JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2020, 60, 890−897

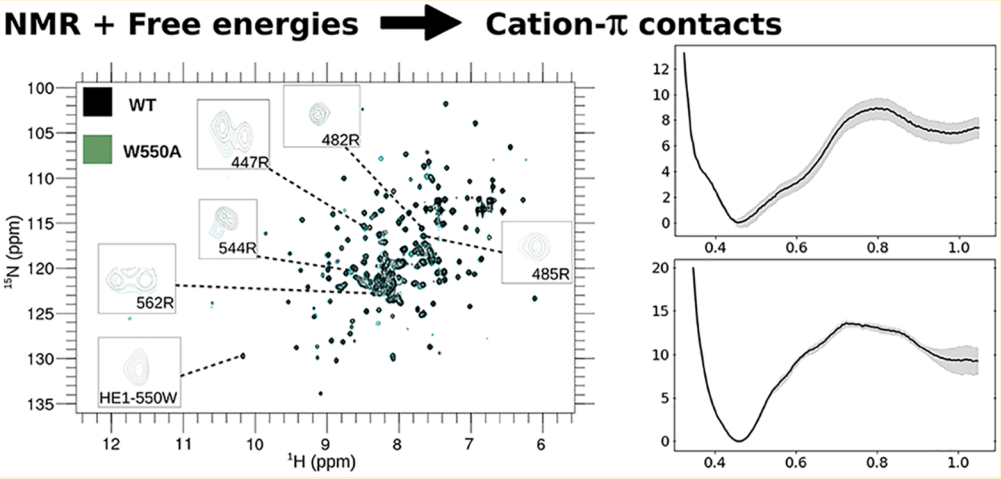
pubs.acs.org/jcim

Combining Free Energy Simulations and NMR Chemical-Shift Perturbation To Identify Transient Cation−π Contacts in Proteins

André A. O. Reis, Raphael S. R. Sayegh, Sandro R. Marana, and Guilherme M. Arantes*

Department of Biochemistry, Instituto de Química, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, 05508-900 São Paulo, São Paulo, Brazil

Supporting Information



ABSTRACT: Flexible protein regions containing cationic and aromatic side-chains exposed to solvent may form transient cation−π interactions with structural and functional roles. To evaluate their stability and identify important intramolecular cation−π contacts, a combination of free energy profiles estimated from umbrella sampling with molecular dynamics simulations and chemical shift perturbations (CSP) obtained from nuclear magnetic resonance (NMR) experiments is applied here to the complete catalytic domain of human phosphatase Cdc25B. This protein is a good model system for transient cation−π interactions as it contains only one Trp residue (W550) in the disordered C-terminal segment and a total of 17 Arg residues, many exposed to solvent. Eight putative Arg-Trp pairs were simulated here. Only R482 and R544 show bound profiles corresponding to important transient cation−π interactions, while the others have dissociative or almost flat profiles. These results are corroborated by CSP analysis of three Cdc25B point mutants (W550A, R482A, and R544A) disrupting cation−π contacts. The proposed validation of statistically representative molecular simulations by NMR spectroscopy could be applied to identify transient contacts of proteins in general but carefully, as NMR chemical shifts are sensitive to changes in both molecular contacts and conformational distributions.

1. INTRODUCTION

Proteins are stabilized by multiple noncovalent intramolecular contacts such as hydrogen bonds, hydrophobic contacts, and ion pairs. Cation−π interactions formed between cationic residues (in Arg and Lys) and aromatic rings (in Phe, Tyr and Trp) have comparable interaction energies mainly of electrostatic origin[1] and are easy to recognize when buried in folded proteins.[2] Cation−π contacts may also be observed between side-chains exposed to solvent due to the competition of aromatic rings with water in binding cations in aqueous environments.[1,3−5]

The lifetime of a noncovalent contact in aqueous solution is determined by its stability or free energy, and weak contacts will survive briefly due to structural fluctuations and water competition.[6] Flexible protein regions such as side-chains exposed to solvent, loops, and intrinsically disordered segments should have their structure represented by a conformational ensemble in which the stability of a possible noncovalent contact is proportional to its population or statistical importance in the configurational distribution.[7] However, ensembles commonly obtained from trajectories of molecular dynamics (MD) simulation often sample the underlying equilibrium distribution inappropriately or incompletely.[8] Not seldom, a single observation of a transient contact during

