The role of zinc finger linkers in zinc finger protein binding to DNA

Mazen Y Hamed*, Reema Siam[1], Roza Zaid[2]

Chemistry Department, Computational chemistry program, Birzeit University, P.O Box 14, Birzeit, Palestine.

*Corresponding author: Dr. Mazen Y Hamed, email address: mhamed@birzeit.edu

1: Research Assistant, Birzeit University, 2: Graduate student, Birzeit University

Abstract

Zinc finger proteins (ZFP) play important roles in cellular processes. The DNA binding region of ZFP consists of 3 zinc finger DNA binding domains connected by amino acid linkers, the sequence TGQKP connects ZF1 and ZF2, and TGEKP connects ZF2 with ZF3. Linkers act to tune the zinc finger protein in the right position to bind its DNA target, the type of amino acid residues and length of linkers reflect on ZF1-ZF2-ZF3 interactions and contribute to the search and recognition process of ZF protein to its DNA target. Linker mutations and the affinity of the resulting mutants to specific and nonspecific DNA targets were studied by MD simulations and MM_GB(PB)SA. The affinity of mutants to DNA varied with type and position of amino acid residue. Mutation of K in TGQKP resulted in loss in affinity due to the loss of positive K interaction with phosphates, mutation of G showed loss in affinity to DNA, WT protein and all linker mutants showed loss in affinity to a nonspecific DNA target, this finding confirms previous reports which interpreted this loss in affinity as due to ZF1 having an anchoring role, and ZF3 playing an explorer role in the binding mechanism. The change in ZFP-DNA affinity with linker mutations is discussed in view of protein structure and role of linker residues in binding.

Key words: zinc finger; protein; MM/GBSA; DNA binding; molecular dynamics; mutation; linkers

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

1. Introduction

Zinc finger protein (ZFP) is present in the transcription factor (TFIIIA) as nine zinc finger domains[1]. Transcription factors containing a Cys_2-His_2 zinc finger domain functions in specific DNA recognition[2–4]. The three finger protein is organized in tandem arrays linked by peptide linkers (Figure 1). Zinc finger proteins play a role in DNA replication and repair, transcription and translation [2] [5].The protein structure is stabilized by a tetrahedrally coordinated zinc ion to two histidine residues in the α-helix and two cysteine residues in the β-sheet.

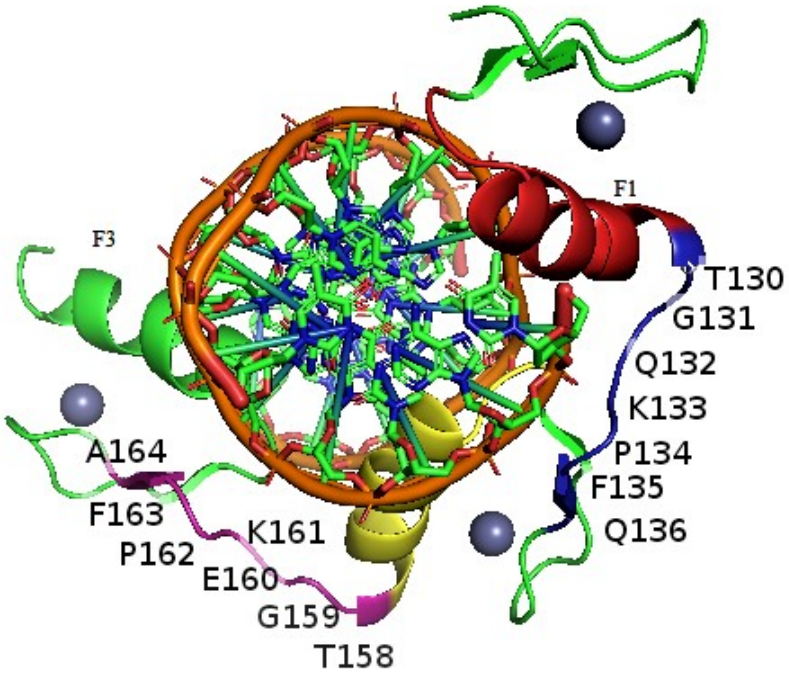


Figure1: Major amino acid residues on each linker in ZFP bound to DNA target (5'-GCG TGG GCG- 3') average structure after MD simulation. Finger1 (Red); linker1 (blue) with labeled residues T28G29E30K31P32). Finger 2 Yellow; linker2 (pink with labeled residues T56G57G58K59P60) and finger three (green).

The zinc finger protein- DNA structure is reported as zif268 (PDB ID 1AAY) [6]. Experimental work revealed that zinc finger protein makes both specific and nonspecific contacts

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

with DNA. In specific contacts, the amino acid residues in positions -1, +2, +3, and +6 in the α-helices bind specifically to the DNA sequence 5'-GCG TGG GCG- 3' [7] [8]. The binding energy of the protein to its specific DNA target is an enthalpy driven process resulting from amino acid side chains contacts with specific DNA bases[7] [8][9][10]. Displacement of bound water molecules from the protein-DNA interface and hydrogen bonds to back bone phosphates also contribute to the entropy of binding [11][12] [13].

The role of linkers in the ZFP- DNA binding process was studied experimentally [14] [15] [16] [17] [18][19][20]. The sequence TGQKP links ZF1 and ZF2 and TGEKP connects ZF2 and ZF3 as shown below[17] [5, 21–24].

F1	TGQKP	F2	TGEKP	F3
	Linker1		Linker2	

The length of linkers one and two are 14.5 Å and 14.4 Å which gives enough flexibility to the Zinc fingers to bind DNA targets. Indeed, Pabo et al [6] used phage display and modeling of three separate fingers docked on B- DNA and obtained a distance of 18.1 Å between F1 and F2 and 17.7 Å between F2 and F3, this led him to conclude that there is a need for finger- finger interaction and DNA twisting to shorten these distances for binding to take place [6] [14] [15][27].

Linking more than three fingers using TGEKP to target more than 10 bases showed reduced affinity[19]. Designed ZF motifs with more flexible linkers were found to reduce the strain and to improve specificity, such designs made the targeting of unique DNA sequences more possible and aided the building of various sets of 6 fingers linked by varying linkers[19] [26][27]. The linkers in the first three fingers in zinc finger protein in TFIIIA are dynamically disordered in solution and they show a well ordered structure upon binding to DNA [16, 20, 28].

Choo et al [21] investigated the effect of mutating the linkers and measured the protein-DNA dissociation constants, the result was that the affinity of point mutants T(L), K(S),P(G) of the linker TGEKP was reduced by a factor ranges from 7 to 13-fold, also reported a 24-fold loss in binding affinity for the $G^{29}P$ mutant**.** The mutations of three amino acids in the linker (TGQKP) using site directed mutagenesis was found to reduce the ZFP binding affinity to DNA by 8-fold [17]. A 50-fold reduction in specific binding was reported upon mutation of linker one (TGEKP) to SEQKP and this was interpreted as an effect on specificity[17]. NMR spectroscopic study of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

zinc fingers showed that the interactions between fingers initiated by linkers contributed to the affinity of binding[29] [20]. Crystal structures and NMR spectra of the ZFP-DNA complex revealed the formation of a hydrogen bonding between the gamma oxygen (O^{γ}) threonine (T1) on the α-helix and the back-bone amide of the linker's amino acid E^{3}. This interaction is known as "C-capping" and acts as a lock on the specific site once the mobile ZFP finds its target[20].

Experiments demonstrated that the intrinsic flexibility of linkers is needed for diffusion of ZFP along the DNA in search of its cognate site[14] [15]. This aids in the dynamic DNA scanning process in which ZFP binds nonspecifically to DNA before it switches to its specific site[14]. Mutating key amino acids in certain positons, namely: -1, 3 and 6 reflects on specificity and affinity[8][7]. kinetic and linker mutations studies proved that linker flexibility and charges of residues involved are expressed in the dynamic nature of the DNA binding process. Positions of residues in the linker sequence enable the fingers to twist to an optimum position for binding.

Linkers establish a balance between the scanning of DNA (nonspecific DNA binding) and other thermodynamic aspects on the way to establish an equilibrium between search and recognition. Linker mutations shifted this equilibrium in a way to influence the affinity and/or to accelerate the kinetically driven search process[14][15]. Experimental studies on mutating T28 to P32 residues in linker one (L1) showed pronounced effect on the protein binding to DNA [18][20][17][30].

In view of the above findings, it became worthwhile to study the energy effect upon linker mutations in relation to DNA binding process. The DNA binding energies of ZF protein with varying linker sequences are expected to show variations [7][8]. The outcome will help scientists to build zinc finger proteins with various linker sequences and tune these motifs to expand the binding to more specific DNA targets [28]. Linker design must make sure that the target DNA remains accessible to ZFP, also to take into consideration the major and minor grooves of DNA which adds to the complexity of ZFP-DNA architecture. This has made the experimental efforts in designing new linkers more difficult.

In this work, we calculated the binding affinity for zinc finger protein-DNA complexes based on the crystal structure of the Zif268-DNA complex and employing molecular dynamics (MD)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

simulations and MM-PB(GB)SA analysis[31][32][33]. Our simulation reproduces the known experimental binding affinity for the three-finger protein-DNA complex. Our results show that the binding affinity changes as the linker residues are mutated. The change is explained as a result of the strong enthalpy change that arises due to a small conformational change in DNA when the three-finger protein is bound.

Mutations of ZFP were performed on residues which are not involved in contact with DNA to maintain specificity. The mutants were designed in a way to test the ZF1-ZF2-ZF3 interactions in relation to ZF-DNA interactions on the basis of ZFP-affinity to specific and nonspecific DNA targets and the relationship to search and recognition mechanism.

2. Computational Methods

2.1 Molecular dynamics simulations: Molecular dynamics simulations were performed on the initial structures based on the 1.60 Å resolution x-ray crystal structure of the zif268-DNA complex, PDB ID 1AAY[6]. Simulations were carried out using the Amber18.0 package [34][35] and employing the AMBER force field ff14SB for proteins and nucleic acids. All atom explicit water molecular dynamics simulations were performed on all systems. PDB file of the Initial structures of zinc finger protein-DNA complex was prepared using **pdb4amber** program, inspected, salt and water were removed. Mutations were performed using either ***PyMOL*** [36]for mutating amino acid residues in the linker region or the UCSF ***Chimera*** package from the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) for mutations in the DNA bases. *Tleap* editor in AMBER 18 was used to prepare the parameter and coordinate files.

In order to preserve the coordination structure of the tetrahedral zinc ion to Cys_2His_2, zinc ion library files and coordination parameters were loaded to the Amber library[37]. The zinc library files were always loaded to the editor prior to loading the protein-DNA complex. The system was always checked for any changes in zinc ion coordination after each run and no significant changes were observed. The complexes were solvated in a TIP3P cubic water box[34] with dimensions of 15.0 Å. Na^+ ions were added to neutralize the system. The structure was checked for errors and then converted to topology and coordinate files.

