

## Structure of the NS3 helicase from Zika virus

Rinku Jain<sup>1</sup>, Javier Coloma<sup>1</sup>, Adolfo García-Sastre<sup>2–4</sup> & Aneel K Aggarwal<sup>1</sup>

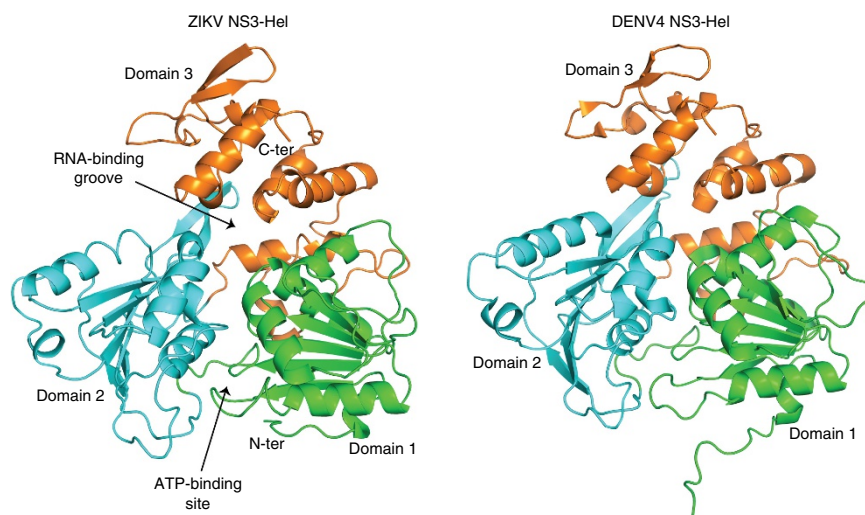
**Zika virus has emerged as a pathogen of major health concern. Here, we present a high-resolution (1.62-Å) crystal structure of the RNA helicase from the French Polynesia strain. The structure is similar to that of the RNA helicase from Dengue virus, with variability in the conformations of loops typically involved in binding ATP and RNA. We identify druggable ‘hotspots’ that are well suited for *in silico* and/or fragment-based high-throughput drug discovery.**

Zika virus (ZIKV) has emerged as a major health hazard over the past year. It has been linked to microcephaly in newborn infants and Guillan–Barré syndrome in adults<sup>1</sup>. ZIKV belongs to the same Flavivirus genus as other mosquito-borne human pathogens, such as dengue virus (DENV1–4), tick-borne encephalitis virus, Japanese encephalitis virus, yellow fever virus, West Nile virus, Murray Valley encephalitis virus and Kokobera virus (KOKV). The public-health emergency posed by ZIKV has invigorated efforts to both develop a vaccine and eradicate the *Aedes* mosquito vectors. In addition to these measures, it is equally important to develop antivirals through targeting of enzymatic activities central to the life cycle and survival of ZIKV. One such enzymatic activity is associated with the C-terminal region of the nonstructural protein NS3 in flaviviruses, namely an RNA helicase (NS3-Hel) involved in genome replication and RNA synthesis<sup>2,3</sup>. NS3-Hel belongs to the superfamily (SF2) helicases<sup>4</sup>, and its inactivation in DENV2 renders the virus incapable of replicating prop-

erly<sup>5</sup>. To help guide the discovery of antiviral compounds against ZIKV, we present a high-resolution (1.62-Å) crystal structure of NS3-Hel from the French Polynesia strain of the virus.

We expressed and purified ZIKV NS3-Hel (residues 171–617) of the H/PF/2013 strain as a soluble protein in *Escherichia coli* (Online Methods). We determined the structure by molecular replacement and refined it to a resolution of 1.62 Å, with  $R_{\text{work}}$  and  $R_{\text{free}}$  values of 16.1% and 19.3%, respectively (Supplementary Table 1). The refined model consists of one NS3-Hel molecule (residues 175–617), one pyrophosphate, six acetate ions and 552 solvent molecules.

