6-23-2023

Leveraging protein dynamics to drug filovirus protein-nucleic acid interactions using simulations and experiments

Matthew Alberto Cruz
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Biochemistry, Biophysics, and Structural Biology

Dissertation Examination Committee:
Gaya K. Amarasinghe, Chair
Gregory R. Bowman, Co-chair
Eric A. Galburt
Carolina B. Lopez
Janice L. Robertson

Leveraging Protein Dynamics To Drug Filovirus Protein-Nucleic Acid Interactions Using Simulations And Experiments

by
Matthew Alberto Cruz

A dissertation presented to
Washington University in St. Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2023
St. Louis, Missouri

Table of Contents

List of Figures

List of Abbreviations

ADP - Adenosine Diphosphate
ATP - Adenosine Triphosphate
ATPase - Adenosine Triphosphatase
CARDS - Correlation of All Rotameric and Dynamical States
CD - Circular Dichroism
CPU - Central Processing Unit
Cryo-EM - Cryo Electron Microscopy
DBD - DNA Binding Domain
DNA - Deoxyribonucleic Acid
DMSO - Dimethyl Sulfoxide
dsRNA - Double Stranded RNA
DVG - Defective Viral Genome
DTNB - 5,5′-Dithiobis-(2-nitrobenzoic Acid)
eVP35 - Zaire Ebolavirus VP35
EC50 - Half Maximal Effective Concentration
FAST - Fluctuation Amplification of Specific Traits
FDA - Federal Drug Administration
FITC - Fluorescein Isothiocyanate
FP - Fluorescence Polarization
F@h - Folding@home
gRNA - Genomic RNA
GP - Glycoprotein
GPCR - G Protein Coupled Receptor
HA - Hemagglutinin
HSQC - Heteronuclear Single Quantum Coherence
IFN - Interferon
IID - Interferon Inhibitory Domain
IMSD - Institute for Maximizing Student Development
IRF-3 - Interferon Regulatory Transcription Factor 3
ITC - Isothermal Titration Calorimetry
L - Large Subunit of the RDRP
LBD - Ligand Binding Domain
MD - Molecular Dynamics
mL - Milliliter
mM - Millimolar
mRNA - Messenger RNA
MSM - Markov State Model
mVP35 - Marburg Marburgvirus VP35
NMR - Nuclear Magnetic Resonance
NIH - National Institutes of Health
NNSV - Negative Sense Non-Segmented Single Stranded Virus
NPBP - Nucleoprotein Binding Peptide
NR - Nuclear Receptor
NSF - National Science Foundation

nL - Nanoliter
nm - Nanometer
nM – Nanomolar
ns – Nanosecond
OD - Oligomerization Domain
PAMP - Pathogen Associated Molecular Pattern
PCR - Polymerase Chain Reaction
PDB - Protein Data Bank
PNI - Protein Nucleic Acid Interaction
PPI - Protein Protein Interaction
RDRP - RNA Dependent RNA Polymerase
RNA - Ribonucleic Acid
RLR - RIG-I Like Receptor
RIG-I - Retinoic Inducible Gene I
SAR - Structure Activity Relationship
SASA - Solvent Accessible Surface Area
SPR - Surface Plasmon Resonance
TCEP - Tris (2-Carboxyethyl) Phosphine
TEV - Tabacco Etch Virus
TF - Transcription Factor
TNB - 2-Nitro-5-Thiobenzoate
µL - Microliter
µM - Micromolar
mM - Millimolar
M - Molar
UV - Ultraviolet
VP24 - Viral Protein 24
VP30 - Viral Protein 30
VP35 - Viral Protein 35
VP40 - Viral Protein 40
WHO - World Health Organization
WUSTL - Washington University in St. Louis
ZINC - ZINC Is Not Commercial

Acknowledgments

Of course, this work is not something one does all alone. First, I give honor to the God and Father of our Lord Jesus Christ through the intercession of Mary, the Blessed Mother and SS. Joseph, Albert the Great, Oscar Romero, and Francis De Sales. We Catholics love to describe saints as our friends in heaven and this is very true, I have felt their prayers carrying me through the hard parts of graduate school. The sweet comfort of Jesus in the eucharist strengthened me for this journey.

Next, I must give thanks to my parents Silvio and Stephanie Cruz. Without them, their support, and their encouragement, I would not have made it through high school, undergraduate or graduate school. My dad, Silvio, is always there to offer a quick kayak trip to relax, make me a carne asada, or just talk about life. My mom, Stephanie, has always been encouraging, reminding me of what is important, goading me about living nearer, and of course indulging my affection for my cat. I hope this accomplishment makes you the proudest parents ever.

To my family, my sisters Andrea and Chasteri, my brothers-in-law Nicholas and Chancey, my nephews and nieces, Jeremiah, Makayla, Julian, Elijah, Johnny, Elliana, Maribella and Dawson, thank you all for the annual trips, the laughter and teasing, the cooking competitions, and constant photos. It is one of my most joyful experiences to watch each of the kids grow as a person and into their distinct personalities. To Granny Lu, I love and miss you; I wish you could be here but I am happy for the time we had. To Grandpa Holley, I love and miss you and am so grateful for your support throughout my life. A mi abuelita Daisy, te quiero mucho, te extraño; ojala que seas orgulloso de mi y deseo de compartir cada exito y prosperidad con ti. To all my aunts and uncles and cousins–your support means the world and you are all rocking it in your worlds too.

In addition to my blood family, I am blessed with a bevy of loving and supportive chosen family members. Claire, you're my sister and George is my nephew and I love you both. Thank you for your constant and meaningful friendship. To Cassy and Alex, Ariyaneh, and Ali, you all probably know the most about my whole graduate school journey because I always felt comfortable complaining to Cherry Emoji Twitter. Thank you for spare beds and mattresses, shared meals, trips, drinks, weddings, and love. I can't imagine my life without our little family, and I look forward to everything we have in the future.

In graduate school, my cohort in Biochemistry, Biophysics, and Structural Biology became my instant family. Jasmine, Jhullian, Paige, Victoria, Rose, Mickey, Kellan, Huiming, and Naomi thank you for keeping me sane, for hearing my complaints, dealing with my flair for the dramatic, driving me places, celebrating my birthdays, honing my scientific skills, making me laugh, cooking me dinners, playing video games, and tolerating my insistence on watching random music videos on YouTube at 1 am. I look forward to the creative ways we'll find to spend time together as we all move on from graduate school.

To my friends from my parish, catholic STL and beyond, Sarah Kate Heyman, Chelsi Creech, Philip Masini, Britt Johansson, and Jane Rhoda. Thank you for supporting me, keeping me normal and sane, laughing with me, and praying for me.

In my professional life, I have had the distinct pleasure of working alongside so many scientists who truly made me a better scientist. My pre-AP and AP biology teacher, Ms. Sansom: thank you for having patience with me, humoring my confusions and my interests, and making biology fun in your iconic way (BAM!). To Dr. Patterson, my pre-AP and AP chemistry teacher, you changed the course of my life with your teaching, being a Rice Alumna, and your excellent labs. Because of you chemistry was not scary, it was exhilarating and concrete; you convinced me that

I could be a scientist, and I am so grateful. Once I got to Rice, I received such incredible mentoring from so many scientists: Kristi Kincaid, Shirley Liu, Margaret Lie, Kendall Burks, Madeline Hewitt, Joseph Xu, Emily Thomas, Jaskeerat Gulati, Joff Silberg, and Matthew Bennett. Thank you all for your various contributions to my scientific development. You each taught me that a scientist does not come from a mold but rather to just be myself. You each exposed me to new ways of thinking, and exciting new research areas. Principally, at Rice, I received incredible mentorship from Drs. Kathleen Matthews and Beth Beason Abmayr. Dr. Matthews taught me how to be resilient in the face of difficult and confusing scientific problems, how to give credit where its due and spread the joy around. Drs. Matthews and Beason taught some of my courses that singularly influenced my scientific trajectory. Thank you for investing in me.

During my graduate work, I have collaborated extensively within and without the lab and I thank everyone who ever helped me along the way. Thank you to my lab mates Sukrit, Max, Justin Porter, Catie, Tom, Neha, Mick, Upasana, Jeff, Borna, Bhakat, Prajna, Justin Miller, Artur, Ayan, Louis, Shahlo and Devin. Each of you really did share in the creation of this knowledge and I am so grateful. Thank you to Rishi Samarth, my intrepid undergraduate mentee, I know it was hectic, but I was happy to teach you what I could.

Catie and Sukrit, I give you special thanks for the role you have played in my graduate development. Without you I would have so much less scientific acumen, I would have not had as much fun, nor would I have made it through graduate school. Thank you for being supporters, challengers, comforters, and all around hype-people.

To Greg, my graduate advisor, thank you so much for taking on the risk that was someone who thought they did not like physics, did not know how to code, and who at the ripe experience level

of a first-year graduate student declared ‘I want to work independently.’ I am grateful for a mentor who is generous, forward when correcting, willing to be disproven, excited with small and big wins, and generally supportive. I know I never made your job easy, yet I never heard you complain about me.

To my thesis committee, Gaya, Eric, Janice, and Carolina. Thank you all for the constant support, encouragement, direction and reality checking when needed. Gaya thank you for accepting and addressing all of my VP35, NMR, and crystallography questions. Eric, thank you for letting me rotate in your lab when I knew nothing about single molecule work or RNA polymerase. Thanks for being there from literally day one of graduate school, on my qualifying exam(S), and my thesis proposal, your feedback and support has been invaluable.

To Dr. Ilagan and Mike in the high-throughput screening center, thank you for everything you did for the screening project to have gone anywhere. Thanks for handling my excessive questions, always having 96-well costar black half area flat bottom microplates when I ran out, and for letting me learn about all the different equipment and what goes into a high-throughput screen.

Many thanks are owed to Drs. John Cooper, and Ben Garcia as the department chairs of the biochemistry and molecular biophysics department at Wash U. Thanks for keeping the ship afloat, valuing graduate students’ opinions, and building a supportive department. Thank you also to Melissa Arenz, Jayma Mikes, Casey Critchlow, Nicholas Caito, and Ginny Ribuado for administrative support, enthusiasm and laughter, and making sure everything keeps moving forward efficiently.

I also owe much gratitude to the Biochemistry and Biophysics department at the University of Pennsylvania, thank you to Kristen Lynch for orchestrating a welcome for the Bowman Lab, to Ronen Marmorstein for letting the Bowman lab crash Structure talks on Friday.
Thanks also to Sebastian Dilones, you have already become such a great friend and I look forward to your successes. Thank you for befriending me and for all the reality TV discussions among others.

This research is made possible through many funding mechanisms and awards I am privileged to have received. Funds awarded to Gregory Bowman: NSF CAREER Award MCB-1552471 and NIH grant R01 GM124007. Funds awarded to myself: NIH grants 5R25GM103757 to WUSTL IMSD program, and NIH F31AI157079 and a microgrant from the WUSTL HTSC. I was also supported with funding from the MilliporeSigma fellowship, and the Elliot Elson fellowship.

Also, in memory and thanks for those who are not with us anymore and for their friendship while they were here: Grandpa Holley, Granny Lu, Kelley Lash, and Colin Kleunder, thank you.

Finally, thank you to my silly cat Monsieur Moreau, you are often annoying, always so sweet, and constantly comforting, for which I am grateful.

Matthew Alberto Cruz

Washington University in St. Louis

August 2023

Dedicated to my parents Silvio and Stephanie Cruz

ABSTRACT OF THE DISSERTATION

Leveraging Protein Dynamics To Drug Filovirus

Protein-Nucleic Acid Interactions Using Simulations And Experiments

by

Matthew Alberto Cruz

Doctor of Philosophy in Biology and Biomedical Sciences

Biochemistry, Biophysics, and Structural Biology

Washington University in St. Louis, 2023

Professor Gregory R. Bowman, Chair

Professor Gaya Amarasinghe, Co-Chair

Drug discovery targeting protein nucleic acid (PNI) and protein-protein interactions (PPI) remains a difficult task. How to identify druggable conformations and discover compounds binding to these conformations is an unsolved problem. Here, I describe how we apply computational and experimental approaches to probe the conformational landscape of a key immune antagonist protein from the Zaire ebolavirus. Viral Protein 35 (VP35) binds to the viral pathogen associated molecular pattern, double stranded RNA (dsRNA), and blocks activation of the interferon (IFN) response. Long time scale molecular dynamics simulations reveal that VP35 adopts an alternative conformation which opens a pocket absent in experimental structures, a cryptic pocket. Simulations predict and we experimentally validate that this pocket exists and is allosterically coupled to the dsRNA binding site. High throughput screening identified new chemical matter that inhibits dsRNA binding. Subsequent structural and biochemical studies show that the inhibitors target the dsRNA binding site. One inhibitor also binds to the VP35

cryptic pocket. Our data shows the usefulness of considering protein conformational heterogeneity for drug discovery targeting PNIs and PPIs.

Chapter 1: Introduction

1.1 Proteins Play A Central Role In All Cellular Processes.

From the elephant to the extremophile archaea surviving off sulphur at hydrothermal vents on the ocean floor[1], proteins are indispensable for all cellular life. Cells utilize proteins to catalyze metabolic reactions, transport molecules across the cell membrane, and to replicate their genetic material. After decades of intensive research, we now understand genetic information stored in DNA is first transcribed to make messenger RNA (mRNA), which then undergoes translation into a polypeptide chain which folds into its three-dimensional protein structure; not all proteins will fold into one discrete fold.[2] Protein folding is essential for proper protein function and remains a complex field of investigation in biophysics.

Though ubiquitous, proteins have evolved into several general classes responsible for performing critical cellular functions. Understanding each of these classes' folding, function, cellular localization, and evolution, is an ongoing effort in biochemistry and biophysics. These major classes of proteins include enzymes, receptors and channels, and transcription factors among others. Enzymes are well known and studied types of proteins that act as catalysts for biological reactions. Enzyme functions range from metabolic breakdown in glycolysis, to managing protein homeostasis via the proteosome or DNA repair in the DNA damage response. Enzymes are found in every cellular context from the mitochondrial membrane to the lysosome and the nucleus. Comparatively, receptors are non-enzymatic proteins usually integrated into the cell membrane where they bind to effector molecules (agonists) which trigger changes in the receptor's structural arrangement activating downstream signaling or other events such as membrane depolarization. Similarly, transcription factors (TFs) act as sensors through binding to signaling molecules and subsequently binding to or releasing from DNA. Whether a TF is bound

or unbound regulates gene expression through recruiting RNA polymerase and associated factors or through preventing binding of such initiation machinery. Occasionally TFs also act to recruit other TFs such as in the case of enhancers that recruit mediator which is required for some kinds of transcription initiation in eukaryotes.[3] While the roles proteins play in cellular processes are increasingly well described, common biophysical models of these interactions often ignore the chemical and physical necessities for function. For example, it is common to describe receptor binding to ligands as a lock and key mechanism where the protein's structure accommodates just the right chemistry of the ligand for proper function. This is patently false since proteins exist in dynamic equilibria between various conformational states whose chemical and physical properties are important for ligand binding and function such as in the well-studied hemoglobin.[2]

1.2 A Protein's Conformational Dynamics Are Required For Function.

In organic chemistry, the relative orientation of atoms in a molecule, along covalent bonds and through space, describe its overall conformation. Through bond vibrations, the orientation of these atoms may change with the probability of orientation changing being associated with the relative energies of the starting and subsequent orientations. In the simplest terms, a conformational change is the transition of a bonded pair's relative orientations to a new orientation. For proteins that have orders of magnitudes more atoms than simple organic molecules, the description of a given conformation is highly complex and often described as a conformational state meaning that state is associated with the various conformations of the individual bonds comprising the protein. Then for a protein, a conformational change is the conversion from a set of conformations for every bond in the protein, to a new set of relative orientations. This may happen locally i.e., not to all bonds, or globally. Most often a global

conformational change will be specified as such while local conformational changes may be referred to simply as conformational changes with the 'local' descriptor assumed. Special cases of such language occur when discussing protein folding. Protein folding describes the transition from the unstructured polypeptide which adopts a variety of high energy conformations, to an ensemble of low energy stably structured states. A global conformational change may also be global unfolding and a local conformational change may also be local unfolding. But not all conformational changes (local or global) are to the unfolded state which specifically refers to a largely unstructured ensemble of states while a global conformational change may be from one folded state to a new, distinct, yet still folded state such as the case of fold-switching proteins, or the simpler, channel open to closed state transition.

Conformational dynamics may denote the *rate* of conversion between two conformational states but equally refers to the thermodynamic perspective wherein a protein adopts several conformational states of Boltzmann distributed population at equilibrium. The relative populations of these states may be important for protein function. Kinetic and thermodynamic descriptions of conformational changes are both important for comprehensive understanding of a protein's function.

The role conformational dynamics play in protein function is well studied in the context of enzymes, receptors, and to some extent, transcription factors.[4] Some well-studied examples of dynamics being important for function include the effects of conformational changes on enzyme function such as in β lactamases[5], Influenza Hemagglutinin, motor proteins, etc. Additionally, in the context of GPCRs[6] and ion channels[7] it is known that conformational changes govern whether the receptor is active or inactive, or the channel is open or closed.

While many metabolic enzymes and their conformational changes associated with function are well studied.[8-10] β lactamase is a useful model system because it is localized in the periplasm of the bacterial cell and there it hydrolyzes β-lactam containing antibiotics to prevent cell death from the antibiotic such as benzylpenicillin. In previous work, the balance between β-lactamase's conformations that can accommodate a larger substrate and those that cannot, described the difference in catalytic profiles between two different substrates.[5] Also, the opening of a cryptic pocket is a local conformational change important for the capacity of β-lactamase to hydrolyze penicillins.[11]

The Influenza A Virus (flu) poses a severe health threat due to its high mutation rate, reassortment and recombination capacity which generate many new and hard to predict variants.[12] Flu variants' pathogenicity naturally varies from season to season and predicting these changes remain an important ongoing area of study. For flu to infect host cells, the viral protein hemagglutinin (HA) interacts with sialic acid triggering clathrin mediated endocytosis. Acidification of the virus containing endosome triggers a conformational change in HA that fuses the viral membrane with the membrane of the endosome releasing the viral capsid containing the genomic RNAs.

The membrane fusion process in flu is dependent on the acidification of the endosome triggering conformational changes in HA. HA contains two subunits HA.1 and HA.2 and HA.2 contains a peptide that in the pre fusion state is bound to a pocket in HA.2. When the endosome pH reaches 5.0-6.0 the fusion peptide releases from the hydrophobic pocket on HA.2 followed by HA.1 and HA.2 dissociation. The fusion peptide then inserts into the membrane acting as a tether of sorts for the HA.2 fusion conformational change in which HA.2 adopts a new fold that is more stable

at lower pH. HA.2's refolding pulls the endosomal and viral membranes together causing fusion and capsid release.[12]

Motor proteins are well studied examples of conformational dynamics being tied to protein function. Myosins are protein superfamily with a wide array of activities including cellular transport and organization, contractility, and motility. Most simply, myosins are motor proteins that hydrolyze adenosine triphosphate (ATP) to generate mechanical force. A few well studied examples of Myosin include beta cardiac myosin, and smooth muscle myosins.[13] In the sarcomere of muscle tissue, Myosin II interacts with actin to pull on the actin filament and generate force for things like muscle twitching or heartbeats. To bind to actin and generate force, a series of steps must occur each with certain associated conformational changes. Myosin II has two important domains for the generation of force. The first is the head domain that binds to actin directly and harbors the ATPase activity. Second is the lever arm which undergoes a conformational change from an 'up' state perpendicular the actin, to a 'down,' state with the lever arm moving closer to the actin (60 degree change). The transition from the lever arm up state to the down state after ATP hydrolysis comprises the power stroke. Myosin bound to available actin will unbind in the presence of ATP, undergo a conformational change and hydrolyze ATP into ADP and inorganic phosphate. Next, myosin bound to ADP and phosphate binds to actin, and phosphate is released followed by the power stroke and finally ADP release. In the presence of ATP, myosin will unbind from actin reinitiating the cycle. For this whole cycle to happen myosin undergoes dramatic conformational rearrangements in the relative orientation of the lever arm, and within the head domain. There are at least nineteen myosin classes and the conformational dynamics described here are proper to the Myosin II class although some features may be ubiquitous. The large number and distribution across various

species and classes of life demonstrate the importance of considering conformational dynamics in protein function.

Conformational dynamics are not just important in the globular protein context but also in the membrane, perhaps even more so. GPCRs are integral membrane proteins comprised of usually 7 transmembrane alpha helices which bind to extracellular agonists (ligands) and undergo conformational changes making them amenable to interactions with their intracellular binding partners, G proteins.[6,14-16] Binding of the GPCRs to G proteins induces guanidine exchange triggering a series of downstream signaling events. Upon binding of a ligand to a receptor, helices 5-7 undergo a variety of changes including shifting, tilting, bending, and twisting to transition the receptor from its inactive to its active state. The receptor can adopt a variety of different states between the active and inactive conformations and the populations of these states are important for the receptor function.[6]

Similarly, ion channels are integral membrane proteins usually composed of a series of alpha helices, similarly to GPCRs.[7] Critically ion channels harbor a central pore through which ions may pass into the cell. Ion channels are capable of discriminating cations independently of atomic charge e.g. Mg^{2+} vs Na^{+} and do so through the unique chemistry of the pore. Many ion channels harbor gating mechanisms to prevent ion influx. One such gating mechanism is the open to close transition. This conformational change enables cellular level control of ion flow in the context of neuron potentiation in the case of the sodium channel. Upon depolarization of the neuronal membrane due to an influx of sodium ions after ligand-mediated sodium channel activation, the helix 4 undergoes a conformational change to trigger opening of the sodium channel.

1.3 The Role Of Protein Dynamics In Protein Nucleic Acid Interactions Is Not Well Understood

Polymerases transcribe DNA into functional RNAs including long non-coding RNAs, and ribosomal RNA, among others and many motor proteins, transcription factors, splicing and ribosomal proteins interact with nucleic acids to function and maintain cell viability. Thus, it becomes the case that the interactions between proteins and nucleic acids are central to all cellular processes. While protein structure, function, and evolution are well studied, the conformational dynamics of proteins and how that relates to their function of protein-nucleic acid interactions remains poorly understood. Much work has been done to understand nucleic acid motors and the interaction between proteins and nucleic acids but the contribution of protein dynamics to non-specific binding, sequence specificity or topological preferences remains unclear. With the increase interest in intrinsically disordered regions and proteins, which are conformationally heterogeneous, new interest is evolving for an understanding of how protein conformational heterogeneity affects protein-nucleic acid interactions.[17] Naturally, understanding this dimension of an important class of interactions will have consequences for drug design and discovery, material sciences, and our understanding of biological processes.

Protein nucleic acid interactions play a key role in several biological processes of clear importance. First, is nucleic acid metabolism which ranges from such processes as replication, transcription, translation, repair, and splicing. The second important context for protein-nucleic acid interactions, is during viral infection. During viral infection, a key process is the replication of the viral genome which may occur using viral machinery, or host machinery. The capacity of host cells to identify the viral genome as exogenous material is an important step in activation of the immune response.

All cells must maintain their genome throughout the cell cycle and during replication and gene expression. All cells are also subject to their local environments which present a variety of insults (UV damage, chemical exposures, reactive oxygen species, etc.) that the cell must overcome. These insults often damage a cell's DNA or cause single or double stranded breaks. To repair this damage and breaks, the cells evolved DNA maintenance machinery such as SSB, uvrD, etc. which perform DSB resolution or Base-excision repair, through protein-nucleic acid interactions. For each of these processes, protein-nucleic acid interactions mediate the recognition of the insult, the recruiting of homologous DNA strands for templating, and the actual repair and replication of nucleic acids. It is known that non-enzymatic function of binding to the nucleic acid is critical for the fidelity of these processes.

A key interaction between viruses and hosts are the innate immune surveillance systems for detecting PAMPs. The RIG-I like receptors (RLRs) are one such class. Importantly, viruses must also engage with the host RNAs to perform other important metabolic processes such as cap snatching. Protein-nucleic acid interactions mediate all these processes.

There are a couple of examples wherein conformational changes in proteins are important for protein-nucleic acid interactions.

1.3.1 Transcription Factors Bind Effectors And DNA

Bacterial transcription factors like LacI were early studied examples of protein-nucleic acid interactions which underwent allosteric changes related to their nucleic acid binding activity. In LacI/GalR family of transcription factors, a small molecule allolactose interacts with and binds to LacI in its ligand binding domain (LBD).[18] Binding of the inducer allosterically changes the conformation of the DNA binding domain (DBD) promoting release of the DNA from the DBD.

This release enables transcription of the *lac* operon which contains genes important for lactate metabolism.

Similarly, in eukaryotes, Nuclear Receptors (NR) have LBDs and DBDs and sense a variety of molecules including vitamins and steroids.[19] While activation varies from class to class, in many classes, binding of the ligand to the NR promotes dissociation of corepressors or NR oligomer. This enables binding of the NRs to DNA and interactions with coactivators that induce gene expression (through recruiting RNA polymerase).

In both the LacI/GalR and NR families, this shared mechanism of allosteric activation to modulate gene expression requires endogenous evolved ligands. This enables straightforward repurposing of native ligands for biomedical or biotechnical purposes. Such examples include IPTG (non-hydrolysable allolactose), or antiandrogens targeting the androgen nuclear receptor. The mechanism of these allosteric transitions remains an active area of study with conformational dynamics-based descriptions of the LacI mechanism only recently published and similar studies on NRs remain ongoing.

1.3.2 Non Enzymatic Factors Play Big Roles Through Binding Alone

Intrinsically disordered proteins generally lack the structure necessary for enzymatic activity and as such their primary function is to bind to other proteins and nucleic acids and in some case nucleate phase separated membranelles organelles. These cellular structures are critical in gene transcription, translation, viral replication, assembly, cell motility, and the stress response. These examples support the importance of studying protein-nucleic acid interactions and the contribution of conformational heterogeneity to protein function

1.4. Many Well Described Techniques Exist For Studying The Relationship Between Protein-Nucleic Acid Interactions And Conformational Dynamics

The techniques available for studying PNIs overlap strongly with those use to study protein folding generally and include many spectroscopic techniques. In this thesis I employ a wide range of spectroscopic techniques but the atomistic detail use to interpret experimental data largely comes from molecular dynamics simulations.

1.4.1 Molecular Dynamics Simulations Enable Rigorous Modeling Of Protein Dynamics And Predictions Of Functional Impacts.

Molecular dynamics (MD) simulations apply physics-based calculations and chemistry informed restraints to model biological molecules' dynamics and function.[20-23] Many types of MD exist ranging from unbiased atomistic simulations to coarse grained and enhanced sampling methods. Atomistic simulations enable our studies of protein conformational changes and functionally relevant changes in protein chemistry (side chain dihedral angles) or overall structural properties (secondary structure movement etc.) A balance that physics-based simulations of proteins must strike is between exploration i.e. sampling new states, and exploitation i.e. sampling states frequently enough to enable statistical analysis.

In unbiased simulations, an experimentally derived protein structure is solvated with water and ions to neutral charge at atomistic detail. Then, energy minimized and equilibrated using algorithms such as gradient descent and then a sufficiently long simulation to allow the system to be equilibrated. Then, several clones of the starting state from equilibration are each perturbed with randomly forces and left to simulate for some time by repeatedly generating forces, calculating velocities and constraints, saving to disk, and repeating at a new time step. With sufficient time, these simulations sample different conformations of the protein of interest.

Further analyses of specific residues' dynamics, application of machine learning algorithms, or visual inspection of the trajectories, can enable hypothesis testing of protein function, or observation of protein folding, and even observation of ligand binding.

An additional note is appropriate here about computational resources for MD. The calculations involved in such simulations are not computationally cheap and best conducted on state-of-the-art graphics processing units (GPUs) and central processing units (CPUs). Toward that goal, the Folding@home (F@H) platform was developed to enable citizen-scientists to donate their compute resources of their home machines toward research using MD. The sum resources of F@H represent the largest computing resource on the planet, capable of performing more floating point operations (FLOPs) than government-run supercomputers such as Frontera.[24] Access to F@H enables researchers to collect hundreds of milliseconds of simulation time for small to medium sized proteins. With this resource, it becomes possible to sample longer timescale events in protein dynamics including protein folding, and ligand unbinding. The conformational changes associated with important protein functions, do not often occur on very fast timescales (though some do) as such sufficiently sampling a protein's conformational landscape, with atomistic unbiased simulations, remains an important and challenging task even with such a resource as F@H.

Another way of using physics-based simulations to sample a protein's energy landscape, is through enhanced sampling methods which prioritize the exploration of a given protein's landscape over frequently visiting a specific state (exploitation). To enable exploration, many simulation strategies have evolved. One approach involves running short (less than 50 ns) simulations, using some criterion to decide which states sampled in first round of simulations to restart as new simulations. Occasionally the new simulations may even be set up in new

conditions such as in replica exchange. The various strategies of enhanced sampling are reviewed elsewhere and for this thesis, I will briefly describe a popular algorithm previously developed in the lab, Fluctuation Amplification of Specific Traits (FAST).[23]

FAST leverages a statistical measure of exploration to balance exploration and exploitation to rapidly explore a protein's conformational landscape while maintaining thermodynamic and kinetic properties. This is accomplished through the application of Markov State models. After an initial round of simulations, the simulation frames are clustered using a hierarchical clustering method such as k means. Then, the clustered data is used to build a Markov State Model and calculate the relative populations of the various states in the model and a user-defined property termed the 'directed' component. According to the states' populations i.e. the statistical component, and the directed component, new states are selected from which new simulations are started. These simulations are then subjected to the same clustering, MSM building and restart for a user-defined number of runs. At that point, a reasonable (0.5 microsecond) dataset of unbiased atomistic simulation data can be clustered a final time, and a final MSM constructed representing the relative populations of states in the protein's conformational landscape. This approach enables rapid comparisons of differences between protein variants' landscapes suggesting hypotheses to explain or predict functional differences.

Molecular dynamics can serve as a computational microscope as such for observing protein dynamics. Then the analyses (such as MSMs) comprise an important step in producing meaningful conclusions from large i.e. F@H datasets, and from smaller (adaptive sampling) datasets. The combination of MSMs and long-timescale simulations enables hypothesis and prediction generation through calculation of experimentally observable changes in protein structure and function.

1.5. Current Work Demonstrates Proteins' Dynamical Changes Are Targetable For Drug Discovery.

Using structures predicted from MD, researchers identified a new conformation in HIV integrase which exposes a trench.[25] Initial docking studies suggested this trench was druggable and later experimental work revealed that inhibitors do bind to the trench. One such inhibitor is now FDA approved under the name raltegravir.

A large wealth of structural information on G-protein coupled receptors (GPCRs) which are responsible for activation of many biological processes exists. These structures reveal that the active and inactive proteins exist in distinct conformational states. Later work revealed that in addition to the agonist binding site, there are a variety of orthosteric (allosteric) binding sites for which some antagonists and inverse agonists exist. These antagonists alter the population of active and inactive states. A high percentage of current FDA approved drugs target GPCRs including some orthosteric binders.[14]

Outside of HIV integrase, targeting enzymes involved in various disease processes at allosteric sites has increased in importance as antibiotic and cancer drug resistance mutations increase. Molecules identified targeting an allosteric site in TEM-1 β-lactamase bind to a previously unidentified cryptic pocket similarly to HIV integrase. [26]

[other examples of Allosteric inhibitors here]

1.6 Filoviruses Are Public Health Threats

Discovered in 1976 and the subject of public hysteria, World Health Organization (WHO) coordinated public health and medical efforts, decades of scientific research, and the popular novel *The Hot Zone*, Filoviruses remain a public health threat. Notably since large outbreaks of

ebolavirus in 2014, and 2018, and smaller outbreaks of marburgviruses in 2023, and the 2020 COVID-19 pandemic, the importance of research into zoonotic viruses has returned to the fore. The Zaire ebolavirus causes the most infections and incurs the most economic damage, though periodic outbreaks of marburgviruses also exist. Upon infection patients develop a severe hemorrhagic disease which results in death if untreated.[27] Standard treatment of oral or intravenous rehydration improves survival rates though case fatality rates range from 50-90% reflecting the severity of disease, and low access to necessary care that infected individuals receive. Currently, there is an FDA approved vaccine (Ervebo) and two antibody-based therapies targeting ebolaviruses (Inmazeb and Ebanga).[28] For marburgviruses the FDA has not approved any specific treatments. When the WHO identifies a new outbreak, they implement ring-vaccination campaigns, but the infection usually spreads for some time before vaccination and other health measures restrict transmission. Even in smaller outbreaks, hundreds of people are affected, may die, and cost millions of dollars in healthcare costs and other economic damage.[29] Yet because viruses are prone to mutation, and new zoonotic threats are constantly emerging in this viral order (*Mononegavirales*), there is much interest in identifying new drug targets and developing our understanding of them for target-based drug design and discovery.

