IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 8, 2019, accepted September 18, 2019, date of publication September 25, 2019,
date of current version October 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2943748

InPrNa: A Tool for Insight Into Protein–Nucleic Acids Interaction Information

WEI WANG[1,2], YUAN ZHAO[1], HONGJUN ZHANG[3], SHIGUANG ZHANG[1], KELIANG LI[1], AND HEHE LV[1]

[1]Department of Computer Science and Technology, College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China
[2]Laboratory of Computation Intelligence and Information Processing, Engineering Technology Research Center for Computing Intelligence and Data Mining, Xinxiang 453007, China
[3]School of Aviation Engineering, Anyang Normal University, Anyang 455000, China

Corresponding author: Wei Wang (pwwei6@163.com)

This work was supported in part by the Natural Science Foundation of Henan province under Grant 182300410368, Grant 182300410130, Grant 182300410306, and Grant 162300410177, in part by the Production and Learning Cooperation and Cooperative Education Project of Ministry of Education of China under Grant 201702115008, in part by the Science and Technology Research Key Project of Educational Department of Henan Province under Grant 16A520016, Grant 17B520002, and Grant 18A520013, in part by the Key Project of Science and Technology Department of Henan Province under Grant 182102210208, and in part by the Ph.D. Research Startup Foundation of Henan Normal University under Grant qd15130, Grant qd15132, and Grant qd15129.

ABSTRACT Protein-nucleic acids bindings play key roles in many biological processes. However, the biological mechanisms underlying these interactions are not fully understood. Understanding the interface features between protein and nucleic acids may offer insights into how proteins are coupled with nucleic acids. There is a lack of tools that insight into the features of interface in a protein at present. In this work, we developed the InPrNa tool, a graphical tool for protein-nucleic acids complexes that works seamlessly within the PyMOL and gives quick results including 3D visualization for the residue's structure. InPrNa provides three distinct visualization modes to highlight interface, detecting interface residues with different distance between protein and nucleic acid, marking physicochemical properties of the interface residues, and displaying three spatial structures of the interface residues. We also demonstrate the effectiveness of InPrNa's algorithm by contrast DNA binding proteins (DBPs) and RNA binding proteins (RBPs). These results show that DBPs and RBPs have significant differences in amino acids distribution and structural distribution. InPrNa may help for analysis the interface characteristics of proteins-nucleic acids in PyMOL, and can be particularly useful in rapidly pinpointing the interaction mode of proteins-nucleic acids interface. Availability: Freely available at https://github.com/HNUBioinformatics

INDEX TERMS Protein-nucleic acids, interaction information, protein structure.

I. INTRODUCTION

Nucleic acids binding proteins (NBPs) are a type of proteins that are composed of amino acids binding domains and thus have specific or general affinities for either. The proteins are essential and ubiquitous proteins involved in many biological processes, such as nucleic acids replication, transcription, nucleic acids repair and gene expression [1]. Now that the specific DNA sequences or certain nucleobases can be specifically identified and bound by nucleic acids binding proteins. It is valuable to develop the tools which will help the researcher to deeply understand how such specific recognition occurs visually [2]. With more and more protein-nucleic acids complexes resolved, it becomes possible to statistically investigate the binding specificity [3]. Since the structure and biological mechanism is not fully understood, the interface of the proteins needs further analysis. The analysis and visualization of protein interface are not only used to find protein-nucleic acids binding features, but also can help to explain the interaction mechanism of protein and nucleic acids.

PyMOL (https://pymol.org) is an open-source software for molecular visualization, and the source code of the latest version of PyMOL is available in Sourceforge. PyMOL has a plugin expansion interface to provide, and there is relatively good scalability. The utilities of PyMOL have been extensively enhanced by many plugins, including protein-ligand docking, homology modeling, macromolecular analysis, and so on [4]. Even though PyMOL has some plugins about dock or NBPs binding domain [5], new challenges

The associate editor coordinating the review of this manuscript and approving it for publication was Bora Onat.

emerge because of the need to insight into the structural and physicochemical information in protein-nucleic acids interface. PDIviz [6] is a plugin to analyze the protein-nucleic acids binding interfaces, but PDIviz lacks structural analysis, and it can't analyze the physicochemical properties of proteins. Autodock plugin [7] is a PyMOL plugin for docking simulations by use of AutoDock Vina. NRGsuite [8] is a PyMOL plugin which performs docking simulations in real time. There are also some protein-nucleic acids interface databases, such as the NBPs interface database (PDIdb) [9] is a database for functional classification of the protein-DNA complexes. And there are some web-servers for docking: SwissDock, istar, DOCK Blaster, and so on [10]–[12]. Currently, most of the tools for the interface of proteins focus primarily on virtual screening and virtual docking. There are few tools focus on visualized analysis of the physicochemical properties and the surface structure in the nucleic acid-binding protein interface. Here, we present a InPrNa tool: an open-source PyMOL plugin (PyMOL versions 1.8.0 and above) which can select NBPs specific interface for NBPs. InPrNa is specifically designed to the visualization of surface structure on the protein-nucleic acid interface and the visual analysis of the physicochemical properties on the interface, and it can calculate the shape of the binding domain protein or a specific interface with the CX method [13], and focus on visual analysis for various residues shapes of protein-DNA interactions. Visual analysis can help the users find out the interface properties of nucleic acid-binding proteins and gain a deeper understanding of how protein and nucleic acid bind.

