SAMN03255391	*Salmonella*	Corvallis	Food	2012
MET S1-31	PRJNA266622	SAMN03255393	*Salmonella*	Salford	Food	2012

Table 3.1 (cont'd)

tMET S1-35	PRJNA266622	SAMN03255392	*Salmonella*	Corvallis	Food	2012
MET S1-63	PRJNA266622	SAMN03255404	*Salmonella*	Telaviv	Food	2012
MET S1-65	PRJNA266622	SAMN03255401	*Salmonella*	Montevideo	Food	2012
MET S1-74	PRJNA266622	SAMN03255405	*Salmonella*	Telaviv	Food	2012
MET S1-87	PRJNA266622	SAMN03255402	*Salmonella*	Othmarschen	Food	2012
MET S1-103	PRJNA266622	SAMN03255406	*Salmonella*	Infantis	Food	2012
MET S1-119	PRJNA266622	SAMN03255403	*Salmonella*	Telaviv	Food	2012
MET S1-163	PRJNA266622	SAMN03255394	*Salmonella*	Hadar	Food	2012
MET S1-248	PRJNA266622	SAMN03255330	*Salmonella*	Anatum	Food	2012
MET S1-258	PRJNA266622	SAMN03255331	*Salmonella*	Anatum	Food	2012
MET S1-272	PRJNA266622	SAMN03255332	*Salmonella*	Anatum	Food	2012
MET S1-277	PRJNA266622	SAMN03255340	*Salmonella*	Montevideo	Food	2012
MET S1-294	PRJNA266622	SAMN03255333	*Salmonella*	Anatum	Food	2012
MET S1-301	PRJNA266622	SAMN03255334	*Salmonella*	Anatum	Food	2012
MET S1-313	PRJNA266622	SAMN03255338	*Salmonella*	Kentucky	Food	2012
MET S1-360	PRJNA266622	SAMN03255339	*Salmonella*	Kentucky	Food	2012
MET S1-391	PRJNA266622	SAMN03255343	*Salmonella*	Telaviv	Food	2012
MET S1-408	PRJNA266622	SAMN03255347	*Salmonella*	Anatum	Food	2012
MET S1-410	PRJNA266622	SAMN03255345	*Salmonella*	Charity	Food	2012
MET S1-416	PRJNA266622	SAMN03255357	*Salmonella*	Chester	Food	2012
MET S1-421	PRJNA266622	SAMN03255348	*Salmonella*	Anatum	Food	2012
MET S1-430	PRJNA266622	SAMN03255359	*Salmonella*	Telaviv	Food	2012
MET S1-440	PRJNA266622	SAMN03255360	*Salmonella*	Telaviv	Food	2012

Table 3.1 (cont'd)

MET S1-443	PRJNA266622	SAMN03255349	*Salmonella*	Anatum	Food	2012
MET S1-456	PRJNA266622	SAMN03255361	*Salmonella*	Telaviv	Food	2012
MET S1-462	PRJNA266622	SAMN03255358	*Salmonella*	Reading	Food	2012
MET S1-465	PRJNA266622	SAMN03255350	*Salmonella*	Anatum	Food	2012
MET S1-482	PRJNA266622	SAMN03255351	*Salmonella*	Anatum	Food	2012
MET S1-512	PRJNA266622	SAMN03255352	*Salmonella*	Montevideo	Food	2012
MET S1-517	PRJNA266622	SAMN03255356	*Salmonella*	Montevideo	Food	2012
MET S1-528	PRJNA266622	SAMN03255363	*Salmonella*	Telaviv	Food	2012
MET S1-530	PRJNA266622	SAMN03255365	*Salmonella*	Telaviv	Food	2012
MET S1-548	PRJNA266622	SAMN03255372	*Salmonella*	Anatum	Food	2012
MET S1-557	PRJNA266622	SAMN03255382	*Salmonella*	Telaviv	Food	2012
MET S1-568	PRJNA266622	SAMN03255379	*Salmonella*	Newport	Food	2012
MET S1-581	PRJNA266622	SAMN03255376	*Salmonella*	Montevideo	Food	2012
MET S1-611	PRJNA266622	SAMN03255377	*Salmonella*	Montevideo	Food	2012
MET S1-625	PRJNA266622	SAMN03255380	*Salmonella*	Typhimurium	Food	2012
MET S1-634	PRJNA266622	SAMN03255378	*Salmonella*	Montevideo	Food	2012
MET S1-637	PRJNA266622	SAMN03255383	*Salmonella*	Telaviv	Food	2012
MET S1-647	PRJNA266622	SAMN03255384	*Salmonella*	Telaviv	Food	2012
MET S1-666	PRJNA266622	SAMN03255329	*Salmonella*	Mbandaka	Food	2012
MET S1-667	PRJNA266622	SAMN03255381	*Salmonella*	Newport	Food	2012
MET S1-670	PRJNA266622	SAMN03255387	*Salmonella*	Newport	Food	2012
MET S1-675	PRJNA266622	SAMN03255367	*Salmonella*	Kentucky	Food	2012

Except MET S1-24, MET S1-31, MET S1-35, MET S1-408, MET S1-410, MET S1-666, MET S1-670 and MET S1-675, the rest of the isolates above were isolated during TUBITAK Project No: 111O192 in 2012 (Acar et al., 2017).

3.3.2 Representative *Salmonella* Isolates

Representative *Salmonella* isolates were selected with the emphasis on clinical importance, PFGE types, isolation source and rareness. These representative isolates which can be seen in Table 3.2 in detail, had already been isolated from Türkiye, stored at -80℃ in METU Food Safety Laboratory of Food Engineering Department and given ID codes. In total 78 *Salmonella* isolates were selected as representative ones.

Table 3.2 Representative *Salmonella* isolates

Isolate ID Code	Genus	Serovar	Source Type	Isolation Year
MET S1-003	*Salmonella*	Virchow	Food	2005
MET S1-005	*Salmonella*	Nchanga	Food	2005
MET S1-008	*Salmonella*	Thompson	Food	2005
MET S1-014	*Salmonella*	Bispebjerg	Food	2005
MET S1-024	*Salmonella*	Corvallis	Food	2012
MET S1-031	*Salmonella*	Salford	Food	2012
MET S1-119	*Salmonella*	Telaviv	Food	2012
MET S1-166	*Salmonella*	Newport	Animal	2012
MET S1-170	*Salmonella*	Montevideo	Animal	2012
MET S1-172	*Salmonella*	Montevideo	Animal	2012
MET S1-175	*Salmonella*	Montevideo	Animal	2012
MET S1-184	*Salmonella*	Paratyphi B	Human	2012
MET S1-185	*Salmonella*	Typhimurium	Human	2012
MET S1-186	*Salmonella*	Paratyphi B	Human	2012
MET S1-187	*Salmonella*	Kentucky	Human	2012
MET S1-188	*Salmonella*	Paratyphi B	Human	2010

Table 3.2 (cont'd)

MET S1-217	*Salmonella*	Enteritidis	Human	2011
MET S1-220	*Salmonella*	Typhi	Human	2011
MET S1-221	*Salmonella*	Enteritidis	Human	2011
MET S1-223	*Salmonella*	Typhimurium	Human	2011
MET S1-227	*Salmonella*	Othmarschen	Human	2012
MET S1-234	*Salmonella*	Typhi	Human	2012
MET S1-313	*Salmonella*	Kentucky	Food	2012
MET S1-324	*Salmonella*	Montevideo	Food	2012
MET S1-372	*Salmonella*	Montevideo	Food	2012
MET S1-391	*Salmonella*	Telaviv	Food	2012
MET S1-392	*Salmonella*	Reading	Animal	2012
MET S1-394	*Salmonella*	Poona	Animal	2012
MET S1-396	*Salmonella*	Montevideo	Animal	2012
MET S1-397	*Salmonella*	Montevideo	Animal	2012
MET S1-398	*Salmonella*	Caracas	Animal	2012
MET S1-399	*Salmonella*	Telaviv	Animal	2012
MET S1-400	*Salmonella*	Montevideo	Animal	2012
MET S1-401	*Salmonella*	Telaviv	Animal	2012
MET S1-405	*Salmonella*	Kentucky	Animal	2012
MET S1-407	*Salmonella*	Telaviv	Animal	2012
MET S1-408	*Salmonella*	Anatum	Food	2012
MET S1-409	*Salmonella*	Mikawasima	Food	2012
MET S1-410	*Salmonella*	Charity	Food	2012
MET S1-411	*Salmonella*	Enteritidis	Food	2012
MET S1-416	*Salmonella*	Chester	Food	2012
MET S1-421	*Salmonella*	Anatum	Food	2012
MET S1-430	*Salmonella*	Telaviv	Food	2012
MET S1-439	*Salmonella*	Montevideo	Food	2012
MET S1-536	*Salmonella*	Kentucky	Animal	2012
MET S1-545	*Salmonella*	Telaviv	Animal	2012

Table 3.2 (cont'd)

MET S1-568	*Salmonella*	Newport	Food	2012
MET S1-581	*Salmonella*	Montevideo	Food	2012
MET S1-611	*Salmonella*	Montevideo	Food	2012
MET S1-625	*Salmonella*	Typhimurium	Food	2012
MET S1-634	*Salmonella*	Montevideo	Food	2012
MET S1-637	*Salmonella*	Telaviv	Food	2012
MET S1- 647	*Salmonella*	Telaviv	Food	2012
MET S1-653	*Salmonella*	Typhimurium	Animal	2012
MET S1-654	*Salmonella*	Anatum	Animal	2012
MET S1-657	*Salmonella*	Typhimurium	Animal	2012
MET S1-658	*Salmonella*	Montevideo	Animal	2012
MET S1-660	*Salmonella*	Enteritidis	Animal	2012
MET S1-662	*Salmonella*	Newport	Animal	2012
MET S1-670	*Salmonella*	Newport	Food	2012
MET S1-702	*Salmonella*	Telaviv	Animal	2012
MET S1-703	*Salmonella*	Hadar	Animal	2012
MET S1-742	*Salmonella*	Enteritidis	Food	2013
MET S1-868	*Salmonella*	Bredeney	Sludge	2015
MET S1-979	*Salmonella*	Muenster	Sludge	2015
MET S1-989	*Salmonella*	Paratyphi B	Sludge	2015
MET S1-998	*Salmonella*	Coeln	Sludge	2015
MET S1-999	*Salmonella*	Virchow	Sludge	2015
MET A2-003	*Salmonella*	Typhimurium	Sludge	2015
MET A2-012	*Salmonella*	Enteritidis	Sludge	2015
MET A2-022	*Salmonella*	Senftenberg	Sludge	2015
MET A2-36	*Salmonella*	Abony	Sludge	2016
MET A2-038	*Salmonella*	Richmond	Sludge	2016
MET A2-072	*Salmonella*	Kentucky	Sludge	2016
MET A2-078	*Salmonella*	Altona	Sludge	2016
MET A2-099	*Salmonella*	Liverpool	Food	2019

Table 3.2 (cont'd)

MET A2-230	*Salmonella*	Enteritidis	Cattle Farm	2021
MET A2-236	*Salmonella*	Typhimurium	Poultry Farm	2021

MET A2-099 was isolated during TÜBİTAK Project No: 118O770 in 2019 and MET A2-230 and MET A2-236 were isolated during TÜBİTAK Project No: 119O345 in 2021. Except MET S1-003, MET S1-005, MET S1-008, MET S1-014, MET S1-024, MET S1-031, MET S1-408, MET S1- 409, MET S1- 410, MET S1-411, MET S1-670, MET S1-868, MET S1-979, MET S1-989, MET S1-998, MET S1-999, MET A2-003, MET A2-012, MET A2-022, MET A2-036, MET A2-038, MET A2-72, MET A2-78 and the other two TÜBİTAK's projects' isolates, the rest of the isolates above were isolated during TUBITAK Project No: 111O192 in 2012 (Acar et al., 2017).

3.3.3 *Salmonella* Isolates from NCBI

SiiE amino acid sequences of different S*almonella* serovars were obtained from NCBI with selection criteria of having both most and least frequent observation rates in different continents among the world and isolation source (Sayers et al., 2021). Additional to the selection criteria, an emphasis on *Salmonella* serovars which are host – restricted, host – unrestricted and host – adapted, were also taken into consideration. Detailed information of the selected *Salmonella* serovars can be seen in Table 3.3. In total, SiiE amino acid sequences of 37 *Salmonella* isolates were obtained from NCBI for this study (Sayers et al., 2021).

Table 3.3 *Salmonella* isolates from NCBI

NCBI Accession	BioProject Accession	BioSample Accession	Genus	Serovar	Source Type	Isolation Year	Isolation Location
Host-unresticted							
QPQ67124.1	PRJNA478278	SAMN16987173	*Salmonella*	Typhimurium	Human	2018	Taiwan

Table 3.3 (cont'd)

QPE00769	PRJNA675428	SAMN16707726	*Salmonella*	Typhimurium	Human	1911	NA
UAD57192.1	PRJNA292661	SAMN20351460	*Salmonella*	Typhimurium	Food	2017	U.S.
QVB93509.1	PRJNA186035	SAMN02383573	*Salmonella*	Typhimurium	Environmental	2010	U.S.
UAF90589.1	PRJNA292666	SAMN07340353	*Salmonella*	Typhimurium	Animal	2017	U.S.
UKR48817	PRJNA803004	SAMN25608808	*Salmonella*	Infantis	Food	2012	Slovenia
QSD73693.1	PRJNA248792	SAMN08732049	*Salmonella*	Infantis	Human	2015	U.K.
UAF72254.1	PRJNA292666	SAMN07410879	*Salmonella*	Infantis	Animal	2017	U.S.
USZ56745.1	PRJNA852665	SAMN29340051	*Salmonella*	Enteritidis	Food	2019	China
UTL53925.1	PRJNA478278	SAMN29431220	*Salmonella*	Enteritidis	Human	2017	Taiwan
QVA53874.1	PRJNA186035	SAMN03165679	*Salmonella*	Enteritidis	Animal	2013	Belgium
QOT44377	PRJNA659233	SAMN15950485	*Salmonella*	Monophasic Typhimurium	NA	2020	U.S.
UAF12800.1	PRJNA292666	SAMN06256079	*Salmonella*	Derby	Animal	2016	U.S.
QVC76329.1	PRJNA186035	SAMN01813422	*Salmonella*	Derby	Environmental	2003	U.S.
UDP37548.1	PRJNA186035	SAMN10251118	*Salmonella*	Javiana	Food	2018	U.S.
QVC41389.1	PRJNA186035	SAMN01832085	*Salmonella*	Javiana	Human	2012	U.S.
QVB49741.1	PRJNA186035	SAMN02335376	*Salmonella*	Javiana	Environmental	2013	Mexico
QUY72529.1	PRJNA186035	SAMN03700983	*Salmonella*	Javiana	Animal	1973	U.S.
UJC60179.1	PRJNA706607	SAMN18138611	*Salmonella*	Agona	Food	2018	NA
UJA74055.1	PRJNA478278	SAMN24782770	*Salmonella*	Agona	Human	2020	Taiwan
UAE19190.1	PRJNA292666	SAMN05440588	*Salmonella*	Agona	Animal	2016	U.S.
QVB45244.1	PRJNA186035	SAMN02335402	*Salmonella*	Agona	Environmental	2012	Mexico
QVC85630.1	PRJNA186035	SAMN03464484	*Salmonella*	Weltevreden	Environmental	2012	Mexico
UDW11782.1	PRJNA292661	SAMN11967932	*Salmonella*	Saintpaul	Food	2018	U.S.

Table 3.3 (cont'd)

UDU71057.1	PRJNA773584	SAMN22488162	*Salmonella*	Saintpaul	Human	2014	China
QWJ55763.1	PRJNA186035	SAMN04961976	*Salmonella*	Saintpaul	Environmental	2012	U.S.
UNE82763.1	PRJNA785151	SAMN23543477	*Salmonella*	Muenchen	Human	2018	Israel
UAE09678.1	PRJNA292666	SAMN04964189	*Salmonella*	Muenchen	Animal	2016	U.S.
UAH66848.1	PRJNA292661	SAMN11968006	*Salmonella*	Muenchen	Food	2018	U.S.
QVY39512.1	PRJNA186035	SAMN02900230	*Salmonella*	Muenchen	Environmental	2013	U.S.
UAH62578.1	PRJNA292661	SAMN11967990	*Salmonella*	Bredeney	Food	2018	U.S.
Host- adapted							
QUZ71173.1	PRJNA186035	SAMN01933086	*Salmonella*	Choleraesuis	NA	NA	Switzerland
UAF35553.1	PRJNA292666	SAMN08115123	*Salmonella*	Dublin	Animal	2017	U.S.
WP_213892561.1	PRJNA224116	SAMN03700895	*Salmonella*	*bongori* 40:z35:-	NA	NA	NA
Host-restricted							
UQY61348.1	PRJNA269675	SAMN23378277	*Salmonella*	Gallinarum	NA	NA	NA
QUZ42214.1	PRJNA186035	SAMN01933155	*Salmonella*	Pullorum	NA	NA	Germany
PSC60333.1	PRJNA422033	SAMN08195664	*Salmonella*	Typhi	Human	2005	Pakistan

3.4 Screening of *siiE* Gene

Whole genome sequenced *Salmonella* isolates' (Table 3.1) FASTA files of *siiE* gene nucleotide sequences were obtained from NCBI (Sayers et al., 2021). By using MEGAX software, these isolates' multiple sequence alignment was conducted and phylogenetic tree was constructed (Stecher et al.,2020). Then, two sets of primers which one set for the beginning of the gene and one set for the end of the gene, were designed by using Primer-BLAST (Ye et al., 2012). Also, specificity and binding sites of the primer sets were double checked by using SnapGene Viewer ("*SnapGene*

Viewer", 2021). Finally, primer-dimer formation of the two primer sets was checked with Multiple Primer Analyzer ("*Multiple primer analyzer"*, n.d.). Detailed information of the two designed primer sets is represented in Table 3.4.

Table 3.4 Properties of the designed two primer sets for *siiE* gene

Nomenclature	Location in the Gene	Sequence 5' to 3'	Length	Melting Temperature	GC Content	Product Length
siiE – I - F	96.bp – 114. bp	TTCCGGGCCAGACATGAAC	19 bp	60℃	57.89%	864 bp
siiE – I - R	940.bp – 959.bp	CCATCAGCAGTCTTGGCCTC	20 bp	60.74℃	60%	
siiE – L - F	15656.bp – 15675.bp	TCTTTACGCCAGGTACACCG	20 bp	59.76℃	55%	583 bp
siiE – L - R	16218.bp – 16238.bp	TCCTTCACCACGCTTTCTTCC	21 bp	60.54℃	52.38%	

To investigate presence of *siiE* gene of *Salmonella* isolates in molecular manner, multiplex PCR was performed. Firstly, the representative *Salmonella* isolates (Table 3.2) which were obtained from -80℃ in METU Food Safety Laboratory of Food Engineering Department were streaked to Brain Heart Infusion (BHI) agars and incubated for 24 hours at 37℃. Next, 95 µL double deionized water (ddH_2O) which had been previously autoclaved were added to 0.2 mL PCR tubes. Afterwards, one colony for each isolate was transferred to 0.2 mL PCR tubes which then microwaved for 30 seconds at 90 V to lyse the cells. Next, master mix which components were presented in Table 3.5 and volume of the components was set according to number of samples, was prepared in an Eppendorf tube and pipetted to new 0.2 mL PCR tubes as each contain 24 µL master mix.

Table 3.5 Multiplex PCR's master mix components for *siiE* gene

Reagent	**Volume (µL)**
ddH_2O	16.5
5xMyTaq Reaction Buffer	5
siiE – I - F	0.5
siiE – I - R	0.5
siiE – L - F	0.5
siiE – L - R	0.5
MyTaq Red DNA Polymerase	0.5

Subsequently, 1 µL of lysed cells are transferred to 24 µL master mix containing PCR tubes. Then, these tubes are put into T100™ Thermal Cycler which thermal cycling conditions are shown in Table 3.6 (BioRad).

Table 3.6 Thermal cycling conditions for multiplex PCR of *siiE* gene

Temperature (℃)	**Time**	**Cycles**
95	1 min	1 x
95	15 sec	34 x
60.5	15 sec	
72	10 sec	
72	5 min	1 x
4	∞	1 x

After PCR products were taken out from the Thermal Cycler, 5 µL of them were loaded to wells of gel which the gel had been prepared beforehand by using 0.5 X Tris Borate EDTA (TBE) Buffer and 1.5% agarose. The gel was submerged in electrophoresis tank which contains 25 mL TBE and 475 mL ddH_2O, after PCR ladder and as a positive control MET S1 – 647 were also loaded as 5 µL to the wells. The gel electrophoresis was run for 50 minutes at 110 V and 400 mA. When the run was finished, the gel was submerged in EtBr solution which contains 0.002 mg EtBr/mL ddH_2O, for 4 minutes as staining step. For destaining, the gel was submerged into ddH_2O for 40 minutes. Finally, the gel was photographed under UV light to obtain the results (BioRad-Gel Doc XR Documentation System, USA).

3.5 Phylogenetic Analysis of *Salmonella* SiiE Protein

Firstly, multiple sequence alignment of bacterial immunoglobulin-like (BIg) domains of SiiE from 50th to 52nd (SiiE BIg50:52) of *Salmonella* serovars which were obtained from NCBI (Table 3.3) and METU whole genome sequenced ones (Table 3.1) was conducted by MEGAX (Sayers et al., 2021; Stecher et al.,2020). After phylogenetic tree was constructed, each cluster of *Salmonella* serovars was assigned to a group number. Then multiple sequence alignment of groups which were 20 in total was conducted by MEGAX (Stecher et al.,2020). To determine amino acid substitutions, the group which consists of *Salmonella* Typhimurium mostly was selected as a reference. Finally, the substitutions of amino acids of SiiE Blg50:52 was determined between the reference group and every other group according to the multiple sequence alignment result.

3.6 Determination of SiiE – MUC1 Interactions

3.6.1 Secondary Structures of SiiE and MUC1

PDB format of *S*. Typhimurium LT2 SiiE BIg50:52 (2YN5 and 2YN3) and *Homo sapiens* MUC1 SEA Domain (6BSC) were obtained from Protein Data Bank (PDB) (Burley et al., 2020). The detailed information of experimentally resolved structures of these proteins can be seen in Table 3.7.

Table 3.7 Experimentally resolved secondary structures of SiiE BIg50:52 and MUC1 SEA domain

Modelled Protein	PDB ID	Method	Length (# of amino acids)	References
SiiE	2YN5	X-Ray Diffraction (1.85 Å)	288	(Griessl et al., 2013).
SiiE	2YN3	X-Ray Diffraction (2.12 Å)	288	(Griessl et al., 2013).
MUC1	6BSC	X-Ray Diffraction (1.30 Å)	57	(Noguera et al., 2020)

Also, secondary structure predictions of *Homo sapiens* (human) MUC1 and *Sus scrofa* (swine) Muc1 were obtained from AlphaFold in PDB format for further analysis (Varadi et al., 2021). The detailed information of AlphaFold prediction of proteins of interest are shown in Table 3.8.

Table 3.8 Secondary structure predictions of Mucin 1 proteins

Modelled Protein	Source Organism	AlphaFold ID	Length (# of amino acids)
MUC1	*Homo sapiens* (human)	AF-P15941-F1	1255
Muc1	*Sus scrofa* (swine)	AF-A0A4X1W201-F1	810

3.6.2 Secondary Structure Prediction and Structure Analysis of SiiE from *Salmonella* Isolates

FASTA files of SiiE BIg50:52 of METU whole genome sequenced *Salmonella* isolates (Table 3.1) and also FASTA files of SiiE BIg50:52 of the selected *Salmonella* serovars obtained from NCBI which are *S. bongori* 40:z35:-, *S.* Gallinarum, *S.* Pullorum, *S.* Choleraesuis and *S.* Enteritidis which have NCBI accession WP_213892561.1, UQY61348.1, QUZ42214.1, QUZ71173.1 and

USZ56745.1 respectively, were submitted to I-TASSER Protein Structure and Function Prediction Server to obtain secondary structures of these isolates (Yang & Zhang, 2015). The secondary structure predictions were obtained in PDB format for further analysis.

Also, to structurally compare experimentally resolved SiiE BIg50:52 (2YN5 and 2YN3) and predicted secondary structures of SiiE BIg50:52 of whole genome sequenced METU *Salmonella* isolates and the selected *Salmonella* isolates obtained from NCBI, and to determine best superimposed chains of experimentally resolved SiiE BIg50:52 with these predicted secondary structures of SiiE BIg50:52 of *Salmonella* isolates, MatchMaker tool of Chimera was used (Meng et al., 2006). After structural alignment, Match -> Align tool of Chimera was also used to determine RMSD value (Meng et al., 2006).

3.6.3 Protein – Protein Docking of SiiE and MUC1

Firstly, AlphaFold prediction of MUC1 (MUC1(AF-P15941-F1)) and experimentally resolved structures of SiiE BIg50:52 (2YN5 and 2YN3) were docked. Then, AlphaFold prediction of MUC1 and best superimposed chains of experimentally resolved structures of SiiE BIg50:52 (2YN5 Chain B and 2YN3 Chain C) were docked. After, AlphaFold prediction of MUC1 and whole genome sequenced METU *Salmonella* isolates SiiE BIg50:52 (WGS METU SiiE BIg50:52) were docked, docking was performed between AlphaFold prediction of MUC1 and *Salmonella* serovars' SiiE BIg50:52 which were obtained from NCBI (NCBI SiiE BIg50:52). Then, docking was performed between experimentally resolved structures of MUC1 (MUC1(6BSC)) and SiiE BIg50:52. After, experimentally resolved structure of MUC1 and whole genome sequenced METU *Salmonella* isolates SiiE BIg50:52 were docked, docking was performed between experimentally resolved structure of MUC1 and *Salmonella* serovars SiiE BIg50:52 which were obtained from NCBI. All the above docking was conducted by using ClusPro Protein – Protein Docking Server by submitting related PDB files (Kozakov et al., 2017). Protein – protein docking scheme of SiiE BIg50:52 and MUC1 can be seen in Table 3.9.

Table 3.9 Protein - protein docking scheme of SiiE BIg50:52 and *Homo sapiens* MUC1

Docked Protein	Docked Protein
SiiE BIg50:52 (2YN5 and 2YN3)	MUC1 (AF-P15941-F1)
SiiE BIg50:52 Chains (2YN5 Chain B and 2YN3 Chain C)	MUC1 (AF-P15941-F1)
WGS METU SiiE BIg50:52	MUC1 (AF-P15941-F1)
NCBI SiiE BIg50:52	MUC1 (AF-P15941-F1)
SiiE BIg50:52 (2YN5 and 2YN3)	MUC1 (6BSC)
WGS METU SiiE BIg50:52	MUC1 (6BSC)
NCBI SiiE BIg50:52	MUC1 (6BSC)

To furtherly investigate the binding characteristics between SiiE BIg50:52 and MUC1, docking was performed between *Sus scrofa* Muc1 (Muc1(AF-A0A4X1W201-F1) and SiiE BIg50:52 which includes experimentally resolved structures of SiiE BIg50:52, best superimposed chains of experimentally resolved SiiE BIg50:52, MET S1-625's SiiE BIg50:52 and *Salmonella* serovars' SiiE BIg50:52 which were obtained from NCBI. All the above docking was conducted by using ClusPro Protein – Protein Docking Server by submitting related PDB files (Kozakov et al., 2017). Protein – protein docking scheme of SiiE BIg50:52 and Muc1 can be seen in Table 3.10.

Table 3.10 Protein - protein docking scheme of SiiE BIg50:52 and *Sus scrofa* Muc1

Docked Protein	Docked Protein
SiiE BIg50:52 (2YN5 and 2YN3)	Muc1 (AF-A0A4X1W201-F1)
SiiE BIg50:52 Chains (2YN5 Chain B and 2YN3 Chain C)	Muc1 (AF-A0A4X1W201-F1)
MET S1-625 SiiE BIg50:52	Muc1 (AF-A0A4X1W201-F1)
NCBI SiiE BIg50:52	Muc1 (AF-A0A4X1W201-F1)

3.6.4 Prediction of Protein – Protein Interactions of SiiE and MUC1

According to the SiiE BIg50:52 – MUC1 docking results, the protein – protein interactions were further investigated by using PyMOL (Schrödinger & Delano, 2021). For determining disulfide bridges, cysteine residues were selected. Then, to determine salt bridges, negatively charged amino acids and positively charged amino acids which are aspartate, glutamate and lysine, arginine, histidine respectively were selected separately and bonding distance was limited to 4 Å. Afterwards, polar contacts were found and hydrogen bonding was determined. Then, for determining hydrophobic interactions, hydrophobic amino acids which are alanine, glycine, valine, isoleucine, leucine, phenylalanine and methionine were selected and bonding distance was limited to 4 Å. Finally, amino acids which are responsible for the related bonding were determined among the protein – protein interface. The PyMOL commands which were used to determine protein – protein interactions are shown in Table 3.11.

Table 3.11 PyMOL commands for determining protein - protein interactions

Interaction Type	Command
Disulfide bridge	sel cysteines, resn cys
	select negative, (resn asp+glu)
Salt bridge	select positive, (resn lys+arg+his)
	distance saltbridge, negative, positive, 4.0, 0
Hydrophonic interaction	select hydrophobes, (resn ala+gly+val+ile+leu+phe+met)
	distance int, hydrophobes, hydrophobes, 4.0, 0

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Prevalence of *siiE* Gene

To investigate presence of *siiE* gene of *Salmonella*, 2 sets of primers were designed by using whole genome sequenced METU *Salmonella* isolates' *siiE* gene nucleotide sequences which were obtained from NCBI (Sayers et al., 2021). In total, 47 whole genome sequenced METU *Salmonella* isolates where their detailed information can be seen in Table 3.1 were used for this purpose in this study. By using MEGAX software, these 47 *Salmonella* isolates' multiple sequence alignment was conducted and phylogenetic tree was constructed where the former can be seen in the Appendix (Figure D.1) and the latter can be seen in Figure 4.1 in order to observe conserved regions of the *siiE* gene since, the goal was designing two primer sets which have the ability to bind different serovars comprehensively (Stecher et al.,2020).

When multiple sequence alignment was conducted with MEGAX, ClustalW system was used. The reason of choosing this system was that ClustalW is a convenient system when dealing with homologous sequences and it uses progressive alignment methods when aligning multiple sequences. With the algorithm of ClustalW, firstly the most similar sequences are aligned then sequences that are more distant are aligned progressively until alignment is achieved globally ("About ClustalW", n.d.). When time and memory constraints are taken into consideration, this heuristic approach of ClustalW is very practical since *siiE* gene is consists of 5559 amino acids and these 47 different isolates' *siiE* gene nucleotide sequences are approximately same in length and share few amino acid substitutions among them. Moreover when phylogenetic tree was constructed, neighbor – joining method was preferred. The method based on the principle on minimum evolution and constructs phylogenetic trees with evolutionary distance data. Neighbor – joining method minimizes total length of branches at every stage of clustering of operational taxonomic units (OTUs / neighbors), for detecting pairs of neighbors. This algorithm of neighbor – joining method saves unignorable amount of time (Saitou & Nei, 1987). When the number of *siiE* nucleotide sequences and their size are taken into consideration, using neighbor – joining method to construct the phylogenetic tree is very useful and time – saving. In addition, Kimura two – parameter (K2P) model was used when the phylogenetic tree was contructed. K2P model is a commonly used model for evoluating genetic distances of nucleotide sequences with considering transitional and transversional rates of substitutions (Nishimaki & Sato, 2019; Kimura, 1980).

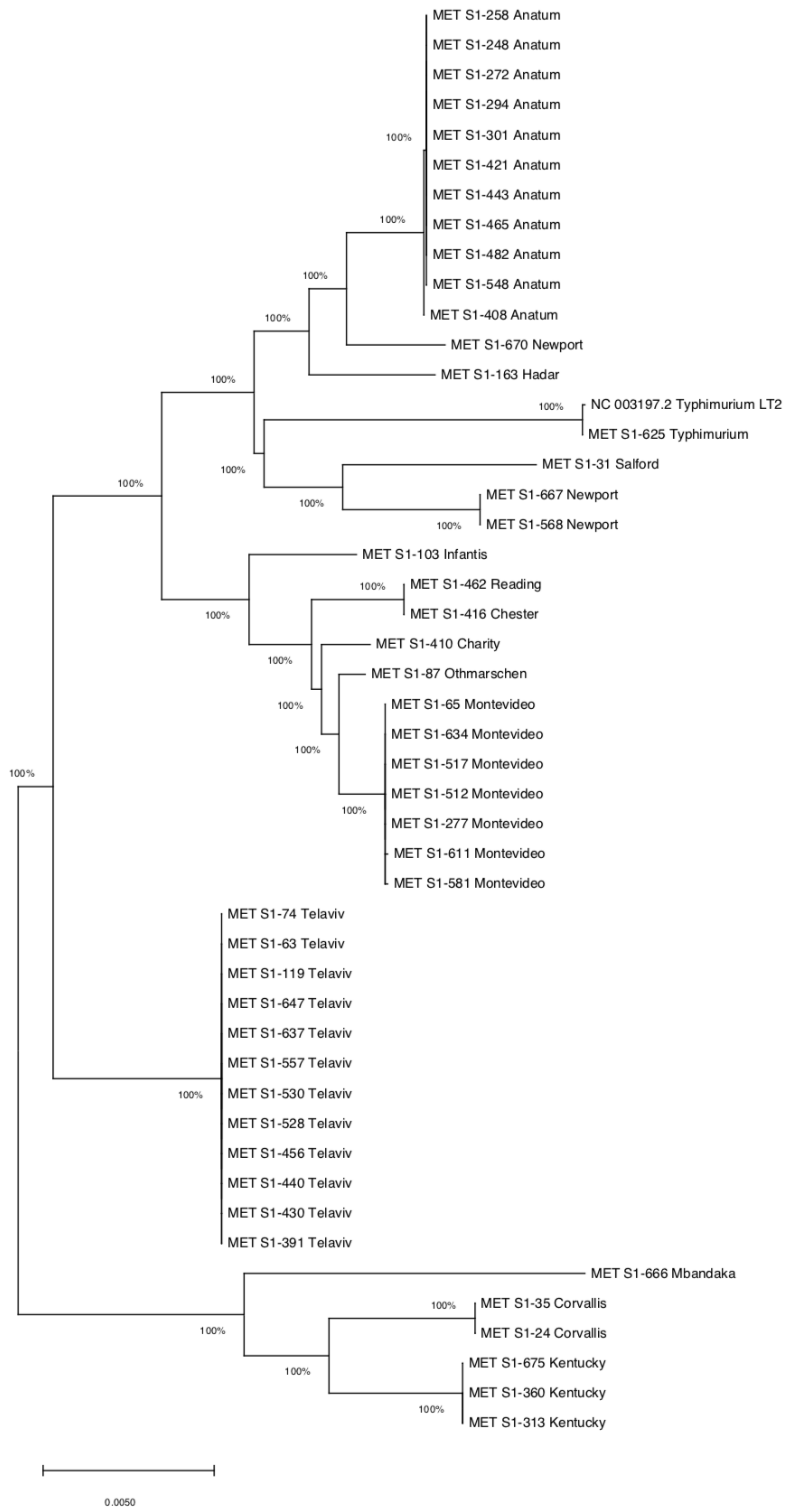


Figure 4.1 Phylogenetic tree of whole genome sequenced METU *Salmonella* Isolates with respect to their *siiE* gene nucleotide sequences

According to Figure 4.1, same type of *Salmonella* serovars constitutes the clusters of the phylogenetic tree. Therefore, *siiE* gene of *Salmonella* is well conserved within the same *Salmonella* serovars.

