Nonstructural proteins (nsP1 to -4), which are expressed as one protein, are considered a replication intermediate.

RNAs are made in vast excess, the negative-sense RNA strand is the template for subgenomic RNA containing the second cistron. The genomic RNA is used as an mRNA for the transcription of the viral replication machinery, and progeny, genomic RNA. Because genomic and subgenomic RNAs are made in vast excess, the negative-sense RNA strand is a complementary negative-sense RNA. The negative-sense strand is a polymerase template for the synthesis of RNA, which is then used as a template for the synthesis of negative-sense RNA. The negative-sense RNA strand is a replication intermediate.

The alphavirus genome consists of single-stranded, positive-sense RNA of ~9–11 kb with a 5’ cap structure and 3’ polyadenosine tail. The genome contains two cistrons. The first, located in the 5’ two-thirds of the genome, encodes the viral replication machinery, termed nonstructural proteins (nsPs), whereas the structural proteins that form virus particles are encoded in the second cistron. The genomic RNA is used as an mRNA for the translation of the nsPs and as a template for the synthesis of complementary negative-sense RNA. The negative-sense strand is a template for subgenomic RNA containing the second cistron and progeny, genomic RNA. Because genomic and subgenomic RNAs are made in vast excess, the negative-sense RNA strand is considered a replication intermediate.

The alphavirus replication machinery is composed of four nonstructural proteins (nsP1 to -4), which are expressed as one of two polyproteins (P123 or P1234). P1234 is expressed as a read-through of an opal termination codon at the end of nsP3. These precursor polyproteins are cleaved by a protease within nsP2 (1, 4). After translation of P1234, cleavage at the P3/4 junction occurs either in cis or trans, followed by the P1/2 junction, which occurs in cis only (4, 5). Both P123+nsP4 and nsP1+P23+nsP4 preferentially synthesize negative strand viral RNA (6, 7). The final cleavage event between P23 produces fully mature nsPs and switches RNA template for synthesis of positive-sense genomic and subgenomic RNAs. The correlation between P23 cleavage and the switch from negative- to positive-sense RNA production is poorly understood.

In mammalian cells, alphavirus infection inhibits host gene expression, causes severe cytopathic effects (CPEs), disrupts the cellular response to IFN, and leads to cell death (8–11). Mature nsP2 contains an amino-terminal RNA helicase domain, a central protease domain that catalyzes all cleavages between the nonstructural proteins, and an inactive RNA methyltransferase-like (MT-like) domain (12). During infection, a portion of nsP2 localizes to the nucleus (13, 14) and plays a role in shutting off host-cell transcription and viral cytopathogenicity (9, 10). Two genetically distinct groups of alphaviruses termed New World and Old World have emerged because geographic isolation (8). In Old World alphaviruses, such as Sindbis virus (SINV) and Semliki Forest virus (SFV), transcriptional shutoff is nsP2-dependent (15). Even in the context of a replicon lacking the structural proteins, the nsPs of SINV and SFV alone can efficiently shut off host transcription and induce severe CPE (16).

Mutations that cause alphaviruses to establish a persistent infection with minimal or no CPE have been identified (16–20). A key feature of these mutants is their inability to inhibit host-cell transcription (10, 15). The best-characterized noncytopathic mutations map to nsP2 proline 726 (P726) within the MT-like domain (10, 16, 17, 21, 22). The P726 mutation has been shown to attenuate the cytopathic effect caused by the expression of wild-type SINV nsP2 protein alone (10). Noncytopathic mutations of nsP2 P726 also reduce viral RNA replication levels and are unable to shut off host transcription and translation processes, showing unexpected insights into viral protein processing and pathogenesis that may be applicable to other polyprotein-encoding viruses such as HIV, hepatitis C virus (HCV), and Dengue virus.

Alphaviruses are small, enveloped RNA viruses that infect various vertebrates, including humans, horses, birds, and rodents (1). Transmission occurs via an invertebrate vector, usually mosquitoes. In humans, typical alphavirus infection can result in rash, arthritis, encephalitis, and death. Within the past decade, infection by Chikungunya virus (CHIKV), a member of the genus alphavirus, caused massive epidemics in Africa and Southeast Asia, with positive cases reported in Europe and the United States, raising the possibility of an emerging threat (2). Additionally, CHIKV often presents with nonclassic symptoms that go undiagnosed, making global pathogenicity difficult to monitor (3).

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4GUA).

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4GUA).


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4GUA).

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 4GUA).
the overall importance of nsP2 to viral replication, as well as host-cell interaction (11, 16). The function(s) of nsP3 in alphavirus replication remains unresolved. Based on sequence conservation, nsP3 is organized into three domains: an amino-terminal macro or X domain (23), a central alphavirus-specific region, and a hypervariable-sequence carboxyl terminus. Macro domains have been shown to bind ADP ribose (24) with certain viral macro domains also retaining ADP ribose-1′′′′-phosphatase activity (25, 26). Recent work in SFV shows that residues just after the macro domain of nsP3 play a role in positioning of the P23 cleavage site (27). Additionally, the carboxyl-terminal region of nsP3 contains numerous phosphorylation sites targeted by yet unknown cellular kinases (28, 29).