1.6.1 Filovirus Classification And Taxonomy

Filoviruses are negative sense non-segmented single stranded RNA viruses in the order *Mononegavirales*. The order *Mononegavirales* contains negative sense, nonsegmented single stranded RNA viruses in such families as *Paramyxoviride* which contains the genus *Henipavirus* which is an emergent health threat. Many members of the order *Mononegavirales* are of great health concern including the measles, mumps, rabies, and respiratory syncytial viruses. Among the members of the family *Filoviridae* that infect humans, the genera *Marburgviruses* and

ebolaviruses are the principal disease-causing genera. The genus *ebolavirus* comprises Zaire, and Reston ebolaviruses among six total species. Of the six ebolavirus species, four cause disease in humans and Zaire ebolavirus is the most common and dangerous species with case fatality rates averaging around 50%. Reston ebolavirus does not cause disease in humans despite being closely related to Zaire ebolavirus. Studies of the important differentiating factors in the family *Filoviridae* are ongoing.

1.6.2 Properties Of Filoviruses

All filoviruses share a common genomic and structural organization. The genome at approximately nineteen kilobases encodes seven genes whose protein products perform all replication and packaging steps of the virus lifecycle. A filovirus enters a host cell by micropinocytosis after which it is enclosed in an endosome which matures triggering fusion of the virion with the endosomal membrane and release of the viral ribonucleocapsid into the cytosol. The ribonucleocapsid carries the virus's genomic RNA (gRNA) and RNA dependent RNA polymerase (L) which first generates subgenomic mRNAs for production of the viral proteins NP, VP35, VP40, GP, VP30, VP24 and L. The cycle switches to replication through a somewhat unknown mechanism likely involving the relative concentrations of the viral proteins. VP30 acts as a transcription factor through its interactions with L. L replicates the gRNA by first generating a positive sense single stranded RNA (ssRNA) which is then used as a template for producing negative sense single stranded gRNA that available NP can encapsidate to initiate viral assembly. The GP glycoprotein is trafficked to the cell membrane and with the matrix (VP40) protein which interacts with the ribonucleocapsid, buds out of the cell via the ESCRT complex. Interestingly, Zaire ebolavirus's genome contains mRNA editing sites that enable production of up to three unique GP variants.

During production of subgenomic RNAs or gRNA replication, the virus may produce double stranded RNA (dsRNA) which is a Pathogen Associated Molecular Pattern (PAMP).[30,31] Produced during genomic replication, the complementary strands of gRNA and positive sense ssRNA can hybridize producing dsRNA which host cells can recognize via the RIG-I like receptors (RLRs) RIG-I and MDA-5. Other mechanisms for producing dsRNAs exist and one such example is the production of defective viral genomes (DVGs). During gRNA replication, it is possible for L to generate copy-back DVGs which are internally self-complementary. DVGs are strongly immunostimulatory through RIG-I specifically. RLRs recognize the dsRNA PAMP through binding to blunt ends or backbone portions of dsRNAs and begin to unwind the duplexes with their ATP-dependent helicase activity. This enzymatic activity activates the interferon (IFN) pathway promoting apoptosis and signaling to nearby cells to do the same. Absent any antagonism, RLR activation helps prevent the spread of viral infections throughout the host. To evade these innate immune systems, filoviruses carry a protein called Viral Protein 35 (VP35) which binds to dsRNA and prevents activation of RLRs and the IFN pathway, inhibiting apoptosis and IFN-β signaling.[32]

1.6.3 VP35 Is Conserved Across The Family *Filoviridae*

All filoviridae have VP35 though with varying sequence identity conservation. Between Marburg marburgvirus VP35 (mVP35) and Zaire ebolavirus (eVP35), there is ~35% sequence identity in the whole sequence though this does vary between domains.[33] VP35 is comprised of three relevant regions: the N-terminal nucleoprotein binding peptide (NPBP), the oligomerization domain (OD), and the interferon inhibitory domain (IID). The NPBP comprises a small (~22 amino acid region) that binds to NP to prevent premature nucleocapsid assembly.[34] The OD forms intermolecular interactions with other VP35 molecules to form coiled coiled

trimers (m, eVP35) and tetramers (other ebolaviruses).[35] The IID is an alpha helical and β-sheet containing domain that interacts with NP, dsRNA, and a variety of host factors.[32] Through its interactions with NP, the IID is involved in the viral replication complex and packaging.[36] Through binding to dsRNA and host factors, the IID acts as an immune antagonist preventing activation of the interferon (IFN) response.

Although its structural organization and functions are largely the same across the family *Filoviridae*, VP35 from different species, have some distinct functions. While not all family members are herein addressed, I will discuss differences between the genera *marburgvirus* and *ebolavirus*. While VP35 from Marburg virus (mVP35) is known to bind only to the backbone of dsRNA, VP35 from Zaire ebolavirus (eVP35) binds to the backbone and blunt ends of dsRNA.[37-40] eVP35's two-dsRNA binding modes comprises a much stronger antagonist to RLR activation than just backbone binding since this antagonizes both RIG-I and MDA-5.[30] mVP35 antagonizes RIG-I activation much less than eVP35 while antagonizing MDA-5 to similar extents.[30] Given that Marburg virus has a case fatality rate that is much lower than that of Zaire ebolavirus, the ability of eVP35 to antagonize immune activation more strongly than mVP35 suggests the second binding mode plays a central role in promoting severe disease through preventing early identification of an ongoing infection. Mutating several residues in the IID that results in weaker dsRNA binding is sufficient to reenable activation of the host immune system and faster viral clearance and less severe disease supporting the model of VP35 as central to immune antagonism and disease severity.[41]

1.6.4 VP35 Is Required For Viral Replication And Packaging

The N-terminus of VP35 contains unstructured regions of unknown importance, the NPBP, and the OD. Truncating VP35 residues 1-52 reduces replication up to 60% relative to full-length

protein. This structural feature is common to mVP35 and eVP35 though with only ~50% sequence identity. This region in eVP35 is necessary and sufficient for replication. The NPBP interacts with the NP C terminal domain and prevents NP oligomerization and RNA binding.[34] NPBP supports gRNA synthesis and inhibits gRNA synthesis depending on whether it is present within full length VP35 or alone.

The OD in VP35 acts to increase dsRNA binding affinity through avidity effects and is required for full IFN antagonism, and genome replication. The ODs of eVP35 and mVP35 only share 22.7% sequence identity but both form coiled coiled trimers.[33,42] The mVP35 trimer is very stable (Tm > 90C) and likely does not form tetramers whereas eVP35 OD can form both trimers and tetramers. In the cryo-EM model of VP35 and L, the tetrameric OD forms some contacts with L though the interaction is primarily through the IID and linker regions. Oligomerization increases the affinity of VP35 for dsRNA 6 fold from 3.3 μM to 500 nM.[43]

1.6.5 VP35 Is Critical For Filovirus Immune Antagonism

The c-terminus of VP35, the IID, is 128 amino acids from 211 to 340. The IID binds to N for gRNA replication and dsRNA to prevent activation of the innate immune response via the RLRs The IID is also responsible for interacting with and inhibiting activation of a variety of other non-RLR host factors such as IRF-3.[31] Critical structural and biochemical data has extensively shown the importance of the IID for severe disease phenotypes.

Three mutations (F239A, K319A, R322A) to the IID are known to reduce dsRNA binding, IFN antagonism, and reenable activation of the host immune system and elicit protective immunity.[41,44] These studies demonstrate the importance of VP35 for virulence. There are important differences in immune antagonism between m and eVP35. mVP35 does not bind to the

blunt ends of dsRNA or inhibit activation of RIG-I while eVP35 both blinds to blunt ends of dsRNA and strongly antagonizes RIG-I.

1.6.6 An Overview Of VP35 Drug Targeting Efforts

VP35 engages in protein-protein and protein-nucleic acid interactions through basic patches on its surface.[45] The geometry of these flat binding surfaces is poorly suited for high affinity drug binding since there is not sufficient interactions available. In its dsRNA bound structure eVP35 binds to the blunt end of the dsRNA via cation-pi interactions. The electrostatic complementarity between the basic (positively charged) patches on eVP35 and the acidic (negatively charged) phosphates on the RNA backbone drives the backbone binding interaction. Both interaction modes seem to be flat interfaces. Still, previous docking studies have predicted that identified compounds likely bind to the RNA-binding interface.[46,47] Given the centrality of VP35's role in immune suppression targeting it with novel small molecules would be therapeutically useful as well as identifying pharmacological tools for future studies.

VP35 has been subject to many drug discovery efforts of varying scale and target. Since VP35 plays multiple roles in the viral lifecycle including at the switch from replication to packaging, and immune antagonism, drug discovery efforts have been targeted at each of these activities. Work targeting the NPBP-NP interaction identified new inhibitors of NPBP binding.[48] Also, studies targeting the IID-NP interaction identified pyrrolidine molecules as useful inhibitors of IID-NP interactions.[49] Studies also have targeting the IID's dsRNA binding activity through molecular docking studies and some experiments though structural information on dsRNA inhibitors and their mechanisms is lacking.[46,47]

1.7 Hypotheses Addressed In This Thesis

1.7.1 Zaire Ebolavirus VP35 Harbors A Cryptic Pocket That Is Coupled To Dsrna Binding

VP35 is known to be an important target for drug discovery to combat filovirus infections. VP35 is conserved across the filovirus family, it suppresses activation of the IFN response and is important for viral replication and packaging. Targeting protein-nucleic acid interactions is not viewed as a tractable drug discovery approach resulting in few drugs targeting non-enzymatic PNIs like VP35.

Furthermore, in early mechanistic studies of VP35's interaction with N found that increasing concentrations of dsRNA led to less VP35-N binding. This negative cooperativity may suggest conformational changes associated with the binding of one upon the binding of the other. Understanding more about VP35's conformational landscape enables us to develop new drugs to target alternative conformations than the current experimentally captured conformation and understand if VP35 undergoes conformational changes coupled to its dsRNA binding activity. Here we develop VP35 into a druggable target through widely sampling its conformational landscape, predicting which conformations are coupled to dsRNA binding, and experimentally validating our predictions. My hypothesis is that VP35 adopts alternative conformations in its native ensemble that are coupled to dsRNA binding.

1.7.2 Conformational Heterogeneity May Be Targetable For Drug Discovery

Given that VP35 is a conformationally heterogenous molecule, alternative conformations may be useful to target for drug discovery. My hypothesis is that a high throughput screen can uncover new chemical matter that binds to the cryptic pocket in VP35. Having new chemical matter is a boon to this biological field as it would set off a new series of drug discovery efforts through showing experimentally what interfaces are available for inhibiting dsRNA binding.

Furthermore, the identification of new chemical matter that binds to the VP35 cryptic pocket enables more complex cell-based or animal-based studies to help understand if targeting this cryptic pocket would produce the same druglike activity as seen *in-vitro*. I will apply multiple biophysical approaches to test this question ranging from NMR to fluorescence and molecular docking studies.

1.8 References

1 Reysenbach, A.-L., Longnecker, K. & Kirshtein, J. Novel Bacterial and Archaeal Lineages from an In Situ Growth Chamber Deployed at a Mid-Atlantic Ridge Hydrothermal Vent. *Applied and Environmental Microbiology* **66**, 3798-3806 (2000). https://doi.org:10.1128/AEM.66.9.3798-3806.2000

2 Nelson, D. L., Cox, M. M. & Hoskins, A. A. *Lehninger principles of biochemistry*. Eighth edition. edn, (Macmillan Learning, 2021).

3 Soutourina, J. Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology* **19**, 262-274 (2018). https://doi.org:10.1038/nrm.2017.115

4 Knoverek, C. R., Amarasinghe, G. K. & Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends in Biochemical Sciences* **44**, 351-364 (2019). https://doi.org:10.1016/j.tibs.2018.11.007

5 Hart, K. M., Ho, C. M. W., Dutta, S., Gross, M. L. & Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nature Communications* **7**, 12965 (2016). https://doi.org:10.1038/ncomms12965

6 Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews* **117**, 139-155 (2017). https://doi.org:10.1021/acs.chemrev.6b00177

7 Tombola, F., Pathak, M. M. & Isacoff, E. Y. How Does Voltage Open an Ion Channel? *Annual Review of Cell and Developmental Biology* **22**, 23-52 (2006). https://doi.org:10.1146/annurev.cellbio.21.020404.145837

8 Hilser, V. J. & Thompson, E. B. Structural Dynamics, Intrinsic Disorder, and Allostery in Nuclear Receptors as Transcription Factors *. *Journal of Biological Chemistry* **286**, 39675-39682 (2011). https://doi.org:10.1074/jbc.R111.278929

9 Aviram, H. Y. *et al.* Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions. **115**, 3243-3248 https://doi.org:10.1073/pnas.1720448115

10 Bhabha, G., Biel, J. T. & Fraser, J. S. Keep on Moving: Discovering and Perturbing the Conformational Dynamics of Enzymes. *Accounts of Chemical Research* **48**, 423-430 (2015). https://doi.org:10.1021/ar5003158

11 Knoverek, C. R. *et al.* Opening of a cryptic pocket in β-lactamase increases penicillinase activity. *Proceedings of the National Academy of Sciences* **118**, e2106473118 (2021). https://doi.org:10.1073/pnas.2106473118

12 Skehel, J. J. & Wiley, D. C. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annual Review of Biochemistry* **69**, 531-569 (2000). https://doi.org:10.1146/annurev.biochem.69.1.531

13 Meller, A. *et al.* Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *eLife* **12**, e83602 (2023). https://doi.org:10.7554/eLife.83602

14 Lu, S. *et al.* Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Communications* **12**, 4721 (2021). https://doi.org:10.1038/s41467-021-25020-9

15 Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356-363 (2009). https://doi.org:10.1038/nature08144

16 Thal, D. M., Glukhova, A., Sexton, P. M. & Christopoulos, A. Structural insights into G-protein-coupled receptor allostery. *Nature* **559**, 45-53 (2018). https://doi.org:10.1038/s41586-018-0259-z

17 Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications* **12**, 1936 (2021). https://doi.org:10.1038/s41467-021-21953-3

18 Xu, J. S. *et al.* Lactose repressor hinge domain independently binds DNA. *Protein Science* **27**, 839-847 (2018). https://doi.org:https://doi.org/10.1002/pro.3372

19 Bain, D. L., Heneghan, A. F., Connaghan-Jones, K. D. & Miura, M. T. Nuclear Receptor Structure: Implications for Function. *Annual Review of Physiology* **69**, 201-220 (2007). https://doi.org:10.1146/annurev.physiol.69.031905.160308

20 Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11681-11686 (2012). https://doi.org:10.1073/pnas.1209309109

21 Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **9**, 71 (2011). https://doi.org:10.1186/1741-7007-9-71

22 Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J. & Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Accounts of Chemical Research*, acs.accounts.9b00613 (2020). https://doi.org:10.1021/acs.accounts.9b00613

23 Zimmerman, M. I. & Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of Chemical Theory and Computation* **11**, 5747-5757 (2015). https://doi.org:10.1021/acs.jctc.5b00737

24 Zimmerman, M. I. *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nature Chemistry* **13**, 651-659 (2021). https://doi.org:10.1038/s41557-021-00707-0

25 Schames, J. R. *et al.* Discovery of a Novel Binding Trench in HIV Integrase. *ACS Publications* **47**, 1879-1881 (2004). https://doi.org:10.1021/jm0341913

26 Hart, K. M. *et al.* Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. **12**, e0178678 https://doi.org:10.1371/journal.pone.0178678

27 Jacob, S. T. *et al.* Ebola virus disease. *Nature Reviews Disease Primers* **6**, 13 (2020). https://doi.org:10.1038/s41572-020-0147-3

28 Cross, R. W., Mire, C. E., Feldmann, H. & Geisbert, T. W. Post-exposure treatments for Ebola and Marburg virus infections. *Nature Reviews Drug Discovery* **17**, 413-434 (2018). https://doi.org:10.1038/nrd.2017.251

29 Huber, C., Finelli, L., infectious, W. S. T. J. o. & 2018. The economic and social burden of the 2014 Ebola outbreak in West Africa. *academic.oup.com* https://doi.org:10.1093/infdis/jiy213","keywords":["ebola

30 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Rep* **14**, 1632-1640 (2016). https://doi.org:10.1016/j.celrep.2016.01.049

31 Ramanan, P. *et al.* Structural basis for Marburg virus VP35-mediated immune evasion mechanisms. *Proc Natl Acad Sci U S A* **109**, 20661-20666 (2012). https://doi.org:10.1073/pnas.1213559109

32 Hartman, A. L., Towner, J. S. & Nichol, S. T. A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. *Virology* **328**, 177-184 (2004). https://doi.org:10.1016/j.virol.2004.07.006

33 Bruhn Jessica, F. *et al.* Crystal Structure of the Marburg Virus VP35 Oligomerization Domain. *Journal of Virology* **91**, e01085-01016 (2017). https://doi.org:10.1128/JVI.01085-16

34 Leung, D. W. *et al.* An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Rep* **11**, 376-389 (2015). https://doi.org:10.1016/j.celrep.2015.03.034

35 Zinzula, L. *et al.* Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species. *Structure* **27**, 39-54.e36 (2019). https://doi.org:https://doi.org/10.1016/j.str.2018.09.009

36 Yuan, B. *et al.* Structure of the Ebola virus polymerase complex. *Nature* **610**, 394-401 (2022). https://doi.org:10.1038/s41586-022-05271-2

37 Bale, S. *et al.* Marburg virus VP35 can both fully coat the backbone and cap the ends of dsRNA for interferon antagonism. *PLoS Pathog* **8**, e1002916 (2012). https://doi.org:10.1371/journal.ppat.1002916

38 Bale, S. *et al.* Ebolavirus VP35 coats the backbone of double-stranded RNA for interferon antagonism. *J Virol* **87**, 10385-10388 (2013). https://doi.org:10.1128/JVI.01452-13

39 Basler, C. F. & Amarasinghe, G. K. Evasion of Interferon Responses by Ebola and Marburg Viruses. *Journal of Interferon & Cytokine Research* **29**, 511-520 (2009). https://doi.org:10.1089/jir.2009.0076

40 Cardenas, W. B. *et al.* Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *J Virol* **80**, 5168-5178 (2006). https://doi.org:10.1128/JVI.02199-05

41 Woolsey, C. *et al.* A VP35 Mutant Ebola Virus Lacks Virulence but Can Elicit Protective Immunity to Wild-Type Virus Challenge. *Cell Rep* **28**, 3032-3046 e3036 (2019). https://doi.org:10.1016/j.celrep.2019.08.047

42 Zinzula, L. *et al.* Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species. *Structure* **27**, 39-54 e36 (2019). https://doi.org:10.1016/j.str.2018.09.009

43 Reid, S. P., Cardenas, W. B. & Basler, C. F. Homo-oligomerization facilitates the interferon-antagonist activity of the ebolavirus VP35 protein. *Virology* **341**, 179-189 (2005). https://doi.org:10.1016/j.virol.2005.06.044

44 Prins, K. C. *et al.* Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. *J Virol* **84**, 3004-3015 (2010). https://doi.org:10.1128/JVI.02459-09

45 Prins, K. C. *et al.* Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. *J Virol* **84**, 10581-10591 (2010). https://doi.org:10.1128/JVI.00925-10

46 Glanzer, J. G. *et al.* In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. *Bioorg Med Chem* **24**, 5388-5392 (2016). https://doi.org:10.1016/j.bmc.2016.08.065

47 Daino, G. L. *et al.* Identification of Myricetin as an Ebola Virus VP35-Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. *Biochemistry* **57**, 6367-6378 (2018). https://doi.org:10.1021/acs.biochem.8b00892

48 Liu, G. *et al.* A Sensitive in Vitro High-Throughput Screen To Identify Pan-filoviral Replication Inhibitors Targeting the VP35-NP Interface. *ACS Infect Dis* **3**, 190-198 (2017). https://doi.org:10.1021/acsinfecdis.6b00209

49 Brown, C. S. *et al.* In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. *J Mol Biol* **426**, 2045-2058 (2014). https://doi.org:10.1016/j.jmb.2014.01.010

Chapter 2: A Cryptic Pocket In Ebola VP35 Allosterically Controls RNA Binding

2.1 Preface

This chapter comprises an adapted and edited form of the publication with the same name published in *Nature Communications* in 2022 (Cruz, M.A., Frederick, T.E., Mallimadugula, U.L. et al. A cryptic pocket in Ebola VP35 allosterically controls RNA binding. Nat Commun 13, 2269 (2022). https://doi.org/10.1038/s41467-022-29927-9). As the author I am the copyright holder of this work which is available under a CC BY creative commons license per *Nature Communications* policy. In collaboration with Thomas E. Frederick[1, ‡] I conducted experiments, manuscript writing, and data visualizations. With Upasana L. Mallimadugula[1], Sukrit Singh[1], Neha Vithani[1], Maxwell I. Zimmerman[1], Justin R. Porter[1] I also contributed to performing simulations, and analyzing simulation data. Other than myself and TEF, Upasana L. Mallimadugula[1] and Katelyn E. Moeder[1] conducted experiments. Myself, Gaya K. Amarasinghe[2], and Gregory R. Bowman[1,3,*] acquired funds to support this work. Finally, GRB, and GKA conceptualized the research direction and strategy.

I am grateful to the citizen scientists who participate in Folding@home for volunteering to run simulations on their personal computers. This work was funded by NSF CAREER Award MCB-1552471 and NIH grant R01 GM124007 (Bowman), as well as NIH grants R01AI123926, P01AI120943, and R01AI143292 (Amarasinghe). GRB holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and a Packard Fellowship for Science and Engineering from The David & Lucile Packard Foundation. I was supported by the NIH grants

5R25GM103757 to WUSTL IMSD program, and NIH F31AI157079. SS was supported by a MilliporeSigma Fellowship. I thank Drs. Timothy M. Lohman and Alexander G. Kozlov for advice on FP assays.

2.2 Abstract

Protein-protein and protein-nucleic acid interactions are often considered difficult drug targets because the surfaces involved lack obvious druggable pockets. Cryptic pockets could present opportunities for targeting these interactions, but identifying and exploiting these pockets remains challenging. Here, we apply a general pipeline for identifying cryptic pockets to the interferon inhibitory domain (IID) of Ebola viral protein 35 (VP35). VP35 plays multiple essential roles in Ebola's replication cycle but lacks pockets that present obvious utility for drug design. Using adaptive sampling simulations and machine learning algorithms, we predict VP35 harbors a cryptic pocket that is allosterically coupled to a key dsRNA-binding interface. Thiol labeling experiments corroborate the predicted pocket and mutating the predicted allosteric network supports our model of allostery. Finally, covalent modifications that mimic drug binding allosterically disrupt dsRNA binding that is essential for immune evasion. Based on these successes, we expect our pipeline is equally applicable to other proteins.

2.3 Introduction

Examining structures available in the protein data bank (PDB) suggests that many protein surfaces that engage in protein-protein interactions (PPIs) and protein-nucleic acid interactions (PNIs) lack druggable pockets.[1,2] As a result, PPIs and PNIs are often considered intractable drug targets even when there is strong evidence that disrupting these interactions would be of great therapeutic value.[3]

Cryptic pockets present opportunities for designing drugs for difficult targets like PPIs and PNIs but identifying and exploiting these pockets remains challenging.[4-6] Cryptic pockets are absent in available experimental structures but form in a subset of excited states that arise due to protein dynamics. These cryptic sites can serve as valuable drug targets if they coincide with key functional sites, or if they are allosterically coupled to distant functional sites.[7,8] Most known cryptic sites were only identified after the serendipitous discovery of a small molecule that binds and stabilizes the open form of the pocket.[8,9] Unfortunately, we currently lack methodology that can decouple pocket discovery from ligand discovery. To overcome this limitation and to increase the number of druggable targets, we have developed a suite of computational and experimental methods for detecting cryptic pockets and allostery, in addition to other available approaches.[10-12,13-24] We have successfully applied subsets of this toolset to a number of enzymes that are established drug targets,[12,25] suggesting that the same tools may be ready for application to challenging targets like PPIs and PNIs.

Here, we present the first integration of our entire pipeline of tools to hunt for cryptic pockets in a difficult, non-enzymatic target that engages in PPIs and PNIs: the interferon inhibitory domain (IID) of Ebola viral protein 35 (VP35). Ebola virus causes a hemorrhagic fever that is often lethal, with case fatality rates approaching 90% in past outbreaks.[26,27] Initial promising results

with the antiviral, remdesivir fell short in a randomized controlled trial so there remains no approved small molecule drugs for treating Ebola.[28] Small molecule antivirals are needed despite recent progress with antibodies[29] because they offer many advantages, including ease of delivery, lower cost, and longer shelf life that are particularly relevant in rural and impoverished regions. The ~120 residue IID of VP35 would be an appealing drug target for combating Ebola and other viruses in the *Filoviridae* family apart from lacking obvious druggable sites that could disrupt its PPI and PNIs. VP35 has a well-conserved sequence and plays multiple essential roles in the viral replication cycle.[30] One of its primary functions is to antagonize the host innate immunity responses, particularly RIG-I-like receptor (RLR)-mediated detection of viral nucleic acids, to prevent an interferon (IFN) response and signaling of neighboring cells to heighten their antiviral defenses.[31-33]

Crystal structures have revealed that VP35's IID binds both the blunt ends and backbone of double-stranded RNA (dsRNA), and that there is a PPI between these dsRNA-binding modes (Fig. 2.1).[34,35] Binding to dsRNA blunt ends plays a dominant role in IFN suppression by Ebola.[36] Indeed, mutations that reduce the IID's affinity for dsRNA blunt ends are sufficient to mitigate IFN antagonism, ultimately attenuating Ebola's pathogenicity.[36-39] Therefore, disrupting this single binding mode could dramatically reduce the impact of an Ebola infection on the host and potentially reduce deleterious effects, including lethality. However, both dsRNA-binding interfaces are large flat surfaces that are difficult for small molecules to bind tightly (Fig. 2.1). As a result, only a few studies have sought to find small molecules targeting VP35, none of which has evolved into a full drug discovery campaign.[40-43] The discovery of cryptic pockets in VP35 could provide new opportunities for drugging this essential viral component.

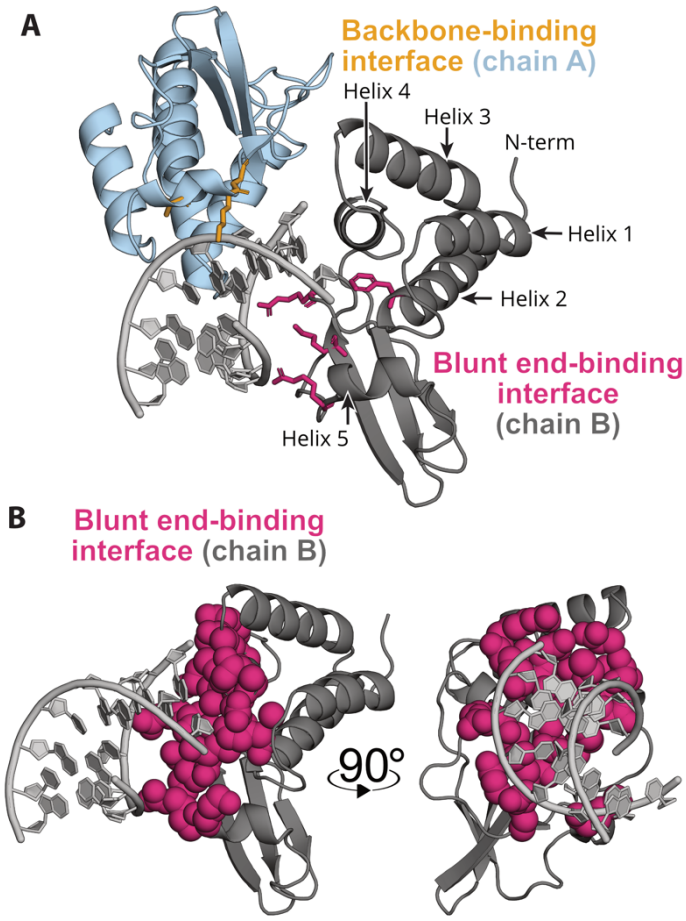

Figure 2.1 VP35 dsRNA interactions occur primarily through flat interfaces.

A) Crystal structure of two copies of VP35's IID (dark gray and light blue) bound to dsRNA (light gray) via two flat interfaces (PDB ID 3L25). Mutations to residues highlighted in pink and yellow sticks eliminate dsRNA binding. B) Isolated chain B from same view as panel A and after 90-degree rotation in the Y axis now highlighting the dsRNA interacting VP35 surface in the blunt end binding protomer. The blunt end-binding interface (pink, 3L25 chain B) is shown as spheres to highlight that VP35 lacks deep pockets amenable to binding small molecules.

2.4 Results

2.4.1 Adaptive Sampling Simulations Reveal A Potentially Druggable Cryptic Pocket.

To discover structures with large pocket volumes that may harbor cryptic pockets, we applied our previously described fluctuation amplification of specific traits (FAST) simulation algorithm.[44] FAST is a goal-oriented adaptive sampling algorithm that exploits Markov state model (MSM) methods to explore regions of conformational space with user-specified structural features. An MSM is a network model of a protein's energy landscape which consists of a set of structural states the protein adopts and the rates of hopping between them.[45,46] After running FAST, we gathered additional statistics by running simulations from each state on the Folding@home distributed computing environment, which brings together the computing resources of hundreds of thousands of citizen scientists who volunteer to run simulations on their personal computers. Our final model has 4,469 conformational states, providing a detailed characterization of the different structures the IID adopts, but making manual interpretation of the model difficult.

To identify cryptic pockets within the large ensemble captured by our MSM, we searched for signatures of cryptic pockets such as groups of residues with highly correlated changes in solvent exposure, referred to as exposons.[12] Exposons are often associated with cryptic sites because the opening/closing of such pockets gives rise to cooperative increases/decreases in the solvent exposure of surrounding residues. Importantly, once an exposon has been identified, our MSM framework provides a facile means to identify the conformational changes that give rise to that exposon.

Our simulations reveal two exposons in the VP35 IID, one of which corresponds to a large cryptic pocket. The blue exposon (Fig. 2.2A and 2.2B) which overlaps with the backbone-binding interface in Fig. 2.1, consists of a set of strongly coupled residues in helix 5 and adjacent loops and secondary structure elements. Visualizing the conformational change that gives rise to this cluster reveals a substantial displacement of helix 5, creating a large cryptic pocket between it and the helical domain (Fig. 2.2C). A number of residues that are displaced along with helix 5 (i.e. A306, K309, and S310) make van der Waals contacts with the dsRNA backbone in the dsRNA-bound crystal structure,[35] so targeting this cryptic pocket could directly disrupt this binding mode.

Retrospective analysis of other validated drug targets suggests cryptic sites created by the movement of secondary structure elements, such as the displacement of helix 5, are often druggable.[47] The potential druggability of this cryptic site is also supported by application of the Fpocket and FTMap algorithms.[48,49] Fpocket predicts this cryptic site to have a high druggability score (0.681) and FTMap highlights a number of hotspots within the pocket where small molecules could form a variety of energetically-favorable interactions (Fig. 2.2E and Fig. S2.7). Unfortunately, disrupting backbone binding is of less therapeutic utility than disrupting blunt end binding and it is unknown whether the contacts between A306, K309, and S310 are essential for backbone binding. Therefore, it is unclear from this analysis alone whether drugging this newly discovered cryptic pocket would be useful.

The second exposon (orange in Fig. 2.2) encompasses portions of both dsRNA-binding interfaces, but it does not correspond to a cryptic pocket. This cluster includes residues that bind

dsRNA's backbone (i.e. S272) and residues that interact with both the blunt ends and backbone of dsRNA (i.e. F239, Q274, and I340).[35] Therefore, altering the conformational preferences of the second exposon could potentially disrupt the blunt end-binding mode and its crucial role in Ebola virus's ability to evade an immune response. However, the largest conformational change involved in the formation of this exposon is a displacement of the loop between helices 3 and 4 (Fig. 2.2D). This rearrangement does not create a cryptic pocket that is large enough to accommodate drug-like molecules, so it is not obvious how to directly manipulate the orange exposon.