Special Issue: Molecular Simulation in Latin America: Coming of Age

Received: September 27, 2019
Published: November 18, 2019

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

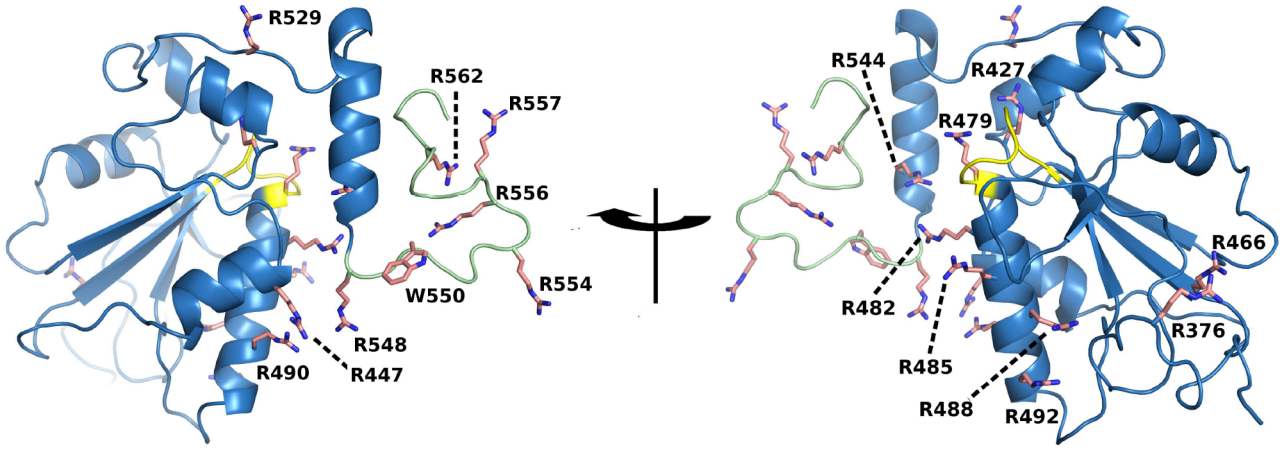


Figure 1. Structure of the complete catalytic domain of human phosphatase Cdc25B taken from a molecular dynamics snapshot and shown in cartoon with indicated Arg and Trp side-chains in pink sticks. The P-loop catalytic site is in yellow, and the disordered C-terminal is in pale green.

a simulation is taken as evidence of its structural or functional role[9−11] but without statistical significance.[12]

Here, the statistical importance of cation−π interactions formed between protein side-chains is investigated with a combination of molecular simulation and experimental spectroscopy. Enhanced sampling free energy simulations are employed. In principle, these may lead to correct contact distributions if a meaningful collective or reaction coordinate describing the transient interaction is chosen and orthogonal degrees of freedom are properly sampled.[8] But, given the approximate description of interaction energies used in biomolecular simulations and finite sampling available in practice, an experimental validation of putative transient contacts will help to confirm the computational predictions.

Nuclear magnetic resonance (NMR) is exquisitely sensitive to the nuclei chemical environment and is routinely used to determine intramolecular contacts of proteins in solution.[13] For instance, cation−π interactions will shift Arg and Lys side-chain resonances upfield because of diamagnetic anisotropic shielding from a nearby indole ring in Trp.[4] But NMR chemical shifts also report on the distribution of protein conformations. The nuclear Overhauser effect (NOE) is frequently used to discriminate shifts induced by molecular contact from those of conformational origin.[13] Here, only standard $^{1}H-^{15}N$ HSQC spectra were explored to avoid additional NMR experiments or extensive assignments of side-chain resonances,[14] so that the proposed experimental validation is kept as simple and general as possible. It should be noted that other spectroscopical methods such as visible absorbance and fluorescence have recently been used to identify cation−π interactions in solution[15] and could also be employed to validate simulations of transient contacts if enough spectral resolution is available.

The complete catalytic domain of human phosphatase Cdc25B was studied here as a model system.[16,17] This enzyme catalyzes the dephosphorylation of pTyr and pThr from other peptides[18−21] and has already been investigated by solution NMR.[22,23] It contains only one Trp residue (W550) located in the disordered C-terminus[23,24] as well as a total of 17 Arg (5 in the disordered C-terminus) and 13 Lys. Many of these Arg side-chains are exposed on the protein surface near the P-loop catalytic site and W550 (Figure 1). The large number of cationic residues in Cdc25B may be related to anionic substrate recognition and product release.[17] Yet, multiple Trp-Arg contacts were previously observed in a long MD simulation of Cdc25B.[23,25]

In the following section the simulation and experimental methods are described with special care to ensure the approximate force-field used in the protein simulations reproduces the cation−π energetics accurately. Then, free energy profiles estimated from umbrella sampling (US) simulations are presented for the formation of eight putative cation−π contacts between side-chains of Trp-Arg pairs in Cdc25B. These results are validated by chemical shift perturbations obtained from NMR spectra of Cdc25B wild-type and point mutants disrupting putative cation−π contacts to identify important transient interactions in the protein conformational ensemble.

2. MATERIALS AND METHODS

2.1. Quantum Chemistry, Molecular Dynamics, and Free Energy Simulations. To evaluate the description of cation−π interactions by approximate force-fields, an isolated complex in the gas phase between blocked tryptophan (Ac-Trp-NHMe, denoted as bTrp) and blocked arginine (bArg) was used as a model system. The interaction energy with side-chains placed in two relevant orientations,[1] sandwich and T-shape (Figure 2), was evaluated quantum mechanically at the MP2/6-311++G** level corrected for the basis set superposition error with the counterpoise correction[26] and compared to CHARMM36[27] and AMBER99-ILDN[28,29] force-fields in the same geometries. The GAUSSIAN 09 rev. A program was used for quantum calculations.[30]

The reaction coordinate used here to describe cation−π interactions for both the bTrp···bArg complex and the Cdc25B free energy profiles below is the distance between the center-of-mass (COM) of all atoms in the Trp and in the Arg side-chains. These COMs roughly correspond to the position of Trp-Cδ2 and Arg-Nϵ atoms.

Molecular dynamics simulations of the complete catalytic domain of the human phosphatase Cdc25B were initiated from the structure in PDB code 1QB0. Sixteen residues (A551-Q566) lacking from the native C-terminal structure[16] and seven residues in the N-terminal part of the experimental construction (see below) were manually added in an extended conformation.[23] Crystallographic water, ions, and β-mercaptoethanol (BMER) were removed. Hydrogen atoms were added, and a standard protonation state at pH = 7 was

