Computational and Structural Biotechnology Journal 20 (2022) 4025–4039

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

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

Virus-host interaction analysis in colorectal cancer identifies core virus network signature and small molecules

Sai Krishna A.V.S [a,1], Swati Sinha [a], Sainitin Donakonda [b,*,2]

[a] Department of Biotechnology, Faculty of Life and Allied Health Sciences, MS Ramaiah University of Applied Sciences, Bengaluru, India
[b] Institute of Molecular Immunology and Experimental Oncology, Klinikum Rechts Der Isar, Technical University of Munich, Munich, Germany

ARTICLE INFO

Article history:
Received 15 April 2022
Received in revised form 23 July 2022
Accepted 23 July 2022
Available online 28 July 2022

Keywords:
Colorectal cancer
Virus-host interactions
Small molecules
Docking
Molecular dynamics

ABSTRACT

Colorectal cancer (CRC) is a significant contributor to cancer-related deaths caused by an unhealthy lifestyle. Multiple studies reveal that viruses are involved in colorectal tumorigenesis. The viruses such as Human Cytomegalovirus (HCMV), Human papillomaviruses (HPV16 & HPV18), and John Cunningham virus (JCV) are known to cause colorectal cancer. The molecular mechanisms of cancer genesis and maintenance shared by these viruses remain unclear. We analysed the virus-host networks and connected them with colorectal cancer proteome datasets and extracted the core shared interactions in the virus-host CRC network. Our network topology analysis identified prominent virus proteins RL6 (HCMV), VE6 (HPV16 and HPV18), and Large T antigen (JCV). Sequence analysis uncovered short linear motifs (SLiMs) in each viral target. We used these targets to identify the antiviral drugs through a structure-based virtual screening approach. This analysis highlighted that temsavir, pimodivir, famotine, and bictegravir bind to each virus protein target, respectively. We also assessed the effect of drug binding using molecular dynamic simulations, which shed light on the modulatory effect of drug molecules on SLiM regions in viral targets. Hence, our systematic screening of virus-host networks revealed viral targets, which could be crucial for cancer therapy.

1. Introduction

Colorectal cancer (CRC) was identified as the third most prevalent cancer worldwide, with continuous increase in instances. It is associated with approximately 9% of cancer deaths [1]. Although the familial history of CRC is demonstrated, many CRC cases are sporadic rather than familial [2]. Multiple aspects such as heredity, epigenetics, lifestyle, and pathogens are identified as potential risk factors. Recent studies have established the role of viruses in many cancers and it is shown that 9.9% of cancers are associated with viral infections [3]. Viruses such as Human Cytomegalovirus (HCMV), Human papillomaviruses (HPV16 & HPV18), and John Cunningham virus (JCV) showed higher prevalence in tumour-associated colorectal tissues [4]. In a study conducted by Cobbs et al., [5] it was shown that 82% of the colorectal polyps had HCMV tumour promoting viral proteins (IE1-72 and pp65) in comparison to with adjacent non-neoplastic colon biopsy samples. A meta-analysis which combined all the outcomes from the majority of the colorectal cancer studies showed statistically significant level of HPV infection present in CRC tumor tissue [6]. Laghi, L et al [7] demonstrated that 81–89% of CRC tissues contained JCV, and the virus was predominant in CRC sections compared to the adjacent normal tissues.

HCMV is a member of the beta-herpes virus subfamily. HCMV can co-exist with its host and can recrudescent at any moment, resulting in substantial morbidity and even death, particularly in immunocompromised patients [8]. According to several reports, HCMV nucleic particles were identified in tissue samples from patients with colorectal cancer [9,10]. Some of these viral proteins enhance cellular processes like proliferation, angiogenesis, and

Abbreviations: CRC, Colorectal cancer; HCMV, Human Cytomegalovirus; HPV16 and 18, Human papillomavirus; JCV, John Cunningham virus; DEPs, Differentially expressed proteins; NH, Network heterogeneity; NC, Network centralization; SLiMs, Short linear motifs; ELM, Eukaryotic Linear Motif Resource; IDR, Intrinsically disordered region; PME, Particle-Mesh Ewald; RMSD, Root-Mean-Square Deviation; RMSF, Root-Mean-Square Fluctuation; hbonds, Hydrogen bonds; MM-PBSA, molecular mechanics – Poisson Boltzmann surface area; SASA, Solvent Accessible Surface Area.

* Corresponding author.
E-mail address: sainitin.donakonda@tum.de (S. Donakonda).
[1] Present address: Chromatin Biology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bengaluru, India.
[2] ORCID ID: orcid.org/0000-0003-3216-8759.

https://doi.org/10.1016/j.csbj.2022.07.040

metastasis, while others suppress the localized immune response [11]. HPV is a family of DNA viruses well known for its role in the occurrence and progression of many cancer types. There are currently 200 HPV genotypes known, with at least 18 belonging to the "high-risk" group primarily responsible for cancer development. Damin et al. [12], showed that the expression of viral (HPV) oncoproteins could induce carcinogenesis. Recent reports have demonstrated the presence of HPV16 [13] and HPV18 [14] DNA or antigen in clinical colorectal samples indicating the possible risk of HPV in CRC. Similarly, JCV being an opportunistic pathogen, is distinctly present in the gastrointestinal tract and remains latent in various organs such as kidneys, and B-lymphocytes [15]. Studies by Mou et al. [16] revealed the persistent association between JCV and colorectal tumor samples while Ksiaa et al. [17] manifested its tumorigenic role. Large T antigen, encoded by JCV, plays a critical role in cancer development by intervening in cellular phenotypes. It is efficient in dysregulating the cell cycle process by interacting with essential proteins (p53 and pRb) and other signalling pathways [18]. The mechanism of such viral tumorigenesis can be attributed to Short Linear Motifs (SLiMs). It is observed that, viral pathogens have developed approached for hijacking host cell machinery via SLiMs [19].

The knowledge of all virus-encoded protein interactions as well as the differential expression of host proteins in infected cells is essential for gaining a deeper understanding of complex diseases. In recent years, meta-analyses of host-virus protein interactions gained prominent importance in providing novel insights for designing comprehensive antiviral treatment strategies. Virtual screening approaches enable us to screen for the antiviral molecules against viral targets and understanding their behaviour using molecular dynamics simulations and permits us to identify the potential therapeutics.

In our study, we chose the four viruses, HCMV, HPV16, HPV18, and JCV. Although the evidence of these viruses in CRC is established, the core molecular interactions through which these viruses target host (human) cellular machinery remains exploratory. Our network analysis identified the core viral-host interaction network involved in CRC and its associated pathway. This study enabled us to screen antiviral molecules against the viral proteins allowing us to discover the potential drug candidates.

2. Materials and methods

2.1. Construction of virus host networks

We retrieved the virus-host interactions related to HCMV from Nobre LV et al. [20]. HPV16 and 18 virus-host interactomes were extracted from STRING viruses v10.5 database (accessed on: 30/04/2021) (https://viruses.string-db.org/) [21]. We used a medium confidence score of 0.4 to retrieve the maximum number of interactions. JCV interactions are extracted from the human-virus interaction database (accessed on: 30/04/2021) (HVIDB, https://zzdlab.com/hvidb/) [22]. We rendered these network diagrams in Cytoscape v3.8.2 [23].

2.2. Literature mining of colorectal cancer proteome

Through literature mining, we culled the differentially expressed proteins (DEPs) in human CRC proteomes. We retrieved the CRC proteins from Maja et al. [24], which gave 900 DEPs. Next, we salvaged 107 and 2778 CRC proteins from Buttacavoli et al. [25] and Saleem et al. [26] respectively. Finally, we combined these lists of CRC proteins, made a unique list of 3092 proteins, and mapped the fold changes to virus-host networks to derive the virus-host CRC protein interaction networks.

2.3. Mining of core virus–host CRC network

We computed the degree for each virus-host CRC interaction network and identified the top 3 viral proteins from each virus type, to derive the subnetworks. Then, we merged these networks and extracted a shared network; this was refined further by selecting the top one virus protein from each virus and identified the core virus-host CRC network.

2.4. Pathway enrichment analysis

We performed the pathway enrichment analysis of proteins from virus-host interaction networks using METASCAPE server (https://metascape.org/gp/index.html#/main/step1) (accessed on: 14/06/2021) [27], which contained the gene ontology biological process. The pathways were considered statistically significant with a p-value $\leq$0.05.

2.5. Short linear motif identification and disorder prediction

We used the Eukaryotic Linear Motif (ELM) resource (accessed on: 15/06/2021) (http://elm.eu.org/) [28] to identify the human proteins having short linear motifs that share a similar region with viral proteins. We used p-value $\leq$0.01 to filter out the statistically significant motifs. Disprot (https://disprot.org/) (accessed on: 16/06/2021 and 21/06/2022) [29] was used to identify the disordered regions in viral proteins. Then, we identified the amino acid residues contributing to the overall disorderness of the protein (disorder score cut-off >0.5).

2.6. Retrieval of three-dimensional protein structures and druggability analysis

We retrieved the 3D X-ray based structures of VE6_HPV16 (PDB id: 6SJA) and VE6_HPV18 (PDB id: 6SJV) from RCSB Protein Data Bank (https://www.rcsb.org/) [30]. We employed transform restrained Rosetta (trRosetta) algorithm (https://yanglab.nankai.edu.cn/trRosetta/) (accessed on: 17/06/2021) which uses deep learning and direct energy minimization to predict the protein structure [31]. While RL6_HCMV structure was modelled based on the de novo approach, large T antigen_JCV was modelled using deep learning approach combined with homology strategy. Its sequence was aligned to the crystal structure of SV40 large T antigen from Macaca mulatta polyomavirus 1 with sequence identity of 76.5 and raw alignment z-score of 30.044. Next, we computed the z-score of homology models to compare them with the experimentally derived structures (X-ray or NMR) using the Protein structure analysis (ProSA-web) tool (https://prosa.services.came.sbg.ac.at/prosa.php) [32]. We used Molprobity server (accessed on: 18/06/2021) [33] to generate a Ramachandran plot and z-score [34] to predict the stereochemical quality of protein structural models. We mapped SLiM binding sites onto 3D protein structures, and computed druggability score using the PockDrug web server (accessed on: 19/06/2021 and 23/06/2022) [35]. 3D structures of HIV proteins (gp120 and integrase) were also retrieved from RCSB Protein Data Bank [30]. From the uniport database [36] annotations (accessed on: 21/06/22), we found that the surface protein gp120 is a part of envelope glycoprotein gp160 (PDB id: 3J70). So, we considered the sequence range corresponding to gp120 (33–511) for further analysis. Similarly, integrase is found to be the part of Gag-Pol Polyprotein (PDB id: 3NFA) that range between 1148 and 1435 amino acids. Therefore, the same region is chosen for SLiM region identification and docking studies.

2.7. Antiviral compounds

We extracted the antiviral compounds from the CheMBL database (accessed on: 23/06/2021) (https://www.ebi.ac.uk/chembl/) [37]. We filtered these antiviral compounds based on the Lipinski rule of 5 (RO5), a measure of drug likeliness [38]. This filtering process gave rise to 190 antiviral compounds. We mapped the CheMBL IDs of these antiviral compounds to PubChem IDs using the Pubchem Identifier Exchange Service. These PubChem IDs were uploaded to the PubChem database (https://pubchem.ncbi.nlm.nih.gov) (accessed on: 23/06/2021) and downloaded the 3D conformers in structural data file (.SDF) format. They were converted to protein data bank format (.PDB) at physiological pH 7 using the OpenBabel v2.4 software [39] and used these formatted files for virtual screening.