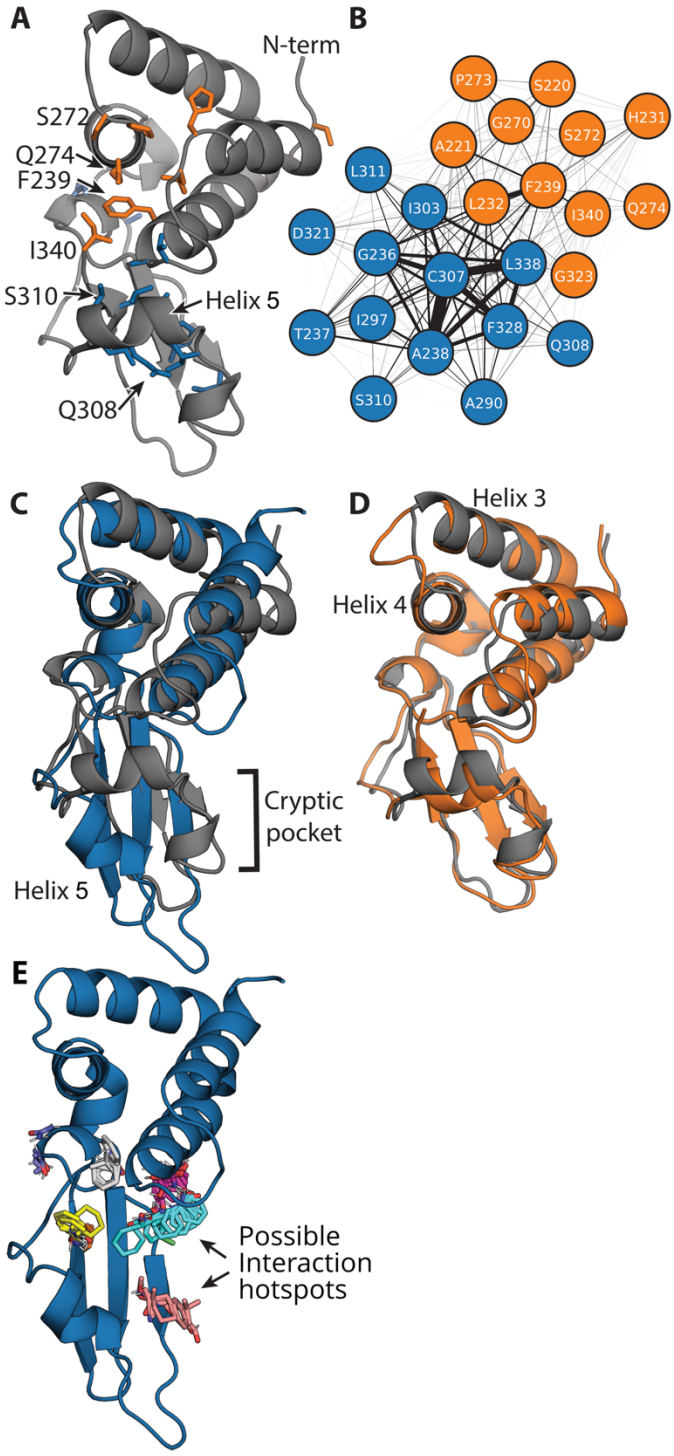

Figure 2.2 Exposons identify a large cryptic pocket and suggest potential allosteric coupling.

A) Structure of VP35's IID highlighting residues in two exposons (blue and orange), the N-terminus (N-term), and C-terminus (I340) (PDB ID 3FKE). B) Network representation of the coupling between the solvent exposure of residues in the two exposons. The edge width between residues is proportional to the mutual information between them. C) Structure highlighting the opening of a cryptic pocket via the displacement of helix 5 that gives rise to the

blue exposon. D) Structure highlighting the conformational change that gives rise to the orange exposon overlaid on the crystal structure (gray) to highlight that the rearrangements are subtler than in the blue exposon. E) FTMap results for the main cryptic pocket as shown in C and hotspots where a variety of small organic probes (multicolored sticks) form energetically favorable interactions. The probe molecules are intended to capture different drug-like interactions (such as hydrogen bonding and van der Waals contacts) and include acetamide, acetonitrile, acetone, acetaldehyde, methylamine, benzaldehyde, benzene, isobutanol, cyclohexane, N,N-dimethylformamide, dimethyl ether, ethanol, ethane, phenol, isopropanol, or urea.[50-53]

2.4.2 The Cryptic Pocket Is Allosterically Coupled To The Blunt End-Binding Interface.

Even though the cryptic pocket does not coincide with the interface of VP35's IID that binds dsRNA blunt ends, it could still serve as a cryptic allosteric site that allosterically controls dsRNA binding. Indeed, the physical proximity of the two exposons and the coupling between them both hint at the possibility for allosteric coupling. Furthermore, our exposons analysis could easily underestimate this coupling given that it focuses on correlated transitions of residues between solvent exposed and completely buried states, leaving it blind to more subtle conformational fluctuations and allostery involving residues that are always buried (or always exposed).

To explore the potential for a broader allosteric network, we quantified the allosteric coupling between every pair of residues using our correlation of all rotameric and dynamical states (CARDS) algorithm.[54] CARDS classifies each dihedral in each snapshot of a simulation as being in one of three rotameric states (gauche+, gauche-, or trans) and one of two dynamical states (ordered or disordered). A mutual information metric is then used to quantify the coupling between the structure and dynamics of every pair of dihedral angles, which can then be coarse-grained to the correlation between every pair of residues. Importantly, CARDS accounts for the potential role of residues that are always buried or always exposed to solvent and subtle conformational changes that do not alter the solvent exposure of residues.

CARDS reveals a broader allosteric network than that identified by our exposons analysis and suggests strong coupling between the cryptic pocket and blunt end-binding interface (Fig. 2.3A and 2.3B). This network consists of five communities of strongly coupled residues, four of which

coincide with large portions of the two dsRNA-binding interfaces. One of these communities (orange) is a hub in the network, having significant coupling to all the other communities. It encompasses part of the orange exposon, particularly residues around the loop between helices 3 and 4. The orange CARDS community and exposon both capture Q274, which engages in both dsRNA-binding interfaces, and S272, which contacts the backbone.[35] However, the CARDS community includes many additional residues not captured by exposons analysis. Examples include I278, which engages in both dsRNA-binding interfaces, and D271, which is part of the PPI between the two binding modes.[35] One of the orange community's strongest allosteric connections is to the green community. This community encompasses the rest of the residues in the orange exposon, including F239 and I340, which are part of both dsRNA-binding interfaces.[35] The green community also captures additional residues, reaching deep into the helical domain. The orange community is also strongly coupled to the blue community, which includes much of helix 5 and nearby residues that move to give rise to the cryptic pocket that was captured by the blue exposon. Notably, the orange and blue communities are both coupled to a cyan cluster that was not hinted at by our exposons analysis because the residues involved are always solvent exposed. It includes R322, which is part of the blunt end-binding interface and the PPI between the two binding modes, and K282, which also contacts dsRNA blunt ends.[35] In addition, this community includes K339, which is an important determinant of the electrostatic favorability of dsRNA binding.[35] Together, these results suggest that opening of the cryptic pocket could strongly impact residues involved in both dsRNA-binding interfaces, as well as the PPI between the two binding modes.

2.4.3 Opening Of The Cryptic Pocket Alters The Structural Preferences Of The Dsrna-Binding Interface

To assess if pocket opening impacts the blunt end-binding interface, we compared the ensembles of structures with the cryptic pocket open or closed. We hypothesized that if pocket opening affects blunt end binding, the dsRNA-binding residues in the ensembles of structures of the open and closed states will have distinct structural features other than pocket opening. To test this hypothesis, we applied our previously described machine learning algorithm, DiffNets, which is a supervised autoencoder architecture designed to identify the key differences between two or more structural ensembles.[10] In this case, we used DiffNets to compare the ensemble of structures with an open cryptic pocket to those with a closed cryptic pocket and assess if there are important differences between the structural preferences of the blunt end-binding interface.

This analysis reveals significant coupling between the opening/closing of the cryptic pocket and the structural preferences of a key blunt end-binding residue, F239. Specifically, we found that the distance between F239 and helix 5 is strongly correlated with the extent of pocket opening. Further investigation revealed that the distribution of $\chi_1$angles for F239 when the pocket is open differs substantially from the distribution when the pocket is closed (Fig. 2.3D). The orientation of F239 observed in available crystal structures is a well-populated when the cryptic pocket is closed. Opening of the cryptic pocket is associated with a reduction in the probability of this Gauche- dsRNA-binding competent rotamer. Therefore, we propose that stabilizing the closed pocket should enhance the affinity between VP35 and dsRNA blunt ends, while stabilizing the open pocket (e.g., via binding of a small molecule) should disrupt dsRNA binding.

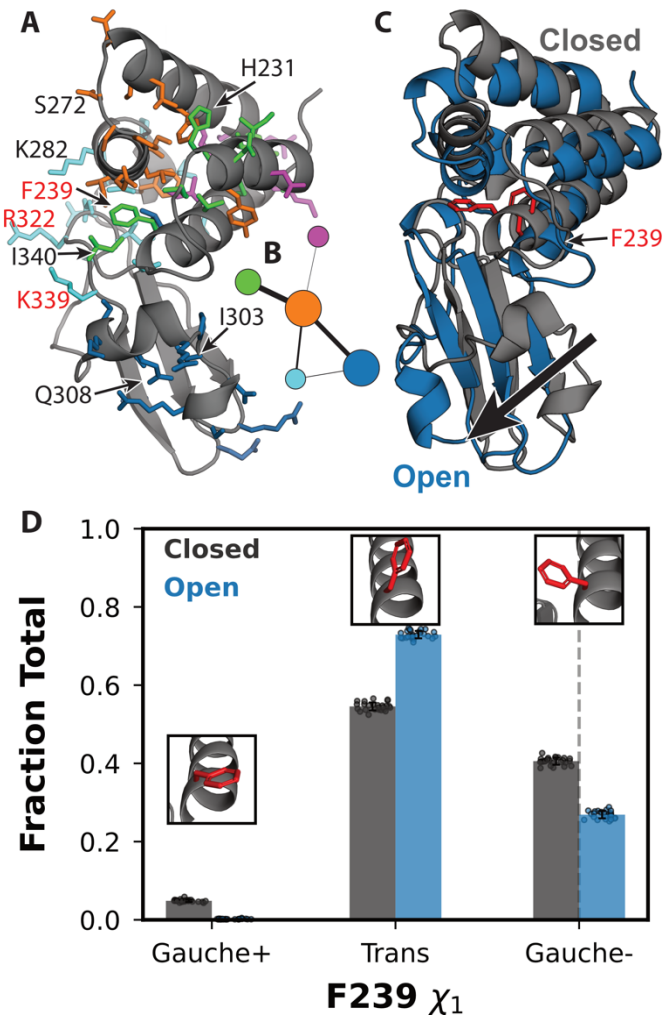

Figure 2.3 Allosteric network revealed by the CARDS algorithm.

A) Structure of VP35's IID with residues in the allosteric network shown in sticks and colored according to which of five communities they belong to. Substitution of residues labeled in red with alanine disrupts binding to dsRNA blunt ends and results in a dramatic reduction in immune suppression. B) Network representation of the coupling between communities of residues, colored as in A. Node size is proportional to the strength of coupling between residues in the community, and edge widths are proportional to the strength of coupling between the communities. C) Representative states of the correlated changes from the DiffNet. In gray is a structure with a closed pocket and in blue is a structure from MD simulation with an open pocket. F239 is shown in red sticks for orientation. D) Distribution of F239 χ_1 from the MSM with respect to states wherein the pocket is open (blue) or closed (black) for the three rotamers. The bar height is mean value from 25 bootstrapped MSMs (dots) of the sum of the population of all states in the MSMs with the specified rotamer. Insets show the conformation of F239 with the highest probability within the region of a given peak in the distribution as sampled in our MSM. The black dashed line at the Gauche– position corresponds to the calculated value of F239 χ_1 from PDB 3L26. Error bars are standard deviation from the mean of bootstrapped values from recalculating the MSM twenty-five times (see Methods).

2.4.4 Thiol Labeling Experiments Corroborate The Predicted Cryptic Pocket.

One way to experimentally test our prediction of a cryptic pocket is to probe for solvent exposure of residues that are buried in all the structures that are currently available in the protein data bank (PDB) but become exposed to solvent upon pocket opening. Cysteines are particularly appealing candidates for such experiments because 1) they have a low abundance and 2) their thiol groups are highly reactive, so it is straightforward to detect exposed cysteines by introducing labeling reagents that covalently bind accessible thiols. Fortuitously, VP35's IID has two cysteines (C307 and C326) that are buried in available crystal structures but become exposed to solvent when the cryptic pocket opens (Fig. 2.4B). There is also a cysteine (C275) that is on the surface of the apo crystal structure[34] and a fourth cysteine (C247) that is buried in the helical bundle. C275 is typically solvent exposed in our simulations, as expected based on the crystallographic data. Examining the solvent exposure of C247 revealed it is sometimes exposed to solvent via an opening of helix 1 relative to the rest of the helical bundle (Fig. S2.8), but FTMap did not identify any hotspots that are likely to bind drug-like molecules in this region. Therefore, we expect to observe labeling of all four cysteines on a timescale that is faster than global unfolding of the protein.

To experimentally test our predicted pocket, we applied a thiol labeling technique that probes the solvent exposure of cysteine residues.[55] For these experiments, 5,5'-dithiobis-(2-nitrobenzoic Acid) (also known as DTNB or Ellman's reagent, Fig. 2.4A) is added to a protein sample. Upon reaction with the thiol group of an exposed cysteine, DTNB breaks into two TNB molecules, one of which remains covalently bound to the cysteine while the other is released into solution. The

accumulation of free TNB can be quantified based on the increased absorbance at 412 nm. We have previously applied this technique to test predicted pockets in β-lactamase enzymes.[12,56]

As expected from our computational model, the observed signal from our thiol labeling experiments is consistent with opening of the cryptic pocket (Fig. 2.4C). Absorbance curves are best fit by four exponentials, each with an approximately equivalent amplitude that is consistent with expectations based on the extinction coefficient for DTNB (Fig. S2.9). To assign these labeling rates to individual cysteines, we systematically mutated the cysteines to serines, performed thiol labeling experiments, and assessed which rates disappeared and which remained (Fig. S2.10 and Table S2.11). For example, labeling of the C275S variant lacks the very fastest rate for wild-type, consistent with the intuition that a residue that is surfaced exposed in the crystal structure (i.e., C275) should label faster than residues that are generally buried. The consistency of the labeling rates between variants also confirms none of the observed labeling events are dependent on labeling of other cysteine residues.

To test whether the observed labeling could be due to an alternative process, such as global unfolding, we determined the population of the unfolded state and unfolding rate of VP35's IID under native conditions (Fig S2.12) and the intrinsic labeling rate for each cysteine (Fig S2.13). As shown in Fig. 2.4C, the observed labeling rates are all considerably faster than the expected labeling rate from the unfolded state at a range of DTNB concentrations. This result confirms that labeling of all four cysteines arises from fluctuations within the native state, consistent with our computational predictions.

That all four cysteines undergo labeling suggests that C247 undergoes local fluctuations that our exposons analysis does not predict will form a pocket. To determine the importance of this

fluctuation, we calculated the equilibrium constant for the exposure of both C247 and C307. Opening of the cryptic pocket is far more probable than the structural fluctuation that exposes C247 (equilibrium constants for the exposure of C247 and C307 are $6.9 \times 10^{-4} \pm 7 \times 10^{-5}$ and $4.0 \times 10^{-1} \pm 1.0 \times 10^{-2}$, respectively). Therefore, a ligand would have to pay a greater energetic cost to stabilize the conformational change that exposes C247 than to stabilize the open state of the cryptic allosteric site created by the motion of helix 5. Taken together with the fact that the motion of helix 5 creates a more druggable pocket than the motion that exposes C247, we continue to focus on the cryptic pocket created by the helix 5 motion.

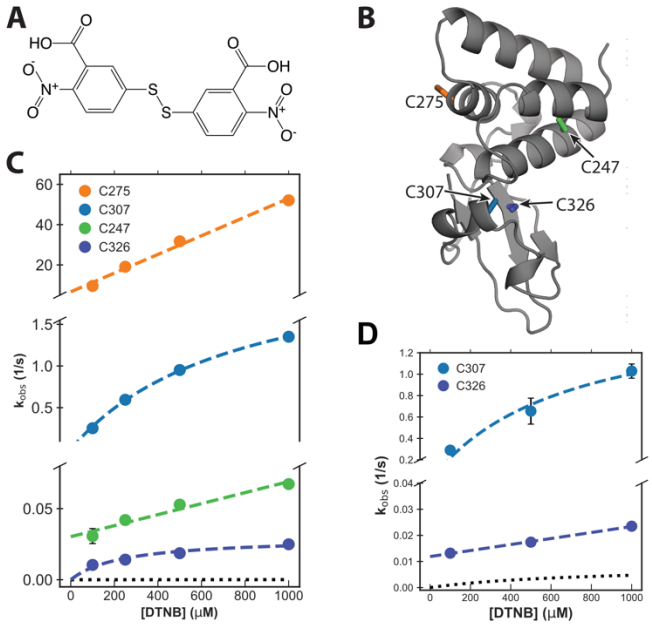

Figure 2.4 Thiol labeling supports the existence of the predicted cryptic pocket.

A) Structure of the DTNB labeling reagent. B) Structure of VP35's IID highlighting the locations of the four native cysteines (sticks). C307 and C326 are both buried and point into the predicted cryptic pocket. C) Observed labeling rates (circles) for WT VP35 at a range of DTNB concentrations. Fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. D) Observed labeling rates (circles) for VP35 C247S/C275S. Fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols.

2.4.5 Mutations Support Our Predicted Allosteric Network.

We sought to test our model of allostery in VP35 by introducing mutations and assessing their impact on the conformations of distal sites. To select mutations, we drew on both our model of allosteric coupling and the published literature. For example, our model's prediction of coupling between the conformation of F239 (in the blunt-end binding interface) and cryptic pocket opening suggests that an F239A mutation is likely to alter pocket opening. Previous work suggests that the linker between the two domains of Reston VP35 confers it with greater stability and rigidity than the Zaire variant of VP35 we focus on in this work.[57] One of the significant differences between the linkers of the two proteins is the presence of a proline in Reston VP35. Given proline is conformationally restricted, we reasoned that substituting A291 for proline in the linker of Zaire VP35 may restrict cryptic pocket opening and enhance dsRNA binding. To test these predictions, we created the relevant variants of VP35 and measured their impact on cryptic pocket opening using our thiol labeling assay.

Thiol labeling of F239A demonstrates that the mutation allosterically increases opening of the cryptic pocket. We find that the observed labeling rates for the cysteines in the cryptic pocket are two-fold faster than in wild-type VP35. Fitting with the Linderstrøm-Lang model reveals that the equilibrium probability of C307 exposure in F239A is approximately double that of wild-type (1.1 ± 0.2 vs $4.0 \times 10^{-1} \pm 1.0 \times 10^{-2}$ respectively) (Fig. S2.11). This thiol labeling data suggests that communication flows to and from the end-cap involved dsRNA binding residues and cryptic pocket.

In contrast, the mutation A291P decreases the probability of pocket opening, which results in a higher affinity for dsRNA. Thiol labeling experiments reveal that A291P dramatically reduces

the labeling rates of the two cysteines in the cryptic pocket (Fig. S2.12). In fact, the labeling rate of C326 in the A291P background is similar to the rate of global protein unfolding (Fig. 2.5A), suggesting that the pocket never opens enough to expose the most deeply buried regions of the cryptic pocket observed in the wild-type protein. The probability that the C307 of the A291P variant is accessible to our DTNB labeling reagent is also significantly smaller than in wild-type ($1.4 \times 10^{-4} \pm 2.0 \times 10^{-4}$ vs $4.0 \times 10^{-1} \pm 1.0 \times 10^{-2}$ respectively).

2.4.6 Stabilizing The Closed Pocket Increases Dsrna Binding.

Based on our predicted allosteric network, stabilizing the closed state of the cryptic pocket should enhance dsRNA binding. Specifically, the fact that the pocket is closed in the co-crystal structure of VP35 with dsRNA (PDB ID: 3L26) implies a closed pocket is favorable for dsRNA binding and a mutation that stabilizes the pocket in its closed form would increase dsRNA binding. Therefore, we should see a higher affinity between A291P and dsRNA.

To test this prediction, we developed a fluorescence polarization (FP) assay for measuring the affinity of VP35 for dsRNA. Paralleling past work on VP35-peptide interactions,[42] we added varying concentrations of VP35 IID to a fixed concentration of 25-bp RNA with a fluorescein isothiocyanate (FITC) conjugation at one the 5' end (Fig. S2.17). Free FITC-dsRNA emits depolarized light upon excitation with polarized light because of the molecule's fast rotation. Binding of one or more VP35 molecules restricts the motion of FITC-dsRNA, resulting in greater emission of polarized light, which is best monitored by the change in anisotropy.[58] This anisotropy based binding measurement recapitulates previously published binding affinities for two different dsRNA end topologies (blunt or overhanging 3' ends) (Fig S2.15).

Our data show that closing the pocket with A291P increases dsRNA binding. To test how A291P binds dsRNA, we repeated the binding assay done with VP35 IID C247S/C275S with A291P and a 25 base pair blunt-ended dsRNA and calculated the apparent affinity to be 1.8 ± 0.1 μM. This corresponds to a twofold *increase* in apparent binding affinity relative to wild-type VP35. We also find that A291P is sensitive to the presence of a 3' overhang as characterized by a rightward shift of the binding curve (Fig. S2.16).

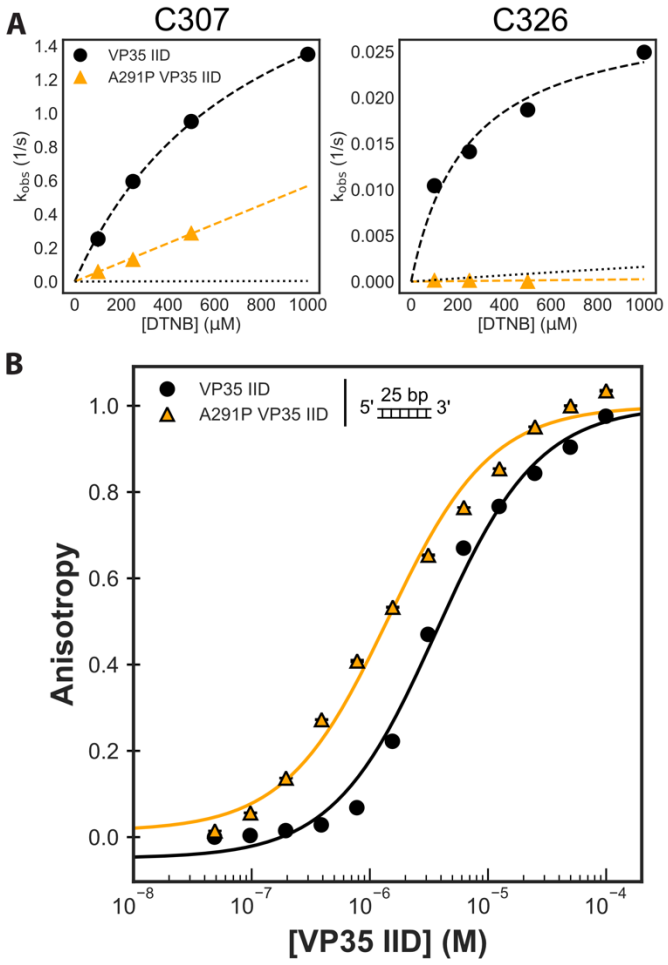

Figure 2.5 An A291P mutation favors both the closed cryptic pocket in VP35's IID and increases dsRNA binding.

A) DTNB observed labeling rates of both wild-type and the A291P mutation for the two cysteines in the pocket. All four cysteines are present, complete data for all four observed rates are in Figure S2.11. B) Binding of both C247S/C275S and A291P VP35 IID to a fluorescently labeled 25-bp double-stranded RNA. The anisotropy was calculated from measured fluorescence polarization and fit to a single-site binding model (black and orange lines). The means and standard deviations from three replicates are shown but error bars are generally smaller than the symbols. Anisotropy was normalized to the max anisotropy for each dataset.

2.4.7 Stabilizing the open cryptic pocket allosterically disrupts binding to dsRNA blunt ends.

We reasoned that covalent attachment of TNB to the cysteine sidechains pointing into the pocket (C307 and C326) would provide a means to capture the open pocket and assess the impact of stabilizing this state with a drug-sized probe on dsRNA binding. Addition of TNB to these cysteines is sterically incompatible with the closed conformation of VP35's RNA bound IID that has been observed crystallographically. TNB's mass of ~198 Da is also similar to many drug fragments used in screening campaigns, making it a reasonable surrogate for the type of effect one might achieve with a fragment hit. Given that we already know DTNB labels the IID's cysteines, a TNB-labeled sample is easily obtainable by waiting until the labeling reaction goes to completion. Finally, we have previously used this same strategy to identify cryptic pockets that exert allosteric control over the activity of β-lactamase enzymes.[12,56]

To specifically probe the behavior of effects of labeling the cryptic pocket, we focus on a C247S/C275S variant that only has cysteines in the cryptic pocket. As with the wild-type protein, thiol labeling of the C247S/C275S variant is consistent with the formation of the predicted cryptic pocket (Fig. 2.4D).

Comparing the dsRNA-binding profile of TNB-labeled protein (TNB-VP35 IID) to unlabeled protein reveals that labeling allosterically reduces the affinity for blunt-ended dsRNA by at least 5-fold (Fig. 2.6A). Solubility limitations prevented us from observing complete binding curves for labeled protein, but the data are sufficient to demonstrate that TNB-labeling has at least as strong an effect on binding as addition of a 3' overhang. As a control to ensure that labeling does not disrupt binding by simply unfolding the protein, we measured the circular dichroism (CD) spectra of labeled and unlabeled protein. The similarity between the CD spectra (Fig. 2.6B) demonstrates that the IID's overall fold is not grossly perturbed. Previous work demonstrated

that VP35's two subdomains do not fold independently[34] supporting our proposal that both domains remain mostly folded. These data indicate that the change in dsRNA binding from TNB-labeled VP35 is unlikely to be due to a local unfolding of the β-sheet subdomain. Furthermore, since past work demonstrated that reducing the blunt end-binding affinity by as little as 3-fold is sufficient to allow a host to mount an effective immune response,[35,36] targeting our cryptic pocket could be of great therapeutic value.

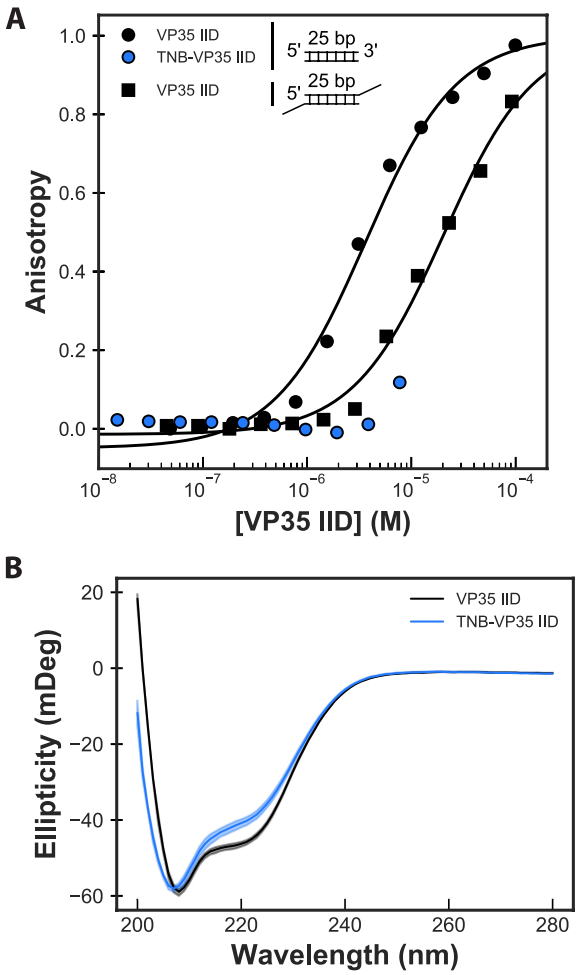

Figure 2.6 Stabilizing the open cryptic pocket in VP35's IID disrupts dsRNA binding.

A) Binding of unlabeled C247S/C275S VP35 IID to two different dsRNA constructs compared to binding of TNB-labeled protein to blunt-ended RNA. The two RNA constructs both have a 25-bp double-stranded segment, and one has 2 nucleotide overhangs on the 3' ends. The anisotropy was measured via a fluorescence polarization assay, converted to anisotropy, fit to a single-site binding model (black lines), and normalized to the fit maximum anisotropy. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. B) Circular dichroism (CD) spectra of labeled and unlabeled protein demonstrate that labeling does not unfold the protein. The opaque and semi-transparent lines represent the mean and standard deviation, respectively, from three replicates. CD spectra were collected in 50 mM Sodium Phosphate pH 7 at 50 μg/mL protein.

2.5 Discussion

We have identified a cryptic allosteric site in the IID of the Ebola VP35 protein that provides a new opportunity to target this essential viral component. Past work identified several sites within the VP35 IID that are critical for immune evasion and viral replication,[30,33,38,39] but structural snapshots captured crystallographically lacked druggable pockets.[35,59] We used adaptive sampling simulations to access more of the ensemble of conformations that VP35 adopts, uncovering an unanticipated cryptic pocket. While the pocket directly coincides with the interface that binds the backbone of dsRNA, it was not clearly of therapeutic relevance since binding dsRNA's blunt ends is more important for Ebola's immune evasion mechanism.[60] However, our simulations also suggested the cryptic pocket is allosterically coupled to the blunt end-binding interface and, therefore, could modulate this biologically-important interaction. Analysis of our computational model suggested that structures with an open cryptic pocket should be less compatible with binding to RNA blunt ends than structures with a closed pocket. Subsequent thiol labeling experiments confirmed that fluctuations within the folded state of the IID expose two buried cysteines that line the proposed cryptic pocket to solvent. Introducing an F239A mutation within the blunt-end binding interface allosterically increases the probability of cryptic pocket opening.[61-63] An A291P mutation allosterically suppresses pocket opening and simultaneously increases the affinity of VP35 for dsRNA. Finally, covalently modifying the pocket facing cysteines to stabilize the open form of the cryptic pocket allosterically disrupts binding to dsRNA blunt ends by at least 5-fold. Previous work demonstrated that reducing the binding affinity by as little as 3-fold is sufficient to allow a host to mount an effective immune response.[35] Therefore, it may be possible to attenuate the impact of viral replication and restrict pathogenicity by designing small molecules to target the cryptic allosteric site we report here.

More generally, our results speak to the power of simulations to provide simultaneous access to both hidden conformations and dynamics with atomic resolution. Such information is extremely difficult to obtain from single structural snapshots or powerful techniques that report on dynamics without directly yielding structures, such as NMR and hydrogen deuterium exchange. As a result, simulations are a powerful means to uncover unanticipated features of proteins' conformational ensembles, such as cryptic pockets and allostery, providing a foundation for the design of further experiments. We anticipate such simulations will enable the discovery of cryptic pockets and cryptic allosteric sites in other proteins, particularly those that are currently considered difficult targets. Furthermore, the detailed structural insight from simulations will facilitate the design of small molecule drugs that target these sites.

2.6 Methods

2.6.1 Molecular Dynamics Simulations And Analysis

Simulations were initiated from the apo protein model of PDB 3FKE [34,35] and run with Gromacs[64] using the amber03 force field[65] and TIP3P explicit solvent[66] at a temperature of 300 K and 1 bar pressure, as described previously.[67] Recombinant VP35 IID is known to be monomeric supporting our choice in system setup. We first applied our FAST-pockets algorithm[44] to balance 1) preferentially simulating structures with large pocket volumes that may harbor cryptic pockets with 2) broad exploration of conformational space. For FAST, we performed 10 rounds of simulations with 10 simulations/round and 80 ns/simulation. To acquire better statistics across the landscape, we performed an RMSD-based clustering using a hybrid k-centers/k-medoids algorithm[68] implemented in Enspara[69] to divide the data into 1,000 clusters. Then we ran three simulations initiated from each cluster center on the Folding@home distributed computing environment, resulting in an aggregate simulation time of 122 μs.

