Supporting Information: Residue level resolution of alphavirus envelope protein interactions in pH dependent fusion

Xiancheng Zeng[a], Suchetana Mukhopadhyay[b], Charles L. Brooks III[a, 1]

[a] Department of Chemistry and Biophysics Program, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109, USA; [b] Department of Biology, Indiana University, Bloomington, 47405-7000 Indiana, USA

[1] To whom correspondence should be addressed. E-mail: brookscl@umich.edu. Phone: 734/647-6682. Fax: 734/647-1604.

Simulation details

The structures of the E2-E1 envelope proteins in different states were built according to the sequence of envelope proteins of Chikungunya virus (CHIKV). The initial state (M) was taken from PDB structure 3J2W (3N43 fitted to EMD-5577).(1, 2) The conformation of the FI state of CHIKV was constructed based on the M conformation with C_α atoms fitted to 3MUW (SINV, 3MUU fitted to EM-1121)(3) with conserved residues aligned.(1) The HT (folded-back) conformation was built based on 1RER (SFV)(4) and non-conserved residues were mutated according to the sequence of CHIKV. Since the glycans are located on the surface of the glycoproteins E2-E1 and not known to introduce pH dependence, the glycosylated-Asn residues were simulated without glycans to reduce the potential artifacts. We constructed and modeled only the ectodomain of the envelope proteins (E1:1-391; E2:8-341; numbering according to Ref. (1)) with icosahedral symmetry for the M and FI states. We note that the application of icosahedral rotational symmetry here enforces the residues from other image units to adopt the same conformations and pK_a values and also limits the exploration of possible symmetry breaking processes in late fusion stages. The symmetry constraints may result in inherent error to microscopic pK_as due to the possible coupling between symmetry-related neighbors. However, H125 is the only conserved His residue having a corresponding image of itself within 10 Å and the coupling coefficient is found to be about 0.3 between the H125 residues from different E1 proteins within the same MAU. Given the couplings are not too strong, our calculations of the microscopic pK_as present a reasonable approximation to the true pK_as for the systems.

Terminal residues were capped with acetylated N-terminus (ACE) and amidated C-terminus (CT2). The HT state of E1 and dissociated E2 monomer were simulated as isolated systems and no symmetry was applied. The structure of the E2 monomer was adapted from the equilibrated FI state. Since the structure of the unbound E2 is not available, we isolated the E2 protein from the E2-E1 heterodimer and carried out 3.2 ns of molecular dynamics (MD) to relax the structure of the dissociated E2 until the $RMSD_{C\alpha}$ reached a plateau with about 11 Å displacement compared to the initial bound structure in the M state of $(E2\text{-}E1)_3$. The HT core formation and DIII fold-back were taken into account together in one step due to the lack of structural information for the HT in extended form.

In all MD simulations and energy minimizations in this work we use the CHARMM package with the CHARMM27 force field(5) with CMAP corrections.(6) For all structures, we restrained the C_α atoms (excluding unresolved E2 DB of the FI state) with respect to the coordinates in the corresponding original PDB structures and carried out minimization for 12 cycles (500 steps of steepest descents and 500 steps of Adopted Basis Newton-Raphson minimization per each cycle, 4 cycles with k = 100 kcal/mol/Å^2, 4 cycles with k = 10 kcal/mol/Å^2, and 4 cycles with k = 1.0 kcal/mol/Å^2). The missing DB of the E2 protein was transplanted from the M state without any restraint. Equilibration MD runs were carried out on all systems in 6 cycles: 1-3) 100 ps/cycle with harmonic restraints of 10 kcal/mol/Å^2, 3 kcal/mol/Å^2, and 1 kcal/mol/Å^2 on each resolved C_α atom, respectively; 4-5) 200 ps/cycle with RMSD restraints of 200 kcal/mol/Å^2 and 50 kcal/mol/Å^2 on the RMSD(7) of the resolved C_α atoms, respectively; 6) 100 ps pH-REX MD (details below) with 50 kcal/mol/Å^2 on the RMSD of the resolved C_α atoms. For the dissociated E2 state there is no evidence that E2 forms a dimer or trimer after dissociation and it is unclear how E2 dissociates from the spike. Therefore, we assume the metastable post-dissociation state of E2 is monomeric and we adapted its conformation from the $(E2\text{-}E1)_3$ structure of the FI state with additional equilibration MD of 3.2 ns without any restraints to further relax the conformation. While we do not assume this is sufficient to achieve a fully equilibrated E2 conformation, because of the lack of any experimental information regarding the nature of this state, we have utilized this structure as a minimal model of dissociated E2. During the relaxation from the heterodimeric to the dissociated state, E2 undergoes a domain compaction, as shown in Figure S1, most likely due to the missing interactions with E1 in the grooves between the domains. The GBSW implicit solvent model(8) and Langevin integration scheme with a 5.0/ps friction coefficient and 2.0 fs time step were used for all MD simulations in this work. Finally, hydrogen-heavy atom bond were constrained with SHAKE.(9)

