 doi:10.20944/preprints202206.0359.v1

Article

Biomolecular Fluorescence Complementation Profiling and Artificial Intelligence Structure Prediction of the Kaposi's Sarcoma-associated Herpesvirus ORF18 and ORF30 Interaction

Yoshiko Maeda [1,#,†], Tadashi Watanabe [1,2,#], Taisuke Izumi [3], Kazushi Kuriyama [2], Shinji Ohno [2] and Masahiro Fujimuro [1,*]

[1] Department of Cell Biology, Kyoto Pharmaceutical University, Kyoto 607-8412, Japan; benirinngo-maimai@yahoo.co.jp, Y.M.; fuji2@mb.kyoto-phu.ac.jp, M.F.
[2] Department of Virology, Graduate school of medicine, University of the Ryukyus, Okinawa 903-0215, Japan; twatanab@med.u-ryukyu.ac.jp, T.W.; k208720@cs.u-ryukyu.ac.jp, K.K.; soono@med.u-ryukyu.ac.jp, S.O.
[3] Department of Biology, Saint Joseph's University, Philadelphia, PA 19104, U.S.A; tizumi@sju.edu
[#] These authors are equally contributed.
[†] Present Address: Division of Pharmacy, Shimonoseki City Hospital, Yamaguchi 750-8520, Japan
[*] Correspondence: fuji2@mb.kyoto-phu.ac.jp

Abstract: Kaposi's sarcoma-associated herpesvirus (KSHV) is the etiologic agent of Kaposi's sarcoma, primary effusion lymphoma (PEL) and multicentric Castleman's disease. During KSHV lytic infection, lytic-related genes, categorized as immediate-early, early, and late genes, are expressed in a temporal manner. The transcription of late genes requires the virus-specific pre-initiation complex (vPIC), which consists of viral transcription factors. However, the characterization of the protein-protein interactions of the vPIC factors has not been completely elucidated. KSHV ORF18 is one of the vPIC factors, and its interaction with other viral factors has not been sufficiently revealed. Here, we analyzed the interaction between ORF18 and another vPIC factor, ORF30, in living cells using the bimolecular fluorescence complementation (BiFC) assay. We identified four amino-acid residues (Leu29, Gln36, His41 and Trp170) on ORF18 that were responsible for its interaction with ORF30. The artificial intelligence (AI) system AlphaFold2 predicted that identified four residues are exposed to the surface of ORF18 and located in proximity to each other in the surface of ORF18. Thus, AI-predicted model supports the importance of four residues for binding of ORF18 to ORF30. Our results indicated that wet experiments in combination with AI may enhance the structural characterization of vPIC protein-protein interactions.

Keywords: Kaposi's sarcoma-associated virus (KSHV); Viral pre-initiation complex (vPIC); Bimolecular fluorescence complementation (BiFC); Artificial intelligence (AI) structure prediction; AlphaFold2

1. Introduction

KSHV, of which DNA fragments were initially detected from a Kaposi's sarcoma lesion from an AIDS patient in 1994, was identified as the 8th human herpesvirus [1]. Currently, KSHV is the causative factor of several malignancies, including Kaposi's sarcoma, PEL, and multicentric Castleman's disease [2-4]. The human herpesviruses are classified into the following three subfamilies: *alphaherpesvirinae* [herpes simplex virus-1 (HSV-1), herpes simplex virus-2 (HSV-2), and varicella-zoster virus (VZV)], *betaherpesvirinae* [human cytomegalovirus (HCMV), human herpesvirus-6A (HHV-6A), human herpesvirus-6B (HHV-6B), and human herpesvirus-7 (HHV-7)], and *gammaherpesvirinae* [Epstein-Barr virus (EBV) and KSHV] [5]. These human herpesviruses have a highly conserved replication system; however, their pathological features are divergent.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

The herpesviral vPIC is the essential transcriptional machinery for late gene expression and it is highly conserved between *betaherpesvirinae* and *gammaherpesvirinae*. Recently, much attention has been focused on the importance of the vPIC for viral replication or production [6]. The vPIC consists of several viral proteins that are recruited to the transcription start site (TSS) of the viral late genes, which encode viral structural proteins (*e.g.*, envelope, capsid, and tegument proteins) that participate in viral assembly. The vPIC recognizes the TATT motif in the TSS instead of a TATA-box and initiates pre-mRNA elongation in conjunction with cellular RNA polymerase II (RNAPII). Although the host cell pre-initiation complex (which consists of various general transcription factors, GTFs) is a functional homolog of vPIC, the amino-acid sequence homology between GTF and vPIC components is quite low. Therefore, the vPIC formation machinery and the precise roles of each vPIC component have remained largely unknown. In the case of KSHV, the vPIC is thought to consist of at least six viral components: ORF18, ORF24, ORF30, ORF31, ORF34, and ORF66 [7-12]. It has been reported that the KSHV ORF24 binds to the promoters of the KSHV late genes along with cellular RNAPII [7]. Furthermore, ORF34 is thought to act as a hub for interaction with other vPIC components including ORF18, ORF24, ORF31, and ORF66 [7-12]. Moreover, several studies have determined that ORF18 is also associated with ORF30, ORF31, ORF34, and ORF66 [7-10]. Thus, these previous studies indicate that ORF18 might serve as a bridge molecule between ORF30 and ORF31/ORF34/ORF66. Although ORF18 is a key element for KSHV vPIC formation, the binding partners of ORF18 have not been fully revealed. Therefore, to gain insight into the molecular interaction involved in ORF18, we applied the following two approaches: the BiFC assay and AI structure prediction. The BiFC assay employs split-fluorescent protein reconstitution to image protein-protein interactions in living cells [13]. Various methods for protein structure prediction in silico have been previously developed. However, the application of AI with the deep learning algorithm AlphaFold2 has led to a paradigm shift, and more accurate whole protein structures were obtained [14-16].

Initially, we analyzed the interactions between KSHV ORF18 and the other components of vPIC using a BiFC assay. Our results showed that apparent BiFC-positive signals were detected between ORF18 and ORF30. Furthermore, we identified four amino-acid residues within ORF18 that were responsible for its interaction with ORF30. Meanwhile, the AI system AlphaFold2 predicted KSHV ORF18 structure with a high confident algorithm score. The AI-predicted ORF18 structure showed that the four identified amino-acids were exposed to the surface of ORF18 and located in proximity to each other. These four residues could be thought to support the surface structure of ORF18 required for binding to ORF30. Thus, the AI-predicted structural model would imply that these four residues of ORF18 might be involved in the interaction with ORF30.

2. Results

2.1. Optimization of the BiFC assay to assess vPIC interactions

Generally, there is a possibility that a tag fused to a protein of interest (P.O.I.) may influence its physical interactions with other proteins. Additionally, in the BiFC assay, inappropriate tag orientation increases the physical distance between the split fluoro-proteins and inhibits fluorescence reconstitution, despite interactions between P.O.I.s. Thus, in the BiFC assay, these fused tags may produce pseudo-negative signals or reduced positive signals. Therefore, to eliminate protein interaction interference by a fused tag and its orientation, all the combinations of fused tags (i.e., the type of tag and its location) were constructed as expression plasmids and analyzed for optimization of the BiFC assay. KSHV vPIC components (ORF24, ORF30, ORF31, ORF34, and ORF66) were screened for ORF18 binding with the BiFC assay. The indicated combination of expression plasmids was transfected into 293T cells and the cells were subsequently subjected to flow cytometry. Harvested cells were gated by forward scatter (FSC) and side scatter (SSC) parameters to exclude debris, and the ratio of mKusabira-green (mKG)-positive cells were further analyzed. We detected a robust ORF18-ORF30 binding signal from all of the transfected cells.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

The combination and orientation of BiFC tags had various effects on the positive scores, ranging from approximately 30% to approximately 70%. Co-expression of N-terminally mKGN-tagged ORF18 (mKGN-ORF18) and C-terminally mKGC-tagged ORF30 (ORF30-mKGC) was the pair with the highest score (Figure 1). However, significant positive signals were not obtained between ORF18 and other vPIC components (Supplemental Figures S1a and S1b). Consequently, we further examined the binding characteristics of ORF18-ORF30.

