Impact of Arg175His mutation on the dynamical patterns of full-length human p53 protein

Morad Mustafa *
Department of Chemistry, School of Science
The University of Jordan, Amman 11942, Jordan

Mohammed Gharaibeh
Department of Chemistry, School of Science
The University of Jordan, Amman 11942, Jordan

January 21, 2021

1 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.

*Correspondence: morad.mustafa@gmail.com

Abstract

p53, a tumor suppressor protein, is essential for preventing cancer development. Although individual domains of the human p53 have been well analyzed, no study has experimentally revealed a full-length structure of human wild-type p53 protein at atomistic level. The presence of all human p53 domains in one structure will help in showing the correlated interactions among these domains, thus, leading to enhance our understanding about the dynamics of this protein and its mutant forms. In this study we have modeled five human p53 forms, namely, inactive, distal-active, proximal-active, distal-mutant, and proximal-mutant forms. These forms have been investigated in this study by gaussian accelerated molecular dynamics simulations in OPC water model at physiological temperature and pH. On the basis of the observed dynamical patterns, all wild-type forms can achieve better conformational stability through dimerization or tetramerization process. The dynamical patterns and free-energy profiles of the wild-type forms highlight the most vulnerable sites to mutations; that is, p53–DNA and p53–p53 interfaces. On the other hand, principal component and clustering analysis methods on Arg175His mutant forms reveal two distinct conformational states (clusters); extended and compact clusters. The two clusters of each mutant form reveal negative cavities near the mutation site, which can be used for drug screening studies. The observed compact structures in the conformations of Arg175His mutant forms may indicate formation of aggregation.
Keywords: p53 monomer · Arg175His mutation · Free-energy · GaMD · PCA

2 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. Because somatic mutations are more frequent than germline mutations, they become significant research targets. 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].

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 p53 tumor-suppressor gene fails to function properly, uncontrolled growth and division of cells arises causing genomics instability, a sign of cancer [4]. Therefore, p53 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 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 Figure 1 according to the discovered p53 isoforms and the pinpointed structural–functional features of human p53 protein. The p53-NTD, spanning residues (1–93) [8], contains 3 subdomains: transcription-activation domain 1 (TAD_1), spanning residues (1–39) [10, 11], transcription-activation domain 2 (TAD_2), spanning residues (40–60) [8], and proline-rich domain (PRD), spanning residues (64–93) [8]. The p53-DBD has spanning residues (94–292) [12]. The p53-CTD, spanning residues (293–393) [8], contains 2 subdomains: oligomerization domain (OD),

spanning residues (325–356) [13], and regulatory domain (RD), spanning residues (358–393) [14–18].

Both p53-TAD and p53-RD are natively unfolded; that is, they are highly flexible and unstructured [5]. The natural disordered state of both p53-TAD and p53-RD facilitates binding p53-TAD to different target proteins or nucleic acid partners, and binding p53-RD to different target proteins or nonspecific DNA [5]. Consequently, both p53-TAD and p53-RD undergo, upon binding, conformational changes causing their disorder-to-order transition [5]. The p53-PRD is essential for the apoptosis function of p53 and for efficient growth suppression; however, it does not participate in p53's upstream regulation by DNA damage [19]. The low-natural-thermodynamic stability of human p53–DBD (with a half-life of 9 minutes) allows p53 to rapidly cycle between folded and unfolded states, facilitating binding of p53 protein to different partner proteins, and makes p53 susceptible to deleterious mutations [5].

In general, several particular challenges contribute in tackling full-length structure of a wild-type form of human p53 protein at the atomic level including [8]: p53 protein has a low-natural stability, p53 protein is expressed in normal cells at very low concentrations, p53 protein has a large size to be studied by conventional NMR spectroscopy method and a small size to be studied by the cryo-electron microscopy, and p53 protein has a high proportion of naturally disordered regions that impedes crystallization.

The low-stability issue of human p53 protein has been solved by designing stabilized forms of human p53 through introducing naturally occurring mutations in the human p53 domains, making this protein more amenable to structural studies [8]. These stabilized forms of human p53 have revealed valuable structural information [20–34]. Nevertheless, the structures of these forms are virtually identical with the structure of the wild-type form away from the mutated sites. On the other hand, numerous computational studies [35–42] have been conducted on the human p53 protein without introducing any stabilized mutation in the p53 domains, yet significant observations and conclusions have been reported. Specifically, structural and dynamical properties of full-length human p53 as well as its interdomain interactions have been studied in a model of full-length p53 using molecular dynamics method at 300 K [41].

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 [43]. 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). The proximal p53-RD allows an additional stabilization of a p53–DNA complex through nonspecific electrostatic interactions between the positively charged proximal p53-RD and a DNA backbone. In this way, the proximal p53-RD aids in docking DNA onto a p53-DBD tetramer; that is, positively regulating p53 binding to its DNA response elements; consequently, the transcriptional activation is enhanced. In a cryo-electron microscopy study [44], the relative arrangement of p53-OD and p53-DBD structures has been revealed; consequently, this orientation facilitates modeling of the natively folded domains.

Determining 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 general, the mutant forms of the human p53 protein are classified into 2 categories: contact site mutants (hotspot mutations: Arg248 and Arg273) or conformational mutants (hotspot mutations: Arg175, Gly245, Arg249, and Arg282) [8]. Contact site mutants are mutated at residues crucial for the p53–DNA interaction, whereas conformational mutants 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 breast cancer, the top three p53 somatic mutations are Arg248Gln, Arg273His, and Arg175His [45]; therefore, Arg175His mutant is one of the designated hotspot mutations in

p53 protein. 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, 46, 47].

Despite several computational studies on p53-DBD, many dynamical and structural aspects of p53 function are remained mysterious. For instance, will introducing both p53-NTD and p53-CTD with p53-DBD in a computational study affect the results obtained from computational studies of p53-DBD alone? How the roles of the individual domains are integrated with each other to facilitate the protein function? What is the impact of maintaining the tetrahedrally coordinated Zn^{2+} ion by Cys176 from L2 loop, His179 from α1 helix, Cys238, and Cys242 from L3 loop on the dynamics of Arg175His mutant?

Therefore, in this study we have built our forms by utilizing the relative orientation of p53-OD to p53-DBD [44]; thus, enhancing the credibility and realisticity of our forms to be solvated with an OPC water model and be studied by Gaussian Accelerated Molecular Dynamics (GaMD) technique at physiological temperature and pH. Our focus in this study is to explore the dynamics of an inactive form, active forms, and mutant forms (Arg175His) to reveal their dynamical patterns for future computer-aided drug design studies.

3 Methods

3.1 Molecular Modeling

The starting structures for the wild-type p53-DBD forms were taken from chain B of PDB entry 2OCJ [48], an inactive (DNA-free) structure, and from chain A of PDB entry 4HJE [49], 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 [50]. In the next three paragraphs, the modeling steps were done using UCSF Chimera [51], version 1.14, and its built in Modeller, version 9.23, functions [52].

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 [44], 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 [53, 54], 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 [55–57]. To remove residue–residue clashes because of the mutation process, minimization in vacume 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) [58] 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 [38], three Cys residues were deprotonated (i.e., Cys176, Cys238, and Cys242). 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

[59, 60].

The LEaP module of Amber 20 [55–57] was used for adding missing atoms, applying ff14SB protein force field [61], solvating the protein in a truncated octahedron box of OPC water molecules [62] with a buffering distance set to 12.0 Å, loading Li/Merz ion parameters (12-6-4 set) for monovalent [63] and +2 [64] ions in the designated water model, and neutralizing the modeled system with charge neutralizing counter ions (i.e., Na^+ ions). Unlike the widely used TIP3P water model [65], the OPC water model was recommended to be used with ff14SB protein force field in biomolecular simulations [57]. Combining OPC water model with ff14SB protein force field in molecular dynamics simulations has improved the accuracy of atomistic simulations and given better modeling sequence-specific behavior, protein mutations, and rational protein-design results [66].

