The Protein Journal (2022) 41:27–43
https://doi.org/10.1007/s10930-022-10041-0

Most Probable Druggable Pockets in Mutant p53-Arg175His Clusters Extracted from Gaussian Accelerated Molecular Dynamics Simulations

Morad Mustafa[1] · **Mohammed Gharaibeh**[1]

Accepted: 5 January 2022 / Published online: 31 January 2022

Abstract
p53, a tumor suppressor protein, is essential for preventing cancer development. Enhancing our understanding of the human p53 function and its modifications in carcinogenesis will aid in developing more highly effective strategies for cancer prevention and treatment. In this study, we have modeled five human p53 forms, namely, inactive, distal-active, proximal-active, distal-Arg175His mutant, and proximal-Arg175His mutant forms. These forms have been investigated using Gaussian accelerated molecular dynamics (GaMD) simulations in OPC water model at physiological temperature and pH. Our observations, obtained throughout 200 ns of production run, are in good agreement with the relevant results in the classical molecular dynamics (MD) studies. Therefore, GaMD method is more economic and efficient method than the classical MD method for studying biomolecular systems. The featured dynamics of the five human p53-DBD forms include noticeable conformational changes of L1 and α1–β5 loops as well as β6–β7 and β7–β8 turns. We have identified two clusters that represent two distinct conformational states in each p53-DBD form. The free-energy profiles of these clusters demonstrate the flexibility of the protein to undergo a conformational transition between the two clusters. We have predicted two out of seven possible druggability pockets on the clusters of the Arg175His forms. These two druggability pockets are near the mutation site and are expected to be actual pockets, which will be helpful for the compound clinical progression studies.

Keywords p53 monomer · Arg175His mutation · Free-energy analysis · Gaussian accelerated molecular dynamics · Principal component analysis · Probable druggable pockets

1 Introduction

Worldwide, lung cancer is the leading cause of cancer deaths in men, whereas breast cancer is the leading cause of cancer deaths in women [1, 2]. To understand the molecular biology of cancer, we have to be able to identify and understand the underlying mutations. The most frequently altered gene in human cancers is *TP53*, which is either directly inactivated by somatic mutations in about 50% of these human cancers or indirectly inactivated in the remainder through binding to viral proteins or having impaired pathways [3, 4].

The p53 protein functions primarily as a transcription factor that can either activate or repress the expression of a large number of DNAs and microRNAs [5, 6] and as a mediator for integrating appropriate cellular signals via protein–protein and protein–DNA interactions [7]. When the p53 protein fails to function properly, uncontrolled growth and division of cells arises causing genomic instability, a sign of cancer [4]. Therefore, the p53 protein is essential for preventing cancer development via complicated interactions, which are mediated by p53 independently folded and fundamentally disordered functional domains [8].

The monomeric form of the p53 protein consists of 393 amino acid residues and has a modular domain structure consisting of three major functional domains: the N-terminal domain (NTD), which functions mainly as a transcription-activation domain [8], the core domain, which functions as a sequence-specific DNA binding domain (DBD) [9], and the C-terminal domain (CTD), which functions as a modulator for the transcriptional activity of p53 protein and as a binding domain to different target proteins or nonspecific DNA [8].

The boundaries of the major p53 domains and their subdomains are demonstrated in Fig. 1 according to the discovered p53 isoforms and the pinpointed structural-functional

✉ Morad Mustafa
morad.mustafa@gmail.com

[1] Department of Chemistry, School of Science, The University of Jordan, Amman 11942, Jordan

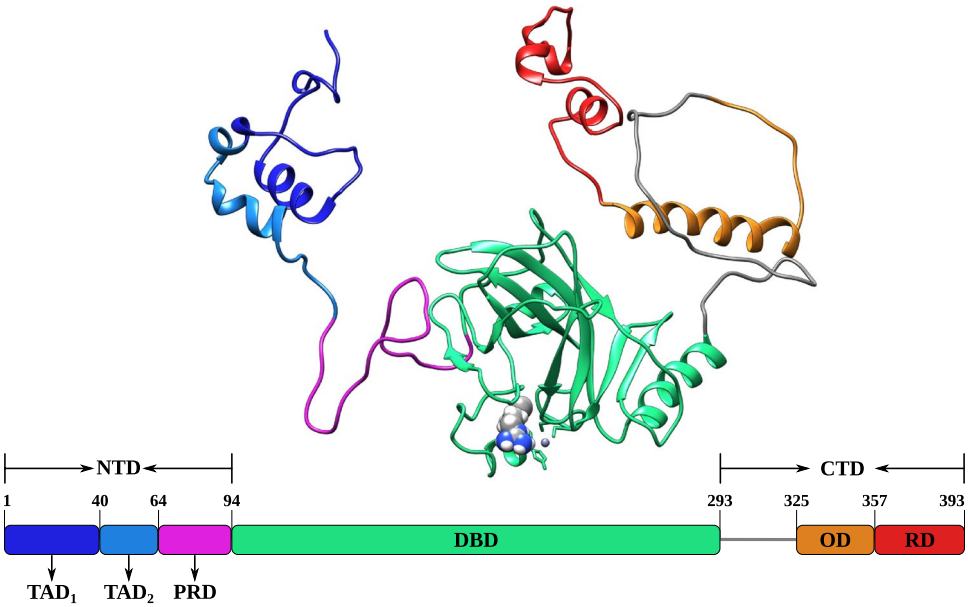

Fig. 1 3D structure of full-length wild-type p53 model. Colors show domains and subdomains: blue, transcription-activation domain 1 (TAD_1); azure, transcription-activation domain 2 (TAD_2); magenta, proline-rich domain (PRD); green, DNA binding domain (DBD); gray, flexible linker; orange, oligomerization domain (OD); red, regulatory domain (RD). The Arg175 is depicted with CPK representation. This model refers to the equilibrated distal-active p53 conformation (Color figure online).

features of human p53 protein [8, 10–18]. The p53-NTD, spanning residues (1–93) [8], contains 3 subdomains: the transcription-activation domain 1 (TAD_1), spanning residues (1–39) [11, 14], the transcription-activation domain 2 (TAD_2), spanning residues (40–60) [8], and the proline-rich domain (PRD), spanning residues (64–93) [8]. The p53-DBD has spanning residues (94–292) [18]. The p53-CTD, spanning residues (293–393) [8], contains 2 subdomains: the oligomerization domain (OD), spanning residues (325–356) [10], and the regulatory domain (RD), spanning residues (358–393) [12, 13, 15–17].

It is known that a biologically active form of the human p53 protein is formed by a homotetramer comprising four identical chains [5]. Accordingly, a model of full-length p53 bound to DNA has been proposed [19]. In the proposed model, two p53-RD units are far from the DNA (distal p53-RD) and the other two are close to the DNA (proximal p53-RD). In a cryo-electron microscopy study [20], the relative arrangement of p53-OD and p53-DBD structures has been revealed; consequently, this orientation facilitates modeling of the natively folded domains.

The full-length structure of a wild-type human p53 protein is not only crucial for understanding the role of p53 protein in the cell cycle and in other activities, but also it is necessary in the design of drugs that target mutant forms of the p53 protein [5]. In breast cancer, the top five substitution mutations in the p53 protein with their frequencies are Arg175His (185), Arg248Gln (156), Arg273His (141), Arg248Trp (122), and Arg273Cys (74) [21]. The Arg175His mutant is one of the designated hotspot mutations in the p53 protein, and the hotspot mutations of Arg175 are classified as conformational mutants, which fail to stabilize the β sandwich in the human p53-DBD; consequently, these mutants lack the appropriate scaffold for the proper interaction with DNA [8]. In addition, the Arg175His mutant perturbs the zinc-binding region, which causes structural instability, rapid exchange between folded and unfolded states, and hydrophobic aggregation at physiological temperature [8, 22, 23].

The ability of the p53 protein to prevent cell proliferation (i.e., suppressing tumor development) is considered a promising therapeutic goal, which can be achieved through the induction of an irreversible exit from the cell cycle or activation of cell death [6]. Small molecules and elimination of mutant p53 are among the approaches that have been adopted in p53-based cancer therapies [4, 24]. The p53-targeted therapies are considered attractive cancer therapies; however, the p53 protein is still a challenging target for drug discovery because p53 does not offer the accessibility of a receptor-ligand interaction or an enzyme active site; thus, there are challenges in the development of p53-targeted therapy [4, 25].

Gaussian accelerated molecular dynamics (GaMD) facilitates unconstrained-enhanced sampling of a biomolecular system by adding a harmonic boost potential to smooth the system's potential energy surface [26]. Relative to the classical molecular dynamics (MD) simulations, the GaMD simulations demonstrate acceleration of biomolecular kinetics [27]. The novel GaMD method has several applications in sampling biomolecular systems and free-energy calculations of biomolecules, such as, predicting drug-receptor interactions [28], deciphering the mechanism of the G-protein-coupled receptors and the G protein interactions [29], and predicting the interactions between the G-protein-coupled receptors and the membrane lipids [30].

All together, the full-length human p53 proteins, wild-type and Arg175His mutant forms, can be used as reference systems in comparing the GaMD results versus the classical MD results. Therefore, in this study we have built p53 forms by utilizing the relative orientation of p53-OD to p53-DBD [20]; thus, enhancing the credibility and realisticity of our forms to be solvated with an OPC water model and be studied by GaMD technique at physiological temperature and pH. In addition, the principal component analysis (PCA) technique has been applied on the GaMD results, and the most probable druggable pockets in the Arg175His mutant forms have been predicted.