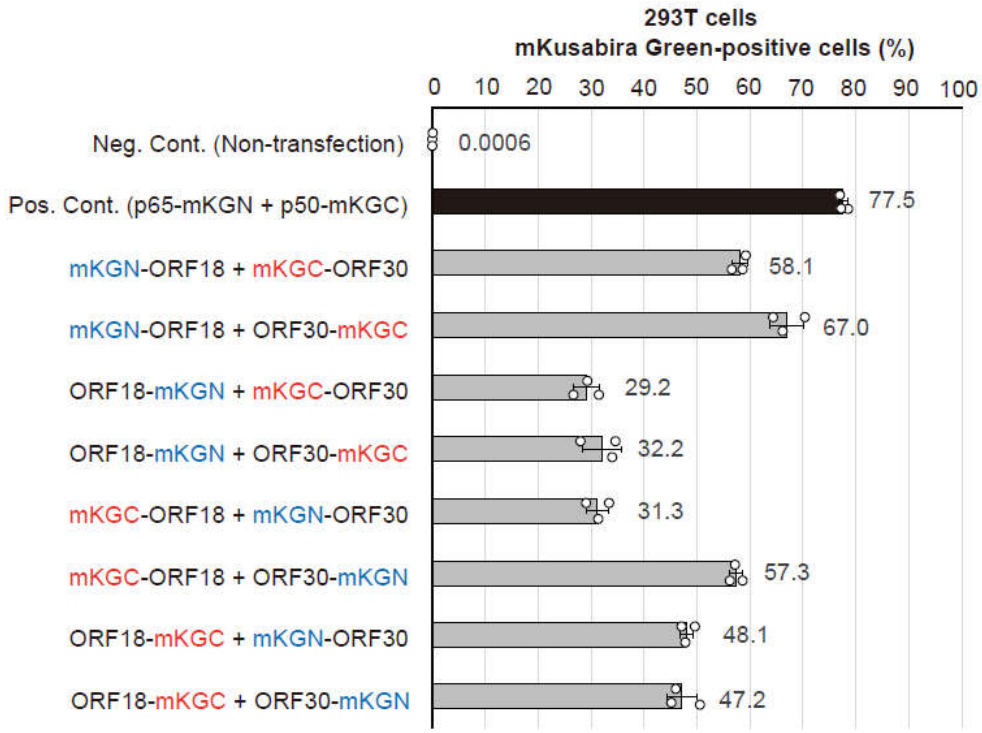

Figure 1. Optimization of the BiFC assay to assess the interaction between KSHV ORF18 and ORF30. In order to optimize the addition of a tag to ORF18 and ORF30, we used different types of split-mKusabira Green (mKG) tag (mKGN or mKGC) and different locations of the tag (N-terminus or C-terminus). Each indicated combination of the expression plasmids was co-transfected into 293T cells by the calcium phosphate method. The mKG-signal was assessed by flow cytometry in three independent samples, which are indicated by the white dots. We utilized non-transfected cells as a negative control (Neg. Cont.). The positive control (Pos. Cont.) was cells co-transfected with p65-mKGN (pCONT-1) and p50-mKGC (pCONT-2) expression plasmids. The average of the mKG-positive cells (%) are noted beside the bars. Each bar and error bar indicate the average and standard deviation, respectively.

2.2. Identification of the regions of ORF18 that interact with ORF30

In order to identify the regions of ORF18 that interact with ORF30, we constructed a total of 13 ORF18 truncated mutants. These truncations consisted of truncated region 1 (TR1) to truncated region 13 (TR13). Each truncation consisted of approximately 20 amino-acid residues that were deleted from the N-terminus to the C-terminus of ORF18 (Figure 2; left panel). If the truncated regions are important for the association with ORF30, the positive BiFC signal might be reduced or lost. Each mKGN-ORF18 truncated mutant was co-transfected with ORF30-mKGC into 293T cells (Figure 2; center panel) or HeLa cells (Figure 2; right panel) and analyzed. Both cell lines showed similar results, ORF18 truncated mutants scoring under 30% in 293T cells and under 5% in HeLa cells were considered as binding-negative mutants. Thus, the following binding-negative mutants were observed: TR1 (Δ1-20 amino-acid (a.a.)), TR2 (Δ 21-40 a.a.), TR3 (Δ 41-60 a.a.), TR8 (Δ 141-160 a.a.), TR9 (Δ 161-60 a.a.), and TR10 (Δ 141-200 a.a.). Therefore, this data suggested that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

amino-acid residues 1 to 60 and amino-acid residues 141 to 200 of ORF18 are essential regions for ORF30 binding.

