Biochemical and Biophysical Research Communications 577 (2021) 152–157

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Protein-DNA complex structure modeling based on structural template

Juan Xie [a, 1], Jinfang Zheng [a, 1], Xu Hong [a], Xiaoxue Tong [a], Xudong Liu [a], Qi Song [b], Sen Liu [b], Shiyong Liu [a, *]

[a] School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China
[b] Key Laboratory of Fermentation Engineering (Ministry of Education), Hubei University of Technology, China

ARTICLE INFO

Article history:
Received 29 July 2021
Received in revised form
5 September 2021
Accepted 6 September 2021
Available online 8 September 2021

Keywords:
Templated based
Protein-DNA modelling
Relationship between sequence and structure

ABSTRACT

DNA-binding is an important feature of proteins, and protein-DNA interaction involves in many life processes. Various computational methods have been developed to predict protein-DNA complex structures due to the difficulty of experimentally obtaining protein-DNA complex structures. However, prediction of protein-DNA complex is still a challenging problem compared with prediction of protein-RNA complex, this may be due to the large conformational changes between bound and unbound structure in both protein and DNA. We extend PRIME 2.0 to PRIME 2.0.1 to model protein-DNA complex structures. By comparing sequence and structure alignment methods, we found that structure-based methods can find more templates than sequence-based methods. The results of all-to-all structure alignments showed that DNA structure plays an important role in prediction of protein-DNA complex structure. By exploring the relationship of sequence and structure, we found that in protein-DNA interaction, numerous structures with dissimilar sequences have similar 3D structures and perform the similar function.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

The interaction of two biological macromolecules, protein and DNA, plays a crucial role in transcriptional regulation [1,2]. Researchers have established a number of experimental technologies to study protein-DNA interactions, such as high-throughput methods and X-ray.

The sequence-based/structure-based computational methods were designed to predict protein-DNA interaction. The Thornton's team [3] reported that more than two-thirds of the interactions in protein-DNA occurred between the backbone of the DNA and the protein, independent of the sequence of the DNA. Thus, many laboratories committed to develop structure-based computational methods to predict protein-DNA interactions or model their structures. At present, the structure-based algorithms for predicting protein-DNA complexes are mainly divided into free docking and template-based docking. The method of free docking mainly includes sampling and scoring. In sampling, the free docking method mainly considers the shape and electrostatic complementarity of the protein-DNA complex, and then outputs possible candidate conformations (decoys). In the scoring process, the scoring function is used to screen decoys and picked out the nearly native structures. The template-based docking method searches the template library to obtain a template structure similar to the target structure at first, and then build a suitable model based on the template structure.

Studies have shown that the template-based docking methods were more accurate than the free docking methods in protein-RNA complexes prediction [4,5], and the template-based methods work better in complexes that involved a big conformational changes [6,7]. HDOCK [8] is a hybrid docking method which combines free docking and sequence template-based docking for protein-DNA complex structure prediction. Their research showed that the hybrid approach outperforms free docking, which suggests that template-based methods play an important role in predicting protein-DNA complexes.

Compared with protein-RNA docking, protein-DNA docking is a more challenge problem. There are two possible reasons. First, the interaction information on the surface of DNA is relatively less [9].

* Corresponding author.
E-mail address: liushiyong@gmail.com (S. Liu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.bbrc.2021.09.018
0006-291X/© 2021 Elsevier Inc. All rights reserved.

Second, like in protein-protein docking, there are large conformational changes on protein and DNA during their binding [6,7]. The template-based docking method can make up for the above defects of free docking.

Although the predecessors' works have made great progress in protein-DNA interactions, no one has completely used the structural template-based docking method for modeling protein-DNA complex structures. It perhaps because of the lack of a suitable program for DNA structure alignment currently. Our team developed RMalign [5], which is well suited for DNA structure alignment (Fig. S1). Therefore, we extended PRIME 2.0 [5], a method for protein-RNA structure modeling, to PRIME 2.0.1, and used it to model protein-DNA complexes.

