An in silico prediction of interaction models of influenza a virus PA and human C14orf166 protein from yeast-two-hybrid screening data

Kadir TURAN[1] and Elif Çağlayan[2]

[1]Marmara Universitesi Eczacilik Fakultesi
[2]Kartal Kosuyolu Yuksek Ihtisas Egitim ve Arastirma Hastanesi

August 18, 2022

Abstract

The human C14orf166 protein, also known as RTRAF, shows positive modulatory activity on the cellular RNA polymerase II enzyme. This protein is a component of the tRNA-splicing ligase complex and is involved in RNA metabolism. It also functions in the nucleo-cytoplasmic transport of RNA molecules. The C14orf166 protein has been reported to be associated with some types of cancer. It has been shown that the C14orf166 protein binds to the influenza A virus RNA polymerase PA subunit and has a stimulating effect on viral replication. In this study, candidate interactor proteins for influenza A virus PA protein were screened with a Y2H assay using HEK293 Matchmaker cDNA. The C14orf166 protein fragments in different sizes were found to interact with the PA. The three-dimensional structures of the viral PA and C14orf166 proteins interacting with the PA were generated using the I-TASSER algorithm. The interaction models between these proteins were predicted with the ClusPro protein docking algorithm and analyzed with PyMol software. The results revealed that the carboxy-terminal end of the C14orf166 protein is involved in this interaction, and it is highly possible that it binds to the carboxy-terminal of the PA protein. Although amino acid residues in the interaction area of the PA protein with the C14orf166 showed distribution from 450th to 700th position, the intense interaction region was revealed to be at amino acid positions 610 to 630.

An *in silico* prediction of interaction models of influenza a virus PA and human C14orf166 protein from yeast-two-hybrid screening data

Running Title: C14ORF166 - PA

Elif Çağlayan[1] and Kadir TURAN[2]*

[1]University of Health Sciences Kartal Koşuyolu High Speciality Educational and Research Hospital, Istanbul, Turkey

[2]Department of Basic Pharmaceutical Sciences, Faculty of Pharmacy, Marmara University, Istanbul, Turkey

*Correspondence: kadirturan@marmara.edu.tr

ORCID

Elif ÇAĞLAYAN: https://orcid.org/0000-0002-4249-0824

Kadir TURAN: https://orcid.org/0000-0003-4175-3423

Abstract

The human C14orf166 protein, also known as RTRAF, shows positive modulatory activity on the cellular RNA polymerase II enzyme. This protein is a component of the tRNA-splicing ligase complex and is involved

in RNA metabolism. It also functions in the nucleo-cytoplasmic transport of RNA molecules. The C14orf166 protein has been reported to be associated with some types of cancer. It has been shown that the C14orf166 protein binds to the influenza A virus RNA polymerase PA subunit and has a stimulating effect on viral replication. In this study, candidate interactor proteins for influenza A virus PA protein were screened with a Y2H assay using HEK293 Matchmaker cDNA. The C14orf166 protein fragments in different sizes were found to interact with the PA. The three-dimensional structures of the viral PA and C14orf166 proteins interacting with the PA were generated using the I-TASSER algorithm. The interaction models between these proteins were predicted with the ClusPro protein docking algorithm and analyzed with PyMol software. The results revealed that the carboxy-terminal end of the C14orf166 protein is involved in this interaction, and it is highly possible that it binds to the carboxy-terminal of the PA protein. Although amino acid residues in the interaction area of the PA protein with the C14orf166 showed distribution from 450th to 700th position, the intense interaction region was revealed to be at amino acid positions 610 to 630.

INTRODUCTION

Influenza A viruses are among the most common causes of human respiratory diseases which cause infections with high mortality, especially in the elderly, infants, and people with chronic diseases and weakened immune systems[1]. These viruses have an RNA genome consisting of eight single-stranded segments with negative polarity [2,3]. Replication and transcription of the viral genome are catalyzed by the RNA-dependent RNA polymerase (RdRP) enzyme, which is a complex composed of PB2, PB1, and PA subunits[4-6]. The RdRP enzyme resides in the virus particle attached to the vRNP complexes consisting of the NP proteins and RNA molecules. The PA subunit of the RdRP enzyme has endonuclease activity, binds to the 5'-cap of the host pre-mRNAs along with the PB2 subunit, and cleaves. It has a role in binding to the genomic RNA promoter along with the PB1 subunit [7]. It is known that the PA protein, which is effective in viral pathogenesis, is associated with several cellular protein factors. Some of the PA protein-related proteins have a negative regulatory effect on influenza virus replication including HAX1 [8] and SNX2[9]; however, some of the PA-associated cellular proteins stimulate viral replication. It has been reported that the pyruvate kinase M2 (PKM2) binds the c-terminal region of the PA subunit and is essential for virus replication[10]. Kawaguchi & Nagata[11] suggest that the minichromosome maintenance protein complex (MCM), a DNA helicase, interacts with the influenza A virus PA protein and stimulates virus replication. The human *Chromosome 14 Open Reading Frame 166* (C14orf166), which is the subject of this study, is a stimulatory factor for the influenza A virus[12]. This protein is encoded by a gene located on chromosome 14 (14q22.1) and has an average weight of 28 kDa. The C14orf166 is also known as RTRAF, CLE, CLE7, hCLE, CGI99, RLLM1, hCLE1, CGI-99, and LCRP369. In recent reports, it was shown that C14orf166 is a member of nucleo-cytosolic shuttle protein complexes involving DDX1, HSPC117, and FAM98B proteins and has a role in the transport of the molecules involved in RNA metabolism between nucleus and cytoplasm[13]. This protein is associated with several factors required for RNA synthesis and processing, including transcription factor 4, heterogeneous nuclear ribonucleoprotein R, poly A binding protein 1, and the nuclear pore complex Nup153[14-16]. C14orf166 is not only involved in the regulation of RNA polymerase activity but is also upregulated in some tumors. Overexpression of the C14orf166 has been found to contribute to oncogenesis and invasive behaviors in various tumors. C14orf166 is a JAK2-related protein that activates STAT3 signaling which may cause cervical and esophageal cancer[17,18]. It has also been known that expression of the C14orf166 is upregulated in breast cancer cell and tissues compared with normal breast tissues. The overexpression of C14orf166 inhibits the expression of cell cycle inhibitors p21 and p27 and increases the phosphorylation level of the Rb protein[19]. From these results, it is suggested that C14orf166 contributes to cell proliferation by regulating the G1/S transition of the cells, but its mechanism has not yet been fully elucidated. Furthermore, the C14orf166 interacts with RNA polymerase II and directly regulates RNA transcription, suggesting that C14orf166 has a vital role in cell growth and organ development [20]. It has been reported that the C14orf166 protein is also associated with some virus replication/transcription processes. Huarte et al. reported the interaction of the C14orf166 protein with influenza A virus PA[21]. Silencing of the *C14orf166* gene expression results

in decreased vRNA transcription/replication, and virus production[12]. The C14orf166 protein involves in acute/chronic hepatitis C virus (HCV) infection by interacting with HCVc174, a core protein of HCV. It is thought that the C14orf166/HCVc174 complex may cause abnormal mitosis of infected hepatocytes, resulting in hepatic carcinoma[22]. These reports show that the c14orf166 protein is a cellular factor involved in the replication of some viruses as well as some cancer types in humans. In this study, full-size and truncated C14orf166 proteins interacting with the influenza A virus PA protein that is used as bait in the yeast two-hybrid (Y2H) screening were defined, and the interaction patterns of these protein with PA were analyzed by in silico tools. It was concluded that some amino acid residues located at the carboxyl terminal end of C14orf166 are significant in the PA protein interaction.

