2017

Selection of functional RNA aptamers against Ebola glycoproteins

Shambhavi Shubham

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Biochemistry Commons

Recommended Citation
https://lib.dr.iastate.edu/etd/16528

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Selection of functional RNA aptamers against Ebola glycoproteins

by

Shambhavi Shubham

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Molecular Cellular and Developmental Biology

Program of Study Committee
Marit Nilsen-Hamilton, Major Professor
W. Allen Miller
Eric Henderson
Drena Dobbs
Walter Moss

Iowa State University
Ames, Iowa

2017

Copyright © Shambhavi Shubham 2017. All rights reserved.
DEDICATION

I dedicate this dissertation to my mother Saroj Shrivastava for her constant support and inspiration.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
</tbody>
</table>

## CHAPTER 1: REVIEW OF LITERATURE

- Filoviruses ........................................................................................................ 1
- Filoviruses Pathogenesis .................................................................................. 2
- Ebola Virus Immune Evasion Strategy .............................................................. 3
- Viral Replication And Transcription ............................................................... 5
- Ebola Glycoprotein GP1, 2 ................................................................................ 7
- Ebola Glycoprotein GP1, 2 And Its Interaction with the Host Cell ................... 9
- Ebola Glycoprotein GP1, 2 Mediating Host and Cell Membrane Fusion .......... 11
- Ebola sGP .......................................................................................................... 12
- Ebola sGP Role In Pathogenesis ....................................................................... 14
- Ebola Therapeutics And Vaccine ..................................................................... 15
- Ebola Diagnostic Kits ....................................................................................... 16
- Aptamers ........................................................................................................... 17
- Methods of Selection In Aptamers ................................................................... 21
- High Throughput Sequencing And Data Analysis ............................................. 25
- Aptamers vs Antibodies And Its Market State ................................................ 27
Aptamers In Diagnosis Of Viral Infections...........................................29
Aptamers In Preventing The Fusion Of Virus Particle With Host Cell........31

CHAPTER 2: A 2’FY RNA EPITOPE FORMS AN APTAMER FOR
EBOLAVIRUS SGP: SELECTION AND CHARACTERIZATION.............60
Abstract.............................................................................................61
Introduction.......................................................................................62
Results...............................................................................................63
Discussion.........................................................................................67
Materials And Methods....................................................................71
Legends To Figures...........................................................................77
References.........................................................................................79

CHAPTER 3: SELECTING A FUNCTIONAL ANTI–VIRAL RNA APTAMER
AGAINST EBOLAVIRUS SURFACE GLYCOPROTEIN......................88
Abstract.............................................................................................89
Introduction.......................................................................................90
Results...............................................................................................92
Discussion.........................................................................................95
Materials and Methods.....................................................................99
Legends to Figures............................................................................104
References.........................................................................................106

CHAPTER 4: CONCLUSION AND DISCUSSION............................115
References.........................................................................................121
# LIST OF FIGURES

## CHAPTER 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Ebola Virus Particle</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2</td>
<td>GP1,2Δmucin X ray Crystallography Structure</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Prefusion and postfusion GP1,2 conformation</td>
<td>36</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Structures of full length GP1,2 and sGP</td>
<td>37</td>
</tr>
<tr>
<td>Figure 5</td>
<td>SELEX protocol to isolate RNA aptamers</td>
<td>38</td>
</tr>
</tbody>
</table>

## CHAPTER 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Next Gen Sequencing Analysis and Consensus Motif Identification</td>
<td>83</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Determining the affinities of 2’FY oligonucleotides against sGP</td>
<td>84</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Determining the binding epitope of sGP on 2’FY RNA aptamer</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Competition Binding Assay to determine the site of interaction of various aptamer sequences</td>
<td>86</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Specificity of 2’FY aptamer for sGP over serum proteins</td>
<td>87</td>
</tr>
</tbody>
</table>

## CHAPTER 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Comparison of top 10 sequences from SELEX experiments to select aptamers that recognize GP1,2Δmucin and full length GP1,2</td>
<td>111</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Selection through a sucrose cushion</td>
<td>112</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Filter capture assay to estimate the affinity of oligonucleotide 4789 for GP1,2Δmucin</td>
<td>113</td>
</tr>
</tbody>
</table>
Figure 4 Filter capture assay to determine the affinity of oligonucleotide 5185 and 5187 for GP1,2
ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Marit Nilsen-Hamilton for her support and guidance all these years. She has taught me how to be a good experimentalist and an independent researcher. Her ideas and knowledge made my Ph.D. experience productive and stimulating. I am thankful to my committee members Dr. Allen Miller, Dr. Drena Dobbs, Dr. Eric Henderson and Dr. Walter Moss for their time, guidance, and insightful questions. I would like to acknowledge Dr. Eric Underbakke for teaching me Mass Spectrometry and being a great mentor.

I would like to thank my lab manager Lee Bendickson for being very patient with me and teaching critical technical details essential for designing good experiments. Thank you to my past and present lab members who have contributed immensely to my personal and professional time at Iowa State University. They have provided me with great guidance and collaboration all these years and have been great friends as well.

My time in Ames was made enjoyable due to my friends, who provided me with a lot of love, support, and guidance. I am grateful for the time spent with them not just discussing science but also life.

Lastly, I would like to thank my parents and my sister for their love and encouragement. This journey wouldn’t have been possible without them. Thanks to my loving, supporting and patient husband Gunjan Pandey for standing by me through thick and thin. And most of all my very loving son Reyansh Pandey, he’s made my life complete and encourages me to be a better person every day. Thank You.
ABSTRACT

Ebola viruses (EBOV) cause severe disease symptoms in humans and in non-human primates in the form of viral hemorrhagic fever. Although EBOV outbreaks have not occurred in the U.S., the virus is of public health concern as a potential bioterrorism organism for which no vaccine or anti-viral is available. Fast-acting and prophylactic therapeutics are needed to reduce mortality due to outbreaks in other countries and for use in the U.S. if the virus is used in a bioterrorism attack. In view of the paucity of current antiviral therapies and diagnostic systems for EBOV, we provide an alternative solution by selection of high affinity 2’FY-stabilized RNA aptamers that bind the EBOV surface exposed glycoprotein, GP1 and soluble glycoprotein (sGP). Aptamers are single stranded short nucleic acid oligonucleotides with sequences that enable high affinity and specificity for their targets. Aptamers have comparable affinity with antibodies, but they are not immunogenic and are raised by in-vitro methods. They can be selected to bind to a precise region of a protein. By this means, an aptamer binding the EBOV surface GP1 would prevent the interaction between the virus and the host cell, disrupt the viral life cycle and an aptamer against sGP can be integrated with a detection platform that can be used as a biosensor to detect Ebola infections.
CHAPTER 1

INTRODUCTION

Filovirus Overview

Filoviridae family is composed of three genera: Ebolaviruses, Marburgviruses and Cueva viruses. There are two species of Marburgvirus, Marburg virus (MARV) and Ravn virus (RAVV) and five species of Ebolavirus, Ebola virus (EBOV), Sudan virus (SUDV), Tai Forest virus (TAFV), Reston Virus (RESTV) and Bundibugyo virus (BDBV) [1]. Ebola virus is an enveloped, non-segmented and negatively stranded RNA virus with genome of roughly 19kb. Ebola virus particle is filamentous in shape and genome encodes seven structural proteins beginning from the 3’ end: the nucleoprotein (NP), viral protein (VP35), VP40, the glycoprotein (GP), VP30, VP24 and the RNA dependent RNA polymerase (L). The viral RNA genome is encapsulated in NP that, along with VP35, VP30, and L form the replication complex [2] (Figure 1). VP 40 is the major matrix protein and is sufficient for budding of virus-like particles (VLPs), VP24 is a minor matrix protein [3, 4]. In addition to these proteins, Ebola virus secretes into the blood stream soluble glycoprotein (sGP). It is a nonstructural, secretory glycoprotein, which shares a 295-amino acid sequence with the glycoprotein GP. Ebolavirus infection can cause hemorrhagic fever with mortality rates as high as 90% [5]. The most recent outbreak was in 2014 when the incidence and prevalence of Ebola infections exceeded any previous outbreak [6]. With its ability to cause severe pathogenicity and a high mortality rate, the potential of being used for bioterrorism and the absence of any licensed vaccines or therapeutics underscores the threat, this virus poses to public health [7].
Filovirus Pathogenesis

Filoviruses cause hemorrhagic fever in human and non-human primates and exhibit high rates of mortality reaching up to 90%. Initial infection occurs through direct transmission from bats which are believed to be the host reservoir [4]. Ebola virus enters the host through mucosal surfaces and breaks through the skin. Most human infections have been caused by direct contact with the human patients or cadavers or infected animal carriers. Infectious virus particles or viral RNA have been detected in the semen and genital secretions of infected patients [4]. The known infectivity titers in non-human primates have been reported to be in the range of $10^7$ to $10^8$ plaque forming units/g [8], any exposure though the oral route can be associated with very high infectious doses. Transmission of the Ebola virus through an aerosol route has not been documented and is thought to be rare. Ebola virus, once it gains entry into the body, targets a range of cells early during infection, including monocytes, macrophages and dendritic cells, and then spreads to a variety of other cell types, including epithelial cells in the visceral organs [10]. Transport of infected immune cells and free virus through the bloodstream is believed to enable the virus to spread from the site of initial infection. The high lethality of Ebola viruses in large part is due to an early and strong inhibition of the innate immune response [20].
Ebola Virus Immune Evasion Strategies

**Phosphatidyl serine dependent**

Ebola virus particles are filamentous in shape and can be up to a micron in length, which would make it difficult for them to be taken in by the cells by clathrin or caveolin mediated endocytosis. Clathrin mediated endocytosis is usually limited to virions of 100nm whereas caveolin is around 50nm. It could be for this reason that Ebola virus is taken up by macropinocytosis. Phosphatidylserine is a lipid found on the outer leaflet of apoptotic cells where it alerts phagocytic cells to engulf the apoptotic cells by macropinocytosis. In a similar fashion, phosphatidylserine on the Ebola virus particles is recognized by the (T-cell immunoglobulin and mucin domain 1) TIM-1 receptors, interaction between TIM-1 receptor and phosphatidylserine mediates the entry of virus into the host cells. Ebola virus acquires phosphatidylserine from the plasma membrane microdomains, called lipid rafts, during the budding process. Lipid rafts are highly enriched for phosphatidylserine on the external leaflet[11]. The Ebola virus particles present themselves to phagocytic cells as apoptotic bodies and avoid initiating the inflammatory response by being engulfed by the cells. [12, 13].
Preventing the interferon response

VP35

VP35 is a multifunctional protein, which is associated with the ribonuclear complex and links the viral RNA polymerase (L) to the nucleoprotein (NP) during viral replication. VP35 also inhibits α and β interferon responses by multiple mechanisms. One mechanism is mediated by inhibiting RIG-I and MDA-5, innate pattern recognition receptors that detect foreign cytoplasmic RNA. RIG-I recognizes 5’-triphosphates of the blunt end of dsRNA and MDA-5 senses long dsRNA. RIG-I and MDA-5 associates with a common downstream adaptor IPS-1, which activates NF-kB, IRF-3, and IRF-7 to trigger the expression of type I interferon and pro-inflammatory cytokines. VP35 is a decoy substrate for IKKε/TBK-1 kinases and impedes the RIG-1 pathway by binding the N-terminal kinase domain of IKKε and prevents IRF-3 phosphorylation. Interactions of IKKε with other proteins such as IRF-1 and IPS-1 are also prevented due to binding of VP35 to IKKε [14]. These interactions successfully block the activation of genes with interferon responsive promoters [15, 16]. In a second mechanism, VP35 competes with RIG-I by binding to the blunt end of dsRNA, which prevents RIG-I and MDA-5 responses [17-19].

VP24

Typically, when innate immunity is functional, the host response to a virus infection is secretion of interferon proteins to generate antiviral responses in neighboring cells, signal hematopoietic cell responses and increase antigen presentation. Secreted interferon binds to type I and II interferon receptors, activating signaling via adaptor proteins, which results in the
phosphorylation and dimerization of the signal transducer and activator of transcription (STAT) proteins. These dimerized transcription factors are transported to the nucleus where they bind to interferon response elements and induce antiviral gene expression [20]. Since, these pathways are so important in fighting viral infections, viruses usually target them. Karyopherin-α1 is the nuclear membrane localized signal receptor that is responsible for nuclear accumulation of the STAT-1 dimers. VP24 interacts with karyopherin-α1 to inhibit the nuclear accumulation of STAT-1 and block interferon signaling. This mechanism prevents antiviral effects in cells expressing Ebola virus [21].

**RNA stability and replication**

The heterogeneous nuclear protein complexes, C1/C2 proteins ((hnRNP C1/C2), reside in the nucleus where they bind poly-U mRNA and assist in splicing before they are transported into the cytoplasm. They also engage in cap independent, IRES (Internal Ribosome Entry Site) dependent translation in the cytoplasm during mitosis [22]. Ebola VP24 causes relocalization of hnRNP C1/C2 from the nucleus to cytoplasm. It is hypothesized that C1/C2 might bind the poly-U tracts in Ebola viral RNA, thereby stabilizing it against degradation and enhancing genomic replication [22]. This suggests that Ebola virus uses the host machinery to prolong the half-life of its viral RNA and optimize transcription from its genome and viral replication.

**Filo viral Replication and Transcription**

Filo virus replication and transcription happens in the cytoplasm of the infected cell, the virus encodes its own RNA-dependent RNA polymerase that uses the negative-strand viral
RNA as a template for transcription and replication. Nucleocapsid assembly for the virus is mediated by three proteins NP, VP35 and VP30. The rate-limiting step for nucleocapsid formation is nucleoprotein (NP) assisted formation of helical bundle. [23]. This is followed by interactions with VP35 and VP30 and L to produce the ribonucleoprotein that encapsulates the viral genome. [24, 25]. For budding and formation of a virion, VP35 interacts with capsid protein VP40 and mediates the packaging of the nucleocapsid into the mature virion on the cell surface [26].

Replication of the RNA genome is initiated by the synthesis of positive sense intermediate (antigenome). The antigenome is the reverse complement of the RNA genome and it serves as a template for the generation of negative stranded RNA. Studies with Marburg and Ebola virus minigenome systems have established NP, VP35 and L as the minimal essential components for viral replication [27, 28][29].

Transcription process begins after the viral RNA is released into cytoplasm, the polymerase complex starts to transcribe it. Viral genome available for transcription is oriented in 3’ to 5’ direction, due to the polymerase complex stopping and reinitiating at each gene junction, transcription of the starting gene NP is favored. Due to sequential transcription, a gradient of transcribed gene products is generated. This leads to a non-homogenous distribution of viral proteins [30]. During transcription the viral RNA is polyadenylated as well as capped for cap-dependent translation of the viral mRNA [31].

Little is known about the mechanism of switching from the transcription to the replication state but its hypothesized that the phosphorylation state of VP30 might play a
critical role in deciding if the polymerase complex will support replication or transcription [32].

