The Dissertation Committee for Shuangluo Xia Certifies that this is

the approved version of the following dissertation:

High Throughput Screening of Inhibitors for Influenza Protein NS1

Committee:

_____
Jon D. Robertus, Supervisor

_____
Kenneth A. Johnson

_____
 Robert M. Krug

_____
Richard J. Meyer

_____
 Y. Whitney Yin

High Throughput Screening of Inhibitors for Influenza Protein NS1

by

Shuangluo Xia, B.E. ; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2009

Dedication

To my wonderful parents

Acknowledgements

I would like to thank my advisor, Dr. Jon Robertus for his guidance, support and encouragement throughout my studies. I would also like to express my gratitude to members of my supervisory committee, Dr. Kenneth Johnson, Dr. Robert Krug, Dr. Richard Meyer and Dr. Whitney Yin; also to members of NS1 research committee, Dr. Andrew Ellington, Dr. Eric Anslyn and Dr. Gaetano Montelione.

I greatly appreciate the help and succinct advice of Dr. Kathryn Kavanagh and Dr. Arthur Monzingo. Many thanks to my lab members past and present: Yan Bai, Matthew Lluis, Kenneth Chang, Beth Eisenhut, Warren Hoe, Huda Suliman, and Gregory Sawyer. I would also like to thank Dr. Eun J Cho for working with me in the high throughput screening assay. A special "thank you" to Natalie Potts for her patience and effort in managing financial and administrative matters.

Finally, I am thankful for the support of my family, especially my parents. They have been a stable, loving influence in my life. I am grateful for their love and encouragement.

High Throughput Screening of Inhibitors for Influenza Protein NS1

Publication No._____

Shuangluo Xia, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Jon D. Robertus

Influenza virus A and B are common pathogens that cause respiratory disease in humans. Recently, a highly virulent H5N1 subtype avian influenza virus caused disease outbreaks in poultry around the world. Drug resistant type A viruses rapidly emerged, and the recent H5N1 viruses were reported to be resistant to all current antiviral drugs. There is an urgent need for the development of new antiviral drugs target against both influenza A and B viruses. This dissertation describes work to identify small molecule inhibitors of influenza protein NS1 by a high throughput fluorescence polarization assay.

The N-terminal GST fusion of NS1A (residue 1-215) and NS1B (residue 1-145) were chosen to be the NS1A and NS1B targets respectively for HT screening. In developing the assay, the concentrations of fluorophore and protein, and chemical additives were optimized. A

total of 17,969 single chemicals from four compound libraries were screened using the optimized assay. Six true hits with dose-response activity were identified. Four of them show an IC_{50} less than 1 μM. In addition, one compound, EGCG, has proven to reduce influenza virus replication in a cell based assay, presumably by interacting with the RNA binding domain of NS1.

High throughput, computer based, virtual screenings were also performed using four docking programs. In terms of enrichment rate, ICM was the best program for virtual screening inhibitors against NS1-RBD. The compound ZINC0096886 was identified as an inhibitor showing an IC_{50} around 19 μM against NS1A, and 13.8 μM against NS1B.

In addition, the crystallographic structures of the NS1A effector domain (wild type, W187A, and W187Y mutants) of influenza A/Udorn/72 virus are presented. A hypothetical model of the intact NS1 dimer is also presented. Unlike the wild type dimer, the W187Y mutant behaved as a monomer in solution, but still was able to binding its target protein, CPSF30, with wild type binding affinity. This mutant may be a better target for the development of new antiviral drugs, as the CPSF30 binding pocket is more accessible to potential inhibitors. The structural information of those proteins would be very helpful for virtual screening and rational lead optimization.

Table of Contents

List of Tables

List of Figures

XIII

XIV

Chapter 1: Introduction

1.1 Influenza virus

Influenza viruses are the causative agents for an infectious and debilitating disease commonly referred to as the flu. The virus has a complex replicative cycle and undergoes rapid evolutionary divergence (Krug 2003). In its most aggressive form, the virus was responsible for one of the most devastating pandemics in human history: the notorious Spanish flu caused 20 to 40 million deaths worldwide in the early twentieth century (Palese 2004). The outbreaks of deadly avian influenza H5N1/1997 and H5N1/2004 in Asia and Europe are reminders of the potential for the emergence of a new pandemic (Webby and Webster 2003; Li, Guan et al. 2004; Jan 2007; Khanna, Kumar et al. 2008). The current swine influenza H1N1 outbreak in North America prompted the World Health Organization (WHO) to set the pandemic alert to phase 5, one level below an official pandemic.

1.1.1 Categories of influenza virus

There are three types of influenza viruses: type A, B and C, which are classified based on the immunological relatedness of nucleocapsid (NP) and matrix (M) proteins (Schulman 1971). Both influenza A and B viruses contain eight RNA genomic segments, whereas influenza C virus contains only seven (Cheung and Poon 2007). All the virus types can naturally infect human beings, although influenza C virus seldom causes disease symptoms. In contrast, influenza A and B viruses are virulent pathogens which will lead to the common flu symptoms in humans; influenza A virus has been responsible for all the known influenza pandemics (Schulman 1971; Cheung and Poon 2007).

Influenza A virus can be further subdivided into different subtypes based on the antigenic variation of the surface hemagglutinin (HA) and neuraminidase (NA) glycoproteins. So far, sixteen subtypes of HA (H1-H16) and nine subtypes of NA (N1-N9) have been identified (Laver, Colman et al. 1984; Fouchier, Munster et al. 2005). Each of these subtypes can be isolated from wild aquatic birds, some may also be found in pigs, seals, horses and whales (Claas, Osterhaus et al. 1998; Fouchier, Schneeberger et al. 2004). The ability of these avian viruses to jump host species barriers makes influenza an almost uneradicable disease (Weber and Stilianakis 2007).

1.1.2 Structure and genome of influenza virus

The genomes of both influenza A and B viruses are composed of eight segments of single stranded RNA (Lamb, 2001). The proteins encoded by the corresponding genome segments of both viruses also share similar functions. In Influenza A virus, the three largest segments encode three viral polymerase subunits, PB1, PB2 and PA respectively (Enami and Enami 1996; Elleman and Barclay 2004; Burleigh, Calder et al. 2005). An alternate reading frame in PB1 encodes a non-structural apoptotic protein PB1-F2. The three medium sized segments encode three structural proteins, hemagglutinin (HA), nucleocapsid protein (NP) and neuraminidase (NA) respectively (Schulman 1971). The last two, smallest, segments, M and NS, encode two proteins each. The M segment encodes the matrix protein (M1) and the ion channel protein (M2), while the NS segment encodes the non-structural (NS) proteins NS1A and NS2A, see figure 1.1.

It is known that PB2, PB1, and PA form a polymerase complex for viral transcription and replication (Huang, Palese et al. 1990; Perales and Ortin 1997; Honda, Mizumoto et al. 2002). PB2 polymerase contains a nuclear localization signal, which helps to transport the polymerase into the nucleus of infected cells (Mukaigawa and Nayak 1991; Perales, de la Luna et al. 1996). It also has a cap-binding protein functioning that aids in generating the cap structures necessary for viral mRNA transcription (Ulmanen, Broni et al. 1981; Blaas, Patzelt et al. 1982; Braam, Ulmanen et al. 1983; Ulmanen, Broni et al. 1983). PB2 was found to associate with the PB1 subunit through amino terminus by mutageness and immunoprecipitation assays (Digard, Blok et al. 1989; Li, Rao et al. 2001; Fechter, Mingay et al. 2003).

PB1 polymerase is encoded by RNA segment 2. It plays an important role in the RNA polymerase complex as the subunit responsible for elongation of the primed nascent viral mRNA (Poole, Elton et al. 2004). PB1 is critical in the assembly of three polymerase protein subunits, and localizes in the nucleus of infected cells (Biswas and Nayak 1994; Gonzalez, Zurcher et al. 1996; Kobayashi, Toyoda et al. 1996; Honda, Mizumoto et al. 2002).

PA subunit is the smallest polymerase, and is encoded by RNA segment 3. The carboxyl terminus of PA is found to be critical for viral transcription by mutagenesis assay (Sanz-Ezquerro, de la Luna et al. 1995; Sanz-Ezquerro, Zurcher et al. 1996; Fodor and Smith 2004). Sequence comparison suggests that it has helicase activity, although the exact functions of PA still remain unknown (de la Luna, Martinez et al. 1989).

2
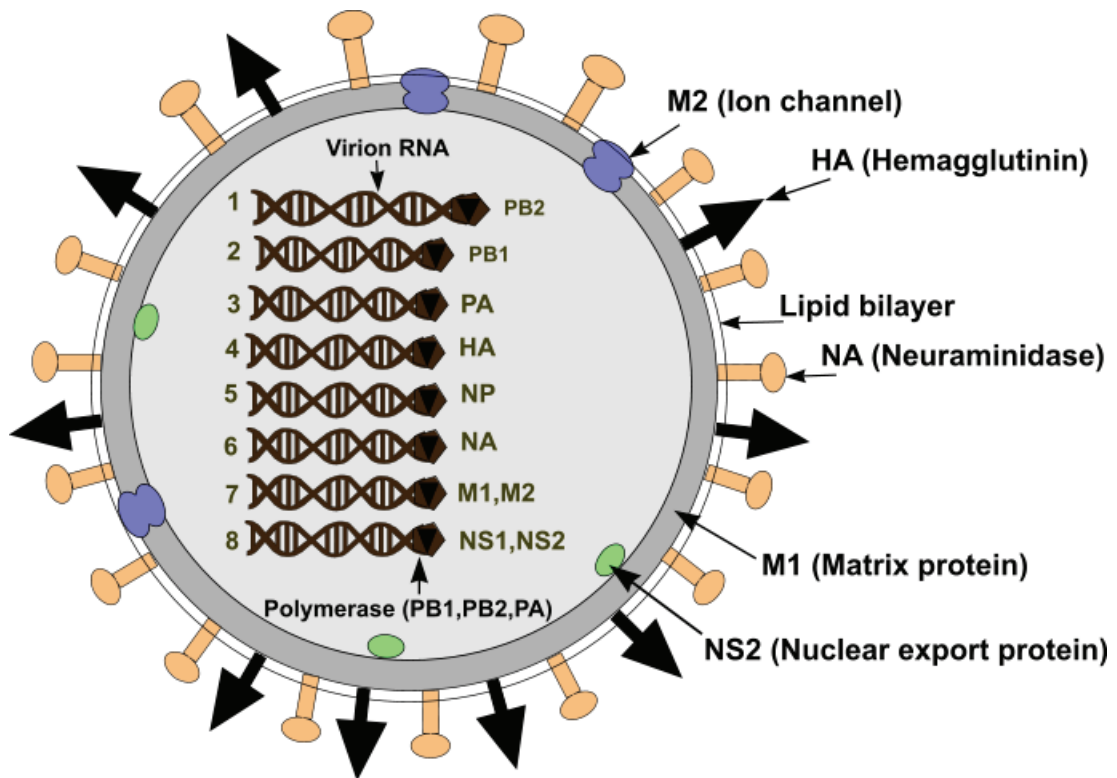
Figure 1.1: Schematic diagram of structure of the influenza A virus. The eight RNA genomic segments are numerically numbered.

Hemagglutinin (HA), encoded by segment 4, is an integral membrane protein responsible for binding the flu virions to sialic acid containing receptors on host cell surfaces (Staudt and Gerhard 1983). HA is initially synthesized as a precursor polypeptide, and then undergoes a series of post-translational modifications (Skehel, Bayley et al. 1982; Horimoto, Nakayama et al. 1994; Webby, Perez et al. 2004). First, the amino-terminal 14 residues are removed, followed by palmitic acidification at the cysteine residue near the carboxyl-terminus. Afterward, the protein is cleaved into two disulphide-linked subunits (HA_1 and HA_2) by trypsin-like proteases of the host cell (Connor, Kawaoka et al. 1994; Vines, Wells et al. 1998). HA, the major flu antigen, is subject to a high rate of mutation due to the viral RNA polymerase error-prone activity. Among all sixteen subtypes of HA (H1-H16) identified so far, the amino acid sequences of HA variants differ by at least 30% from each other (Staudt and Gerhard 1983; Connor, Kawaoka et al. 1994; Vines, Wells et al. 1998).

Nucleoprotein (NP) is encoded by segment 5, and is actively synthesized in the infected cells. The binding of NP to single-stranded RNA is not sequence specific (Kistner, Muller et al. 1989). It is estimated that one NP molecules binds 15-20 nucleotides (Winter and Fields 1981; Kobayashi, Toyoda et al. 1994; Albo, Valencia et al. 1995). Together with viral polymerase, NP and vRNA form a so called viral ribonucleoprotein (vRNP) complex, which is a supercoiled ribbon structure with the polymerase complex at the end of viral genome (Beaton and Krug 1986; Shapiro and Krug 1988; Biswas, Boutz et al. 1998). It is also believed that the phosphorylated NP plays an important role in "switching" of RNA polymerase activity from mRNA synthesis to complementary (cRNA) and viral RNA synthesis (Shapiro and Krug 1988; Kistner, Muller et al. 1989). NP is also found to be involved in the viral RNA nuclear transport (Martin and Helenius 1991; O'Neill, Jaskunas et al. 1995; Whittaker, Bui et al. 1996; Neumann, Castrucci et al. 1997). The detailed mechanisms of switching and transportation are still waiting to be revealed. Like HA, NP is also a major target of host immune response (Shapiro and Krug 1988).

Encoded by segment 6, Neuraminidase (NA) is also an integral membrane glycoprotein. The structure of NA shows that the monomer protein is composed of four domains: a mushroom-shaped globular head, a thin stalk, a transmembrane domain and a cytoplasmic domain (Varghese and Colman 1991). NA is not required for virus replication, as demonstrated by a NA-deficient virus (Li, Schulman et al. 1993; Hausmann, Kretzschmar et al. 1997). The

actual role of NA is to remove terminal sialic acids from the glycoproteins on the virion surface, thereby helping to release the progeny virions from the infected cell (Palese, Tobita et al. 1974; Palese and Compans 1976; Liu, Eichelberger et al. 1995).

Segment 7 of influenza A virus encodes two proteins: the matrix protein (M1) and the ion channel protein (M2) (Shih, Nemeroff et al. 1995). M1 is a collinear transcript of segment 7, whereas M2 is encoded by a spliced mRNA variant of segment 7 (Ye, Liu et al. 1999). Collinear transcripts arise with common initiation but different termini. Being the most abundant influenza virion protein, M1 forms a shell underneath the viral lipid membrane to interact with the virus-specific ribonucleoprotein (vRNP) and cytoplasmic domains of various integral membrane proteins (Wakefield and Brownlee 1989; Ye, Baylor et al. 1989; Elster, Fourest et al. 1994). It is reported that M1 binds to viral RNA in a sequence nonspecific manner. Once M1 binds to vRNP, it accelerates the nuclear export of vRNP, and at the same time inhibits the import of vRNP (Ye, Robinson et al. 1995; Bui, Whittaker et al. 1996; Huang, Liu et al. 2001; Sakaguchi, Hirayama et al. 2003). In 2004, Elleman and Barclay demonstrate that the M1 matrix protein alone controls the filamentous phenotype of influenza A virus (Elleman and Barclay 2004). Unlike M1, M2 is an integral membrane protein (Lamb, Zebedee et al. 1985). The native protein exists in a homotetramer conformation, and acts as an ion channel that modulates the pH of the Golgi during HA synthesis (Pinto, Holsinger et al. 1992; Wang, Lamb et al. 1994). Thus it helps to stabilize the native conformation of newly synthesized HA during viral assembly. M2 also plays an important role in the acidification of the interior of the virion particle during virus uncoating (Henkel and Weisz 1998; Park, Castrucci et al. 1998).

Like segment 7, segment 8 also encodes two proteins: nonstructural proteins NS1 and NS2. NS1 message is a collinear transcription of segment 8, whereas NS2 is encoded by a spliced mRNA of segment 8 (Alonso-Caplen and Krug 1991; Nemeroff, Utans et al. 1992). NS1 is not incorporated into virions, and is the only nonstructural protein in influenza virus (Greenspan, Palese et al. 1988; Nemeroff, Qian et al. 1995). NS1 has been shown to have multiple functions, including binding dsRNA and interfering with host translation and mRNA processing (Hale, Randall et al. 2008). After expression it localizes in the host cell nucleus. The details will be discussed in a separate chapter in the following section.

The name of NS2 is somewhat misleading as NS2 is found to incorporate into viral particles in low amounts (Richardson and Akkina 1991; Yasuda, Nakada et al. 1993). Unlike NS1, NS2 primarily localizes in the cytoplasm (Richardson and Akkina 1991; Yasuda, Nakada et al. 1993). The role of NS2 is to promote replication of vRNA and to mediate the export of newly synthesized vRNP from the nucleus to the cytoplasm, although the detailed mechanism still remains unknown (O'Neill, Talon et al. 1998; Neumann, Hughes et al. 2000).

In summary, influenza A virus is a small enveloped pleomorphic particle with a diameter of 80-120nm. The viral particles contain a lipid bilayer envelope derived from the host cell membrane. HA, NA, and M2 are embedded in that lipid envelope, and project like spikes. Under the lipid envelope, there is an electron dense layer composed of M1 protein. Inside the virion, the genome of influenza virus is composed of eight unique segments of single stranded RNA. The RNA is encapsidated by multiple NP molecules to form RNP complexes. The viral RNA polymerase complex is composed of PB2, PB1, and PA. NS2 is also present in the virion in low amounts.

1.1.3 Transcription and replication of influenza virus

Transcription and replication of the influenza viral genome takes place in the nucleus of infected cells (Herz, Stavnezer et al. 1981; Jackson, Caton et al. 1982). Like other negative-stranded RNA viruses, the RNA-dependent RNA polymerase encoded by the three largest flu genome segments converts the single-stranded virion RNA (vRNA) into viral mRNAs. The viral mRNAs are exported into the cytoplasm, and translated by host proteins into the proteins of new virus particles, see figure 1.2.

The viral mRNA requires a type I cap structure (m7GpppNm, where N is any nucleotide) and a polyadenylated tail (Braam, Ulmanen et al. 1983). The whole transcription process starts by the recognition of the cap structure of host pre-mRNA by PB2 subunit of viral RNA polymerase. Once PB2 recognizes and binds the cap structure of host pre-mRNA, it cleaves about 10 nucleotides downstream from the cap structure. The resulting capped RNA fragments are used by viral polymerase as transcription primers. Transcription is initiated by incorporating a G or C onto the 3' end of the primers which based pair with viral RNA template. The elongation of the mRNA chain proceeds until reaching a polyadenylation signal, which consists of 5–7 U residues approximately 17 nucleotides from the 5' end of vRNA. Finally, the viral polymerase generates a poly (A) tail at the end of mRNA by stuttering and slipping between the
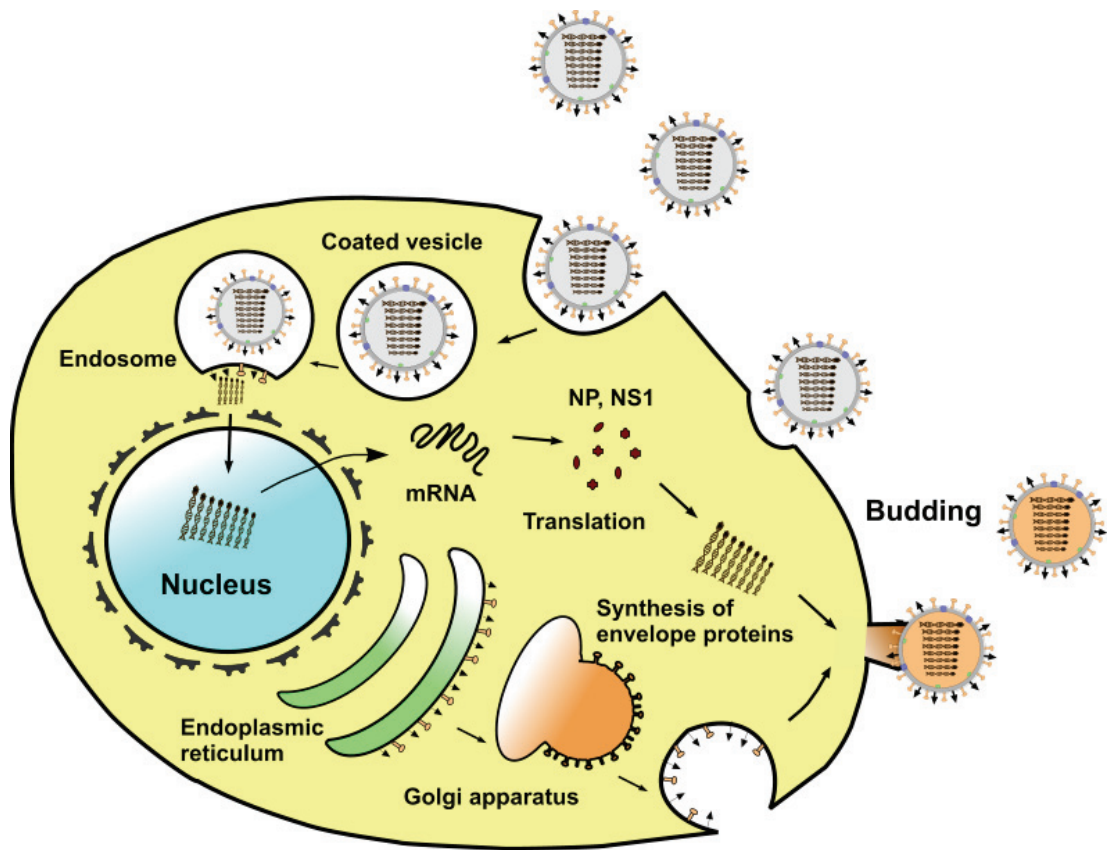
Figure 1.2: Schematic diagram of influenza A virus replication cycle. Redrawn and modified from http://www.influenzareport.com/influenzareport.pdf

template and elongating RNA chain (Krug, Broni et al. 1979; Plotch, Bouloy et al. 1979). In other words, it iteratively copies the U-track at the end of vRNA to nascent mRNA. It is well demonstrated that the 5' cap structure, together with poly(A) tail are essential for mRNA stabilization and nuclear export (Bouloy, Plotch et al. 1980).

A second population of RNA transcripts is complementary RNA (cRNA); these are full-length copies of the vRNA that can be used as templates for progeny vRNA synthesis. Unlike viral mRNA synthesis, the synthesis of cRNA is initiated on its own promoter without a primer, and is not terminated with a polyadenylated tail (Hay, Skehel et al. 1982). Instead, dinucleotides are generated by viral RNA polymerase to serve as primers for the chain elongation. As observed from in vitro replication system, the RNA synthesis is significantly increased in the presence of additional dinucleotide primers (Biswas, Boutz et al. 1998). It is also observed that the amount of vRNA synthesis from cRNA template is 10 times higher than that of cRNA synthesis from vRNA template (Vreede, Jung et al. 2004). This phenomenon could be due to the difference in the initiation activity. In addition, the synthesis of cRNA is dependent on the continued production of functional viral proteins, whereas the synthesis of mRNA continues in the absence of protein synthesis. Therefore, it was suggested that cRNA is unstable and tends to be degraded until there are enough copies of newly synthesized polymerases or NP, which stabilize cRNA by forming RNP structure (Nagata, Takeuchi et al. 1989; Shimizu, Handa et al. 1994; Nagata, Kawaguchi et al. 2008). However, the question remains unsolved as to what form of polymerase complexes are required for the switch from the capped RNA primer dependent mRNA synthesis to unprimed cRNA synthesis, and how this switch is initiated (Nagata, Kawaguchi et al. 2008). The replicated vRNA and newly synthesized viral proteins are then assembled into a virion, and further bud from the cell surface.

1.2 Vaccination and antiviral therapies

Currently, the most effective way to prevent influenza infection is vaccination with killed or attenuated virus (Nichol 2006; Ada 2007). Almost all flu vaccines on the market are produced from virus grown in fertile hens' eggs, and then inactivated by either β-propiolactone or formaldehyde (Nichol, Lind et al. 1995). The resulting vaccines consist of whole virus, detergent-treated split virus product, or purified hemagglutinin (HA) and neuraminidase (NA) surface antigens. Due to adverse affect on young children, whole-virus vaccines are not widely

available (Heikkinen, Ruuskanen et al. 1991; Nichol, Lind et al. 1995; Block 2004). In contrast, split virus products and purified HA/NA surface antigen are well tolerated and safe (Laver 1985). Ideally, the vaccine will be most effective when the strains included in the vaccine are a good antigenic match with the current circulating strains of influenza. Currently licensed influenza vaccines are trivalent, inactivated formulations that consist of HA proteins from H1N1, H3N2 and influenza B strains respectively (Laver 1985). Those three strains were chosen by the WHO Global Influenza Surveillance Network based on their best predicition of which strains will be predominant in the coming year. The vaccine is designed to block the function of HA and NA proteins, as both of them are the primary targets of the protective antibody response. As a result of vaccination, antibodies against HA can neutralize subsequent viral infection, and antibodies against NA can modify the severity of disease. Unfortunately, the HA protein mutates from year to year, by the process called "antigenic drift" (Luther, Bergmann et al. 1984; Yamada, Brown et al. 1984; Lipkind and Shihmanter 1986). By randomly accumulating amino acid changes, HA is said to "drift" from one shape to another, thereby making the human immune system less able to recognize the new strain. As a result, antiviral drugs targeting non-surface proteins are gaining much more attention in the war against influenza virus (Ferguson, 2005; Germann, 2006).

Currently, there are two classes of antiviral drugs available: the M2 inhibitors (amantadine and rimantadine), and the neuraminidase inhibitors (zanamivir and oseltamivir). Amantadine and rimantadine were developed to block the ion channel activity of M2 protein (Rees, Harkins et al. 1997; Ellis and Zambon 2002). Therefore, viral replication is inhibited by the blockade of hydrogen ion flow. It has been reported that the inhibition of viral replication could be achieved at micromolar concentration (Rees, Harkins et al. 1997). However, these drugs are ineffective against influenza B virus because it doesn't encode M2 protein (Gubareva, Kaiser et al. 2000). Both drugs are found to be 70–90% effective in preventing and relieve the symptoms caused by naturally occurring influenza A, but amantadine has more adverse effects, including insomnia, headache and vomiting, which are troublesome in elderly people and very young children (Nicholson 1996; Margo and Shaughnessy 1998). The main drawback to the use of M2 inhibitors is that drug resistance emerges rapidly during treatment (Suzuki, Saito et al. 2003). Sadly, those drug-resistant variants are fully pathogenic and transmissible to others. Both amantadine and rimantadine resistant strains have been characterized by mutations in the

9

transmembrane domain of the M2 protein (Suzuki, Saito et al. 2003). The most common one is a single nucleotide change in codon 31, which results in a single amino acid substitution in the membrane spanning region (Intharathep, Laohpongspaisan et al. 2008; Monto 2008). Recently, the highly virulent H5N1 strain was reported to be resistant to both amantadine and rimantadine (Harrod, Emery et al. 2006).

Other drugs, like Zanamivir and Oseltamivir, have been developed to target the NA protein, which is expressed in both influenza A and B viruses (Woods, 1993; Ryan, 1994; Kim, 1997; Mendel, 1998). Although NA is not required for replication, it is required for releasing newly assembled virons from the infected cell surface by cleaving terminal sialic acid residues from glycoconjugates (Oxford, Mann et al. 2003). As a result, NA inhibitors inhibit the release of virions from infected cells, cause aggregation of virions at the cell surface and retard further spreading to other healthy cells. Zanamivir, also called Relenza or GG167, is the first neuraminidase inhibitor commercially developed (Monto, Webster et al. 1999). The debut of zanamivir is an important milestones of the rational drug design. Zanamivir was designed based on the structure of the transition state analogue of NA, DANA (2-deoxy-2,3-dehydro-N-acetylneuraminic acid). The resulting binding affinity increased 5000 fold compare to that of sialic acid, the native substrate of NA (von Itzstein, Wu et al. 1993; Zimmerman, Ruben et al. 1997). Because of its poor oral bioavailability, zanamivir has to be administered through inhalation (Englund 2002).

Oseltamivir, also called Tamiflu or GS4104, was designed based on another potent transition state analogue of sialic acid hydrolysis. The drug has good oral bioavailability (Li, Escarpe et al. 1998). In 1999, Oseltamivir was approved as the first orally effective drug for the treatment of flu infection (Sidwell, Bailey et al. 1999). Mutant strains resistance to zanamivir and Oseltamivir have been reported, although the frequency is much less than those to amantadine and rimantadine (Moscona 2004). In 2005, H5N1 virus that is resistant to the Oseltamivir has been isolated (Puthavathana, 2005). During the 2007-2008 flu season, 11% of H1N1 virus was found to be Oseltamivir resistant (Lowen and Palese 2007). Two mechanisms of resistance have been identified: the NA dependent resistance and the NA independent resistance (Aoki, Boivin et al. 2007; Ferraris and Lina 2008; Lackenby, Thompson et al. 2008). The NA dependent resistance involves single amino acid substitution in the active site of the NA, which alters its sensitivity to the drug. The NA independent resistance involves mutations

in the HA receptor-binding site, thereby reducing the efficiency of virus binding to cellular receptors. Therefore, virus can free from the infected cells without the need for NA. The emergence of drug resistant strains emphasizes the need for new antiviral drugs

1.3 Antiviral defense mechanism

When influenza virus invades humans, the innate immune responses of the host provide immediate defense against infection. Several components are involved in the innate immune system, including mucus, macrophage, interferon α/β (INFα/β), cytokines (IL-1, IL-6 and TNF α), natural killer (NK) cells (Hartshorn, Karnad et al. 1990). Representing the first barrier against infection, the innate immune response also provides the appropriate signals required for the subsequent adaptive immune response, in which antigen-specific memory cells (T and B cells) and their products functions as antigen-specific effectors, such as cytotoxic T lymphocytes and antibodies, to target the virus (Hartshorn, Karnad et al. 1990; Tamura and Kurata 2004).

The detection of influenza viral RNA in cytoplasm is accomplished by two related cytosolic RNA helicases, termed retinoic acid-induced gene I (RIG-I, also know as DDX58) and melanoma differentiation-associated gene 5 product (MDA5, also known as IFIH1) (Andrejeva, Childs et al. 2004; Yoneyama, Kikuchi et al. 2004). Binding to viral dsRNA by RIG-I and MDA5 initiates the conformational changes which leads to recruiting additional cellular factors, including the recently identified interferon β promoter stimulator 1(IPS-1), mitochondrial antiviral signaling (MAVS) protein, and virus-induced signaling adaptor (VISA) (Kawai, Takahashi et al. 2005; Seth, Sun et al. 2005). As a result, different cellular kinases, including interferon regulatory factor-3 (IRF3) kinase, TNF-receptor-associated factor family member-associated nuclear factor-κB activator binding kinase 1(TBK1) and inhibitory κBα protein kinas (IKK-ε), are activated. Then, the activated IRF3 kinase, together with nuclear factor-κB (NF-κB) and activating protein 1(AP-1), bind to the promoter of IFNβ promoter and initiate transcription (Garcia-Sastre 2006).

Once IFNα/β has been synthesized, it is secreted and binds to the IFNα/β receptor on neighboring cells. As a result, two members of the Janus tyrosine kinase family (JAK1 and Tyk2) are activated. JAK1 and Tyk2 in turn phosphorylate STAT1 (signal transducers and activators of transcription) and STAT2 transcription factors (Stark, Kerr et al. 1998; Le Bon, Schiavoni et al. 2001). The phosphorylated STAT1 interacts with STAT2 and p48/IRF-9 to

11

form the transcription factor IFN-stimulated gene factor 3 (ISGF3). The ISGF3 complex then binds to specific DNA sequences called IFN-stimulated regulatory elements (ISREs), promoting the transcription of approximately 100 to 300 genes. Among these genes, those encoding the dsRNA-activated protein kinase (PKR), the 2',5'-oligoadenylate synthetases (2'-5'A synthetases), and the Mx proteins are demonstrated to interfere with viral replication by different mechanisms, including binding to viral nucleocapsids, translation inhibition, RNA degradation, RNA editing, and apoptosis induction (Biron 2001; Weber, Kochs et al. 2004). Overall, the secretion of IFNα/β by virus-infected cells contributes to the induction of an antiviral state in neighboring uninfected cells.

The serine-threonine protein kinase PKR is one of the most extensively studied interferon-induced protein kinase (Ludwig and Planz 2008). Although the level of PKR increases in cell in response to IFN, it remains inactive unless it interacts with viral dsRNA. The binding of dsRNA to PKR causes a conformational change of the protein, which causes the C-terminal kinase domain of PKR to release the N-terminal dsRNA binding domain (Williams 1999). As a result, PKR becomes autophosphorylated and dimerized as an active kinase. The activated PKR phosphorylates α-subunit of the translation initiation factor eIF-2 α (Williams 1999). This leads to the inactivation of eIF-2 α, which in turn results in the inhibition of protein synthesis. Like PRK, the 2'-5'(A) synthetases need to be activated by viral dsRNA (Silverman 2007). Once activated, 2'-5'(A) synthetases polymerize ATP into 2'-5' linked oligoadenylates of different length. The resulting 2'-5' A molecules bind to and activate a latent ribonuclease, RNase L. Activated RNase L induces the degradation of RNAs, including viral mRNAs and rRNAs, therefore consequently inhibits viral replication and protein synthesis. Finally, the Mx proteins were discovered as IFN-induced proteins which inhibit replication of specific groups of viruses, including influenza virus, rhabdovirus and orthomyxovirus. The Mx proteins bind GTP and have an intrinsic GTPase activity that is necessary for their intracellular antiviral actions. Although the detailed mechanism of that action is still unknown, it was proposed that Mx proteins interfere with virus replication through a dynamin-like force-generating mechanism by wrapping around viral nucleocapcids (Horisberger 1995; Haller, Staeheli et al. 2007).

Despite the host's sophisticated innate immune system, influenza viruses continue to successfully infect them and cause disease. The success of influenza viruses is partly due to the

acquisition of viral genes that antagonize the host immune response. In the case of influenza A virus, this IFNα/β antagonistic function is encoded by the NS1 gene.

1.4 The influenza virus NS1 protein

1.4.1 NS1 from influenza A virus (NS1A)

The NS1 protein of influenza A virus (NS1A) is not a structural component of the virion, but is highly expressed in infected cells (Krug 1993). It has a strain specific length of 230 to 237 residues, and an approximate molecular mass of 26 kDa. The NS1A protein is a multifunctional protein that participates in both protein-protein and protein-RNA interactions (Hale, Randall et al. 2008). Two functional domains identified are the N-terminal 73 amino acids RNA-binding domain (RBD) and C-terminal effector domain (residue 74-230), see figure 1.3. The full length NS1A exists as a homodimer with both the RNA binding domain and effector domain contributing to multimerization (Chien, Tejero et al. 1997; Liu, Lynch et al. 1997). The RBD binds non-specifically to double-stranded RNA (dsRNA), and indirectly protects the virus against the antiviral state induced by interferon α/β. This protection is achieved by blocking the activation of the 2'-5'-oligo(A) synthetase/RNase L pathway, which inhibits virus replication by degradation of viral RNA (Min and Krug 2006). It was also found that out-competing 2'-5'-oligo(A) synthetase for interaction with dsRNA is a predominant function of NS1A RNA-binding domain (Min and Krug 2006). Mutation studies of NS1A RBD show that Arg38 and Lys41 are two amino acids critical for the dsRNA binding activity (Chien, Xu et al. 2004). A recombinant influenza A virus expressing the mutant (R38A)-NS1A protein is attenuated 1000-fold in replication (Min and Krug 2006). These results suggest that the NS1A RBD is a valid target for the development of antiviral drug. The RBD also interacts with the retinoic acid-inducible gene I (RIG-I) binding domain (Samuel 2007; Uematsu and Akira 2007). The RIG-I protein functions as an intracellular viral sensor, which can be activated by influenza virus ssRNA/dsRNA. The binding of RIG-I to RBD of NS1A inhibits the ability of viral detection by the cell (Hornung, Ellegast et al. 2006; Pichlmair, Schulz et al. 2006).

The C-terminal effector domain of NS1 was found to interact with at least four different proteins: elongation initiation factor 4GI (eIF4GI), protein kinase R (PKR), cleavage and polyadenylation specificity factor (CPSF) and poly(A)-binding protein II (PAB II) (Hale, Randall et al. 2008). The first 113 amino acids, but not the first 81 residues, of the NS1A

13
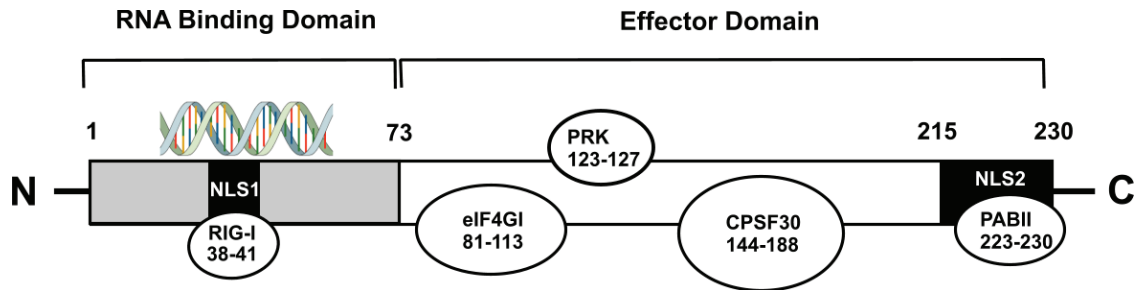
Figure 1.3: Structural domains of NS1A protein.
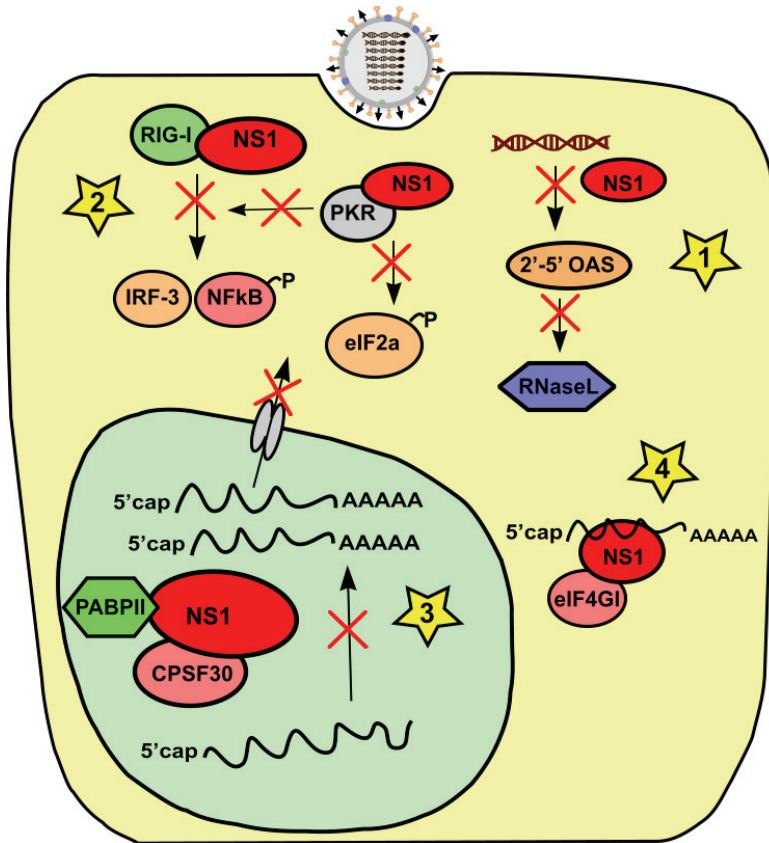
Figure 1.4: Schematic diagram of NS1 functions. Star1: Directly limits the antiviral state by blocking the activation of OAS/RNaseL; Star2: Inhibits the induction of IFN-beta at pre-transcriptional level; Star3: Post-translational block of the processing and nuclear export of all cellular mRNAs; Star4: Enhancement of viral mRNA translation. Redrawn and modified from (Hale, Randall et al. 2008).

protein are sufficient to bind eIF4GI (Aragon, de la Luna et al. 2000; Burgui, Aragon et al. 2003). It was suggested that NS1 recruits eIF4GI specifically to the 5' untranslated region of viral mRNA, thus preferentially enhances the translation of the viral mRNA (Burgui, Aragon et al. 2003). The direct binding of PKR to residues 123 to 127 of NS1A is necessary and sufficient to block the activation of PKR, and thereby inhibit the antiviral effect of PKR (Li, Min et al. 2006; Min, Li et al. 2007).

The effector domain also binds to the 30 kDa subunit of CPSF, thus inhibiting the maturation and exportation to the cytoplasm of the host cellular antiviral mRNAs (Li, Rao et al. 2001; Noah, Twu et al. 2003; Twu, Noah et al. 2006). The second and third zinc fingers (F2F3) of CPSF were shown to interact with NS1A (Twu, Noah et al. 2006). A recombinant Udorn virus expressing NS1A with mutations spanning the CPSF binding sites (residues 184 to 188) induced a high level of interferon β (INF β) mRNA, and was attenuated 1000 fold in replication (Noah, Twu et al. 2003; Twu, Noah et al. 2006). An engineered MDCK cell lines which constitutively expressing epitope-tagged F2F3 in the nucleus effectively blocked the binding of endogenous CPSF to NS1A, thereby selectively inhibited the influenza virus A replication (Noah, Twu et al. 2003; Twu, Noah et al. 2006). All these observations suggest that the CPSF30 binding site can be targeted for development of new antiviral drugs.

The function of PABII in cells is to facilitate the elongation of poly(A) tails during the generation of mRNAs. The binding of PABII to residues 215 to 230 of NS1A prevents PABII from properly extending the ploy(A) tail of cellular pre-mRNA, thus retarding pre-mRNA export from the nucleus (Chen, Li et al. 1999).

Finally, the effector domain is also crucial for the function of the RBD. It was revealed that dimerization of these two domains is essential for the NS1A protein to interact with RNA or other cellular proteins (Wang, Basler et al. 2002; Fernandez-Sesma 2007). Taken together, NS1A is a very potent inhibitor of host cell innate immunity and allows influenza A virus to efficiently escape host defenses and to further establish infection (see figure 1.4).

Well before the very recent structure of full length NS1A from an H5N1 virus became available, the structures of the two function domains were solved independently. The NMR and X-ray structure of the 73-residue N-terminal domain of NS1A shows that the protein is active as a dimer (Qian, Chien et al. 1995; Chien, Xu et al. 2004). The structural analysis revealed a unique six-helix structure for the dimer, which differs from that of other known dsRNA-

binding proteins. The dimeric protein binds dsRNA as part of its function in the viral cycle. According to the sedimentation equilibrium measurements, the NS1A73 dimer binds to a 16-bp synthetic dsRNA with a 1:1 stoichiometry (Chien, Xu et al. 2004). In 2008, the crystal structure of NS1A73 dimer bound to a self-complementary 21-nucleotide RNA (with 19 base pairs) was determined (Cheng, Wong et al. 2009). The complex structure reveals that RBD recognizes the A-form dsRNA in a major groove binding mode. The critical Arg38 was found to penetrate into the dsRNA helix and to form hydrogen bonds with the phosphate groups of two nucleotides from two different RNA strands. Furthermore, a hydrogen bond network is formed by Arg38 and its symmetrically related molecules to anchor dsRNA along the binding surface. The structure of the RBD and dsRNA complex also explains why the RBD of NS1 has the ability to distinguish between dsRNA and dsDNA. The inter-molecular contacts are directed towards the phosphate backbone of the A helix, and the 2′-OH groups of the RNA strands.

The X-ray structure of the NS1A effector domain (residues 79-205) from the mouse-adapted influenza A/PR8/34(PR8) virus strain was solved in 2006 (Bornholdt and Prasad 2006). Like the N-terminal domain, the effector domain forms a dimer in solution. Each monomer consists of seven β-stands and three α-helices. The structure is believed to be a novel fold, which can be described as an α-helix β-crescent fold, as the β-strands form a crescent-like shape around the α-helix. The NS1A from the PR8 virus strain does not bind CPSF30 because it does not have the required consensus human recognition sequence at residue 103 (it has S instead of F) and residue 106 (I instead of M) (Das, Ma et al. 2008). The CPSF30 binding site is crucial for virus replication and is a proven target for new antiviral drug development. As a result, the structure of effector domain that is capable of binding CPSF30 would be extremely helpful for structure based drug design.

Recently, the x-ray structure of the full length NS1A (R38A,K41A) from an H5N1 strain (A/Vietnam/1203/2004) was solved by Bornholdt and Prasad (Bornholdt and Prasad 2008). R38 and K41 of the protein were mutated to alanine. This double mutant completely abrogated aggregation and allowed the protein to be sufficiently concentrated for crystallization. Compared to the individually determined structures of RNA binding domain and effector domain, the overall polypeptide folding of RBD and effector domain from H5N1 showed only minor structural alterations. The most striking difference is the dimer interaction between the two domains. The two domains of each NS1A molecule separately interact with their respective

16

domains from the neighboring NS1A molecules, related by crystallographic two-fold symmetry. This results in the formation of a chain of NS1A molecules with alternating dimers of RBD and dimers of effector domain. The authors proposed a model for full length NS1 interacting with dsRNA in which NS1A cooperatively oligomerizes in the presence of dsRNA to from a tubular structure with dsRNA in the hollow central tunnel. This model explains why NS1A is able to sequester varying lengths of dsRNA in infected host cells. It was further substantiated by cryo-electron microscopy images of native full-length NS1 from influenza A virus H0N1 strain (A/PR8/34) with different length of dsRNA.

1.4.2 NS1 from influenza B virus (NS1B)

The NS1 encoded by influenza B virus (NS1B) shares little sequence homology with NS1A protein, and has a larger size (281 residues) with molecular mass around 32kDa. Like NS1A, NS1B also has two functional domains: the N-terminal dsRNA-binding domain (residue 1-93) and C-terminal effector domain (residue 94-281) (Wang and Krug 1996; Yuan and Krug 2001). Although the sequence identity of RBD between the two viruses is only 20%, the x-ray structures show that they share conserved surface features (Wang, Riedel et al. 1999; Yuan, Aramini et al. 2002). The dsRNA-binding surface of both NS1A and NS1B proteins have highly conserved patches of basic and hydrophilic amino acids which are complementary to the polyphosphate backbone conformation of the A-form dsRNA. The NS1B RBD would also be expected to protect influenza B virus against the interferon α/β-induced antiviral state by out-competing 2'-5'-oligo(A) synthetase for interaction with dsRNA. The direct binding of interferon stimulated gene protein 15 (ISG15) to the residues 1 to 104 of NS1B is necessary and sufficient to block the antiviral ability of ISG15 (Yuan, Aramini et al. 2002). The ISG15 is an IFN-inducible protein that enhances the IFN-mediated antiviral response, including RNA splicing, antiviral ability and cytoskeleton regulation (Ritchie and Zhang 2004). Although the size is larger, the effector domain of NS1B does not bind either CPSF30 or PABII (Wang and Krug 1996; Yuan and Krug 2001). There is also no evidence to show direct interactions between the NS1B effector domain and eIF4GI or PKR. This may explain why NS1B does not inhibit the posttranscriptional processing of cellular mRNAs. As NS1 is the major IFNα/β antagonist, the difference in NS1A and NS1B very likely contributes to the different biological properties of influenza A and B viruses.

1.5 Project goals

       The goal of this project is to identify small molecule ligands of the NS1 protein of influenza A/B viruses. The ligands are to be targeted to the dsRNA binding region of the RBD and the F2F3 binding region of the effector domain. Such ligands will serve as chemical "platforms" for future rational based drug design. Research in Dr.Krug's laboratory has validated both the N-terminal RNA-binding domain and C-terminal effector domains as efficacious targets for new antiviral drugs.

       A fluorescence polarization (FP) assay was developed for the high throughput screening of small compounds library against the NS1 RBD protein. The focus of my study is to refine and optimize this FP assay to screen several commercially available small-compound libraries, including ChemBridge fragment library (4000 compounds), ChemBridge kinase library (13000 compounds), NCI clinical trial compound library (400 compounds), and natural product library (4000 compounds). The "hit" compounds with K_d <100 μM were further validated and refined using structure-based analyses. Several confirmed "hit" compounds were soaked or co-crystallized with appropriate protein receptors in an effort to elucidate their binding modes. Another goal of this project was to crystallize related NS1 proteins, including different domains of the NS1 protein as well as the full length NS1. The structural information of those proteins should be extremely helpful for high throughput virtual screening and rational lead optimization.

Chapter 2: Verification and Optimization of the Fluorescence Polarization (FP) Assay as a High Throughput Screening (HTS) Method for NS1 Inhibitors

2.1 Introduction

Structural and biochemical data show that the RNA binding domain of NS1A is a valid target for the development of antiviral drugs (Min and Krug 2006). A fluorescence polarization (FP) assay has been developed for the high throughput screening of small compounds library against RBD of NS1 protein. Fluorescence polarization is a spectroscopic method that is based on the molecular movement of the fluorescent molecules in solution (Nasir and Jolley 1999; Roehrl, Wang et al. 2004). When excited with polarized light, the fluorescent molecules emit light in the same polarized plane if the molecule remains stationary. If the fluorescent dye tumbles significantly during the fluorescent life time, the emitted light is depolarized relative to the excitation plane; this would be the case if the dye were attached to a small, mobile molecule. Conversely, if the fluorescent dye is attached to a very large, slowly rotating molecule during excitation state, the emitted light remains highly polarized. Therefore, fluorescence polarization provides a direct measurement of the extent of fluorescent dye binding to macromolecules, including proteins, nucleic acids and other biopolymers. Experimentally, the degree of polarization is determined by measuring the fluorescence intensities parallel and perpendicular to the plane of the excitation light. To simplify the expression mathematically, fluorescence polarization is expressed in terms of anisotropy. As fluorescence polarization provides direct measurement of the binding of a fluorescent labeled molecule to a target molecule, it is widely used to determine the binding affinity of molecular interactions.

Here we determine the equilibrium dissociation constant, K_d, for binding of fluorescein-labeled dsRNA to different NS1A /NS1B constructs by the FP assay. To facilitate our search for ligand/inhibitors, we were aiming to find the NS1 construct with highest binding affinity to dsRNA. The rotation of fluorescent labeled molecules is influenced by temperature, molecular volume, and most importantly, by the solution viscosity which can be affected by the presence

of additive chemicals. Therefore, the impact of additive chemicals on the FP assay, including tRNA, BSA and DMSO, will be tested and optimized.

2.2 Materials and methods

2.2.1 Construction of various NS1 variants

The NS1A(1-215)-pGEX4T3, NS1B(1-104)-pGEX3X, NS1B(1-145)-pGEX3X, and NS1B(1-218)-pGEX3X plasmids were provided by Dr.Robert Krug. The 215 residue NS1A(1-215) was PCR amplified with primers GGGAATTCCATATGGATTCCAACACTGTGTCAA GTTTTCAGGTAG and CCGCTCGAGAGTAAGTGGAGGTCTCCCATTCTCATTACTGCT TCC. And the product was digested with NdeI and Xhol restriction endonucleases, and cloned into pET43.1a plasmid DNA (Novagen) cleaved with NdeI and Xhol. The first 73 amino acids of NS1A was PCR amplified with primers CATGCCATGGATTCCAACACTGTGTCAAGTT TTCAGGTAG and CCGCTCGAGGCCCTGAAAATACAGGTTTTCAGATTCTTCCTTC. The TEV protease recognition site was engineered in the 3' primer. Then the PCR product was digested with NcoI and Xhol restriction endonucleases, and cloned into pET28b plasmid DNA (Novagen) cleaved with NcoI and Xhol. The integrity of all cloned DNA(s) were confirmed by DNA sequencing.

2.2.2 Construction of plasmid mutations

The mutations were introduced by site-directed mutagenesis, according to the Stratagene protocol (Stratagene). Around 50 ng of plasmid and 150 ng of each primer were combined with reaction buffer (20 mM Tris–HCl (pH 7.5), 8 mM MgCl2, 7.5 mM DTT, 50 μg/ml of bovine serum albumin (BSA)), 150 μM dNTP mix, 1 units of KOD Hot Start DNA Polymerase (Novagen), and deionized water to final volume of 50 μl. Primer sequences of NS1A mutant (R38A) and NS1B mutants (R50A, R53A and R50A-R53A) are shown in table 2.1. Reaction mixture was further treated with 10 units of DpnI (New England Biolabs) at 37 °C for 2hrs. Then 1 μl of treated reaction mixture was transformed into E. coli DH5α competent cells. The presence of the expected mutations was confirmed by DNA sequencing.

2.2.3 Protein purification

This purification procedure is the same for all proteins expressed from genes cloned in pGEX vector. *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) containing the plasmid was grown overnight in broth medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v)

NS1A mutant	Primer sequence
R38A-5'	CATTTCTTGATCGGCTTCGCGCGGATCAGAAGTCCCTAAG
R38A-3'	CTTAGGGACTTCTGATCCGCGCGAAGCCGATCAAGAAATG

NS1B mutants	Primer sequence
R50A-5'	CAAGACCGCCTAAACGCACTAAAGAGAAAATTAG
R50A-3'	CTAATTTTCTCTTTAGTGCGTTTAGGCGGTCTTG
R53A-5'	CCTAAACAGACTAAAGGCAAAATTAGAGTCAAG
R53A-3'	CTTGACTCTAATTTTGCCTTTAGTCTGTTTAGG
R50A-R53A-5'	CAAGACCGCCTAAACGCACTAAAGGCAAAATTAGAGTCAAG
R50A-R53A-3'	CTTGACTCTAATTTTGCCTTTAGTGCGTTTAGGCGGTCTTG

Table 2.1 Primer sequences of NS1A and NS1B mutants

NaCl) containing ampicillin (100 μg/ml) at 37 °C. The cells were then diluted 1:100 in 4 × 0.5 l medium and grown to a cell density of approximately 4×10^8 per ml. IPTG was added to final 1 mM and the cells then grown overnight at 25 °C. Protein was purified according to a procedure developed by Sigma Aldrich. Briefly, cells were collected by centrifugation, resuspended in 50 ml column buffer (CB: 50 mM HEPES (pH 7.5), 50 mM NaCl) and disrupted in a French pressure cell. Cellular debris was pelleted by centrifugation at 5,000g for 60 min at 4 °C. The supernatant was applied to a column containing glutathione agarose beads (3 ml bed volume, equilibrated with CB), and washed with 100 ml of CB, then eluted with 10 ml elution buffer (EB: 10mM reduced glutathione, 50 mM Tris, pH 8.0). Fractions containing the GST fusion protein were pooled together and dialyzed against FP binding buffer (50 mM Tris (pH 8.0), 50 mM KCl, 1 mM DTT, 0.02% NaN_3). The protein was concentrated by using an Amicon bioseparator fitted with a 25 mm YM-3 membrane and applying 60–70 lb/in2 N2 for 7–10 h at 4 °C.

This purification procedure is the same for all proteins expressed from genes cloned into the pET28b vector. *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) containing the plasmid was grown overnight in broth medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) at 37 °C. The cells were then diluted 1:100 in 2 l medium and grown to a cell density of approximately 4×10^8 per ml. IPTG was added to 1 mM and the cells then grown overnight at 25 °C. Protein was purified essentially according to a procedure developed by Novagen. Briefly, cells were collected by centrifugation, resuspended in 25 ml column buffer (CB: 50 mM Hepes, pH 8.0, 300 mM NaCl, 20 mM Imidazole) and disrupted in a French pressure cell. Cellular debris was pelleted by centrifugation at 5000 xg for 60 min at 4 °C. The supernatant was applied to a column containing Ni-NTA His binding beads (2 ml bed volume, equilibrated with CB), and washed with 100 ml of CB, then eluted with 10 ml elution buffer (EB: 50 mM Hepes, pH 8.0, 300 mM NaCl, 250 mM Imidazole). The C-terminal His tag were removed by digestion with TEV protease (50 Unit) overnight at room temperature. The sample then underwent size-exclusion chromatography with a 50 mM Tris (pH 8.0), 300 mM KCl, 1 mM DTT, 0.02% NaN_3 buffer. Fractions containing the target protein were pooled together and dialyzed against FP binding buffer (50 mM Tris (pH 8.0), 50 mM KCl, 1 mM DTT, 0.02% NaN_3). The protein

was concentrated to 1–2 ml using an Amicon bioseparator fitted with a 45 mm YM-3 membrane and applying 60–70 lb/in^2 N$_2$ at 4 °C.

2.2.4 Fluorescence polarization (FP) assay

All fluorescence measurements were made using an Envision (Perkin Elmer Inc) spectrofluorometer. The two 16 mer RNA oligonucleotides with sequence CCAUCCUCUA CAGGCG (sense) and CGCCUGUAGAGGAUGG (antisense) were (5′-fluorescein) labeled. Approximate 1:1 molar ratios of sense RNA and antisense RNA were mixed in 50 mM Tris (pH 8.0), 100 mM NaCl buffer. The mixed solutions were heated to 90 °C for 2 min and then slowly cooled to room temperature to anneal the duplexes. Samples were excited at 490 nm, and emitted light was collected through an orange glass filter (OG 515, Schott). The binding affinity for the NS1 and double stranded (ds)RNA complexes were determined by measurement of the steady-state anisotropy of fluorescence as a function of added protein. The concentration of dye labeled dsRNA was 5 nM. The binding buffer contains 50 mM Tris (pH 8.0), 50 mM KCl, 1 mM DTT, 0.02% NaN$_3$. The concentration of protein was plotted against anisotropy of fluorescence and fit to a hyperbola equation to get the dissociation constants. The nonlinear regression analysis was performed in the program GraFit 5.0 (Erithacus Software).

