GREATER THAN THE SUM OF ITS PARTS: UNDERSTANDING ALPHAVIRUS INFECTION THROUGH CHARACTERIZATION OF NONSTRUCTURAL POLYPROTEINS

By

SAMANTHA ALEXIS YOST

A dissertation submitted to the

Graduate School-New Brunswick

And

The Graduate School of Biomedical Sciences

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Microbiology and Molecular Genetics

Written under the direction of

Joseph Marcotrigiano, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2016

ABSTRACT OF THE DISSERTATION

Greater Than the Sum of Its Parts: Understanding Alphavirus Infection through Characterization of Nonstructural Polyproteins

By SAMANTHA ALEXIS YOST

Dissertation Director:
Joseph Marcotrigiano, Ph.D.

Alphaviruses are enveloped, positive strand RNA viruses that are capable of causing a variety of symptoms including arthritis, encephalitis, and even death in humans. Chikungunya virus, a member of the alphavirus genus, is a recently emerging pathogen in the Americas with over one million suspected cases in the Americas in 2014. There is no specific anti-viral treatment or vaccine against any member of the alphavirus genus. The viral nonstructural proteins (nsP1-4) responsible for viral RNA replication are produced as a polyprotein. RNA replication complex template usage is regulated by proteolytic processing of the nonstructural polyproteins. Information about the rearrangement of viral proteins before and after cleavage is essential to understanding of the progression of viral pathogenesis. After cleavage, the mature nsPs each assume individual functions in the RNA replication complex or elsewhere for productive infection.

We described a portion of alphavirus polyprotein P23 from Sindbis virus and chikungunya virus. The structures show four domains each occupying a vertex of a rectangle. A 40 Å distance between the nsP2 protease and P2/3 cleavage site supports the proposed *trans* cleavage mechanism. Our P23 structure was used for ligand docking and will be a useful tool in the development of unique anti-viral therapies. Thus far, nsP3 is the least understood alphavirus protein. In the polyprotein, nsP3 forms an extended linker, encircling nsP2,

connecting its two domains, important in efficient polyprotein processing and the progression of infection. The P23 structure also revealed a novel zinc binding domain of nsP3 necessary for viral infection. Further biochemical study indicates a role for nsP3 in regulating nsP2 ATP hydrolysis activity. Additionally, nsP3 macrodomain has been shown to bind ADP-ribose but the significance of this during infection is unknown. Abolishing ADP-ribose binding has a significant negative effect on replication in insect cells, whereas previously only the nsP3 C-terminal region was known to mediate infection in various host cells.

Overall, our strategy of describing alphavirus nsPs with regard to their structure and activities pre- and post-cleavage via biophysical, biochemical, and virological methods has significantly contributed to a better understanding of viral pathogenesis and present opportunities for novel anti-viral therapy design.

ACKNOWLEDGMENTS

Dr. Joseph Marcotrigiano is an outstanding mentor, scientist, and friend. I am exceedingly grateful he chose to take a chance on me, teach me, and push me beyond my limits ("Serenity now!"). Through the years, his dedication to science and seemingly never ending stream of new ideas kept me both awed and inspired.

As a second generation graduate student of the Marcotrigiano lab, I must acknowledge those that came before me: Gyehwa Shin, Dr. Fuguo Jiang, Dr. Jillian Whidby, and Ankita Basant. This thesis would not have been possible without your hard work to grow the lab from its infancy. To Alicja Cygan, Lexi Zatorski, and our many other past members who lent support to me during the past six years: thank you. My current lab mates have been a pleasure to work with and dear friends to me; thank you Jennifer Casiano, Dr. Abdul Khan, Amoli Kulkarni, Ryan Levy, Dr. Matthew Miller, Lynda Tuberty, Chen Wang, and Yuanyuan Wang. The students, faculty, and staff of Center for Advanced Biotechnology and Medicine are absolutely wonderful.

The guidance and wisdom provided by my committee members Drs. Michael Hampsey, Vikas Nanda, and Smita Patel have been indispensable throughout this journey. Thank you for encouraging and challenging me. To our collaborators Drs. Arash Grakoui, Craig Cameron, Stefano Forli, and many others: thank you for the insight, reagents, and support contributed to this thesis.

I must give recognition to my friends and family both near and far: MK Hunter, Sara Marin, Anna Harbom, Donald Schumacher Jr., Katie Harabin, and so many others. Without the home away from home provided by my sister, Rachael Yost, and brother-in-law, Bruce H. Voge III, I would have burned out years ago. You two always keep me laughing, even when I try my hardest to keep a straight face. Having a graduate student/bridezilla daughter is not easy, but

my mother Maureen Yost seems to have come out on the other side just fine. Her strong will pushed me to fight harder through every day. Thank you for always believing in me.

Lastly, to my husband Brian: we did it! This dissertation is as much yours as it is mine. You supported me with a level head for the last six years from bringing me dinner and keeping me company in lab late at night, to listening to my rants about failed experiments. I can't wait to see where our life together goes next. We are finally Dr. and Dr.!

DEDICATION

This thesis is dedicated to two immeasurably wonderful people who made this world a much brighter place and helped me become the woman I am today. First, to my incredibly kind and gentle father, William H. Yost, who told me at every opportunity how proud he was of me: I would not have been able to get through some very long, tough days without remembering your encouragement and love. Second, to my enthusiastic and driven cousin, Grace R. Mann: your excitement infectious and smile simply inspirational. Your positivity helped me believe in myself, and your steadfastness taught me to stand up for what is right. I am immensely grateful to have had these people supporting me and am greatly indebted to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. Positive Strand RNA Viruses

Viruses are infectious, biological machines which require a host in order to replicate. They are ubiquitous in nature, found to affect all kingdoms of life (1). In general, viruses are classified by: 1. the type of nucleic acid used to code their genome, 2. the structure of their genome, and 3. their method of genome replication. Viruses can code their genetic information within RNA or DNA, and in either single or double stranded form. Additionally, some viral genomes consist of several different RNA molecules versus a single one (i.e. segmented versus non-segmented). Some viruses, such as HIV, carry their genome as RNA, but use a DNA intermediate to replicate the information, while others replicate exclusively through RNA.

Positive-sense RNA viruses are classified in Group IV by the Baltimore classification system. This group maintains their genomes as a ssRNA that can be directly used as mRNA for translation of viral proteins. Group IV viruses do not use a DNA intermediate in their life cycle and replicate their genome via an RNA-dependent, RNA polymerase (RdRp). Positive strand RNA viruses are the largest classification of viruses; among this group are several pathogens of human importance which cause illnesses such as the “common cold”, SARS, Hepatitis A, C, and E, and Dengue fever.

2. Alphavirus Background

In 1931, the first alphavirus, western equine encephalitis virus, was described (2). Eighty-five years later, the genus alphavirus has grown to include thirty-one recognized members (3). Alphaviruses are enveloped, positive strand RNA viruses that are mainly mosquito-borne and capable of causing a variety of symptoms including rash, arthritis,

encephalitis, and even death in humans. Alphaviruses are classified by the NIH/NIAID as Category B bioterrorism agents due to their ease of dissemination and morbidity rates (4).

Large outbreaks of Chikungunya virus (CHIKV), a member of the alphavirus family, have been reported in parts of southeast Asia in 2005-2007 and in Europe in 2007 (5). Currently CHIKV is considered an emerging pathogen in the western hemisphere. Viral life cycles are typically maintained in a mosquito-bird-mosquito cycle except in urban areas where humans are a primary reservoir. Tropical mosquito populations which spread alphaviruses are thriving and expanding into more subtropical and temperate regions due to global climate change, causing major concern (6, 7). Within the past few years, CHIKV has been spreading rapidly in the Americas. In 2014 alone, there were over 25,000 confirmed and one million suspected cases of CHIKV in the Americas according to the Pan American Health Organization (8). To compound matters, infection rates and spread are likely severely underreported for three reasons. First, patients are often misdiagnosed with Dengue fever due to the similarity of common symptoms (9). Second, symptoms of CHIKV infection can be non-classical allowing many cases to go undiagnosed (10). Lastly, surveillance of mosquito populations is not maintained in many high-risk areas. A 2013 report in The Times of India states that 65% of mosquito surveillance posts in India are vacant and no surveillance is being done in these areas (11). As of 2014, it seems the issue had not yet been resolved as mosquito surveillance in at least twelve districts of India was not being performed (12).

There is no specific anti-viral treatment or vaccine against any member of the alphavirus genus. This is a clear motivating factor in alphavirus research. In addition to being a pathogen of interest due to their impact on human health, alphaviruses are heavily studied as vectors for delivering targeted gene therapy and anti-cancer treatments. Alphaviruses have a simple genome, are easy and safe to manipulate in the lab, infect a wide range of mammalian hosts,

and have been successfully used for expression of recombinant proteins in cell culture and *in vivo* making them ideal for use in gene therapy (13-15).

3. Alphavirus Virology

Alphaviruses are a diverse genus of enveloped arboviruses classified within the family *Togaviridae.* They are grouped into two main divisions called "Old World" and "New World" by their evolutionary divergence due to distribution in opposite hemispheres and further into seven "complexes" based on serological cross-reactivity (16, 17). These viruses span the globe and have been found on every continent except Antarctica. The origin of alphaviruses is still being studied based on phylogenetic analysis and protein structural similarity studies; however, it has been hypothesized that this group emerged in the southern Pacific Ocean (18).

3.a. Virus Entry

Alphavirus virions are coated in a lipid bilayer derived from host cell membrane. Within this membrane, viral envelope glycoproteins E1 and E2 are embedded as "spikes" composed of trimeric E1 and E2 heterodimers (19, 20). Natural resistance-associated macrophage protein (NRAMP) has been identified as the host cell receptor recognized by E2 for entry (21). Alphaviruses enter the cell via clathrin-mediated endocytosis and E1 is then responsible for membrane fusion (22, 23). Within the endosome, a change in environmental pH leads to conformational change in the envelope proteins. E2 is then freed and the trimerization of E1 exposes the fusion peptide ultimately leading to host-virus membrane fusion and release of the viral RNA (24, 25).

3.b. Genome and Replication

Alphaviruses contain a positive-sense, ssRNA genome of approximately 9-11kb with a 5' m7G(5')ppp(5')N (cap-0) structure and 3' polyadenosine tail (26). The genome contains two cistrons. The first, located in the 5' two-thirds of the genome, encodes the RNA replication machinery, termed nonstructural proteins (nsP1-4), while the structural proteins that form progeny virus particles are encoded in the second cistron (E1, E2, E3, 6K, and capsid).

After infection, the viral genomic RNA is used directly as mRNA for translation of the nsPs and as a template for the synthesis of a complementary negative-sense RNA. The negative-sense strand is a template for progeny, genomic RNA and for subgenomic RNA containing the cistron of structural genes. The nsPs responsible for viral RNA replication are produced as a one of two, large polyproteins (P123 or P1234) depending on the virus. A majority of alphaviruses produce P123 in higher proportion to P1234, as P1234 is only made by a read-through of an opal termination codon at the end of nsP3. Others, such as Semliki Forest virus (SFV), lack this opal termination codon and only make P1234 (27). Precursor polyproteins are cleaved by a protease domain within nsP2 in a highly regulated manner (28-30). Studies suggest that the RNA replication complex template usage (plus or minus strand RNA) is regulated by this proteolytic processing mechanism of the nonstructural proteins (Figure 1) (29, 31-39).

If nsP4 is present in the polyprotein, nsP2 first cleaves at the P3/4 junction either in *cis* or *trans*. Next, the P1/2 junction is cleaved in *cis* (31, 39). During the intermediate stage of cleavage when P123+nsP4 and nsP1+P23+nsP4 are present, negative strand genomic RNA is being produced. The final cleavage event at the P2/3 junction yielding fully mature nsPs signals a switch in template to the newly formed minus strands, thus switching synthesis to positive sense genomic and sub-genomic RNAs (34, 35). This strategy provides temporal regulation of viral RNA replication, providing that negative-sense RNA is preferentially produced early in

infection when replication complexes are in limited number, and positive-sense subgenomic and genomic RNA is made later when structural proteins are abundant for packaging of RNA into budding virions. The underlying mechanism of action behind RNA template recognition is unknown, as is what factors dictate the timing of polyprotein cleavage.

The structural proteins are translated from subgenomic RNA initially as a single polyprotein, similar to the nsPs. The structural polyprotein is cleaved by viral capsid protein and cellular proteases (signalase and furin) to yield the mature envelope and capsid proteins (16).

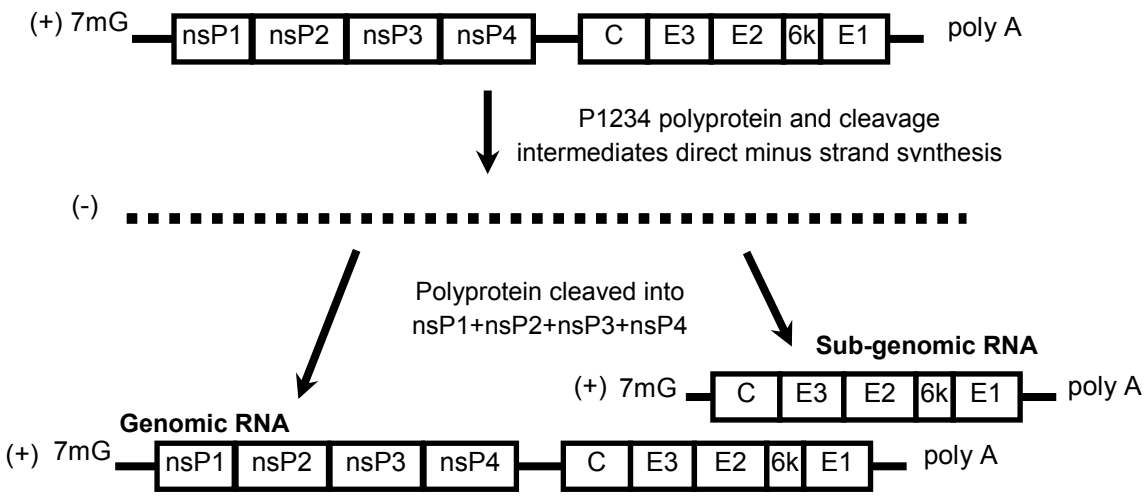

Figure 1. Alphavirus Nonstructural Polyprotein Processing in Relation to Viral RNA Synthesis. Plus strand viral RNA is used as a template to translate nonstructural polyprotein P1234. This precursor protein directs the synthesis of negative strand RNA. nsP2 cleaves the polyprotein, switching the template of RNA to produce plus strand RNA only after the final cleavage of P23.

4. Nonstructural Proteins

The alphavirus RNA replication complex is composed of the four nsPs; however, beyond the replication complex, the nsPs play a role in other important pathways including host immune evasion, and shut off of host transcription and translation. The nsPs serve multiple roles and they are each absolutely necessary in viral RNA replication. Previous to our studies,

only one viral nonstructural polyprotein structure had been described in the literature (40). Overall, very little is known about how the structures and activities of the nsPs change pre- and post-cleavage from the polyprotein.

4.a. nsP1

Alphavirus RNA replication complexes are associated with the cytoplasmic surface of modified endosome and lysosome membranes (41, 42). Palmitoylation of nsP1 and an amphipathic helix within the protein are necessary for tight association of the replication complex with the membrane (43, 44). nsP1 may play a role in minus strand synthesis based on previously described temperature sensitive viral mutants; however, the main function of nsP1 within the replication complex is RNA capping (45-47). Subgenomic and genomic viral RNAs of alphaviruses have a cap on their 5' end in order to mimic cellular mRNAs, allowing for translation of viral proteins via the ribosome, and protecting from RNase degradation. Cellular mRNAs are generally capped in the nucleus by: 1. removing the 5' γ-phosphate from the RNA molecule, 2. guanylyltransferase and GTP forming a covalent guanylyltransferase-GMP complex, 3. transferring GMP to the 5' end of the RNA, and finally 4. methylating by nucleoside-2'-*O*-methyltransferase. Alphavirus nsP1 uses a unique method for RNA capping in the cytoplasm and first acts as a methyltransferase to transfer a methyl group from S-adenosylmethionine to GTP, then forms a covalent complex with the newly formed m^7GMP (47-50). The triphosphatase activity necessary to remove the 5' γ-phosphate from the viral RNA before nsP1 transfers the m^7GMP is actually found in nsP2 (26).

4.b. nsP2

The multi-tasking nsP2 of alphaviruses serves many roles during infection. As described above, the centrally located nsP2 protease domain is responsible for all of the nonstructural polyprotein cleavage events (Figure 2). The very N-terminal portion of nsP2 is also required as a "cofactor" for efficient cleavage (39, 51). The protease domains of several alphaviruses have been characterized as papain-like, with a cysteine-histidine catalytic dyad; however, a 2015 study by Saisawang *et al.* shows that this is not the case for CHIKV nsP2 protease (29-31, 52-57). The role of the cysteine residue of the active site in CHIKV nsP2 protease can also be performed by an adjacent serine (56). The CHIKV protease is also unique compared to Sindbis virus (SINV) or SFV in that it can recognize small peptide substrates (58). CHIKV protease is shown to recognize and cleave peptides as small as nine amino acids in length whereas SINV and SFV protease require at least the first 170 amino acids of nsP3 to cleave at the P2/3 junction (51, 58). This indicates some differences in protease substrate specificity and cleavage requirements within different alphavirus species. Additionally, nsP2 proteases are not always interchangeable between species; the SINV protease is able to process SFV polyprotein, but not vice versa (51). The factors contributing to substrate specificity are not fully understood for alphaviruses.

nsP2 also contains a helicase/NTPase domain and a non-functional methyltransferase-like (MT-like) domain. The N-terminal helicase/NTPase domain is theorized to unwind viral RNA secondary structures during replication. Helicase activity of nsP2 is driven by the NTPase activity of two RecA-like domains (59, 60). CHIKV nsP2 cannot unwind dsDNA substrates or dsRNA substrates with a 3' overhang, blunt ends, or a short 5' overhang. Full length CHIKV nsP2 is, however, able to unwind dsRNA with a longer, 12-base 5' overhang (61). Truncated versions of nsP2 that do not include the MT-like domain lack helicase activity, making this seemingly nonfunctional domain synergistically important to helicase function (61, 62).

nsP2 plays a vital role in pathogenicity and host immune evasion. About 50% of mature nsP2 localizes to the nucleus through two localization signals and a nucleolus-targeting signal (63-65). In Old World alphaviruses such as SINV and SFV, mature nsP2 causes host transcriptional shutoff (17, 66). nsP2 induces degradation of a catalytic subunit of the RNA polymerase II (RNAPII) complex, Rpb1, via induction of ubiquitination. This causes a halt in host transcription, preventing essential anti-viral genes from being made in response to infection. NTPase activity of nsP2 is required for the rapid degradation of Rbp1. Researchers suggest that nsP2 may be able to bind dsDNA, stall the RNAPII complex and induce ubiquitination of Rpb1 via the transcription-coupled repair pathway (67). Previously described C-terminal domain mutations which cluster on the surface of nsP2 are also essential for degradation of Rpb1 and cytopathogenicity in mammalian cells (68-73).

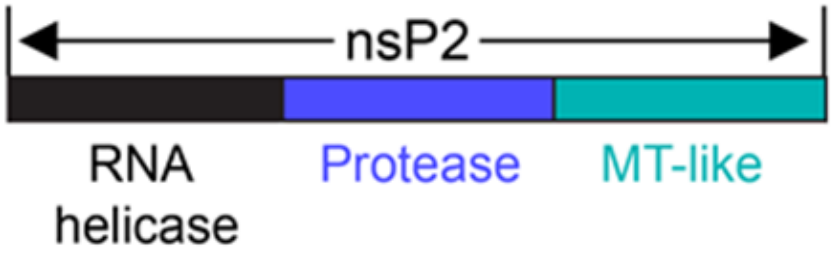

Figure 2. Domain Organization of nsP2. Schematic of the domain organization of nsP2 showing the N-terminal helicase domain (black), central protease domain (blue), and C-terminal methyltransferase-like (MT-like) domain (cyan).

4.c. nsP3

The functions of alphavirus nsP3 are still unclear. Mutational studies have shown a role for nsP3 in minus-strand synthesis and subgenomic RNA synthesis, but the mechanism is still not understood (37, 74, 75). A 2013 study found that nsP3 may be responsible for vector specificity. O'nyong nyong virus (ONNV) is a unique alphavirus in that it is able to infect *Anopheles gambiae*

mosquitoes. ONNV is the only alphavirus known to be transmitted by these mosquitoes. However, when the CHIKV nsP3 protein was switched to ONNV nsP3, infection rates in *An. Gambiae* jumped from 0% to a surprising 63.5% (76).

nsP3 is predicted to have three domains: an N-terminal macrodomain (or X domain), a central alphavirus unique domain (AUD), and a C-terminal phosphorylated hypervariable carboxyl terminus (Figure 3). Residues necessary for neurovirulence have been found in all three domains of nsP3 (77-79). The centrally located AUD is highly conserved among alphavirus species, but its structure cannot be predicted due to lack of sequence similarity to proteins of other organisms.

Macrodomains are highly conserved and found across all domains of life as well as in viruses such as rubella virus, hepatitis E virus, and coronaviruses (80-83). Human macrodomain containing proteins mainly localize to the nucleus and participate in transcriptional regulation via alteration of chromatin structure (84-87). Although its function in viral replication is unknown, evidence shows that residues N10 and N24 of SINV nsP3 macrodomain are necessary for neurovirulence in mice (81). Some alphavirus macrodomains have been shown to bind DNA, RNA, and polyADP-ribose with varying affinities and have relatively poor ADP-ribose-1'' phosphatase activity compared to yeast and mammalian macrodomains (83, 88, 89). As mentioned above, in SINV and SFV, the macrodomain of nsP3 is also necessary for efficient polyprotein cleavage through proper positioning of the P2/3 cleavage site (51, 58).

The C-terminal, hypervariable region (HVR) of nsP3 is not well conserved among alphaviruses. It is predicted to be a region of high flexibility and likely unstructured (74). The HVR is phosphorylated by yet known cellular kinases on serine and threonine residues within a span of fifty amino acids near the beginning of the HVR in SFV (90-93). The phosphorylation sites are not completely mapped for each alphavirus, as this process is still quite challenging due

to the high number of phosphorylation sites in close proximity to one another. In SFV nsP3, between eleven and sixteen threonine and serine residues are phosphorylated (93). Phosphorylation levels of this protein have been shown to change over the course of alphavirus infection (91, 92). Mutation of sites within this domain often show little or no effect in highly permissive tissue culture models such as BHK-21 cells; however, in mice, virulence is diminished or abolished (78, 90, 94-96). Studies in mosquito cells indicate a role for this domain in a species-specific manner, meaning this domain may be involved in virus-host interaction based on specific host kinases (97).

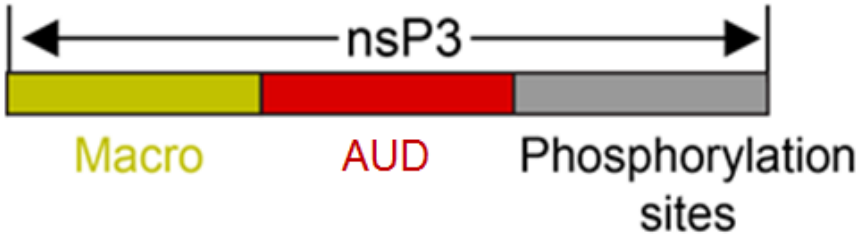

Figure 3. Domain Organization of nsP3. Schematic of the domain organization of nsP3 showing the N-terminal macrodomain (yellow), central AUD (red), and C-terminal hypervariable, phosphorylated region (gray).

4.d. nsP4

nsP4 is the alphavirus RdRp responsible for RNA synthesis. Cleavage of nsP4 is necessary for polymerase activity (34). The variable N-terminal region of nsP4 is unique to alphaviruses and predicted to be disordered. This domain may be necessary for association with other alphavirus proteins, but this theory is currently only based on evidence obtained by second site mutation (98). Although there is no published structure for this protein, the remaining portion of nsP4 is predicted to be a classical RdRp with a GDD active site within a palm, surrounded by fingers and thumb domains (99, 100).

5. Rationale

The polyprotein translation strategy is common across many viruses families, including *Flaviviridae*, *Picornaviridae*, and *Retroviridae* representing several important human pathogens (101). Encoding and translating large viral polyprotein precursors is a strategy with many benefits. It is evolutionarily beneficial for viruses to have a compact genome size. By encoding their genes in tandem, viruses can eliminate features such as promoter or enhancer elements necessary for individual protein expression. As stated above with regard to the nsPs of alphaviruses, cleavage site usage can be coordinated with regulation of protein activity (i.e. turn function "on" or "off"). Proteins can also perform alternative functions in their mature versus precleavage forms—function "A" or function "B" (102).

Despite their presence in many viral families, there are only a handful of structures available of precleavage polyprotein intermediates, and only one for a viral nonstructural polyprotein previous to our study (40). Structural information about the rearrangement of viral proteins before and after cleavage is essential to understanding of the regulation, function, and progression of viral pathogenesis. The RNA template switching mechanism of alphavirus nsPs is linked to polyprotein cleavage progression, temporally regulating the steps of viral pathogenesis. The "point of no return" in this process is the cleavage step between P2/3, but the mechanism behind this is poorly understood. Furthermore, of the nonstructural proteins, nsP3 itself is the least understood. Perhaps our lack of knowledge of nsP3 stems from a lack of studies within polyprotein P23, if the main functions of nsP3 are actually carried out within the polyprotein, and not after cleavage. With this in mind, **our goal is to describe alphavirus nsPs with regard to their structure and activities pre- and post-cleavage via biophysical, biochemical, and virological methods in order to better understand viral pathogenesis and present opportunities for novel anti-viral therapy design.**

MATERIALS AND METHODS

Section I. Sindbis Virus P23$^{pro\text{-}zbd}$ Polyprotein

1. SINV P23$^{pro\text{-}zbd}$ Expression and Purification

P23$^{pro\text{-}zbd}$ of the Sindbis virus strain Toto1101, encompassing amino acids 1011-1675 (residues 471-807 of nsP2 and 1-328 of nsP3) of the P1234 polyprotein, were expressed overnight at 18°C in *Escherichia coli* as a fusion with glutathione S-transferase (GST). After lysis, the soluble fraction was loaded onto a GSTrap (GE healthcare) column in a low salt buffer [20 mM HEPES (pH 7.5), 150 mM KCl, and 5% glycerol]. After equilibration in the low salt buffer, the column was extensively washed with 20 mM HEPES (pH 7.5), 1 M KCl, and 5% glycerol to remove bound RNA. The fusion protein was eluted with 0.1 M Tris-HCl (pH 8), 150 mM KCl, 15 mM glutathione, and 5% glycerol. The recovered protein was then digested overnight with Prescission Protease (GE healthcare) to remove the GST tag. The digested material was purified by hydroxyapatite column (CHT-II, Bio-Rad) using a linear gradient of increasing phosphate concentration (5-200 mM) (pH 7.5) with 150 mM KCL and 5% glycerol. The protein containing fractions were pooled, desalted into the low salt buffer and purified by heparin sepharose column (GE Healthcare) following the manufacturer's protocol.

2. Protein Crystallography

The purified protein was pooled, buffer exchanged into 20 mM HEPES (pH 7.5), 150 mM KCl, 5% glycerol and 1 mM Tris(2-carboxyethyl)phosphine (TCEP) and concentrated to 15 mg/mL. Crystals of P23$^{pro\text{-}zbd}$ were grown by the vapor diffusion method against 2 M ammonium sulfate and 0.1 M 2-(N-morpholino)ethane sulfonic acid (MES) (pH 6.5). Crystals were transferred to a cryoprotectant solution containing 2 M ammonium sulfate, 0.1 M MES (pH 6.5), 35% xylitol and 5% trimethylamine N-oxide and then flash cooled in liquid nitrogen. Diffraction

data were collected at 1.2824 Å and 100 Kelvin to 2.85 Å resolution at beam line X25 at Brookhaven National Synchrotron Light Source (Table 1). The data were integrated and scaled with MOSFLM and Scala (103, 104). Phases were determined by molecular replacement using the $nsP2^{pro}$ (53) and nsP3 macro structures (88). The location of the endogenous zinc ion was determined by SHELX (105) using anomalous difference maps. The model was improved with several rounds of manual building using Coot (106) and Phenix (107). The final model contains three molecules of $P23^{pro\text{-}zbd}$, three zinc ions, 31 sulfate ions and five molecules of MES with R_{work}/R_{free} factors of 0.220/0.251. MOLPROBITY (108) revealed no unfavorable (φ, ψ) combinations with main-chain and side-chain parameters consistently better than or within the average for structures refined to 2.85 Å. Graphics were generated using the program PyMOL (109). APBS (110) was used for calculating surface potentials.

3. Mammalian Cell Culture

BHK-J cells (provided by Charles Rice, Rockefeller University) were maintained in Minimum Essential Medium (MEM) (Life Technologies) supplemented with 7.5% fetal bovine serum (FBS) at 37°C and 5% CO_2.

4. Plasmid Constructs

Mutations were introduced into full-length SINV cDNA clone pToto1106 (74) via PCR site-directed mutagenesis. To minimize the possibility of reversions, mutated codons were designed to introduce at least two nucleotide changes away from the wild type codon when possible.

5. RNA Transcription

In preparation for transcription, mutant SINV cDNA clones were linearized with Xho1, purified via standard phenol/chloroform extraction, and ethanol precipitated. RNA was transcribed *in vitro* using the mMessage mMachine SP6 kit (Life Technologies) and purified using the RNeasy Mini Kit (Qiagen). Non-denaturing gel electrophoresis was used to verify RNA integrity. Aliquots of 1µg were stored at -80°C until transfection.

6. RNA Transfection

BHK-J cells were harvested and centrifuged at 1200 rpm for 5 minutes at 4°C. Cell pellets were washed twice with 1x DPBS (Life Technologies), counted by trypan blue exclusion, and resuspended in 1x DPBS to $2x10^7$ cells/mL. 400µL of the cell suspension ($8x10^6$ cells) was mixed with 1µg RNA and immediately placed in a 2-mm gap cuvette for electroporation. Samples were subjected to five, 99 µs pulses at 550 V in an ECM 830 Electro Square Porator (Harvard Apparatus) and allowed to rest ten minutes at room temperature. 10-fold serial dilutions of electroporated cells were made in MEM containing 1% FBS for infectious center assay. Remaining cells were resuspended in MEM containing 7.5% FBS and plated in 150 mm dishes. Virus was harvested 24 hours after electroporation and aliquoted for storage at -80°C.

7. Infectious Center Assay

BHK-J subconfluent monolayers were prepared in 6-well plates at least 12 hours in advance at $5x10^5$ cells per well for infectious center assay. 10-fold serial dilutions of electroporated cells were seeded on these monolayers and incubated for 1 hour at 37°C before being overlaid with 1.2% agarose LE containing MEM, 2% FBS, and 1x pen/strep. After 48 hours of incubation at 37°C, cells were fixed with 7% formaldehyde and stained with crystal violet.

8. Plaque Assay

Serial dilutions of harvested virus were made in PBS with 1% FBS and used to infect BHK-J monolayers. Hereafter, this assay was performed similarly to the infectious center assay.

Section II. Mutational analysis of Sindbis Virus Polyprotein P23 Interface

1. P2/3 Interface Mutant Generation

pToto110 cDNA constructs were cloned as described in Materials and Methods Section I. cDNAs were transcribed and transfected into BHK-J cells as described Materials and Methods Section I.

2. Infectious Center Assay and Plaque Assay

Infectious center assays to measure RNA infectivity and plaque assays to measure viral titer were performed as described in Materials and Methods Section I.

3. TSG/PAC Cytopathogenicity Selection

Mutations were introduced into the pTSG/PAC replicon (provided by Charles Rice, Rockefeller University) via traditional cloning methods. Fragments from the previously made pToto1106 containing the mutated bases were cloned into the TSG/PAC background and sequenced for verification. In preparation for transcription, mutant SINV cDNA clones were linearized with Xho1, purified via standard phenol/chloroform extraction, and ethanol precipitated. RNA was transcribed *in vitro* using the mMessage mMachine SP6 kit (Life Technologies) and purified using the RNeasy Mini Kit (Qiagen). Non-denaturing gel electrophoresis was used to verify RNA integrity. Aliquots of 1µg were stored at -80°C until transfection.

1 mg/mL puromycin stocks were made in MEM and sterile filtered before use. Wild type SINV and a mutant unable to process nonstructural polyproteins [T*SG/PAC (69)] were used as controls. Transfection followed the protocol indicated in Materials and Methods Section I and cells were allowed to recover for five hours before addition of puromycin at 5

µg/mL. Medium was replaced every other day with fresh puromycin-MEM and 7.5% FBS to select for foci. After eight days of growth, cells were washed with DPBS, fixed with 7% formaldehyde, and stained with crystal violet to visualize foci.

