Computational study of the effect of protonation states of PSA protein zinc fingers on its DNA binding

Nguyen Viet Duc, Ly Hai Nguyen, Hien T. T. Lai, Toan T. Nguyen
VNU Key Laboratory for Multiscale Simulation of Complex Systems, VNU University of Science, Vietnam National University, 334 Nguyen Trai street, Thanh Xuan, Hanoi, 11416, VIETNAM
E-mail: toannt@hus.edu.vn

Abstract. In this study, we investigate the binding of the Zinc finger (ZF) structure on a short DNA molecule. The zinc finger of a protein where a Zn$^{2+}$ ion binds to 4 cysteine or histidine amino acids in a tetrahedral structure is a very common motif of nucleic acid binding proteins. This structure is ubiquitous and the corresponding interaction model is present in 3% of the genes of human genome. ZF has been shown to be extremely useful in various therapeutic and research capacities, as well as in biotechnology. A recent computational study has shown that isolated zinc finger structure is stable if the cysteine amino acids are in deprotonated state. Here, we investigate how this deprotonated state influences protein structure, dynamics, and function in binding of ZF to short DNA molecules using molecular dynamics simulations in sub-microsecond range. Our results show that the Zn$^{2+}$ ion and the deprotonated state of cysteine is essential for mechanical stabilization of the functional, folded conformation. Not only this state stabilizes the ZF structure, it also stabilizes the DNA-binding structure. Our result has potential impact on better design of zinc fingers for various biotechnological applications.

1. Introduction
Zinc finger proteins are among the most abundant proteins in eukaryotic genomes. Their functions are extraordinarily diverse and include DNA recognition, RNA packaging, transcription activation, regulation of apoptosis, protein folding and assembly, and lipid binding. Recently, in the area of biotechnology, they are also employed for biosensor design. For example, the PSA protein is a common marker for prostate cancer [1]. It has a zinc finger structure for nucleic acid binding. Therefore, one can detect the PSA presence in a sample by using a substrate that is functionalized with aptamers, short DNA molecules that only PSA protein can recognize [2–6]. Upon binding of PSA protein to the aptamers, the electrochemical properties of the substrate will change and can be measured accurately using the companion electric circuit. The strength of the perturbation is a measure of the PSA concentration in the sample. Thus one can detect and measure rather accurate the PSA concentration allowing for early detection of prostate cancer.

Zinc finger structures are as diverse as their functions. However, they all shares the same motif of a short α-helix, two β-strands and a loop (see Fig. 1). The amino acid residues of this protein segment arrange in three-dimensional space such that the zinc ions would coordinate with 4 or 6 residues, Cys2His2, Cys4 or Cys6, to maintain the rigidity of the structure. The helix group then would bind to the major groove of the DNA double helix. The rest of the residues...
form hydrogen bonds to appropriate nucleic acid residues. The amino acid sequence of the zinc finger would bind to specific DNA nucleic acid sequence. It is this genome specificity that makes zinc finger, either natural or artificial, a very promising molecules for biotechnological application for gene therapy or recognition. In this work, we focus on the binding of the aptamer to PSA zinc finger structure used in the prostate cancer detection mentioned earlier. This is a Zn$^{2+}$–Cys4 structure. Its atomic structure has been resolved using X-ray crystallography method [7]. More specifically, we investigate the influence of the deprotonated state of cysteine amino acids on stability as well as nucleic acid binding of the protein. The previous study of isolated zinc finger structure has shown that the deprotonated state of cysteine is more stable compared to natural state [8]. Here, we show that this is also the case in the presence of a DNA molecule that binds to it. This suggests that the deprotonated state of cysteine is essential for this binding mechanism.

This paper is organized as follows. After the introduction in Section 1, the detail of the computational procedure is presented in Section 2. In Section 3, the results are presented and discussed. We conclude in Section 4.

