TPT Trends in Pharmacology and Toxicology

OPEN ACCESS

Bioinformatics-Guided Analysis of *MLH1* Gene Variants: Structural Modeling, Dynamics, and Drug Interaction Studies

[1]*Amina Islam*, [2]*Kainat Ramzan and* [3]*Izwa Islam*

[1]*Department of Microbiology and Molecular Genetics, University of Okara, 56130, Renala Khurd, Pakistan*

[2]*Department of Molecular Biology, University of Okara, 56130, Renala Khurd, Pakistan*

[3]*Department of Biotechnology, University of Okara, 56130, Renala Khurd, Pakistan*

Corresponding Author: Kainat Ramzan (kainat.ramzan@uo.edu.pk)

ABSTRACT

Background and Objective: The *MLH1* gene encodes a key protein of the DNA mismatch repair (MMR) system responsible for maintaining genomic stability. Mutations in MLH1 can disrupt DNA repair, leading to microsatellite instability and an increased risk of hereditary cancers such as Lynch syndrome and colorectal cancer. This study aimed to identify deleterious missense variants of the *MLH1* gene and evaluate their structural, dynamic, and therapeutic implications using integrated bioinformatics analysis, Normal Mode Analysis (iMODS), and molecular docking studies. **Materials and Methods:** Missense variants were screened using twelve *in silico* prediction tools to identify potentially damaging nsSNPs. Mutation3D was used to analyze the spatial clustering of variants. A full-length three-dimensional structure of MLH1 (756 amino acids) was predicted using AlphaFold and validated using QMEAN and ProSA scores. Structural dynamics were examined using iMODs server. Molecular docking was performed in PyRx to assess interactions between wild-type and mutant MLH1 proteins and selected FDA-approved anticancer drugs. **Results:** Twenty-six nsSNPs were predicted to be highly deleterious, with eight variants located in structurally important regions. Structural validation of the MLH1 model yielded a QMEAN Z-score of -4.17 and a ProSA Z-score of -7.15. Four variants (L11V, R27Q, I115T, and E319K) showed notable structural deviations from the wild-type protein. Normal Mode Analysis indicated that the L11V variant exhibited slightly increased rigidity compared with the wild type, while R27Q, I115T, and E319K showed marginally increased flexibility. Docking analysis revealed stable interactions between MLH1 proteins and several anticancer drugs, including Cabozantinib, Irinotecan, Crizotinib, and Olaparib, although mutations altered binding orientations and interaction patterns. **Conclusion:** The findings highlight the structural and dynamic effects of MLH1 missense mutations and their potential influence on therapeutic interactions. This study provides insights into MLH1 variant pathogenicity and may support future experimental validation and targeted therapeutic strategies for colorectal cancer.

KEYWORDS

AlphaFold modeling, CRC, DNA mismatch repair, FDA-approved anticancer drugs, Lynch syndrome, *MLH1* gene, molecular docking, nsSNPs

ISSN: 2789-9446 (Online)
ISSN: 2789-9438 (Print)
https://doi.org/10.21124/tpt.2026.141.164

Received: 13 Apr. 2026
Accepted: 24 Jul. 2026
Published: 30 Sep. 2026

INTRODUCTION

Colorectal cancer (CRC) is a malignant neoplasm of the colon or rectum characterized by uncontrolled cellular proliferation, frequently arising from precancerous polyps. Clinically, CRC manifests with a range of symptoms, including alterations in bowel habits, rectal bleeding, abdominal discomfort, and unintended weight loss[1]. Globally, CRC represents a significant public health burden, ranking as the third most commonly diagnosed cancer and the 3rd leading cause of cancer-related mortality[2]. In 2020, more than 1.9 million new CRC cases and over 930,000 deaths were reported[3]. Notably, early-onset CRC is increasing, particularly in New Zealand, Chile, and England, a trend attributed to unhealthy dietary patterns, physical inactivity, and rising obesity prevalence[4]. According to the National Cancer Registry of Pakistan, CRC accounted for approximately 4.9% of all cancer cases between 2015 and 2019, ranking the most prevalent cancers in the country[5].

Moreover, CRC develops through a multifactorial process involving both genetic susceptibility and environmental influences. Inherited disorders such as Lynch syndrome (LS) and familial adenomatous polyposis (FAP) markedly increase the likelihood of developing CRC by impairing essential DNA repair systems and tumor-suppressing mechanisms. Lynch syndrome, also referred to as hereditary nonpolyposis colorectal cancer (HNPCC), is primarily caused by alterations in DNA mismatch repair (MMR) genes, including MLH1, MSH2, MSH6, and PMS2[6,7].

In LS, defective MMR leads to the accumulation of DNA replication errors, significantly increasing the CRC cases and other malignancies, such as endometrial and gastric cancers[8]. To date, more than 3,000 distinct MLH1 variants have been documented, including missense mutations (Leu749Pro, Gly67Arg), splice-site mutations (c.588+5G>A), and frameshift mutations (c.1852_1853delAG). Germline mutations, such as c.677G>A (Val226Ile), destabilize the protein, reduce DNA repair efficiency, and increase the risk of colorectal cancer[9].

In sporadic colorectal cancers, MLH1 promoter hypermethylation is a common mechanism of gene silencing, further reducing MMR efficiency and promoting tumorigenesis. In Lynch syndrome, MLH1 mutations account for approximately 25-40% of all identified mismatch repair gene mutations, making it one of the leading genetic contributors to hereditary nonpolyposis colorectal cancer[10]. The *MLH1* gene is located on human chromosome 3p22.2 and spans about 57 kilobases, comprising 19 exons that encode the full-length protein. Exon 1 contains the untranslated region (UTR), while exons 2-19 encode the functional domains of the protein. The MLH1 protein consists of 756 amino acids and forms a heterodimer with PMS2, generating the MutLα complex, which coordinates the repair of base-base mismatches and insertion-deletion loops. The N-terminal ATPase domain (residues 1-340) mediates ATP binding and hydrolysis, providing the energy required for conformational changes during DNA repair. The C-terminal domain (residues 492-756) is responsible for dimerization with PMS2, a critical step for MutLα complex formation and recruitment to the mismatch site[11].

Mutations in the ATPase or dimerization domains can severely compromise the protein's function, resulting in MMR deficiency and genomic instability. MLH1 is associated with germline and somatic alterations, which can disrupt transcription or protein activity[12]. This study aimed to identify deleterious missense variants of the *MLH1* gene using integrated *in silico* approaches. Functional nsSNPs were evaluated for their disease association, effects on protein stability, and structural flexibility. Three-dimensional modeling and refinement were performed to assess mutation-induced structural changes, while protein dynamics were examined using Normal Mode Analysis through the iMODs server. In addition, potential small-molecule ligands were retrieved from the ZINC and PubChem databases, and protein-ligand interactions were analyzed using molecular docking and Discovery Studio to explore possible therapeutic targets.

MATERIALS AND METHODS

Datasets screening: The reference sequence of the human *MLH1* gene was retrieved from the National Center for Biotechnology Information database, and corresponding protein information was obtained from UniProt[13]. Human SNP data were retrieved from the dbSNP database of the National Center for Biotechnology Information (accessed on 2 October 2025). For downstream analysis, only non-synonymous (missense) SNPs with available protein-level information were selected. These datasets provided the foundation for computational screening and the identification of potentially deleterious variants (Table 1).

SNPs annotation and disease analysis: SNPnexus was used as an integrated platform incorporating multiple prediction tools, including SIFT and PolyPhen. PolyPhen-2, ranging from 0 (benign) to 1 (probably damaging), was used to assess the pathogenicity of variants based on sequence conservation and structural features. PANTHER was employed to analyse evolutionary conservation and Gene Ontology-based functional annotations, supporting SNP pathogenicity prediction. In addition, PredictSNP was used to generate confidence-weighted consensus predictions by integrating multiple algorithms. The potential disease association of the variants was evaluated using PhD-SNP, SNP & GO. SNP & GO combines Gene Ontology annotations with a probability score between 0 and 1; variants with a score>0.5 are considered likely disease-associated variants. SUSPECT was employed to assess the potential impact of these missense mutations on protein stability[14,15].

Analysis of MLH1 stability: The impact of nsSNPs on MLH1 protein stability was analysed using multiple computational tools. MuPro predicts stability changes based on machine learning models, where negative ΔΔG values indicate destabilization. The i-Stable applies a meta-prediction approach to assess sequence-based stability alterations. For more rigorous predictions, DUET was employed[16].

STRUCTURE-BASED ALGORITHMS

Swiss modeling: The 3D structure of the wild-type (WT) MLH1 protein was generated using the SWISS-MODEL server The FASTA sequence of MLH1 was submitted, templates were selected based on sequence

Table 1:Computational workflow and tools employed for variant and structural analysis

Analysis type	Tool/Database	Primary application
Variant data acquisition	NCBI dbSNP	Retrieval of genetic variants
	UniProt	Protein sequence and functional annotation
Functional impact assessment	SNPnexus	Variant annotation and functional inference
	PolyPhen-2	Prediction of structural and functional effects
	PANTHER-PSEP	Evolutionary preservation scoring
	PredictSNP	Consensus-based variant effect prediction
Disease association prediction	SNP & GO	Gene Ontology-based disease scoring
	PhD-SNP	Machine learning-based disease classification
	SUSPECT	Functional and structural impact evaluation
Protein stability and dynamics	MuPro	ΔΔG-based stability prediction
	iStable 2.0	Meta-prediction of stability alterations
	DynaMut2	Stability and flexibility analysis
	DUET	Integrated FoldX-mCSM stability estimation
Structural modeling	SWISS-MODEL	Homology-based 3D structure generation
Model refinement	GalaxyRefine	Atomic-level structural optimization
Structure quality validation	SAVES Server	Stereochemical and interaction validation
Model quality evaluation	QMEAN	Global and local quality scoring
	ProSA-web	Z-score-based structural assessment
Structural comparison	TM-align	Structural alignment and RMSD calculation
Mutation clustering analysis	Mutation3D	Identification of mutational hotspots
Protein dynamics analysis	iMODS	Normal Mode Analysis for protein flexibility and stability
Ligand library preparation	PubChem	Small-molecule compound retrieval
Molecular docking	PyRx (AutoDock Vina)	Protein–ligand interaction analysis
Interaction visualization	Discovery Studio Visualizer	2D/3D interaction profiling
	Ligplot+	

similarity, and the best-fitting models were constructed. The resulting structures were visualized in PyMOL, with site-specific amino acid substitutions incorporated. Atomic-level optimization and minimization of steric clashes were performed using GalaxyRefine[13,17].

Template validation: Model quality was evaluated using the SAVESv6.1 server, including PROCHECK, ERRAT, and VERIFY-3D modules. Ramachandran plots generated by PROCHECK assessed torsional angles, with >90% of residues expected in favorable regions; the MLH1 models showed 95.3% of residues in these regions. ERRAT evaluated non-bonded interactions, while VERIFY-3D compared the 3D models against a high-quality protein database for consistency[18]. Model quality was further assessed using QMEAN, where higher scores indicate greater similarity to experimentally resolved structures. ProSA-web was used to calculate Z-scores, providing an overall quality estimate compared with native proteins[14,17].

Structural alteration analysis: Structural alignment of WT and mutant proteins was performed using TM-align. Template Modeling (TM) scores and Root-Mean-Square Deviation (RMSD) values were calculated, with TM scores near 1 indicating high structural similarity, while higher RMSD values reflected significant conformational change[15]. The spatial distribution of amino acid substitutions in MLH1 was analysed using Mutation3D. This tool allows identification of clustered or hotspot mutations within experimental or modeled protein structures, providing insight into potential effects on folding and function[19].

INTERACTION ANALYSIS AND VISUALIZATION

The toxicity profiles of the selected compounds were evaluated using the ProTox-II web server[20]. Molecular docking studies were conducted using the PyRx platform, which integrates AutoDock and AutoDock Vina algorithms. Protein and ligand structures were converted into PDBQT format, and grid boxes were defined to encompass the active site region. Docking simulations were carried out using the Lamarckian Genetic Algorithm (LGA), and the most favorable binding poses were selected based on minimum binding energy values. The docked protein-ligand complexes were evaluated using Discovery Studio Visualizer and LigPlot+. Interaction profiles were generated in both 2D and 3D formats to identify hydrogen bonds, hydrophobic interactions, and the key amino acid residues contributing to ligand binding within the active site[13,14].

NORMAL MODE ANALYSIS OF MLH1 STRUCTURAL DYNAMICS

To evaluate the structural flexibility and dynamic behavior of the protein encoded by the *MLH1* gene, normal mode analysis (NMA) was performed using the iMODS server. The analysis was conducted with default parameters. This approach provides insight into protein stability, flexibility, and deformation patterns, allowing assessment of how specific amino acid substitutions may influence the structural behavior of both wild-type and mutant MLH1 proteins[18].