In this study, we mainly discuss the results of applying PRIME 2.0.1 to model protein-DNA complex structure. We found that in many cases the structural alignment-based method can find more templates than the sequence alignment-based method. Considering the protein and DNA structure at the same time can eliminate more non-near-native structures than considering the protein structure alone. In addition, in many structures, their sequences are dissimilar, but they share similar 3D structures. In this case, homologous sequence-based methods for predicting complex structures are often not advantageous.

2. Methods

2.1. The process of protein-DNA interaction modeling

Our structural template-based approach PRIME 2.0.1 for modeling protein-DNA interactions mainly includes three parts: building the dataset, predicting the model and evaluating the predicted model.

The method of building the dataset is similar to the method for modeling protein-RNA interactions in our group [4,5]. The protein-DNA complexes were downloaded from PDB (a total of 3963 protein-DNA complexes were downloaded on June 5, 2017). 3287 complex structures were kept whose resolution were better than or equal to 3 Å. These high-resolution structures were divided into binary complexes. The so-called binary complex means that the structure contains one protein chain and one DNA chain. As a result, a total of 6,704 protein chains with a length of more than 30 amino acids and 1585 DNA chains with a length of at least 20 bases were reserved as the same as the study in protein-RNA [4]. A binary complex structure with a maximum distance of 4.5 Å between any two heavy atoms of the protein-DNA interaction chains was kept. This eventually left 4631 binary complex structures, of which contained 2342 protein chains and 1583 DNA chains. In order to make the results more fair, we used the CD-HIT [10] software to remove redundant DNA with an identity threshold of 0.99. 1583 DNA chains were clustered into 683 classes. We selected a representative structure in each cluster. The final representative structure was the one who owns the highest resolution and the highest sequence identity with the representative structure selected by CD-HIT in each cluster. Finally, a total of 1536 protein-DNA non-redundant binary complexes (NRBC1536) were preserved for subsequent all-to-all alignments and benchmarking. In order to verify the predictive power of our program, we divided the NRBC1536 into two groups: 330 newly solved structures (NRBC330) were designated as targets, and 1206 "older" structure (NRBC1206) served as templates.

2.2. Targets/templates alignment

When performing all-to-all alignments in the dataset of NRBC1536, we carried out global sequence alignment and structure alignment.

The first method was global sequence alignment. We used the needle in the EMBOSS software [11] for sequence alignment with the default parameters. The sequence identity of the complex structure was the smaller sequence identity in the binary complex monomer. That is, needle calculated the sequence identity for protein and DNA respectively, and we used the smaller sequence identity between protein and DNA as the final sequence identity of the binary complex structure.

The second method was structural alignment. The binary complex structure was split into protein and DNA monomers that were aligned to the protein and DNA in the template respectively. For protein structure alignment, TM-align [12] was employed as the aligner like previous studies for protein-protein [13–18] and protein-RNA interactions [4,5]. The TM-score which shows the results of TM-align has values varying from 0 for completely dissimilar structures, to 1 for identical structures. The DNA was aligned with RMalign, which was based on a scoring function RMscore that is independent of the size of the molecule. RMalign was developed for RNA 3D structural alignment, but it is still effective in DNA 3D structural alignment. Similar to TM-score, RMscore also used a normalized value to describe the similarity of DNA. The larger the RMscore, the more similar the two DNA structures are. The smaller of TM-score and RMscore was chosen as the final complex structure score to describe the structural similarity of the binary complex.

Both of the above alignment methods were employed to the NRBC1536 dataset to test whether the target can find the correct template (correct template is defined as the complex structural score greater than transition point, which describes the transition from random to similar binding mode). Like the studies on protein-RNA interaction [4,5], the atoms (Cα and C3' atoms represent protein and DNA respectively) in both alignment methods were used to calculate the interaction RMSD (iRMSD), which was applied to characterize the similarity of binding mode between two protein-DNA complex structures.

In order to evaluate the quality of the built model, we calculated the ligand RMSD (lRMSD) between the model and the native complex. In protein-DNA docking, if lRMSD $\leq$ 5.0 Å, the predicted model was considered to be medium quality; and if lRMSD $\leq$ 10.0 Å, the predicted model was deemed to be acceptable. In this study, we also adopted the same lRMSD evaluation criteria.