2. Computational details

2.1. Simulation systems preparation

The structure of the PSA protein zinc fingers and the DNA segment it binds to is downloaded from the Protein Data Bank (https://www.rcsb.org/), PDB code 1R4I. This structure was resolved using X-ray crystallography method with a resolution of 3.1Å [7]. The complex contains a DNA segment and two proteins segments called chain A and chain B, and four Zinc ions. On each protein chain, the four Cys542, Cys545, Cys559, and Cys562 amino acids bind to the first Zinc ion (Zn$_1$) and the four Cys578, Cys584, Cys594, Cys597 amino acids bind to the second Zinc ion (Zn$_2$) in a tetrahedral structure (see Fig. 1). There are totally 4 zinc finger structures on this complex. To investigate the difference between the CYS-ZN complex with cysteine amino acids in their natural state and the CYN-ZN complex with cysteine amino acids in deprotonated state (CYN is the short name for deprotonated cysteine), we manually remove the hydrogen from the thiol group of those 16 zinc-binding cysteine amino acids (all other cysteine residues in the protein remain in its natural state).
2.2. Molecular dynamics (MD) simulation
The periodic boundary condition is used in our simulation. After setting up the coordinates of the atoms, the periodic simulation box size is chosen such that the protein and DNA on neighboring image boxes are at least 3nm apart. This is significantly larger than the screening length of the solution (about 1nm). This is large enough to eliminate the finite size effect due to the long-range electrostatic interactions, yet maintain a small enough system to have the simulation run in reasonable time with the available computing resource. The systems are then solvated with water molecules in an explicit solvent simulation. After solvation, Na\(^+\) and Cl\(^-\) ions are added to the system at the physiological concentration of 150mM by randomly replace water molecules by ions. The total charge of the system is zero to maintain the neutrality. The specific number of molecular for each system is listed in Table 1. The systems are then subjected to an energy minimization procedure using a steepest descent method to remove potentially high energy contact and overlapping atoms before doing molecular dynamic simulation.

All-atom molecular dynamics simulation with the explicit solvent model is carried out in this work. The forcefield AMBER 99-ILDN [9] is used to parameterize the protein molecules. The newest state of the art, PARMBSC1 [10] is used to parameterize the DNA molecule. Water molecules are parameterized using the TIP3P forcefield [11], a common and highly compatible forcefield for the chosen protein and DNA forcefields. The GROMACS version 2016.3 software package [12] is used for molecular dynamics simulation of the systems. Each system is subjected to equilibration in NVT and NPT ensembles at temperature 298K and pressure of 1 atm for 5ns. After that, a long production run of 500ns is used for taking statistics. The Nose-Hoover thermostat [13,14] is used to maintain the temperature of the systems. The Parrinello-Rahman barostat [15,16] is used to maintain the pressure of systems. Both electrostatics and van de Waals interactions are cut off at 1.2nm. The long-range part of the electrostatic interactions among charges is calculated in the \(k\)-space using the Ewald summation via Particle Mesh Ewald method [17] at fourth order interpolation. The long-range part of the van de Waals interactions among atoms is approximated as appropriate corrections to the energy and pressure. All the covalent bonds are constrained using the LINCS (Linear Constraint Solver) algorithm in order to increase the simulation time step to 2 fs [18].

### Table 1. the molecules simulated for each systems.

| System name      | CYN-ZN | CYS-ZN |
|------------------|--------|--------|
|                  | # residues | # atoms | # residues | # atoms |
| H\(_2\)O         | 79326   | 79383   |
| Na\(^+\)         | 103     | 87      |
| Cl\(^-\)         | 77      | 77      |
| Protein chain A  | 73      | 1120    | 73         | 1124    |
| Protein chain B  | 70      | 1079    | 70         | 1083    |
| DNA              | 18      | 1142    | 18         | 1142    |