The particle mesh Ewald method was used for treating long range electrostatics, a 9Å cutoff was set for long range interactions. The force field energy of each structure was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

minimized by progressively relaxing the system before starting the MD simulations. Minimizations were performed employing steepest descent followed by conjugate gradient minimizations (1000 cycles in tandem). After relaxation of the system it was heated to 300K applying harmonic restraint (10 Kcal/Å^2.mol) on solute. This was followed by an unrestrained 2ns MD simulation at 300K and 1 atm to equilibrate the system and adjust the density. The Shake algorithm was used to constrain hydrogen atoms in order to enable a longer time step (2fs) in the simulation. A Langevin thermostat with 2 ps^{-1} collision frequency and weak coupling barostat with 2 ps of relaxation time were employed. Production MD simulations were carried out for 50, 100 and 150 ns and gave converged trajectory. Trajectories were collected at 2 ps intervals, these trajectories were used to calculate the binding free energy using 'MMPBSA.py' script, 50 or 100 frames were used in calculations (for details see supplements S-methods-1). The binding energy of the Zif268-DNA complex was calculated using both MM_GBSA /PBSA methods, entropy contributions were calculated using the normal mode. All properties were monitored during simulations to avoid any sudden jumps in system properties (see Figures 1-S, 2-S,3-S and S-an -6 in supplements), [38]. The temperature rises regularly from 0 K to an equilibrium value of 300K during the simulation indicating that Langevin dynamics worked effectively. Over the remaining part of the simulation the system was kept under constant pressure to make sure that the system has reached the equilibrium state [39][40][41]. Both binding sites used in this study (5'-GCG TGG GCG- 3') and (5'-GCG GGG GCG- 3') were reported as consensus high-affinity binding sites for Zif268, with a slight difference in their binding affinity with ZF in favor of the former [42–44]. The nonspecific binding was studied using the nonspecific DNA site (5'—GCA GAT TCC—3') studied previously by ^{1}H NMR as by [15][14].

2.2 Analysis of MD trajectory:

In all trajectory files *cpptraj facility in Amber* was used to strip water and Na^+ ions. RMSD analysis on wild type protein and its mutants with and without complex with DNA showed stability of all systems and it also showed that both the proteins and their complexes are well equilibrated. All RMSD plots indicate reasonable stability for all mutants and WT protein bound to DNA after 100 ns MD simulation.

Pairwise RMSD for specific snapshots was Computed using *pytraj* in Amber. The RMSD to the experimental structure reference was computed, then, pairwise RMSD for first 50 snapshots and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

skipping every 10 frames was computed (For output see Figure 1-S, 2-S and S-An-6 in supplements)

Clustering:

Free energy calculations of a system require well-converged sampling of the ensemble of structures. Analysis of MD simulations is necessary to give reliable calculated energies. Principal component analysis (PCA) measures the dynamic nature of a structure and complements cluster analysis (Clustering-1-S and Clustering 2-S in supplements). While clustering gives information about adopted conformations during simulation, PCA provides information about motions of the structure undergone to achieve conformation. Input to PCA coordinate covariance matrix, entries to this matrix are covariance between X, Y and Z components of each atom. PCs can be visualized by converting to NetCDF pseudo-trajectory file of motion. All pairwise RMSD's between conformations were computed and distances were transformed to similarity scores. Clustering analysis [45][46] was performed using mdtraj, numpy and scipy[47] and pytraj (Amber.jupyter notebook) in AMBER. Finding centroid (representative structure for group conformations, this group might potentially come from clustering using Ward hierarchal clustering ways to define centroid. Using an *md.rmsd* [47] algorithm to compute all pairwise RMSD's between conformations, transform these distances to similarity scores:

To check the effect of MD simulation on the protein structure: The average structure was created after simulation and converted to pdb files (using ambpdb program in Amber) and structures superimposed on the original structure using pymol [48], no evident changes in the structure were observed. (for details, see S-methods-4 in supplements).

Zinc Parameters and binding

In order to preserve the coordination structure of the tetrahedral zinc ions to Cys_2His_2, amino acids were manually modified to enable the binding to zinc ions, by manually modifying CYS to the Amber deprotonated CYM and histidine bound to zinc where modified from HIE to the Amber recognized deprotonated HID, zinc ion library files and coordination parameters were prepared and loaded to the Amber library. The zinc library files were always loaded to the editor prior to loading the protein-DNA complex. Add atom types for the ZAFF metal center then the library for atomic ions loaded using the command *Load ZAFF prep* file and *Load ZAFF frcmod* file then the zinc bonded to the CYM and HID using the command bond.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

2.3 Calculation of the Free Energy of Zinc finger interaction with DNA bases:

Before using MM-GBSA[31][32][49][50] the system equilibration was verified by considering temperature, density, total energy and root mean squared deviation of coordinates (RMSD). An RMSD value relative to the crystal structure of 1.5Å was deemed acceptable. Extensive analysis of trajectory was performed to make sure the energies calculated are reliable depending on the snapshots (see Analysis section). The resulting trajectories were analyzed using MMPBSA.py script [33][51][52].

MM_GBSA method in AMBER18 was employed for the calculations:

$\Delta G_{binding(Solvated)} = \Delta G_{binding(vac)} + \Delta G_{Solvation(complex)} - (\Delta G_{solvation(Lig)} + \Delta G_{solvation(receptor)})$

$\Delta G_{solvation} = G_{electrostatic(\varepsilon=80)} - G_{electrostatic(\varepsilon=1)} + \Delta G_{hydrophobic}$,

The free energy of binding contains three major contributions: Van der Waals, electrostatic and solvation energies. The solvation energy contribution was calculated using both the polar contribution to solvation (GB) and the nonpolar contribution to solvation (calculated based on accessible surface area (SA)). Normal mode calculation was used to calculate entropy contributions, $T\Delta S$, from the same snapshots as were used for calculating $\Delta G_{binding}$. Then the absolute free energy of binding was calculated from

$\Delta G = \Delta G_{binding} - T\Delta S$

$\Delta G_{binding} = \Delta H - T\Delta S$

$\sim \Delta E_{MM} + \Delta G_{sol} - T\Delta S$

and ΔE_{MM} consists of:

$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw}$

$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$

3. Results and Discussion

3.1 RMSD and clustering Analysis: In order to check the convergence of the MD simulations, the root mean square deviations (RMSDs) of heavy atoms are plotted in Figure 2A. The RMSD values of the ZFP-DNA (Blue) and the unbound ZFP (green line in 2A) and the unbound DNA (Figure 2B) all remain stable around 2 Å during the entire simulation time. The systems showed stability, Figures 3A shows examples of the RMSD plots of the WT zinc finger protein complex

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

and its unbound protein and DNA and 3B shows the DNA complex of K59Pmutant (green) and the unbound protein (red), (See also supplements S-An-6, Figure 1-S Clustering and Figure 2-S).

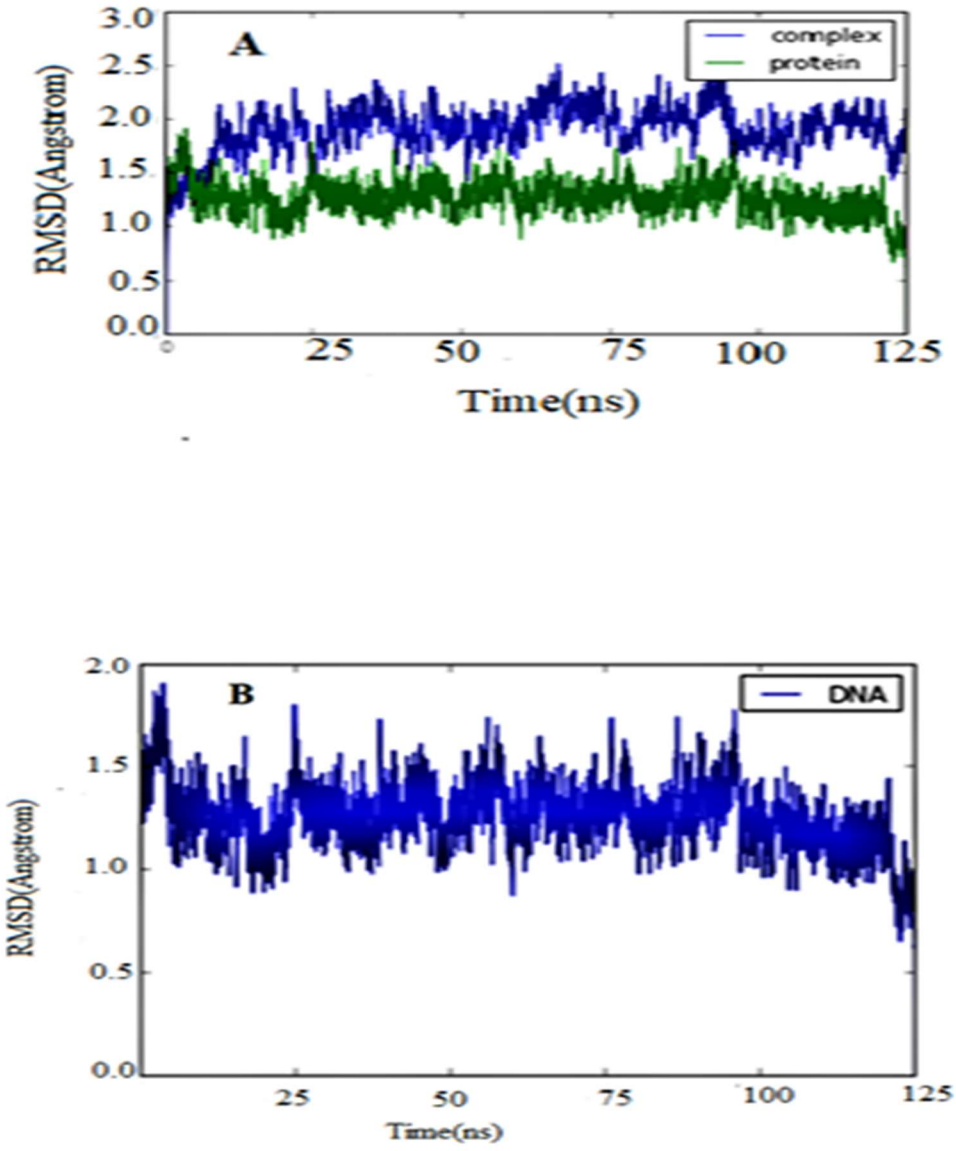


Figure 2: RMSD plots of Zinc finger protein; A): RMSD of wild type ZFP-DNA complex (blue); unbound ZFP protein (Green). B) RMSD of DNA (Blue)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

The RMSD analysis showed that both the wild type protein and its mutants reached a well equilibrated state.

