Sequence-structure functional implications and molecular simulation of high deleterious nonsynonymous substitutions in *IDH1* revealed the mechanism of drug resistance in glioma

Muhammad Suleman[1†], Syeda Umme-I-Hani[2],
Muhammad Salman[3], Mohammed Aljuaid[4], Abbas Khan[5,6†],
Arshad Iqbal[1], Zahid Hussain[1], Syed Shujait Ali[1], Liaqat Ali[7],
Hassan Sher[8], Yasir Waheed[9]* and Dong-Qing Wei[5,6,10,11]*

[1]Centre for Biotechnology and Microbiology, University of Swat, Swat, Khyber Pakhtunkhwa, Pakistan, [2]Punjab Medical College, Faisalabad, Punjab, Pakistan, [3]Rashid Latif Medical College, Lahore, Punjab, Pakistan, [4]Department of Health Administration, College of Business Administration, King Saud University, Riyadh, Saudi Arabia, [5]Department of Bioinformatics and Biological Statistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [6]Zhongjing Research and Industrialization Institute of Chinese Medicine, Zhongguancun Scientific Park, Meixi, Henan, China, [7]Division of Biology, Kansas State University, Manhattan, KS, United States., [8]Centre for Plant Science and Biodiversity, University of Swat, Charbagh, Pakistan, [9]Office of Research, Innovation and Commercialization, Shaheed Zulfiqar Ali Bhutto Medical University (SZABMU), Islamabad, Pakistan, [10]State Key Laboratory of Microbial Metabolism, Shanghai-Islamabad-Belgrade Joint Innovation Center on Antibacterial Resistances, Joint Laboratory of International Cooperation in Metabolic and Developmental Sciences, Ministry of Education and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China, [11]Peng Cheng Laboratory, Shenzhen, Guangdong, China

In the past few years, various somatic point mutations of isocitrate dehydrogenase (IDH) encoding genes (IDH1 and IDH2) have been identified in a broad range of cancers, including glioma. Despite the important function of *IDH1* in tumorigenesis and its very polymorphic nature, it is not yet clear how different nsSNPs affect the structure and function of IDH1. In the present study, we employed different machine learning algorithms to screen nsSNPs in the *IDH1* gene that are highly deleterious. From a total of 207 SNPs, all of the servers classified 80 mutations as deleterious. Among the 80 deleterious mutations, 14 were reported to be highly destabilizing using structure-based prediction methods. Three highly destabilizing mutations G15E, W92G, and I333S were further subjected to molecular docking and simulation validation. The docking

results and molecular simulation analysis further displayed variation in dynamics features. The results from molecular docking and binding free energy demonstrated reduced binding of the drug in contrast to the wild type. This, consequently, shows the impact of these deleterious substitutions on the binding of the small molecule. PCA (principal component analysis) and FEL (free energy landscape) analysis revealed that these mutations had caused different arrangements to bind small molecules than the wild type where the total internal motion is decreased, thus consequently producing minimal binding effects. This study is the first extensive *in silico* analysis of the *IDH1* gene that can narrow down the candidate mutations for further validation and targeting for therapeutic purposes.

1 Introduction

In primary brain tumor, glioblastoma, also known as grade IV glioma, is the most common and deadly form of brain tumor (Wirsching and Weller, 2017). In malignant gliomas, the primary GBMs account for 90% while the secondary GBMs that emerge from lower-grade gliomas (LGGs) in younger individuals account for less than 10% of clinical reports (Ohgaki and Kleihues, 2013). After the initial diagnosis, the survival for glioma individuals is from 14 to 16 months. Recent investigations revealed some metabolic features that are shared by virtually all GBMs and help to differentiate tumors from the normal brain (Tan et al., 2020). The GBM metabolic features are the excess generation of lactate in conjunction with the acetate and glucose oxidations to provide macromolecular precursors and energy. In low-grade glioma, secondary glioblastoma, and acute myelogenous leukemia, the oncogenic mutations in the two-isocitrate dehydrogenase (IDH) encoding genes (IDH1 and IDH2) have been identified (Agnihotri et al., 2013; Zhou and Wahl, 2019).

Normally in the Krebs cycle, the isocitrate is converted into a-ketoglutarate (a-KD) by isocitrate dehydrogenases (IDHs) in a NAD(P)-dependent manner. IDH1, IDH2, and IDH3 are three IDH isozymes that function in different subcellular compartments. Various somatic point mutations of IDH1 or IDH2 have been discovered in a variety of malignancies in recent years, such as gliomas and AMLs (acute myeloid leukemias) (Yan et al., 2009; Zhao et al., 2009). Identified mutations such as IDH1 R132H/C/Q, IDH2 R140Q/W/L, and R172K/T/S/G/M adversely affect the normal function of IDH protein and initiate the abnormal activity of protein with IDH mutations that produced oncometabolite 2-hydroxyglutarate from the a-KG (Frezza et al., 2010; Huang, 2019). 2-HG may accumulate to horrifically high levels of 5–35 mmol/g in human glioma samples with IDH1/2 mutations, which is 100-fold higher than its normal level in the brain (Dimitrov et al., 2015).

Single nucleotide polymorphisms (SNPs), which affect both coding and noncoding regions of DNA, are the most common genetic changes. SNPs are found every 200–300 bp in the human genome and account for around 90% of all genetic differences in the human genome. Nonsynonymous SNPs cause genetic alteration in the exonic regions of the protein and disturb their sequence, structure, and normal function by enhancing the abnormal transcription and translation mechanisms. Previously, several *in silico* computational techniques were developed to quickly and precisely assess the functional implications of nonsynonymous variation on protein structure and function (Junaid et al., 2018; Khan et al., 2020a; Khan et al., 2020b; Khan et al., 2021). Until now, a total of 298 SNPs with 207 missense mutations in the human IDH1 gene has been described and deposited to the gnomAD database.

Although IDH1 plays a crucial role in carcinogenesis (gliomas) and has a polymorphism character, it is still unclear how identified nsSNPs alter the protein's structure and biological activity. In this study, we employed a number of computational approaches to find nsSNPs in the IDH1 gene that are extremely detrimental to the structure and function of the IDH1 protein.

2 Materials and methods

2.1 Collection of data

The available data on human IDH1 were obtained from available online databases. The online database gnomAD (https://gnomad.broadinstitute.org/) was used to retrieve all predicted SNPs in the human *IDH1*gene (Karczewski and Francioli, 2017). The amino acid sequence (UniProt: O75874) and previously deposited 3D structure (PDB ID: 6BKX) of the protein that expresses the *IDH1* gene were obtained from the UniProt online database (http://www.uniprot.org/) (Rose et al., 2010; Consortium, 2015). The whole workflow of the work is given in Figure 1.

FIGURE 1
Methodological workflow of the work. Each tool used in each step is also given.
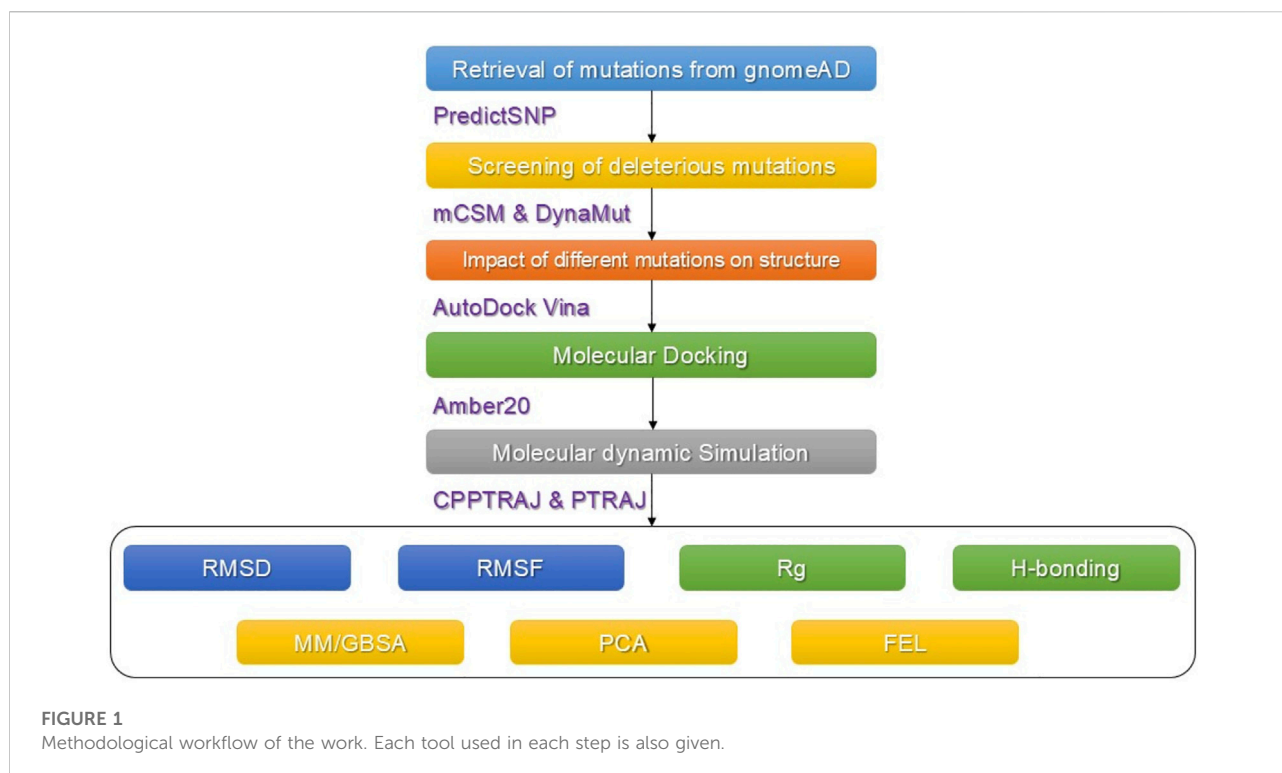
2.2 Disease-related single nucleotide polymorphism predictions