4. Protein Synthesis during SINV Infection

BHK-J cells were prepared in 6-well plates at $5x10^5$ cells per well and infected with virus at a multiplicity of infection (MOI) of 10 PFU per cell in MEM supplemented with 1% FBS for one hour at 37°C. Media was then replaced and cells were incubated until specified time points. At 2, 3, 6, 9, and 12 hours post-infection, cells were washed three times with PBS and media replaced with 1mL DMEM (lacking methionine and cysteine) with 1% FBS and 20µCi of [^{35}S] in EasyTag EXPRE^{35}S^{35}S Protein Labeling Mix (Perkin Elmer). After one hour, cells were washed three times with PBS and lysed with 200µL SDS-PAGE loading buffer. 10µL samples were run on 10% SDS-PAGE gels, stained with Coomassie Blue, dried and autoradiographed.

5. RNA Synthesis during SINV Infection

BHK-J cells were prepared in 6-well plates at $5x10^5$ cells per well and infected with virus at a MOI of 10 PFU per cell in MEM supplemented with 1% FBS for one hour at 37°C. Media was then replaced and cells were incubated until specified time points. At 2, 4, 8, and 12 hours post-infection, media was replaced with 0.5 mL MEM with 7.5% FBS and 30 µCi of [5,6-^{3}H] uridine (Perkin Elmer). After three hours, cells were washed three times with PBS and lysed with 300 µL TRIzol LS Reagent (Invitrogen). Samples were frozen at -80°C until RNA extraction following TRIzol protocol and resuspending in a final volume of 20 µL water.

6. Denaturing RNA Gel Electrophoresis

5 µL [5,6-^{3}H] uridine labeled RNA samples were denatured with 25 µL glyoxal reaction mixture (6 mL DMSO, 2 mL deionized glyoxal, 1.2 mL 10x BPTE electrophoresis buffer, 0.6 mL 80% glycerol, and 0.2 mL water) incubated for 1 hour at 55°C. Before loading, samples were chilled for 10 minutes on ice and mixed with 2 µL RNA loading buffer. Denatured RNA was loaded on a 0.8% agarose LE 1x BTPE gel and run in 1x BPTE buffer (10 mM PIPES, 30 mM BIS-TRIS, 1 mM EDTA) at 5 V/cm. Ethidium bromide staining was used to assess RNA integrity before gels were dried and autoradiographed.

7. Immunization and Antibody Generation

nsP2-protease domain was produced and purified via the method in Section I similar to the P23$^{pro\text{-}zbd}$ protein. BALB/c mice were injected intraperitoneally with 50 µg of nsP2-protease protein emulsified in complete Freund's adjuvant. Two additional immunizations were carried out every 14 days with the same dosage in incomplete Freund's adjuvant. A booster injection was given intravenously 3 days prior to cell fusion. Splenocytes were harvested and mixed with myeloma cells (P3Ag8.6.5.3) at a ratio of 4:1 before fusion in polyethylene glycol 1500 (Roche). The fused cells were re-suspended in HAT IMDM medium (HyClone) plus 10% fetal bovine serum (HyClone). Peritoneal cells from unimmunized BALB/c mice were added as feeder cells. The resulting cell suspension was seeded in 96 well plates and supernatants of proliferating cells were screened by ELISA and western blot.

8. Western Blotting

BHK-J cells were prepared in 6-well plates at 5x10^{5} cells per well and infected with virus at a multiplicity of infection (MOI) of 10 PFU per cell in MEM supplemented with 1% FBS for one

hour at 37°C. Media was then replaced with 1 mL MEM with 7.5% FBS. At 2, 4, 6, 9, and 12 hours post-infection, media was removed and cells lysed with 200 µL standard loading buffer. 12 µL samples were separated on 8% SDS-page gels and transferred to nitrocellulose membranes using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) in 25 mM Tris, 192 mM glycine, pH 8.3, 0.025% SDS, and 10% methanol for 2 hours at 0.3 Amp. Membranes were blocked for 1 hour at room temperature in 5% skim milk in 0.05% Tween-20 in PBS and incubated with primary antibody overnight at 4°C. Anti-nsP2-protease (2A11) and anti-nsP3 antibodies [WU136; (111)] were used at 1:1000 dilution. GAPDH served as a loading control and was detected with purified anti-GAPDH antibody (Millipore) at a dilution of 1:10,000.

9. Expression of P23$^{pro\text{-}zbd}$ Polyprotein with Interface Mutations

SINV P23$^{pro\text{-}zbd}$ with interface mutations was cloned into the pGEX-6P-1 expression construct. Expression of GST-P23$^{pro\text{-}zbd}$ with interface mutations was performed as described for wild type P23 in Materials and Methods Section I.

10. Hydrogen-Deuterium Exchange (HDX)

SINV P23$^{pro\text{-}zbd}$ proteins were diluted with D_2O buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol) and incubated for 10, 30, 60, 300, 900, and 3600s on ice. At the indicated times, the resulting mixtures were quenched by addition of ice-cold quench solution (100 mM TCEP, pH 2.4), immediately frozen on dry ice, and stored at −80 °C. Non-deuterated control was prepared similarly in H_2O buffer (50 mM 29 HEPES, pH 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol). Samples were injected into an immobilized pepsin column, followed by 0.05% trifluoroacetic acid. After pepsin digestion, peptide fragments were separated by an online C18

HPLC column (Supelco, Sigma-Aldrich) using a linear acetonitrile gradient of 2−50%. Eluate was then analyzed by mass spectrometry (MS).

Section III. Chikungunya Virus P23$^{pro-zbd}$ Structure

1. CHIKV P23$^{pro-zbd}$ Expression and Purification

CHIKV nonstructural polyprotein DNA constructs were obtained from Dr. Craig Cameron, Pennsylvania State University. P23$^{pro-zbd}$ of Chikungunya virus was expressed overnight at 18°C in *Escherichia coli* as a N-terminal fusion with six-histidine and small ubiquitin-like modifier (SUMO) tags. After lysis, the soluble fraction was loaded onto a HisTrap (GE healthcare) column in a low imidazole buffer [50 mM sodium phosphate (pH 8.0), 150 mM KCl, 25 mM imidazole, and 5% glycerol]. After equilibration in the low imidazole buffer, the column was extensively washed with 50 mM sodium phosphate (pH 8.0), 1 M KCl, and 5% glycerol to remove bound RNA. The fusion protein was eluted with an increasing gradient to buffer B (50 mM sodium phosphate (pH 8.0), 150 mM KCl, 250 mM imidazole, and 5% glycerol). The recovered protein was then digested overnight with Ubiquitin-like-specific protease 1 (ULP1) to remove the six-histidine-SUMO tag. The digested material was purified by hydroxyapatite column (CHT-II, Bio-Rad) using a linear gradient of increasing phosphate concentration (5-225 mM) (pH 7.5) with 150 mM KCL and 5% glycerol. The protein containing fractions were pooled, desalted into the low salt buffer and purified by heparin sepharose column (GE Healthcare) following the manufacturer's protocol.

2. Protein Crystallography

The purified protein was pooled, buffer exchanged into 20 mM HEPES (pH 7.5), 150 mM KCl, 5% glycerol and 1 mM TCEP and concentrated to 8 mg/mL. Crystals of P23$^{pro-zbd}$ were grown by the vapor diffusion method against 0.1 M 2,2-Bis(hydroxymethyl)-2,2',2''-nitrilotriethanol (BIS-TRIS) pH 6.5, 1% Tacsimate pH 6.0 (Hampton Research) and 20% polyethylene glycol 3350. Initially obtained crystal stacks were broken up with a cat whisker and used to streak seed new

drops to promote single crystal growth. Crystals were transferred to a cryoprotectant solution containing 17% 2-methyl-2,4-pentanediol and then flash cooled in liquid nitrogen. Diffraction data were collected at 1.264 Å and 100 Kelvin to 2.20 Å resolution at beam line 12-2 at the Stanford Synchrotron Radiation Laboratory (Table 3). The data were integrated and scaled with MOSFLM and Scala (103, 104). Phases were determined by molecular replacement using the SINV $P23^{pro\text{-}zbd}$ structure (68). The model was improved with several rounds of manual building using Coot (106) and Phenix (107). The final model contains four molecules of $P23^{pro\text{-}zbd}$, four zinc ions, and 15 glycerol molecules with R_{work}/R_{free} factors of 0.19/0.24. MOLPROBITY (108) revealed no unfavorable (φ, ψ) combinations with main-chain and side-chain parameters in the top percentile when compared to all other structures. Graphics were generated using the program PyMOL (109). APBS (110) was used for calculating surface potentials.

3. Ligand Docking with CHIKV $P23^{pro\text{-}zbd}$

In silico docking experiments were performed by Dr. Stefano Forli at The Scripps Research Institute. The protein solvent-accessible surface of P23 was calculated using the program MSMS (112). Cysteine residues in contact with the surface were selected for reactive docking calculations. A grid box of approximately 25 Å per side was centered on each cysteine and a small library of reactive fragments were docked into the space using AutoDock (113). Results were filtered to identify ligands positioned with the reactive groups near the cysteine side chains and corresponding docking energy was used to rank the cysteine residues ligand binding potential.

Section IV. Biophysical and Virological Studies of nsP3 Macrodomain

1. Mammalian and Insect Cell Culture

BHK-J cells (provided by Charles Rice, Rockefeller University) were maintained in Minimum Essential Medium (MEM) (Life Technologies) supplemented with 7.5% fetal bovine serum (FBS) at 37°C and 5% CO_2. L929 cells were maintained in MEM and 10% equine serum, 1X non-essential amino acids, and 1 mM sodium pyruvate. C7-10 and C6/36 mosquito cell lines (provided by Victor Stollar, Rutgers University) were maintained in MEM supplemented with 10% FBS and 1X non-essential amino acids.

2. nsP3 Macrodomain Mutant Generation

pToto110 cDNA constructs were cloned as described in Materials and Methods Section I. cDNAs were transcribed and transfected into BHK-J cells as described Materials and Methods Section I. Viral titers were measured by plaque assay performed as described in Materials and Methods Section I.

3. Western Blotting

Western blotting was performed as described in Materials and Methods Section I.

4. SINV pTE/Luc Reporter Construct

Infectious SINV cDNA pTE/5'2J (provided by Charles Rice, Rockefeller University) contains a second subgenomic promoter between the nonstructural and structural ORFs (114-116). A firefly luciferase gene was synthesized (Genscript) and cloned into pTE/5'2J with XbaI and Bsp120 to create the pTE/Luc construct. pToto1106 constructs with macrodomain mutations were made with site-directed mutagenesis initially. These were used as a template

for cloning into the pTE/Luc vector. *In vitro* transcription, electroporation, and plaque assay were performed as described in Materials and Methods Section I.

5. Luciferase Assay

Cells were plated in 96 well plates at $2x10^4$ cells per well at least 12 hours before infection. Cells were infected with pTE/Luc wild type and mutant viruses at an MOI of 10. Mock infected cells and pTE/5'2J wild type infected cells were used as negative controls. At time points post infection, media was removed and cells were lysed with 100 µL 1X Glo Lysis Buffer (Promega) for five minutes before transferring to 96-well opaque, black plates. 100 µL reconstituted Bright-Glo Luciferase Assay Reagent (Promega) was added and luminescence read immediately on a SpectraMax M3 (Molecular Devices) at 500 ms integration time.

6. Purification of nsP3 Macrodomain

nsP3 macrodomain of the Sindbis virus strain Toto1101, encompassing amino acids 1-175 of nsP3 were expressed overnight at 18°C in *Escherichia coli* as a fusion with glutathione S-transferase (GST). After lysis, the soluble fraction was loaded onto a GSTrap (GE healthcare) column in a low salt buffer [20 mM HEPES (pH 7.5), 150 mM KCl, and 5% glycerol]. After equilibration in the low salt buffer, the column was extensively washed with 20 mM HEPES (pH 7.5), 1 M KCl, and 5% glycerol to remove bound RNA. The fusion protein was eluted with 0.1 M Tris-HCl (pH 8), 150 mM KCl, 15 mM glutathione, and 5% glycerol. The recovered protein was then digested overnight with Prescission Protease (GE healthcare) to remove the GST tag. The digested material was further purified via heparin sepharose column (GE Healthcare) following the manufacturer's protocol.

7. Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a Microcal iTC_{200}. Experiments were carried out at 20°C in 10 mM HEPES pH 7.5 and 150 mM NaCl. Macrodomain protein in the cell was 60 μM, and ADP-ribose concentration in the syringe was 1000 μM. Heats of dilution were measured by injecting the ligand into the protein solution. Background heat of dilution curve was subtracted from the experimental curves prior to data analysis. Titration curves were analyzed with MicroCal Origin software, assuming one set of sites.

Section V. Full Length P23 Polyproteins

1. Expression and Purification of Full Length P23 Polyproteins

Full length P23 of SINV, CHIKV, and VEEV encompassing the entirety of nsP2 and the macrodomain and ZBD of nsP3, were expressed overnight at 13°C in *Escherichia coli* as a N-terminal fusion with six-histidine and small ubiquitin-like modifier (SUMO) tags. SINV nonstructural polyprotein constructs were codon optimized and obtained from Dr. Craig Cameron, Pennsylvania State University. After lysis, the soluble fraction was loaded onto a HisTrap (GE healthcare) column in a low imidazole buffer [50 mM sodium phosphate (pH 8.0), 150 mM KCl, 25 mM imidazole, and 5% glycerol]. The fusion protein was eluted with an increasing gradient to buffer B (50 mM sodium phosphate (pH 8.0), 150 mM KCl, 250 mM imidazole, and 5% glycerol). The recovered protein was then digested overnight with Ubiquitin-like-specific protease 1 (ULP1) to remove the six-histidine-SUMO tag. The digested material of SINV and CHIKV was purified by hydroxyapatite column (CHT-II, Bio-Rad) using a linear gradient of increasing phosphate concentration (5-200 mM) (pH 7.5) with 150 mM KCL and 5% glycerol. The protein containing fractions were pooled, desalted into the low salt buffer and all three species were finally purified by heparin sepharose column (GE Healthcare) following the manufacturer's protocol.

2. CHIKV P23 Cleavage Assay

CHIKV FL nsP2 was incubated with CHIKV FL P23 or P23pro-zbd at a 1:1 molar ratio at room temperature. Time points were taken at 30m, 1h, 2h, 4h, and overnight and mixed with SDS loading buffer. Samples were boiled for five minutes at 95˚C before storage at -20˚C.

3. ATP Hydrolysis Assay

The ATP hydrolysis assays were performed in 1X Buffer-A [50 mM MOPS-Na (pH 7.4), 5 mM MgCl2, 5 mM DTT, 0.01% Tween 20] at 37°C. The ATP hydrolysis time course (0-30 min) was measured using 5nM or 10 nM protein, 1 mM ATP spiked with [γ-32P]ATP and RNA/DNA substrates (1 μM). The reactions were stopped at desired time points using 4N formic acid (HCOOH) and analyzed by PEI-Cellulose-F TLC (Merck) developed in 0.4 M potassium phosphate buffer (pH 3.4). The TLC plates were exposed to a phosphor-imager screen, which was imaged on a Typhoon phosphor-imager and quantified using ImageQuant software. The ATPase turnover rate was determined from the plots of [Pi] produced versus time and dividing the ATPase turnover rate by the respective enzyme concentration.

4. Small Angle X-Ray Scattering

Small angle X-ray scattering (SAXS) data were collected at the CHESS beamline G1 using a Finger Lakes CCD X-ray detector system with a sample-to-detector distance 1450mm to make scattering vectors q range from 0.007 to 0.7 $Å^{-1}$, where $q = 4\pi\sin\theta/\lambda$ (2θ is the scattering angle and $\lambda = 1.244$ Å). Samples were gel filtered into 50 mM phosphate pH 7.5, 500 mM KCl, and 2% glycerol via Superdex200 column (GE Healthcare) and flow through buffer from protein concentration was used in background subtraction. ATSAS was used to process and evaluate scattering data (117). Radius of gyration (Rg) was analyzed using the Guinier approximation with low angle data ($q < 1.3/Rg$). The probability distribution of distances between scattering atoms within the macromolecule, P(r), and the maximum atom pair distance, D_{max}, were determined from the scattering data using the GNOM algorithm.

EXPERIMENTAL RESULTS

Section I. Sindbis Virus P23$^{pro\text{-}zbd}$ Polyprotein

Prior to this study, only one viral nonstructural polyprotein structure was available in the literature. Poliovirus is a non-enveloped virus with a positive-sense RNA genome (Figure 4A). It expresses its viral proteins as one large polyprotein encoded by a single open reading frame (ORF) (118-120). Cleavage of the precursor polyprotein is performed by one of three viral proteins: mature 2A, mature 3C, or 3CD polyprotein (Figure 4A). The 3CD polyprotein is made up of 3C (protease) and 3D (RNA polymerase) and its structure resembles a beads-on-a-string model with the two proteins connected by a flexible, solvent exposed linker containing the accessible cleavage site (Figure 4B) (40).

Poliovirus uses the polyprotein encoding strategy to its benefit in several ways. First, 3C protease activity is regulated by its cleavage state. Mature 3C protease recognizes substrates preferentially; 3CD cleaves viral capsid precursor polyprotein substrates at a higher rate compared to mature 3C (121, 122). Additionally, the 3D polymerase activity is altered pre- and post- cleavage; 3D polymerase is inhibited while in the pre-cleavage 3CD polyprotein (123). Polyprotein 3CD is also solely responsible for the redistribution of cellular ADP-ribosylation factor proteins involved in membrane trafficking, which in turn promotes RNA synthesis (124, 125). Polyprotein processing represents an essential mechanism of gene regulation for poliovirus.

In order to further understand strategies used by alphaviruses for regulating protein function, structural studies are essential. For these studies, we chose to focus our efforts on polyprotein P23 for several reasons: 1. P23 is the last cleavage performed of P1234 and signals a “point of no return” template switching event to minus strand RNA, the mechanism of which is unknown, 2. the overall function of nsP3 is undefined and structural data could provide insight,

and 3. structures of nsP2 protease domain and nsP3 macrodomain provide a starting point for construct design. Our SINV P23$^{pro\text{-}zbd}$ structure was published in the *Proceedings of the National Academy of Sciences* in 2012 (68).

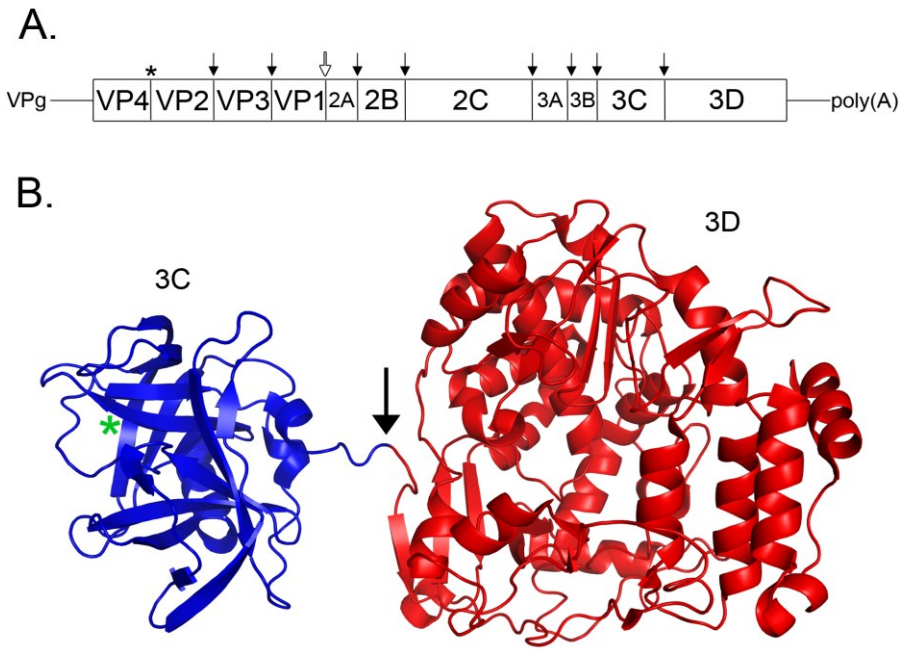

Figure 4. Poliovirus Genome and Polyprotein 3CD Structure. (A) RNA genome schematic of Poliovirus. Polyprotein cleavage sites are denoted by an asterisk (autoprotease), solid black arrow (3C or 3CD), or a hollow arrow (2A protease). (B) Structure of 3CD (PDB ID: 2IJD). A green asterisk shows the 3C protease active site and an arrow shows the cleavage site. Figure adapted from (102).

1. SINV P23$^{pro\text{-}zbd}$ Expression and Purification

A portion of the SINV P23 precursor protein from the nsP2 protease through the nsP3 AUD was expressed in bacteria (termed P23$^{pro\text{-}zbd}$; Figure 5) with an N-terminal GST tag. This construct was initially designed without the nsP2 helicase or nsP3 HVR to eliminate predicted areas of high flexibility not conducive to structural studies. P23$^{pro\text{-}zbd}$ is purified initially via GST affinity column and is shown by SDS-PAGE to be fully intact protein despite containing a

competent nsP2 protease (Figure 6A-B). Further purity of the protein is achieved via hydroxyapatite and heparin chromatography. $P23^{pro\text{-}zbd}$ also co-purifies with another molecule which elutes in high salt conditions (1M) from the GST column (Figure 6A). Based on the SDS-PAGE analysis, the high salt wash does not contain coomassie-stainable protein (Figure 6B). However, on an agarose gel, a high concentration of nucleic acid appears to be present in the fraction eluted with high salt buffer. Digestion with DNaseI or RNase A verified that bacterial RNA co-purifies with $P23^{pro\text{-}zbd}$ (Figure 6C).

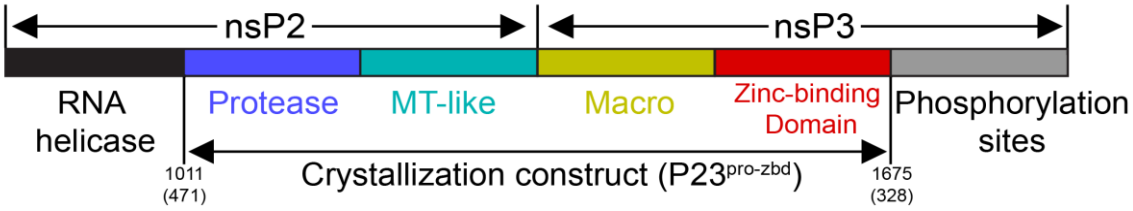

Figure 5. Domain Organization of P23 and Crystallization Construct Boundaries. Expression construct of SINV $P23^{pro\text{-}zbd}$ from the nsP2 protease domain through the AUD of nsP3, newly described as a zinc binding domain (ZBD). Amino acid numbering provided from the beginning of the P123 polyprotein provided with numbering from the N-terminus of nsP2 and nsP3 also noted in parenthesis. Figure adapted from (68).

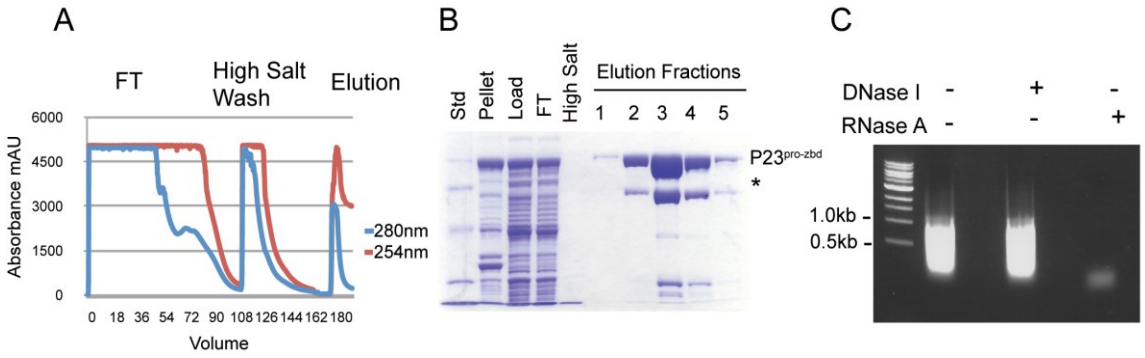

Figure 6. Purification of $P23^{pro\text{-}zbd}$ and Analysis of Co-Purifying Nucleic Acids. (A) This protein is initially purified from crude bacterial cell lysate using GST

affinity column. The red and blue represent absorbance in milliabsorbance units (mAU). Bacterial lysate is first loaded onto the column (FT) and the column is washed with high salt buffer to release nucleic acids. Bound GST-P23$^{pro\text{-}zbd}$ protein is then eluted with buffer containing glutathione. (B) SDS-PAGE analysis of the GST- P23$^{pro\text{-}zbd}$ purification in A. The asterisk (*) denotes an nsP3 fragment degradation product not resulting from authentic P2/3 cleavage. (C) Agarose gel visualization of nucleic acids released in the high salt wash digested by DNase I or RNase A. Figure adapted from (68).

2. Overview of SINV P23$^{pro\text{-}zbd}$ Structure

SINV P23$^{pro\text{-}zbd}$ crystals diffracted to 2.85 Å at Brookhaven National Synchrotron Light Source (PDB ID: 4GUA). The structure was determined by molecular replacement with nsP2-protease and nsP3 macrodomain (Figure 7; Table 1) (53, 88). There are three molecules of P23$^{pro\text{-}zbd}$, 31 sulfate ions, and five molecules of MES in each asymmetric unit (68). The protein is fairly compact, with each of the four domains occupying a vertex of a rectangle.

nsP2 and nsP3 share over 3,000 Å^2 of surface area. The two nsP3 domains are connected by a forty amino acid linker which lacks secondary structure, but wraps in a ring-like structure around the MT-like domain of nsP2 (Figure 8A-C). nsP2 residue Arg781 protrudes through the opening created by nsP3 (Figure 8C). The interface between the two proteins is charged, with nsP2 being overall basic and nsP3 overall acidic (Figure 8B-C).

The P2/3 cleavage site is solvent exposed; however, it is located within a cleft 11-13 Å wide that does not allow modeling of the nsP2 protease onto the cleavage site without significant steric clashes (Figure 9). Furthermore, the protease active site is 40 Å away from the cleavage site, indicating that the proposed *trans* cleavage mechanism is likely correct (29, 36, 39, 68).

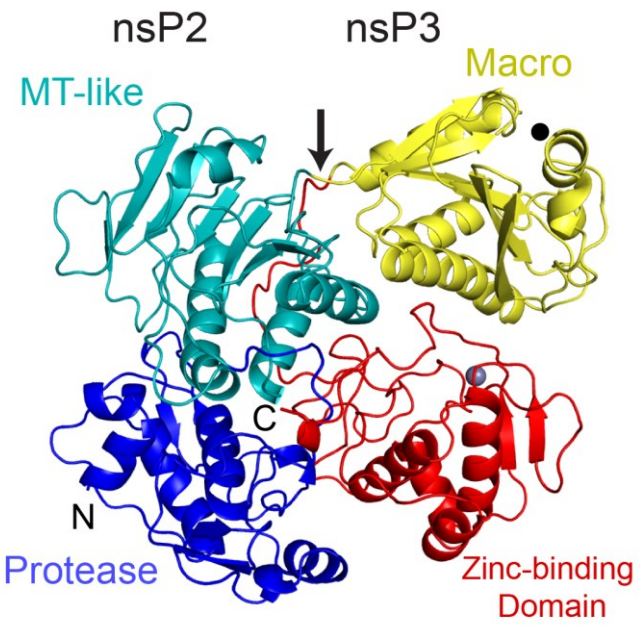

Figure 7. Structure of SINV P23$^{pro\text{-}zbd}$ Ribbon Model. Domain coloring of P23$^{pro\text{-}zbd}$ shown is the same as in Figure 5. A grey sphere represents a zinc ion, the P2/3 cleavage site an arrow, ADP-ribose binding site a filled circle, and the nsP2 protease active site an asterisk (*). Figure adapted from (68).

Table 1. Data Collection and Refinement Statistics for SINV P23$^{pro\text{-}zbd}$.

Data Collection	
Space group	$P3_121$
Cell dimensions	
a, b, c (Å)	147.3, 147.3, 360.5
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	52.05-2.85 (3.01-2.85)*
R_{merge}	0.137 (1.103)
R_{pim}**	0.044 (0.353)
$I/\sigma I$	15.6 (2.5)
Completeness (%)	99.9 (99.7)
Redundancy	10.7 (10.5)
Refinement	
Resolution (Å)	51.51-2.85
No. reflections	105713
R_{work}/R_{free}	0.220/0.251
No. atoms	
Protein	15594
Ligand/ion	30
B-factors	
Protein	63.7
Ligand/Ion	89.6
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.841
Ramachandran Plot	
Most favored (%)	98.1
Allowed (%)	1.9
Disallowed (%)	0

*Highest resolution shell is shown in parenthesis.

** $R_{p.i.m.} = \sum_{hkl}\sqrt{\frac{1}{N-1}}\sum_i \left| I_i(hkl) - \overline{I(hkl)} \right| / \sum_{hkl}\sum_i I_i(hkl)$, where I_i(hkl) = Intensity of the *i*th observation of the reflection *hkl* and *N* is the redundancy of multiplicity of the observed reflection, calculated using the program SCALA (104).

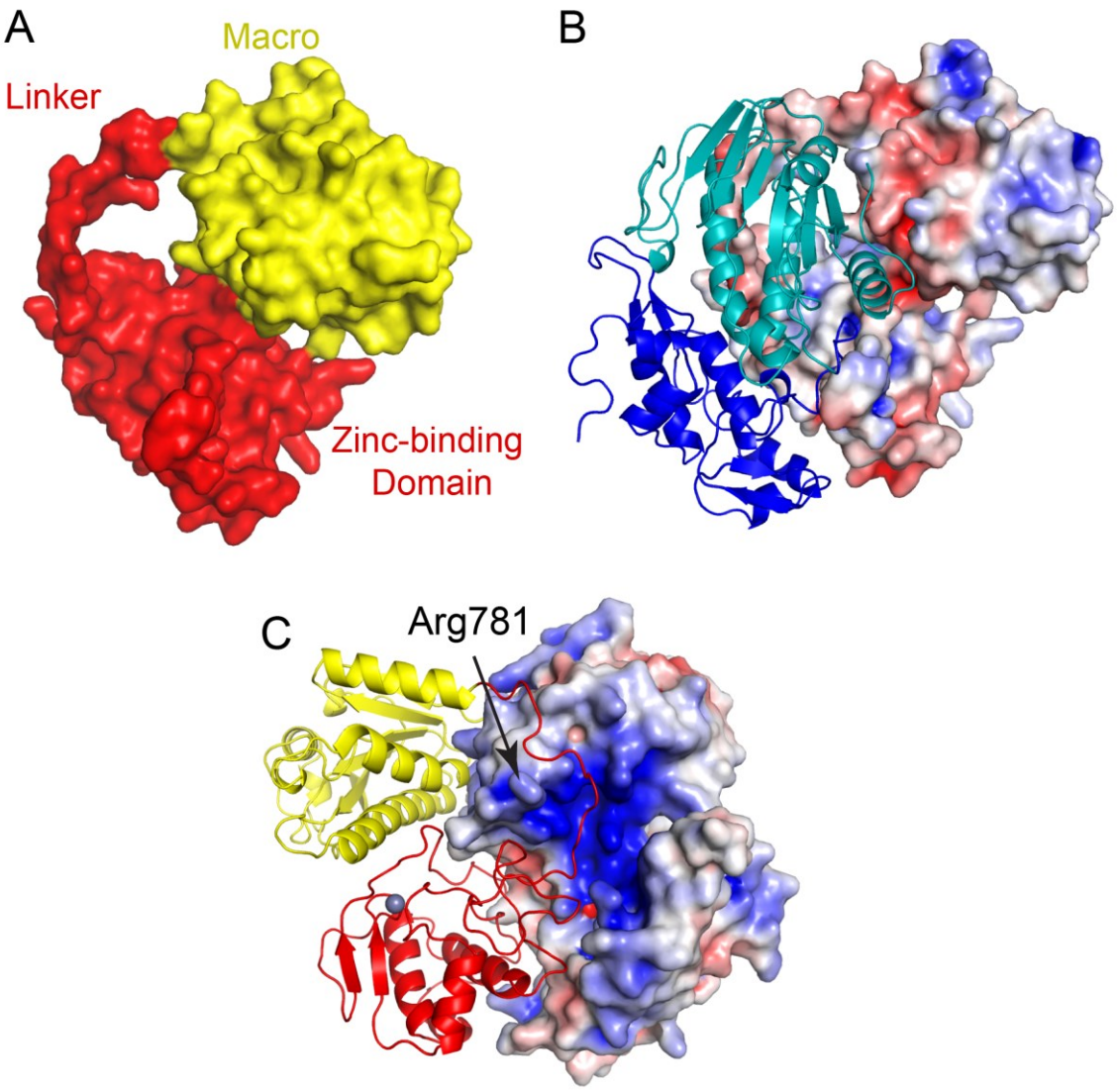

Figure 8. nsP3 Forms a Ring-like Structure Around nsP2. (A) Surface of nsP3 shown alone with the macrodomain in yellow and the linker and ZBD in red. Surface of nsP3 (B) or nsP2 (C) colored for electrostatic potential ±5kT/e; blue (basic), red (acidic), and white (neutral) with ribbon diagram of nsP2 (B) or nsP3 (C). Figure adapted from (68).

