Cryo-EM structure of the mature and infective Mayaro virus at 4.4 Å resolution reveals features of arthritogenic alphaviruses

SUPPLEMENTARY INFORMATION



Supplementary Fig. 1. Fourier shell correlation profile of MAYV Cryo-EM density map. FSC curve calculated from two MAYV 3D volumes is shown in blue, the ½--bit curve in red and the 0.143 crossing line in green. The global resolution determined based on the ½--bit or 0.143 criterion is indicated60.



Supplementary Fig. 2. Purified MAYV stocks used in Cryo-EM experiments are viable and infective. (A) Representative pictures of plaque assays performed in Vero CCL81 cells to assess infective MAYV load in stock samples. Viable Vero CCL81 cells were stained in Methylene Blue 1% w/ v. MAYV replication leads to cell death, creating transparent lysis plaques in the blue Vero CCL81 monolayer. (B) Viral load of stocks used in Cryo-EM experiments, before and after the purification process. (C) An example of MAYV Cryo-EM micrograph from set of 8792 movie stacks. The micrograph was processed by applying a band-pass filter in Imagic software to suppress very low-frequency information (less than 0.02 of Nyquist frequency), removing ramps of background fluctuations, and to suppress very high-frequency information (higher than 0.4 of Nyquist frequency), allowing a better visualization of low frequency content. After filtering, we resized the micrograph to 512x512 dimensions. Inset presents individually raw picked particles used in 3D reconstruction.



Supplementary Fig. 3. Absence of the E3 protein indicates that purified MAYV samples are composed of mature virions. (A) Superposition of the MAYV electron density with the SINDV particle cryo-EM structure (entry 6IMM). The SIDNV E3 protein is shown in orange. MAYV electron density is shown in grey contoured at 2.0 sigma-level. (B) Analysis of purified MAYV in 15% SDS-PAGE gel, stained with Coomassie brilliant blue. Lane 1 is carried with 30 µg of denatured purified MAYV and shows no visible bands compatible with E3 protein size (7 kDa). Lanes 2 and 3 are carried with 1 µg and 9 µg of a 5 kDa synthetic peptide, used as a positive control. This gel is representative of 7 experiments with similar results.



Supplementary Fig. 4. List of the main interfaces among MAYV structural proteins of the asymmetric unit. The interfaces were identified using the PDBePISA server (https://www.ebi.ac.uk/pdbe/prot_int/pistart.html). Only interfaces with area larger than 100 Å were considered. The structures are rendered as spacefill and the interaction partners are colored in blue or cyan. Interaction residues are colored in red or green. Nhb and Nsb represent the average number of hydrogen bonds and salt-bridges in each interface, respectively.

A - Multiple Sequence Alignment of the Alphaviral Capsid

	1					1	0							20	2									З	0									4 0					
MAYV	MD	F	LE	T	QV	F	ΥG	RF	R W I	R.	P	RI	и.	ΡE	R		P	W			. R	PI	RP	ΡΊ	Ί		•	. Q	RI	ΡD	QQ		R	2 M	QÇ	2 L	ΙA	A	VS
CHIKV	ΜE	F	ΙE	ΡT	QI	F	ΥN	RF	R Y (Q.	₽	RΙ	PW	ΤF	R		. P	T	ΙQ	V	ΙR	PI	RP	R.	•				. 1	ΡÇ	R	2 <mark>A</mark>	G	ΩL	AC	2 L	IS	A	VN
SFV	ΜN	ΙY	IE	T	QI	F	ΥG	RF	R 🕅 1	R.	₽	RI	2.	AA	AR		. P	W	ΡL	Q.	ΑT	P۱	/A	ΡV	V				. 1	PD	F	2 <mark>A</mark>	Q	QM	QC	2 L	IS	A	ИV
RRV	ΜN	ΙY	IE	T	QI	F	ΥG	RF	R W I	R.	₽	RI	2.	ΑF	R		. P	W	QV	S	МQ	P	ΓP	ΤM	IV	ΤP	М	LQ	AI	ΡD	L	2 <mark>A</mark>	Q	ΩM	QÇ	2 L	IS	A	VS
SINDV	ΜN	Ι.	• •	•	RG	F	FN	ΜI	G	RF	₹P	FΙ	PA	ΡΊ	ΓA		. М	W		•	. R	PI	RR	RF	Q.	AA	P	ΜP	AI	RN	[G]	L A	S	ΩI	QÇ	2 L	ТΤ	A	VS
VEEV	FΡ	F	QE	M	ΥF	M	QP	ΜĐ	Y	RN	ΙP	FΖ	ΑA	ΡF	R		. P	W	FΡ	R	ΤD	₽.	• •		•		•		•		F1	LA	М	2V	QE	Ľ	ΤF	S	ΜA
EEEV	LN	IY	PE	M	AF	γI	NP	ΜÆ	Y	RD	P	ΝI	2.	ΡF	۲Q	VP	A P	F		•	. R	P					•		•		P I	LA	A	ΩI	ΕĽ) L	RF	S	ΙA