2.2.1 Prediction of Functional Consequences of nsSNPs

Different online servers such as PredictSNP (https://loschmidt.chemi.muni.cz/predictsnp1/), Polymorphism Phenotyping version 2 (Polyphen-2) (http://genetics.bwh.harvard.edu/pp2), Sorting Intolerant from Tolerant (SIFT) (http://sift.bii.a-star.edu.sg), Screening of nonacceptable polymorphism (SNAP) (https://rostlab.org/services/snap), Protein Analysis Through Evolutionary Relationship (PANTHER) (http://www.pantherdb.org/tools/csnpScoreForm.jsp), Multivariate Analysis of Protein Polymorphism (MAPP) (http://mendel.stanford.edu/SidowLab/downloads/MAPP/), and predictor of Human Deleterious Single Nucleotide Polymorphism (PhD-SNP) (http://snps.biofold.org/phd-snp/phd-snp.html) were used to predict the functional effect of all nonsynonymous single nucleotide polymorphisms (nsSNPs) (Johnson et al., 2008; Sim et al., 2012; Adzhubei et al., 2013; Landis et al., 2014; Bendl et al., 2015; Capriotti and Fariselli, 2017). All of the nsSNPs that were verified as highly deleterious by all of the aforementioned web servers were selected for further analysis.

2.2.2 Structure-based stability calculation

For structure-based stability prediction, mCSM and DynaMut webservers were used to estimate the impact of each substitution on the structural stability and flexibility (Pires et al., 2014; Rodrigues et al., 2018). The highly deleterious mutations were processed for the prediction of structure-based stability calculation. These servers use graph-based signatures to estimate the impact of each mutation on the protein's structure. The top three mutations were selected based on the mCSM and DynaMut results together for further analysis.

2.2.3 Modeling of mutants of *IDH1* protein

The crystal structure of the *IDH1* protein was extracted from PDB (Entry ID: 6BKX). Both ligands and water molecules were separated from the protein structure, and the Chimera software was used to minimize the wild-type structure of *IDH1* protein. Moreover, the predicted most deleterious mutations such as G15E, W92G, and I333S were modeled in the wild-type structure of *IDH1* protein using the Chimera software.

2.2.4 Molecular docking of DWP with the wild type and mutant *IDH1*

The impact of selected substitutions on the binding of DWP ((6aS,7S,9R, 10aS)-7,10a-dimethyl-8-oxo-2-(phenylamino)-5,6,6a,7,8,9,10,10a-octahydrobenzo[h]quinazoline-9-carbonitrile with the wild type and mutant was also evaluated using the molecular docking approach. For this estimation, a previously described protocol was employed using AutoDock Vina (Eberhardt et al., 2021).

TABLE 1 List of highly deleterious and destabilizing mutations in IDH1. Among the 80 mutations, 14 highly destabilizing ones are shown in bold.

Index	Mutation	ΔΔG mCSM	Outcome
1	**G15E**	**−3.094**	**Highly Destabilizing**
2	D16H	−1.439	Destabilizing
3	**W23C**	**−2.149**	**Highly Destabilizing**
4	Y34C	−1.848	Destabilizing
5	V35A	−1.926	Destabilizing
6	Y42C	−1.287	Destabilizing
7	D43A	−0.396	Destabilizing
8	**R49H**	**−2.437**	**Highly Destabilizing**
9	R49C	−1.957	Destabilizing
10	R49P	−1.427	Destabilizing
11	**I76T**	**−2.36**	**Highly Destabilizing**
12	**R82S**	**−2.362**	**Highly Destabilizing**
13	R82M	−1.294	Destabilizing
14	V83F	−1.491	Destabilizing
15	E85G	−1.128	Destabilizing
16	L88F	−1.714	Destabilizing
17	M91R	−0.476	Destabilizing
18	M91T	−1.317	Destabilizing
19	**W92G**	**−3.644**	**Highly Destabilizing**
20	W92R	−1.562	Destabilizing
21	N96H	−0.752	Destabilizing
22	T98N	−1.142	Destabilizing
23	N101Y	−0.377	Destabilizing
24	T106M	−0.611	Destabilizing
25	**F108V**	**−2.441**	**Highly Destabilizing**
26	R109K	−1.318	Destabilizing
27	I113S	−1.972	Destabilizing
28	G150R	−0.548	Destabilizing
29	**V152G**	**−2.065**	**Highly Destabilizing**
30	I154R	0.027	Stabilizing
31	D160Y	0.05	Stabilizing
32	G177D	−0.073	Destabilizing
33	Y183C	−0.838	Destabilizing
34	A193T	−0.975	Destabilizing
35	L207W	−1.744	Destabilizing
36	**Y208H**	**−2.648**	**Highly Destabilizing**
37	Y208C	−1.847	Destabilizing
38	T214S	−0.522	Destabilizing
39	Y219H	−0.962	Destabilizing
40	Y219C	−0.98	Destabilizing
41	D220G	−1.683	Destabilizing
42	**Y231H**	**−2.007**	**Highly Destabilizing**
43	Y235C	−1.24	Destabilizing
44	**Y246H**	**−2.036**	**Highly Destabilizing**
45	A256V	−0.53	Destabilizing
46	K260N	0.221	Stabilizing

(Continued in next column)

TABLE 1 (*Continued*) List of highly deleterious and destabilizing mutations in IDH1. Among the 80 mutations, 14 highly destabilizing ones are shown in bold.

Index	Mutation	ΔΔG mCSM	Outcome
47	G263E	−0.546	Destabilizing
48	D273G	−0.293	Destabilizing
49	G274S	−0.962	Destabilizing
50	V276M	−0.5	Destabilizing
51	S278P	−0.225	Destabilizing
52	S278L	−0.228	Destabilizing
53	D279H	−0.517	Destabilizing
54	M291T	−1.405	Destabilizing
55	T292I	−0.114	Destabilizing
56	S293I	0.169	Stabilizing
57	P298L	−0.358	Destabilizing
58	G300V	−0.604	Destabilizing
59	G300D	−1.749	Destabilizing
60	E306A	0.57	Stabilizing
61	H309R	−0.568	Destabilizing
62	H309Q	0.139	Stabilizing
63	G310R	−0.445	Destabilizing
64	R314C	0.301	Stabilizing
65	H315D	−0.52	Destabilizing
66	R317C	−0.995	Destabilizing
67	R317L	−0.082	Destabilizing
68	T325M	0.467	Stabilizing
69	N328S	−1.25	Destabilizing
70	N328K	−0.122	Destabilizing
71	**I330T**	**−2.449**	**Highly Destabilizing**
72	**I333S**	**−3.298**	**Highly Destabilizing**
73	G339R	−1.576	Destabilizing
74	L346P	−1.191	Destabilizing
75	**I367T**	**−2.845**	**Highly Destabilizing**
76	G370V	−0.276	Destabilizing
77	M372T	−1.565	Destabilizing
78	T373I	−0.732	Destabilizing
79	T373N	−1.561	Destabilizing
80	L401P	-1.626	Destabilizing