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

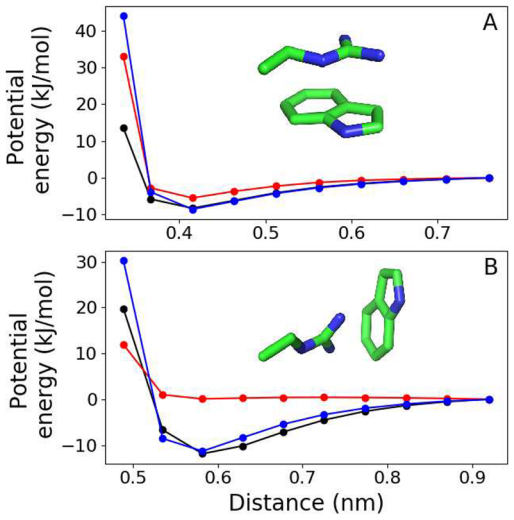


Figure 2. Potential energy for the interaction between Trp and Arg side-chains (bTrp···bArg complex) in sandwiched (panel A) and T-shaped (panel B) orientations, as shown in inserts. Black dots show the quantum chemical reference (MP2/6-311++G**), blue dots are the CHARMM36 force field, and red dots show the AMBER99-ILDN force field. Lines are guides to the eye.

assumed for side-chains. This is equivalent to the setup we used previously.[23] Two inorganic phosphate ions ($HOPO_3^-$ form) were added manually to the catalytic P-loop (R479) and to the secondary phosphate binding site near R488 and R492 side-chains as previous data such as phosphate titrations strongly support that these sites will bind phosphate ions under our experimental conditions.[17,23,25] This model was solvated in a dodecahedral simulation box of 13234 water molecules, with 4 Na^+ and 2 Cl^- ions added to neutralize the system and mimic the experimental salt concentration. This system was relaxed and equilibrated during a 100 ns free MD simulation. The CHARMM36 force-field for protein and ions and the TIP3P water model[31] were used. Simulations were performed with a 2 fs integration step, 300 K temperature, and 1 bar pressure (NPT ensemble) using the Bussi thermostat[32] and Parrinello–Rahman barostat both with a period of 0.5 ps. PME method with 1.2 nm real space cutoff and 0.12 nm Fourier spacing was used to treat long-range electrostatics. All classical force-field calculations, MD simulations, and analysis were performed with the GROMACS 4.6.7 program.[33]

Free energy profiles were estimated with US.[34] Initial configurations for each window were obtained after model equilibration described above. Seven US windows with reference reaction coordinate separated by 0.1 nm, from 0.4 to 1.0 nm, were used to estimate the profiles for each of Trp-Arg pair. A harmonic potential with force constant $k_{umb,dCOM}$ = 2000 kJ mol^{-1} nm^{-2} was used. Each US window was sampled for 500 ns with a sample collected every ps. Thus, the total aggregate simulation time was 20 μs. Potentials of mean-force were obtained with WHAM,[34] and the statistical uncertainty was estimated as 95% confidence intervals by bootstrap analysis with 50 resampling steps.[35] The initial 50 ns of each window was discarded to allow equilibration of orthogonal degrees of freedom in all analysis presented. The overlap of reaction coordinate distribution between adjacent windows was confirmed.

2.2. Point Mutants, Protein Expression, and Purification. Cdc25B W550A, R482A, and R544A point mutants were generated from pET-28a plasmids with the complete wild-type Cdc25B catalytic domain (S373–Q566) cloned and fused to His-tag and thrombin cleavage sites. The N-terminal sequence after cleavage was GSHMEFQ_(373)SDHRELI···, where the first seven residues are not native. Point mutations were introduced by using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent Tech. #210518).

Expression of wild-type Cdc25B and point mutants was made after *Escherichia coli* BL21-Gold (DE3) transformation with the respective plasmids and growth in M9 minimal ($^{15}NH_4Cl$-labeled) medium at 37 °C until OD_{600nm} 0.6 (exponential growth phase). At this time, 0.8 mM isopropyl β-thiogalactopyranoside (IPTG) solution was added for induction of protein expression, and temperature was maintained at 20 °C for 18 h when bacteria were then collected by centrifugation (7000 × *g*-20 min). Bacteria were lysed by sonication in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (BMER), 1 mM phenylmethylsulfonylfluoride-PMSF, pH = 7.4). The lysate was subjected to centrifugation, and supernatant was subjected to column affinity (Ni-NTA) and size-exclusion (Superdex 75 HR 10/30) chromatographies to protein purification as previously described in detail.[23,36]

2.3. NMR Spectra. All samples were prepared using 0.2 mM concentration of uniformly ^{15}N-labeled proteins in the same phosphate buffer preparation (20 mM NaH_2PO_4, 50 mM NaCl, 5 mM BMER, 2 mM dithiotreitol, 5% ^2H_2O, pH = 6.7) to obtain equivalent environmental conditions. NMR spectra were acquired on Bruker Ascend III 800 MHz NMR apparatus with 4-channel TCI cryoprobe and Avance III console. Initial processing of the spectra was done with NMRPipe program,[37] and assignment transfer for the main and side-chain resonances was performed with CCPNMR Analysis Assign 2.4 program.[38]

The chemical shift perturbation (CSP) analyzed for each assigned N–H system was defined according to

$$\mathrm{CSP} = \sqrt{\frac{1}{j_\alpha}(\Delta\delta^1\mathrm{H})^2 + \frac{1}{k_\alpha}(\Delta\delta^{15}\mathrm{N})^2}$$

where $\Delta\delta$ is the chemical shift change upon mutation, j_α and k_α are scaling factors (Table S1) that correspond to standard deviation of ^1H and ^{15}N chemical shifts of N–H system α (thus, unique to each amino acid type, main and side-chains) as observed in the RefDB database.[39] Statistical analysis was performed by the iterated removal of outliers defined as the values above the mean ($\bar{x}$) plus 3 standard deviations (σ).[40] CSP was considered significant when larger than $\bar{x} + \sigma$.[41] Wild-type and mutant chemical shifts obtained here were deposited in the BMRB databank (accession numbers from 50030 to 50033).