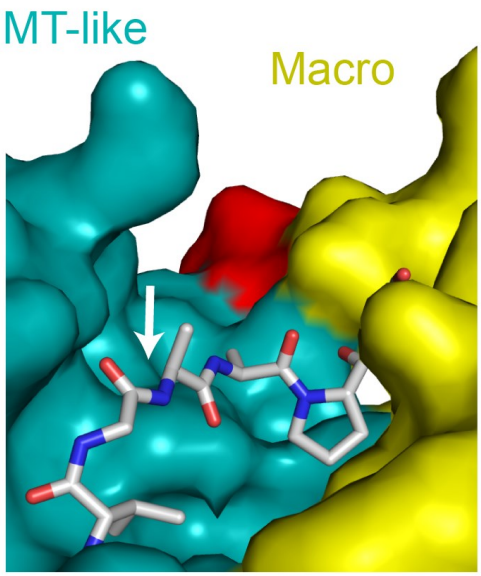

Figure 9. P2/3 Cleavage Site Cleft. The P2/3 cleavage site spans a cleft between the nsP2 MT-like domain and nsP3 macrodomain. The cleavage site is indicated by an arrow and amino acids on either side of the site are shown in stick representation. Figure adapted from (68).

3. nsP3 Alphavirus Unique Domain is an Essential Zinc Binding Domain

The $P23^{pro\text{-}zbd}$ structure contains the previously uncharacterized AUD which we revealed to be a novel zinc binding domain (68). Sequence alignment of the AUD shows several absolutely conserved cysteine residues within alphavirus sequences, which indicate a metal ion binding site. The AUD of nsP3 was expressed alone and analyzed by quantitative X-ray fluorescence for common transition metals found in metalloproteins and were shown to have one zinc ion per molecule (126, 127). The conserved nsP3 AUD contains an antiparallel, alpha helical bundle, two parallel beta strands, and zinc-coordination site (Figure 10A). The zinc ion location was confirmed by anomalous difference maps and is coordinated by C263, C265, C288, and C306 of nsP3 (Figure 10B-C). Analysis of current Protein Data Bank structures using PDBe Fold and Dali servers did not identify any statistically significant similarities with known metalloprotein folds (128, 129). The first two cysteine residues are in the same loop between

alpha helix three and four, while C288 and C306 are at the C-terminal ends of the two parallel beta strands, which represents a novel structural configuration for zinc ion coordination (Figure 10A) (130). Additionally, there are two highly conserved serine residues (S289 and S290) located below the zinc atom which force the cysteine coordination into a distorted tetragon formation (Figure 10C).

Site-directed mutagenesis was used to mutate each zinc-coordinating cysteine residue in nsP3 individually to alanine residues in a SINV infectious clone. SINV missing any one of the four zinc coordinating cysteines failed to achieve productive infection in highly permissive BHK cell lines as evaluated by infectious center and plaque assays (Table 2). A conserved nsP3 cysteine residue (C247) unrelated to zinc coordination was also mutated to serve as a control and replicated similar to wild type SINV levels. Although the two serines near the zinc binding site are highly conserved throughout alphaviruses, mutations of those residues still resulted in near wild type levels of RNA infectivity and viral titers. Only after a double mutant (S289A + S290A) was created was a significant reduction seen in RNA infectivity levels; although the viral titers ultimately return to wild type levels.

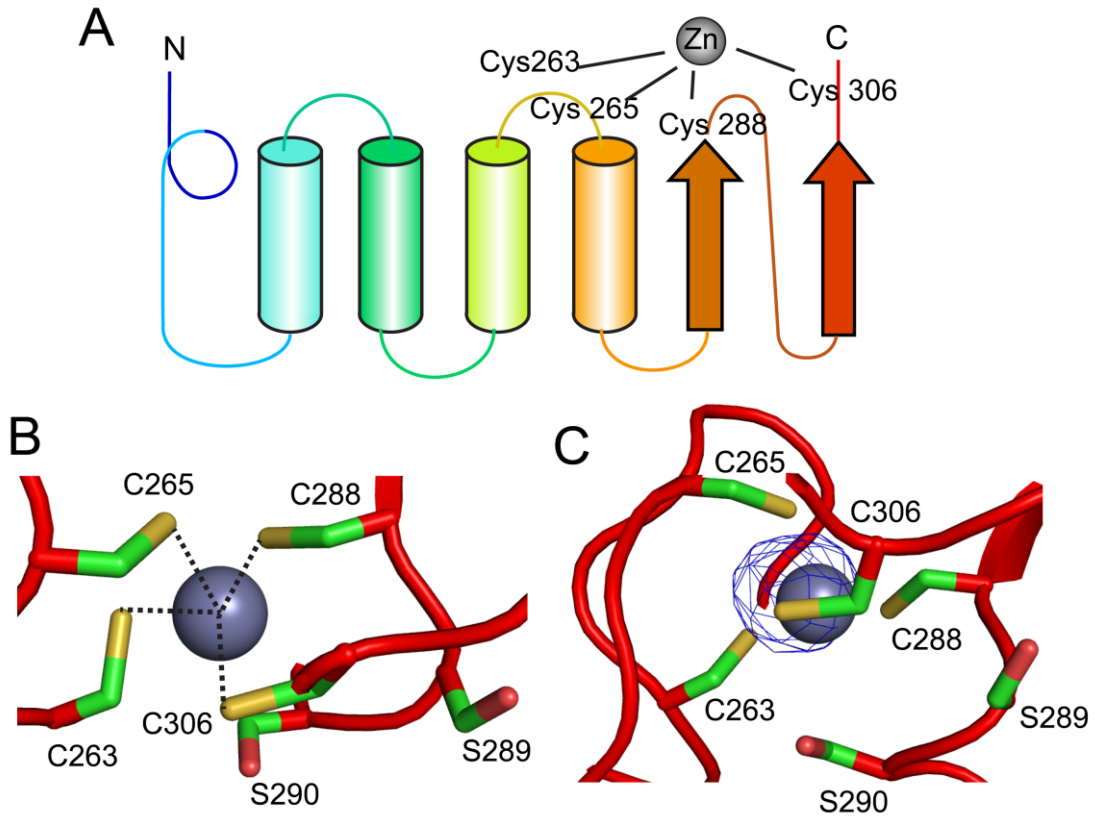

Figure 10. Previously Unknown nsP3 Zinc Binding Domain. (A) Topology model in rainbow coloring scheme of nsP3 linker and ZBD. Four zinc-coordinating cysteine residues of nsP3 are noted. (B) Stick representation of the nsP3 ZBD zinc coordination site with the four coordinating cysteines and a grey sphere denoting the zinc ion. Two highly conserved serine residues located near the site are also noted. (C) An alternate view of the zinc ion with an anomalous difference map contoured at 9σ. Figure adapted from (68).

Table 2. Viral Replication of SINV with Zinc Coordinating Cysteine Mutants.

Mutation in SINV nsP3	RNA Infectivity (IU/µg RNA)	Standard Error of the Mean	Titer (PFU/mL)	Standard Error of the Mean
Wild type	$1.4x10^6$	$4.1x10^5$	$6.0x10^8$	$1.8x10^8$
C263A	<50	-	<50	-
C265A	<50	-	<50	-
C288A	<50	-	<50	-
C306A	<50	-	<50	-
C247A	$6.9x10^5$	$1.6x10^5$	$4.7x10^8$	$2.0x10^8$
S289A	$1.8x10^6$	$1.8x10^5$	$5.2x10^8$	$4.3x10^7$
S290A	$1.3x10^6$	$2.5x10^5$	$1.8x10^9$	$7.1x10^8$
S289A+S290A	$8.2x10^4$	$5.4x10^4$	$1.2x10^8$	$4.0x10^7$

4. Discussion

The SINV P23$^{pro\text{-}zbd}$ structure described here is only the second viral nonstructural precleavage structure in literature. The previous structure 3CD from poliovirus introduced above shows two globular domains with no intramolecular contacts and a flexible linker connection containing the cleavage site. Our SINV P23$^{pro\text{-}zbd}$ structure shows an entirely different conformation, with the arrangement of four domains (protease, MT-like, macrodomain, and ZBD) each occupying a vertex of a rectangle in a highly compact manner. This is the first evidence that viral replicative polyproteins do not all resemble "beads on a string" but can be intertwined and probably undergo large structural rearrangements upon cleavage. Additionally, the previously mysterious AUD contains a novel zinc binding fold which is absolutely necessary for viral replication.

As discussed earlier, a portion of nsP2 localizes to the nucleus while nsP3 stays in the cytoplasm after P2/3 cleavage (63-65). Because of the extensive surface area shared by the two proteins, it is unclear how the two dissociate post-cleavage. Although our construct contained an active protease domain, the P2/3 bond was not cleaved. This is consistent with previous evidence showing that the N-terminal portion of nsP2 is necessary for P2/3 cleavage (39, 51). The cleavage site, although solvent exposed, is located in a cleft and inaccessible to the protease. This indicates that access to the cleavage site is likely tightly regulated and only a free nsP2 N-terminus after cleavage at the P1/2 junction can expose the P2/3 cleavage site. The 40 Å distance between the protease active site and the cleavage site combined with the inflexible, compact nature of the protein indicates that cleavage of this bond is likely done in *trans*, supporting current models (29, 36, 39). Overall, our structure implies that a tightly coordinated cleavage mechanism governs viral polyprotein processing and, in turn, RNA replication.

Nonstructural polyprotein cleavage status dictates RNA template usage of the alphavirus replication machinery. SINV P23$^{pro\text{-}zbd}$ co-purified with bacterial RNA. A large, basic surface that extends across nsP3 and nsP2 may be responsible for this (Figure 11A-B). This observation combined with the fact that many zinc binding proteins are also nucleic acid-binding suggests a role in RNA binding in the replication complex (131). Molecules of sulfate and MES that were used in crystallization were bound to the protein along the basic surface. These may be potential sites for phosphate groups of nucleic acids binding. Our P23$^{pro\text{-}zbd}$ structure expands our knowledge of viral polyprotein processing and how it relates to alphavirus RNA replication.

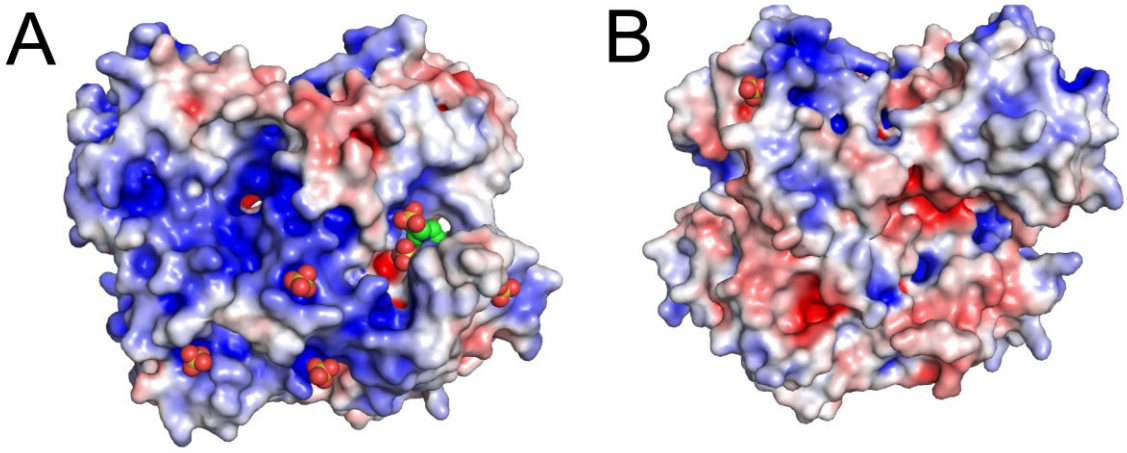

Figure 11. Potential RNA Binding Surface of P23$^{pro\text{-}zbd}$. (A) Surface of P23$^{pro\text{-}zbd}$ colored for electrostatic potential at ±5kT/e. Location of sulfate and MES ligands are represented by spheres. (B) Rotation of 180° around a vertical axis from A. Figure adapted from (68).

Section II. Mutational analysis of Sindbis Virus Polyprotein P23 Interface

Several previously identified mutations in nsP2 result in noncytopathic virus and persistent infection in mammalian cells (69-73). Viruses carrying these mutations have reduced viral RNA replications levels and are unable to shut off host gene expression but still retain the ability to replicate to high titers. A well characterized non-cytopathic mutation site is SINV nsP2 P726. The phenotype produced by mutation of P726 of nsP2 in SINV is dependent on which amino acid residue the proline is replaced with. Large side chain amino acids such as phenylalanine, tyrosine, leucine, or arginine cause a dramatic decrease in viral RNA replication. These viruses are also non-cytopathic. Amino acid substitutions with small side chains such as alanine or valine give a phenotype similar to wild type SINV. Other amino acids like glycine and serine give intermediate phenotypes (69). Transcription is halted during infection by nsP2 causing degradation of Rpb1, and essential RNAPII subunit (67). The non-cytopathic P726G mutant has been shown to be necessary for degradation of Rpb1. This mutation clearly affects the downstream activities of mature nsP2 in the nucleus, but it is unknown how this mutation affects the properties of nsP2, precleavage or mature, while participating in RNA replication in the viral replication complex.

To this end, we mapped the non-cytopathic mutations in our SINV $P23^{pro\text{-}zbd}$ structure and performed structure-directed mutational analysis *in vivo* to investigate the role of non-cytopathic mutation during viral infection.

1. Non-Cytopathic Mutations Map to P2/3 Interface

Our SINV $P23^{pro\text{-}zbd}$ structure revealed a long, extended linker connecting the two domains of nsP3 and encircling nsP2 (68). All of the previously described non-cytopathic mutations were found to map to the surface of nsP2 at the P2/3 interface following the path of

the nsP3 linker (Figure 12A). P726 is in an nsP2 loop connecting α9 and β11 and makes direct contact with the nsP3 linker (Figure 12B). Since there is a direct correlation between the size of the residue located at position 726 and the severity of phenotype, we hypothesized that the interface of P2/3 may be playing a role in the non-cytopathic phenotype based on the disruption of the P2/3 interface. Modeling studies showed that mutating nsP2 P726 to large, bulky residues resulted in steric clashes with the nsP3 linker. Of particular interest are V162 and L165 of nsP3, as they interact directly with the surface of nsP2 (Figure 12B-C).

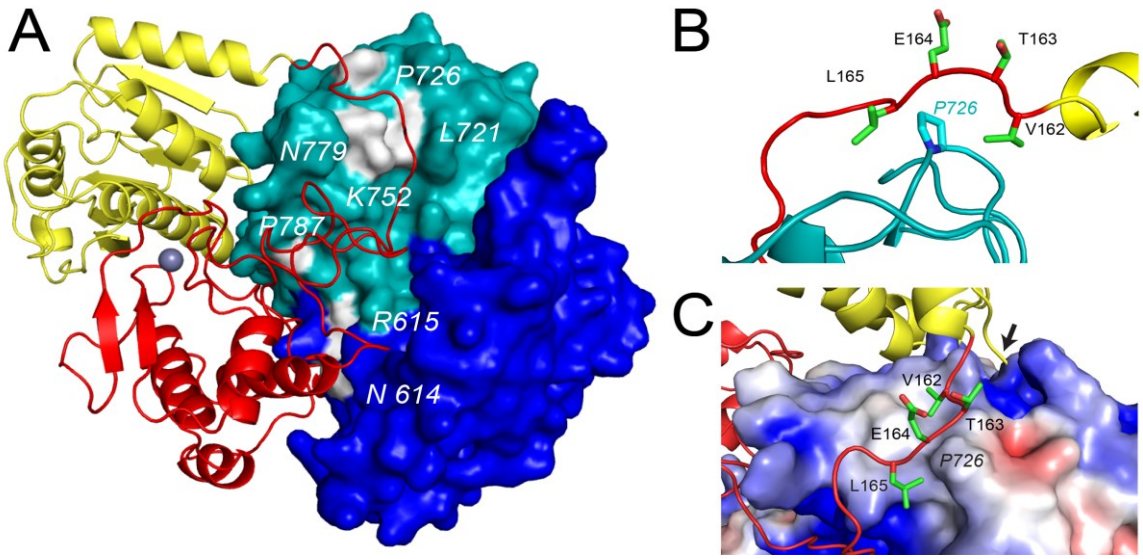

Figure 12. Non-Cytopathic Mutations of nsP2 Map to the P2/3 Interface. (A) Surface of nsP2 highlighting the location of nsP2 non-cytopathic mutations in white. Numbering corresponds to SINV nsP2 sequence. Domain coloring is identical to Figure 5. (B) Location of nsP2 P726 with the adjacent nsP3 residues pictured. The last helix in nsP3 macrodomain is depicted in yellow. (C) The proximity of nsP2 noncytopathic mutation P726 to the P2/3 cleavage site. The surface of nsP2 surrounding P726 is colored for electrostatic potential at ±4kT/e with adjacent nsP3 linker region in stick format. Residues labeled in italics are in nsP2 and the arrow indicates the P2/3 cleavage site. Figure adapted from (68).

2. Effect on Viral RNA Infectivity and Viral Replication

Site-directed mutagenesis to SINV nsP3 V162 and L165 was performed to determine the effect of mutations which may destabilize the P2/3 interface in proximity to the nsP2 P726 region. The sidechains of V162 and L165 are approximately 4.6 and 3.3 Å away from the nsP2 P726 sidechain, respectively. nsP3 residues were mutated to a variety of amino acids: Ala, Asp, Glu, Phe, Arg, or Trp. SINV containing nsP2 P726 mutations to Ser, Gly, or Leu were used as a comparison as their phenotypes have been studied extensively in the literature already.

RNA infectivity was measured by infectious center assay on BHK cells. P726L (nsP2), L165D, L165E, and L165R (nsP3) mutants had reduced viral RNA infectivity (Figure 13A). The SINV virus with a P726L mutation in nsP2 only produced 2% of RNA infectivity levels compared with wild type virus, while the three nsP3 L165 mutants averaged 8% of wild type (68). These SINV mutants also showed a small plaque phenotype in infectious center assay (Figure 14A-D). 24 hours post-electroporation, viral titers were measured for each virus. L165A, L165D, L165E, and L165R mutants were significantly lower viral titers at an average of 34% wild type levels (Figure 13B). The nsP3 mutant with the poorest viral replication capabilities was L165D, which only produced 26% of wild type titers. Interestingly, all of the V162 mutants were fully viable and had wild type phenotypes in both RNA infectivity and viral titer. These mutants were not investigated further. P726 mutations with severe RNA replication deficiencies such as P726L are known to initially have very low replication levels, but also quickly revert to recover wild type-like replication; therefore, it is unsurprising that the initially poor P726L virus recovered to wild type titer levels (69).

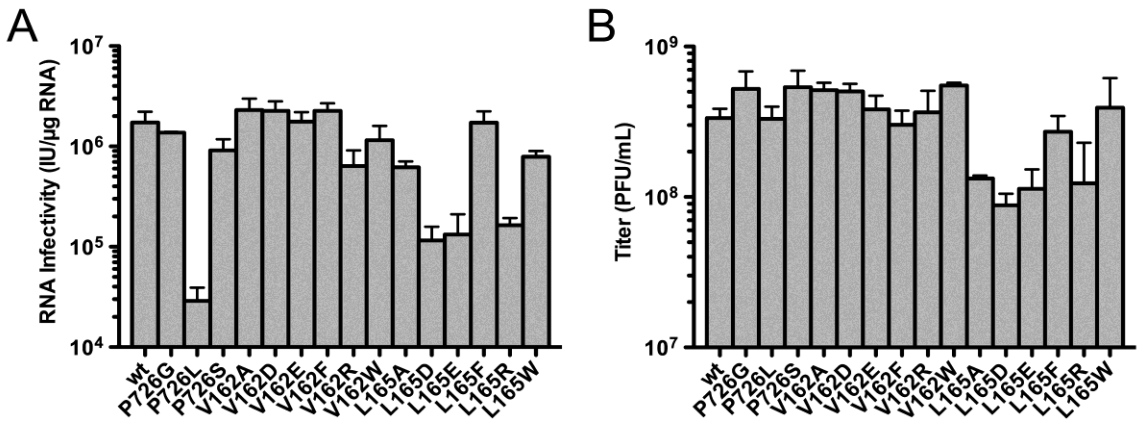

Figure 13. *In Vivo* Effects of nsP3 Linker Region Mutation. (A) Viral RNA infectivity [infectious units (IU)/µg RNA) as measured by infectious center assay. (B) Viral titer [plaque forming unites (PFU)/mL] 24 hours after electroporation of BHK cells as determined by plaque assay over two or more independent experiments. Error bars represent standard error of the mean (SEM). Figure adapted from (68).

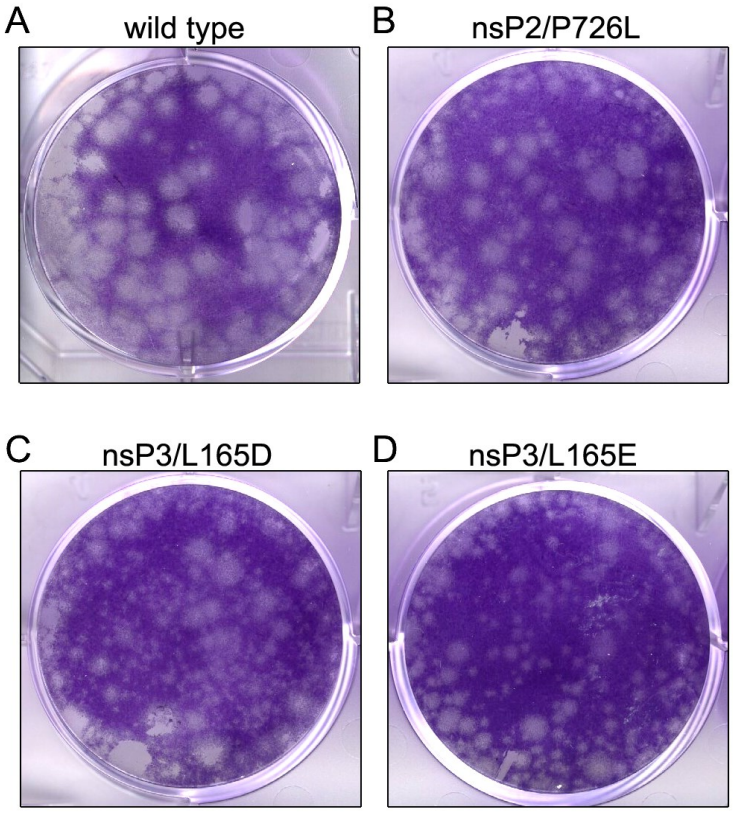

Figure 14. Small Plaque Phenotype of P2/3 Interface Mutations. Infectious center assay was performed in SINV RNA transfected BHK cells plated at increasing dilutions on a BHK monolayer. (A) Wild type SINV RNA transfected

cells at a dilution of 10^3. (B) SINV nsP2 P726L RNA transfected cells at a dilution of 10^1. (C) SINV nsP3 L165D RNA transfected cells at a dilution of 10^1. (D) SINV nsP3 L165E RNA transfected cells at a dilution of 10^2. Figure adapted from (68).

3. Cytopathogenicity of nsP3 Linker Region Mutants

Previous evidence shows that the P726 residue is necessary for cytopathic viral infection. Our SINV P23 structure shows that the cytopathogenicity conferring residues of nsP2 are located on the surface of nsP2. In order to asses SINV with nsP3 linker region mutations and their ability to confer a non-cytopathic phenotype, a SINV replicon with a selectable marker was used (Figure 15) (69). The TSG/PAC replicon background allows for identification of puromycin resistant foci through replacement of the SINV structural ORF with a puromycin resistance gene. Mutations in the nsPs which allow productive but non-cytopathic replication will allow foci to grow under puromycin treatment. Positive control P726G mutation produced puromycin resistant foci (Figure 16); however, none of the nsP3 linker region mutants produced a noncytopathic phenotype in the TSG/PAC background (68).

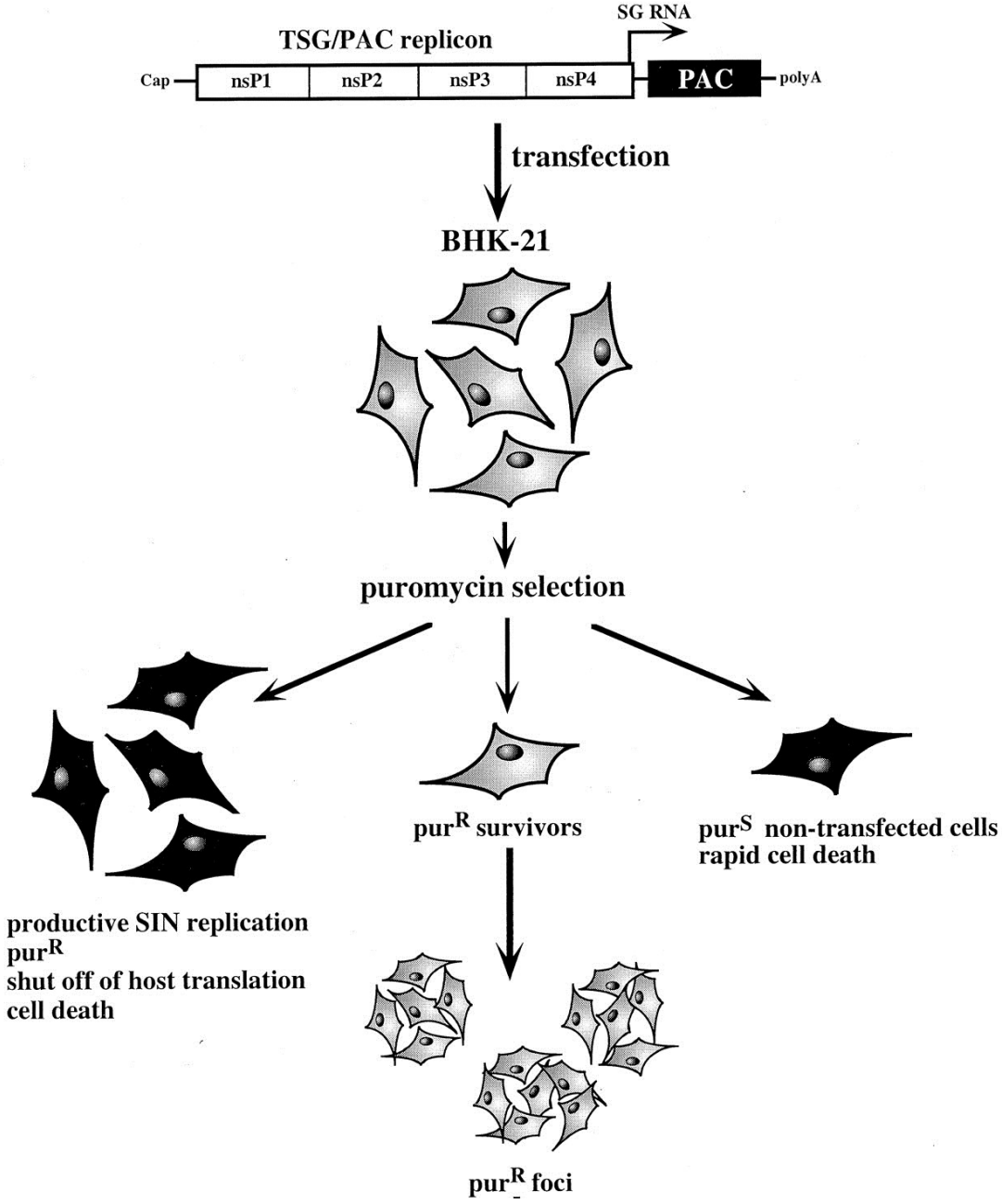

Figure 15. Schematic of SINV RNA Replicon for Identification of Non-Cytopathic Replication. The TSG/PAC replicon was used to identify non-cytopathic replication of SINV nsP mutations. The structural ORF of SINV is replaced with a puromycin selectable marker resistance gene. Transfected cells are treated with puromycin. Cells which do not take up the RNA or are transfected with productive, cytopathic SINV replicons quickly die. Those transfected and producing SINV nsPs with noncytopathic mutations survive and form puromycin resistant foci. Figure adapted from (69).

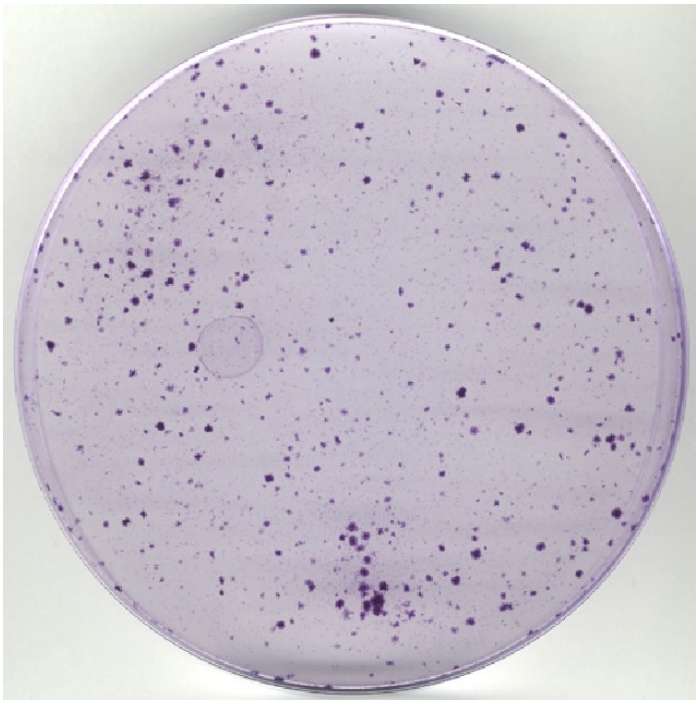

Figure 16. nsP2 P726G Mutation Causes Non-Cytopathic, Persistent Infection. SINV TSG/PAC construct with nsP2 P726G mutation form puromycin resistant foci, confirming persistent, non-cytopathic infection.

4. Production of Viral Proteins and Shutoff of Host Protein Synthesis

Alphaviruses are capable of host cell transcription and translation shut off. To evaluate viral and host protein production over the course of infection, and the efficiency of translational shut-off, radiolabeling experiments with ^{35}S labeled methionine and cysteine were performed. Cellular translation was evaluated via actin production. Wild type SINV infection halts host cell translation of actin completely within six hours post infection at an MOI of 10 PFU per cell (Figure 17). In contrast, non-cytopathic SINV with nsP2 mutations P726S and P726G were unable to do shut off host translation even up to 12 hours post-infection. nsP2 P726L and the four selected nsP3 L165 mutants (Asp, Glu, Arg, and Trp) halted host translation completely, but appears delayed compared to wild type. Viral capsid was produced strongly in wild type and all mutant viruses by four hours post infection with levels increasing over the course of infection in

wild type SINV, P726L, and all four nsP3 L165 mutants examined (Figure 17). P726G and P726S mutants produced capsid at a consistent level over the course of infection.

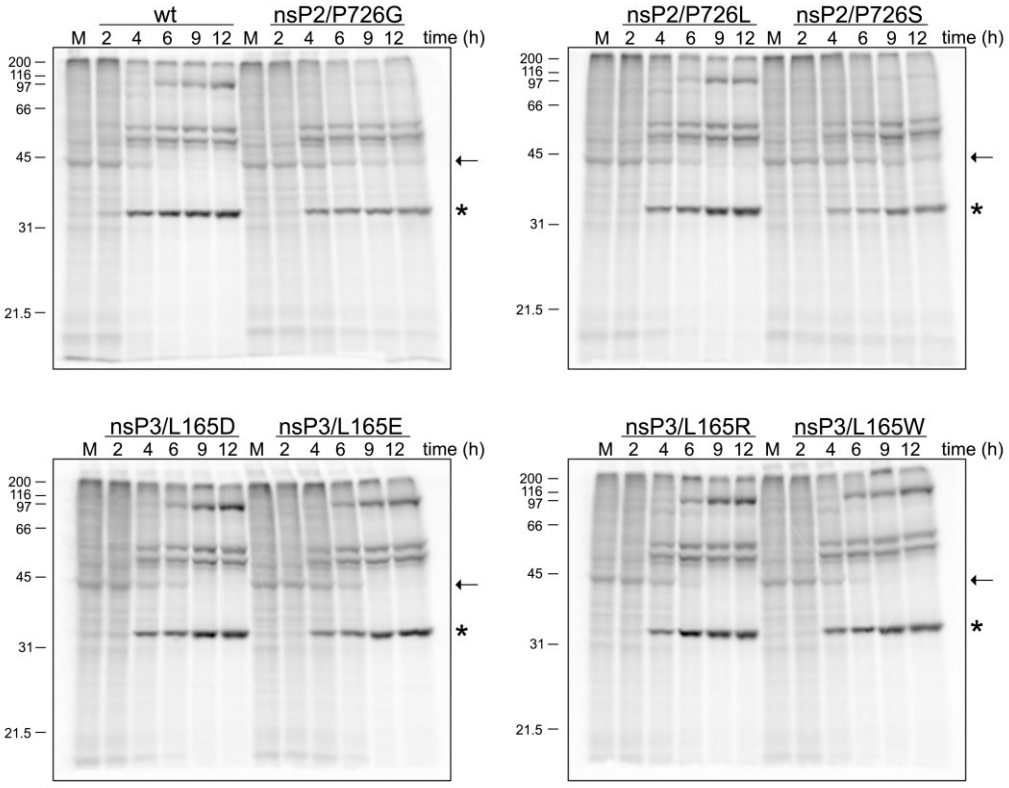

Figure 17. Host Translational Shut Off and Viral Protein Production during SINV Infection. ^{35}S labeled host and viral proteins during SINV infection at an MOI of 10. Cellular actin is denoted with an arrow and SINV capsid is denoted with an asterisk (*). Mock infected cells are indicated by an M. Figure adapted from (68).