2.2.5 Molecular dynamics simulation

The highly destabilizing and functional substitutions were evaluated for the dynamic properties using the AMBER2.0 molecular simulation tool. For this purpose, ff14SB force field was recruited for uniformity with the previous parameters (Salomon Ferrer et al., 2013). A TIP3P water box (cutoff = 10.0 Å) was employed for solvation, while neutralizations were performed by adding sodium ions. Each complex was minimized well in two steps: the first for 6000 steps, while the second was run for 3,000 steps. The further protocol used in the previous study was employed. Lastly, for each

complex, a 100 ns production run under constant pressure was completed. To control the temperature, a Langevin thermostat with 1 atm pressure and 300 K was used (Zwanzig, 1973). The particle mesh Ewald (PME) algorithm was used to compute long-range interactions (Darden et al., 1993; Essmann et al., 1995). The cutoff distances were set to 10 A˚. For the covalent bonds involving hydrogen, the SHAKE algorithm was used (Ryckaert et al., 1977). GPU-accelerated simulation (PMEMD.CUDA) was used for all of the processes. Post simulation analyses including dynamic stability calculated as RMSD (root mean square deviation), residual flexibility estimated as RMSF(root mean square fluctuation), hydrogen bonding analysis over the simulation time, and the radius of gyration (Rg) for protein packing assessment were calculated using CPPTRAJ and PTRAJ modules of AMBER1.9 (Roe and Cheatham, 2013).

2.2.6 Binding free energy calculation

For the calculation of binding free energy, a whole trajectory of each complex was subjected to MM/GBSA analysis by utilizing MMPBSA.PY script (Hou et al., 2011). This widely applicable approach, which has been previously used to characterize the binding of various biological complexes, was used for estimation by employing the following equation:

$$\Delta Gbind = \Delta Gcomplex - [\Delta Greceptor + \Delta Gligand] \quad (1)$$

Each term in the binding free energy was estimated using the following equation:

$$G = Gbond + Gele + GvdW + Gpol + Gnpol \quad (2)$$

2.2.7 Clustering of MD trajectories using principal component analysis

To comprehend the motion of MD trajectories, an unsupervised learning method known as principal component analysis (PCA) (Pearson, 1901; Wold et al., 1987) was performed to acquire knowledge regarding the internal motion of the system. For this purpose, an Amber module known as CPPTRAJ was used. The spatial covariance matrix was determined for eigenvector and their atomic co-ordinates. Using an orthogonal coordinate transformation, a diagonal matrix of eigenvalues was generated. Based on the eigenvectors and eigenvalues, the principal components were extracted. Using these PCs, the dominant motions during simulation were plotted (Balsera et al., 1996; Ernst et al., 2015).

3 Results and discussion

3.1 Identification of deleterious nsSNPs

The online public resources were used to retrieve all of the available data on the human *IDH1* gene. According to the information obtained from the online gnomAD database,

there were a total of 298 SNPs in the *IDH1* protein. Of these, 207 SNPs were identified as nonsynonymous. These 207 SNPs were submitted to a different online server to identify the deleterious mutations. First, the SNPs were submitted to PredictSNP and MAPP servers, and only 141 and 140 SNPs were found to be deleterious, respectively (Supplementary Table S1). The nsSNPs were then submitted to PhD-SNP and SNAP online tools and found 63 and 55 SNPs as deleterious, respectively (Supplementary Table S2). The other online servers such as PolyPhen-1, PolyPhen-2, SIFT, and PANTHER analyzed the nsSNPs and predicted that out of 119 SNPs only 51, 46, 68, and 80 were deleterious, respectively (Supplementary Tables S3, S4). All of the nsSNPs were selected for further analysis and was predicted as highly deleterious together by all of the abovementioned online servers. Only 80 mutations were selected for the structure-based stability analysis using mCSM as shown in Table 1. mCSM predicted that only 14 mutations (G15E(-3.09), W23C(-2.14), R49H(-2.43), I76T(-2.36), R82S(-2.36), W92G(-3.64), F108V(-2.44), V152G(-2.06), Y208H(-2.64), Y231H(-2.00), Y246H(-2.03), I330T(-2.44), I333S(-3.29), and I367T(-2.84)) out of 80 were highly destabilizing. These 14 mutations were further verified using the DynaMut web server.

The reported 14 highly destabilizing mutations were then processed using the DynaMut server to determine the effect of these 14 mutations on the flexibility of protein structure. Among these 14 mutations, 12 mutations induced higher flexibility while the other two mutations, G15E and V246H, demonstrated structural rigidity. These changes in flexibility shown in red and blue were mapped onto the corresponding protein structure and are presented in Figure 2. Among these 14 mutations, only three mutations were reported to be consistently highly destabilizing, which were then selected for further analysis (Table 2).

3.2 Molecular docking analysis of the wild type and mutants

Molecular docking-based investigation of the binding variations caused by these mutations in contrast to the wild type revealed significant differences. The docking score for the wild type was calculated to be −8.76 kcal/mol. The interaction analysis revealed multiple hydrogen bonds including Ser326, Lys374, and Asp375, while various pi–pi interactions and pi–alkyl interactions were observed. The interaction pattern of the wild type is given in Figure 3A. On the other hand, despite the significant reduction in the number of bonding contacts, the His314 (correspond to His315) hydrogen bond remained conserved, which has been reported to be associated with the inhibitory properties of this drug. With a single hydrogen bond and various other interactions, the docking score for this complex was calculated to be −7.35 kcal/mol. It can be seen that this

FIGURE 2
Effect of mutations on the flexibility of different residues. Different colors represent different levels of flexibility. The red regions demonstrate that the flexibility is increased, the blue regions show that the flexibility is decreased due to the mutations, and gray represents no change in the flexibility.
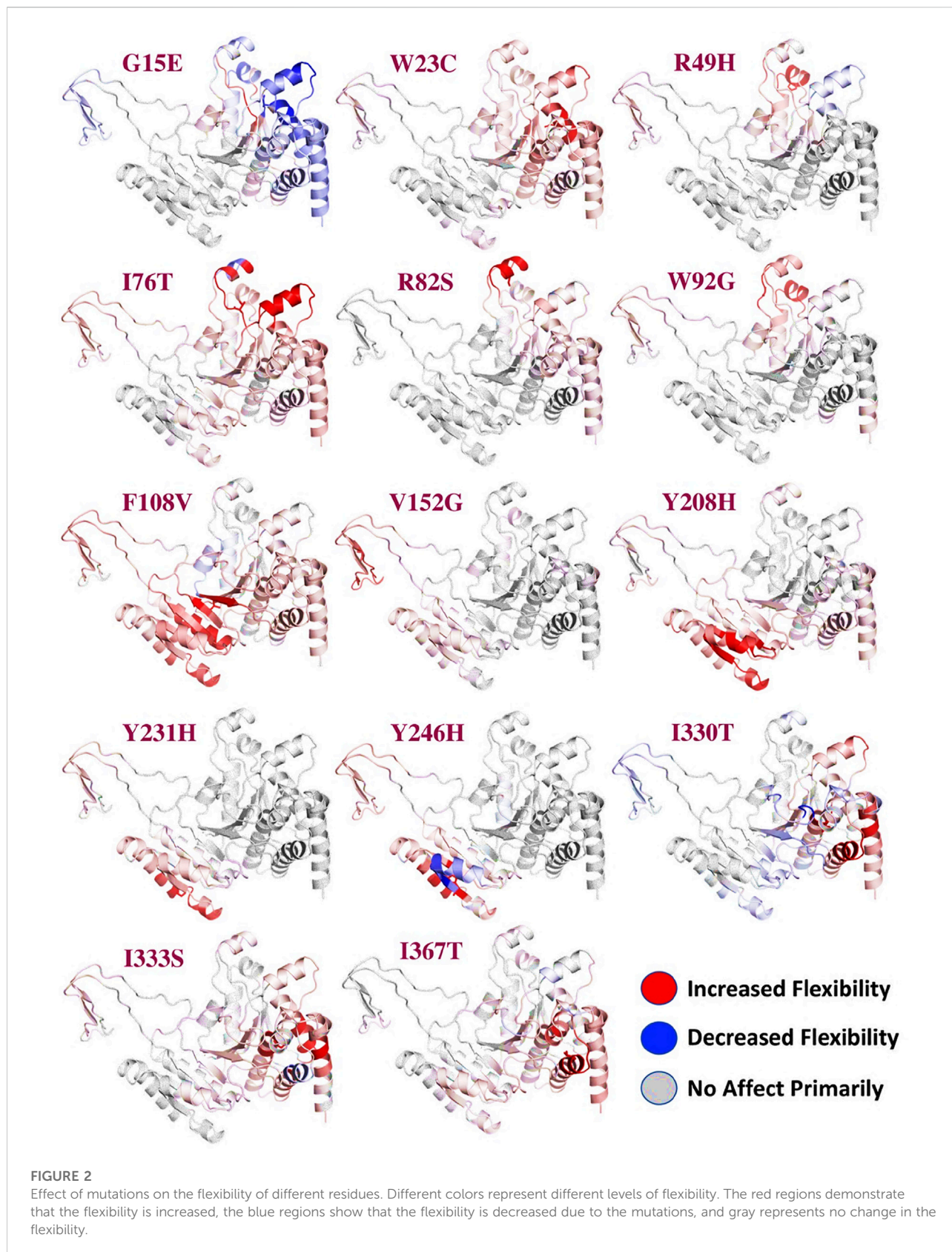
TABLE 2 List of highly deleterious and destabilizing mutations in IDH1. Among the 14 mutations, three are highly destabilizing reported by both the servers (mCSM and DynaMut) shown in bold.

