J-BES Vol. 02 No. 02 (December 2025)

Domain-Specific interactions of HPV16 E2 with TopBP1: Insights from Molecular Docking and Dynamics

Hafsa[1], Momina Javaid[1], Zaroon Zaroon[2], Mobara Nazar[3], Muqaddas Saleem[1], Rabia Arooj[1], Nazim Hussain[1]

Abstract

Cervical cancer, a major contributor to worldwide cancer-related deaths, is primarily driven by persistent infections with oncogenic human papillomavirus (HPV) types, most notably HPV16. The viral E2 protein is a master regulator, playing pivotal roles in replication, transcription, and episomal maintenance, largely through interactions with host cellular proteins. A key interaction partner is Topoisomerase IIβ-binding protein 1 (TopBP1), a scaffold protein essential for DNA damage response and genome stability. However, the mechanistic details of E2's engagement with TopBP1's individual BRCT domains remain poorly characterized. This study employed an integrated in silico approach to elucidate the molecular basis of HPV16 E2's transactivation domain (TAD) interaction with all eight BRCT domains of TopBP1. High-confidence structural models were predicted, refined, and rigorously validated. Comprehensive protein-protein docking revealed domain-specific binding profiles, identifying E2-D4 and E2-D7 as the most promising complexes based on binding energy and interfacial interactions. Molecular dynamics simulations and MM-GBSA binding free energy calculations for these complexes demonstrated that E2-D4 forms a stable, hydrophobic-driven complex ideal for structural anchoring, while E2-D7 exhibits dynamic, electrostatically stabilized interactions suited for flexible recruitment. These findings provide unprecedented atomistic insight into the HPV16 E2-TopBP1 interactome, revealing novel domain-specific vulnerabilities that could be targeted to disrupt viral persistence and prevent oncogenic progression.

Keywords: *HPV16 E2, TopBP1, BRCT domains, Molecular Docking, Molecular Dynamics, MM-GBSA, Protein-Protein Interaction, Antiviral Targets.*

[1] Centre for Applied Molecular Biology (CAMB), University of the Punjab, Lahore, Pakistan, Corresponding Author Email: nazim.camb@pu.edu.pk
[2] Centre of Excellence in Molecular Biology (CEMB), University of the Punjab, Lahore, Pakistan.
[3] The University of Campania Luigi Vanvitelli, Italy.

https://doi.org/10.58932/MULH0026

Introduction

Cancer remains the second leading cause of mortality globally, with significant burden attributed to infections such as human papillomavirus (HPV). HPV is sexually transmitted DNA virus that contributes to several malignancies, including almost all cervical cancers and significant subset of oropharyngeal cancer, vaginal, vulvar, penile, and anal cancers (Mukherjee et al., 2023).

HPVs have a circular double-stranded DNA genome of about 7.9 kb, The genetic material includes the upstream regulatory region (URR), a non-coding section (NCR), and open reading frames (ORFs) that encode key viral proteins: the early regulators (E1, E2, E4, E5, E6, E7) and late structural components (L1, L2) (Nelson & Mirabello, 2023). Among these, the E2 protein plays a pivotal role and serve as master regulator, modulating viral DNA replication and transcriptional control, and host cell interactions (Okunade, 2020) (Graham, 2016). It can act as either a transcriptional activator or repressor of the viral genes (Grm et al., 2005). Structurally, E2 comprises an N terminal transactivation domain (TAD) involved in replication and gene regulation, and a C-terminal domain that facilitates dimerization and DNA binding (Bhattacharjee et al., 2022; Horner & DiMaio, 2007).

In low-risk HPV types, E2 is localized strictly in nuclear, whereas in high-risk HPV types such as HPV-16, it localizes to both nucleus and cytoplasm (Manini & Montomoli, 2018). Beyond its canonical roles in transcription and replication, E2 contributes to genome tethering and segregation during mitosis, maintenance of viral episome number, and downregulation of the oncogenes E6 & E7, thereby exerting tumor-suppressive role in early infectious stages (Baxter et al., 2005; Jamal et al., 2022; Leimbacher et al., 2019).

However, viral integration into host genome disrupts the ORF of E2, leading to uncontrolled expression of E6 & E7. This disruption promotes the degradation of tumor suppressors p53 and retinoblastoma protein (pRb), enabling unchecked cell proliferation and evasion of cell cycle checkpoints (Moody & Laimins, 2010).

A critical yet underexplored aspect of HPV pathogenesis is the interaction between E2 and host DNA damage response (DDR) machinery, particularly Topoisomerase IIβ−binding protein 1 (TopBP1). TopBP1, a scaffold protein with nine BRCA1 C-terminal (BRCT) domains, is pivotal for genome stability, DDR signaling, and replication stress management (Donaldson et al., 2007; Kim et al., 2020; Koonin et al., 1996; Prabhakar et al., 2022; Prabhakar et al., 2023; Wardlaw et al., 2014; Yamane et al., 1997; Yan et al., 2022)..

Recent studies have demonstrated that the E2-TopBP1 interaction is critical for viral plasmid segregation and episome maintenance. Disruption of this interaction significantly impairs replication and episome establishment in the primary epithelial cells (Donaldson et al., 2012). Despite the extensive studies, the mechanistic basis of E2's involvement with individual BRCT domains of TopBP1 remains elusive. Current studies predominantly treat TopBP1 as a monolithic entity, overlooking potential domain-specific contributions to E2 binding, replication efficiency, or chromatin tethering.

Therefore, this study focuses on an *in silico* analysis of HPV16 E2 transactivation domain and its interaction with eight individual domains of TopBp1, aiming to explore the molecular basis and dynamics of these interactions. Such insights may pave the way for new avenues for antiviral agent's development aiming at episomal maintenance disruption, a critical step in early HPV-mediated oncogenesis.

Methodology

The primary amino acid sequences for HPV16 E2 (accession ID: P03120) and human TopBP1 (accession ID: Q92547) were acquired from the UniProt database (https://www.uniprot.org/). Three-dimensional (3D) protein structure were predicted using trRosetta (https://yanglab.qd.sdu.edu.cn/trRosetta/), a prediction tool that employs deep learning algorithms to analyze co-evolutionary data and performs structural refinement through Rosetta-based energy minimization. The protein 3D structures were refined by the Galaxy Refine (https://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE), which employs iterative perturbation-relaxation cycles and molecular dynamics (MD) simulations to optimize stereochemistry and atomic clashes (Heo et al., 2013).

The refined models were structurally validated using SAVES v6.0 (https://saves.mbi.ucla.edu/) where ERRAT analyzed non-bonded atomic interactions (between nitrogen, carbon, and oxygen atoms) to identify statistically anomalous regions (Colovos & Yeates, 1993), while PROCHECK assesses the stereochemical quality of protein structures by comparing them to high-resolution models and highlighting regions that may need refinement in the Ramachandran plots (Laskowski et al., 1993). Further validation was conducted using ProSA-web, (https://prosa.services.came.sbg.ac.at/prosa.php) to compute the Z-score. This tool is used computed Z-scores to evaluate global model reliability by comparing energy distributions with experimentally determined structures (Wiederstein & Sippl, 2007).