3.2 Gaussian Accelerated Molecular Dynamics Simulations

GaMD simulations were performed using PMEMD engine within the Amber 20 software suite [55–57]. 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 [37], 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, 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 [38, 41], 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 [67]. 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 [68] with a collision frequency of 5 ps^{-1}. The pressure was maintained at 1 bar with isotropic position scaling using Berendsen barostat [69] with pressure relaxation time of 1 ps. The nonbonded cutoff was assigned to 9.0 Å. The Particle Mesh Ewald method [70] 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 [71] was used in all simulation processes, but minimization processes, to omit bond interactions involving 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 GaMD simulation [72]. 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 molecular dynamics (no statistics were collected), 3.0 ns of initial stage as conventional molecular dynamics (potential statistics were calculated for GaMD pre-equilibration stage), 2.0 ns of GaMD pre-equilibration stage as preparation biasing molecular dynamics simulation (boost potential was applied but boost parameters were not updated), and 3.0 ns of GaMD equilibration stage as biasing molecular dynamics 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.

3.3 Data Analysis

Both R [73], version 3.6.3, and Python [74], version 3.7.3, languages 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, principal component analysis (PCA), and free-energy profiles.

3.3.1 Root-Mean-Square Deviation

RMSD reflects the degree of similarity between three-dimensional protein structures. It can be computed [75–77] 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).

3.3.2 Root-Mean-Square Fluctuation

RMSF reveals the conformational variance of the protein. It can be measured [75–77] by calculating the deviation between the position of atom i (usually C_α) with respect to its average position over the whole simulation trajectory.

3.3.3 Radius of Gyration

In biological molecules, radius of gyration indicates the protein structure compactness [78]. It can be determined [75–77] 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 [78]. Maintaining a relatively steady value of radius of gyration over time reveals the stability of the protein folding state.

3.3.4 Principal Component Analysis

Conducting PCA reveals the most important motions of a biological system over a broad range of time and spatial scales [79]. 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 [79]. The first principal component, the eigenvector with the highest corresponding eigenvalue, reflects the most identifying motion patterns in the simulation. The eigenvalues show the contribution of the corresponding eigenvectors to the global fluctuations of a biological system.

In our GaMD simulations, PCA was performed [75–77] 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.

3.3.5 Clustering Analysis

Clustering analysis is a technique that finds patterns within data by locating clusters of geometrically similar conformers in ensembles of chemical conformations [80]. Most cluster-

ing 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 [75–77] has been used because of its one of the fastest and most widely used techniques, and it has apparent good performance in analyzing molecular dynamics trajectory data [80]. 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 [80]. The quality of a k-means partition is evaluated by calculating the percentage as indicated by Eq. 1 [80], 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 [81] built in R language [73]. 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.

3.3.6 Free-Energy Analysis

GaMD method facilitates unconstrained-enhanced sampling of a biomolecular system by adding a harmonic boost potential to smooth the system potential energy surface [57]. By constructing a harmonic boost potential that follows Gaussian distribution, potential of mean force (PMF) (i.e., a free-energy profile) can be extracted by accurate reweighting of the GaMD simulations [57].

In the reweighting method, cumulant expansion can be used to approximate the ensemble-averaged Boltzmann factor, $\langle \mathrm{e}^{\beta \Delta V(r)} \rangle$, where $\beta = \frac{1}{k_{\mathrm{B}} T}$, k_{B} is Boltzmann constant, and T is the system temperature [82]. 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 [82], and $\langle \mathrm{e}^{\beta \Delta V(r)} \rangle_i$ be the ensemble-averaged Boltzmann factor of $\Delta V(r)$ for simulation frames found in the i^{th} bin. Thus, the cumulant expansion to the third order is given by Eq. 2 [83].

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

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 i^{th} 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 $F^{\mathrm{b}}[A(r)]$, for each bin i as illustrated in Eq. 6.

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

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

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

To identify distinct low-energy states of our biomolecule systems, free energy profiles were obtained by performing reweighting method along PC1 and PC2.

4 Results

4.1 Model Quality Assessment

The quality of our models has been assessed using SWISS-MODEL server tools [84–86] and Structure Analysis and Verification Server tools [87–91]. Assessment results (Tables S1–S3 and Figure 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.

4.2 Stability of Molecular Structures

In general, the RMSD results for p53-DBD structures (see Figure S2) show reasonable stability in those regions throughout the first 100 ns production run time scale. Beyond 100 ns, the RMSD values are slightly higher than those before 100 ns indicating some structural variations in p53-DBD structures. On the other hand, the RMSD results for the full-length p53 forms (data not shown) have high values, which are attributed, as expected, to the free movements of both p53-NTD and p53-CTD [41].

The RMSF results (Figure 2) for all forms confirm the high thermal fluctuations of p53-NTD and p53-CTD. The most significant thermal fluctuations in p53-NTD regions include p53-TAD$_2$ and p53-PRD. The trend of RMSF results for wild-type p53 forms (Figure 2a) agrees with a previously reported RMSF trend of a full-length p53 [41]. For p53-DBD thermal fluctuations, we highlight the differences among the wild-type forms, and list the regions with the highest thermal fluctuations in mutant forms. As shown in Figures 2a and S3a–c, L1 loop, beginning of L2 loop, $\beta6$–$\beta7$ loop, and $\beta9$–$\beta10$ loop have higher thermal fluctuations in inactive and proximal-active forms than in distal-active form. The observed similarity in the RMSF values for an inactive and a proximal-active p53-DBD forms might be due to both structures have the same DBD–OD orientation. In Figures 2b and S3d–e, the p53-DBDs in the mutant forms have high thermal fluctuations in $\beta1$–$\beta2$ loop, L1 loop, $\beta3$–$\beta4$ loop, L2 loop, $\beta6$–$\beta7$ loop, $\beta7$–$\beta8$ loop, L3 loop, and $\beta9$–$\beta10$ loop.

The protein structure compactness reveals variations in packing character (Figure 3). In the first 100 ns production run time scale, the radius of gyration has a steady value in the wild-type systems (Figure 3a). The steady values of radius of gyration indict stability in the

protein folding state. In contrast and after 100 ns, the tightest packing character in both an inactive and a proximal-active forms is increased, whereas in the distal-active form is decreased. Visual inspection of inactive and proximal-active trajectories beyond 100 ns reveals emerging of compact conformations of p53 proteins. On the other hand, the radius of gyration decreases in mutant forms (Figure 3b) indicating an increase in the tightest packing character. By visually inspecting the mutant trajectories, the p53-NTD comes in a direct contact with the p53-CTD forming a compact structure.

4.3 Dynamics of Protein Domains

Figure S4 shows the PCA scree plot, which indicates the proportion of variance against its mode index. The first two components together make up 60.9%, 56.8%, 74.8%, 73.2%, and 69% of the variance of the inactive, distal-active, proximal-active, distal-mutant, and proximal-mutant p53 forms, respectively. The major dynamics of p53 protein can be captured by PC1; therefore, Movies S1–S5 demonstrate the trajectories of atomic displacements along PC1 for p53 forms. In all p53 forms, Movies S1–S5 show projection of p53-TAD away from the surface of p53-DBD by p53-PRD motion, and this observation is consistent with the observation in an experimental study [92].

