Supporting Information

Gu et al. 10.1073/pnas.0913380107

SI Text

SI Materials and Methods. ***Recombinant protein production and purification.*** The hepatitis C virus (HCV) NS3 helicase domain, NS3h/NS3(188-625), was PCR amplified from Con1 HCV replicon (genotype 1b) cDNA (1). The amplified fragment was inserted into the pET-SUMO vector (Invitrogen) and transformed into Rosetta (DE3) cells (Novagen). One colony was inoculated into 800 mL LB culture and grown overnight. The overnight culture was diluted 10 times into fresh LB. The cells were grown for about 150 min at 37 °C to reach $OD_{600} \sim 1.0$. The cells were chilled on ice and induced for protein expression with 1 mM IPTG. After shaking at 25 °C for 4 h, the cells were harvested by centrifugation. Cell pellets were resuspended in buffer (20% sucrose and 50 mM Tris pH 8.5), flash frozen in liquid nitrogen, and stored at −80 °C.

To purify NS3h, the cells were fully lysed by sonication in the presence of 500 mM NaCl, 1 mM PMSF, 0.1% IGEPAL (Octylphenyl-polyethylene glycol), 10 mM imidazole, 1 mM β-ME (β-mercaptoethanol), 50 μg/mL lysozyme, 5 μg/mL DNase I, and 2.5 μg/mL RNase A. Insoluble materials were removed by centrifugation. The supernatant was passed through an Econo-Pac column (Bio-Rad) prepacked with Ni Sepharose 6 Fast Flow (GE Healthcare). Unbound and nonspecifically bound materials were washed off the column by 25 column volumes of washing buffer (500 mM NaCl, 20 mM imidazole, 20 mM Tris pH 8.5, and 1 mM β-ME). The bound protein, His_6-Smt3-NS3h, was eluted with elution buffer (500 mM NaCl, 240 mM imidazole, 20 mM Tris pH 8.5, and 1 mM β-ME). The protein was sequentially purified with ion-exchange and gel-filtration columns (HiTrap Q HP 5 mL and HiPrep 26/60 Sephacryl S-200 HR; GE Healthcare). His_6-Smt3 was cleaved by Ulp1 protease (2) and separated from NS3h by gel filtration. Pure NS3h fractions were collected, and the protein concentrated to 20 mg/mL in buffer containing 50 mM NaCl, 20 mM Tris (pH 8.5), and 2 mM DTT. Protein aliquots were flash frozen in liquid nitrogen and stored at −80 °C.

Crystallization. The NS3h-ssDNA binary complexes were reconstituted by mixing 0.2 mM NS3h with 0.4 mM synthetic ssDNA (dA_6 or dT_6;) Integrated DNA Technologies). NS3h-ADP · AlF_4^--ssDNA ternary complexes were reconstituted by mixing 0.2 mM NS3h with 1 mM ADP, 2.5 mM $MgCl_2$, 10 mM NaF, 1.5 mM $AlCl_3$, and 0.4 mM ssDNA (dT_6; Integrated DNA Technologies). The mixture was incubated at room temperature for about 2 h. The NS3h-ADP · BeF_3-ssDNA (dT_6) ternary complexes were reconstituted in the same way, except that 0.4 mM $MnCl_2$ and 1.5 mM $BeCl_2$ were used in place of $MgCl_2$ and $AlCl_3$, respectively. In addition, dT_{12} (Integrated DNA Technologies) was used in the place of dT_6 for the reconstitution of the NS3h-ADP · BeF_3-ssDNA(dT_{12}) ternary complexes.

Crystallization conditions were identified through diverse reagent screens (Hampton Research) followed by successive condition optimization attempts. Additive screens (Hampton Research) were carried out to search for conditions supporting better crystal growth. The NS3h-ssDNA(dA_6 or dT_6) complexes were crystallized at room temperature, in 0.1 M MES [2-(N-morpholino)ethanesulfonic acid] pH 6.5, 22% (vol/vol) glycerol, and 10% (wt/vol) PEG (polyethylene glycol) 8000. The NS3h-ADP · AlF_4^--ssDNA(dT_6) complex was crystallized at 4 °C, in 10 mM NaF, 50 mM trimethylamine *n*-oxide, and 20% (wt/vol) PEG MME (monomethyl ether) 550. The NS3h-ADP · BeF_3-ssDNA(dT_6) complex was crystallized at room temperature, in 10 mM NaF, 0.1 M Bis-Tris [Bis(2-hydroxyethyl)amino-tris-(hydroxymethyl)methane] pH 5.5, 25 mM trimethylamine *n*-oxide, and 12% (wt/vol) PEG 400. The NS3h-ADP · BeF_3-ssDNA(dT_{12}) complex was crystallized at room temperature, in 10 mM NaF, 2.5% (wt/vol) PEG 4000, and 15% (wt/vol) PEG MME 550. The same type of crystals for the NS3h-ADP · BeF_3-ssDNA (dT_{12}) complex can also be obtained in a different condition, 10 mM Bistris pH 6.5, 10% (vol/vol) isopropanol, and 8.5% (wt/vol) PEG MME 5000, indicating that the crystallization solutions do not affect the overall structures of the complexes. Crystals grew to full size within 5 days.