2.8. Molecular docking

We used the virtual screening software PyRx v0.9.9 [40] to perform the docking. We used the make macromolecule option to prepare protein structures by removing water, adding hydrogens and partial charges. Make ligand command is utilized to prepare antiviral compounds and converted them to PDBQT format. Then, we initiated the docking using AutoDockVina, which is an integral part of PyRx tool docking to calculate the binding affinities. Keeping the exhaustiveness at 8, we centred the search space and receptor grids around the binding site (SLiM region in the protein targets). This was followed by the construction of XYZ center coordinates and grid boxes for each protein target: RL6_HCMV (center coordinates: −0.48 Å, 0.23 Å, 1.50 Å, grid box: 28.03 Å, 28.36 Å, 23.26 Å), VE6_HPV16 VE6 (center coordinates: −33.13 Å, 61.79 Å, −28.46 Å, grid box: 27.92 Å, 25.22 Å, 25.87 Å), VE6_HPV18 (center coordinates: −30.07 Å, 103.80 Å, 22.41 Å, grid box: 22.41 Å, 18.68 Å, 48.95 Å) and large T antigen_JCV (center coordinates: 103.319 Å,1 6.620 Å, 26.142 Å, grid box: 19.888 Å, 21.509 Å, 25.421 Å). GP120_HIV (center coordinates: 265.19 Å, 159.01 Å, 201.73 Å, grid box: 30.18 Å, 30.17 Å, 33.94 Å) and Integrase_HIV (center coordinates: 265.19 Å, 159.01 Å, 201.73 Å, grid box: 30.18 Å, 30.17 Å, 33.94 Å).

2.9. Molecular dynamics simulation analysis and binding free energy calculations

The molecular dynamics (MD) simulations were carried out using the Gromacs 2020 package [41]. To initiate MD, we retrieved the apo form of proteins as well as target-drug complexes from docked poses. Using GROMACS and the SwissParam web server (accessed on: 12/07/2021) [42], we adopted the CHARMM27 force fields for protein targets and drug molecules. We performed five simulations for the viral target-drug complex including the one with apo-form of the protein target (without the drug). The pdb2gmx command and the SwissParam server [42] were used to construct the topologies for protein and drug. Once, the drug molecule topologies are re-joined to the protein structures, they were inserted into a cubic box using the explicit TIP3 water model with a buffering distance of 1.2 nm. To neutralize the net charges in the system, we added Na and Cl ions to the system. Next, these complexes were subjected to a minimization procedure with the Steepest Descent method for 2000 steps. The systems equilibrated in two steps: NVT and NPT ensembles run with the Brendsen coupling algorithm in the periodic boundary conditions. The temperature was kept constant at 310 K with pressure at 1 bar. The electrostatic interactions with an interpolation order of four and grid spacing of 0.12 nm and constrained using the LINCS algorithm computed with Particle-Mesh Ewald (PME). All the simulations were set at 2 fs and ran for 100 ns. We analysed the MD trajectories with built-in GROMACS, gmx rms, gmx rmsf, gmx hbond commands. The Molecular Mechanics energies combined with Poisson-Boltzmann (MM-PBSA) approach was used to compute the binding free energies of protein-drug complexes at last 50 ns MD trajectories with the g_mmpbsa tool [43].

2.10. Data visualization and statistical analysis

The statistical analysis and visualization in the study was done using the R statistical software v4.0 (https://www.r-project.org) unless stated otherwise. We used Cytoscape plug-in, Network analyser [44], to compute the network heterogeneity and centralization parameters. The bar plots, topology parameters such as degree, network centralization, heterogeneity, and binding affinities were visualized using the ggplot2 R package. We generated the random network using the Erdos Renyi G(n,p) model, part of the Network Randomizer Cytoscape plug-in. We plotted number of common and unique proteins as a barplot using the UpSet R package. MD trajectories were visualized as line plots using xmgrace tool. Statistical analysis of RMSF values rendered as Boxplots, Wilcoxon-test was calculated using the ggpubr R package, and p-value ≤ 0.05 was considered as significant. We visualized Protein structures and docked complexes in Pymol v2.7 software. The discovery studio visualizer (https://discover.3ds.com/discovery-studio-visualizer-download) was used to visualize the molecular interactions between the drug and protein structure.

3. Results

3.1. In-silico workflow to understand the virus-host relationship

Viruses such as HCMV, HPV16, HPV18, and JCV are known to be the risk factors for the occurrence and development of CRC. In this study, we aimed to understand the virus-host interactions in CRC, and for this purpose, we designed an *in-silico* workflow (Fig. 1A). First, we retrieved the virus-host protein interactions and constructed the networks. Following the literature search, we extracted the CRC associated proteins and mapped to the virus-host interactomes to isolate the virus-host CRC networks. We analyzed the topological parameters of the network and performed pathway enrichment analysis to extract the core virus-host CRC network. Then, we identified the SLiMs the highest ranked protein for each of the four viruses and predicted the disorder regions. We extracted antiviral molecules from the ChemBL database and performed molecular docking with the viral protein targets identified from the core virus-host CRC network. Finally, we analyzed the molecular dynamic interactions of virus-protein drug complexes.

3.2. Retrieval of virus-host networks and impact of viruses on host

Viruses are known to interact with host cellular machinery, rewire molecular pathways, and promote their replication [45]. Therefore, we sought to understand the HCMV, HPV16, HPV18, and JCV relationship with the host (human) concerning the development and advancement of CRC pathobiology. To do this, we retrieved the virus-host interactions of these viruses (Supplementary Fig. 1A–D). We constructed the networks composed of virus-specific proteins and their interactions. As HCMV has the largest genome size, this virus thus has the highest number of proteins (n = 167) connecting with 2175 host proteins (Fig. 1B). While the HPV16 has seven proteins interacting with 100 host proteins (Fig. 1C), HPV18 and JCV viruses have five proteins each collaborating with 47 and 51 host proteins respectively (Fig. 1D-E). Since biological networks are known to be densely connected and participate in various functions, we computed topological parame-

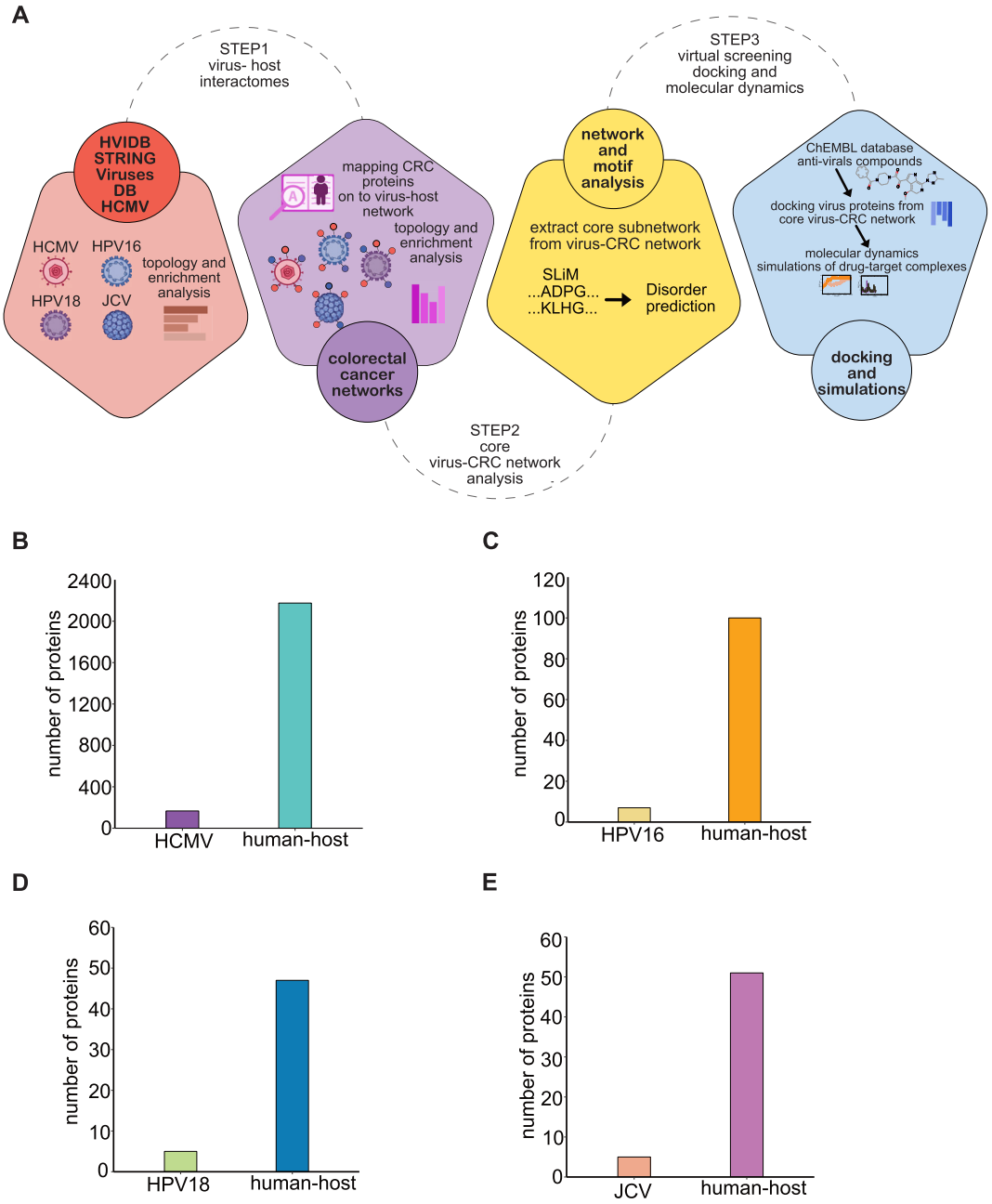

Fig. 1. Virus – host network analysis in CRC proteome. (A) The schematic representation of integrative data analysis. (B–E) This represents the HCMV, HPV16, HPV18 and JCV interactions with the human host.

ters; network heterogeneity and centralization of the virus-host networks to compare with random networks. Network heterogeneity (nh) measures the degree of network distribution and it shows that biological network tends to have central nodes [46]. The network centralization (nc) index measures the degree of dispersion of all node centrality scores in a network. This analysis revealed that virus-host networks have high scores of nh and nc affirming that biological networks are heterogeneous with hubs and are densely connected in comparison to random networks (Supplementary Fig. 2A–D).

Next, we performed pathway enrichment analysis for the host proteins corresponding to each virus network (Supplementary Fig. 3A–D). Most of the proteins from the HCMV network are associated with protein metabolism and localization. Viruses are known to utilize the host metabolic machinery to replicate in the cellular environment [47]. On the other hand, the proteins from HPV16 are involved in the G1/S phase transition of the cell cycle and cellular stress response. Similarly, most of the HPV18 network proteins are engaged in cell cycle G1/S phase transition and positively regulate the cell cycle process. Proteins from the JCV network are enriched in negative regulation of G1/S transition mitotic phase. It is shown that viruses interact with cell cycle machinery and enhance their replication [48]. Overall, we found that host proteins from all their corresponding virus networks are enriched in pathways associated with cell cycle regulation.