2.2.5 Unlabelled dsRNA competition assay

The FP assay was performed with 5 nM fluorescein labeled dsRNA, 300 nM GST-NS1A(1-215) and different concentrations of unlabeled dsRNA (ranging from 0 nM to 400 nM).

2.2.6 Effect of DMSO, tRNA, BSA coating

The FP assay was performed with 5 nM fluorescein labeled dsRNA, 300 nM GST-NS1A(1-215), and in the presence/absence of 1% DMSO, 50 ng/μl tRNA, and 2% BSA coating respectively.

2.2.7 Z' factor calculation for high throughput assay

For a 384 well plate, the first 12 columns are loaded with a 300 μM protein and 5 nM dsRNA mix, and the last 12 columns are loaded with dsRNA only. FP measurements were made, and the Z' factor calculated according to this equation: Z' = 1-(3*SD$_f$+3*SD$_b$)/(U$_b$-U$_f$). Here, SD stands for the standard deviation of the free (SD$_f$) and bound (SD$_b$) samples respectively. U stands for average value of the free (U$_b$) and bound (U$_f$) samples respectively.

23

2.3 Results

2.3.1 FP results for binding to various NS1A constructs

Figure 2.1 shows the results of the FP assay for binding a 5'-fluorescein labeled dsRNA to various NS1A constructs, including NS1A(1-73), N-terminal GST fusion NS1A(1-215), and C-terminal His-tagged NS1A(1-215). All FP values were expressed in millipolarization (mP) units, and plotted against protein concentration (μM). As expected from previous studies, the R38A mutant of both NS1A(1-73) and GST fusion NS1A(1-215) did not cause any change in the FP signal of the dye labeled dsRNA. That is, the mutant protein failed to bind dsRNA.

Table 2.2 summarizes the binding affinities of dsRNA to the three NS1A constructs, at two pH values. The N-terminal GST fusion NS1A(1-215) protein show the tightest binding with a K_d around 0.16 μM. While the C-terminal His-tagged NS1A(1-215) and the NS1A(1-73) bind to dsRNA almost 100 fold more weakly than the GST fusion protein. By lowering the pH from 8.0 to 6.0, the NS1A(1-73) could achieve the same bind affinity toward dsRNA as the GST fusion protein. The GST fusion protein also leads the best yield of expression from E. coli. The yield of NS1A(1-73) peptide was not as high and required more purification steps, in addition to removing the C-terminal his tag. The yield of the C-terminal His-tagged NS1A(1-215) was very low because most of protein was found in the insoluble fraction, and presumably denatured.   Finally, in establishing our FP assay, we showed that the 5'-fluorescein labeled dsRNA bound to the GST-NS1A(1-215) could be successfully displaced by unlabelled dsRNA, as shown in figure 2.2; this showed that in principle at least, the construct could reveal the binding of ligands that compete with dsRNA. Overall, the GST-NS1A(1-215) construct was the best candidate for FP assay against NS1A.

24

(a)

(b)

(c)
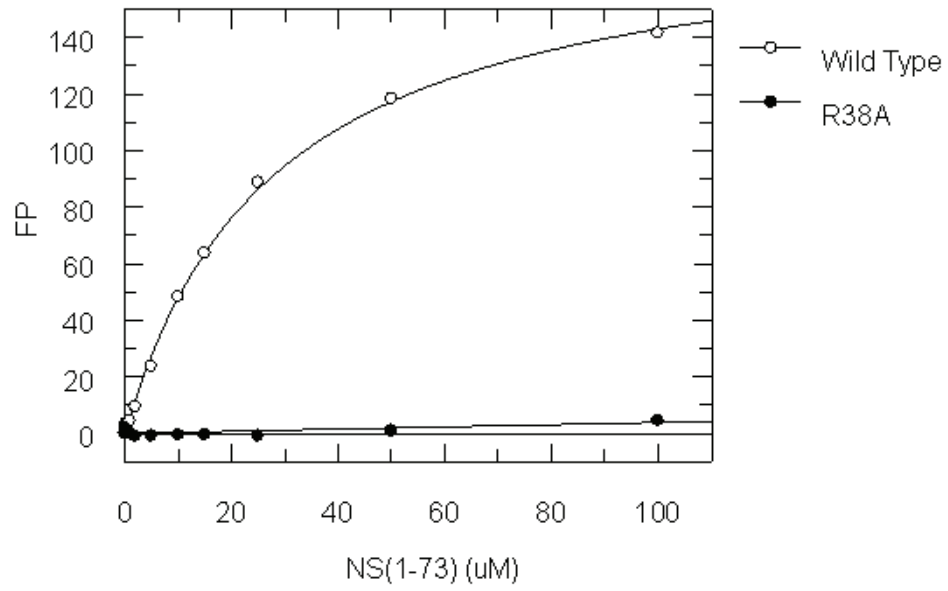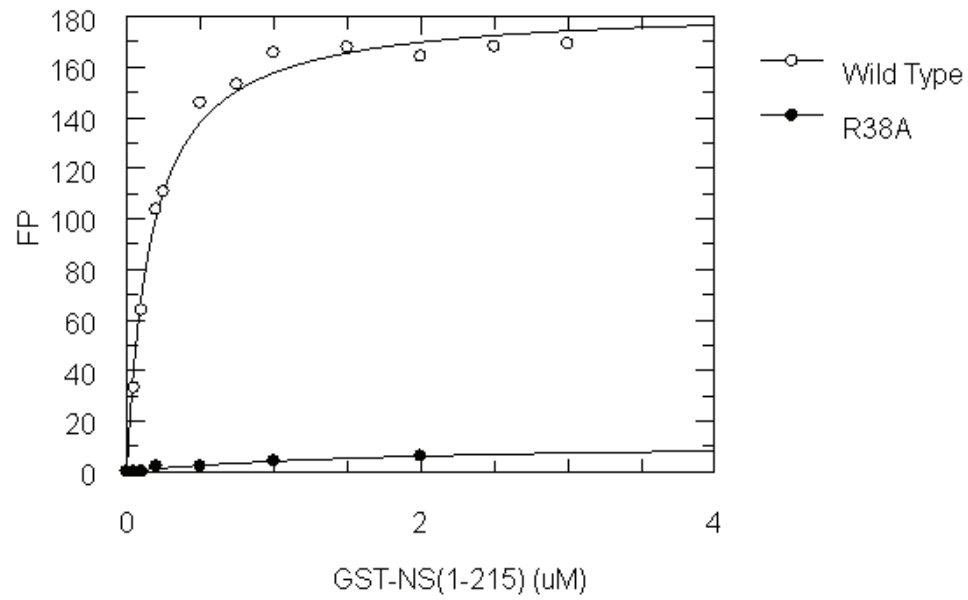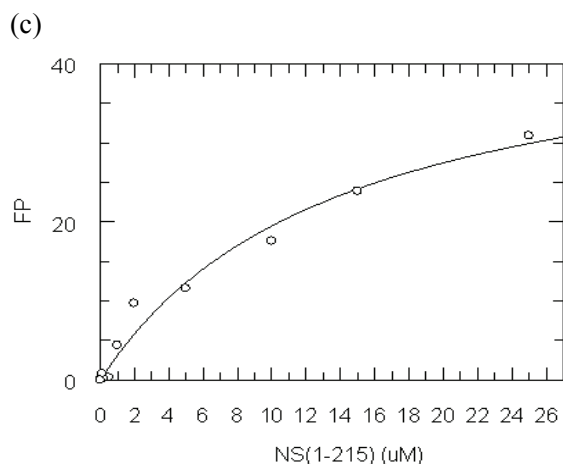
Figure 2.1: FP signal of 5'-fluorescein labeled dsRNA binding to the (a) NS1A(1-73), (b) GST-NS1A(1-215), (c) NS1A(1-215). The wild type is shown as open circle and the corresponding R38A mutant as a filled circle.

	GST-NS1A(1-215) (pH 8.0)	NS1A(1-215) (pH 8.0)	NS1A(1-73) (pH 8.0)	NS1A(1-73) (pH 6.0)
K_d (μM)	0.16	13	14	0.2
Yield	++	--	+	+

Table 2.2: Binding affinity for dsRNA and expression yield in E.coli of various NS1A constructs.
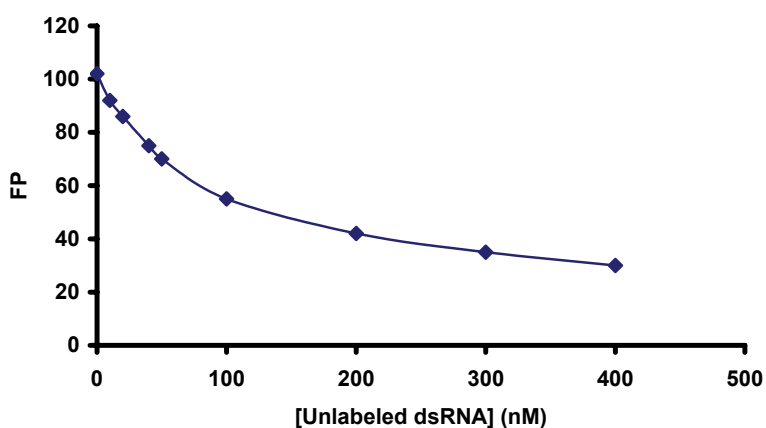
Figure 2.2: Competitive assay between labeled (5nM) and unlabeled dsRNA with GST-NS1A(1-215) (300nM).

26

2.3.2 FP results of various NS1B constructs (including wild type and mutants)

Figure 2.3 showed the results of FP assays for binding a 5'-fluorescein labeled dsRNA to various N-terminal GST fusion NS1B constructs, including GST-NS1B(1-104), GST-NS1B(1-145) and GST-NS1B(1-218). All FP values were expressed in millipolarization (mP) units, and plotted against protein concentration (μM). As expected, both the R50A mutant of GST-NS1B(1-104) and the double mutant R50A-R53A of GST-NS1B(1-145) failed to cause any change in FP signal, that is did not bind the oligonucleotide. Table 2.3 summarizes the binding affinities of dsRNA to all three NS1B constructs. The N-terminal GST fusion NS1B(1-145) protein showed the tightest binding with a K_d around 0.05μM. While the GST-NS1B(1-104) not only bind dsRNA around 30 fold weaker, but also showed much weaker FP signal than the GST-NS1B(1-145). As shown in figure 2.3 panel a, the maximum change of mP unit is too small to constitute a significant positive signal. Although GST-NS1B(1-218) showed similar binding affinity toward dsRNA as that of GST-NS1B(1-145), the yield of that protein was extremely low. In addition, the impurity of GST-NS1B(1-218) made it hard to accurately quantify the amount of the protein used in the assay. Overall, the GST-NS1B(1-145) construct was the best candidate for FP assay against NS1B.

2.3.3 The effect of additives on the FP assay (including DMSO, tRNA and BSA)

All the compounds in our library collections are dissolved in 100% DMSO to a final concentration of ~10mM. As a consequence, DMSO will be transferred to all inhibitors assays. The effect of DMSO concentration on fluorescence polarization assay was tested. As shown in Figure 2.4, up to 1% of DMSO is tolerated by the fluorescence polarization assay. Above 1%, a significant decrease of polarization was observed, probably due to the denaturation of the double stranded RNA. This means that compounds from the libraries must be diluted at least 100 fold, producing a maximum assay concentration of ~100 μM. As shown in figure 2.5 panel a, the titration curves of FP signal vs. protein concentration in the presence and absence of 1% DMSO can almost overlap with each other.

In a common protein dsRNA binding assay, 50 ng/μl tRNA was used to prevent non-specific binding of dsRNA to protein (Chien, 2004). Our FP assay showed that there was no strong impact of tRNA on the assay, see figure 2.5 panel b. There is also a concern that the protein-dsRNA complex may stick to the surface of the well, which may give a false positive signal. To address this concern, two percent BSA was used to coat the 384 well plate before

(a)

(b)

(c)
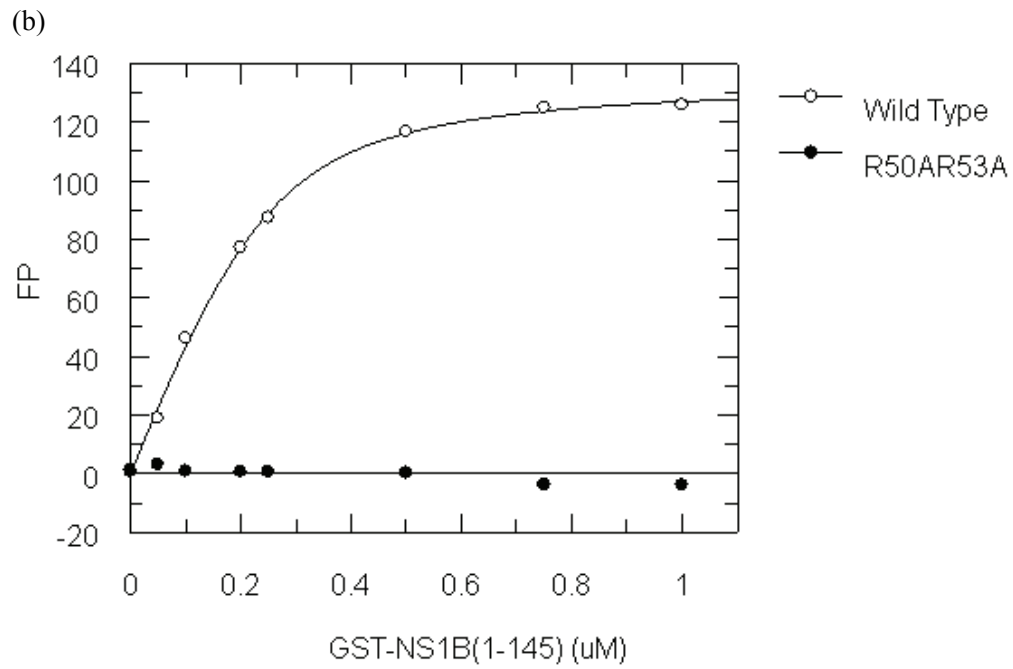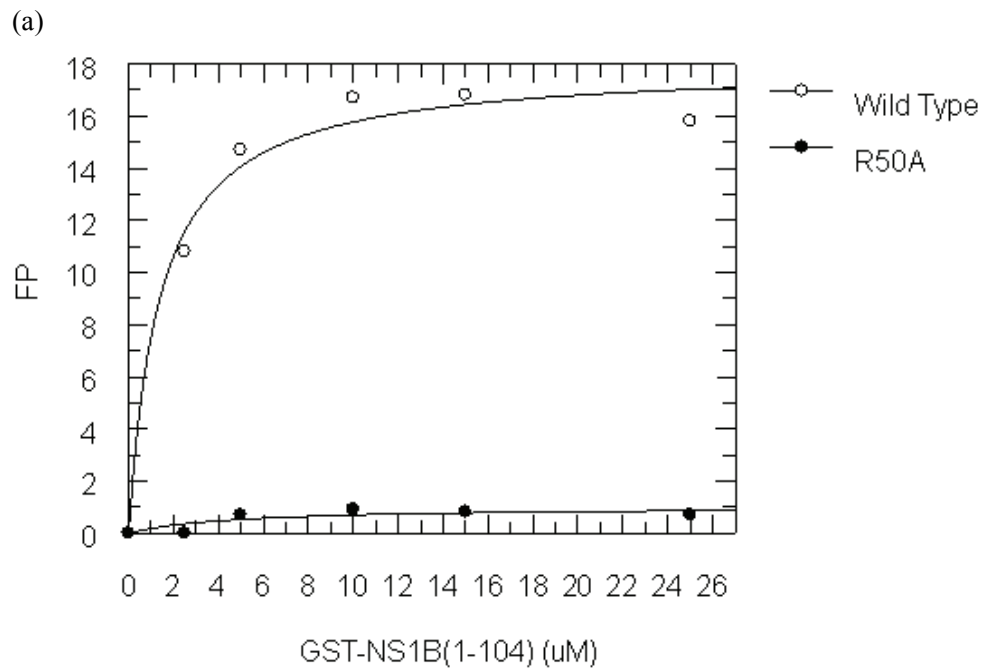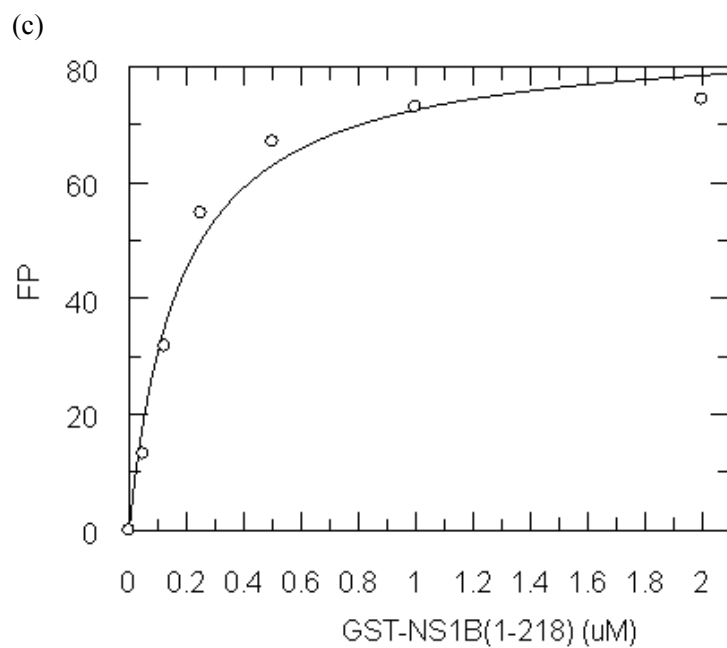
Figure 2.3: FP signal of 5'-fluorescein labeled dsRNA binding to the (a) GST-NS1B(1-104), (b) GST-NS1B(1-145), (c) NS1B(1-218). The wild type is shown as open circle and the corresponding R38A mutant as a filled circle.

	GST-NS1B(1-104)	**GST-NS1B(1-145)**	**GST-NS1B(1-218)**
K_d (μM)	1.4	0.05	0.18
Yield	+	++	--

Table 2.3: Binding affinity for dsRNA and expression yield in E.coli of various NS1B constructs.

(a)
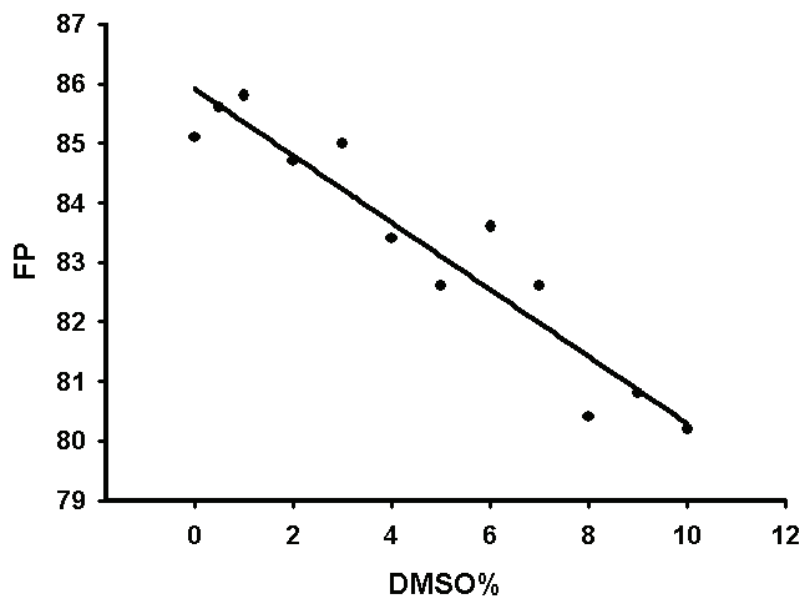
(b)
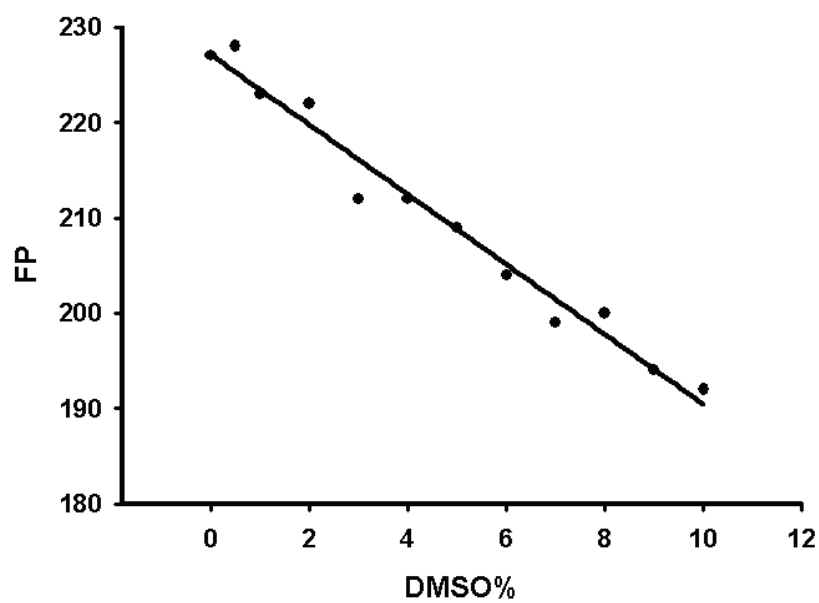
Figure 2.4: Effect of DMSO concentration on FP signal of (a) dsRNA and (b) complex of dsRNA with GST-NS1A(1-215)

(a)

(b)

(c)
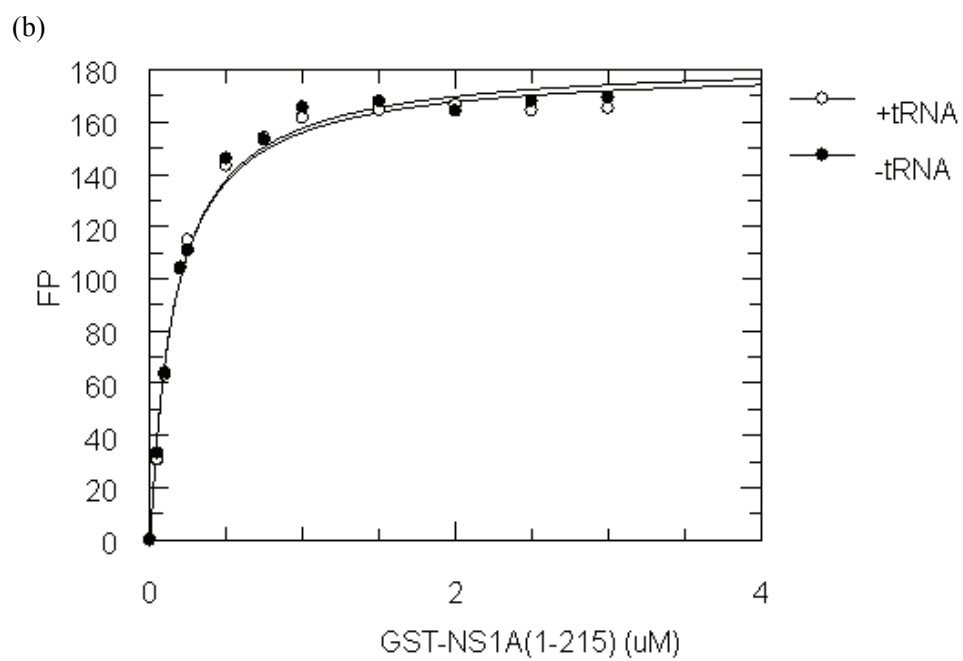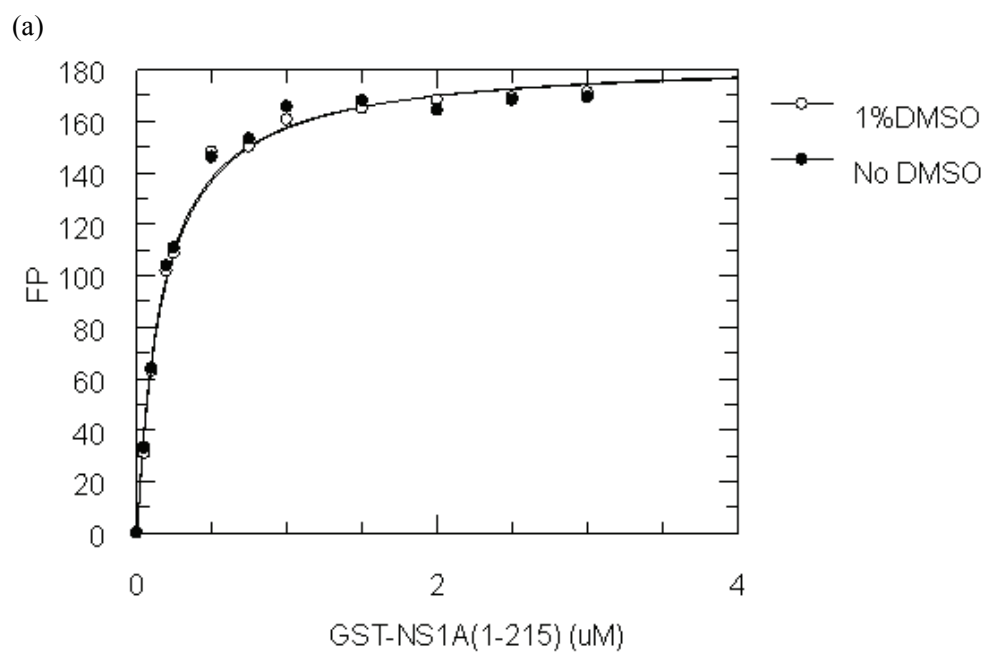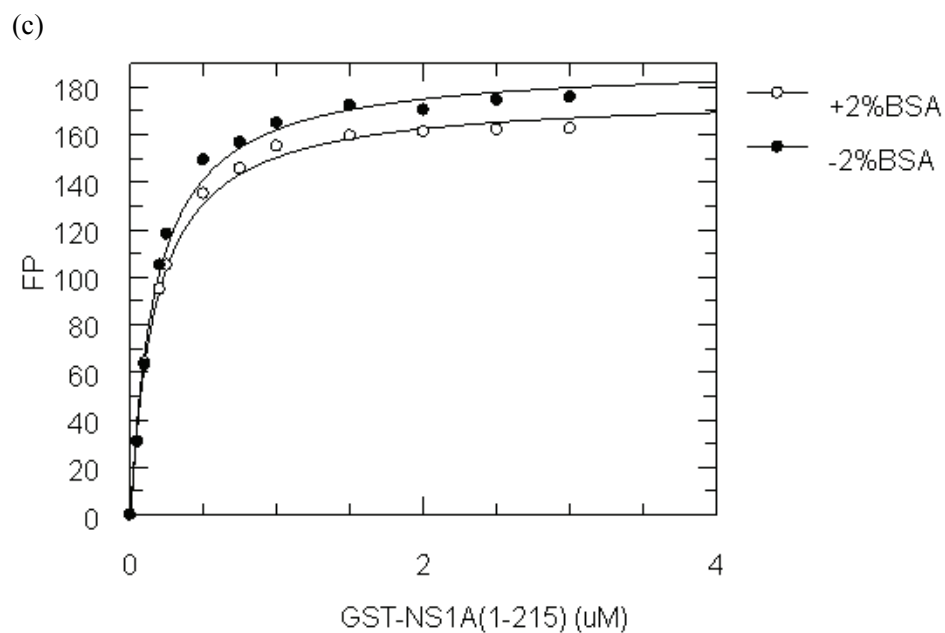
Figure 2.5: FP signal of 5'-fluorescein labeled dsRNA binding to the GST-NS1A(1-215) in the presence and absence of (a) 1%DMSO, (b) 50ng/μl tRNA, and (c) 2%BSA coating. The open circles represent the FP signal in the presence of corresponding additives, and the filled circles represent the FP signal in the absence of corresponding additives.

adding the binding solution mix; the FP assay results showed that the change of binding constants with or without BSA coating were negligible, see figure 2.5 panel c.

2.3.4 Z' factor for the high throughput FP assay

Before an experimental screening is carried out it is essential to quantify the robustness of the assay. For this purpose the Z' factor of this high throughput assay was calculated. The Z' factor is computed from experimental observations, via the expression $Z = 1 - (3*SD_f + 3*SD_b)/(U_b - U_f)$. Basically, Z' factor is a measure of the quality of a high-throughput screening (HTS) assay (Arai 2001). An assay can be considered validated for high throughput screening after three independent experiments have been shown to result in reproducible and suitable Z' factor values. If the Z' factor is above 0.9, it means the assay is an excellent one. If it is between 0.7 and 0.9, then it is a good one. If Z' factor is below 0.5, that means the assay is not a working one at all. The Z' factors of our assay from three independent experiments are 0.87, 0.88 and 0.85 respectively. Therefore, the average Z' factor for our assay is 0.87, which means our high throughput assay is a good one.

2.4 Discussion

Fluorescence polarization (FP) assays are among the most widely used high throughput screening methods for drug discovery (Roehrl, Wang et al. 2004; Roehrl, Wang et al. 2004; Rishi, Potter et al. 2005). It is compatible with the requirements of high throughput assay, which includes: (1) minimal experiment steps; (2) stable and homogeneous experiment reagents; (3) highly sensitive and rapid readout; (4) amenable to automation (Roehrl, Wang et al. 2004; Zhang, Huang et al. 2006). Here we used FP assay to determine the binding affinity of fluorescein-labeled dsRNA to different NS1A /NS1B constructs. Usually, the concentrations of the fluorophore and its binding partner are the major factor determining the interference of the potential inhibitors. A high concentration of fluorescein-labeled dsRNA and NS1 complex would result in reduced sensitivity for weak inhibitors. To maximize the effective concentration of potential inhibitors, we were looking for the NS1 construct that allowed us to use a low protein concentration. In other words, we chose the NS1 construct with the highest binding affinity to dsRNA.

Our FP assay showed that the N-terminal GST fusion NS1A(1-215) is the best choice among all the NS1A constructs. There is a serious aggregation problem for C-terminal His-tagged NS1A(1-215) protein during preparation, which makes it impossible for large scale

33

purification. The RNA binding domain of NS1A, NS1A(1-73), binds to dsRNA almost 100 fold weaker than does the GST-NS1A(1-215) fusion construct. Decreasing the pH of the buffer from 8.0 to 6.0 could achieve the same binding for the NS1A(1-73) toward dsRNA as that of the GST fusion protein. However, fluorescein is strongly quenched at pH 6, reducing its utility as a probe. As shown in Figure 2.6, the intensity of fluorescence dropped by three fold when pH is 6.0, which means that we would have to use three fold more of reagents at pH6.0 to achieve the same amount of signal intensity at pH8.0. It's unclear why the GST-NS1A(1-215) binds dsRNA much tighter than NS1A(1-73). Since the R38A mutant of GST-NS1A(1-215) didn't bind dsRNA at all, the N-terminal GST domain is obviously not directly involved in the interaction with dsRNA. It is known that GST behaves as a dimer in solution. It's very likely that the N-terminal GST protein facilitates the dimer formation of NS1A(1-215). That could explain why the GST fusion protein has the highest binding affinity toward dsRNA. The GST-NS1B(1-145) showed the tightest binding affinity toward dsRNA with a K_d around 0.05μM. The other two GST fusion NS1B proteins were excluded mainly because of the poor purity and low yield.

The FP assay conditions were optimized with different additive chemicals, including DMSO, tRNA and BSA. DMSO is frequently used as a solvent for commercial chemical compound library due to its excellent solvating power. To maximize the effective concentration of potential inhibitors, we were looking for the highest DMSO concentration that our assay could tolerate. It turned out that up to 1% DMSO can be tolerated in our assay. Other potential assay complications, like non specific RNA binding, were assessed using tRNA addition and BSA coating. We established a useful assay with an excellent Z' factor of 0.87. In summary, the fluorescence polarization assay was optimized and miniaturized for high throughput screening.
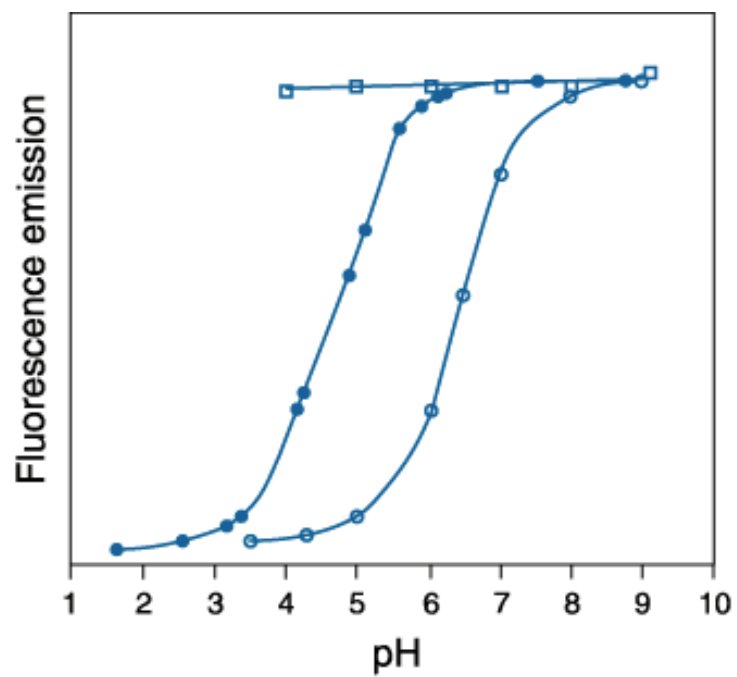
Figure 2.6: Comparison of pH-dependent fluorescence of the Oregon Green 488 (●),
carboxyfluorescein (◔) and Alexa Fluor 488 (☐) fluorophores. Fluorescence intensities were
measured for equal concentrations of the three dyes using excitation/emission at 490/520 nm.
Cited from http://probes.invitrogen.com/handbook/figures/0495.html

Chapter 3: High Throughput Screening (HTS) of Compound Libraries for NS1 Inhibitors

3.1 Introduction

The high throughput fluorescence polarization assay is typically designed as competitive equilibrium binding assay. It usually detects changes in polarization caused by changes in the effective mass of the fluorescently labeled molecule.   As molecular mass increases, as when the labeled molecule binds to a target protein, polarization decreases. When a small ligand competes successfully with the fluorescently labeled molecule, the dye is displaced, tumbles more quickly, and polarization increases. Here, a high throughput FP assay was developed to screen small compound libraries against dsRNA binding by the NTD domain of NS1 protein. Fluorescein-labeled dsRNA is pre-incubated with the N-terminal GST fusion to NS1A (residue 1-215) protein. Then the decrease of fluorescence polarization indicated the displacement of dsRNA by potential inhibitors. Our goal was to find a small molecule that can bind to NS1A in such a way that it precludes RNA binding. The compound libraries chosen to be screened are: ChemBridge fragment library (4000 compounds), ChemBridge kinase library (11250 compounds), NIH clinical trial collection (446 compounds), and Microsource Discovery spectrum collection (2000 compounds).

3.2 Materials and methods

3.2.1 Protein purification

The GST fusion protein NS1A(1-215) and NS1B(1-145) were expressed and purified according to the method described in chapter 2.

3.2.2 Fluorescence polarization and high-throughput screening

High-throughput FP assays were performed on an Envision spectrofluorometer (Perkin Elmer) using an excitation filter of 490 nm, and an emission filter of 540 nm. The blank correction and fluorescence gain were calibrated. Assays were conducted in 384-ShallowWell black polystyrene plates (Nunc). The "blueprint" of an assay plate is shown in figure 3.1. The FP assay was performed with 5 nM fluorescein labeled dsRNA, 300 nM GST-NS1A(1-215) in 50 mM Tris (pH 8.0), 50 ng/μl tRNA, 50 mM KCl, 1 mM DTT, 0.02% NaN$_3$ buffer. A liquid
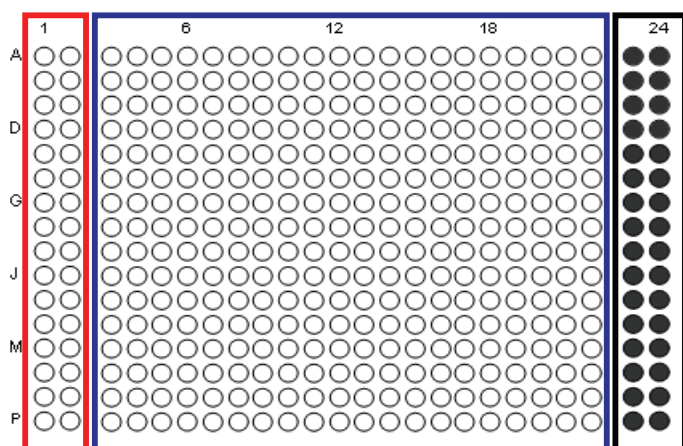
Figure 3.1: Blueprint of assay plate. The dsRNA and protein mixtures were loaded in the first two columns to serve as positive controls (marked in red frame). The dsRNA alone was loaded in the last two columns to serve as negative control (marked in black frame). The compounds were transferred into the dsRNA protein mixture in the middle section (column 3 to 22; marked in blue frame). Each plate can accommodate 320 ligand tests.

handling automatic working station, a Janaus robot (Perkin Elmer), was used to aliquot 20 μl protein-dsRNA mixtures into each well, and to transfer 0.1 μl of test compounds from their storage plates into the 20 μl mixture. The final inhibitor concentration was a nominal 50 μM, and the final concentration of DMSO will be 0.5%. It must be noted however, that the solubility of the compounds in the new aqueous environment is unknown and it is likely that most compounds are actually present at concentrations below the nominal value. Data analysis and data mining were performed using software developed by Collaborative Drug Discovery Inc. For secondary confirmation, FP assays were performed with cherry-picked compounds at two different concentrations (10 μM and 50 μM).

3.2.3 Dose response assay for "hit" compounds

A compound giving a decrease in FP in a HT assay is a possible ligand or "hit". Hit compounds were subsequently tested in a dose response assay; they were diluted serially in binding buffer (50 mM Tris (pH 8.0), 50 ng/μl tRNA, 50 mM KCl, 1 mM DTT, 0.02% NaN$_3$), at a final concentration of DMSO of 0.5%. Like in the HTS assay, the concentration of fluorescein labeled dsRNA was 5 nM, and the concentration of protein (NS1A/NS1B) was 300 nM. We control for positional variation during HTS and dose response assay to minimize positional artifacts such as edge effects.

A concern is that many compounds in the libraries contain aromatic rings which might intercalate into dsRNA and give a false positive reading in this fashion. Hit compounds were also tested for their ability to intercalate by competing them against ethidium bromide, a well known intercalator, according to the method described by Boger (Boger, Fink et al. 2001). This assay assumes that if the compound binds to DNA/RNA, it will displace prebound ethidium bromide that can be measured by a decrease of fluorescence (Boger, Fink et al. 2001; Rishi, Potter et al. 2005). One DNA hairpin was prepared by IDT Inc, with sequence: TGACGTCAAAAAATGACGTCA. The final concentrations of hairpin and ethidium bromide were 2 μM and 1 μM respectively. Test compounds were added to the test plate with same concentration gradient as in the dose response assay.

3.2.4 Compound 4141340-spectrum scan

The spectrum scan of compound 4141340 and its derivatives was performed on a NanoDrop 2000c instrument (Thermo Fisher).

3.2.5 Circular dichroism measurements

Circular dichroism spectra were recorded using a Jasco 715 spectropolarimeter, equipped with a thermostated cell holder and a NesLab-111 circulating water bath. CD spectra were recorded in cells with an optical path length of 0.1cm. Experiments were performed in 25 mM NaH_2PO_4 (pH 6.5), 25 mM NaCl, and 25 mM NaN_3 buffer. Three scans were repeated for each experimental condition.

3.2.6 Thermal denaturation experiments

Thermal denaturation profiles were obtained by recording the temperature dependence of the ellipticity at 227 nm in the range 6–80 °C. The temperature was continuously changed at a rate of 0.5 °C/min. Experiments were performed in 25 mM NaH_2PO_4 (pH 6.5), 25 mM NaCl, and 25 mM NaN_3 buffer. Tm was determined by locating the maxima/minima of the first derivative of the curve describing the melting profile (CD versus T).

3.3 Results

3.3.1 High throughput screening results

A total of 17,969 compounds from four libraries have been screened against the RBD of NS1. From the ChemBridge fragment library, a total 52 hits were identified showing inhibition greater than 20% based on millipolarization change and +/- 30% fluorescence intensity change. As shown in table 3.1, six compounds among these showed inhibition greater than 50%, without any apparent intercalation into dsRNA. Among those six hits from the fragment library, only compound 4141340 (ChemBridge library number) and compound 5792605 showed 100% inhibition in the initial screen and a robust dose response activity, see Figure 3.5.

For the ChemBridge kinase library, a total 40 hits were identified showing inhibition greater than 50% based on polarization change and +/- 30% fluorescence intensity change. Among the 40 hits, there are six compounds showing inhibition greater than 60% at concentration of 50 μM (table 3.2), and one compound, 7869182, showing an IC_{50} around 0.8 μM against NS1A, and 0.6 μM against NS1B (Figure 3.2).

As shown in table 3.3, seven compounds were identified from the NIH clinical collection showing inhibition greater than 40% based on polarization change and +/- 30% fluorescence intensity change. The compound SAM001247031 is the only one from those seven hits giving an IC_{50} around 0.29 μM against NS1A, and 0.18 μM against NS1B without disrupting dsRNA.

As for the Microsource Discovery spectrum collection, there were eight compounds showing almost 100% inhibition at concentration of 50 μM (table 3.4), but only two
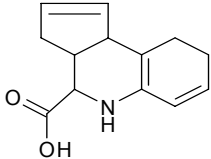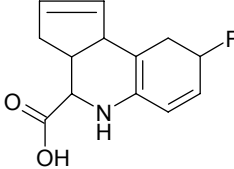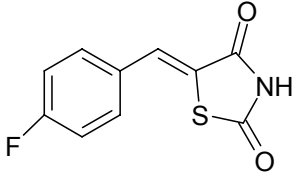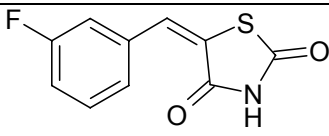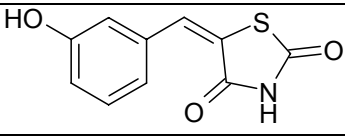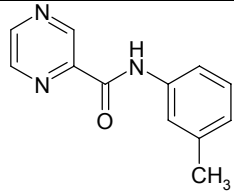
Compound ID	Structure	Inhibition% at 50μM
4141340		100
5792605		100
5265182		51
5378142		51
5378417		49
5556613		49

Table 3.1: Structures and inhibitory activities of selected compounds from ChemBridge fragment library.
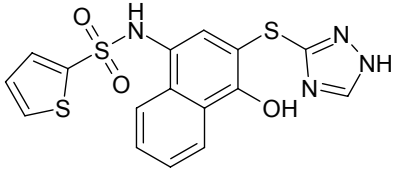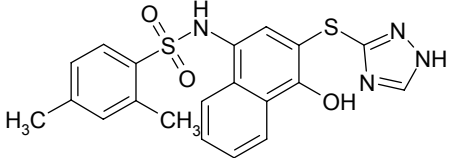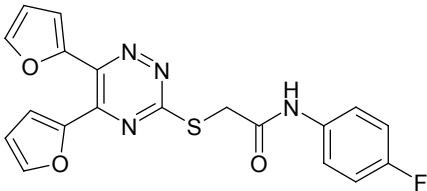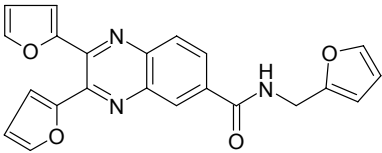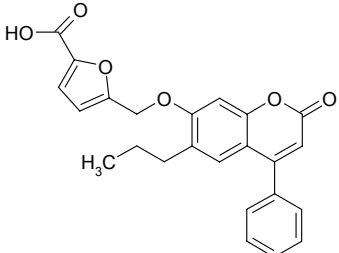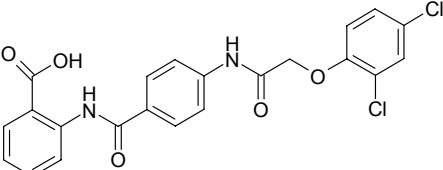
Compound ID	Structure	Inhibition% at 50μM
7796127		100
7788159		88
7560017		85
7869182		74
6628963		68
7971347		68

Table 3.2: Structures and inhibitory activities of selected compounds from ChemBridge kinase library.

Compound ID	Structure	Inhibition% at 50μM
SAM001246577		100
SAM001247031		89
SAM001246657		89
SAM001246818		86
SAM001246816		75
SAM001246892		65
SAM001246737		49

Table 3.3: Structures and inhibitory activities of selected compounds from NIH clinical collection.
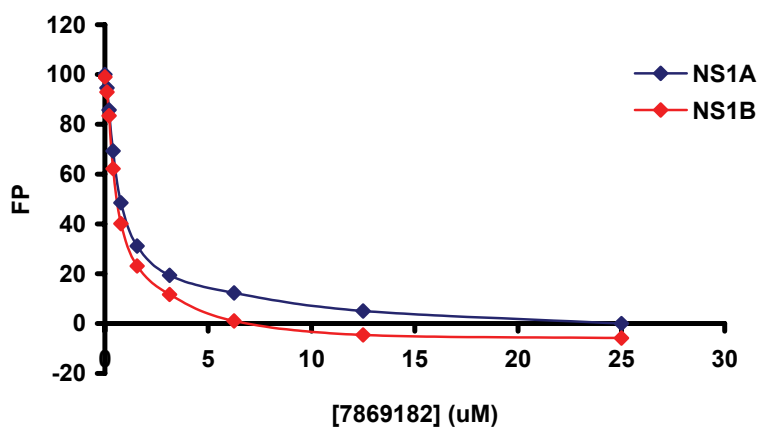
Figure 3.2: FP signal of dsRNA displacement by compound 7869182. The dose-response against GST-NS1A(1-215) is shown in blue, and the dose-response against GST-NS1B(1-145) is shown in red.
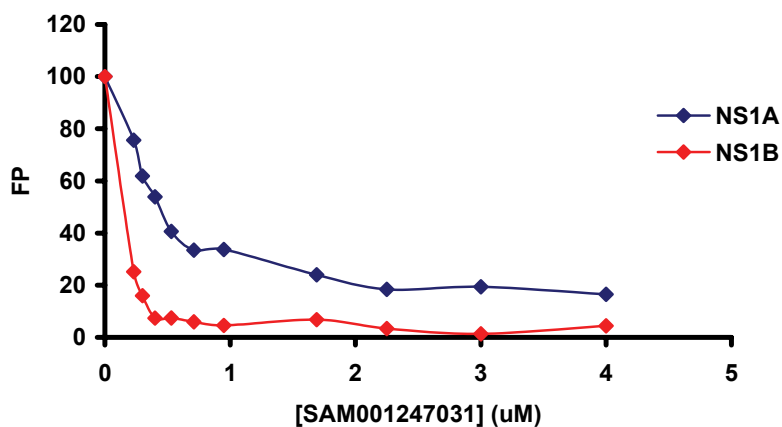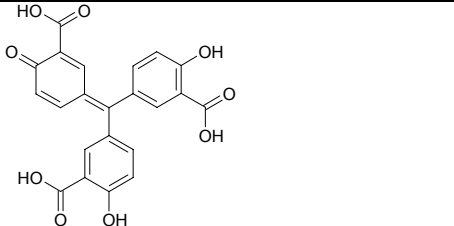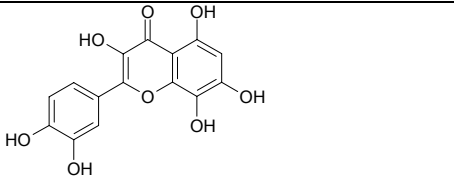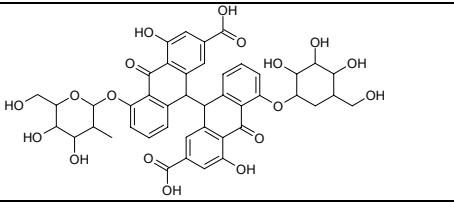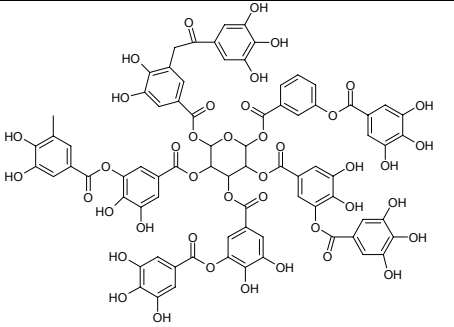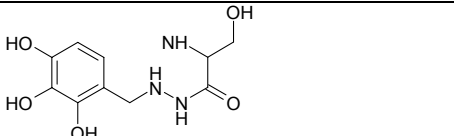
Figure 3.3: FP signal of dsRNA displacement by compound SAM001247031. The dose-response against GST-NS1A(1-215) is shown in blue, and the dose-response against GST-NS1B(1-145) is shown in red.

Compound Name	Structure	% Inhibition at 50μM
Aurothioglucose		100
Aurin tricarboxylic acid		100
Juglone		100
Gossypetin		100
Sennoside B		100
Tannic acid		100
Benserazide hydrochloride		100
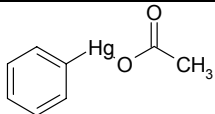
Phenylmercuric acetate		100

Table 3.4: Structures and inhibitory activities of selected compounds from Microsource Discovery spectrum collection.

(a)
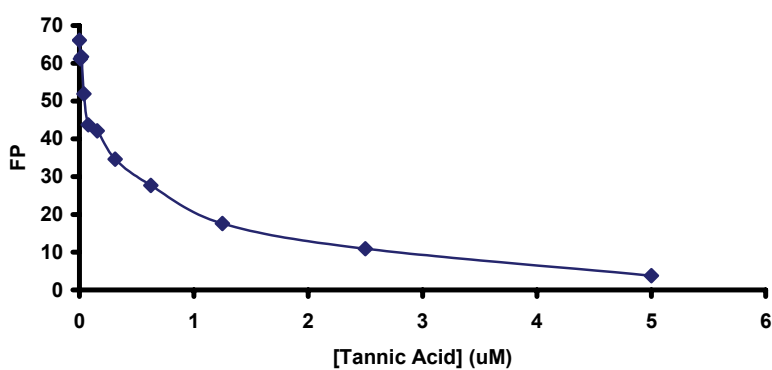
(b)
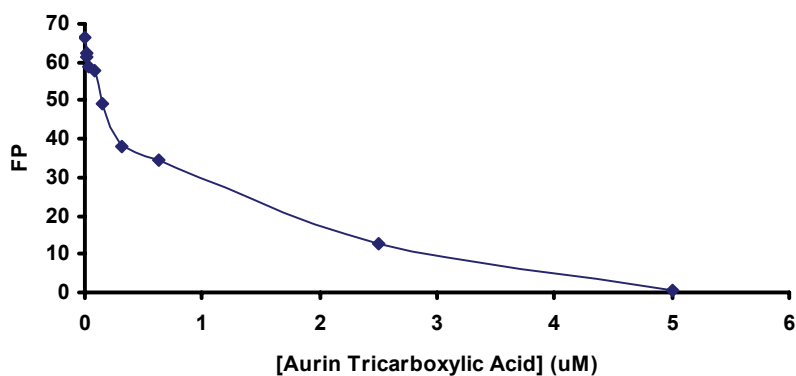
Figure 3.4: FP signal of dsRNA displacement against GST-NS1A(1-215) by compound (a) tannic acid, (b) aurin tricarboxylic acid.

compounds among those showed a significant dose response and did not disrupt dsRNA. They are tannic acid, with IC_{50} 0.48 μM against NS1A, and aurin tricarboxylic acid with IC_{50} 0.5 μM against NS1A, see Figure 3.4.

3.3.2 The dose response of compound 4141340 and its derivatives

The scientific name of compound 4141340 is 3a,4,5,9b-Tetrahydro-3H-cyclopenta[c] quinoline-4-carboxylic acid. Compound 5792605, is 8-Fluoro-3a,4,5,9b-Tetrahydro -3H-cyclopenta[c] quinoline-4-carboxylic acid; it is a derivative compound of 4141340. Three more derivative compounds of 4141340 were found using the program SciFinder Scholar: 8-Benzyl-3a,4,5,9b-tetrahydro-3H-cyclopenta[c] quinoline-4-carboxylic acid (4B), 8-Carboxyl-3a,4,5,9b-tetrahydro-3H-cyclopenta[c] quinoline-4-carboxylic acid (4C) and 8-Methyl-3a,4,5,9b-tetra-hydro-3H-cyclopenta [c] quinoline-4-carboxylic acid (4M). As shown in table 3.5, those five compounds have different functional groups at the position 8 of quinoline. All five compounds can displace dsRNA fully at around 20 μM concentration, as shown in the dose-response curves (Figure 3.5). The calculated IC_{50} against NS1A and NS1B were summarized in table3.5. Basically, they all gave an IC_{50} around 10 μM against either NS1A or NS1B.

When 4141340 was freshly dissolved in 100% DMSO, it did not show the same color as seen in the compound plate when purchased from the vendor. The freshly dissolved 4141340 showed a light brown color, while the one in the compound plate was green. Interestingly, the freshly dissolved 4141340 didn't show any inhibition effect, nor did any of the four derivatives. As shown in Figure 3.6, compound 4141340 and its derivatives changed color from light brown to dark green within a week. The dose response experiments shown in the above section were carried by using those aged compounds. A full spectrum scan from 220nm to 750nm was conducted for the freshly dissolved 4141340 and one-week old 4141340 respectively. A new absorbance peak at 624nm showed up in the spectrum profile of aged compound, which indicates that the compound might undergo some chemical changes (see Figure 3.7). The same absorbance peak at 624nm appeared for all four derivatives within four days.
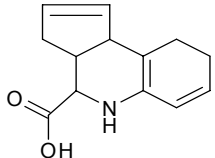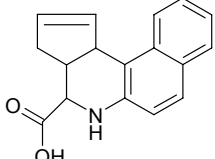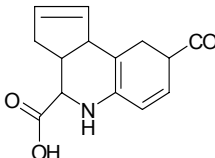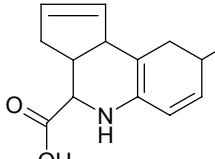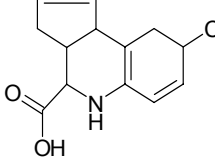
Compound Name	Structure	IC$_{50}$ (μM) [NS1A]	IC$_{50}$ (μM) [NS1B]
4141340 3a,4,5,9b-Tetrahydro -3H-cyclopenta[c] quinoline-4-carboxylic acid		8	8
4B 8-Benzyl- 3a,4,5,9b-tetrahydro -3H-cyclopenta[c] quinoline-4-carboxylic acid		5	5
4C 8-Carboxyl- 3a,4,5,9b-tetrahydro -3H-cyclopenta[c] quinoline-4-carboxylic acid		N/A	2.5
4F 8-Fluoro-3a,4,5,9b-tetrahydro -3H-cyclopenta[c] quinoline-4-carboxylic acid		7	5
4M 8-Methyl-3a,4,5,9b-tetrahydro -3H-cyclopenta [c] quinoline-4-carboxylic acid		9	5

Table 3.5: Structures and IC$_{50}$ of compound 4141340 and its four derivatives (4B, 4C, 4F and 4M) against GST-NS1A(1-215) and GST-NS1B(1-145).

(a)
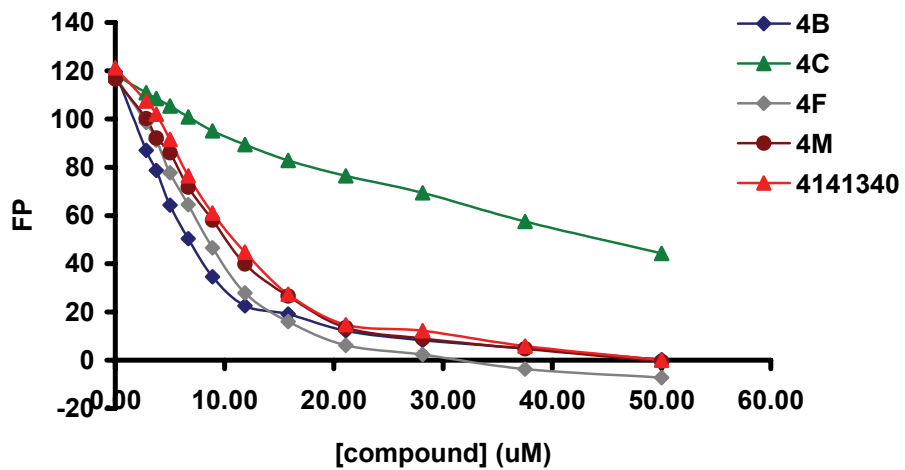
(b)
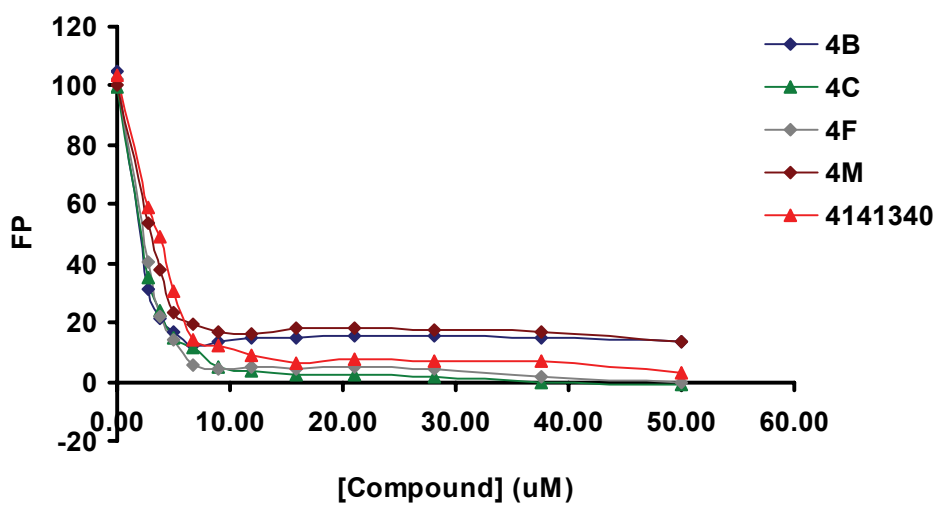
Figure 3.5: FP signal of dsRNA displacement against (a) GST-NS1A(1-215), (b) GST-NS1B(1-145) by compound 4141340 and its four derivatives (4B, 4C, 4F and 4M).