All Asp, Glu and His residues (total 404) were allowed to titrate. Since the trans-membrane segment or endodomain was not included in the current simulations, residues with their titrating groups within 4 Å of the terminal residues were not used for pH analysis. Our analysis is focused on His residues, since they are responsible for most of the large thermodynamic contributions to the pH-dependent conformational changes. The conservation frequency of a certain residue is obtained by counting the number of viral species that have conserved His residues at the corresponding site from sequences for 13 alphaviruses (Table S2). A conservation frequency of 13/13 indicates that the specific His residue is strictly conserved

among all 13 alphaviral species we surveyed here. The pH-REX simulations were carried out for the M, FI, and HT states and the dissociated E2 monomer. The pK_a values of the same residues in a viral spike $(E2\text{-}E1)_3$ were averaged for the free energy calculations and analysis. In the pH-REX simulations, 64 replicas covering the pH range from 1.0 to 8.875 with an increment of 0.125 were performed for the M and FI states. Each replica was simulated for 1 ns and total trajectories of 64 ns were generated for both the M and FI states, respectively. 36 replicas covering the pH range from 2.0 to 9.0 with an increment of 0.2 were used for the HT state with 2 ns per replica, generating 72 ns trajectories in total. 24 replicas covering the pH range from 2.0 to 8.9 with an increment of 0.3 were used for the dissociated E2 monomer with 4 ns per replica, generating 96 ns trajectories in total. The convergence of the simulations was evaluated by the standard error of pK_a from the covariance matrix in the non-linear least-squares fitting to the Henderson-Hasselbach equation. The average standard errors of the M, FI, HT(E1) and Dis(E2) were 0.08, 0.08, 0.06, 0.04 pH units, respectively. The standard errors of pK_a of individual titrating residues are shown in Figure S2. The Hill coefficient *n* related to the cooperativity between titrating residues was set to be 1.0, since the changes in *n* only lead to deviations less than 0.02 pK_a units in our systems.