Viral assembly and budding are mediated by VP24 and VP40. To prepare the virus for budding VP24, one of the matrix proteins, associates with the RNP complex and functions as a signal to switch from the replication-transcription state to the viral assembly state [33]. These mechanisms ensure that the machinery for viral transcription and replication is tightly regulated.

**Ebola Virus Glycoprotein GP1, 2**

The fourth gene from the 3’ end of the filovirus genome encodes the viral envelope GP1, 2 (a glycoprotein). Viral glycoprotein (GP1) is a heterotrimer that forms spikes on the surface of the virus and is the only protein on the EBOV viral particle surface. Folding and assembly of GP1 occurs independently of other viral proteins [34]. It is initially synthesized as a precursor that is later cleaved by furin, a proprotein convertase, into GP1 (140kD) and GP2 (26kD) polypeptides that are linked by disulfide bonds [35]. The GP2 subunit contains two heptad repeat regions that facilitate assembly of GP into trimers, a transmembrane anchor sequence, and the fusion loop [34]. The heterotrimer (GP1 and GP2) assembles as a 450 KDa complex on the surface of the nascent virion [36]. The GP1 subunit contains the cell surface receptor binding site and a heavily glycosylated mucin domain. The residues responsible for mediating interactions between GP1 and the host cell are within the 230-residue N-terminal domain of GP1.
Deletion of the C-terminal mucin domain enhances viral infectivity \textit{in vitro} [37-39]. This observation is explained by the crystal structure of EBOV GP (Figure 2), in which the receptor binding site of GP1 is masked by a glycan cap created by the mucin domain [36]. The GP1 subunit responsible for interaction with the cell contains four different domains: base, head, glycan cap and mucin like domain (MLD). The first three domains of GP1 are important for expression and function of the prefusion state of the glycoprotein. The GP1 base forms a hydrophobic, semicircular surface that interacts with the internal fusion loop and heptad repeat of GP2, thereby acting as a clamp that holds the prefusion GP2 form. The head part of GP1 is believed to contain the receptor binding domain (RBD) with a significant portion of this domain exposed to the solvent.

The structure of the GP2 subunit, which is responsible for the fusion of viral and host membrane, includes a hydrophobic internal fusion loop, two heptad repeats (HR1 and HR2) and a CX6CC disulfide bond motif, membrane proximal external region and a transmembrane anchor [40]. The internal fusion loop is wrapped around the outside of the GP trimer in the pre-fusion state, with its hydrophobic side chains sequestered from the solvent due to packing into a GP1 monomer.

Although Ebola virus species are antigenically distinct, the GP1 RBD is highly conserved among them. The receptor binding domain of the Ebola viruses is masked by the highly glycosylated domains of the mainly N-glycosylated glycan cap and a mucin like domain, which is both N- and O-glycosylated [41]. Glycosyl groups protect proteins from degradation but can also obstruct interaction of the glycosylated protein with another molecule. Removal of the 7 glycan sites in the GP1 core had no significant effect on GP expression but
increased transduction compared with the wild type. Similar results were observed when the mucin domain lacked its N-glycosylation [41]. These results demonstrated that the N linked glycosyl groups on GP1 are not required for viral entry in Vero cells [41].

**Ebola Virus Glycoprotein (GP1,2) Interaction With The Host Cell**

Ebola viruses infect a broad range of cells and a host of proteins are known to enhance viral entry into the host cells including the C-type lectins L-SIGN, DC-SIGN and the tyrosine kinase receptor Axl [42-45]. Infectivity involves at least two recognition events 1) at the host cell surface that mediates endocytosis and 2) in the endosome to mediate cytoplasmic entry. The cell surface receptor for EBOV is believed to be the T-cell immunoglobulin and mucin domain 1 (TIM-1) receptor [46]. The usual function of the TIM-1 receptor is to bind to phosphatidyl serine (Ptd-Ser) on the surface of apoptotic cells and thereby facilitate phagocytosis of these cells [47]. It has been shown that TIM-1 facilitates the entry of Dengue Virus by directly interacting with the virion-associated phosphatidylserine [48]. Phosphatidylserine is the most highly represented anionic lipid in inner leaflet of plasma membrane, where it usually is around 15-20% of anionic lipids [49]. Ebola Virus also gains entry through interaction of the TIM-1 cell surface receptor and Ptd-Ser residue on the viral envelope. The TIM-1 receptor enhances the internalization of pseudovirions that do not express GP1, 2, confirming that TIM-1 and Ptd-Ser mediate the interaction for the viral entry into the cell. [50]. After its initial interaction with the cell surface receptor, the virus enters the endosomes through macropinocytosis. When the endosomes containing the virus particles mature to late endosomes, the viral GP is further processed and it then interacts with the
endoosomal receptor. However, although TIM-1 and DCSIGN have been identified as cellular receptors for viral entry, the expression of TIM-1 and DCSIGN is not sufficient to make certain cells permeable to Ebola virus entry. This result suggests that there may be distinct mechanisms for Ebola mediated entry depending on which attachment factor is used for entry [51, 52]. Ebola virus requires endosomal processing before the final entry of the virus particle into the cell. The protease, cathepsin B initiates entry which cleaves off the mucin domain and glycan cap to expose the RBD [53-56].

The endosomal receptor for EBOV is believed to be the Niemann-Pick C1 (NPC1), a lysosomal cholesterol transporter [57]. This conclusion is consistent with the observation that Npc1−/− mice demonstrated significantly reduced viremia in comparison to wild type mice when tested in a mouse model of Ebola disease [58]. NPC1 is a 13-pass transmembrane protein found in the membranes of late endosomes and lysosomes of all cells [59]. It works together with another protein, NPC2, to mediate the transport of cholesterol between cellular compartments [60]. Binding site for C1 (NPC1) is exposed after the endosomal cysteine proteases cleave EBOV GP1 to remove the heavily glycosylated C-terminal sequences, which generates an entry intermediate comprised of the N terminal regions of GP1 and GP2 [54, 57]. The importance of the identified C1 (NPC1) binding site on GP1 was demonstrated with virions that contained mutations in the C1 receptor binding residues, which are vital for viral entry and for maintaining the conformation of the glycoprotein [57]. Recent studies have shown that TIM-1 and NPC1-mediated interaction is responsible for viral fusion in the late endosomes. Although, TIM-1 is primarily expressed on the cell surface, it is also found in early endosomes and can cycle between the endosome and the cell surface [61]. A monoclonal
antibody (MabM224/1) against the TIM-1 receptor inhibits the internalization of Ebola virus in Vero6 cells, which express the TIM-1 receptor. Thus, it is postulated that TIM-1 interacts with the Ptd-Ser on Ebola virus, is taken into the endosome, and along with the cleaved fragment of GP1, interacts with NPC-1 and initializes viral fusion with the late endosome lipid membrane [62].

**GP2-Mediated Viral and Host Membrane Fusion**

The process of fusion of the viral and host cell membranes has a very high energy barrier. To lower the activation barrier, viral proteins must undergo huge conformational changes to generate sufficient released free energy. The first step for viral-host cell membrane fusion in the endosome is to prime the envelope GP to a metastable conformation that can be subsequently used for viral fusion. The crystal structure of EBOV GP1,2Δmucin (mucin domain deleted) illustrates that the hydrophobic fusion residues are within the antiparallel β scaffold and the structure is not conformationally restricted [63]. The N terminal of the fusion loop has conformational freedom and is believed to aid fusion. In influenza and flaviviruses, the main trigger for fusion is low pH, which results in protonation of histidine residues located near positively charged residues in the prefusion state. These protonated histidines are found in the post fusion conformation engaged in electrostatic interactions with the negatively charged residues [64, 65]. The trigger for Ebola virus fusion is not well defined. Although the low pH of the endosome is required for the cleavage of GP1,2 and for its subsequent interaction with NPC1, the change in pH due to endosome maturation is not involved in triggering fusion. However, it is unknown as yet if the cleavage of GP triggers the fusogenic conformation of the
virus [54] or the cleavage of GP by cathepsin B triggers another cellular factor that initiates the fusion [55]. Even though the fusogenic trigger is unknown, it is evident from the crystal structure that the GP2 hydrophobic patch, which clamps around GP1, must be released before fusion. Comparing the pre- and post-fusion conformations of GP1,2 and GP2 respectively, shows that the heptad repeat segments of GP2 unwinds from its prefusion state of a ring around GP1, straightens and assembles into a 44 residue long helical rod like structure [63, 66, 67]. In response to the rotational and translational movement of heptad repeat, the internal fusion loop arranges itself to the top of the trimeric GP2 and adopts a 3<sub>10</sub> helical conformation, which is stabilized by membrane interaction with hydrophobic residues in the fusion loop that mediate penetration into the host membrane [68]. After formation of this intermediate conformation, GP2 folds on itself bringing the two heptad repeats HR1 and HR2 close together to form a six-helix bundle. This collapse of the two GP2 heptad repeats brings the host and viral membranes closer causing the two bilayers to merge into a hemifusion stalk that eventually opens to a fusion pore. (Figure 3).

**Soluble Glycoprotein (sGP)**

GP gene in the Ebola virus genome produces three different sized proteins: full length 676 aa GP1,2, pre-sGP which is 364 aa in length and a 298 residue small secreted protein (ssGP) [69, 70]. These three proteins are produced because of transcriptional slippage of the GP gene. GP1, 2 and ssGP are produced when slippage of the viral polymerase occurs at seven consecutive uridines within the GP gene. With slippage of one U, an additional adenosine residue is incorporated into the mRNA, which results in two sequential open reading frames
becoming a continuous reading frame that encodes full length GP1,2. sGP is produced from the unedited version of the mRNA in which there is a stop codon at the end of the first reading frame. Slippage along two Us and incorporation of two extra adenosines results in the synthesis of ssGP, a secreted homodimer [69, 71]. The ratio of transcripts of these three proteins have been shown to be (sGP:GP1,2:ssGP) of 71:24:5 in Vero E6 cells [71] suggesting that sGP is the main viral protein product of the GP gene. Although the time of release of sGP into the blood stream is unknown, sGP in 15 µL of serum, was readily detected by western blot in comparison to GP1,2, which could not be detected [72].

Soluble glycoprotein (sGP) is encoded by the GP gene in all species of Ebola Virus. It is initially synthesized as pre-GP which is shunted into the golgi complex via the signal recognition complex. The sGP precursor undergoes post translational cleavage at the C-terminus to yield the mature form of the protein [73]. After its proteolytic cleavage it forms a 103 KDa homodimer linked by two disulfide bonds at Cys 53 and Cys 306 [74, 75] with six N-glycosylation sites [75]. sGP shares the first 295 amino acids with GP1, 2. A recently determined cryo-EM structure of sGP showed that the monomeric structures of GP1,2 and sGP are similar [Figure 4] [76]. Co-expression of sGP and GP2 by HEK293T cells resulted in an sGP:GP2 protein complex that was recognized by the known neutralizing antibody KZ52 [77]. KZ52 is an antibody that was isolated from an Ebola survivor that binds to the interface of GP1 and GP2 [63]. Using Circular Dichroism, Fluorescence Spectroscopy and MALLS it was demonstrated that sGP, which is secreted as a disulfide linked homodimer, is mostly beta sheet with 15% alpha helix content, and is highly stable at 37°C for up to 72 h [78].
Role Of sGP In Ebola Pathogenesis

sGP viral protein is secreted in large amounts by HEK293T cells and is present in the blood of infected individuals [77], due to its high concentrations in blood various studies were performed to determine its role in Ebola pathogenesis. One study suggested that sGP could inactivate neutrophils by the binding to CD16b, which is a neutrophil specific Fc receptor III [79]. However, these studies were challenged because the initial study overlooked the fact that sGP could bind the neutrophil receptor indirectly through the Fc region of anti-sGP antibody [80]. A later study confirmed that sGP immune complexes bound CD16b by demonstrating that sGP complexed with an Fab fragment binds to neutrophils [81]. These results argued against the potential role of sGP as a neutrophil inactivator.

Another potential role of sGP was to promote apoptosis of B and T cells during viral infection. Earlier reports had shown that lymphocyte apoptosis during Ebola infection could be mediated through Fas/FasL interactions [82]. But, later it was demonstrated that sGP alone or in combination with Fas ligands had no effect on frequency of Jurkat T cell death [82]. Investigators concluded that sGP does not induce apoptosis via extrinsic pathway. Thus, the current evidence supports the conclusion that sGP is not involved in stimulating apoptosis of lymphocytes.

A possible function of sGP in vascular dysregulation was also investigated, because the expression of cellular adhesion molecules had earlier been shown to increase in the presence of Ebola GP1,2. But, sGP and its degradation products failed to activate expression of adhesion molecules in endothelial cells or to alter the integrity of endothelial cell barrier [83]. Instead, when endothelial cells were treated with sGP and TNF-alpha (an inflammatory
signal that decreases endothelial barrier function), sGP restored barrier function, which suggested an anti-inflammatory role for sGP [83]. Further studies explored the relationship between the structure of sGP and its anti-inflammatory role. Mutations of Cys53Gly, Cys306Gly, and the double mutant failed to restore the barrier function. Therefore, a fully functional sGP homodimer is required for the protein to restore barrier function [75]. Overall, these studies show that Ebola sGP might not have a role in the vascular deregulation but it might have an anti-inflammatory function.

**Immune modulation of host responses by sGP**

Antibodies in the sera of human Ebola hemorrhagic fever survivors cross reacted with sGP rather than GP1,2 [84]. Because sGP and GP1,2 are structurally similar they can be bound by antibodies that recognize shared epitopes. However, due to the higher rate of production of sGP in comparison to GP1,2, B cells that recognize sGP specific epitopes or shared epitopes outcompete the B cells that recognize GP1,2 specific antibodies. This is believed to be the reason for the biased population of antibodies that target sGP specific epitopes and GP1,2 shared epitope. Thus, the presence of high concentrations of sGP in the blood directs antibody production towards sGP, which acts as a decoy to allow the viral infection to continue unabated [85].

**Vaccine and Therapeutics For Ebola**

Until the 2014 Ebola outbreak no vaccines or therapeutics were available because the infected population was very small and few companies were interested in the market. However,
with the magnitude of the outbreak in 2014, various vaccine formulations were proposed and clinical trials were initiated. Clinical trials with rVSV-ZEBOV vaccine found that it was 70-100% effective in providing protection therefore can be used in future outbreaks [86]. ChAd3-ZEBOV, developed by GlaxoSmithKline in collaboration with the US national institute of allergy and infectious diseases is still in clinical trial. The Novavax recombinant Ebola vaccine is currently undergoing Phase II clinical trial. Other products in development include an oral adenovirus (Vaxart), a different vesicular stomatitis virus candidate (Profectus Biosciences, recombinant protein (Protein Sciences), DNA vaccine (Invovia) and a recombinant rabies vaccine from (Jefferson University) [87].

Apart from the vaccines, there are many candidate therapeutics being evaluated including small molecule inhibitors [88, 89] and siRNA based therapeutics [90]. ZMapp is a cocktail of humanized antibodies that target three different sites on the Ebola glycoprotein [91]. These three antibodies are produced by genetically modified tobacco plants and clinical trials for Zmapp were concluded in 2016 [92]. The drug was well received by patients; however it did not produce definitive results with the small number of patients in the trial. Therefore, the FDA allowed Mapp Biopharmaceuticals to make Zmapp available to a larger number of patients to obtain definitive answers.