3.3. Colorectal cancer proteins are enriched in virus-host network modules

Once we had virus-host networks in hand, we aimed to deduce the virus's relationship with colorectal cancer. We extracted the CRC proteins and constructed a virus-host CRC network (Fig. 2A).

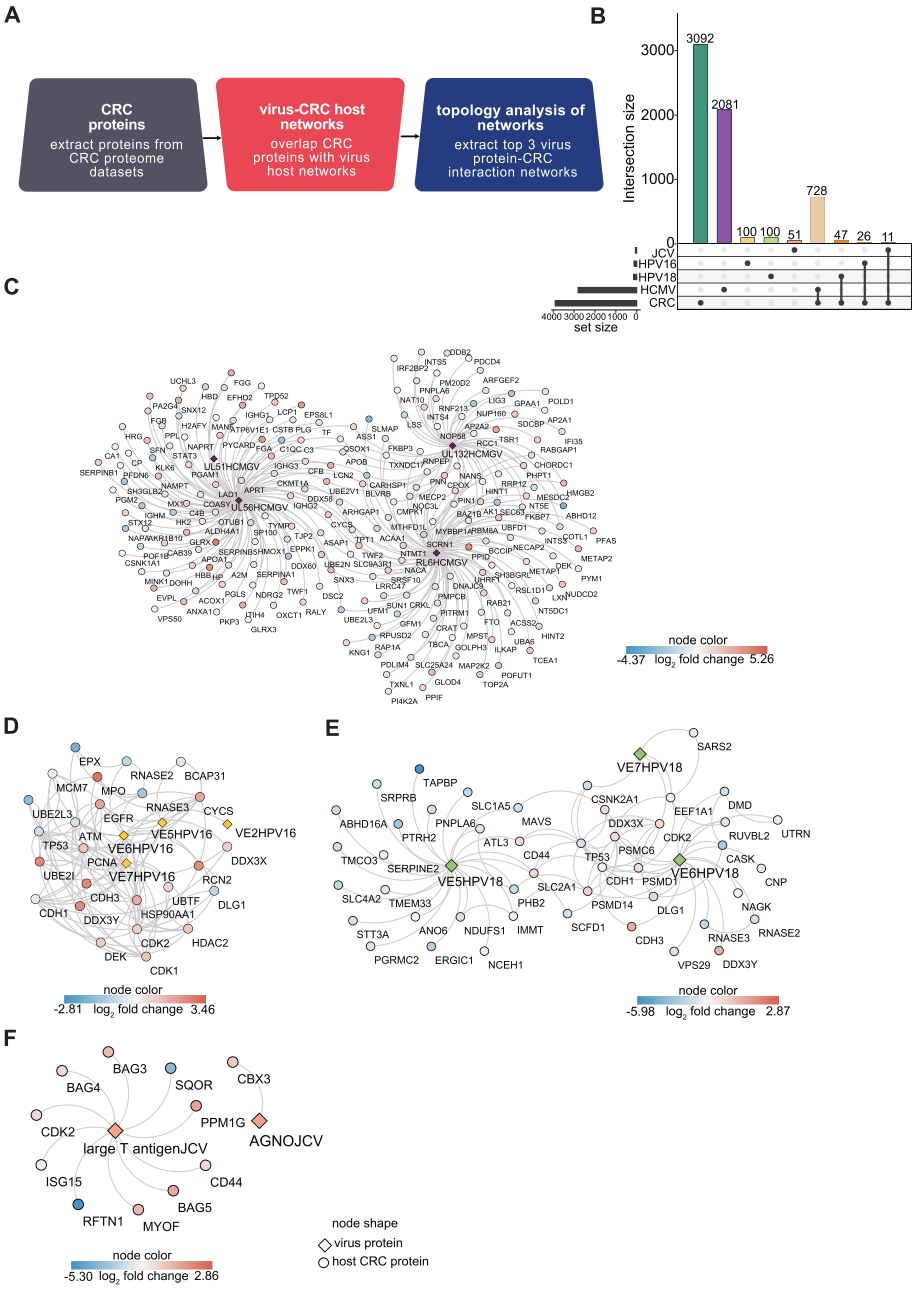

Fig. 2. Virus-host CRC proteome networks. (A) The outline of the workflow to generate the virus-CRC networks. (B) The UpSet plot displays the exclusive (single black dots) and shared (connecting black dots) virus-host and CRC proteins. (C–F) HCMV, HPV16, HPV18 and JCV networks represent the top three virus proteins connected with CRC proteins in the host. Note: Red and blue colors represent the up and down regulation of CRC proteins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Specifically, we mined 3092 CRC proteins from various literature sources (see methods) and mapped them onto the virus-host network of each virus type (HCMV, HPV16, HPV18, and JCV). We found that the HCMV network consists of 728 CRC proteins while the HPV16, HPV18, and JCV networks were included with 26, 47, and 11 (Fig. 2B); this enabled us to create individual virus-host specific CRC networks (Supplementary Fig. 4A–C). Our pathway enrichment analysis revealed that HCMV-CRC network regulates exocytosis, apoptotic signaling, and proteolysis pathways. The HPV16-CRC network is enriched in the G1/S transition of mitotic cell cycle, apoptosis, and DNA repair pathways. HPV18-CRC network participates in myeloid leukocyte mediated immunity, Wnt signaling, and cell cycle pathways. JCV-CRC network is engaged with cell cycle and catabolic processes (Supplementary Fig. 5A-D).

Next, we aimed to derive the key virus targets from each virus-host CRC network. To do this, we computed the degree i.e., number of interactions for each virus protein in the virus-CRC networks. It allowed us to extract the top 3 virus protein targets and their interactions in each network. The HCMV network comprises 247 interactions between 3 HCMV proteins (RL6, UL56, and UL132) and 202 CRC proteins (Fig. 2C). Similarly, the three VE7, VE6, and VE5 proteins of HPV16 & HPV18 made 95 and 82 interactions with 25 and 45 CRC proteins, respectively (Fig. 2D-E). In the JCV network, we were able to extract only two viral proteins instead of the top three due to the lack of abundant virus-host interaction data. We found 11 interactions between two JCV proteins (Large T antigen and Agnoprotein) (Fig. 2F). Altogether, these results indicated that the proteins involved in CRC are enriched among these virus-host networks.

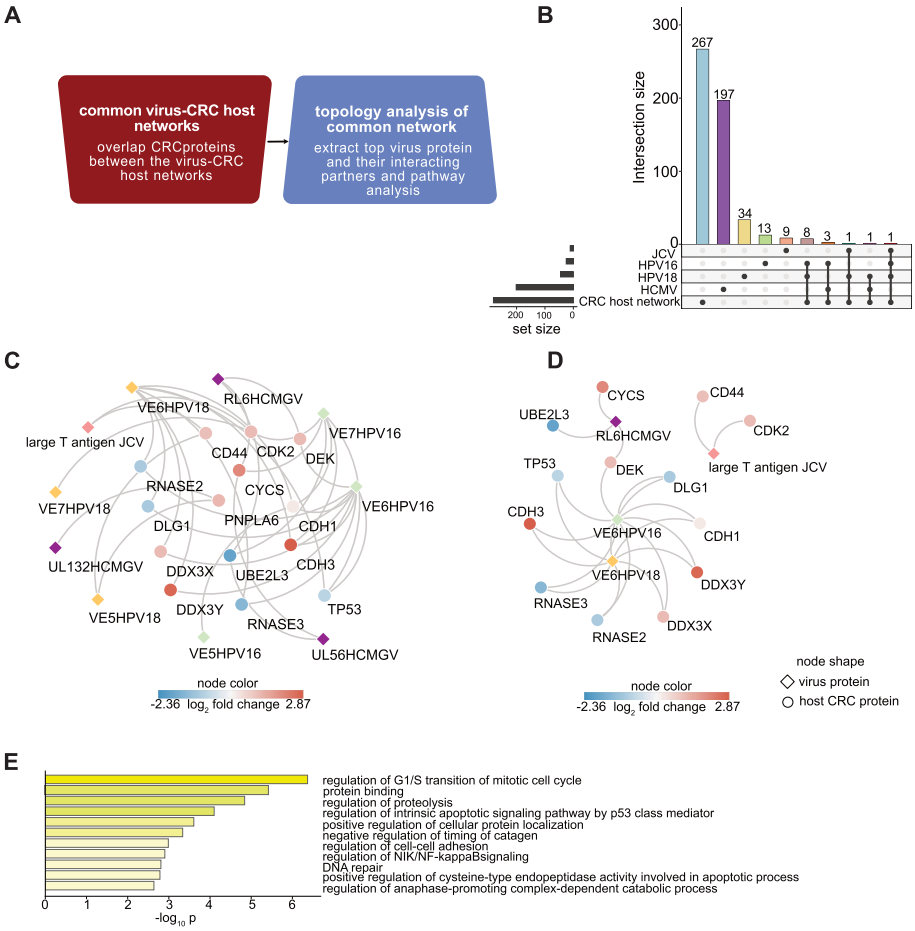

Fig. 3. Core virus-host CRC proteins between the virus networks. (A) Schematic representation of the workflow to spawn the common CRC proteins across the virus networks. (B) The common and unique CRC proteins across the virus-host CRC networks. (C) Common CRC proteins between the virus-host CRC networks. (D) Network shows the top virus proteins and their interactions with core virus-host CRC network. (E) Statistically significant biological processes enrichment of the core virus-host CRC network. Note: Red and blue colors represent the up and down regulation of CRC proteins. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Viruses modulates core colorectal cancer protein network

Our CRC network analysis illuminated the virus's connections to cancer-specific proteins in the host, as discussed above. Next, we sought to identify the network signature from virus-host CRC networks. So, we integrated all four virus-host CRC networks and overlapped them to isolate the CRC proteins which are present in more than one network (Fig. 3A). It identified 14 CRC proteins in total (Fig. 3B), among which 8 (CDH3, DLG1, TP53, RNASE3, DDX3Y, CDH1, RNASE2, DDX3X) were present in both HPV16 and HPV18 networks, while we found 3 proteins (CYCS, DEK, UBE2L3) to be common in HPV16 and HCMV networks. PNPLA6 was identified as the common protein between HCMV and HPV18 networks. Similarly, CD44 serves as the mutual one between HPV18 and JCV networks. CDK2 is the only protein found to be associated with HPV16, HPV18, and JCV networks (Fig. 3C).

With this CRC network, we aimed to extract a core virus-host CRC network. To do this, we took advantage of topology analysis and derived top one protein from each virus network based on their degree; this divulged RL6_HCMV, VE6_HPV16, VE6_HPV18, and large T antigen proteins (Fig. 3D). Functional enrichment analysis revealed that the CRC protein interactors of RL6_ HCMV (CYCS-up, DEK- up, UBE2L3 - down) are involved in the G1/S transition of mitotic cell cycle and apoptosis. The DEK (up), CDH3(up), CDH1 (up), DDC3Y(up), DDC3X(up), TP53(down), DLG1(down), RNASE2(-down), and RNASE3(down) protein interactors of VE6_HPV16 and VE6_HPV18 were associated with G1/S transition of the mitotic cell cycle, proteolysis and apoptosis respectively. The interacting proteins CD44 (up), CDK2 (up) of large T antigen_JCV were found to regulate the G1/S transition of the mitotic cell cycle (Fig. 3E, Supplementary Table S1). Overall, our analysis revealed the core cell growth-related virus-host CRC network.

