MOLECULAR BASIS OF COMPLEMENT FACTOR-H RECRUITMENT BY THE LYME DISEASE PATHOGEN *BORRELIA BURGDORFERI*

By

Jagannath Silwal

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2018

ABSTRACT

MOLECULAR BASIS OF COMPLEMENT FACTOR-H RECRUITMENT BY THE LYME DISEASE
PATHOGEN *BORRELIA BURGDORFERI*

By

Jagannath Silwal

Lyme disease is the most common vector borne illness around the globe, caused by
species of spirochetes in the *Borrelia* genus*.* In the United States, *Borrelia burgdorferi* sensu
stricto (hereafter termed *B. burgdorferi*) is the major causative agent of Lyme disease, whereas
Borrelia afzelii, *Borrelia garinii* and many other related *Borrelia* species are major pathogens
responsible for Lyme disease in other parts of the world. *B. burgdorferi* is transmitted by hard-
bodied tick species mainly from genus *Ixodes* and, like many other infectious pathogens, it has
evolved many mechanisms to circumvent a highly sophisticated and tightly regulated host
immune system, causing persistent infection. One of the key mechanisms that allows *Borrelia* to
cause such infection is its ability to manipulate the host's complement system by recruiting host
regulator proteins to its own cell surface.

The complement system is an integral part of the innate immune system, tightly
regulated via interactions of several regulator proteins. The complement factor H (FH) protein is
one of the key complement regulator proteins and plays a major role in avoiding complement
attack on self-cells. FH consists of 20 complement control protein (CCP) domains and inhibits
both complement activation and amplification on host cells by promoting inhibition and
degradation of enzymes associated with innate immunity. *B. burgdorferi* recruits FH protein to
its surface by expressing multiple surface proteins. This mechanism allows *B. burgdorferi* to evade
host immune attack, leading to pathogen infections in the host. As the key immune evasion tool,

B. burgdorferi expresses multiple FH-binding proteins on its surface during various stages of infection. So far, five different FH-binding surface proteins (CspA, CspZ, ErpA, ErpP and ErpC) are known in *Borrelia*. The structures of three of these surface proteins, CspA, CspZ and ErpA, in complex with different CCPs of human complement FH (hFH) were recently solved in our lab and are presented here.

My research work focused on characterization of each of these three host-pathogen protein complexes using a combination of biochemical and biophysical methods. Interestingly, our results for CspA and CspZ protein complexes with hFH protein revealed unique binding sites in borrelial as well as FH protein, contradicting previous reports. Many results have been published in the last decades regarding the binding of FH with CspA protein, assigning 1) CspA dimer as the major FH binding site and 2) hFH CCP 7 as the predominant binding site for CspA. In contrast, our results show that hFH CCP 5 alone is sufficient for binding CspA, with nanomolar dissociation constant (K_d) in a different region from the previously published dimeric cleft in CspA. Similarly, hFH CCP 7 and CCP 20 bind to borrelial protein CspZ and ErpA, respectively. I will present the structures of all these protein complexes along with results from characterization of each of these protein complexes as determined by alanine scanning mutagenesis, isothermal titration calorimetry (ITC) and extensive computational analysis.

To my dear grandma, **Ganga Silwal,** and grandpa, Late *Shankhar Silwal*,
with immense love and gratitude

ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. Honggao Yan for all his support throughout my graduate school journey. Thank you, Dr. Yan, for always guiding and encouraging me to think critically and more importantly, for training me to think 'outside of the box'. You were always patient with me when we had issues with experiments and you guided me to solve problems successfully. Without your continuous support and encouragement to participate in many activities during my graduate career, I wouldn't have achieved many valuable skills that are proving to be very helpful as I step out to expand my scientific career beyond graduate school. You gave me the flexibility to enjoy my graduate school beyond the lab. My graduate life wouldn't have been the same without your passion and professionalism for the research. So, thank you for all you did for me.

I am grateful to Profs. Gary Blanchard (Chemistry) and Erich Grotewold (Biochemistry) for extending their support when I needed the most during this journey. I am highly indebted to my committee members, Profs. Heedeok Hong, Kevin Walker and Jian Hu, for all the help and support during this journey. I am thankful to members of Yan lab, Dr. Yue Li, Dr. Aizhuo Liu and Yan Wu, for teaching and training me on all the basic molecular biology techniques. I will greatly miss your friendship. A great thanks to my friend Dan Lybrand (BMB) for stepping up when I needed the most.

I am beyond grateful to my family back in my country. I can't imagine my academic journey this far without your selfless love and support. Thank you, mom (Ranju Silwal) and dad (Rajaram Silwal), for always motivating me to think and aim higher. You sacrificed many things in

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

MAC	Membrane attack complexes
AMD	Age related macular degeneration
MBL	Mannose binding lectin
MASP	MBL-associated proteins
CP	Classical pathway
AP	Alternative pathway
LP	Lectin pathway
RCAs	Regulators of complement activation
FH	Factor H
CFH	Complement Factor H
FHL-1	Factor H-like protein-1
CCP	Complement control protein
BSK	Barbour-Stonner-Kelly
PBMC	Peripheral blood mononuclear cells
VlsE	Variable major protein-like sequence expressed
IgM	Immunoglobulin M
CRASP	Complement regulator-acquiring surface proteins
CFHR	Complement factor H-related proteins
hFH	Human factor H
mFH	Mouse factor H

MD	Molecular Dynamics
MM/GBSA	Molecular Mechanics/Generalized Born Surface Area
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
ITC	Isothermal Titration Calorimetry
IPTG	Isopropyl β-thiogalactoside
ELISA	Enzyme-linked immunosorbent assay
TEV	Tobacco Etch Virus
PMSF	Phenylmethylsulphonyl fluoride
MBP	Maltose binding protein

CHAPTER 1

Human Complement System

1.1. Introduction

The complement system is an integral part of the innate immune system and considered the first line of defense against microbial intruders [1-3]. It plays a very important role in tuning immune responses to discriminate among healthy host cells, apoptotic cells and foreign pathogens. The system consists of more than 30 circulating soluble and membrane-associated regulator proteins that interact with one another to induce cascades of responses to selectively kill foreign pathogens, while tightly regulating to avoid immune attack on 'self' cells [4]. The system is strictly controlled and regulated by an intricate network of effectors, receptors and regulators. The major functions attributed to the complement system are (i) phagocytosis of opsonized microbes [5, 6], (ii) amplification of inflammatory responses by recruiting macrophages and neutrophils, and (iii) lysis of foreign pathogens by forming membrane attack complexes (MAC) on the pathogen surface [7-9]. Recent research has also shown that the complement system plays an active role in processes such as maintaining immunologic memory to prevent re-invasion of pathogens [9, 10], tissue regeneration, and tumor growth [11] as well as in many human pathological conditions, such as age related macular degeneration (AMD), systemic lupus erythematosus, rheumatoid arthritis, atypical hemolytic syndrome [12, 13].

1.1.1. Complement Activation

The complement system is activated through three different pathways: lectin, classical and alternative pathway (Fig. 1.1). The three pathways are activated by different mechanisms and involve sequential cleavage and activation of protein complexes leading to the formation of complement activation and amplification protein complexes [14].
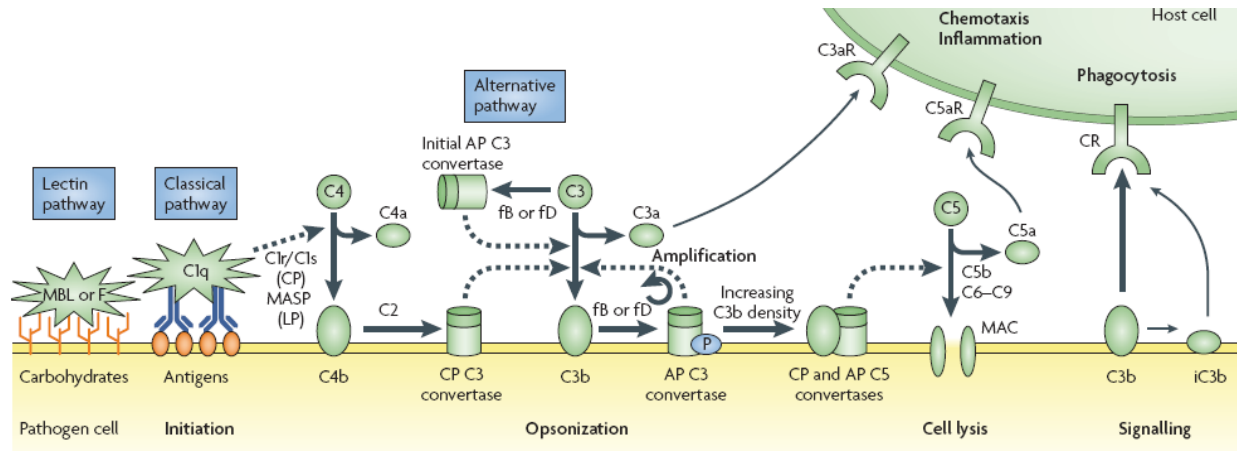
Figure 1.1. Schematic representation of three different pathways of complement activation. The figure is taken from reference [14].

The initial stages that trigger these three complement pathways differ significantly. The lectin pathway (LP) is activated when specific carbohydrate moieties on the pathogen surface are recognized by pattern recognition receptors such as mannose binding lectin (MBL) or Ficolin (F) [14]. MBL and Ficolin circulate in serum in complex with MBL-associated proteins (MASPs). Recognition and binding to specific carbohydrate receptors on the pathogen cell surface induces conformational changes resulting in the activation of the MASP [15, 16]. The activated MASP complex then cleaves the complement component 4 (C4) protein (~200 kDa) yielding two fragments, a small (~9 kDa) anaphylatoxin peptide C4a and the bigger (~190 kDa) C4b [14]. C4b then attaches to the pathogen cell surface and induces C2 protein to bind, making the CPC3 protein complex [14]. C2 is further cleaved by MASP proteins to form fragments C2a and C2b. C2b together with C4b has enzymatic activity, forming LP C3 convertase complex C4bC2a [17].

The classical pathway (CP) is initiated when antibodies recognize pathogens or other non-self-antigens in the host. The precursor of the classical pathway, the multimeric C1 protein complex consisting of C1q, C1r and C1s, specifically recognizes and binds to the Fc region of the

antibody complex [17]. C1 protein further cleaves C4 and C2 to form CPC3 convertase, C4bC2a [17].

Unlike LP and CP, the alternative pathway (AP) does not depend on a specific recognition-activation system [17]. Instead the precursor protein C3 undergoes a spontaneous conformational change and is constantly hydrolyzed at a low level to form two fragments, C3a and C3b [17]. C3a plays an important role as an anaphylatoxin, whereas C3b binds to the pathogen surface and recruits Factor B followed by Factor D. Factor D then cleaves Factor B to form the C3 convertase C3bBb [17]. The AP C3 convertase is also responsible for amplification of the activation of the complement system by all three pathways.

Complement activation from all three pathways results in opsonization of the pathogen by deposition of C3b on the pathogen cell surface, leading to phagocytosis. However, during sustained complement activation, the C3 convertases C4bC2a of LP and CP and C3bBb of the AP further cleave C3 into C3a and C3b. C3b functions as an opsin and plays an important role in phagocytosis as well as in amplification of the complement activation [17]. Additionally, C3b combines with C3 convertase to form a new multimeric complex, C5 convertase [17]. The C5 convertase cleaves C5 into two smaller components, C5a and C5b. C5a triggers inflammatory responses whereas C5b then binds to C6, C7, C8 and C9 to form the membrane attack complex (MAC). MAC inserts itself into the pathogen cell membranes, forming a pore in the membrane, resulting in cell death [17]. The complement system kills Gram-negative bacteria by both phagocytosis as well as by forming MAC complex [2]. Phagocytosis is the major way the complement system attacks Gram-positive bacteria because the thick peptidoglycan layer prevents complement attack by MAC [2, 4].

1.1.2. Complement Evasion

The complement system is an integral part of the immune system and hence is tightly controlled to avoid complement attack on self-cells. Various stages of complement activation are regulated by many membrane-bound and fluid phase regulators of complement activation (RCAs) [18]. RCAs then selectively bind to the human cell surface and downregulate complement activation on the self-cell, avoiding autoimmune attack.

Despite intricate control and tight regulation of complement activation, many pathogens have developed clever ways to circumvent this system. Many pathogens express proteins that affect the regulatory function of C3 convertase by either blocking the enzyme directly or by adjusting the enzyme activity, resulting in the attenuation of complement amplification loop [19-21]. Some microbes express and secrete endogenous proteases, such as metalloprotease aureolysin from *Staphylococcus aureus,* that can effectively cleave and inactivate many components of the complement system including C3 convertase, leading to complement evasion [19, 20].

Acquisition and binding of many soluble human RCAs, such as complement Factor H (FH) proteins, to the pathogen surface is another common immune evasion strategy employed by a wide range of pathogens [22]. The major RCA proteins from human serum that are frequently 'hijacked' by microbial pathogens include Factor H (FH) and Factor H-like protein-1 (FHL-1), regulator of the alternative pathway, and C4BP, as well as many other regulators of the alternative pathway [14, 19, 23, 24].

1.2. Complement Factor H Protein and its Roles in Complement Regulation

1.2.1. Structure and Function of FH

Out of three complement activation pathways, the alternative pathway (AP) of the complement system is the major pathway for complement activation [14]. However, unlike LP and CP, the AP doesn't have a specific recognition mechanism and is always active at a low level inside the host system but is rapidly amplified upon detection of pathogens [25]. Activation of complement via AP in self-cells can have catastrophic effects such as serious cases of autoimmunity. One major way normal healthy cells avoid autoimmune attack from the complement system is by producing many complement regulator proteins that selectively inhibit or downregulate complement activation on self-cells, preventing C3b deposition on host cell surfaces [22]. One of the key complement regulator proteins, Factor H (FH), plays both roles by effectively tagging and protecting self-cells from complement attack [26].

FH is an abundant serum glycoprotein that is constitutively expressed in human tissues (mainly in liver). It is a large (~155 kDa) protein consisting of 20 complement control protein (CCP) modules that give FH protein a 'beads-on-a-string' arrangement. Each CCP is ~60 residues long with a ~3-8-residue linker in between each pair of CCPs. Sequence alignment of all 20 CCPs shows four invariant Cys residues and a single almost invariant Trp residue (Fig. 1.2). These four conserved Cys residues form two disulfide bonds on each CCP, stabilizing the tertiary structure with Cys(I)-Cys(III) and Cys(II)-Cys(IV) disulfide bond arrangement (Fig. 1.3) [26, 27]. A shorter version of FH protein termed as Factor H-like protein 1 (FHL-1) is a splicing variant of FH that consists of only first 7 CCPs with four amino acid modifications at the C-terminus [28].
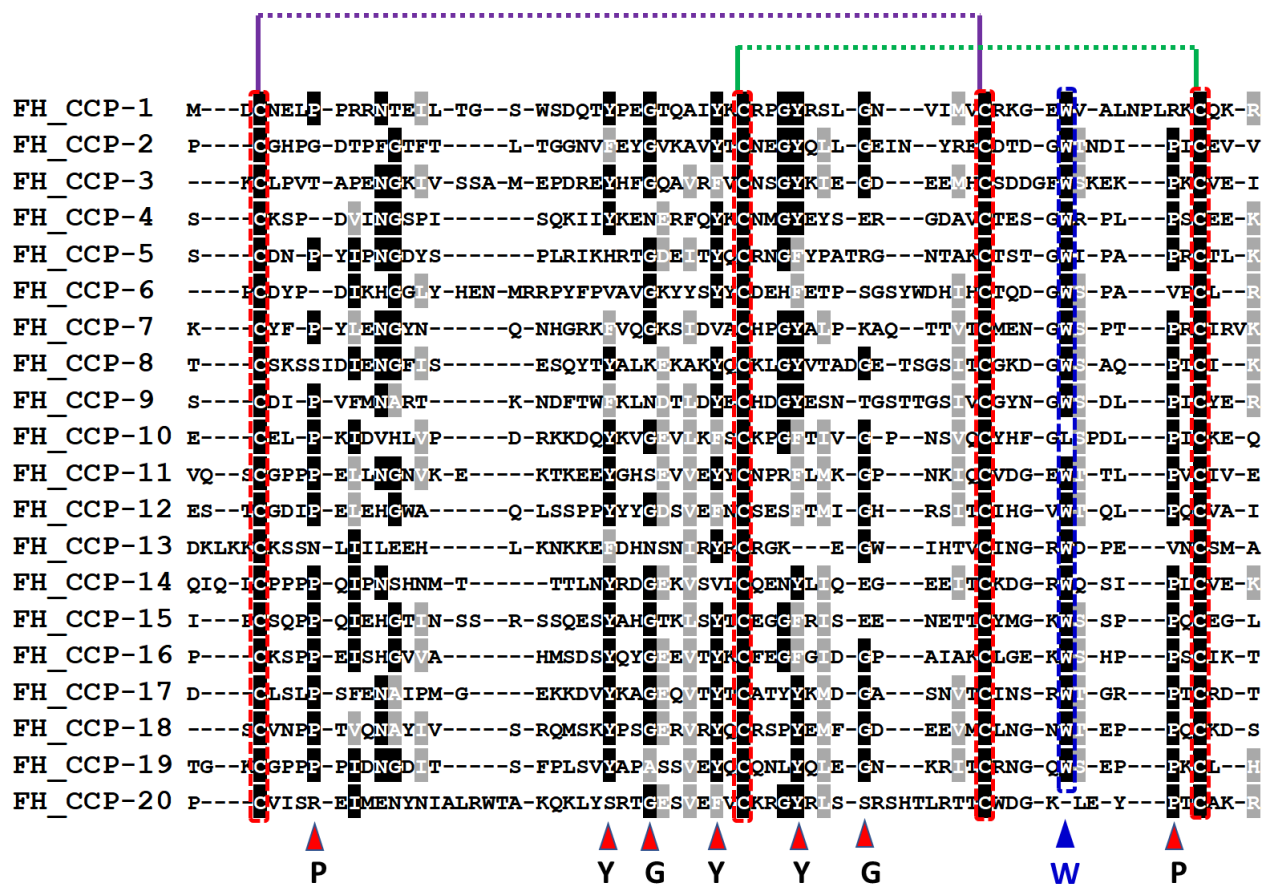
Figure 1.2. Multiple sequence alignment of 20 CCPs of complement FH protein. Highly conserved Cys and Trp residues are highlighted. Many well-conserved residues are indicated with arrows at the bottom. Horizontal lines indicate disulfide bond arrangement between Cys residues [26].

High resolution NMR and/or X-ray crystal structures are now available for all CCPs except for CCPs 14 and 17 [25, 29-35]. A typical FH CCP is predominantly rich in β-sheet and ovoid in shape with approximate dimensions of 40 Å x 15 Å x 10 Å [26]. Five extended stretches of β-strands run back and forth forming a head-to-tail arrangement with adjacent modules [26]. Despite a very high degree of structural similarity between CCPs, different regions of FH display distinct functions (Fig. 1.5). A previous study suggests that such variable functions between CCPs are due to both the diversity in the sequences of the individual CCPs and the different relative orientation of CCPs via intermolecular interactions between neighboring CCPs [36].
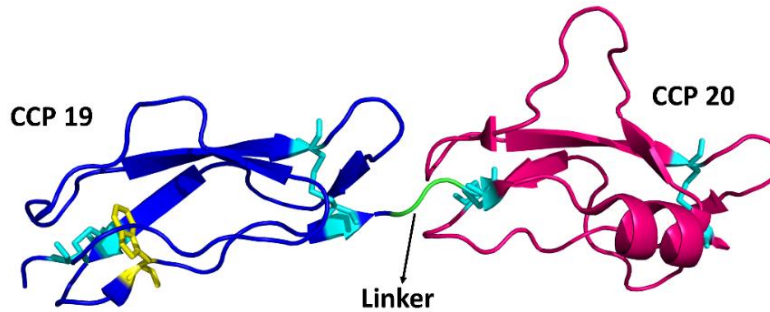
Figure 1.3. Cartoon representation of human complement Factor H (hFH) CCPs 19-20. Cys side chains making disulfide linkages are colored cyan. The almost invariant Trp residue is colored yellow in CCP 19. CCP 20 lacks this Trp residue and has a slightly more elongated structure than CCP 19. CCP 19 is a typical representation of all CCPs of hFH [26].

FH is a versatile and key regulator protein for the proper control of the alternative pathway of complement. It can bind to the host cells and hence protect them from complement mediated damage. Major mechanisms by which FH inhibits the complement attack on host cells are (i) inhibiting the formation of major alternative pathway precursor molecule C3 convertase, C3bBb, (ii) accelerating the decay of C3bBb, (iii) degradation and inactivation of C3b, and (iv) deactivation or disintegration of C5 convertase (Fig. 1.4) [27].
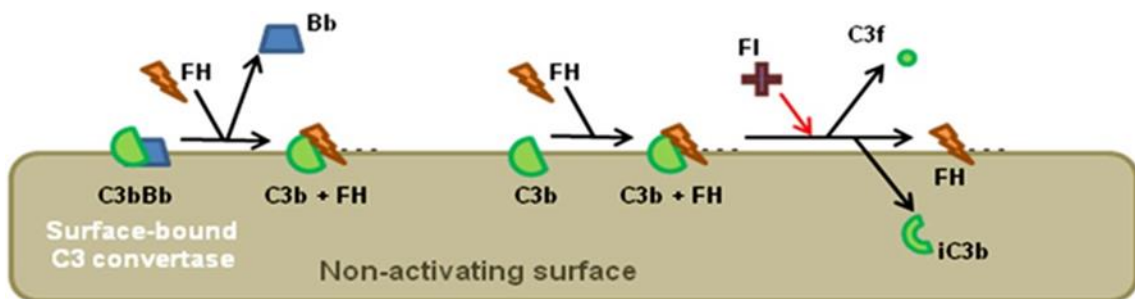
Figure 1.4. Schematic representation of various roles of FH protein in preventing complement attack on host cells.

Major functions of FH mentioned above largely depend on the deposition of C3b molecules on the pathogen cell surface. The serum concentration of the FH protein is around 500 mg/L or 3.2 μM, although it varies widely from 116−810 mg/L (0.8−5 μM) depending on genetic

8

and environmental factors. The plasma concentration of FH is similar to the C3 convertase concentration but is much higher than that of C3b concentration (0.1 µM) in human serum [27, 37]. FH CCP 1-4 and CCP 19-20 are two regions that bind to surface-bound C3b with a K_d value of ~ 12 µM and ~ 4 µM, respectively [27]. Three-dimensional structures are available for the complexes of CCP 1-4 with C3b and CCP 19-20 with C3d, one of the domains of C3b complex [29, 35, 38].

One of the key steps in downregulating or inhibiting complement activation on host cells and tissues involves FH selectively attaching to the host cell surface [27]. FH recognizes and preferentially binds to polyanionic markers such as sialic acid, heparin and glycosaminoglycans that are only present on host cell surfaces and not on pathogens [26, 27]. The mechanism of recognition and binding of such polyanionic markers on the host cells by FH serves as the distinguishing factor for the activation of complement on the pathogen cell surface without harming the host cells. The two regions of FH, CCP 6-8 and CCP 19-20, bind to polyanionic markers on the host cell surface (Fig. 1.4) [37]. Although the binding affinities of FH for polyanionic markers mentioned above are not reported yet, several binding studies of FH with commercial heparin have been published with reported K_d of 9.2 nM for FH [39], 9 µM for FH CCP 19-20 (with shorter heparin fragments) [40] and about 14 µM for FH CCP 6-8 (with longer heparin fragments) [41].
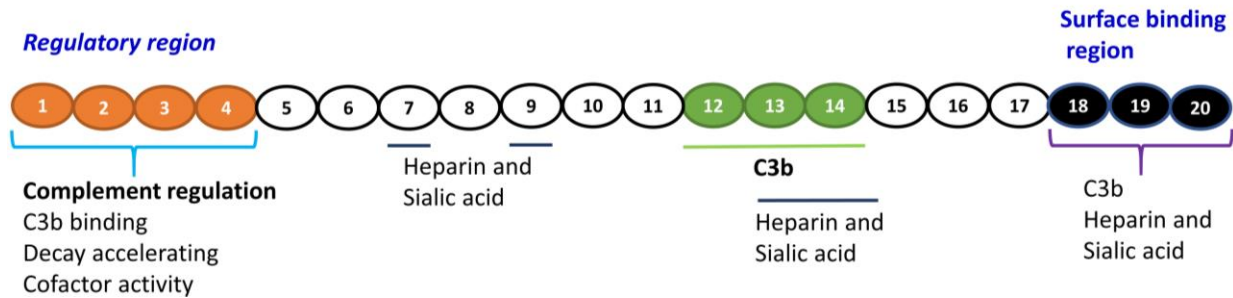
Figure 1.5. Schematic representation of different functional domains (CCPs) of complement Factor H (FH).

1.2.2. FH Recruitment as a Mechanism of Complement Evasion

As described earlier, the complement system is a powerful and tightly regulated system that facilitates destruction of foreign objects and pathogens from the host. Even though host cells are strictly protected by a combination of various membrane-associated and soluble regulatory proteins (e.g. FH protein) that block complement attack on self-cells, a wide range of pathogens have evolved various mechanisms to manipulate and modulate different stages of the complement activation process, successfully bypassing complement-mediated killing and hence effectively causing persistent infection of the host [1, 3, 14].

Interference at the early stages of complement activation is an efficient evasion strategy used by a wide range of pathogens that leads to blockage of downstream complement system activation. Capturing the host's Fc-tail of Ab or degrading Ab, suppressing LP and CP by capturing the components of C1, degrading and inactivating C3 and C5 convertase by proteolytic activity and inhibiting formation of MAC complex are some of the common strategies many pathogens use as immune evasion mechanisms. For example, many proteins from *Pseudomonas aeruginosa* such as elastase (PaE) and alkaline protease (PaAP) as well as proteins from *Staphylococcus*

Streptococcus spp. degrade and/or inhibit IgG and C1q, preventing complement activation by the CP [42, 43]. Likewise, several viruses, fungi and animal parasites have also evolved to use similar mechanisms of immune evasion [19, 44, 45]. Among many of these strategies used by microbes, acquisition and binding of soluble human regulators of complement activation (RCAs) to the surface of the microbial pathogen is the most common complement evasion strategy and is used by wide range of microbial pathogens (Fig. 1.6) [18, 22].
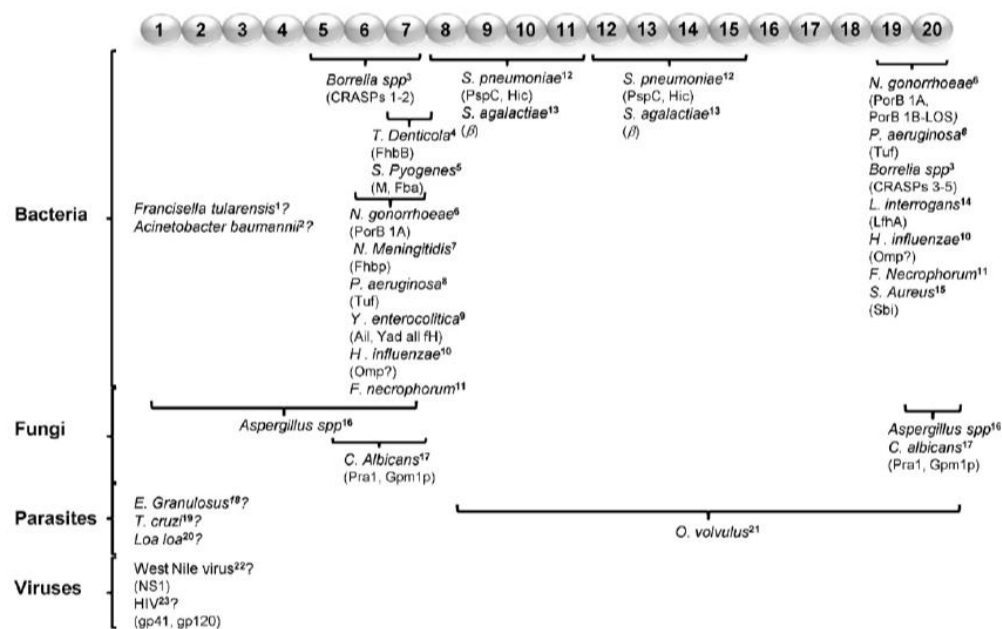
Figure 1.6. Interaction of complement Factor H (FH) protein with multiple pathogens. Pathogen-specific FH binding proteins are shown in parentheses. Pathogens for which FH interacting domains are unknown are indicated with a question mark. The figure is reproduced from reference [22].

Taking advantage of the crucial functions of FH, a wide variety of microbes express proteins that bind FH, 'hijacking' FH proteins from the host and recruiting them to their own surfaces to avoid complement attack, similar to how host cells utilize various functions of FH to avoid complement attack on self-cells.

Many prominent bacteria and viruses are known to use FH recruitment as a major immune evasion mechanism, leading to persistent infection of the host. The OspE protein of *Borrelia burgdorferi* [46], Sbi protein of *Staphylococcus aureus* [47], PspC of *Streptococcus Pneumoniae* [48], and fHbp of *Neisseria meningitidis* [49] are examples of surface proteins expressed by pathogens to recruit FH from the host. Similarly, the fungus *Candida albicans* and various *Echinococcus spp.* are some examples of diverse pathogens which use FH recruitment as the major immune evasion tool for pathogenesis [22]. In the past decade, significant progress has been made in gaining insights into molecular and structural details of protein complexes between FH and FH-binding proteins from various microbes. The crystal structure of the FH binding protein of *B. burgdorferi*, OspE, with human CCP19-20 [50] and the crystal structures of the complex of choline binding protein A (CbpA) of *Streptococcus pneumoniae* with human CCP9 [51] are a few examples of available structures of the protein complexes between human FH and microbial FH-binding proteins. Characterization of protein complexes between different CCPs of hFH and three of the FH-binding surface proteins of *B. burgdorferi,* CspA, CspZ and ErpA, is the focus of this work.

1.3. Host Specificity

Many pathogens, including *B. burgdorferi*, are able to infect multiple hosts, but some are highly specific and adapted to a single host species. In-depth understanding of the molecular basis of host specificity of pathogens can be crucial in deciphering pathogenic mechanisms and designing new therapeutics against pathogens [52]. Due to the complex molecular composition of bacteria, the molecular details of host specificity is poorly understood in bacteria relative to viruses [52]. However, in the last two decades, there has been significant progress in

characterization of the molecular determinants responsible for host specificity in many bacterial pathogens [52]. It is now clear that multiple molecular interactions that take place between the host and the pathogen during different stages of infection are responsible for bacterial host specificity [52].

The host specificity of any particular baterial strain is determined largely by the ability of the specific bacteria to (i) adhere to the host cells, (ii) replicate and colonize in host cells, and (iii) evade host immune attack. For example, the Opa protein of *Neisseria gonorrhoeae* is able to bind to human CEACAM1, an integral membrane glycoprotein, but not its canine, bovine, or murine orthologs, a factor making gonorrhoea a strictly human-specific disease [53]. Similarly, *Streptococcus pneumoniae* and *Haemophilus influenzae* induce an extracellular serine-type protease that specifically recognizes and cleaves IgA1 at the hinge region from human but not the IgA from mouse and other mammals due to significant sequence differences at the hinge region [52]. This effectively abolishes or significantly reduces all functions mediated by the Fc region in a human-specific manner [52, 54-56].

Among all the determinants of host specificity mentioned above, the ability of the pathogen to successfully evade the host immune system is one of the key determinants of host specificity. Unlike many human-specific pathogens, *Borrelia* is able to adhere to, colonize and infect a wide range of hosts including humans, other mammals, birds, and even reptiles. How *Borrelia* is able to survive and cause persistent infection on such a wide range of hosts is still unclear. Given the fact that there are no potent vaccines or specific therapeutics for Lyme disease, understanding the molecular basis of host specificity of *B. burgdorferi* is of eminent importance. In our lab, we hypothesized that the ability of *Borrelia* to infect such a wide range of

13

hosts is, at least in part, due to the multiple FH-binding surface proteins that *Borrelia* expresses to evade host immune attack. In total there are five FH-binding proteins known for *Borrelia.* Understanding the molecular basis of FH binding to three of these surface proteins is the focus of my dissertation work.

1.4. Lyme Borreliosis

1.4.1. Overview and Epidemiology

Lyme disease or Lyme borreliosis, caused by the tick-borne spirochete *Borrelia burgdorferi* (sensu lato), is the most common vector-borne disease in the United States and Europe [57]. If untreated, it can lead to a multisystemic chronic illness, particularly in skin, joints, nervous system, heart, or a combination thereof [58]. The reported cases of Lyme disease in the United States has drastically increased over the years, with more than 300,000 cases noted yearly [59]. The substantial spread of the disease in the U.S has primarily been in the Northeast from Maine to North Carolina; in the Midwest in Wisconsin, Minnesota and Michigan; and in the West, primarily in California [58]. In the U.S, the incidence of Lyme disease is highest among children between 5–15 years of age and adults 45-55 years of age, with a higher rate among men than among women in age groups less than 60 years of age [57, 60]. In some European countries, like Germany and Slovenia, the incidence of Lyme disease is slightly higher among women (55%) than among men (45%) [60-62]. In most parts of Europe and northeastern U.S., June and July are the peak months for the onset of disease, which corresponds to the feeding habits of *Ixodes* nymphal ticks [58].

1.4.2. Causative Pathogens

B. burgdorferi is the only species known to causes Lyme disease in North America. However, at least five different genospecies of *Borrelia (B. burgdorferi, Borrelia garinii, Borrelia afzelii, Borrelia spielmanii, and Borrelia bavariensis)* have been detected in Europe, leading to a wide range of possible clinical manifestations in Europe than in United States [63]. Infections from *B. garinii and B. afzellii strains* account for most Lyme disease cases in Europe, wheras *B. garinii* is the predominant species in Asia [63].
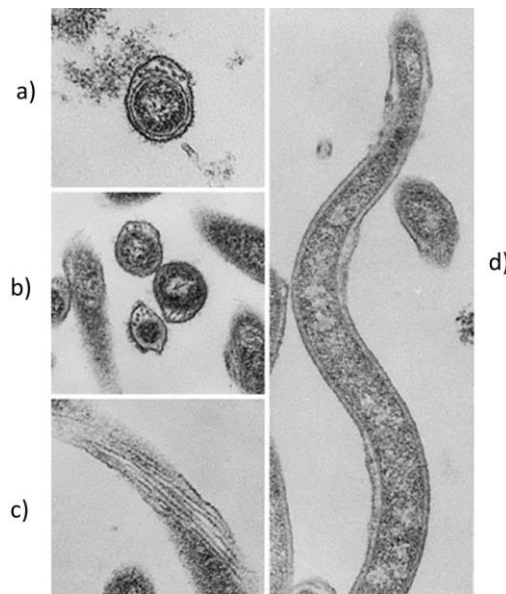
Figure 1.7. Electron micrographs of *Borrelia burgdorferi*. (a,b and c) The spirochetes have a transverse diameter of about 0.2mm with 7 to 11 flagella. (d) Longitudinal cross section of Borrelia burgdorferi. The spirochete has a length of about 11 to 39 mm with an outer membrane, flagellae, a cell wall, and cytoplasmic contents. The figure is reproduced from references [64, 65].

B burgdorferi was the first spirochete for which the complete genome was sequenced [66]. Due to complete absence of certain biosynthetic pathways in *Borrelia*, it solely depends on the host and surrounding environment for survival and nutritional requirements [63]. Nevertheless, borrelial species can be grown *in vitro* using highly nutrient-rich culture media such as Barbour-Stonner-Kelly (BSK) [67, 68].

1.4.3. Vector of Transmission and Animal Hosts

Two species of hard-bodied ticks, *Ixodes scapularis* and *Ixodes pacificus* are the main tick-vectors resposible for the transmission of *Borrelia* to a wide range of hosts in the U.S.; *Ixodes ricinus* in Europe and and *Ixodes persulcatus* in Asia are other common *Ixodes* tick species that transmit *Borrelia* in different hosts [69]. The life cycle of Lyme disease spirochetes is depicted in Fig. 1.7. The tick *Ixodes* goes through a three-stage life cycle — larva, nymph and adult – with one blood meal in each cycle [70]. The larva can feed on multiple hosts, including many mammals and birds [70]. An uninfected larva acquires *B. burgdorferi* by feeding on an infected host and the bacteria are retained within the tick midgut during subsequent blood meals and molts across all stages [71-73]. Nymphs feed on a similar range of hosts to larvae and transmit spirochetes to another competent reservoir host, perpetuating the enzootic cycle for the next generation of larval ticks [70, 73]. Adult ticks are generally not important in maintaining the *B. burgdorferi* life cycle, as they predominantly feed on larger incompetent hosts such as deer [73]. However, deer are an important part of the enzootic life cycle, as ticks mate on them. While all three stages of ticks can feed on human, nymphs are responsible for the vast majority of spirochete transmission to humans. Humans are generally considered the dead-end host, as it is unclear whether an uninfected feeding tick is able to acquire spirochetes from infected humans [70].
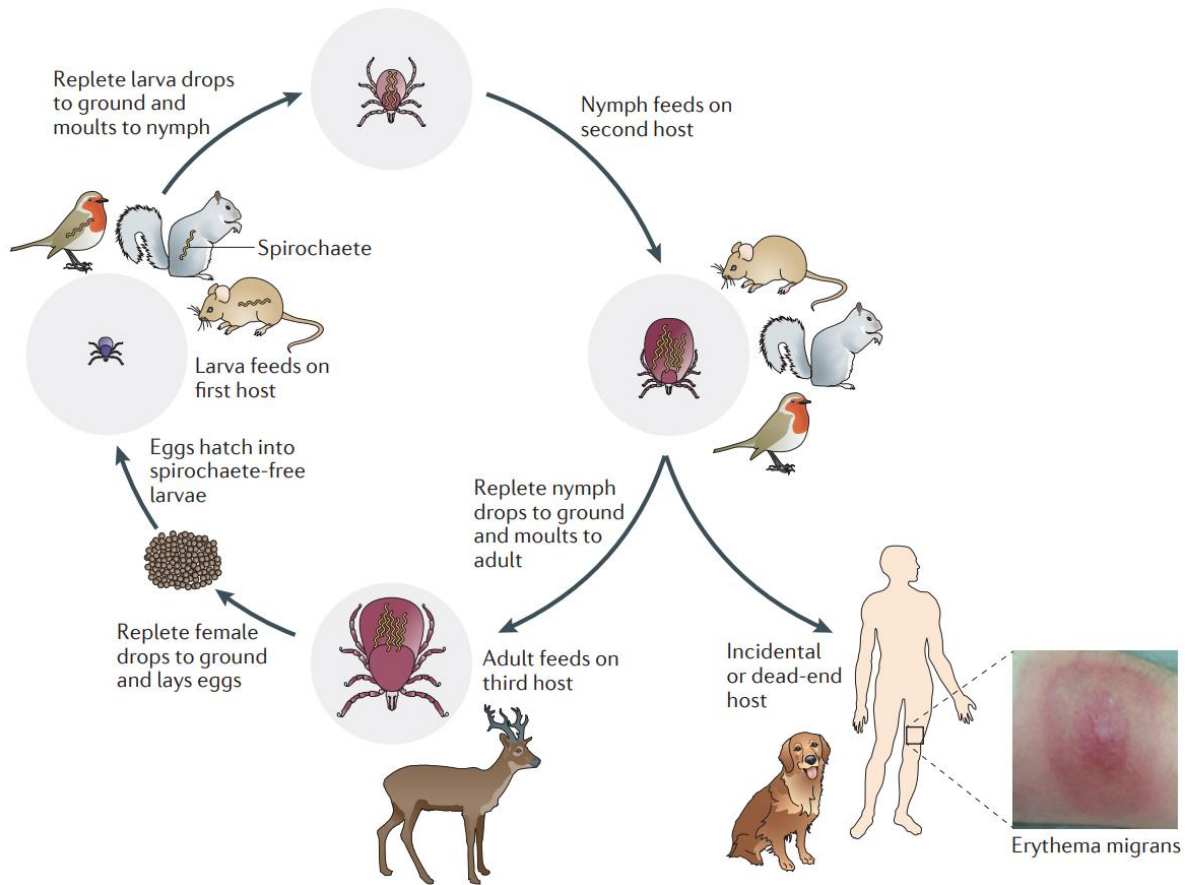
Figure 1.8. The enzootic cycle of Borrelia burgdorferi. The figure is reproduced from reference [70].

1.4.4. Pathogenesis

In order to maintain a complex enzootic cycle, *B. burgdorferi* must adapt to two drastically distinct environments: the tick and a mammalian or avian host. After the uninfected nymph acquires the spirochete, the spirochete survives in a dormant state in the tick midgut, expressing abundant OspA protein [74]. OspA protein plays major role in shielding the spirochete from host antibodies during initial uptake of a blood meal [75]. However, OspA protein is not important for persistence of spirochetes inside the tick vector [75]. After the blood meal, the OspA protein is significantly downregulated and another set of proteins, including OspC, are expressed and upregulated [76-78]. OspC protein binds to mammalian plasminogen as well as many tick salivary

17

gland proteins, helping the spirochete to invade the host tick's salivary glands. Also, expression of OspC at this stage plays an important role in successful evasion of initial immune attack from the mammalian host [74, 76].

During feeding, an infected tick injects *B. burgdorferi*, usually at the local site of the bite into the skin of the host, and starts to multiply. In general, the host's immune cells first encounter *B. burgdorferi* at this site and *in vitro* experiments show that dendritic cells derived from the dermis readily engulf *B. burgdorferi* [64, 79]. Soon after the infection, *B. burgdorferi* induces and stimulates many inflammatory and peripheral blood mononuclear cells (PBMCs), resulting in production of cytokines, particularly interferons (IFN)-γ [80-82]. Thus, infection of *B. burgdorferi* in humans and other animals elicits immune attack from both the innate and adaptive immune systems, resulting in both macrophage-mediated and antibody-mediated killing and clearance of the pathogen [63]. Within a few days to weeks, *B. burgdorferi* can disperse to many sites in the host, including the myocardium, retina, muscle, bone, spleen, liver, meninges, and even brain [64]. Spreading through the skin and other tissues inside the host is facilitated by binding of OspC protein to mammalian plasminogen and its activator proteins [83].

Despite a robust and highly regulated cellular and humoral immune response from the host, *B. burgdorferi* can survive during dissemination and cause persistent infection. Major virulence factors in *Borrerlia* that aid in persistence include (i) ability of the pathogen to change or downregulate expression of certain immunogenic surface exposed proteins, such as OspC, and (ii) ability to alter rapidly and extensively by recombining antigenic properties of the surface lipoprotein known as variable major protein-like sequence expressed (VlsE) [63]. The spirochete also differentially expresses various other lipoproteins and their paralogs, including OspE/F, that

can further contribute significantly in maintaining antigenic diversity [64, 84]. In addition, *B. burgdorferi* expresses up to five different complement regulator-acquiring surface proteins (CRASPs) that bind to the host's regulators of complement activation (RCAs), such as FH and FHL-1, protecting the spirochete from complement-mediated killing [63].

Since *B. burgdorferi* does not produce any toxins or any extracellular matrix-degrading proteases, major clinical manifestations of Lyme disease, including tissue damage, at each stages result primarily from the host inflammatory responses that can vary depending on the different genospecies of *Borrelia* that causes infection [63]. In human, specific immunoglobulin M (IgM) is generally associated with polyclonal activation of B cells, resulting in elevated levels of IgM and circulation of key immune regulatory complexes in serum [64, 85]. Complement fixation and opsonization by the host B cell responses seems to play a major role in killing the spirochete [86]. Studies in mice have shown that *B. burgdorferi*-specific CD4$^+$ Th1 cells, which mainly secretes IFN-γ, are the major primers to induce T-cell dependent B cell responses [87].

1.5. Complement Regulator Acquiring Surface Proteins (CRASPs) of *B. burgdorferi*

As discussed earlier, one of the key mechanisms that borrelial species use to avoid complement-mediated immune attack from the host is by acquiring host immune regulator proteins, such as FH and FHL-1, to their own surface. Recruitment and binding of FH protein in *B. burgdorferi* is primarily mediated by surface proteins on the bacteria, collectively termed complement regulator-acquiring surface proteins (CRASPs). So far, there are five such surface proteins from 3 distinct genetically unrelated groups known in *Borrelia*. CRASPs are significantly different in their primary and tertiary structures and interact with FH and FHL-1 with different

binding modes. Each of these proteins (CspA, CspZ, ErpA, ErpC and ErpP) are briefly discussed below.

1.5.1. Characteristics of *Borrelia* Surface Protein *CspA*

The CspA protein (also referred as BbCRASP-1, CRASP-1 or BBA68 in literature) is a 25.9 kDa surface-exposed lipoprotein in *B. burgdorferi* [88]. It can bind to human FH (hFH) and FHL-1 protein via CCP 5. The cspA gene is located on the linear lp54 replicon of *B. burgdorferi* B31 strain, which is a part of a large PFam54 gene family [88]. Sequence analysis reveals that *B. burgdorferi* B31 strain contains 11 PFam54 genes, located in 4 different linear plasmids [88-91].

The crystal structure of CspA revealed a homodimer, each monomer with 'helical-lollipop' like arrangements with 5 crossing α-helices (αA-αE)[92]. Previously, it was hypothesized that CspA contained a coiled-coil element that potentially served as the binding site for FH protein [92, 93]. However, the published structure of the CspA dimer disproved the hypothesis and suggested a different potential binding site for FH [94]. All the studies published so far, based on the crystal structure of the FH-free CspA, have suggested the dimeric cleft between two paired monomers as the binding site for FH [88, 95]. However, crystal structure and other characterization studies from our lab revealed a completely different FH binding site in CspA. These results are discussed in chapter 2.
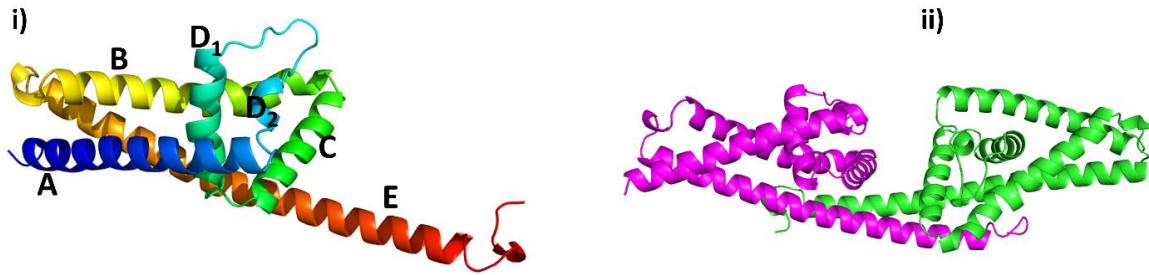
Figure 1.9. Overall structure of CspA. (i) The cartoon representation of CspA monomer secondary structure. Helices are labeled from N terminus to the C terminus with A-E. (ii) Cartoon representation of CspA homodimer showing the dimeric cleft in between the two paired CspA monomers.

