Cryo-EM structure of the CBC-ALYREF complex

Bradley P. Clarke[1], Alexia E. Angelos[1], Menghan Mei[1], Pate S. Hill[1], Yihu Xie[1,*], and Yi Ren[1,2,*]

[1]Department of Biochemistry, Vanderbilt University School of Medicine Basic Sciences, Nashville, TN, USA
[2]Center for Structural Biology, Vanderbilt University School of Medicine Basic Sciences, Nashville, TN, USA

*Corresponding authors

Yi Ren: yi.ren@vanderbilt.edu
Yihu Xie: yihu.xie@vanderbilt.edu

ABSTRACT

In eukaryotes, RNAs transcribed by RNA Pol II are modified at the 5' end with a 7-methylguanosine (m^7G) cap, which is recognized by the nuclear cap binding complex (CBC). The CBC plays multiple important roles in mRNA metabolism including transcription, splicing, polyadenylation and export. It promotes mRNA export through direct interaction with ALYREF, which in turn links the TRanscription and EXport (TREX) complex to the 5' end of mRNA. However, the molecular mechanism for CBC mediated recruitment of the mRNA export machinery is not well understood. Here, we present the first structure of the CBC in complex with a mRNA export factor, ALYREF. The cryo-EM structure of CBC-ALYREF reveals that the RRM domain of ALYREF makes direct contacts with both the NCBP1 and NCBP2 subunits of the CBC. Comparison of CBC-ALYREF to other CBC and ALYREF containing cellular complexes provides insights into the coordinated events during mRNA transcription, splicing, and export.

INTRODUCTION

The nuclear cap binding complex (CBC) binds to the m^7G cap of RNAs transcribed by RNA pol II. It is comprised of NCBP1 (also known as CBP80) and NCBP2 (also known as CBP20). This heterodimeric CBC can form a variety of interactions with different proteins to promote mRNA processing and influence the fate of the transcript. As such, the CBC regulates gene expression at multiple levels, ranging from transcription and splicing to nuclear export and translation[1,2].

The CBC is one of a myriad of protein factors that associate with newly synthesized transcripts. These factors package mRNAs into compacted ribonucleoprotein particles (mRNPs). While the overall structural arrangement of mRNPs is not known, factors such as serine/arginine-rich (SR) proteins are suggested to contribute to mRNP compaction. Prior to export, mRNPs acquire the export receptor NXF1-NXT1 to gain access to the nuclear pore complex (NPC)[3]. NXF1-NXT1 can interact with the FG repeats of the nucleoporin proteins in the NPC to mediate mRNA export[4,5]. Intriguingly, early electron microscopic studies on the Balbiani ring mRNPs in *C. tentans* showed that mRNPs translocate through the NPC in a 5' to 3' direction[6]. Recent single molecule work on human mRNPs also suggests mRNA is exported 5' end first[7].

The CBC promotes nuclear mRNA export through its interaction with a key mRNA export factor ALYREF. ALYREF in turn binds to DDX39B (also known as UAP56) and by extension, the entire TRanscription-EXport (TREX) complex at the 5' end of the RNA[8]. The TREX complex is conserved from yeast to humans[9-13]. The human TREX complex is composed of THOC1, 2, 3, 5, 6, 7, and the DEAD-box helicase DDX39B[14]. Interestingly, ALYREF was also identified as the THOC4 subunit of the TREX complex, indicating that the function of ALYREF is tightly integrated with the TREX complex. ALYREF and its yeast ortholog Yra1 contain UBMs (UAP56-binding motifs) that mediate the interaction with DDX39B and Sub2, respectively[10,15-17]. The yeast CBC was also shown to facilitate the recruitment of Yra1 onto nascent RNA[18]. ALYREF and TREX play central roles in mRNA export through direct interactions with various factors including the export receptor NXF1-NXT1[19-21]. Their association with the 5' cap of mRNAs in particular is a key step in

the export process. However, thus far the molecular mechanism underlying how ALYREF bridges the CBC and TREX remains unclear.

ALYREF and the associated TREX complex are not only required for cellular mRNA export but also can be hijacked for nuclear export of some viral mRNAs, such as those of herpes viruses. Two well studied examples are Herpes simplex virus (HSV-1) ICP27 and Herpesvirus saimiri (HVS) ORF57[22,23]. Both HSV-1 ICP27 and HVS ORF57 directly target the host ALYREF protein[24,25]. Structural studies show that they recognize overlapping surfaces on the RRM domain of ALYREF. How these viral factors affect host CBC-ALYREF interaction and function is not known.

The highly integrated nuclear mRNA processing and mRNP packaging also require the actions of multi-functional splicing factors, such as the SR protein SRSF1 and the exon junction complex (EJC). SRSF1 couples transcription, splicing, and export through direct interactions with the CBC, spliceosome, NXF1-NXT1, and RNA[26-28]. Of note, ALYREF also has binding activities to these factors. Therefore, the functions of ALYREF and SRSF1 in mRNA processing and export are likely interconnected. In higher eukaryotes, the EJC is deposited 20–24 nucleotides upstream of the exon junctions during splicing[29-31]. The EJC was initially shown to associate with ALYREF through a WxHD motif in the N-terminal unstructured region of ALYREF[32]. More recently, multiple binding interfaces were shown between ALYREF and the EJC[12]. How the CBC-ALYREF connection affects the function of EJC-ALYREF remains elusive.

To better understand the multiple functions of ALYREF and the CBC in RNA metabolism, we carried out structural and biochemical studies. We present cryo-EM structures of the CBC and CBC-ALYREF complexes at 3.51 Å and 3.16 Å resolution, respectively. The CBC-ALYREF structure reveals that both the NCBP1 and NCBP2 subunits of the CBC interact with ALYREF. Conformational changes in the CBC are observed that accommodate the interaction. Both HSV-1 ICP27 and HVS ORF57 target the binding interface between the RRM domain of ALYREF and the CBC. We suggest that these viruses not only hijack host pathways to export their own RNA but could also inhibit host

RNA metabolism through their interactions with ALYREF. Structural overlay of CBC-ALYREF and EJC-ALYREF reveals that both the CBC and the EJC bind to the RRM domain of ALYREF in a mutually exclusive manner. This suggests that ALYREF's interaction with the EJC is favored after ALYREF dissociates from the CBC, or as an independent event.

RESULTS AND DISCUSSION

ALYREF directly interacts with CBC

Recombinant human ALYREF protein was shown to interact with the CBC in RNase-treated nuclear extracts[8]. Here we used purified recombinant proteins to further investigate the molecular interactions between ALYREF and the CBC (Figure 1). It is well known that RS domain containing proteins, including ALYREF, exhibit low solubility and are prone to aggregation. The addition of glutamic acid and arginine to the buffer can increase protein solubility and stability[33]. Indeed, with a 1:1 mixture of glutamic acid and arginine, we were able to purify a GST-tagged ALYREF (residues 1-183) protein containing the N-terminal region and the RRM domain of ALYREF (Figure 1A). This construct includes the WxHD motif (residues 87-90), whose mutation was shown to affect interaction with the CBC in an immunoprecipitation study[32]. GST pull down assays showed that ALYREF directly interacts with the CBC (Figure 1B).

To aid in structural studies, we tested solubility of several ALYREF orthologs and found that mouse ALYREF2 (mALYREF2, residues 1-155) exhibited better solubility than the ALYREF construct utilized above. The conserved UBM motifs, the WxHD motif, and the RRM domain are nearly identical between ALYREF and mALYREF2 (Figure 1–figure supplement 1). Indeed, like ALYREF, mALYREF2 directly interacts with the CBC (Figure 1B). Since the ALYREF/mALYREF2 interaction with the CBC is conserved and mALYREF2 exhibits better solubility, we focused on mALYREF2 in the cryo-EM investigations.

Cryo-EM structures of CBC and CBC-ALYREF complexes.

We first determined the cryo-EM structure of the human CBC complex at 3.51 Å (Figure 2A, Figure 2–figure supplement 1 and 2, Table 1). The cap analog m^7GpppG was added in our cryo-EM study and the electron density of the cap moiety is clearly visible bound to NCBP2. Compared to unliganded CBC, the cap analog induces significant rearrangements in both the N-terminal extension and the C-terminal tail of NCBP2 (Figure 2B) to form critical interactions. For example, the N-terminal extension (residues 16-28) swings toward the central globular domain of NCBP2 and positions the Y20 residue to sandwich the cap analog with Y43 of NCBP2. These conformational changes were also observed in the crystal structures of ligand bound CBC[34,35]. Overall, the cryo-EM structure of the CBC determined here resembles the previously reported crystal structures of the liganded CBC, with root mean squared deviation (RMSD) of 0.89 Å and 0.87 Å for NCBP1 and NCBP2, respectively[34].

