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.

BBA - Biomembranes 1865 (2023) 184174

Contents lists available at ScienceDirect

BBA - Biomembranes

journal homepage: www.elsevier.com/locate/bbamem

Understanding how transmembrane domains regulate interactions between human BST-2 and the SARS-CoV-2 accessory protein ORF7a

Madison M. Mann [a,1], Min-Kang Hsieh [b,1], James D. Tang [a], William S. Hart [a], Matthew J. Lazzara [a], Jeffery B. Klauda [b,c,*], Bryan W. Berger [a,d,**]

[a] Department of Chemical Engineering, University of Virginia, United States of America
[b] Department of Chemical and Biomolecular Engineering, University of Maryland College Park, United States of America
[c] Institute for Physical Science and Technology, Biophysics Program, University of Maryland College Park, United States of America
[d] Department of Biomedical Engineering, University of Virginia, United States of America

ARTICLE INFO

Keywords:
Transmembrane domain
RNA virus
Glycosylation
Molecular dynamics
Cellular immune response
SARS-CoV-2

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID, replicates at intracellular membranes. Bone marrow stromal antigen 2 (BST-2; tetherin) is an antiviral response protein that inhibits transport of viral particles after budding within infected cells. RNA viruses such as SARS-CoV-2 use various strategies to disable BST-2, including use of transmembrane 'accessory' proteins that interfere with BST-2 oligomerization. ORF7a is a small, transmembrane protein present in SARS-CoV-2 shown previously to alter BST-2 glycosylation and function. In this study, we investigated the structural basis for BST-2 ORF7a interactions, with a particular focus on transmembrane and juxtamembrane interactions. Our results indicate that transmembrane domains play an important role in BST-2 ORF7a interactions and mutations to the transmembrane domain of BST-2 can alter these interactions, particularly single-nucleotide polymorphisms in BST-2 that result in mutations such as I28S. Using molecular dynamics simulations, we identified specific interfaces and interactions between BST-2 and ORF7a to develop a structural basis for the transmembrane interactions. Differences in glycosylation are observed for BST-2 transmembrane mutants interacting with ORF7a, consistent with the idea that transmembrane domains play a key role in their heterooligomerization. Overall, our results indicate that ORF7a transmembrane domain interactions play a key role along with extracellular and juxtamembrane domains in modulating BST-2 function.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified to be the causative agent of a fatal respiratory illness that emerged in the city of Wuhan, China at the end of 2019 [1]. Since then, the outbreak of COVID-19, the disease caused by SARS-CoV-2, has infected >36.5 million individuals and claimed over one million lives worldwide [2]. The pathogenesis of SARS-CoV-2 infection in humans has been well documented; SARS-CoV-2 enters the respiratory tract through the binding of the viral structural spike (S) protein to the angiotensin-converting enzyme 2 (ACE2) receptor that is present on the surface of host cells [3–7]. Its genome organization is shared with other betacoronaviruses, and has a 79 % and 50 % genome sequence identity with SARS-CoV and Middle Eastern Respiratory Syndrome (MERS)-CoV, respectively [8]. In addition to the functional open reading frames (ORFs) such as the replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M) and nucleocapsid (N), the SARS-CoV-2 genome also contains several putative ORFs encoding accessory and nonstructural proteins interspersed between structural genes that promote SARS-CoV-2 virulence and replication [9].

One of these accessory proteins, ORF7a, is a type-I transmembrane protein comprised of a N-terminal signal peptide, an ectodomain, a transmembrane region, and a cytoplasmic di-lysine motif (KRKTE) for ER localization [10]. This protein shares 85.2 % identity and 95.9 %

* Correspondence to: J.B. Klauda, Department of Chemical and Biomolecular Engineering, University of Maryland, College Park, MD 20742, United States of America.
** Correspondence to: B. W. Berger, Department of Chemical Engineering, University of Virginia, Charlottesville, VA 22903, United States of America.
E-mail addresses: jbklauda@umd.edu (J.B. Klauda), bryan.berger@virginia.edu (B.W. Berger).
[1] These authors contributed equally

https://doi.org/10.1016/j.bbamem.2023.184174
Received 20 February 2023; Received in revised form 12 April 2023; Accepted 10 May 2023
Available online 19 May 2023

sequence similarity with SARS-CoV ORF7a [11]. It was previously shown that SARS-CoV-2 ORF7a has the ability to antagonize the interferon (IFN-I) response by suppressing nuclear translocation of phosphorylated signal transducers and activators of transcription 1 (STAT1) during IFN-I signaling [12]. Similarly, Cao et al. reported that upon IFN-α stimulation, ORF7a suppressed STAT2 but not STAT1 phosphorylation in a dose-dependent manner [13], suggesting that ORF7a does not effectively block nuclear transport, which may explain the increase in sensitivity of SARS-CoV-2 to IFN pretreatment. This inhibition of STAT2 phosphorylation may be attributed to ORF7a polyubiquination which subsequently enhances IFN-I antagonism. A recent study has shown that the SARS-CoV-2 ORF7a ectodomain binds to CD14+ monocytes in human peripheral blood with high efficiency, leading to decreased antigen-presenting ability and inducing a dramatic expression of proinflammatory cytokines by human immune cells [14]. These cytokines, including IL-1β, IL-6, IL-8, and TNF are dictated by NF-κB, and linked to cytokine release syndrome, implicating the positive associations with severe disease outcome [15]. A previous study demonstrated that ORF7a of SARS-CoV-2 can activate NF-κB function and increases these proinflammatory cytokine expressions [16]. Since IL-1β is in part responsible for the cytokine storm by SARS-CoV and MERS-CoV [17], ORF7a may play a significant role in the clinical severity of COVID-19, however, further studies are needed to characterize the molecular details and biological functions related to how ORF7a initiates viral pathogenesis.

Bone marrow stromal antigen 2 (BST-2; also known as CD317 or tetherin) is an IFN-inducible gene. As viral infections trigger expression of IFN, BST-2 expression is increased due to IFN responsive regulatory elements in BST-2's promoter region [18,19].

BST-2 is known for its ability to block the egress of enveloped viruses from infected cells. In the absence of an antagonist, BST-2 is incorporated into budding viral particles and homodimerizes with other resident BST-2 molecules present on the cell surface [20], thereby 'tethering' the viral particle to the host plasma membrane and restricting cell-to-cell virus spread. BST-2's antiviral function was first identified for HIV-1 [21], where it was initially thought to restrict HIV-1 virus release. Further studies revealed that BST-2 can also impair the release of various enveloped viruses belonging to *Retroviridae*: HIV type 2 (HIV-2) such as simian immunodeficiency viruses (SIVs) [22], equine infectious anemia virus (EIAV) [23], feline immunodeficiency virus (FIV) [24], prototype foamy virus (PFV), and Mason-Pfizer monkey virus (MPMV) [25].

The ability of BST-2 to impair the release of a broad spectrum of viruses would suggest that various viruses have evolved a way to antagonize the cellular restriction factor(s) of BST-2. For instance, the widely studied Vpu accessory protein of HIV-1 is known to physically interact with the transmembrane (TM) domain of BST-2 [26]. Infection of cells by the virus results in the internalization of BST-2 from the plasma membrane through a clathrin-dependent endocytosis pathway compared to ΔVpu viruses [27]. This suggests that Vpu may affect resupply or surface delivery of BST-2 [27]. Vpu was also found to reduce total cellular levels of endogenous as well as exogenously expressed BST-2 [28], although the mechanism of cell surface BST-2 downregulation is unknown [29]. Interestingly, a recent analysis of the TM domains of human and rhesus BST-2 revealed a number of differences including deletions and non-synonymous mutations that affect Vpu sensitivity [30–32]. Previous studies indicated that a deletion of a GI amino acid motif present in human BST-2, but absent in non-human BST-2 variants [26,33], combined with mutation of T45I, resulted in complete loss of Vpu sensitivity [31]. Similarly, a I48T mutation induced partial sensitivity of rhesus BST-2 to Vpu [30], suggesting that the BST-2 TM domain contains the determinants responsible for the species-specific sensitivity to Vpu.

Biochemical analyses reveal that BST-2 is post-translationally modified by N-linked glycosylation, and forms stable cysteine-linked homodimers [34]. In a recent study, BST-2 was found to significantly block human coronavirus 229E progeny virus release [35], indicating that BST-2 is also capable of inhibiting virus budding at intracellular membranes. Similarly, SARS-CoV ORF7a was demonstrated to preferentially bind to unglycosylated BST-2, suggesting that the blocking of glycosylation by ORF7a is directly responsible for the antagonism of BST-2 [36]. Several studies have implicated BST-2 dimerization as essential for inhibition of HIV-1 release [20,37]. A recent investigation into the role of BST-2 dimer formation in the release of viral particles showed the mutation of multiple cysteine residues (C53A, C63A, and C91A) prevented both dimer formation and BST-2 function [20,37]. However, single and double substitutions had no effect, suggesting that promiscuous dimer formation is important for BST-2 anti-viral activity [20,37]. Thus, the dimerization interface of BST-2 and SARS-CoV-2 may play a unique role in viral release and BST-2 antagonism.

The present study focuses on identifying specific interactions between BST-2 and SARS-CoV-2 ORF7a, with particular emphasis on the TM domain interactions between BST-2 and ORF7a. Based on dimerization assays in cell membranes using the AraTM and DN-AraTM systems, we demonstrate the sensitivity of homodimerization to varying BST-2 domain truncations as well as point mutations within the TM domain. Similarly, we observed preferential heterodimeric interactions of the TM domains within ORF7a and BST-2 as well as the I28T BST-2. Our results also show that in expression of the full length constructs, SARS-CoV-2 ORF7a disrupts BST-2 glycosylation for all variants studied which has been previously linked to its viral restriction activity [36]. Molecular dynamics (MD) simulations have revealed the BST-2 and ORF7a protein-protein interface and revealed the details of the heterodimerizations. By surveying the heterodimeric configurations, the MD simulations provide molecular insights that the I28S BST-2 variant has enhanced its helix-helix hydrophobic interaction, compared to the WT, to support our observation from *Ara-TM and DN-AraTM* experiments.

2. Results

2.1. AraTM and DN-AraTM dimerization assays

The Ara-C based transcriptional reporter assay, AraTM [38], was used to study the effects of BST-2 domains (cytoplasmic, transmembrane, and extracellular) on dimerization. In this assay (Fig. 1A), varying domains of BST-2 are fused at the N-terminus to maltose binding protein (MBP) which allows for placement in the bacterial inner membrane. The C-terminus is fused to the AraC transcriptional factor which induces expression of eGFP through activation of the P_{BAD} promotor upon dimerization. eGFP fluorescence is quantified from culture and directly correlates with the extent of dimerization of the inserted BST-2 domains in the chimera [38]. Note that BST-2 is a Type II integral membrane protein, whereas ORF7a is a Type I integral membrane protein; constructs were cloned into AraTM assays to reflect the appropriate Type I and Type II configurations for each construct tested.

