JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING
Cite This: J. Chem. Inf. Model. 2019, 59, 3316−3329
Article
pubs.acs.org/jcim

Characterization of Hydration Properties in Structural Ensembles of Biomolecules

Mohan R. Pradhan,[†,‡] Minh N. Nguyen,[†] Srinivasaraghavan Kannan,[†] Stephen J. Fox,[†] Chee Keong Kwoh,[‡] David P. Lane,[§] and Chandra S. Verma*[,†,∥,⊥]

[†]Bioinformatics Institute, A*STAR (Agency for Science, Technology and Research), 30 Biopolis Street, #07-01 Matrix, Singapore 138671

[‡]School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

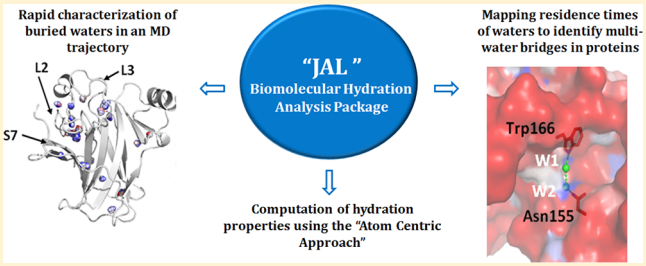
[§]p53 Laboratory, A*STAR (Agency for Science, Technology and Research), 8A Biomedical Grove, #06-04/05, Neuros/Immunos, Singapore 138648

[∥]School of Biological Sciences, Nanyang Technological University, 50 Nanyang Drive, Singapore 637551

[⊥]Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543

Supporting Information

ABSTRACT: Solute−solvent interactions are critical for biomolecular stability and recognition. Explicit solvent molecular dynamics (MD) simulations are routinely used to probe such interactions. However, detailed analyses and interpretation of the hydration patterns seen in MD simulations can be both complex and time-consuming. A variety of approaches/tools to compute and interrogate hydration properties in structural ensembles of proteins, nucleic acids, or in general any molecule are available and are complemented here with a new and free software package ("JAL"). Central to "JAL" is an intuitive atom centric approach of computing hydration properties. In addition to the standard metrics commonly used to understand hydration, "JAL" introduces two nonstandard utilities: a program to rapidly compute buried waters in an MD trajectory and a new method to compute multiwater bridges around a solute. We demonstrate the utility of the package by probing the hydration characteristics of the tumor suppressor protein p53 and the translation initiation factor eif4E. "JAL" is hosted online and can be accessed for free at http://mspc.bii.a-star.edu.sg/minhn/jal.html.



INTRODUCTION

Optimal levels of hydration are critical for maintaining the native states of proteins and in mediating their interactions with other molecules. Aberrant solvation during protein folding may lead to a misfolded or partially folded protein, resulting in aggregation.[1] Additionally, water interactions with backbone hydrogen bonds in proteins have been shown to result in charge-hydrophobe interactions at protein−protein/ligand interfaces.[2−7] Water can also act as a catalyst in sampling protein conformations. Their small size enables them to rearrange quickly, thus facilitating conformational rearrangements in proteins, i.e., smoothening the potential energy surface of proteins.[8] Hence, detailed pictures of protein hydration are necessary to obtain insights into the thermodynamics and kinetics of intramolecular and intermolecular interactions.[9−13]

Unsurprisingly, protein hydration is extremely sensitive to surface chemistry and topology[14−16] which have very diverse characteristics, making the study of hydration very challenging. Several experimental approaches[17] have been used to study biomolecular hydration. For example, the mining of multiple, independently determined, high resolution crystal structures of proteins in the Protein Data Bank (PDB)[18] have led to the identification of consensus hydration sites.[19] However, crystal structures represent an average conformation, and while B-values have provided some guidance on flexibility, a more dynamic picture is necessary for comprehensively understanding biological function.[20,21] This need has been complemented by developments in experimental techniques such as Nuclear Magnetic Resonance (NMR),[22] femtosecond spectroscopy,[23] vibrational spectroscopy,[24] ultrafast optical Kerr-Effect spectroscopy,[25] and 4D Ultrafast Electron Crystallography[26] and more recently by advances in computational techniques such as molecular dynamics (MD) simulations.[27−31] MD provides a detailed spatiotemporal picture of conformational variability and associated dynamic interactions, including those with solvent molecules.

A range of computational tools has been developed in the past to predict water locations around a protein structure and has provided excellent insights into structure−function relationships and mechanisms (Table S3). Very early on, GRID[32] was the method of choice; this involved placing a probe such as

Received: July 9, 2018
Published: May 29, 2019

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

water, methyl group, amine nitrogen, carboxy oxygen, and hydroxyl molecule around the protein and evaluating its interaction potential to identify hydration sites. Energy levels for each probe were then contoured and filtered to identify important interaction sites. Next appeared a method by Rashin and Honig[33] who developed a rapid method to find cavities based on SASA and predict buried waters by evaluating the ability of waters to form hydrogen bonds; their method successfully reproduced 80% of the buried waters in a test set of 12 crystallographically resolved proteins. This was followed by AQUARIUS[34] which utilized information mined regarding amino acid specific hydration patterns from structures in the PDB to compute solvent structures around proteins by computing likelihood weights for grid points; they were able to achieve ∼60% accuracy. Zhang and Hermans developed Dowser[35] where the energetics of solvation was calculated for cavities in proteins; given the nature of the calculations and the computer hardware available during that period, this was somewhat slow; however, they achieved around 90% success for locating buried waters in the protein subtilisin. The program SuperStar[36] appeared next from Verdonk and colleagues, creating probability density maps for various functional groups based on mining nonbonded interactions in structures from the PDB[37] and CSD (Cambridge Structural Database)[38] and find that they are generally good predictors. In 2005, the FOLD-X force field[39] was used to evaluate the energetics of protein−water interactions using a force field and identified 76% of the waters that make at least two hydrogen bonds with the protein. A rigorous method developed by Hirata and co-workers[40] used 3D-RISM theory and was able to correctly predict all six buried waters that were observed to be located in protein cavities in crystal structures of hen egg-white lysozyme. In 2009, a sophisticated statistical mechanics based approach called JAWS[41] was reported which computed the thermodynamics of water binding at sites in proteins and could successfully locate hydration sites in high resolution crystal structures of five different proteins. HyPred[42] was developed in 2010 to predict solvent density using radial distribution functions for a set of atom types around a constrained protein in an all atom explicit solvent MD simulation and was successful in predicting 50% of the crystallographically observed waters in three different proteins. The same year also witnessed the publication of a method that involved the iterative placement of water sites on a lattice and solving the solvent distribution using a semiheuristic cellular automata approach for computation of water propensities at protein sites by evaluating interactions using a mean field approach.[43] Mobywat[44] used MD simulations to compute hydration patterns based on water dynamics by factoring all interatomic interactions and demonstrated 80% agreement across 1500 crystallographic waters in 20 proteins. More recently, a new robust and rapid method based on a semiexplicit, discrete solvation model was proposed by Setny[45] which produced results that are in very good agreement with the rigorous and highly computationally demanding double decoupling free energy calculations (the energetics computed by these two methods are within 2 kcal/mol of each other). In 2016, Dowser++[46] was reported which was a semiempirical modification of the earlier program Dowser;[35] it combined two methods, AutoDock Vina[47] and WaterDock (which uses the docking program AutoDock Vina to dock water molecules in proteins and achieves 97% accuracy[48]), and correctly predicted most of the buried waters in 14 crystal structures together with providing a quantitative estimate for the quality of water placement.

In parallel, plugins to commonly used modeling/graphics programs were developed. Watclust is a new VMD plugin that can determine and analyze hydration sites and can directly utilize this information in subsequent docking studies in Autodock.[49] WATsite[50] and PyWATER[51] are plugins to the Pymol visualizer[52] and predict invariant waters as well as hydration sites in binding interfaces based on analyses of superimposed structures using clustering. WATsite[50] further elucidates the thermodynamic profile of a water molecule and its potential contribution to ligand binding.

Other efforts include setting up of publicly available Web resources for different aspects of the analysis of water interactions. 3DSS[53] is a Web server for superimposition of multiple structures of a protein and computation of invariant water molecules. Similarly, Water Analysis Package (WAP)[54] and Protein Structure Analysis Package (PSAP)[55] are Web servers that compute and analyze water interactions with proteins in a single input structure. A variety of programs for computing solvent properties in an ensemble of conformations are embedded in the standard trajectory analysis modules of the MD simulation packages CHARMM,[56] Amber,[57] and GROMACS.[58]

With increasing access to high performance computing, MD simulations in aqueous solvents are becoming common. While there are several resources available that can enable the computation of various characteristics of motions of proteins and of protein−ligand (including water) interactions (see above), analysis of hydration characteristics has been more challenging. Sophisticated and computationally demanding techniques of course have been developed including Grand Canonical Monte Carlo[59] and free energy methods,[60] but these preclude their use for rapid analysis. Of course, continuum models of water as a dielectric have been used successfully in several cases,[61] yet these methods average out any specific hydrogen bond effects and localized dielectric effects. These issues are important because waters not only occupy cavities in proteins and affect protein stabilities[62,63] but also mediate bridging interactions between ligands and proteins;[64] these issues will impact upon protein and ligand design;[65,66] for example knowledge of the geometry of a hydrogen bond made by a buried water will be critical in deciding whether a polar side chain would replace it appropriately in the design of a mutation in a protein or in the sculpting of a new functionality in a ligand toward greater stability/affinity. The robust and rapid detection of buried waters and water bridges from MD simulations was given a boost by the development of two sophisticated and yet rapid methods by the groups of Setny[45] and of Gilson and colleagues.[67] The former method is based on a semiexplicit, discrete solvation model that is used to compute the energetics of interactions of water molecules from an MD simulation. The latter method combines standard MD with translational Metropolis Monte Carlo (MC) moves in which water molecules transit between the bulk solvent and the protein interior, allowing for stabilization of water in cavities that have the potential to be hydrated. Complementing these methods, we present here a new software package ("JAL") that enables the computational detection of buried waters and water bridges in MD simulations; JAL also computes the standard metrics (solvent density, residence times, closest-water distance measures) from MD simulations. Using as examples the tumor suppressor p53 DNA Binding Domain (DBD) and the