We next determined the cryo-EM structure of the CBC-mALYREF2 complex in the presence of the cap analog m^7GpppG at 3.16 Å resolution. The structure shows that the RRM domain of mALYREF2 binds to both NCBP1 and NCBP2 subunits (Figure 2C and 2D). The N terminal region of mALYREF2 (residues 1 to 73) does not show traceable density and is possibly disordered. Compared to the CBC alone cryo-EM structure, conformational changes are observed at the mALYREF2-CBC interface. Notably, a loop formed by residues 38-45 of NCBP1 becomes ordered and moves toward mALYREF2 (Figure 2D).

CBC-ALYREF interfaces

The interfaces between mALYREF2 and the CBC involve 19, 14, and 5 residues of mALYREF2, NCBP1, and NCBP2, respectively (Figure 3). The RRM domain of mALYREF2 assumes a canonical β1α1β2β3α2β4 topology (Figure 3A), forming an α-helical surface and a β-sheet surface. The α-helical surface recognizes the CBC through extensive hydrophilic and hydrophobic interactions (Figure 3B). The α1 helix of mALYREF2 is enriched with acidic residues and makes key hydrophilic interactions with

NCBP1. For example, E97 forms salt bridges with K330 and K381 of NCBP1. Y135 on the α2 helix of mALYREF2 makes a hydrogen bond with K330 of NCBP1. The importance of this interface between ALYREF and NCBP1 is highlighted by a K330N mutation found in human uterine corpus endometrial carcinoma[36-38]. In addition, the loop between α2 and β4 of mALYREF2 forms hydrophobic interactions with NCBP1. V138, P139, and L140 of mALYREF2 bind to a hydrophobic pocket on NCBP1 formed by A334, V337, and L382. Furthermore, at the NCBP1-mALYREF2 interface, the conformational change of the NCBP1 loop (residues 38-45), observed when compared to the CBC alone structure, enables K41 of NCBP1 to interact with mALYREF2 (Figure 2D and 3B).

The interface between mALYREF2 and NCBP2 is near the m^7G binding pocket (Figure 3C). The α2 helix of mALYREF2 contacts S13 and Y14 in the N-terminal extension of NCBP2. S13 and Y14 also directly interact with NCBP1 and are thought to enable the hinged motion of the N-terminal extension (residues 16-28) upon binding to the cap[35]. In addition, the α1 helix of mALYREF2 is in proximity to the R105 and I110 residues of NCBP2. Mutations of these interacting residues are associated with human cancers. For example, the NCBP2 R105C mutation has been reported in colorectal cancer and the NCBP2 I110M mutation has been found in head and neck cancer[36-38]. NCBP2 exhibits a positively charged groove extending from the cap binding site which is suggested to be an RNA-binding site[34]. Upon mALYREF2 binding, this groove is buried. Interestingly, mALYREF2 features a positively charged surface near the m^7G site (Figure 3D). Conceivably, this positively charged surface on mALYREF2 could serve as an RNA binding site for the nucleotides following the cap.

Based on the CBC-mALYREF2 structure, we generated mutations (mut-1 and mut-2) on the RRM domain of human ALYREF (ALYREF-RRM, residues 103-183) to validate its interaction with the CBC. For ALYREF-RRM-mut-1 (Y166R/V169R/P170R), mutated residues are localized on the α2-β4 loop and correspond to residues Y135/V138/P139 in mALYREF2 (Figure 3A, B). For ALYREF-RRM-mut-2 (E124R/E128R), mutated residues are localized on the α1 helix and correspond to residues E93/E97 in mALYREF2 (Figure 3A, B). In agreement with the CBC-ALYREF structure, we found that the RRM domain of

ALYREF directly interacts with the CBC, albeit with weaker interaction compared to ALYREF (residues 1-183) (Figure 1B and 4A). The difference likely results from the WxHD motif (residues 87-90) localized in the N-terminal region of ALYREF. Evidence suggests that mutation of the WxHD motif reduces ALYREF's interaction with the CBC[32]. The WxHD motif may represent a second binding site for the CBC that remains to be characterized. Importantly, compared to the wild type protein, both ALYREF-mut-1 and mut-2 show reduced binding to the CBC (Figure 4A). Together, the mutagenesis studies validate the CBC-ALYREF interfaces observed in the structure.

The CBC-mALYREF2 structure reveals that the interaction between ALYREF and the CBC mainly involves the NCBP1 subunit (Figure 3B and 3C). We further dissected the interaction between ALYREF and individual NCBP1 and NCBP2 subunits using GST pull down assays. NCBP1 can be efficiently pulled down by GST-ALYREF, whereas NCBP2 did not show detectable interaction (Figure 4B). These results are consistent with the structural observations and indicate that NCBP1 is the major subunit of the CBC to interact with ALYREF.

CBC-ALYREF and 5' cap dependent mRNP export.

ALYREF recruits the mRNP export machinery TREX complex to the 5' end of mRNA through direct interactions with both the CBC and TREX[8,10]. The UBMs of ALYREF directly interact with the DDX39B component of the TREX complex[10]. The N-terminal UBM is included in the ALYREF construct used for our cryo-EM studies but did not show visible electron density. Thus, this UBM is likely exposed and available to interact with DDX39B, which further connects to the entire TREX complex (Figure 5A). Consistently, ALYREF, DDX39B, THOC1 and THOC2 are present in NCBP1 immunoprecipitations from RNase-treated HeLa cell nuclear extracts[8]. In yeast, mutually exclusive interactions were shown between Yra1 with Sub2 and the NXF1-NXT1 ortholog Mex67-Mtr2[17]. So, the ALYREF dependent NXF1-NXT1 loading on mRNA likely occurs after DDX39B dissociates from ALYREF. The CBC could also function as a landing pad for ALYREF as previously proposed[39]. After recruitment to the 5' end of mRNA by the CBC, ALYREF

could then transfer away from the 5' end, to other sites enriched with export factors and participate in different complexes located along the mRNA. In addition to the ALYREF-NXF1-NXT1 complex, some other ALYREF containing complexes could exist on the same mRNP, such as the complex of ALYREF/DDX39B/SARNP, which facilitates high order mRNP assembly[40,41].

The process of mRNP export, with the 5' end exiting first from the NPC, has been shown in both insect and human systems using electron microscopy and single molecule imaging techniques[6,7]. Interestingly, in the latter study, several adjacent NPCs were found to engage in the export of the same mRNA[42]. This observation is reminiscent of the gene gating hypothesis which suggested that transcriptionally active genes are physically tethered to the site of mRNA export at the NPC[42]. Gated genes have been shown in yeast, worms, flies and humans[43,44]. For these gated genes, the 5' directionality of mRNA export could be primarily driven by the key placement of crucial RNA export factors at the 5' end of the gene as illustrated here (Figure 5A), and this localization of export factors could greatly increase the efficiency of co-transcriptional processing and export.

CBC-ALYREF and viral hijacking of host mRNA export pathway.

HSV-1 ICP27 and HVS ORF57 hijack the host mRNA export pathway through interactions with ALYREF. The RRM domain of ALYREF is targeted by both HSV-1 ICP27 and HVS ORF57 with overlapping interfaces (Figure 5B, Figure 5–figure supplement 1). Structural comparison between CBC-ALYREF, ALYREF-ICP27, and ALYREF-ORF57 reveals that the interface between ALYREF's RRM domain with the CBC is not compatible with the ICP27/ORF57-ALYREF interactions (Figure 5B, Figure 5–figure supplement 1). However, in vivo data shows that the ORF57 orthologue from Kaposi's Sarcoma–Associated Herpesvirus (KSHV) can still form a complex with ALYREF and the CBC[22]. So, it is likely that although ALYREF's RRM domain interface with the CBC could be disrupted by ORF57, ALYREF can still use the WxHD motif to interact with the CBC (Figure 5C). Using this strategy, a virus can hijack the host pathway while at the same time also disrupt host interactions and processes. It should also be noted that the CBC, ALYREF, ICP27, and

ORF57 are all RNA binding proteins. In addition to the protein mediated interactions discussed above, RNA interactions should be considered, especially under the in vivo setting (Figure 5C). NXF1-NXT1 and DDX39B, the cellular ALYREF interacting proteins, are also hijacked by other factors from viruses. NXF1-NXT1 is targeted by influenza A virus NS1 protein[45] and SARS-CoV-2 Nsp1 protein[46]. DDX39B is targeted by influenza A virus NP protein[47,48]. The molecular mechanisms revealed here and from previous studies pave the way for new useful targets in anti-viral therapeutics.

Functional interplay of CBC-ALYREF and mRNP export factors

Transcription, splicing, and export are all tightly linked processes. The CBC promotes splicing through its binding partners, such as SRSF1[49,50]. The detailed molecular interaction between the CBC and SRSF1 is revealed in a human pre-B^{act-1} spliceosome structure[28]. The NCBP2 subunit of the CBC is the major binding site for SRSF1. Structural overlay of this structure with the CBC-ALYREF structure shows no significant steric hindrance between SRSF1 and ALYREF (Figure 6A). However, whether the CBC-ALYREF-SRSF1 complex exits in vivo and how their functions might be coordinated require further studies. After splicing, SRSF1 is mainly deposited on exons[51,52]. SRSF1 functions in mRNA export through interaction with the export receptor NXF1-NXT1[26,53,54]. Of note, the interaction between both ALYREF and SR proteins with NXF1-NXT1 is regulated by phosphorylation. Only hypo-phosphorylated SR proteins can bind to NXF1-NXT1 efficiently[26,55,56]. Interestingly, SR proteins, such as Gbp2 and Hrb1 in *S. cerevisiae*, also interact directly with the TREX complex[11,57], suggesting coordinated actions of TREX and SR proteins in mRNA export.