Figures

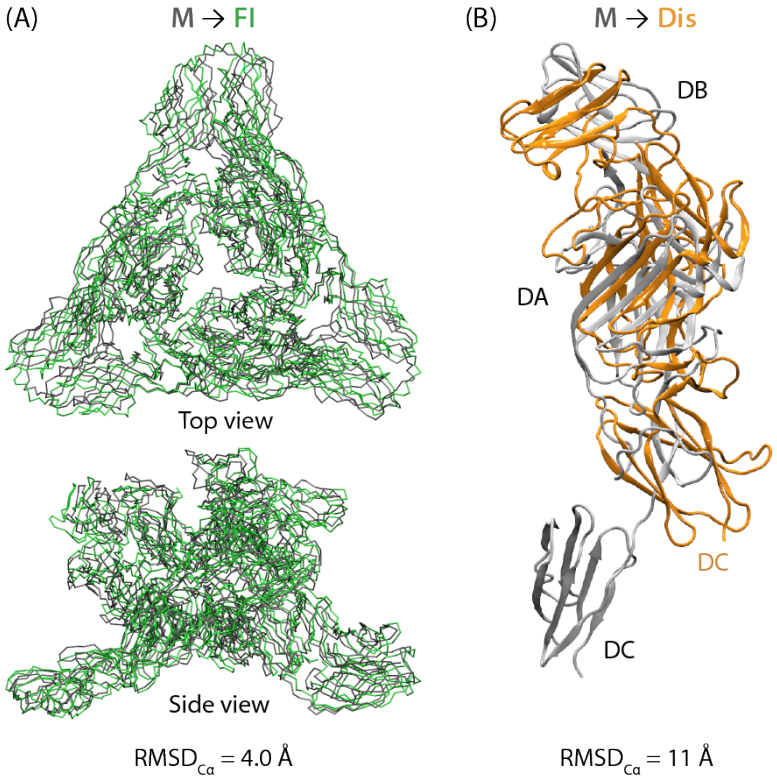

Figure S1. Overlap of structures of E2 and E1 proteins in different states. (A) Overlap of $(E2\text{-}E1)_3$ in the M (gray) and FI (green) states, showing $RMSD_{C\alpha}$ = 4.0 Å. (B) Overlap of E2 domain A, B and C (DA, DB, and DC) in the M (gray) and Dis (orange) states, showing $RMSD_{C\alpha}$ = 11 Å.

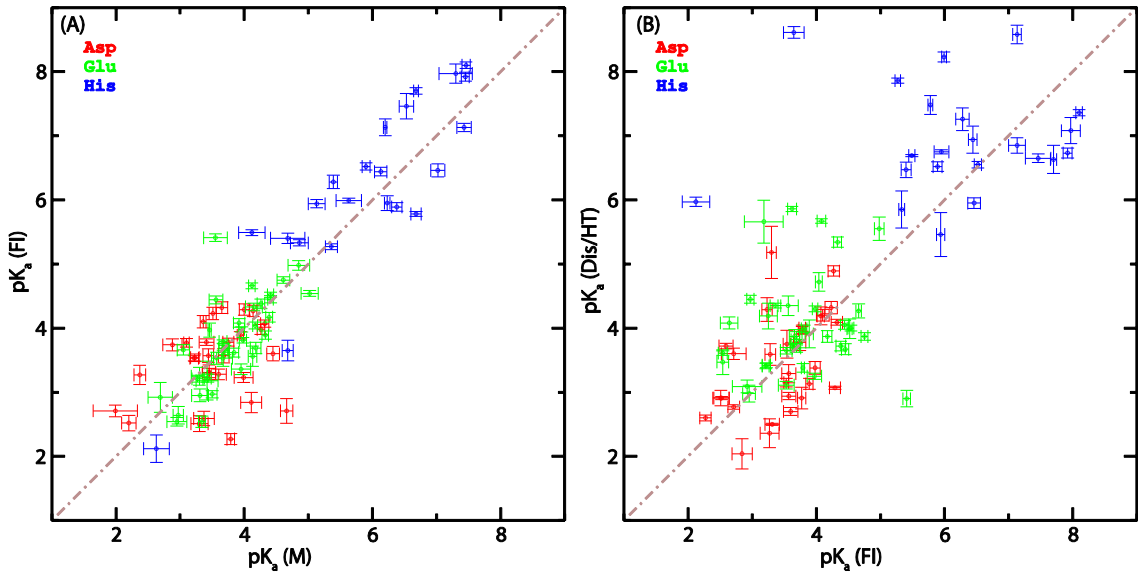

Figure S2. Comparison of averaged pK_a values of all titrating residues in the same spike (A) between the M and FI states, and (B) between the FI and Dis/HT states. Error bars of pK_a in X- and Y-axis reflect the pK_a errors estimated based on block-averages, using 100 ps per block. Red, green, and blue symbols represent Asp, Glu and His residues, respectively.