ZIKV NS3-Hel is composed of three domains of roughly similar size (Fig. 1). Domains 1 and 2 (residues 182–327 and 328–480, respectively) comprise the tandem  $\alpha/\beta$  RecA-like folds characteristic of superfamily 1 (SF1) and SF2 helicases<sup>4</sup>. Domain 1 contains the classical motifs I (P loop or Walker A), Ia, II (or Walker B) and III, whereas domain 2 contains motifs IV, IVa, V and VI (Supplementary Fig. 1). These helicase motifs are typically associated with ATP binding and/or hydrolysis (motifs I, II and VI), and interdomain communication and RNA binding (motifs Ia, IV and V), and they line a cleft at the interface of domains 1 and 2 (Fig. 1 and Supplementary Fig. 1). As with the closely related DENV4 helicase (Fig. 1), for which there are structures in different ligand-bound states<sup>6</sup>, ATP is expected to bind the ‘bottom’ side of the cleft between domains 1 and 2, whereas the RNA would be accommodated in a tunnel that separates domains 1 and 2 from domain 3. Notably, we observed electron density of approximately the size of a pyrophosphate near the P loop in the



**Figure 1** Structure comparison of ZIKV NS3-Hel with DENV4 NS3 helicase. Overall fold of ZIKV NS3-Hel residues 171–617 (left) and DENV4 helicase<sup>6</sup> (PDB 2JLQ, right). RecA-like domains 1 and 2 are colored green and cyan, and domain 3 is colored orange. The ATP-binding pocket and RNA-binding groove are indicated. Ter, terminus.

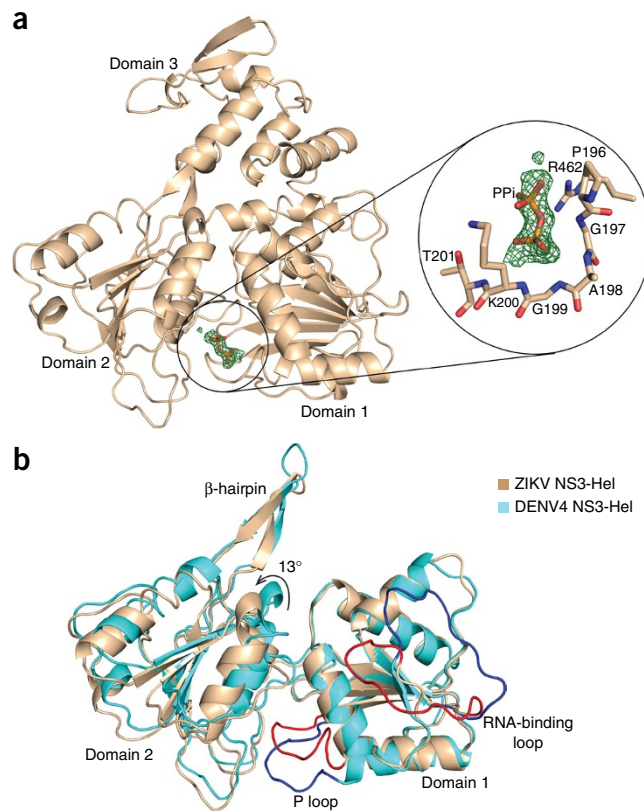
<sup>1</sup>Department of Structural & Chemical Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. <sup>2</sup>Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. <sup>3</sup>Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. <sup>4</sup>Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. Correspondence should be addressed to A.K.A. (aneel.aggarwal@mssm.edu)

Received 1 May; accepted 14 June; published online 11 July 2016; doi:10.1038/nsmb.3258

**Figure 2** Features of the ZIKV NS3-Hel structure. (a) Difference electron density (contoured at  $3\sigma$ ) near the P loop. A magnified view is shown on the right. Residues of the P loop in the vicinity of the density are shown as tan sticks; modeled pyrophosphate (PPI) is shown in orange sticks. (b) Superimposition of ZIKV NS3-Hel with DENV4 NS3-Hel via domain 1. Domain 3 has been omitted for clarity. ZIKV NS3-Hel is shown in tan, with the P loop and RNA-binding loop in red. DENV4 helicase is shown in cyan, with the P loop and RNA-binding loop in blue.