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

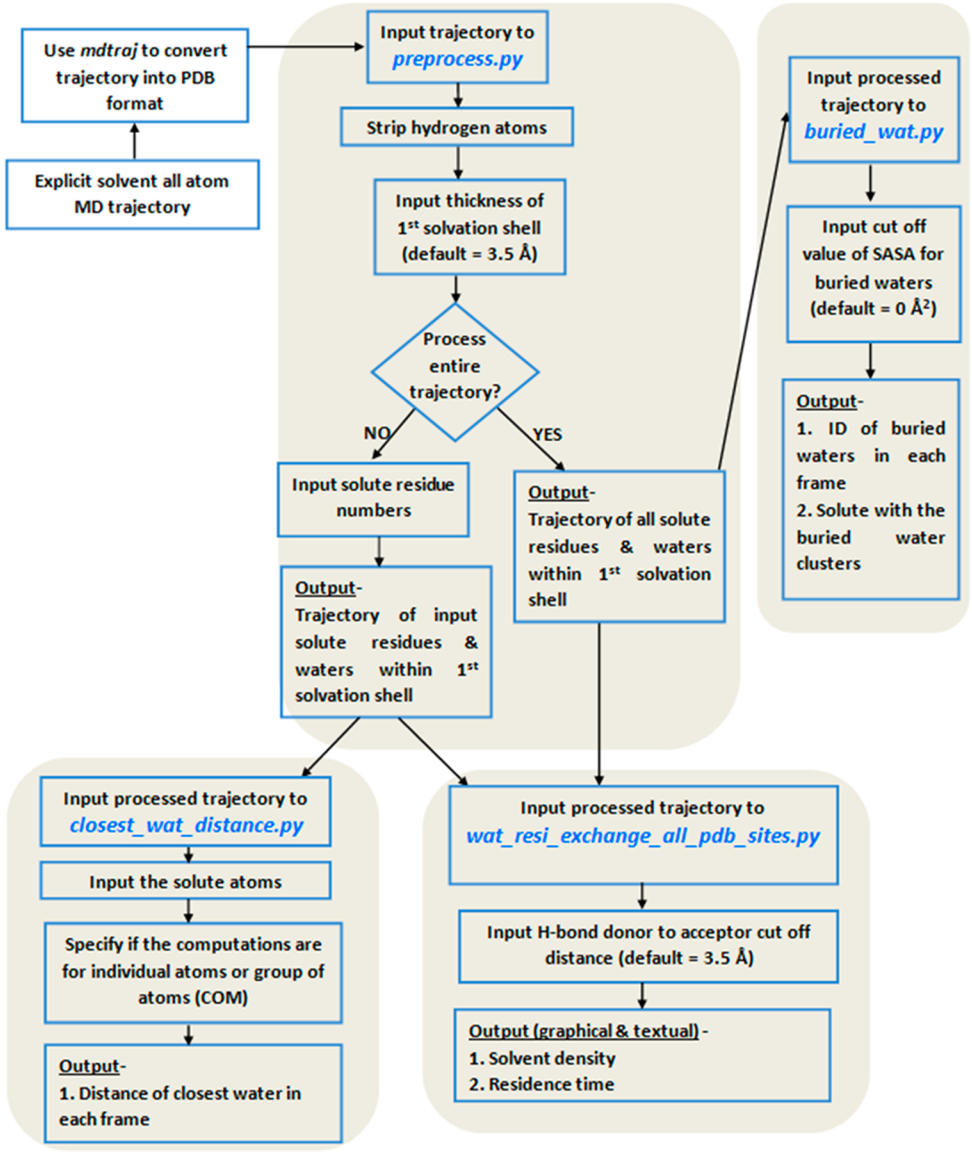


Figure 1. Hydration metrics embedded in JAL. Keyword "COM" refers to the Center of Mass.

translation initiation factor eIF4E, we demonstrate how such analyses can reveal interesting insights into protein structure and function. In summary, JAL facilitates detailed characterization of hydration properties of solutes from MD simulations. The programs in the "JAL" package, together with detailed documentation, are available at http://mspc.bii.a-star.edu.sg/minhn/jal.html.

METHODS

Hydration Properties. The programs developed here enable a rigorous characterization and an intuitive analysis of hydration properties in MD simulation trajectories using the metrics described in Figure 1. The program takes as input a single trajectory file in PDB format consisting of frames extracted from an MD trajectory. This single file can be created using the mdtraj module[68] of python by extracting structures from MD trajectories in diverse formats (xtc, trr, dcd, binpos, netcdf, mdcrd, etc.). The input PDB file is first preprocessed ('preprocess.py') by removing the hydrogen atoms and saving the coordinates of the solute (which is either the whole molecule or a set of atoms that is defined by the user) and the water (or solvent) molecules (which are either those in the first hydration shell, defined as within 3.5 Å of any solute heavy atom, or a cut off specified by the user). The default 3.5 Å cut off is optimal because it is taken as the maximum hydrogen bond donor to hydrogen bond acceptor distance for a hydrogen bond.[69] Hydration metrics (closest water distance, residence time, and solvent density) are computed by tracking the water molecules that lie within a sphere of 3.5 Å radius from the reference (non-hydrogen protein) atoms in each snapshot of the input structural ensemble. This (protein) atom centric approach of computing the solvent properties is different from a grid based approach (for instance the "grid" command in ptraj module of Amber). A grid based approach creates a 3D grid over the water box and then counts the number of the solvent molecules within each grid box (usually of 0.5 Å dimension). In the protein (atom) centric approach, we monitor the protein (non-hydrogen) atoms that interact with the solvent (water oxygen). Although the current approach is computationally more expensive than the grid based approach, it does have several advantages (elaborated in the Results and Discussion sections). The hydration programs produce two kinds of output for analysis -

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

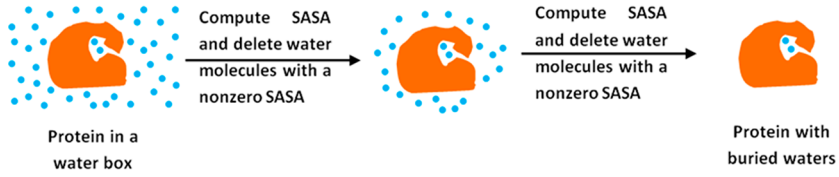


Figure 2. Schematic diagram to demonstrate the approach used in JAL to compute buried waters in an MD trajectory. The algorithm to compute buried waters entails "evaporation" of the exposed waters (based on accessible surface area) until only buried waters remain in protein cavities. The computations shown in the figure are performed for every frame in the MD trajectory.

graphical and textual. The detailed nature of the output and the associated analyses will be discussed in the Results section, and the documentation is provided on the "JAL" Web site.

Along with the standard hydration metrics (solvent density, residence time, and closest distance measure), "JAL" introduces two new methods: (1) a method to compute buried waters, which is benchmarked using BPTI (Figure S2) and subsequently used to explore the structural/functional aspects of hydration in the DNA binding domain (DBD) of the human tumor suppressor protein p53, and (2) a method to compute multiwater bridges around a solute based on residence time calculations, which we use to probe the hydration of the human translation initiation factor eIF4E. We carry out rigorous double decoupling free energy calculations to substantiate our findings.

The hydration programs embedded in the "JAL" package compute the following hydration properties:

1. Solvent density: The number of times a water oxygen atom is observed within a distance of 3.5 Å from the reference atom. The reference atom can be input by the user or can be all the non-hydrogen solute atoms. It is measured as percent hydration across the ensemble.
2. Residence time: It is an estimate of how stable/buried a water molecule is at a particular site. The input for the program is an evenly spaced (snapshots at fixed time interval τ) MD trajectory. The value of τ used for the computation of residence time in this study is 100 ps. The residence time is defined as the length of time a distinct water molecule stays continuously within 3.5 Å of the solute site. This produces a list of residence time values for that site across the MD trajectory. The summation of all such values at a site gives the total residence time, the average of this gives the average residence time, and the maximum value from this list is reported as the maximum residence time (the total residence time will depend on the length of the trajectory). Effectively, the program computes the residence time autocorrelation function $P(\tau)$ (as also done in ref 70) but does not fit an exponential or multiexponential model.
3. Distance of closest water: It computes the distance between a solute atom and the closest solvent molecule. The metric estimates the accessibility of a site on the solute to the water and hence provides a likely estimate of the strength of the interaction.
4. Buried waters: The core of a protein is usually well packed but can occasionally contain cavities. Hydration of the cavity depends on the volume and residues lining it. These (buried) waters, which are usually found in the protein cavities and core, have little solvent accessibility and are referred to as buried waters. Buried waters are computed using a novel methodology based on iterative calculations of Solvent Accessible Surface Area (SASA) of water molecules and deleting those waters which have SASA more than zero after every iteration (analogous to "evaporating" waters from the protein surface) (Figure 2). Ultimately waters left with a SASA value of zero are the ones deemed to be buried. We benchmarked our calculations of buried waters using bovine trypsin inhibitor (PDB id 4PTI, resolution 1.5 Å[71]). We correctly identify the four buried waters (Figure S2) which are characterized by low B-values and engage in a strong interaction network with the protein.

The program to compute buried waters takes as input an MD trajectory of the solute (protein) with the water box. For every frame, the program computes the SASA of all the heavy atoms in the system using the rolling ball algorithm.[72] This entails rolling a probe of the size of a water molecule (radius of 1.4 Å) on the system (protein + water box). In the next step, all the waters with a nonzero SASA value are deleted. Computation of SASA of all heavy atoms followed by deletion of the waters with a nonzero SASA value is iterated until only waters with a SASA value of zero (buried waters) are left; this value can be changed to nonzero values to reflect almost completely buried waters.