RESULTS

Functional assessment of MLH1 variants: The analysis of the *MLH1* gene in the NCBI dbSNP database identified 23,356 SNPs. These included 1,657 synonymous variants and 5,167 missense (non-synonymous) variants. This study found 15,754 were located in coding regions and 1,154 in non-coding regions. Genomic region distribution showed 8,930 intronic SNPs, 529 exonic SNPs, and 2,615 SNPs within untranslated regions, comprising 1,366 variants in the 3-UTR and 1,249 in the 5-UTR. In addition, 1,092 SNPs were identified in the 3 downstream region and 2,741 in the 5 upstream region (Fig. 1). Functional annotation of missense SNPs using SNPnexus and SIFT predicted 2,962 variants as deleterious, 193 as low-confidence deleterious, 2,144 as tolerated, and 161 as low-confidence tolerated (Fig. 2a). PolyPhen analysis classified 1,934 variants as probably damaging, 927 as possibly damaging, and 2,574 as benign (Fig. 2b). Comparative analysis of SIFT and PolyPhen predictions identified 76 common ns SNPs consistently classified as deleterious or probably damaging (Table 2).

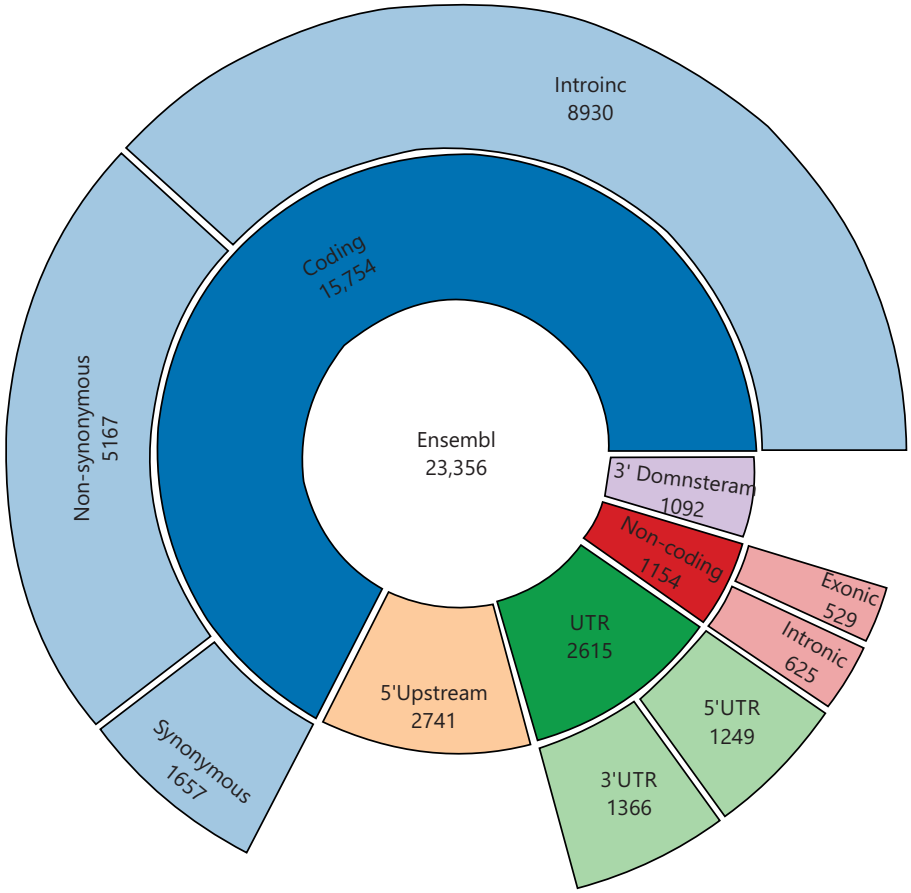

Fig. 1: A circular representation illustrates the SNP distribution identified in the *MLH1* gene across different genomic regions

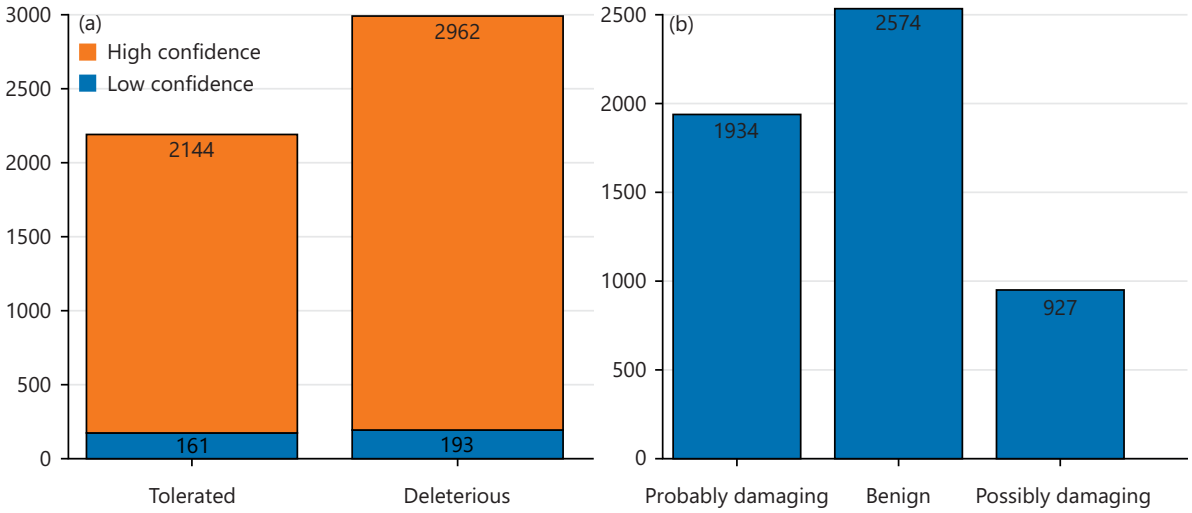

Fig. 2(a-b): Prediction of deleterious variations by (a) SIFT and (b) Polyphen
Y-axis: Number of Variants (nsSNPs)

PANTHER analysis predicted three functional outcome categories for the 76 nsSNPs, such as probably damaging, possibly damaging, and benign. Of these, 74 nsSNPs were classified as probably damaging, while I219F and I115T were assigned a score of 0.27 (Table 2). PolyPhen-2 analysis categorized 68 nsSNPs as probably damaging, whereas I513T, S332F, D737V, A31S, L135R, H112R, and V16L were classified as possibly damaging. In addition, A129T (rs63750866) was predicted to be benign by PolyPhen-2 (Table 2). PredictSNP analysis indicated deleterious effects for all 76 nsSNPs on protein function, as summarized in Table 2.

Table 2: Computational analysis of high-risk missense variants using diverse functional prediction tools

			SIFT		Polyphen		Panther		Predict SNP		Polyphen 2	
Variation ID	Variant	AA	Eff	Sc	Eff	Sc	Eff	Sc	Eff	Sc (%)	Eff	Sc
rs1193754562	C/G	R10G	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	76	Probably damaging	1
rs1224959447	C/G	P747R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	55	Probably damaging	0.997
rs1312172811	T/G	C672W	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs1315572872	G/A	G54R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs1366752604	G/T	G6W	Deleterious	0	Probably damaging	0.961	Probably damaging	0.74	Deleterious	76	Probably damaging	1
rs138705565	G/A	R27Q	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs1399177067	A/G	I94V	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	63	Probably damaging	0.972
rs1413395234	T/G	Y625D	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	76	Probably damaging	1
rs141688321	T/C	I513T	Deleterious	0	Probably damaging	0.961	Probably damaging	0.74	Deleterious	74	Possibly damaging	0.69
rs1430103195	T/G	F88V	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	75	Probably damaging	1
rs1443882824	C/T	T386I	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	55	Probably damaging	1
rs147342421	A/C	Q48H	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	60	Probably damaging	0.998
rs1482654951	G/C	D304H	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	1
rs151119913	C/T	H315Y	Deleterious	0	Probably damaging	0.961	Probably damaging	0.86	Deleterious	63	Probably damaging	0.999
rs1799977	A/T	I219F	Deleterious	0	Probably damaging	0.961	Probably benign	0.27	Deleterious	62	Probably damaging	0.967
rs182963667	T/G	F88L	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	75	Probably damaging	0.998
rs191257018	C/T	S332F	Deleterious	0	Probably damaging	0.961	Probably damaging	0.74	Deleterious	60	Possibly damaging	0.544
rs200076893	T/A	Y126N	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	51	Probably damaging	0.999
rs267607706	C/G	N38K	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	1
rs267607725	G/A	G98S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	0.999
rs267607808	C/T	P309S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	62	Probably damaging	1
rs267607885	A/T	D737V	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	87	Possibly damaging	0.942
rs267607886	A/G	Y684C	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	76	Probably damaging	1
rs35001569	A/G	K618E	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	65	Probably damaging	0.998
rs35045067	A/G	Y548C	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs367654552	C/G	R18G	Deleterious	0	Probably damaging	0.961	Probably damaging	0.86	Deleterious	87	Probably damaging	1
rs377241633	T/C	F626L	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	60	Probably damaging	1
rs41295280	G/C	G22A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	0.994
rs41295284	T/A	L607H	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	72	Probably damaging	1
rs550914672	G/A	E331K	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	68	Probably damaging	0.983
rs56185292	C/G	S577W	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	61	Probably damaging	0.994

Table 2: Continue

			SIFT		Polyphen		Panther		Predict SNP		Polyphen 2	
Variation ID	Variant	AA	Eff	Sc	Eff	Sc	Eff	Sc	Eff	Sc (%)	Eff	Sc
rs63750555	A/T	E23D	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	87	Probably damaging	0.999
rs63750641	A/G	K84E	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	1
rs63750650	A/G	E268G	Deleterious	0	Probably damaging	0.961	Probably damaging	0.57	Deleterious	60	Probably damaging	0.999
rs63750726	C/T	P654L	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	61	Probably damaging	1
rs63750760	C/T	R385C	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs63750792	C/T	P28L	Deleterious	0	Probably Damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs63750796	G/A	E319K	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	0.994
rs63750866	G/A	A120T	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	61	Benign	0.001
rs63750877	A/C	K57Q	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	63	Probably damaging	1
rs63750899	C/T	P648S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs63750952	A/G	N64S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs63751109	C/T	S44F	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	87	Probably damaging	0.999
rs63751275	C/T	R687W	Deleterious	0	Probably damaging	0.961	Probably damaging	0.74	Deleterious	76	Probably damaging	1
rs63751597	C/T	H264Y	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	87	Probably damaging	0.997
rs63751715	G/C	Q346H	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	72	Probably damaging	1
rs730882127	C/G	A31G	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	76	Probably damaging	1
rs746206527	C/A	T310K	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs748417604	G/A	V152M	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	76	Probably damaging	1
rs749671520	G/T	A31S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	77	Possibly damaging	0.924
rs752850761	G/T	R79S	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs753233578	T/G	L135R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	51	Possibly damaging	0.862
rs756398627	C/T	R27W	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs759287108	C/G	T45R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs764120517	T/C	I115T	Deleterious	0	Probably damaging	0.961	Probably benign	0.27	Deleterious	61	Probably damaging	1
rs770276731	A/G	T96A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	1
rs773647920	A/G	Y97C	Deleterious	0	Probably damaging	0.961	Probably damaging	0.86	Deleterious	87	Probably damaging	1
rs775914775	A/G	H112R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	74	Possibly damaging	0.909
rs776643257	G/C	V16L	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	65	Possibly damaging	0.867
rs777054430	T/C	Y750H	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	72	Probably damaging	1
rs779581111	A/G	T310A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	63	Probably damaging	0.998
rs780141938	C/G	I68M	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1

Table 2: Continue

			SIFT		Polyphen		Panther		Predict SNP		Polyphen 2	
Variation ID	Variant	AA	Eff	Sc	Eff	Sc	Eff	Sc	Eff	Sc (%)	Eff	Sc
rs780232692	T/C	C142R	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs786202396	G/T	G634V	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	72	Probably damaging	1
rs786203362	A/G	R79G	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs786203583	C/A	L741M	Deleterious	0	Probably damaging	0.961	Probably damaging	0.85	Deleterious	75	Probably damaging	0.983
rs786203745	A/G	T81A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	65	Probably damaging	0.996
rs863224637	T/G	N30K	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs863224638	T/C	V307A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs863224639	G/T	V314F	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs864622396	T/C	V15A	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs864622485	G/C	E37D	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	76	Probably damaging	1
rs864622596	C/G	L11V	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs876659860	A/C	K196T	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	61	Probably damaging	1
rs876660301	G/A	V15M	Deleterious	0	Probably damaging	0.961	Probably damaging	0.89	Deleterious	87	Probably damaging	1
rs876660775	G/T	K311N	Deleterious	0	Probably damaging	0.961	Probably damaging	0.95	Deleterious	87	Probably damaging	1