Crystals of the NS3h-ssDNA binary complexes were clustered thin plates, and those of the ternary complexes were clustered needles. Single crystals for diffraction were carefully isolated from the clusters. The crystals of the ternary complexes were soaked in the presence of ADP, magnesium/manganese, and BeF_3/AlF_4^-. Before flash-freezing in liquid nitrogen, crystals were cryoprotected in their own crystallization solutions plus 5–10% (vol/vol) glycerol.

Data collection and structure determination. Crystals were screened by using Rigaku/MSC MicroMax-007HF (the Rockefeller University Structural Biology Resource Center), which was made possible by the National Center for Research Resources of the National Institutes of Health. Datasets for all types of crystals were collected with a synchrotron x-ray source at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY). The datasets were processed with DENZO, SCALEPACK, and CCP4 (3, 4). Molecular replacement experiments were carried out with MOLREP and PHASER (CCP4 program suite). The structure of NS3h alone [Protein Data Bank (PDB) ID: 1HEI] was used as a searching model in the initial molecular replacement for the NS3h-ssDNA complex. Before structural refinement, the sequence of the searching model was changed to that of Con1 genotype 1b. To find the helicases in the NS3h-ADP · BeF_3-ssDNA and NS3h-ADP · AlF_4^--ssDNA ternary complexes, domains 1 and 3 of the NS3h-ssDNA complex were used as a searching model, and domain 2 (without motif V) was docked into the $2F_o - F_c$ electron density calculated in CNS (5).

Rigid body refinement and simulated annealing (5) were carried out sequentially. Model building was done with the O program (6). The crystals of NS3h-ssDNA and NS3h-ADP · AlF_4^--ssDNA contain two superimposable complexes in their asymmetric units (rmsd ∼0.36 and ∼0.17 Å, respectively, for the Cα atoms of the aligned protomers). NCS (noncrystallographic symmetry) restraints were applied to the NS3h-ADP · AlF_4^--ssDNA complexes in the initial refinement. No NCS restraints were applied to the NS3h-ssDNA complexes. The poly(dA) chain in the NS3h-ssDNA(dA_6) structure was changed to poly(dT) to generate an initial model of NS3h-ssDNA(dT_6) for refinement.

The crystals of the NS3h-ADP · BeF_3-ssDNA(dT_6) ternary complex in a $P6_1$ space group contain a single complex in the asymmetric unit. The structure of this complex shows that the 3′ end of one DNA strand binds to domain 2 and the 5′ end of the other DNA strand binds to domain 1. To visualize one continuous DNA strand in the substrate-binding groove, we solved the structure of NS3h-ADP · BeF_3-ssDNA(dT_{12}). Six ternary complexes were refined in a P1 space group with NCS restraints in the protein chains. Each ssDNA strand is shared by three NS3h molecules. We built four alternate conformations for the bound

DNA strands (different regions of ssDNA bound to NS3h) before the last round of structural refinement to better interpret the electron density. As a result, two types of DNA binding were observed in the substrate-binding groove. One type is like that in the NS3h-ADP · BeF_3-ssDNA(dT_6) complex, and the other is binding of a continuous DNA strand. No alternate conformations were observed in protein chains. NS3h in all six NS3h-ADP · BeF_3-ssDNA(dT_{12}) complexes is structurally superimposable to that in NS3h-ADP · BeF_3-ssDNA(dT_6) (rmsd ~0.25 Å for the Cα atoms between the aligned protein chains). The structure of NS3h in complex with ADP · BeF_3 and a continuous DNA strand was the source of the ssDNA structure and binding model presented in Figs. 1*B*, 4, and 5*B*.

Fluorescence anisotropy titration. Experiments were carried out on a FluoroLog-3 spectrofluorometer (HORIBA Scientific). NS3h and 5′ fluorescein-labeled poly(dT_8) were mixed in a series of tubes in which the amount of DNA was kept constant at 20 nM while the amount of NS3h was incrementally increased from 0 to 10.24 μM. The protein and DNA were incubated in a buffer of 25 mM NaCl, 20 mM Tris pH 8.5, 1 mM DTT, 1 mM ADP, 0.8 mM $MnCl_2$, and 10 mM NaF, at room temperature for 30 min. $AlCl_3$/$BeCl_2$ was added at a concentration of 1.5 mM when necessary to reconstitute the ATP mimics. To determine the apparent dissociation constants (K_d), measured fluorescence anisotropy (A) was plotted as a function of total NS3h concentration (N_t), and the data were fit to the following quadratic equation:

$$A = (A_{\max} - A_0) \times \frac{K_d + D_t + N_t - \sqrt{(K_d + D_t + N_t)^2 - 4 \times D_t \times N_t}}{2 \times D_t} + A_0, \quad \textbf{[S1]}$$

where D_t is the concentration of total DNA, $A_{\max}$ is the maximum anisotropy, and A_0 is anisotropy in the absence of NS3h.

1. Blight KJ, Kolykhalov AA, Rice CM (2000) Efficient initiation of HCV RNA replication in cell culture. *Science* 290:1972–1974.
2. Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* 5:865–876.
3. CCP4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
4. Otwinowski Z, Minor W (1997) Processing of x-ray diffraction data collected in oscillation mode. *Method Enzymol* 276:307–326.
5. Brunger AT, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
6. Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47(Pt 2):110–119.