**Diagnostic Kits for EBOLA**

WHO reported that, between 15th Feb 2015 and 19th April 2015, more than 26044 people were infected and around 10808 died from Ebola Virus [93]. As just discussed, several potential drugs and vaccines are being tested, yet no FDA-approved treatment has yet been
commercialized. To minimize spread of the virus a means of early detection of Ebola virus is needed so that infected individuals can be quarantined. By 2014 WHO had received 17 applications from diagnostic companies for devices to detect Ebola virus [94]. These included reverse transcriptase polymerase chain reaction (RT-PCR) methods for detecting genomic sequences, immunochromatographic strip formats and ELISAs. RT-PCR is the standard diagnostic method for Ebola virus and is highly sensitive. However, it requires trained personnel and specialized equipment in addition to high level biocontainment measures. In addition, RT-PCR involves several steps during PCR amplification and requires constant monitoring of the temperature. A constant power source is needed to limit the assay to specific laboratories. The WHO approved a gold strip procedure for detecting Ebola virus. It is less sensitive than RT-PCR and therefore would not detect an Ebola infection immediately after the symptoms appear. It also has a 15% false positive rate, which limits its application as a diagnostic kit. Current methods for Ebola detection dependent on the availability of necessary reagents, require close contact with infected samples, and a constant supply of power. This motivates interest in developing a robust, sensitive, and in-house diagnostic kit.

Aptamers

The term aptamer is derived from the latin word “aptus” which means “fit” and in Greek the word meros, which means “particles”. Aptamers are short oligonucleotides with sequences that enable them to bind to their selected targets with high affinity and specificity. Aptamers have access to a repertoire of tertiary structural elements, including bulges, loops, pseudoknots, and hairpin loops. Due to their varied structures and their capability to snugly
bind to the target molecule, the aptamer targets range from prokaryotic/eukaryotic cells, viral particles, small molecules, and proteins. The interactions between the aptamer and target consist mainly of a combination of stacking of aromatic rings, electrostatic and van der Waals interactions and hydrogen bonding [95, 96].

SELEX (Systematic Evolution of Ligands through Exponential enrichment) is a process that involves progressive siphoning of oligonucleotides (aptamers) through iterative rounds of partitioning and amplification [97]. A randomized pool of RNA or ssDNA is incubated with target molecules under specific buffer conditions. Oligonucleotides that bind the target are separated from the unbound and are later amplified. The process is repeated until the pool randomness is significantly reduced [96 168].

**SELEX**

**Library Generation**

The first step in SELEX process is to generate a randomized sequence oligonucleotide library, which consists of ssDNA oligonucleotides comprised of a 20-60 nt random region flanked by constant primer regions to be used for amplification. For selection of DNA aptamers the synthesized pool can be directly used. Whereas for RNA aptamer selection, ssDNA pool is converted to dsDNA and modified to include a T7RNA polymerase promoter using a primer. The maximum number of random sequences in the library can be determined using $4^n$ with $n$ being the number of positions in the random sequence. For our SELEX experiments we used 53 nucleotides in the random regions which translates to $\sim 10^{32}$ possible sequences. However, our starting pool size includes only $10^{15}$ oligonucleotides. It has been reported that the
probability of an aptamer for its target ranges from $10^{-13}$ to $10^{-14}$ with mean/mode probability of $10^{-11}$ [98]. Therefore, a pool of $10^{15}$ complexity is frequently used for aptamer selection. Things that need to be considered in the pool design are pool length, which will determine the complexity of the pool, and nucleotide composition of the random sequences and constant regions. Long oligonucleotides can be expected to adopt multiple secondary structures, which might impede the correct folding of the desired binding motif. From studies to determine the optimal length of the random region in oligonucleotide libraries it was suggested that longer random pools (70-200nts) should be used to select for ribozymes, whereas shorter oligonucleotides (<70nt) would serve the purpose for selecting aptamers [99]. Experiments with pools containing different lengths of random region (16, 22, 26 50, 70 and 90) were used to select for an aptamer that binds isoleucine, results showed that the optimum pool length for isolating the desired aptamer was 50-70 nt [100]. A second important factor in designing pools is the complexity of the pool. This depends on the ratio of the four nucleotides used to create the ssDNA library, which can be determined in the manufacturing process [101, 102].

Chemically modified nucleotides can be used to introduce new features to aptamers such as enhancing the sites of interaction with the target, improving the aptamer’s structural stability or providing enhanced nuclease resistance [103, 104]. Two approaches can be used to generate modified nucleotides. The first is to incorporate modified nucleotides at the time of pool generation. However, this approach could be a problem if DNA or RNA polymerase will not recognize the desired modified nucleotide as substrate. A second option is to modify the already selected aptamer. However, a modification to an already selected aptamer might alter the binding affinity or completely abolish binding [102].
Different techniques have been utilized to obtain modified aptamers conforming to a desired function. For example, 2’F and 2’NH2 modifications in the 2’ ribose have been included to improve the aptamer structural stability and nuclease resistance [105, 106]. Nucleotides containing photoactivable functional groups (e.g. 5 iodo,5-bromo and 4 thiouridine), which can covalently crosslink with a target protein, have also been used to increase the affinity and specificity of the aptamers for their targets [107]. Modified aptamers containing fluorescent groups have been used to study the binding of the aptamer to the protein [108]. Amino acid side chain, like groups on 5-benzylaminocarbonyl-dU, 5-napthylmethylaminocarbonyl-dU, 5-tryptaminocarbonyl-dU and 5-isobutylaminocarbonyl-dU, confer a significant increase in chemical diversity, thereby increasing the chances of interaction with proteins. For example, SOMAmers (Slow off rate modified aptamers) are ssDNA aptamers with modified uracils and 5-methyl-dC [109].

Another important aspect of pool design is the design of constant regions. The sequence of the constant region affects the amplification efficiency during PCR and RT (reverse transcription) in case of RNA. The constant regions should be designed so as to not form stable secondary structures [100]. Primers corresponding to the constant regions should also not form strong secondary structures or include sequences that can cause primer dimers to form. To improve amplification efficiency, the 3’ primer end should have WSS (W=T or A and S= G or C) because polymerases can extend well with ACC at the 3’ end. A/T rich regions at the 3’ end can be used to avoid mispriming [110].
The selection process

SELEX process (Figure 5) starts with the incubation of the target with a pool of oligonucleotides with random central sequences followed by a selection step in which the bound oligonucleotide is separated from the unbound using a partitioning method. After this separation for an RNA aptamer the selected RNA is reverse transcribed and further amplified through PCR. In case of selection for a DNA aptamer the bound ssDNA is PCR-amplified. Multiple rounds are performed in succession, which involve in vitro transcription for RNA aptamer selection and separation of single stranded DNA for DNA aptamers. After multiple rounds have been completed, the pools are analyzed by Next Gen Sequencing or Sanger sequencing and evaluated for their ability to bind to the target with high affinity and specificity.

Aptamer Selection Methods

Filter capture

The filter capture method has been used extensively for selection of aptamers. The principle factor in the method is the nonspecific binding of the target protein to the nitrocellulose filter with the oligonucleotides passing through. Oligonucleotides bound to the protein are trapped on the filter. This method has been widely successful but has the following limitations 1) the efficiency of protein capture varies with the protein, 2) small molecules cannot be captured on the membrane, and 3) certain aptamers bind to nitrocellulose membranes in the absence of their protein target [111]. The filter capture method relies on the true equilibrium binding of aptamers to the target protein in solution in comparison to other methods in which either the target or the aptamer is immobilized on a solid support [112]. In
all these methods, including the filter capture method, the washing steps that follow the partitioning can distort the equilibrium achieved in binding if the complex is characterized by a high off-rate. However, the washing step is the shortest in the filter capture method compared to others and thus least likely to influence the outcome. The filter capture method is also widely used for evaluating the binding of the selected aptamers to the target protein [113].

**Bead based SELEX**

Immobilization of target molecules on a solid support is another method for selection [114], various affinity tags (6X his, GST and MBP) can be used, in addition to coupling chemistries such (amine, thiol or carboxyl). In bead based selection, target molecules can either be pre-immobilized then incubated with the target library or the molecules can be incubated with the target library and then captured by the affinity column. Various versions of affinity chromatography have been used for selections such as 1) MEDUSA (Microplate based enrichment device used for the selection of aptamers) [115], 2) the particle display method [116] in which emulsion PCR is used to immobilize $10^5$ copies of a single clonally amplified ssDNA aptamer on each bead, the beads then incubated with fluorescent labeled target molecules and sorted by flow cytometry, and 3) MonoLEX, which has been shown to generate ssDNA aptamers in a single round that bind whole vaccinia virus particles [117]. The limitations of Bead-based SELEX are the restrictive interaction surface, requirement for electronic instruments and flow pumps and increased non-specific interactions due to density-dependent cooperative target binding [118].
Electrophoretic SELEX

EMSA (Electrophoretic Mobility shift assay) has long been used for downstream applications during SELEX, mostly for evaluating the binding of potential aptamers to the target protein. However, this method has also been used for aptamer selection [119]. Although this method offers advantages such as equilibrium binding and separation of the bound and unbound populations, it can be used with only some proteins and under limited buffer conditions. As well, the volumes that can be accommodated in the gels used for separation are too small to include the number of oligonucleotides required for the maximum possible sequence diversity for SELEX. Therefore, it is wiser to use this method for downstream applications.

A modification of EMSA is capillary electrophoresis which also relies on the electrophoretic mobility of the aptamer, but separation is in the capillary as opposed to the gel [120]. However, this method is limited to molecules that are within a certain size bracket and that are charged so they can be resolved on the capillary.

Microfluidic SELEX

In the recent times, microfluidic SELEX has been used for selection of aptamers. There are two different kinds of platforms 1) the target molecule is immobilized on micromagnetic beads [121] and 2) the target molecule is encapsulated in sol-gel, gel like porous silica material [122]. Microfluidic platforms allow fewer rounds of selection due to reduction in the available target molecule along with function to perform continuous washing. This continuous washing step allows removal of weak and nonspecific binders [123]. Since they are small, the use of
reagents is considerably reduced for microfluidic methods and the sol-gel based microfluidic system can be utilized to select against multiple targets in one experiment. Even though the advantages of this method are many, the usage of this method might be limited due to the requirement for fabricated devices and electronic instruments. In sol gel based platform there is also the chance of undesired effects of a chemical reaction during sol-gel formation on the target molecule.

**Cell SELEX**

Cell-SELEX is used to select for aptamers against cells or cell fragments rather than targeting a specific protein. In all SELEX procedures, negative selection against proteins or cells other than the desired biomarker helps in gaining specificity towards the targeted biomarker. Cell SELEX has been used to identify aptamers that can differentiate between normal or diseased cell types [124, 125] and therefore the negative selective steps are generally against normal cells. Cell SELEX has been used to select aptamers with therapeutic potential. In terms of experiments, positive and negative selections can be done under conditions of steady state equilibrium binding. Due to the possibility of nucleic acid aptamers being either degraded or internalized during selection, Cell-SELEX protocols usually employ shorter incubation periods than necessary for reaching equilibrium. Since cell surfaces also display non-target cellular proteins, it is difficult to achieve very high specificity towards the target molecules. TECS SELEX addresses this problem by overexpressing the desired target protein for selections. An aptamer that recognizes the TGF-β type III receptor with nanomolar affinity was successfully selected using this method [126]. Another complication of Cell-SELEX is
that it is prone to artifacts resulting from the presence of dead cells in the selection population. Cell-SELEX protocols use FACS to separate bound and unbound aptamers [127]. This enables sorting of live and dead cells along with selecting cells with bound aptamers [128].

High Throughput Sequencing and Data Analysis

Incorporation of high throughput sequencing into the SELEX protocol for the analysis of enriched aptamer libraries and identification of candidate aptamers has been the most informative recent change in SELEX applications. Previously, enriched aptamer pools were cloned into a plasmid and a few hundred individual clones were sequenced to identify high affinity aptamers. In contrast, high throughput sequencing analysis has a readout of over a hundred million sequences. Candidate aptamers can be selected at much earlier rounds by studying the enrichment of an oligonucleotide with a particular sequence across various rounds, whereas conventional cloning-based aptamer identification required multiple rounds until an ensured enriched population was observed [129]. Illumina platforms are a preferred platform for NGS as they offer a significantly higher number of sequence reads with read lengths, which would cover the random regions and also the constant regions of most libraries [130, 131]. A large number of reads helps in gaining a more comprehensive insight into the sequence and structural features of the aptamers selected and a broad perspective on the progress of each SELEX process. Information gathered through NGS has also helped to predict secondary structure of the RNA aptamer and to identify appropriate truncations that yielded a better binding aptamer [132]. Given the ability of NGS to sequence multiple SELEX rounds simultaneously, it can be used to monitor the selection process as it proceeds. For
example, in a recent study SELEX pools derived from a genomic library selected against the *E. coli* Hfq protein were run in parallel with a control SELEX (Neutral SELEX) that did not include the target during the selection steps and was subjected to the same PCR and RT treatments. The analysis of various rounds revealed that selection with the target progressed independently of the Neutral SELEX, with the former resulting in substantial enrichment of certain sequences whereas with the latter there was no enrichment of sequences [133].

It is widely accepted now that even highly enriched libraries contain many thousands of potential aptamer sequences, many of them present in the populations in lower numbers. It has also become more evident that the most abundant oligonucleotides obtained in the final SELEX pools might not have the sequences for the highest target affinities. Other than selection for affinity to the target, reasons for high abundance of an oligonucleotide with a sequence in a selected pool could be PCR amplification bias or affinity of the oligonucleotide with that sequence to the partitioning matrix. Therefore, the fold enrichment over one or more SELEX rounds rather than abundance is a much better determinant of high affinity aptamers [131]. Tracking fold-enrichment is a good way to identify aptamers especially if the starting library is biased.

The current capability of generating an immense amount of data and its contribution to better understanding the SELEX process has resulted in high throughput sequencing being used by most aptamer research groups to analyze the SELEX results. Despite its popularity, the development of specific bioinformatics tools for identification of candidate aptamers from the huge amounts of sequencing data is limited. Current bioinformatics tools based on clustering identify top clusters and the aptamer population can be traced to identify the
enriched sequences. Current tools also allow sorting of the putative aptamer sequences via their sequence structure motifs [134-136]. However, bioinformatics tools are still needed that better predict the target binding aptamer candidates based on structure or sequence classifications.