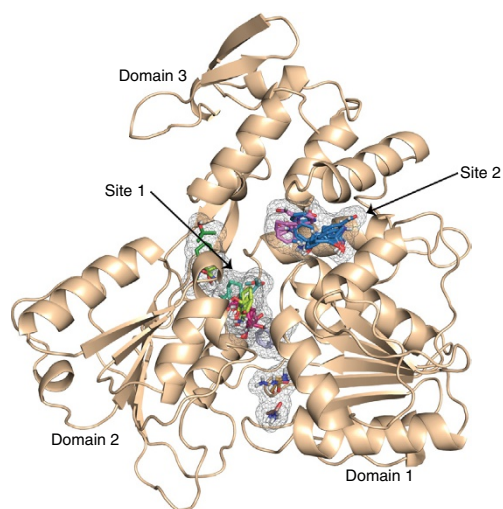
ATP-binding cleft (**Fig. 2a**). Guided by its shape and size, we modeled this extra density as both glycerol (present in the purification buffer and cryoprotectant) and pyrophosphate, but we obtained the best fit with a pyrophosphate (**Supplementary Fig. 2a**), which may have been acquired during expression of the protein in *E. coli*<sup>7</sup>. The putative pyrophosphate is in a similar position as the  $\beta$ - and  $\gamma$ -phosphates of AMPNP in the DENV4 NS3-Hel-AMPNP structure<sup>6</sup> (**Supplementary Fig. 2b**) and has the capacity to form hydrogen bonds with the main chain and side chain atoms of the conserved P loop (**Fig. 2a**). It is not uncommon for an NTPase to copurify with a pyrophosphate, for example, as in structures of KOKV NS3-Hel<sup>8</sup> and the *Helicobacter pylori* GTP cyclohydrolase<sup>9</sup>. A  $\beta$ -hairpin extends from domain 2 and interacts with domain 3 and has been proposed to act as a 'wedge' facilitating separation of the RNA strands of double-stranded RNA in the unwinding reaction<sup>6</sup> (**Fig. 1**). Interestingly, the  $\beta$ -hairpin in our structure (residues 431–444) is slightly smaller than that in DENV4 helicase (430–445) (**Fig. 2b**) and may alter the kinetics of the RNA unwinding reaction. Domain 3 is predominantly  $\alpha$ -helical. In addition to contacting RNA, this domain has also been implicated in interactions with the RNA-dependent RNA polymerase NS5 in other flaviviruses<sup>10</sup>.

Overall, all three domains superimpose very well on equivalent domains in other flaviviruses (**Supplementary Fig. 3a**). ZIKV domains 1, 2 and 3 superimpose with r.m.s. deviations of 0.60 Å, 0.31 Å and 0.40 Å, respectively, on the equivalent domains of DENV4. However, there are some qualitative differences in the electrostatic surfaces of ZIKV NS3-Hel compared with other flaviviruses, including an RNA-binding groove that is somewhat less basic (**Supplementary Fig. 4**). An additional comparison between ZIKV and the other flavivirus NS3-Hel structures highlights the flexibility in the relative orientation of domains 1 and 2 (**Fig. 2b**) and in the conformation of loops implicated in binding ATP and RNA (**Supplementary Fig. 3b,c**). Compared with the apo-DENV4



helicase, domain 2 of the ZIKV helicase is rotated away from domain 1 by  $\sim 13^\circ$ , thus leading to a wider ATP-binding cleft (**Fig. 2b**). Interestingly, apo-West Nile virus, yellow fever virus, KOKV and Japanese encephalitis virus also adopt the ZIKV-like partially open conformation<sup>11,12</sup>, whereas DENV2 and Murray Valley encephalitis virus assume the DENV4-like closed conformation<sup>13,14</sup>. The P loop (residues 196–203) and the RNA-binding loop (residues 244–255) in our structure adopt conformations similar to that when DENV4 NS3-Hel binds ATP and single-stranded RNA, rather than that of the apoenzyme<sup>6</sup>. The P loops in other flavivirus apo structures also assume conformations similar to that in ZIKV (**Supplementary Fig. 3b**), whereas their RNA-binding loops are partially disordered (**Supplementary Fig. 3c**). Altogether, the P loop and the RNA-binding loop are the most flexible segments in NS3-Hel structures and intermittently sample conformations of the ligand-bound state<sup>15</sup>.

Is ZIKV NS3-Hel a druggable target? We used the program FTMap to derive druggable hotspots on the ZIKV helicase surface<sup>16</sup>. FTMap maps clusters of small organic molecules on a protein surface as putative drug- or ligand-binding sites. The two most prominent sites on ZIKV NS3-Hel are between domains 1 and 2 (site 1) and at the junction of domains 1 and 3 (site 2) (**Fig. 3**). Site 1 is close to the extra 'pyrophosphate' density in our structure, whereas site 2 is within the RNA-binding groove, close to the putative 3' end of bound RNA. Both pockets possess polar and hydrophobic characteristics and appear to be well suited for *in silico* high-throughput screening with drug-like molecules and/or fragment-based screening<sup>17,18</sup>.