*Eff: Effect and Sc: Scorz

Table 3: Prediction of disease-associated nsSNPs using SNPS&GO, SusPect, and PhD-SNP

		SNP and GO		SuSPect		PhD-SNP	
Variation ID	A.A	Effect	Score	Score	Prediction	Effect	Score (%)
rs1193754562	R10G	Disease	1	54	Disease	Disease	73
rs1224959447	P747R	Disease	0	35	Disease	Disease	68
rs1312172811	C672W	Disease	7	96	Disease	Disease	88
rs1315572872	G54R	Disease	0	57	Disease	Disease	88
rs1366752604	G6W	Neutral	4	52	Disease	Disease	77
rs138705565	R27Q	Disease	1	74	Disease	Disease	88
rs1399177067	I94V	Neutral	8	65	Disease	Disease	89
rs1413395234	Y625D	Neutral	1	93	Disease	Disease	88
rs141688321	I513T	Neutral	3	38	Disease	Disease	58
rs1430103195	F88V	Disease	2	67	Disease	Disease	55
rs1443882824	T386I	Neutral	1	49	Disease	Disease	86
rs147342421	Q48H	Neutral	7	73	Disease	Disease	66
rs1482654951	D304H	Disease	5	97	Disease	Disease	86
rs151119913	H315Y	Neutral	1	45	Disease	Disease	66
rs1799977	I219F	Disease	3	86	Disease	Disease	66
rs182963667	F88L	Neutral	1	70	Disease	Disease	66
rs191257018	S332F	Neutral	5	56	Disease	Disease	66
rs200076893	Y126N	Disease	6	77	Disease	Disease	68
rs267607706	N38K	Disease	4	99	Disease	Disease	88
rs267607725	G98S	Disease	6	96	Disease	Disease	82
rs267607808	P309S	Disease	4	94	Disease	Disease	55
rs267607885	D737V	Neutral	1	50	Disease	Disease	82
rs267607886	Y684C	Disease	5	93	Disease	Disease	88
rs35001569	K618E	Neutral	3	71	Disease	Disease	61
rs35045067	Y548C	Disease	3	88	Disease	Disease	82
rs367654552	R18G	Disease	1	87	Disease	Disease	86
rs377241633	F626L	Disease	1	84	Disease	Disease	86
rs41295280	G22A	Neutral	1	81	Disease	Disease	77
rs41295284	L607H	Disease	0	79	Disease	Disease	68
rs550914672	E331K	Neutral	6	42	Disease	Disease	78
rs56185292	S577W	Disease	2	45	Disease	Disease	82
rs63750555	E23D	Neutral	1	93	Disease	Disease	88
rs63750641	K84E	Disease	7	95	Disease	Disease	88
rs63750650	E268G	Neutral	1	59	Disease	Disease	66
rs63750726	P654L	Disease	7	96	Disease	Disease	88
rs63750760	R385C	Disease	3	63	Disease	Disease	86
rs63750792	P28L	Neutral	3	87	Disease	Disease	86
rs63750796	E319K	Neutral	4	74	Disease	Disease	77
rs63750866	A120T	Neutral	3	30	Disease	Disease	59
rs63750877	K57Q	Neutral	8	55	Disease	Disease	58
rs63750899	P648S	Disease	5	93	Disease	Disease	86
rs63750952	N64S	Neutral	1	86	Disease	Disease	86
rs63751109	S44F	Disease	0	97	Disease	Disease	88
rs63751275	R687W	Neutral	3	13	Neutral	Disease	45
rs63751597	H264Y	Neutral	0	77	Disease	Disease	58
rs63751715	Q346H	Neutral	2	48	Disease	Disease	73
rs730882127	A31G	Neutral	3	96	Disease	Disease	58
rs746206527	T310K	Disease	4	77	Disease	Disease	68
rs748417604	V152M	Disease	1	88	Disease	Disease	86
rs749671520	A31S	Neutral	1	95	Disease	Disease	76
rs752850761	R79S	Disease	5	84	Disease	Disease	82
rs753233578	L135R	Neutral	4	58	Disease	Disease	73
rs756398627	R27W	Disease	3	90	Disease	Disease	88
rs759287108	T45R	Neutral	2	75	Disease	Disease	82
rs764120517	I115T	Disease	2	95	Disease	Disease	51
rs770276731	T96A	Disease	0	91	Disease	Disease	73
rs773647920	Y97C	Disease	5	86	Disease	Disease	77

Table 3: Continue

		SNP and GO		SuSPect		PhD-SNP	
Variation ID	A.A	Effect	Score	Score	Prediction	Effect	Score (%)
rs775914775	H112R	Neutral	7	39	Disease	Disease	55
rs776643257	V16L	Neutral	2	89	Disease	Disease	86
rs777054430	Y750H	Disease	4	92	Disease	Disease	88
rs779581111	T310A	Neutral	6	46	Disease	Disease	78
rs780141938	I68M	Neutral	7	77	Disease	Disease	59
rs780232692	C142R	Disease	2	79	Disease	Disease	77
rs786202396	G634V	Disease	1	79	Disease	Disease	88
rs786203362	R79G	Disease	6	90	Disease	Disease	86
rs786203583	L741M	Neutral	7	69	Disease	Disease	72
rs786203745	T81A	Neutral	6	62	Disease	Disease	89
rs863224637	N30K	Neutral	0	78	Disease	Disease	86
rs863224638	V307A	Disease	2	92	Disease	Disease	88
rs863224639	V314F	Disease	5	96	Disease	Disease	82
rs864622396	V15A	Neutral	4	86	Disease	Disease	73
rs864622485	E37D	Neutral	4	73	Disease	Disease	88
rs864622596	L11V	Disease	0	90	Disease	Disease	86
rs876659860	K196T	Disease	2	75	Disease	Disease	51
rs876660301	V15M	Neutral	5	89	Disease	Disease	73
rs876660775	K311N	Disease	4	96	Disease	Disease	77

Table 4: Integrated *in silico* evaluation of variants affecting protein structural stability

		i-Stable		MUpro		Duet	
Variation ID	A.A	Effect	Score	Effect	Score	Effect	Score
rs1193754562	R10G	Decrease	0.775588	Decrease	-1	Stabilising	0.138
rs1224959447	P747R	Decrease	0.711756	Decrease	-1	Destabilising	-0.144
rs1312172811	C672W	Decrease	0.746844	Decrease	-0.09647	Destabilising	-1.565
rs1315572872	G54R	Increase	0.611656	Increase	0.131007	Destabilising	-1.372
rs1366752604	G6W	Increase	0.558945	Increase	0.24325	Destabilising	-1.093
rs138705565	R27Q	Decrease	0.645462	Decrease	-1	Destabilising	-0.415
rs1399177067	I94V	Decrease	0.646576	Decrease	-0.80259	Destabilising	-1.214
rs1413395234	Y625D	Decrease	0.599322	Decrease	-0.44977	Destabilising	-1.469
rs141688321	I513T	Decrease	0.671393	Decrease	-1	Destabilising	-2.999
rs1430103195	F88V	Increase	0.665845	Increase	0.300372	Destabilising	-0.267
rs1443882824	T386I	Increase	0.696248	Increase	0.039758	Stabilising	0.405
rs147342421	Q48H	Decrease	0.738391	Decrease	-0.32425	Destabilising	-0.478
rs1482654951	D304H	Decrease	0.738391	Decrease	-0.32425	Destabilising	-0.249
rs151119913	H315Y	Increase	0.641638	Increase	0.227079	Stabilising	1.259
rs1799977	I219F	Decrease	0.697505	Decrease	-1	Destabilising	-1.087
rs182963667	F88L	Increase	0.683275	Increase	0.250489	Destabilising	-0.148
rs191257018	S332F	Decrease	0.637513	Decrease	-0.19562	Destabilising	-0.448
rs200076893	Y126N	Decrease	0.842354	Decrease	-1	Destabilising	-1.663
rs267607706	N38K	Decrease	0.705164	Decrease	-0.29669	Stabilising	0.079
rs267607725	G98S	Decrease	0.834635	Decrease	-0.87681	Destabilising	-1.338
rs267607808	P309S	Decrease	0.817149	Decrease	-0.59174	Destabilising	-0.186
rs267607885	D737V	Decrease	0.691426	Decrease	-0.36223	Stabilising	0.254
rs267607886	Y684C	Decrease	0.674578	Decrease	-0.28919	Destabilising	-0.459
rs35001569	K618E	Increase	0.569278	Increase	0.229491	Destabilising	-1.256
rs35045067	Y548C	Decrease	0.693891	Decrease	-1	Destabilising	-1.812
rs367654552	R18G	Decrease	0.692444	Decrease	-1	Destabilising	-1.558
rs377241633	F626L	Decrease	0.791955	Decrease	-0.17678	Destabilising	-0.392
rs41295280	G22A	Increase	0.631844	Increase	0.429184	Destabilising	-0.098
rs41295284	L607H	Decrease	0.707733	Decrease	-0.62096	Destabilising	-1.565
rs550914672	E331K	Decrease	0.713918	Decrease	-0.4841	Stabilising	0.008
rs56185292	S577W	Decrease	0.555761	Decrease	-0.11195	Destabilising	-0.422
rs63750555	E23D	Increase	0.631876	Increase	0.206299	Destabilising	-0.722

Table 4: Continue

		i-Stable		MUpro		Duet	
Variation ID	A.A	Effect	Score	Effect	Score	Effect	Score
rs63750641	K84E	Increase	0.627028	Increase	0.503611	Stabilising	0.66
rs63750650	E268G	Decrease	0.704084	Decrease	-1	Destabilising	-0.705
rs63750726	P654L	Increase	0.63126	Increase	0.193497	Stabilising	0.322
rs63750760	R385C	Decrease	0.738606	Decrease	-0.16808	Destabilising	-0.436
rs63750792	P28L	Decrease	0.738606	Decrease	-0.16808	Stabilising	0.513
rs63750796	E319K	Decrease	0.739107	Decrease	-0.12521	Destabilising	-0.287
rs63750866	A120T	Decrease	0.693477	Decrease	-0.55338	Destabilising	-0.042
rs63750877	K57Q	Decrease	0.693477	Decrease	-0.55338	Destabilising	-0.364
rs63750899	P648S	Decrease	0.83609	Decrease	-1	Destabilising	-2.312
rs63750952	N64S	Decrease	0.67731	Decrease	-0.02465	Stabilising	0.052
rs63751109	S44F	Decrease	0.526191	Increase	0.361924	Destabilising	-1.058
rs63751275	R687W	Decrease	0.741712	Decrease	-0.98293	Destabilising	-0.595
rs63751597	H264Y	Increase	0.699031	Increase	0.014046	Stabilising	1.178
rs63751715	Q346H	Increase	0.588122	Increase	0.610292	Stabilising	0.261
rs730882127	A31G	Decrease	0.822508	Decrease	-0.31982	Destabilising	-1.949
rs746206527	T310K	Increase	0.649609	Increase	0.109074	Destabilising	-1.108
rs748417604	V152M	Decrease	0.784881	Decrease	-0.69358	Destabilising	-1.587
rs749671520	A31S	Decrease	0.8087	Decrease	-0.59011	Destabilising	-2.477
rs752850761	R79S	Decrease	0.748267	Decrease	-0.94111	Destabilising	-0.772
rs753233578	L135R	Decrease	0.737886	Decrease	-1	Destabilising	-0.677
rs756398627	R27W	Decrease	0.712872	Decrease	-0.60613	Destabilising	-0.109
rs759287108	T45R	Increase	0.755264	Increase	0.349333	Stabilising	0.205
rs764120517	I115T	Decrease	0.673349	Decrease	-1	Destabilising	-3.215
rs770276731	T96A	Increase	0.507296	Decrease	-0.82159	Destabilising	-0.442
rs773647920	Y97C	Decrease	0.668154	Decrease	-0.0317	Destabilising	-1.549
rs775914775	H112R	Increase	0.6027	Increase	0.37764	Destabilising	-0.641
rs776643257	V16L	Increase	0.554638	Increase	0.351501	Destabilising	-0.309
rs777054430	Y750H	Decrease	0.621652	Decrease	-1	Destabilising	-0.763
rs779581111	T310A	Decrease	0.703168	Decrease	-0.1827	Destabilising	-0.559
rs780141938	I68M	Decrease	0.797461	Decrease	-1	Destabilising	-0.849
rs780232692	C142R	Decrease	0.751534	Decrease	-0.3283	Destabilising	-0.007
rs786202396	G634V	Increase	0.530842	Increase	0.03186	Destabilising	-1.059
rs786203362	R79G	Decrease	0.644622	Decrease	-1	Destabilising	-0.714
rs786203583	L741M	Decrease	0.761662	Decrease	-0.15319	Stabilising	-0.437
rs786203745	T81A	Decrease	0.644493	Decrease	-1	Destabilising	-0.39
rs863224637	N30K	Decrease	0.583433	Decrease	-0.44567	Stabilising	0.002
rs863224638	V307A	Decrease	0.66852	Decrease	-1	Destabilising	-1.098
rs863224639	V314F	Decrease	0.705583	Decrease	-1	Destabilising	-1.46
rs864622396	V15A	Decrease	0.794747	Decrease	-0.68189	Destabilising	-0.602
rs864622485	E37D	Decrease	0.714411	Decrease	-0.32641	Destabilising	-2.053
rs864622596	L11V	Decrease	0.786045	Decrease	-1	Destabilising	-0.103
rs876659860	K196T	Decrease	0.850422	Decrease	-0.69323	Destabilising	-0.989
rs876660301	V15M	Decrease	0.850887	Decrease	-0.2174	Destabilising	-0.194
rs876660775	K311N	Decrease	0.735552	Decrease	-0.23809	Destabilising	-0.797