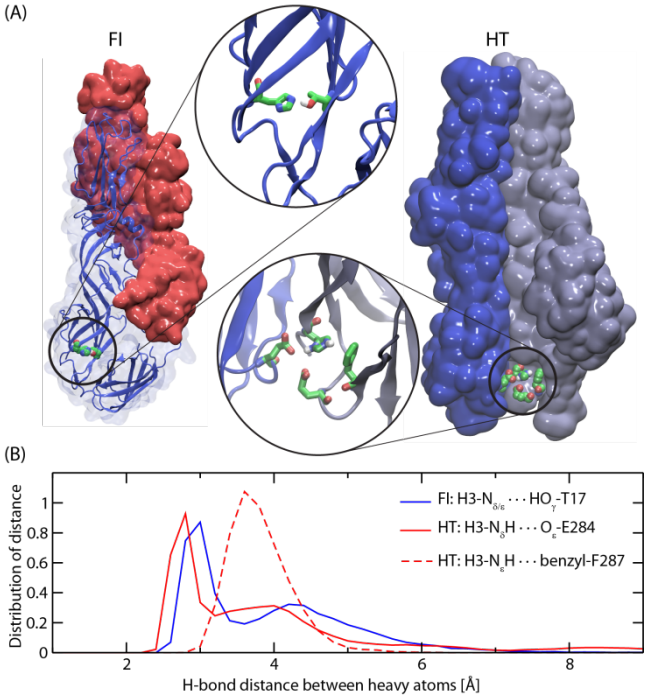

Figure S3. Salt-bridge and H-bond involving H3(E1) in the FI and HT states. (A) H3(E1)-T17(E1) and E1 H3(E1)-E284(E1')/F287(E1) (HT) interactions in the FI and HT states, respectively. (B) Distributions of the H-bond distances of H3(E1)–T17(E1)/E284(E1') / benzyl-F287(E1) (between heavy atoms) in the FI and HT states.

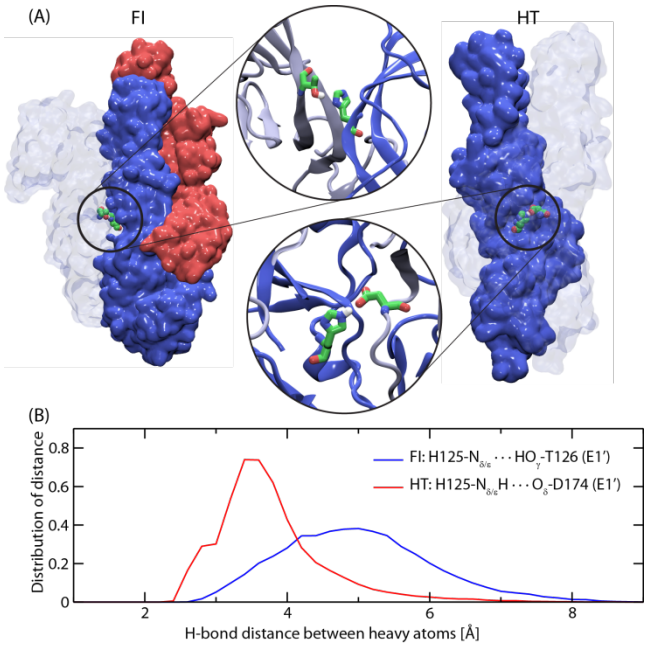

Figure S4. Salt-bridge and H-bond involving E1 H125 in the FI and HT states. (A) H125(E1)–T126(E1) and H125(E1)–D174(E1') interactions in the FI and HT state, respectively. (B) Distributions of the H-bond distances of H125(E1)–T126(E1')/D174(E1') (between heavy atoms) in the FI and HT states. Neighboring E1' is shown in a transparent outline in the overall structure and colored in light blue in the zoomed-in view.