Molecular Dynamics Simulations. We probed these properties in all-atom explicit solvent MD simulations of the wild type (WT) and the V143A mutant of the p53 DNA Binding Domain (DBD) (PDB id of WT is 2AHI, chain A, resolution 1.85 Å[73]) and separately of the translational initiation factor eIF4E (PDB id 2W97, chain B, resolution 2.29 Å[74]) in their apo states. V143A is one of the most common destabilizing mutations in the DBD of p53 and is known to result in the unfolding and aggregation of p53 under physiological conditions. There exists a crystal structure of this mutant (PDB id 2J1W), but this was generated against a background of stabilizing mutations; we wished to examine the effects of this mutation in the WT protein and so introduced the mutation in the WT crystal structure, allowing its effects to evolve during the subsequent MD simulations. Simulations were carried out using the Amber ff99SB force field.[75] The N and C termini of the proteins were capped with ACE and NME moieties, respectively. All the crystallographic waters were retained. Zinc is coordinated to four residues in the loops L2 and L3 of the DBD. These include Cys176 (atom SG), Cys238 (atom SG), Cys242 (atom SG), and His179 (atom ND1). We used the bonded model of Zinc[76] and covalently bonded it to these four residues as it preserves the tetrahedral coordination which has been reported to be crucial for the stability of loops L2 and L3 and also for the interactions of the DBD with DNA.[77] The cysteines 176, 238, and 242 were deprotonated at their sulfur atoms thus enabling them to be covalently linked to the zinc. The zinc was also covalently linked to the ND1 atom of His179, and the proton was placed on the NE2 atom and corresponds to its protonation state at neutral pH. The proteins were solvated in a cubic water box with a minimum separation of 10 Å between the protein atoms and the edge of the water box. The TIP3P[78]

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

water model was used to model the water molecules. This was followed by neutralization of the systems with the addition of an appropriate number of counterions. The solute−solvent systems were subjected to an energy minimization involving 1000 steps each of steepest descent followed by an equal number of conjugate gradient steps. We gradually heated the systems to 310 K in an NVT ensemble followed by 2 ns of equilibration in the NPT ensemble. Finally, a 100 ns production run in NPT was carried out for each system using the GPU implementation of PMEMD.[79] A cut off of 8 Å was used to evaluate the Lennard-Jones (LJ) and short-range electrostatic interactions, while the long-range electrostatic interactions were evaluated using the particle-mesh Ewald (PME) method. Langevin thermostat[80] was used to regulate the temperature. A constant pressure (1.0 bar) was maintained during the simulations using a weak coupling with a relaxation time of 1 ps. The SHAKE algorithm[81] was used to constrain the lengths of the covalent bonds associated with every hydrogen atom, thus enabling an integration time step of 2 fs to be used in the simulations; coordinates were saved at intervals of 5 ps. Hydration analysis is carried out on the last 50 ns sampled. The appropriate values for these parameters are dependent on the protein system being studied and the purpose of the calculations. Simulations were performed in triplicates.

JAL is also compared with a recently published method to predict buried waters.[45] The author of that study tested their method on eight proteins (PDB ids: 1AHO chain A, 1BAS chain A, 1SNB chain A, 1UOY chain A, 1W8V chain A, 2YGS chain A, 3HVV chain A, 3Q7Y chain A). We have additionally carried out MD simulations on these eight proteins using the protocol described in the previous paragraph.

Buried Water Clusters in p53 DBD. To demonstrate the utility of JAL in the computation of buried waters, we analyzed the crystallographic waters in crystal structures of the DBD with resolution ≤2.5 Å from the PDB.[37] Multiple chains of the DBD within a PDB entry were considered as separate structures. This resulted in a total of 119 unique structures (Table S1). For each structure, the iterative procedure to "evaporate" the nonburied waters was carried out. Structures of the DBD with buried waters (if any) were next superimposed using the "align" function in PyMOL[52] on to the structure of the DBD with an extended N-terminus (PDB id 2XWR; 1.68 Å; chain A). We find that several structures of DBD have buried waters but with varying locations within the protein. For the purposes of this study we decided to focus on those crystallographic buried water molecules whose positions were invariant in at least 50% of the superimposed structures (Figure 3A). Protein structures being dynamic in MD simulations, we identify the buried water clusters in the MD simulations as the ones which are present in at least 5% of the simulation (Figure 3C). Waters in the superimposed structures which are within 1.7 Å of each other were considered as belonging to the same cluster.

Double Decoupling Method. A systematic analysis of the buried waters in the MD simulations of the WT p53 DBD revealed one highly persistent buried water molecule in loop L3. However, this water molecule is lost in the simulations of the V143A mutant where loop L3 assumes a non-native conformation. To quantify the free energy of binding of this water molecule to WT p53 DBD, we use the double decoupling method.[60]

The double decoupling method is a two-step process for computing the free energy of the ligand (water) by gradually switching off the van der Waals and electrostatic interactions

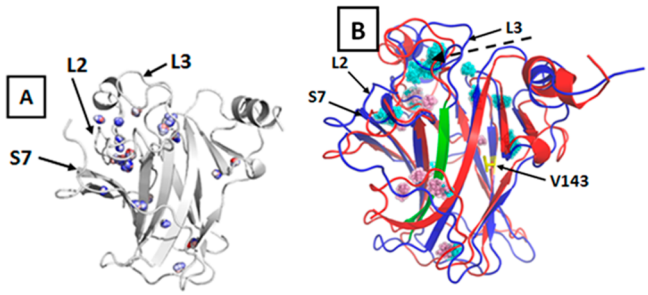


Figure 3. (A) Clusters of buried waters (colored according to the B-value; a color gradient of blue-white-red denotes low to high B-values) present in at least 50% of the 119 superimposed crystal structures of human p53 DBD. (B) Clusters of buried waters, shown as cyan and pink spheres, as present in at least 5% of the MD simulations of WT (PDB id 2AHI chain A) and the V143A (constructed in this study) in blue and red cartoons, respectively. The site of mutation (V143A) is shown with yellow sticks. Strand colored in green represents the Aggregation Prone Region (APR). A buried water cluster marked with a dashed line is the prominent difference between the two forms of the protein since it is present almost exclusively in WT.

(referred to as decoupling) with the rest of the system in a two-step process (as shown below):

$$W_{sol} \rightarrow W_{gas} \qquad \Delta G^{w}_{dec}$$

$$PW_{sol} \rightarrow P_{sol} + W_{gas} \qquad \Delta G^{pw}_{dec}$$

$$W_{sol} + P_{sol} \rightarrow PW_{sol} \qquad \Delta\Delta G_{abs} = \Delta G^{w}_{dec} - \Delta G^{pw}_{dec}$$

In the first step, the free energy of decoupling a water molecule (ΔG^{w}_{dec}) from the bulk solvent was computed and found to be −6.1 kcal/mol, which agrees very well with values computed earlier by others (−6.0 kcal/mol)[60,65] and with experiment (−6.3 kcal/mol),[82] suggesting that our protocol is robust. In the second step, the bound water molecule is decoupled from the protein−water system (ΔG^{pw}_{dec}). The absolute binding free energy of the water ($\Delta\Delta G_{abs}$) is computed using equations listed here. Decoupling is carried out over a span of 101 windows (λ values = 0.00, 0.01, . . ., 1.00) at the simulation temperature of 310 K. During gradual decoupling of the electrostatic and van der Waals interactions of the water molecule from the protein−water system, an appropriate restraint weight needs to be imposed on the water molecule (to prevent it from flying off from its binding site). This restraint is proportional to the dynamics of the water molecule at that site in a canonical MD simulation of WT DBD and is computed from its mean square displacement (MSD) as follows:

$$k = \frac{3RT}{\langle \delta r^2 \rangle}$$

MSD for the buried water molecule in loop L3 of WT DBD was 3.74 Å^2, and the corresponding restraint weight (which was calculated as 0.49 kcal/mol) ensures that the water molecule remains bound to the protein during the decoupling process and the restraint is sufficient for the water molecule to explore the conformational space accessible in the binding pocket. The same restraint (0.49 kcal/mol) is also imposed on the protein Cα atoms to avoid protein rotations and translations during the simulation. We use the MBAR (Multistate Bennett Acceptance Ratio) method for calculating the free energy differences.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

RESULTS

Case Study 1: The Role of Buried Waters in the Stability of the p53 DNA Binding Domain. To evaluate the method for identifying buried waters in an MD trajectory using 'JAL', we investigate the hydration of p53 DBD. DBD is involved in numerous interactions with DNA and proteins and is often mutated in tumor cells.[83−86] A subset of these mutations is known to lower the melting temperature of the protein, resulting in unfolding and aggregation.[87−90] The mechanisms that underpin the stability of p53 have been extensively studied; however, a coherent picture, especially one that integrates the role of hydration, still remains incomplete. Experiments report that ligand (DNA) binding and dehydration in p53 should stabilize the native folded state, although the underlying molecular mechanisms remain unknown.[91] The hydration shell around the DBD, defined as being composed of "tight" and "dynamical" water molecules, has been hypothesized to make the DBD structure more labile.[87] However, there have been no reports of the role of buried water on the stability of the DBD, and we explore this in WT and the V143A mutants of p53; the latter is found with high frequency in tumors and is known to induce destabilization and unfolding of p53 under physiological conditions.[89]

Most of the buried water clusters in the crystal structures of the DBD are present near loops L2 and L3 and strand S7 and have low B-values, thereby suggesting that they are tightly bound (Figure 3A). The MD simulations of DBD carried out here suggest that the positions of buried water clusters in the simulations of WT and V143A DBD are quite different (Figure 3B). The positions of buried waters in the MD simulations of the WT DBD, but not the V143A mutant form of DBD, correspond well to the positions of the crystallographic buried waters in the WT. It appears that in the V143A mutant of DBD there are fewer buried water clusters around the loops L2 and L3 as compared to WT DBD (shown with dashed line in Figure 3B). Therefore, we investigate if the dehydration of the loops L2 and L3 implies loss of stability in DBD. The importance of buried waters for the stability of protein structures has been abundantly suggested in the literature. For instance, it has been suggested that buried waters fulfill the unsatisfied interaction valencies of the protein atoms through water mediated interaction networks.[92] Alternatively, hydration of a hydrophobic cavity in the protein core may prevent its collapse and help a protein retain its native structure.[93] Protein hydration has also been shown to be associated with increasing protein flexibility and hence entropic stabilization.[11,62]

A closer inspection of the positions of buried waters in the DBD reveals the possibility of differential conformational sampling of the loop L3 between the WT and V143A (Figure 4A and 4B). The loop L3 in V143A samples a non-native conformation in one of the three replicates. It appears that the crystal-like native conformation of loop L3 in the WT DBD is stabilized by a buried water molecule which is lost when the mutant DBD assumes a non-native conformation. This buried water makes interactions with the backbone of loop L2 (oxygen of Arg174) and loop L3 (nitrogen of Met246) as well as with the side chain of Cys238 (given that this is a sulfur, this interaction is likely to be weak), which is also coordinated to zinc (Figure 4C). The total residence time of a water molecule at this position (computed by taking the backbone nitrogen of Met246 as the reference atom) is 73.33 (±9.6) ns for the WT, while it is only 34 (±20.5) ns for V143A across the triplicate MD simulations, each