Expression levels of *cspA* gene in spirochetes isolated from the midgut of unfed ticks and from ticks isolated two weeks after transmission to mammal is largely undetectable [96-99]. However, spirochetes show very high levels of cspA gene transcription and CspA protein production during tick-mammal feeding [97, 100, 101]. *Borrelia* produces CspA again during transmission from infected mammals to feeding uninfected ticks. Hence, transcription of cspA increased rapidly during tick-to-mammal and mammal-to-tick transmission processes but is significantly downregulated at all other stages of infection [97, 101]. Because the spirochete expresses cspA only for a short period of time during the mammalian infection, cspA is unable to generate a robust immunogenic and antibody response from the human immune system [96, 102]. Hence, whether CspA can be an effective target for vaccine design for humans and other animals against *Borrelia* is still unclear.

B. burgdorferi spirochetes expressing CspA protein alone on their surface can escape complement mediated attack from the host [103, 104], whereas spirochetes lacking CspA on their surface are highly susceptible to complement attack. Both of these observations suggest that by expressing CspA alone on the surface, *B. burgdorferi* is able to mediate resistance against human complement attack [104]. This makes CspA a very important FH-binding surface protein

for Lyme disease pathogenesis and possibly a significant target for developing therapeutics against *B. burgdorferi*.

1.5.2. Characteristics of *Borrelia* Surface Protein *CspZ*

CspZ (also referred as BbCRASP-2, CRASP-2, or BBH06 in literature) is a 23.2 kDa surface lipoprotein in *B. burgdorferi* [88]. It binds to both FH and FHL-1 proteins via CCP 7 [105, 106]. Located on the lp28-3 replicon of *B. burgdorferi* B31 strain [88, 105, 106], the cspZ gene is unique within the entire *B. burgdorferi* B31 strain genome [66, 88, 91, 106]. Sequence alignment shows a well-conserved cspZ sequences among different genospecies of *Borrelia* that cause Lyme disease [88, 107-109]. However, the CspZ protein sequences vary significantly due to species-specific polymorphism among different genospecies of *Borrelia* [88, 108, 109]. Interestingly, native CspZ protein is insensitive to proteinase *K* and trypsin-related proteolytic degradation [88, 106, 110].

The expression profile of CspZ protein during the *Borrelia* infection cycle is opposite of CspA. The cspZ gene is undetected in *Borrelia* in tick midguts or during transmission from feeding ticks to mammals [88, 101, 110]. The cspZ transcript level increases significantly after two weeks of the infection and stays at high levels throughout the dissemination and persistent infection in human [88, 110]. Further, unlike CspA, patients with infection of Lyme disease during early or late stage usually show very robust antibody response to CspZ, further suggesting the significant production of CspZ protein by *Borrelia* during natural infection [88, 95, 101, 108].

The crystal structure of CspZ protein consists of a single domain with a single hydrophobic core [111]. The tertiary structure also indicates two different lobes, the bigger N-terminal lobe and the smaller C-terminal lobe (Fig. 1.9) [111], with both lobes consisting almost entirely of α-

helices. There are 6 helices (αA - αF) in the N-terminal lobe and 3 helices in the C-terminal lobe

with helical arrangement similar to a right-turn four-helix bundle for the N-terminal helices [111].

However, unlike in a regular four-helix bundle arrangement where helices run continuously,

helices in the N-terminal lobe show little bent structure with loops between helices A and B as

well as helices C and D [111]. The C-terminal lobe consists of three α-helices (αG- αI) with αG and

αH forming a helix-turn-helix motif in an antiparallel fashion [111]. The αI helix is connected to

αH helix by a long loop running parallel to αC of the N-terminal lobe [111]. Structural alignment

analysis suggests that the three dimensional fold of CspZ protein is unique among all other FH-

binding proteins [111].
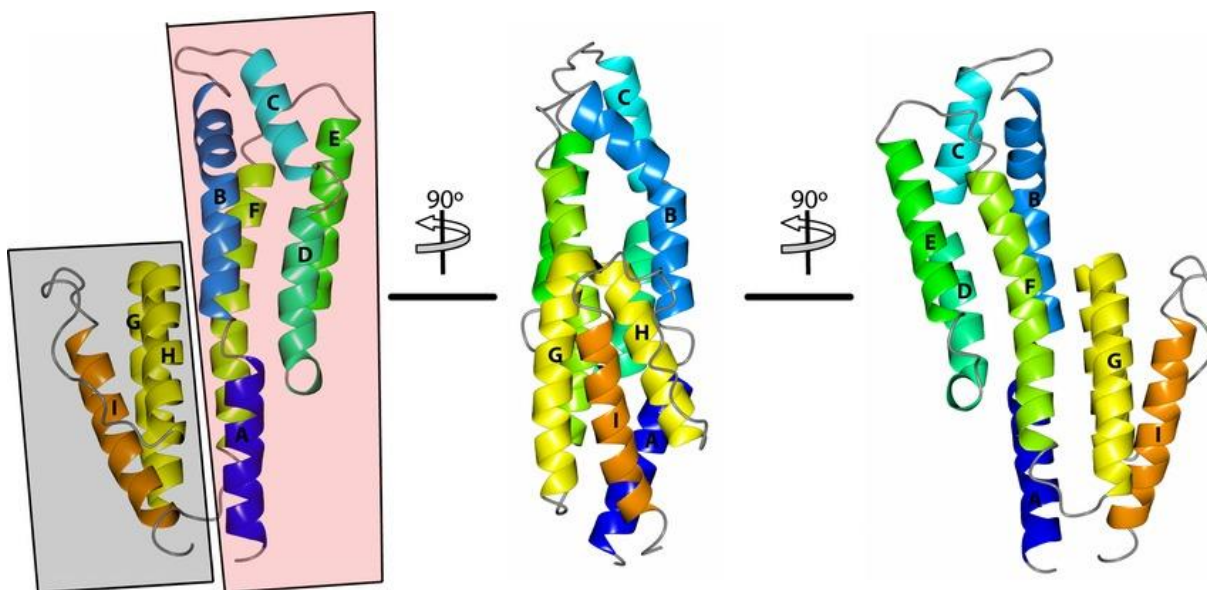
Figure 1.10. The structure of CspZ protein in its free form. The three different orientations of CspZ, rotated by 90° over the vertical plane. The N terminal lobes (pink shade) are labeled A to F and C-terminal lobe (grey shade) are labeled G to I. The figure is reproduced from reference [111].

Several studies identified the N-terminal lobe as the potential binding site for FH protein

and several specific amino acid residues were identified as key residues important for FH binding

[105, 109]. However, based on the crystal structure of the complex of CspZ with FH CCP 7 from

our lab, a majority of the specific amino acid residues identified by previous mutagenesis studies as important for binding FH appear incorrect. Instead, most of the residues identified thus far are involved in maintaining the structural integrity of the protein itself, rather than directly involved in binding to FH. These results are discussed in chapter 3.

1.5.3. Characteristics of *Borrelia* Erp Family Proteins

ErpP, ErpC, and ErpA (also termed BbCRASP-3, BbCRASP-4 and BbCRASP-5 respectively) are ~20 kDa outer membrane lipoproteins in *B. burgdorferi* B31 strain that bind to human FH protein, predominantly via CCP 20 [50, 88, 112]. However, unlike CspA and CspZ, Erp family proteins do not bind to FHL-1 from human [112-114]. The B31 strain of *Borrelia* encodes more than 13 unique Erp-related proteins and each of these proteins bind to FH with varying degrees of affinity under certain conditions [91, 115]. However, whether such binding with FH has any biological relevance is not understood.

Erp proteins are largely expressed by the borrelial spirochete during mammalian infection but are greatly repressed during tick colonization within the host [88]. Out of all the Erp proteins, ErpP, ErpA and ErpC are studied quite extensively in the past decade. Although all three proteins are thought to play an important role in the bacterial immune evasion process by binding to FH from the host, detailed molecular mechanisms of FH recruitment by Erp proteins and the biological significance of such interactions are yet to be investigated. Moreover, spirochetes lacking cspA and cspZ genes with only erpP, erpC and erpA genes did not bind to human FH but were able to bind to smaller complement factor H-related proteins (CFHRs) [113, 116], indicating different binding properties of these proteins *in vivo* and *in vitro*. Although it has been hypothesized that binding of FH to Erp proteins might be hindered by the presence of many other

24

large membrane associated proteins on the bacterial surface *in vivo*, there is no clear evidence

for this and it adds additional challenge to expand further studies on Erp proteins [88].

The crystal structure of one of the Erp family proteins, ErpP, in free form as well as in

complex with human FH CCP 20 was recently published [50]. Previous computational studies

suggested the presence of a coiled-coil element within Erp family proteins as an important

structural feature important for FH binding [93]. However, the published structure of ErpP

protein as well as crystal structure of ErpA protein solved in our lab did not show any such

structural features, contradicting this hypothesis.
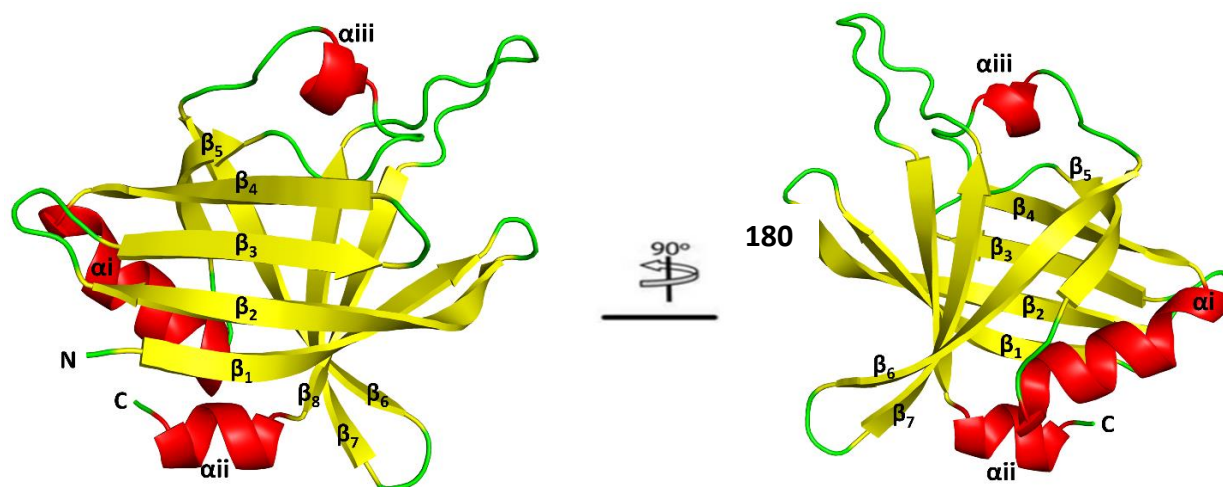
Figure 1.11. Cartoon representation of the overall structure of ErpA protein, rotated around 180° in the vertical plane.

Typical Erp protein contains a single large globular domain with eight anti-parallel β-

strands and two α-helices [50]. The backbone hydrogen bonding between residues in β_1 and β_8

results in an asymmetric β-barrel arrangement along the center of the protein [50]. β_2 and β_6 are

highly twisted and do not form a regular β-barrel [50]. Erp proteins are attached to the outer

membrane of *Borrelia* via a highly flexible N-terminal region [50]. This flexibility on the N-

terminus is thought to allow free movement of Erp proteins about the lipid anchor, although the anchoring mechanism is not fully understood [50].

Erp proteins are well-exposed to the bacterial surface and a wide range of Erp proteins are expressed on the bacterial surface during mammalian infection [88]. This induces a variety of counter-responses from the host immune system and produces strong antibody responses. However, due to significant differences in Erp protein sequences among different genospecies of *Borrelia*, the Erp protein family does not seem to be a suitable target for a therapeutic approach against Lyme disease [88].

1.6. Scope and Significance of this Study

Lyme disease is one of the most common vector-borne diseases in the United States and Europe and is caused by different genospecies of *Borrelia burgdorferi* sensu lato, including *B. burgdorferi*, *B. garinii* and *B. afzelii [70, 117].* Unlike many pathogens that have very strict hosts, one key and unique feature of *B. burgdorferi* genospecies is their ability to survive, colonize and cause persistent infection in a wide range of hosts, including humans and other mammals, birds and even reptiles [70, 117]. The ability of *Borrelia* spirochetes to survive and cause disease in many hosts is mainly associated with their ability to successfully circumvent the host's immune system via FH recruitment to their own surface using FH-binding surface proteins [72, 88]. Understanding the molecular details of FH recruitment mechanisms from *Borrelia* would provide valuable insights into the pathogenesis of Lyme disease. Further, in the long run, this information can be used to gain a detailed understanding of the molecular mechanism involved in host specificity of Lyme disease. This could open a new door for Lyme disease therapeutics, as there are currently no specific therapeutics or vaccines available against this disease.

Research in the last decade shows that it is not a simple problem to connect the role of FH binding ability of *Borrelia* to its pathogenesis and host specificity. A highly complex borrelial genome with variable redundant sequences that make up different genospecies of *Borrelia* and existence of multiple FH binding surface proteins, including CspA, CspZ and several Erp proteins, add additional challenges in understanding the significance of FH recruitment by *Borrelia* in pathogenesis and host specificity of Lyme disease. Also, data presented in various published studies have been collected using non-equilibrium methods such as ELISA [105, 106]. Further, the use of crude serum sample without proper quantification of the FH content make the validation of the results challenging [101, 108, 118]. In addition, computational and extensive mutagenesis studies have been carried out without any knowledge of binding sites on either protein [95]. Therefore, these studies have suggested contradicting and incorrect information on protein-protein interactions between human FH and borrelial surface proteins.

The crystal structure of the complex between i) CspA and FH CCP 5, ii) CspZ and FH CCP 7, iii) ErpA and FH CCP 20 were recently solved in our lab and each of the structures are presented here in subsequent chapters. Using rigorous biochemical, biophysical and computational approaches, we have characterized these protein-protein interactions and identified key residues important for FH binding on borrelial as well as FH proteins for each protein complex. Availability of these structures and identification of 'hot-spot' residues at the protein interfaces would constitute major progress in understanding Lyme disease pathogenesis. Further, insights presented here could be crucial in designing better animal models that can be key in designing novel therapeutics against Lyme disease.

REFERENCES

REFERENCES

1. Ricklin, D., et al., *Complement: a key system for immune surveillance and homeostasis.* Nat Immunol, 2010. **11**(9): p. 785-97.

2. Serruto, D., et al., *Molecular mechanisms of complement evasion: learning from staphylococci and meningococci.* Nat Rev Microbiol, 2010. **8**(6): p. 393-9.

3. Zipfel, P.F., et al., *The complement fitness factor H: role in human diseases and for immune escape of pathogens, like pneumococci.* Vaccine, 2008. **26 Suppl 8**: p. I67-74.

4. Walport, M.J., *Complement. First of two parts.* N Engl J Med, 2001. **344**(14): p. 1058-66.

5. Brown, E.J., *Complement receptors and phagocytosis.* Curr Opin Immunol, 1991. **3**(1): p. 76-82.

6. Brown, E.J., *Complement receptors, adhesion, and phagocytosis.* Infect Agents Dis, 1992. **1**(2): p. 63-70.

7. Bloch, E.F., et al., *C5b-7 and C5b-8 precursors of the membrane attack complex (C5b-9) are effective killers of E. coli J5 during serum incubation.* Immunol Invest, 1997. **26**(4): p. 409-19.

8. Tomlinson, S., et al., *Killing of gram-negative bacteria by complement. Fractionation of cell membranes after complement C5b-9 deposition on to the surface of Salmonella minnesota Re595.* Biochem J, 1989. **263**(2): p. 505-11.

9. Dunkelberger, J.R. and W.C. Song, *Complement and its role in innate and adaptive immune responses.* Cell Res, 2010. **20**(1): p. 34-50.

10. Molina, H., et al., *Markedly impaired humoral immune response in mice deficient in complement receptors 1 and 2.* Proc Natl Acad Sci U S A, 1996. **93**(8): p. 3357-61.

11. Qu, H., D. Ricklin, and J.D. Lambris, *Recent developments in low molecular weight complement inhibitors.* Mol Immunol, 2009. **47**(2-3): p. 185-95.

12. Wagner, E. and M.M. Frank, *Therapeutic potential of complement modulation.* Nat Rev Drug Discov, 2010. **9**(1): p. 43-56.

13. Robson, M.G. and M.J. Walport, *Pathogenesis of systemic lupus erythematosus (SLE).* Clin Exp Allergy, 2001. **31**(5): p. 678-85.

14. Lambris, J.D., D. Ricklin, and B.V. Geisbrecht, *Complement evasion by human pathogens.* Nat Rev Microbiol, 2008. **6**(2): p. 132-42.

15. Wallis, R., *Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway.* Immunobiology, 2007. **212**(4-5): p. 289-99.

16.   Sorensen, R., S. Thiel, and J.C. Jensenius, *Mannan-binding-lectin-associated serine proteases, characteristics and disease associations.* Springer Semin Immunopathol, 2005. **27**(3): p. 299-319.

17.   Sarma, J.V. and P.A. Ward, *The complement system.* Cell Tissue Res, 2011. **343**(1): p. 227-35.

18.   Zipfel, P.F., T. Hallstrom, and K. Riesbeck, *Human complement control and complement evasion by pathogenic microbes--tipping the balance.* Mol Immunol, 2013. **56**(3): p. 152-60.

19.   Zipfel, P.F., R. Wurzner, and C. Skerka, *Complement evasion of pathogens: common strategies are shared by diverse organisms.* Mol Immunol, 2007. **44**(16): p. 3850-7.

20.   Laarman, A., et al., *Complement inhibition by gram-positive pathogens: molecular mechanisms and therapeutic implications.* J Mol Med (Berl), 2010. **88**(2): p. 115-20.

21.   Rooijakkers, S.H., et al., *Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases.* Nat Immunol, 2005. **6**(9): p. 920-7.

22.   Ferreira, V.P., M.K. Pangburn, and C. Cortes, *Complement control protein factor H: the good, the bad, and the inadequate.* Mol Immunol, 2010. **47**(13): p. 2187-97.

23.   Blom, A.M., T. Hallstrom, and K. Riesbeck, *Complement evasion strategies of pathogens-acquisition of inhibitors and beyond.* Mol Immunol, 2009. **46**(14): p. 2808-17.

24.   Simon, N., et al., *Malaria parasites co-opt human factor H to prevent complement-mediated lysis in the mosquito midgut.* Cell Host Microbe, 2013. **13**(1): p. 29-41.

25.   Wu, J., et al., *Structure of complement fragment C3b-factor H and implications for host protection by complement regulators.* Nat Immunol, 2009. **10**(7): p. 728-33.

26.   Schmidt, C.Q., et al., *Translational mini-review series on complement factor H: structural and functional correlations for factor H.* Clin Exp Immunol, 2008. **151**(1): p. 14-24.

27.   Achila, D.O., *Molecular basis of pneumococcal adherence and complement evasion : structural and biochemical studies of pneumococcal virulence factor, CbpA*. 2013. 1 online resource (xiv, 162 pages).

28.   Clark, S.J. and P.N. Bishop, *Role of Factor H and Related Proteins in Regulating Complement Activation in the Macula, and Relevance to Age-Related Macular Degeneration.* J Clin Med, 2015. **4**(1): p. 18-31.

29.   Kajander, T., et al., *Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2897-902.

30.   Blaum, B.S., et al., *Structural basis for sialic acid-mediated self-recognition by complement factor H.* Nat Chem Biol, 2015. **11**(1): p. 77-82.

31. Makou, E., A.P. Herbert, and P.N. Barlow, *Functional anatomy of complement factor H.* Biochemistry, 2013. **52**(23): p. 3949-62.

32. Makou, E., et al., *Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H.* J Mol Biol, 2012. **424**(5): p. 295-312.

33. Schmidt, C.Q., et al., *The central portion of factor H (modules 10-15) is compact and contains a structurally deviant CCP module.* J Mol Biol, 2010. **395**(1): p. 105-22.

34. Morgan, H.P., et al., *Structural analysis of the C-terminal region (modules 18-20) of complement regulator factor H (FH).* PLoS One, 2012. **7**(2): p. e32187.

35. Morgan, H.P., et al., *Structural basis for engagement by complement factor H of C3b on a self surface.* Nat Struct Mol Biol, 2011. **18**(4): p. 463-70.

36. Makou, E., A.P. Herbert, and P.N. Barlow, *Creating functional sophistication from simple protein building blocks, exemplified by factor H and the regulators of complement activation.* Biochem Soc Trans, 2015. **43**(5): p. 812-8.

37. Perkins, S.J., et al., *Complement factor H-ligand interactions: self-association, multivalency and dissociation constants.* Immunobiology, 2012. **217**(2): p. 281-97.

38. Herbert, A.P., et al., *Structural and functional characterization of the product of disease-related factor H gene conversion.* Biochemistry, 2012. **51**(9): p. 1874-84.

39. Yu, J., et al., *Biochemical analysis of a common human polymorphism associated with age-related macular degeneration.* Biochemistry, 2007. **46**(28): p. 8451-61.

40. Herbert, A.P., et al., *Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure.* J Biol Chem, 2006. **281**(24): p. 16512-20.

41. Fernando, A.N., et al., *Associative and structural properties of the region of complement factor H encompassing the Tyr402His disease-related polymorphism and its interactions with heparin.* J Mol Biol, 2007. **368**(2): p. 564-81.

42. Hong, Y.Q. and B. Ghebrehiwet, *Effect of Pseudomonas aeruginosa elastase and alkaline protease on serum complement and isolated components C1q and C3.* Clin Immunol Immunopathol, 1992. **62**(2): p. 133-8.

43. Schenkein, H.A., et al., *Increased opsonization of a prtH-defective mutant of Porphyromonas gingivalis W83 is caused by reduced degradation of complement-derived opsonins.* J Immunol, 1995. **154**(10): p. 5331-7.

44. Favoreel, H.W., et al., *Virus complement evasion strategies.* J Gen Virol, 2003. **84**(Pt 1): p. 1-15.

45. Cooper, N.R., *Complement evasion strategies of microorganisms.* Immunol Today, 1991. **12**(9): p. 327-31.

46. Hellwage, J., et al., *The complement regulator factor H binds to the surface protein OspE of Borrelia burgdorferi.* J Biol Chem, 2001. **276**(11): p. 8427-35.

47. Haupt, K., et al., *The Staphylococcus aureus protein Sbi acts as a complement inhibitor and forms a tripartite complex with host complement Factor H and C3b.* PLoS Pathog, 2008. **4**(12): p. e1000250.

48. Janulczyk, R., et al., *Hic, a novel surface protein of Streptococcus pneumoniae that interferes with complement function.* J Biol Chem, 2000. **275**(47): p. 37257-63.

49. Pizza, M., J. Donnelly, and R. Rappuoli, *Factor H-binding protein, a unique meningococcal vaccine antigen.* Vaccine, 2008. **26 Suppl 8**: p. I46-8.

50. Bhattacharjee, A., et al., *Structural basis for complement evasion by Lyme disease pathogen Borrelia burgdorferi.* J Biol Chem, 2013. **288**(26): p. 18685-95.

51. Achila, D., et al., *Structural determinants of host specificity of complement Factor H recruitment by Streptococcus pneumoniae.* Biochem J, 2015. **465**(2): p. 325-35.

52. Pan, X., Y. Yang, and J.R. Zhang, *Molecular basis of host specificity in human pathogenic bacteria.* Emerg Microbes Infect, 2014. **3**(3): p. e23.

53. Voges, M., et al., *CEACAM1 recognition by bacterial pathogens is species-specific.* BMC Microbiol, 2010. **10**: p. 117.

54. Reinholdt, J. and M. Kilian, *Lack of cleavage of immunoglobulin A (IgA) from rhesus monkeys by bacterial IgA1 proteases.* Infect Immun, 1991. **59**(6): p. 2219-21.

55. Kilian, M., J. Mestecky, and R.E. Schrohenloher, *Pathogenic species of the genus Haemophilus and Streptococcus pneumoniae produce immunoglobulin A1 protease.* Infect Immun, 1979. **26**(1): p. 143-9.

56. Male, C.J., *Immunoglobulin A1 protease production by Haemophilus influenzae and Streptococcus pneumoniae.* Infect Immun, 1979. **26**(1): p. 254-61.

57. Mead, P.S., *Epidemiology of Lyme disease.* Infect Dis Clin North Am, 2015. **29**(2): p. 187-210.

58. Steere, A.C., et al., *Lyme borreliosis.* Nat Rev Dis Primers, 2016. **2**: p. 16090.

59. Hinckley, A.F., et al., *Lyme disease testing by large commercial laboratories in the United States.* Clin Infect Dis, 2014. **59**(5): p. 676-81.

60. Bacon, R.M., et al., *Surveillance for Lyme disease--United States, 1992-2006.* MMWR Surveill Summ, 2008. **57**(10): p. 1-9.

61. Fulop, B. and G. Poggensee, *Epidemiological situation of Lyme borreliosis in germany: surveillance data from six Eastern German States, 2002 to 2006.* Parasitol Res, 2008. **103 Suppl 1**: p. S117-20.

62.   Berglund, J., et al., *An epidemiologic study of Lyme disease in southern Sweden.* N Engl J Med, 1995. **333**(20): p. 1319-27.

63.   Stanek, G., et al., *Lyme borreliosis.* Lancet, 2012. **379**(9814): p. 461-73.

64.   Steere, A.C., *Lyme Disease (Lyme Borreliosis) Due to Borrelia burgdorferi*, in *Principles and Practice of Infectious Diseases*, J.E. Bennett, R. Dolin, and M.J. Blaser, Editors. 2017, Elsevier: Philadelphia, PA. p. 2725-2735.

65.   Steere, A.C., et al., *The spirochetal etiology of Lyme disease.* N Engl J Med, 1983. **308**(13): p. 733-40.

66.   Fraser, C.M., et al., *Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.* Nature, 1997. **390**(6660): p. 580-6.

67.   Preac-Mursic, V., B. Wilske, and G. Schierz, *European Borrelia burgdorferi isolated from humans and ticks culture conditions and antibiotic susceptibility.* Zentralbl Bakteriol Mikrobiol Hyg A, 1986. **263**(1-2): p. 112-8.

68.   Barbour, A.G., *Isolation and cultivation of Lyme disease spirochetes.* Yale J Biol Med, 1984. **57**(4): p. 521-5.

69.   Gray, J.S., *Review The ecology of ticks transmitting Lyme borreliosis.* Experimental & Applied Acarology, 1998. **22**(5): p. 249-258.

70.   Radolf, J.D., et al., *Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes.* Nat Rev Microbiol, 2012. **10**(2): p. 87-99.

71.   Kurtenbach, K., et al., *Fundamental processes in the evolutionary ecology of Lyme borreliosis.* Nat Rev Microbiol, 2006. **4**(9): p. 660-9.

72.   Kurtenbach, K., et al., *Host association of Borrelia burgdorferi sensu lato--the key role of host complement.* Trends Microbiol, 2002. **10**(2): p. 74-9.

73.   Sparagano, O., *Samuels DS, Radolf JD: Borrelia: Molecular Biology, Host Interaction and Pathogenesis*. Vol. 3. 2010.

74.   Steere, A.C., J. Coburn, and L. Glickstein, *The emergence of Lyme disease.* The Journal of clinical investigation, 2004. **113**(8): p. 1093-1101.

75.   Battisti, J.M., et al., *Outer surface protein A protects Lyme disease spirochetes from acquired host immunity in the tick vector.* Infect Immun, 2008. **76**(11): p. 5228-37.

76.   Rosa, P., *Lyme disease agent borrows a practical coat.* Nat Med, 2005. **11**(8): p. 831-2.

77.   Schwan, T.G. and J. Piesman, *Temporal changes in outer surface proteins A and C of the lyme disease-associated spirochete, Borrelia burgdorferi, during the chain of infection in ticks and mice.* J Clin Microbiol, 2000. **38**(1): p. 382-8.

78. Montgomery, R.R., et al., *Direct demonstration of antigenic substitution of Borrelia burgdorferi ex vivo: exploration of the paradox of the early immune response to outer surface proteins A and C in Lyme disease.* J Exp Med, 1996. **183**(1): p. 261-9.

79. Filgueira, L., et al., *Human dendritic cells phagocytose and process Borrelia burgdorferi.* J Immunol, 1996. **157**(7): p. 2998-3005.

80. Mullegger, R.R., et al., *Differential expression of cytokine mRNA in skin specimens from patients with erythema migrans or acrodermatitis chronica atrophicans.* J Invest Dermatol, 2000. **115**(6): p. 1115-23.

81. Salazar, J.C., et al., *Coevolution of markers of innate and adaptive immunity in skin and peripheral blood of patients with erythema migrans.* J Immunol, 2003. **171**(5): p. 2660-70.

82. Glickstein, L., et al., *Inflammatory cytokine production predominates in early Lyme disease in patients with erythema migrans.* Infect Immun, 2003. **71**(10): p. 6051-3.

83. Lagal, V., et al., *Borrelia burgdorferi sensu stricto invasiveness is correlated with OspC-plasminogen affinity.* Microbes Infect, 2006. **8**(3): p. 645-52.

84. Hefty, P.S., et al., *Changes in temporal and spatial patterns of outer surface lipoprotein expression generate population heterogeneity and antigenic diversity in the Lyme disease spirochete, Borrelia burgdorferi.* Infect Immun, 2002. **70**(7): p. 3468-78.

85. Steere, A.C., et al., *Lyme arthritis: correlation of serum and cryoglobulin IgM with activity, and serum IgG with remission.* Arthritis Rheum, 1979. **22**(5): p. 471-83.

86. Montgomery, R.R., et al., *Human phagocytic cells in the early innate immune response to Borrelia burgdorferi.* J Infect Dis, 2002. **185**(12): p. 1773-9.

87. Keane-Myers, A. and S.P. Nickell, *T cell subset-dependent modulation of immunity to Borrelia burgdorferi in mice.* J Immunol, 1995. **154**(4): p. 1770-6.

88. Kraiczy, P. and B. Stevenson, *Complement regulator-acquiring surface proteins of Borrelia burgdorferi: Structure, function and regulation of gene expression.* Ticks Tick Borne Dis, 2013. **4**(1-2): p. 26-34.

89. Wywial, E., et al., *Fast, adaptive evolution at a bacterial host-resistance locus: the PFam54 gene array in Borrelia burgdorferi.* Gene, 2009. **445**(1-2): p. 26-37.

90. Casjens, S.R., et al., *Genome stability of Lyme disease spirochetes: comparative genomics of Borrelia burgdorferi plasmids.* PLoS One, 2012. **7**(3): p. e33280.

91. Casjens, S., et al., *A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi.* Mol Microbiol, 2000. **35**(3): p. 490-516.

92. Cordes, F.S., et al., *A novel fold for the factor H-binding protein BbCRASP-1 of Borrelia burgdorferi.* Nat Struct Mol Biol, 2005. **12**(3): p. 276-7.

93. McDowell, J.V., et al., *Demonstration of the involvement of outer surface protein E coiled coil structural domains and higher order structural elements in the binding of infection-induced antibody and the complement-regulatory protein, factor H.* J Immunol, 2004. **173**(12): p. 7471-80.

94. Cordes, F.S., et al., *Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of Borrelia burgdorferi.* Int J Med Microbiol, 2006. **296 Suppl 40**: p. 177-84.

95. Kraiczy, P., et al., *Mutational analyses of the BbCRASP-1 protein of Borrelia burgdorferi identify residues relevant for the architecture and binding of host complement regulators FHL-1 and factor H.* Int J Med Microbiol, 2009. **299**(4): p. 255-68.

96. Rossmann, E., et al., *Borrelia burgdorferi complement regulator-acquiring surface protein 1 of the Lyme disease spirochetes is expressed in humans and induces antibody responses restricted to nondenatured structural determinants.* Infect Immun, 2006. **74**(12): p. 7024-8.

97. Bykowski, T., et al., *Borrelia burgdorferi complement regulator-acquiring surface proteins (BbCRASPs): Expression patterns during the mammal-tick infection cycle.* Int J Med Microbiol, 2008. **298 Suppl 1**: p. 249-56.

98. Lederer, S., et al., *Quantitative analysis of Borrelia burgdorferi gene expression in naturally (tick) infected mouse strains.* Med Microbiol Immunol, 2005. **194**(1-2): p. 81-90.

99. Wallich, R., et al., *Artificial-infection protocols allow immunodetection of novel Borrelia burgdorferi antigens suitable as vaccine candidates against Lyme disease.* Eur J Immunol, 2003. **33**(3): p. 708-19.

100. von Lackum, K., et al., *Borrelia burgdorferi regulates expression of complement regulator-acquiring surface protein 1 during the mammal-tick infection cycle.* Infect Immun, 2005. **73**(11): p. 7398-405.

101. Bykowski, T., et al., *Coordinated expression of Borrelia burgdorferi complement regulator-acquiring surface proteins during the Lyme disease spirochete's mammal-tick infection cycle.* Infect Immun, 2007. **75**(9): p. 4227-36.

102. McDowell, J.V., et al., *Evidence that the BBA68 protein (BbCRASP-1) of the Lyme disease spirochetes does not contribute to factor H-mediated immune evasion in humans and other animals.* Infect Immun, 2006. **74**(5): p. 3030-4.

103. Brooks, C.S., et al., *Complement regulator-acquiring surface protein 1 imparts resistance to human serum in Borrelia burgdorferi.* J Immunol, 2005. **175**(5): p. 3299-308.

104. Kenedy, M.R., et al., *CspA-mediated binding of human factor H inhibits complement deposition and confers serum resistance in Borrelia burgdorferi.* Infect Immun, 2009. **77**(7): p. 2773-82.

105. Siegel, C., et al., *Deciphering the ligand-binding sites in the Borrelia burgdorferi complement regulator-acquiring surface protein 2 required for interactions with the human immune regulators factor H and factor H-like protein 1.* J Biol Chem, 2008. **283**(50): p. 34855-63.

106. Hartmann, K., et al., *Functional characterization of BbCRASP-2, a distinct outer membrane protein of Borrelia burgdorferi that binds host complement regulators factor H and FHL-1.* Mol Microbiol, 2006. **61**(5): p. 1220-36.

107. Kraiczy, P., et al., *Borrelia burgdorferi complement regulator-acquiring surface protein 2 (CspZ) as a serological marker of human Lyme disease.* Clin Vaccine Immunol, 2008. **15**(3): p. 484-91.

108. Rogers, E.A., et al., *Comparative analysis of the properties and ligand binding characteristics of CspZ, a factor H binding protein, derived from Borrelia burgdorferi isolates of human origin.* Infect Immun, 2009. **77**(10): p. 4396-405.

109. Rogers, E.A. and R.T. Marconi, *Delineation of species-specific binding properties of the CspZ protein (BBH06) of Lyme disease spirochetes: evidence for new contributions to the pathogenesis of Borrelia spp.* Infect Immun, 2007. **75**(11): p. 5272-81.

110. Coleman, A.S., et al., *Borrelia burgdorferi complement regulator-acquiring surface protein 2 does not contribute to complement resistance or host infectivity.* PLoS One, 2008. **3**(8): p. 3010e.

111. Brangulis, K., et al., *Structural characterization of CspZ, a complement regulator factor H and FHL-1 binding protein from Borrelia burgdorferi.* FEBS J, 2014. **281**(11): p. 2613-22.

112. Kraiczy, P., et al., *Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi.* Infect Immun, 2001. **69**(12): p. 7800-9.

113. Hammerschmidt, C., et al., *Contribution of the infection-associated complement regulator-acquiring surface protein 4 (ErpC) to complement resistance of Borrelia burgdorferi.* Clin Dev Immunol, 2012. **2012**: p. 349657.

114. Haupt, K., et al., *Binding of human factor H-related protein 1 to serum-resistant Borrelia burgdorferi is mediated by borrelial complement regulator-acquiring surface proteins.* J Infect Dis, 2007. **196**(1): p. 124-33.

115. Stevenson, B., et al., *Borrelia burgdorferi erp proteins are immunogenic in mammals infected by tick bite, and their synthesis is inducible in cultured bacteria.* Infect Immun, 1998. **66**(6): p. 2648-54.

116. Siegel, C., et al., *Complement factor H-related proteins CFHR2 and CFHR5 represent novel ligands for the infection-associated CRASP proteins of Borrelia burgdorferi.* PLoS One, 2010. **5**(10): p. e13519.117. Mannelli, A., et al., *Ecology of Borrelia burgdorferi sensu lato in Europe: transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change.* FEMS Microbiol Rev, 2012. **36**(4): p. 837-61.

118.   Kraiczy, P., et al., *Binding of human complement regulators FHL-1 and factor H to CRASP-1 orthologs of Borrelia burgdorferi.* Wien Klin Wochenschr, 2006. **118**(21-22): p. 669-76.

CHAPTER 2

Molecular Dynamics Analysis of *B. burgdorferi* Outer Surface Protein CspA

2.0. Abstract

Many pathogenic bacteria have evolved robust immune evasion strategies to evade host immune attack and hence can cause persistent infection. One common mechanism used by a wide range of pathogens to avoid host immune attack is to 'hijack' immune regulators from the host cells to the pathogen surface, allowing the pathogen to remain undetected by the host's immune system for kill and clearance. *B. burgdorferi*, the causative agent of Lyme disease, expresses up to five different surface proteins that recruit a key immune regulator protein in human, complement factor H (FH), to its own surface. This gives *Borrelia* the ability to successfully disguise itself and circumvent immune attack from many hosts, including human, other mammals, reptiles and even birds. Gaining insights into the molecular mechanism of FH recruitment by *Borrelia* using each of these surface proteins is crucial for understanding the pathogenesis of Lyme disease. In this chapter, I present computational studies of one of the key borrelial surface proteins, CspA, with detailed analysis of its dimeric form. Molecular analysis of the complex between CspA dimer and FH protein will be discussed in Chapter 3.

2.1. Introduction

CspA is one of the key surface-exposed lipoproteins in *Borrelia* that binds and recruits human FH (hFH) protein to the borrelial surface. The crystal structure of CspA revealed a homodimer, each monomer with 'helical-lollipop' like arrangements with 5 crossing α-helices (αA-αE) [1]. Previously, it was hypothesized that CspA contained coiled-coil elements that served as the potential binding site for FH protein [2]. However, the published structure of the CspA dimer disproved this hypothesis and suggested a different potential binding site for FH [3].

When the crystal structure of dimeric form of CspA protein was first published, several studies suggested the cleft region between the two CspA monomers as a potential binding site for FH (Fig. 2.1) [1, 4]. However, based on the crystal structure of the complex between CspA and hFH CCP5 solved in our lab (unpublished), the CspA dimer cleft region does not appear to be the FH binding site. Nevertheless, understanding the dynamics of the cleft region could provide some insights into other potential roles and biological relevance of CspA protein.
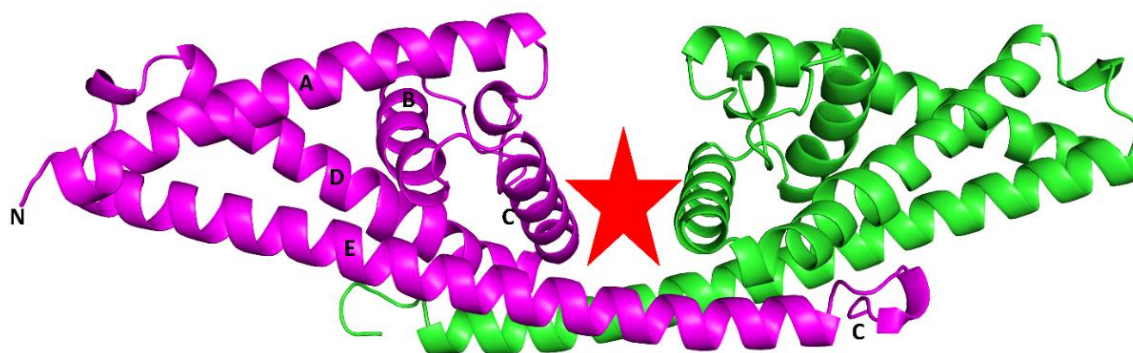
Figure 2.1. Ribbon representation of the CspA dimer. Two monomers of CspA are colored purple and green, and helices are labelled A-E. The red star at the dimeric cleft represents the suggested FH binding site based on published results [1, 3-5].

The CspA dimer is predominantly stabilized by the long C-terminal helical tail (helix E) that protrudes outwards from each CspA monomer [3]. The residues in the C-terminal end of each monomer are extensively involved in interactions between monomer, with approximately 2350 $Å^2$ of total protein surface area buried at dimer interface [3, 6]. This buried surface area at the dimer interface is larger than average buried area found in common biologically relevant protein complexes, indicating potential unique biological relevance of CspA dimer in *Borrelia* [3, 7]. Although the association of two CspA monomers at the dimeric interface is not strong, with reported K_d of 33 ± 5 μM [3], given the limitation of diffusion for membrane associated proteins and very strong binding of CspA with FH, the dimeric form of CspA indicates a biologically relevant protein form in *Borrelia* [3].

Previous studies showed that the deletion of the last ten residues from the C-terminal end of CspA completely abolished its FH binding ability, leading to the hypothesis that residues at the C-terminal end were directly involved in FH and FHL binding [1, 3, 4]. However, other studies have revealed that CspA C-terminal deletion mutants are structurally unstable and easily aggregate in nonspecific fashion [3], suggesting the importance of these residues to maintain structural integrity of CspA rather than direct involvement in FH binding. Based on the crystal structure of the CspA:hFH CCP5 complex solved in our lab, the previous hypothesis and the proposition of direct involvement of C-terminal residues in FH binding are incorrect. These results are discussed in chapter 3 in detail.

Several extensive mutagenesis studies performed targeting residues at the dimeric cleft region of CspA showed reduced binding to FH upon mutation of certain residues [4, 8]. However, based on our study, those residues are most likely important in maintaining structural integrity

of the CspA protein itself and none of the residues identified and reported so far as key residues for FH binding are involved in direct binding to FH. The binding affinity between two CspA dimers is relatively weak with a K_d of ~ 33 ± 5 µM [1]. The two CspA dimer crystal structures published (PDB IDs: 1W33 and 4BL4) show greater flexibility at the long C-terminal dimer interface, with approximately 16.8° intermonomer angle increase over both copies of CspA dimer in 4BL4 when compared to the 1W33 structure published earlier [1, 6]. Due to this flexibility at the dimer interface, the size of the dimeric cleft changes by 6 Å when the conformation is switched between 1W33 and 4BL4 (Fig. 2.2) [6].
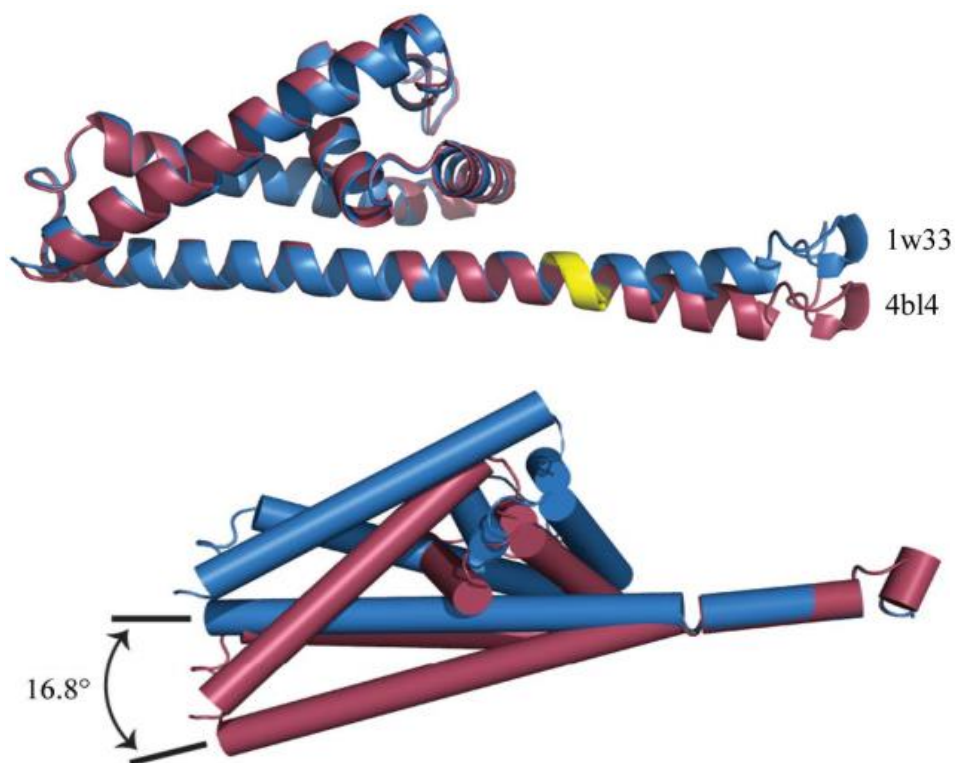
Figure 2.2. Different conformations of CspA dimer. (Top) Two reported CspA dimeric structures 1W33 (blue) and 4BL4 (red) superimposed along chain E. The residues colored in yellow function as the pivot point for conformation change between 1W33 and 4BL4. Only one monomer is shown for easiness. (Bottom) Superposition of two conformations over the last 10 residues of Chain E. This gives an intermonomeric angle of 16.8°, suggesting great deal of flexibility around dimeric interface. The figures are reproduced from the reference [6].

Although the initial hypothesis suggesting that change in the size of the CspA dimeric cleft functions as the clamping mechanism for FH binding is wrong, it would be interesting to explore more on the effects of the conformational change on the dynamics of the CspA cleft region. While the differences between these two CspA dimeric conformations could arise from variation in crystal packing, further study is required to get some prospective on two different reported conformations of the CspA dimer and to understand any biological relevance of the dimeric cleft. We employed the MD simulation approach to gain insights on two CspA dimer conformations.

2.2. Methods

2.2.1. MD Simulation and MM/GBSA Analysis of CspA Dimer

To investigate the dynamics of the CspA dimer, MD simulations for wild type and mutant CspA protein complexes were carried out. All MD simulations were performed using AMBER 16 Molecular Dynamics simulations package and the ff14SB force field was used to describe the protein [9, 10]. The crystal structure of the CspA dimer (PDB ID: 1W33) was used as the starting structure to generate coordinate files for subsequent MD simulations and analysis [3]. All mutant protein for *in silico* mutagenesis study were generated with the assumption that the structural integrity of the proteins is not affected by the single amino acid substitutions. Before running the simulation, all protein complexes were processed through the H^{++} program which automatically computes pK_a values of ionizable groups in protein and adds missing hydrogen atoms according to specified pH of the environment [11]. Special attention was given to the protonation state of histidine suggested by this program and adjustment of the ionization state of histidine was made accordingly.

The MD simulations were carried out with the AMBER 16 platform using the ff14SB force field [9, 10]. Original water molecules from the crystal structure were removed and hydrogen atoms were added using the TLEAP of the AMBER package [9]. The standard protonation states were further checked using H^{++} online software and verified manually for consistency and accuracy [11]. Four Na$^+$ counterions were added to balance the resulting charge of -4 of the dimer and the structure was solvated in a rectangular TIP3P water box [9], with the boundaries of at least 12 Å away from the protein atoms. Parameter files for MD, including coordinate and topology files, were generated for further processing. The solvated structure of the complex was

44

energy minimized using steepest descent and conjugated gradient methods [9], each for 5000 steps, first with positional restrain force constant of 100 kcal/ (mol. Å^2) on all heavy atoms of the protein and then without any positional restraints on the whole system. The minimized system was heated linearly to 300K for 120 ps under constant volume periodic boundary conditions (NVT) with no positional restraints. The system was further simulated for 750 ns under constant pressure and temperature conditions (NPT). Temperature was controlled using Langevin dynamics with collision frequency of 1 ps^{-1} during heating, equilibration and production steps [9, 12, 13]. All covalent bonds to hydrogen atoms were constrained by the SHAKE algorithm, with numerical integration time step of 2 fs [12, 13]. Long-range electrostatic attractions were computed using the Particle Mesh Ewald (PME) method with a cutoff of distance of 10 Å [14].

