Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Viral precursor polyproteins: keys of regulation from replication to maturation
Samantha A Yost and Joseph Marcotrigiano

Many viruses use a replication strategy involving the translation of a large polyprotein, which is cleaved by viral and/or cellular proteases. Several of these viruses severely impact human health around the globe, including HIV, HCV, Dengue virus, and West Nile virus. This method of genome organization has many benefits to the virus such as condensation of genetic material, as well as temporal and spatial regulation of protein activity depending on polyprotein cleavage state. The study of polyprotein precursors is necessary to fully understand viral infection, and identify possible new drug targets; however, few atomic structures are currently available. Presented here are structures of four recent polyprotein precursors from viruses with a positive sense RNA genome.

Addresses
Center for Advanced Biotechnology and Medicine, Department of Chemistry and Chemical Biology, Rutgers University, 679 Hoes Lane West, Piscataway, NJ 08854, United States

Corresponding author: Marcotrigiano, Joseph
(jmarco@cabm.rutgers.edu)

Precursors from viral genome replication proteins
Alphavirus P23
Alphaviruses are enveloped viruses of the Togaviridae family whose genome consists of a positive-sense RNA of ~9–11 kb with a 5’ cap structure and 3’ polyadenosine tail (Figure 1a). The genome contains two cistrons: the first encodes the viral replication machinery, or nonstructural proteins (nsPs), and the latter contains the structural proteins that form the virion particle. The alphavirus replication machinery is composed of four nonstructural proteins (nsP1 to nsP4), which are expressed as one of two polyproteins (P123 or P1234) caused by read through of an opal codon at the end of nsP3. The nonstructural precursor polyproteins are cleaved in a highly regulated manner by a protease at the C-terminus of nsP2 [2,3]. Shortly after translation of P1234, cleavage at the P3/4 junction occurs either in cis or trans, followed by the P1/2 junction in cis only [3,4]. Cleavage intermediates P123 + nsP4 and nsP1 + P23 + nsP4 preferentially use the genomic RNA to synthesize a negative-sense, intermediate RNA [5,6]. The final cleavage event between P23 yields fully mature nsPs and switches the template to the negative-sense RNA to synthesize positive-sense genomic and subgenomic RNAs.

The structure of an uncleaved P23 precursor protein from Sindbis virus spanning the nsP2 protease domain through to the central, zinc-binding domain (ZBD) of nsP3 (P23pro-zbd) has been recently determined [7**].
P23\textsuperscript{pro-zbd} shows a solvent exposed but inaccessible P2/3 cleavage site located at the base of an 11–13 Å wide cleft (Figure 1b) \cite{7**}. The nsP2 protease active site is 40 Å away from the P2/3 cleavage site, which supports a trans only cleavage mechanism \cite{4,7**,8,9}. Attempts to model the nsP2 protease onto the P2/3 cleavage site resulted in steric clashes, indicating that accessibility of the cleavage site is highly regulated by an unknown mechanism. P23 has an extensive buried surface area of 3000 Å\textsuperscript{2} due to a ring-shape structure created by nsP3, which encircles nsP2 (Figure 1c). Several previously described mutations in nsP2 that confer a noncytopathic phenotype \cite{10–14} map to the surface of nsP2 and make contact with nsP3, implicating this interaction in viral pathogenesis. During the alphavirus lifecycle, nsP2 localizes to the nucleus; however, the mechanism by which nsP2 dissociates from nsP3 post cleavage is unknown. A potential RNA binding surface, which extends over the surface of both nsP2 and nsP3 has been proposed. It is hypothesized that the RNA binding surface may be altered after cleavage, resulting in the RNA template switching that occurs after P23 cleavage.