After conserved regions of whole genome sequenced METU *Salmonella* isolates' *siiE* gene were detected according to the multiple sequence alignment result, two sets of primers were designed by using Primer-BLAST (Ye et al., 2012). These primer sets' binding sites for whole genome sequenced METU *Salmonella* isolates were highlighted in the multiple sequence alignment which were shown in the Appendix (Figure E.1 - Figure E.4). Also, specificity and binding sites of the primer sets which can be seen in Figure 4.2, were double checked by using SnapGene Viewer ("*SnapGene Viewer",* 2021). Finally, primer-dimer formation of the two primer sets was checked with Multiple Primer Analyzer ("*Multiple primer analyzer"*, n.d.). Detailed information of the two designed primer sets was represented in Table 3.4. Also by designing two primer sets rather than one, detection of both beginning and end part of the gene was attained which helped to infer its expression and visualization in the gel was achieved.

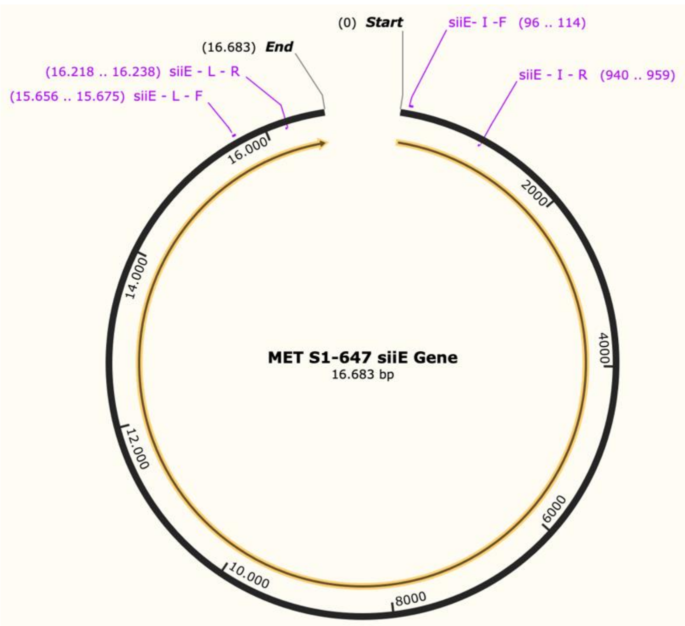


Figure 4.2 SnapGene Viewer representation of binding sites of *siiE*-I-F, *siiE*-I-R, *siiE*-L-F and *siiE*-L-R for MET S1-647

To symbolise *Salmonella* diversity extensively, representative *Salmonella* isolates where their detailed information can be seen in Table 3.2 were selected with the emphasis on clinical importance, PFGE types, isolation source and rareness. In this study, in total 78 *Salmonella* isolates which had been isolated and belonged to METU Food Safety Laboratory were selected as representative ones. For investigating presence of *siiE* gene of *Salmonella* isolates in molecular manner, multiplex PCR was performed with proper thermal cycling conditions (see Table 3.6).

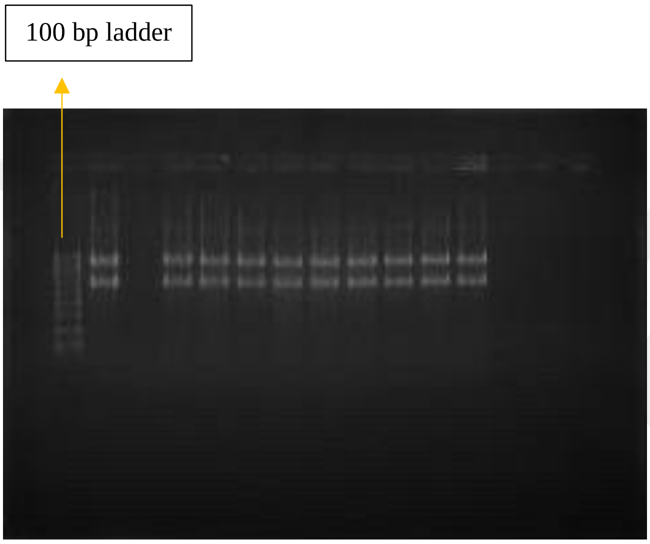


Figure 4.3 Multiplex PCR gel electrophoresis image of *Salmonella* isolates. L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-409; 4: MET S1-979; 5: MET S1-005; 6: MET S1-394; 7: MET S1-392; 8: MET A2-022; 9: MET A2-038; 10: MET S1-31; 11: MET S1-008

There are two bands present in the multiplex PCR gel electrophoresis image. The upper band has product length of 864 bp, while the lower band has the product length of 583 bp. The former amplifies the beginning part of *siiE* gene, despite the latter amplifies the end part of the gene. The rest of the representative *Salmonella* isolates' multiplex PCR gel electrophoresis images are shown in the Appendix (Figures F.1). It was seen that all the isolates that were undergone multiplex PCR, have *siiE* gene. Hence, it can be clearly concluded that *siiE* gene is highly prevalent among different *Salmonella* serovars.

Similarly, Wang and his colleagues (Wang et al., 2020b), conducted characterization of *Salmonella* which were isolated from chickens in central China and investigated virulence associated genes of them with PCR. The researchers observed that *siiE* gene is present in all the 84 isolated *Salmonella* isolates. Another research which distribution of virulence genes of *Salmonella* Pullorum isolates from chickens in China between 1953 and 2015 were screened by PCR, also detected *siiE* gene in all the 304 *Salmonella* Pullorum isolates (Zhang et al., 2018). Moreover, when in total 138 *Salmonella* Typhimurium isolates which were isolated from chickens in China between 2007 and 2012 were investigated for the presence of 30 different virulence genes, the researchers concluded the most frequent detected gene as *siiE* (Li et al., 2017). In addition, Qiao and his colleagues (Qiao et al., 2018) were determined 30 virulence genes of extended - spectrum β – lactamases (ESBLs) – producing *Salmonella* and the researchers pointed out that *marT* (a transcriptional regulator) was the most frequent detected gene which was followed by *siiE*. Finally, high conservation of *siiE* among different *Salmonella* serovars was highlighted according to bioinformatic studies (Latasa et al., 2005). These researchers' findings corroborate with the prevalent occurrence of representative METU *Salmonella* isolates' *siiE* gene.

4.2 Amino Acid Substitutions of *Salmonella* SiiE BIg50:52

According to Europen Union One Health 2020 Zoonoses Report and a research where worlwide epidemiology of *Salmonella* serovars were investigated, *Salmonella* serovars which have both the most and the least frequent observation rates in different continents among the world were determined (EFSA, 2020; Ferrari et al., 2019). Additional to the observation incidence, an emphasis on *Salmonella* serovars which are host – restricted, host – unrestricted and host – adapted, were also taken into consideration. In total, 37 *Salmonella* serovars' where their detailed information can be seen in Table 3.3, SiiE (*Salmonella* intestine infection) amino acid sequences were obtained from NCBI for this study (Sayers et al., 2021).

Protein domains that resemble immunoglobulin (Ig) domains in terms of amino acid sequence and structure are named as immunoglobulin (Ig)–like domain. These Ig-like domains are consist of approximately 100 amino acids and their structure involves β-sheets (Horstkorte & Fuss, 2012). Moreover, bacterial Ig-like (BIg) domains which play a role in different processes such as biofilm formation and adhesion, can be extensively observed in bacteria (Hüttener et al., 2019). In other words, BIg domains of bacteria have a special influence on molecular recognition which includes binding properties (Bodelón et al., 2013). The giant adhesin SiiE of *Salmonella* has 53 repetetive bacterial immunoglobulin-like (BIg) domains (Griessl et al., 2013). According to a research where phylogenetic analysis of BIg domains of SiiE were conducted revealed that domains 51 and 52 can be regarded as interpreters of other SiiE domains, while domain 50 can be considered as a unique one. Hence, BIg domains of SiiE from 50^{th} to 52^{nd} (SiiE BIg50:52) which lie between residues from 5078^{th} to 5365^{th} and consist of 288 amino acids in total, have a special importance to gain an understanding on whole SiiE domains (Griessl et al., 2013). Therefore, in this study SiiE BIg50:52 of *Salmonella* were furtherly investigated to highlight amino acid variations among different *Salmonella* serovars and to gain insight of these variations influence on binding to SiiE's host receptor mucin 1 (MUC1).

By using MEGAX, multiple sequence alignment of SiiE BIg50:52 of *Salmonella* serovars which were obtained from NCBI and METU whole genome sequenced ones was conducted (Stecher et al.,2020). The multiple sequence alignment result can be seen in the Appendix (Figure G.1). Then phylogenetic tree which is shown in Figure 4.4 was constructed by using MEGAX (Stecher et al.,2020). For the MSA and construction of the phylogenetic tree, ClustalW system and neighbor – joining method which the reason of choosing these algorithms was explained in detail previously were used respectively.

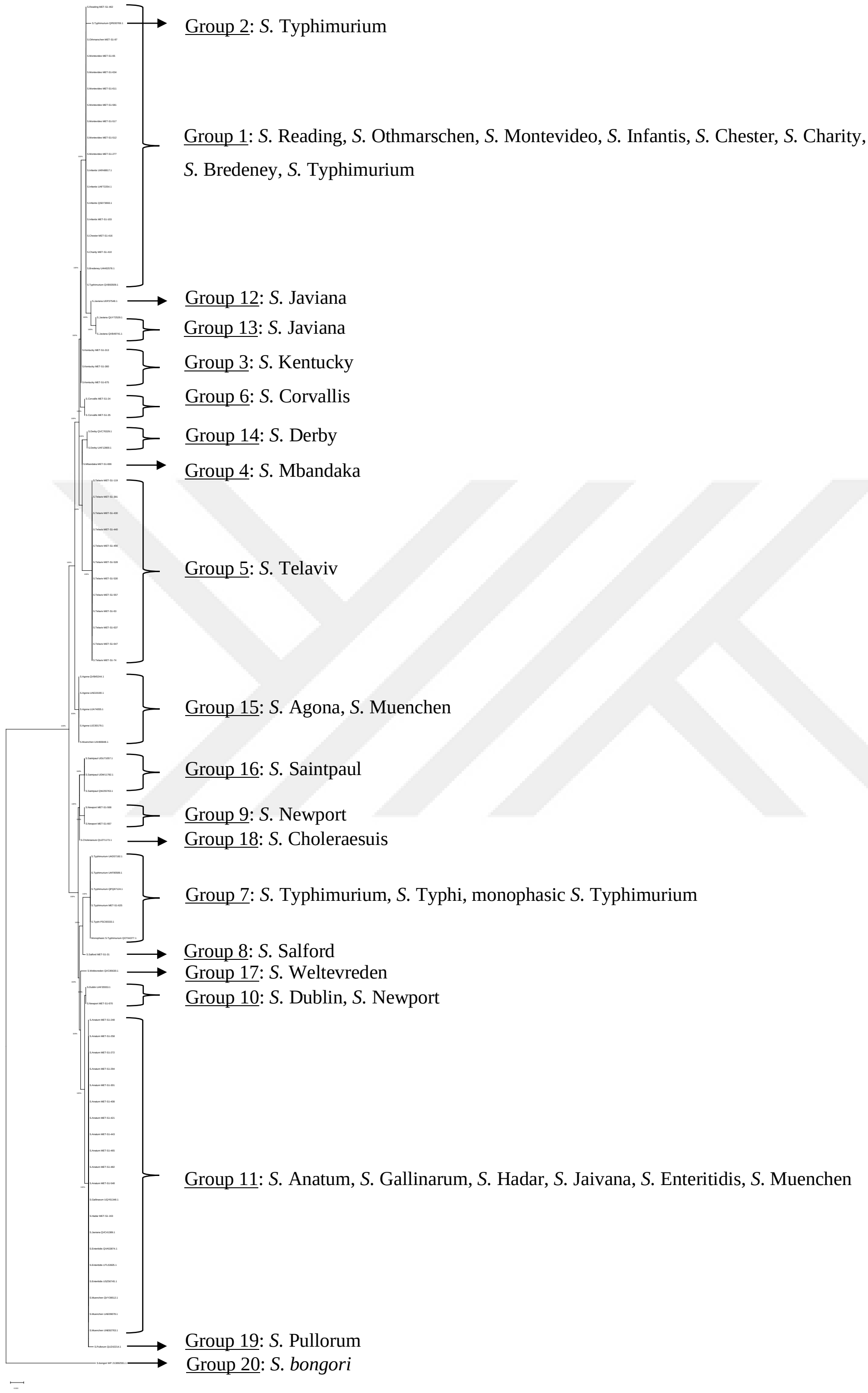


Figure 4.4 Phylogenetic tree of SiiE Blg50:52 of *Salmonella* serovars which were obtained from NCBI and METU whole genome sequenced ones

According to Figure 4.4, there are two main branches of the phylogenetic tree where one branch is furtherly subdivided to *enterica* species and the other to *bongori* species. When examined in more detail, there exist 20 different clusters which most of them have the pattern of containing same kind of *Salmonella* serovars. For example, *S*. Telaviv constitute one cluster while, *S*. Newport constitute another. This motif of the phylogenetic tree was an expected result since, genus *Salmonella* consists of 2 species which are furtherly categorized in serovar level.

After each cluster of the phylogenetic tree was assigned to a group number, multiple sequence alignment of the groups which are 20 in total was conducted by MEGAX with ClustalW system which the reason of choosing this algorithm was explained in detail previously (Stecher et al.,2020). The detailed information of the groups and multiple sequence alignment result of the groups can be seen in the Appendix (Table G.1 and Figure G.2 respectively). Then, group number 7 which consists of *Salmonella* Typhimurium LT2 mostly was selected as a reference and amino acids substitutions of SiiE Blg50:52 were determined between the reference group and every other group according to the multiple sequence alignment result. Incidence of amino acid substitutions of SiiE BIg50:52 is represented in Table 4.1. For example, the amino acid substitution on residue 5081 of SiiE was only observed in group 20, while for residue 5113 the amino acid substitution was observed in every group except group 20. Also, the positions of these substitutions are highlighted for MET S1-625's SiiE BIg50:52 which is shown in Figure 4.5.

Table 4.1 Incidence of amino acid substitutions of SiiE BIg50:52

SiiE BIg 50:52 AA Location	**Whole SiiE AA Location**	**Substitution***	**Incidence Among Groups**	**Group Number of Observed AA Substitution In**
4	5081	N - G	1	20
36	5113	E - N	18	1,2,3,4,5,6,8,9,10,11,12,13,14,15,16,17,18,19
41	5118	T - M	1	17
44	5121	V - A	1	19
49	5126	K - N	1	20
52	5129	M - T	1	20
70	5147	I - M	1	20
73	5150	R - K	1	20

Table 4.1 (cont'd)

76	5153	D - V	1	20
78	5155	P - S	4	5,11,19,20
84	5161	E - D	1	20
85	5162	V - T	8	1,2,3,4,5,6,12,14
91	5168	S - A	1	20
92	5169	P - V	1	20
97	5174	G - N	1	20
98	5175	T - I	1	20
99	5176	V - A	3	4,5,14
102	5179	N - D	1	9
105	5182	R - Q	1	20
107	5184	K - N	10	1,2,3,4,5,12,13,14,15,20
108	5185	Q - K	1	20
122	5199	V – I & V - T	1 & 1	15 & 20
131	5208	V - L	1	20
136	5213	K - Q	1	20
137	5214	N - R	1	20
141	5218	V - I	1	2
149	5226	P - A	1	20
151	5228	A - T	1	16
168	5245	Q - K	1	20
194	5271	D - G	4	1,2,12,13
197	5274	A - V	5	10,11,12,13,19
227	5304	V - I	1	20
230	5307	T - A	1	20
232	5309	N - S	1	20
234	5311	V - G	1	20
237	5314	I - T	7	5,8,10,11,17,19,20
243	5320	G - S	1	20
259	5336	N - T	18	1,2,3,4,5,6,9,10,11,12,13,14,15,16,17,18,19,20
269	5346	R - D	1	20
273	5350	S - K	1	20
274	5351	Q – L & Q - K	9 & 1	1,2,3,4,5,6,12,13,15 & 20
275	5352	Q - H	1	20
280	5357	A – E & A - T	10 & 1	1,2,3,4,5,6,12,13,14,15 & 20

*: N: Aspargine, G: Glycine, E: Glutamate, T: Threonine, M: Methionine, V: Valine, A: Alanine, K: Lysine, I: Isoleucine, R: Arginine, D: Aspartate, P: Proline, S: Serine, Q: Glutamine, L: Leucine, H: Histidine

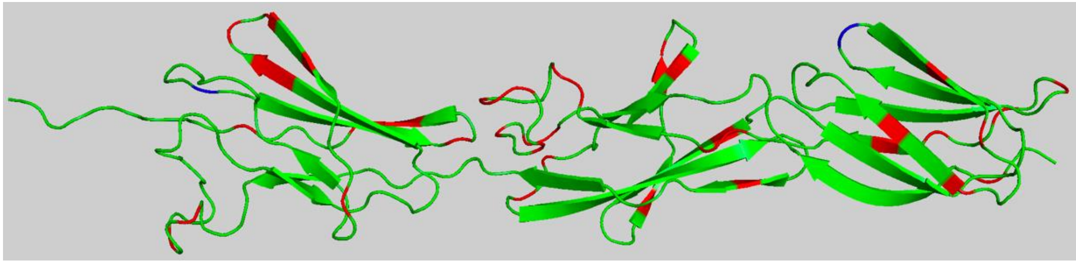


Figure 4.5 Representation of positions of amino acid substitutions (red) for MET S1-625's SiiE BIg50:52. (blue: the most frequently observed substitutions)

According to incidence of amino acid substitutions of SiiE BIg50:52, it can be clearly understood that the most frequently observed substitutions are on residues 5113 and 5336 where *S*. Typhimurium's glutamate (E) is substituted to aspargine (N) and aspargine (N) is substituted to threonine (T) respectively. The most frequent substitutions are followed by residues 5184, 5351 and 5357. Detecting these amino acid substitutions have a special significance since, protein – protein interactions (PPIs) occur when particular types of amino acids interact at a specified distance (Libretexts, 2022). For instance, SiiE of *S*. Typhimurium's residue 5113 is glutamate while for different serovar it is aspargine. The former is a negatively charged amino acid while, the latter is a polar one. These types of residue substitutions affect the type of PPIs and even can be resulted in interaction loss.

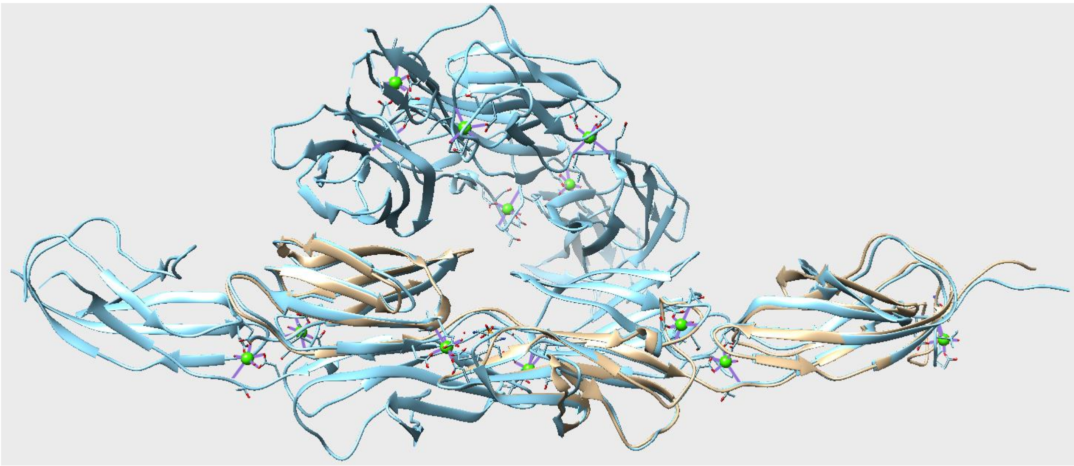
4.3 Protein – Protein Interactions Between SiiE and MUC1

To obtain secondary structure predictions of SiiE BIg50:52 of METU whole genome sequenced *Salmonella* isolates and also *S. bongori* 40:z35:-, *S*. Gallinarum, *S*. Pullorum, *S*. Choleraesuis and *S*. Enteritidis' FASTA files were submitted to I-TASSER Protein Structure and Function Prediction Server (Yang & Zhang, 2015).

I – TASSER (Iterative Threading ASSEmbly Refinement) is a server which builds 3D structures of proteins from their amino acid sequences. It's hierarchical approach firstly matches the proteins' structure to known proteins' in PDB by multiple threading of reassembled fragments. Also, I – TASSER is considered as the best server for structure prediction of proteins with respect to CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, CASP13 and CASP14 experiments (Yang & Zhang, 2015; "*I-tasser server for protein structure and function prediction"*, n.d.). Hence, preferring the algorithm of I – TASSER was logical and suitable for predicting secondary structure of SiiE BIg50:52 of *Salmonella*, since the unknown structures of different *Salmonella* serovars' SiiE BIg50:52 were able to be predicted by using threading approach. According to the secondary structure predictions of in total 47 METU whole genome sequenced *Salmonella* serovars' BIg50:52, it was observed 11 different models where their detailed information can be seen in the Appendix (Table H.1) and dominance of β – sheet structure. The predictions of related *Salmonella* serovars' SiiE BIg50:52 can be seen in the Appendix (Figure H.1 and Figure H.2).

After experimentally resolved structures of SiiE BIg50:52 (2YN5 and 2YN3) and MUC1 SEA domain (6BSC) from PDB, and Alphafold prediction of MUC1 (MUC1(AF-P15941-F1)) where their detailed information can be seen in Table 3.7 and Table 3.8 respectively were obtained, secondary structure predictions of METU WGS *Salmonella* serovars and *Salmonella* serovars obtained from NCBI were furtherly analyzed by MatchMaker tool of Chimera (Varadi et al., 2021; Meng et al., 2006). 2YN3 consists of 4 chains (A, B, C, D) and 2YN5 consists of 2 chains (A, B). Structural comparison of experimentally resolved SiiE BIg50:52 and predicted secondary structures showed that METU WGS and NCBI *Salmonella* serovars' SiiE BIg50:52 were best superimposed with chain C of 2YN3 and chain B of 2YN5. Hence, in this study best superimposed chains of 2YN3 and 2YN5 were furtherly used. Also, structural comparison of predicted model 1 with experimentally resolved SiiE BIg50:52 of *S*. Typhimurium LT2 is shown in Figure 4.6, while the rests' can be seen in the Appendix (Figure I.1 – Figure I.15).

A



B

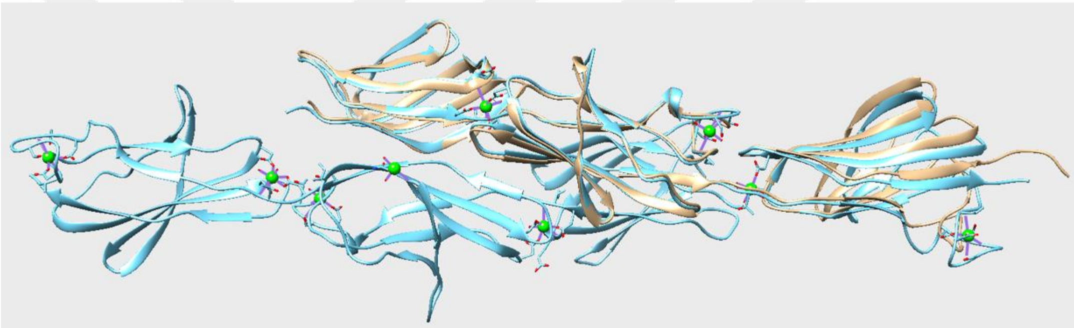


Figure 4.6 Structural comparion of model 1 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

After structural alignment, Match -> Align tool of Chimera was also used to determine root mean square deviation (RMSD) value of the models with respect to 2YN3 and 2YN5 where can be seen in the Appendix (Table I.1) (Meng et al., 2006). Two proteins’ corresponding atoms’ average deviation is measured with RMSD value, in other words it measures two protein structures’ similarity. In detail, the smaller the RMSD, the more similar the two structures (Reva et a., 1998). Also, RMSD values less than 1.5 Å which can be up to 2 Å dependent upon the structure size, are considered as very close structures (Hevener et al., 2009). In this case, almost all the RMSD values of the models with respect to 2YN3 and 2YN5 are less than 1.5 Å indicating very close structure of predicted SiiE BIg50:52 with experimentally resolved SiiE BIg50:52.

According to the docking scheme of SiiE BIg50:52 and *Homo sapiens* MUC1 which can be seen in Table 3.9, protein – protein docking was performed with ClusPro Protein – Protein Docking Server (Kozakov et al., 2017). ClusPro tool firstly by sampling billions of conformations performs rigid body docking, then for representing the most likely models finds the largest clusters by clustering of 1000 lowest energy structures according to RMSD, finally based on Gibbs free energy minimization refines the selected structures (Kozakov et al., 2017). Hence, using the algorithm of ClusPro which is based on direct docking was a reasonable preference for investigating the interaction of SiiE BIg50:52 and MUC1 since, the exact mechanism between them and their active sites had not known.

MUC1 consists of variable tandem repeat region (VNTR), sperm protein-enterokinase-agarin (SEA) domain, transmembrane domain (TM) and cytoplasmic tail (CT) where approximately located in its residues 128^{th} to 928^{th}, 1041^{st} to 1152^{nd}, 1158^{th} to 1181^{st} and 1184^{th} to 1255^{th} respectively (Chen et al., 2021). Also, in this study AlphaFold prediction of MUC1 (MUC1(AF-P15941-F1)) which is the whole structure of MUC1 is abbreviated as MUC1(AF), while experimentally resolved structure of MUC1 (6BSC) which includes only SEA domain is abbreviated as MUC1(PDB).

With further analysis by PyMOL, protein – protein interactions between SiiE BIg50:52 and MUC1 were determined (Schrödinger & Delano, 2021). The particular amino acids that are responsible for specific types of the interactions is represented in Table 4.2 and Table 4.3. Also, their visualizations can be seen in the Appendix (Figure J.1 – Figure J.40).

Table 4.2 Residues of SiiE BIg50:52 and whole *Homo sapiens* MUC1(AF) that are responsible for specific interactions respectively

Docking Scheme*	Docked Proteins	H – bond Residues	Salt Bridge Residues	Hydrophobic Interaction Residues
SiiE BIg50:52 (PDB) & MUC1 (AF)	2YN3 & MUC1 (AF)	(5282, 5283-511), (5279-529,549), (5266-551), (5260-612), (5081-656), (5152-698), (5242-756), (5201-796), (5210-816,836), (5209-856), (5212-858)	-	(5224-877), (5198-778,797,798), (5199,5241-777), (5241-757,778), (5280-509)
	2YN5 & MUC1 (AF)	(5241-296), (5201-356), (5208-376), (5225-396), (5334-640), (5359-660,680), (5361-700), (5175-1095), (5298-1093), (5314-1091,1092), (5312-1087,1089), (5315-1070), (5317-1067,1062)	(5343-1093)	(5261-189), (5198-317,337), (5211-357), (5224-397, 417), (5355-598, 618)
SiiE BIg50:52 (PDB) & MUC1 (AF)	2YN3 Chain C & MUC1 (AF)	(5310-6), (5364-24), (5285-1161), (5314-1149,1150), (5301,5314-1147), (5263-460), (5294-500), (5170-1070,1091,1092,1093), (5216-1071), (5214-1067)	(5216-1071), (5170-1093)	(5308-9,10), (5309-10), (5333-13,17)
	2YN5 Chain B & MUC1 (AF)	(5092-696), (5295-520), (5317-501), (5319-483), (5291-440), (5353-6), (5301-1147), (5183-1091), (5170-1089), (5185-1070), (5212,5213-1071), (5215-1075)	(5212-1071)	(5355-10), (5280-13)
METU WGS SiiE BIg50:52 & MUC1 (AF)	Model 1 & MUC1 (AF)	-	-	(5260-12), (5350,5353-9), (5355-10), (5263-13)

Table 4.2 (cont'd)

METU WGS SiiE BIg50:52 & MUC1 (AF)	Model 2 & MUC1 (AF)	-	(5170-1093)	(5308,5309-9)
	Model 3 & MUC1 (AF)	-	(5123-1217), (5089-365,385)	(5176-20), (5174-17), (5151,5344-14), (5299,5301-10), (5320-4), (5321-1)
	Model 4 & MUC1 (AF)	-	(5276-836,856)	-
	Model 5 & MUC1 (AF)	-	(5093-676,696)	(5310-10), (5335-13,14)
	Model 6 & MUC1 (AF)	-	(5216-1071), (5170-1093)	(5309-9), (5333-13)
	Model 7 & MUC1 (AF)	-	(5161-776), (5259-916,939), (5281-796,816)	-
	Model 8 & MUC1 (AF)	-	(5113-716,696,676), (5093-656,636)	(5355-10,13), (5280-17)
	Model 9 & MUC1 (AF)	-	(5215-1071), (5092,5160-1082), (5169,5171-1093)	(5330-12)
	Model 10 & MUC1 (AF)	-	(5215-1071), (5169-1093), (5092,5160-1082)	(5330-12)
	Model 11 & MUC1 (AF)	-	(5276-676,696), (5320-776,796), (5294-796), (5296-816,836), (5178-856), (5093,5161-956), (5196,5216-980)	(5319-778,798), (5318-777,797)
NCBI SiiE BIg 50:52 & MUC1 (AF)	*S. bongori* & MUC1 (AF)	-	(5169-1093), (5213-1068), (5215-1071)	(5332-13,17), (5308-10)
	S. Gallinarum & MUC1 (AF)	-	(5276-1093,1095), (5270-1093), (5343-376), (5178,5296-356), (5178-336), (5322,5302,5282-12)	(5268,5269,5274-1)
	S. Pullorum & MUC1 (AF)	-	(5320-836,856), (5157-596)	-
	S. Choleraesuis & MUC1 (AF)	-	(5320-716), (5296-656)	(5147-17), (5148-13)
	S. Enteritidis & MUC1 (AF)	-	(5216-1071), (5170,5183-1093)	(5355-13), (5333-14)

*: In the parenthesis, data source of the protein of interest is represented. PDB: Protein Data Bank, AF: AlphaFold

According to the Table 4.2, there was no H – bonding between METU WGS BIg50:52 & MUC1 (AF) and NCBI BIg50:52 & MUC1 (AF), even H atoms were added later to these two proteins and docking was performed again to be certain. PyMOL did not show polar contact between these two proteins, although other PPIs were present. This situation can be resulted from the shorter distance of H – bond (less than 2.5 Å) than the other PPIs (Hubbard & Kamran Haider, 2010). Another reason can be explained with the phenomenon that when the size of a protein increases, the accuracy of the bioinformatic methods decreases (María Hernández-Domínguez et al., 2020). Also, importance of SEA domain of MUC1 was described previously as having a function against invasive pathogens by actioning as a decoy receptor through its self-cleavage (Chen et al., 2021; Dhar & McAuley, 2019). Therefore, to obtain more precise information and further research, MUC1 SEA domain was furtherly investigated.

Table 4.3 Residues of SiiE BIg50:52 and SEA domain of *Homo sapiens* MUC1 (PDB) that are responsible for specific interactions respectively

Docking Scheme*	Docked Proteins	H – bond Residues	Salt Bridge Residues	Hydrophobic Interaction Residues
SiiE BIg50:52 (PDB) & MUC1 (PDB)	2YN3 & 6BSC	(5311,5312-1140), (5314-1143), (5298-1145), (5344,5126-1109), (5174,5172,5169-1108)	(5216,5344-1109), (5172-1108)	-
	2YN5 & 6BSC	(5157-1108), (5344-1091), (5256,5296-1093), (5256,5258,5294-1095), (5292-1096), (5292,5293,5322-1059), (5357-1060), (5320-1056)	(5292-1059), (5294,5256,5229-1095), (5296,5343,5256-1093), (5150-1075), (5157-1108)	-

Table 4.3 (cont'd)

SiiE BIg50:52 (PDB) & MUC1 (PDB)	2YN3 Chain C & 6BSC	(5311,5312-1108), (5311-1109), (5300,5314-1090), (5297-1091), (5298-1104), (5295,5296,5344-1093), (5298,5175-1102), (5178,5296-1095), (5344-1100), (5175-1046), (5174-1140)	(5311-1108), (5178-1095), (5296-1093,1095)	(5176-1139)
	2YN5 Chain B & 6BSC	(5314-1091), (5174-1046), (5175,5178-1095)	(5311-1108), (5178-1095)	-
METU WGS SiiE BIg50:52 & MUC1 (PDB)	Model 1 & 6BSC	(5311-1109), (5312,5314-1108), (5315-1106), (5299,5174,5177-1102)	(5312-1108), (5177-1095), (5297-1093,1095)	-
	Model 2 & 6BSC	(5093-1140), (5161,5162-1102), (5192-1104), (5169-1108)	(5170,5172-1108), (5161-1095)	(5191-1144), (5090-1097)
	Model 3 & 6BSC	(5178-1102), (5174,5176-1140), (5313-1108), (5316-1106)	(5298-1093,1095), (5313-1108)	-
	Model 4 & 6BSC	(5175-1102), (5182-1145), (5350-1108)	(5178-1095)	-
	Model 5 & 6BSC	(5175-1102), (5316-1104), (5313,5315-1108), (5312-1109)	(5313-1108), (5178-1048), (5298-1095)	(5176-1139)
	Model 6 & 6BSC	(5252-1145)	(5170-1093,1095), (5093-1082)	(5177-1144)
	Model 7 & 6BSC	(5295-1108), (5158-1102), (5157-1140)	(5296-1108), (5281-1095)	(5159-1139)
	Model 8 & 6BSC	(5303-1145), (5310-1109), (5311,5313-1108), (5314-1104), (5174,5344-1102), (5174-1140)	(5296-1095), (5311-1108)	-

Table 4.3 (cont'd)

METU WGS SiiE BIg50:52 & MUC1 (PDB)	Model 9 & 6BSC	(5090-1102), (5091-1139), (5092-1140), (5161-1104), (5191-1145), (5166,5171,5178-1108)	(5171,5177-1108), (5160-1095)	(5090-1100)
	Model 10 & 6BSC	(5172-1142), (5174-1102), (5175-1140), (5349-1145), (5311-1108)	(5297-1095)	-
	Model 11 & 6BSC	(5174-1140), (5178-1102), (5310-1109), (5311-1108)	(5311-1108), (5296-1095)	-
NCBI SiiE BIg50:52 & MUC1 (PDB)	*S. bongori* & 6BSC	(5168,5171-1108), (5299-1145), (5311-1142), (5310-1140), (5309-1102)	(5171-1108)	-
	S. Gallinarum & 6BSC	(5311-1108), (5314-1104), (5175-1102), (5174-1140)	(5296-1095), (5311-1108)	(5176-1139)
	S. Pullorum & 6BSC	(5311,5313-1108), (5310-1109), (5174-1140), (5178-1102)	(5311-1108), (5296-1095)	-
	S. Choleraesuis & 6BSC	(5311,5313-1108), (5314-1105), (5303-1145), (5175,5298-1102), (5174-1140)	(5311-1108), (5178-1048), (5296-1095)	(5176-1139)
	S. Enteritidis & 6BSC	(5311,5313-1108), (5310-1109), (5314-1105), (5350-1145), (5344-1102), (5174-1140)	(5178-1048,1095), (5296-1093,1095), (5311-1108)	(5176-1139)

*: In the parenthesis, data source of the protein of interest is represented. PDB: Protein Data Bank, AF: AlphaFold

When residues that are involved in specific interactions between SiiE BIg50:52 and MUC1 (AF) which is the whole structure of MUC1, and SiiE BIg50:52 and MUC1 (PDB) which includes only SEA domain of MUC1 is compared, it was noticed that some of the residues that are involved in interactions between SiiE BIg50:52 and MUC1 (PDB) are not present in the interactions of SiiE BIg50:52 and MUC1 (AF).