2 Methods

2.1 Molecular Modeling

The starting structures of the wild-type p53-DBD forms were taken from chain B of PDB entry 2OCJ [31], an inactive (DNA-free) structure, and from chain A of PDB entry 4HJE [32], an active structure. An initial structure for a wild-type p53-OD form was chosen from the first conformer of chain D of PDB entry 3SAK [33]. In the next three paragraphs, the modeling steps were done using UCSF Chimera [34], version 1.15, and its built in Modeller, version 9.23, functions [35].

The two PDB structures of p53-DBD were superposed along with the PDB structure of p53-OD on a Cryo-EM structure, the PDB entry 5XZC [20], using C_α atoms. Superposition on chain B of PDB entry 5XZC generated two models, inactive and proximal-active models, whereas superposition on chain C of PDB entry 5XZC generated distal-active model. After the superposition process, a flexible loop (FL) between a p53-DBD and the p53-OD structures was modeled. Because of the natural disordered state of human p53-TAD, p53-PRD, and p53-RD [5], these parts were modeled at Robetta web server [36, 37], a de novo protein-structure-prediction server. The generated p53-TAD, p53-PRD, and p53-RD models were superposed on each p53-DBD–p53-OD model using C_α atoms. Loop refinement was applied on each model; thus, three full-length p53 forms were obtained, namely, inactive, distal-active, and proximal-active p53 forms.

Two mutant forms were generated using wild-type distal-active and proximal-active p53 forms. The mutation process, mutating Arg175 residue to His175 residue, was achieved with the pdb4amber program in Amber 20 [38–40]. To remove residue–residue clashes because of the mutation process, minimization in vacuum was applied on each mutant. The minimization process consisted of 50 cycles; 20 cycles of steepest-descent energy minimization process followed by 30 cycles of conjugate-gradient energy minimization process.

To fulfill the physiological pH condition (i.e., pH 7.4) [41] in each full-length p53 form, Asp and Glu residues were deprotonated, Arg and Lys residues were protonated, and His residues kept neutral. To model the coordinated Zn^{2+} complex state [42], three Cys residues were deprotonated (i.e., Cys176, Cys238, and Cys242 were assigned a −1 charge). This zinc model, with the protonation state of −1, was found to be the only model that kept the zinc interface structure intact (i.e., maintaining the tetrahedral structure) [42]. The detailed atomic charge calculations for the amino acids residues in the zinc model, using B3LYP density functional method with the 6-311+G** basis set, were reported [42]. Both N- and C-termini of each full-length p53 form were capped with acetyl (ACE) and N-methyl (NME) groups, respectively to keep them neutral. The ACE and NME groups were obtained from RCSB Protein Data Bank [43, 44].

The LEaP module of Amber 20 [38–40] was used for adding missing atoms, applying ff14SB protein force field [45], solvating the protein in a truncated octahedron box of OPC water molecules [46] with a buffering distance set to 12.0 Å, loading Li/Merz ion parameters (12–6–4 set) for monovalent [47] and +2 [48] ions in the designated water model, and neutralizing the modeled system with charge neutralizing counter ions (i.e., Na^+ ions). The Li/Merz ion parameters were selected because they were designed for the monovalent, divalent, trivalent, and tetravalent ions [38–40]; that is, the same force field was used for all ions. Unlike the widely used TIP3P water model [49], the OPC water model was recommended to be used with ff14SB protein force field in biomolecular simulations [40]. Combining OPC water model with ff14SB protein force field in MD simulations has improved the accuracy of atomistic simulations and given better modeling sequence-specific behavior, protein mutations, and rational protein-design results [50].

2.2 GaMD Simulations

Gaussian accelerated molecular dynamics simulations were performed using PMEMD engine within the Amber 20 software suite [38–40]. The total preparation simulation time for each system with positional and distant restraints was 11.5 ns (Table 1). In Table 1, system preparation protocol included several conjugate-gradient energy minimization processes, heating from 0 to 310.15 K, and several equilibration processes at different periodic boundary conditions.

The applied position restraints were relative to the initial coordinates of the modeled system. In the DNA-free form and at 300 K [51], it was observed that the Zn^{2+} binding site did not have a stable structure; the spontaneous Zn^{2+} release leads to dissociation of His179 from Zn^{2+}; thus,

Table 1 Summary of system preparation protocol for active and mutant forms

Process	Ensemble[b]	Restraint type [k_f[a]]					Steps
		Position			Distance with Zn		
		Protein	PB_1[c]	PB_2[d]	Cys-SG	His-ND1	
Minimization 1	NVT	60.0					2000
Minimization 2	NVT	40.0					1000
Minimization 3	NVT	20.0					1000
Minimization 4	NVT		20.0		56.36	51.53	1000
Minimization 5	NVT			20.0	56.36	51.53	2000
Heating[e]	NV			20.0	56.36	51.53	500000
Equilibration 1[e]	NVT			20.0	56.36	51.53	100000
Equilibration 2[e]	NPT			20.0	56.36	51.53	100000
Equilibration 3[e]	NPT			15.0	56.36	51.53	100000
Equilibration 4[e]	NPT			10.0	56.36	51.53	100000
Equilibration 5[e]	NPT			5.0	56.36	51.53	100000
Equilibration 6[f]	NPT			1.0	56.36	51.53	100000
Equilibration 7[f]	NPT			0.5	56.36	51.53	150000
Equilibration 8[f]	NPT				56.36	51.53	5000000

[a]k_f: force constant in units of kcal/(mol · Å^2)

[b]The symbols in this column stand for; *N:* constant number of atoms, *V*: constant volume, *T:* constant temperature, and *P:* constant pressure

[c]PB_1: Protein backbone atoms

[d]PB_2: p53-DBD and p53-OD backbone atoms

[e]Time step was assigned to 1 fs

[f]Time step was assigned to 2 fs

increasing the thermal fluctuation of L2 loop. Therefore and to maintain four-ligand coordinations of zinc in the protein at physiological temperature and during long-time simulations, the distance restraint approach was chosen [42, 52], but the bond lengths were selected relative to the bond lengths in the PDB structure of the modeled system, and the force constant parameters were interpolated from the zinc Amber force field data [53]. For active and mutant forms, the utilized force constant for Cys-SG-Zn bond was 56.36 kcal/(mol · Å^2), and for His-ND1-Zn bond was 51.53 kcal/(mol · Å^2). For inactive form, the utilized force constant for Cys-SG-Zn bond was 25.93 kcal/(mol · Å^2), and for His-ND1-Zn bond was 58.98 kcal/(mol · Å^2).

The temperature was maintained at physiological value of 310.15 K using Langevin dynamics [54] with a collision frequency of 5 ps^{-1}. The pressure was maintained at 1 bar with isotropic position scaling using Berendsen barostat [55] with pressure relaxation time of 1 ps. The nonbonded cutoff was assigned to 9.0 Å. The Particle Mesh Ewald method [56] with its default parameters was used to calculate the full electrostatic energy of the unit cell in a macroscopic lattice of repeating images. The SHAKE algorithm [57] was used in all simulation processes, but not in minimization processes, to constrain hydrogen atoms. Consequently, the time step was assigned to 2 fs for dynamics integration except at specific processes mentioned in Table 1, where it was assigned to 1 fs to maintain system stability.

In Table 1, the last equilibration process was necessary for conducting the GaMD simulation [26]. A dual boost on both dihedral and total potentials was applied on the last equilibration process and production runs. The last equilibration process involved 4 stages: 2.0 ns of preparatory stage as conventional MD (no statistics were collected), 3.0 ns of initial stage as conventional MD (potential statistics were calculated for GaMD pre-equilibration stage), 2.0 ns of GaMD pre-equilibration stage as preparation biasing MD simulation (boost potential was applied but boost parameters were not updated), and 3.0 ns of GaMD equilibration stage as biasing MD simulation (boost potential was applied and boost parameters were updated). The average and standard deviation of potential energies were calculated every 0.5 ns, and the rest of GaMD parameters were assigned their default values. To enhance sampling simulations and free energy calculation of biomolecules, GaMD was applied on each system for a simulation time of 200 ns. The GaMD production trajectory files were written every 2.0 ps.

2.3 Data Analysis

R language [58], version 3.6.3, python language [59], version 3.7.3, cpptraj [60], version 4.25.6, and pytraj [60], version 2.0.6, were used for composing analysis scripts and for generating analysis figures. Analysis scripts included root-mean-square displacement (RMSD), root-mean-square fluctuation (RMSF), radius of gyration, PCA, clustering, and free-energy profiles.

2.3.1 Root-Mean-Square Deviation

RMSD reflects the degree of similarity between three-dimensional protein structures. It can be computed [61–63] by measuring the RMSD between backbone atoms of superimposed protein structures (the protein structure in trajectory frame i and the restart frame from the last equilibration process).

2.3.2 Root-Mean-Square Fluctuation

RMSF reveals the conformational variance of the protein. It can be measured [61–63] by calculating the deviation between the position of atom i (usually C_α) with respect to its average position over the whole simulation trajectory.

2.3.3 Radius of Gyration

In biological molecules, radius of gyration indicates the protein structure compactness [64]. It can be determined [61–63] by measuring the root mean square distance from protein atoms (usually C_α) to their center of mass. Among the major protein classes (i.e., α, β, α/β, and $\alpha+\beta$) and when protein size is larger than 300 amino acid residues, α proteins have the highest radius of gyration indicating the least tight packing character as compared with the character of other classes, whereas α/β proteins have the lowest radius of gyration indicating the tightest packing character as compared with the character of other classes [64]. Maintaining a relatively steady value of radius of gyration over time reveals the stability of the protein folding state.