\section*{Poliovirus 3CD}

Poliovirus is a nonenveloped virus of the \textit{Picornaviridae} family with a positive-sense RNA genome of \(\sim 7–9\) kb. Poliovirus RNA lacks a 5' cap structure and uses an internal ribosomal entry site (IRES) for translation initiation \cite{15,16}. The genome contains a single, open reading frame encoding both structural and nonstructural genes as a large polyprotein \cite{17–19} (Figure 2a). For simplicity, the proteins are grouped into viral capsid proteins (P1), and proteins involved in viral genome replication and protein processing (P2 and P3). Cleavage of the polyprotein is performed by virus-encoded proteases 2A, 3C, or 3CD precursor. 3CD polyprotein is made up of the 3C protease and the viral RNA polymerase (3D) that is responsible for genome replication. The efficiency of 3C protease and 3D polymerase activity is altered depending on the cleavage state. The 3CD polyprotein preferentially cleaves the P1 precursor compared to mature 3C alone \cite{20,21} while the 3D polymerase activity is inhibited in the precursor \cite{22}. Furthermore, 3CD promotes RNA synthesis by a redistribution of cellular Arf proteins involved in membrane trafficking, an activity not triggered by the cleaved forms \cite{23,24**}. These activities highlight the importance of polyprotein cleavage as a mechanism of gene regulation.

The atomic structure of 3CD resembles beads-on-a-string with 3C and 3D connected by a flexible, solvent exposed linker region containing the cleavage site \cite{25**} (Figure 2b). Overall, the structures of 3C and 3D alone are very similar to their counterparts in the 3CD precursor. In sharp contrast to alphavirus P23, 3C and 3D have no intramolecular contacts. The 3CD structure supports a trans cleavage mechanism due to the location...
of the protease active site and the cleavage site [7,25**,26]. In the post cleavage structure of 3D, the N-terminal residues fold inward and insert into a binding pocket where they participate in several hydrogen bonding interactions, while these residues are extended in the 3CD structure. The rearrangement of the N-terminal residues of 3D post cleavage is proposed to be the mechanism for activation of the 3D polymerase activity [25**].

Precursors from viral structural proteins

Flavivirus prM-E

Flaviviruses are enveloped viruses with a positive-sense RNA genome of ~11 kb. The flavivirus structural and nonstructural genes are contained within the same open reading frame and are translated as a large polyprotein (Figure 3a). Both viral and cellular encoded proteases are involved in specific cleavage of the polyprotein. The surface of the virion is composed of the envelope (E) and membrane (M) proteins [27]. E interacts with cellular receptors and contains a loop responsible for membrane fusion [27,28]. M is expressed as a precursor protein, prM, which prevents premature fusion by E and is cleaved by furin during egress to release the pr fragment, resulting in a mature particle. In addition to its role in virion morphogenesis, the prM precursor may assist in proper folding of the E protein [29,30].

The structure of a dengue virus prM-E polyprotein was published in 2008 (Figure 3b) [31**]. The construct included prM residues 1 through 91, an eight amino acid linker in place of a transmembrane region, E residues 1 through 394, and a mutated furin cleavage site between pr and M. E contains three distinct domains, termed DI,
DII, and DIII, with the fusion loop located at the tip of DII [32–37]. The pr peptide has a novel beta barrel fold positioned at the distal end of DII, blocking the fusion loop, consistent with its role in preventing membrane fusion [31,38]. Similar inaccessibility of the fusion loop in immature virus has also been described in Chikungunya structural precursor polyprotein p62 (discussed by Vaney, Duquerroy, and Ray in current issue of Current Opinion in Virology) [39*]. Cryo-electron microscopy shows that immature flaviviruses at neutral pH have a ‘spiky’ appearance, while mature flaviviruses have a ‘smooth’ surface [40–43]. These appearances are altered depending on the oligomeric state of immature prM-E or mature M and E proteins. The hinge angle between DI and DII of E alone is very similar to the hinge angle of the prM-E heterodimer [31,37]. However, the difference in the angle between DI and DII in immature versus mature virus is 23°, which may determine the oligomeric state of E. Fitting the X-ray structure to cryoEM density shows that immature prM-E at neutral pH would prevent furin cleavage of prM due to steric hindrance [31**]. During egress, immature virions are transported through the trans Golgi Network where exposure to more acidic conditions causes a conformational change in prM-E, allowing furin access to the cleavage site [44].