MATERIAL AND METHODS

Yeast cells

The yeast strain Saccharomyces cerevisiae PJ69-4A was used to detect the proteins interacting with the influenza A virus PA protein by the yeast dual-hybrid method (Genotype: MATa trp1-901 leu2-3, 112 ura3-52 his3-200 gal4(deletion) gal80(deletion) LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ). The cells were grown in YPAD (yeast extract- peptone-dextrose plus adenine) rich medium at 30°C at 175 rpm shaking and stored at -80°C in a YPD (yeast extract peptone dextrose) medium containing 20% glycerin.

cDNA library

The cDNA library constructed by cloning human embryonic kidney cells (HEK293) cDNA after the sequence encoding the yeast GAL4 activation domain (GAL4-AD) in pAKT2 Y2H plasmid (pAKT2-cDNAn) was purchased as transformed in *E. coli* cells (BNN132) (Clontech #638826)). The cells were grown on LB agar (+amp) plates according to the manufacturer's instructions, and the plasmid DNA was isolated.

Plasmids

Construction of the pGBD-PA plasmid vector encoding the viral PA as a bait protein has been previously described[9]. Briefly, the pGBD-C1 plasmid DNA was digested with EcoRI restriction enzyme just after the sequence coding the yeast GAL4 DNA binding domain (GAL4-BD), was blunted with the Klenow enzyme (New England Biolabs, UK), and then was dephosphorylated with shrimp alkaline phosphatase (*Thermo Fisher Scientific*, USA). The influenza A/duck/Pennsylvania/10,218/84/H5N2 (DkPen) virus PA gene open reading frame (ORF) was generated from the pCAGGS-PA (DkPen) plasmid[23] with PCR by using phosphorylated oligonucleotide primers 5'-CGGAGGATCTGGA*ATG* GAAGACTTTGTGCGACAATG-3' and 5'-CTATTTCAGTGCATGTGCGAG-3'. The PCR was carried out with high-fidelity KOD DNA polymerase. The PCR amplified DNA was purified with a gel extraction kit (Invitrogen # K210012, USA) and ligated with linear pGBD-C1 using a T4 DNA ligase kit (Ligation High v.II, TaKaRa, Japan). To construct a mammalian expression vector coding the human C14orf166 protein, the ORF of the gene was cloned into the EcoRV site of pCHA plasmid DNA[24]. C14orf166 ORF was generated with PCR using the Y2H cDNA library as a template and phosphorylated primers having 5'-ATC*ATG* TTCCGACGCAAGTTG-3' and 5'-AT*CTA* TCTTCCAACTTTTCCCAG-3' sequences. The PCR amplified DNA was purified by a gel extraction kit and ligated with the pCHA plasmid which was digested with EcoRV restriction enzyme (New England Biolabs, UK) and dephosphorylated with SAP. The resulting plasmid was named pCHA-C14orf166. The plasmid constructs were verified by colony PCR, restriction digestion, and finally checked by Sanger sequencing.

Yeast two-hybrid screening

A small-scale yeast culture (5 ml) was prepared in the YPAD medium and transformed with the pGBD-PA plasmid DNA coding the PA bait protein by the lithium acetate/polyethylene glycol (LiAc/PEG) method. Transformants were selected on a yeast synthetic drop-out (SD) agar medium (without Trp). One of the colonies harboring the bait plasmid was grown in the YPAD medium, transformed with the Y2H M Matchmaker cDNA library and seeded on SD agar plates (without Her, Leu, Ade, and Trp) for selection of positive colonies.

The yeast colonies grown on the selective medium were tested for reporter β-galactosidase activity, which is encoded from the second reporter gene. The plasmids carrying cDNA were isolated from the colonies having high β-galactosidase activity with a yeast plasmid DNA extraction kit. The cDNAs carried on the plasmids were amplified with PCR using the oligonucleotide primers 5'-AATACCACTACAATGGATGATGT-3' and 5'-CCAAGATTGAAACTTAGAGGAGT-3'. The PCR products were purified with a gel extraction kit and the sequence of the cDNAs was defined with Sanger sequencing. DNA sequence results were evaluated with the Basic Local Alignment Search Tool (BLAST), and the genes were identified.

β-γαλαχτοσιδασε ασσαψ

The yeast cells in a 250 μl of saturated yeast cultures were precipitated with centrifugation at 3500 rpm for 5 minutes, and the precipitates were suspended in 300 μl of Z buffer/β-ME (100 mM Phosphate Buffer, pH. 7, 10 mM KCl, 1mM $MgSO_4$, 50 mM 2-β-Mercaptoethanol). The samples were subjected to 10 cycles of freezing (in liquid nitrogen) and thawing (at 37°C). Then, 60 μl of an ONPG substrate solution (4 mg/ml) was added to the samples and incubated at 30°C for 60 minutes. After incubation, the reaction was terminated by adding 300 μl of 0.5 M Na_2CO_3 to the samples. The cell debris was precipitated with centrifugation at 15000 rpm for 5 minutes, and OD_{420} of the supernatants was determined.

Minireplicon assay

The effect of C14orf166 overexpression on the human type influenza A/WSN/33/H1N1 (WSN) and the avian type DkPen virus RNA polymerase enzyme was investigated with a minireplicon model in HEK293 cells. The assays were performed using minireplicon plasmid DNAs, as previously described[23,25].

In silico predictions

The shortest C14orf166 peptide (consisting of 69 amino acid residues at the carboxy-terminal end) having a positive interaction with the viral PA protein in yeast cells was used as a basic structure for the in silico prediction of three-dimensional (3D) structures and protein interactions. The 3D models of the peptide/proteins were generated with the online I-TASSER algorithm (*https://zhanggroup.org/I-TASSER/*)[26]. The best-predicted models were determined by using the c-score [-5 - +2], TM-score (>0.5), and root-mean-square deviation (RMSD) that were used to predict the quality of the models. A reference 3D model of C14orf166 was received in PDB format from the AlphaFold Protein Structure Database*(https://alphafold.ebi.ac.uk/entry/Q9Y224)*. A homology model of the influenza A DkPen PA protein predicted with the I-TASSER algorithm was used as a reference model in all protein-protein docking analyses. The docking analyses were carried out using the ClusPro protein docking algorithm (*https://www.cluspro.org*)[27]. The most likely C14orf166/PA interaction models were defined by considering the experimental data of the Y2H screening, the reports of the structure of the influenza A virus RdRP enzyme complex consisting of PB2-PB1 and PA proteins and 3D models of this complex[28] (PDB # 6QPF), binding energies and/or clustering rates revealed by docking algorithms, and amino acid residues with possible binding potential in the models. PyMOL software and an online PDBsum server (*http://www.ebi.ac.uk*) were used in the analysis of selected models for PA and C14orf166 interaction[29].

RESULTS

Identification of human C14orf166 proteins interacting with influenza a virus PA protein in yeast cells

Several host proteins that interact with influenza virus PA were identified using the Y2H screening assay carried out with transformation of the HEK293 Matchmaker cDNA library into the yeast cells harboring the pGBD-PA bait plasmid. Among the cDNAs isolated from the positive yeast colonies, the most often encountered cDNA belonged to the*C14orf166* gene. It was also noteworthy that the bait proteins with different sizes encoded from differently truncated C14orf166 cDNAs gave positive reactions to the viral PA protein. From the sequencing results of cDNAs isolated from the yeast cells, seven different sizes of C14orf166 cDNA were defined (four samples are given in Figure 1).