5. Shutoff of Host RNA Replication and Viral RNA Replication

To evaluate viral and host RNA production over the course of infection, and the efficiency of transcriptional shut-off, radiolabeling experiments with tritium (^{3}H) labeled uridine in the presence and absence of actinomycin D were performed. Actinomycin D inhibits RNA transcription by preventing elongation of RNAs thus allowing a clearer visualization of viral RNA transcription via the viral RNA replication complex by eliminating background cellular RNAs

(132). In the presence of actinomycin D, genomic and subgenomic viral RNA production begins in wild type SINV infection around two hours post infection and appears to remain steady over 12 hours of infection (Figure 18A). Levels of subgenomic and genomic RNAs are much lower in the nsP2 P726G and P726S mutants. Viral RNA production in P726L is significantly lower at two hours post infection, but quickly recovers to wild type levels later in infection. This data is consistent for this quickly reverting mutation. Of the nsP3 mutations observed, all four viruses show an increasing amount of viral RNA production over time of both genomic and subgenomic RNAs, surpassing the wild type levels of RNA replication (Figure 18A).

In the absence of actinomycin D, the same trend is reiterated for viral RNA replication as without actinomycin D. Cellular 28S and 18S rRNA production was completely halted by eight hours post-infection with wild type SINV (Figure 18B). P726G and P726S SINV mutants were not able to shut off host cell transcription showing rRNA being produced up to twelve hours post infection. The nsP2 P726L and four nsP3 L165 mutants were able to inhibit transcription, although the effect is delayed compared to wild type SINV. These mutants achieved complete shut off of rRNA production by twelve hours post infection (Figure 18B).

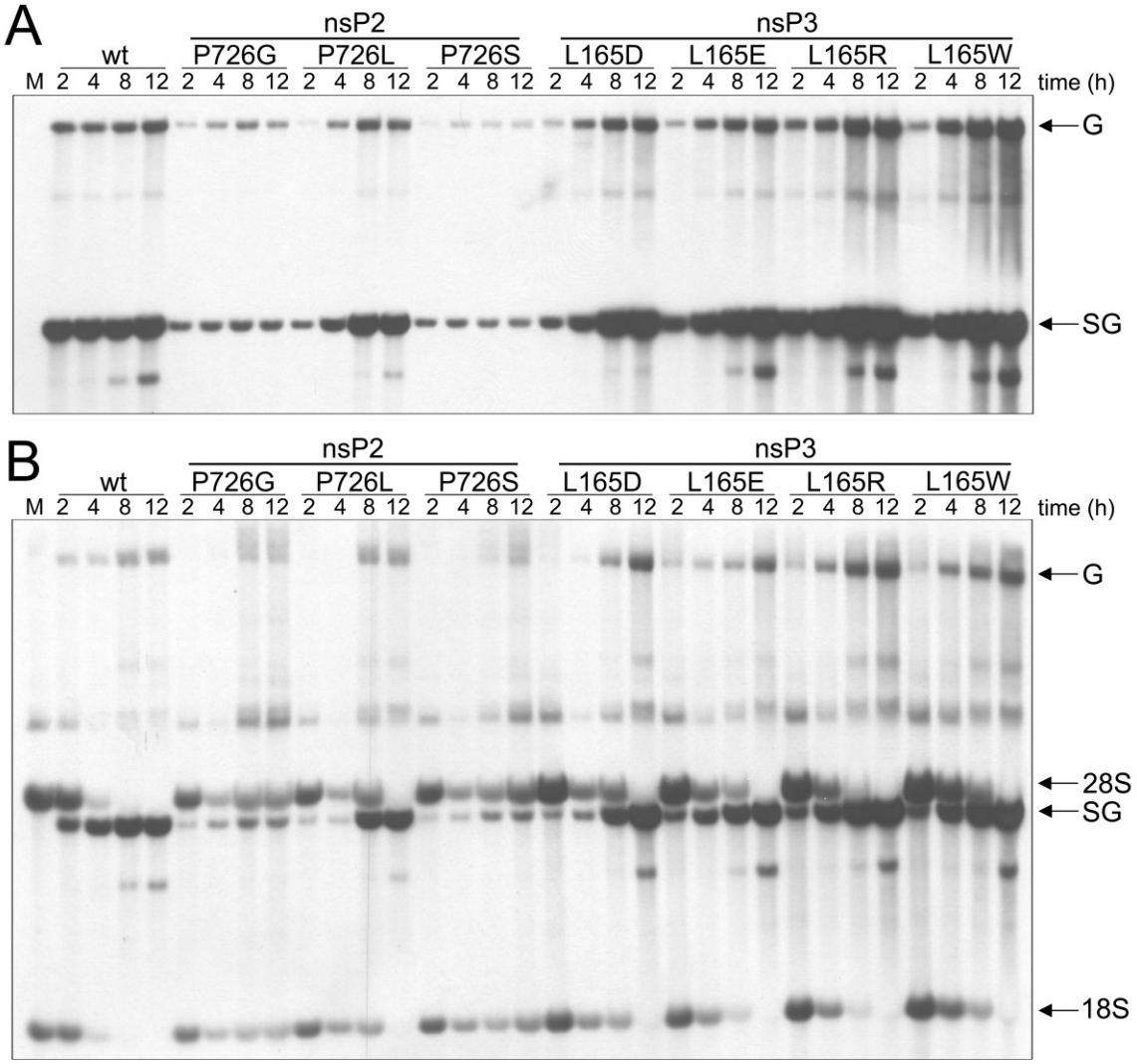

Figure 18. Host Transcriptional Shut Off and Viral RNA Production during SINV Infection. Tritium labeled host and viral RNAs during SINV infection at an MOI of 10. At the indicated time points above, RNAs were labeled in the presence (A) or absence (B) of actinomycin D. Total RNAs were analyzed by denaturing agarose gel electrophoresis. Viral genomic (G) and subgenomic (SG) and cellular rRNAs (28S and 18S) are noted. Mock infected cells are indicated by an M. Figure adapted from (68).

6. Efficiency of Polyprotein Processing in Vivo

Previous studies indicate that in SFV and SINV, the first 170 amino acid residues of nsP3 macrodomain are required for efficient cleavage at the P2/3 junction. Furthermore, residues R159 and E163 of SFV nsP3 specifically have been shown to affect P2/3 cleavage *in vitro* (51).

We used western blotting to evaluate the effect of the SIV nsP3 mutations on cleavage activity during active SINV infection. nsP3 linker mutants L165D and L165E had drastic inefficiency in P2/3 cleavage that was not seen in the other L165 mutants (Figure 19A-B). Uncleaved P23 accumulated in nsP3 L165D and L165E mutants starting six hours post-infection. Wild type SINV, non-cytopathic nsP2 P726 mutants, and nsP3 L165W or L165R mutants did not show cleavage deficiencies. Anti-nsP3 antibodies were used to confirm the presence of P23 (Figure 19B). nsP3 typically appears as a smear due to multiple phosphorylation states (96).

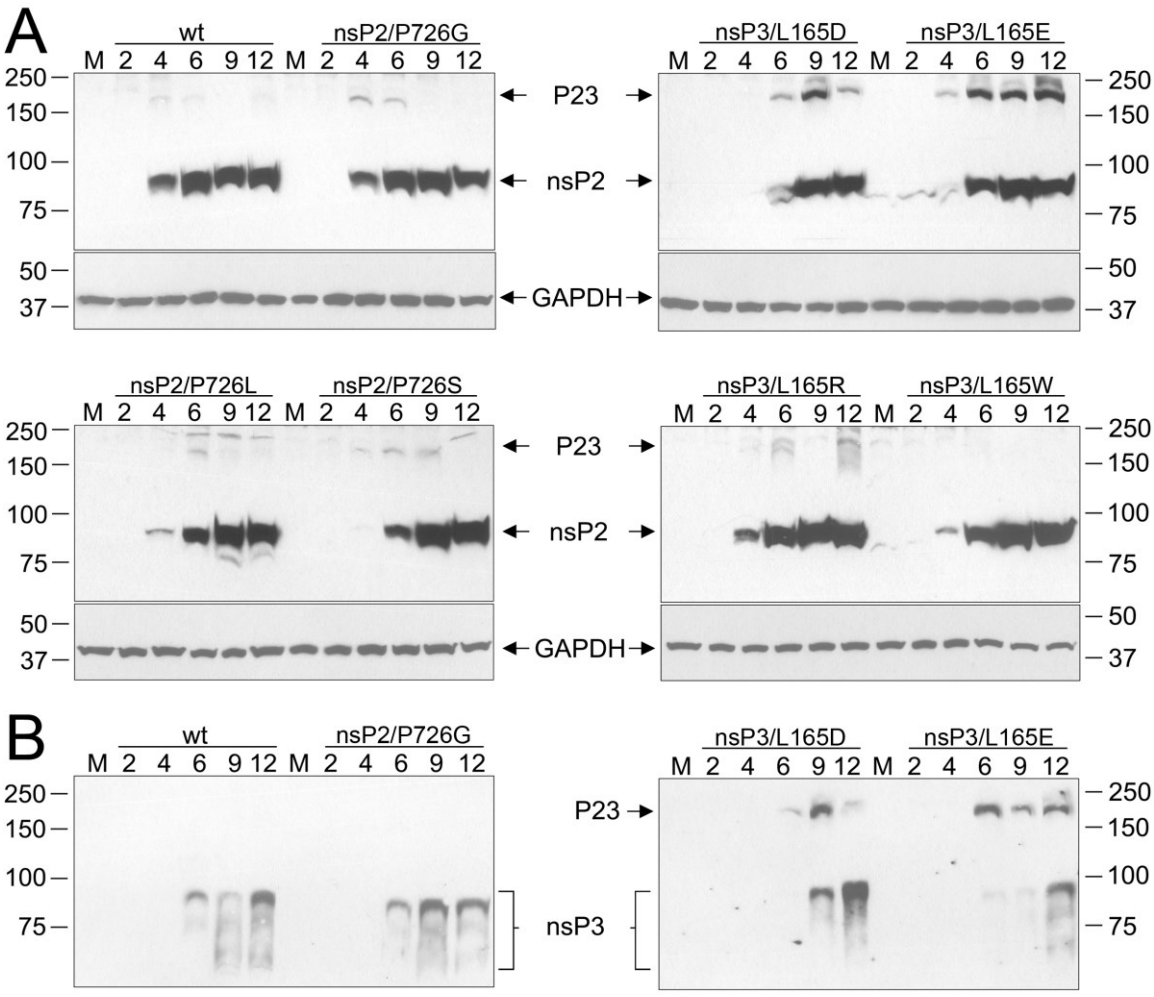

Figure 19. nsP3 Linker Region Mutants are P2/3 Cleavage Inefficient. Analysis of P2/3 cleavage *in vivo* via western blot on total protein harvested from BHK

cells infected with wt or mutant SINV. Membranes were probed using anti-nsP2 (A) and anti-nsP3 (B) antibodies. Figure adapted from (68).

7. Expression of P23$^{pro\text{-}zbd}$ Polyprotein with Interface Mutations

In order to perform further *in vitro* analyses on the effects of P2/3 interface mutations, GST fusion expression constructs were designed (as described in Section I) with nsP2 mutation P726G and nsP3 mutation L165E. Purification was performed initially using GST affinity, then hydroxyapatite and heparin chromatography. P23$^{pro\text{-}zbd}$ P726G was shown to be expressed and fully intact protein (Figure 20A). P23$^{pro\text{-}zbd}$ L165E, however, was expressed at much lower levels and resulting protein was highly insoluble (Figure 20B).

Although mass spectrometry (MS) analysis previously revealed cleavage of P23 was not at the authentic P2/3 site, to rule out the possibility of degradation by promiscuous nsP2 protease activity, the active site cysteine of nsP2 was mutated to a serine to abolish protease activity in wild type P23, P726G, and L165E constructs. SDS-PAGE analysis shows that the major degradation product is still present during expression of all three of these constructs, indicating that cleavage is likely due to a bacterial protease (Figure 21A). Additionally, nsP2 residue R802, which was implicated by MS analysis to be the site of cleavage, was mutated to alanine in the hopes of increasing the amount of protein obtainable for further experiment. This tactic was unsuccessful as the degradation product of P23$^{pro\text{-}zbd}$ is still made (Figure 21B).

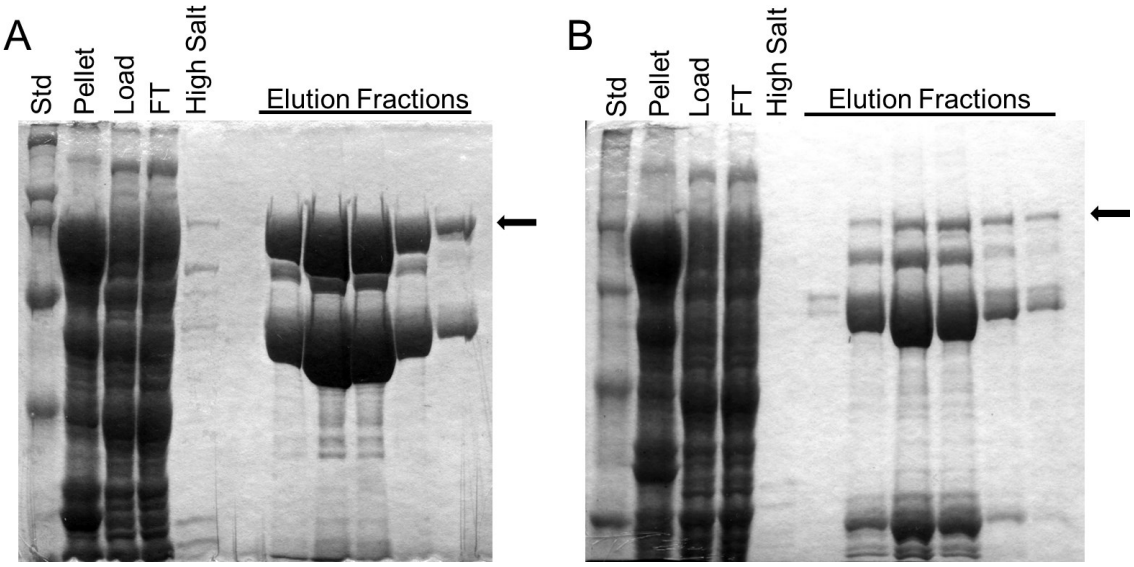

Figure 20. Expression of SINV P23$^{pro\text{-}zbd}$ with P2/3 Interface Mutations. SDS-PAGE analysis of GST affinity purification of SINV P23$^{pro\text{-}zbd}$ with (A) nsP2 P726G mutation or (B) nsP3 L165E mutation. The arrow indicates the P23 protein. Lower bands are degradation products.

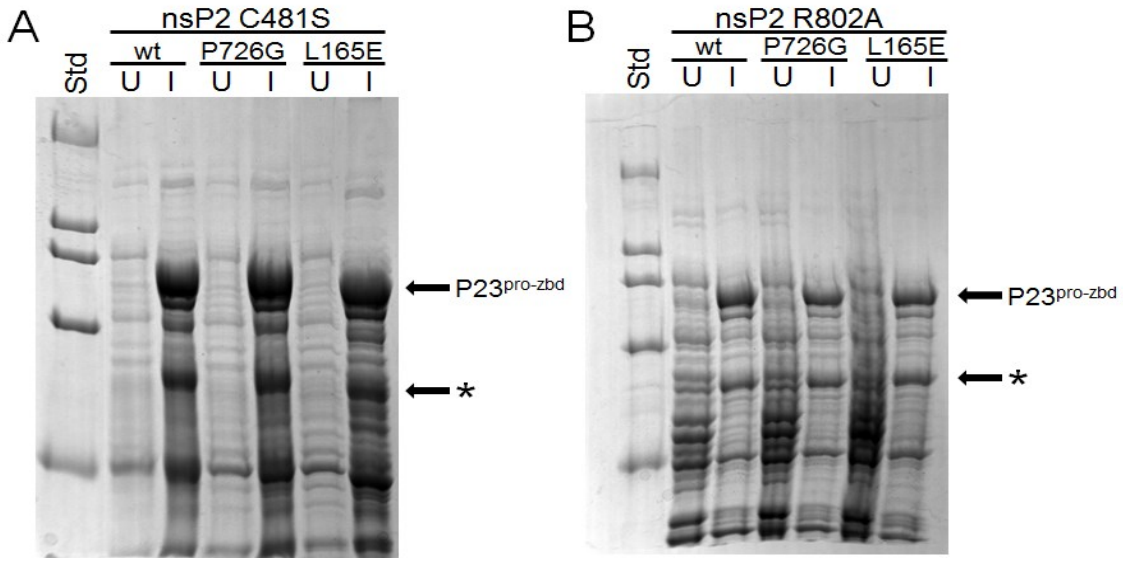

Figure 21. SINV P23$^{pro\text{-}zbd}$ Degradation Product Analysis. In wild type, P726, and L165E expression constructs, (A) nsP2 active site C481 was mutated to a serine residue or (B) the alternative cleavage site in nsP2, R802, was mutated to alanine. Uninduced bacterial cultures (U) were compared with induced samples (I) expressing the construct. Full length P23$^{pro\text{-}zbd}$ is noted as well as the degradation product (*).

8. Flexibility of the P23 Interface

As X-ray crystallography only gives a static molecular structure, hydrogen-deuterium exchange (HDX) in combination with MS was used to gain insight into any conformational dynamics within the SINV $P23^{pro\text{-}zbd}$ protein. Exposed areas of the protein which do not have stable hydrogen bonding will rapidly exchange hydrogen for deuterium. More tightly folded or buried areas of the protein will exchange more slowly. By exposing the protein to D_2O over measured periods of time, then analyzing by MS-based peptide mapping, information can be gained about which areas of the protein exchange more rapidly (high deuterium, more solvent exposed, more dynamic) or more slowly (low deuterium, less solvent exposed, less dynamic) (133).

For wild type $P23^{pro\text{-}zbd}$, HDX-MS shows regions of high flexibility mainly in the nsP3 macrodomain (Figure 22). The P2/3 cleavage site is buried in the crystal structure; however the cleavage site and surrounding strands leading into and out of the cleavage site (residues 325-373 in Figure 22) are available for deuterium exchange at about 60-70%. There is also a highly (80%+ deuterium exchange) solvent accessible loop in nsP3 macrodomain (residues 450-458 of Figure 22) located in proximity to the ADP-ribose binding site (Figure 23). With regard to the P2/3 interface, the nsP2 surface is very inaccessible at the nsP2 P726 position at only a six percent deuterium incorporation for the peptide (residue 256 of Figure 22). Conversely, the nsP3 linker residues corresponding to the region in contact with nsP2 P726, including our L165 residue of interest, is highly accessible with about 80%+ deuterium exchange (residues 498-502 in Figure 22). In comparing wild type and P726G mutant, $P23^{pro\text{-}zbd}$ proteins shows a minor increase (5%) in accessibility of the cleavage site in the mutant, a 17% increase in the nsP2 P726-containing peptide, and a 10% increase in the corresponding nsP3 linker region (Figure 24).

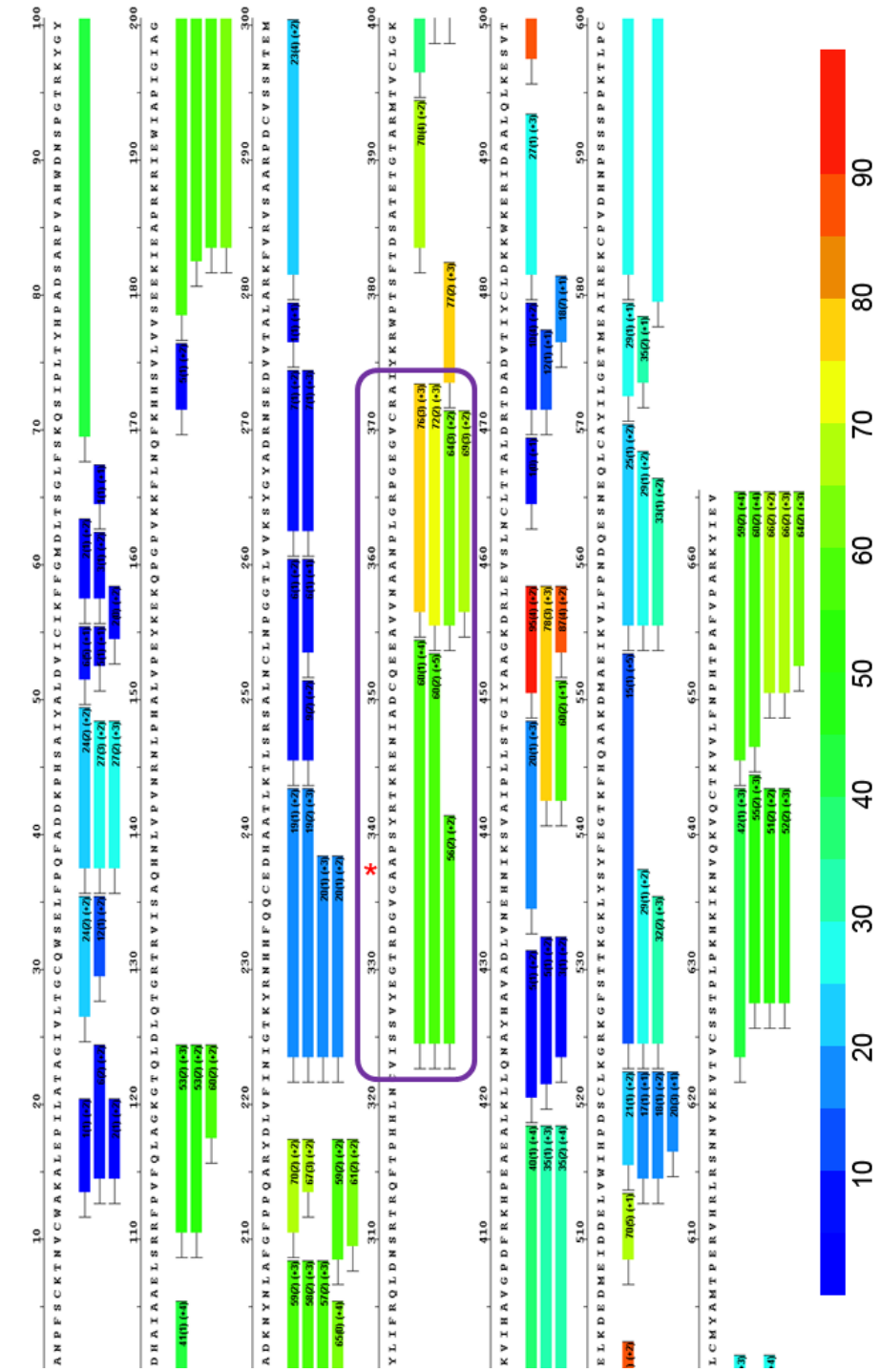

Figure 22. Dynamic Properties of SINV P23$^{pro\text{-}zbd}$. HDX and MS results for P23 show areas of high flexibility in the macrodomain and nsP3 linker. The cleavage site is denoted with a red asterisk and the strand leading into and out of the cleavage site emphasized by a purple box. Each horizontal bar represents a peptide analyzed. The color code represents the average percent of deuterium incorporation across six time points: 10, 30, 60, 300, 900, and 3600 seconds.

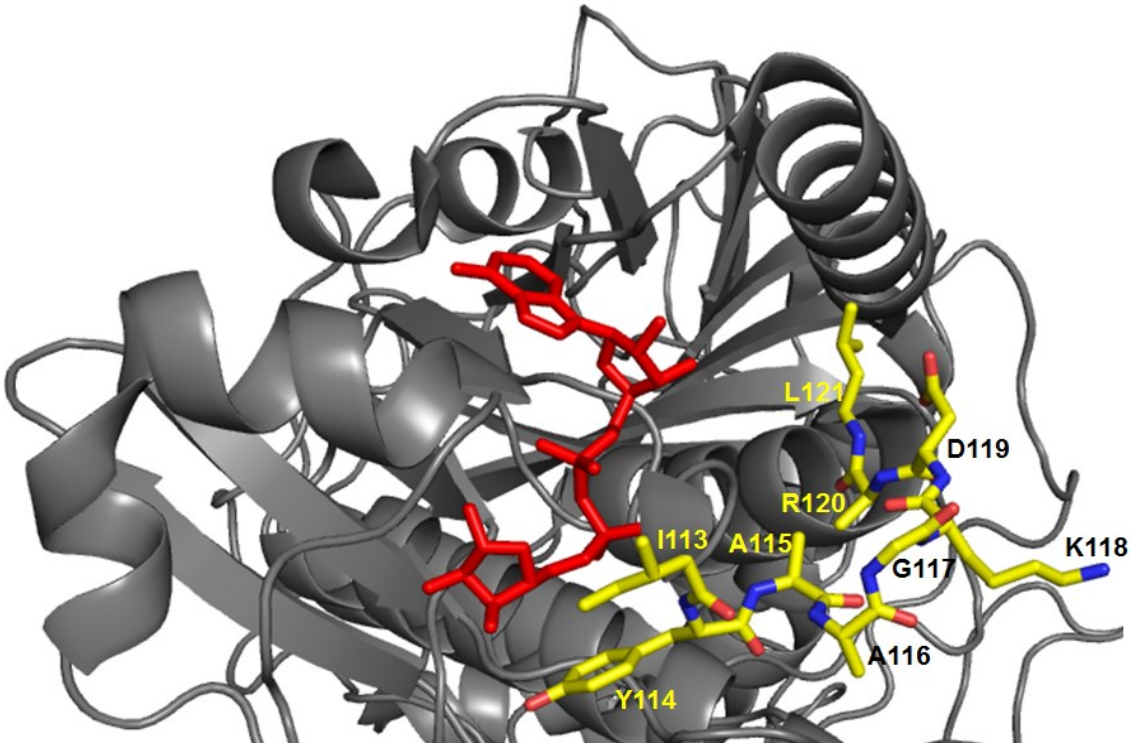

Figure 23. ADP-Ribose Binding Site and Flexible Macrodomain Loop. SINV $P23^{pro\text{-}zbd}$ is shown in grey ribbon format with the flexible loop of nsP3 macrodomain highlighted in yellow stick format. ADP-ribose in red is overlaid to show the ADP-ribose binding pocket (PDB ID: 3GPO) (88).

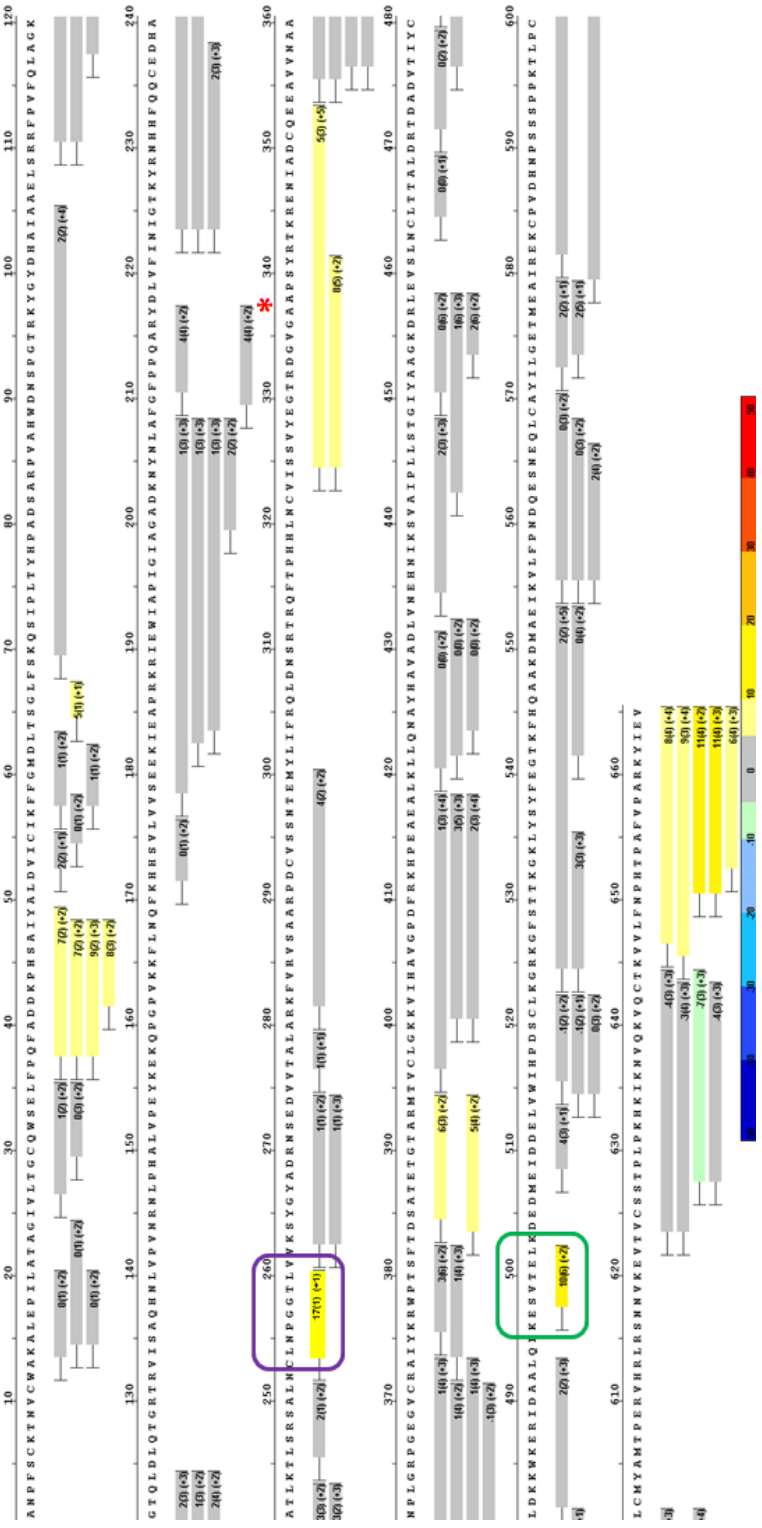

Figure 24. Protein Dynamics Comparison between Wild Type and P726G Mutant. HDX and MS results comparison showing the change in deuterium exchange percent compared to wild type P23$^{pro\text{-}zbd}$. P2/3 cleavage site is noted with a red asterisk. The nsP2 P726G-containing peptide is highlighted with a purple box, and the corresponding nsP3 linker region contact is in a green box.

9. Discussion

Previously described non-cytopathic mutations at the surface of nsP2 contacting the nsP3 linker region appear by molecular modeling to destabilize P23 interaction, decreasing efficiency of the RNA replication complex during virus infection. Altering the corresponding nsP3 residue (L165), which contacts nsP2 P726 did not produce virus with the same non-cytopathic phenotype. The actions of mature nsP2 alone are likely responsible for the non-cytopathic effect in mammalian cells. However, in contrast to the nsP2 P726G and P726S mutants, nsP3 L165 mutants had enhanced RNA replication levels, indicating an apparent stabilization of the replication complex. Despite the large, bulky nature of a tryptophan substitution at L165, the virus remained similar in phenotype to wild type. The L165D and L165E mutants exhibited the most extreme phenotype in terms of reduced RNA infectivity and inefficient P23 cleavage. The surface of nsP2 surrounding P726 is both basic and hydrophobic (Figure 12C), suggesting that the P23 interface may be stabilized by the introduced acidic amino acid at L165 through interaction with the surface of nsP2. Given the increased flexibility of the P726G mutant $P23^{pro\text{-}zbd}$ and lack of P2/3 cleavage inefficiency, the stabilization of the L165D or L165E mutants could cause the interface to become “locked in”, preventing the flexibility needed for conformational change and thus disallowing nsP2 access to the P2/3 cleavage site.

The build-up of uncleaved P23 is evidence that the nsP3 linker region, which encircles nsP2, may function in positioning and recognition of the P2/3 cleavage site. Given the nsP3 linker residues and nsP2 P726 are within reasonable distance to the P2/3 cleavage site (21 Å and 17 Å from C_α of L165 or P726 to the cleavage site, respectively), it is plausible that mutation of the nsP3 linker region shifts local structural features altering recognition of the cleavage site (Figure 12C). A previously described E163R mutation in SFV nsP3 which altered P2/3 cleavage is

completely solvent exposed (Figure 12B) (51). These results may indicate that the nsP3 linker residues surrounding nsP2 P726 are recognized by the nsP2 protease or another factor necessary for P2/3 cleavage.

Section III. Chikungunya Virus P23$^{pro\text{-}zbd}$ Structure

The CHIKV protease is able to recognize peptide substrates as small as nine amino acids. This characteristic is unique compared to SINV and SFV which require the first 170 amino acids of nsP3 for efficient P2/3 cleavage (58). Therefore, there must be differences in substrate specificity and promiscuity between alphavirus species, even within these closely related Old World alphaviruses. Additionally, nsP2 proteases are not always interchangeable between species; the SINV protease is able to process SFV polyprotein, but not vice versa (51). The factors contributing to substrate specificity and promiscuity of alphavirus proteases are not fully understood. Further studies of other nsP2 protease active site structures and P2/3 cleavage sites may show differences responsible for substrate specificity preferences.