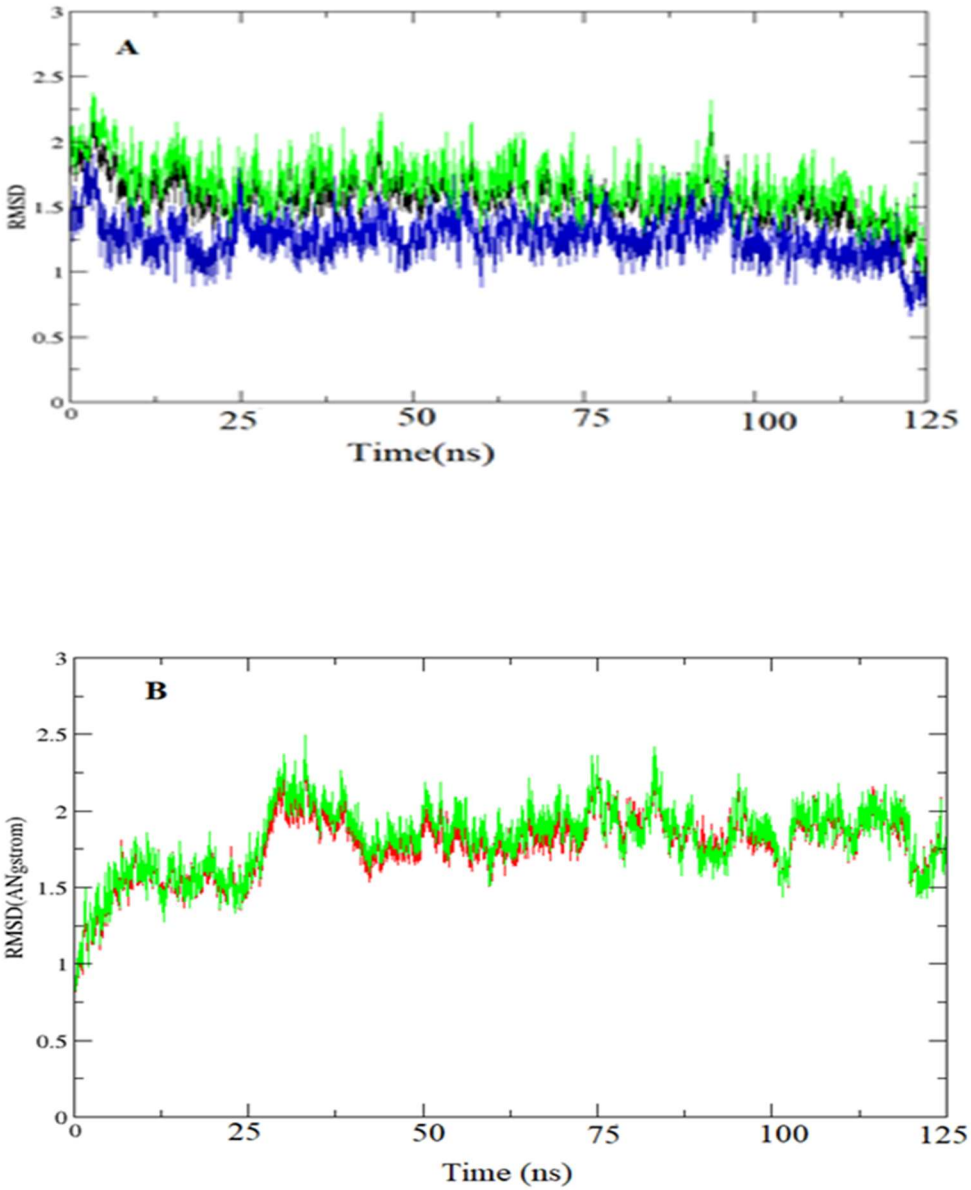


Figure 3: RMSD (Å) after 125ns MD simulation in reference to the original structure calculated (back bone only). A) ZF protein-DNA complex (Black); unbound ZF protein (Green); DNA (Blue). B) Comparison of RMSD of the mutant protein K59P in complex with DNA (Green) with RMSD of the unbound protein K59P (Red).

When each simulated mutant structure was superimposed on the structure of WT protein, the mutant structures show a slight movement in both zinc finger 1 and zinc finger 3 towards the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

major groove of DNA. Finger one is shifted by a distance of 3.5 to 4.5 Å closer to DNA compared to that in WT protein. This shift was also observed for finger 3. To the contrary, in the K31N mutant F1 moved away from the DNA groove, this is in line with the loss in affinity of the mutant. (See Figures S 3-C in supplements compare the simulated WT zinc finger with the its starting structure from X-ray)

Clustering the molecular configurations from MD trajectory[45][46] resulted in distinct sets of groups of similar molecular configurations (See supplements, Figure 1-S, 2-S, 3-S), this process allowed a refined view of how the protein molecule is sampling conformational space and makes direct characterization of separate conformational sub-states visited by MD. It is known that large-scale conformational changes during MD can lead to high variance for calculation of properties such as MM-PBSA/GBSA estimates of free energy and entropy calculations. Clustering of trajectory into sub-states minimizes the variance. The average linkage hierarchical clustering, where two major clusters of configurations resulted (one is more populated than the other with 0.25 nm distance. Which allowed the assignment of configurations to the closest centroid, it was reported that this method performs well up to 500 ps [45]. (Plots in supplements S-An-6)

3.2 Binding of ZFP which contains mutated linkers to a specific DNA target (5'-A GCG TGG GCG T- 3'):

The affinity of WT ZFP and its mutants in Table 1 to a specific DNA target are used to study the structural role in the ZFP DNA scanning before achieving specific binding. The DNA binding process starts with scanning (search followed by recognition (specific binding)).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

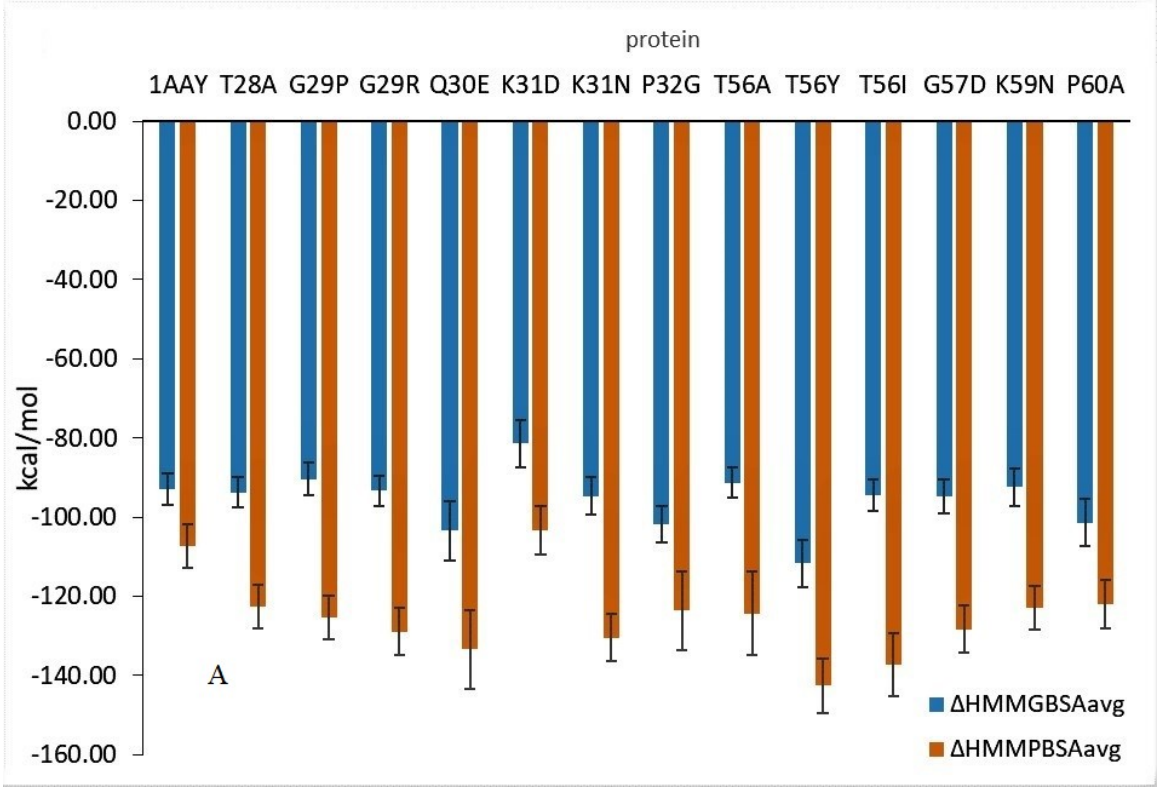
Table1: ZFP linker point mutations. The point mutant residues shown in red color

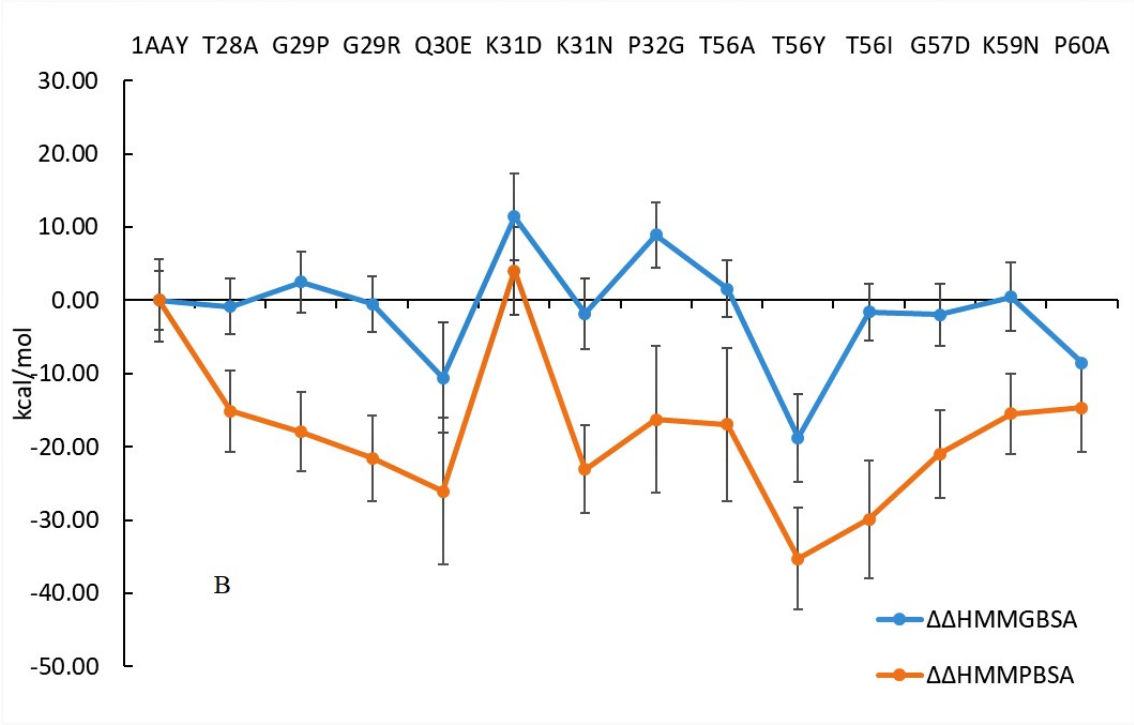
		L1		L2	
WT	F1	TGQKP	F2	TGEKP	F3
MUT1		AGQKP		TGEKP	
MUT2		TPQKP		TGEKP	
MUT3		TGEKP		TGEKP	
MUT4		TGQDP		TGEKP	
MUT5		TGQNP		TGEKP	
MUT6		TGQKG		TGEKP	
MUT1		TGQKP		AGEKP	
MUT2		TGQKP		YGEKP	
MUT3		TGQKP		IGEKP	
MUT4		TGQKP		TDEKP	
MUT5		TGQKP		TGENP	
MUT6		TGQKP		TGEKA	

The equilibrium state is a result of a balance between dynamic (nonspecific) binding and a specific binding. Any loss in affinity is interpreted as loose binding of ZF1 due to conformational changes in the protein resulted from mutations which confirms the anchoring role of ZF1 in the binding process. Weak nonspecific ZFP-DNA interaction shifts the conformation towards the search mode, while strengthening this interaction shifts the conformation towards recognition mode.

point mutants of linker one (L1) and linker two (L2) were built using pymol as shown in Table 1.The enthalpies, free energies of binding of mutants to the specific DNA seqence (5'-A GCG TGG GCG T- 3') were calculated (see Tables 1-S, 1-SB, 2-S, 3-S and 4-S in supplements) and plotted along with energies of the WT protein binding. (Figures 4 and 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1





Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

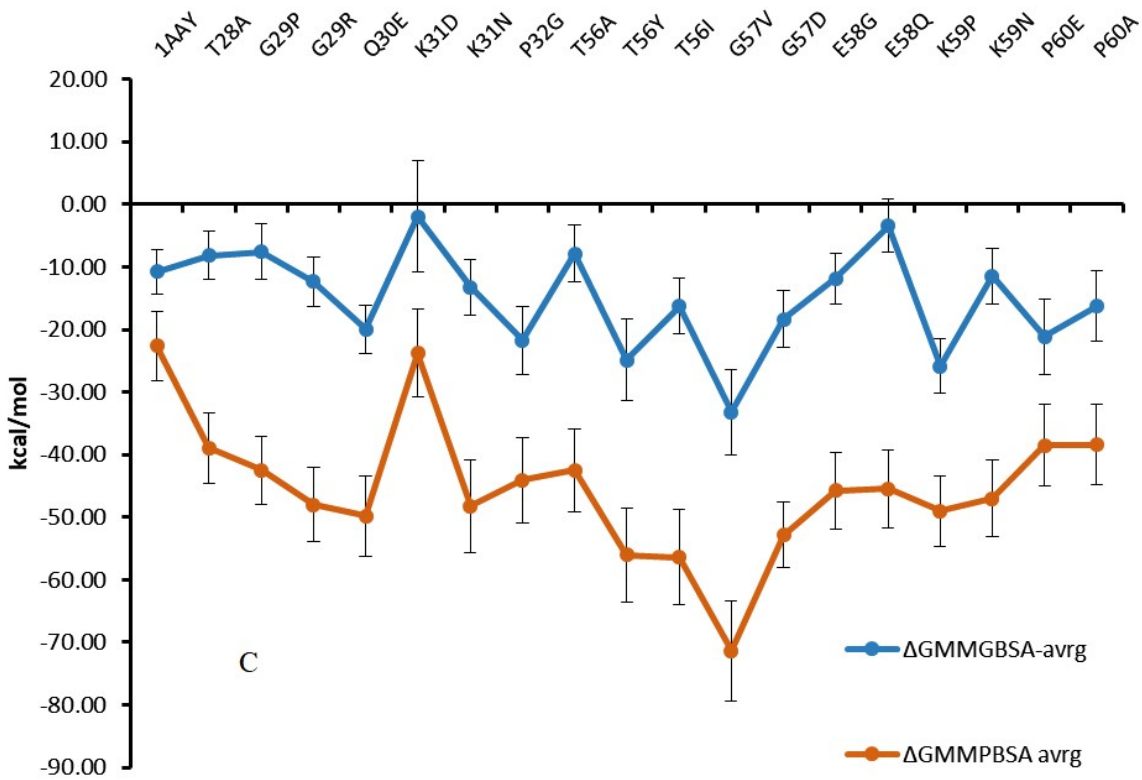


Figure4 A: The average values of binding Energies (ΔH) of WT zinc finger protein (1AAY) and its single residue mutants of linker 1 and linker 2 bound to its specific DNA target 5'-GCG TGG GCG- 3'as calculated by MM_ GBSA (Blue) and MM_PBSA (brown). [For values see tables 1-S to 4-S]

Figure 4B: The change in binding energies (ΔΔH) values of single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3'measured relative to wild type ZFP ΔH value in Figure 1B. Average values calculated by MM_ GBSA (Blue) and from MM_PBSA (orange-brown).

Figure4C: Free Energies of binding (ΔG) values of single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3' as calculated by MM_ GBSA (Blue) and MM_PBSA (brown) using the equation: $\Delta G_{binding} = \Delta H - T\Delta S$.

The energy values obtained by MM/GBSA are more reliable and confirms the reported binding energy values[18] [49] [49] , these values are used in the interpretation of changes in affinity measured by ΔH and ΔG. The reference energy is that of WT zinc finger protein binding specifically to its target. Contributing energy components are: van der Waals energy, polar interaction energy (Coulomb energy plus polar solvation free energy), nonpolar solvation free

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

energy, and internal energy. The Coulomb and polar solvation terms have large absolute values with opposite signs, and the sum of the two terms represents the net effect of polar interaction, For plots of energy values see Plots-S in supplements)

Each amino acid in the linker region of the ZF protein plays a role in the DNA binding along with its neighboring residues in addition to DNA bases[21][44], their role in affinity and specificity of ZF protein binding was extensively discussed in view of experimentally determined K_d values for ZFP mutants with DNA[44][6][17]. In the present study, each linker residue was mutated separately to produce point mutants of the protein (Table 1)[44] [6]. Three trials on each point were used for the calculated energies at 50, 100 and 150 ns simulations, and the average energy was calculated in all trials (some energy shifts shown in Table 2).

Table 2: Calculated energies of WT ZFP and linker mutants compared to that calculated by experiments, the shift in binding energies of some mutants from ΔG of WT binding (ΔΔG) is shown

protein	ΔG kcal/mol	ΔΔG
1AAY Exp. value reported[19]	-10.87(3) -13.1 -10.1(3.4)	0
T28A	-8.13(4)	+2.74
G29P	-7.57(4.5)	+3.3
G29R	-12.5(4)	-1.63
K31D	-1.9(5)	+8.97
E58Q	-3.3(3)	+7.57

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

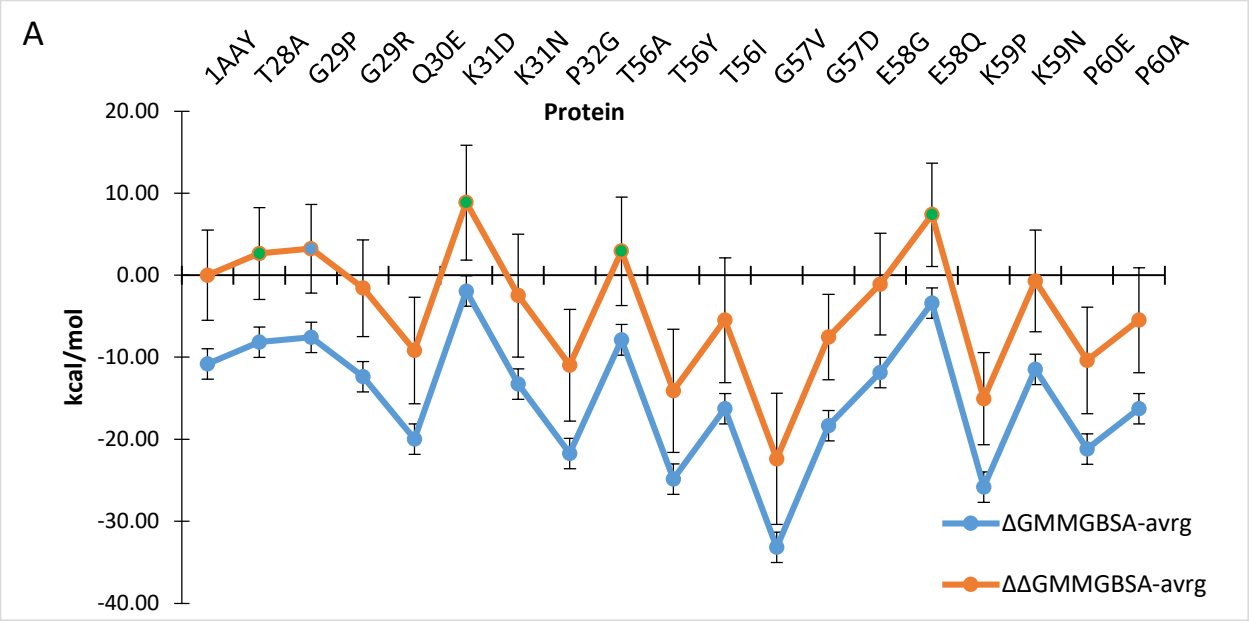
Figures 4A and 4B show the resulting ΔH values of binding based on both GBSA and PBSA calculations. Both methods showed a parallel behavior with difference in numerical values due to the inherent nature of how each method deals with energy calculations but the trend in energy change is parallel [44] [6][54]. Some mutants in both L1 and L2 showed loss in affinity except Q30E in L1 and T56Y, P60 mutants in L2, see plots of $\Delta\Delta H_{GBSA}$ in Figure 4B and $\Delta\Delta G_{GBSA}$ 5A, these results are in agreement with the reported site directed mutagenesis results where three linker amino acids were substituted in L1 and L2, with corresponding amino acids from p43 [17]. Wild-type zinc fingers 1–3, bind DNA with a dissociation constant of ~3 nM. Substitution of amino acids TGE in the conserved Krüppel-like linker, TGEKPF between zinc fingers 1 and 2, with amino acids SEQ from p43, reduced the affinity to DNA by 5-fold. A similar reduction in affinity (3-fold) was observed when the second linker, TGE, was substituted with LAL residues. The DNA binding affinity in which both linkers were substituted, was reduced by 8-fold (Kd = 24 ± 4 nM) compared with the native TFIIIA fragment[44][17]. The disruption of the hydrogen bond in linker one; between Threonine (T1) in linker with the gamma oxygen (O^{γ}) of back-bone amide of E3 in the mutant T28A; reduced the affinity of ZF protein to DNA[28][6][56]. Indeed, the mutation of K31 in linker one reduced the affinity of ZFP to DNA, this loss was explained as due to the loss of water- mediated contacts to the 5′ phosphate of base 5 and the same was observed for K59 in linker2 in agreement with Pabo et al[6] (see Figure 4B and 5A)

In general, L1 mutations reduced the affinity to DNA specific target in agreement with experimental work while mutations in L2 is divided,T5Y and P60A enhanced the affinity and T56A, T56I, G57D and K59N resulted in loss in affinity, all mutants compared to the affinity of the wild type protein [6][17][44]. Studying the intact Protein structure gives more insight on contribution of the protein tertiary structure to the affinity and recognition of ZFP to its specific DNA target, contrary to studying single finger binding because linkers are known to play an important role in the zinc finger protein binding to its target DNA by a determined balance of interactions between ZF1-ZF2-ZF3 on one side and ZFs-DNA which was ignored in single finger studies[6][57].

Linkers (L1 and L2) are reported to function in the ZFP recognition to its specific site after a dynamically driven nonspecific site search. They also influence the way that the ZFP folds itself around the target DNA[14][15]. A typical ZF protein binds DNA by a series of H-bonds between

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

amino acid side chains in positions -1, 3 and 6 and DNA base pairs, in addition to nonspecific hydrogen bonding with the back-bone phosphates of DNA. Zinc fingers are connected with the conserved linker TGEKP which plays an important role in specific binding and this binding is sensitive to residue mutations in the linkers. In view of these results, designs with more flexible linkers were found to improve DNA binding specificity and recognition. The importance of type of residues in linker one (L1) is clear with special emphasis on residue number 5 in linker one (K31) followed by position 4 in L2 (E58) followed in importance by positions one (T28) and two (G29) in L1 and to less importance position one in L2. The type of amino acid residue plays a role in affinity as replacing the same residue with more than one type of amino acid gives different affinities, for example G29P shows loss in affinity while G29R shows increased affinity. The same can be said upon mutation of T56 in L2 when mutated to A, Y and I respectively. The mutants E58G and E58Q and both K59P and K59N; showed variations in affinities with varying the incoming amino acid residue to replace the same original residue. This change in affinity can be seen in the plot in Figure 5A in the form of ΔΔG values relative to the WT protein. K31D and E58Q show the highest loss in affinity amongst mutants.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