Day 1
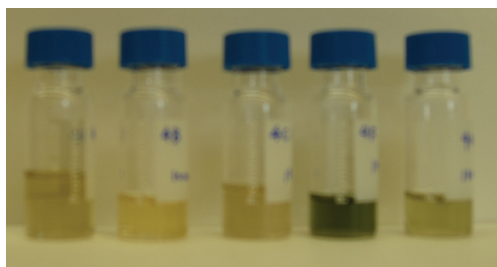
Day 2
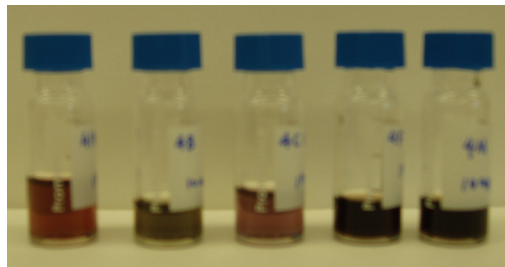
Day 3
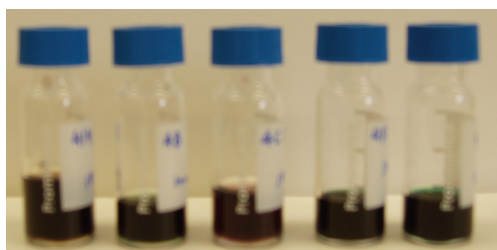
Day 4
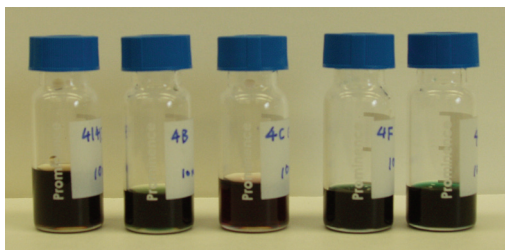
Day 5
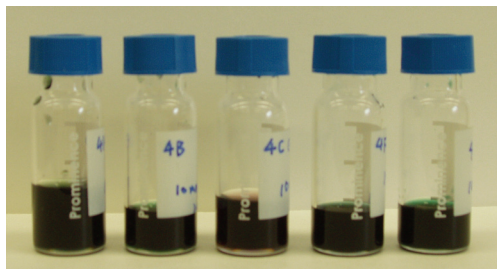
Day 6
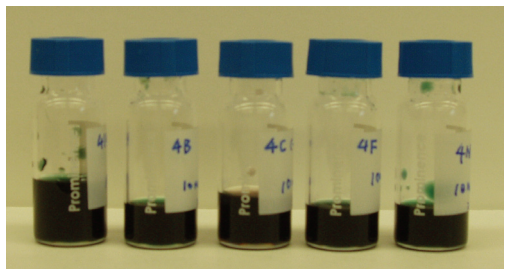
Figure 3.6: The color changing profile of compound 4141340 and its derivatives in six days. (From left to right are: 4141340, 4B, 4C, 4F and 4M.)
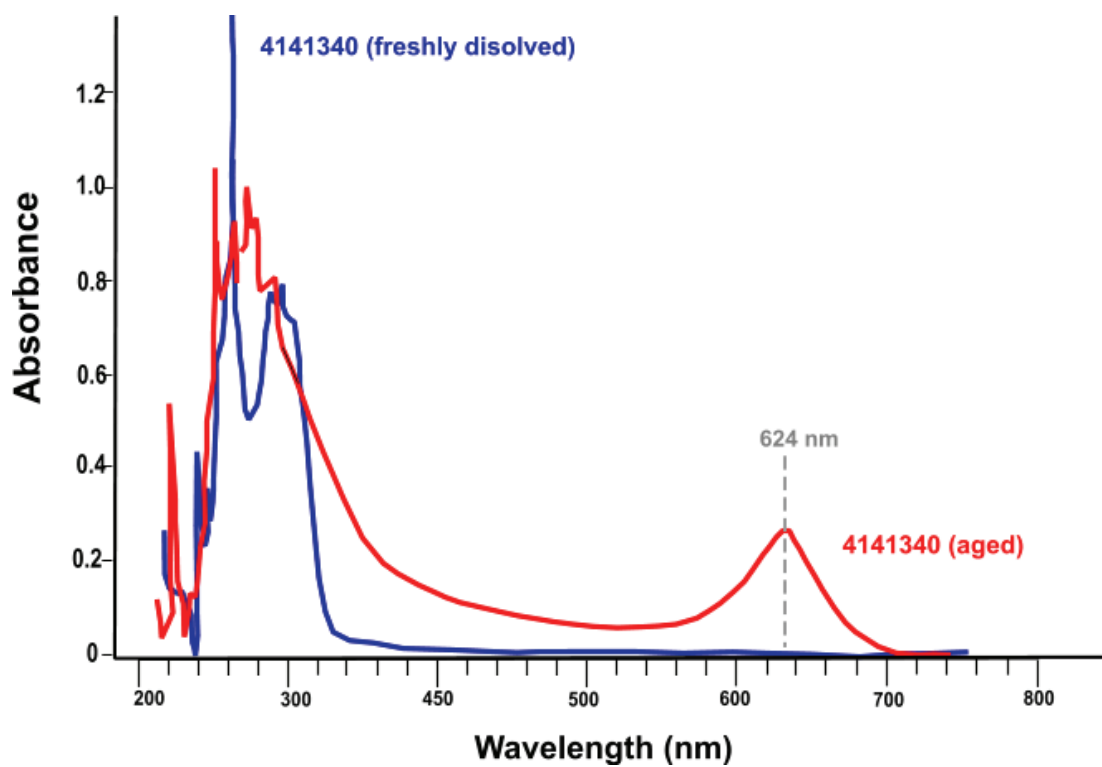
Figure 3.7: The full spectrum scan of freshly dissolved (blue) and aged (red) compound 4141340 from 220 nm to 750nm.

3.3.3 CD spectra and thermal melting profile of NS1A in the presence of SAM001247031

The structure of NS1A(1-73) was characterized chiroptically by circular dichroism. A quantitative analysis of the experimental CD profile suggests the structure is 68.6% α–helix and 10.7% random coil. This is in reasonable agreement with the observed X-ray structure of NS1A, which is 80% α-helix and 18% random coil. We then assessed the effect of compound SAM001247031 on the structure. As shown in Figure 3.8, SAM001247031 did not induce significant change in CD spectra even at concentration of 400 μM.

To explore the compound induced structural changes more fully, we carried out thermal denaturation experiments on both wild type and the R38A mutant NS1A(1-73) in the presence of different concentration of SAM001247031 (Figure 3.9). Monitoring ellipticity at 227 nm, we observed that SAM001247031 appears to stabilize the wild type NS1A(1-73), increasing the Tm by 11.05 °C at a concentration of 400 μM, see table 3.6. In contrast, SAM001247031 did not alter the Tm of the R38A mutant NS1A(1-73) significantly, only 1.15 °C even at a concentration of 400 μM. This huge Tm difference on wild type and R38A mutant proteins suggests that SAM001247031 stabilizes the structure of NS1A by interacting with the critical residue Arg38.
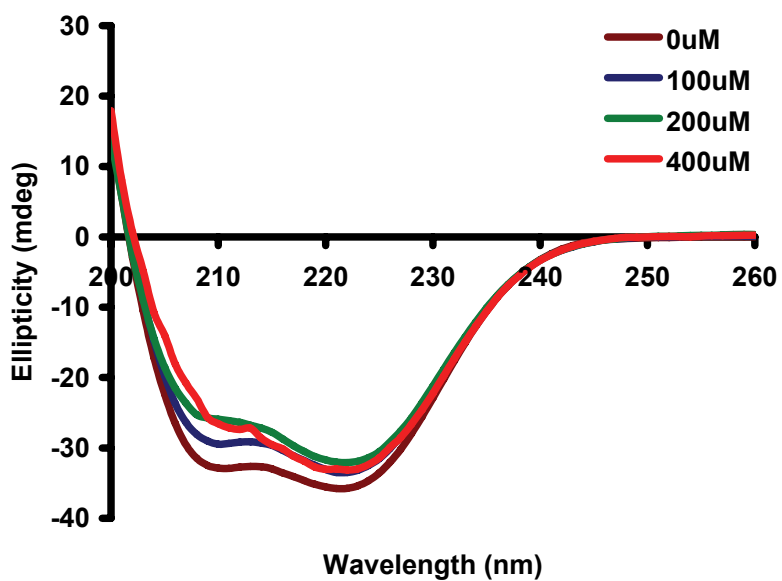
Figure 3.8: Effects of different concentration of compound SAM001247031 on NS1A(1-73) CD spectra at room temperature.

(a)
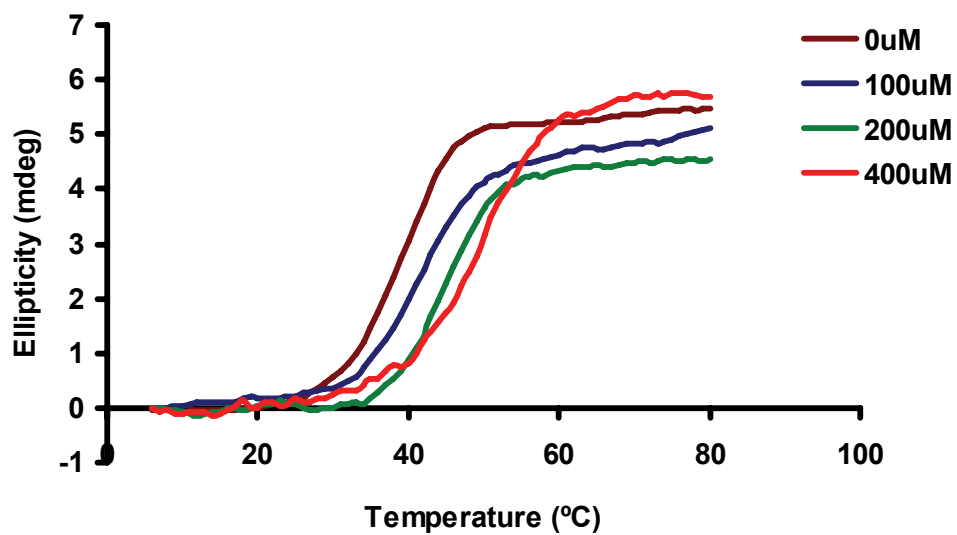
(b)
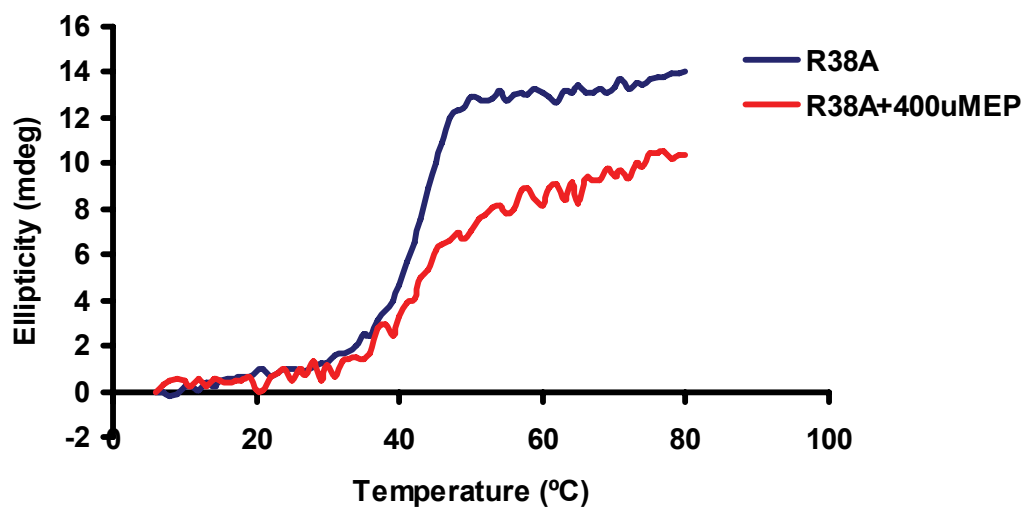
Figure 3.9: Effects of different concentration of compound SAM001247031 on the 227nm thermal melting profile of (a) wild type NS1A(1-73), and (b) NS1A(1-73) R38A mutant.

NS1A(1-73)	[SAM001247031] (μM)	Tm (°C)	ΔTm (°C)
Wild type	0	38.72	-
Wild type	100	41.16	+ 2.44
Wild type	200	44.58	+ 5.86
Wild type	400	49.77	+ 11.05
R38A	0	41.05	-
R38A	400	42.62	+ 1.15

Table 3.6: The shift of Tm of the thermal melting profile of wild type and R38A mutant NS1A(1-73) in the presence of different concentration of SAM001247031.

3.4 Discussion

A total of 17,969 single chemicals from four compound libraries (ChemBridge fragment library, ChemBridge kinase library, NIH clinical collection and MicroSource Discovery spectrum collection) were screened using the optimized FP-based high throughput screening assay. To minimize false positives due to compound-induced fluorescence enhancement or quenching, hits that had greater than a +/- 30% change in total fluorescence compared with controls were removed from consideration. After the primary and secondary screening, 27 compounds were identified showing greater than 40% dsRNA displacement, without disrupting dsRNA. Among them, 6 true hits were identified with dose-response activity and average IC_{50} values less than 10 μM. They are: compounds 4141340 and 5792605 from fragment library, compound 7869182 from kinase library, tannic acid and aurin tricarboxylic acid from the spectrum collection, and compound SAM001247031 from the clinical collection. The overall hit rate is around 0.03%.

The compounds 4141340 and 5792605 are related by a different functional group at the position 8 of quinoline. The full name of 4141340 is 3a,4,5,9b-Tetrahydro-3H-cyclopenta[c] quinoline-4-carboxylic acid. The same compound was also identified as a potent inhibitor of a family of human MAPK-specific protein tyrosine phosphatases using a high throughput screening (Eswaran, von Kries et al. 2006). Three other derivative compounds of 4141340 found through SciFinder Scholar show similar IC_{50} against NS1A and NS1B. That indicates the carboxyl group in this inhibitor class may interact in a similar way with the active-site residues of NS1, probably arginines or lysines. An unsettling aspect of this inhibitor class is that only the aged compounds have inhibitory effects; the freshly dissolved ones can not displace dsRNA. We did notice a color change and additional absorbance peak at 624nm in the aged compounds. However, the color change could be misleading in that only a tiny fraction of colored chemicals could easily change the appearance of compound. As shown in Figure 3.10, a very distinctive ^{1}H NMR spectra of freshly dissolved and aged compound 4141340 strongly suggests that some chemical changes happened when the compound aged. Unfortunately, we were unable to identify the transformed structure. We tried soaking the aged 4141340 into NS1(1-73) crystals, as well as cocrystallization with NS1(1-73). Even though the crystal turned green, we still did not see any sign of 4141340 in the electron density map.

54

(a)
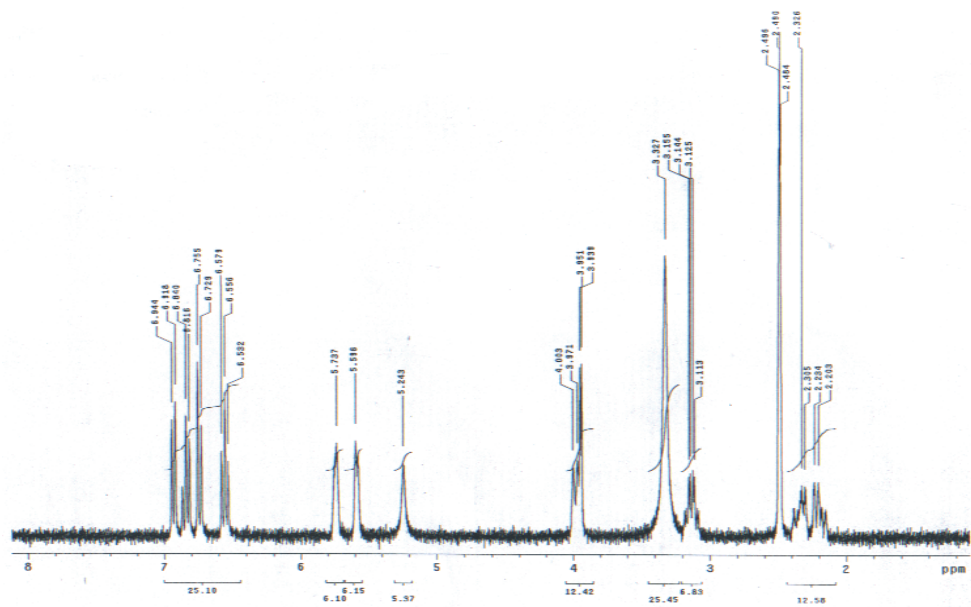
(b)
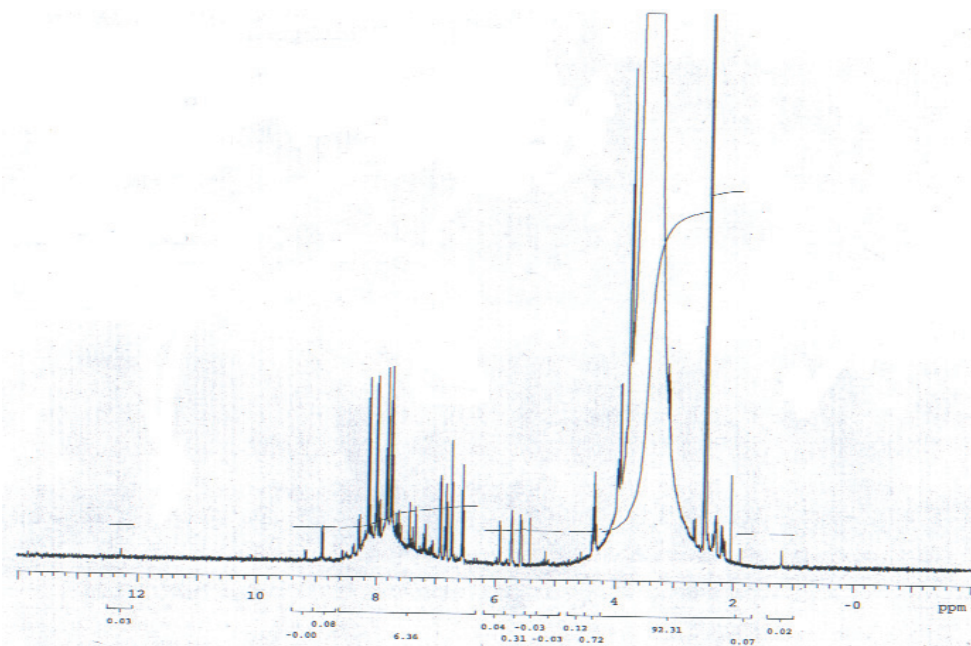
Figure 3.10: The ^{1}H NMR spectra (300MHz) of (a) freshly dissolved and (b) aged compound 4141340.

55

Figure 3.11 shows the docking mode of compound 4141340 in NS1-RBD by the program ICM. It shows good shape complenentarity with the NS1 dimer interface. The docking model predicts a binding mode with hydrogen bonds from between the carboxylic acid group of 4141340 and side chains of Arg19, Arg35 from each monomer. The predicted binding mode of 5792605 can almost superimpose to that of 4141340 with the fluorine at the position 8 of quinoline pointing toward the solvent. From our FP assay results, the substitution at position 8 on the quioline ring to larger moiety (such as benzyl or carboxyl) did not affect the IC_{50} of the compound. The presentation of this position 8 group toward solvent in the docking model may rationalize this observation. Because of the aging effect of compound 4141340 series, we do not know the actual composition of compound that displaced the dsRNA binding. In this regard, a recent paper reported an incorrect structure of a registered compound in the ChemBridge compound library (Inglese, Shamu et al. 2007). The compound "Mirin" was identified from a chemical screening of 10,000 compounds from ChemBridge by Eastman's laboratory. It turned out the structure of "Mirin" provided by ChemBridge is not the real structure of the compound in the screening plate. In other words, the compound in the screening plate is not exactly what the vendor claims, and this makes us wonder about the true identity of compound 4141340.

The compound 7869182 was identified as a hit from the ChemBridge kinase library; it had an IC_{50} around 0.8 μM against NS1A, and 0.6 μM against NS1B. The scientific name of compound is 2,3-di-2-furyl-N-(2-furylmethyl)-6-quinoxalinecarboxamide. The same compound was also identified as an inhibitor of tau fibrillization with an IC_{50} in the low micro molar range (l–3 μM) (Crowe, Ballatore et al. 2007). As shown in table 3.7, five derivates of compound 7869182 were found through a search using SciFinder Scholar. We were surprised to find that only the derivative D4 showed slight inhibition at a nominal concentration of 50 μM, since all five derivatives share the same 2,3-di(furan-2yl)-quinoxalines scaffold. This phenomenon may suggest a possible structure–activity relationship. Figure 3.12 shows the docking mode of compound 7829182 and its derivative D4 in NS1-RBD by ICM. Because of the structure similarity, all derivatives have very similar docking poses to that of 7829182. It is very hard to rationalize the differences in inhibition effect from the predicted docking modes. In addition, the docking scores of 7829182 and its derivatives were very poor (>0), considering a value of -35 is the default threshold value set by ICM for strong binding. An X-ray structure of NS1 with

(a).
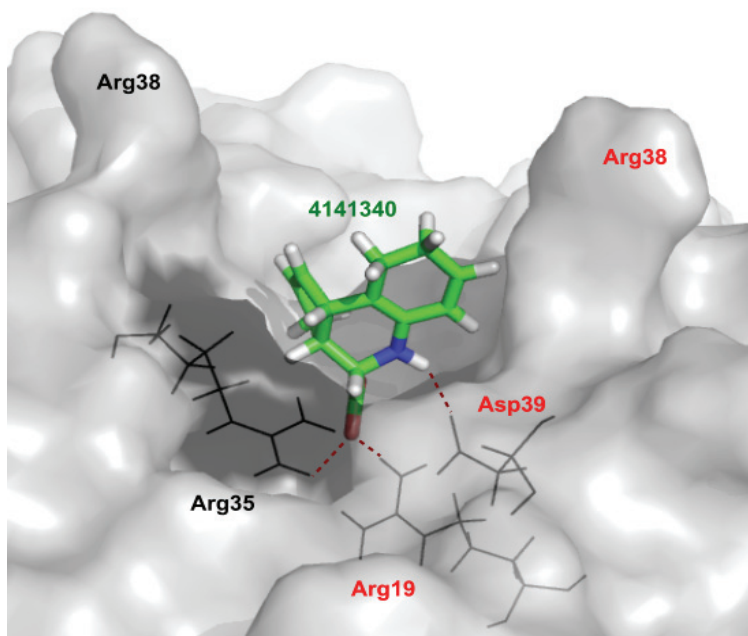
(b).
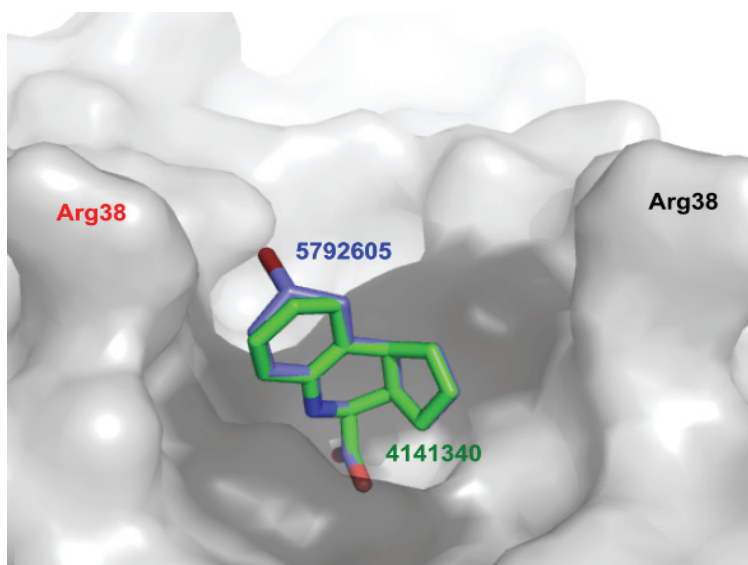
Figure 3.11: The docking model of (a) 4141340; (b) 4141340 and 5792605 in NS1-RBD by ICM. The residues of different monomer were labeled in red and black respectively. The carbon backbones of 4141340 and 5792605 were labeled in green and blue respectively.
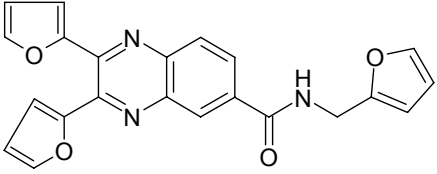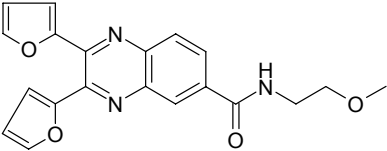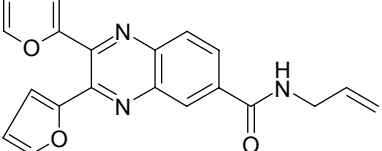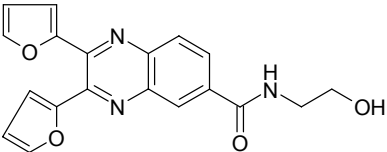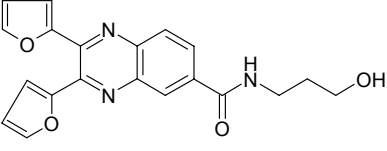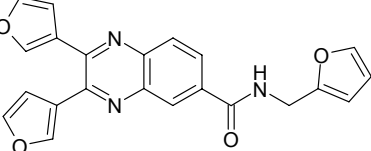
Compound 7829182	Compound Derivatives
2,3-di-2-furyl-N-(2-furylmethyl)-6-quinoxalinecarboxamide 	 D1
	 D2
	 D3
	 D4
	 D5

Table 3.7: The chemical structure of compound 7829182 and its derivatives.
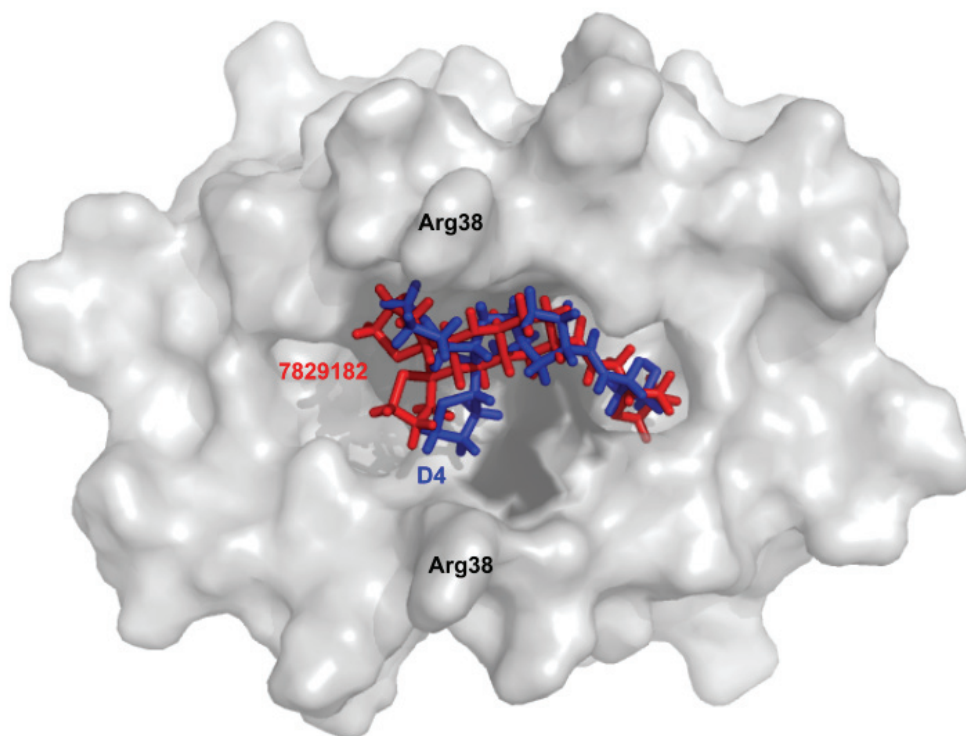
Figure 3.12: The docking model of 7829182 and its derivative D4 in NS1-RBD by ICM. 7829182 was marked in red; and D4 was marked in blue.

7829182 would help to reveal the actual interaction between the two molecules. A cell based plaque reduction assay was carried on by Dr.Krug's laboratory; compound 7829182 did not affect virus replication.

One common name for the compound SAM001247031, identified from the NIH clinical collection, is epigallocatechin gallate (EGCG). It is one of the most abundant polyphenolic antioxidant metabolites in green tea, and has been extensively studied for its antiviral, cancer preventive activities (Brown 1999; Chen, Daniel et al. 2004; Chen and Zhang 2007; Wolfram 2007). EGCG had an IC_{50} around 0.29 μM against NS1A, and 0.18 μM against NS1B, without disrupting dsRNA. More importantly, the thermal melting profile of EGCE with wild type and R38A mutant NS1A(1-73) suggested that EGCG stabilizes the structure of NS1A by interacting with the critical residue Arg38. Exhaustive soaking and co-crystallization of EGCG with NS1A(1-73) have been tried, but were unsuccessful. Figure 3.13 show the predicted docking mode of EGCG in NS1-RBD by the program Surflex. It shows good shape complenentarity with the NS1 dimer interface. The docking model predicts a binding mode with hydrogen bonds from between the hydroxyl groups of 3-galloyl group of EGCG and side chains of Asp12, Ser42 and Arg45; hydroxyl groups of catechin skeleton and side chain of Arg19 and Arg45. Modeling results suggest that EGCG binding to the NS1-RBD is plausible, so we are unsure why complexes have not formed. The reason may be the interaction between the two is very dynamic, or the concentration of EGCG in the crystalline environment is not high enough to populate the site. That is, the micro environment for a successful crystal may not always be consistent with ligand binding.

A cell based plaque reduction assay was carried on by Dr. Krug's laboratory; it showed that EGCG can reduce influenza viral replication by 30%. The same anti-influenza viral activity of EGCG has been observed against three major types of human influenza viruses, including H3N2, H2N2 and H9N2 by Seong's lab in 2007 (Song, Park et al. 2007). Those authors evaluated EGCG and three derivative compounds regarding their ability to inhibit influenza virus replication in cell culture, see figure 3.14. They found that the 3-galloyl group of catechin skeleton plays an important role on the observed antiviral activity, whereas the 5' hydroxyl group at the trihydroxy benzyl moiety at 2-position plays a minor role (Song, Lee et al. 2005; Song, Park et al. 2007). It is very interesting to see that our docking model is consistent with Seong' observations. (1): The 2-trihydroxyl benzyl group did not interact with

(a).
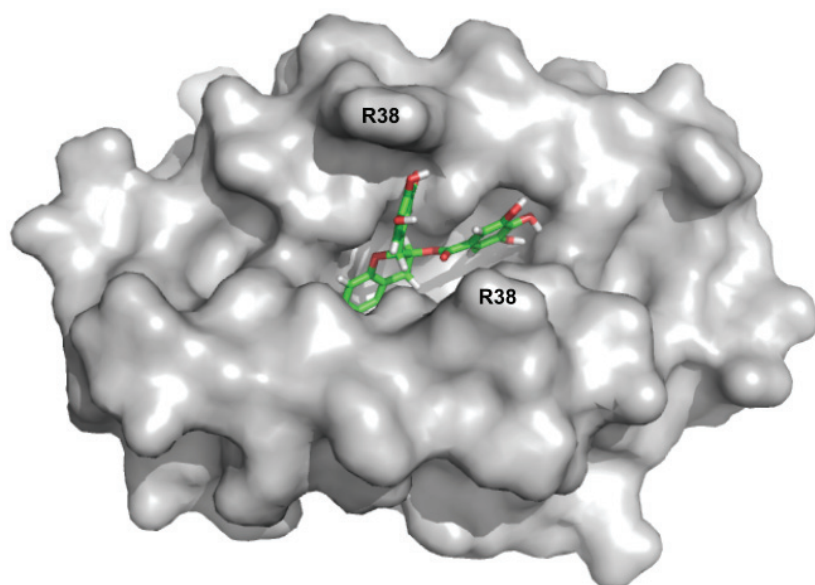
(b).
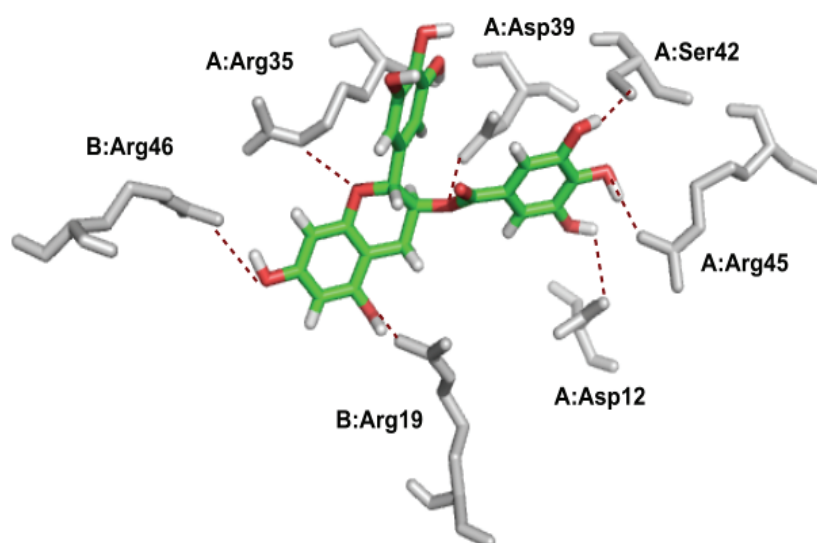
Figure 3.13: The docking model of (a) EGCG in NS1-RBD, (b) the predicted interaction of EGCG and NS1 by Surflex.
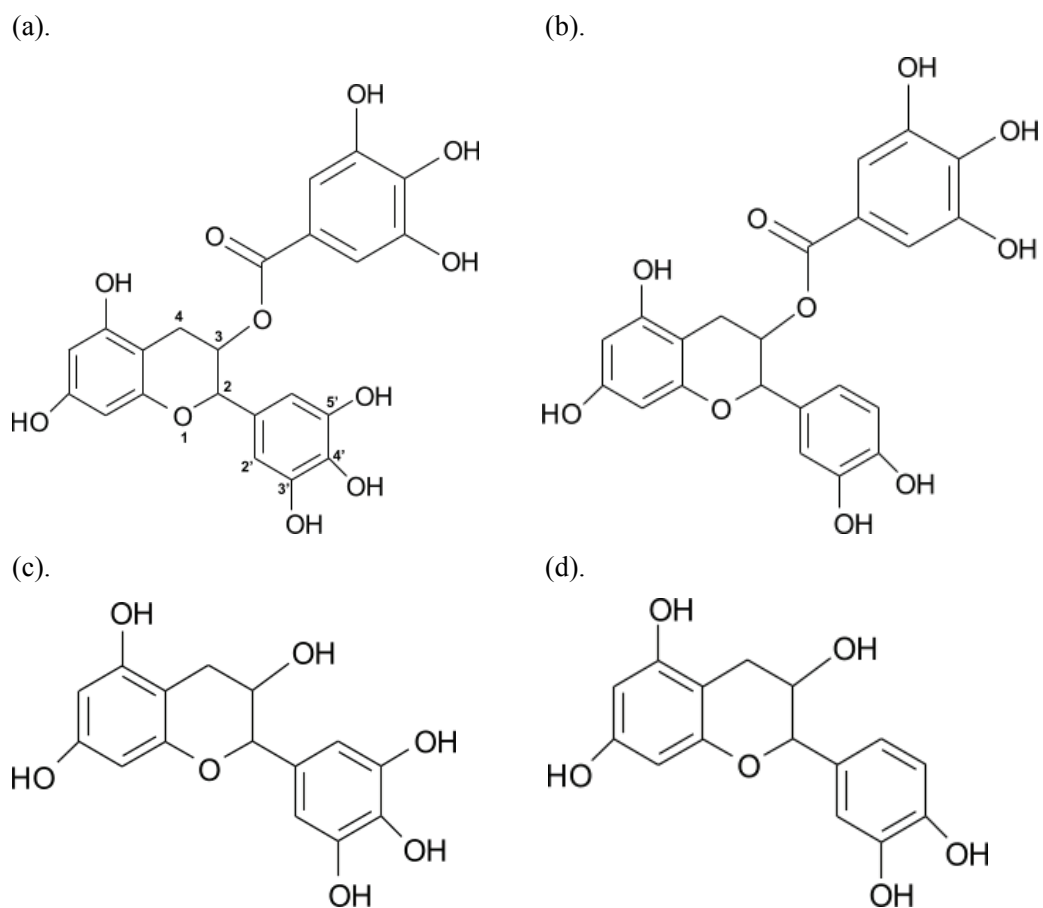
Figure 3.14: The chemical structures of (a) epigallocatechin gallate (EGCG), (b) epicatechin gallate (ECG), (c) epigallocatechin (EGC), and (d) epicatechin (EC).

62

NS1 directly in the predicted binding mode. Besides, the docking pose of ECG was nearly identical to that of EGCG. This may explain why EGCG and ECG have similar inhibition effects on virus replication. (2): The docking model predicts four hydrogen pairs formed between the 3-galloyl group of catechin skeleton and NS1-RBD. The binding affinity would significantly decrease without the entire 3-galloyl group. That could explain why EGC and EC only have slight inhibitory effects in plaque inhibition assay.

We can not rule out the possibility that EGCG may interact with some proteins other than NS1 to cause this anti-influenza viral activity, since EGCG has been reported to interact with many proteins (Ichikawa, Matsui et al. 2004; Wolfram 2007; Kim 2008; Shin, Park et al. 2008). Our results strongly suggest that the anti-influenza viral activity of EGCG is at least partially due to the fact that EGCG interacts with the RBD of NS1 in a way that precludes RNA binding.

Tannic acid and aurin tricarboxylic acid were identified from spectrum collection showing IC_{50} values of 0.48 μM and 0.5 μM respectively. Both tannic acid and aurin tricarboxylic acid have been well studied. Tannic acid is a polyphenol found in the bark of redwood, and functions as a natural defense against infestation, wild fire and decomposition. It is known to exert an anti-cancer activity, and recently reported to be a potent inhibitor of epidermal growth factor receptor tyrosine kinase (Marienfeld, Tadlock et al. 2003; Yang, Wei et al. 2006). Because of the bulky structure (10 benzyl rings) and massive mass (1.7Kd), Tannic acid does not seem to interact with NS1 specifically. It is more likely that tannic acid non-specifically coats the protein surface in a way that precludes the dsRNA binding. Therefore, tannic acid is not a good candidate for further drug development.

Aurin tricarboxylic acid (ATA) is a general endonuclease inhibitor which inhibits protein biosynthesis at the initial stages. It is also known as a platelet adhesion inhibitor and an anti-AIDS compound (Gan, Weaver et al. 1990; Kim, Kim et al. 2008). Interestingly, ATA has been recently discovered to be an inhibitor of influenza virus neuraminidase (Hung, Tseng et al. 2009). The authors found that ATA reduced the influenza viral yield by more than 1000 fold. Because the neuraminidase inhibition assay in Hung's paper was a cell-based assay, itis possible that the displacement of RNA to NS1 by ATA may also contribute to the inhibition effect of ATA on influenza virus replication. Figure 3.15a shows the docking mode of ATA in NS1-RBD by ICM. It shows good shape complenentarity with the NS1 dimer interface. The three-ring structure of ATA somehow resembles that of EGCG; they even share similar

(a).
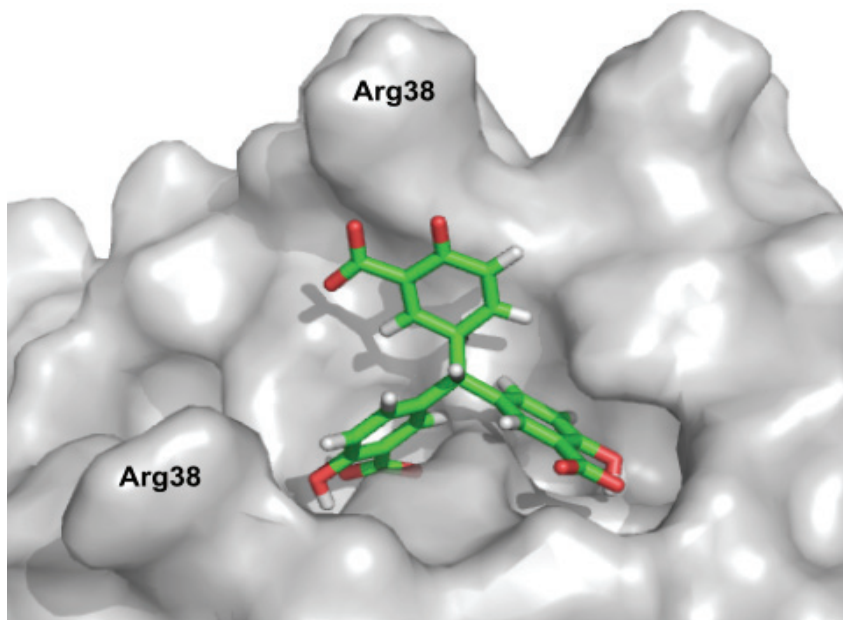
(b).
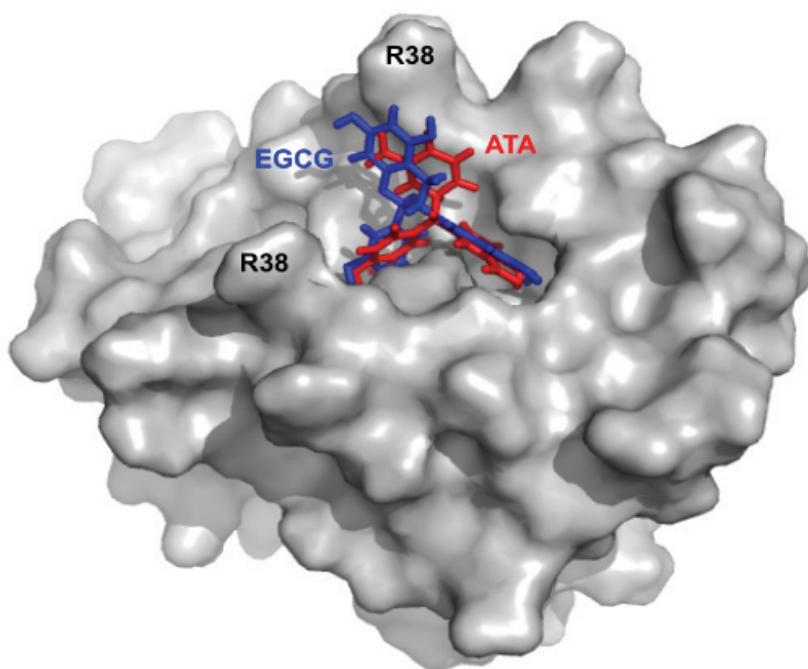
Figure 3.15: The docking model of (a) ATA, (b) ATA (red) and EGCG (blue) in NS1-RBD by ICM.

docking modes predicted by ICM (figure 3.15b). That could explain why the two compounds have close IC_{50}(s) against NS1-RBD. Although ATA has three carboxylic groups on each of benzyl rings, it has very poor solubility in aqueous solution. This could be the reason we failed to get a complex structure either by soaking or co-crystallization.

In summary, we have identified six true ligand hits for the RBD of NS1, with reasonable dose-response activity, and without disrupting dsRNA itself. Four of these show an average IC_{50} less than 1 μM. In addition, one compound, EGCG, has been shown to reduce influenza virus replication in a biological assay, presumably by interacting with the RNA binding domain of NS1.

Chapter 4: Virtual screening of NS1 inhibitors

4.1 Introduction

Virtual screening, also called docking, is a computational technique widely used in the field of structure-based drug discovery. It uses high performance computing programs to analyze large databases of chemical compounds in order to identify possible drug candidates. Basically, virtual screening places a ligand into the binding site of a receptor in a manner that is appropriate for optimal interactions with the receptor. The compounds that are virtually screened, or docked, may be commercially available or may be simply model compounds. Virtual screening in fact does not replace experimental research, and to be useful, the selected compounds from virtual screening need to be validated by real experiments. The goal is to identify some active compounds with non-promiscuous-binding behavior, and then refine them into a series of structures with relevant biological and drug-like activity, known as leads which will be served as drug design platform.

Research in Dr.Krug's laboratory showed that the inhibition of dsRNA binding to NS1A inhibits viral replication (Min and Krug 2006). Our goal was to find a molecule that can bind to NS1A RBD in such a way that it precludes dsRNA binding. The x-ray structure of NS1 RBD reveals a small, nonpolar pocket within the NS1 RNA binding channel; this pocket sits on the NS1 molecular two-fold axis. The bottom of this small pocket is largely defined by the side chains of symmetrically related Leu 15 residues. Preliminary docking of 30 aromatic compounds against the entire surface of the NS1A73 dimer by the program eHiTs, showed that most of the compounds fall into that deep hydrophobic pocket. One of the top scoring compounds, fluorene, lies in the center of the pocket as shown in Figure 4.1. Interestingly, no other cavity on the dimer surface was found to bind those compounds. In addition, docking with 2000 compounds from the Sigma Aldrich compound library revealed that of the top 5% of compounds, the vast majority docked into the hydrophobic pocket. Some top ranked compounds display symmetric, or pseudo symmetric, properties. Their molecular centers aligned near the 2-fold of the NS1A dimer. These results establish that this hydrophobic pocket is a very promising target for computer based virtual drug screening.

66
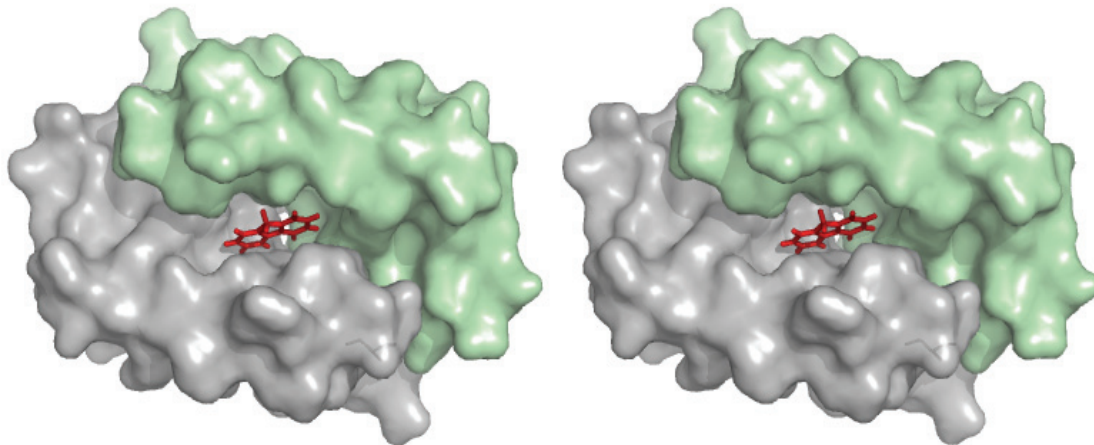
Figure 4.1: Predicted orientation of fluorene to the NS1A dimer. The NS1 dimer is represented as a surface on this stereogram. Each monomer has a distinct color, and the tricyclic fluorene is shown in red.

From previous docking experience, different docking programs show different performance on different proteins (Kellenberger, Rodrigo et al. 2004; McInnes 2007; Hawkins, Warren et al. 2008). Currently, there is a wide range of docking soft ware available for screening virtual compound libraries f. In order to choose the best docking program(s) suitable for virtual screening of NS1A inhibitors, we will compare the known binding affinity of compounds from the physical high throughput FP screening with the ranking given by the virtual screening from four docking programs: eHiTs, GOLD, ICM and Surflex. The program giving the best correlation between computed ranking score and experimental ligand binding affinity was to be used for larger library virtual screenings, such as the ChemBridge Diversity set which contains around 50,000 compounds. Highest-scoring compounds could be validated by FP assay and put into an inhibitor design pipeline.

4.2 Materials and methods

4.2.1 Compound database

Five databases from the five different vendors are available in our data base. These include: ChemBridge (Fragment library, 3963 compounds), National Institutes of Health (Clinical collection, 446 compounds), Maybridge (Diversity library, 53038 compounds), Sigma-Aldrich (Diversity library, 49020 compounds), and MicroSource Discovery Inc (Spectrum collection, 2000 compounds) totaling 106667 potentially available compounds. Most of the libraries were downloaded from corresponding vendor web sites as 2D SDF representations. The 2D SDF format files were converted to 3D SDF format by the program Internal Coordinate Mechanics (ICM). For convenience of docking results analysis, the compound name/ID was copied to first line of each compound SDF file by an in-house Perl script. Open Babel was used for manipulating the various chemical formats of ligands. PyMol from DeLano Scientific was used for visual inspection of results and graphical presentations.

4.2.2 Receptor X-ray structure

The 3D coordinate of the crystal structure of N-terminal domain of NS1 is available as PDB code: 1AIL. The symmetric related molecule was generated by using the molecular graphics program O. The dimer molecule was used as receptor model in virtual screening programs. All water molecules were removed.

4.2.3 Computer workstation

Virtual screening was carried out on TI3D Drug Discovery Cluster. The cluster contains 16 HP Proliant BL35P blade servers, each with 2 dual core AMD Opteron 2.4 GHz processors for a total of 64 processors. Each blade contains 8 GB of memory and a 6 GB ATA hard disk drive. The front-end of the cluster is an HP xw9300 Workstation, equipped with a dual core AMD Opteron 2.4 GHz processor, an NVIDIA Quadro FX4500 graphics card, 4 GB memory, and two 500 GB SATA hard drives. Also attached to the cluster is an HP Proliant DL380 G4 storage server with an Intel Xeon Processor (3.4 GHz) which supports a RAID5 network attached storage system with five 500 GB SATA disk drives.

4.2.4 Docking with eHiTs

eHiTs automatically evaluates all of the possible protonation states for ligands and enzymes, therefore no special preparation of the NS1 RBD dimer was carried out (Zsoldos, Reid et al. 2006; Eitner, Gaweda et al. 2007; Zsoldos, Reid et al. 2007). Initially, docking proceed with 30 ring structures (our in house drug platforms) against the entire surface of NS1A73 dimer without defining a particular active site using the 'receptor' and 'ligand' parameter. Five percent of the best scoring compounds from the first run were then re-docked with a higher accuracy setting (set to 6) and scored with eHiTS_Score, which is included in the eHiTS software package. The eHiTS score mimics the free energy of binding and so the more negative scores reflect better predicted binding.

4.2.5 Docking with GOLD

GOLD uses a genetic algorithm to explore the possible binding modes for protein-ligand docking (Verdonk, Chessari et al. 2005; Olsen, Jost et al. 2006; Thomas, McInnes et al. 2006). The receptor was prepared with all hydrogens added by using Openbabel. The active site was centered around the dyadic pocket near the Leu15 on both monomers. The default setting was used; therefore, 10 dockings were performed for each compound with a total of 10000 genetic algorithm operations. Each docking is followed by Simplex minimization in which ligand orientations are refined to the nearest local optimum. Here 'Goldscore' is used as main scoring function to rank the affinity of the ligand for the receptor. GOLD is optimized for parallel execution on processor networks. The GOLD score is adjusted to a positive value, so higher scores are better; experience with other systems suggests a score of 40 is good.

4.2.6 Docking with ICM

ICM employs a Monte-Carlo Minimization algorithm to find the optimal binding modes of a ligand in the active site represented by a pre-calculated potential grid (Abagyan, Totrov et al. 1994; Borchert, Kishan et al. 1995; Totrov and Abagyan 1996). The coordinates of NS1 RBD dimer were converted to ICM objects with hydrogen atoms added. Then the receptor was optimized using Monte Carlo simulation and energy optimizations. A binding site was defined and a grid map (5 Å) that included the active site amino acids was generated. The compound database was converted into an indexed ICM object for batch docking. Like eHiTS, the ICM score mimics the free energy of binding and so the more negative scores reflect better predicted binding.; the authors suggest -35 as a cut off for a good ligand.

4.2.7 Docking with Surflex

Surflex is a new implementation of the Hammerhead methodology described by Jain (Jain 2003; Jain 2004; Krishnan, Caligaris et al. 2004).The active site of NS1 RBD domain was defined by generating a protomol, a pseudo binding site, using the 'proto' parameter. The definition of the protomol is a tricky step, as parameter 'proto_bloat' defines how far from a potential ligand the site should extend and parameter 'proto_thresh' defines how deep the atomic probe should go into the protein. Surflex-Dock was operated with parameters '-premin -remin'and '-ndock_final 1'. The Surflex-Dock poses were sorted and extracted according to the highest Surflex Raw Score by an in-house Perl script. Like GOLD, the Surflex score is adjusted to a positive value, so higher scores are better; experience with other systems suggests a score of 4 is good.

4.2.8 Data analysis

All the data analysis and file manipulations, including sorting scores, extracting interesting compound poses, extracting common compound coordinates, are done by in-house Perl scripts. Correlation diagrams are generated by Sigmaplot.

4.3 Results

4.3.1 Virtual screening of the fragment library

The ChemBridge fragment library was designed by the distributors based on the commonly accepted Astex Rule of Three (MW $\leq$ 300, H-bond donors $\leq$ 3, H-bond acceptors $\leq$ 3, cLogP $\leq$ 3). As one of the most lead-like fragment libraries, it contains fragments with both free and protected functionality. The initial goal of virtual screening of the fragment library

(3,963 compounds) was to select a docking program that showed the best correlation between computer ranking scores and ligand binding affinity as observed in the HT FP assay.

Figure 4.2 compares the results of docking the fragment library by eHiTs, Surflex, Gold and ICM. The Y axis is the "inhibition percentage", that is the % of probe displacement, for the top 50 compounds derived from HT screening; the X axis is the corresponding docking score for each compound generated by the various docking programs. The calculated correlation coefficients are: -0.09 for eHiTs, 0.01 for GOLD, -0.3 for ICM, and 0.15 for Surflex. These numbers are too low to be considered significant. The inhibition percentage was calculated based on only one fixed compound concentration, 50 μM. Therefore, the Y axis in the correlation diagram is not the ideal binding affinity of compound to protein, but it is still a useful indicator of receptor affinity.

A second form of analysis was carried out by comparing docking scores for the 18 compounds confirmed as hits from secondary FP screening. As shown in table 4.1, the two compounds (4141340 and 5792605) showing 100% inhibition in the HT screening were found in the top 5% of compounds ranked by ICM. For eHiTs to find these hits 66% of the list must be included, 44% for Surflex and 21% for GOLD. In terms of enrichment rate, ICM also did the best for the other five compounds that showed around 50% inhibition in HT screening. Therefore, ICM seems to be the best docking program for NS1 RBD in terms of enrichment rate for top hits.