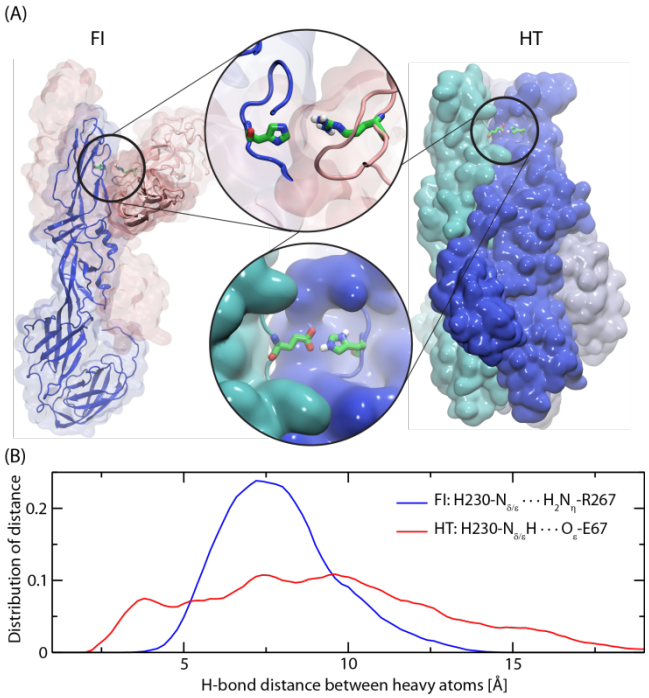

Figure S5. H-bond and Salt-bridge involving H230(E1) in the FI and HT states. (A) H230(E1)-R267(E2') and E1 H230(E1)-E67(E1') interactions in the FI and HT state, respectively. E2 protein is shown in red cartoon and E1 protein is in blue. Neighboring E1s in HT are shown as solid surfaces in light blue and cyan. (B) Distributions of the H-bond distances of H230(E1)-R267(E2')/E67(E1') (between heavy atoms) in the FI and HT states.

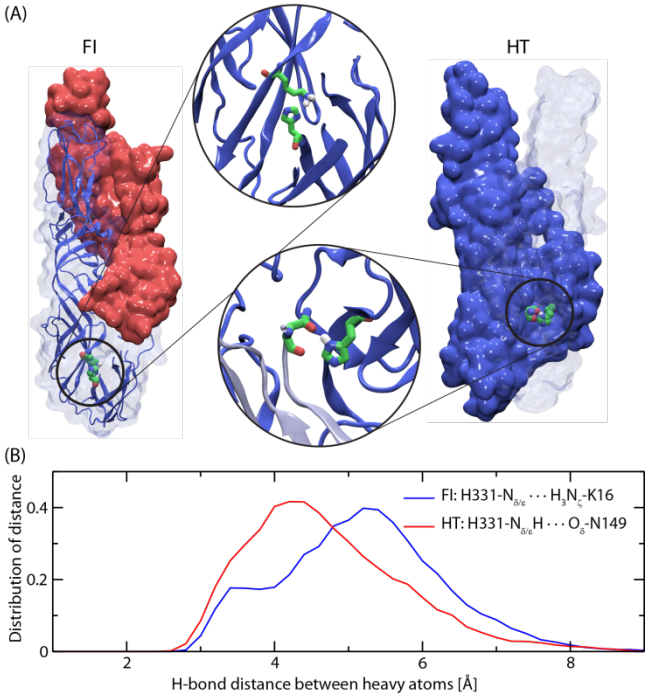

Figure S6. H-bonds involving H331(E1) in the FI and HT states. (A) H331(E1)-K16(E1) and H331(E1)- N149(E1') interactions in the FI and HT state, respectively. (B) Distributions of the H-bond distances of H331(E1)-K16(E1) /N149(E1') (between heavy atoms) in the FI and HT states.