II. CALCULATION METHOD

A. DEFINITION OF PROTEIN INTERFACE

The previous studies [14]–[17] on NBPs binding site prediction have used various definitions of binding sites, and the utilization of Euclidean distance methods to determine binding sites is a kind of rapid and precise method [18], [19]. So we used the Euclidean distance between the residues atoms and the nucleic acid molecules to determine these binding sites. When the distance is less than the threshold between the any heavy atom of the residue and any one atom of the nucleic acids, the residue will be considered one of the interfaces.

B. DETERMINATION OF INTERFACE SURFACE SHAPE

The surface shape of the nucleic acid binding protein is irregular and various grooves. InPrNa divides the spatial shape of protein surface into three categories by means of CX algorithm. The CX algorithm was proposed by *Alessandro Pintar*, which determines the shape of protrusions and depressions on the surface of the protein by calculating the ratio of the occupied volume and the free volume of the protein in the sphere.

InPrNa analyzes the shape of the interface through a CX algorithm: The *Ca* atom of each residue is used as the center of a sphere with radius *R*(default *R* is set to 12 Å). The volume of the sphere is *Vint*, which counts the number of non-hydrogen atoms in the sphere as *Natom*(set the volume of each non-hydrogen atom as 20.1 $Å^3$). The volume of all non-hydrogen atoms in the sphere is *Vint*[20].

Shape index (*SI*) is the average value of the CX values of non-hydrogen atoms in this component, and we judge the shape of the component by the value of *SI*: valley ($SI <= -0.2$), flat ($-0.2 < SI < 0.2$), and peak ($SI >= 0.2$), the *SI* threshold values were set based on our previous experiment [21].

$$Vext = Vsphere - Vint \tag{1}$$

$$CX = (Vext - Vint)/Vsphere \tag{2}$$

III. IMPLEMENTATION

InPrNa is a PyMOL plugin developed in Python, so it can be easily installed using PyMOL's install plugin function. When InPrNa is installed in PyMOL, the users can find InPrNa from PyMOL's plugin menu. Opening the InPrNa plugin will bring up an import dialog box at first. Click the browse button to select a PDB file to import. The protein-nucleic acids complex PDB file loaded in InPrNa is automatically calculated and the interface will be shown in a variety of different color after the calculation is completed. There are three different tabs in the tool: (i) 'Visualization' tab, containing interface parameters for calculation and visualization, and setting the color and display mode of the specific interface. (ii) 'Amino acid' tab: containing a table listing calculated physicochemical properties in the interface, and setting the color and display mode of all residues with certain physicochemical properties in the protein. (iii) 'Bump' tab: containing spatial structures of each interface residue, and setting the color and display mode of different structures of residues.

A. 'VISUALIZATION' TAB

InPrNa provides the function to display the interface states for the convenience of analysis interface. When the drop-down box of 'Interface distance' is set in the 'Visualization' tab (Fig. 1), the user can select any Euclidean distances for calculation. After selecting, click on the 'Apply' button, the user can see the composition of the interface in PyMOL. The different Euclidean distance (The default distance is 6 Å) between any heavy atom of residues and any heavy atom of nucleic acid bases are less than the setting value, which can be recognized to be the specific interface domain of the protein. The distance can also be set any other non-zero value in the 'visualization' tab, and the display color and display mode of the interface field can also be selected in the 'visualization' tab. When the user clicks on the drop-down menu of 'interface distance', there are 15 distances to choose from (These distances are a number between 3 Å and 10 Å). After choosing a distance, click the 'Apply ' button and the user can immediately see the interface directly in PyMOL window. The amount of amino acid that the interface contains will be displayed in the 'Visualization' tab. The user clicks 'interface color' button will pop up the specific 'Color' window.

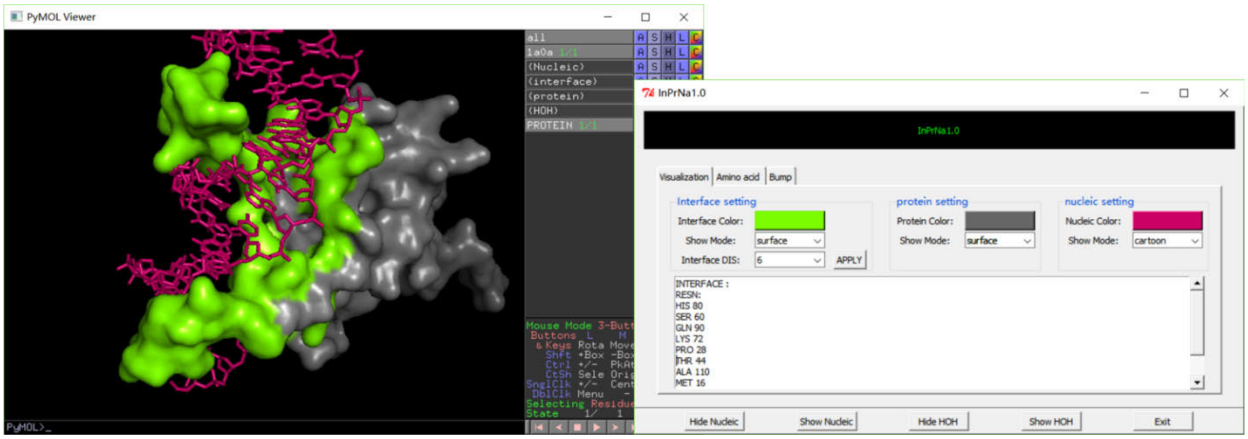

FIGURE 1. Proteins-nucleic acids interface visualizations are created by InPrNa in the 'Visualization' tab. The phosphate system positive regulatory protein PHO4/DNA complex (1a0a) is presented in the left-hand side. InPrNa can be chosen to set the interface color and display mode for protein and nucleic acid, and it can show the residues numbers in the interface.