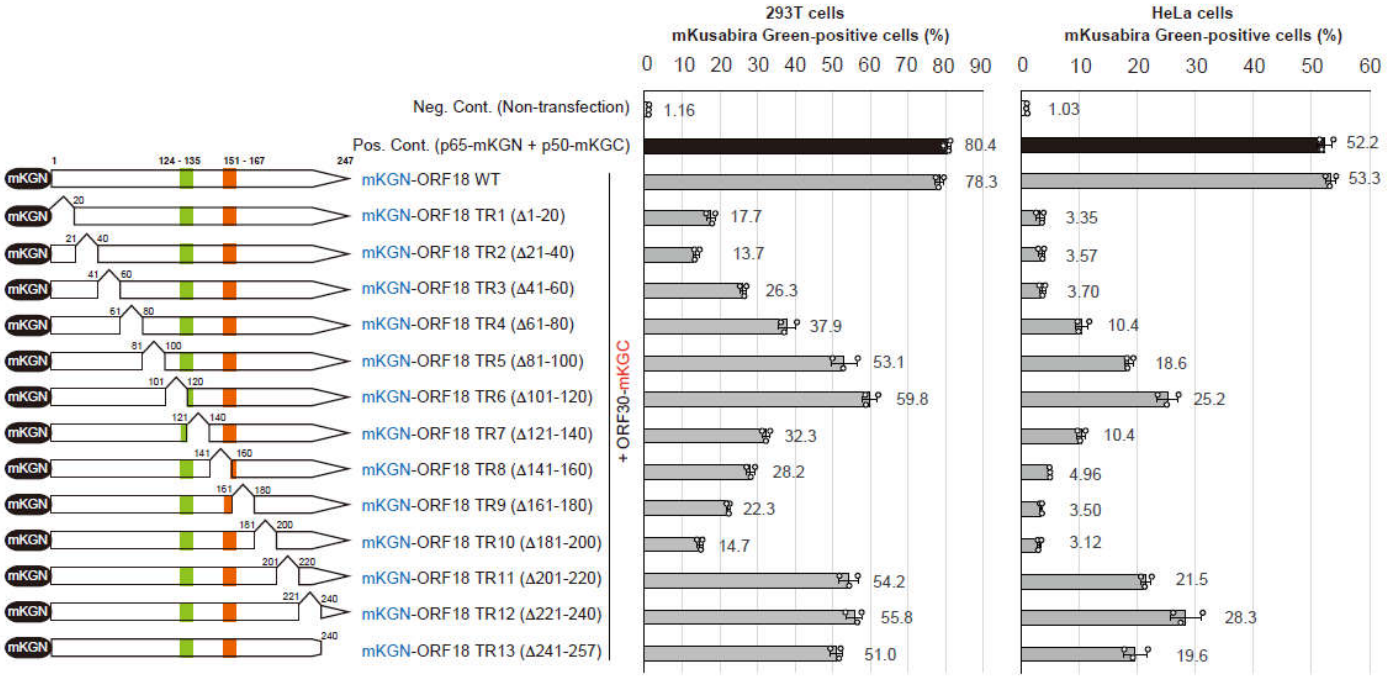

Figure 2. Identification of the ORF30-binding site on ORF18 using truncated mutants. The N-terminally mKGN-tagged ORF18 truncated region (TR) mutants (mKGN-ORF18 TR1 to TR13) are graphically displayed. The green block on ORF18 (a.a. 124-135) indicates the basic region, and the orange block (a.a. 151-167) indicates the hydrophobic region. Each mKGN-ORF18 truncated mutant expression plasmid was co-transfected with ORF30-mKGC into 293T and HeLa cells by a lipofection method and three independent samples were assessed by flow cytometry. We utilized non-transfected cells as a negative control (Neg. Cont.). The positive control (Pos. Cont.) was cells co-transfected with p65-mKGN (pCONT-1) and p50-mKGC (pCONT-2) expression plasmids. The average of the mKusabira Green-positive cells (%) are noted beside the bars. Each bar and error bar indicate the average and standard deviation, respectively. The protein expression levels of the assessed mutants are shown in Supplemental Figure S2.

2.3. Alignment of KSHV ORF18 to its viral homologs

In order to survey the amino-acid residues of ORF18 which facilitate its binding to ORF30, whole amino-acid sequences of KSHV ORF18 were compared to its homologs from the betaherpesviruses (HCMV and HHV6B) as well as the gammaherpesviruses (MHV68 and EBV) (Figure 3). Conserved amino-acid sequences among all homologs or those exclusively conserved in gammaherpesviral homologs were indicated in the alignment. The regions (TR1-3 and TR8-10) identified in Figure2 are also illustrated. Based on this information, we selected several adjacent conserved amino-acid residues in the narrowed regions and designated them as blocks. Each of these 14 blocks (ASC1-ASC14) were subjected to alanine-scanning mutagenesis, where the conserved amino-acid residues in each block were substituted with alanine residues, which is denoted as ASC1mut-ASC14mut underneath the alignment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

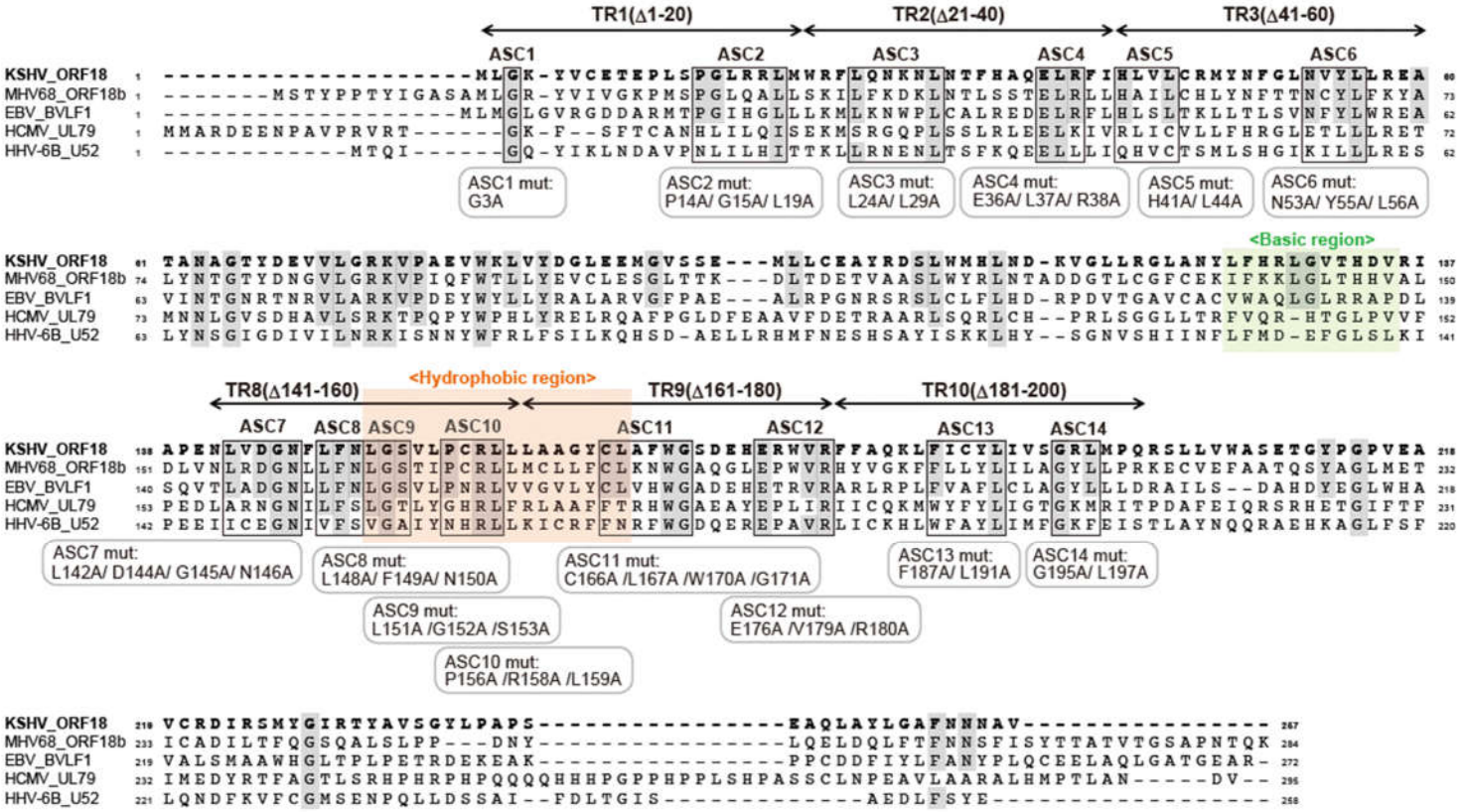

Figure 3. KSHV ORF18 amino-acid sequence alignment. Each herpesvirus homologous amino-acid was translated from the nucleotide sequences from the NCBI database [KSHV ORF18 (JSC-1-BAC16; Accession number GQ994935), MHV68 ORF18b (strain WUMS; NC_001826), EBV BVLF1 (strain Akata; KC207813), HCMV UL79 (strain Towne; FJ616285), and HHV-6B U52 (strain japan-a1; KY239023)]. The alignment raw data were obtained using Clustal Omega (EMBL-EBI; https://www.ebi.ac.uk/Tools/msa/clustalo/). The regions of TR1-3 and TR8-10 were essential for the interaction with ORF30 in the BiFC interaction assay (Figure 2). The conserved amino-acid residues between homologs, at least completely conserved in the gammaherpesvirus family (KSHV, MHV68, and EBV), are indicated with a gray background. Based on the amino-acid homology, several conserved amino-acid residues were split into blocks (ASC1-ASC14) and each conserved amino-acid was substituted with an alanine thus generating the alanine-scanning ORF18 mutants, ASC1mut-ASC14mut. These mutants contain the following substitutions: ASC1mut; ORF18 G3A, ASC2mut; ORF18 P14A/G15A/L19A, ASC3mut; ORF18 L24A/L29A, ASC4mut; ORF18 E36A/L37A/R38A, ASC5mut; ORF18 H41A/L44A, ASC6mut; ORF18 N53A/Y55A/L56A, ASC7mut; ORF18 L142A/D144A/G145A/N146A, ASC8mut; ORF18 L148A/F149A/N150A, ASC9mut; ORF18 L151A/G152A/S153A, ASC10mut; ORF18 P156A/R158A/L159A, ASC11mut; ORF18 C166A/L167A/W170A/G171A, ASC12mut; ORF18 E176A/V179A/R180A, ASC13mut; ORF18 F187A/L191A, and ASC14mut; ORF18 G195A/L197A. The basic region is shown in green, and the hydrophobic region is shown in orange.