3. RESULTS AND DISCUSSION

3.1. Benchmark of the Empirical Force Field. The accuracy of a molecular simulation will depend on both energetic description and conformational sampling. Before addressing the importance of transient cation–π interactions in the protein Cdc25B, we benchmarked the interactions between Trp and Arg side-chains by comparing two widely employed protein force fields to high-level quantum chemical calculations.

Figure 2 shows the CHARMM36 force field correctly describes the cation–π interaction in both sandwiched (or stacking, parallel) and T-shaped (or perpendicular) orienta-

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890–897

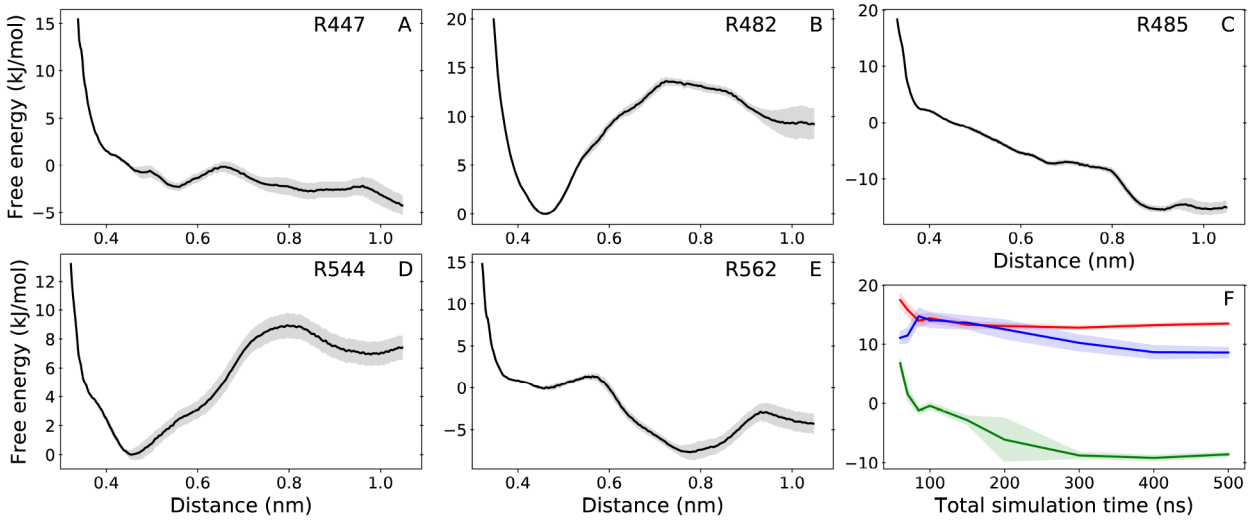


Figure 3. Free energy profiles for the cation−π interaction between W550 side-chain and R447 (panel A), R482 (B), R485 (C), R544 (D), and R562 (E) side-chains. Shadows around lines in all panels show the statistical uncertainty. (F) Convergence with total simulation time of the free energy difference between distance 0.45 and 0.80 nm for the cation−π complex with R482 in red line, R485 in green, and R544 in blue.

tions, with minimum energies within 0.5 kJ/mol of the MP2 energy reference. On the other hand, the AMBER99-ILDN energy function only agrees with the MP2 reference for the sandwiched orientation, with a difference of 3 kJ/mol for the minimum energy. Strikingly, the T-shaped orientation is fully nonbinding with the AMBER force field. Thus, CHARMM36 was employed for US simulations presented below, as this comparison suggests it should give a reliable energetic description for the mainly electrostatic interactions between Trp and Arg side-chains.[1]

Discrepancies between quantum and classical energy descriptions in small distances (0.3 nm for sandwiched and 0.5 nm for T-shaped orientations, Figure 2) are due to the steep (and formally approximate) behavior of the repulsive component employed in force fields.[42] This short-range discrepancy does not affect significantly the stability of cation−π interactions simulated here.

3.2. Free Energy Profiles for Trp-Arg Cation−π Interaction. Human Cdc25B catalytic domain contains 17 Arg, 13 Lys, and only one Trp residue. To reduce the number of simulations performed and simplify our analysis, we discarded Trp-Lys interactions and concentrated in Trp-Arg pairs, which interactions were shown above to be well described by the CHARMM36 force field. Arg residues visiting a distance of <8 Å from W550 during a previous long MD simulation (total of 6 μs, Table S2)[23,25] were considered. Partially buried residues (R479, R488 and R490) performing stable contacts with phosphate groups and R548 (next to W550 and impossible to check with the CSP analysis) were discarded. Thus, eight putative Trp-Arg pairs had the free energy for interaction between their side-chains estimated using US simulations. Results are shown in Figure 3 and Figure S1.

There are three kinds of free energy profiles. For R482 and R544 (Figure 3), the profile clearly shows a bound minimum at 0.45 nm distance and a dissociation barrier at 0.7−0.8 nm of 14 kJ/mol for R482 and 9 kJ/mol for R544, characterizing an important transient cation−π interaction with W550. For R485 (Figure 3) and R490 (Figure S1), the profile is downhill or dissociative, meaning that no stable interaction should be observed. For R447, R562 (Figure 3), R554, and R556 (Figure S1), the profile shows very weak minima and is almost flat in bound distances (0.4−0.8 nm), characterizing possible but short-lived and statistically unimportant interactions with W550.

The free energy difference between distance 0.45 and 0.80 nm in the profiles of Figure 3, which corresponds to the dissociation barrier of bound profiles, converges to within one statistical uncertainty in <300 ns of total simulation time per US window (Figure 3F). Thus, free energy profiles were estimated here with twice more sampling time than necessary for barrier convergence. Note the pair between W550-R548 was not studied in detail. Their neighbor side-chains will trivially engage in a cation−π contact, but it is impossible to validate this contact with the experimental CSP analysis proposed below.