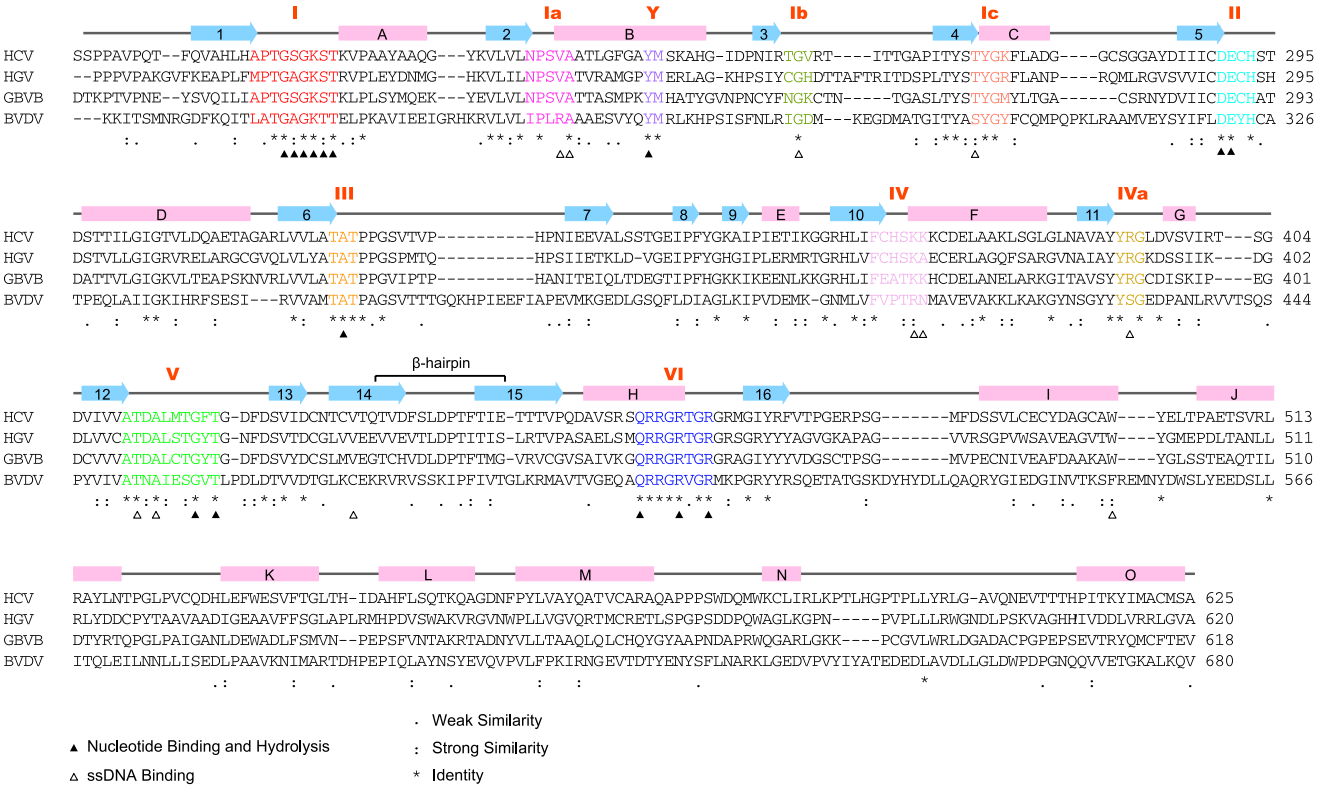


Fig. S1. Sequence alignment. The aligned sequences belong to the viral helicases encoded by HCV (Con1 genotype 1b), hepatitis G virus (HGV), hepatitis GB virus B (GBVB), and bovine viral diarrhea virus 1 (BVDV). The helicase motifs are labeled and color coded as in Fig. 1*B* and 1*C*. Secondary structures are noted above the HCV sequence. The pink rectangles are α-helices, and the blue arrows are β-strands. The asterisks and dots below the sequences represent identity and similarity, respectively. The black and white triangles point to the protein residues involved in nucleotide-binding pocket formation and ssDNA binding, respectively.