Here, we present the structure of an uncleaved P23 precursor protein spanning the nsP2 protease domain through to the central, zinc-binding domain (ZBD) of nsP3 from SINV (Fig. 1A). The P23 cleavage site is located in a narrow cleft formed between nsP2 and nsP3 that is inaccessible for proteolysis. Surprisingly, all of the previously reported nsP2 noncytopathic mutants lie at the interface between nsP2 and nsP3. Mutagenesis of residues in nsP3 opposite of nsP2 P726 showed inefficient cleavage of P23, delayed onset of pathogenesis and host gene shutoff, and altered viral RNA synthesis. Because P23 was copurified with large amounts of RNA, we propose a potential RNA-binding surface that extends across both nsP2 and nsP3 and hypothesize that cleaving the P2/3 junction may alter the RNA-binding surface, thereby contributing to template switching by the replication machinery. Our results provide a better understanding of the mechanism of viral polyprotein processing and pathogenesis.

Results

Alphavirus P23pro-zbd Structure. The SINV P23 precursor protein extending from the nsP2 protease domain to the nsP3 central domain was shown to be fully intact by SDS/PAGE and mass spectrometry despite the presence of a competent nsP2 protease active site (Fig. S1 A and B). Sequence alignment of the central domain of nsP3 shows numerous absolutely conserved cysteine residues (Fig. S2), suggesting the presence of a metal ion-binding site. Various purified preparations of the middle region of nsP3 were analyzed by quantitative X-ray fluorescence for common trace metals found in metalloproteins (30, 31) and shown to have one zinc ion per molecule. Crystals of P23pro-zbd diffracted to 2.85-Å resolution and the structure was determined by molecular replacement with structures of nsP2pro (12) and nsP3 macro domain (32) (Table S1). The resulting electron density allowed for an unambiguous trace of the entire polypeptide chain, including the previously uncharacterized nsP3 central region containing the zinc-binding site. The location of the zinc ion was confirmed by anomalous difference maps (Table S1 and Fig. S4B).

SINV P23pro-zbd shows a highly compact structure, with the four domains (protease and MT-like from nsP2; macro and ZBD from nsP3) arranged with each domain occupying a vertex of a rectangle (Fig. 1 B–D). The polypeptide chain progresses around the perimeter of the rectangle such that the nsP2 protease and nsP3 ZBD are adjacent to one another. The carboxyl terminus of the P23pro-zbd construct is pointing toward the protease active site (Fig. 1 B–D). A previously described SINV mutant (ts7) defective in RNA synthesis at nonpermissive temperatures contains a F312S substitution (33). The cleavage site is located at the base of a narrow cleft formed by the MT-like and macro domains. Three amino acids on either side of the scissile bond are solvent-exposed, attempts to position rounding the scissile bond are sterically hindered. In a previously described nsP2pro structure, the six carboxyl-terminal amino acids were disordered (12). In the P23pro-zbd structure presented here, the entire polypeptide between nsP2 and nsP3 is ordered, permitting modeling of the P2/3 cleavage site. The P2/3 cleavage site is located at the base of a narrow cleft about 11–13 Å wide formed by the MT-like domain of nsP2 and the macro domain of nsP3 (Fig. 1 F and G). The cleavage site is 40 Å away from the nsP2 protease active site, supporting a trans cleavage mechanism (5, 34, 35). Although the 6 aa surrounding the scissile bond are solvent-exposed, attempts to position the protease domain onto the P2/3 cleavage site using the proposed model (36) resulted in numerous steric clashes.
ZBD in nsP3 Is Essential for Viral Replication. After the P2/3 cleavage site, the polypeptide chain continues into the nsP3 macro domain. The ADP ribose–binding site within the macro domain is solvent-exposed and points away from the other four domains. The polyprotein exits the last helix of the macro domain adjacent to the P2/3 cleavage site and continues into a stretch of nsP3 that lacks secondary structure but forms a twist (Fig. S3). The conserved nsP3 central domain contains an antiparallel α-helical bundle, two parallel β-strands, and a previously unknown zinc-coordination site (Fig. S4B). The zinc ion is coordinated by C263, C265, C288, and C306 of nsP3. C263 and C265 are located in the loop between the last two α-helices, whereas C288 and C306 are at the carboxyl termini of the two parallel β-strands (Fig. S3C). Each of the zinc coordinating cysteines is invariant among all alphavirus sequences (Fig. S2). Structural comparison of the nsP3 ZBD against all known folds in the Protein Data Bank using PDBfolds (37) and Dali servers (38) failed to produce any statistically significant hits, and no metalloprotein folds were identified. The spatial arrangement of secondary structural elements that contribute to nsP3 zinc-binding site, i.e., two cysteines in a loop between alpha helices and the other two cysteines at the end of two parallel β-strands, represents a structural scaffold for zinc-ion coordination (39).