3.5. Sequence analysis of viral proteins from common colorectal cancer protein network

As discussed above, our layered network analysis identified the core virus-CRC network. It identified four proteins RL6_HCMV, VE6_HPV16, VE6_HPV18, and large T antigen_JCV. Viruses tend to mimic the Short Linear Motifs (SLiMs) of host proteins to exploit the various cellular processes for their advantage. This comprehensive mimicking renders them as possible therapeutic targets [49]. Therefore, we identified the probable SLiM regions for the top protein of each virus using the Eukaryotic Linear Motif (ELM) database. Considering the value of the probability score as less than 0.01, we identified 17 classes of ELMs for RL6_HCMV, 25 for VE6_HPV16, 32 for VE6_HPV18, and 46 for large T antigen_JCV.

The intrinsically disordered regions (IDRs) provide functional plasticity to the viral proteins actively involved in the replication and immune evasion of host cells [19]. Since these unstructured regions have a substantial importance in drug discovery, we predicted the IDRs of each viral protein using Disprot, a manually curated database of intrinsically disordered regions from the literature. We identified that various regions of viral proteins were found to be disordered. As shown in Supplementary Table S2, sequence range of top protein from each virus with disorderliness and its corresponding SLiM is represented. It also shows the its corresponding ELM class and probability score of each SLiM. We merged the residues from SLiM and IDRs. Altogether, these results allowed us to choose the SLiM region of viral protein having the highest disorder score.

3.6. Virtual screening identified top antivirals against the viral proteins

The SLiMs in the pathogen protein structures are beneficial for therapeutic purposes [50]. Our in-depth sequence analysis of the four viral targets (RL6_HCMV, VE6_HPV16, VE6_HPV18, and large T antigen_JCV) identified statistically significant SLiMs and IDRs. To trigger the structure-based virtual screening of antiviral compounds (Fig. 4A), we extracted the experimentally derived protein structures of VE6_HPV16 and VE6_HPV18 from the protein data bank. However, we did not find the structures of RL6_HCMV and Large T antigen_JCV; thus, we modelled the 3-dimensional protein structures using trRosetta (Supplementary Fig. S6A-B). The model for RL6_HCMV was predicted using the de novo folding approach by rosetta with a very high template modeling (TM) score of

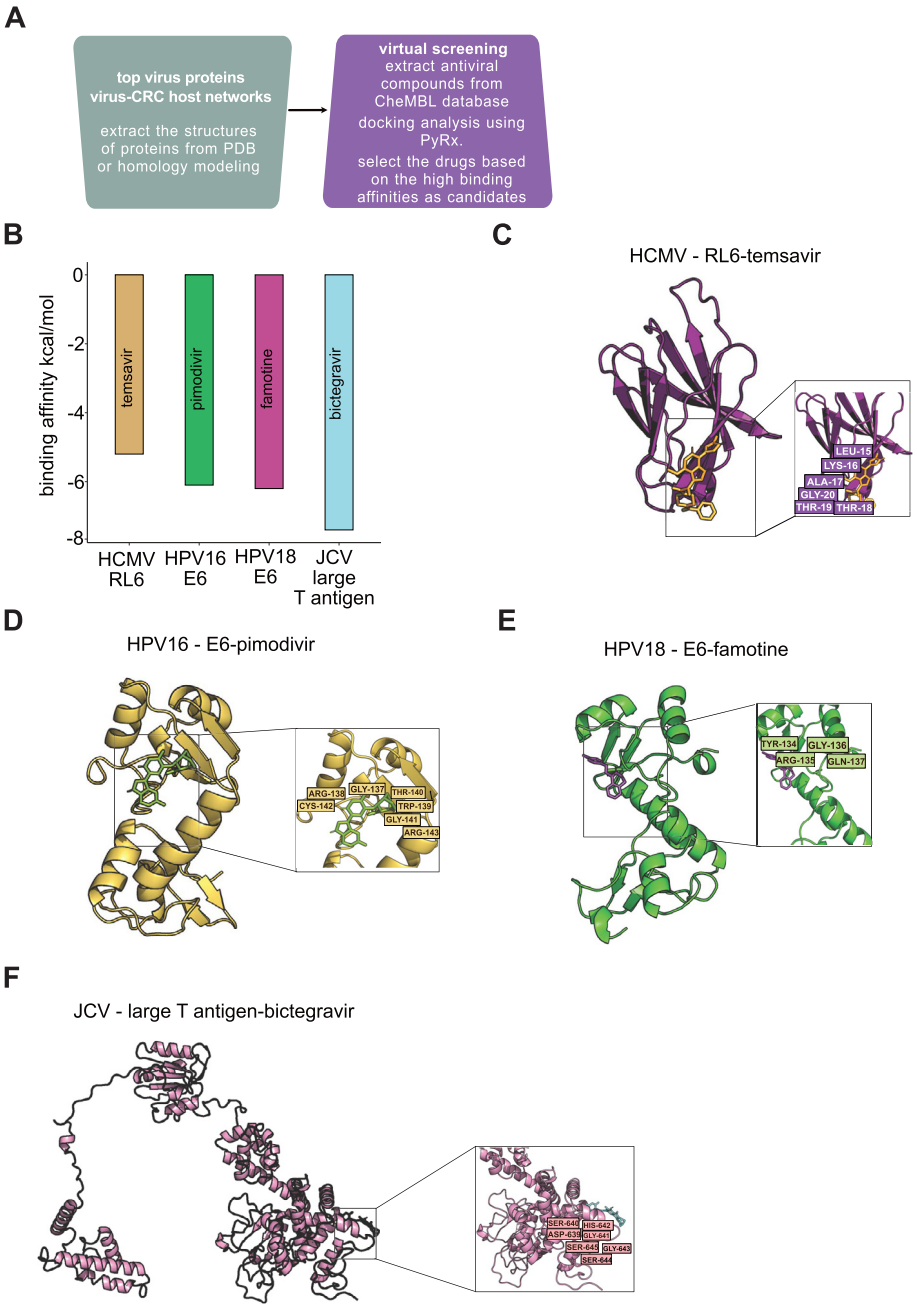

Fig. 4. Virtual screening of the antivirals and molecular docking. (A) The scheme of workflow to identify the drug candidates against the virus proteins. (B) The binding affinities of top drug molecules against the virus targets. (C–F) Illustration of binding modes temsavir, pimodivir, famotine, and bictegravir with their respective viral target proteins.

0.758. The large T antigen_JCV was modelled based on homologous templates, which revealed a high TM score of 0.610. We subjected these computational structural models to ProSA validation analysis to compare them with existing X-ray or NMR 3D structures. This analysis revealed that RL6_HCMV and Large T antigen_JCV models are in the range of NMR (z-score: −4.47) and X-ray (z-score: −11.96) based structures respectively (Supplementary Fig. S6C-D). Then we performed the Ramachandran plot analysis to check the rotations of backbone bonds in the protein residues. It revealed that 99.1% of residues are in allowed region out of which 94.5% are in favorable region of RL6_HCMV protein structure (Supplementary Fig. S6E). We observed that the allowed region consists of 97.5% of the residues in which 91.1% is the favorable region in the large T antigen_JCV (Supplementary Fig. S6F). Lately, the Rama-Z score is proposed to validate the distributions in Ramachandran plots [34]. If Rama-Z score >3, it indicates the improbable geometry of the backbone, whereas, Rama-Z score < 2 suggests favorable backbone geometry. Thus, we computed this score for the protein structure models. It revealed a low Rama Z-score of 1.28 for RL6_HCMV and very low for large T antigen_JCV (Rama Z-score −0.56). Therefore, it is confirmed that both structural models are geometrically reliable.

Once we had the structural data in hand, we sought to screen the antiviral compounds from the CheMBL database (see methods) against the virus target proteins. To do this, we utilized the AutoDock Vina algorithm, a part of PyRx virtual screening software. As discussed above, SLiM regions could be targeted therapeutically thus we assessed the druggability of each SLiM region. We observed 0.9 ± 0.02 (RL6_HCMV), 0.88 ± 0.05 (VE6_HPV16), 0.94 ± 0.03 (VE6_HPV18) and 0.93 ± 0.03 (Large T antigen_JCV) this analysis revealed high probabilities close to 1. It confirmed that these SLiM sites can be druggable and used for docking.

Further, we performed a functional SLiM site-specific docking assay with antiviral compounds against the viral targets. Then we ranked antiviral compounds from highest to lowest binding affinities. We took the compounds which showed the highest binding affinities for further analysis. It identified the temsavir (RL6_HCMV), pimodivir (VE6_HPV16), famotine (VE6_HPV18), and bictegravir (Large T antigen_JCV). This analysis showed that temsavir, pimodivir, famotine, and bictegravir binds to their targets with −5.2 kcal/mol, −6.1 kcal/mol, −6.2 kcal/mol, and −6.4 kcal/mol binding affinities (Fig. 4B). Our docking analysis showed that temsavir binds to the RL6_HCMV SLiM site (Fig. 4C), and we observed that LYS16 forms pi-Alkyl bonds whereas THR18 forms van der Waals interactions with the compound (Supplementary Fig. 7A). Temsavir is known to target GP120_HIV protein and this interaction foils the binding between the virus cellular CD4 receptors which in-turn prevents the viral entry to host cell [51,52]. It is reported that temsavir interacts with following residues ILE108, ILE109, ASP113, ASN425, MET426, TRP427, GLN428, GLN432, ALA433, VAL255, SER256, THR257, GLU370, SER375 in GP120_HIV [53]. Given this premise of temsavir we compared the docking results of this compound with RL6_HCMV. In order to do this, we followed similar procedure as RL6_HCMV as we focused on SLiM and disorder region in the docking, we predicted those regions in GP120_HIV protein structure (Supplementary Table S3). The SLiM binding site (druggability score: 1) specific docking of temsavir and GP120_HIV showed −7.2 kcal/mol binding affinity (Supplementary Fig. 8A). Further analysis of temsavir binding to GP120_HIV showed it binds to SLiM binding site and ASN156, ASN160 forms van der Waals interactions with the compound (Supplementary Fig. S8B-C). The interaction analysis of pimodivir revealed that it binds to the VE6_HPV16 SLiM site (Fig. 4D), and ARG142 form hydrogen bonds with the pimodivir (Supplementary Fig. 7B). A close look at the docked pose of famotine with VE6_HPV18 SLiM site (Fig. 4E) and interaction analysis showed that the compound forms a hydrogen bond with TYR134 (Supplementary Fig. 7C). Bictegravir binds to Large T antigen_JCV (Fig. 4F) in which SER640 forms hydrogen bonds, whereas GLY641, HIS642 forms van der Waals interactions with the compound (Supplementary Fig. 7D). Integrase_HIV is crucial for viral DNA integration into host DNA. This process is vital for replication of virus and bictegravir is a well-known inhibitor which interacts with ASN117 and GLY118 residues of Integrase_HIV [54] which inhibits virus integration in host cells [55,56]. In order to compare the docking results of JCV we predicted the SLiM site in Integrase_HIV protein (Supplementary Table S3). The druggability analysis revealed probability score as 0.95 ± 0.02. SLiM site-specific and disorder region docking assay revealed bictegravir and Integrase docking revealed −8.2 kcal/mol binding affinity (Supplementary Fig. 8A). Further analysis of docked complex revealed that Bictegravir binds to SLiM site of Integrase_HIV and ASN144, SER147 forms van der Waals interactions (Supplementary Fig. 8D-E). In sum, our virtual screening revealed antiviral compounds that bind to the viral targets.