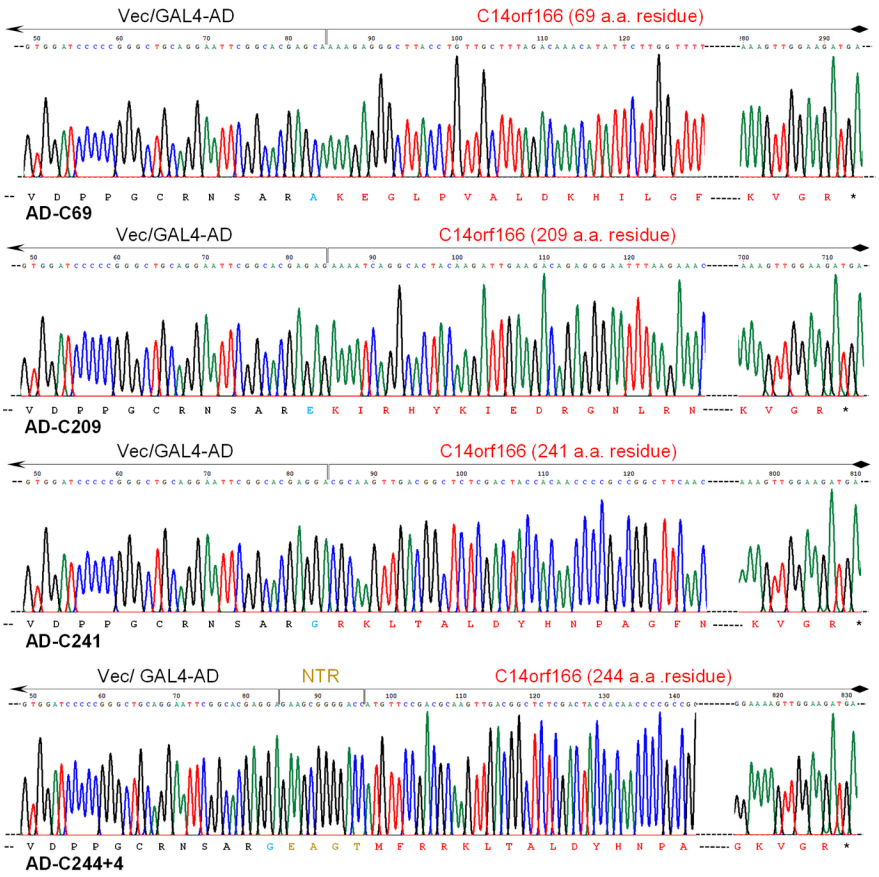


FIGURE 1. Sanger sequencing chromatograms of *C14orf166*cDNAs fused with GAL4-AD isolated from the yeast cells and amino acid sequences of the proteins encoded from these cDNAs. **NTR** , non-translated region of *C14orf166* gene.

Therefore, in this work, we focused on the interaction patterns of human C14orf166 with the viral PA protein. Representative structures of the proteins with different sizes encoded from the cDNAs are shown in Figure 2A. It was determined that some of the cDNAs encode the C14orf166 carboxy-terminal region, which consists of only 69 amino acid residues fused with GAL4-AD (AD-C69). Some of the cDNAs encode the C14orf166 protein fused with GAL4-AD, which is composed of 209 amino acids lacking 35 residues (AD-C209), while some encode the protein lacking only three amino acids at the amino-terminal end (AD-C14orf166/241). Sequencing results showed that some of the cDNAs encode the full-size C14orf166 proteins having extra regions at the amino-terminal end consisting of 4 or more irrelevant amino acid residues corresponding to the non-translated region (NTR) of the gene before the initial methionine residue in the protein (AD-C244+4). Although in different sizes, the C14orf166 proteins fusing with GAL4-AD caused an increase in reporter β-galactosidase activity by interacting with bait PA protein in the yeast cells (Figure 2B).

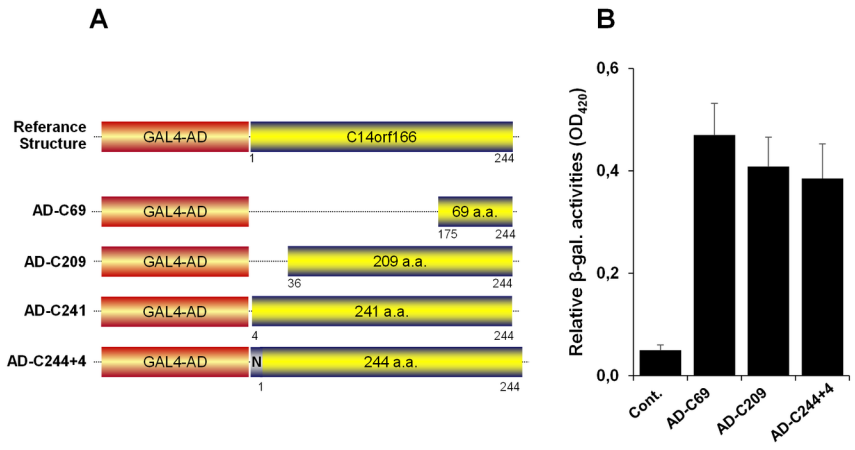


FIGURE 2 . Different sizes of C14odf166 proteins in fusion with yeast GAL4-AD, encoded by C14orf166 cDNAs isolated from positive yeast colonies (**A**), and reporter β-galactosidase activities measured in the yeast cells harboring these proteins along with the GAL4-BD. PA bait protein (**B**). **N** , amino acid residues corresponding to the non-translated region of the gene.

The Y2H screening assay was also performed using a bait protein consisting of the amino-terminal moiety (359 amino acid residues) of influenza A virus PA fused to yeast GAL4-BD as bait (data not shown). Although many positive colonies were obtained on the SD selective medium, no colonies carrying C14orf166 cDNA were found among them, suggesting that the amino-terminal half of the PA protein was not a target for the C14orf166 protein.

Predicted models for viral PA protein and C14orf166 interaction

Protein-protein interactions allow complex metabolic events to take place harmoniously in the cells. Therefore, determining the interactions between proteins and the interaction models allows us to understand biological events. Deciphering the interrelationships of viral and cellular proteins is of vital importance for revealing the replication/transcriptional mechanisms of viruses, which depend on the host cells for all biosynthesis events, and developing strategies to cope with pathogenic viruses. However, it is quite difficult to predict the three-dimensional (3D) structures of proteins based on their primary structures and protein-protein interactions. Peptide bonds between amino acids are strong and highly rigid chemical bonds. In contrast, the bonds connecting the alpha carbon (α-carbon) of the amino acids to the carboxyl group (-COOH) and the amino group (-NH2) are very flexible, and the polypeptide can rotate freely around these two bonds[30]. The flexibility of these groups allows variable three-dimensional structures for a protein containing 20 different amino acids and makes predicting the folding patterns much more complicated, depending on the size of the protein. Therefore, the carboxy-terminal part of the C14orf166 protein covering 69 amino acid residues (C69) which interacts with the viral PA in the yeast cells was used as the basic structure for the prediction of the PA protein interaction patterns. *De novo* 3D fold models of the C69 peptide were generated using an online I-TASSER server. The top three models with the highest C-score are given in Figure 3. Despite bending in different directions, the most prominent structural features of the models are the alpha helix structure at the carboxy-terminal ends.