2.4. Narrowing the ORF18 interaction blocks to single amino-acid residues

293T cells were co-transfected with plasmids expressing each ORF18 alanine-scanning mutant block [as indicated in the above experiment (Figure 3)] and mCherry expression plasmid (which was used as a transfection control). The transfected cells were subjected to flow cytometry. A red fluoro-protein, mCherry, was used as the transfection control for acquiring more accurate signal intensity. Harvested cells were initially gated by FSC as well as SSC and then mCherry-positive cells were also gated to extract the transfected cells. The ratio of mKG-positive cells was analyzed (Figure 4). The ORF18 block alanine-scanning mutants scoring under 20% were considered as binding-negative mutants because the negative control exhibited approximately 20% scoring in Fig4a. The substituted residues in ASC2mut, ASC3mut, ASC4mut, ASC5mut, and ASC11mut potentially contained essential residues for ORF30 interaction. The ASC10mut was not included

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

in the list of potential candidates because its protein expression was not confirmed by Western blotting (Supplemental Figure S3b). Another experiment without mCherry showed almost the same results (Supplemental Figure S3a).

Finally, plasmids encoding the ORF18 single alanine-scanning mutants were constructed. Every single conserved residue in the blocks of ASC2-5mut and ASC10-11mut were substituted with alanine. We included the ASC10mut in this analysis because the ASC10mut expression was deficient in transfected cells. Therefore, it was necessary to evaluate the residues in the ASC10mut block. ORF18 single alanine-scanning mutants scoring under 60% were considered as binding-negative mutants because the positive control and most mutants exhibited over 90% scoring. These serial results indicated that Leu(L)29, Gln(E)36, His(H)41, and Trp(W)170 were the responsible residues for ORF30 binding. On the other hand, affinity decreases were not observed in the single mutations of the ASC2 and ASC10 blocks [Pro(P)14, Gly(G)15, and Leu(L)19 of ASC2 and Pro(P)156, Arg(R)158, and Leu(L)159 of ASC10].

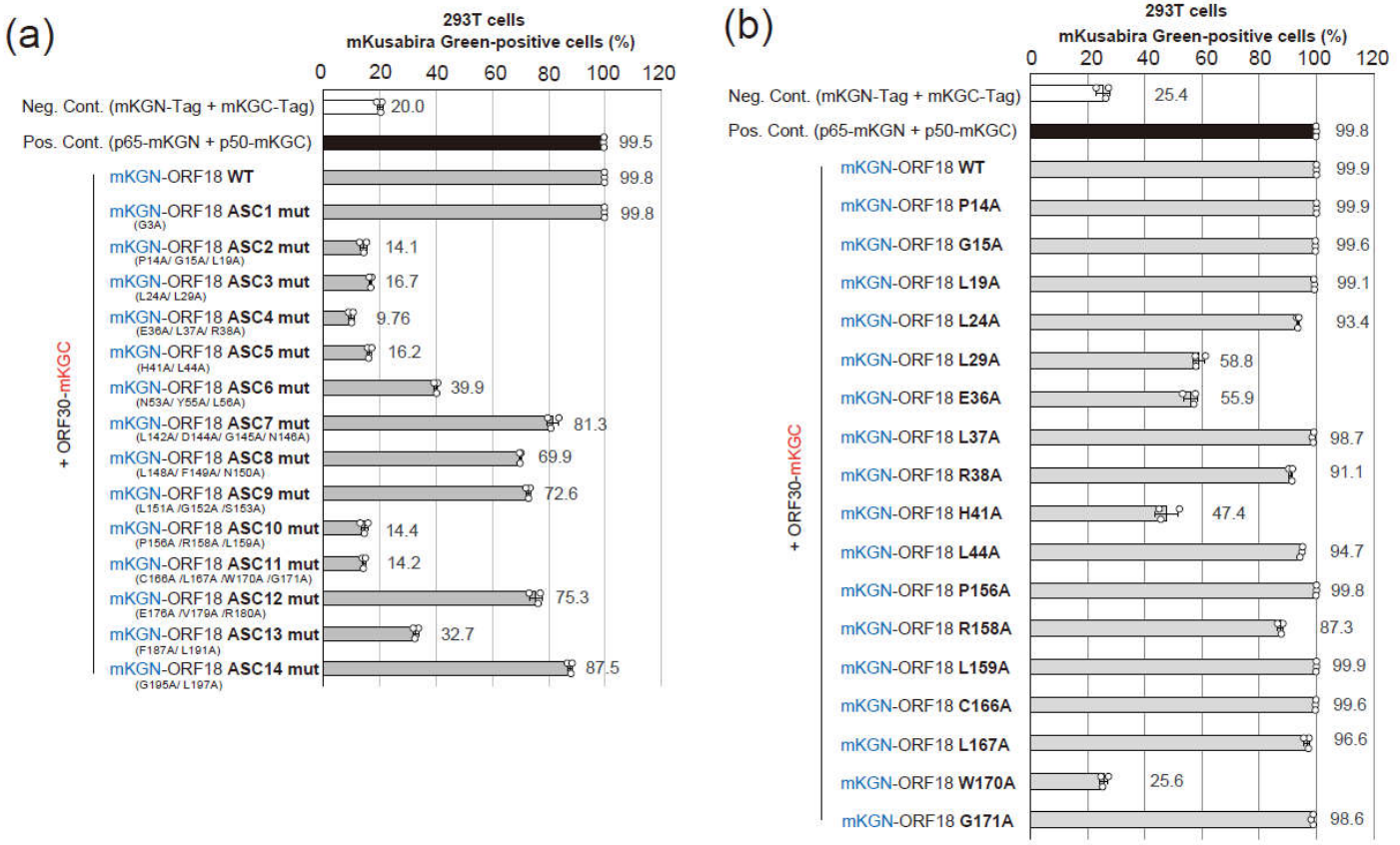

Figure 4. Identification of the ORF30-binding site on ORF18 using block alanine-scanning and single alanine-scanning mutants. (a) The indicated mKGN-ORF18 block alanine-scanning mutant (mKGN-ORF18 ASC1mut to ASC14mut) expression plasmid was co-transfected with ORF30-mKGC and mCherry, which was used as a transfection control. (b) The indicated mKGN-ORF18 single alanine-scanning mutant expression plasmid was co-transfected with ORF30-mKGC and mCherry, which was used as a transfection control. For both (a) and (b), 293T cells were transfected by a lipofection method and three independent samples were assessed by flow cytometry. The individual mutations that comprise the block alanine-scanning mutation are shown under the mKGN-ORF18 expression plasmid name. The negative control (Neg. Cont.) was a combination of tag-only expression plasmids (mKGN and mKGC), and the positive control (Pos. Cont.) was a combination of p65-mKGN (pCONT-1) and p50-mKGC (pCONT-2) expression plasmids. Each bar and error bar indicate the average and standard deviation, respectively. The mutant protein expression levels in the assay are shown in Supplemental Figure S4.