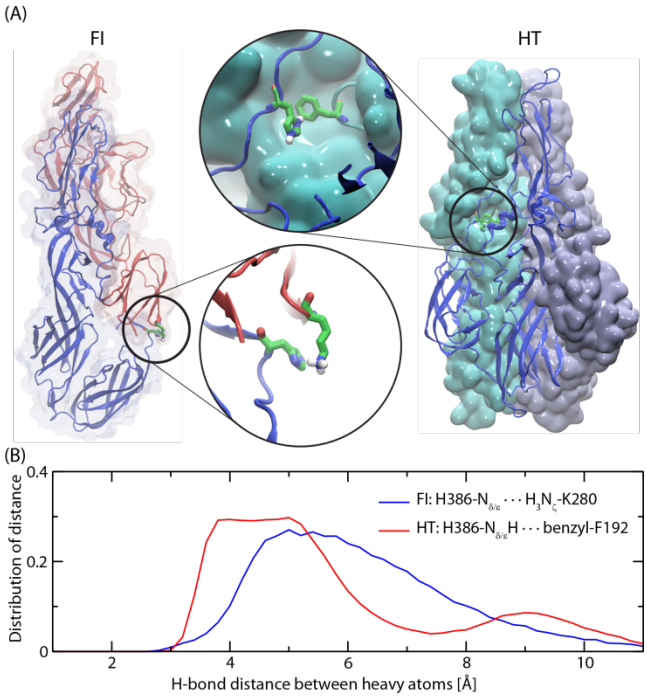

Figure S7. Salt-bridge and H-bond involving H386(E1) in the FI and HT states. (A) H386(E1)- K280(E2) and H386(E1)- F192(E1') interactions in the FI and HT state, respectively. E2 protein is shown in red cartoon and E1 protein is in blue. Neighboring E1s in HT are shown as solid surfaces in light blue and cyan. (B) Distributions of the H-bond distances of H386(E1)-K280(E2)/F192(E1') (between heavy atoms) in the FI and HT states.

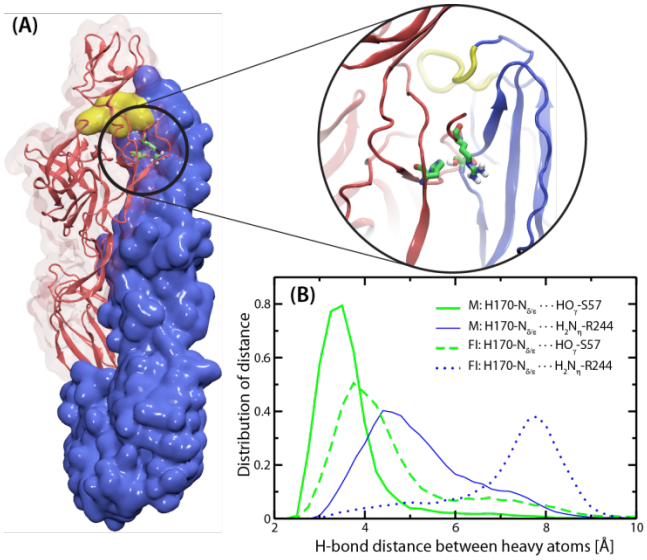

Figure S8. H-bond interactions between H170(E2) and S57(E1). (A) The location of H170(E2)-S57(E1) and the interaction details are shown in the zoomed-in structure. (B) Distributions of the H-bond distances of H170(E2)-S57(E1) and H170(E2)-R244(E2) (between heavy atoms) in the M and FI states.