4.3.2 Virtual screening of Sigma-Aldrich diversity library

A more diverse collection of drug-like molecules, the Sigma-Aldrich diversity library (with 49,020 compounds) was chosen for virtual screening. The Sigma-Aldrich diversity library set is selected based on 3D pharmacophore analysis to cover the broadest part of biologically relevant pharmacophore diversity space while still maintaining drug-likeness, as all compounds satisfy the standard of Lipinski's rule of 5 (MW ≤ 500, H-bond donors ≤ 5, H-bond acceptors ≤ 10, cLogP ≤ 5 and rotatable bonds ≤ 5). Because computational speed is critical while docking with large database, the program eHiTs was ruled out due to its slow docking algorithm and poor performance on the Fragment library test. The 49,020 compounds were docked against NS1 RBD by GOLD, ICM and Surflex on the TI-3D 64-CPU cluster. Comparing the top 5000 compounds from each run (~top 10%), there are 502 overlaps between ICM and GOLD, 482

(a)
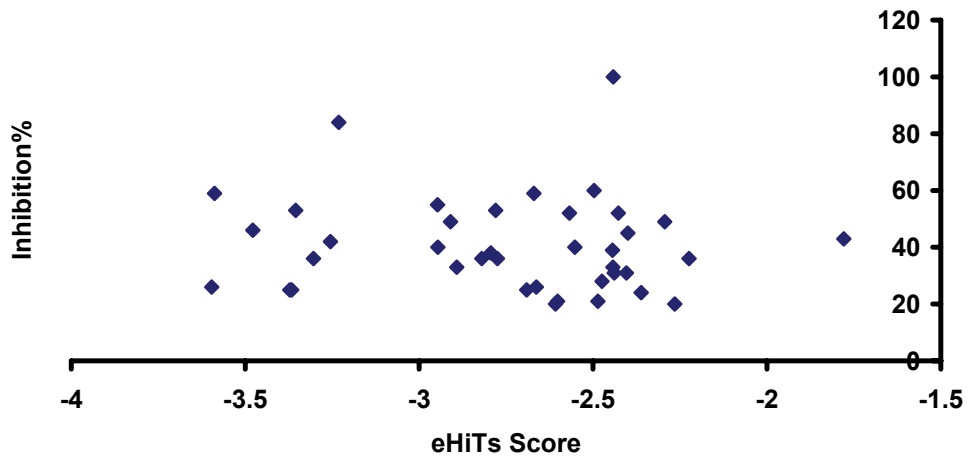
(b)
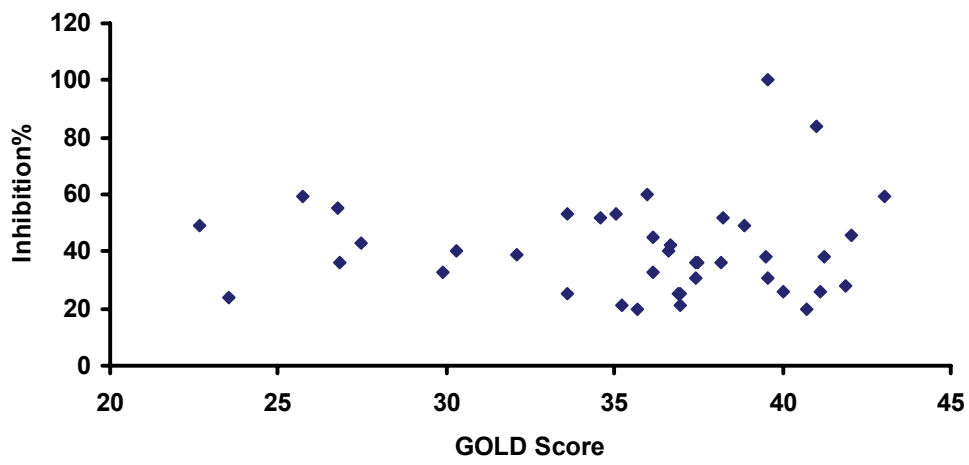
(c)
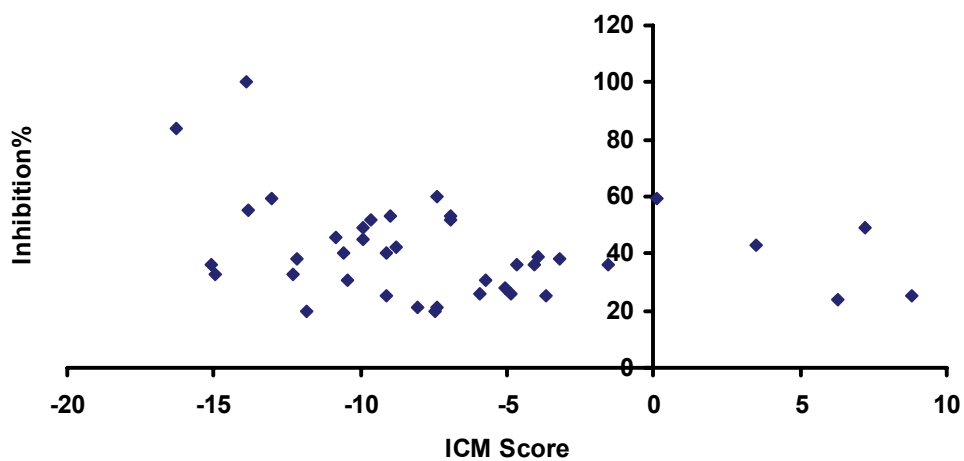
(d)
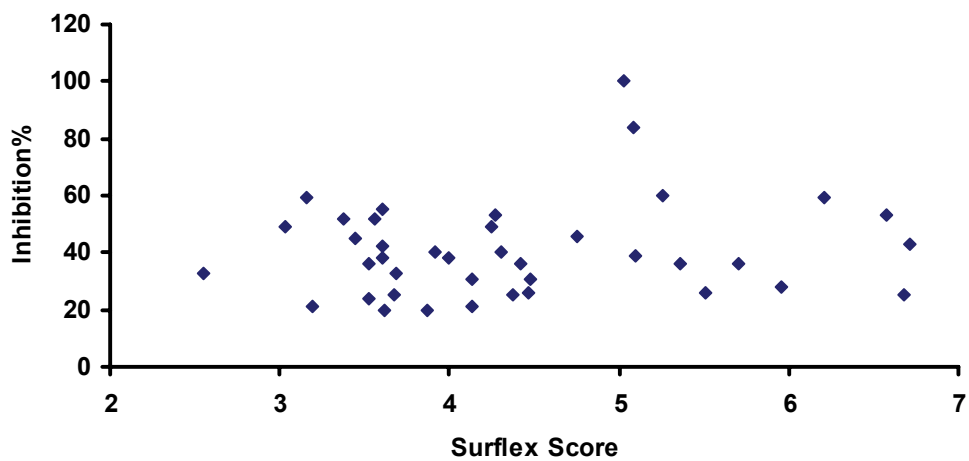
Figure 4.2: Correlation between the inhibition percentages derived from HT screening of ChemBridge fragment library and docking scores generated by virtual screening program (a) eHiTs, (b) GOLD, (c) ICM and (d) Surflex.

(a)

Compound ID	I% 50uM	I% 10uM	eHiTs Score	Ranking	Enrichment Rate
5792605	100	9	-2.44	2423	0.66
4141340	100	95	-3.23	623	0.17
5377792	54	44	-2.29	2782	0.75
5556613	51	9	-2.69	1816	0.49
5265182	49	40	-2.40	2528	0.69
5378142	49	38	-2.95	1144	0.31
7683034	47	47	-3.48	318	0.09
5465122	45	27	-1.78	3479	0.94
6190191	43	22	-2.79	1549	0.42
5212524	42	28	-2.44	2415	0.66
9071210	41	30	-2.55	2137	0.58
9008236	39	33	-3.30	520	0.14
7702036	33	25	-3.60	226	0.06
5738566	33	2	-2.44	2414	0.65
Average ranking for I% at 50uM > 45				1662	0.45

(b)

Compound ID	I% 50uM	I% 10uM	GOLD Score	Ranking	Enrichment Rate
5792605	100	9	39.58	849	0.21
4141340	100	95	41.01	639	0.16
5377792	54	44	38.89	990	0.25
5556613	51	9	36.91	1416	0.36
5265182	49	40	36.16	1593	0.40
5378142	49	38	36.60	1490	0.38
7683034	47	47	42.04	504	0.13
5465122	45	27	27.49	3143	0.79
6190191	43	22	41.25	616	0.16
5212524	42	28	29.89	2929	0.74
9071210	41	30	30.31	2866	0.72
9008236	39	33	26.85	3183	0.80
7702036	33	25	41.13	627	0.16
5738566	33	2	32.11	2575	0.65
Average ranking for I% at 50uM > 45				1069	0.27

(c)

Compound ID	I% 50uM	I% 10uM	ICM Score	Ranking	Enrichment Rate
5792605	100	9	-13.90	203	0.05
4141340	100	95	-16.30	72	0.02
5377792	54	44	-9.89	750	0.19
5556613	51	9	-9.09	922	0.23
5265182	49	40	-9.93	742	0.19
5378142	49	38	-10.57	624	0.16
7683034	47	47	-10.84	576	0.15
5465122	45	27	3.46	3465	0.88
6190191	43	22	-3.19	2470	0.62
5212524	42	28	-12.31	371	0.09
9071210	41	30	-9.08	925	0.23
9008236	39	33	-4.64	2130	0.54
7702036	33	25	-4.87	2074	0.52
5738566	33	2	-3.93	2317	0.59
Average ranking for I% at 50uM > 45				556	0.14

(d)

Compound ID	I% 50uM	I% 10uM	Surflex Score	Ranking	Enrichment Rate
5792605	100	9	4.73	1756	0.44
4141340	100	95	4.82	1643	0.41
5377792	54	44	2.7	3807	0.96
5556613	51	9	3.43	3384	0.85
5265182	49	40	2.96	3703	0.93
5378142	49	38	3.56	3272	0.83
7683034	47	47	4.76	1714	0.43
5465122	45	27	6	476	0.12
6190191	43	22	4.17	2492	0.63
5212524	42	28	3.84	2940	0.74
9071210	41	30	3.86	2909	0.73
9008236	39	33	4.65	1856	0.47
7702036	33	25	5.09	1301	0.33
5738566	33	2	5.12	1264	0.32
Average ranking for I% at 50uM > 45				2754	0.69

Table 4.1: Enrichment profiles of 18 hits from ChemBridge fragment library by docking program (a) eHiTs, (b) GOLD, (c) ICM and (d) Surflex.

(a)
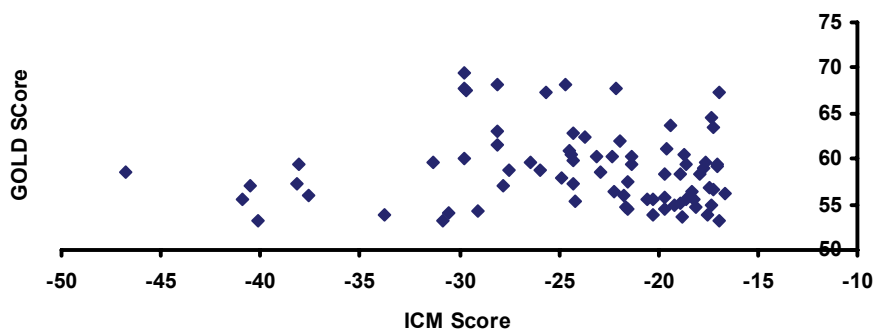
(b)
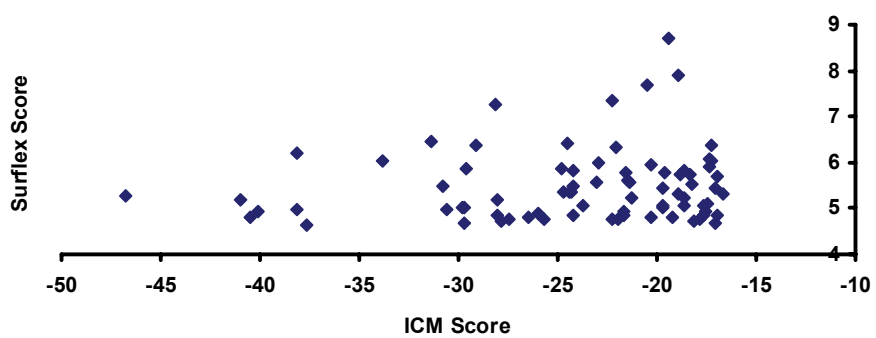
(c)
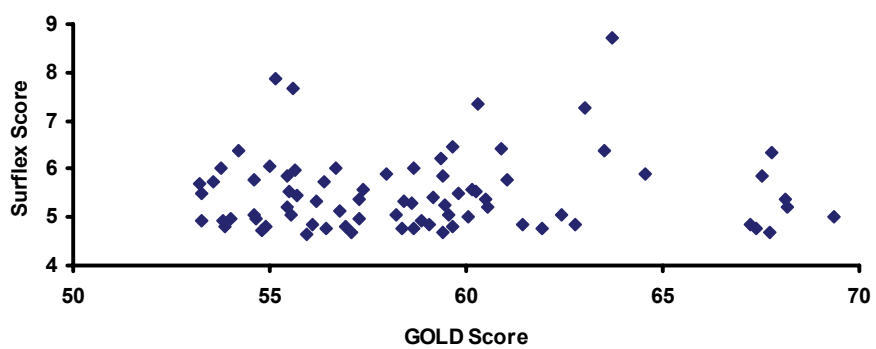
Figure 4.3: Correlation between docking scores of 77 compounds generated by (a) GOLD and ICM, (b) ICM and Surflex, (c) GOLD and Surflex.

overlaps between ICM and Surflex, 976 overlaps between GOLD and Surlfex, and 77 overlaps among all three programs. The coordinates and docking scores of the overlap compounds were extracted from final docking results generated by each program. Among those 77 compounds, the binding mode of each compound generated by three programs looks similar in most cases, but the docking scores do not agree with each other very well. As show in Figure 4.3, the docking score generated by one program was plotted against the docking score of the same compound generated by another program to see if there is any sign of correlation. The calculated correlation coefficients are: -0.027 for ICM and GOLD, 0.041 for Surflex and GOLD, and 0.124 for ICM and Surlfex. Therefore, none of them is considered as statistically significant. In other words, the scoring algorithms of those three programs are uncorrelated. This indicates the importance of choosing the right docking program for the specific targets.

4.3.3 Virtual screening of NIH clinical collection, MicroSource Discovery spectrum collection and MayBridge diversity library by ICM

As shown in the results of virtual screening and high throughput screening of fragment library, ICM did best in terms of enrichment rate for top hits. Therefore, ICM was used in virtual screening of NIH clinical collection (446 compounds), MicroSource Discovery spectrum collection (2000 compounds) and MayBridge diversity library (53,038 compounds). The compounds in NIH clinical collection are drugs that have been in phase I-III clinical trials, and have highly developed properties of drug-likeness. The 2000 compounds in the spectrum collection consist of known drugs (50%), experimental bioactives (20%), and pure natural products (30%). The MayBridge diversity library is a highly diverse pharmacophore-rich collection of hit-like and drug-like compounds. Compounds with ICM score lower than -30 were pooled together.

A total of 224 compounds were extracted; 208 from MayBridge diversity library; 15 from MicroSource Discovery spectrum collection and 1 from NIH clinical collection. A visual inspection of those compounds revealed three symmetric molecules:ZINC0096886, ZINC01045105 and ZINC01324187, see table4.2. They all bound over the NS1A RBD center of symmetry, and made reasonable polar and hydrophobic interactions. Two derivatives were found for each of the three compounds through SciFinder Scholar, as shown in table 4.2. The FP assay shows that only compound ZINC0096886 displaced the dsRNA probe without
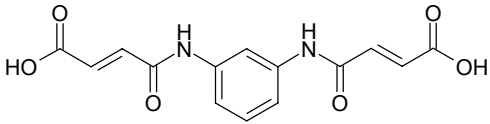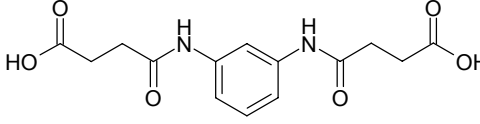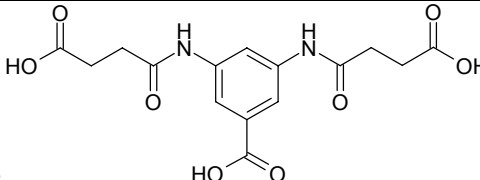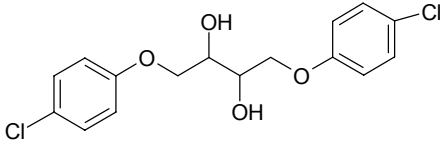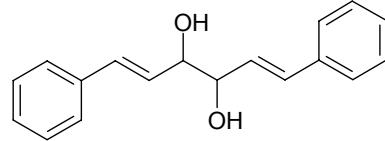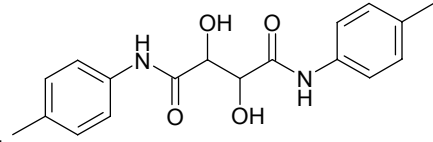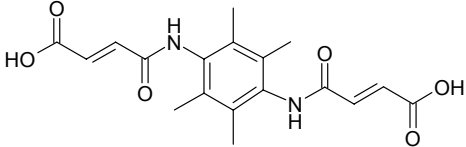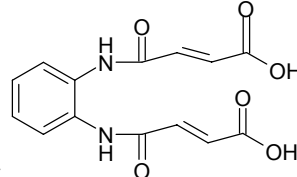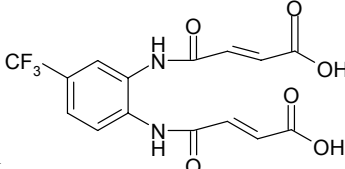
Compound selected from VS	Compound Derivatives
ZINC0096886 4-{3-[(3-carboxyacryloyl)amino]anilino}-4-oxo-2-butenoic acid 	 Z1 Z2
ZINC01324187 1,4-bis-(4-chlorophenoxy)-2,3-butanediol 	 Z3 Z4
ZINC01045105 4-{4-[(3-carboxyacryloyl)amino]-2,3,5,6-tetramethylanilino}-4-oxobut-2-enoic acid 	 Z5 Z6

Table 4.2: The chemical structure of ZINC0096886, ZINC01324187, ZINC01045105, and their corresponding derivatives.
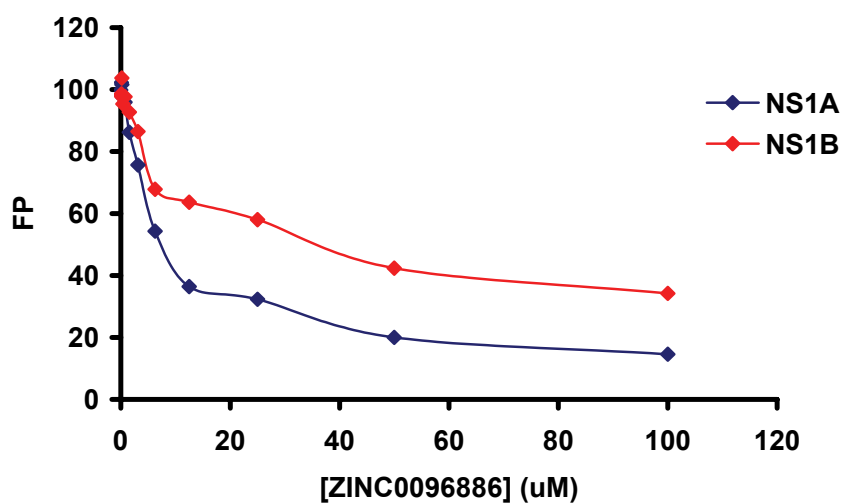
Figure 4.4: FP signal of dsRNA displacement by compound ZINC0096886. The dose-response against GST-NS1A(1-215) is shown in blue, and the dose-response against GST-NS1B(1-145) is shown in red.

disrupting dsRNA. It has an IC_{50} around 19 μM against NS1A, and 13.8 μM against NS1B, see Figure 4.4.

4.4 Discussion

It's a well known fact that inventing and developing a new drug is a long, costly and highly risky process. On average, it takes 1.4 billion dollars and 12 to 24 years for a single new medicine, from starting a project to the launch of a drug product. Since we already know that drugs interact with their target protein in a highly specific manner, virtual screening or docking can, in principle, serve as a computational filter. It may significantly reduce the size of a chemical library needed to be screened experimentally in order to reveal promising inhibitors (Sun 2008). The docking results can also be used to select scaffolds and to help design the actual chemical library to be synthesized. Compared with large scale experimental screening, virtual screening is much more affordable in an academic environment, as most compound library databases are free, and physical assay costs are greatly reduced. Virtual screening can be used with any target, as along as there is a structure or reasonable model. Some virtual screening strategies have already proven to be a successful in a few drug targets, including 5-lipoxygenae and the ATP-sensitive potassium channel (Fechner, Franke et al. 2003; Mikhailov, Campbell et al. 2005).

Here, aiming to find potential compounds that will displace dsRNA binding on NS1, we used four docking programs (eHiTs, GOLD, ICM and Surflex) to dock small compounds to RNA binding domain (RBD) of NS1. eHiTs is an exhaustive docking program that takes a systematic divide and conquer approach to explore the full pose and conformational space (Zsoldos, Reid et al. 2006; Eitner, Gaweda et al. 2007; Zsoldos, Reid et al. 2007). The program splits the ligands into groups of rigid parts and flexible chains, then docks each rigid parts into the receptor independently. A fast graph matching algorithm was applied to score all matching solutions to reconstruct the original molecules.

GOLD (Genetic Optimization for Ligand Docking) used a genetic algorithm that mimics the process of evolution by applying genetic operators to a collection of putative poses for a given ligand (Verdonk, Chessari et al. 2005; Olsen, Jost et al. 2006; Thomas, McInnes et al. 2006). It provides two consensus docking protocols, "Goldscore" and "Chemscore", in terms of docking accuracy and speed.

ICM (Internal Coordinate Modeling) is a combination of the internal coordinate docking methodology with a global optimization scheme (Abagyan, Totrov et al. 1994; Borchert, Kishan et al. 1995; Totrov and Abagyan 1996). With its fast and empirically adjusted scoring functions, the program claims to have an average yield of lead candidates in top 10% of scored compounds.

Surflex uses a systematic approach docking methodology that combines Hammerhead's empirical scoring function with a molecular similarity method to generate docking poses of ligand (Jain 2003; Jain 2004). Those four programs use distinct docking and scoring algorithms, and each one has its own unique features. We did an in-house evaluation of the programs by using eleven known binding affinity complexes of trypsin and its inhibitors; the docking modes and true K_d values are known for all the inhibitors. It turned out all four programs can reproduce the X-ray binding modes of ligands correctly. In addition, both GOLD and Surflex give decent correlation between the docking scores and experimental binding affinities. It is possible that the scoring functions of GOLD and Surflex were already trained by those trypsin-inhibitor complexes. To choose a program works best for NS1 target, we conducted a pilot HT screening of 4,000 compounds from ChemBridge fragment library. The screening results were used to select a docking program showing the best correlation between ranking score and ligand binding affinity.

The actual physical screening results for the fragment library did not correlate well with the docking results by any of the four programs. The successful experience with trypsin and its inhibitors might be exceptional. Perisoff and Head's research group had evaluated 10 docking programs and 37 scoring functions for their rank-ordering by affinity for lead optimization (Warren, Andrews et al. 2006). None of the docking programs or scoring functions made a useful prediction of ligand binding affinity (Warren, Andrews et al. 2006). In fact, it is still very difficult for docking programs to accurately predict the binding free energy of protein ligand complex. One major reason is that the binding free energy includes a loss in configurational entropy upon ligand binding (Chang, Chen et al. 2007; Gilson and Zhou 2007). The decrease in freedom of ligand and protein will cancel much of the energy gained that drives binding. Therefore, neglecting the loss of configurational entropy will overestimate the binding affinities. Gilson reported that the energy terms in a physics based scoring function have to be

scaled down around 10 fold to bring the results into range with experimental binding affinities (Chen, Chang et al. 2004).

As an aid to the practical identification of inhibitors, the enrichment rate calculation is a more appropriate measure for the comparison of programs. ICM showed the best performance in terms of enrichment rate for top hits in a HTP screen for NS1-NTD. The two compounds (4141340 and 5792605) showing 100% inhibition in the HT screening were found in the top 5% of compounds ranked by ICM. That indicates ICM could be the best docking program for virtual screening inhibitors against NS1(1-73).

We were disappointed to find out that there was no correlation among the docking scores given by GOLD, ICM or Surflex in the virtual screening experiments of Sigma-Aldrich diversity library. That indicates that not all docking programs are equally valuable for a given target. As a result, choosing the right docking program(s) for the specific target could be critical. Accurately predicting how a small molecule binds to a protein is difficult, and no program can guarantee success. Correctly scoring and ranking ligands according to their experimental binding affinities is even more difficult, and that's why some program provides trainable scoring function to improve the accuracy and enrichment results by taking advantage of available experimental data for a particular target of interest. eHiTs was acclaimed to have this attractive feature, but our in-house evaluation didn't reveal that benefits. Because of its slow docking speed, eHiTs was not selected in the high throughput screening of NS1 inhibitors.

ICM has shown great success with protein kinases, and has also been used to identify antagonists for the thyroid hormone receptor and EGFR (Schapira, Raaka et al. 2003; Kovacs, Chacon et al. 2004; Cavasotto, Kovacs et al. 2005; Nicola, Smith et al. 2007). Because of its higher enrichment rate in the virtual screening of fragment library, we used ICM to screen a total of 55,484 compounds from three libraries (NIH clinical collection, MicroSource Discovery spectrum collection and MayBridge diversity library). As mentioned above, three compounds were selected for further testing based on their ICM scores and symmetric properties. ZINC0096886 (4-{3-[(3-carboxyacryloyl)amino]anilino}-4-oxo-2-butenoic acid) showed an IC_{50} around 20uM in FP assay. That compound was also identified as LPA (lysophosphatidic acid) antagonist against LPA3 receptor through virtual screening (Fells, Tsukahara et al. 2008). The virtual docking of ZINC0096886 produced an ICM score of "-37". No fully documented inhibitor of NS1 RBD is available, so we could not use its docking score

82

as a threshold value, as is commonly practiced (Cavasotto, Kovacs et al. 2005; Nicola, Smith et al. 2007). As shown in Figure 4.5, the two carboxyl groups of compound ZINC0096886 are predicted by ICM to interact with Arg38 of one NS1 monomer and Arg19 of the other monomer respectively. Without an X-ray structure it is hard to tell how accurate this docking is. We were surprised to find out that the two derivatives of compound ZINC0096886, especially the 4-{3-[(3-carboxyacryloyl)amino] anilino} -4-oxo-butanoic acid (Z1 in table 4.2), could not displace dsRNA at concentration of 100 μM. The only difference between compound ZINC0096886 and compound Z1 is the bond between C2 and C3. The C2-C3 double bond makes compound ZINC0096886 very rigid, while the C2-C3 single bond gives compound Z1 more rotational freedom at carboxyl end. If the docking conformation of ZINC0096886 is correct, the C2-C3 double might play a critical role to hold the Arg38 and Arg19 in a fixed conformation to preclude the binding of dsRNA.

Most docking programs on the market treat the protein as a rigid object, while the truth is protein flexibility often plays an important role in accommodating ligands. Using a fixed receptor structure may be a great limitation if the protein undergoes an induced fit upon ligand binding. The recent X-ray structure of NS1(1-73) with dsRNA complex (Cheng, Wong et al. 2009) revealed that a dramatic conformational change occurred in the orientation of the side chain of Arg38 before and after dsRNA binding (Cheng, Wong et al. 2009). As shown in Figure 4.6, the Arg38 pair acts as a lid to cover the deep pocket upon dsRNA binding, whereas they project into the solvent in the absence of dsRNA. All four docking programs (eHiTs, GOLD, ICM and Surflex) address the problem of ligand flexibility by a variety of algorithms, but the issue of receptor flexibility remains a great challenge.

ICM is able to implement partial receptor flexibility, but the computation rate is very slow, and only a single ligand at a time can be tested. Furthermore, the user is responsible for choosing the specific static receptor conformations generated by program. It's very likely the actually conformation of receptor upon ligand binding is not included in the selection list. In addition, this partial receptor flexibility was implemented in ICM through rotating of receptor side-chain torsion angles by biased-probability sampling method (Abagyan and Totrov 1994; Abagyan, Totrov et al. 1994). However, even if the global optimization of ligand side-chain interaction is convergent, the deformations of the protein backbone may still be crucial to accurate prediction of protein ligand association. The lack of a flexible binding model could be
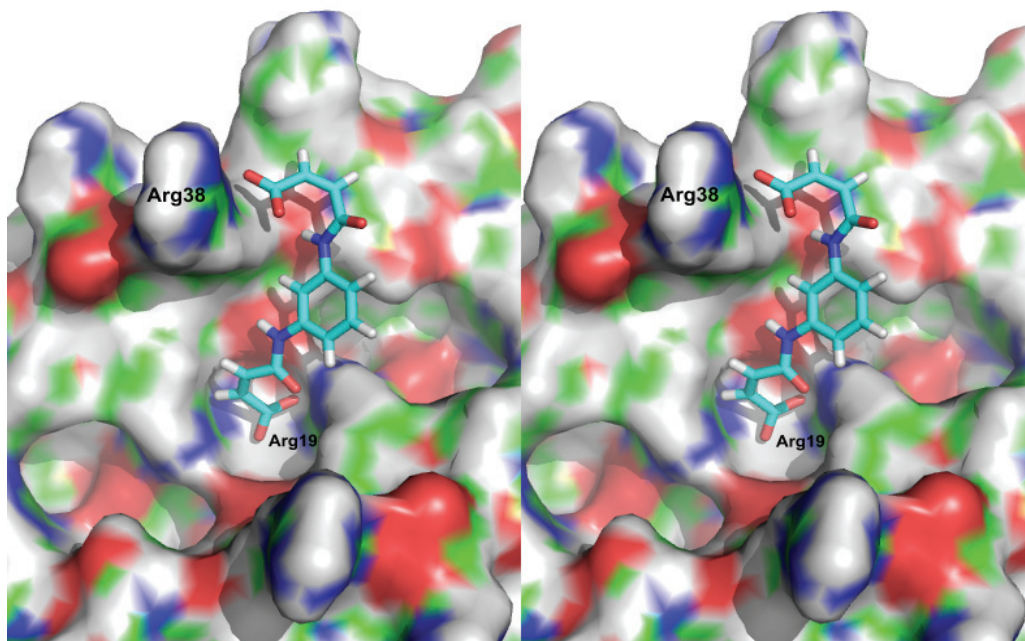
Figure 4.5: Stereo images of the docking pose of compound ZINC0096886 in NS1(1-73). The surface of protein is colored white (carbon), red (oxygen) and blue (nitrogen).
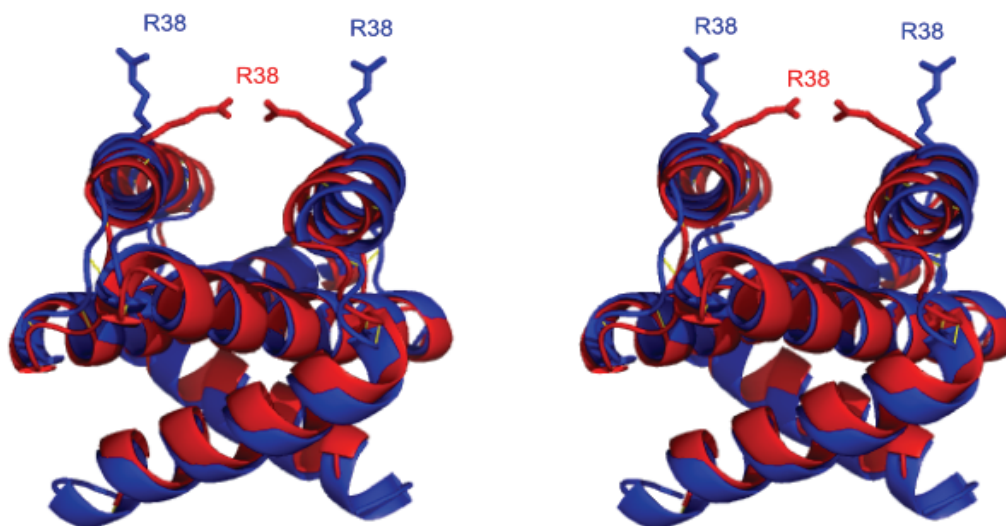
Figure 4.6: Stereo images of superposition of NS1A-RBD dimer in dsRNA bound state (red) and in dsRNA free state (blue).

the reason we get so few hits from virtual screening 55,484 compounds. Docking with full protein flexibility is currently not feasible because of the vast computational increase required.

Epigallocatechin gallate (EGCG), which did well in the HTP assay, was among the 55,484 compounds virtually screened by ICM. We were disappointed to find out that ICM failed to dock EGCG to NS1-RBD even with a threshold value +35 (default is -35). In ICM, the user can adjust the size or position of docking box during receptor preparation. This docking box represents the region in which maps will be generated; it needs to be large enough to encompass the binding pocket but not including regions of receptor which are not relevant for the ligand to bind. By adjusting the size of docking box, ICM was able to dock EGCG to NS1-RBD with a final fitness score -13. This result raised a concern regarding the general utility of docking by ICM. ICM employs a Monte-Carlo Minimization algorithm to find the optimal binding modes of a ligand in the active site (Abagyan, Totrov et al. 1994). The Monte-Carlo Minimization algorithm, also called Monte-Carlo simulation, was firstly introduced as a minimization procedure in molecular dynamics application (Fishman 1995). It was reported to be one of the best algorithms to accurately determine the binding constants for protein ligand interaction (Schapira, Totrov et al. 1999; Huang, Kalyanaraman et al. 2006). In ICM, this search algorithm randomly selects a conformation in internal coordinate space inside the docking box and then makes a step to a new position by a pseudo-Brownian move according to a predefined continuous probability distribution followed by a force-field based energy minimization (Abagyan, Totrov et al. 1994). The minimal energy conformations are returned to the user. This algorithm assumes that the convergence to the best structure should occur from a large class of initial conformations, and only one of which needs to be considered (Abagyan, Totrov et al. 1994; Totrov and Abagyan 1997). This Monte Carlo method is not guaranteed to search exhaustively, so that the best conformation may be missed. That could explain why we did not get the docking pose for EGCG at the first trial. In contrast to EGCG, ICM was able to successfully rank aurin tricarboxylic acid, the proven hit identified from HT FP screening, among the top 2% of 2,000 compounds in MicroSource Discovery Spectrum collection.

The five proven hits (except tannic acid) derived from the HT FP screening were pooled together and docked against NS1-RBD by GOLD, ICM and Surflex respectively. The docking scores were listed in table 4.3. The corresponding correlation coefficients between docking scores and IC_{50}(s) were calculated, see table 4.3. It is surprising to find out that GOLD showed

better correlation (r = -0.84) than that of ICM (r = -0.45) or Surflex (r = -0.53). Although five data points might not truly reveal the correlation between docking scores and IC_{50}(s), it at least suggests that GOLD may be more valuable than we originally thought. The docking pose of EGCG by ICM showed better shape complementary than that for either GOLD or Surflex (figure 4.7). As shown in figure 4.7 panel b, the EGCG was completely buried in the deep hydrophobic pocket between the dimer interface; did not interact with Arg38 from either monomer. But our CD experiment proves that Arg38 is directly involved in the interaction between EGCG and NS1-RBD. It further reminds us to understand the limits and scope of any virtual screening program.

In conclusion, the four docking programs (eHiTs, GOLD, ICM and Surflex) differ in their performance on docking small compound to NS1-RBD. The compound ZINC0096886 was identified as an inhibitor showing IC_{50} around 19 μM against NS1A, and 13.8 μM against NS1B. Currently, we still do not have direct structure evidence to guide us choose the docking program most suitable for HT virtual screening of inhibitors against NS1-RBD. The dynamic motion of critical residue Arg38 on protein receptor surface could be the bottleneck for accurate docking of small compounds to NS1-RBD. Although ICM did a better job in terms of enrichment rate, we should not give up docking with GOLD or Surflex in the future.

	IC$_{50}$ (μM)	GOLD	ICM	Surflex
4141340	8	41.01	-16.3	3.00
5792605	7	39.58	-13.9	2.76
7869182	0.8	52.353	1.69	2.81
ATA	0.5	60.263	-14.73	3.61
EGCG	0.3	48.45	-13.19	6.13
Correlation coefficient (r)		-0.84	-0.45	-0.53

Table 4.3: Docking score profiles of 5 hits from HT FP screening by docking program: GOLD, ICM and Surflex.

(a)
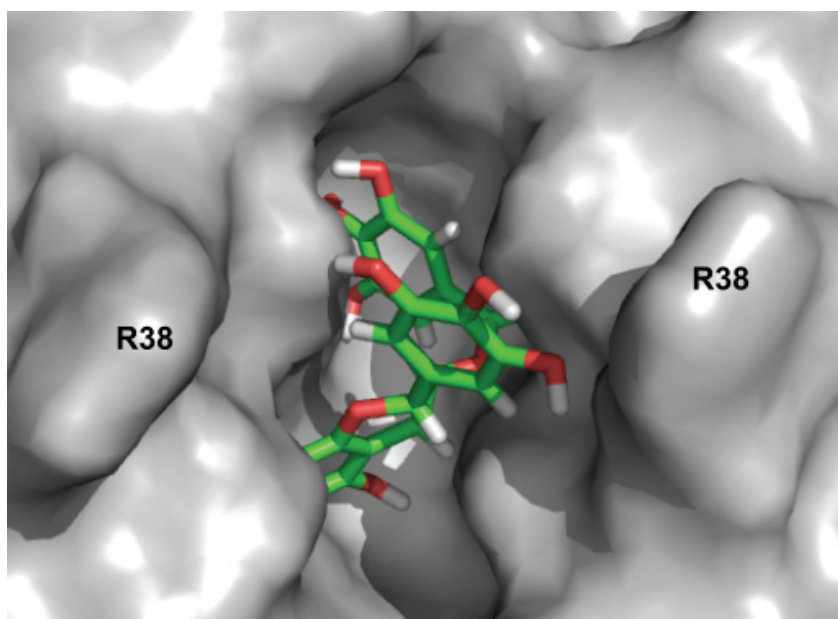
(b)

(C)
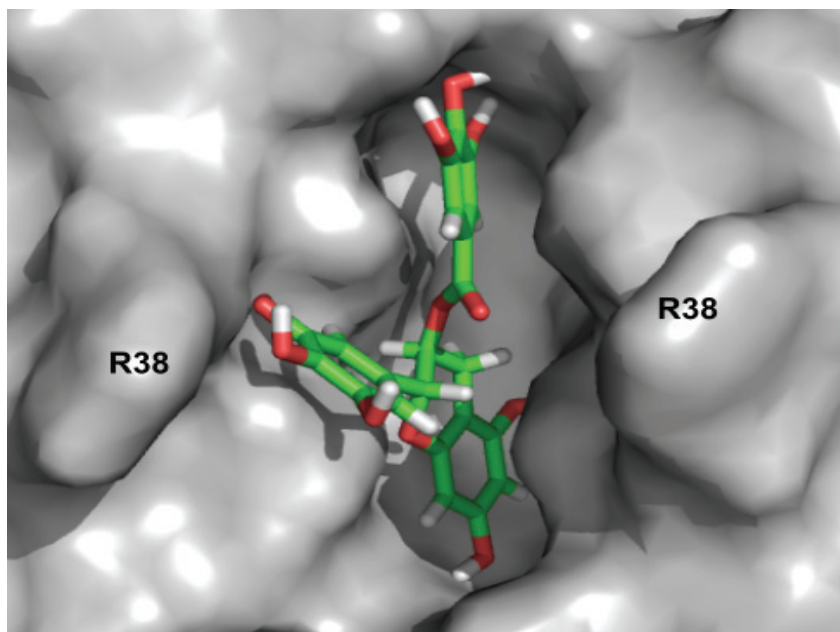
Figure 4.7: The docking pose of compound EGCG in NS1(1-73) by (a) GOLD, (b) ICM and (C) Surflex.

Chapter 5: Crystallization of NS1 related proteins

5.1 Introduction

The most well known benefit of using crystallography in drug discovery is the development of the structure based drug design cycle: structure, synthesis, testing, and back to structure. Compounds identified as inhibitors can be soaked or co-crystallized with the corresponding target protein. An X-ray structure of the complex will reveal the binding mode of inhibitors and serve as basis for further lead refinement.

The X-ray structure of the N-terminal 73-residue RNA binding domain of NS1A was reproduced from the published crystallization conditions. Hit compounds are used in experiments to form useful complexes with it.

The X-ray structure of the NS1A effector domain (residues 79-205) of the mouse-adapted influenza A/PR8/34(PR8) virus strain was solved in 2006 (Bornholdt and Prasad 2006). However that particular NS1A strain does not bind CPSF30. It is likely that this arises from two mutations (F103S and M106I) in the F2F3 recognition site. As the CPSF30 binding site is crucial for virus replication, and is a proven target for new antiviral drug development, it is important to analyze the structure of an NS1effector domain that does interact with CPSF30 (Twu, Noah et al. 2006). Besides the effector domain, crystallization efforts will be made toward crystallizing the nearly full length NS1A protein from Udorn strain.

From previous results, the N-terminal GST tagged NS1A(1-215) fusion protein has a much tighter binding affinity to dsRNA than the protein without a GST tag. Therefore, it would be very interesting, and useful, to see the structure of the fusion protein, since the GST-NS1A(1-215) has been chosen as the HT screening target. Such a structure would be very helpful for virtual screening and rational lead optimization.

5.2 Materials and methods

5.2.1 Construction of various NS1 constructs

The 126 amino acids (residue 79-205) of NS1A was PCR amplified with primers GACGACGACAAGATGACCATGGCCTCCACACC and GAGGAGAAGCCCGGTTTAGC TTCCCCAAGCGAATC. The product was cloned into pET-46 Ek/LIC (Novagen). The first 145 amino acids of CPSF30, was purified by first amplifying the DNA with primers

CGCGGATCCATGCAGGAAATCATCGCCAGC and CCGCTCGAGTTATCTCCGTGTGT GCCGG. And the product was digested with BamHI and Xhol restriction endonucleases, and cloned into pGEX4T3 plasmid DNA (GE Healthcare) cleaved with BamHI and Xhol. The 205 residues NS1A(1-205) was PCR amplified with primers TACTTCCAATCCATGGATTCCAA CACTGTGTCAAGTTTTC and TATCCACCTTTACTGTTAGCTTCCCCAAGCGAATC. And the PCR product was cloned into pNIC28-Bsa4 (Plasmid pNIC28-Bsa4 was kindly provided by Dr.Opher Gileadi at University of Oxford).

5.2.2 Construction of plasmid mutations

The mutations were introduced in plasmid pET46 and pNIC28-Bsa4 by site-directed mutagenesis, according to the Stratagene protocol (Stratagene). Around 50 ng of plasmid and 150 ng of each primer were combined with reaction buffer (20 mM Tris–HCl (pH 7.5), 8 mM MgCl2, 7.5 mM DTT, 50 μg/ml of bovine serum albumin (BSA)), 150 μM dNTP mix, 1 units of KOD Hot Start DNA Polymerase (Novagen), and deionized water to final volume of 50 μl. Primer sequences of effector domain mutant (W817A and W187Y) and NS1A(1-205) mutants (R38A-K41A and R38A-K41A-W187A) are shown in table 5.1. Reaction mixture was further treated with 10 units of DpnI (New England Biolabs) at 37 °C for 2hrs. Then 1 μl of treated reaction mixture was transformed into *E. coli* DH5α competent cell. The presence of the expected mutations was confirmed by DNA sequencing.

Eff(79-205) mutant	Primer sequence
W187A-5'	CATCGGAGGACTTGAAGCGAATGATAACACAGTTC
W187A-3'	GAACTGTGTTATCATTCGCTTCAAGTCCTCCGATG
W187Y-5'	CATCGGAGGACTTGAATATAATGATAACACAGTTC
W187Y-3'	GAACTGTGTTATCATTATATTCAAGTCCTCCGATG
NS(1-205) mutants	**Primer sequence**
R38A-K41A-5'	CTTGATCGGCTTCGCGCGGATCAGGCGTCCCTAAGGGGAAG
R38A-K41A-3'	CTTCCCCTTAGGGACGCCTGATCCGCGCGAAGCCGATCAAG

Table 5.1 Primer sequences of effector domain and NS1A(1-205) mutants

5.2.3 Protein purification

The GST fusion protein, the wild type and mutants, of the NS1A effector domain and NS1A(1-205) were expressed and purified according to the methods described in chapter 2.

5.2.4 Fluorescence polarization (FP) assay

The FP assay was performed as described in chapter 2. Here the 46 mer RNA oligonucleotides with sequence AAGACUCUUGGGGUUUCUGAUAGGCACUGACUCUCUC UGCCUAUUG (sense) and CAAUAGGCAGAGAGAGUCAGUGCCUAUCAGAAACCCAA GAGUCUU (antisense) was used.

5.2.5 GST-pull down assay

The protein purification procedure is the same for all genes cloned in the pGEX vector. *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) containing the CPSF30(1-145)-pGEX/pGEX plasmid were grown and harvested as described in Chapter 2. The cell pellet was resuspended in 50 ml column buffer (CB: 50 mM HEPES (pH 7.5), 50 mM NaCl) and disrupted in a French pressure cell. Cellular debris was pelleted by centrifugation at 5,000g for 60 min at 4 °C. The supernatant was applied to a column containing glutathione agarose beads (3 ml bed volume, equilibrated with CB), and washed with CB until the OD_{280} of the flow through was equal to the background reading. Then 5 ml of purified effector domain protein (2 mg/ml) was passed over the column twice. The column was washed with CB till the OD_{280} of flow through equal to the background reading, then eluted with 10 ml elution buffer (EB: 10mM reduced glutathione, 50 mM Tris, pH 8.0). As a control, the purified effector domain protein was also applied to a column containing the fresh glutathione agarose beads (3 ml bed volume, equilibrated with CB), followed by washing with CB till the OD_{280} of flow through equal to the background reading, then eluted with 10 ml elution buffer (EB: 10mM reduced glutathione, 50 mM Tris-HCl, pH 8.0). All eluted fractions were analyzed on 12% SDS-PAGE.

5.2.6 Screening for crystallization condition

A total of 698 conditions were manually screened for crystallization using the sitting drop method. Among those conditions, 458 came from commercially available crystallization screening kits (Hampton Research), including: Screens I and II, Index, PEG/Ion I and II, Natrix, Grid screen PEG 6000/AS/NaCl/MPD, and Quick screen phosphate. The other 240 conditions came from 10 in-house grid-screening kits. The wild type effector domain and two mutants of NS1A(1-205) were also sent to the Hauptman-Woodward Institute (Buffalo NY) for a high-throughput screening with 1,536 conditions.

5.2.7 NS1-RBD and NS1-ED crystallization conditions

Crystals of the RBD of Udorn NS1A were grown from a solution of 50 mM NaH_2PO_4, 100 mM NaCl, 1mM NaN_3, 10% (w/v) PEG6000 and 20 mg/ml protein by hanging drop method. This condition was optimized based on the published crystallization condition.

5.2.8 NS1-ED wild type and mutant (W817A, W187Y) crystallization conditions.

One crystal form (I) of the influenza A/Udorn/72 NS1A effector domain (wild type) was grown at room temperature using the batch method by mixing 5 µl 10 mg ml^{-1} NS1A effector domain in 50 mM HEPES pH 8.0, 0.1 M NaCl with 5 µl 20%(w/v) PEG 400, 0.1 M sodium acetate pH 5.5, 0.1 M $MgSO_4$. This condition was identified from screens carried out in the high-throughput crystallization screening laboratory at the Hauptman-Woodward Medical Research Institute (Luft *et al.*, 2003). A second crystal form (II) was grown at 277 K using the sitting-drop method. 5 µl protein solution (10 mg ml^{-1} in 50 mM HEPES pH 8.0, 0.1 M NaCl) was mixed with 5 µl reservoir solution and equilibrated against a reservoir containing 800 µl 12%(w/v) PEG 3350 and 4%(v/v) Tacsimate pH 6.0. This condition was found using the Hampton Research PEG/Ion HT Screen kit.

Both crystals of the W187A and W187Y mutant NS1-ED were grown at 277 K using the sitting-drop method. 5 µl protein solution (10 mg ml^{-1} in 50 m*M* HEPES pH 8.0, 0.1 *M* NaCl) was mixed with 5 µl reservoir solution and equilibrated against a reservoir containing 800 µl 0.3 M Magnesium formate and 0.1 M Bis-Tris pH 5.5. This condition was found using the Hampton Research Index HT Screen kit.

5.2.9 Data collection and processing

Crystal form I of the influenza A/Udorn/72 NS1A effector domain was cryoprotected by dipping it into an artificial mother liquor containing 30% PEG 400. Crystal form II was similarly cryoprotected with an artificial mother liquor containing 20% PEG 3350. Crystals of mutant effector domain were also cryoprotected by dipping into an artificial mother liquor containing 30% PEG4000. Crystals mounted in a cryoloop (Hampton Research) were flash-frozen by dipping them into liquid nitrogen andthen placing the frozen loops in the liquid nitrogen cold stream on the X-ray detector goniostat. Diffraction data were collected at 100 K on a MAR 345 image-plate detector (MAR Research) with X-rays generated by a Rigaku Micromax007 rotating-anode generator (Rigaku, The Woodlands, Texas, USA) operated at 40 mV and 30 mA. Diffraction data were collected using a crystal-to-detector distance of

200 mm. Diffraction data were integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

5.2.10 Structure determination and refinement

The initial molecular replacement model for the primitive cell crystal was obtained using the PR8 protein as a model (Bornholdt and Prasad 2006); the search was conducted using the program MOLREP (Vagin and Teplyakov 1997). Subsequent density modification and refinement were carried using CNS (Brunger, Adams et al. 1998) and Refmac (CCP4 1994). Individual atomic isotropic temperature factors were refined. Molecular visualization and rebuilding were done using COOT (Emsley and Cowtan 2004). Water and other solvent molecules (sulfate) were identified using COOT based on an Fo-Fc difference map. Five percent of the diffraction data were set aside throughout refinement for cross-validation (Brunger 1993). PROCHECK was used to make the Ramachandran plots (Laskowski, MacArthur et al. 1993). Model pictures were made using PYMOL (Delano Scientific, San Carlos, CA).

5.3 Results and discussions

5.3.1 RNA binding domain of NS1A

The N-terminal 73 amino RNA binding domain of influenza A/Udorn/72 NS1A was cloned into the pET28b vector. The protein was purified, and digested with tobacco etch virus (TEV) protease to remove the 6xHis tag at the C-terminus of the protein. The protein was crystallized using published crystallization condition with an additional 10% PEG6000. The crystal was found to belong to space group $P4_32_12$, at pH 6.0, with cell constants, a = 41.07, b = 41.07, c =77.88Å. There is one molecule per asymmetric unit, giving a V_m value of 2.21 $Å^3$/Da. X-ray data and refinement statistics for the structure are shown in Table 5.2. A section of the final 2Fo-Fc electron density map used for model construction from the pH 6.0 crystal is shown in Figure 5.1. Following refinement, a Ramachandran plot of the primitive cell crystal showed 96.8% of residues to be in the most favorable region and 3.2% in additional allowed space. The refined structure includes 23 solvent molecules.

The structure of the NS1A(1-73) monomer contains three continuous α-helices. The first helix is almost anti-parallel to the second helix, and the third helix crosses the first two helices with a crossing angle of 90° and 60° respectively. The six additional C-terminal amino acids, which are part of the TEV protease recognition site, are left after TEV protease digestion. There

Data Collection	
Wavelength (Å)	1.5418
Space group	$P4_32_12$
Cell dimensions	
a, b, c (Å)	41.07, 41.07, 77.88
Resolution (Å) (last shell)	30 – 2.4 (2.36-2.44)
R_{merge} (%) (last shell)	4.5 (15.6)
$<I/\sigma I>$ (last shell)	58.1 (5.8)
Completeness (last shell)	97.9 (83.7)
Redundancy	12.0 (5.0)
Refinement	
No.reflections	2732
$R_{working}$ (last shell)	0.208 (0.229)
R_{free} (last shell)	0.298 (0.306)
Average B factor for protein atom ($Å^2$)	36.8
R.m.s deviation from ideality	
Bond lengths (Å)	0.013
Bond angles (°)	1.331

Table 5.2: Data collection and model refinement statistics of NS1A(1-73)
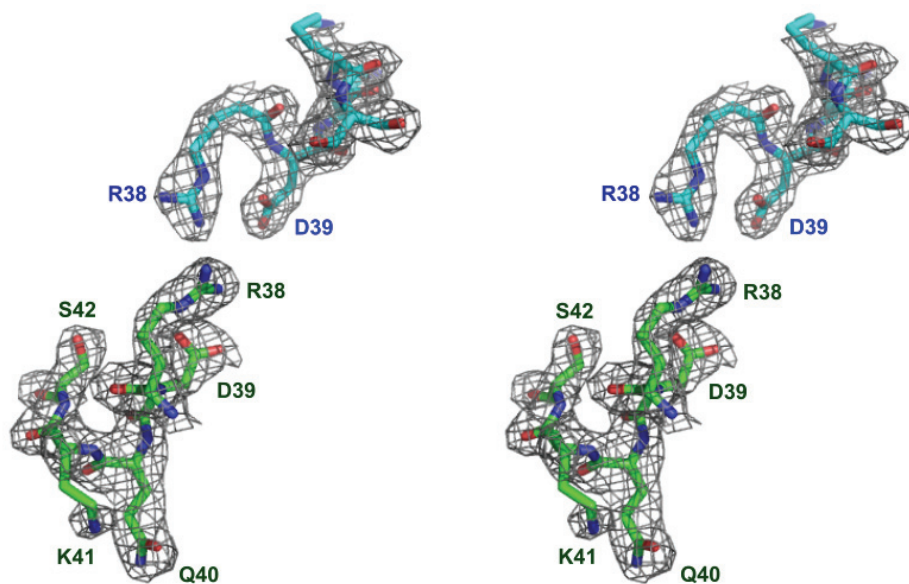
Figure 5.1: Electron density for the NS1A(1-73). This is a section of a $2F_0$-F_c map contoured at 1σ to show the interface of the RBD dimer.

is no electron density for those 6 amino acids, probably because they are flexible and do not pack uniformly in the crystal. The overall structure is almost identical to the published X-ray structure of NS1A-RBD (1ail.pdb). A superposition of NS1A(1-73) with the published NS1A-RBD gives an rms distance between 70 equivalent C_α atoms of 0.18 Å. The most striking difference is that the side chain of Arg38 in our structure has shifted toward the dimer cleft by almost 90° compared to the published structure of NS1A-RBD (1ail.pdb), and 45° compared to the published structure of NS1A-RBD dsRNA bound state (2zko.pdb), see Figure 5.2. This shift could be due to the presence of the additional 10% PEG6000 in crystallization buffer, which makes the solution more viscous than the original crystallization condition.

The Arg38 pair in our NS1(1-73) X-ray structure covers the nonpolar pocket beneath, see figure 5.3. This could explain why soaking of identified inhibitors from our HT screening did not form complexes, as the compounds may not have access to the binding pocket beneath.
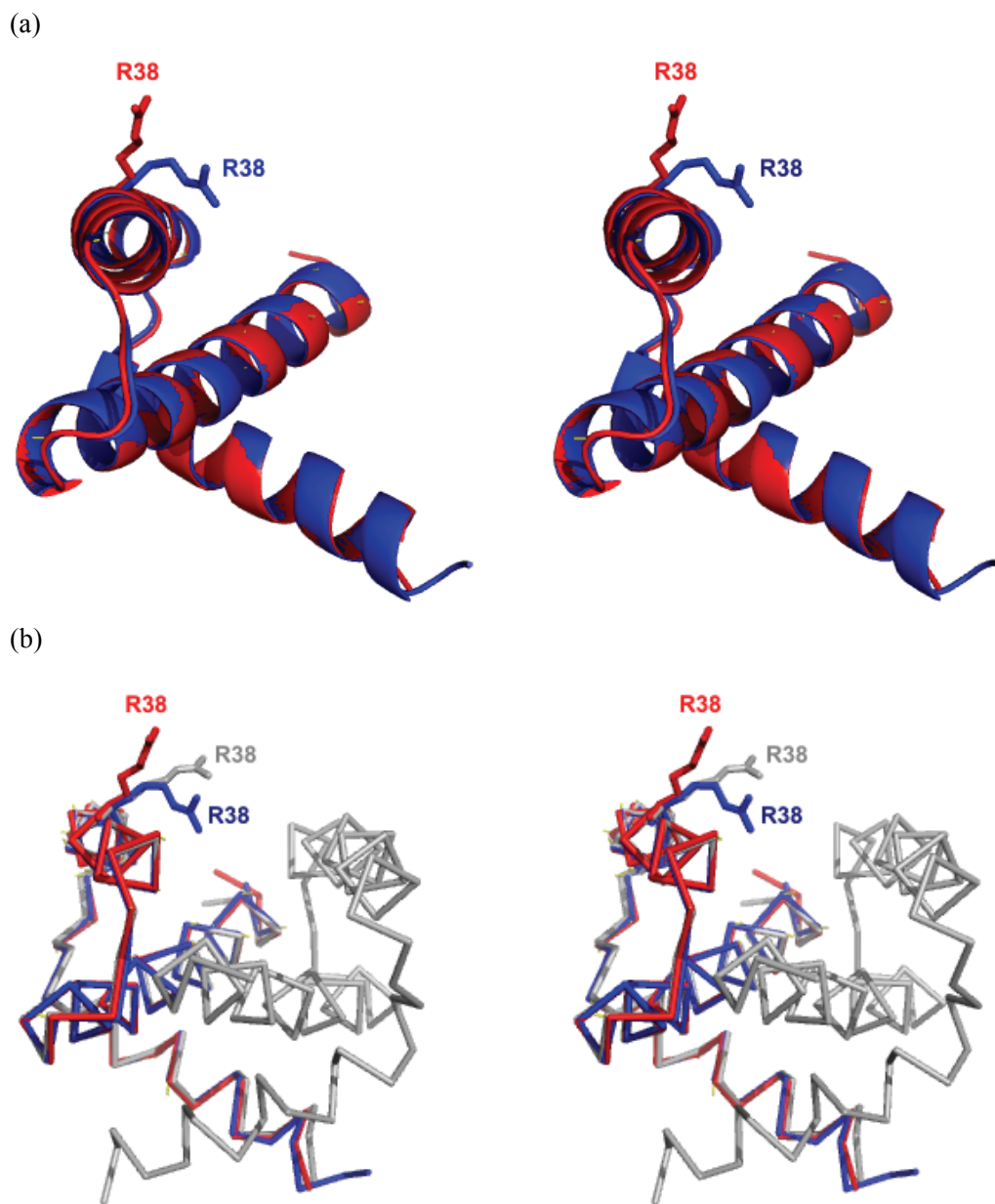
Figure 5.2: Stereo images of superposition of (a) NS1A(1-73) with published NS1A-RBD structure (1ail.pdb), (b) NS1A(1-73) with published NS1A-RBD (1ail.pdb) and NS1A-RBD dsRNA bound state (2zko.pdb). Our NS1A(1-73) is shown in blue, 1ail.pdb is shown in red, and 2zko.pdb is shown in gray.
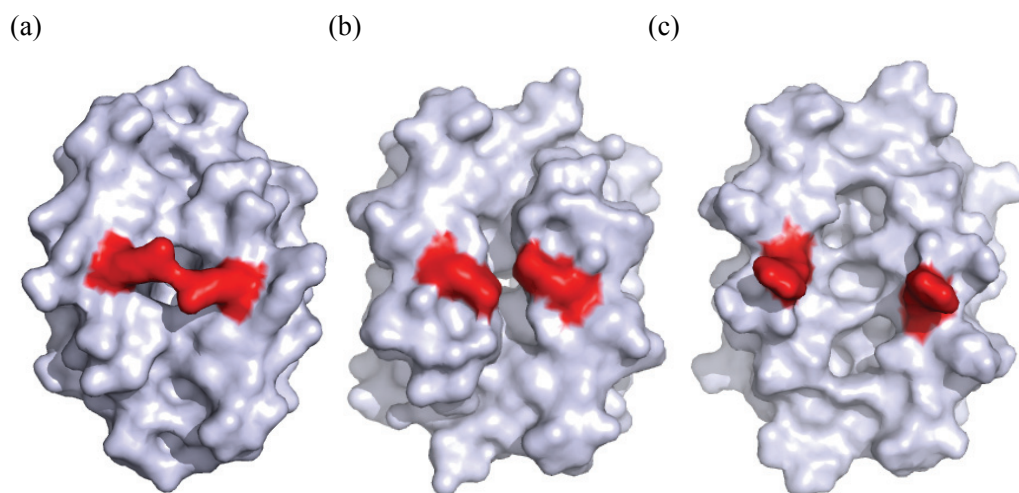
Figure 5.3: Surface presentation of NS1A dimer from (a) dsRNA bound state (2zko.pdb), (b) our structure, and (c) dsRNA unbound state (1ail.pdb). Arg38 is highlighted in red.