	50	eò	7 <u>0</u>	8 Q	эo
MAYV	TLALRQ.	NAAAPQRGR	KKQPR <mark>RK</mark> KPKPQP.	EK <mark>PK</mark> KQEQKP	K QKKT P
CHIKV	K <mark>L</mark> TMRA.	VPQQKP	RRN <mark>R</mark> KN <mark>K</mark> KQKQKQ.	QAPQNNTNQK	KQPPKKKPAQ
SFV	A <mark>L</mark> TMRQ.	NAIAPARPPI	KPK <mark>K</mark> K <mark>K</mark> TTKPKPK.	TQPKKINGKT	QQQKK <mark>K</mark> DKQAD
RRV	A <mark>L</mark> TTKQ.	NVKAP.KGQ	RQK <mark>K</mark> QQ <mark>K</mark> PKEKKE.	NQKKKPTQKK	KQQQK <mark>P</mark> KPQA
SINDV	A <mark>L</mark> VIGQA	TRPQ <mark>P</mark> PRPRI	PPP <mark>RQKK</mark> QAPKQP.	PKPKKPKTQE	<mark>К</mark> КККQ <mark>Р</mark>
VEEV	N <mark>L</mark> TFKQ.	RRDAPPEGP	SAK <mark>K</mark> P <mark>KK</mark> EASQKQK	GGGQGKK <mark>K</mark> KNQGKKK	A K T G P <mark>P</mark> N P K A Q N G N K
EEEV	N <mark>L</mark> TLKQ.	RAPNPPAGP	PAK <mark>R</mark> K <mark>K.</mark>	PAPSLSLETK	KKRPPPPAKKQ

		100	110	120	130	140
MAYV	KK	K P G R R E R M C M F	KIEHDCIFEVK	HE.GKVTGY2	ACLVGDKVMKP	A HV P G V ID NID L AF
CHIKV	KKK	K P G R <mark>R E R M</mark> C M F	IENDCIF EVK	HE. <mark>G</mark> KVT <mark>G</mark> Y <mark>2</mark>	ACLVGDKVMK <mark>P</mark>	A H V K G T I D N A D L A F
SFV	. KK <mark>KK</mark>	K P G K <mark>R E R M</mark> C M F	KIENDCIFEVK	HE. <mark>G</mark> KVT <mark>G</mark> Y <mark>2</mark>	ACLVGDKVMK <mark>P</mark>	A HV K G V ID NAD L AK
RRV	KKK	K P G R <mark>R E R M</mark> C M F	KIENDCIFEVK	LD. <mark>G</mark> KVT <mark>G</mark> Y <mark>2</mark>	ACLVGDKVMK <mark>P</mark>	A HV K G T ID NPD L AK
SINDV	A <mark>k</mark> p	K P G K <mark>R</mark> Q R M A L B	KLEADRLFDVK	NEDGDVIGH	ALAMEGKVMK <mark>P</mark>	L HV K G T ID HPV L SK
VEEV	KKTNK	K P G K R Q R M V M P	K L E S D K T F P I M	ILE. <mark>G</mark> KIN <mark>G</mark> Y <mark></mark>	ACVVGGKLFR <mark>P</mark>	MHVEGKIDNDVLAA
EEEV	. KR <mark>K</mark> P	K P G K R Q R M C M F	L E S D K T F P I M	ILN. <mark>G</mark> QVN G Y <mark>2</mark>	ACVVGGRVFKP	LHVEGRIDNEQLAP

	150	160	170	180	190	200	
MAYV	LSYKKS	SKYDLECAQI	P V A M K S D A S K	Y T H E K P E G H Y	NWHYGAVQY	TGG <mark>RFT</mark> V <mark>P</mark> T <mark>G</mark> V	GKPG
CHIKV	LAFKRS	SKYDLECAQI	P V H M K S D A S K I	FTHEKPEGYY	NWHHGAVQY	SGG <mark>RFT</mark> I <mark>PTG</mark> A	GKPG
SFV	LAFKKS	SKYDLECAQI	PVHMRSDASK	Y T H E K <mark>P</mark> E G H Y	NWHHGAVQY	SGG <mark>RFT</mark> I P T <mark>G</mark> A	GKPG
RRV	LTYKKS	SKYDLECAQI	P V H M K S D A S K '	Y <mark>T</mark> H E K <mark>P E G H Y</mark>	NWHHGAVQY	SGG <mark>RFT</mark> I PTG A	GKPG
SINDV	LKFTKS	SAYDMEFAQL	P V N M R S E A F T '	Y <mark>T</mark> SEH <mark>PEG</mark> FY	NWHHGAVQY	SGG <mark>RFT</mark> I <mark>P</mark> R <mark>G</mark> V	GRG
VEEV	LKTKKA	SKYDLEYADV	P QN <mark>M</mark> RADTFK	Y T H E K P Q G Y Y	S <mark>WH</mark> H <mark>GAVQY</mark>	E N G <mark>R F T</mark> V <mark>P</mark> K <mark>G</mark> V	GAKG
EEEV	IKLKKA	SIYDLEYGDV	PQCMKSDTLQ	Y T S D K P P G F Y	NWHHGAVQY	ENN RFTVP R <mark>G</mark> V	GGKG