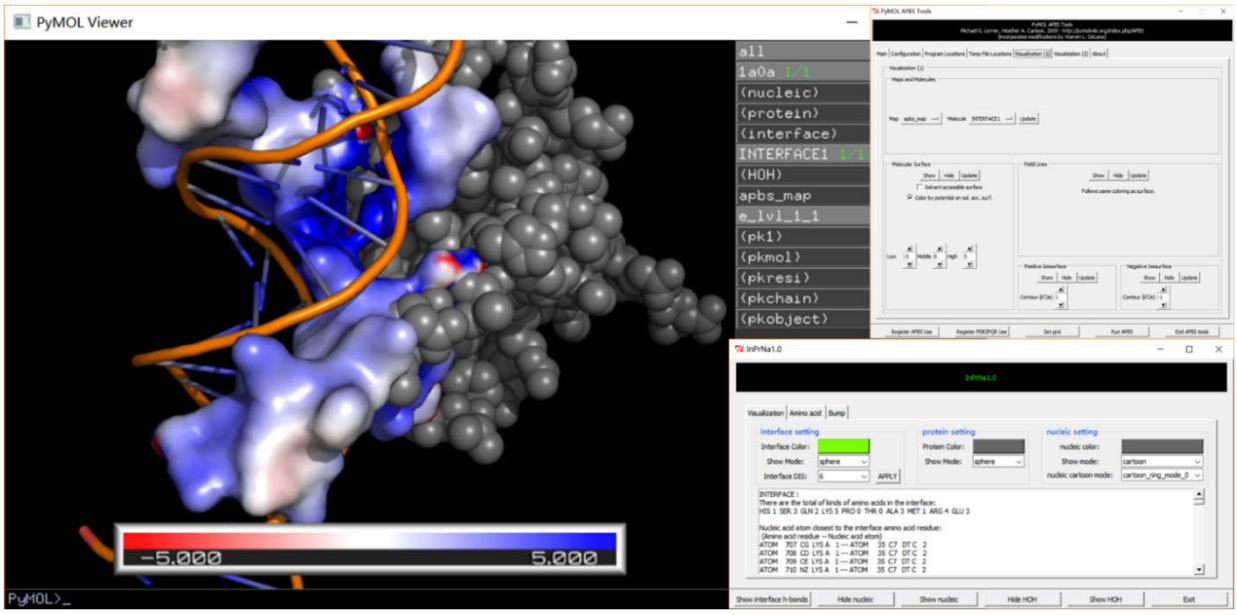

FIGURE 2. InPrNa and PyMOL's ABPS plugin can make an electrostatic map of the interface. The binding region of nucleic acid to protein in the PHO4/DNA complex (1a0a) is basically positively charged.

After the user selects the color in this window, the interface color will convert immediately to the selected color after confirms, then the user clicks the drop-down menu after 'interface show mode', the user can select the display mode of the interface. Electrostatic surface maps can show the charged nature of the interface, and be usually calculated using APBS. Users can create an electrostatic map of the interface using InPrNa and PyMOL's ABPS plugin (Fig 2).

B. 'AMINO ACID' TAB

InPrNa also provides the function of displaying the physicochemical properties of residues: in the 'Amino acid' tab (Fig. 3), the user can select the color and display mode of the selected physicochemical property of amino acids. This tool can visually analyze the physicochemical properties of the binding protein interface. At the same time, the number of residues with the physicochemical property in the protein will also be shown in 'Amino acid' tab.

C. 'BUMP' TAB

The spatial structure of specific interface between nucleic acids and proteins has attracted much attention. InPrNa divides the spatial patterns presented into valley, peak and flat shape base on the CX algorithm. The user can set the color and display mode of the three spatial forms of different regions in the 'Bump' tab (Fig. 4).

IV. ANALYSIS AND RESULT

A. MATERIALS

To verify the effectiveness of the InPrNa algorithm, we downloaded DNA binding proteins (DBPs) and RNA binding proteins (RBPs) from PDB database for analyzing