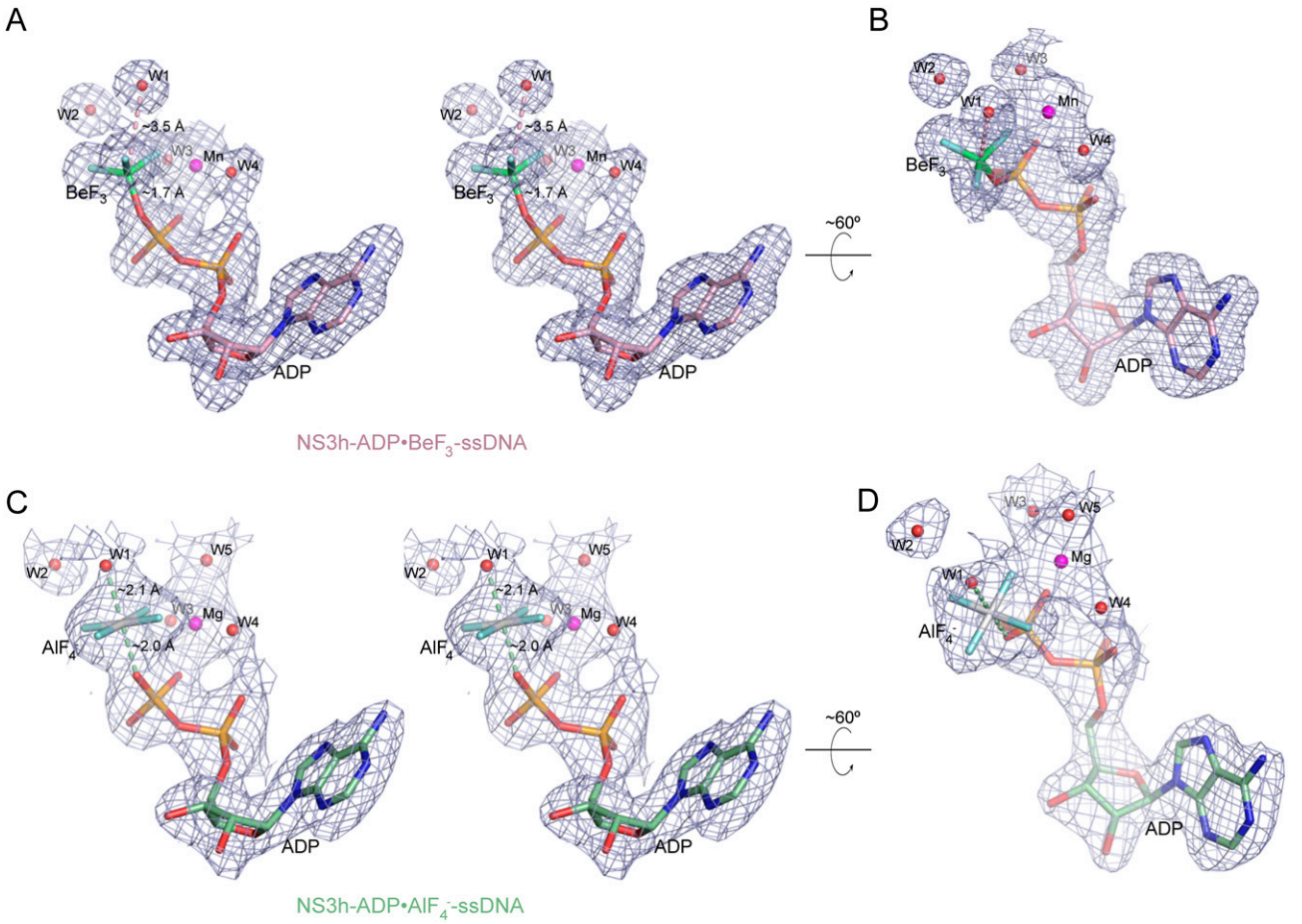


Fig. S2. Electron density covering nucleotides. Simulated annealing composite omit electron density calculated in CNS is contoured at 1.0σ around the ground-state and transition-state mimics of ATP. (*A*) Stereo view of the ground-state mimic of ATP in NS3h-ADP · BeF_3-ssDNA. (*B*) An alternate view of (*A*). (*C*) Stereo view of the transition-state mimic of ATP in NS3h-ADP · AlF_4^--ssDNA. (*D*) An alternate view of (*C*). The electron density is shown as blue mesh. The ATP mimics are shown with sticks. The water molecules and metal ions are modeled with spheres. The atoms are color coded according to elements. The dashed lines represent the distances between atoms.