2.3.4 Principal Component Analysis

Conducting PCA reveals the most important motions of a biological system over a broad range of time and spatial scales [65]. PCA is a multivariate statistical technique that reduces the number of dependent motions needed to describe the dynamics of a biological system into a smaller number of independent motions called principal components [65]. The first principal component, the eigenvector with the highest corresponding eigenvalue, reflects the most identifying motion patterns in the simulation [65]. The eigenvalues show the contribution of the corresponding eigenvectors to the global fluctuations of a biological system.

In our GaMD simulations, PCA was performed [61–63] on all heavy atoms of the protein after removing all global translations and rotations about the center of mass and orienting all structures with respect to the restart frame from the last equilibration process. The porcupine plots were generated using the porcupine plot plugin in VMD [66].

2.3.5 Clustering Analysis

Clustering analysis is a technique that finds patterns within data by locating clusters of geometrically similar conformers in ensembles of chemical conformations [67]. Most clustering algorithms measure *distance* between objects to compute the dissimilarity matrix; thus, clustering algorithms can be divided into partitional and hierarchical clustering methods.

In this study, k-means clustering method [61–63] has been used because of its one of the fastest and most widely used techniques, and it has apparent good performance in analyzing MD trajectory data [67]. The k-means clustering method aims at dividing n observations into non-overlapping k clusters in which each observation belongs to the cluster with the closest centroid, and each centroid depicts the conformation that best represents the conformations within a cluster [67]. The quality of a k-means partition is evaluated by calculating the percentage as indicated by Eq. 1 [67], where *BSS* stands for between sum of squares, *TSS* stands for total sum of squares, and *CQ* stands for cluster quality.

$$CQ = \frac{BSS}{TSS} \times 100\% \tag{1}$$

The higher the percentage, the better the score (and thus the quality). Thus, the optimal number of clusters for a k-means approach has been determined by NbClust function [68] built in R language [58]. The utilized arguments in NbClust function included: *scores* of the supplied coordinates on the PCs as dataset, *euclidean* as the distance measure to be used to compute the dissimilarity matrix, *k-means* as the cluster analysis method to be used, *silhouette* as the index to be calculated. In k-means function, *Hartigan-Wong* was used as an algorithm for k-means calculations.

2.3.6 Free-Energy Analysis

The GaMD method facilitates unconstrained-enhanced sampling of a biomolecular system by adding a harmonic boost potential to smooth the system potential energy surface [26]. By constructing a harmonic boost potential that follows a Gaussian distribution, potential of mean force (PMF) (i.e., a free-energy profile) can be extracted by accurate reweighting of the GaMD simulations [26].

In the reweighting method, cumulant expansion can be used to approximate the ensemble-averaged Boltzmann factor, $\langle e^{\beta \Delta V(r)} \rangle$, where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann constant, and T is the system absolute temperature [69]. To recover the most accurate free-energy profile, let $\Delta V(r)$ be the added non-negative boost energy to the system when the system potential is lower than a reference energy, where r denotes the atomic positions [69], and $\langle e^{\beta \Delta V(r)} \rangle_i$ be the ensemble-averaged Boltzmann factor of $\Delta V(r)$ for simulation frames found in the ith bin. Thus, the cumulant expansion to the third order is given by Eq. 2 [70].

$$\begin{aligned} \langle e^{\beta \Delta V(r)} \rangle_i &= \left[\exp\left(\sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right) \right]_i \\ &= \left[\exp\left(\beta C_1 + \frac{\beta^2}{2} C_2 + \frac{\beta^3}{6} C_3 + \ldots \right) \right]_i \end{aligned} \tag{2}$$

where the first three cumulants can be calculated by

$$C_1 = \langle \Delta V(r) \rangle_i \tag{3}$$

$$C_2 = \langle \Delta V^2(r) \rangle_i - \langle \Delta V(r) \rangle_i^2 \equiv \sigma^2_{\Delta V(r)} \tag{4}$$

$$C_3 = \langle \Delta V^3(r) \rangle_i - 3 \langle \Delta V^2(r) \rangle_i \langle \Delta V(r) \rangle_i + 2 \langle \Delta V(r) \rangle_i^3 \tag{5}$$

where σ is the standard deviation of the $\Delta V(r)$ distribution in the ith bin. For a GaMD simulation of a biomolecular system, the probability distribution along a reaction coordinate $A(r)$ is denoted as $p[A(r)]$, which can be used to calculate the biased PMF, which is denoted as $F^b[A(r)]$, for each bin i as illustrated in Eq. 6.

$$F_i^b[A(r)] = -\frac{1}{\beta} \ln \left\{ p_i[A(r)] \right\} \tag{6}$$

Finally, the reweighted PMF, which is denoted as $F[A(r)]$, can be evaluated by Eq. 7 for each bin i.

$$F_i[A(r)] = F_i^b[A(r)] - \frac{1}{\beta} \ln \left(\langle e^{\beta \Delta V(r)} \rangle_i \right) \tag{7}$$

To identify distinct low-energy states of our biomolecule systems, free energy profiles were obtained by performing reweighting method along PC1 and PC2. To obtain a reasonable bin resolution, a bin width of 4.0 was used.

2.3.7 Pocket Druggability Prediction

The druggable pockets in the Arg175His mutant forms were predicted using the protein druggability prediction method provided in the PockDrug server [71]. After feeding the structure of the designated protein to the PockDrug server, the pockets on that protein were estimated on the basis of the amino acid atoms that form the surface of potential binding cavities [71]. The selected estimation method, which is called fpocket estimation method, extracted all the possible pockets from the protein surface using spheres of varying diameters, then prioritized these pockets for compound development in computer-aided drug design [71].

3 Results

3.1 Model Quality Assessment

The quality of our models, which have been constructed prior to the energy minimization processes, has been assessed using SWISS-MODEL server tools [72–74] and Structure Analysis and Verification Server tools [75–79]. Assessment results (Tables S1–S3 and Fig. S1) demonstrate through quality estimate, molprobity results, structure analysis, and verification server tools that the three models are reliable and can be studied via GaMD simulations.

3.2 Stability of Molecular Structures

The molecular structures of the wild-type and the Arg175His mutant forms are shown in Fig. S2. In general, the RMSD results of the p53-DBD structures (see Fig. S3) show acceptable domain stability throughout the production run time scale, and are consistent with RMSD results obtained from classical MD simulation [80]. On the other hand, the RMSD results of the full-length p53 forms (data not shown) have high values (in the range of $5 - 20$ Å), which are attributed, as expected, to the free movements of both the p53-NTD and the p53-CTD [52]. Therefore, our next analyses are mainly focused on investigating the dynamics of p53-DBD forms in the presence of the p53-NTD and the p53-CTD.

The radius of gyration results of the p53-DBD structures (see Fig. S4) have steady values in all forms throughout the production run time scale. The steady values of the radius of gyration indicate acceptable stability in the p53-DBD folding state.

The RMSF results of the full-length p53 forms (see Fig. S5) confirm the high thermal fluctuations of the p53-NTD and the p53-CTD relative to the thermal fluctuations of the p53-DBD. As seen in Fig. S5, including the p53-NTD and the p53-CTD in the RMSF calculation will affect the RMSF values of the p53-DBD. On the other hand, RMSF results of only the p53-DBD forms (see Fig. 2) show changes in thermal fluctuations of the L1, α1–β5, and L3 loops as well as the β6–β7 turn. These highlighted thermal fluctuations agree with the observed thermal fluctuations in the classical MD study [81]. In addition, it is observed that the α1 helix in the middle of the L2 loop spontaneously unfolds