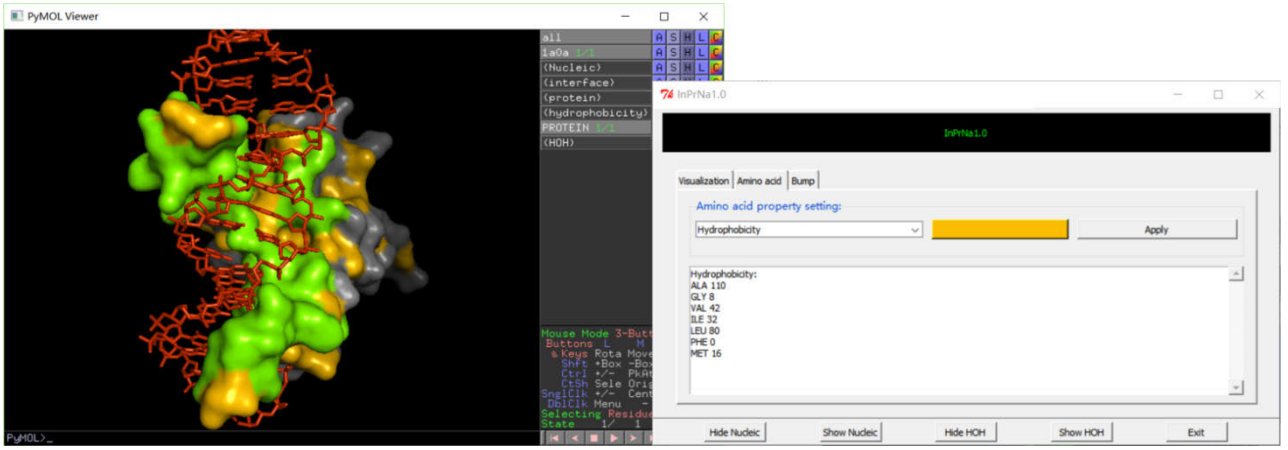

FIGURE 3. The physicochemical properties of interface visualizations are created by the 'Amino acid' tab. The upper left panel gives the physicochemical properties of PHO4/DNA complex (1a0a) is presented in the different color, and the number of these amino acids.

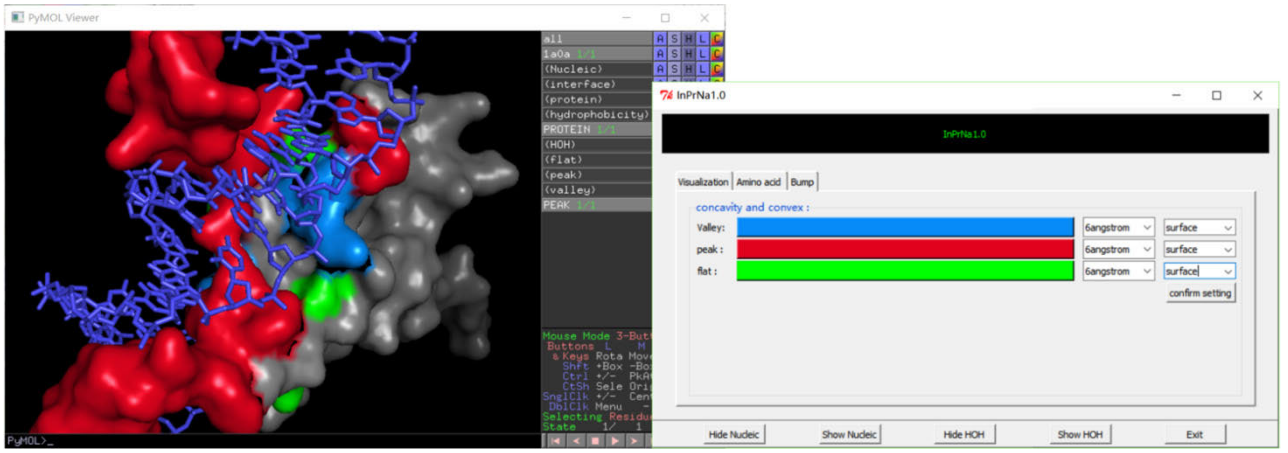

FIGURE 4. The spatial structures (valley, peak and flat) in the interface visualizations are created by InPrNa in the 'Bump' tab. The PHO4/DNA complex (1a0a) is presented mode of three spatial structures with different color in the interface, and the InPrNa can display the spatial structures of interface at the different Euclidean distance between proteins and nucleic acids.

physicochemical and structure features of specific interface. The 8021 DBPs and the 5660 RBPs have been collected from the PDB database [22]. The CD-HIT program is used to remove homology redundancy for the collection of protein data [23]. Finally, we obtained 912 DBPs and the 573 RBPs as test datasets (Stable 1 in the Supplementary File). We statistically analyze the filtered data to verify that the InPrNa plugin can assist the researcher to more intuitively understand the physicochemical properties of the interface between protein and nucleic acid.

B. ANALYSIS OF INTERFACE

Here, we adopted the criteria proposed in the previous study to determine whether the residues belong to the nucleic acid binding interfaces [20]. Kuznetsov *et al.* [18] and Si *et al.* [16] used Euclidean distance to distinguish binding and nonbinding residues. Shandar Ahmad considers the interface residues have a cut off distance of 3.5 Å between any heavy atom of the residue and any atom in the nucleic acid molecules. However, some computational methods of interface used the distance

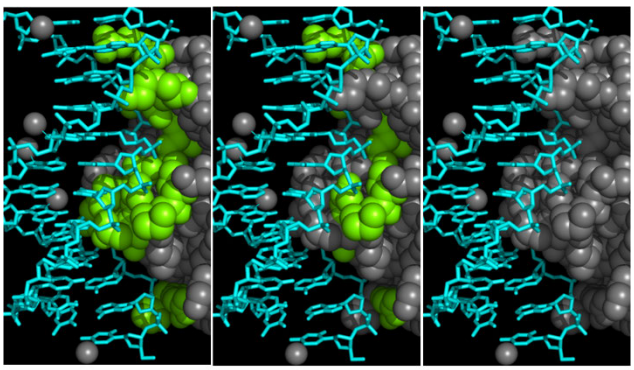

FIGURE 5. The binding sites at different Euclidean distances. The light blue part is the DNA molecule, the green part is the binding sites, and the gray part is the other protein structures.

less than 6 Å [19]. The Euclidean distance between protein and nucleic acid molecule that lack of uniform metrics. We analyzed these NBPs complex and statistically analyzed