The simulation results were analyzed using the CPPTRAJ program of the AMBER16 package and the PyMOL program [15-17]. Root mean square deviation (RMSD) was monitored throughout the simulation to ensure equilibration of the system. Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and energy decomposition analyses were performed using the MMPBSA.py python script incorporated in AmbertTools16 [9, 18]. A generalized Born implicit solvent model with 0.15 M salt concentration was used for all MM/GBSA analyses [18]. The results from MM/GBSA run were calculated every 25 ns snapshot for a total of 30 snapshots ranging from 0 to 750 ns of total simulation time. The average from all the snapshots was calculated to estimate the interaction energies of the residues, particularly focusing on residues at CspA dimer interface. Since we were only interested in relative mutational effect on the interactions at the protein interface, the entropic contribution to the free energy change was not included in the calculation. Further, entropic calculations using currently

available platforms are computationally expensive and tend to have large margins of error, introducing significant uncertainty in the results.

After each successful MD simulation run, criteria such as density, temperature, pressure, root mean square deviation (RMSD), root mean square fluctuation (RMSF), KE, PE and total energy were analyzed to confirm system convergence to the equilibrium. The CPPTRAJ module from AMBER16 was used to calculate backbone RMSD values to monitor the stability of the protein complex during the entire simulation [15]. Root mean square fluctuation (RMSF) was calculated to measure fluctuation of individual residues during simulation. Similarly, the CPPTRAJ module was used to analyze various hydrogen bond properties, such as hydrogen bond distance and occupancies of each hydrogen bond interaction between residues at the dimer interface. A distance cutoff of 3.3 Å (between acceptor and donor heavy atoms) and angle cutoff of 120° (between donor heavy atom-hydrogen-acceptor heavy atom) was used to track all hydrogen bonds. Special attention was given to the stable hydrogen bonds, with occupancy of at least 50%.

2.2.2. Computational Alanine Scanning Mutagenesis

Alanine mutagenesis was also performed using the MD simulation data for the CspA dimer. *In silico* alanine scanning mutagenesis is a quick way to calculate the binding energy contribution of residues in protein-protein interaction interfaces. Because of the size, main chain conformation and electrostatic properties of alanine, it is usually the best choice for mutagenesis studies. MM/GBSA is a widely used and one of the most accurate and reliable *in-silico* approaches for predicting binding free energies of the protein complexes.

2.3. Results

2.3.1. RMSD and RMSF of the CspA Dimer

Backbone RMSD were calculated using the original crystal structure of dimeric CspA (1W33) as the reference coordinates. As indicated by the plot of backbone RMSD versus time, all systems are stable after the equilibrium point, with mean RMSD of 2.98 ± 0.54 Å over 750 ns of simulation time (Fig. 2.3). Considering the high degree of flexibility in CspA dimer, this RMSD value is within the acceptable range. Interestingly, right around 390 ns, there is a sudden decrease in total binding energy at the dimer interface, indicating that many interactions at the dimer interface are weakening at this point. To gain better insights into the types of interactions that are breaking around this time, contribution from all energy components, including Van der Waals, electrostatic, non-polar solvation and polar solvation energy, were analyzed before and after 390 ns (Table 2.1). The highest decrease in total binding energy is contributed by the weakening of non-polar interactions at the interface, as indicated by ~ 8-fold decrease in Van der Waals energy contribution after 390 ns. This indicates that the weaker non-polar interactions, including many hydrophobic pockets that are stabilizing the dimeric interface, are weakening at this point. Similarly, the hydrogen bond analysis shows that the two major hydrogen bonding interactions, N155 - N340 and R92 - K353, between two monomers also weakens significantly around this time. This is further supported by ~ 3.5-fold decrease in electrostatic energy at the dimer interface.
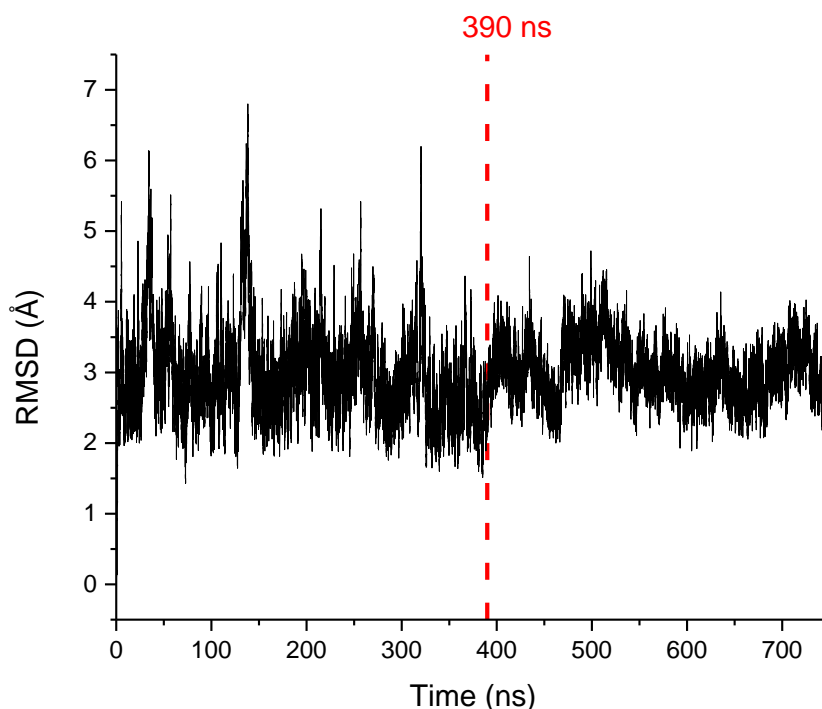
Figure 2.3. Backbone RMSD for the 750 ns simulation. As evident from the RMSD plot, the dimer remains stable throughout the simulation. The average RMSD for the entire simulation was 2.98 ± 0.54 Å.

Table 2.1. Binding free energies of free form CspA dimer averaged over 0 to 390 ns and 391 to 750 ns of simulation time

Simulation time	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{total}
0 to 390 ns	-245.3 ± 43.0	-435.2 ± 69.1	512.6 ± 78.2	-34.2 ± 5.9	-202.1 ± 38.9
391 ns – 750 ns	-31.4 ± 1.4	-123.0 ± 27.2	148.0 ± 28.5	-5.0 ± 0.3	-11.3 ± 1.4

ΔE_{vdw}= van der Waals contribution, ΔE_{ele} = electrostatic energy, ΔG_{GB} = solvation free energy , ΔG_{SA} = non polar contribution to the solvation energy

The flexibility of all CspA residues was monitored by the calculation of RMSF (Fig. 2.4). The red vertical line in the graph separates residues between two CspA monomers. There are five major peaks in the RMSF plot indicating high fluctuation within these regions. The corresponding regions of high fluctuation are colored red and are numbered from (i) to (vi) on the CspA monomer. All major fluctuations correspond to residues in either the loop regions or the residues at or near terminal ends. Residues corresponding to α-helices have relatively low fluctuations.

Particularly, the small fluctuation of the last 20 residues within helix E further indicates the tight interactions of these residues to stabilize the dimeric interface.
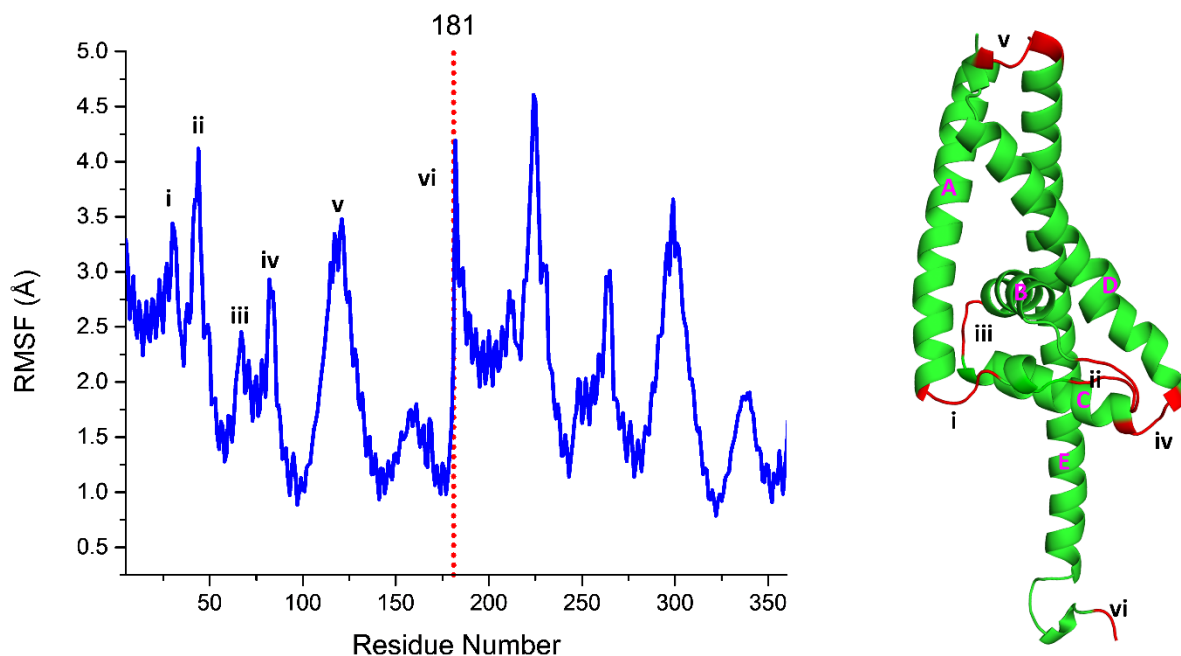
Figure 2.4. Backbone RMSF of CspA dimer residues (left). Some peak fluctuations correspond to residues at the loop regions in each monomer. Corresponding regions with maximum fluctuations, numbered (i) to (vi) in RMSF plot, are colored in red in the cartoon structure of the cspA monomer (right). As expected, the major fluctuations come from residues at the flexible loop regions.

2.3.2. Analysis of the Dynamics of CspA Dimeric Cleft

Since the two CspA dimer showed great deal of flexibility along helix E, I was interested to learn how this conformation changes over the time of simulation and how this conformational change affects the size of the dimeric cleft. Based on our analysis, the PDB snapshot from the last frame of 700 ns simulation showed similar conformation change with ~ 16.8° change in angle along helix E (Fig. 2.5). So, it is highly likely that the two conformations of CspA-dimer, as shown by the previous studies, are not just due to crystal packing. Two CspA monomers move along the vertical plane with about 17° angle with respect to helix E, changing the overall conformation and the size of the cleft region.

49
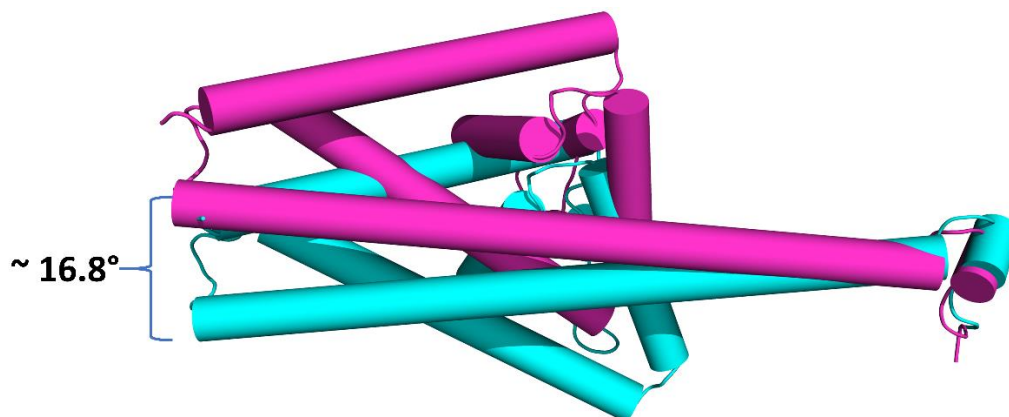
Figure 2.5. Structural alignment of initial CspA dimer (PDB ID: 4BL4) with PDB snapshot from the last frame of 700 ns simulation along helix E of CspA dimer. Starting with crystal structure of CspA dimer (PDB:4BL4, cyan), the conformation changes to the one shown in pink color in about 400ns simulation time.

In order to understand the dynamics of the conformational change at the CspA dimer interface, two simple calculations were made, focusing on residues from helices C and D of each monomer. First, the time evolution of the angle between helix C and helix D was monitored for the entire simulation. The fluctuation in angle between these two helices directly impacts not only the size of the base of the cleft but also affects the movement of helix E that forms the main interacting interface in each monomer. Second, the relative change in distance between residues in helix C of monomer-1 and monomer-2 was also monitored (Fig. 2.6). Time evolution of the distance between α-carbons of residues K127-K320, K141-K305 and E91-E269 from two CspA monomers was tracked throughout the simulation. K127, K141, K320 and K305 are from helix C and E91-E269 are from helix A of two CspA monomers. Tracking the relative change of distance between α-carbons of K127-K320 and K141-K305 can give some insights into the regions within CspA dimer that cause the reported change in dimeric cleft size. The two published crystal structures reported that the cleft size changes by 6 Å between two conformations. However, it is

50

not known what regions within the dimeric cleft cause change in the cleft size as there are many helices and loops in close vicinity of the cleft. Since helix C from two monomers forms the base and helix A forms the top of the cleft, tracking the relative distance change between the selected residue-pairs from two CspA monomers can help to identify the regions within the dimeric cleft that contribute the most to fluctuation of the cleft size.
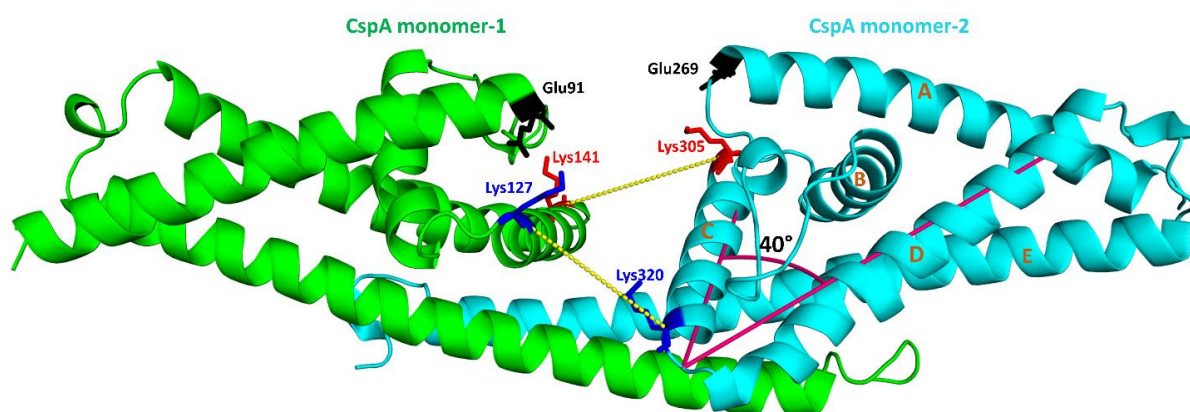
Figure 2.6. Ribbon representation of CspA dimer. Monomer-1 and monomer-2 are colored green and cyan respectively. Residue pair selected to track the α-carbon distances are represented with the same color, connected with yellow dash line. The angle between helix C and helix D is shown in pink with starting angle of 40°.

Since previous studies indicated that the change in conformation of the CspA dimer causes the change in size of the dimeric cleft, we were interested to find any specific movement at the cleft region due to helix C from two monomers that can potentially affect the cleft size. Since helix C from both monomers forms the base of the cleft, the size of the cleft is directly affected by the movement of this helix.

51
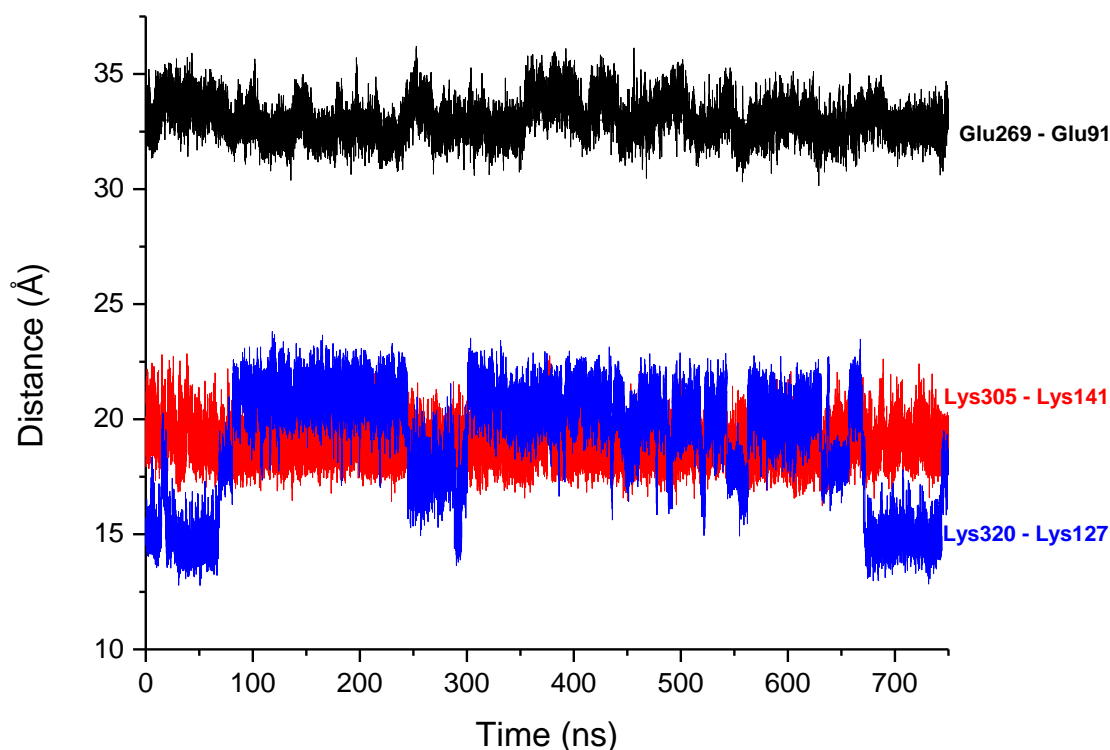
Figure 2.7. Time evolution of distance between α-carbons of residues from helix C and helix A of the CspA dimer. K141-K305 and K127-K320 are residue pairs from helix C of two monomers. E91-E269 is residue pair from Helix A that forms the top of the cleft. As evident by the change of the distance as the simulation proceeds, the change in dimeric cleft size is predominantly due to the movement of helix C along the base of the cleft. As suggested before, our simulation suggests little to no involvement of helix A in changing the cleft size.

Out of two residue pairs selected from helix C of each monomer, the distance between

K141- K305 remained largely constant for the duration of the simulation (Fig. 2.7). Also, since the

top portion of helix A forms the entry point into the cleft region, the distance between E91

(monomer-1) and E269 (monomer-2) was calculated. This distance also remains fairly constant

indicating little to no contribution of helix A in changing the cleft size. However, the distance

between α-carbons of K127 (monomer-1) and K320 (monomer-2) showed drastic fluctuations

over the time, with an increase of ~6 Å distance starting at about 80 ns. This 6 Å increase in

distance between helix C of two monomers suggests that the change in the size of the cleft

between two conformations of CspA dimer is most likely predominantly controlled by helix C without much involvement of other helices at the cleft region. While this result correlates with the 6 Å difference in cleft size that was observed between two conformations of CspA dimer (Fig. 2.4), where this shift in cleft size originates was not clear. Our results indicate that the C-terminal region of the helix C causes this fluctuation in cleft size at the dimeric interface. Interestingly, the fluctuation of distance between α-carbons of K127-K320 converges back to initial distance of ~ 15 Å around 690 ns, suggesting potential correlation between conformational change and the change in the size of the cleft.

Further, since movement along helix C was quite substantial, we expected a similar level of fluctuation in helix D relative to helix C. Based on the crystal structure of the CspA dimer, helix C is connected to helix D by a short loop of 3 amino acids with an angle of ~40°. We investigated how this angle changes with respect to the fluctuation in helix C.
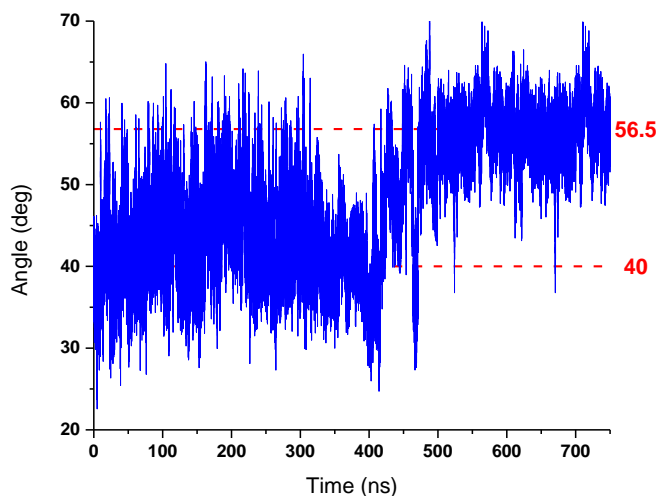
Figure 2.8. The time evolution of the angle between helix C and D. The angle at the start of the simulation is ~ 40° and increases to ~ 56.5°.

As shown in Figure 2.8, the angle between helices C and D changes by ~ 16.5° over the simulation time. Interestingly, this difference in angle between helices C and D correlates with the intermonomer angle difference between two reported structures of CspA dimer (Fig. 2.2). It is also interesting to note that helix D is connected to helix E, which forms the main interacting region for CspA dimer. Our results show that the difference between the intermonomer angle between the reported structures of CspA dimer is due to the movement of helix C which in turn affects the position of helix E which forms the main dimeric interface. This gives rise to a different conformation of CspA dimer where helix E of one dimer makes an angle of ~ 16.8° relative to the other structure. In contrast to previous studies, our computational analysis shows that the change in the dimeric cleft is predominantly driven by movement of helix C and not by helix A.

Hence, using MD simulation data, we shed some light on the dynamics of the CspA dimeric structure and the flexibility in conformations. However, any biological significance of such flexibility around the dimeric cleft has yet to be uncovered.

2.3.3. Identification of Key Residues Important for CspA Dimer Formation

The interaction between two CspA monomers is greatly stabilized by the C terminal end of helix E. The two CspA monomers associate weakly with a K_d of ~ 33 ± 5 µM [3]. To identify key residues and interactions involved in CspA dimer formation, we further analyzed MD simulation data and performed decomposition energy calculation using the MM/GBSA method [19].

Based on the decomposition of the binding energy (Fig. 2.9), the dimer is stabilized by many hydrophobic interactions and some hydrogen bonds. Many hydrophobic residues, mainly repeats of L, I, V and F, are deeply buried at the dimeric interface, forming major hydrophobic binding pockets at the center of the dimer interface. Previous studies showed that L177D

mutation completely abolished the dimer formation [4] and this is further supported by the large

decomposition energy for L177 from our computational analysis. Based on our results, mutating

L177 to D introduces a polar group in a highly hydrophobic pocket surrounded by residues V161,

L163, F167 and F174. This disrupts the hydrophobic pocket at the center of the dimer interface

and hence L177D greatly destabilizes the dimer interface.
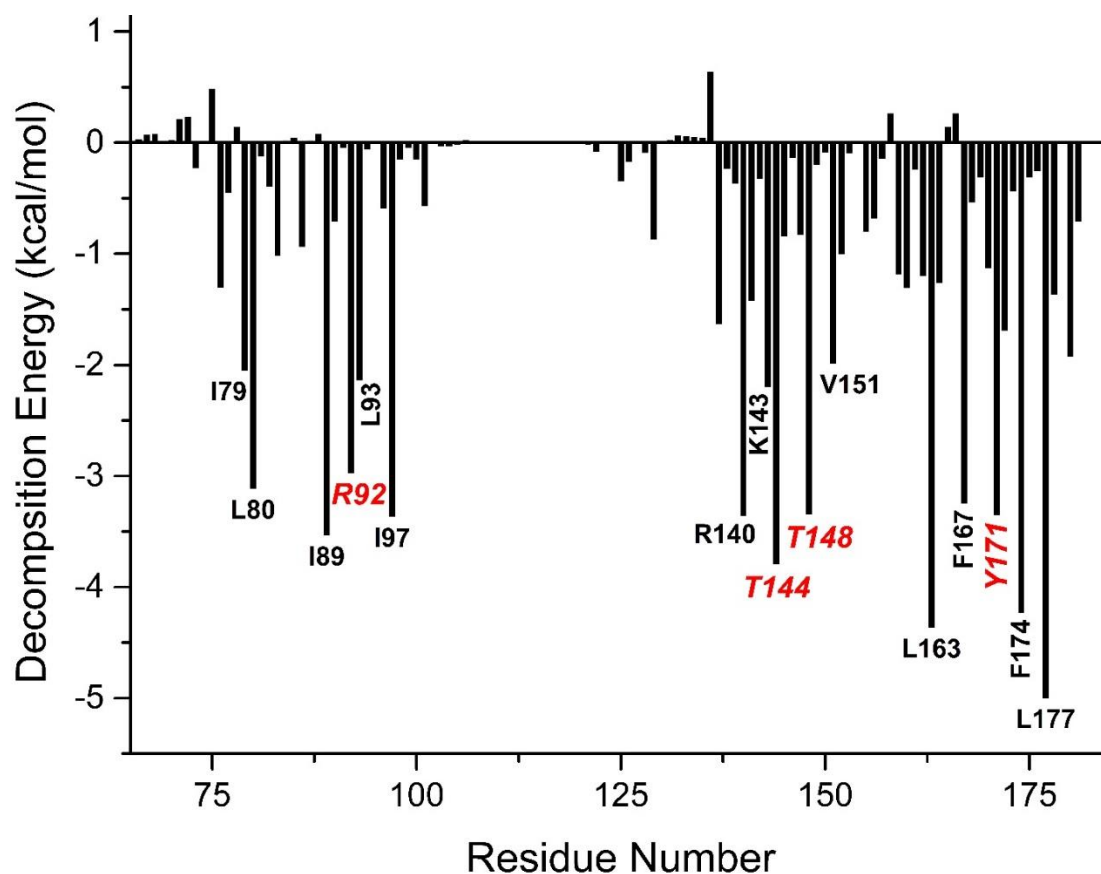
Figure 2.9. Decomposition energy of residues at the dimer interface calculated using MM/GBSA analysis.

Based on the analysis, the CspA dimer interface is largely stabilize by non-polar

interactions with very few hydrogen-bonding interactions. Most of these hydrogen bonds are at

the C-terminal end of helix E, with only one hydrogen bond stabilizing the center of the interface

(Fig. 2.10). In order to gain a better understanding of the role of this hydrogen bond in overall

stability of the CspA dimer, we calculated the pattern of all intermolecular hydrogen bonds at the dimer interface using MD simulation and analyzed those interactions that showed occurrence frequency of at least 40% of the simulation time (Table 2.2).
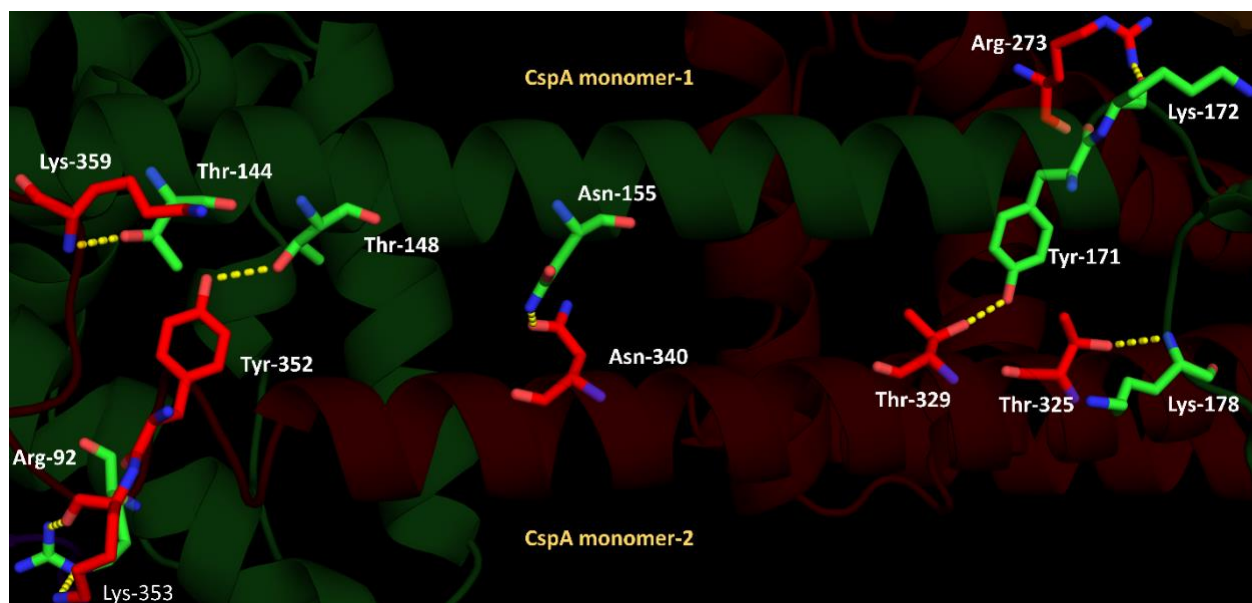
Figure 2.10. Hydrogen bonds at the CspA dimer interface with interaction frequency of at least 40% of the simulation time. The C-terminal helix that forms the main interacting region for the dimer from monomer-1 and 2 are colored green and red respectively.

A distance cutoff of 3.3 Å (between acceptor and donor heavy atoms) and angle cutoff of 120° (between donor heavy atom-hydrogen-acceptor heavy atom) was used to track all hydrogen bonds at the CspA interface. Based on the MD simulation analysis, there are six stable hydrogen bonds at the CspA interface among which hydrogen bonds involving Lysine and Threonine residues contribute the most towards interface stability. The stability contribution of N155-ND2…..OD1-N340 and R92-NH2…..O-K353 hydrogen bonds are minimal. Further, the only hydrogen bond at the center of the dimer interface, N155-ND2……OD1-N340, that acts as a pivot point for the fluctuation of long helix E occurs for only for ~ 44% of the simulation, suggesting relatively little contribution of this residue to dimer stabilization. When this central interaction

weakens, the two long E helices from each monomer can fluctuate more and the dimer starts to

dissociate. Interestingly, when we analyzed the dimer interface in the initial crystal structure,

there are more CspA intermonomer residues within 3.3 Å distance that can form hydrogen bonds.

Most of these residues are from the loop regions connecting helix D and E. However, simulation

results indicate that these hydrogen bonds do not exist for long periods due to drastic

fluctuations in loop structure as indicated by the RMSF plot.

Table 2.2. Intermolecular hydrogen bonds at the CspA dimer interface evaluated from MD
simulation

Monomor-1……. Monomor-2	Occurrence (%)	Distance (Å)	Crystal Structure Distance (Å)
K178N……………OG1T325	98.90	2.93 ± 0.13	3.11
T144-OG1………N-K359	98.57	2.94 ± 0.13	3.31
T148-OG1……..OH-Y352	98.39	2.84 ± 0.16	3.22
Y171-OH……….OG1-T329	97.80	2.85 ± 0.18	2.77
R92 NE………….O-K353	89.41	2.89 ± 0.55	3.50
K172-O………….NE-R273	87.74	2.86 ± 0.58	3.74
N155-ND2……..OD1-N340	44.08	3.88 ± 1.67	2.24
R92-NH$_2$……….O- K353	41.51	3.93 ± 1.92	2.41
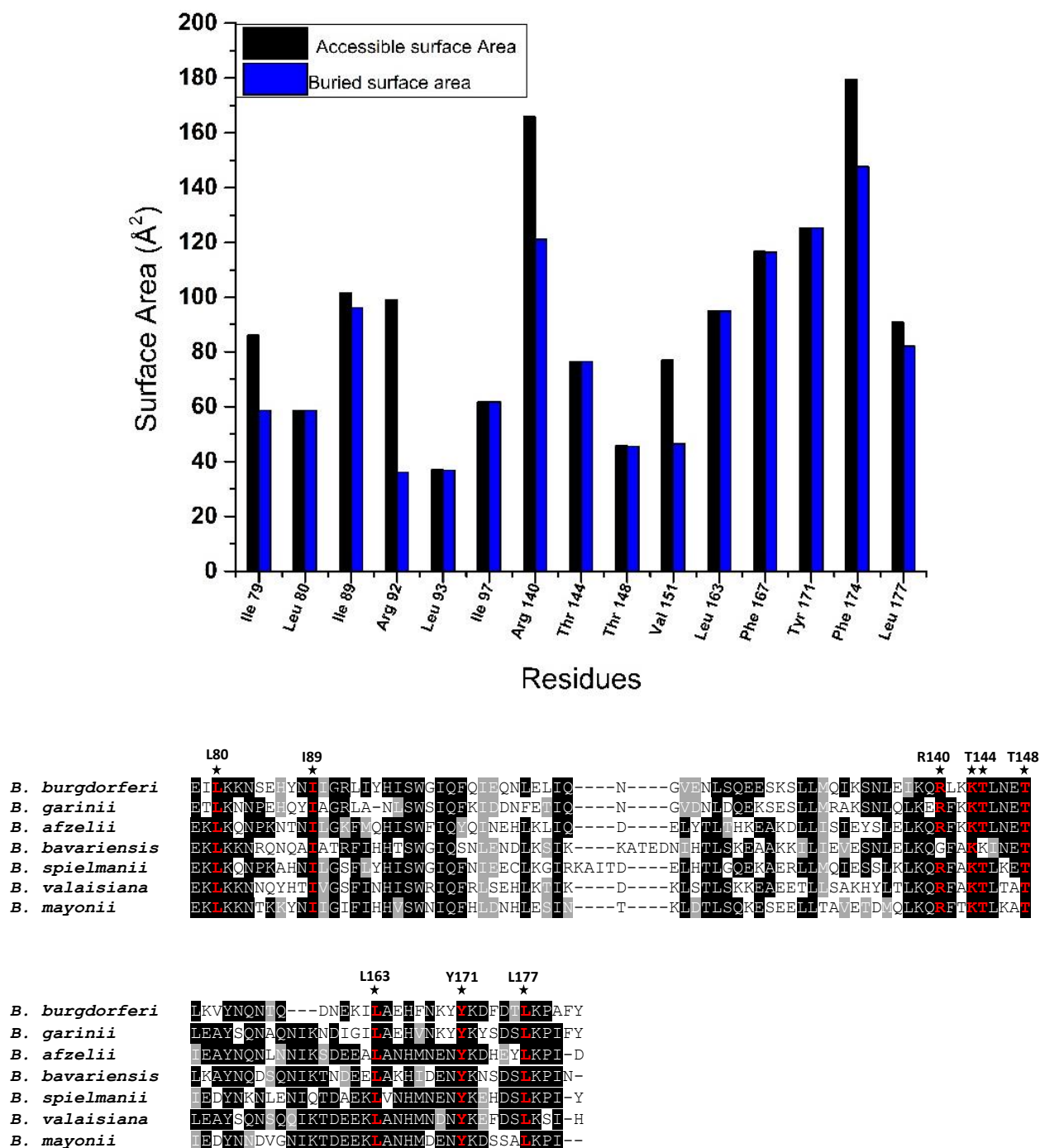
Figure 2.11. (Top) Comparison of accessible surface area versus buried area of key residues contributing to dimer formation. (Bottom) Sequence analysis of CspA among different genospecies of *Borrelia.* Key residues with highest contribution in binding at the dimer interface are colored in red. Most of the hydrophobic residues are invariantly conserved among all borrelial genospecies.

Sequence analysis of major borrelial genospecies shows that the hydrophobic core at the dimer interface is well conserved among all major borrelial genospecies. L80, I89, L163 and L177 are invariantly conserved among all genospecies of *Borrelia*, further supporting the importance of non-polar interactions for the stability of the CspA dimer. As expected, the key residues at the dimer interface indicated by the energy decomposition energy calculation are deeply buried in the complex, further stabilizing the dimer interface (Fig. 2.11).

2.4. Discussion

In this study, we have gained some insights into the dynamics of the unbound CspA dimer. It was suggested based one the two independent crystal structures of CspA that the dimer shows a great deal of flexibility along the long E helix, which forms the main interaction surface for CspA dimerization [3]. The architecture of the cleft has been assigned as a suitable site for accommodation of a single FH domain with a clamping mechanism while excluding antibodies [1, 6]. Recently, several mutagenesis studies have been performed on several residues at the cleft region and dimeric interface that are proposed to bind directly to FH protein [4]. However, our crystal structure and characterization of the complex between CspA and FH do not support the proposition of the cleft region as the FH binding site.

The study presented in this chapter focused on analysis of the free form of CspA dimer. The CspA dimer is largely stabilized by hydrophobic interactions with few hydrogen bonding interactions. There is a single hydrogen bond between Asn155-Asn340 at the center of the helix E that acts as a pivot point for the fluctuation of the interaction surface. Hydrogen bond analysis shows that this central hydrogen bond only exists for ~ 44% of the time, indicating that this interaction has much less contribution to the stability of the dimeric interface and weakens around 390 ns, causing high fluctuation across the major interface helix E.

The two conformations shown in two published crystal structures of CspA dimer are supported by our simulation study. However, the change in distance between helix C of the two monomers as shown in our study supports the notion change in dimeric cleft size is predominantly due to the movement in helix C at the base of dimeric cleft and not due to the movement of helix A covering the opening of the cleft. The subsequent change in angle between

helix C and D further supports the movement of helix C as the major initiation factor for the conformational change at the dimeric interface.

The relevance of the dimeric interface of CspA to FH binding is currently not clear, as our studies show that the dimeric cleft is not strictly required for FH binding. However, gaining insights into the CspA dimeric interface and the cleft region of CspA might be beneficial to improve study design and to explore additional biological relevance of CspA protein that can greatly aid in developing Lyme disease therapeutics in the future.

REFERENCES

REFERENCES

1. Cordes, F.S., et al., *Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of Borrelia burgdorferi.* Int J Med Microbiol, 2006. **296 Suppl 40**: p. 177-84.

2. McDowell, J.V., et al., *Demonstration of the involvement of outer surface protein E coiled coil structural domains and higher order structural elements in the binding of infection-induced antibody and the complement-regulatory protein, factor H.* J Immunol, 2004. **173**(12): p. 7471-80.

3. Cordes, F.S., et al., *A novel fold for the factor H-binding protein BbCRASP-1 of Borrelia burgdorferi.* Nat Struct Mol Biol, 2005. **12**(3): p. 276-7.

4. Kraiczy, P., et al., *Mutational analyses of the BbCRASP-1 protein of Borrelia burgdorferi identify residues relevant for the architecture and binding of host complement regulators FHL-1 and factor H.* Int J Med Microbiol, 2009. **299**(4): p. 255-68.

5. Kraiczy, P., et al., *Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi.* Infect Immun, 2001. **69**(12): p. 7800-9.

6. Caesar, J.J., et al., *Further structural insights into the binding of complement factor H by complement regulator-acquiring surface protein 1 (CspA) of Borrelia burgdorferi.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2013. **69**(Pt 6): p. 629-33.

7. Lesk, A.M., *Introduction to Protein Science: Architecture, Function, and Genomics*, in *Biochemistry and Molecular Biology Education*. 2006, Oxford University Press: Oxford, United Kingdom.

8. Kraiczy, P., et al., *Complement resistance of Borrelia burgdorferi correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins.* J Biol Chem, 2004. **279**(4): p. 2421-9.

9. D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, et al., *AMBER 2016*. 2016, University of California, San Francisco.

10. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.

11. Gordon, J.C., et al., *H++: a server for estimating pKas and adding missing hydrogens to macromolecules.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W368-71.

12. Coleman, T.G., H.C. Mesick, and R.L. Darby, *Numerical integration: a method for improving solution stability in models of the circulation.* Ann Biomed Eng, 1977. **5**(4): p. 322-8.

13. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

14. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**.

15. Roe, D.R. and T.E. Cheatham, 3rd, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013. **9**(7): p. 3084-95.

16. *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*

17. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, Palo Alto, CA, USA: DeLano Scientific.

18. Miller, B.R., 3rd, et al., *MMPBSA.py: An Efficient Program for End-State Free Energy Calculations.* J Chem Theory Comput, 2012. **8**(9): p. 3314-21.

19. Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* J Chem Inf Model, 2011. **51**(1): p. 69-82.

CHAPTER 3

Molecular Analysis of FH Recruitment by the Outer Surface Protein CspA of *Borrelia burgdorferi*

Author Contributions

Jagannath Silwal expressed and purified all proteins, carried out molecular dynamics simulations, MM-GBSA analysis, performed alanine scanning mutagenesis and did all ITC experiments; Aizhuo Liu performed crystallization screening and optimization, structure determination, and X-ray diffraction data collection; Yue Li made the overexpression systems; Honggao Yan conceived the project and supervised this work.

3.0. Abstract

CspA is a major surface exposed lipoprotein expressed by *B. burgdorferi.* Expression of CspA is directly corelated with the ability of *Borrelia* to successfully evade the host immune system, leading to persistent Lyme infection in humans and many animals. *B. burgdorferi* spirochetes expressing CspA protein alone on their surface can escape complement-mediated attack from the host, making CspA one of the most important surface proteins in *Borrelia*. One of the major ways *Borrelia* avoids immune attack from the host is by 'hijacking' immune regulator proteins, such as complement factor H (FH) protein, binding them on their own surface using CspA. However, the molecular details of protein-protein interactions between CspA and hFH (hereafter designated as CspA:hFH) is not clear. Many published results are contradictory, adding further challenges in getting clear insights on details of host-pathogen interactions. Our study reveals that out of the 20 CCPs that constitute hFH, CCP5 alone is sufficient for tight binding of CspA, with K_d of 1.3 nM. Similarly, the high resolution structure and extensive characterization study of CspA:hFH CCP5 complex in our lab reveals a unique FH binding site in CspA that is different from previously suggested cleft region between two CspA monomers. Instead, FH binds at the 'head' region of CspA aligning parallel to helix C. Our results invalidate many previously published results on FH binding to CspA and provides new insights into host specificity of Lyme disease.

3.1. Introduction

Pathogens must survive attack from the human immune system to successfully cause persistent infection. The human complement system is a key part of the innate immune system and is the first line of defense against invading pathogens [1, 2]. While the complement system plays an important role in human immunity, the major complement activation pathway (the alternative pathway) cannot distinguish self from non-self-cells [3]. Therefore, the host must rely on other mechanisms to protect its own cells from complement-mediated attack. One of the major mechanisms to protect host cells from auto immune attack is to recruit complement factor H (FH) to the surface of self-cells [4-6]. FH is a large complement control protein with 20 individual domains commonly termed complement control protein (CCP) modules. FH inhibits and/or downregulates both activation and amplification of complement activation by inhibiting and promoting disintegration of interactions between key immune regulator enzymes. Taking advantage of this key function of FH, a wide range of pathogens have evolved with various mechanisms to recruit FH protein to their own surface. Hence, pathogen can successfully evade host immune system, eventually leading to persistent infection of the host. *Borrelia* spp [7], *Neisseria gonorrhoeae [8-13]*, *Haemophilus parasuis* [14], *Streptococcus Pneumoniae [15-24]*, *Staphylococcus aureus [25, 26]*, *Rickettsia conorii [27]*, *Salmonella* spp [28, 29] and *Neisseria meningitidis [30]* are some examples of bacteria that hijack FH from the host to cause infection. In addition, some species of fungi (*Aspergillus fumigatus* [31] and *Candida albicans* [32, 33]*)* and parasites (*Echinococcus granulosus* [34], *Loa loa* [35] and *Onchocerca volvulus* [36]) also use FH recruitment as the major immune evasion strategy.

Host specificity is a key factor in determining pathogenesis and epidemiology of any disease [37, 38]. One of the key determinants of host-specificity of a pathogen is its ability to successfully circumvent immune attack from the host system [38]. Host specificity of FH recruitment is corelated to the host specificity of infection in a wide range of pathogens. Pathogens such as *S. pneumoniae, N. meningitidis* and *N. gonorrhoeae* recruit FH in human-specific manner and cause infection only to humans, indicating FH recruitment as a key contributing factor to the host specificity of these pathogens [9, 39, 40]. In contrast, pathogens such as *B. burgdorferi* can survive and infect a wide range of hosts. The broad host range character of *Borrelia* is attributed partly to its ability to express multiple FH binding proteins on its surface [41].

Lyme disease is the most common vector-borne illness in United States and Europe [42]. Bacterial pathogen from *Borrelia burgdorferi* sensu lato family including *B. burgdorferi* sensu stricto (commonly referred as *B. burgdorferi), B. afzelii, B. spielmanii, and B. bavariensis* are the most common causative agents of Lyme disease [43]. A salient feature of this zoonotic bacterial group is the ability to survive and cause persistent infection in a wide range of hosts, including human, birds, reptiles and other large mammals [42, 43]. The host association of individual borrelial genospecies is corelated with their ability to escape complement-mediated attack from the host [41]. The key mechanism employed by the species of *Borrelia* to avoid such complement attack from the host is by recruiting FH on their surface using FH binding proteins [7]. However, discerning the role of FH recruitment in borrelial pathogenesis and host specificity is challenging for several reasons. First, a single *Borrelia* pathogen expresses up to five different FH binding surface proteins, including CspA, CspZ, ErpA, ErpC, and ErpP. Second, *Borrelia* has one of the most

68

complex prokaryotic genomes with long linear chromosome and up to 21 linear and circular plasmids. These plasmids contain many redundant sequences that can vary significantly, resulting in many genospecies of *Borrelia*. Third, many published studies are performed using non-equilibrium methods such as enzyme-linked immunosorbent assay (ELISA) and far western blotting without proper quantification of FH protein in sera samples.

In our lab, we have been employing rigorous biochemical and biophysical methods and well-defined systems to overcome the above issues in *Borrelia* research. During this process, the crystal structures of CspA, CspZ and ErpA in complex with specific CCPs of hFH were solved and characterized. The structure of the CspA and hFH complex along with results from extensive characterization of this protein complex using computational studies, site-directed mutagenesis and ITC are presented in this chapter.

3.2. Material and Methods

3.2.1. CspA Production and Purification

The DNA encoding the outer surface protein CspA of *Borrelia burgdorferi* strain B31 was amplified by PCR from the genomic DNA. The following reaction conditions were used for PCR: 95 °C for 30 s, 56 °C for 30 s and 72 °C for 60 s for 30 cycles. The amplified product was cloned into the lab-made overexpression vector pET17bHR (derived from the Novagen pET17b vector, with addition of an N-terminal His-tag and a TEV protease cleavage site) by digestion with the restriction enzymes BamHI and NdeI, followed by ligation. The cloned DNA fragment was sequenced to ensure the correct coding sequences.