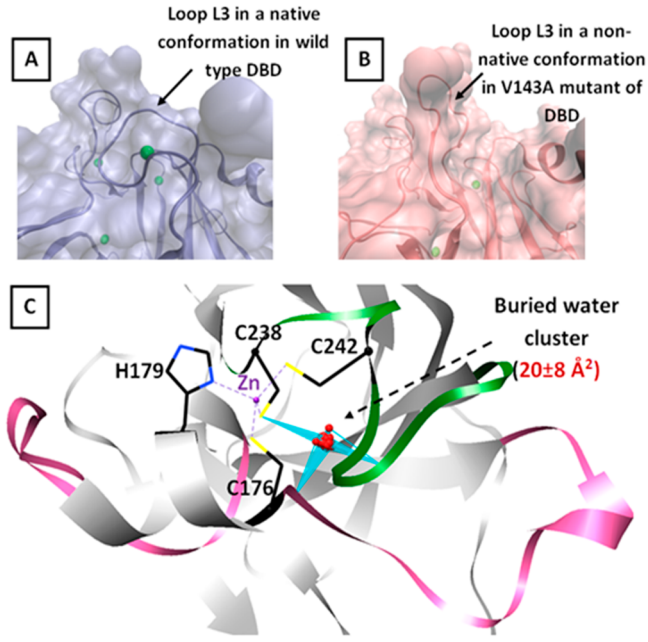


Figure 4. Buried waters (green spheres) in a snapshot from simulations of the DBD of (A) WT (PDB id 2AHI chain A) and (B) the V143A mutant (constructed in this study). It appears that the native conformation of loop L3 is stabilized by the buried water in WT, while this water is lost when the loop L3 assumes a non-native conformation in one of the three replicates of the V143A mutant. (C) Analysis of the crystal structures of all the human p53 DBD entries from the Protein Data Bank[37] confirms the presence of this water molecule (Table S1). A cluster of water molecules (red spheres) at this position is present in 100 out of a total of 119 chains of DBD that were superimposed (on to the crystal structure of WT human p53, PDB id 2XWR chain A which is shown). The average B-values of water molecules contributing to these clusters are shown in brackets. Lines in cyan represent hydrogen bonds. Loops L2 and L3 are shown in pink and green, respectively.

with a total sampling time of 100 ns. Water at this position is also conserved in the crystal structures. Out of the 119 chains (Table S1) of DBD that were superimposed, buried water at this position in loop L3 is present in 100 structures (Figure 4C). The average B-value of the waters contributing to this cluster is 20 ± 8 Å^2, implying that this water is indeed tightly bound to the protein.[94] Furthermore, we extended the simulations of the two replicates of V143A that did not show loop L3 in a non-native conformation, to 500 ns. We now see that the loop L3 indeed assumes a non-native conformation, making this observation consistent in all three replicates (Figure 5).

Next, in a control study of WT DBD, we remove all the crystal waters from the protein immersed in the box of waters and ran an MD simulation. A water molecule from the bulk solvent is observed to occupy this position near loop L3 almost instantaneously, clearly suggesting that this site prefers to be hydrated.

Figure S3 suggests that the three atoms, which interact with the buried water, are only engaged in interactions with waters and/or zinc and not with any protein atoms: Cys238(SG) is coordinated to the zinc ion and covalently bonded to it in the simulations; Arg174(O) can interact with another water molecule (apart from the buried water in loop L3); Met246(N) interacts exclusively with the buried water in loop L3. These suggest that it is unlikely that the non-native conformation of loop L3 in V143A is caused by the loss of protein−water and not

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

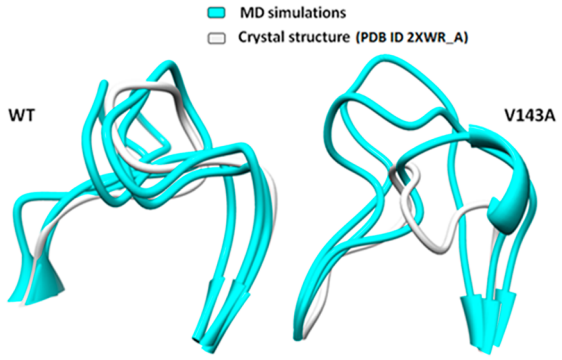


Figure 5. Conformation of loop L3 in the crystal structure (PDB id 2AHI chain A) and seen in the MD simulations of the DBD of WT type and the V143A mutant (constructed in this study). Last frames from the simulations are superimposed onto the crystal structure to highlight the differences.

of protein−protein interactions, although the role of long-range intraprotein interactions may play some role.

We further looked at the amino acid conservation of the three residues (whose atoms are engaged in making interactions with the buried water) across the homologues of p53 in the IARC database (http://p53.iarc.fr/SequenceAlignment.aspx)[95] (Figures S4A and S4B). We observe that these residues are conserved. Additionally, the spatial orientations of the backbone atoms of Arg174(N) and Met246(O), that are hydrogen bonded to the buried water in loop L3, are strictly conserved in the 119 crystal structures of human DBD studied (Figure S5). Together these make a compelling case for the importance of these interactions in maintaining the conformation of loop L3 which appears to be destabilized in V143A with the loss of the water. While a causal relationship between the loss of buried water and the native conformation of loop L3 is not easy to establish unambiguously in such complex systems, nevertheless the strong correlation observed between the presence of buried water in loop L3 and the conformation of loop L3 is compelling and underscores the importance of this buried water for the stability of the DBD.

Given the fact that most of the commonly occurring cancer mutations (referred to as "hot spot" mutations) are located in loops L2 and L3,[84,96,97] the change in the dynamics of these regions resulting from the loss of the buried water molecule could have implications in tumorigenesis. The increased flexibility also correlates well with the available experimental data on the NMR of the destabilizing mutants including, V143A, which suggests that residues in the loop L3 are affected between the WT and the destabilizing mutants.[90] A limitation of the current model, where zinc is coordinated irreversibly to loops L2 and L3 through C176, H179, C238, and C242, is whether the water is related to zinc binding (Figure 4C). Zinc binding is known to stabilize the DBD, and the loss of zinc is reported to be a major factor in the destabilization of DBD.[98−100]

To further probe the importance of this buried water, we carried out double decoupling free energy calculations. Table 1 shows the ΔG_{abs} values for the water molecule bound to loop L3 when each λ window is simulated for 10 ps, 100 ps, and 1 ns. The ΔG_{abs} is highly consistent across the different lengths each λ window is simulated for; however, the standard deviation decreases with an increase in the sampling, implying good convergence. Regardless, our findings suggest that for such studies, a 10 ps window length should technically suffice. A

Table 1. Free Energy of Binding of the Water Molecule in Loop L3 of DBD Computed Using the Double Decoupling Method[60]

	ΔG_{abs} (kcal/mol)		
each λ window	10 ps	100 ps	1 ns
MBAR	−7.8 ± 0.17	−7.8 ± 0.05	−7.8 ± 0.02

ΔG_{abs} value of −7.8 kcal/mol (compared to the free energy of hydration of the water of −6.1 kcal/mol) indicates that the water molecule prefers to be bound to the loop in the WT.

To further investigate the MD trajectory at λ = 0 (buried water molecule is present with full electrostatics and van der Waals potential) and λ = 1 (buried water molecule has no electrostatic or van der Waals potential) we ran an additional simulation (production run = 15 ns to compare with the 10 ns TI (Thermodynamic Integration) calculations at these two points of λ by selectively restraining regions of DBD (as shown in Figure 6A). We removed the restraints in the loop regions around the buried water in loop L3 (while restraining other regions) so as to investigate the influence of this water molecule on the dynamics of these loop regions. We examine the RMS fluctuations of the backbone atoms between the simulations of WT DBD at λ = 0 and λ = 1. Simulation of the DBD with the interaction potential of the buried water molecule in L3 switched off exhibits higher fluctuations, especially in loop L3, as compared to the simulation where the interaction potential of the buried water molecule in L3 is switched on (Figure 6B). This clearly suggests that the buried water in loop L3 has a significant impact on the structure, interactions, and dynamics of the neighboring regions and that this buried water stabilizes loop L3 of DBD through van der Waals and electrostatic interactions.

Interestingly the G245S mutation is also known to destabilize the DBD, and a longer serine side chain is thought to displace this structural water molecule.[89] To test this hypothesis, we simulated the G245S mutant of p53 DBD (∼1.5 μs), without removing the water. In our simulations (Figure 7), the buried water in loop L3 is only partially displaced, and loop L3 adopts a non-native conformation. This further strengthens our hypothesis that this water molecule is indeed important for maintaining the WT-like conformation of loop L3 and may well be implicated in several DBD mutations.

Case Study 2: Residence Time of Waters at the eIF4E-mRNA Cap Binding Interface. Identifying multiwater bridges (water bridges involving multiple persistent/high residence time waters) in MD simulations remains a yet unsolved problem. Presently, it has been possible to track only single water molecules that are bridging different regions of the protein or bridging two different molecules. This is achieved by tracking the hydrogen bonds formed between the water molecule and the solute atoms. However, the tracking of multiple waters does not appear to be straightforward. Multiwater bridges are often seen in clusters of conserved water molecules, linking the solutes at the protein−protein/ligand interfaces, and are thought to have critical roles in molecular recognition and interaction.[64,101] Here, we demonstrate that probing the residence times of water molecules[80,81] using the atom centric approach (see Methods section) can help identify multiwater bridges in proteins.