When the whole MUC1 is intact in docking, the frequency of binding to SEA domain of MUC1 is lower when it compared to the frequency of binding to only SEA domain. This situation is hypothesized as when the SEA domain is not self-cleaved, consequently MUC1 protein is not divided into two subunits, binding of SiiE BIg50:52 to MUC1 SEA domain is harder.

Additionally, when Table 4.2 and Table 4.3 were inspected in detail, the most prevalent residue of SiiE BIg50:52 which is involved in H – bonding is residue 5174 followed by residues 5311 and 5314. For salt bridge, the most prevalent residue of SiiE BIg50:52 is 5296 followed by residues 5178 and 5311, while for hydrophobic interactions, the most prevalent residue of SiiE BIg50:52 is 5176 followed by residue 5355. Also, there is no disulfide bridges between these two proteins, since SiiE BIg50:52 doesn't have cysteine residue.

Moreover, when the binding patterns were examined comprehensively by considering different regions of *Homo sapiens* MUC1, it was observed that H – bonding and salt bridges between SiiE BIg50:52 and MUC1, occur mostly in VNTR region followed by SEA domain, while hydrophobic interactions take place mostly in the region before VNTR followed by VNTR region of MUC1. According to a study, which showed MUC1 is a receptor for SiiE adhesin of *Salmonella*, binding between SiiE and MUC1 is proposed as arising from due to interactions between repetitive domains of proteins of interest (Li et al., 2019). Similarly, the observed salt bridges and H – bonding between SiiE BIg50:52 and MUC1 in this study are mostly in the VNTR of MUC1. Also, the VNTR region of MUC1 has hydrophilic characteristic (Apostolopoulos et al., 2015). Therefore, the observed hydrophobic interactions between SiiE BIg50:52 and MUC1 in this study mostly in the region before VNTR is reasonable, since hydrophobic interactions take place between nonpolar side chains of amino acids (Libretexts, 2022).

To furtherly investigate binding characteristics of proteins of interest in terms of host specificity of *Salmonella*, secondary structure prediction of *Sus scrofa* (swine) Muc1 (nomenclature for human MUC1 for animals Muc1) was obtained from AlphaFold

(Varadi et al., 2021). The detailed information of *Sus scrofa* Muc1 protein can be seen in Table 3.8.

Protein docking was also performed according to the docking scheme of SiiE BIg50:52 and *Sus scrofa* Muc1 which can be seen in Table 3.10, by using ClusPro Protein – Protein Docking server (Kozakov et al., 2017). The reason of choosing this tool was explained previously in detail. Protein – protein interactions between SiiE BIg50:52 and *Sus scrofa* Muc1 were determined with further analysis by PyMOL (Schrödinger & Delano, 2021). The specific interactions and particular amino acids which are responsible for these interactions is shown in Table 4.4. Their visualizations can also be seen in the Appendix (Figure K.1 - Figure K.10).

Table 4.4 Residues of SiiE BIg50:52 and *Sus scrofa* Muc1 that are responsible for specific interactions respectively

Docking Scheme*	**Docked Proteins**	**H – bond Residues**	**Salt Bridge Residues**	**Hydrophobic Interaction Residues**
SiiE BIg50:52 (PDB) & Muc1 (AF)	2YN3 & Muc1 (AF)	(536-540,541), (5359-543), (5282-629,633), (5285-80), (5263,5265-72), (5182-734)	-	(5302-75,78), (5266-76), (5266,5337-79), (5304-82)
	2YN5 & Muc1 (AF)	-	(5170-757)	-
SiiE BIg50:52 (PDB) & Muc1 (AF)	2YN3 Chain C & Muc1 (AF)	(5344-510), (5298-511), (5205-502), (5252-469), (5166,5167-474), (5216-478), (5215-482), (5093-366), (5092-330), (5098-286), (5126-212,249), (5124-214), (5113-370,338), (5114-302), (5115-298), (5116-266), (5118-262,226), (5152-238,274)	(5216-478), (5113-370,334), (5111-334)	-
	2YN5 Chain B & Muc1 (AF)	(5245-78), (5116-177), (5106,5118-179), (5092-209), (5128-215), (5177-719)	-	(5237-82,85), (5162-75), (5208-89), (5206-85,88), (5249-77)

Table 4.4 (cont'd)

METU WGS SiiE BIg50:52 & Muc1 (AF)	Model 8 (*S.* Typhimurium) & Muc1 (AF)	-	(5216-478), (5236,5194-442), (5113-406,370), (5111-370)	-
NCBI SiiE BIg50:52 & Muc1 (AF)	*S. bongori* & Muc1 (AF)	-	(5088-40), (5262-406,442), (5275-334,298), (5270-334)	(5364-200)
	S. Gallinarum & Muc1 (AF)	-	(5194,5216-540), (5212-534), (5263-478)	(5364-291,292)
	S. Pullorum & Muc1 (AF)	-	(5296,5178-478), (5178-442), (5089-226,262)	-
	S. Choleraesuis & Muc1 (AF)	-	(5161-744), (5178-757)	-
	S. Enteritidis & Muc1 (AF)	-	(5216-478), (5194,5136-442), (5111-370)	-

*: In the parenthesis, data source of the protein of interest is represented. PDB: Protein Data Bank, AF: AlphaFold

According to the Table 4.4, again PyMOL did not show polar contact between METU WGS BIg50:52 (Model 8, including *S.* Typhimurium) & Muc1 (AF) and NCBI BIg50:52 & Muc1 (AF) for H-bonding. The reason of encountering this situation was explained previously in detail. Additionally, when the residues that are involved in interactions between SiiE BIg50:52 and *Sus scrofa* Muc1 was inspected in detail, the most prevalent residue of SiiE BIg50:52 which is involved in salt bridge is residue 5113 and 5216 followed by residues 5111 and 5194 and for hydrophobic interactions, the most prevalent residue of SiiE BIg50:52 is 5364 which is the only residue that is responsible for hydrophobic interactions between NCBI SiiE BIg50:52 & *Sus scrofa* Muc1. Also, again there is no disulfide bridges between these two proteins, since SiiE BIg50:52 doesn't have cysteine residue.

Also, when the binding patterns were examined by considering different regions of *Sus scrofa* Muc1, a similar fashion with *Homo sapiens* MUC1 was noticed.

Likewise, H – bonding and salt bridges between SiiE BIg50:52 and *Sus scrofa* Muc1, take place mostly in VNTR region, while hydrophobic interactions occur mostly in the region before VNTR followed by VNTR region of MUC1. According to a study which VNTR polymorphism of *Muc1* gene and litter size of pigs were investigated, the interspecies homology between repeat units of VNTR region of human and pig *Muc1* gene was estimated as 63% and cytoplasmic tail of MUC1 is well conserved among different species (Xiao et al., 2012; Dhar & McAuley, 2019). Therefore, the binding of SiiE BIg50:52 to the similar regions of *Homo sapiens* MUC1 and *Sus scrofa* Muc1 is sensible.

Additionally the residues of SiiE BIg50:52 which are responsible for PPIs between SiiE BIg50:52 and *Homo sapiens* MUC1 were compared to the incidence of amino acids substitutions of SiiE BIg50:52 which can be seen in Table 4.1, it was seen that residues 5081, 5113, 5126, 5147, 5150, 5161, 5162, 5168, 5169, 5174, 5175, 5176, 5182, 5185, 5199, 5208, 5213, 5214, 5274, 5309, 5311, 5314, 5320, 5350 and 5357 are both responsible for PPIs and present in incidence of amino acids substitutions of SiiE BIg50:52. The most prevalent residue that takes place in PPIs among them is residue 5311 followed by 5174, then 5314 and 5175. For residue 5311, the type of amino acid is same for *S.* Typhimurium and other serovars since both valine (V) and glycine (G) is nonpolar. However, for residue 5174, *S.* Typhimurium has glycine (G) which is nonpolar, while other serovars have asparagine (N) which is polar. In addition, for residue 5175, *S.* Typhimurium has threonine (T) which is polar, while other serovars have isoleucine (I) which is nonpolar and for residue 5314 vice versa substitution takes place. Therefore, residues 5174, 5175 and 5314 may have a vital role on host – receptor interaction between SiiE BIg50:52 of *Salmonella* and MUC1 of *Homo sapiens*, since these are the most prevalent residues of SiiE BIg50:52 that are responsible for PPIs and their properties are dependent to the amino acid substitutions among different *Salmonella* serovars. When the same comparison was also conducted for the residues that are responsible for PPIs between SiiE BIg50:52 and *Sus scrofa* Muc1, and the incidence of amino acid substitutions of SiiE BIg50:52, only residue that can be pronounced is 5113. For the residue, *S.* Typhimurium has glutamate (E) which is negatively charged, while other serovars have asparagine (N), which is polar. Hence, residue 5113 may have a significant role during host – receptor interaction between SiiE BIg50:52 of *Salmonella* and Muc1 of *Sus scrofa.* According to a research where function of Ca^{2+} binding sites of SiiE were investigated, the role of type I Ca^{2+} binding sites is highlighted for secretion and expression of SiiE, while the role of type II Ca^{2+} binding sites in the C terminal is pointed out for proper function of SiiE during adhesion and invasion (Peters et al., 2017). Griessl et al. (2013) stated type II Ca^{2+} binding sites for SiiE BIg50:52 as 5172, 5173, 5175, 5178 and 5180 in the domain 51. Therefore, the residue 5175 which thought to have a significant role on the interaction between SiiE BIg50:52 - MUC1, is corroborant with these researcher's findings.

Also, the general overview of the number of the PPIs between SiiE BIg50:52 and *Homo sapiens* MUC1(AF), *Homo sapiens* MUC1(PDB), *Sus scrofa* Muc1 respectively is shown in Table 4.5.

Table 4.5 Number of PPIs between SiiE BIg50:52 and *Homo sapiens* MUC1(AF), *Homo sapiens* MUC1(PDB), *Sus scrofa* Muc1 individually

		***Homo sapiens* MUC1(AF)**			***Homo sapiens* MUC1(PDB)**			***Sus scrofa* Muc1(AF)***		
		H - bond	Salt Bridge	Hydro-phobic Int.	H - bond	Salt Bridge	Hydro-phobic Int.	H - bond	Salt Bridge	Hydro-phobic Int.
SiiE BIg50:52 (PDB)	2YN3	14	0	9	9	3	0	9	0	6
	2YN5	17	1	8	13	9	0	0	1	0
SiiE BIg50:52 (PDB) Chains	2YN3 Chain C	15	2	5	17	4	1	23	4	0
	2YN5 Chain B	13	1	2	4	2	0	7	0	7
METU WGS SiiE BIg50:52	Model 1 – Model 11	-	0 - 12	0 - 8	1 - 8	1 -4	0 -2			
	Model 8 (*S.* Typhimurium)	-	5	3	8	2	0	-	6	0
NCBI SiiE BIg50:52	*S.* Enteritidis	-	3	2	7	5	1	-	4	0
	S. bongori	-	3	3	6	1	0	-	6	1
	S. Gallinarum	-	10	3	4	2	1	-	4	2
	S. Pullorum	-	3	0	5	2	0	-	5	0
	S. Choleraesuis	-	2	2	7	3	1	-	2	0

*: In the parenthesis, data source of the protein of interest is represented. PDB: Protein Data Bank, AF: AlphaFold

MUC1 is considered as having a dual function for pathogens, in other words, it can function as anti-inflammatory or proinflammatory during infections (Bose & Mukherjee, 2020). For the protective activity of MUC1, McAuley and his colleagues (2017) observed higher morbidity and mortality rates during influenza A virus infection for *Muc1*$^{-/-}$ mice. Anti-inflammatory role of MUC1 on *H. pylori* was also demonstrated (McGuckin et al., 2007; Lindén et al., 2009). Additionally, McAuley et al. (2007) identified *C.jejuni* in the organs of *Muc1*$^{-/-}$ mice, but not in the *Muc1*$^{+/+}$ mice and stated more damage in the intestine of *Muc1*$^{-/-}$ mice during oral infection. In the same study, the researchers also pointed out that no observed difference for *S.*

Typhimurium which targets M cells nevertheless, whether MUC1 is expressed in the M cells or not is not known (McAuley et al., 2007). For the proinflammatory role of MUC1 on pathogens, Boll and his colleagues (2017) stated MUC1 function as a host receptor for aggregative adherence fimbriae of enteroaggregative *E.coli* and observed higher bacterial adhesion which is facilitated by MUC1. According to a more recent study which revealed MUC1 is a host receptor for *Salmonella* adhesin SiiE for the first time, MUC1 facilitates apical invasion of *Salmonella* on epithelial cells of intestine through glycans on MUC1. The researchers highlighted that *S.* Enteritidis hardly invade MUC1 knockout HT29-MTX cells, consequently MUC1 does not function as a defensive barrier against *Salmonella* (Li et al., 2019). Hence, considering these researcher's findings, the function of MUC1 against *Salmonella* was investigated in this study by taking into account host specificity and diversity of *Salmonella* comprehensively. Since *S. bongori* is most commonly observed in cold blooded animals, *S*. Gallinarum and *S*. Pullorum are host – specific to poultry and *S*. Choleraesuis is host – adapted to swine. In other words, these serovars of *Salmonella* do not cause disease in humans as prevalent as host generalist serovars of *Salmonella* which are *S*. Typhimurium and *S*. Enteritidis (Jajere, 2019). When the binding characteristics between SiiE BIg50:52 and *Homo sapiens* MUC1 which is shown in Table 4.2 and Table 4.3 is considered in terms of host specificity of *Salmonella*, it was noticed that host adapted and host restricted *Salmonella* serovars bind to MUC1 in a similar fashion with host generalist ones. Same situation is also valid for binding characteristics between SiiE BIg50:52 and *Sus scrofa* Muc1 which is presented in Table 4.4. Put differently, no relation was observed for interaction of MUC1 and different *Salmonella* serovars in terms of host specificity in this study. There might be another mechanism responsible for adhesion and invasion in terms of *Salmonella* host specificity into the host species. Grzymajlo et al. (2017) stated that calreticulin which is expressed in intestinal cells of swine and receptor for FimH adhesin of *S.* Choleraesuis, plays a part in host specifity of *S.* Choleraesuis. The researchers observed binding of FimH of *S.* Choleraesuis to calreticulin, while binding of FimH of *S.* Enteritidis to calreticulin did not observed. Therefore, the researchers pointed out that host specificity of *Salmonella* also arises from particular receptors which are expressed by specific hosts, not only expressed proteins and particular mechanisms of *Salmonella* (Grzymajlo et al., 2017).

CHAPTER 5

CONCLUSION

Salmonella enterica subsp. *enterica* (*Salmonella*) has been a prominent foodborne pathogen due to its prevalence and consequences on public health among the world. *Salmonella* has the facility of presence in various kind of foods and can be encountered during any stage of farm to fork food supply chain. Also, the burden of *Salmonella* on public health in terms of both mortality and socioeconomically cannot be underestimated. Hence, to reduce and prevent salmonellosis, investigations and actions taken have a vital role. One of the main concerns during supression of *Salmonella* is its diversity. Not all serovars cause the same disease or the same symptomps and even cannot persist in the same host. Therefore, investigating the different *Salmonella* serovars' effect on invasiveness have a vital role on enlightening the impact of salmonellosis, when its pathogenesis, disease severity and host specificity are taken into consideration. Recently, *Salmonella* intestine infection's (SiiE) which is a giant adhesin of *Salmonella* and responsible for first contact on polarized epithelial cells, host receptor on mucous which is mucin 1 (MUC1) is discovered. Consequently, investigating the host – receptor relation between SiiE and MUC1 has a crucial role on clarifying the invasiveness and even preventing salmonellosis from the very first stage.

In this study, two sets of primers were designed to conduct multiplex PCR by using whole genome sequenced METU *Salmonella* isolates. To symbolize *Salmonella* diversity, representative *Salmonella* isolates were selected with putting emphasis on clinical importance, prevalence, molecular subtype and isolation source. It was concluded that *siiE* gene is well conserved among different *Salmonella* serovars, since all of the isolates undergone molecular conformation were positive for the gene. Also, genomic datas of important *Salmonella* serovars in terms of host specificity were obtained from NCBI to furtherly used in *in silico* analysis. By using phylogenetic analysis, amino acid substitutions of different *Salmonella* serovars' SiiE bacterial immunoglobulin-like domains from 50^{th} to 52^{nd} (BIg50:52), representative of the whole SiiE protein, were highlighted. The clusters of the phylogenetic tree and its branches were consistent with the species and serovars of the *Salmonella*. Moreover, by using threading approach secondary structures of whole genome sequenced METU *Salmonella* serovars and important *Salmonella* serovars which were obtained from NCBI were predicted. After secondary structure analysis was conducted, protein – protein docking was performed for SiiE BIg50:52 and MUC1, and then protein – protein interactions were investigated. It was observed that except disulfide bridges other interactions which are hydrogen bonds, salt bridges and hydrophobic interactions have a role on binding. Also, no correlation was observed for interaction of MUC1 and SiiE BIg50:52 of different *Salmonella* serovars in terms of host specificity. Finally, it was remarked that SiiE's residues 5174 and 5175 may have a crucial role on host – receptor interaction between SiiE BIg50:52 of *Salmonella* and MUC1 of *Homo sapiens*, when incidence of amino acid substitutions and residues that are responsible for PPIs were taken into consideration.

This study facilitated to an improved comprehension of *Salmonella* –host interaction and gave a better understanding of the role of *Salmonella* diversity on binding characteristics and invasiveness.

Consequently, different mechanisms during host – pathogen interaction should also be taken into account, considering sole adhesin – host receptor interaction in this case SiiE – MUC1 interaction may not be enough to infer host specificity of *Salmonella*. Nevertheless, further investigations are required to obtain a better understanding of host – receptor relation of *Salmonella*, to obtain an extensive basis and to prevent salmonellosis from the very first stage. Since, bioinformatic studies are hypothetical until experimentally proven. Hence, the need of obtaining experimentally resolved structures of the whole SiiE and MUC1 is still required in future with improved technology.

REFERENCES

About ClustalW. (n.d.). Retrieved from https://www.megasoftware.net/web_help_10/Part_II_Assembling_Data_For_Analysis/Building_Sequence_Alignments/ClustalW/About_ClustalW.htm

Abraham, S. N., Sharon, N., Ofek, I., & Schwartzman, J. D. (2015). Adhesion and colonization. *Molecular Medical Microbiology*, 409–421. https://doi.org/10.1016/b978-0-12-397169-2.00024-x

Acar, S., Bulut, E., Durul, B., Uner, I., Kur, M., Avsaroglu, M.D., Kirmaci, H.A., Tel, Y.O., Zeyrek, F.Y., Soyer, Y., 2017. Phenotyping and genetic characterization of *Salmonella* enterica isolates from Turkey revealing arise of different features specific to geography. Int. J. Food Microbiol. 241, 98–107. https://doi.org/10.1016/j.ijfoodmicro.2016.09.031

Agnihotry, S., Pathak, R. K., Singh, D. B., Tiwari, A., & Hussain, I. (2022). Protein structure prediction. *Bioinformatics*, 177–188. https://doi.org/10.1016/b978-0-323-89775-4.00023-7

Apostolopoulos, V., Stojanovska, L., & Gargosky, S. E. (2015). Muc1 (CD227): A multi-tasked molecule. *Cellular and Molecular Life Sciences*, *72*(23), 4475–4500. https://doi.org/10.1007/s00018-015-2014-z

Azimi, T., Zamirnasta, M., Sani, M. A., Soltan Dallal, M. M., & Nasser, A. (2020). molecular mechanisms of *salmonella*effector proteins: A comprehensive review. *Infection and Drug Resistance, Volume 13*, 11–26. https://doi.org/10.2147/idr.s230604

Bakshani, C. R., Morales-Garcia, A. L., Althaus, M., Wilcox, M. D., Pearson, J. P., Bythell, J. C., & Burgess, J. G. (2018). Evolutionary conservation of the antimicrobial function of mucus: A first defence against infection. *Npj Biofilms and Microbiomes*, *4*(1). https://doi.org/10.1038/s41522-018-0057-2

Barlag, B., & Hensel, M. (2015). The Giant Adhesin SiiE of Salmonella enterica. *Molecules*, *20*(1), 1134-1150. https://doi.org/10.3390/molecules20011134

Baumler, A., & Fang, F. C. (2013). Host specificity of bacterial pathogens. *Cold Spring Harbor Perspectives in Medicine*, *3*(12). https://doi.org/10.1101/cshperspect.a010041

Bell, C., & Kyriakides, A. (2009). Salmonella. *Foodborne Pathogens*, 627–674. https://doi.org/10.1533/9781845696337.2.627

Besser, J. M. (2018). Salmonella epidemiology: A whirlwind of change. *Food Microbiology*, *71*, 55–59. https://doi.org/10.1016/j.fm.2017.08.018

Birhanu, B. T., Park, N.-H., Lee, S.-J., Hossain, M. A., & Park, S.-C. (2018). Inhibition of salmonella typhimurium adhesion, invasion, and intracellular survival via treatment with methyl gallate alone and in combination with marbofloxacin. *Veterinary Research*, *49*(1). https://doi.org/10.1186/s13567-018-0597-8

Braulio, V. B. (2005). The process of bacterial translocation and implications for health in Alcoholics. *Comprehensive Handbook of Alcohol Related Pathology*, 949–964. https://doi.org/10.1016/b978-012564370-2/50076-3

Brenner, F. W., Villar, R. G., Angulo, F. J., Tauxe, R., & Swaminathan, B. (2000). *salmonella* nomenclature. *Journal of Clinical Microbiology*, *38*(7), 2465–2467. https://doi.org/10.1128/jcm.38.7.2465-2467.2000

Brown, R. B., Hollingsworth, M. A., & Schaffer, J. E. (2021). Protein modifications | mucin family of glycoproteins. *Encyclopedia of Biological Chemistry III*, 167–172. https://doi.org/10.1016/b978-0-12-819460-7.00210-3

Bodelón, G., Palomino, C., & Fernández, L. Á. (2013). Immunoglobulin domains in*escherichia coli*and other enterobacteria: From pathogenesis to applications in Antibody Technologies. *FEMS Microbiology Reviews*, *37*(2), 204–250. https://doi.org/10.1111/j.1574-6976.2012.00347.x

Boll, E. J., Ayala-Lujan, J., Szabady, R. L., Louissaint, C., Smith, R. Z., Krogfelt, K. A., Nataro, J. P., Ruiz-Perez, F., & McCormick, B. A. (2017). Enteroaggregative *escherichia coli* adherence fimbriae drive inflammatory cell recruitment via interactions with epithelial MUC1. *mBio*, *8*(3). https://doi.org/10.1128/mbio.00717-17

Bose, M., & Mukherjee, P. (2020). Microbe–muc1 crosstalk in cancer-associated infections. *Trends in Molecular Medicine*, *26*(3), 324–336. https://doi.org/10.1016/j.molmed.2019.10.003

Broz, P., Ohlson, M. B., & Monack, D. M. (2012). Innate immune response to*salmonella typhimurium*, a model enteric pathogen. *Gut Microbes*, *3*(2), 62–70. https://doi.org/10.4161/gmic.19141

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G. V., Christie, C. H., Dalenberg, K., Di Costanzo, L., Duarte, J. M., Dutta, S., Feng, Z., Ganesan, S., Goodsell, D. S., Ghosh, S., Green, R. K., Guranović, V., Guzenko, D., Hudson, B. P., … Zhuravleva, M. (2020). RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and Applied Research and education in fundamental biology, biomedicine, biotechnology, Bioengineering and Energy Sciences. *Nucleic Acids Research*, *49*(D1). https://doi.org/10.1093/nar/gkaa1038

CDC (Centers for Disease Control and Prevention)., 1985. *Milk-borne salmonellosis -- Illinois*. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/00000520.htm

CDC (Centers for Disease Control and Prevention),. 2009. *Multistate outbreak of salmonella infections associated with peanut butter and peanut butter--containing products --- United States, 2008--2009*. Retrieved from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm58e0129a1.htm#fig1

CDC (Centers for Disease Control and Prevention)., 2018. *Burden of Foodborne Illness: Findings | Estimates of Foodborne Illness |*. Retrieved from https://www.cdc.gov/foodborneburden/2011-foodborne-estimates.html

CDC (Centers for Disease Control and Prevention)., 2022. *Salmonella homepage*. Retrieved from https://www.cdc.gov/salmonella/index.html

Chan, K., Baker, S., Kim, C. C., Detweiler, C. S., Dougan, G., & Falkow, S. (2003). Genomic comparison of *salmonella enterica* serovars and *salmonella bongori* by use of an *S. enterica* serovar typhimurium DNA microarray. *Journal of Bacteriology*, *185*(2), 553–563. https://doi.org/10.1128/jb.185.2.553-563.2003

Chen, W., Zhang, Z., Zhang, S., Zhu, P., Ko, J. K.-S., & Yung, K. K.-L. (2021). Muc1: Structure, function, and clinic application in epithelial cancers. *International Journal of Molecular Sciences*, *22*(12), 6567. https://doi.org/10.3390/ijms22126567

Cheng, R. A., Eade, C. R., & Wiedmann, M. (2019). Embracing Diversity: Differences in Virulence Mechanisms, Disease Severity, and Host Adaptations Contribute to the Success of Nontyphoidal Salmonella as a

Foodborne Pathogen. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.01368

Cui, L., Wang, X., Zhao, Y., Peng, Z., Gao, P., Cao, Z., Feng, J., Zhang, F., Guo, K., Wu, M., Chen, H., & Dai, M. (2021). Virulence comparison of salmonella enterica subsp. enterica isolates from chicken and whole genome analysis of the high virulent strain S. Enteritidis 211. *Microorganisms*, *9*(11), 2239. https://doi.org/10.3390/microorganisms9112239

Dhar, P., & McAuley, J. (2019). The role of the cell surface mucin MUC1 as a barrier to infection and regulator of inflammation. *Frontiers in Cellular and Infection Microbiology*, *9*. https://doi.org/10.3389/fcimb.2019.00117

Dharmani, P., Srivastava, V., Kissoon-Singh, V., & Chadee, K. (2008). Role of intestinal mucins in innate host defense mechanisms against pathogens. *Journal of Innate Immunity*, *1*(2), 123–135. https://doi.org/10.1159/000163037

Dokholyan, N. V. (2020). Experimentally-driven protein structure modeling. *Journal of Proteomics*, *220*, 103777. https://doi.org/10.1016/j.jprot.2020.103777

Dolan, B., & Hansson, G. C. (2021). Mucins. *Reference Module in Life Sciences*. https://doi.org/10.1016/b978-0-12-821618-7.00005-5

Duangnumsawang, Y., Zentek, J., & Goodarzi Boroojeni, F. (2021). Development and functional properties of intestinal mucus layer in poultry. *Frontiers in Immunology*, *12*. https://doi.org/10.3389/fimmu.2021.745849

EFSA (European Food Safety Authority)., 2020. The European Union one health 2020 zoonoses report. (2021). *EFSA Journal*, *19*(12). https://doi.org/10.2903/j.efsa.2021.6971

Eswarappa, S. M., Janice, J., Nagarajan, A. G., Balasundaram, S. V., Karnam, G., Dixit, N. M., & Chakravortty, D. (2008). Differentially evolved genes of salmonella pathogenicity islands: Insights into the mechanism of host specificity in salmonella. *PLoS ONE*, *3*(12). https://doi.org/10.1371/journal.pone.0003829

Farooq, Q. ul, Shaukat, Z., Aiman, S., & Li, C.-H. (2021). Protein-protein interactions: Methods, databases, and applications in virus-host study. *World Journal of Virology*, *10*(6), 288–300. https://doi.org/10.5501/wjv.v10.i6.288

Fattinger, S. A., Sellin, M. E., & Hardt, W.-D. (2021). Salmonella effector driven invasion of the gut epithelium: Breaking in and setting the house on fire. *Current Opinion in Microbiology*, *64*, 9–18. https://doi.org/10.1016/j.mib.2021.08.007

Ferrari, R. G., Rosario, D. K., Cunha-Neto, A., Mano, S. B., Figueiredo, E. E., & Conte-Junior, C. A. (2019). Worldwide epidemiology of*salmonella*serovars in animal-based foods: A meta-analysis. *Applied and Environmental Microbiology*, *85*(14). https://doi.org/10.1128/aem.00591-19

FDA (Food and Drug Administration)., 2019. *Salmonella (Salmonellosis)*. Retrieved from https://www.fda.gov/food/foodborne-pathogens/salmonella-salmonellosis

FDA (Food and Drug Administration)., 2022. *Outbreak investigation of salmonella: Seafood (October 2022)*. Retrieved from https://www.fda.gov/food/outbreaks-foodborne-illness/outbreak-investigation-salmonella-seafood-october-2022

Foodborne illness outbreaks. MN Department of Health. (2022). Retrieved from https://www.health.state.mn.us/diseases/foodborne/outbreak/foodborne.html

Fookes, M., Schroeder, G. N., Langridge, G. C., Blondel, C. J., Mammina, C., Connor, T. R., Seth-Smith, H., Vernikos, G. S., Robinson, K. S., Sanders, M., Petty, N. K., Kingsley, R. A., Bäumler, A. J., Nuccio, S.-P., Contreras, I., Santiviago, C. A., Maskell, D., Barrow, P., Humphrey, T., … Thomson, N. R. (2011). Salmonella bongori provides insights into the evolution of the Salmonellae. *PLoS Pathogens*, *7*(8). https://doi.org/10.1371/journal.ppat.1002191

Gerlach, R. G., Cláudio, N., Rohde, M., Jäckel, D., Wagner, C., & Hensel, M. (2008). Cooperation of *salmonella* pathogenicity islands 1 and 4 is required to breach epithelial barriers. *Cellular Microbiology*, *10*(11), 2364–2376. https://doi.org/10.1111/j.1462-5822.2008.01218.x

Gerlach, R. G., Jäckel, D., Stecher, B., Wagner, C., Lupas, A., Hardt, W.-D., & Hensel, M. (2007). Salmonella pathogenicity island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. *Cellular Microbiology*, *9*(7), 1834–1850. https://doi.org/10.1111/j.1462-5822.2007.00919.x

Ghosh, S., Chakraborty, K., Nagaraja, T., Basak, S., Koley, H., Dutta, S., Mitra, U., & Das, S. (2011). An adhesion protein of *salmonella enterica* serovar typhi is required for pathogenesis and potential target for vaccine development. *Proceedings of the National Academy of Sciences*, *108*(8), 3348–3353. https://doi.org/10.1073/pnas.1016180108

Griessl, M. H., Schmid, B., Kassler, K., Braunsmann, C., Ritter, R., Barlag, B., Stierhof, Y.-D., Sturm, K. U., Danzer, C., Wagner, C., Schäffer, T. E., Sticht, H., Hensel, M., & Muller, Y. A. (2013). Structural insight into the giant ca2+-binding adhesin siie: Implications for the adhesion of salmonella enterica to polarized epithelial cells. *Structure*, *21*(5), 741–752. https://doi.org/10.1016/j.str.2013.02.020

Grzymajlo, K., Ugorski, M., Suchanski, J., Kedzierska, A. E., Kolenda, R., Jarzab, A., Biernatowska, A., & Schierack, P. (2017). The novel type 1 fimbriae FIMH receptor calreticulin plays a role in salmonella host specificity. *Frontiers in Cellular and Infection Microbiology*, *7*. https://doi.org/10.3389/fcimb.2017.00326

Guo, F., Li, S. C., Wang, L., & Zhu, D. (2012). Protein-protein binding site identification by enumerating the configurations. *BMC Bioinformatics*, *13*(1). https://doi.org/10.1186/1471-2105-13-158

Hald, T. (2013). Pathogen update: Salmonella. *Advances in Microbial Food Safety*, 25–46. https://doi.org/10.1533/9780857098740.2.25

Hallstrom, K. N., & McCormick, B. A. (2015). Pathogenicity Islands. *Molecular Medical Microbiology*, 303–314. https://doi.org/10.1016/b978-0-12-397169-2.00016-0

Hansmeier, N., Miskiewicz, K., Elpers, L., Liss, V., Hensel, M., & Sterzenbach, T. (2017). Functional expression of the entire adhesiome of salmonella enterica serotype typhimurium. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-10598-2

Harris, G. (2009, February 11). *Peanut products sent out before tests*. The New York Times. Retrieved from https://www.nytimes.com/2009/02/12/health/policy/12peanut.html

Hato, T., & Dagher, P. C. (2015). How the innate immune system senses trouble and causes trouble. *Clinical Journal of the American Society of Nephrology*, *10*(8), 1459–1469. https://doi.org/10.2215/cjn.04680514

Hennessy, T. W., Hedberg, C. W., Slutsker, L., White, K. E., Besser-Wiek, J. M., Moen, M. E., Feldman, J., Coleman, W. W., Edmonson, L. M., MacDonald, K. L., & Osterholm, M. T. (1996). A national outbreak of*salmonella enteritidis*infections from ice cream. *New England Journal of*

Medicine, *334*(20), 1281–1286. https://doi.org/10.1056/nejm199605163342001

Hevener, K. E., Zhao, W., Ball, D. M., Babaoglu, K., Qi, J., White, S. W., & Lee, R. E. (2009). Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *Journal of Chemical Information and Modeling*, *49*(2), 444–460. https://doi.org/10.1021/ci800293n

Horstkorte, R., & Fuss, B. (2012). Cell adhesion molecules. *Basic Neurochemistry*, 165–179. https://doi.org/10.1016/b978-0-12-374947-5.00009-2

Houlihan, C. F., & Whitworth, J. A. G. (2019). Outbreak science: Recent progress in the detection and response to outbreaks of infectious diseases. *Clinical Medicine*, *19*(2), 140–144. https://doi.org/10.7861/clinmedicine.19-2-140

Huang, K., Herrero-Fresno, A., Thøfner, I., Skov, S., & Olsen, J. E. (2019). Interaction differences of the avian host-specific salmonella enterica serovar gallinarum, the host-generalist*s* . Typhimurium, and the cattle host-adapted*s* . Dublin with chicken primary macrophage. *Infection and Immunity*, *87*(12). https://doi.org/10.1128/iai.00552-19

Hubbard, R. E., & Kamran Haider, M. (2010). Hydrogen bonds in proteins: Role and strength. *ELS*. https://doi.org/10.1002/9780470015902.a0003011.pub2

Hüttener, M., Prieto, A., Aznar, S., Bernabeu, M., Glaría, E., Valledor, A. F., Paytubi, S., Merino, S., Tomás, J., & Juárez, A. (2019). Expression of a novel class of bacterial ig-like proteins is required for IncHI plasmid conjugation. *PLOS Genetics*, *15*(9). https://doi.org/10.1371/journal.pgen.1008399