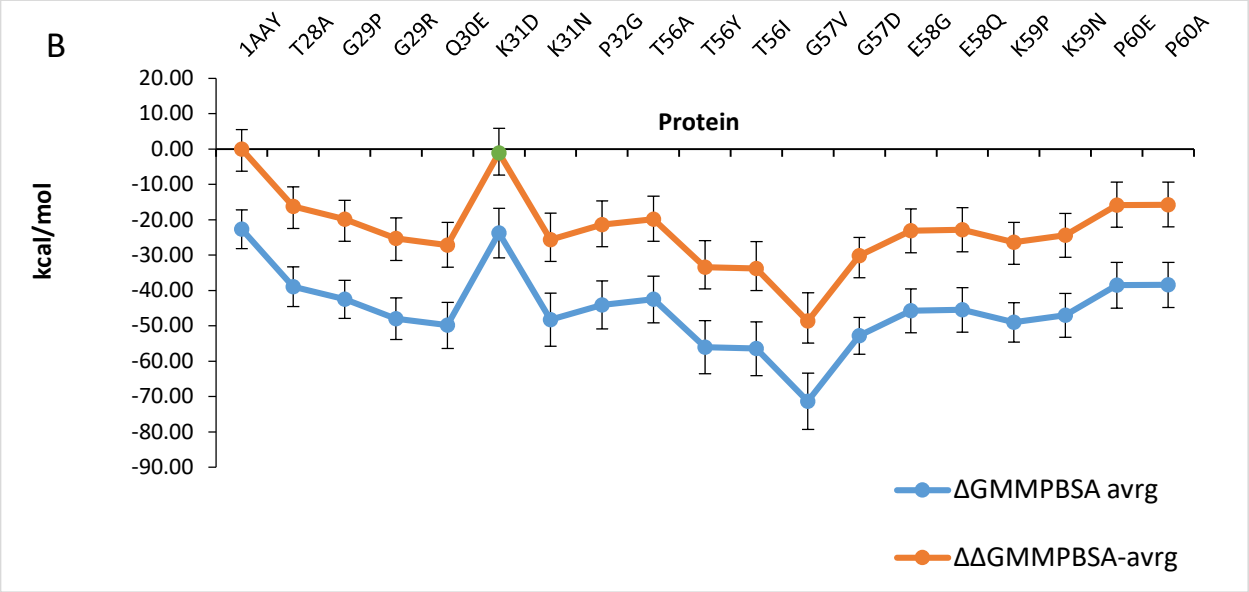


Figure 5

A: Free Energies of binding (ΔG) values of single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3' as calculated by MM_ GBSA (Blue) compared to ΔΔG values relative to the WT value as reference (brown). [for values see Table 4-S]

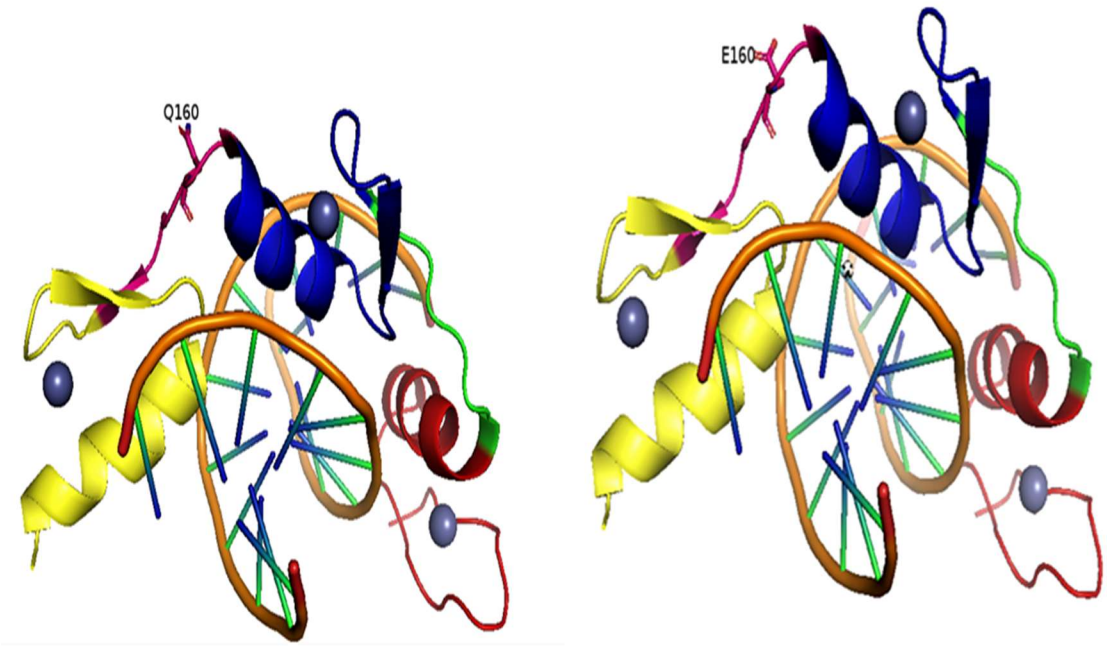
B: Free Energies of binding (ΔG) values of single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3' as calculated by MM_ PBSA (Blue) compared to ΔΔG values relative to the WT value as reference (brown).

The effect of residue type, vdw volume, charge and position in linker on ZF protein affinity to DNA is evident (see data in supplements Table 5-S). In K31D mutant the loss in affinity took place upon the change from positive to negative residue with a smaller radius, where in L2 E58Q suffered maximum loss in affinity amongst L2 residues and the change is from negative E to neutral Q with comparable volume. The loss in affinity to DNA upon mutation of either T1 or E3 (Figure 5A) confirms the previous finding of the presence of hydrogen bonding between the backbone amide of E3 and the side-chain O^{γ} of T1 in the linker[20]. It was suggested that these DNA-induced C-capping interactions provide a means whereby the ZF protein- complex, which showed flexibility in the unbound state when searching for its target DNA sequence, once the target DNA sequence is found the protein locks in place. These observations support a rationale for the conservation of the TGEKP linker sequences in zinc finger proteins

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

The free energy ΔG was calculated for each mutant and the change in energy relative to the wild type ZFP (ΔΔG) was plotted versus mutant type (Figure 5). The behavior in energy change is parallel to that of ΔH indicating a minimum entropic contribution to binding.

The enhanced affinity of ZFP to DNA upon certain mutations is an indicator of increased linker stability which in turn enables a better fit of ZF protein to its specific site. This might have resulted from the increasing speed of nonspecific site search, thus aiding to anchor the ZF1 to its target. The more drastic the change in affinity of ZFP to its target DNA upon mutation of a certain amino acid residue, the more important this residue's role in the DNA binding process. This observation is clear in average structures (see Str-Figures-S) where finger one is shifted closer to the DNA groove by 3 to 4 Å. in both E58Q and T56Y mutants (see also Figures in supplements for all mutants Str-Figures-S). On the other hand, ZF1 in the mutant K31N shifted away from DNA (Str-Figures-S), this is in agreement with the reduced affinity of K31N to DNA (Figure 6). This observation proves that there is a relation between change in affinity of the mutant to DNA and finger one and finger three.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

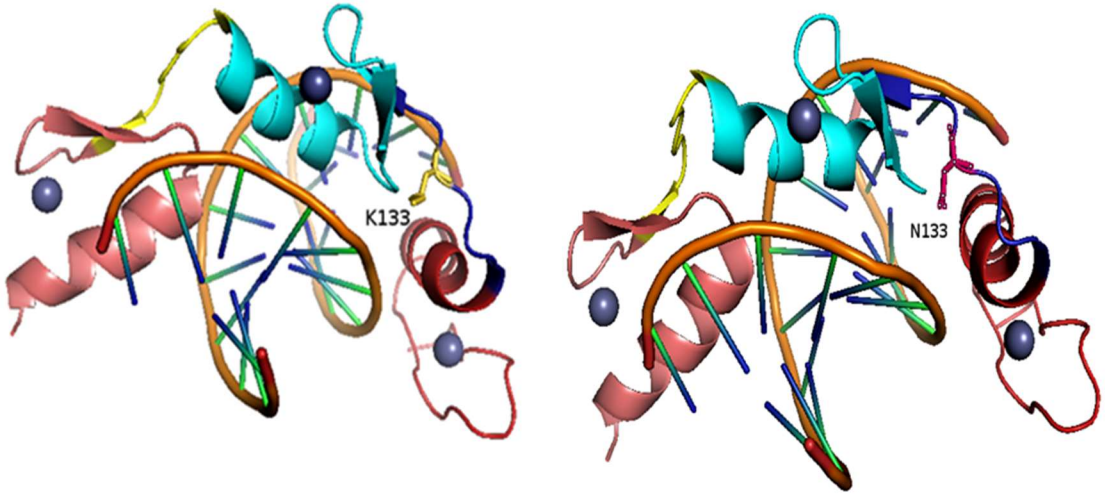


Figure 6: Top: The E58Q mutant in L2 (left) which showed a loss in affinity to DNA-ZFP complex (right), F2 shifted away from DNA by 5Å.

Bottom: Structures of The K31N mutant (left) which showed a loss in affinity to DNA relative to the wild type ZFP complex (right).

The effect of mutating amino acids on binding (Figures 5, Table 1-S) varies depending on the position of mutations (linker1 or linker 2), along with the position of the mutant in the linker 1 to 5). The optimum affinity enhancement took place when T56 was mutated to Y and G57 to V in linker 2 (L2) and to less extent for Q30E mutant in linker 1 (L1). It can be said that the largest affinity changes took place for positions 1 and 2 in linker 2 and position 3 in linker1. This behavior is a consequence of the change in linker flexibility and the ability of linkers to twist the fingers in an optimum position for DNA binding. This also reflects on DNA contacts. Mutating the same position with more than one amino acid by replacing a single amino acid with a different residue one at a time, i.e. T56Y, T56I and T56A. The affinity, as measured by ΔH, showed variations depending on the type of the incoming residue. For example, the affinities of the T56Y, T56I and T56A increased by 36%, 11% and 8% respectively, while it increased by 9% for G29P and by 11% for G29R. For K31D the affinity decreased by 5% while in K31N remained the same as for

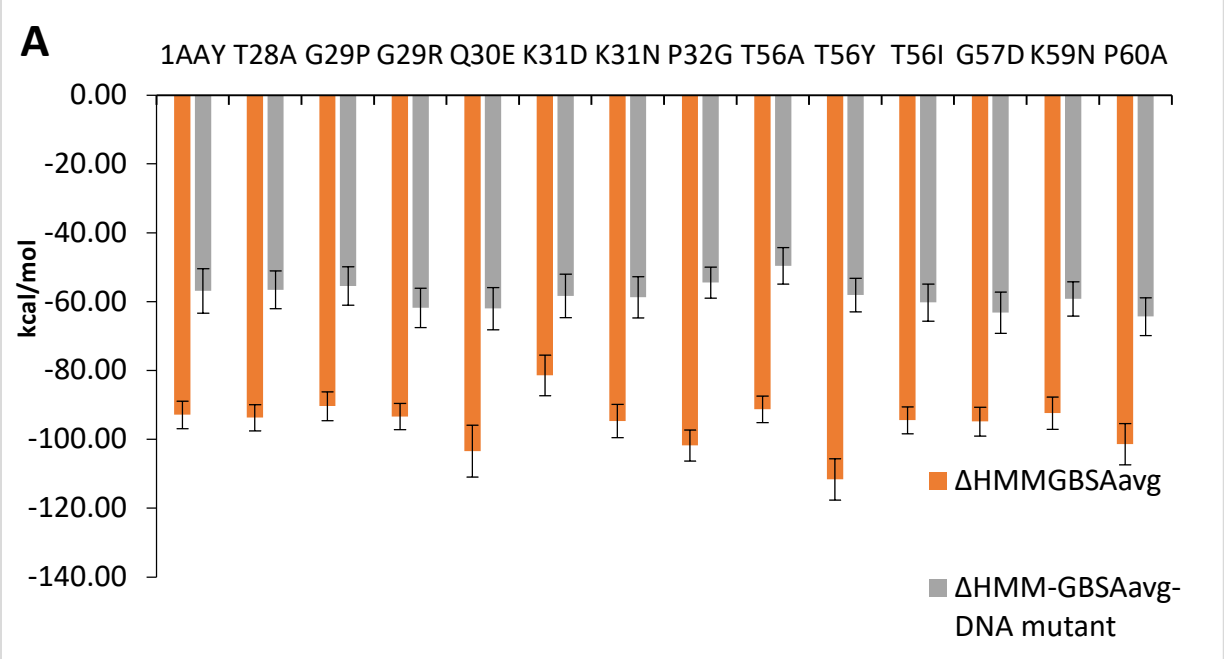
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