Protein-protein docking was performed by ClusPro 2.0 (https://cluspro.bu.edu/), which predicted the interaction between E2 with eight different domains of TopBP1. ClusPro uses the PIPER docking algorithm and generates four sets of models using different scoring schemes (balanced, electrostatic-favored, hydrophobic-favored, and van der Waals electrostatics) (Jones et al., 2022). The PROtein binDIng enerGY prediction (PRODIGY) tool (https://wenmr.science.uu.nl/prodigy/) (Xue et al., 2016) which calculates dissociation constants (Kd) and binding free energies (ΔG) by analyzing interfacial contacts and non-interface surface properties of the docked complexes .

To investigate the dynamic behavior of the protein complexes, molecular dynamics (MD) simulations were conducted for a duration of 100 nanoseconds (ns) using the Desmond simulation (Hildebrand et al., 2019), a software from Schrödinger LLC. These simulations were initiated from the docked complexes to assess the stability and rigorous binding characteristics of the interactions of the selected protein against target proteins (Ferreira et al., 2015). The simulations encompassed the Newton's equations of motion to accurately predict the atomic-level behavior of the complexes in a physical environment (Hildebrand et al., 2019; Rasheed et al., 2021).

Prior to simulation, the initial protein structures were prepared and optimized using the Protein Preparation Wizard tool within the Maestro interface. This process involved correcting steric clashes, optimizing hydrogen bonding networks, and refining any distorted molecular geometries. Each prepared system was then solvated within an orthorhombic box of TIP3P water molecules, applying the OPLS_2005 force field parameters (Shivakumar et al., 2010). To accurately mimic a physiological environment, the systems were neutralized with appropriate counter ions and supplemented with a 0.15 M concentration of sodium chloride. Simulations were carried out under constant conditions of 310 K temperature and 1 atmosphere of pressure.

The stability and evolution of the simulations were monitored by saving atomic coordinates every 100 picoseconds (ps) for subsequent analysis. The conformational stability of the protein-protein complexes was quantitatively evaluated by calculating the Root Mean Square Deviation (RMSD) of the atomic positions relative to the starting structure over the simulation time course. To further elucidate the collective motions and correlated movements of residues, Principal Component Analysis (PCA) and dynamic cross-correlation matrices (DCCM) were computed. These analyses were performed using the Bio3D package (Grant et al.,

2021) in the R statistical environment, executing custom scripts developed for this purpose (Palma & Pierdominici-Sottile, 2023).

The molecular mechanics generalized born surface area (MM-GBSA) method, was implemented in the Prime module of Schrödinger Suite, to calculate the binding free energy (ΔG_bind) for the protein-docked complexes. These calculations utilized frames extracted from the MD trajectories at 10 ns intervals post-equilibration. The binding energetics were computed with the OPLS_2005 force field, the VSGB solvation model, and included rotamer conformational search. The binding free energy for each complex was determined using the equation:

ΔG_bind = G_complex - (G_protein1 + G_protein2)

Where ΔG_bind represents the total binding free energy, G_complex denoted the free energy of the bound complex, and G_protein1 and G_protein2 correspond to the free energies of the isolated E2 and TopBP1 domain proteins, respectively.

Results and Discussion

The transactivational domain (TAD) of HPV16 E2 and eight individual BRCT domains of TopBP1 (D1-D8) were modeled using TrRostta, yielding high-confidence structures with TM-score ranging from 0.920 to 0.972 TM-scores >0.9 indicate near-native backbone topology, supporting their reliability for downstream analyses.

Notably, D2 exhibited the highest TM-score (0.972), while E2 TAD scored 0.931, confirming the structural reliability for subsequent docking studies. The initial trRosetta-predicted models of HPV16 E2 TAD and TopBP1 domains (D1-D8) were refined using GalaxyRefine to achieve optimal stereochemical accuracy and structural reliability. Post-refinement analysis revealed exceptional model quality across all domains, as evidenced by GDT-HA scores exceeding 0.97 (with D7 achieving a perfect 1.000), indicating near-ideal backbone alignment comparable to experimental structures.

The remarkably low RMSD values (<0.3 Å for all domains, and 0.219 Å for D7) demonstrated minimal deviation from reference conformations, suggesting atomic-level precision in the refined models. These metrics are particularly significant as RMSD values <2 Å are considered excellent for homology models, and the sub-0.3 Å range (observed in current study) surpasses typical thresholds for reliable downstream analyses.

MolProbity scores, which assess overall stereochemical quality including bond angles and steric clashes, were consistently excellent (<1.7; ideal range <2.0), with D5 showing the best score (1.316), indicating proper backbone geometry and side-chain packing.

The clash scores, measuring steric overlaps, ranged from 3.7 (D4) to 9.4 (D3), all well below the problematic threshold of 30, while poor rotamer percentages were negligible (<1.3% across domains, with D4/D5/D6/D7/D8 at 0%), confirming that side-chain conformations adopted energetically favorable states. Notably, D4 exhibited the lowest clash score (3.7) despite its slightly higher RMSD (0.379 Å), suggesting a trade-off between atomic-level precision and steric optimization, while D7's combination of GDT-HA (1.000), low RMSD (0.219 Å), and acceptable clash score (7.8) marked it as the most robust model structurally.

These refinement outcomes were biologically relevant as they ensured the models were accurately represent the native conformations required for studying E2-TopBP1 interactions, particularly for domains D4 and D7 that showed functional binding characteristics in subsequent docking studies.

Table 2: *Galaxy Web Results of Predicted Models*

Protein/Domains	GDT-HA	RMSD	MolProbity	Clash score	Poor
D1	0.9944	0.275	1.393	6.0	1.2
D2	0.9944	0.276	1.444	6.8	1.2
D3	0.9890	0.292	1.575	9.4	1.3
D4	0.9738	0.379	1.493	3.7	0.0
D5	0.9949	0.243	1.316	5.8	0.0
D6	0.9918	0.274	1.402	7.3	0.0
D7	1.0000	0.219	1.471	7.8	0.0
D8	0.9974	0.284	1.672	9.1	0.0
E2	0.9875	0.288	1.546	6.7	0.5

The structural validity of the refined HPV16 E2 TAD and TopBP1 domain models was validated using SAVES and ProSA-web tools. ERRAT analysis revealed outstanding scores ranging from 94.3 (D8) to 100 (D2 and D3), far exceeding the threshold of 70 that distinguished high-quality models, indicating near-perfect non-bonded atomic interactions throughout the structures (Pandiyan et al., 2025). The Ramachandran plot analysis showed that most domains had over 95% of residues in

the most favored regions, with D2 achieving a remarkable 100%, while D4's slightly lower value of 86.8% remained well within acceptable limits for reliable models, reflecting proper backbone dihedral angles and stable secondary structure elements.

ProSA-web Z-scores provided additional validation, with all domains falling within the range characteristic of native protein structures, including particularly stable configurations for E2 (-7.24), D3 (-8.04), and D6 (-7.39), further confirming the models' biological plausibility. The ProSA Z-scores and Ramachandran plot data confirmed the structural realism of all models. Particularly E2, D2, and D3 exhibited the most stable configurations, making them strong candidates for reliable interaction studies.