The minimum distance and dissociation barrier of bound profiles (Figure 3B,D) are in good agreement with the distance and interaction energy of the minima for the bTrp···bArg complex found in our benchmark calculations (Figure 2A). The distribution of the angle formed between the planes defined by the indole group of W550 and the guanidinum of R482 shown in Figure 4A spreads out and shifts to higher angle values as the US reference coordinate assumes longer distances. The distribution is narrower in short distances because the T-shaped orientation is sterically hindered. In agreement with previous studies,[1] a combination of mostly sandwiched but also T-shaped orientations is found for important transient cation−π interactions within the bound state (interval from 0.4 to 0.8 nm).

Given the large number of Arg residues in Cdc25B, many flanking W550 and the catalytic site, we asked which interactions would W550 perform when unbound from one of its important Arg partner. With R482 restricted to a long distance, Figure 4B shows that W550 side-chain spontaneously forms complexes with R544 and R548, with very short-lived contacts with other Arg residues (R562 in this case). Similarly, when R544 is held away, R482 and R548 form complexes with W550 (Figure S4). Thus, there is a competition of transient cation−π complexes with W550 in Cdc25B.

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

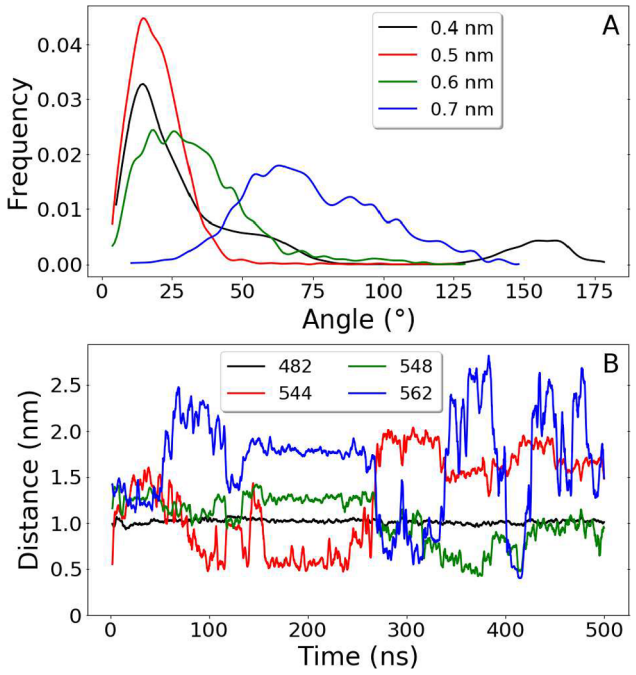


Figure 4. Analysis of Trp-Arg interactions during R482 US simulations. (A) Distribution of the angle formed between the planes defined by the indole group of W550 and the guanidinum of R482 with US reference coordinate centered in 0.4 (black), 0.5 (red), 0.6 (green), and 0.7 (blue) nm. Lower angles represent a sandwiched conformation and angles near 90 represent a T-shaped conformation, as in Figure 2. Data were smoothed with a spline interpolation. (B) Distance trajectories between the COM of W550 side-chain and the COM of R482 (black), R544 (red), R548 (green), and R562 (blue) side-chains, obtained with US reference coordinate in 1.0 nm.

3.3. Validation by NMR Chemical Shift Perturbation.

We attempted to experimentally validate the importance of transient cation−π interactions by comparing standard ^{1}H−^{15}N HSQC spectra acquired for Cdc25B wild-type and point mutants W550A (Figure 5), R482A (Figure S2), and R544A (Figure S3) in solution. Our premise was that a significant perturbation (i.e., larger than $\bar{x} + \sigma$) would be observed in the chemical shifts of residues performing important cation−π interactions by abolishing such contact upon a point mutation.

Main chain ^{1}H−^{15}N and Trp ^{15}Nϵ_1 side-chain signals were analyzed here as they are generally available and easy to acquire. We avoided additional NMR experiments (for instance, 2D HE(NE)HGHH)[14] that would be necessary to assign Arg and Lys side-chain resonances. Possible environmental perturbations were minimized by acquiring all spectra in the same batch and buffer solution. Assignments were transferred for 144 peaks out of the 150 originally assigned to wild-type Cdc25B,[23] and CSP was calculated for each point mutant, as shown in Table 1.

The CSP of cationic residues upon Trp mutation should be examined first. Insignificant perturbations found for R376, R427, R485, R488, R490, R492, and R562 in Table 1 imply that these residues do not perform important or long-lived cation−π interactions with W550. A similar conclusion is valid for most of the Lys residues (Table S3).

These observations are consistent with the Cdc25B crystal structure, where R376 and R427 are very distant from W550, but this flexible Trp residue approaches the other Arg located on the protein surface during a long MD simulation (Table S2). Why are their CSP insignificant? Side-chains of R488 and R492 make strong electrostatic interactions and remain buried within the secondary phosphate binding pocket.[17,23] Side-chains of R485, R490, and R562 are fully exposed to solvent but do not perform a stable cation−π interaction with W550, as shown by the dissociative free energy profiles estimated here (Figures 3 and S1).

A significant CSP is observed for R482 and R544 in the W550A mutant, in line with their bound free energy profiles. However, a significant CSP is also observed for R447, R529, R554, R556, and R557, in apparent disagreement with the almost flat free energy profiles for R447, R554, and R556 and with the long distances to R529 and R557 (Table S2) that suggest no important interactions with W550 are formed. How can this be explained?