 doi:10.20944/preprints202206.0359.v1

2.5. AI-predicted structure model of ORF18 and the identified residues that mediate its interaction with ORF30

We attempted to predict the whole protein structure of KSHV ORF18 in order to gain further insight to function of the four amino-acid residues (L29, E36, H41, and W170) on ORF18 in the interaction with ORF30. To construct a structural model of ORF18, we used the AI deep learning algorithm AlphaFold2. Although we had no prior proof for the utility of AlphaFold2 to predict viral proteins, a predicted structure of KSHV ORF18 was obtained with a high confidence score (Figure 5a). This predicted structure model may be used to evaluate the biological characteristics of the interaction between ORF18 and ORF30. The ORF18 model suggested that all four amino-acid residues which we identified as being responsible for ORF30 binding are exposed to the surface of ORF18 and are located in proximity to each other. In particular, H41, W170, and L29 are in very close proximity and form a straight line on the predicted model surface of ORF18. Although E36 is not located on this H41-W170-L29-straight line, E36 and H41 are closely located on the same helix2 of ORF18 (a.a. 34-63) (Figures 5a-c and Supplemental Figure S6). Thus, our data obtained with the BiFC assay generally supported the correctness of the AI-predicted structural model. Furthermore, we predicted the ORF18-ORF30 complex model using AlphaFold2. AlphaFold2 has been frequently updated, and a multimer prediction algorithm has recently been implemented. The utilization of a multimer model is not yet fully established. Therefore, the ORF18-ORF30 complex model is introduced in the discussion section (Figures 6a-d and Supplemental Figures S7, S8, and S9).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

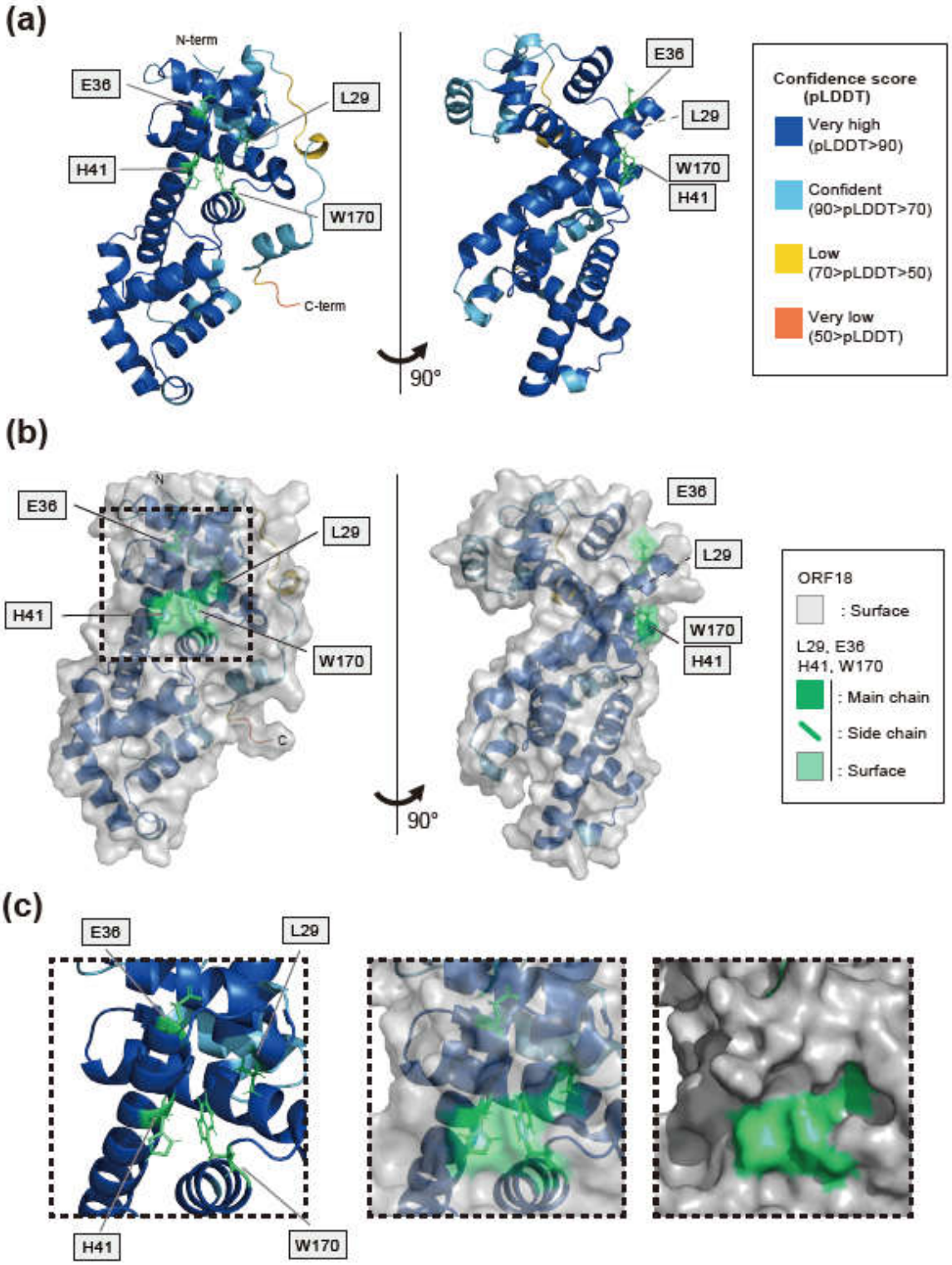

Figure 5. The predicted structure of ORF18 using the AlphaFold2 AI system with the ORF18 amino-acid residues responsible for the interaction with ORF30, as identified with a BiFC assay. (a) The whole predicted structure of ORF18 is shown as a cartoon model and two side angles. The cartoon color (blue to orange) indicates the confidence score (pLDDT) calculated by AlphaFold2. The L29, E36, H41, and W170 amino-acid side chains are shown as green stick models. (b) Two side views of the cartoon model and stick models in (a) were superimposed with the protein surface shown in gray. The surface-exposed L29, E36, H41, and W170 amino-acid side chains were visualized by green transparent shading. The square dotted line area indicates the enclosed regions in Fig. 5c. (c) The region enclosed within the dotted line in (b) is shown. The positions of the L29, E36, H41, and W170 amino-acid residues in the predicted model were enclosed with a cartoon (right panel),

 doi:10.20944/preprints202206.0359.v1

cartoon/transparent surface (center panel), and surface model (left panel). Various angles of the predicted models of ORF18 are shown in Supplemental Figure S6.

3. Discussion

In this study, four amino-acid residues of KSHV ORF18 required for its binding to ORF30 were identified using a BiFC assay (Figures 2-4). Additionally, we utilized the AI deep learning algorithm to predict the structure of the viral protein ORF18 without wet structural data (e.g., X-ray crystallography, NMR, and Cryo-TM). The predicted KSHV ORF18 model, which has a high confidence score, demonstrated that the four identified amino-acid residues responsible for the ORF18-ORF30 interaction were located close together on the surface of ORF18 (Figures 5a-c and Supplemental Figure S6). The AI-predicted structural models of ORF18 single-molecule and ORF18-ORF30 complex generally supported the molecular interaction data that was obtained with the BiFC assay.