	210	220	230	240	250	
MAYV	DSGRP	IFDNKGRVVA	IVLGGANEGAF	RTALSVVTWN	. KDMVTKI TPE GT	ΕEŴ
CHIKV	DSGRP	IFDNKGRVVA	IVLGGANEGAF	RTALSVVTWN	. KDIVTKI TPE GA	EEW
SFV	DSGRP	IFDNKGRVVA	IVLGGANEGSF	RTALSVVTWN	. KDMVTRV TPE GS	EEW
RRV	DSGRP	IFDNKGRVVA	IVLGGANEGAF	RTALSVVTWT	. KDMVTRV TPE GT	ΕEW
SINDV	DSGRP	IMDNSGRVVA	IVLGGADEGTF	RTALSVVTWNS	S <mark>K</mark> GKTIKT TPE GT	ΕEW
VEEV	DSGRP	ILDNQGRVVA	IVLGGVNEGSE	RTALSVVMWNI	EKGVTVKY TPE NC	EQW
EEEV	DSGRP	ILDNKGRVVA	IVQGGVNEGSE	RTALSVVTWN	Q <mark>K</mark> GVTVKD TPE GS	EPW

B - Multiple Sequence Alignment of the Alphaviral E1

	i 1	ιọ	20	30	40	50	60	70	80
MAYV CHIKV SFV RRV SINDV VEEV EEEV	YEHTAVIPI YEHVTVIPI YEHSTVMPI YEHTATIPI YEHATTVPI YEHATTMPS YEHTAVMPI	VQVGFPYKA VTVGVPYKT VVGFPYKA VVVGFPYKA VVPQIPYKA QAGISYNT VKVGIPYKA	HVAREGYSP LVNRPGYSP HIERPGYSP HIERNGFSP LVERAGYAP IVNRAGYAP LVERPGYAP	LTLQMQVVE MVLEMELLS LTLQMQVVE MTLQLEVVE LNLEITVMS LPISITPTK VHLQIQLVN	TSLEPTLNLEY VTLEPTLSLDY TSLEPTLNLEY TSWEPTLNLEY SEVLPSTNQEY IKLIPTVNLEY TRIIPSTNLEY	ITCDYKTKVE ITCEYKTVIE ITCEYKTVVE ITCEYKTVVE ITCKFTTVVE VTCHYKTGME ITCKYKTKVE	SPYVKCCGTA SPYVKCCGTA SPYVKCCGAS SPFIKCCGTS SPKIKCCGSL SPAIKCCGSQ SPVVKCCGAT	ECRTQDKPEY ECKDKNLPDY ECSTKEKPDY ECSSKEQPDY ECOPAAHADY ECTPTYRPDE QCTSKPHPDY	KCAV SCKV QCKV QCKV TCKV QCKV QCV
MAYV CHIKV SFV RRV SINDV VEEV EEEV	FTGVYPFMU FTGVYPFMU YTGVYPFMU FGGVYPFMU FTGVYPFMU FTGVYPFMU	O NGGAYCFCD NGGAYCFCD NGGAYCFCD NGGAYCFCD NGGAYCFCD NGGAYCFCD	100 SENTQMSEA AENTQLSEA SENTQLSEA SENTQLSEA SENSQMSEA TENTQVSKA TENTQMSEA	110 YVERADVC HVEKSESCK YVDRSDVCR YVDRSDVCR YVELSADCA YVMKSDDCL YVERSEECS	120 HDYAAAYRAHT TEFASAYRAHT HDHASAYKAHT SDHAQAIKVHT ADHAEAYKAHT IDHAKAYKVHT	130 ASLRAKIKVI ASASAKLRVI ASLKAKVRVN ASLKATIRIS AAMKVGLRIV ASVQAFLNII GTVQAMVNIJ	140 YGTVN.QTVE YQGNN.ITVT YQGNVN.QTVD YGTIN.QTTE YGTIN.QTTE YGNTT.SFLD VGEHS.IVTT YGSVTWRSAD	150 AYVNGDHAVT YANGDHAVT YYVNGDHAVT AFVNGEHAVN YYVNGTPGT YYVNGETPVN YYVNGETPAK	I A G T VKDA I GGT VGGS SKDL FNGV I GDA