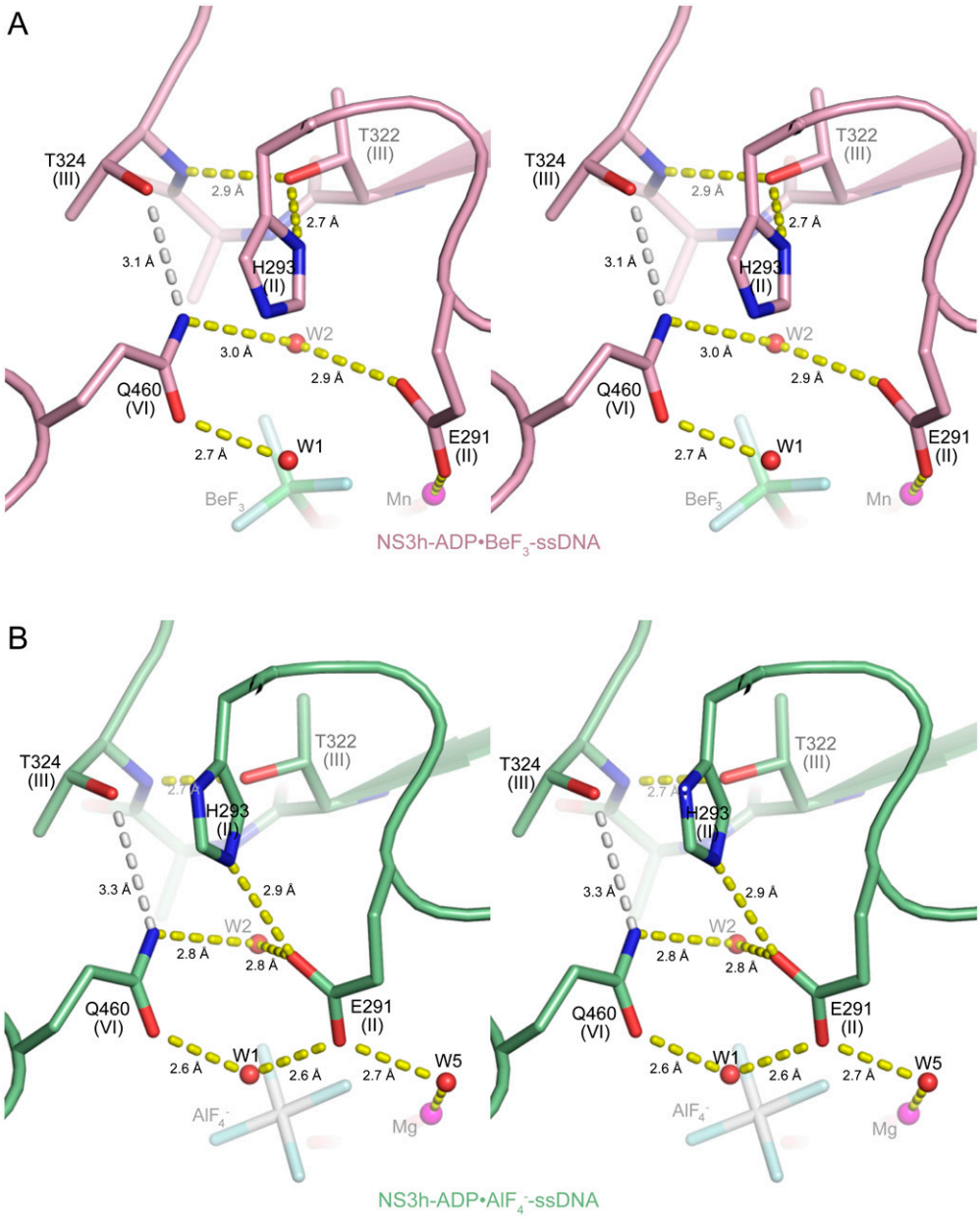


Fig. S3. Stereo view of H293 and adjacent residues. The (*A*) and (*B*) panels are prepared and arranged as those in Fig. 2. The views focus on H293 and its adjacent residues. The side chains are labeled and the motif codes are attached. The dashed lines represent distances between atoms. The yellow dashed lines represent the hydrogen-bond interactions. The distances are labeled.

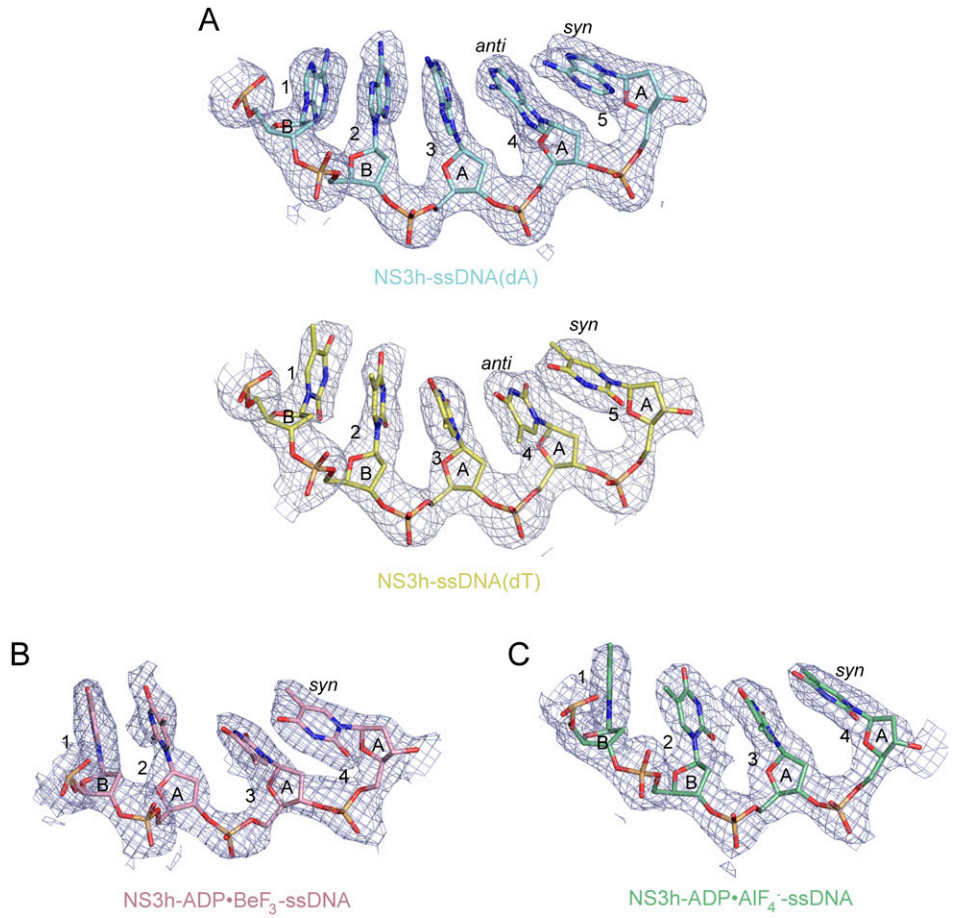


Fig. S4. Electron density covering DNA strands. Simulated annealing composite omit electron density calculated in CNS is contoured at 0.8σ around the DNA strands in (*A*) NS3h-ssDNA (dA_6 and dT_6), (*B*) NS3h-ADP · BeF_3-ssDNA(dT_{12}), and (*C*) NS3h-ADP · AlF_4^--ssDNA(dA_6). The panels are prepared as those in Fig. 5. Only the DNA residues captured between the "bookend" residues are shown. The electron density is shown as blue mesh. The protein components are omitted. The DNA residues are numbered. The sugar groups labeled with A are in C3′-endo. The ones labeled with B are in C2′-endo. The DNA bases are either "syn" or "anti".