Previous studies have shown that ORF18 and its gammaherpesvirus homologs are essential for viral replication [17-19]. Furthermore, the amino-acid residues responsible for KSHV ORF18 function have been identified [10]. Three ORF18 amino-acid residues, L29, E36, and W170 residue are involved in its binding to ORF30 [10]. These residues reported by other groups were consistent with the data from our study. These consistent results confirmed the utility of using the BiFC assay to analyze vPIC interactions. However, several interactions between ORF18 and other vPIC factors (e.g., ORF18-ORF34, ORF18-ORF66, and ORF18-ORF31) reported in previous studies, which were analyzed via split-luciferase and pull-down/immunoprecipitation assays, were not observed with our BiFC assay. Thus, this method may exhibit low sensitivity for weak interactions. Our serial strategy to determine the ORF30-binding site on ORF18 identified an additional novel amino-acid residue, H41 that was required for the ORF18-ORF30 interaction. H41 residue was likely detected because our single alanine substitution targets did not only include the amino-acid residues conserved among betaherpesviruses and gammaherpesviruses, but also among KSHV, MHV68 and EBV. Our BiFC experiments revealed two additional observations. First of all, the sequentially conserved amino-acid residues (P14/G15/L19 residues on the ASC2mut target) are responsible for the interaction of ORF18-ORF30, but mutation of each single amino-acid had no effect on the ORF18-ORF30 interaction (Figure 4). This result indicated that the partial structure of P14/G15/L19 residues in ORF18 contributed to its interaction with ORF30. Secondly, we did not detect any protein expression from the Ala-substituted mutant ASC10mut (P156A/R158A/L159A). It is not clear whether the lack of protein expression was due to a direct effect such as destabilization of the protein or an indirect effect such as reduced binding to ORF30 that resulted in a destabilization of the protein. Each single Ala-substituted mutation of P156, R158, or L159 expresses protein and retains the ORF30-binding ability. The requirement of several ORF18 mutations for the loss of ORF30 binding is speculated to provide ORF18 with a safety margin against the failure of binding due to a single missense mutation that may occur upon viral genome replication.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

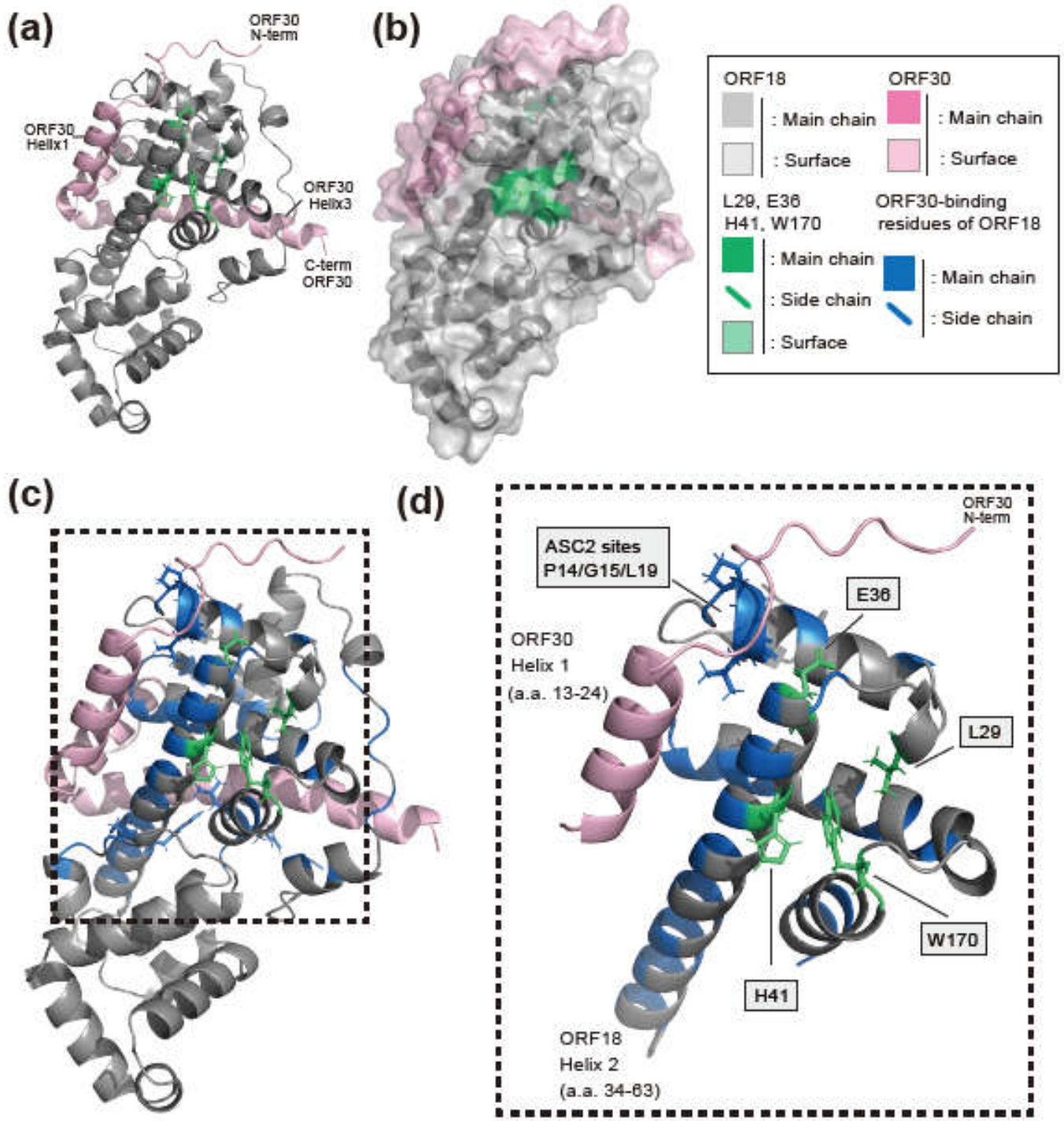

Figure 6. The predicted ORF18-ORF30 complex structure using the multimer prediction mode of AlphaFold2. (a) The predicted model is shown as a cartoon. The L29, E36, H41, and W170 amino-acid residues were visualized by green stick models. (b) The protein surface was superimposed onto the model shown in Fig. 6a. The ORF18 and ORF30 surfaces were visualized by transparent gray and pink shading, respectively. (c) The ORF18 amino-acid residues in contact with ORF30 are shown in blue. The square dotted line area indicates the enclosed regions in Fig. 6d. (d) The extracted structures are in close proximity to ORF30 helix1. The ASC2 mutation target residues (P14/G15/L19) are indicated as blue stick models. The L29, E36, H41, and W170 are indicated by green stick models. Various angles of the predicted ORF18-ORF30 complex with several markers are shown in Supplemental Figures S7, S8, and S9.