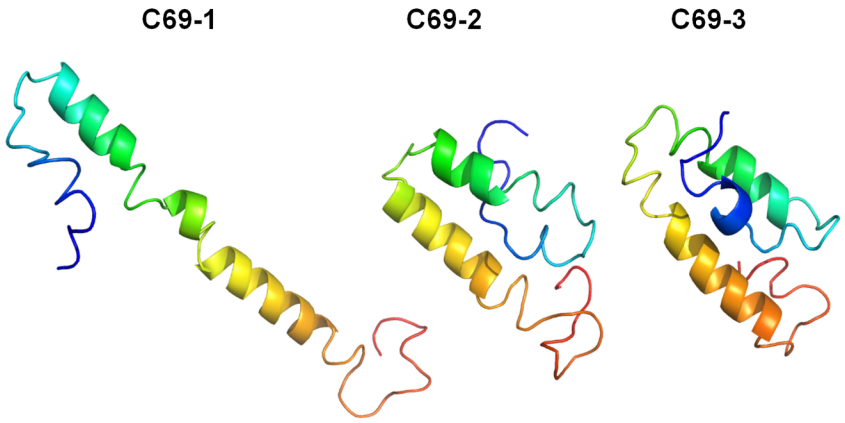


FIGURE 3. *De novo* 3D models of the carboxyl-terminal end of C14orf166 consisting of 69 amino acid residues (C69) determined with the I-TASSER algorithm. The carboxyl-terminal and amino-terminal ends of the peptides are shown in red and blue, respectively. C-scores of the peptides are C69-1: -3.05, C69-2: -2.78, C69-3: -3.24.

The C14orf166 3D, a reference model predicted by AlphaFold, also has a longer alpha helix structure at the c-terminal end (Figure 4A). The c-terminal end of this model has a similar folding pattern to the model given in Figure 3, C69-1. The homology model of influenza A DkPen virus PA protein generated with I-TASSER was used as a reference model for PA:C14orf166 docking analyses (Figure 4B).

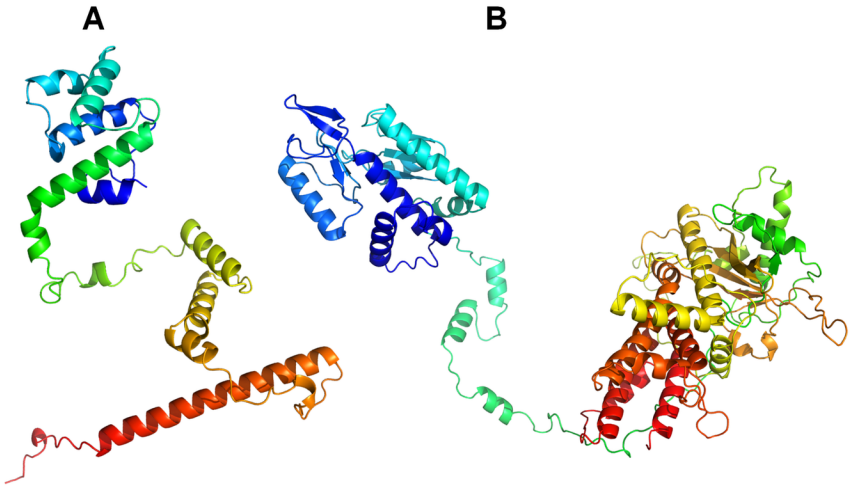


FIGURE 4. 3D models of C14orf166 (**A**) and influenza DkPen PA proteins (**B**) predicted by AlphaFold (C14orf166 # Q9Y224) and I-TASSER algorithm, respectively. The carboxyl-terminal and amino-terminal ends of the proteins are shown in red and blue, respectively.

Model C69-1 (Figure 3), the c-terminal peptide of C14orf166 consisting of 69 amino acid residues and the model of the DkPen virus PA protein (Figure 4B) were matched with the online protein-protein docking server using the ClusPro docking algorithm. Among the many prominent interaction models, considering the influenza A virus RdRP complex structure [28] (PDB # 6QPF) (the regions of the PA protein interacting with PB2 and PB1 were eliminated) and the inability to detect C14orf166 among the host proteins interacting with the amino-terminal of viral PA in the Y2H screening assay (in which we used amino-terminal half of

the PA protein as a bait-data not shown), the region with the highest accumulation for C69-1 on the PA protein was determined. It was observed that model C69-1 accumulates in the carboxyl-terminal half of the viral PA protein involving amino acid residues mostly between positions 610 to 630 in most of the docking models as shown in Figure 5. One salt bridge and 11 hydrogen bonds were defined in the interaction area of the models.

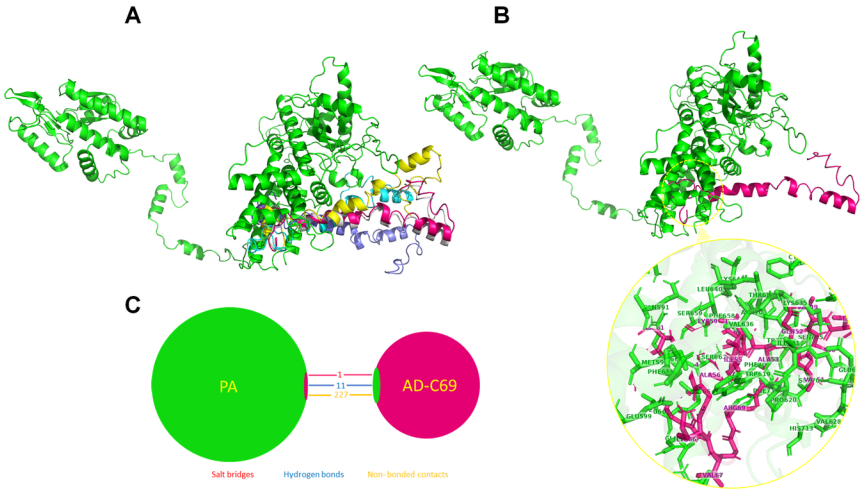


FIGURE 5 . 3D models of influenza virus PA protein and C69 peptide, and selected docking models. **A** : several C69 peptides clustered on the PA protein, **B** : single C69 peptide-PA model (the interaction area was presented in enlarged circle), **C:**chemical bounds defined between viral PA and C69 peptide with PDBsum algorithm.

In the Y2H method, the candidate interactor protein is synthesized in the cells as fused with the activation domain (AD) of the yeast GAL4 transcription activator. Therefore, the 3D models of the C69 peptide fused with GAL4 AD (AD-C69) (Figure 2) were predicted by I-TASSAR. Two models with higher c-scores were subjected to docking analysis with the PA protein by ignoring the AD of the protein. In many interaction patterns, the AD-C69 protein was found to be clustered in the same area as the C69 peptide (Figure 5) on the bait PA protein in both docking patterns (Figures 6A and 6C). Figure 6B and 6D show the predicted interaction patterns between a single AD-C69 protein and viral PA.

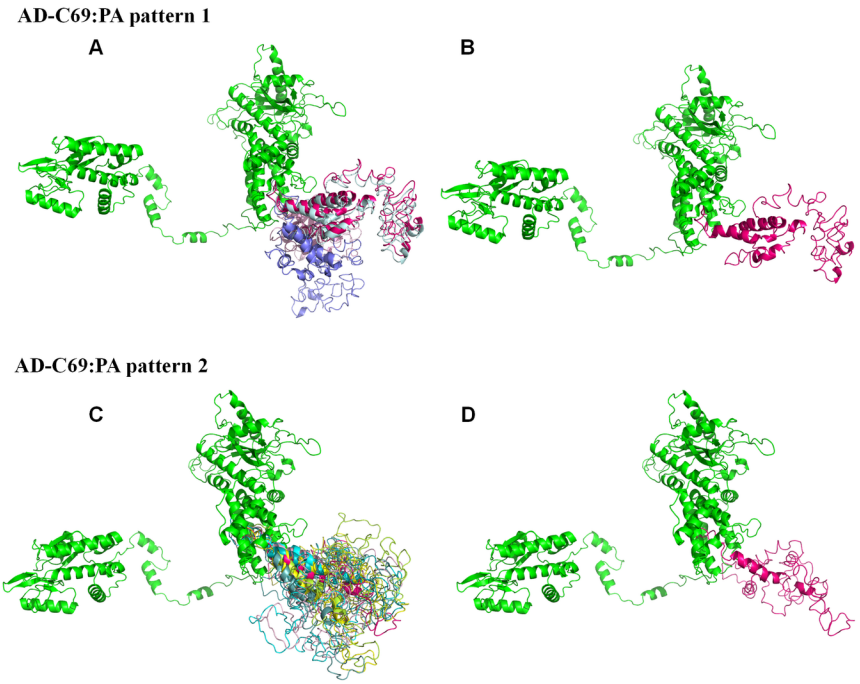


FIGURE 6. 3D models of influenza virus PA and AD-C69 proteins, and selected two docking patterns. **A/C** : several AD-C69 proteins clustered on the PA model, **B/D** : single AD-C69 protein-PA model.