The MBP-BST2-AraC constructs containing various forms of the BST-2 domains were transformed into the AraC-deficient *E. coli* strain SB1676 (Fig. 2) containing the engineered *araBAD::GFP* reporter. Interestingly, domains containing the first 52 amino acids, which include the cytosolic domain and transmembrane domain of BST-2, exhibit the highest dimerization in the AraTM assay. Moreover, addition of extracellular domain regions to the TM and juxtamembrane (JM) region of BST-2 (corresponding to amino acids 80–147) do not cause a substantial increase in observed homodimerization, indicating that the primary signal observed in the AraTM assay is due to the TM and JM interactions. Previous crystal structures show that BST-2 forms a continuous α-helix through the cytoplasmic domain that dimerizes through interactions at the C-terminus, consistent with our observed results [64,65]. While prior structures of ORF7a and BST-2 indicate that potential regions within the soluble domains are also responsible for oligomerization, our primary focus was on the role of TM and juxtamembrane interactions. Thus, the cytoplasmic tail and TM domain

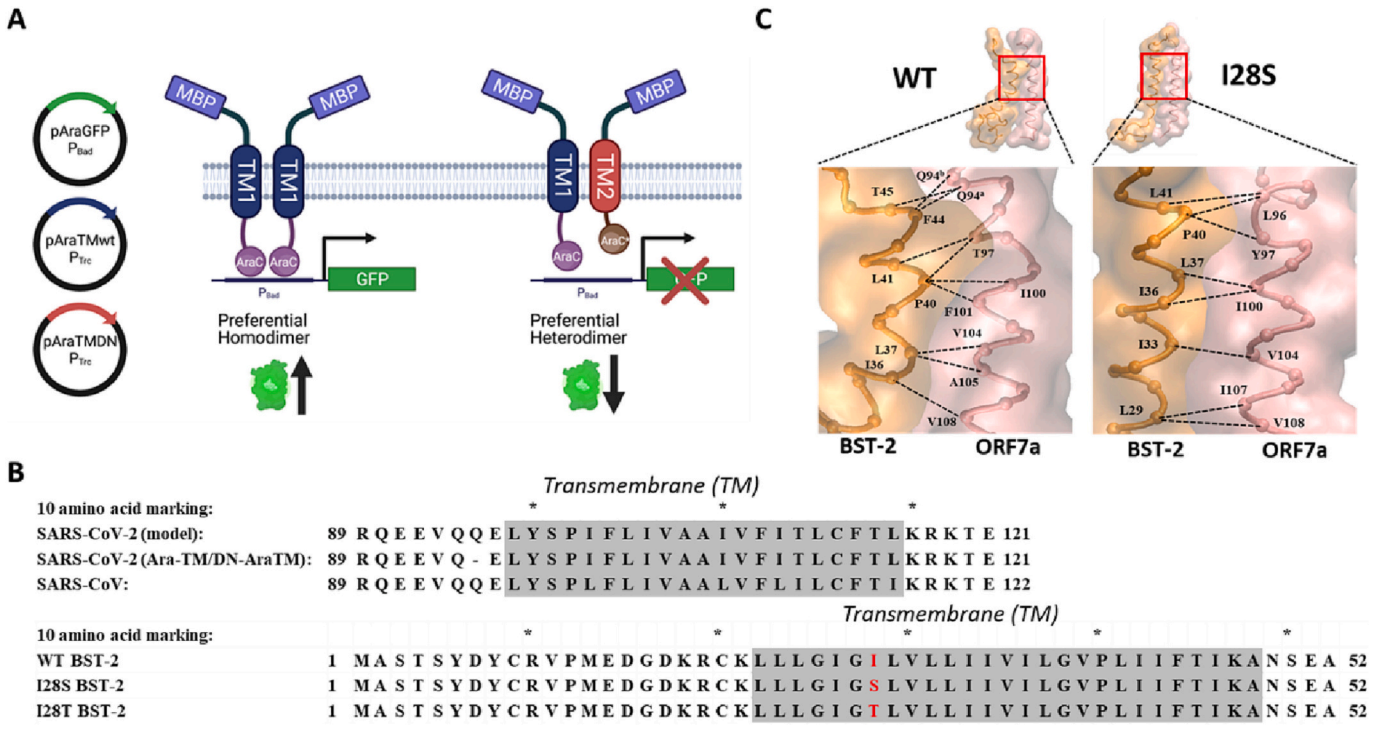

Fig. 1. (**A**) AraTM and DN-AraTM assay. In the AraTM assay, pAraTMwt expresses the full-length AraC fused to a target sequence under control of the PTrc promoter. This plasmid is co-expressed with the pAraGFP plasmid that contains an eGFP reporter gene under control of the PBAD promoter. In the DN-AraTM assay, addition of the pAraTMDN plasmid expresses the truncated AraC (AraC*) fused to a competitor sequence under control of the PTrc promoter. Co-expression of both constructs with pARAGFP in *E. coli* leads to a distribution of interactions between AraC and AraC* fusions, and the relative affinity for homo- versus heterooligomerization can be independently quantified in terms of eGFP expression; reduced GFP expression is indicative of preferential heterooligomerization, whereas increased GFP expression is indicative of preferential homooligomerization. **(B)** Truncated amino acid sequences of SARS-Cov-2 ORF7a and BST-2 domains used for the current work. **(C)** Key residue contacting pairs of both WT and I28S heterodimerization revealed from MD simulations.

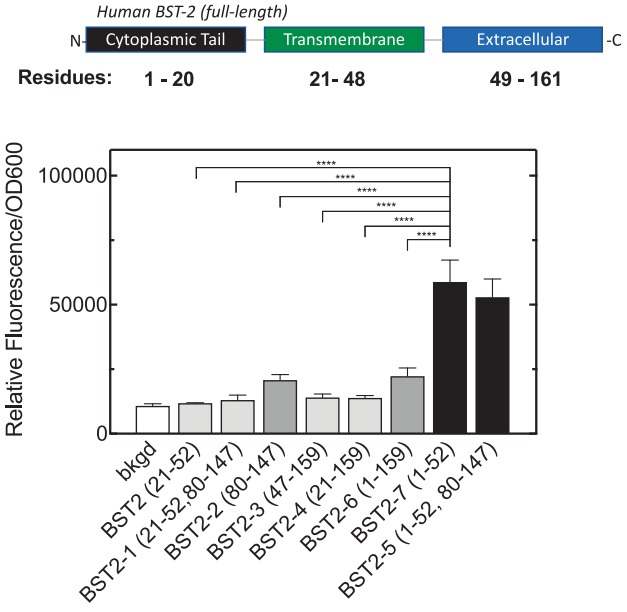

Fig. 2. AraTM homodimerization assay for varying BST-2 domains. Ratios of GFP fluorescence intensity vs. cell concentration (OD_{600}) for each construct are compared with the background (bkgd). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold plasmid that does not express the MBP-AraC-construct. (Upper) Representation of full-length human BST-2 protein and residue numbers used for all BST-2 domain chimeras (Accession Number Q10589). ($N = 5$, **** represents $p < 0.0001$).

containing BST-2 chimera (BST2–7) was used in the proceeding heterodimer assays.

As discussed previously, BST-2 exhibits evidence of positive, adaptive selection in response to exposure of viral pathogens, and therefore there are a wide range of identified variations in the TM sequence for human BST-2. To investigate the effects of identified sequence variants available in dbSNPs [66], mutants of the cytoplasmic-TM domain containing BST-2 chimera (BST2–7) were created. To assess self-assembly and heterodimerization interactions of the cytoplasmic-TM domain containing wild type (WT) BST-2 chimera (BST2–7) as well as BST2–7 SNP mutants I28S and I28T with the TM domain of the ORF7a accessory protein from SARS-CoV-2 (C19), a dominant-negative AraC-based transcriptional reporter assay (DN-AraTM) was used in addition to the homodimer AraTM assay [39]. The DN-AraTM assay measures the relative affinity of two TM-containing domains to heterodimerize and enables relative comparisons to TM-containing domain homodimerization. This process is illustrated in Fig. 1A, where a wild-type AraC chimera containing the TM and JM truncated form of BST-2 (BST2–7) is co-expressed with an inactivated, dominant-negative form (AraC*) containing ORF7a or BST-2. When these chimeras are co-expressed from unique plasmids (pAraTMwt and pAraTMDN), the AraC*-containing ORF7a or BST-2 chimeras act as competitors to BST-2 self-association, thereby decreasing the GFP signal.

As expected, Fig. 3 shows co-expression of WT BST-2 in both wild-type and dominant-negative AraC constructs causes a significant decrease in GFP signal, confirming the strong self-association seen in the AraTM assay (Fig. 2). While BST-2 WT and I28T exhibit this self-association, co-expression of these proteins with the AraC*-containing ORF7a also showed a significant reduction in GFP indicating dimerization competition with the accessory protein. However, the I28S mutant does not show self-association in the DN-AraTM assay. While

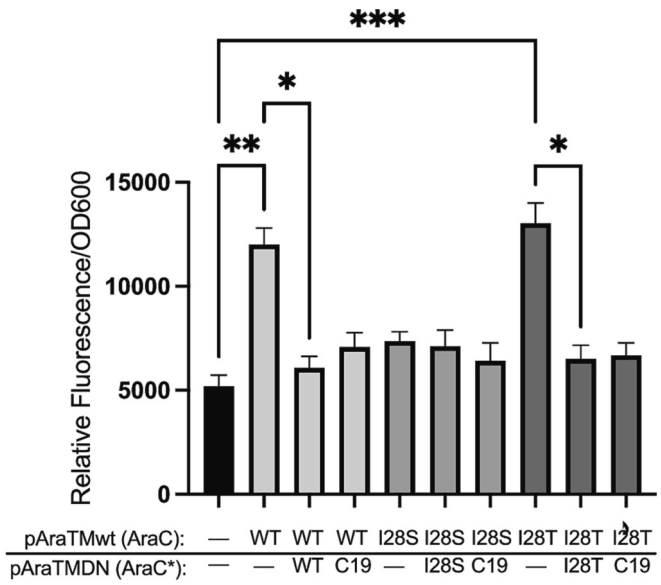

Fig. 3. DN-AraTM dimerization assay for BST-2 SNPs with SARS-CoV-2 ORF7a. Ratios of GFP fluorescence intensity vs. cell concentration (OD_{600}) for each construct are compared with the background (denoted as – in both the AraC and AraC* containing construct table). In the background sample, *E. coli* was transformed with the empty pTrcRSF scaffold for pAraTMwt and pTrc99a scaffold for pAraTMDN plasmids that do not express the MBP-AraC or MBP-AraC* constructs. ($N = 5$, ***,**, and * represent $p < 0.001$, $p < 0.01$, and $p < 0.05$ respectively).

heterodimer competition with ORF7a cannot be studied in this system upon this loss of self-association of the I28S variant, this lack of homodimerization upon a single nucleotide substitution in the transmembrane domain of BST-2 is interesting to note; the lack of self-association of BST-2 I28S mutant relative to BST-2 WT renders it less susceptible to heterodimerization. Specifically, while promiscuous in mechanism, dimerization of BST-2 has been previously linked to viral response function [20,37]. Collectively, these results are consistent with heterodimerization with the transmembrane domain of SARS-CoV-2 ORF7a being significant relative to BST-2 WT and I28T homodimerization.

2.2. Coarse-grained molecular dynamics simulations

Further evidence to support the experimental observations of differences in ORF7a-BST-2 interactions was provided through coarse-grained molecular dynamics (CG MD) simulations. ORF7a-BST-2 heterodimers were separated initially in a POPC bilayer, and CG MD simulations with MARTINI force field were performed (Fig. S1). Within 100 ns, the helical dimers associated from their initial separation and remained tightly packed for the 5 μs simulation. The contact maps of the WT heterodimer, calculated by the distances of inter-residue backbone (BB) units between two monomers (Fig. S2A–C) and the end snapshots (Fig. S2D–F), show slight variations between replicas but consistently show interacting transmembrane (TM) and cytoplasmic (CYTO) domains. Similarly, the BB contact maps between the ORF7a and the I28S BST-2 (Fig. S3A–C) as well as the end snapshots (Fig. S3D–F) show a consistent interaction in the TM and CYTO domains. Based on these CG-MD simulations, some slight variations in the dimer interactions exist, but examples of complete dissociation are lacking. Thus, heterodimerization is consistently observed over multiple interactions of specific TM and CYTO interfaces of the ORF7a and BST-2 in both WT and I28S.