**Aptamers vs. Antibodies and The Current State Of The Market**

Antibodies, have been widely used for various purposes. Apart from being used in the lab for research, they are used as therapeutics drugs, are also functionalized on biosensors for detection purposes. Although, antibodies have been fundamental to several applications, there are limitations associated with its usability. Synthesis of an antibody is time consuming and there is always an issue of reproducibility between batches. Neutralizing antibodies have been traditionally used to prevent the spread of viral infections. But, no neutralizing antibodies have yet been commercialized for EBOV. Also, antibodies used in passive immunity must be “humanized” to prevent their immunogenicity. Thus, even if there were an effective anti-EBOV antibody available, it would be very expensive and time consuming to humanize the antibody. Antibodies due to their short shelf life, require proper storage conditions, therefore when used for a detection platform, conditions amenable for preserving antibodies is needed. In addition, to perform techniques like RT-PCR or ELISAs trained personnel is required. Usage of antibodies adds up to additional costs and require infrastructure, regions affected by Ebola viruses have poor infrastructure therefore storage of these drugs and kits require additional costs and efforts.

Aptamers have been established as a potent alternative for antibodies. Aptamers have been selected against several targets to function either as a therapeutic drug or as a biosensor.
Aptamers have notable advantages over antibodies that include non-immunogenicity, \textit{in-vitro} production, and stability. Aptamers can be rapidly reconstituted and the \textit{in vivo} pharmacokinetics and tissue distribution can be tuned by chemical modifications including conjugation to high molecular weight molecules such as PEG. Conjugation can increase the half-life of an aptamer from 40 min to 50 h. Modifications such as with fluorine in the 2’ position also increase stability in serum [137].

Even though aptamers were discovered in 1990 they still haven’t effectively transcended from research labs into the market full-fledged even though they hold many advantages over antibodies (Table 1). Only one aptamer-based product Macugen has cleared clinical trials and is available in the market. The huge investment of pharmaceutical companies in the antibody market may be partly responsible for the resistance against aptamers. Even though few aptamers are in the market, the research on aptamers hasn’t receded. A handful of companies have been developing and marketing aptamers as antibody-like binding reagents for research and development. These companies are Soma Logic (Boulder, CO, USA), OTC Biotech (San Antonio, TX, USA), Aptagen (Jacobus, PA, USA), Base Pair Biotechnologies (Houston, TX, USA), NeoVentures Biotechnologies (London, ON, Canada), Aptamer Sciences (Pohang, Korea). A few companies marketed aptamers as components of assay kits or concentrating devices, such as NeoVentures Ochratoxin A. Soma Logic, a company at the forefront of aptamer research, has been pushing its technology, SOMAmers and SOMAscan aptamer platform sensors, in the diagnostic market [138].
Proper diagnosis of any viral infection is important for preventing its spread. Symptoms of an acute infection are at first non-diagnostic and associated with a variety of viral infections. To prevent its spread, it is imperative that the infectious agent be identified in the early stages.

Aptamer development to date has been concentrated mostly on the well-known viral diseases such as HIV, Influenza virus, SARS and HPV [155-158]. Little research has been done on a wider variety of other viral diseases, including Ebola virus. Very few aptamers have been reported that bind Ebola viral proteins or are directed towards blocking an antiviral function [159, 160].

There is a wide range of aptasensor options designed based on different principles for detection of viral particles. An aptasensor was designed to detect avian influenza virus, based on the principle of quartz crystal microbalance [161]. Gold microelectrodes with impedimetric
properties allows one to distinguish between active and inactive forms of virus. Using specific DNA aptamers viable virus was detected by using the impedimetric sensor [162]. Other approaches to detect viral antigens include an RNA aptamer sensor to the HCV core antigen in which fluorescent dye (Cy3) conjugated RNA aptamer was the means of detection.[163]

In another study an aptamer to viral glycoprotein E2 was selected using CS SELEX and demonstrated to inhibit HCV infection in Huh7.5.1 cells [144].

**Comparison of aptamer based tests with other analytical methods**

Most current methods to detect virus infections rely on Enzyme linked immunosorbent assay (ELISA) or RT-PCR [161, 164, 165]. However, ELISA is reported to have low sensitivity and very high rates of false positives [166-168]. It also cannot be broadly applied because of the need for specific antibodies that are difficult to obtain for several viral diseases [169]. RT-PCR can be performed without application of antibodies and can detect the viral genome efficiently. It is also a sensitive method. Although the advantages are many for RT-PCR, it requires expensive enzymes and trained personnel to do these analyses.

Aptamers could be used to circumvent bottlenecks in the current detection systems for viral diseases. Properly developed and authenticated aptamers could efficiently differentiate between infected and non-infected cells.

On a Hydrogel coated QCM (Quartz crystal microbalance aptasensor) biosensor, anti-H5 antibodies were coated in parallel with anti-H5 aptamers. Although, same concentration of antibody was coated as the aptamer, the detection limit of H5N1 with anti-H5 antibody was 0.128U whereas with the aptamer the detection limit was 0.0128U. [161, 167]. In addition to
providing higher sensitivity, detection using aptamers takes less time than with antibodies or PCR. For example, to detect avian influenza virus the time for detection by ELISA or qPCR would be 3 hours and 5 hours respectively, whereas with an aptasensor it only takes 1.5 hours [165-168].

**Aptamers In Preventing The Fusion Of Virus Particle To The Host Cell**

Apart from diagnosis, another obvious function of an aptamer could be as a therapeutic. An aptamer could act as a therapeutic agent by preventing the interaction between the viral proteins and the host cell receptor proteins. Many aptamers have been selected that inhibit viral entry. An RNA aptamer (B40) that binds HIV gp120 inhibits viral interaction with T cell co-receptor CCR5. Application of the aptamer resulted in a decreased concentration of p24 antigen in the supernatants from virus-infected cultures of human peripheral mononuclear blood cells as measured by ELISA. Further characterization of the aptamer showed that it binds to the core conserved region of gp120 [170-172].

E2 protein is a potential target for creating blocking aptamers against HCV infection. E2 is a coreceptor of human CD81, which is presented on hepatocytes and B lymphocytes. ZE2, a DNA aptamer selected against, E2 blocked E2 binding of a wide range of HCV serotypes to Huh7.5.1, human established cells line of hepatocellular carcinoma cells. The presence of the aptamer reduced viral RNA levels was tested by qPCR. [144].

A DNA aptamer A22 to the influenza virus HA protein, blocked influenza virus entry into MDCK (Madin-Darby Canine Kidney). Cell viability increased in proportion to A22 aptamer concentrations from 50 to 100 pM. The therapeutic potential of this aptamer was also
demonstrated in an animal model in which mice infected with A/Texas/1/7 influenza strain and simultaneously treated with the aptamer A22, lost weight slower that the non-treated control group [173]. C7-35M, a DNA aptamer that binds to receptor binding region of hemagglutinin binds to the virus particles and inhibits the binding of the virus to the host cell receptors. [174].

Similar studies tested the effect of aptamers against human cytomegalovirus infections. Two RNA aptamers that recognize HCMV viral glycoproteins B and H, reduced the infectivity of cytomegalovirus in a human foreskin fibroblast cell line culture [175]. Aptamers against HSV-1 bound viral glycoprotein D, a ligand of Nectin-1, which mediates HSV entry into the host cell. Reduced infectious potential of HSV-1 was observed in proportion to the dose of this aptamer. The treatments showed no evidence of cytotoxicity and the aptamer could distinguish between HSV-1 and HSV-2 strains [176].

In my thesis, data from SELEX experiments to select 2’Fluoro pyrimidine modified RNA aptamers against Ebola glycoproteins GP1,2 and sGP is presented.
Figure 1: Ebola Virus Particle.

Figure 2: X-ray crystallography determined structure of GP1,2 ΔmucinΔtm.

The structure of Ebola virus glycoprotein bound to neutralizing antibody KZ52. GP1,2 is a heterotrimer expressed on the surface of the viral membrane. The cartoon depicts the mucin domain that shrouds the protein along with glycan cap. The red region is supposed to be the receptor binding domain. Reprinted by permission from [63].
Figure 3: Prefusion and postfusion conformational switch of GP1,2 glycoprotein.

The cartoon depicts the interaction of GP1,2 with host cell receptors followed by internalization via micropinocytosis. Once the virus is internalized into the endosomes, low pH dependent proteases cleave the glycan cap and mucin domain to expose the receptor binding domain 19-KDa core bonded to GP2. The receptor binding domain interacts with a receptor in the endosomes and this interaction allows a conformational change within GP2. This conformational change allows GP2 to release its internal fusion loop (IFL) and insert into host cell membrane. Interaction between the heptad repeats 2 and 1 brings host cell and viral membrane into contact and forms a hemifusion stalk. The hemifusion stalk causes the formation of a fusion pore. Reprinted by permission from [36].
Figure 4: CryoEM determined structures of full length GP1,2 glycoprotein and soluble glycoprotein (sGP). Cryo EM determined quaternary structure of GP1,2. The figure also depicts the residues of N linked glycosylation and mucin domain sites. Cryo EM determined structure of sGP shows a parallel homodimer connected by disulfide bonds at the N and C terminus. The monomer structures of GP1,2 and sGP are similar. Reprinted by permission from [76]
Figure 5: SELEX protocol for the selection of RNA aptamer that binds the target protein.

A DNA pool with a complexity of $10^{15}$ was PCR amplified, followed by transcription to prepare an RNA pool. The RNA pool was incubated with target protein and non-binders were separated by capture on a nitrocellulose filter. Retained RNA was eluted and reverse transcribed followed by PCR to amplify the selected sequences. This process was performed for 8-10 times before submitting for NGS.
References


160. Huang, Z., X. Wang, and G. Gao, Analyses of SELEX-derived ZAP-binding RNA aptamers suggest that the binding specificity is determined by both structure and sequence of the RNA. Protein & cell, 2010. 1(8): p. 752-759.


CHAPTER 2

A 2’FY-RNA EPITOPE FORMS AN APTAMER FOR EBOLAVIRUS sGP:

SELECTION AND CHARACTERIZATION

Shambhavi Shubham\textsuperscript{a,b}, Jan Hoinka\textsuperscript{c}, Emma Swanson\textsuperscript{a}, Teresa M. Przytycka\textsuperscript{c}, Wendy Maury\textsuperscript{d}, Soma Banerjee\textsuperscript{a,b} and Marit Nilsen-Hamilton\textsuperscript{a,b}

\textsuperscript{a}Iowa State University, Ames IA, \textsuperscript{b}Aptalogic Inc., Ames IA, \textsuperscript{c}National Institutes of Health, Bethesda, DC, \textsuperscript{d}University of Iowa, Iowa City IA

Short title: A 2’FY-RNA Epitope Selectively Binds Ebola Virus sGP

Experimental contributions to this manuscript:

SELEX against sGP, identification of sequence structure motif, binding assays to determine poly U and 2’FY modification dependent high affinity binding, truncation of parent aptamer to determine the RNA binding epitope, antisense hybridization assay and binding assays to determine the Kd of aptamer 39SGP1A against sGP were performed by Shambhavi Shubham

Analysis of Next Gen Sequencing data and the identification of top 10 clusters was done by Jan Hoinka.

Emma Swanson repeated the binding assays of 39SGP1A against sGP and performed binding assays to determine the affinity of 39SGP1A against serum proteins.

Soma Banerjee determined the binding affinity of 39SGP1A against sGP using an Aptamer Sandwich Binding Assay that she developed.
ABSTRACT

Nucleic acid aptamers have advantages over antibodies for incorporation into sensing devices. These advantages include stability to storage in the dehydrated state, structural stability for repetitive use, and compatibility with many sensing/transducing mechanisms. Such capability is important for detecting Ebola virus (EBOV) for which epidemics frequently start in locations that lack sophisticated equipment. sGP (soluble glycoprotein), a nonstructural Ebola viral glycoprotein, is an EBOV biomarker that is secreted in abundance in the bloodstream even during the early stages of infection. Here, we report the selection of 39SGP1A, a 2’fluoro pyrimidine (2’FY)-modified RNA aptamer that specifically binds to sGP and does not bind human serum albumin, the most abundant protein in human blood. We demonstrate by computational and biochemical analysis that the recognition motif of 39SGP1A is a novel polypyrimidine-rich sequence. The replacement of the F group in the 2’ position of the ribose with an OH group resulted in complete loss of affinity for sGP. Computational structural modeling suggests that this motif could form a compact surface for protein interaction for which modification by fluorine increases molecular hydrophobicity.
INTRODUCTION

Ebola virus, in the family of Filoviridae, is a highly virulent enveloped RNA virus that causes hemorrhagic fever with 50-90% fatality and for which early detection is the key to controlling epidemics such as have occurred in the recent past. The Ebola virus GP gene encodes two glycoproteins with the surface exposed trimeric glycoprotein (GP1,2) being the minor product. The major translated product is sGP (soluble glycoprotein), which is conserved amongst the five Ebola virus species (1,2). sGP shares its first 295 amino acids with GP1,2 but has a unique 28 amino acid carboxy terminus. Monomers are disulphide linked (at C53 and C306) in a parallel orientation and appear in the blood as homodimers. It is N glycosylated at 6 positions and C-mannosylated at W288. Due to their sequence homology sGP and GP1,2 have similar monomeric structures (3). Although, the role of sGP in pathogenesis is not yet established, antibodies elicited during infection cross react with sGP and GP1,2. This observation supports role for sGP as a decoy for antibodies against GP1,2 on the virus particle (4). Consistent with this hypothesis is the observation that sGP is secreted in abundance in the early stages of infection. Thus, it is an excellent choice as an Ebola virus biomarker.

Despite its clear advantage as a biomarker for Ebola virus infections, sGP has been so far overlooked for this purpose. Here we describe the application of SELEX (5-7) combined with Next Gen Sequencing (NGS) and analysis by APTAGUI to select a 2’FY-RNA aptamer (39SGP1A) with high affinity and specificity for sGP. Aptamer 39SGP1A binds sGP tightly with a Kd of 27nM. By contrast, 39SGP1A does not bind human serum albumin. The requirement for structure in the binding of 39SGP1A to sGP is evident by the lack of affinity for sGP of the RNA (2’OH) version of this aptamer, which distinguishes the 39SGP1A motif
from the RNA recognition domain of the polypyrimidine tract binding protein (8-10). 2’F modification, which provides resistance against endonucleases (11) also results in conformational differences from the 2’OH RNA. The 2’F-modified ribose sugar tends to adopt the C3’ endo conformation and stabilizes the RNA structure (12). In addition, the inclusion of fluorine modifications increases molecular hydrophobicity. Here, we describe the aptamer and characterize its interaction with sGP, identifying a novel 2’FY-RNA binding domain as a poly-2’F-U loop with GAGC in the stem.

RESULTS

Selection of 2’FY-RNAs with High Affinity for sGP

2’FY-RNAs with high affinity for sGP were selected by SELEX from an oligonucleotide pool, consisting of a 53nt random region sequence flanked by 25 bases 5’ and 3’ constant regions with a starting complexity of $10^{15}$ molecules of RNA. The protocol followed a mathematically defined approach of starting with a relatively high molar ratio of 2:1 (RNA: sGP) followed by harmonic reductions in the sGP concentration (13) to increase the stringency of selections (Fig.1B). To eliminate non-specific sequences binding to the nitrocellulose membrane, negative selections were performed at rounds 2 and 4. After 9 rounds of selection, enrichment of the pools was analyzed by nitrocellulose filter capture assay and EMSA. The binding curve of the initial pool did not reach saturation, whereas the binding curve of the final pool was biphasic with saturation achieved at ~100 nM and ~1μM respectively suggesting selection of high and low affinity RNA binders (data not shown). The filter binding results were corroborated by the results from electrophoretic mobility shift assay
(EMSA) results using the pools from the 1\textsuperscript{st} and 9\textsuperscript{th} rounds RNA pools. The 1\textsuperscript{st}, 4\textsuperscript{th}, 6\textsuperscript{th} and 9\textsuperscript{th} rounds pools were given for Next Gen Sequencing Analysis and simultaneously 9\textsuperscript{th} round pool was sub-cloned and sequenced by TOPO-TA cloning.