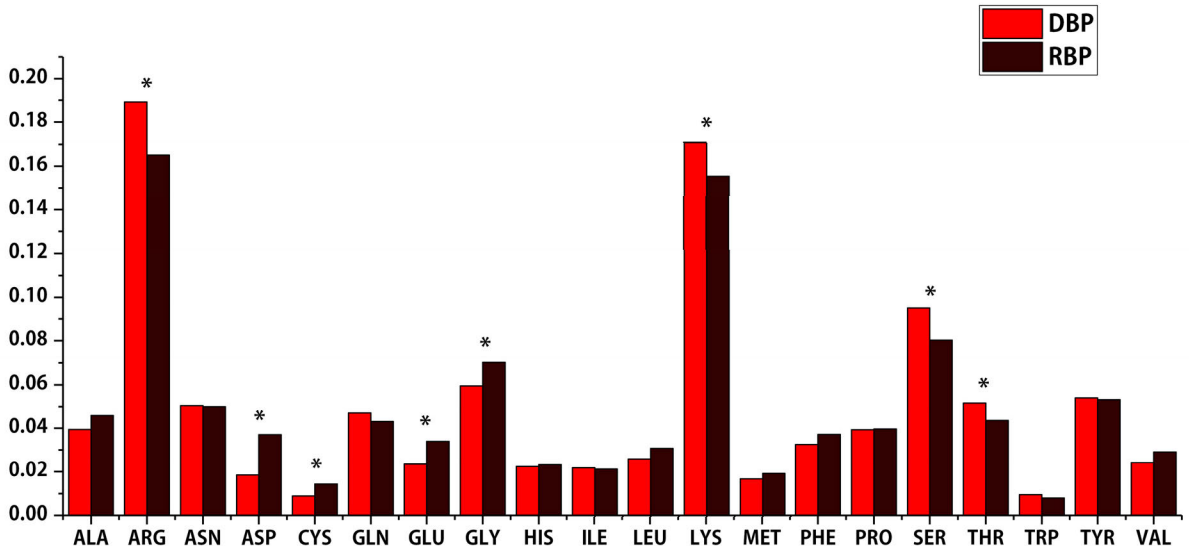

FIGURE 6. The distribution of 20 amino acids in the interfaces (Euclidean distance is 4.5 Å), and the asterisk marked in the Fig. indicates a significant difference with P-value <0.05.

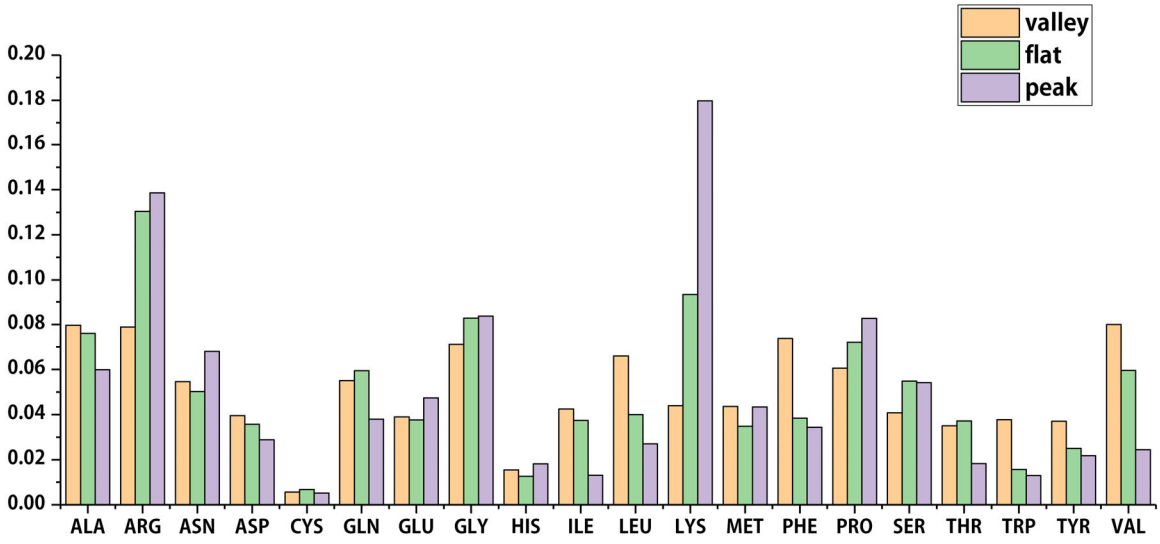

FIGURE 7. The amino acid distribution of the three spatial shapes in the DBPs interface.

the number of amino acids contained in the interface at different distances. After contrasted the interface distribution by using the different Euclidean distance values, we can find the interface significant distributional difference at different distances. The Fig. 5 shows a DBPs complex (the PBD ID is 1PP7) as an example to observe the shape of the interface at different distances. Therefore, by comparing the statistical results and referring to the work of Kuznetsov *et al.* [18], we choose 4.5Å as the Euclidean distance to calculate the interface, and statistically analyze the physicochemical properties of the interfaces. When the Euclidean distance between any atom of a residue and any atom of a nucleic acid molecule is less than 4.5 Å, this residue is considered to interact with the nucleic acid, and the residue is considered to belong to the interface.