Exposons were identified using our previously described protocols,[12] as implemented in Enspara.[69] Briefly, the solvent accessible surface area (SASA) of each residue's side-chain was calculated using the Shrake-Rupley algorithm[70] implemented in MDTraj[71] using a drug-sized probe (2.8 Å sphere). Conformations were clustered based on the SASA of each residue using a hybrid k-centers/k-medoids algorithm, using a 2.7 $Å^2$ distance cutoff and 5 rounds of k-medoids updates. A Markov time of 6 ns was selected based on the implied timescales test (Fig. S2.17). The center of each cluster was taken as an exemplar of that conformational state, and residues were classified as exposed if their SASA exceeded 2.0 $Å^2$ and buried otherwise. The mutual information between the burial/exposure of each pair of residues was then calculated based on

the MSM (i.e., treating the centers as samples and weighting them by the equilibrium probability of the state they represent). Finally, exposons were identified by clustering the matrix of pairwise mutual information values using affinity propagation.[72]

The CARDS algorithm[54] was applied to identify allosteric coupling using our established protocols,[73] as implemented in Enspara.[69] Briefly, each dihedral angle in each snapshot of the simulations was assigned to one of three rotameric states (gauche+, gauche-, or trans) and one of two dynamical states (ordered or disordered). The total coupling between each pair of dihedrals $X$and Y was then calculated as $I(X_R, Y_R) + I(X_R, Y_D) + I(X_D, Y_R) + I(X_D, Y_D)$, where I is the mutual information metric, X_R is the rotameric state of dihedral X, and X_D is the dynamical state of dihedral X. The term $I(X_R, Y_R)$ is the purely structural coupling, while the sum of the other three terms is referred to as the disorder-mediated coupling. The dihedral level couplings were coarse-grained into residue-level coupling by summing the total coupling between all the relevant dihedrals. The network was subsequently filtered to only retain significant edges.[74] Finally, communities of coupled residues were identified by clustering the residue-level matrix of total couplings using affinity propagation.[72] These algorithms are available at github.com/bowman-lab.

We processed, trained and analyzed our DiffNet as previously described.[10] Briefly, we isolated coordinates of the heavy atoms (all protein atoms excluding hydrogens) for trajectories of our two ensembles of pocket open and closed states using a 1.5Å cutoff for the distance between the center of mass of residues 305 to 310 (helix 5) and the center of mass of residues 238 to 245 (helix 2). We then centered the atom coordinates at the origin and aligned to PDB 3FKE. Next, we mean shifted then whitened the coordinates. Finally, we used the resulting data to train the neural net for 20 epochs with 30 latent space variables with a batch size of 32. Frames with the

pocket closed were initially assigned a classification label of zero while frames with the pocket open were assigned a label of one. For expectation maximization, we set the initial bounds as 10 – 40% for closed frames then 60 – 90% for open frames. We then visualized the top 250 correlated distance changes in PyMol.

We used the calc_chi1 function in MdTraj to calculate the F239 χ_1 in our original MSM. We then binned the χ_1 values as guache+, trans, or gauche- using previously described cuttoffs.[75] To estimate the error in our rotamer distribution in figure 2.3D, we randomly selected N trajectories from the original dataset where N = number of total original trajectories in our dataset, with replacement. We then refit the MSM as described above, keeping the same state space but with the resampled trajectories, twenty-five total times. Then we calculated the total population of each rotamer in the resampled datasets, and the mean population of that rotamer across all resampled MSMs. The error bars are then the standard deviation of the mean of the resampled population values for each rotamer in the open and closed ensembles with respect to the refit MSMs.

2.6.2 Protein Expression And Purification

All variants of VP35's IID were purified from the cytoplasm of *E. coli* BL21(DE3) Gold cells (Agilent Technologies).[34,36,61-63] Variants were generated using the site directed mutagenesis method and confirmed by DNA sequencing. Transformed cells were grown at 37°C until OD 0.3 then grown at 18°C until induction at OD 0.6 with 1 mM IPTG (Gold Biotechnology, Olivette, MO). Cells were grown for 15 hours then centrifuged after which the pellet was resuspended in 20 mM Sodium Phosphate pH 8, 1 M sodium chloride, with 5.1 mM β-mercaptoethanol. Resuspended cells were subjected to sonication at 4°C followed by centrifugation. The

supernatant was then subjected to Ni-NTA affinity (BioRad Bio-Scale Mini Nuvia IMAC column), TEV digestion, cation exchange (BioRad UNOsphere Rapid S column), and size exclusion chromatography (BioRad Enrich SEC 70 column or Cytiva HiLoad 16/600 Superdex 75) into 10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 2 mM TCEP.

2.6.3 Thiol Labeling

We monitored the change in absorbance over time of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Ellman's reagent, Thermo Fisher Scientific). Various concentrations of DTNB were added to protein and change in absorbance was measured in either an SX-20 Stopped Flow instrument (Applied Photophysics, Leatherhead, UK), or an Agilent Cary60 UV-vis spectrophotometer at 412 nm until the reaction reached steady state (~300 s). Data were fit with a Linderstrøm-Lang model to extract the thermodynamics and/or kinetics of pocket opening, as described in detail previously.[12] As a control, the equilibrium constant for folding and the unfolding rate were measured (Fig. S2.12) and used to predict the expected labeling rate from the unfolded state. The equilibrium constant was inferred from a two-state fit to urea melts monitored by fluorescence and unfolding rates were inferred from exponential fits to unfolding curves monitored by fluorescence after the addition of urea, as described previously.[12,56,76] Fluorescence data were collected using a Jasco FP-8300 Spectrofluorometer with Jasco ETC-815 Peltier and Koolance Exos2 Liquid Coolant-controlled cuvette holder.

2.6.4 Fluorescence Polarization Binding Assay

Binding affinities between variants of VP35's IID and dsRNA were measured using fluorescence polarization in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$. A 25 base pair FITC-dsRNA

(Integrated DNA Technologies) substrate with and without a 2 nucleotide 3' overhang was included at 100 nM. The sample was equilibrated for one hour before data collection. Data were collected on a BioTek Synergy2 Multi-Mode Reader as polarization and were converted to anisotropy as described previously.[58] TNB-labeled samples were generated by allowing DTNB and VP35's IID to react for 3 minutes and then removing excess DTNB with a Zeba spin desalting columns (Thermo Fisher Scientific). Data were analyzed in Jupyter Notebook using Scipy 1.3.2, NumPy 1.14.x and 1.19.5, Matplotlib 3.5, Pandas 0.25.3, and Seaborn 0.11.2. A single-site binding model was sufficient to fit the data:

$$r_{obs} = r_0 + (r_{max} - r_0)\left(\frac{K_A \cdot [VP35]}{1 + K_A \cdot [VP35]}\right) \text{ (Equation 1)}$$

2.7 References

1 Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nature Reviews Drug Discovery* **3**, 301-317, doi:10.1038/nrd1343 (2004).

2 Johnson, D. K. & Karanicolas, J. Computational Screening and Design for Compounds that Disrupt Protein-protein Interactions. *Current Topics in Medicinal Chemistry* **17**, 2703-2714, doi:10.2174/1568026617666170508153904 (2017).

3 Hopkins, A. & Groom, C. The druggable genome. *Nat Rev Drug Discov* **1**, 727-730 (2002).

4 Wilson, C. *et al.* Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882-886 (2015).

5 Schames, J. *et al.* Discovery of a novel binding trench in HIV integrase. *J. Med. Chem* **47**, 1879-1881 (2004).

6 Jakubík, J. & El-Fakahany, E. E. Allosteric Modulation of Muscarinic Acetylcholine Receptors. *Pharmaceuticals* **3**, 2838-2860 (2010).

7 Knoverek, C. R., Amarasinghe, G. K. & Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends Biochem Sci* **44**, 351-364, doi:10.1016/j.tibs.2018.11.007 (2019).

8 Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Current opinion in chemical biology* **44**, 1-8, doi:10.1016/j.cbpa.2018.05.003 (2018).

9 Hardy, J. A. & Wells, J. A. Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* **14**, 706-715 (2004).

10 Ward, M. D. *et al.* Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. *Nature Communications* **12**, 3023, doi:10.1038/s41467-021-23246-1 (2021).

11 Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11681-11686, doi:10.1073/pnas.1209309109 (2012).

12 Porter, J. R. *et al.* Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophysical Journal* **116**, 818-830, doi:10.1016/j.bpj.2018.11.3144 (2019).

13 Johnson, D. K. & Karanicolas, J. Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface. *PLoS Computational Biology* **9**, e1002951, doi:10.1371/journal.pcbi.1002951 (2013).

14 Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *Journal of the American Chemical Society* **138**, 14257-14263, doi:10.1021/jacs.6b05425 (2016).

15 Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J Mol Biol* **428**, 709-719, doi:10.1016/j.jmb.2016.01.029 (2016).

16 Schmidt, D., Boehm, M., McClendon, C. L., Torella, R. & Gohlke, H. Cosolvent-Enhanced Sampling and Unbiased Identification of Cryptic Pockets Suitable for Structure-Based Drug Design. *J Chem Theory Comput* **15**, 3331-3343, doi:10.1021/acs.jctc.8b01295 (2019).

17 Cuchillo, R., Pinto-Gil, K. & Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *J Chem Theory Comput* **11**, 1292-1307, doi:10.1021/ct501072t (2015).

18 Ghanakota, P. & Carlson, H. A. Moving Beyond Active-Site Detection: MixMD Applied to Allosteric Systems. *The Journal of Physical Chemistry B* **120**, 8685-8695, doi:10.1021/acs.jpcb.6b03515 (2016).

19 Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J. & Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Accounts of Chemical Research*, acs.accounts.9b00613, doi:10.1021/acs.accounts.9b00613 (2020).

20 Wassman, C. D. *et al.* Computational identification of a transiently open L1/S3 pocket for reactivation of mutant p53. *Nat Commun* **4**, 1407, doi:10.1038/ncomms2361 (2013).

21 Ni, D. *et al.* Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. *Chemical Science* **12**, 464-476, doi:10.1039/D0SC05131D (2021).

22 Lu, S. *et al.* Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Communications* **12**, 4721, doi:10.1038/s41467-021-25020-9 (2021).

23 Raich, L. *et al.* Discovery of a hidden transient state in all bromodomain families. *Proceedings of the National Academy of Sciences* **118**, e2017427118, doi:10.1073/pnas.2017427118 (2021).

24 Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Current Opinion in Chemical Biology* **44**, 1-8, doi:10.1016/j.cbpa.2018.05.003 (2018).

25 Zimmerman, M. I. *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nature Chemistry* **13**, 651-659, doi:10.1038/s41557-021-00707-0 (2021).

26 Cross, R. W., Mire, C. E., Feldmann, H. & Geisbert, T. W. Post-exposure treatments for Ebola and Marburg virus infections. *Nat Rev Drug Discov* **17**, 413-434, doi:10.1038/nrd.2017.251 (2018).

27 Keshwara, R., Johnson, R. F. & Schnell, M. J. Toward an Effective Ebola Virus Vaccine. *Annual review of medicine* **68**, 371-386, doi:10.1146/annurev-med-051215-030919 (2017).

28 Mulangu, S. *et al.* A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *New England Journal of Medicine* **381**, 2293-2303, doi:10.1056/NEJMoa1910993 (2019).

29 Mulangu, S. *et al.* A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. *N Engl J Med* **381**, 2293-2303, doi:10.1056/NEJMoa1910993 (2019).

30 Messaoudi, I., Amarasinghe, G. K. & Basler, C. F. Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. *Nature Reviews. Microbiology* **13**, 663-676, doi:10.1038/nrmicro3524 (2015).

31 Basler, C. F. *et al.* The Ebola virus VP35 protein inhibits activation of interferon regulatory factor 3. *J Virol* **77**, 7945-7956, doi:10.1128/jvi.77.14.7945-7956.2003 (2003).

32 Cárdenas, W. B. *et al.* Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. *Journal of virology* **80**, 5168-5178, doi:10.1128/JVI.02199-05 (2006).

33 Hartman, A. L., Towner, J. S. & Nichol, S. T. A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. *Virology* **328**, 177-184, doi:10.1016/j.virol.2004.07.006 (2004).

34 Leung, D. W. *et al.* Structure of the Ebola VP35 interferon inhibitory domain. *Proceedings of the National Academy of Sciences* **106**, 411-416, doi:10.1073/pnas.0807854106 (2009).

35 Leung, D. W. *et al.* Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nat Struct Mol Biol* **17**, 165-172, doi:10.1038/nsmb.1765 (2010).

36 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Reports* **14**, 1632-1640, doi:10.1016/j.celrep.2016.01.049 (2016).

37 Hartman, A. L., Dover, J. E., Towner, J. S. & Nichol, S. T. Reverse genetic generation of recombinant Zaire Ebola viruses containing disrupted IRF-3 inhibitory domains results in attenuated virus growth in vitro and higher levels of IRF-3 activation without inhibiting viral transcription or replication. *J Virol* **80**, 6430-6440, doi:10.1128/JVI.00044-06 (2006).

38 Prins, K. C. *et al.* Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. *J Virol* **84**, 3004-3015, doi:10.1128/JVI.02459-09 (2010).

39 Prins, K. C. *et al.* Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. *Journal of virology* **84**, 10581-10591, doi:10.1128/JVI.00925-10 (2010).

40 Brown, C. S. *et al.* In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. *J Mol Biol* **426**, 2045-2058, doi:10.1016/j.jmb.2014.01.010 (2014).

41 Glanzer, J. G. *et al.* In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. *Bioorganic & medicinal chemistry* **24**, 5388-5392, doi:10.1016/j.bmc.2016.08.065 (2016).

42 Liu, G. *et al.* A Sensitive in Vitro High-Throughput Screen To Identify Pan-filoviral Replication Inhibitors Targeting the VP35–NP Interface. *ACS Infectious Diseases* **3**, 190-198, doi:10.1021/acsinfecdis.6b00209 (2017).

43 Daino, G. L. *et al.* Identification of Myricetin as an Ebola Virus VP35–Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. *Biochemistry* **57**, 6367-6378, doi:10.1021/acs.biochem.8b00892 (2018).

44 Zimmerman, M. I. & Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of chemical theory and computation* **11**, 5747-5757, doi:10.1021/acs.jctc.5b00737 (2015).

45 Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99-105, doi:10.1016/j.ymeth.2010.06.002 (2010).

46 Prinz, J.-H. *et al.* Markov models of molecular kinetics: generation and validation. *J Chem Phys* **134**, 174105, doi:10.1063/1.3565032 (2011).

47 Beglov, D. *et al.* Exploring the structural origins of cryptic sites on proteins. *Proc Natl Acad Sci U S A* **115**, E3416-E3425, doi:10.1073/pnas.1711490115 (2018).

48 Ngan, C. H. *et al.* FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* **40**, W271-275, doi:10.1093/nar/gks441 (2012).

49 Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* **10**, 733-755, doi:10.1038/nprot.2015.043 (2015).

50 Kozakov, D. *et al.* Structural conservation of druggable hot spots in protein–protein interfaces. *Proceedings of the National Academy of Sciences* **108**, 13528-13533, doi:10.1073/pnas.1101835108 (2011).

51 Kozakov, D. *et al.* The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature Protocols* **10**, 733-755, doi:10.1038/nprot.2015.043 (2015).

52 Brenke, R. *et al.* Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **25**, 621-627, doi:10.1093/bioinformatics/btp036 (2009).

53 Bohnuud, T. *et al.* Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Research* **40**, 7644-7652, doi:10.1093/nar/gks519 (2012).

54 Singh, S. & Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. *Journal of chemical theory and computation* **13**, 1509-1517, doi:10.1021/acs.jctc.6b01181 (2017).

55 Bernstein, R., Schmidt, K., Harbury, P. & Marqusee, S. Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 10532-10537 (2011).

56 Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2734-2739, doi:10.1073/pnas.1417811112 (2015).

57 Leung, D. W. *et al.* Structural and Functional Characterization of Reston Ebola Virus VP35 Interferon Inhibitory Domain. *Journal of Molecular Biology* **399**, 347-357, doi:10.1016/j.jmb.2010.04.022 (2010).

58 Kozlov, A. G., Galletto, R. & Lohman, T. M. SSB-DNA binding monitored by fluorescence intensity and anisotropy. *Methods Mol Biol* **922**, 55-83, doi:10.1007/978-1-62703-032-8_4 (2012).

59 Leung, D. W. *et al.* Structure of the Ebola VP35 interferon inhibitory domain. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 411-416, doi:10.1073/pnas.0807854106 (2009).

60 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell reports* **14**, 1632-1640, doi:10.1016/j.celrep.2016.01.049 (2016).

61 Leung, D. W. *et al.* Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nature Structural & Molecular Biology* **17**, 165-172, doi:10.1038/nsmb.1765 (2010).

62 Leung, D. W. *et al.* Crystallization and preliminary X-ray analysis of Ebola VP35 interferon inhibitory domain mutant proteins. *Acta Crystallographica Section F* **66**, 689-692, doi:doi:10.1107/S1744309110013266 (2010).

63 Leung, D. W. *et al.* Expression, purification, crystallization and preliminary X-ray studies of the Ebola VP35 interferon inhibitory domain. *Acta Crystallographica Section F* **65**, 163-165, doi:doi:10.1107/S1744309108044187 (2009).

64 van der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* **26**, 1701-1718, doi:10.1002/jcc.20291 (2005).

65 Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **24**, 1999-2012, doi:10.1002/jcc.10349 (2003).

66 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926, doi:10.1063/1.445869 (1983).

67 Hart, K. M., Ho, C. M. W., Dutta, S., Gross, M. L. & Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nature Communications* **7**, 12965, doi:10.1038/ncomms12965 (2016).

68 Beauchamp, K. A. *et al.* MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J Chem Theor Comp* **7**, 3412-3419, doi:10.1021/ct200463m (2011).

69 Porter, J. R., Zimmerman, M. I. & Bowman, G. R. Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *The Journal of chemical physics* **150**, 044108, doi:10.1063/1.5063794@jcp.2019.MMMK.issue-1 (2019).

70 Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351-371 (1973).

71 McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528-1532, doi:10.1016/j.bpj.2015.08.015 (2015).

72 Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-976, doi:10.1126/science.1136800 (2007).

73 Sun, X., Singh, S., Blumer, K. J. & Bowman, G. R. Simulation of spontaneous G protein activation reveals a new intermediate driving GDP unbinding. *eLife* **7**, 19, doi:10.7554/eLife.38465 (2018).

74 Dianati, N. Unwinding the hairball graph: Pruning algorithms for weighted complex networks. *Physical Review E* **93**, 012304, doi:10.1103/PhysRevE.93.012304 (2016).

75 Roland L. Dunbrack Jr., F. E. C. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science : A Publication of the Protein Society* **6**, 1661-1681, doi:10.1002/pro.5560060807 (2008).

76 Zimmerman, M. I. *et al.* Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS Central Science* **3**, 1311-1321, doi:10.1021/acscentsci.7b00465 (2017).

2.8 Supplementary Materials

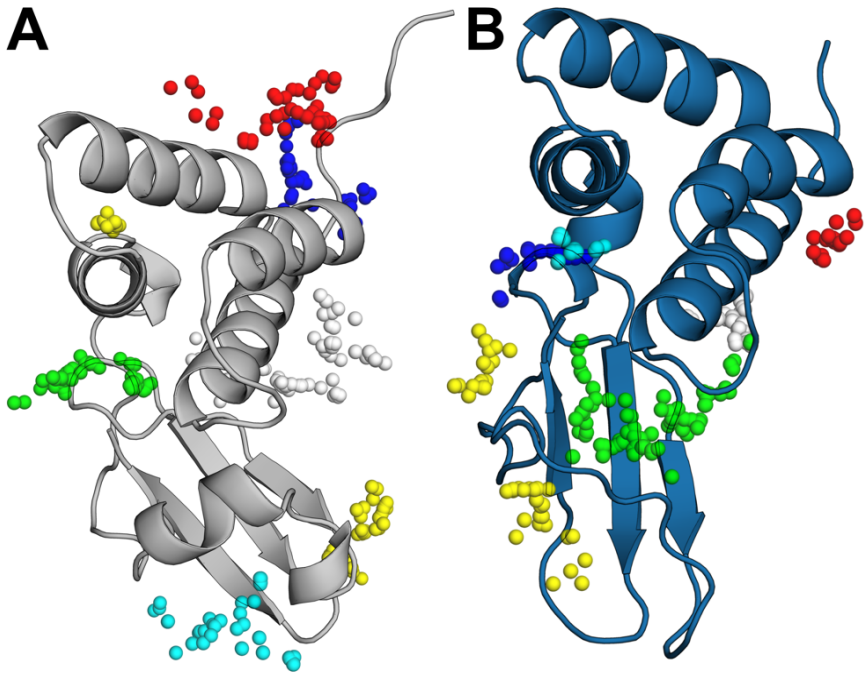

Figure S2.7 Fpocket algorithm highlights the cryptic pocket in VP35.

Fpocket result for the 3FKE crystal structure (gray) and for an open state from our MSM (blue) highlighting the opening of a druggable pocket. The groups of multicolored spheres are pockets computed using fpocket https://github.com/Discngine/fpocket which captures pockets in protein structures. In A) we applied the fpocket algorithm to detect pockets in the VP35 crystal structure. The shown pockets all have druggability scores less than 0.5 excepting the pocket highlighted with white spheres which has a druggability score of 0.578 and is known binding site for an inhibitor of nucleoprotein-VP35 interaction. B) Fpocket applied to a highly populated, open state from our MSM. The green spheres highlight a pocket where our methods also report a pocket. The druggability score of this pocket is 0.681 higher than the proposed druggability cutoff of 0.5.

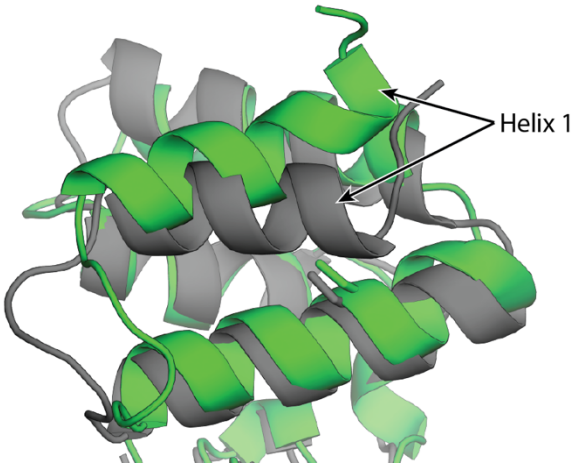

Figure S2.8 States from VP35 simulations highlighting the motion of helix 1 exposing C247.

Motion of helix 1 (green vs gray structures) sometimes exposes C247 (sticks) to solvent. However, the resulting pocket is small and FTMAP does not identify any hotspots in this region that are likely to bind drug-like molecules. Therefore, we focus our attention on the cryptic pocket created by the displacement of helix 5.

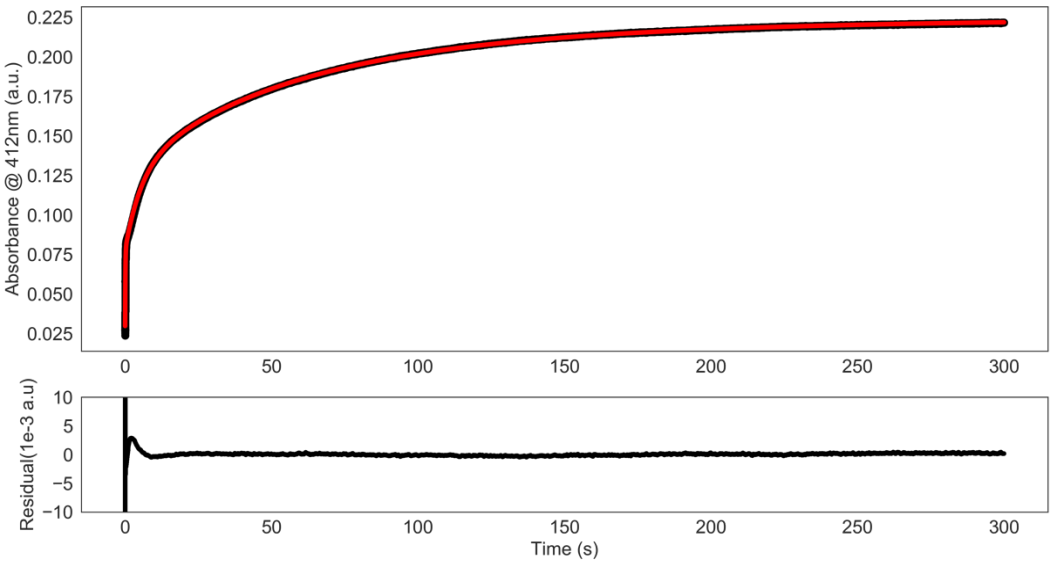

Figure S2.9 A representative time trace and fit line from a thiol labeling with residuals.

A representative time trace from a thiol labeling experiment (black) performed at 100 µM DTNB and a quadruple exponential fit (red).The data are background subtracted (e.g. the average absorbance from three runs with DTNB but no protein were subtracted) to account for spontaneous hydrolysis of DTNB. Source data are provided as a Source Data file.

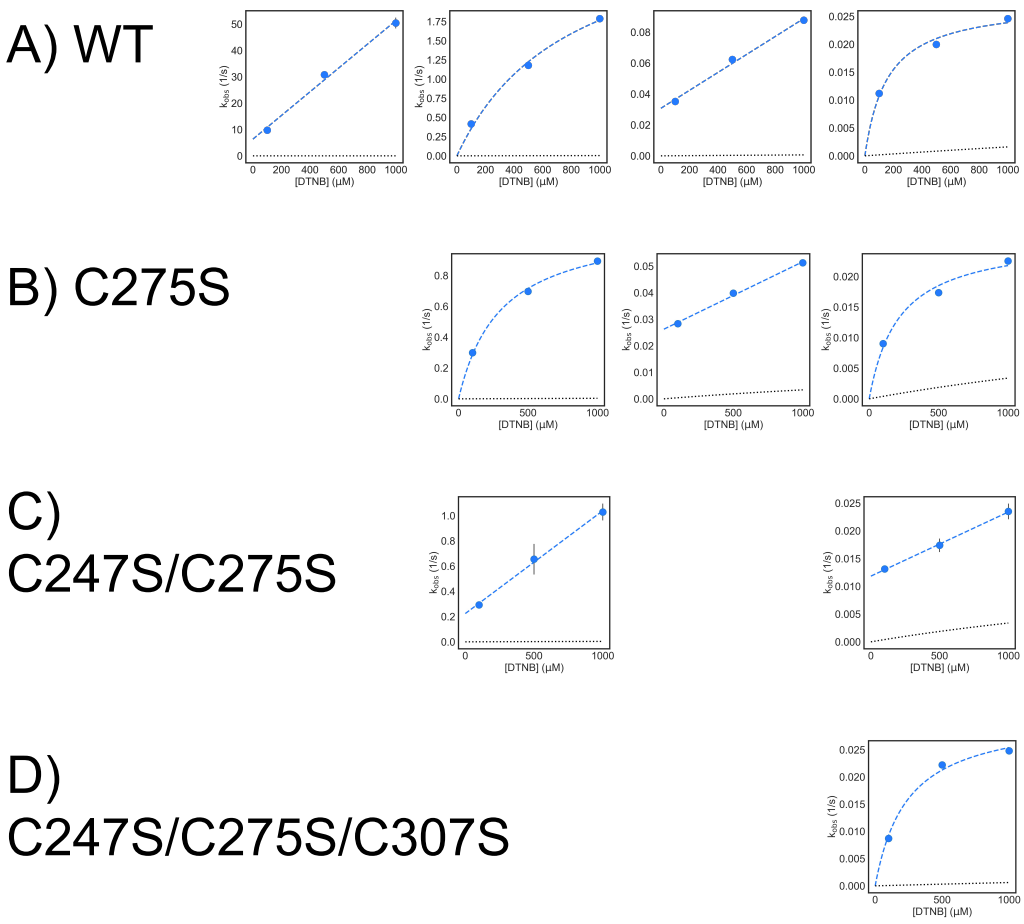

Figure S2.10 Representative Kobs vs [DTNB] plots from thiol labeling experiments for VP35 with fits of the Linderstrøm-Lang Model.

Representative Kobs vs [DTNB] plots from thiol labeling experiments performed at 100. 500 and 1000 μM DTNB (blue circles) and fit of the Linderstrøm-Lang Model (dashed lines). A) wild-type observed rates. B) C275S observed rates C) C247S/C275S observed rates. D) C247S/C275S/C307S observed rates. For all panels, fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. Source data are provided as a Source Data file.

Variant	C275 rate (s^{-1})	C307 rate (s^{-1})	C247 rate (s^{-1})	C326 rate (s^{-1})
Wild-type	1.4 ± 0.02	0.201 ± 0.002	0.024 ± 0.0004	0.011 ± 0.000030
C275S	-	0.30 ± 0.002	0.028 ± 0.0003	0.0090 ± 0.00003
C247S/C275S	-	0.29 ± 0.01	-	0.013 ± 0.0004
C247S/C275S/C307S	-	-	-	0.0087 ± 0.0003
C247S/C275S/C326S	-	0.20 ± 0.007	-	-

Figure S2.11 Observed labeling rates for VP35 cysteine substituted variants.

Observed labeling rates at 100 μM DTNB for a set of variants with different cysteines mutated to serines to uncover which rate in the wild-type fit corresponds to which cysteine residue. Error is standard deviation from three replicates. Dash represents rates not measured due to the absence of that cysteine residue. Source data are provided as a Source Data file.

Variant	K	k_{unfold} (s^{-1})
Wild-type	$6.57 \pm 4.0 \times 10^{-5}$	0.0175
C247S/C275S	$4.01 \pm 0.8 \times 10^{-4}$	0.0083
F239A	$4.67 \pm 1.0 \times 10^{-4}$	0.0179
A291P	$3.83 \pm 4.0 \times 10^{-5}$	0.00274

Figure S2.12 Measured VP35 variant equilibrium constants and unfolding rates.

Characterization of the folding/unfolding of VP35's IID used to test whether the observed thiol labeling is due to fluctuations within the native state or global unfolding of the protein. K is the equilibrium constant between the folded and unfolded state determined from denaturation data, k_{unfold} is the unfolding rate of the respective variants measured by intrinsic tryptophan fluorescence. Source data are provided as a Source Data file.

Residue	k_{int} ($\mu M^{-1} s^{-1}$)
C247	0.0566 ± 0.0007
C275	0.00254 ± 0.001
C307	0.0290 ± 0.002
C326	0.395 ± 0.02

Figure S2.13 Intrinsic labeling rates (kint) for each cysteine residue.

Intrinsic labeling rates were measured using either urea unfolded variants containing only the specified cysteine, or peptides containing the specified cysteine and its surrounding residues using an SX 20 stopped-flow uv-vis apparatus. Source data are provided as a Source Data file.

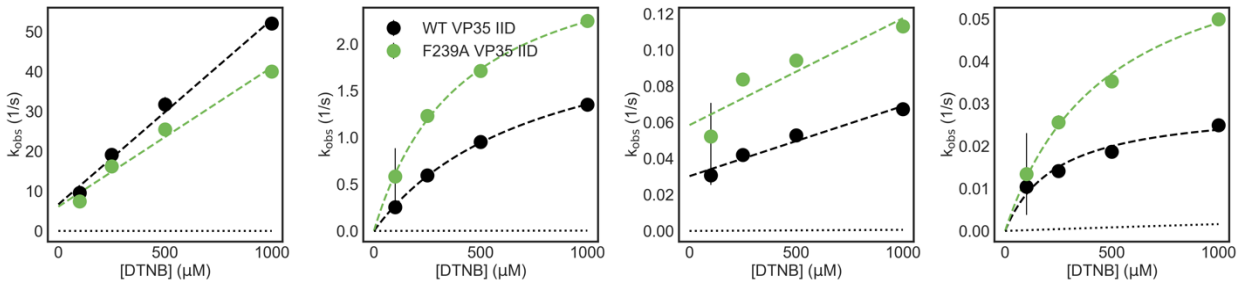

Figure S2.14 Representative Kobs vs [DTNB] plots from thiol labeling experiments for wild-type eVP35 and F239A and fit of the Linderstrøm-Lang Model

Representative Kobs vs [DTNB] plots from thiol labeling experiments performed at 100, 250, 500 and 1000 μM DTNB for wild-type eVP35 (blue circles) and F239A (green circles) and fit of the Linderstrøm-Lang Model (dashed blue and green lines).For all panels, fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. Source data are provided as a Source Data file.

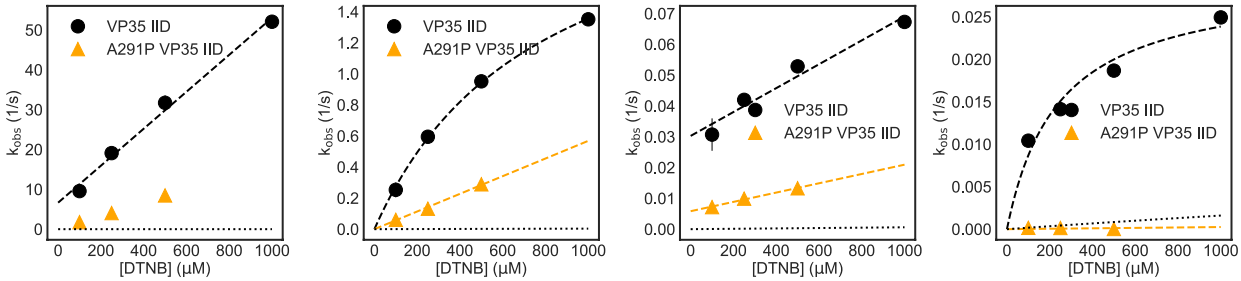

Figure S2.15 Representative Kobs vs [DTNB] plots from thiol labeling experiments for wild-type and A291P VP35 IID and fits of the Linderstrøm-Lang Model.