Index	Mutation	ΔΔG ENCoM	ΔΔG DynaMut
1	**G15E**	**1.101**	**−3.347**
2	W23C	−0.977	−0.288
3	R49H	−0.348	−1.701
4	I76T	−0.198	−1.926
5	R82S	−0.607	−2.139
6	**W92G**	**−1.741**	**−2.633**
7	F108V	−0.341	−2.520
8	V152G	−0.894	−1.667
9	Y208H	−0.662	−0.591
10	Y231H	−0.369	−0.945
11	Y246H	−0.136	−0.100
12	I330T	−0.171	−0.733
13	**I333S**	**−0.478**	**−3.413**
14	I367T	−0.318	−2.610

complex has lost important hydrogen bonds of essential residues, thus reducing the bonding energy and contributing to resistance to chemotherapy. The interaction pattern of G15E is given in Figure 3B. Moreover, the docking score for W92G was estimated to be −8.62 kcal/mol. This complex retained some important hydrogen contacts, that is, Ser326 and His314 (correspond to His315), which gives comparable results to the wild type. With various hydrogen bonding contacts, many pi–pi, pi–alkyl, and salt bridge contacts were also reported in this complex. The interaction pattern of W92G is given in Figure 3C. I333S lies near the active site and reports a substantial decrease in the bonding pattern. With only one hydrogen bond of His314 (corresponding to His315) and pi–alkyl interaction with Val311, this complex reported a significant decrease in the docking score. The docking score for this complex was calculated to be −6.87 kcal/mol. The interaction pattern of I333S is given in Figure 3D.
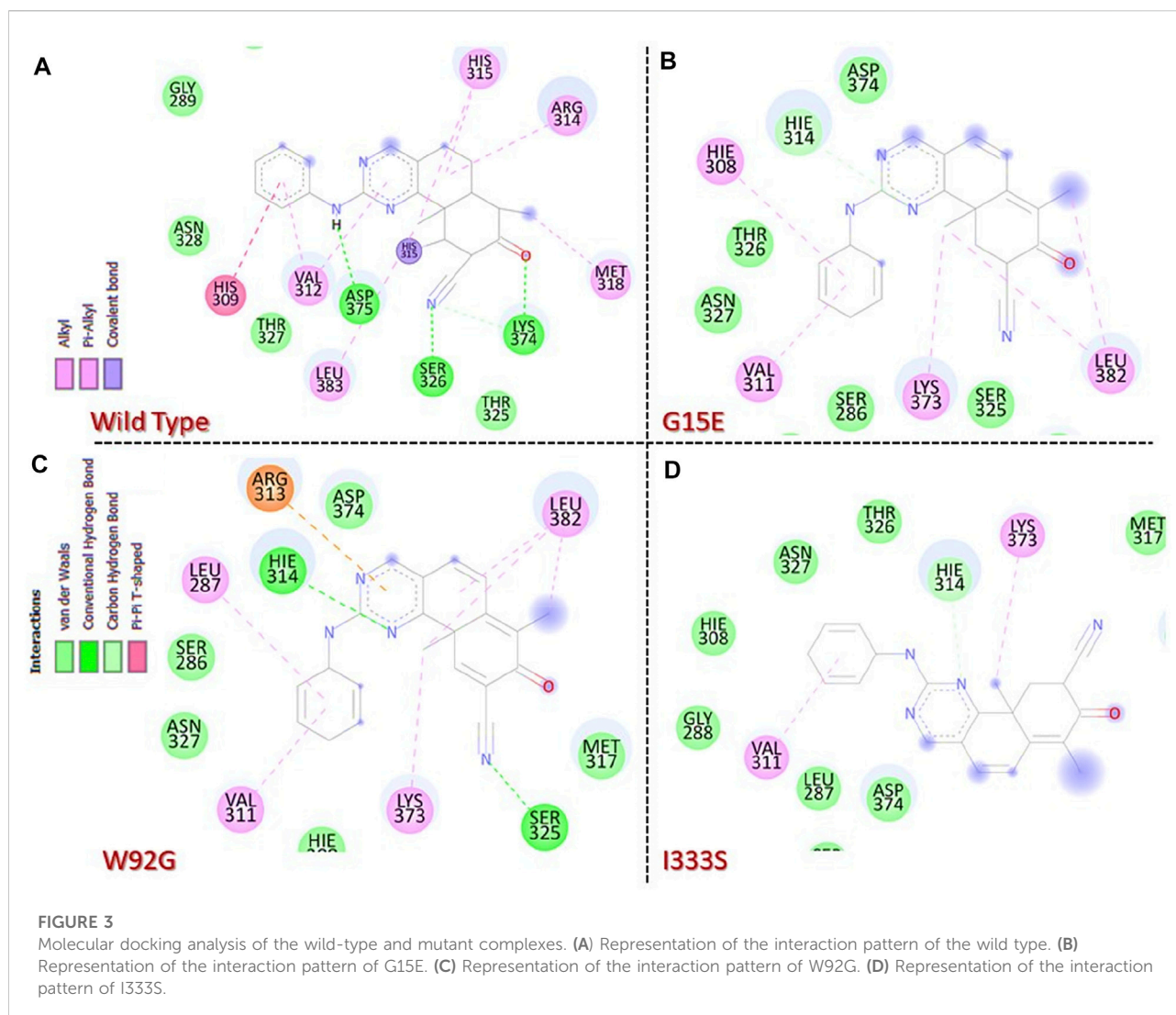
3.3 Investigation of the dynamic behavior of the wild-type and mutant complexes

To provide worthy insights into the impact of any particular mutation on the structure and function of a protein comprehension of key dynamic features is essential. For instance, dynamic stability [root mean square deviation (RMSD)] can be used to estimate the stability of a biological complex in a dynamic environment. To assess the structural stability, we herein also calculated RMSD as a function of time. The wild type and the three mutants were compared and are shown in Figures 4A–C. The wild type presented more stable behavior than the three mutants. As shown in Figure 4A, the wild type initially demonstrated a higher RMD for a short period

(2–6 ns); thereafter, the complex equilibrated and attained stability at 2.5 Å. The RMSD continued to follow the same trend until 55 ns and then decreased to 2.1 Å until the end of simulation. No significant perturbation was observed during the simulation, and the average RMSD was calculated to be 2.30 Å. Comparatively, the G15E mutant initially demonstrated significant perturbations in the RMSD, particularly between 5 and 20 ns. The RMSD then stabilized for a period between 21 and 55 ns and then continued to increase gradually until the end of the simulation. The complex reported a higher RMSD between 56 and 73 ns, then abruptly decreased, and then increased back with major deviation until the end of simulation. The complex reached the complete stability state after 80 ns. This shows that this mutation, despite its location away from the binding cavity, allosterically affects the binding affinity by compromising the stability of the protein. Since it has been previously reported that mutations that alter the protein stability result in radical function, thus this corroborates with the current findings (Dehury et al., 2020a). The RMSD graph for G15E mutant is given in Figure 4A. Unlike the wild type and G15E, the W92G complex demonstrated significant structural instability from the start of the simulation. The RMSD during the first 60 ns reported minor deviations at different time intervals, thus resulting in continuous destabilization effects of mutation. The RMSD increased to 6.0 Å at 60 ns and then gradually decreased until 100 ns. An average RMSD for W92G was recorded to be 3.60 Å and is shown in Figure 3B. On the other hand, the I333S mutant was reported to be the most destabilizing mutation among the shortlisted top deleterious mutations. First, the complex reported significant deviations until 55 ns and then gradually increased the RMSD and followed the same pattern to demonstrate significant deviations until the end of the simulation. An average RMSD of 3.2 Å was calculated for the I333S complex and is shown in Figure 4C. It has been reported that mutations that increase the stability may also increase the binding while destabilizing mutations decrease the binding. The current findings strongly corroborate with the previous reports where the filtration of mutations to obtain the most deleterious mutations showed that R132C, R132G, R132H, R132L, and R132S decrease the stability of IDH1 (Kumar et al., 2018). The stability feature is also reported to be affected by mutations in other diseases, which results in deleterious effects (Dehury et al., 2020a). Thus, herein, the mutations demonstrated destabilizing effects in contrast to the wild type.