**HIV Gag**

The *Retroviridae* genome consists of two copies of positive-sense RNA, which is reverse-transcribed early during infection. The resulting DNA is inserted into the host genome and used for production of viral RNA and proteins as infection progresses [45]. Retroviruses vary in the number of genes encoded, but they all have at least four open reading frames (gag, pol, and env), which may or may not be in the same reading frame. HIV is a retrovirus whose genome encodes nine (sometimes 10) genes. The *gag* gene is located at the 5′ end of all retrovirus genomes and is translated as a polyprotein precursor containing three distinct domains: matrix (MA), capsid (CA), and nucleocapsid (NC) (Figure 4a) [46]. The Gag precursor is associated with the inner surface of the cell’s plasma membrane, where it participates in forming the immature virion and selectively packaging viral RNA [45–47]. Cleavage of precursor protein causes mature CA proteins to organize into cone-like structures and ultimately leads to maturation of the virus [48**].

To discern differences between immature and mature virion structures, the NMR structure for the N-terminal portion of immature HIV-1 Gag, including MA and the N-terminal portion of the CA domain (CA^N), was determined (Figure 4b) [48**]. Structures of CA^N in the immature polyprotein and mature post cleavage form demonstrate that the N-terminus undergoes a disorder to order transition upon cleavage, resulting in the formation of a β-hairpin, while immature and mature MA structures do not show significant changes. Integral residues located in the CA^N hairpin region have previously been shown to prevent proper capsid core formation, leading to formation of noninfectious virus [49–51]. Therefore, cleavage triggers the formation of the β-hairpin in CA^N, and may be responsible for proper capsid core particle assembly.

**Conclusions**

Presented here are four recent atomic structures of viral polyprotein precursors. Several underlying mechanisms for viral gene regulation can be gleaned from these
Structures of viral precursor polyproteins Yost and Marcotrigiano

structures. The first involves accessibility of the polyprotein cleavage site to control the gene’s activity, function, and location at specific times post infection. Cleavage of these polyproteins is regulated by different methods. The alphavirus P2/3 cleavage site is solvent exposed but not readily accessible to the protease and the requirement for cleavage is still unknown. Access to the Poliovirus 3CD cleavage site has been postulated to be dependent on the intermolecular scaffolding thought to form between 3CD molecules [25**]. Cleavage of flavivirus pM-E is regulated by the low pH environment found late in viral egress, preventing immature exposure of the fusion loop. Secondly, cleavage events themselves can cause conformational changes in the newly formed termini of the mature, cleaved protein, permitting new functions post cleavage. The 3D polymerase activity is initiated by inserting the N-terminus of the mature protein into a hydrophobic pocket only after cleavage. The N-terminus of HIV CA undergoes the formation a β-hairpin, resulting in maturation of the core particle. Further structures regarding virus particle maturation and the involvement of precursor proteins not discussed here can be found for Chikungunya virus p62 [39*], poliovirus VP0 [52*], influenza HA0 [53*], and human parainfluenza virus 3 fusion protein [54*].

Viral polyproteins represent a largely unexplored structural aspect of the viral lifecycle, especially those involved in viral genome replication. Despite the prevalence of these proteins throughout virus families, very few viral precursor polyprotein structures are currently available. On the basis of the diversity of the four structures and cleavage site regulation presented here, it is likely that many more regulatory mechanisms are yet to be described. Because of their regulatory roles ranging from genome replication to virion maturation, viral polyproteins will be valuable tools for the design of uniquely targeted drugs.