Representative Kobs vs [DTNB] plots from thiol labeling experiments performed at 100, 250, 500 and 1000 μM DTNB for wild-type (black circles) and 100, 250, and 500 μM DTNB for A291P VP35 IID (orange circles) and fit of the Linderstrøm-Lang Model (dashed black and orange lines). For all panels, fits to the Linderstrøm-Lang model are shown in dashed colored lines and the expected labeling rate from the unfolded state is shown as black dotted lines. The mean and standard deviation from three replicates is shown but error bars are generally smaller than the symbols. Source data are provided as a Source Data file.

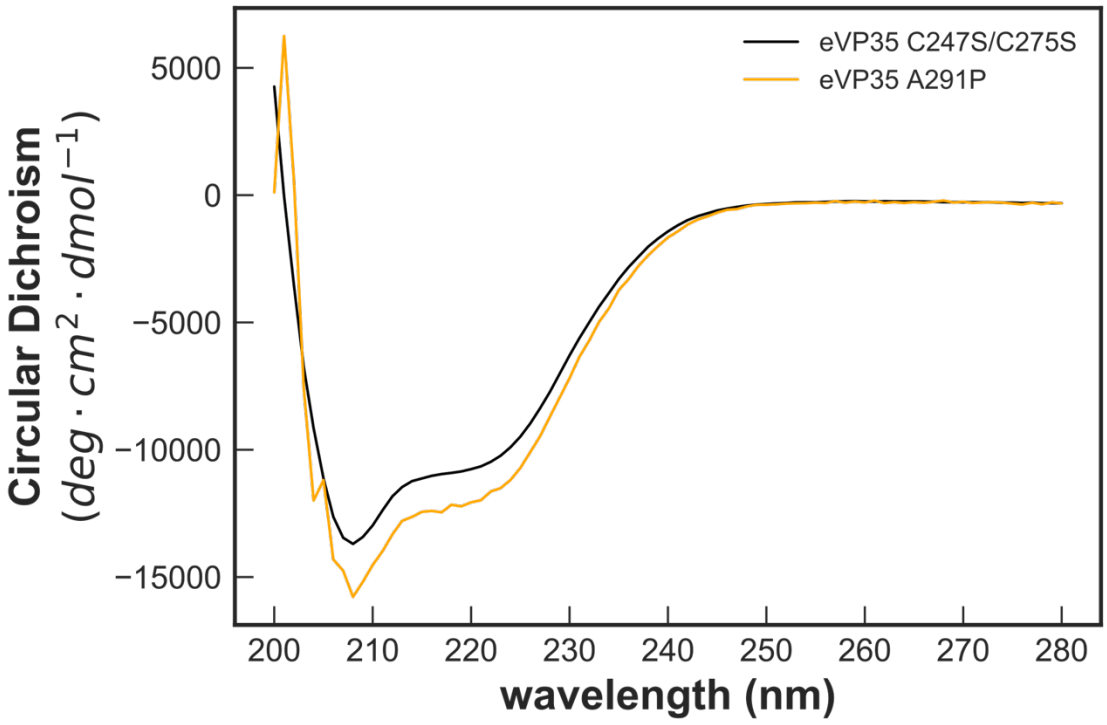

Figure S2.16 Circular dichroism spectra of C247S/C275S and A291P VP35 IID.

Circular dichroism (CD) spectra of C247S/C275S (black) and A291P VP35 IID (orange) protein demonstrate that A291P substitution does not unfold the protein. The opaque and semi-transparent lines represent the mean and standard deviation, respectively, from three replicates. CD spectra were collected in 50 mM Sodium Phosphate pH 7 at 50 µg/mL protein (C247S/C275S) or 20 mM Tris pH 8, 150 mM NaCl at 35 µg/mL (A291P). Source data are provided as a Source Data file.

Length	Sense Strand	Antisense Strand
25mer	56-FAM-rArArArCrUrGrArArArGrGrGrArGrArArArGrUrGrArArArGrUrG	rCrArCrUrUrUrCrArCrUrUrCrUrCrCrCrUrUrUrCrArGrUrUrU
25mer with 2nt 3' overhang	56-FAM-rArArArCrUrGrArArArGrGrGrArGrArArArGrUrGrCrU	rCrArCrUrUrUrCrArCrUrUrCrUrCrCrCrUrUrUrCrArGrUrUrUrCrU

Figure S2.17 RNA sequences used in fluorescence polarization binding assays.

The sense and antisense strands were annealed in a 1:1 molar ratio.[1]

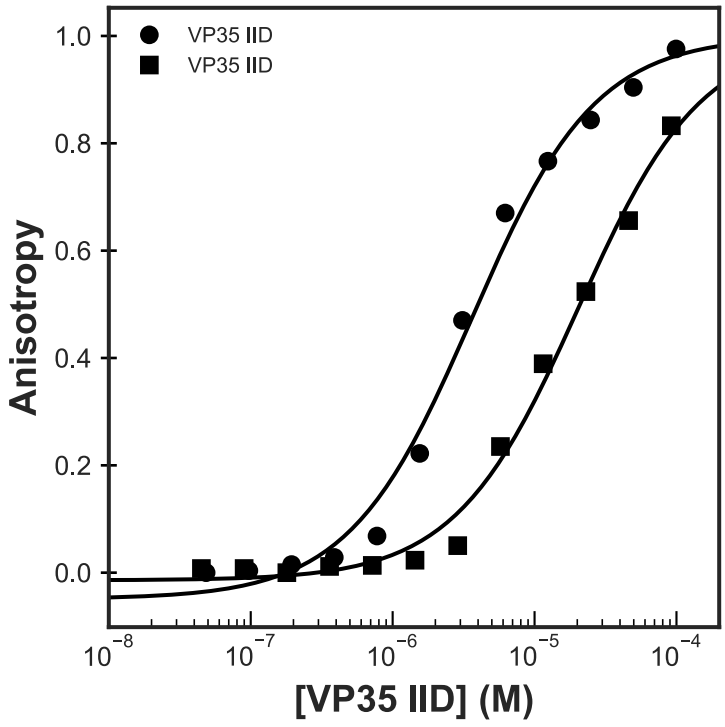

Figure S2.18 Binding of C247S/C275S VP35 IID to a fluorescently labeled 25-bp double-stranded RNAs.

Binding of C247S/C275S VP35 IID to a fluorescently labeled 25-bp double-stranded RNA with (squares) and without a two nucleotide overhang on the 3' end (circles). The anisotropy was calculated from measured fluorescence polarization and fit to a single-site binding model (black lines). The means and standard deviations from three replicates are shown but error bars are generally smaller than the symbols. Anisotropy was normalized to the max anisotropy for each data set. Source data are provided as a Source Data file.

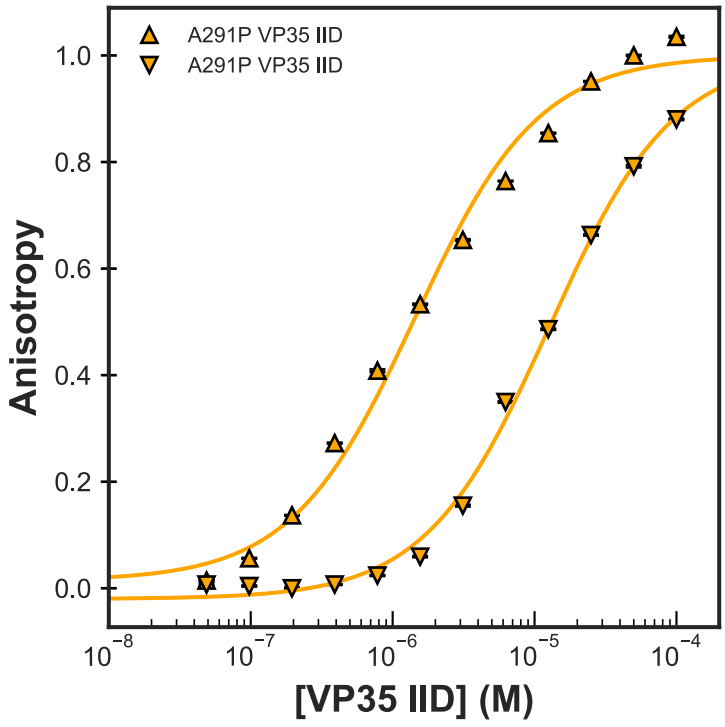

Figure S2.19 Binding of A291P VP35 IID to a fluorescently labeled 25-bp double-stranded RNAs.

Binding of A291P VP35 IID to a fluorescently labeled 25-bp double-stranded RNA with (upward pointing triangles) and without a two nucleotide overhang on the 3' end (downward pointing triangles). The anisotropy was calculated from measured fluorescence polarization and fit to a single-site binding model (orange lines). The means and standard deviations from three replicates are shown but error bars are generally smaller than the symbols. Anisotropy was normalized to the max anisotropy for each data set. Source data are provided as a Source Data file.

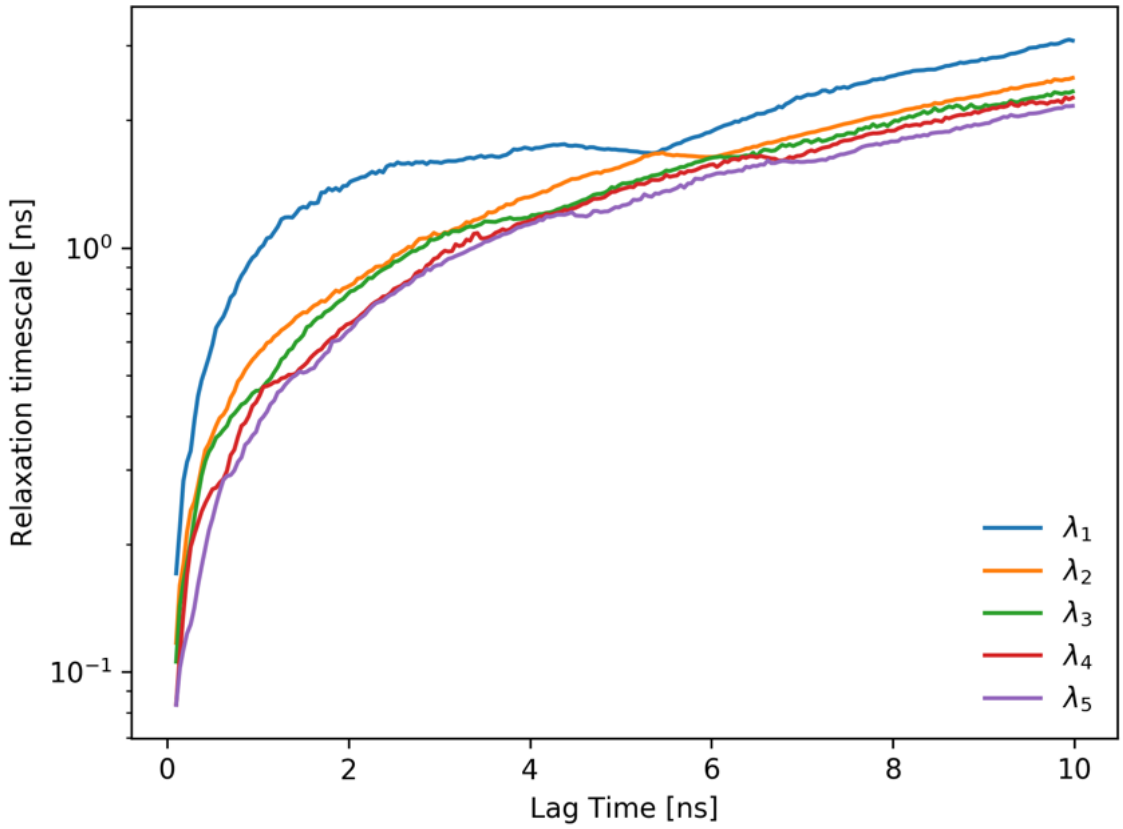

Figure S2.20 Implied timescales test for the VP35 IID MSM suggests the kinetics are stable from 3-6ns.

Analysis in the main text uses a Markov time of 6 ns. Key results were consistent for lag times from 3-6 ns.

2.9 Supplemental Methods

We performed all data analysis and curve fitting using Python 3, Scipy and Numpy, in Jupyter Notebook. These notebooks are available upon request.

2.9.1 Intrinsic Tryptophan Fluorescence Denaturation Experiments

To determine the denaturation midpoint and free energy of folding, eVP35 iid was equilibrated at 0.035 mg/mL in various concentrations of urea (MidSci, St. Louis, MO) from 0 to 8 M in 20 mM Tris pH 8, 150 mM NaCl. At 25ºC using a Jasco FP-8300 Spectrofluorometer the sample was excited with 280 nm light and we collected an emission spectrum from 300-400 nm. The max fluorescence emission at 0 M urea occurs at 322 nm. We recorded the fluorescence value at 322 nm for each concentration of urea in triplicate. Then, we used a six parameter fit for a two-state model of protein unfolding to determine the C_M and folding free energy.[2] The folding free energy (ΔG) was used to calculate the fraction folded and fraction unfolded at 0 M urea. This was repeated for WT eVP35 iid, eVP35 iid C247S/C275S, eVP35 iid F239A, and eVP35 iid A291P.

$$Unfolded \overset{K}{\leftrightarrow} Native \quad \text{(Equation 2)}$$

$$Obs\ Fluor = \frac{Fluor_{U,H2O} + \beta_U[Urea] + \left(Fluor_{N,H2O} + \beta_N[Urea]\right)e^{-\frac{\Delta G_{H2O}+m[Urea]}{RT}}}{1 + e^{-\frac{\Delta G_{H2O}+m[Urea]}{RT}}}$$

(Equation 3)

Fluor is the baseline fluorescence values for the native and unfolded states absent any denaturant. β values are the slopes of the native and unfolded baseline signals.

To determine the unfolding rates, WT eVP35 iid, C247S/C275S, F239A, or A291P were manually injected into a solution of 20 mM Tris pH 8, 150 mM NaCl to a final concentration of 0.03 mg/mL at varying concentrations of urea equal to or above the C_M measured in the equilibrium experiments. Then we collected fluorescence measurements exciting at 280 nm and detecting at 322 nm until the resulting curve reached steady state ~500s or more. The resulting curve was fit to an exponential model to extract the observed unfolding rate at that concentration of urea. Then we plotted the log of the observed rates as a function of urea concentration and linearly extrapolated to the unfolding rate in 0 M urea.

2.9.2 Circular Dichroism Spectra

We evaluated the effect of amino acid substitutions on protein folding and overall secondary structure formation via circular dichroism spectrophotometry. Protein samples were buffer exchanged into the buffer at which they were measured and diluted to 50 or 35 µg/mL (see specific figure legends for variant specific information). We collected spectra at 25˚C in an Applied Photophysics Chirascan equipped with a Quantum Northwest Inc. TC125 Peltier-controlled cuvette holder, reading every integer wavelength between 200 and 280 nm, averaging data for each point for either one or five seconds. Data reported are in either millidegrees, or converted into molar ellipticity using the following equations[3].

$$\theta = \frac{mDeg \cdot MRW}{10 \cdot l \cdot C} \text{ (Equation 4)}$$

$$MRW = \frac{M}{N-1} \text{ (Equation 5)}$$

Here, θ is molar ellipticity, mDeg is millidegrees CD, MRW is the mean residual weight, l is the path length in centimeters, C is the protein concentration in molar, M is the protein molecular weight in g/mol and N is the number of amino acids in the protein.

2.9.3 Rna Fluorescence Anisotropy Experiments

Monitoring the binding of protein to 25-bp dsRNA gives affinities that are consistent with past work. Previous work used a dot-blot assay to measure binding and reported an apparent dissociation constant (K_d) for blunt-ended dsRNA of 3.4±0.07 μM.[1] Similarly, our FP assay gives an apparent K_d of 3.64 ± 0.34 μM for VP35 IID binding blunt-ended dsRNA (Fig. 2.5B). Furthermore it was previously reported that the addition of 2-nucleotide overhang to the 3' end of the RNA reduces VP35 dsRNA-binding affinity by 10-fold.[4] Even when the presence of an overhang inhibits blunt end binding the VP35 IID is still able to bind to the dsRNA backbone though this interaction is weaker than with the blunt ends. Similarly to published data, our FP assay gives an apparent K_d of 20.4±1.1 μM corresponding to at least a five fold reduction in apparent binding to the overhang dsRNA relative to the blunt ended dsRNA. However, an upper baseline could not be captured due to limitations in the protein's solubility, so this apparent K_d is a lower bound. These data are also fit well assuming an apparent K_d of 30.1±7.2 μM that was reported previously.[4] Monitoring the binding of VP35 to 25-bp dsRNA with two blunt ends or 3' overhangs demonstrates that our FP assay is sensitive to both dsRNA-binding modes and gives affinities that are consistent with past work (Fig. 2.5B).

We buffer exchanged and concentrated to ~300 μM VP35 iid then serially diluted by half across the rows of Costar 96 well half area black flat bottom polystyrene plates with a non-binding surface, in triplicate, in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$. Then we added

dsRNA to each well at a final concentration of 100 nM 25 bp FAM-dsRNA. For every experiment we included the following control wells: buffer only, dsRNA only, dsRNA with VP35 at the expected K_D, and dsRNA with VP35 and a previously described dsRNA binding inhibitor[5]. The plates were protected from light and allowed to equilibrate for at least one hour before reading as described in the main-text methods. Data were routinely checked for fluorescence anomalies in the control and experimental wells.

2.9.4 Supplemental References

1 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Reports* **14**, 1632-1640, doi:10.1016/j.celrep.2016.01.049 (2016).

2 Street, T. O., Courtemanche, N. & Barrick, D. in *Methods in Cell Biology* Vol. 84 295-325 (Academic Press, 2008).

3 Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876-2890, doi:10.1038/nprot.2006.202 (2006).

4 Ramanan, P. *et al.* Structural basis for Marburg virus VP35–mediated immune evasion mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20661-20666, doi:10.1073/pnas.1213559109 (2012).

5 Glanzer, J. G. *et al.* In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. *Bioorganic & Medicinal Chemistry* **24**, 5388-5392, doi:10.1016/j.bmc.2016.08.065 (2016).

Chapter 3: Allosteric RNA Binding Inhibition As A Way To Combat Ebolavirus Infection

3.1 Preface

In this chapter, I describe my efforts to discover and characterize the mechanism of action of dsRNA binding inhibitors of VP35. This work has been a great personal expedition into Nuclear Magnetic Resonance experiments and working with unique small molecules. These results are unpublished though very much on the way toward publication. I worked on this project quite independently though I had some important sources of support along the way. First, Justin Miller and Louis Smith have been critical in honing my thoughts and scientific thought over the course of this project as well as being expert mentors in small molecule and drug characterization both experimentally and computationally. I also mentored an undergraduate, Rishi Samarth, who made important contributions to this work. This work was supported by my NIAID NIH F31AI157079, the MilliporeSigma fellowship from WashU, and a mini grant from the WUSTL Center for Drug Discovery. I also cannot neglect to thank and acknowledge Dr. Maxene Ilagan and Mike Prinsen for both of their work on the high throughput screening section of this work. Jian Gao also conducted the virtual screening briefly mentioned in the supplemental information though I performed the experiments.

3.2 Abstract

Protein-protein (PPI) and protein-nucleic acid interactions (PNI) are difficult to target classes of interactions that occur in many available protein structures. For many PPIs and PNIs of known therapeutic importance, few attempts have been made toward drug development. Cryptic pockets

are absent in available experimental structures but open due to protein fluctuations making them attractive targets for drug discovery. VP35 from the Zaire ebolavirus harbors a cryptic pocket which affects dsRNA binding but there are no known ligands that bind to the pocket. Filoviruses such as the ebolaviruses are negative sense, non-segmented, single strand RNA viruses that cause a highly lethal hemorrhagic fever. VP35's interactions with dsRNA inhibit activation of the host cells' immune response promoting viral replication promoting severe disease. Here, we report the use of high-throughput screening and nuclear magnetic resonance spectroscopy to discover and an allosteric inhibitor of VP35's dsRNA binding activity. Through combining experimental high throughput screening, spectroscopic approaches, and molecular docking we obtain high resolution insight into the molecular mechanism of this allosteric small molecule inhibitor and show it acts through a 2:1 mechanism to potently reduce dsRNA binding. These findings should enable further drug development targeting VP35's cryptic pocket and encourage additional drug discovery attempts targeting therapeutically important PPIs and PNIs.

3.3 Introduction

Protein-protein (PPI) and protein-nucleic acid interactions (PNI) are difficult to target classes of interactions that occur in many available protein structures.[1,2] PPIs and PNIs often occur through primarily flat interfaces making drug design and discovery difficult since these flat surfaces lack binding pockets. PPIs and PNIs are often important drug targets amplifying the importance of studying them and developing drug targeting strategies.[3]

For many PPIs and PNIs of known therapeutic importance, few attempts have been made toward drug development. The identification of protein hotspots that promote PPIs has been a somewhat successful approach for targeting PPIs with drugs.[4] Such hotspots do not have natural small molecule ligands making the identification of chemical matter with druglike potencies difficult.[4] Additionally, many PPIs and PNIs contain conserved residues at these interfaces making specific inhibitors difficult to design.[2] Yet some progress has been made through combining high throughput screening with high-resolution structural information to improve and elaborate on hit molecules.[2] Few approaches though consider protein dynamics in their discovery efforts which limits the available sites for molecular docking through ignoring potentially useful alternative conformations.

Cryptic pockets are absent in available experimental structures but open due to protein fluctuations. Cryptic pockets are exciting therapeutic opportunities for PPIs and PNIs and identifying such pockets has become easier with recent advances in machine learning algorithms and other approaches such as enhanced sampling molecular dynamics simulations.[5-13] After identification, knowing whether a given pocket affects protein function is not always obvious. To address this challenge, many tools we and others have developed can predict the functional impacts of a cryptic pocket to determine whether to target a cryptic pocket for drug

discovery.[8,10,14-22] Despite the availability of information about cryptic pockets in a variety of proteins, and tools to predict the utility of targeting them with drugs, there remains few ways to identify novel chemical matter predicted to target these pockets. One groundbreaking example of considering a protein's dynamics giving rise to a therapeutic is that of HIV integrase in which the discovery of a cryptic pocket and molecular docking produced novel chemical matter targeting this important protein.[23] Targeting cryptic pockets offers great opportunity for discovering therapeutics targeting PPIs and PNIs in a variety of diseases.

We previously described a cryptic pocket in VP35 and demonstrated that stabilizing the pocket in its open state with a drug fragment like modification, decreases dsRNA binding (Fig. 3.21A).[10] VP35's interactions with dsRNA inhibit activation of the host cells' immune response promoting viral replication.[30,31,33-39] Unchecked viral replication in early stages of infection is one of the causes of severe disease.[33,40,41] Previous structural and functional work on VP35 demonstrated its structure and the central role of dsRNA binding in immune antagonism (Fig. 3.21B).[33,39,42-44] It is also now shown that decreasing dsRNA binding by 3-5-fold is sufficient to reenable activation of the interferon response through enabling RIG-I activation supporting VP35 as an important drug target.[33,44,45]

Filoviruses such as the ebolaviruses are negative sense, non-segmented, single strand RNA viruses that cause a highly lethal hemorrhagic fever. In past and recent outbreaks of ebolaviruses case fatality rates approach 66% despite the use of FDA approved vaccines, and antibody therapies.[26,27] There remains no approved small-molecule therapies despite initial promising results.[28,29] Ebolaviruses cause infection even outside of large outbreaks and are potential zoonotic agents. All filoviruses' genomes encode viral protein 35 (VP35) which engages in

protein-nucleic acid and protein-protein interactions for immune antagonism and viral replication.[30-32]

Here, we report the use of high-throughput screening and nuclear magnetic resonance (NMR) spectroscopy to discover and learn the mechanism of dsRNA inhibition of small molecules that bind to Zaire ebolavirus VP35. Existing experimental efforts to discover small molecule inhibitors of dsRNA binding produced modest leads from natural sources.[24] Few docking-based studies exist for VP35 targeting dsRNA binding but have identified some potentially useful molecules[25]. To discover small molecule inhibitors of dsRNA binding that bind to the cryptic pocket we conducted a high throughput screen of a diverse chemical library. Then, we used nuclear magnetic resonance to experimentally map the interactions of our hit compounds with VP35. Finally, combining molecular docking and biochemical mutation data, we show that one hit molecule binds to the cryptic pocket to contribute to allosterically inhibit dsRNA binding.

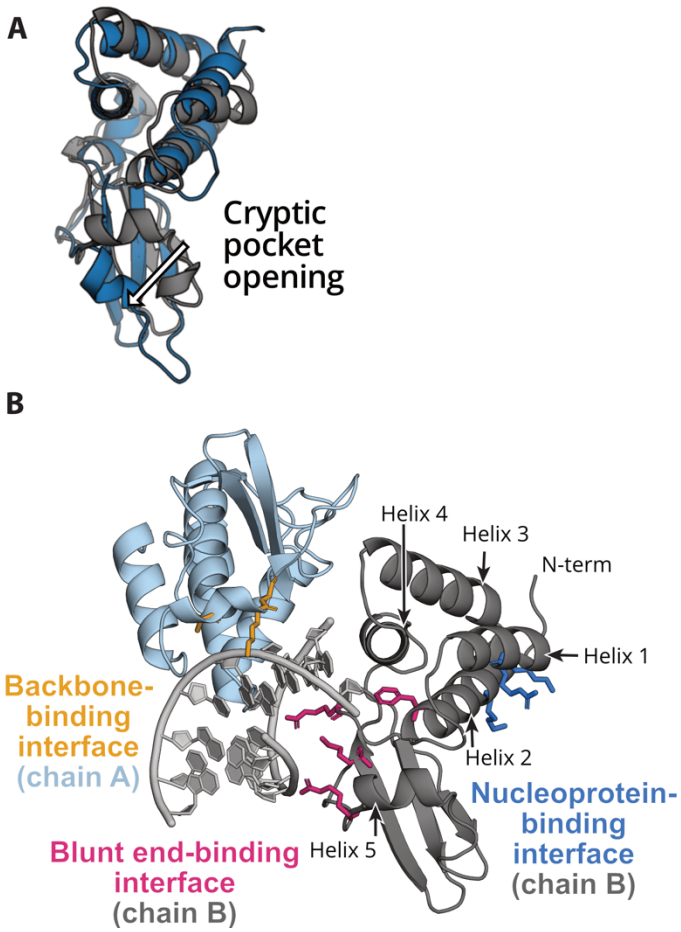

Figure 3.21 The VP35 cryptic pocket and VP35's interactions with dsRNA.

In A) VP35 forms a cryptic pocket as sampled from molecular dynamics simulations. This data appears in Chapter 2. The dark gray structure is the crystal-like structure and the structure in blue cartoon shows a representative conformation of the pocket open state. B) VP35 interacts with dsRNA through both the blunt ends (gray cartoon) and the backbone (light blue cartoon). Residues in magenta and goldenrod sticks, are required for dsRNA binding. Residues shown in blue sticks are required for binding to nucleoprotein. PDB ID 3L25.

Combining experimental high throughput screening, spectroscopic approaches, and molecular docking provides exquisite detail into the molecular mechanism of small molecule inhibitors. We identify several inhibitors and show that these compounds interact with the dsRNA binding interface in VP35. One compound additionally binds to the VP35 cryptic pocket demonstrating its druggability and opening up new options for drug development. Our data supports our predicted allosteric models in VP35 and shows that considering the protein ensemble in drug design and discovery can provide critical advances for therapeutics.

3.4 Results

3.4.1 A High-Throughput Screen Identifies Inhibitors Of Dsrna Binding.

We reasoned that a high throughput screen targeting VP35 would yield hits based on our recent discovery that VP35 harbors a cryptic pocket that is allosterically coupled to dsRNA binding. To perform a screen, we miniaturized our previously reported a fluorescence polarization-based assay for dsRNA binding into a high-throughput screen (Fig. 3.22). In high-throughput screening, the Z' score denotes the signal to noise ratio and a good assay that can distinguish true positives from false positives and false negatives has a score greater than 0.5 and less than 1. We found that over the course of the screening campaign, the assay maintained a high signal to noise ratio ($Z' > 0.5$) enabling identification of true positives from the data set. Similarly to the Z' score, when calculated from the experimental data rather than the positive and negative controls, the Z score denotes the change in signal from a specific molecule relative to the mean and standard deviations of all compounds' signal. Prior to screening we defined a hit as a molecule that changes FP by 3 standard deviations above the mean FP of all samples assayed, or a Z score

of 3. For our screen, a Z score of greater than or equal to 3 corresponds to ~20% inhibition and 126 compounds met this criterion for a hit rate of 0.88%. As a check on our hit list, we conducted a confirmation screen where we remeasured the FP of all 126 compounds, at three concentrations. From these results, we refined our hit list to any compounds that were active at the three assay concentrations. Our final hit list had 63 compounds and we find 47% of our preliminary hits were confirmed at 10 micromolar compound resulting in a 0.43% hit rate.

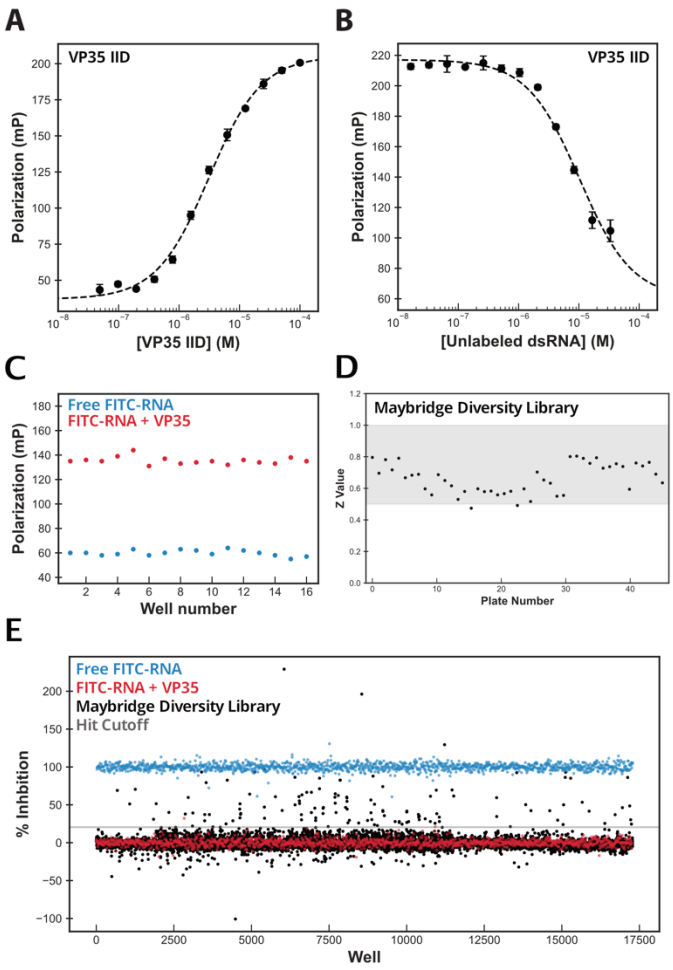

Figure 3.22 Fluorescence Polarization as a high throughput screen for dsRNA inhibitors.

While we previously reported using FP to measure dsRNA binding, here we show that this assay has a large dynamic range that is sensitive to competitive binding. A) VP35 titration curve with measuring polarization of 5' fluorescein labeled 25 base-pair RNA. B) Titration curve showing a decrease in fluorescently labeled dsRNA binding with increasing concentration of unlabeled dsRNA. C) VP35 and a 25 base-pair RNA have well-separated bound and unbound species when measuring FP. D) across a large screen of 14,400 compounds, the assay retains a robust signal to noise ratio. E) Data are showed as calculated percent inhibition for every data point in the screen. A majority of compounds tested have no effect on dsRNA binding (some black points) and overlap with the control values (red) while others inhibit over a range of potencies. The positive RNA-only controls are shown in blue. The cutoff used to designate a hit is the gray line across the plot.

To determine the potency of our newly identified inhibitors, we measured the EC_{50} for our compounds. For dose dependence experiments, we serially diluted each compound and added a fixed concentration of VP35 and 5FAM-dsRNA then measured fluorescence polarization of the labeled dsRNA (Fig. 3.23 and S3.26). Our dose dependence experiments demonstrate that multiple hits from our screen exhibit low micromolar EC_{50}s for VP35-dsRNA binding. Compounds MAC-02, MAC-26, MAC-41, and MAC-61 have EC_{50}s of 3.3 ± 1.5 μM, 1.05 ± 0.1 μM, 1.1 ± 0.04 μM, and 5.59 ± 0.70 μM, respectively. MAC-26 inhibits only 14.1 ± 0.6 percent of binding at maximal concentrations while MAC-02, MAC-41, and MAC-61 each inhibit 62.8 ± 1.7, 86.7 ± 1.7, and 79.8 ± 0.3 percent. Some hits only partially inhibit dsRNA binding while others nearly completely inhibit binding suggesting a variety of inhibitory mechanisms.