Unlike the CBC and SRSF1 interfaces discussed above, the interaction between the ALYREF RRM domain and the CBC is not compatible with the interface between the ALYREF RRM domain and the EJC subunit MAGOH[12] (Figure 6B). Mutation of the ALYREF WxHD motif affects its interaction with both the CBC and the EJC subunit eIF4A3[32]. Since both the WxHD motif and the RRM domain of ALYREF are mutually exclusive binding sites for the CBC and the EJC, the formation of the EJC-ALYREF

complex likely happens after ALYREF dissociates from the CBC. It is also possible that the EJC-ALYREF interaction is independent of the CBC (Figure 6C). This possibility is supported by the report that ALYREF can be recruited to RNA by both CBC dependent and independent mechanisms[58]. The resulting mRNP with multiple copies of ALYREF and NXF1-NXT1, each recruited through different mechanisms and at different sites on the mRNP (Figure 6C), could exhibit increased export efficiency.

MATERIALS AND METHODS

Plasmids

Human NCBP1 (UniProt Q09161) was cloned into the pFastBac HTc vector with an N-terminal TEV cleavable His tag. Human NCBP2 (UniProt P52298) was cloned into a modified pFastBac1 vector containing an N-terminal TEV cleavable GST tag. Human ALYREF (UniProt Q86V81) and mouse ALYREF2 (UniProt Q4KL64) constructs were cloned into a modified pGST-4T-1 vector containing an N-terminal TEV cleavable GST tag.

Protein expression and purification

The NCBP1-NCBP2 complex was expressed in High-Five insect cells (Invitrogen) by coinfection of recombinant baculoviruses. Individual NCBP1 and NCBP2 subunits were expressed in High-Five insect cells infected with the respective recombinant baculovirus. High-Five cells were harvested 48 hrs after infection and lysed in a buffer containing 50 mM Tris pH 8.0, 300 mM NaCl, 0.2 mM AEBSF, 2 mg/L aprotinin, 1 mg/L pepstatin, 1 mg/L leupeptin, and 0.5 mM TCEP. Proteins were purified using Glutathione Sepharose 4B resin (Cytiva) for NCBP2 and NCBP1-NCBP2, or Ni Sepharose 6FF resin (Cytiva) for NCBP1 alone. The proteins were further purified on a mono Q column (Cytiva). The expression tags were removed by overnight incubation with GST-TEV (for NCBP2 and NCBP1-NCBP2) or His-TEV (for NCBP1) at 4 °C. The samples were passed through Glutathione Sepharose 4B resin (for NCBP2 and NCBP1-NCBP2) or Ni Sepharose 6FF resin (for NCBP1) to remove undigested protein and TEV. The proteins were further

purified on a Superdex 200 column (Cytiva) in 10 mM Tris pH 8.0, 300 mM NaCl, and 0.5 mM TCEP.

GST tagged ALYREF (residues 1-183), ALYREF-RRM (residues 102-183). ALYREF-RRM-mut-1 (Y166R/V169R/P170R), ALYREF-RRM-mut-2 (E124R/E128R), and mALYREF2 (residues 1-155) were expressed in *E. coli* Rosetta cells (Sigma-Aldrich). Protein expression was induced with 0.5 mM IPTG at 16°C overnight. Cells were lysed in a buffer containing 50 mM Tris, pH 8.0, 500 mM NaCl, 50 mM glutamic acid, 50 mM arginine, 0.5 mM TCEP, 0.2 mM AEBSF, and 2 mg/L aprotinin. The GST-tagged proteins were pulled down using Glutathione Sepharose 4B resin. GST-tagged ALYREF-RRM wild type and mutant proteins were purified on a Superdex 200 column in 10 mM Tris pH 8.0, 300 mM NaCl, and 5 mM DTT. GST-tagged ALYREF (residues 1-183) was purified by a cation exchange column (source 15S, Cytiva), followed by a Superdex 200 column equilibrated with 10 mM Tris, pH 8.0, 500 mM NaCl, 50 mM glutamic acid, 50 mM Arginine, and 0.5 mM TCEP. GST-tagged mALYREF2 (residues 1-155) was purified by an anion exchange column (Q Sepharose, Cytiva), followed by a cation exchange column (SP Sepharose, Cytiva). For untagged mALYREF2 used in cryo-EM studies, the GST tag was removed by overnight incubation with GST-TEV at 4 °C. Untagged mALYREF2 was further purified on a cation exchange HiTrap SP column (Cytiva), followed by a Superdex 200 column equilibrated with 10 mM Tris, pH 8.0, 150 mM NaCl, 50 mM glutamic acid, 50 mM Arginine, and 0.5 mM TCEP.

All purified proteins were flash frozen in liquid nitrogen, and stored at -80°C.

Cryo-EM sample preparation and data collection

NCBP1-NCBP2 at 1.2 μM was incubated with mALYREF2 (residues 1-155) at 3.6 μM in the presence of the 5' cap analog m^7GpppG (NEB) at 500 μM at 4 °C for 0.5 hr. The sample was deposited on glow-discharged UltrAuFoil R 1.2/1.3 grids (Quantifoil). Grids were blotted for 6 s with a blotting force of 6 at 4 °C and 100% humidity and plunged into liquid ethane using a FEI Vitrobot Mark IV (Thermo Fisher). The data were collected with a Glacios Cryo-TEM (Thermo Fisher) equipped with a Falcon 4i detector (Thermo Fisher).

Movies were collected with EPU at a magnification of 190,000x, corresponding to a calibrated pixel size of 0.732 Å/pixel. A total of 5858 movies recorded in EER format were collected with a defocus range from 1.0 µm to 2.0 µm. A full description of the cryo-EM data collection parameters can be found in Table 1.

Cryo-EM image processing and Model building

Cryo-EM data were processed with cryoSPARC[59]. Movies in EER format were gain normalized, aligned and dose-weighted using patch motion correction, followed by patch CTF estimation. About 2.5 million particles were initially picked with templates generated from the NCBP1-NCBP2 structure (PDB 1H6K). Multiple rounds of 2D classification and heterogenous refinement were carried out to obtain 745k good particles. Refinement of these particles showed unambiguous density corresponding to mALYREF2. To select mALYREF2 containing particles, heterogenous refinement was performed with two classes corresponding to NCBP1-NCBP2-mALYREF2 and NCBP1-NCBP2 alone. The class corresponding to NCBP1-NCBP2 was subjected to another round of heterogenous refinement to get a cleaner set of particles, followed by homogenous refinement and local refinement. The NCBP1-NCBP2 map was refined with 143k particles and achieved a resolution of 3.51 Å as assessed by an FSC threshold of 0.143. The class corresponding to NCBP1-NCBP2-mALYREF2 was subjected to homogenous refinement and local refinement. The NCBP1-NCBP2-mALYREF2 map was refined with 440k particles and achieved a resolution of 3.16 Å as assessed by an FSC threshold of 0.143.

An initial model of NCBP1-NCBP2 was obtained by docking AlphaFold models of NCBP1 (AF-Q09161-F1) and NCBP2 (AF-P52298-F1) into the cryo-EM density map sharpened using a B-factor of -121 $Å^2$ estimated from the Guinier plot. An initial model of NCBP1-NCBP2-mALYREF2 was obtained by docking AlphaFold models of NCBP1, NCBP2, and mALYREF2 into the cryo-EM density map sharpened using a B-factor of -130 $Å^2$ estimated from the Guinier plot. The models were adjusted in Coot[60], followed by real-space refinement in Phenix[61]. The final NCBP1-NCBP2 model contains NCBP1, NCBP2, and the m^7G moiety of m^7GpppG. The final NCBP1-NCBP2-mALYREF2 model contains NCBP1, NCBP2, the RRM domain of mALYREF2, and the m^7G moiety of m^7GpppG.

Figures were prepared using PyMOL (Molecular Graphics System, Schrodinger, LLC) and Chimera[15].