In an inactive form, its dynamics is demonstrated by Movie S1. PC1 describes drastic motions of p53-NTD and p53-CTD, which are consistent with their high thermal fluctuations (Figure 2a), and a periodic motion of p53-DBD. Both p53-TAD and p53-DBD move closer to each other, whereas p53-RD moves closer to p53-DBD while p53-DBD moves away from p53-RD. Thus, the observed correlated motions produce a compact structure of an inactive p53 form, which explains the radius of gyration results for an inactive p53 form (Figure 3a).

In a distal-active form, its dynamics is illustrated by Movie S2. PC1 depicts substantial motions of p53-NTD and p53-CTD, which are consistent with their high thermal fluctuations (Figure 2a), and a weak periodic motion of p53-DBD. In addition, there is a noticeable motion of L2 loop (comprising residues 164–194) in p53-DBD. Moreover, the same movie shows denaturation of p53-OD and p53-RD regions.

In a proximal-active form, its dynamics is displayed by Movie S3. PC1 shows drastic motions of p53-NTD and p53-CTD, which are consistent with their high thermal fluctuations (Figure 2a), and a periodic motion of p53-DBD. Both p53-TAD and p53-DBD move closer to each other, whereas p53-OD and p53-RD move closer to p53-DBD while p53-DBD moves away from them. Thus, the observed correlated motions produce a compact structure of a proximal-active p53 form, which explains the radius of gyration results (Figure 3a). In addition, the motions of L1 (comprising residues 113–123) and L3 (comprising residues 237–250) loops in p53-DBD have been observed. Moreover, the same movie shows unfolding states of p53-TAD (i.e., losing their secondary structures).

The dynamics of a distal-mutant form is illustrated by Movie S4. PC1 displays extreme motions of p53-NTD and p53-CTD, which are consistent with their high thermal fluctuations (Figure 2b), and a periodic motion of p53-DBD. Surprisingly, both p53-TAD and p53-CTD move closer to each other in a way that resembles an opening and closing motion. Thus, the observed correlated motions explain the compactness of the distal-mutant form (Figure 3b). In addition, the motions of L1 loop and $\beta 7$–$\beta 8$ loop (comprising residues 220–229) in p53-DBD have been observed.

Movie S5 demonstrates the dynamics of a proximal-mutant form. PC1 describes heavy motions of p53-NTD and p53-CTD, which are consistent with their high thermal fluctuations (Figure 2b), and a periodic motion of p53-DBD. Unexpectedly, both p53-TAD and p53-CTD move closer to each other in a way that mimics an opening and closing motion. Thus, the observed correlated motions explain the compactness of the proximal-mutant form (Figure 3b).

Moreover, the motion $\beta7$–$\beta8$ loop in p53-DBD has been observed. Also, the same movie shows unfolding states of p53-TAD.

Projecting the trajectory frames onto the plane formed by PC1 and PC2 (Figure S5) shows how similar structures are clustered and reveals random diffusion in a high-dimensional harmonic potential. The observed patterns can be interpreted as thermal motion along a shallow free-energy landscape [93]. Even though 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.

Contrarily, conformational states analysis can be achieved by clustering the PC data; thus, allowing comparisons of all conformers sampled during the apparent thermal diffusion. According to NbClust function [81] built in R language [73], Figure 4 shows that PC data of each form consists of two clusters. This observation is also supported by the cluster quality results (see Figure S6), which shows an identified *kink* at cluster count of two. In addition, Figure 4 illustrates that approximately the first 60%, 50%, 60%, 30%, and 70% of PCA time scale correspond to the *cluster 1* in the inactive, distal-active, proximal-active, distal-mutant, and proximal-mutant p53 forms, respectively.

Two representative-clustered conformations have been extracted for distal- and proximal-mutant forms as shown in Figure 5 and 6, respectively. In Figure 5a, *cluster 1* in distal-mutant contains a deep negative cavity near the Arg175His mutation, whereas in Figure 5b, *cluster 2* in distal-mutant contains a wide negative cavity near the mutation site. On the other hand, *cluster 1* in proximal-mutant contains a deep and narrow negative cavity near the Arg175His mutation (see Figure 6a), whereas in Figure 6b, *cluster 2* in proximal-mutant contains a wide negative cavity near the mutation site.

4.4 Free Energy Profiles

The free energy profiles explain the relationship between the conformational change and energetic change. Figure 7 shows the free energy profiles for p53 forms. More energetic wells (blue regions) represent the protein undergo larger conformational change. Noticeably, distal-active form experience large conformational changes as revealed by the deeper well (see Figure 7b). Despite the observed interactions between p53-TAD and p53-CTD (Movies S4 and S5), both distal- and proximal-mutant forms encounter medium conformational changes as revealed by Figure 7d and 7e. However, as illustrated in Figure 7, all p53 forms are not confined to a major energetic deep well throughout the simulation, highlighting ensembles of different conformational states distributed over a large free energy landscape. Under physiological conditions, these results, obtained by GaMD simulations, indicates that high flexibility conformations have high thermal fluctuations, which leads to high elastic transition from one conformation to another.

5 Discussion

It has been suggested that maintaining the stability of the Zn^{2+}-binding site decreases the thermal fluctuation of L2 loop and prevents aggregation [37]. By maintaining the stability of Zn^{2+} binding site, the structural and dynamical properties of full-length human p53 as well as its interdomain interactions have been studied by classical molecular dynamics method at 300 K for 850 ns [41]. In a good agreement with classical molecular dynamics results [41], our study shows strong stability in all p53-DBD forms with major flexibility in both p53-NTD and p53-CTD (2), and shows strong correlated motions involving p53-TAD, p53-PRD, p53-DBD, p53-OD, and p53-RD (Movies S1–S5). The observed stability of p53-TAD secondary structure in this study is consistent with the observation in another study at 310 K [40] and at 298 and

400 K [36]. That is, p53-TAD_1 has natively compact and folded structure that possesses a stable α-helix composed by residues Phe19–Leu26 [40]. Moreover, it is observed that α1 helix in the middle of L2 loop spontaneously unfolds in all simulations, and this observation is consistent with the other observation in a previous computational study at 300 K [37].

We are able to observe several amazing conformational changes in a short period of simulation time scale using novel GaMD method. For instance, the noticeable motion of part of L2 loop in distal-active form may demonstrate how L2 loop explores the right conformation for forming dimer–dimer interface [94]. Also, the observed motions of L1 and L3 loops in proximal-active form may illustrate how L3 loop searches for the right conformation that is useful for dimer–dimer interface [94], and how L1 loop seeks the right rearrangement to allow DNA binding [48]. For instance, the observed compact structures in an inactive and proximal-active forms may indicate a possibility of a loss in p53 functionality due to aggregation. Our findings could indicate that proximal-active form will dock on the DNA surface to prevent formation of compact structures then followed by docking of distal-active form.

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 (L1 loop, β2–β2′ turn, the 4 C-terminal residues of β10 strand, and the α2 helix) [95, 96]. In a molecular dynamics simulation study [97] of Arg175His mutant, a hydrophobic patch has been suggested as a druggable site to prevent unfolding and aggregation by stabilizing the zinc binding region.

Contrarily, conducting GaMD simulations on a hotspot mutation (Arg175His) while maintaining the stability of Zn^{2+} binding site can reveal further information. The observed compact structures in the conformations of Arg175His mutant forms (Movies S4 and S5) may indicate formation of aggregation. The association of structure compactness with negative cavities near the mutation site may reveal conformations that are suitable for drug screening.

6 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. Most of our wild-type observations are in good agreement with those from another study [41] conducted on full-length p53 protein with TIP3P water model suggesting that OPC water model can give credible results and GaMD method can give faster results.