As seen in Figure 2, some of the C14orf166 proteins interacting with the bait PA protein in the yeast cells consisted of 209 amino acid residues fused with GAL4-AD (AD-C209) while some were full-size (AD-C244+4) and carried additional amino acid residues corresponding to the NTR region of the gene. The interaction models of these two proteins of different sizes with the viral PA protein were predicted by in silico tools. For these proteins, 3D models were generated with I-TASSER, prominent 3D models were matched with the reference PA model using ClusPro, and interaction models of AD-C209:PA and AD-C244+4:PA are given in Figure 7.

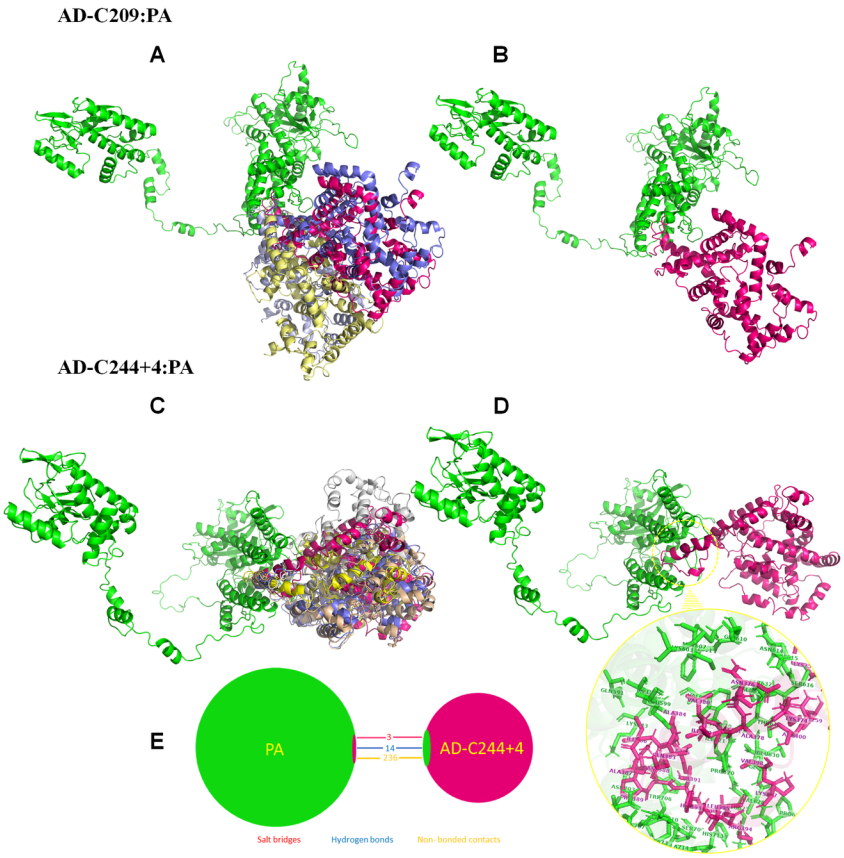


FIGURE 7. 3D models of influenza virus PA, AD-C209 and AD-C244+4 proteins, and selected docking models. **A/C** : several AD-C209 and AD-C244+4 proteins clustered on the PA model, **B/D** : single - AD-C209:PA and AD-C244+4:PA models, **E:** chemical bounds between viral PA and C244+4 proteins defined with PDBsum algorithm.

Among the docking patterns generated by using both the AD-C209 and the AD-C244+4 protein models, there are lots of matches showing an interaction between the carboxy-terminal end of the C14orf166 and PA, which is similar to the models shown in Figure 6. This result increased the interaction possibility of human C14orf166 and influenza A virus PA proteins with these regions in cells. Among these docking models, one of the PA:C244+4 interaction models, which is prominent in terms of binding properties, and the chemical bonds of interaction defined by the PDBsum algorithm are given in Figure 7C. In the evaluation made with the algorithm, fourteen hydrogen bonds and three salt bridges were detected in the interaction area between these two proteins.

Effect of C14orf166 over-expression on a viral RdRP enzyme in mammalian cells

It is known that the C14orf166 protein has a positive regulatory effect on influenza A virus replication. In this study, since it is associated with viral RdRP PA subunit, the effects of the over-expression of the C14orf166 protein on human (WSN) and avian (DkPen) type virus RdRP enzymes were investigated in HEK293 cells using minireplicon models. The results showed an average increase of 70% in the WSN-type virus polymerase enzyme (Figure 8). On the other hand, the C14orf166 over-expression stimulated the DkPen RdRP activity at a slightly lower than that of the WSN type. This result suggested that the C14orf166 protein may affect the viral RdRP activity at differ levels depending on the virus type.