GST pull down

GST or GST tagged proteins were pre-incubated with GST resin in binding buffer (10 mM Tris pH 8.0, 500 mM NaCl, 50 mM glutamic acid, 50 mM arginine, 0.5 mM TCEP) on ice for 0.5 hour and was mixed with gentle tapping every 3-5 minutes. The beads were then washed three times with 600 µl of buffer containing 10 mM Tris pH 8.0, 100 mM NaCl (Figure 1B and 4B) or 50 mM NaCl (Figure 4A), and 0.5 mM TCEP. NCBP1, NCPB2, or the NCBP1-NCBP2 complex was incubated with the 5' cap analog m^7GpppG at 20 µM, adjusted salt concentration to 100 mM NaCl (Figure 1B and 4B) or 50 mM NaCl (Figure 4A), and then added to beads. The samples were incubated on ice for 0.5 hour and mixed with gentle tapping every 3-5 minutes. Beads were then washed three times with 600 µl of wash buffer (10 mM Tris pH 8.0, 50 mM NaCl, 0.5 mM TCEP) before bound proteins were eluted in 10 mM Tris pH 8.0, 500 mM NaCl (Figure 1B and 4B) or 150 mM NaCl (Figure 4A), 25 mM Glutathione, and 0.5 mM TCEP. 6% (Figure 1B and 4B) or 3% (Figure 4A) of the input and 60% of the eluted proteins were analyzed using Coomassie stained SDS-PAGE gels. The experiments were repeated three times independently.

DATA AVAILABILITY

The cryo-EM models and corresponding maps have been deposited in the RCSB database and in the EMDB.

ACKNOWLEDGEMENTS

We thank Melissa Chambers, Scott Collier, and Mariam Haider at the Center for Structural Biology Cryo-EM Facility at Vanderbilt University for assistance in Cryo-EM data collection. We acknowledge the use of the Glacios cryo-TEM, which was acquired by NIH grant S10 OD030292-01.

FUNDING

This work was supported by NIH R35 GM133743 (Y.R.). B.P.C. was in part supported by NIH/NCI training grant T32CA119925.

CONFLICT OF INTEREST

The authors declare they have no conflict of interest.

REFERENCES

1. Gonatopoulos-Pournatzis, T. & Cowling, V.H. Cap-binding complex (CBC). *Biochem J* **457**, 231-42 (2014).
2. Rambout, X. & Maquat, L.E. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev* **34**, 1113-1127 (2020).
3. Xie, Y. & Ren, Y. Mechanisms of nuclear mRNA export: A structural perspective. *Traffic* **20**, 829-840 (2019).
4. Fribourg, S., Braun, I.C., Izaurralde, E. & Conti, E. Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. *Mol Cell* **8**, 645-56 (2001).
5. Grant, R.P., Neuhaus, D. & Stewart, M. Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1A resolution. *J Mol Biol* **326**, 849-58 (2003).
6. Mehlin, H., Daneholt, B. & Skoglund, U. Translocation of a specific premessenger ribonucleoprotein particle through the nuclear pore studied with electron microscope tomography. *Cell* **69**, 605-13 (1992).
7. Ashkenazy-Titelman, A., Atrash, M.K., Boocholez, A., Kinor, N. & Shav-Tal, Y. RNA export through the nuclear pore complex is directional. *Nat Commun* **13**, 5881 (2022).
8. Cheng, H. et al. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389-400 (2006).
9. Strasser, K. et al. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304-8 (2002).
10. Ren, Y., Schmiege, P. & Blobel, G. Structural and biochemical analyses of the DEAD-box ATPase Sub2 in association with THO or Yra1. *Elife* **6**(2017).
11. Xie, Y. et al. Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. *Elife* **10**(2021).
12. Pacheco-Fiallos, B. et al. mRNA recognition and packaging by the human transcription-export complex. *Nature* **616**, 828-835 (2023).
13. Schuller, S.K. et al. Structural insights into the nucleic acid remodeling mechanisms of the yeast THO-Sub2 complex. *Elife* **9**(2020).
14. Masuda, S. et al. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* **19**, 1512-7 (2005).
15. Luo, M.L. et al. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644-7 (2001).

16. Chang, C.T. et al. Chtop is a component of the dynamic TREX mRNA export complex. *EMBO J* **32**, 473-86 (2013).
17. Strasser, K. & Hurt, E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature* **413**, 648-52 (2001).
18. Sen, R. et al. Distinct Functions of the Cap-Binding Complex in Stimulation of Nuclear mRNA Export. *Mol Cell Biol* **39**(2019).
19. Viphakone, N. et al. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun* **3**, 1006 (2012).
20. Strasser, K. & Hurt, E. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* **19**, 410-20 (2000).
21. Stutz, F. et al. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**, 638-50 (2000).
22. Boyne, J.R., Colgan, K.J. & Whitehouse, A. Recruitment of the complete hTREX complex is required for Kaposi's sarcoma-associated herpesvirus intronless mRNA nuclear export and virus replication. *PLoS Pathog* **4**, e1000194 (2008).
23. Sandri-Goldin, R.M. The many roles of the regulatory protein ICP27 during herpes simplex virus infection. *Front Biosci* **13**, 5241-56 (2008).
24. Tunnicliffe, R.B. et al. Structural basis for the recognition of cellular mRNA export factor REF by herpes viral proteins HSV-1 ICP27 and HVS ORF57. *PLoS Pathog* **7**, e1001244 (2011).
25. Tunnicliffe, R.B., Hautbergue, G.M., Wilson, S.A., Kalra, P. & Golovanov, A.P. Competitive and cooperative interactions mediate RNA transfer from herpesvirus saimiri ORF57 to the mammalian export adaptor ALYREF. *PLoS Pathog* **10**, e1003907 (2014).
26. Huang, Y., Yario, T.A. & Steitz, J.A. A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A* **101**, 9666-70 (2004).
27. Clery, A. et al. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc Natl Acad Sci U S A* **110**, E2802-11 (2013).
28. Townsend, C. et al. Mechanism of protein-guided folding of the active site U2/U6 RNA during spliceosome activation. *Science* **370**(2020).
29. Le Hir, H., Izaurralde, E., Maquat, L.E. & Moore, M.J. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**, 6860-9 (2000).
30. Andersen, C.B. et al. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**, 1968-72 (2006).
31. Bono, F., Ebert, J., Lorentzen, E. & Conti, E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713-25 (2006).
32. Gromadzka, A.M., Steckelberg, A.L., Singh, K.K., Hofmann, K. & Gehring, N.H. A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res* **44**, 2348-61 (2016).
33. Golovanov, A.P., Hautbergue, G.M., Wilson, S.A. & Lian, L.Y. A simple method for improving protein solubility and long-term stability. *J Am Chem Soc* **126**, 8933-9 (2004).
34. Calero, G. et al. Structural basis of m7GpppG binding to the nuclear cap-binding protein complex. *Nat Struct Biol* **9**, 912-7 (2002).

35. Mazza, C., Segref, A., Mattaj, I.W. & Cusack, S. Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J* **21**, 5548-57 (2002).
36. Tate, J.G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947 (2019).
37. Heath, A.P. et al. The NCI Genomic Data Commons. *Nat Genet* **53**, 257-262 (2021).
38. Zhang, J. et al. The International Cancer Genome Consortium Data Portal. *Nat Biotechnol* **37**, 367-369 (2019).
39. Viphakone, N. et al. Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol Cell* **75**, 310-323 e8 (2019).
40. Dufu, K. et al. ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. *Genes Dev* **24**, 2043-53 (2010).
41. Xie, Y. et al. Structural basis for high order complex of SARNP and DDX39B to facilitate mRNP assembly. *Cell Rep* (accepted).
42. Blobel, G. Gene gating: a hypothesis. *Proc Natl Acad Sci U S A* **82**, 8527-9 (1985).
43. Burns, L.T. & Wente, S.R. From hypothesis to mechanism: uncovering nuclear pore complex links to gene expression. *Mol Cell Biol* **34**, 2114-20 (2014).
44. Scholz, B.A. et al. WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating. *Nat Genet* **51**, 1723-1731 (2019).
45. Zhang, K. et al. Structural basis for influenza virus NS1 protein block of mRNA nuclear export. *Nat Microbiol* **4**, 1671-1679 (2019).
46. Zhang, K. et al. Nsp1 protein of SARS-CoV-2 disrupts the mRNA export machinery to inhibit host gene expression. *Sci Adv* **7**(2021).
47. Momose, F. et al. Cellular splicing factor RAF-2p48/NPI-5/BAT1/UAP56 interacts with the influenza virus nucleoprotein and enhances viral RNA synthesis. *J Virol* **75**, 1899-908 (2001).
48. Morris, A.K. et al. Cellular mRNA export factor UAP56 recognizes nucleic acid binding site of influenza virus NP protein. *Biochem Biophys Res Commun* **525**, 259-264 (2020).
49. Lenasi, T., Peterlin, B.M. & Barboric, M. Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive transcription elongation factor b (P-TEFb). *J Biol Chem* **286**, 22758-68 (2011).
50. Pabis, M. et al. The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* **19**, 1054-63 (2013).
51. Das, S. & Krainer, A.R. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Mol Cancer Res* **12**, 1195-204 (2014).
52. Pandit, S. et al. Genome-wide analysis reveals SR protein cooperation and competition in regulated splicing. *Mol Cell* **50**, 223-35 (2013).
53. Reed, R. & Cheng, H. TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol* **17**, 269-73 (2005).
54. Muller-McNicoll, M. et al. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* **30**, 553-66 (2016).
55. Okada, M., Jang, S.W. & Ye, K. Akt phosphorylation and nuclear phosphoinositide association mediate mRNA export and cell proliferation activities by ALY. *Proc Natl Acad Sci U S A* **105**, 8649-54 (2008).