5.3.2 NS1A effector domain (wild type)

Recombinant, His tagged, effector domain (residue 79 to 205) can be expressed in good yield from *E. coli* (~10 mg/ liter of cell culture). A screen of crystallization conditions showed two useful forms. Both types are orthorhombic but one cell is primitive and the other centered. The best crystals of NS1A effector domain (residues79-205) formed in space group $P2_12_12_1$, at pH 5.5, with cell constants, a = 47.9, b = 61.5, c =132.1Å. There are two molecules per asymmetric unit, giving a Vm value of 3.04 Å3/Da. The second form is space group $C222_1$, at pH 6.0, with a = 62.8, b = 74.0, and c = 121.9 Å. Like the pH5.5 crystal, there are two molecules in the asymmetric unit, giving a Vm value of 2.21 Å3/Da. X-ray data and refinement statistics for the two structures are shown in Table 5.3. A section of the final 2Fo-Fc electron density map used for model construction from the pH 5.5 crystal is shown in Figure 5.4.

Following refinement, a Ramachandran plot of the primitive cell crystal, at pH 5.5, showed 87.1% of residues to be in the most favorable region and 12.9% in additional allowed space. The refined structure includes two sulfate ions and 136 solvent molecules. The centered cell crystal, at pH 6.0, had a Ramachandran plot with 87.3% of residues in the most favorable region and 12.7% in additional allowed space. That refined structure includes 20 solvent molecules.

The structure of the Udorn NS1A effector domain shows an α-helix β-crescent fold generally similar to that of PR8 NS1A effector domain (Bornholdt and Prasad 2006), and of the recently solved effector domain from an avian influenza virus NS1 (Hale et al, 2008). A ribbon drawing of an effector domain monomer is shown in Figure 5.5. There are seven β-stands and three α-helices in each monomer. Six of the β-stands surround a long central α-helix and make an extensive network of hydrophobic interactions with it. It is these interactions around one side of the helix that gives rise to the rough crescent shape.

As show in Figure 5.5, the first β-strand (a) lies on the convex side of crescent and is connected by a short helix A to strand b. Then the five β-strands (b, c, d, e and f) sequentially connect to each other to form an antiparallel twisted β-sheet that surrounds the central helix B. Helix B is the longest α-helix of the structure; it connects strand f to strand g, which lies adjacent and antiparallel to strand d. Strand g is followed by the last helix C lying on the sharp end of crescent. One sulfate ion is observed in each monomer, and forms hydrogen bonds with

Data Collection		
Wavelength (Å)	1.5418	1.5418
Space group	$P2_12_12_1$	$C222_1$
Cell dimensions		
a, b, c (Å)	47.92, 61.46, 132.13	62.82, 74.03, 121.88
Resolution (Å) (last shell)	30 – 2.1 (2.13-2.21)	30 – 2.6 (2.59-2.68)
R_{merge} (%) (last shell)	6.6 (22.9)	6.0 (14.9)
$<I/\sigma I>$ (last shell)	28.5 (2.6)	32.0 (3.7)
Completeness (last shell)	87.7 (29.7)	85.7 (26.3)
Redundancy	6.1	6.8
Refinement		
No.reflections	17766	7261
$R_{working}$ (last shell)	0.197 (0.234)	0.197 (0.284)
R_{free} (last shell)	0.232 (0.287)	0.235 (0.316)
Average B factor for protein atom ($Å^2$)	35.4	38.8
R.m.s deviation from ideality		
Bond lengths (Å)	0.011	0.014
Bond angles (°)	1.236	1.482

Table 5.3: Data collection and model refinement statistics of effector domain of NS1
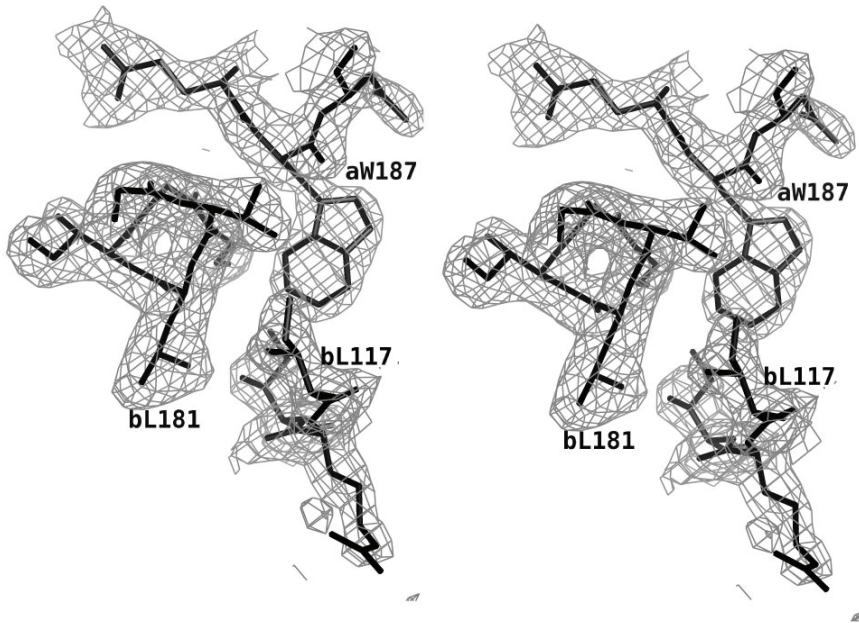
Figure 5.4: Electron density for the NS1 effector domain
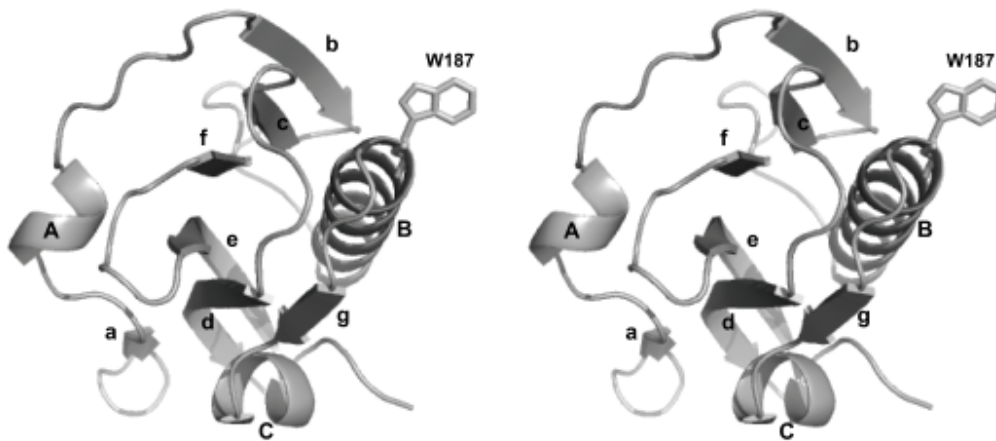
Figure 5.5: Ribbon drawing of the effector domain of Udorn NS1A.

the side chains of Q199 and R193.

The influenza A/Udorn/72 NS1A effector domain and PR8 effector domain share 89% sequence identity and consequently (Figure 5.6), the overall folding of the two monomer structures is quite similar. The superposition of the two structures gives an rms distance between 81 equivalent Cα atoms of 0.29 Å; the superposition is shown in Figure 5.7. The most obvious differences between two monomer structures lie at the N-terminus and in two loops regions; these are indicated by labels. In Figure 5.7, S87 is near the N terminus and marks that region; residues 135-143 (including labeled F138) are shifted up to 5Å toward strands a and d . The second loop region, L163 to G168 (labeled with P167), is shifted toward strand b ~3 Å in the Udorn structure. W187, crucial to effector domain function, is shown for reference.

The NS1 effector domain behaves as a dimer in solution based on its chromatographic properties, and the asymmetric unit of the crystals is also a dimer (Nemeroff, Qian et al. 1995). In both of our crystal structures, the main stabilizing interaction for non-crystallographic dimer formation appears to be the pseudo-symmetrical insertion of the W187 indole ring into the hydrophobic F2F3 binding pocket of its dimer partner; the dimer interface is shown in Figure 5.8.   Figure 5.8a shows a space filling model of a monomer, with the W187 side chain of the other monomer fitting into it; Figure 5.8b shows more details about the interactions of the buried W187 side chain with the receptor pocket. The pocket is formed by main chain residues G183 to N188 together with side chains from residue K108, K110, I117, Q121, and V180. The hydrophobic pocket is located at the base of the central long helix, with the side chain of W187 pointing outward from the periphery. The recently obtained X-ray structure of Udorn effector domain with F2F3 complex (Das, Ma et al. 2008) proves that this hydrophobic pocket is indeed the CPSF binding pocket. This is consistent with the observation that residues G184 to N188 are crucial for binding to the CPSF subunit and are highly conserved among human influenza A virus, including the A/Hongkong/493/1997 virus and A/Vietnam/1203/2004 virus (Twu, Kuo et al. 2007).

The X-ray structure of an avian virus NS1 effector domain also exhibits a non-crystallographic dimer very similar to that seen in the two Udorn crystals (Hale, Barclay et al. 2008). The authors also realized the importance of the interaction of W187 with the CPSF binding site. They mutated three residues near the dimeric interface to alanines, M106A,

```
A/HongKong/483/1997    MDSNTVSSFQVDCFLWHVRKRFADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIRTATRE 60
A/VietNam/1203/2004    MDSNTVSSFQVDCFLWHVRKRFADQELGDAPFLDRLRRDQKSLRGRGNTLGLDIETATRA 60
2GX9-A/PR8/34          MDPNTVSSFQVDCFLWHVRKRVADQELGDAPFLDRLRRDQKSLRGRGSTLGLDIETATRA 60
A/udorn/1972           MDSNTVSSFQVDCFLWHVRKQVVDQELGDAPFLDRLRRDQKSLRGRGSTLGLNIEAATHV 60
                       **.*****************:..*********************.****:*.:**:

A/HongKong/483/1997    GKHIVERILEEESDEALKMTIASVPAPRYLAEMTLEEMSRDWLMLIPKQKVTGSLCIRMD 120
A/VietNam/1203/2004    GKQIVERILEGESDKALKM-----PASRYLTDMTLEEMSRDWFMLMPKQKVAGSLCIKMD 115
2GX9-A/PR8/34          GKQIVERILKEESDEALKMTMASVPASRYLTDMTLEEMSRDWSMLIPKQKVAGPLCIRMD 120
A/udorn/1972           GKQIVEKILKEESDEALKMTMASTPASRYITDMTIEELSRDWFMLMPKQKVEGPLCIRID 120
                       **:***:**: ***:****      **.**:::**:**:**** **:***** *.***::*

A/HongKong/483/1997    QAIMDKDIILKANFSVIFNRLEALILLRAFTDEGAIVGEISPLPSLPGHTEEDVKNAIGV 180
A/VietNam/1203/2004    QAIMDKTIILKANFSVIFDRLETLILLRAFTEEGAIVGEISPLPSLPGHTGEDVKNAIGV 175
2GX9-A/PR8/34          QAIMDKNIILKANFSVIFDRLETLILLRAFTEEGAIVGEISPLPSLPGHTAEDVKNAVGV 180
A/udorn/1972           QAIMDKNIMLKANFSVIFDRLETLILLRAFTEEGAIVGEISPLPSFPGHTIEDVKNAIGV 180
                       ****** *:*********:***:********:***************:**** ******:**

A/HongKong/483/1997    LIGGLEWNNNTVRVSENLQRFTWRSSDENGRSLLPPKQKRKMERTIEPEV------- 230
A/VietNam/1203/2004    LIGGLEWNDNTVRVTETIQRFAWRNSDEDGRLPLPPNQKR---------------- 215
2GX9-A/PR8/34          LIGGLEWNDNTVRVSETLQRFAWRSSNENGRPPLTPKQKREMAGTIRSEV------- 230
A/udorn/1972           LIGGLEWNDNTVRVSKTLQRFAWGSSNENGRPPLTPKQKRKMARTARSKVRRDKMAD 237
                       ********:*****::..:***:* .*:*:** *.*:***
```

Figure 5.6: Alignment of the protein sequence of NS1A from influenza A/Hong Kong/483/1997, A/VieNam/1203/2004, A/PR8/34 and A/Udorn/1997.
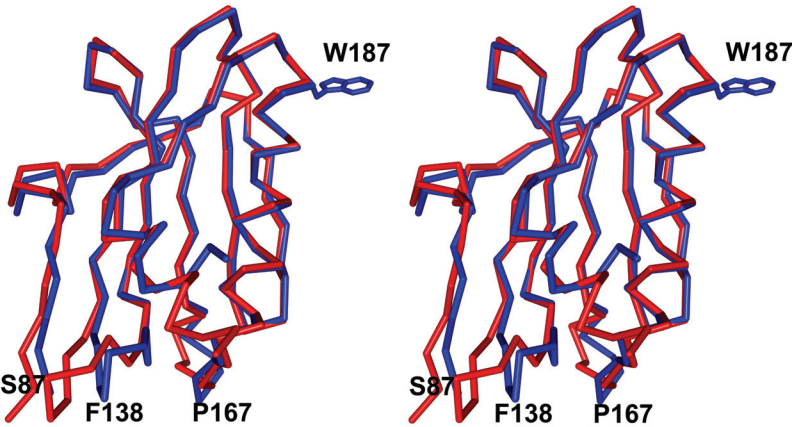
Figure 5.7: Superposition of Udorn effector domain and PR8 effector domain(2GX9). The C_α trace of Udorn effector domain is shown in blue bonds, and that of PR8 effector domain is shown in red bonds. The side chain of W187, indicating the area of the effector domain binding pocket, is also shown.

(a)

(b)
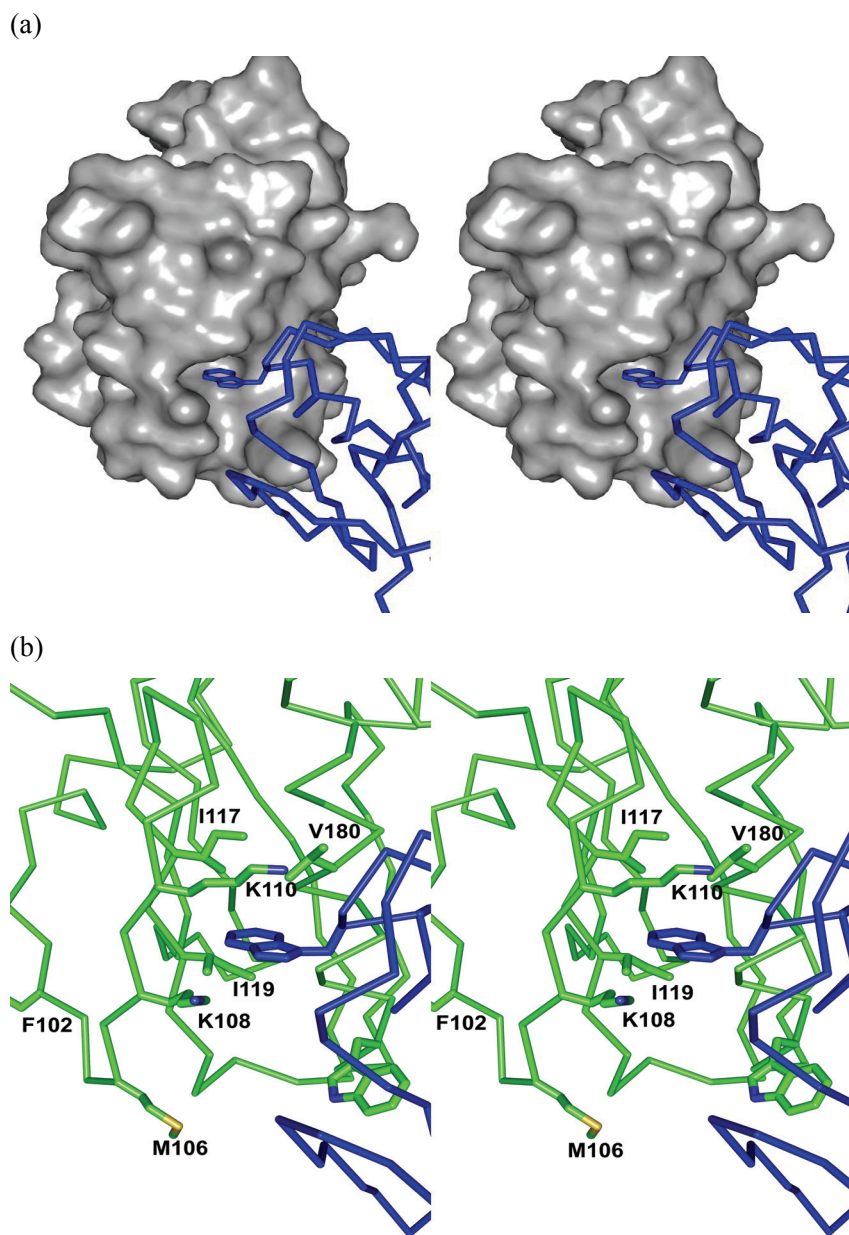
Figure 5.8: The dimer interface for NS1-effector domain. (a). A space filling model of monomer A reveals a distinct, largely hydrophobic pocket that has evolved to bind aromatic residues from the F2F3 domain of CPSF30. Here W187 from effector domain monomer B binds in that pocket. (b). A detailed view of the interactions in the pocket of W187 side chain with binding cleft residues.

Q121A, and W187A, but only the tryptophan alteration caused the protein to act as a monomer in solution.

The PR8 effector domain also forms a non-crystallographic dimer and has the same kind of interface described here (Bornholdt and Prasad 2006). The authors of that study focused their attention not on the W187 contact, but on the formation of a segment of antiparallel β sheet between the N-terminal strands of the neighboring molecules. That β sheet pairing is not seen in either of our structures nor in the avian virus structure, and appears to result from adventitious crystal packing forces unique to the PR8 crystal. The pseudo dimeric interaction of W187 with the F2F3 binding site is common to all four structures and is clearly the main interaction driving effector domain dimerization, as confirmed by the recent mutagenic studies.

Figure 5.9 shows the relative orientations of the non-crystallographic dimers from the two crystal forms of Udorn NS1 effector domain. One monomer from each structure is superimposed in a least squares sense with an rms deviation of 0.28 Å for Cα atoms. In each case its dimeric partner makes interactions like those described above, burying the W187 side chain in the neighboring binding site (the side chains are shown and labeled). However, subtle differences at the interface generate rather large differences for remote regions of the molecule. To superpose the second monomer from the two systems requires a rotation of 44°. The central B helices of the monomers are shown as cartoons in Figure 5.9 to facilitate the comparison. The orientation of the A PR8 effector domain dimer is similar, but not identical, to that of the pH 6 Udorn crystal (red in Figure 5.9). Again, the dimer interface is basically the same in all the crystal structures; however, subtle differences at the interface still allow for larger differences in the crystal packing of the entire dimer.

In discussing the F2F3 binding site of the effector domain, it is interesting to note that the pocket residues K108, K110, I117, Q121, V180, and G183 to W187 are conserved among almost all influenza A viruses, including those having an intrinsic defect in binding CPSF. It appears that proteins, like PR8, that are defective in CPSF binding lack the consensus amino acids Phe 103 and Met 106. The recent structure of the effector domain with F2F3 fragment of CPSF bound reveals that F103 and M106 are not part of the CPSF binding pocket but are involved in intermolecular interactions that stabilize the complex at a site remote from the hydrophobic pocket. F102 and Met 106 are not involved in the dimerization of the Udorn effector domain either, as shown in Figure 5.9 panel b.
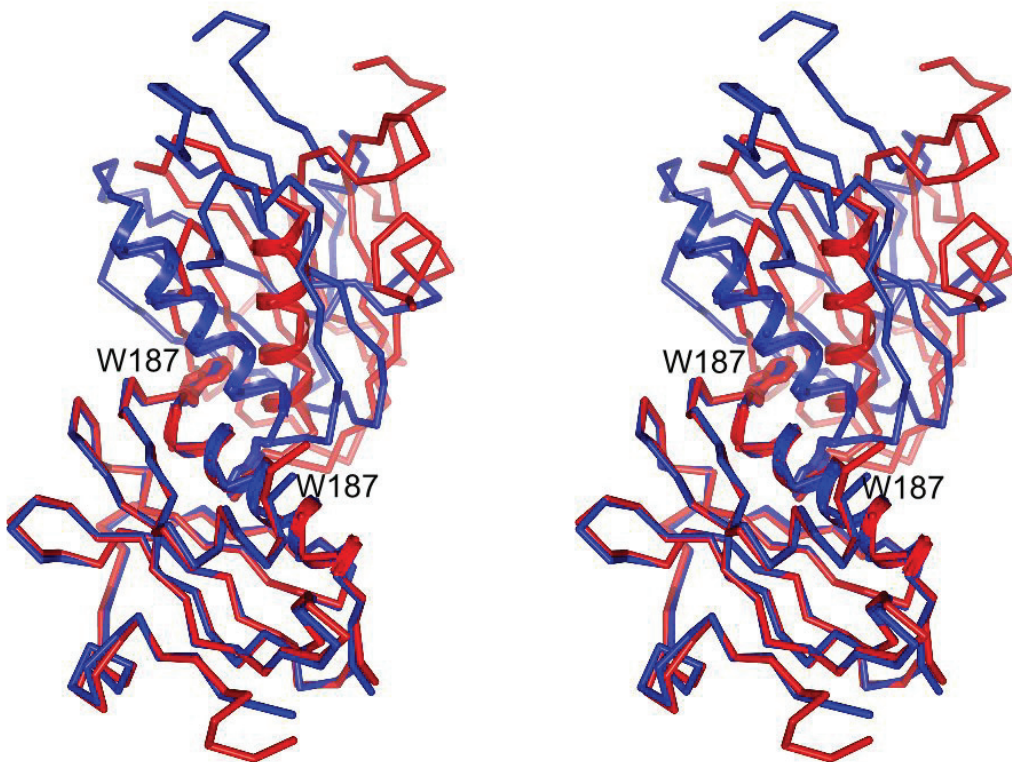
Figure 5.9: Superposition of the Udorn effector domain at pH5.5 (blue ribbon) and the PR8 effector domain (red ribbon)

As described in the introduction section, NS1 is a functional dimer with two major domains (Nemeroff et al., 1995). We recently solved the structure of the NS1A N-terminal domain, residues 1-79, but found that it was the same as that described previously for residues 1-73 (Liu et al., 1997). In our structure we note that residues beyond 73 are not observed, presumably because they are disordered. In the C-terminal effector domain reported here, residues 79-83 are not observed in either crystalline form, again because they are likely to be disordered. This suggests that there is a flexible linker region of about ten amino acids (residues 74-83) between the two domains.

The NS1 dimer is maintained by dimerization of the N-terminal domain (Wang et al., 1999) and it may be that the effector domains of NS1 also participate in NS1 dimerization (Wang et al., 2002). It is likely that in the NS1 dimer the effector domains dimerize in a manner similar to that which we and others have observed crystallographically. To help understand the action of the NS1 effector-domain dimers, in which the biologically important CPSF30-binding sites are buried, we constructed a hypothetical model of the intact NS1 dimer. This model is illustrated in Figure 5.10. To build the model, we aligned the twofold axes of the N-terminal and C-terminal dimers and rotated them so as to bring the C-terminal residues of the N-terminal domain near the N-terminal residues of the C-terminal domain. We found that this was very straightforward and allowed the domains to be linked plausibly by a ten-residue linker. We minimized the energy of the model using CNS (Brünger et al., 1998). In Figure 5.10, the N-terminal domains are shades of blue for chain 1 and cyan for chain 2. The effector domains are shades of red for chain 1 and orange for chain 2. The ten-residue linkers are colored yellow. The Trp187 side chains are shown as van der Waals structures and chain 1 is labeled; the indole ring binds deep in the pocket of chain 2. We have also added a plausible binding site for dsRNA to the NS1 model. RNA is thought to bind in the prominent channel between symmetrical helices 2 of the N-terminal dimer. The RNA binding is stabilized by a number of ionic interactions (Wang et al., 1999). This notion has been given strong support from site-directed mutations which show, among other things, that Arg38 is essential to dsRNA binding. This key side chain is shown as stick bonds pointing to the dsRNA; Arg38 on chain 1 is labeled and also serves to identify helix 2 which forms the binding channel.

In the NS1 dimer, it is likely that the effector domains dimerize such that for each subunit the Trp187 side chain is buried in the hydrophobic pocket of its dimer partner. The binding of

106
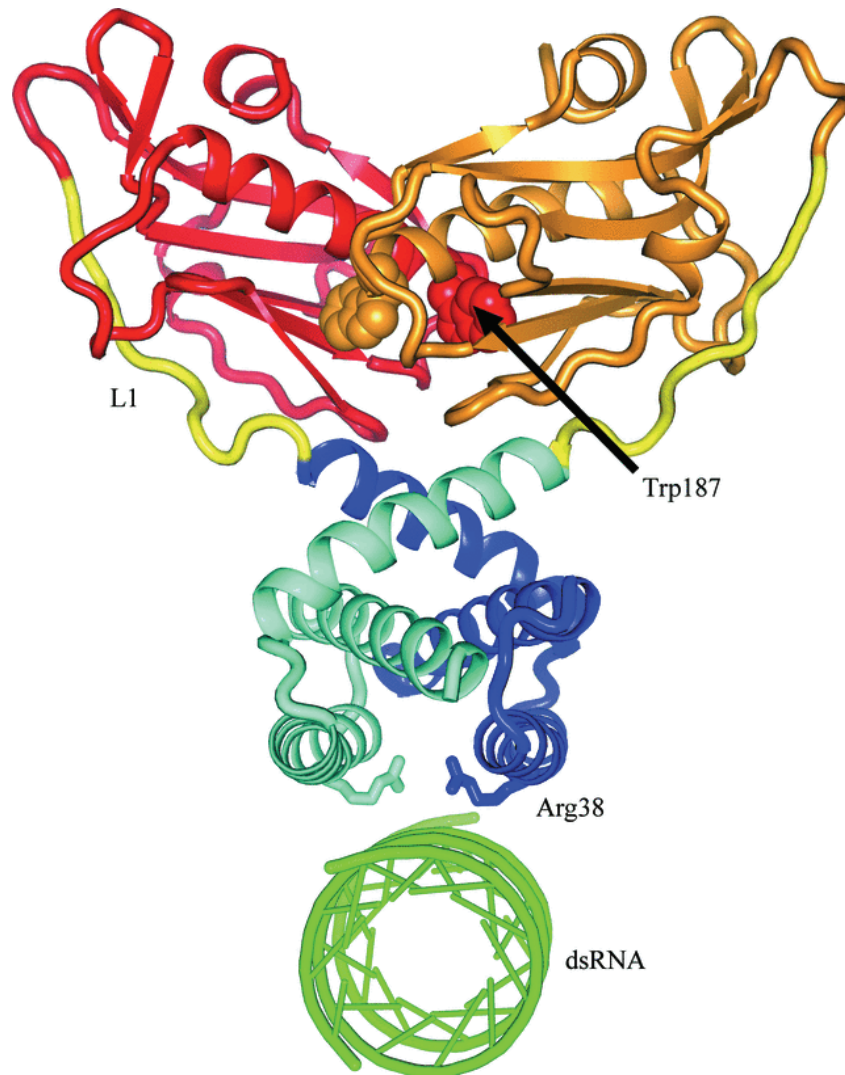
Figure 5.10: Hypothetical model of intact NS1. One chain has the N-terminal domain colored blue and the effector domain colored red; the second chain is colored cyan and orange, respectively. A ten-residue linker (yellow) joins the domains which have been observed crystallographically.

CPSF30 is likely to require a large quaternary structural change to the effector-domain dimer. Initially, the CPSF30-binding pocket is probably blocked by the Trp187 side chain of its dimeric partner, as seen in Figure 5.10. In order to bind CPSF30, the Trp187 side chain must be displaced and the effector domains rotate to accommodate CPSF30 phenylalanines 97, 98 and 103. It is clear from our model that the long unstructured linkers would allow the effector domains to rotate freely into the solvent from the stable N-terminal domain platform. Presumably, there is an equilibrium between the dimerized state that we observe in the crystal structure and a more open conformation. If that open form interacts with the F2F3 domain of CPSF30, those interactions are presumably much stronger than the internal dimer. The effector domains of NS1 proteins that do not bind CPSF30, such as the PR8 NS1, also form dimers with Trp187 pointing into hydrophobic pocket of an adjacent monomer and presumably also have some open conformation. However, without Phe at position 103 and Met at position 106 to help stabilize the F2F3 interactions, the internal effector-domain dimer may be more stable. This would explain why all influenza viruses have this conserved hydrophobic pocket, but only viruses with Phe at position 103 and Met at position 106 are able to break their internal dimer and bind CPSF30 (Twu et al., 2007).

5.3.3 NS1A effector domain mutants (W187A, W187Y)

The C-terminal effector domain (ED) has been reported to interact with the 30 kDa subunit of the cleavage and the polyadenylation specificity factor (CPSF30), and resulted in the inhibition of the maturation and export of host cellular antiviral mRNAs (Noah, Twu et al. 2003). NS1 mutations spanning the CPSF30-binding site (residues 184-188) reduce CPSF30 binding. The recombinant virus expressing these mutations induced a high level of host-cell interferon β, while viral replication was attenuated 1000-fold (Noah, Twu et al. 2003). In addition, an engineered MDCK cell line which constitutively expresses epitope-tagged F2F3 in the nucleus effectively blocked the binding of endogenous CPSF30 to NS1A and thereby selectively inhibited influenza virus A replication (Twu, Noah et al. 2006). All these suggest that the CPSF30 binding site can be targeted for the development of new antiviral drugs.

As shown in figure 5.11, the effector domain (residue 79 to 205) of NS1 was a dimer in solution. And from the X-ray structure, we know this dimer was stabilized by the pseudo-symmetrical insertion of Trp187 indole ring into the CPSF30 binding pocket of its dimer partner. In order to take the best advantage of this potential drug target, we tried to engineer a
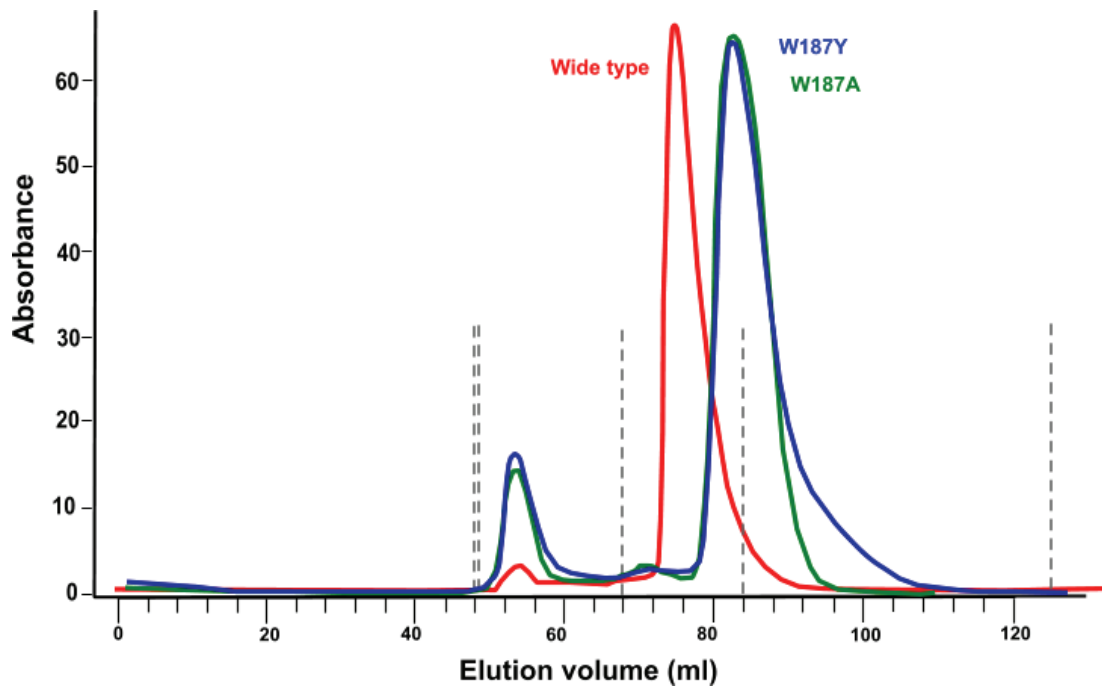
Figure 5.11: HPLC profiles of wild type and mutant NS1A-ED. The wild type elution profile is shown as a red line, the W187A mutant is shown in green, and the W187Y mutant is shown in blue. Absorbance was measured at 280 nm, and the gray dashed lines indicate the position of protein standards. From left to right, they are: thyroglobulin (670kD), bovine gamma globulin (158kD), chicken ovalbumin (44 kD), equine myoglobin (17kD) and Vitamin B12 (1.35kD).

mutant protein of effector domain that still retains the wild type binding affinity toward CPSF30 but remains monomeric in solution. In that way, this CPSF30 binding pocket of effector domain will be exposed to the solvent and to potential inhibitors.

The Trp187 residue was reported to be essential for dimerization of NS1-ED of avian influenza virus A/Duck/Albany/76 NS1 (Hale, Randall et al. 2008), which share 70% sequence identity with NS1-Ed of influenza A/Udorn/1997 (Figure 5.12). A superposition of the Udorn NS1-ED with Duck NS1-ED gives an rms deviation between 123 C_α atoms of 0.43 Å. As shown in 5.13, the Trp187 adopts a different conformation in each structure, as the orientation of indole ring is flipped 180° compared to each other. Interestingly, Trp187 is highly conserved in all NS1 protein of influenza A virus. Therefore, two mutant NS1-ED(s) were constructed with Trp187 converted to alanine and phenylalanine respectively.

Both recombinant, His tagged, effector domain mutants (W187A and W187Y) can be expressed in good yield from *E. coli* (~10 mg/ liter of cell culture). The multimeric state of the NS1-ED mutants was estimated by gel filtration. As shown in figure 5.11, both W187A and W187Y mutant proteins eluted in a volume equivalent to the size of a monomer. The SDS-PAGE analysis showed that the denatured wild type and two mutant proteins all exhibited the identical mobility, see figure 5.14.

A screen of crystallization conditions showed that both W187A and W187Y mutants form diffraction quality crystals in the same condition. Both types are orthorhombic with space group $P2_12_12_1$. The cell constants for W187A mutant is a = 47.78, b = 48.19, c =154.59Å. The W187Y mutant has a slight larger unit cell, a = 48.0, b = 60.24, and c = 132.66 Å. There are two molecules in the asymmetric unit, giving a Vm value of 2.78 Å^3/Da for W187A mutant, and 3.0 Å^3/Da for W187Y mutant. X-ray data and refinement statistics for the two structures are shown in Table 5.4. A section of the final 2Fo-Fc electron density map used for model construction is shown in Figure 5.15.

Following refinement, a Ramachandran plot of the W187A crystal showed 89.2% of residues to be in the most favorable region and 10.3% in the additionally allowed space. The refined structure includes 230 solvent molecules. The W87Y crystal had a Ramachandran plot with 87.3% of residues in the most favorable region and 12.7% in additionally allowed space. That refined structure includes 94 solvent molecules.

110

```
A/Duck/Alb     7 IAIASSPAPRYITDMSIEEISREWYMLMPRQKITGGLMVKMDQAIMDKRITLKANFSVLF
A/Udorn/72     1 MTMASTPASRYITDMTIEELSRDWFMLMPKQKVEGPLCIRIDQAIMDKNIMLKANFSVIF
                 ** ** ****** *** ** * **** **   * *     ******* * ******* *

A/Duck/Alb    67 DQLETLVSLRAFTDDGAIVAEISPIPSMPGHSTEDVKNAIGILIGGLEWNDNSIRASENI
A/Udorn/72    61 DRLETLILLRAFTEEGAIVGEISPLPSFPGHTIEDVKNAIGVLIGGLEWNDNTVRVSKTL
                 * ****  *****   **** **** ** ***  ******* **********   * *

A/Duck/Alb   127 QRFAWG
A/Udorn/72   121 QRFAWG
                 ******
```

Figure 5.12: Alignment of the protein sequence of NS1A effector domain from influenza A/Udorn/1997 and avian influenza A/Duck/Albany/76 NS1.
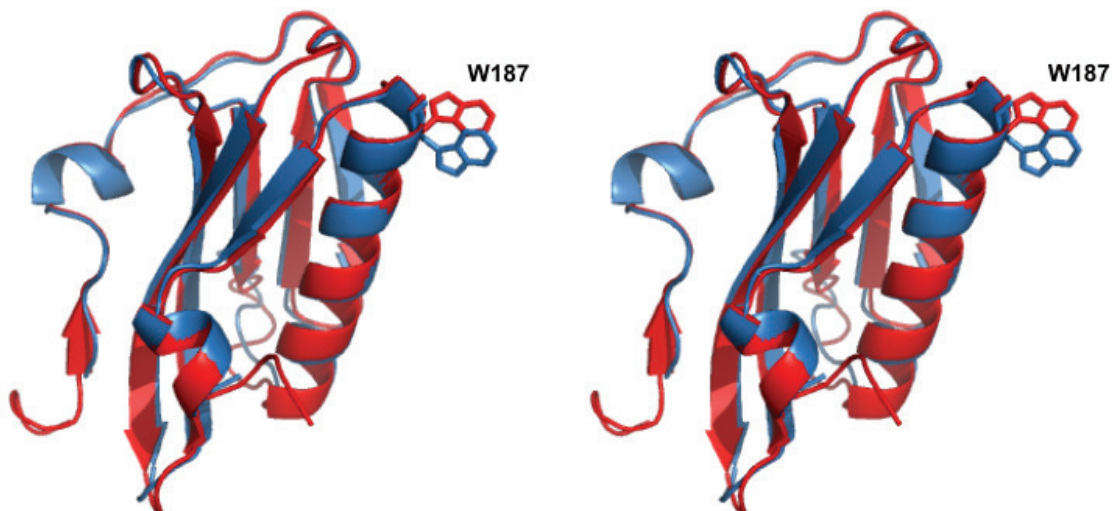
Figure 5.13: Superposition of the influenza virus A/Udorn/1997 NS1 effector domain (red ribbon) and the avian influenza A/Duck/Albany/76 NS1 effector domain (blue ribbon)
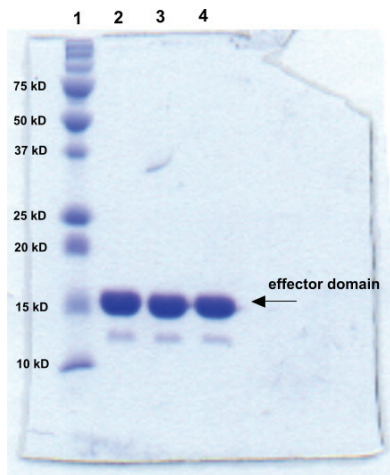
Figure 5.14: SDS-PAGE analysis of wild type and mutant NS1A-ED.

Data Collection	W187A mutant	W187Y mutant
Wavelength (Å)	1.5418	1.5418
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	47.78, 48.19, 154.69	47.99, 60.24, 132.66
Resolution (Å) (last shell)	30 – 2.21 (2.21-2.29)	30 – 2.21 (2.21-2.29)
R_{merge} (%) (last shell)	3.7 (9.8)	4.8 (24.2)
$<I/\sigma I>$ (last shell)	78.9 (23.9)	51.4 (6.0)
Completeness (last shell)	99.3 (92.8)	99.0 (90.5)
Redundancy	13.4	10.0
Refinement		
No.reflections	17384	18310
$R_{working}$ (last shell)	0.187 (0.202)	0.226 (0.295)
R_{free} (last shell)	0.233 (0.285)	0.258 (0.305)
Average B factor for protein atom ($Å^2$)	29.4	37.9
R.m.s deviation from ideality		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.138	1.190

Table 5.4: Data collection and model refinement statistics of effector domain of NS1 mutants.

(a)

(b)
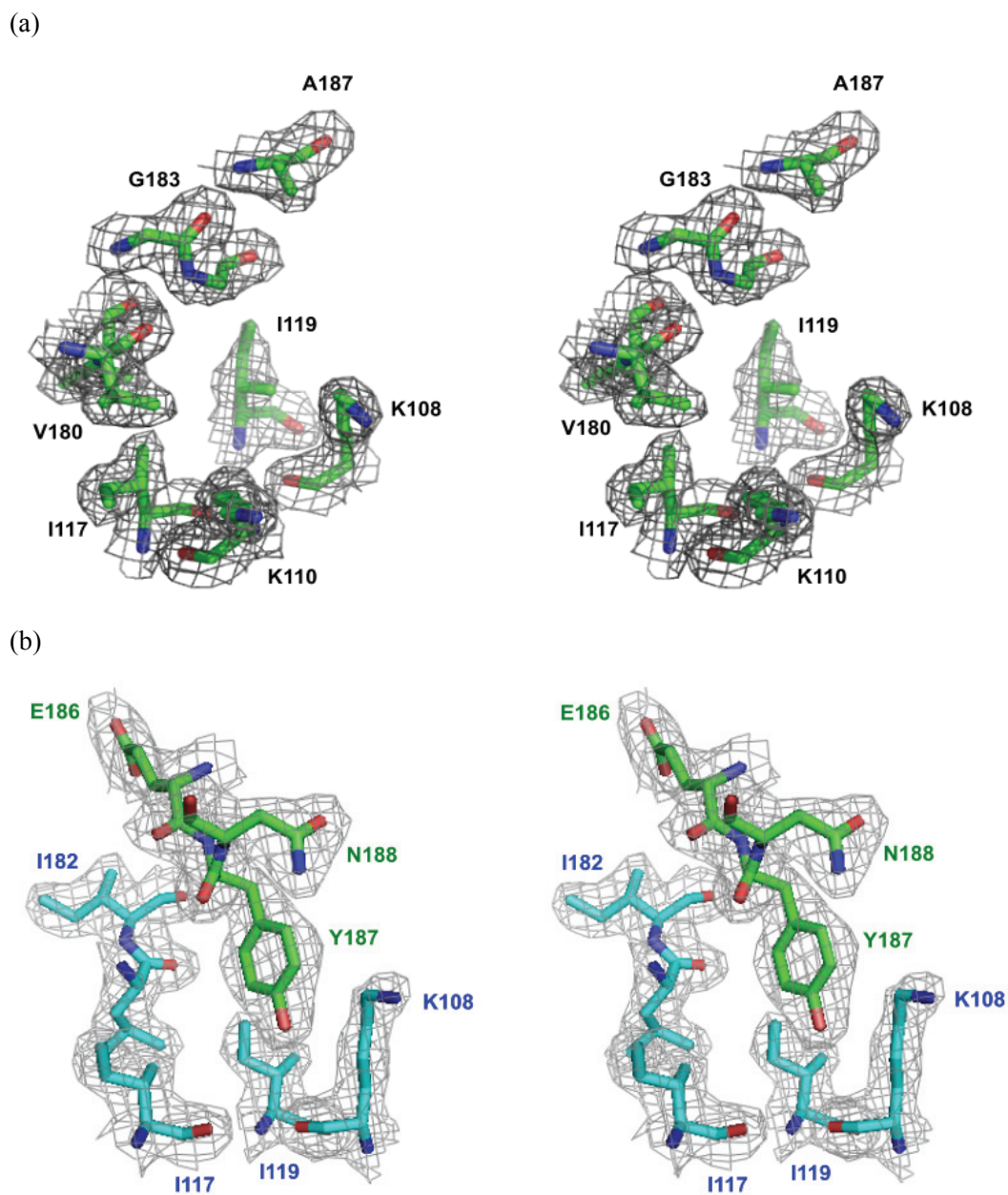
Figure 5.15: Electron density for the (a) W187A and (b) W187Y mutant of NS1A effector domain. (a) This is a section of a $2F_0$-F_c map contoured at 1σ to show the exposed CPSF30 binding pocket. (b) This is a section of a $2F_0$-F_c map contoured at 1σ to show the interface of the ED dimer due to crystallographic packing.

The structures of the mutant NS1-ED, including W187A and W187Y, show an α-helix β-crescent fold generally identical to that of wild type protein (Xia, Monzingo et al. 2009). As shown in figure 5.16, a superposition of the mutant ED with wild type ED gives an rms distance between 123 C^{α} atoms of 0.16Å and 0.39Å for W187A mutant and W187Y mutant respectively.

Although both mutants of NS1-ED behave as monomers in solution, based on chromatographic properties, the asymmetric unit in both crystals has two molecules. The two molecules of the W187Y mutant form a similar dimer to the wild type NS1-ED. As shown in figure 5.17a, the main interaction for this noncrystallographic dimer formation appears to be the pseudo-symmetrical insertion of Tyr187 phenol group into the hydrophobic F2F3-binding pocket of its dimer partner. This pocket is located at the base of long central helix, and formed by the side chain residues Lys108, Lys110, Ile117, Gln121 and Val180. Superposition of the W187Y mutant dimer and wild type NS1-ED dimer a revealed similar dimer interface, see figure 5.18a. The phenol group of the mutant protein Tyr187 lies in the same plane as the indole ring of wild typeTrp187 (Figure 5.18b). The hydroxyl end of phenol group shifts 10° toward the edge of the hydrophobic pocket of its dimer partner as compared to the orientation of indole ring. As phenylalanine is 30% less hydrophobic than tryptophan based on the hydrophobicity index of amino acid measured at pH7.0, there is certainly less attraction force of this hydrophobic interaction. Because W187Y mutant exits as a monomer in solution, this dimer interaction observed is probably very weak and be supported by other crystal packing interactions. In contrast, the two molecules of W187A form a different dimer than the wild type and W187Y mutant. Illustrated in figure 5.17b, the hydrophobic F2F3-binding pocket is exposed to the solvent in the W187A mutant.

CPSF30 contains five C3H zinc finger repeats, from which the second and third finger repeats (F2F3) alone are sufficient for efficient binding to NS1A. (Barabino, Hubner et al. 1997; Twu, Noah et al. 2006). To see if the two mutant NS1A-EDs are able to bind CPSF30 or not, we expressed GST fusions of the N-terminal 145 residues of CPSF30, which includes the first four zinc finger repeats, in bacteria and used this GST fusion in pull-down assays with wild type and mutant NS1-EDs. As expected, the GST-CPSF30(1-145) can bind wild type NS1-ED effectively, see figure 5.19. As shown in the SDS-PAGE profile, there are three bands
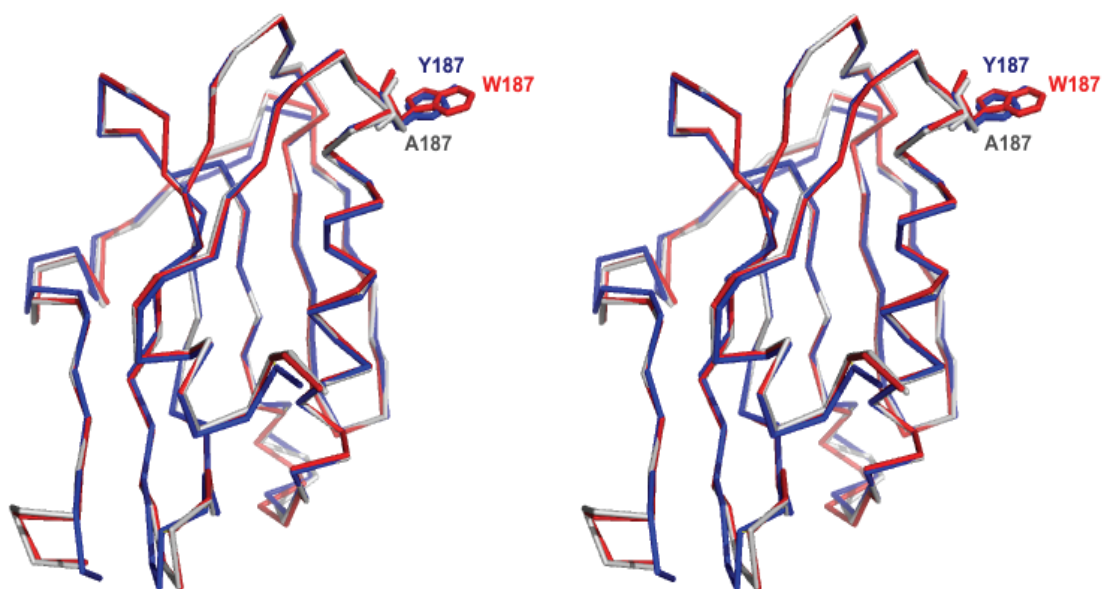
Figure 5.16: Superposition of wild type effector domain with W187Y and W187A mutant effector domains: (a) overview and (b) detailed view of residue 187. The C^α trace of wild type effector domain is shown in red bonds, the C^α trace of W187Y mutant effector domain is shown in blue bonds, and that of W187A mutant effector domain is shown in grey bonds. The side chain of W187, Y187 and A187 are highlighted in stick mode.
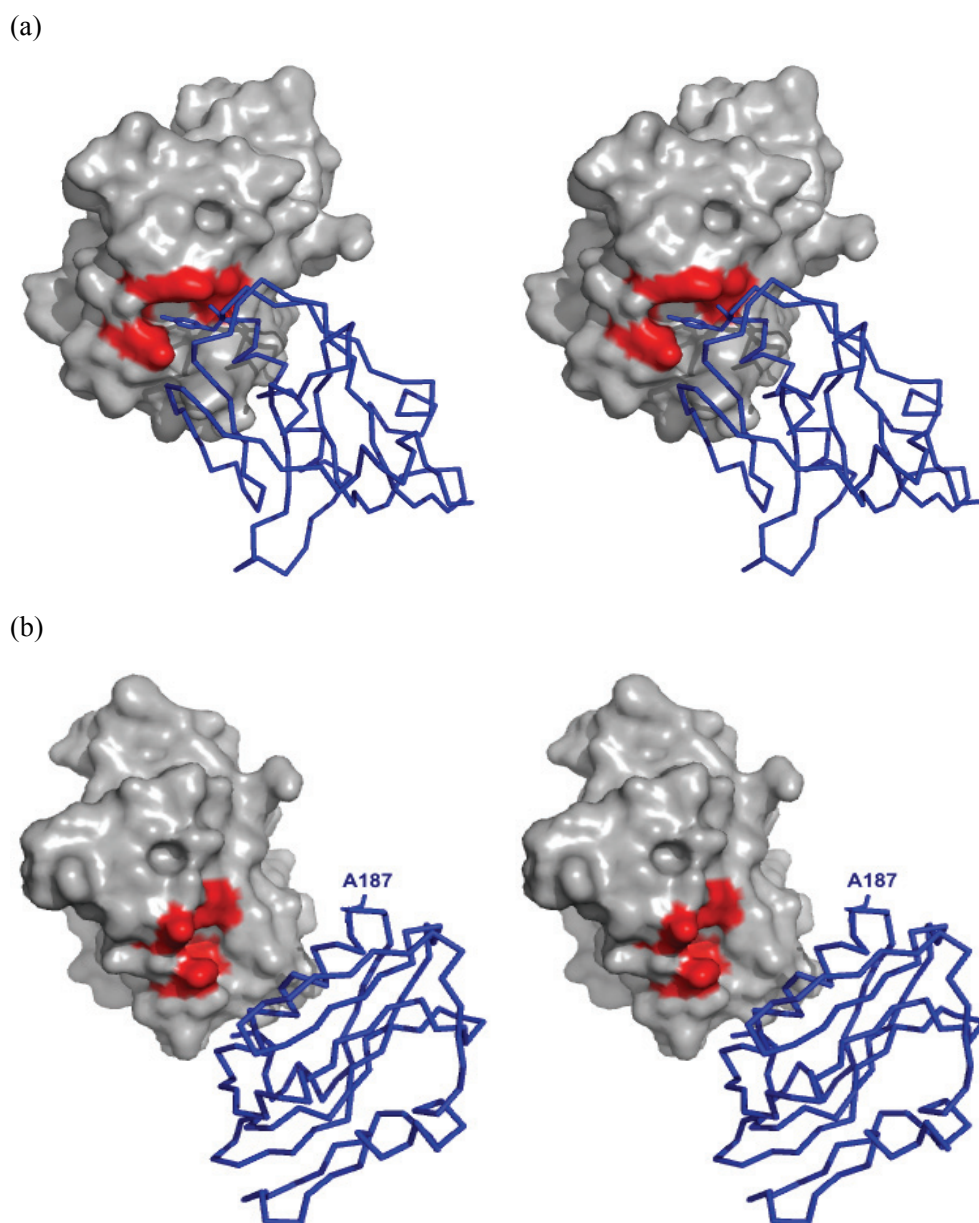
Figure 5.17: The crystallographic dimer interface for (a) W187Y and (b) W187A mutant NS1-effector domain. A space filling model of crystallographic monomer A reveals a distinct, largely hydrophobic pocket that has evolved to bind aromatic residues from the F2F3 domain of CPSF30. Here this pocket is filled by the Y187 of crystallographic monomer A of W187Y mutant, and is exposed to solvent in W187A mutant.

(a)

(b)
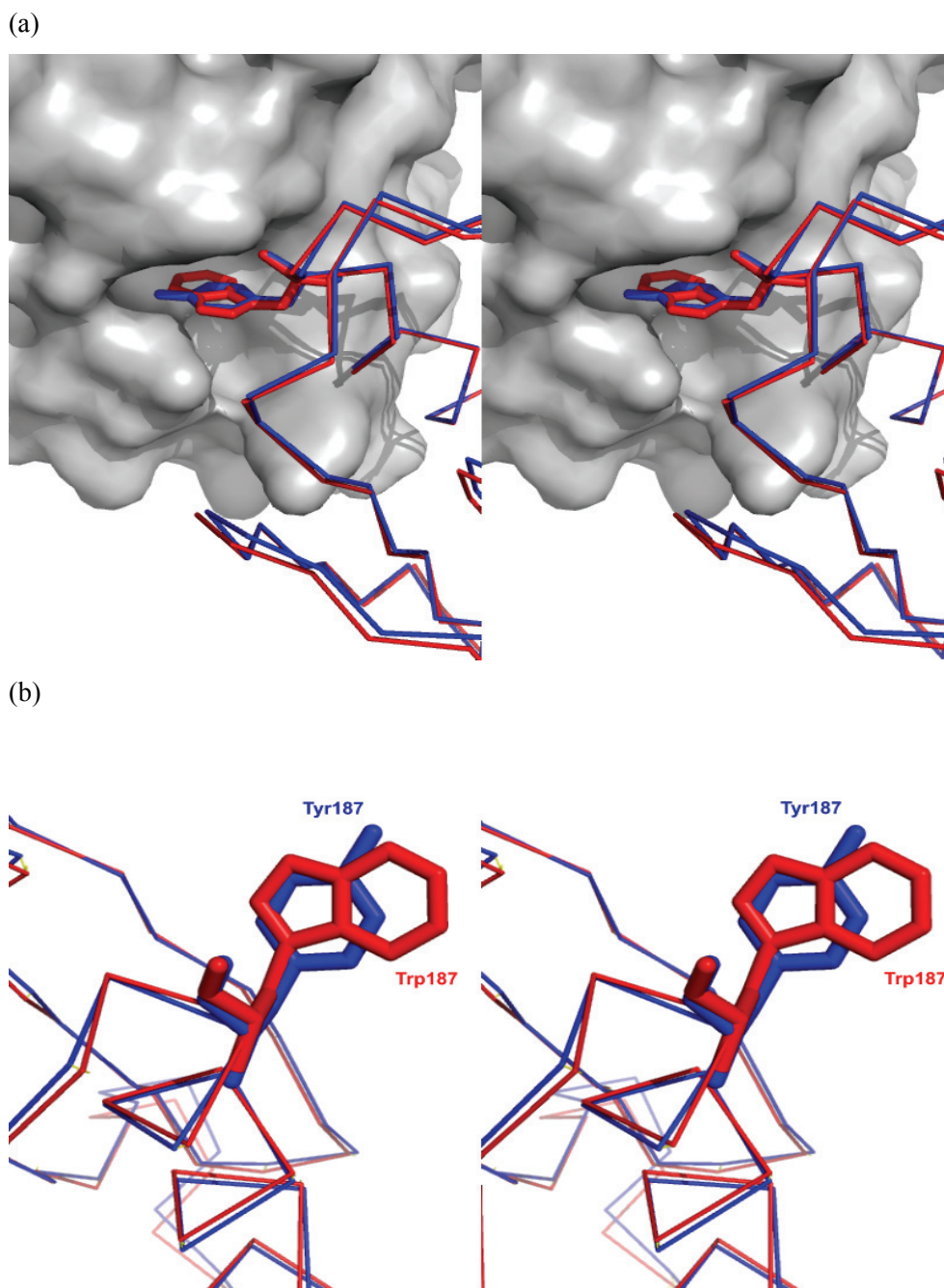
Figure 5.18: Superposition of the dimer interface of wild type NS1-ED with W187Y mutant ED: (a) overview and (b) detailed view of residue 187. One monomer was displaced in space-filling mode, while the other monomer is shown in ribbon, with the wild type in red ribbon and the W187Y mutant in blue ribbon.