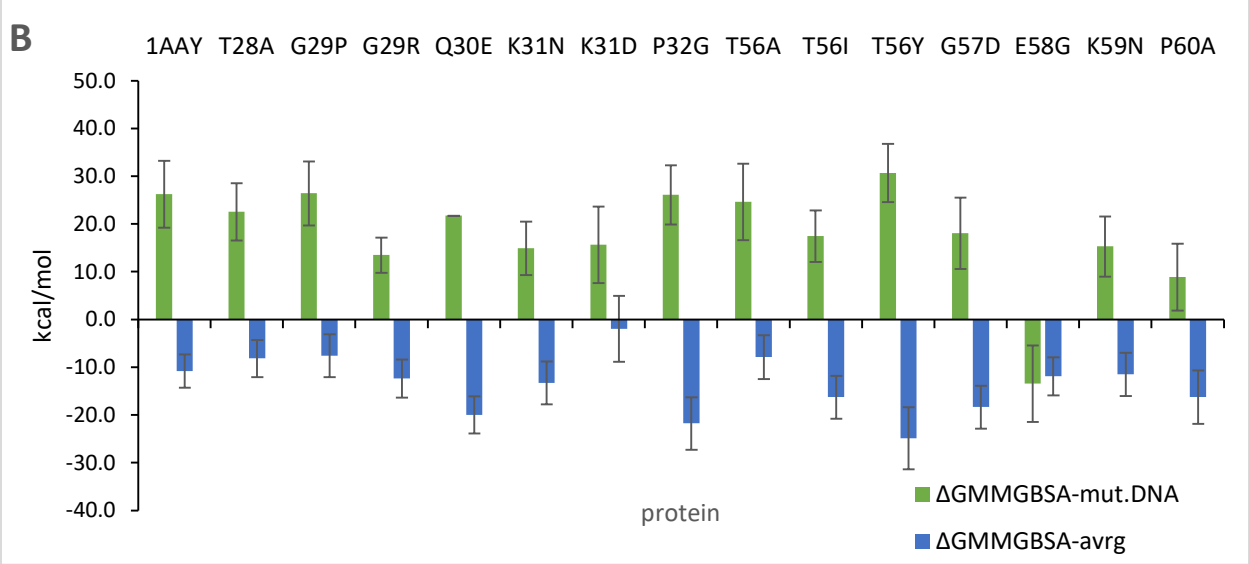
WT ZFP within the error margin. Inspecting the changes which took place in both ΔH and ΔG in Figures 6 and 7, Tables 1-S, 2-S and 3-S. The binding energy values correlate well with the total electrostatic energy (correlation coefficient of 0.86). Optimum stabilization upon mutation took place after mutation of residues 1 and 2 in linker 2. Hydrophobicity of the incoming residue along with the large vdW radius contributed to the enhancement of binding to the specific DNA site. Mutating residue 3 in linker 1 from neutral to a negative with comparable vdW volume enhanced the binding. Mutations in positions 4 in linker 1and position 3 in linker 2 destabilized the binding of the protein to its specific DNA target.

The average structures for the three-finger protein show interesting conformational changes in the "TGQKP" linker region between finger 1 and finger 2, as illustrated in Figure 6. In both proteins, the large conformational change of the linker results in close contact between the helix of finger 1 and the sheet region of finger 2. The polar interaction between R of finger 1 and Q30 is regarded as a key interaction responsible for this change. The second linker connecting finger 2 and finger 3 of the three-finger protein does not show any significant conformational change. This difference between the two linkers can be understood from the fact that finger 3 has alanine, A, at the position equivalent to Q30, and close interaction with arginine is not possible. From these observations, it is suggested that a mutation of Q30 may affect the binding affinity of the complex by changing the stability of the unbound zinc finger protein.

3.3 Binding of zinc finger protein and its mutants to a nonspecific DNA target 5'-GCA GAT TCC-3': The WT ZF protein and its mutants showed lower affinity to the nonspecific DNA target 5'-GCA GAT TCC-3' as measured by binding energies shown in figures 7. This binding to nonspecific target was studied experimentally to uncover the role of each finger in the process supported by the presence of linkers in the search mode and the conformational switching which takes place on the way to achieve recognition mode, the lower affinity was reported as an evidence of search mode [14][15].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1





Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

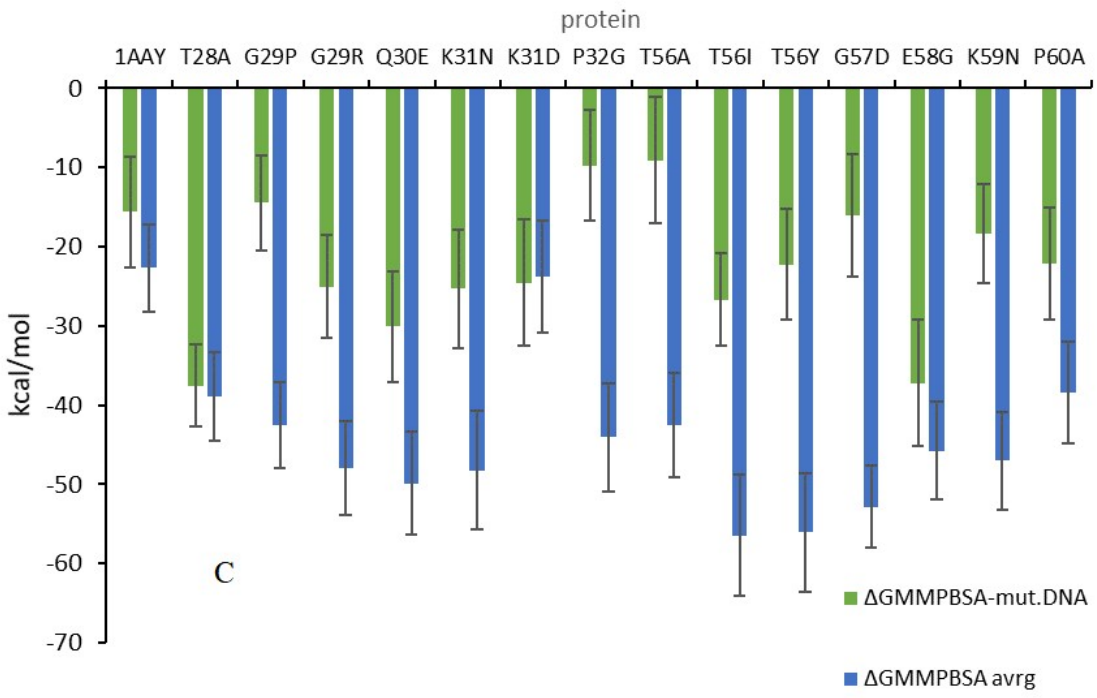


Figure 7A: Binding Energies (ΔH) values for single residue mutants of linker 1 and linker 2 in ZFP to its specific DNA target 5'-GCG TGG GCG- 3'calculated by MM_ GBSA (brown). Binding Energies (ΔH) for the same mutants bound to a nonspecific DNA sequence 5'—GCA GAT TCC—3' (grey). [for values see Table 3-S]

Figure 7B: Free energies of binding (ΔG) values for single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3'calculated by MM_ GBSA (Blue). ΔG values for the same mutants bound to a nonspecific DNA sequence 5'—GCA GAT TCC—3' (Green).

Figure 7C: Free energies of binding (ΔG) values for single residue mutants of linker 1 and linker 2 in ZFP bound to its specific DNA target 5'-GCG TGG GCG- 3'calculated by MM_ PBSA (Blue). ΔG values for the same mutants bound to a nonspecific DNA sequence 5'—GCA GAT TCC—3' (Green).

In search mode, the positively charged ZF2 and ZF3 are bound to DNA while ZF1 remains dissociated from nonspecific DNA target. Indeed, the nmr study revealed the dynamic nature of the binding process expressed as a shift between search and recognition modes. The study of WT zinc finger protein and T23/Q32E linker mutants to the non-specific DNA target revealed that ZF1 is most mobile amongst the three fingers in the search process before binding the target[15]. T28K and Q30E mutants showed

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

higher energy barrier than the WT protein resulting in a kinetically slower binding process, i.e, process shifts towards search[15]. The loss in binding to non-specific DNA target showed that ZF1 dissociates and causes the loss in affinity in linker mutants, this change in binding is attributed to changes in the composition and distribution of charged residues in the binding interface.

Values of the binding energies of the wild type zinc finger protein and its single point mutants to DNA were calculated, see plots in Figures 7 (numerical values for energies are listed in Tables 1-S, 2-S, 3-S in supporting documents). The WT ZFP and its mutants showed lower affinity to the non-specific DNA target compared to that of the specific target, the loss in affinity represented by ΔH calculated by MM/GBSA and PBSA. It should be noted that the calculated energy values for WT ZFP are consistent with the reported values obtained using MM_PBSA [18][18][58]. As anticipated; binding of the linker mutants to a nonspecific DNA target with the sequence 5'-GCAGATTCC-3'resulted in a loss in binding energy, a similar loss in binding affinity was reported by Chen et al for T28 and Q30 mutants by measuring Kd for protein-DNA complex [14][15][17][44]. The extent to which binding is lost or enhanced upon linker mutation varies depending on the type and position of residue replaced, as well as the incoming residue (Table 5-S in supplements). The change in ZFP affinity to DNA depends on the composition and distribution of charged residues in the binding interface, i.e., the type of amino acids in linkers plays an important role in directing the fingers to proper binding positions in the DNA groove.

Upon binding to nonspecific DNA target, the balance between ZF1-ZF2-ZF3 interactions and ZFP-DNA affects the binding process, upon weakening of finger interactions the ZF1 binding is lost and moves away from the DNA, this is clear in the P32G mutant (see Supplements F-S Figure 9B.