The observed similarity behaviors of the conformational stability in both an inactive and a proximal-active forms explains the observation in previous studies [38, 98] that the DNA-free p53-DBD structure is similar to the DNA-bound p53-DBD structure [95] but with higher mobility and flexibility. Moreover, the higher thermal fluctuations of both inactive and a proximal-active forms might indicate that their relative orientation of p53-OD to p53-DBD is more vulnerable to mutations.

In this study we have also used PCA combined with clustering to study the dynamics of full-length human p53 protein. On the basis of the observed dynamical patterns, all wild-type forms can achieve better conformational stability through dimerization or tetramerization process. The dynamical patterns of the three wild-type forms highlight the most vulnerable sites to mutations; that is, p53–DNA and p53–p53 interfaces.

By a subsequent clustering using the k-means algorithm of the structural frames in the subspace spanned by PC1 and PC2, we have identified two clusters that represent two distinct conformational states, which are useful in drug screening studies for Arg175His mutation. Free-energy profiles demonstrate ensembles of different conformational states distributed over a large free energy landscape. The observed compact structures in the conformations of Arg175His mutant forms may indicate formation of aggregation.

7 Conflict of interest

The authors declare that they have no conflict of interest.

8 Abbreviations

The following abbreviations are used in this manuscript:

NTD	N-terminal domain
DBD	DNA binding domain
CTD	C-terminal domain
TAD_1	Transcription-activation domain 1
TAD_2	Transcription-activation domain 2
PRD	Proline-rich domain
OD	Oligomerization domain
RD	Regulatory domain
FL	Flexible loop
ACE	Acetyl group
NME	N-methyl group
NVT	Constant number of particles (N), volume (V), and temperature (T)
NPT	Constant number of particles (N), pressure (P), and temperature (T)
GaMD	Gaussian Accelerated Molecular Dynamics
RMSD	Root-mean-square displacement
RMSF	Root-mean-square fluctuation
Rgyr	Radius of Gyration
PCA	Principal component analysis
PMF	Potential of mean force

References

[1] Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, 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., M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, 2019. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* 144:1941–1953.

[3] Vogelstein, B., D. Lane, and A. J. Levine, 2000. Surfing the p53 network. *Nature* 408:307–310.

[4] Hong, B., A. P. J. van den Heuvel, V. V. Prabhu, S. Zhang, and W. S. El-Deiry, 2014. Targeting tumor suppressor p53 for cancer therapy: strategies, challenges and opportunities. *Curr. Drug Targets* 15:80–89.

[5] Joerger, A. C., and A. R. Fersht, 2010. The tumor suppressor p53: from structures to drug discovery. *Cold Spring Harb Perspect. Biol.* 2:a000919.

[6] Kruiswijk, F., C. F. Labuschagne, and K. H. Vousden, 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., and W. Gu, 2009. Modes of p53 regulation. *Cell* 137:609–622.

[8] Joerger, A. C., and A. R. Fersht, 2008. Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.* 77:557–582.

[9] Pavletich, N. P., K. A. Chambers, and C. O. Pabo, 1993. The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev.* 7:2556–2564.

[10] Marcel, V., P. L. T. Tran, C. Sagne, G. Martel-Planche, L. Vaslin, M.-P. Teulade-Fichou, J. Hall, J.-L. Mergny, P. Hainaut, and E. Van Dyck, 2011. G-quadruplex structures in TP53 intron 3: role in alternative splicing and in production of p53 mRNA isoforms. *Carcinog.* 32:271–278.

[11] Feng, H., L. M. M. Jenkins, S. R. Durell, R. Hayashi, S. J. Mazur, S. Cherry, J. E. Tropea, M. Miller, A. Wlodawer, E. Appella, and Y. Bai, 2009. Structural basis for p300 Taz2–p53 TAD1 binding and modulation by phosphorylation. *Structure* 17:202–210.

[12] Martinez-Zapien, D., F. X. Ruiz, J. Poirson, A. Mitschler, J. Ramirez, A. Forster, A. Cousido-Siah, M. Masson, S. V. Pol, A. Podjarny, G. Travé, and K. Zanier, 2016. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature* 529:541–545.

[13] Jeffrey, P. D., S. Gorina, and N. P. Pavletich, 1995. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* 267:1498–1502.

[14] Chuikov, S., J. K. Kurash, J. R. Wilson, B. Xiao, N. Justin, G. S. Ivanov, K. McKinney, P. Tempst, C. Prives, S. J. Gamblin, N. A. Barlev, and D. Reinberg, 2004. Regulation of p53 activity through lysine methylation. *Nature* 432:353–360.

[15] Sheng, Y., V. Saridakis, F. Sarkari, S. Duan, T. Wu, C. H. Arrowsmith, and L. Frappier, 2006. Molecular recognition of p53 and MDM2 by USP7/HAUSP. *Nat. Struct. Mol. Biol.* 13:285–291.

[16] Schumacher, B., J. Mondry, P. Thiel, M. Weyand, and C. Ottmann, 2010. Structure of the p53 C-terminus bound to 14-3-3: implications for stabilization of the p53 tetramer. *FEBS Lett.* 584:1443–1448.

[17] Wang, L., L. Li, H. Zhang, X. Luo, J. Dai, S. Zhou, J. Gu, J. Zhu, P. Atadja, C. Lu, E. Li, and K. Zhao, 2011. Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation. *J. Biol. Chem.* 286:38725–38737.

[18] Tong, Q., S. J. Mazur, H. Rincon-Arano, S. B. Rothbart, D. M. Kuznetsov, G. Cui, W. H. Liu, Y. Gete, B. J. Klein, L. Jenkins, G. Mer, A. G. Kutateladze, B. D. Strahl, M. Groudine, E. Appella, and T. G. Kutateladze, 2015. An acetyl-methyl switch drives a conformational change in p53. *Structure* 23:322–331.

[19] Walker, K. K., and A. J. Levine, 1996. Identification of a novel p53 functional domain that is necessary for efficient growth suppression. *Proc. Natl. Acad. Sci. U.S.A.* 93:15335–15340.

[20] Joerger, A. C., M. D. Allen, and A. R. Fersht, 2004. Crystal structure of a superstable mutant of human p53 core domain. Insights into the mechanism of rescuing oncogenic mutations. *J. Biol. Chem.* 279:1291–1296.

[21] Joerger, A. C., H. C. Ang, D. B. Veprintsev, C. M. Blair, and A. R. Fersht, 2005. Structures of p53 cancer mutants and mechanism of rescue by second-site suppressor mutations. *J. Biol. Chem.* 280:16030–16037.

[22] Joerger, A. C., H. C. Ang, and A. R. Fersht, 2006. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc. Natl. Acad. Sci. U.S.A.* 103:15056–15061.

[23] Khoo, K. H., A. C. Joerger, S. M. Freund, and A. R. Fersht, 2009. Stabilising the DNA-binding domain of p53 by rational design of its hydrophobic core. *Protein Eng. Des. Sel.* 22:421–430.

[24] Basse, N., J. L. Kaar, G. Settanni, A. C. Joerger, T. J. Rutherford, and A. R. Fersht, 2010. Toward the rational design of p53-stabilizing drugs: probing the surface of the oncogenic Y220C mutant. *Chem. Biol.* 17:46–56.

[25] Natan, E., C. Baloglu, K. Pagel, S. M. Freund, N. Morgner, C. V. Robinson, A. R. Fersht, and A. C. Joerger, 2011. Interaction of the p53 DNA-binding domain with its N-terminal extension modulates the stability of the p53 tetramer. *J. Mol. Biol.* 409:358–368.