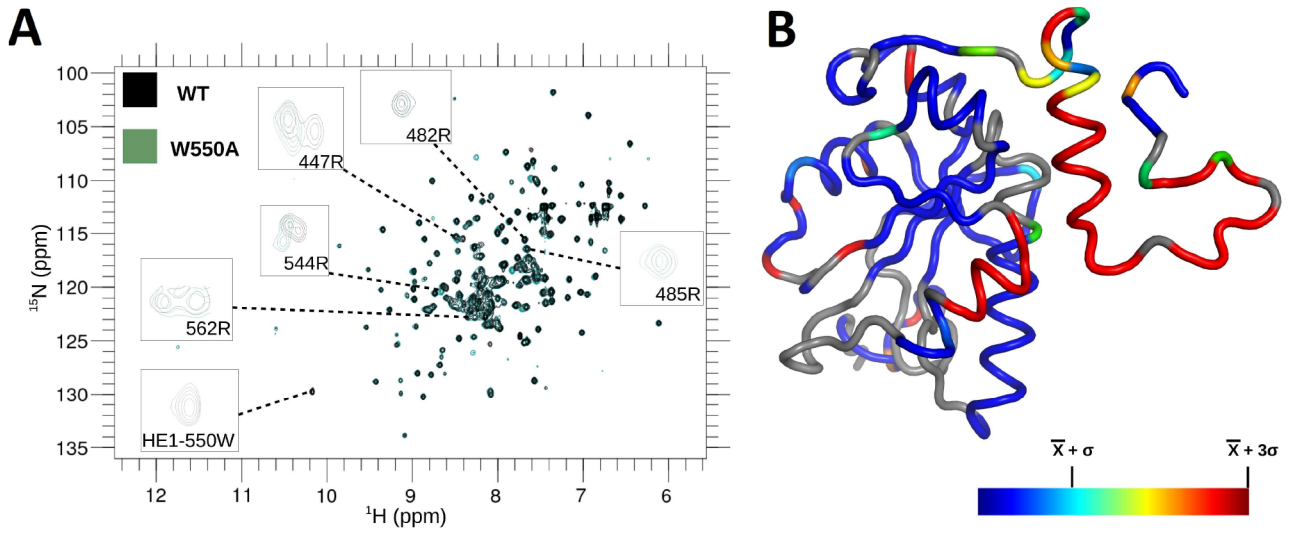


Figure 5. NMR spectra of Cdc25B catalytic domain and chemical shift perturbation upon mutation. (A) ^{1}H−^{15}N HSQC for Cdc25B wild-type and W550A point mutant and (B) CSP projected according to colorbar over the same structure shown in Figure 1. Gray segments were not assigned.

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

Table 1. CSP of Arg and Trp Main Chain N−H Centers for W550A, R482A, and R544A Point Mutants of the Cdc25B Phosphatase[a]

	mutant			
residue	W550A	R482A	R544A	PMF
Trp				
550	n.a.	**0.103**	**0.105**	
550ϵ_1	n.a.	**0.045**	**0.027**	
Arg				
376	0.011	0.014	0.018	
427	0.007	**0.036**	**0.042**	
447	**0.087**	**0.093**	0.011	F
466	n.a.	n.a.	n.a.	
479	n.a.	n.a.	n.a.	
482	**0.016**	n.a.	**0.029**	B
485	0.009	**0.062**	0.013	D
488	0.002	0.022	0.026	
490	0.004	**0.033**	0.006	D
492	0.004	0.021	0.006	
529	**0.016**	0.012	**0.041**	
544	**0.167**	**0.249**	n.a.	B
548	**0.170**	**0.083**	**0.100**	
554	**0.409**	0.019	**0.028**	F
556	**0.082**	**0.037**	**0.027**	F
557	**0.016**	**0.049**	0.015	
562	0.008	0.014	**0.070**	F
$\bar{x} + \sigma$	0.012	0.030	0.026	

[a]The Nϵ_1-H system in W550 side-chain is also shown. Unassigned centers are marked as n.a. Mean plus standard deviation ($\bar{x} + \sigma$) were calculated from around 115 CSP values for each point mutant after removal of outliers. Significative CSP are shown in bold. The PMF column codes whether the simulated free energy profile for the respective Arg-W550 cation−π interaction (Figures 3 and S1) shows a bound minimum for letter B, a dissociative profile for letter D, and an almost flat profile for letter F.

Chemical shifts are exquisitely sensitive to the distribution of protein conformations and to their interconversion rates.[13] In Cdc25B, R554, R556, and R557 are close to W550 in sequence and placed in the same disordered C-terminal segment.[23] If the W550A mutation alters the distribution or dynamics of main-chain configurations, it will lead indirectly to significant CSP for these Arg even if they do not perform important cation−π contacts. Although R447 and R529 are placed in the folded Cdc25B protein core, their segments have considerable flexibility as indicated by NMR order parameters ($S^2 \leq 0.95$ for D448-L453 and $S^2 \leq 0.90$ for D527-M531) and exchange rates ($R_{ex} = 17\ s^{-1}$ for E534) previously measured.[23] Allosteric modulation of the flexibility along these segments by the W550A mutation may also lead indirectly to significant CSP.[43,44]

These indirect effects may also account for the significant CSP observed in protein regions close to mutation sites of R482A and R544A constructions as well as for the CSP in R544 for the W550A mutant, although this Arg residue is in a stable α-helix. Thus, a significant CSP upon mutation of the π-system containing side-chain, here W550, is necessary but not sufficient to prove the importancy of a transient cation−π contact.