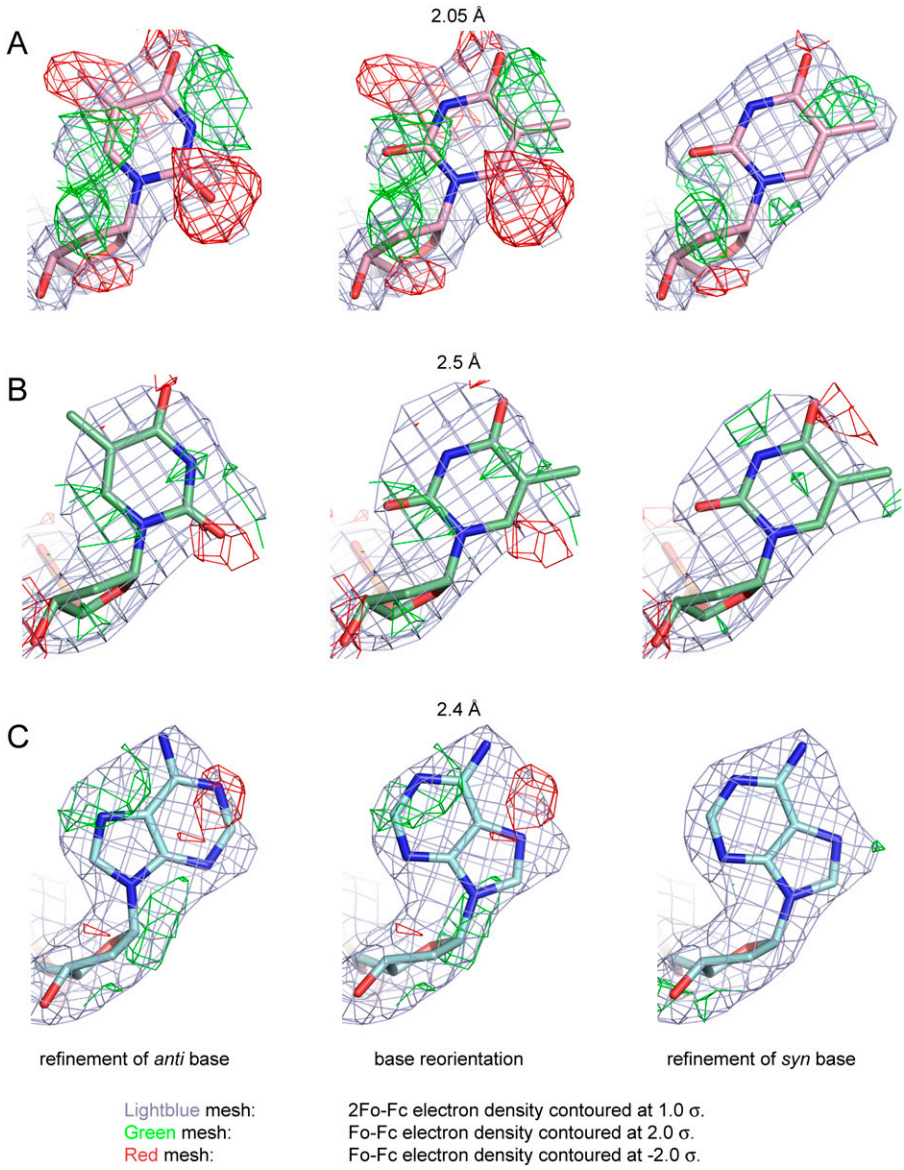


Fig. S5. Base orientations. The DNA bases stacking to W501 were initially refined in the anti orientations (*Left*). The bases were changed to the syn orientations to fit better into the $2F_o - F_c$ and $F_o - F_c$ electron density (*Center*). The bases were refined in the syn orientations, and the new $2F_o - F_c$ and $F_o - F_c$ electron density maps are shown in the right panels. This test was done for the (*A*) 2.05-Å-refined NS3h-ADP · BeF_3-ssDNA(dT_6) complex, (*B*) 2.5-Å-refined NS3h-ADP · AlF_4^--ssDNA (dT_6) complex, and (*C*) 2.4-Å-refined NS3h-ssDNA(dA_6) complex.