Site-directed mutagenesis was performed to examine the functionality of the four conserved cysteine residues involved in zinc coordination in nsP3 (Fig. S4A). All four SINV mutants (C263A, C265A, C288A, and C306A) failed to produce productive infection in BHK cells, as evaluated by infectious center and plaque assays (Fig. S4G). A conserved cysteine residue in nsP3 (C247) unrelated to the zinc-binding site served as a control and C247A mutant produced infectious particles to near wild-type levels. Zinc-coordination mutants failed to express the nonstructural polyprotein and replicate virus, thereby supporting the structural role of the zinc ion (Fig. S4D).

The coordination site forms a distorted tetragon due to the presence of two highly conserved serine residues (S289 and S290) below the zinc atom (Fig. S2). Mutation of these serine residues did not alter viral replication levels. Furthermore, a double serine to alanine mutant (S289A+S290A) resulted in decreased viral RNA infectivity, but overall viral titers were comparable to wild type (Fig. S4C).

Importance of P23 Interface for Polyprotein Processing and Viral Pathogenesis. The interface between the macro and ZBD of nsP3 has a buried surface area of 570 Å², whereas the linker connecting the two domains is extended, making nsP3 into a ring-like structure with an inner diameter of 15–18 Å (Fig. 2A). Together, nsP2 and nsP3 share an extensive interface with 3,000 Å² surface area. nsP3 encircles the MT-like domain of nsP2, with residue R781 of nsP2 protruding through the opening (Fig. 2C). The interface between nsP2 and nsP3 is charged, with the nsP2 surface being mostly basic, whereas nsP3 is generally acidic (Fig. 2B and C).

Several groups have identified mutations in the carboxyl-terminal portion of nsP2 that result in noncytopathic virus and persistent infection (16–20). These mutations reduce viral RNA replication levels with minimal effects on host gene expression and viral yields. All of the noncytopathic mutations map to the surface of nsP2 at the nsP3 interface (Fig. 2D). The mutations follow the encircling path made by the macro to ZBD linker. Mutagenesis of P726 in SINV to amino acids with large side chains (Phe, Tyr, Leu, Arg, and Gln) caused a dramatic decrease in RNA replication and were noncytopathic, whereas substitutions with small side chains (Ala or Val) behaved like wild-type virus, with Gly, Ser, and Thr substitutions giving an intermediate effect (16). P726 is located in the loop connecting α9 with β11 and makes direct contact with nsP3 linker just after the macro domain (Fig. 2E). There is a striking, direct correlation between cytopathogenicity and the level of RNA replication with the size of the mutant side chain. Molecular modeling analyses demonstrated that mutating P726 to Phe, Tyr, Leu, Arg, or Gln resulted in numerous steric clashes with nsP3, whereas Ser, Thr, Ala, and Val had fewer conflicts. Moreover, the P726G mutation could destabilize the α9-β11 loop, causing it to interfere with P23 interaction. Attempts to purify P23protein with a P726S or P726L mutation resulted in poorly soluble and aggregated protein, supporting the importance of this position.

To investigate how destabilizing P23 interface may contribute to changes in viral pathogenesis and replication, site-directed mutagenesis to the corresponding surface of SINV nsP3 in close proximity to nsP2 P726 were created (Fig. 2E). The side chains of V162 and L165 of nsP3 are ~4.6 and 3.3 Å away from nsP2 P726 side chain, respectively (Fig. 2F). V162 and L165 of nsP3 were mutated to one of the following: Ala, Asp, Glu, Phe, Arg, or Trp, nsP2 P726 to Ser, Gly, or Leu mutants were used for comparison. RNA infectivity measured by infectious center assay on BHK cells were reduced in P726L, L165D, L165E, and L165R mutants compared with the wild-type SINV (Fig. 3A). P726L mutation in nsP2 reduced RNA infectivity to only 2% of wild-type levels, whereas the three nsP3 L165 mutants averaged an RNA infectivity at 8% of wild type. These four mutants also displayed a small size plaque phenotype in infectious center assay (Fig. S5A–D). In addition,
indicated hours postinfection and analyzed by Western blot using anti-nsP2 harvested from wild-type and indicated mutant SINV-infected BHK cells at cleavage of nsP2 P726 and nsP3 L165 mutants in vivo. Total protein was measured independent experiments. Error bars represent the SEM. (A) and anti-nsP3 (C) antibodies.

L165A, L165D, L165E, and L165R mutants produced lower viral titers at 24 h after electroporation, only reaching an average of 34% of wild-type levels (Fig. 3B). Of the nsP3 mutants generated, L165D had the lowest titers, only producing 26% compared with wild-type titers. All others, including all three P726 mutants, reached wild-type levels of virus titers. All V162 mutants behaved similarly to wild type in both RNA infectivity and viral titer and were not pursued further. None of the nsP3 mutants displayed a noncytopathic phenotype in the TSG/puromycin N-acetyltransferase (PAC) replicon background (16), although onset of CPE in L165 to D, E, and R mutants was delayed.