Acknowledgements
This study was supported by the NIH through a training grant (T32AI007403-18) to S.A.Y. and an R01 (AI080659) and P50 (GM103368) to J.M.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Andres G, Alejo A, Salas J, Salas ML: African swine fever virus polyproteins pp250 and pp62 assemble into the core shell. J Virol 2002, 76:12473-12482.

2. Griffin DE: Alphaviruses. In Fields Virology, vol 1. Edited by David M, Knipe DMH. Lippincott Williams & Wilkins; 2007:1023-1067.

3. de Groot RJ, Hardy WR, Shirako Y, Strauss JH: Cleavage-site preferences of Sindbis virus polyproteins containing the non-structural proteinase. Evidence for temporal regulation of polyprotein processing in vivo. EMBO J 1990, 9:2631-2638.

4. Vasiljeva L, Merits A, Golubitskaya A, Sizemskaja V, Klääräinen L, Ahola T: Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J Biol Chem 2003, 278:41636-41645.

5. Shirako Y, Strauss JH: 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 1994, 68:1874-1885.

6. Lemm JA, Rümenapf T, Strauss EG, Strauss JH, Rice CM: Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. EMBO J 1994, 13:2925-2934.

7. Shin G, Yost SA, Miller MT, Eliod EJ, Grakoui A, Marcotrigiano J: Structural and functional insights into alphavirus polyprotein processing and pathogenesis. Proc Natl Acad Sci U S A 2012, 109:16534-16539.

This reference describes the structure of alphavirus nonstructural polyprotein precursor, P230(C15). The importance of regulation in polyprotein cleavage and relation to viral pathogenesis is discussed.

8. Shirako Y, Strauss JH: Cleavage between nsP1 and nsP2 initiates the processing pathway of Sindbis virus nonstructural polyprotein P123. Virology 1990, 177:54-64.

9. Hardy WR, Strauss JH: 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 1989, 63:4653-4664.

10. Frolov I, Agapov E, Hoffman TA, Pragai BM, Lippa M, Schlesinger S, Rice CM: Selection of RNA replicons capable of persistent noncytopathic replication in mammalian cells. J Virol 1999, 73:3854-3865.

11. Dryga SA, Dryga OA, Schlesinger S: 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 1997, 228:74-83.

12. Mayuri, Geders TW, Smith JL, Kuhn RJ: 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 2008, 82:7284-7297.

13. Perri S, Driver DA, Gardner JP, Sherrill S, Belli BA, Dubensky TW, Polo JM: Replicon vectors derived from Sindbis virus and Semliki forest virus that establish persistent replication in host cells. J Virol 2000, 74:9802-9807.

14. Petrikova O, Volkova E, Gorochakov R, Paessler S, Kinney RM, Frolov I: Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in mammalian cells. J Virol 2005, 79:7597-7608.

15. Hewlett MJ, Rose JK, Baltimore D: 5’-Terminal structure of poliovirus polyribosomal RNA is pU. Proc Natl Acad Sci U S A 1976, 73:327-330.

16. Nomoto A, Lee YF, Wimmer E: The 5’ end of poliovirus mRNA is not capped with m7G[5’]ppp[5’]p. Proc Natl Acad Sci U S A 1976, 73:375-380.

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

18. Kitamura N, Semler BL, Rothberg PG, Larsen GR, Adler CJ, Doner AJ, Emin Murphy, Hancea R, Lee JJ, van der Werf S et al.: Primary structure, gene organization and polypeptide expression of poliovirus RNA. Nature 1981, 291:547-553.

19. Racanelli VR, Baltimore D: Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. Proc Natl Acad Sci U S A 1981, 78:4887-4891.

20. Jore J, De Geus B, Jackson RJ, Pouwels PH, Enger-Valk BE: Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 in vitro. J Gen Virol 1988, 69(Pt 7):1627-1636.
39. Voss JE, Vaney MC, Duquerroy S, Vonhein C, Girard-Blanc C, Cruziel E, Thompson A, Bricogne G, Ray FA: Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. Nature 2010, 468:709-712.