After the recent outbreaks in the Americas, and predicted future spread, our attention turned to CHIKV as the focus of our next polyprotein structure. There are currently no vaccines or specific treatments for alphaviruses; with a high resolution structure, structure-based drug design can proceed in the hopes of creating novel anti-viral therapies targeting alphavirus polyprotein processing.

1. CHIKV P23$^{pro\text{-}zbd}$ Expression and Purification

A portion of the CHIKV P23 precursor protein from the nsP2 protease through the nsP3 ZBD was expressed in bacteria (termed P23$^{pro\text{-}zbd}$; domain organization pictured in Figure 5) with an N-terminal six-histidine and small ubiquitin-like modifier (SUMO) tag (134). This construct was designed to be homologous to the SINV P23$^{pro\text{-}zbd}$ structure (68). CHIKV P23$^{pro\text{-}zbd}$ is purified initially via hisTrap column and is shown by SDS-PAGE to be fully intact protein (Figure 25A). Further purity of the protein is achieved via hydroxyapatite and heparin chromatography (Figure 25B). CHIKV P23$^{pro\text{-}zbd}$ also co-purifies with nucleic acid similar to the SINV protein.

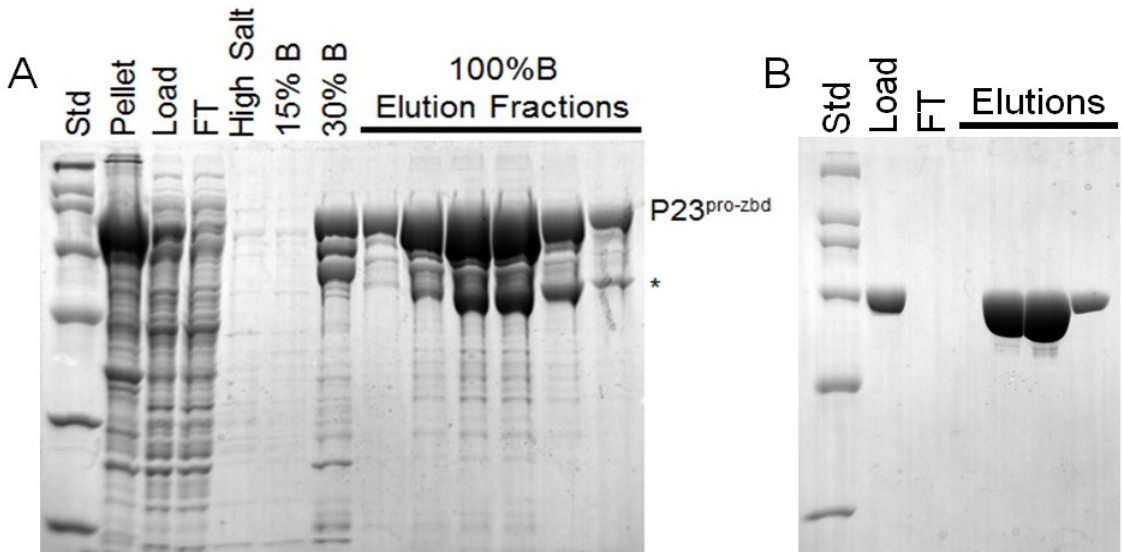

Figure 25. Purification of CHIKV P23$^{pro\text{-}zbd}$. (A) This protein is initially purified from crude bacterial cell lysate using HisTrap column. Bacterial lysate is first loaded onto the column (FT) and the column is washed with high salt buffer to release nucleic acids. Bound SUMO- P23$^{pro\text{-}zbd}$ is eluted with increasing concentrations of imidazole. A P23 degradation product is denoted with an asterisk (*). (B) The final step of purification over heparin column gives very pure CHIKV P23$^{pro\text{-}zbd}$.

2. CHIKV P23$^{pro\text{-}zbd}$ Structure Determination and Comparison

CHIKV P23$^{pro\text{-}zbd}$ crystals diffracted to 2.20 Å at Stanford Synchrotron Radiation Lightsource. The structure was determined by molecular replacement with SINV P23$^{pro\text{-}zbd}$ (Figure 26A; Table 3) (68). There are four molecules of P23$^{pro\text{-}zbd}$ in each asymmetric unit as well as 15 glycerol and 1098 water molecules from the crystallization solution. At this resolution, water molecules could also to be modeled within and around the protein (Figure 26B). The previous SINV P23 structure at 2.8 Å led us to believe this protein was fairly rigid and compact; however, with the increased resolution presented here, there are several pockets of solvent-accessibility within the protein itself, especially between the macrodomain and the rest of the protein. The two proteins share 50% sequence identity and their overall root mean square deviation (RMSD) is 1.55 Å (Figure 27). Between the two species, the protease domains differ

the most with an RMSD of 0.98 Å. The nsP2 MT-like domains, and nsP3 macrodomains and ZBDs have RMSDs of 0.98 Å, 0.63 Å, and 0.72 Å, respectively. The RMSD of the two molecules as a whole accounts for the orientation of each domain in relation to each other, as well, giving a higher difference due to a shift away from the rest of the molecule by the CHIKV macrodomain compared to SINV (Figure 27).

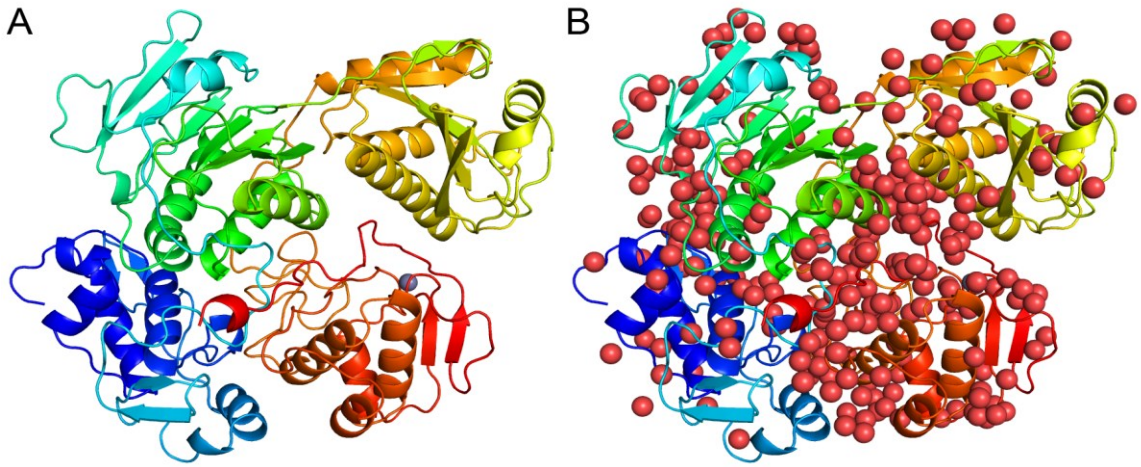

Figure 26. Structure of CHIKV P23$^{pro\text{-}zbd}$ Ribbon Model. CHIKV P23$^{pro\text{-}zbd}$ structure in rainbow coloring from N-terminus (blue) to C-terminus (red) without (A) and with (B) solvent molecules shown as red spheres. Zinc is shown as a grey sphere in both panels.

Table 3. Data Collection and Refinement Statistics for CHIKV P23$^{pro\text{-}zbd}$.

Data Collection	
Space group	$P2_1$
Cell dimensions	
a, b, c (Å)	90.6, 106.0, 145.7
α, *β*, *γ* (°)	90.0, 104.4, 90.0
Resolution (Å)	23.35-2.20 (2.24-2.20)
R_{merge}	0.092 (0.471)
R_{pim}**	0.086 (0.429)
$I/\sigma I$	7.7 (2.3)
Completeness (%)	95.5 (95.9)
Redundancy	3.2 (3.2)
Refinement	
Resolution (Å)	23.35-2.20 (2.23-2.20)
No. reflections	124421 (3632)
R_{work}/R_{free}	0.19/0.24
No. atoms	
Protein	19773
Ligand/ion	1098
B-factors	
Protein	31.82
Ligand/Ion	39.19
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.582
Ramachandran Plot	
Most favored (%)	96.5
Allowed (%)	3.5
Disallowed (%)	0

*Highest resolution shell is shown in parenthesis.

** $R_{p.i.m.} = \sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_i \left| I_i(hkl) - \overline{I(hkl)} \right| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ = Intensity of the *i*th observation of the reflection *hkl* and *N* is the redundancy of multiplicity of the observed reflection, calculated using the program SCALA (104).

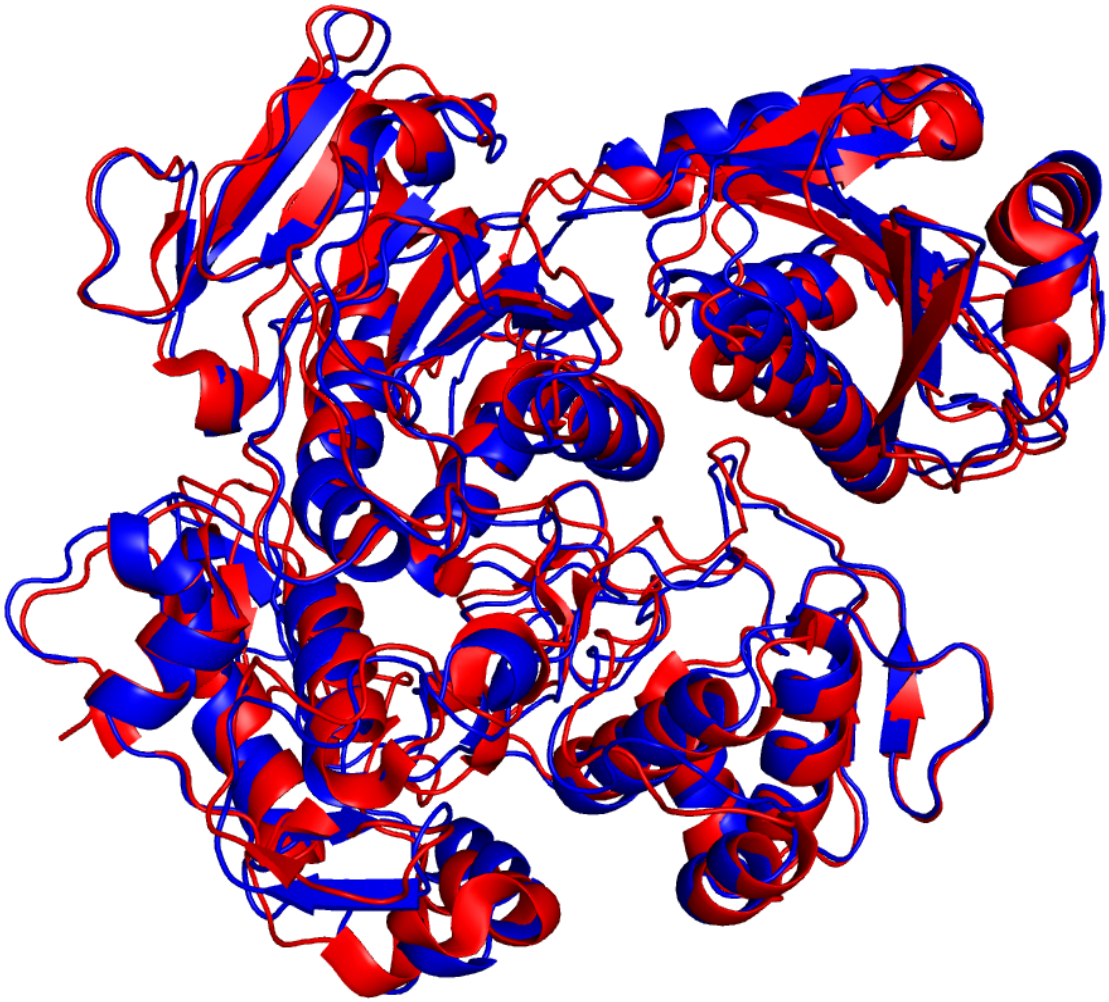

Figure 27. Overall Comparison of SINV and CHIKV P23$^{pro\text{-}zbd}$ Structures.
Overlay of the SINV structure (PDB ID: 4GUA) shown in red and CHIKV structure in blue.

3. P2/3 Cleavage Site Flexibility

Three of four CHIKV P23$^{pro\text{-}zbd}$ molecules in the asymmetric unit lack electron density for residues immediately preceding the P2/3 cleavage site, indicative of inherent flexibility. The P23 molecule with density in that area is near a symmetry mate, likely stabilizing this region via crystal packing. In contrast, all three of the molecules in the SINV P23 asymmetric unit had very well defined density in the cleavage site area. Comparing the one CHIKV molecule with cleavage

site density and the SINV site shows that these two are very different with an RMSD of 1.6 Å (Figure 28).

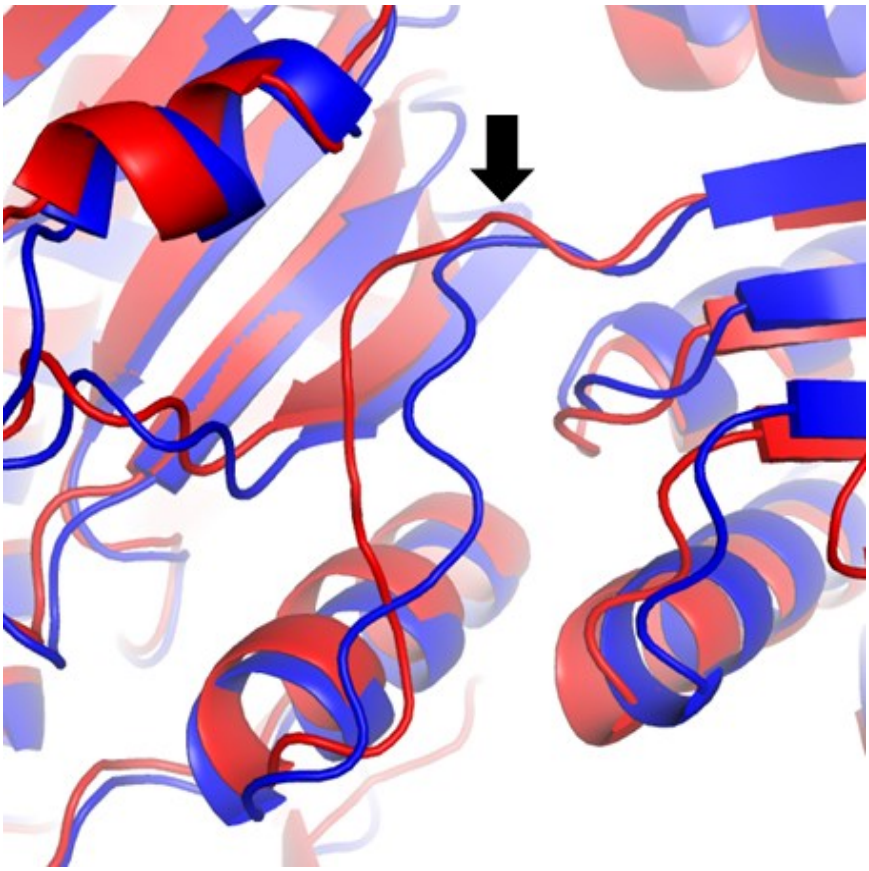

Figure 28. CHIKV and SINV P2/3 Cleavage Site Comparison. A view looking down into the cleft where the P2/3 cleavage site is located. Overlay of the strand leading into the P2/3 cleavage site. SINV structure (PDB ID: 4GUA) is shown in red and CHIKV structure in blue. The P2/3 cleavage site is denoted with an arrow.

4. nsP2 Protease Loop

The protease domains of CHIKV and SINV nsP2 have the highest RMSD out of the four domains compared. This is due to a loop located near the catalytic site of the protease (Figure 29). Comparing thee three known alphavirus protease domain structures shows that this loop is variable between species. Not only does SINV contain a larger loop, the shorter loops of VEEV

(PDB ID: 2HWK) and CHIKV are angled such that there is a 3.9 Å distance between them (Figure 30A-B) (53, 68). Additionally, the loop itself has variable sequence conservation across these species (Figure 31).

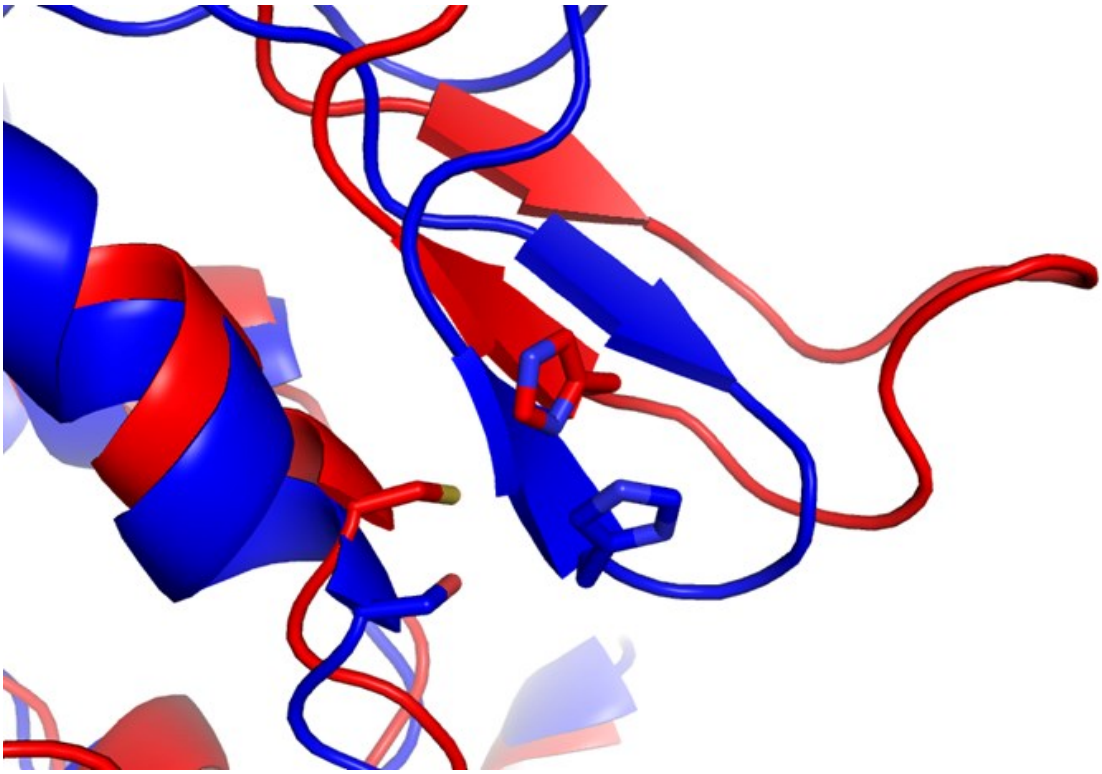

Figure 29. SINV Extended nsP2 Protease Loop. Overlay of CHIKV (blue) and SINV (red) nsP2 protease domains shows an extended loop near the protease active site that may be responsible for substrate specificity. Catalytic dyad residues cysteine and serine are also shown.

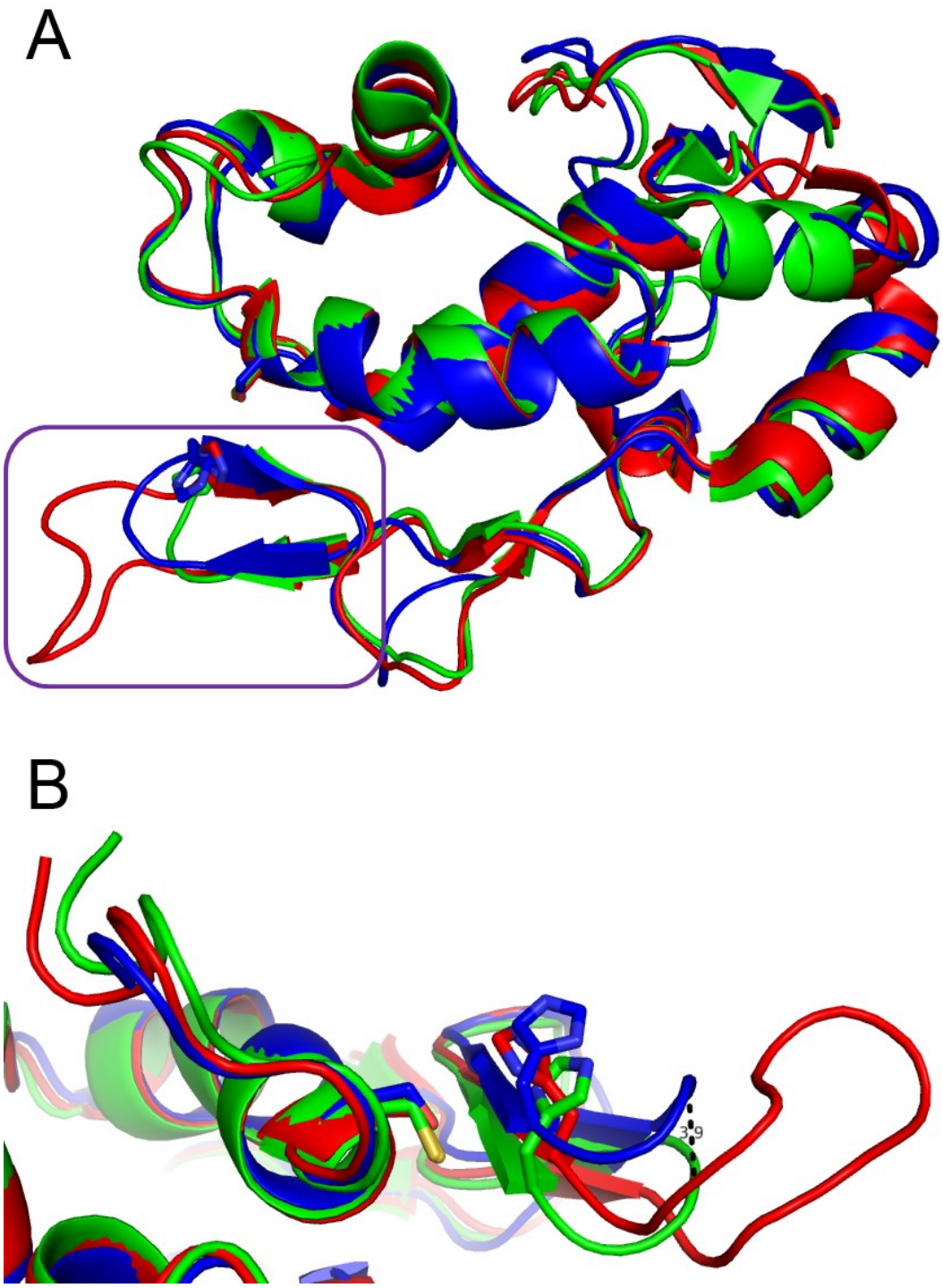

Figure 30. nsP2 Protease Loop Comparison Between Species. Structures of protease domains from three alphaviruses superimposed in ribbon format. Catalytic dyad residues are shown. The arm of the domain containing the active site is highlighted with a purple box in (A). An alternate and zoomed view of the protease loop-active site region is shown in (B) with 3.9 Å measurement between the CHIKV (blue) and VEEV (green) loops. SINV protease is depicted in red.

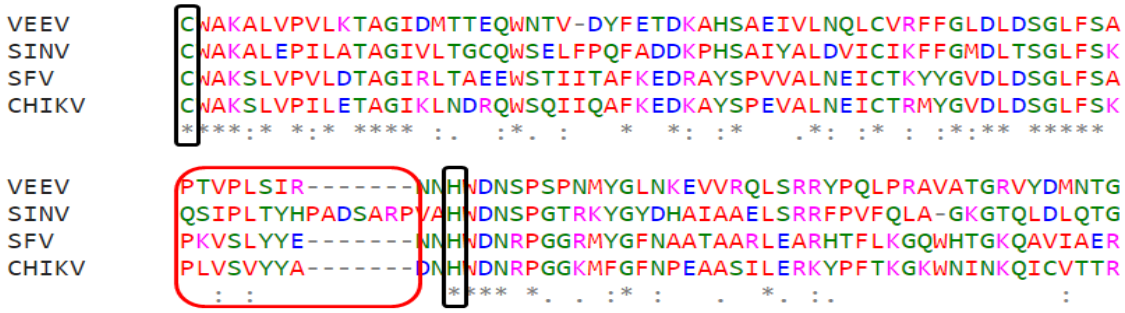

Figure 31. Sequence Alignment of Protease Loop. Sequence alignment with VEEV (Accession: AAU89533.1), SINV (Accession: NP_740671.1), SFV (Accession: ABP58684.1), and CHIKV (Accession: ADZ47896.1) nsP2 protease loop with the catalytic cysteine and histidine residues denoted in a black box and the protease loop in a red box.

5. Modeling of Accessible Cavities for Drug-like Fragment Screening

Structure-based drug design is an important method to predict possible ligand binding pockets of a protein. *In silico* screenings of massive compound libraries minimizes time and cost necessary for high throughput trials. This drug discovery method has been used to predict compounds *de novo* or to improve current drug-like compounds to successfully treat a variety of diseases from cancers to viral and fungal infections (135-137). The high resolution structure of CHIKV $P23^{pro\text{-}zbd}$ allows us to begin designing small molecule inhibitors in the hopes of creating novel anti-viral therapies.

AutoLigand was used to predict all targetable pockets of $P23^{pro\text{-}zbd}$ (138). This program identifies the optimal volume of pockets in order to maximize ligand interaction and predicts an ideal ligand for that pocket. Several pockets overlap with the water channels within the protein that our structure revealed. Pockets of interest include sites within the ADP-ribose binding site, protease active site, and near the P2/3 cleavage site. To confirm ligand binding pocket availability predicted *in silico*, the protein was screened via activity-based protein profiling (ABPP) and a library of click chemistry probes (139, 140). Initial data confirms two cysteines of

particular interest are accessible for ligand binding: nsP2 C798, which is the last residue of nsP2, and nsP3 C24, located in the ADP-ribose binding site.

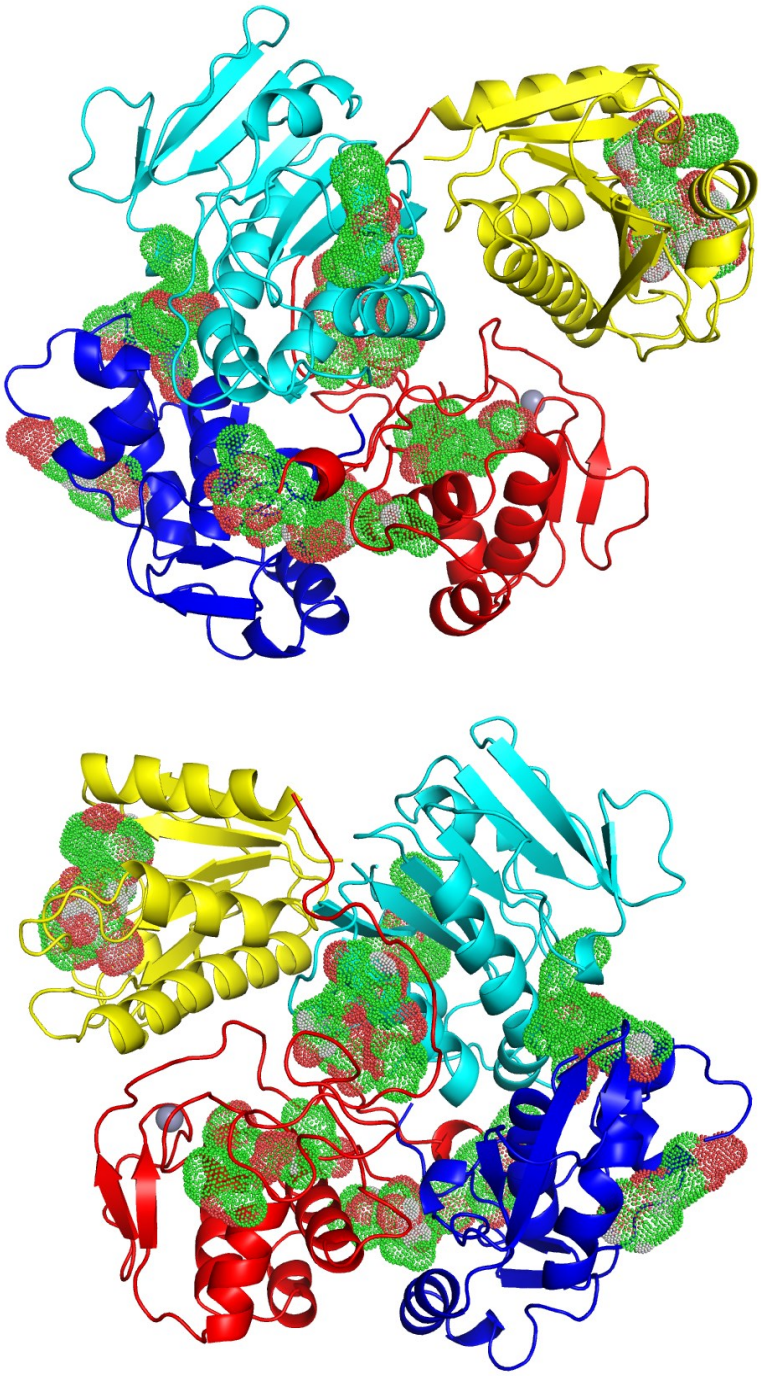

Figure 32. Predicted Ligand-Binding Pockets of CHIKV P23$^{pro\text{-}zbd}$. Ribbon representation of CHIKV P23 with domain coloring as in Figure 5. Predicted ligand-binding targets shown as shells calculated by AutoLigand (138). Shell coloring corresponds to interaction type: green is hydrophobic, red is hydrogen bond acceptor, and white is hydrogen bond donor. Bottom panel is a 180° rotation about a vertical axis. Zinc ion is pictured as a grey sphere.

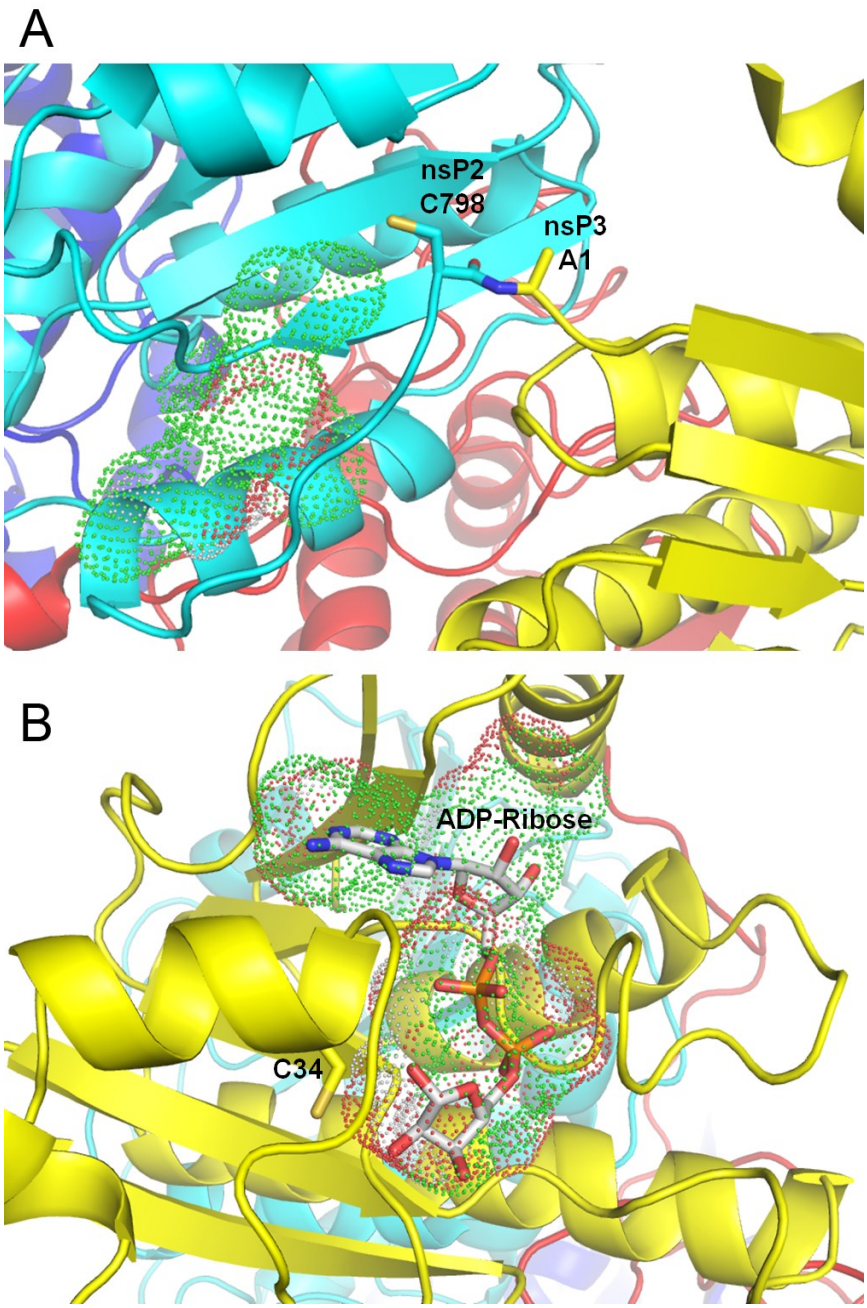

Figure 33. Confirmed Accessible Cysteines for Ligand Binding Targets. Ribbon representation of CHIKV P23 with domain coloring as in Figure 5. Predicted ligand-binding targets shown as shells calculated by AutoLigand (138). Shell coloring corresponds to interaction type: green is hydrophobic, red is hydrogen bond acceptor, and white is hydrogen bond donor. Activity-based protein profiling with a library of click chemistry compounds identified (A) nsP2 cysteine 798, the last residue before the P2/3 cleavage site and (B) nsP3 C34, in the ADP-ribose binding pocket, as accessible. ADP-ribose is overlaid in the macrodomain binding pocket for reference.