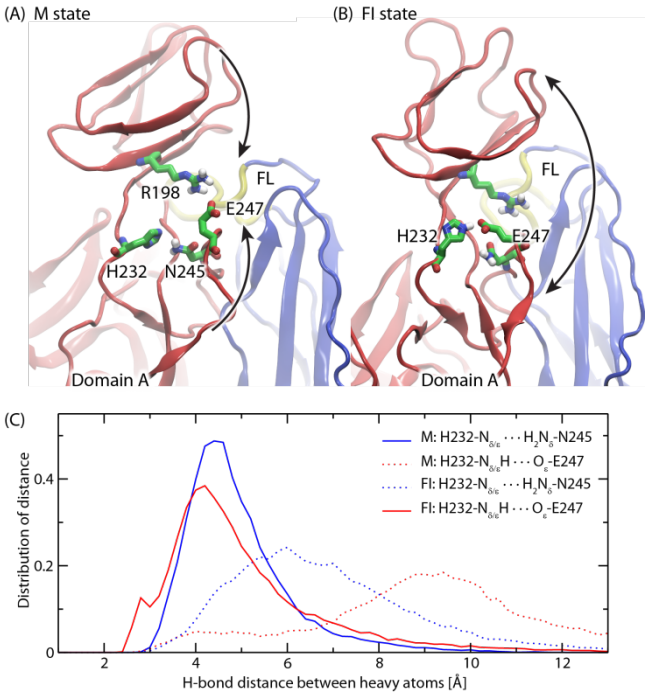

Figure S9. Salt-bridges and H-bonds involving H232(E2) in the M and FI states. (A) H232(E2)-N245(E2) and the nearby E247(E2)-R198(E2) interactions in the M state; (B) new H232(E2)-E247(E2) interactions formed in the FI state that disrupts the previous E247(E2)-R198(E2) interactions between D A-B; (C) Distributions of the H-bond distances of H232(E2)-N245(E2)/E247(E2) (between heavy atoms) in the M and FI states.

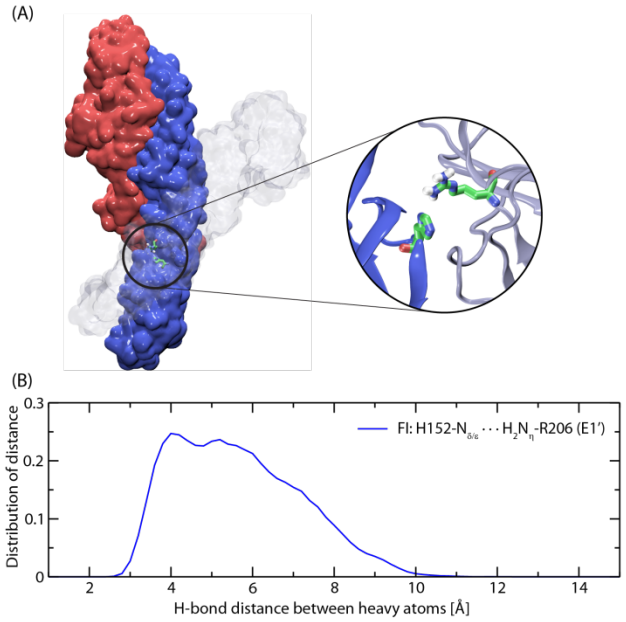

Figure S10. H-bond interaction involving H152(E1) in the FI state. H152(E1) interacts with R206(E1') in the viral spike. (A) locations and structures of H152(E1)-R206(E1'); (B) distributions of the H-bond distances of H152(E1)-R206 (E1') (between heavy atoms).

Tables

Table S1. Residues with |ΔΔG|>1.0 kcal/mol per monomer in the pH dependent activation (M→FI), HT formation and E2 dissociation (FI→Dis/HT).

(A) M→FI					Conservation frequency [b]
Residue	pK_a^M	pK_a^{FI}	ΔpK_a	-ΔΔG [a]	
H170(E2)	6.4	7.5	1.1	1.0	13/13
H232(E2)	6.2	7.1	0.9	1.0	8/13
(B) FI→Dis					Conservation frequency [b]
Residue	pK_a^{FI}	pK_a^{Dis}	ΔpK_a	-ΔΔG [a]	
H226(E2)	5.8	7.5	1.7	1.7	6/13
H131(E2)	6.3	7.3	1.0	1.0	8/13
H127(E2)	5.4	6.5	1.1	1.0	2/13
(C) FI→HT					Conservation frequency [b]
Residue	pK_a^{FI}	pK_a^{HT}	ΔpK_a	-ΔΔG [a]	
H331(E1)	3.6	8.6	5.0	2.7	13/13
H3(E1)	5.3	7.9	2.6	2.2	13/13
H125(E1)	6.0	8.2	2.2	1.8	13/13
H152(E1)	5.5	6.7	1.2	1.2	6/13
H386(E1)	6.6	7.7	1.1	1.0	13/13

[a.] Note –ΔΔG (per monomer) was reported to maintain consistent signs with ΔpK_a. [b.] Conservation frequency shows the number of species that have conserved His on specific site among 13 alphaviruses listed in Table S2(C). A conservation frequency of 13/13 indicates that the specific His residue is strictly conserved among all 13 alphaviral species we surveyed here.