I-tasser server for protein structure and function prediction. Zhang Lab. (n.d.). Retrieved from https://zhanggroup.org/I-TASSER/

Jajere, S. M. (2019). A review of salmonella enterica with particular focus on the pathogenicity and virulence factors, host specificity and antimicrobial resistance including multidrug resistance. *Veterinary World*, *12*(4), 504–521. https://doi.org/10.14202/vetworld.2019.504-521

Ke, C.-C., & Liu, R.-S. (2014). Lentiviral-encoded sodium iodide symporter-mediated cancer gene therapy. *Gene Therapy of Cancer*, 463–478. https://doi.org/10.1016/b978-0-12-394295-1.00032-9

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/bf01731581

Kiss, T. Ã., Morgan, E., & Nagy, G. Ã. (2007). Contribution of SPI-4 genes to the virulence of*salmonella enterica*. *FEMS Microbiology Letters*, *275*(1), 153–159. https://doi.org/10.1111/j.1574-6968.2007.00871.x

Klemm, P., & Schembri, M. A. (2000). Bacterial adhesins: Function and structure. *International Journal of Medical Microbiology*, *290*(1), 27–35. https://doi.org/10.1016/s1438-4221(00)80102-2

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Kufe, D. W. (2009). Mucins in cancer: Function, prognosis and therapy. *Nature Reviews Cancer*, *9*(12), 874–885. https://doi.org/10.1038/nrc2761

Latasa, C., Roux, A., Toledo-Arana, A., Ghigo, J.-M., Gamazo, C., Penadés, J. R., & Lasa, I. (2005). Bapa, a large secreted protein required for biofilm formation and host colonization of salmonella enterica serovar enteritidis. *Molecular*

Microbiology, *58*(5), 1322–1339. https://doi.org/10.1111/j.1365-2958.2005.04907.x

Leach, A. R., & Thomas, P. J. (2017). Protein structure prediction and homology modeling. *Comprehensive Medicinal Chemistry III*, 120–144. https://doi.org/10.1016/b978-0-12-409547-2.12350-9

Li, K., Ye, S., Alali, W. Q., Wang, Y., Wang, X., Xia, X., & Yang, B. (2017). Antimicrobial susceptibility, virulence gene and pulsed-field gel electrophoresis profiles of salmonella enterica serovar typhimurium recovered from retail raw chickens, China. *Food Control*, *72*, 36–42. https://doi.org/10.1016/j.foodcont.2016.07.032

Li, X., Bleumink-Pluym, N. M., Luijkx, Y. M., Wubbolts, R. W., van Putten, J. P., & Strijbis, K. (2019). MUC1 is a receptor for the salmonella siie adhesin that enables apical invasion into enterocytes. *PLOS Pathogens*, *15*(2). https://doi.org/10.1371/journal.ppat.1007566

Libretexts. (2022). *1.17: Protein structure*. Biology LibreTexts. Retrieved from https://bio.libretexts.org/Courses/University_of_California_Davis/BIS_2A%3A_Introductory_Biology_(Britt)/01%3A_Readings/1.17%3A_Protein_Structure

Lillehoj, E. P., Kato, K., Lu, W., & Kim, K. C. (2013). Cellular and Molecular Biology of Airway Mucins. *International Review of Cell and Molecular Biology*, 139–202. https://doi.org/10.1016/b978-0-12-407697-6.00004-0

Lindén, S. K., Sheng, Y. H., Every, A. L., Miles, K. M., Skoog, E. C., Florin, T. H., Sutton, P., & McGuckin, M. A. (2009). Muc1 limits helicobacter pylori infection both by steric hindrance and by acting as a Releasable decoy. *PLoS Pathogens*, *5*(10). https://doi.org/10.1371/journal.ppat.1000617

Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V., & McGuckin, M. A. (2008). Mucins in the mucosal barrier to infection. *Mucosal Immunology*, *1*(3), 183–197. https://doi.org/10.1038/mi.2008.5

Longo, A., Petrin, S., Mastrorilli, E., Tiengo, A., Lettini, A. A., Barco, L., Ricci, A., Losasso, C., & Cibin, V. (2019). Characterizing salmonella enterica serovar choleraesuis, var. Kunzendorf: A comparative case study. *Frontiers in Veterinary Science*, *6*. https://doi.org/10.3389/fvets.2019.00316

Main-Hester, K. L., Colpitts, K. M., Thomas, G. A., Fang, F. C., & Libby, S. J. (2008). Coordinate regulation of *salmonella* Pathogenicity Island 1 (SPI1) and spi4 in *salmonella enterica* serovar typhimurium. *Infection and Immunity*, *76*(3), 1024–1035. https://doi.org/10.1128/iai.01224-07

María Hernández-Domínguez, E., Sofía Castillo-Ortega, L., García-Esquivel, Y., Mandujano-González, V., Díaz-Godínez, G., & Álvarez-Cervantes, J. (2020). Bioinformatics as a tool for the structural and evolutionary analysis of proteins. *Computational Biology and Chemistry*. https://doi.org/10.5772/intechopen.89594

Marshall, J. S., Warrington, R., Watson, W., & Kim, H. L. (2018). An introduction to immunology and immunopathology. *Allergy, Asthma & Clinical Immunology*, *14*(S2). https://doi.org/10.1186/s13223-018-0278-1

McAuley, J. L., Corcilius, L., Tan, H.-X., Payne, R. J., McGuckin, M. A., & Brown, L. E. (2017). The cell surface mucin MUC1 limits the severity of influenza A virus infection. *Mucosal Immunology*, *10*(6), 1581–1593. https://doi.org/10.1038/mi.2017.16

McAuley, J. L., Linden, S. K., Png, C. W., King, R. M., Pennington, H. L., Gendler, S. J., Florin, T. H., Hill, G. R., Korolik, V., & McGuckin, M. A. (2007). Muc1 cell surface mucin is a critical element of the mucosal barrier to infection. *Journal of Clinical Investigation*, *117*(8), 2313–2324. https://doi.org/10.1172/jci26705

McGuckin, M. A., Every, A. L., Skene, C. D., Linden, S. K., Chionh, Y. T., Swierczak, A., McAuley, J., Harbour, S., Kaparakis, M., Ferrero, R., & Sutton, P. (2007). Muc1 mucin limits both helicobacter pylori colonization of the murine gastric mucosa and associated gastritis. *Gastroenterology*, *133*(4), 1210–1218. https://doi.org/10.1053/j.gastro.2007.07.003

Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, *7*(1). https://doi.org/10.1186/1471-2105-7-339

Monte, D. F., & Sellera, F. P. (2020). *salmonella*. *Emerging Infectious Diseases*, *26*(12). https://doi.org/10.3201/eid2612.et2612

Mullins, J. G. L. (2012). Structural modelling pipelines in Next Generation Sequencing Projects. *Challenges and Opportunities of Next-Generation Sequencing for Biomedical Research*, 117–167. https://doi.org/10.1016/b978-0-12-394287-6.00005-7

Multiple primer analyzer. Thermo Fisher Scientific - US. (n.d.). Retrieved from https://www.thermofisher.com/tr/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html

Mumy, K. L. (2014). Salmonella. *Encyclopedia of Toxicology*, 211–212. https://doi.org/10.1016/b978-0-12-386454-3.00537-6

Nicastro, L. K., Tursi, S. A., Le, L. S., Miller, A. L., Efimov, A., Buttaro, B., Tam, V., & Tükel, Ç. (2019). Cytotoxic curli intermediates form during *salmonella* biofilm development. *Journal of Bacteriology*, *201*(18). https://doi.org/10.1128/jb.00095-19

Nishimaki, T., & Sato, K. (2019). An extension of the kimura two-parameter model to the natural evolutionary process. *Journal of Molecular Evolution*, *87*(1), 60–67. https://doi.org/10.1007/s00239-018-9885-1

Noguera, M. E., Jakoncic, J., & Ermácora, M. R. (2020). High-resolution structure of intramolecularly proteolyzed human mucin-1 sea domain. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics*, *1868*(3), 140361. https://doi.org/10.1016/j.bbapap.2020.140361

O'Sullivan, N. L., & Montgomery, P. C. (2015). Ocular mucosal immunity. *Mucosal Immunology*, 1873–1897. https://doi.org/10.1016/b978-0-12-415847-4.00099-9

Peters, B., Stein, J., Klingl, S., Sander, N., Sandmann, A., Taccardi, N., Sticht, H., Gerlach, R. G., Muller, Y. A., & Hensel, M. (2017). Structural and functional dissection reveals distinct roles of ca2+-binding sites in the giant adhesin siie of salmonella enterica. *PLOS Pathogens*, *13*(5). https://doi.org/10.1371/journal.ppat.1006418

Pollard, T. D. (2010). A guide to simple and informative binding assays. *Molecular Biology of the Cell*, *21*(23), 4061–4067. https://doi.org/10.1091/mbc.e10-08-0683

Popa, G. L., & Popa, M. I. (2021). Salmonella spp. infection – a continuous threat worldwide. *GERMS*, *11*(1), 88–96. https://doi.org/10.18683/germs.2021.1244

Qiao, J., Alali, W. Q., Liu, J., Wang, Y., Chen, S., Cui, S., & Yang, B. (2018). Prevalence of virulence genes in extended-spectrum β-lactamases (esbls)-

producing *salmonella* in retail raw chicken in China. *Journal of Food Science*, *83*(4), 1048–1052. https://doi.org/10.1111/1750-3841.14111

Raghunathan, D., Wells, T. J., Morris, F. C., Shaw, R. K., Bobat, S., Peters, S. E., Paterson, G. K., Jensen, K. T., Leyton, D. L., Blair, J. M., Browning, D. F., Pravin, J., Flores-Langarica, A., Hitchcock, J. R., Moraes, C. T., Piazza, R. M., Maskell, D. J., Webber, M. A., May, R. C., … Henderson, I. R. (2011). Sada, a trimeric autotransporter from salmonella enterica serovar typhimurium, can promote biofilm formation and provides limited protection against infection. *Infection and Immunity*, *79*(11), 4342–4352. https://doi.org/10.1128/iai.05592-11

Rehman, T., Yin, L., Latif, M., Chen, J., Wang, K., & Geng, Y. et al. (2019). Adhesive mechanism of different Salmonella fimbrial adhesins. *Microbial Pathogenesis*, *137*, 103748. https://doi.org/10.1016/j.micpath.2019.103748

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an RMSD of 6 Å? *Folding and Design*, *3*(2), 141–147. https://doi.org/10.1016/s1359-0278(98)00019-4

Römling, U., Bian, Z., Hammar, M., Sierralta, W. D., & Normark, S. (1998). Curli fibers are highly conserved between *salmonella typhimurium* and *escherichia coli* with respect to operon structure and regulation. *Journal of Bacteriology*, *180*(3), 722–731. https://doi.org/10.1128/jb.180.3.722-731.1998

Ryan, M. P., O’Dwyer, J., & Adley, C. C. (2017). Evaluation of the complex nomenclature of the clinically and veterinary significant pathogen*salmonella*. *BioMed Research International*, *2017,* 1–6. https://doi.org/10.1155/2017/3782182

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Salmonellosis. Department of Health. (2022). Retrieved from https://www.health.ny.gov/diseases/communicable/salmonellosis/fact_sheet.htm

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2021). Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *50*(D1). https://doi.org/10.1093/nar/gkab1112

Schaeffer A. J. (1993). Die Rolle der Bakterienadhärenz bei Harnwegsinfekten [The role of bacterial adhesion in urinary tract infections]. *Der Urologe. Ausg . A*, *32*(1), 7–15. PMID: 8447048

Schrödinger, L., & Delano, W. (2021). *PyMOL*. Retrieved from https://www.pymol.org/pymol

Sharma, A. K., Dhasmana, N., Dubey, N., Kumar, N., Gangwal, A., Gupta, M., & Singh, Y. (2016). Bacterial virulence factors: Secreted for survival. *Indian Journal of Microbiology*, *57*(1), 1–10. https://doi.org/10.1007/s12088-016-0625-1

Shelobolina, E. S., Sullivan, S. A., O'Neill, K. R., Nevin, K. P., & Lovley, D. R. (2004). Isolation, characterization, and u(vi)-reducing potential of a facultatively anaerobic, acid-resistant bacterium from low-ph, nitrate- and u(vi)-contaminated subsurface sediment and description of *salmonella* *subterranea* SP. Nov. *Applied and Environmental Microbiology*, *70*(5), 2959–2965. https://doi.org/10.1128/aem.70.5.2959-2965.2004

Silakari, O., & Singh, P. K. (2021). Hotspot and binding site prediction: Strategy to target protein–protein interactions. *Concepts and Experimental Protocols of Modelling and Informatics in Drug Design*, 267–284. https://doi.org/10.1016/b978-0-12-820546-4.00012-x

Singh, V. (2013). Salmonella serovars and their host specificity. *Journal of Veterinary Science & Animal Husbandry*, *1*(3). https://doi.org/10.15744/2348-9790.1.301

Singh, Y., Saxena, A., Kumar, R., & Saxena, M. K. (2018). Virulence system of *salmonella* with special reference to *salmonella enterica*. *Salmonella - A Re-Emerging Pathogen*. https://doi.org/10.5772/intechopen.77210

SnapGene Viewer. (2021). Version 5.3.2. Retrieved from https://www.snapgene.com

Soleymani, F., Paquet, E., Viktor, H., Michalowski, W., & Spinello, D. (2022). Protein–protein interaction prediction with Deep Learning: A Comprehensive Review. *Computational and Structural Biotechnology Journal*, *20*, 5316–5341. https://doi.org/10.1016/j.csbj.2022.08.070

Stanzione, F., Giangreco, I., & Cole, J. C. (2021). Use of molecular docking computational tools in Drug Discovery. *Progress in Medicinal Chemistry*, 273–343. https://doi.org/10.1016/bs.pmch.2021.01.004

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular evolutionary genetics analysis (MEGA) for macos. *Molecular Biology and Evolution*, *37*(4), 1237–1239. https://doi.org/10.1093/molbev/msz312

Su, L. H., & Chiu, C. H. (2007). Salmonella: clinical importance and evolution of nomenclature. *Chang Gung medical journal*, *30*(3), 210–219.

The New York Times. (1985, April 17). *Salmonella outbreak is traced*. Retrieved from https://www.nytimes.com/1985/04/17/garden/salmonella-outbreak-is-traced.html

Thompson, A. E. (2015). The immune system. *JAMA*, *313*(16), 1686. https://doi.org/10.1001/jama.2015.2940

Tradigo, G., Rondinelli, F., & Pollastri, G. (2019). Algorithms for structure comparison and analysis: Docking. *Encyclopedia of Bioinformatics and Computational Biology*, 77–80. https://doi.org/10.1016/b978-0-12-809633-8.20485-8

Urrutia, I. M., Fuentes, J. A., Valenzuela, L. M., Ortega, A. P., Hidalgo, A. A., & Mora, G. C. (2014). Salmonella typhi shda: Pseudogene or allelic variant? *Infection, Genetics and Evolution*, *26*, 146–152. https://doi.org/10.1016/j.meegid.2014.05.013

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2021). Alphafold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, *50*(D1). https://doi.org/10.1093/nar/gkab1061

Voynow, J. A., & Fischer, B. M. (2006). Mucins. *Encyclopedia of Respiratory Medicine*, 56–62. https://doi.org/10.1016/b0-12-370879-6/00248-9

Wagner, C., & Hensel, M. (2011). Adhesive mechanisms of salmonella enterica. *Advances in Experimental Medicine and Biology*, 17–34. https://doi.org/10.1007/978-94-007-0940-9_2

Wagner, C., Barlag, B., Gerlach, R. G., Deiwick, J., & Hensel, M. (2014). The*sAlmonella Enterica*giant adhesin siie binds to polarized epithelial cells in a lectin-like manner. *Cellular Microbiology*, *16*(6), 962–975. https://doi.org/10.1111/cmi.12253

Wang, M., Qazi, I. H., Wang, L., Zhou, G., & Han, H. (2020a). Salmonella virulence and immune escape. *Microorganisms*, *8*(3), 407. https://doi.org/10.3390/microorganisms8030407

Wang, S., Yang, D., Wu, X., Wang, Y., Wang, D., Tian, M., Li, T., Qi, J., Wang, X., Ding, C., & Yu, S. (2018). Autotransporter Misl of salmonella enterica serotype typhimurium facilitates bacterial aggregation and biofilm formation. *FEMS Microbiology Letters*, *365*(17). https://doi.org/10.1093/femsle/fny142

Wang, X., Wang, H., Li, T., Liu, F., Cheng, Y., Guo, X., Wen, G., Luo, Q., Shao, H., Pan, Z., & Zhang, T. (2020b). Characterization of salmonella spp. isolated from chickens in central China. *BMC Veterinary Research*, *16*(1). https://doi.org/10.1186/s12917-020-02513-1

WHO (World Health Organization)., 2022. *Food safety*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/food-safety

WHO (World Health Organization)., 2019. *Foodborne diseases*. Retrieved from https://www.who.int/health-topics/foodborne-diseases#tab=tab_1

WHO (World Health Organization)., 2018. *Salmonella (non-typhoidal)*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal)

WHO (World Health Organization)., 2015. *WHO estimates of the global burden of foodborne diseases: Foodborne disease burden epidemiology reference group*

2007-2015. World Health Organization. Retrieved from https://apps.who.int/iris/handle/10665/199350

Wu, S.-yan, Wang, L.-dan, Li, J.-ling, Xu, G.-mei, He, M.-ling, Li, Y.-yuan, & Huang, R. (2016). Salmonella SPV locus suppresses host innate immune responses to bacterial infection. *Fish & Shellfish Immunology*, *58*, 387–396. https://doi.org/10.1016/j.fsi.2016.09.042

Xiao, C., Jinluan, F., & Aiguo, W. (2012). Effect of VNTR polymorphism of the MUC1 gene on litter size of pigs. *Molecular Biology Reports*, *39*(5), 6251–6258. https://doi.org/10.1007/s11033-012-1445-y

Xu, F., Newby, J. M., Schiller, J. L., Schroeder, H. A., Wessler, T., Chen, A., Forest, M. G., & Lai, S. K. (2019). Modeling barrier properties of intestinal mucus reinforced with IGG and secretory IGA against motile bacteria. *ACS Infectious Diseases*, *5*(9), 1570–1580. https://doi.org/10.1021/acsinfecdis.9b00109

Yang, J., & Zhang, Y. (2015). I-Tasser Server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1). https://doi.org/10.1093/nar/gkv342

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-Blast: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1). https://doi.org/10.1186/1471-2105-13-134

Yuste, J. (2017). Mucin 1 is a novel glycoprotein involved in host defense against invasive pneumococcal disease. *Virulence*, *8*(8), 1475–1477. https://doi.org/10.1080/21505594.2017.1356971

Zhang, D., Zhuang, L., Wang, C., Zhang, P., Zhang, T., Shao, H., Han, X., & Gong, J. (2018). Virulence gene distribution of salmonella Pullorum isolates recovered from chickens in China (1953–2015). *Avian Diseases*, *62*(4), 431. https://doi.org/10.1637/11927-071318-resnote.1

APPENDICES

A. PREPARATION OF MEDIA

Table A.1 Brain Heart Infusion (BHI) Agar

BHI Medium	18.5 g
Agar Bacteriological	7.5 g
dH_2O	500 mL

B. CHEMICALS AND MATERIALS

Table B.1 The list of chemicals and materials with related manufacturer information

Chemicals	**Manufacturers**
American Bacteriological Agar	Condalab (Madrid, Spain)
Brain Hearth Infusion Medium	Merck (Darmstadt, Germany)
Trizma Base	Sigma-Aldrich (St. Lois, MO, USA)
Boric Acid	Sigma-Aldrich (St. Lois, MO, USA)
EDTA	Sigma-Aldrich (St. Lois, MO, USA)
5xMyTaq Reaction Buffer	Bioline (London, UK)
siiE Primer Sets	Metabion (Planegg, Germany)
MyTaq Red DNA Polymerase	Bioline (London, UK)
Agarose	Sigma-Aldrich (St. Lois, MO, USA)
DirectLoad PCR 100 bp Low Ladder	Sigma-Aldrich (St. Lois, MO, USA)
EtBr	Merck (Darmstadt, Germany)

C. DATABASES AND BIOINFORMATIC TOOLS

Table C.1 The list of databases and bioinformatic tools with their versions

Databases and Bioinformatic Tools	Versions
NCBI	-
MEGAX	X
NCBI Primer - BLAST	-
SnapGene Viewer	5.3.2
ThermoFisher Multiple Primer Analyzer	-
PDB	-
AlphaFold	-
I - TASSER	5.1
ClusPro	2.0
PyMOL	2.5.2
UCSF Chimera	1.16
UCSF Chimera MatchMaker	-
UCSF Chimera Match -> Align	-

D. MULTIPLE SEQUENCE ALIGNMENT OF *siiE* GENE

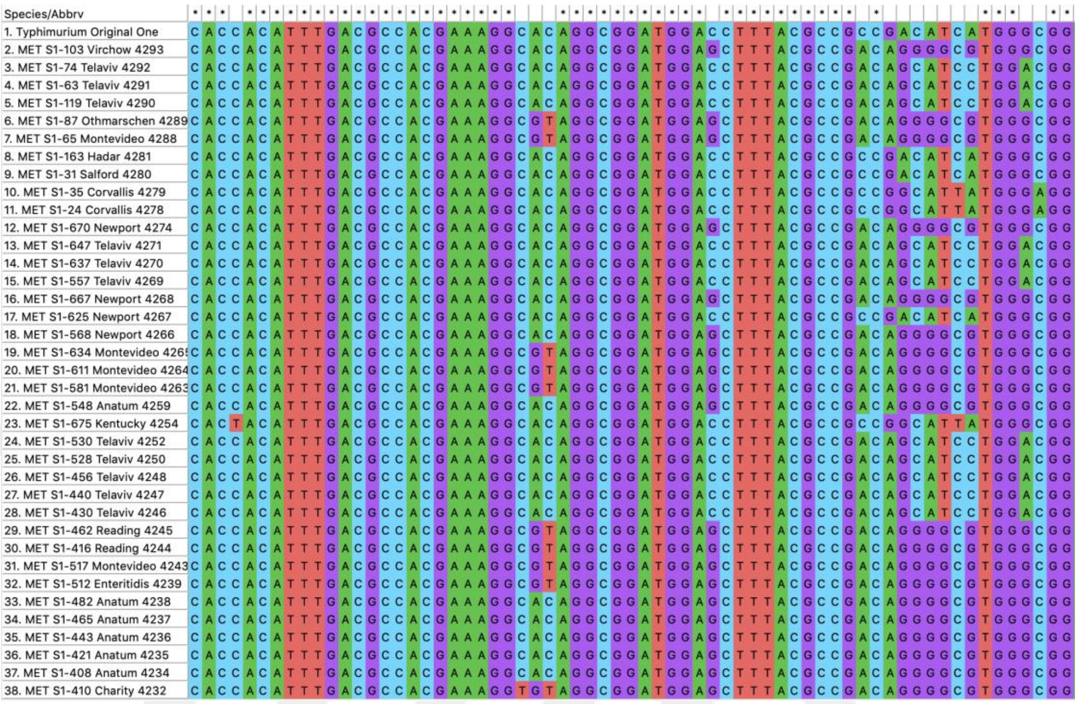


Figure D.1 Part of multiple sequence alignment of WGS METU *Salmonella* Isolates' *siiE* gene nucleotide sequences (residues 4116-4180)

E. BINDING SITES OF PRIMER SETS

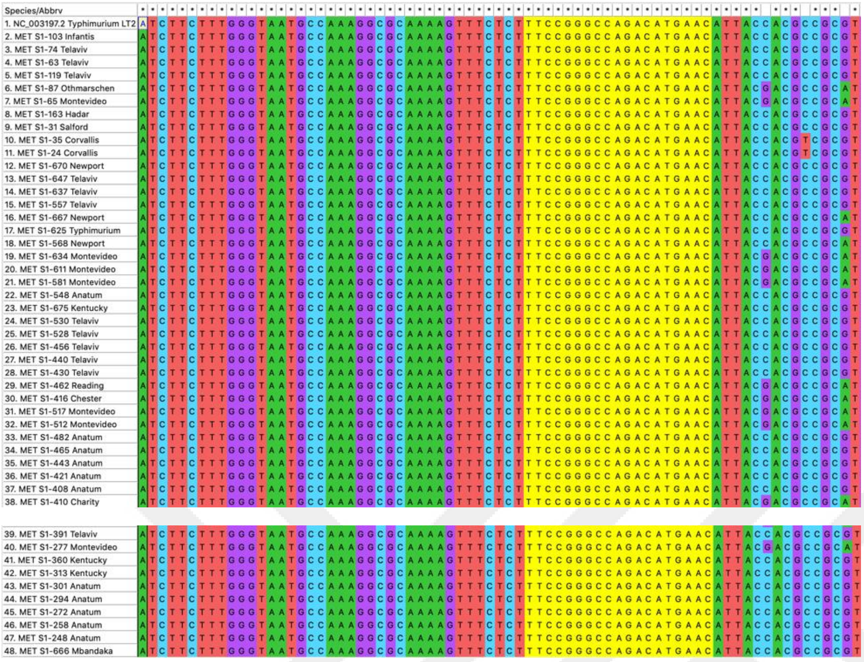


Figure E.1 Binding site of *siiE*-I-F for WGS METU *Salmonella* isolates

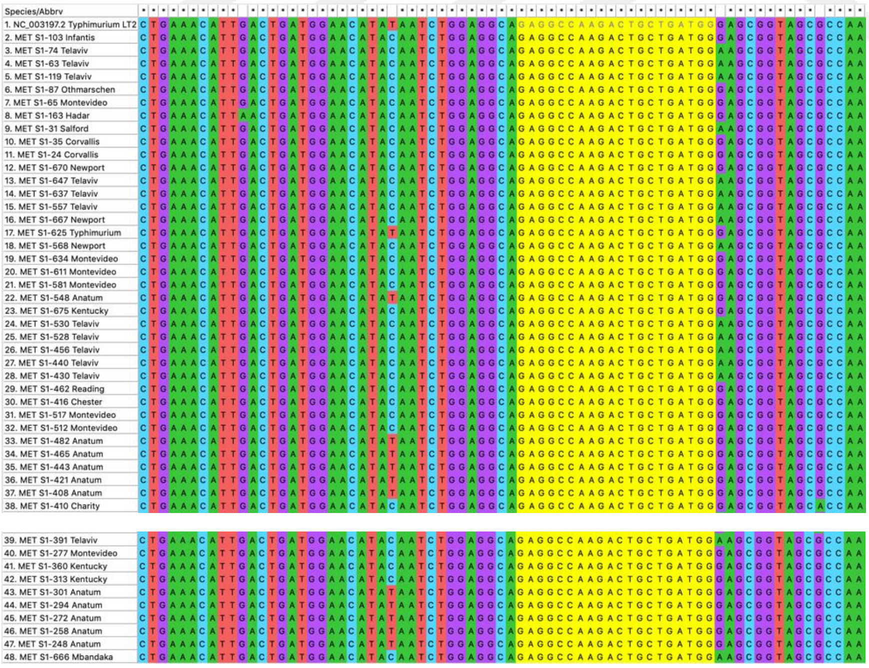


Figure E.2 Binding site of *siiE*-I-R for WGS METU *Salmonella* isolates

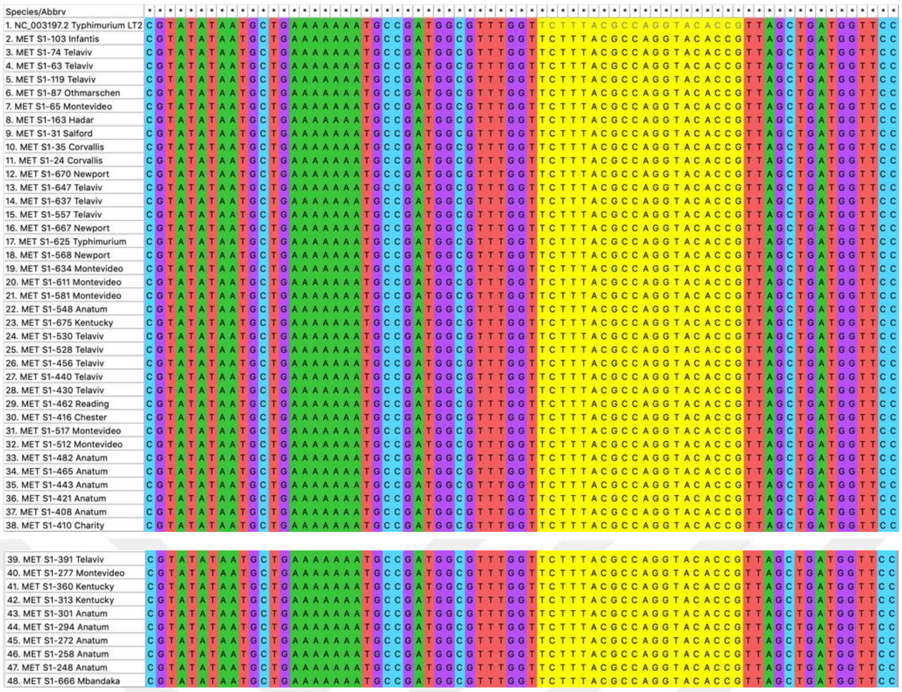


Figure E.3 Binding site of *siiE*-L-F for WGS METU *Salmonella* isolates

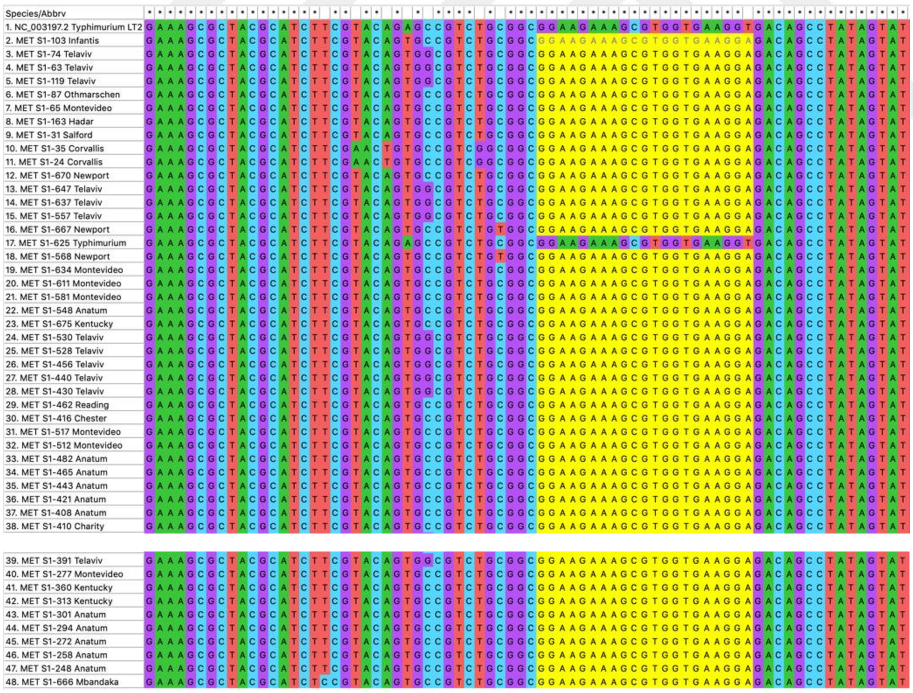
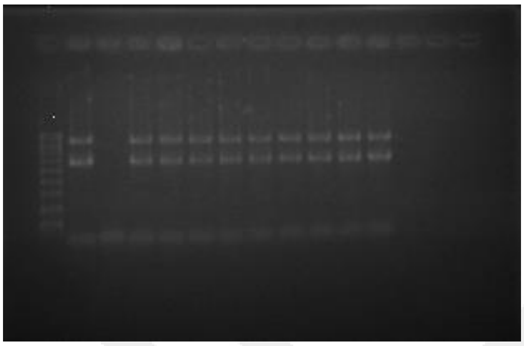


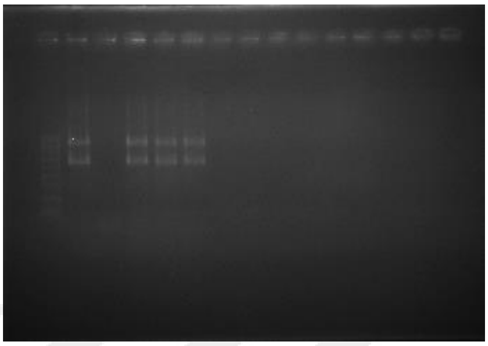
Figure E.4 Binding site of *siiE*-L-R for WGS METU *Salmonella* isolates

F. MULTIPLEX PCR GEL ELECTROPHORESIS IMAGES OF *SALMONELLA* ISOLATES

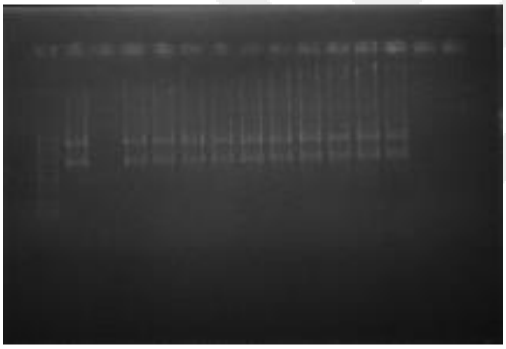
A



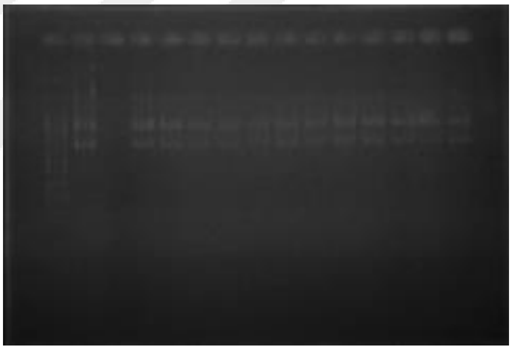
B



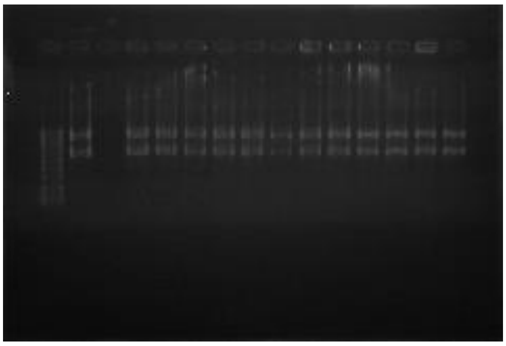
C



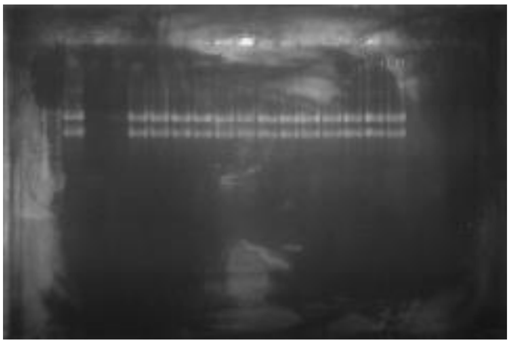
D



E



F



G

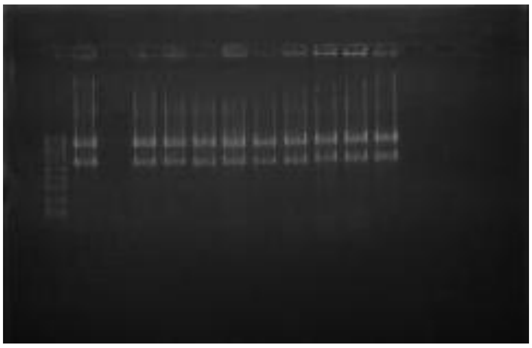


Figure F.1 Multiplex PCR gel electrophoresis images of *Salmonella* isolates. (A) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-611; 4: MET S1-625; 5: MET S1-634; 6: MET S1-637; 7. (B) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-324; 4: MET S1-372; 5: MET S1-581. (C) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-184; 4: MET S1-188; 5: MET S1- 185; 6: MET S1-223; 7: MET S1-220; 8: MET S1-234; 9: MET S1-187; 10: MET S1-217; 11: MET S1-221; 12: MET S1-227. (D) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-170; 4: MET S1-172; 5: MET S1-175; 6: MET S1-396; 7: MET S1-397; 8: MET S1- 400; 9: MET S1-439; 10: MET S1-658; 11: MET S1-166; 12: MET S1-568; 13: MET S1-662; 14: MET S1-670. (E) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-119; 4: MET S1-391; 5: MET S1-399; 6: MET S1-401; 7: MET S1-407; 8: MET S1-545; 9: MET S1-702; 10: MET S1-411; 11: MET S1-660; 12: MET S1-742; 13: MET A2-12; 14: MET A2-230. (F) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET S1-313; 4: MET S1-405; 5: MET S1-536; 6: MET A2-72; 7: MET S1-653; 8: MET S1-657; 9: MET A2-3; 10: MET A2-236; 11: MET S1-186; 12: MET S1-989; 13: MET S1-654; 14: MET S1-3; 15: MET S1-999; 16: MET A2-99. (G) L: DNA ladder; 1 (+): MET S1-647; 2 (-); 3: MET A2-36; 4: MET A2-78; 5: MET S1-14; 6: MET S1-868; 7: MET S1-398; 8: MET S1-998; 9: MET S1-24; 10: MET S1-703; 11: MET S1-313