6. Discussion

Alphavirus P2/3 cleavage sites likely have varying degrees of accessibility complementary to the protease substrate specificity needs. Perhaps the SINV nsP2 requires an extra-long loop to recognize a more buried SINV P2/3 cleavage site, and the CHIKV or VEEV nsP2 does not because their sites are more accessible in general. Additionally, it has been previously shown that SINV is able to process SFV polyproteins, but SFV cannot cleave SINV polyproteins (51). This may be due to the longer loop of SINV which is able to access to the SFV cleavage site. According to sequence alignment, SFV lacks this loop and therefore, according to our hypothesis, could be a reason why it cannot process SINV polyproteins. Additional sequence analysis indicates that only viruses in the SINV branch (i.e. SINV, Babanki virus, Ockelbo virus, A.A.AR86 virus, Girdwood virus, etc.) of the Western Equine Encephalitis (WEE) Complex contains this loop. Other very closely related viruses to SINV within the WEE Complex such as WEEV and Buggy Creek virus do not contain an extended the loop, nor do alphaviruses outside of the WEE Complex (141). Further research is necessary to further describe the function this protease loop serves. Overall, it seems that within alphaviruses there is a two-fold approach to substrate specificity: the accessibility and flexibility of the cleavage site and the structure of the loop adjacent to the protease active site.

The higher resolution of the CHIKV P23 structure allows us to model solvent accessible channels within the protein itself, giving insight into how densely packed the molecule actually is. *In silico* experiments suggest several pockets that could be targeted for ligand binding, including areas of interest near the ADP-ribose binding site and the P2/3 cleavage site. Drug-like molecules are currently being designed for testing in X-ray crystallography experiments. These molecules, once confirmed in binding assays, may be used as a basis for further computer-based modeling to increase binding affinities. Through several rounds of modeling and

experimentation, we hope to eventually design the first effective anti-viral therapy for alphaviruses.

Section IV. Biophysical and Virological Studies of nsP3 Macrodomain

nsP3 is the least understood protein of alphaviruses. A longstanding goal of our project is to further characterize the activities of nsP3. As our CHIKV P23 protein structure indicated the strong potential of the nsP3 macrodomain ADP-ribose binding site for drug targeting, we sought to further the knowledge of this domain using both biophysical and virological studies.

Macrodomains are highly conserved and found across all domains of life (80-83). As described earlier, human macrodomain containing proteins mainly localize to the nucleus and participate in transcriptional regulation via alteration of chromatin structure (84-87). Alphavirus nsP3 does not localize to the nucleus so the function of the alphavirus macrodomain still remains elusive in the cytoplasm. There is limited data in the literature regarding the function of nsP3 macrodomain both *in vivo* and *in vitro*. Residues N10 and N24 of SINV nsP3 macrodomain have been shown to be necessary for neurovirulence in mice (81). Some alphavirus macrodomains have been shown to bind DNA, RNA, and poly(ADP-ribose) with varying affinities and have relatively poor ADP-ribose-1'' phosphatase activity compared to yeast and mammalian macrodomains (83, 88, 89). CHIKV and VEEV macrodomains have binding efficiency K_ds of 5 ± 0.4 µM and 3.9 ± 0.65 µM respectively with ADP-ribose. Mutation of the CHIKV D10 residue abolishes ADP-ribose binding (88). Conversely, SFV macrodomain shows absolutely no binding to ADP-ribose, but does bind to poly(ADP-ribose) (89).

Our studies focused on SINV due to lack of data on the macrodomain from this alphavirus and the ease of tissue culture handling of the virus.

1. Macrodomain Mutations are Polyprotein Cleavage Competent

Due to previous evidence that the macrodomain is involved in position the P2/3 cleavage site, it was reasonable to first check mutations in the ADP-ribose binding domain for

competency of polyprotein cleavage. We chose to initiate our study with two main nsP3 mutants: H67A and double mutant NN21/24AA. These are implicated in viral replication or ADP-ribose binding (81, 142). Mutations were made in infectious SINV cDNA clone pTOTO1106 and transcribed into RNA for transfection into BHK cells. Western blotting shows no indication of cleavage deficiencies post-infection in BHK cells, but does show a delay in nsP2 production in the H67A mutant (Figure 34). Likely the nsP3 linker region is the only portion of nsP3 which directly affects P23 cleavage efficiency, not the ADP-ribose binding pocket.

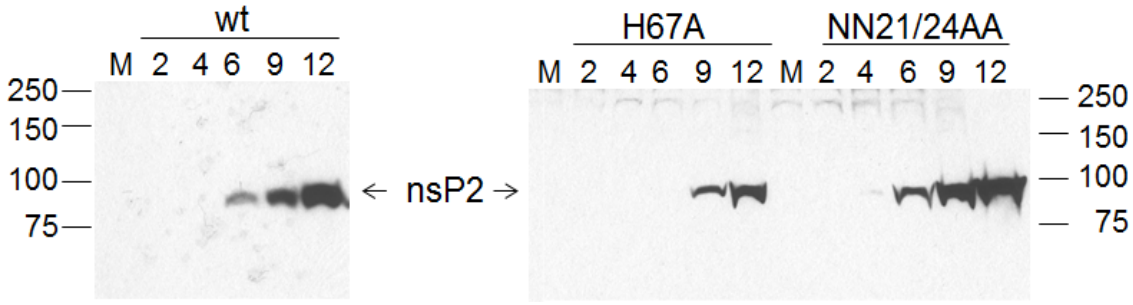

Figure 34. Macrodomain Mutants Cleave Polyproteins Efficiently. Analysis of P2/3 cleavage *in vivo* via western blot on total protein harvested from BHK cells infected with wt or mutant SINV at hour post infection. Membranes were probed using anti-nsP2 antibodies. M indicates mock infected cells as a negative control.

2. The Limitations of Plaque Assay

Previous evidence suggests that the functions of nsP3 may be host-cell dependent. Therefore, we sought to measure viral replication in multiple mammalian and insect cell types. Alphavirus studies are regularly performed in BHK cells, which lack a competent interferon response to infection, thus allowing the virus to replicate to high titers (143-147). We selected an interferon competent cell line, L929, to compare with (148). Initial plaque assay results show a clear difference in plaque morphology of wild type SINV as measured on BHK versus L929 cells (Figure 35). On BHK cells, wild type plaques are large and clear, whereas on L929 cells, they are

very small. Furthermore, for the nsP3 mutant H67A, individual plaques are indistinct and cannot be reliably counted on the L929 cell line, although a reduction in measured viral titer can be inferred (Figure 35). The double nsP3 mutant NN21/24AA has a similar plaque size as wild type SINV on both cell lines (data not shown). This method of measuring viral replication levels is inherently flawed due to its reliance on cytopathogenesis and plaque morphology.

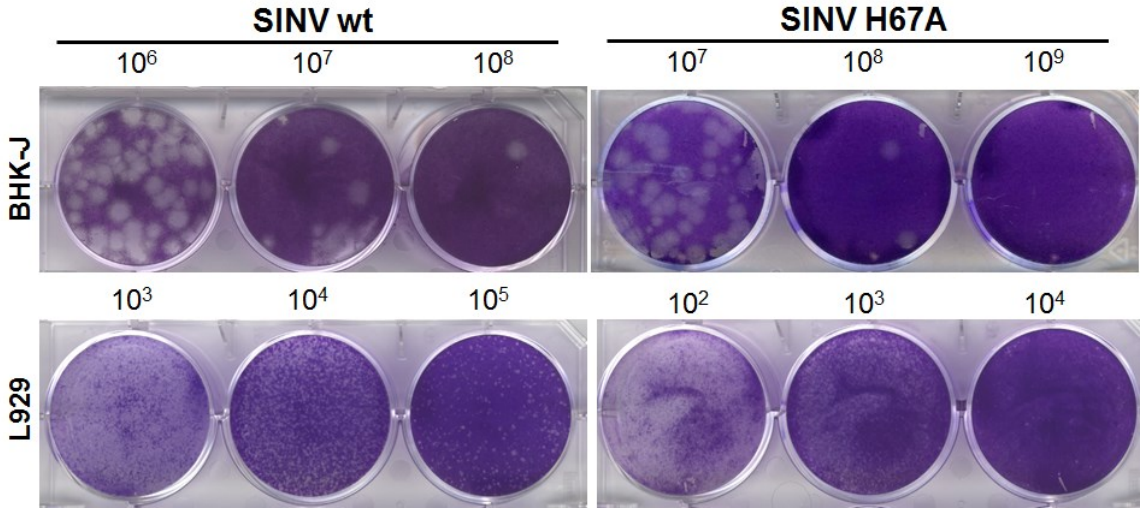

Figure 35. Plaque Assay of SINV on BHK-J and L929 Cell Lines. Plaque assay stained with crystal violet of SINV wt and nsP3 H67A mutant on BHK-J (top) and L929 (bottom) cells. Dilution factor is indicated above each well.

3. Functions of Macrodomain May be Host Dependent

To further pursue the difference seen between SINV macrodomain mutant infection in alternate cell types independent of the plaque assay, we used a SINV infectious cDNA clone pTE/5'2J designed with a second subgenomic promoter expressing firefly luciferase between the nonstructural and structural ORFs (114-116). This gives the advantage of being able to measure luciferase production via luminescence assay throughout infection which directly correlates to levels of viral replication without relying on cytopathogenicity necessary in the plaque assay. Our pTE/Luc design is also a higher throughput system, allowing the screening of many cell types

and SINV mutants simultaneously. This allowed us to expand our macrodomain mutants to include several other residues of possible interest.

pTE/Luc constructs with nsP3 macrodomain mutations were: NN21/24AA, G32E, C34A, and H67A (Figure 36). N21/N24, and H67 have been previously studied as described above in other viruses. Although H67 is not in the ADP-ribose binding pocket, it is in the hydrophobic core of the macrodomain. We chose to mutate C34 of macrodomain due to its proximity to the ADP-ribose binding pocket and the fact that cysteine residues often indicate an enzymatic active site. Lastly, the G32 residue was chosen due to new data from human macroD2 protein. Mutation of G100E of human macroD2 protein prevents binding to ADP-ribosylated proteins (149). The G100 residue correlates to G32 in SINV macrodomain. All of these residues are highly conserved throughout alphaviruses.

It is thought that the contribution of nsP3 to host specificity is mainly a function of the C-terminal HVR (76). Our findings show a contribution from the macrodomain as well, specifically the ADP-ribose binding site. In interferon-deficient BHK cells, all macrodomain mutants replicate to very high levels on par with the wild type SINV virus (Figure 37A). NN21/24AA mutant replicates at even higher rates than wild type up to six hours post-infection. The same is true for interferon competent L929 cells (Figure 37B). The G32E mutant replicates less efficiently early in infection in L929s, but ultimately reaches wild type levels by twelve hours post infection. Two mosquito cell lines were used in this analysis, both from *Aedes albopictus*. C7-10 and C6/36 *A. albopictus* cell lines were both derived from larval isolates (150). Viral growth in both mosquito cell lines shows moderate deficiencies in NN21/24AA and severe deficiencies in G32E mutant replication (Figure 37C-D). In C7-10 cells, NN21/24AA and G32E mutants reach 55% and 13% of wild type levels, respectively, by twenty four hours post-infection. In C6/36 cells, NN21/24AA and G32E mutants replicated to 64% and 0.25%.

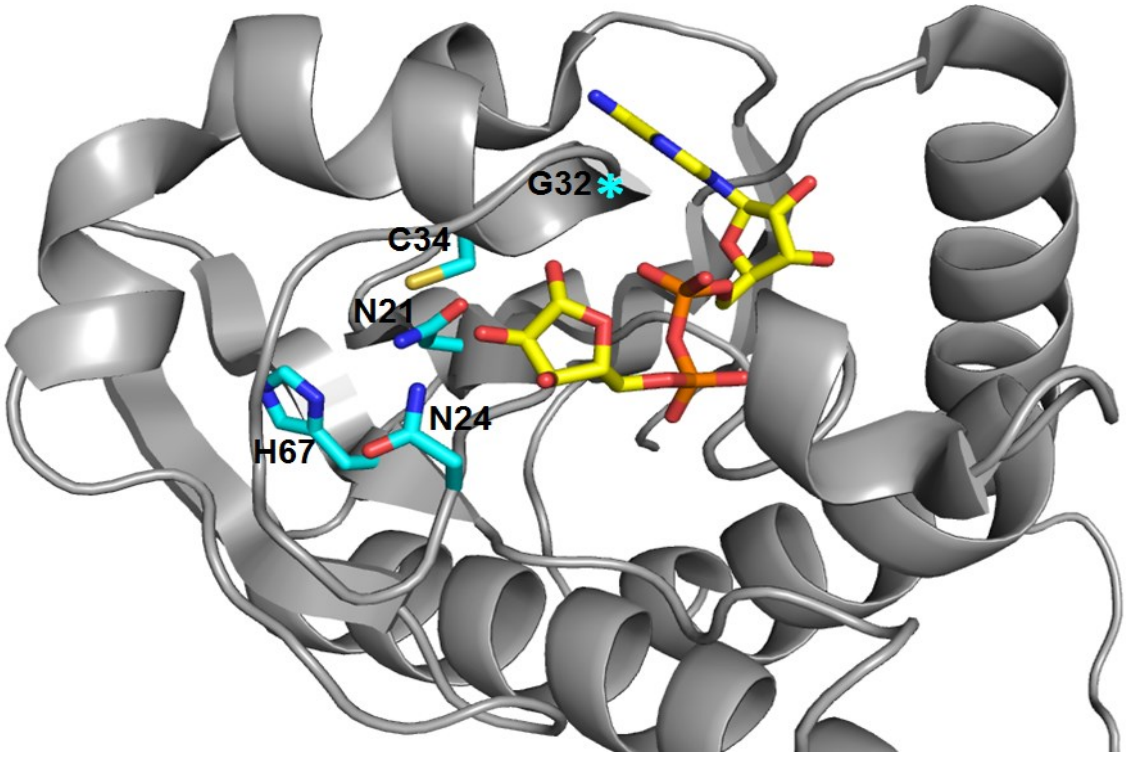

Figure 36. nsP3 Macrodomain ADP-Ribose Binding Pocket. SINV nsP3 (PDB ID: 4GUA) macrodomain is shown in grey ribbon format with ADP-ribose (PDB ID: 3GPO) overlaid in the binding pocket. Side chains of nsP3 residues of interest are labeled are shown in cyan, with the glycine location as an asterisk.

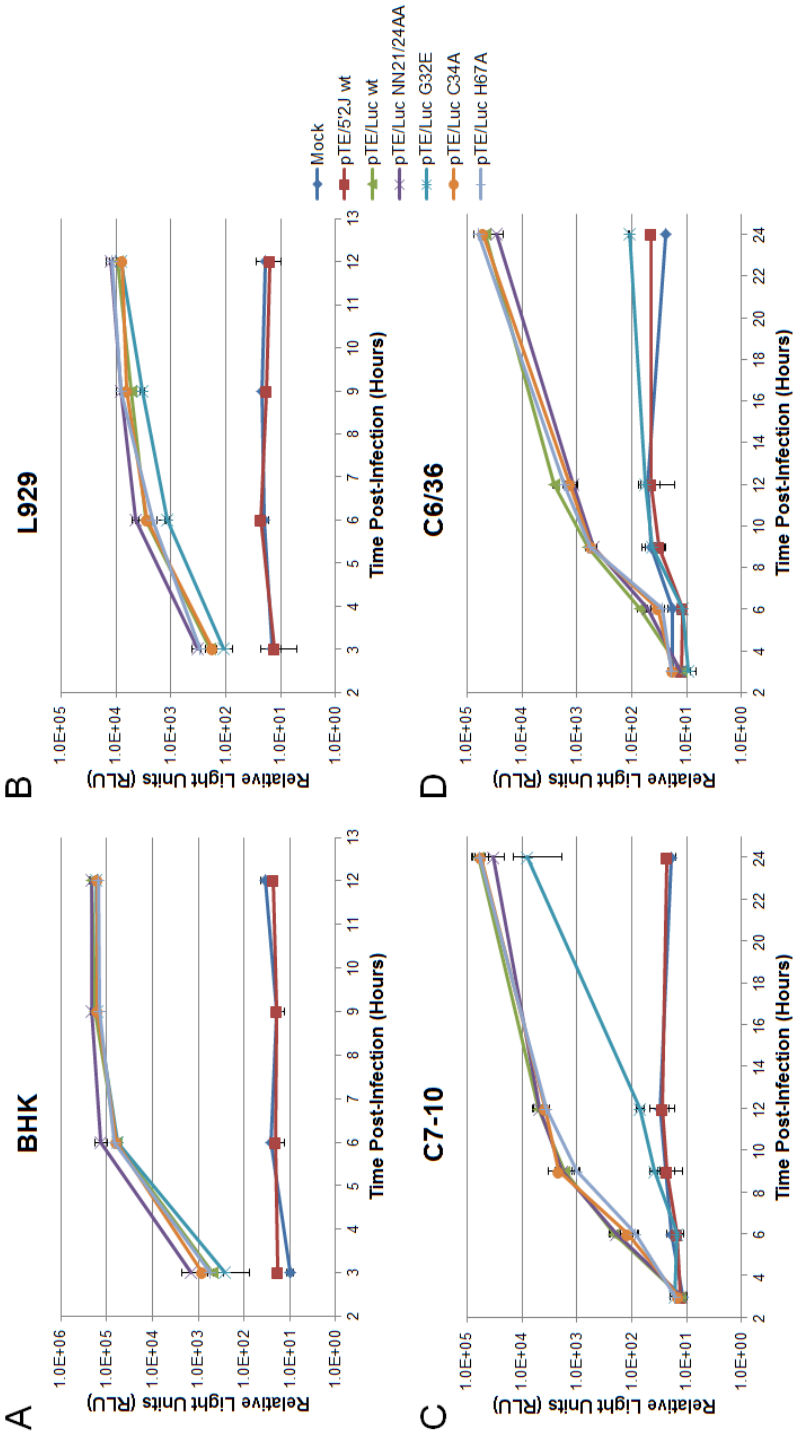

Figure 37. Replication of SINV Macrodomain Mutants in Mammalian and Insect Systems. Luciferase production was measured in relative light units (RLU) of SINV cells infected at MOI 10 at time points post infection. Mock and SINV replicon without the luciferase gene (pTE/5'2J wt) were negative controls. Mammalian cell lines were BHK (A) cells and L929 cells (B). *Aedes albopictus* cell lines C7-10 (C) and C6/36 (D) were also infected. Error bars represent standard error of the mean.

4. ADP-Ribose Binding of SINV Macrodomain Mutants

SINV nsP3 macrodomains have not been previously measured for ADP-ribose binding affinity. As our experiments indicate that the macrodomain (especially G32) may be important in host specificity, it is important to examine whether this phenotype correlates with ADP-ribose binding. First, nsP3 macrodomain wild type, G32E, and H67A mutants were purified as GST fusion proteins similarly to SINV $P23^{pro\text{-}zbd}$ shown previously (Figure 38). All proteins co-eluted with nucleic acids, indicating that the ADP-ribose binding site may not be necessary for RNA binding. Macrodomain wild type, G32E, and H67A expressed at high levels and were gel filtered before ITC experiments.

Given our previous finding implicating G32 as an important residue for host specificity, we measured the ADP-ribose binding of wild type and mutant SINV macrodomain proteins via ITC. Wild type SINV macrodomain has a K_d of 8.8 ± 0.4 µM for ADP-ribose. H67A mutant showed a ten times lower affinity at 91.7 ± 18.3 µM (Figure 39). Lastly, G32E had no detectable binding to ADP-ribose by ITC. NN21/24AA protein is still in analysis.

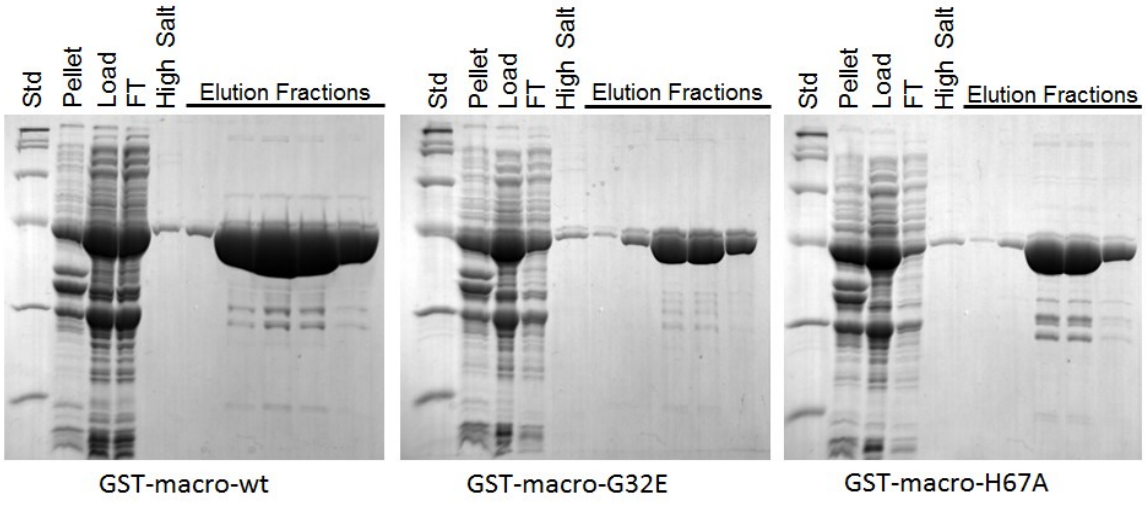

Figure 38. SINV Macrodomain Purification. SDS-PAGE analysis of the purification of GST-macrodomain proteins. Cell lysate was loaded onto the

column and flow through (FT) of unbound proteins was collected. The column was washed with buffer containing 1M KCl to elute bound nucleic acids before protein elution.

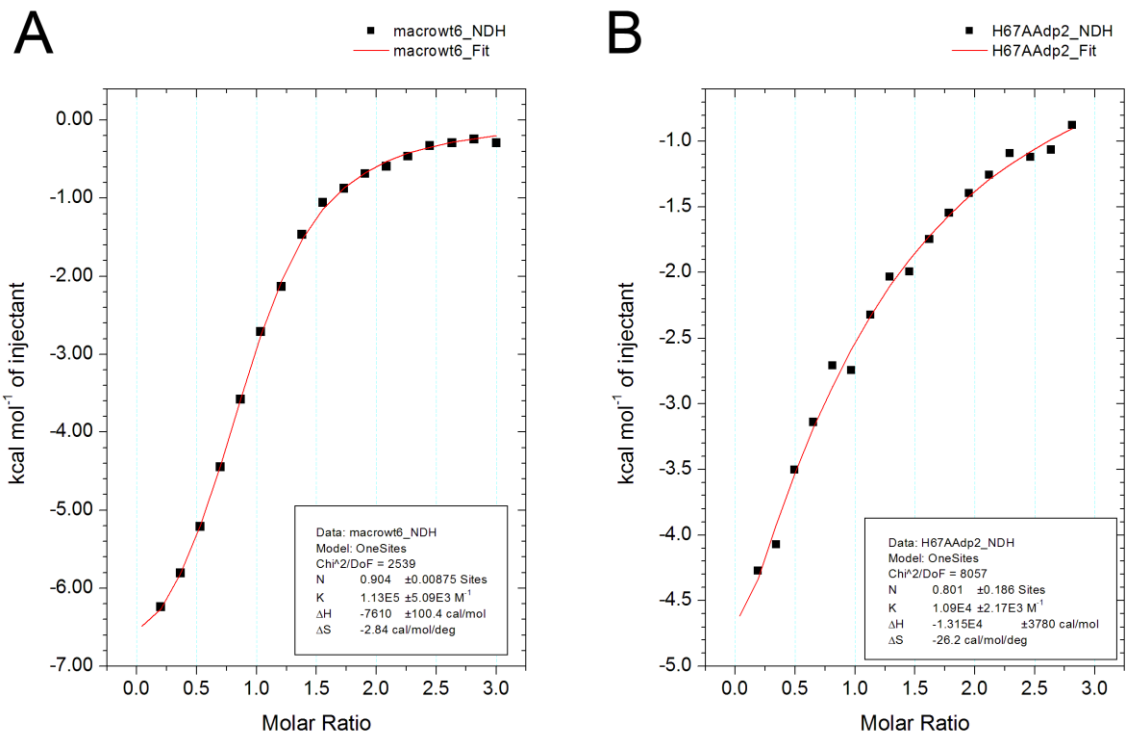

Figure 39. ADP-Ribose Binding of SINV Macrodomain. Isothermal titration calorimetry profile for binding of ADP-ribose to SINV nsP3 macrodomain wild type (left) and H67A (right).

5. Discussion

nsP3 is the least understood of the alphavirus nonstructural proteins. Studies show variable binding affinities for ADP-ribose, poly(ADP-ribose), and polyA as well as poor ADP-ribose-1'' phosphatase activity. The significance of these findings is unknown, and the lack of consistency in the literature with regard to alphavirus species and mutations analyzed makes it difficult to assess the importance of this domain in alphavirus infection. Mutation of residues in the ADP-ribose binding pocket (NN21/24AA) or near the pocket but not directly interacting with ADP-ribose (H67A) do not affect polyprotein cleavage efficiency during infection. Many of the

mutations we made in the ADP-ribose binding pocket had little or no effect in mammalian systems. N24 of macrodomain is necessary for neurovirulence in mice, but in our mammalian systems, viruses with NN21/24AA mutations grow at wild type levels, or better than wild type at early hours post-infection. While H67A showed a delayed production of nsP2 by western blotting, but there was minimal negative effect in replication in mammalian and insect cells via luciferase assay.

Mutations in the ADP-ribose binding site did not prevent the co-purification of nucleic acids bound to nsP3 macrodomain after bacterial expression. Combined with ITC data indicating that H67A has lower ADP-ribose binding affinity, and G32E has no detectable binding, this ADP-ribose binding pocket may not be the site of RNA binding. Instead, we hypothesize that nsP3 macrodomain interacts with ADP-ribosylated proteins over the course of infection. Therefore, host cell differences in ADP-ribosylation of proteins in mammalian versus insect cells could account for the differences in viral replication in the G32E mutant between hosts. Experiments are ongoing to measure the ADP-ribose binding affinity of the NN21/24AA macrodomain mutant.

Overall, much more consistent studies are necessary in this area than what is available in current literature to determine the binding partners of alphavirus nsP3 macrodomain during active infection.

Section V. Full Length P23 Polyproteins

The N-terminal portion of nsP2 is required as a "cofactor" for efficient P2/3 cleavage (39, 51). The addition or deletion of a single amino acid residue at the N-terminus of nsP2 reduces P2/3 cleavage efficiency significantly, and the deletion of two residues abolishes cleavage ability (51). It is unknown how the extreme N-terminus of nsP2 participates in efficient P2/3 cleavage, but studies show it must be present in the same molecule as the protease domain to be functional (51).

nsP2 also contains a helicase/NTPase domain. The N-terminal helicase/NTPase domain is theorized to unwind viral RNA secondary structures during replication. Helicase activity of nsP2 is driven by the NTPase activity of two RecA-like domains (59, 60). CHIKV nsP2 cannot unwind dsDNA substrates or dsRNA substrates with a 3' overhang, blunt ends, or a short 5' overhang. Full length CHIKV nsP2 is, however, able to unwind dsRNA with a longer, 12-base 5' overhang (61). The structure of the helicase domain and its position in relation to the rest of the polyprotein is unknown. Additionally, the activities of nsP2 have only been characterized in a post-cleavage, mature form. Post-cleavage activities have not been fully characterized and may be regulated based on polyprotein state.

Our attempts to characterize full length P23 are detailed below in the hopes of understanding how alphaviruses regulate the activities of nonstructural proteins through the use of polyprotein intermediates.

1. Purification of SINV P23 Polyprotein

SINV FL P23 with protease active site cysteine mutation C481A was expressed from the nsP2 helicase domain through the nsP3 ZBD as a fusion with SUMO (Figure 40). Purification initially was performed over HisTrap column and SDS-PAGE analysis revealed many degradation

products contributing to poor yields of FL P23 (Figure 41A). Degradation products were confirmed to be truncations of FL P23 by a gel shift after SUMO tag cleavage. After further purification, FL P23 was determined to be aggregated and unsuitable for further biophysical studies via size exclusion chromatography (Figure 41B).

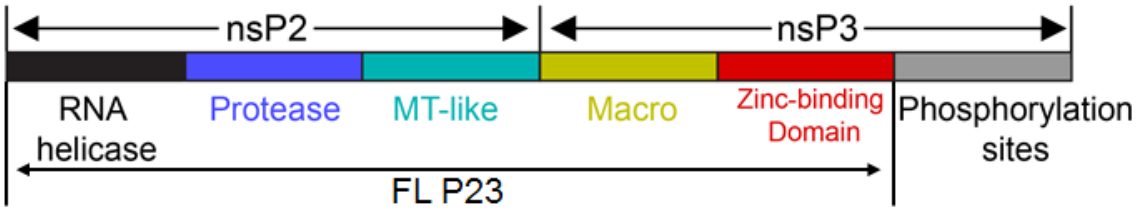

Figure 40. Domain Organization of P23 and FL P23 Expression Construct Boundaries. Expression construct of SINV FL P23 from the nsP2 helicase domain through the ZBD of nsP3.

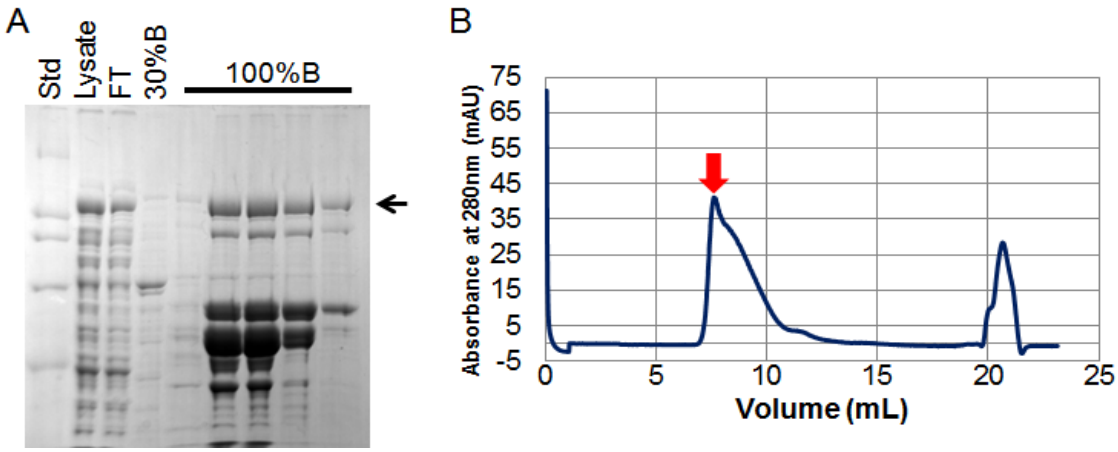

Figure 41. Expression and Purification of SINV FL P23. (A) SDS-PAGE analysis of SINV FL P23 protein purification from HisTrap (GE Healthcare) column. Protein is loaded onto the column (FT) and eluted with increasing concentrations of imidazole (25-250 mM). FL P23 is denoted with an arrow. (B) Chromatogram of absorbance at 280nm of purified FL P23 from size exclusion chromatography. Void volume is denoted with a red arrow.

2. Novel Expression Method for CHIKV nsP2

Further studies with SINV FL P23 were not feasible due to poor protein expression, low solubility, and aggregation; therefore, our focus shifted back to CHIKV proteins. CHIKV FL P23 with an active protease was expressed from the nsP2 helicase domain through the nsP3 ZBD as a fusion with SUMO. Although previous studies indicate that the nsP2 N-terminus must be free to achieve efficient P2/3 cleavage, our construct with an N-terminal SUMO fusion achieved complete cleavage during expression (Figure 42A) (51). Cleaved nsP3 was insoluble. Well folded, monomeric FL CHIKV nsP2 was able to be purified in this manner (Figure 42B). CHIKV, SFV, and VEEV nsP2 have been historically difficult to produce (60, 151, 152). More recently this has been overcome by expressing nsP2 has a fusion with maltose binding protein (MBP) and reliance on glutamic acid and arginine in buffer solutions to improve solubility (58). Expression with a portion of nsP3 ultimately provides higher yield and solubility of FL nsP2 and eliminates the need for MBP and buffer additives. Thus far, attempts to crystalize FL CHIKV nsP2 have produced no crystals.