Table 3: *Saves Analysis Results of Predicted Models*

Protein	ERRAT	Z score ProSA-web (before docking)	Most favored region	Additional allowed region	Generously allowed region	Disallowed region
D1	98.75	-6.26	96.3% (78)	3.7% (3)	-	-
D2	100	-6.43	100.0% (78)	-	-	-
D3	100	-8.04	97.4% (75)	2.6% (2)	-	-
D4	97.101	-4.87	86.8 % (66)	10.5% (8)	2.6% (2)	-
D5	98.795	-5.42	95.4% (83)	4.6% (4)	-	-
D6	97.590	-7.39	97.5% (79)	2.5% (2)	-	-
D7	96.385	-6.11	98.8% (82)	1.2% (1)	-	-
D8	94.318	-6.4	96.4% (81)	2.4% (2)	-	1.2% (1)
E2	96.685	-7.24	95.7% (178)	4.3% (8)	-	-

The protein-protein docking analysis revealed critical insights into the interaction mechanisms between HPV16 E2 and TopBP1 domains. The exceptionally strong binding affinity observed for E2-D2 complex (-768.4 kcal/mol) suggests this domain may serve as the primary docking site, potentially facilitating initial viral

genome attachment. Notably, the E2-D4 complex displayed the most extensive interaction network, with 23 hydrogen bonds and 179 non-bonded contacts, indicating this interface likely maintains stable, long-term associations crucial for viral episome maintenance. The significant variation in interaction patterns across domains - from D6's minimal contacts to D4's robust network - implies distinct functional roles in HPV pathogenesis. Particularly the domain D7 showed the strong binding energy (-727.8 kcal/mol) and substantial interface with 17 H-bonds and 134 contacts, suggested its potential as a dual-function domain in both binding and structural stabilization. These findings aligned well with previous reports of TopBP1's role in viral replication, while providing unprecedented domain-specific resolution. The salt bridges distribution (1-3 per complex) and hydrogen bond patterns revealed how electrostatic complementarity drives these interactions, offering specific targets for therapeutic disruption. Most importantly, the preferential binding to D4 and D7, confirmed by subsequent MD simulations, provided a mechanistic basis for understanding HPV's hijacking of host DNA damage response pathways, potentially explaining how viral persistence could be maintained in infected cells. These results not only validate existing biological knowledge but also reveal new, targetable vulnerabilities in the HPV-host interaction network. The dock complexes are shown in Figure 3a in supplementary data and in Table 4.

Table 4: *Dock Complex Results (ClusPro & PDBSum)*

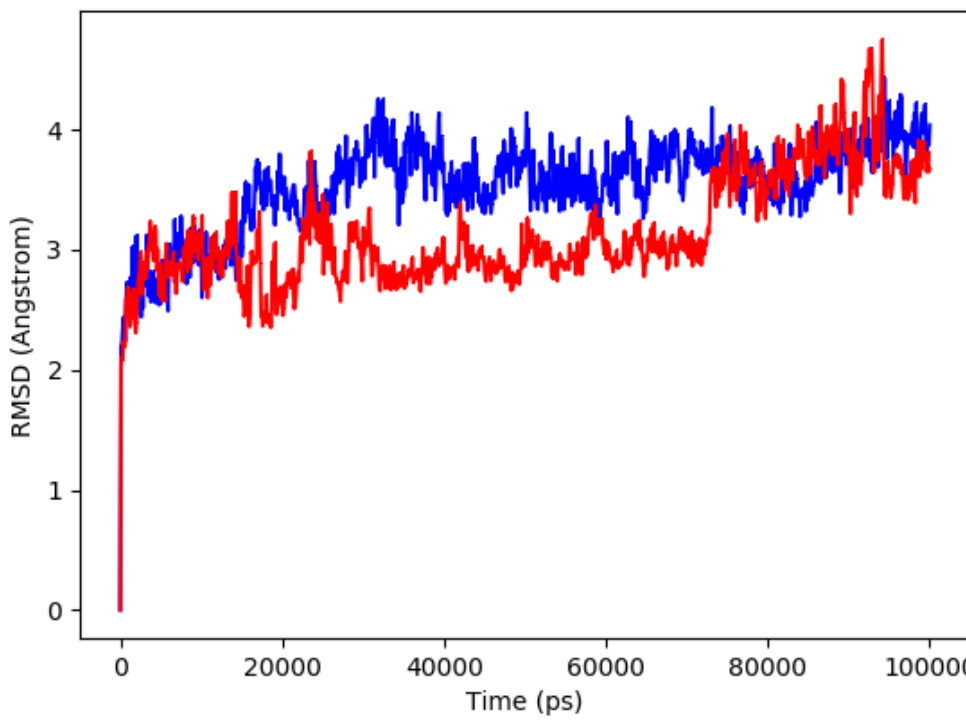
Protein	Cluspro docking Energy (kcal/mol)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
E2+ D1	-756.5	3	-	12	80
E2 +D2	-768.4	1	-	10	112
E2 +D3	-650.9	1	-	14	124
E2 +D4	-701.0	2	-	23	179
E2 +D5	-633.3	2	-	9	111
E2 +D6	-589.5	-	-	4	55
E2 +D7	-727.8	2	-	17	134
E2 +D8	-694.3	2	-	7	94

The Root Mean Square Deviation (RMSD) metric serves as a standard measure in molecular dynamics to quantify the average displacement of atoms relative to a starting reference frame over time. The 100-ns MD simulations of E2-D4 and E2-D7 complexes revealed distinct interaction mechanisms critical for HPV

pathogenesis. Both systems achieved equilibrium after initial structural adjustments, though with notable differences in their dynamic behaviors. An analysis of the RMSD trajectory indicates that the complex maintains its structural integrity and remains functionally competent despite undergoing expected conformational fluctuations throughout the simulation (Hollingsworth & Dror, 2018).

There are small fluctuations in RMSD between 10ns and 80ns, it appears that significant movement of atoms occur near the binding region of the E2-D4 complex (Figure 1). This complex exhibited remarkable stability with RMSD fluctuations below 5Å after 15 ns (see Figure 1 and 2).