[26] Wilcken, R., X. Liu, M. O. Zimmermann, T. J. Rutherford, A. R. Fersht, A. C. Joerger, and F. M. Boeckler, 2012. Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J. Am. Chem. Soc.* 134:6810–6818.

[27] Liu, X., R. Wilcken, A. C. Joerger, I. S. Chuckowree, J. Amin, J. Spencer, and A. R. Fersht, 20. Small molecule induced reactivation of mutant p53 in cancer cells. *Nucleic Acids Res.* 41:6034–6044.

[28] Joerger, A. C., M. R. Bauer, R. Wilcken, M. G. J. Baud, H. Harbrecht, T. E. Exner, F. M. Boeckler, J. Spencer, and A. R. Fersht, 2015. Exploiting transient protein states for the design of small-molecule stabilizers of mutant p53. *Structure* 23:2246–2255.

[29] Wilcken, R., M. O. Zimmermann, M. R. Bauer, T. J. Rutherford, A. R. Fersht, A. C. Joerge, and F. M. Boeckle, 2015. Experimental and theoretical evaluation of the ethynyl moiety as a halogen bioisostere. *ACS Chem. Biol.* 10:2725–2732.

[30] Bauer, M. R., A. C. Joerger, and A. R. Fersht, 2016. 2-Sulfonylpyrimidines: Mild alkylating agents with anticancer activity toward p53-compromised cells. *Proc. Natl. Acad. Sci. U.S.A.* 13:E5271–E5280.

[31] Bauer, M. R., R. N. Jones, M. G. J. Baud, R. Wilcken, F. M. Boeckler, A. R. Fersht, A. C. Joerger, and J. Spencer, 2016. Harnessing fluorine–sulfur contacts and multipolar interactions for the design of p53 mutant Y220C rescue drugs. *ACS Chem. Biol.* 11:2265–2274.

[32] Veprintsev, D. B., S. M. Freund, A. Andreeva, S. E. Rutledge, H. Tidow, J. M. Cañadillas, C. M. Blair, and A. R. Fersht, 2006. Core domain interactions in full-length p53 in solution. *Proc. Natl. Acad. Sci. U.S.A.* 103:2115–2119.

[33] Tidow, H., R. Melero, E. Mylonas, S. M. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle, and A. R. Fersht, 2007. Quaternary structures of tumor suppressor p53 and a specific p53–DNA complex. *Proc. Natl. Acad. Sci. U.S.A.* 104:12324–12329.

[34] Huang, F., S. Rajagopalan, G. Settanni, R. J. Marsh, D. A. Armoogum, N. Nicolaou, A. J. Bain, E. Lerner, E. Haas, L. Ying, and A. R. Fersht, 2009. Multiple conformations of full-length p53 detected with single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* 106:20758–20763.

[35] Chong, L. T., C. D. Snow, Y. M. Rhee, and V. S. Pande, 2005. Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *J. Mol. Biol.* 345:869–878.

[36] Espinoza-Fonseca, L. M., and J. G. Trujillo-Ferrara, 2006. Transient stability of the helical pattern of region F19–L22 of the N-terminal domain of p53: a molecular dynamics simulation study. *Biochem. Biophys. Res. Commun.* 343:110–116.

[37] Duan, J., and L. Nilsson, 2006. Effect of Zn^{2+} on DNA recognition and stability of the p53 DNA-binding domain. *Biochemistry* 45:7483–7492.

[38] Lu, Q., Y. H. Tan, and R. Luo, 2007. Molecular dynamics simulations of p53 DNA-binding domain. *J. Phys. Chem. B* 111:11538–11545.

[39] Lwin, T. Z., J. J. Durant, and D. Bashford, 2007. A fluid salt-bridging cluster and the stabilization of p53. *J. Mol. Biol.* 373:1334–1347.

[40] Espinoza-Fonseca, L. M., 2009. Leucine-rich hydrophobic clusters promote folding of the N-terminus of the intrinsically disordered transactivation domain of p53. *FEBS Lett.* 583:556–560.

[41] Chillemi, G., P. Davidovich, M. D'Abramo, T. Mametnabiev, A. V. Garabadzhiu, A. Desideri, and G. Melino, 2013. Molecular dynamics of the full-length p53 monomer. *Cell Cycle* 12:3098–3108.

[42] Lukman, S., D. P. Lane, and C. S. Verma, 2013. Mapping the structural and dynamical features of multiple p53 DNA binding domains: insights into loop 1 intrinsic dynamics. *PLoS One* 8:e80221.

[43] Kitayner, M., H. Rozenberg, N. Kessler, D. Rabinovich, L. Shaulov, T. E. Haran, and S. Z., 2006. Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* 22:741–753.

[44] Ghosh, R., S. Kaypee, M. Shasmal, T. K. Kundu, S. Roy, and J. Sengupta, 2019. Tumor suppressor p53-mediated structural reorganization of the transcriptional coactivator p300. *Biochemistry* 58:3434–3443.

[45] Tate, J. G., S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, and S. A. Forbes, 2018. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47:D941–D947.

[46] Butler, J. S., and S. N. Loh, 2003. Structure, function, and aggregation of the zinc-free form of the p53 DNA binding domain. *Biochemistry* 42:2396–2403.

[47] Butler, J. S., and S. N. Loh, 2006. Folding and misfolding mechanisms of the p53 DNA binding domain at physiological temperature. *Protein Sci.* 15:2457–2465.

[48] Wang, Y., A. Rosengarth, and H. Luecke, 2007. Structure of the human p53 core domain in the absence of DNA. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 63:276–281.

[49] Chen, Y., X. Zhang, A. C. D. Machado, Y. Ding, Z. Chen, P. Z. Qin, R. Rohs, and L. Chen, 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.

[50] Kuszewski, J., A. M. Gronenborn, and G. M. Clore, 1999. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* 21:2337—2338.

[51] Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, 2004. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25:1605–1612.
UCSF Chimera Home Page. http://www.rbvi.ucsf.edu/chimera/.

[52] Šali, A., and T. L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.

[53] Raman, S., R. Vernon, J. Thompson, M. Tyka, R. Sadreyev, J. Pei, D. Kim, E. Kellogg, F. Dimaio, O. Lange, L. Kinch, W. Sheffler, B. H. Kim, R. Das, N. V. Grishin, and D. Baker, 2009. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins Struct. Funct. Bioinf.* 77:89–99.

[54] Song, Y., F. DiMaio, R. Y.-R. Wang, D. Kim, C. Miles, T. Brunette, J. Thompson, and D. Baker, 2013. High-resolution comparative modeling with RosettaCM. *Structure* 21:1735–1742.

[55] Case, D. A., T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, 2005. The amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.

[56] Salomon-Ferrer, R., D. A. Case, and R. C. Walker, 2013. An overview of the amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* 3:198–210.

[57] Case, D. A., K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny,

T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, and P. A. Kollman, 2020. AMBER 2020. University of California, San Francisc.

[58] Wright, J. D., S. Y. Noskov, and C. Lim, 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.

[59] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242. RCSB PDB: Homepage. https://www.rcsb.org/.

[60] Burley, S. K., H. M. Berman, C. Bhikadiya, C. Bi, L. Chen, L. D. Costanzo, C. Christie, K. Dalenberg, J. M. Duarte, S. Dutta, Z. Feng, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranović, D. Guzenko, B. P. Hudson, T. Kalro, Y. Liang, R. Lowe, H. Namkoong, E. Peisach, I. Periskova, A. Prlić, C. Randle, A. Rose, P. Rose, R. Sala, M. Sekharan, C. Shao, L. Tan, Y.-P. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Woo, H. Yang, J. Young, M. Zhuravleva, and C. Zardecki, 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. RCSB PDB: Homepage.
https://www.rcsb.org/.