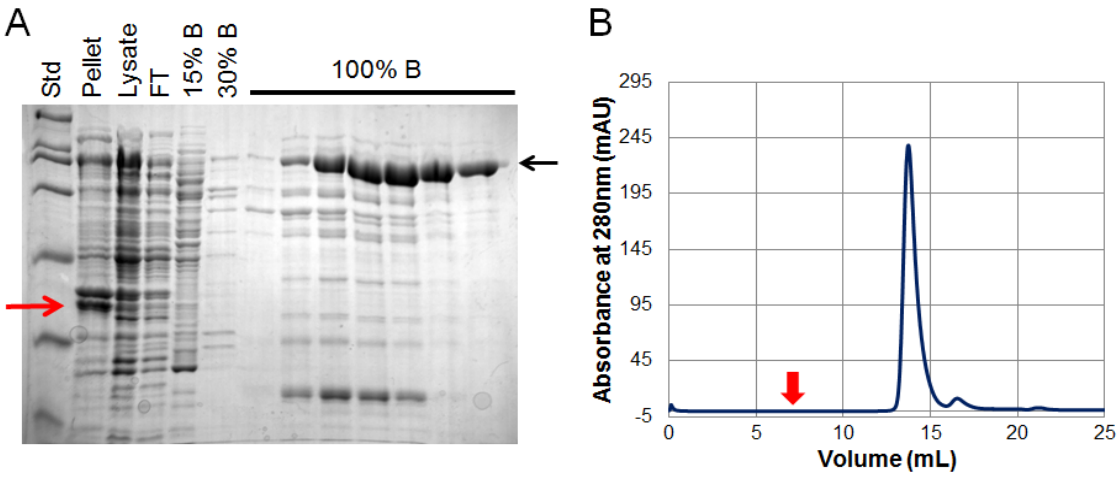

Figure 42. Expression and Purification of FL CHIKV nsP2. (A) Cleavage competent FL CHIKV P23 expressed in bacterial cells and purified over HisTrap column. SDS-PAGE analysis shows a band corresponding to cleaved nsP3

(macrodomain and ZBD) in the insoluble fraction (red arrow). Free, FL nsP2 was eluted at 250 mM imidazole (black arrow). (B) Size exclusion chromatography shows FL nsP2 to be well folded and monomeric. Void volume is denoted with a red arrow.

3. Expression and Purification of CHIKV P23 Polyprotein

To produce FL CHIKV P23 polyprotein, we mutated C478 of the catalytic dyad. CHIKV FL P23 C478S was expressed from the nsP2 helicase domain through the nsP3 ZBD as a fusion with SUMO (Figure 43A). Resulting protein was shown to be well folded and monomeric via size exclusion chromatography (Figure 43B). Solubility of this protein is increased by use of higher concentrations of salt (up to 500 mM KCl) and phosphate buffer over others (e.g. HEPES), although concentrations above 5 mg/mL are difficult to obtain. Purified CHIKV nsP2 is able to cleave FL P23 and P23$^{pro\text{-}zbd}$ proteins, meaning our polyproteins are in a native, relevant conformation (Figure 44). Crystallization trials of the FL P23 polyprotein have yielded no crystals.

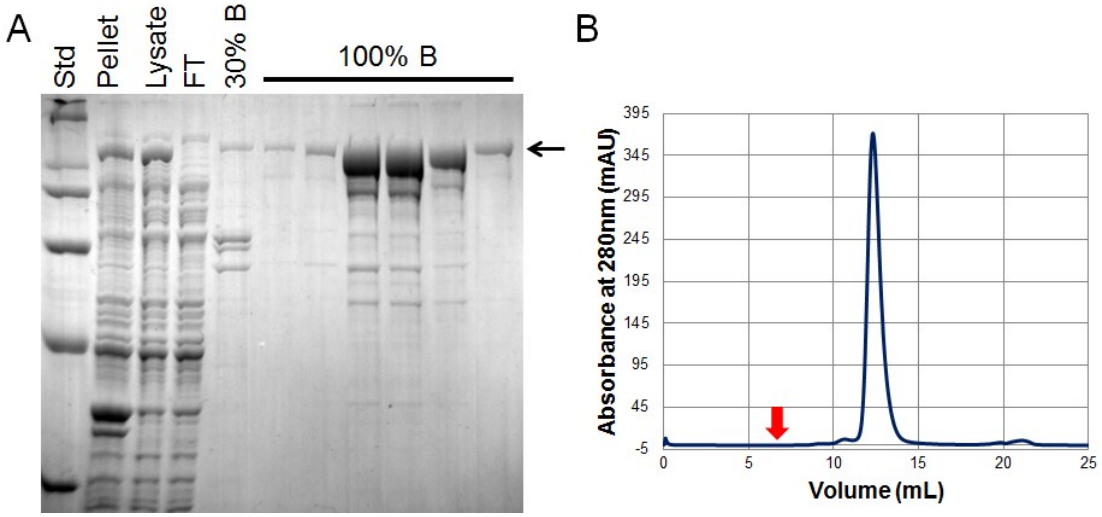

Figure 43. Expression and Purification of CHIKV FL P23. (A) SDS-PAGE analysis of CHIKV FL P23 protein purification from HisTrap (GE Healthcare) column. Protein is loaded onto the column (FT) and eluted with increasing concentrations of imidazole (25-250 mM). FL P23 is denoted with an arrow. (B)

Chromatogram of absorbance at 280nm of purified FL P23 from size exclusion chromatography. Void volume is denoted with a red arrow.

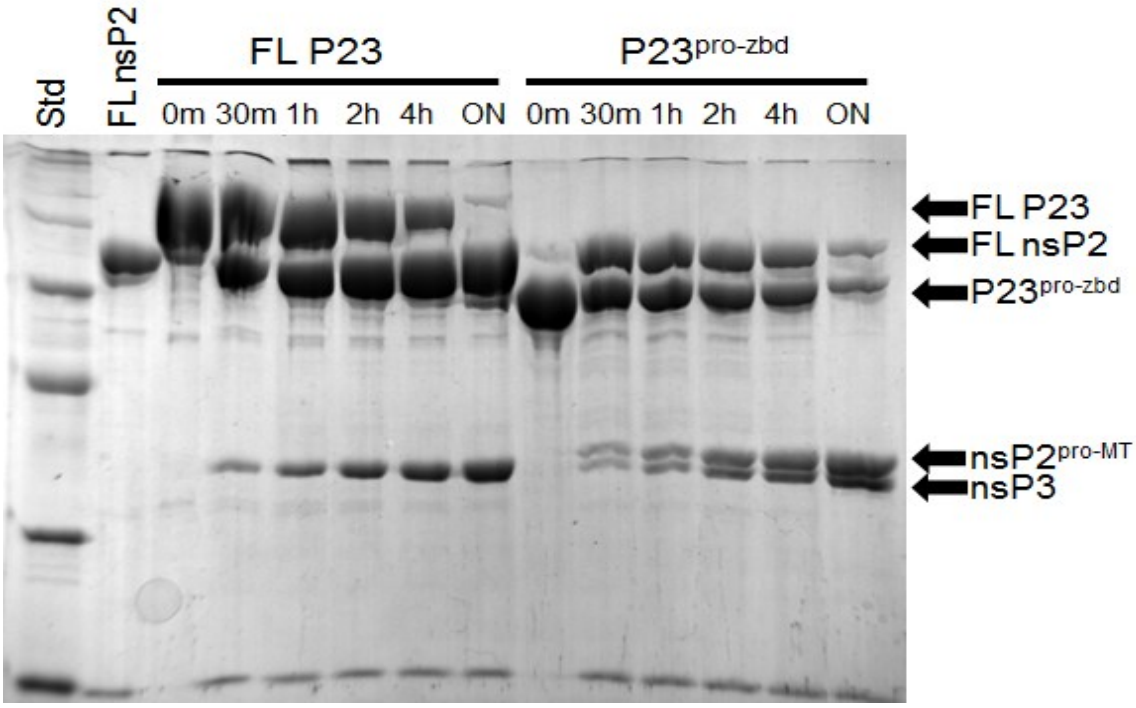

Figure 44. Cleavage of Purified CHIKV Proteins by nsP2. nsP2 is able to process the P2/3 site in both FL P23 and P23$^{pro\text{-}zbd}$ proteins. nsP2 and indicated substrates were added at a 1:1 molar ratio and incubated at room temperature for indicated time points. SDS loading buffer was added and samples were boiled for five minutes to stop the reaction.

4. ATP Hydrolysis Rates of CHIKV nsP2 and P23

There is no data in the current literature about the rate of ATP hydrolysis by P23. Typically helicase domains which are composed of two RecA-like domains lack ATP hydrolysis activity without the presence of a ligand as the ATPase active site is brought together by the two domains when a ligand is bound. Mature nsP2, however, does have intrinsic ATPase activity in the absence of ligand (Figure 45). The overall trend for both nsP2 and P23 is that ATPase activity increases in the presence of single stranded overhangs and non-base paired forks. ATPase activity also increases with increasing length of substrates. The activity of nsP2 seems to be inhibited in the presence of nsP3 as the P23 activity is consistently lower than mature nsP2.

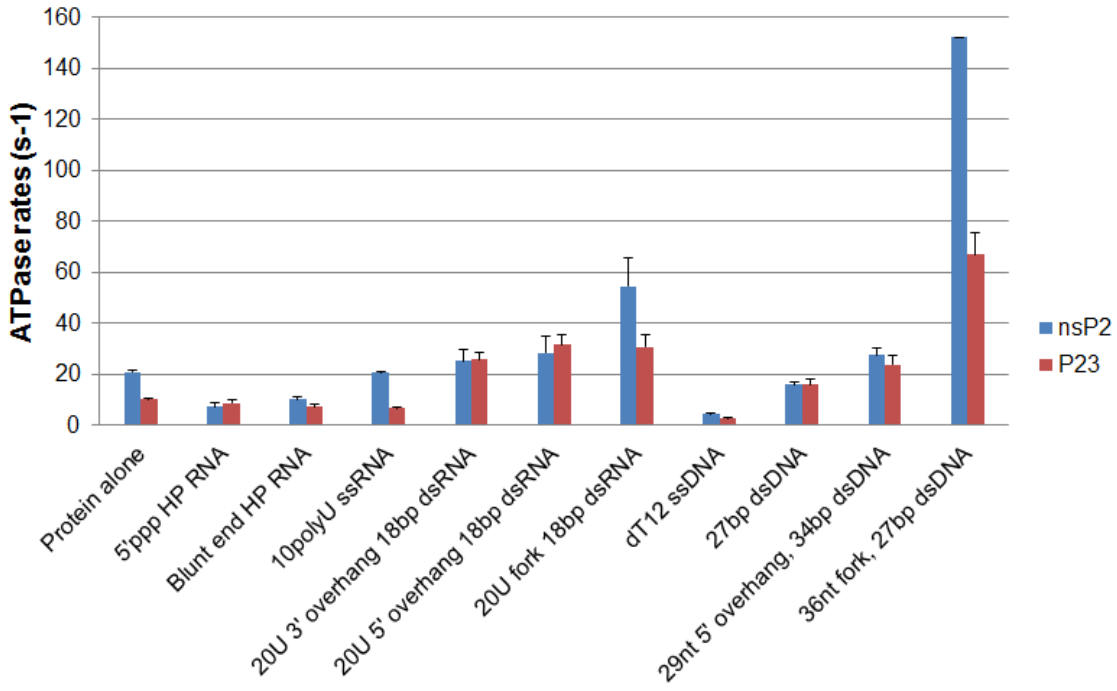

Figure 45. ATP Hydrolysis Analysis of CHIKV nsP2 and P23. ATPase activity of mature nsP2 (blue) and polyprotein P23 (red) was measured in the presence and absence of nucleic acids ligands varying in length, number of strands, and fork/overhang structures. HP indicates hairpin structures. nsP2 has intrinsic ATP hydrolysis in the absence of ligand, which is about double the activity of P23.

5. Expression and Purification of VEEV nsP2 and P23

Studies show variation between alphavirus species with regard to nonstructural protein activity levels: the SINV protease is able to process SFV polyprotein, but not vice versa, macrodomain has varying affinity for ADP-ribose between species, etc. To investigate whether different species of alphaviruses have varying ATP hydrolysis rates, or respond differently to nucleic acid substrates, we expressed and purified nsP2 and FL P23 from VEEV (Figure 46A-B). VEEV and CHIKV nsP2 and P23 have 56% sequence identity and 15% sequence similarity. nsP2 was produced similarly to CHIKV as a cleavage competent P23 construct. During protein production, nsP2 cleaves the P2/3 bond and nsP3 becomes insoluble leaving FL nsP2 for purification. P23 was made with active site mutation C477A. VEEV proteins expressed at much

higher levels versus CHIKV proteins and were much more soluble, even at lower ionic concentrations. Both proteins were found to be monomeric via size exclusion chromatography (Figure 47). Crystallization trials of VEEV FL P23 and nsP2 have yielded no results.

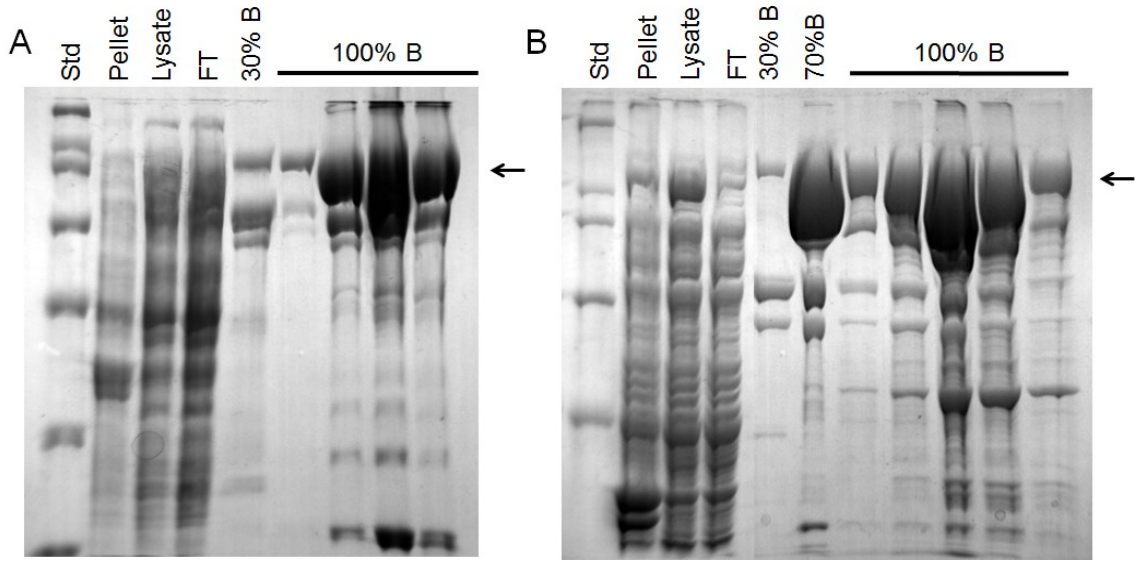

Figure 46. Expression and Purification of VEEV nsP2 and P23. (A) SDS-PAGE analysis of VEEV nsP2 expressed as a cleavage competent P23 fusion with SUMO. Arrow indicates eluted nsP2. (B) SDS-PAGE analysis of VEEV P23 expressed with a mutated protease active site and fused with SUMO. Arrow indicates eluted P23.

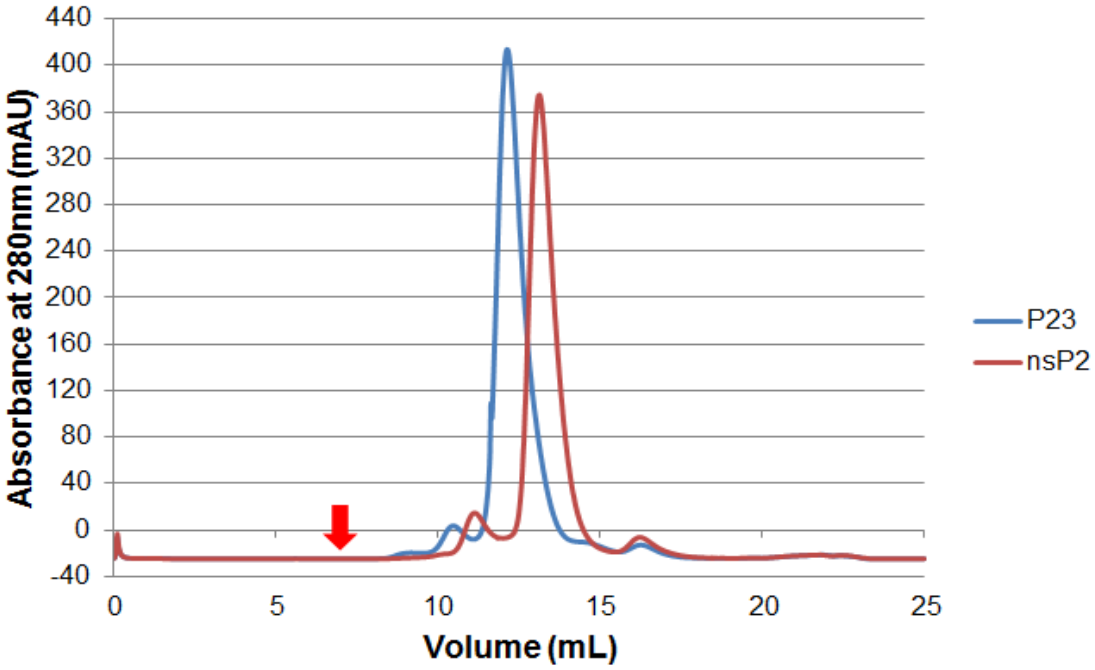

Figure 47. Size Exclusion Chromatography of VEEV P23 and nsP2.
Chromatogram of absorbance at 280nm of purified FL P23 (blue) and nsP2 (red) from size exclusion chromatography. Void volume is denoted with a red arrow.

6. ATP Hydrolysis Rates of VEEV nsP2 and P23

As there is no data in the current literature about the rate of ATP hydrolysis by P23 in the literature to compare with, we expressed nsP2 and P23 from VEEV for analysis. Like CHIKV, the VEEV mature nsP2 does have intrinsic ATPase activity (Figure 48). The overall trend for both nsP2 and P23 is that ATPase activity increases with increasing length of substrates with non-base paired forks and overhangs; however, unlike CHIKV, P23 has consistently higher ATP hydrolysis activity than nsP2 alone. This is in sharp contrast to the CHIKV proteins, which seemed to indicate that the presence of nsP3 may be regulating the activity of nsP2.

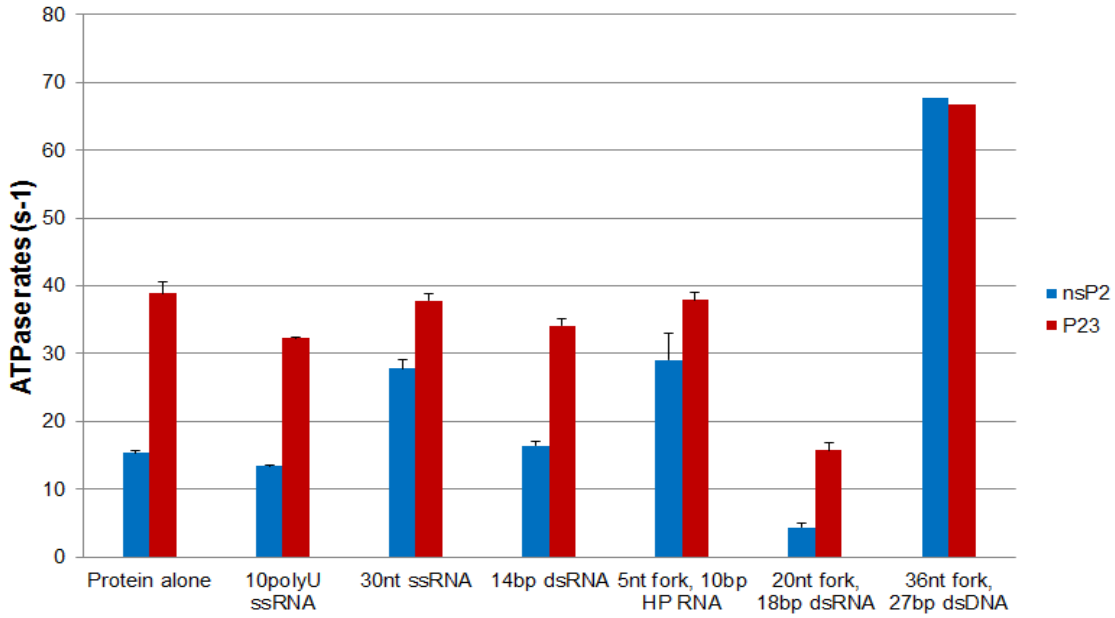

Figure 48. ATP Hydrolysis Analysis of VEEV nsP2 and P23. ATPase activity of mature nsP2 (blue) and polyprotein P23 (red) was measured in the presence and absence of nucleic acids ligands varying in length, number of strands, and fork/overhang structures. HP indicates hairpin structures.

7. Alphavirus Helicase Flexibility

Because all trials with nsP2 and P23 from both CHIKV and VEEV failed to produce any crystals for X-ray crystallography studies, we sought to gain information about our proteins via small angle X-ray scattering (SAXS). SAXS is a method used to study overall conformation of a protein in a low resolution manner. SAXS provides scattering data that can be used to obtain the radius of gyration (R_g) and overall shape information about a protein. R_g is the root-mean-square of the distances of all regions to the center of mass of the particle while P(r) represents the probability distribution of distances between scattering atoms within a particle. Data is typically represented in two ways. First, a pairwise distribution plot [R_g versus the P(r)] which for globular proteins will have a Gaussian distribution (a bell shaped curve) (Figure 49A). Second is in a Kratky plot [q versus $q^2I(q)$] in which a compact, globular protein gives a bell-shaped curve

and becomes flat at higher values of q (Figure 49B). Using these types of analyses, we can analyze whether our proteins are partially unfolded and/or flexible (153).

The CHIKV P23$^{pro\text{-}zbd}$ construct used for X-ray crystallography was used as a reference in this data collection since we know the protein is folded and globular. By pairwise distribution comparison, this protein appears globular (Figure 50A). CHIKV helicase domain (amino acid residues 1-477 of nsP2) indicates some amount of unfoldedness as the pairwise distribution is not a symmetric peak (Figure 50B). The nsP2 and P23 pairwise distributions are very broad and non-symmetric, indicating likely a multidomain protein (Figure 50B). Based on what we already know about the P23$^{pro\text{-}zbd}$ structure, it seems the helicase may not be closely associated with the rest of the complex, and may form a somewhat separated domain structure. Kratky plots show folded P23$^{pro\text{-}zbd}$ while helicase, nsP2, and FL P23 are partially unfolded or flexible (Figure 51A-D). VEEV nsP2 and P23 data is similar showing asymmetric functions by pairwise distribution and partially unfolded or flexible protein by Kratky plot (Figure 52A-C).

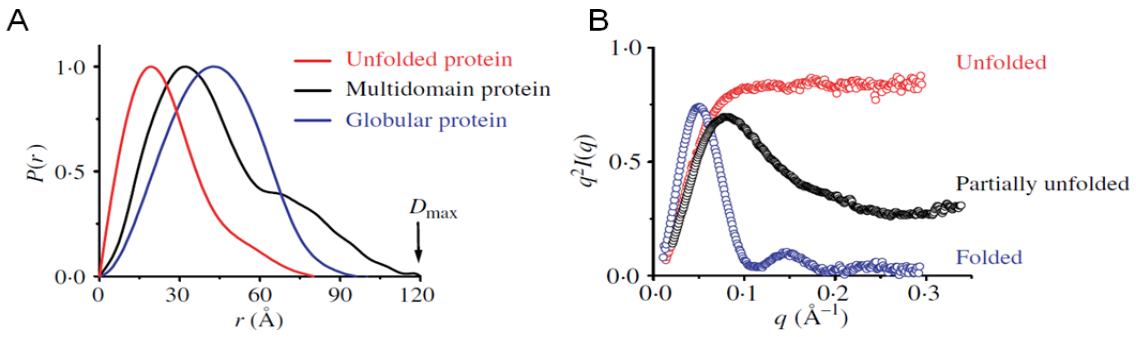

Figure 49. Sample SAXS Data Analysis. Radius of gyration versus probability of distribution and (B) Kratky plot for proteins of different conformation. Figure adapted from (153).

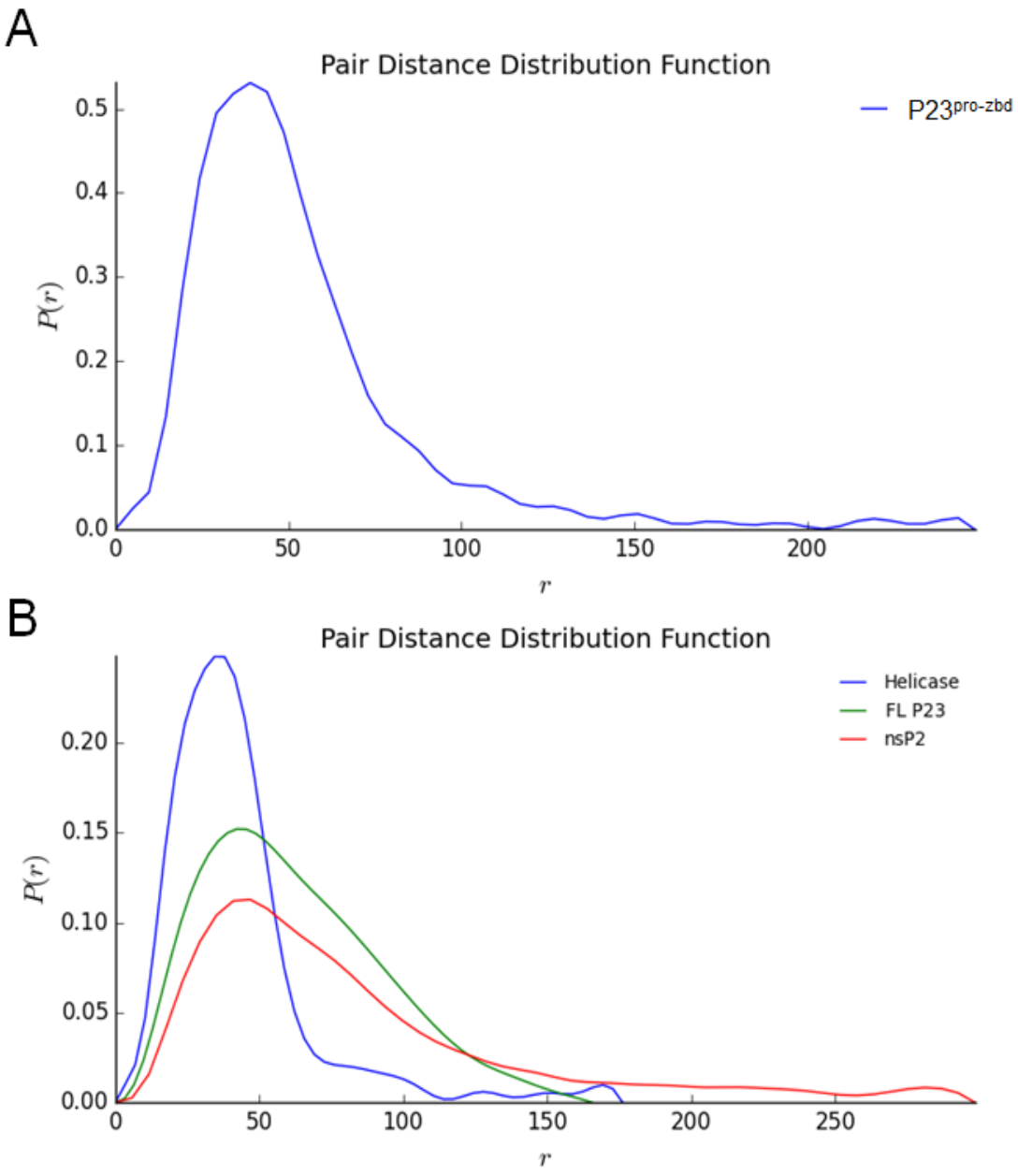

Figure 50. Pairwise Distance Distribution Function of CHIKV Proteins. R_g (r) plotted against the P(r) for (A) CHIKV $P23^{pro\text{-}zbd}$, and (B) CHIKV helicase, FL P23, and nsP2.

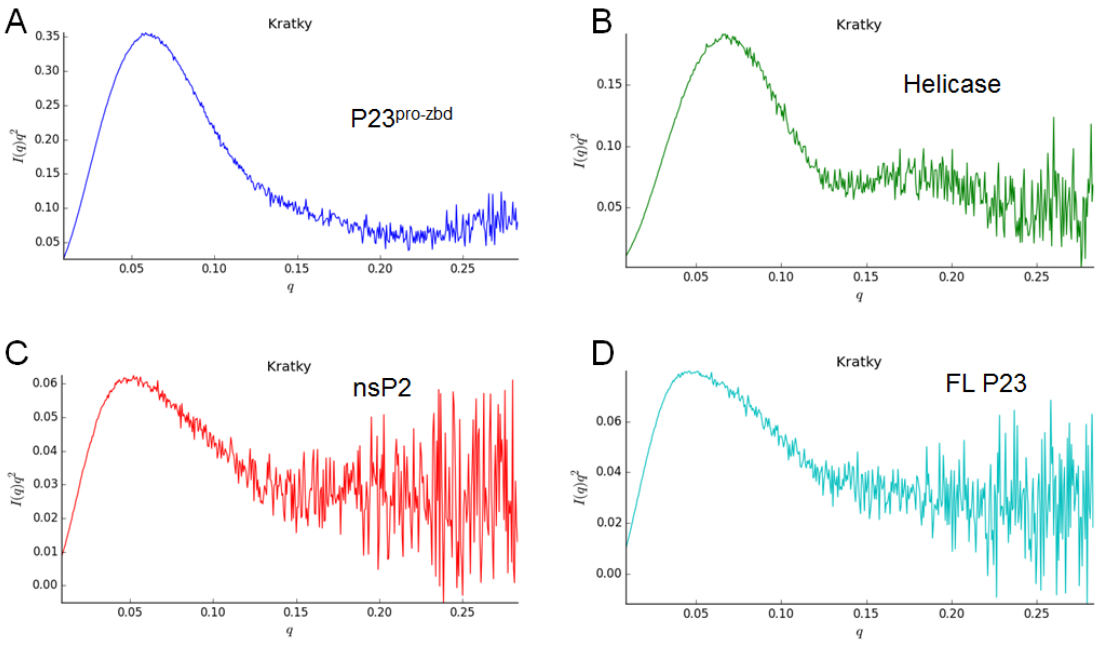

Figure 51. Kratky Plots for CHIKV Proteins. Kratky plot of scattering for CHIKV proteins (A) $P23^{pro\text{-}zbd}$, (B) helicase alone, (C) nsP2, and (D) FL P23.

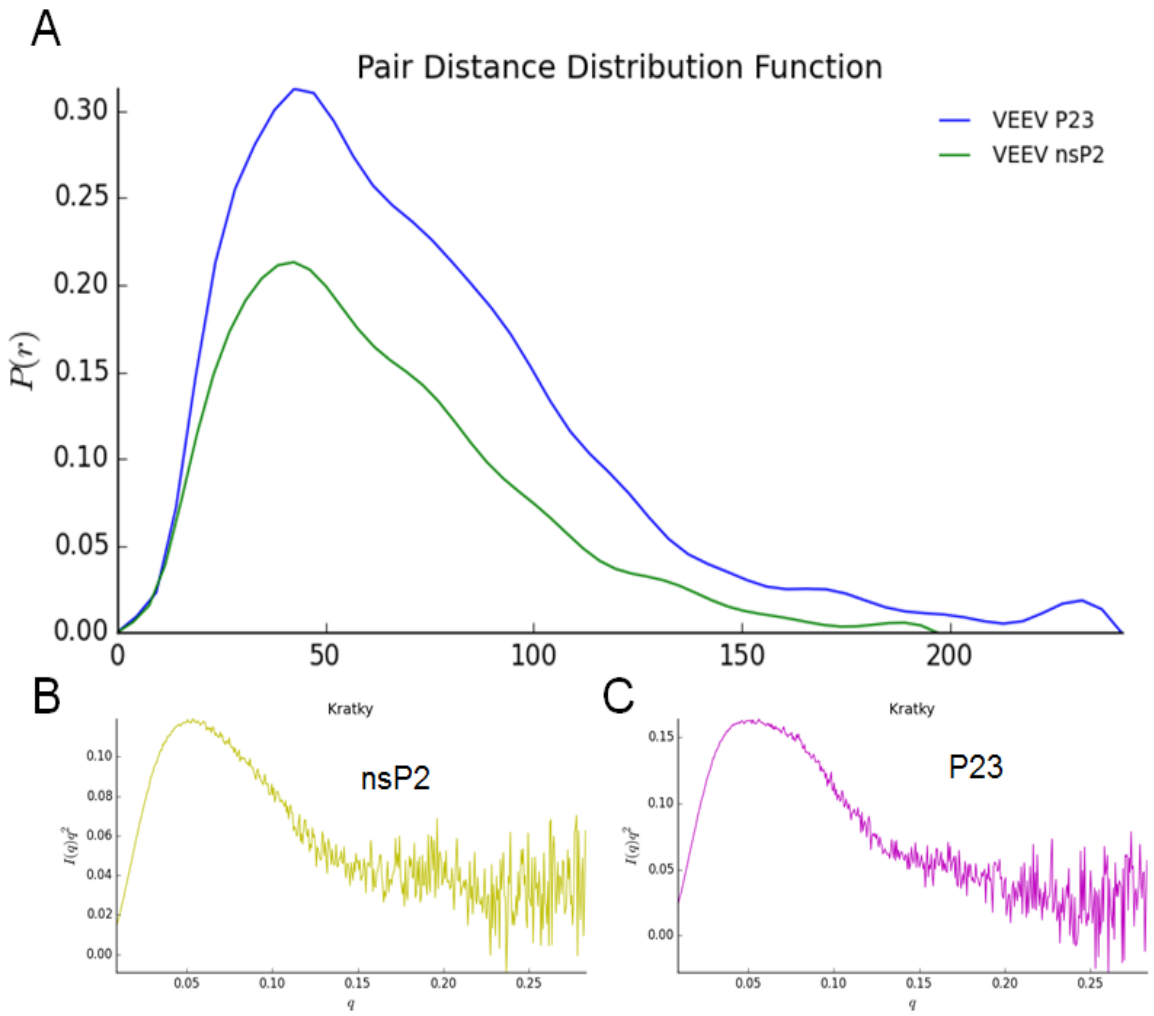

Figure 52. Pairwise Distance Distribution Function and Kratky Plots for VEEV Proteins. (A) R_g (r) plotted against the P(r) for VEEV nsP2 and FL P23. Kratky plot of scattering for VEEV proteins (B) nsP2 and (C) FL P23.