Mutations of cationic residues may also be explored. The CSP for W550 main-chain N−H and $^{15}N\epsilon_1$-H in the indole side-chain was significant for both R482A and R544A mutants (Table 1). This may be a relatively easy and general probe of transient cation-Trp interactions in other proteins, as the Trp $^{15}N\epsilon_1$ signal usually appears isolated and can be readily identified even for unassigned HSQC spectra.[13]

Given that a competition of transient cation−π interactions is present (Figure 4), abolishing one contact by mutation of the involved cationic residue would also perturb the molecular interactions (and the resonance signals) of the remaining Arg performing important contacts with the same π system. An indication of this competition in Cdc25B is the very high and localized CSP found for R544 upon R482A mutation (Figure S2), even though these residues are structurally distant (Figure 1). CSP is also significant for R548 and R556 in the R482A mutant and for R482, R548, and R556 (by only 0.001 CSP units) in the R544A mutant. On the other hand, insignificant perturbations are found for R529 and R554 in the R482A mutant and for R447 and R557 in the R544A mutant, indicating these four Arg do not perform important interactions with W550.

We may conclude that all CSP data (Table 1) and comparisons described here consistently suggest that *only* R482, R544, R548, and R556 may perform important transient cation−π interactions with W550 out of 15 assigned Arg residues in Cdc25B. Evidence is stronger for R482 and R544 placed in stable helices. Their mutation lead to significant CSP in W550 (Figures S2 and S3), presumably by abolishing a direct (side-chain) contact. The same analysis of the W550 CSP upon R548 and R556 mutations could not be carried out because these residues are placed in the same disordered C-terminus (pale green in Figure 1) and mutations would easily perturb the backbone configurational distribution.

Despite the various structural and dynamical sources of CSP upon point mutations which complicate the identification of transient contacts, another possible shortcoming of our NMR analysis is the non-unique formula used for calculation of CSP.[41] Here we found it essential to iteratively remove outliers[40] and to scale chemical shift changes by their natural standard deviation obtained from a database.[39] This is particularly relevant to be able to compare main-chain to side-chain CSP, as the dispersion of chemical shift for side-chain Trp-$^{15}N\epsilon_1$ is much narrower than the dispersion of all main-chain shifts (Table S1). For the free energy simulations, recent force-field parametrizations[45,46] and modified water models[47] could improve sampling of the backbone configurational distribution of the disorded C-terminal region in Cdc25B.

4. CONCLUSIONS

Free energy profiles estimated from US with MD simulations and perturbation of chemical shifts obtained from standard HSQC experiments in solution were combined here to identify transient cation−π interactions of solvent-exposed Arg-Trp pairs in proteins. The phosphatase Cdc25B, containing several Arg residues and only one Trp located in a disordered segment, was used as a model system.

Three kinds of free energy profiles were found: bound profiles (R482 and R544) with a clear dissociation barrier of 9−14 kJ/mol, characterizing an important transient cation−π interaction; dissociative profiles (R485 and R490), meaning that no cation−π interaction should be observed; and almost flat profiles (R447, R554, R556 and R562) with weak minima (≤6 kJ/mol) in bound distances, characterizing possible but

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

very short-lived and statistically unimportant cation−π interactions. The same three kinds of profiles should be found generally for transient cation−π interactions between solvent-exposed side-chains in other proteins.

The free energy profiles obtained for eight putative cation−π pairs in Cdc25B are in qualitative agreement with the CSP analysis, confirming that R482 and R544 perform important transient interactions with W550 in solution, even though their side-chains are separated by more than 8 Å in the crystal structure. The only exception is R556 which shows low but significant CSP and an almost flat free energy profile, with a shallow minimum. Although transient contacts of W550 with R447 and R485 were observed in previous MD simulations,[23] results presented here clearly show that these contacts are not important in the Cdc25B conformational ensemble. Thus, the statistical significance of transient contacts in solution should not be relied upon single observations over crystal structures or molecular simulations.

The experimental validation proposed here could be generally applied to other proteins. It is relatively simple to conduct once the protein constructs are purified and the (wild-type) $^{1}H-^{15}N$ HSQC spectra is assigned. However, NMR chemical shifts are sensitive to changes in both molecular contacts and conformational distributions. Disruption of a cation−π interaction upon mutation will directly result in a significant CSP by abolishing this molecular contact. But a significant CSP may also appear due to indirect changes in the residue conformational distribution. Thus, the proposed CSP analysis should be applied carefully and in combination with other statistically representative data.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00859.

Three tables with CSP scaling factors and side-chain distances to Arg and to Lys residues. Four figures with additional free energy profiles, NMR spectra, and distance trajectories (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: garantes@iq.usp.br.

ORCID

Guilherme M. Arantes: 0000-0001-5356-7703

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Funding from FAPESP (scholarships 2012/00543-1 and 2016/24096-5 to R.S.R.S. and research grants 2016/22365-9, 2018/25952-8, and 2018/08311-9), CNPq (scholarship 141683/2016-3 to A.A.O.R.), CAPES (finance code 001), and computational resources from the SDumont cluster in the National Laboratory for Scientific Computing (LNCC/MCTI) are gratefully acknowledged.

■ REFERENCES

(1) Ma, J. C.; Dougherty, D. A. The Cation-π Interaction. *Chem. Rev.* **1997**, *97*, 1303−1324.

(2) Burley, S. K.; Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **1986**, *203* (2), 139−143.

(3) Dougherty, D. A. Cation-π interactions in chemistry and biology: A new view of benzene, Phe, Tyr and Trp. *Science* **1996**, *271*, 163−168.

(4) Jing, W.; Demcoe, A. R.; Vogel, H. J. Conformation of a Bactericidal Domain of Puroindoline a: Structure and Mechanism of Action of a 13-Residue Antimicrobial Peptide. *J. Bacteriol.* **2003**, *185*, 4938−4947.