Estimation of disease-associated nsSNPs: According to SNP & Go, 42 nsSNPs were predicted as disease-associated, while 35 were classified as neutral, as shown in Table 3. PhD-SNP indicating that 76 nsSNPs are disease-causing, with predicted probabilities ranging from 51% to 89%, as presented in Table 3. Variants such as C672W (rs1312172811), G54R (rs1315572872), R27Q (rs138705565), and Y684C (rs267607886) showed the highest disease probabilities (≥88%), highlighting their potential pathogenic significance. SuSPect predicted that all 77 nsSNPs are disease-associated, with scores ranging from 13 to 99. The algorithm indicated a high likelihood of functional impact for most variants, with particularly high scores observed for C672W (rs1312172811, score 96), N38K (rs267607706, score 99), and P654L (rs63750726, score 96). Only R687W (rs63751275) showed a low score of 13 and was classified as neutral, as shown in Table 3.

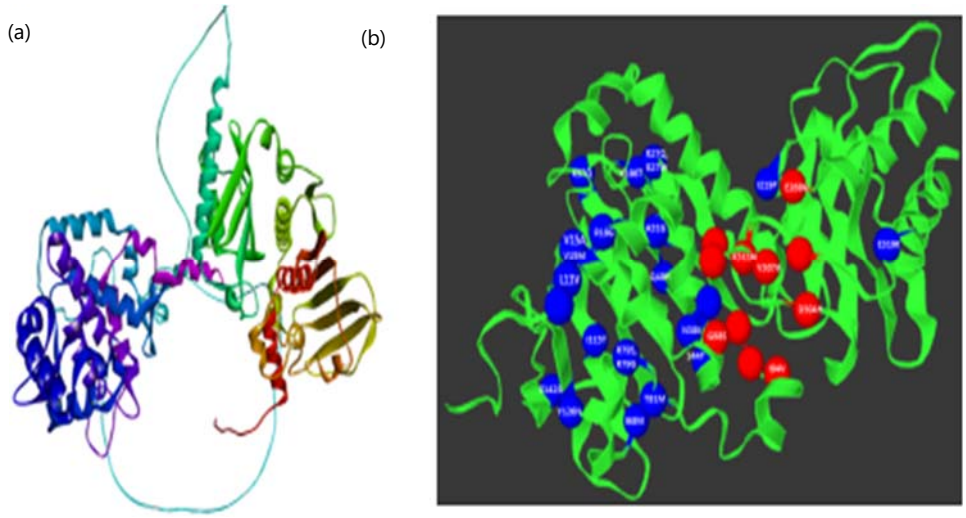

Fig. 3(a-b): (a) Three-dimensional structure of the MLH1 protein and (b) Predicted structural impact of the identified nsSNPs on MLH1

Table 5: Functional prediction and visualization of the *MLH1* gene by mutation 3D

rs ID	Model	Prediction	rs ID	Model	Prediction
rs138705565	R27Q	Clustered mutation	rs876660775	K311N	Covered mutation
rs1399177067	I94V	Covered mutation	rs752850761	R79S	Covered mutation
rs147342421	Q48H	Covered mutation	rs756398627	R27W	Clustered mutation
rs1482654951	D304H	Covered mutation	rs764120517	I115T	Covered mutation
rs1799977	I219F	Covered mutation	rs779581111	T310A	Covered mutation
rs200076893	Y126N	Covered mutation	rs780141938	I68M	Covered mutation
rs267607706	N38K	Covered mutation	rs780232692	C142R	Covered mutation
rs267607725	G98S	Covered mutation	rs786203362	R79G	Clustered mutation
rs367654552	R18G	Covered mutation	rs864622396	V15A	Clustered mutation
rs63750650	E268G	Covered mutation	rs864622596	L11V	Clustered mutation
rs63750796	E319K	Covered mutation	rs876659860	K196T	Covered mutation
rs63750877	K57Q	Clustered mutation	rs876660301	V15M	Clustered mutation
rs63751109	S44F	Covered mutation	rs730882127	A31G	Clustered mutation

Evaluation of protein stability: We assessed the impact of nsSNPs on protein stability using three computational tools. The i-Stable analysis indicated that 19 nsSNPs could increase the stability of the mutant protein, while 57 nsSNPs were predicted to reduce stability, which may negatively affect protein function (Table 4). MUpro results were similar, showing 57 nsSNPs as destabilizing and 19 as stabilizing (Table 4). DUET identified 61 nsSNPs as destabilizing and 15 as stabilizing, demonstrating some differences between the tools (Table 4). By comparing all prediction tools, we identified 26 nsSNPs as highly deleterious, which were chosen for further detailed analysis.

Prediction of mutation clusters: The MLH1 protein was analysed using the Mutation3D web server with all 26 selected nsSNPs as input. Among these, eight missense mutations formed clusters, while the remaining 18 were classified as covered mutations (Table 5). For structural analysis, a 3D model of MLH1 was generated using homology modeling, based on 23 templates obtained from the RCSB PDB database (Fig. 3a). These models, however, did not span the full length of the protein, which consists of 756 amino acids (UniProt ID: P40692). The clustered mutations were R27Q, K57Q, A31G, R27W, R79G, V15A, L11V, and V15M, whereas the covered mutations included I94V, Q48H, D304H, I219F, Y126N, N38K, G98S, R18G, E268G, E319K, S44F, R79S, I115T, T310A, I68M, C142R, K196T, and K311N (Fig. 3b).

SWISS modeling of MLH1 protein: Structural modeling of the human MLH1 protein was performed to evaluate the impact of selected missense variants on protein conformation. A total of 26 non-synonymous substitutions (R27Q, I94V, Q48H, D304H, I219F, Y126N, N38K, G98S, R18G, E268G, E319K, K57Q, S44F, A31G, R79S, R27W, I115T, T310A, I68M, C142R, R79G, V15A, L11V, K196T, V15M, and K311N) were

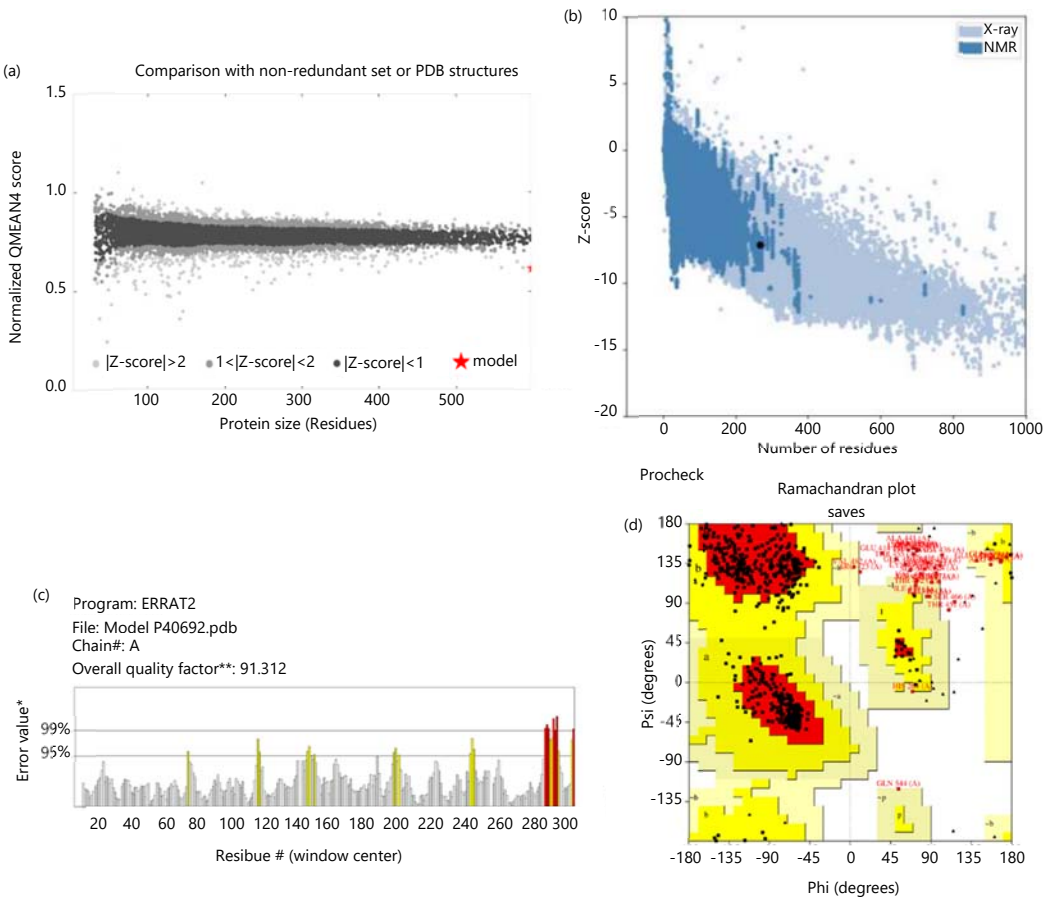

Fig. 4(a-d): Structural modeling and validation of the native MLH1 -predicted model, (a) QMEAN analysis showing overall model quality, (b) ProSA Z-score evaluation of structural reliability, (c) ERRAT analysis assessing non-bonded atomic interaction and (d) PROCHECK Ramachandran plot illustrating the stereochemical quality of the model

included for comparative homology modeling. Fifty potential templates were identified, among which P40692 showed the highest suitability. The three-dimensional structure of MLH1 was generated using the AlphaFold database model for MLH1_HUMAN (Homo sapiens), based on template P40692.1, covering residues 1-756 with complete sequence coverage. The selected template exhibited 97.98% sequence identity with the query protein. The predicted oligomeric state of the modeled structure was monomeric, and the final model was saved in PDB format for validation.

Moreover, Model 1 (P40692.1) demonstrated higher structural quality and was selected for downstream analyses. Model validation using the SAVES server yielded an ERRAT score of 91.312%, indicating reliable non-bonded atomic interactions. Ramachandran plot analysis performed through SWISS-MODEL showed that 90.30% of residues were located in the most favored regions. Additional stereochemical assessment using PROCHECK further confirmed the structural reliability of the model, with detailed statistics summarized in Table 6. The selected structure was refined using GalaxyRefine to improve side-chain conformations and overall structural geometry. Model quality evaluation resulted in a QMEAN Z-score of -4.17, suggesting reduced structural stability (Fig. 4a), while ProSA analysis produced a Z-score of -7.15, supporting the presence of structural instability (Fig. 4b). The optimized P40692.1. A model is highlighted in Fig. 4c and the Procheck Ramachandran Plot (Fig. 4d). Root-Mean-Square Deviation (RMSD) analysis identified L11V (0.57 Å), R27Q (0.57 Å), I115T (0.58 Å), and E319K (0.57 Å), as exhibiting the highest conformational deviations from the wild-type structure. These variants were therefore selected for subsequent protein-ligand docking studies (Table 6).