G. AMINO ACID SUBSTITUTIONS OF *SALMONELLA* SiiE BIg50:52

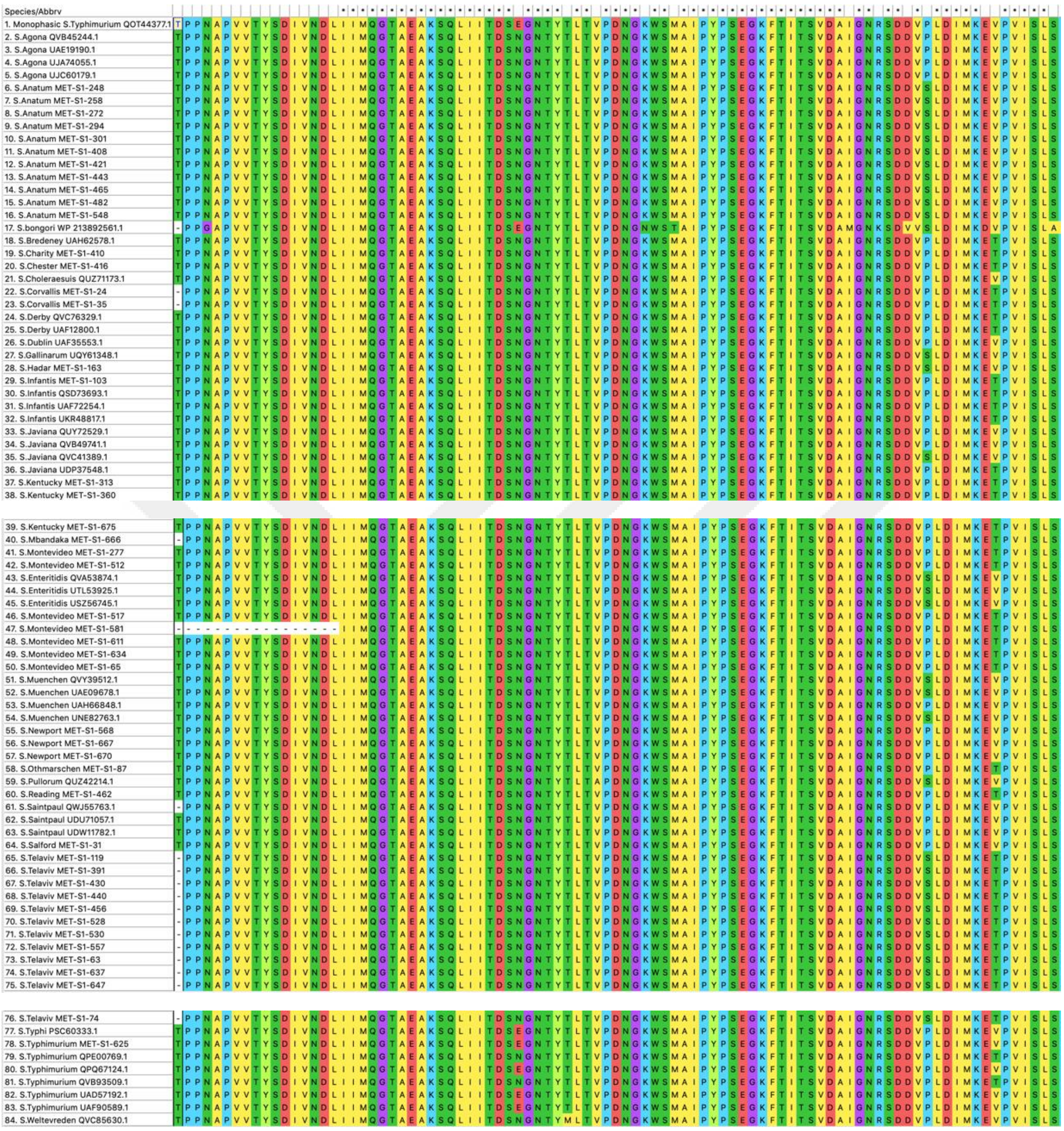


Figure G.1 Multiple sequence alignment of SiiE Blg50:52 of *Salmonella* isolates from NCBI and WGS METU isolates (residues 1-91)

Table G.1 Groups which are determined according to the phylogenetic tree clusters

Group Number	***Salmonella* Serovars Included**
1	Reading, Othmarschen, Montevideo, Infantis, Chester, Charity, Bredeney, Typhimurium
2	Typhimurium
3	Kentucky
4	Mbandaka
5	Telaviv
6	Corvallis
7	Typhimurium, Typhi, monophasic Typhimurium
8	Salford
9	Newport
10	Dublin, Newport
11	Anatum, Gallinarum, Hadar, Javiana, Enteritidis, Muenchen
12	Javiana
13	Javiana
14	Derby
15	Agona, Muenchen
16	Saintpaul
17	Weltevreden
18	Choleraesuis
19	Pullorum
20	Bongori

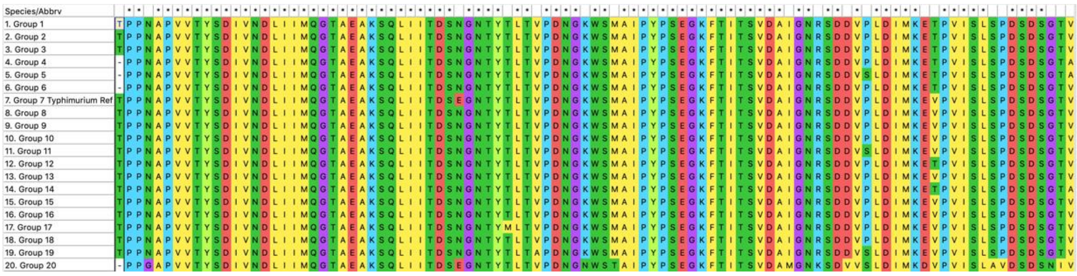


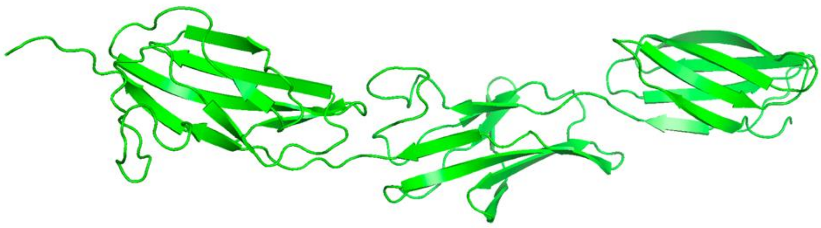
Figure G.2 Multiple sequence alignment of the groups (residues 1-99)

H. SECONDARY STRUCTURE PREDICTIONS OF SiiE BIg50:52

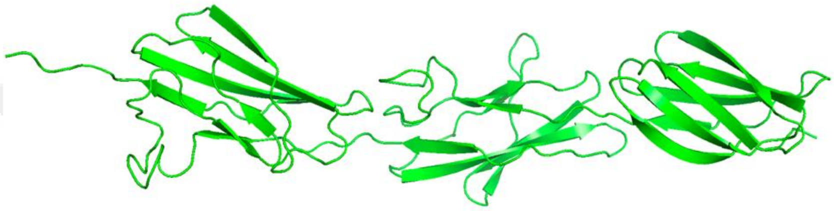
Table H.1 Models which are determined according to secondary structure predictions

Model Number	METU ID Code
1	MET S1-24, MET S1-35
2	MET S1-31
3	MET S1-65, MET S1-87, MET S1- 103, MET S1- 277, MET S1- 410, MET S1-416, MET S1-462, MET S1-512, MET S1-517, MET S1-611, MET S1-634
4	MET S1-163, MET S1-248, MET S1-258, MET S1-272, MET S1-294, MET S1-301, MET S1-408, MET S1-421, MET S1-443, MET S1- 465, MET S1-482, MET S1-548
5	MET S1-313, MET S1-360, MET S1-675
6	MET S1-568, MET S1-667
7	MET S1-581
8	MET S1-625
9	MET S1-647, MET S1-63, MET S1-74, MET S1-119, MET S1-391, MET S1-430, MET S1-440, MET S1-456, MET S1-528, MET S1-530, MET S1-557, MET S1-637
10	MET S1-666
11	MET S1-670

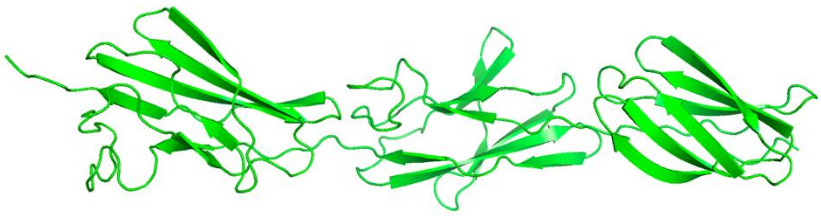
A



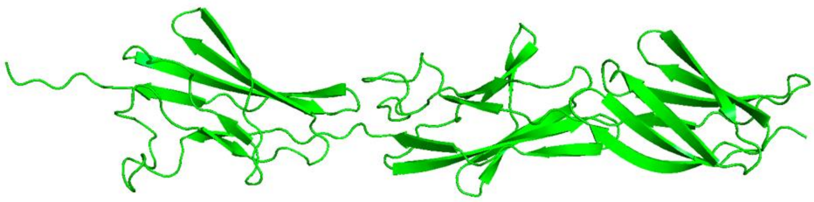
B



C

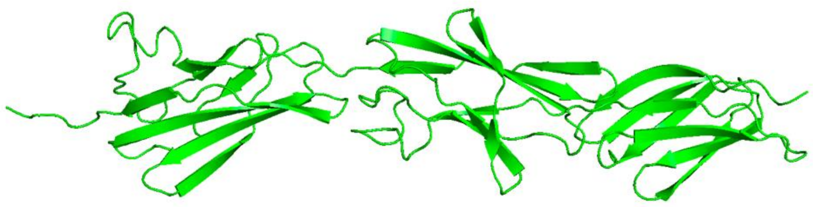


D

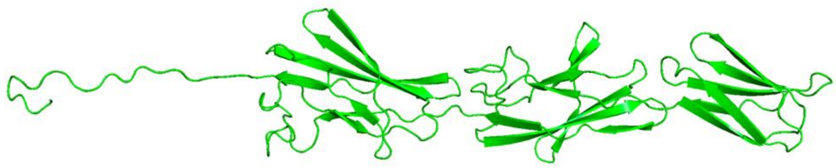


E

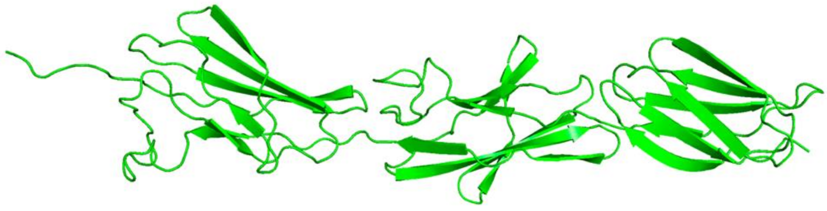
F



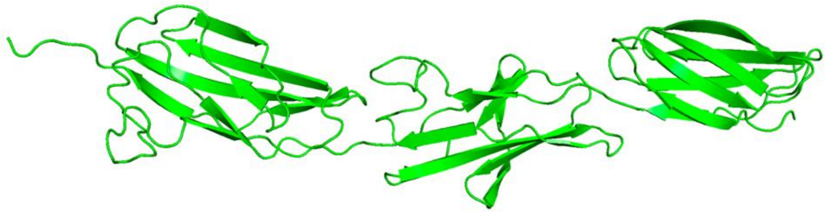
G



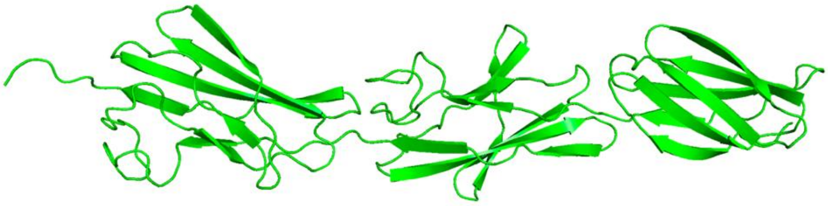
H



I



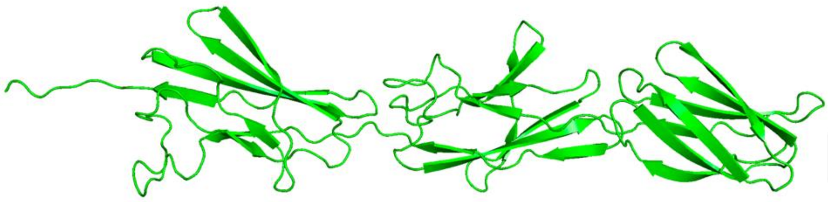
J



K

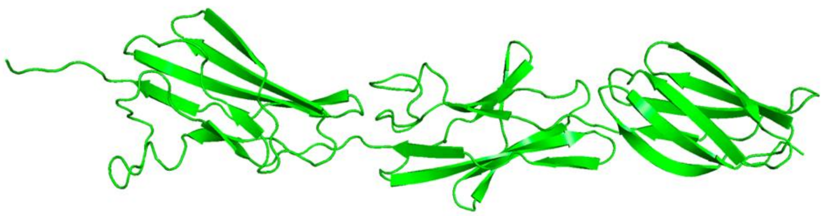
Figure H.1 Secondary structure predictions of SiiE BIg50:52 of WGS METU *Salmonella* isolates. (A) Model 1. (B) Model 2. (C) Model 3. (D) Model 4. (E) Model 5. (F) Model 6. (G) Model 7. (H) Model 8. (I) Model 9. (J) Model 10. (K) Model 11.

A

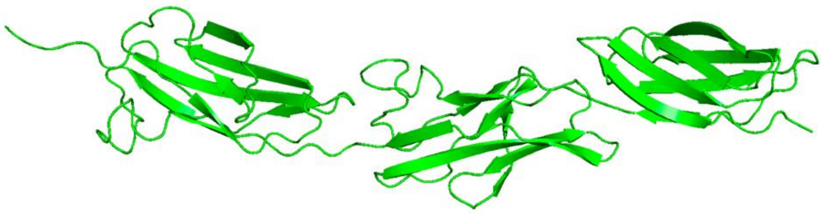


B

C



D



E

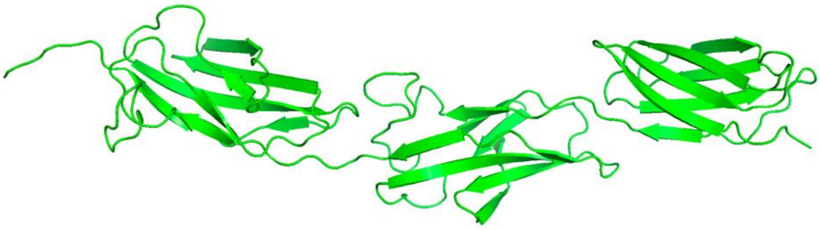
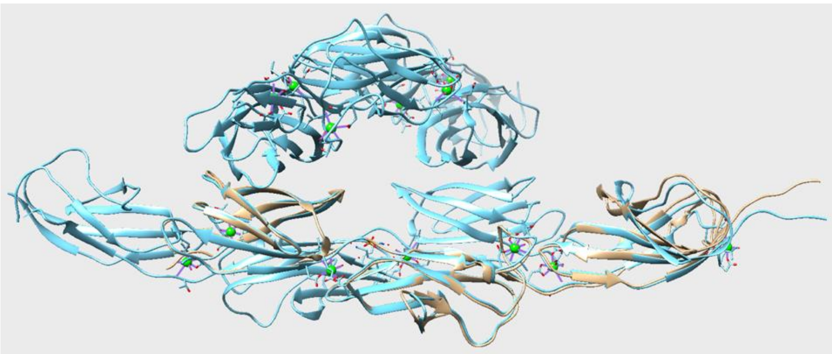


Figure H.2 Secondary structure predictions of SiiE BIg50:52 of *Salmonella* isolates obtained from NCBI. (A) *S*. bongori 40:z35:-. (B) *S.* Gallinarum. (C) *S*. Pullorum. (D) *S*. Choleraesuis. (E) *S*. Enteritidis

İ. STRUCTURAL COMPARISON OF *SALMONELLA* ISOLATES' SiiE BIg50:52

A



B

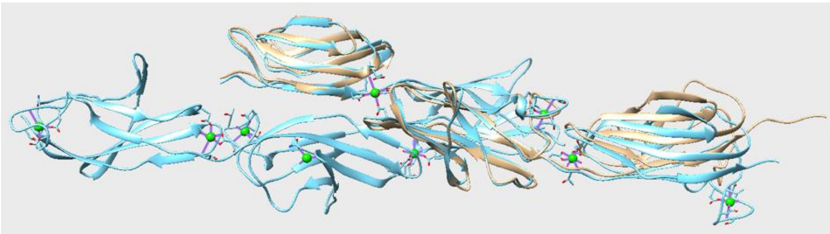
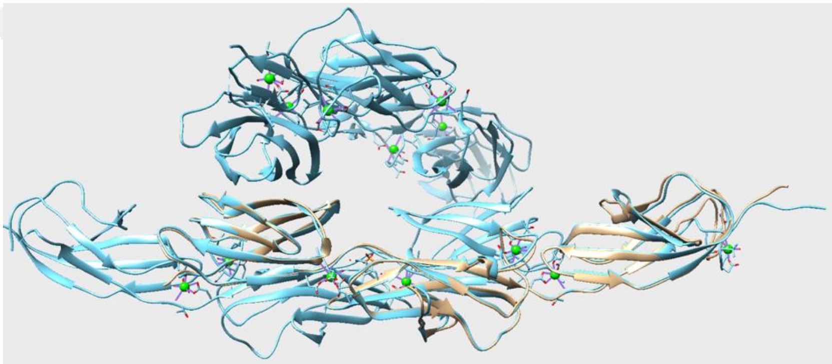


Figure I.1 Structural comparison of model 2 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

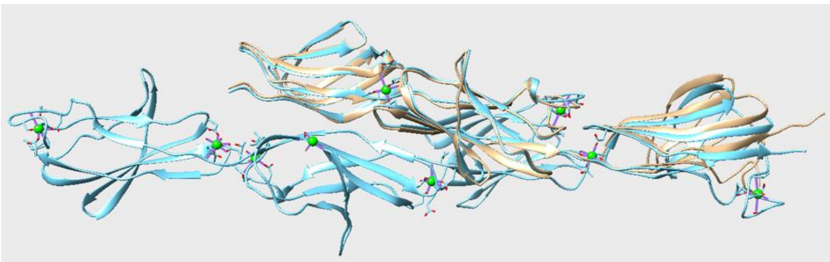
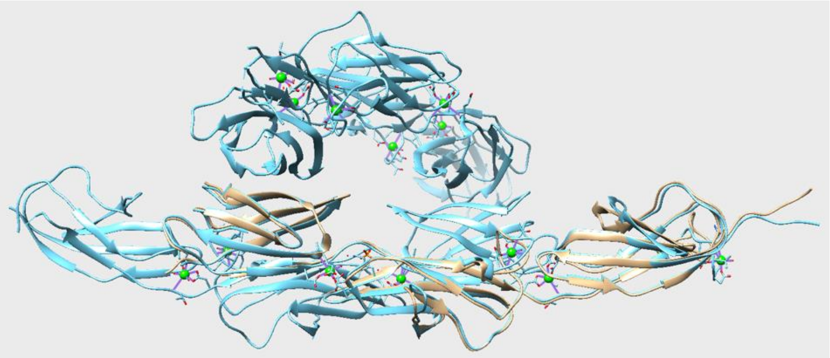


Figure I.2 Structural comparison of model 3 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

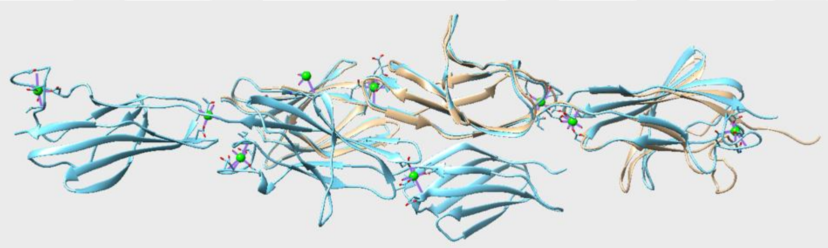
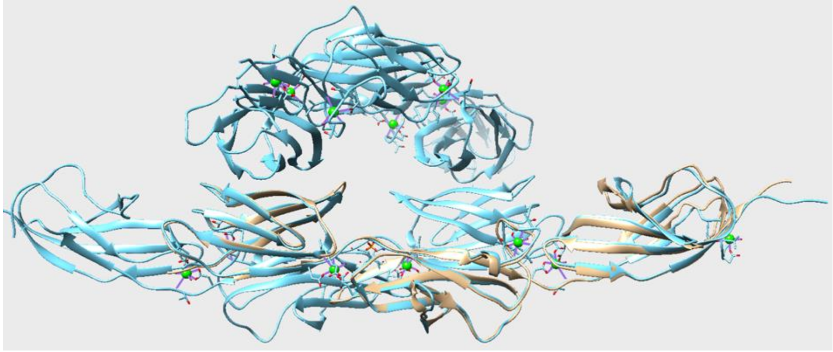


Figure I.3 Structural comparison of model 4 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

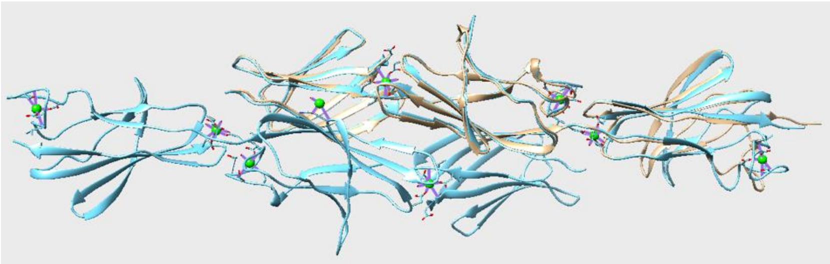
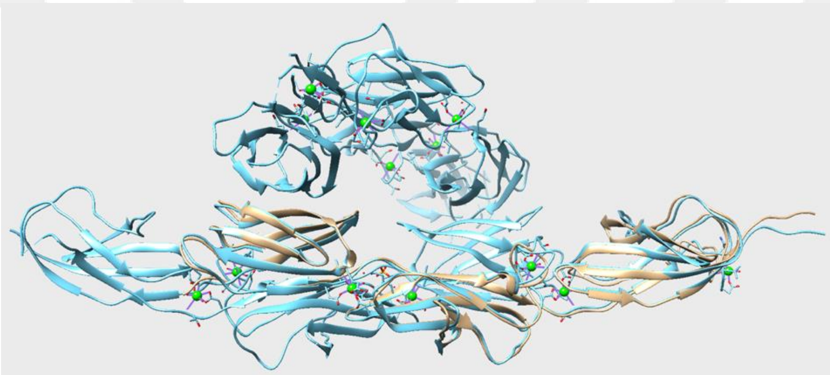


Figure I.4 Structural comparison of model 5 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

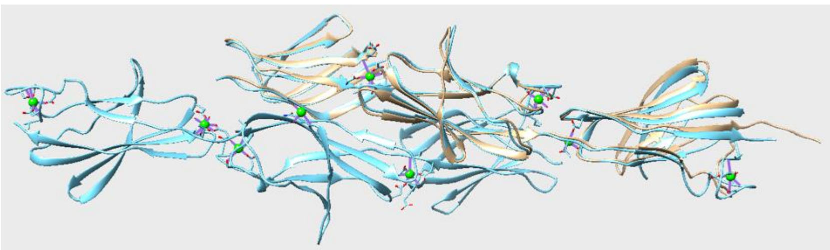
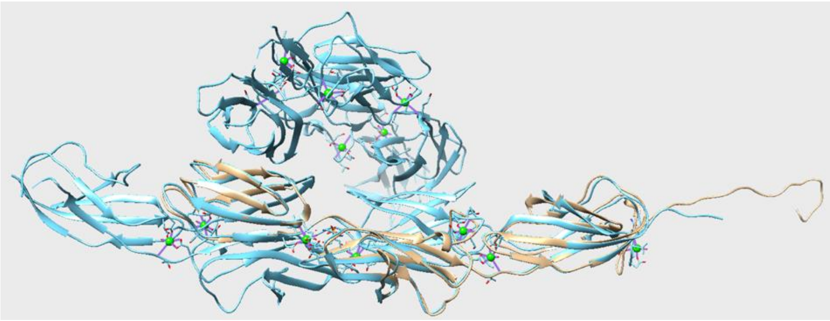


Figure I.5 Structural comparison of model 6 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

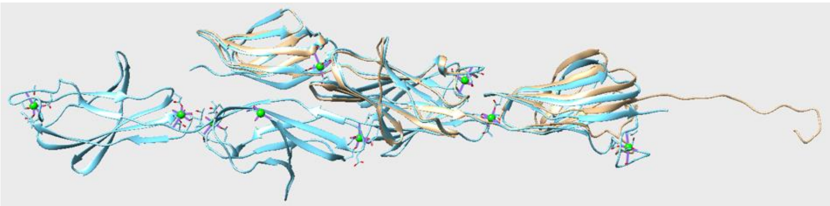
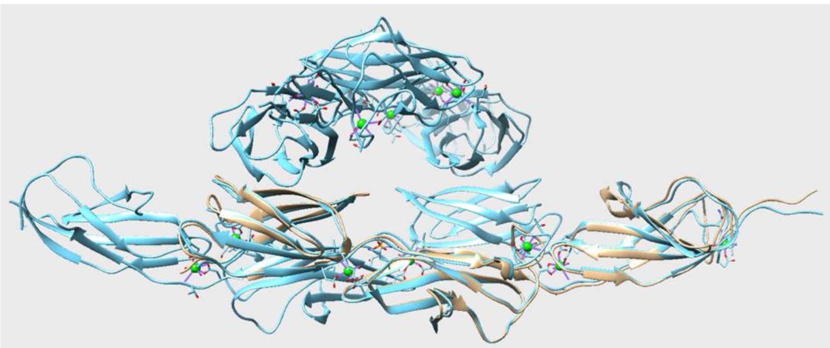


Figure I.6 Structural comparison of model 7 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

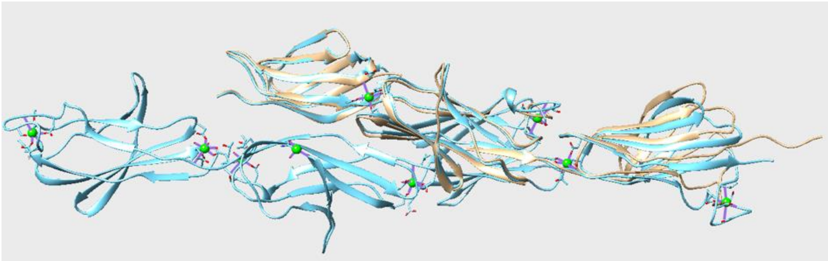
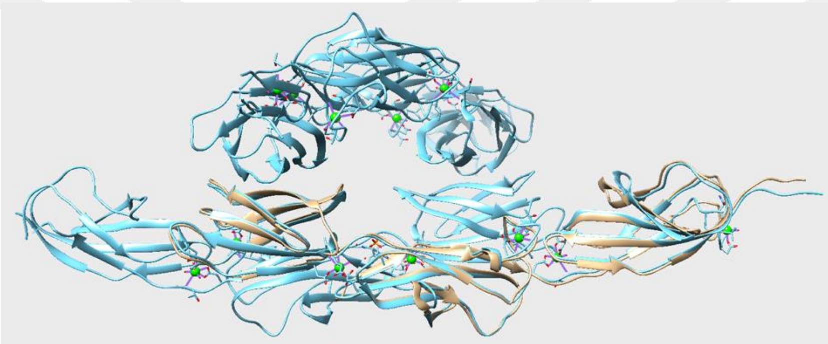


Figure I.7 Structural comparison of model 8 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

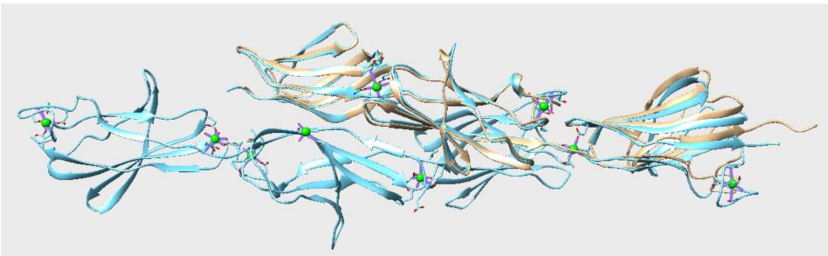
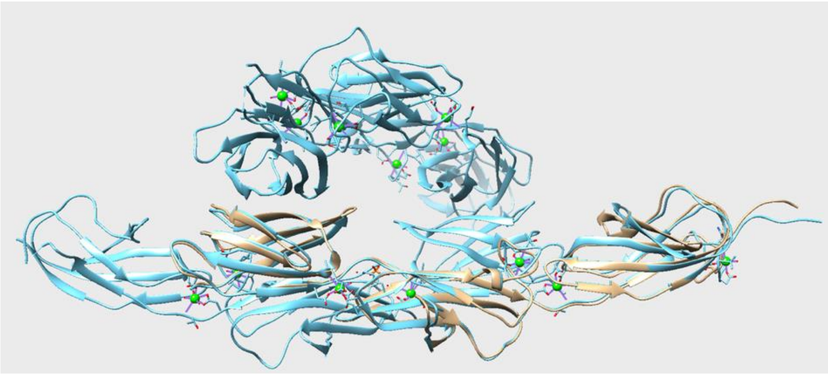


Figure I.8 Structural comparison of model 9 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

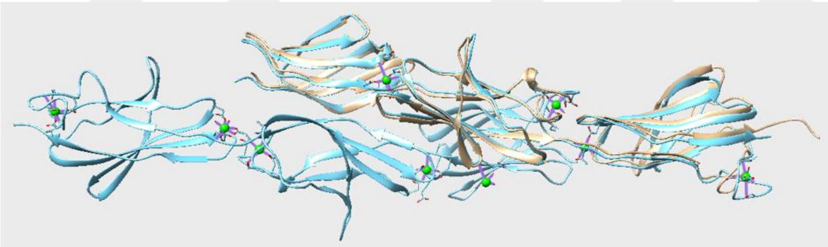
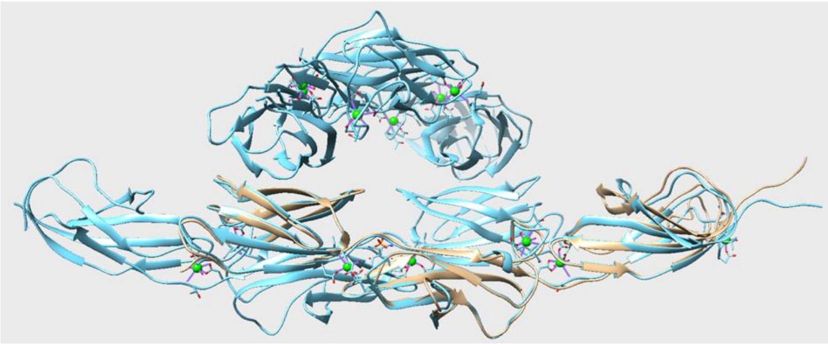


Figure I.9 Structural comparison of model 10 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

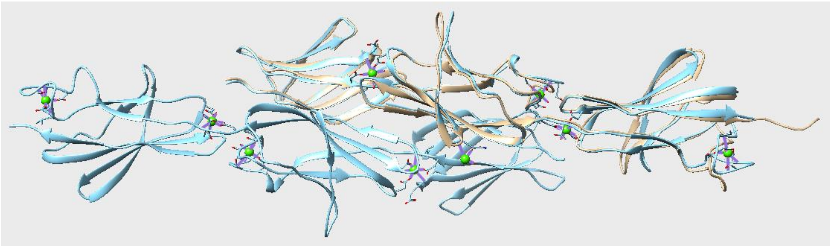
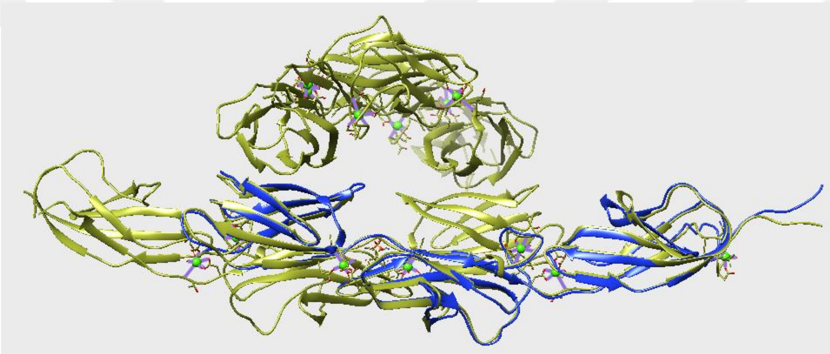


Figure I.10 Structural comparison of model 11 (beije). (A) With respect to 2YN3 (blue). (B) With respect to 2YN5 (blue).