56. Xie, Y. et al. Structure and activation mechanism of the yeast RNA Pol II CTD kinase CTDK-1 complex. *Proc Natl Acad Sci U S A* **118**(2021).
57. Hurt, E., Luo, M.J., Röther, S., Reed, R. & Strässer, K. Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. *Proc Natl Acad Sci U S A* **101**, 1858-62 (2004).
58. Nojima, T., Hirose, T., Kimura, H. & Hagiwara, M. The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J Biol Chem* **282**, 15645-51 (2007).
59. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
60. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
61. Afonine, P.V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-67 (2012).
62. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-4 (2014).

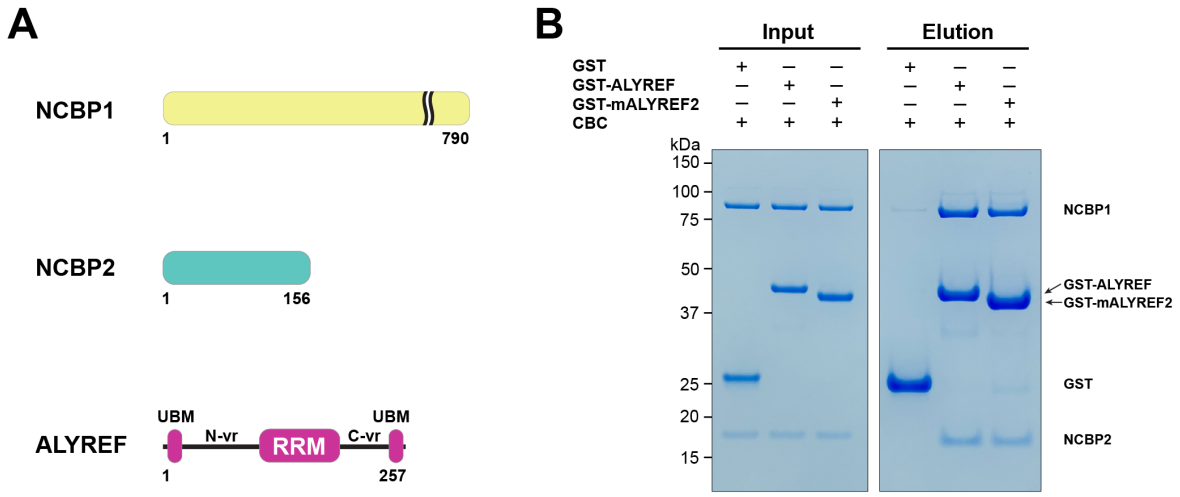

Figure 1. ALYREF directly binds to CBC.
(A) Schematics of the NCBP1 and NCBP2 subunits of the human CBC and ALYREF. ALYREF contains a central RRM domain. The RRM domain is connected to two conserved motifs (UBMs) at both termini through variable regions (N-vr and C-vr). (B) ALYREF directly interacts with the CBC. In vitro GST pull down assays were performed with purified recombinant human CBC and GST-tagged ALYREF (residues 1-183) or mALYREF2 (residues 1-155). Results are representative of three independent experiments.

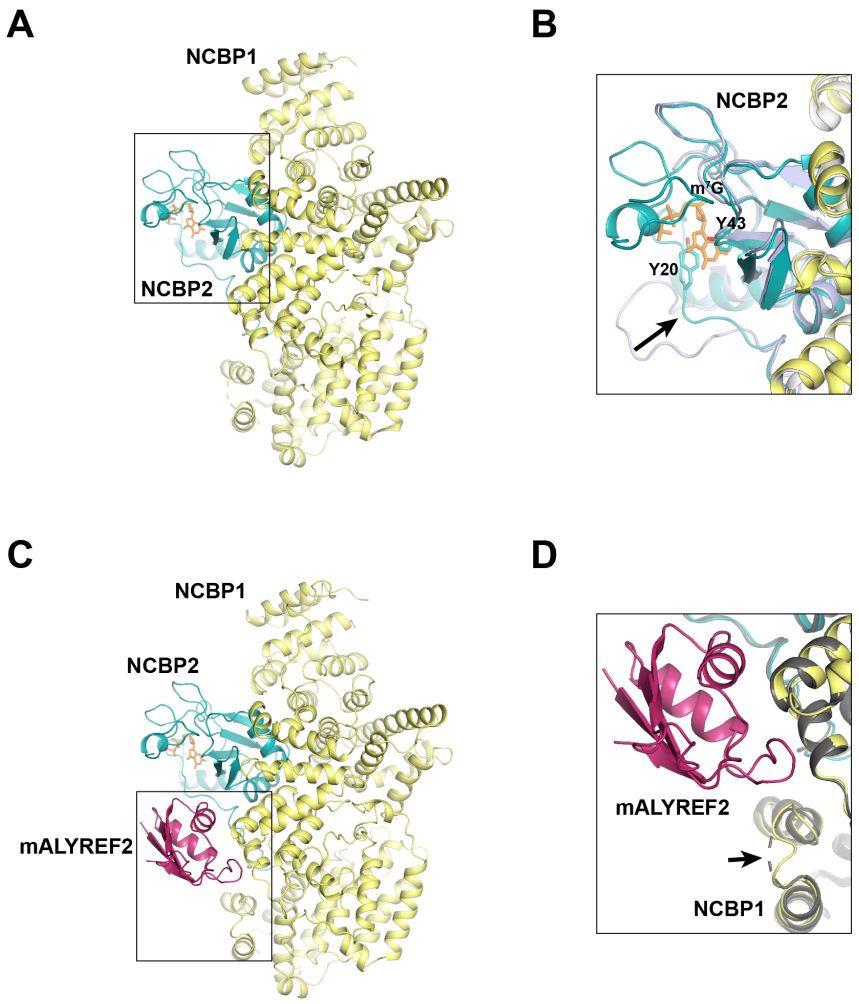

Figure 2. Cryo-EM structures of CBC and CBC-mALYREF2
(A) Overall architecture of the CBC complex. NCBP1 and NCBP2 are colored in yellow and teal, respectively. The cap analog is shown as orange sticks. (B) Comparison of the CBC cryo-EM structure to the unliganded CBC crystal structure (PDB ID 1N54). The cryo-EM structure is colored as in (A). Unliganded CBC crystal structure is colored in light blue. Arrow indicates the conformational change in the N-terminal extension of NCBP2 upon cap analog binding. (C) Overall architecture of the CBC-mALYREF2 complex. NCBP1, NCBP2, and mALYREF2 are colored in yellow, teal, and red, respectively. The RRM domain of mALYREF2 binds to the CBC. (D) Comparison of the CBC-mALYREF2 and the CBC cryo-EM structures. CBC structure is colored in gray. Arrow indicates conformational changes in a loop (residues 38-45) of NCBP1 upon mALYREF2 binding.