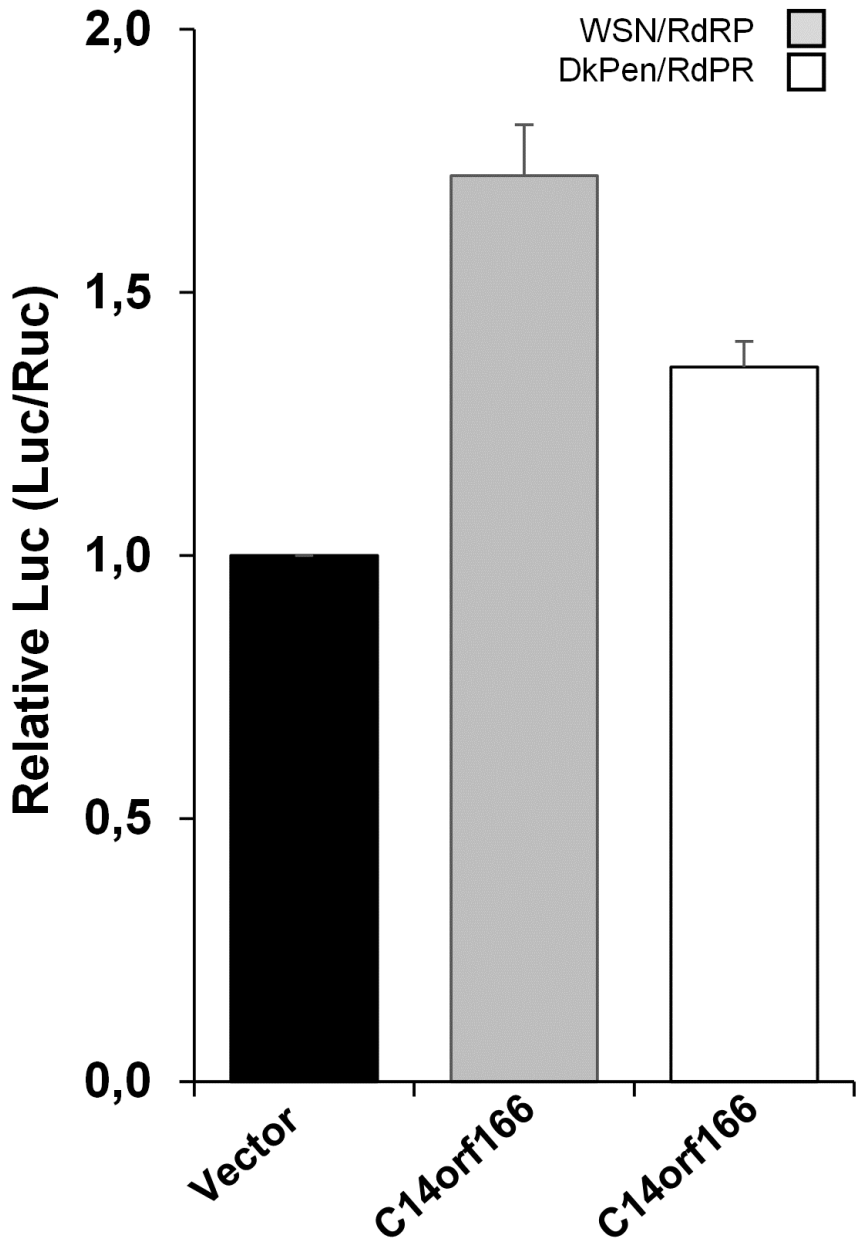


FIGURE 8. The effects of overexpression of C14orf166 proteins on influenza A WSN and DkPen virus RdRp activity. The HEK293 cells were grown in a 24-well plate and co-transfected with a 200 ng of pCHA-C14orf166 expression plasmid and influenza A virus mini-replicon plasmids, and the reporter enzyme activities were defined as previously described[23].

DISCUSSION

The C14orf166 protein is referred to as *RNA Transcription,Translation and Transport Factor* (RTRAF) according to recent data. This protein is a nucleo-cytoplasmic protein having a binding activity to RNA molecules and the RNA polymerase II complex[31]. As a result of these activities, it has a positive regulatory

effect on RNA Pol II transcription and is involved in processes related to RNA metabolism such as tRNA splicing as part of the tRNA splicing ligase complex. In addition, it is thought that it also functions in RNA transport. The C14orf166, a multifunctional protein with these properties, is associated with some cancer types. It has been suggested that the overexpression of C14orf166 has a role in the development of non-small-cell lung cancer (NSCLC), and it would be beneficial to consider it in the treatment of high-risk patients[32]. It has been reported that this protein is overexpressed in breast cancer[33]. The expression of both the C14orf166 mRNA and protein was also found to be upregulated in cervical cancer cell lines and tissues[18]. Therefore, it has been suggested that it can be used as a biomarker for the diagnosis and prognosis of some cancer types[34].

It has been determined that the C14orf166 protein, which has important functions in cellular RNA metabolism, also participates in the replication/transcription processes of some viruses. The C14orf166 protein interacts with the hepatitis C virus core protein and upregulates the viral infection[22]. Another important virus type with which the C14orf166 protein is associated is the influenza A virus. It has been stated that this protein binds to the influenza A virus RbRP enzyme PA subunit and stimulates the virus replication[21]. Our results with the minireplicon models also showed that C14orf166 stimulates the influenza viral polymerases (Figure 8).

In this study, we demonstrated that the full-size and truncated C14orf166 proteins in different sizes interacted with the influenza virus PA protein in the Y2H assay using the PA protein as a bait and the HEK293 Matchmaker cDNA (Figures 1 and 2). Among these proteins, the smallest fragment that interacts with PA protein consists of 69 amino acid residues that form the carboxy-terminal end of the C14odf166 (AD-C69). Therefore, it was concluded that the carboxy-terminal end of C14orf166 is important and sufficient for interaction with the viral PA protein. In a previous report, it was shown that the C14orf166 protein interacts not only with PA alone but also with the influenza virus RdRP complex and co-localizes with viral ribonucleoproteins[12]. Thus, the C14orf166 protein is not thought to bind to the regions of PA that interact with PB2 and PB1 subunits. In the Y2H assay where we used the amino-terminal half of the viral PA protein as a bait, no C14orf166 protein could be detected among the host interactor proteins, which led us to conclude that the amino-terminal region of PA is most likely not a target for this protein. Therefore, we focused on the carboxy-terminal moiety of PA for binding of the C14orf166 protein in docking analyses. Although the most probable 3D structures were selected in terms of free energies among the AD-C14orf166 protein models interacting with PA, these models may not reflect the real structures of the proteins. Here, we focused on the C14orf166:PA interaction model rather than the 3D structures. The docking results of the PA with the predicted 3D structures of the C69 peptide (non-fused with GAL4-AD) and AD-C69 (fused with GAL4-AD) expressed in the yeast cells indicated that this peptide/protein binds to the ligand (peptide) recognition site[6] located in the carboxy-terminal region of the PA. In docking models, it was observed that C69 and AD-C69 clustered in this region, mostly covering the 610th to 630th amino acid residues on the PA (Figures 5A, 6A, and 6C). The docking results of AD-C209:PA and AD-C244+4:PA proteins synthesized in yeast cells showed that the carboxy-terminal ends of these fusion proteins have a high affinity for the peptide binding site of the PA protein such as C69 and AD-C69 (Figure 7). One the docking models of AD-C244+4, which involves full-length C14orf166 and the PA protein, the prominent amino acids in the interaction areas of these two were evaluated with PyMol software and PDBsum algorithm. In this model, there is the possibility of forming three salt bridges and fourteen hydrogen bonds between AD-C244+4 and the PA protein (Figure 7 E). Although the amino acid residues in the interaction area of the PA protein show a distribution within the 450th to 700th amino acids, the intense interaction site is seen in the 610th to 630th amino acid positions. Huarte et al., as a result of their mapping study, suggested that amino acid residues in positions 493 to 512 and 557 to 574 on the PA protein are important for binding to C14orf166[21]. Although this region of PA agrees with our results, docking analysis data point to amino acids within the 610th to 630th position as the most intense interaction site in our models.

In conclusion, possible interaction models between influenza A virus PA and human C14orf166 protein were revealed by reconciling the experimental results of the Y2H assay carried out with the influenza A virus PA as a bait, existing literature information, and the results of in silico 3D prediction/protein-protein docking

analysis. It is very difficult to accurately predict protein-protein interaction patterns similar to those in the natural environment with in silico tools ignoring the factors in the cellular environment. It was concluded that the data of the Y2H assay will make an important contribution to predicting the 3D structures and possible interaction models in a form close to its native structure. The interaction models presented here for human C14orf166 and viral PA protein will allow us to better understand the influenza A virus replication and virus-host relationship.

Acknowledgments

This work was supported by a grant from the Scientific and Technological Research Council of Turkey (TÜBİTAK; grant no: *SBAG-112S518*).

REFERENCES

1. Peteranderl C, Herold S, Schmoldt C. Human Influenza Virus Infections. Semin Respir Crit Care Med 2016;37(4):487-500.

2. Apostolov K, Flewett TH. Further observations on the structure of influenza viruses A and C. J Gen Virol 1969;4(3):365-370.

3. Bouvier NM, Palese P. The biology of influenza viruses. Vaccine 2008;26 Suppl 4:D49-53.

4. Boivin S, Cusack S, Ruigrok RW, Hart DJ. Influenza A virus polymerase: structural insights into replication and host adaptation mechanisms. J Biol Chem 2010;285(37):28411-28417.

5. Noda T, Sagara H, Yen A, et al. Architecture of ribonucleoprotein complexes in influenza A virus particles. Nature 2006;439(7075):490-492.

6. Obayashi E, Yoshida H, Kawai F, et al. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. Nature 2008;454(7208):1127-1131.

7. Dias A, Bouvier D, Crepin T, et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature 2009;458(7240):914-918.

8. Hsu WB, Shih JL, Shih JR, et al. Cellular protein HAX1 interacts with the influenza A virus PA polymerase subunit and impedes its nuclear translocation. J Virol 2013;87(1):110-123.