We examine the residence times of waters in eIF4E, a translation initiation factor which binds to the 5′ mRNA cap (m7GTP) and triggers a molecular signaling cascade that culminates in translation. The mRNA cap binding interface of eIF4E in the cap-unbound state is very well hydrated.[74,102]

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

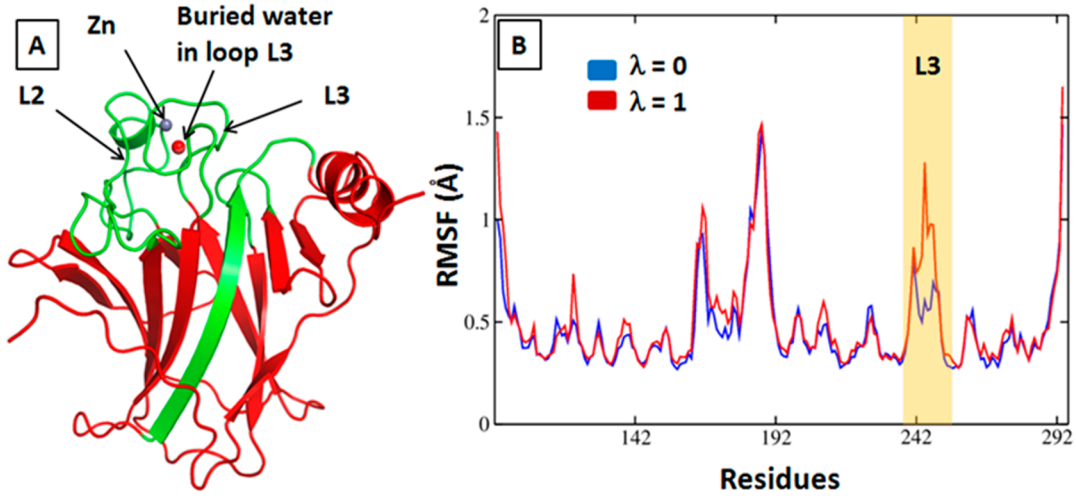


Figure 6. WT p53 DBD (PDB id 2AHI chain A) is simulated with regions colored in red (in (A)) having their Cα atoms restrained. (B) RMSF of residues from simulations of DBD at different λ points. $\lambda = 0$ and $\lambda = 1$ indicates that electrostatic and van der Waals interactions of the buried water molecule are switched on and off, respectively. Highlighted region in the plot shows differences in fluctuations in residues at loop L3.

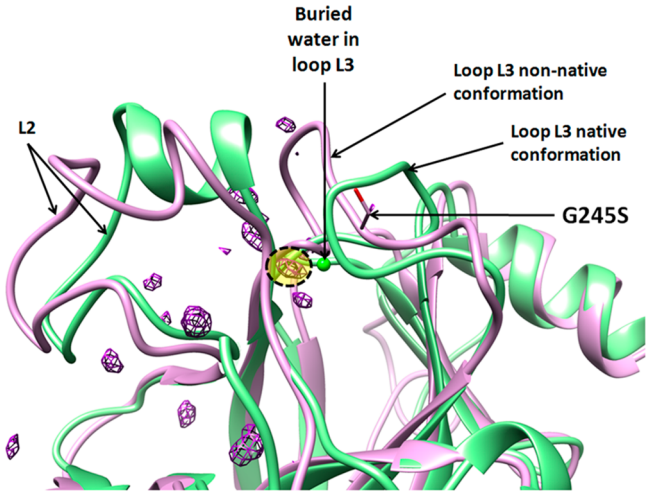


Figure 7. Structures of p53 DBD from crystal structure (PDB id 2AHI chain A) and from G245S mutant (constructed in this study) MD simulations are superimposed and shown in green and purple cartoon representations, respectively. Buried water in loop L3 from the crystal structure is shown as a green sphere. Solvent density from the MD simulations, contoured at 3 times the bulk solvent density, is shown as purple mesh. Highlighted region represents the solvent density of buried water in loop L3 from the MD simulations.

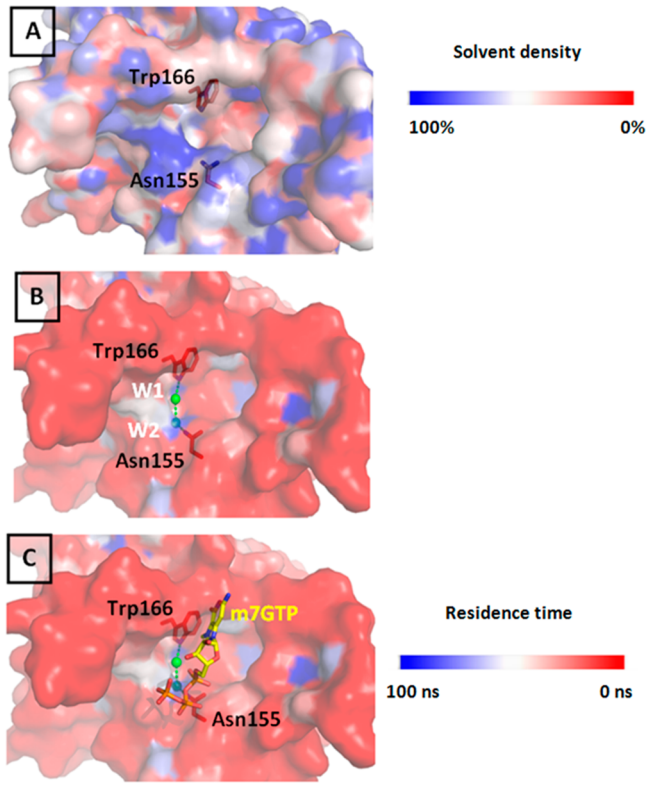


Figure 8. mRNA cap binding site of eIF4e taken from PDB id 2W97 chain A is colored according to (A) hydration and (B, C) residence times of water molecules with a color gradient from blue to white to red regions representing high to low values. W1 and W2 (green spheres) positions show markedly high residence times (distinct blue patches on the surface). Residues shown in sticks are Trp166 and Asn155 that form two ends of the water bridge. (C) mRNA cap m7GTP is shown for perspective.

Recently, a network of 14 invariant interfacial water molecules was identified as mediating interactions between eIF4E and mRNA cap.[64] A water bridge involving two solvent molecules lies at the core of this network. This water bridge engages in hydrogen bonding interactions with Trp166 and Asn155 and has been hypothesized to be vital for mRNA cap binding in eIF4E.[64]

We compute the solvent density and the residence times of waters around eIF4E by tracking water interactions with all the respective non-hydrogen protein atoms within a sphere of radius 3.5 Å. Figures 8A and 8B show solvent density and residence time values mapped onto the protein. The solvent density clearly shows that the mRNA cap binding site of eIF4E is indeed well hydrated. However, it is apparent that not many sites in the binding pocket have high residence times. Thus, although the mRNA binding site of eIF4E is well solvated, it appears that most of the solvent molecules are rapidly exchanging with bulk solvent. However, we see two sites in the binding pocket with extremely high residence times (seen as blue patches on the surface). The two sites with high residence times correspond to the waters involved in forming the water bridge (labeled as W1 and W2 in Figure 8B). This water bridge motif has been reported

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

to be central to the mRNA cap recognition of eIF4E (Figure 8C) and so may play a crucial role in the kinetics and thermodynamics of the process.[64] This example demonstrates how mapping the residence times of waters on the protein surface can help identify water bridges involving multiple waters in proteins. This nicely complements modules such as the "hbond" function in the cpptraj module of AmberTools15[57] which identifies water bridges mediated by a single water molecule.

DISCUSSION

JAL Complements Existing Approaches To Compute Hydration Properties. JAL analyzes all atom explicit solvent MD simulations to detect waters that are deemed to be buried based on the criteria of their accessibility to an external probe. Subsequently it analyzes several properties associated with these buried waters. Together with two unique utilities, namely the detection of buried waters and water bridges, it complements existing approaches to compute hydration properties.

Grid Based vs Protein (Solute) Atom Centric Approach. Hydration around a protein is a function of the chemistry and topology of the protein surface,[14−16] thus making analysis challenging, most notably because of the small size of a water molecule enabling it to engage in multiple interactions that can rapidly interconvert even on the time scales of protein motions.[14] To approach this issue, we use an atom centric approach (that is central to all JAL programs) which maps hydration metrics (solvent density or residence times) onto protein atoms, thereby linking hydration to specific sites (atoms) of the proteins. In contrast, representing solvent density as grid boxes in 3D (as represented by the "grid" function in Amber's ptraj module for instance) obscures the identity of the solute atoms whose properties determine the particular solvent densities. Additionally, the grid based approach is extremely sensitive to solute dynamics and therefore affects the accuracies of the solvent properties computed in the 3D grids in cases where the solute exhibits large amplitude backbone fluctuations and/or large scale conformational changes as is often the case for intrinsically disordered proteins (IDPs). The grid-based method (for instance the hydration program *grid* in the ptraj module of Amber) provides a hydration measure (e.g., solvent density) that is properly normalized by voxel volume and has a clear interpretation, enabling it to be transferable between different locations and different systems. In a similar manner, the hydration measure computed using the atom centric approach is also transferrable, with the constraint that the same (color) spectrum scale, frame interval, and number of frames in a trajectory are used when comparing hydration properties across two states of the molecule (Figure S1) or two different molecules.

Buried Waters and Multiwater Bridges. While JAL provides utilities that are also provided by other programs, it additionally offers the following two novel features:

(1) Computing Buried Waters. JAL complements existing programs (outlined in the Introduction) as a rapid and computationally relatively inexpensive method to detect buried waters in MD simulation trajectories. It computes the exposure (solvent accessible surface area or SASA) of waters to a probe followed by an iterative "evaporation" of exposed waters until convergence is reached (i.e., only those waters remain whose SASA is 0 Å^2). We compare the ability of JAL to find buried waters with the recent method of Setny[45] (Table S4). This data is further detailed in Table S5 which shows the fraction of buried crystallographic waters predicted by JAL compared to the other recent method.[45] The results demonstrate that both JAL and the other method are able to identify most of the crystallographic buried waters (Figure S6).[45]

It is clear from Table S4 that water molecules that make three or four hydrogen bonds are largely well-defined in the crystal structures, have low B-values, and are likely to be buried based on the SASA metric, thus making their prediction by either method straightforward. Only two water molecules that satisfied these criteria were not detected by either method. These were Wat 2094 in the structure with PDB id 1W8V and Wat 170 in the structure with PDB id 3Q7Y (Table S4). These water molecules have low B-values (14.62 Å^2 and 21.73 Å^2) and engage in three and four hydrogen bonds, respectively. Examination of the structural and chemical environments of these water molecules (Figure S8) suggests that both water molecules are very close to the protein surface and are soon lost to bulk solvent. Given their low B-values, it is possible that other water molecules will rehydrate the site but at time scales longer than those simulated; this is an example of the limitation of such methods. However, Wat 165 with a B-value of 20.73 Å^2 in 3Q7Y is predicted by JAL. A look at the structure shows that Wat 165 and Wat 170 are both buried in the crystal structure and are spatially adjacent to each other, making a hydrogen bond with each other and three hydrogen bonds with the protein. During the simulation, Wat 170, which is closer to the surface escapes, resulting in a conformational change whereby the two loop regions which were hydrogen bonded to Wat 170 now "collapse" over Wat 165 to compensate for the lost Wat 165−Wat 170 hydrogen bond, thus caging Wat 165 and preventing its escape into solvent.