Table 6: Structural validation and comparative assessment of wild-type and mutant MLH1 protein models

		ERRAT	Precheck-ramachandran plot				Verify	TM align	
Protein	A.A	Score	Core (%)	Allow (%)	Generously (%)	Disallowed (%)	Score (%)	TM Score	RMSD
rs138705565	R27Q	93.0116	95.10	4.10	0.60	0.10	66.40	0.99644	0.57
rs1399177067	I94V	94.5364	96.20	3.10	0.30	0.40	67.06	0.99652	0.56
rs147342421	Q48H	93.3665	95.10	4.30	0.30	0.30	66.80	0.99651	0.56
rs1482654951	D304H	93.6982	95.40	3.70	0.60	0.30	64.55	0.99668	0.55
rs1799977	I219F	93.75	94.90	4.60	0.10	0.40	66.67	0.99652	0.56
rs200076893	Y126N	94.6755	95.90	3.50	0.10	0.40	66.27	0.99668	0.55
rs267607706	N38K	94.7811	96.00	3.40	0.30	0.30	66.67	0.99649	0.56
rs267607725	G98S	91.6667	95.20	4.10	0.40	0.30	65.34	0.99662	0.55
rs367654552	R18G	94.5364	95.60	3.80	0.30	0.30	63.89	0.99663	0.55
rs63750650	E268G	94.3709	95.70	3.80	0.30	0.10	61.64	0.99659	0.56
rs63750796	E319K	95.0413	96.00	3.40	0.30	0.30	65.08	0.9965	0.57
rs63750877	K57Q	93.0693	96.00	3.20	0.60	0.10	64.55	0.99667	0.55
rs63751109	S44F	95.1505	95.70	3.70	0.30	0.30	61.77	0.99654	0.56
rs730882127	A31G	96.8333	96.00	3.40	0.10	0.40	64.29	0.99667	0.55
rs752850761	R79S	90.9091	95.70	3.70	0.10	0.40	67.86	0.99672	0.55
rs756398627	R27W	94.3615	95.60	3.70	0.30	0.40	63.62	0.99658	0.56
rs764120517	I115T	94.5455	95.60	3.80	0.30	0.30	67.06	0.99635	0.58
rs779581111	T310A	94.2339	95.60	3.70	0.30	0.40	66.67	0.99661	0.55
rs780141938	I68M	92.7393	96.20	3.40	0.30	0.10	66.01	0.99657	0.56
rs780232692	C142R	95.0331	96.20	3.40	0.30	0.10	62.43	0.9966	0.56
rs786203362	R79G	94.3615	95.90	3.50	0.30	0.30	67.59	0.99663	0.55
rs864622396	V15A	93.3555	95.90	3.40	0.30	0.40	66.67	0.99658	0.56
rs864622596	L11V	93.1894	95.30	4.10	0.30	0.30	68.65	0.99644	0.57
rs876659860	K196T	93.75	95.60	4.10	0.10	0.10	63.76	0.99657	0.56
rs876660301	V15M	94.1176	95.70	3.70	0.40	0.10	66.93	0.99665	0.55
rs876660775	K311N	95.5224	95.90	3.40	0.30	0.40	69.18	0.99661	0.55

Table 7: Docking score for 30 ligands with wild-type and mutant MLH1 proteins

Ligands	Wild type (P40692)	R27Q	E319K	I115T	L11V
5Fluorouracil	-4.6	-5	-4.6	-4.8	-4.4
Adagrasib	-7.8	-7.5	-7.5	-7.3	-7.7
Afatinib	-7.4	-7.5	-7.6	-7.6	-7.4
Bevacizumab	-5.2	-5.4	-5.3	-5.3	-5.4
Cabozantinib	-8.8	-9	-9.1	-8.2	-8.8
Camptosar	-8.6	-8.6	-9.1	-8.9	-8.8
Capecitabine	-6.4	-6.3	-6	-6.6	-6.3
Crizotinib	-8.5	-8.4	-4.2	-7.6	-8.2
Doxorubicin	-7.2	-7.6	-8.6	-8.3	-7.6
Erlotinib	-7.3	-7	-7	-6.8	-6.7
Fruquintinib	-7.6	-7.4	-7.8	-7.3	-8
Hydroxyurea	-3.9	-3.9	-3.9	-3.9	-3.9
Irinotecan	-7.7	-7.3	-7.3	-7.7	-7.9
Lenvatinib3D	-7.7	-7.1	-7.4	-7.7	-7.1
Leucovorin	-7.3	-7.6	-7.5	-7.5	-7.4
Methotrexate	-7.5	-7.4	-7.5	-7.5	-7.8
Mitomycin	-6.5	-6.4	-6.5	-7.2	-6.4
Nintedanib	-7.5	-8.7	-8.6	-8.6	-8.7
Niraparib	-7.6	-7.6	-8.1	-8	-7.5
Olaparib	-8.5	-8.3	-8	-9.1	-7.6
Osimertinib	-8.3	-7.7	-6.9	-7.8	-8.1
Panobinostat	-6.6	-6.3	-8.8	-6.3	-6.4
Regorafenib	-8	-8.3	-7.8	-8.5	-7.8
Rucaparib	-8	-6.8	-7.5	-6.9	-8
Selpercatinib	-8	-7.9	-8.7	-7.7	-8.2
Sorafenib	-8.3	-8.1	-8.1	-8.3	-8.4
Sunitinib	-7.3	-7.3	-7.3	-6.8	-7
Tegafur	-5.9	-6	-5.9	-5.4	-5.9
Temozolomide	-6.2	-6.3	-6.1	-6	-5.9
Tivozanib	-7.7	-7.6	-7.6	-7.9	-7.7

Table 8: Toxicity prediction of selected anticancer ligands using the ProTox-II server

Ligands	Carcinogenicity	Immunotoxicity	Mutagenicity	Cytotoxicity	BBB Barrier	Ecotoxicity	Clinical toxicity	Nutritional toxicity
Cabozantinib	Active (0.52)	Active (0.89)	Inactive (0.59)	Inactive (0.66)	Active (0.56)	Active (0.55)	Active (0.75)	Inactive (0.70)
Camptosar	Inactive (0.62)	Active (0.96)	Inactive (0.97)	Inactive (0.93)	Inactive (1.00)	Active (0.73)	Inactive (0.56)	Inactive (0.74)
Crizotinib	Active (0.55)	Active (0.65)	Inactive (0.56)	Inactive (0.64)	Active (0.67)	Active (0.54)	Active (0.70)	Inactive (0.70)
Nintedanib	Inactive (0.52)	Inactive (0.79)	Inactive (0.63)	Inactive (0.56)	Inactive (0.71)	Inactive (0.68)	Active (0.67)	Inactive (0.58)
Olaparib	Inactive (0.57)	Inactive (0.95)	Inactive (0.54)	Inactive (0.65)	Active (0.85)	Inactive (0.52)	Active (0.71)	Inactive (0.71)
Sorafenib	Inactive (0.50)	Active (0.92)	Inactive (0.79)	Active (0.77)	Active (0.65)	Active (0.50)	Active (0.69)	Inactive (0.70)
Regorafenib	Inactive (0.50)	Active (0.99)	Inactive (0.79)	Active (0.77)	Active (0.65)	Active (0.50)	Active (0.69)	Inactive (0.70)
Panobinostat	Active (0.57)	Inactive (0.72)	Active (0.55)	Inactive (0.63)	Active (0.78)	Inactive (0.55)	Active (0.87)	Inactive (0.55)
Selpercatinib	Inactive (0.57)	Active (0.97)	Inactive (0.58)	Inactive (0.61)	Active (0.59)	Active (0.56)	Active (0.57)	Inactive (0.72)

***In silico* protein-ligand docking study:** Molecular docking studies were performed using the PyRx platform to investigate the binding interactions between the MLH1 protein and a library of small-molecule compounds. A total of 30 ligands were screened against the MLH1 structure, with detailed compound information presented in Table 7. The selected compounds are well-known or emerging anticancer agents with reported relevance in colorectal cancer therapy. These results highlight the potential therapeutic significance of these molecules in MLH1-mutated or mismatch repair-deficient colorectal cancer and support their further investigation through molecular dynamics simulations and experimental validation.

The docking analysis revealed that Cabozantinib, Camptosar, Crizotinib, Nintedanib, Olaparib, Sorafenib, Regorafenib, Panobinostat, and Selpercatinib exhibited strong binding affinities toward both wild-type and mutant MLH1 proteins. Toxicity prediction using the ProTox-II server indicated generally low mutagenic and nutritional toxicity profiles, although variations were observed in immunotoxic, carcinogenic, cytotoxic, and clinical toxicity endpoints. Among the screened compounds, Nintedanib and Olaparib demonstrated comparatively favorable toxicity profiles, as given in Table 8.

These ligands were further evaluated against MLH1 wild-type and mutant variants (R27Q, L11V, I115T, and E319K), which were identified through RMSD-based structural deviation analysis. Interaction analyses were performed to assess mutation-induced changes in binding affinity, residue contacts, and complex stability. All compounds showed binding energies below -3.9 kcal/mol, indicating stable interactions with both native and mutant MLH1 structures. Notably, Cabozantinib exhibited the highest binding affinity (up to -9.1 kcal/mol), followed by Camptosar and Olaparib, which consistently showed strong docking performance across all protein variants. Furthermore, Nintedanib, Sorafenib, Regorafenib, Panobinostat, and Selpercatinib maintained stable and robust binding interactions with both wild-type and mutant MLH1, supporting their potential as promising inhibitory candidate.

As summarized in Table 9, Cabozantinib established extensive van der Waals interactions with key MLH1 residues, including PRO300, SER95, ARG265, PHE261, LEU260, LEU296, VAL303, ASP304, ILE298, GLU297, GLU313, SER269, SER271, LEU317, and GLU319. Similarly, Camptosar engaged with THR96, ILE94, TYR97, GLU297, LEU296, PHE261, ASN306, LEU260, ILE298, ASN302, SER269, VAL307, LEU317, and SER271. Crizotinib was observed to interact with VAL307, GLU319, GLU313, SER269, GLU268, ILE298, LEU260, ASN302, LEU296, GLU297, HIS264, and PHE261, while Olaparib occupied the binding pocket through contacts with GLU297, LEU296, HIS264, ILE298, LEU266, PRO300, PHE261, LEU260, ASN306, ASN302, VAL307, LEU317, GLU313, SER269, and SER271, as illustrated in Fig. 5. As shown in Fig. 6, Cabozantinib, Camptosar, Crizotinib, and Olaparib demonstrated stable and well-oriented binding within the MLH1 active site. The corresponding docking conformations and LigPlot+ interaction profiles revealed multiple hydrogen bonding and hydrophobic interactions with surrounding residues, indicating strong and energetically favorable protein-ligand complexes.

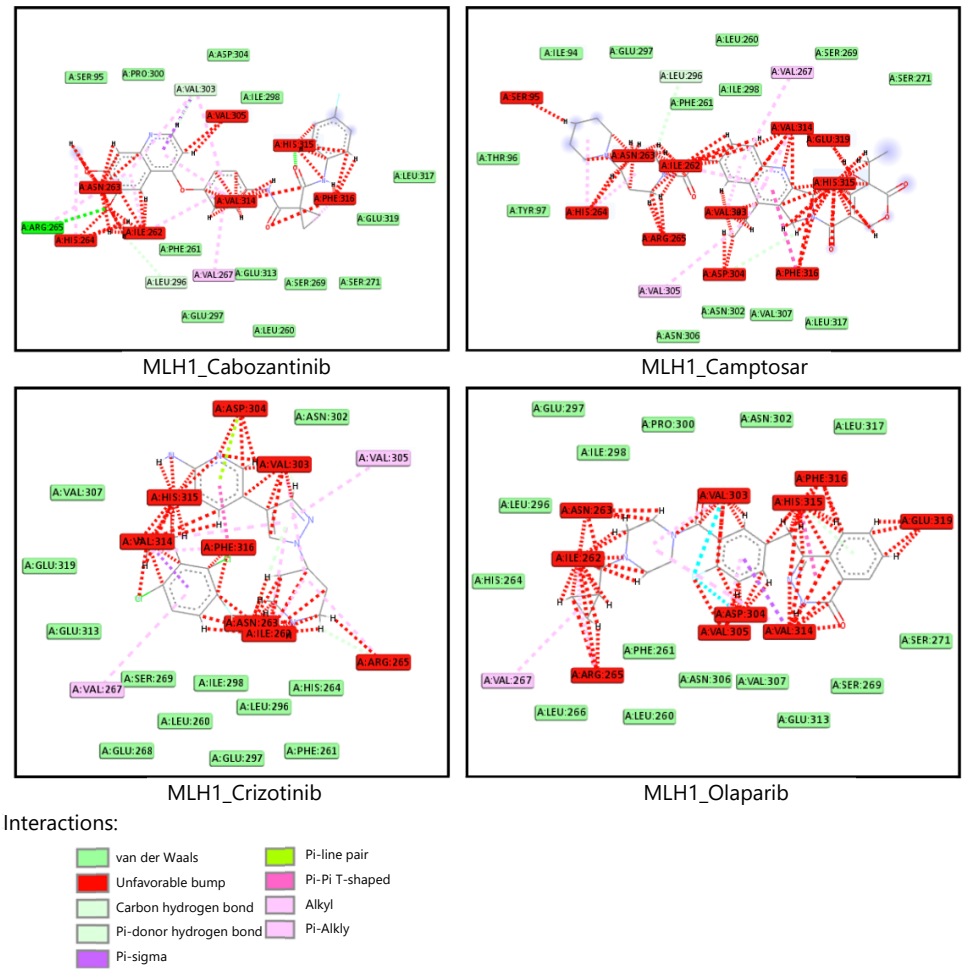

Fig. 5: Illustration of docking interactions between the wild-type MLH1 protein and selected ligands

For the L11V mutant, Cabozantinib interacted with ILE298, ARG265, SER95, ILE94, PRO300, ASP304, LEU296, PHE261, GLU297, GLU313, SER269, LEU272, SER271, LEU323, and LEU317. Camptosar formed interactions with ILE94, GLU297, LEU296, PHE261, ILE298, SER269, LEU260, SER271, LEU317, ASN302, VAL307, THR96, and TYR97. Nintedanib engaged VAL307, ASP304, GLU313, ARG265, ILE94, PRO300, SER95, GLU297, GLU268, LEU260, HIS312, ILE298, LEU296, GLU319, and SER271. Sorafenib interacted with GLU297, LEU296, ILE298, VAL305, ASN306, LEU317, VAL307, THR270, SER269, SER271, LEU272, and GLU313 as shown in Fig. 7.