117

in the elution fraction, corresponding to the 43 kD GST-CPSF30(1-145), the 16kD NS1-ED and the 26kD GST protein. The expression of GST alone is probably due to a misreading of the stop condon or the falling off of the ribosome at the end of GST expression. This leaky expression of GST has been observed in the production of other GST fusion proteins in the laboratory. The negative control indicates that NS1-ED alone is not trapped in the glutathione affinity column. To exclude the possibility that GST itself might interact with NS1-ED, we also conducted a pull-down assay with purified GST protein. As shown in figure 5.20, GST did not bind NS1-ED. In other words, GST did not contribute to the interaction observed between the GST-CPSF30(1-145) and NS1-ED. The GST-pull down assay of mutant NS1-ED(s) showed that both mutants are able to bind GST-CPSF30(1-145), see figure 5.21. The binding of CPSF30 to W187A mutant might be weaker than that to W187Y mutant, but the conversion of tryptophan to alanine did not abolish the interaction between them. It's very likely that the W187Y mutant has the same binding affinity toward CPSF30 as does the wild type protein.

In conclusion, we have engineered a W187Y mutant of NS1-ED, which behaves as a monomer in solution but is still able to binding CPSF30 with similar binding affinity as the wild type protein. This mutant may be a better target for a HT screening assay since the binding pocket is more accessible to potential inhibitors.

5.3.4 NS1A "full length" (residue 1 -205) mutant

Exhaustive crystal screening has been conducted for the N-terminal GST tagged NS1A(1-215) fusion protein, but no useful crystallization condition has been identified. Our initial goal was to analyze the structure of this GST fusion protein, since the GST-NS1A(1-215) has been chosen as the screening target for the HT FP assay. Such a structure would be very helpful for virtual screening and rational lead optimization.

The recent X-ray structure of R38A-K41A double mutant of full length NS1A from an H5N1 strain (A/Vietnam/1203/2004) revealed new insights into the dimer interaction between the N-terminal RNA binding domain and the C-terminal effector domain (Bornholdt and Prasad 2008). It turns out that the two domains of each NS1A molecule separately interact with their respective domains from the neighboring NS1A molecules, and form a chain of NS1A molecules with alternating dimers of RBD and dimers of effector domain. The authors proposed that NS1A cooperatively oligomerizes in the presence of dsRNA to form a tubular
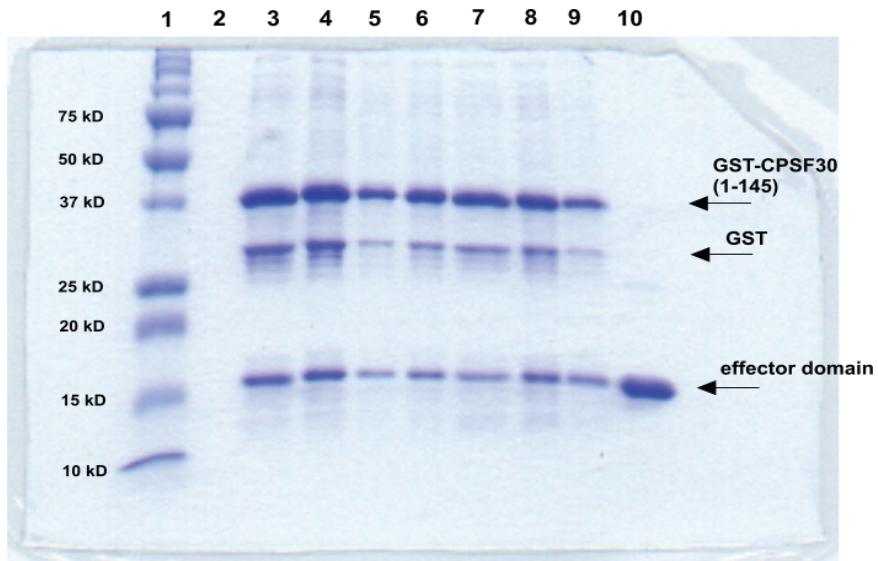
Figure 5.19: SDS-PAGE analysis of binding of GST-CPSF30(1-F4) to wild type NS1-ED. Lane 1 is the protein standard marker; Lane 2 is the negative control to see if the effector domain will be trapped in the glutathione beads. Lanes 3 to 9 are fractions of eluted complex from the column, they are fraction 2,3,7,5,6,4,8 (from left to right). Lane 10 is the purified wild type NS1-ED.
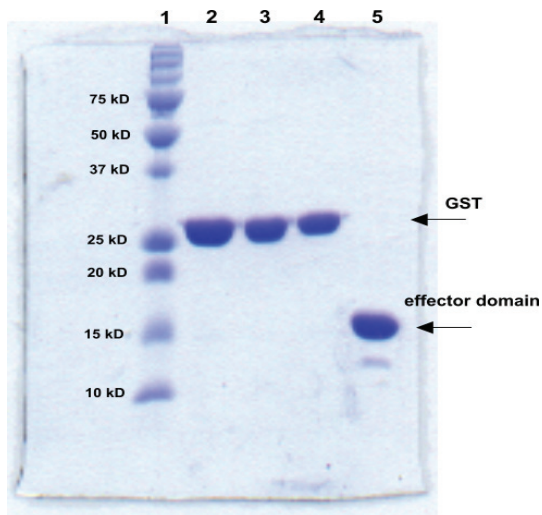
Figure 5.20: SDS-PAGE analysis of binding of GST to NS1-ED. Lane 1 is the protein standard marker. Lane 2 to lane 4 are the eluted fractions of GST; and they are: fraction 2,3 and 4 (from left to right). Lane 5 is the purified NS1-ED protein.
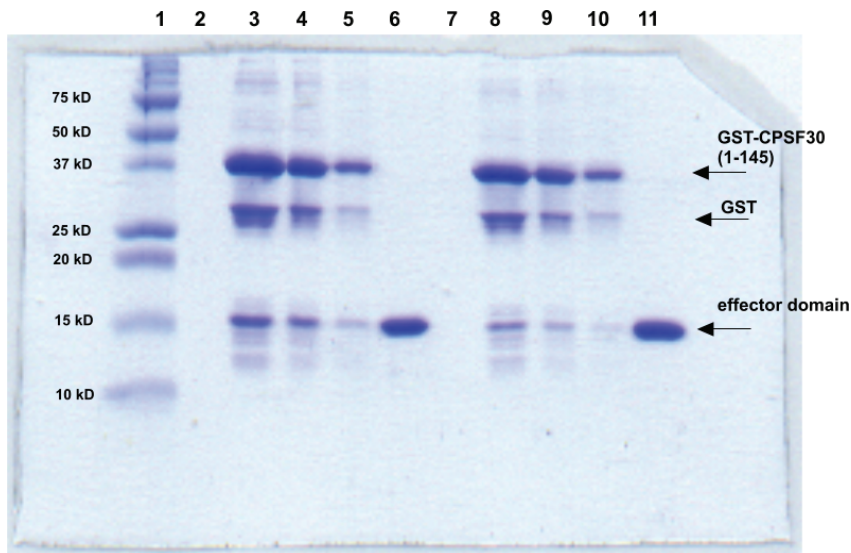
Figure 5.21: SDS-PAGE analysis of binding of GST-CPSF30(1-F4) to W187Y and W187A mutant NS1-ED. Lane 1 is the protein standard marker; Lane 2 is the negative control to see if W187Y mutant ED will be trapped in the glutathione beads. Lanes 3 to 5 are the eluted fractions of GST-CPSF30(1-F4) and W187Y complex; they are fraction 2,3 and 4 (from left to right). Lane 6 is the purified W187Y mutant NS1-ED. Lane 7 is the negative control to see if W187A mutant ED will be trapped in the glutathione beads. Lanes 8 to10 are the eluted fractions of GST-CPSF30(1-F4) and W187A complex; they are fraction 2,3,4 (from left to right). Lane 11 is the purified W187A mutant NS1-ED.

tructure with dsRNA in the hollow central tunnel. This model explains why NS1A is able to sequester varying lengths of dsRNA in infected host cells. To explore this model, we tested a 46mer dsRNA with NS1A(1-205) and GST-NS1A(1-215) respectively by fluorescence polarization. As shown in figure 5.22, the FP signal reached its plateau once the concentration of GST-NS1A(1-215) was above 2 μM. Interestingly, the FP signal continuously increases when the concentration of NS1A(1-205) was tested. Judging from the binding curve of NS1A(1-205) to dsRNA, it seems there are multiple events going on. First, the 46mer dsRNA recruits NS1A(1-205) dimer from the solution; then, incoming NS1A molecules gradually form a chain along the dsRNA. For the GST fusion protein, the N-terminal GST tag appears to prevent the formation of alternating dimers and the titration curve reaches a plateau as the structure self limits. The binding of GST-NS1A(1-215) to 46mer dsRNA give a K_d around 0.35 μM, which is comparable to the binding affinity of that to 16mer dsRNA (0.16μM). Our FP assay with 46mer dsRNA suggests that the model Prasad proposed is probably correct.

The N-terminal 205 residue protein of NS1A from A/Udorn/72 was also cloned into a His-tag vector pNIC, with Arg38 and Lys41 both converted to alanine; this should prevent aggregation seen with the wild type protein (Bornholdt and Prasad 2008). The mutant protein was expressed in bacteria with good yield, and behaved as a homo dimer as estimated by gel filtration, see figure 5.23. Interestingly, once the Trp187 was also converted to alanine, the triple mutant (R38A-K41A-W187A) behaved as a monomer in solution (Figure 5.23). That indicates that the Trp187 is not only essential for the dimerization of effector domain, but is also a major force in the dimerization of the whole protein. We have tried crystal screening for the R38A-K41A double mutant and the R38A-K41A-W187A triple mutant of NS1A(1-205), but no crystallization conditions have been identified so far.
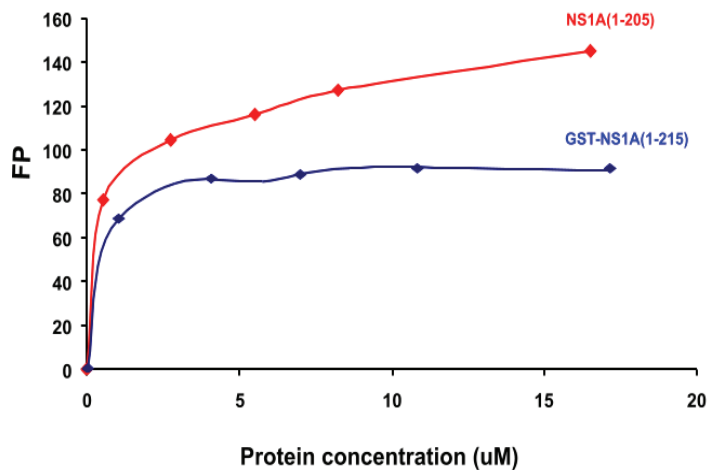
Figure 5.22: FP signal of 5'-fluorescein labeled 46mer-dsRNA binding to the GST-NS1A(1-215) and NS1A(1-205).
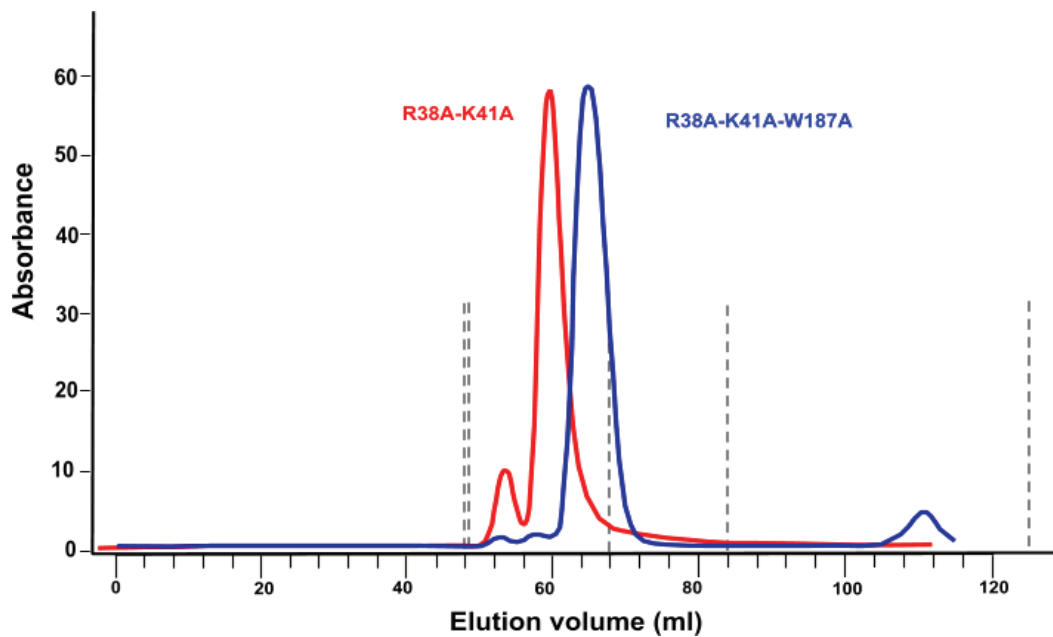
Figure 5.23: HPLC profiles of R38A-K41A mutant (red line) and R38A-K41A-W187A mutant (blue line) of NS1A(1-205). Absorbance was measure at 280nm, and the gray dashed lines indicate the position of protein standards. From left to right, they are: thymoglobulin (670kD), bovine gamma globulin (158kD), chicken ovalbumin (44 kD), equine myoglobin (17kD) and Vitamin B12 (1.35kD).

Project II: Structure and Kinetics Studies of the MobA Protein

MobA is an enzyme catalyzing the conjugative mobilization of the plasmid R1162, a mobilization plasmid that transfers drug resistance genes across species. Although it is clear that MobA catalyzes a single strand cleavage of the plasmid DNA at the oriT site by forming a covalent bond between Tyr25 and 5' phosphate of the nicked strand, the overall reaction mechanism is still unknown. This dissertation describes work to further investigate some details of this important and unusual enzyme mechanism, including the role of metallic cations.

The X-ray structure of the N-terminal 184-residue relaxase domain, called minMobA, has been used to guide the mutagenesis studies aimed at identifying key amino acids for the transesterification mechanism. The DNA nicking assay results from a series of minMobA mutations revealed no obvious candidate for the role of a general base to polarize the active site Tyr 25. On the other hand, MobA is a very unusual enzyme, in that it carries out a single reaction turnover. The nicking reaction is not the rate-limiting step compared with DNA transfer during bacteria conjugation, which means the activation of the tyrosine nucleophile is not mechanistically critical. Therefore, it is not biologically important for MobA to enhance the rate of transesterification.

In addition, the effects of different divalent metallic cations on minMobA were investigated by means of circular dichroism and protein denaturation experiments. The results suggest that divalent metallic ions may affect the conformations of active site residues of minMobA; they may organize key side chains and the local backbone structure to favor DNA cleavage.

123

Chapter 6: Introduction

6.1 Overview of horizontal gene transfer

Horizontal gene transfer (HGT), also known as lateral gene transfer, refers to the transfer by and organism of its own genetic materials to another organism which is not the offspring of that organism. By contrast, vertical gene transfer occurs when parental traits pass to progeny by sexual reproduction.

HGT is particularly common among bacteria, and is thought to be the major factor in accelerating the rate of bacteria evolution. HGT is also a significant cause of increased drug resistance. Although bacteria can develop drug resistance by mutation of existing genes, it is far more common for resistance to be spread by transferring of existent resistant genes (Tenover 1995; Davies 1996; Wright 2007). Additionally, by HGT, genetic materials can be exchanged promiscuously between a broad spectrum of bacteria, archaea and eukarya, including plants, fungi and mammalian cells. Thus, it significantly influences both our health and the environment, and receives centered attention from a wide range of scientific disciplines.

In prokaryotes, there are three common mechanisms for horizontal gene transfer: transformation, transduction and conjugation (Thomas and Nielsen 2005). In transformation, a cell is genetically altered by uptake, recombination and expression of foreign genetic materials. This stable genetic change, brought by transformation, could gain new functions for the cell or lose the activity of the gene which is replaced due to the recombination (Griffith 1928; Avery and Macleod 1944). The state of being able to take up exogenous genetic materials (DNA or RNA) from the environment is referred to as competence. Bacteria which are naturally competent to act as recipients have been discovered to exist in several genera including Acinetobacter, Bacillus, Neisseria, haemophilus, Pneumococcus and Pseudomonas (Lorenz and Wackernagel 1994). In addition, some bacteria can become competent artificially. In other words, cells can be made passively permeable to exogenous genetic materials by laboratory procedures, such as electroporation or chemical treatment. (Droge, Puhler et al. 1998).

Transduction is a process involving transfer of foreign genetic materials from one bacterium to another with the help of phage, virus or virus vector. Transduction can occur through either the lysogenic cycle or the lytic cycle (Parkinson 1975). In a lysogenic cycle, the

phage or virus chromosome is integrated into the host bacterial chromosome, and thus can be transmitted to daughter cells at subsequent cell divisions. In lytic cycles, usually induced by UV radiation, lysis of the infected host cell occurs, and results in the release of new phage particles or progeny viruses. Because viruses can transfer genetic materials across species boundaries, transduction by virus can occur in eukaryotic cell as well. In addition, the HGT by viruses is essential for evolutionary progress (Sorensen, Bailey et al. 2005; Thomas and Nielsen 2005).

Unlike transformation and transduction, conjugation is a mechanism involving unidirectionally transfer of genetic material between bacteria through direct cell-to-cell contact. During conjugation, a piece of DNA from one bacterium (donor) is copied and then transferred to another bacterium (recipient) via a temporary connection. In nature, the genetic information transferred is often beneficial to the recipient. Those benefits may include resistance to new antibiotic, tolerance to new xenobiotic, or the ability to usee a new metabolite. The process of conjugation is facilitated by diverse mobilization plasmids and conjugative transposons, and it constitutes the major route for HGT (Sorensen, Bailey et al. 2005; Thomas and Nielsen 2005). Remarkably, conjugation can occur not only in one species but also between species, in some cases, from bacteria to fungal, plant or mammalian cells(Stachel and Zambryski 1986; Droge, Puhler et al. 1998; Waters 2001). These properties make conjugation an important source of genetic plasticity.

6.2 Bacteria conjugation

Although bacterial conjugation is a highly specific process, the mechanism of conjugative transfer appears to be remarkably conserved across a broad spectrum of plasmids in both Gram-negative and Gram-positive bacteria. It generally involves the following steps: (1) establishment of direct cell-to-cell contact; (2) occurrence of some kind of signal that transfer should begin; (3) relaxosome formation at the origin of transfer; (4) nicking of one DNA strand at the nic site with the nickase enzyme covalently attached to the 5' end of DNA; (5) unwinding of the cleaved strand from the duplex plasmid; (6) transfer of the cleaved strand in the 5' to 3' direction to the recipient cell; (7) recircularization of the transferred strand; (8) synthesis of the complementary strand in both the donor and recipient cell (Lorenz and Wackernagel 1994). A simplified scheme of bacteria conjugation is illustrated in figure 6.1. The relaxosome is a complex of proteins that recognize a specific DNA sequence (called nicking site) and facilitate
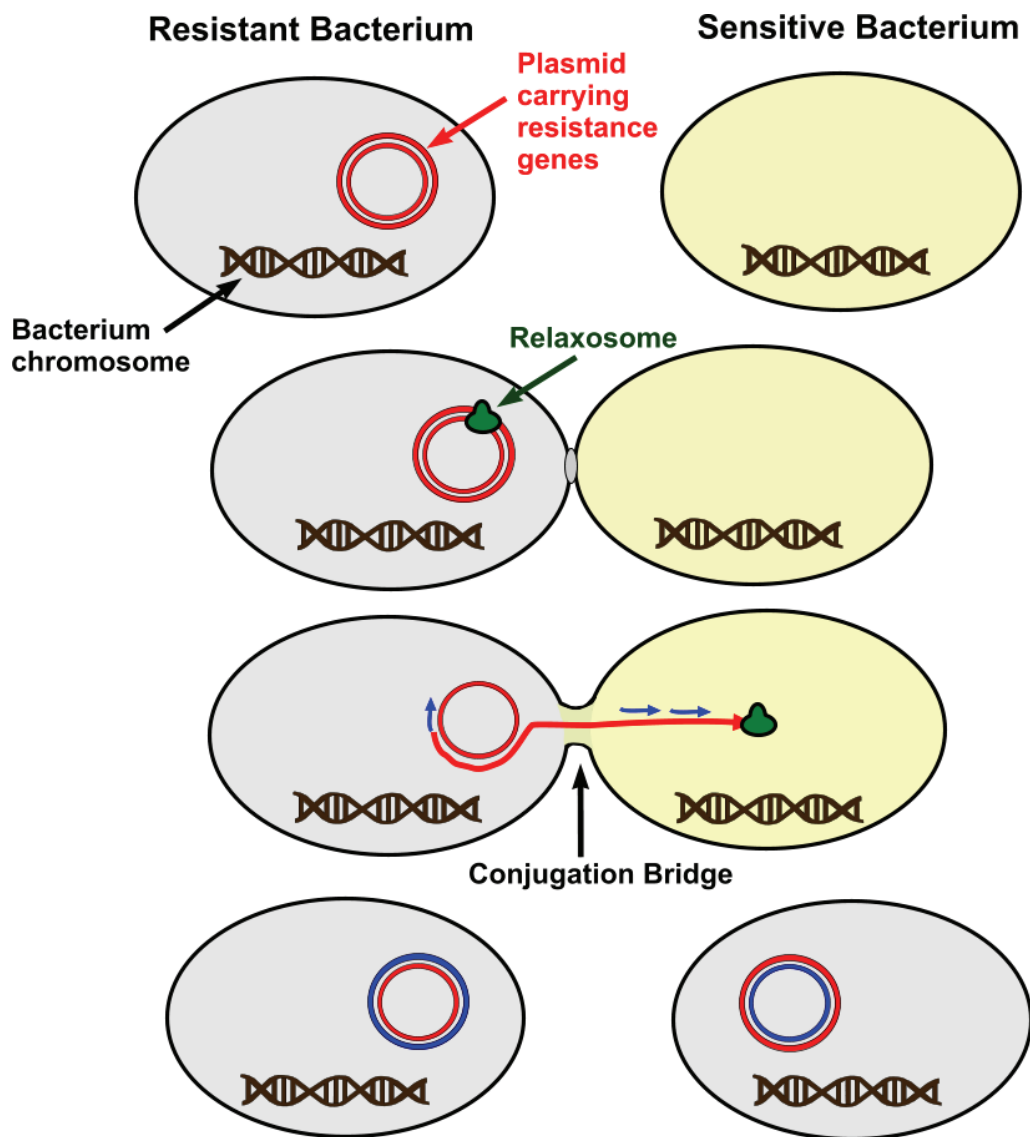
Figure 6.1: Transfer of antibiotic resistance genes by bacterial conjugation

plasmid transfer during bacteria conjugation. The direct cell-to-cell contact during conjugation is mediated by a specialized multiprotein-complex, called conjugation apparatus. In Gram-negative bacteria, this physical contact is established by sex pili, a hair like appendage found on the surface of bacteria. Whereas, for the majority of Gram-positive bacteria, the means to achieve this direct cell-to-cell contact still remains to be identified.

6.2.1 Conjugation in Gram-positive bacteria

In 2000, Zechner and Zatyka divided conjugative plasmids from Gram-positive bacteria into four groups: broad host-range plasmids, pheromone responding plasmids, conjugative transposons and plasmids from mycelium-forming *streptomycetes* (Grohmann, Muth et al. 2003). The broad host-range plasmids are exemplified by pSK41 family plasmids and pIP501 plasmid (Allignet and El Solh 1999; Steinmetzer, Kuhn et al. 2002). The pSK41 family plasmids were identified in both *S. aureus* and coagulase-negative *staphylococci* (Firth, Ridgway et al. 1993). They encode resistance genes against neomycin, gentamicin, tobramycin, kanamycin, and some antiseptics and disinfectants (Grohmann, Muth et al. 2003). The pIP501 plasmid, originally isolated from *streptococcus agalactiae*, encodes resistance genes against chloramphenicol and erythromycin (Evans and Macrina 1983). The plasmid can be conjugated into *Streptococcus*, *Staphylococcus*, *Clostridium*, *Listeria* and *Pediococcus* species. The pAD1 plasmid isolated from *Enterococcus faecalis* is a representative of pheromone responding plasmids. It is highly conjugative and encodes multiple antibiotics resistance genes, such as erythromycin, tetracycline, streptomycin and kanamycin (Nakayama and Suzuki 1997; Spataro, Toda et al. 1997). Conjugative transposons combine features of transposons, plasmids and bacteriophages, in sense that they excise from and integrate into DNA. They are very common among the Gram-positive *streptococci and enterococci*. The 18kb Tn916 is the most extensively studied conjugative transposon; it encodes a resistance gene against tetracycline (Marra and Scott 1999; Marra, Smith et al. 1999). During the life cycle of *streptomyces*, the bacteria interact conjugally to promote the transfer of their own DNA. It was reported that all plasmids from mycelium-forming *streptomyces* are conjugative (Tiffert, Gotz et al. 2007). Among them, the pIJ101 has been used as a model system to study conjugative plasmids in this category. Originally purified from *Streptomyces lividans*, pIJ101 uses a rolling-circle replication (RCR) mechanism, a process that involves a Rep protein and a double-strand origin (DSO) of replication. The efficient transfer of pIJ101 is dependent on both the membrane

127

protein encoding Tra gene, and a cis-acting transfer locus (Servin-Gonzalez 1993; Dyson and Evans 1998). In 2003, Espinosa proposed to put this type of multicellular bacteria conjugation mechanism into a separate category, in contrast to the conjugation mechanism taking place in unicellular gram-positive bacteria (Grohmann, Muth et al. 2003).

6.2.2 Conjugation in Gram-negative bacteria

Compared to Gram-positive bacteria, conjugation in Gram-negative bacteria has been studied in great detail over the last decades, and most thoroughly through the analyses of antibiotic resistance plasmids of Gram-negative bacteria. The plasmids involved in conjugative transfer can be categorized into two classes: the self-transmissible plasmids and the mobilizable plasmids.

6.2.2.1 Self-transmissible plasmid

Conjugative plasmids are considered to be self-transmissible if they encode a self-sufficient conjugative transfer system. On the basis of genetic relatedness and pilus structure, there are four classes of self-transmissible plasmids: IncF-like plasmids, IncP-like plasmid, Ti group plasmids, and IncI-like plasmids (Reddy, Battisti et al. 1987; Waters 1999). The plasmids in the first two groups appear to account for most of the drug resistance gene transfer. Here, "Inc" is an abbreviation for incompatibility, which is the inability of two plasmids to be propagated stably in the same cell line. Incompatibility is a manifestation of relatedness, in other words, the sharing of common elements involved in plasmid replication and control.

The first group is the IncF-like plasmids. Being a representative of that family, IncF F plasmid is the prototype for conjugative plasmids, and is usually referred to as F factor or sex factor. In a given bacterium, there can only be one copy of the F plasmid. The 100kb-length F plasmid was originally obtained from a host E. coli K-12 strain (Finlay, Frost et al. 1986). It has an origin of replication (oriV), an origin of transfer (oriT), a Tra and a Trb locus, which together are about 33 kb long and contain about 40 genes (Taylor, Harrison et al. 1990). When conjugation is initiated, a relaxase enzyme called TraI creates a nick in one strand of the F plasmid at oriT. The cleaved strand is unwound from the duplex plasmid, and then transferred in the 5' to 3' direction to the recipient bacterium. Once inside the recipient bacterium, the transferred strand is recircularized. The complementary strand is synthesized in the donor and recipient bacterium, both of which are then capable of plasmid transfer (Finlay, Frost et al.

1986; Saul, Spiers et al. 1989). The entire process of F plasmid transfer takes approximately 5 min to complete (Ippen-Ihler and Minkley 1986).

The second group consists of the IncP-like plasmids, including IncP, IncU and IncW (Frost, Ippen-Ihler et al. 1994). They are believed to be the most widely distributed broad-host-range plasmids carrying multiple antibiotic resistance genes. The 60kbRP4 plasmid is the prototype of IncP plasmids. Conjugative functions of plasmid RP4 are encoded in two distinct regions on the plasmid known as Tra1 and Tra2. Those Tra regions were originally identified and evaluated by deletion analysis (Lessl, Pansegrau et al. 1992). There are three genes in the Tra1 operon: the primase, the relaxase, and the leader operon. The function of the RP4 plasmid can be divided into two parts: DNA transfer and replication (Dtr), and mating pair formation (Mpf) (Pansegrau, Lanka et al. 1994). All Dtr genes locate exclusively in Tra1, and function in the initial relaxosome formation. The Mpf system is the establishment of direct cell-to-cell contact between a donor and a recipient cell during conjugation. The Mpf genes map mostly in Tra2 region, and include one Tra1 gene, traF. Transfer of genetic material is thought to be initiated via TraI-piloted single-stranded intermediates (Thorsted, Macartney et al. 1998). Coated with TraC protein, the protein nucleotide complex is transported through a channel or pore at the mating bridge between the donor and the recipient cells.

The IncW plasmids are the smallest self-transmissible plasmids naturally found in Gram negative bacteria (Bolland, Llosa et al. 1990), but are in the IncP family. The 33kb-length plasmid R388 is the prototype of IncW plasmids. Because of its small size and broad host range, R388 has been a good example for the analysis of the genetic organization of conjugative plasmid (Llosa, Bolland et al. 1991). R338 has the shortest Dtr region, which consists of one OriT and three genes (trwA, trwB and trwC). TrwA is a nic-cleavage accessory protein that binds to sites around OriT and enhances the relaxase activity. TrwB is a typical integral membrane protein that couples the relaxosome to the nucleotide transport system. TrwC is the actual relaxase responsible for both double strand nucleotides unwinding and DNA nicking. The Mpf region is composed of 10 genes, responsible for mating pair and DNA-transport complex formation. Another characteristic of IncP-like plasmids is that they contain several regions which are sites for interaction with other transfer systems. For example, the osa region interferes with the DNA transfer complex of the Ti plasmid (Llosa, Bolland et al. 1994).

The third group, similar to the IncP-like plasmid but with a different pilus structure, is referred to as the Ti plasmid group, including IncX and IncN. The 38kb-length plasmid R6K is the prototype of IncX plasmid (Bastia, Germino et al. 1981). It contains three origins of vegetative replication (a-oriV, b-oriV and g-oriV) and two antibiotic resistance genes (ampicillin and streptomycin). Both a-oriV region and b-oriV region contain a palindromic sequence, which is involved in the initiation of replication. When used in vitro, all three ori sites are used at equal frequency. Interestingly, there two specific nic-cleavage sites observed in the vicinity of a-oriV and b-oriV regions. Those properties indicate a coupled regulatory system for replication and conjugation (Kolter and Helinski 1982; Shafferman, Kolter et al. 1982; Stalker, Kolter et al. 1982). The 51kb-length plasmid R46 is the prototype of IncN plasmid. It is originally isolated from Salmonella typhimurium, and carries antibiotics resistance genes to ampicillin, tetracycline, streptomycin, and sulfonamides; it also confers protection against UV induced damage (Waleh and Stocker 1979; Attfield and Pinney 1982). The specific relaxase encoded by R46 is TraI, which is able to catalyze DNA-strand transfer reactions with specific a nic sequence. R46 also encodes another accessory protein, TraJ, which helps unwind super coiled dsDNA. Together with TraK and TraH, those proteins form the relaxation complex (Zatyka, Jagura-Burdzy et al. 1994).

The forth self transmissible group is the IncI-like plasmids, including IncI, IncB, and IncK (Droge, Puhler et al. 1998). They have a unique pilus structure compared to the other three groups (Komano, Kim et al. 1994). In addition, the organization of Tra genes, their specificity and function are quite distinct from the other groups. Plasmid ColIb-P9 is the prototype of IncI plasmid. It is known to encode two morphologically distinct types of pilus: a thick rigid one and a thin flexible one (Howland and Wilkins 1988). The thick pilus supports conjugation on a semi-solid surface; while the thin pilus allows conjugation occur in liquid media. The variation of thin pilus in this plasmid group affects the conjugation efficiency of different enterobacteria in a liquid environment, possibly by changing the ability of the pilus to recognize the surface of different recipient cells. The core region of ColdIb plasmid also includes a DNA primase gene (sog) and exclusion determinant (exc) (Howland, Rees et al. 1989; Hama, Takizawa et al. 1990). The sog gene encodes two sequence-related polypeptides with the N-terminal domain having DNA primase activity and C-terminal domain facilitating transfer of single-stranded DNA following the initiation of transfer at oriT site.

6.2.2.2 Mobilizable plasmids

Conjugative plasmids are considered to be mobilizable if they are not self-transmissible, but can be transferred when a helper self-transmissible plasmid is also present in the donor cell. In other words, mobilizable plasmids carry the genetic information necessary for relaxosome formation, but don't encode any genes required for the establishment of direct cell-to-cell contact between a donor and a recipient cell. Therefore, the success of conjugative transfer depends on the co presence of a self-transmissible plasmid. Mobilizable plasmids usually carry an origin of transfer (oriT) and a mobilization region (mob) encoding specific relaxase components. Thanks to the development of DNA sequencing technology, the number of known mobilizable plasmids increased rapidly and sequence similarity has replaced the traditional incompatibility testing in classifying these plasmids. In 2004, Francia and Cruz classified mobilizable plasmid into four main super families based on the amino acid sequence similarity of their relaxases (Francia, Varsaki et al. 2004). Those plasmids family includes: MOB_Q family, ColE1-superfamily, pMV158-superfamily and CloDF13 family.

The MOB_Q family shares a common domain structure consisting of an N-terminal relaxase domain and a C-terminal primase domain. It is believed that the linkage between both domains promote the initiation of complementary strand synthesis in the recipient cell. The 9kb plasmid RSF1010 is the archetype of this family (Guerry, van Embden et al. 1974). It belongs to the incompatibility group Q, and encodes antibiotics resistance genes to streptomycin and sulfonamides. RSF1010 is known for its efficiency of mobilization with the helper plasmids from different incompatibility groups, such as IncP, IncI and IncX (Willetts and Crowther 1981). The RSF1010 enables its replicon is able to interact with the replication machineries of a variety of hosts. The replication of RSF1010 can proceed either unidirectionally or bidirectionally from a unique oriV with the help of three plasmid-specified proteins, RepA, RepB, and RepC, which function as a DNA helicase, primase, and initiator protein, respectively (Scholz, Haring et al. 1985; Scherzinger, Kruft et al. 1993). The RepA protein is hexamer of 30kDa. It has two enzymatic activities: a single-stranded DNA-dependent ATPase and also a DNA helicase. The RepB gene encodes two polypeptides of 36kDa and 78 kDa by using two alternative start codons and the same stop codon. The 36kDa RepB is a novel type of DNA primase by which it does not require ribonucleotide triphosphates for the priming reaction. The 31kDa RepC protein is a dimer, and binds specifically to the direct repeats in the oriV region.

The second mobilizable family is the ColE1-superfamily, which is best represented by the ColE1 plasmid (Scholz, Haring et al. 1985). The 6 kb ColE1 is a multicopy plasmid, and its replication requires only proteins from its host bacterium E. coli. It can replicate in E. coli even in the presence of protein synthesis inhibitors, such as chloramphenicol, which makes it perfect for the construction of bacterial cloning vectors (Engleberg, Cianciotto et al. 1988; Inoue and Uchida 1991; Lin-Chao and Cohen 1991). Because of these characteristics, ColE1 have been the subject of extensive study and used as vectors of gene therapy (Engleberg, Cianciotto et al. 1988). ColE1 can also be mobilized by members of many incompatibility groups, like IncP, IncF and IncW. The replication of ColE1 plasmid is exclusively unidirectional and relies only on host proteins, partially because the plasmid does not encode any enzyme for replication (Scholz, Haring et al. 1985). In ColE1 DNA, there is a 600bp region 500bp upstream of oriV, which is necessary for initiation of replication. The promoter is also located in this region, and is responsible for initiation of the replication primer RNA (called RNAII). The ColE1 mobilization region is composed of five genes (MbeA, MbeB, MbeC, MbeD and MbeE), of which the first four are essential for plasmid mobilization. The MbeE doesn't play any essential role during bacteria conjugation.

The third family, is the pMV158-superfamily, contains mobilization plasmids from both Gram-negative and Gram-positive bacteria. The 5.5kb-length plasmid pMV158 is the prototype of this family (Burdett 1980). It is non-conjugative, but can be mobilized by conjugative plasmids of the pIP501/pAMβ1 family as well as by helper plasmids from different incompatibility groups, such as IncP, IncF and IncW (del Solar, Diaz et al. 1987; Priebe and Lacks 1989). It is known that the relaxase of pMV158 is only encoded by the oriT region and MobM. The pMV158 encoded MobM, also called cladeA, cleaves pMV158 DNA at the cleaves the 5P-GpT-3P dinucleotide between co-ordinates 3591 and 3592 within the plasmid oriT site, then forms a covalent adduct with the target DNA via Tyr49 (Burdett 1980). The DNA region surrounding the oriT of pMV158 is conserved among a group of rolling circle replication (RCR) plasmids in Gram-positive bacteria, including an inverted repeat (IR) with 7 to 10-nucleotides long stem and a six-nucleotide long loop. The nic site is located in the loop of the inverted repeat. The cleavage reaction is dependent of the presence of divalent cations, but not the presence of additional proteins. Genetic experiments proved that disturbing or removal

the oriT region of MobM completely abolished the mobilization (Guzman and Espinosa 1997; Moscoso, Eritja et al. 1997).

The forth plasmid family is the relatively small sized ColDF13 family. The 9kb plasmid ColDF13 is the archetype of this family. It originates from Enterobacter cloacae, but is also stably maintained in E. coli (Nijkamp, de Lang et al. 1986). Like all plasmids in the pMV158 super family, ColDF13 can be mobilized by helper plasmids from different incompatibility groups, such as IncP, IncF and IncW (Cabezon, Sastre et al. 1997). Unlike other mobilization plasmids, mobilization of CloDF13 does not require the presence of coupling protein encoded by the helper conjugative plasmid but only the Mpf gene products (Nunez and De La Cruz 2001). CloDF13 encodes two proteins: the 61kDa MobB and the 24kDa MobC. The MobC protein is a relaxase; while the MobB protein is made to enhance the relaxase activity of MobC. Interestingly, the MobB and MobC protein of cloDF13 are quite different from other characterized plasmid Mob proteins in both sequence and biochemical properties (van Putten, Jochems et al. 1987; Francia and Clewell 2002). For example, Cruz et al found that both the 5' and 3' ends of the *nic* site are not blocked during nicking process. In other words, the relaxase of cloDF13 process the cleavage reaction quite differently from the other Mob plasmids. In all these respects, CloDF13 seemed to be rather atypical mobilizable plasmid, which deserved to be classified as a separate group (Nunez and De La Cruz 2001).

6.3 R1162

Work in this dissertation focuses on the mobilizable plasmid R1162; it belongs to the MOB_Q family of the mobilization plasmid. It is almost identical to plasmid RSF1010, except that it is isolated from Pseudomonas aeruginosa rather than Escherichia coli (Meyer, Hinds et al. 1982; Meyer, Lin et al. 1985). The R1162 plasmid is the archetype for the incompatibility group Q (IncQ), which are characterized by their broad host range and relatively small size (Rawlings and Tietze 2001). The plasmid encodes resistance to the antibiotics streptomycin (strA and strB) and sulfonamide (sulII). It also contains an origin of transfer (oriT), and encodes 3 proteins: mobA, B and C, which are all required for plasmid mobilization, see figure 6.2.

The oriT DNA sequence of R1162 consists of a 35 base oligonucleotide, which was mapped by Becker and Meyer and shown in Figure 6.2 (Becker and Meyer 2002; Becker and Meyer 2003). Sequence alignments of oriT with some of the IncQ family members reveals that

133
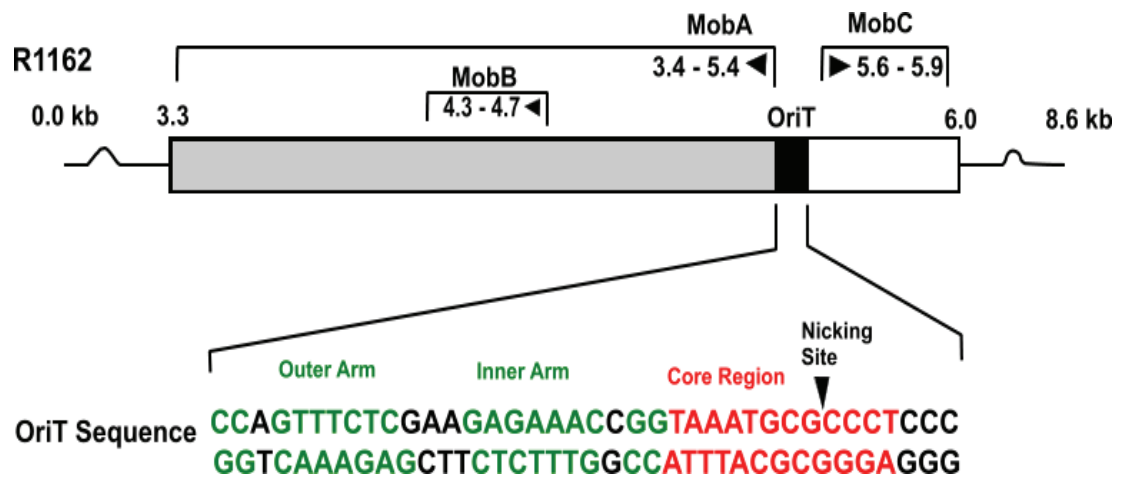
Figure 6.2: The organization of mob genes of R1162 (in kilobases) and the nucleotide sequence of oriT. The vertical arrow indicates the nicking site.

a 12 base core region is highly conserved. The 23-base oligonucleotide 5' to the core region is an imperfect inverted repeated (IR), which is commonly found in the R1162 family of plasmids. However, the sequence and size of the repeats of the R1162 family varies considerably (Becker and Meyer 2000; Becker and Meyer 2002). For R1162, the presence of this inverted repeat significantly enhances the binding by MobA, although the core and inner arm of IR is sufficient for initiation. Interestingly, Mob A can bind mutated variants of oriT to MobA can be tolerated by complementary mutation at IR region, which suggested that the sequence at IR serves as more of a structural role (Parker, Zhang et al. 2002; Zhang, Zhang et al. 2003; Parker and Meyer 2007). The nic site is 8bp from the inverted repeat as shown.

The Mob proteins assemble at oriT to form a complex called the relaxosome (Meyer, Hinds et al. 1982). The most important protein of the relaxosome is the 78 kDa (709 residues) MobA protein which has two functional domains. The C-terminal domain encodes a 43 kDa protein, called primase, which can also be expressed separately. Both forms of primase have been found in host cells. The function of the primase is to lay down primers within oriV, which can be further extended by host cell DNA polymerases to replicate the plasmid.

The N-terminal domain is a relaxase consisting of about 250 amino acids. It cleaves one of the DNA strands within oriT, and forms a covalent adduct with the 5' end. Tyr25 of the N-terminal domain has been identified to be the critical residue (Becker, 2002). It carries out a nucleophilic attack on the phosphodiester bond at the nick site, and forms a phosphodiester bond with the 5' end. Once the covalent adduct is formed, the complex is recognized by a T4S system, and the cleaved strand is unwound from its complement strand, then transferred in the 5' to 3' direction to the host cell. When the transfer is completed, the MobA protein carries out a second transesterification reaction which rejoins of the two ends and releases the protein. Once the circular plasmid DNA is formed in the host cell, a complementary strand is synthesized in the new cell by the host polymerase. MobA complex is then released (Becker, 2002).

The MobB and MobC are accessory proteins of relaxosome complex. MobB, a 19kD protein, is found to stabilize the assembly of MobA and MobC at oriT both in vivo and in vitro (Perwez, 1996; Zhang, 1997). It was reported that the frequency of mobilization of R1162 decreases two to three orders of magnitude in the absence of MobB (Zhang, 2003), providing strong evidence that the function of MobB is to stabilize the relaxosome complex. Parker and

Meyer (2007) proposed that MobB inserts into the cell membrane and thus stabilizes the association between the relaxase and the type IV transfer apparatus. The function of 10kD MobC protein is to enhance strand separation at the site of cleavage, thereby facilitate the strand cleavage by MobA. It is reported that multiple copies of MobB and MobC join with MobA to assemble on oriT (Zhang, 2003). While, it is still unknown that how relaxosome is assembled. Interestingly, not all mobilization plasmids have homologs of MobB or MobC. For example the 371 residue pSC101 from *E.coli* is a homolog of the relaxase domain of MobA form R1162, but the plasmid does not code for any homologs of MobB or c (Meyer, 2000). Obviously, the relaxase is the most important key to understand the mobilization process.

The X-ray structure of N-terminal 186 residue MobA (minMobA) was solved in our laboratory in 2007 (Monzingo, Ozburn et al. 2007). It consists of five antiparallel beta-sheets connecting by four helices lying on both sides of the sheet. The critical residue tyr25 lies on the first helix on the front side of the molecule, close to a metal ion. As also shown in figure 6.3, this metal ion is bound on the front side of the beta sheet, and chelated by His112, His120 and His122. Interestingly, minMobA share a common fold with two other relaxase enzymes, TraI and TrwC (Grandoso, Avila et al. 2000; Datta, Larkin et al. 2003). TraI is from the self-transmissible F factor plasmid, while TrwC is encoded by plasmid R338, which belongs to IncW plasmid family (Llosa, Grandoso et al. 1995; Llosa, Grandoso et al. 1996; Matson, Sampson et al. 2001; Street, Harley et al. 2003). Superposition of minMobA to the equivalent 106 residue domain of TraI gives an RMS distance of 2.3 Å between alpha-carbon atoms. Similarly, superimpose of minMobA to the equivalent 293 residue domain of TrwC gives an RMS distance of 2.4 Å. The relaxase domains of all three proteins share same overall structural features and have same biological functions. All three proteins have both a catalytic tyrosine residue and an active site metal ion bound by three histidine residues (Grandoso, Avila et al. 2000; Datta, Larkin et al. 2003). Although there are only 12% and 10% sequence identities of minMobA with TraI and TrwC respectively, the conserved structure and function imply that all three proteins may share a common ancestor.

A hypothetical model of minMobA bound to 33mer single stranded DNA in the stem-loop conformation has been made in our laboratory, based on the DNA binding observed for TrwC, see figure 6.4. The 33mer DNA in this model is composed of a 23 base double stranded stem loop followed by a 10 residue core region of oriT; the TrwC sequence was modified in
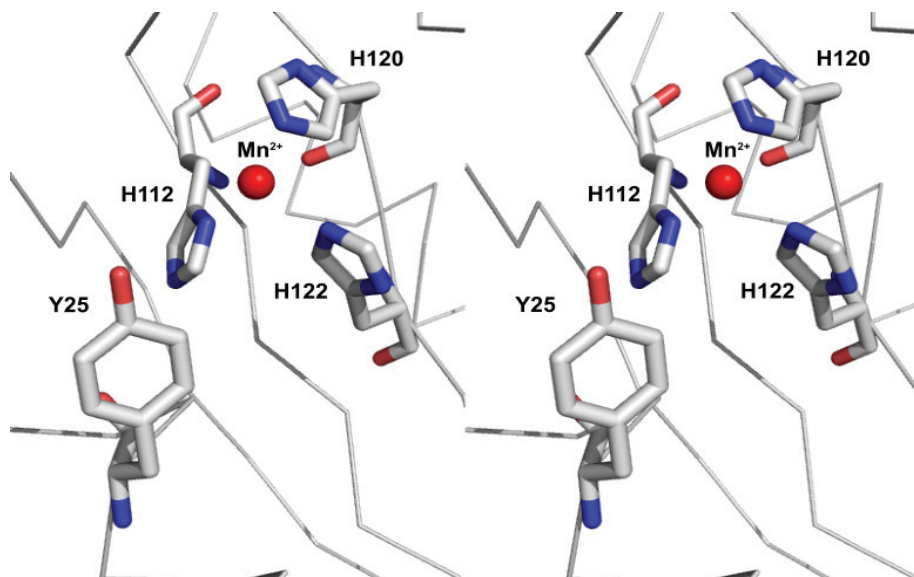
Figure 6.3: Active site of minMobA. Y25, H112, H120 and H122 are highlighted in stick mode; Mn^{2+} is labeled in red.
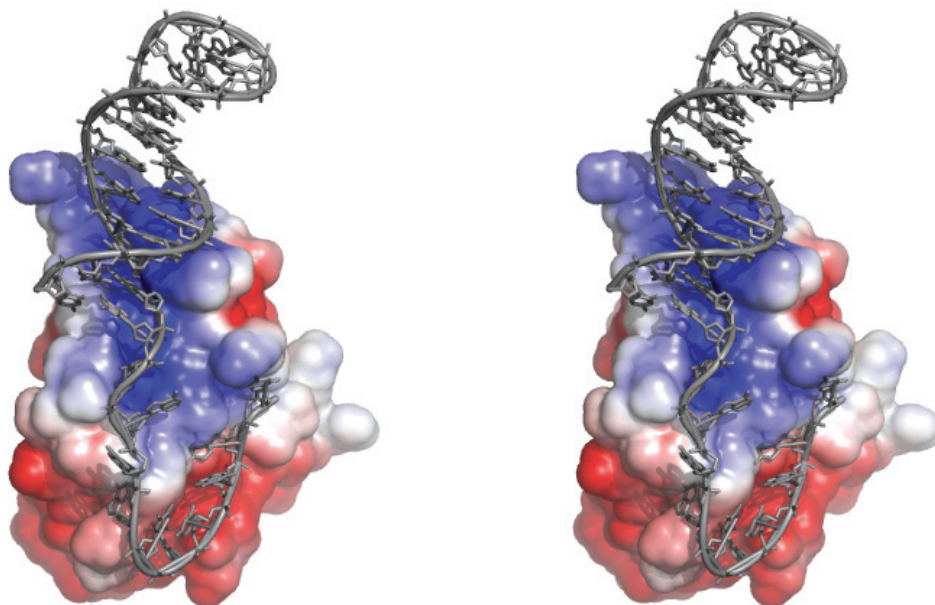
Figure 6.4: Stereo image of the surface electrostatic potential of minMobA-33mer model complex. Electronegative charges are shown in red, and positive charges are shown in blue. The bound DNA is represented in a cartoon mode.

this region to reflect that of R1162. As expected, the negatively charged phosphate backbone of the 33mer is stabilized by a patch of positively charged amino acids on the protein surface. The binding characteristics of the interaction between MobA and the 33mer are in agreement with those deduced from experimental studies. For example, the model shows that only the first nine nucleotides at the base of the stem interact with protein. This is consistent with previous studies that only seven bases of the inner arm of oriT make specific interactions with MobA. With this model, it will be helpful for us to design mutagenesis experiments to assign roles to residues which are critical for substrate-binding or catalysis.

6.4 Project goals

6.4.1 Identify a potential active site base

The first goal of this project was to use our structural knowledge of minMobA to guide mutagenesis studies aimed at identifying key amino acids for the transesterification mechanism. We looked for the chemically interesting residues near Tyr25, which carries out the nucleophilic attack on the phosphate of G31 of OriT. It is possible that a carboxylate may function as base to deprotonate the Tyr25 hydroxyl and thereby increase its nucleophilicity.

6.4.2 Analyze the role of minMobA active site metal

MobA is known to have a requirement for divalent cations, initially thought to be Mg^{2+}, although Mn^{2+}, Ca^{2+}, or Ba^{2+} could substitute at lesser efficiencies (Scherzinger, Lurz et al. 1992). However, the physiologically relevant metal for relaxase is still unknown. It is uncertain if the metal is structural or catalytic in nature. To further explore the role of metal ion in mobA's cleavage reaction, we will assess the ability of minMobA to cleave single-strand oligo DNA substrate, to bind ssDNA, and to change the circular dichroism spectrum in the presence of Ni^{2+}, Mg^{2+}, Zn^{2+}, Mn^{2+} and Ca^{2+}.

6.4.3 Express and crystallize related MobA proteins

The third goal of this project was to grow useful crystals of a MobA oligonucleotide complex, and to crystallize other, related, MobA proteins. This effort involved several subprojects. (1): To visualize protein DNA interactions, different lengths of oriT DNAs were used to form complexes with the minMobA Y25F mutant enzyme. (2): The minMobA protein is able to carry out the initial DNA cleavage, but can not interact with T4SS in gene transfer. A longer version of MobA protein named MobA*, which is active in both nicking and transfer, was screened for crystallization condition. (3): Efforts were made to crystallize the 371 residue

138

pSC101 enzyme that is homologous to R1162. The pSC101 protein is fully functional in nicking and transfer, but naturally lacks the primase domain of MobA.

Chapter 7: Materials and methods

7.1 Construction of various MobA plasmids

The minMobA-pTYB2, MobA*-pTYB2 and pSC101-pTYB2 plasmids were kindly provided by Dr. Richard Meyer. The 341 residue MobA* was PCR amplified with primers GACGACGACAAGATGGCGATTTATCACCTTACGGCG and GAGGAGAAGCCCGGTT TAACCTCCCGCAACTC. The product was cloned into pET-46 Ek/LIC (Novagen). The MobA* gene was also PCR amplified with primers CATGCCATGGATGGCGATTTATCACC TTACGGC and CCGCTCGAGACCTCCCGCAACTC. The product was digested with NdeI and XhoI restriction endonucleases, and cloned into pET28b plasmid DNA (Novagen) cleaved with NdeI and XhoI. The 371 residue pSC101 was PCR amplified with primers GGAATTCCA TATGGCATCCTATCATC and CCGCTCGAGTCGTGAGAATGACC. The product was cloned into pET28b plasmid DNA cleaved with NdeI and XhoI. The first 193 amino acid fragment of pSC101 was purified by first amplifying the DNA with primers GGAATTCCATATGGCATCCTATCATC and TCCCCCCGGGGGTGCTGATACCGTTC. The product was digested with NdeI and XmaI restriction endonucleases, and cloned into pTYB2 plasmid DNA (New England Biolabs) cleaved with NdeI and XmaI. The integrity of all cloned DNA(s) were confirmed by DNA sequencing.

7.2 Site directed mutagenesis

Various mutations were introduced by site-directed methods, according to the Stratagene protocol (Stratagene). Around 50 ng of plasmid and 150 ng of each primer were combined with reaction buffer (20 mM Tris–HCl (pH 7.5), 8 mM $MgCl_2$, 7.5 mM DTT, 50 µg/ml of bovine serum albumin (BSA)), 150 µM dNTP mix, 1 units of KOD Hot Start DNA Polymerase (Novagen), and deionized water to final volume of 50 µl. Primer sequences of minMobA mutants (Y25F, E74A, E74Q, E76A, E74A-E76A, E38A, H112A, H120A and H112A) and pSC101 mutants (Y22F, E73A and E75A) are shown in table 7.1. Reaction mixtures were further treated with 10 units of DpnI (New England Biolabs) at 37 °C for 2hrs. Then 1 µl of treated reaction mixture was transformed to E. coli DH5α competent cell. The presence of the expected mutations was confirmed by DNA sequencing.

minMobA Mutants	Primer sequence
Y25F-5'	GGGCCAAGGCCGACTTCATCCAGCGCGAAGGC
Y25F-3'	GCCTTCGCGCTGGATGAAGTCGGCCTTGGCCC
E74A-5'	GCGGCTGTTCAAGGCGGTCGAATTTGCCCTG
E74A-3'	CAGGGCAAATTCGACCGCCTTGAACAGCCGC
E74Q-5'	GCGGCTGTTCAAGCAGGTCGAATTTGCCCTG
E74Q-3	CAGGGCAAATTCGACCTGCTTGAACAGCCGC
E76A-5'	CTGTTCAAGGAGGTCGCGTTTGCCCTGCCGG
E76A-3'	CCGGCAGGGCAAACGCGACCTCCTTGAACAG
E74A-E76A-5'	GCGGCTGTTCAAGGCGGTCGCATTTGCCCTG
E74A-E76A-3'	CAGGGCAAATGCGACCGCCTTGAACAGCCGC
E38A-5'	CCCGCGACATGGATGCGGTCTTGCACGCCG
E38A-3'	CGGCGTGCAAGACCGCATCCATGTCGCGGG
H112A-5'	GTATACGCTGGCTATCGCGGCCGGTGGCGGCGAG
H112A-3'	CTCGCCGCCACCGGCCGCGATAGCCAGCGTATAC
H120A-5'	GCGGCGAGAACCCGGCGTGCCACCTGATGATC
H120A-3'	GATCATCAGGTGGCACGCCGGGTTCTCGCCGC
H122A-5'	GAGAACCCGCACTGCGCGCTGATGATCTCCGAG
H122A-3'	CTCGGAGATCATCAGCGCGCAGTGCGGGTTCTC
pSC101 mutants	**Primer sequence**
Y22F-5'	CTCCTCATGCCGATTTTATTGCGCGTGAGG
Y22F-3'	CCTCACGCGCAATAAAATCGGCATGAGGAG
E73A-5'	CTGTACCTACCGTGCGATCGAGATTGCCCTG
E73A-3'	CAGGGCAATCTCGATCGCACGGTAGGTACAG
E75A-5'	CTACCGTGAAATCGCGATTGCCCTGCCGCG
E75A-3'	CGCGGCAGGGCAATCGCGATTTCACGGTAG

Table 7.1 Primer sequences of minMobA and pSC101 mutants

141

7.3 Protein purification

The protein purification procedure is the same for all genes cloned into and expressed from the pTYB2 vector. *Escherichia coli* K-12 strain ER2566 (New England Biolabs) containing the plasmid was grown overnight in broth medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) containing ampicillin (100 μg/ml) at 37 °C. The cells were then diluted 1:100 in 4×0.5 l medium and grown to a cell density of approximately 4×10^8 per ml. IPTG was added to final 0.3 mM and the cells then grown overnight at 25 °C. Protein was purified according to a procedure developed by New England Biolabs. Briefly, cells were collected by centrifugation, resuspended in 50 ml column buffer (20 mM Tris (pH 8.0), 500 mM NaCl, 1 mM EDTA) and disrupted in a French pressure cell. Cellular debris was pelleted by centrifugation at 5,000g for 60 min at 4 °C. The supernatant was applied to a column containing chitin beads pre- equilibrated with column buffer. Intein cleavage was initiated by rapidly flushing the column with 30 ml of CB containing 40 mM DTT, and left at 4 °C for 48 hours. Protein was then eluted from the column by washing the beads with 15ml column buffer. The protein was concentrated by using an Amicon bioseparator fitted with a 25 mm YM-3 membrane and applying 60–70 lb/in^2 N$_2$ for 7–10 h at 4 °C.