2.3. All-atom MD and clustering of dimer structures

Since the MARTINI FF used in this work is known to be in favor of residue affinity resulting in an overestimation of protein-protein contacts [67], MD simulations with an all-atom representation were performed to refine the modeled CG-based structure [68]. The equilibrated AA trajectory data were prepared by dropping the first 200 ns of each replica and then combining three heterodimeric replicas for both WT and the I28S sets. The approach to determining the hyperparameters for heterodimeric clustering is described in previous work [63]. The chosen hyperparameters for both dimers were optimal based on a cross-validated reconstruction loss, shown in Table S1. During training, the data were split into a train/validation set with a fraction of 0.8 for the training set and 0.2 for the validation set. The total loss, cross-entropy loss, and reconstruction loss decreased during 100 epochs and remained at stable values for both dimerization data (Figs. S4–S5). Fig. S6 shows the reconstruction loss as a function of the latent space dimension for both systems, indicating higher dimensions result in better reconstruction loss. The latent space dimension of 5 was chosen for minimizing the reconstruction loss of the validation set for both systems. The cluster size was determined using the threshold scheme that has membership possibility higher than 0.95 cut-off. A total of 9 and 13 clusters were identified for the dimerization of ORF7a with the WT and the I28S mutant BST-2, respectively. The t-distributed stochastic neighborhood embedding (T-SNE) [69] was applied to visualize the 5D latent space by transforming the five-dimensional embedding into two dimensions (Fig. S7), showing the clusters are highly separated on this landscape with membership possibility over 0.75 for both systems. The distribution of data (Tables S2–S3) shows the even fractions among the WT clusters from single or two resource replicas for individual clusters while higher fractions in two clusters than others for the I28S clusters only from single replica for each cluster. 91 % and 96 % of data points have been identified for the WT and I28S clusters, respectively, with the membership probability >0.75. This GMAVE clustering approach demonstrates its capability to classify the dimer configurations from AA MD dimerization simulations and sort data across different independent runs by using the C_α distances between two peptides only. Since the intra-peptide C_α distances and other detail sidechain structural information have not been included, the structural variation within a cluster were expected. These GMAVE clusters were taken to further build our structural model for the protein-protein interface, where we are looking for mainly TM helix-helix packing of the heterodimeric system. In examining the contact map of these clusters (Figs. S8–S9), several have a similar profile of contacting pairs since this GMAVE was not trained using criteria of contacts between two peptides. Therefore, we further classified the GMVAE clusters into groups by using the contacting occupancy within a whole contact map.

2.3.1. Contact Map

Figs. 4 and S8 show the contact maps (frequency of contacting distance < 9 Å) of the GMVAE identified clusters for the ORF7a/WT BST-2 dimerization, showing the contacting residue pairs occupy 0.21– 2.3 % of the whole contact map (Table S4) for the identified GMVAE clusters. The contacting occupancies for each division were listed as well. These clusters were further grouped by the distribution of their regional contact occupancy. Specifically, the regional contact domains on BST-2 were defined by ranges of residue positions: EC domain (residue 49–52), top of the TM domain (residue 40–48), middle of the TM domain (residue 31–39), bottom of the TM domain (residue 22–30) and CYTO domain (residue 1–21). Figs. 4 and S10 also show the corresponding superpositions of the cluster snapshots which were generated from the top 50 secondary configurations of peptides referring to the center of individual clusters (sorting by argmax of RMSD) for visualizing the dimeric structures. The centers of structural configuration represent the average C_α coordinate over the whole cluster. Table 1 summarizes the grouping classification based on domain contacts for dimer clusters of

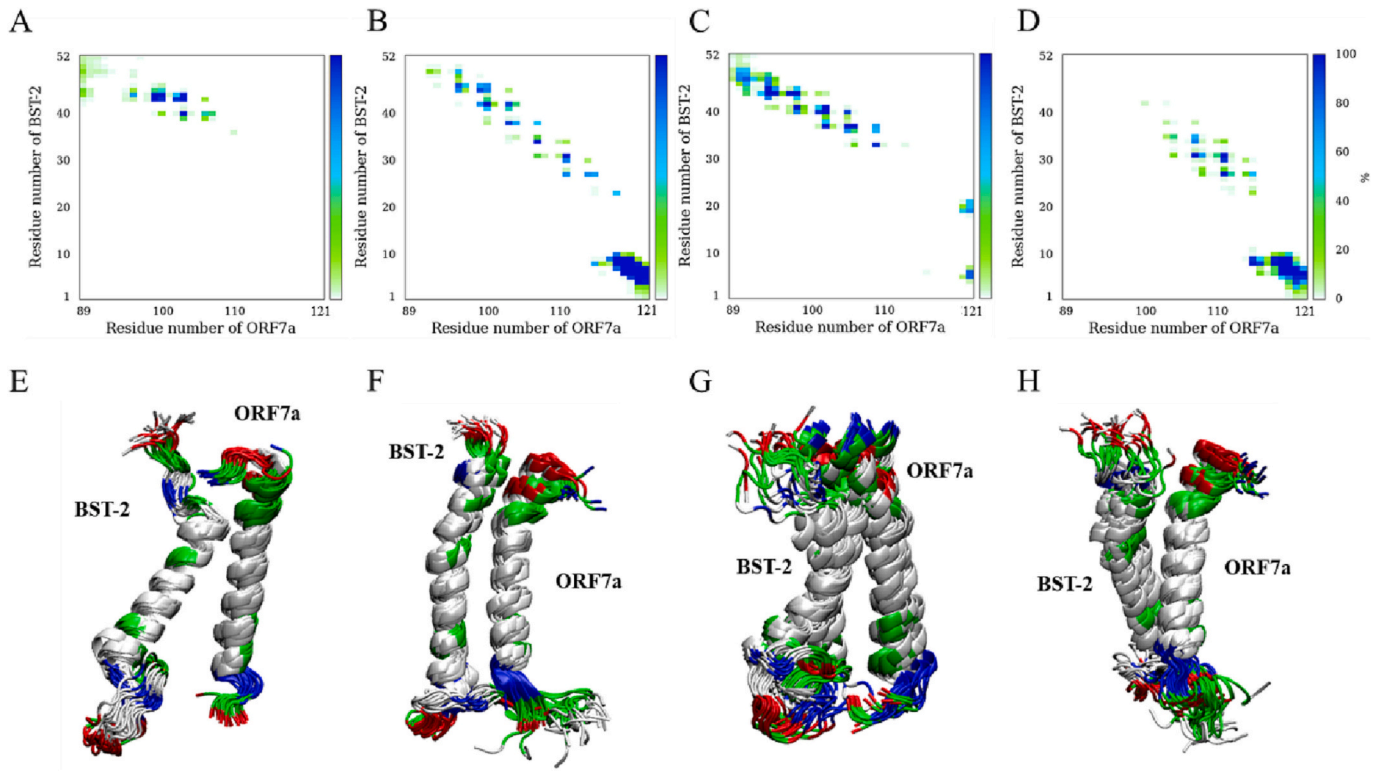

Fig. 4. Results from AA-MD for ORF7a dimerization with WT BST-2.
(**A**–**D**): Contact maps of four groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance < 9 Å. (**E**–**H**): Snapshots of dimer groups made by the superimposed top 50 configurations from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Grouping of the GMVAE identified dimer clusters.

Group[a]	GMVAE clusters	Contacting domains	% of data
WT Group 1	Cluster 1, Cluster 3	Top TM	19.66
WT Group 2	Cluster 2, Cluster 6	Top and middle TM, CYTO	22.98
WT Group 3	Cluster 7, Cluster 8	Top and middle TM	21.57
WT Group 4	Cluster 4, Cluster 5	Middle and bottom TM, CYTO	17.36
I28S Group 1	Cluster 1, Cluster10	Middle and bottom TM	13.29
I28S Group 2	Cluster 3	Middle and bottom TM, CYTO	6.83
I28S Group 3	Cluster 2, Cluster 4, Cluster 6, Cluster 13	Top and middle TM, CYTO	32.23
I28S Group 4	Cluster 5, Cluster 9	Middle TM	9.99
I28S Group 5	Cluster 12	EC, top TM	12.77

[a] WT groups classify the GMVAE clusters from ORF7a/WT BST-2 dimerization while I28S groups classify the GMVAE clusters from ORF7a/I28S BST-2 dimerization.

ORF7a/WT BST-2 as well as ORF7a/I28S BST-2. Note that the ORF7a/WT BST-2 cluster with an overall occupancy of <0.3 %, i.e., Cluster 9, was not classified due to minimal contacts of the dimer.

The same grouping approach for ORF7a/I28S BST-2 were applied after the dimer configurations were clustered by GMAVE. Figs. 5 and S9 show the contact maps of the ORF7a/I28S BST-2 dimer clusters with the contacting residue pair occupies 0.01–1.1 % of whole contact map (Table S5) of the identified clusters. Note that Cluster 7, 8, and 11 were not included in the grouping due to a low occupancy of their contact maps (Table 1). Therefore, a total of 4 and 5 structural groups with various contacting behaviors were identified for WT and I28S BST-2 respectively. The variation in residue contacts indicates the heterodimeric dimerization of BST-2 and ORF7a has multiple states of association.

2.3.2. *Helix-helix C_α backbone packing*

To examine the helical packing of each group, a tight packing form (the top 5000 sampling from the center of each group) was analyzed and represent visually for individual groups. To further examine the interaction of residues, the whole configurations (with a membership probability of >0.75) will be included to obtain the energetic behavior of individual groups.

To construct the protein-protein interface structural model of these heterodimeric type I transmembrane proteins, the lateral association of helix-helix crossing angle (Ω), packing distance ($D_{Helix\text{-}Helix}$) [70], and packing motifs were examined using the C_α coordinates of individual clusters. Then, residue-residue interactions, including sidechain contacts, were further assessed by residue binding energetics. The above clustering of atomic configurations enhancing their regional C_α contacts were used to assess the lateral helix-helix packing. Figs. S12 and S13 demonstrate the distribution of the Ω and $D_{Helix\text{-}Helix}$ for each cluster in which the top 5000 samples of each group (and cluster) were taken and the helices of each configuration were defined as the largest continuously helical domain (including 3-helix, α-helix, and 5-helix structures) by using DSSP-2.2.0 [71]. Tables 2, S6, and S7 summarizes the size (L_{Helix}, number of residues) of the helices, Ω and $D_{Helix\text{-}Helix}$ for each heterodimeric cluster. In general, a single helical domain per peptide was found, but a short second helix domain was found in I28S Group 4. A

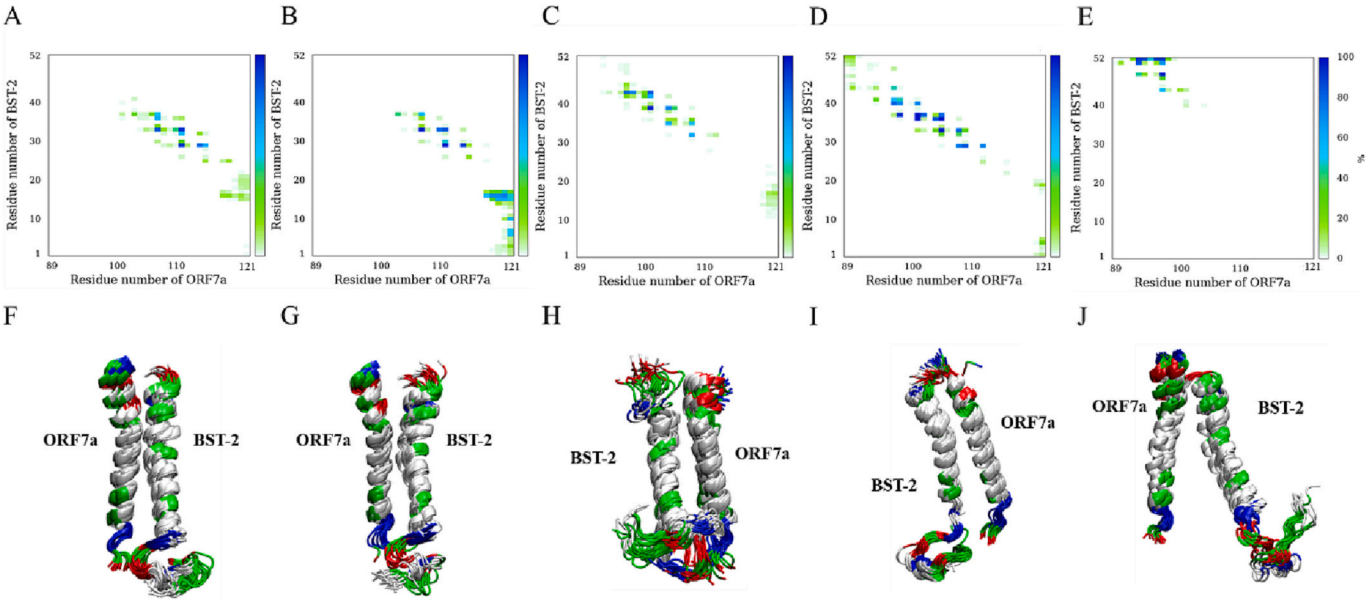

Fig. 5. Results from AA-MD for ORF7a dimerization with mutant I28S BST-2.
(**A**–**E**): Contact maps of five groupings calculated by using the top 5000 configurations from the center of GMVAE clustering groups. The color bars represent the percentage of contacting distance < 9 Å. (**F**–**J**): Snapshots of dimer groups made by the superimposed top 50 configurations from the center of the group. Nonpolar, polar, acidic, and basic residues are colored white, green, red, and blue, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Length, distance, and crossing angle of helices. The mean and standard error have been provided.