1 MAYV CHIKV SFV RRV SINDV VEEV EEEV	60 KFIFGPVS VFIVGPMS VFIFGPIS KFIFGPIS KVIAGPIS KVIAGPIS KLIIGPLS	LTO FAWTPFDTK SAWTPFDNK SAWTPFDNK SAWTPFDNK ASFTPFDHK SAWSPFDNK SAWSPFDNK	180 IVVYKGEVY IVVYKGDVY IVVYKDEVF IVVYKDEVF VVIHRGLVY IVQYAGEIY VVYGHEVY	190 NQDFEPFGA NMDYPFGA NQDFPPYGS NQDFPEYGS NYDFPEYGA NYDFPEYGA NYDFPEYGA	200 GQFGRFGDIQS GRPGQFGDIQS GQPGRFGDIQS MKPGAFGDIQA GQPGAFGDIQA GQPGAFGDIQA GQPGAFGDIQA GQPGAFGDIQA	210 RTLDSKDLYA RTPESKDVYA RTVESKDLYA RTVESKDLYA TSLTSKDLIA RTVSSSLYA RTSJNDLYA	220 NTGLKLARPA NTQLVLQRPA NTALKLARPS STDIRLLKPS NTNLVLQRPK NTNLVLQRP	230 AGNIHVPYTQ AGTVHVPYSQ PGMVHVPYTQ PGVVHVPYTQ AKNVHVPYTQ AGAIHVPYTQ AGIVHTPFTQ	T P S G A P S G T P S G A S S G A P S G A P S G
2 MAYV CHIKV SFV RRV SINDV VEEV EEEV	40 FKTWQKDRI FKYWLKERC FKYWLKEKC FKYWLKEKC FEMWKNNSC FEQWKKDKA FEQWKRDKA	250 SPLNAKAP SASLQHTAP TALNTKAP SSLNTKAP SRPLQETAP SAPLNDVAP	260 FGCTIQTNP FGCQIATNP FGCQIKTNP FGCKIKTNP FGCKIAVNP FGCCSIALEP	270 VRAMNCAVG VRAMNCAVG VRAMDCAVG IRAENCAVG LRPENCAVG	280 NIPUSMDIADS NMPISIDIPDA NIPUSMDIPDS NIPISIDIPNA SIPLAFDIPDA SIPISIDIPDA	290 AFTRLTDAPI AFTRVVDAPS AFTRIVEAPI AFTRVSDAPI LFTRVSETPI AFTRISETPI	300 ISELLCTVST LTDMSCEVPA IIDLTCTVAT VTDLSCQVVV VSTVKCEVSE LSAAECTLNE VSDLECKITE	310 CTHSSDFGGV CTHSSDFGGV CTHSSDFGGV CTHSSDFGGV CTYSSDFGGM CTYSSDFGGI CTYSSDFGGI	AVLS AIIK LTLT ATLS ATLQ ATVK ATLP
3: MAYV CHIKV SFV RRV SINDV VEEV EEEV	20 YKVEKAGR YAVSKKGK YKTNKNGD YKTDKPGQ YVSDREGQ YSASKSGKC TNPVKQET	B 3 0 C DVHSHSNV C AVHSHSNV SVHSHSNV C AVHSHSNV C PVHSHSST C AVHVPSGT Z OF I VHQVL	340 AVLQE.VS VTIREAEIE ATLQEATAK ATLQEATVD ATLQESTVH ATLKEAAVE QLLKRMTSP	350 IEAEGRSVI VEGNSQLQI VKTAGKVTL VKEDGKVTV VKEDGKVTV LTEQGSATI LLRAGSFTF	360 HFSTASAAPSF SFSTALASAEF HFSTASASPSF HFSTASASPAF HFSTASPQANF HFSTANIHPEF HFSTANIHPEAF	370 IVSVCSSRAT RVQVCSTQVF VVSLCSARAT KVSVCDAKTT IVSLCGKKT RLQICTSYVT KLQVCTSGIT	380 CTAKCEPPKD CAAECHPPKD CSASCEPPKD CTAACEPPKD CNAECKPPKD CKGDCHPPKD CKGDCKPPKD	390 HVVT YPANHN HIVNYPASHT HIVPYAASHS HIVPYGASHN HIVSTPHKND HIVTHPQYHA HIVDYPAQHT	GITL TLGV NVVF QEFQ QTFTT ESFT
MAYV CHIKV	400 PDLSSTAM QDISATAMS	410 IWAQHLAGC WVQKITGG	420 VGLLIALAV VGLVVAVAA	430 Lilvivitc. Lilivvic.	. ITLRR . VSFSRH.				