**Sequencing Analysis and Identification of Conserved Sequence Motifs**

APTA GUI (14), an open source graphical user interface was used for the Next Gen Sequencing data analysis. APTAGUI provides essential information regarding the quality of the data by separating the reads into several rounds and extracting randomized sequences. APTA GUI also generates enrichment ratios across the selection cycles, which helps in selecting prospective sequences for further analysis. In addition to individual sequences it also provides information about various aptamer clusters and sequence families. The APTAGUI platform allows the user to study the mutation count on each individual sequence. Due to its extensive analysis of the aptamer families, it greatly facilitates the identification of aptamers. Next Gen Sequencing Analysis (Figs. 1A, C, D). The top 10 Sequence clusters obtained by TOPO TA cloning were identical to the Top 10 clusters obtained from Next Gen Sequencing analysis.

Analysis of the sequences in the top 10 clusters from NGS with MEME (15) identified a conserved sequence motif of a stretch of U’s and GAGC (Fig. 1E). No other sequence structure motifs were found in other enriched clusters. In all the secondary structures predicted for the sequence motif by MFold (16) the poly 2’F-U’s were in a loop followed by GAGC in a stem.
The poly2’F-U-GAGC Sequence Motif Identifies A High Affinity Aptamer

The affinities of 2’FY-RNA oligonucleotides 5183, 5177, 5179 and 5182 containing the poly 2’F-U-GAGC sequence motif were compared with oligonucleotides 4789, 5181 from the top sequence cluster that did not contain the motif. In the initial titrations, we observed saturation at 100nM for all sequences containing U rich motif suggesting high affinity binding. The results, demonstrated for 2’FY-RNA-5183 and 2’FY-RNA-4789 revealed high affinities (Kd = 20-50 nM) for oligonucleotides containing the poly 2’F-U-GAGC sequence motif and low affinities (Kd = 0.5 to 1μM) for sequences that did not contain the motif (Fig. 2 A-D). That the 2’FY modification is essential for 2’FY-RNA-5183 binding to sGP is demonstrated by the observation that the same sequence in the form of an RNA (with 2’OH) did not bind sGP (Fig. 2A). sGP binding by 2’FY-RNA 5183 was confirmed by electrophoretic mobility shift assay (EMSA, Fig. 2E).

The specificity of 2’FY-RNA-5183 binding to sGP in the presence and absence of an unrelated RNA aptamer (RNA-569) that recognizes mouse Lcn2 (17). RNA-569, in 10 molar excess, did not compete for 2’FY-RNA-5183 binding to sGP (Fig. 2E). These results are consistent with the hypothesis that 2’FY-RNA-5183 binds sGP with specificity and high affinity. Cluster containing 2’FY-RNA-5183 was examined for base frequency at each position by APTAGUI to establish a “mutation rate” through the selections (14). 5’ end stem loop, which contains the poly 2’F-U-GAGC sequence motif, was identified as the most stable sequence. Truncations of the 2’FY-RNA-5183 were tested for binding sGP (Fig. 3A, B, D). These results showed that the high affinity binding element in 2’FY-RNA-5183 contains the poly 2’F-U-GAGC sequence motif.
A 2’FY-RNA-5199, a 39nt truncated version of 5199 containing the poly 2’F-U-GAGC sequence motif that binds sGP with Kd of 27nM was given the aptamer name of 39SGP1A. Further truncation to 2’FY-RNA-5199 containing only 6 U’s within the loop (Oligo 5015) had lower affinity for sGP, supporting the hypothesis that the structure of the U rich loop, rather than just the presence of a series of 2’FY modified Us, is important for binding (Fig. 3F). We also tested the binding of 39SGP1A with sGP by performing UV crosslinking and observed a clear mobility shift when the RNA and protein were combined. We confirmed that the shifted band was an RNA-protein photo-adduct by digestion with Proteinase K (Fig. 3C). We also tested involvement of the poly 2’F-U loop by hybridizing it with antisense oligonucleotide DNA-5196. The presence of a 10-fold molar excess of DNA-5196 over 39SGP1A resulted in 50% reduction in the sGP-RNA shifted band after crosslinking. These results demonstrate that 39SGP1A is a 2’FY-RNA aptamer that bind sGP with high affinity.

**How Many Epitopes On sGP Are Recognized By The Selected 2’FY-RNA Oligonucleotides?**

Several RNA oligonucleotide families were identified from the NGS results that are not homologous in sequence. Whereas members of the family identified by the poly 2’F-U-GAGC sequence motif bound sGP with high affinity, members of other families bound with lower affinity (example in Fig. 2). To establish if these oligonucleotides bound to the same or different protein epitopes, we tested the ability of a representative of another family to compete for binding of 39sGP1A with sGP.
The 39SGP1A aptamer, 2’FY RNA oligonucleotides 4789, 5179 and oligonucleotide 6011, a DNA oligo that binds sGP, were used as competitors of $^{32}$P-39SGP1A for binding sGP. Competition was observed with all the oligonucleotides, which varied in affinity for sGP. This observation suggests that sGP has one favored nucleic acid binding epitope and the high affinity and low affinity binding is determined by the tertiary structure of the RNA (Fig.4).

**Specificity of 39SGP1A**

To establish if 39SGP1A is suitable as a sensor for early diagnosis of sGP in patient serum, we tested its affinity for human serum albumin (HSA). HSA is highly abundant in blood, reaching concentrations up to 300 μM. The 2’FY-RNA oligonucleotide-5183, from which 39SGP1A was derived, bound HSA with a Kd of 400 nM (Fig. 5A). However, once isolated from the remainder of 2’FY-RNA-5183, 39SGP1A did not bind HSA up to 68 μM (Fig. 5B). 39SGP1A was also found not to bind two other highly abundant serum proteins, alpha 2 macroglobulin and fibrinogen (Fig. 5C).

**DISCUSSION**

In this study, we identified an RNA aptamer that binds sGP with high affinity and specificity. sGP is a non-structural Ebola virus 70 KDa glycoprotein that forms a parallel homodimer linked by disulfide bonds, which is responsible for immune modulation during infection (4). The choice of sGP as a SELEX target was motivated by the fact that it is secreted in abundance in blood early during infection, which makes it an excellent biomarker for Ebola infection (2). An aptamer that binds sGP could be incorporated onto a biosensor platform for
early infection detection. 39SGP1A is the first RNA aptamer reported to be specific for sGP. For our selections, we began with a complexity of $10^{15}$ 2'FY modified RNA oligonucleotides. 2’F ribose modification provides nuclease resistance and has widely been used in aptamer selections. After eight rounds of selection we compared the binding of the final pool with the starting pool by the filter capture assay. The initial RNA pool showed no evidence of saturation at 5 μM, whereas in the final round RNA pool the sGP titration indicated the presence of two groups of oligonucleotides identified by high and low affinities. We used conventional TOPO TA cloning for high throughput sequencing of the 8th round and performed Next Gen Sequencing on 1st, 4th, 6th and 8th Round pools. Clusters obtained from high throughput sequencing utilizing TOPO TA cloning were identical to the clusters identified by Next Gen Sequencing.

To identify potential aptamers, we searched for conserved motifs within the top 10 clusters and identified a consensus sequence of poly U and GAGC. In the predicted secondary structures, the poly U stretch of 18 nts resided in the loop region followed by GAGC’s in the stem region. Oligonucleotides containing the poly U motif demonstrated high affinity binding ($K_d = 20-34$ nM), whereas oligonucleotides from the top ten selected clusters that did not contain the poly U motif bound sGP with low affinity. Certain cellular proteins to bind to poly(U) rich RNA sequences (18-20). However, the selected oligonucleotides in this study contained 2’F- modified pyrimidines, which alters the structure by holding the sugar ribose in the C3’ endo conformation. This conformation is usually favored in RNA duplexes, increasing the helical stability. When oligonucleotide 5183 was prepared with 2’OH uracil and 2’F
cytosine or 2’ OH cytosine binding to sGP was lost, showing that 2’F-U modification is essential for this oligonucleotide to bind sGP with high affinity.

A comparison of selected aptamers from libraries of RNA molecules containing 2'amino and 2'fluoro deoxy pyrimidines resulted in higher affinity aptamers from the 2’FY selection. From this it was speculated that the increased affinity of 2’FY modified RNA is due to formation of thermodynamically stable helices and rigid structures in 2’FY RNA compared to the 2'amino modified RNA (12). However, X ray crystallography studies with 2’F modified and unmodified duplexes demonstrated that the presence of 2’F modification does not alter the overall helical structure of the duplex (21). The increased thermodynamic stability of the helices is due to the highly electronegative fluorine polarizing the nucleobases, which enhances the Watson-Crick pairing. Osmotic stress and X ray crystallography studies show that the presence of 2’Fluorine also significantly reduces the hydration of the duplex, increasing its hydrophobicity (22) (21). In our study also, we observed increased retention of 2’F poly U sequences to the hydrophobic nitrocellulose membranes in the absence of target protein in comparison to 2’OH RNA. Increased retention of nucleic acids to nitrocellulose membranes was predominantly observed with oligonucleotides containing the poly U rich sequences whereas past studies have reported the presence of Multi G Motifs (MGM’s) that exhibit significant nonspecific nitrocellulose retention (23,24).

To localize the aptamer sequence in oligo 5183, we tested truncations, which identified the 5’ poly U hairpin region, which was truncated to 39 nucleotides and tested for binding to sGP by filter capture and EMSA assays and for the ability of a DNA oligo (5196)
complementary to the poly U loop reduce binding to sGP. These results confirmed that oligonucleotide 5199 is an aptamer with high affinity for sGP and it was name 39SGP1A.

and hydrogen bonding interactions among uracil (25) in comparison to GNRA (N any nucleotide and R purine) loops or UUCG loops, which have interactions within loop bases (26). 39SGP1A contains an 18-nucleotide loop region comprised of 77% uracils with intermittent purines and pyrimidines. The effect of 2’Fluoro modification on the structure of a loop is unknown but, based on the reported effects of 2’FY substitution on helical stability due to base polarization and increased hydrophobicity, we speculate that the high affinity of the 2’FY modified 39SGP1A aptamer might be aided by strong intermolecular hydrogen bonding interactions and π-π stacking interactions between the available 2’F uracils with sGP residues. In addition, the 2’F- dependent structural organization within the loop is also expected to contribute to enhanced affinity.

We investigated several of the selected oligonucleotides to determine if they bind the same epitope on sGP in a competition assay with 32P-39SGP1A and unlabeled low affinity (oligo 4789), and high affinity (oligo 5179) and a DNA aptamer independently isolated (manuscript in preparation). These three oligonucleotides competed with 39SGP1A for binding sGP, which suggests that the protein has an epitope that is preferred for RNA binding.

39SGP1A was selected as a potential sensor for early EbolaVirus infection (27), which would be performed with samples of patient blood. To obtain a significant signal to noise ratio the aptamer should not bind other serum proteins. The most abundant serum protein is albumin which is present in blood at concentrations as high as 300 μM. Although the selected
oligonucleotide, 5183, from which 39SGP1A was derived, bound to HSA with a Kd of 500 nM, 39SGP1A did not bind HSA. Thus, truncating the oligonucleotide to a minimal critical sequence, appears to have eliminated structures that could bind HSA. 39SGP1A also did not bind α2 macroglobulin and fibrinogen, two other serum proteins which could contribute to the background noise.

In summary, we report the selection of a 2’FY-RNA aptamer that binds the Ebola virus sGP glycoprotein with high affinity and specificity and demonstrate that the 2’F modification is required for binding to sGP.

**MATERIALS AND METHODS**

**SELEX library construction**

A DNA single stranded oligonucleotide library named 487 was synthesized by IDT (Integrated DNA Technologies, Coralville, IA) with the following sequence: 5’CCTGTGTTGTAGCCTCCTGTCGAA (53N) TTGAGCGTTTATTCTTGTCTCCC 3’, and N symbolizes equimolar mixture of A, C, G and T. The following primers were used for reverse transcription and PCR reactions: Oligo484:5’-TAATACGACTTATAGGGAGACGCAATAAACGCTCAA-3’ and Oligo 485: 5’-GCCTGTGTTGAGCCTCCTGTCGAA-3.'
Conversion of ssDNA SELEX library to dsDNA

To generate starting RNA pool, an extension reaction was performed using primer 484 and the starting ssDNA pool. 10 reactions of 1000uls each containing 2 μM 487 pool, 3.3 μM 484 primer, 0.5 mM dNTP mix, 0.03 U/μl DNA Taq Polymerase (GenScript) in reaction buffer containing 50 mM KCl, 10 mM Tris-HCl (pH 8.55), 1.5 mM MgCl$_2$, 0.1% TritonX-100 was incubated at 94°C for 5 min, 65°C for 15 min and 72°C for 99 min using the Multi GENE II PCR minicycler. The generated dsDNA was run on 2% agarose gel and purified using Qiagen PCR purification kit. Purified DNA was quantified via Nanodrop.

RNA synthesis and purification

RNA was prepared by *in-vitro transcription* using Durascript$^{TM}$ T7 transcription kit from (Epicentre, Madison). Around 2nmoles of dsDNA from pool generated from extension is incubated with 5 mM ATP, 5 mM GTP, 5 mM 2’FY UTP, 5 mM 2’FYCTP, 5 mM DTT, 0.2U/μL T7 Enzyme in a total volume of 2.8 mL and 5% DMSO and incubated at 37°C for 4 h. DNA mixture was then digested with 1 MBU DNase I. The RNA was resolved through a 7M urea 8% (19:1 acrylamide: bisacrylamide) gel in 1X Tris Borate EDTA buffer, pH 9.1 to separate the residual NTPs and abortive transcripts. Transcribed RNA was eluted in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0) for 18 h at 37°C.

*In-vitro selection of RNA aptamers against sGP (soluble glycoprotein)*

In vitro selection of an aptamer for sGP was carried out by SELEX (Systematic Evolution of Ligands by EXponential enrichment). Prior to the first round of selection, a
dsDNA pool was generated from a library of ssDNAs (487D pool) by primer extension. Starting RNA pool was generated by in-vitro transcription using dsDNA (487D) pool as template. RNA pool (1X10^15 molecules) 2nmoles was incubated with sGP and the non-binding RNA sequences were partitioned by passing it through nitrocellulose membrane. For each positive selection, to increase the stringency of selection ratio of RNA pool: sGP was increased. To remove nonspecific membrane binders negative selections were performed against the nitrocellulose membrane. Reverse transcribed polymerase chain reaction (RT-PCR) amplification was used to generate DNA for subsequent selection rounds. 8 rounds of selections were performed including negative and positive selections. Enriched pool from the 10th SELEX round was cloned into the TOPO XL PCR cloning plasmid for sequencing, in addition PCR libraries were prepared for Next Gen Sequencing using Illumina sequencing primers.