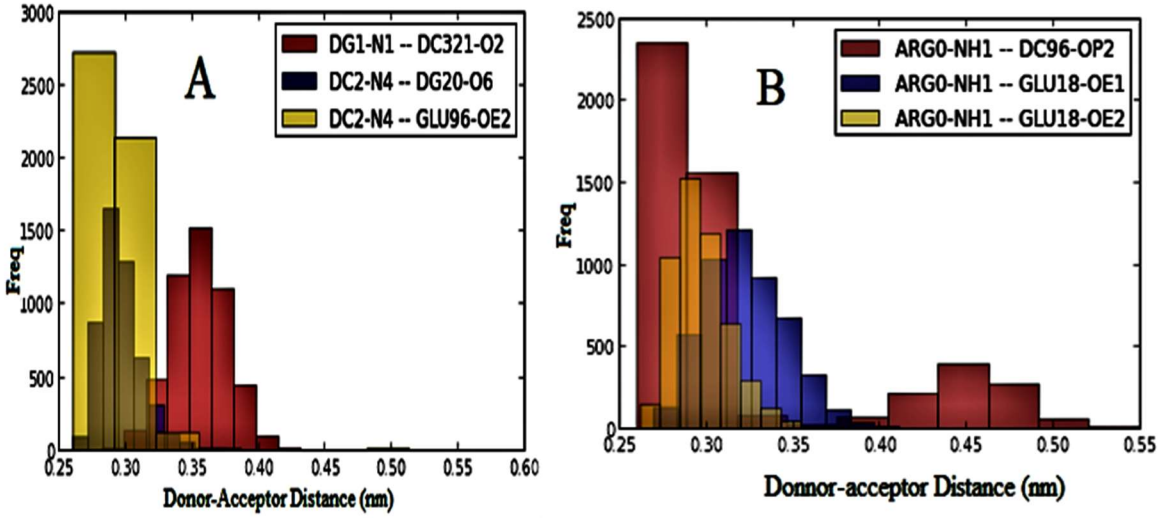
The positive ΔG values in Figure 7 for the binding of WT ZFP and its mutants to the nonspecific DNA target 5'-GCA GAT TCC-3' confirms the specificity of WT ZFP and its mutants to the specific target for zinc finger protein (5'-GCG TGG GCG- 3'). This observation also confirms the recognition process to a specific target. . Zandarashvili *et al* explained this behavior as due to changes in electrostatic affinity of ZF1 to DNA[14][15]. This nonspecific binding explains the ZFP binding mechanism in which the domain with higher affinity to the DNA target has an anchoring role, while the domain with lower affinity, i.e. ZF3 takes the role

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

of an explorer in the DNA binding process. Indeed an experimental study of WT ZFP and its mutants to specific and nonspecific 12-bp DNA targets showed similar behavior[14]. K_d value for WT ZFP specific binding was 0.5mM and for nonspecific binding 0.2 µM , this was interpreted as ZF2 and ZF3 remain bound to nonspecific DNA and ZF1 dissociated from the target, T23K/Q32E mytants gave K_d value for specific binding 2.5mM and for nonspecific binding 0.9 µM, this was also interpreted as a shift in binding mechanism to the recognition mode. All mutants T23K/Q32E/E60Q/K79T gave K_d specific values of 0.1mM, and K_d nonspecific of 0.15 µM while the mutant E60Q/K79T gave K_d specific = 0.1mM and K_d nonspecific 0.1 µM both were interpreted as a shift towards the search mode[14].

3.4 Hydrogen bonding:

The hydrogen bonds between ZF protein and DNA were discussed in previous report[8]. The method used here is ***Baker Hubbard*** which identifies hydrogen bonds based on cutoffs for the Donor-acceptor distance and the angles. The criterion employed is θ>120º and **r** H-acceptor <2.5Å in at least 10% of the trajectory. The return value is a list of the indices of the atoms (donor, H, acceptor) that satisfy this criteria, an example of the plot of frequency of hydrogen bonds variation with donor-acceptor distance is shown in Figure 8.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

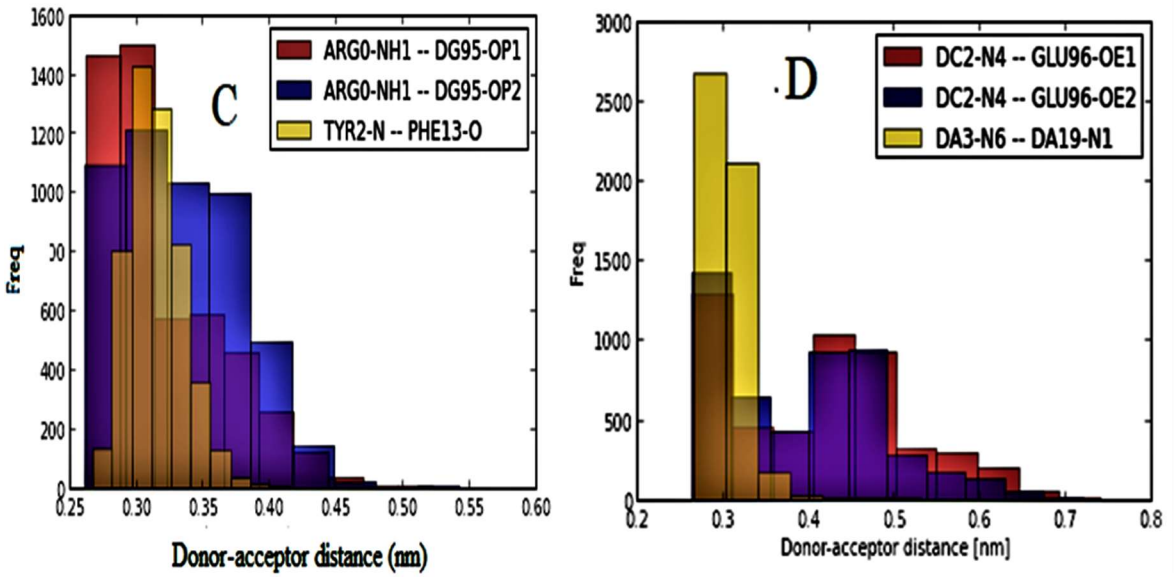


Figure 8: The hydrogen bonds based on cutoffs for the Donor-H-acceptor distance and angle calculated by Baker-Hubbard method (8-S in supplements). Plots show the variation in H-bond frequency with distance in nm. The effect of mutation can be seen in the differences in hydrogen bond frequencies between plots: A) WT zinc finger protein bound to specific DNA sequence; B) E58G mutant bound to the specific DNA sequence; C) E58G mutant bound to non-specific sequence D) K31N mutant bound to specific DNA sequence. For detailed listing of hydrogen bonds, see supplements Table 8-S

For detailed listing of hydrogen bonds of amino acid residues in ZF protein with DNA bases and each other see list of hydrogen bonds 8-S in supplements. WT zinc finger protein showed 6 specific hydrogen bonds where the DNA acts as donor and amino acids act as acceptors (C2-Glu92, two A12-Asp39 contacts, C15-Asp67 and two A18-Asp95 contacts**).** Mutants E58G and K31N showed 3 and 2 hydrogen bonds respectively (C base to Glu). Arginine plays the major role in H-bonding in the wild type protein and its mutants with variation in the number and type of contacts. R_{-1} in ZF 1 shows four specific contacts with G in both WT and K31N mutant while a new R-G(N) appeared in E58G mutant [8]. R-1 in ZF2 in the WT shows 2 specific contacts with G, these contacts changed to 3 nonspecific with phosphates in linker mutants. In K31N mutant, one specific contact was preserved. Residue S1 of ZF2 showed nonspecific contact with G and another with C in the wild type disappeared in mutant proteins (see Table 3)[7][8]. R_{-1} of ZF3 gives one nonspecific contact with G in WT protein, this contact (R_{-1} from ZF3 bond to G) has changed in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

Table 3: The changes in hydrogen bond frequency and length of ZFP and its mutants upon binding to DNA specific and nonspecific DNA targets

protein	DNA target	Highest frequency Hydrogen bond	Intermediate frequency Hydrogen bonds	Lowest frequency Hydrogen bonds	ΔH, AG Kcal/mol
Wild type ZFP	specific	C2(N4)—Glu96(O) 0.2-0.3 nm	C2(N4)—G20(O) 0.3 nm	G1(N1)—C321(O) 0.35 NM	-92.5(4), -10.8(3)
Wild type ZFP	nonspecific	Arg (0)N2-Glu18(OE1) 0.27-0.35 nm	Arg(0)NH2-C96(OP2) 0.27-0.32nm	Arg(0)NH2-G95(OP2) 0.35 nm	-56.9(6), 25.5(6)
E58G	specific	Arg(0)NH1-C96(OP2) 0.25-0.35 nm	Arg(0)NH1-Glu18(OE2) 0.3nm	Arg(0)NH1-Glu18(OE1) 0.35nm	-86.6(5), -12 (6)
E58G	nonspecific	Arg(0)NH1-G95(OP1) 0.25-0.35 nm	Arg(0)NH1-G95(OP2) 0.25-0.35 nm	TyR2(N)-PHE13-O 0.3 nm	-90.5, -14
K31N	specific	A3(N6)-A19(N1) 0.3 nm	C2(N4)- Glu(96) OE1 0.45 nm	C2(N4)-Glu(96) OE2 0.45 nm	-77.9(5) -11(6)

K31N mutant to a specific contact with G(O). The nonspecific contacts of histidine and lysine were preserved in both the WT and mutant proteins. The highest frequency of short H-bonds in the WT protein was observed for C2 as donor to Glu as acceptor (Figure 8 and Table 3). While in both E58G and K31N mutants, this high frequency of hydrogen bonds shifted to Arginine nonspecific contacts with C and G Bases (see Figure 8B and C) see 8-S in supplements for hydrogen bond listing.

Direct recognition is due mainly to hydrogen bonding between residues and DNA bases and the indirect recognition is due to hydrogen bonding to the back bone phosphates, other contributions to binding like electrostatic attractions, water mediated hydrogen bonds, hydrophobic forces and ion release in addition to DNA deformation. Figure 9 shows how the hydrogen bonding between ZF protein and a nonspecific DNA target varies with simulation time which implies a relationship between establishing the equilibrium position and H-bonding.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 November 2020 doi:10.20944/preprints202011.0524.v1

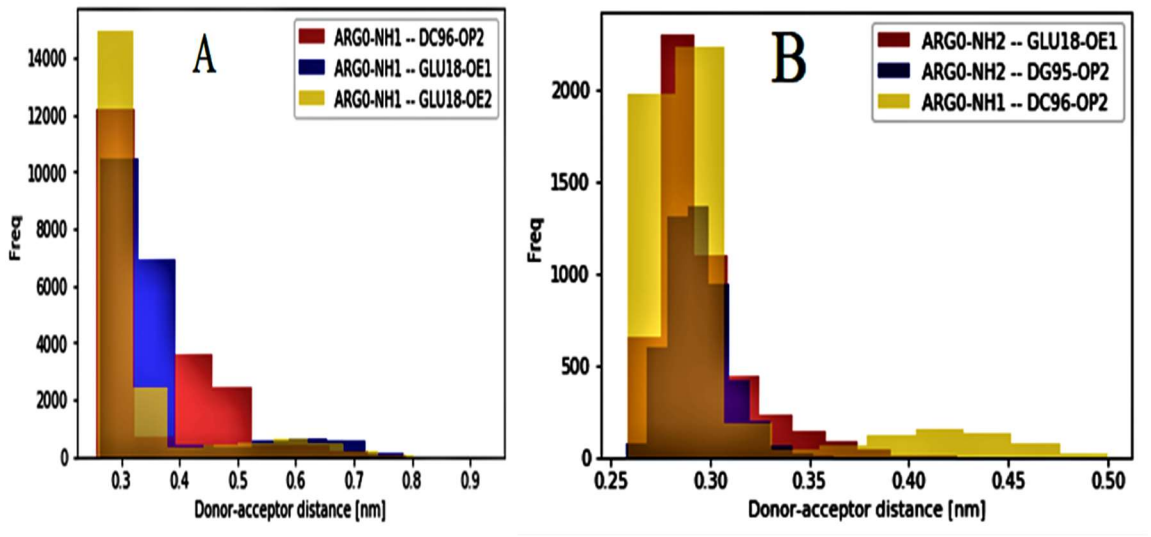


Figure 9: Nonspecific Binding: Variation of the H-bond (length and frequency) for ZF binding with mutant-DNA with simulation time. A) after 62 ns the bonds were (Arg22-C95 phosphate, two Arg22- Glu18 bonds) B) H bonds after 102 ns (one Arg22-Glu18, two Arg22 bonds with phosphates of C95 and C96. The shifts in positions on the x axis shows how bonds change length with simulation time.