MAYV PDLSSTAMIWAQHLAGGVGLLIALAVLILVIVIC..ITLRR.. CHIKV QDISATAMSWVQKIITGGVGLVVAVAALILIVVIC..VSFSRH. SFV PDMSGTALSWVQKISGGLGAFAIGAILVLVVVVTC..IGLRR.. RRV PDMSGTAMTWVQRLASGLGGLALIAVVVLVVVTC..IIMRR.. SINDV AAISKTSWSWLFALFGGASSLLIIGLMIFACSMM..LTSTRR. VEEV AAVSKTAWTWLTSLLGGSAVIIIIGLVLATIVAMYVLTNQKHN EEEV SAISATAWSWLKVLVGGTSAFIVLGLIATAVVAL.VLFFHRH.

C - Multiple Sequence Alignment of the Alphaviral E2

	1 10	20	30	40	50	60	70
MAYV CHIKV SFV RRV SINDV VEEV EEEV	STANHFNAYKLT STKDNFNVYKAT SVSQHFNVYKAT SVTEHFNVYKAT SVIDDFTLT STEELFNEYKLT DLDTHFTQYKLA	RPYVAYCADCGM RPYLAHCPDCGE RPYLAHCPDCGE RPYLAYCADCGD SPYLAYCADCGD SPYLGTCSYCHH RPYMARCIRCAV RPYIADCPNCGH	GHSCHSPAMI GHSCHSPVAL GHSCHSPVAL GYFCYSPVAL TVPCFSPVAL G.SCHSPIAL S.RCDSPIAL	SNVQADATDGTI SRIRNEATDGTI AVRSEATDGMI KIRDEAPDGMI QVWDEADDNTI AVKSDGHDGYV SEVRGDAHAGVI	KIQFASQIGI KIQVSLQIGI KIQVSAQIGI KIQVSAQIGI RIQTSAQFGY RIQTSAQFGY RIQTSAMFGL	TKTDTHDHTKI GTDDSHDWTKL DKSDNHDYTKI DKAGTHAHTKI DQSGAASANKY DSSGNLKGRTM KRHGV.DLAYM	RYAEGHDIAE RYMDNHIPAD RYADGHAIEN RYMAGHDVQE RYMSLKQDHTVKE RYDMHGTIKE SFMNGKTQKS
MAYV CHIKV SFV RRV SINDV VEEV EEEV	8099 AARSTLKVHSSS AGRAGLFVRTSA AVRSSLKVATSG SKRDSLRVYTSA GTMDDIKISTSG IPLHQVSLYTSR IKIDNLHVRTSA	• 100 ECAVIGIMGHFI PCTITGTMGHFI CFVHGTMGHFI ACSIHGTMGHFI PCRRLSYKGYFL PCHIVDGHGYFL PCSLVSHHGYYI	110 LAKCPPGEVII LAKCPPGETLI VAHCPPGDYLI LAKCPPGDYLI LAKCPPGDSII LARCPAGDSII	120 SVSFVDSKNEQF IVGFTDSRKISH VSFEDADSHVK SVSFEDADSHVK IVSIVSS.NSAT IMEFKKD.SVRH IVGFHDG.PNRH	130 TCRIAYHHEQ SCTHPFHHDP ACRIQYHHDP SCTLARKIKP SCVPYEVKF TCRLAHKVEF	140 RLIGRERFTVR PVIGREKFHSR QPVGREKFTIR LPVGREKFVVR KFVGREKYDLP NPVGRELYTHP RPVGREKYRHP	150 PHHGIELPCTTYQ PQHGKELPCTTYQ PHYGKEIPCTYQ PHFGVELPCTSYQ PVHGKKIPCTVYD PEHGVEQACQVYA PEHGVELPCNRYT