	BST-2 L_{Helix}[a]	ORF7a L_{helix}	$D_{helix\text{-}helix}$ (Å)	Ω (deg.)
WT Group 1	21.5 ± 0.02	26.43 ± 0.03	11.1 ± 0.03	44.85 ± 0.39
WT Group 2	26.37 ± 0.03	25.9 ± 0.02	11.67 ± 0.02	−13.07 ± 0.08
WT Group 3	20.57 ± 0.04	27.36 ± 0.04	9.38 ± 0.02	−35.01 ± 0.13
WT Group 4	25.21 ± 0.05	25.28 ± 0.03	10.56 ± 0.01	−44.04 ± 0.09
I28S Group 1	29.18 ± 0.04	27.51 ± 0.03	11.24 ± 0.01	−37.41 ± 0.06
I28S Group 2	29.89 ± 0.03	27.9 ± 0.03	11.98 ± 0.01	−38.5 ± 0.06
I28S Group 3	21.38 ± 0.02	24.36 ± 0.04	11.97 ± 0.02	16.49 ± 0.15
I28S Group 4	23.1 ± 0.02	23.82 ± 0.03	11.13 ± 0.01	−21.53 ± 0.06
I28S Group 5	28.8 ± 0.06	28.17 ± 0.03	12.71 ± 0.04	−33.05 ± 0.11

[a] L_{Helix} is the number of the residue in the DSSP identified biggest helix domain.

short β-β contact has been found in the CYTO domain in WT Group 2. A variation of Ω distributions indicate various orientations of both helices, and their associations could be either right-handed (a negative crossing angle) or left-handed (a positive crossing angle) crossing, though the right-handed crossing is the majority. Fig. S13B shows a distinguishable $D_{Helix\text{-}Helix}$ across WT groups while Fig. S13D demonstrates the similarity of the $D_{Helix\text{-}Helix}$ distributions among the I28S groups. The results show helix packing variations between groups demonstrating what appears to be high entropic heterodimeric dimerization of BST-2 and ORF7a. It is worth noting that in Group 3 with the shortest $D_{Helix\text{-}Helix}$, BST-2 has a short L_{Helix} compared to other groups while the ORF7a L_{Helix} is longest among all WT groups, resulting in ~7 residue differences between the helixes in the WT packing. Also, the average L_{Helix} of BST-2 over I28S groups is about 3 residues longer, compared to the WT BST-2 while the average L_{Helix} of ORF7a is the same in both heterodimers.

2.3.3. *Binding energetics and heli-packing motifs*

Tables S8 and S9 list the residue pairing with a distance less than the cutoff (i.e., 9 Å) indicating that the WT heterodimerization is more likely driven by the interactions within the CYTO and EC domains, compared to the I28S heterodimerization. Examination of the contacts within the TM domain demonstrate that various pairing residues in the helix-packing region neither correlate to the length of the L_{Helix} nor the strength of interactions in the CYTO and EC domains, indicating the TM contacts result from a complex mechanism of protein-protein and protein-lipid interactions. The hydrophobic interaction of these TM contacts is the main contributor to helix-helix associations within the membrane. These contacts can vary from helix-helix packing that spans the TM region to packing that is limited to shorter ranges. Extended TM contact exists for the WT Group 2 and 3 (Table S8) and I28S Group 4 (Table S9), compared to more localized regions with other classified groups. WT Group 3 has a tighter packing, having a shorter distance average among the TM contacting pairs, compared to WT Group 2, as expected due to the shortest $D_{Helix\text{-}Helix}$ of WT Group 3. According to the C_α pairing, the helix-helix packing motif could be identified as the "knobs-into-holes" heptad [63] repeats within the closest helix packing configurations (Fig. 1C), but it was not clear for other groups due to a short contacting region. The WT Group 3 helices associate through $I^{36}L^{37}xxP^{40}L^{41}xxF^{44}T^{45}$ and $V^{108}xxA^{105}V^{104}xxF^{101}I^{100}xxY^{97}xxQ^{94b}Q^{94a}$ of BST-2 and ORF7a respectively and form a packing complex. On the other hand, the I28S Group 4 TM helices pack with heptad repeats through $L^{29}xxxI^{33}xxI^{36}L^{37}xxP^{40}L^{41}$ and $V^{108}I^{107}xxV^{104}xxxI^{100}xxY^{97}L^{96}$ of BST-2 and ORF7a respectively, and form TM heterodimer structural complexes. Both involve stacking interactions among F^{44}, Y^{97}, and Q^{94a} at the EC/TM interface. Figs. 6, 7, S14, and S15 visualize the helix-helix contacting motif including side-chain contacts identified based on a contact distance of <3 Å in the center configuration of individual clusters. The contacting pairs in the WT Group 3 located across the top and middle of the TM domain while the contacting motif in the I28S Group 4 dimerization shifts closer to the cytoplasmic side across the middle and bottom of the TM domain, indicating the I28S mutation initiates helix-helix lateral association close to the mutated position (at the bottom of the TM domain).

Figs. 8, S16, and S17 show the energetics of dimerization contributions from individual residues, supporting the observation in the C_α pairing. The energetics also revealed the binding enthalpic contributions of mainly sidechain contacts quantitatively.

These TM helix pairings within helix-packing motif are stabilized by

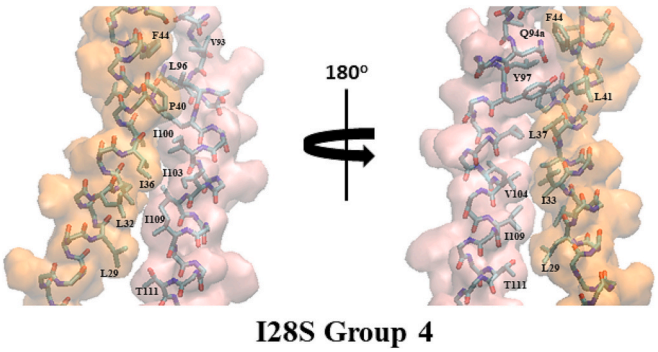

Fig. 6. Molecular structure of the helix-packing motif with detailed side chain contacts in the WT Group 3 Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

van der Waals interactions of large side chain residues (valine, leucine, and isoleucine) which contribute to residue binding energetics. Moreover, energetically strong interacting residues have been found in the CYTO domain of loose packing groups. In WT Group 2 and 4 (Fig. S16), energetic residues in the CYTO domains of WT BST-2 (S^5, Y^6, D^7, Y^8, and C^9) interact with the CYTO residues of ORF7a (F^{114}, T^{115}, K^{117}, R^{118}, K^{119}, and T^{120}). On the other hand, in I28S Group2 (Fig. S11), the different interacting residues in the CYTO domain of I28S BST-2 (D^7, R^{10}, E^{14}, D^{15}, G^{16}, and D^{17}) interact with the same region of the ORF7a CYTO domain (K^{117}, K^{119}, T^{120}, and E^{121}).

In the most extended packing conformations (WT Group 3 and I28S Group 4), E^{121} (ORF7a) in the CYTO domain forms a salt bridge with R^{19} (Fig. 9A) or K^{21} (Fig. 9B) in the WT with the formation probability of 0.83 and 0.76, respectively; E^{121} (ORF7a) also forms a salt bridge with R^{19} (Fig. 9C) in the I28S mutant with the formation probability of 1.10. Moreover, the EC domain can dimerize with the formation of varied contacts. In WT Group 3, either a salt bridge formed by E^{91} (ORF7a) and K^{47} (BST-2) (Fig. 10A) with the formation probability of 0.21 or hydrogen bond between Q^{94a} (ORF7a) and K^{47} (BST-2) (Fig. 10B) with the formation probability of 0.12 promotes an amino-π interaction between Q^{94a} (ORF7a) and F^{44} (BST-2) and a π-π stacking between Y^{97} (ORF7a) and F^{44} (BST-2) with the formation probability of 0.19 and 0.15, respectively. On the other hand, in I28S Group 4, the contacting complex is formed by Q^{94a}, L^{96}, Y^{97} (ORF7a) and F^{44} (BST-2) (Fig. 10C) where Q^{94a} (ORF7a) and F^{44} (BST-2) form an amino-π stacking with its formation probability of 0.22 as Q^{94a} and Y^{97} form a backbone hydrogen bonding. L^{96} (ORF7a) has a hydrophobic interaction in the complex which contributed to its residual energy (Fig. 8).

These CYTO and EC interactions involving polar/charged residues (i.e., glutamine, glutamic acid, lysine) can stabilize protein-protein binding by creating strong hydrogen bonding or salt-bridges. Moreover, these residues present within a TM helix drive a highly stable helix-helix association (I28S Group 4) [64] and form a stacking interaction in a helix-packing complex even enhancing the stability of both associations [65]. It has also been observed that the polar residues contribute to drive stable heterodimerization, but they also interrupt the helix packing pattern. Interestingly, although dimerization profiles vary in the BST-2 WT and I28S mutant, it is consistently observed that the contacting region of the I28S dimerization extends deeper in the bilayer and away from polar residues of both peptides, indicating interactions within the hydrophobic region could essentially affect the helix-helix association of ORF7a and BST-2 heterodimer. According to the known structural and biofunctional features of ORF7a of SARS-Cov-2, K119, as a ubiquitination position [66], shows a strong interactive energy in the WT packing (Groups 2 and 4), compared to the I28S packing, indicating tight dimeric packing would reduce the activity of the K^{119} and may prohibit ubiquitination. Additionally, one of the identified key contact residues on ORF7a, A^{105}, has been reported as mutating to valine, resulting in a strong interaction among L^{102}, I^{103}, V^{104}, and A^{105} which is associated with increased severity and lethality of the infection in a group of Romanian patients [67]. A^{105} is listed as a key contacting residue (Table S8) in the WT packing (Group 3), matching the observation in the study of Romanian cases. On the other hand, several identified TM contact residues on BST-2 (Tables S8–S9): I^{34}, L^{37}, P^{40}, L^{41}, and T^{45} match those reported when this protein forms a heterodimer with the HIV-1 viral protein U [68].

2.3.4. *Summary of MD simulations*

A structural model for the protein-protein interface has been presented using multi-scale MD simulations. CG MD provided conformational and AA MD refined these structures. The heterodimeric structures have been classified by multilayer clustering including GMVAE and contact map matching approaches, resulting in 4 and 5 configuration clusters for the WT and I28S heterodimers, respectively. Among these clusters, residue contacts between the two peptides show variations in the domains and intensity. Although many strong residue pairs have been observed in the EC and CYTO domains, the tight helical packing structures revealed that coexisting interactions within the CYTO domain (the salt bridges) and at the TM/EC interface (the interacting complex) are likely the primary driving forces for stabilizing TM contacts in both heterodimers. The tight packing structures also revealed a longer TM helix-helix contact region for the I28S dimer compared to the WT.