We next look at water molecules which exhibit large B-values (more than 30 Å^2). Three water molecules belong to this category: Wat 224 (B-value = 30.21 Å^2) and Wat 238 (B-value = 45.85 Å^2) in PDB id 1BAS and Wat140 (B-value = 69.41 Å^2) in PDB id 2YGS. While the former two water molecules engage in four hydrogen bonds and are predicted using JAL (Wat 238 was only predicted by JAL), Wat 140 is clearly weakly held at the surface with just one hydrogen bond, is lost to bulk solvent in the simulation, and hence unsurprisingly, is not predicted by either program. Wat 224 and Wat 238 have high B-values and yet remain hydrogen bonded to the protein atoms (whose B-values are in the range of 12−23 Å^2), and both waters are caged. Both programs are able to predict waters that make two hydrogen bonds with the protein (Wat 2046 and Wat 2122 in PDB id 1W8V, Wat 304 and Wat 307 in PDB id 3HVV) with B-values ranging between 6.96 and 15.44 Å^2. However, waters such as Wat 150 in PDB id 1SNB (with a B-value of 23.15 Å^2) make one hydrogen bond with the protein and one with loosely bound water and are not predicted because it escapes into bulk solvent during the simulation. The examples studied in Table S4 refer to waters that engage in hydrogen bonds with the protein or with the protein and other water molecules. The buried waters that are difficult to predict by either method (Figure S7) are the ones near the protein surface and prone to escape into bulk solvent; they may exchange but on time scales much larger than those simulated. Another limitation of these SASA based methods is in the detection of "buried" water in cavities that are large since these cavities may contain waters and yet may accommodate a rolling probe which would flag those waters as solvent accessible; the likelihood of hydration of large cavities in proteins has been a matter of some debate;[103−106] a solution to this problem may lie in using probes with larger radii.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

(2) Computing Multiwater Bridges. JAL computes intra- or intermolecular multiwater bridges around a protein by calculating the residence times of water molecules at sites in proteins (non-hydrogen atoms) followed by linking the sites within hydrogen bonding distance which have persistent waters (high residence times). One of the existing programs, AmberTools, enables the calculation of water-bridges (based on hydrogen bonds); however it can only identify single-water mediated water bridges, while JAL can detect multiple water mediated bridges and has been used to identify and speculate on some functional water networks in eIF4E.[64]

Other Functionalities of JAL. Most MD packages enable the computation of solvent density. For instance, AmberTools uses a grid based approach to compute solvent density. Similarly, trjorder, h2order, g(x), and sasa are some of the programs in GROMACS that can potentially be used for hydration analysis but require scripting to extract information such as the residence times of water molecules, indirectly from hydrogen bond analysis; in contrast, JAL produces these properties without the need for any scripting; to the best of our knowledge JAL is unique in this ability, along with the ability to compute buried waters and multiwater bridges around proteins.

Performance Assessment. Table S2 shows the computational cost associated with the hydration analysis and how it varies with the system size. Several measures were taken to optimize the speed of the calculations. For instance, a preprocessing step is introduced to reduce the size of the system. It trims the hydrogens and all water molecules beyond the first hydration shell. Tracking water molecules for solvent density/residence time involves measuring distances between solute atoms and all solvent atoms which can be extremely time-consuming. We partly overcome this problem by sorting the solvent molecules to be searched for according to the X, Y, and Z coordinates and then searching only for waters that are close to the solute atoms. We are implementing further optimization of the programs, and the updates will be made available at http://mspc.bii.a-star.edu.sg/minhn/jal.html. The program to identify buried waters in an MD trajectory was quite efficient and largely independent of the protein size, requiring a computing time of ∼1 s per snapshot. Hardware and memory requirements for all the JAL programs are quite minimal (the 1 s benchmark was for a 1.5 GHz CPU requiring less than 100 Mb memory).

Potential Limitations. A parameter central to the atom centric approach is the distance cut off used for scanning a water oxygen atom from a reference solute heavy atom. We explored several different cut offs including 3, 3.5, and 4 Å and found 3.5 Å to be optimal; it is also the maximum hydrogen bond donor to hydrogen bond acceptor distance for a hydrogen bond (of average strength). A limitation of JAL is that it cannot report easily on the origin of the solvent densities (expressed as percent solvation). The solvent densities around solute atoms could arise from (1) the ability of solute atom to interact strongly/weakly with water oxygen atoms and (2) the high or low solvent accessibility of solute atoms. This is a problem also faced by the grid based methods and is a general problem because of the complex relationship between the relatively small size of the water molecule and the complex shape and chemistry of the protein.

We would also like to highlight that the precise location of the densities of individual water molecules is not directly accessible in the atom centric approach presented here. While this is a limitation of the atom centric approach, there are grid based methods already available that allow one to visualize the precise location of waters (as has already been suggested in the Methods section). The caveat is that the precise location of waters around a solute is a function of the topology and chemistry of the solute surface[14−16] which is obscured in the grid based methods. Thus, JAL complements the grid based methods, in that, the grid based methods allow one to visualize the precise location of waters while JAL enables a detailed understanding of the physicochemical nature of the associated solute atoms. The precise location of waters around a solute can also be found using JAL; however, it does require some additional work that has not been automated here as it requires graphical visualization: for example, in Figure 8, the solute is shown in surface representation and colored by the solvation measure using the output of JAL, and the blue patches show the exact location of water molecules.

It can also be argued that Radial Distribution Function (RDF) provides the same information as solvent density and distance of closest measures combined in the "JAL" package. The RDF can provide detailed information but only for a single site on the protein surface. Visualization of the nature and extent of hydration of the entire ligand binding region, for instance, can conveniently be carried out using JAL (as in Figure 8), but several RDFs ('*n*' numbers of plots with '*n*' being the number of sites/atoms on proteins) will need to be combined and processed to produce an equivalent image.

Current MD simulations are unable to sample the long time-scale water exchange processes, and hence these waters will not be flagged as buried using the methods described, although a recent method[67] has overcome some of these limitations. Finally, methods that rely on SASA as a metric are of little utility in cases where the size of the cavity that contains a water molecule is large, as the SASA calculation will result in the incorrect identification of the buried water as exposed to solvent.

A summary of comparison of JAL against existing approaches to probe biomolecular hydration is presented in the Supporting Information (Table S3).

■ CONCLUSION

With the surge in computational power, MD simulations are increasingly being used to simulate biomolecular processes.[27] Subsequently, powerful analytical programs are required to process the trajectory data. Motivated by the scarcity of programs that allow efficient computation of hydration properties, particularly residence times and buried waters in an MD trajectory, we have developed a software suite of programs ("JAL") that enable the computation of hydration properties of a solute from its conformational ensemble. The command line programs are interactive and take, as input from user, an MD trajectory, in addition to the optional parameters required for the calculation. More emphasis is given on making the program user-friendly and the computations fast and accurate. We have benchmarked the calculation of several hydration properties against available experimental data. For instance, we have benchmarked the buried water program within JAL against the known buried waters in the protein BPTI. Additionally, we have compared the buried water computation in JAL with a recent buried water prediction method which is based on semiexplicit solvent simulations[45] and demonstrate that the results of JAL are quite comparable. A major limitation of all such methods is the general lack of sampling of water exchange processes in MD simulations, although a recent method[67] has made a breakthrough, and also in the use of SASA of waters in large cavities which will erroneously yield large SASA values and flag the waters as exposed.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

We demonstrate the utility of "JAL" using as examples p53 and eIF4E, which also resulted in the suggestion of useful biological insights. For instance, using the example of eIF4E−mRNA cap binding interface, we demonstrate a simple technique where mapping the residence times of waters on the protein atoms can be used to map water bridges involving multiple water molecules with high residence times that bridge sites across proteins. A systematic study of buried water clusters, performed using JAL, in MD simulations of WT and a V143A mutant form of DBD provides interesting insights on the role of water in structural stabilization. We identify one highly persistent (high residence time) buried water molecule that coordinates with the loops L2 and L3 in the WT DBD but is lost in the V143A oncogenic mutant form as loop L3 assumes a non-native conformation. Loss of this buried water in the mutant could destabilize p53. We establish the significance of this buried water to the structural viability in p53 DBD using double decoupling free energy calculations which predict a very favorable binding energy for this water in loop L3.

Central to JAL is the atom centric approach used to compute hydration properties. Though computationally more expensive than the grid-based approaches, the atom centric approach provides an understanding of which regions (or atoms) on the protein govern the nature and extent of solvation observed around it.

Together with the closest distance measure, the hydration properties computed using JAL allow detailed characterization of the nature of hydration sites. The calculated information can also be used to investigate most persistent hydrogen bonds between solvent and protein. Textual output is available in a format that can be easily analyzed visually (through plots or in 3D through PyMOL) as well as a detailed summary, allowing easy and detailed analysis of the hydration properties. "JAL" will complement existing hydration analysis tools in providing insights into the role of water in biology. Our future effort will be focused on making the programs faster, modularizing the JAL programs and adding more functionalities to the JAL package.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00453.