The structures of Chikungunya virus P62-E1 heterodimer and mature E3-E2-E1 complexes are described in this reference. These provide an insight on the role of polyprotein cleavage and viral maturation.

40. Zhang Y, Corver J, Chipman PR, Zhang W, Pletnev SV, Sediak D, Baker TS, Strauss JH, Kuhn RJ, Rossmann MG: Structures of immature flavivirus particles. *EMBO* J 2003, 22:2604-2613.

41. Zhang Y, Kaufmann B, Chipman PR, Kuhn RJ, Rossmann MG: Structure of immature West Nile virus. *J Virol* 2007, 81:6141-6145.

42. Kuhn RJ, Zhang W, Rossmann MG, Pletnev SV, Corver J, Lenches E, Jones CT, Mukhopadhyay S, Chipman PR, Strauss EG et al.: Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell 2002, 108:717-725.

43. Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, Baker TS, Strauss JH, Rossmann MG, Kuhn RJ: Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nat Struct Biol* 2003, 10:907-912.

44. Yu IM, Zhang W, Holdaway HA, Li K, Kostyuchenko VA, Chipman PR, Kuhn RJ, Rossmann MG, Chen J: Structure of the immature dengue virus at low pH primes proteolytic maturation. Science 2008, 319:1834-1837.

45. Goff SP: Retroviridae: the retroviruses and their replication. In *Fields Virology*, vol 2. Edited by David M, Knipe PMH, Lippincott Williams & Wilkins; 2007:1101-1152.

46. Lu K, Heng X, Summers MF: Structural determinants and mechanism of HIV-1 genome packaging. *J Mol Biol* 2011, 410:609-633.

47. Vogt VM, Simon MN: Mass determination of rous sarcoma virus virosomes by scanning transmission electron microscopy. *J Virol* 1999, 73:7050-7055.

48. Tang C, Nidassa Y, Summers MF: Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nat Struct Biol* 2002, 9:537-543.

This reference describes the structure of a fragment of HIV-1 Gag polyprotein and suggests that formation of a β-hairpin at the N-terminus of the CA domain post cleavage is responsible for correct capsid core assembly.

49. von Schwedler UK, Stemmler TL, Klishko VY, Li S, Albertine KH, Davis DR, Sundquist WJ: Proteolytic refolding of the HIV-1 capsid protein amino-terminal fragment facilitates viral core assembly. *EMBO* J 1998, 17:1555-1568.

50. Fitzon T, Leschonsky B, Bieler K, Paulus C, Schroder J, Wolf H, Wagner R: Proline residues in the HIV-1 N22-terminal capsid domain: structure determinants for proper core assembly and subsequent steps of early replication. *Virology* 2000, 265:294-307.

51. Tang S, Murakami T, Agreste BE, Campbell S, Freed EO, Levin JG: Human immunodeficiency virus type 1 N-terminal capsid mutants that exhibit aberrant core morphology and are blocked in initiation of reverse transcription in infected cells. *J Virol* 2001, 75:9375-9366.

52. Basavappa R, Syed R, Fiore O, Incogne JP, Filman DJ, Hogle JM: Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: structure of the empty capsid assembly intermediate at 2.9Å resolution. *Protein Sci* 1994, 3:1651-1668. The structure of poliovirus VP0 precursor polyprotein (which yields mature capsid proteins VP4 and VP2) is described in the context of empty viral capsids in this reference.

53. Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ, Wiley DC: Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* 1998, 95:409-417.

The structure of influenza hemagglutinin precursor HA0 ectodomain is described in this reference, giving an insight into influenza pathogenicity and regulation of membrane fusion.

54. Yin HS, Paterson RG, Wen X, Lamb RA, Jardetzky TS: Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. *Proc Natl Acad Sci U S A* 2005, 102:9288-9293. Uncleaved paramyxovirus fusion (F) protein is described in this reference.