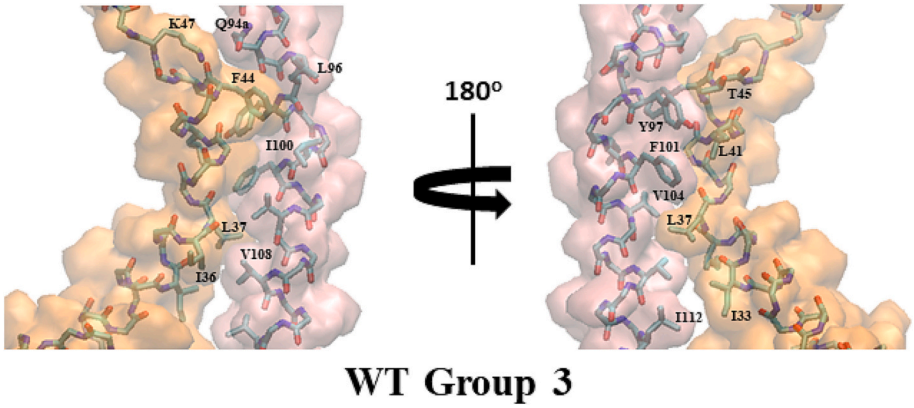

Fig. 7. Molecular structure of the helix-packing motif with detailed side chain contacts in the I28S Group 4. Oxygen, nitrogen, and carbon atoms colored red, blue, and gray, respectively and I28S BST-2 and ORF7a colored orange and pink in QuickSurf style. The labeled residues have contacts with the other peptide chain within 3 Å. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

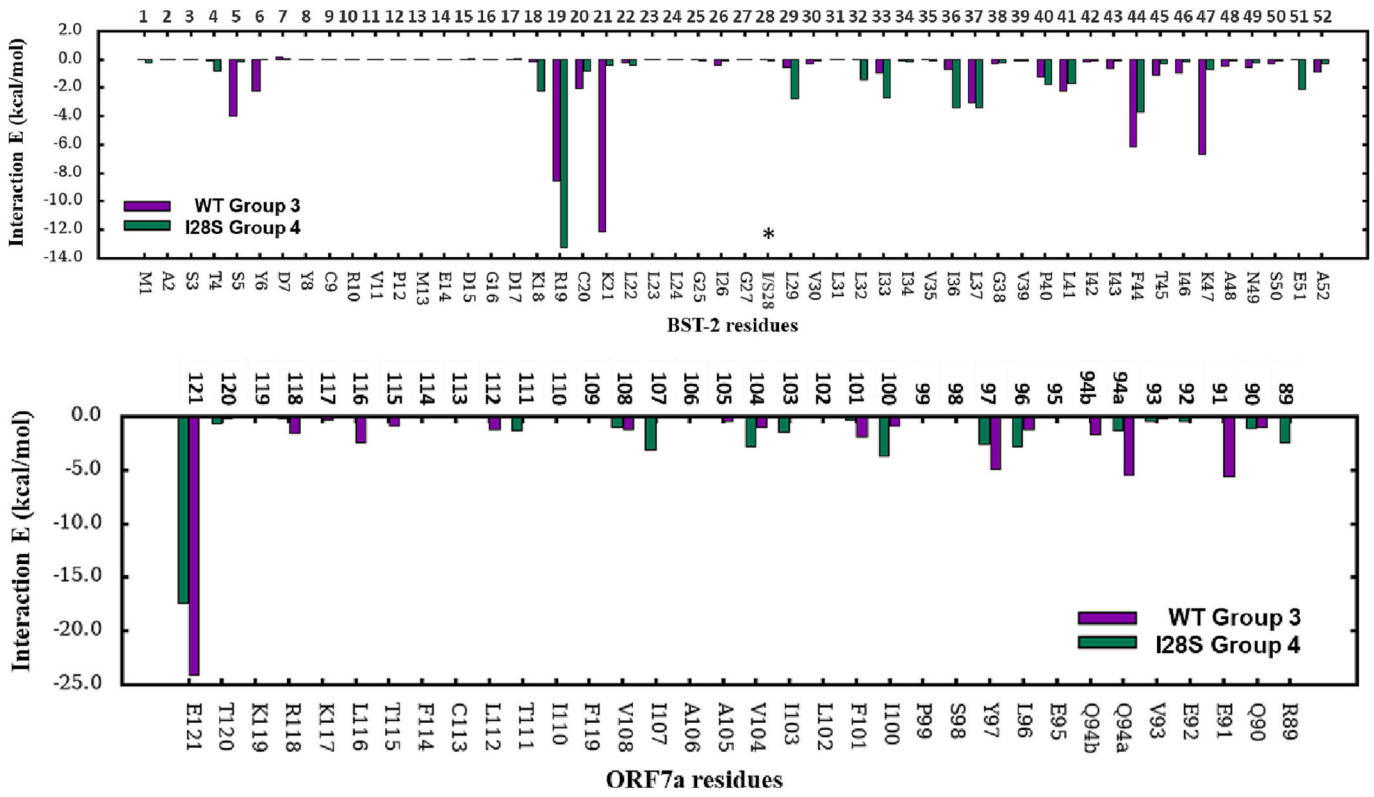

Fig. 8. Interaction energies for amino acid residues.
(**A**) BST-2 and (**B**) ORF7a computed over the WT Group 3 and I28S Group 4. The individual membership possibility of heterodimeric configuration is >0.75 for its own group. The mutated residue is marked with an "*".

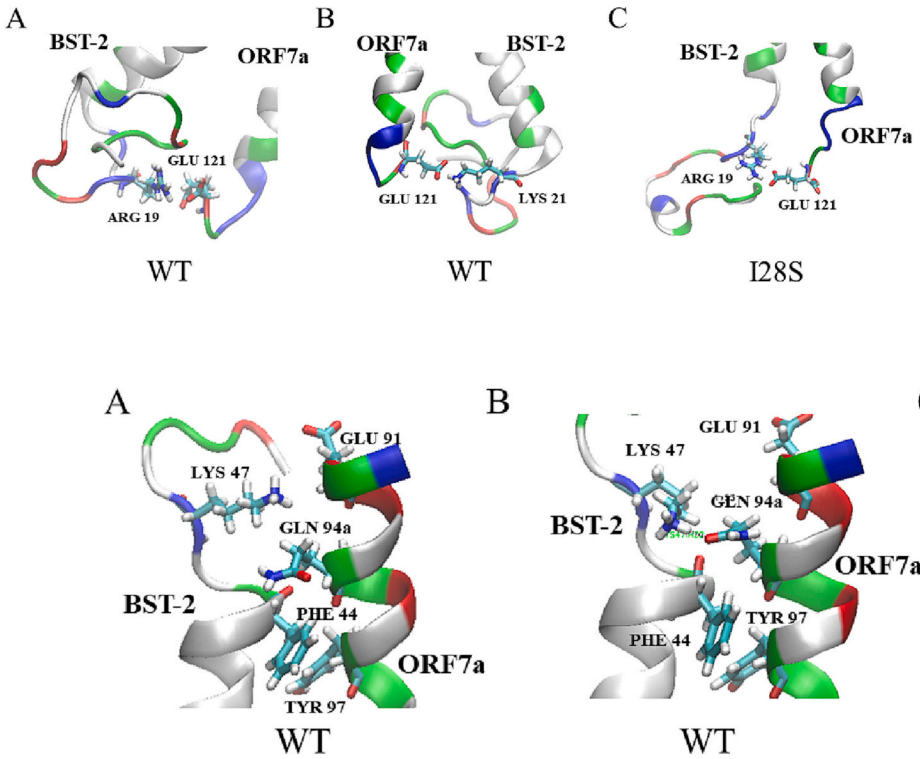

Fig. 9. Salt bridges formed in WT Group 3 and I28S Group 4.
Salt bridges are formed by (**A**) E^{121} (ORF7a) and R^{19} (WT BST-2), (**B**) E^{121} (ORF7a) and K^{21} (WT BST-2) and (**C**) E^{121} (ORF7a) and R^{19} (I28S BST-2). Nonpolar, polar, acidic, and basic residues in secondary structure are colored white, green, red, and blue, respectively. Oxygen, nitrogen, carbon and hydrogen atoms in highlight molecular structures are colored red, blue, cyan, and white, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

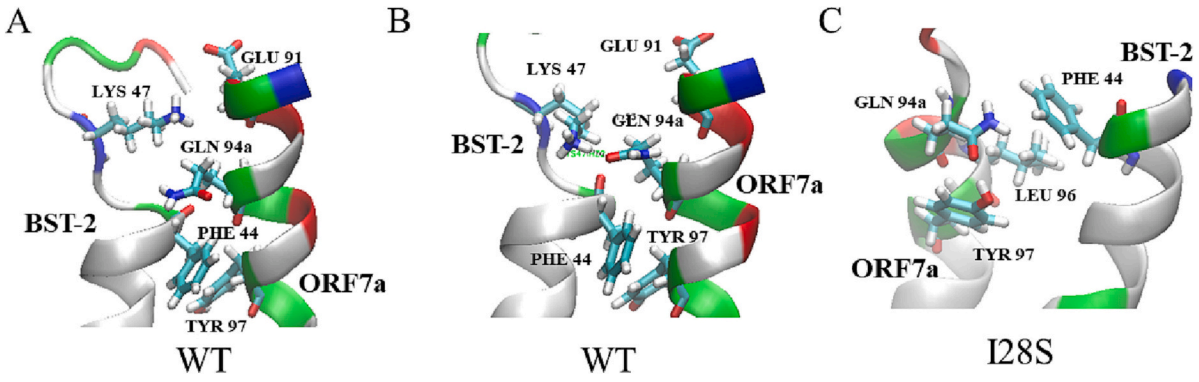

Fig. 10. Contacting complexes of ORF7a/BST-2 heterodimerization at the EC/TM interface
In WT packing, (**A**) a salt bridge formed by E^{91} (ORF7a) and K^{47} (BST-2) or (**B**) a hydrogen bond between Q^{94a} (ORF7a) and K^{47} (BST-2) promotes an amino-π interaction between Q^{94a} (ORF7a) and F^{44} (BST-2) and π-π stacking between Y^{97} (ORF7a) and P^{44} (BST-2). In I28S packing, (**C**) an ORF7a intra-hydrogen bond formed by Q^{94a} and Y^{97} in the extended helix domain promotes the same residues to form amino-π stacking and a π-π stacking interactions.

Fig. 11 shows that the I28S mutation affects the backbone structure by forming an extra sidechain-backbone hydrogen bond (formation probability over 0.99) between S^{28} and L^{24} of BST-2 which does not occur in the WT BST-2. I28S appears to alter the rigidity of the helix structure, which affects the heterodimeric packing. Examination of contacting water within 3 Å of each BST-2 residue shows that contacting water stops at G25 from the cytoplasm in the I28S Group 4 while it goes further to L^{29} in the WT Group 3. This indicates the extra hydrogen bond reduces the steric restriction which prevents water molecules from penetrating deeper into the hydrophobic region and prompts the tight helix-helix packing. Moreover, the increase in rigidity of the I28S BST-2 helix structure affects the whole domain. By creating better alignment of the backbone atoms, the extended L_{Helix} of I28S BST-2 about 3 residues (averaging over all groups) longer than WT BST-2 can promote hydrophobic interactions between the two helices. These MD simulations provide molecular insights on how mutations at I28 can influence the dimerization profile with ORF7a.

2.4. Coexpression of SARS-CoV-2 ORF7a leads to lower-molecular weight BST-2 products in HEK293T cells

As explained in a previous section, heterodimerization of I28S BST-2 with ORF7a could not be assessed via the bacterial transmembrane DN-AraTM assay. However, the MD simulations presented here have not only elucidated the probability of heterodimerization but also extensively characterized these interaction mechanisms and compared them to wild type BST-2. Therefore, protein-protein interactions were further assessed using the full-length sequences of both BST-2 and its variants as well as SARS-CoV-2 ORF7a in HEK293T cells. Previously, it has been shown that SARS-CoV-2 ORF7a interferes with glycosylation of BST-2 [36]. Here, we show that cells cotransfected with SARS-CoV-2 ORF7a also led to a higher abundance of lower molecular weight bands for BST-2 (Fig. 12). This increase in unglycosylated BST-2 also occurs in I28S and I28T variants despite the inhibition of self-association for the I28S transmembrane domain seen in our bacterial assays.

3. Discussion

Here, we employed BST-2 mutational studies to show that oligomerization is regulated through specific TM residue interactions. As alluded to previously, BST-2 exhibits evidence of positive, adaptive selection in response to viral pathogen exposure. Therefore, there are a wide range of identified single nucleotide polymorphisms (SNPs) within the TM sequence for human BST-2. By analogy to HIV-1 Vpu, it is intriguing to speculate that these mutations may enhance or diminish the heterooligomerization of BST-2 with SARS-CoV-2 ORF7a, thereby enhancing or diminishing the immunomodulatory effect of ORF7a.