The overexpression plasmid construct was transformed into the *E. coli* strain BL21(DE3) for the production of histidine-tagged CspA. The expression system was cultured in LB medium containing 100 µg/ml ampicillin with vigorous shaking (225 rpm) at 37 °C until OD_{600} reached 0.8. IPTG was added to a final concentration of 0.4mM to induce heterologous protein production. The culture was incubated with shaking for another 6 hours at 37 °C.

The cells were harvested by centrifugation at 5,000 rpm for 10 minutes and the cell pellets were resuspended in buffer A containing 20 mM Tris-HCl and 150 mM NaCl, pH 7.5, supplemented with 5 mM $MgCl_2$ and 5 µg/ml of DNase I. Sonication was used for cell lysis and the lysate was cleared by centrifugation at 15,000 rpm for 30 minutes before loading onto a Ni-NTA agarose column, pre-equilibrated with over 10 times of the column volume of buffer A. After washing with 5 mM imidazole in buffer A, the column was eluted with a linear imidazole gradient of 5-250 mM in buffer A. Fractions containing pure CspA were analyzed by SDS-PAGE, pooled and concentrated to the final volume of ~ 20 mL using an Amicon concentrator. The histidine tag was

cleaved with TEV (tobacco etch virus) protease using 2 OD_{280} of TEV per 100 OD_{280} of CspA protein, supplemented with 1.0 mM DTT and 0.5 mM EDTA. Progress of the cleavage reaction was tracked by running SDS-PAGE and after verifying the completeness of the cleavage reaction, the cleaved protein was separated from uncleaved protein by loading onto a second Ni-NTA column pre-equilibrated with buffer A. The flow through was collected, concentrated and loaded onto a Sephadex G-75 gel filtration column preequilibrated with buffer A. The fractions containing pure CspA were pooled, concentrated and dialyzed against buffer A and was stored at 4 °C. The final protein yield was ~200 mg per liter of LB culture.

3.2.2. hFH CCP5 Production and Purification

Synthetic DNA, with codons optimized for expression in *E. coli*, coding for human FH (hFH) and mouse FH (mFH) were synthesized by Biomatik. The expression vector pET17bHMHT (derived from pET17b) was used to clone the DNA fragments encoding various CCPs and the correct coding sequences were verified by Sanger sequencing. The overexpression plasmid constructs were transformed to the *E. coli* strain SHuffle T7 *LysY*. The FH proteins were produced as fusion proteins with maltose binding protein (MBP), with two histidine-tags, one before the MBP and another between the MBP and CCPs. A thrombin cleavage site was also present between the second histidine tag and the CCPs to allow for the cleaving of the fusion partner MBP. The expression system was cultured in LB medium containing 20 µg/ml chloramphenicol and 100 µg/ml ampicillin with vigorous shaking (225 rpm) at 37 °C until the OD_{600} reaches 0.8. The culture was then cooled on ice to 16 °C and 0.5 mM IPTG was added to induce the FH protein production. The culture was incubated for another 24 hours with shaking at 16 °C.

Cells were harvested, resuspended in buffer A containing 20 mM tris-HCl and 150 mM NaCl, pH 7.5, lysed, centrifuged and loaded onto a Ni-NTA column as described earlier. The loaded Ni-NTA column was washed with 20 mM imidazole in buffer A until OD_{280} of the flow-through was less than 0.05 and the column was eluted with 20-250 mM imidazole gradient. The fractions containing pure CCPs were pooled and concentrated using an Amicon concentrator. An equimolar amount of lab-made histidine-tagged *E. coli* DsbC (a prokaryotic disulfide bond isomerase), was added to the concentrated protein for disulfide bond reshuffling. After dialyzing overnight against buffer A at room temperature, 0.75 U thrombin per mg of fusion protein along with $CaCl_2$ to the final concentration of 0.125 mM was added to facilitate the MBP cleavage reaction. After incubation at room temperature for 3 hours, the progress of MBP cleavage was checked by SDS PAGE and upon completion of the reaction, phenylmethylsulfonyl fluoride (PMSF) was added to a final concentration of 0.5 mM to terminate the cleavage reaction. The cleaved FH protein was then passed through a second Ni-NTA column to separate cleaved FH protein from the uncleaved fusion protein and flow through containing only cleaved FH protein was collected, concentrated and loaded on to a Sephadex G-75 gel filtration column for further purification (if not pure). The fractions containing pure FH were pooled, concentrated and dialyzed first against buffer containing 5 mM sodium phosphate, pH 8.0, and then against buffer containing 2 mM sodium phosphate, pH 8.0. The dialyzed protein was lyophilized and stored at -80 °C.

Table 3.1. Primers for PCR cloning and mutagenesis of CspA and FH CCP5

Name	Primer Sequence
CspAY47Af	5'-C ACT TTT AAA GTT GGT CCT GCC GAT CTT ATT GAT GAA GAT-3'
CspAY47Ar	5'- ATC TTC ATC AAT AAG ATC GGC AGG ACC AAC TTT AAA AGT G -3'
CspAD48Af	5'- TTT AAA GTT GGT CCT TAC GCT CTT ATT GAT GAA GAT ATC C -3'
CspAD48Ar	5'- G GAT ATC TTC ATC AAT AAG AGC GTA AGG ACC AAC TTT AAA-3'
CspAL49Af	5'- TTT AAA GTT GGT CCT TAC GAT GCT ATT GAT GAA GAT ATC CAA ATG-3'
CspAL49Ar	5'-CAT TTG GAT ATC TTC ATC AAT AGC ATC GTA AGG ACC AAC TTT AAA -3'
CspAD51Af	5'- GGT CCT TAC GAT CTT ATT GCT GAA GAT ATC CAA ATG AA -3'
CspAD51Ar	5'- TT CAT TTG GAT ATC TTC AGC AAT AAG ATC GTA AGG ACC -3'
CspAI54Af	5'- TAC GAT CTT ATT GAT GAA GAT GCC CAA ATG AAA ATA AAA AGA ACG-3'
CspAI54Ar	5'- CGT TCT TTT TAT TTT CAT TTG GGC ATC TTC ATC AAT AAG ATC GTA-3'
CspAY90Af	5'-CTT AAA AAA AAT TCC GAA CAT GCC AAT ATA ATT GGA AGA TTG ATT-3'
CspAY90Ar	5'- AAT CAA TCT TCC AAT TAT ATT GGC ATG TTC GGA ATT TTT TTT AAG -3'
CspAR95Af	5'-GAA CAT TAC AAT ATA ATT GGA GCA TTG ATT TAT CAC ATA TCA TGG-3'
CspAR95Ar	5'- CCA TGA TAT GTG ATA AAT CAA TGC TCC AAT TAT ATT GTA ATG TTC -3'
CspAY98Af	5'-T ATA ATT GGA AGA TTG ATT GCT CAC ATA TCA TGG GGC ATT C -3'
CspAY98Ar	5'- G AAT GCC CCA TGA TAT GTG AGC AAT CAA TCT TCC AAT TAT A-3'
CspAH99Af	5'-ATA ATT GGA AGA TTG ATT TAT GCC ATA TCA TGG GGC ATT CAA TTC -3'
CspAH99Ar	5'-GAA TTG AAT GCC CCA TGA TAT GGC ATA AAT CAA TCT TCC AAT TAT -3'
CspAW102Af	5'-GA TTG ATT TAT CAC ATA TCA GCG GGC ATT CAA TTC CAA ATA G -3'
CspAW102Ar	5'- CTA TTT GGA ATT GAA TGC CCG CTG ATA TGT GAT AAA TCA ATC-3'
CspAF106Af	5'-C ATA TCA TGG GGC ATT CAA GCC CAA ATA GAG CAA AAT TTA -3'
CspAF106Ar	5-TAA ATT TTG CTC TAT TTG GGC TTG AAT GCC CCA TGA TAT G -3'

Table 3.1 (Cont'd)

Name	Primer Sequence
hFH5D258Af	5'- G TAT ATT CCG AAT GGT GCC TAC AGC CCG CTG CGC-3'
hFH5D258Ar	5'- GCG CAG CGG GCT GTA GGC ACC ATT CGG AAT ATA C -3'
hFH5S260Af	5'-G AAT GGT GAC TAC GCC CCG CTG CGC ATT AAA C-3'
hFH5S260Ar	5'-G TTT AAT GCG CAG CGG GGC GTA GTC ACC ATT C -3'
hFH5L262Af	5'-GT GAC TAC AGC CCG GCG CGC ATT AAA CAT C -3'
hFH5L262Ar	5'-G ATG TTT AAT GCG CGC CGG GCT GTA GTC AC -3'
hFH5R263Af	5'-GAC TAC AGC CCG CTG GCC ATT AAA CAT CGT AC -3'
hFH5R263Ar	5'- GT ACG ATG TTT AAT GGC CAG CGG GCT GTA GTC-3'
hFH5R267Af	5'- G CTG CGC ATT AAA CAT GCT ACG GGC GAT GAA ATC -3'
hFH5R267Ar	5'- GATTTCATCGCCCGTAGCATGTTTAATGCGCAGC-3'
hFH5D270Af	5'- CAT CGT ACG GGC GCT GAA ATC ACC TAT C-3'
hFH5D270Ar	5'- GATAGGTGATTTCAGCGCCCGTACGATG-3'
hFH5E271Af	5'- CGT ACG GGC GAT GCA ATC ACC TAT CAG-3'
hFH5E271Ar	5'- CTGATAGGTGATTGCATCGCCCGTACG-3'
hFH5T273Af	5'- GGC GAT GAA ATC GCC TAT CAG TGC CG-3'
hFH5T273Ar	5'-CGGCACTGATAGGCGATTTCATCGCC -3'
hFH5Y274Af	5'- GC GAT GAA ATC ACC GCT CAG TGC CGT AAT GG-3'
hFH5Y274Ar	5'- CCATTACGGCACTGAGCGGTGATTTCATCGC-3'
hFH5Q275Af	5'-GGC GAT GAA ATC ACC TAT GCC TGC CGT AAT GGT TTT TAC -3'
hFH5Q275Ar	5'-GTAAAAACCATTACGGCAGGCATAGGTGATTTCATCGCC -3'
hFH5R277Af	5'- GAA ATC ACC TAT CAG TGC GCT AAT GGT TTT TAC CCG G-3'
hFH5R277Ar	5'- CCGGGTAAAACCATTAGCGCACTGATAGGTGATTTC-3'
hFH5N278Af	5'- CC TAT CAG TGC CGT GCT GGT TTT TAC CCG G-3'

Table 3.1 (Cont'd)

hFH5N278Ar	5'- CCGGGTAAAAACCAGCACGGCACTGATAGG-3'
hFH5N287Af	5'-TT TAC CCG GCA ACC CGC GGC GCT ACG GCT AAA TGT ACC AGC A -3'
hFH5N287Ar	5'-TGCTGGTACATTTAGCCGTAGCGCCGCGGGTTGCCGGGTAAA -3'
hFH5T288Af	5'-CA ACC CGC GGC AAC GCT GCT AAA TGT ACC AGC -3'
hFH5T288Ar	5'-GCTGGTACATTTAGCAGCGTTGCCGCGGGTTG -3'

3.2.3. Site-Directed Mutagenesis

Site-directed mutagenesis performed according to a Quik-Change ™ protocol from Strategene. The mutagenic primers employed for this study are listed in Table 3.1. PCR reactions were carried out for 18 cycles (95 °C for 30s, 56 °C for 30 s and 72 °C for 8 minutes) using 40 ng of the expression construct for CspA or hFH CCP 5, with each mutagenic primer to a final concentration of 10 µM, dNTPs to the final concentration of 10 mM and 2.5U p*futurbo* DNA polymerase (Strategene) in a final reaction volume of 50 µl. The PCR product was digested with 10 U DpnI at 37 °C for an hour to eliminate the parental plasmid. The digested product was transformed into the *E. coli* strain DH5α, and plasmid DNA was extracted and purified following the PCR clean-up protocol from Promega.

3.2.4. Isothermal Titration Calorimetry (ITC)

Isothermal titration Calorimetry (ITC) experiments were performed at 25 °C using VP-ITC system (Microcal Inc.) [44]. The lyophilized proteins were dissolved in buffer C containing 50 mM HEPES and 50 mM KCl, pH 7.5 and were extensively dialyzed against the same buffer. Before each

ITC experiment, all protein solutions were filtered, degassed to prevent bubble formation, and equilibrated to 25 °C. For a typical reaction, the sample cell was loaded with FH protein and the syringe was loaded with CspA. The experimental parameters and protein concentrations were adjusted and optimized for each mutant protein. In general, ITC experiments were carried out with 12 µl of CspA injection at 6-minute interval for a total of 20 injections. The resulting binding isotherms were analyzed and processed to obtain final thermodynamic parameters using the Microcal Origin 5.0 (OriginLab Corporation).

3.2.5. Molecular Dynamics (MD) Simulation and MM-GBSA Analysis

To investigate the dynamics of the CspA:hFH CCP5 complex, MD simulations for wild type and mutant protein complexes were carried out. All MD simulations were performed using the AMBER 16 Molecular Dynamics package and the ff14SB force field was used to describe the proteins [45, 46]. The crystal structure of the complex of CspA and hFH CCP 5 solved in our lab was used as the starting structure to generate coordinate files for subsequent MD simulations and analysis. Since the sequence identity between hFH CCP5 and mouse FH (mFH) CCP5 is close to 60%, the structure of the complex between hFH CCP5 and CspA was used as template to obtain a homology model of the mFH CCP5 using the SWISS-MODEL server [47]. Using the structure of the bound hFH CCP5 as the template, a model of mFH CCP5 bound to CspA was built using the SWISS-MODEL server [47, 48]. All mutant proteins for *in-silico* mutagenesis study were generated with the assumption that the structural integrity of the proteins is not affected by a single amino acid substitution. Before running the simulation, all protein complexes were processed through the H $^{++}$ program, which computes the pK values of ionizable groups of the protein and adds missing hydrogen atoms according to the specified pH of the environment [49]. Special attention

was given to the protonation state of histidine suggested by this program and adjustment was made accordingly.

Original water molecules from the crystal structure were removed and hydrogen atoms were added using TLEAP package from AMBER [45]. The standard protonation states were further checked using H++ online software and verified manually for consistency and accuracy [49]. An appropriate number of counterions was added to balance the resulting charge of the protein complex and the structure was solvated in a rectangular TIP3P water box, with the boundaries of at least 12 Å away from the protein atoms. Parameter files for MD, including coordinate and topology files, were generated for further processing [45, 50]. The solvated structure of the complex was energy minimized using steepest descent and conjugated gradient methods, each for 5000 steps, first with positional restraint force constant of 100 kcal/ (mol. $Å^2$) on all heavy atoms of the protein and then without any positional restraint on the whole system. The minimized system was heated linearly to 300K in 120 ps under constant volume conditions (NVT) with no position restraints. The system was further simulated for 750 ns under constant pressure and temperature conditions (NPT). Temperature was controlled using Langevin dynamics with a collision frequency of 1 ps^{-1} during heating, equilibration and production steps [51]. All covalent bonds to hydrogen atoms were constrained by the SHAKE algorithm, with a numerical integration time step of 2 fs [50]. Long range electrostatic interactions were computed using Particle Mesh Ewald (PME) method with a cutoff of distance of 10 Å [52].

The simulation results were analyzed using the CPPTRAJ package [53] from AMBER16 and the PyMOL program [45, 54]. Root mean square deviation (RMSD) was monitored throughout the simulation to ensure equilibration of the system. Molecular Mechanics/Generalized Born

Surface Area (MM/GBSA) and energy decomposition analyses were performed using MMPBSA.py

python script incorporated in AmbertTools16 [45]. A generalized Born implicit solvent model with

0.15 M salt concentration was used for all the MM/GBSA analysis [55]. The results from

MM/GBSA run were calculated using 500 snapshots for each simulation and were averaged to

estimate global free energy of binding between CspA and hFH CCP5. Since we were only

interested in relative mutational effect in interaction at the protein interface, the entropic

contribution to the free energy change was not included in the calculation. Further, entropic

calculations using currently available platform are computationally very expensive and tend to

have large margin of errors, introducing significant uncertainty in the result.

After successful MD simulation run, criteria such as density, temperature, pressure, root

mean square deviation (RMSD), root mean square fluctuation (RMSF), KE, PE and total energy

were analyzed to confirm system convergence to the equilibrium. CPPTRAJ module from

AMBER16 was used to calculate backbone RMSD values to monitor the stability of the protein

complex during the entire simulation. Root mean square fluctuation (RMSF) was calculated to

measure fluctuation of individual residues during simulation. Similarly, CPPTRAJ module was

used to analyze various hydrogen bonding properties, such as hydrogen bond distance and

occurrence of each hydrogen bond between interface residues of the two proteins. Distance

cutoff of 3.3 Å (between acceptor and donor heavy atoms) and angle cutoff of 120° (between

donor heavy atom-hydrogen-acceptor heavy atom) was used to track all hydrogen bonds. Special

attention was given to the stable hydrogen bonds, which occupied at least 50% of MD simulation

time. Hydrogen bond occurrence of more than 50 % of the total simulation time are considered

present and are discussed in the result section.

3.3. Results

3.3.1. Purification of CspA and hFH CCP5 Proteins

Both CspA and hFH CCP5 proteins were purified to homogeneity using various affinity and gel filtration chromatographic techniques described earlier (Fig. 3.1). Protocols for expression, production and purification of different CCPs from FH protein were not clearly established. In order to carry out experiments to figure out specific FH domain(s) responsible for binding to CspA, it was crucial to establish a robust protocol to produce active FH CCPs. In our lab, we successfully engineered, overexpressed and produced truncated FH proteins and hence created a library consisting of many CCPs of FH.

The presence of two disulfide bonds in each CCP of FH poses some challenges at the beginning, as correct formation of these two disulfide bonds is crucial to maintain the activity and solubility of the FH protein during the production and purification steps. The *E. coli* strains SHuffle T7 Express $_{lysY}$ was used to express all the CCPs of FH. This strain is engineered to facilitate correct disulfide bond formation by mutating genes responsible for reductive pathways such as thioredoxin and glutathione. This helps to avoid reduction of disulfide bonds in the cytoplasm. Expression of all FH CCPs as a fusion protein with maltose binding protein (MBP) and addition of *DsbC* (disulfide bond isomerase) protein further aided in the yield of functional protein by enhancing correct disulfide bond formation.
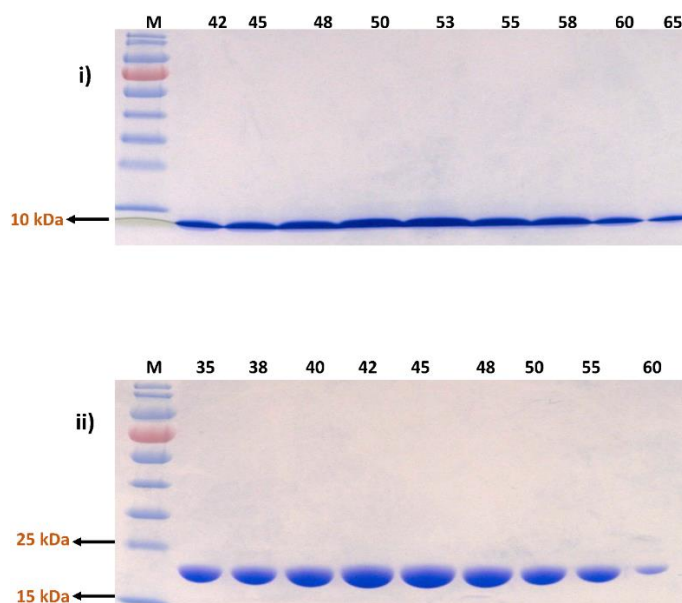
Figure 3.1. Coomassie blue (15 %) gel stain of (i) CspA and (ii) hFH CCP5 fractions collected from final sephadex-G-75 gel filtration column. M is the protein molecular weight marker and subsequent numbers indicate the protein fractions collected.

3.3.2. Localization of the CspA-Binding Region of FH

Out of all 20 CCPs of FH, which CCP binds to CspA is still unclear, with contradicting and inconsistent results in the literature. For example, one study has assigned FH CCP5-7 as key CCPs for CspA binding, whereas another study suggests only CCP 7 as the key region of FH that binds to CspA [56, 57]. In order to address this issue, we created a library of various truncated version of FH, CCP6, CCP7, CCP5-7, CCP6-7 and CCP5 and measured their respective binding affinities to CspA using ITC. Based on our ITC results, neither CCP6-7 nor CCP6 showed any binding affinity to CspA, as previously suggested by other groups (Table 3.2). CCP5-7 and CCP5 both showed very tight binding to CspA with K_d of ~ 1.3 nM, indicating that CCP5 alone is sufficient for tight binding of CspA (Fig. 3.2). Hence, our results from ITC and the crystal structure of the complex invalidate previous studies that claimed CCP7 as the FH domain binding CspA.

Table 3.2. Binding of different CCPs of FH to CspA as determined by ITC.

Proteins	Kd (nM)
hFH CCPs	
CCP6-7	No Binding
CCP6	No Binding
CCP7	No Binding
CCP5-7	1.27 ± 0.03
CCP5	1.30 ± 0.02
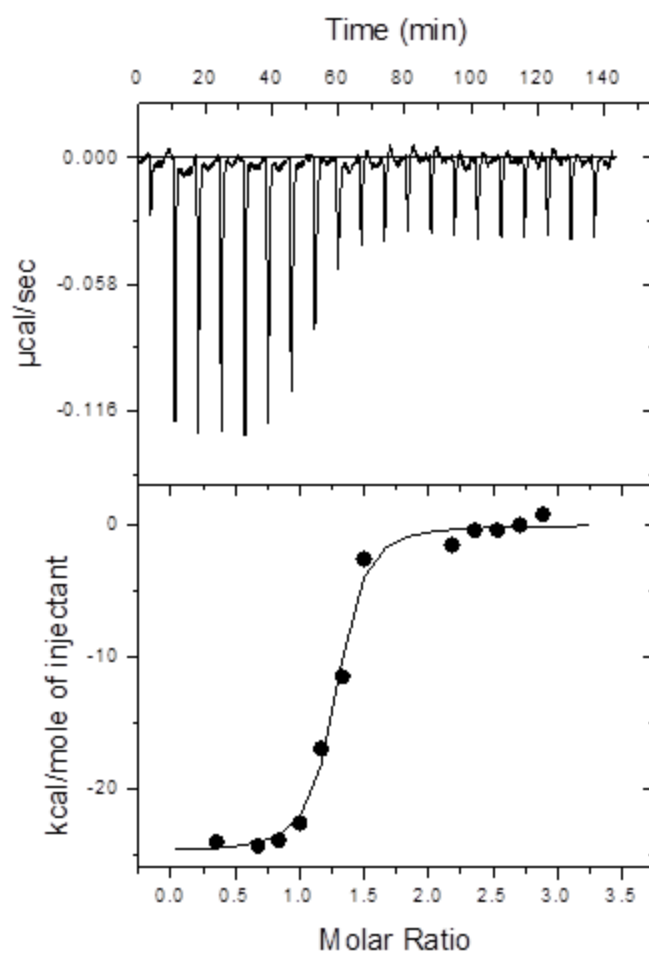
Figure 3.2. Binding of CspA and hFH CCP5. ITC binding isotherm obtained from the interaction of CspA and hFH CCP5.

Although the crystal structure of the free form of CspA was solved more than a decade ago [58], the structure of the CspA:hFH protein complex was not available. Without structure of the complex, published studies not only assigned the wrong FH CCP as CspA-binding region but also predicted the wrong FH binding region within CspA [58-60]. When the published crystal structure identified the dimeric form of CspA protein [58], several published studies suggested the cleft region between two monomeric CspA as the binding site for FH (Fig. 3.3) [59, 60]. Also, many residues at the dimeric interface and the cleft region were identified as key residues important for FH binding [60]. However, based on the crystal structure of the complex between CspA and hFH CCP5 solved in our lab (unpublished), the suggested cleft region on CspA dimer is not the FH binding site in CspA. Here, we present the crystal structure and the biophysical and computational analysis of CspA:hFH CCP5 complex.
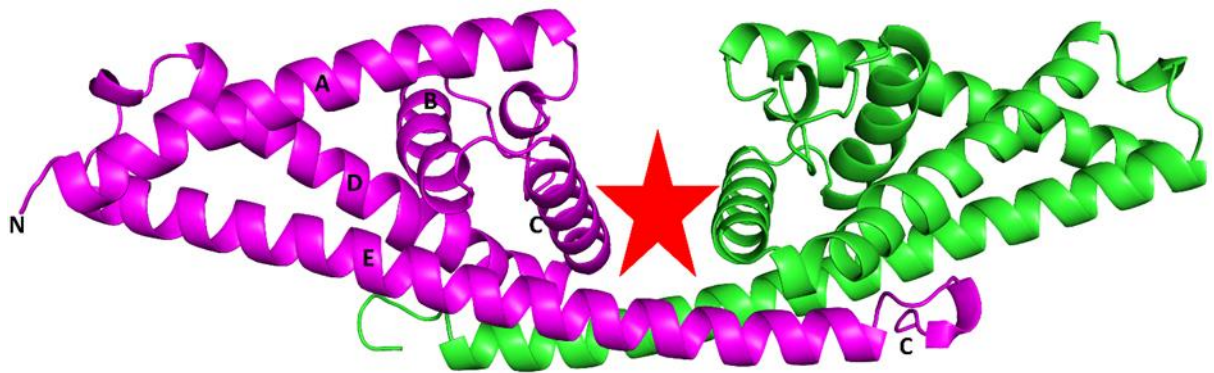
Figure 3.3. Ribbon representation of the CspA dimer. Two monomers of CspA are colored purple and green and helices are labelled A-E. The red star at the dimeric cleft represents the suggested FH binding site based on the published results.

82

3.3.3. Structure of the Complex of CspA and hFH CCP5

ITC experiments show that the binding of CspA and hFH CCP5 is very tight, with K_d of ~ 1.3 nM. Based on the structure of the complex, hFH CCP5 binds to the N-terminal 'head' region of CspA at an angle almost parallel to helix D (Fig. 3.4). The interface between CspA and hFH CCP5 contains 22 residues from CspA and 23 residues from CCP5. All the interface residues in CspA that are stabilizing the complex with hydrogen bond interaction are either from helix D or from the loop region connecting helix A and B. The key interface residues in hFH CCP5 span from Leu262 to Asn278, including a 10 amino acid long loop region connecting two β-strands.
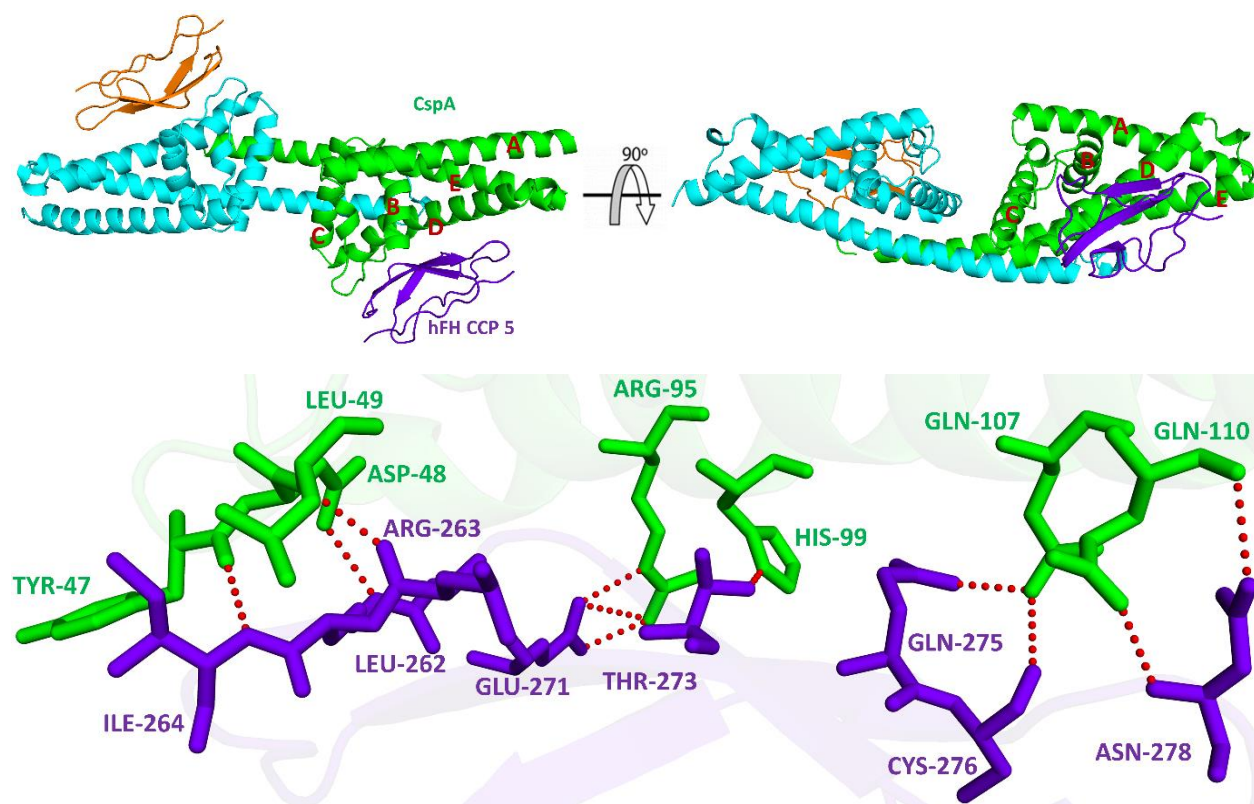
Figure 3.4. (Top) Overall structure of the complex of CspA and hFH CCP5. One subunit of the dimeric CspA is colored in cyan and the other in green. hFH CCP5 is colored in orange and purple. (Top-right) Alternate view of the complex with 90° rotation along the horizontal axis. (Bottom) Zoomed in view of the CspA:hFH CCP5 interface. Analysis of interface residues showing extensive hydrogen bonding interactions between CspA (green) and hFH CCP5 (purple) residues.

3.3.4. MD Simulation Analysis of CspA:hFH CCP5 Complex

3.3.4.1. RMSD & RMSF Analysis of CspA:hFH CCP5 Complex

The CPPTRAJ module provided by Amber16 was used to evaluate RMSD values (Fig. 3.5). The backbone RMSD was first calculated for individual proteins in the complex by superimposing individual proteins. The RMSD values of the individual proteins were low, with an average RMSD of only 0.78 ± 0.02 Å for two hFH CCP5 and 2.39 ± 0.12 Å for CspA subunits. However, the RMSD of the whole complex showed some distinct fluctuations and increases rapidly at round 390 ns. Interestingly, the 390 ns simulation time also corresponds to our earlier studies where free form CspA dimer starts to dissociate around this time. The average binding free energies before and after 390 ns for free CspA dimer as well as hFH CCP5 bound CspA dimer drastically decrease after 390 ns. Contribution from both Van der Waals and electrostatic energy significantly decreases after 390 ns with ~ 3.5-fold decrease in total binding energy. This indicates that the binding of hFH CCP5 has little to no effect in dynamics of the CspA dimerization, further invalidating the idea that CspA dimerization is required for FH binding. The absence of detectable FH binding in previous studies of C-terminal deleted CspA is most likely caused by the large global conformational changes in CspA protein due to the deletion, making the FH binding site in CspA inactive.

Table 3.3. Binding free energies of bound form CspA dimer over 0 to 390 ns and 391 to 750 ns of simulation time

Simulation time	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{total}
0 to 390 ns	-199.1 ± 3.0	-442.5 ± 41.3	504.4 ± 41.0	-28.2 ± 0.3	-165.3 ± 2.3
391 ns – 750 ns	-68.7 ± 9.6	-265.5 ± 19.2	297.5 ± 20.4	-10.2 ± 1.2	-46.8 ± 6.3

ΔE_{vdw} = van der Waals contribution, ΔE_{ele} = electrostatic energy, ΔG_{GB} = solvation free energy , ΔG_{SA} = non polar contribution to the solvation energy
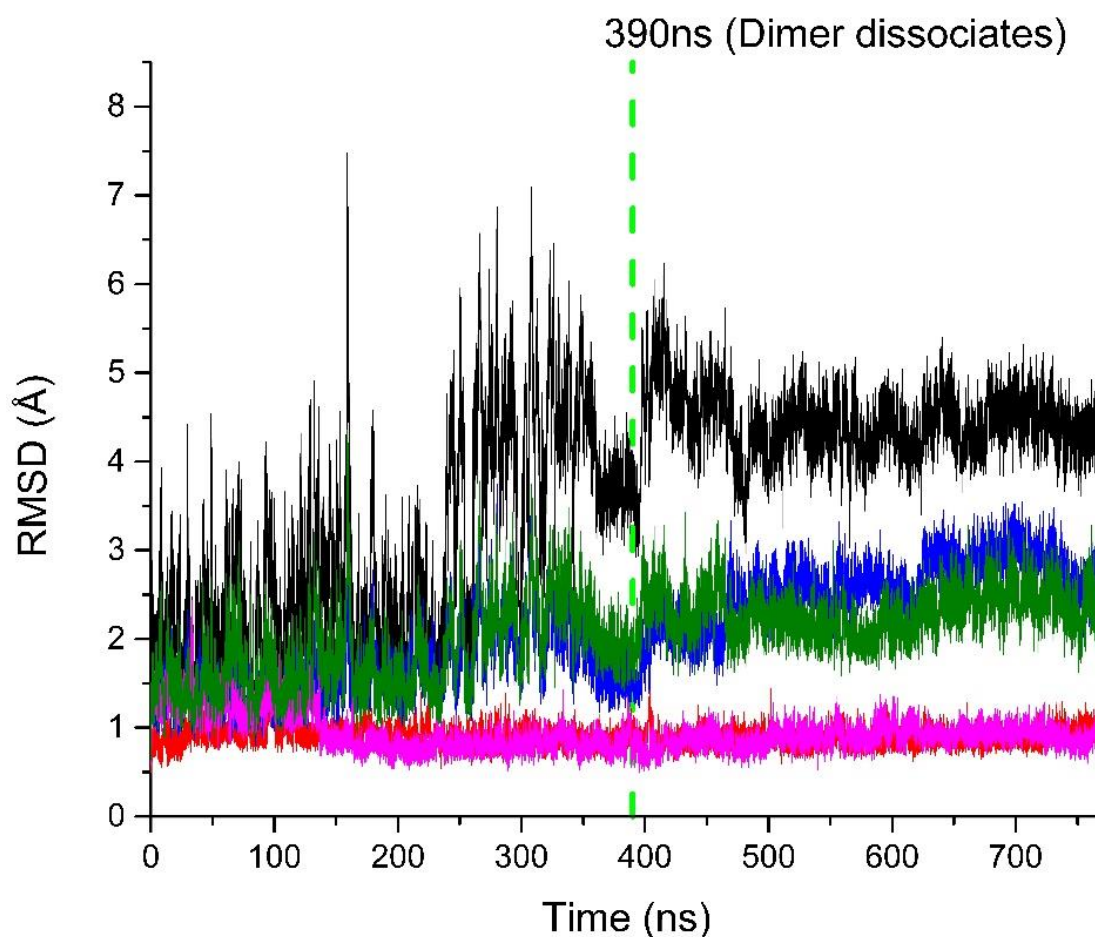
Figure 3.5. Backbone RMSD plot. Red and pink plot represent the backbone RMSD of the two hFH CCP5 and green and blue color plots represent RMSD of each CspA monomer. The RMSD plot for the whole complex is colored in black.

Similar to the RMSF plot from free form CspA dimer, the RMSF plot of the bound form CspA:hFH CCP5 complex also showed maximum fluctuations around the loop regions (Fig. 3.6). However, the loop connecting helix A and helix B was significantly stabilized in the bound form compared to the free form CspA dimer. This is expected because three residues in this loop region, Tyr47, Asp48 and Leu49, are making hydrogen bonding interactions with CCP5 that restrict the fluctuation of this region in the bound form. No other major differences in RMSF were noted between the bound and free form CspA.
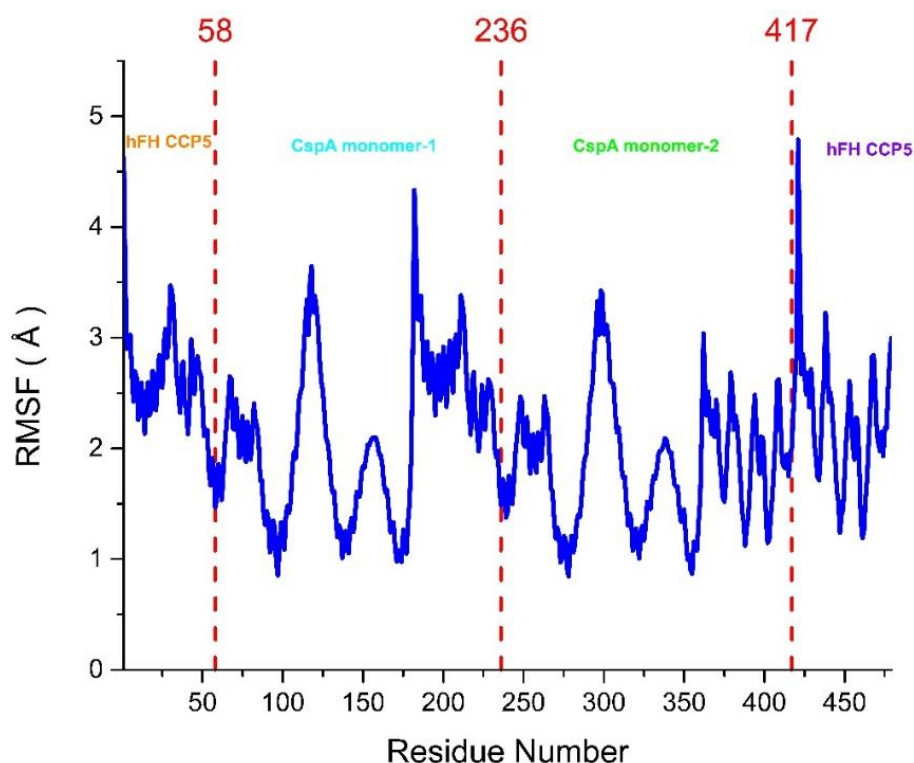
Figure 3.6. Backbone RMSF of CspA:hFH CCP5 complex. Red lines indicate cutoff residues number for each protein in the complex.

3.3.4.2. Identification of Key CspA Residues Important for FH Binding

After validating convergence and stability of the simulation system, the binding energy contribution of interface residues in CspA were analyzed (Fig. 3.7). Out of 22 total interface residues in CspA, the energy decomposition calculation shows that 11 residues are contributing significantly to binding of FH CCP5. While a majority of the CspA residues contribute to the binding via hydrogen bond interactions, there are a few hydrophobic pockets that are contributing significantly to binding as well. Specifically, residues L49, I54, W102 and F106 have the highest non-polar contribution to binding of hFH CCP5.
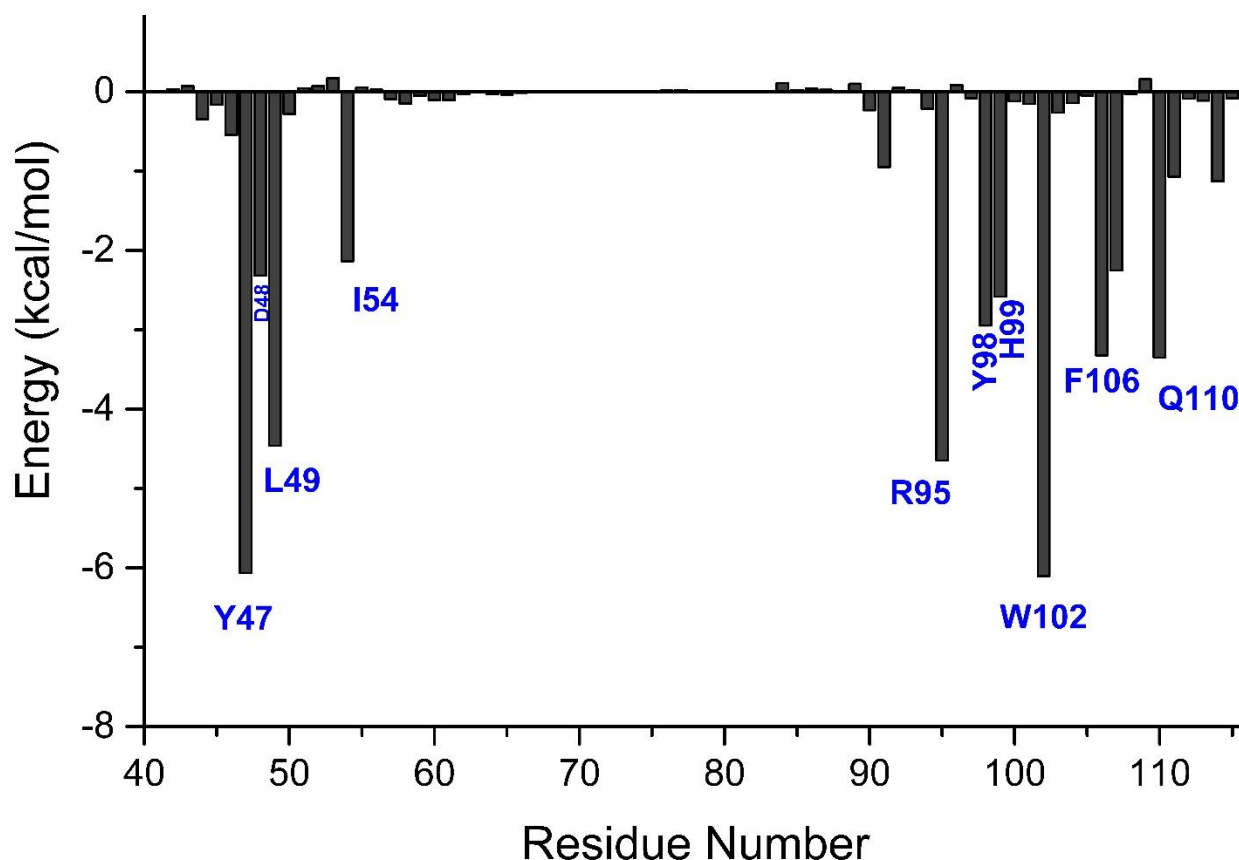
Figure 3.7. Binding energy decomposition of CspA residues. Residues are labeled and numbered in blue.

Energy decomposition calculation shows two distinct clusters of residues in CspA that are interacting with hFH CCP5. Residues from the loop region between helix A and B, Y47, D48 and Leu49, are all involved in hydrogen bonding interactions with hFH CCP5. More than 80 % of accessible surface area of Ile54 is buried upon complex formation. I54 forms a small hydrophobic pocket with residues from FH, contributing to the overall binding energy. Similarly, residues R95, Y98, H99 and Q110 are all from helix D and are also making hydrogen bonding interactions with hFH CCP5. More than 90% of F106 accessible surface area is buried with its hydrophobic ring centered at the binding interface surrounded by residues of FH CCP5 in a hydrophobic pocket. So, both clusters of key residues in CspA are mainly stabilized by many hydrogen bonding interactions and a key hydrophobic interaction. Hydrogen bond analysis from computational

studies shows that hydrogen bond between L49-L262 and Y47-I264 are the most stable

interactions at the interface with occurrence of more than 90% of the simulation time (Table 3.4).

Three hydrogen bonds exist for more than 80% of the time and five other hydrogen bonds exist

for at least 60% of the time. The standard deviation of hydrogen bond distance between C276-

Q107 and N278-Q110 during the simulation is higher, indicating bigger fluctuation between these

residue pairs. This is expected, as Q107 and Q110 are both on the long loop region of hFH CCP5,

creating more fluctuations than other residues forming hydrogen bonds at the interface.

Table 3.4. Intermolecular hydrogen bond at the interface of the complex of CspA and hFH CCP5

hFH CCP5……… CspA	Occurrence (%)	Distance (Å)	Crystal structure (Å)
L262-O…………N-L49	97.10	2.97 $\pm$ 0.21	2.81
I264-N…………O-Y47	94.60	2.96 $\pm$ 0.37	2.85
Q275-OE1……NE2-Q107	85.20	3.09 $\pm$ 0.71	2.71
R263-NE………OD1-D48	83.27	3.15 $\pm$ 0.58	2.83
T273OG1……..NE2-H99	81.32	2.85 $\pm$ 0.69	2.79
C276-O…………NE2-Q107	78.10	3.47 $\pm$ 1.14	3.14
Glu271-OE2…NH2-R95	70.01	3.09 $\pm$ 0.35	2.92
N278-ND2……O-Q110	66.40	3.85 $\pm$ 1.46	3.02
E271-OE1…….NH2-R95	60.16	3.17 $\pm$ 0.36	3.50
E271-E1………..NE-R95	59.46	3.22 $\pm$ 0.48	3.02

Further, W102 of CspA has one of the highest energy contributions to the binding of FH

CCP5. However, the crystal structure as well as hydrogen bond analyses of the MD simulation

trajectory shows that the sidechain of W102 (-NH from indole ring) is not involved in any

hydrogen bonding interactions with CCP5. Further analysis of the side chain and backbone energy

contribution of the W102 residue showed that most of the energy contribution of this residue is

due to the Van der Waals interactions of its bulky side chain with residues from hFH CCP5 (Table

3.5). A closer look at the interface shows W102 deeply buried at the center of the interface and

surrounded by residues from hFH CCP5 (Fig. 3.8). The large hydrophobic side chain of W102 is positioned like a 'key', where residues from hFH CCP5 are locking its large hydrophobic side chain via non-polar interactions at the interface. When W102 is mutated to alanine, this interaction pocket is broken, leaving a big void at the center of the interface. Hence, the W102A mutation collapses most of the interactions at the dimer interface, drastically decreasing the binding affinity of the complex.
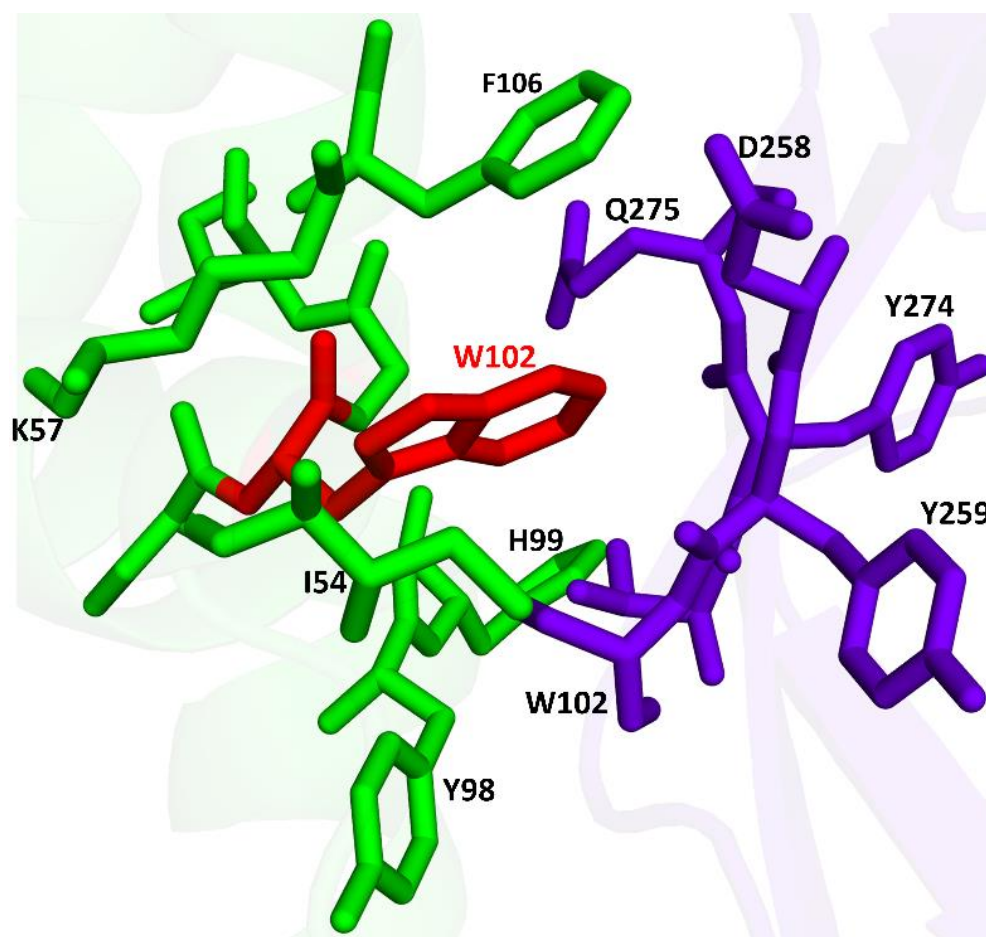
Figure 3.8. Closeup view of the binding pocket involving W102 residues of CspA in CspA:hFH CCP5 complex. W102 is colored in red and all other residues from CspA are colored in green. Residues from hFH CCP5 are colored in purple.