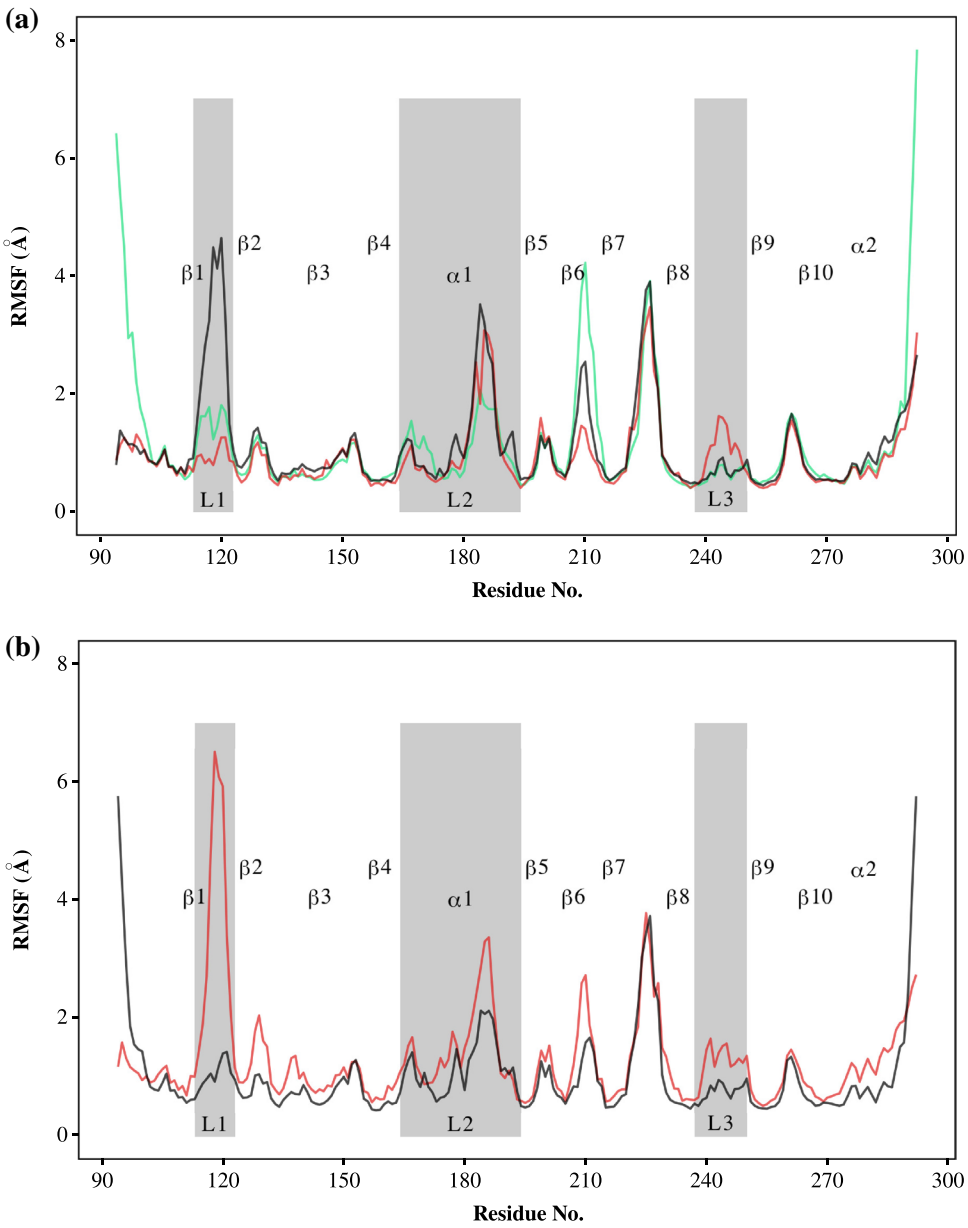

Fig. 2 Root-mean-square fluctuation (RMSF) of the p53-DBD forms: (**a**) inactive; green, distal-active; red, proximal-active; black. (**b**) Distal-Arg175His mutant; red, proximal-Arg175His mutant; black. The location of the secondary structures are labeled (Color figure online).

in all simulations, and this observation is consistent with the observation in the classical MD study at 300 K [51].

3.3 Dynamics of Protein Structures

Figure S6 shows the PCA scree plot, which indicates the proportion of variance against its mode index. The first two components together make up 45.4%, 20.4%, 29.5%, 38.2%, and 27.1% of the variance of the inactive, distal-active, proximal-active, distal-Arg175His mutant, and proximal-Arg175His mutant forms, respectively.

Most of the significant dynamics of p53-DBD forms can be captured by PC1 [65]; therefore, Figs. 3 and 4 show the porcupine plots of the PC1 obtained by performing PCA on GaMD trajectories of the p53-DBD forms.

In the inactive p53-DBD form, its dynamics illustrates drastic motions in the L1 loop as well as in the $\beta 6$–$\beta 7$ and $\beta 7$–$\beta 8$ turns (see Fig. 3a). On the other hand, the dynamics of the distal-active p53-DBD form contains moderate motions in the $\alpha 1$–$\beta 5$ loop and the $\beta 7$–$\beta 8$ turn (see Fig. 3b), whereas the dynamics of the proximal-active p53-DBD form embraces significant motions in the L1 loop as well as in the $\alpha 1$–$\beta 5$ loop and the $\beta 7$–$\beta 8$ turn (see Fig. 3c). In addition, the dynamics of the distal-Arg175His mutant form shows extreme motions in the L1 loop as well as in the $\beta 6$–$\beta 7$ and $\beta 7$–$\beta 8$ turns (see Fig. 4a), whereas the dynamics

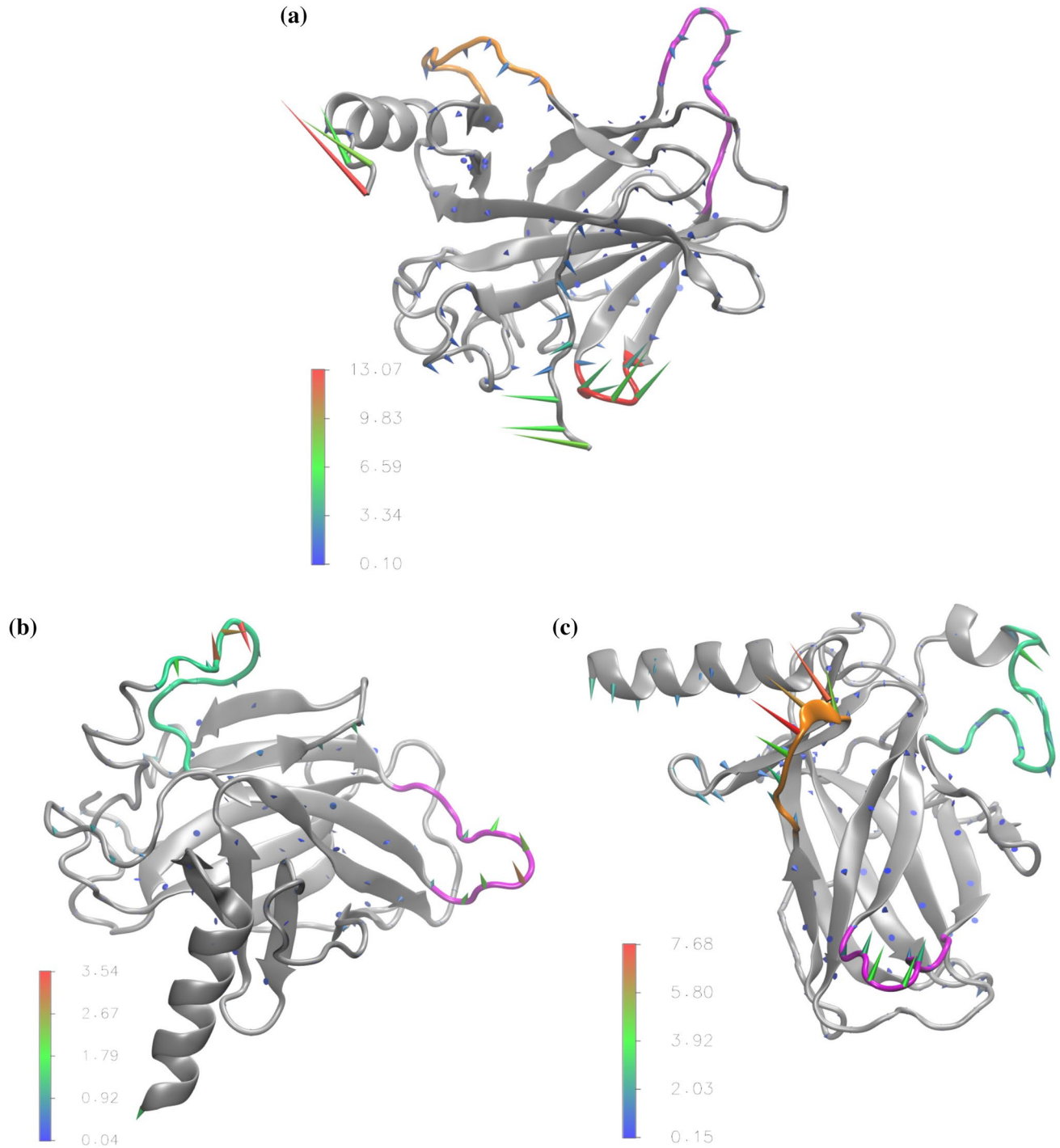

Fig. 3 Porcupine plot of the PC1 for the wild-type p53-DBD forms: **(a)** inactive, **(b)** distal-active, and **(c)** proximal-active. The arrows indicate direction of the eignvector and its magnitude in units of Å. Colors show the significant movements of the regions: orange, L1 loop; green, $\alpha 1$–$\beta 5$ loop; red, $\beta 6$–$\beta 7$ turn; magenta, $\beta 7$–$\beta 8$ turn (Color figure online).

of the proximal-Arg175His mutant form exhibits moderate motions in the $\beta 7$–$\beta 8$ turn (see Fig. 4b).

Projecting the trajectory frames onto the plane formed by PC1 and PC2 (see Fig. S7) reveals random diffusion in a high-dimensional harmonic potential. The observed patterns can be interpreted as thermal motion along a shallow free-energy landscape [82]. Even though the PCA is a powerful tool for finding global-correlated motions in atomic simulations of biomolecules, yet it does not partition the frames into distinct conformational states.