Radiolabeling experiments were performed to evaluate both viral and host protein and RNA production over the course of infection and to analyze efficiency of host-cell transcriptional and translational shutoff caused by SINV nsP3 mutations. Wild-type SINV is able to halt host translation completely within 6 h postinfection, whereas P726S and P726G mutants were unable to do so even up to 12 h postinfection, as indicated by persistent actin production (Fig. S6). nsP2 P726L and the four selected nsP3 L165 mutants (Asp, Gln, Arg, Trp) were able to shut off host translation, although at a slightly delayed time postinfection compared with wild type (Fig. S6).

Tritium-labeled uridine was used to track viral and host RNA production over the course of infection. In the presence of actinomycin D, both genomic and subgenomic viral RNA production can be clearly observed in wild-type SINV starting as early as 2 h postinfection and remaining stable until 12 h (Fig. 3C). Levels of subgenomic and genomic RNA in P726G and P726S mutants are reduced compared with wild type. nsP2 P726L and the four nsP3 L165 mutants exhibit a slow initial rate of transcription especially of genomic RNA; however, these mutants reach wild-type levels by 4 h postinfection. Thereafter, the levels of RNA replication in the L165 mutants continue to increase, and even surpass, wild-type RNA replication, especially subgenomic RNA (Fig. 3C). RNA labeling in the absence of actinomycin D showed complete inhibition of host cellular transcription by 8 h postinfection as demonstrated by decreasing 28S and 18S cellular rRNA (Fig. 3D). P726G and P726S mutants exhibited a lack of host-cell transcriptional shutoff with rRNA persisting up to 12 h postinfection. The nsP2 P726L mutant and the four nsP3 L165 mutants showed a delay in host transcription inhibition but eventually achieved complete shutoff 12 h postinfection (Fig. 3D).

Given that residues R159 and E163 of SFV nsP3 have been shown to affect P2/3 cleavage (27), we assessed the effect of the SINV nsP3 mutations on cleavage activity. nsP3 linker region mutants L165D and L165E revealed drastic inefficiency in P23 cleavage that was not seen in L165W and L165R mutants (Fig. 3E and F and Fig. S7). The presence of uncleaved P23 in the nsP3 L165D and L165E mutants started 6 h postinfection and persisted up to 12 h, whereas P726 mutants, as well as nsP3 L165R and nsP3 L165W, showed substantially lower amounts of uncleaved P23. Western blots probing with anti-nsP2 antibodies were used to confirm the presence of uncleaved P23 (Fig. 3F). nsP3 appears as a smear, indicative of alternate phosphorylation states and possibly other post-translational modifications (40). The nsP2 P726 mutants may also accumulate precleavage forms but at very low concentrations that do not build up over time, as in the case of the nsP3 linker mutants.

Discussion

The structure of alphavirus precursor polyprotein P23pro-zbd presented here reveals the arrangement of four domains (protease, MT-like, macro, and ZBD) in a precleavage form. Based on the structure of P23pro-zbd, several inferences can be made about the mechanism of alphavirus polyprotein cleavage. First, the carboxyl terminus of the P23pro-zbd structure is pointing toward the nsP2 protease active site about 40–45 Å away. It is possible that the large, flexible carboxyl region of nsP3 could span the distance, placing the P3/4 junction in the protease active site for cis cleavage (34). Second, the 40 Å distance between the P2/3 cleavage site and the nsP2 protease active site supports the current trans model of cleavage (5, 34, 35). Third, the inaccessibility of the P23 cleavage site itself indicates that access is tightly regulated. It is possible the activator sequence present in the extreme

Fig. 3. In vivo effects of nsP3 linker region mutations. (A and B) Viral RNA infectivity [infectious units (IU)/μg RNA] as measured by infectious center assay (A) and viral titers [plaque-forming units (pfu)/mL] (B) 24 h after electroporation of BHK cells, as determined by plaque assay over two or more independent experiments. Error bars represent the SEM. (C and D) Cells were infected with wild-type or mutant SINV at a multiplicity of infection (MOI) of 10. At indicated time points postinfection, RNAs were labeled for 3 h in the presence (C) and absence (D) of actinomycin D. Total RNAs were harvested and analyzed by denaturing agarose gel electrophoresis. The position of genomic (G), subgenomic (SG), 28S, and 18S RNAs are labeled. M indicates mock-infected cells labeled for 3 h. (E and F) Analysis of P23 cleavage of nsP2 P726 and nsP3 L165 mutants in vivo. Total protein was harvested from wild-type and indicated mutant SINV-infected BHK cells at indicated hours postinfection and analyzed by Western blot using anti-nsP2 (E) and anti-nsP3 (F) antibodies.
During the course of alphavirus infection, a portion of nsP2 is localized to the nucleus, whereas nsP3 remains cytoplasmic (13, 14, 41). At the moment, it is unclear how nsP2 would dissociate post-cleavage from the extensive, encircling grasp of nsP3. Comparison of the domains in P23pro-zbd with the isolated structures of nsP2pro (12) and nsP3 macro domain (32) demonstrated that the protease and macro domains were highly similar with a root mean square deviation (rmsd) of <0.6 Å, whereas the MT-like domains have an rmsd of 1.3 Å for similar α carbon. A superposition of the nsP2 structures shows the discrepancy between nsP2pro and P23pro-zbd is limited to specific regions within the MT-like domain (Fig. 4A and B). α-Helices α8, α10, and α11, as well as β-strands β11, β12, and β13, between the two structures superimpose well with an rmsd of <1 Å, although the β13-α10 loop moves 5 Å because of interactions with the nsP3 linker. In addition, the nsP3 linker interacts with α9, causing the helix to rotate 8° with the carboxyl-terminal end of the helix moving 3.8 Å to form a single continuous helix with α8. This movement of the α9 helix causes the β-sheet to become more concave. The distortion of the MT-like domain in the P23pro-zbd becomes more pronounced the further the secondary structures are from α9, with α7 and β9 moving about 4 Å from their corresponding positions in nsP2pro. Perhaps movement of the MT-like domain facilitates separation of nsP2 and nsP3 post-cleavage, which is triggered by binding viral proteins or RNA to the concave side of the nsP2 MT-like β-sheet.