Table 3.5. Side chain and backbone binding free energies of <u>**CspAW102**</u>

Energy Contribution	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{total}
Side Chain	-5.48 ± 0.10	-2.28 ± 0.08	2.41 ± 0.12	-0.63 ± 0.00	-5.98 ± 0.08
Backbone	-0.18 ± 0.00	0.42 ± 0.06	-0.38 ± 0.06	0.00	-0.15 ± 0.00

ΔE_{vdw} = van der Waals contribution, ΔE_{ele} = electrostatic energy, ΔG_{GB} = solvation free energy , ΔG_{SA} = non polar contribution to the solvation energy

3.3.5. Experimental Analysis of CspA: hFH CCP5 Complex

To further validate the results from the computational study, experimental alanine scanning mutagenesis study of CspA interface residues was carried out. ITC experiments were carried out for each mutant protein to obtain thermodynamic binding parameters. Based on the results from ITC measurements, W102A mutation decreases the binding affinity of the complex by more that 2000-fold, with K_d of ~ 2.89 µM compare to ~1.3 nM for the wild type. Similarly, Y98A mutation decreases the binding affinity by a factor of 100. Our analysis shows that the -OH group from Tyr98 doesn't make any hydrogen bond interactions at the interface, suggesting that this hydrophobic sidechain contributes to the complex stability via non-polar interactions. D48A, L49A, I54A, R95A and H99A all showed at least 10-fold decrease in binding affinity, further supporting results from the computational analysis.

ITC data showed a significant loss of enthalpy and positive gain of entropy for the W102A mutation, supporting the results from computational analysis that this mutation collapses significant interactions at the binding interface (Table 3.6). The final $\Delta\Delta$G for most of the residues selected for mutagenesis are ~ 2 kcal/mol, which correlates with average energy of a single hydrogen bond. The W102A mutation showed $\Delta\Delta$G of ~ 4.65 kcal/mol indicating quite significant non-polar stabilization of protein complex by W102.

Table 3.6. Thermodynamics binding parameters for CspA interface residues

CspA Mutants	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	T ΔS (kcal/mol)
WILD	*1.3 ± 0.02*	*-12.21 ± 0.02*	*-21.71 ± 1.02*	*-9.5 ± 0.16*
Y47A	6.89 ± 1.72	-11.17 ± 0.16	-15.20 ± 0.65	-4.05 ± 0.71
D48A	46.41 ± 8.24	- 10.04 ± 0.09	-16.95 ± 1.95	-6.91 ± 1.99
L49A	14.74± 3.74	-10.74 ± 0.15	-19.54 ± 0.38	-8.80 ± 0.22
D51A	7.92 ± 1.03	-11.11 ± 0.16	-16.87 ± 2.70	-5.70 ± 2.81
I54A	15.29 ± 2.49	-10.66 ± 0.11	-19.69 ± 0.91	-9.03 ± 1.01
Y90A	2.97 ± 0.63	-10.37 ± 0.13	-19.64 ± 1.68	-7.97 ± 1.54
R95A	30.20 ± 6.3	-10.35 ± 0.11	-18.62 ± 1.46	-8.32 ± 1.57
Y98A	161.65 ± 23	-9.27 ± 0.08	-20.2 ± 1.32	-11.34 ± 1.33
H99A	13.41 ± 3.32	-10.84 ± 0.26	-21.43 ± 0.11	-10.55 ± 0.14
W102A	2890 ± 369	-3.64 ± 0.07	-5.88 ± 2.11	-2.24 ± 0.22
F106A	11.58 ± 1.20	-10.83 ± 0.06	-13.50 ± 1.39	-2.67 ± 1.41

Sequence analysis shows that D48, H99 and W102 are invariantly conserved among all genospecies of *Borrelia* (Fig. 3.9). In addition to the contribution of these residues in stabilizing the complex of CspA:hFH CCP5, based on the sequence conservation, these residues might be important to maintain the structural integrity of the CspA itself. Y47 is conserved in *B. burgdorferi* and *B. afzelii* but not in *B. garinii*. Similarly, Leu49 is conserved in *B. burgdorferi* and *B. garinii* but not in *B. afzelii*. Hydrophobic residue F106 is present in *B. burgdorferi* but absent in both *B. garinii* and *B. afzelii*. Overall, there is a good agreement between relative energy contribution of interface residues between computational and experimental results.
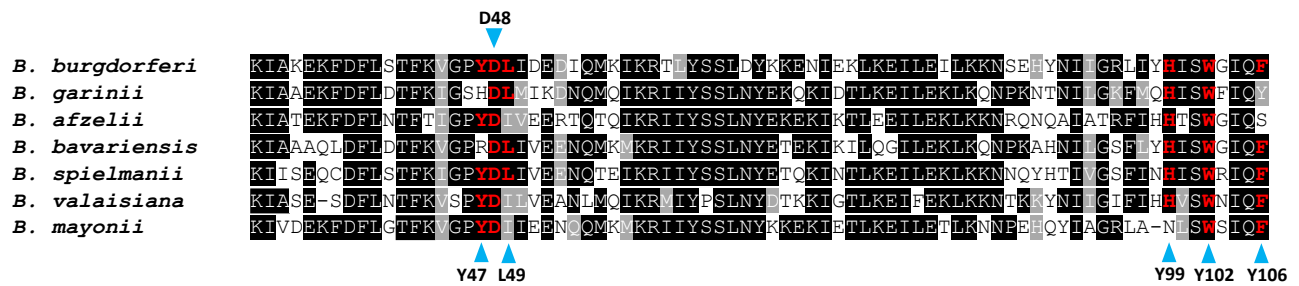
Figure 3.9. Sequence alignment of CspA residues from seven different genospecies of *Borrelia*. Residues colored in red are key residues important for binding in CspA from *B. burgdorferi*. Key residues conserved are also colored red.

3.3.6. Identification of Key FH Residues Important for Binding to CspA

After identifying key CspA residues important for formation of the CspA:hFH CCP5 complex, the same approach was taken to identify key residues in hFH CCP5 that are important for binding of CspA. The energy decomposition analysis for the interface residues of hFH CCP5 showed that R263 contributes significantly in CspA binding (Fig. 3.10). R263 forms a salt bridge with D48 of CspA and this salt bridge is quite stable with more than 80% occurrence as indicated by the hydrogen bond analysis. The backbone oxygen of L262 and backbone nitrogen of I264 are involved in strong hydrogen bonding interactions with CspA residues L49 and Y47, respectively. In addition, the hydrophobic ring of Y47 is making Van der Waals interactions with sidechain of I264, which further stabilizes the complex. L262 and I264 are from the loop between two β-strands in hFH CCP5. Y47 and L49 are also from the loop region between helix A and helix B in CspA. Interactions between these pairs of residues greatly stabilizes loop regions for both proteins, contributing to overall greater stability at the interface. T273 and Q75 also make stable hydrogen bonding with His99 and Q107 of CspA with occurrences of ~ 80% for both interactions.

Interestingly, the backbone oxygen of C276 is also involved in hydrogen bond interaction with Q107 from CspA. C276 forms a disulfide linkage with C302 within hFH CCP5 and is crucial in

92

maintaining the structural integrity and activity of CCP5. So, not only does C276 stabilizes hFH

CCP5, it also plays important role in stabilizing the protein complex with CspA.
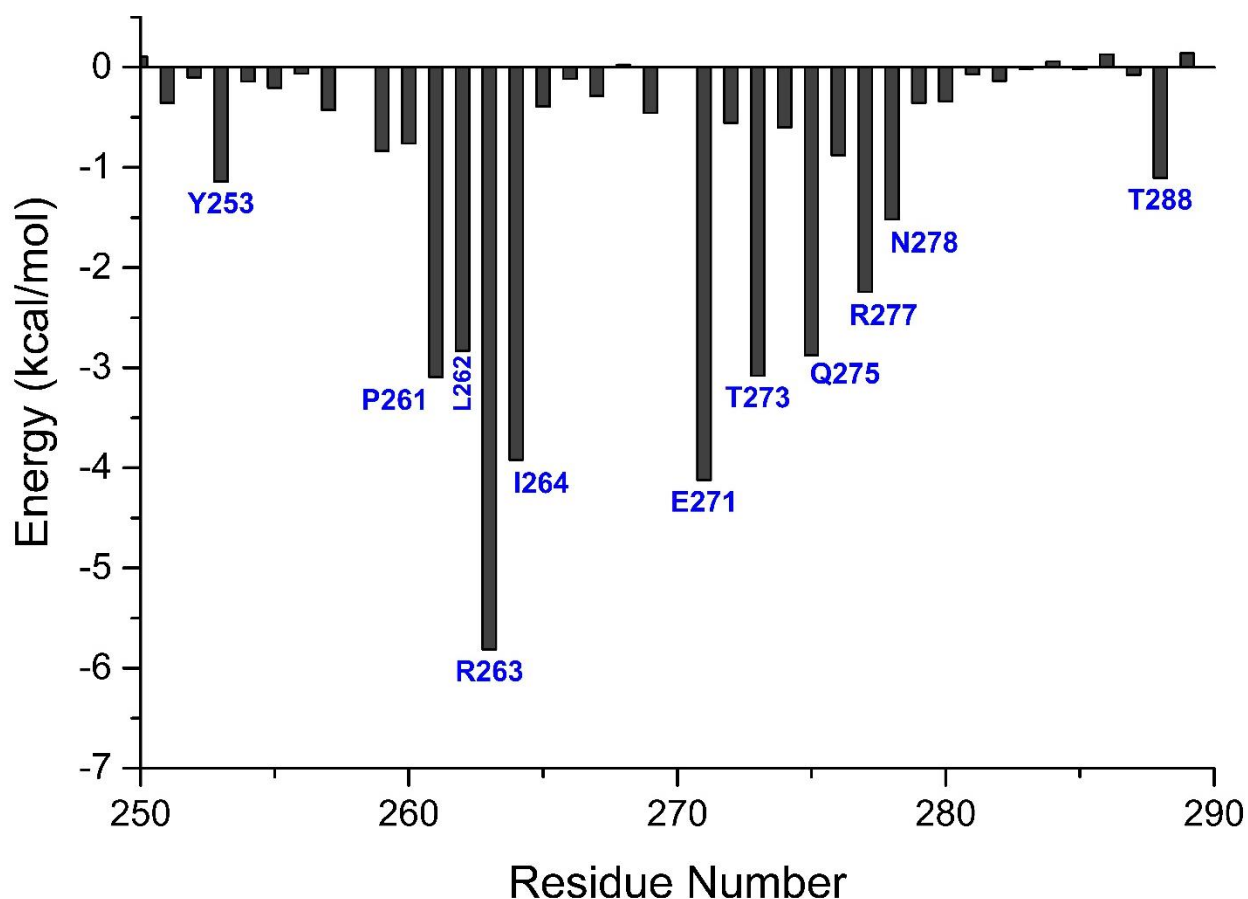
Figure 3.10. Binding energy decomposition of hFH CCP5 interface residues. Residues are labelled at the end of the plot.

To further validate results from computational analysis, alanine scanning mutagenesis

and thermodynamics analysis using ITC were carried out (Table 3.7). Experimental results show

that the R263A mutation decreases the binding affinity by more than 500-fold, with loss of

enthalpy and increase in entropy. R263 is one of the three residues at the center of the long loop

connecting two β-strands in hFH CCP5 and greatly stabilizes the complex via hydrogen bond and

salt bridge interactions with R95 of CspA from the loop region. R263 is located in between two

93

important non-polar residues, L262 and I264, both of which are making backbone hydrogen bond interactions with CspA residues at the interface. In addition, I264 is further stabilizing the complex by making non-polar interaction with Y98 from CspA. When R263 is mutated, not only the key interaction that is stabilizing the interface is broken, this also destabilizes the loop region that is making key interactions with CspA. As a result, there is a high possibility that the interactions of L262 and I264 with CspA residues at the loop region are also weakened significantly which can decrease the overall binding affinity and stability of the complex.

Table 3.7. Thermodynamics binding parameters for hFH CCP5 interface residues

hFH-5 Mutants	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	T ΔS (kcal/mol)
WILD	1.30 ± 0.02	-12.21 ± 0.02	-21.72 ± 1.02	-9.5 ± 0.16
D258A	1.57 ± 0.30	-12.80 ± 0.65	-30.80 ± 2.85	-18.00 ± 2.48
S260A	2.76 ± 0.10	-11.77 ± 0.01	-33.33 ± 0.47	-21.6 ± 0.48
L262A	9.35 ± 0.84	-10.96 ± 0.05	-33.28 ± 2.07	-22.31 ± 3.02
R263A	670 ± 28.35	-8.43 ± 0.03	-15.64 ± 4.03	-7.19 ± 4.00
R267A	1.35 ± 0.17	-12.12 ± 0.08	-32.96 ± 3.76	-20.79 ± 3.74
D270A	3.62 ± 0.03	-11.50 ± 0.01	-31.40 ± 4.30	-19.90 ± 4.32
E271A	16.88 ± 0.57	-10.62 ± 0.15	-38.77 ± 1.92	-28.14 ± 1.96
T273A	10.40 ± 0.32	-10.08 ± 0.18	-31.25 ± 3.79	-21.17 ± 3.94
Y274A	9.31 ± 0.96	-10.96 ± 0.18	-22.58 ± 2.43	-11.61 ± 2.29
Q275A	2.15 ± 0.10	-11.73 ± 0.09	-46.33 ± 3.70	-34.59 ± 4.69
R277A	2.52 ± 0.25	-11.51± 0.29	-31.20 ± 1.61	-19.69 ± 1.51
N278A	1.11 ± 0.15	-12.24 ± 0.04	-41.08 ± 0.76	-28.83 ± 0.71
N287A	2.91 ± 0.54	-11.66 ± 0.13	-66.10 ± 3.23	-51.43 ± 3.09
T288A	6.04 ± 1.80	-11.22 ± 0.28	-25.12 ± 2.41	-13.90 ± 2.10

3.3.7. Insights into the Host Specificity of *B. burgdorferi*

Host specificity is the range of organisms that a pathogen can colonize or infect. Some pathogens have a wide host range and can infect humans and many animals, whereas others may have a very strict host range and infect only one or a few related organisms. Host specificity is an important issue in infectious diseases, because it affects not only the epidemiology of the pathogen but also development of animal models and vaccines. Elucidation of the molecular basis of host specificity not only will greatly enhance our knowledge of pathogenesis and epidemiology of bacterial infection but also will be invaluable for development of strategies to improve animal models for studying their pathogenesis and simulating human diseases and development of therapeutics and vaccines.

One of the major factors determining host specificity is whether the pathogen can successfully escape the attack from the host immune system. Many human pathogens recruit complement factor H (FH) to escape human complement attack, a critical response from the innate immune system of the host. Host specificity of FH recruitment is correlated with host specificity of bacterial infection. Recent studies have revealed that human pathogens, *S. pneumoniae* [61], *N. meningitidis* [62], *N. gonorrhoeae* [13] and nontypeable *H. influenzae* (NTHi) [63], recruit FH in a human-specific manner, indicating FH recruitment as a contributing factor to the host specificity of these pathogens. Conversely, the broad host range of the bacterial pathogen *Borrelia burgdorferi*, is attributed partly to its ability to express many FH binding proteins, CspA being one of those major ones. Since there was no reliable information on the binding of CspA to FH until now, there has not been any major progress in understanding

95

the molecular level correlation between differential FH binding ability of CspA among various animals and potential role of such binding in host specificity of *Borrelia*.

Table 3.8. Resistance of *Borrelia* genospecies to sera from human and other animals [64]

	B. burgdorferi s.s.	*B. afzelii*	*B. garinii*	*B. bavariensis*
Human	+/–	+	–	+
Rodents	+/–	+	–	+
Birds	+/–	–	+	–
Dog	+	+	+/–	+
Wolf	+	+/–	+/–	+
Cat	+	+	+	+
Horse	+/–	–	–	–
Deer/cattle	–	–	–	–

Abbreviations: +, serum resistant; +/–, moderately serum resistant; –, serum sensitive; ND, not determined

One of the most interesting and complicated features of *Borrelia* is its ability to survive and cause persistent infection in wide range of hosts. Many studies have attempted to analyze the host-specificity of *Borrelia* based on the sensitivity of borrelia genospecies to serum from human and other animals (Table 3.8) [37, 41]. However, existence of many genospecies of *Borrelia* and its ability to express at least five different FH binding proteins on its surface make it complicated to get molecular level insights into host-specificity. Nevertheless, based on serum sensitivity studies on many *Borrelia* genospecies, large animal such as cattle and deer are incompetent dead-end hosts for *Borrelia*, where the borrelial spirochete is most likely killed by complement-mediated immune attack from these host [65]. Small animals such as dog, cat, mouse and sheep are considered competent hosts for *Borrelia* as most of the genospecies are resistant to complement mediated killing by the serum from these mammals [38]. Serum resistivity and sensitivity not only varies among different animals but also varies among different genospecies of *Borrelia* within the same animal, complicating the matter further. In Europe, *B. garinii* uses birds as the major competent host for the transmission of spirochete whereas *B.*

afzelii and *B. bavariensis* are predominantly transmitted by rodents. *B. burgdorferi* is carried and transmitted by a wide range of large animals and birds including human and other mammals. The ability of *Borrelia* to interact with and 'hijack' FH protein from the host is directly correlated to the serum sensitivity. Understanding the underlying basis of such species-specific adaptation among various genospecies of borrelia may reveal specific adaptations that borrelia uses to avoid complement mediated killing in different hosts, leading to potentially new therapeutics discovery.

Based on our study in this chapter, we have demonstrated that CspA of *B. burgdorferi* binds very tightly to hFH CCP5 and have successfully identified key residues important for binding in CspA as well as hFH CCP5. Sequence alignment of hFH CCP5 with FH CCP5 from many mammalian species closely related to human, such as monkey and chimpanzee shows that key hFH CCP5 residues that are important for binding of CspA are absolutely conserved (Fig. 3.11). Small animals like rodent, mouse, mole and rabbit show at least partial conservation of the key residues.
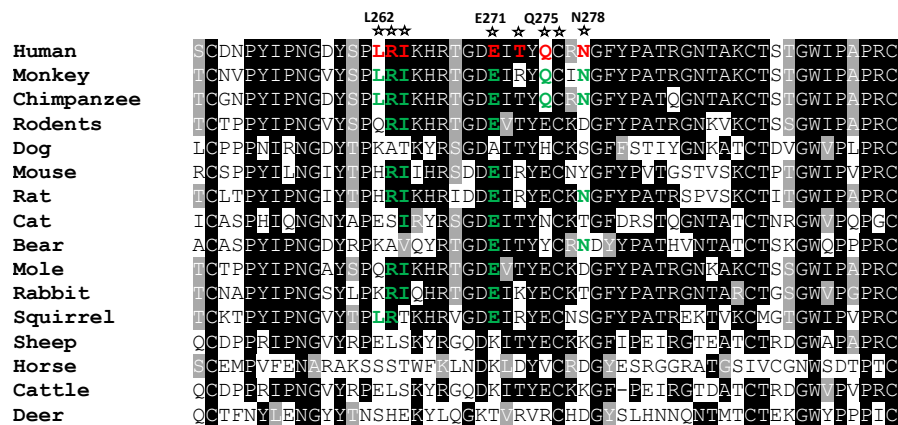
Figure 3.11. FH CCP5 sequence alignment of various animal. Important hFH CCP5 residues contributing in binding of CspA are colored in red. Subsequent conserved residue in other mammals are colored in green. Monkey and chimpanzee show the best alignment among all the mammals. Whereas alignment score for sheep, cattle and deer is the lowest.

97

Interestingly, FH CCP5 sequences from species that are determined to be incompetent hosts for *Borrelia*, such as, horse, cattle and deer, show the largest variation in FH CCP5 sequence compare to the hFH CCP5. The key residues from hFH CCP5 that we characterized as the most important residues for CspA binding are not conserved in sheep, horse, cattle and deer. Based on this study, most likely the incompetent nature of animals like sheep, cattle and deer is the outcome of the inability of *Borrelia* to bind and recruit FH CCP5 via CspA-mediated interactions as FH CCP5 of these animals lack key residues that are required for the interaction. Although, dogs are considered a dead-end host for the Lyme disease, studies show that most of the genospecies are resistant to serum from dog, making dogs susceptible to Lyme disease. However, sequence analysis of dog and human FH showed that none of the key residues from hFH CCP5 are conserved in dog. Whether *Borrelia* spirochete uses different mechanisms or different surface proteins to gain resistance against dog serum is yet to be investigated.

In order to gain better insights into the binding of FH from other animals, we briefly studied the binding of mouse FH (mFH) CCP5 to CspA. The binding affinity of mFH CCP5 to CspA is ~ 2500-fold lower, in comparison with binding of CspA with hFH CCP (Table 3.9). The significant decrease in binding energy is due to a dramatic loss of enthalpy.

Table 3.9. Thermodynamic parameters of binding between hFH and mFH CCP5 with CspA

	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)
hFH CCP5	1.30 ± 0.02	-12.21 ± 0.02	-21.71 ± 1.02	-9.50 ± 0.16
mFH CCP 5	3300 ± 244	-8.80 ± 0.88	-0.84 ± 0.08	7.94 ± 0.80

Hydrogen bond analysis of the CspA:mFH CCP5 complex during the simulation revealed key differences in the interactions of mFH and hFH CCP5 with CspA. Unlike hFH CCP5, there are

only two hydrogen bonding interactions with occupancy of more than 90% in mFH CCP5 and all

other remaining hydrogen bonding interactions occupy close to or below 60%. So, there are only

two key hydrogen bonds stabilizing the interface with all other weak hydrogen bonds dissociating

at certain point during the simulation (Table 3.10). There are only four residues from CspA that

are involved in stabilizing mFH CCP5 complex, which further indicates a weak association at the

interface.

Although our study lays down the foundation to better understand the host specificity of

Borrelia, further extensive studies are required to fully understand the role of FH binding ability

of different genospecies of *Borrelia* and their correlation to the host specificity.

Table 3.10. Intermolecular hydrogen bonds observed during the MD simulation of the complex
of mFH CCP5 and CspA

mFH5...........CspA	Occurrence (%)	Distance (Å)	Modelled Structure (Å)
H262-O………N-L48	95.64	2.93±0.18	2.91
I264-N……….O-Y46	95.16	1.99±0.20	3.04
Q271-OE2… NH2-R94	65.30	3.24±0.69	2.85
Q271-OE1… NH2-R94	61.23	3.260.69 ±	2.83
Q271-OE2… NE-R94	57.35	3.41±0.87	2.72
R263-NE……OD2-D47	56.97	4.01 ±1.53	2.92
Q271-OE1…NE-R94	51.66	3.47 ±0.86	2.74

3.4. Discussion

In this study, using the crystal structure of CspA:hFH CCP5, computational analysis and ITC experiments, we have demonstrated that a single hFH module (CCP5) alone is sufficient for tight binding of the key borrelial surface protein, CspA. All studies published so far have reported the dimeric cleft region between two CspA monomers as the FH binding site [58, 60, 66]. Also, many studies suggested CCP6 or CCP5-7 as the predominant CspA binding region in FH [56, 57]. However, our results invalidate all those published results and hence have given a new direction in understanding the molecular basis of FH recruitment by *B. burgdorferi*.

Further, we have identified hot spot residues at the CspA:hFH CCP5 interface. Out of more than 20 interface residues on each protein, we have identified 7 residues in CspA and 8 residues in hFH CCP5 that are important for formation of the complex. Among these residues, Trp102 in CspA and Arg263 in hFH CCP5 are identified as the most important residues for CspA and hFH CCP5 binding. Studies have shown that irrespective of the borrelial genospecies, the complement-mediated killing of borrelial spirochetes is higher from serum of cattle and deer [67]. In other animals and humans, the complement mediated killing is intermediate and *Borrelia* species dependent [67]. Sequence analysis of FH CCP5 from various animals revealed that the key hFH CCP5 residues identified as the most important for CspA binding are completely variant in cattle and deer whereas some or a majority of these key residues are conserved among other small animals that are competent host for *Borrelia*. Based on these results and analyses, it is further validated that the complement-mediated killing of a borrelial spirochete is inversely correlated to its ability to bind and recruit FH protein on their surface. From these preliminary results, cattle and deer lack all the residues that are important for binding to CspA. So, when

borrelial species are exposed to cattle or deer, the interaction between CspA and FH is inefficient. As a result, *Borrelia* is not able to circumvent complement-mediated killing by the immune system of these animals and hence they are effectively killed and cleared. While more studies need to be done to further support these results, our preliminary data show some promising direction in understanding the pathogenesis and host specificity of *Borrelia*. Similar characterization and analysis from all other FH binding surface proteins from *Borrelia* can shed more light on this issue.

Understanding the molecular basis of host specificity would greatly enhance the progress in Lyme disease vaccine design. Although a borrelial outer surface protein A (OspA) based vaccine showed promising results against Lyme disease in late 90's, several factors led to its failure after only 4 years. Experiments on mouse and other animals suggest that there is insufficient essential antibody response generated during natural infection that is insufficient for long term protection against the disease [68]. Further complicating this matter, *Borrelia* spirochetes alter antigen expression during different stages of infection that enables them to successfully evade the antibody-mediated immune response from the host [69, 70]. Therefore, successful identification and characterization of suitable antigens in *Borrelia* is the biggest challenge to advance vaccine development against Lyme disease.

Borrelial FH binding proteins such as CspA and CspZ can be potential targets for therapeutics against Lyme disease. We have successfully identified unique host-pathogen interaction that can potentially be valuable information in advancing our knowledge and understanding of host-specificity and pathogenesis of Lyme disease.

REFERENCES

REFERENCES

1.   Ricklin, D., et al., *Complement: a key system for immune surveillance and homeostasis.* Nature Immunology, 2010. **11**(9): p. 785-797.

2.   Serruto, D., et al., *Molecular mechanisms of complement evasion: learning from staphylococci and meningococci.* Nature Reviews Microbiology, 2010. **8**(6): p. 393-399.

3.   Harboe, M. and T.E. Mollnes, *The alternative complement pathway revisited.* Journal of Cellular and Molecular Medicine, 2008. **12**(4): p. 1074-1084.

4.   Makou, E., A.P. Herbert, and P.N. Barlow, *Functional Anatomy of Complement Factor H.* Biochemistry, 2013. **52**(23): p. 3949-3962.

5.   Ferreira, V.P., M.K. Pangburn, and C. Cortes, *Complement control protein factor H: The good, the bad, and the inadequate.* Molecular Immunology, 2010. **47**(13): p. 2187-2197.

6.   Parente, R., et al., *Complement factor H in host defense and immune evasion.* Cellular and Molecular Life Sciences, 2017. **74**(9): p. 1605-1624.

7.   Kraiczy, P. and B. Stevenson, *Complement regulator-acquiring surface proteins of Borrelia burgdorferi: Structure, function and regulation of gene expression.* Ticks and Tick-Borne Diseases, 2013. **4**(1-2): p. 26-34.

8.   Ram, S., et al., *Binding of complement factor H to loop 5 of porin protein 1A: A molecular mechanism of serum resistance of nonsialylated Neisseria gonorrhoeae.* Journal of Experimental Medicine, 1998. **188**(4): p. 671-680.

9.   Ngampasutadol, J., et al., *Human factor H interacts selectively with Neisseria gonorrhoeae and results in species-specific complement evasion.* Journal of Immunology, 2008. **180**(5): p. 3426-3435.

10.   Shaughnessy, J., et al., *Functional Comparison of the Binding of Factor H Short Consensus Repeat 6 (SCR 6) to Factor H Binding Protein from Neisseria meningitidis and the Binding of Factor H SCR 18 to 20 to Neisseria gonorrhoeae Porin.* Infection and Immunity, 2009. **77**(5): p. 2094-2103.

11.   Shaughnessy, J., et al., *Molecular Characterization of the Interaction between Sialylated Neisseria gonorrhoeae and Factor H.* Journal of Biological Chemistry, 2011. **286**(25): p. 22235-22242.

12.   Jongerius, I., et al., *Distinct Binding and Immunogenic Properties of the Gonococcal Homologue of Meningococcal Factor H Binding Protein.* Plos Pathogens, 2013. **9**(8): p. e1003528.

13.   Chen, A. and H.S. Seifert, *Structure-Function Studies of the Neisseria gonorrhoeae Major Outer Membrane Porin.* Infection and Immunity, 2013. **81**(12): p. 4383-4391.

14.   Zhang, B., et al., *Serum resistance in Haemophilus parasuis SC096 strain requires outer membrane protein P2 expression.* FEMS Microbiology Letters, 2012. **326**(2): p. 109-115.

15.   Dave, S., et al., *PspC, a pneumococcal surface protein, binds human factor H.* Infection and Immunity, 2001. **69**(5): p. 3435-3437.

16.   Dave, S., et al., *Dual roles of PspC, a surface protein of Streptococcus pneumoniae, in binding human secretory IgA and factor H.* Journal of Immunology, 2004. **173**(1): p. 471-477.

17.   Hammerschmidt, S., et al., *The host immune regulator factor H interacts via two contact sites with the PspC protein of Streptococcus pneamoniae and mediates adhesion to host epithelial cells.* Journal of Immunology, 2007. **178**(9): p. 5848-5858.

18.   Janulczyk, R., et al., *Hic, a novel surface protein of Streptococcus pneumoniae that interferes with complement function.* Journal of Biological Chemistry, 2000. **275**(47): p. 37257-37263.

19.   Neeleman, C., et al., *Resistance to both complement activation and phagocytosis in type 3 pneumococci is mediated by the binding of complement regulatory protein factor H.* Infection and Immunity, 1999. **67**(9): p. 4517-4524.

20.   Agarwal, V., et al., *Complement regulator Factor H mediates a two-step uptake of Streptococcus pneumoniae by human cells.* Journal of Biological Chemistry, 2010. **285**(30): p. 23486-95.

21.   Mohan, S., et al., *Tuf of Streptococcus pneumoniae is a surface displayed human complement regulator binding protein.* Molecular Immunology, 2014. **62**(1): p. 249-264.

22.   Kohler, S., et al., *Binding of vitronectin and Factor H to Hic contributes to immune evasion of Streptococcus pneumoniae serotype 3.* Thrombosis and Haemostasis, 2015. **113**(1): p. 125-142.

23.   Achila, D., et al., *Structural determinants of host specificity of complement Factor H recruitment by Streptococcus pneumoniae.* Biochemical journal, 2015. **465**(2): p. 325-35.

24.   Lu, L., Y.Y. Ma, and J.R. Zhang, *Streptococcus pneumoniae recruits complement factor H through the amino terminus of CbpA.* Journal of Biological Chemistry, 2006. **281**(22): p. 15464-15474.

25.   Haupt, K., et al., *The Staphylococcus aureus Protein Sbi Acts as a Complement Inhibitor and Forms a Tripartite Complex with Host Complement Factor H and C3b.* Plos Pathogens, 2008. **4**(12).

26.   Sharp, J.A. and K.M. Cunnion, *Disruption of the alternative pathway convertase occurs at the staphylococcal surface via the acquisition of factor H by Staphylococcus aureus.* Molecular Immunology, 2011. **48**(4): p. 683-690.

27.   Riley, S.P., J.L. Patterson, and J.J. Martinez, *The Rickettsial OmpB beta-Peptide of Rickettsia conorii Is Sufficient To Facilitate Factor H-Mediated Serum Resistance.* Infection and Immunity, 2012. **80**(8): p. 2735-2743.

28.   Ho, D.K., et al., *The Yersinia pseudotuberculosis Outer Membrane Protein Ail Recruits the Human Complement Regulatory Protein Factor H.* Journal of Immunology, 2012. **189**(7): p. 3593-3599.

29.   Ho, D.K., H. Jarva, and S. Meri, *Human Complement Factor H Binds to Outer Membrane Protein Rck of Salmonella.* Journal of Immunology, 2010. **185**(3): p. 1763-1769.

30.   McNeil, L.K., R.J. Zagursky, and S.L. Lin, *Role of factor H binding protein in Neisseria meningitidis virulence and its potential as a vaccine candidate to broadly protect against meningococcal disease.* Microbiol Mol Biol Rev, 2013. **77**: p. 234-252.

31.   Behnsen, J., et al., *The opportunistic human pathogenic fungus Aspergillus fumigatus evades the host complement system.* Infection and Immunity, 2008. **76**(2): p. 820-827.

32.   Luo, S.S., et al., *Immune evasion of the human pathogenic yeast Candida albicans: Pra1 is a Factor H, FHL-1 and plasminogen binding surface protein.* Molecular Immunology, 2009. **47**(2-3): p. 541-550.

33.   Luo, S., et al., *Sequence Variations and Protein Expression Levels of the Two Immune Evasion Proteins Gpm1 and Pra1 Influence Virulence of Clinical Candida albicans Isolates.* PloS one, 2015. **10**(2): p. e0113192.

34.   Diaz, A., A. Ferreira, and R.B. Sim, *Complement evasion by Echinococcus granulosus - Sequestration of host factor H in the hydatid cyst wall.* Journal of Immunology, 1997. **158**(8): p. 3779-3786.

35.   Haapasalo, K., T. Meri, and T.S. Jokiranta, *Loa loa Microfilariae Evade Complement Attack In Vivo by Acquiring Regulatory Proteins from Host Plasma.* Infection and Immunity, 2009. **77**(9): p. 3886-3893.

36.   Meri, T., et al., *Onchocerca volvulus microfilariae avoid complement attack by direct binding of factor H.* Journal of Infectious Diseases, 2002. **185**(12): p. 1786-1793.

37.   Baeumler, A. and F.C. Fang, *Host Specificity of Bacterial Pathogens.* Cold Spring Harbor Perspectives in Medicine, 2013. **3**(12).

38.   Pan, X., Y. Yang, and J.-R. Zhang, *Molecular basis of host specificity in human pathogenic bacteria.* Emerg Microbes Infect, 2014. **3**: p. e23.

39.   Langereis, J.D., M.I. de Jonge, and J.N. Weiser, *Binding of human factor H to outer membrane protein P5 of non-typeable Haemophilus influenzae contributes to complement resistance.* Molecular Microbiology, 2014. **94**(1): p. 89-106.

40.   Granoff, D.M., J.A. Welsch, and S. Ram, *Binding of Complement Factor H (fH) to Neisseria meningitidis Is Specific for Human fH and Inhibits Complement Activation by Rat and Rabbit Sera.* Infection and Immunity, 2009. **77**(2): p. 764-769.

41.   Kurtenbach, K., et al., *Host association of Borrelia burgdorferi sensu lato - the key role of host complement.* Trends in Microbiology, 2002. **10**(2): p. 74-79.

42.   Radolf, J.D., et al., *Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes.* Nature Reviews Microbiology, 2012. **10**(2): p. 87-99.

43.   Mannelli, A., et al., *Ecology of Borrelia burgdorferi sensu lato in Europe: transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change.* Fems Microbiology Reviews, 2012. **36**(4): p. 837-861.

44.   Velazquez-Campoy, A. and E. Freire, *Isothermal titration calorimetry to determine association constants for high-affinity ligands.* Nat Protoc, 2006. **1**(1): p. 186-91.

45.   D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, et al., *AMBER 2016*. 2016, University of California, San Francisco.

46.   Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.

47.   Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.

48.   Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

49.   Gordon, J.C., et al., *H++: a server for estimating pKas and adding missing hydrogens to macromolecules.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W368-71.

50.   Coleman, T.G., H.C. Mesick, and R.L. Darby, *Numerical integration: a method for improving solution stability in models of the circulation.* Ann Biomed Eng, 1977. **5**(4): p. 322-8.

51.   Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

52.   Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**.

53.   Roe, D.R. and T.E. Cheatham, 3rd, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013. **9**(7): p. 3084-95.

54.   *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*

55.   Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* J Chem Inf Model, 2011. **51**(1): p. 69-82.

56.   Kraiczy, P., et al., *Complement resistance of Borrelia burgdorferi correlates with the expression of BbCRASP-1, a novel linear plasmid-encoded surface protein that interacts with human factor H and FHL-1 and is unrelated to Erp proteins.* J Biol Chem, 2004. **279**(4): p. 2421-9.

57.   Herzberger, P., et al., *Identification and characterization of the factor H and FHL-1 binding complement regulator-acquiring surface protein 1 of the Lyme disease spirochete Borrelia spielmanii sp. nov.* Int J Med Microbiol, 2009. **299**(2): p. 141-54.

58.   Cordes, F.S., et al., *A novel fold for the factor H-binding protein BbCRASP-1 of Borrelia burgdorferi.* Nat Struct Mol Biol, 2005. **12**(3): p. 276-7.

59.   Cordes, F.S., et al., *Structure-function mapping of BbCRASP-1, the key complement factor H and FHL-1 binding protein of Borrelia burgdorferi.* Int J Med Microbiol, 2006. **296 Suppl 40**: p. 177-84.

60.   Kraiczy, P., et al., *Mutational analyses of the BbCRASP-1 protein of Borrelia burgdorferi identify residues relevant for the architecture and binding of host complement regulators FHL-1 and factor H.* Int J Med Microbiol, 2009. **299**(4): p. 255-68.

61.   Achila, D., et al., *Structural determinants of host specificity of complement Factor H recruitment by Streptococcus pneumoniae.* Biochem J, 2015. **465**(2): p. 325-35.

62. Serruto, D., et al., *Molecular mechanisms of complement evasion: learning from staphylococci and meningococci.* Nat Rev Microbiol, 2010. **8**(6): p. 393-9.

63. Fleury, C., et al., *Identification of a Haemophilus influenzae factor H-Binding lipoprotein involved in serum resistance.* J Immunol, 2014. **192**(12): p. 5913-23.

64. de Taeye, S.W., et al., *Complement evasion by Borrelia burgdorferi: it takes three to tango.* Trends Parasitol, 2013. **29**(3): p. 119-28.

65. Kurtenbach, K., et al., *Host association of Borrelia burgdorferi sensu lato--the key role of host complement.* Trends Microbiol, 2002. **10**(2): p. 74-9.

66. Kraiczy, P., et al., *Further characterization of complement regulator-acquiring surface proteins of Borrelia burgdorferi.* Infect Immun, 2001. **69**(12): p. 7800-9.

67. Bhide, M.R., et al., *Sensitivity of Borrelia genospecies to serum complement from different animals and human: a host-pathogen relationship.* FEMS Immunol Med Microbiol, 2005. **43**(2): p. 165-72.

68. Felsenfeld, O. and R.H. Wolf, *Reinfection of vervet monkeys (Cercopithecus aethiops) with Borrelia hermsii.* Res Commun Chem Pathol Pharmacol, 1975. **11**(1): p. 147-50.

69. Nowakowski, J., et al., *Culture-confirmed infection and reinfection with Borrelia burgdorferi.* Ann Intern Med, 1997. **127**(2): p. 130-2.

70. Golde, W.T., et al., *Culture-confirmed reinfection of a person with different strains of Borrelia burgdorferi sensu stricto.* J Clin Microbiol, 1998. **36**(4): p. 1015-9.

CHAPTER 4

Molecular Analysis of FH Recruitment by the Outer Surface Protein CspZ of Borrelia burgdorferi

4.0. Abstract

The complement system is the first line of host defense and an intricate system for immune surveillance against microbial intruders. To cause human disease with persistent infection, pathogen must survive attack from the host complement system. A key complement regulator, complement factor H (FH), is a 155-kDa protein containing 20 domains, commonly referred as complement control protein (CCP) modules. FH inhibits both complement activation and amplification. Since one of the key complement activation pathways, alternative pathway, is always active at low level, the host cells avoid autoimmunity by covering themselves with FH molecules. Taking advantage of this important function of FH, a wide range of pathogens have evolved multiple mechanisms to recruit FH on their own cell surface, effectively evading the host complement attack. Lyme disease is the most common vector borne illness in United States and Europe. The causative pathogen of Lyme disease, *Borrelia burgdorferi,* can survive and infect wide range of hosts, including human, animals, birds and reptiles. The host-association of individual borrelial spirochetes is corelated with their ability to escape host complement attack. One key mechanism *Borrelia* employs to circumvent host complement attack is by recruiting FH protein on their surface using their surface proteins. However, how these borrelial proteins interact with FH and correlation of such interactions to host specificity are largely unknown. In this chapter, we present crystal structure and the extensive biophysical and computational characterization of the protein complex between another key borrelial surface protein, CspZ, and hFH CCP7. In addition, we have studied binding of CspZ to mouse FH (mFH) CCP7 by isothermal titration calorimetry (ITC). Together, we have provided new insights to the specificity of FH recruitment by borrelial surface protein CspZ.

4.1. Introduction

The complement system is the first line of defense against foreign pathogen and plays a crucial role in proper regulation of the innate immune system [1, 2]. However, excessive or inappropriate activation of complement system can lead to severe autoimmunity. To avoid complement attack on self-cells, many immune regulator proteins play important role in regulating complement activation [3]. Complement factor H (FH) protein is one of the key regulators of complement activation and is a major player in protecting self-cells from complement mediated attack by inhibiting and downregulating pathways of complement activation [3, 4]. FH is a 155-kDa protein with 20 homologous domains, commonly termed as complement control protein (CCP) modules [2, 3]. Taking advantage of this key function of FH, a wide range of pathogens have evolved to recruit FH to their own surface to circumvent the host complement attack. Many pathogens, including *S. pneumoniae*, *N. gonorrhoeae* and *N. meningitidis*, recruit FH in a human specific manner and such specificity in FH recruitment is correlated with the host specificity of these pathogens [5-7]. In contrast, the main causative agent of Lyme disease, *B. burgdorferi*, can survive and infect a wide range of hosts. The broad host range of *B. burgdorferi* is attributed to its ability to recruit FH from a variety of hosts by expressing multiple FH binding proteins [8-10].

Lyme disease is the most common vector borne illness in the United States and Europe [11] and is caused by different genospecies of *Borrelial burgdorferi* sensu lato family, including *B. burgdorferi* sensu stricto (*B. burgdorferi*), *B. afzelii*, *B. garinii*, *B. spielmanii* and *B. bavariensis* [10, 11]. As mentioned before, one of the key features of *Borrelia* is its ability to infect a wide range of hosts, including human and other large animals, reptiles and birds [10]. The competency of

110

borrelial host and the successful host association is correlated with the ability of the pathogen to avoid complement mediated attack from the host [8]. The major mechanism employed by *Borrelia* to survive complement attack from the host is by recruiting FH protein on their surface via protein-protein interaction using their surface proteins [12]. So far, five different FH binding surface proteins, CspA, CspZ, ErpA, ErpC and ErpP, are known in *Borrelia*.

CspZ is a surface exposed 23.2-kDa lipoprotein that *Borrelia* produces mainly within two weeks of infection [12]. Extensive studies have shown that CspZ can provide resistance against host complement mediated attack by recruiting FH from the host [12]. Although molecular details of interaction between CspZ and FH proteins have emerged recently [13, 14], most of the studies are done without proper knowledge of the binding interface in CspZ. When the crystal structure of CspZ was published, many studies have attempted to map the binding site for its interaction with hFH [14-16]. Also, hFH CCP7 was mapped previously as the predominant binding region for CspZ [15, 16]. However, the crystal structures of the individual free form of CspZ and hFH CCP7 are inadequate for identifying exactly which residues are involved in binding of FH. In our lab, we have determined the structure of *B. burgdorferi* strain B31 CspZ in complex with human FH (hFH) CCP7 by X-ray crystallography. In addition, we have measured the thermodynamic parameters for binding of CspZ to hFH and mouse FH (mFH) CCP7s by isothermal titration calorimetry (ITC). In conjunction of extensive computational analyses, we provide new insights to the specificity of FH recruitment by CspZ.

4.2. Material and Methods

4.2.1. CspZ Production and Purification

The DNA encoding the matured CspZ, residues N25-L236, was amplified by PCR from the genomic DNA of *Borrelia burgdorferi* strain B31 (a gift from Dr. Fang-Ting Liang). The forward and reverse primers used for the PCR reaction were 5'- G GAA TTC CAT ATG AAT CAG AGA AAT ATT AAT GAG CTT-3' and 5'- G GGA TCC CTA TAA TAA AGT TTG CTT AAT AGC-3', respectively. The amplified DNA fragment was cloned into the lab-made expression vector pET17bHR by digestion with the restriction enzymes BamHI and NdeI and ligation. The expression vector was derived from the Novagen vector pET17b with the addition of an N-terminal His-tag and a TEV protease cleavage site. The correct coding sequence was verified by DNA sequencing.

The expression construct was transformed into the *E. coli* strain BL21(DE3). A liter of LB media with 100 μg/mL ampicillin was inoculated with a single colony of transformants and shacked at 225 rpm and 37 °C overnight. The cells were harvested by centrifugation at 4 °C. The cell pallet was re-suspended in buffer A (20 mM Tris and 150 mM NaCl, pH 7.5) supplemented with proper amount of DNase I and $MgCl_2$. Then, the cells were lysed by French Pressure and the supernatant of the lysate from centrifugation was loaded onto the Ni-NTA column pre-equilibrated with buffer A. The column was washed with buffer A until the reading of OD_{280} was less than 0.05 and eluted with a linear gradient of 0-250 mM imidazole in buffer A with over 10 times of the column volume. Fractions containing His-tagged CspZ, as monitored by SDS-PAGE, were pooled and concentrated to ~10 mL. Proper amount of TEV protease was added into the protein solution with 1.0 mM DTT and 0.5 mM EDTA, and the solution was incubated for 2 h at room temperature and overnight at 4 °C. The completeness of the His-tag cleavage was

112

confirmed by SDS-PAGE. The protein solution was then dialyzed thoroughly against buffer A to get rid of EDTA and DTT. After centrifugation, the supernatant was loaded onto the re-equilibrated Ni-NTA agarose column, and the CspZ protein without His-tag was eluted with buffer A. CspZ-containing fractions were pooled and concentrated to ~15 mL. After centrifugation, the protein solution was loaded onto a Sephadex G-75 gel filtration column pre-equilibrated with buffer A. The column was developed with the same buffer. The fractions containing pure CspZ as monitored by SDS-PAGE were pooled and concentrated to ~15 mL and stored at 4 °C with addition of 150 mM NaN_3 for later use. The protein yield was ~185 mg/per liter LB culture.