Table S2. Alignment of residue sequences in different human alphaviruses (Ref. (1)) for the highly conserved pH-sensitive sites and the corresponding H-bond partners.

(A) Critical residues from E1 protein and their corresponding H-bond partners.

E1 Residues	H3				H125			H230			H331			H386			H152	
HB partners	3	17	284	287	125	126	174	230	267[a]	67	331	16	149	386	280[a]	192	152	206
CHIKV	H	T	E	F	H	T	D	H	R	E	H	K	N	H	K	F	H	R
ONNV	H	T	D	F	H	T	D	H	R	E	H	K	N	H	K	F	H	R
RRV	H	A	D	F	H	T	D	H	R	E	H	K	N	H	K	Y	H	R
SAGV	H	A	D	F	H	T	D	H	R	E	H	K	N	H	K	Y	H	R
SFV	H	A	D	F	H	T	D	H	R	E	H	K	N	H	K	Y	H	R
MAYA	H	A	D	F	H	T	D	H	R	E	H	K	N	H	K	Y	H	R
MIDD	H	A	D	F	H	T	D	H	R	E	H	R	N	H	K	Y	S	R
BFV	H	A	D	F	H	T	D	H	A	E	H	K	N	H	L	Y	S	R
AURA	H	A	D	F	H	T	D	H	T	E	H	K	N	H	Y	F	T	S
SINV	H	A	D	F	H	T	D	H	M	E	H	K	N	H	F	Y	T	T
WEEV	H	A	D	F	H	T	D	H	P	E	H	K	N	H	F	Y	T	S
VEEV	H	T	D	F	H	T	D	H	T	E	H	N	N	H	F	Y	T	R
EEEV	H	A	D	F	H	T	D	H	I	Q	H	K	N	H	H	Y	T	R

a. R267 and K280 are residues on E2 protein.

(B) Critical residues from E2 protein and their corresponding H-bond partners.

E2 Residues	H170		H232		
HB partners	170	57[b]	232	245	247
CHIKV	H	S	H	N	E
ONNV	H	S	H	N	E
RRV	H	S	H	.	D
SAGV	H	S	H	.	D
SFV	H	S	H	.	D
MAYA	H	S	H	.	E
MIDD	H	S	H	.	A
BFV	H	S	N	.	E
AURA	H	S	H	.	T
SINV	H	S	Q	.	D
WEEV	H	S	Q	.	T
VEEV	H	S	N	.	A
EEEV	H	S	Q	.	E

b. S57 is on E1 protein.

(C) Name codes of 13 alphaviruses used for residue conservation analysis.

Virus	Code	Virus	Code
Chikungunya	CHIKV	Barmah Forest	BFV
O'nyong-nyong	ONNV	Aura	AURA
Ross River	RRV	Sindbis	SINV
Sagiyama	SAGV	Western equine encephalitis	WEEV
Semliki forest	SFV	Venezuelan equine encephalitis	VEEV
Mayaro	MAYA	Eastern equine encephalitis	EEEV
Middleburg	MIDD		

References

1. Voss JE*, et al.* (2010) Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468(7324):709-712.
2. Sun S*, et al.* (2013) Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife* 2.
3. Li L, Jose J, Xiang Y, Kuhn RJ, & Rossmann MG (2010) Structural changes of envelope proteins during alphavirus fusion. *Nature* 468(7324):705-708.
4. Gibbons DL*, et al.* (2004) Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427(6972):320-325.
5. MacKerell AD*, et al.* (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* 102(18):3586-3616.
6. Mackerell AD, Feig M, & Brooks CL, III. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11):1400-1415.
7. Woo H-J & Roux B (2005) Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 102(19):6825-6830.
8. Im W, Lee MS, & Brooks CL, III. (2003) Generalized born model with a simple smoothing function. *J. Comput. Chem.* 24(14):1691-1702.
9. Ryckaert J-P, Ciccotti G, & Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23(3):327-341.