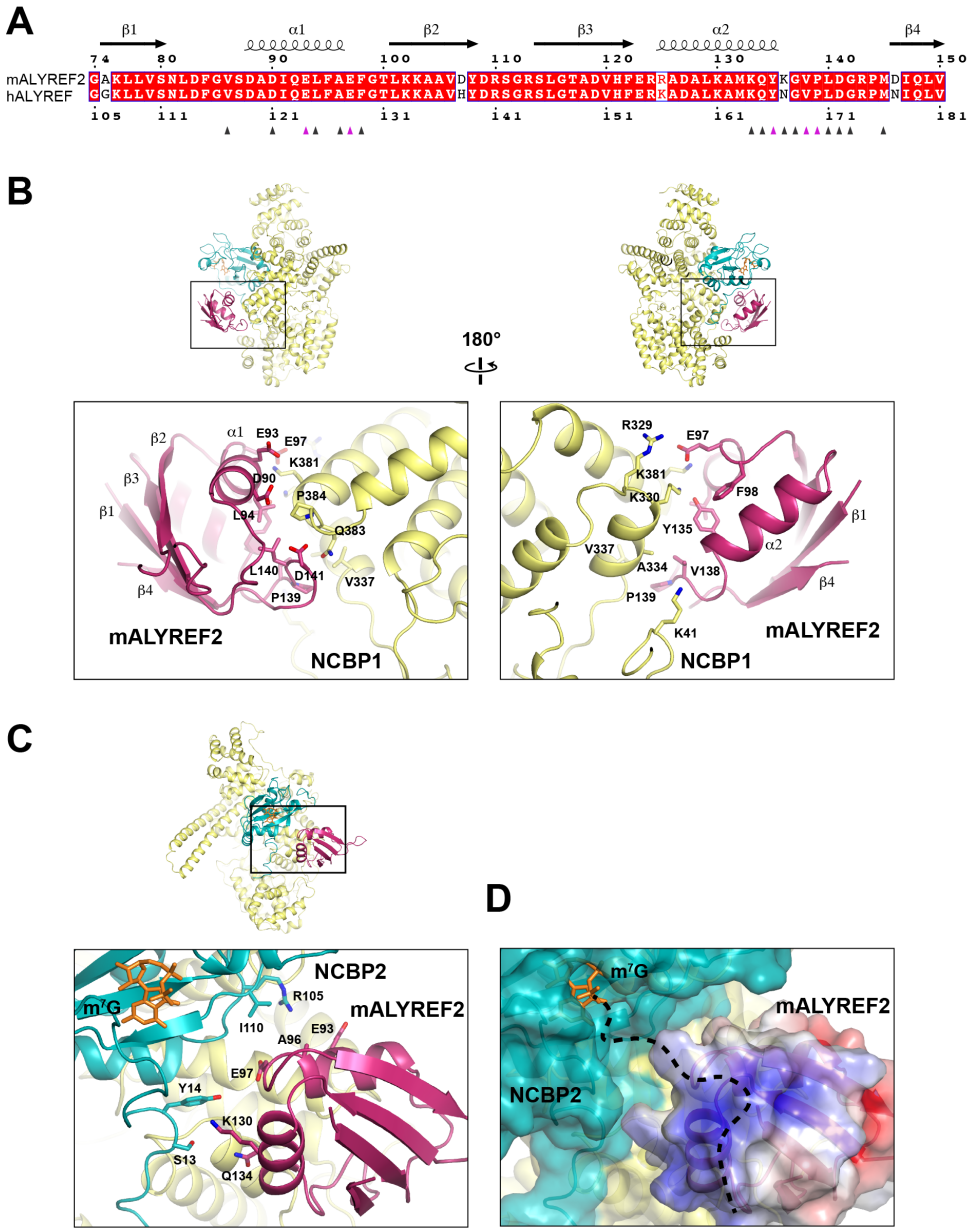

Figure 3. ALYREF binds to both NCBP1 and NCBP2.
(A) Sequence alignment of the RRM domain of mALYREF2 and ALYREF[62]. The residues that interface with the CBC are indicated by triangles below the sequence. The triangles colored in purple correspond to residues subjected to mutagenesis in Figure 4. (B) Details of the interaction between NCBP1 and mALYREF2. As a reference, an overall model is shown on the top to indicate the zoomed-in area. (C) Details of the interaction between NCBP2 and mALYREF2. (D) mALYREF2 features a positively charged surface near the cap analog bound to NCBP2. Surface of mALYREF2 is colored according to the electrostatic potential, ranging from red (-5 k_BT/e) to blue (+5 k_BT/e). Dotted line indicates putative RNA binding path.

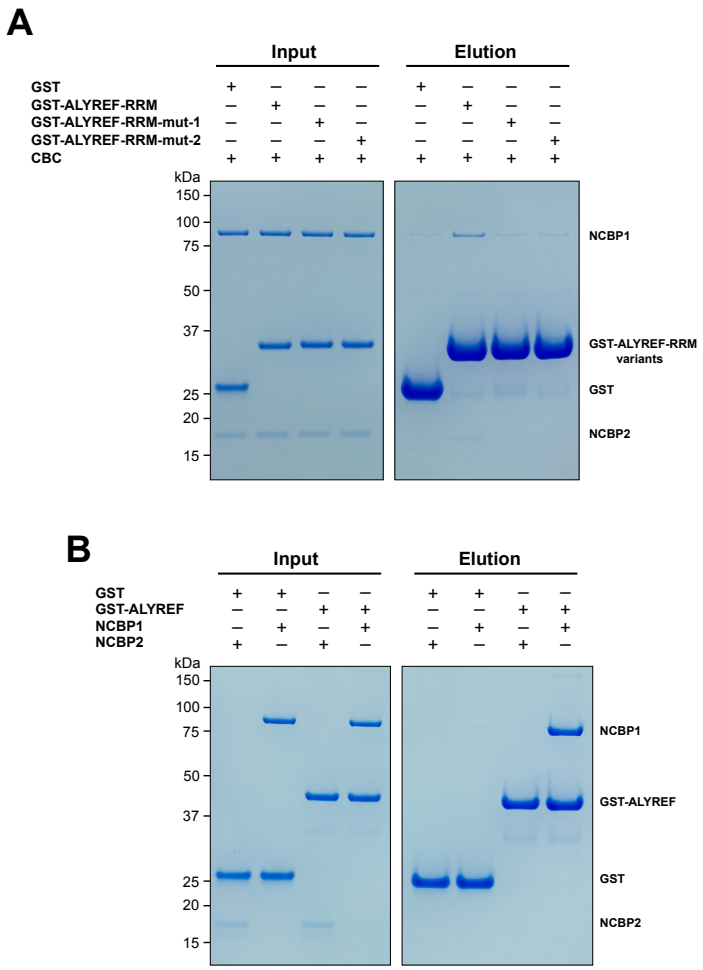

Figure 4. Dissection of the ALYREF and CBC interfaces.
(A) Mutations of key interface residues on ALYREF reduced CBC binding. In vitro GST pull down assays were performed with purified recombinant human CBC and GST-tagged ALYREF-RRM wild type or mutants (mut-1, Y166R/V169R/P170R; mut-2, E124R/E128R). (B) NCBP1 is sufficient to interact with ALYREF. In vitro GST pull down assays were performed with purified recombinant GST-tagged ALYREF and individual CBC subunits. Results are representative of three independent experiments.

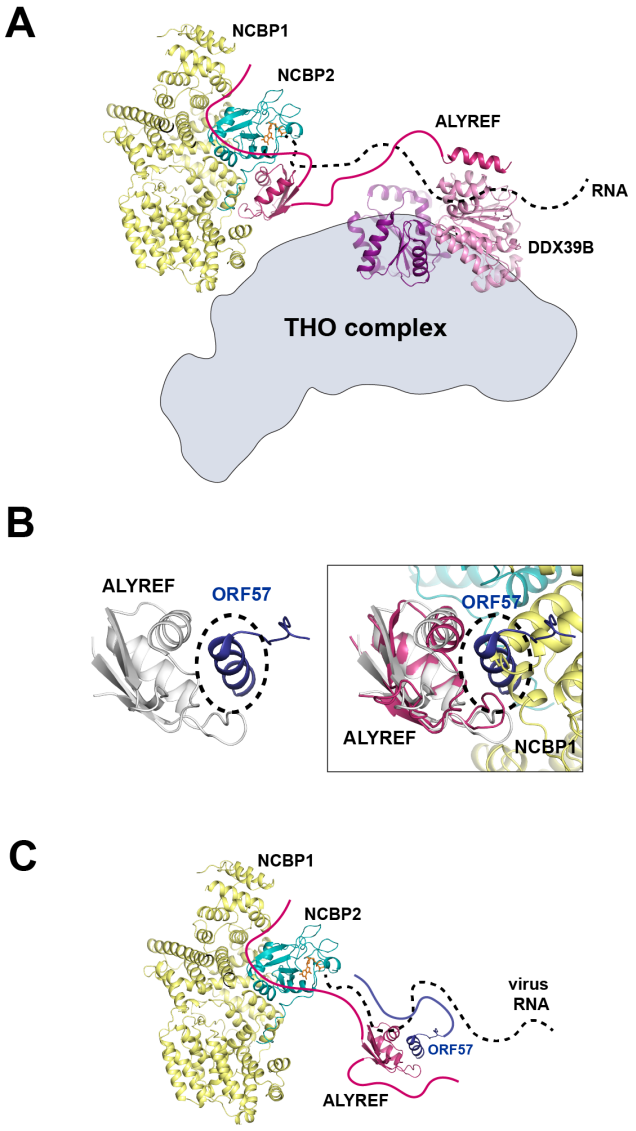

Figure 5. Functional implications for ALYREF and CBC in 5' cap dependent mRNP export.
(A) ALYREF links the TREX complex to the CBC. The RRM domain of ALYREF recognizes the CBC at the 5' end of mRNA. The UBM of ALYREF binds to the DDX39B subunit of TREX; their complex is represented using their yeast orthologs Yra1 and Sub2, respectively (PDB ID 5SUP). DDX39B in turn associates with the THO subcomplex of TREX. (B) HVS ORF57 binds to the RRM domain of ALYREF and interferes with the CBC-ALYREF interaction. (left) NMR structure of the ALYREF-ORF57 complex (PDB ID 2YKA). (right) The CBC-ALYREF structure is overlayed with the ALYREF-ORF57 structure. (C) Model of viral mRNA export mediated by herpes viral ORF57 homologs. On viral transcripts, ALYREF associates with the CBC and ORF57 via the WxHD motif and the RRM domain, respectively. Both ALYREF and ORF57 feature RNA binding regions to form contacts with the RNA. ALYREF recruits the other TREX complex components to facilitate nuclear export of viral mRNAs.