**Figure 3** Hotspots predicted by FTMap for the ZIKV NS3-Hel domain. 11 hotspots predicted by FTMap are shown in gray mesh, and the cluster of probes at each spot is shown in sticks of various colors. These hotspots coalesce into two broad and contiguous sites that map between domains 1 and 2 (site 1) and at the junction of domains 1 and 3 (site 2).

Additionally, the proximity of the two pockets may enable the design of inhibitors that can span both sites. Several inhibitors have been reported for DENV helicase<sup>19</sup>. It will be interesting to determine whether these inhibitors bind the ZIKV helicase and mitigate the ability of the virus to replicate.

In conclusion, we report a high-resolution crystal structure of the ZIKV RNA helicase that should aid in the discovery of antiviral compounds against this pernicious emerging pathogen. We note that a study reporting the structure of ZIKV NS3-Hel at 1.8-Å resolution has very recently been published<sup>20</sup>, showing a similar structure to that described here but without the pyrophosphate, thus providing independent support for our own findings.

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

**Accession codes.** Coordinates and structure factors have been deposited in the Protein Data Bank under accession code [5JRZ](#).

*Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).*

## ACKNOWLEDGMENTS

We thank K. Rajashankar and the staff at NECAT-24ID (APS) for facilitating data collection. These studies were partly funded by a supplement to the NIAID NIH grant U19AI118610 (to A.G.-S.).

## AUTHOR CONTRIBUTIONS

R.J., A.K.A. and A.G.-S. designed the experiments; R.J. performed protein expression, purification and crystallization. R.J. and J.C. performed crystallography. A.K.A., R.J. and J.C. wrote the paper. All authors reviewed the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Petersen, L.R., Jamieson, D.J., Powers, A.M. & Honein, M.A. *N. Engl. J. Med.* **374**, 1552–1563 (2016).
- Frick, D.N. & Lam, A.M. *Curr. Pharm. Des.* **12**, 1315–1338 (2006).
- Luo, D., Vasudevan, S.G. & Lescar, J. *Antiviral Res.* **118**, 148–158 (2015).
- Singleton, M.R., Dillingham, M.S. & Wigley, D.B. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
- Matusan, A.E., Pryor, M.J., Davidson, A.D. & Wright, P.J. *J. Virol.* **75**, 9633–9643 (2001).
- Luo, D. *et al. EMBO J.* **27**, 3209–3219 (2008).
- Kukko, E. & Heinonen, J. *Eur. J. Biochem.* **127**, 347–349 (1982).
- Speroni, S. *et al. Proteins* **70**, 1120–1123 (2008).
- Yadav, S. & Karthikeyan, S. *J. Struct. Biol.* **192**, 100–115 (2015).
- Tay, M.Y. *et al. J. Biol. Chem.* **290**, 2379–2394 (2015).
- Wu, J., Bera, A.K., Kuhn, R.J. & Smith, J.L. *J. Virol.* **79**, 10268–10277 (2005).
- Mastrangelo, E. *et al. J. Mol. Biol.* **372**, 444–455 (2007).
- Mancini, E.J. *et al. Protein Sci.* **16**, 2294–2300 (2007).
- Xu, T. *et al. J. Virol.* **79**, 10278–10288 (2005).
- Boehr, D.D., Nussinov, R. & Wright, P.E. *Nat. Chem. Biol.* **5**, 789–796 (2009).
- Kozakov, D. *et al. Nat. Protoc.* **10**, 733–755 (2015).
- Rees, D.C., Congreve, M., Murray, C.W. & Carr, R. *Nat. Rev. Drug Discov.* **3**, 660–672 (2004).
- Shoichet, B.K. *Nature* **432**, 862–865 (2004).
- Sweeney, N.L. *et al. ACS Infect. Dis.* **1**, 140–148 (2015).
- Tian, H. *et al. Protein Cell* **7**, 450–454 (2016).