A



B

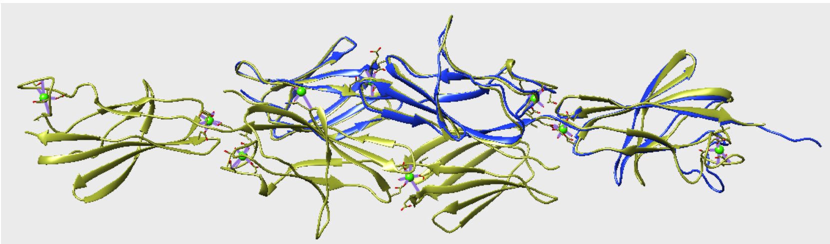
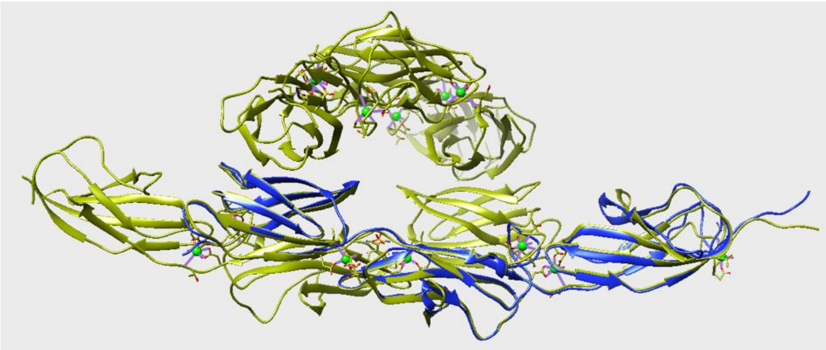


Figure I.11 Structural comparison of *S*. bongori 40:z35:- (blue). (A) With respect to 2YN3 (yellow). (B) With respect to 2YN5 (yellow).

A



B

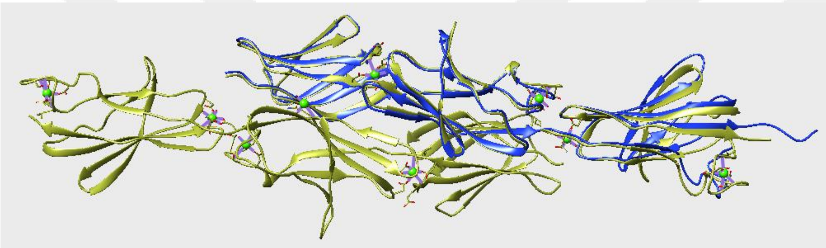
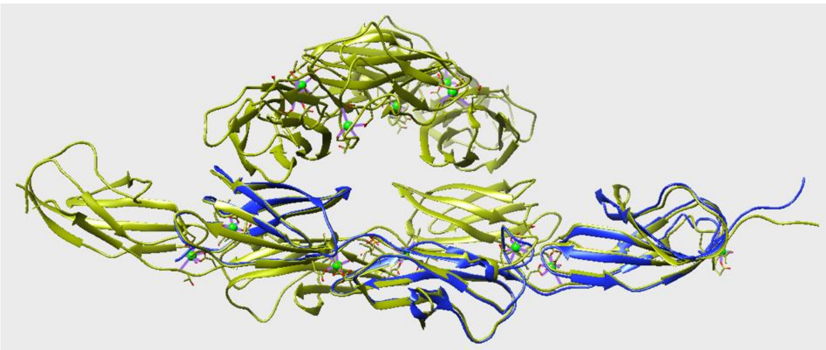


Figure I.12 Structural comparison of *S*. Gallinarum (blue). (A) With respect to 2YN3 (yellow). (B) With respect to 2YN5 (yellow).

A



B

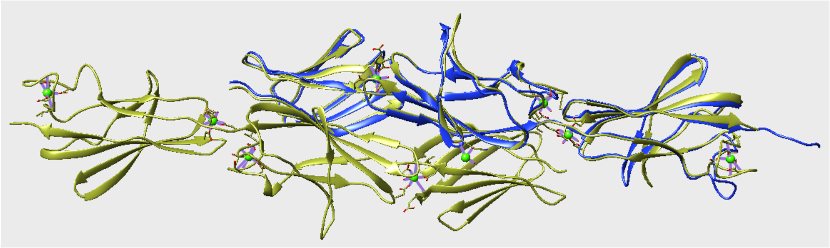
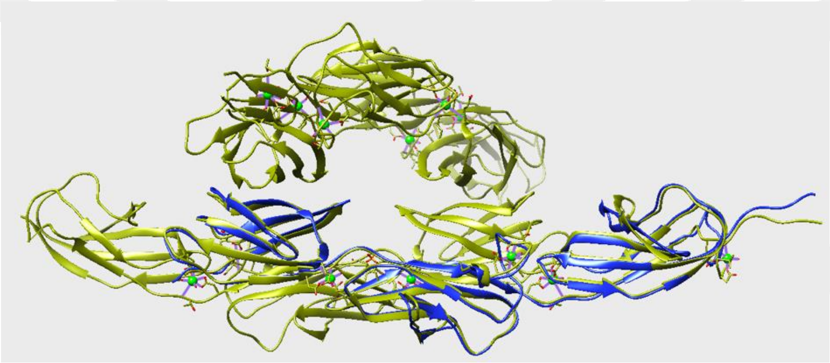


Figure I.13 Structural comparison of *S*. Pullorum (blue). (A) With respect to 2YN3 (yellow). (B) With respect to 2YN5 (yellow).

A



B

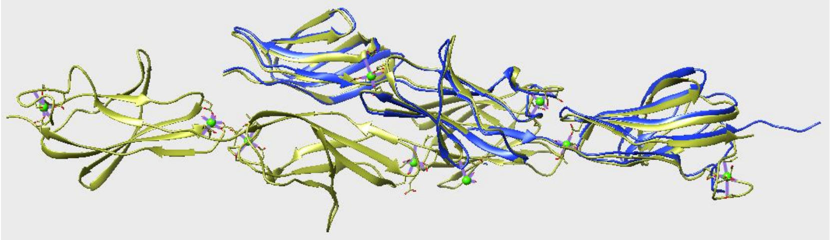
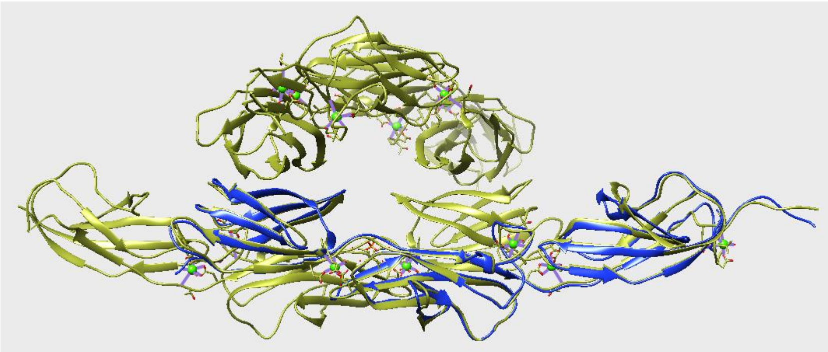


Figure I.14 Structural comparison of *S*. Choleraesuis (blue). (A) With respect to 2YN3 (yellow). (B) With respect to 2YN5 (yellow).

A



B

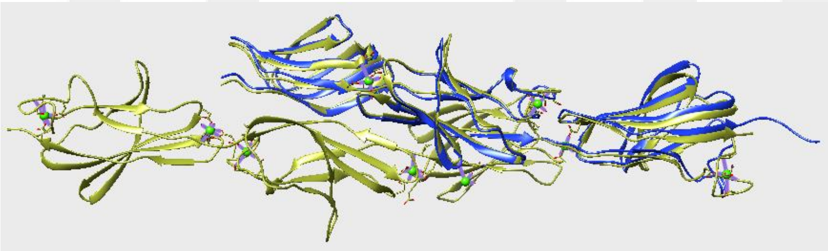


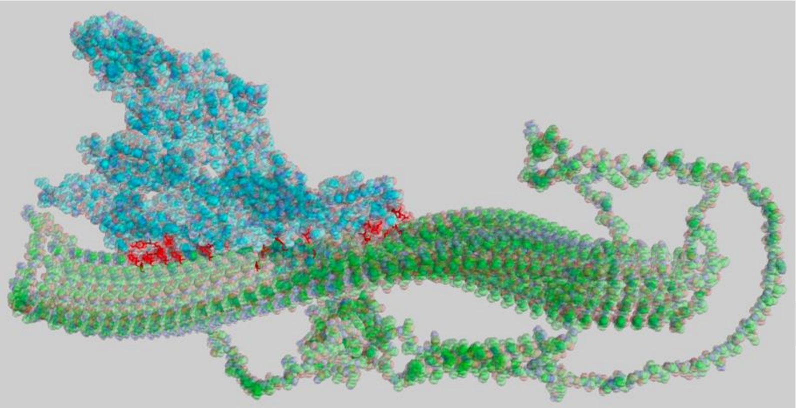
Figure I.15 Structural comparison of *S*. Enteritidis (blue). (A) With respect to 2YN3 (yellow). (B) With respect to 2YN5 (yellow).

Table I.1 RMSD values of the models and NCBI isolates

Model Number	**RMSD Value (wrt 2YN3 / 2YN5) (Å)**
1	1.267 (C) / 1.698 (B)
2	1.266 (C) / 1.684 (B)
3	1.023 (C) / 1.721 (B)
4	1.156 (C) / 1.742 (B)
5	1.026 (C) / 1.632 (B)
6	1.201 (C) / 1.767 (B)
7	1.359 (C) / 1.688 (B)
8	1.177 (C) / 1.718 (B)
9	1.151 (C) / 1.652 (B)
10	1.318 (C) / 1.633 (A)
11	1.397 (C) / 1.606 (A)
S. bongori 40:z35:-	1.091 (C) / 1.674 (B)
S. Gallinarum	1.040 (C) / 1.738 (B)
S. Pullorum	1.148 (C) / 1.657 (A)
S. Choleraesuis	1.130 (C) / 1.548 (A)
S. Enteritidis	1.065 (C) / 1.500 (A)

J. PROTEIN – PROTEIN INTERACTIONS BETWEEN SiiE BIg50:52 AND *Homo sapiens* MUC1

A



B

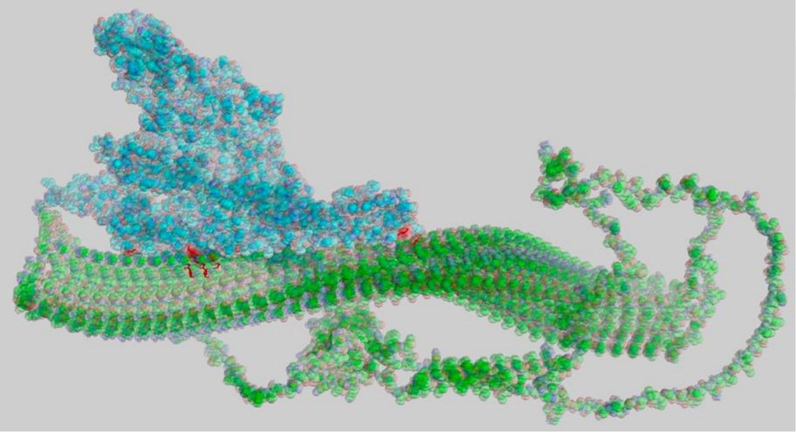
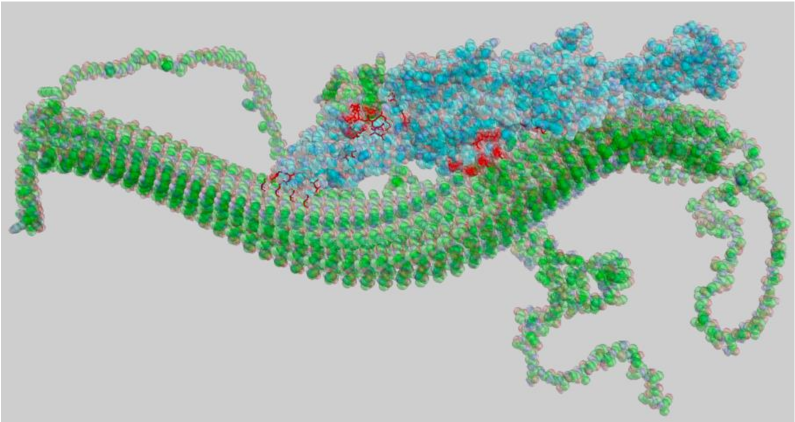
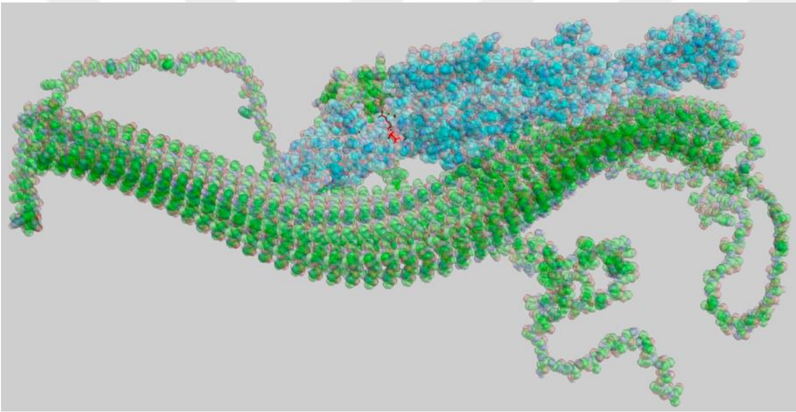


Figure J.1 PPIs (red) between 2YN3 (blue) & MUC1(AF) (green). (A) H – bonding. (B) Hydrophobic Interactions.

A



B



C

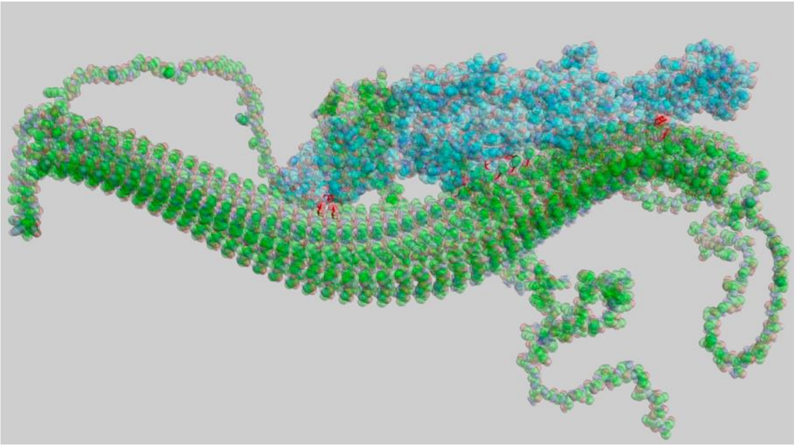
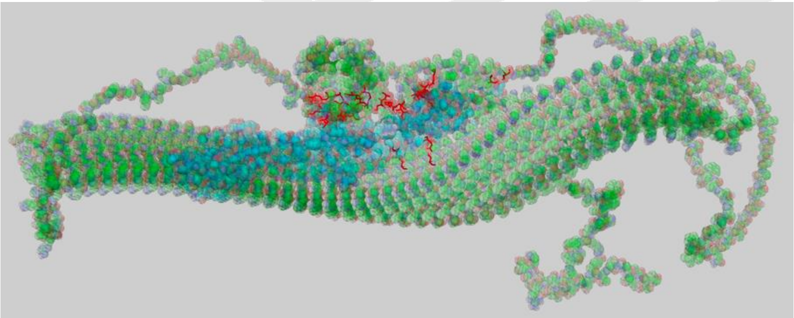
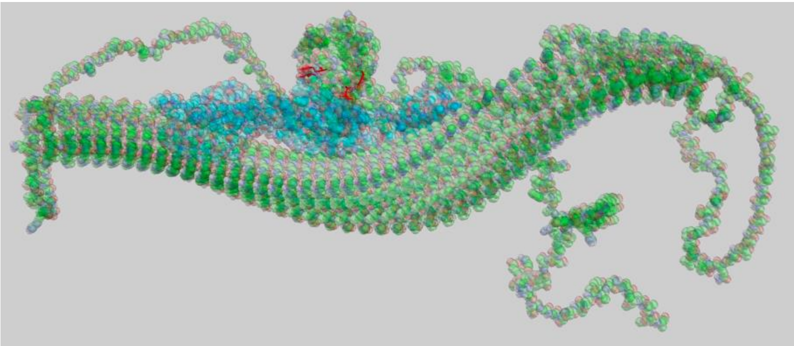


Figure J.2 PPIs (red) between 2YN5 (blue) & MUC1(AF) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B



C

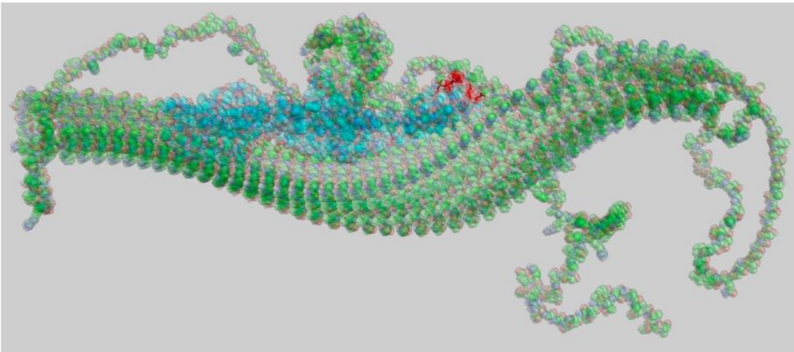
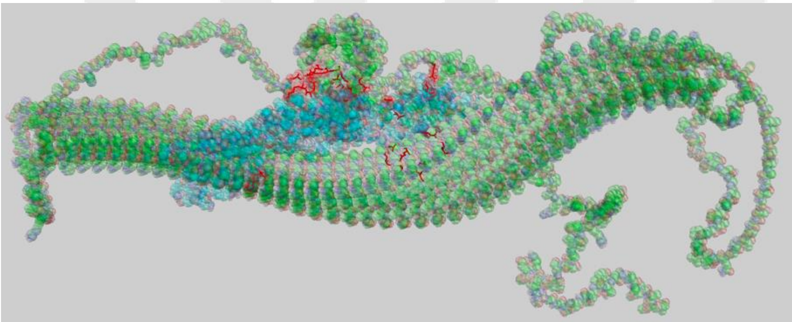
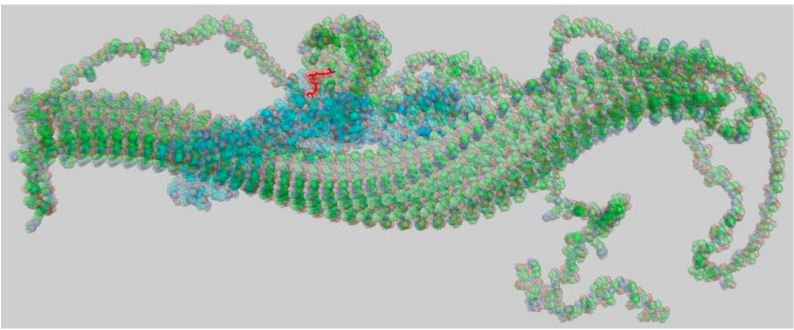


Figure J.3 PPIs (red) between 2YN3 Chain C (blue) & MUC1(AF) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B



C

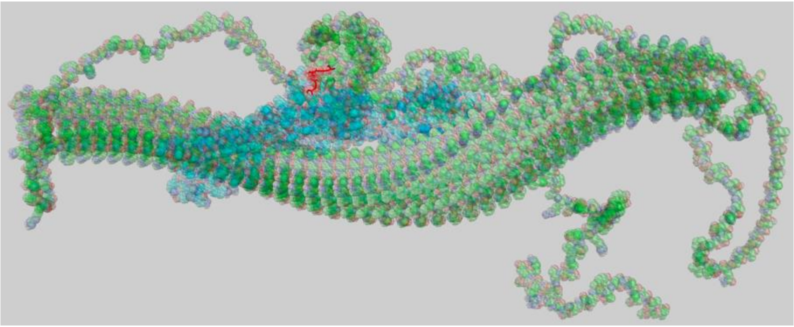


Figure J.4 PPIs (red) between 2YN5 Chain B (blue) & MUC1(AF) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A

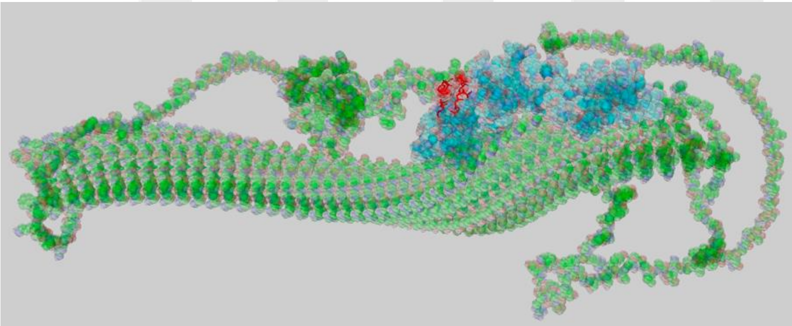
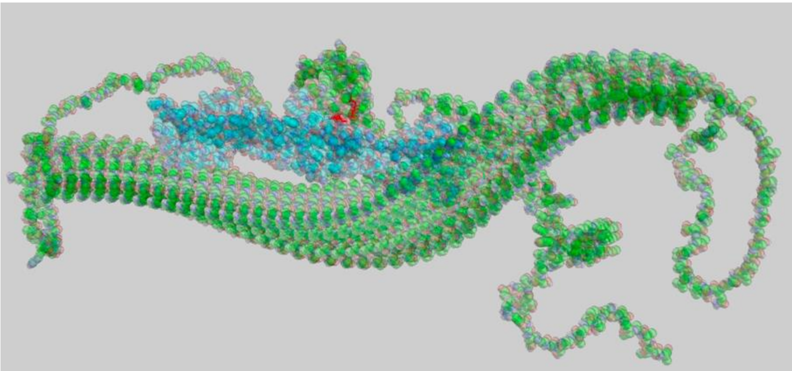


Figure J.5 PPIs (red) between Model 1 (blue) & MUC1(AF) (green). (A) Hydrophobic Interactions.

A



B

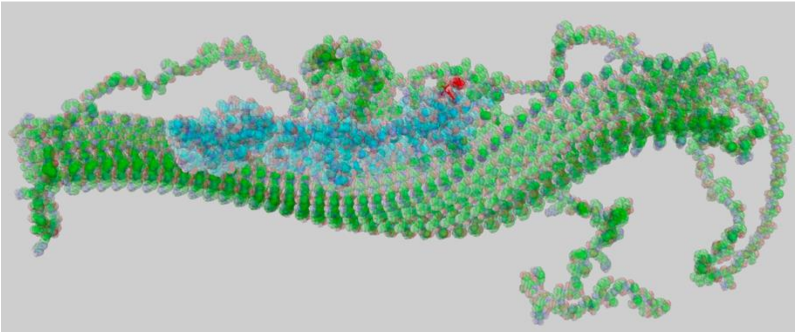
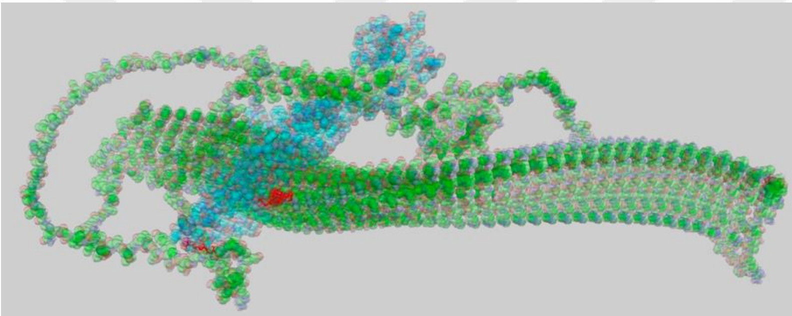


Figure J.6 PPIs (red) between Model 2 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

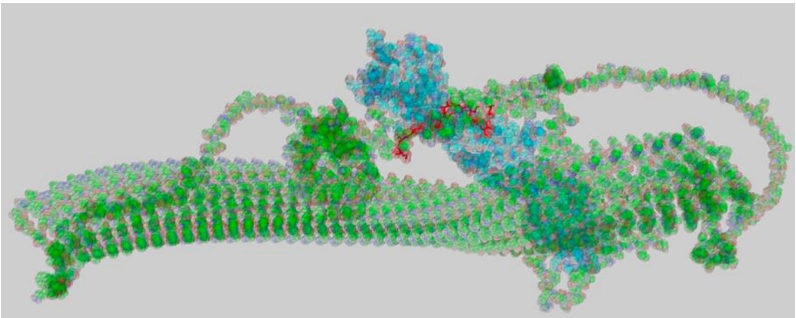


Figure J.7 PPIs (red) between Model 3 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A

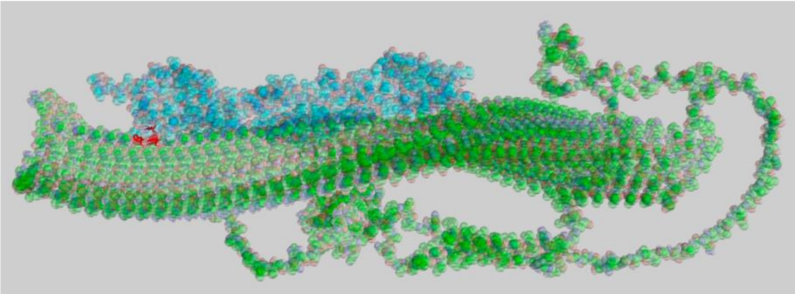
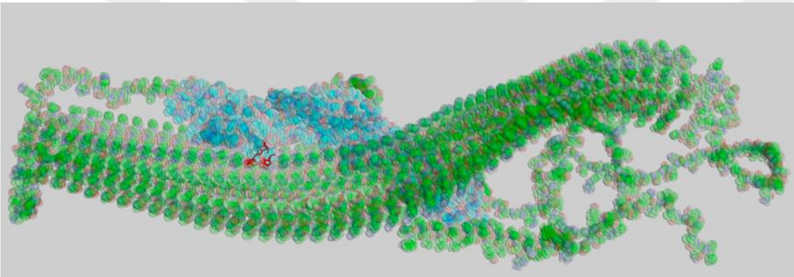


Figure J.8 PPIs (red) between Model 4 (blue) & MUC1(AF) (green). (A) Salt Bridges.

A



B

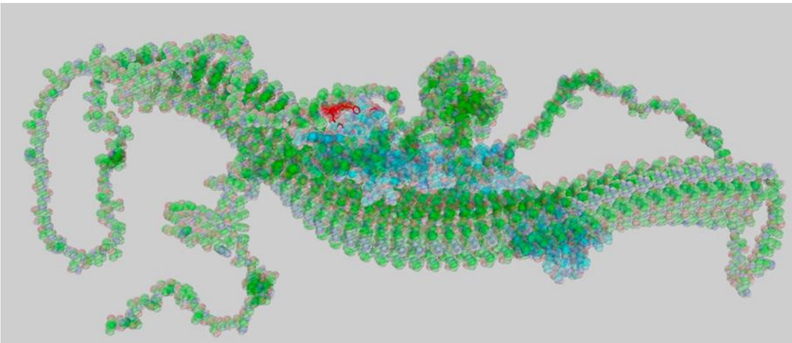
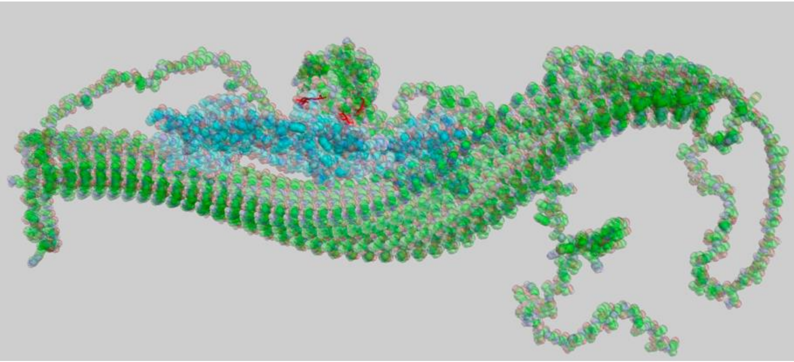


Figure J.9 PPIs (red) between Model 5 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

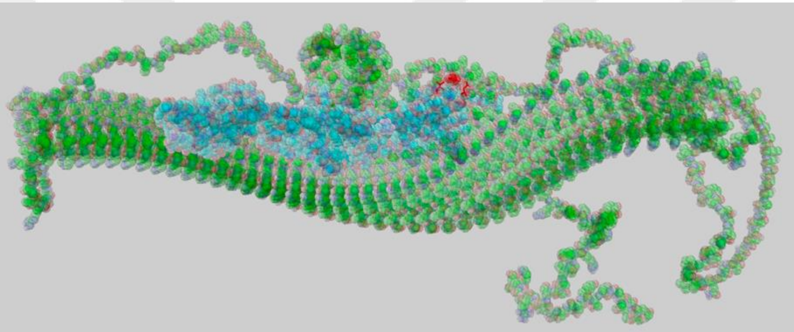


Figure J.10 PPIs (red) between Model 6 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A

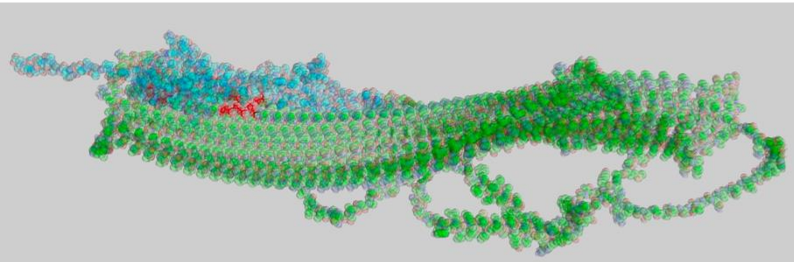
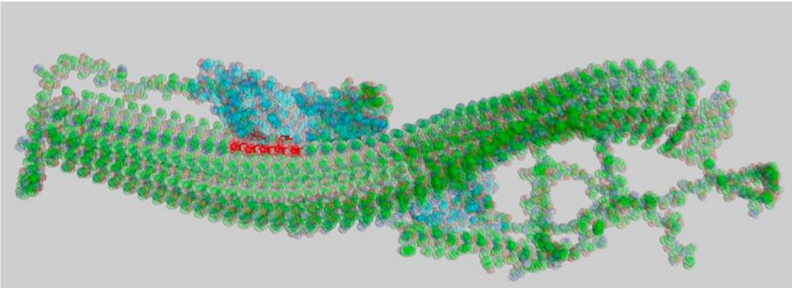


Figure J.11 PPIs (red) between Model 7 (blue) & MUC1(AF) (green). (A) Salt Bridges.