3.7. Molecular dynamics simulation demonstrates the effect of antivirals on virus proteins

The docking analysis gave us the best drug molecules against the proteins from each virus type. But docking revealed only static poses of binding pockets of drugs within the protein structure. The dynamics of stability are essential for the biological function of the proteins [57]. Thus, we performed the molecular dynamics (MD) simulations to understand the dynamics of viral target-drug complexes from the docking analysis. We ran 100 ns simulations for apo form (protein only) and bound form (drug-target complexes). Then, we analysed MD trajectories with various parameters like Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), hydrogen bonds to apprehend the conformational changes and stability of complexes. Along with these, we also computed the binding free energies of protein-drug complexes using the MM-PBSA method (Fig. 5A).

First, we analyzed the stability of drug-protein complexes using the RMSD. The RMSD of RL6_HCMV and temasvir complex rise to 0.4 nm at around 20 ns and remained stable throughout the simulation. Apo form (only protein) of RL6 showed an initial rise to 0.3 nm, and it was steady until 75 ns, but later it rose again till the end of the simulation (Fig. 5B). The pimodivir - VE6_HPV16 complex increased to 0.51 nm at 21 ns and further rose to 1.2 nm at 45 ns while it remained stable throughout the simulation. In contrast, the apo form of VE6 had an initial rise at 20 ns to 0.5 nm and a further increase at 60 ns to 0.8–1 nm, where it remained stable till the end (Fig. 5C). The VE6_HPV18 - famotine complex started at 1.1 ns and remained stable till 65 ns but gradually dropped at the simulation's end. The apo form of VE6_HPV18 showed very low RMSD while maintaining stability throughout the simulation (Fig. 5D). The RMSD of the JCV large T antigen- bictegravir complex rose to 2.1 nm at 18 ns and remained stable with slight fluctuations through the simulation. The apo form of JCV showed low RMSD compared to a drug-protein complex (Fig. 5E). We observed low RMSD 0.1–0.3 nm of ligands (drugs) along the MD trajectory; this suggested that they were stable during the simulation (Supplementary Fig. 9A–C). We also performed MD on GP120_HIV complexed with temsavir, RMSD of the complex rose to 1.2 nm at 30 ns and remained steady throughout the simulation and apo form of GP120_HIV divulged an initial rise to 0.9 nm at 30 ns and declined afterwards (Supplementary Fig. 10A). We observed low RMSD of temsavir ranging from 0.2 to 0.3 nm along the MD simulation suggesting the stability of the drug molecule (Supplementary Fig. 10B). The RMSD of Integrase_HIV and bictegravir complex shows initial rise to 0.25 nm and steadily declined

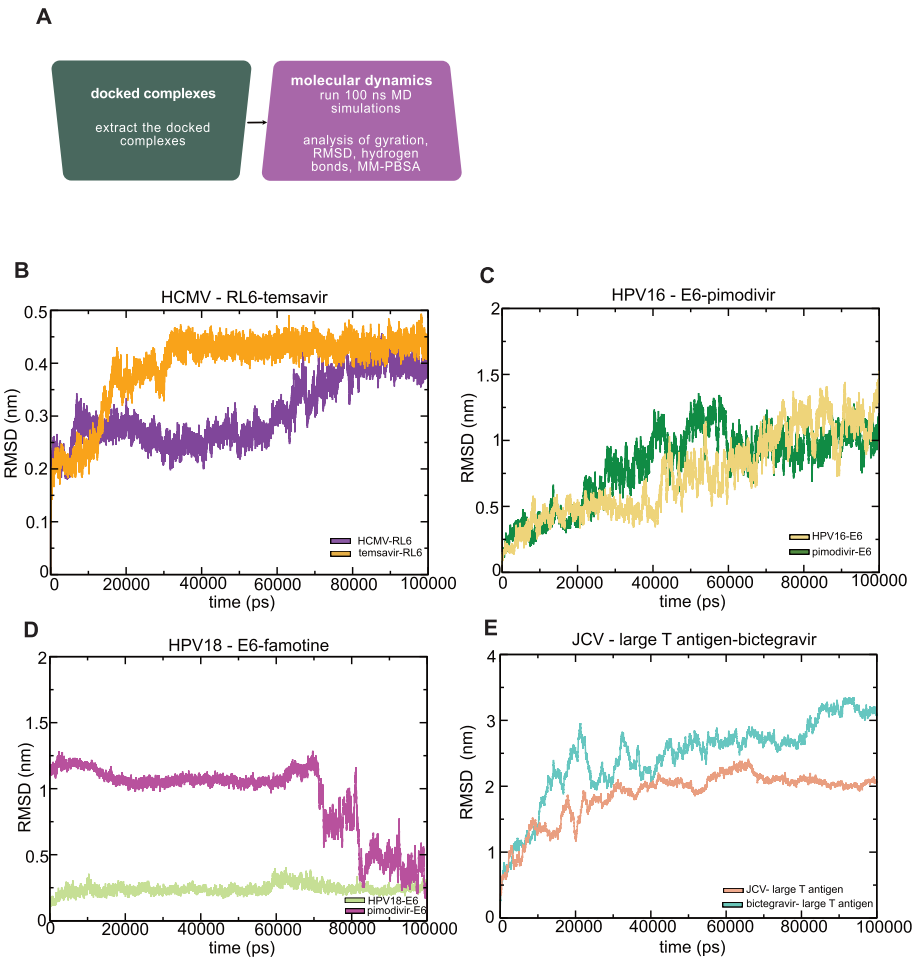

Fig. 5. Molecular dynamics simulation analysis of drug-virus target complexes. (A) Outline of molecular dynamics simulations. (B–E) Root mean square deviation (RMSD) profiles of the drug molecules after binding to the virus targets over 100 ns.

from 20 ns and rose again at 65 ns and remained high when compared to apo form of Integrase_HIV (Supplementary Fig. 11A). The bictegravir drug RMSD remained low (0.05–0.25 nm) (Supplementary Fig. 11B). It revealed that the temsavir and bictergravir effects HIV proteins structures and comparable to the RL6_HCMV and Large T antigen_JCV MD trajectories. Altogether, the RMSD analysis revealed that the drugs impact virus protein backbone structure.

3.8. Residue level analysis revealed that drug related changes in SLiM regions

We analyzed the RMSF profile of protein-drug complexes to understand the effect of the drugs on aminoacids. The RMSF of RL6_HCMV revealed that temsavir induced significant fluctuations at the SLiM region (Fig. 6A). The RMSF assessment of GP120_HIV divulged that temsavir produced noteworthy fluctuations at the SLiM region (Supplementary Fig. 10C-D). The RMSF calculations of pimodivir revealed that it caused changes in the SLiM area of VE6_HPV16. (Fig. 6B). The RMSF profile of VE6_HPV18 - famotine showed that the SLiM binding region is significantly affected by the drug (Fig. 6C). The RMSF analysis of large T antigen_JCV - bictegravir complex divulged the drug-induced compelling variations in the SLiM site (Fig. 6D). The RMSF assessment of Integrase_HIV divulged that bictegravir created notable fluctuations at the SLiM region (Supplementary Fig. 11C-D). Overall, the detailed residue level analysis suggested that drugs impact the residues related to SLiM region.

3.9. Hydrogen bond analysis showed interactions of drugs with viral targets

Hydrogen bonds (hbonds) are vital molecular interactions, and they guide the protein-drug interactions. The MD simulations provide a variety of conformations and they can be used to understand the interaction patterns of hydrogen bonds between the protein and drug. We computed the number of hydrogen bonds formed throughout the 100 ns MD simulation.

The complex of RL6_HCMV-temsavir divulged that it formed four hydrogen bonds on an average during the simulation that remained stable until 90 ns, and formed 5 to 6 hydrogen bonds (Fig. 7A). Temsavir form 3 to 6 hydrogen bonds with GP120_HIV and remained stable throughout simulation (Supplementary Fig. 10E). The VE6_HPV16 - pimodivir complex formed three hydrogen bonds until the end of the simulation (Fig. 7B). On average, Famotine formed one hydrogen bond during the simulation with minor fluctuations and two hydrogen bonds (Fig. 7C). Bictegravir interacted with JCV large T antigen to form 2 hydrogen bonds on an average until the end of the simulation. Whereas, we observed 3–4 hydrogen bonds both at the beginning (20–40 ns) and end of simulations (80–100 ns) (Fig. 7D). We observed that bictegravir forms 2 to 3 hydrogen bonds with the Inte-