Previously described nontypopathetic 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 nontypopathetic phenotype. In contrast to the nsP2 P726G and P726S mutants, nsP3 L165 mutants had enhanced RNA replication levels, indicating the 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, however, 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 (Fig. 2F), suggesting that the P23 interface may be stabilized by the introduced acidic amino acid at L165 through interaction with the surface of nsP2. This stabilization 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 and 17 Å from Cα 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 (Fig. 2F). A previously described E163R mutation in SFV nsP3 that altered P2/3 cleavage is completely solvent-exposed (Fig. 2E) (27). 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.

Cleavage between P2/3 dictates template use by the alphavirus replication machinery. During purification, P23pro-zbd bound extensive amounts of bacterial RNA, which could only be removed by high concentrations of salt (Fig. S1A and C). The electrostatic potential of the P23pro-zbd structure identified a large basic surface that centers on the zinc-binding site of nsP3 and extends to nsP2 (Fig. 4A). Many zinc metalloproteins are involved in nucleic acid binding and gene expression (42); the location of a basic patch in close proximity to the zinc-binding site suggests a potential role in RNA binding. Several molecules of sulfate and 2-(N-morpholino)ethane sulfonic acid (MES), used in crystallization, were bound to P23pro-zbd along the basic surface, which may indicate potential binding sites for the phosphate groups of nucleic acids (Fig. 4C–E). In fact, a molecule of MES occupies the ADP ribose–binding site in each of the three macro domains in P23pro-zbd in the asymmetric unit, as well as in the macroH2A1.1 structure (Fig. 4E) (43). The alphavirus genome contains conserved sequence elements (CSEs) that serve as promoters for production of positive- and negative-sense genomic RNA, as well as subgenomic RNA. Mutations in a 31-nt-long CSE in the nsP1-coding region of SINV disrupted the RNA structure without affecting the coding sequence and demonstrated reduced replication in mosquito cells but not in mammalian cells (44). Adaptive mutations (E118K of nsP2 and T286I of nsP3) were shown to restore replication of the mutated CSE. T286 is 2 aa before one of the zinc-coordinating cysteine residues (C288) (Fig. 4F) and is located on the basic surface of the ZBD. T286 is 6.1 Å away from a sulfate ion that is coordinated by three highly conserved amino acids in nsP3: Y267, R273, and R276 (Fig. 4F). Taken together, these data strongly implicate this surface as a potential RNA-binding site.

Many important human viral pathogens encode a polyprotein that is cleaved by viral or cellular proteases. Given the wide prevalence of regulated polyprotein processing seen in a number...
of viral families, there is lack of structural information regarding the precursor forms. Currently, alphavirus P23pro-zbd and poliovirus 3CD (45) are the only structures of polyprotein precursor from a viral replication machinery. Both alphavirus P23pro-zbd and poliovirus 3CD structures contain the protease domain responsible for cleaving the precursor, which is thought to proceed by a trans mechanism. However, 3CD has no intramolecular contacts and a solvent accessible cleavage site that is regulated by intermolecular contacts, whereas P23pro-zbd has an extensive interface and an inaccessible cleavage site. The results presented here further expand our current understanding of viral polyprotein processing, replication, and cytopathogenesis, which may be applicable to other positive-sense RNA viruses.

Materials and Methods

Complete materials and methods are provided in SI Materials and Methods.