2.3. Analysis the results of MD simulation
Analysis of the simulation results is performed using the corresponding tools provided within the GROMACS package, such as the root mean square deviation (RMSD) for backbone as well as the root mean square fluctuation (RMSF) for C\(_\alpha\) for both protein and DNA on each
3. Results and discussions

3.1. Deviation of complexes

First, we investigate the overall stability of our complex by measuring the deviation of the protein as well as the DNA from its initial PDB structure. The backbone RMSD value for the wild type protein structure CYS-ZN and the deprotonated CYN-ZN are calculated as a function of time simulated. This value is an important parameter to analyze the equilibration of MD trajectories as well as its stability. The RMSD of the backbone atoms relative to the corresponding starting structure is plotted in Fig. 2. In Figs. 2(a) and 2(b), the RMSDs of protein chain A and B of the CYS-ZN system (purple line) clearly deviate more than those of the CYN-ZN complex (green line). The structure of protein chain B is more stable than the protein chain A structure on both the complexes. In the CYS-ZN complex, we can observe that the proteins in the CYN−ZN complex. The visualization of 3D structure on the systems are performed using VMD version 1.9.3 program [19]. Some in-house scripts are used for various tasks and combining different analysis software.

Figure 2. The backbone root mean square deviation (RMSD) of the two protein chains A (a) and B (b) and two complementary strands of the DNA molecules. For each subfigure, the CYN-ZN (deprotonated cysteine residues) and CYS-ZN (natural cysteine residues) are plotted for comparison. Stability of CYN−ZN system is clearly demonstrated.
system are quite stable within the sub-microsecond time simulated with the value of equilibrium deviation of about 0.2 nm for protein A and slightly less for protein B. On the other hands, the proteins in the CYS–ZN system are much less stable and seem to still in the process of moving to a different equilibrium structure.

Similar RMSD analysis of the DNA strands, DNA strand 1 from 3-end to 5-end, and DNA strand 2 (complementary strand) from 5-end to 3-end, are plotted in Figs. 2(c) and 2(d). In both systems, the DNA molecule is always stable as one would expect. Both strands show near identical RMSD as a function of time suggest they always remain in double helix binding state and move together. However, the CYS–ZN system clearly shows higher RMSD for the DNA strands with large variation with time. As we will see later that this is due to less stable DNA–binding structure for this system. Overall, our RMSD analysis shows that the CYN–ZN complex where the zinc-binding cysteine amino acids are in the deprotonated state is more stable than the CYS–ZN complex for both the protein component and the DNA component.

3.2. Fluctuation of each residues on the complexes

![Figure 3](image-url)  
**Figure 3.** The root mean square fluctuation (RMSF) of backbone Cα atoms of the protein chain A (a) and B (b) and of the two complementary strands (c) and (d) of the DNA molecule.

To investigate the stability of individual residue in the binding state of the complexes, the RMSFs are calculated for each residue. In Fig. 3(a) and 3(b), the average fluctuation value of
Co backbone atom of each amino acid residue varies between 0.08 and 0.4nm. For CYS−ZN system, four residues that bind to zinc ions show higher fluctuation than other regions (about 0.4nm). The same residues in the CYN−ZN complex show much smaller fluctuations. Once again, this confirms that deprotonated cysteine residue stabilizes zinc-finger structure even in the preference of negatively charged DNA molecule, a similar conclusion to that of the RMSD analysis.

Fig. 3(d) and 3(c) show the fluctuations of each nucleic acid in both strands in the DNA molecule of complexes. The fluctuation values of DNA strands in the CYN-ZN complex are between 0.1 and 0.4nm. Specifically, the 8th nucleic acid in strand 2 and the 13th nucleic acid in strand 1 have the fluctuation values of about 0.1nm, they are smaller in value than other residues (Fig. 3(c)). This is very easily understandable because the 8th nucleic acid in strand 2 and 13th nucleic acid in strand 1 are the main residues that bind to the two chains of protein, and the free energies for binding between DNA and protein chains on this complex are large.