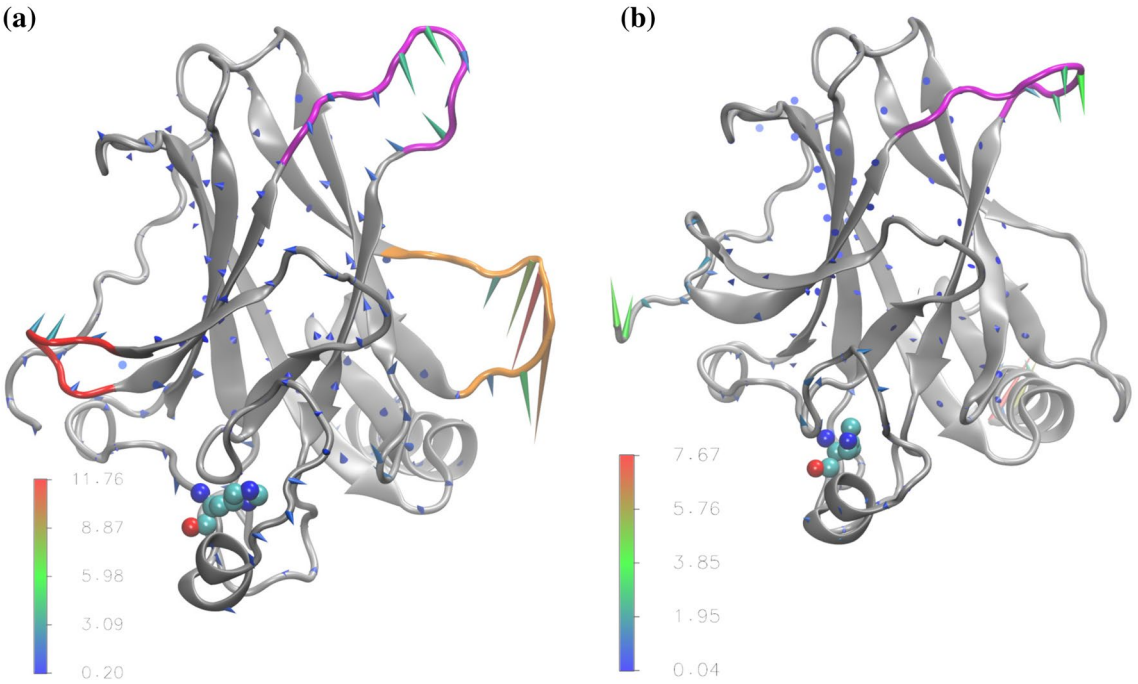

Fig. 4 Porcupine plot of the PC1 for the mutant p53-DBD forms: **(a)** distal-Arg175His mutant and (**b**) proximal-Arg175His mutant. The arrows indicate direction of the eignvector and its magnitude in units of Å. Colors show the significant movements of the regions: orange, L1 loop; red, β6–β7 turn; magenta, β7–β8 turn. The Arg175His mutation is depicted with CPK representation (Color figure online).

Contrarily, conformational states analysis, which can be achieved by clustering the PC data, allows comparisons of all conformers sampled during the apparent thermal diffusion. By using the NbClust function [68] built in R language [58], the PC data of each form consists of two clusters (see Fig. 5). This observation is also supported by the cluster quality results (see Fig. S8), which show an identified *kink* at cluster count of two. In addition, Fig. 5 illustrates that approximately 50% of PCA time scale corresponds to *cluster 1* in most p53-DBD forms.

3.4 Free Energy Profiles

The free energy profiles of the p53-DBD forms are shown in Fig. 6. Clearly, the inactive p53-DBD form is confined to an energetic well (see Fig. 6a) that is lower than those for the distal-active and proximal-active p53-DBD forms (see Fig. 6b and c, respectively). In addition, the energetic well for the distal-active p53-DBD form (see Fig. 6b) is lower than that for the proximal-active p53-DBD form (see Fig. 6c). On the other hand, the distal-Arg175His mutant form is confined to a lower energetic well (see Fig. 6d) than that for the proximal-Arg175His mutant form (see Fig. 6e).

3.5 Probable Druggable Pockets

Prediction of the protein ability to bind drug-like molecules with high affinity is an interesting approach that has been adopted in p53-based cancer therapies [25]. As seen in Fig. 5, there are two clusters for each Arg175His mutant form. Therefore, the most probable druggable pockets for each cluster are listed in Table S4. The first and the last frames of the GaMD trajectories have been extracted to represent cluster 1 and 2, respectively. Fig. 7 shows the probable druggable pockets on the mutant forms. Specifically, the probable druggable pocket on the distal or proximal Arg175His conformation, which has been constructed prior to the minimization processes, is shown in Fig. 7a, whereas the probable druggable pockets on each cluster of the Arg175His mutant forms are shown in Fig. 7b–e.

4 Discussion

It has been suggested that maintaining the stability of the Zn^{2+}-binding site decreases the thermal fluctuation of the L2 loop and prevents aggregation [51]. By maintaining the stability of Zn^{2+}-binding site, the structural and dynamical properties of the full-length human p53 as well as its

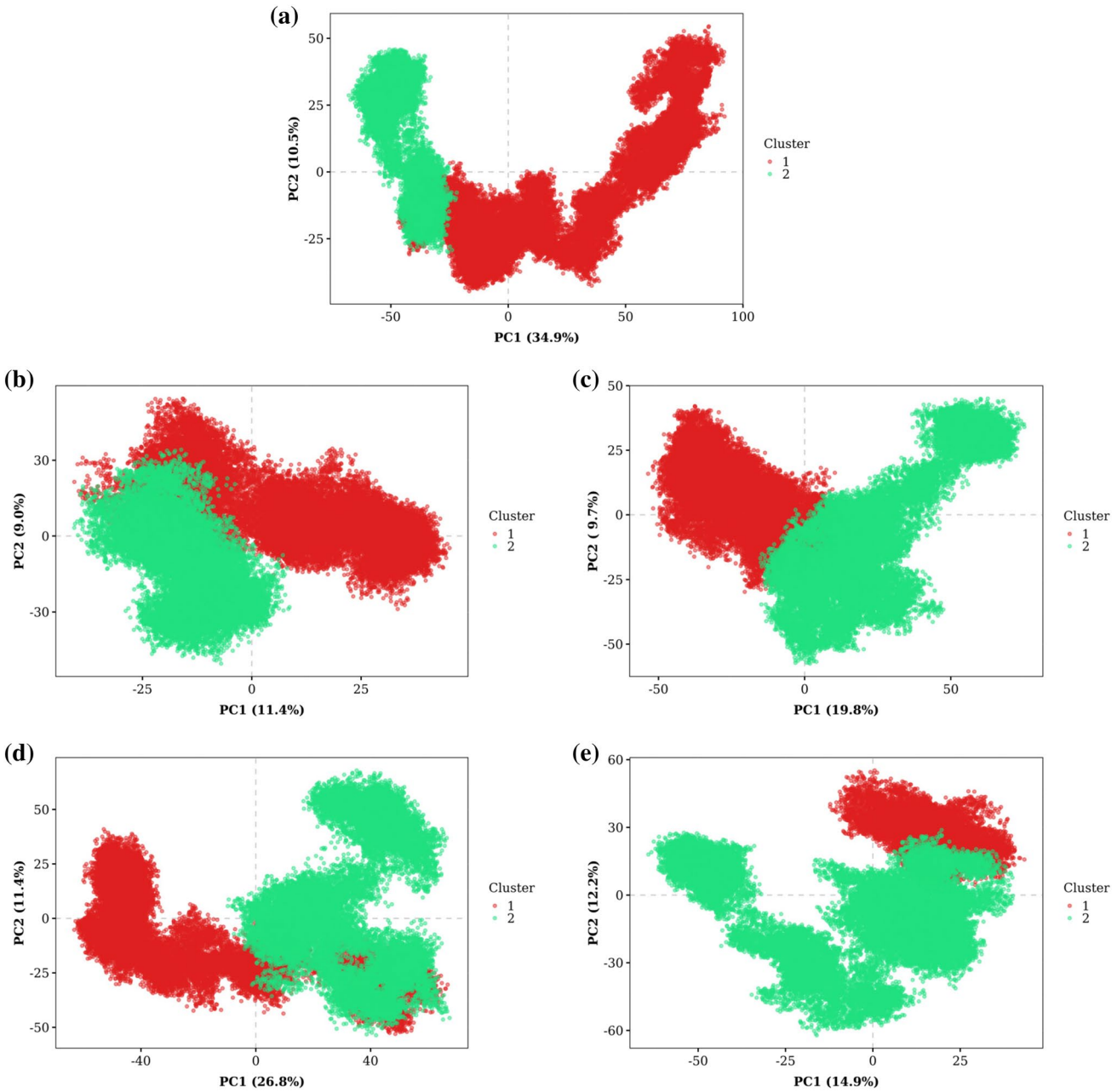

Fig. 5 Clustering results of k-means algorithm on subspace dimension projected on the 2D plane formed by the PC1 and PC2 of the p53-DBD forms: **(a)** inactive, **(b)** distal-active, **(c)** proximal-active, **(d)** distal-Arg175His mutant, and **(e)** proximal-Arg175His mutant.

interdomain interactions have been studied by the classical MD method at 300 K for 850 ns [52]. Another classical MD simulation, maintained the stability of the Zn^{2+}-binding site and used TIP3P water model, has been conducted on p53-DBD monomers to investigate the dynamics of the wild-type and some aggregating mutant forms, including the Arg175His mutant form, at 310 K for 500 ns [80].