4.2.2. hFH CCP7 Production and Purification

The DNA encoding hFH CCP7 (K370-K428) was amplified by PCR from a codon-optimized synthetic hFH gene [17]. The forward and reverse primers used for the PCR reaction were 5'- G GAA TTC CAT ATG AAA TGT TAC TTT CCG TAT CTG GAA AAC-3' and 5'- G GGA TCC TCA TTT GAC ACG AAT GCA GCG-3', respectively. The PCR reaction, cloning, protein production and purification were performed as previously described in Chapter 3 [17].

4.2.3. Co-crystallization of CspZ and hFH CCP7

The complex of CspZ and hFH CCP7 were made by mixing the two proteins with a molar ratio of 1 to 1.1 in buffer A. Protein concentrations were determined by measuring OD_{280} and using extinction coefficients calculated with the ExPASy web server [18]. The complex was purified by FPLC using a HiLoad 26/60 Superdex-75 gel filtration column. The chromatogram showed two peaks, a large peak containing both proteins and a small peak containing only CCP7, indicating that the two protein formed a 1:1 complex. Fractions containing the complex were

pooled and concentrated. After dialysis against 10 mM Tris buffer at pH 7.5, the protein solution was further concentrated to 25 mg/ml for crystallization trials. Crystallization screening was performed using the sitting-drop vapor-diffusion method on a Cryphon robot (Art Robbins Instrument) at room temperature with Hampton screening kits. The crystallization conditions were optimized manually by the hanging-drop vapor-diffusion method at 20 °C. The solution in the best condition contained 200 mM $MgCl_2$, 100 mM Tris-HCl (pH 7.8), and 22-24% (w/v) PEG 3350. Crystals appeared in a week and continued to grow for over a month. They were then flash-frozen in liquid nitrogen using 25% glycerol as the cryo-protestant.

4.2.4. X-ray Data Collection and Structure Determination

X-ray diffraction data were collected at the LS-CAT 21-ID-G beamline of Advanced Photon Source, Argonne National Laboratory. The data were indexed and scaled with HKL2000 [19] and XDS [20, 21]. The processed data were phased by molecular replacement using the program Phaser-MR [22] with the structures of the free CspZ (PDB entry 4CBE) [15] and the CCP7 module in hFH CCP 6-8 (PDB 2UWN) [23] as searching models. The structure was built with Coot 7.2 [24] and refined with PHENIX [25]. Figures of the crystal structure were prepared with PyMOL [26]. A summary of data collection and structure refinement statistics of CspZ:hFH CCP7 is given in Table 4.2.

4.2.5. Isothermal Titration Calorimetry (ITC)

ITC measurements were performed at 25 °C on a VP-ITC isothermal titration calorimeter (MicroCal) according to the protocol of Velazquez-Campoy and Freire [27]. All protein solutions were dialyzed against a buffer containing 50 mM HEPES and 50 mM KCl, pH7.5. The concentrations of the dialyzed protein solutions were determined by measuring OD_{280} and using

extinction coefficients calculated by the method of Gill and von Hippel [28] using the ProtParam tool of the ExPASy web server [29]. The sample cell was loaded with a FH CCP7 solution and the syringe with a *CspZ* solution. The binding enthalpies were obtained by injecting the CspZ solution into the sample cell under stirring conditions. A typical ITC experiment consisted of 28 injections with 10 µl each at 6 min intervals. The ITC data were analyzed using Origin 5.0.

4.2.6. Site-Directed Mutagenesis

Single amino acid substitution mutation was carried out using site directed mutagenesis approach as described by Quik-Change ™ protocol from Strategene. The mutagenic primers employed for this study are listed in Table 4.1. PCR reactions were carried out for 18 cycles (95 °C for 30s, 56 °C for 30 s and 72 °C for 8 minutes) using 40 ng expression vector pET17bHR, with each mutagenic primer to a final concentration of 10 µM, dNTPs to the final concentration of 10mM and 2.5U p*futurbo* DNA polymerase (Strategene) in a final reaction volume of 50 µl. The PCR product was digested with 10 U DpnI at 37 °C for an hour to eliminate the parent plasmid. The digested product was transformed in DH5α (*E. coli* strain) and plasmid DNA was extracted and purified following the PCR clean-up protocol from Promega.

Table 4.1. Primers for PCR cloning and mutagenesis of CspZ

Seq. Name	Seq 5' —————————————→ 3'
CspZD47Af	G TAT TAT TCT ATA AAA TTA G<u>CC</u> GCT ATT TAT AAC GAA TGT AC
CspZD47Ar	GTACATTCGTTATAAATAGCGGCTAATTTTATAGAATAATAC
CspZY50Af	CT ATA AAA TTA GAC GCT ATT <u>GC</u>T AAC GAA TGT ACA GGA GC
CspZY50Ar	GCTCCTGTACATTCGTTAGCAATAGCGTCTAATTTTATAG
CspZN51Af	GAC GCT ATT TAT <u>GCC</u> GAA TGT ACA GGA GC
CspZN51Ar	GCTCCTGTACATTCGGCATAAATAGCGTC
CspZT54Af	GAC GCT ATT TAT AAC GAA TGT <u>GCA</u> GGA GCA TAT AAT G
CspZT54Ar	CATTATATGCTCCTGCACATTCGTTATAAATAGCGTC
CspZN58Af	C GAA TGT ACA GGA GCA TAT <u>GC</u>T GAT ATT ATG ACT TAT TCG
CspZN58Ar	CGAATAAGTCATAATATCAGCATATGCTCCTGTACATTCG
CspZT62Af	GGA GCA TAT AAT GAT ATT ATG <u>GC</u>T TAT TCG GAA GGT AC
CspZT62Ar	GTACCTTCCGAATAAGCCATAATATCATTATATGCTCC
CspZR139Af	GCT GAT TCT TAT AAA AAA CTT <u>GC</u>A AAA TCT GTT GTA TTA GCC
CSPZR139Ar	GGC TAA TAC AAC AGA TTT TGC AAG TTT TTT ATA AGA ATC AGC
CspZN180Af	GAG TTT GTA GAG GAA **<u>GCT</u>** GAT CTT ATA GCT CTT GAG
CspZN180Ar	GAGCTATAAGATCAGCTTCCTCTACAAACTCTTTAGC
CspZY211Af	GA AGC AGG TAT AAT AAT TTT GCT AAA AAA GAA GCA G
CspZY211Ar	CTGCTTCTTTTTTAGCAAAATTATTATACCTGCTTC

4.2.7. Molecular Dynamics Simulation and MM-GBSA Analysis

To investigate the dynamics of the CspZ:hFH CCP7 complex, MD simulations for wild type and mutant protein complexes were carried out. All MD simulations were performed using AMBER 16 Molecular Dynamics package and the ff14SB force field was used to describe the proteins [30, 31]. The crystal structure of the complex of CspZ and hFH CCP7 solved in our lab

was used as the starting structure to generate coordinate files for subsequent MD simulations and analysis. The structure of the complex between hFH CCP7 and CspZ was used as template to obtain homology model of the structure of the complex between mFH CCP7 and CspZ using SWISS-MODEL server [32]. Using the structure of the bound hFH CCP7 as the template, a model of the bound mFH CCP7 was built using the SWISS-MODEL server [32, 33]. All mutant proteins for *in-silico* mutagenesis study were generated with assumption that the structural integrity of the proteins is not affected by the single amino acid change. Before running the simulations, all protein complexes were processed the H $^{++}$ program, which is an automated system that computes pK_a values of ionizable groups in protein and adds missing hydrogen atoms according to specified pH of the environment [34]. Special attention was given to the protonation state of histidine suggested by this program and adjustment was made accordingly.

The MD simulation was carried out on AMBER 16 platform using ff14SB force field [30]. Original water molecules from the crystal structure was removed and hydrogen atoms were added using TLEAP package from AMBER [30]. The standard protonation states were further checked using H^{++} online software and verified manually for the consistency and accuracy [34]. Appropriate number of counterions were added to balance the resulting charge of the protein complex and the structure was solvated in a rectangular TIP3P water box, with the boundaries of at least 12 Å away from the protein atoms [35]. Parameters files for MD, including coordinate and topology files, were generated for further processing. The solvated structure of the complex was energy minimized using steepest descent and conjugated gradient methods, each for 5000 steps, first with positional restrain force constant of 100 kcal/ (mol. Å^2) on all heavy atoms of the protein and then without any positional restrain on the whole system [35]. The minimized system

was heated linearly to 300 K for 120 ps under constant volume conditions (NVT) with no position restraints. The system was further simulated for 1000 ns under constant pressure and temperature conditions (NPT). Temperature was controlled using Langevin dynamics with collision frequency of 1 ps^{-1} during heating, equilibration and production steps. All covalent bonds to hydrogen atoms were constrained by the SHAKE method, with numerical integration time step of 2 fs [35]. Long range electrostatic attractions were computed using Particle Mesh Ewald (PME) method with and a cutoff of distance of 10 Å [36].

The simulation results were analyzed using the CPPTRAJ package from AMBER16 and the PyMOL [30, 37, 38]. Root mean square deviation (RMSD) was monitored throughout the simulation to ensure equilibration of the system. Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and energy decomposition analyses were performed using MMPBSA.py python script incorporated in AmbertTools16 [30, 39, 40]. A generalized Born implicit solvent model with 0.15 M salt concentration was used for all MM/GBSA analysis [40]. The results from MM/GBSA run were calculated using 500 snapshots for each simulation and were averaged to estimate global free energy of binding between CspA and hFH CCP5. Since we were only interested in the relative contributions of the interface residues to the formation of the complex, the entropic contribution to the free energy change was not included in the calculation. Further, entropic calculations using currently available platform are computationally very expensive and tend to have large margin of errors, introducing significant uncertainty in the result.

After successful MD simulation run, criteria such as density, temperature, pressure, root mean square deviation (RMSD), root mean square fluctuation (RMSF), KE, PE and total energy were analyzed to confirm system convergence to the equilibrium. CPPTRAJ module from

AMBER16 was used to calculate backbone RMSD values to monitor the stability of the protein complex during the entire simulation [37]. Root mean square fluctuation (RMSF) was calculated to measure fluctuation of individual residues during simulation. Similarly, CPPTRAJ module was used to analyze various hydrogen bonding properties, such as hydrogen bond distance and occupancies of each hydrogen bond interactions between interface residues from CspZ and hFH CCP7. A distance cutoff of 3.3 Å (between acceptor and donor heavy atoms) and an angle cutoff of 120° (between donor heavy atom-hydrogen-acceptor heavy atom) were used to track all hydrogen bonds. Special attention was given to the stable hydrogen bonds, which occupied at least 50% of the MD simulation time. Hydrogen bond occurrence of more than 50 % of the total simulation time are considered present and are discussed in the result section.

4.2.8. Computational Alanine Scanning Mutagenesis

Alanine mutagenesis was also performed using the MD simulation data for CspZ-hFH CCP7 protein complex. *In-silico* alanine scanning mutagenesis is quick way to calculate the binding energy contribution of residues in protein-protein interaction interfaces. Because of the size, main chain conformation and electrostatic properties of alanine, it is usually the best choice for mutagenesis study. MM/GBSA is a widely used and one of the most accurate and reliable in-silico approaches for predicting binding free energies of the protein complexes.

4.3. Results

4.3.1. Purification of CspZ and hFH CCP7 Proteins

Both CspZ and hFH CCP7 proteins were purified to homogeneity using various affinity and gel filtration chromatographic techniques described earlier (Fig. 4.1). Final protein yield was 185 mg/ml for CspZ and 5 mg/ml for hFH CCP7.
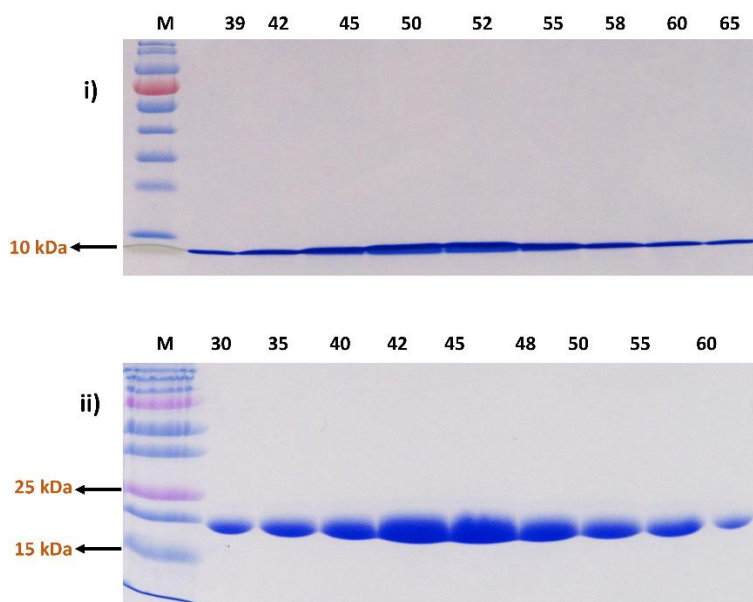
Figure 4.1. Coomassie blue (15 %) gel stain of (i) CspA and (ii) hFH CCP5 fractions collected from final sephadex-G-75 gel filtration column. M is the protein molecular weight markers and subsequent numbers indicate the protein fractions collected.

4.3.2 Structure of CspZ in Complex with hFH CCP7

The crystal structure of the complex of CspZ and hFH CCP7 was solved at 1.8 Å resolution by molecular replacement using the structure of free form CspZ (PDB entry: 4CBE) [15] and the structure of the CCP7 domain in hFH CCP 6-8 (PDB entry: 2UWN) [23] as the search models (Fig. 4.2). The structure is of good quality as indicated by the statistics for the crystallographic data and structure refinement (Table 4.2). Each asymmetric cell contains two CspZ (chains B and C)

120

and two hFH CCP7 (chains A and D) molecules with two complex forms, each consisting of one CspZ and one hFH CCP7 molecule. The electron density maps for all molecules are well defined, as illustrated in Fig. 4.2a for part of the CspZ–CCP7 interface. Since the two molecules of the same protein in the asymmetric cell are very similar to each other except the C-terminal three residues of CspZ, which are far away from the interface, we will hereafter use chains A and B to represent CCP7 and CspZ subunits, respectively, for structural analysis.
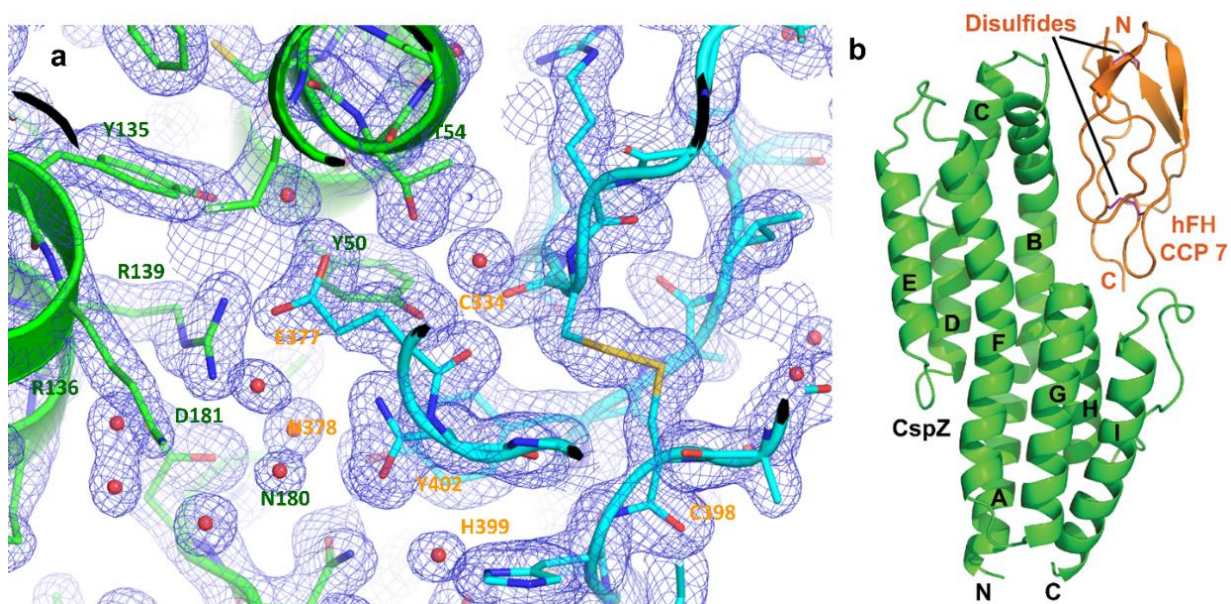
Figure 4.2. The overall structure of the CspZ:hFH CCP7 complex. (a) Part of the electron density map ($2F_o - F_c$) of the interface. The carbon atoms of the CspZ residues are colored in green and those of the CCP7 residues in orange. (b) Cartoon representation of the complex structure. The two disulfide bonds in CCP7 are represented with pink sticks. CspZ is in green and hFH CCP7 in orange. The nine helices of CspZ are labeled with letters A-I, and the N- and C-termini are indicated for both molecules.

Table 4.2. Data collection and refinement statistics for the structure of CspZ:hFH CCP7 complex

Data collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	43.75, 53.88, 116.94
A, β, γ (°)	90.0, 93.0, 90.0
Wavelength (Å)	0.97856
Resolution (Å)	39.60-1.80 (1.82-1.80) *
R_{sym} or R_{merge}	0.116 (0.388)
$I / \sigma I$	6.10 (1.80)
Completeness (%)	98.0 (99.4)
Redundancy	4.1 (4.0)
Refinement	
Number of reflections	205745
Number of unique reflections	49925
R_{work}	0.1792
R_{free}	0.2225
Number of atoms	4915
Protein	4414
Water	499
B-factors	
Protein	29.40
Water	36.90
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.040
Ramachandran outliers (%)	
Residues in most favored regions	99.0
Residues in allowed regions	1.0
Residues in not allowed regions	0.0

*Values in parentheses are for highest-resolution shell.

The structures of both CspZ and CCP7 molecules in their free forms are largely maintained upon formation of the complex. The RMSD values are 1.07 Å for 206 pairs of Cα atoms between the free CspZ in the published structure [15] and the complexed CspZ in our structure, and 0.82 Å for 59 pairs of Cα atoms between the free CCP7 in the published structure [23] and the complexed CCP7 in our structure. The result indicates that the structures of the free and the CCP7-bound CspZ are very similar and the structure of the CspZ-bound CCP7 is essentially the same as that of the free CCP7 (Fig. 4.3). Like the free CspZ, the CspZ molecule in the complex consists of nine helices (Fig. 4.2b). Although CspZ is a single-domain protein with a single hydrophobic core, the protein can be divided into two parts. Both parts are essentially helical bundles: the first part consists of the first five helices (A−F) and the second the last three helices (G−I) and a long loop that connects helices H and I. The first helical bundle is much longer than the second, and the two bundles of helices are packed against each other, forming a chair-like structure. The seat of the chair is formed by one end of the shorter second bundle of helices, and the back is made of helices B and F of the longer first bundle of helices. CCP7 sits snugly on this CspZ chair, interacting with residues that form the seat and back of the chair (Figs. 4.2b and 4.3).
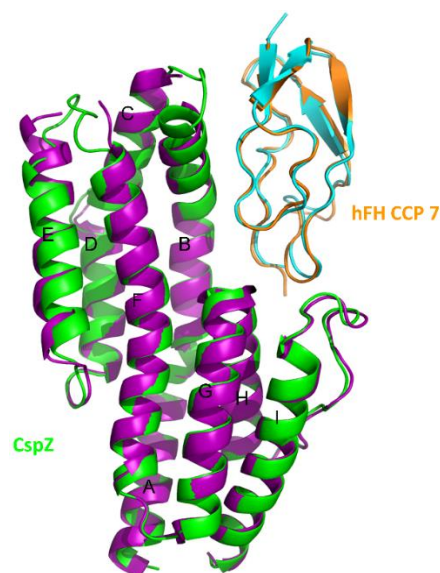
Figure 4.3. Structural comparison between the complexed and the free forms of CspZ and hFH CCP7. CspZ is in green in the complex and magenta in the free form. hFH CCP7 is in orange in the complex and cyan in the free form.

Although the overall structures of the free and the CCP-7 bound CspZ are very similar, there are significant local differences between the structures of the two forms of the protein (Fig. 4.3). The most significant differences are located in two regions, residues 63–69 and 121–127. The first region encompasses the N-terminus of helix B and the loop that connects helix B with helix C, and the second region constitutes most part of the loop that links helix E with helix F. The two regions pack against each other and form one end of the longer helical bundle. In addition to the conformational changes as reflected in the Cα RMSDs, the formation of the complex also makes these two regions more rigid, as the electron density map is well refined in the complex (this work) but poor in the free CspZ (published work [15]). There is no electron density for residues E65-F68 and V125 in the free CspZ. It should be noted that these two regions are not involved in binding CCP7 and the functional significance of these conformational differences is not clear.

4.3.3. Analysis of the Key Interactions at CspZ:hFH CCP7 Binding Interface

The formation of the complex buries a total of 1692 $Å^2$ of solvent-accessible area, with about equal contributions from CspZ (815 $Å^2$) and hFH CCP7 (877 $Å^2$). This size of the buried solvent-accessible area is typical of those of protein antigen-antibody complexes [41]. With 63.7% of the buried solvent-accessible area nonpolar and 36.3% polar, the interface is largely hydrophobic. Figure 4.4 shows the spatial arrangement of the interface residues on both sides of the complex. Although the interface is largely hydrophobic, there are many polar and electrostatic interactions, including 14 intermolecular hydrogen bonds and two salt bridges (Table 4.3).
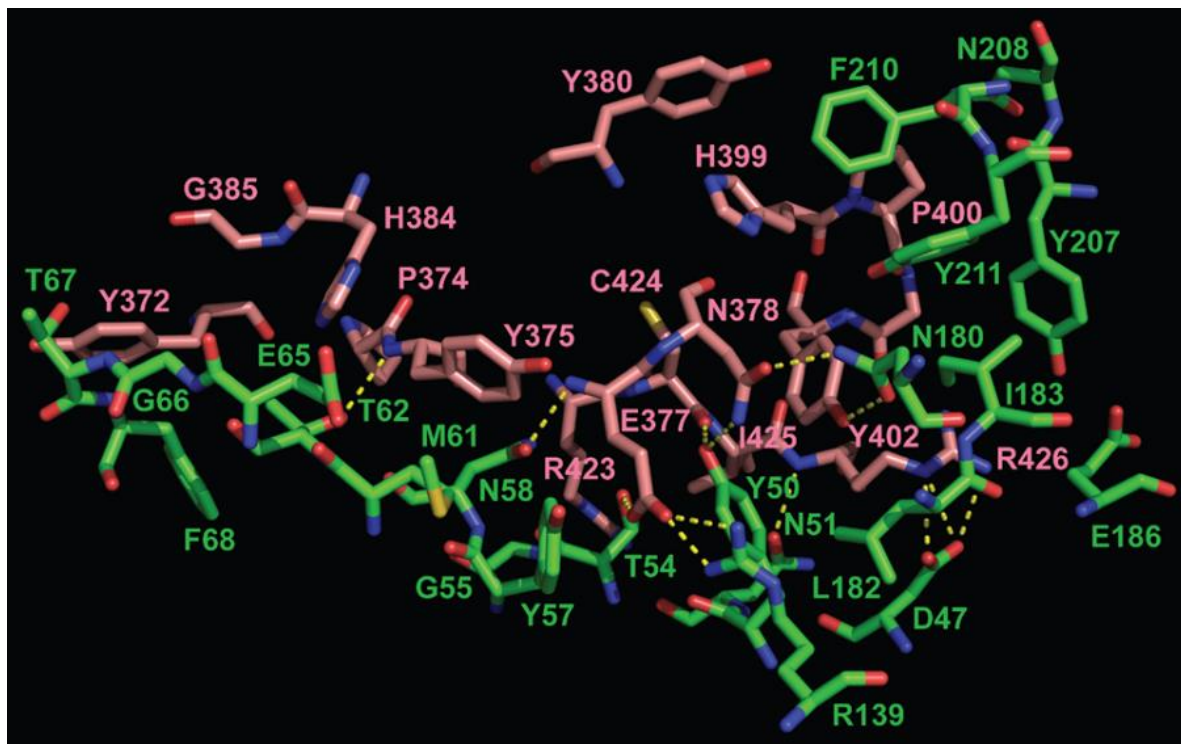
Figure 4.4. Specific interactions between CspZ and hFH CCP7. All residues are shown in the stick presentation. Intermolecular hydrogen bonds are indicated by dashed yellow lines. The carbon atoms of the CspZ residues are colored in green and those of CCP7 in pink.

Table 4.3. Intermolecular hydrogen bonds in the complex of *CspZ* and hFH CCP7

CspZ heavy atoms	FH heavy atoms	Distance (Å)*
D47–OD1	R426–NE	3.11 (3.16)
D47–OD2	R426–NH2	3.05 (3.28)
Y50–OH	E377–O	3.52 (3.52)
Y50–OH	C424–O	2.75 (2.79)
Y50–OH	N378–ND2	2.88 (2.94)
N51–ND2	R426–O	3.76
N51–O	R423–NH1	(3.29)
N51–O	R423–NH2	3.46
N51–OD1	R426–N	2.82 (2.85)
T54–OG1	E377–OE1	2.97
N58–OD1	E377–N	(2.82)
T62–OG1	Y375–N	2.90 (2.89)
T67–N	Y372–OH	(3.33)
T67OG1	Y372–OH	(3.73)
R139–NH1	E377–OE1	2.41 (3.18)
R139–NH2	E377–OE2	(2.68)
N180–ND2	N378–OD1	2.91 (2.86)
N180–OD1	Y402–OH	2.73 (2.77)
Y211–OH	P400–O	2.87 (2.75)

*Distances in parentheses are measured for the complex composed of chain C and chain D, otherwise are measured for the complex composed of chain A and chain B from the crystal structure.

Most of the polar and salt bridge interactions at the binding interface are found in the back of the chair formed by the first bundle of the helices, nine of which involve the side chains of helix B, including a salt bridge between D47 of CspZ and R426 of hFH CCP7. The other salt bridge forms double hydrogen bonds between R139 located in the middle of helix F of CspZ and E377 of hFH CCP7. Only three hydrogen bonds involve the second bundle of helices of CspZ, two of which are from the side chain of N180 residing at the sharp turn between helices G and H, and the last one from the side chain of Y211 at the N-terminus of helix I. Hydrophobic interactions involve mainly the residues from the second helical bundle of CspZ that form the seat of the chair,

including residues L182, I183, Y207, F210 and Y211. This cluster of hydrophobic interactions also involves Y50 from helix B. These residues are packed against Y380, H399, P400 and Y402 in the C-terminal region of hFH CCP7. In addition, there are two small hydrophobic patches. One patch is formed by Y57 and M61 from the first helical bundle of CspZ, packed against Y375 of CCP7, and the other by F68 of CspZ, packed against Y372 of CCP7. At the level of amino acid sequence, intermolecular interactions are located in four regions of CspZ, D47-K73, Y135-R139, N180-E186 and Y207-Y211, and 3 regions in CCP7, Y372-K387, H399-Y402 and T421-K428. To assess the relative importance of the interface residues to the formation of the complex, we performed computational alanine scanning mutagenesis of the interface residues using two structure-based methods, the BeAtMuSiC method [42] and the MutaBind method [43], developed for rapid estimation of the effect of an amino acid substitution on the binding free energy of a protein-protein complex. The very fast BeAtMuSiC method is based on a machine learning algorithm with a combination of different statistical potentials and has outperformed many other methods in the 26th Critical Assessment of Predicted Interactions (CAPRI) [44]. The recently developed MutaBind method employs a combination of molecular mechanics force fields, statistical potentials and fast side-chain optimization algorithms and had even better performances [43]. The results obtained by the two methods are remarkably consistent for the most part (Fig. 4.5). The main difference is in the estimation of the contributions of electrostatic interactions to the formation of the complex. The estimations of the contributions of D47 of CspZ (Fig. 4a) and E377 and R426 of hFH by the MutaBind method are much higher than those by the BeAtMuSiC method. However, the two estimations of the contribution of R139 are very similar. As described earlier, D47 of CspZ forms a salt bridge with R426 of hFH, and R139 of CspZ forms a salt bridge

with E377 of hFH. According to the MutaBind method, the salt bridge between D47 of CspZ and R426 of hFH is very important for the formation of the complex. Based on one or both methods, with a cutoff of 1 kcal/mol, 16 residues of CspZ, D47, Y50, N51, T54, Y57,N58, M61, T62, F68, R139, N180, L182, I183, Y207, F210 and Y211, and 10 residues of hFH, Y372, P374, Y375, L376, E377, N378, P400, Y402, I425 and R426, make significant contributions to the formation of the complex. E65, T67, E186, and N208 of CspZ and Y380A, H384, H399 and R423 of hFH are not important for the formation of the complex.
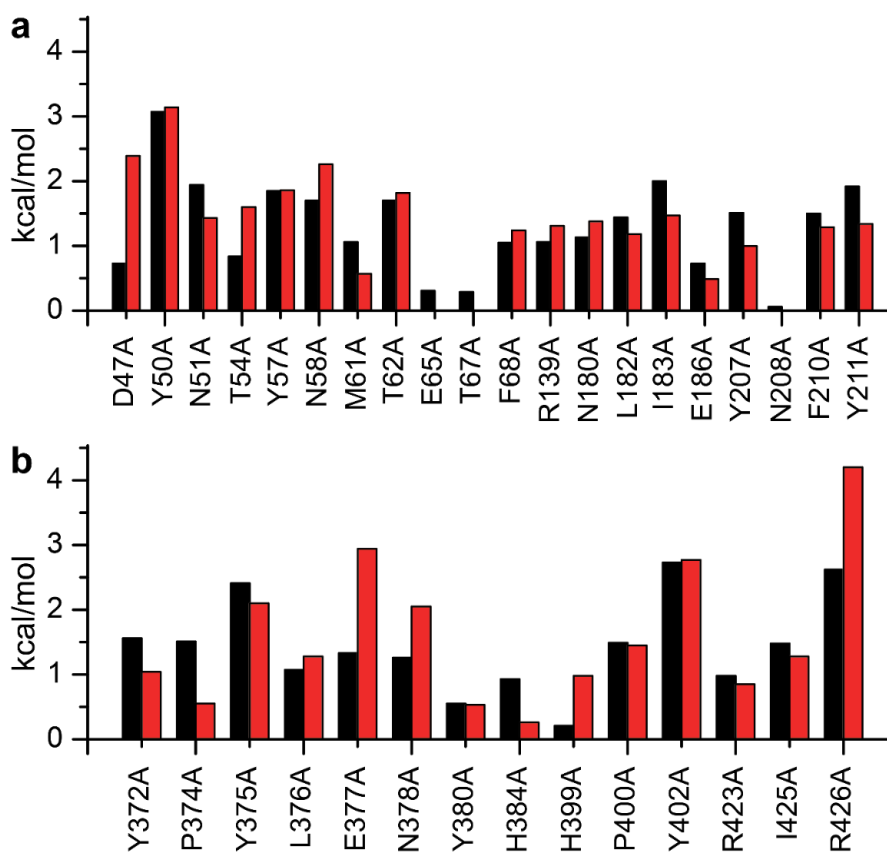
Figure 4.5. Computational alanine scanning mutagenesis of CspZ (a) and hFH CCP7 (b). The changes in binding energy caused by the mutations were calculated by the BeAtMuSiC method (black columns) [42] and the MutaBind method (red columns) [43].

4.3.4. MD Simulation Analysis of CspZ:hFH CCP7 Complex

4.3.4.1. RMSD and RMSF Analysis

The two popular computational methods described in the previous section are fast and effective way to calculate the relative effect of mutations. However, in order to further analyze the interactions of the two proteins, we performed MD simulation of the CspZ:hFH CCP7 complex. The MD simulation trajectory was analyzed using the CPPTRAJ module of Amber16[30, 37]. The backbone RMSD was first calculated for individual proteins of the complex by superimposing only the protein of interest. The RMSDs of the individual proteins were low, with an average RMSD of 1.14 ± 0.02 Å for hFH CCP7 and 1.53 ± 0.12 Å for CspZ. The RMSD of the whole complex is 1.53 ± 0.10 Å (Fig. 4.6). Based on the RMSD analysis, the complex of CspZ and hFH CCP7 showed no significant change in global conformation upon complex formation. The RMSF plot also shows that residue fluctuation upon complex formation is quite stable (Fig. 4.6). There are three regions, labelled (i) to (iii), in hFH CCP7 and five regions, labelled (iv) to (viii) in CspZ with bigger fluctuations as indicated by sharp peaks in RMSF plots. However, all of these residues are from the loop regions in CCP7 and CspZ. The RMSF of all other regions stayed below 1 Å, indicating no major fluctuations during complex formation. This is also consistent with the structural alignment between bound and free form CspZ and hFH CCP7 (Figure 4.2).
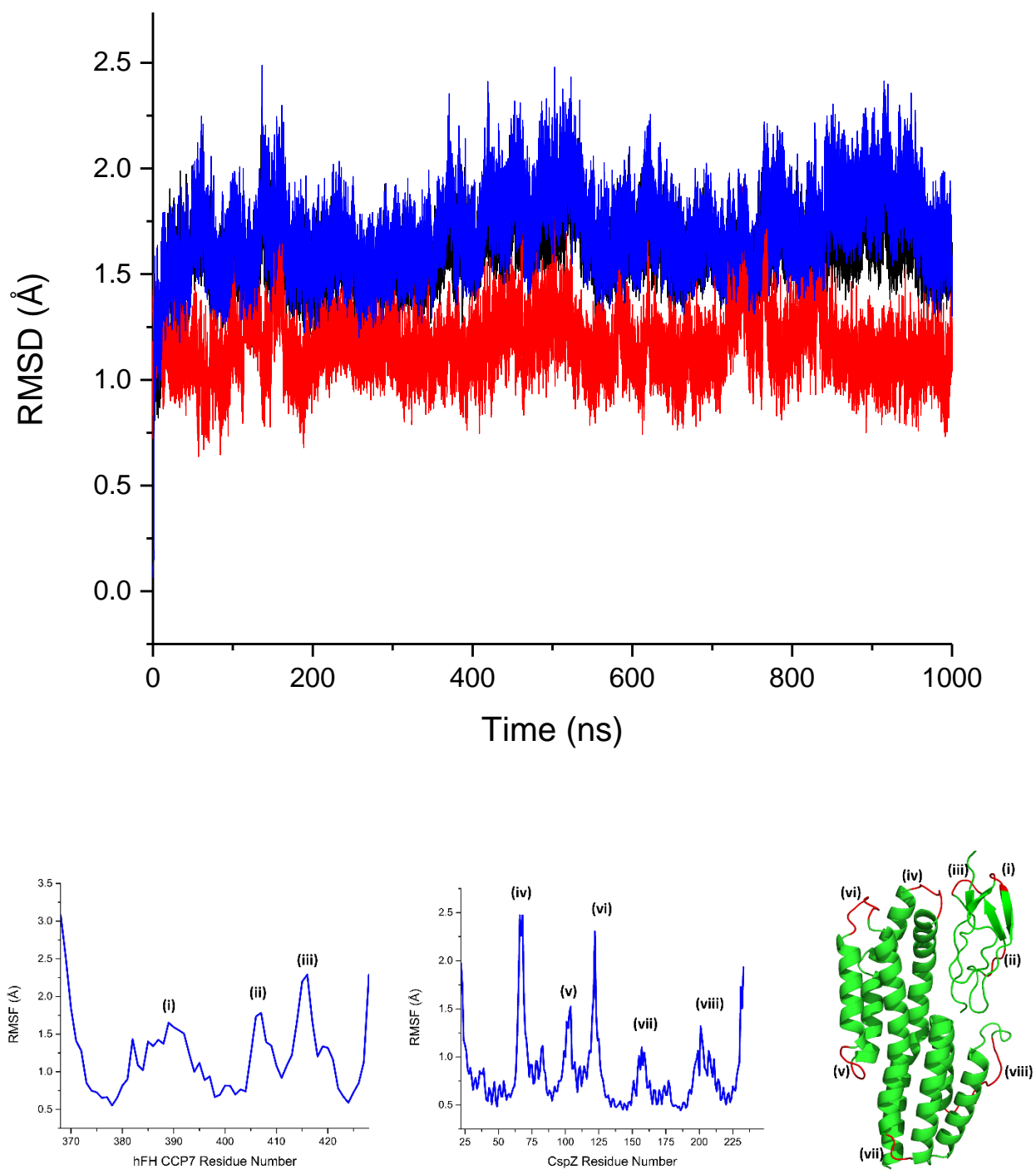
Figure 4.6. (Top) Backbone RMSD plot of free form hFH CCP7 (red), free form CspZ (blue) and the complex of CspZ and hFH CCP7. As evident by the plot, there are no major change in global conformation upon complex formation. (Bottom) RMSF of hFH CCP7 and CspZ. All major fluctuations in RMSF for CCP7 are labelled (i) to (iii) and (iv) to (viii) for fluctuation peaks from CspZ. As depicted in the figure on the right, all fluctuations are originating from the loop region on the right.

4.3.4.2. Determining Key Interface Residues of CspZ and hFH CCP7

The crystal structure of the CspZ:hFH CCP7 show some extensive hydrogen bonds as well as hydrophobic interactions between the two proteins. The relative contributions of the individual interface residues to the formation of the complex were further analyzed (Figs. 4.7 and 4.8).
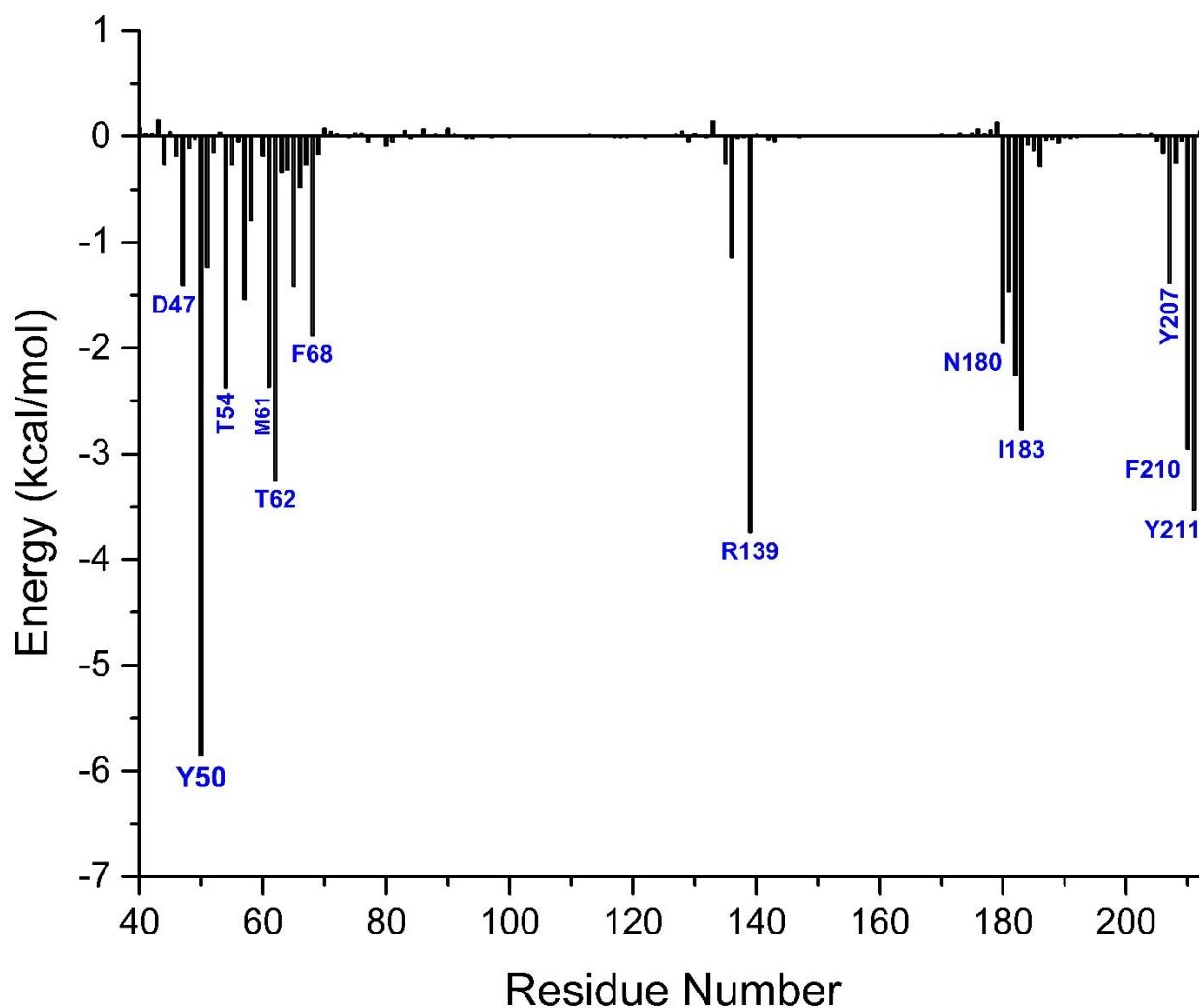
Figure 4.7. Decomposition energy calculation of CspZ residues. Residue name and number are labelled at the end of the bars with blue color.

Based on the energy decomposition analyses, Y50 of CspZ and R426 of hFH CCP7 showed the largest contribution in stabilizing the CspZ:hFH CCP7. Given that D47 of CspZ form salt bridge and hydrogen bonds with R426 of hFH CCP7, we expected a high binding energy contribution from D47. However, based on the decomposition energy of D47, the energy contribution is relatively low with ~ 1.4 kcal/mol. To better understand this very low energy contribution of D47, MD simulation of the D47A mutant protein was performed.
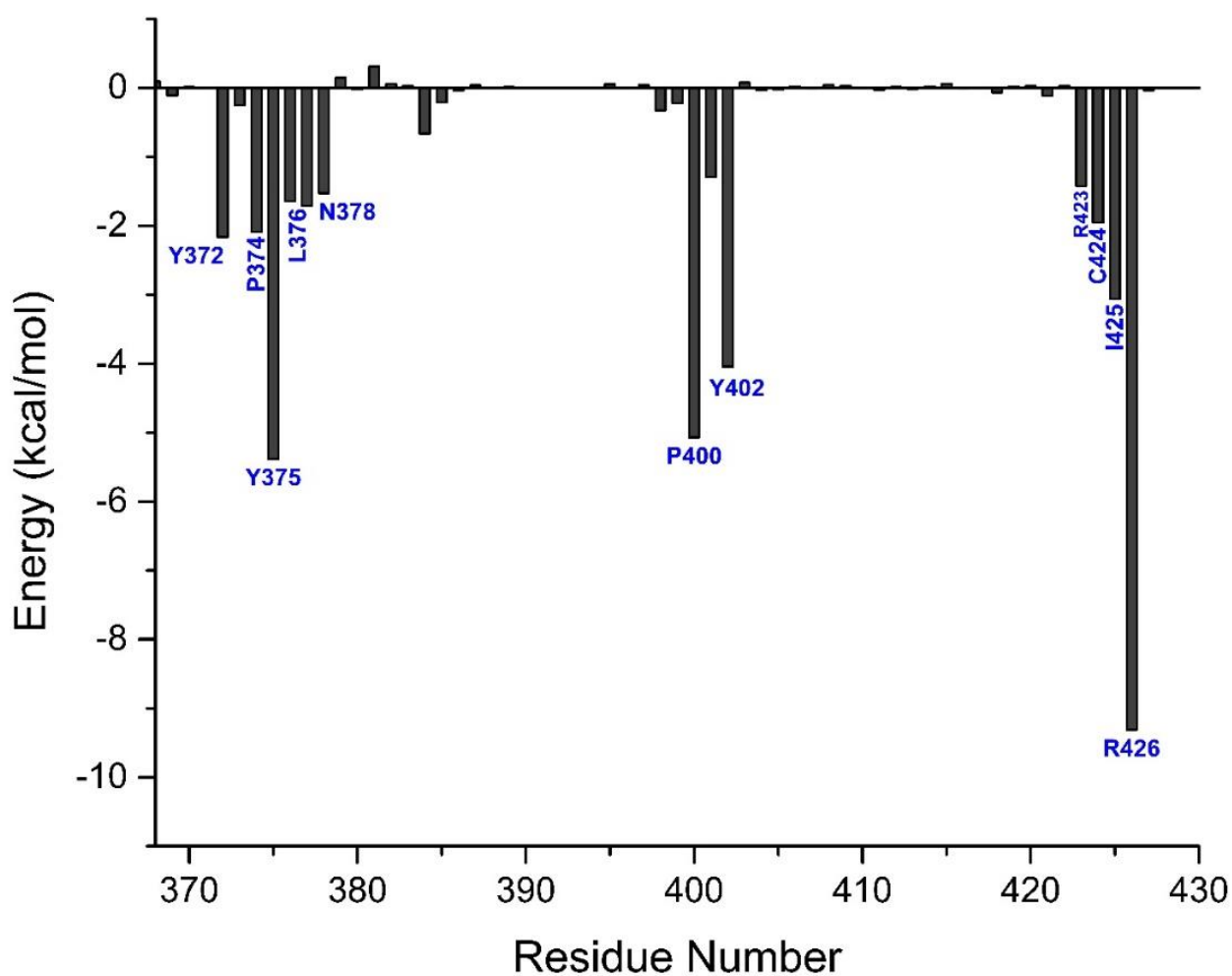
Figure 4.8. Relative contributions of hFH CCP7 residues to binding of CspZ based on the energy decomposition analysis. Residue name and number are labelled at the end of the bars with blue color.

Upon closer look at the binding interface involving of D47A mutation, the long side chain of R426 that forms hydrogen bond and salt bridge interactions with D47 moves freely, with its non-polar part of the side chain coming in close contact with residues Y50, L188 and I183, forming a stabilizing hydrophobic pocket at the interface. So, most likely, even when hydrogen bonds from D47 are broken upon mutation, there are additional non-polar interactions that results in less binding energy contribution from Asp47. Mutating R426A from hFH CCP7 not only abolishes hydrogen bond and salt bridge interactions with D47 of CspZ but also abolishes another stable hydrogen bond interaction with N51. As a result, R426 from hFH CCP7 has the highest energy contribution to the binding. N180 and Y211 of CspZ are also making hydrogen bonds with Y402 and P400 of hFH CCP7, respectively. N180 and Y211 both show greater than -2 kcal/mol of decomposition energy. In addition to making hydrogen bond with P400, the side chain ring of Y211 of CspZ is packed with other hydrophobic residues, F210, I183 and Y207 of hFH CCP7. This interaction stabilizes the position of Y211 which in turn makes hydrogen bond with hFH CCP5.Decomposition energy and alanine scanning mutagenesis studies both show ~ -4 kcal/mol of binding energy for P400. Binding energy for Y402 is ~ -6 kcal/mol. In addition to a single hydrogen bond, Y402 is also making hydrophobic interaction with Y50, L182 and I183 of CspZ, accounting for higher energy contribution than P400.