High frequency and short distance of C(2) –Glu 96 hydrogen bond in ZF -specific DNA complex (Figure 9(A)) to that in ZF-nonspecific DNA complex, a loss in bonding is observed and the bonding was replaced with several Arg –backbone phosphate bonds in the nonspecific DNA complex

Concluding remarks:

In previous reports[7] [8], we have established that the affinity of ZFP to a specific DNA target is larger than the sum of affinities of individual fingers (no linkers involved). This difference was attributed to protein structure including linkers. Also we reported that a ZFP consisting of F1F2 which lacks F3 is destabilized more than an F2F3 protein which lacksF1, also mutating F1 had a maximum effect on reducing the ZFP affinity to DNA amongst the three fingers, this finding gives a special role to ZF1 and consequently L1, this is followed in importance by the role of ZF3.

 doi:10.20944/preprints202011.0524.v1

The affinity of ZF protein to its specific DNA target showed sensitivity to linker mutations. The binding energy of the ZFP to DNA target results from complex interactions in the major and minor groves, i.e. hydrogen bonding of amino acid side chains in fingers with DNA bases (specific) and to backbone phosphates (nonspecific). ZFP structure has a direct effect on binding to DNA, the finger-finger and finger-DNA interactions are important factors contributing to hydrogen bonding and hence binding energy (affinity). Mutations in the protein sequence and measuring the affinity to specific and nonspecific DNA targets give insight on the mechanism of binding, the protein DNA binding process starts with scanning the DNA in search of the specific sight (low affinity for nonspecific binding), then upon recognition of the target the specific binding takes place (ZF1 anchoring and both ZF2 and ZF3 bound, high affinity for specific binding). The affinity of WT ZFP and its mutants to nonspecific DNA target is reduced indicating a shift to search mode[14].

Binding of ZFP to both specific and nonspecific DNA targets showed sensitivity to mutations in linkers between ZF1, ZF2 and ZF3, this finding confirms the role of linkers in tuning the zinc fingers to an optimum position[23]. The largest reduction in affinity to specific DNA target was for K31D (residue 4 in L1) and E58Q (residue 3 in L2). The reduction in affinity is related to the ZF1 search and recognition process. These results give more insights on the effect of energy of ZF protein DNA binding upon mutation of both protein Fingers and the DNA target [7][8].

References

1. Miller J, McLachlan AD, Klug A, et al (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 4:1609
2. Krishna SS, Majumdar I, Grishin N V (2003) Structural classification of zinc fingers SURVEY AND SUMMARY. Nucleic Acids Res 31:532–550
3. Jacobs GH (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J 11:4507
4. Pellegrino GR, Berg JM (1991) Identification and characterization of" zinc-finger" domains by the polymerase chain reaction. Proc Natl Acad Sci 88:671–675
5. Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: New insights into structural and functional diversity. Curr Opin Struct Biol 11:39–46 . doi: 10.1016/S0959-440X(00)00167-6
6. Elrod-Erickson M, Rould M a, Nekludova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. Structure 4:1171–80

7. Hamed MY, Arya G (2016) Zinc Finger Protein Binding to DNA: An Energy Perspective Using Molecular Dynamics Simulation and Free Energy Calculations on Mutants of both Zinc Finger Domains and their Specific DNA bases. J Biomol Struct Dyn 34:919–934 . doi: 10.1080/07391102.2015.1068224

8. Hamed MYMY (2018) Role of protein structure and the role of individual fingers in zinc finger protein–DNA recognition: a molecular dynamics simulation study and free energy calculations. J Comput Aided Mol Des 32:657–669 . doi: 10.1007/s10822-018-0119-9

9. Nardelli J, Gibson T, Charnay P (1992) Zinc finger-DNA recognition: analysis of base specificity by site-directed mutagenesis. Nucleic Acids Res 20:4137–44

10. Pabo CO, Sauer RT (1984) Protein-DNA recognition. Annu Rev Biochem 53:293–321 . doi: 10.1146/annurev.bi.53.070184.001453

11. Temiz AN, Camacho CJ (2009) Experimentally based contact energies decode interactions responsible for protein-DNA affinity and the role of molecular waters at the binding interface. Nucleic Acids Res 37:4076–4088 . doi: 10.1093/nar/gkp289

12. Hamilton TB, Borel F, Romaniuk PJ (1998) Comparison of the DNA binding characteristics of the related zinc finger proteins WT1 and EGR1. Biochemistry 37:2051–2058 . doi: 10.1021/bi9717993\rbi9717993 [pii]

13. Elrod-Erickson M, Benson TE, Pabo CO (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. Structure 6:451–64

14. Chen C, Esadze A, Zandarashvili L, et al (2015) Dynamic equilibria of short-range electrostatic interactions at molecular interfaces of protein-DNA complexes. J Phys Chem Lett 6:2733–2737 . doi: 10.1021/acs.jpclett.5b01134

15. Zandarashvili L, Vuzman D, Esadze a., et al (2012) PNAS Plus: Asymmetrical roles of zinc fingers in dynamic DNA-scanning process by the inducible transcription factor Egr-1. Proc Natl Acad Sci 109: . doi: 10.1073/pnas.1121500109

16. Foster MP, Wuttke DS, Radhakrishnan I, et al (1997) Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. Nat Struct Biol 4:605–608

17. Ryan RF, Darby MK (1998) The role of zinc finger linkers in p43 and TFIIIA binding to 5S rRNA and DNA. Nucleic Acids Res 26:703–709 . doi: 10.1093/nar/26.3.703

18. Lee J, Kim J-S, Seok C (2010) Cooperativity and specificity of Cys2His2 zinc finger protein-DNA interactions: a molecular dynamics simulation study. J Phys Chem B 114:7662–71 . doi: 10.1021/jp1017289

19. Wolfe S a, Ramm EI, Pabo CO (2000) Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. Structure 8:739–50

20. Laity JH, Dyson HJ, Wright PE (2000) DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys 2-His 2 zinc fingers. J Mol Biol 295:719–727

21. Choo Y, Klug A (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. Nucleic Acids Res 21:3341–3346 . doi: 10.1093/nar/21.15.3341

22. Nagaoka M, Nomura W, Shiraishi Y, Sugiura Y (2001) Significant effect of linker sequence on DNA recognition by multi-zinc finger protein. Biochem Biophys Res Commun 282:1001–1007 . doi: 10.1006/bbrc.2001.4672

23. Peisach E, Pabo CO (2003) Constraints for Zinc Finger Linker Design as Inferred from X-ray Crystal Structure of Tandem Zif268–DNA Complexes. J Mol Biol 330:1–7 . doi: 10.1016/S0022-2836(03)00572-2

24. Wu H, Yang WP, Barbas CF (1995) Building zinc fingers by selection: toward a therapeutic application. Proc Natl Acad Sci U S A 92:344–348 . doi: 10.1073/pnas.92.2.344

25. Kim J-S, Pabo CO (1998) Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. Proc Natl Acad Sci 95:2812–2817

26. Smith NC, Matthews JM (2016) Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. Curr Opin Struct Biol 38:68–74

27. Liu Q, Segal DJ, Ghiara JB, Barbas CF (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc Natl Acad Sci U S A 94:5525–5530 . doi: 10.1073/pnas.94.11.5525

28. Anand P, Schug A, Wenzel W (2013) Structure based design of protein linkers for zinc finger nuclease. FEBS Lett 587:3231–5 . doi: 10.1016/j.febslet.2013.08.015

29. Laity JH, Chung J, Dyson HJ, Wright PE (2000) Alternative splicing of Wilms' tumor suppressor protein modulates DNA binding activity through isoform-specific DNA-induced conformational changes. Biochemistry 39:5341–5348

30. Laity JH, Dyson HJ, Wright PE (2000) Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. Proc Natl Acad Sci 97:11932–11935

31. Gohlke H, Case D a (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. J Comput Chem 25:238–50 . doi: 10.1002/jcc.10379

32. Hayes JM, Archontis G (2012) MM-GB (PB) SA calculations of protein-ligand binding free energies. INTECH Open Access Publisher

33. Chen F, Liu H, Sun H, et al (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein--protein binding free energies and re-rank binding poses generated by protein--protein docking. Phys Chem Chem Phys 18:22129–22139 . doi: 10.1039/C6CP03670H

34. Case DA, Ben-Shalom IY, Brozell SR, et al AMBER 2018; 2018. Univ California, San Fr

35. D. A. Case, D. A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang WSR, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vencent M, Crawley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pietera, I. Massova GLS, U. C.

Singh, P.K. Weiner PAK (2006) Amber 9

36. DeLano WL (2002) The PyMOL Molecular Graphics System

37. Peters MB, Yang Y, Wang B, et al (2010) Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). J Chem Theory Comput 6:2935–2947

38. D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham S, DeBolt, I.I.I. D. Ferguson, G. Seibel PK (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculation to simulate the structural and energetic properties of molecules. 1–41

39. Nurisso A, Daina A, Walker RC (2012) A practical introduction to molecular dynamics simulations: applications to homology modeling. In: Homology Modeling. Springer, pp 137–173

40. Wolfe SA, Nekludova L, Pabo CO (2000) Dna r c z f p. 183–212

41. Raval A, Piana S, Eastwood MP, et al (2012) Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins Struct Funct Bioinforma 80:2071–2079

42. Christy B, Nathans D (1989) DNA binding site of the growth factor-inducible protein Zif268. Proc Natl Acad Sci 86:8737–8741

43. Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science (80-) 252:809–817

44. Wilson TE, Day ML, Pexton T, et al (1992) In vivo mutational analysis of the NGFI-A zinc fingers. J Biol Chem 267:3718–3724

45. Shao J, Tanner SW, Thompson N, Cheatham TE (2007) Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. J Chem Theory Comput 3:2312–2334

46. Roe DR, Bergonzo C, Cheatham III TE (2014) Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. J Phys Chem B 118:3543–3552

47. McGibbon RT, Beauchamp KA, Harrigan MP, et al (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J 109:1528–1532 . doi: 10.1016/j.bpj.2015.08.015

48. Rother K (2005) Introduction to PyMOL. Methods Mol Biol Clift Nj 635:0–32 . doi: 10.1213/ANE.0b013e3181e9c3f3

49. Hou T, Wang J, Li Y, et al (2010) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J Chem Inf Model 51:69–82 . doi: 10.1021/ci100275a

50. Genheden S, Kuhn O, Mikulskis P, Ryde U (2012) The Normal-Mode Entropy in the MM/GBSA Method: E ff ect of System Truncation, Bu ff er Region, and Dielectric

Constant. J Chem Inf Model

51. Miller III BR, McGee Jr TD, Swails JM, et al (2012) MMPBSA. py: an efficient program for end-state free energy calculations. J Chem Theory Comput 8:3314–3321 . doi: 10.1021/ct300418h

52. Galindo-Murillo R, Roe DR, Cheatham TE (2015) Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). Biochim Biophys Acta - Gen Subj 1850:1041–1058 . doi: 10.1016/j.bbagen.2014.09.007

53. Kuhn B, Gerber P, Schulz-Gasch T, Stahl M (2005) Validation and use of the MM-PBSA approach for drug discovery. J Med Chem 48:4040–8 . doi: 10.1021/jm049081q

54. Wang J, Morin P, Wang W, Kollman P a (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 123:3986–3994

55. Brown RS (2005) Zinc finger proteins: Getting a grip on RNA. Curr Opin Struct Biol 15:94–98 . doi: 10.1016/j.sbi.2005.01.006

56. Wyszko E, Radłowski M, Bartkowiak S, Barciszewska MZ (1997) Maize TF IIIA--the first transcription factor IIIA from monocotyledons. Purification and properties. Acta Biochim Pol 44:579–589

57. Dreier B, Segal DJ, Barbas CF (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. J Mol Biol 303:489–502 . doi: 10.1006/jmbi.2000.4133

58. Spyrakis F, Cozzini P, Bertoli C, et al (2007) Energetics of the protein-DNA-water interaction. BMC Struct Biol 7:4 . doi: 10.1186/1472-6807-7-4