3.4 Protein structure packing analysis

An assessment of protein packing reveals information regarding the binding and unbinding events that occurred during simulation. These events are steered by the bonding of

FIGURE 3
Molecular docking analysis of the wild-type and mutant complexes. **(A)** Representation of the interaction pattern of the wild type. **(B)** Representation of the interaction pattern of G15E. **(C)** Representation of the interaction pattern of W92G. **(D)** Representation of the interaction pattern of I333S.

small molecules with the protein cavity. For instance, this approach has been used previously to see the impact of mutations on the structural compactness when IDH1 binds to its homodimer (Yuan et al., 2017). Herein, to understand the structural compactness, radius of gyration (Rg) was calculated as a function of time. Consistent with the RMSD results, the wild-type complex reported a uniform pattern of Rg during the simulation. There was a slight increase in Rg between 40 and 60 ns, but it then stabilized again until the end of the simulation. The average Rg for this complex was calculated to be 22.5 Å. On the other hand, for the G15E mutant, despite its similar Rg value, , wild-type deviations at different time intervals were observed. This trend can be seen for the whole simulation time period (0–100 ns), which shows maximum unbinding events induced by the mutation. The W82G and I333S mutants demonstrated a similar pattern of Rg for the first 40 ns. With slightly higher Rg

values, the two complexes reported patterns similar to those of the wild type. However, W92G experienced a continuous increase in the Rg between 41 and 60 ns and then a decreased back effect was observed until 70 ns. The Rg value for the remaining simulation time remained lower; however, major deviations were reported. On the other hand, I333S reported a gradual increase in the Rg value after 45 ns and continued to follow this pattern until 75 ns. The Rg then again decreased and remained consistent until the end of the simulation. The average Rg values for W92G and I333S were calculated to be 22.80 Å and 22.78 Å, respectively. Interestingly the other reported mutations in the interface site, that is, R132C, R132G, R132H, R132L, and R132S, also demonstrated higher radius of gyration values; thus, further validating our findings in terms of protein compactness (Yuan et al., 2017; Bendahou et al., 2020). The Rg graphs for each complex are shown in Figures 5A–C.
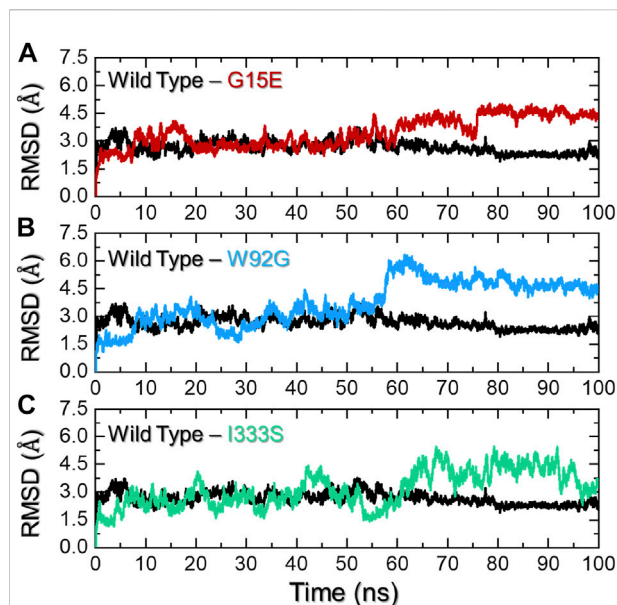
FIGURE 4
Dynamic stability analysis of the wild-type and mutant complexes. **(A)** Representation of the RMSD of the wild type and G15E. **(B)** Representation of the RMSD of the wild type and W92G. **(C)** Representation of the RMSD of the wild type and I333S. The x-axis represents time in nanoseconds while the y-axis represents RMSD in Å.

FIGURE 5
Radius of gyration analysis of the wild-type and mutant complexes. **(A)** Representation of the Rg of the wild type and G15E. **(B)** Representation of the Rg of the wild type and W92G. **(C)** Representation of the Rg of the wild type and I333S. The x-axis represent time in nanoseconds while the y-axis represent Rg in Å.
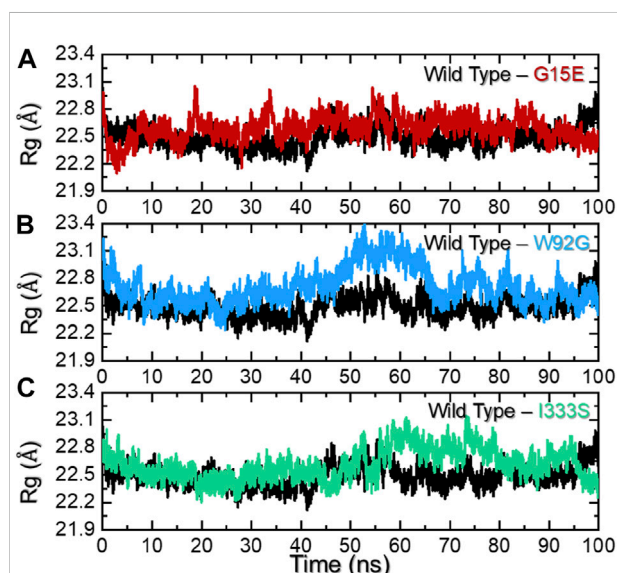
3.5 Residues' flexibility indexing

Residues' flexibility indexing is a key assessment to understand the role of each residue in different biological functions. The flexibility can be applied in different domains, such as molecular recognition, drug binding, cascade signaling, protein coupling, enzyme engineering, and protein designs. To estimate the residual flexibility of each complex, we calculated root mean square fluctuation (RMSF). As shown in Figure 5, all of the complexes demonstrated a more similar pattern of RMSF except for the regions between 125 and175 for all of the complexes. The flexibility at this portion (125–175) revealed a different pattern, which shows the impact of a particular mutation on the protein's internal dynamics. The wild type in this region demonstrates minimal RMSF, while the three mutations reported maximum RMSF. This consequently shows the altered dynamics and residue flexibility by the induced mutations in the structure, thus altering the binding of small molecules and function. This overall higher flexibility with the loss of compactness and intramolecular hydrogen bonds makes these mutations more deleterious than the other mutations. The findings are prevalidated by the previous literature where increased flexibility was observed mediated by different mutations. Moreover, other diseases mutations are reported to decrease the binding either due to increasing the cavity space or affecting the on/off switch, which consequently increases/decreases the distance between essential atoms (Dehury et al., 2020b; Dehury et al., 2020c). The RMSF of each complex is shown in Figure 6.