Structure Determination. P23pro-zbd of the SINV strain Toto101 was expressed in Escherichia coli as a fusion with GST. The protein was purified by sequential chromatography over GST affinity, hydroxyapatite, and heparin columns. Crystals of P23pro-zbd were grown by vapor diffusion against 2 mM ammonium sulfate and 0.1 M MES (pH 6.5). Crystals were transferred to a cryoprotectant solution containing 2 mM ammonium sulfate, 0.1 M MES (pH 6.5), 35% (vol/vol) xylitol, and 5% (vol/vol) trimethylamine N-oxide and then flash-cooled in liquid nitrogen. Diffraction data were collected at the zinc, K absorption edge (1.2824 Å) to 2.85 Å resolution at beam line X25 at Brookhaven National Synchrotron Light Source (Table S1). Phases were calculated by molecular replacement using the nsP29 (12) and nsP3 macro structures (32). The final model contains 3 molecules of P23pro-zbd, 3 zinc ions, 31 sulfite ions, and 5 molecules of MES. Atomic coordinates have been deposited in the Protein Data Bank under PDB id code 4GUA.

Virological Studies. Infectious center and plaque assays were performed similarly to those described in ref. 46. Protein and RNA radiolabeling were performed as described in ref. 9. Harvested protein from BHK-J cells at time points postinfection were used in Western blotting experiments, and membranes were probed with anti-nsP2, anti-nsP3, and anti-GAPDH antibodies.

ACKNOWLEDGMENTS. We thank J. Bonanno for assistance with the data processing. C. Rice and M. MacDonald for providing BHK-J cells, Sindbis virus DNA, and nsP3 antibody. J. Millong and P. Matteson for the GAPDH antibody, and A. Basant, J. Chiu, F. Jiang, A. Khan, C. Rice, A. Stock, V. Stollar, and J. Whidby for insightful discussions. S.A.Y. was supported by the NIH-NIAID training Grant T32AA070433-11. This study was supported by National Institutes of Health Grants AI080659 (to J.M.) and DK08353 and AI070101 (to A.G.), New Jersey Commission on Cancer Research Grant 10-1962-COR-E (to J.A.G.), and PHS Merck Research Grant RR-00165 (to A.G.). We wish to express our profound gratitude to the late Dr. Aaron Shatkin for his unwavering support, sage advice, and insightful discussions.

Supporting Information

Shin et al. 10.1073/pnas.1210418109

SI Materials and Methods

Protein Expression and Purification. P23<sup>pro-<em>zbd</em></sup>-<em>b</em> 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, was expressed overnight at 18 °C in <em>Escherichia coli</em> as a fusion with GST. After lysis, the soluble fraction was loaded onto a GSTrap (GE Healthcare) column in a low-salt buffer [20 mM Heps (pH 7.5), 150 mM KCl, and 5% glycerol]. After equilibration in the low-salt buffer, the column was extensively washed with 20 mM Heps (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 on a heparin Sepharose column (GE Healthcare) according to the manufacturer’s protocol.

Protein Crystallography. The purified protein was pooled, buffer-exchanged into 20 mM Heps (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<sup>pro-<em>b</em></sup> were grown by the vapor diffusion method against 2 M ammonium sulfate and 0.1 M 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 oxide and then flash-cooled in liquid nitrogen. Diffraction data were collected at 1.282 Å and 100 K to 2.85 Å resolution at beam line X25 at Brookhaven National Synchrotron Light Source (Table S1). The data were integrated and scaled with MOSFLM and Scala (1, 2). Phases were determined by molecular replacement using the nsP2<sup>pro</sup> (3) and nsP3 macro structures (4). The location of the endogenous zinc ion was determined by SHELX (5) using anomalous difference maps. The model was improved with several cycles of refinement.

Cell Culture. BHK-J cells (provided by Charles Rice, Rockefeller University, New York, NY) were maintained in Minimum Essential Medium (MEM) (Life Technologies) supplemented with 7.5% FBS at 37 °C and 5% CO<sub>2</sub>.

Plasmid Constructs. Mutations were introduced into full-length SINV cDNA clone pToto1106 (11) 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.

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). Nondenaturing gel electrophoresis was used to verify RNA integrity. Aliquots of 1 μg were stored at −80 °C until transfection.

RNA Transfection. BHK-J cells were harvested and centrifuged at 335 × g for 5 min at 4 °C. Cell pellets were washed twice with 1× Dulbecco’s phosphate-buffered saline (DPBS) (Life Technologies), counted by trypan blue exclusion, and resuspended in 1× DPBS to 2 × 10<sup>5</sup> cells/mL. Four hundred microliters of the cell suspension (8 × 10<sup>5</sup> cells) were mixed with 1 μg of RNA and immediately placed in a 2-mm gap cuvette for electroporation. Samples were subjected to five 99-us pulses at 550 V in an ECM 830 Electro Square Porator (Harvard Apparatus) and allowed to rest 10 min at room temperature. Ten-fold serial dilutions of electroporated cells were made in MEM containing 1% FBS for infectious center assay. The remaining cells were resuspended in MEM containing 7.5% FBS and plated in 150 mm dishes. Virus was harvested 24 h after electroporation and aliquoted for storage at −80 °C.