[61] Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, 2015. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

[62] Izadi, S., R. Anandakrishnan, and A. V. Onufriev, 2014. Building water models: a different approach. *J. Phys. Chem. Lett.* 5:3863–3871.

[63] Li, P., L. F. Song, and K. M. Merz Jr., 2015. Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* 11:1645–1657.

[64] Li, P., and K. M. Merz Jr., 2014. Taking into account the ion-induced dipole interaction in the nonbonded model of ions. *J. Chem. Theory Comput.* 10:289–297.

[65] Jorgensen, W. L., J. Chandrasekhar, and J. D. Madura, 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

[66] Tian, C., K. Kasavajhala, K. A. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, , and C. Simmerling, 2020. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* 16:528–552.

[67] Peters, M. B., Y. Yang, B. Wang, L. Füsti-Molnár, M. N. Weaver, and K. M. Merz Jr., 2010. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* 6:2935–2947.

[68] Pastor, R. W., B. R. Brooks, and A. Szabo, 1988. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* 65:1409–1419.

[69] Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

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

[71] van Gunsteren, W. F., and H. J. C. Berendsen, 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34:1311–1327.

[72] Miao, Y., V. A. Feher, and J. A. McCammon, 2015. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* 11:3584–3595.

[73] R Core Team, 2020. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

[74] Van Rossum, G., and F. L. Drake, 2009. Python 3 reference manual. CreateSpace, Scotts Valley, CA.

[75] Grant, B. J., A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, 2006. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22:2695–2696.

[76] Skjæven, L., G. Yao, Xin-Qiu andScarabelli, and B. J. Grant, 2014. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinf.* 15:399.

[77] Skjæærven, L., S. Jariwala, X.-Q. Yao, and B. J. Grant, 2016. Online interactive analysis of protein structure ensembles with Bio3D-web. *Bioinformatics* 32:3510–3512.

[78] Lobanov, M. Y., N. S. Bogatyreva, and O. V. Galzitskaya, 2008. Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* 42:623–628.

[79] David, C. C., and D. J. Jacobs, 2014. Principal component analysis: a method for determining the essential dynamics of proteins. *In* D. R. Livesay, editor, Protein Dynamics: Methods and Protocols, Humana Press, Totowa, NJ, volume 1084 of *Methods in Molecular Biology*, 193–226.

[80] Wolf, A., and K. N. Kirschner, 2013. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J. Mol. Model.* 19:539–549.

[81] Charrad, M., N. Ghazzali, V. Boiteau, and A. Niknafs, 2014. NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Software* 61:1–36.

[82] Miao, Y., W. Sinko, L. Pierce, D. Bucher, R. C. Walker, and J. A. McCammon, 2014. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *J. Chem. Theory Comput.* 10:2677–2689.

[83] Rodríguez, A., and C. Tsallis, 2010. A generalization of the cumulant expansion. Application to a scale-invariant probabilistic model. *J. Math. Phys.* 51:073301.

[84] Chen, V. B., W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson, 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* D66:12–21.

[85] Benkert, P., M. Biasini, and T. Schwede, 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27:343–350.

[86] Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, and T. Schwede, 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:W296–W303.

[87] Bowie, J. U., R. Lüthy, and D. Eisenberg, 1991. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164–170.

[88] Roland Lüthy, D. E., James U. Bowie, 1992. Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85.

[89] Colovos, C., and T. O. Yeates, 1993. Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci.* 2:1511–1519.

[90] Pontius, J., J. Richelle, and S. J. Wodak, 1996. Deviations from standard atomic volumes as a quality measure for protein crystal structures. *J. Mol. Biol.* 264:121–136.

[91] Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton, 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

[92] Wells, M., H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, and A. R. Fersht, 2008. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U.S.A.* 105:5762–5767.

[93] Hess, B., 2002. Convergence of sampling in protein simulations. *Phys. Rev. E: Stat. Nonlin. Soft Matter Phys.* 65:031910.

[94] Chen, Y., R. Dey, and L. Chen, 2010. Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. *Structure* 18:246–256.

[95] Cho, Y., S. Gorina, P. D. Jeffrey, and N. P. Pavletich, 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265:346–355.

[96] Eldar, A., H. Rozenberg, Y. Diskin-Posner, R. Rohs, and Z. Shakked, 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.

[97] Thayer, K. M., and T. R. Quinn, 2016. p53 R175H hydrophobic patch and H-bond reorganization observed by MD simulation. *Biopolymers* 105:176–185.

[98] Pérez-Cañadillas, J. M., H. Tidow, S. M. V. Freund, T. J. Rutherford, H. C. Ang, and A. R. Fersht, 2006. Solution structure of p53 core domain: structural basis for its instability. *Proc. Natl. Acad. Sci. U.S.A.* 103:2109–2114.

9 Tables

Table 1 Summary of system preparation protocol

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		25.93	58.98	1000
Minimization 5	NVT			20.0	25.93	58.98	2000
Heating[e]	NV			20.0	25.93	58.98	500000
Equilibration 1[e]	NVT			20.0	25.93	58.98	100000
Equilibration 2[e]	NPT			20.0	25.93	58.98	100000
Equilibration 3[e]	NPT			15.0	25.93	58.98	100000
Equilibration 4[e]	NPT			10.0	25.93	58.98	100000
Equilibration 5[e]	NPT			5.0	25.93	58.98	100000
Equilibration 6[f]	NPT			1.0	25.93	58.98	100000
Equilibration 7[f]	NPT			0.5	25.93	58.98	150000
Equilibration 8[f]	NPT				25.93	58.98	5000000

[a] k_f: force constant in units of kcal/(mol $\cdot$ 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.

10 Figure Legends

Figure 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); spring green, DNA binding domain (DBD); gray, flexible linker; orange, oligomerization domain (OD); red, regulatory domain (RD). The mutated residue is depicted with CPK representation. This model refers to the equilibrated distal-mutant conformation.

Figure 2

Root-Mean-Square Fluctuation (RMSF) for p53 forms: (a) inactive p53; red, distal-active p53; green, proximal-active p53; blue. (b) distal-mutant p53; green, proximal-mutant p53; blue. The rectangular region represents the p53-DBD region.

Figure 3

Radius of Gyration (Rgyr) for p53 forms: (a) inactive p53; red, distal-active p53; green, proximal-active p53; blue. (b) distal-mutant p53; green, proximal-mutant p53; blue.

Figure 4

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

Figure 5

Electrostatic potential surface around distal-mutant form: (a) cluster 1, (b) cluster 2. The mutated residue is depicted with ball and stick representation.

Figure 6

Electrostatic potential surface around proximal-mutant form: (a) cluster 1, (b) cluster 2. The mutated residue is depicted with ball and stick representation.