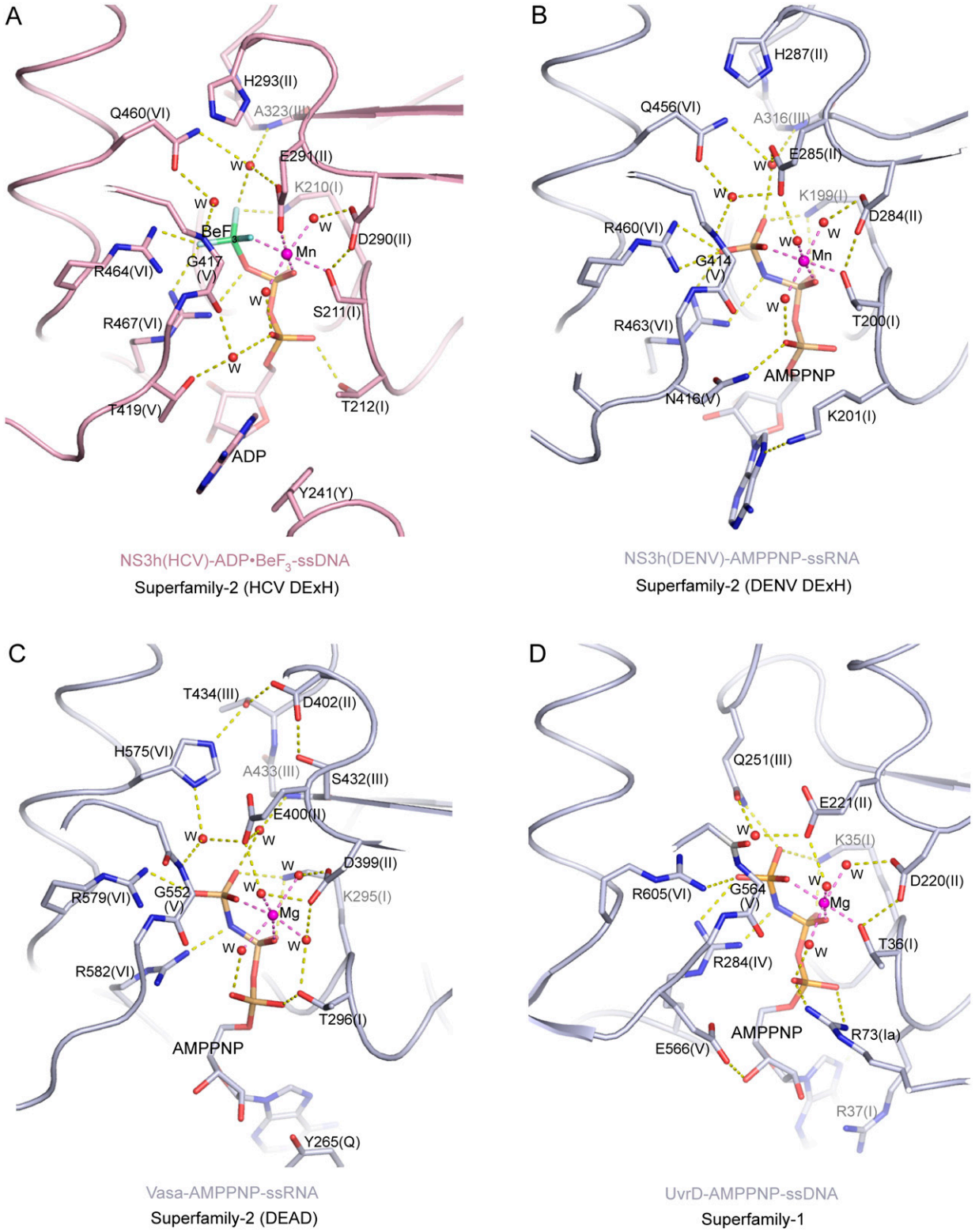


Fig. S6. Nucleotide-binding sites of different helicases. The panels are prepared as those in Fig. 2. Views of ground-state nucleotide binding are presented for the (*A*) HCV NS3h helicase, (*B*) DENV (dengue virus) NS3h (PDB ID: 2JLV), which represents flavivirus-encoded helicases, (*C*) Vasa DEAD-box helicase (PDB ID: 2DB3), which is required for germ cell development and function, and (*D*) UvrD helicase (PDB ID: 2IS4), which plays roles in DNA repair.