## ONLINE METHODS

**Purification, crystallization and structure determination.** ZIKV NS3-Hel (171–617) from the H/PF/2013 strain was expressed in and purified from *E. coli* strain B834(DE3) with an N-terminal His<sub>6</sub>-SUMO tag. Cell pellets containing the recombinant protein were resuspended in buffer containing 50% B-PER (Thermo Scientific), 25 mM Tris, pH 8.0, 500 mM NaCl, 10% sucrose and 5 mM 2-mercaptoethanol (BME). Cells were lysed by sonication, and the filtered lysate was loaded on a 5-mL Ni-NTA column (Qiagen). Protein bound to the Ni-NTA column was eluted with buffer containing 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 5 mM BME and 300 mM imidazole. Eluted protein was dialyzed into buffer containing 50 mM Tris, pH 8.0, 150 mM NaCl, 1 mM DTT and 0.01% IGEPAL CA-630. The His<sub>6</sub>-SUMO tag was cleaved with Ulp protease, and the protein was reloaded on the Ni-NTA column to remove the cleaved His<sub>6</sub>-SUMO tag and any uncleaved protein. The cleaved protein was purified further by ion-exchange chromatography on an anion-exchange column or by size-exclusion chromatography on an SD75 column (GE Healthcare Life Sciences). Before crystallization, the protein was concentrated to 10 mg/ml in buffer containing 25 mM Tris, pH 8.0, 100 mM NaCl and 2 mM TCEP.

Initial screens were set up with an Oryx robot at 4 °C. Thin plate-like crystals were obtained in conditions containing PEG 8000, MOPS buffer and 200 mM magnesium acetate from the Protein Complex suite (Qiagen). Crystals were optimized by varying the concentration of PEG 8000 and microseeding. Stock solutions for crystal optimization consisted of 50% (w/v) PEG 8000 (Sigma-Aldrich), 0.1 M MOPS (NaOH), pH 7.0, and 2 M magnesium acetate (Fluka). Crystals used for data collection grew from drops containing 4–8% PEG 8000, 0.1 M MOPS, pH 7.0, and 100 mM magnesium acetate.

For data collection, crystals were cryoprotected by quick dipping in mother liquor containing 30% glycerol alone, or in a mixture of 9% sucrose, 2% glucose, 8% glycerol and 8% ethylene glycol, and flash-cooled in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (beamline 24-ID-E) under cryogenic conditions at a wavelength of 0.97918 Å, and indexed with HKL2000 (ref. 21).

The ZIKV NS3-Hel structure was solved by molecular replacement by using the Auto-Rickshaw web server (<http://webapps.embl-hamburg.de/cgi-bin/Auto-Rick/arinitAR1.cgi/>)<sup>22</sup>. The model obtained from the Auto-Rickshaw pipeline was improved by iterative manual building and

refinement with Coot<sup>23</sup> and Phenix<sup>24</sup>, respectively. After the protein chain was built, difference electron density (greater than 3σ) was visible near the P loop (Fig. 2a and Supplementary Fig. 2), thus indicating the presence of a ligand. Automated ligand identification with Phenix suggested the presence of bulky amino acids (arginine, glutamate, lysine, tryptophan and methionine), nucleotides with the phosphate moiety in the difference density, pyrophosphate or glycerol. Visual inspection of the fit of these putative ligands into the difference density, their residual density after refinement with Phenix and their interactions with the neighboring protein residues suggested that the density was best described as a pyrophosphate. The pyrophosphate moiety is engaged in putative hydrogen-bonding interactions with the P-loop residues Gly197, Gly 199 and Lys200, as well as with the side chain of Arg462.

The final model was refined to 1.62 Å and has good stereochemistry, with 98% of the residues in the most favored regions of the Ramachandran plot, and 0.2% in the disallowed regions. Figures were prepared with PyMOL (<https://www.pymol.org/>). Qualitative surface electrostatic potential for the Flavivirus NS3-helicase domains (Supplementary Fig. 4) were computed with PyMOL. The potential range was set the same for all the structures; positive potential is shown in blue, and negative potential is shown in red.

**Mapping druggable hotspots with FTmap.** The FTmap server (<http://ftmap.bu.edu/>) was used to identify hotspots of ZIKV NS3-Hel and to determine its druggability. The hotspots are cavities on the protein surface that represent potential ligand-binding sites. The program was run with default parameters. 11 hotspots were predicted, and the numbers of probe clusters at these sites were 28, 26, 11, 8, 5, 4, 4, 3, 2, 2 and 2 from the highest- to the lowest-ranked hotspot. These hotspots coalesced into two broad and contiguous sites (sites 1 and 2; Fig. 3) in the vicinity of the RNA-binding groove and ATP-binding cleft.

21. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307–326 (1997).

22. Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. & Tucker, P.A. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 449–457 (2005).

23. Emsley, P. & Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

24. Adams, P.D. *et al. Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).