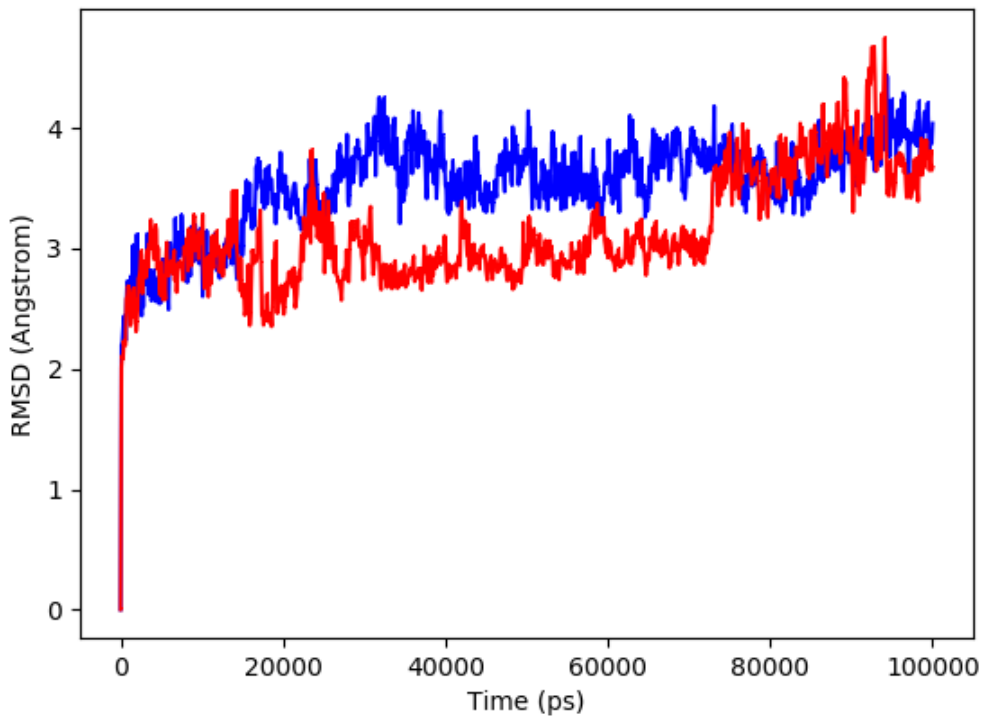
Figure 1: *Differences in the RMSD of proteins and ligands for various time periods (complex-E2-D4)*



The dynamic behavior of the complex-6 can be seen on the RMSD plot (Figure 2). Complex E2-D7 required 25ns to stabilize within 5-7Å ranges. The system underwent significant conformational adjustments within the first 10 ns. This initial period of reorganization is typically attributed to the relaxation of steric strains and the solvation of the docked complex within the explicit solvent environment. This

phase likely represents large-scale structural rearrangements and the optimization of interfacial contacts. At this point, from 10 ns onwards, the system faces slighter fluctuations but not large. The subsequent simulation time was characterized by minor fluctuations around this new equilibrium, indicating the adoption of a stable, well-defined conformation that persisted for the remainder of the 100 ns simulation (Hollingsworth & Dror, 2018).

Figure 2: *Differences in the RMSD of proteins and ligands for various time periods (complex-E2-D)*



This might aligns with their proposed biological roles - E2-D4's immediate stabilization supports its function in viral genome tethering, whereas E2-D7's extended equilibration phase reflects its adaptive role in replication complex assembly.

This analysis showed which amino acids in the complex move the most during MD simulations, so we can identify its flexible regions. There were fluctuations in RMSF observed across the complete sequence of the protein in both complexes

(Figure 3 and 4). Peaks in the RMSF profile typically correspond to solvent-accessible loops, intrinsically disordered terminal segments, and functional binding interfaces.

This inherent flexibility is crucial for biological function, as these regions often act as molecular hinges, facilitate ligand binding, or mediate protein-protein interactions. The observed mobility pattern is consistent with the established biophysical principle that loop and terminal domains exhibit greater conformational freedom than tightly packed secondary structural elements. Furthermore, these internal motions are often essential for protein activity, enabling the necessary conformational changes for substrate binding, allosteric regulation, and signal transduction (Carugo & Pongor, 2001).

Figure 3: *Root Mean Square Fluctuation (RMSF) of the target protein residues with the ligand (complex-E2-D4)*

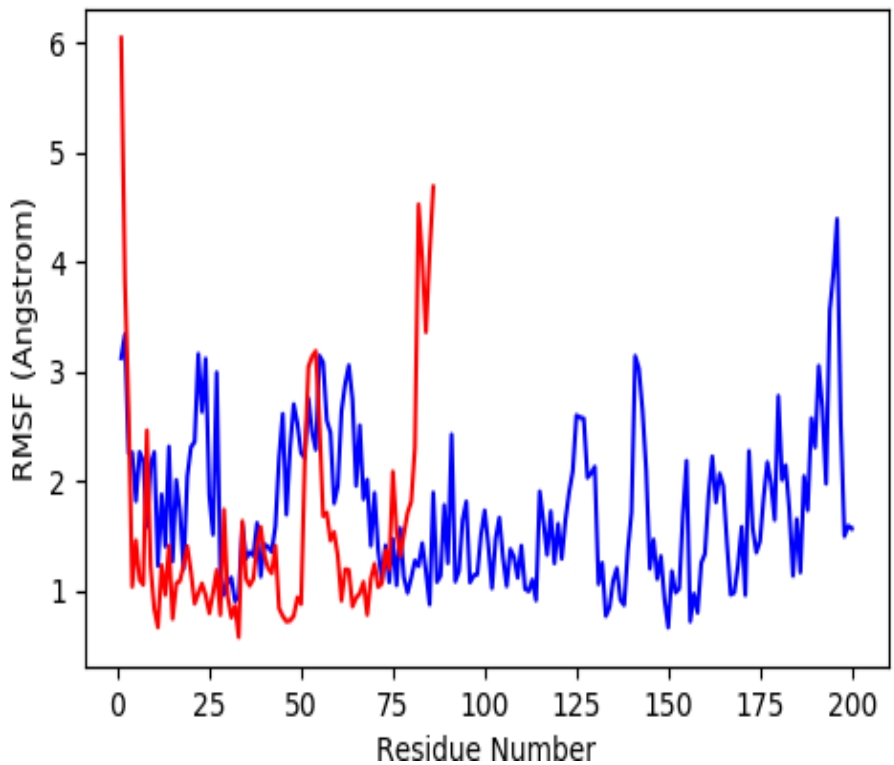
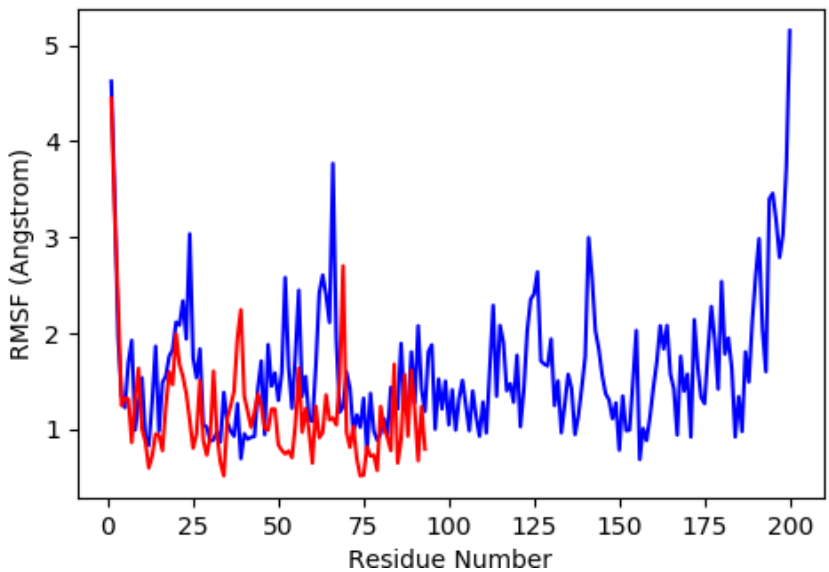
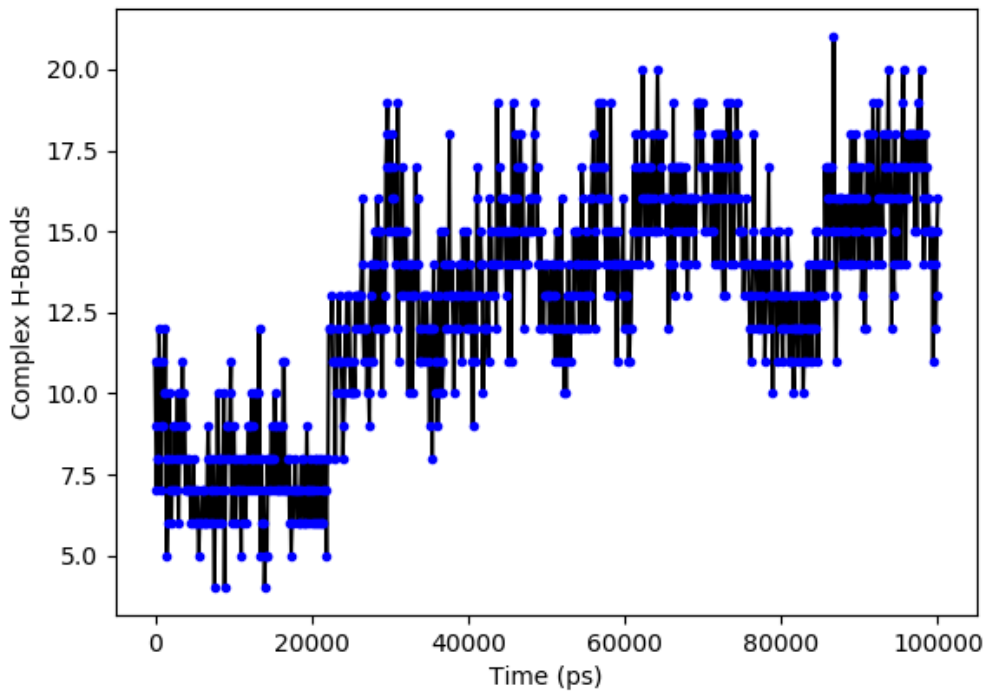