A



B

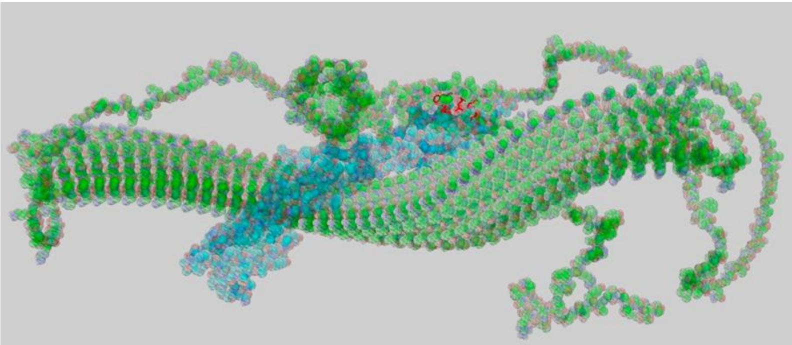
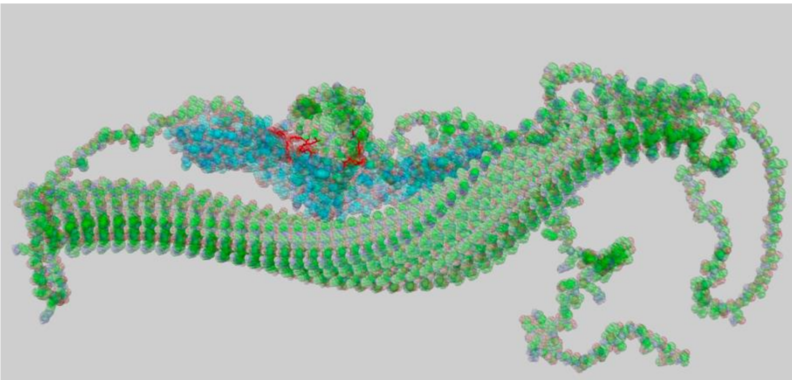


Figure J.12 PPIs (red) between Model 8 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

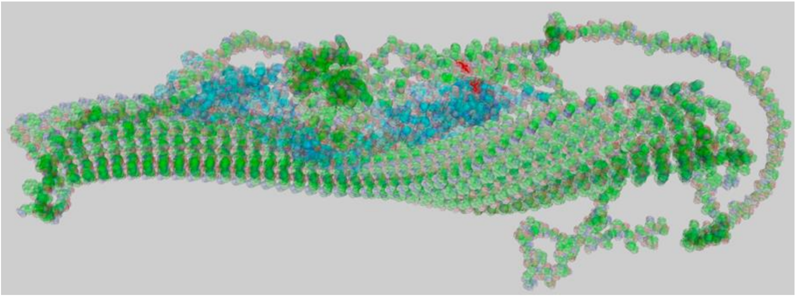
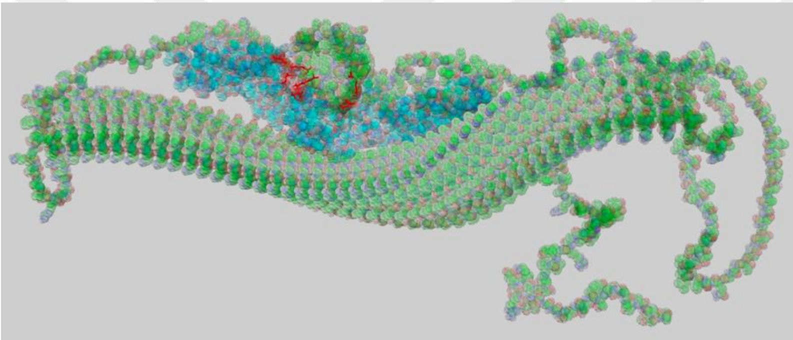


Figure J.13 PPIs (red) between Model 9 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

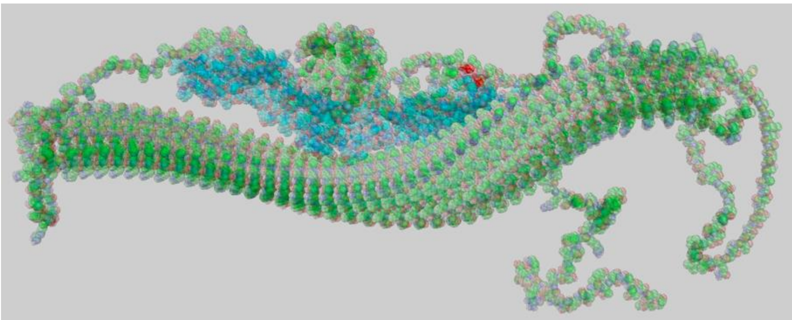
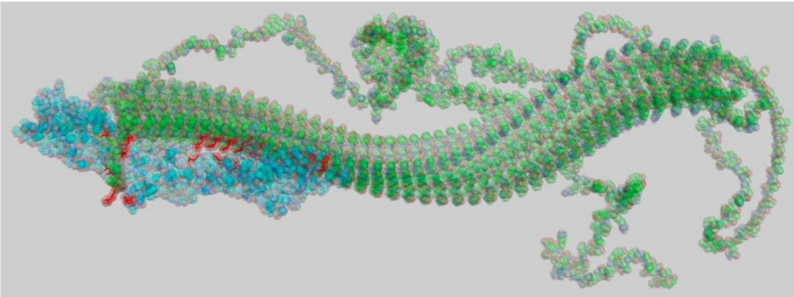


Figure J.14 PPIs (red) between Model 10 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

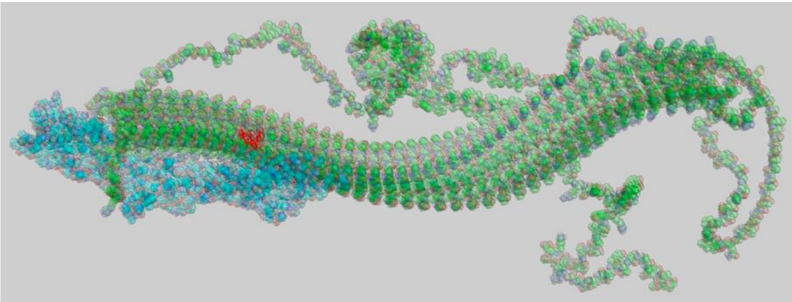
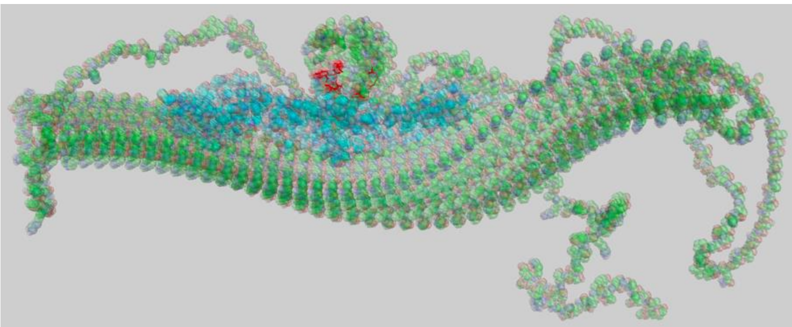


Figure J.15 PPIs (red) between Model 11 (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

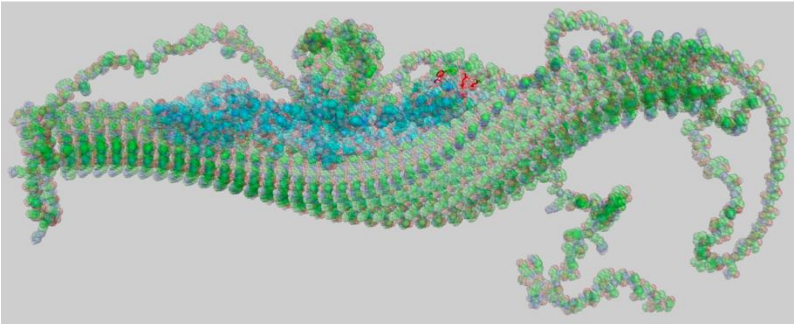
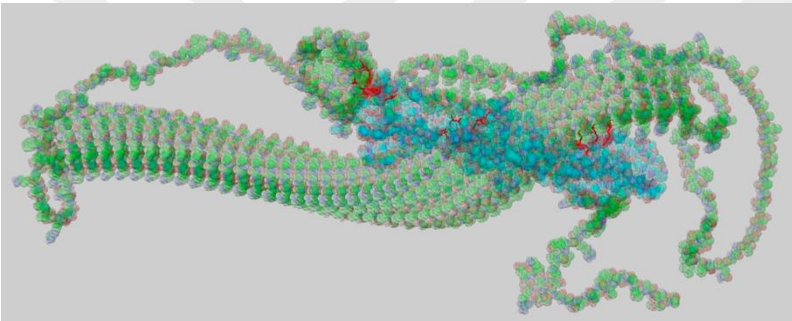


Figure J.16 PPIs (red) between *S. bongori* (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

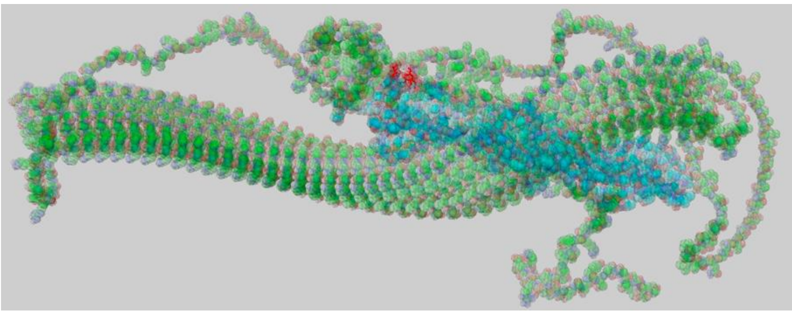


Figure J.17 PPIs (red) between *S*. Gallinarum (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A

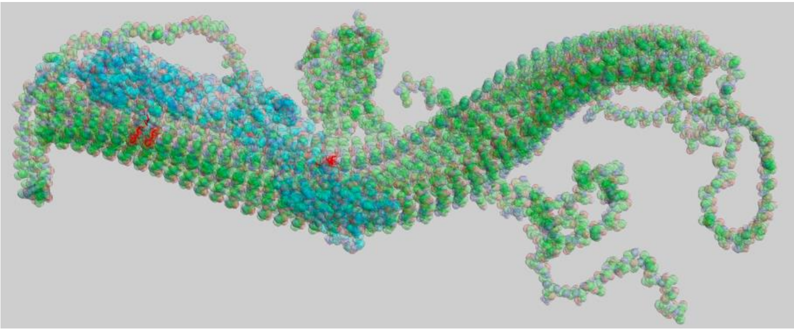
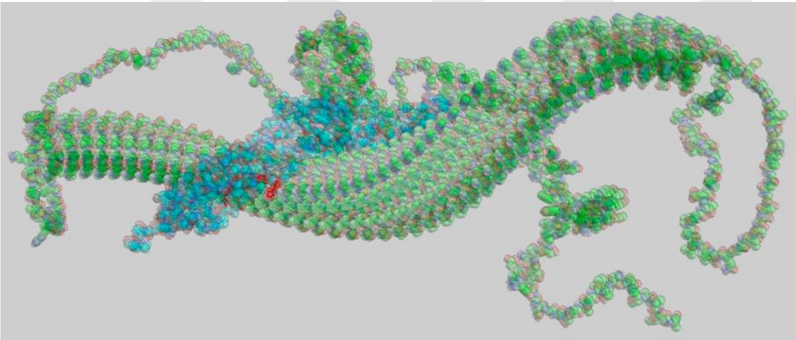


Figure J.18 PPIs (red) between *S*. Pullorum (blue) & MUC1(AF) (green). (A) Salt Bridges.

A



B

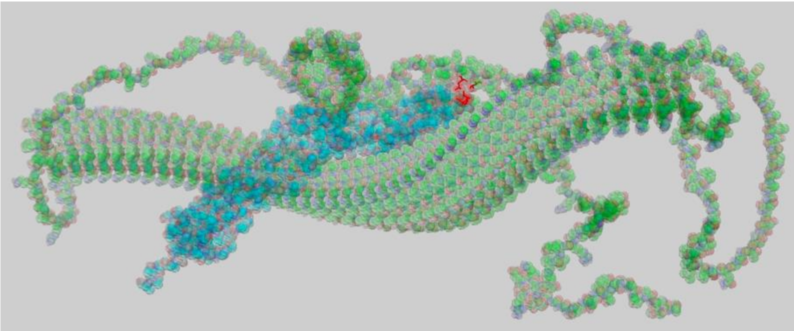
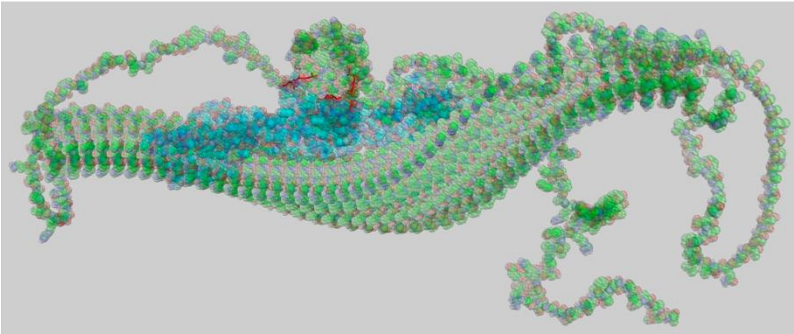


Figure J.19 PPIs (red) between *S*. Choleraesuis (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

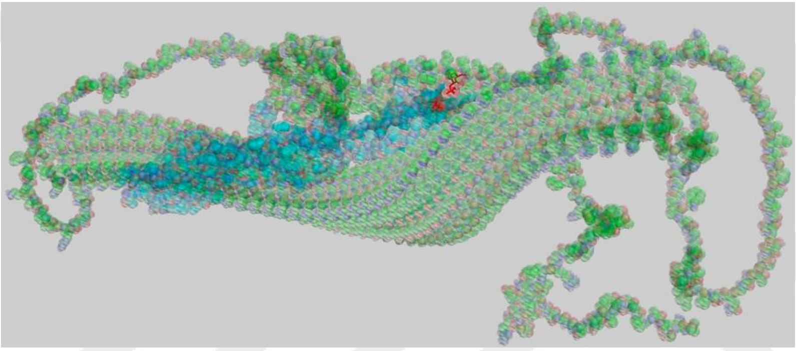
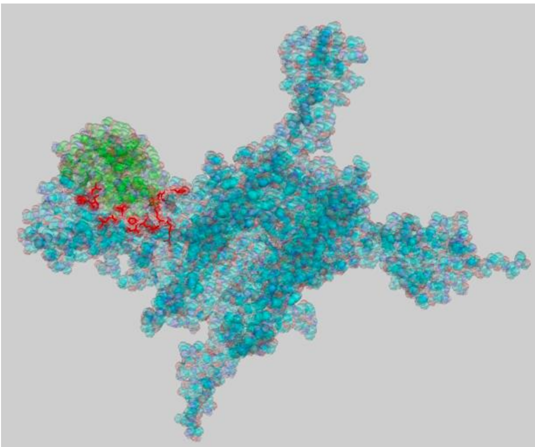


Figure J.20 PPIs (red) between *S*. Enteritidis (blue) & MUC1(AF) (green). (A) Salt Bridges. (B) Hydrophobic Interactions.

A



B

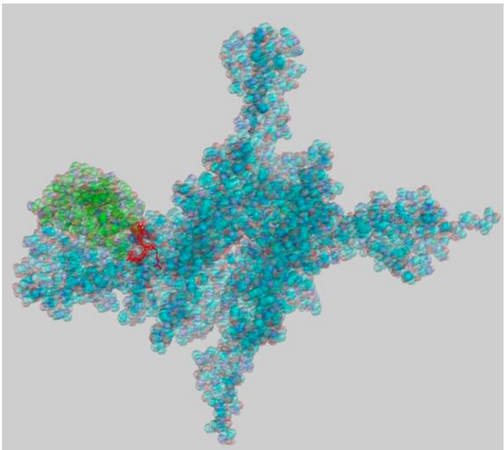
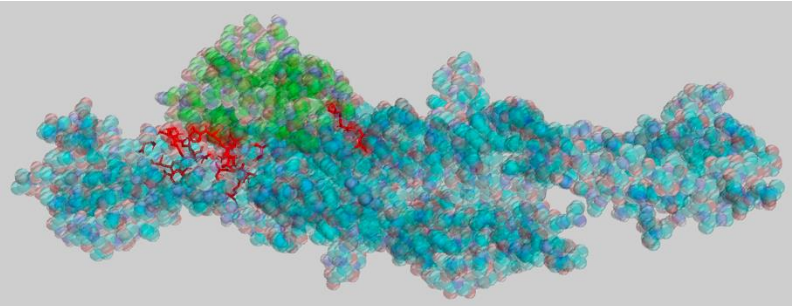


Figure J.21 PPIs (red) between 2YN3 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges

A



B

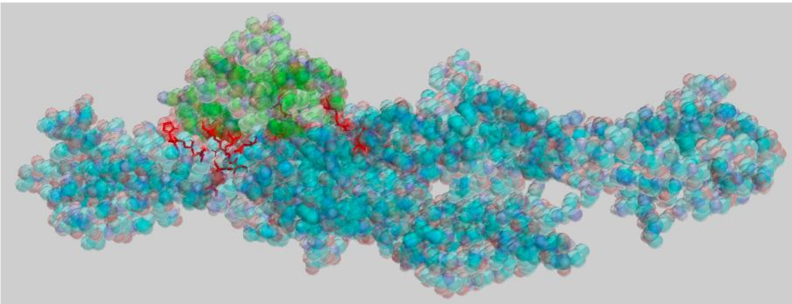
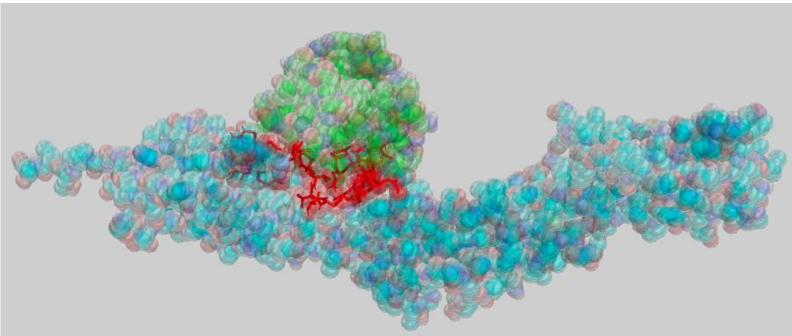
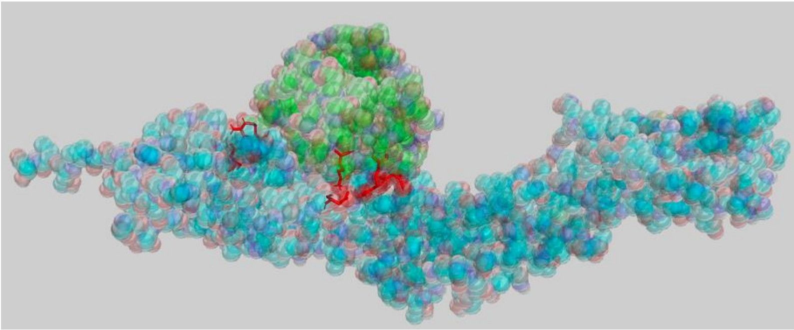


Figure J.22 PPIs (red) between 2YN5 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges

A



B



C

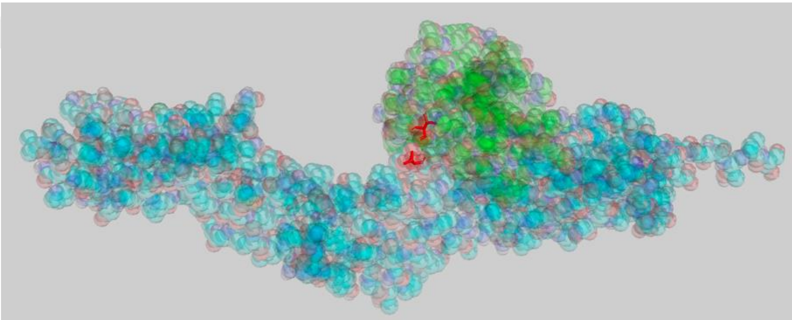
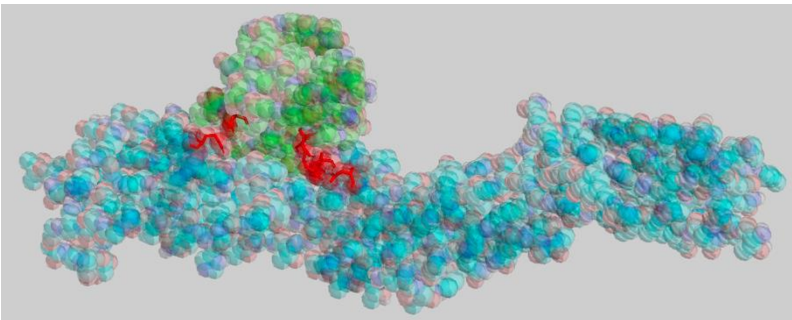


Figure J.23 PPIs (red) between 2YN3 Chain C (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B

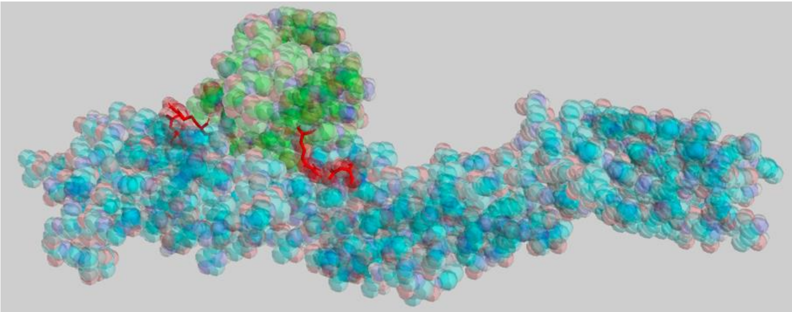
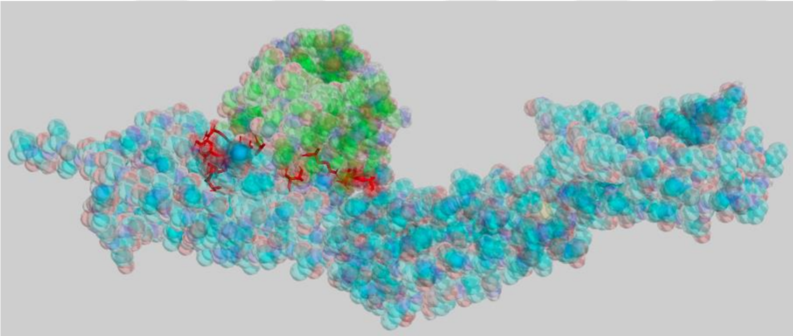


Figure J.24 PPIs (red) between 2YN5 Chain B (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B

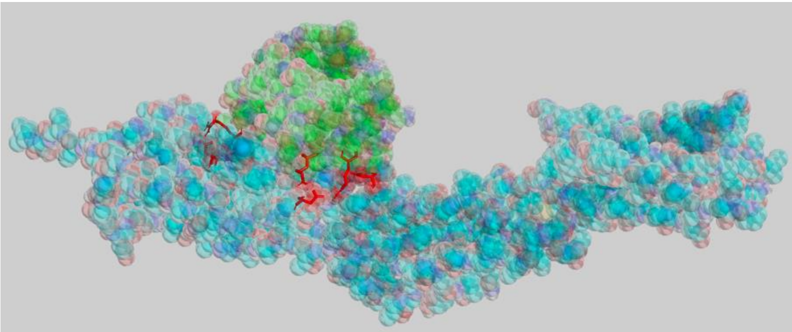
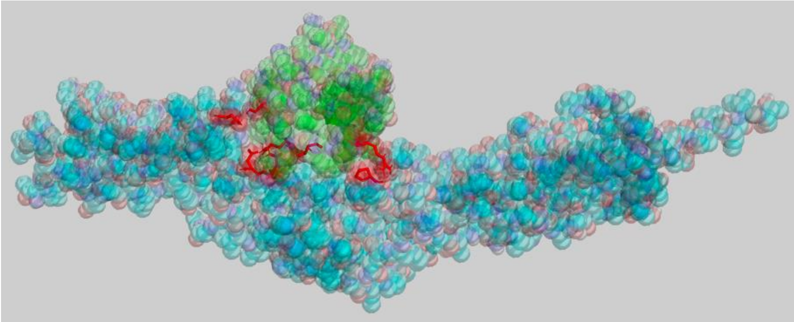
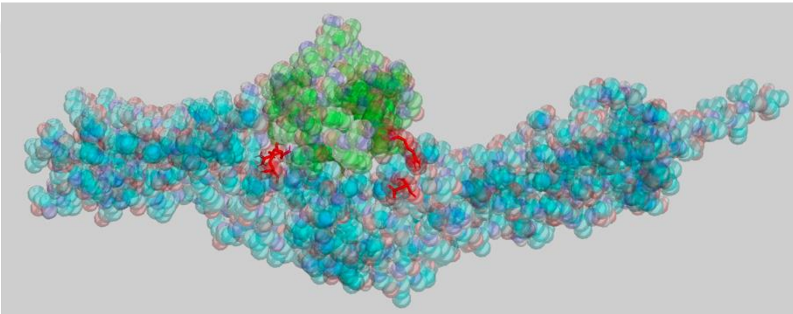


Figure J.25 PPIs (red) between Model 1 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B



C

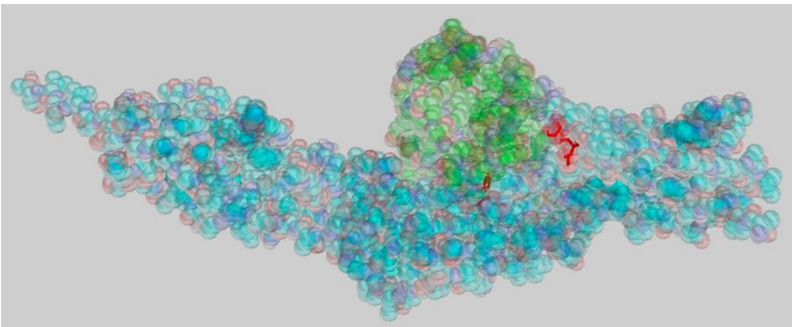
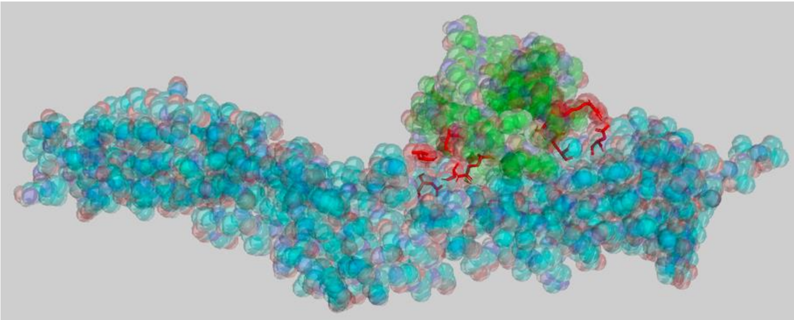


Figure J.26 PPIs (red) between Model 2 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B

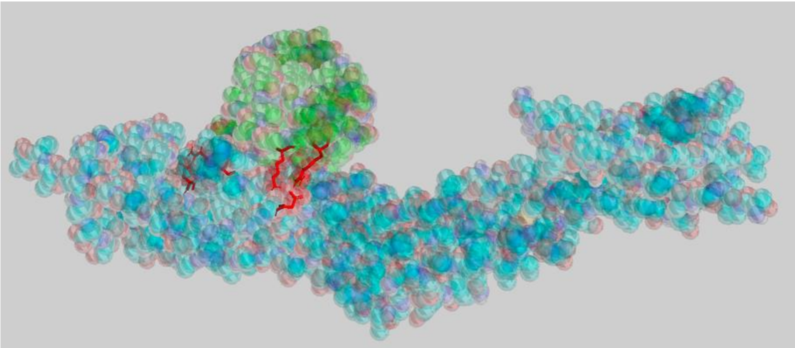
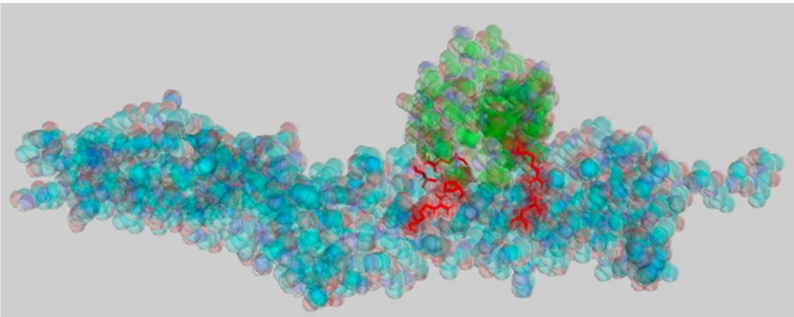


Figure J.27 PPIs (red) between Model 3 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B

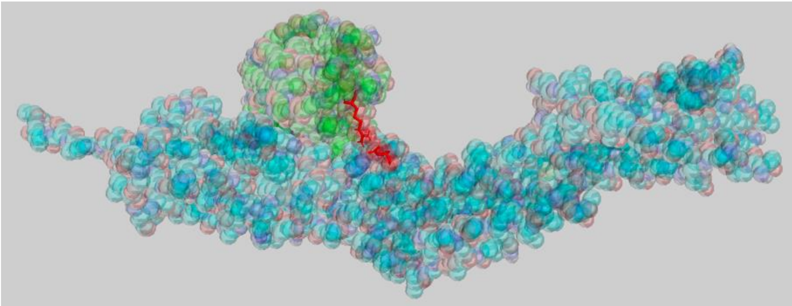
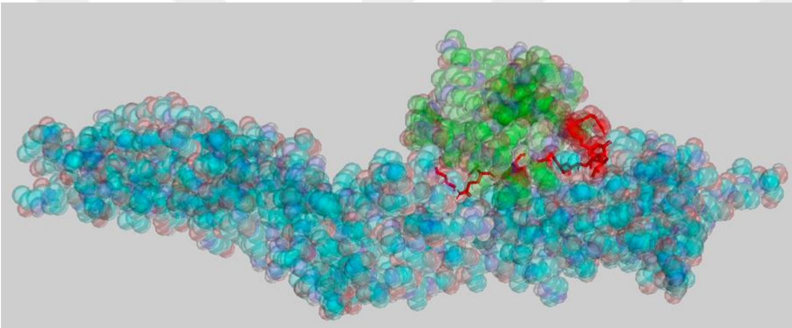
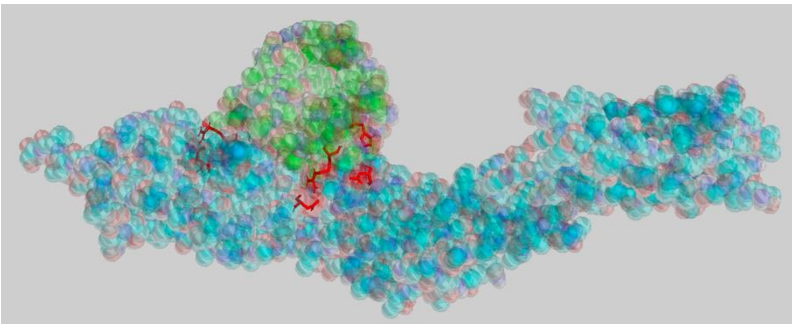


Figure J.28 PPIs (red) between Model 4 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B



C

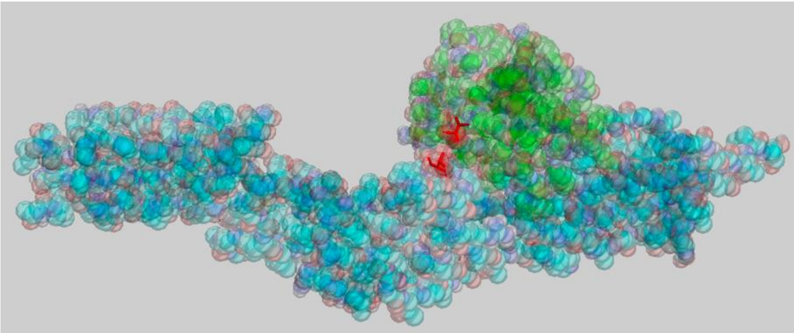
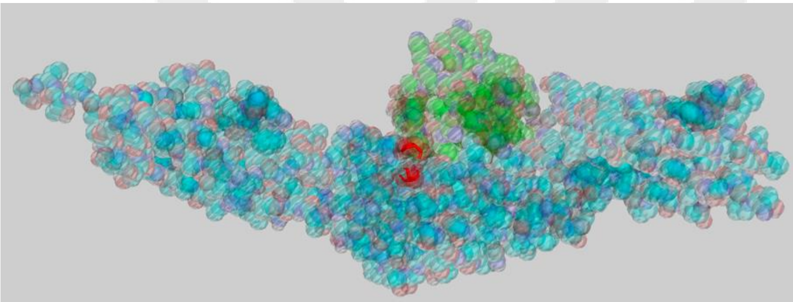
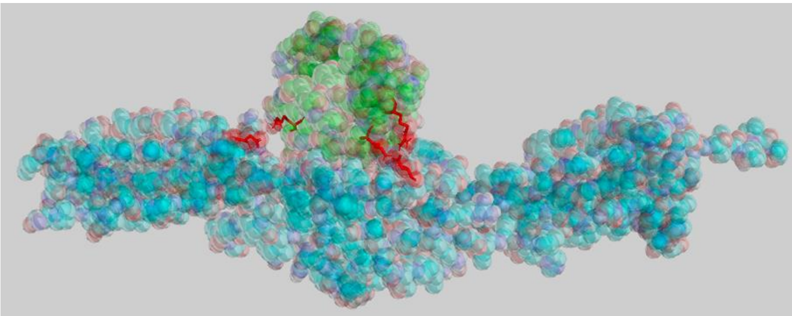


Figure J.29 PPIs (red) between Model 5 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B



C

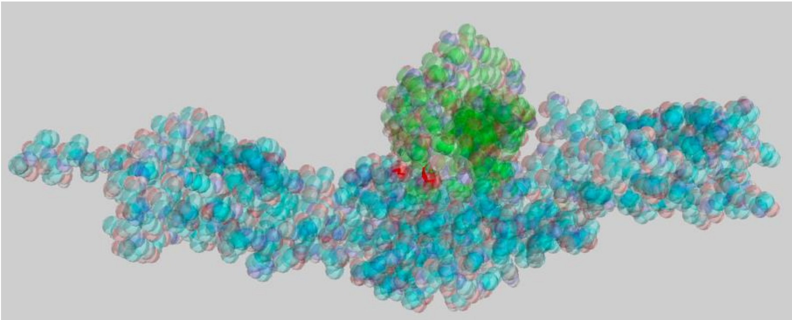
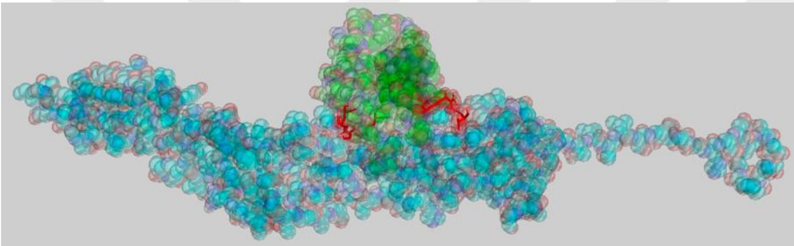
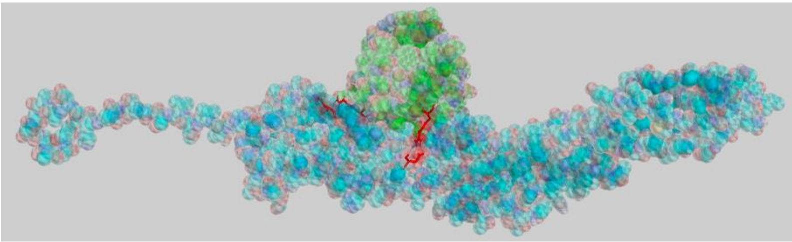


Figure J.30 PPIs (red) between Model 6 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B



C

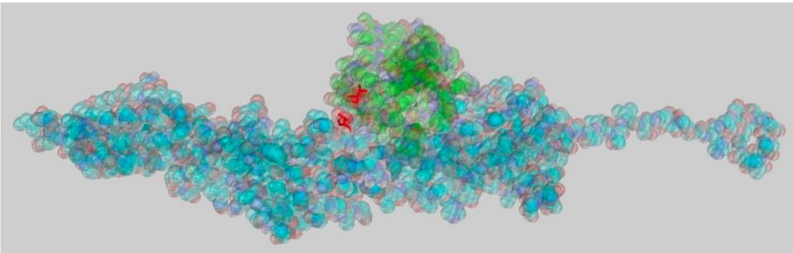
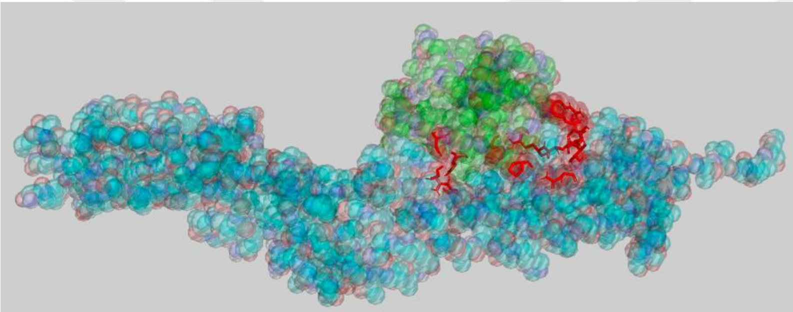


Figure J.31 PPIs (red) between Model 7 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B

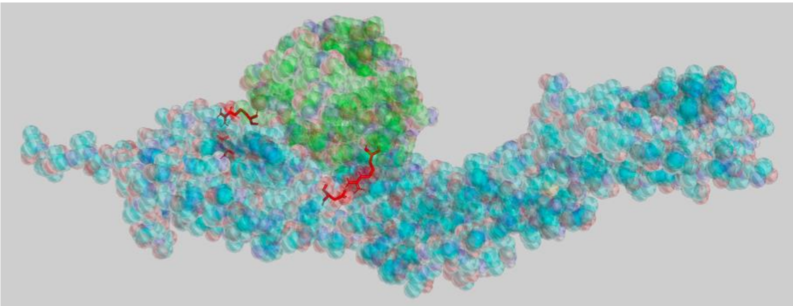
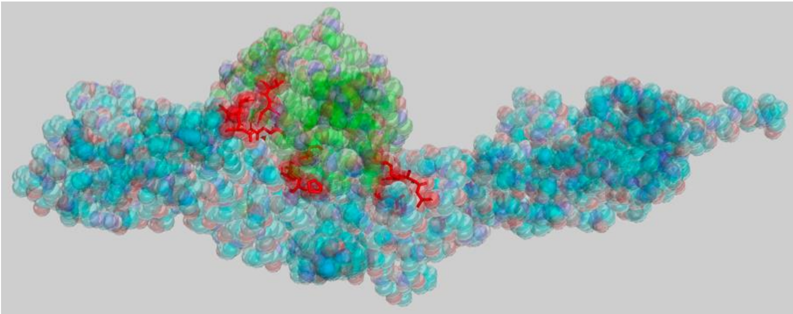
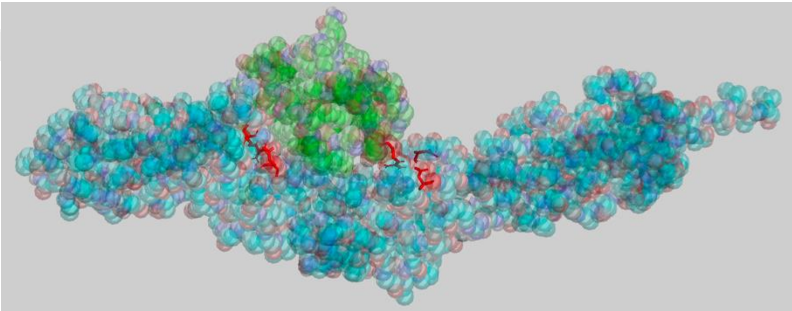