MAYV CHIKV SFV RRV SINDV VEEV EEEV	160 17 LTTAETSEEIDM QSNAATAEEIEV QTTAETVEEIDM LTTAPTDEEIDM RLKETTAGYITM HDAQNRGAYVEM HKRADQGHYVEM	0 180 HM P PDIPDRTLL HM P PDTPDRTLL HM P PDTPDRTLL HTP PDIPDRTLL HRP RPHAYTSYL HLP GSEVDSSLV HQP GLVGDHSLL	190 SQQSGNVKIT SQQSGNVKIT SQQSGNVKIT SQTAGNVKIT EESSGKVYAKI SLSGSSVTVT SIHSAKVKIT	200 V.NGRTVKYSCS V.NGRTVRYKCN V.GGKKVKYNCT A.GGRTIRYNCT PPGKNITYECK PPDGTSALVECE VPSGAQVKYYCK	210 CGSKPSG.TI CGSNEG.LI CGTGNVG.TT CGRDNVG.TT CGDYKTG.TV CGTKISETI CDYREG.IT	220 TTDKTINSCT. TTDKVINNCK. NSDMTINTCL. STDKTINTCK. STRTEITGCTA NKTKQFSQCTK SSDHT.TTC	230 VDKCQAYVTSHTK VDQCHAAVTNHKK IEQCHVSVTDHKK IDQCHAAVTSHDK IKQCVAYKSDQTK KEQCRAYRLQNDK VKQCRAYLIDNKK
MAYV CHIKV SFV RRV SINDV VEEV EEEV	240 WQFNSPFVPR.A WQYNSPLVPRNA WQFNSPFVPR.A WQFTSPFVPR. WVFNSPDLIR.H WVYNSDKLPK.A WVYNSDKLPK.A	250 EQAERKGKVHIP ELGDRKGKIHIP DEPARKGKVHIP DQTARRGKVHVP DDHTAQGKLHVP AGATLKGKLHVP EGDTFKGKLHVP	260 FPLINTICRVI FPLANVICMVI FPLDNITCRVI FPLJNTCRVI FKLIPSTCMVI FLLADGKCTVI FVPVKAKCIA	270, 28 PLAPEALVRSGK PKARNPTVTYGK MAREPTVIHGK PLARAPDVTYGK VAHAPNVIHGF PLAPEPMITFGF TLAPEPLVEHKH	Q 29 REATLSTHPI NQVIMLLYPD REVTLHLHPD KEVTLRLHPD KHISLOLDTD RSVSLKLHPK RTLILHLHPD	0.300 HPTLLSYRTLG HPTLFSYRSLG HPTLFSYRSLG NPTYLITRRLG NPTYLITRSLG	310 REPVFDEQWITTQ EEPNYQEEWVTHK EDPQYHEEWVTAA AEPHPYEEWVDKF ANPEPTTEWIVFK DEPHYTHELISEP SDANPTRQWIERP
MAYV CHIKV SFV RRV SINDV VEEV EEEV	320 TEVTIPVPVEGV KEVVLTVPTEGL VERTIPVPVPGM SERIIPVTEGI TVRNFTVTEKGW AVRNFTVTEKGW TTVNFTVTGEGI	330 EYRWGNHKPQRL EVTWGNNEPYKY EYHWGNNDPVRL EYLWGNHEPVRV EFVWGNHEPKRV EFVWGNHPPKRF	340 WSQLTTEGRA WPQLSANGTA WSQLTTEGRP WSQLTTEGRP WSQLTTEGRP WSQLTTEGRP WSQETAPGNP WAQESGEGNP	350 GWPHEIIEYYY GHPHEIILYYY GWPHEIIQYYY GWPHEIIQYYY HGWPHEIVQHYY HGWPHEVVVVYYY	0 37 GLHPTTTIVV ELYPTMTVVV GLYPAATVVA GLYPAATIAA HRHPVYTILA HRHPVYTILA NRYPMSTILG NRYPLTTILG	0 380 VVAVSVVVLLS VSVASFILLSM VVGMSLLALIS VSGASLMALLT VASATVAMMIG LSICAAIATVS LCTCVALIMVS	390 VAASVYMCVVARN VGMAVGMCMCARR IFASCYMLVAARS LAATCCMLATARR VTVAVLCACKARR VAASTWLFCRSRV VAASTWLFCRSRV ODHPCGSFSGLRN
MAYV CHIKV SFV RRV SINDU	400 KCLTPYALTPGA KCLTPYALTPGA KCLTPYALTPGA KCLTPYALTPGA	410 VVPVTIGVLCCA TVPFLLSLICCI AVPWTLGILCCA VVPLTLGLLCCA	420 PKAHA RTAKA PRAHA PRANA PRANA				, , , , , , , , , , , , , , , , , , ,