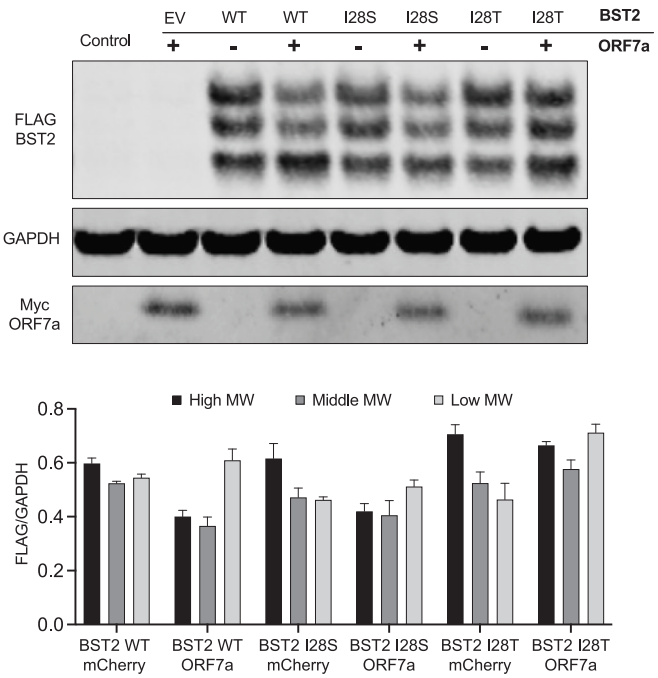

Fig. 12. HEK293T cells expressing BST-2 variants exhibit changes in glycosylation patterns when cotransfected with SARS-CoV-2 ORF7a. (**Top**) Western blot of cells transfected with wild-type as well as I28S and I28T BST-2 exhibit an increase in intensity for lower molecular weight band upon coexpression with ORF7a. (**Bottom**) Densitometry analysis of protein bands normalized to GAPDH internal controls. ($n = 3$) This lower molecular weight band represents unglycosylated BST-2 which suggests ORF7a expression leads to decreased levels of glycosylated BST-2 in all variants.

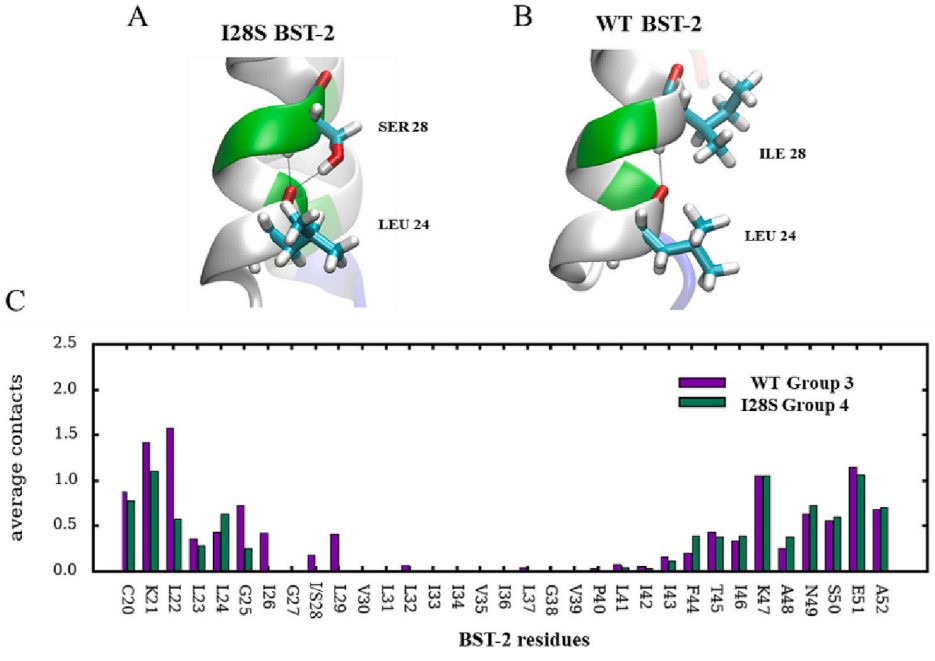

Fig. 11. I28S mutation effects.
The I28S SNP in BST-2 alters the local molecular interactions and water contacts per residues. The configurations demonstrate that (**A**) S^{28} forms an extra sidechain-backbone hydrogen bond with L^{24}, which does not occur for (**B**) I^{28}. (**C**) The water contacts for each BST-2 residue computed water contacting residues within 3 Å over the WT Group 3 and I28S Group 4 which the individual membership possibility of heterodimeric configuration is >0.75 for its own group.

Based on our bacterial *AraC* transcription-based assay, we observed that preferential heterodimeric interactions of the TM domains within ORF7a and BST-2 occur but self-association of BST-2 may be affected by an I28 variant. Our AA structural model for protein-protein interface, however, shows heterodimerization of ORF7a with I28S BST-2. Furthermore, it reveals the helix-packing motif involves more residue contacting pairs in the I28S mutant that extend to the bottom TM domain compared to the WT which only exists at the top and middle TM domains. This primary change alters the hydrophobic interaction of the TM association promoting a strong helix-packing in the I28S heterodimeric configuration. We also show that HEK293T cells co-transfected with SARS-CoV-2 ORF7a and BST-2 display different glycosylation patterns. These patterns are consistent with the previously studied SARS-CoV ORF7a – WT BST-2 interactions [36]. As heterodimerization of SARS-CoV ORF7a with unglycosylated BST-2 is preferred [36], these results are consistent with a mechanism by which SARS-CoV-2 ORF7a could also be responsible for BST-2 antagonism, including both wild-type and BST-2 variants studied.

4. Experimental procedures

4.1. Subcloning

Unless otherwise stated, all molecular biology procedures were performed according to standard methods supplied by manufacturers. For site-directed mutagenesis, primers were designed using PrimerX, and mutations were introduced using the QuikChange II site-directed mutagenesis kit (Agilent). Mutated sequences were verified by DNA sequencing (GeneWiz). *E. coli* strain DH5α was used for DNA maintenance.

4.2. Plasmids

For bacterial assays, plasmids pAraTMwt and pAraTMDN were generated as previously described from pTrcRSF and pTrc99a scaffolds respectively [38,39]. The reporter plasmid, pAraGFP was previously derived from plasmid pDS439 [38]. The DNA sequences coding for the BST-2 domains of interest were cloned into the pAraTMwt (coding for AraC) plasmid, and the ORF7-a transmembrane domains of interest for both COVID and SARS were cloned into the pAraTMDN (coding for the inactive form of AraC, AraC*) plasmid. For mammalian cell expression, pCAGGS-mCherry as a gift from Phil Sharp (Addgene plasmid # 41583; http://n2t.net/addgene:41583; RRID:Addgene_41583) [40], and pcDNA3.1(+) was purchased from Invitrogen. BST-2-FLAG was constructed where the FLAG epitope was fused to the N-terminal of human BST-2 (UniProtKB – Q10589) and amplified using forward primer (5′-AAACTTAAGCTTGGTACCGCCACCATGGATTACAAGGA-3′) and reverse primer (5′-CTCTAGACTCGAGCGGCCG-3′). PCR products were digested with *Kpn*I and *Not*I and cloned into pcDNA3.1(+). For the SARS-CoV-2 ORF7a-myc construct, forward primer (5′-ATCATTTTGGCAAAGAATTCTAGCGCC-3′) and reverse primer (5′-GATCGAGATCTGAGTCCGGATTACAGA-3′) were used to amplify the sequence. A c-myc tag was fused in-frame to the C-terminal of ORF7a to replace the stop codon at the 3′ terminal end. The PCR product was then digested with *Eco*RI and *Bgl*II and cloned into pCAGGS-mCherry.

4.3. Ara-TM and DN-AraTM dimerization assays

Ara-TM and DN-AraTM dimerization assays were conducted as previously described [38,39]. The constructs in the pAraTMwt plasmids and the reporter plasmid (pAraGFPCDF) were co-transformed with or without the pAraTMDN constructs for the Ara-TM homodimerization and DN-AraTM heterodimerization assays respectively into the AraC-deficient *E. coli* strain SB1676 and streaked onto selective LB plates (100 μg/mL ampicillin, 50 μg/mL kanamycin, and 100 μg/mL spectinomycin). Colonies were picked for each construct and grown in 2 mL of selective lysogeny broth (LB) for 12 h at 37 °C and 250 rpm. Cultures were then diluted into selective media with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for protein induction and grown in a 2.0-mL-deep, 96-well PP plate (PlateOne) for additional 6 h at 37 °C and 250 rpm. We then transferred 200 μL of each culture to a black 96-well, clear bottom plate (Greiner). Absorbance measurements at 600 nm as well as GFP fluorescence emission measurements at 530 nm after excitation at 485 nm were collected using a Synergy Neo2 Hybrid Multi-Mode Microplate Reader (Biotek). The results are reported as the ratio of fluorescence emission at 530 nm to absorbance at 600 nm.

4.4. BST-2 and SARS-CoV-2 ORF7a protein co-transfection

HEK293T cells were plated in 6-well plates at a density of 4e5 cells per well and allowed to adhere overnight before transfection. Co-transfection was performed using Lipofectamine 3000 (Invitrogen) according to manufacturer's instructions. A total of 5 μg of DNA was used, with 2.5 μg of pcDNA3.1(+)-BST2-FLAG (or empty vector control) and 2.5 μg of pCAGGS-ORF7A-myc (or control). Cells were lysed 48 h after transfection.

4.4.1. Lysis and Western Blotting

Cells were lysed in standard cell lysis buffer (Invitrogen) supplemented with protease and phosphatase inhibitors (Sigma-Aldrich, P8340, P5726, P0044). Lysates were clarified by centrifugation at 20,000 ×*g* for 10 min at 4 °C, and total protein concentrations were determined by micro-bicinchoninic assay (Pierce). Approximately 20 μg of denatured protein was loaded per lane onto 4–12 % gradient polyacrylamide gels (Thermo Fisher Scientific). Samples were transferred to 0.2 μm nitrocellulose membranes (Bio-Rad). Membranes were probed with antibodies and imaged using a LICOR Odyssey CLx system or a Bio-Rad ChemiDoc. Antibodies used included FLAG (Cell Signaling Technology, #8146), GAPDH (Santa Cruz, sc-32233), and myc-HRP conjugate (Cell Signaling Technology, #9B11).

4.5. Homology modeling

A model of the ORF7a (amino acids 89–121 of NCB reference sequence: NC_ 004718.3; (RQEEVQ^{94a}Q^{94b}ELYSPIFLIVAAIVFITLCFTLKRKTE) (Fig. 1A) was generated by ab initio modeling with ROBETTA [41]. This is a chimeric sequence where the N-terminus (EC domain) is from SARS-CoV while the remaining sequence is SARS-CoV-2. The only difference is in the two consecutive glutamines starting at the 94th residue, while only a single Glutamine at position in SARS-CoV-2. This mistake should have minimal influence on our structures obtained below. Hence, we labeled the tthtwo consecutive Glutamine as Q^{94a} and Q^{94b}, for keeping consistency of the sequence number with ORF7a of SARS-Cov-2. A model of the BST-2 (amino acids 1–52 of NCB accession number BAD96844.1) (MASTSYDYCRVPMEDGDKRCKLLLGIGILVLLIIVILGVPLIIFTIKANSEA) (Fig. 1B) was generated using known crystal structures (4P6Z (X-ray; 3A) (https://www.rcsb.org/structure/4p6z), 2LK9 (NMR) (https://www.rcsb.org/structure/2LK9 BMRB: 17985), and 3MQC (X-ray; 2.8 A) with MODELLER. The mutated BST-2 (I28S) was also generated by MODLLER after the WT BST-2 model.

4.6. Coarse-grained molecular dynamics (CG-MD)

MD simulations were performed on the TM interface between ORF7a and BST-2 and the approaches was summarized in Fig. S1. The wild-type structure of ORF7a was taken to form dimers with the wild-type and the I28S mutant structures of BST-2 to identify the effect of the mutation on dimerization. The initial structure was constructed using the CHARMM-GUI *Martini Bilayer Maker* [42–44]with a separation of ~2 nm based on their TM helical axis. Three selected placements (Fig. S1) were set for three replicas in both WT and I28S. The simulation box also includes a

15 Å layer of explicit nonpolarizable water as a buffer on the top and the bottom of the system, a bilayer of 100 lipids of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) per leaflet. MD simulations were conducted using GROMACS 2019.4 [45,46] with the Martini 2.2 [47,48] on Intel® Ivy Bridge E5-2680v2 processors at the Deepthought2 High-Performance Computing (HPC) cluster. First, the systems were energy minimized using 5000 steps of a steepest descent algorithm followed by equilibrium runs. Then, 5 μs production runs (unscaled time) were performed in the isothermal–isobaric ensemble (NPT) ensemble at 310 K and 1 bar. The Berendsen method [49] as applied with semi-isotropic scaling at a pressure of 1 bar for pressure coupling. The potential-shift [50] was used for both electrostatics and van der Waals calculations with cutoff at 11 Å. A 25 fs time step was used, and trajectories were saved every 125 ps. The last 2.5 μs of simulations (after equilibrium) was taken to perform distance calculations and create inter-residue contact maps using CHARMM scripts.