In the R27Q mutant, Cabozantinib formed interactions with SER95, ARG265, ILE94, PHE261, PRO300, LEU296, ILE298, GLU297, ASP304, GLU313, SER269, SER271, LEU272, LEU317, and LEU323, while Camptosar bound with THR96, TYR97, ILE94, GLU297, LEU296, PHE261, LEU260, ILE298, SER269, LEU317, ASN302, GLU313, GLU319, VAL307, and SER271. Crizotinib interacted with GLU319, LEU317, VAL307, GLU313, SER269, GLU268, LEU260, ILE298, GLU297, LEU272, ASN302, PRO300, LEU296, HIS264, PHE261, and ARG265. Nintedanib engaged VAL307, ASP304, GLU313, ILE94, ARG265, PRO300, SER95, GLU297, GLU268, HIS312, LEU260, ILE298, SER269, LEU296, and GLU319as shown in Table 9.

In the I115T mutant, Camptosar interacted with ILE94, THR96, TYR97, GLU297, LEU296, PHE261, ILE298, LEU260, SER269, SER271, VAL307, ASN302, and LEU317. Nintedanib formed van der Waals interactions with VAL307, ASP304, GLU319, LEU296, ILE298, LEU260, HIS312, GLU313, GLU268, GLU297, ARG265, ILE94, SER95, and PRO300. Olaparib interacted through HIS264, LEU296, GLU297, ILE298, PRO300, ASN302, LEU317, LEU266, PHE261, LEU260, ASN306, VAL307, GLU313, SER269, and SER271. Regorafenib engaged LEU317, SER95, THR96, HIS264, GLU297, LEU296, VAL305, ILE298, VAL267, VAL307, and SER269, as shown in Fig. 7.

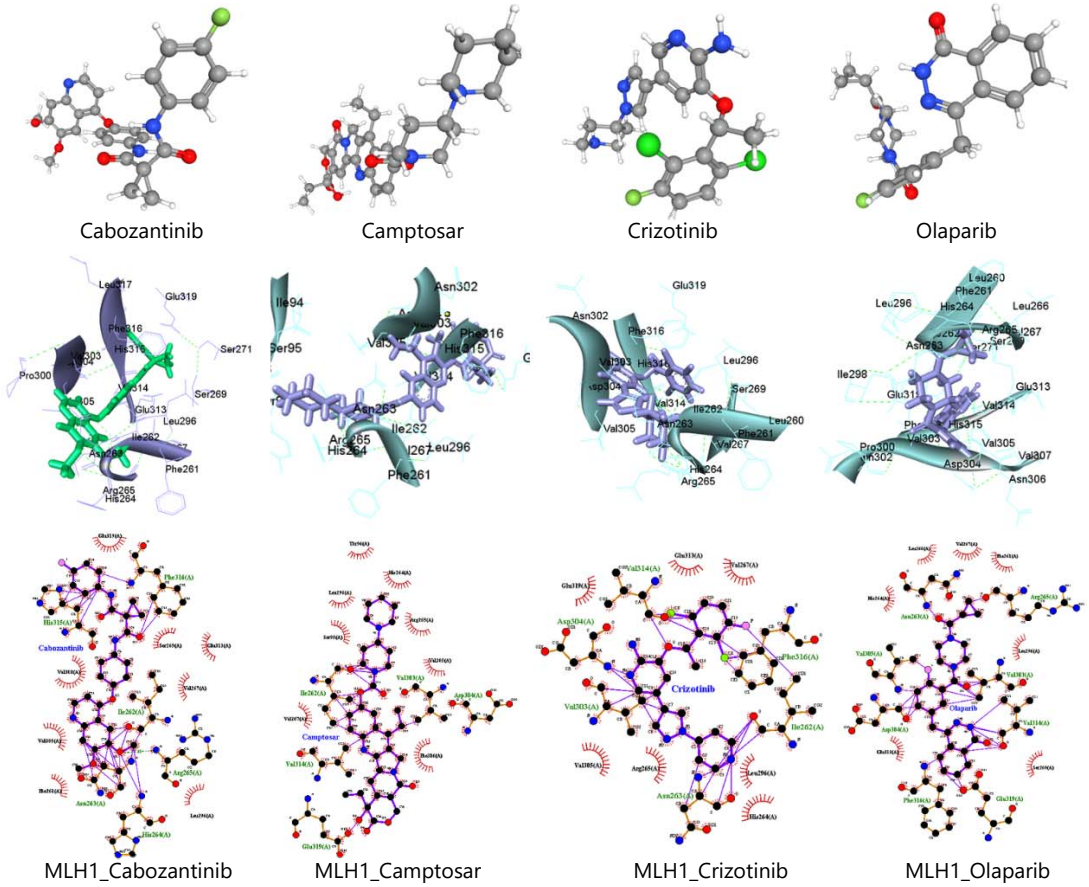

Fig. 6: Molecular docking interactions and 3D structural visualization of ligands within the active site of the MLH1 protein

Table 9: Interacting residues obtained from docking studies. Variants of the wild protein, R27Q, L11V, I115T and E319K as well as the ligands' hydrophobic interactions and binding residues

Protein-ligands	Hydrophilic interactions	Hydrophobic interactions	Stearic hinderance
WT-Cabozantinib	PRO300, SER95, ARG265, PHE261, LEU260, LEU296, VAL303, ASP304, ILE298, GLU297, GLU313, SER269, SER271, LEU317, GLU319	VAL267	VAL262, ASN263, HIS264, HIS315, VAL305, VAL314, PHE316
WT-Camptosar	THR96, ILE94, TYR97, GLU297, LEU296, PHE261, ASN306, LEU260, ILE298, ASN302, SER269, VAL307, LEU317, SER271	VAL305, VAL267	SER95, HIS264, ASN263, ILE262, ARG265, VAL303, ASP304, VAL314, GLU319, HIS315, PHE316
WT-Crizotinib	VAL307, GLU319, GLU313, SER269, GLU268, ILE298, LEU260, ASN302, LEU296, GLU297, HIS264, PHE261	VAL305, VAL267	VAL314, HIS315, ASP304, PHE316, VAL303, ASN263, ILE262, ARG265
WT-Olaparib	GLU297, LEU296, HIS264, ILE298, LEU266, PRO300, PHE261, LEU260, ASN306, ASN302, VAL307, LEU317, GLU313, SER269, SER271	VAL267	ASN263, ILE262, ARG265, VAL303, ASP303, VAL305, PHE316, HIS315, VAL314, GLU319
R27Q-Cabozantinib	SER95, ARG265, ILE94, PHE261, PRO300, LEU296, ILE298, GLU297, ASP304, GLU313, SER269, SER271, LEU272, LEU317, LEU323	VAL303, VAL267, GLU319	ASN263, ILE262, HIS264, VAL305, VAL314, HIS315, PHE316
R27Q-Camptosar	THR96, TYR97, ILE94, GLU297, LEU296, PHE261, LEU260 ILE298, SER269, LEU317, ASN302, GLU313, GLU319, VAL307, SER271	VAL305, VAL267	SER95, HIS264, ASN263, ARG265, ILE262, VAL303, ASP304, VAL314, PHE316, HIS315

Table 9: Continue

Protein-ligands	Hydrophilic interactions	Hydrophobic interactions	Stearic hinderance
R27Q-Crizotinib	GLU319, LEU317, VAL307, GLU313, SER269, GLU268, LEU260, ILE298, GLU297, LEU272, ASN302, PRO300, LEU296, HIS264, PHE261, ARG265	VAL267, VAL305	ASP304, HIS315, VAL303, VAL314, PHE316, ASN263, ILE262
R27Q-Nintedanib	VAL307, ASP304, GLU313, ILE94, ARG265, PRO300, SER95, GLU297, GLU268, HIS312, LEU260, ILE298, SER269, LEU296, GLU319	NULL	HIS315, PHE316, VAL314, VAL303, VAL305, ILE262, VAL267, HIS264, ASN263
L11V- Cabozantinib	ILE298, ARG265, SER95, ILE94, PRO300, ASP304, LEU296, PHE261, GLU297, GLU313, SER269, LEU272, SER271, LEU323, LEU317	VAL267, GLU319	VAL305, VAL303, ASN263, HIS264, ILE262, VAL314, HIS315, PHE316
L11V-Camptosar	ILE94, GLU297, LEU296, PHE261, ILE298, SER269, LEU260, SER271, LEU317, ASN302, VAL307, THR96, TYR97	VAL305, VAL267	SER95, ASN263, ILE262, ARG265, HIS264, VAL303, ASP304, VAL314, GLU319, PHE316, HIS315
L11V-Nintedanib	VAL307, ASP304, GLU313, ARG265, ILE94, PRO300, SER95, GLU297, GLU268, LEU260, HIS312, ILE298, LEU296, GLU319, SER271	NULL	HIS315, PHE316, VAL314, VAL303, SER269, ILE262, VAL267, HIS264, ASN263, VAL305
L11V-Sorafenib	GLU297, LEU296, ILE298, VAL305, ASN306, LEU317, VAL307, THR270, SER269, SER271, LEU272, GLU313	NULL	HIS264, ASN263, ILE262, ARG265, PHE316, VAL303, ASP304, HIS315, VAL314, GLU319
I115T-Camptosar	ILE94, THR96, TYR97, GLU297, LEU296, PHE261, ILE298, LEU260, SER269, SER271, VAL307, ASN302,LEU317	VAL305, VAL267	SER95, HIS264, ASN263, ILE262, ARG265,VAL303, ASP304, VAL314, PHE316, GLU319, HIS315, PHE316
I115T-Nintedanib	VAL307, ASP304, GLU319, LEU296, ILE298, LEU260, HIS312, GLU313, GLU268, GLU297, ARG265, ILE94, SER95, PRO300	NULL	HIS315, PHE316, VAL314, VAL303, VAL305, SER269, ILE262, VAL267, HIS264, ASN263
I115T-Olaparib	HIS264, LEU296, GLU297, ILE298, PRO300, ASN302, LEU317, LEU266, PHE261, LEU260, ASN306, VAL307, GLU313, SER269, SER271	VAL267	ASN263, ILE262, ARG265, VAL303, ASP304, VAL305, HIS315, PHE316, VAL314, GLU319
I115T-Regorafenib	LEU317, SER95, THR96, HIS264, GLU297, LEU296, VAL305, ILE298, VAL267, VAL307. SER269	ILE94, ARG265, VAL303, GLU313, ASP304, GLU319	ASN263, ILE262, PHE316, VAL314, HIS315
E319K-Cabozantinib	THR96, ILE94, GLU297, LEU296, PHE261, ILE298, SER269, SER271, LEU317, ASN302, VAL307	VAL267, TYR97, VAL305	SER95, HIS264, ASN263, ARG265, ILE262, VAL303, ASP304, VAL314, PHE316, HIS315, LYS319
E319K-Crizotinib	ASN302, PRO300, VAL307, LYS319, GLU313, SER269, ILE298, HIS264, LEU296, LEU272, GLU297, GLU268, PHE261	VAL305, VAL267	ASP304, HIS315, VAL303, PHE316, VAL314, ASN263, ILE262, ARG265
E319K-Panobinostat	ILE94, SER95, PHE261, VAL305, ASP304, VAL307, GLU313, SER269, SER271	ARG265, VAL267	ILE262, ASN263, HIS264, VAL303, VAL314, PHE316, HIS315, LYS319
E319K-Selpercatinib	SER271, THR270, LEU272, SER269, GLU313, TYR97, VAL305, THR96, ASP304, VAL307, ILE298, GLU297, LEU296, PHE261, HIS264, SER95, PRO300	LEU317, VAL303, ARG265, VAL267	LYS319, HIS315, VAL314, PHE316, ILE262, ASN262, ASN263, ILE94

For the E319K mutant, Cabozantinib interacted with THR96, ILE94, GLU297, LEU296, PHE261, ILE298, SER269, SER271, LEU317, ASN302, and VAL307. Camptosar engaged ASN302, PRO300, VAL307, LYS319, GLU313, SER269, ILE298, HIS264, LEU296, LEU272, GLU297, GLU268, and PHE261. Panobinostat ligand occupied the E319K binding pocket through standard Van der Waals interactions involving residues ILE94, SER95, PHE261, VAL305, ASP304, VAL307, GLU313, SER269, and SER271. Similarly, Selpercatinib binding to the E319K variant occurred through interactions with SER271, THR270, LEU272, SER269, GLU313, TYR97, VAL305, THR96, ASP304, VAL307, ILE298, GLU297, LEU296, PHE261, HIS264, SER95, and PRO300, as detailed in Table 9.