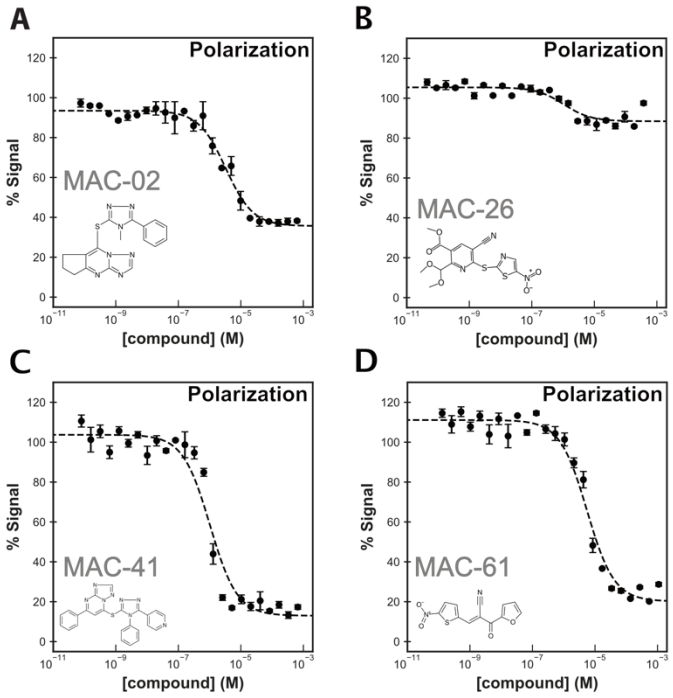

Figure 3.23 Hit molecules have a range of chemistries, potencies, and inhibitory effects.

To assess the strength of hit compounds inhibition, we conducted dose-dependence experiments. In the presence of a fixed concentration of VP35 and fluorescently labeled dsRNA, we varied the concentration of hit compound then measured FP. A-D) Data are reported as percent change relative to the controls with out compound present. Each data point represents the average of three measurements. A depiction of the molecule tested is shown with each plot.

To understand the importance of different chemical moieties in our hit list, we conducted hit list chemical substructure searches. We identified no maximum common substructure in our hit list suggesting there's not a certain required scaffold for inhibition of VP35-dsRNA binding. Next, we conducted substructure searches to understand the chemistries present in the hit list and whether the screening selectively enriched for specific moieties. Substructure searches revealed our screen selected for nitrile, and triazole moieties and for multiple scaffolds from the larger library including purines, pyrrolidines, and thiophenes. The variety of chemistries in our hit list suggests many different mechanisms of action.

3.4.2 NMR Chemical Shift Perturbation Experiments Show All Hits Impact The Dsrna Binding Interface.

We reasoned that understanding the mechanism of action for sufficiently strong hits i.e. $K_D < 10$ μM will enable optimization of our hits and clarify previous literature on VP35's dynamics and predicted druggable sites. Previous docking studies predicted identified dsRNA-binding inhibitors of VP35 bind at the RNA binding interface while other work suggested alternative ways to inhibit RNA binding such as targeting alternative conformations. To test the mechanism of inhibition for our ligands, we turned to chemical shift perturbation experiments which previously successfully identified ligand binding sites in VP35 (Fig. 3.24). First, we purified wildtype ^{15}N VP35 and collected an unliganded VP35 HSQC spectrum. Our unliganded spectrum matches previously published spectra and peak lists. We find the addition of up to 5% d6-DMSO induces no significant perturbations to the peaks.

To map the compounds' binding sites onto VP35, we collected the HSQC spectra for VP35 in the presence of hit compounds individually. Compounds were individually added to a sample of

VP35 at 80 μM with 10% D_2O, at a molar ratio according to their solubility (usually at least 2 moles compound to 1 mole of VP35). Then the same HSQC parameters were used to collect the spectrum of VP35 in the presence of the compound. From HSQCs in the presence of the compounds, we calculated chemical shift perturbations using the apo VP35 peak assignments (Fig. 3.24).

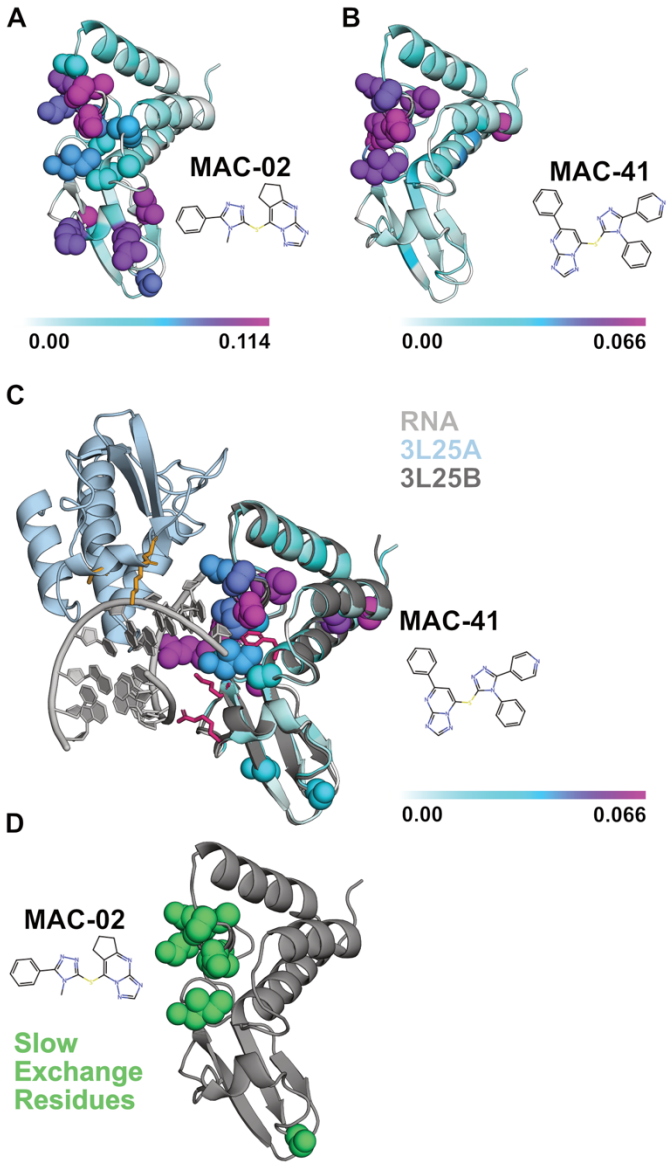

Figure 3.24 Hit compounds bind directly to VP35's blunt-end interface.

Nuclear Magnetic Resonance chemical shift perturbation experiments show the interacting residues on VP35 with a given compound. Two example molecules' results are shown: A, D) MAC-02 and B,C) MAC-41. Here, perturbations are calculated as described in the methods section. Then mapped onto the VP35 crystal structure and colored on a gradient from white to cyan to magenta according to the magnitude of the perturbation (see color bars). The maximum value for magenta is set to 3X greater than the average of all CSPs for that given compound and that value is shown on the colorbar. Compound names and structures are shown next to the data. MAC-02 shows

scattered difficult to interpret perturbations while MAC-41 has tightly clustered and strong perturbations. C) The MAC-41 perturbations are shown in spheres overlaid onto the 3L25 co-crystal of VP35 with dsRNA highlighting the overlap between the blunt-end binding interface and the compound binding site. D) NMR titration experiments with MAC-02 reveal a cluster of slow-exchange residues at the dsRNA interface. The slow exchange residues are highlighted in green spheres with the same orientation as other structures in panels A-C.

NMR perturbation experiments show that many hits induce perturbations at the dsRNA binding interface. Previous docking studies drug targeting VP35 predicted modest binding energies for natural products and ZINC database compounds to the dsRNA binding interface suggesting the interface as a possible binding site for small molecules.[24,25] Many of the compounds in this study map to the same region containing basic residues predicted in previous studies to bind small molecules and required to dsRNA such as R312, R322, and K339. Other residues thought to be important for dsRNA competitive inhibition are K309, F239, A238, P233, T237, P273, Q274, S272, C275, Q279, I278, and I340. MAC-02 (EC_{50} 3.3 ± 1.5 μM) induces strong perturbations (>0.05) at L232, G234, A238, D271, S272, Q274, A276, Q279, T281, K282, A290, A291, I295, S299, R305, K309, G323, V325, G337, and I340. MAC-41 (EC_{50} 1.1 ± 0.04 μM) induces strong perturbations at A221, S272, Q274, I278, Q279, I280, T281, D321, G323, and I340. Residues Q274, Q279, and I340 are important for dsRNA binding suggesting a directly competitive inhibitory mechanism. Each of these bind with EC_{50} in the micromolar range but have multiple scaffolds for downstream optimization.

3.4.3 Slow Exchanging Residues In NMR Titrations Of VP35 With MAC-02 Demonstrate A High Affinity Binding Site.

Chemical shift perturbations for one compound, MAC-02, suggested a more complex mechanism than competitive binding to the dsRNA interface. The hit compound MAC-02 induced widespread difficult to interpret perturbations. Residues at the RNA binding interface, as well as near the previously predicted cryptic pocket are both strongly perturbed. Perturbations at the RNA binding interface may arise from allosteric effects since the cryptic pocket is coupled to the RNA binding interface. Interestingly, the converse is also true that the RNA binding interface can induce changes at the cryptic pocket. It is possible MAC-02 binds to the cryptic pocket

inducing chemical shift perturbations at the RNA-binding interface or that binding of MAC-02 to the interface, induces chemical shift perturbations in the cryptic pocket. Perturbations that do not map well to a distinct site cannot unambiguously indicate a conformational change or higher than one-to-one stoichiometry but may suggest either or both mechanisms.

To offer clarity to MAC-02 binding site, we conducted a chemical shift perturbation titration experiment (Fig. 3.24D). If a ligand binds with high affinity, the exchange between the bound and the unbound states is significantly slowed. This is called the slow-exchange regime and is characterized by a peak decreasing in intensity at a given chemical shift, increasing in intensity at a different chemical shift as the concentration of ligand increases. In the slow exchange regime k_{off} is much slower than the difference in frequency between the bound and free states, which suggests a low micromolar K_d. Given that MAC-02 has a low micromolar EC_{50} we reasoned that some observed chemical shift perturbations may be in the slow exchange regime. Then, a titration allows us to observe the exchange regime into which a given residue's perturbations fall which may suggest whether a set of perturbations comprise a primary binding site.

Slow exchanging residues cluster on helix 4 near the RNA binding site. In the titration experiment, we add increasing amounts of MAC-02 to a sample of VP35 and collect the HSQC spectrum at each concentration. We varied the concentration from substoichiometric to stoichiometric ratios (12:1 VP35:MAC-02 up to 1:4). Unambiguously, in VP35 residues Q274, A276, Q279, K281, and I340 that undergo perturbations in the slow exchange regime, cluster on helix 4 near the RNA binding site. Other residues undergo strong perturbations in the fast and intermediate exchange regimes such as I295, R305, Q308, K309, and V325, are spatially distinct from the slow-exchanging residues. Since strong perturbations can indicate changes due to allosteric effects, or higher than one stoichiometry, identification of a principal binding site does

not rule out a secondary binding site. So chemical shift perturbation titrations reaffirm that MAC-02's mechanism of action is more complex than one to one binding.

3.4.4 MAC-02 Binds Both The RNA-Binding Interface, And The VP35 Cryptic Pocket With An Overall 2:1 Stoichiometry To VP35

Amino acid substitutions enable probing of specific VP35 structural regions to identify MAC-02's binding site. To clarify the NMR binding site data, we hypothesized that alanine substitutions to residues that are in regions that the NMR data suggests interact with MAC-02 would disrupt the ligand-protein interaction reducing MAC-02's inhibitory effect and verify MAC-02's binding site. To test this hypothesis, we generated two variants, F239A and C326A VP35 and tested their dose dependent responses in comparison to wild-type VP35 (Fig. 3.25). I340 is a highly perturbed residue upon compound binding but is recalcitrant to substitution but is near F239, a residue of known importance for dsRNA binding. Interestingly I340 is hypothesized to be important for blunt-end binding, more than F239 so mutating F239 is a good way to disrupt binding to the RNA-binding interface rather than directly substituting I340. While C326 is a buried residue, our previous work demonstrated that it slowly exposes to solvent when the VP35 cryptic pocket opens suggesting it could interact with a bound ligand if the pocket was open. MAC-02 binding modestly perturbs V235 along with other residues whose backbone amides point into the VP35 cryptic pocket suggesting their environment changes due to MAC-02 binding. Given the localization of chemical shift perturbations to the cryptic pocket, and its correlation to pocket opening, we chose C326 for substitution.

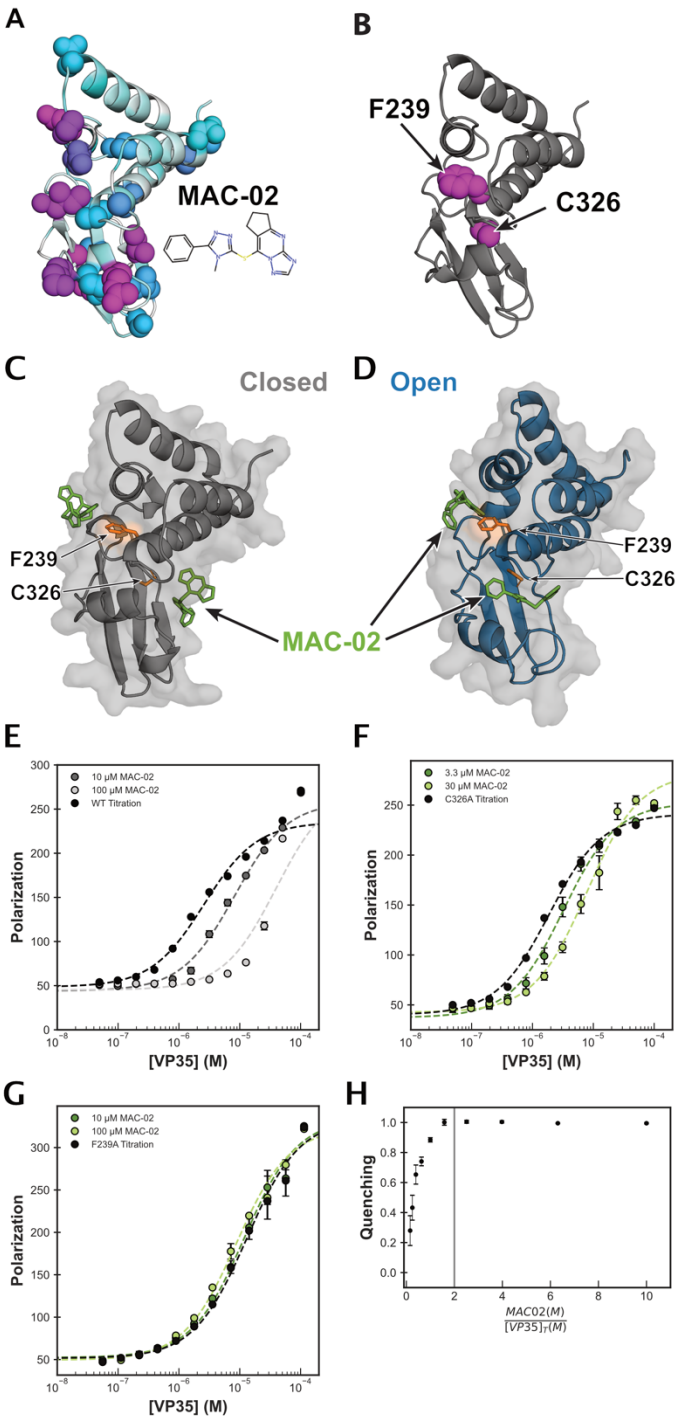

Figure 3.25 Docking, Mutations, and stoichiometry experiments suggest MAC-02 binds two sites on VP35.

MAC-02's mechanism of action was unclear from A) NMR chemical shift perturbation experiments. B) Structure highlighting the location of F239 and C326 in spheres for comparison to the patterns of perturbation in A). C,D) Results from molecular docking where MAC-02 (green) was docked against the cryptic pocket and the dsRNA binding interface to the crystal-like pocket closed state C) and the cryptic pocket open state D) from MD simulation. F239 and C326 are highlighted in orange sticks and surface. VP35 can accommodate two bound molecules of MAC-

02. E, F, G) MAC-02-based dsRNA inhibition is reduced upon mutation of C326 or F239 to alanine suggesting MAC-02 interacts with both sites: the cryptic pocket and the dsRNA binding interface. H) Intrinsic Tryptophan Fluorescence Quenching with increasing concentrations of MAC-02 shows that a molar ratio of at least 2 MAC-02 to 1 VP35 is required to saturate binding suggesting MAC-02's inhibition acts through both sites on VP35.

Mutating F239 and the distant C326 both disrupt MAC-02 mediated inhibition of dsRNA binding (Fig. 3.25 E, F and Fig. S3.28-31, S3.34). To test the effect of mutating F239 or C326 on MAC-02 inhibition, we performed dsRNA binding experiments at multiple fixed, but increasing concentrations of MAC-02. Fitting the dsRNA binding titration curve to a generalized single site binding model, in the absence of MAC-02, produces a K_d of wild-type for 25 base-pair RNA of 3.6 μM (Fig. S3.27) as previously reported. Then, after repeating the titration with 3.3 and 30 μM MAC-02, we observe a ten-fold decrease in apparent binding affinity for the 25 base-pair dsRNA. After repeating this experiment with the variant C326A, we observe that C326A has significantly reduced response to the addition of MAC-02 with a four-fold reduction in apparent binding affinity for dsRNA at 30 μM MAC-02. More dramatically, when we perform this experiment with F239A, we find this variant is not at all sensitive to MAC-02 activity, even at higher overall concentrations of MAC-02 up to 100 μM MAC-02 (>30 fold the wild-type EC_{50}). While F239A seems to eliminate all MAC-02 mediated inhibition, the C326A data suggests an additional binding site that may be useful for further inhibitor development. These data from C326A and F239A suggest that MAC-02 acts mostly through direct interactions with the dsRNA binding interface (F239) but with a significant contribution from interactions with C326 and surrounding residues. In combination with the chemical shift perturbation data, alanine substituted variants support a greater-than-stoichiometry mechanism for MAC-02 and suggest the possibility of an indirect mechanism for dsRNA inhibition, compatible with previously described conformational dynamics.

To identify MAC-02's mechanism of action, we measured the stoichiometry of binding to VP35 in the absence of dsRNA via intrinsic tryptophan fluorescence quenching (Fig. 3.25H). Given that substituting F239 and C326, two distant residues, decreases inhibition by MAC-02, and

MAC-02 is not large enough for one molecule to induce all observed chemical shift perturbations or contact both residues, we hypothesized that more than one molecule of MAC-02 binds to VP35 to inhibit dsRNA binding. Through binding to both the interface and the cryptic pocket, MAC-02 would achieve greater inhibition than binding to the pocket alone but this 2:1 stoichiometry remains hypothetical. An alternative hypothesis is that MAC-02 is forming micelles or aggregates and affecting VP35 chemical shifts but over concentrations tested, MAC-02 is soluble and unlikely aggregated (Fig. S3.35). VP35 harbors one tryptophan residue and the addition of MAC-02 under stoichiometric results in a characteristic decrease in intrinsic tryptophan fluorescence. Using intrinsic tryptophan fluoresce, we measured the binding of VP35 to MAC-02 in the stoichiometric concentration regime. We varied the concentration of MAC-02 from a molar ratio of 0.1 to 8 in the presence of saturating concentrations of VP35 and measured protein fluorescence using a microplate reader. Under these stoichiometric conditions, intrinsic tryptophan fluorescence quenching saturates at a molar ratio of compound to VP35 of greater than 2 (Fig. 3.26). At stoichiometries up to 8:1 MAC-02 to VP35, we observe no further binding demonstrating that the stoichiometry is not greater than 2:1. These data in combination with the NMR experiments, support a model whereby MAC-02 binds to the cryptic pocket near C326 and a second molecule of MAC-02 can independently bind near the blunt-end binding interface forming interactions with F239.

To test the geometric constraints of this 2:1 model, we docked MAC-02 to conformational states obtained from molecular dynamics as reported in our previous study (Fig. 3.26 C,D). We used Vina to dock MAC-02 against the protein-RNA interface, and in the cryptic pocket for each state in our simulation dataset. Next, we manual inspected the poses for MAC-02 in the case where the pocket is closed similarly to the crystal structure (apo or bound to dsRNA) and where it is

open (see methods). We identify that when docked to the RNA binding interface, Vina predicts, MAC-02 makes a cation-pi stacking interaction with F239 similarly to the terminal nucleotide in dsRNA (PDB: 3l26). When docked into the cryptic pocket, MAC-02 is well accommodated without disrupting the F239-MAC-02 interaction. The docking data shows that MAC-02 is of the correct size and chemistry to bind to both the RNA binding interface and cryptic pocket. The mutations to F239 at the RNA binding interface, and C326 in the pocket support these poses since substituting these residues with Alanine results in a decrease in MAC-02's inhibitory effect. From our quenching data, we can say the stoichiometry is certainly no more than two molecules of MAC-02 per VP35 molecule. We also compared our docking results to our observed CSPs and find that docked poses of MAC-02 well match the observed CSPs. Combining the data from NMR experiments, alanine substitutions, stoichiometry measurements, and molecular docking, our data supports a model of MAC-02 binding to VP35 at a 2:1 ratio and allosterically inhibiting RNA binding. MAC-02 confirms our previous prediction that the VP35 cryptic pocket is a druggable allosteric site and MAC-02 can now be leveraged to target ebolavirus infections or as a pharmacological tool to probe the effect of disrupting conformational dynamics in more complex contexts.

3.5 Discussion

We demonstrate the discovery of small-molecule inhibitors of the ebolavirus VP35-RNA interaction from a large diverse library. In our past work we identified a cryptic pocket in VP35 and predicted it was a druggable site based on existing structural and functional data.[10,31,33,39,42,45] But we had not yet identified new chemical matter binding to VP35 or its cryptic pocket until now. We attempted virtual screening to identify new matter, with little success (Fig. S3.33). To achieve this, we conducted a routine high throughput screen by fluorescence polarization. Our

screen was sensitive, with high signal to noise ratio, and it identified ~60 hits for follow up studies. Yet simply inhibiting dsRNA binding does not require a compound bind to the VP35 cryptic pocket. To determine the binding site for hit molecules, we conducted NMR chemical shift perturbation experiments. Our NMR experiments demonstrated that MAC-02 and MAC-41, two hits from our screen, bind to the VP35-dsRNA binding interface. Yet the results for MAC-02 also suggested that it was interacting with VP35 at two different binding sites—the dsRNA binding interface, and the cryptic pockets. To test this, we introduced mutations to each site and observed that the mutations F239A, and C326A reduced MAC-02 based inhibition. To further test if VP35 could accommodate binding of more than one molecule of MAC-02 we computationally docked MAC-02 to the RNA binding interface and to the cryptic pocket. We find that MAC-02 has a favorable binding energy to the pocket and that VP35 is large enough to accommodate two distinct binding modes for MAC-02. Finally, we measured the amount of MAC-02 required to saturate VP35 binding and find that two molecules of MAC-02 is sufficient to saturate VP35. These data support a model whereby MAC-02 has a high affinity binding site for the dsRNA binding interface and a lower affinity binding site in the cryptic pocket. MAC-02 is a rule of three molecule with favorable properties for cell permeability and drug development. Through attempting to target the cryptic pocket, we identified novel chemical matter which will be important for future studies of the druggability of the VP35 cryptic pocket in the disease context i.e. cellular studies. Medicinal chemistry approaches to optimize MAC-02 may also yield a higher-affinity inhibitor for development into a therapeutic to combat ebolavirus infections. These results demonstrate the importance of considering a protein's conformational dynamics in the context of drug discovery especially for hard-to-drug targets.

3.6 Methods

3.6.1 Protein Expression And Purification

All variants of VP35's IID were purified from the cytoplasm of E. coli BL21(DE3) Gold cells (Agilent Technologies).[35,46] Variants were generated using the site directed mutagenesis method and confirmed by DNA sequencing. Transformed cells were grown at 37°C until OD 0.3 then grown at 18°C until induction at OD 0.6 with 1 mM IPTG (Gold Biotechnology, Olivette, MO). Cells were grown for 15 hours then centrifuged after which the pellet was resuspended in 20 mM Sodium Phosphate pH 8, 1 M sodium chloride, with 5.1 mM β-mercaptoethanol. Resuspended cells were subjected to sonication at 4°C followed by centrifugation. The supernatant was then subjected to Ni-NTA affinity (BioRad Bio-Scale Mini Nuvia IMAC column), TEV (ProTEV Plus, Promega, Madison, WI) digestion, cation exchange (BioRad UNOsphere Rapid S column), and size exclusion chromatography (Cytiva HiLoad 16/600 Superdex 75) into 10 mM Hepes pH 7, 150 mM NaCl, 1 mM MgCl2, 2 mM TCEP.

[^{15}N]-VP35 expression and purification

Overnight seed cultures of E. coli BL21(DE3) Gold transformed with the same VP35 gene containing plasmid as above, were grown in LB Miller media (Thermo Fisher). M9 media was prepared as followed: ~990 mL of ddH_2O or MilliQ H_2O was added to baffled 2.5 L Erlenmeyer flasks. To the flasks was also added 0.3 mM $CaCl_2$, 1 mM $MgSO_4$, 1 g/L [^{15}N]-NH_4Cl, Trace Metals (13 mg/L $CuCl_2$,830 mg/L $FeCl_3(6H_2O)$, 10 mg/L $CoCl_2(6H_2O)$,10 mg/L HBO_3, 1.6 mg/L $MnCl_2(6H_2O)$, 84 mg/L $ZnCl_2$), and 4 g/L Glucose, 100 mg/mL ampicillin. The media was then shaken and incubated at 37C until all components were dissolved. Each flask was then inoculated with 5 mL of high-density overnight culture. The OD_{600} was monitored, and

expression was induced at greater than OD_{600} 0.6 with 1 mM IPTG as above. The remainder of the expression and purification is the same as above excepting the addition of a second purification via size exclusion column.

3.6.2 FP Assay

We adapted our previously described[10] fluorescence polarization-based assay for high-throughput screening, and dose dependence experiments as described below.

High throughput screening:

VP35 purified as above, was diluted to 9 μM in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$, and 0.01% Triton-X. Two individual expression batches were pooled before dilution for screening. 25 base-pair FITC-dsRNA (Integrated DNA Technologies) was diluted to 300 nM in the same buffer as VP35 above. Sense strand sequence: 56-FAM-rArArArCrUrGrArArArGrGrGrArGrArArGrUrGrArArArGrUrG, antisense strand sequence rCrArCrUrUrUrCrArCrUrUrCrUrCrCrCrUrUrUrCrArGrUrUrU. 200 nL of 1 mM compounds or DMSO from the Maybridge Hit Finder v13 were spotted into 384 well costar black plates (ID 3575) with the Digilab Hummingbird XL 96 robotics system. Then the Multidrop Combi (Thermo Fisher) equipped with small tube dispensing cassettes was used to sequentially add first 10 μL of VP35 then 10 μL dsRNA to the plates such that each plate had two negative and two positive control columns (dsRNA+VP35, and dsRNA only respectively) at 1% final DMSO in addition to the 20 columns of sample. After 1 hour of incubation fluorescence polarization was read on the BioTek Synergy2 Multi-Mode Reader. Raw data were used to calculate Z' value as shown below, and percent inhibition per compound, based on dsRNA fluorescence polarization, by plate.

$$Z' = 1 - \frac{3 \cdot (\sigma_+ - \sigma_-)}{|\mu_+ - \mu_-|}$$ (Equation 6)

$$\% \, inhibition = \left[1 - \frac{(\mathrm{mp_i} - \mu_+)}{(\mu_- - \mu_+)}\right] \cdot 100$$ (Equation 7)

Here μ_+, μ_- correspond to the mean value of the negative and positive control values, respectively. Similarly, σ_+, σ_- correspond to the standard deviation of the negative and positive controls. For calculating percent inhibition, $\mathrm{mp_i}$ corresponds to the observed fluorescence polarization of the *i* compound sample. On average, the assay had a Z' factor of 0.75 ± 0.05 demonstrating good signal to noise ratio per plate. All positive and negative controls were analyzed using GraphPad Prism for outliers based on plate-normalized values. In total, 6 negative and 6 positive control values were removed.

A Z-score as defined below, of greater than or equal to 3 ~20.5% inhibition, was used to classify if a compound was a hit. From the screen, 126 compounds met this criterion resulting in an initial hit rate of 0.88%. These 126 compounds were then subjected to the confirmation screen.

Similarly to the high-throughput screen, DMSO or compounds were spotted in 50, 100, or 600 nL volumes at 1 mM concentration resulting in final concentrations of 2.5, 10, or 30 µM compound. Protein and RNA reagents were dispensed as above, plates were incubated and read the same as above. The resulting data was processed the same as described above resulting in a Z' of 0.78 ± 0.03 and one positive control was removed.

Compound dose-response experiments:

For determining a given compound's EC_{50}, powdered compound was obtained from Molport at least 95% purity and resuspended in d_6-DMSO to usually 10 mg/mL. The concentrated compound was then serially diluted in DMSO in 8-strip PCR tubes, then buffer (Hepes, NaCl,

$MgCl_2$, Triton-X as above) was added and DMSO was at 3% (v/v) final volume and a multichannel pipettor was used to transfer 25 µL of compound onto a 96 well plate, in triplicate. To each well was then added 9 µM VP35, and 300 nM FITC-dsRNA to final concentrations of 3 µM and 100 nM. For every experiment we included the following control wells: buffer only, dsRNA only, dsRNA with VP35 at the expected KD, and dsRNA with VP35, dsRNA with MAC-02, VP35 binding titration in DMSO, without compound, and all controls were at 1% DMSO final. The plate was light protected, and fluorescence polarization was read on either a BioTek Synergy2 Multi-Mode Reader, PerkinElmer EnVision Xcite, or TECAN Spark microplate reader after one hour. Data are shown as percent polarization as calculated from the RNA, Protein DMSO control well. Fits are to a general single site binding model as described below[47]:

$$p_{obs} = p_0 + (p_{max} - p_0)\left(\frac{EC_{50}\cdot[Compound]}{1+EC_{50}\cdot[Compound]}\right) \text{ (Equation 8)}$$

VP35-dsRNA binding assays with MAC-02:

We buffer exchanged and concentrated to ~300 µM VP35 iid variants then serially diluted by half across the rows of Costar 96 well half area black flat bottom polystyrene plates with a non-binding surface, in triplicate, in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM MgCl2, 0.01% Triton-X. MAC-02 was resuspended in d_6-DMSO at 10mg/mL and diluted to 3X assay concentration (3% protonated DMSO) in the same Hepes, NaCl, $MgCl_2$ and Triton-X buffer as above before added to each well containing VP35 dilutions. Then we added dsRNA to each well at a final concentration of 100 nM 25 bp FAM-dsRNA. For every experiment we included the following control wells: buffer only, dsRNA only, dsRNA with VP35 at the expected KD, and dsRNA with VP35, dsRNA with MAC-02, and all controls were at 1% DMSO final. The plates

were protected from light and allowed to equilibrate for at least one hour before reading as described above. Data were routinely checked for fluorescence anomalies in the control and experimental wells. The sample plate was equilibrated for one hour before data collection. Data were collected on a BioTek Synergy2 Multi-Mode Reader as polarization. Data were analyzed in Jupyter Notebook using Scipy 1.3.2, NumPy 1.14.x and 1.19.5, Matplotlib 3.5, Pandas 0.25.3, and Seaborn 0.11.2. A single-site binding model was sufficient to fit the data similarly as above:

$$p_{obs} = p_0 + (p_{max} - p_0)(\frac{K_A \cdot [VP35]}{1 + K_A \cdot [VP35]})$$ (Equation 9)

3.6.3 Chemical Informatics

The Maybridge Hit Finder v13 is available from Thermo Fisher as a structure data file (sdf) enabling quick queries using freely available chemical informatics software. The starting library and hit list were subjected to chemical substructure searches in Jupyter Notebook using RDKit, Scipy 1.3.2, NumPy 1.14.x and 1.19.5, Matplotlib 3.5, Pandas 0.25.3, and Seaborn 0.11.2.