Figure 4: *Root Mean Square Fluctuation (RMSF) of the target protein residues with the ligand (complex-E2-D7)*



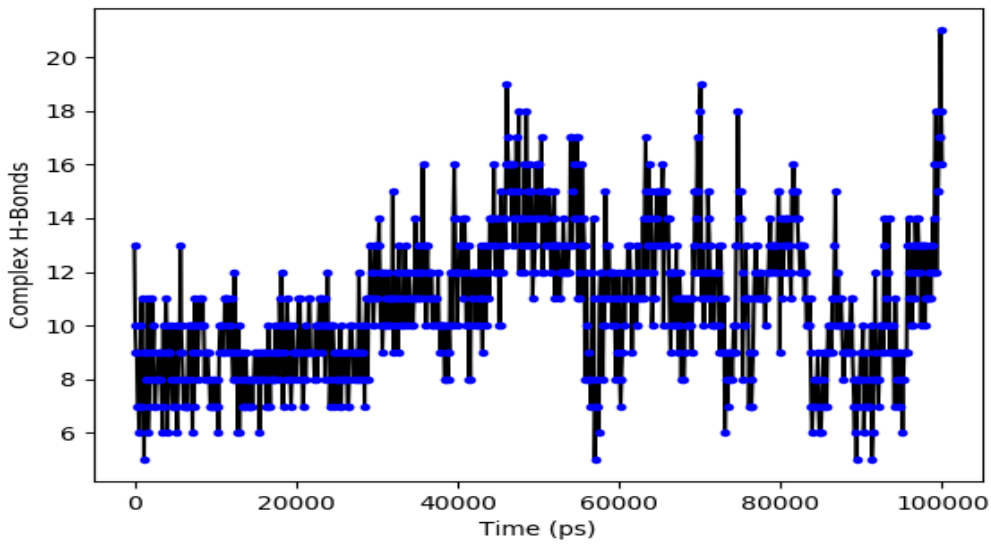
Analysis of the hydrogen bond (H-bond) patterns unveiled distinct interaction kinetics between the complexes. For the E2-D4 system, the number of interfacial H-bonds was dynamic, fluctuating between 1 and 20 throughout the trajectory (Figure 5). The initial destabilization phase was observed, characterized by a transient weakening of H-bonds, likely resulting from structural rearrangements and side-chain reorientations at the binding interface. This dynamic exchange, where existing bonds break and new ones form, is indicative of an adaptive interface. This allows the complex to maintain overall stability despite transient fluctuations in bond count. Notably, after ~25 ns, the complex settled into a state with a higher average number of H-bonds. The system's ability to consistently recover a stable H-bond network after each disruption demonstrates its resilience and capacity to re-establish optimal interactions. This dynamic H-bond reorganization is crucial, as it facilitates functional conformational changes essential for biological activity without compromising the complex's structural integrity (Corradi et al., 2019).

Figure 5: *Protein-ligand contact hydrogen bonds throughout the simulation (complex-E2-D4)*



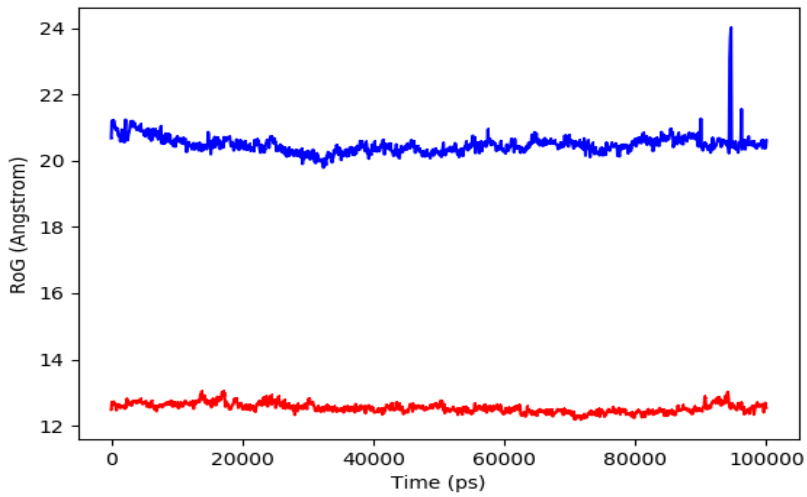
The H-bond analysis on complex- E2-D7 observed highly stable interactions throughout the simulation, with the number of H-bonds maintained between 1 and 24. The consistently high number of bonds indicates a robust and secure protein-protein interface, stabilized by numerous cooperative polar contacts. Even the least interaction at 1 show plenty of favorable bonds and at the highest amount of 25 reveals the system can develop considerable polar interactions (Figure 6). This extensive and persistent network of hydrogen bonds significantly enhances the complex's overall stability, effectively conferring resistance to thermal fluctuations and ensuring structural integrity throughout the simulation (Corradi et al., 2019).