Figure 7

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

11 Figures

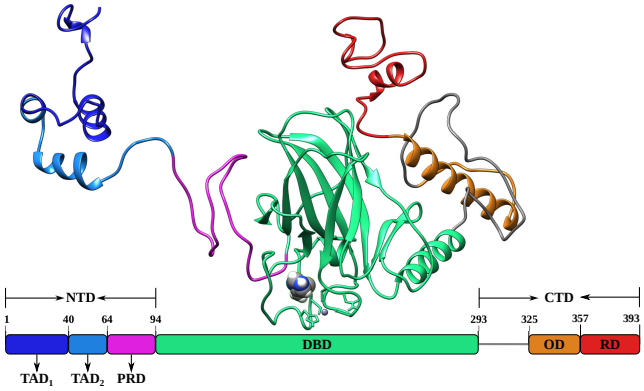


Fig. 1

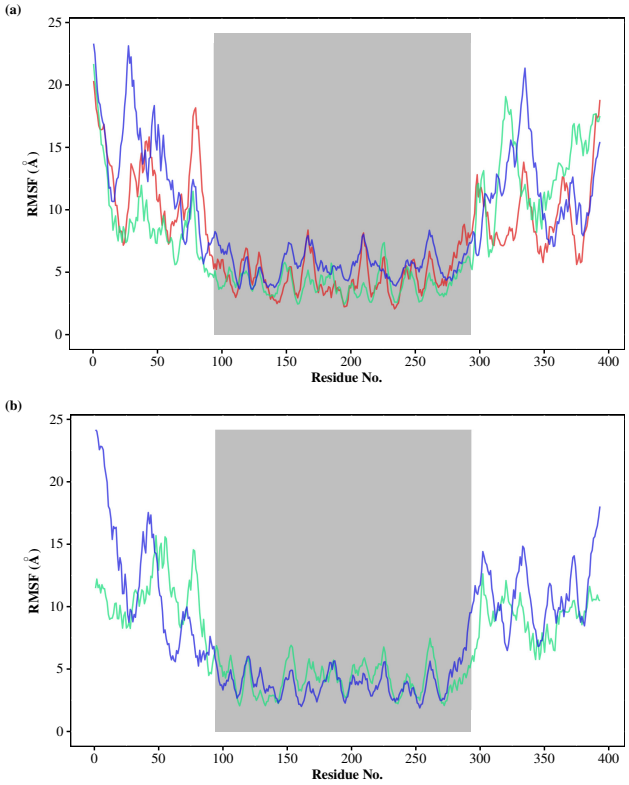


Fig. 2

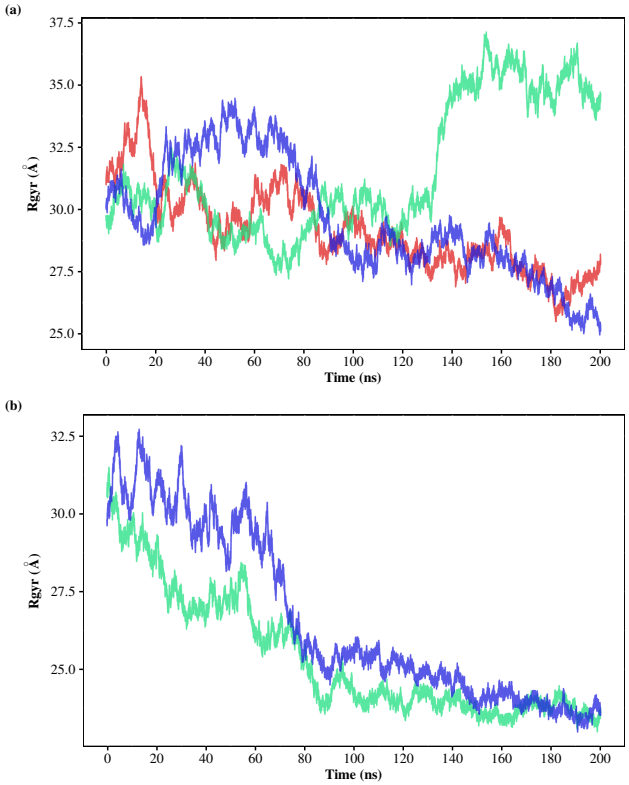

(a)
37.5
35.0
32.5
30.0
27.5
25.0
Rgyr (Å)
0
20
40
60
80
100
120
140
160
180
200
Time (ns)
(b)
32.5
30.0
27.5
25.0
Rgyr (Å)
0
20
40
60
80
100
120
140
160
180
200
Time (ns)

Fig. 3

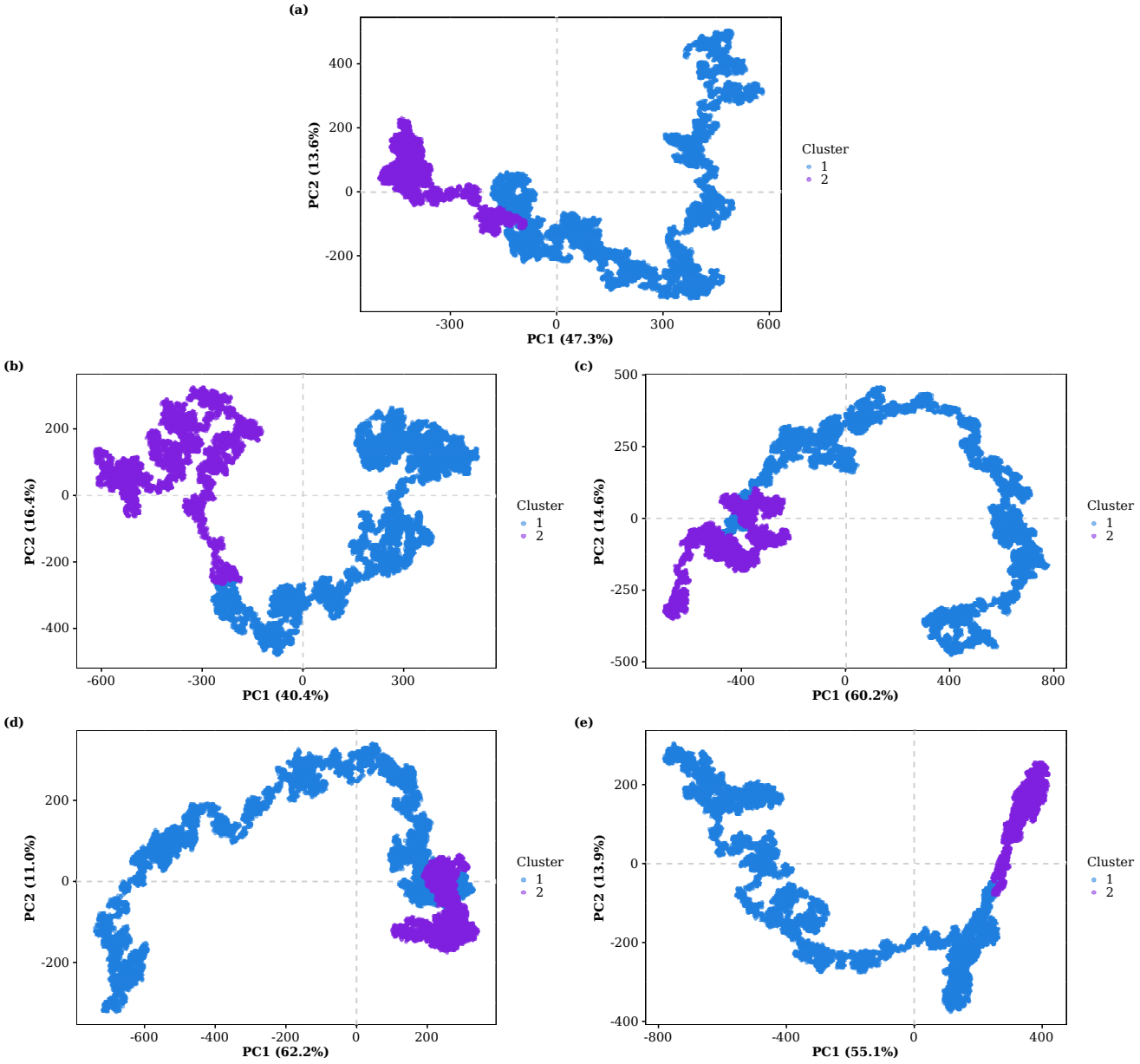

(a)
PC2 (13.6%)
PC1 (47.3%)
Cluster
1
2
(b)
PC2 (16.4%)
PC1 (40.4%)
(c)
PC2 (14.6%)
PC1 (60.2%)
(d)
PC2 (11.0%)
PC1 (62.2%)
(e)
PC2 (13.9%)
PC1 (55.1%)