**Nitrocellulose capture assay**

To determine the binding affinity of sGP to RNA aptamers, 5’P32 end labeled RNA was firstly prepared. 2’FY modified RNA was transcribed using Durascribe transcription kit (Epicentre Technologies), transcribed RNA was electrophoresed in an 8% polyacrylamide–7 M urea gel, and eluted from the gel. RNA was dephosphorylated with calf intestine phosphatase for 1 h at 37°C, extracted with phenol, and precipitated with ethanol. The RNA was end-labeled with [γ-32P] ATP and T4 polynucleotide kinase. To determine the binding affinity of the RNA aptamer 5183 to sGP, binding studies were performed by first denaturing the 2nM RNA at 95 degrees for 5 mins followed by refolding in binding buffer [137 mM NaCl,
2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, 5 mM MgCl$_2$, pH 7.4] at RT for 30 min. RNA was incubated with varying concentrations of sGP (10 nM-1 μM) at RT for 30 min. The RNA sGP complex was passed through nitrocellulose membranes (HAWP 02500) and filters were washed with 3 mL binding buffer and later quantified using Liquid Scintillation Counter. The data was fit to $F=F_{min}+(F_{max}*L^n)/(L^n+K_d^n)$ to determine the Kd.

**Electrophoretic mobility shift assay (EMSA)**

Binding reactions performed by firstly by denaturing of $^{32}$P- 5’ end labeled RNA at 95 degrees for 5mins followed by slow cooling in binding buffer [137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, 5 mM MgCl$_2$, pH 7.4] for 30 min. Refolded RNA was incubated with sGP in binding buffer for 30 min at room temperature. The mixtures were analyzed for complex formation by resolution through a 6% native polyacrylamide gels run in Tris HEPES EDTA buffer (33 mM Tris, 66 mM HEPES, 0.1 mM EDTA-Na, pH 7.5) at 200V and the $^{32}$P was visualized using a phosphor imager.

**Competition binding assay**

Competitive binding reactions were performed by incubating 4 μM sGP with 18 μM competitor RNA sequences for 30 min at RT in binding buffer. Following the preincubation period, $^{32}$P-labeled 39sGP1A 2’FY-RNA was added to the reaction mix to a final concentration of 200 nM and the samples were incubated for 15 min. The RNA protein complexes were then UV crosslinked at 365 nm for 15 min. Crosslinked products were run on 10% SDS PAGE reducing gel, the gel was dried and later imaged by Typhoon scanner and quantification was
performed using Image J software. 39sGP1A RNA aptamer was tested individually against 39sGP1A, 5179, 4789 and 6011 oligonucleotides.

**Incorporation of 4 Thio UTP and RNA protein UV crosslinking**

RNA was *in vitro* transcribed in the presence of 2’F-UTP and 4-thio-uridine 5’triphosphate (Trilink Biotechnologies) at a molar ratio of 1:1 and radiolabeled with $^{32}$P-$\gamma$-ATP by T4 polynucleotide kinase. 3 pmol of RNA was denatured at 95°C for 5 min followed by snap cooling on ice for 5 min. The RNA was folded in PS Buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, 5 mM MgCl$_2$, pH 7.4) for 10 min at RT. To link sGP to the 2’FY-RNA aptamer 5183, 3 pmol 2’FY-RNA was incubated with 15 mol sGP in 15 μL binding buffer for 20 min at RT for complex formation. This mixture was irradiated for 6 min and 12 min under a UV lamp (365 nm) with the samples placed 3 cm from the lamps. RNA-sGP samples were treated with proteinase K for 30 min. The samples were later resolved through a 10% reducing SDS PAGE gel to separate the 2’FY-RNA-protein complex, free 2’FY-RNA and free protein. The $^{32}$P on the gel was visualized using a phosphor imager.

**Antisense competition assay with DNA oligo**

39sGP1A RNA aptamer was transcribed using Durascribe T7 transcription kit, RNA transcript was dephosphorylated using alkaline phosphatase (Promega 1U) at 37°C for 1 h, followed by phenol chloroform extraction. The dephosphorylated 2’FY-RNA was 5’ end labeled with $^{32}$P by incubating with T4 polynucleotide kinase (NEB) at 37°C for 30 min followed by 65°C for 15 min to inactivate the enzyme. The labeled RNA aptamer was
denatured at 95°C for 5 min followed by snap cooling on ice for 5 min and then incubated at RT for 10 min in (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 2 mM KH₂PO₄, 5 mM MgCl₂, pH 7.4). The folded RNA was incubated in the presence and absence of 5-fold excess of oligonucleotide 5196 for 20 min then sGP was added and the samples incubated for 30 min. The samples were then irradiated for 15 min UV (365nm) at 3 cm distance from the lamp. The samples were later resolved through a 10% reducing SDS PAGE gel to separate the 2’FY-RNA aptamer-protein complex, free RNA and free protein. The ³²P on the gel was visualized using phosphor imager.

ACKNOWLEDGEMENT

We thank Lee Bendickson and Muslum Ilgu for helpful discussions and advice. This work was funded by NIH grant 5R43AI118139-02 to Aptalogic Inc.
LEGENDS TO FIGURES

Figure 1. Next Gen Sequencing Analysis and Consensus Motif Identification. A) The frequency of singletons, enriched and unique sequences for each SELEX round B) The SELEX protocol for selecting 2’FY-RNA aptamers with high affinity for sGP C, D) The frequency of base occurrence in each position for the 1st and 9th round pools showing evidence of sequence enrichment during selection E) A conserved sequence motif is identified among the top identified clusters.

Figure 2. The Affinities of Selected 2’FY-oligonucleotides for sGP. A) The concentration dependence of sGP binding to 2 nM ³²P-5183 as a 2’FY-RNA and an RNA oligonucleotide. The estimated Kd was for the 2’FY-RNA was 50 nM. B) An extended sGP titration up to 3 μM against 2nM nM oligonucleotide ³²P-5183 to validate saturation of the binding curve, C) The concentration dependence of sGP binding to 20 nM ³²P- oligonucleotide 4789. The estimated Kd was 500 nM. D) M fold predicted secondary structures of oligonucleotide s 4789 and 5183 E) Electrophoretic mobility shift assay with 100 nM ³²P- oligonucleotide 5183. Buff, buffer only. sGP, buffer plus 2 μM sGP. comp, buffer plus 2 μM sGP and 10 μM Lcn2 aptamer.

Figure 3. Determining the Binding Epitope for sGP on the 2’FY RNA aptamer. A) A mutation count chart created in APTAGUI to determine the mutation frequency for 5183 at each nucleotide position, B) The M-fold predicted secondary structures of truncations of 5183, C) UV crosslinking of 200 nM ³²P 39SGP1A to 2.5 uM sGP followed by treatment with or without proteinase K, D) Binding of 10nM ³²P-oligo 5183 truncated oligonucleotides to 50nM and 500nM sGP, E) Binding of 0.5 uM ³²P-39SGP1A to 10uM sGP in the presence of
5 μM antisense oligonucleotide 5196, which is complimentary to the loop region of oligonucleotide 39SGP1A. F) Concentration dependence of sGP binding to 10nM \( ^{32}\text{P}-\text{oligonucleotide 39SGP1A} \) and 10 nM \( ^{32}\text{P}-\text{oligonucleotide 5015} \). The estimated Kd was 27nM for 39SGP1A.

**Figure 4. Competition Binding Assay to Determine if the Selected Aptamers Bind to the Same Protein Epitope.** \( ^{32}\text{P}-\text{oligonucleotide 39SGP1A} \) (0.2 μM) was incubated for 15 min at 23°C with a 10-fold excess of the identified oligonucleotides. The samples were stabilized by UV crosslinking and resolved by SDS-PAGE.

**Figure 5. Specificity of 2’FY-5183 and 39SGP1A for sGP.** A) Binding of 2 nM 2’FY 5183 protoaptamer to human serum albumin (HSA) B) Binding of the 10nM \( ^{32}\text{P} \) 2’FY 39SGP1A aptamer to HSA C) Lack of affinity of 10 nM 2’FY 39SGP1A for human serum albumin (200 nM to 50 μM), fibrinogen (200 nM to 4 μM) or α2macroglobulin (200 nM to 5 μM).
REFERENCES


Next Gen Sequencing Analysis And Consensus Motif Identification
The Affinities of Selected 2’FY-oligonucleotides for sGP.
Determining the Binding Epitope for sGP on the 2’FY RNA Aptamer.
Figure 4

Competition Binding Assay to Determine if the Selected Aptamers Share a Protein Epitope
Specificity of 2’FY-5183 and 39SGP1A for sGP.
CHAPTER 3

SELECTING A FUNCTIONAL ANTI-VIRAL RNA APTAMER AGAINST

EBOLAVIRUS SURFACE GLYCOPROTEIN

Shambhavi Shubham\textsuperscript{a,b}, Jan Hoinka\textsuperscript{c}, Emma Swanson\textsuperscript{a}, Teresa M. Przytycka\textsuperscript{c},

Wendy Maury\textsuperscript{d}, and Marit Nilsen-Hamilton\textsuperscript{a,b}

\textsuperscript{a}Iowa State University, Ames IA, \textsuperscript{b}Aptalogic Inc., Ames IA, \textsuperscript{c}National Institutes of Health, Bethesda, DC, \textsuperscript{d}University of Iowa, Iowa City IA

Experimental contributions to manuscript:

SELEX and identification of potential sequences and binding assays to identify oligonucleotide sequences that bind to GP1,2 Δmucin was done by Shambhavi Shubham.

Analysis of Next Gen Sequencing data was performed by Jan Hoinka along with the identification of top ten clusters.

Emma Swanson performed the binding assays for oligonucleotide sequences that bind to GP1,2 full length.
ABSTRACT

Ebola viruses (EBOV) cause severe disease in humans and in non-human primates in the form of viral hemorrhagic fever. Ebola virus is also of public health concern as a potential bioterrorism organism for which no vaccine or anti-viral is available. Fast-acting and prophylactic therapeutics are needed to reduce mortality. In view of the paucity of current antiviral therapies for EBOV we undertook to select an aptamer against the EBOV surface-exposed glycoprotein, GP1. Here we described the selection and characterization of several 2’FY RNA aptamers that recognize the GP1,2 heterotrimer. One aptamer recognizes the heterotrimer with the mucin domain included, which is the form of the protein found on the circulating virion. Another aptamer recognizes the GP1,2 heterotrimer lacking the mucin domain of GP1, which is the form of the protein in the endosomes that mediates entry into the cytoplasm. Thus, these aptamers target the virion in two stages of infection and together might be effective in preventing the interaction between virus and the host cell and viral entry into the cytoplasm, thereby reducing viremia and spread of the virus to others.
INTRODUCTION

Ebola Virus (EBOV) and Marburg Virus (MARV) are members of the family *Filoviridae* of enveloped viruses with a non-segmented negative strand RNA genome. Filoviruses cause sporadic outbreaks of hemorrhagic fever in human and non-human primates in Africa with fatality rates up to 90%. Filovirus hemorrhagic fever is associated with high levels of inflammatory cytokines and coagulation disorders resulting in septic shock and multiorgan failure (1). The virus is transmitted through contact with bodily fluids and can infect various cell types across different host species. They typically infect hepatocytes, dendritic cells, endothelial cells, macrophages, and monocytes (2). The viral genome encodes seven structural proteins: envelope glycoprotein (GP), major matrix protein (VP40), nucleoprotein (NP), polymerase cofactor (VP35), replication/transcription protein (VP30), minor matrix protein (VP24), and RNA dependent DNA polymerase (L) (3). The viral replication cycle involves attachment of the virus to its receptors, uptake of the virus, intracellular trafficking of the virus in the endosome and release of the nucleocapsid into the cytoplasm, synthesis of viral proteins, assembly and budding from the cell surface (4).

Viral glycoprotein (GP1) is a homotrimer, forming spikes on the surface of the virus and is the only protein on the EBOV viral particle surface. Folding and assembly of GP1 occurs independently of other viral proteins (5). It is initially synthesized as a precursor that is later cleaved by the proprotein convertase, furin, into GP1 (140kD) and GP2 (26kD) that are linked by disulfide bonds (6). The GP2 subunit contains two heptad repeat regions, which facilitate assembly of GP into trimers, a transmembrane anchor sequence, and the fusion loop (5). The GP1 subunit contains the receptor binding site and a heavily glycosylated mucin domain. The
residues responsible for mediating interactions between GP1 and the host cell are in the 230-residue N-terminal domain of GP1. The crystal structure of EBOV GP demonstrates that the receptor binding site of GP1 is masked by the glycan cap created by the mucin domain (7). Consistent with this structure is the observation that deletion of the C-terminal mucin domain enhances viral infectivity in vitro (8-10).

Ebola viruses infect a broad range of cells and a host of proteins are known to enhance viral entry into the host cells including the C-type lectins L-SIGN, DC-SIGN and the tyrosine kinase receptor Axl (11-14). Infectivity involves at least two recognition events 1) endocytosis at the host cell surface and 2) cytoplasmic entry in the endosome. The cell surface receptor for EBOV is believed to be the T-cell immunoglobulin and mucin domain 1 (TIM-1) receptor (15). The usual function of the TIM-1 receptor is to bind phosphatidyl serine on the surface of apoptotic cells and facilitate their phagocytosis (16). TIM-1 also facilitates the entry of Dengue Virus by directly interacting with the virion-associated phosphatidylserine (17). Structural studies of the TIM immunoglobulin domain have demonstrated that phosphatidyl serine binds in a cavity built up by the CC’ and FG loops termed the metal ion-dependent ligand binding site (MILIBS) (17). By contrast with the Dengue virus interaction, TIM-1-mediated Ebola infection depends on a direct interaction between the viral glycoprotein GP through residues in the ARD5 domain, which is outside of the TIM-1 phosphatidylserine binding pocket and does not include the MILIBS (15). This specificity was demonstrated by showing that the anti-TIM antibody, which binds the ARD5 epitope, completely blocked EBOV entry to Vero cells, whereas A6G2 antibody, which prevents TIM-1 binding to phosphatidyl serine, was much less effective in blocking the viral transduction (15). Co-immunoprecipitation studies also showed that there is
direct interaction between Ebola GP and TIM-1 that depends on the receptor binding domain (15). However, later it was discovered that phosphatidyl serine on the virus surface directly binds TIM-1 thereby promoting virion internalization (18).

The endosomal receptor for EBOV is believed to be Niemann-Pick C1(NPC1), a lysosomal cholesterol transporter (19). The binding site for C1(NPC1) is exposed after endosomal cysteine proteases cleave EBOV GP1 to remove the heavily glycosylated C terminal region to generate an entry intermediate comprising of the N terminal of GP1 and GP2 (19,20). Mutations in the C1(NPC1) binding site on GP1 decrease viral entry (19).