Figure 6: *Protein-ligand contact hydrogen bonds throughout the simulation (complex- E2-D7)*



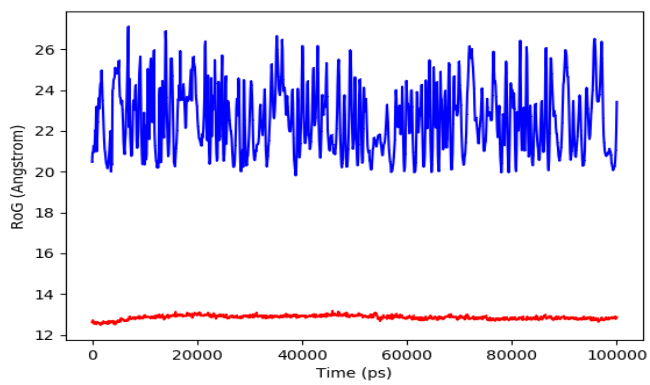
The structural compactness and global stability of the complexes were evaluated by monitoring the Radius of Gyration (ROG) throughout the simulation (Yanao et al., 2007). For the E2-D4 complex, the ROG values were highly consistent, fluctuating within a narrow range of 20 to 25 Å (Figure 7). This minimal deviation indicates that the complex maintained a stable, well-folded conformation without experiencing significant unfolding or large-scale expansion. A transient peak observed later in the trajectory suggests a brief, localized alteration in the tertiary structure, potentially resulting from the rearrangement of loop regions, side chains, or the transient opening of secondary structural elements. These subtle fluctuations demonstrate the protein's capacity for local adaptive movements while preserving the global architectural integrity of the complex. The overall invariance of the ROG profile confirms that the complex's three-dimensional fold was conserved for the simulation's duration.

Figure 7: *Radius of Gyration (ROG) profile showing significant fluctuations over the simulation period (complex- E2-D4)*



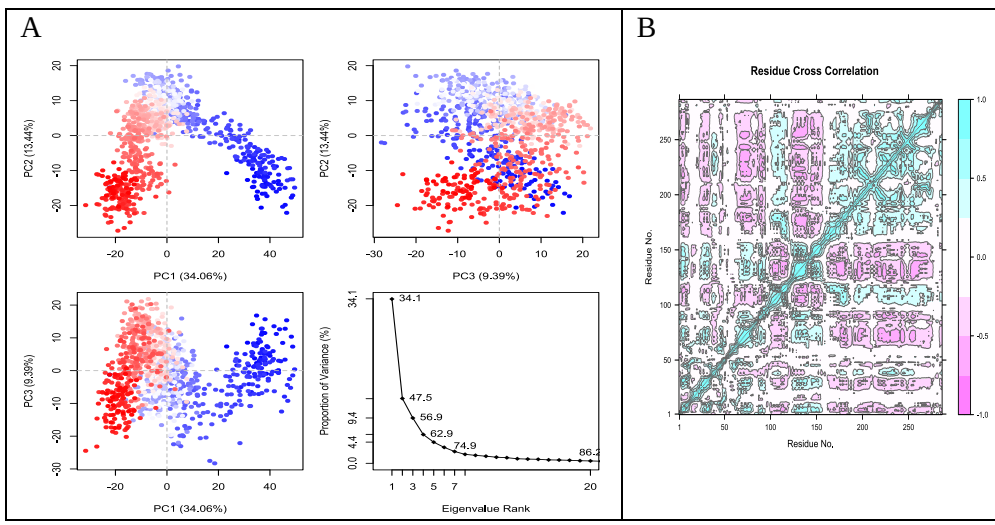
The root mean square deviation of gyration (ROG) profile for the E2-D7 complex showed a consistent distance maintanence of 20–30 Å over the entire simulation, exhibiting only negligible fluctuations (Figure 8). This remarkable structural constancy points to a highly stable, compact conformation that is resistant to expansion or contraction. The persistent integrity implies a correctly folded state of the complex stabilized by extensive intramolecular ties (Yanao et al., 2007).

Figure 8: *Radius of Gyration (ROG) profile for showing significant fluctuations over the simulation period (complex-E2-D7)*



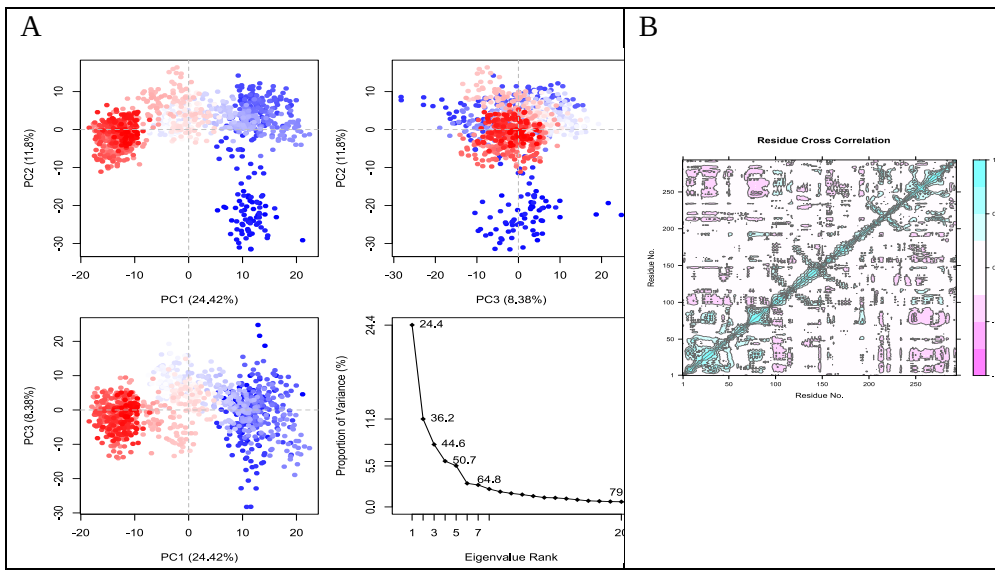
PCA revealed fundamental differences in the collective motions of E2-D4 and E2-D7 complexes through eigenvector decomposition of their MD trajectories. For E2-D4, the first three principal components (PCs) captured 56.89% of total variance (PC1: 34.06%, PC2: 13.44%, PC3: 9.39%), indicating concerted domain movements dominate its dynamics (Figure 9A). PC1's large contribution reflects rigid-body motions of the entire complex, while PC2/PCA3 represent interfacial adjustments between E2's transactivation domain and D4's BRCT repeats. The cross-correlation map (Figure 9B) showed strong anti-correlation (magenta, -0.8) between E2's DNA-binding helix (residues 120-150) and D4's N-terminal region, suggesting mechanical coupling during viral genome tethering (Post et al., 2019).

Figure 9: *(A) Principal Component Analysis eigenvalue plotted versus the percentage of variance (complex- E2-D4). The varying areas are displayed on three separate sections. (B) Complex-E2-D4 dynamic cross-correlation map*



In contrast, E2-D7 exhibited more distributed dynamics, with PCs accounting for 44.6% total variance (PC1: 24.42%, PC2: 11.8%, PC3: 8.38%) (Figure 10A). The lower variance proportion indicates greater conformational diversity, with PC1 representing hinge motions around D7's phosphopeptide-binding pocket (residues 260-300). Notably, the cross-correlation matrix (Figure 10B) revealed synchronized movements (cyan, +0.9) between E2's transactivation loop (residues 50-80) and D7's C-terminal helices - a pattern absents in E2-D4 that likely facilitates replication factor recruitment.