2.3. Modeling and evaluation model

In practice, the target is usually unaware of its template structure. Therefore, the template was needed to be found out for the target structure before modeling by TM-align and RMalign. After the structure alignment (see Supplementary Material), PRIME 2.0.1 outputs the predicted model according to the template structure. During this process, the target protein and DNA are superimposed on the protein and DNA respectively in the template structure. Undergoing rotation and translation, TM-align [12] outputs the matrix when the target and template proteins share the maximum TM-score. Similarly, RMalign [5] also outputs a rotating matrix while the target and template DNA structures have the maximum RMscore. And then, the model is built with the matrixes generated by TM-align and RMalign.

3. Results

3.1. The relationship between sequence/structure and the binding mode

The relationship between iRMSD and complex sequence

identity/complex structural score in the protein-DNA complexes is explored by performing an all-to-all pairwise alignment on the NRBC1536 like previous studies [4,20]. In Fig. 1, it shows the relationship between complex sequence identity and iRMSD when all-to-all alignment is conducted on the NRBC1536. From the illustration of Fig. 1, we find that the noise is very large when the sequence identity is less than 0.25. That is the binding mode of the complex structures with low similarity. In comparison, when the sequence identity $\geq$ 0.25, the noise is reduced, and more complex structures share a similar binding mode. Therefore, in this study, we take the sequence identity with 0.25 as the threshold to distinguish the true binding mode from random binding mode when the sequence alignment is performed.

In Fig. 2, we plot the relationship between structural similarity and iRMSD with an all-to-all alignment on the NRBC1536. We studied the relationship between the complex structural score/TM-score and iRMSD. The inset of Fig. 2(a)/b shows that the transition point is 0.4/0.5 for complex structural score/TM-score. That is, the ratio of the number of structures with iRMSD $\leq$ 5 Å increases when the complex structural score/TM-score changes into 0.4/0.5.

As can be seen from Fig. 2(b), a partial of protein-DNA complex structures can find suitable templates (TM-score $\geq$ 0.5) when the TM-score was considered alone. However, relatively large noise still exists at TM-score $\geq$ 0.7, and we found that most of the noise was removed after the DNA structure alignment was considered (Fig. 2(a)). For instance, the target structure 5CMX has chains H and A [21], and the template structure 4I7Y has chains H and D [22], with iRMSD 52.72 Å and TM-score 0.988, ranking first in the case of considering the TM-score alone. However, the complex structure score is 0.287, ranking 72nd. So, the point representing the relationship between 5CMX_HA and 4I7Y_HD was moved to the left of the transition point, and therefore noise can be reduced. Previous study shows that relying solely on TM-score is not sufficient in predicting DNA-binding proteins because one-third of non-DNA-binding proteins have TM-score $\geq$ 0.55 with DNA-binding proteins [23]. Therefore, both protein and DNA are essential for finding the template for protein-DNA complex structure modeling. The maximum value of the iRMSD is smaller than that of the protein-RNA complexes [4], showing that the binding mode between the protein-DNA complexes is more similar than the protein-RNA complexes. It is probably because RNA can form a more complex secondary structure. Although protein or DNA has large conformational changes if they are forming a complex, most of the DNA structure forms a relatively fixed double helix, so there may be more similar binding interfaces between DNAs such that their final complex structure may be similar.

3.2. Comparison of sequence identity and structural similarity

The previous section reveals the transition point is 0.25/0.4 for sequence/structure. Both structure and sequence can describe the similarity of binding mode. So, what is the relationship between structure and sequence in protein-DNA interaction? To this end, we explored the relationship for protein-DNA complex structures.

In Fig. 3, it depicts the relationship between sequence identity and structural similarity on the NRBC1536. The relationship between sequence identity and structure similarity was divided into four quadrants by X = 0.4 (transition point of structure score) and