Fig. 4

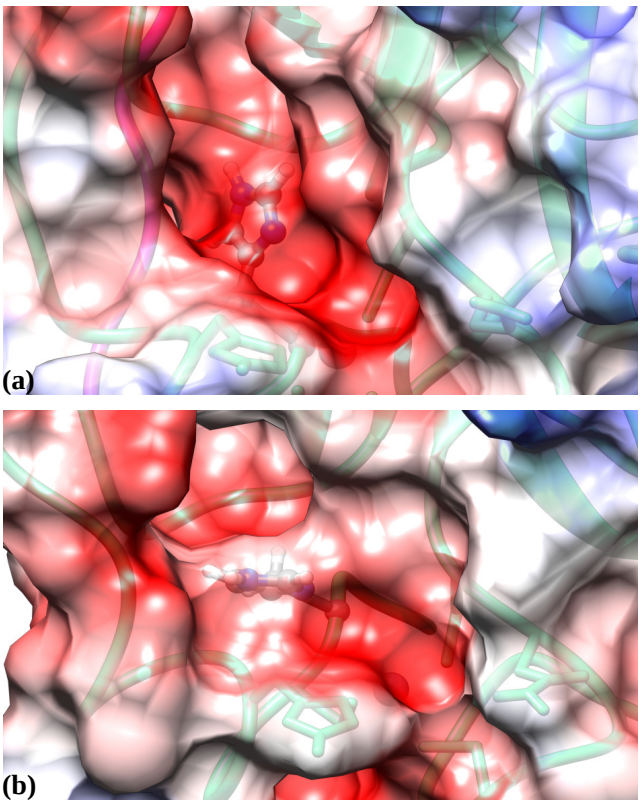
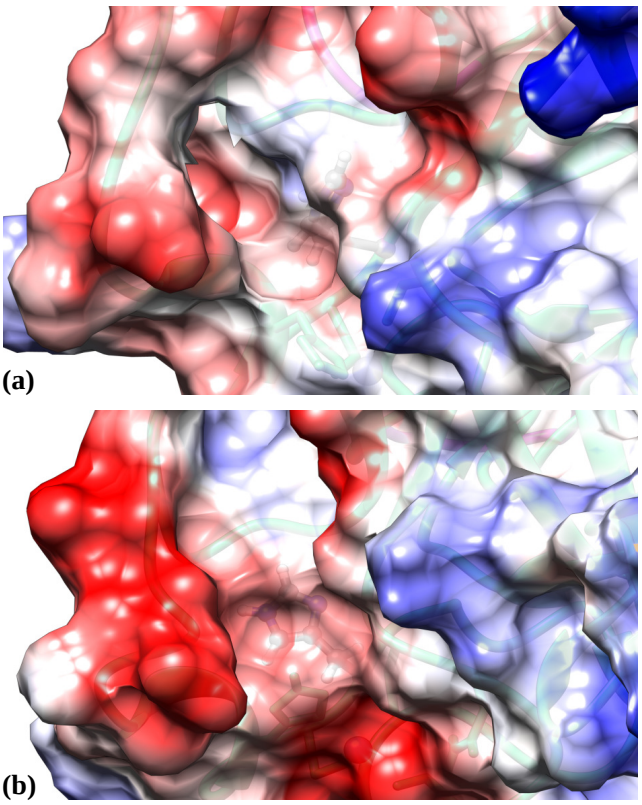


Fig. 5



(a)

(b)

Fig. 6

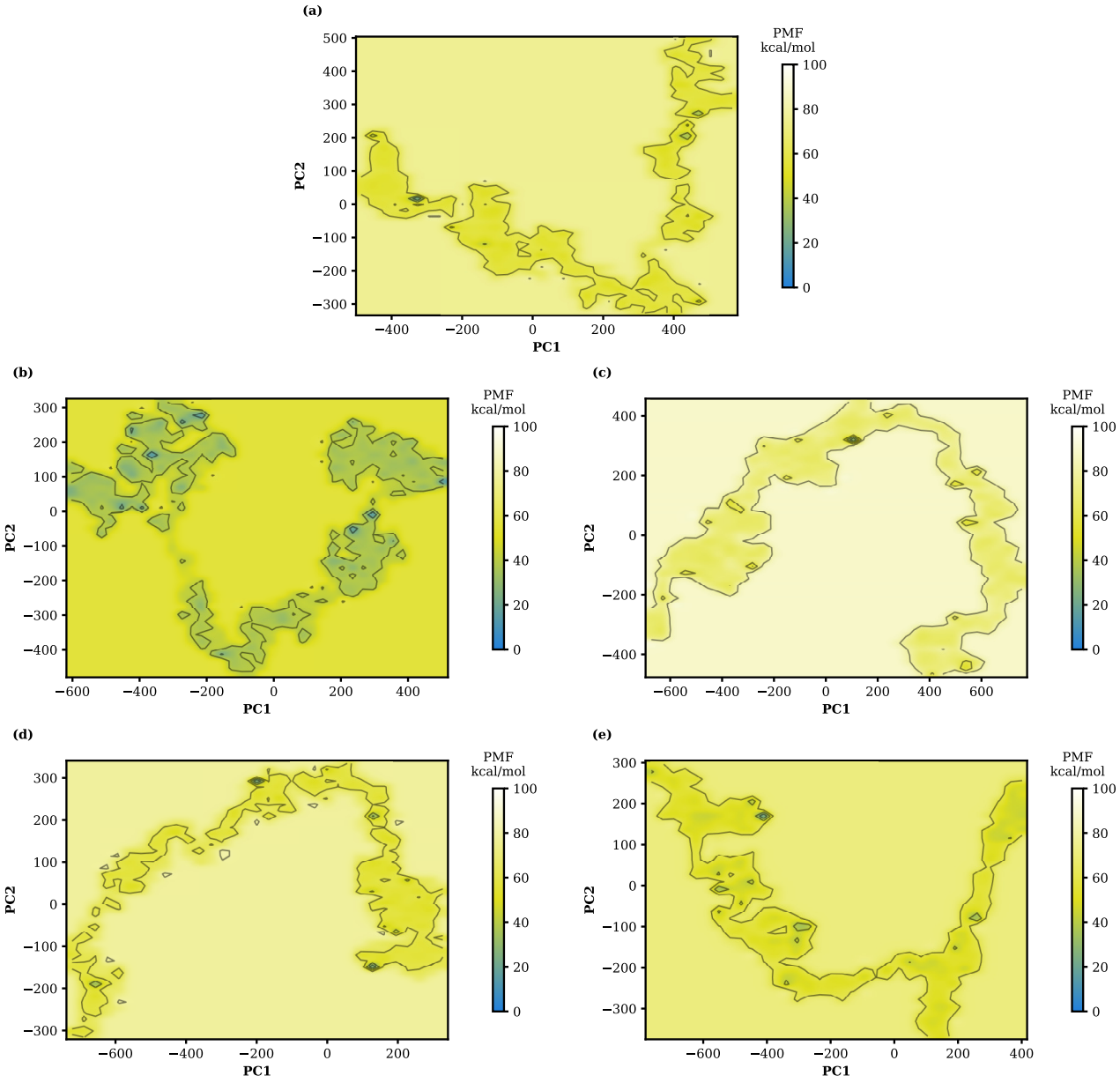


Fig. 7