Electrostatic complementarity is considered to be one of the main reasons for the non-specific binding of proteins to DNA [24]. The electrostatic charge of the NBPs interfaces is one of the most influential roles in protein-nucleic acid recognition and interaction, and many studies have confirmed this role. According to previous studies [24], [25], nucleic acids have a higher negative charge and NBP-binding sites have higher positive electrical properties. In Fig. 6, the positively charged amino acids (Arg and Lys) in the distribution statistics of amino acids in the interface have a higher ratio. Generally, the nucleic acids backbone is carried a net negative

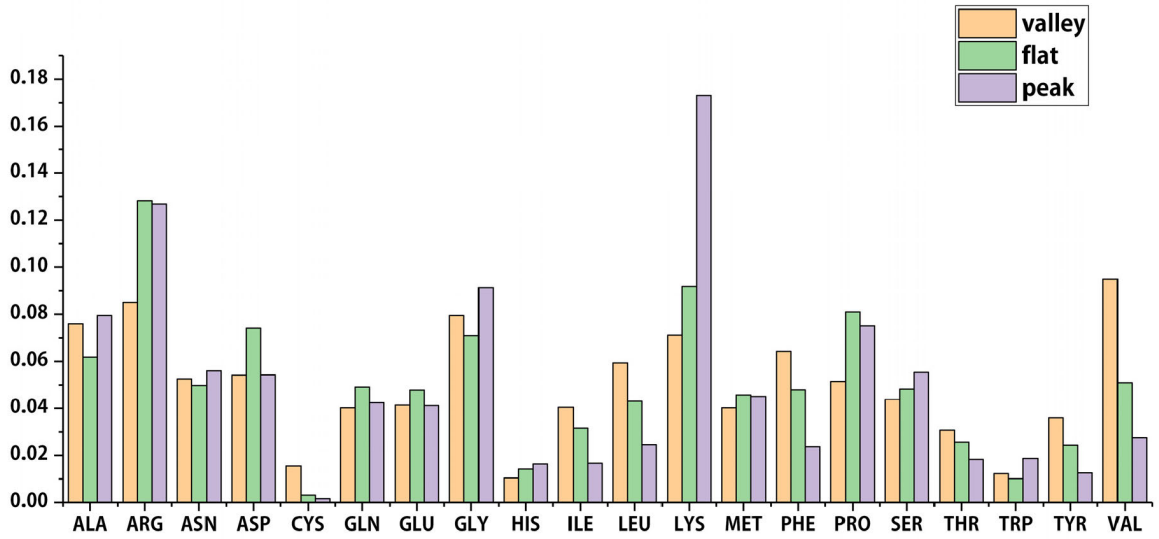

FIGURE 8. The amino acid distribution of the three spatial shapes in the RBPs interface.

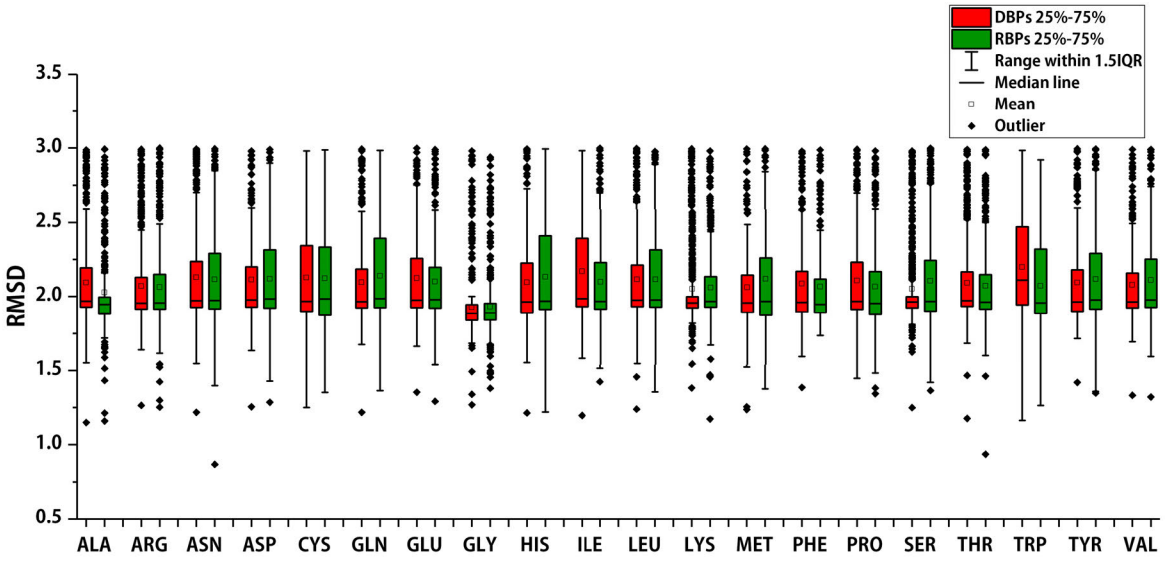

FIGURE 9. Comparison of residues of RMSD in the Interfaces of DBPs and RBPs.

charge because of the numerous phosphate groups in their structure. The positively charged amino acid sequence in the 20 amino acids is considered to be a characteristic region capable of binding nucleic acids. This is also considered to be a condition for protein and nucleic acids binding: the amino acids binding domain must be a multipolar positively charged amino acid domain. Therefore, a specific interface region will inevitably satisfy some physicochemical properties: (i) Positive electricity, (ii) Hydrophilicity. The amino acids binding domain tends to be exposed to solvents to adapt to nucleic acids [25], [26].