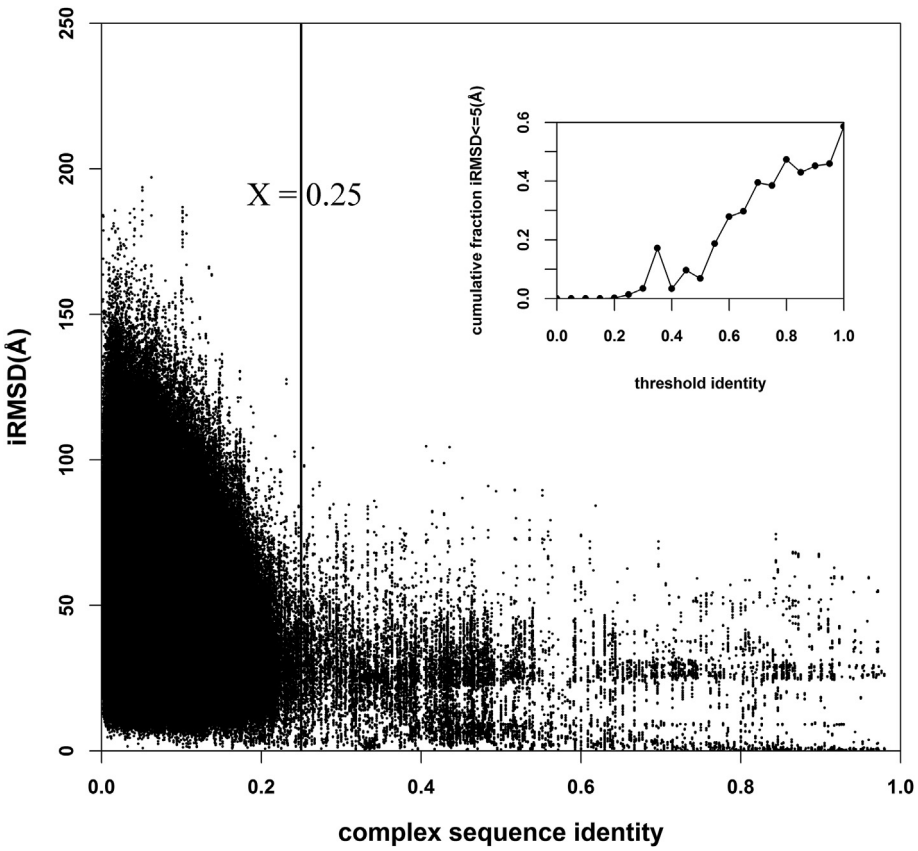

Fig. 1. The relationship between the binding mode and the sequence identity of the complex structures. The X-axis denotes the smaller value of the monomer sequence identity of the complex structure. The Y-axis represents iRMSD. A pairwise comparison of all the complexes in NRBC1536 is described. In the inset, statistics are taken at intervals of 0.05, the ratio of complexes with iRMSD $\leq$ 5 Å in this interval to show the phase transition. The vertical line represented the transition point with the sequence identity 0.25.