3.6.4 NMR Chemical Shift Perturbation Experiments

Sample preparation

We measured the concentration of [^{15}N]-VP35 after purification per the Edelhoch method and prepared samples at 80 μM VP35 (10 mM Hepes pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 2 mM TCEP), 10% D_2O, and either 5% d_6-DMSO or 5% compound resuspended in 100% d_6-DMSO. Samples were prepared at ~340 μL and loaded into 5 mm shigemi tubes (Millipore Sigma BMS-005B). Before loading, samples were gently vortexed and centrifuged at 14.8 krpm in a benchtop centrifuge for 10 minutes. VP35 is robust to high concentrations of DMSO (Fig. S3.32).

Data collection

Samples were gauged and loaded into a 600.13 MHz Bruker Avance spectrometer affixed with a cryoprobe and temperature set to 25°C, corrected. Samples were manually tuned, then subjected to 3D shimming using Z6 with shigemi parameters. A 90° pulse was used to calibrate the ^{1}H pulse. All data was collected in Bruker Topspin. 1H–15N HSQC spectra were collected for every sample shown in the figures with 16 dummy scans, 8 scans, 128 complex t1 points, 2048 total t2 points. We used the HSQCETFPF3GPSI pulse sequence from Bruker.

Data analysis

All free induction decay data were transformed and phased using TopSpin. All spectra's peaks were picked, assigned, and tracked across titrations or spot-check experiments using Poky.[48] Peak lists were processed and CSPs were calculated using in Jupyter Notebook using Scipy 1.3.2, NumPy 1.14.x and 1.19.5, Matplotlib 3.5, Pandas 0.25.3, and Seaborn 0.11.2. per the equation below and following published protocols.[49]

$$d = \sqrt[2]{\frac{1}{2}[\delta_H^2 + (\alpha \cdot \delta_N)^2]} \text{ (Equation 10)}$$

We show calculated CSPs using alpha = 0.14. After calculation, CSPs were mapped onto VP35 structure by encoding the CSP per residue as a B-factor where all atoms in a given residue are colored according to the calculated perturbation. Only significant perturbations > 0.05 ppm (>5X) the ^{1}H resolution are shown in spheres and considered for binding site mapping. The B-factor coloring ranges from 0.0 to 3X the standard deviation of all peaks' perturbation along the white, cyan, and magenta color spectrum.

3.6.5 Docking Approach

To better understand the binding of MAC-02 to VP35, MAC-02 was docked to every state in our MSM (previously reported[10]) using Vina.[50,51] The output predicted binding energy for a given state was weighted by that state's equilibrium probability and used to calculate an overall free energy of binding as previously described[52]. The docking box was centered on 57.863, 59.645, 7.277 and 58, 41, 5 corresponding to docking to the RNA endcap interface, or the cryptic pocket (see Appendix). Poses for all states were mass loaded into PyMOL with a representative pocket open state shown. We then visually inspected the MAC-02 poses for given open states to identify poses with the compound inside the predicted pocket.

3.6.6 Intrinsic Tryptophan Fluorescence Quenching

To identify the stoichiometric ratio needed to saturate VP35 with MAC-02 we used VP35's intrinsic tryptophan fluorescence. VP35 was purified as described above and diluted in 10 mM Hepes pH 7, 150 mM NaCl, 1 mM MgCl2, 2 mM TCEP to 40 μM. MAC-02 was serially diluted by half in 8-strip PCR tubes then transferred onto a microplate (Costar black, half-area 96-well plate). VP35 was added to MAC-02 dilutions and incubated to equilibrium for one hour at room temperature (21-23°C). Controls were included of VP35 alone, VP35 with a saturating concentration of MAC-02, and MAC-02 alone at every concentration measured. Then, in a Tecan SPARK, fluorescence was measured with the temperature controlled at 25°C. Each sample was excited at 280 nm with a 10 nm bandwidth. Then, emission was detected as spectrum from 322 nm and 20 nm bandwidth. Gain was set automatically based on the intensity of the

unliganded VP35. Read height was set to 20 mm. Data were processed from raw fluorescence to quenching as described.[53] Briefly described,

$$Q_{obs} = \frac{F_0 - F_{obs}}{F_0} \text{ (Equation 11)}$$

$$\boldsymbol{Q_{max} = \frac{F_0 - F_{min}}{F_0}} \text{ (Equation 12)}$$

Stoichiometry of binding was then estimated from the molar ratio of MAC-02 to VP35 total that produced saturated binding.

3.7 References

1 Arkin, M. R. & Wells, J. A. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. Nature Reviews Drug Discovery 3, 301-317 (2004). https://doi.org:10.1038/nrd1343

2 Johnson, D. K. & Karanicolas, J. Computational Screening and Design for Compounds that Disrupt Protein-protein Interactions. Current Topics in Medicinal Chemistry 17, 2703-2714 (2017). https://doi.org:10.2174/1568026617666170508153904

3 Hopkins, A. L. & Groom, C. R. The druggable genome. Nature Reviews Drug Discovery 1, 727-730 (2002). https://doi.org:10.1038/nrd892

4 Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. Nature 450, 1001-1009 (2007). https://doi.org:10.1038/nature06526

5 Meller, A. et al. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. Nature Communications 14, 1177 (2023). https://doi.org:10.1038/s41467-023-36699-3

6 Artur, M., Soumendranath, B., Shahlo, S. & Gregory, R. B. Accelerating cryptic pocket discovery using AlphaFold. bioRxiv, 2022.2011.2023.517577 (2022). https://doi.org:10.1101/2022.11.23.517577

7 Lu, S. et al. Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. Nature Communications 12, 4721 (2021). https://doi.org:10.1038/s41467-021-25020-9

8 Porter, J. R. et al. Cooperative changes in solvent exposure identify cryptic pockets, conformational switches, and allosteric coupling. bioRxiv, 323568 (2018). https://doi.org:10.1101/323568

9 Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. Current Opinion in Chemical Biology 44, 1-8 (2018). https://doi.org:10.1016/j.cbpa.2018.05.003

10 Cruz, M. A. et al. A cryptic pocket in Ebola VP35 allosterically controls RNA binding. Nat Commun 13, 2269 (2022). https://doi.org:10.1038/s41467-022-29927-9

11 Morando, M. A. et al. Conformational Selection and Induced Fit Mechanisms in the Binding of an Anticancer Drug to the c-Src Kinase. Scientific Reports 6, 98-99 (2016). https://doi.org:10.1038/srep24439

12 Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J. & Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. Accounts of Chemical Research, acs.accounts.9b00613 (2020). https://doi.org:10.1021/acs.accounts.9b00613

13 Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. Journal of the American Chemical Society 138, 14257-14263 (2016). https://doi.org:10.1021/jacs.6b05425

14 Knoverek, C. R., Amarasinghe, G. K. & Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. Trends in Biochemical Sciences 44, 351-364 (2019). https://doi.org:10.1016/j.tibs.2018.11.007

15 Ward, M. D. et al. Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. Nature Communications 12, 3023 (2021). https://doi.org:10.1038/s41467-021-23246-1

16 Hart, K. M. et al. Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. 12, e0178678 https://doi.org:10.1371/journal.pone.0178678

17 Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. Proceedings of the National Academy of Sciences 112, 2734-2739 (2015). https://doi.org:10.1073/pnas.1417811112

18 Hart, K. M., Ho, C. M. W., Dutta, S., Gross, M. L. & Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. Nature Communications 7, 12965 (2016). https://doi.org:10.1038/ncomms12965

19 Singh, S. & Bowman, G. R. Quantifying Allosteric Communication via Both Concerted Structural Changes and Conformational Disorder with CARDS. Journal of Chemical Theory and Computation 13, 1509-1517 (2017). https://doi.org:10.1021/acs.jctc.6b01181

20 Zimmerman, M. I. et al. SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. Nature Chemistry 13, 651-659 (2021). https://doi.org:10.1038/s41557-021-00707-0

21 Brenke, R. et al. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics 25, 621-627 (2009). https://doi.org:10.1093/bioinformatics/btp036

22 Kozakov, D. et al. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nature Protocols 10, 733-755 (2015). https://doi.org:10.1038/nprot.2015.043

23 Schames, J. R. et al. Discovery of a Novel Binding Trench in HIV Integrase. ACS Publications 47, 1879-1881 (2004). https://doi.org:10.1021/jm0341913

24 Daino, G. L. et al. Identification of Myricetin as an Ebola Virus VP35-Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. Biochemistry 57, 6367-6378 (2018). https://doi.org:10.1021/acs.biochem.8b00892

25 Glanzer, J. G. et al. In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. Bioorg Med Chem 24, 5388-5392 (2016). https://doi.org:10.1016/j.bmc.2016.08.065

26 Cross, R. W., Mire, C. E., Feldmann, H. & Geisbert, T. W. Post-exposure treatments for Ebola and Marburg virus infections. Nature Reviews Drug Discovery 17, 413-434 (2018). https://doi.org:10.1038/nrd.2017.251

27 Keshwara, R., Johnson, R. F. & Schnell, M. J. Toward an Effective Ebola Virus Vaccine. Annual Review of Medicine 68, 371-386 (2017). https://doi.org:10.1146/annurev-med-051215-030919

28 Warren, T. K. et al. Therapeutic efficacy of the small molecule GS-5734 against Ebola virus in rhesus monkeys. Nature 531, 381-385 (2016). https://doi.org:10.1038/nature17180

29 Mulangu, S. et al. A Randomized, Controlled Trial of Ebola Virus Disease Therapeutics. New England Journal of Medicine 381, 2293-2303 (2019). https://doi.org:10.1056/NEJMoa1910993

30 Messaoudi, I., Amarasinghe, G. K. & Basler, C. F. Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus. Nature Reviews Microbiology 13, 663-676 (2015). https://doi.org:10.1038/nrmicro3524

31 Edwards, M. R. et al. Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. Cell Rep 14, 1632-1640 (2016). https://doi.org:10.1016/j.celrep.2016.01.049

32 Woolsey, C. et al. A VP35 Mutant Ebola Virus Lacks Virulence but Can Elicit Protective Immunity to Wild-Type Virus Challenge. Cell Rep 28, 3032-3046 e3036 (2019). https://doi.org:10.1016/j.celrep.2019.08.047

33 Prins, K. C. et al. Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. J Virol 84, 10581-10591 (2010). https://doi.org:10.1128/JVI.00925-10

34 Zinzula, L., Esposito, F., Pala, D. & Tramontano, E. dsRNA binding characterization of full length recombinant wild type and mutants Zaire ebolavirus VP35. Antiviral Res 93, 354-363 (2012). https://doi.org:10.1016/j.antiviral.2012.01.005

35 Leung, D. W. et al. Expression, purification, crystallization and preliminary X-ray studies of the Ebola VP35 interferon inhibitory domain. Acta Crystallogr Sect F Struct Biol Cryst Commun 65, 163-165 (2009). https://doi.org:10.1107/S1744309108044187

36 Brown, C. S. et al. In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. J Mol Biol 426, 2045-2058 (2014). https://doi.org:10.1016/j.jmb.2014.01.010

37 Zinzula, L. et al. Purification and functional characterization of the full length recombinant Ebola virus VP35 protein expressed in E. coli. Protein Expr Purif 66, 113-119 (2009). https://doi.org:10.1016/j.pep.2009.02.008

38 Leung, D. W. et al. Structural and functional characterization of Reston Ebola virus VP35 interferon inhibitory domain. J Mol Biol 399, 347-357 (2010). https://doi.org:10.1016/j.jmb.2010.04.022

39 Leung, D. W. et al. Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. Nat Struct Mol Biol 17, 165-172 (2010). https://doi.org:10.1038/nsmb.1765

40 Muhlberger, E., Weik, M., Volchkov, V. E., Klenk, H. D. & Becker, S. Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems. J Virol 73, 2333-2342 (1999). https://doi.org:10.1128/JVI.73.3.2333-2342.1999

41 Basler, C. F. & Amarasinghe, G. K. Evasion of Interferon Responses by Ebola and Marburg Viruses. Journal of Interferon & Cytokine Research 29, 511-520 (2009). https://doi.org:10.1089/jir.2009.0076

42 Leung, D. W. et al. Structure of the Ebola VP35 interferon inhibitory domain. Proc Natl Acad Sci U S A 106, 411-416 (2009). https://doi.org:10.1073/pnas.0807854106

43 Hartman, A. L., Towner, J. S. & Nichol, S. T. A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. Virology 328, 177-184 (2004). https://doi.org:10.1016/j.virol.2004.07.006

44 Cardenas, W. B. et al. Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling. J Virol 80, 5168-5178 (2006). https://doi.org:10.1128/JVI.02199-05

45 Prins, K. C. et al. Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. J Virol 84, 3004-3015 (2010). https://doi.org:10.1128/JVI.02459-09

46 Leung, D. W. et al. Crystallization and preliminary X-ray analysis of Ebola VP35 interferon inhibitory domain mutant proteins. Acta Crystallogr Sect F Struct Biol Cryst Commun 66, 689-692 (2010). https://doi.org:10.1107/S1744309110013266

47 Liu, G. et al. A Sensitive in Vitro High-Throughput Screen To Identify Pan-filoviral Replication Inhibitors Targeting the VP35-NP Interface. ACS Infect Dis 3, 190-198 (2017). https://doi.org:10.1021/acsinfecdis.6b00209

48 Lee, W., Rahimi, M., Lee, Y. & Chiu, A. POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. Bioinformatics 37, 3041-3042 (2021). https://doi.org:10.1093/bioinformatics/btab180

49 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. Progress in Nuclear Magnetic Resonance Spectroscopy 73, 1-16 (2013). https://doi.org:https://doi.org/10.1016/j.pnmrs.2013.02.001

50 Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling 61, 3891-3898 (2021). https://doi.org:10.1021/acs.jcim.1c00203

51 Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31, 455-461 (2010). https://doi.org:https://doi.org/10.1002/jcc.21334

52 Meller, A. et al. Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. eLife 12, e83602 (2023). https://doi.org:10.7554/eLife.83602

53 Kozlov, A. G., Galletto, R. & Lohman, T. M. 55-83 (Humana Press, 2012).

3.8 Supplementary Materials

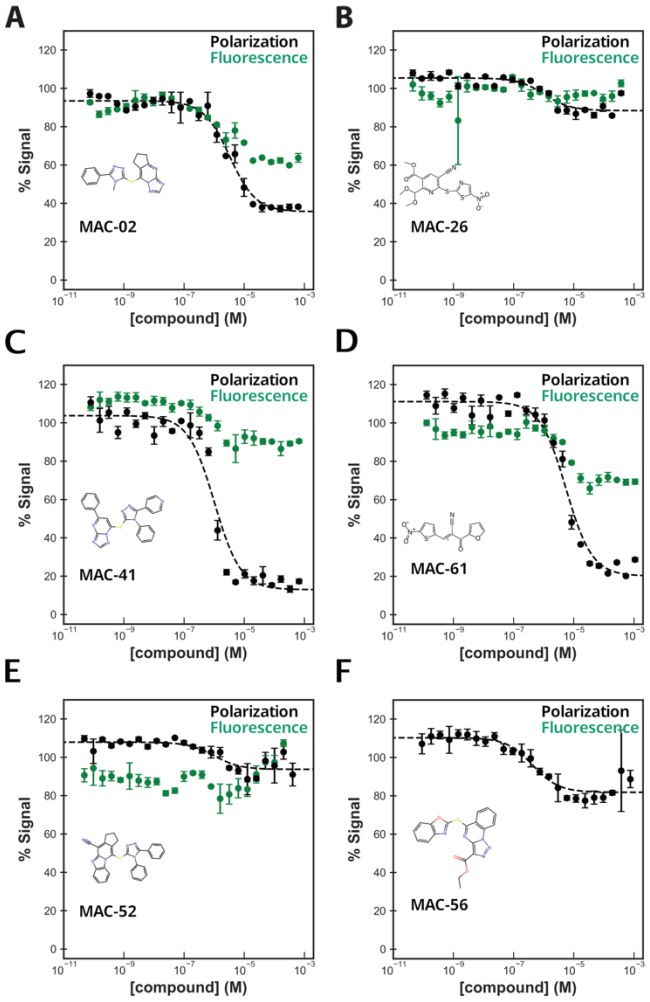

Figure S3.26 Dose dependence experiments for six hit compounds including fluorescence intensity control

Fluorescence (green) and Polarization (black) percent signal change over a range of MAC compound concentrations. Dashed black line is a fit to the polarization data to extract the EC_{50} value. In order A-F are MAC-02, MAC-26, MAC-41, MAC-61, MAC-52, and MAC-56.

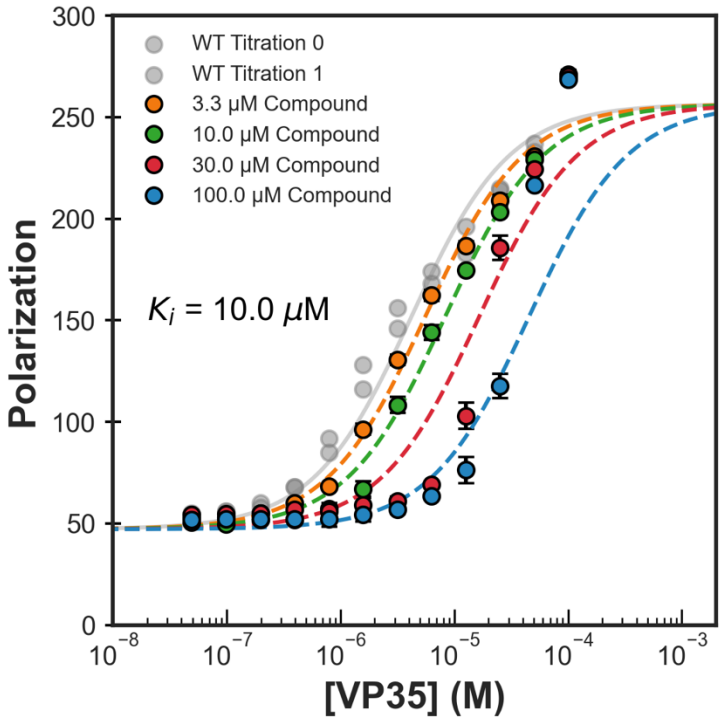

Figure S3.27 MAC-02 inhibition of wild-type VP35 dsRNA binding

Global fit of MAC-02 inhibition of VP35 dsRNA binding at 3.3, 10, 30, 100 µM MAC-02. dsRNA binding was measured as a function of VP35 concentration in the presence of a fixed concentration of MAC-02. Data were globally fit to 1:1:1 competition model to extract Ki as shown on the graph.

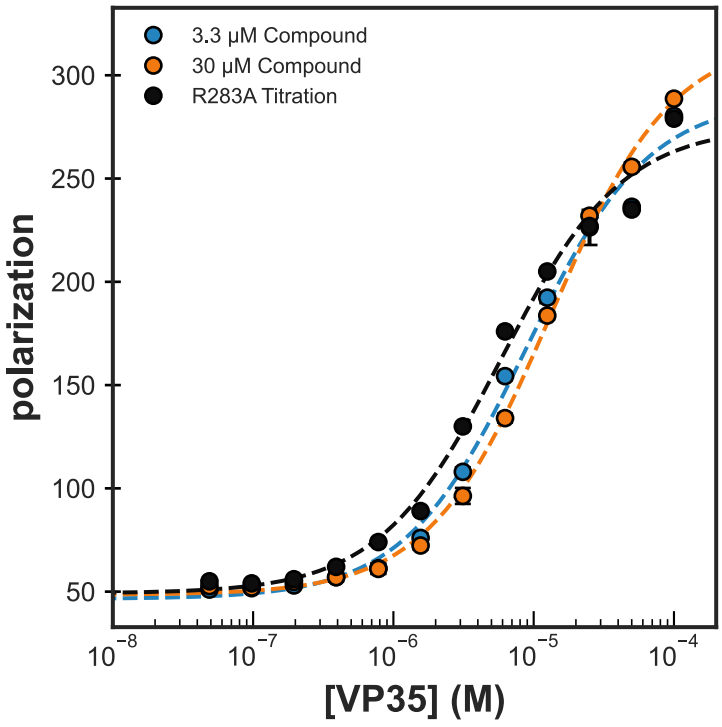

Figure S3.28 Effect of R283A on MAC-02 binding

Binding titration of R283A VP35 with increasing concentrations of MAC-02. Effect of R283A on MAC-02 binding is less potent than F239A but still strong which matches our model since R283 is near slow-exchanging residues.

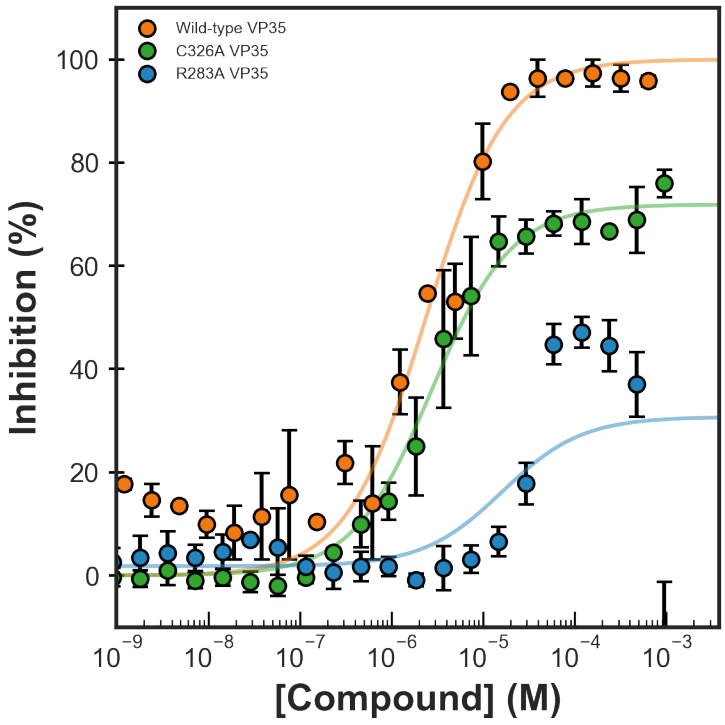

Figure S3.29 Dose dependence of MAC-02 on wild-type, R283A and C326A.

MAC-02 mediated inhibition of wild-type R283A and C326A VP35 variants. Each variant has a different effect on MAC-02 based inhibition. Both R283A and C326A decrease the total inhibition from MAC-02 but only R283A seems to shift the midpoint of the transition. This may suggest that C326A's effect may be indirect or allosteric.

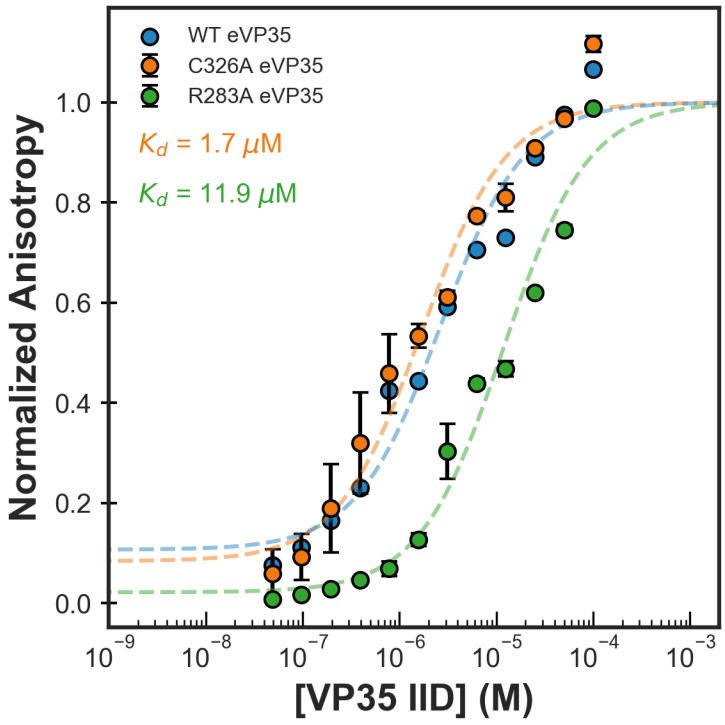

Figure S3.30 Baseline dsRNA binding of wild-type VP35, R283A, and C326A.

Baseline dsRNA binding of R283A, and C326A showing reduced dsRNA binding for R283A but not C326A. Estimated apparent binding affinity shown is from fitting a single site binding model to the data as in Chapter 2.

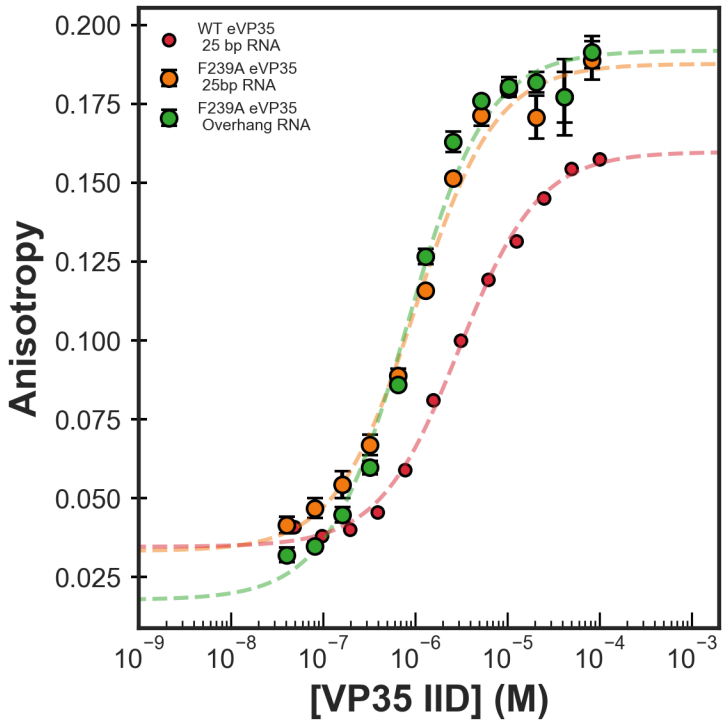

Figure S3.31 Baseline dsRNA binding for F239A

F239A baseline dsRNA binding is stronger than wild-type and insensitive to the topology of dsRNA ends. Binding was measured to two different 25 base-pair RNAs: one with a 3' two nucleotide overhang, the other with blunt ends. F239A does not bind differently to the two different RNAs.

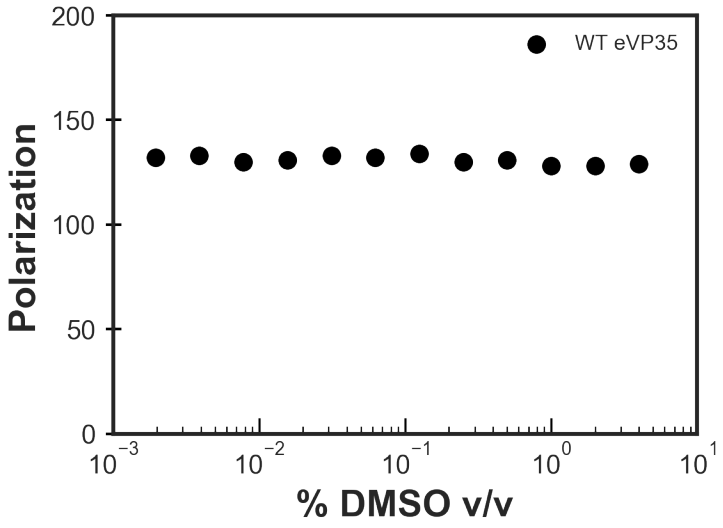

Figure S3.32 DMSO does not disrupt Wild-type VP35 binding to dsRNA

WT VP35 binding to dsRNA in tolerant to a large range of % DMSO values. Polarization of the 25 base-pair RNA was measured in the presence of a fixed concentration of VP35 near its K_d for RNA. Then, the concentration of DMSO was varied and FP read.

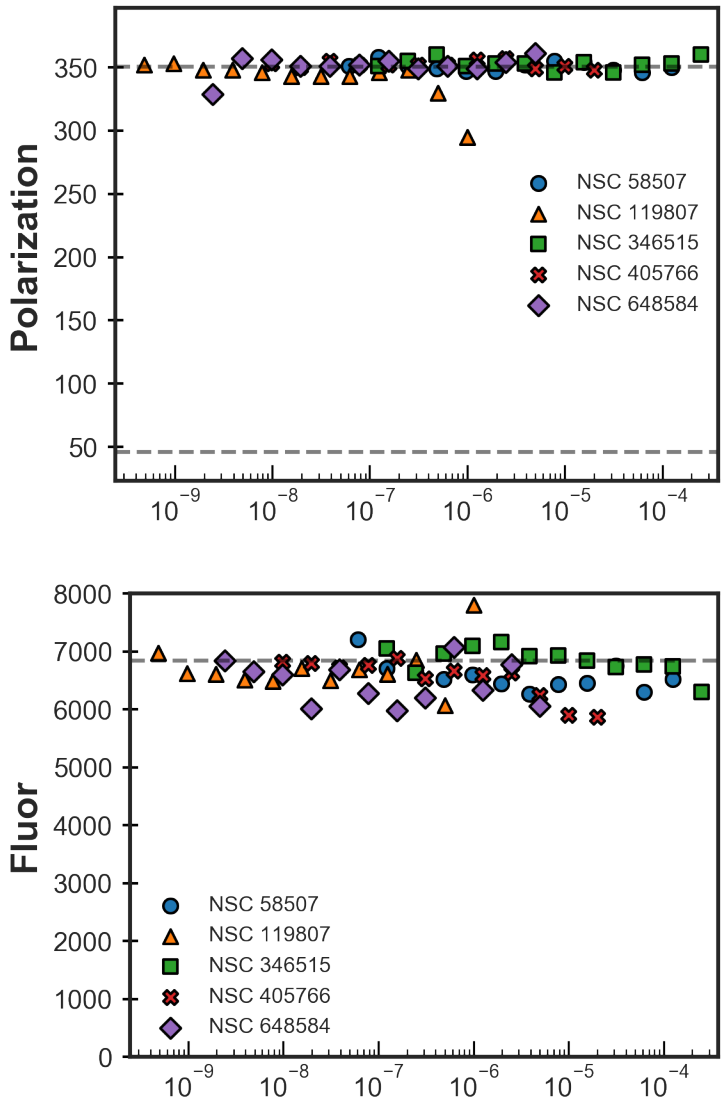

Figure S3.33 Virtual screening cannot identify inhibitors alone.

A filtered version of the ZINC database was docked to states from the VP35 MSM and reweighted by averaging scores across all states. Then, top scoring candidates from virtual screening were ordered and tested for inhibition of VP35-dsRNA binding. None of the tested compounds experimentally inhibit dsRNA binding and that effect is not due to fluorescence interference. The dotted line in each plot represents the polarization or intensity of the labelled 25 base-pair RNA without VP35 present and with VP35 present.

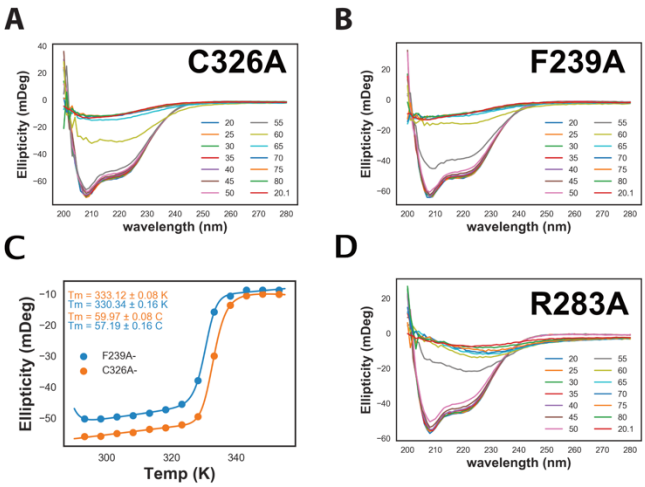

Figure S3.34 Selected mutations do not dramatically disturb VP35's thermodynamic stability.

Temperature induced unfolding experiments by circular dichroism spectroscopy with R283A, C326A and F239A. All three proteins are well folded with similar thermodynamic properties.

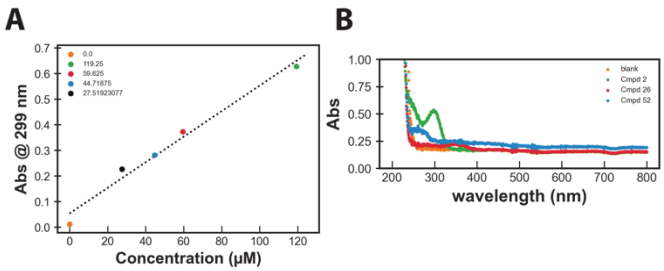

Figure S3.35 Absorbance spectroscopy experiments of MAC-02

The absorbance profile B) of MAC-02 was measured over a range of wavelengths to identify its maximal absorbance peak (299 nm). A) absorbance of MAC-02 is linear with concentration suggesting MAC-02 does not form micelles or aggregates for its inhibition mechanism and is soluble over experimental concentration ranges.