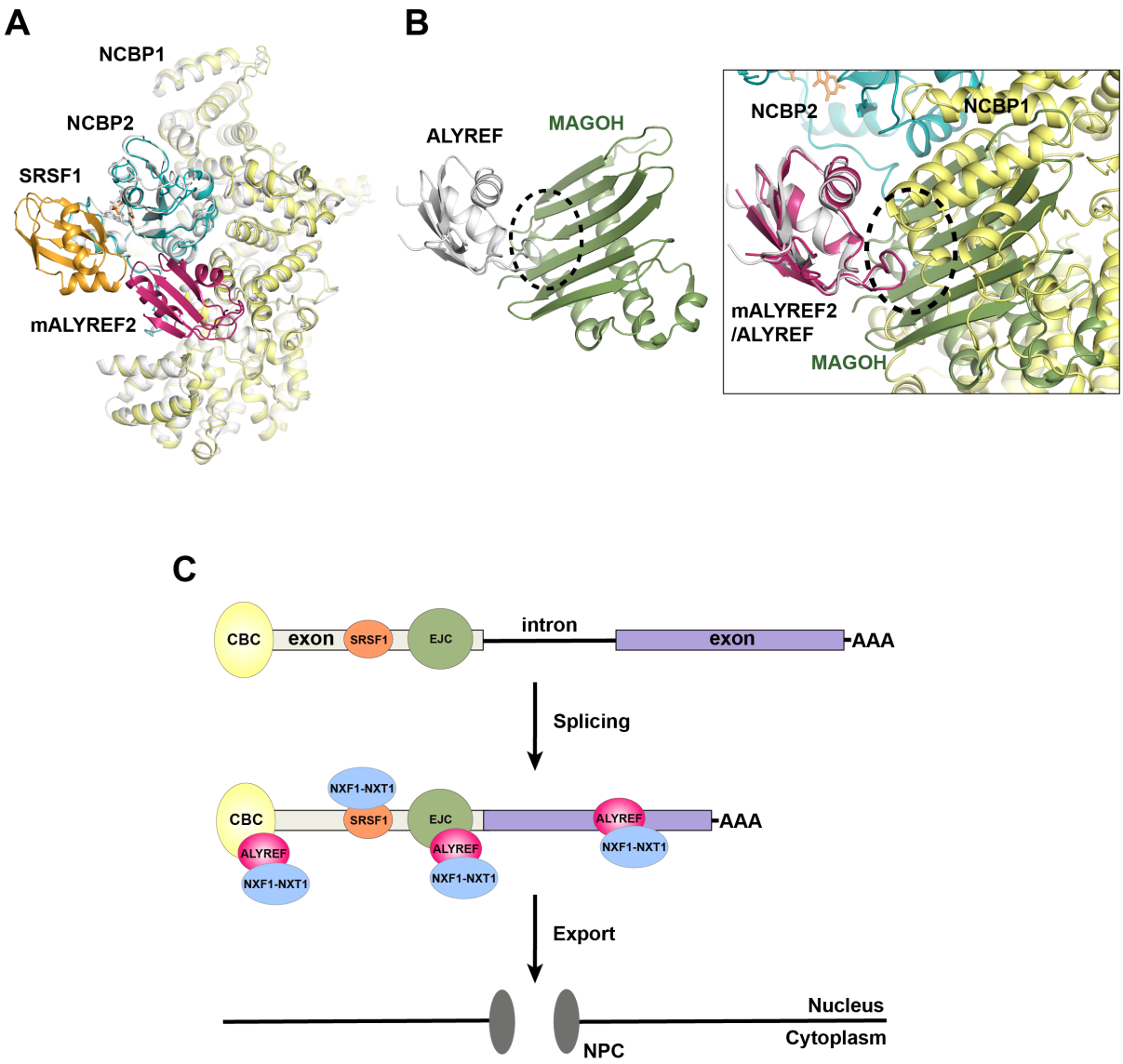

Figure 6. ALYREF and CBC in splicing and export
(A) Overlay of the CBC-ALYREF structure with the CBC-SRSF1 structure (PDB ID 7ABG). CBC-ALYREF is colored as in Figure 2B. CBC-SRSF1 is colored with the CBC in white and SRSF1 in orange. (B) ALYREF interaction with the CBC and the EJC is mutually exclusive. (left) ALYREF binds to the MAGOH subunit of the EJC (PDB ID 7ZNJ). (right) The CBC-ALYREF structure is overlayed with the ALYREF-MAGOH structure. (C) Model of the mRNP export receptor NXF1-NXT1 recruitment by CBC-ALYREF and other factors during mRNA maturation.

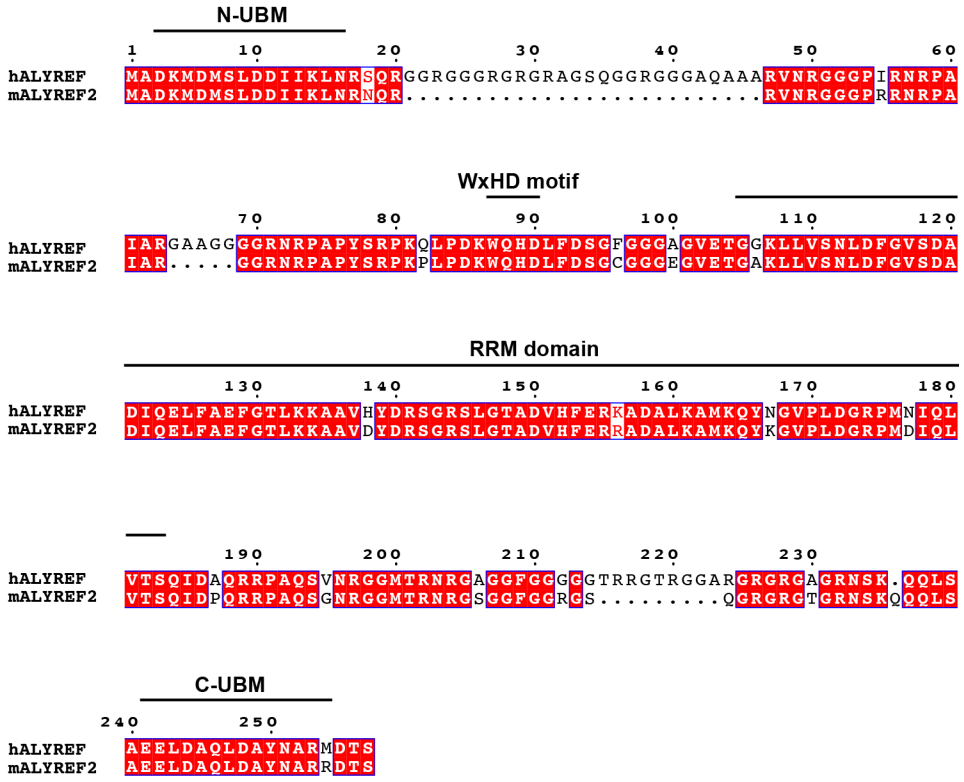

Figure 1 – figure supplement 1. Human ALYREF and mouse ALYREF2 are conserved.
Sequence alignment of ALYREF and mALYREF2[62]. The DDX39B binding motifs (N-UBM and C-UBM), the WxHD motif, and the RRM domain are indicated.