4.7. All-atom MD (AA-MD)

Each ORF7a/BST-2 dimer structures obtained from the last time step of CG simulations were taken as initial configurations of AA-MD simulations. CG to AA conversion of a whole simulation system including proteins, POPC lipids, and water, was conducted using the CHARMM-GUI *Martini to All-atom Converter* [51]. The NAMD 2.14 [52,53] simulation package was used with the CHARMM36 (C36) force field for protein and lipids [54,55] and TIP3P [56,57] for water, and NPT ensemble at 310 K and 1 atm, performed on NVIDIA® Tesla® K80 and P100 GPUs at the Bluecrab HPC cluster. The Lennard-Jones potential was used to describe van der Waals interactions, and a force-based switching function in the range of 10–12 Å was chosen [58]. Langevin dynamics maintained the temperature, and the Nosé-Hoover Langevin piston algorithm was applied to maintain the pressure [59,60]. Hydrogen atoms were constrained by using the RATTLE algorithm [61]. Simulations were run for 800 ns with a time step was 2 fs.

4.8. Clustering for AA configurations

A machine learning method, Gaussian mixture variational autoencoder (GMVAE) [62], that can perform an unsupervised clustering of protein conformations using the distance matrix of C_α as inputs was applied to identify the main dimer clusters. The GMVAE has shown its capacity for identifying configurations and used to study several proteins in their folding mechanism [63]. Here, we will demonstrate the potential application of the GMVAE for clustering helix-helix dimerization. The GMVAE clustering by the distance matrix of C_α between two peptide chains could sort the backbone configurations with structural similarity, which would not limit contacting configurations. All three replicas (dropping the first 200 ns each) were combined for GMAVE clustering. To further investigate the contacting of two peptides, we grouped the cluster with the occupancy of contact map to remove the clusters with weak contacts and combined the clusters with high similarity in their contact maps. These identified dimer clusters were further examined for their contacts between ORF7a and BST-2 to construct the protein-protein interface structural model for dimerization to provide the insight of molecular interaction at the atomic level.

Funding and additional information

This work was supported by the National Science Foundation grants CHE-2029895 (B.W.B., M.M.M., J.D.T.) and CHE-2029900 (J.B·K, M-K. H.). Additional support was provided through the University of Virginia School of Engineering and Applied Science (UVA SEAS) Phil Parrish Fellowship (J.D.T.). Computation support was provided through the Extreme Science and Engineering Discovery Environment (XSEDE) (grant # MCB-100139) (Comet), MARCC (JHU/UMD) and Deepthought2 (UMD).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bryan Berger and Jeff Klauda reports financial support was provided by National Science Foundation. The co-author Jeff Klauda serves on the editorial board of BBA Biomembranes. Bryan Berger and Jeff Klauda reports financial support was provided by National Science Foundation. The co-author Jeff Klauda serves on the editorial board of BBA Biomembranes.

Data availability

Data for this manuscript is provided in the provided in Figs. 1–12 and Supporting Information.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbamem.2023.184174.

References

[1] M. Cai, L. Xun, Organization and regulation of pentachlorophenol-degrading genes in Sphingobium Chlorophenolicum ATCC 39723, J. Bacteriol. 184 (17) (2002) 4672–4680, https://doi.org/10.1128/JB.184.17.4672-4680.2002.
[2] Z.Y. Wu, J.M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention, J. Am. Med. Assoc. 323 (13) (2020) 1239–1242.
[3] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C. L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X. R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F. X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin (Vol 579, Pg 270, 2020), Nature 588 (7836) (2020) E6.
[4] J. Shang, G. Ye, K. Shi, Y.S. Wan, C.M. Luo, H. Aihara, Q.B. Geng, A. Auerbach, F. Li, Structural basis of receptor recognition by SARS-CoV-2. *Nature*, Nature 581 (7807) (2020) 221–+.
[5] A. Chandrashekar, J.Y. Liu, A.J. Martinot, K. McMahan, N.B. Mercado, L. Peter, L. H. Tostanoski, J.Y. Yu, Z. Maliga, M. Nekorchuk, K. Busman-Sahay, M. Terry, L. M. Wrijil, S. Ducat, D.R. Martinez, C. Atyeo, S. Fischinger, J.S. Burke, M.D. Slein, L. Pessaint, A. Van Ry, J. Greenhouse, T. Taylor, K. Blade, A. Cook, B. Finneyfrock, R. Brown, E. Teow, J. Velasco, R. Zahn, F. Wegmann, P. Abbink, E.A. Bondzie, G. Dagotto, M.S. Gebre, X. He, C. Jacob-Dolan, N. Kordana, Z.F. Li, M.A. Lifton, S. H. Mahrokhian, L.F. Maxfield, R. Nityanandam, J.P. Nkolola, A.G. Schmidt, A. D. Miller, R.S. Baric, G. Alter, P.K. Sorger, J.D. Estes, H. Andersen, M.G. Lewis, D. H. Barouch, SARS-CoV-2 infection protects against rechallenge in Rhesus macaques, Science 369 (6505) (2020) 812–+.
[6] J. Shang, Y.S. Wan, C.M. Luo, G. Ye, Q.B. Geng, A. Auerbach, F. Li, Cell entry mechanisms of SARS-CoV-2, Proc. Natl Acad. Sci. USA 117 (21) (2020) 11727–11734.
[7] C.B. Jackson, M. Farzan, B. Chen, H. Choe, Mechanisms of SARS-CoV-2 entry into cells, Nat. Rev. Mol. Cell Biol. 23 (1) (2022 Jan) 3–20, https://doi.org/10.1038/s41580-021-00418-x.
[8] R.J. Lu, X. Zhao, J. Li, P.H. Niu, B. Yang, H.L. Wu, W.L. Wang, H. Song, B.Y. Huang, N. Zhu, Y.H. Bi, X.J. Ma, F.X. Zhan, L. Wang, T. Hu, H. Zhou, Z.H. Hu, W.M. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J.Y. Yuan, Z.H. Xie, J.M. Ma, W.J. Liu, D.Y. Wang, W.B. Xu, E.C. Holmes, G.F. Gao, G.Z. Wu, W.J. Chen, W.F. Shi, W. J. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (10224) (2020) 565–574.
[9] J.F.W. Chan, K.H. Kok, Z. Zhu, H. Chu, K.K.W. To, S. Yuan, K.Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan (Vol 23, Pg 2511, 2019), Emerg. Microbes Infec. 9 (1) (2020) 540.
[10] R. Arya, S. Kumari, B. Pandey, H. Mistry, S.C. Bihani, A. Das, V. Prashar, G. D. Gupta, L. Panicker, M. Kumar, Structural insights into SARS-CoV-2 proteins, J. Mol. Biol. 433 (2) (2021).
[11] F.K. Yoshimoto, The proteins of severe acute respiratory syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19, Protein J 39 (3) (2020) 198–216.
[12] X.B. Lei, X.J. Dong, R.Y. Ma, W.J. Wang, X. Xiao, Z.Q. Tian, C.H. Wang, Y. Wang, L. Li, L.L. Ren, F. Guo, Z.D. Zhao, Z. Zhou, Z.C. Xiang, J.W. Wang, Activation and evasion of Type I interferon responses by SARS-CoV-2, Nat. Commun. 11 (1) (2020).
[13] Z.G. Cao, H.J. Xia, R. Rajsbaum, X.Z. Xia, H.L. Wang, P.Y. Shi, Ubiquitination of SARS-CoV-2 ORF7a promotes antagonism of interferon response, Cell. Mol. Immunol. 18 (3) (2021) 746–748.

[14] Z.L. Zhou, C.L. Huang, Z.C. Zhou, Z.X. Huang, L.L. Su, S.S. Kang, X.X. Chen, Q. Y. Chen, S.H. He, X. Rong, F. Xiao, J. Chen, S.D. Chen, Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14(+) monocytes, Iscience 24 (3) (2021).

[15] C. Lucas, P. Wong, J. Klein, T.B.R. Castro, J. Silva, M. Sundaram, M.K. Ellingson, T. Mao, J.E. Oh, B. Israelow, T. Takahashi, M. Tokuyama, P. Lu, A. Venkataraman, A. Park, S. Mohanty, H. Wang, A.L. Wyllie, C.B.F. Vogels, R. Earnest, S. Lapidus, I. M. Ott, A.J. Moore, M.C. Muenker, J.B. Fournier, M. Campbell, C.D. Odio, A. Casanovas-Massana, Yale I.M.P.A.C.T. Team, A. Obaid, A. Lu-Culligan, A. Nelson, A. Brito, A. Nunez, A. Martin, A. Watkins, B. Geng, C. Kalinich, C. Harden, C. Todeasa, C. Jensen, D. Kim, D. McDonald, D. Shepard, E. Courchaine, E.B. White, E. Song, E. Silva, E. Kudo, G. DeIuliis, H. Rahming, H.-J. Park, I. Matos, J. Nouws, J. Valdez, J. Fauver, J. Lim, K.-A. Rose, K. Anastasio, K. Brower, L. Glick, L. Sharma, L. Sewanan, L. Knaggs, M. Minasyan, M. Batsu, M. Petrone, M. Kuang, M. Nakahata, M. Campbell, M. Linehan, M.H. Askenase, M. Simonov, M. Smolgovsky, N. Sonnert, N. Naushad, P. Vijayakumar, R. Martinello, R. Datta, R. Handoko, S. Bermejo, S. Prophet, S. Bickerton, S. Velazquez, T. Alpert, T. Rice, W. Khoury-Hanold, X. Peng, Y. Yang, Y. Cao, Y. Strong, R. Herbst, A.C. Shaw, R. Medzhitov, W.L. Schulz, N.D. Grubaugh, C. Dela Cruz, S. Farhadian, A.I. Ko, S. B. Omer, A. Iwasaki, Longitudinal analyses reveal immunological misfiring in severe COVID-19, Nature 584 (7821) (2020) 463–469, https://doi.org/10.1038/s41586-020-2588-y.

[16] C.-M. Su, L. Wang, D. Yoo, Activation of NF-KB and induction of Proinflammatory cytokine expressions mediated by ORF7a protein of SARS-CoV-2, Sci. Rep. 11 (1) (2021) 13464, https://doi.org/10.1038/s41598-021-92941-2.

[17] C.Y. Cheung, L.L.M. Poon, I.H.Y. Ng, W. Luk, S.-F. Sia, M.H.S. Wu, K.-H. Chan, K.-Y. Yuen, S. Gordon, Y. Guan, J.S.M. Peiris, Cytokine responses in severe acute respiratory syndrome coronavirus-infected macrophages in vitro: possible relevance to pathogenesis, J. Virol. 79 (12) (2005) 7819–7826, https://doi.org/10.1128/JVI.79.12.7819-7826.2005.

[18] T. Ohtomo, Y. Sugamata, Y. Ozaki, K. Ono, Y. Yoshimura, S. Kawai, Y. Koishihara, S. Ozaki, M. Kosaka, T. Hirano, M. Tsuchiya, Molecular cloning and characterization of a surface antigen preferentially overexpressed on multiple myeloma cells, Biochem. Biophys. Res. Commun. 258 (3) (1999) 583–591.

[19] R.E. Randall, S. Goodbourn, Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures, J. Gen. Virol. 89 (2008) 1–47.

[20] A.J. Andrew, E. Miyagi, S. Kao, K. Strebel, The formation of cysteine-linked dimers of BST-2/Tetherin is important for inhibition of HIV-1 virus release but not for sensitivity to Vpu, Retrovirology (2009) 6.