3.6 Hydrogen bonding analysis

Macromolecular complexes, particularly protein–protein coupling, are primarily driven by numerous factors, among which hydrogen bonding and hydrophobic contacts are essential. The environment of protein interfaces is enriched with water molecules that work with the residues to form hydrogen bonds (Chen et al., 2016). The mechanisms underlying protein–protein interaction as well as the ramifications for hydrogen bonding are unclear (Chodera and Mobley, 2013). Whether hydrogen bonds govern protein–protein docking, in particular, is a long-standing concern, and the mechanism is poorly understood (Patil et al., 2010; Olsson et al., 2011). Therefore, it is important to understand the hydrogen bonding landscape in protein–protein association. For instance, previously, hydrogen bonding was predicted to estimate the strength of the association between two molecules. As shown in Figure 7A, the wild type reported an average of 215 hydrogen bonds, while the G15E complex reported 212 hydrogen bonds over the simulation time. On the other hand, the W92G complex reported 209 average hydrogen bonds in contrast to the wild type. The hydrogen bonds in the wild type and I333S were comparable. In the case of I333S, average hydrogen bonds were calculated to be 215,
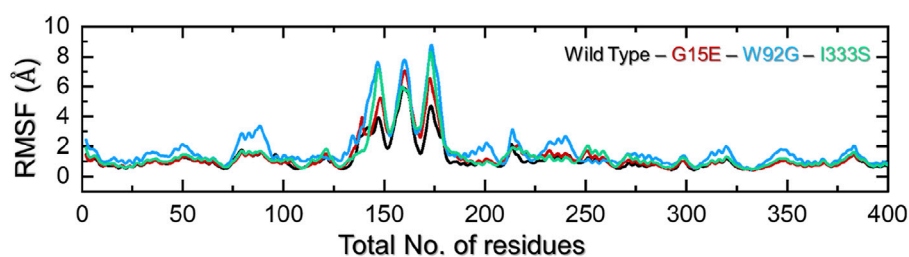
FIGURE 6
Residues' flexibility analysis of the wild-type and mutant complexes. The x-axis represents time in total number of residues while the y-axis represent RMSF in Å.
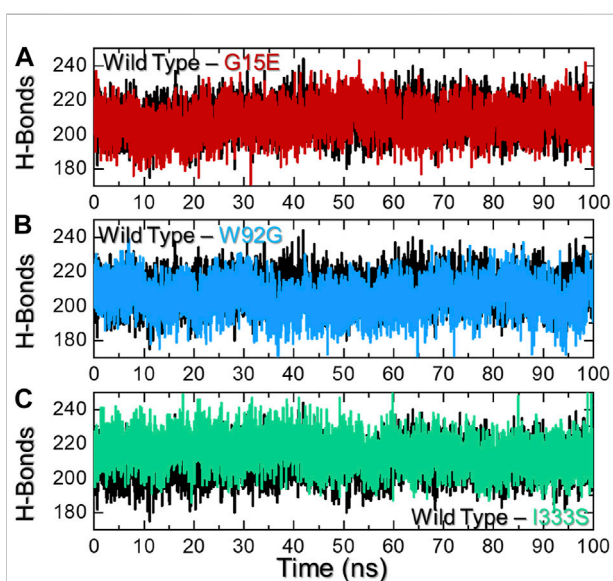

FIGURE 7
Hydrogen bonding analysis of the wild-type and mutant complexes. **(A)** Representation of H-bonds of the wild type and G15E. **(B)** Representation of H-bonds of the wild type and W92G. **(C)** Representation of H-bonds of the wild type and I333S. The x-axis represents time in nanoseconds, while the y-axis represents H-bond population.
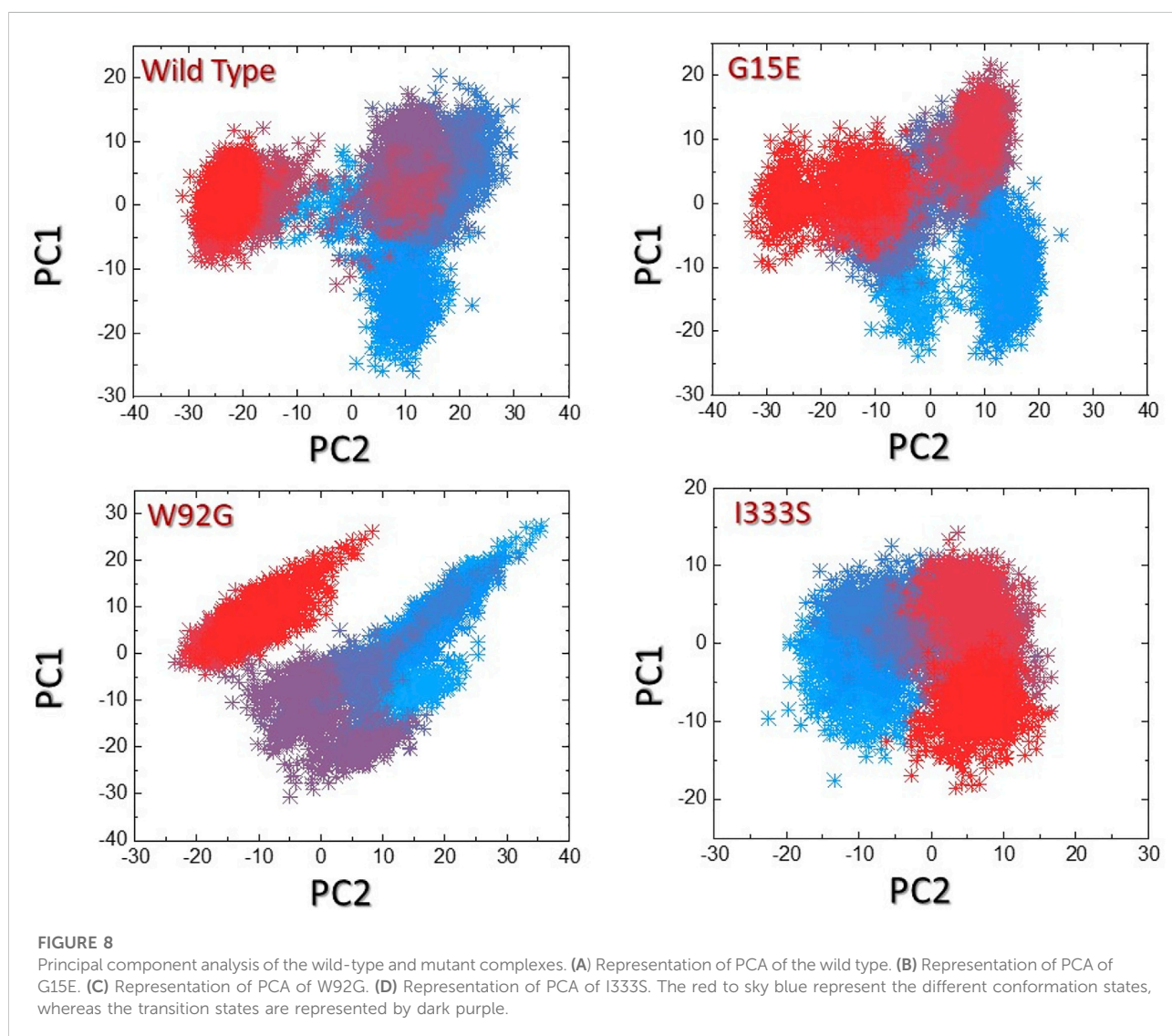
reliability than the conventional docking and alchemical methods. It is a widely used approach to explore the interaction strength and reveal the key binding feature, which steers the overall binding. Considering the applicability of the MM/GBSA approach, we also estimated the total binding energy for the wild-type and mutant complex. As shown in Table 3, the vdW for the wild type was estimated to be −40.78 ± 0.045 kcal/mol, while for the mutant it was −35.13 ± 0.054 kcal/mol reported by G15E mutations, −38.07 ± 0.053 kcal/mol reported by W92G, and −36.46 ± 0.06 kcal/mol calculated for the I333S mutant. This shows the loss of important interacting contacts that remained conserved in the wild type. On the other hand, the electrostatic energy determined an inverted trend. In the case of wild type, the electrostatic energy was calculated to be 3.55 ± 0.034 kcal/mol, while for the mutations −1.23 ± 0.057 kcal/mol (G15E) and −2.53 ± 0.052 kcal/mol (W92G), and −0.99 ± 0.06 kcal/mol for I333S mutant was calculated. This shows that due to these mutations the electrostatic contacts are increased, which may consequently alter the binding. Moreover, ΔG total was reported to be −34.77 ± 0.036 kcal/mol for the wild type and −34.07 ± 0.051 kcal/mol for W92G. The total binding energy of the wild type and G15E is comparable. The two mutants, that is, G15E and I333S, demonstrated a significant decrease in the ΔG total. The ΔG total for the G15E mutant was calculated to be −30.90 ± 0.041 kcal/mol, while for I333S, the ΔG total was estimated to be −31.91 ± 0.04 kcal/mol. This consequently shows the impact of these deleterious substitutions on the binding of the small molecule. The other parameters of the total binding free energy are given in Table 3.

same as that of the wild type. This demonstrate that these mutations alter the internal dynamics, consequently altering the hydrogen bonding network and causes resistance to the drug. The hydrogen bonding graphs of all of the complexes are shown in Figures 7A–C.