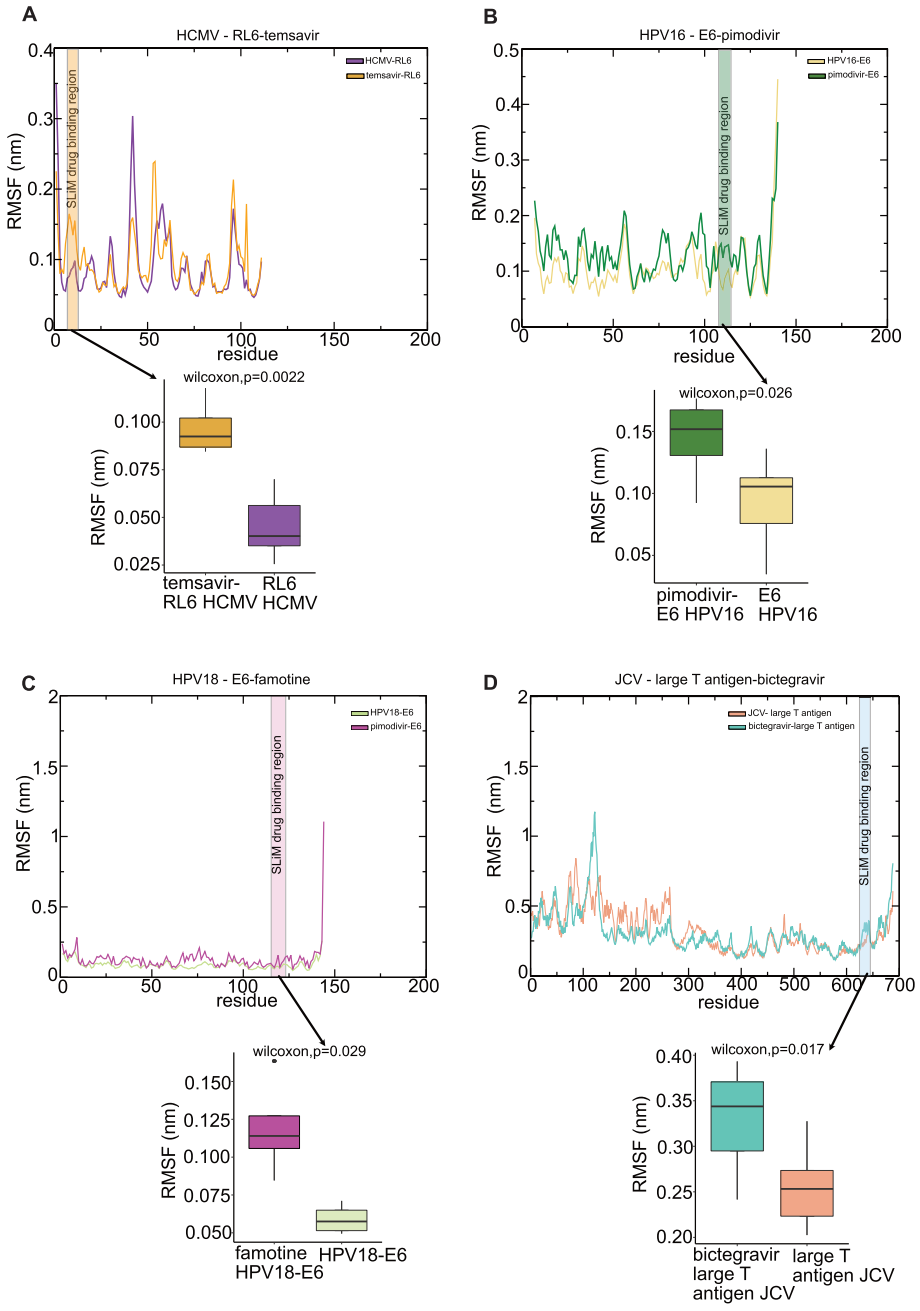

Fig. 6. Root mean square fluctuation analysis. (A–D) The line plots represent the root mean square fluctuation profiles of drug molecules with their virus targets. Boxplots shows the distribution of RMSF with significance of the differences between the drug binding SLiM region in the complex and apo forms. Significance ($p \leq 0.05$) was calculated using the Wilcoxon-test. The boxes indicate the 25th percentile, median and 75th percentile.

grase_HIV and stayed stable until the end of simulation (Supplementary Fig. 11E). In sum, this analysis revealed that stable hydrogen bonds formed between the drug and protein complexes.

3.10. Analysis of MM-PBSA free binding energies of drugs with virus targets

The virus protein-drug complexes were analyzed to estimate the free energies of binding using the molecular mechanics-Poisson-Boltzmann surface area method (MM-PBSA). The MM-PBSA provides factual estimates of free energies of drug binding to proteins [38]. We computed the following energy terms: van der Waals energy, electrostatic energy, polar solvation energy, Solvent Accessible Surface Area (SASA), and binding energy from 50 to 100 ns. Except for the polar solvation energy, all forms of energy are favorably contributed to the interactions between the drugs and viral targets. The complex of RL6_HCMV - temsavir showed −28.269 kJ/mol (Fig. 8A) binding energy. The temsavir bound to GP120_HIV with −102.179 kJ/mol binding energy (Supplementary Fig. 10F). The pimodivir is bound to VE6_HPV16 with binding energy of −44.410 kJ/mol (Fig. 8B). Famotine complexed with VE6_HPV18 with the binding energy of −55.237 kJ/mol (Fig. 8C). Finally, we observed that bictegravir was constrained with the large T antigen_JCV and has a binding energy of −101.559 kJ/mol (Fig. 8D). The Integrase_HIV-bictegravir complex revealed −95.092 kJ/mol binding energy (Supplementary Fig. 11F). Taken together, binding free energy analyses showed that drugs are bound to the viral proteins with good binding energies.

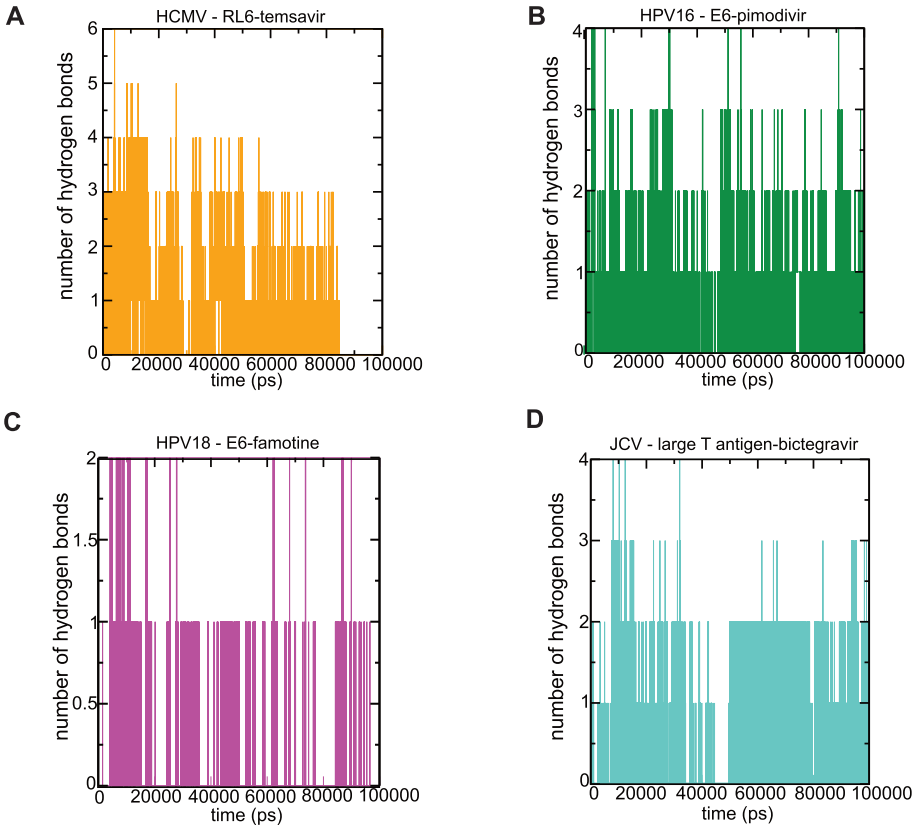

Fig. 7. Hydrogen bond formation of drugs bound form complexes. (A-D) Number of hbonds formed between the virus protein-drug complexes throughout the 100 ns simulation period.

4. Discussion

Virus infections trigger about 12% of human cancers. They hijack the host cellular pathways like cell cycle, metabolism, angiogenesis, cell proliferation to advance their replication and survival in the host cells [58]. Thus, it is imperative to understand the interplay between the viruses and host cellular machinery to get insight into the mechanism employed by viruses to initiate cancer pathobiology. Recent studies suggested that HCMV, HPV16, HPV18, and JCV play a key role in colorectal cancer pathology [5–7]. In the current study, we took advantage of publicly available virus-host interactome datasets related to these viruses and constructed virus-specific networks (Supplementary Fig. 1A–D). The pathway enrichment analysis revealed that these viruses affect the host's cell cycle and various signaling pathways. Furthermore, topology analyses showed that these networks are biologically important when compared to random networks (Supplementary Fig. 2A–D).

We extracted differentially expressed proteins in colorectal cancer patients from the published studies [24–26]. We mapped them to the virus-host networks to gain further insight into the links between colorectal tumorigenesis and viruses. This allowed us to derive the virus specific-host CRC networks; the functional enrichment analysis showed that viruses regulate apoptosis, cell cycle, and signaling processes (Supplementary Fig. 5A–D). Next, we aimed to identify the key virus proteins contributing to cancer pathobiology for each of the four viruses individually. We considered the top 3 virus proteins based on their number of interactions for further analysis. In HCMV we found that, RL6, UL56, and UL132 have high highest number of interactions with host CRC proteins. We identified VE7, VE6, and VE5 in HPV16 and HPV18. The JCV network analysis highlighted two top virus proteins, Large T antigen and agnoprotein (Fig. 2C–F).

We further analysed each virus-host CRC subnetworks, and integrated all these four types of viral networks to explore core CRC network signature between the viruses (Fig. 3C). Through this analysis, we first identified the shared CRC proteins between the viruses' networks. Additionally, we refined this network based on the topology parameter and identified one top virus protein connected with CRC proteins commonly present in each virus type except JCV, as it did not share common CRC proteins with other virus types (Fig. 3D). Specifically, we identified the following viral proteins in the core virus-host CRC network: RL6 contain RL11D domain and belongs to RL11 gene family of HCMV. RL11 family genes are believed to code for putative transmembrane glycoproteins [59]. Though it is believed that RL11 family genes are dispensable [60], recently Katie et al showed that RL1, an RL11 family protein, targets an endonuclease SLFN11 by ubiquitin mediated degradation in order to benefit viral replication [61]. Similar studies show that other RL11 family proteins such as UL4, RL11 encode envelope glycoprotein and IgG-Fc binding glycoprotein respectively [62]. Therefore, we hypothesize that RL6 protein might encode glycoprotein with crucial importance in promoting viral propagation.

VE6 in HPV16 and HPV18 promotes cell cycle deregulation leading to malignancy [63]. The large T antigen supports viral replication and carcinogenesis [64,65]. Functional enrichment analysis of this refined network showed that it is enriched in the cell cycle, apoptosis, and signaling pathways (Supplementary Table S1). All these analyses affirmed that viruses HCMV, HPV16, HPV18, and JCV are linked to tumorigenesis in CRC phenotype.

We identified that many of the CRC proteins in the core network have various functions in common. They are involved in cancer-related pathways such as cell cycle regulation, apoptotic signaling, and cell–cell adhesion regulation. Interestingly, our analysis