To further validate our computational results, we performed experimental alanine scanning mutagenesis study in CspZ and analyzed the binding thermodynamics parameters using ITC (Table 4.3). While the results from the computational analyses predicted the relative binding, energy associated with interface residues, there were some discrepancies between computation and experiment. While the computational decomposition energy for CspZ indicates that Y50 had

the highest binding energy, the results from ITC showed that Y211 contributed the most to the stability at the interface (Table 4.4). Y211A mutation decreased the binding affinity by ~ 1,800-fold, with a decrease in binding free energy by ~ 4.5 kcal/mol. T50A also drastically decreased the binding affinity, with a decrease in binding free energy by 3.7 kcal/mol. Residues making important hydrogen bonds and hydrophobic interactions at the interface, D47, Y50, N51, T54 and F210 of CspZ, were also characterized by significant decreases in binding free energy, indicating greater stabilizing effect of these residues at the interface. The order of mutational effects with the highest effect on the binding energy to the lowest is Y211A > Y50A > Y207A > R139A > T62A > N58A > N51A > D47A > L182A > N180A > F210A in CspZ and R426A > Y375A > Y402A > P400A > Y372A > E77A > I425A in hFH CCP7.

Table 4.4. Thermodynamics binding parameters for CspZ mutagenesis from ITC

CspZ Mutants	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	T ΔS (kcal/mol)
WILD	**0.30 ± 0.10**	**-13.01 ± 0.23**	**-8.62 ± 0.21**	**4.41 ± 0.22**
D47A	26.10 ± 1.64	-10.33 ± 0.03	-3.13 ± 0.31	7.21 ± 0.35
Y50A	149.01 ± 2.32	-9.31 ± 0.28	-2.83 ± 0.12	6.49 ± 0.86
N51A	31.21 ± 3.54	-10.24 ± 0.85	-4.52 ± 0.18	5.69 ± 0.79
T54A	6.34 ± 0.29	-11.18 ± 0.89	-6.38 ± 0.36	4.79 ± 0.28
Y57A	4.50 ± 1.56	-11.40 ± 0.21	-8.23 ± 0.50	3.17 ± 0.71
N58A	34.45 ± 2.86	-10.18 ± 0.19	-6.31 ± 0.46	3.86 ± 0.39
M61A	6.48 ± 1.67	-11.18 ± 0.15	-12.66 ± 1.54	-2.68 ± 0.18
T62A	47.10 ± 3.69	-9.99 ± 0.24	-9.05 ± 1.01	0.94 ± 0.16
E65A	2.74 ± 0.17	-11.8 ± 0.17	-9.15 ± 0.05	2.66 ± 0.66
F68A	3.68 ± 0.97	-11.5 ± 0.20	-9.20 ± 0.02	2.32 ± 0.29
R139A	53.07 ± 0.27	-9.92 ± 0.03	-10.88 ± 0.57	-0.96 ± 0.42
N180A	8.01 ± 0.02	-11.04 ± 0.19	-7.42 ± 0.92	3.62 ± 0.92
L182A	15.85 ± 0.82	-10.67 ± 0.03	-10.75 ± 1.17	-0.08 ± 0.01
I183A	4.01 ± 0.37	- 11.45 ± 0.05	- 13.04 ± 1.98	-1.58 ± 0.19
Y207A	104 ± 5.36	-9.56 ± 0.03	-8.86 ± 0.06	0.70 ± 0.09
F210A	6.13 ± 2.07	-11.21 ± 0.20	-3.73 ± 0.14	7.47 ± 0.06
Y211A	546 ± 18.65	-8.54 ± 0.09	-4.06 ± 0.71	4.48 ± 0.75

Also, in addition to identifying important residues at the interface, we studied the stability of all possible hydrogen bonds at the interface (Table 4.5). Although the initial crystal structures show extensive hydrogen bonding interaction stabilizing the CspZ:hFH CCP7 complex, not all hydrogen bonds have the same energy contribution to the stability of the complex. Our MD analysis of hydrogen bonds in the 1000 ns simulation trajectory indicates that there are fourteen hydrogen bonds that occupy at least 80 % of the time. The hydrogen bond between D47-R426, Y50-C444, N51-R426 and T62-Y375 are close to 100 % occupancy, implying the most stable interactions at the interface. The least stable hydrogen bond is between E65-H384. E65 from CspZ is in the loop region connecting helix B and C, and H384 in hFH CCP7 is located also in the loop region connecting the first and second β-strands. Interestingly, the distance between the backbone oxygen of E65 and the NE2 of H384 in the crystal structure is more than 5 Å. This indicates that this hydrogen bond only forms intermittently because the two loops fluctuate in a relatively large amplitude during the simulation.

Table 4.5. Intermolecular hydrogen bonds at the interface of the complex of CspA and hFH CCP5

CspZ……… hFH$_7$	Occurrence (%)	Distance (Å)
D47-OD1…..NE-R426	99.48	2.84 ± 0.12
D47-OD2…..NH2-R426	95.16	2.82 ± 0.15
Y50-OH…..O-C444	98.87	2.74 ± 0.13
Y50-OH…..ND2-N378	91.08	3.00 ± 0.18
N51-OD1…..N-R426	98.21	2.90 ± 0.15
D59-OD1…..NH1-R423	92.43	2.82 ± 0.18
D59-OD1…..NH2-R423	83.48	2.98 ± 0.21
D59-OD2…..NH2-R423	85.99	2.94 ± 0.22
T62-OG1…..N-Y375	98.02	2.95 ± 0.13
E65-O…..NE2-H384	66.54	2.92 ± 1.60
R139-NH1…..OE1-E377	89.92	2.93 ± 0.25
R139-NH1…..OE2-E377	80.24	2.97 ± 0.31
N180-ND2…..OD1-N378	81.14	2.91 ± 0.22
N180-OD1…..HO-Y402	94.06	2.78 ± 0.19
Y211-OH…..O-P400	99.82	2.73 ± 0.13

4.3.5. Structural Comparison with the Published Results

The effort to identify the binding interface on CspZ began long before the structure determination of the free CspZ [15] and the structure determination of the complex of CspZ and hFH CCP7 (this work). Based on a peptide mapping analysis using a library of synthetic peptides derived from CspZ, four regions have been identified to bind to FH, residues K34−E52, N127−A145 and N202−G226 [45]. The first region (K34−E52) identified in the peptide mapping overlaps with the N-terminal residues of the first region (D47−K73) identified in our crystallographic analysis. The second region (N127−A145) identified in the peptide mapping analysis encompasses the second region (Y135−R139) identified in our analysis. The third region (N202−G226) identified in the peptide mapping analysis contains the fourth region (Y207−Y211) identified in our analysis. An additional region, D77−V88, has been identified in the peptide

mapping analysis to bind hFH-like protein 1 (FHL1), a protein consisting of the first seven domains of hFH and generated by alternative splicing of the hFH gene, but not hFH. This region overlaps with the first region (D47−K73) identified in our crystallographic analysis. The peptide analysis has not identified the third region (N180−E186) which is identified in our analysis.

In addition, it has been shown using deletion mutagenesis that the 16 C-terminal residues are important for binding hFH [45]. The 16 C-terminal residues are not part of the interface of the complex. Most likely they play a role in maintaining the structural integrity of CspZ rather than a role in binding of hFH.

Extensive site-directed mutagenesis studies were performed to identify specific residues that are involved in binding FH [46, 47]. Alanine scanning mutagenesis was performed on nearly every residue of the three regions of CspZ identified by peptide mapping [47]. Single point mutations causing impaired binding of hFH include Q71A, N75A, S79A, F81A, D84A, R129A, H130A, R139A, R204A, R206A, Y207A, F210A, Y211A and E214A, suggesting that these mutated residues are at the interface of the complex. However, when these residues are mapped onto the CspZ molecule in our crystal structure of the complex (Fig. 4.9), it is clear that only four residues, R139, Y207, F210 and Y211, are at the interface. Since the binding residues identified by the mutagenesis study are distributed in several regions of the protein surface, most of these residues are unlikely to be directly involved in binding FH, as noted earlier when the crystal structure of the free CspZ was determined [15].

a

```
25
NQRNINELKI  FVEKAKYYSI  KLDAIYNECT

55
GAYNDIMTYS  EGTFSDQSKV  NQAISIFKKD

85
NKIVNKFKEL  EKIIEEYKPM  FLSKLIDDFA

115
IELDQAVDND  VSNARHVADS  YKKLRKSVVL

145
AYIESFDVIS  SKFVDSKFVE  ASKKFVNKAK

175
EFVEENDLIA  L CIVKTIGD  MVNDREINSR

205
SRYNNFYKKE  ADFLGAAVEL  EGAYKAIKQT

235
LL
```

b
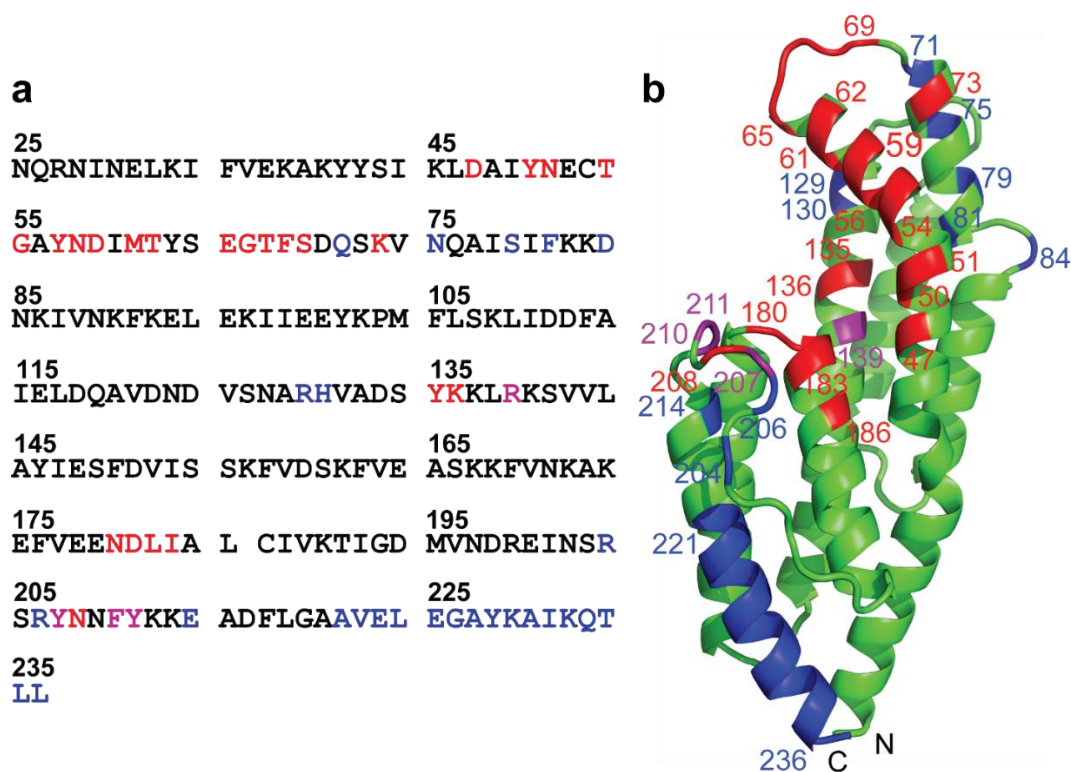
Figure 4.9. Comparison of the crystallographic and mutagenesis analyses of the binding of CspZ and FH. (a) The amino acid sequence of CspZ, and (b) the cartoon representation of CspZ in the complex with hFH CCP7. Interface residues identified by both crystallographic and mutagenesis analyses are colored in magenta, those only by crystallographic analysis (this work) in red, and those only by alanine-scanning mutagenesis analysis in blue. The numbering of residues in the structure is the same as that in the sequence.

Most of the interface residues identified in the site-directed mutagenesis study are involved in maintaining the structural integrity or the stability of the protein rather than in binding of FH (Fig. 4.9). In particular, Q71 is located at the N-terminal end of helix C and its side-chain amide forms a hydrogen bond with the hydroxylic group of Y63 at the C-terminal end of helix B. The side-chain amide of N75 makes a hydrogen bond with the backbone amide of N123 in the loop between helices E and F. F81 is a completely buried in the hydrophobic core of the protein. The guanidinium group of R129 participates in the hydrogen bond network formed by the side-chains of D33, S64 and E65. The mutagenesis study identified 14 interface residues or 30 if the 16 C-terminal residues are included. The fact that only four of these 14 or 30 residues

are found at the interface of the complex suggests that one must be cautious in using mutagenesis analysis alone to identify interface residues, and structural information is required to distinguish a structural role from a functional role that can be played by a mutated residue.

To explain binding between CspZ and hFH, a coiled-coli hypothesis has been suggested and tested using two double mutations on CspZ, one with I29 mutated to threonine and L32 to arginine and the other with F91 mutated to S and L94 to R [46]. The failure of both variants to bind hFH was taken as an indication that these residues are important for binding hFH and as a support for the hypothesis. However, according to the crystal structures of CspZ (Brangulis and coworkers [15] and this work), all four hydrophobic residues are buried and are unlikely to be involved in binding hFH. Most likely these drastic mutations, substituting two hydrophobic residues with a polar and a charged residue, caused a major damage to the structural integrity of the protein and consequently abolished the binding of CspZ to hFH. These residues may be important for maintaining the structural integrity of CspZ but are unlikely to be involved in binding of FH. Furthermore, based on the crystal structures of CspZ (Brangulis and coworkers [15] and this work), CspZ does not contain any coiled-coil structure as previously suggested [45].

4.3.6. Insights into the Host Specificity of *Borrelia* Based on FH Recruitment by CspZ

Because FH binding is correlated with borrelial host range [8], it is important to clarify the host specificity of FH recruitment by CspZ. So far western blotting has been used to detect CspZ binding with FH proteins from sera of human and various animal sources [46]. Strong FH bands have been detected for human and mouse sera, moderate FH bands for guinea pig and cow sera, and weak FH bands for monkey, pig, rabbit, and duck sera. No FH band was detected for dog, cat, goat, horse, rat and chicken sera [8]. However, it is difficult to rank the affinities of CspZ for these

FH proteins, because the serum FH concentrations were not controlled, and the experimental procedure used SDS-PAGE to separate FH from other proteins. In this work, we measured the thermodynamics parameters CspZ binding of hFH and mFH (Fig. 4.10 and Table 4.6). mFH was chosen because mouse is a major borrelial host and an excellent model system for studying borrelial pathogenesis. CspZ binds both CCP7 domain tightly, with a sub-nM K_d for hFH CCP7 and a 2-nM K_d for mFH CCP7. A large favorable enthalpy change drives the formation of both complexes, but the magnitude of the enthalpy change for the formation of the complex with mFH CCP7 is significantly larger than that for the formation of the complex with hFH CCP7. The smaller enthalpy change for the formation of the complex with hFH CCP7 is compensated by a large favorable change in entropy.

Table 4.6. Thermodynamics parameters of binning of CspZ to hFH and mFH CCP7

	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	TΔS (kcal/mol)
hFH CCP7	0.3 ± 0.1	-13.0 ± 0.2	-8.6 ± 0.2	4.4 ± 0.2
mFH CCP7	1.9 ± 0.1	-11.9 ± 0.4	-12.4 ± 0.1	-0.5 ± 0.1
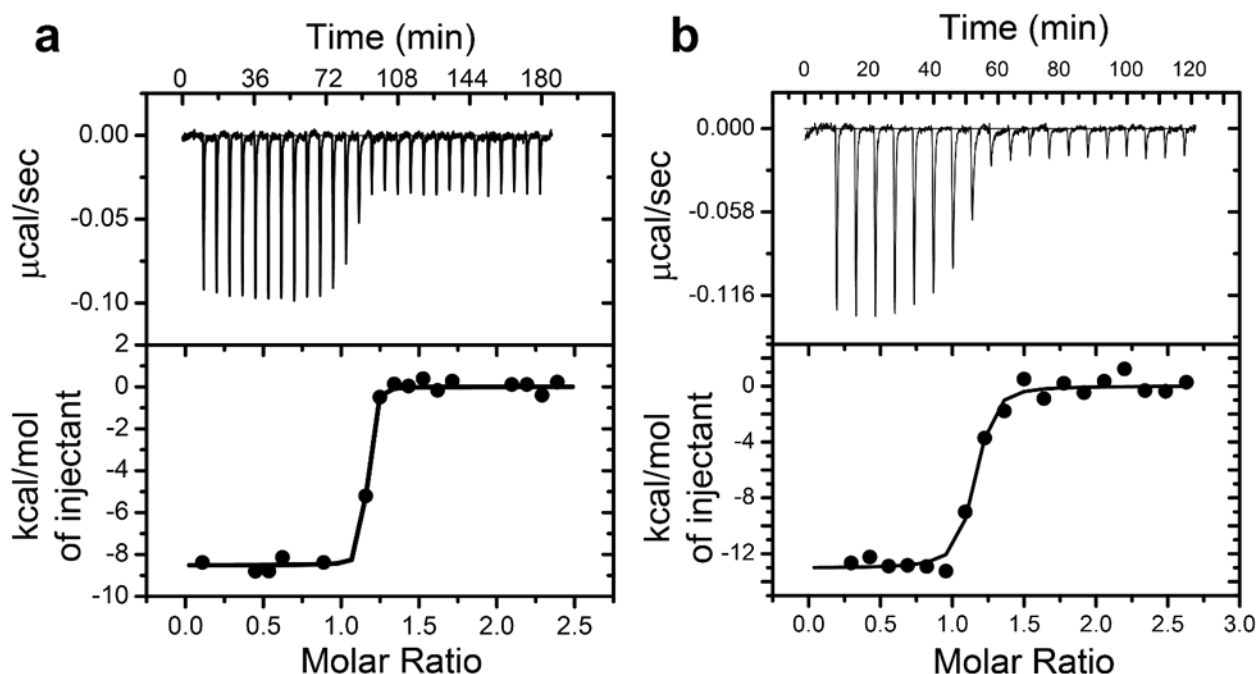
Figure 4.10. ITC measurement of the binding of CspZ to hFH CCP7 (a) and mFH CCP7 (b). The ITC experiments were carried out at 25 °C. The injections were made over a period of 180 min with a 6 min interval between subsequent injections. The sample cell was stirred at 310 rpm.

To understand the structural basis for the host specificity of FH recruitment by CspZ, we aligned eight FH CCP7 amino acid sequences (Fig. 4.11). Surprisingly, the sequence identities between hFH CCP7 and other CCP7 domains are relatively low, ranging from 47% to 61%. Many interface residues are not conserved. To assess the impacts of the natural variations on binding of CspZ, we performed structure-based computational analyses using the BeAtMuSiC method [42] and the MutaBind method [43]. Based on the computational analyses (Fig. 4.5b), substitutions with significant effects on binding of CspZ include Y372V, Y372I, Y372N, Y372T, P374D, P374H, P374N, P374H, N378H, N378Y, G385E, P400E, P400N, P400S, R426H, and R426L. All animal FH proteins have at least three significant substitutions (Table 4.7). Based on the calculations by the BeAtMuSiC method, the relative affinities of various FH proteins for CspZ are

as follows in descending order: human > rabbit > rat, pig, sheep, bovine > mouse > horse. The

relative affinities calculated by the MutaBind method are similar: human > rabbit > rat, mouse,

pig, sheep, bovine > horse > pig. The most significant difference is pig FH, mainly due to the

Arg426Leu substitution, which has a much higher energetic cost by the MutaBind method. A

weak or no FH band for rabbit and rat in the far western blotting assay [46] are not a true

reflection of their affinities for CspZ. Our results predict that the affinities of rabbit and rat FH for

CspZ are likely as high as that of mFH.
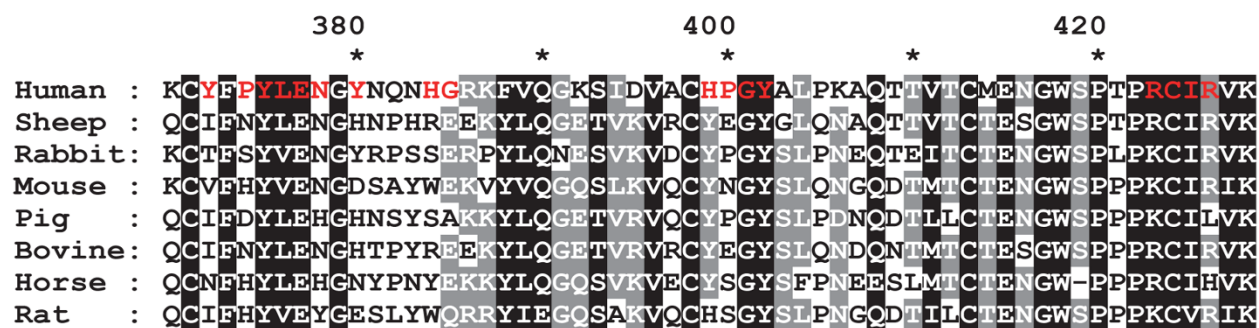
Figure 4.11. Amino acid sequence alignment of the CCP7 domain of hFH with those of orthologous FH proteins. The conserved residues among all species are shaded in black and the interface residues in the complex of hFH CCP7 with CspZ are colored in red. The amino acid numbering is that of hFH, excluding the N-terminal signal peptide as commonly done in the FH literature.

Based on the sequence analysis, we further performed reciprocal mutagenesis study

between hFH and mFH CCP7. Key residues from hFH CCP7, Y372, P374, P400 and R423 were

mutated to mFH residues V, H, N and K respectively. Similarly, mFH CCP7 residues, V372, H374,

N400 and K423 are mutated to Y, P, P and R. Our computational analysis shows that even

introducing mutations on four key residues in hFH CCP7 that corresponds to mFH residues, the

relative free energy change of stays the same (Table 4.7) with almost identical energy

contributions from polar and non-polar interactions. Mutating these four residues in wild type

mFH CCP7 increases the overall binding free energy that is close to the binding free energy change of the wild type hFH CCP7. These mutations in hFH CCP7 and mFH CCP7 has very little effect in binding affinity to CspZ (Table 4.8). This is the key reason for very tight binding of both hFH and mFH CCP7 to CspZ with low nanomolar dissociation constant.

Table 4.7. Sequence variations in binding of CspZ and FH CCP7 from various animals

	Residue Variations	$\Delta\Delta G_{bind}$ (kcal/mol)[a]	$\Delta\Delta G_{bind}$ (kcal/mol)[b]
Sheep FH	**Y372I**, **P374N**, Y380H, H384R, **G385E**, H399Y, **P400E**	4.8	2.3
Rabbit FH	**Y372T**, **P374S**, L376V, H384S, **G385E**, H399Y, R423K	4.4	1.6
Mouse FH	**Y372V**, **P374H**, L376V, Y380D, H384W, **G385E**, H399Y, **P400N**, R423K	5.3	2.3
Pig FH	**Y372I**, **P374D**, **N378H**, Y380H, **H384S**, G385A, H399Y, R423K, **R426L**	4.7	7.2
Bovine FH	**Y372I**, **P374N**, Y380H, H384R, **G385E**, H399Y, **P400E**	4.8	2.3
Horse FH	**Y372N**, **P374H**, Y380N, H384Y, **G385E**, H399Y, **P400S**, **R426H**	7.2	6.3
Rat FH	**Y372I**, **P374H**, L376V, Y380E, H384W, **G385Q**, **P400S**, R423K, I425V	4.7	2.2

*Variations with a significant effect on binding CspZ are highlighted in bold.
[a]The sum of the effects of the highlighted variations on binding of CspZ. Calculated by the BeAtMuSiC method.
[b]The sum of the effects of the highlighted variations on binding of CspZ. Calculated by the MutaBind method.

Table 4.8. Binding free energies of hFH and mFH CCP7 calculated by MM/GBSA (kcal/mol)

System	ΔE_{vdw}	ΔE_{ele}	ΔG_{GB}	ΔG_{SA}	ΔG_{cal}
WT hFH CCP7	-98.80	381.49	414.20	-14.37	-80.47 ± 1.31
MT hFH CCP7	-98.35	-382.17	414.70	-14.38	-80.21± 1.12
WT mFH CCP7	-86.60	-242.34	267.15	-13.39	-75.19 ± 2.36
MT mFH CCP7	-91.04	-259.06	278.28	-13.91	-85.73± 1.78

ΔE_{vdw} = van der Waals contribution, ΔE_{ele} = electrostatic energy, ΔG_{GB} = solvation free energy , ΔG_{SA} = non polar contribution to the solvation energy

4.4. Discussion

One of the salient features of borrelial spirochete is its ability to survive and infect a wide range of hosts, including human, large mammals, reptiles and birds [8]. The host association of borrelia is corelated to its ability to successfully evade host complement attack and hence causing persistence infection [8]. One of the key mechanisms *Borrelia* employs to evade immune attack from the host is by recruiting FH protein on its surface via interactions with its surface proteins [2, 3]. There are five different FH binding proteins known so far and, in this chapter, we presented molecular basis of FH recruitment by one of the key surface proteins, CspZ.

CspZ is a major FH binding surface protein that protects *Borrelia* from complement mediated attack from the host [13]. CspZ binds tightly to hFH CCP7 with nanomolar K_d [15]. The structure of the free form of CspZ has been known for more than a decade and extensive mutagenesis and computational studies have predicted specific residues in CspZ as key residues directly involved in binding to hFH CCP7 [15, 47]. Peptide mapping analysis experiments have mapped four key regions within CspZ that bind to hFH CCP7 [47]. However, we identified an additional region (D47-K73) that is directly involved in FH binding. In addition, the 16 C-terminal residues in CspZ is also shown to directly involved in FH binding [47]. As seen in the binding interface in CspZ, these residues are not making any direct contact with hFH CCP7. Instead, the loss in binding affinity observed is most likely due to the compromise in the structural integrity of the protein upon mutation. Further, extensive mutagenesis studies have been performed in almost every CspZ residues to assess the effect in binding to hFH CCP7 [45, 47]. Fourteen CspZ residues have drastically impaired FH binding and hence suggested as the residues directly interacting with FH [45, 47]. However, it is clear from our studies that out of all the residues

identified so far as important for binding, only four residues, R139, Y207, F210 and Y211, are at the binding interface. Since most of the residues studied for mutagenesis studies are distributed in several regions of CspZ, it is unlikely that any of these previously identified residues are directly involved in FH binding.

Sequence analysis and computational studies of FH from human and various animals show that unlike CspA, CspZ most likely can bind to a wider range of hosts with relatively equal affinity. Our analysis shows that CspZ binds to both hFH (K_d ~0.3 nM) and mFH (K_d ~1.9 nM) CCP7 tightly. Further, computational reciprocal mutagenesis studies show that there are only 4 residues in mFH CCP7 that increases the binding affinity to the level similar to that of hFH CCP7 when mutated to corresponding hFH CCP7 residues. As mouse is the most frequently used animal model for disease study, our results can potentially provide valuable information in improving and designing better animal models for Lyme disease study.

REFERENCES

REFERENCES

1. Ricklin, D., et al., *Complement: a key system for immune surveillance and homeostasis.* Nature Immunology, 2010. **11**(9): p. 785-797.

2. Serruto, D., et al., *Molecular mechanisms of complement evasion: learning from staphylococci and meningococci.* Nature Reviews Microbiology, 2010. **8**(6): p. 393-399.

3. Ferreira, V.P., M.K. Pangburn, and C. Cortes, *Complement control protein factor H: The good, the bad, and the inadequate.* Molecular Immunology, 2010. **47**(13): p. 2187-2197.

4. Zipfel, P.F. and C. Skerka, *Complement regulators and inhibitory proteins.* Nat Rev Immunol, 2009. **9**(10): p. 729-40.

5. Granoff, D.M., J.A. Welsch, and S. Ram, *Binding of Complement Factor H (fH) to Neisseria meningitidis Is Specific for Human fH and Inhibits Complement Activation by Rat and Rabbit Sera.* Infection and Immunity, 2009. **77**(2): p. 764-769.

6. Ngampasutadol, J., et al., *Human factor H interacts selectively with Neisseria gonorrhoeae and results in species-specific complement evasion.* Journal of Immunology, 2008. **180**(5): p. 3426-3435.

7. Lu, L., et al., *Species-Specific Interaction of Streptococcus pneumoniae with Human Complement Factor H.* Journal of Immunology, 2008. **181**(10): p. 7138-7146.

8. Kurtenbach, K., et al., *Host association of Borrelia burgdorferi sensu lato - the key role of host complement.* Trends in Microbiology, 2002. **10**(2): p. 74-79.

9. Stevenson, B., et al., *Differential binding of host complement inhibitor factor H by Borrelia burgdorferi Erp surface proteins: a possible mechanism underlying the expansive host range of Lyme disease spirochetes.* Infect Immun, 2002. **70**(2): p. 491-7.

10. Radolf, J.D., et al., *Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes.* Nature Reviews Microbiology, 2012. **10**(2): p. 87-99.

11. Mannelli, A., et al., *Ecology of Borrelia burgdorferi sensu lato in Europe: transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change.* Fems Microbiology Reviews, 2012. **36**(4): p. 837-861.

12. Kraiczy, P. and B. Stevenson, *Complement regulator-acquiring surface proteins of Borrelia burgdorferi: Structure, function and regulation of gene expression.* Ticks and Tick-Borne Diseases, 2013. **4**(1-2): p. 26-34.

13. Kraiczy, P., et al., *Borrelia burgdorferi complement regulator-acquiring surface protein 2 (CspZ) as a serological marker of human Lyme disease.* Clin Vaccine Immunol, 2008. **15**(3): p. 484-91.

14. Rogers, E.A., et al., *Comparative Analysis of the Properties and Ligand Binding Characteristics of CspZ, a Factor H Binding Protein, Derived from Borrelia burgdorferi Isolates of Human Origin.* Infection and Immunity, 2009. **77**(10): p. 4396-4405.

15. Brangulis, K., et al., *Structural characterization of CspZ, a complement regulator factor H and FHL-1 binding protein from Borrelia burgdorferi.* Febs Journal, 2014. **281**(11): p. 2613-2622.

16. Hartmann, K., et al., *Functional characterization of BbCRASP-2, a distinct outer membrane protein of Borrelia burgdorferi that binds host complement regulators factor H and FHL-1.* Mol Microbiol, 2006. **61**(5): p. 1220-36.

17. Achila, D., et al., *Structural determinants of host specificity of complement Factor H recruitment by Streptococcus pneumoniae.* Biochemical journal, 2015. **465**(2): p. 325-35.

18. Gasteiger, E., et al., *ExPASy: the proteomics server for in-depth protein knowledge and analysis.* Nucleic Acids Research, 2003. **31**(13): p. 3784-3788.

19. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods in Enzymology, 1997. **276**: p. 307-326.

20. Kabsch, W., *XDS.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 125-132.

21. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 133-144.

22. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**: p. 658-674.

23. Prosser, B.E., et al., *Structural basis for complement factor H linked age-related macular degeneration.* Journal of Experimental Medicine, 2007. **204**(10): p. 2277-83.

24. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 2126-2132.

25. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 213-221.

26. DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, Palo Alto, CA, USA: DeLano Scientific.

27. Velazquez-Campoy, A. and E. Freire, *Isothermal titration calorimetry to determine association constants for high-affinity ligands.* Nat. Protocols, 2006. **1**(1): p. 186-191.

28. Gill, S.C. and P.H. Vonhippel, *Calculation of protein extinction coefficients from amino-acid sequence data.* Analytical Biochemistry, 1989. **182**(2): p. 319-326.

29.   Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*, in *The Proteomics Protocols Handbook*, J.M. Walker, Editor. 2005, Humana Press: Totowa, NJ. p. 571-607.

30.   D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, et al., *AMBER 2016*. 2016, University of California, San Francisco.

31.   Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.

32.   Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Res, 2018. **46**(W1): p. W296-W303.

33.   Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

34.   Gordon, J.C., et al., *H++: a server for estimating pKas and adding missing hydrogens to macromolecules.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W368-71.

35.   Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

36.   Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**.

37.   Roe, D.R. and T.E. Cheatham, 3rd, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013. **9**(7): p. 3084-95.

38.   *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*

39.   Hou, T., et al., *Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations.* J Chem Inf Model, 2011. **51**(1): p. 69-82.

40.   Miller, B.R., 3rd, et al., *MMPBSA.py: An Efficient Program for End-State Free Energy Calculations.* J Chem Theory Comput, 2012. **8**(9): p. 3314-21.

41.   Lo Conte, L., C. Chothia, and J. Janin, *The atomic structure of protein-protein recognition sites.* Journal of Molecular Biology, 1999. **285**(5): p. 2177-2198.

42.   Dehouck, Y., et al., *BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations.* Nucleic Acids Research, 2013. **41**(W1): p. W333-W339.

43.   Li, M.H., et al., *MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions.* Nucleic Acids Research, 2016. **44**(W1): p. W494-W501.

44.   Moretti, R., et al., *Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions.* Proteins-Structure Function and Bioinformatics, 2013. **81**(11): p. 1980-1987.

45.   Hartmann, K., et al., *Functional characterization of BbCRASP-2, a distinct outer membrane protein of Borrelia burgdorferi that binds host complement regulators factor H and FHL-1.* Molecular Microbiology, 2006. **61**(5): p. 1220-1236.

46.   Rogers, E.A. and R.T. Marconi, *Delineation of species-specific binding properties of the CspZ protein (BBH06) of Lyme disease spirochetes: Evidence for new contributions to the pathogenesis of Borrelia spp.* Infection and Immunity, 2007. **75**(11): p. 5272-5281.

47.   Siegel, C., et al., *Deciphering the Ligand-binding Sites in the Borrelia burgdorferi Complement Regulator-acquiring Surface Protein 2 Required for Interactions with the Human Immune Regulators Factor H and Factor H-like Protein 1.* Journal of Biological Chemistry, 2008. **283**(50): p. 34855-34863.

CHAPTER 5

Molecular Dynamics (MD) Analysis of FH Recruitment by Outer Surface Protein ErpA of
Borrelia burgdorferi

5.0. Abstract

One of the key mechanisms attributed to persistent infection from Lyme disease pathogen, *Borrelia burgdorferi*, is its ability to successfully evade complement mediated killing by recruiting complement factor H (FH) proteins from the host. The Erp family of outer surface proteins expressed by *Borrelia* are an important protein family that aid in complement evasion by recruiting FH protein to borrelial surface. However, largely due to the lack of structural and biophysical understanding of the protein-protein interactions between Erp family proteins and FH, biological relevance and other functions of many Erp proteins are still unknown. Here, we present the crystal structure and molecular dynamics (MD) simulations studies of the protein complex between ErpA and hFH CCP20. The structure of the complex and the analysis presented in this chapter further aid in deciphering the role of FH recruitment in immune evasion by *Borrelia*. Such information could be beneficial in exploring and discovering Erp protein-based vaccine candidate against Lyme disease.

5.1. Introduction

Despite facing a very sophisticated and highly regulated immune surveillance system in many hosts, a wide range of pathogens can cause persistent infection [1]. One of the key mechanisms commonly used by pathogenic microbes to cause infection is by evading complement mediated immune attack from the host [1, 2]. While complement system serves as the first line of defense against any microbial intruders, one of the key complement activation pathways, the alternative pathway, is always active at low level and ready to attack all cells in non-selective way [3]. Hence, the activation of complement can act like a double-edged sword: while its activation protects against microbial pathogens, it can also kill self-cells causing a serious case of autoimmunity [3, 4]. In order to avoid such a disaster, the host immune system has many immune regulator proteins circulating in plasma that selectively bind to self-cells and inhibit complement activation [5, 6]. Complement factor H (FH) protein is one of the key immune regulator proteins involved in inhibition of complement mediated attack on self-cells [7].

FH is a large (155k-Da) complement regulator proteins, consisting of 20 domains called complement control module (CCP). It inhibits both complement activation and amplification in self-cell by inhibiting and promoting dissociation of many key protein assemblies in complement activation pathways [6]. Taking advantage of this key function of FH, a wide range of pathogens have evolved with ways to 'hijack' FH to their own cell surface so that they are protected against complement mediated attack, in a similar way self-cells are protected against complement attack by FH [6, 8]. The main causative bacteria of Lyme disease, *Borrelia burgdorferi*, expresses many FH binding proteins and recruits FH to its own surface as a key immune evasion mechanism.

Lyme disease is one of the most common vector-borne illness in the United States and Europe [9]. Three genospecies of *B. burgdorferi* sensu lato group, *B. burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii*, are the causative agents for most of the Lyme disease cases around the globe [10]. One of the salient features of *Borrelia* is its ability to survive and cause persistent infection in a wide range of hosts, including human, other large and small mammals, birds and even reptiles [9, 10]. The wide range of host association of *Borrelia* is corelated with its ability to avoid complement mediated attack from the host via FH recruitment to their own surface [11]. However, deciphering the role of FH recruitment in pathogenesis and host specificity of borrelia is complicated due to many factors. The key factors hindering advancement of Lyme disease study include complex borrelial genomes consisting of many genospecies, ability of *Borrelia* to express multiple FH binding proteins, and non-equilibrium methods used in binding studies to understand such mechanisms are some.

In this chapter, we present the crystal structure of the complex between one of the key FH binding surface proteins, ErpA, and human FH (hFH) CCP 20. Further, we have analyzed and identified 'hot-spot' residues at the ErpA and hFH CCP20 interface that are important for formation of the complex. Information presented in this chapter provides new insights into FH recruitment mechanism by yet another borrelia protein.

5.2. Methods

5.2.1. Molecular Dynamics Simulation and MM-GBSA Analysis

All MD simulations were performed using AMBER 16 Molecular Dynamics package and the ff14SB force field was used to describe the proteins [12, 13]. All mutant proteins for *in-silico* mutagenesis study were generated with assumption that the structural integrity of the proteins is not affected by the single amino acid substitution. Before running the simulation, all protein complexes were processed with the H $^{++}$ program, which computes the pK_a values of ionizable groups in protein and adds missing hydrogen atoms according to specified pH of the environment [14]. Special attention was given to the protonation state of histidine suggested by this program and adjustment was made accordingly.

Original water molecules from the crystal structure were removed and hydrogen atoms were added using the TLEAP module of AMBER 16 [12]. The structure was solvated in a rectangular TIP3P water box [12], with the boundaries of at least 12 Å away from the protein atoms. Parameters files for the MD simulations, including coordinate and topology files, were then generated. The solvated structure of the complex was energy minimized using steepest descent and conjugated gradient methods [12], each for 5000 steps, first with positional restraint force constant of 100 kcal/ (mol. $Å^2$) on all heavy atoms of the protein and then without any positional restrain on the whole system. The minimized system was heated linearly to 300K in 120 ps under constant volume conditions (NVT) with no position restraints. The system was further simulated for 750 ns under constant pressure and temperature conditions (NPT). Temperature was controlled using Langevin dynamics with a collision frequency of 1 ps^{-1} during heating, equilibration and production steps [12, 15, 16]. All covalent bonds to hydrogen atoms

were constrained by the SHAKE method, with numerical integration time step of 2 fs [15, 16]. Long range electrostatic interactions were computed using Particle Mesh Ewald (PME) method with a cutoff of distance of 10 Å [17].

The simulation results were analyzed using the CPPTRAJ module of AMBER 16 and the PyMOL program [18-20]. Root mean square deviation (RMSD) was monitored throughout the simulation to ensure equilibration of the system. Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) and energy decomposition analyses were performed using MMPBSA.py python script incorporated in AmbertTools 16 [12, 21]. A generalized Born implicit solvent model with 0.15 M salt concentration was used for all the MM/GBSA analysis. The results from MM/GBSA run were calculated using 500 snapshots for each simulation and were averaged to estimate global free energy of binding between CspZ and hFH CCP20. The average from all the snapshots were calculated to estimate the interaction energies of the residues, particularly focusing on residues at the ErpA and hFH CCP20 interface. Since we were only interested in relative mutational effects on binding, the entropic contribution to the free energy change was not included in the calculation. Further, entropic calculations using currently available methods are computationally very expensive and tend to have large margin of errors, introducing significant uncertainty in the result.

After successful MD simulation run, criteria such as density, temperature, pressure, root mean square deviation (RMSD), root mean square fluctuation (RMSF), KE, PE and total energy were analyzed to confirm system convergence to the equilibrium. CPPTRAJ module from AMBER16 was used to calculate backbone RMSD values to monitor the stability of the protein complex during the entire simulation [18]. Root mean square fluctuation (RMSF) was calculated

to measure fluctuation of individual residues during simulation. Similarly, CPPTRAJ module was used to analyze various hydrogen bond properties, such as hydrogen bond distance and occupancies of each hydrogen bond interactions between residues at the dimer interface [18]. Distance cutoff of 3.3 Å (between acceptor and donor heavy atoms) and angle cutoff of 120° (between donor heavy atom-hydrogen-acceptor heavy atom) were to track all hydrogen bonds. Special attention was given to the stable hydrogen bonds, with occurrence of at least 50%.

5.2.2. Computational Alanine Scanning Mutagenesis

Alanine mutagenesis was also performed using the MD simulation data for ErpA:hFH CCP20. *In-silico* alanine scanning mutagenesis is a quick way to calculate the binding energy contribution of residues at protein-protein interaction interfaces. Because of the size, main chain conformation and electrostatic properties of alanine, it is usually the best choice for mutagenesis study.

5.3. Results

5.3.1. Structure of ErpA:hFH CCP20 Complex

The structure of bound form ErpA protein has typical characteristics of other Erp proteins with known structures [22]. It has a rigid fold of eight β-strands and three short α helices connected in the order β1-β2-β3-β4-αi-β5-αii-β6-β7-β8-αiii. β-strands are arranged in antiparallel fashion with β1 and β4 crossing each other almost perpendicularly. Highly twisted β2, and β6 do not form typical β- barrel structure. β1 and β2 are the longest β-sheets with ten and thirteen residues respectively. The shortest β-sheets, β5 and β8 contain eight and seven residues respectively. Based on computational analysis of the Erp family proteins, previous studies suggested the presence of coiled-coil structural element in the Erp family proteins [23]. However, the structure of ErpA protein presented here shows that there are no coiled-coil elements present in ErpA proteins as previously suggested.
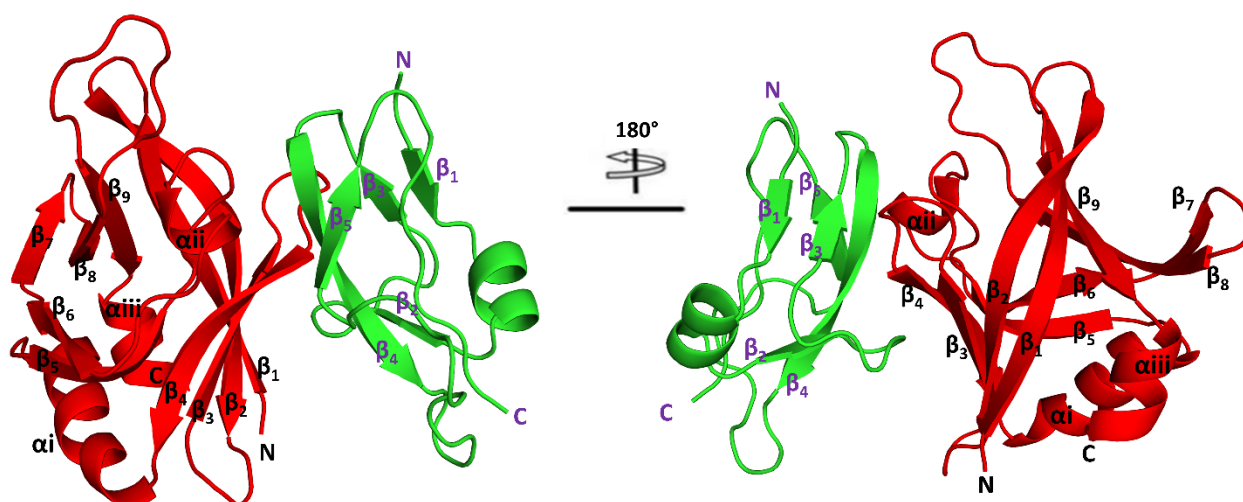

Figure 5.1. (Left) Cartoon representation of the structure of the complex ErpA and hFH CCP20. ErpA protein is colored in red and hFH CCP20 is colored in green. (Right) Alternative view of the protein complex with 180° rotation along the vertical plane.

The complex between ErpA and hFH CCP 20 is stabilized by hydrogen bonds as well non-polar interactions. Out of 19 and 16 interface residues from ErpA and hFH CCP 20, respectively, six residues from each protein are involved in a network of hydrogen bonds at the interface. All residues involved in hydrogen bonds in ErpA are within β2, β3 and β4, including two residues, G81 and H82, from the loop connecting β3 and β4.
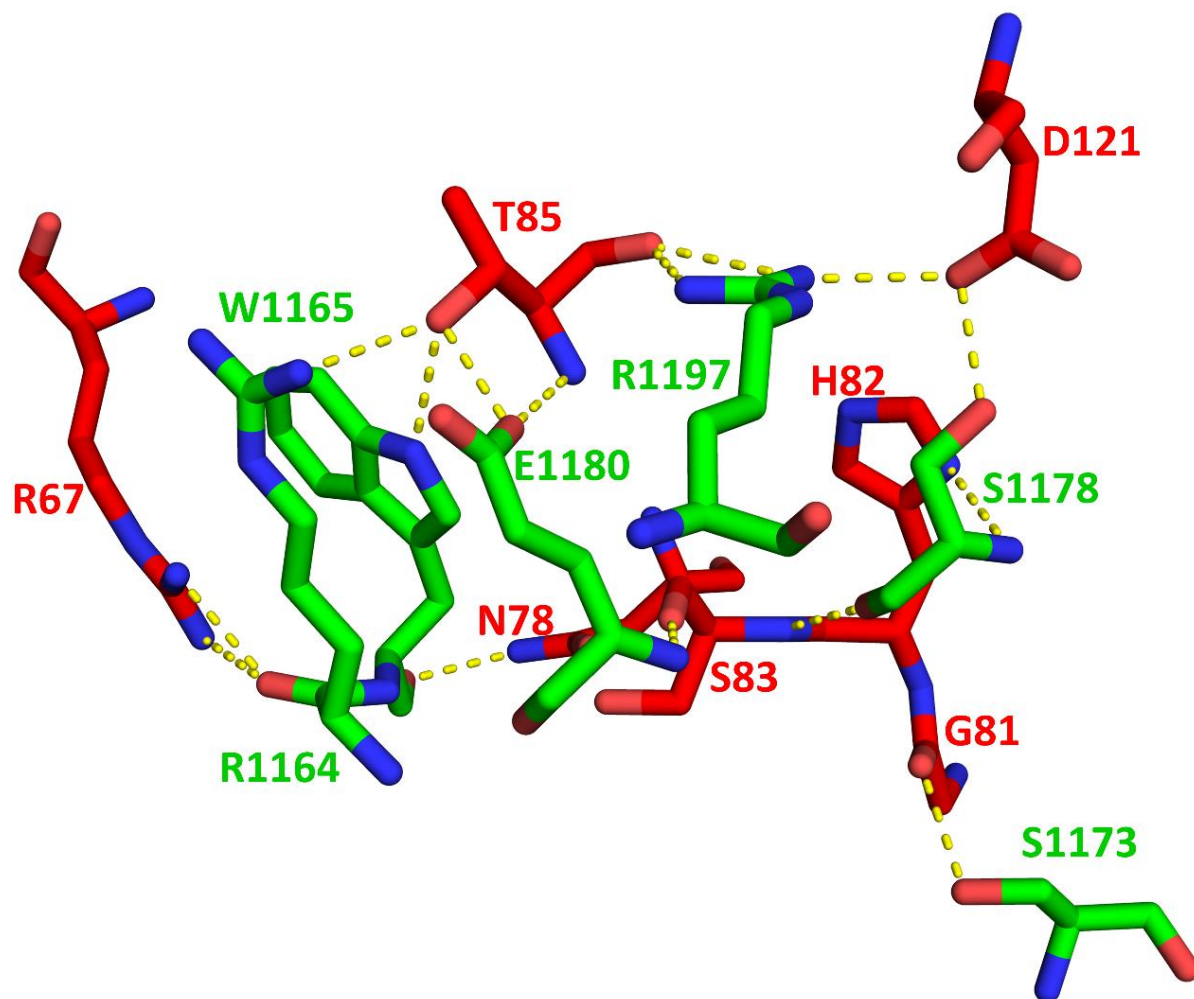

Figure 5.2. Hydrogen bonding interactions at the ErpA:hFH CCP 20 interface. Residues from ErpA and hFH CCP20 are colored in red and green, respectively. Hydrogen bonds between two residues are indicated by yellow dashed line.