The regions of binding between DNA and protein are the same on both the CYN-ZN and CYS-ZN systems, but the fluctuation of two strands in the structure of DNA on the CYS-ZN complex are higher (Fig. 3(d)). The nucleic acids of this DNA are about 0.2nm fluctuation and it is higher at the ends of the strands. Two nucleic acids that bind to protein chains have the fluctuation values like its complementary nucleic acids, so the free energies for binding between DNA and protein on this complex are smaller. From this analysis, we can see that deprotonated cysteine amino acids not only stabilize zinc finger structure, they also stabilize DNA-binding structure.

4. Conclusion
In this paper, we have used the molecular dynamics simulation to investigate the binding of zinc finger structure of PSA protein to a short DNA. Previous study has shown that, in its own entity, zinc-finger structure is more stable if the cysteine amino acids that bind the zinc ions are in the deprotonated state. Within the scope of this work, our preliminary results show that this is also the case with the presence of DNA, the deprotonated state creates more stable DNA-protein complex. More comprehensive analysis and structural studies of these complexes together with genome sequence specificity are being conducted and will be published in near future work. Our results provide insights into the DNA binding state of PSA protein and have potential application in designing specialized biosensor for prostate cancer screening.

The authors acknowledge the financial support of the Vietnam National University for the Key Laboratory of Multiscale Simulations of Complex System, annual grant number TXTCN.18.04. We thank Dr. Lien Truong of the Hanoi University of Science and Technology for useful discussion and for sharing her preliminary work on manufacturing biosensors of prostate cancer screening.

References
[1] Balk S P, Ko Y J and Bubley G J 2003 J. Clin. Oncol. 21 383–391
[2] Formisano N, Jolly P, Bhalla N, Cromhout M, Flanagan S P, Fogel R, Limson J L and Estrela P 2015 Sens. Actuators B Chem. 220 369–375
[3] Aus G, Damber J E, Khatami A, Lilja H, Stranne J and Hugosson J 2005 Arch. Intern. Med. 165 1857–1861
[4] Botchorishvili G, Matikainen M P and Lilja H 2009 Curr. Opin. Urol. 19 221
[5] Lilja H, Ulmert D and Vickers A J 2008 Nat. Rev. Cancer 8 268
[6] Liu B, Lu L, Hua E, Jiang S and Xie G 2012 Microchim. Acta 178 163–170
[7] Shaffer P L, Jivan A, Dollins D E, Claessens F and Gewirth D T 2004 Proc. Natl. Acad. Sci. U.S.A. 101 4758–4763
[8] Godwin R C, Melvin R L, Gmeiner W H and Salsbury Jr F R 2017 Biochemistry 56 623–633
[9] Showalter S A and Brüschweiler R 2007 J. Chem. Theory Comput. 3 961–975
[10] Ivani I, Dans P D, Noy A, Pérez A, Faustino I, Hospital A, Walther J, Andrio P, Goñi R, Balaceanu A et al. 2016 Nat. Methods 13 55
[11] Price D J and Brooks III C L 2004 *J. Chem. Phys.* **121** 10096–10103
[12] Hess B, Kutzner C, Van Der Spoel D and Lindahl E 2008 *J. Chem. Theory Comput.* **4** 435–447
[13] Nosé S 1984 *Mol. Phys.* **52** 255–268
[14] Hoover W G 1985 *Phys. Rev. A* **31** 1695
[15] Parrinello M 1985 *Phys. Rev. Lett.* **45** 1196
[16] Parrinello M and Rahman A 1981 *J. Appl. Phys.* **52** 7182–7190
[17] Darden T, York D and Pedersen L 1993 *J. Chem. Phys.* **98** 10089–10092
[18] Hess B, Bekker H, Berendsen H J and Fraaije J G 1997 *J. Comput. Chem.* **18** 1463–1472
[19] Humphrey W, Dalke A and Schulten K 1996 *J Mol Graph* **14** 33–38