3.7 Binding free energy calculation

Binding free energy calculation determines the accurate binding strength and conformation of the small molecule. It is an essential estimation to re-evaluate the docking predictions by considering the highest accuracy and

3.8 Clustering of protein's motion

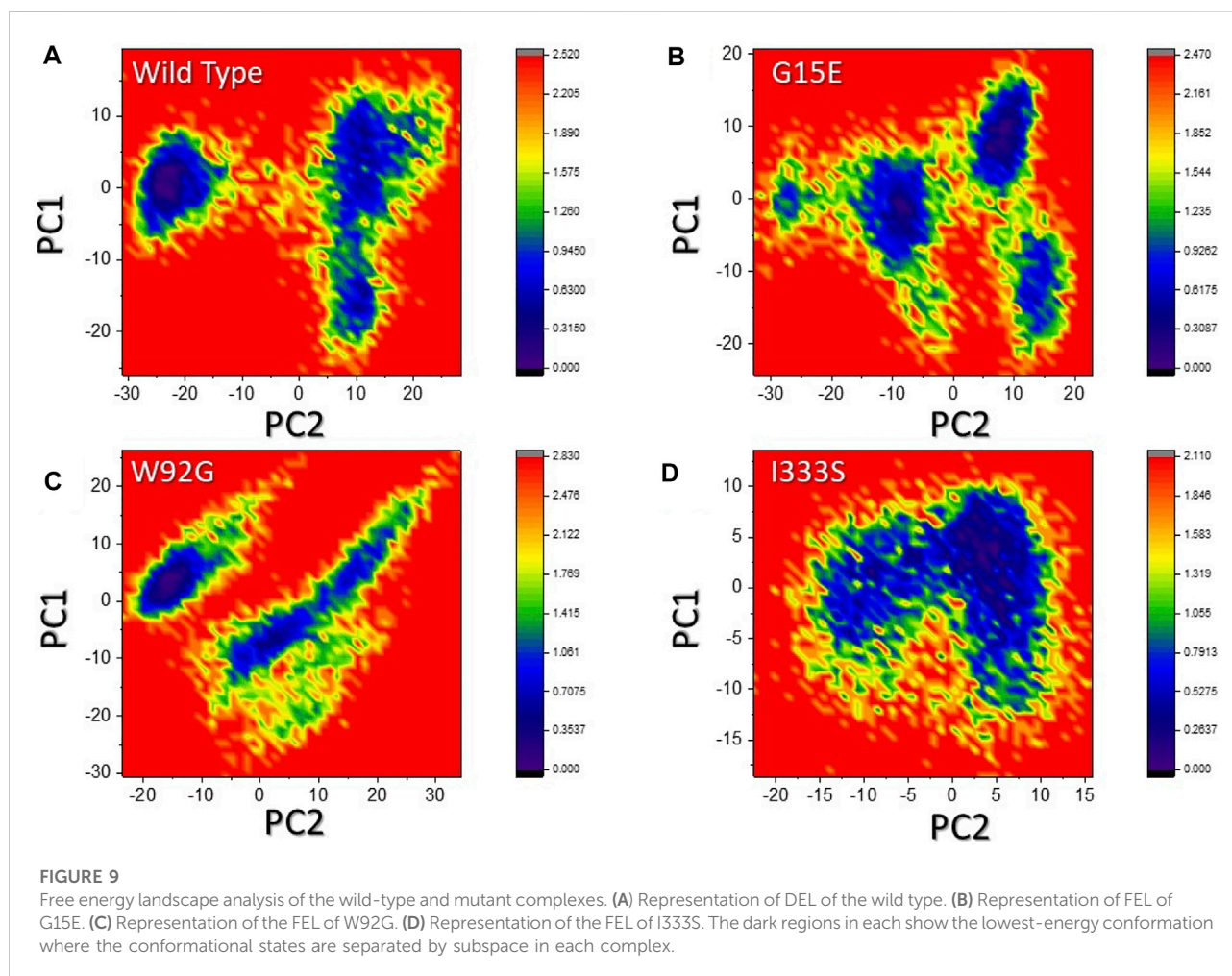We used the principal component analysis (PCA) to cluster the protein motions in the simulation trajectories.

TABLE 3 Total binding free energy for the wild-type, G15E, W92G, and I333S complexes. All of the values are calculated in kcal/mol.

Parameters	Wild type	G15E	W92G	I333S
VDWAALS	−40.78 ± 0.045	−35.13 ± 0.054	−38.07 ± 0.053	−36.46 ± 0.06
EEL	3.55 ± 0.034	−1.23 ± 0.057	−2.53 ± 0.052	−0.99 ± 0.06
EGB	7.22 ± 0.023	9.64 ± 0.040	10.55 ± 0.056	9.73 ± 0.04
ESURF	−4.77 ± 0.003	−4.18 ± 0.007	−4.02 ± 0.005	−4.19 ± 0.008
Delta G Gas	−37.23 ± 0.037	−36.36 ± 0.046	−40.61 ± 0.071	−37.45 ± 0.005
Delta G Solv	2.45 ± 0.023	5.46 ± 0.043	6.53 ± 0.056	5.53 ± 0.005
Delta Total	−34.77 ± 0.036	−30.90 ± 0.041	−34.07 ± 0.051	−31.91 ± 0.04

FIGURE 8
Principal component analysis of the wild-type and mutant complexes. (A) Representation of PCA of the wild type. (B) Representation of PCA of G15E. (C) Representation of PCA of W92G. (D) Representation of PCA of I333S. The red to sky blue represent the different conformation states, whereas the transition states are represented by dark purple.

The two PCs, that is, PC1 and PC2, reflected these motions in two dimensions. The distributed principal components for each complex are given in Figures 8A–D. The first three eigenvectors contributed 54% of the total motion, while the rest was contributed by the other eigenvectors. In contrast, the three mutants, that is, G15E the first three eigenvectors contributed 49%, W92G reported 48%, and I333S reported 44% of the total motion by the first three eigenvectors. The rest

FIGURE 9
Free energy landscape analysis of the wild-type and mutant complexes. **(A)** Representation of DEL of the wild type. **(B)** Representation of FEL of G15E. **(C)** Representation of the FEL of W92G. **(D)** Representation of the FEL of I333S. The dark regions in each show the lowest-energy conformation where the conformational states are separated by subspace in each complex.

demonstrated localized motion by these complexes. The conformational transition can be easily separated from each other by sky blue and red colors. This shows that these mutations had caused different arrangements to bind small molecules compared to the wild type, where the total internal motion is decreased, thus consequently producing minimal binding effects.

3.9 Free energy landscape analysis

The two PCs were then mapped to identify the stable and metastable states for each complex. As shown in Figures 9A–D, the wild type attained one lowest-energy conformation, while G15E and W92G attained two lowest-energy conformations. I333S also attained a single lowest-energy state. The conformational transition in each complex is separated by a subspace. This shows that the mutations had caused a different arrangement to bind small molecules

compared to the wild type, thus consequently producing minimal binding effects.

4 Conclusion

The current study used genomic mutation screening and molecular simulation methods to identify the most detrimental mutations in the IDH1 gene. The investigation of a large number of mutations revealed that three mutations, G15E, W92G, and I333S, are the most deleterious and highly destabilizing, which can affect the binding of a drug. These mutations primarily affect the binding of the drug with IDH, thus consequently reducing the efficacy of the already approved drug. Further validations such as molecular docking and dynamics simulation demonstrated that these mutations do not only affect the stability but also alter the bonding network. In addition, the BFE was also observed to have been reduced due to conformational changes mediated by these mutations. In sum, the current mutations contribute to drug resistance in glioma, and the atomic features

explored in this study could be used for structure-based drug designing against resistant glioma.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

Conceptualization: MSu, SU-I-H, AK, YW, and D-QW; data curation: MSu, SU-I-H, MSa, MA, and AI; formal analysis: MSu, SU-I-H, AK, and ZH; funding acquisition: YW; investigation: MSa, AI, ZH, and SA; methodology: MS, AK, AI, ZH, and SA; project administration: AK, SA, and YW; resources: MA and YW; software: D-QW; supervision: ZH and D-QW; validation: MA; visualization: SA; writing—original draft: AK, YW, and D-QW; and writing—review and editing: D-QW.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.927570/full#supplementary-material

References

Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 76, 7. doi:10.1002/0471142905.hg0720s76

Agnihotri, S., Burrell, K. E., Wolf, A., Jalali, S., Hawkins, C., Rutka, J. T., et al. (2013). Glioblastoma, a brief review of history, molecular genetics, animal models and novel therapeutic strategies. *Arch. Immunol. Ther. Exp.* 61, 25–41. doi:10.1007/s00005-012-0203-0

Balsera, M. A., Wriggers, W., Oono, Y., and Schulten, K. (1996). Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100, 2567–2572. doi:10.1021/jp9536920

Bendahou, M. A., Arrouchi, H., Lakhlili, W., Allam, L., Aanniz, T., Cherradi, N., et al. (2020). Computational Analysis of IDH1, IDH2, and TP53 mutations in low-grade gliomas including oligodendrogliomas and astrocytomas. *Cancer Inf.* 19, 1176935120915839. doi:10.1177/1176935120915839

Bendl, J., Musil, M., Štourač, J., Zendulka, J., Damborský, J., and Brezovský, J. (2015). *PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions*. Plos computational biology.