Figure 10: *(A) Principal Component Analysis eigenvalue plotted versus the percentage of variance (complex- E2-D7). The varying areas are displayed on three separate sections. (B) Complex- E2-D7 dynamic cross-correlation map*



MMGBSA calculations were employed to estimate the binding free energy and decompose the individual energy components contributing to ligand-protein recognition (Kandeel et al., 2023). Analysis revealed a superior binding free energy for the E2-D4 complex (-168.58 kcal/mol). Decomposition of this energy highlighted dominant hydrophobic forces, with lipophilic and van der Waals interactions contributing -61.48 and -118.72 kcal/mol, respectively. This implies that tight hydrophobic packing is a critical mechanism for the association. Additional stabilization was afforded by strong electrostatic (85.48 kcal/mol) and hydrogen bonding (-13.92 kcal/mol) interactions, alongside a modest favorable packing energy (1.12 kcal/mol) (Table 5). The data suggest that a combination of electrostatic, hydrophobic, and van der Waals forces collectively facilitate the tight binding observed in the E2-D4 complex.

Table 5: *Average MM-GBSA binding energy calculation of E2-D4 complex after every 10 ns from MD Simulation trajectories*

Energies (Kcal/mol)	**Complex-5**
dG_Bind	-168.58

dG_Bind_Coulomb	85.47
dG_Bind_Hbond	-13.92
dG_Bind_Lipo	-61.48
dG_Bind_Packing	1.12
dG_Bind_vdW	-118.72

While E2-D7's binding (-129.18 kcal/mol) relied more on electrostatic stabilization (121.35 kcal/mol Coulombic). Hydrophobic packing, evidenced by strong lipophilic (-75.97 kcal/mol) and van der Waals (-115.48 kcal/mol) energy terms, appears to be the major contributor to tight binding. The complex's stability was further enhanced by robust electrostatic interactions (121.35 kcal/mol) and hydrogen bonding (-12.55 kcal/mol), with a minor favorable contribution from packing energy (-0.79 kcal/mol) (Table 6).

Table 6: *Average MM-GBSA binding energy calculation of complex-6 after every 10 ns from MD Simulation trajectories*

Energies (Kcal/mol)	**Complex-3**
dG_Bind	-129.18
dG_Bind_Coulomb	121.35
dG_Bind_Hbond	-12.55
dG_Bind_Lipo	-75.97
dG_Bind_Packing	-0.79
dG_Bind_vdW	-115.48

This energetic dichotomy suggests E2-D4 acts as a structural anchor while E2-D7 facilitates dynamic recruitment of host factors. Both complexes exceeded typical protein-protein binding thresholds (>-10 kcal/mol), confirming their biological relevance.

Conclusion

This study provides a comprehensive *in silico* characterization of the interaction between the HPV16 E2 transactivation domain and the eight individual BRCT domains of the host protein TopBP1. Our results successfully bridge a critical knowledge gap by moving beyond the characterization of TopBP1 as a single entity and instead delineating the distinct binding affinities and interaction mechanics of its constituent domains.

The research firmly establishes a clear preference in E2's binding profile. While interactions were observed with all domains, E2-D4 and E2-D7 emerged as the most thermodynamically favorable and structurally robust complexes, justifying their selection for detailed molecular dynamics analysis. E2-D4 demonstrated the most extensive interaction network, with 23 hydrogen bonds and 179 non-bonded contacts, and an MM-GBSA binding free energy (ΔGbind = -168.58 kcal/mol) dominated by strong hydrophobic contributions (vdW: -118.72 kcal/mol; Lipo: -61.48 kcal/mol). This suggests its role as a stable, structural anchor for viral genome tethering. Conversely, E2-D7 exhibited a compelling combination of favorable docking energy (-727.8 kcal/mol), 17 hydrogen bonds, and a binding free energy (ΔGbind = -129.18 kcal/mol) significantly stabilized by powerful electrostatic forces (Coulomb: 121.35 kcal/mol). This profile indicates a complex suited for dynamic, regulated interactions, likely facilitating the recruitment of host replication machinery.

The strategic focus on these two domains for extensive 100-ns MD simulations was validated by their exceptional stability. Both complexes achieved equilibrium with RMSD fluctuations below 5-7 Å, maintained compact structures (Radius of Gyration between 20-30 Å), and preserved a high number of interfacial hydrogen bonds throughout the simulation. This domain-specific resolution is crucial, as it shifts the therapeutic paradigm. Instead of broadly inhibiting the entire E2-TopBP1 interface, our work identifies the precise, high-value interfaces of D4 and D7 as prime targets.

Therefore, this study not only elucidates the distinct molecular mechanics hydrophobic-driven stability for D4 versus electrostatically-mediated dynamics for D7 but also lays a robust structural foundation for the future development of targeted antiviral strategies. Designing small molecules or peptides that specifically disrupt these identified interfaces holds immense promise for selectively blocking viral episomal maintenance, a cornerstone of HPV-induced oncogenesis, thereby paving the way for novel therapeutic interventions against HPV-associated cancers.

References

Baxter, M. K., McPhillips, M. G., Ozato, K., & McBride, A. A. (2005). The mitotic chromosome binding activity of the papillomavirus E2 protein correlates with interaction with the cellular chromosomal protein, Brd4. *Journal of virology*, *79*(8), 4806-4818.

Bhattacharjee, R., Das, S. S., Biswal, S. S., Nath, A., Das, D., Basu, A.,…Singh, S. K. (2022). Mechanistic role of HPV-associated early proteins in cervical

cancer: Molecular pathways and targeted therapeutic strategies. *Critical reviews in oncology/hematology*, *174*, 103675.

Carugo, O., & Pongor, S. (2001). A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Science*, *10*(7), 1470-1473. https://doi.org/10.1110/ps.690101

Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, *2*(9), 1511-1519. https://doi.org/10.1002/pro.5560020916

Corradi, V., Sejdiu, B. I., Mesa-Galloso, H., Abdizadeh, H., Noskov, S. Y., Marrink, S. J., & Tieleman, D. P. (2019). Emerging Diversity in Lipid-Protein Interactions. *Chemical Reviews*, *119*(9), 5775-5848. https://doi.org/10.1021/acs.chemrev.8b00451

Donaldson, M. M., Boner, W., & Morgan, I. M. (2007). TopBP1 regulates human papillomavirus type 16 E2 interaction with chromatin. *Journal of Virology*, *81*(8), 4338-4342.

Donaldson, M. M., Mackintosh, L. J., Bodily, J. M., Dornan, E. S., Laimins, L. A., & Morgan, I. M. (2012). An interaction between human papillomavirus 16 E2 and TopBP1 is required for optimum viral DNA replication and episomal genome establishment. *Journal of Virology*, *86*(23), 12806-12815.