Similarly, six residues from hFH CCP 20 that form hydrogen bonds with ErpA residues are distributed over *β3, *β4, *β5 and loop region connecting *β2 and *β3 of hFH CCP 20. ErpA proteins are attached to the outer membrane of *Borrelia* via a highly flexible N-terminal region. This flexibility on the N-terminus is thought to allow free movement of Erp proteins about the lipid anchor [24], although the anchoring mechanism is not fully understood

Based on the structure of the complex, G81, H82, S83, T85, R67, N78 and D79 from ErpA and R1164, W1165, S1173, S1178, R1197 and E1180 from hFH CCP20 are involved in hydrogen bonding interactions at the interface. Interestingly, ErpA residue T85 is involved in six hydrogen bonding interactions and most likely one of the key stabilizing interface residues. T85 alone is interacting with four residues, R1164, W1165, E1180 and R1197, from hFH CCP20. Similarly, W1165 from hFH CCP 20 not only makes hydrogen bond involving its side chain, it also forms a big hydrophobic pocket towards the edge of the binding interface, completely burying its hydrophobic ring at the interface. Hydrophobic and Van der Waals interactions along with hydrogen bond interaction from W1165 greatly stabilizes the interface. Salt bridge interactions between E74-R1164 further stabilizes the interaction between ErpA and hFH CCP20.

5.3.2. Structural Analysis of ErpA:hFH CCP20 Using MD Simulation

Based on the structural analysis, key hydrogen bonds are identified above. However, not all hydrogen bonds contribute equally in stabilizing the complex. Furthermore, there are other interactions that may stabilize the protein complex. In order to further characterize the intermolecular interactions, we performed MD simulation of the ErpA:hFH CCP20 complex. CPPTRAJ module provided by Amber16 was used to evaluate RMSD and RMSF of the MD simulation of the complex of ErpA and hFH CCP 20. The backbone RMSD was first calculated for

the individual proteins of the complex by fitting individual proteins. The RMSDs of the individual

proteins were low, with an average RMSD of 0.94 ± 0.12 Å for hFH CCP 20 and 1.38 ± 0.22 Å for

ErpA. The RMSD of the whole complex is 1.51 ± 0.23 Å. Based on the RMSD analysis, the complex

of ErpA and hFH CCP20 stayed stable conformation upon complex formation. RMSD plot shows

good convergence and stability of the system, and the coordinates and trajectory files from the

MD simulation are used for further analysis of the complex.
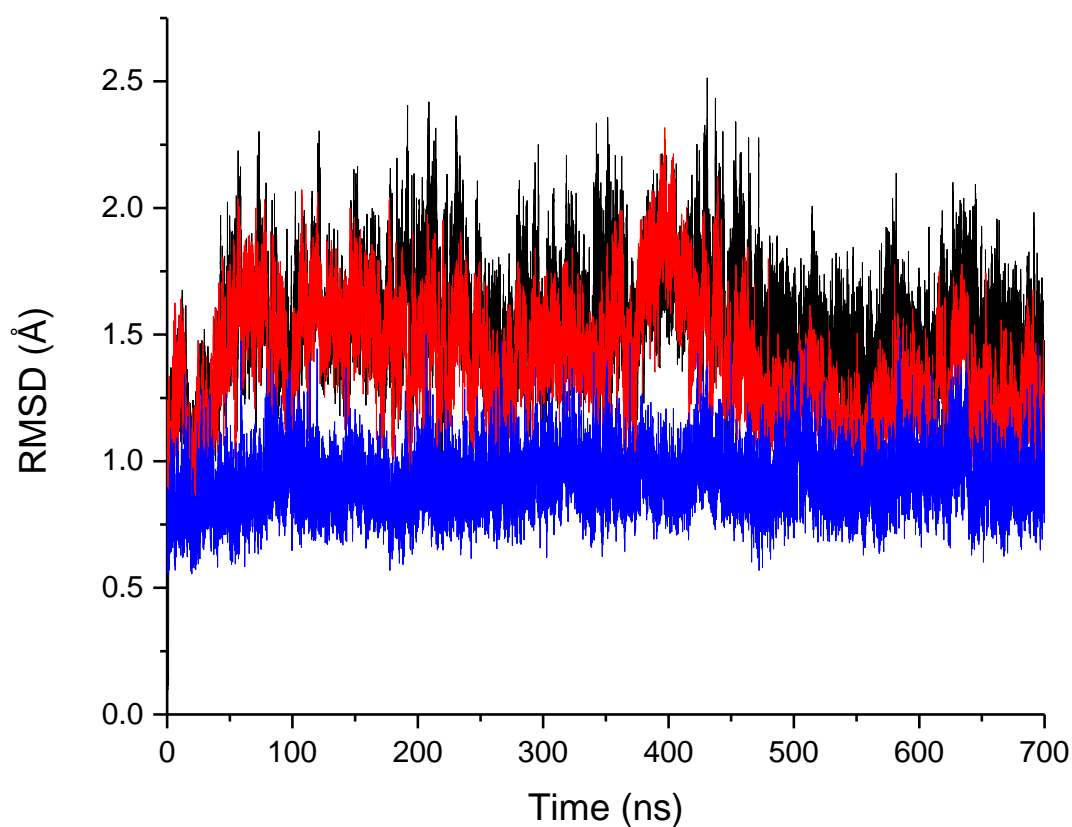
Figure 5.3. Time evolution of the backbone RMSD for the complex of ErpA and hFH CCP20. The blue plot represents the RMSD change of hFH CCP20 and the red plot the RMSD change of ErpA. The black plot represents the RMSD change of the complex of ErpA and hFH CCP 20.

While there is no structure of the free form ErpA available, the structure of the free form

of other closely related Erp proteins (OspE family proteins) are currently available in PDB. The

structure of the free form of OspE as well as the structure of the complex between hFH CCP19-

20 and OspE protein was recently deposited in PDB with PDB IDs 2M4F and 4J38 respectively
[22]. The structure of the free form CCP19-20 is also available with PDB ID 2G7I [25]. Since there
is more than 85% of sequence similarity between OspE and ErpA, a high degree of structural
similarity between these two proteins is expected.

In order to see if there is any conformational change in any part of the ErpA during the
complex formation, the free form of OspE was aligned with the bound form ErpA and the free
form of hFH CCP 20 was aligned with bound hFH CCP 20 using PyMOL [19, 20]. The RMSD for the
structure alignment was 0.66 Å and 0.73 Å for hFH CCP20 and ErpA, respectively. The structural
alignment shows no significant change upon complex formation, especially at the binding
interface. Most of the conformation change is limited to the flexible loops.
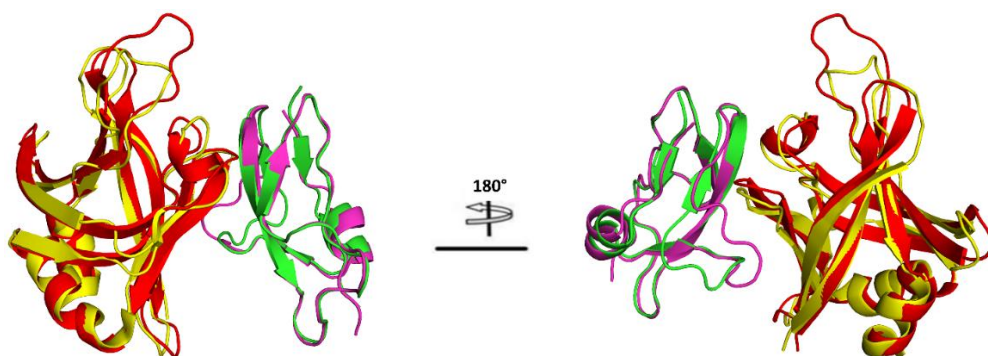
Figure 5.4. Structural comparison between the free form of OspE (yellow, PDB ID: 2M4F) and free form of hFH CCP 20 (magenta, PDB ID: 2G7I) with the bound hFH CCP 20 (green) and ErpA (red). The alternative view with 180° rotation along the horizontal plane is shown on the right. The RMSD for alignment was low with 0.66 Å for FH and 0.73 Å for ErpA.

Further, the OspE:hFH CCP19-20 was also superimposed with ErpA:hFH CCP20 protein
complex for structural comparisons between bound form of these two protein complexes. The
RMSD for the alignment of these two complexes was 0.75 Å, which is almost identical to ErpA-
OspE alignment as described earlier. This indicates that there is very little conformational change

upon complex formation in ErpA and FH CCP20 and most of the small conformational changes are confined within the flexible loop regions.
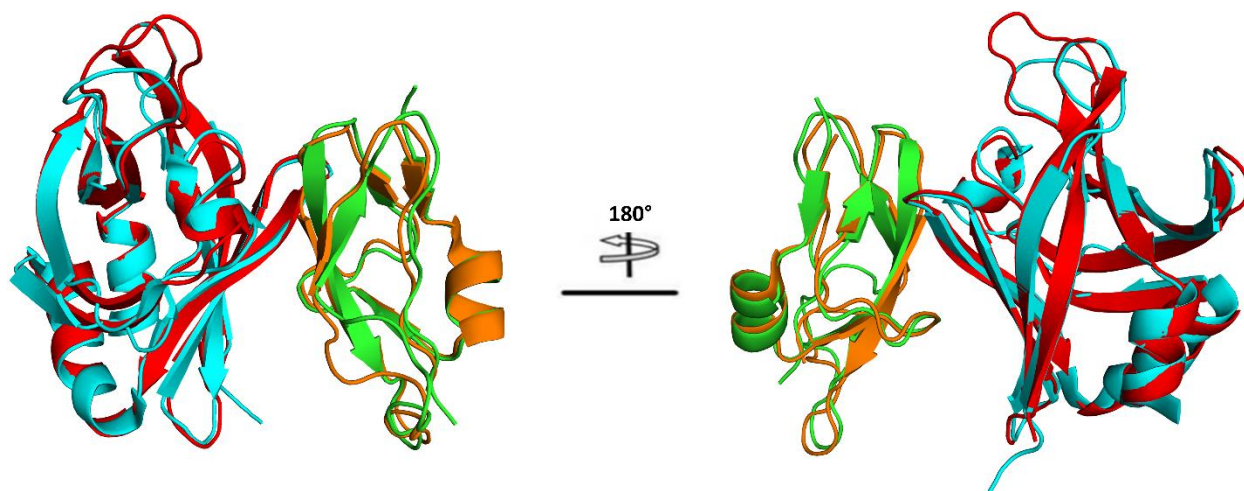

Figure 5.5. Structural alignment between OspE:hFH CCP19-20 (PDB ID: 4J38) and ErpA:hFH CCP20 protein complexes. OspE and hFH CCP20 are colored cyan and orange, respectively whereas ErpA in complex with hFH CCP20 are colored in red and green, respectively. The alternative view with 180° rotation along the horizontal plane is shown on the right. The RMSD for alignment was low with 0.75 Å.

The RMSF plot also shows stable residue fluctuations upon complex formation except for the loop regions in both ErpA and hFH CCP20. The RMSF of all other regions stayed below 0.5 Å, with an average RMSF of 0.87 ± 0.55 Å. The high standard deviation associated with RMSF is due to the large variations in RMSF in the loop regions. While the RMSF for all the loop regions stayed below 2 Å, the longest loop region within ErpA comprising 12-residue span from D148 to I159 showed the maximum fluctuation with RMSF of ~ 4.96 Å.
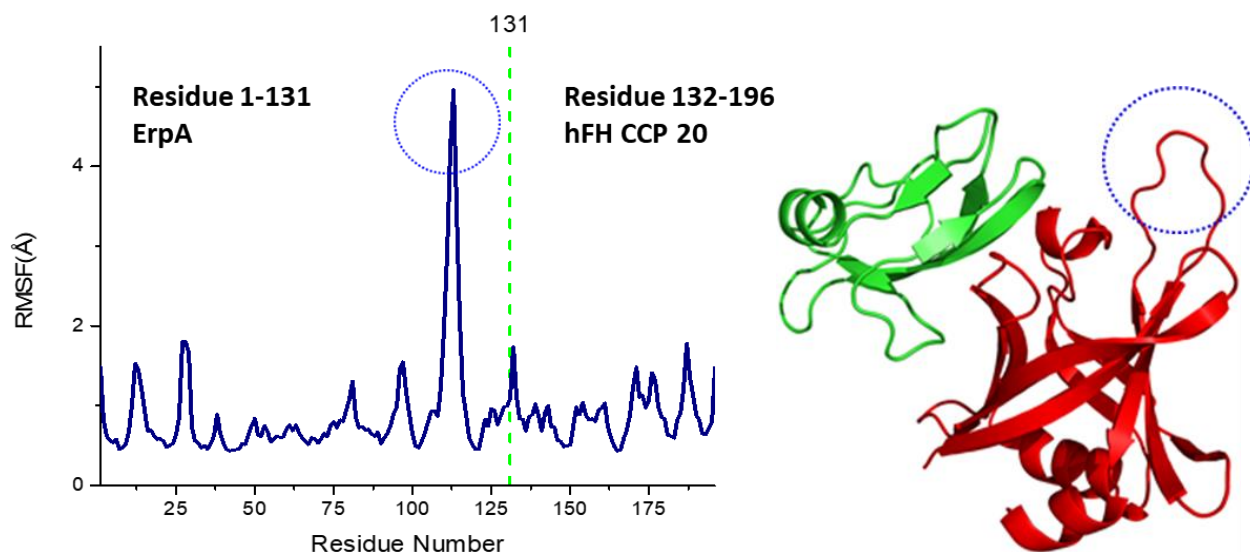
Figure 5.6. (Left) Time evolution of backbone RMSF during MD simulation for the complex of ErpA and hFH CCP20. The RMSF of loop regions are the highest while all RMSF for all other regions stayed below 0.5 Å with an overall average RMSF of 0.87 ± 0.55 Å. The high standard deviation associated with RMSF is due to the large variation in RMSF due to loop regions. The highest fluctuation indicated by dashed circle in RMSF plot corresponds to the long loop region in ErpA spanning residues D148 to I159.

After verifying good convergence of the simulation system, stability of all the hydrogen bonds at the interface was analyzed by calculating percentage occurrence of each hydrogen bond. Hydrogen bonds with occurrence of at least 50% were considered present and are listed in table 5.2. Among the hydrogen bonds, the ones between S1178-S83, W1165-N78 and E1180-S83 are the most stable and contribute the most in complex stabilization. Most of the hydrogen bonds with less than 70% occurrence show large fluctuation in acceptor-donor distance over the time as indicated by the large standard deviations associated with the distance change between hydrogen bond donors and acceptors.

Table 5.1. Hydrogen bond occupancy of residues at the ErpA:hFH CCP20 interface

hFH20...........ErpA	Occurrence (%)	Distance (Å)	Crystal Structure (Å)
S1178-O........N-S83	98.75	2.92 ± 0.13	3.0
W1165-O......ND2-N78	97.00	2.92±0.18	2.9
E1180-N........O-S83	86.12	2.96±0.14	2.9
S1173-OG.....O-G81	81.94	3.02±0.56	2.8
S1178-N........ND1-H82	72.66	3.24±0.41	3.2
E1180-OE1....OG1-T85	68.95	2.98±0.38	2.6
R1197-NH1...O-T85	67.97	3.63±1.28	2.8
E1180-OE2....OG1-T85	64.91	3.02±0.39	2.6
R1164-O........NH1-R67	63.14	3.88±1.72	2.7
E1180-OE1....N-T85	55.99	3.69±0.96	2.9
E1180-OE2....NH2-T85	51.53	3.78±0.98	4.1

5.3.3. Identification of Key Interface Residues in ErpA:hFH CCP20 Complex

To further identify interface residues important for binding, the binding energy contributions for residues from ErpA and hFH CCP20 were calculated using MM/GBSA module from the AMBER program. A generalized Born implicit solvent model with 0.15 M salt concentration was used for all the MM/GBSA analysis. The results from MM/GBSA run were calculated using 500 snapshots for each simulation and were averaged to estimate global free energy of binding between ErpA and hFH CCP20. The average from all the snapshots were calculated to estimate the interaction energies of the residues, particularly focusing on interface residues.
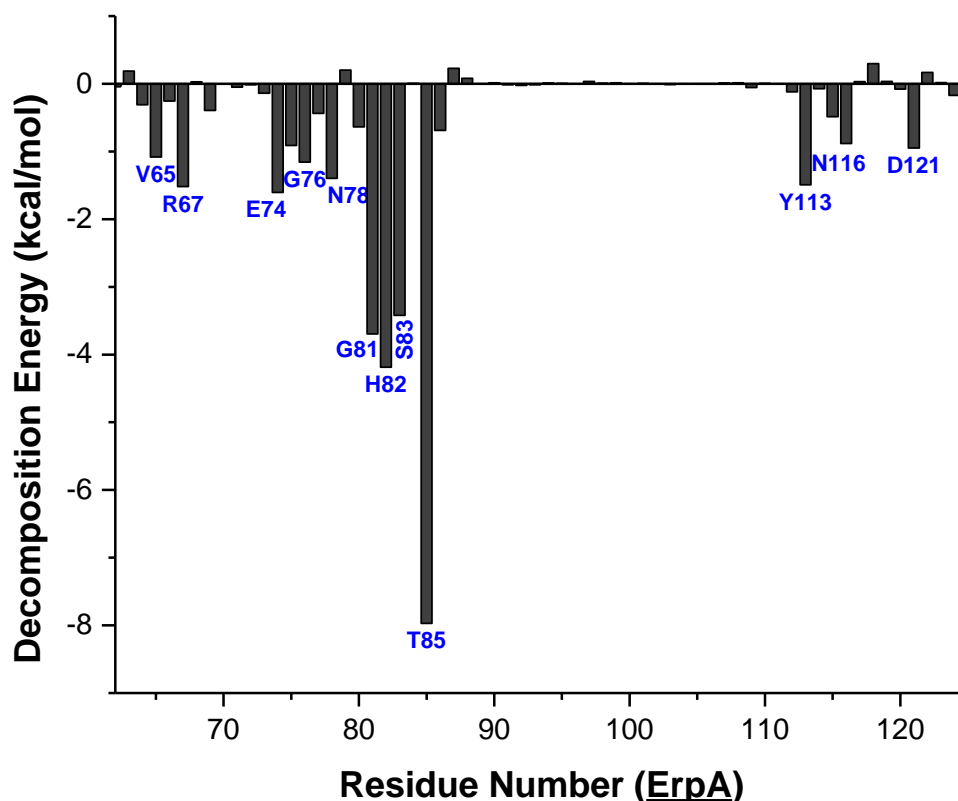
Figure 5.7. Energy decomposition calculation of ErpA residues.

Based on the binding energy calculation, four residues, G81, H82, S83 and T85 of ErpA are important for binding, with ~4 kcal/mol energy contribution from G81, H82 and S83 and ~ 8 kcal/mol of binding energy contribution from T85 alone. Based on the hydrogen bond analysis from the MD simulation, T85 can form five hydrogen bonds, each with hFH CCP20 residues W1165, R1164 and R1197 and two with E1180. Backbone nitrogen and oxygen of T85 is involved in hydrogen bonding interactions with E1180 and R1197. The structural analysis shows an additional hydrogen bonding interaction between T85 and R1197 with distance of 3.35 Å between donor and acceptor atom. However, this hydrogen bond is excluded from our analysis as the distance cutoff of 3.30 Å is used for the MD simulation analysis. Nevertheless, this

additional and relatively weaker hydrogen bond also contributes to the complex formation. G81, H82 and S83 each form one hydrogen bond at the interface. Although R67 from ErpA forms hydrogen bond, the energy contribution from this residue is relatively low with ~ 1.5 kcal/mol of energy. Most likely, since R67 is making a hydrogen bond with backbone oxygen atom of R1164 that is in a highly flexible long loop of FH CCP20, this interaction weakens over time and hence its relative contribution to the overall stability of the complex is low. The low hydrogen bond occurrence, 63 %, for this hydrogen bond further supports the low binding energy contribution from R67.

Interestingly, based on the crystal structure of the complex, D121 of ErpA can form a hydrogen bond with S1178 and a salt bridge with R1197 of FH CCP20. While a large binding energy contribution was expected from D121, the energy decomposition analysis shows less than 1 kcal/mol of binding energy contribution from D121. To investigate this further, the initial crystal structure was aligned with the representative PDB extracted at 700 ns simulation time. D121 is in the loop region between β5 and β6 in ErpA and structural alignment shows that there is a large shift in the position of D121 during the simulation (denoted by D121*). This causes a large change in distance between atoms of D121 that can potentially contribute to hydrogen bonding interactions and the atoms in hFH CCP20. Due to such large fluctuation with D121, the distance between D121-OD1 and OG1-S1178 and D121-OD1-NH2R1197 increases by 2.2 Å. Similarly, the distance between D121-O and NE-R1197 increases by 3.4 Å. All the potential hydrogen bonding and salt bridge interactions with D121 are weakened and hence a very little contribution from this residue at the interface is observed. All the hydrogen bond interactions associated with D121

showed less than 10 % of occurrence, further supporting the idea that these interactions do not contribute much to the stabilization of the complex.
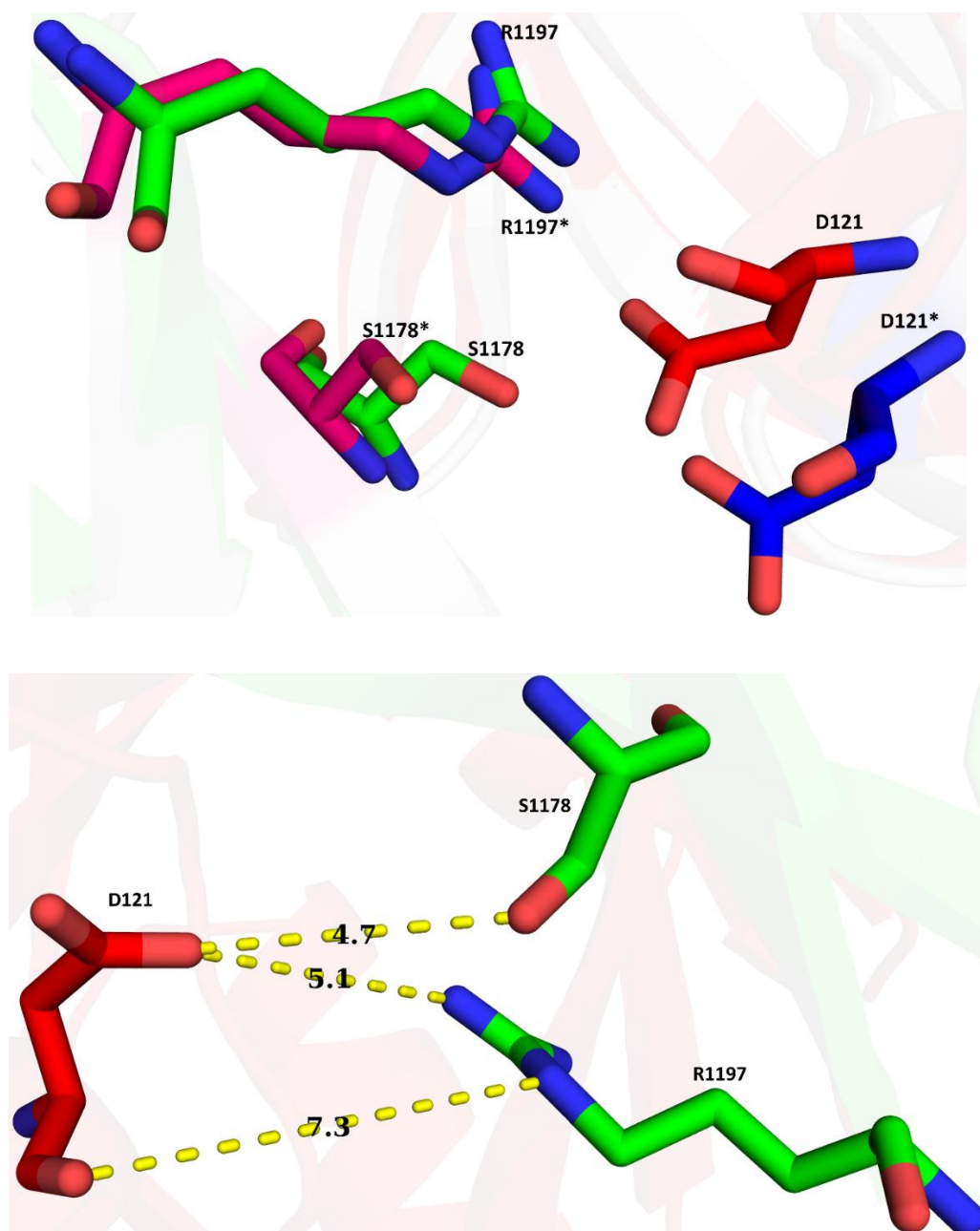

Figure 5.8. (Top) Structural alignment of ErpA:hFH CCP20 crystal structure with the PDB structure from a single snapshot at 700 ns. (Bottom) During simulation, there is large increase in distance between S1178 and R1197 with D112 that weakens important hydrogen bond interactions between ErpA and hFH CCP20.

Sequence alignment of many FH-binding Erp family proteins among different *B. burgdorferi* strains and genospecies as well as OspE from *B. garinii* and *B. afzelii* shows a good conservation of residues that make hydrogen bonds with FH. R67 and N78 are invariantly conserved. T85 is conserved in all species except for *B. afzelii*. H82 is almost conserved with some conservative replacement with S82Y in some proteins.

Figure 5.9. Sequence alignment of the FH binding regions of ErpA, ErpP, ErpC and other Erp paralog proteins encoded by different strains of *B. burgdorferi*. Erp family protein, OspE, from *B. garinii* and *B. afzelii* are also compared.

Similarly, binding energy analysis of residues from hFH CCP20 showed three residues, R1164, W1165 and E1180, as the key residues at the interface, with more than 4 kcal/mol of energy contribution from R1164 and E1180 and more than 11 kcal/mol of energy contribution from W11659 (Fig. 5.10). Not only side chain of W1165 is making a stabilizing hydrogen bond with T85 of ErpA but also its large hydrophobic aromatic ring is tightly packed, with 100 % of its accessible surface area deeply buried at the center of the interface, further stabilizing the complex by hydrophobic interactions. R1164 makes a hydrogen bond with T85 and its long non-polar part of the side chain is in close contact with W1165 aiding in complex stabilization. E1180 makes two hydrogen bonds with T85 contributing ~ 6 kcal/mol to the binding energy.
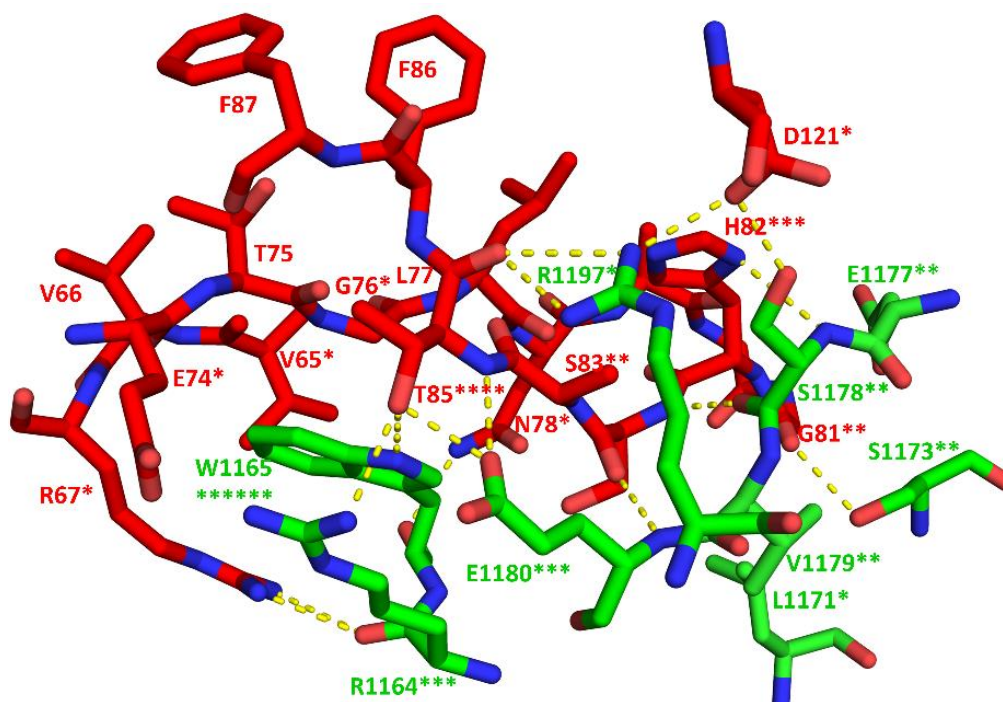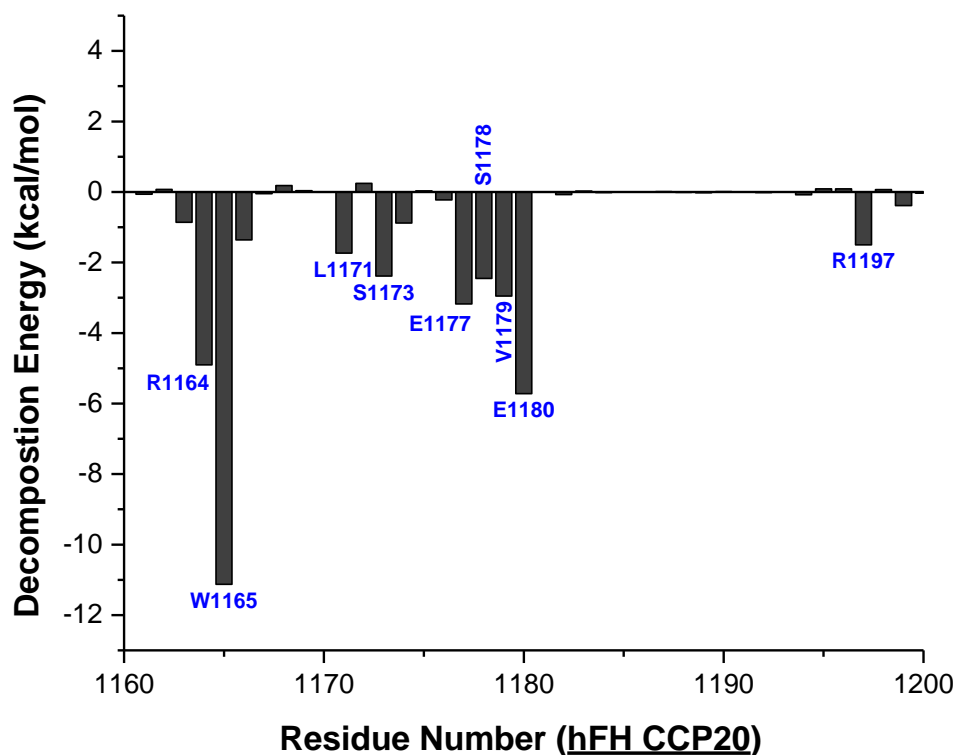
Figure 5.10. (Top) Energy decomposition analysis of hFH CCP20 residues. (Bottom) Stick representations of the specific interactions between ErpA and hFH CCP20. ErpA residues are colored and labelled in red whereas hFH CCP20 residues are colored and labelled in green. Each asterisk (*) sign represents about 2 kcal/mol of binding energy contribution. All hydrogen bond interactions between ErpA and hFH CCP20 residues are shown by yellow dash.

Interface residues from ErpA and hFH CCP20 contributing the highest binding energy at the interface are mostly buried (Fig. 5.11). The big hydrophobic aromatic chain of W1165 is completely buried which is consistent with the high stabilization of W1165 at the interface.
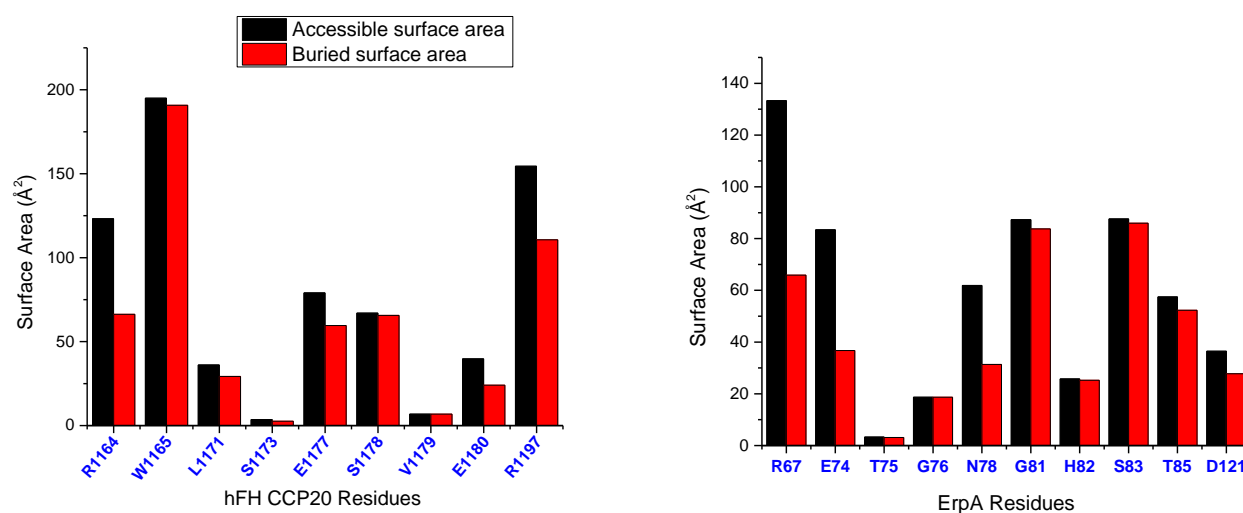

Figure 5.11. Comparison of accessible and buried surface area of the important interface residues in hFH CCP20 (left) and ErpA (right) upon complex formation.

5.3.4. Analysis of Binding of ErpA and Mouse FH (mFH) CCP20

One of the salient features of borrelial species is their ability to survive and infect multiple hosts [11]. Deciphering the role of FH binding ability of *Borrelia* in host specificity and pathogenesis is further complicated by the presence of multiple FH-binding surface proteins within a single borrelial species. In this study, we have presented yet another structure and characterizations of the complex between hFH and borrelial surface protein, ErpA. To understand the host specificity nature of *Borrelia*, such studies need to be expanded to a wide range of hosts.

Interactions between mouse FH (mFH) CCP20 and ErpA protein was also studied using ITC. The K_d of binding between mFH CCP20 and ErpA was ~ 398 nM compare to ~ 4 nM for hFH CCP20. Based on mFH studies with CspA, CspZ and ErpA proteins, the binding of mFH with ErpA showed intermediate level of affinity, with very low affinity to CspA and very high affinity to CspZ.

Table 5.2. Thermodynamics binding parameters of binding of ErpA with hFH and mFH CCP20

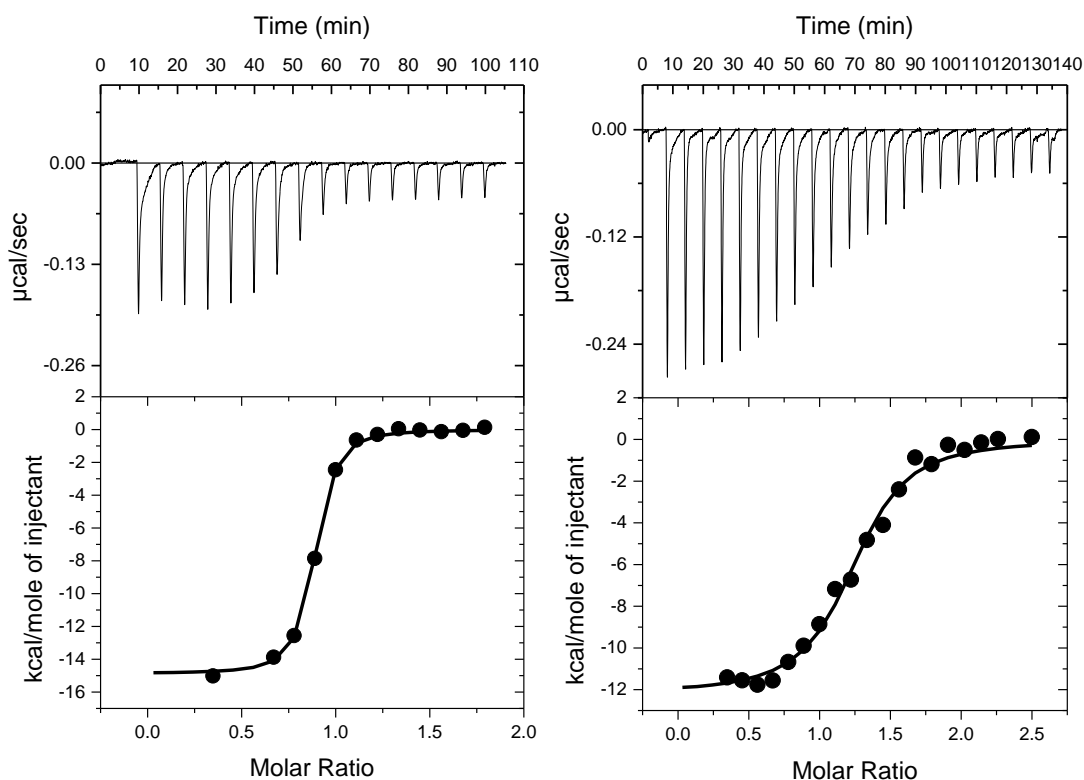|	K_d **(nM)**	**ΔG (kcal/mol)**	**ΔH (kcal/mol)**	**TΔS (kcal/mol)**
hFH CCP 20	4.0±0.7	−11.45±0.16	−9.05±0.3	2.4±0.19
mFH CCP 20	398±15.5	-8.74±0.08	−14.7±0.5	−5.95±0.6

Figure 5.12. ITC isotherms obtained from binding studies of ErpA and hFH CCP20 (left) and ErpA and mFH CCP20

172

5.4. Discussion

This study presents the three-dimensional structure of the complex between hFH and one of the surface proteins of Borrelia, ErpA. Although previous studies suggest that expression of Erp protein alone on bacterial surface does not confer serum resistance against borrelial species, structural information of Erp proteins, whether the free or the bound form can potentially be useful to explore other functions of Erp proteins that are currently unknown. With the help of the structure of the complex between ErpA and FH CCP20 along with computational analyses presented in this chapter, we have successfully identified residues at the binding interface that may play an important role in stabilizing the protein complex. W1165 from hFH CCP20 and T85 from ErpA stabilize the ErpA:hFH CCP20 complex with the highest binding energy associated with them. Out of more than 19 interface residues in ErpA, four residues, G81, H82, S83 and T85 are most important for stabilizing the protein complex with FH. Similarly, out of 16 interface residues in hFH CCP 20, R1164, W1165, S1173, E1177, S1178 and E1180 are crucial for binding. Hence, we have narrowed down to few residues that are key determinants in binding of ErpA and FH. Further, based on the structure of ErpA protein presented here, the previous hypothesis that Erp proteins contain coiled coil structural element is not correct.

While the functions or Erp proteins are largely unknown at this point, our studies could provide some insights into the functions of not only ErpA protein but also the functions of other Erp family proteins. Due to high degree of sequence similarities among different Erp paralog proteins among different genospecies of borrelia, deciphering structure-function relationship among Erp proteins could be more efficient based on the information provided in this chapter.

Studies have shown that enough response from humoral immune system can provide protection against Lyme disease [26]. The entire repertoire of Erp proteins are produced by *Borrelia* during the mammalian infection, inducing wide range antibody response against these proteins during mammalian infection [24, 27-31]. Although, another borrelial outer surface protein, OspA, based vaccine was a failure due to many side effects [32], there have been significant progress in identifying FH binding proteins as potential vaccine candidates against infectious diseases. FH-binding protein from *N. meningitidis* serotype B has recently shown promising results in clinical trials in safeguarding against meningococcal disease [33]. Hence, similar FH-binding protein-based vaccine against Lyme disease would be highly desirable.

Surface exposure, high sequence similarities between different strains and genospecies, production during mammalian infection and ability to raise high immune response in human are some of the key features of the suitable vaccine candidate and Erp family proteins demonstrate all of these characteristics. We have successful identified key residues required for ErpA binding which are highly conserved among most species and strains of Borrelia. This may be valuable for using ErpA and other Erp family proteins as potential vaccine candidate for Lyme disease.

REFERENCES

REFERENCES

1. Ricklin, D., et al., *Complement: a key system for immune surveillance and homeostasis.* Nature Immunology, 2010. **11**(9): p. 785-797.

2. Lambris, J.D., D. Ricklin, and B.V. Geisbrecht, *Complement evasion by human pathogens.* Nat Rev Microbiol, 2008. **6**(2): p. 132-42.

3. Harboe, M. and T.E. Mollnes, *The alternative complement pathway revisited.* Journal of Cellular and Molecular Medicine, 2008. **12**(4): p. 1074-1084.

4. Schmidt, C.Q., et al., *Translational mini-review series on complement factor H: structural and functional correlations for factor H.* Clin Exp Immunol, 2008. **151**(1): p. 14-24.

5. Kraiczy, P. and R. Wurzner, *Complement escape of human pathogenic bacteria by acquisition of complement regulators.* Mol Immunol, 2006. **43**(1-2): p. 31-44.

6. Makou, E., A.P. Herbert, and P.N. Barlow, *Functional Anatomy of Complement Factor H.* Biochemistry, 2013. **52**(23): p. 3949-3962.

7. Parente, R., et al., *Complement factor H in host defense and immune evasion.* Cellular and Molecular Life Sciences, 2017. **74**(9): p. 1605-1624.

8. Makou, E., A.P. Herbert, and P.N. Barlow, *Creating functional sophistication from simple protein building blocks, exemplified by factor H and the regulators of complement activation.* Biochem Soc Trans, 2015. **43**(5): p. 812-8.

9. Radolf, J.D., et al., *Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes.* Nature Reviews Microbiology, 2012. **10**(2): p. 87-99.

10. Mannelli, A., et al., *Ecology of Borrelia burgdorferi sensu lato in Europe: transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change.* Fems Microbiology Reviews, 2012. **36**(4): p. 837-861.

11. Kurtenbach, K., et al., *Host association of Borrelia burgdorferi sensu lato - the key role of host complement.* Trends in Microbiology, 2002. **10**(2): p. 74-79.

12. D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, et al., *AMBER 2016.* 2016, University of California, San Francisco.

13. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB.* J Chem Theory Comput, 2015. **11**(8): p. 3696-713.

14. Gordon, J.C., et al., *H++: a server for estimating pKas and adding missing hydrogens to macromolecules.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W368-71.

15. Coleman, T.G., H.C. Mesick, and R.L. Darby, *Numerical integration: a method for improving solution stability in models of the circulation.* Ann Biomed Eng, 1977. **5**(4): p. 322-8.

16.	Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

17.	Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* The Journal of Chemical Physics, 1993. **98**.

18.	Roe, D.R. and T.E. Cheatham, 3rd, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013. **9**(7): p. 3084-95.

19.	*The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*

20.	DeLano, W.L., *The PyMOL Molecular Graphics System*. 2002, Palo Alto, CA, USA: DeLano Scientific.

21.	Miller, B.R., 3rd, et al., *MMPBSA.py: An Efficient Program for End-State Free Energy Calculations.* J Chem Theory Comput, 2012. **8**(9): p. 3314-21.

22.	Bhattacharjee, A., et al., *Structural basis for complement evasion by Lyme disease pathogen Borrelia burgdorferi.* J Biol Chem, 2013. **288**(26): p. 18685-95.

23.	McDowell, J.V., et al., *Demonstration of the involvement of outer surface protein E coiled coil structural domains and higher order structural elements in the binding of infection-induced antibody and the complement-regulatory protein, factor H.* J Immunol, 2004. **173**(12): p. 7471-80.

24.	Lam, T.T., et al., *Outer surface proteins E and F of Borrelia burgdorferi, the agent of Lyme disease.* Infect Immun, 1994. **62**(1): p. 290-8.

25.	Jokiranta, T.S., et al., *Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome.* EMBO J, 2006. **25**(8): p. 1784-94.

26.	Vaz, A., et al., *Cellular and humoral immune responses to Borrelia burgdorferi antigens in patients with culture-positive early Lyme disease.* Infect Immun, 2001. **69**(12): p. 7437-44.

27.	Akins, D.R., et al., *Evidence for in vivo but not in vitro expression of a Borrelia burgdorferi outer surface protein F (OspF) homologue.* Mol Microbiol, 1995. **18**(3): p. 507-20.

28.	Miller, J.C., et al., *Temporal analysis of Borrelia burgdorferi Erp protein expression throughout the mammal-tick infectious cycle.* Infect Immun, 2003. **71**(12): p. 6943-52.

29.	Zhang, B., et al., *Serum resistance in Haemophilus parasuis SC096 strain requires outer membrane protein P2 expression.* FEMS Microbiol Lett, 2012. **326**(2): p. 109-15.

30.	Stevenson, B., et al., *Borrelia burgdorferi erp proteins are immunogenic in mammals infected by tick bite, and their synthesis is inducible in cultured bacteria.* Infect Immun, 1998. **66**(6): p. 2648-54.

31.	Wallich, R., et al., *Artificial-infection protocols allow immunodetection of novel Borrelia burgdorferi antigens suitable as vaccine candidates against Lyme disease.* Eur J Immunol, 2003. **33**(3): p. 708-19.

32.	Rose, C.D., P.T. Fawcett, and K.M. Gibney, *Arthritis following recombinant outer surface protein A vaccination for Lyme disease.* J Rheumatol, 2001. **28**(11): p. 2555-7.

33.   Giuliani, M.M., et al., *A universal vaccine for serogroup B meningococcus.* Proc Natl Acad Sci U S A, 2006. **103**(29): p. 10834-9.