9. Kocmar T, Caglayan E, Rayaman E, Nagata K, Turan K. Human sorting nexin 2 protein interacts with Influenza A virus PA protein and has a negative regulatory effect on the virus replication. Mol Biol Rep 2022;49(1):497-510.

10. Miyake Y, Ishii K, Honda A. Influenza Virus Infection Induces Host Pyruvate Kinase M Which Interacts with Viral RNA-Dependent RNA Polymerase. Front Microbiol 2017;8:162.

11. Kawaguchi A, Nagata K. De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. EMBO J 2007;26(21):4566-4575.

12. Rodriguez A, Perez-Gonzalez A, Nieto A. Cellular human CLE/C14orf166 protein interacts with influenza virus polymerase and is required for viral replication. J Virol 2011;85(22):12062-12066.

13. Perez-Gonzalez A, Pazo A, Navajas R, Ciordia S, Rodriguez-Frandsen A, Nieto A. hCLE/C14orf166 associates with DDX1-HSPC117-FAM98B in a novel transcription-dependent shuttling RNA-transporting complex. PLoS One 2014;9(3):e90957.

14. Amiel J, Rio M, de Pontual L, et al. Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. Am J Hum Genet 2007;80(5):988-993.

15. Fukuda A, Shimada M, Nakadai T, Nishimura K, Hisatake K. Heterogeneous nuclear ribonucleoprotein R cooperates with mediator to facilitate transcription reinitiation on the c-Fos gene. PLoS One 2013;8(8):e72496.

16. Makise M, Mackay DR, Elgort S, Shankaran SS, Adam SA, Ullman KS. The Nup153-Nup50 protein interface and its role in nuclear import. J Biol Chem 2012;287(46):38515-38522.

17. Chen X, Ying Z, Lin X, et al. Acylglycerol kinase augments JAK2/STAT3 signaling in esophageal squamous cells. J Clin Invest 2013;123(6):2576-2589.

18. Zhang W, Ou J, Lei F, et al. C14ORF166 overexpression is associated with pelvic lymph node metastasis and poor prognosis in uterine cervical cancer. Tumour Biol 2016;37(1):369-379.

19. Cheang TY, Zhou HY, Chen W, et al. C14orf166 overexpression correlates with tumor progression and poor prognosis of breast cancer. J Transl Med 2016;14:54.

20. Elvira G, Wasiak S, Blandford V, et al. Characterization of an RNA granule from developing brain. Mol Cell Proteomics 2006;5(4):635-651.

21. Huarte M, Sanz-Ezquerro JJ, Roncal F, Ortin J, Nieto A. PA subunit from influenza virus polymerase complex interacts with a cellular protein with homology to a family of transcriptional activators. J Virol 2001;75(18):8597-8604.

22. Lee JW, Liao PC, Young KC, et al. Identification of hnRNPH1, NF45, and C14orf166 as novel host interacting partners of the mature hepatitis C virus core protein. J Proteome Res 2011;10(10):4522-4534.

23. Pham PTV, Turan K, Nagata K, Kawaguchi A. Biochemical characterization of avian influenza viral polymerase containing PA or PB2 subunit from human influenza A virus. Microbes Infect 2018;20(6):353-359.

24. Nagata K, Saito S, Okuwaki M, et al. Cellular localization and expression of template-activating factor I in different cell types. Exp Cell Res 1998;240(2):274-281.

25. Turan K, Mibayashi M, Sugiyama K, Saito S, Numajiri A, Nagata K. Nuclear MxA proteins form a complex with influenza virus NP and inhibit the transcription of the engineered influenza virus genome. Nucleic Acids Res 2004;32(2):643-652.

26. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008;9:40.

27. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. Nat Protoc 2017;12(2):255-278.

28. Fan H, Walker AP, Carrique L, et al. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature 2019;573(7773):287-290.

29. Laskowski RA. Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. Bioinformatics 2007;23(14):1824-1827.

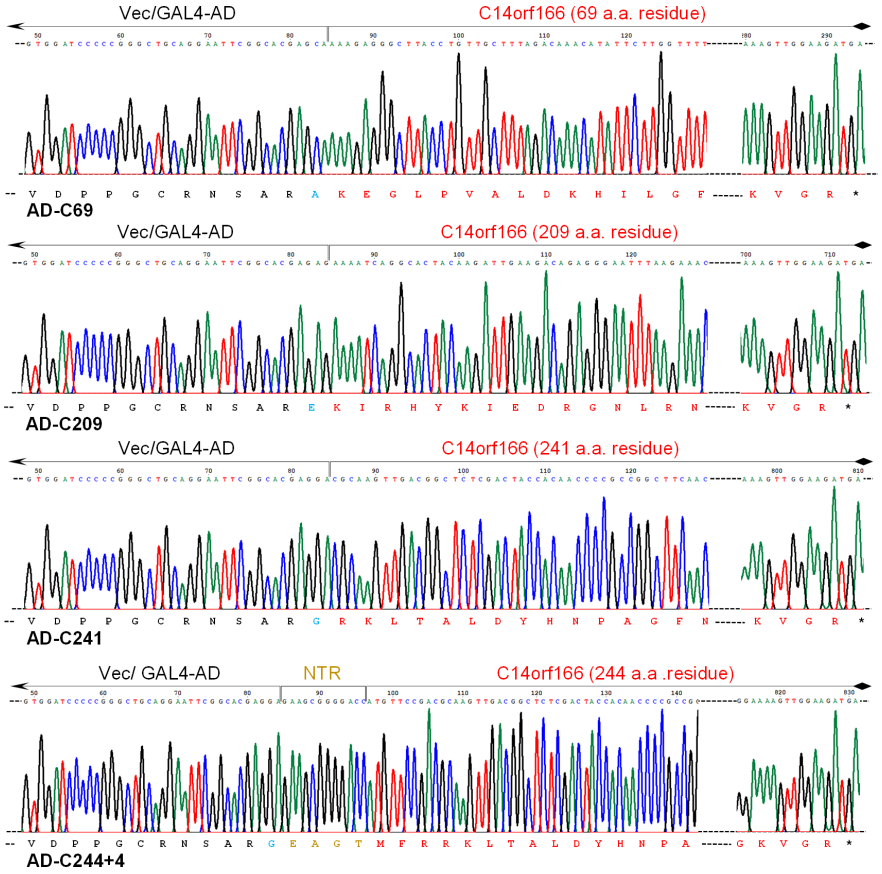
30. Berg JM, Tymoczko JL, Stryer L 2001 Biochemistry, 5th ed. edn, W. H. Freeman and CO, New York.

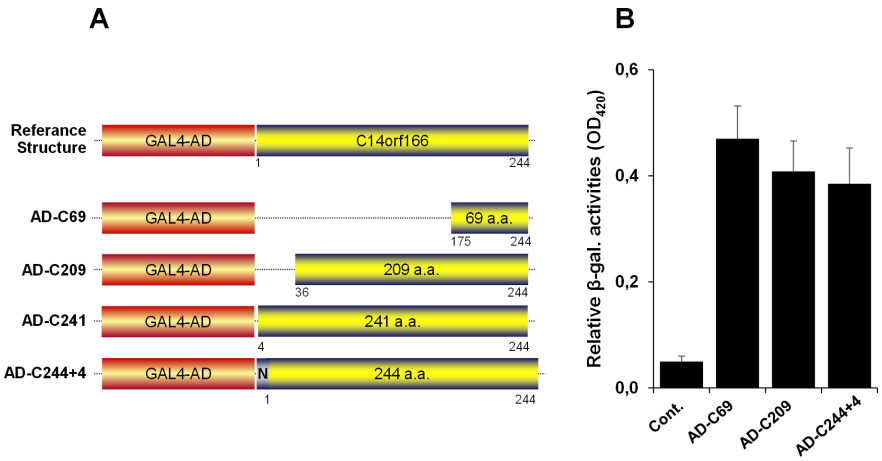
31. Perez-Gonzalez A, Rodriguez A, Huarte M, Salanueva IJ, Nieto A. hCLE/CGI-99, a human protein that interacts with the influenza virus polymerase, is a mRNA transcription modulator. J Mol Biol 2006;362(5):887-900.