Chapter 4: Conclusions and Future Directions

4.1 Main Findings

Combining computational biophysics tools with experiments enables high resolution understandings of proteins, their functions, and therapeutic opportunities. Much work has been done to understand the importance of conformational dynamics for enzyme function while non-enzymatic proteins such as those that primarily engage in protein-protein (PPI) and protein-nucleic acid interactions (PNIs) have been understudied.[1] Thus, it is necessary to develop new frameworks for how to tackle such targets and transform them into druggable targets. Many PPIs and PNIs occur through flat interfaces necessitating considering each target's conformational dynamics to discover new therapeutic opportunities. Here, I have developed VP35 from the Zaire ebolavirus as a model system for understanding the relationship between a protein's conformation and its function.

First, in chapter 2, we took VP35 which had available structural information and functional information available demonstrating its importance in ebolavirus immune antagonism and replication. Given its central role in viral pathogenesis, VP35 makes an attractive drug target yet it lacks good deep ligand binding sites such as those present in enzymes.[2-7] We hypothesized that VP35 may adopt alternative druggable conformations within its native ensemble. To test this, we conducted long timescale unbiased molecular dynamics simulations seeded from a previous round of enhanced sampling simulations.[8-10] From our simulations we identified that VP35 is a dynamic molecule and adopts a wide variety of conformations not observed in the experimentally derived structure.[11] One in particular opens a cryptic pocket, that is a pocket

absent in available structures but which opens due to protein fluctuations. We then applied a supervised machine learning approach, Diffnets, to identify whether the opening of the cryptic pocket is coupled to functionally important residues and find that the algorithm predicts cryptic pocket opening would remodel residues involved dsRNA binding.[12]

Next, we experimentally tested our predictions by probing the dynamics of VP35 in native conditions using thiol labeling. We find VP35 undergoes structural rearrangements that expose buried cysteines to solvent sufficiently to react with a thiol reactive probe. This reaction takes place independently of protein unfolding which we confirmed through measuring the population of the unfolded state in these conditions and VP35's unfolding rate.

Finally, to assess if this cryptic pocket is druggable, we measured VP35's dsRNA binding activity after increasing the population of the closed state, then after increasing the population of the open state. To increase the population of the closed state, we leveraged a mutation that introduces rigidity on a hinge loop between the two domains involved in pocket opening and confirmed it is more closed with the thiol labeling experiment. When the population of the cryptic pocket open states decreases, VP35 binds dsRNA with higher apparent binding affinity. This finding demonstrates that the closed state is favored for dsRNA binding as suggested in our simulation data. Then, we covalently modified the cysteine residues in the cryptic pocket with a drug fragment sized probe to stabilize the pocket in its open conformation. Structural experiments demonstrate that the protein remains folded after this modification. When we measured the modified protein's dsRNA binding properties, we find a significant decrease in apparent dsRNA binding affinity. Decreased dsRNA binding confirms that the VP35 pocket is coupled to dsRNA binding and stabilizing the pocket in its open state similarly to if a compound

was bound, supports our hypothesis that targeting VP35's alternative conformations is a great therapeutic opportunity.

Furthermore, through combining computer simulations and experiments, we generated testable detailed hypotheses that reveal VP35's important druggable conformations. The synergy between MD simulations, machine learning approaches, and biophysical and functional assays enable a novel perspective on VP35's druggability and function. We also believe our approach can serve as a general strategy for identifying and testing the importance of alternative conformations in any protein target of interest to expand the druggable proteome or even rapidly identify novel therapies targeting new pathogens.[13,14]

After developing VP35 as a drug target, the next step is to attempt to target VP35 with small molecules. To do this, as described in Chapter 3, we conducted high throughput screening against VP35 to identify inhibitors of dsRNA binding. We based our screening strategy on a similar study which aimed to discover inhibitors of a PPI.[15] Using fluorescence polarization of a fluorescently labeled dsRNA, we screened the Maybridge Hit Finder library and identified 62 hit compounds. This represents a wide portion of chemical space as many different scaffolds and functional groups are present in our hit list.

Initial biochemical experiments demonstrate that our hit list comprises a range of inhibitory effects from as little as 15% inhibition to as much as 85% and maybe more in the untested compounds. Toward our original purpose of identifying small molecules that bind to the VP35 cryptic pocket, we conducted SAR studies by NMR chemical shift perturbation experiments.[16] We find that compounds that inhibit dsRNA binding induce perturbations at the dsRNA binding interface suggesting they are binding and directly competing with dsRNA. This may also explain

why so many of the tested compounds have only modest EC_{50}s in the micromolar range since the dsRNA binding interface lacks the three-dimensional structure which is necessary for forming enough interactions for high affinity binding.

Interestingly, one compound does seem to bind to the cryptic pocket. This compound MAC-02 induces widespread perturbations by NMR but these can be grouped into two categories. The slow exchange residues as evidenced by titration experiments fall at the dsRNA binding interface. The second category are the intermediate and fast exchange residues which are near the cryptic pocket. We introduced mutations to the dsRNA binding interface and to residues in the cryptic pocket and find that both disrupt inhibition of dsRNA binding by MAC-02 though to different extents. To assess if it is possible for a compound to bind to both the interface and to the pocket, we conducted a molecular docking study with MAC-02. The docking results show that MAC-02 is sufficiently small that two molecules can interact with both the dsRNA binding interface and the cryptic pocket independently. In fact, intrinsic tryptophan fluorescence quenching experiments confirm that MAC-02 binding to VP35 saturates at 2 moles of MAC-02 per mole of VP35. Combining atomistic resolution computational tools with biochemical and. biophysical experiments enabled us to learn that VP35 harbors a cryptic pocket that is coupled to dsRNA binding and show that this site is labile for drug discovery through traditional means.

4.2 Closing Thoughts

VP35 is a relatively labile protein to work with as it contains minimal intrinsically disordered regions in its IID, is generally robust to mutation, it can be expressed to high concentrations and purified in milligram quantities from one liter of culture. Yet, less than ten years ago, its prospects as a drug target seemed slim. VP35 lacks enzymatic activity so substrate analogs are not a possible therapeutic direction. RNA aptamers are an option but developing aptamers into

therapies is still an ongoing challenge so there seemed to be few possibilities.[17] In 2016 then again in 2018, the idea of targeting the dsRNA binding interface directly started to show up in publications suggesting an ongoing interest in targeting the VP35's immune antagonist functions.[18,19]

What remained exciting about VP35 and the studies we undertook is that we were constantly at this interface of known approaches such as high-throughput screening or analysis of functional conformations, yet certain things remained elusive. Principal of these is the lack of a straightforward way of identifying if a given inhibitor acts allosterically, competitively, or non-competitively. Enzyme targeting has a great advantage in the framework of Michaelis-Menten kinetics, but no such framework exists for systems such as PPIs and PNIs. Regardless, applying well known techniques and theories has been sufficient to characterize VP35's conformational landscape and identify the mechanism whereby hit molecules inhibit dsRNA binding. My hope is that the biophysics and infectious disease fields find this work helpful for future drug targeting efforts in systems considered intractable for drug discovery.

4.3 Future Directions

Inasmuch as I have been attempted to do in this work, much remains to be done to transform VP35 into not just a druggable target but a drugged target and gain a deeper understanding of the contribution of conformational dynamics to VP35's other functions. This work describes new chemical matter that binds to VP35 and inhibits dsRNA binding. Generating a real lead compound from this work will require much work.

First, higher resolution structural information is essential for medicinal chemistry endeavors moving forward to both optimize these compounds for use as pharmacological tools and drugs.

The high-resolution x-ray diffraction structures of VP35 in complex with these hit compounds will go far toward confirming the results in this work and providing medicinal chemists necessary information for deriving new molecules that bind with high affinity to the cryptic pocket. While structures are in progress, there are available molecules through the ZINC database with analogous chemical structures.[20] Both NMR chemical shift perturbation and X-ray crystallography experiments with these analogs should be of highest priority. The medicinal chemistry steps are necessarily secondary but vitally important to both improve the utility of these compounds for further study and if any are to advance toward lead status.

Next, even with their relatively low EC_{50}s, all molecules in the hit list are Rule of 3 compliant and have favorable cell permeabilities. As such, each hit should be tested for if it can restore activation of the interferon response in model systems of infection. These experiments are crucial to demonstrate whether the cryptic pocket has relevance in the infection context and if these molecules are useful for the investment that would be required to truly transform them into lead molecules. In addition to cellular-based assays, one biochemical assay in particular stands out as being particularly important to be done with these compounds. VP35 specifically antagonizes activation of RIG-I like receptors and these are ATPases whose activity can be measured *in-vitro*.[21] RLR ATPase activity should be measured in the presence and absence of compounds that inhibit VP35's binding to dsRNA. If compounds *in-vitro* reenable RLR activation, the case for continued study of VP35 as a drug target and these compounds as potential molecules, is much stronger.

Targeting VP35's dsRNA binding activity is one therapeutic strategy targeting VP35. But VP35 plays an important role in replication of the viral genome, and viral packaging through its interactions with nucleoprotein (NP).[5,15,22-25] The effect of VP35's conformational dynamics on

the VP35-NP interaction remains unknown. While we show that the pocket closed state in VP35 favors binding to dsRNA, it is possible though untested, that NP can bind to VP35 in the pocket open state. In this model, the cryptic pocket would act as a switch between different binding modes in VP35 from dsRNA bound to NP bound. This switch may occur once there are high levels of VP35 around after early stages of infection such that VP35 not bound to dsRNA is now available to interact with NP to modulate packaging and replication through L (the RNA dependent RNA polymerase).

Studies to test the effect of VP35's conformation on its interactions with NP should include hydrogen deuterium exchange mass spectrometry experiments on VP35-NP complexes in the wild-type case and with mutations in VP35 that increase the population of the closed state (A291P VP35). With increasing concentrations of NP, if wild-type VP35 hydrogen-deuterium exchange in the pocket region decreases, that would suggest NP does also bind to the pocket closed state of VP35. In the context of A291P, one would expect very little exchange to start with providing a good control for exchange in wild type. Also, it would be important to conduct binding studies such as by Surface Plasmon Resonance (SPR) or isothermal titration calorimetry (ITC) of the A291P VP35 variant binding to NP to determining if varying the population of the pocket open state affects binding to NP.

In cell studies of the VP35 mutants generated during this study would be of great utility for correlating the *in-vitro* observed biophysics to the resulting cellular phenotypes. Previous studies have utilized a minigenome assay which couples viral replication to luciferase activity.[26] The mutations A291P, F239A and C326A should individually be introduced to the minigenome and their effects on viral replication could be measured. If A291P increases dsRNA binding in the cellular environment as it does in the biochemical assays, and the result is a decrease in viral

replication, that may support the hypothesis that the cryptic pocket is targetable in the cellular context as well. Similarly, if F239A (which we show has a higher open pocket population in Chapter 2) demonstrates less viral replication, that supports that the pocket open state is a good therapeutic target. For the C326A mutation, if interactions between other binding partners are modulated through the pocket as the MAC-02 interactions are, it is informative toward demonstrating that the pocket is not only coupled to functional sites but may itself be a functional site.

Similarly, in mammalian cells, it is possible to measure the extent of interferon antagonism for various VP35 mutants. If decreasing dsRNA binding through mutation or small-molecule binding is of therapeutic use, it would need to result in increased interferon signaling. As done previously, to parse if the effect is due to a change in dsRNA binding or VP35's interactions with other host factors, the interferon signaling assay can be performed in an RNA dependent or independent manner.[4]

The translation of these biophysical results into the cellular context is the next important step in demonstrating the therapeutic utility of VP35's cryptic pocket. Through measuring viral replication, interferon signaling, RIG-I activation and NP interactions as a function of the VP35 cryptic pocket, it is possible to begin understanding the importance this alternative conformation may play in viral pathogenesis.

4.4 References

1 Johnson, D. K. & Karanicolas, J. Computational Screening and Design for Compounds that Disrupt Protein-protein Interactions. *Current Topics in Medicinal Chemistry* **17**, 2703-2714 (2017). https://doi.org:10.2174/1568026617666170508153904

2 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Reports* **14**, 1632-1640 (2016). https://doi.org:https://doi.org/10.1016/j.celrep.2016.01.049

3 Hartman, A. L., Towner, J. S. & Nichol, S. T. A C-terminal basic amino acid motif of Zaire ebolavirus VP35 is essential for type I interferon antagonism and displays high identity with the RNA-binding domain of another interferon antagonist, the NS1 protein of influenza A virus. *Virology* **328**, 177-184 (2004). https://doi.org:10.1016/j.virol.2004.07.006

4 Leung, D. W. *et al.* Structural basis for dsRNA recognition and interferon antagonism by Ebola VP35. *Nat Struct Mol Biol* **17**, 165-172 (2010). https://doi.org:10.1038/nsmb.1765

5 Prins, K. C. *et al.* Basic residues within the ebolavirus VP35 protein are required for its viral polymerase cofactor function. *J Virol* **84**, 10581-10591 (2010). https://doi.org:10.1128/JVI.00925-10

6 Prins, K. C., Cardenas, W. B. & Basler, C. F. Ebola virus protein VP35 impairs the function of interferon regulatory factor-activating kinases IKKepsilon and TBK-1. *J Virol* **83**, 3069-3077 (2009). https://doi.org:10.1128/JVI.01875-08

7 Prins, K. C. *et al.* Mutations abrogating VP35 interaction with double-stranded RNA render Ebola virus avirulent in guinea pigs. *J Virol* **84**, 3004-3015 (2010). https://doi.org:10.1128/JVI.02459-09

8 Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proceedings of the National Academy of Sciences* **112**, 2734-2739 (2015). https://doi.org:10.1073/pnas.1417811112

9 Bowman, G. R. & Geissler, P. L. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proceedings of the National Academy of Sciences* **109**, 11681-11686 (2012). https://doi.org:10.1073/pnas.1209309109

10 Hart, K. M., Ho, C. M. W., Dutta, S., Gross, M. L. & Bowman, G. R. Modelling proteins' hidden conformations to predict antibiotic resistance. *Nature Communications* **7**, 12965 (2016). https://doi.org:10.1038/ncomms12965

11 Cruz, M. A. *et al.* A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nat Commun* **13**, 2269 (2022). https://doi.org:10.1038/s41467-022-29927-9

12 Ward, M. D. *et al.* Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets. *Nature Communications* **12**, 3023 (2021). https://doi.org:10.1038/s41467-021-23246-1

13 Knoverek, C. R., Amarasinghe, G. K. & Bowman, G. R. Advanced Methods for Accessing Protein Shape-Shifting Present New Therapeutic Opportunities. *Trends in Biochemical Sciences* **44**, 351-364 (2019). https://doi.org:10.1016/j.tibs.2018.11.007

14 Zimmerman, M. I. *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nature Chemistry* **13**, 651-659 (2021). https://doi.org:10.1038/s41557-021-00707-0

15 Liu, G. *et al.* A Sensitive in Vitro High-Throughput Screen To Identify Pan-filoviral Replication Inhibitors Targeting the VP35-NP Interface. *ACS Infect Dis* **3**, 190-198 (2017). https://doi.org:10.1021/acsinfecdis.6b00209

16 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* **73**, 1-16 (2013). https://doi.org:https://doi.org/10.1016/j.pnmrs.2013.02.001

17 Binning, J. M. *et al.* Development of RNA Aptamers Targeting Ebola Virus VP35. *Biochemistry* **52**, 8406-8419 (2013). https://doi.org:10.1021/bi400704d

18 Glanzer, J. G. *et al.* In silico and in vitro methods to identify ebola virus VP35-dsRNA inhibitors. *Bioorg Med Chem* **24**, 5388-5392 (2016). https://doi.org:10.1016/j.bmc.2016.08.065

19 Daino, G. L. *et al.* Identification of Myricetin as an Ebola Virus VP35-Double-Stranded RNA Interaction Inhibitor through a Novel Fluorescence-Based Assay. *Biochemistry* **57**, 6367-6378 (2018). https://doi.org:10.1021/acs.biochem.8b00892

20 Irwin, J. J. *et al.* ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling* **60**, 6065-6073 (2020). https://doi.org:10.1021/acs.jcim.0c00675

21 Edwards, M. R. *et al.* Differential Regulation of Interferon Responses by Ebola and Marburg Virus VP35 Proteins. *Cell Rep* **14**, 1632-1640 (2016). https://doi.org:10.1016/j.celrep.2016.01.049

22 Brown, C. S. *et al.* In silico derived small molecules bind the filovirus VP35 protein and inhibit its polymerase cofactor activity. *J Mol Biol* **426**, 2045-2058 (2014). https://doi.org:10.1016/j.jmb.2014.01.010

23 Dong, S. *et al.* Insight into the Ebola virus nucleocapsid assembly mechanism: crystal structure of Ebola virus nucleoprotein core domain at 1.8 Å resolution. *Protein & Cell* **6**, 351-362 (2015). https://doi.org:10.1007/s13238-015-0163-3

24 Leung, Daisy W. *et al.* An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Reports* **11**, 376-389 (2015). https://doi.org:https://doi.org/10.1016/j.celrep.2015.03.034

25 Trunschke, M. *et al.* The L-VP35 and L-L interaction domains reside in the amino terminus of the Ebola virus L protein and are potential targets for antivirals. *Virology* **441**, 135-145 (2013). https://doi.org:10.1016/j.virol.2013.03.013

26 Muhlberger, E., Weik, M., Volchkov, V. E., Klenk, H. D. & Becker, S. Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems. *J Virol* **73**, 2333-2342 (1999). https://doi.org:10.1128/JVI.73.3.2333-2342.1999

Appendix: MSM Reweighted Docking Of Hit Compounds To VP35.

An ongoing area of research for drug discovery is that of virtual screening. Virtual screening via molecular docking is the calculation of a ligand's binding affinity for a given protein receptor from physical principles such as the energy of hydrogen bonding, among others. This technique would prove exceptionally useful for the high-throughput screening of millions of molecules against newly identified pockets on a known target or against any available binding site on a new target. Yet virtual screening remains somewhat underutilized as its predictive capacity falls short. For example, in work unreported in this thesis but done as a matter of course, virtual screening predicted a compound to have a binding energy of ~-10 kcal per mole but experimental testing found that compound to have no activity against the target. Other approaches such as alchemical free energy calculations are more predictive but also more computationally expensive.

It has been found that the predictive capacity of docking studies can be improved through considering the whole ensemble of states a protein can adopt.[1] In the course of the work reported in Chapter 3, we found ourselves in possession of a longer hit list of compounds than is feasible to structurally characterize one at a time. So we had a need to prioritize compounds for testing. While our principal interest was in identifying compounds that bind to the VP35 cryptic pocket, the need for reliable virtual screening approaches is even greater if one is interested in identifying which is the highest affinity binding compound in a list.

Thus, here I briefly report a small development in the application of molecular docking for prioritizing on which small molecules from a list of hit compounds to perform structural studies.

Method

Molecular dynamics simulations and the Markov State Model (MSM) as described in Chapter 2, as well as the 62 hit molecules from the high throughput screen described in Chapter 3 were used to conduct molecular docking using the Vina energy function.[2,3] All hit molecules were docked against every state (4469 states) from our MSM with exhaustiveness set to 32, number of output poses set to 1, and two different docking boxes were used. The first, centered around the blunt-end binding site had center 57.863, 59.645, 7.277, and size 16, 16, 16. The second, centered on the cryptic pocket had center 58, 41, 5 and size 20, 20, 20. We then extracted output docking scores and reweighted them by the equilibrium probability associated with the state from which the score was calculated according to the previously described equation:$\Delta G = -RTln\left(\sum \pi_i K_{eqi}\right)$. Where R is the gas constant, T is temperature in kelvin, $\pi_i$is the equilibrium probability of the MSM state, and K is the equilibrium constant calculated from the docking score to the ith state. These calculations were done with R in kcal/mol and T=300 K. Finally with the reweighted scores for every compound per docking site (pocket or interface), we calculated a $\Delta\Delta G$ defined as $\Delta\Delta G = \Delta G_{interface} - \Delta G_{pocket}$ to identify which compounds have more favorable binding energy to the cryptic pocket over the RNA binding interface. In this case compounds with positive $\Delta\Delta G$ are predicted to bind more favorably to the cryptic pocket.

Results

We find that there is a very similar distribution of binding affinities between the two sites before $\Delta\Delta G$ calculation though different in magnitude (Fig. A36). The mean binding energy for binding to the dsRNA binding interface is -4.75 kcal/mol while the approximate mean value for binding to the pocket is ~ -5.5 kcal/mol across all compounds. Some molecules are predicted to bind with

modestly strong binding affinity to the cryptic pocket ~ -7 kcal/mol but none have predicted binding energies of less than -8 kcal/mol suggesting few of these compounds are strongly binding.

After calculating $\Delta\Delta G$, we do find some compounds do have different binding affinities to the two sites such that binding favors the pocket but none of our molecules favor the pocket over the interface by more than 1.5 kcal/mol suggesting that none of the compounds would be strongly binding to the pocket over the interface.

A binding energy of ~-6-8 kcal/mol corresponds to weak binding and docking algorithms have the pathology of being able to calculate a docking energy even for non-binders. Our approach is likely more predictive for compounds for which binding is significantly preferred at one site ($\Delta\Delta G$ greater than 2 kcal/mol) over the other. In the regime in which either site is approximately equally favored, new virtual screens should be conducted with analogs to begin identifying useful substitutions to favor binding to the site of interest.

Future work is needed to better understand how to deploy useful tools such as docking to improve the efficiency of high-throughput screening and follow up studies. Also, improvements to the way compounds are docked (rigid, flexible, or ensemble), or how their binding energies are calculated may improve results such as these.

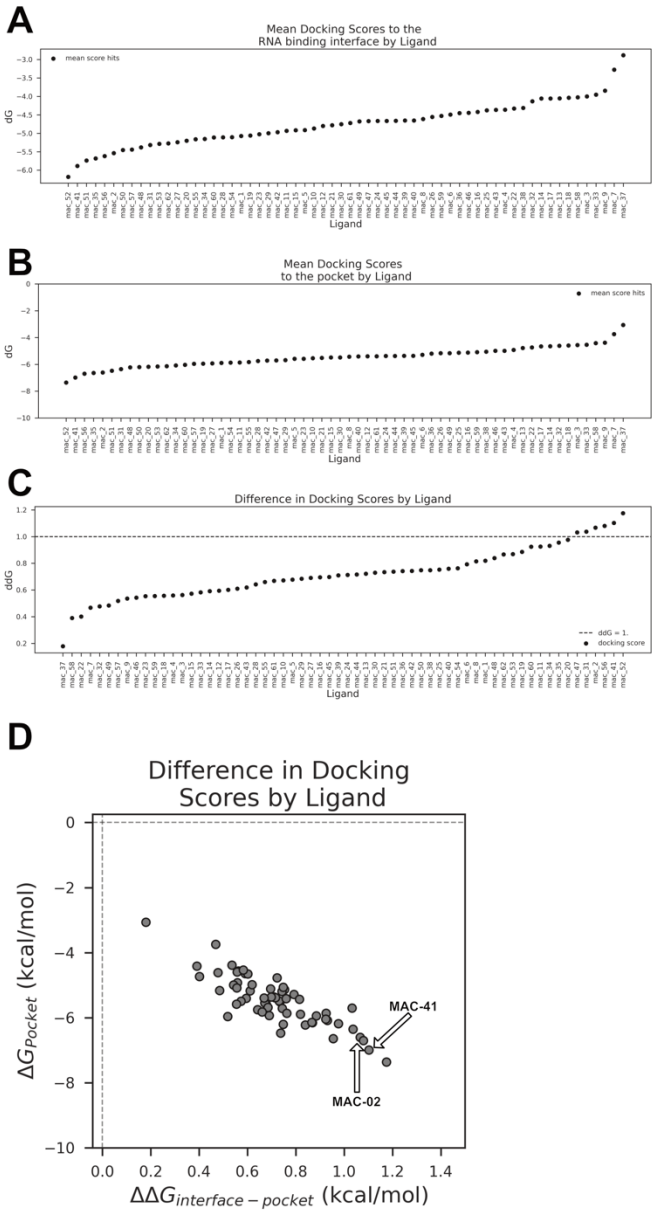

Figure A36 Reweighted docking scores for compounds to the two VP35 binding sites.

A), B) MSM reweighted docking scores of all 62 compounds from the HTS in chapter 3 ordered from most to least favorable, to the interface, and cryptic pocket, respectively. C) $\Delta\Delta G$ values for all compounds ranked from least favorable for binding to the cryptic pocket, to most. The dotted line shows a 1 kcal/mol difference threshold. D) Demonstration that cryptic pocket binding favorability qualitatively correlates with the $\Delta\Delta G$ calculation. The values on the y-axis are the same as in panel B and the values on the x-axis are the same as in panel C. The dotted lines are for binding energy of 0 kcal/mol.

References

1 Meller, A. *et al.* Drug specificity and affinity are encoded in the probability of cryptic pocket opening in myosin motor domains. *eLife* **12**, e83602 (2023). https://doi.org:10.7554/eLife.83602

2 Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling* **61**, 3891-3898 (2021). https://doi.org:10.1021/acs.jcim.1c00203

3 Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455-461 (2010). https://doi.org:https://doi.org/10.1002/jcc.21334

Curriculum Vitae

Matthew A. Cruz

PhD level biophysicist with eight years' experience in protein structure-function relationships, and protein-nucleic acid interactions. Expertise in molecular dynamics simulations, programming and data visualization in Python, high throughput spectroscopic techniques, and stopped-flow kinetics.

Education

Ph.D. Biochemistry, Biophysics, and Structural Biology 2017–2023

Washington University in St. Louis (WUSTL)

Laboratory of Gregory R. Bowman

Qualifying Exam Passed February 2019, GPA: 3.5/4.0

Thesis: 'Leveraging protein dynamics to drug filovirus protein-nucleic acid interactions using simulations and experiments'

B.S. Biochemistry and Cell Biology 2013–2017

Rice University, Rice Emerging Scholar

Honors and Awards

MilliporeSigma Fellowship December 2021

$25,00 one time

NIH Kirschstein National Research Service Award September 2021

NIAID F31 AI157079 "Leveraging protein dynamics to drug filovirus protein-nucleic acid interactions using simulations and experiments"

$32k research stipend and institutional allowance, renewable for three years

Elson Fellowship November 2020

$1,500 one time

NSF Graduate Research Fellowship April 2019

Honorable Mention

WUSTL Center for Drug Discovery Matching Grant May 2019

'Small molecule disruption of Ebolavirus immune evasion'

$7,500 match

WUSTL Institute for Maximizing Student Development Scholar May 2019–2020

$30,500 research stipend

Suggs Undergraduate Research Fellowship May–August 2016

$3000 summer research stipend

Publications

1. **Cruz MA***, Frederick TE*, Mallimadugula, UL, Singh S, Vithani N, Zimmerman MI, Porter JR, Moeder KE, Amarasinghe GK, and Bowman GR (2022), A cryptic pocket in Ebola VP35 allosterically controls RNA binding. **Nature Communications**, 13: 2269. doi: 10.1038/s41467-022-29927-9
2. Xu JS*, Hewitt MN*, Gulati JS*, **Cruz MA***, Zhan H, Liu S and Matthews KS (2018), Lactose repressor hinge domain independently binds DNA. **Protein Science**, 27: 839-847. doi:10.1002/pro.3372

Teaching and Mentorship

Undergraduate Guided Research Mentor June 2020–June 2022

Washington University in St. Louis

In-Lab Research mentor for undergraduate student Rishi Samarth

Graduate Rotation Research Mentor November–January 2020

Washington University in St. Louis

In-Lab Research mentor for graduate student Bejan Mahmud

Macromolecular Interactions January–May 2019

Washington University in St. Louis

Tutoring Session Coordinator

Chemistry and Physics of Biomolecules August–December 2018

Washington University in St. Louis

Assistant to the Instructor, under Jay Ponder Ph.D.

Young Scientist Program Summer Focus Summers 2018–2019

Washington University in St. Louis

Tutor

Advanced Experimental Biosciences August 2016–May 2017

Rice University

Laboratory Teaching Assistant

Research Experience

Graduate Research, under Gregory R. Bowman Ph.D.

February 2018–Present

Washington University in St. Louis, Department of Biochemistry and Molecular Biophysics, St. Louis, Missouri

Disovered an allosteric small molecule inhibitor of VP35 dsRNA binding

- Conduct a high throughput screen of a large library to identify small molecule inhibitors of dsRNA binding to VP35 using fluorescence polarization.
- Determined the potency of screening hits via dose-dependence experiments.
- Identified the enriched-for chemistry from the screening library via chemoinformatics using Python and RDKit.
- Determined the binding site for hit compounds to VP35 using Nuclear Magnetic Resonance spectroscopy.
- Conducted biophysical experiments testing the stoichiometry of compounds binding to VP35 to assess the mechanism of action.
- Performed molecular docking against the VP35 ensemble to predict how given compounds interact with VP35 to inhibit dsRNA binding.

Demonstrate the existence of a cryptic pocket in the marburgvirus VP35 homolog

- Analyze molecular dynamics data to predict the presence and coupling of a cryptic pocket in mVP35 to functional sites using Markov state models
- Conduct side-chain based allosteric network analysis from side-chain fluctuations
- Experimentally probe side-chain solvent accessibility with thiol-labeling assay
- Measure RNA binding of VP35 in pocket-open and pocket-closed states to test allosteric effect of pocket dynamics

- Measure VP35 variants' thermodynamic stability using chemical and temperature denaturation experiments
- Perform circular-dichroism spectroscopy to assess VP35's fold before and after covalently stabilizing the pocket open state
- Screen large drug-like library for small molecules that bind to the Ebola or Marburg VP35 cryptic pocket

Discovered an allosteric Cryptic Pocket in ebolavirus VP35

- Demonstrated the experimental existence of a fluctuation in VP35's native state that exposes buried cysteines to solvent using a thiol labeling assay
- Developed and optimized an anisotropy based binding assay for double stranded RNA including determining best fit model for binding data
- Assessed the ability of VP35 to bind double stranded RNA substrates before and after opening of the cryptic pocket
- Demonstrated an allosteric decrease in RNA binding after stabilizing the VP35 pocket open state
- Measured two VP35 variants' thermodynamic stability from intrinsic tryptophan fluorescence chemical denaturation
- Demonstrated retention of the protein's fold using circular-dichroism spectroscopy

Undergraduate Research, under Kathleen Matthews Ph.D.

January 2015–August 2017

Rice University Department of Biosciences, Houston, Texas

Binding Assessment of LacI Hinge-Helix Binding Domain

- Examined the DNA binding activity of a dimeric peptide homologous to the hinge region of the E. coli lac repressor helix-turn-helix DNA binding domain.
- Demonstrated nonspecific binding of a LacI hinge peptide to both operator and scrambled DNA sequences using filter binding assay with radiolabeled DNA.
- Showed the necessity of hinge peptide dimerization for DNA binding revealing the importance of this structural motif for DNA binding.
- Measured DNA binding using fluorescence and pull-down methods
- Drafted, edited, and published a report describing the independent binding of the LacI hinge domain to DNA
- Consulted with course instructors to generate teaching lab project to further understand the structural determinants of DNA binding and looping in LacI and the purine repressor (PurR) using chimeric protein constructs.

Presentations including Posters

- **Biochemistry and Molecular Biophysics Retreat, UPenn, Poster** October 2022
- **The Protein Society annual meeting Poster** July 2022

- **Biophysics at the Dawn of Exascale Computers Poster** May 2022
- **Protein Folding Consortium Poster** June 2021, 2019
- **Department of Biochemistry and Molecular Biophysics** February 2020, 2018
- **Biophysical Society Annual Meeting Poster, Platform Talk** February 2020, 2022
- **Gibbs Biothermodynamics Conference Poster** October 2019
- **Biochemistry, Biophysics and Structural Biology** October 2019

  Program Retreat Poster
- **Division of Biology and Biomedical Sciences Poster** August 2018
- **Gulf Coast Undergraduate Research Symposium** October 2016
- **Rice University Institute of Biosciences and Bioengineering Poster** August 2016
- **Rice Undergraduate Research Symposium Poster** April 2015, April 2016

Committee Service

Student Liaison Committee, *Member, Chair* May 2018–May 2022

Department of Biochemistry and Molecular Biophysics

Washington University in St. Louis, St. Louis, Missouri

- Represents the graduate students in the department and advises the department chair on graduate student concerns and interests
- Organizes a weekly Science Friday seminar as a Science Friday subcommittee member including introducing speakers and moderating question-answer sessions

Relevant Courses

- Chemistry and Physics of Biomolecules
- Nucleic Acid and Protein Synthesis
- Macromolecular Interactions
- Introductory Biostatistics
- Computational Biophysics Journal Club