Figure J.32 PPIs (red) between Model 8 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B



C

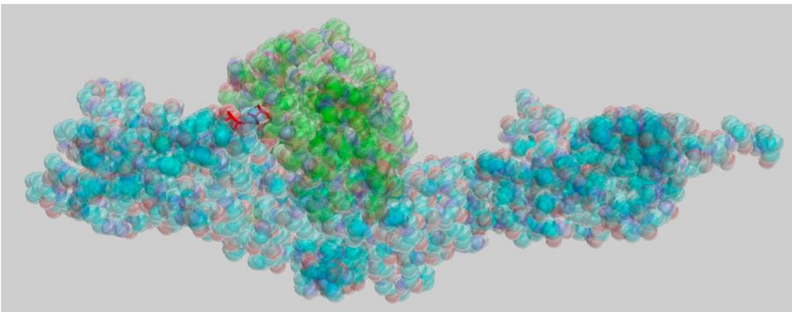
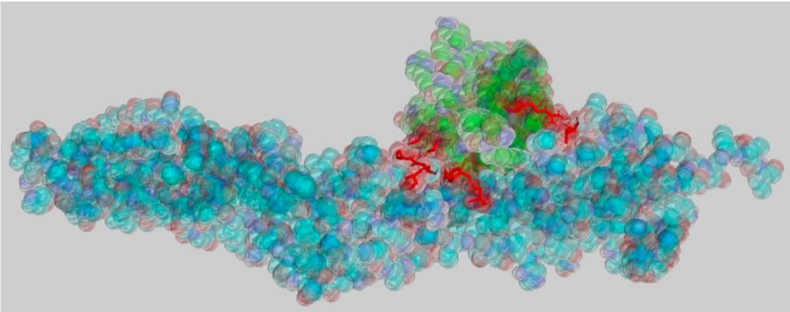


Figure J.33 PPIs (red) between Model 9 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B

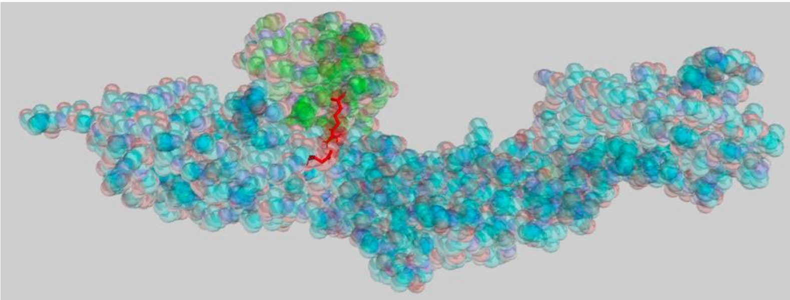
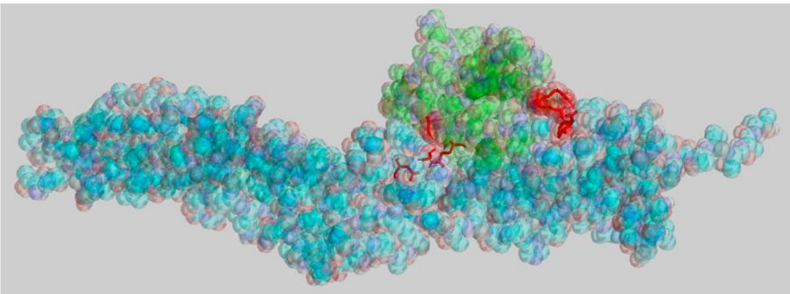


Figure J.34 PPIs (red) between Model 10 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges

A



B

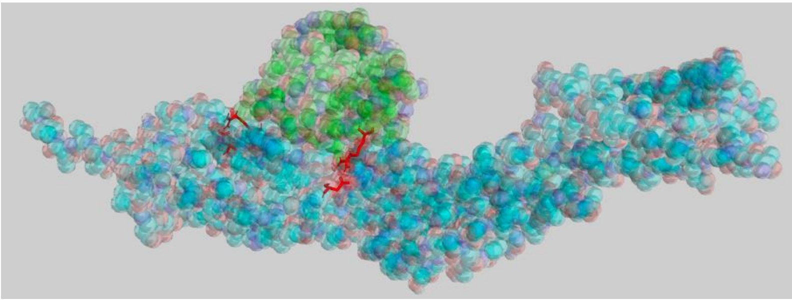
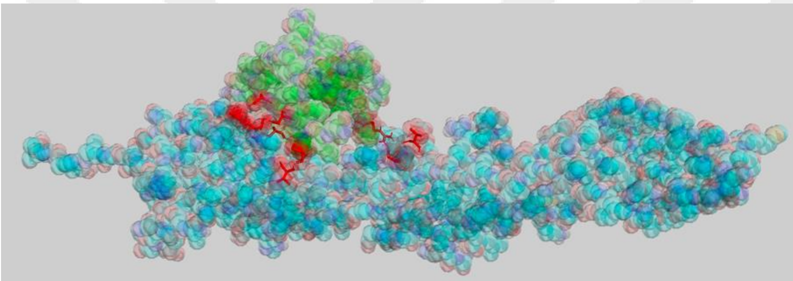


Figure J.35 PPIs (red) between Model 11 (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges

A



B

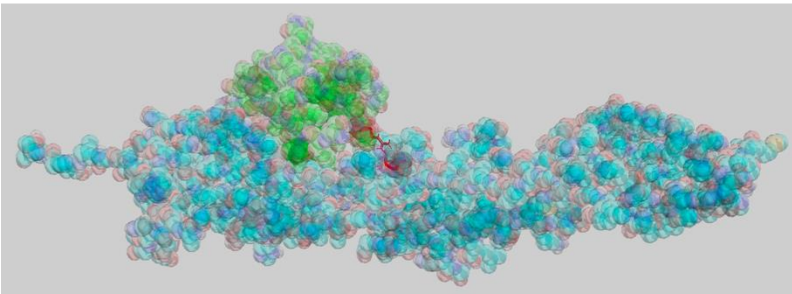
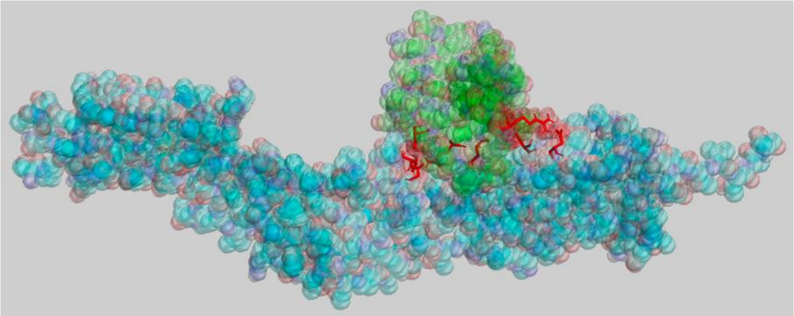
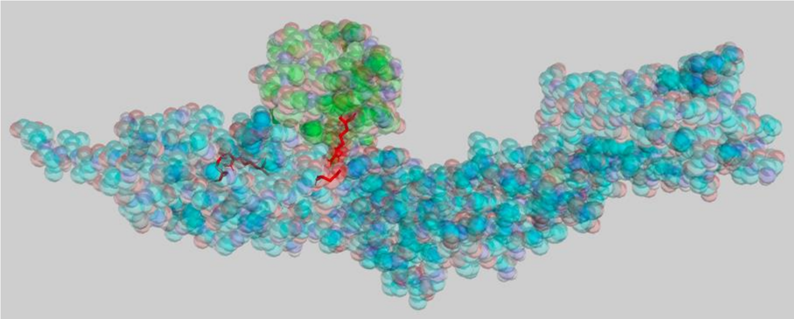


Figure J.36 PPIs (red) between *S. bongori* (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges

A



B



C

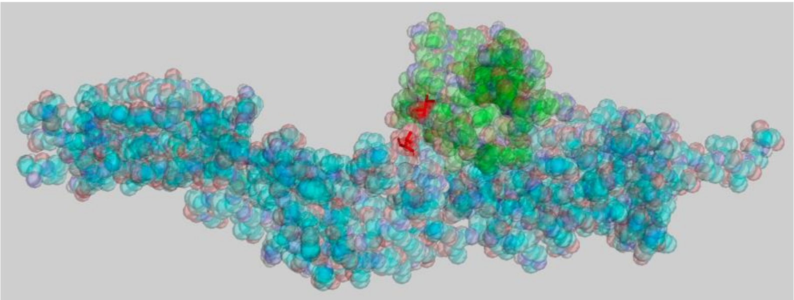
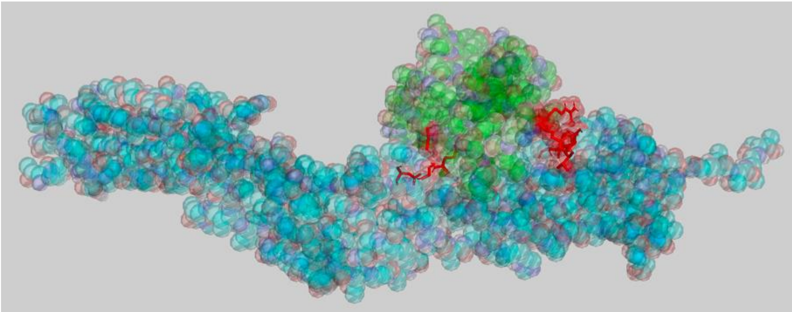


Figure J.37 PPIs (red) between *S.* Gallinarum (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B

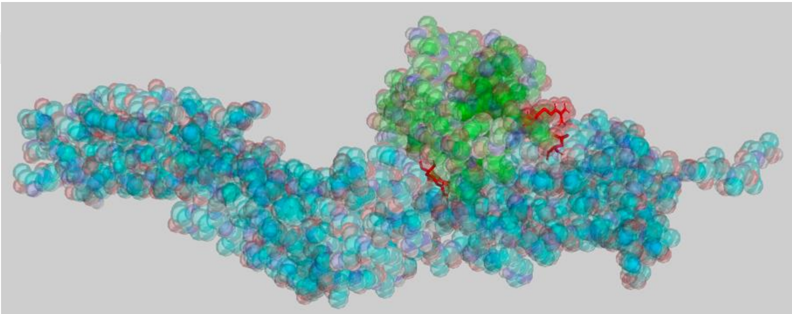
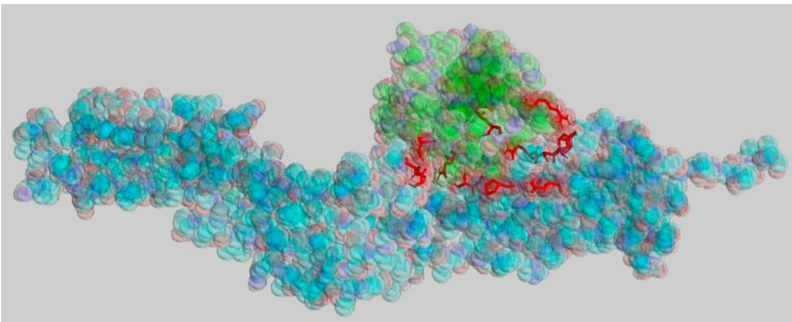
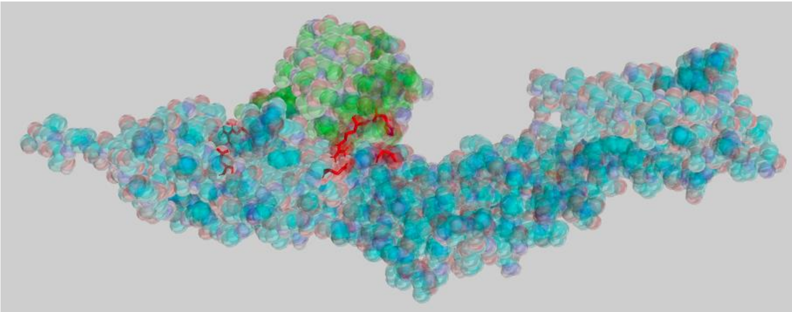


Figure J.38 PPIs (red) between *S*. Pullorum (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges.

A



B



C

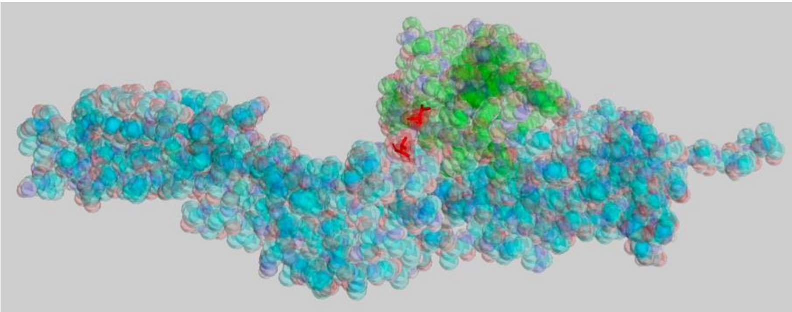
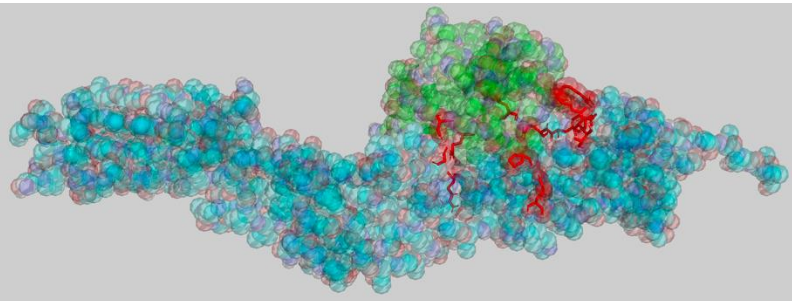
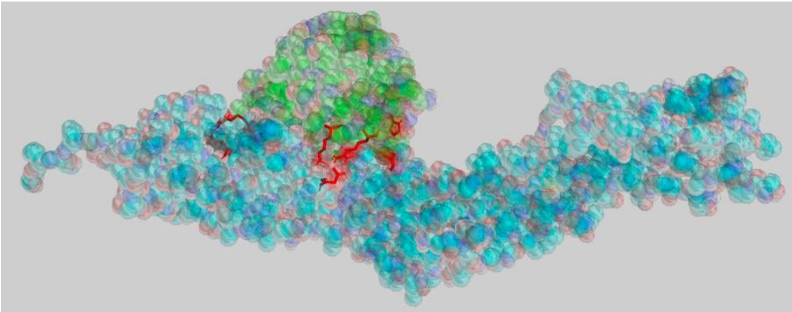


Figure J.39 PPIs (red) between *S*. Choleraesuis (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

A



B



C

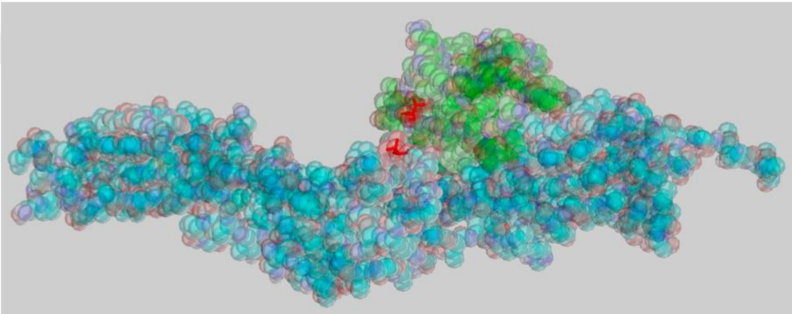
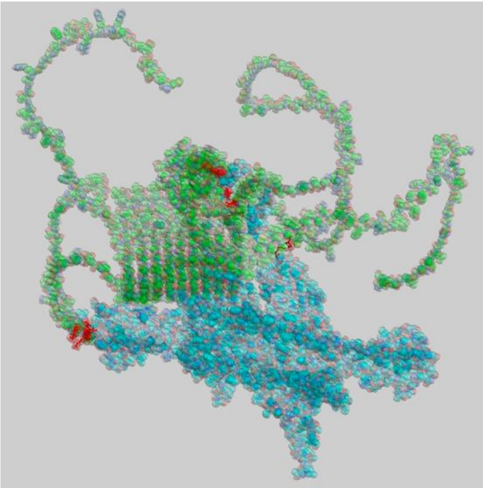


Figure J.40 PPIs (red) between *S*. Enteritidis (blue) & MUC1(PDB) (green). (A) H – bonding. (B) Salt Bridges. (C) Hydrophobic Interactions.

K. PROTEIN – PROTEIN INTERACTIONS BETWEEN SiiE BIg50:52 AND *Sus scrofa* Muc1

A



B

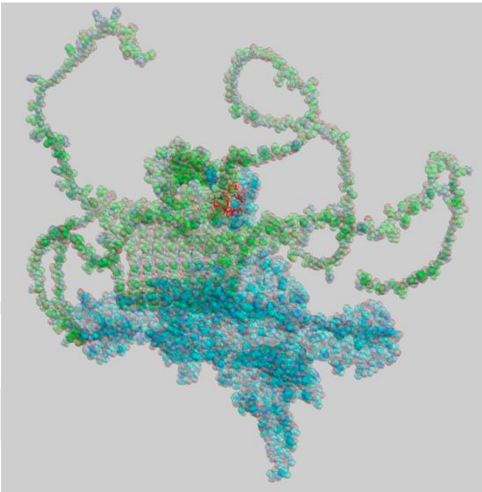


Figure K.1 PPIs (red) between 2YN3 (blue) & Muc1(AF) (green). (A) H – bonding. (B) Hydrophobic Interactions.

A

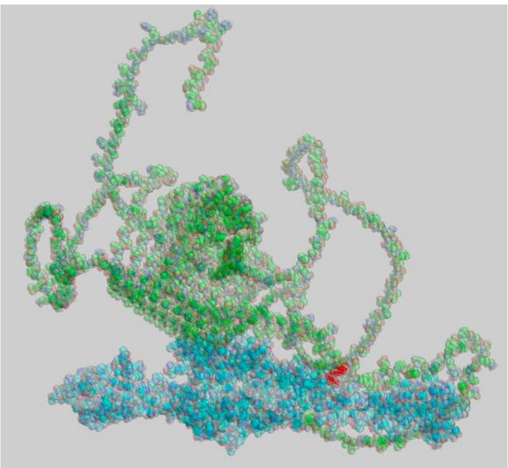
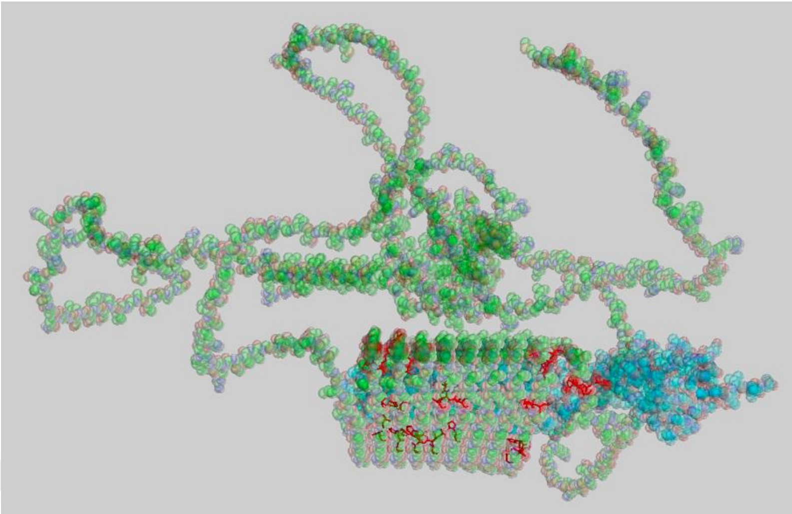


Figure K.2 PPIs (red) between 2YN5 (blue) & Muc1(AF) (green). (A) Salt bridges

A



B

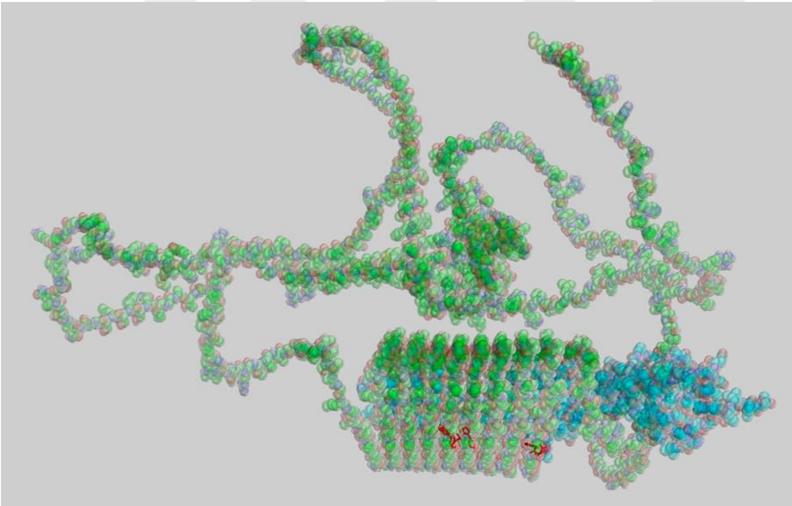
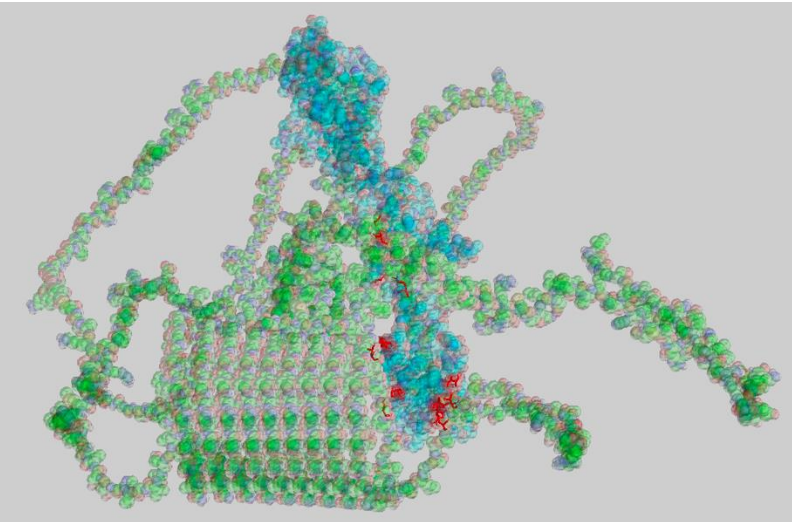


Figure K.3 PPIs (red) between 2YN3 Chain C (blue) & Muc1(AF) (green). (A) H – bonding. (B) Salt bridges.

A



B

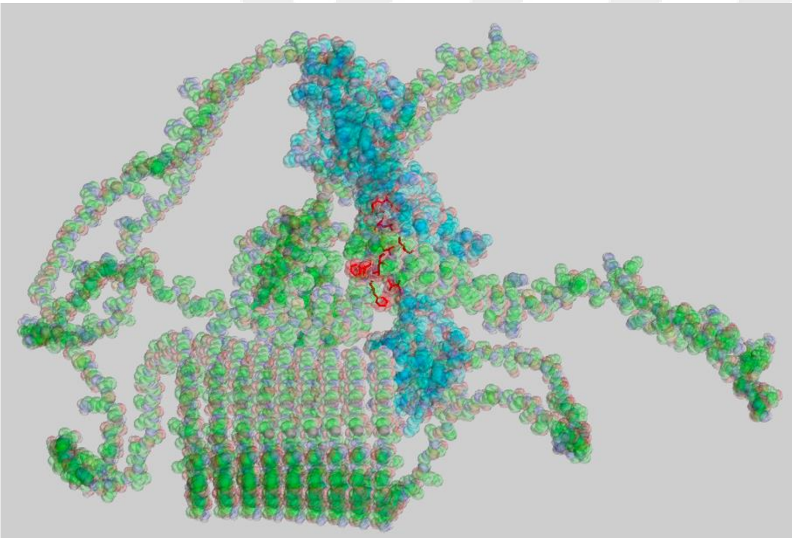


Figure K.4 PPIs (red) between 2YN5 Chain B (blue) & Muc1(AF) (green). (A) H – bonding. (B) Hydrophobic Interactions.

A

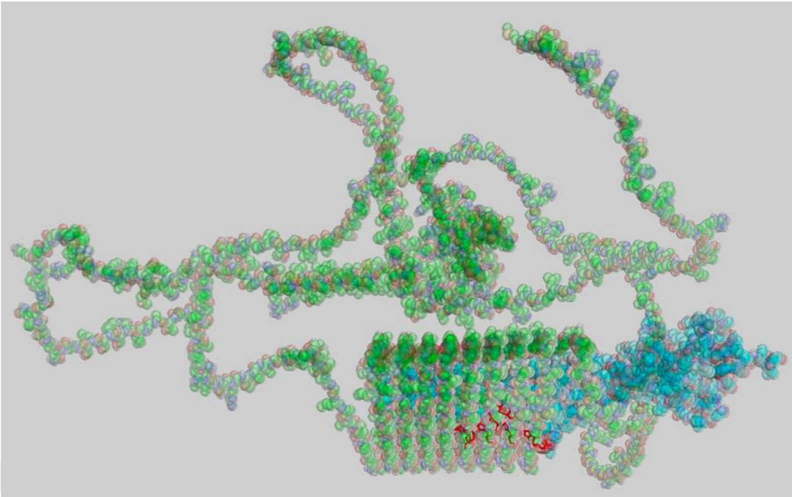
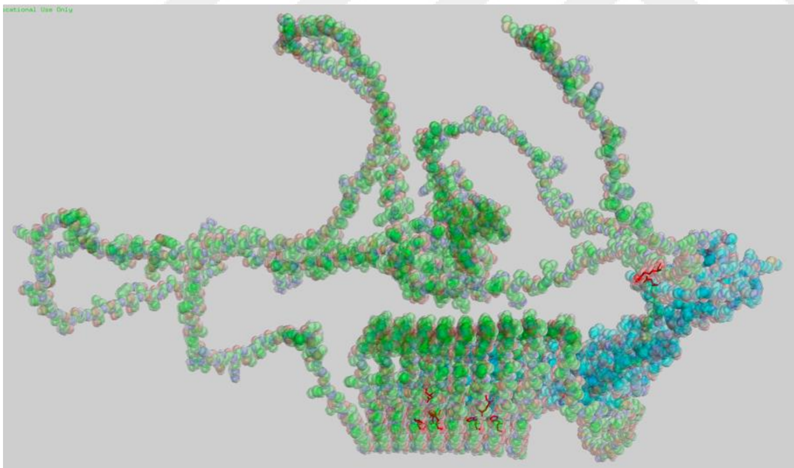


Figure K.5 PPIs (red) between Model 8 (blue) & Muc1(AF) (green). (A) Salt bridges

A



B

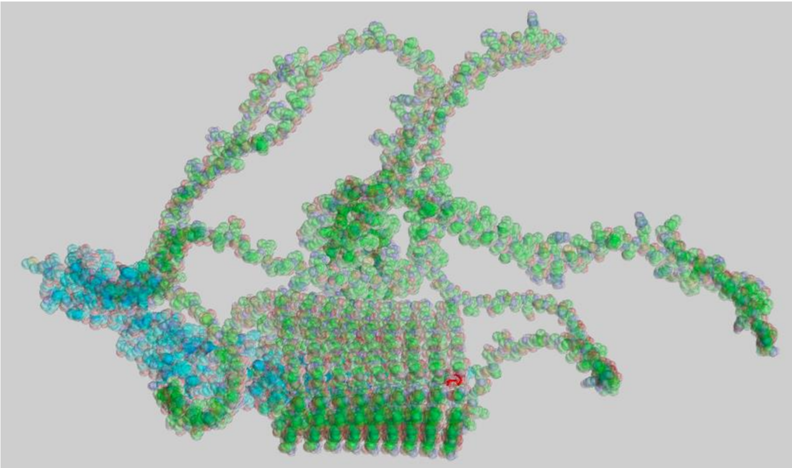
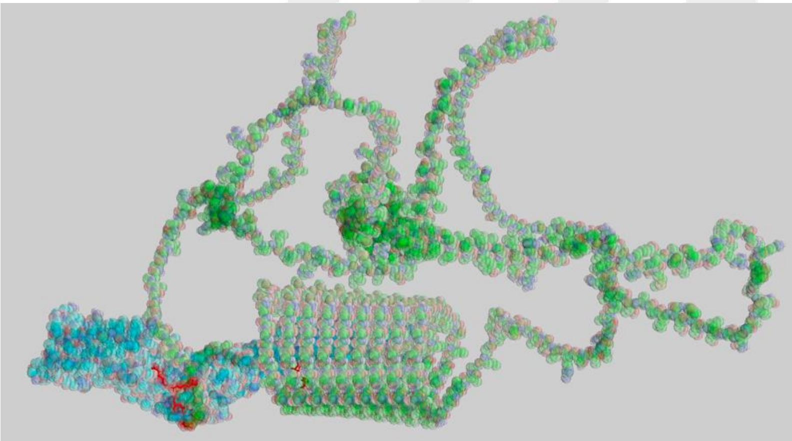


Figure K.6 PPIs (red) between *S. bongori* (blue) & Muc1(AF) (green). (A) Salt bridges. (B) Hydrophobic Interactions.

A



B

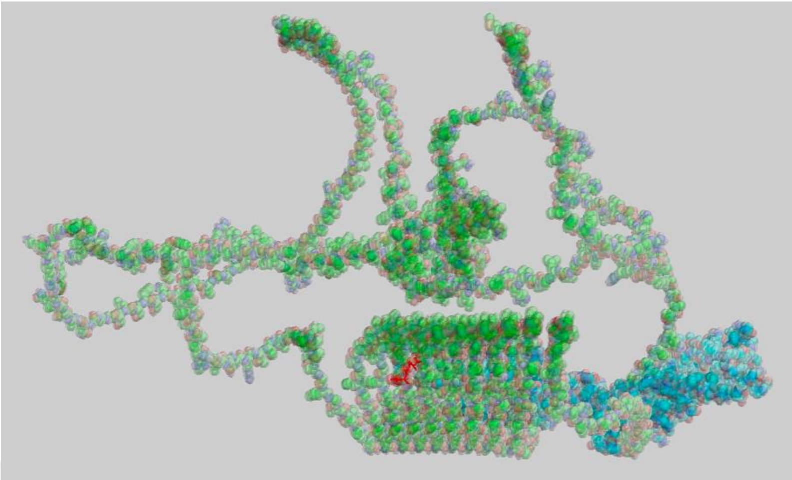


Figure K.7 PPIs (red) between *S*. Gallinarum (blue) & Muc1(AF) (green). (A) Salt bridges. (B) Hydrophobic Interactions.

A

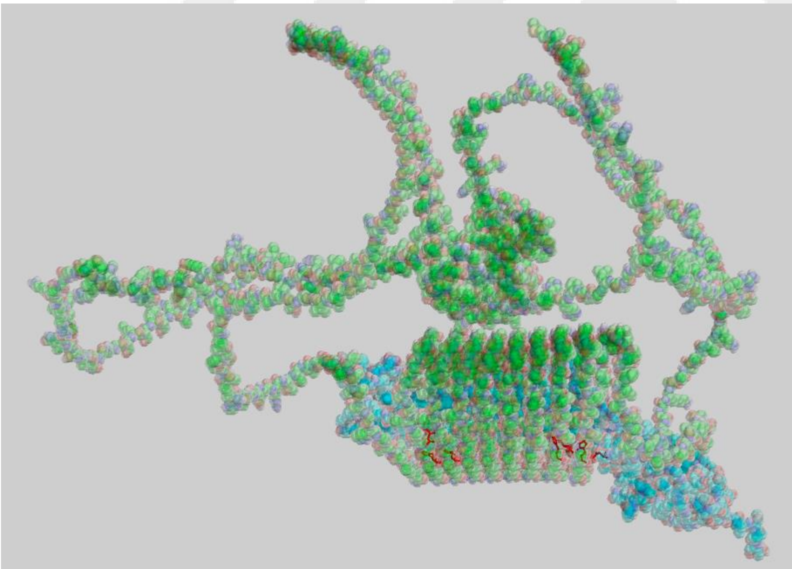


Figure K.8 PPIs (red) between *S*. Pullorum (blue) & Muc1(AF) (green). (A) Salt bridges.

A

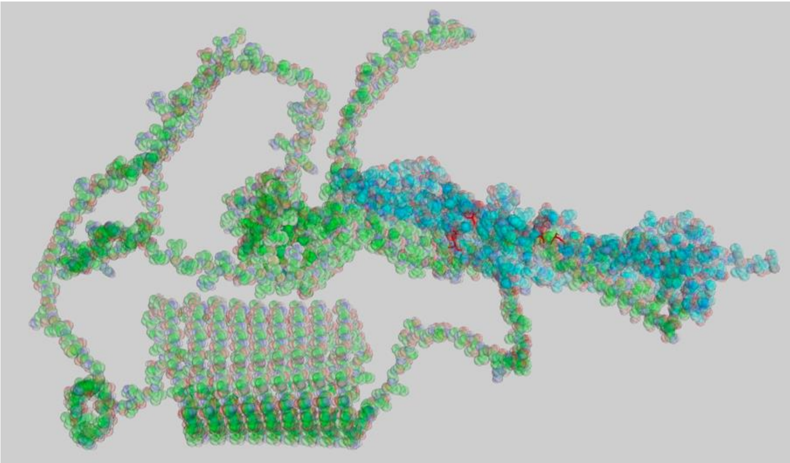


Figure K.9 PPIs (red) between *S*. Choleraesuis (blue) & Muc1(AF) (green). (A) Salt bridges.

A

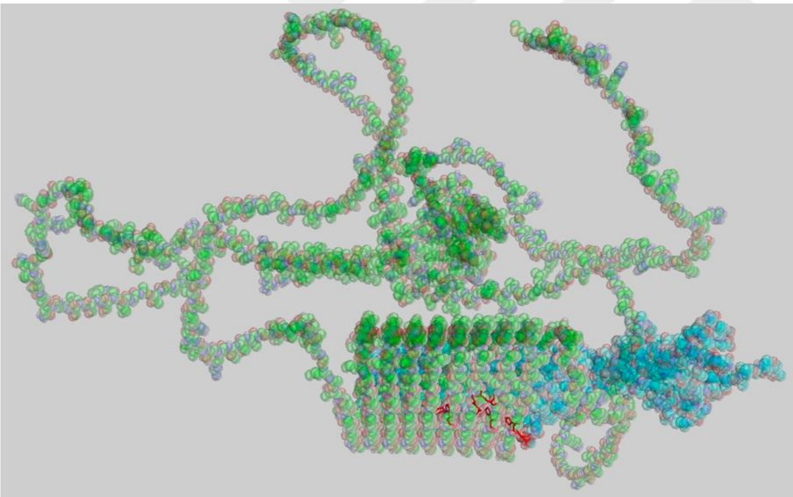


Figure K.10 PPIs (red) between *S*. Enteritidis (blue) & Muc1(AF) (green). (A) Salt bridges.