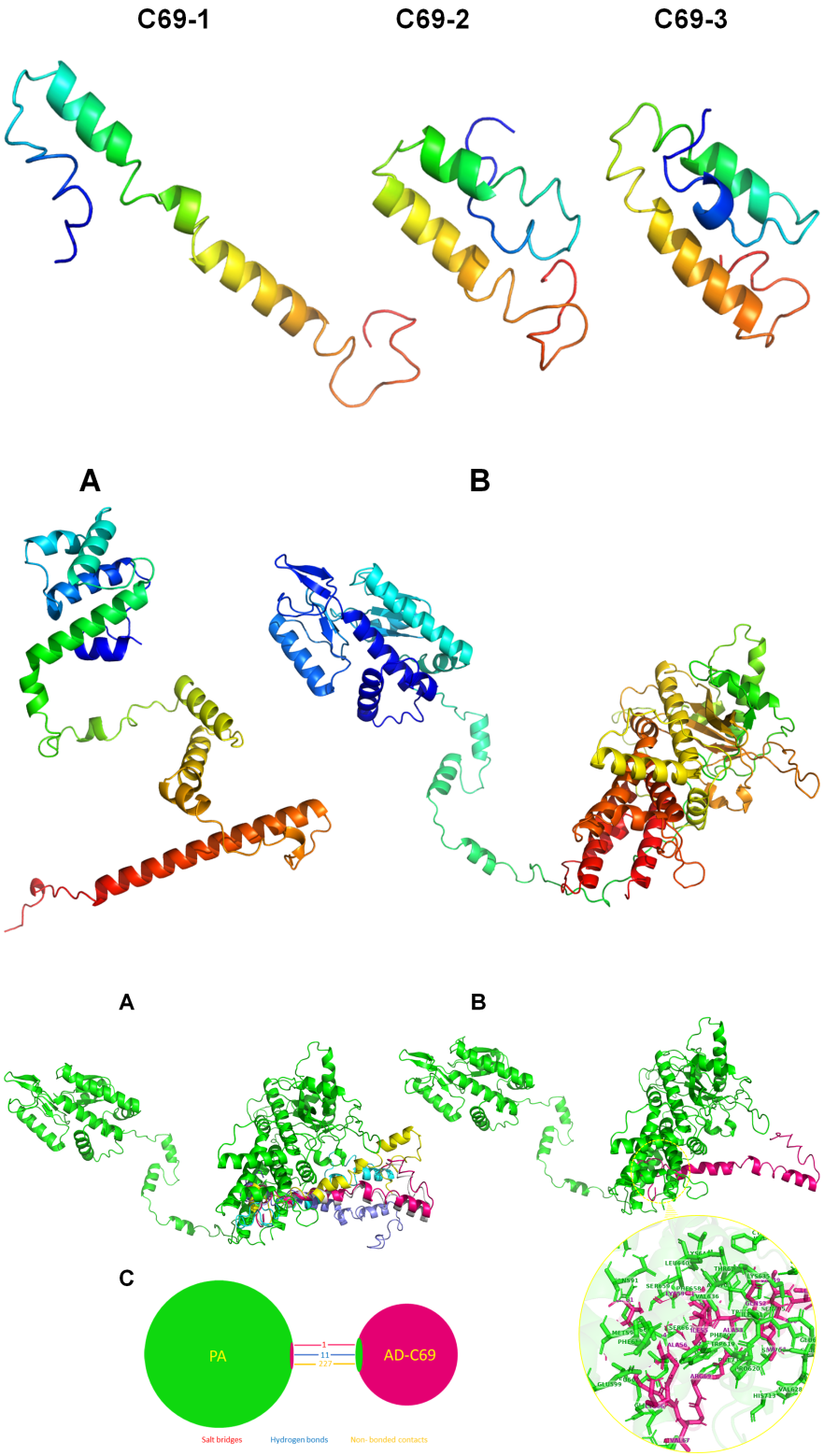
32. Zhou YW, Li R, Duan CJ, Gao Y, Cheng YD, Zhang CF. Overexpressed C14orf166 associates with disease progression and poor prognosis in non-small-cell lung cancer. Biosci Rep 2018;38(5).

33. Lin C, Liao W, Jian Y, et al. CGI-99 promotes breast cancer metastasis via autocrine interleukin-6 signaling. Oncogene 2017;36(26):3695-3705.

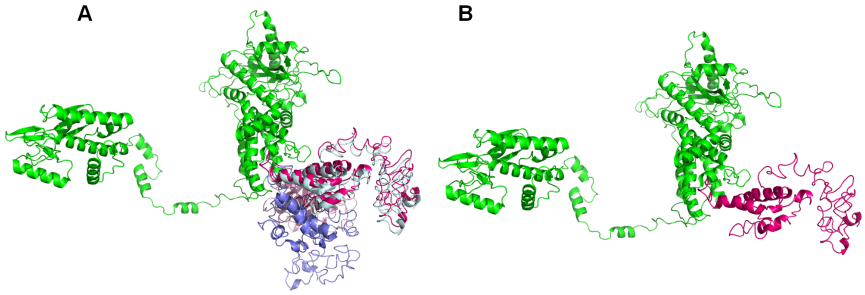
34. Chen Q, Li R. Role of C14orf166 in viral infection and RNA metabolism and its relationship with cancer (Review). Mol Med Rep 2021;23(6).



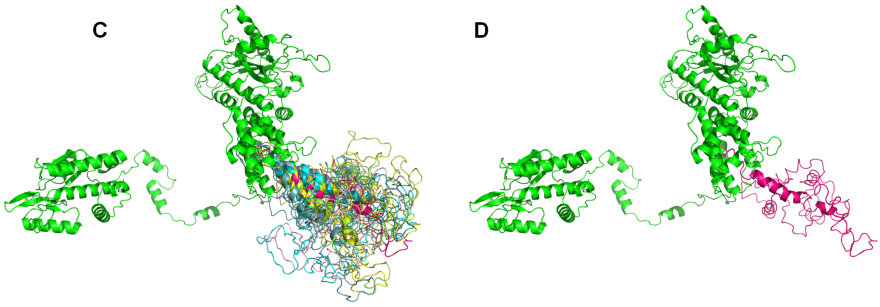




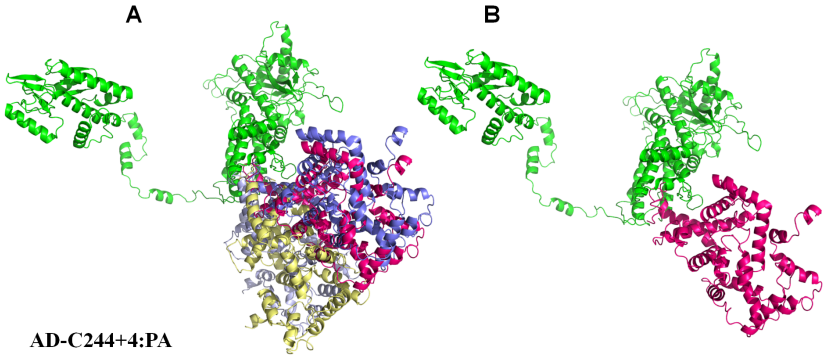
AD-C69:PA pattern 1



AD-C69:PA pattern 2



AD-C209:PA



AD-C244+4:PA