In this work, we have conducted GaMD simulations with an OPC water model at physiological temperature and pH to study the dynamics of five full-length human p53 proteins, wild-type and Arg175His mutant forms, for 200 ns. The molecular structures of these forms are shown in Fig. S2 and the observed conformational changes are illustrated in Figs. 3, 4, 5. The noticeable conformational changes of the L1 loop in the inactive p53-DBD (see Fig. 3a), distal-active p53-DBD (see Fig. 3b), and distal-Arg175His mutant (see Fig. 3d) forms indicate high-conformational flexibility of the L1 loop [31, 42, 52, 81, 83, 84]. In a classical MD simulation [84], it has been reported that the L1 loop occupies two major conformational states; an extended and a recessed

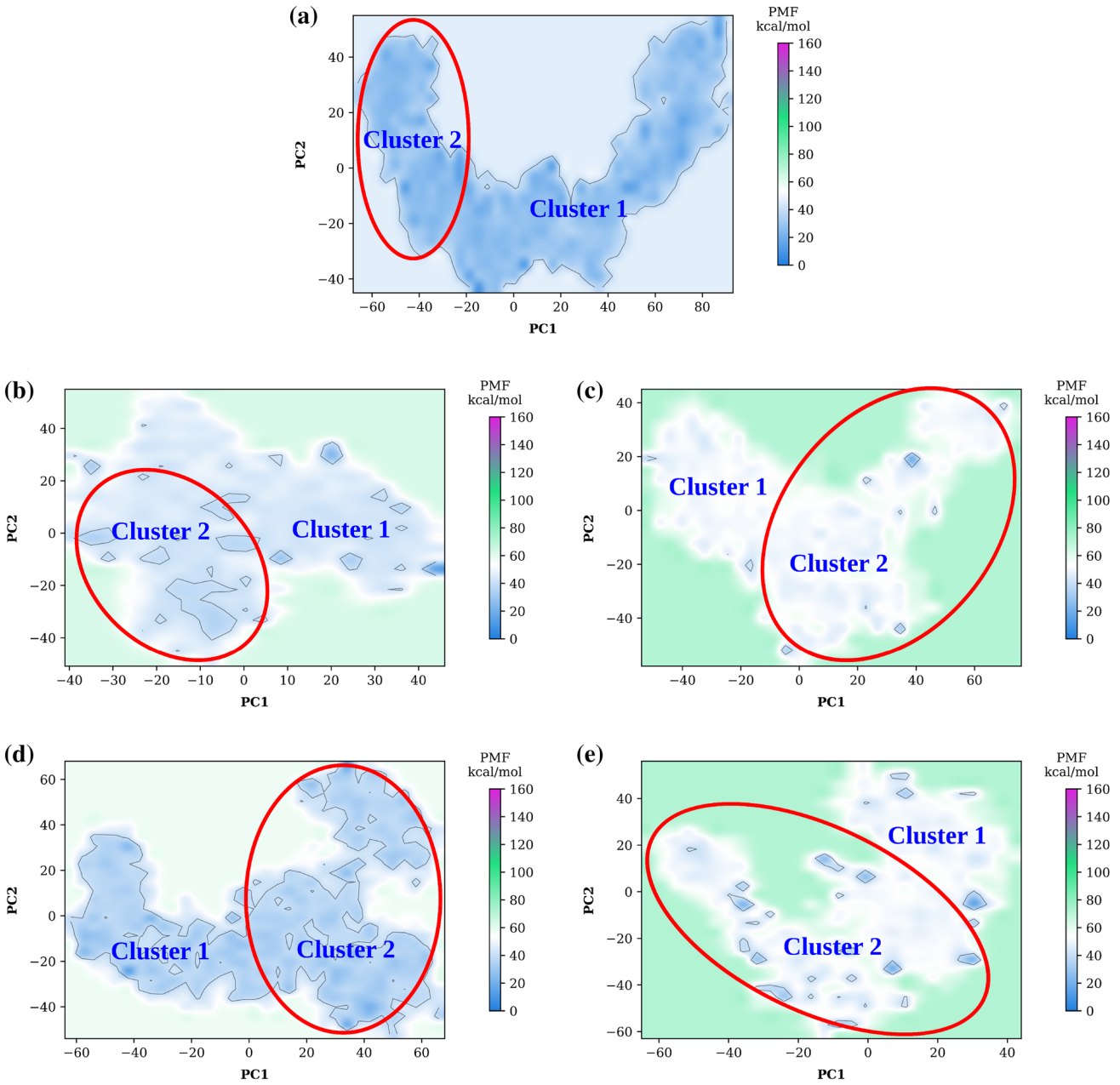

Fig. 6 Two-dimensional free energy profiles along PC1 and PC2 calculated for the p53-DBD forms: (**a**) inactive, (**b**) distal-active, (**c**) proximal-active, (**d**) distal-Arg175His mutant, and (**e**) proximal-Arg175His mutant.

states, which are crucial for binding to DNA elements. In our simulations, a change in the conformational states of the L1 loop has been observed in two forms, proximal-active and distal-Arg175His mutant forms as shown in Fig. S9, where the extended conformation is represented by a distance of ≈10 Å and the recessed conformation is represented by a distance of ≈25 Å. As seen in Fig. S9, the dominant L1 loop conformation in our simulation is the extended conformation. The observed thermal flexibility of the L1 loop can be quenched by forming hydrogen bonds between the backbone atoms of the L1 loop and the N-terminus atoms of the $\alpha 2$ helix in the presence of a DNA element [81, 84–86].

Another dynamical feature that has been observed in p53-DBD forms is the thermal motion of the $\alpha 1$–$\beta 5$ loop (part of the L2 loop) in the distal-active and proximal-active p53-DBD forms (see Fig. 3b and c, respectively). Our observations are consistent with those in an X-ray study [31] and other classical MD studies [51, 52, 81, 84]. The observed thermal flexibility of the $\alpha 1$–$\beta 5$ loop can be stabilized in the presence of dimer-dimer interface [19, 81, 86].

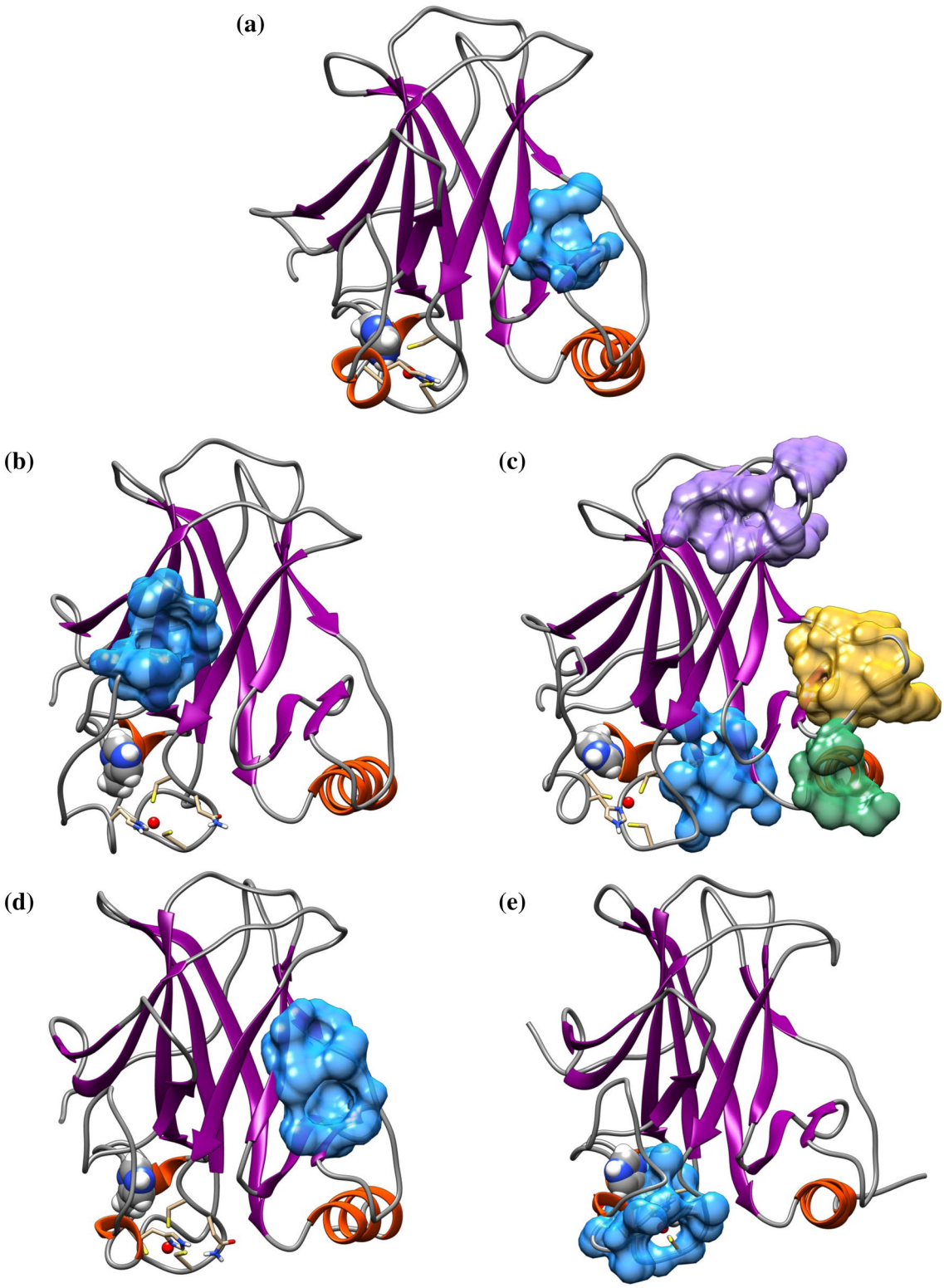

Fig. 7 Probable druggable pockets in the Arg175His mutant p53-DBD forms: **(a)** prior minimization processes, **(b)** cluster 1; distal conformation, (**c**) cluster 2; distal conformation, (**d**) cluster 1; proximal conformation, (**e**) cluster 2; proximal conformation. Colored surfaces show pockets: blue, pocket 1; green, pocket 2; gold, pocket 3; violet, pocket 4. The Arg175His mutation is depicted with CPK representation, and Zn^{2+} atom is colored in red color (Color figure online).