Capriotti, E., and Fariselli, P. (2017). PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Res.* 45, W247–W252. doi:10.1093/nar/gkx369

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S., and Savidge, T. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci. Adv.* 2 (3), e1501240. doi:10.1126/sciadv.1501240

Chodera, J. D., and Mobley, D. L. (2013). Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys.* 42, 121–142. doi:10.1146/annurev-biophys-083012-130318

Consortium, U. (2015). UniProt: A hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi:10.1093/nar/gku989

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092. doi:10.1063/1.464397

Dehury, B., Somavarapu, A. K., and Kepp, K. P. (2020). A computer-simulated mechanism of familial Alzheimer's disease: Mutations enhance thermal dynamics and favor looser substrate-binding to γ-secretase. *J. Struct. Biol.* 212, 107648. doi:10.1016/j.jsb.2020.107648

Dehury, B., Tang, N., and Kepp, K. P. (2020). Insights into membrane-bound presenilin 2 from all-atom molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 38, 3196–3210. doi:10.1080/07391102.2019.1655481

Dehury, B., Tang, N., Mehra, R., Blundell, T. L., and Kepp, K. P. (2020). Side-by-side comparison of Notch- and C83 binding to γ-secretase in a complete membrane model at physiological temperature. *RSC Adv.* 10, 31215–31232. doi:10.1039/d0ra04683c

Dimitrov, L., Hong, C. S., Yang, C., Zhuang, Z., and Heiss, J. D. (2015). New developments in the pathogenesis and therapeutic targeting of the IDH1 mutation in glioma. *Int. J. Med. Sci.* 12, 201–213. doi:10.7150/ijms.11047

Eberhardt, J., Santos-Martins, D., Tillack, A. F., and Forli, S. (2021). AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *J. Chem. Inf. Model.* 61, 3891–3898. doi:10.1021/acs.jcim.1c00203

Ernst, M., Sittel, F., and Stock, G. (2015). Contact-and distance-based principal component analysis of protein dynamics. *J. Chem. Phys.* 143, 244114. doi:10.1063/1.4938249

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117

Frezza, C., Tennant, D. A., and Gottlieb, E. (2010). IDH1 mutations in gliomas: When an enzyme loses its grip. *Cancer Cell* 17, 7–9. doi:10.1016/j.ccr.2009.12.031

Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 51, 69–82. doi:10.1021/ci100275a

Huang, L. E. (2019). Friend or foe—IDH1 mutations in glioma 10 years on. *Carcinogenesis* 40, 1299–1307. doi:10.1093/carcin/bgz134

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., and De Bakker, P. I. (2008). Snap: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24, 2938–2939. doi:10.1093/bioinformatics/btn564

Junaid, M., Shah, M., Khan, A., Li, C.-D., Khan, M. T., Kaushik, A. C., et al. (2018). Structural-dynamic insights into the *H. pylori* cytotoxin-associated gene A (CagA) and its abrogation to interact with the tumor suppressor protein ASPP2 using decoy peptides. *J. Biomol. Struct. Dyn.* 37, 4035–4050. doi:10.1080/07391102.2018.1537895

Karczewski, K., and Francioli, L. (2017). *The genome aggregation database (gnomAD)*. Melbourne: MacArthur Lab.

Khan, A., Junaid, M., Li, C.-D., Saleem, S., Humayun, F., Shamas, S., et al. (2020). Dynamics insights into the gain of flexibility by Helix-12 in ESR1 as a mechanism of resistance to drugs in breast cancer cell lines. *Front. Mol. Biosci.* 6, 159. doi:10.3389/fmolb.2019.00159

Khan, A., Khan, M. T., Saleem, S., Junaid, M., Ali, A., Ali, S. S., et al. (2020). Structural Insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Comput. Struct. Biotechnol. J.* 18, 2174–2184. doi:10.1016/j.csbj.2020.08.006

Khan, A., Khan, S., Saleem, S., Nizam-Uddin, N., Mohammad, A., Khan, T., et al. (2021). Immunogenomics guided design of immunomodulatory multi-epitope subunit vaccine against the SARS-CoV-2 new variants, and its validation through *in silico* cloning and immune simulation. *Comput. Biol. Med.* 133, 104420. doi:10.1016/j.compbiomed.2021.104420

Kumar, D. T., Sneha, P., Uppin, J., Usha, S., and Doss, C. G. P. (2018). Investigating the influence of hotspot mutations in protein–protein interaction of IDH1 homodimer protein: A computational approach. *Adv. Protein Chem. Struct. Biol.* 111, 243–261. doi:10.1016/bs.apcsb.2017.08.002

Landis, J. R., Williams, D. A., Lucia, M. S., Clauw, D. J., Naliboff, B. D., Robinson, N. A., et al. (2014). The MAPP research network: Design, patient characterization and operations. *BMC Urol.* 14, 58–17. doi:10.1186/1471-2490-14-58

Ohgaki, H., and Kleihues, P. (2013). The definition of primary and secondary glioblastoma. *Clin. Cancer Res.* 19, 764–772. doi:10.1158/1078-0432.CCR-12-3002

Olsson, T. S., Ladbury, J. E., Pitt, W. R., and Williams, M. A. (2011). Extent of enthalpy–entropy compensation in protein–ligand interactions. *Protein Sci.* 20, 1607–1618. doi:10.1002/pro.692

Patil, R., Das, S., Stanley, A., Yadav, L., Sudhakar, A., and Varma, A. K. (2010). Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PloS one* 5, e12029. doi:10.1371/journal.pone.0012029

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philosophical Mag. J. Sci.* 2, 559–572. doi:10.1080/14786440109462720

Pires, D. E., Ascher, D. B., and Blundell, T. L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30, 335–342. doi:10.1093/bioinformatics/btt691

Rodrigues, C. H., Pires, D. E., and Ascher, D. B. (2018). DynaMut: Predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Res.* 46, W350–W355. doi:10.1093/nar/gky300

Roe, D. R., and Cheatham, T. E., III (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095. doi:10.1021/ct400341p

Rose, P. W., Beran, B., Bi, C., Bluhm, W. F., Dimitropoulos, D., Goodsell, D. S., et al. (2010). The RCSB protein data bank: Redesigned web site and web services. *Nucleic Acids Res.* 39, D392–D401. doi:10.1093/nar/gkq1021

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* 23, 327–341. doi:10.1016/0021-9991(77)90098-5

Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An overview of the Amber biomolecular simulation package. *WIREs. Comput. Mol. Sci.* 3, 198–210. doi:10.1002/wcms.1121

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40, W452–W457. doi:10.1093/nar/gks539

Tan, A. C., Ashley, D. M., López, G. Y., Malinzak, M., Friedman, H. S., and Khasraw, M. (2020). Management of glioblastoma: State of the art and future directions. *Ca. Cancer J. Clin.* 70, 299–312. doi:10.3322/caac.21613

Wirsching, H.-G., and Weller, M. (2017). Glioblastoma. *Malignant Brain Tumors*, 265. Handbook of clinical neurology by Elsevier.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemom. intelligent laboratory Syst.* 2, 37–52. doi:10.1016/0169-7439(87)80084-9

Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., et al. (2009). IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* 360, 765–773. doi:10.1056/NEJMoa0808710

Yuan, S., Chan, H. S., and Hu, Z. (2017). Using PyMOL as a platform for computational drug design. *WIREs Comput. Mol. Sci.* 7, e1298. doi:10.1002/wcms.1298

Zhao, S., Lin, Y., Xu, W., Jiang, W., Zha, Z., Wang, P., et al. (2009). Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1α. *Science* 324, 261–265. doi:10.1126/science.1170944

Zhou, W., and Wahl, D. R. (2019). Metabolic abnormalities in glioblastoma and metabolic strategies to overcome treatment resistance. *Cancers* 11, 1231. doi:10.3390/cancers11091231

Zwanzig, R. (1973). Nonlinear generalized Langevin equations. *J. Stat. Phys.* 9, 215–220. doi:10.1007/bf01008729