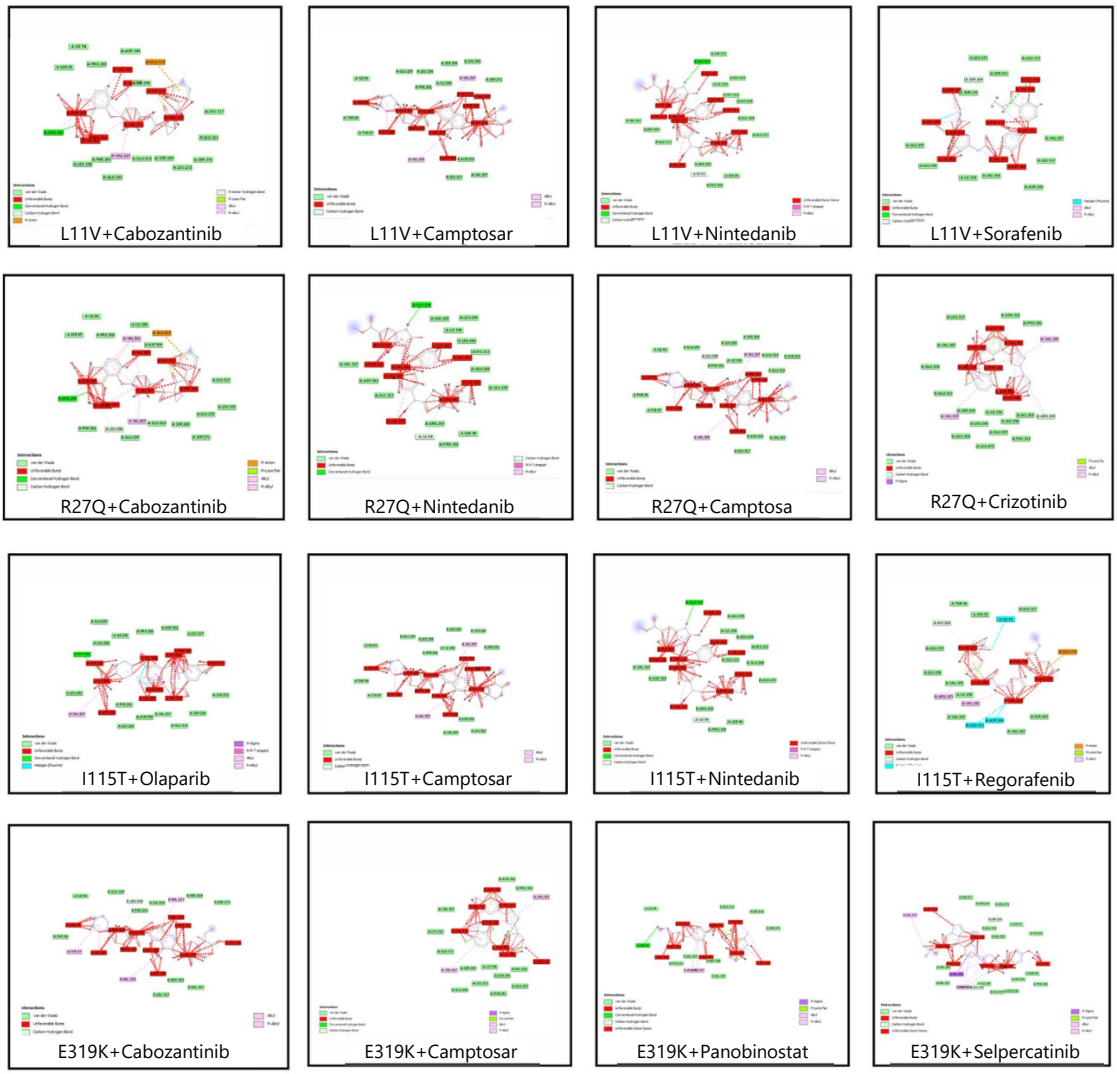

Fig. 7: Two-dimensional interaction diagrams illustrating the binding modes of selected ligands within the active sites of MLH1 mutant variants

The interaction residues identified via docking differ significantly between the mutant and wild-type forms, as listed in Table 9. These differences suggest that the E319K mutation alters the protein's functional properties. Specifically, the polymorphisms appear to affect the protein structure by modifying hydrophilic interactions, hydrogen bonding, and the total number of interacting residues.

Molecular docking analyses revealed key hydrophilic, hydrophobic, and steric interactions between selected ligands and the wild-type MLH1 protein (P40692) as well as its variants R27Q, L11V, I115T, and E319K (Fig. 7). In the wild-type structure, Cabozantinib, Camptosar, Crizotinib, Irinotecan, and Olaparib showed strong binding through multiple polar and non-polar interactions involving residues such as PRO300, SER95, ARG265, GLU297, GLU313, PHE261, LEU260, and VAL267. In contrast, steric clashes were commonly observed near VAL262, HIS264, VAL305, and PHE316. Overall, these ligands maintained a balanced interaction network of hydrogen bonding and hydrophobic contacts, supporting stable binding within the MLH1 active site (Table 9).

In mutant variants, ligand interaction patterns were notably altered due to structural changes in the binding pocket. Although key hydrogen bonds and hydrophobic contacts were largely retained, new interactions and increased steric hindrance were observed across R27Q, L11V, and I115T variants, particularly near residues such as HIS315, VAL303, and GLU319 (Fig. 7). Panobinostat and Selpercatinib

also demonstrated strong binding in the E319K variant through extensive van der Waals and polar interactions; however, enhanced steric clashes indicated reduced binding compatibility in certain regions. Collectively, these findings suggest that MLH1 mutations significantly influence ligand binding behavior, potentially affecting drug efficacy and protein functionality.

Normal mode analysis (NMA): NMA was performed to investigate the structural stability and intrinsic dynamic behavior of the MLH1 wild-type protein and its variants (L11V, R27Q, I115T, and E319K). The iMODS platform generated key parameters, including deformability profiles, B-factor mobility, eigenvalues, covariance matrices, and elastic network models. Within the framework of NMA, the eigenvalue serves as an indicator of structural stiffness, representing the energy required to induce conformational deformation. Consequently, higher eigenvalues correspond to increased rigidity, whereas lower values reflect enhanced flexibility. The wild-type MLH1 protein exhibited an eigenvalue of 6.39×10^{-8}, accompanied by low deformability and consistent B-factor fluctuations, indicative of a structurally stable and compact conformation (Fig. 7). Among the analyzed variants, the L11V mutant demonstrated a marginally elevated eigenvalue of 6.65×10^{-8}, suggesting increased structural rigidity and reduced global mobility relative to the wild type. In contrast, the R27Q, I115T, and E319K mutants displayed eigenvalues comparable to or slightly lower than those of the wild type (~$6.2\text{-}6.4\times10^{-8}$), implying a modest increase in structural flexibility.

The deformability and B-factor analyses revealed that residue fluctuations in all variants were predominantly localized to loop and surface regions, while the structural core remained largely unaffected. Covariance analysis indicated a balanced distribution of correlated and anti-correlated motions, reflecting the preservation of residue interaction networks and coordinated domain movements. Furthermore, elastic network modeling demonstrated that all structures retained a compact and well-connected architecture, with only minor variations in peripheral flexibility. Moreover, the L11V variant was selected for subsequent molecular dynamics (MD) simulation. This selection was justified by its relatively higher eigenvalue, indicative of increased rigidity and altered dynamic behavior compared to the wild type. Given that even subtle changes in structural stiffness can influence protein function and conformational dynamics, L11V represents the most structurally distinct variant and is therefore an appropriate candidate for in-depth dynamic characterization through MD simulation.

DISCUSSION

The *MLH1* gene plays a central role in preserving genomic integrity through its involvement in the DNA mismatch repair (MMR) pathway. By forming a heterodimer with PMS2, MLH1 generates the MutLα complex, which coordinates the correction of base-base mismatches and insertion-deletion loops that arise during DNA replication. This repair system acts as a molecular defense, ensuring that replication errors do not accumulate and compromise cellular stability[8,21]. When MLH1 function is disrupted, the consequences are profound. Germline mutations in MLH1 are among the most well-established causes of Lynch syndrome, a hereditary cancer predisposition disorder primarily associated with colorectal and endometrial cancers[22]. Defective MMR leads to microsatellite instability (MSI), a condition marked by widespread replication errors in repetitive DNA regions. Over time, this genomic instability accelerates the accumulation of oncogenic mutations, driving malignant transformation[23].

Missense mutations affecting critical functional domains of MLH1 can significantly impair its activity[24]. Alterations within the N-terminal ATPase domain may hinder ATP binding and hydrolysis, processes essential for conformational changes during repair. Similarly, mutations in the C-terminal region can weaken its interaction with PMS2, destabilizing the MutLα complex and reducing repair efficiency[25]. Variants such as G67R and K618A have been shown to interfere with these interactions, underscoring how subtle structural changes can translate into substantial functional loss[26].

In addition to hereditary mutations, MLH1 is frequently inactivated in sporadic colorectal cancers through promoter hypermethylation[27]. Although the underlying mechanism differs, the outcome remains the same: reduced MMR capacity, increased mutation rates, and tumor progression. This convergence highlights MLH1's role as a critical tumor suppressor gene[28]. In Lynch syndrome, MLH1 mutations account for approximately 25-40% of all identified mismatch repair gene mutations, making it one of the leading genetic contributors to hereditary nonpolyposis colorectal cancer[29]. To date, more than 3,000 distinct MLH1 variants have been documented, including missense, nonsense, frameshift, splice-site mutations, large genomic rearrangements, and promoter alterations. Among these, over 1,000 are classified as pathogenic or likely pathogenic, while a significant proportion of missense variants remain categorized as variants of uncertain significance (VUS)[30].

The continuous expansion of next-generation sequencing efforts across diverse populations is further increasing the number of reported MLH1 variants and deepening our understanding of their contributions to cancer risk. The present study was to identify missense variants in the *MLH1* gene with potential pathogenic effects and to assess their impact on protein structure and function using *in silico* prediction tools. Additionally, protein-ligand docking analyses were conducted to explore potential therapeutic targets. By integrating computational predictions with existing literature and GWAS data, the study aimed to clarify the pathogenicity of MLH1 variants and provide insights for precision diagnostics and targeted interventions in Lynch syndrome and sporadic colorectal cancers. The functional assessment of MLH1 variants revealed a high prevalence of genetic alterations with potential deleterious effects. Out of 23,356 identified SNPs, 5,167 were missense variants, and analysis using SNPnexus, PolyPhen-2, PANTHER, and PredictSNP highlighted 76 nsSNPs consistently predicted to be deleterious or probably damaging.

Disease association prediction further prioritized variants such as C672W, G54R, R27Q, and Y684C, which showed the highest probabilities of causing functional disruption, suggesting a strong link to Lynch syndrome and colorectal cancer susceptibility. Although these mutations have been previously reported, our study provides novel insights by demonstrating their structural consequences, clustering within critical functional domains, and effects on protein-ligand interactions, offering a mechanistic understanding of how these variants drive disease development. Protein stability analyses indicated that most of these nsSNPs were destabilizing, potentially impairing MLH1 function and compromising mismatch repair efficiency. Mutation clustering revealed eight mutations (R27Q, K57Q, A31G, R27W, R79G, V15A, L11V, and V15M) forming structural clusters, highlighting local conformational changes, disrupting hydrogen bonding networks, and altering hydrophobic interactions, potentially destabilizing the protein's tertiary structure.