Table S1, crystal structures of p53 DBD; Table S2, computational cost associated with hydration calculations using JAL; Table S3, existing approaches to compute hydration properties; Table S4, comparison of JAL and Setny methods in abilities to predict buried waters; Table S5, fraction of crystallographic buried waters identified using Setny and JAL methods; Figure S1, color spectrum scale to map hydration properties on solute; Figure S2, JAL module to identify buried waters; Figure S3, interactions of atoms hydrogen bonded to buried water cluster loop L3 of DBD of human p53; Figure S4, multiple sequence alignment of homologues of p53 in IARC database; Figure S5, spatial orientation of Arg174 and Met246; Figure S6, buried waters as computed using JAL and Setny methods; Figure S7, buried waters not predicted using either Setny or JAL methods (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (65) 6478 8273. Fax: (65) 6478 9048. E-mail: chandra@bii.a-star.edu.sg. Corresponding author address: Bioinformatics Institute, A*STAR (Agency for Science, Technology and Research), 30 Biopolis Street, #07-01 Matrix, Singapore 138671.

ORCID

Mohan R. Pradhan: 0000-0001-9759-1676
Minh N. Nguyen: 0000-0003-4949-2571
Srinivasaraghavan Kannan: 0000-0002-9539-5249
Chandra S. Verma: 0000-0003-0733-9798

Funding

M.R.P. was funded by the Research Scholarship Award by Bioinformatics Institute-NTU/SCE Joint Ph.D. Program. M.N.N. would like to thank A*STAR Joint Council Office (JCO) Career Development Award [15302FG145] for support. S.K. is partly funded by IAF-PP grant H18/01/a0/015 (A*STAR/NRF/EDB); S.J.F. is funded by IAF-PP grant H17/01/a0/0W9 (A*STAR/NRF/EDB). Support by the A*STAR Biomedical Research Council (BMRC) and computing facility by A*STAR Computing Resource Centre (A*CRC), Agency for Science Technology and Research (A*STAR), Singapore, and NSCC, Singapore is gratefully acknowledged.

Notes

The authors declare the following competing financial interest(s): C.S.V. and S.K. are the founder directors of Sinopsee Therapeutics, a biotech company developing molecules for therapeutic purposes; the current work has no conflict with the company.

■ ACKNOWLEDGMENTS

We thank Dr Tom Oldfield, Dr Leo Caves, Dr Mike Hartshorn, and Dr Paul Elmsley for late night discussions in the late 1980s at the now named Structural Biology Labs, University of York, UK, on the "evaporation" of waters.

■ REFERENCES

(1) Thirumalai, D.; Reddy, G.; Straub, J. E. Role of water in protein aggregation and amyloid polymorphism. *Acc. Chem. Res.* **2012**, *45* (1), 83−92.

(2) Fernandez, A. Functionality of wrapping defects in soluble proteins: what cannot be kept dry must be conserved. *J. Mol. Biol.* **2004**, *337* (2), 477−83.

(3) Fernandez, A.; Berry, R. S. Extent of hydrogen-bond protection in folded proteins: a constraint on packing architectures. *Biophys. J.* **2002**, *83* (5), 2475−81.

(4) Fernandez, A.; Crespo, A. Protein wrapping: a molecular marker for association, aggregation and drug design. *Chem. Soc. Rev.* **2008**, *37* (11), 2373−82.

(5) Fernandez, A.; Scheraga, H. A. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 113−8.

(6) Fernandez, A.; Scott, R. Dehydron: a structurally encoded signal for protein interaction. *Biophys. J.* **2003**, *85* (3), 1914−28.

(7) Selkoe, D. J. Folding proteins in fatal ways. *Nature* **2003**, *426* (6968), 900−4.

(8) Chaplin, M. Do we underestimate the importance of water in cell biology? *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (11), 861−6.

(9) Ball, P. Water as a biomolecule. *ChemPhysChem* **2008**, *9* (18), 2677−85.

(10) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108* (1), 74−108.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

(11) Csermely, P. Water and cellular folding processes. *Cell Mol. Biol. (Noisy-le-grand)* **2001**, *47* (5), 791−800.
(12) Levy, Y.; Onuchic, J. N. Water and proteins: a love-hate relationship. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (10), 3325−6.
(13) Chong, S. H.; Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **2015**, *48* (4), 956−65.
(14) Dahanayake, J. N.; Mitchell-Koch, K. R. Entropy connects water structure and dynamics in protein hydration layer. *Phys. Chem. Chem. Phys.* **2018**, *20* (21), 14765−14777.
(15) Hua, L.; Huang, X.; Liu, P.; Zhou, R.; Berne, B. J. Nanoscale dewetting transition in protein complex folding. *J. Phys. Chem. B* **2007**, *111* (30), 9069−9077.
(16) Fogarty, A. C.; Laage, D. Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation. *J. Phys. Chem. B* **2014**, *118* (28), 7715−7729.
(17) Ball, P. Portrait of a molecule. *Nature* **2003**, *421* (6921), 421−2.
(18) Thanki, N.; Thornton, J. M.; Goodfellow, J. M. Distributions of water around amino acid residues in proteins. *J. Mol. Biol.* **1988**, *202* (3), 637−57.
(19) Savage, H.; Wlodawer, A. Determination of water structure around biomolecules using X-ray and neutron diffraction methods. *Methods Enzymol.* **1986**, *127*, 162−83.
(20) Madhusudhan, M. S.; Vishveshwara, S. Deducing hydration sites of a protein from molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **2001**, *19* (1), 105−14.
(21) Hamelberg, D.; McCammon, A. Dealing with bound waters in a site: Do they leave or stay? In *Computational and structural approaches to drug discovery*; Stroud, R. M.; Finer-Moore, J. Eds.; 2008; pp 95−106, DOI: 10.1039/9781847557964-00095.
(22) Otting, G.; Liepinsh, E.; Wuthrich, K. Protein hydration in aqueous solution. *Science* **1991**, *254* (5034), 974−80.
(23) Zhong, D.; Pal, S. K.; Zewail, A. H. Biological water: A critique. *Chem. Phys. Lett.* **2011**, *503* (1), 1−11.
(24) Panman, M. R.; Bakker, B. H.; den Uyl, D.; Kay, E. R.; Leigh, D. A.; Buma, W. J.; Brouwer, A. M.; Geenevasen, J. A.; Woutersen, S. Water lubricates hydrogen-bonded molecular machines. *Nat. Chem.* **2013**, *5* (11), 929−34.
(25) Hunt, N. T.; Kattner, L.; Shanks, R. P.; Wynne, K. The dynamics of water-protein interaction studied by ultrafast optical Kerr-effect spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (11), 3168−72.
(26) Yang, D. S.; Zewail, A. H. Ordered water structure at hydrophobic graphite interfaces observed by 4D, ultrafast electron crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (11), 4122−6.
(27) Dodson, G. G.; Lane, D. P.; Verma, C. S. Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Rep.* **2008**, *9* (2), 144−50.
(28) Helms, V.; Wade, R. C. Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study. *Biophys. J.* **1995**, *69* (3), 810−24.
(29) Helms, V.; Wade, R. C. Hydration energy landscape of the active site cavity in cytochrome P450cam. *Proteins: Struct., Funct., Genet.* **1998**, *32* (3), 381−96.
(30) Wade, R. C.; Mazor, M. H.; McCammon, J. A.; Quiocho, F. A. A molecular dynamics study of thermodynamic and structural aspects of the hydration of cavities in proteins. *Biopolymers* **1991**, *31* (8), 919−31.
(31) Lazaridis, T.; Karplus, M. New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science* **1997**, *278* (5345), 1928−31.
(32) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849−57.
(33) Rashin, A. A.; Iofin, M.; Honig, B. Internal cavities and buried waters in globular proteins. *Biochemistry* **1986**, *25* (12), 3619−3625.
(34) Pitt, W. R.; Goodfellow, J. M. Modelling of solvent positions around polar groups in proteins. *Protein Eng., Des. Sel.* **1991**, *4* (5), 531−537.
(35) Zhang, L.; Hermans, J. Hydrophilicity of cavities in proteins. *Proteins: Struct., Funct., Genet.* **1996**, *24* (4), 433−438.
(36) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **1999**, *289* (4), 1093−108.
(37) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−42.
(38) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58* (Pt 3 Pt 1), 380−388.
(39) Schymkowitz, J. W.; Rousseau, F.; Martins, I. C.; Ferkinghoff-Borg, J.; Stricher, F.; Serrano, L. Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (29), 10147−52.
(40) Imai, T.; Hiraoka, R.; Kovalenko, A.; Hirata, F. Locating missing water molecules in protein cavities by the three-dimensional reference interaction site model theory of molecular solvation. *Proteins: Struct., Funct., Genet.* **2007**, *66* (4), 804−813.
(41) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337−46.
(42) Virtanen, J. J.; Makowski, L.; Sosnick, T. R.; Freed, K. F. Modeling the hydration layer around proteins: HyPred. *Biophys. J.* **2010**, *99* (5), 1611−9.
(43) Setny, P.; Zacharias, M. Hydration in discrete water. A mean field, cellular automata based approach to calculating hydration free energies. *J. Phys. Chem. B* **2010**, *114* (26), 8667−75.
(44) Jeszenoi, N.; Horvath, I.; Balint, M.; van der Spoel, D.; Hetenyi, C. Mobility-based prediction of hydration structures of protein surfaces. *Bioinformatics* **2015**, *31* (12), 1959−65.
(45) Setny, P. Prediction of water binding to protein hydration sites with discrete, semi-explicit solvent model. *J. Chem. Theory Comput.* **2015**, *11* (12), 5961−5972.
(46) Morozenko, A.; Stuchebrukhov, A. A. Dowser++, a new method of hydrating protein structures. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1347−1357.
(47) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455−461.
(48) Ross, G. A.; Morris, G. M.; Biggin, P. C. Rapid and accurate prediction and scoring of water molecules in protein binding sites. *PLoS One* **2012**, *7* (3), No. e32036.
(49) López, E. D.; Arcon, J. P.; Gauto, D. F.; Petruk, A. A.; Modenutti, C. P.; Dumas, V. G.; Marti, M. A.; Turjanski, A. G. WATCLUST: a tool for improving the design of drugs based on protein-water interactions. *Bioinformatics* **2015**, *31* (22), 3697−36999.
(50) Hu, B.; Lill, M. A. WATsite: hydration site prediction program with PyMOL interface. *J. Comput. Chem.* **2014**, *35* (16), 1255−60.
(51) Patel, H.; Gruning, B. A.; Gunther, S.; Merfort, I. PyWATER: a PyMOL plug-in to find conserved water molecules in proteins by clustering. *Bioinformatics* **2014**, *30* (20), 2978−80.
(52) Schrodinger, LLC, *The PyMOL Molecular Graphics System*, Version 1.4r1; 2010.
(53) Sumathi, K.; Ananthalakshmi, P.; Roshan, M. N.; Sekar, K. 3dSS: 3D structural superposition. *Nucleic Acids Res.* **2006**, *34* (Web Server), W128−W132.
(54) Praveen, S.; Ramesh, J.; Sivasankari, P.; Sowmiya, G.; Sekar, K. WAP (version 2.0): an updated computing and visualization server for water molecules. *J. Appl. Crystallogr.* **2008**, *41* (5), 952−954.
(55) Balamurugan, B.; Roshan, M. N. A., Md.; Shaahul Hameed, B.; Sumathi, K.; Senthilkumar, R.; Udayakumar, A.; Venkatesh Babu, K. H.; Kalaivani, M.; Sowmiya, G.; Sivasankari, P.; Saravanan, S.; Vasuki Ranjani, C.; Gopalakrishnan, K.; Selvakumar, K. N.; Jaikumar, M.; Brindha, T.; Michael, D.; Sekar, K. PSAP: protein structure analysis package. *J. Appl. Crystallogr.* **2007**, *40* (40), 773−777.
(56) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.;