8. Discussion

Here we discussed FL P23 expression, a novel method for producing mature nsP2, and the changes in nsP2 ATP hydrolysis rates in the P23 polyprotein versus cleaved protein. Full length alphavirus nsP2 has been historically difficult to produce. Our method of production with a portion of nsP3 increases expression levels and solubility over expression of nsP2 alone. Purified nsP2 was verified to be in an active conformation via testing in cleavage assays with P23 polyproteins.

Expression of genes as a polyprotein has been shown several times in literature to serve as a mechanism for protein activity regulation. Our studies show ATP hydrolysis activity of nsP2 changes depending on polyprotein state. In CHIKV, nsP3 downregulates the ATP hydrolysis activity of nsP2. It is the opposite for VEEV. Further study is necessary to understand why this is the case, but it may be due to differences the progression of replication in different hosts. Additionally, nsP2 activity increased in the presence of long, forked substrates, especially DNAs. This fact seems contradictory with other studies showing lack of dsDNA unwinding ability of nsP2 (61).

Expanding the $P23^{pro\text{-}zbd}$ structure was our initial goal for this section. Further structural information is necessary to understand how the extreme N-terminal portion of nsP2 participates in efficient P2/3 cleavage. Unfortunately, none of the proteins containing the nsP2 N-terminal helicase domain yielded any crystals. A ligand may be necessary for the stability of the protein, which would facilitate crystal growth; however, all trials with ATP analogues and nucleic acid substrates also failed. Under many conditions, nsP2 and P23 precipitate in the presence of nucleic acid ligand. It is unknown whether this is due to a multimerization event, or the protein simply undergoes a conformational change causing aggregation. However, in the presence of phosphate buffer, these proteins fail to bind nucleic acids, presumably because the phosphate ions out-compete nucleic acid backbone-protein interactions. Data implies that the helicase domain is functional, as it has ATP hydrolysis activity, but is too flexible to allow crystal formation. Further trials with alternate ATP analogues and buffer conditions are ongoing.

CONCLUSION

The polyprotein translation strategy is common across many viruses families, including several of importance to human health (101). Encoding and translating large viral polyprotein precursors is a strategy with many benefits. Despite their presence in many viral families, there are only a handful of structures available of precleavage polyprotein intermediates, and only one for a viral nonstructural polyprotein previous to our study (40). Our studies produced two alphavirus polyprotein structures, $P23^{pro\text{-}zbd}$ from SINV and CHIKV (68). The "point of no return" in the RNA template switching process during RNA replication is the cleavage step between P2/3. The structures imply that this cleavage event is likely tightly regulated as the cleavage site is not accessible to the protease. Additionally, this site is processed in *trans* by a protease which must contain the N-terminal domain to function. The high resolution CHIKV $P23^{pro\text{-}zbd}$ structure allowed for *in silico* ligand docking experiments. Accessible pockets are available which may interrupt the protease active site, ADP-ribose binding site, and the P2/3 cleavage site. Antiviral drugs capable of disrupting P23 cleavage may be a promising avenue as engineered alphavirus constructs unable to process the P23 polyprotein cannot produce cytopathic effects (69).

We proposed an RNA binding surface that spans across both nsP2 and nsP3. The cleavage of P23 may alter this surface and affect RNA template selection. Ideally, the structures of FL P23 and fully mature nsP2 and nsP3 would be necessary to fully understand how these proteins change post-cleavage. Their extensive shared surface area begs the question of how these two separate post-cleavage. In our expression studies of P23, nsP3 remains soluble; however, post-cleavage nsP3 is highly insoluble. This clearly makes X-ray crystallography study of mature nsP3 quite difficult. Expression in mammalian or insect systems or the addition of highly soluble tags, such as MBP, may help produce quantities of nsP3 necessary for further experimentation.

Our studies of FL P23 show that the N-terminal domain is highly flexible, making it unsuitable for crystallization. nsP2 helicase has altered ATP hydrolysis activity based on its cleavage status. Further studies need to be conducted to verify whether ATPase activity correlates with increased overall unwinding activity, and how this relates to the activities of nsP2 during stages of infection. Non-hydrolyzable ATP analogues and nucleic acid substrates may allow for crystal growth by stabilizing flexible regions.

nsP3 is, by far, the least understood alphavirus nonstructural protein. We not only showed nsP3 to be important in efficient P2/3 cleavage, structures also revealed a novel zinc binding domain not previously described in nsP3. The zinc ion is coordinated by four highly conserved cysteine residues which are each absolutely necessary for viral replication. To further our understanding of nsP3, we examined the contributions of ADP-ribose binding of nsP3 macrodomain in viral infection and discovered that there is a significant effect during replication in mosquito cells. ADP-ribosylation of proteins is a post-translational modification with importance in a wide range of cellular activities from DNA repair to apoptosis (154, 155). Perhaps nsP3 plays a role in host cell interaction and progression of infection through ADP-ribosylated proteins.

Overall, the data presented here successfully furthers our goal of describing alphavirus nsPs with regard to their structure and activities pre- and post-cleavage via biophysical, biochemical, and virological methods in order to better understand viral pathogenesis and present opportunities for novel anti-viral therapy design.

REFERENCES

1. Koonin EV, Senkevich TG, & Dolja VV (2006) The ancient Virus World and evolution of cells. *Biology direct* 1:29.

2. Meyer K, Haring C, & Howitt B (1931) The Etiology of Epizootic Encephalomyelitis of Horses in the San Juaquin Valley, 1930. *Science (New York, N.Y* 74(1913):227-228.

3. ICTV (2014) Virus Taxonomy: 2014 Release. (Montreal, Canada).

4. NIAID (2016) NIAID Emerging Infectious Diseases/Pathogens. (NIH).

5. Pulmanausahakul R, Roytrakul S, Auewarakul P, & Smith DR (2011) Chikungunya in Southeast Asia: understanding the emergence and finding solutions. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 15(10):e671-676.

6. Tatem AJ, Hay SI, & Rogers DJ (2006) Global traffic and disease vector dispersal. *Proc Natl Acad Sci U S A* 103(16):6242-6247.

7. San Martin JL*, et al.* (2010) The epidemiology of dengue in the americas over the last three decades: a worrisome reality. *The American journal of tropical medicine and hygiene* 82(1):128-135.

8. PAHO/WHO (2014) Number of reported cases of Chikungunya Fever in the Americas - Cumulative Cases (May 15, 2015). (Pan American Health Organization).

9. Dash PK*, et al.* (2008) Development and evaluation of a 1-step duplex reverse transcription polymerase chain reaction for differential diagnosis of chikungunya and dengue infection. *Diagnostic microbiology and infectious disease* 62(1):52-57.

10. Nkoghe D*, et al.* (2012) Clinical forms of chikungunya in Gabon, 2010. *PLoS Negl Trop Dis* 6(2):e1517.

11. Sharda S (2013) UP disabled to hear killer mosquito's buzz. The Times of India.

12. Isalkar U (2014) Dengue spurt in state caused by poor vector surveillance. The Times of India.

13. Lundstrom K (2015) Alphaviruses in gene therapy. *Viruses* 7(5):2321-2333.

14. Liljestrom P & Garoff H (1991) A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Biotechnology (N Y)* 9(12):1356-1361.

15. Ehrengruber MU*, et al.* (1999) Recombinant Semliki Forest virus and Sindbis virus efficiently infect neurons in hippocampal slice cultures. *Proc Natl Acad Sci U S A* 96(12):7041-7046.

16. Strauss JH & Strauss EG (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiological reviews* 58(3):491-562.

17. Garmashova N*, et al.* (2007) The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. *J Virol* 81(5):2472-2484.

18. Forrester NL*, et al.* (2012) Genome-Scale Phylogeny of the Alphavirus Genus Suggests a Marine Origin. *Journal of Virology* 86(5):2729-2738.

19. Lescar J*, et al.* (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105(1):137-148.

20. Zhang W*, et al.* (2002) Placement of the structural proteins in Sindbis virus. *J Virol* 76(22):11645-11658.

21. Rose PP*, et al.* (2011) Natural resistance-associated macrophage protein is a cellular receptor for sindbis virus in both insect and mammalian hosts. *Cell host & microbe* 10(2):97-104.

22. Omar A & Koblet H (1988) Semliki Forest virus particles containing only the E1 envelope glycoprotein are infectious and can induce cell-cell fusion. *Virology* 166(1):17-23.

23. Wahlberg JM & Garoff H (1992) Membrane fusion process of Semliki Forest virus. I: Low pH-induced rearrangement in spike protein quaternary structure precedes virus penetration into cells. *The Journal of cell biology* 116(2):339-348.

24. Li L, Jose J, Xiang Y, Kuhn RJ, & Rossmann MG (2010) Structural changes of envelope proteins during alphavirus fusion. *Nature* 468(7324):705-708.

25. Voss JE*, et al.* (2010) Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468(7324):709-712.

26. Vasiljeva L, Merits A, Auvinen P, & Kääriäinen L (2000) Identification of a novel function of the alphavirus capping apparatus. RNA 5'-triphosphatase activity of Nsp2. *J Biol Chem* 275(23):17281-17287.

27. Strauss EG, Levinson R, Rice CM, Dalrymple J, & Strauss JH (1988) Nonstructural proteins nsP3 and nsP4 of Ross River and O'Nyong-nyong viruses: sequence and comparison with those of other alphaviruses. *Virology* 164(1):265-274.

28. Ding MX & Schlesinger MJ (1989) Evidence that Sindbis virus NSP2 is an autoprotease which processes the virus nonstructural polyprotein. *Virology* 171(1):280-284.

29. Hardy WR & Strauss JH (1989) Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *J Virol* 63(11):4653-4664.

30. Strauss EG, De Groot RJ, Levinson R, & Strauss JH (1992) Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology* 191(2):932-940.

31. de Groot RJ, Hardy WR, Shirako Y, & Strauss JH (1990) Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. *The EMBO journal* 9(8):2631-2638.

32. Lemm JA & Rice CM (1993) Assembly of functional Sindbis virus RNA replication complexes: requirement for coexpression of P123 and P34. *J Virol* 67(4):1905-1915.

33. Lemm JA & Rice CM (1993) Roles of nonstructural polyproteins and cleavage products in regulating Sindbis virus RNA replication and transcription. *J Virol* 67(4):1916-1926.

34. Lemm JA, Rümenapf T, Strauss EG, Strauss JH, & Rice CM (1994) Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *The EMBO journal* 13(12):2925-2934.

35. Shirako Y & Strauss JH (1994) Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas cleaved products from P123 are required for efficient plus-strand RNA synthesis. *J Virol* 68(3):1874-1885.

36. Shirako Y & Strauss JH (1990) Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. *Virology* 177(1):54-64.

37. Wang YF, Sawicki SG, & Sawicki DL (1994) Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *J Virol* 68(10):6466-6475.

38. Vasiljeva L, Valmu L, Kääriäinen L, & Merits A (2001) Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. *J Biol Chem* 276(33):30786-30793.

39. Vasiljeva L*, et al.* (2003) Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *J Biol Chem* 278(43):41636-41645.

40. Marcotte L*, et al.* (2007) Crystal Structure of Poliovirus 3CD Protein: Virally Encoded Protease and Precursor to the RNA-Dependent RNA Polymerase. *The Journal of Virology* 81(7):3583.

41. Friedman RM, Levin JG, Grimley PM, & Berezesky IK (1972) Membrane-associated replication complex in arbovirus infection. *J Virol* 10(3):504-515.

42. Froshauer S, Kartenbeck J, & Helenius A (1988) Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *The Journal of cell biology* 107(6 Pt 1):2075-2086.

43. Ahola T*, et al.* (2000) Effects of palmitoylation of replicase protein nsP1 on alphavirus infection. *J Virol* 74(15):6725-6733.

44. Spuul P*, et al.* (2007) Role of the amphipathic peptide of Semliki forest virus replicase protein nsP1 in membrane association and virus replication. *J Virol* 81(2):872-883.

45. Hahn YS, Strauss EG, & Strauss JH (1989) Mapping of RNA- temperature-sensitive mutants of Sindbis virus: assignment of complementation groups A, B, and G to nonstructural proteins. *J Virol* 63(7):3142-3150.

46. Wang YF, Sawicki SG, & Sawicki DL (1991) Sindbis virus nsP1 functions in negative-strand RNA synthesis. *J Virol* 65(2):985-988.

47. Ahola T & Kaariainen L (1995) Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc Natl Acad Sci U S A* 92(2):507-511.

48. Laakkonen P, Hyvonen M, Peranen J, & Kaariainen L (1994) Expression of Semliki Forest virus nsP1-specific methyltransferase in insect cells and in Escherichia coli. *J Virol* 68(11):7418-7425.

49. Ahola T, Laakkonen P, Vihinen H, & Kaariainen L (1997) Critical residues of Semliki Forest virus RNA capping enzyme involved in methyltransferase and guanylyltransferase-like activities. *J Virol* 71(1):392-397.

50. Li C*, et al.* (2015) mRNA Capping by Venezuelan Equine Encephalitis Virus nsP1: Functional Characterization and Implications for Antiviral Research. *J Virol* 89(16):8292-8303.

51. Lulla A, Lulla V, & Merits A (2012) Macromolecular assembly-driven processing of the 2/3 cleavage site in the alphavirus replicase polyprotein. *J Virol* 86(1):553-565.

52. Merits A, Vasiljeva L, Ahola T, Kaariainen L, & Auvinen P (2001) Proteolytic processing of Semliki Forest virus-specific non-structural polyprotein by nsP2 protease. *The Journal of general virology* 82(Pt 4):765-773.

53. Russo AT, White MA, & Watowich SJ (2006) The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 14(9):1449-1458.

54. Russo AT, Malmstrom RD, White MA, & Watowich SJ (2010) Structural basis for substrate specificity of alphavirus nsP2 proteases. *J Mol Graph Model* 29(1):46-53.

55. ten Dam E, Flint M, & Ryan MD (1999) Virus-encoded proteinases of the Togaviridae. *The Journal of general virology* 80 (Pt 8):1879-1888.

56. Saisawang C*, et al.* (2015) Chikungunya nsP2 protease is not a papain-like cysteine protease and the catalytic dyad cysteine is interchangeable with a proximal serine. *Scientific reports* 5:17125.

57. Golubtsov A, Kaariainen L, & Caldentey J (2006) Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis. *FEBS letters* 580(5):1502-1508.

58. Saisawang C*, et al.* (2015) Full length and protease domain activity of chikungunya virus nsP2 differ from other alphavirus nsP2 proteases in recognition of small peptide substrates. *Bioscience reports* 35(3).

59. Gomez de Cedron M, Ehsani N, Mikkola ML, Garcia JA, & Kaariainen L (1999) RNA helicase activity of Semliki Forest virus replicase protein NSP2. *FEBS letters* 448(1):19-22.

60. Rikkonen M, Peranen J, & Kaariainen L (1994) ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. *J Virol* 68(9):5804-5810.

61. Das PK, Merits A, & Lulla A (2014) Functional cross-talk between distant domains of chikungunya virus non-structural protein 2 is decisive for its RNA-modulating activity. *J Biol Chem* 289(9):5635-5653.

62. Karpe YA, Aher PP, & Lole KS (2011) NTPase and 5'-RNA triphosphatase activities of Chikungunya virus nsP2 protein. *PloS one* 6(7):e22336.

63. Peranen J, Rikkonen M, Liljestrom P, & Kaariainen L (1990) Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. *J Virol* 64(5):1888-1896.

64. Rikkonen M, Peranen J, & Kaariainen L (1992) Nuclear and nucleolar targeting signals of Semliki Forest virus nonstructural protein nsP2. *Virology* 189(2):462-473.

65. Rikkonen M, Peranen J, & Kaariainen L (1994) Nuclear targeting of Semliki Forest virus nsP2. *Arch Virol Suppl* 9:369-377.

66. Frolova EI*, et al.* (2002) Roles of nonstructural protein nsP2 and Alpha/Beta interferons in determining the outcome of Sindbis virus infection. *J Virol* 76(22):11254-11264.

67. Akhrymuk I, Kulemzin SV, & Frolova EI (2012) Evasion of the innate immune response: the Old World alphavirus nsP2 protein induces rapid degradation of Rpb1, a catalytic subunit of RNA polymerase II. *J Virol* 86(13):7180-7191.

68. Shin G*, et al.* (2012) Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proc Natl Acad Sci U S A* 109(41):16534-16539.

69. Frolov I*, et al.* (1999) Selection of RNA replicons capable of persistent noncytopathic replication in mammalian cells. *J Virol* 73(5):3854-3865.

70. Dryga SA, Dryga OA, & Schlesinger S (1997) Identification of mutations in a Sindbis virus variant able to establish persistent infection in BHK cells: the importance of a mutation in the nsP2 gene. *Virology* 228(1):74-83.

71. Mayuri, Geders TW, Smith JL, & Kuhn RJ (2008) Role for conserved residues of sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection. *J Virol* 82(15):7284-7297.

72. Perri S*, et al.* (2000) Replicon vectors derived from Sindbis virus and Semliki forest virus that establish persistent replication in host cells. *J Virol* 74(20):9802-9807.

73. Petrakova O*, et al.* (2005) Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in Mammalian cells. *J Virol* 79(12):7597-7608.

74. LaStarza MW, Lemm JA, & Rice CM (1994) Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. *J Virol* 68(9):5781-5791.

75. Rupp JC, Jundt N, & Hardy RW (2011) Requirement for the amino-terminal domain of sindbis virus nsP4 during virus infection. *J Virol* 85(7):3449-3460.

76. Saxton-Shaw KD*, et al.* (2013) O'nyong nyong virus molecular determinants of unique vector specificity reside in non-structural protein 3. *PLoS Negl Trop Dis* 7(1):e1931.

77. Suthar MS, Shabman R, Madric K, Lambeth C, & Heise MT (2005) Identification of adult mouse neurovirulence determinants of the Sindbis virus strain AR86. *J Virol* 79(7):4219-4228.

78. Tuittila M & Hinkkanen AE (2003) Amino acid mutations in the replicase protein nsP3 of Semliki Forest virus cumulatively affect neurovirulence. *The Journal of general virology* 84(Pt 6):1525-1533.

79. Tuittila MT, Santagati MG, Roytta M, Maatta JA, & Hinkkanen AE (2000) Replicase complex genes of Semliki Forest virus confer lethal neurovirulence. *J Virol* 74(10):4579-4589.

80. Allen MD, Buckle AM, Cordell SC, Lowe J, & Bycroft M (2003) The crystal structure of AF1521 a protein from Archaeoglobus fulgidus with homology to the non-histone domain of macroH2A. *Journal of molecular biology* 330(3):503-511.

81. Park E & Griffin DE (2009) The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. *Virology* 388(2):305-314.

82. Pehrson JR & Fuji RN (1998) Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Res* 26(12):2837-2842.

83. Egloff MP*, et al.* (2006) Structural and functional basis for ADP-ribose and poly(ADP-ribose) binding by viral macro domains. *J Virol* 80(17):8493-8502.

84. Aguiar RC, Takeyama K, He C, Kreinbrink K, & Shipp MA (2005) B-aggressive lymphoma family proteins have unique domains that modulate transcription and exhibit poly(ADP-ribose) polymerase activity. *J Biol Chem* 280(40):33756-33765.

85. Ahel D*, et al.* (2009) Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science (New York, N.Y* 325(5945):1240-1243.

86. Gottschalk AJ*, et al.* (2009) Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc Natl Acad Sci U S A* 106(33):13770-13774.

87. Han WD*, et al.* (2007) Estrogenically regulated LRP16 interacts with estrogen receptor alpha and enhances the receptor's transcriptional activity. *Endocrine-related cancer* 14(3):741-753.

88. Malet H*, et al.* (2009) The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J Virol* 83(13):6534-6545.

89. Neuvonen M & Ahola T (2009) Differential activities of cellular and viral macro domain proteins in binding of ADP-ribose metabolites. *Journal of molecular biology* 385(1):212-225.

90. Vihinen H, Ahola T, Tuittila M, Merits A, & Kääriäinen L (2001) Elimination of phosphorylation sites of Semliki Forest virus replicase protein nsP3. *J Biol Chem* 276(8):5745-5752.

91. Li GP, La Starza MW, Hardy WR, Strauss JH, & Rice CM (1990) Phosphorylation of Sindbis virus nsP3 in vivo and in vitro. *Virology* 179(1):416-427.

92. Peranen J, Takkinen K, Kalkkinen N, & Kaariainen L (1988) Semliki Forest virus-specific non-structural protein nsP3 is a phosphoprotein. *The Journal of general virology* 69 (Pt 9):2165-2178.

93. Vihinen H & Saarinen J (2000) Phosphorylation site analysis of Semliki forest virus nonstructural protein 3. *J Biol Chem* 275(36):27775-27783.

94. Davis NL, Willis LV, Smith JF, & Johnston RE (1989) In vitro synthesis of infectious venezuelan equine encephalitis virus RNA from a cDNA clone: analysis of a viable deletion mutant. *Virology* 171(1):189-204.

95. Galbraith SE, Sheahan BJ, & Atkins GJ (2006) Deletions in the hypervariable domain of the nsP3 gene attenuate Semliki Forest virus virulence. *The Journal of general virology* 87(Pt 4):937-947.

96. LaStarza MW, Grakoui A, & Rice CM (1994) Deletion and duplication mutations in the C-terminal nonconserved region of Sindbis virus nsP3: effects on phosphorylation and on virus replication in vertebrate and invertebrate cells. *Virology* 202(1):224-232.

97. Foy NJ, Akhrymuk M, Shustov AV, Frolova EI, & Frolov I (2013) Hypervariable domain of nonstructural protein nsP3 of Venezuelan equine encephalitis virus determines cell-specific mode of virus replication. *J Virol* 87(13):7569-7584.

98. Fata CL, Sawicki SG, & Sawicki DL (2002) Modification of Asn374 of nsP1 suppresses a Sindbis virus nsP4 minus-strand polymerase mutant. *J Virol* 76(17):8641-8649.

99. O'Reilly EK & Kao CC (1998) Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* 252(2):287-303.

100. Rubach JK*, et al.* (2009) Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* 384(1):201-208.

101. Korant BD (1990) Strategies to inhibit viral polyprotein cleavages. *Annals of the New York Academy of Sciences* 616:252-257.

102. Yost SA & Marcotrigiano J (2013) Viral precursor polyproteins: keys of regulation from replication to maturation. *Current opinion in virology* 3(2):137-142.

103. Leslie AGW & Powell HR (2007) Processing Diffraction Data with MOSFLM. *Evolving Methods for Macromolecular Crystallography: The Structural Path to the Understanding of the Mechanisms of Action of CBRN Agents*, eds Read RJ & Sussman JL (Springer Verlag), Vol 245, pp 41-51.

104. 4 CCPN (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760-763.

105. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112-122.

106. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.

107. Adams PD*, et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.

108. Chen VB*, et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12-21.

109. DeLano WL (2002) The PyMOL molecular graphics system.

110. Baker NA, Sept D, Joseph S, Holst MJ, & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 98(18):10037-10041.

111. Cristea IM*, et al.* (2006) Tracking and elucidating alphavirus-host protein interactions. *J Biol Chem* 281(40):30269-30278.

112. Sanner MF, Olson AJ, & Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38(3):305-320.

113. Morris GM*, et al.* (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry* 30(16):2785-2791.

114. Hahn CS, Hahn YS, Braciale TJ, & Rice CM (1992) Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. *Proc Natl Acad Sci U S A* 89(7):2679-2683.

115. Pierro DJ, Myles KM, Foy BD, Beaty BJ, & Olson KE (2003) Development of an orally infectious Sindbis virus transducing system that efficiently disseminates and expresses green fluorescent protein in Aedes aegypti. *Insect molecular biology* 12(2):107-116.

116. Seabaugh RC, Olson KE, Higgs S, Carlson JO, & Beaty BJ (1998) Development of a chimeric sindbis virus with enhanced per Os infection of Aedes aegypti. *Virology* 243(1):99-112.

117. Petoukhov MV*, et al.* (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* 45:342-350.

118. Summers DF & Maizel JV, Jr. (1968) Evidence for large precursor proteins in poliovirus synthesis. *Proc Natl Acad Sci U S A* 59(3):966-971.

119. Kitamura N*, et al.* (1981) Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* 291(5816):547-553.

120. Racaniello VR & Baltimore D (1981) Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc Natl Acad Sci U S A* 78(8):4887-4891.

121. Jore J, De Geus B, Jackson RJ, Pouwels PH, & Enger-Valk BE (1988) Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 in vitro. *The Journal of general virology* 69 (Pt 7):1627-1636.

122. Ypma-Wong MF, Dewalt PG, Johnson VH, Lamb JG, & Semler BL (1988) Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology* 166(1):265-270.

123. Harris KS, Reddigari SR, Nicklin MJ, Hammerle T, & Wimmer E (1992) Purification and characterization of poliovirus polypeptide 3CD, a proteinase and a precursor for RNA polymerase. *J Virol* 66(12):7481-7489.

124. Xiang W, Harris KS, Alexander L, & Wimmer E (1995) Interaction between the 5'-terminal cloverleaf and 3AB/3CDpro of poliovirus is essential for RNA replication. *J Virol* 69(6):3658-3667.

125. Belov GA, Habbersett C, Franco D, & Ehrenfeld E (2007) Activation of cellular Arf GTPases by poliovirus protein 3CD correlates with virus replication. *J Virol* 81(17):9259-9267.

126. Chance MR*, et al.* (2004) High-throughput computational and experimental techniques in structural genomics. *Genome Res* 14(10B):2145-2154.

127. Shi W*, et al.* (2005) Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics. *Structure* 13(10):1473-1486.

128. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2256-2268.

129. Holm L & Sander C (1996) Mapping the protein universe. *Science (New York, N.Y* 273(5275):595-603.

130. Krishna S, Majumdar I, & Grishin N (2003) Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Research* 31(2):532.

131. Berg JM (1990) Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J Biol Chem* 265(12):6513-6516.

132. Sobell HM (1985) Actinomycin and DNA transcription. *Proc Natl Acad Sci U S A* 82(16):5328-5331.

133. Konermann L, Pan J, & Liu YH (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society reviews* 40(3):1224-1234.

134. Peroutka Iii RJ, Orcutt SJ, Strickler JE, & Butt TR (2011) SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods Mol Biol* 705:15-30.

135. van Montfort RL & Workman P (2009) Structure-based design of molecular cancer therapeutics. *Trends in biotechnology* 27(5):315-328.

136. Schneider G & Fechner U (2005) Computer-based de novo design of drug-like molecules. *Nature reviews. Drug discovery* 4(8):649-663.

137. Talele TT, Khedkar SA, & Rigby AC (2010) Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. *Current topics in medicinal chemistry* 10(1):127-141.

138. Harris R, Olson AJ, & Goodsell DS (2008) Automated prediction of ligand-binding sites in proteins. *Proteins* 70(4):1506-1517.

139. Berger AB, Vitorino PM, & Bogyo M (2004) Activity-based protein profiling: applications to biomarker discovery, in vivo imaging and drug discovery. *American journal of*

pharmacogenomics : genomics-related research in drug development and clinical practice 4(6):371-381.

140. Speers AE & Cravatt BF (2009) Activity-Based Protein Profiling (ABPP) and Click Chemistry (CC)-ABPP by MudPIT Mass Spectrometry. *Current protocols in chemical biology* 1:29-41.

141. Nasar F*, et al.* (2012) Eilat virus, a unique alphavirus with host range restricted to insects by RNA replication. *Proc Natl Acad Sci U S A* 109(36):14622-14627.

142. Neuvonen M (2011) Functions of Alphavirus Macrodomain-Containing Protein nsP3. (University of Helsinki, Helsinki).

143. Otsuki K, Maeda J, Yamamoto H, & Tsubokura M (1979) Studies on avian infectious bronchitis virus (IBV). III. Interferon induction by and sensitivity to interferon of IBV. *Archives of virology* 60(3-4):249-255.

144. Andzhaparidze OG, Bogomolova NN, Boriskin YS, Bektemirova MS, & Drynov ID (1981) Comparative study of rabies virus persistence in human and hamster cell lines. *J Virol* 37(1):1-6.

145. Clarke JB & Spier RE (1983) An investigation into causes of resistance of a cloned line of BHK cells to a strain of foot-and-mouth disease virus. *Veterinary microbiology* 8(3):259-270.

146. Stanwick TL & Hallum JV (1974) Role of interferon in six cell lines persistently infected with rubella virus. *Infection and immunity* 10(4):810-815.

147. Truant AL & Hallum JV (1977) A persistent infection of baby hamster kidney-21 cells with mumps virus and the role of temperature-sensitive variants. *Journal of medical virology* 1(1):49-67.

148. Rosztoczy I, Papos M, & Megyeri K (1986) Different interferon-producing capacities of L929 cell sublines and the enhancement of interferon production by priming are controlled pretranslationally. *FEBS letters* 208(1):56-58.

149. Rosenthal F*, et al.* (2013) Macrodomain-containing proteins are new mono-ADP-ribosylhydrolases. *Nat Struct Mol Biol* 20(4):502-507.

150. Miller ML & Brown DT (1992) Morphogenesis of Sindbis virus in three subclones of Aedes albopictus (mosquito) cells. *J Virol* 66(7):4180-4190.

151. Pastorino BA*, et al.* (2008) Expression and biochemical characterization of nsP2 cysteine protease of Chikungunya virus. *Virus research* 131(2):293-298.

152. Russo AT & Watowich SJ (2006) Purification, crystallization and X-ray diffraction analysis of the C-terminal protease domain of Venezuelan equine encephalitis virus nsP2. *Acta*

crystallographica. Section F, Structural biology and crystallization communications 62(Pt 6):514-517.

153. Putnam CD, Hammel M, Hura GL, & Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics* 40(3):191-285.

154. Berger F, Ramirez-Hernandez MH, & Ziegler M (2004) The new life of a centenarian: signalling functions of NAD(P). *Trends in biochemical sciences* 29(3):111-118.

155. Corda D & Di Girolamo M (2003) Functional aspects of protein mono-ADP-ribosylation. *The EMBO journal* 22(9):1953-1958.

ABBREVIATIONS

ABPP: activity-based protein profiling

AUD: alphavirus unique domain

bp: base pair

CHIKV: chikungunya virus

dsDNA: double stranded DNA

dsRNA: double stranded RNA

E1: envelope protein 1

E2: envelope protein 2

E3: envelope protein 3

FL: full length

FT: flow through

GST: glutathione S-transferase

HDX: hydrogen-deuterium exchange

HIV: human immunodeficiency virus

HP: hairpin

HVR: hypervariable region

IPTG: isopropyl β-D-1-thiogalactopyranoside

ITC: isothermal titration calorimetry

IU: infectious units

MBP: maltose binding protein

MOI: multiplicity of infection

mRNA: messenger RNA

MS: mass spectrometry

MT-like: methyltransferase-like domain

NIAID: National Institute of Allergy and Infectious Diseases

NIH: National Institutes of Health

NRAMP: natural resistance-associated macrophage protein

nsP: nonstructural protein

ONNV: O'nyong nyong virus

ORF: open reading frame

$P23^{pro\text{-}zbd}$: polyprotein P23 containing nsP2 protease domain through nsP3 zinc binding domain

PFU: plaque forming units

RdRp: RNA-dependent, RNA polymerase

RMSD: root mean square deviation

RNAPII: RNA polymerase II

SARS: severe acute respiratory syndrome

SAXS: small angle X-ray scattering

SEM: standard error of the mean

SFV: Semliki Forest virus

SINV: sindbis virus

ssRNA: single stranded RNA

Std: standard

SUMO: small ubiquitin-like modifier

TCEP: Tris(2-carboxyethyl)phosphine

VEEV: Venezuelan equine encephalitis virus

WEEV: western equine encephalitis virus

wt: wild type

ZBD: zinc binding domain