[21] E.F. Terwilliger, E.A. Cohen, Y.C. Lu, J.G. Sodroski, W.A. Haseltine, Functional-role of human immunodeficiency virus Type-1 Vpu, Proc. Natl Acad. Sci. USA 86 (13) (1989) 5163–5167.

[22] E.S. Lim, H.S. Malik, M. Emerman, Ancient adaptive evolution of Tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses, J. Virol. 84 (14) (2010) 7124–7134.

[23] X. Yin, Z. Hu, Q.Y. Gu, X.L. Wu, Y.H. Zheng, P. Wei, X.J. Wang, Equine Tetherin blocks retrovirus release and its activity is antagonized by equine infectious Anemia virus envelope protein, J. Virol. 88 (2) (2014) 1259–1270.

[24] I. Dietrich, E.L. McMonagle, S.J. Petit, S. Vijayakrishnan, N. Logan, C.N. Chan, G. J. Towers, M.J. Hosie, B.J. Willett, Feline Tetherin efficiently restricts release of feline immunodeficiency virus but not spreading of infection, J. Virol. 85 (12) (2011) 5840–5852.

[25] M. Bray, S. Prasad, J.W. Dubay, E. Hunter, K.T. Jeang, D. Rekosh, M. L. Hammarskjold, A small element from the Mason-Pfizer monkey virus genome makes human-immunodeficiency-virus Type-1 expression and replication rev-independent, Proc. Natl Acad. Sci. USA 91 (4) (1994) 1256–1260.

[26] M.W. McNatt, T. Zang, T. Hatziioannou, M. Bartlett, I. Ben Fofana, W.E. Johnson, S.J.D. Neil, P.D. Bieniasz, Species-specific activity of HIV-1 Vpu and positive selection of Tetherin transmembrane domain variants, Plos Pathog. 5 (2) (2009).

[27] A. Le Tortorec, S.J.D. Neil, Antagonism to and intracellular sequestration of human Tetherin by the human immunodeficiency virus type 2 envelope glycoprotein, J. Virol. 83 (22) (2009) 11966–11978.

[28] R.S. Mitchell, C. Katsura, M.A. Skasko, K. Fitzpatrick, D. Lau, A. Ruiz, E. B. Stephens, F. Margottin-Goguet, R. Benarous, J.C. Guatelli, Vpu antagonizes BST-2-mediated restriction of HIV-1 release via Beta-TrCP and Endo-lysosomal trafficking, Plos Pathog. 5 (5) (2009).

[29] J.L. Douglas, K. Viswanathan, M.N. McCarroll, J.K. Gustin, K. Fruh, A.V. Moses, Vpu directs the degradation of the human immunodeficiency virus restriction factor BST-2/Tetherin via a Beta TrCP-dependent mechanism, J. Virol. 83 (16) (2009) 7931–7947.

[30] T. Yoshida, S. Kao, K. Strebel, Identification of residues in the BST-2 TM domain important for antagonism by HIV-1 Vpu using a gain-of-function approach, Front. Microbiol. (2011) 2, https://doi.org/10.3389/fmicb.2011.00035.

[31] D. Gupta, P. Sharma, M. Singh, M. Kumar, A.S. Ethayathulla, P. Kaur, Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants, Cell. Mol. Life Sci. 78 (24) (2021) 7967–7989.

[32] C. Goffinet, I. Allespach, S. Homann, H.M. Tervo, A. Habermann, D. Rupp, L. Oberbremer, C. Kern, N. Tibroni, S. Welsch, J. Krijnse-Locker, G. Banting, H. G. Krausslich, O.T. Fackler, O.T. Keppler, HIV-1 antagonism of CD317 is species specific and involves Vpu-mediated proteasomal degradation of the restriction factor, Cell Host Microbe 5 (3) (2009) 285–297.

[33] L. Rong, J. Zhang, J. Lu, Q. Pan, R.-P. Lorgeoux, C. Aloysius, F. Guo, S.-L. Liu, M. A. Wainberg, C. Liang, The transmembrane domain of BST-2 determines its sensitivity to Down-modulation by human immunodeficiency virus type 1 Vpu, J. Virol. 83 (15) (2009) 7536–7546, https://doi.org/10.1128/JVI.00620-09.

[34] J. Ishikawa, T. Kaisho, H. Tomizawa, B.O. Lee, Y. Kobune, J. Inazawa, K. Oritani, M. Itoh, T. Ochi, K. Ishihara, T. Hirano, Molecular-cloning and chromosomal mapping of a bone-marrow stromal cell-surface gene, Bst2, that may be involved in pre-B-cell growth, Genomics 26 (3) (1995) 527–534.

[35] S.M. Wang, K.J. Huang, C.T. Wang, BST2/CD317 counteracts human coronavirus 229E productive infection by tethering Virions at the cell surface, Virology 449 (2014) 287–296.

[36] J.K. Taylor, C.M. Coleman, S. Postel, J.M. Sisk, J.G. Bernbaum, T. Venkataraman, E.J. Sundberg, M.B. Frieman, Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 Virion tethering through a novel mechanism of glycosylation interference, J. Virol. 89 (23) (2015) 11820–11833, https://doi.org/10.1128/JVI.02274-15.

[37] D. Perez-Caballero, T. Zang, A. Ebrahimi, M.W. McNatt, D.A. Gregory, M. C. Johnson, P.D. Bieniasz, Tetherin inhibits HIV-1 release by directly tethering Virions to cells, Cell 139 (3) (2009) 499–511, https://doi.org/10.1016/j.cell.2009.08.039.

[38] P.-C. Su, B.W. Berger, Identifying key Juxtamembrane interactions in cell membranes using AraC-based transcriptional reporter assay (AraTM), J. Biol. Chem. 287 (37) (2012) 31515–31526, https://doi.org/10.1074/jbc.M112.396895.

[39] P.-C. Su, B.W. Berger, A novel assay for assessing Juxtamembrane and transmembrane domain interactions important for receptor Heterodimerization, J. Mol. Biol. 425 (22) (2013) 4652–4658, https://doi.org/10.1016/j.jmb.2013.07.022.

[40] A.M. Gurtan, V. Lu, A. Bhutkar, P.A. Sharp, In vivo structure-function analysis of human dicer reveals directional processing of precursor MiRNAs, Rna 18 (6) (2012) 1116–1122, https://doi.org/10.1261/rna.032680.112.

[41] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the robetta server, Nucleic Acids Res. 32 (Web Server) (2004) W526–W531, https://doi.org/10.1093/nar/gkh468.

[42] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (11) (2008) 1859–1865, https://doi.org/10.1002/jcc.20945.

[43] Y. Qi, H.I. Ingólfsson, X. Cheng, J. Lee, S.J. Marrink, W. Im, CHARMM-GUI Martini maker for coarse-grained simulations with the Martini force field, J. Chem. Theory Comput. 11 (9) (2015) 4486–4494, https://doi.org/10.1021/acs.jctc.5b00513.

[44] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (10) (2009) 1545–1614, https://doi.org/10.1002/jcc.21287.

[45] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M.R. Shirts, J. C. Smith, P.M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics 29 (7) (2013) 845–854, https://doi.org/10.1093/bioinformatics/btt055.

[46] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25, https://doi.org/10.1016/j.softx.2015.06.001.

[47] D.H. de Jong, G. Singh, W.F.D. Bennett, C. Arnarez, T.A. Wassenaar, L.V. Schäfer, X. Periole, D.P. Tieleman, S.J. Marrink, Improved parameters for the Martini coarse-grained protein force field, J. Chem. Theory Comput. 9 (1) (2013) 687–697, https://doi.org/10.1021/ct300646g.

[48] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, S.-J. Marrink, The MARTINI coarse-grained force field: extension to proteins, J. Chem. Theory Comput. 4 (5) (2008) 819–834, https://doi.org/10.1021/ct700324x.

[49] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external Bath, J. Chem. Phys. 81 (8) (1984) 3684–3690, https://doi.org/10.1063/1.448118.

[50] D. van der Spoel, P.J. van Maaren, The origin of layer structure artifacts in simulations of liquid water, J. Chem. Theory Comput. 2 (1) (2006) 1–11, https://doi.org/10.1021/ct0502256.

[51] T.A. Wassenaar, K. Pluhackova, R.A. Böckmann, S.J. Marrink, D.P. Tieleman, Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models, J. Chem. Theory Comput. 10 (2) (2014) 676–690, https://doi.org/10.1021/ct400617g.

[52] J.C. Phillips, D.J. Hardy, J.D.C. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M.C.R. Melo, B.K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot, E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, J. Chem. Phys. 153 (4) (2020), 044130, https://doi.org/10.1063/5.0014475.

[53] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (16) (2005) 1781–1802, https://doi.org/10.1002/jcc.20289.

[54] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, J. Phys. Chem. B 114 (23) (2010) 7830–7843, https://doi.org/10.1021/jp101759q.

[55] J. Huang, A.D. MacKerell, CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data, J. Comput. Chem. 34 (25) (2013) 2135–2145, https://doi.org/10.1002/jcc.23354.

[56] S.R. Durell, B.R. Brooks, A. Ben-Naim, Solvent-induced forces between two hydrophilic groups, J. Phys. Chem. 98 (8) (1994) 2198–2202, https://doi.org/10.1021/j100059a038.
[57] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (2) (1983) 926–935, https://doi.org/10.1063/1.445869.
[58] P.J. Steinbach, B.R. Brooks, New spherical-cutoff methods for long-range forces in macromolecular simulation, J. Comput. Chem. 15 (7) (1994) 667–683, https://doi.org/10.1002/jcc.540150702.
[59] S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, Constant pressure molecular dynamics simulation: the Langevin piston method, J. Chem. Phys. 103 (11) (1995) 4613–4621, https://doi.org/10.1063/1.470648.
[60] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms, J. Chem. Phys. 101 (5) (1994) 4177–4189, https://doi.org/10.1063/1.467468.
[61] H.C. Andersen, Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations, J. Comput. Phys. 52 (1) (1983) 24–34, https://doi.org/10.1016/0021-9991(83)90014-1.
[62] N. Dilokthanakul, P.A.M. Mediano, M. Garnelo, M.C.H. Lee, H. Salimbeni, K. Arulkumaran, M. Shanahan, Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders, 2016, https://doi.org/10.48550/ARXIV.1611.02648.
[63] M. Ghorbani, S. Prasad, J.B. Klauda, B.R. Brooks, Variational embedding of protein folding simulations using Gaussian mixture Variational autoencoders, J. Chem. Phys. 155 (19) (2021), 194108, https://doi.org/10.1063/5.0069708.
[64] H. Yang, J. Wang, X. Jia, M.W. McNatt, T. Zang, B. Pan, W. Meng, H.-W. Wang, P. D. Bieniasz, Y. Xiong, Structural insight into the mechanisms of enveloped virus tethering by Tetherin, Proc. Natl. Acad. Sci. 107 (43) (2010) 18428–18432, https://doi.org/10.1073/pnas.1011485107.
[65] W. Deng, R. Li, Juxtamembrane contribution to transmembrane signaling: transmembrane signaling, Biopolymers 104 (4) (2015) 317–322, https://doi.org/10.1002/bip.22651.
[66] S.T. Sherry, DbSNP: the NCBI database of genetic variation, Nucleic Acids Res. 29 (1) (2001) 308–311, https://doi.org/10.1093/nar/29.1.308.
[67] M. Javanainen, H. Martinez-Seara, I. Vattulainen, Excessive aggregation of membrane proteins in the Martini model, PLoS One 12 (11) (2017), e0187936, https://doi.org/10.1371/journal.pone.0187936.
[68] R. Barton, P. Khakbaz, I. Bera, J.B. Klauda, M.K. Iovine, B.W. Berger, Interplay of specific trans- and Juxtamembrane interfaces in Plexin A3 dimerization and signal transduction, Biochemistry 55 (35) (2016) 4928–4938, https://doi.org/10.1021/acs.biochem.6b00517.
[69] L. Van der Maaten, G. Hinton, Visualizing Data Using T-SNE, J. Mach. Learn. Res. 1 (2008) 1–48.
[70] J. Lee, W. Im, Implementation and application of Helix–Helix distance and crossing angle restraint potentials, J. Comput. Chem. 28 (3) (2007) 669–680, https://doi.org/10.1002/jcc.20614.
[71] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637, https://doi.org/10.1002/bip.360221211.