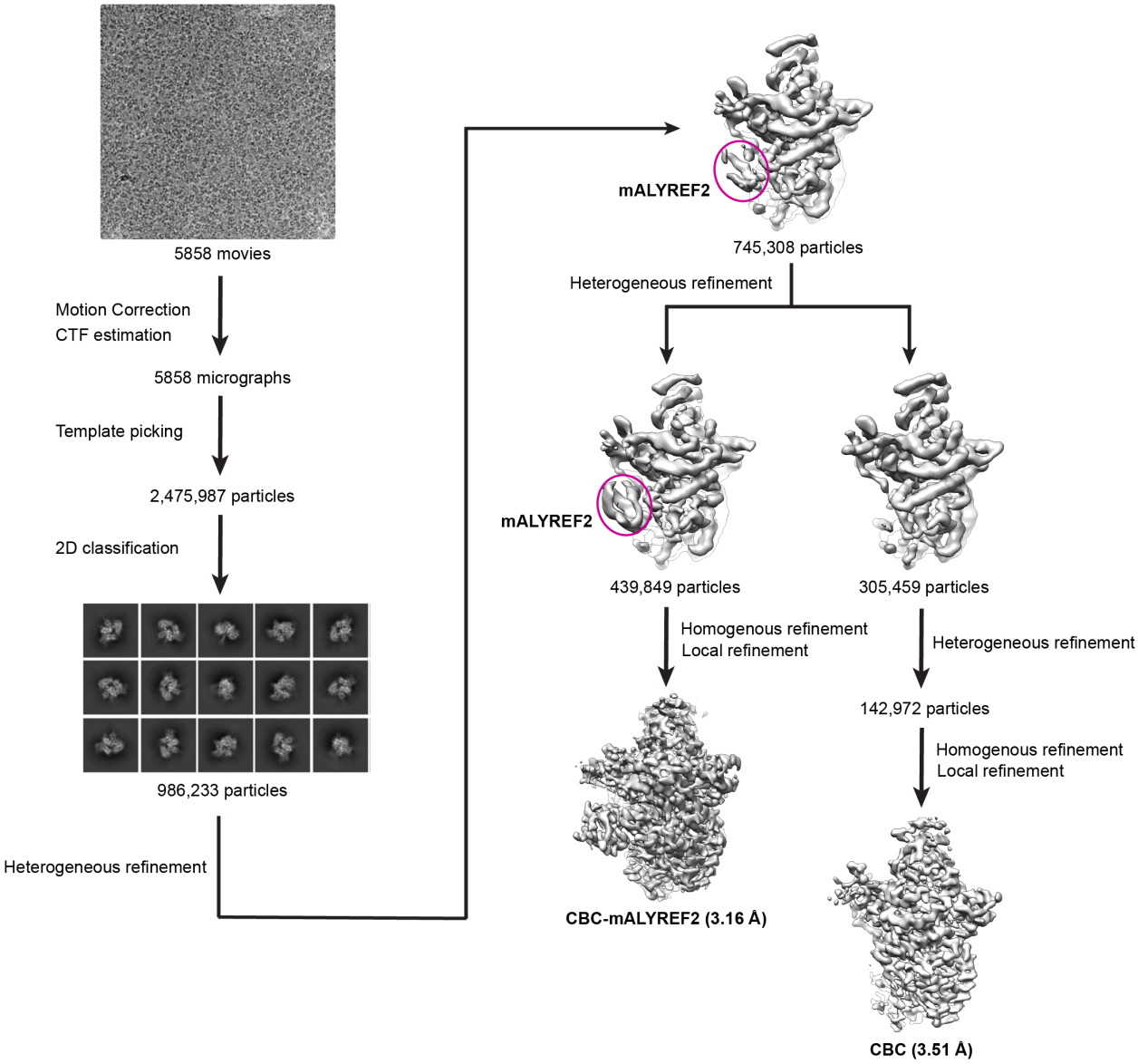

Figure 2 – figure supplement 1. Workflow for cryo-EM data processing.

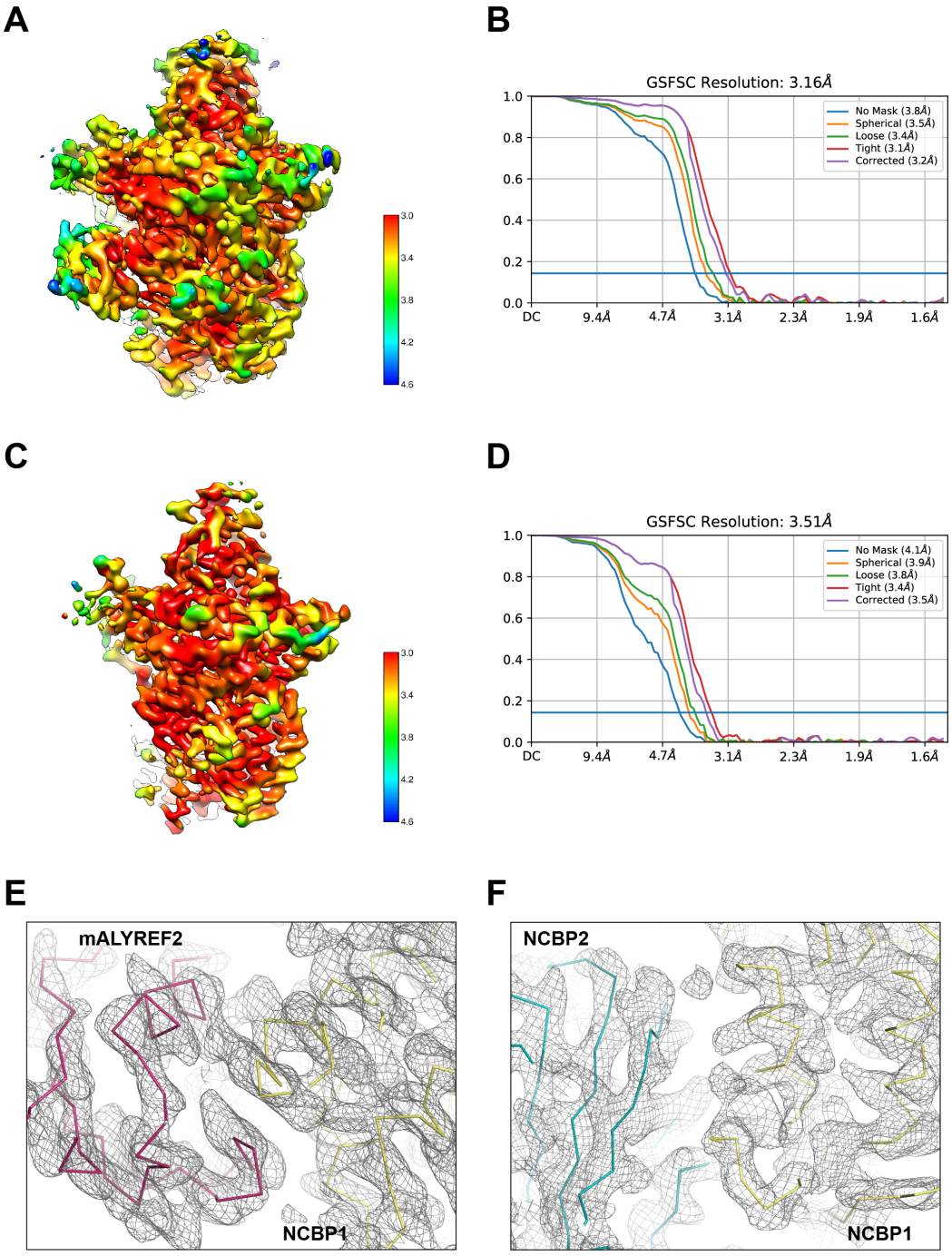

Figure 2 – figure supplement 2. Local resolution estimation and FSC curves. (A) Local resolution of the CBC density map. (B) FSC curves between the two half maps of CBC. (C) Local resolution of the CBC-mALYREF2 density map. (D) FSC curves between the two half maps of CBC-mALYREF2. (E) Electron density map of CBC-mALYREF2 at the interface between NCBP1 and mALYREF2. (F) Electron density map of CBC-mALYREF2 at the interface between NCBP1 and NCBP2.

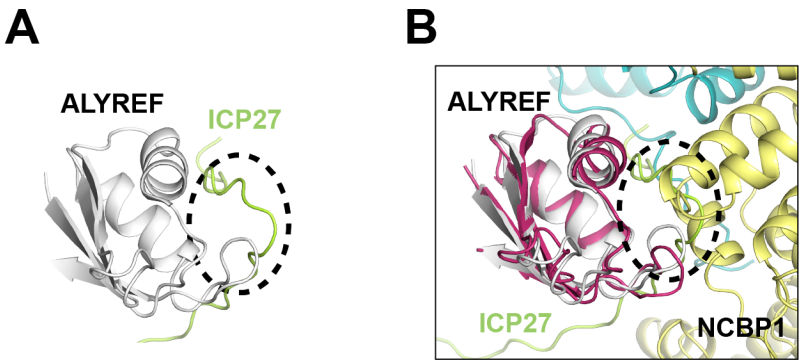

Figure 5 – figure supplement 1. HSV-1 ICP27 binds to the RRM domain of ALYREF and interferes with the CBC-ALYREF interaction. (A) NMR structure of the ALYREF-ICP27 complex (PDB ID 2KT5). (B) The CBC-ALYREF structure is overlayed with the ALYREF-ICP27 structure.

Table 1. Cryo-EM data collection, refinement and validation statistics

	CBC-mALYREF2 (EMDB EMD-####) (PDB ####)	CBC (EMDB EMD-####) (PDB ####)
Data collection and processing		
Microscope/Camera	Glacios/Falcon 4i	
Voltage (kV)	200	
Electron exposure (e–/Å^2)	52	
Defocus range (μm)	1.0 to 2.0	
Pixel size (Å)	0.732	
Symmetry imposed	C1	C1
Final particle images (no.)	439,849	142,972
Resolution at 0.143 FSC (unmasked, Å)	3.8	4.1
Resolution at 0.143 FSC (masked, Å)	3.16	3.51
Refinement		
Model composition		
Protein residues	974	880
Ligands	1	1
B factors (Å^2)		
Protein	208.5	229.1
Ligand	215.8	239.8
R.m.s. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.436	0.464
Validation		
MolProbity score	1.59	1.64
Clashscore	10.09	11.29
Poor rotamers (%)	0	0
Ramachandran plot		
Favored (%)	97.7	97.7
Allowed (%)	2.3	2.3
Disallowed (%)	0	0