The DBPs and RBPs are measured using the quantitative ratios of the physicochemical properties of the amino acids contained in the interfaces (Sfig.1 and Sfig.2). As shown in additional Sfig.1 and Sfig.2, the results show that the residues with positive charge occupy a relatively high proportion in the interface, and the residues with negative charge occupy a relatively low ratio in the interface, which is the same as the theoretical analysis. We found that the hydrophilic residues occupy relatively high proportions with respect to the hydrophobic residues, which is consistent with the theory that the interface should be hydrophilic. As can also be seen from the figures, the statistical results for RBPs and DBPs are not much different for positive electricity, hydrophilic and hydrophobic. The statistical analysis of the distribution was performed using T-test to measure the significance of

the difference. Here, the distribution of residues in the RBPs and DBPs binding regions was calculated by the T-test. The marked asterisks in Fig.6 indicate significant differences with P-values <0.05 (ARG, ASP, CYS, GLU, GLY, LYS, SER and THR). The positive electricity amino acids (ARG, LYS) have the highest proportion in their interfaces. So there are significant differences and some common interface features between DNA and RNA binding proteins. In addition, we believe the surface shapes of the DBPs and RBPs should be different to bind with different nucleic acid ligands. Thus the interface residues are calculated the proportion of three spatial shapes of amino acids. Here, the DBPs and RBPs are measured the shapes of amino acids in the interfaces. The Fig. 7 and Fig. 8 show that the residues of valley shape have higher proportion for ILE, LEU, PHE, TYR and VAL in the DBPs and the RBPs. However, the positive electricity amino acids (ARG and LYS) have more the residues of peak shape in two kinds of proteins. These interesting results may reveal the principle of the lock key model from another side (Supplementary data in Stable 2 and Stable 3). The most commonly used metric in this category is the root-mean-square deviation, RMSD, in which the root-mean-square distance between corresponding residues is calculated after one structure to another. So we compared the interfaces RMSD values between DBPs and RBPs. The results show there are some slight differences between DBPs and RBPs in Fig. 9. We find that ALA, GLN, ILE, LYS, SER and TRP present more differences than others.

V. CONCLUSION

In this work, we have developed the InPrNa tool, an easy-to-use PyMOL plugin that can analyze the physicochemical properties of NBPs and present them in an easy to understand manner. The users can visually observe some of the physicochemical properties of these proteins by InPrNa, which will help researchers to understand in depth how this specific identification occurs. InPrNa has three main functions: (i) It can show the interface of the nucleic acid binding protein. (ii) It can show the physicochemical properties and count the number of residues with different physicochemical properties in the interface. (iii) The spatial structure of interface can be divided into three types (valley, flat, peak) by the CX algorithm, the user can set the color and display mode of these three spatial structures. After we analyzed the interface information of DBPs and RBPs based on the InPrNa tool. The results show that the interfaces of DBPs have significant positively charged, and there are some differences between DBPs and RBPs. Moreover, the spatial structure of the interface is mostly different shapes residues, which will be helpful to understand the principle of lock-key model between protein and nucleic acid. Therefore, the InPrNa will be helpful to learn the details about the protein binding specificity. Further refinement of the method will be applied to the detection of specific interfaces between proteins and drugs, and it may also be applied to designing and quantitatively analyzing drug compatibility.

REFERENCES

[1] N. Pokhrel, C. C. Caldwell, E. I. Corless, E. A. Tillison, J. Tibbs, N. Jocic, S. M. A. Tabei, M. S. Wold, M. Spies, and E. Antony, "Dynamics and selective remodeling of the DNA-binding domains of RPA," *Nature Struct. Mol. Biol.*, vol. 26, no. 2, pp. 129–136, 2019.

[2] L.-L. Zheng, K.-R. Zhou, S. Liu, D.-Y. Zhang, Z.-L. Wang, Z.-R. Chen, J.-H. Yang, and L.-H. Qu "dreamBase: DNA modification, RNA regulation and protein binding of expressed pseudogenes in human health and disease," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D85–D91, 2018.

[3] W. Zhang, X. Yue, G. Tang, W. Wu, F. Huang, and X. Zhang, "SFPEL-LPI: Sequence-based feature projection ensemble learning for predicting LncRNA-protein interactions," *PLoS Comput. Biol.*, vol. 14, no. 12, 2018, Art. no. e1006616.

[4] S. Yuan, H. C. S. Chan, and Z. Hu, "Using PyMOL as a platform for computational drug design," *Wiley Interdiscipl. Rev., Comput. Mol. Sci.*, vol. 7, no. 2, p. e1298, 2017.

[5] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, "PLIP: Fully automated protein–ligand interaction profiler," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W443–W447, 2015.

[6] J. Ribeiro, F. Melo, and A. Schüller, "PDIviz: Analysis and visualization of protein–DNA binding interfaces," *Bioinformatics*, vol. 31, no. 16, pp. 2751–2753, 2015.

[7] D. Seeliger and B. L. de Groot, "Ligand docking and binding site analysis with PyMOL and Autodock/Vina," *J. Comput.-Aided Mol. Des.*, vol. 24, no. 5, pp. 417–422, 2010.

[8] F. Gaudreault, L.-P. Morency, and R. J. Najmanovich, "NRGsuite: A PyMOL plugin to perform docking simulations in real time using FlexAID," *Bioinformatics*, vol. 31, no. 23, pp. 3856–3858, 2015.

[9] T. Norambuena and F. Melo, "The Protein-DNA Interface database," *BMC Bioinf.*, vol. 11, no. 1, 2010, Art. no. 262.