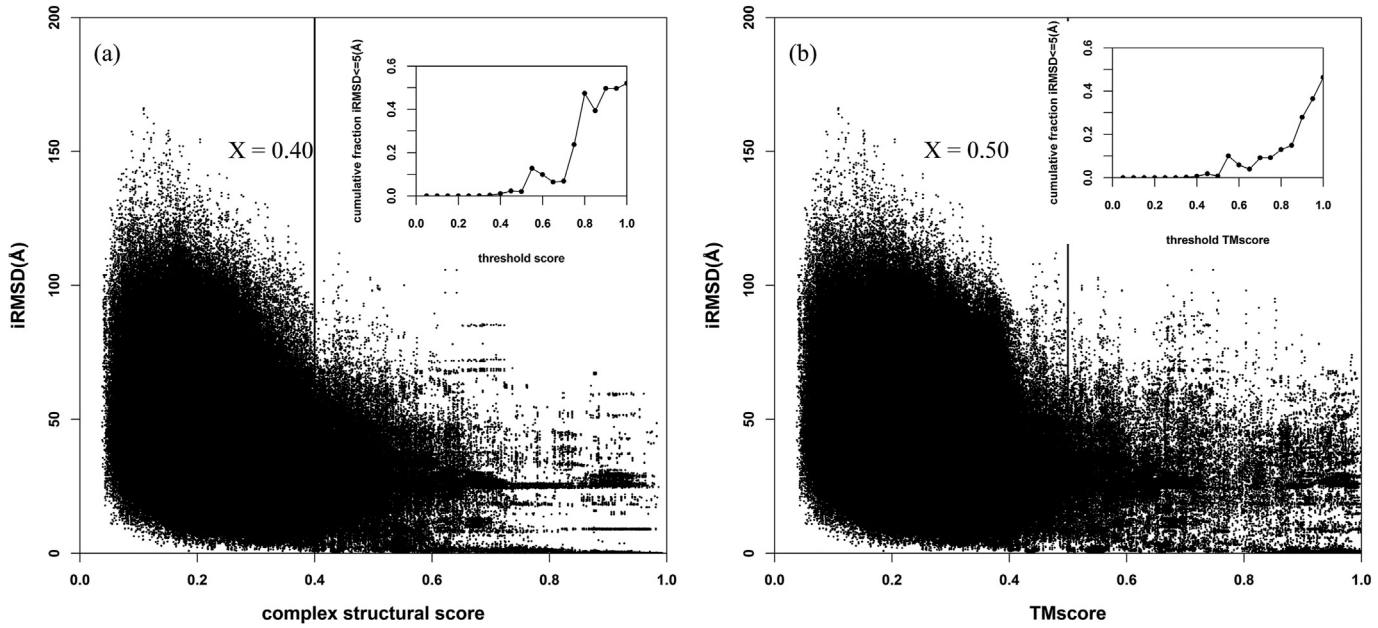

Fig. 2. The relationship between the binding mode and the similarity of the complex structures. The iRMSD is plotted against the lowest complex structures score (a) and protein structural score (b) in the protein-DNA binary complex structures. A pairwise comparison of all binary complexes in NRBC1536 is depicted. In the illustration, according to the complex structure score(a) and protein structure score(b), statistics are taken at intervals of 0.05, the ratio of the complexes with iRMSD $\leq$ 5 Å in this bins to show the phase transition. The vertical lines showed that the transitions are occurred when the similarities of complex structure are 0.40(a) and protein structure is 0.50(b).

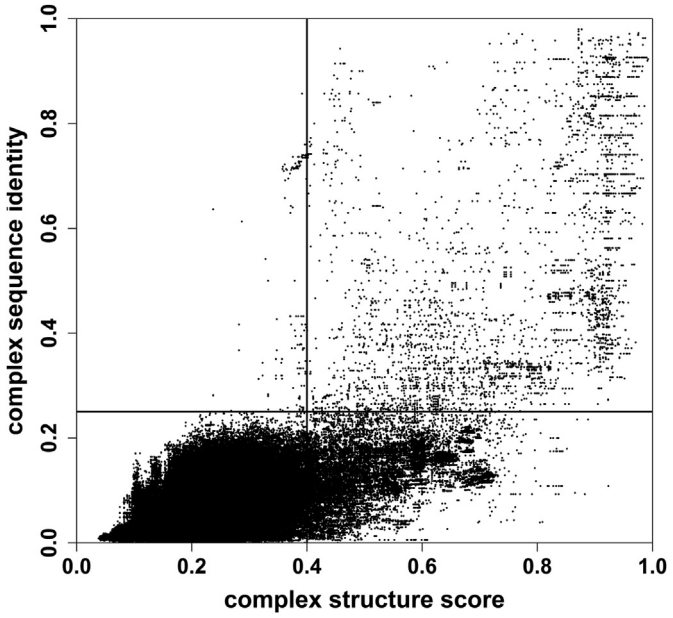

Fig. 3. The relationship between structure similarity and sequence identity in protein-DNA complex structures. The relationship between structural similarity and sequence identity of the protein-DNA binary complex on the NRBC1536 dataset. Two vertical lines divide the sequence identity and structural similarity into four quadrants according to the threshold (see text).

Y = 0.25 (transition point of sequence identity). Points in the first quadrant represent complex structures that can be used to find the correct template by both sequence-based and structure-based methods. These complexes with relatively conserved sequences and structures not only have similar sequences but also similar structures. Points in the second quadrant represent complex structures with similar sequences but lower structural similarity. The target can discover template using sequence-based modeling methods in this quadrant, but structural modeling methods are relatively difficult to search the template correctly. In contrast to the first quadrant, points in the third quadrant represent the complexes whose sequence are dissimilar and 3D structures are also dissimilar. The complex structures located in this quadrant are currently unable to find the template quite correctly with sequence-based and structure-based methods. The points in the fourth quadrant represent complex structures with low similarity on sequence but high similarity on structure. In supplementary table 1, we enumerated 1267 structures with sequence similarity below 0.25, structural similarity greater than 0.4, iRMSD and lRMSD less than 10 Å. It is difficult to find the correct template with sequence-based methods, but conversely, a structure-based approach makes it easier to find the correct template. Studies have shown that proteins and DNAs undergo structural changes while they are docking [24–26]. Structures located in the fourth quadrant may be due to the large conformation changes in protein or DNA structure resulting in their structure to be similar.