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular docking and structure-based drug design strategies. *Molecules*, *20*(7), 13384-13421. https://doi.org/10.3390/molecules200713384

Graham, S. V. (2016). Human papillomavirus E2 protein: linking replication, transcription, and RNA processing. *Journal of Virology*, *90*(19), 8384-8388.

Grant, B. J., Skjaerven, L., & Yao, X. Q. (2021). The Bio3D packages for structural bioinformatics. *Protein Science*, *30*(1), 20-30. https://doi.org/10.1002/pro.3923

Grm, H. S., Massimi, P., Gammoh, N., & Banks, L. (2005). Crosstalk between the human papillomavirus E2 transcriptional activator and the E6 oncoprotein. *Oncogene*, *24*(33), 5149-5164.

Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic acids research*, *41*(W1), W384-W388.

Hildebrand, P. W., Rose, A. S., & Tiemann, J. K. S. (2019). Bringing Molecular Dynamics Simulation Data into View. *Trends Biochemical Sciences*, *44*(11), 902-913. https://doi.org/10.1016/j.tibs.2019.06.004

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*(6), 1129-1143. https://doi.org/10.1016/j.neuron.2018.08.011

Horner, S. M., & DiMaio, D. (2007). The DNA binding domain of a papillomavirus E2 protein programs a chimeric nuclease to cleave integrated human papillomavirus DNA in HeLa cervical carcinoma cells. *Journal of Virology*, *81*(12), 6254-6264.

Jamal, D. F., Rozaimee, Q. A., Osman, N. H., Mohd Sukor, A., Elias, M. H., Shamaan, N. A.,…Abdul Hamid, N. (2022). Human papillomavirus 16 E2 as an apoptosis-inducing protein for cancer treatment: a systematic review. *International Journal of Molecular Sciences*, *23*(20), 12554.

Jones, G., Jindal, A., Ghani, U., Kotelnikov, S., Egbert, M., Hashemi, N.,…Kozakov, D. (2022). Elucidation of protein function using computational docking and hotspot analysis by ClusPro and FTMap. *Biological Crystallography*, *78*(6), 690-697.

Kandeel, M., Iqbal, M. N., Ali, I., Malik, S., Malik, A., & Sehgal, S. A. (2023). Comprehensive in silico analyses of flavonoids elucidating the drug properties against kidney disease by targeting AIM2. *PLoS One*, *18*(5), e0285965. https://doi.org/10.1371/journal.pone.0285965

Kim, A., Montales, K., Ruis, K., Senebandith, H., Gasparyan, H., Cowan, Q., & Michael, W. M. (2020). Biochemical analysis of TOPBP1 oligomerization. *DNA repair*, *96*, 102973.

Koonin, E. V., Altschul, S. F., & Bork, P. (1996). … Functional motifs…. *Nature Genetics*, *13*(3), 266-268.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Applied Crystallography*, *26*(2), 283-291.

Leimbacher, P.-A., Jones, S. E., Shorrocks, A.-M. K., de Marco Zompit, M., Day, M., Blaauwendraad, J.,…Fink, D. (2019). MDC1 interacts with TOPBP1 to maintain chromosomal stability during mitosis. *Molecular Cell*, *74*(3), 571-583. e578.

Manini, I., & Montomoli, E. (2018). Epidemiology and prevention of Human Papillomavirus. *Annali di Igiene*, *30*(4), 28-32.

Moody, C. A., & Laimins, L. A. (2010). Human papillomavirus oncoproteins: pathways to transformation. *Nature Reviews Cancer*, *10*(8), 550-560.

Mukherjee, A. G., Ramesh Wanjari, U., Valsala Gopalakrishnan, A., Jayaraj, R., Katturajan, R., Kannampuzha, S.,…Vellingiri, B. (2023). HPV-associated cancers: insights into the mechanistic scenario and latest updates. *Medical Oncology*, *40*(8), 212.

Nelson, C. W., & Mirabello, L. (2023). Human papillomavirus genomics: Understanding carcinogenicity. *Tumour Virus Research*, *15*, 200258.

Okunade, K. S. (2020). Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology*, *40*(5), 602-608.

Palma, J., & Pierdominici-Sottile, G. (2023). On the Uses of PCA to Characterise Molecular Dynamics Simulations of Biological Macromolecules: Basics and Tips for an Effective Use. *Chemical Physics and Physical Chemistry*, *24*(2), e202200491. https://doi.org/10.1002/cphc.202200491

Pandiyan, S., Ruan, T., Zhong, Z., Yao, M., & Wang, L. (2025). Modeling 3D structures of PIK3CA and PIK3R1 genes based on homology modeling, molecular docking, molecular dynamics and MM-GBSA study against breast cancer: Insights from an in-silico approach. *Journal of Molecular Structure*, 141580.

Post, M., Wolf, S., & Stock, G. (2019). Principal component analysis of nonequilibrium molecular dynamics simulations. *The ournal of Chemical Physice*, *150*(20), 204110. https://doi.org/10.1063/1.5089636

Prabhakar, A. T., James, C. D., Das, D., Fontan, C. T., Otoa, R., Wang, X.,…Morgan, I. M. (2022). Interaction with TopBP1 is required for human papillomavirus 16 E2 plasmid segregation/retention function during mitosis. *Journal of Virology*, *96*(16), e00830-00822.

Prabhakar, A. T., James, C. D., Fontan, C. T., Otoa, R., Wang, X., Bristol, M. L.,…Morgan, I. M. (2023). Human papillomavirus 16 E2 interaction with TopBP1 is required for E2 and viral genome stability during the viral life cycle. *Journal of Virology*, *97*(3), e00063-00023.

Rasheed, M. A., Iqbal, M. N., Saddick, S., Ali, I., Khan, F. S., Kanwal, S.,…Awais, M. (2021). Identification of Lead Compounds against Scm (fms10) in Enterococcus faecium Using Computer Aided Drug Designing. *Life (Basel)*, *11*(2). https://doi.org/10.3390/life11020077

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., & Sherman, W. (2010). Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *Journal of Chemical Theory and Computation*, *6*(5), 1509-1519. https://doi.org/10.1021/ct900587b

Wardlaw, C. P., Carr, A. M., & Oliver, A. W. (2014). TopBP1: A BRCT-scaffold protein functioning in multiple cellular pathways. *DNA Repair*, *22*, 165-174.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids Research*, *35*(suppl_2), W407-W410.

Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics*, *32*(23), 3676-3678.

Yamane, K., Kawabata, M., & Tsuruo, T. (1997). A DNA-Topoisomerase-11–Binding Protein with Eight Repeating Regions Similar to DNA-repair Enzymes and to a Cell-Cycle Regulator. *European Journal of Biochemistry*, *250*(3), 794-799.

Yan, Z., Ge, X., Wang, J., Xu, F., Li, M., Liu, X.,…Wu, C. (2022). The impact of TOPBP1 mutations in human cancers on the DNA damage response. *Genome Instability & Disease*, *3*(3), 144-162.

Yanao, T., Koon, W. S., Marsden, J. E., & Kevrekidis, I. G. (2007). Gyration-radius dynamics in structural transitions of atomic clusters. *The Journal of Chemical Physics*, *126*(12), 124102. https://doi.org/10.1063/1.2710272