[10] A. Grosdidier, V. Zoete, and O. Michielin, "SwissDock, a protein-small molecule docking Web service based on EADock DSS," *Nucleic Acids Res.*, vol. 39, no. suppl_2, pp. W270–W277, 2011.

[11] H. Li, K.-S. Leung, P. J. Ballester, and M.-H. Wong, "istar: A Web platform for large-scale protein-ligand docking," *PLoS ONE*, vol. 9, no. 1, 2014, Art. no. e85678.

[12] J. J. Irwin, B. K. Shoichet, M. M. Mysinger, N. Huang, F. Colizzi, P. Wassam, and Y. Cao, "Automated docking screens: A feasibility study," *J. Medicinal Chem.*, vol. 52, no. 18, pp. 5712–5720, 2009.

[13] A. Pintar, O. Carugo, and S. Pongor, "CX, an algorithm that identifies protruding atoms in proteins," *Bioinformatics*, vol. 18, no. 7, pp. 980–984, 2002.

[14] J. Si, Z. Zhang, B. Lin, M. Schroeder, and B. Huang, "MetaDBSite: A meta approach to improve protein DNA-binding sites prediction," *BMC Syst. Biol.*, vol. 5, Jun. 2011, Art. no. S7.

[15] S. Hwang, Z. Gou, and I. B. Kuznetsov, "DP-Bind: A Web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins," *Bioinformatics*, vol. 23, no. 5, pp. 634–636, 2007.

[16] J. Si, R. Zhao, and R. Wu, "An overview of the prediction of protein DNA-binding sites," *Int. J. Mol. Sci.*, vol. 16, no. 12, pp. 5194–5215, 2015.

[17] Y. Xiong, J. Liu, and D.-Q. Wei, "An accurate feature-based method for identifying DNA-binding residues on protein surfaces," *Proteins, Struct., Function, Bioinf.*, vol. 79, no. 2, pp. 509–517, 2011.

[18] I. B. Kuznetsov, Z. Gou, R. Li, and S. Hwang, "Using evolutionary and structural information to predict DNA-binding sites on DNA-binding proteins," *Proteins, Struct., Function, Bioinf.*, vol. 64, no. 1, pp. 19–27, 2006.

[19] S. Ahmad, M. M. Gromiha, and A. Sarai, "Analysis and prediction of DNA-binding proteins and their binding residues based on composition, sequence and structural information," *Bioinformatics*, vol. 20, no. 4, pp. 477–486, 2004.

[20] W. Wang, J. Liu, Y. Xiong, L. Zhu, and X. Zhou, "Analysis and classification of DNA-binding sites in single-stranded and double-stranded DNA-binding proteins using protein information," *IET Syst. Biol.*, vol. 8, no. 4, pp. 176–183, 2014.

[21] W. Wang, J. Liu, and L. Sun, "Surface shapes and surrounding environment analysis of single- and double-stranded DNA-binding proteins in protein-DNA interface," *Proteins-Structure Function Bioinf.*, vol. 84, no. 7, pp. 979–989, 2016.

[22] H. Berman, K. Henrick, and H. Nakamura, "Announcing the worldwide Protein Data Bank," *Nature Struct. Biol.*, vol. 10, no. 12, p. 980, 2003.

[23] Y. Huang, B. Niu, Y. Gao, L. Fu, and W. Li, "CD-HIT Suite: A Web server for clustering and comparing biological sequences," *Bioinformatics*, vol. 26, no. 5, pp. 680–682, 2010.

[24] S. Jones, H. P. Shanahan, H. M. Berman, and J. M. Thornton, "Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins," *Nucleic Acids Res.*, vol. 31, no. 24, pp. 7189–7198, 2003.
[25] Y. Xiong, J. Xia, W. Zhang, and J. Liu, "Exploiting a reduced set of weighted average features to improve prediction of DNA-binding residues from 3D structures," *PLoS ONE*, vol. 6, no. 12, 2011, Art. no. e28440.
[26] S. Shazman, G. Elber, and Y. Mandel-Gutfreund, "From face to interface recognition: A differential geometric approach to distinguish DNA from RNA binding surfaces," *Nucleic Acids Res.*, vol. 39, no. 17, pp. 7390–7399, 2011.

WEI WANG received the Ph.D. (Eng.) degree in computer software and theory from Wuhan University, Wuhan, China, in 2014. He is currently a Teacher and a Researcher with the College of Computer and Information Engineering, Henan Normal University, China. His research interests include bioinformatics, data mining, and machine learning.

YUAN ZHAO is currently pursuing the bachelor's degree with the School of Computer and Information Engineering, Henan Normal University. His research interests include data mining, bioinformatics software development, and web tools.

HONGJUN ZHANG received the M.S. degree from the Department of Computer Science and Technology, Henan Normal University, Xinxiang, China, in 2009. He is currently an Associate Professor with Anyang Normal University, Anyang, China. His main research interests include computer networks and big data.

SHIGUANG ZHANG received the Ph.D. degree in applied mathematics from Hebei Normal University, in 2014. He is currently with the College of Computer and Information Engineering, Henan Normal University, China. His research interests include optimization algorithm, machine learning, and knowledge discovery for regression.

KELIANG LI received the B.S. degree from the Wanfang Science and Technology Institute, Henan Technological University, Zhengzhou, China, in 2014. He is currently pursuing the M.S. degree with Henan Normal University, Xinxiang, China. His research interests include biological information and data mining.

HEHE LV is currently a graduate student with the School of Computer and Information Engineering, Henan Normal University. His research interests include artificial intelligence and bioinformatics.