3.3. The relationship between sequence, structure and function

We picked out four complex structures that are in the fourth quadrant described in the previous section (Fig. 4). In these examples, the sequence identity between the target and its template ranges from 0.162 to 0.205. The dissimilar sequence makes it difficult to identify their templates by the sequence-based method. But their structure can be aligned well, so it is convenient for structural template-based approach to get the template.

Proteins in the same family may bind with DNA in similar ways. In the first example, we examine the target structure 1R8D_AC (chains A and C for protein and DNA respectively) [27] and its template structure 4WLW_AY (protein chain is A, DNA chain is Y) [1]. Their sequence identity is 0.198, but they have similar 3D structures with the structure score 0.630. These two transcription factors belonged to the MerR family can activate or inhibit

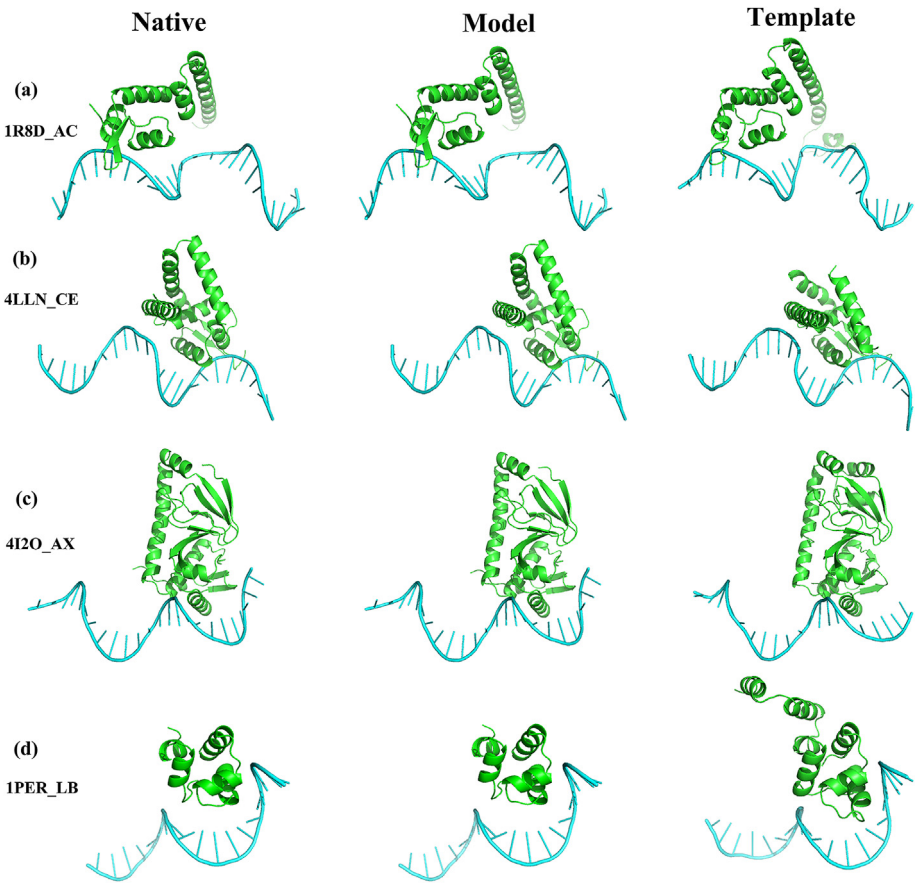

Fig. 4. Four examples with dissimilar sequences but similar structure between target and template. (a) The target structure 1R8D_AC (protein chain A, DNA strand C) hits its template structure 4WLW_AY. Their sequence identity is 0.198, and structural score is 0.630. The lRMSD between 4WLW_AY and its native structure is 3.17 Å. (b) The target structure 4LLN_CE takes 5HLG_AI21 as the template. Their sequence identity is 0.205 but their 3D structures are very similar, and their complex structure score equals to 0.775, 4LLN_CE obtains a lRMSD 2.82 Å with its native structure. (c) 4I2O_AX employs 3MZH_BC as the template, their sequence identity is 0.162, and the structural similarity is 0.739. Besides, their lRMSD is 1.84 Å (d) Target structure 1PER_LB used 4JCX_BD as the template. Although they are dissimilar in sequence, their structural similarity is 0.623. The lRMSD between 1PER_LB and its native structure is 1.26 Å.

transcription after binding to DNA. They have similar binding patterns. The N-terminal of their structure has a winged helix-turn-helix (wHTH) motif that is a DNA-binding domain. The HTH motif consists of a four-helix bundle and a three-stranded antiparallel β sheet [1,27].

The second example involves the target structure 4LLN_CE [2], and the template 5HLG_AI [28], their sequence identity is 0.205 (0.236 and 0.205 for protein and DNA identity respectively). Their 3D structures can be well matched. The TM-score is 0.801 and the RMscore is 0.775. They are the members of the MarR family and they bind to DNA at the wHTH DNA-binding domain to participate in transcriptional repression.