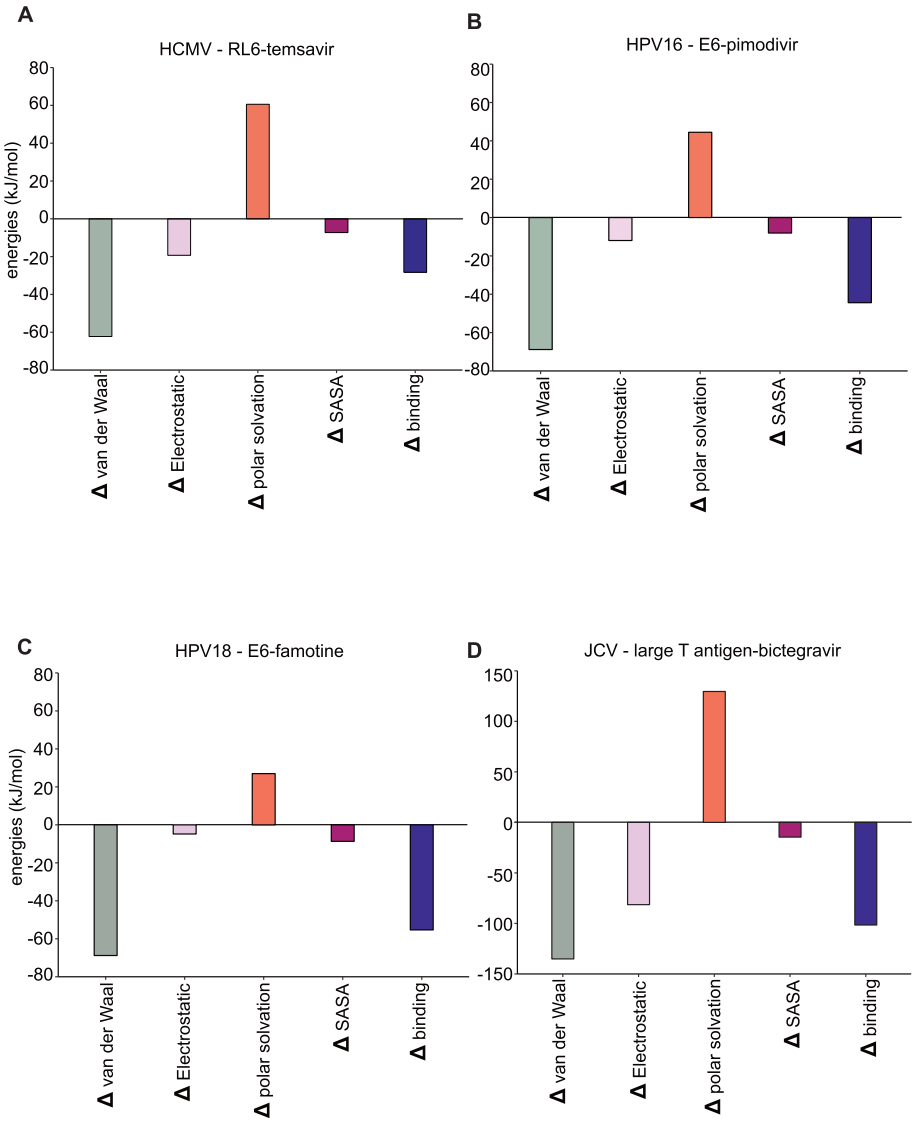

Fig. 8. Binding energy calculations of drug complexes. (A–D) The free energy terms acquired from MM-PBSA calculations of four drugs temsavir, pimodivir, famotine and bictegravir complexed with HCMV- RL6, HPV16-E6, HPV18-E6 and JCV- large T antigen.

yielded proteins associated with the regulation of proteolysis, which is an essential contributor in the process of invasion and metastasis.

The interactors of JCV large *T*-antigen, the CD44, and CDK2 are known to play a crucial role in CRC tumorigenesis. CD44 is known to activate cell signaling pathways to induce cell proliferation, increase cell survival and enhance cellular motility [66]. Its association with MMPs, plays a role in proteolytic cleavage of collagen IV, thereby promoting cell invasion [67]. Moreover, Geol et al. observed that CD44 is significantly altered following JCV *T*-Ag expression [68]. CDK2 is one of the cyclin-dependent kinases that directly mediate the G1/S phase transition of the cell cycle, and its overexpression in primary CRC tumors is linked to lymph nodes metastasis [69].

The proteins interacting with the VE6 protein of HPV16 and HPV18 are associated with various signaling processes. The VE6 protein of these two viruses interacts with DDX3X and its paralog DDX3Y, that belongs to the large DEAD-box protein family which promotes viral replication [70]. Over-expression of DDX3X closely correlated with nuclear β-catenin expression, that activates the Wnt signaling pathway, thereby promoting cancer cell proliferation in CRC[71]. Cadherin related proteins CDH1 and CDH3 interacts with both HPV16 and HPV18 VE6 proteins. Cadherins are significant contributors to cell–cell adhesion in epithelial tissues. The expression of CDH1 (E-cadherin) and CDH3 (P-cadherin) is upregulated in CRC. It supports the notion that their expression plays a role in cell adhesion in CRC [72,73]. VE6 protein from both HPV16 and HPV18 viruses also interacts with TP53 which is down regulated in CRC and this revealed DNA repair and apoptosis pathways are suppressed in CRC. The p53 tumor suppressor is the first described and best-known target of VE6. Since VE6 binds to p53, causing ubiquitination, hence preventing growth arrest or apoptosis of infected cells [74]. VE6_HPV16 interacts with DLG1 and it shown that it is targeted by human viral oncoproteins such as VE6 of HPV, suggesting the potential association [75].

We identified that an oncogene, DEK, interacts with the VE6 protein of HPV16 and RL6 of HCMV. Interestingly, DEK is known to inhibit senescence and apoptosis via the destabilization of p53 [76]. The interactors of RL6_HCMV proteins are UBE2L3 and CYCS. UBE2L3 is known to promote cell growth in non-small cell lung cancer (NSCLC) [77]. Also, its interaction with GSK3β stimulates the development of certain tumors including colorectal cancer

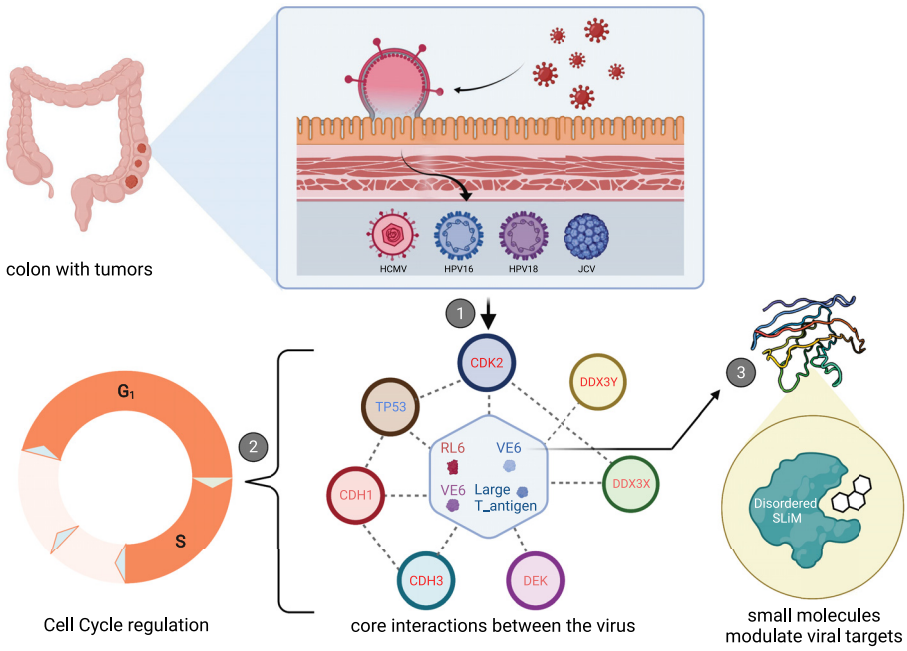

Fig. 9. Schematic representation of core network. The viruses-host CRC core network regulating cell cycle pathway and small molecules modulating the virus protein targets. This diagram is created with BioRender.com.

[78]. Similarly, CYCS known to to regulate tumor growth in renal cell carcinoma [79].

As discussed above, it is clear that RL6_HCMV, VE6_HPV16, VE6_HPV18 and Large T antigen_JCV have clear interactions with the tumor growth related proteins indicating that that the viral proteins have functional impact in cancer phenotype. Thus, it is imperative to identify antivirals against the virus proteins. Viruses use a variety of mechanisms to take over the host cells. Among them protein mimicry and SLiMs which facilitate the viruses to interact with host proteins and hijack molecular machinery for proliferation in the host cells [80]. The disorder property in virus sequences emerged as a key feature of pathogenicity [81,82] and SLiMs could be exploited for therapeutic purposes [49]. So, we predicted the SLiMs and disorder regions in each virus target from the refined core CRC network signature. Then, we merged both the regions to extract binding sites having disorderliness and with at least a class of SLiM. In our study, we aimed to identify the drug molecules against the viral proteins which could target SLiMs. We used a structure-based virtual screening approach. In brief, we used antiviral drug molecules from the ChEMBL database and filtered them based on the Lipinski Rule of five (RO5) [38]. The molecular docking assay yielded one antiviral compound for each virus targets. We found temsavir targets RL6 in HCMV; and it is known to block HIV entry [83]. Pimodivir binds to VE6 in HPV16 and is known to reduce the virus influenza replication [84]. Famotine targets VE6_HPV18; it is noted for antiviral activity [85]. Bictegravir binds to Large T antigen in JCV. This drug molecule known to block replication of HIV [56].

We further performed the molecular dynamics simulations to support the docking results. In the RMSD profile of antiviral drug target complexes, we observed that all compounds had induced changes in protein backbone structures in the viral targets. It is corroborated with RMSF analysis, highlighting that antiviral significantly affected the SLiM sites in the viral targets. The binding free energies of MD trajectories confirmed that the antivirals identified in our study are strong binders to the viral targets. The temsavir and bictegravir revealed comparable effect on GP120_HIV and Integrase_HIV in docking and MD assays. We believe that inhibitors identified in this study against the viral targets RL6_HCMV, VE6_HPV16 and HPV18, Large T antigen_JCV will have functional impact on tumor promoting proteins.

In summary, we combined multiple virus-host network datasets with CRC proteome and identified core virus-host CRC network signature. It revealed that HCMV, HPV16, HPV18 and JCV commonly regulate cell cycle related proteins in CRC. Structure-based drug screening identified vital antiviral compounds (Fig. 9), which could pave a path for detailed experimentation to treat virus-driven colorectal cancer in future.

5. Conclusion

The hypothesis that viral infections have a role in CRC development and progression is biologically plausible and potentially relevant for disease amelioration. In this study, virus-host networks unveiled the relationship between HCMV, HPV16, HPV18, JCV, and human proteins. Our network analysis identified the core CRC network between the viruses, which enabled us to identify the top viral proteins and their biologically relevant pathways. Through this analysis, we could deduce the possible relationship between the viral interacting proteins and their role in colorectal carcinogenesis. Our understanding of the motif mimicry mediated protein interactions provided us with a strategy that can be used for developing antiviral molecules that target such interactions. Finally, our structure-based virtual screening analysis identified promising antiviral compounds to demolish the virulence proteins involved in human CRC. Overall, our findings open new intriguing research options for the future, leading to a better comprehension of viral oncogenesis and more efficient CRC treatments strategies.

CRediT authorship contribution statement

Sai Krishna A.V.S: Software, Investigation, Resources, Writing – original draft, Validation, Formal analysis, Data curation, Visualization. **Swati Sinha:** Project administration, Writing - review & editing. **Sainitin Donakonda:** Conceptualization, Methodology, Supervision, Software, Investigation, Resources, Writing – original draft, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.040.