The protein purification procedure is the same for all genes cloned into and expressed from the pET28b vector. *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) containing the plasmid was grown overnight in broth medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl) containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) at 37 °C. The cells were then diluted 1:100 in 2 l medium and grown to a cell density of approximately 4×10^8 per ml. IPTG was added to 0.5mM and the cells then grown overnight at 25 °C. Protein was purified essentially according to a procedure developed by Novagen. Briefly, cells were collected by centrifugation, resuspended in 25 ml column buffer (CB: 50 mM Hepes, pH 8.0, 300 mM NaCl, 20 mM Imidazole) and disrupted in a French pressure cell. Cellular debris was pelleted by centrifugation at 5000 xg for 60 min at 4 °C. The supernatant was applied to a column containing Ni-NTA His binding beads (2 ml bed volume, equilibrated with CB), and washed with 100 ml of CB, then eluted with 10 ml elution buffer (EB: 50 mM Hepes, pH 8.0, 300 mM NaCl, 250 mM Imidazole). The protein was concentrated to 1–2 ml using an Amicon bioseparator fitted with a 45 mm YM-3 membrane and applying 60–70 lb/in^2 N$_2$ at 4 °C.

7.4 Fluorescence polarization assay

All fluorescence measurements were made using an Envision (Perkin Elmer Inc) spectrofluorometer. Six different lengths of oligonucleotides, with sequences listed in table 7.2, were (3′-fluorescein) labeled. Samples were excited at 490 nm, and emitted light was collected through an orange glass filter (OG 515, Schott). The binding affinity for the minMobA (Y25F) and oligo DNA complex were determined in the presence and absence of 1mM metallic cations by measurement of the steady-state anisotropy of fluorescence as a function of added protein. The binding buffer contains 10 mM Tris–HCl (pH 7.5), 20 mM NaCl. The concentration of protein was plotted against anisotropy of fluorescence and fit to a hyperbola equation to get the dissociation constants. The nonlinear regression analysis was performed in the program GraFit 5.0 (Erithacus Software).

Oligonucleotide	Sequences
D35	CCAGTTTCTCGAAGAGAAACCGGTAAGTGCGCCCT
D31	CCAGTTTCTCGAAGAGAAACCGGTAAGTGCG
D24	CCAGTTTCTCGAAGAGAAACCGGT
D19	CCGGTAAGTGCGCCCT
D12	TAAGTGCGCCCT
D12m	TAAATGCGCCCT

Table 7.2: Sequence of oligonucleotides used in fluorescence polarization assay.

7.5 DNA cleavage assay

The DNA cleavage assay was performed as described below. A 35mer oligonucleotide (Integrated DNA Technologies) with sequence CCAGTTTCTGAAGAGAAACCGGTAAATG CGCCCT was (3′-^{33}P)-labeled with terminal transferase, according to the manufacturer's instructions (New England Biolabs). The assay reaction mixtures contained 0.03 μM labeled 35mer oligonucleotide, 0.1 μM minMobA, 40 mM Tris–HCl (pH 8.0), 20 mM NaCl, and in the presence and absence of 1mM metallic cations. For minMobA mutants, the reaction mixtures contained additional 50 mM $MgCl_2$. The reactions were terminated by addition of EDTA to 40 mM after 0, 1, 2, 4, 6, 9, 11, 15, 20, 25, 28, 36, 45, 55, and 60 min. Reaction products were separated by SDS–12% PAGE and imaged on a Molecular Imager FX system (BioRad Laboratories, Inc., Hercules, CA). The concentration of protein-DNA covalent adduct was

plotted against time and fit to a single exponential equation: [protein-DNA] =A*(1-exp(−k_{obs}*t))+C. The nonlinear regression analysis was performed in the program GraFit 5.0 (Erithacus Software). The protocol for DNA cleavage assay by pSC101 is almost the same as that of minMobA, except that the substrate is 37mer oligonucleotide with sequence TTTCTGAACGA AGTGAAGAAACGTCTAAGTGCGCCC T, and the reactions were terminated after 0, 0.5, 1, 2, 4, 6, 9, 11, 15, and 20 min.

7.6 Isothermal titration calorimetry (ITC)

A MicroCal VP-ITC calorimeter was used to measure the binding affinity of minMobA to metallic cation. Protein was dialyzed against 10 mM Na-HEPES (pH 7.5), 25 mM NaCl, with 10 g/liter of Chelex beads in the dialysis buffer. Protein, metal and buffer solutions were degassed before loading into the calorimeter. For the measurement of Mn^{2+} binding, 4mM $MnCl_2$ was injected into 158 μM minMobA, with a stirring speed at 200 rpm. Data analysis was carried out by fitting to both single site and multiple binding sites models as defined in the Origin software package.

7.7 Inductively coupled plasma mass spectrometry (ICP-MS)

ICP-MS studies were carried out on an Agilent 7500ce ICP mass spectrometer with an Octopole Reaction System for interference reduction, and an electron multiplier detector that operates simultaneously in pulse counting and analog modes. All the samples were prepared in a class 100 clean laboratory with all reagents being ultra pure. The supernatant of cell lysate was dissolved in high concentration of HNO_3 and diluted to final 1% HNO_3 before analysis. Concentrations of all elements were calculated from the calibration curve after proper correction for control blank, matrix and drift effects.

7.8 Circular dichroism measurements

Circular dichroism spectra were recorded using a Jasco 715 spectropolarimeter, equipped with a thermostated cell holder and a NesLab-111 circulating water bath. CD spectra were recorded in cells with an optical path length of 0.1cm. Experiments were performed in 10 mM Tris–HCl (pH 7.5), 20 mM KCl, in the presence or in the absence of different concentrations of metal ions. Three scans were repeated for each experimental condition.

7.9 Thermal denaturation experiments

Thermal denaturation profiles were obtained by recording the temperature dependence of the ellipticity at 220 nm in the range 6–80 °C. The temperature was continuously changed at a

rate of 0.5 °C/min. Experiments were performed in 10 mM Tris–HCl (pH 7.5), 20 mM KCl, in the presence or in the absence of different concentrations of metal ions. Tm was determined by locating the maxima/minima of the first derivative of the curve describing the melting profile (CD versus T).

7.10 Screening for crystallization condition

A total of 698 conditions were manually screened for crystallization condition using the sitting drop method. Among those conditions, 458 come from the commercially available crystallization screening kits (Hampton Research), including: Screens I and II, Index, PEG/Ion I and II, Natrix, Grid screen PEG 6000/AS/NaCl/MPD, and Quick screen phosphate. The rest 240 conditions came from 10 in-house Grid-screening kits. MobA* and pSC101 were also sent to the Hauptman-Woodward Institute for a high-throughput screening with 1,536 conditions.

Chapter 8: Results

8.1 Nicking activity of wild type and mutant minMobA(s)

To get a quantitative measurement of the transesterification activity of MobA protein, a nicking assay was developed by incubating minMobA with 3' radiolabeled oligonucleotide. The reactions were terminated by addition of EDTA at various time points over one hour period, and then the reaction products were separated by SDS-PAGE. As shown in figure 8.1 panel a, the top bands corresponded to the protein-oligonucleotide adduct, and the bottom bands represented the unreacted oligonucleotide. The fractions of radiolabeled covalent adduct formed were calculated, and plotted against reaction time. The resulting curve was fitted into a single exponential equation. The parameters provided the observed reaction rate and the fraction of covalent adduct. The observed rate constants of nicking reactions are summarized in table 8.1.

As expected, Tyr25 is critical for catalysis; no covalent adduct was observed once it was mutated to alanine (Figure 8.1, panel b). Surprisingly, the conversion of Glu74 to Ala or Gln decreases the covalent adduct formation in one hour to 50% and 70% respectively, of wild-type activity (Figure 8.1, panel c and d). The conversion of the nearby Glu76 to Ala decreases the covalent adducts formation by 45% (Figure 8.1, panel e). Even the double mutant E74A-E76A reduces activity only tenfold (Figure 8.1, panel f). This suggests E74 plays only the most minor role in catalysis. The same phenomena were observed for pSC101, in which the conversion of E73 (corresponding in position to E74 of minMobA) to Ala, or conversion of E75 (corresponding in position to E76 of minMobA) to Ala only partially reduced enzyme activity (40% and 42% of wild-type activity respectively), see figure 8.2 panel c and d. This indicates the phenomenon we observed in minMobA is not an artifact of protein truncation.

8.2 Effects of metallic cations on nicking activity of minMobA

The metal ions were replaced in minMobA by an initial exhaustive dialysis against EDTA followed by dialysis against the divalent cation of choice. As expected, no covalent DNA adduct formed in the absence of any divalent cations; that is, nicking activity has an absolute requirement for cations (Figure 8.3, panel a).

(a). Wild type minMobA

(b). Y25F

(c). E74A

(d). E74Q

(e). E76A

(f). E74A-E76A
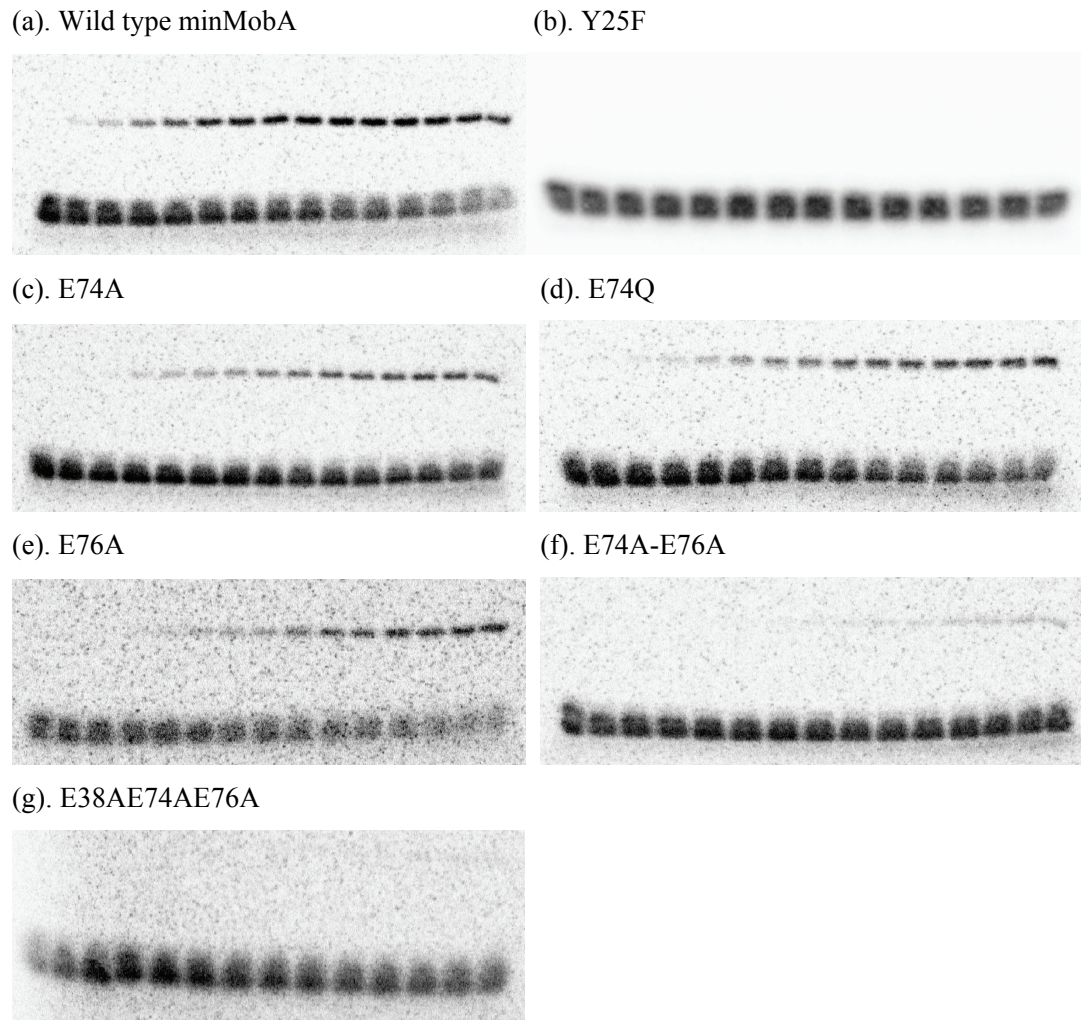
(g). E38AE74AE76A

Figure 8.1: In vitro cleavage of 35mer oligonucleotide by minMobA (a)wild type and mutants: (b) Y25F, (c) E74A, (d) E74Q, (e) E76A, (f) E74AE76A and (g) E38AE74AE76A. In each gel, each lane contains the separated components from the reaction stopped after a time ranging from 0min (left) to 60 min (right). The top bands corresponds to the protein-oligonucleotide adduct, and the bottom bands represent the unreacted oligonucleotide.

minMobA	K_{obs} (uM/min)
Wild type	0.1036
Y25F	0
E74A	0.0431
E74Q	0.0390
E76A	0.0294
E74AE76A	0.0124
E38AE74AE76A	0

Table 8.1: Observed rate constants of nicking reaction of minMobA wild type and mutant proteins.

(a). pSC101 wild type

(b). Y22F
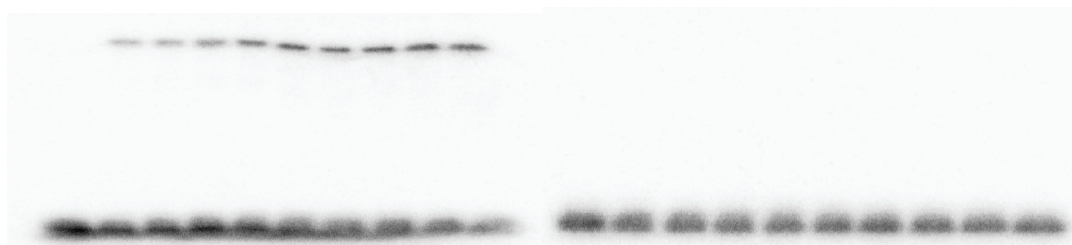
(c). E73A

(d). E75A
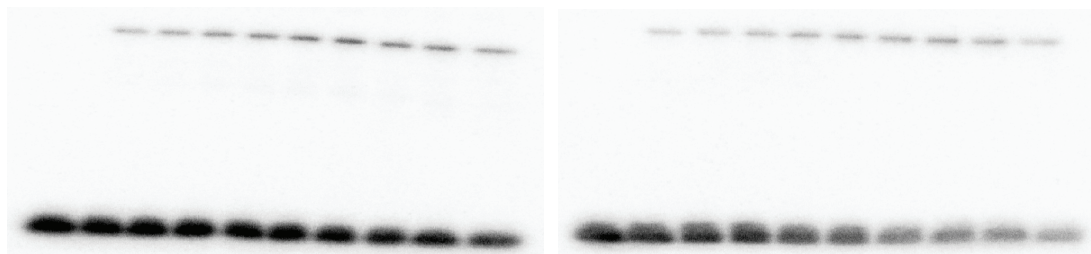
Figure 8.2: In vitro cleavage of 37mer oligonucleotide by pSC101 (a) wild type and mutants: (b) Y22F, (c) E73A and (d) E75A. In each gel, each lane contains the separated components from the reaction stopped after a time ranging from 0min (left) to 36 min (right). The top bands corresponds to the protein-oligonucleotide adduct, and the bottom bands represent the unreacted oligonucleotide.

pSC101	K_{obs} (uM/min)
Wild type	0.3945
Y22F	0
E73A	0.1562
E75A	0.1696

Table 8.2: Observed rate constants of nicking reaction of pSC101wild type and mutant proteins

(a). minMobA (H112A)
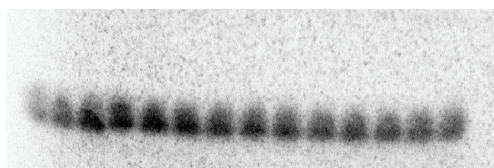
(b). minMobA (H120A)
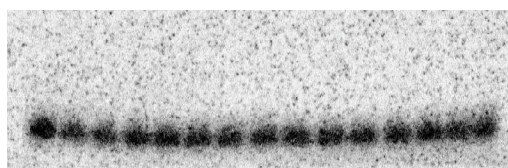
(c). minMobA (H122A)
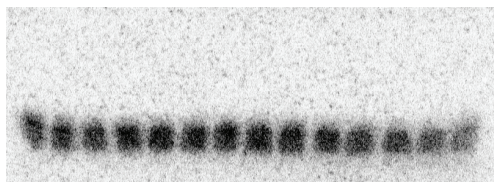
(d). Mn^{2+}
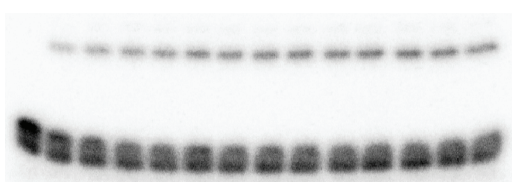
(e). Ni^{2+}
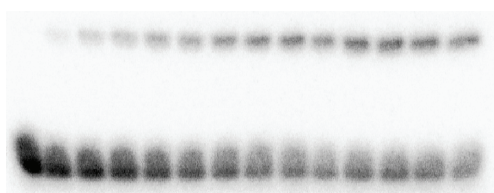
(f). Mg^{2+}

(g). Ca^{2+}
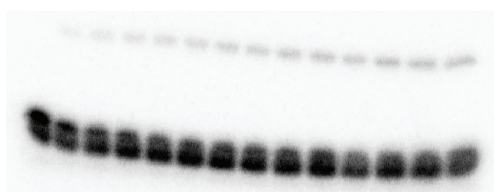
(h). Zn^{2+}

(i). No metal
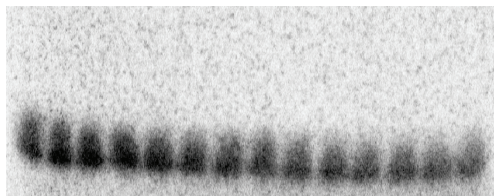
Figure 8.3: In vitro cleavage of 35mer oligonucleotide by minMobA mutant (a) H112A, (b) H120A, (c) H122A in the presence of 50mM Mg^{2+}, and by minMobA wild type in the presence of different metallic cations: (d) 1mM Mn^{2+}, (e) 1mM Ni^{2+}, (f) 1mM Mg^{2+}, (g) 1mM Ca^{2+}, (h) 1mM Zn^{2+}, (i) 20mM EDTA; and In each gel, each lane contains the separated components from the reaction stopped after a time ranging from 0min (left) to 60 min (right). The top bands corresponds to the protein-oligonucleotide adduct, and the bottom bands represent the unreacted oligonucleotide.

Metallic Cation	K_{obs} (uM/min)
Mn^{2+}	0.2272
Ni^{2+}	0.0723
Mg^{2+}	0.0628
Ca^{2+}	0.0519
Zn^{2+}	0
No metal	0

Table 8.3: Observed rate constants of nicking reaction in the presence of different metallic cations.

As shown in Figure 8.3, Mg^{2+}, Mn^{2+}, Ni^{2+} and Ca^{2+} are all able to restore the enzyme activity. At 1 mM concentration, Mn^{2+} gives the highest reaction rate, while Ni^{2+} gives the most complete reaction (Table 8.3). Surprisingly, 1 mM Zn^{2+} can not restore enzyme activity, as no detectable covalent adduct is formed. The crystal structure of minMobA shows that the metal binds with three Histidine residues (H112, H120 and H122) and a water molecule in a tetrahedral conformation. The mutation of any of these three Histidine residues to Alanine completely abolishes the enzyme activity, in spite of the presence of high concentration of Mg^{2+} (Figure 8.3, panel G-I). This suggests that the loss of any, or all, of the Histidines precludes metal binding at this site.

8.3 Effects of metallic cations on ssDNA binding activity of minMobA

To measure the equilibrium binding strength of ssDNA to MobA independently of covalent adduct formation, we mutated Tyr 25 to Phe (Y25F). A fluorescence polarization assay was used to analyze the binding affinity of the 35mer single-strand DNA to mutant minMobA (Y25F), in the presence of 1mM Mg^{2+}, Zn^{2+}, Mn^{2+}, Ca^{2+} and 0.2mM Ni^{2+} respectively. As shown in figure 8.4, the binding of oligonucleotide to minMobA does not require added divalent cations, but the presence of any of the divalent cation improved binding affinity by 2 to 3 fold. The oligonucleotide D_{35} binds to minMobA(Y25F) with $K_d{\sim}20$ nM, in the presence of Mg^{2+}, Mn^{2+}, Ca^{2+}, Ca^{2+} or Zn^{2+} (Table 8.4). We used 0.2mM Ni^{2+} in our experiments because Ni^{2+} at concentration greater than 0.2mM causes significant fluorescence quenching in this system.

8.4 minMobA Mn^{2+} binding affinity by ITC

Affinities of wild-type and variant minMobA proteins for Mn^{2+} were measured by isothermal titration calorimetry (ITC). As shown in figure 8.5, the stoichiometry of Mn^{2+} to minMobA appears to be 3:1. That is consistent with what we found in X-ray structure, in which three metal ions were bound to minMobA; one is chelated by three histidine residues in the active site and the other two are chelated by His46 and His98 respectively. The integrated curve was fitted to an independent multiple sites binding model, and gave a K_d of 0.2 μM for the active site Mn^{2+}, and 40 μM and 45 μM for the Mn^{2+} ions that bind outside the active site. Although the triple mutant E38A-E74A-E76A abolished the nicking activity completely (see Figure 8.1, panel g), it showed the same binding affinity and stoichiometry to Mn^{2+} as the wild type protein.

151
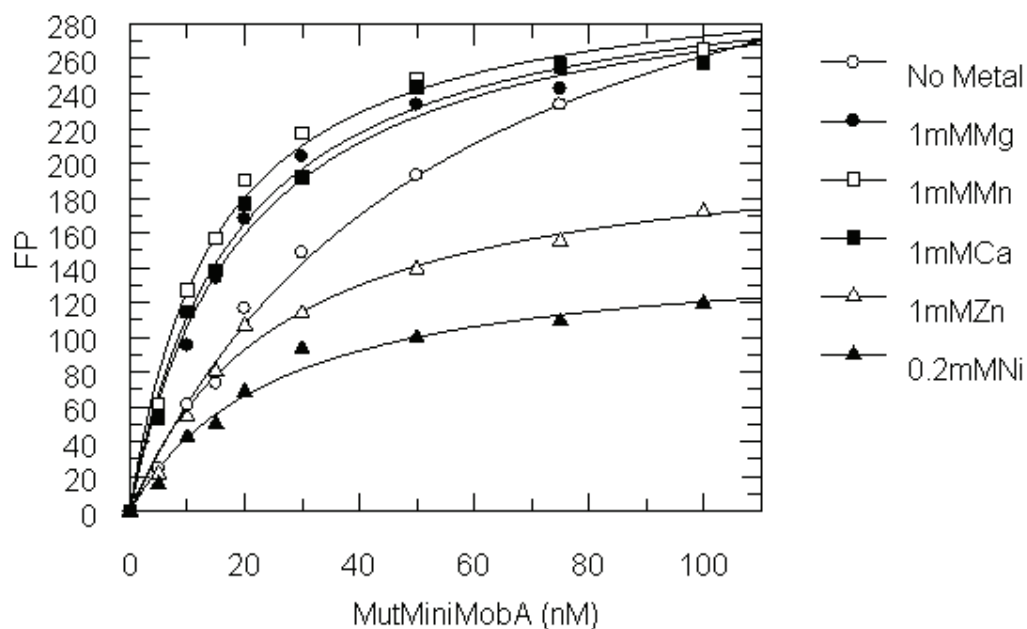
Figure 8.4: Fluorescence anisotropy assay of binding of minMobA (Y25F) to 3'-Fluorescein labeled oligonucleotide in the presence of different metallic cations.

Metallic Cation	K_d (nM)
Mn^{2+}	13
Ni^{2+}	24
Mg^{2+}	14
Ca^{2+}	15
Zn^{2+}	27
No metal	57

Table 8.4: Dissociation constants of minMobA (Y25F) to 3'-Fluorescein labeled oligonucleotide in the presence of different metallic cations.

(a).

(b).

153
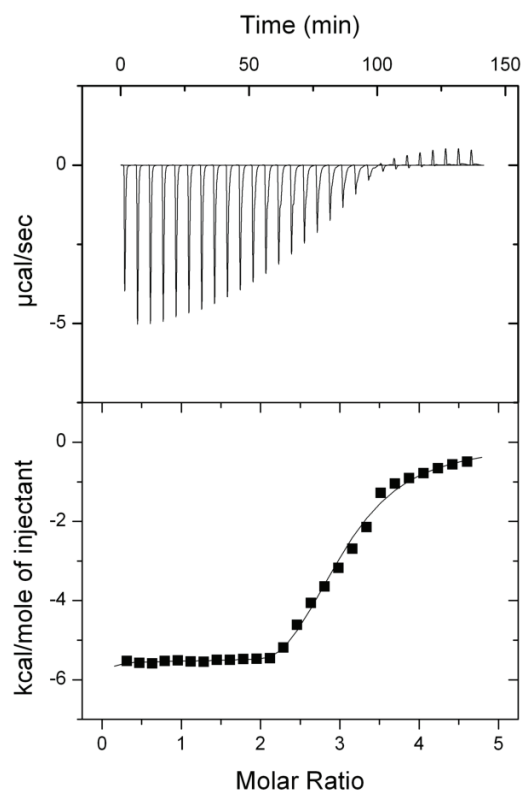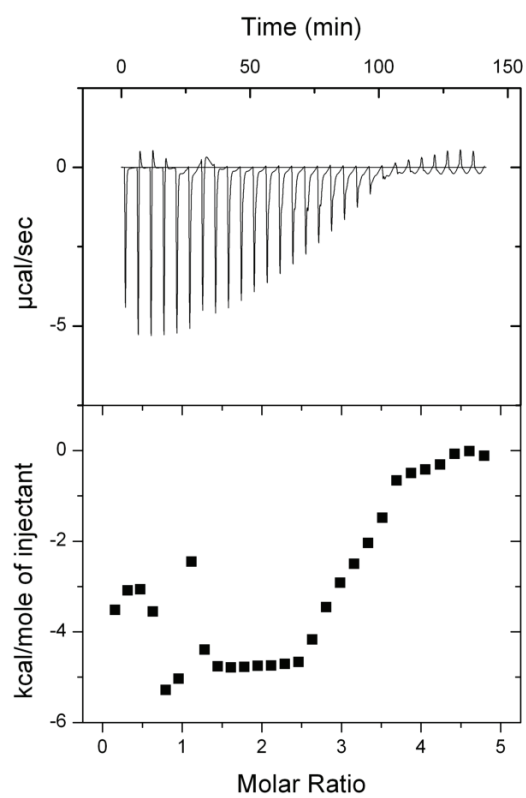
(c)
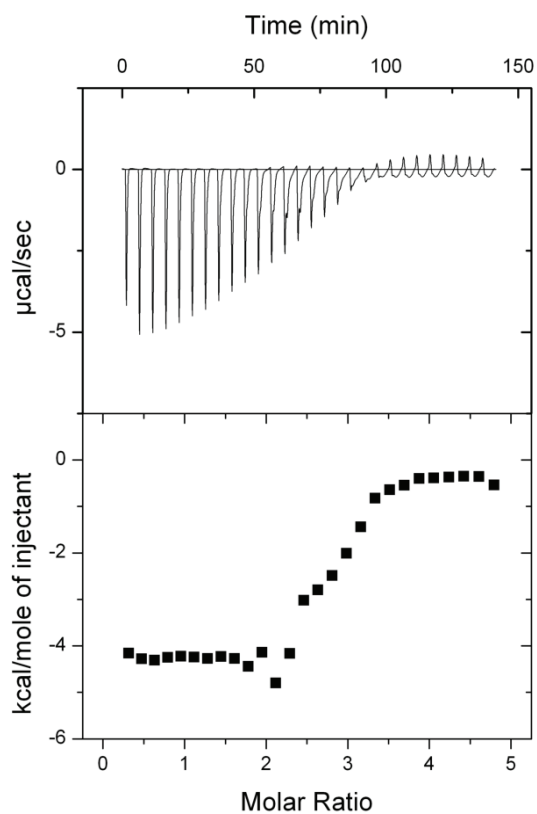
Figure 8.5: The representative data for titrations of 4mM MnCl$_2$ into 180μM (a) wild type minMobA, (b) E74A mutant minMobA, (c) E38A-E74A-E76A triple mutant minMobA. The top panels represent the raw heats of injection; and the bottom panels show the result after integration of the peaks using the Origin software package.

8.5 Inductively coupled plasma mass spectrometry (ICP-MS) analysis

ICP-MS studies were used to identify the most abundant metallic cations in the minMobA protein solution, cell lysate, and LB growth media. As shown in Figure 8.6 panel a, Mg^{2+} and Ca^{2+} are the two most abundant metallic cations in LB media, while Mg^{2+} is the highest metallic cation in the supernatant of cell lysates. Although the absolute quantity of Ca^{2+} is higher than that of Mg^{2+} in minMobA protein solution, the molar ration of metal to protein is almost the same, around 1.4 :1. In other words, Mg^{2+} and Ca^{2+} are equally likely to the metallic cation which is bound to MobA protein in vitro, see Figure 8.6 panel b.

8.6 Conformational change by circular dichroism

The structure of minMobA was characterized chiroptically by circular dichroism. A quantitative analysis of the experimental CD profile suggests the structure is 35.6% α–helix and 20.7% β-sheet. This is in reasonable agreement with the observed X-ray structure of minMobA, which is 31% α-helix and 17% β-sheet(Monzingo, Ozburn et al. 2007). We then assessed the effect of divalent cations on the structure. Ions like Mg^{2+} and Ca^{2+} at a concentration of 1mM did not induce any changes in the CD spectra at 25°C. However, 1 mM Mn^{2+} and Zn^{2+} led to slight perturbations of the CD spectra, as shown in Figure 8.7, panel a. By contrast, 1 mM Ni^{2+} produced a significant decrease in the apparent α-helical content (~8%) at room temperature (figure 8.7, panel b). However there was no apparent alteration of structure at concentrations less than 0.5 mM.

8.7 Thermal denaturation experiments

To explore metal induced structural changes more fully, we carried out thermal denaturation experiments in the presence of different cations (Figure 8.8). Monitoring ellipticity at 220 nm, we observed that Mn^{2+} appears to stabilize the structure, increasing the Tm by 6 °C at a concentration of 1 mM. Other cations like Mg^{2+}, Ca^{2+} and Zn^{2+} at 1 mM or Ni^{2+} at 0.5 mM, do not alter the Tm of minMobA significantly.

8.8 minMobA ssDNA binding activity

A fluorescence polarization assay was used to analyze the binding affinity of six different oligonucleotides (D_{12}, D_{12m}, D_{19}, D_{24}, D_{31} and D_{35}) to mutant minMobA (Y25F), as shown in figures 8.9 and 8.10. D_{12} and D_{12m} are the basically the core sequence of oriT, with only one nucleotide difference between the two. D_{24} is the stem loop part of 35mer oriT, and D_{31}, is the D_{19} is the core sequence with part of the inner arm sequence. The FP assay shows that the

(a)

(b)
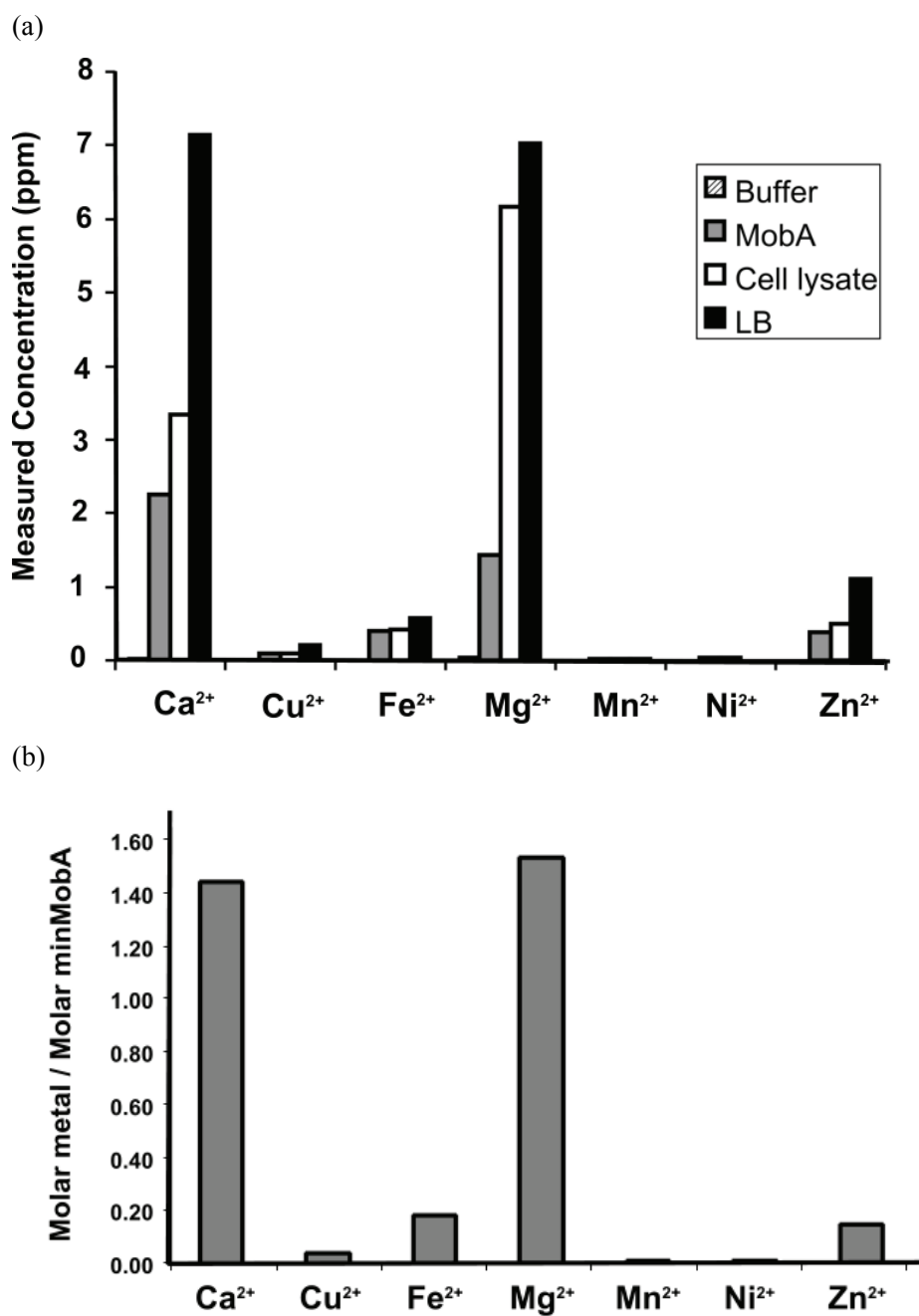
Figure 8.6: (a) Measured concentrations of different metallic cations (Ca2+, Cu2+, Fe2+, Mg2+, Mn2+, Ni2+ and Zn2+) in lysis buffer, minMobA, supernatant of cell lysate and LB media. (b) The calculated mole metal per mole minMobA for corresponding metallic cations. The ppb stands for parts per billion weight, also known as nanograms per gram.

156

(a)

(b)
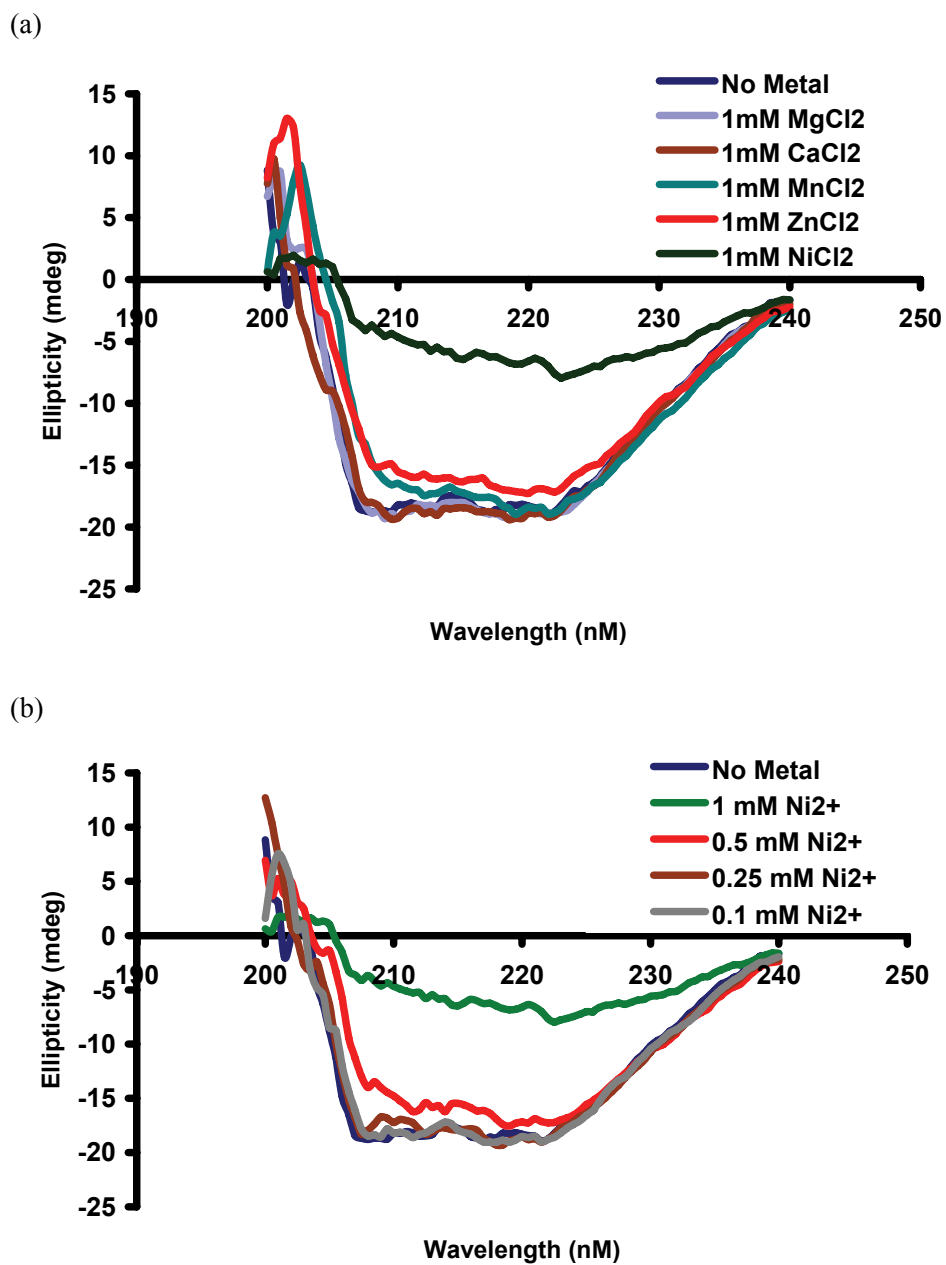
Figure 8.7: Effects of (a) different metallic cations (1mM) and (b) increasing concentrations of Ni^{2+} on minMobA (10μM) CD spectra at room temperature.

157
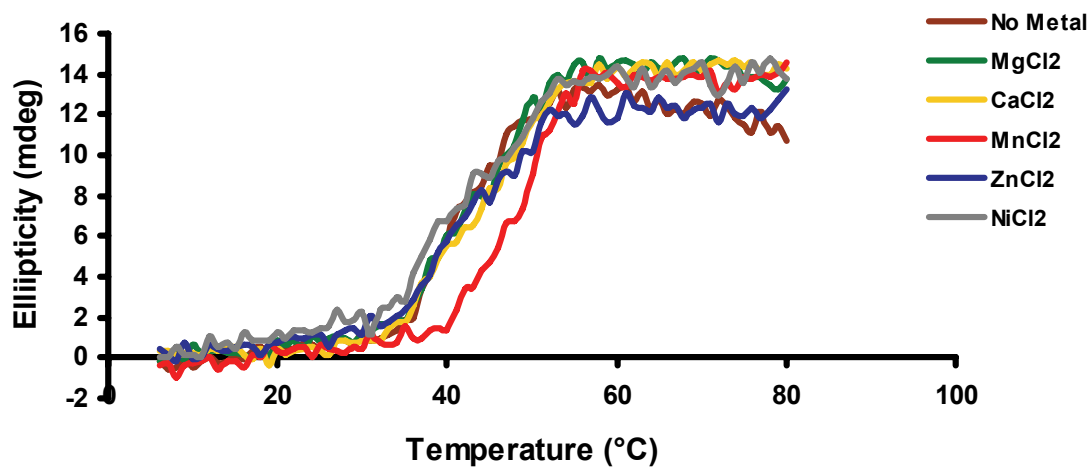
Figure 8.8: Effects of different metallic cations (1mM Mn^{2+}, 1mM Ca^{2+}, 1mM Mg^{2+}, 1mM Zn^{2+} and 0.5mM Ni^{2+}) on the 220nm thermal melting profile of minMobA (10μM) recorded in 10mM HEPES (pH 7.5), 20mM KCl.

Metallic Cation	Tm (°C)	ΔTm (°C)
Mn^{2+}	48.57	+ 5.86
Ni^{2+}	42.56	- 0.15
Mg^{2+}	44.28	+ 1.57
Ca^{2+}	43.69	+ 0.98
Zn^{2+}	42.34	- 0.37
No metal	42.71	

Table 8.5: Shift of Tm of the thermal melting profile of minMobA in the presence of different metallic cations.

Figure 8.9 Different length of oligonucleotides. The outer and inner arm regions are marked in green, and the core region is marked in red. The nicking site is marked with a star.

Figure 8.10: Fluorescence anisotropy assay of binding of minMobA (Y25F) to different length of 3'-Fluorescein labeled oligonucleotides.

Oligonucleotide	K_d (nM)
D35	18
D31	114
D24	6
D19	50
D12	95
D12m	39

Table 8.6: Dissociation constants of minMobA (Y25F) to different length of 3'-Fluorescein labeled oligonucleotides.

oligonucleotide with the stem loop binds to protein tighter than the ones without. Although there is only one nucleotide difference between D_{12m} and D_{12}, it's surprising to see that D_{12m} binds to minMobA two fold tighter than D_{12}, see table 8.6.

Chapter 9: Discussion

9.1 Potential base in the reaction mechanism

The structural data of minMobA shows that it shares a common fold with those of TraI and TrwC. A mechanism for the transesterification reaction of TrwC has been proposed by Roeland. It was inspired by the structure of TrwC complexed with 27mer DNA comprising the hairpin and nicking site (Llosa, Grandoso et al. 1995; Grandoso, Avila et al. 2000; Cesar, Machon et al. 2006). Acting as base, the Asp85 of TrwC abstracts a proton from the catalytic Tyr18, and activates it for nucleophilic attacking the scissile diphosphate bond (Llosa, Grandoso et al. 1995; Grandoso, Avila et al. 2000; Cesar, Machon et al. 2006). As shown in figures 9.1 and 9.2, the structure-based sequence alignment reveals that Glu74 in MobA corresponds in position to Asp85 of TrwC. It was plausible that Glu74 in Mob A might serve as the general base in the mechanism for transesterification. Surprisingly, the conversion of Glu74 to Ala or Gln only decreases the covalent adduct formation in one hour to 50% and 70%, respectively, of wild-type activity. Even the double mutant E74A-E76A still retains some activity. This suggests E74 plays only the most minor role in catalysis. The same phenomena were observed for pSC101 which naturally lack the primase domain. This indicates the phenomenon we observed in minMobA is not an artifact of truncation of protein.

Usually, the cleavage of the phosphodiester bond between the phosphate and the oxygen at the 3' position of the deoxyribose sugar would be an S_N2 reaction. The completion of this reaction requires three chemical entities: a Lewis acid to stabilize the phosphor anion transition state, a general acid to protonate the 3' leaving group and a general base to activate the nucleophile. Unlike Asp85 in TrwC, Glu74 does not act directly as a general base in nicking reaction of MobA. This is consistent with isothermal titration calorimetry (ITC) experiments, in which the mutant minMobA E74A showed the same binding affinity and stoichiometry to Mn^{2+} as wild type protein. Glu76 and Glu38 are two other potential bases in the active sites beside Glu74, as shown in figure 9.3. But as for Glu74, neither Glu76 nor Glu38 acts directly as a general base. The triple mutant E38A-E74A-E76A abolished the nicking activity completely, but still has the same binding affinity and stoichiometry to Mn^{2+} as the wild type protein. This may suggest that those carboxylate residues help stabilize the active configuration of the

```
           1        :             :             :             :          50
           LLLLEEEEEEELLLLLHHHHHHHHHHLLHHHHHLLLLLEEEEEEEELLLLLLLL
MobA       AIYHLTAKTGSGGQSARAKADYIQREGKYARDMDEVLHAESGHMPEFVER
TraI       ---MMSIAQV---RSAGSAGNYYTD-------KDNYMGERWAGAEQLG-D
TrwC       ---XLSHXVLTRQDI--GRAASYYE-------DDYYDASEWQAEELG-KR

           51       :             :             :             :          100
           HHHHHHHHHHHHLLLLLLLLEEEEEEEELLLLLLHHHHHHHHHHHHHHHHHHLLL
MobA       PADYWDAADLYERANGRLFKEVEFALPVELTLDQQKALASEFAQHLTGAE
TraI       KDVFTRLLEDLSRNRHRPGYDLTFSAPKSVLIDAHNQAVDFAVRQVEALG
TrwC       FRELLAG--IXRSSKERIGLDLTFSAPKSVIIKAHDRAVARTLEQAEARG

           101      :             :             :             :          150
           LLLEEEEEEEELLLEEEEEEEELLLLLLLLLLLLHHHHLLLLLLLLLLHHHLL
MobA       RLPYTLAIHAGGGENPHCHLMISERINDGIERPAAQWFKRYNGKTPEKGG
TraI       NLVMALFNHDTQEPQLHTHAVVANVTQHNG------------------E
TrwC       NLVIGKFRHETSDPQLHTHAVILNXTKR-----------------SDGQ

           151      :             :             :
           LLLLLLLLLLLHHHHHHHHHHHHHHHHHHHHHHLLLL
MobA       AQKTEALKPKAWLEQTREAWADHANRALERAGH
TraI       WKTLGFIENYANQIAFGRLYREKLKEQVEALGY
TrwC       WRAL-KNDE-KATRYLGAVYNAELAHELQKLGY
```

Figure 9.1: The structure-based sequence alignment of minMobA with TraI and TrwC. The sequence numbering and secondary structural elements of minMobA are shown on top (H: α helix, E: β strand, and L: random coil).

162
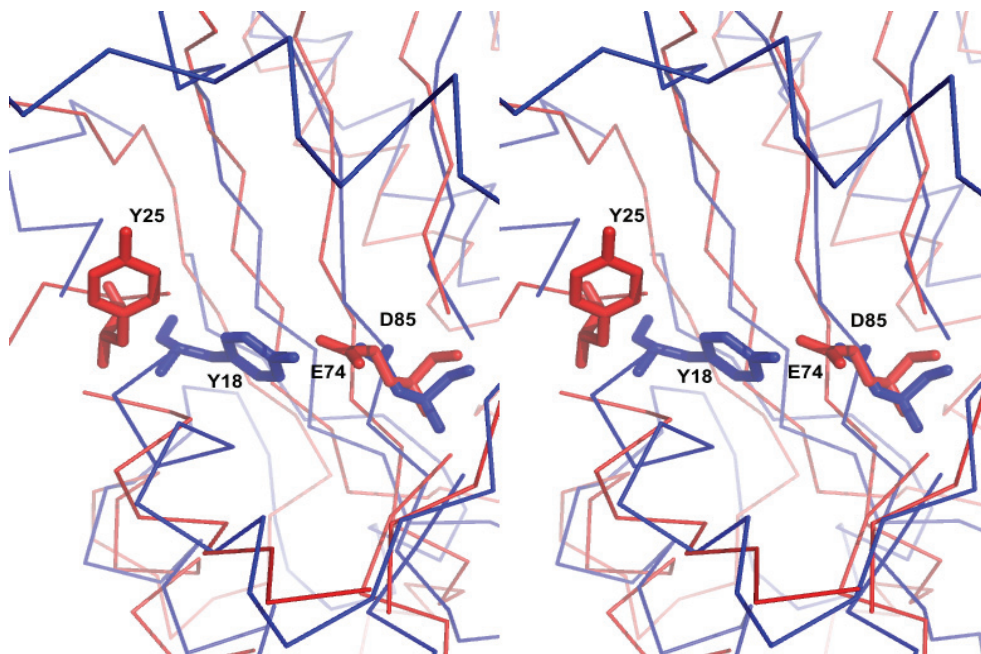
Figure 9.2: Superposition of minMobA (red) and the Trwc (blue). Tyr25 and Glu74 from minMobA are shown as red sticks, while Tyr18 and Asp85 of Trwc are shown as blue sticks.
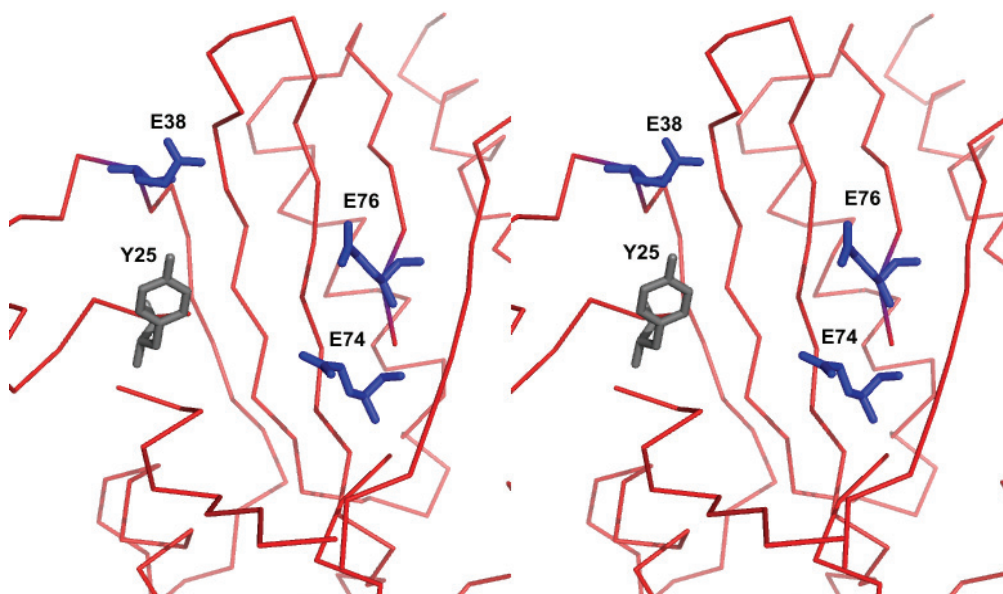
Figure 9.3: Stereo image of active site minMobA. Tyr25 is shown in grey stick, and the nearby Glu38, Glu74 and Glu76 are shown in blue stick.

enzyme. However, it must be recalled that MobA is effectively a single turnover enzyme and there is probably no selective pressure to make it a more efficient enzyme. The normal pKa of Tyr is probably sufficient to generate the necessary nucleophile for the transesterification reaction.

9.2 Role of the metallic cation in MobA

MobA carries out metal ion-dependent DNA cleavage and rejoining reactions as part of its mobilization function. Previously, Scherzinger reported that the nicking activity of MobA, in vitro, depended on the presence of a divalent cation, such as Mg^{2+}, Mn^{2+}, Ca^{2+}, or Ba^{2+}. From the X-ray structure of minMobA, the metal ion is chelated by three histidine residues and a water molecule in a tetrahedral geometry. This arrangement is commonly observed for Zn^{2+} or Ni^{2+} binding proteins(Chivers and Tahirov 2005). However, the biologically relevant cation, and the role of that cation in the nicking reaction is uncertain.

In this project we observed that a bound cation is not necessary for minMobA to physically bind DNA, but we see that such cation-independent binding does not allow the nicking reaction to proceed. Addition of any of a variety of cations increased the magnitude of binding at least 3-fold, and generally allows nicking to occur (Zn^{2+} being the exception). We cannot say if DNA binds in a different orientation in the presence of cations. However, it seems likely that cations organize the MobA active site in a way that improves DNA binding and favors an orientation that is potentially productive. Based on our CD measurements the magnitude of this organization may be small, and have little effect on the backbone structure. (Ni^{2+} is the exception, causing significant structural changes). In the X-ray structure of minMobA, the critical Tyr25 lies at the edge of an α-helix, surrounded by four acidic residues (Asp35, Asp37, Glu38 and Glu74) and four basic residues (Arg28, Arg34, Lys22 and Lys31). The presence of the divalent metallic cations could easily change the conformation of those residues, altering the local protein conformation, stability, and reactivity.

It is also possible that the bound cation participates in the chemistry of the nicking reaction. It may function to align the scissile bond for nucleophilic attack by Tyr 25 or it may polarize that bond to aid cleavage. The fact that the triple mutant E38A-E74A-E76A abolished the nicking activity completely but still had the same binding affinity and stoichiometry to Mn^{2+} also suggests that the role of metallic cation can be purely structural either. It is important to recall that MobA is a very poor enzyme, and indeed, its action requires it to carry out a single

turnover, forming a covalent adduct that persists during transmission out of the cell. As a consequence, there may not be any pressure to optimize the rate of this reaction. Our data show that Mn^{2+} gives the fastest reaction, the strongest DNA complex and the greatest thermal stabilization. However, its advantage over Mg^{2+} and Ca^{2+} is small, and it may be that any of these cations will function well in the mobilization scheme. We showed that Ni^{2+} seems to create structural aberrations that diminish substrate binding. Most surprisingly we found that Zn^{2+}, a cation that should bind strongly to the tetrahedral geometry of the MobA active site, is unable to support the catalytic reaction.

It's still unknown what the physiologically relevant cation is for the relaxase, or indeed if several cations can be used efficaciously. The ICP-MS results show that Mg^{2+} is the metallic cation most available to MobA protein in vitro. Although the tetrahedral geometry with three Histidine residues and a water molecule, as observed in x-ray structure, is more appropriate for Zn^{2+} and Ni^{2+}, we can not neglect the fact that the presence of a phosphate group of nucleotide could provide a better environment for Mn^{2+} or Mg^{2+} in a typical octahedral geometry.

Metal ligation is affected by the hardness of the metals and the ligands. The known order of the hardness of metallic cations used here is: $Mg^{2+} > Ca^{2+} > Mn^{2+} > Ni^{2+} > Zn^{2+}$ (Zechner, Eisenbrandt et al. 2000). The nitrogens of the histidine ligands are soft and should favor the softer metals, like Zn^{2+}, while oxygen (from water or phosphate) is a hard ligand. Our work is consistent with the idea that the cation site is not sharply defined, but may be optimized to Mn^{2+}. This is consistent with our crystallization studies as well. Our largest and most strongly diffraction crystals (Monzingo, Ozburn et al. 2007) grow in the presence of Mn^{2+}. We have grown crystals in the absence of cations, but they are too small to diffract, which suggests they may not maintain a homogeneous structure.

In conclusion, divalent metallic ions affect the conformations of active site residues of minMobA; they may organize key side chains and the local backbone structure to favor DNA cleavage.

9.3 Crystallization of related MobA proteins

Although a hypothetical model of minMobA with 31-mer oligonucleotide has been proposed, the X-ray structure of MobA and oligonucleotide complex would reveal more accurate information of how the nucleotide interacts with protein, and possibly greater insight into the reaction mechanism of DNA nicking activity.