Supplementary Fig. 5. Multiple sequence alignment of alphaviruses structural proteins. The alignments were performed for proteins C (A), E1 (B) and E2 (C) using the Muscle Algorithm in EBI server (https://www.ebi.ac.uk/Tools/msa/muscle/) and visualized with ESPript version 3.0 server.

VEEV EEEV

AC

PY<mark>RLTPNARIPFCLAVLCC</mark>ARTARA PYKLAPNAQVPILLALLCCIKPTRA



Supplementary Fig. 6. MAYV E1 and E2 transmembrane domains. Overall view of E1 and E2 TM domains inserted into the lipid membrane. The 3D atomic model is colored by hydrophobicity using Kyte-Doolittle scale. Important arginine and lysine residues in E2 are highlighted in spheres, as well as glycine (E1) and serine (E2) residues. The electron density is shown in grey surface.



Supplementary Fig. 7. 3D model fitting of a C18 hydrocarbon (Octadecane) in the extra density from two MAYV E1-E2 heterodimers (chains B/C and chains L/M). Octadecane was built and fitted into the density map using Coot. The density map was rendered at 2.5 sigma contour level.



Supplementary Fig. 8. Structural features extracted from the cavity between E1 and E2 TM helices in MAYV and other alphaviruses. (A) Cavity volume estimated for the four E1-E2 heterodimers (n = 4 independent heterodimer structures) in asymmetric unit. One-way ANOVA with Tukey's multiple comparison test was used for comparing MAYV cavity volume with other alphaviruses (* indicates adj. p < 0.01 when comparing the alphaviruses to MAYV). (B) Number of residues in each four E1-E2 heterodimers (n = 4 independent heterodimer structures) in asymmetric unit separated by classes. One-way ANOVA with Tukey's multiple comparison test was used for comparing the number of residues in aliphatic apolar class with the other classes in the same alphavirus species (* indicates adj. p < 0.01 when comparing the aliphatic apolar class with the other classes). All data are presented as mean values +/- SD. Aliphatic apolar: ALA, VAL, ILE, LEU, GLY, PRO; Aromatic: PHE, TYR, TRP; Polar uncharged: SER, THR, CYS, MET, ASN, GLN; Negatively charged: GLU, ASP; Positively charged: ARG, LYS, HIS.



Supplementary Fig. 9. Structural features extracted from the C-protein cavity that binds to E2 C-terminal. (A) Boxplot of cavity volume estimated for the four capsids (n = 4 independent capsids structures) in asymmetric unit. In the boxplot, the box portion defines the interquartile range (IQR) (67.5 Å3) and the 75th (Q3) (526.2 Å3) and 25th (Q1) (458.7 Å3) percentiles. The central line indicates the median (486.3 Å3) and the mean (498.6 Å3) is indicated by a dot. The whiskers with minimum (395.7 Å3) and maximum (626.2 Å3) are determined using Q1-1.5 x IQR and Q3+1.5 x IQR, respectively. (B) Number of residues in each four capsids (n = 4 independent capsids structures) in asymmetric unit separated by classes. One-way ANOVA with Tukey's multiple comparison test was used for comparing the number of residues in aliphatic apolar class with the other classes) All data are presented as mean values +/- SD. Aliphatic apolar: ALA, VAL, ILE, LEU, GLY, PRO; Aromatic: PHE, TYR, TRP; Polar uncharged: SER, THR, CYS, MET, ASN, GLN; Negatively charged: GLU, ASP; Positively charged: ARG, LYS, HIS.



Supplementary Fig. 10. The electrostatic potential of at the surface of MAYV capsid proteins. The electrostatic potential was calculated using ABPS and visualized in Pymol. Top, bottom and side view from capsid hexameric organization is shown. Charges are presented in a gradient of blue (positive) to red (negative), white being neutral local charge.







Supplementary Fig. 11. Spectra of identified peptides, after Endo H and trypsin digestions, annotated by Proteome Discoverer and manually verified. (A) Spectrum of the peptide VTYGTVNQTVEAYVNGDHAVTIAGTK (m/z 971.14655, +3) from E1 protein with the N-acetylglucosamine in N141 residue, and (B) VHIPFPLINTTCR (m/z 590.97968, +3) E2 peptide with the N-acetylglucosamine in the N262 residue. The blue square represents the N-Acetyl-D-Glucosamine monosaccharide.



Supplementary Fig. 12. Base peak chromatogram derived from UPLC-MS/MS analyses of N-glycans released from E1/E2 glycoproteins of MAYV and glycan composition interpretation based on MS2 data. (A) Nine peaks were integrated (#1 to #9) and characterized as N-glycans composed by Nacetyl hexosamine (HexNAc); hexose (Hex) and fucose (Fuc) - HexNAc(2-5)Hex(5-8)Fuc(0-1). These glycan were proposed based on MS1 isotopic pattern (presence of double and triple charge forms in a variation of adducts) and respective carbohydrate unit compositions were proposed based on MS2 fragmentation pattern, as exemplified for peak #3 in (B). (B) MS2 (fragmentation) spectrum of m/z 808.8387, the most intense MS1 peak detected in the chromatographic peak #3. Through the interpretation of different MS1 peaks, according to the isotopic pattern for double and triple charge and also the variation of adducts (please see Supplementary Table 5 or details), the m/z 808.8387 was identified as [M+2H]2+, thus representing a glycan of 1,615.6597 Da composed by HexNAc(2)Hex(6). In the fragmentation spectrum, it was possible to recognize neutral losses of 162.05 (relative to hexose residues) and 203.07 (N-acetyl-hexosamine residues), in addition to fragments corresponding to procainamide derivatives (orange and blue sequence paths, respectively). Furthermore, MS2 double charge peaks fragmentation sequence (green) allowed for the visualization of losses of 81.02 (another way to verify the presence of the hexose units). For the other six chromatogram peaks (peaks #4 to #9), the same rational was used to investigate the fragmentation pattern and to attribute the respective compositions of the glycans (Figure 4D). For peaks #1 and #2, it composition was predicted mainly using MS1 data, due to the poor fragmentation in the MS2 data. All N-glycans compositions were confirmed using Glycoworkbench software and data was deposited at UniCarb-DB.