Our AlphaFold2 prediction model of KSHV ORF18 showed that the newly identified H41 residue is located on the same helix (helix2) with the previously identified E36

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

residue and directly contacts the previously identified W170 residue (Figures 5a-c and Supplemental Figure S6) [10]. Additionally, the AlphaFold2 prediction model of the KSHV ORF18-ORF30 complex revealed several insights into the interaction mode of this complex (Figures 6a-d and Supplemental Figures S7, S8, and S9). KSHV ORF30 is a short-length protein, and it is predicted to contain three helices. The model showed that all three helices of ORF30 fill the groove of ORF18 when ORF30 is bound to ORF18. The ORF30 N-terminus and helix1 (a.a. 13-24) are estimated to be in close proximity to our identified four residues of ORF18 (L29/E36/H41/W170) that are required for the ORF18-ORF30 interaction (Figure 6d). The ORF18 helix2 harboring E36 and H41 is located along and parallel with the ORF30 helix1. E36 is located on ORF18 helix2 and directly contacts the N-terminal tail of ORF30, but H41 does not contact the ORF30 helix1. However, it is possible that H41 contributes to the proper orientation of both ORF18 helix2 and ORF30 helix1. The H41 in ORF18 form a straight line with W170 and L29, which may support the ORF18 helix2 as a scaffold (Figure 6d). Direct contact of P14/G15/L19 (ASC2mut target) to the ORF30 N-terminal tail was also observed. The combined results of the AI prediction model and BiFC data provided further insight regarding the interaction residues of N53/Y55 (partial ASC6mut target) and R158/L159 (partial ASC10mut target). These residues may be involved in supporting the ORF30 helix3 (Supplemental Figure S9). The ORF18 ASC6mut exhibited mildly decreased binding to ORF30 and the ORF18 ASC10mut protein was not expressed in transfected cells (Figures 4a-b). The importance and interaction characteristics of E36 and R158/L159 were already reported [10] and agree well with our experimental data and prediction model. However, in order to reveal the real interactions between ORF18 and ORF30, characterization by X-ray crystallography is required.

AI-prediction protein modelling has been used in our vPIC analysis, for assessment of appropriateness the real experimental data observed by pull-down and Immuno-precipitation assays [11]. It has been accepted that homology modeling is a useful method for predicting protein structure. However, the prediction of viral proteins that lack related homologous structures may not be reliable because of the subjective nature of selecting the specific protein based on the prediction. Expansion of protein modelling in silico may lead to a paradigm shift in the study of protein structure in various life science fields, including virology. With the exclusion of intensely studied viruses, a massive amount of viral protein structures remain unknown. Thus, these innovations would enhance the data-driven study of the viral proteins whose structure is unknown.

In sum, we analyzed the interactions between KSHV ORF18 and the other components (ORF24, ORF30, ORF31, ORF34, and ORF66) of vPIC in living cells using a BiFC assay. As a result, apparent BiFC-positive signal was detected only between ORF18 and ORF30. The four amino-acid residues (L29, E36, H41, and W170) in ORF18 were necessary for its interaction with ORF30. Moreover, AI-predicted structure model of ORF18 suggested that these amino-acid residues are located in proximity to each other in order to form the surface structure of ORF18 required for binding to ORF30. Thus, AI-predicted model supported the importance of these four residues for binding of ORF18 to ORF30.

4. Materials and Methods

4.1. Cell culture and reagents

293T and HeLa cells (RCB2202 and RCB0007 respectively; RIKEN Bio Resource Center, Tsukuba, Japan) were cultured in DMEM supplemented with 10% fetal calf serum. The transfection reagent, PEI-MAX MW40000 (Polysciences, Inc., Warrington, PA, USA) was dissolved at a concentration of 2mg/mL in distilled water and filtered.

4.2. Plasmids

KSHV ORF18, ORF24, ORF30, ORF34, and ORF66 coding fragments were obtained by PCR from KSHV BAC16 (GQ994935; KSHV BAC clone based on strain JSC-1)[20] using primer sets noted in the Supplemental Table S1. These fragments were cloned into the phmKGN-MC, phmKGC-MC, phmKGN-MN, and phmKGC-MN vectors. These empty

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0359.v1

backbone vectors were included in the CoralHue™ Fluo-chase Kit (MBL, Nagoya, Japan). The KSHV ORF18 truncated mutants and alanine-scanning mutant coding fragments were obtained by 2-step PCR using primer sets noted in the Supplemental Table S1 and they were cloned into the phmKGN-MC vector. The mCherry coding fragment was obtained by PCR from pEF6.mCherry-TSG101[21] (Addgene plasmid # 38318; gifted from Dr. Quan Lu) using primer sets noted in the Supplemental Table S1 and it was cloned into the pCI-neo mammalian expression vector (Promega, Madison, WI, USA).

4.3. BiFC interaction assay

Our BiFC assays were conducted according to a previous method reported by Yoshida et al. [22]. Briefly, $3x10^5$ of 293T cells were seeded into each well of a 12 well plate. After an overnight incubation, the cells were transfected with mKGN-tagged protein expression plasmids and mKGC-tagged protein expression plasmids. The calcium phosphate method was used to transfect a total of 3 mg of plasmid (1.5 mg of each expression plasmid). The PEI (Polyethylenimine) method was also used to transfect a total of 1.6 mg of plasmid (0.8 mg of each expression plasmid or 0.5 mg of each expression plasmid and 0.6 mg of mCherry expression plasmid, which was utilized as a transfection control). After two days following the transfection, the cells were harvested and fixed with 1% formaldehyde in PBS and subjected to flow-cytometry analysis. In order to confirm the protein expression, cells transfected in parallel were used to prepare protein extracts, which were subjected to Western blotting.

4.4. Western blotting

Western blots were performed as described previously [8]. Anti-monomeric Kusabira-Green N-terminal fragment antibodies (M148-3M; MBL) and anti-actin antibodies (AC-15; Santa-Cruz, CA, USA) were used as the primary antibodies. HRP linked anti-mouse IgG antibody (GE healthcare UK Ltd., Buckinghamshire, UK) was used as the secondary antibody. Antibody-bound proteins were visualized with ECL Western Blotting Detection Reagents (GE healthcare UK Ltd.).

4.5. Protein structure prediction

The KSHV ORF18 protein structure model and the KSHV ORF18-ORF30 complex model were predicted using AlphaFold 2.1.2 and 2.2.0 (https://github.com/deepmind/alphafold) in its own local environment, respectively [16]. The following databases were utilized for AlphaFold2: Uniclust30 (version 2018_08), MGnify (version 2018_12), pdb70 (downloaded on Feb 15th, 2022), PDB/mmCIF (downloaded on Feb 18th, 2022), and pdb_seqres (downloaded on Feb 18th, 2022). It is noted that several versions or frequently updated versions of these databases were utilized. The predicted model was visualized with the molecular visualization open-source software, PyMOL. In addition, the pLDDT score coloring of the protein model was visualized with the python module PSICO (https://github.com/YoshitakaMo/pymol-psico, https://github.com/speleo3/pymol-psico).

Supplementary Materials:

Supplemental Figure S1: Screening and optimization of BiFC assays for the determination of vPIC component interactions.

Supplemental Figure S2: Confirmation of ORF18 truncated mutant protein expression.

Supplemental Figure S3: BiFC assay of ORF18 alanine-scanning block mutants and mutant protein expression.

Supplemental Figure S4: Confirmation of ORF18 single alanine mutant protein expression.

Supplemental Figure S5: Original Western blotting data from Supplemental Figures S2, S3, and S4.

Supplemental Figure S6: KSHV ORF18 structure model predicted by AlphaFold2.

Supplemental Figure S7: KSHV ORF18-ORF30 complex structure model predicted by AlphaFold2.

Supplemental Figure S8: KSHV ORF18-ORF30 complex structure model with ORF18 L29, E36, H41, and W170 amino-acid residues denoted as green stick models.

Supplemental Figure S9: KSHV ORF18-ORF30 complex structure model with ORF30-binding amino-acid residues of ORF18 highlighted.

Supplemental Table S1: Primers used for the construction of expression plasmids.

Author Contributions: Conceptualization, Y.M. and T.W.; methodology, Y.M. and T.W.; formal analysis, Y.M., T.W., K.K. and T.I.; investigation, Y.M.; data curation, T.W., T.I. and M.F.; writing—original draft preparation, T.W., Y.M., T.I., S.O. and M.F. ; writing—review and editing, T.W., T.I., S.O. and M.F.; visualization, Y.M., T.W. and T.I.; funding acquisition, T.W., T.I. and M.F. All authors have read and agreed to the published version of the manuscript. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation.