In this study, we report the selection and characterization of 2’FY RNA aptamers against GP1,2 delta mucin and full length GP1,2 glycoproteins. Together these aptamers are expected to thwart interactions between virion and host at two stages. The GP1,2 delta mucin conformation with exposed RBD interacts with NPC-1 in the endosome. Therefore, an RNA aptamer against the GP1,2 delta mucin could intercept the viral host interaction at that step. In addition, TIM-1 is not expressed by all EBOV permissive cell lines (18), which suggests the occurrence of other unidentified cellular receptors for EBOV. In this situation, an RNA aptamer against of GP1,2 full length will be effective in preventing viral replication.

RESULTS

SELEX and Analysis

To identify aptamers that bind Ebola GP1,2 we performed two SELEX experiments in which early rounds selected against the purified recombinant protein and later rounds were against the protein in the context of the viral particle (Fig. 1). GP1,2 lacking the mucin domain
(GP1,2Δmucin) was used in the early rounds of SELEX 1280 and the full length GP1,2 was used in the early rounds of SELEX 1281 (Fig. 1A, D). In both protocols, later rounds were selected against Vesicular Stomatitis Virus (VSV) pseudotyped with GP1,2Δmucin (SELEX1280) or GP1,2 (SELEX1281). Next Gen Sequencing results from both protocols showed evidence of expanding potential aptamer populations with a reduction in unique sequences from early to late rounds, which indicates population expansions (Fig. 1B, E). The progress of oligonucleotide selection evaluated from frequency plots for each round in which the number (counts) of each base (Y axis) is plotted for each base position in the oligonucleotide length (X axis). The frequency plots show that a different population emerged in SELEX1280 after switching the selection from the protein to the virus particles, whereas this did not happen in SELEX1281 (Fig. 1C, F). We selected the top 10 enriched classes from each SELEX experiment for further study.

From the SELEX experiments we identified clusters of selected sequences families using the APTAGUI web application. To further identify which clusters likely contained high affinity aptamers, oligonucleotides representing the top 10 sequence clusters found for SELEX 1280 were mixed in equal amounts and incubated for 15 min with VSV alone or VSV pseudotyped with GP1,2Δmucin (RNA/Protein =10:1). The samples were then spun through a sucrose cushion to separate the RNAs bound to the virus particles from the unbound RNAs. The collected RNAs were cloned and sequenced. Oligonucleotides 4789 and 4796 were the most highly represented of oligonucleotides that remained associated with the virus particles through the sucrose cushion, summing to 47% of the clones from the GP1,2Δmucin-VSV population (Fig. 2). These oligonucleotides also had the highest selective binding to the GP1,2Δmucin-
VSV captured population, being represented 3-4.5 times as much with GP1,2Δmucin-VSV than with the VSV control particles.

**The Dissociation Constant (Kd) of Selected Proto-Aptamers**

Oligonucleotides 4789 and 4796 (SELEX 1280), identified from the sucrose centrifugation assay, were tested for their affinities for the GP1,2 delta mucin recombinant and GP1,2 full length recombinant proteins using the nitrocellulose filter capture assay. Oligonucleotide 4789 bound GP1,2Δmucin with a Kd of 140nM but did not bind the full length GP1,2 glycoprotein (Fig. 3A). These results suggest that 4789 binds to an epitope on GP1,2 that is inaccessible in the presence of mucin domain. The interaction between oligonucleotide 4789 and GP1,2 was also demonstrated by EMSA in which a shifted band was observed when the RNA was incubated with GP1,2Δmucin. The specificity of oligonucleotide 4789 binding for GP1,2Δmucin was also tested with BSA for which no binding was observed. (Fig. 3B).

4797 oligonucleotide sequence appeared in the top clusters of both SELEX 1280 and 1281 and was tested for its binding to GP1,2 full length and GP1,2 delta mucin recombinant protein. Interestingly, it bound neither protein. This sequence appeared in the top clusters, which were performed over different time periods. This might be an example of the evolution of a “sticky” oligonucleotide that binds nonspecifically to many surfaces such as the nitrocellulose membrane used for particle capture. The “sticky” nature of this oligonucleotide is also consistent with the observation that 4797 pelleted predominantly with VSV control particle when passed through a sucrose gradient.
Oligonucleotides 5185, 5186, 5187 and 5188 sequences selected in SELEX 1281 in but not in SELEX 1280 were tested for binding to the GP1,2 full length protein. The affinities (Kd) of these oligonucleotides were 50 nM, 120 nM, 560 nM and 300 nM for 5185, 5186, 5187 and 5188 respectively. These oligonucleotides tested for binding to the isolated GP1,2 proteins were further evaluated for their ability to bind virus particles pseudotyped with full-length GP1,2 and with control virus particles lacking GP1,2. Oligonucleotide 5185 bound GP1,2 full length glycoprotein with a Kd of 50 nM and it also bound the GP1,2 pseudotyped virus particles in a titration that achieved saturation. (Fig.4).

**DISCUSSION**

Aptamers have notable advantages over antibodies that include non-immunogenicity, *in-vitro* production and stability (21). A variety of aptamers have been selected to counter host-virus interactions by targeting different viral proteins (21-23). Examples include: 1) An RNA aptamer that recognizes HIV-GP120 that effectively neutralizes clinical isolates of HIV (24), 2) Viral RNA-dependent RNA polymerases with basic patches that permit easier selection of an RNA aptamer (25), and 3) the internal ribosome entry site (IRES), a structured RNA region in the viral mRNA that binds the ribosome and initiates cap-independent translation to favors viral protein translation. Aptamers targeting the IRES inhibit IRES-dependent translation of HCV proteins (26). These cited aptamers were selected to counter host–virus interactions by targeting viral components.

In this study 2’FY modified RNA aptamers were identified that bind to two conformations of Ebola glycoproteins which could be used to thwart viral host cell interactions. SELEX 1280
was performed against GP1,2Δmucin and SELEX1281 was performed against full length GP1,2. We employed Next Gen Sequencing (NGS) of the random region of several SELEX populations to identify expanding potential aptamer populations. NGS has the advantage over conventional sequencing that it provides up to 130 million sequence reads and, with barcoding, multiple rounds can be sequenced simultaneously. In these SELEX experiments, reduction in unique sequences from 87% to 23% of the total between the 2nd and 5th rounds demonstrated selection for a subset of sequences. The 9th round involved selection against virus particles and the frequency plot showed that a different population emerged in this round for SELEX 1280. This change in dominant populations may be due to the elimination of oligonucleotides that bound to surfaces on the viral protein other than the protein surface exposed on the virus particles. In contrast, there wasn’t much increase in the percentage of unique sequences in round 9. The unique oligonucleotide sequences are expected to be oligonucleotides that bind weakly to the protein or that do not bind and that were therefore retained through the washing during filter capture.

For SELEX 1281, 1st, 4th, 5th, 8th and 10th rounds were barcoded by PCR and sequenced. 27 million reads were obtained with same number of sequences obtained from each round. The uniqueness declined from each round, with the first round having 99.6% in the unique fraction and the percent of unique sequences reduced with each cycle. The last round of selection had 25.9% unique sequences. Reduction in the unique fraction in both SELEX experiments clearly suggests that sequences were enriched as we moved along in the selection process. In contrast to SELEX1281, in which there was change in the frequency plot when the target was changed
from protein to virus particle expressing protein, no change in population was observed when the target was changed from recombinant GP1,2 to virus particles expressing GP1,2.

Our overall goal was to obtain RNA aptamers that can be applied in the blood in which they must remain associated with the viral particles in the presence of shear stress. Therefore, selection was performed under conditions that apply shear to the aptamer-viral complexes. The RNAs bound to viral particles were spun through a sucrose cushion to identify RNAs retained on the virus particles after subjection to shear stress. For this assay the top ten potential aptamers identified from SELEX 1280 were mixed at concentrations to achieve high RNA/Protein (10:1). Centrifugation was for an hour to select for aptamers that form complexes with slower koffs (27). The kinetic half-lives of aptamer protein complexes with pM Kd’s range from 6 min to 60 mins.

Using this method, we selected for oligonucleotides with the sequences 4789 and 4796 that preferentially bound to the GP1,2Δmucin pseudotyped VSV virus particles. Oligonucleotide 4789 bound to GP1,2 Δmucin with a Kd of 143 nM and it showed concentration dependent saturation binding against GP1,2Δmucin virus particles. We did not observe any binding with the VSV control virus particles. From the top clusters obtained from SELEX 1281, we tested the binding of oligonucleotides 5185, 5186, 5187 and 5188 to full length GP1,2. All the oligonucleotides bound to full length GP1,2 with different affinities. Among these, oligonucleotide 5185 bound full length GP1,2 with high affinity (50 nM) Kd and also bound the GP1,2 pseudotyped virus particles. Thus, from the binding assay and sucrose centrifugation assay we can conclude that oligonucleotides 4789 and 5185 are sequences that bind to
GP1,2Δmucin and full length GP1,2 conformations, in recombinant form and when expressed on virus particles.

GP1,2 interacts with the NPC-1 receptor in endosome in low pH conditions therefore the next step in the development of these aptamers will be to test the sensitivity of the aptamer-GP1,2 interaction at low pH. The identified oligonucleotides that bind full-length GP1,2 pseudotyped virus particles will be tested to determine if they prevent viral replication using an \textit{in vitro} transduction assay. Once an aptamer that reduces viral infectivity \textit{in vitro} has been identified, the efficacy of these aptamers against the viruses will be determined in an \textit{in vivo} assay. If these identified oligonucleotides that bind to the GP1,2 proteins do not prevent viral transduction they could still be used as drug and anti-viral agent carriers to tag along with the virus particles and target the cells that are also targeted by the virus (28,29). They could also be developed for diagnostic purposes (21,30). For EBOV, a rapid diagnosis of infection is required because, after the appearance of symptoms, the patient succumbs to infection in 10 days. The current methods for diagnosis of EBOV are RT-PCR and ELISA (31). But these are time consuming assays that require sophisticated and expensive equipment. The necessary equipment for these assays is generally lacking in regions of EBOV occurrence. Aptamers are compatible with inexpensive analytical equipment that could be readily used in field to detect EBOV and other analytes.
MATERIALS AND METHODS

SELEX library construction

DNA single stranded oligonucleotide library named 487 was generated 5’CCTGTTTGTGAGCCTCCTGTCGAA (53N) TTGAGCGTTTATTCTTGTCTCCC 3’, and N symbolizes an equimolar mixture of A, C, G and T. It was synthesized by Integrated DNA Technologies, Coralville, IA. Primers given below were used for reverse transcription and PCR reactions.

484: 5’TAATACGACTCACTATAGGGAGACAAGAATAAACGCTCAA-3’ and
485: 5’-GCCTGTTTGTGAGCCTCCTGTCGAA-3.

Conversion of ssDNA SELEX library to dsDNA

To generate a starting RNA pool, an extension reaction was performed using primer 484 and the starting ssDNA pool. 10 reactions of 1000 μL each containing 2μM 487 pool, 3.3μM 484 primer, 0.5 mM dNTP mix, 0.03 U/μL DNA Taq Polymerase (GenScript) in reaction buffer containing 50 mM KCl, 10 mM Tris-HCl (pH 8.55), 1.5 mM MgCl₂, 0.1% TritonX-100 was incubated at 94°C for 5 min, 65°C for 15 min and 72°C for 99 min using the Multi GENE II PCR minicycler. The generated dsDNA was run on 2% agarose gel and purified using Qiagen PCR purification kit. Purified DNA was quantified using a Nanodrop spectrophotometer.
RNA Synthesis and Purification

RNA was prepared by *in-vitro* transcription using a Durascript™ T7 transcription kit (Epicentre, Madison). Around 2 nmole of dsDNA from a pool generated from extension was incubated with 5 mM ATP, 5 mM GTP, 5 mM 2’FY UTP, 5 mM 2’FY CTP, 5 mM DTT, 0.2 U/μL of T7 polymerase in a total volume of 2.8 mL and 5% DMSO and incubated at 37°C for 4 h. The DNA was then digested with 1MBU of DNAse I and the remaining RNA was resolved through an 8% (19:1 acrylamide:bisacrylamide) acrylamide gel in 7M urea, Tris Borate EDTA buffer (TBE buffer 89 mM Tris, 89 mM Boric acid and 2 mM EDTA) pH 9.1 to separate residual NTPs and abortive transcripts. The transcribed RNA was eluted in 10 mM Tris-HCl, 1 mM EDTA pH 8.0 for 18 h at 37°C and concentrated by ethanol precipitation.

*In-Vitro* Selection of RNA Aptamers That Bind Ebola Virus Particles and Glycoproteins

SELEX was performed against GP1,2 full length glycoprotein SELEX 1281 and GP1,2 mucin domain deleted glycoprotein (SELEX1280). A ssDNA pool with a complexity of 4^{53} sequences from IDT Technologies was used as starting pool. This pool was designated as 487, with each draw from the tube identifying the subset of oligonucleotides with a different letter (487A and 487D were used here). Constant regions that flank the 53-base central region of random sequence in pool 487 are complementary to primers 484 and 485. In the first round of SELEX, DNA polymerase catalyzed extension copied the ssDNA to produce dsDNA from the forward primer TAATACGACTCCTATAGGGAGACGAAGAATATAGCTCA (484), which contains a T7 promoter. The DNA was *in-vitro* transcribed using T7 RNA polymerase, 2’ fluoro-labeled pyrimidine triphosphates, and unmodified purine triphosphates. The resulting
complexity was about $10^{15}$ RNA sequences. The RNA pool was incubated with the GP1 full length and mucin domain deleted trimer at a 1:1 ratio and the complexes collected on a nitrocellulose filter that allowed the free RNA (non-binders) to pass through. RNA bound to the protein was eluted from nitrocellulose membranes by 7M urea was purified through ethanol precipitation. Purified RNA was reverse transcribed and the cDNA was amplified by PCR under conditions that promote low fidelity transcription.

For SELEX 1280 in each of the successive rounds the concentration of the protein was reduced by 10% to increase the selection stringency. Nine rounds of selection were performed of which the first 7 rounds were done with soluble trimers followed by three rounds of negative selections and the final selection was done using GP1,2 pseudotyped VSV virus particles. The 2nd, 5th, 8th and 9th rounds were barcoded by PCR amplification using primers containing the barcode sequences and obtained 97 million sequences from the sequencing.