	CHIKV strain S27	SFV	RRV strain T48	SINDV	EEEV strain Florida 91- 469	VEEV strain TC-83
Capsid – disorder N- terminal (1-102)	53 %	56 %	63 %	43 %	33 %	37 %
Capsid - Structured domain (103-258)	89 %	90 %	90 %	65 %	63 %	64 %
E1	62 %	72 %	71 %	48 %	52 %	52 %
E2	56 %	67 %	62 %	37 %	41 %	38 %

Supplementary Table 1. Identity between structural proteins from MAYV and other alphaviruses.

Supplementary Table 2. Cryo-EM data collection and processing.

Magnification (x)	59,000
Voltage (kV)	300
Electron exposure (e- Å-2)	30
Nominal defocus (µm)	-1.5
Pixel size (Å)	1.1074
Symmetry imposed	lcosahedral
Number of micrographs	175840 (8792 movie stacks)
Initial number of particles images	79000
Number of particle images used in the final 3D reconstruction	40179
Map global resolution	4.4 (1⁄2-bit threshold) 4.3 (0.143 threshold)

Supplementary Table 3. Overall MAYV 3D atomic model quality

evaluated by MolProbity.

Structural composition	Structural composition							
Chains	12							
Residues	4004							
MolProbity statistics								
Bonds length (Å) (RMSD - # > 4 sigmas)	0.006 (0)							
Bond angles (°) (RMSD - #	1.070 (43)							
> 4 sigmas)								
MolProbity score	2.0							
Clash score	8.7							
Ramachandran plot (%)								
Outliers	0.1							
Allowed	11.4							
Favored	88.6							
Rotamer outliers (%)	0.59							
Cβ outliers (%)	0							
Model vs. Data								
CC (mask)	0.69							

Supplementary Table 4. List of alphaviruses PDB files and Cryo-EM maps used for comparative studies.

Alphavirus	PDB ID	EMD ID	Method	Resolution	Reference
сніку	6NK5, 6NK6	9393	Cryo-EM	~4.5 A	Basore,
					2019
SINDV (only	6IMM	9693	Cryo-EM	~3.5 A	Chen, 2018
E1 and E2)					
VEEV	3J0C	5275	Cryo-EM	~4.4 A	Zhang, 2011
EEEV	6MX4	9280	Cryo-EM	~4.4 A	Hasan, 2018

Supplementary Table 5. Summary of MS¹ analyses of N-glycans released from E1/E2 MAYV and identified by UPLC-MS/MS analyses. The nine chromatographic peaks are presented with featured MS¹ variations (double and triple charge and adducts). The glycan composition was proposed in combination with MS² data.

Peaks	RT (min)	Peak Area	[M+2H] ²⁺	[M+HK] ²⁺	[M+2K] ²⁺	[M+3H] ³⁺	[M+2HK] ³⁺	[M+2KH] ³⁺	[M+3K] ³⁺	Ma	Proposed Composition ^b
#1	17.1	14739	727.8114	746.7888	ND	ND	ND	ND	ND	1453.6066	HexNAc(2)Hex(5)
#2	19.1	25597	829.3501	848.3246	ND	ND	565.8864	578.5398	ND	1656.6851	HexNAc(3)Hex(5)
#3	20.5	298547	808.8383	827.8154	846.7937	ND	552.213	564.8646	ND	1615.6597	HexNAc(2)Hex(6)
#4	21.2	78873	930.8921	949.8696	968.8477	620.9288	633.5809	646.2322	658.8851	1859.7648	HexNAc(4)Hex(5)
#5	21.8	25492	1032.4332	1051.4114	ND	688.622	701.2741	713.9259	726.5785	2062.8460	HexNAc(5)Hex(5)
#6	22.6	60892	1003.9234	1022.9008	1041.8813	669.6158	682.2672	694.9192	707.5719	2005.8260	HexNAc(4)Hex(5)Fuc(1)
#7	23.2	27692	1105.4638	1124.4417	ND	737.3091	749.9607	762.6121	775.2670	2208.9063	HexNAc(5)Hex(5)Fuc(1)
#8	23.9	195538	889.8666	908.844	927.822	ND	606.2306	618.8828	631.5339	1777.7148	HexNAc(2)Hex(7)
#9	26.9	76399	970.8955	989.8735	1008.8516	ND	660.2494	672.9011	685.5527	1939.7717	HexNAc(2)Hex(8)

^aFor N-glycans exact mass, it was considered the average of the most intense MS¹ peaks ^bDerived from MS and MS² peak interpretation of N-glycans reacted with procainamide ND - not detected