Funding: This research was partially funded by the JSPS Grant-in-Aid for Scientific Research (C) (grant number 18K06642, M.F.; grant number 21K08509, T.W.), the JSPS Grant-in-Aid for Young Scientists (grant number 18K14910, T.W.), and University of the Ryukyus Research Project Promotion Grant for Young Researchers (grant number 20SP04102, T.W.). This project was also supported in part by a start-up from Saint Joseph's University to T.I. and Advanced Medical Research Support Project from the University of the Ryukyus to T.W.

Data Availability Statement: Plasmid DNA sequence data are available from the corresponding author upon reasonable request. Other data are presented in the manuscript.

Acknowledgments: The BAC16, KSHV BAC clone, was kindly gifted from Dr. Jae U Jung (Lerner Research Institute of Cleveland Clinic, U.S.).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chang, Y.; Cesarman, E.; Pessin, M.S.; Lee, F.; Culpepper, J.; Knowles, D.M.; Moore, P.S. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **1994**, *266*, 1865-1869.
2. Nador, R.G.; Cesarman, E.; Chadburn, A.; Dawson, D.B.; Ansari, M.Q.; Sald, J.; Knowles, D.M. Primary effusion lymphoma: a distinct clinicopathologic entity associated with the Kaposi's sarcoma-associated herpes virus. *Blood* **1996**, *88*, 645-656.
3. Soulier, J.; Grollet, L.; Oksenhendler, E.; Cacoub, P.; Cazals-Hatem, D.; Babinet, P.; d'Agay, M.F.; Clauvel, J.P.; Raphael, M.; Degos, L.; et al. Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. *Blood* **1995**, *86*, 1276-1280.
4. Russo, J.J.; Bohenzky, R.A.; Chien, M.C.; Chen, J.; Yan, M.; Maddalena, D.; Parry, J.P.; Peruzzi, D.; Edelman, I.S.; Chang, Y.; et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 14862-14867, doi:10.1073/pnas.93.25.14862.
5. Krug, L.T.; Pellet, P.E. *The family Herpesviridae: A brief introduction*; Damania, B.A., Cohen, J.I., Eds.; Lippincott Williams & Wilkins: 2022; Volume 2.
6. Gruffat, H.; Marchione, R.; Manet, E. Herpesvirus Late Gene Expression: A Viral-Specific Pre-initiation Complex Is Key. *Front. Microbiol.* **2016**, *7*, 869-869, doi:10.3389/fmicb.2016.00869.
7. Davis, Z.H.; Hesser, C.; Park, J.; Glaunsinger, B.A. Interaction Between ORF24 and ORF34 in the Kaposi's Sarcoma-Associated Herpesvirus Late Gene Transcription Factor Complex is Essential For Viral Late Gene Expression. *J. Virol.* **2015**, doi:10.1128/jvi.02157-15.
8. Nishimura, M.; Watanabe, T.; Yagi, S.; Yamanaka, T.; Fujimuro, M. Kaposi's sarcoma-associated herpesvirus ORF34 is essential for late gene expression and virus production. *Sci. Rep.* **2017**, *7*, 329, doi:10.1038/s41598-017-00401-7.
9. Brulois, K.; Wong, L.Y.; Lee, H.R.; Sivadas, P.; Ensser, A.; Feng, P.; Gao, S.J.; Toth, Z.; Jung, J.U. Association of Kaposi's Sarcoma-Associated Herpesvirus ORF31 with ORF34 and ORF24 Is Critical for Late Gene Expression. *J. Virol.* **2015**, *89*, 6148-6154, doi:10.1128/jvi.00272-15.
10. Castañeda, A.F.; Glaunsinger, B.A. The Interaction between ORF18 and ORF30 Is Required for Late Gene Expression in Kaposi's Sarcoma-Associated Herpesvirus. *J. Virol.* **2019**, *93*, e01488-01418, doi:10.1128/jvi.01488-18.
11. Watanabe, T.; Nishimura, M.; Izumi, T.; Kuriyama, K.; Iwaisako, Y.; Hosokawa, K.; Takaori-Kondo, A.; Fujimuro, M. Kaposi's Sarcoma-Associated Herpesvirus ORF66 Is Essential for Late Gene Expression and Virus Production via Interaction with ORF34. *J. Virol.* **2020**, *94*, doi:10.1128/jvi.01300-19.

12. Didychuk, A.L.; Castaneda, A.F.; Kushnir, L.O.; Huang, C.J.; Glaunsinger, B.A. Conserved CxnC Motifs in Kaposi's Sarcoma-Associated Herpesvirus ORF66 Are Required for Viral Late Gene Expression and Are Essential for Its Interaction with ORF34. *J. Virol.* **2020**, *94*, doi:10.1128/JVI.01299-19.
13. Kodama, Y.; Hu, C.D. Bimolecular fluorescence complementation (BiFC): a 5-year update and future perspectives. *Biotechniques* **2012**, *53*, 285-298, doi:10.2144/000113943.
14. David, A.; Islam, S.; Tankhilevich, E.; Sternberg, M.J.E. The AlphaFold Database of Protein Structures: A Biologist's Guide. *Journal of molecular biology* **2022**, *434*, 167336, doi:10.1016/j.jmb.2021.167336.
15. Fersht, A.R. AlphaFold - A Personal Perspective on the Impact of Machine Learning. *Journal of molecular biology* **2021**, *433*, 167088, doi:10.1016/j.jmb.2021.167088.
16. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583-589, doi:10.1038/s41586-021-03819-2.
17. Gong, D.; Wu, N.C.; Xie, Y.; Feng, J.; Tong, L.; Brulois, K.F.; Luan, H.; Du, Y.; Jung, J.U.; Wang, C.Y.; et al. Kaposi's sarcoma-associated herpesvirus ORF18 and ORF30 are essential for late gene expression during lytic replication. *J. Virol.* **2014**, *88*, 11369-11382, doi:10.1128/jvi.00793-14.
18. Arumugaswami, V.; Wu, T.T.; Martinez-Guzman, D.; Jia, Q.; Deng, H.; Reyes, N.; Sun, R. ORF18 is a transfactor that is essential for late gene transcription of a gammaherpesvirus. *J. Virol.* **2006**, *80*, 9730-9740, doi:10.1128/JVI.00246-06.
19. Wu, T.-T.; Park, T.; Kim, H.; Tran, T.; Tong, L.; Martinez-Guzman, D.; Reyes, N.; Deng, H.; Sun, R. ORF30 and ORF34 are essential for expression of late genes in Murine Gammaherpesvirus 68. *J. Virol.* **2009**, *83*, 2265-2273, doi:10.1128/JVI.01785-08.
20. Brulois, K.F.; Chang, H.; Lee, A.S.; Ensser, A.; Wong, L.Y.; Toth, Z.; Lee, S.H.; Lee, H.R.; Myoung, J.; Ganem, D.; et al. Construction and manipulation of a new Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone. *J. Virol.* **2012**, *86*, 9708-9720, doi:10.1128/jvi.01019-12.
21. Nabhan, J.F.; Hu, R.; Oh, R.S.; Cohen, S.N.; Lu, Q. Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4146-4151, doi:10.1073/pnas.1200448109.
22. Yoshida, T.; Ebina, H.; Koyanagi, Y. N-linked glycan-dependent interaction of CD63 with CXCR4 at the Golgi apparatus induces downregulation of CXCR4. *Microbiol. Immunol.* **2009**, *53*, 629-635, doi:10.1111/j.1348-0421.2009.00167.x.