Infectious Center Assay. BHK-J subconfluent monolayers were prepared in six-well plates at least 12 h in advance at 5 × 10<sup>5</sup> cells per well for infectious center assay. Ten-fold serial dilutions of electroporated cells were seeded on these monolayers and incubated for 1 h at 37 °C before being overlaid with 1.2% agarose containing MEM, 2% FBS, and 1x penicillin/streptomycin. After 48 h of incubation at 37 °C, cells were fixed with 7% formaldehyde and stained with crystal violet.

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

Protein Synthesis During Sindbis Virus Infection. BHK-J cells were prepared in six-well plates at 5 × 10<sup>5</sup> cells per well for infectious center assay. Ten-fold serial dilutions of electroporated cells were seeded on these monolayers and incubated until specified time points. At 2, 3, 6, 9, and 12 h postinfection, cells were washed three times with PBS, and media were replaced with 1 mL of Dulbecco’s modified eagle medium (lacking methionine and cysteine) with 1% FBS and 20 μCi of [<sup>35</sup>S] in EasyTag EXPRE<sup>®</sup>S<sup>35</sup>S Protein Labeling Mix (Perkin-Elmer). After 1 h, cells were washed three times with PBS and lysed with 200 μL of SDS/PAGE loading buffer. Ten-microliter samples were run on 10% SDS/PAGE gels, stained with Coomassie blue, dried, and autoradiographed.

RNA Synthesis During Sindbis Virus Infection. BHK-J cells were prepared in six-well plates at 5 × 10<sup>5</sup> cells per well and infected with virus at 10 pfu per cell in MEM supplemented with 1% FBS for 1 h at 37 °C. Media were then replaced, and cells were incubated until specified time points. At 2, 4, 8, and 12 h postinfection, cells were washed three times with PBS, and media were replaced with 1 mL of Dulbecco’s modified eagle medium (lacking methionine and cysteine) with 1% FBS and 20 μCi of [<sup>35</sup>S] in EasyTag EXPRE<sup>®</sup>S<sup>35</sup>S Protein Labeling Mix (Perkin-Elmer). After 1 h, cells were washed three times with PBS and lysed with 300 μL of 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 of water.

Denaturing RNA Gel Electrophoresis. Five-microliter [<sup>5</sup>H]-uridine-labeled RNA samples were denatured with 25 μL of glyoxal reaction mixture (6 mL of DMSO, 2 mL of deionized glyoxal, 1.2
mL of 10x BPTE electrophoresis buffer, 0.6 mL of 80% glycerol, and 0.2 mL of water) incubated for 1 h at 55 °C. Before loading, samples were chilled for 10 min on ice and mixed with 2 μL of RNA loading buffer. Denatured RNA was loaded on a 0.8% agarose 1x BTPE gel and run in 1x BTPE 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.

**Immunization and Antibody Generation.** nsP2-protease domain was produced and purified via the method described above, similar to the P23apo-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 d with the same dosage in incomplete Freund’s adjuvant. A booster injection was given i.v. 3 d before 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 resuspended in Iscove’s modified Dulbecco’s media (HyClone) supplemented with 1x hypoxanthine-aminopterin-thymidine (Sigma) plus 10% FBS.


Fig. S5. Purification and RNA analysis of P23apo-zbd. (A) Chromatogram showing the purification of GST- P23apo-zbd over glutathione affinity column. The red and blue lines represent the absorbance in milliabsorbance units (mAU) at 254 and 280 nm, respectively. The expressed protein was loaded onto the column in a low-salt buffer with 150 mM KCl and washed until the absorbance returned to baseline (FT). Then, the column is washed with buffer containing 1 M KCl (high-salt wash). The protein is eluted off the column with a low-salt buffer containing reduced glutathione. (B) SDS-PAGE analysis of the purification in A. The position of full-length, intact P23apo-zbd protein is indicated. The asterisk (*) denotes a degradation product of P23apo-zbd resulting from an authentic cleavage at the P2/3 site. Mass spectrometry identified the protein to be nsP2 and a portion of the nsP3 macro domain, terminating at a solvent exposed loop. This cleavage reaction is not catalyzed by the nsP2 protease, because expression of a protease defective mutant in nsP2 also produces the same band. (C) The high absorbance at 254 nm relative to 280 nm and the lack of protein in the high-salt wash by SDS-PAGE analysis suggested the presence of nucleic acids. An aliquot of the high-salt wash was incubated in the presence of DNase I or RNase A for 30 min at 37 °C, loaded onto a 1% agarose gel, and stained with ethidium bromide. An identical amount of the high-salt wash was used for both analyses in B and C.
**Fig. S2.** Sequence alignment of the conserved central domain of nsP3. Amino acid sequence alignment of SINV nsP3 central domain (amino acids 161–325 from the beginning of nsP3; amino acids 1508–1672 from translation initiation site of polyprotein P1234) with orthologs in Old World, New World, and aquatic alphaviruses, showing the conserved zinc ion–coordinating cysteine residues [denoted by an asterisk (*)]. This domain appears to be unique to the alphavirus genus because a similar domain is not present in the closely related rubella virus genome. Cylinders and arrows above the alignment represent the location of α-helices and β-strands, respectively. Secondary structures were assigned using the program DSSP (2). Sequences used in the alignment are SIN [Sindbis virus; National Center for Biotechnology Information (NCBI) reference sequence NP_740672], AUR (Aura virus; NCBI reference sequence NP_819012), RRV (Ross River virus; NCBI reference sequence NP_041254), BFV (Barmah Forest virus; NCBI reference sequence NP_819897), CHT (Chikungunya virus; NCBI reference sequence ADJ19187), VEE (Venezuelan equine encephalitis; NCBI reference sequence NP_740698), EEE (Eastern equine encephalitis; NCBI reference sequence NP_740651), SPD (Salmon pancreas disease virus; NCBI reference sequence NP_740637), and SDV (sleeping disease virus; NCBI reference sequence NP_740655). Sequences were aligned using multiple alignment using fast fourier transform (MAFFT) (3) and edited for purpose of presentation using JalView version 2 (4).