165

We tried to co crystallize minMobA (Y25F) with six different length oligo nucleotides: D_{12}, D_{12m}, D_{19}, D_{24}, D_{31} and D_{35}. So far, only the complex of minMobA (Y25F) with D_{12m} gave some hexagon-shaped crystals after three months. Unfortunately, the crystal didn't diffract well, with the best resolution only 5.0Å. Molecular replacement methods were not successful in phasing the X-ray data, because there were eight molecules in one asymmetric unit. All the molecular replacement software we tried failed to predict more then four molecules in one asymmetric unit. We are still in the process of refine the crystallization condition of this complex, hopefully, it may give a crystal which diffracts better.

The minMobA protein is able to carry out the initial DNA cleavage, but can not interact with T4SS in gene transfer. A longer version of MobA protein called MobA*, which can not only cleave oriT, but also support conjugative transfer of the covalent DNA-protein complex into a new host, has also been screened for crystallization condition (Becker and Meyer 2002). MobA* is a 341 residue N-terminal fragment of MobA; and it has been expressed and partially characterized in Dr. Meyer's laboratory. Because MobA* supports conjugative transfer, it is likely that MobA* may have a 150 residue domain that interacts with the Type IV secretion system besides the minMobA catalytic domain. From genetic experiments, we know that this interaction is vital for initiating the physical movement of the covalent DNA-complex through the secretion pore (Becker and Meyer 2002; Becker and Meyer 2003). So far, the structure of this putative interaction domain is still unknown. So, the structure of MobA* will reveal some important information on the mechanism of drug resistance transfer. MobA* was cloned into three different vectors: pTYB2 (C-terminal intein fusion vector), pET46 (N-terminal His-tag vector) and pET28b (C-terminal His-tag vector). The protein with either N-terminal His-tag or C-terminal His-tag was mis-folded and went into the cell pellet after centrifugation. MobA* with a self-splicing C-terminal intein tag was soluble, but no crystallization condition has been identified so far.

Another protein we produced and attempted to crystallize is the intact MobA from pSC101. It is a 371 residue relaxase homologous to R1162, and naturally lacking the primase domain of the full length MobA from R1162 (Meyer 2000). It has been shown to have a more restrictive specificity than MobA. MobA can bind, cleave and mobilize the pSC101 oriT, but the reverse is not true. The pSC101 plasmid has an accessory protein called MobX, which is a homologous of MobB protein. The pSC101 plasmid is competent for conjugal transfer. In

166

other words, it contains a recognition domain for the type IV secretion system. The crystallization effort will provides additional opportunity to see this important class of structural elements. The pSC101 gene was cloned into both an intein fusion vector and a His-tag vector. Like MobA*, the His-tagged pSC101 was insoluble. The only crystal found in the pSC101screening turned out to be a salt crystal. A truncated version of pSC101 corresponds to minMobA, a 193-residue minPSC101, was engineered for crystallization screening. Surprisingly, this minpSC101 is unable to carry out the nicking reaction, and no crystallization condition has been identified for minpSC101 either.

Reference

Abagyan, R. and M. Totrov (1994). "Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins." J Mol Biol **235**(3): 983-1002.

Abagyan, R. A., M. M. Totrov, et al. (1994). "Icm: A New Method For Protein Modeling and Design: Applications To Docking and Structure Prediction From The Distorted Native Conformation." J.Comp.Chem **15**: 488-506.

Ada, G. (2007). "The importance of vaccination." Front Biosci **12**: 1278-90.

Albo, C., A. Valencia, et al. (1995). "Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein." J Virol **69**(6): 3799-806.

Allignet, J. and N. El Solh (1999). "Comparative analysis of staphylococcal plasmids carrying three streptogramin-resistance genes: vat-vgb-vga." Plasmid **42**(2): 134-8.

Alonso-Caplen, F. V. and R. M. Krug (1991). "Regulation of the extent of splicing of influenza virus NS1 mRNA: role of the rates of splicing and of the nucleocytoplasmic transport of NS1 mRNA." Mol Cell Biol **11**(2): 1092-8.

Andrejeva, J., K. S. Childs, et al. (2004). "The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter." Proc Natl Acad Sci U S A **101**(49): 17264-9.

Aoki, F. Y., G. Boivin, et al. (2007). "Influenza virus susceptibility and resistance to oseltamivir." Antivir Ther **12**(4 Pt B): 603-16.

Aragon, T., S. de la Luna, et al. (2000). "Eukaryotic translation initiation factor 4GI is a cellular target for NS1 protein, a translational activator of influenza virus." Mol Cell Biol **20**(17): 6259-68.

Arai, K. (2001). "[Assay in high throuput screening]." Nippon Yakurigaku Zasshi **118**(2): 81-8.

Attfield, P. V. and R. J. Pinney (1982). "Plasmid R46-mediated protection against bleomycin is poLA+-dependent." J Gen Microbiol **128**(3): 539-47.

Avery, O. R. and C. M. Macleod (1944). "Studies on the chemical nature of the substance inducing transformation in pneumococcal types." Journal of Experimental Medicine **79**: 137-159.

Barabino, S. M., W. Hubner, et al. (1997). "The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins." Genes Dev **11**(13): 1703-16.

Bastia, D., J. Germino, et al. (1981). "The nucleotide sequence surrounding the replication terminus of R6K." Proc Natl Acad Sci U S A **78**(4): 2095-9.

Beaton, A. R. and R. M. Krug (1986). "Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end." Proc Natl Acad Sci U S A **83**(17): 6282-6.

Becker, E. C. and R. J. Meyer (2000). "Recognition of oriT for DNA processing at termination of a round of conjugal transfer." J Mol Biol **300**(5): 1067-77.

Becker, E. C. and R. J. Meyer (2002). "MobA, the DNA strand transferase of plasmid R1162: the minimal domain required for DNA processing at the origin of transfer." J Biol Chem **277**(17): 14575-80.

Becker, E. C. and R. J. Meyer (2003). "Relaxed specificity of the R1162 nickase: a model for evolution of a system for conjugative mobilization of plasmids." J Bacteriol **185**(12): 3538-46.

Biron, C. A. (2001). "Interferons alpha and beta as immune regulators--a new look." Immunity **14**(6): 661-4.

Biswas, S. K., P. L. Boutz, et al. (1998). "Influenza virus nucleoprotein interacts with influenza virus polymerase proteins." J Virol **72**(7): 5493-501.

Biswas, S. K. and D. P. Nayak (1994). "Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1." J Virol **68**(3): 1819-26.

Blaas, D., E. Patzelt, et al. (1982). "Identification of the cap binding protein of influenza virus." Nucleic Acids Res **10**(15): 4803-12.

Block, S. L. (2004). "Role of influenza vaccine for healthy children in the US." Paediatr Drugs **6**(4): 199-209.

Boger, D. L., B. E. Fink, et al. (2001). "A simple, high-resolution method for establishing DNA binding affinity and sequence selectivity." J Am Chem Soc **123**(25): 5878-91.

Bolland, S., M. Llosa, et al. (1990). "General organization of the conjugal transfer genes of the IncW plasmid R388 and interactions between R388 and IncN and IncP plasmids." J Bacteriol **172**(10): 5795-802.

Borchert, T. V., K. V. Kishan, et al. (1995). "Three new crystal structures of point mutation variants of monoTIM: conformational flexibility of loop-1, loop-4 and loop-8." Structure **3**(7): 669-79.

Bornholdt, Z. A. and B. V. Prasad (2006). "X-ray structure of influenza virus NS1 effector domain." Nat Struct Mol Biol **13**(6): 559-60.

Bornholdt, Z. A. and B. V. Prasad (2006). "X-ray structure of influenza virus NS1 effector domain." Nat Struct Mol Biol. **13**: 559-60.

Bornholdt, Z. A. and B. V. Prasad (2008). "X-ray structure of NS1 from a highly pathogenic H5N1 influenza virus." Nature **456**(7224): 985-8.

Bouloy, M., S. J. Plotch, et al. (1980). "Both the 7-methyl and the 2'-O-methyl groups in the cap of mRNA strongly influence its ability to act as primer for influenza virus RNA transcription." Proc Natl Acad Sci U S A **77**(7): 3952-6.

Braam, J., I. Ulmanen, et al. (1983). "Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription." Cell **34**(2): 609-18.

Brown, M. D. (1999). "Green tea (Camellia sinensis) extract and its possible role in the prevention of cancer." Altern Med Rev **4**(5): 360-70.

Brunger, A. T. (1993). "Assessment of phase accuracy by cross validation: the free R value. Methods and applications." Acta Crystallogr D Biol Crystallogr **49**(Pt 1): 24-36.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54 (Pt 5)**: 905-21.

Bui, M., G. Whittaker, et al. (1996). "Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins." J Virol **70**(12): 8391-401.

Burdett, V. (1980). "Identification of tetracycline-resistant R-plasmids in Streptococcus agalactiae (group B)." Antimicrob Agents Chemother **18**(5): 753-60.

Burgui, I., T. Aragon, et al. (2003). "PABP1 and eIF4GI associate with influenza virus NS1 protein in viral mRNA translation initiation complexes." J Gen Virol **84**(Pt 12): 3263-74.

Burleigh, L. M., L. J. Calder, et al. (2005). "Influenza a viruses with mutations in the m1 helix six domain display a wide variety of morphological phenotypes." <u>J Virol</u> **79**(2): 1262-70.

Cabezon, E., J. I. Sastre, et al. (1997). "Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation." <u>Mol Gen Genet</u> **254**(4): 400-6.

Cavasotto, C. N., J. A. Kovacs, et al. (2005). "Representing receptor flexibility in ligand docking through relevant normal modes." <u>J Am Chem Soc</u> **127**(26): 9632-40.

CCP4 (1994). "The CCP4 suite: programs for protein crystallography." <u>Acta Cryst</u> **D5**: 760-763.

Cesar, C. E., C. Machon, et al. (2006). "A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences." <u>Mol Microbiol</u> **62**(4): 984-96.

Chang, C. E., W. Chen, et al. (2007). "Ligand configurational entropy and protein binding." <u>Proc Natl Acad Sci U S A</u> **104**(5): 1534-9.

Chen, D., K. G. Daniel, et al. (2004). "Green tea and tea polyphenols in cancer prevention." <u>Front Biosci</u> **9**: 2618-31.

Chen, L. and H. Y. Zhang (2007). "Cancer preventive mechanisms of the green tea polyphenol (-)-epigallocatechin-3-gallate." <u>Molecules</u> **12**(5): 946-57.

Chen, W., C. E. Chang, et al. (2004). "Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design." <u>Biophys J</u> **87**(5): 3035-49.

Chen, Z., Y. Li, et al. (1999). "Influenza A virus NS1 protein targets poly(A)-binding protein II of the cellular 3'-end processing machinery." <u>Embo J</u> **18**(8): 2273-83.

Cheng, A., S. M. Wong, et al. (2009). "Structural basis for dsRNA recognition by NS1 protein of influenza A virus." <u>Cell Res</u> **19**(2): 187-95.

Cheung, T. K. and L. L. Poon (2007). "Biology of influenza a virus." <u>Ann N Y Acad Sci</u> **1102**: 1-25.

Chien, C. Y., R. Tejero, et al. (1997). "A novel RNA-binding motif in influenza A virus non-structural protein 1." <u>Nat Struct Biol</u> **4**(11): 891-5.

Chien, C. Y., Y. Xu, et al. (2004). "Biophysical characterization of the complex between double-stranded RNA and the N-terminal domain of the NS1 protein from influenza A virus: evidence for a novel RNA-binding mode." <u>Biochemistry</u> **43**(7): 1950-62.

Chivers, P. T. and T. H. Tahirov (2005). "Structure of Pyrococcus horikoshii NikR: nickel sensing and implications for the regulation of DNA recognition." J Mol Biol **348**(3): 597-607.

Claas, E. C., A. D. Osterhaus, et al. (1998). "Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus." Lancet **351**(9101): 472-7.

Connor, R. J., Y. Kawaoka, et al. (1994). "Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates." Virology **205**(1): 17-23.

Crowe, A., C. Ballatore, et al. (2007). "High throughput screening for small molecule inhibitors of heparin-induced tau fibril formation." Biochem Biophys Res Commun **358**(1): 1-6.

Das, K., L.-C. Ma, et al. (2008). "Structural basis for suppression of a host antiviral response by influenza A virus." Proceedings of the National Academy of Sciences: -.

Das, K., L. C. Ma, et al. (2008). "Structural basis for suppression of a host antiviral response by influenza A virus." Proc Natl Acad Sci U S A **105**(35): 13093-8.

Datta, S., C. Larkin, et al. (2003). "Structural insights into single-stranded DNA binding and cleavage by F factor TraI." Structure **11**(11): 1369-79.

Davies, J. (1996). "Origins and evolution of antibiotic resistance." Microbiologia **12**(1): 9-16.

de la Luna, S., C. Martinez, et al. (1989). "Molecular cloning and sequencing of influenza virus A/Victoria/3/75 polymerase genes: sequence evolution and prediction of possible functional domains." Virus Res **13**(2): 143-55.

del Solar, G., R. Diaz, et al. (1987). "Replication of the streptococcal plasmid pMV158 and derivatives in cell-free extracts of Escherichia coli." Mol Gen Genet **206**(3): 428-35.

Digard, P., V. C. Blok, et al. (1989). "Complex formation between influenza virus polymerase proteins expressed in Xenopus oocytes." Virology **171**(1): 162-9.

Droge, M., A. Puhler, et al. (1998). "Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern." J Biotechnol **64**(1): 75-90.

Dyson, P. and M. Evans (1998). "Novel post-replicative DNA modification in Streptomyces: analysis of the preferred modification site of plasmid pIJ101." Nucleic Acids Res **26**(5): 1248-53.

Eitner, K., T. Gaweda, et al. (2007). "eHiTS-to-VMD interface application. The search for tyrosine-tRNA ligase inhibitors." J Chem Inf Model **47**(2): 695-702.

Elleman, C. J. and W. S. Barclay (2004). "The M1 matrix protein controls the filamentous phenotype of influenza A virus." Virology **321**(1): 144-53.

Ellis, J. S. and M. C. Zambon (2002). "Molecular diagnosis of influenza." Rev Med Virol **12**(6): 375-89.

Elster, C., E. Fourest, et al. (1994). "A small percentage of influenza virus M1 protein contains zinc but zinc does not influence in vitro M1-RNA interaction." J Gen Virol **75 (Pt 1)**: 37-42.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.

Enami, M. and K. Enami (1996). "Influenza virus hemagglutinin and neuraminidase glycoproteins stimulate the membrane association of the matrix protein." J Virol **70**(10): 6653-7.

Engleberg, N. C., N. Cianciotto, et al. (1988). "Transfer and maintenance of small, mobilizable plasmids with ColE1 replication origins in Legionella pneumophila." Plasmid **20**(1): 83-91.

Englund, J. A. (2002). "Antiviral therapy of influenza." Semin Pediatr Infect Dis **13**(2): 120-8.

Eswaran, J., J. P. von Kries, et al. (2006). "Crystal structures and inhibitor identification for PTPN5, PTPRR and PTPN7: a family of human MAPK-specific protein tyrosine phosphatases." Biochem J **395**(3): 483-91.

Evans, R. P., Jr. and F. L. Macrina (1983). "Streptococcal R plasmid pIP501: endonuclease site map, resistance determinant location, and construction of novel derivatives." J Bacteriol **154**(3): 1347-55.

Fechner, U., L. Franke, et al. (2003). "Comparison of correlation vector methods for ligand-based similarity searching." J Comput Aided Mol Des **17**(10): 687-98.

Fechter, P., L. Mingay, et al. (2003). "Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding." J Biol Chem **278**(22): 20381-8.

Fells, J. I., R. Tsukahara, et al. (2008). "Identification of non-lipid LPA3 antagonists by virtual screening." Bioorg Med Chem **16**(11): 6207-17.

Fernandez-Sesma, A. (2007). "The influenza virus NS1 protein: inhibitor of innate and adaptive immunity." Infect Disord Drug Targets **7**(4): 336-43.

Ferraris, O. and B. Lina (2008). "Mutations of neuraminidase implicated in neuraminidase inhibitors resistance." J Clin Virol **41**(1): 13-9.

Finlay, B. B., L. S. Frost, et al. (1986). "Nucleotide sequences of five IncF plasmid finP alleles." J Bacteriol **167**(2): 754-7.

Firth, N., K. P. Ridgway, et al. (1993). "Analysis of a transfer region from the staphylococcal conjugative plasmid pSK41." Gene **136**(1-2): 13-25.

Fishman, G. S. (1995). Monte Carlo: Concepts, Algorithms, and Applications. New York, Springer.

Fodor, E. and M. Smith (2004). "The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex." J Virol **78**(17): 9144-53.

Fouchier, R. A., V. Munster, et al. (2005). "Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls." J Virol **79**(5): 2814-22.

Fouchier, R. A., P. M. Schneeberger, et al. (2004). "Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome." Proc Natl Acad Sci U S A **101**(5): 1356-61.

Francia, M. V. and D. B. Clewell (2002). "Transfer origins in the conjugative Enterococcus faecalis plasmids pAD1 and pAM373: identification of the pAD1 nic site, a specific relaxase and a possible TraG-like protein." Mol Microbiol **45**(2): 375-95.

Francia, M. V., A. Varsaki, et al. (2004). "A classification scheme for mobilization regions of bacterial plasmids." FEMS Microbiol Rev **28**(1): 79-100.

Frost, L. S., K. Ippen-Ihler, et al. (1994). "Analysis of the sequence and gene products of the transfer region of the F sex factor." Microbiol Rev **58**(2): 162-210.

Gan, Y. X., J. L. Weaver, et al. (1990). "Aurin tricarboxylic acid, the anti-AIDS compound, prevents the binding of interferon-alpha to its receptor." Biochem Biophys Res Commun **172**(3): 1298-303.

Garcia-Sastre, A. (2006). "Antiviral response in pandemic influenza viruses." Emerg Infect Dis **12**(1): 44-7.

Gilson, M. K. and H. X. Zhou (2007). "Calculation of protein-ligand binding affinities." Annu Rev Biophys Biomol Struct **36**: 21-42.

Gonzalez, S., T. Zurcher, et al. (1996). "Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure." Nucleic Acids Res **24**(22): 4456-63.

Grandoso, G., P. Avila, et al. (2000). "Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation." J Mol Biol **295**(5): 1163-72.

Greenspan, D., P. Palese, et al. (1988). "Two nuclear location signals in the influenza virus NS1 nonstructural protein." J Virol **62**(8): 3020-6.

Griffith, F. (1928). "The significance of pneumococcal types." Journal of Hygenics **27**: 113-159.

Grohmann, E., G. Muth, et al. (2003). "Conjugative plasmid transfer in gram-positive bacteria." Microbiol Mol Biol Rev **67**(2): 277-301, table of contents.

Gubareva, L. V., L. Kaiser, et al. (2000). "Influenza virus neuraminidase inhibitors." Lancet **355**(9206): 827-35.

Guerry, P., J. van Embden, et al. (1974). "Molecular nature of two nonconjugative plasmids carrying drug resistance genes." J Bacteriol **117**(2): 619-30.

Guzman, L. M. and M. Espinosa (1997). "The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT." J Mol Biol **266**(4): 688-702.

Hale, B. G., W. S. Barclay, et al. (2008). "Structure of an avian influenza A virus NS1 protein effector domain." Virology **378**(1): 1-5.

Hale, B. G., R. E. Randall, et al. (2008). "The multifunctional NS1 protein of influenza A viruses." J Gen Virol **89**(Pt 10): 2359-76.

Haller, O., P. Staeheli, et al. (2007). "Interferon-induced Mx proteins in antiviral host defense." Biochimie **89**(6-7): 812-8.

Hama, C., T. Takizawa, et al. (1990). "Organization of the replication control region of plasmid ColIb-P9." J Bacteriol **172**(4): 1983-91.

Harrod, M. E., S. Emery, et al. (2006). "Antivirals in the management of an influenza pandemic." Med J Aust **185**(10 Suppl): S58-61.

Hartshorn, K. L., A. B. Karnad, et al. (1990). "Influenza A virus and the neutrophil: a model of natural immunity." J Leukoc Biol **47**(2): 176-86.

Hausmann, J., E. Kretzschmar, et al. (1997). "Biosynthesis, intracellular transport and enzymatic activity of an avian influenza A virus neuraminidase: role of unpaired cysteines and individual oligosaccharides." J Gen Virol **78 (Pt 12)**: 3233-45.

Hawkins, P. C., G. L. Warren, et al. (2008). "How to do an evaluation: pitfalls and traps." J Comput Aided Mol Des **22**(3-4): 179-90.

Hay, A. J., J. J. Skehel, et al. (1982). "Characterization of influenza virus RNA complete transcripts." Virology **116**(2): 517-22.

Heikkinen, T., O. Ruuskanen, et al. (1991). "Influenza vaccination in the prevention of acute otitis media in children." Am J Dis Child **145**(4): 445-8.

Henkel, J. R. and O. A. Weisz (1998). "Influenza virus M2 protein slows traffic along the secretory pathway. pH perturbation of acidified compartments affects early Golgi transport steps." J Biol Chem **273**(11): 6518-24.

Herz, C., E. Stavnezer, et al. (1981). "Influenza virus, an RNA virus, synthesizes its messenger RNA in the nucleus of infected cells." Cell **26**(3 Pt 1): 391-400.

Honda, A., K. Mizumoto, et al. (2002). "Minimum molecular architectures for transcription and replication of the influenza virus." Proc Natl Acad Sci U S A **99**(20): 13166-71.

Horimoto, T., K. Nakayama, et al. (1994). "Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses." J Virol **68**(9): 6074-8.

Horisberger, M. A. (1995). "Interferons, Mx genes, and resistance to influenza virus." Am J Respir Crit Care Med **152**(4 Pt 2): S67-71.

Hornung, V., J. Ellegast, et al. (2006). "5'-Triphosphate RNA is the ligand for RIG-I." Science **314**(5801): 994-7.

Howland, C. J., C. E. Rees, et al. (1989). "The ssb gene of plasmid ColIb-P9." J Bacteriol **171**(5): 2466-73.

Howland, C. J. and B. M. Wilkins (1988). "Direction of conjugative transfer of IncI1 plasmid ColIb-P9." J Bacteriol **170**(10): 4958-9.

Huang, N., C. Kalyanaraman, et al. (2006). "Molecular mechanics methods for predicting protein-ligand binding." Phys Chem Chem Phys **8**(44): 5166-77.

Huang, T. S., P. Palese, et al. (1990). "Determination of influenza virus proteins required for genome replication." J Virol **64**(11): 5669-73.

Huang, X., T. Liu, et al. (2001). "Effect of influenza virus matrix protein and viral RNA on ribonucleoprotein formation and nuclear export." Virology **287**(2): 405-16.

Hung, H. C., C. P. Tseng, et al. (2009). "Aurintricarboxylic acid inhibits influenza virus neuraminidase." Antiviral Res **81**(2): 123-31.

Ichikawa, D., A. Matsui, et al. (2004). "Effect of various catechins on the IL-12p40 production by murine peritoneal macrophages and a macrophage cell line, J774.1." Biol Pharm Bull **27**(9): 1353-8.

Inglese, J., C. E. Shamu, et al. (2007). "Reporting data from high-throughput screening of small-molecule libraries." Nat Chem Biol **3**(8): 438-41.

Inoue, N. and H. Uchida (1991). "Transcription and initiation of ColE1 DNA replication in Escherichia coli K-12." J Bacteriol **173**(3): 1208-14.

Intharathep, P., C. Laohpongspaisan, et al. (2008). "How amantadine and rimantadine inhibit proton transport in the M2 protein channel." J Mol Graph Model **27**(3): 342-8.

Ippen-Ihler, K. A. and E. G. Minkley, Jr. (1986). "The conjugation system of F, the fertility factor of Escherichia coli." Annu Rev Genet **20**: 593-624.

Jackson, D. A., A. J. Caton, et al. (1982). "Influenza virus RNA is synthesized at fixed sites in the nucleus." Nature **296**(5855): 366-8.

Jain, A. N. (2003). "Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine." J Med Chem **46**(4): 499-511.

Jain, A. N. (2004). "Ligand-based structural hypotheses for virtual screening." J Med Chem **47**(4): 947-61.

Jan, K. (2007). "Avian flu: pandemic preparedness." Home Healthc Nurse **25**(10): 637-42; quiz 643-4.

Kawai, T., K. Takahashi, et al. (2005). "IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction." Nat Immunol **6**(10): 981-8.

Kellenberger, E., J. Rodrigo, et al. (2004). "Comparative evaluation of eight docking tools for docking and virtual screening accuracy." Proteins **57**(2): 225-42.

Khanna, M., P. Kumar, et al. (2008). "Emerging influenza virus: a global threat." J Biosci **33**(4): 475-82.

Kim, H. K., J. E. Kim, et al. (2008). "Aurintricarboxylic acid inhibits endothelial activation, complement activation, and von Willebrand factor secretion in vitro and attenuates

hyperacute rejection in an ex vivo model of pig-to-human pulmonary xenotransplantation." <u>Xenotransplantation</u> **15**(4): 246-56.

Kim, M. H. (2008). "Protein phosphatase 1 activation and alternative splicing of Bcl-X and Mcl-1 by EGCG + ibuprofen." <u>J Cell Biochem</u> **104**(4): 1491-9.

Kistner, O., K. Muller, et al. (1989). "Differential phosphorylation of the nucleoprotein of influenza A viruses." <u>J Gen Virol</u> **70 (Pt 9)**: 2421-31.

Kobayashi, M., T. Toyoda, et al. (1994). "Molecular dissection of influenza virus nucleoprotein: deletion mapping of the RNA binding domain." <u>J Virol</u> **68**(12): 8433-6.

Kobayashi, M., T. Toyoda, et al. (1996). "Influenza virus PB1 protein is the minimal and essential subunit of RNA polymerase." <u>Arch Virol</u> **141**(3-4): 525-39.

Kolter, R. and D. R. Helinski (1982). "Plasmid R6K DNA replication. II. Direct nucleotide sequence repeats are required for an active gamma-origin." <u>J Mol Biol</u> **161**(1): 45-56.

Komano, T., S. R. Kim, et al. (1994). "DNA rearrangement of the shufflon determines recipient specificity in liquid mating of IncI1 plasmid R64." <u>J Mol Biol</u> **243**(1): 6-9.

Kovacs, J. A., P. Chacon, et al. (2004). "Predictions of protein flexibility: first-order measures." <u>Proteins</u> **56**(4): 661-8.

Krishnan, R., M. Caligaris, et al. (2004). "Removal of the superficial zone of bovine articular cartilage does not increase its frictional coefficient." <u>Osteoarthritis Cartilage</u> **12**(12): 947-55.

Krug, R. M. (1993). "The regulation of export of mRNA from nucleus to cytoplasm." <u>Curr Opin Cell Biol</u> **5**(6): 944-9.

Krug, R. M. (2003). "The potential use of influenza virus as an agent for bioterrorism." <u>Antiviral Res</u> **57**(1-2): 147-50.

Krug, R. M., B. A. Broni, et al. (1979). "Are the 5' ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs?" <u>Cell</u> **18**(2): 329-34.

Lackenby, A., C. I. Thompson, et al. (2008). "The potential impact of neuraminidase inhibitor resistant influenza." <u>Curr Opin Infect Dis</u> **21**(6): 626-38.

Lamb, R. A., S. L. Zebedee, et al. (1985). "Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface." <u>Cell</u> **40**(3): 627-33.

Laskowski, R. A., M. W. MacArthur, et al. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." <u>J. App. Cryst.</u> **26**: 283-291.

Laver, W. G. (1985). "Immunochemistry of variants of influenza virus hemagglutinin and neuraminidase." <u>Adv Exp Med Biol</u> **185**: 149-74.

Laver, W. G., P. M. Colman, et al. (1984). "Influenza virus neuraminidase with hemagglutinin activity." <u>Virology</u> **137**(2): 314-23.

Le Bon, A., G. Schiavoni, et al. (2001). "Type i interferons potently enhance humoral immunity and can promote isotype switching by stimulating dendritic cells in vivo." <u>Immunity</u> **14**(4): 461-70.

Lessl, M., W. Pansegrau, et al. (1992). "Relationship of DNA-transfer-systems: essential transfer factors of plasmids RP4, Ti and F share common sequences." <u>Nucleic Acids Res</u> **20**(22): 6099-100.

Li, K. S., Y. Guan, et al. (2004). "Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia." <u>Nature</u> **430**(6996): 209-13.

Li, M. L., P. Rao, et al. (2001). "The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits." <u>Embo J</u> **20**(8): 2078-86.

Li, S., J. Y. Min, et al. (2006). "Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA." <u>Virology</u> **349**(1): 13-21.

Li, S., J. Schulman, et al. (1993). "Glycosylation of neuraminidase determines the neurovirulence of influenza A/WSN/33 virus." <u>J Virol</u> **67**(11): 6667-73.

Li, W., P. A. Escarpe, et al. (1998). "Identification of GS 4104 as an orally bioavailable prodrug of the influenza virus neuraminidase inhibitor GS 4071." <u>Antimicrob Agents Chemother</u> **42**(3): 647-53.

Lin-Chao, S. and S. N. Cohen (1991). "The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids in vivo." <u>Cell</u> **65**(7): 1233-42.

Lipkind, M. and E. Shihmanter (1986). "Antigenic relationships between avian paramyxoviruses. I. Quantitative characteristics based on hemagglutination and neuraminidase inhibition tests." <u>Arch Virol</u> **89**(1-4): 89-111.

Liu, C., M. C. Eichelberger, et al. (1995). "Influenza type A virus neuraminidase does not play a role in viral entry, replication, assembly, or budding." <u>J Virol</u> **69**(2): 1099-106.

Liu, J., P. A. Lynch, et al. (1997). "Crystal structure of the unique RNA-binding domain of the influenza virus NS1 protein." <u>Nat Struct Biol</u> **4**(11): 896-9.

179

Llosa, M., S. Bolland, et al. (1991). "Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46." Mol Gen Genet **226**(3): 473-83.

Llosa, M., S. Bolland, et al. (1994). "Genetic organization of the conjugal DNA processing region of the IncW plasmid R388." J Mol Biol **235**(2): 448-64.

Llosa, M., G. Grandoso, et al. (1995). "Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388." J Mol Biol **246**(1): 54-62.

Llosa, M., G. Grandoso, et al. (1996). "Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function." J Mol Biol **264**(1): 56-67.

Lorenz, M. G. and W. Wackernagel (1994). "Bacterial gene transfer by natural genetic transformation in the environment." Microbiol Rev **58**(3): 563-602.

Lowen, A. C. and P. Palese (2007). "Influenza virus transmission: basic science and implications for the use of antiviral drugs during a pandemic." Infect Disord Drug Targets **7**(4): 318-28.

Ludwig, S. and O. Planz (2008). "Influenza viruses and the NF-kappaB signaling pathway - towards a novel concept of antiviral therapy." Biol Chem **389**(10): 1307-12.

Luther, P., K. C. Bergmann, et al. (1984). "An investigation of antigenic drift of neuraminidases of influenza A (H1N1) viruses." J Hyg (Lond) **92**(2): 223-9.

Margo, K. L. and A. F. Shaughnessy (1998). "Antiviral drugs in healthy children." Am Fam Physician **57**(5): 1073-7.

Marienfeld, C., L. Tadlock, et al. (2003). "Inhibition of cholangiocarcinoma growth by tannic acid." Hepatology **37**(5): 1097-104.

Marra, D. and J. R. Scott (1999). "Regulation of excision of the conjugative transposon Tn916." Mol Microbiol **31**(2): 609-21.

Marra, D., J. G. Smith, et al. (1999). "Excision of the conjugative transposon Tn916 in Lactococcus lactis." Appl Environ Microbiol **65**(5): 2230-1.

Martin, K. and A. Helenius (1991). "Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import." Cell **67**(1): 117-30.

Matson, S. W., J. K. Sampson, et al. (2001). "F plasmid conjugative DNA transfer: the TraI helicase activity is essential for DNA strand transfer." J Biol Chem **276**(4): 2372-9.

McInnes, C. (2007). "Virtual screening strategies in drug discovery." <u>Curr Opin Chem Biol</u> **11**(5): 494-502.

Meyer, R. (2000). "Identification of the mob genes of plasmid pSC101 and characterization of a hybrid pSC101-R1162 system for conjugal mobilization." <u>J Bacteriol</u> **182**(17): 4875-81.

Meyer, R., M. Hinds, et al. (1982). "Properties of R1162, a broad-host-range, high-copy-number plasmid." <u>J Bacteriol</u> **150**(2): 552-62.

Meyer, R. J., L. S. Lin, et al. (1985). "Broad host-range plasmid R1162: replication, incompatibility, and copy-number control." <u>Basic Life Sci</u> **30**: 173-88.

Mikhailov, M. V., J. D. Campbell, et al. (2005). "3-D structural and functional characterization of the purified KATP channel complex Kir6.2-SUR1." <u>Embo J</u> **24**(23): 4166-75.

Min, J. Y. and R. M. Krug (2006). "The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2'-5' oligo (A) synthetase/RNase L pathway." <u>Proc Natl Acad Sci U S A</u> **103**(18): 7100-5.

Min, J. Y., S. Li, et al. (2007). "A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis." <u>Virology</u> **363**(1): 236-43.

Monto, A. S. (2008). "Antivirals and influenza: frequency of resistance." <u>Pediatr Infect Dis J</u> **27**(10 Suppl): S110-2.

Monto, A. S., A. Webster, et al. (1999). "Randomized, placebo-controlled studies of inhaled zanamivir in the treatment of influenza A and B: pooled efficacy analysis." <u>J Antimicrob Chemother</u> **44 Suppl B**: 23-9.

Monzingo, A. F., A. Ozburn, et al. (2007). "The structure of the minimal relaxase domain of MobA at 2.1 A resolution." <u>J Mol Biol</u> **366**(1): 165-78.

Moscona, A. (2004). "Oseltamivir-resistant influenza?" <u>Lancet</u> **364**(9436): 733-4.

Moscoso, M., R. Eritja, et al. (1997). "Initiation of replication of plasmid pMV158: mechanisms of DNA strand-transfer reactions mediated by the initiator RepB protein." <u>J Mol Biol</u> **268**(5): 840-56.

Mukaigawa, J. and D. P. Nayak (1991). "Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2." <u>J Virol</u> **65**(1): 245-53.

Nagata, K., A. Kawaguchi, et al. (2008). "Host factors for replication and transcription of the influenza virus genome." Rev Med Virol **18**(4): 247-60.

Nagata, K., K. Takeuchi, et al. (1989). "In vitro synthesis of influenza viral RNA: biochemical complementation assay of factors required for influenza virus replication." J Biochem **106**(2): 205-8.

Nakayama, J. and A. Suzuki (1997). "Genetic analysis of plasmid-specific pheromone signaling encoded by pPD1 in Enterococcus faecalis." Biosci Biotechnol Biochem **61**(11): 1796-9.

Nasir, M. S. and M. E. Jolley (1999). "Fluorescence polarization: an analytical tool for immunoassay and drug discovery." Comb Chem High Throughput Screen **2**(4): 177-90.

Nemeroff, M. E., X. Y. Qian, et al. (1995). "The influenza virus NS1 protein forms multimers in vitro and in vivo." Virology **212**(2): 422-8.

Nemeroff, M. E., U. Utans, et al. (1992). "Identification of cis-acting intron and exon regions in influenza virus NS1 mRNA that inhibit splicing and cause the formation of aberrantly sedimenting presplicing complexes." Mol Cell Biol **12**(3): 962-70.

Neumann, G., M. R. Castrucci, et al. (1997). "Nuclear import and export of influenza virus nucleoprotein." J Virol **71**(12): 9690-700.

Neumann, G., M. T. Hughes, et al. (2000). "Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1." Embo J **19**(24): 6751-8.

Nichol, K. L. (2006). "Improving influenza vaccination rates among adults." Cleve Clin J Med **73**(11): 1009-15.

Nichol, K. L., A. Lind, et al. (1995). "The effectiveness of vaccination against influenza in healthy, working adults." N Engl J Med **333**(14): 889-93.

Nicholson, K. G. (1996). "Use of antivirals in influenza in the elderly: prophylaxis and therapy." Gerontology **42**(5): 280-9.

Nicola, G., C. A. Smith, et al. (2007). "Discovery of novel inhibitors targeting enoyl-acyl carrier protein reductase in Plasmodium falciparum by structure-based virtual screening." Biochem Biophys Res Commun **358**(3): 686-91.

Nijkamp, H. J., R. de Lang, et al. (1986). "The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13." Plasmid **16**(2): 135-60.

Noah, D. L., K. Y. Twu, et al. (2003). "Cellular antiviral responses against influenza A virus are countered at the posttranscriptional level by the viral NS1A protein via its binding to a cellular protein required for the 3' end processing of cellular pre-mRNAS." Virology **307**(2): 386-95.

Nunez, B. and F. De La Cruz (2001). "Two atypical mobilization proteins are involved in plasmid CloDF13 relaxation." Mol Microbiol **39**(4): 1088-99.

O'Neill, R. E., R. Jaskunas, et al. (1995). "Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import." J Biol Chem **270**(39): 22701-4.

O'Neill, R. E., J. Talon, et al. (1998). "The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins." Embo J **17**(1): 288-96.

Olsen, L., S. Jost, et al. (2006). "New leads of metallo-beta-lactamase inhibitors from structure-based pharmacophore design." Bioorg Med Chem **14**(8): 2627-35.

Oxford, J. S., A. Mann, et al. (2003). "A designer drug against influenza: the NA inhibitor oseltamivir (Tamiflu)." Expert Rev Anti Infect Ther **1**(2): 337-42.

Palese, P. (2004). "Influenza: old and new threats." Nat Med **10**(12 Suppl): S82-7.

Palese, P. and R. W. Compans (1976). "Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA): mechanism of action." J Gen Virol **33**(1): 159-63.

Palese, P., K. Tobita, et al. (1974). "Characterization of temperature sensitive influenza virus mutants defective in neuraminidase." Virology **61**(2): 397-410.

Pansegrau, W., E. Lanka, et al. (1994). "Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis." J Mol Biol **239**(5): 623-63.

Park, E. K., M. R. Castrucci, et al. (1998). "The M2 ectodomain is important for its incorporation into influenza A virions." J Virol **72**(3): 2449-55.

Parker, C. and R. J. Meyer (2007). "The R1162 relaxase/primase contains two, type IV transport signals that require the small plasmid protein MobB." Mol Microbiol **66**(1): 252-61.

Parker, C., X. L. Zhang, et al. (2002). "Conjugative DNA synthesis: R1162 and the question of rolling-circle replication." Plasmid **48**(3): 186-92.

Parkinson, J. S. (1975). "Genetics of chemotactic behavior in bacteria." Cell **4**(3): 183-8.

Perales, B., S. de la Luna, et al. (1996). "Mutational analysis identifies functional domains in the influenza A virus PB2 polymerase subunit." J Virol **70**(3): 1678-86.

Perales, B. and J. Ortin (1997). "The influenza A virus PB2 polymerase subunit is required for the replication of viral RNA." J Virol **71**(2): 1381-5.

Pichlmair, A., O. Schulz, et al. (2006). "RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates." Science **314**(5801): 997-1001.

Pinto, L. H., L. J. Holsinger, et al. (1992). "Influenza virus M2 protein has ion channel activity." Cell **69**(3): 517-28.

Plotch, S. J., M. Bouloy, et al. (1979). "Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro." Proc Natl Acad Sci U S A **76**(4): 1618-22.

Poole, E., D. Elton, et al. (2004). "Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites." Virology **321**(1): 120-33.

Priebe, S. D. and S. A. Lacks (1989). "Region of the streptococcal plasmid pMV158 required for conjugative mobilization." J Bacteriol **171**(9): 4778-84.

Qian, X. Y., C. Y. Chien, et al. (1995). "An amino-terminal polypeptide fragment of the influenza virus NS1 protein possesses specific RNA-binding activity and largely helical backbone structure." Rna **1**(9): 948-56.

Rawlings, D. E. and E. Tietze (2001). "Comparative biology of IncQ and IncQ-like plasmids." Microbiol Mol Biol Rev **65**(4): 481-96, table of contents.

Reddy, A., L. Battisti, et al. (1987). "Identification of self-transmissible plasmids in four Bacillus thuringiensis subspecies." J Bacteriol **169**(11): 5263-70.

Rees, W. A., J. D. Harkins, et al. (1997). "Amantadine and equine influenza: pharmacology, pharmacokinetics and neurological effects in the horse." Equine Vet J **29**(2): 104-10.

Richardson, J. C. and R. K. Akkina (1991). "NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells." Arch Virol **116**(1-4): 69-80.

Rishi, V., T. Potter, et al. (2005). "A high-throughput fluorescence-anisotropy screen that identifies small molecule inhibitors of the DNA binding of B-ZIP transcription factors." Anal Biochem **340**(2): 259-71.

Ritchie, K. J. and D. E. Zhang (2004). "ISG15: the immunological kin of ubiquitin." Semin Cell Dev Biol **15**(2): 237-46.

184

Roehrl, M. H., J. Y. Wang, et al. (2004). "Discovery of small-molecule inhibitors of the NFAT--calcineurin interaction by competitive high-throughput fluorescence polarization screening." Biochemistry **43**(51): 16067-75.

Roehrl, M. H., J. Y. Wang, et al. (2004). "A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization." Biochemistry **43**(51): 16056-66.

Sakaguchi, A., E. Hirayama, et al. (2003). "Nuclear export of influenza viral ribonucleoprotein is temperature-dependently inhibited by dissociation of viral matrix protein." Virology **306**(2): 244-53.

Samuel, C. E. (2007). "Innate immunity minireview series: making biochemical sense of nucleic acid sensors that trigger antiviral innate immunity." J Biol Chem **282**(21): 15313-4.

Sanz-Ezquerro, J. J., S. de la Luna, et al. (1995). "Individual expression of influenza virus PA protein induces degradation of coexpressed proteins." J Virol **69**(4): 2420-6.

Sanz-Ezquerro, J. J., T. Zurcher, et al. (1996). "The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis." J Virol **70**(3): 1905-11.

Saul, D., A. J. Spiers, et al. (1989). "Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids." J Bacteriol **171**(5): 2697-707.

Schapira, M., B. M. Raaka, et al. (2003). "Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking." Proc Natl Acad Sci U S A **100**(12): 7354-9.

Schapira, M., M. Totrov, et al. (1999). "Prediction of the binding energy for small molecules, peptides and proteins." J Mol Recognit **12**(3): 177-90.

Scherzinger, E., V. Kruft, et al. (1993). "Purification of the large mobilization protein of plasmid RSF1010 and characterization of its site-specific DNA-cleaving/DNA-joining activity." Eur J Biochem **217**(3): 929-38.

Scherzinger, E., R. Lurz, et al. (1992). "In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins." Nucleic Acids Res **20**(1): 41-8.

Scholz, P., V. Haring, et al. (1985). "Replication determinants of the broad host-range plasmid RSF1010." Basic Life Sci **30**: 243-59.

Schulman, J. L. (1971). "Influenza." Postgrad Med **50**(5): 171-5.

Servin-Gonzalez, L. (1993). "Relationship between the replication functions of Streptomyces plasmids pJV1 and pIJ101." Plasmid **30**(2): 131-40.

Seth, R. B., L. Sun, et al. (2005). "Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3." Cell **122**(5): 669-82.

Shafferman, A., R. Kolter, et al. (1982). "Plasmid R6K DNA replication. III. Regulatory properties of the pi initiation protein." J Mol Biol **161**(1): 57-76.

Shapiro, G. I. and R. M. Krug (1988). "Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer." J Virol **62**(7): 2285-90.

Shih, S. R., M. E. Nemeroff, et al. (1995). "The choice of alternative 5' splice sites in influenza virus M1 mRNA is regulated by the viral polymerase complex." Proc Natl Acad Sci U S A **92**(14): 6324-8.

Shimizu, K., H. Handa, et al. (1994). "Regulation of influenza virus RNA polymerase activity by cellular and viral factors." Nucleic Acids Res **22**(23): 5047-53.

Shin, E. S., J. Park, et al. (2008). "Catechin gallates are NADP+-competitive inhibitors of glucose-6-phosphate dehydrogenase and other enzymes that employ NADP+ as a coenzyme." Bioorg Med Chem **16**(7): 3580-6.

Sidwell, R. W., K. W. Bailey, et al. (1999). "Influence of treatment schedule and viral challenge dose on the in vivo influenza virus-inhibitory effects of the orally administered neuraminidase inhibitor GS 4104." Antivir Chem Chemother **10**(4): 187-93.

Silverman, R. H. (2007). "Viral encounters with 2',5'-oligoadenylate synthetase and RNase L during the interferon antiviral response." J Virol **81**(23): 12720-9.

Skehel, J. J., P. M. Bayley, et al. (1982). "Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion." Proc Natl Acad Sci U S A **79**(4): 968-72.

Song, J. M., K. H. Lee, et al. (2005). "Antiviral effect of catechins in green tea on influenza virus." Antiviral Res **68**(2): 66-74.

Song, J. M., K. D. Park, et al. (2007). "Biological evaluation of anti-influenza viral activity of semi-synthetic catechin derivatives." Antiviral Res **76**(2): 178-85.

Sorensen, S. J., M. Bailey, et al. (2005). "Studying plasmid horizontal transfer in situ: a critical review." Nat Rev Microbiol **3**(9): 700-10.

Spataro, V., T. Toda, et al. (1997). "Resistance to diverse drugs and ultraviolet light conferred by overexpression of a novel human 26 S proteasome subunit." J Biol Chem **272**(48): 30470-5.

Stachel, S. E. and P. C. Zambryski (1986). "Agrobacterium tumefaciens and the susceptible plant cell: a novel adaptation of extracellular recognition and DNA conjugation." Cell **47**(2): 155-7.

Stalker, D. M., R. Kolter, et al. (1982). "Plasmid R6K DNA replication. I. Complete nucleotide sequence of an autonomously replicating segment." J Mol Biol **161**(1): 33-43.

Stark, G. R., I. M. Kerr, et al. (1998). "How cells respond to interferons." Annu Rev Biochem **67**: 227-64.

Staudt, L. M. and W. Gerhard (1983). "Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. I. Significant variation in repertoire expression between individual mice." J Exp Med **157**(2): 687-704.

Steinmetzer, K., K. Kuhn, et al. (2002). "Plasmid pIP501 encoded transcriptional repressor CopR: single amino acids involved in dimerization are also important for folding of the monomer." Plasmid **47**(3): 201-9.

Street, L. M., M. J. Harley, et al. (2003). "Subdomain organization and catalytic residues of the F factor TraI relaxase domain." Biochim Biophys Acta **1646**(1-2): 86-99.

Sun, H. (2008). "Pharmacophore-based virtual screening." Curr Med Chem **15**(10): 1018-24.

Suzuki, H., R. Saito, et al. (2003). "Emergence of amantadine-resistant influenza A viruses: epidemiological study." J Infect Chemother **9**(3): 195-200.

Tamura, S. and T. Kurata (2004). "Defense mechanisms against influenza virus infection in the respiratory tract mucosa." Jpn J Infect Dis **57**(6): 236-47.

Taylor, I. M., J. L. Harrison, et al. (1990). "The TraT lipoprotein as a vehicle for the transport of foreign antigenic determinants to the cell surface of Escherichia coli K12: structure-function relationships in the TraT protein." Mol Microbiol **4**(8): 1259-68.

Tenover, F. C. (1995). "The best of times, the worst of times. The global challenge of antimicrobial resistance." <u>Pharm World Sci</u> **17**(5): 149-51.

Thomas, C. M. and K. M. Nielsen (2005). "Mechanisms of, and barriers to, horizontal gene transfer between bacteria." <u>Nat Rev Microbiol</u> **3**(9): 711-21.

Thomas, M. P., C. McInnes, et al. (2006). "Protein structures in virtual screening: a case study with CDK2." <u>J Med Chem</u> **49**(1): 92-104.

Thorsted, P. B., D. P. Macartney, et al. (1998). "Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncP backbone." <u>J Mol Biol</u> **282**(5): 969-90.

Tiffert, Y., B. Gotz, et al. (2007). "Conjugative DNA transfer in Streptomyces: SpdB2 involved in the intramycelial spreading of plasmid pSVH1 is an oligomeric integral membrane protein that binds to dsDNA." <u>Microbiology</u> **153**(Pt 9): 2976-83.

Totrov, M. and R. Abagyan (1996). "The contour-buildup algorithm to calculate the analytical molecular surface." <u>J Struct Biol</u> **116**(1): 138-43.

Totrov, M. and R. Abagyan (1997). "Flexible protein-ligand docking by global energy optimization in internal coordinates." <u>Proteins</u> **Suppl 1**: 215-20.

Twu, K. Y., R. L. Kuo, et al. (2007). "The H5N1 influenza virus NS genes selected after 1998 enhance virus replication in mammalian cells." <u>J Virol</u> **81**(15): 8112-21.

Twu, K. Y., D. L. Noah, et al. (2006). "The CPSF30 binding site on the NS1A protein of influenza A virus is a potential antiviral target." <u>J Virol</u> **80**(8): 3957-65.

Uematsu, S. and S. Akira (2007). "Toll-like receptors and Type I interferons." <u>J Biol Chem</u> **282**(21): 15319-23.

Ulmanen, I., B. Broni, et al. (1983). "Influenza virus temperature-sensitive cap (m7GpppNm)-dependent endonuclease." <u>J Virol</u> **45**(1): 27-35.

Ulmanen, I., B. A. Broni, et al. (1981). "Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription." <u>Proc Natl Acad Sci U S A</u> **78**(12): 7355-9.

Vagin, A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." <u>J. Appl. Cryst.</u> **30**: 1022-1025.

van Putten, A. J., G. J. Jochems, et al. (1987). "Structure and nucleotide sequence of the region encoding the mobilization proteins of plasmid CloDF13." <u>Gene</u> **51**(2-3): 171-8.

Varghese, J. N. and P. M. Colman (1991). "Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 A resolution." J Mol Biol **221**(2): 473-86.

Verdonk, M. L., G. Chessari, et al. (2005). "Modeling water molecules in protein-ligand docking using GOLD." J Med Chem **48**(20): 6504-15.

Vines, A., K. Wells, et al. (1998). "The role of influenza A virus hemagglutinin residues 226 and 228 in receptor specificity and host range restriction." J Virol **72**(9): 7626-31.

von Itzstein, M., W. Y. Wu, et al. (1993). "Rational design of potent sialidase-based inhibitors of influenza virus replication." Nature **363**(6428): 418-23.

Vreede, F. T., T. E. Jung, et al. (2004). "Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates." J Virol **78**(17): 9568-72.

Wakefield, L. and G. G. Brownlee (1989). "RNA-binding properties of influenza A virus matrix protein M1." Nucleic Acids Res **17**(21): 8569-80.

Waleh, N. S. and B. A. Stocker (1979). "Effect of host lex, recA, recF, and uvrD genotypes on the ultraviolet light-protecting and related properties of plasmid R46 in Escherichia coli." J Bacteriol **137**(2): 830-8.

Wang, C., R. A. Lamb, et al. (1994). "Direct measurement of the influenza A virus M2 protein ion channel activity in mammalian cells." Virology **205**(1): 133-40.

Wang, W. and R. M. Krug (1996). "The RNA-binding and effector domains of the viral NS1 protein are conserved to different extents among influenza A and B viruses." Virology **223**(1): 41-50.

Wang, W., K. Riedel, et al. (1999). "RNA binding by the novel helical domain of the influenza virus NS1 protein requires its dimer structure and a small number of specific basic amino acids." Rna **5**(2): 195-205.

Wang, X., C. F. Basler, et al. (2002). "Functional replacement of the carboxy-terminal two-thirds of the influenza A virus NS1 protein with short heterologous dimerization domains." J Virol **76**(24): 12951-62.

Warren, G. L., C. W. Andrews, et al. (2006). "A critical assessment of docking programs and scoring functions." J Med Chem **49**(20): 5912-31.

Waters, V. L. (1999). "Conjugative transfer in the dissemination of beta-lactam and aminoglycoside resistance." Front Biosci **4**: D433-56.

189

Waters, V. L. (2001). "Conjugation between bacterial and mammalian cells." <u>Nat Genet</u> **29**(4): 375-6.

Webby, R. J., D. R. Perez, et al. (2004). "Responsiveness to a pandemic alert: use of reverse genetics for rapid development of influenza vaccines." <u>Lancet</u> **363**(9415): 1099-103.

Webby, R. J. and R. G. Webster (2003). "Are we ready for pandemic influenza?" <u>Science</u> **302**(5650): 1519-22.

Weber, F., G. Kochs, et al. (2004). "Inverse interference: how viruses fight the interferon system." <u>Viral Immunol</u> **17**(4): 498-515.

Weber, T. P. and N. I. Stilianakis (2007). "Ecologic immunology of avian influenza (H5N1) in migratory birds." <u>Emerg Infect Dis</u> **13**(8): 1139-43.

Whittaker, G., M. Bui, et al. (1996). "Nuclear trafficking of influenza virus ribonuleoproteins in heterokaryons." <u>J Virol</u> **70**(5): 2743-56.

Willetts, N. and C. Crowther (1981). "Mobilization of the non-conjugative IncQ plasmid RSF1010." <u>Genet Res</u> **37**(3): 311-6.

Williams, B. R. (1999). "PKR; a sentinel kinase for cellular stress." <u>Oncogene</u> **18**(45): 6112-20.

Winter, G. and S. Fields (1981). "The structure of the gene encoding the nucleoprotein of human influenza virus A/PR/8/34." <u>Virology</u> **114**(2): 423-8.

Wolfram, S. (2007). "Effects of green tea and EGCG on cardiovascular and metabolic health." <u>J Am Coll Nutr</u> **26**(4): 373S-388S.

Wright, G. D. (2007). "The antibiotic resistome: the nexus of chemical and genetic diversity." <u>Nat Rev Microbiol</u> **5**(3): 175-86.

Xia, S., A. F. Monzingo, et al. (2009). "Structure of NS1A effector domain from the influenza A/Udorn/72 virus." <u>Acta Crystallogr D Biol Crystallogr</u> **65**(Pt 1): 11-7.

Yamada, A., L. E. Brown, et al. (1984). "Characterization of H2 influenza virus hemagglutinin with monoclonal antibodies: influence of receptor specificity." <u>Virology</u> **138**(2): 276-86.

Yang, E. B., L. Wei, et al. (2006). "Tannic acid, a potent inhibitor of epidermal growth factor receptor tyrosine kinase." <u>J Biochem</u> **139**(3): 495-502.

Yasuda, J., S. Nakada, et al. (1993). "Molecular assembly of influenza virus: association of the NS2 protein with virion matrix." <u>Virology</u> **196**(1): 249-55.

Ye, Z., T. Liu, et al. (1999). "Association of influenza virus matrix protein with ribonucleoproteins." J Virol **73**(9): 7467-73.

Ye, Z., D. Robinson, et al. (1995). "Nucleus-targeting domain of the matrix protein (M1) of influenza virus." J Virol **69**(3): 1964-70.

Ye, Z. P., N. W. Baylor, et al. (1989). "Transcription-inhibition and RNA-binding domains of influenza A virus matrix protein mapped with anti-idiotypic antibodies and synthetic peptides." J Virol **63**(9): 3586-94.

Yoneyama, M., M. Kikuchi, et al. (2004). "The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses." Nat Immunol **5**(7): 730-7.

Yuan, W., J. M. Aramini, et al. (2002). "Structural basis for ubiquitin-like ISG 15 protein binding to the NS1 protein of influenza B virus: a protein-protein interaction function that is not shared by the corresponding N-terminal domain of the NS1 protein of influenza A virus." Virology **304**(2): 291-301.

Yuan, W. and R. M. Krug (2001). "Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein." Embo J **20**(3): 362-71.

Zatyka, M., G. Jagura-Burdzy, et al. (1994). "Regulation of transfer genes of promiscuous IncP alpha plasmid RK2: repression of Tra1 region transcription both by relaxosome proteins and by the Tra2 regulator TrbA." Microbiology **140 (Pt 11)**: 2981-90.

Zechner, E. L., F. R. Eisenbrandt, et al. (2000). Conjugative-DNA transfer processes. The horizontal gene pool bacterial plasmids and gene spread. C. M. Thomas. Amsterdam, Harwood Academic Publishers**:** 87-174.

Zhang, T. T., Z. T. Huang, et al. (2006). "High-throughput fluorescence polarization method for identifying ligands of LOX-1." Acta Pharmacol Sin **27**(4): 447-52.

Zhang, X., S. Zhang, et al. (2003). "Molecular handcuffing of the relaxosome at the origin of conjugative transfer of the plasmid R1162." Nucleic Acids Res **31**(16): 4762-8.

Zimmerman, R. K., F. L. Ruben, et al. (1997). "Influenza, influenza vaccine, and amantadine/rimantadine." J Fam Pract **45**(2): 107-22; quiz 123-4.

Zsoldos, Z., D. Reid, et al. (2006). "eHiTS: an innovative approach to the docking and scoring function problems." Curr Protein Pept Sci **7**(5): 421-35.

Zsoldos, Z., D. Reid, et al. (2007). "eHiTS: a new fast, exhaustive flexible ligand docking system." J Mol Graph Model **26**(1): 198-212.

Vita

Shuangluo Xia was born in Shanghai, People's Republic of China. He received his Bachelor of Engineering in Biochemical Engineering in Shanghai University in 1997. After graduation, he was enrolled in a joint graduate program between Shanghai University and Shanghai Institute of Biochemistry, Chinese Academy of Science; received his Master of Engineering in 2000. In 2003, he received his Master of Science in Computer Science at University of Texas at Dallas. In September of 2004, he entered the Graduate School of University of Texas at Austin.

Permanent address: 1620 W 6th street, Apt W, Austin, TX 78703

This dissertation was typed by the author.