Although some similar structures are already known to be in the same family and share similar functions, we have found that there are some structures with unknown functions but have very similar structures. For example, we investigate the target 4I2O_AX [29] and the template 3MZH_BC. Their sequence identity is 0.162, but the structural similarity is 0.739. The function of 3MZH_BC is unknown. 4I2O is a FixK2-DNA complex. FixK 2, an import regulatory protein for α-proteobacterium Bradyrhizobium japonicum, belonging to the cAMP receptor protein (CRP) superfamily, is negatively controlled by oxidation of its single cysteine located next to the DNA-binding domain. 4I2O_AX and 3MZH_BC share similar 3D structures, so they may perform similar regulating functions.

In the last example, the target complex structure 1PER_LB [30], hits the template 4JCX_BD [31]. Although their sequence identity is low (protein identity is 0.198 and DNA identity is 0.111), they share similar 3D structures with structural similarity of 0.623. The functions of these two complexes are currently unknown, but according to the relationship between sequence, structure, and function, these two complexes may perform a similar function.

4. Discussion

Structural template-based methods are proposed for predicting protein-protein [6,7,20,32–35] and protein-RNA interactions [4,5]. In this study, PRIME 2.0.1 provides a structural template-based protein-DNA complex structure modeling method. The key point of this approach is structural similarity. Similar to the results in protein-RNA modeling, the all-to-all alignment in protein-DNA modeling also revealed that the predicted structure changed from random to similar binding mode due to the similarity between monomers. In many cases, structural alignment methods are more suitable than sequence-based methods, and DNA structures play an important role in protein-DNA complex structure prediction. Numerous protein-DNA complexes are dissimilar in sequences but with similar 3D structures which participate in the same biological processes and perform the same functions, such as transcription

factor-DNA complexes.

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China [31100522]; National High Technology Research and Development Program of China[2012AA020402]; the Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) [U1501501] and the Fundamental Research Funds for the Central Universities [2016YXMS017].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbrc.2021.09.018.

Author contributions

Conceptualization, JX, JFZ and SYL; Investigation JX, JFZ, XXT, XH and SYL; PRIME 2.0.1 software development, JFZ, JX and SYL; Data Curation, JX and JFZ; Writing - Original Draft, JX; Writing - Review & Editing, JX, JFZ, XH, XXT, XDL, QS, SL and SYL; Funding Acquisition, SYL; Supervision, SYL.

Accession codes

The source code PRIME 2.0.1 was freely available at http://www.rnabinding.com/PRIME2.0.1/PRIME2.0.1.html.