**Fig. S3.** ZBD of nsP3. (**A** and **B**) Ribbon diagrams of the middle region of nsP3, showing the linker and ZBD. The polypeptide chain is rainbow colored from the amino terminus (blue) to carboxyl terminus (red). The coordinated zinc atom is represented by a gray sphere surrounded by the four coordinating cysteine residues in stick format. The view in **B** is rotated 90° along the vertical axis from **A**. (**C**) Topology model of the nsP3 linker and ZBD, highlighting the location of the cysteine-coordinating residues. The coloring follows the ribbons diagram in **A** and **B**.

---

**Shin et al. www.pnas.org/cgi/content/short/1210418109**
Fig. S4. Mutations in nsP3 ZBD. (A) Stick representation of the nsP3 zinc-binding site. A gray sphere represents zinc. The four conserved cysteines, which directly coordinate the zinc ion are marked along with two also highly conserved serine residues. (B) Another view of the zinc ion with an anomalous difference map contoured at 9σ. (C) Results of infectious center (RNA infectivity), and plaque (viral titer) assays of nsP3 mutants in the zinc coordinating cysteine and highly conserved serine residues. (D) Total protein harvested from cells five and 10 h after electroporation with wild-type or indicated mutant SINV RNA was analyzed by Western blot using anti-nsP2 antibodies. The identical samples were probed with anti-GAPDH as a loading control and shown in the panel below. M indicates mock electroporated cells.
Fig. S5. SINV nsP2 or nsP3 mutants result in smaller plaque sizes. Infectious center assay was performed with BHK cells transfected with SINV RNA, plated on a monolayer of BHK cells at increasing dilutions, and overlaid with agarose. (A) Wild-type SINV RNA–transfected cells at a dilution of $10^3$. (B) SINV nsP2 P726L RNA–transfected cells at a dilution of $10^3$. (C) SINV nsP3 L165D RNA–transfected cells at a dilution of $10^3$. (D) SINV nsP3 L165E RNA–transfected cells at a dilution of $10^2$.

Fig. S6. Host translational shutoff and production of viral proteins after infection with SINV. 35S-labeled proteins harvested from BHK cells over time points postinfection with SINV wild type, nsP2 mutants, or nsP3 mutants at an MOI of 10. Cellular actin and SINV capsid are denoted by an arrow and asterisk (*), respectively. M indicates mock-infected cells.
Fig. S7. Analysis of P23 cleavage of nsP2 P726 and nsP3 L165 mutants in vivo. Total protein was harvested from wild-type and indicated mutant SINV-infected BHK cells at the indicated hours postinfection and analyzed by Western blot using anti-nsP2 antibodies. The identical samples were probed with anti-GAPDH as a loading control and shown in the lower gels. M indicates mock-infected cells.

Table S1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>BNLX25 Zn K-Edge</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P3,21</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>147.3, 147.3, 360.5</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>52.05–2.85 (3.01–2.85)*</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.137 (1.103)</td>
</tr>
<tr>
<td>Rpim</td>
<td>0.044 (0.353)</td>
</tr>
<tr>
<td>I/σI</td>
<td>15.6 (2.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (99.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>10.7 (10.5)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>51.51–2.85</td>
</tr>
<tr>
<td>No. reflections</td>
<td>105,713</td>
</tr>
<tr>
<td>Rwork/Rfree</td>
<td>0.2200/0.251</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>15,594</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>304</td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>63.7</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>89.6</td>
</tr>
<tr>
<td><strong>rmsds</strong></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.004</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>0.841</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
</tr>
<tr>
<td>Most favored (%)</td>
<td>98.1</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>1.9</td>
</tr>
<tr>
<td>Disallowed (%)</td>
<td>0</td>
</tr>
</tbody>
</table>

*Highest resolution shell is shown in parentheses.

\[ R_{p.i.m.} = \frac{1}{N} \sum_{hkl} \sqrt{\frac{\sum_{i=1}^{N} [I_i(hkl) - \langle I(hkl) \rangle]}{\sum_{hkl} I(hkl)}} \]

where I(hkl) is 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 (2).