Among the other important outcomes of our simulations is the dynamics of the β6–β7 and β7–β8 turns. The β6–β7 turn has drastic motions in the inactive (see Fig. 3a) and distal-Arg175His mutant (see Fig. 4a) forms, whereas the β7–β8 turn has various motion strengths in all p53-DBD forms (see Figs. 3 and 4). The observed dynamics of the β6–β7 and β7–β8 turns agrees with that in classical MD studies [42, 80, 81]. It has been reported [80] that the β6–β7 turn exhibits two conformational states; open and closed states, which are related to the thermodynamic stability of the p53-DBD; the open state represents a destabilizing (or destabilized) state of the p53-DBD, whereas the closed state represents the stable state of the p53-DBD. Figs. S10 and S11 confirm the existence of these states in our simulations, where the closed state is represented by an Arg209CA–Asp259CA distance of ≈13 Å or an Arg209CZ–Asp259O distance of ≈10 Å and the open state is represented by an Arg209CA–Asp259CA distance of ≈23 Å or an Arg209CZ–Asp259O distance of ≈20 Å. Furthermore, the inactive and proximal-active p53-DBD forms explore both the open and closed states of the β6–β7 turn, whereas the distal-active p53-DBD form explores an intermediate state between the open and closed states. On the other hand, the distal-Arg175His mutant form explores both the open and closed states of the β6–β7 turn, whereas the proximal-Arg175His mutant form explores the closed state. The observed thermal flexibility of the β6–β7 turn can be quenched through formation of dimer-dimer interface [86] and hydrogen bonding with PRD region [52], whereas the thermal flexibility of the β7–β8 turn can be stabilized by forming dimer-dimer interface [81, 86].

These remarkable dynamics can be partitioned into distinct conformational states by performing clustering analysis on the PC data [67]; thus, allowing comparisons of all conformers sampled during the apparent thermal diffusion. As seen in Fig. 5, the PC data of each form consists of two clusters (distinct conformations). The presence of two clusters in each p53-DBD form is worth investigation by other analysis techniques, such as free energy analysis (see Fig. 6). The shape of the plots in Fig. 6 resembles the shape of the plots in Figs. 5 and S7; therefore, these free energy profiles represent the energy of the conformational clusters. Consequently, the flatness of the energetic wells in the free energy profiles (see Fig. 6) probably means an easy process for the protein to undergo a conformational change from cluster 1 to cluster 2. In addition, it is observed that the distal forms have lower free-energy profiles than those of the proximal forms. This observation might indicate, in general, that the distal conformation is more thermally stable than the proximal conformation. The proximal conformation is stabilized by the presence of the DNA elements [19].

The p53-DBD contains a region called aggregation prone region (APR), spanning residues (251–257) within the β9 strand [32], which participates in the aggregation behavior of the Arg175His mutant [87]. Our results (see Fig. 2, 3, 4) show very stable APR with no significant dynamical behavior. In addition, the solvent accessible surface area (SASA) of the β9 strand has been calculated using Connolly surface area method [88] and is shown in Fig. S12. There is no significant difference between the SASA of the wild-type forms (see Fig. S12a) and that of the Arg175His mutant forms (see Fig. S12b), and this observation agrees with that in the classical MD simulations study [80]. The similarity of the SASA results, Fig. S12a and b, might be attributed to the maintained stability of the Zn^{2+}-binding site in our simulations. Therefore, it is not expected to see aggregation behavior of the Arg175His mutant forms in our simulations.

The most frequently mutated residues in the human p53 protein are at or near the p53-DNA interface, specifically, the 2 large loops and the loop-sheet-helix motif (the L1 loop, the β2–β2'turn, the 4 C-terminal residues of β10 strand, and the α2 helix) [85, 89]. In the classical MD simulation study of Arg175His mutant [90], a hydrophobic patch has been suggested as a druggable site to prevent unfolding and aggregation by stabilizing the zinc binding region. On the basis of the two conformational Agr175His forms and the cluster states, we have predicted 7 druggable pockets on the Agr175His mutant forms (see Fig. 7b–e). These predicted pockets (Table S4) are different from the reported ones [80, 90]. To rule out the pseudo pockets, the probable druggable pockets on the active p53 forms are shown in Table S5 and Fig. S13. The pockets on the Agr175His forms (see Fig. 7a, c colored with yellow, and d) are expected to be pseudo pockets because they exist on the active p53 forms (see the blue colored pockets in Fig. S13a, b, and d). Moreover, the pocket on the Agr175His form (see Fig. 7b) is expected to be a pseudo pocket because it exists on the active p53 forms (see Fig. S13c and d colored with green). In addition, the pocket on the Agr175His form (see the violet colored pocket in Fig. 7c) is expected to be a pseudo pocket because it exists on the active p53 forms (see the yellow colored pocket in Fig. S13a). Consequently, there are three remaining pockets on the Agr175His form (see Fig. 7c colored with blue or green and e) that have to be clarified.

The two representative clusters of each Arg175His mutant forms are depicted with electrostatic potential surface (EPS) (see Fig. S14), whereas the EPS of the two representative clusters of each active p53-DBD forms are shown in Fig. S15. Clearly, there is more negative EPS and deeper cavity around the mutation point in the mutant forms than that in the active wild-type forms. Therefore, we expect that the pockets near the mutation point to have desirable docking affinity. Specifically, the first pocket in the Arg175His mutant forms (see the blue colored pockets in Fig. 7c and e) can be used in a drug screening study as one big pocket. Therefore, the blue colored pockets in Fig. 7c and e are the

unique pockets that are expected to be the targeted pockets in a drug screening study.

5 Conclusions

The novel GaMD method has several applications in sampling biomolecular systems and free-energy calculations of biomolecules. Five full-length human p53 forms have been investigated by GaMD simulations in OPC water model at physiological temperature and pH. our observations, obtained throughout 200 ns of production run, are in good agreement with the relavent results in the classical MD studies [52, 80, 84, 90]. Therefore, GaMD method is more economic and efficient method than the classical MD method for studying biomolecular systems.

The featured dynamics of the five human p53-DBD forms include noticeable conformational changes of the L1 and α1–β5 loops as well as the β6–β7 and β7–β8 turns. The observed thermal flexibility of these regions can be stabilized either by binding to DNA element or by forming dimer–dimer interface [19, 81, 84–86].

By a subsequent clustering analysis of the structural frames in the subspace spanned by PC1 and PC2, we have identified two clusters that represent two distinct conformational states. The free-energy profiles of these clusters in each p53-DBD form demonstrate the flexibility of the protein to undergo a conformational transition between the two clusters. However, the aggregation behavior of the Arg175His mutant forms is not expected to be observed because of the maintained stability of the Zn^{2+}-binding site in our simulations.

The utilized bonded approach for maintaining the Zn^{2+}-binding site in our simulation might be considered inappropriate technique for comparing the results of the wild-type and mutant forms of p53 protein. Therefore, we encourage using nonbonded approach for maintaining the Zn^{2+}-binding site by electrostatic and van der Waals interactions, which allow studying the catalytic function of the p53 protein for only a short period of simulation time scale [42].

By using a representative structure for each cluster of the Arg175His forms, we have predicted seven druggability pockets. Four druggability pockets on the Arg175His forms have been ruled out because there are similar pockets to them that exist in the active p53 forms. Furthermore, the druggability pockets near the mutated residue are expected to have high docking affinity because of the EPS influence. Consequently, the two druggability pockets near the mutation site are expected to be actual pockets, which will be helpful for the compound clinical progression studies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10930-022-10041-0.

Acknowledgements We acknowledge the Center for Computational Sciences at University of Kentucky (Lexington, KY, USA) for allocations of compute time on the high performance computing facility (Lipscomb Cluster). Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Author Contributions Not applicable.

Funding The authors declare that no funding was received for this work.

Data Availability The authors declare that the data that supports the findings of this study are available in the supplementary material of this article.

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68:394–424
2. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Pineros M, Znaor A, Bray F (2019) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 144:1941–1953
3. Vogelstein B, Lane D, Levine AJ (2000) Surfing the p53 network. Nature 408:307–310
4. Hong B, van den Heuvel APJ, Prabhu VV, Zhang S, El-Deiry WS (2014) Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. Curr Drug Targets 15:80–89
5. Joerger AC, Fersht AR (2010) The tumor suppressor p53: from structures to drug discovery. Cold Spring Harb Perspect Biol 2:a000919
6. Kruiswijk F, Labuschagne CF, Vousden KH (2015) p53 in survival, death and metabolic health: a lifeguard with a licence to kill. Nat Rev Mol Cell Biol 16:393–405
7. Kruse J-P, Gu W (2009) Modes of p53 regulation. Cell 137:609–622
8. Joerger AC, Fersht AR (2008) Structural biology of the tumor suppressor p53. Annu Rev Biochem 77:557–582
9. Pavletich NP, Chambers KA, Pabo CO (1993) The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. Genes Dev 7:2556–2564
10. Jeffrey PD, Gorina S, Pavletich NP (1995) Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. Science 267:1498–1502