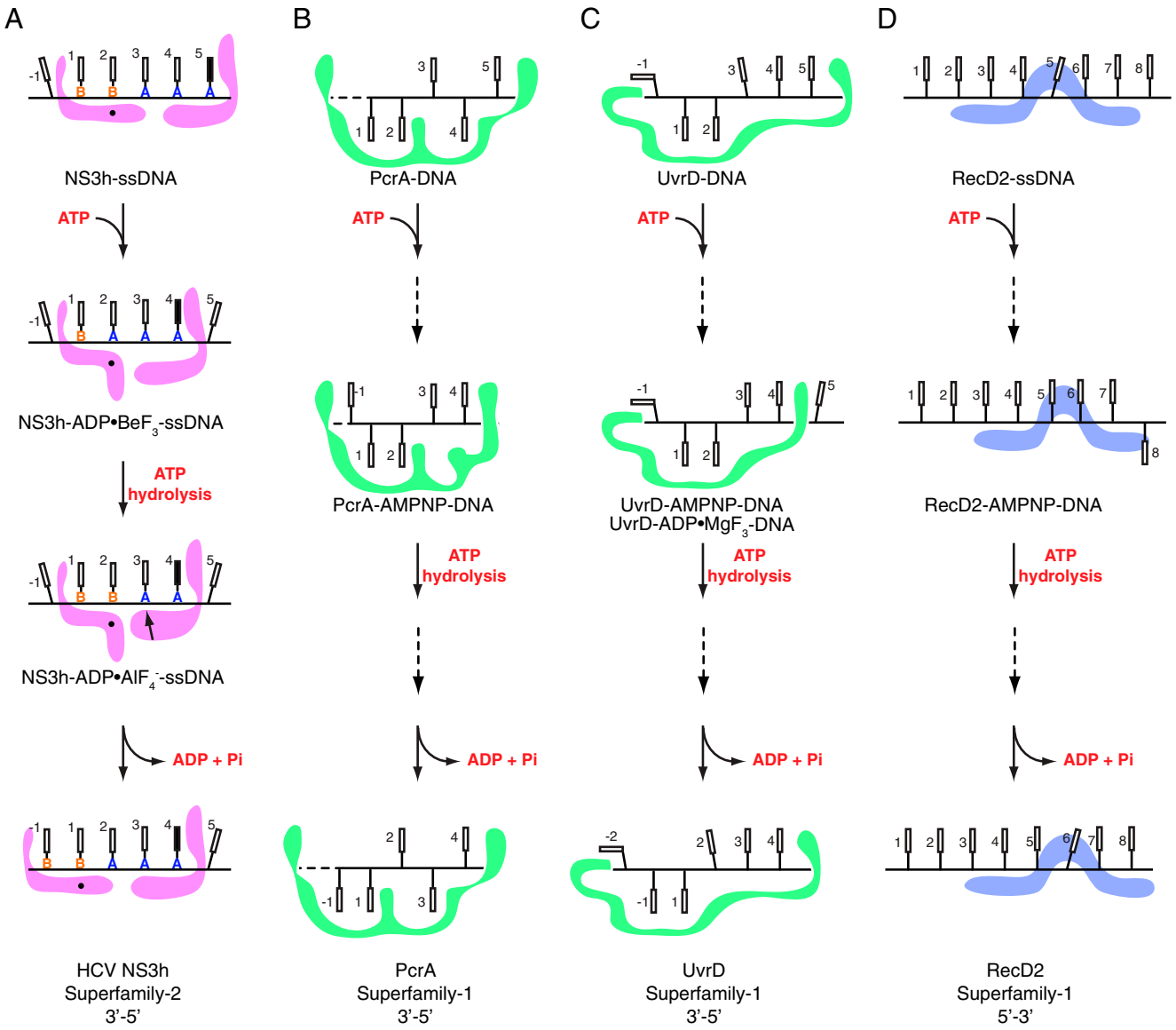
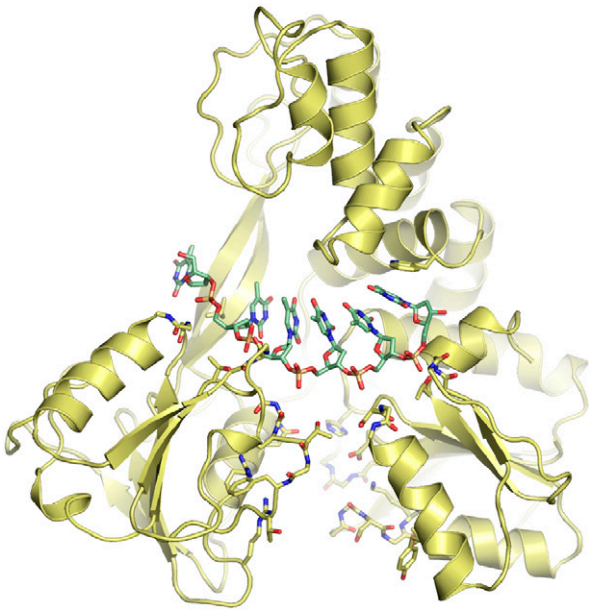


Fig. S7. Comparison of translocation modes for SF1 and SF2 helicases. The panels highlight the features of helicase translocation for the (*A*) HCV SF2 3′-5′ DNA/RNA helicase, (*B*) PcrA SF1 3′-5′ DNA helicase, (*C*) UvrD SF1 3′-5′ DNA helicase, and (*D*) RecD2 SF1 5′-3′ DNA helicase. The DNA backbones are simplified as lines and the bases are rectangles. In (*A*), the solid-black bases are in syn orientation. The blue "A" letters and orange "B" letters stand for the C3′-endo and C2′-endo deoxyribose groups, respectively. The pink and green modules represent the components of proteins involved in contacting the nucleic-acid substrates.



Movie S1. Animation of nucleotide-dependent helicase motion. The structures we determined for NS3h-ssDNA, NS3h-ADP · BeF_3-ssDNA, and NS3h-ADP · AlF_4^--ssDNA were used as starting states to create morphed structures. Movie frames were generated in Pymol (http://pymol.sourceforge.net/). NS3h is presented with yellow ribbons. The residues involved in nucleotide and nucleic-acid binding are shown with sticks and color coded according to elements. The DNA strand and nucleotides are modeled with sticks. The DNA carbon atoms are colored in green, and the ADP carbon atoms are colored in pink. The other atoms are color coded according to elements.

Movie S1 (MOV)