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545−614.
(57) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 2015*; University of California: San Francisco, 2015.
(58) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845−54.
(59) Woo, H. J.; Dinner, A. R.; Roux, B. Grand canonical Monte Carlo simulations of water in protein environments. *J. Chem. Phys.* **2004**, *121* (13), 6392−6400.
(60) Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *J. Am. Chem. Soc.* **2004**, *126* (24), 7683−7689.
(61) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33* (12), 889−897.
(62) Fischer, S.; Verma, C. S. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (17), 9613−5.
(63) Takano, K.; Yamagata, Y.; Yutani, K. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng., Des. Sel.* **2003**, *16* (1), 5−9.
(64) Lama, D.; Pradhan, M. R.; Brown, C. J.; Eapen, R. S.; Joseph, T. L.; Kwoh, C. K.; Lane, D. P.; Verma, C. S. Water Bridge Mediates Recognition of mRNA Cap in eIF4E. *Structure* **2017**, *25* (1), 188−194.
(65) Barillari, C.; Taylor, J.; Viner, R.; Essex, J. W. Classification of water molecules in protein binding sites. *J. Am. Chem. Soc.* **2007**, *129* (9), 2577−87.
(66) Jung, S. W.; Kim, M.; Ramsey, S.; Kurtzman, T.; Cho, A. E. Water Pharmacophore: Designing Ligands using Molecular Dynamics Simulations with Water. *Sci. Rep.* **2018**, *8*, 10400.
(67) Ben-Shalom, I. Y.; Lin, C.; Kurtzman, T.; Walker, R. C.; Gilson, M. K. Simulating Water Exchange to Buried Binding Sites. *J. Chem. Theory Comput.* **2019**, *15* (4), 2684−2691.
(68) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109* (8), 1528−32.
(69) Baker, E.; Hubbard, R. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44* (2), 97−179.
(70) De Simone, A.; Dodson, G. G.; Verma, C. S.; Zagari, A.; Fraternali, F. Prion and water: tight and dynamical hydration sites have a key role in structural stability. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (21), 7535−40.
(71) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes with Inhibitors. *Acta Crystallogr., Sect. B: Struct. Sci.* **1983**, *39* (4), 480−490.
(72) Shrake, A.; Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **1973**, *79* (2), 351−71.
(73) Kitayner, M.; Rozenberg, H.; Kessler, N.; Rabinovich, D.; Shaulov, L.; Haran, T. E.; Shakked, Z. Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* **2006**, *22* (6), 741−53.
(74) Brown, C. J.; Verma, C. S.; Walkinshaw, M. D.; Lane, D. P. Crystallization of eIF4E complexed with eIF4GI peptide and glycerol reveals distinct structural differences around the cap-binding site. *Cell Cycle* **2009**, *8* (12), 1905−11.
(75) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65* (3), 712−25.
(76) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M., Jr. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935−2947.
(77) Bullock, A. N.; Henckel, J.; Fersht, A. R. Quantitative analysis of residual folding and DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer therapy. *Oncogene* **2000**, *19* (10), 1245−56.
(78) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926−935.
(79) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542−1555.
(80) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32* (5), 523−35.
(81) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327−341.
(82) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016−2027.
(83) Brown, C. J.; Cheok, C. F.; Verma, C. S.; Lane, D. P. Reactivation of p53: from peptides to small molecules. *Trends Pharmacol. Sci.* **2011**, *32* (1), 53−62.
(84) Brown, C. J.; Lain, S.; Verma, C. S.; Fersht, A. R.; Lane, D. P. Awakening guardian angels: drugging the p53 pathway. *Nat. Rev. Cancer* **2009**, *9* (12), 862−73.
(85) Schuijer, M.; Berns, E. M. TP53 and ovarian cancer. *Hum. Mutat.* **2003**, *21* (3), 285−91.
(86) Vogelstein, B.; Lane, D.; Levine, A. J. Surfing the p53 network. *Nature* **2000**, *408* (6810), 307−10.
(87) Xu, J.; Reumers, J.; Couceiro, J. R.; De Smet, F.; Gallardo, R.; Rudyak, S.; Cornelis, A.; Rozenski, J.; Zwolinska, A.; Marine, J. C.; Lambrechts, D.; Suh, Y. A.; Rousseau, F.; Schymkowitz, J. Gain of function of mutant p53 by coaggregation with multiple tumor suppressors. *Nat. Chem. Biol.* **2011**, *7* (5), 285−95.
(88) Boeckler, F. M.; Joerger, A. C.; Jaggi, G.; Rutherford, T. J.; Veprintsev, D. B.; Fersht, A. R. Targeted rescue of a destabilized mutant of p53 by an in silico screened drug. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (30), 10360−5.
(89) Joerger, A. C.; Ang, H. C.; Fersht, A. R. Structural basis for understanding oncogenic p53 mutations and designing rescue drugs. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (41), 15056−61.
(90) Wong, K. B.; DeDecker, B. S.; Freund, S. M.; Proctor, M. R.; Bycroft, M.; Fersht, A. R. Hot-spot mutants of p53 core domain evince characteristic local structural changes. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (15), 8438−42.
(91) Silva, J. L.; Vieira, T. C.; Gomes, M. P.; Bom, A. P.; Lima, L. M.; Freitas, M. S.; Ishimaru, D.; Cordeiro, Y.; Foguel, D. Ligand binding and hydration in protein misfolding: insights from studies of prion and p53 tumor suppressor proteins. *Acc. Chem. Res.* **2010**, *43* (2), 271−9.
(92) Park, S.; Saven, J. G. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins: Struct., Funct., Genet.* **2005**, *60* (3), 450−63.
(93) Das, P.; Kapoor, D.; Halloran, K. T.; Zhou, R.; Matthews, C. R. Interplay between drying and stability of a TIM barrel protein: a combined simulation-experimental study. *J. Am. Chem. Soc.* **2013**, *135* (5), 1882−90.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329

(94) Martz, E.; Decatur, W. Temperature value. In *Proteopedia*; ISPC at the Weizmann Institute of Science in Israel: 2014.
(95) Olivier, M.; Eeles, R.; Hollstein, M.; Khan, M. A.; Harris, C. C.; Hainaut, P. The IARC TP53 database: new online mutation analysis and recommendations to users. *Hum. Mutat.* **2002**, *19* (6), 607−14.
(96) Bullock, A. N.; Fersht, A. R. Rescuing the function of mutant p53. *Nat. Rev. Cancer* **2001**, *1* (1), 68−76.
(97) Joerger, A. C.; Fersht, A. R. Structural biology of the tumor suppressor p53. *Annu. Rev. Biochem.* **2008**, *77*, 557−82.
(98) Blanden, A. R.; Yu, X.; Wolfe, A. J.; Gilleran, J. A.; Augeri, D. J.; O'Dell, R. S.; Olson, E. C.; Kimball, S. D.; Emge, T. J.; Movileanu, L.; Carpizo, D. R.; Loh, S. N. Synthetic metallochaperone ZMC1 rescues mutant p53 conformation by transporting zinc into cells as an ionophore. *Mol. Pharmacol.* **2015**, *87* (5), 825−31.
(99) Meplan, C.; Richard, M. J.; Hainaut, P. Metalloregulation of the tumor suppressor protein p53: zinc mediates the renaturation of p53 after exposure to metal chelators in vitro and in intact cells. *Oncogene* **2000**, *19* (46), 5227−36.
(100) Yu, X.; Blanden, A. R.; Narayanan, S.; Jayakumar, L.; Lubin, D.; Augeri, D.; Kimball, S. D.; Loh, S. N.; Carpizo, D. R. Small molecule restoration of wildtype structure and function of mutant p53 using a novel zinc-metallochaperone based mechanism. *Oncotarget* **2014**, *5* (19), 8879−92.
(101) Nguyen, M. N.; Pradhan, M. R.; Verma, C.; Zhong, P. The interfacial character of antibody paratopes: analysis of antibody-antigen structures. *Bioinformatics* **2017**, *33* (19), 2971−2976.
(102) Brown, C. J.; McNae, I.; Fischer, P. M.; Walkinshaw, M. D. Crystallographic and mass spectrometric characterisation of eIF4E with N7-alkylated cap derivatives. *J. Mol. Biol.* **2007**, *372* (1), 7−15.
(103) Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. *Science* **1995**, *267* (5205), 1813−1817.
(104) Yu, B.; Blaber, M.; Gronenborn, A. M.; Clore, G. M.; Caspar, D. L. Disordered water within a hydrophobic protein cavity visualized by x-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (1), 103−108.
(105) Heidary, D. K.; Roy, M.; Daumy, G. O.; Cong, Y.; Jennings, P. A. Long-range coupling between separate docking sites in interleukin-1beta. *J. Mol. Biol.* **2005**, *353* (5), 1187−1198.
(106) Somani, S.; Chng, C. P.; Verma, C. S. Hydration of a hydrophobic cavity and its functional role: a simulation study of human interleukin-1beta. *Proteins: Struct., Funct., Genet.* **2007**, *67* (4), 868−885.

DOI: 10.1021/acs.jcim.8b00453
J. Chem. Inf. Model. 2019, 59, 3316−3329