11. Marcel V, Tran PLT, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou M-P, Hall J, Mergny J-L, Hainaut P, Van Dyck E (2011) G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. Carcinogenesis 32:271–278
12. Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, Barlev NA, Reinberg D (2004) Regulation of p53 activity through lysine methylation. Nature 432:353–360
13. Sheng Y, Saridakis V, Sarkari F, Duan S, Wu T, Arrowsmith CH, Frappier L (2006) Molecular recognition of p53 and MDM2 by USP7/HAUSP. Nat Struct Mol Biol 13:285–291
14. Feng H, Jenkins LMM, Durell SR, Hayashi R, Mazur SJ, Cherry S, Tropea JE, Miller M, Wlodawer A, Appella E, Bai Y (2009) Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. Structure 17:202–210
15. Schumacher B, Mondry J, Thiel P, Weyand M, Ottmann C (2010) Structure of the p53 C-terminus bound to 14-3-3: implications for stabilization of the p53 tetramer. FEBS Lett 584:1443–1448
16. Wang L, Li L, Zhang H, Luo X, Dai J, Zhou S, Gu J, Zhu J, Atadja P, Lu C, Li E, Zhao K (2011) Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation. J Biol Chem 286:38725–38737
17. Tong Q, Mazur SJ, Rincon-Arano H, Rothbart SB, Kuznetsov DM, Cui G, Liu WH, Gete Y, Klein BJ, Jenkins L, Mer G, Kutateladze AG, Strahl BD, Groudine M, Appella E, Kutateladze TG (2015) An acetyl-methyl switch drives a conformational change in p53. Structure 23:322–331
18. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, Cousido-Siah A, Masson M, Pol SV, Podjarny A, Travé G, Zanier K (2016) Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature 529:541–545
19. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, Shakked Z (2006) Structural basis of DNA recognition by p53 tetramers. Mol Cell 22:741–753
20. Ghosh R, Kaypee S, Shasmal M, Kundu TK, Roy S, Sengupta J (2019) Tumor suppressor p53-mediated structural reorganization of the transcriptional coactivator p300. Biochemistry 58:3434–3443
21. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA (2018) COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 47:D941–D947
22. Butler JS, Loh SN (2003) Structure, function, and aggregation of the zinc-free form of the p53 DNA binding domain. Biochemistry 42:2396–2403
23. Butler JS, Loh SN (2006) Folding and misfolding mechanisms of the p53 DNA binding domain at physiological temperature. Protein Sci 15:2457–2465
24. Lane DP, Cheok CF, Lain S (2010) p53-based cancer therapy. Cold Spring Harb Perspect Biol 2:a001222
25. Hussein HA, Geneix C, Petitjean M, Borrel A, Flatters D, Camproux A-C (2017) Global vision of druggability issues: applications and perspectives. Drug Discovery Today 22:404–415
26. Miao Y, Feher VA, McCammon JA (2015) Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. J Chem Theory Comput 11:3584–3595
27. Miao Y (2018) Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. J Chem Phys 149:072308
28. Bhattarai A, Miao Y (2018) Gaussian accelerated molecular dynamics for elucidation of drug pathways. Expert Opin Drug Discovery 13:1055–1065
29. Wang J, Miao Y (2019) Mechanistic insights into specific G protein interactions with adenosine receptors. J Phys Chem B 123:6462–6473
30. Bhattarai A, Wang J, Miao Y (2020) G-protein-coupled receptor-membrane interactions depend on the receptor activation state. J Comput Chem 41:460–471
31. Wang Y, Rosengarth A, Luecke H (2007) Structure of the human p53 core domain in the absence of DNA. Acta Crystallogr Sect D: Biol Crystallogr 63:276–281
32. Chen Y, Zhang X, Machado ACD, Ding Y, Chen Z, Qin PZ, Rohs R, Chen L (2013) Structure of p53 binding to the BAX response element reveals DNA unwinding and compression to accommodate base-pair insertion. Nucleic Acids Res 41:8368–8376
33. Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. J Am Chem Soc 21:2337–2338
34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
35. Šali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
36. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Kim D, Kellogg E, Dimaio F, Lange O, Kinch L, Sheffler W, Kim BH, Das R, Grishin NV, Baker D (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins Struct Funct Bioinf 77:89–99
37. Song Y, DiMaio F, Wang RY-R, Kim D, Miles C, Brunette T, Thompson J, Baker D (2013) High-resolution comparative modeling with RosettaCM. Structure 21:1735–1742
38. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26:1668–1688
39. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the amber biomolecular simulation package. WIREs Comput Mol Sci 3:198–210
40. Case DA, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R, Izadi S, Izmailov SA, Kasavajhala K, Kovalenko A, Krasny R, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Man V, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, Onufriev A, Pan F, Pantano S, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Skrynnikov NR, Smith J, Swails J, Walker RC, Wang J, Wilson L, Wolf RM, Wu X, Xiong Y, Xue Y, York DM, Kollman PA (2020) AMBER 2020. University of California, Oakland
41. Wright JD, Noskov SY, Lim C (2002) Factors governing loss and rescue of DNA binding upon single and double mutations in the p53 core domain. Nucleic Acids Res 30:1563–1574
42. Lu Q, Tan YH, Luo R (2007) Molecular dynamics simulations of p53 DNA-binding domain. J Phys Chem B 111:11538–11545
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
44. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Costanzo LD, Christie C, Dalenberg K, Duarte JM, Dutta S, Feng Z, Ghosh S, Goodsell DS, Green RK, Guranović V, Guzenko D, Hudson BP, Kalro T, Liang Y, Lowe R, Namkoong H, Peisach E, Periskova I, Prlić A, Randle C, Rose A, Rose P, Sala R, Sekharan M, Shao C, Tan L, Tao Y-P, Valasatava Y, Voigt M, Westbrook J, Woo J, Yang H, Young J, Zhuravleva M, Zardecki C (2019) RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res 47:D464–D474

45. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11:3696–3713
46. Izadi S, Anandakrishnan R, Onufriev AV (2014) Building water models: a different approach. J Phys Chem Lett 5:3863–3871
47. Li P, Song LF, Merz KM Jr (2015) Systematic parameterization of monovalent ions employing the nonbonded model. J Chem Theory Comput 11:1645–1657
48. Li P, Merz KM Jr (2014) Taking into account the ion-induced dipole interaction in the nonbonded model of ions. J Chem Theory Comput 10:289–297
49. Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935
50. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C (2020) ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J Chem Theory Comput 16:528–552
51. Duan J, Nilsson L (2006) Effect of Zn^{2+} on DNA recognition and stability of the p53 DNA-binding domain. Biochemistry 45:7483–7492
52. Chillemi G, Davidovich P, D'Abramo M, Mametnabiev T, Garabadzhiu AV, Desideri A, Melino G (2013) Molecular dynamics of the full-length p53 monomer. Cell Cycle 12:3098–3108
53. Peters MB, Yang Y, Wang B, Füsti-Molnár L, Weaver MN, Merz KM Jr (2010) Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). J Chem Theory Comput 6:2935–2947
54. Pastor RW, Brooks BR, Szabo A (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol Phys 65:1409–1419
55. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
56. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
57. van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34:1311–1327
58. R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
59. Van Rossum G, Drake FL (2009) Python 3 reference manual. CreateSpace, Scotts Valley, CA
60. Roe DR, Cheatham TE III (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 9:3084–3095
61. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD (2006) Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22:2695–2696
62. Skjaeven L, Yao X-Q, Scarabelli G, Grant BJ (2014) Integrating protein structural dynamics and evolutionary analysis with Bio3D. BMC Bioinf 15:399
63. Skjærven L, Jariwala S, Yao X-Q, Grant BJ (2016) Online interactive analysis of protein structure ensembles with Bio3D-web. Bioinformatics 32:3510–3512
64. Lobanov MY, Bogatyreva NS, Galzitskaya OV (2008) Radius of gyration as an indicator of protein structure compactness. Mol Biol 42:623–628
65. David CC, Jacobs DJ (2014) Principal component analysis: a method for determining the essential dynamics of proteins. In: Livesay DR (ed) Protein dynamics: methods and protocols, vol 1084. Humana Press, Totowa, NJ, pp 193–226
66. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Gr 14:33–38
67. Wolf A, Kirschner KN (2013) Principal component and clustering analysis on molecular dynamics data of the ribosomal L11-23S subdomain. J Mol Model 19:539–549
68. Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36
69. Miao Y, Sinko W, Pierce L, Bucher D, Walker RC, McCammon JA (2014) Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. J Chem Theory Comput 10:2677–2689
70. Rodríguez A, Tsallis C (2010) A generalization of the cumulant expansion. Application to a scale-invariant probabilistic model. J Math Phys 51:073301
71. Hussein HA, Borrel A, Geneix C, Petitjean M, Regad L, Camproux A-C (2015) PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Res 43:W436–W442
72. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr Sect D: Biol Crystallogr D66:12–21
73. Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27:343–350
74. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303
75. Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170
76. Roland Lüthy DE, Bowie James U (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
77. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
78. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264:121–136
79. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
80. Pradhan MR, Siau JW, Kannan S, Nguyen MN, Ouaray Z, Kwoh CK, Lane DP, Ghadessy F, Verma CS (2019) Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket. Nucleic Acids Res 47:1637–1652
81. Offutt TL, Ieong PU, Demir O, Amaro RE (2018) Dynamics and molecular mechanisms of p53 transcriptional activation. Biochemistry 57:6528–6537
82. Hess B (2002) Convergence of sampling in protein simulations. Phys Rev E: Stat Nonlinear Soft Matter Phys 65:031910
83. Pérez-Canadillas JM, Tidow H, Freund SMV, Rutherford TJ, Ang HC, Fersht AR (2006) Solution structure of p53 core domain: structural basis for its instability. Proc Natl Acad Sci USA 103:2109–2114
84. Lukman S, Lane DP, Verma CS (2013) Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. PLoS ONE 8:e80221
85. Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 265:346–355

86. Chen Y, Dey R, Chen L (2010) Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. Structure 18:246–256
87. Xu J, Reumers J, Couceiro JR, Smet FD, Gallardo R, Rudyak S, Cornelis A, Rozenski J, Zwolinska A, Marine J-C, Lambrechts D, Suh Y-A, Rousseau F, Schymkowitz J (2011) Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. Nat Chem Biol 7:285–295
88. Connolly ML (1983) Analytical molecular surface calculation. J Appl Crystallogr 16:548–558
89. Eldar A, Rozenberg H, Diskin-Posner Y, Rohs R, Shakked Z (2013) Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein-DNA interactions. Nucleic Acids Res 41:8748–8759
90. Thayer KM, Quinn TR (2016) p53 R175H hydrophobic patch and H-bond reorganization observed by MD simulation. Biopolymers 105:176–185