The full-length MLH1 protein structure is not available in experimental databases such as the RCSB PDB, although partial domain structures have been reported. UniProt and NCBI provide complete sequence and functional annotations, while a full-length 3D model was generated using Swiss-Model. This predicted structure was used to evaluate the impact of missense mutations on protein conformation and ligand interactions, providing insights that are not accessible from existing experimental structures. Homology modeling and structural validation indicated that variants such as R27Q, L11V, I115T, and E319K induce notable conformational changes, particularly within ATPase and PMS2-interacting domains. These substitutions are predicted to destabilize protein structure, impair mismatch repair function, and contribute to genomic instability. The identified variants are located within structurally and functionally critical regions, suggesting strong pathogenic potential.

Normal Mode Analysis (NMA) using the iMODS server further supported these findings by revealing alterations in protein dynamics and stability. The wild-type MLH1 exhibited a stable conformational profile with low deformability and balanced residue fluctuations, whereas mutant variants showed shifts in flexibility and rigidity patterns. Among them, L11V displayed increased structural rigidity (higher

eigenvalue), while R27Q, I115T, and E319K showed slight increases in flexibility, indicating altered dynamic behavior compared to the wild type. Covariance and elastic network analyses confirmed that although the overall structural framework remained intact, subtle changes in residue motion and connectivity were introduced by mutations, potentially affecting functional dynamics.

Molecular docking studies further demonstrated that these structural alterations influence ligand-binding interactions. The wild-type MLH1 maintained stable hydrophilic and hydrophobic contacts, whereas mutant variants exhibited altered interaction profiles, including new contacts, steric clashes, and changes in hydrogen bonding networks. FDA-approved compounds such as Cabozantinib, Camptosar (Irinotecan), Crizotinib, Irinotecan, and Olaparib showed stable binding with both wild-type and mutant proteins. Key interacting residues, including LEU260, PHE261, LEU296, ILE298, GLU297, SER269, VAL307, and LEU317, were conserved across all complexes; however, mutations introduced variations in binding orientation and interaction strength, suggesting potential effects on drug efficacy and protein-ligand stability.

However, these predictions are computational and require experimental validation through in vitro and in vivo studies to confirm both the functional impact of the mutations and the therapeutic efficacy of the compounds. The novelty of this study lies in identifying domain-specific mutations that directly disrupt the structural integrity of MLH1, providing mechanistic insight into cancer development. Future research should integrate population-based analyses, wet-lab experiments, molecular dynamics simulations, and clinical studies to validate these findings and support the development of targeted MLH1 therapies. This approach can advance personalized treatment strategies for patients with Lynch syndrome and colorectal cancer.

CONCLUSION

This study presents a comprehensive computational analysis of the *MLH1* gene, a key component of the DNA mismatch repair pathway, where germline mutations are strongly linked to Lynch syndrome and colorectal cancer. From 23,356 SNPs retrieved, 5,167 were identified as missense variants, among which 76 nsSNPs were consistently predicted to be deleterious across multiple *in silico* tools. Disease association analyses highlighted C672W, G54R, R27Q, and Y684C as high-risk pathogenic variants. Mutation clustering revealed key structural hotspots, including R27Q, K57Q, A31G, R27W, R79G, V15A, L11V, and V15M, suggesting disruption of local structural stability, hydrogen bonding, and hydrophobic interactions. Homology modeling indicated that these mutations induce conformational changes in critical ATPase and PMS2-interacting domains, potentially impairing mismatch repair function. Normal Mode Analysis (iMODS) further demonstrated altered protein dynamics, where the wild-type structure maintained stable flexibility patterns, whereas variants such as L11V showed increased rigidity and others exhibited altered flexibility profiles, indicating changes in global motion and structural stability. Molecular docking with FDA-approved compounds, including Cabozantinib, Camptosar (Irinotecan), Crizotinib, Irinotecan, and Olaparib, showed stable binding with both wild-type and mutant MLH1 proteins. Key residues such as LEU260, PHE261, LEU296, ILE298, GLU297, SER269, VAL307, and LEU317 were consistently involved in ligand interactions; however, mutations introduced additional contacts, steric clashes, and altered hydrogen bonding patterns, suggesting potential modulation of drug binding affinity and efficacy. Overall, these findings highlight the structural and functional consequences of MLH1 missense mutations and their impact on protein stability, dynamics, and therapeutic interactions. The study provides a mechanistic link between genetic variation and disease risk, supporting the potential of these variants as biomarkers for colorectal cancer susceptibility. Further validation through molecular dynamics simulations, experimental studies, and population-level analyses is required to confirm these *in silico* predictions and advance precision medicine strategies for Lynch syndrome and colorectal cancer.

SIGNIFICANCE STATEMENT

This study provides a comprehensive bioinformatics analysis of deleterious *MLH1* gene variants associated with colorectal cancer and Lynch syndrome. By integrating structural modeling, protein dynamics, and molecular docking, it highlights mutation-induced functional alterations and potential therapeutic interactions. The findings offer valuable insights for future experimental validation and precision-based therapeutic development.

REFERENCES

1. McCulloch, S.M., I. Aziz, A.V. Polster, A.B. Pischel and H. Stålsmeden *et al*., 2020. The diagnostic value of a change in bowel habit for colorectal cancer within different age groups. United Eur. Gastroenterol. J., 8: 211-219.
2. Benarba, B. and A. Pandiella, 2018. Colorectal cancer and medicinal plants: Principle findings from recent studies. Biomed. Pharmacother., 107: 408-423.
3. Shamsi, U., S. Qureshi, M. Moosaji, I. Azam and R. Idrees, 2024. Colorectal cancer survival and its prognostic factors in Karachi, Pakistan. J. Coll. Physicians Surgeons Pak., 34: 688-692.
4. Abebe, Z., M.M. Wassie, A.C. Reynolds and Y.A. Melaku, 2025. Burden and trends of diet-related colorectal cancer in OECD countries: Systematic analysis based on Global Burden of Disease study 1990-2021 with projections to 2050. Nutrients, Vol. 17. 10.3390/nu17081320.
5. Ikram, A., S. Pervez, M.T. Khadim, M. Sohaib and Hafeez Uddin *et al*., 2023. National cancer registry of Pakistan: First comprehensive report of cancer statistics 2015-2019. J. Coll. Physicians Surgeons Pak., 33: 625-632.
6. Bocchi, M., E.V. Fernandes, N. de Sousa Pereira and M.K. Amarante, 2025. Etiopathogenesis and treatment of colorectal cancer. Immuno, Vol. 31. 10.3390/immuno5030031.
7. Stoffel, E.M. and F. Kastrinos, 2014. Familial colorectal cancer, beyond lynch syndrome. Clin. Gastroenterol. Hepatol., 12: 1059-1068.
8. Dąbrowska, A., J. Mastalerz, Z. Łapińska, I. Deszcz, A. Chwiłkowska and N. Rembiałkowska, 2025. Mismatch repair deficiency and the role of non-canonical functions in cancer: Diagnosis and therapeutic implications. Int. J. Mol. Sci., Vol. 26. 10.3390/ijms26199312.
9. Köger, N., L. Paulsen, F. López-Kostner, A.D. Valle and C.A. Vaccaro *et al*., 2018. Evaluation of *MLH1* variants of unclear significance. Genes Chromosomes Cancer, 57: 350-358.
10. Lynch, H.T. and T.G. Shaw, 2013. Practical genetics of colorectal cancer. Chin. Clin. Oncol., Vol. 2. 10.3978/j.issn.2304-3865.20.
11. Ghaedi, H., S.M. Ramsheh, M.E. Omidvar, A. Labbaf and E. Alehabib *et al*., 2020. Whole-exome sequencing identified a novel mutation of *MLH1* in an extended family with lynch syndrome. Genes Dis., 7: 614-619.
12. Andersen, S.D., S.E. Liberti, A. Lützen, M. Drost and I. Bernstein *et al*., 2012. Functional characterization of *MLH1* missense variants identified in lynch syndrome patients. Hum. Mutat., 33: 1647-1655.
13. Noman, A., N. Sardar, A. Islam, K. Ramzan and M.Z. Ali *et al*., 2025. Integrative computational analysis of CFTR mutations linked to cystic fibrosis. Int. J. Mol. Microbiol., 8: 52-69.
14. Sardar, N., A. Noman, K. Ramzan, I. Bilal and A. Islam *et al*., 2025. Exploring the impact of STAT4 non-synonymous mutations on hepatitis B virus susceptibility: A bioinformatics approach. Int. J. Mol. Microbiol., 8: 79-102.w
15. Tariq, H., M. Asif, M. Saleem, K. Ramzan, M. Zulfiqar, A. Amira and A.R. Asif, 2024. Evaluation of detrimental missense SNPs of human *CXCL6* gene by combining algorithms, homology modeling, and molecular docking. Int. J. Biol. Res., 4: 92-106.
16. Aslam, T., H. Tariq, M. Saleem, K. Ramzan and K. Aaliya *et al*., 2024. Bioinformatic analysis of human TLR4 coding variations associated with ocular infection: A structural prediction and molecular docking studies. Int. J. Pharm. Sci., 2: 930-954.
17. Tariq, H., M. Saleem, A.H. Ali, M. Fatima and A. Fatima *et al*., 2026. Integrative computational approaches to protein structure and drug design in POEMS syndrome. Int. J. Mol. Microbiol., 9: 17-35.

18. Khatoon, M., Y.S. Sekar, S. Rani, V. Ramesh and M. Shijili *et al*., 2025. Computational analysis of non-synonymous SNP effects on human PLVAP gene structure and function. J. Appl. Genet., 10.1007/s13353-025-01038-3.
19. Gómez, L.D.O., H.J.C. Martínez, D.A.G. Pareja, S.V. Gómez, J.E.S. Flórez, F.P. Monroy and R.G.P. Sánchez, 2024. Mutations in the *PIK3C2B*, *ERBB3*, *KIT*, and *MLH1* genes and their relationship with resistance to temozolomide in patients with high-grade gliomas. Biomedicines, Vol. 12. 10.3390/biomedicines12122777.
20. Banerjee, P., A.O. Eckert, A.K. Schrey and R. Preissner, 2018. ProTox-II: A webserver for the prediction of toxicity of chemicals. Nucleic Acids Res., 46: W257-W263.
21. Trost, H., F.W. Lopezcolorado, A. Merkell and J.M. Stark, 2025. Functions of PMS2 and MLH1 important for regulation of divergent repeat-mediated deletions. DNA Repair, Vol. 145. 10.1016/j.dnarep.2024.103791.
22. Cohen, S.A. and A. Leininger, 2014. The genetic basis of Lynch syndrome and its implications for clinical practice and risk management. Appl. Clin. Genet., 7: 147-158.
23. Awosika, J.A., J.L. Gulley and D.M. Pastor, 2025. Deficient mismatch repair and microsatellite instability in solid tumors. Int. J. Mol. Sci., Vol. 26. 10.3390/ijms26094394.
24. Dubanowicz, M., 2015. The pathological significance of selected *MLH1* missense mutations based on the theoretical *in silico* predictions. Hereditary Cancer Clin. Pract., Vol. 13. 10.1186/1897-4287-13-S1-A9.
25. Fishwick, K.M., D.G. Vieito, G. Greco, G. Collotta and M. Gatti *et al*., 2025. Disruption of protein-protein interaction hotspots in the C-terminal domain of MLH1 confers mismatch repair deficiency. NAR Cancer, Vol. 7. 10.1093/narcan/zcaf055.
26. Couch, F.J., L.J. Rasmussen, R. Hofstra, A.N.A. Monteiro, M.S. Greenblatt and N. de Wind, 2008. Assessment of functional effects of unclassified genetic variants. Hum. Mutat., 29: 1314-1326.
27. Helderman, N.C., K.D. Andini, M.E. van Leerdam, L.P. van Hest and D.R. Hoekman *et al*., 2024. *MLH1* promotor hypermethylation in colorectal and endometrial carcinomas from patients with lynch syndrome. J. Mol. Diagn., 26: 106-114.
28. Walker, R., K. Mahmood, J. Como, M. Clendenning and J.E. Joo *et al*., 2023. DNA mismatch repair gene variant classification: Evaluating the utility of somatic mutations and mismatch repair deficient colonic crypts and endometrial glands. Cancers, Vol. 15. 10.3390/cancers15204925.
29. Li, X., X. Yao, Y. Wang, F. Hu and F. Wang *et al*., 2013. *MLH1* promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features. PLoS ONE, Vol. 8. 10.1371/journal.pone.0059064.
30. Chen, E., F.M. Facio, K.W. Aradhya, S. Rojahn and K.E. Hatchell *et al*., 2023. Rates and classification of variants of uncertain significance in hereditary disease genetic testing. JAMA Netw. Open, Vol. 6. 10.1001/jamanetworkopen.2023.39571.