References

[1] Sung H et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209–49.
[2] Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348(10):919–32.
[3] Zapatka M et al. The landscape of viral associations in human cancers. Nat Genet 2020;52(3):320–30.
[4] Chen H et al. Viral infections and colorectal cancer: a systematic review of epidemiological studies. Int J Cancer 2015;137(1):12–24.
[5] Harkins L et al. Specific localisation of human cytomegalovirus nucleic acids and proteins in human colorectal cancer. Lancet 2002;360(9345):1557–63.
[6] Ibragimova MK, Tsyganov MM, Litviakov NV. Human papillomavirus and colorectal cancer. Med Oncol 2018;35(11):140.
[7] Laghi L et al. JC virus DNA is present in the mucosa of the human colon and in colorectal cancers. Proc Natl Acad Sci U S A 1999;96(13):7484–9.
[8] Fulop T, Larbi A, Pawelec G. Human T cell aging and the impact of persistent viral infections. Front Immunol 2013;4:271.
[9] Chen HP et al. Human cytomegalovirus preferentially infects the neoplastic epithelium of colorectal cancer: a quantitative and histological analysis. J Clin Virol 2012;54(3):240–4.
[10] Bai B et al. Human cytomegalovirus infection and colorectal cancer risk: a meta-analysis. Oncotarget 2016;7(47):76735–42.
[11] Teo WH et al. Human cytomegalovirus infection enhances cell proliferation, migration and upregulation of EMT markers in colorectal cancer-derived stem cell-like cells. Int J Oncol 2017;51(5):1415–26.
[12] Damin DC, Ziegelmann PK, Damin AP. Human papillomavirus infection and colorectal cancer risk: a meta-analysis. Colorectal Dis 2013;15(8):e420–8.
[13] Ghabreau L et al. High-risk human papillomavirus infections in colorectal cancer in the Syrian population and their association with Fascin, Id-1 and P-cadherin expressions: A tissue microarray study. Clin Cancer Invest J 2012;1(1):26.
[14] Salepci T et al. Detection of human papillomavirus DNA by polymerase chain reaction and southern blot hybridization in colorectal cancer patients. J BUON 2009;14(3):495–9.
[15] Uleri E et al. Multiple signatures of the JC polyomavirus in paired normal and altered colorectal mucosa indicate a link with human colorectal cancer, but not with cancer progression. Int J Mol Sci 2019;20(23).
[16] Mou X et al. Prevalence of JC virus in Chinese patients with colorectal cancer. PLoS ONE 2012;7(5):e35900.
[17] Ksiaa F et al. Assessment and biological significance of JC polyomavirus in colorectal cancer in Tunisia. J BUON 2015;20(3):762–9.
[18] Shavaleh R et al. Association between JC virus and colorectal cancer: systematic review and meta-analysis. Infect Dis (Lond) 2020;52(3):152–60.
[19] Davey NE, Trave G, Gibson TJ. How viruses hijack cell regulation. Trends Biochem Sci 2011;36(3):159–69.
[20] Nobre LV et al. Human cytomegalovirus interactome analysis identifies degradation hubs, domain associations and viral protein functions. Elife 2019;8.
[21] Cook HV et al. Viruses.STRING: A virus-host protein-protein interaction database. Viruses 2018;10(10).
[22] Yang X et al. HVIDB: a comprehensive database for human-virus protein-protein interactions. Brief Bioinform 2021;22(2):832–44.
[23] Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.
[24] Ludvigsen M et al. Proteomic characterization of colorectal cancer cells versus normal-derived colon mucosa cells: approaching identification of novel diagnostic protein biomarkers in colorectal cancer. Int J Mol Sci 2020;21(10).
[25] Buttacavoli M et al. Proteomic profiling of colon cancer tissues: discovery of new candidate biomarkers. Int J Mol Sci 2020;21(9).
[26] Saleem S et al. Proteomics analysis of colon cancer progression. Clin Proteomics 2019;16:44.
[27] Zhou Y et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10(1):1523.
[28] Kumar M et al. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res 2020;48(D1):D296–306.
[29] Hatos A et al. DisProt: intrinsic protein disorder annotation in 2020. Nucleic Acids Res 2020;48(D1):D269–76.
[30] Burley SK et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. Nucleic Acids Res 2021;49(D1):D437–51.
[31] Yang J et al. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci U S A 2020;117(3):1496–503.
[32] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res, 2007. 35(Web Server issue): p. W407-10.
[33] Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010;66(Pt 1):12–21.
[34] Sobolev OV et al. A global ramachandran score identifies protein structures with unlikely stereochemistry. Structure 2020;28(11):1249–1258 e2.
[35] Hussein HA et al. PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 2015;43(W1):W436–42.
[36] Wang Y et al. A crowdsourcing open platform for literature curation in UniProt. PLoS Biol 2021;19(12):e3001464.
[37] Gaulton A et al. The ChEMBL database in 2017. Nucleic Acids Res 2017;45(D1):D945–54.
[38] Lipinski CA. Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 2004;1(4):337–41.
[39] O'Boyle NM et al. Open Babel: An open chemical toolbox. J Cheminform 2011;3:33.
[40] Dallakyan S, Olson AJ. Small-molecule library screening by docking with PyRx. Methods Mol Biol 2015;1263:243–50.
[41] Abraham MJ et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1:19–25.
[42] Zoete V et al. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 2011;32(11):2359–68.
[43] Kumari R et al. g_mmpbsa–a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 2014;54(7):1951–62.
[44] Doncheva NT et al. Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc 2012;7(4):670–85.
[45] Lozach PY. Early virus-host cell interactions. J Mol Biol 2018;430(17):2555–6.
[46] Zhang P et al. A protein network descriptor server and its use in studying protein, disease, metabolic and drug targeted networks. Brief Bioinform 2017;18(6):1057–70.
[47] Sanchez EL, Lagunoff M. Viral activation of cellular metabolism. Virology 2015;479–480:609–18.
[48] Emmett SR et al. The cell cycle and virus infection. Methods Mol Biol 2005;296:197–218.
[49] Samano-Sanchez H, Gibson TJ. Mimicry of short linear motifs by bacterial pathogens: A drugging opportunity. Trends Biochem Sci 2020;45(6):526–44.
[50] Hraber P et al. Resources to discover and use short linear motifs in viral proteins. Trends Biotechnol 2020;38(1):113–27.
[51] Gartland M et al. Susceptibility of global HIV-1 clinical isolates to fostemsavir using the PhenoSense(R) entry assay. J Antimicrob Chemother 2021;76(3):648–52.
[52] Lai YT. Small molecule HIV-1 attachment inhibitors: discovery, mode of action and structural basis of inhibition. Viruses 2021;13(5).
[53] Pancera M et al. Crystal structures of trimeric HIV envelope with entry inhibitors BMS-378806 and BMS-626529. Nat Chem Biol 2017;13(10):1115–22.
[54] Smith SJ et al. HIV-1 integrase inhibitors with modifications that affect their potencies against drug resistant integrase mutants. ACS Infect Dis 2021;7(6):1469–82.
[55] Craigie R. The molecular biology of HIV integrase. Future Virol 2012;7(7):679–86.
[56] Tsiang M et al. Antiviral activity of bictegravir (GS-9883), a novel potent HIV-1 integrase strand transfer inhibitor with an improved resistance profile. Antimicrob Agents Chemother 2016;60(12):7086–97.
[57] Bhardwaj VK, Purohit R. A new insight into protein-protein interactions and the effect of conformational alterations in PCNA. Int J Biol Macromol 2020;148:999–1009.
[58] Mesri EA, Feitelson MA, Munger K. Human viral oncogenesis: a cancer hallmarks analysis. Cell Host Microbe 2014;15(3):266–82.
[59] Rigoutsos I et al. In silico pattern-based analysis of the human cytomegalovirus genome. J Virol 2003;77(7):4326–44.
[60] Dunn W et al. Functional profiling of a human cytomegalovirus genome. Proc Natl Acad Sci U S A 2003;100(24):14223–8.
[61] Nightingale K et al. Human cytomegalovirus protein RL1 degrades the antiviral factor SLFN11 via recruitment of the CRL4 E3 ubiquitin ligase complex. Proc Natl Acad Sci U S A 2022;119(6).
[62] Shikhagaie M et al. The human cytomegalovirus-specific UL1 gene encodes a late-phase glycoprotein incorporated in the virion envelope. J Virol 2012;86(8):4091–101.
[63] Pappa KI et al. Novel structural approaches concerning HPV proteins: Insight into targeted therapies for cervical cancer (Review). Oncol Rep 2018;39(4):1547–54.
[64] Orba Y et al. Large T antigen promotes JC virus replication in G2-arrested cells by inducing ATM- and ATR-mediated G2 checkpoint signaling. J Biol Chem 2010;285(2):1544–54.
[65] Zheng HC et al. The oncogenic roles of JC virus T antigen in breast carcinogenesis. Front Mol Biosci 2021;8:687444.
[66] Chen C et al. The biology and role of CD44 in cancer progression: therapeutic implications. J Hematol Oncol 2018;11(1):64.

[67] Medrano-Gonzalez PA et al. Proteolytic processing of CD44 and its implications in cancer. Stem Cells Int 2021;2021:6667735.
[68] Link A et al. JC virus mediates invasion and migration in colorectal metastasis. PLoS ONE 2009;4(12):e8146.
[69] Thoma OM, Neurath MF, Waldner MJ. Cyclin-dependent kinase inhibitors and their therapeutic potential in colorectal cancer treatment. Front Pharmacol 2021;12:757120.
[70] Riva V, Maga G. From the magic bullet to the magic target: exploiting the diverse roles of DDX3X in viral infections and tumorigenesis. Future Med Chem 2019;11(11):1357–81.
[71] Zhao L et al. Multifunctional DDX3: dual roles in various cancer development and its related signaling pathways. Am J Cancer Res 2016;6(2):387–402.
[72] Adib E et al. CDH1 germline variants are enriched in patients with colorectal cancer, gastric cancer, and breast cancer. Br J Cancer 2022;126(5):797–803.
[73] Kumara H et al. P-Cadherin (CDH3) is overexpressed in colorectal tumors and has potential as a serum marker for colorectal cancer monitoring. Oncoscience 2017;4(9–10):139–47.
[74] Ruttkay-Nedecky B et al. Relevance of infection with human papillomavirus: the role of the p53 tumor suppressor protein and E6/E7 zinc finger proteins (Review). Int J Oncol 2013;43(6):1754–62.
[75] Marziali F et al. Differential expression of DLG1 as a common trait in different human diseases: an encouraging issue in molecular pathology. Biol Chem 2019;400(6):699–710.
[76] Wise-Draper TM et al. Overexpression of the cellular DEK protein promotes epithelial transformation in vitro and in vivo. Cancer Res 2009;69(5):1792–9.
[77] Bui QT et al. Ubiquitin-conjugating enzymes in cancer. Cells 2021;10:6.
[78] Zhang X et al. Mechanism and disease association with a ubiquitin conjugating E2 enzyme: UBE2L3. Front Immunol 2022;13:793610.
[79] Liu Z et al. Cytochrome C inhibits tumor growth and predicts favorable prognosis in clear cell renal cell carcinoma. Oncol Lett 2019;18(6):6026–32.
[80] Via A et al. How pathogens use linear motifs to perturb host cell networks. Trends Biochem Sci 2015;40(1):36–48.
[81] Marin M, Uversky VN, Ott T. Intrinsic disorder in pathogen effectors: protein flexibility as an evolutionary hallmark in a molecular arms race. Plant Cell 2013;25(9):3153–7.
[82] Xue B et al. Structural disorder in viral proteins. Chem Rev 2014;114 (13):6880–911.
[83] Meanwell NA et al. Inhibitors of HIV-1 attachment: the discovery and development of temsavir and its prodrug fostemsavir. J Med Chem 2018;61 (1):62–80.
[84] Mifsud EJ, Hayden FG, Hurt AC. Antivirals targeting the polymerase complex of influenza viruses. Antiviral Res 2019;169:104545.
[85] Williamson GM, Jackson D. The antiviral activity of the isoquinolines famotine and memotine in respiratory infections in man. Bull World Health Organ 1969;41(3):665–70.