(5) Qamar, S.; Wang, G.; Randle, S. J.; Ruggeri, F. S.; Varela, J. A.; Lin, J. Q.; Phillips, E. C.; Miyashita, A.; Williams, D.; Strohl, F.; Meadows, W.; Ferry, R.; Dardov, V. J.; Tartaglia, G. G.; Farrer, L. A.; Schierle, G. S. K.; Kaminski, C. F.; Holt, C. E.; Fraser, P. E.; Schmitt-Ulms, G.; Klenerman, D.; Knowles, T.; Vendruscolo, M.; St George-Hyslop, P. FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-π Interactions. *Cell* **2018**, *173*, 720−734.

(6) Díaz-Moreno, I.; de la Rosa, M. Transient interactions between biomolecules. *Eur. Biophys. J.* **2011**, *40*, 1273−1274.

(7) Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction*, 1st ed.; CRC Press: Boca Raton, FL, 2010.

(8) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.

(9) Robustelli, P.; Stafford, K. A.; Palmer, A. G. Interpreting Protein Structural Dynamics from NMR Chemical Shifts. *J. Am. Chem. Soc.* **2012**, *134*, 6365−6374.

(10) Gu, Y.; Li, D.-W.; Bruschweiler, R. NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment. *J. Chem. Theory Comput.* **2014**, *10*, 2599−2607.

(11) Sharma, V.; Belevich, G.; Gamiz-Hernandez, A. P.; Róg, T.; Vattulainen, I.; Verkhovskaya, M. L.; Wikström, M.; Hummer, G.; Kaila, V. R. I. Redox-induced activation of the proton pump in the respiratory complex I. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11571−11576.

(12) Knapp, B.; Ospina, L.; Deane, C. M. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* **2018**, *14*, 6127−6138.

(13) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Academic Press: New York, 2007.

(14) Pellecchia, M.; Wider, G.; Iwai, H.; Wuthrich, K. Arginine side chain assignments in uniformly ^{15}N-labeled proteins using the novel 2D HE(NE)HGHH experiment. *J. Biomol. NMR* **1997**, *10*, 193−197.

(15) Juszczak, L. J.; Eisenberg, A. S. The Color of Cation-π Interactions: Subtleties of Amine-Tryptophan Interaction Energetics Allow for Radical-like Visible Absorbance and Fluorescence. *J. Am. Chem. Soc.* **2017**, *139*, 8302−8311.

(16) Reynolds, R. A.; Yem, A. W.; Wolfe, C. L.; Deibel, M. R.; Chidester, C. G.; Watenpaugh, K. D. Crystal structure of the catalytic subunit of Cdc25B required for G(2)/M phase transition of the cell cycle. *J. Mol. Biol.* **1999**, *293*, 559−568.

(17) Rudolph, J. Cdc25 Phosphatases: Structure, Specificity, and Mechanism. *Biochemistry* **2007**, *46*, 3595−3604.

(18) Rudolph, J. Catalytic mechanism of CDC25. *Biochemistry* **2002**, *41*, 14613−14623.

(19) Arantes, G. M.; Loos, M. Specific Parametrisation of a Hybrid Potential to Simulate Reactions in Phosphatases. *Phys. Chem. Chem. Phys.* **2006**, *8*, 347−353.

(20) Arantes, G. M. Free energy profiles for catalysis by Dual-specificity Phosphatases. *Biochem. J.* **2006**, *399*, 343−350.

(21) Arantes, G. M. The Catalytic Acid in the Dephosphorylation of the Cdk2-pTpY/CycA Protein Complex by Cdc25B Phosphatase. *J. Phys. Chem. B* **2008**, *112*, 15244−15247.

(22) Lund, G.; Cierpicki, T. Solution NMR studies reveal no global flexibility in the catalytic domain of CDC25B. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 2889−2895.

(23) Sayegh, R. S. R.; Tamaki, F. K.; Marana, S. R.; Salinas, R. K.; Arantes, G. M. Conformational Flexibility of the Complete Catalytic Domain of Cdc25B Phosphatases. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 1567−1575.

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897

(24) Arantes, G. M. Flexibility and inhibitor binding in Cdc25 phosphatases. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3017−3032.
(25) Sayegh, R. S. R. Conformational flexibility of the catalytic domain of Cdc25B phosphatase. *Ph.D. Thesis*, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, 2016.
(26) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94*, 1873−1885.
(27) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671−690.
(28) 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*, 712−725.
(29) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.
(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, A.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
(31) 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*, 926−935.
(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(33) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(34) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(35) Johnson, R. W. An Introduction to the Bootstrap. *Teaching Statistics* **2001**, *23*, 49−54.
(36) Souza, V. P.; Ikegami, C. M.; Arantes, G. M.; Marana, S. R. Protein thermal denaturation is modulated by central residues in the protein structure network. *FEBS J.* **2016**, *283*, 1124−1138.
(37) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6*, 277−293.
(38) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687−696.
(39) Zhang, H.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR* **2003**, *25*, 173−195.
(40) Schumann, F.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* **2007**, *39*, 275−289.
(41) Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1−16.
(42) Leach, A. R. *Molecular modelling: principles and applications*, 2nd ed.; Prentice Hall: Harlow, 2001.
(43) Tzeng, S.-R.; Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **2011**, *21*, 62−67.
(44) Lisi, G. P.; Loria, J. P. Allostery in enzyme catalysis. *Curr. Opin. Struct. Biol.* **2017**, *47*, 123−130.
(45) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.
(46) Liu, H.; Song, D.; Zhang, Y.; Yang, S.; Luo, R.; Chen, H.-F. Extensive tests and evaluation of the CHARMM36IDPSFF force field for intrinsically disordered proteins and folded proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 21918−21931.
(47) Boonstra, S.; Onck, P. R.; van der Giessen, E. CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State. *J. Phys. Chem. B* **2016**, *120*, 3692−3698.

DOI: 10.1021/acs.jcim.9b00859
J. Chem. Inf. Model. 2020, 60, 890−897