Similarly, for SELEX 1281 for each successive round the concentration of protein was reduced by 10%, 4 rounds of positive selections were performed including negative selections against the nitrocellulose membrane. 8th and 9th round involved selection against VSV virus particles pseudotyped with GP1,2 full length glycoprotein. Final 10th round selection was against virus particles lacking GP1,2 to eliminate nonspecific binders. 1st, 4th, 5th, 8th and 10th rounds were barcoded by PCR amplification and we obtained 24 million sequences from sequencing.
Filter Capture Assay

To determine the binding affinity of GP1,2 recombinant proteins to RNA aptamers, 5’\(^{32}\)P end labeled RNA was prepared. The 2’FY modified RNA was transcribed using a Durascribe transcription kit (Epicentre Technologies), separated by electrophoresis through an 8% polyacrylamide gel in the presence of 7 M urea, and eluted from the gel. The RNA was dephosphorylated by incubation with calf intestine phosphatase for 1 h at 37°C, extracted with phenol, and precipitated with ethanol. The RNA was 5’ end-labeled with [\(\gamma\)\(^{32}\)P] ATP and T4 polynucleotide kinase. In preparation to determine its binding affinity for GP1,2 and other proteins, the RNA (2 nM) was incubated at 95°C for 5 min followed by refolding in binding buffer [137 mM NaCl, 2.7 mM KCl, 10 mM Na\(_2\)HPO\(_4\), 2 mM KH\(_2\)PO\(_4\), 5 mM MgCl\(_2\) pH 7.4] at room temperature (RT; 22-24 degree Celsius) for 30 min. The folded RNA was incubated with varying concentrations of GP1,2 (10 nM-1 \(\mu\)M) at RT for 30 min. The RNA GP1,2 complex was captured by passing the reaction mixture through nitrocellulose membranes (HAWP 02500) and washing with filters 3 mL of binding buffer. Quantification of RNA/protein captured on the filter was by liquid scintillation counting. To determine the Kd the data was fit to F=F\(_{\text{min}}\)+(F\(_{\text{max}}\)L^n)/(L^n+Kd^n) (32).

Electrophoretic Mobility Shift Assay (EMSA)

Binding reactions performed with 5’ end labeled RNA that is first denatured at 95°C for 5 min followed by slow cooling in binding buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na\(_2\)HPO\(_4\), 2 mM KH\(_2\)PO\(_4\), 5 mM MgCl\(_2\) pH 7.4) for 30 min. 5\(\mu\)M Refolded RNA was incubated with 0.5\(\mu\)M of recombinant GP1,2 delta mucin or 0.5\(\mu\)M BSA in binding buffer for
30 min at room temperature. Mixtures were analyzed for complex formation by resolving them through a 6% (37.5:1 acrylamide:bisacrylamide) native polyacrylamide gels run in Tris Acetate EDTA buffer (6.75 mM TrisCl, 1 mM EDTA, 3.3 mM sodium acetate) (pH 7.4) @ 200V followed by quantification of the $^{32}$P on the gel using an X ray film.

**Determination of Total Protein and Percent Glycoprotein on Virus Particles**

Total viral protein content was measured using Bradford assay using a standard series of BSA concentrations 1:150 dilutions of GP1,2 pseudotyped virus particles (in duplicates). The pseudotyped virus particles were also resolved by 10% SDS PAGE along with a known concentration of recombinant EBOV GP1, 2 (Purchased from BPS biosciences). The gels were stained with Coomassie blue and the bands quantified by Image J. The amount of GP1,2 in the particles was determined relative to the control the GP1,2 control and then converted to a percent of total protein on the virus particles using the results from the Bradford (Coomassie blue) assay for total protein.

**ACKNOWLEDGEMENT**

Financial support for this study was provided by NIH grant R21AI106329-01 to MNH.
LEGENDS TO FIGURES

Figure 1 Comparison of top 10 sequences from SELEX experiments to select aptamers that recognize GP1,2Δmucin and full length GP1,2

A,D) Flow charts depicting the sequential selection steps for each SELEX experiment. Downward arrows represent positive selection steps and horizontal arrows represent negative selections. B,E) Frequency chart of SELEX 1280 (B) and SELEX 1281 (E) depicting the frequency of singletons, enriched and unique fraction. C,F) Frequency chart of the random region of each round of SELEX 1280 (C) and SELEX 1280 (F)

Figure 2 Selection through a sucrose cushion

The oligonucleotides listed in the abscissa were spun through a 20% sucrose cushion in the presence of either control VSV particles or VSV particles pseudotyped with GP1,2Δmucin. The oligonucleotides in the pellet were sequenced to determine the fractional frequency of each oligonucleotide. The dashed at 0.1 represents the fractional frequency expected if an oligonucleotide were not preferentially pelleted with the particle. The asterisks identify two oligonucleotides that preferentially pelleted with the VSV particles pseudotyped with GP1,2Δmucin

Figure 3 Filter capture assay to estimate the affinity of oligonucleotide 4789 for GP1,2Δmucin

A) Titration of full length GP1,2 and GP1,2Δmucin recombinant proteins against 2 nM 32P-4789 2’FY-RNA by filter capture assay to establish the Kds, which was ND (not determined) for GP1,2 and 143nM GP1,2Δmucin, respectively B) EMSA analysis with 5uM 32P-4789
2’FY-RNA and 0.5 μM BSA, 0.5 μM GP1,2 delta mucin or 64 ng/ul VSV –GP1,2Δmucin virus particles. The arrows show the positions of the mobility shifted RNA associated with virus particles (V) and recombinant protein (P). C) Titration of GP1,2Δmucin-pseudotyped VSV virus particles or VSV virus particles lacking GP1,2 against 10nM ³²P-4789 2’FY-RNA by filter capture.

Figure 4 Filter capture assay to determine the affinity of oligonucleotide 5185 and 5187 for GP1,2

A) Titration of full length GP1,2 against 10 nM ³²P-5185 2’FY RNA measured by the filter capture assay from which the estimated Kd was 50nM. B) Titration of full length GP1,2 pseudotyped VSV particles against 10 nM ³²P-5185 2’FY RNA measured by the filter capture assay. C) Titration of full length GP1,2 against 10 nM ³²P-5187 2’FY RNA measured by the filter capture assay from which the estimated Kd was 560nM.
REFERENCES


Comparison of top 10 sequences from SELEX experiments to select aptamers that recognize GP1,2Δmucin and full length GP1,2.
Selection through a sucrose cushion
Figure 3

Filter capture assay to estimate the affinity of oligonucleotide 4789 for GP1,2Δmucin
Filter capture assay to determine the affinity of oligonucleotide 5185 and 5187 for GP1,2
CHAPTER 4

CONCLUSION AND DISCUSSION

Ebola viruses are highly pathogenic viruses that infect humans and non-human primates with very high mortality rates. The first Ebola virus outbreak was reported in 1976 and there have been sporadic outbreaks reported over the years. However, there still aren’t any commercially available vaccine or therapeutic drugs against Ebola viruses. During the devastating outbreak of Ebola Viruses in 2014, many candidate vaccines were introduced and are still in clinical trials. In addition to vaccines, several therapeutic drugs against Ebola virus were also expeditiously put under investigation for efficacy that are still in clinical trials. Aside from the lack of tested therapeutics for Ebola virus, another pressing concern is the inability to detect early virus infections. The available diagnostic kits are mostly antibody based and detect the serum antibodies against Ebola virus, which take time to appear, or the viral RNA, which requires correct timing of blood sampling. [1, 2]. Due to limitations, such as those just mentioned, associated with the current detection kits there is still a need for robust, cost effective and sensitive technique for early detection of Ebola viruses. Aptamers have been established as a potent alternative to antibodies. Several aptamers have been selected by others to function either as an antiviral or for biomolecule detection [3-5].

This thesis describes efforts to select and characterize 2’FY-RNA aptamers to Ebola Glycoproteins (sGP and GP1,2) with the expectation that the aptamers will have applications for intervention of viral host interaction and/or integration onto a biosensor platform. We chose the full-length GP1,2 recombinant protein, GP1,2 mucin domain deleted glycoprotein and sGP
as the target proteins for selection to obtain aptamers that could target both the surface exposed
GP1,2 and the intracellular mucin domain deleted GP1,2. Three SELEX experiments were
performed 1280 (GP1,2Δmucin), 1281 (full-length GP1,2) and 1282 (sGP). Each SELEX
experiment was performed using an oligonucleotide pool with a 53-nucleotide random region.
Three separate pools were used and at different time intervals. Enrichment of unique clusters
was evident in all three SELEX experiments when the frequency plots of initial round pool and
final round pool was compared. A sequence comparison of the results of the three SELEX
experiments revealed identical sequences in all three. Most of the sequences in SELEX 1281
(Chapter 3) and SELEX 1282 (Chapter 2) were identical. Oligonucleotide (4797) was found
in all three selections but bound none of the targets. Thus, oligonucleotide 4797 is
likely to
bind the background matrix that was common to all three SELEX experiments. One reason for
the appearance of identical sequences in 1281 and 1282 was the similar tertiary structures of
full length GP1,2 and sGP.

The observation of the same sequences appearing in two SELEX experiments was
unexpected and led to the question of the probability that same sequence could appear in two
separate DNA pools in the absence of evolution of the sequences driven by selection. The
probability of having oligonucleotides with identical sequences in separate pools is very low.
The frequency plots of the RNA pools were 99% unique, which eliminates the hypothesis that
sequences of oligonucleotides in the starting pools were redundant due to bias in the synthesis
step. Another source of bias could be in the sequencing step. The complexity of the pool is
$10^{15}$ oligonucleotides representing a subset of the possible $10^{31}$ sequences and the NGS results
average a million reads per pool. Other points of bias that could increase the representation of
certain sequences are the transcription and PCR steps. However, none of these sources of bias provide a credible explanation for the fact that many of the same sequences were isolated in two independently run SELEX experiments, which started with different RNA pools. The results could be explained by the combination of selection against target and the variations in sequence imposed by low fidelity amplification protocol. If this is the explanation, it also suggests that the initial selections (round 1 and 2) are very efficient in capturing oligonucleotides with some affinity for the target. The remaining rounds of SELEX may be dominated by the evolution of sequences due to errors in replication and transcription, which expands the oligonucleotide populations that bind with high affinity. These results also suggest that the selected aptamers are either unique or amongst very few possible oligonucleotides that are capable of binding sGP with high affinity.

Chapter 2 discusses the selection of high affinity 2’F RNA aptamers that bind to sGP with high affinity and specificity. MEME software was used to search for sequence structure motifs in the oligonucleotides for which sequences were found in the SELEX 1281 and SELEX1282 top clusters. We identified polyU rich sequences in the selected oligonucleotides from both SELEX experiments. Poly U rich sequences are known to bind intracellular regulatory RNA binding proteins [6]. We found that the oligonucleotides 5177, 5182, 5183 and 5179 with poly U rich sequences bound sGP with high affinity, whereas oligonucleotides 4789 and 5181 lacking poly U rich sequences bound with low affinity. We tested the dependency on 2’FY modification for high affinity binding and found that unmodified 2’OH, 5183 RNA did not bind sGP. From this result, we concluded that binding of oligonucleotide 5183 to sGP depends on the 2’fluoro modification of the pyrimidines. This observation is
consistent with earlier reports of enhanced affinity of RNA aptamers to target proteins due to 2′F modification [7].

To identify the aptamer sequence in the larger 5183 oligonucleotide, oligonucleotides with sequences that were truncations of the parent 5183 sequence were tested for binding to sGP. These results confirmed the central role of the polyU-rich loop in binding sGP as the truncated 39mer containing the hairpin U loop sequence bound sGP most effectively. This 39mer was demonstrated to bind sGP with high affinity and was given the aptamer name 39SGP1A. Our long-term goal is to integrate the selected aptamer on a detection platform. Therefore, we tested the affinities of this aptamer for the serum proteins, human serum albumin, fibrinogen and α2-macroglobulin.

It is important for the application of 39SGP1A to diagnostics that it not bind to other proteins in the blood. The parent proto-aptamer, 5183 binds sGP with a Kd of 400 nM compared with a Kd of 40 nM for sGP. The concentration of human serum albumin (HSA) in blood is 300 μM, which means that the 5183 oligonucleotide would be saturated with HSA if incubated with a sample of serum. However, once the sGP binding element was isolated to the poly U stem and 5183 truncated to create the aptamer 39SGP1A there was no remaining affinity for HSA.

In chapter 3, for selections against GP1,2 in SELEX 1280 and 1281 no conserved sequence structural motif was found. Therefore, the affinity of oligonucleotides with sequences in the SELEX 1280 selected families and unique sequences from SELEX 1281 were tested for binding to GP1,2 with and without the mucin domain attached. From these tests, oligonucleotide 4789 was found to bind GP1,2Δmucin with high affinity whereas it did not
bind the full length GP1,2 glycoprotein. This oligonucleotide also bound sGP (Chapter 1). Therefore, oligonucleotide 4789 appears to bind a domain that is common between sGP and GP1,2Δmucin, which is in the N terminal portion of the GP1,2 protein sequence. The N terminal region of GP1,2 and sGP is the conserved receptor binding domain suggesting that 4789 binds the receptor binding domain of the glycoprotein. This aptamer might have the capability of intercepting viral host interactions. Tests of other oligonucleotides selected in SELEX 1281 for binding to the full length GP1,2 (including the mucin domain) identified two sequences 5185 and 5187 that bound the full length GP1,2 and virus particles pseudotyped with the full length GP1,2.

My thesis describes the development and characterization of sGP and GP1,2 specific 2′FY-modified RNA aptamers. Additional studies are needed to determine if the aptamers selected against GP1,2 might provide therapeutic options. First, the specificity of these 2′FY-RNA aptamers must be determined by testing their ability to bind other proteins and to bind GP1,2 in the presence of blood proteins. The selected aptamers to full length GP1,2 could target extracellular virus and the aptamers that recognize GP1,2 lacking the mucin domain might target the cleaved GP1,2 protein in the endosome. Because Ebola viruses have GP1,2 independent mechanisms for entry, full length GP1,2 and GP1,2Δmucin aptamer chimeras could be effective. The GP1,2 RNA aptamer when bound with full length GP1,2 glycoprotein can facilitate the entry of both aptamers into the endosomes where the second aptamer to GP1,2Δmucin could thwart viral entry into the cytoplasm. The challenge at this step could be if these aptamers are unable to survive low pH endosomal conditions, which has not yet been tested. An alternative approach could be to conjugate these aptamers to known therapeutic
drugs for intervention of viral entry. An aptamer that neutralizes EBOV will provide a means for rapid, albeit transient, protection as a stop-gap measure prior to further treatment including isolation of the infected individual from others who might be susceptible.

Our future goal for the selected aptamer 39SGP1A is to use it to detect sGP. The aptamer will be integrated onto microcantilever a platform that is equipped with a system to detect RNA aptamer-sGP interaction with high sensitivity and specificity such as was previously demonstrated with mouse lipocalin-2 [8]. Although, we have tested the aptamers for its specificity against human serum albumin, there are many known proteins present at various concentrations in the serum. Therefore, binding of these aptamers to sGP in the presence of serum should be evaluated. Aptamer efficiency is also determined by its tertiary folding, therefore following the integration of the selected aptamer onto a platform, it must be ensured that folding of the aptamer hasn’t been compromised.

In summary, the work presented in this thesis focuses on the selection and characterization of 2’FY-RNA aptamers against Ebola viral glycoproteins. Results from the selections reveal various important factors responsible for the selection of high affinity aptamers. These include considerations with respect to designing the pool, ensuring evolution of oligonucleotide sequences and the importance of the 2’F modification in mediating high affinity interactions with proteins. These studies resulted in the identification of aptamers that bind sGP and GP1,2, which can be developed as antivirals or for detection of Ebola virus.
REFERENCES