References

[1] S.J. Philips, M. Canalizo-Hernandez, I. Yildirim, G.C. Schatz, A. Mondragon, T.V. O'Halloran, TRANSCRIPTION. Allosteric transcriptional regulation via changes in the overall topology of the core promoter, Science 349 (2015) 877–881.
[2] I. Birukou, S.M. Seo, B.D. Schindler, G.W. Kaatz, R.G. Brennan, Structural mechanism of transcription regulation of the Staphylococcus aureus multidrug efflux operon mepRA by the MarR family repressor MepR, Nucleic Acids Res. 42 (2014) 2774–2788.
[3] N.M. Luscombe, R.A. Laskowski, J.M. Thornton, Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level, Nucleic Acids Res. 29 (2001) 2860–2874.
[4] J. Zheng, P.J. Kundrotas, I.A. Vakser, S. Liu, Template-based modeling of protein-RNA interactions, PLoS Comput. Biol. 12 (2016), e1005120.
[5] J. Zheng, J. Xie, X. Hong, S. Liu, RMalign: an RNA structural alignment tool based on a novel scoring function RMscore, BMC Genom. 20 (2019) 276.
[6] T. Vreven, H. Hwang, B.G. Pierce, Z. Weng, Evaluating template-based and template-free protein-protein complex structure prediction, Briefings Bioinf. 15 (2014) 169–176.
[7] A. Szilagyi, Y. Zhang, Template-based structure modeling of protein-protein interactions, Curr. Opin. Struct. Biol. 24 (2014) 10–23.
[8] Y. Yan, D. Zhang, P. Zhou, B. Li, S.Y. Huang, HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy, Nucleic Acids Res. 45 (2017) W365–W373.
[9] M. van Dijk, A.M. Bonvin, A protein-DNA docking benchmark, Nucleic Acids Res. 36 (2008) e88.
[10] L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, Cd-Hit, Accelerated for clustering the next-generation sequencing data, Bioinformatics 28 (2012) 3150–3152.
[11] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, Trends Genet. 16 (2000) 276–277.
[12] Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score, Nucleic Acids Res. 33 (2005) 2302–2309.
[13] I.A. Vakser, Low-resolution structural modeling of protein interactome, Curr. Opin. Struct. Biol. 23 (2013) 198–205.
[14] R. Sinha, P.J. Kundrotas, I.A. Vakser, Protein docking by the interface structure similarity: how much structure is needed? PloS One 7 (2012), e31349.
[15] R. Sinha, P.J. Kundrotas, I.A. Vakser, Docking by structural similarity at protein-protein interfaces, Proteins 78 (2010) 3235–3241.
[16] P.J. Kundrotas, I.A. Vakser, J. Janin, Structural templates for modeling homodimers, Protein Sci. 22 (2013) 1655–1663.
[17] P.J. Kundrotas, I.A. Vakser, Global and local structural similarity in protein-protein complexes: implications for template-based docking, Proteins 81 (2013) 2137–2142.
[18] M.F. Lensink, S. Velankar, A. Kryshtafovych, S.Y. Huang, D. Schneidman-Duhovny, A. Sali, J. Segura, N. Fernandez-Fuentes, S. Viswanath, R. Elber, S. Grudinin, P. Popov, E. Neveu, H. Lee, M. Baek, S. Park, L. Heo, G. Rie Lee, C. Seok, S. Qin, H.X. Zhou, D.W. Ritchie, B. Maigret, M.D. Devignes, A. Ghoorah, M. Torchala, R.A. Chaleil, P.A. Bates, E. Ben-Zeev, M. Eisenstein, S.S. Negi, Z. Weng, T. Vreven, B.G. Pierce, T.M. Borrman, J. Yu, F. Ochsenbein, R. Guerois, A. Vangone, J.P. Rodrigues, G. van Zundert, M. Nellen, L. Xue, E. Karaca, A.S. Melquiond, K. Visscher, P.L. Kastritis, A.M. Bonvin, X. Xu, L. Qiu, C. Yan, J. Li, Z. Ma, J. Cheng, X. Zou, Y. Shen, L.X. Peterson, H.R. Kim, A. Roy, X. Han, J. Esquivel-Rodriguez, D. Kihara, X. Yu, N.J. Bruce, J.C. Fuller, R.C. Wade, I. Anishchenko, P.J. Kundrotas, I.A. Vakser, K. Imai, K. Yamada, T. Oda, T. Nakamura, K. Tomii, C. Pallara, M. Romero-Durana, B. Jimenez-Garcia, I.H. Moal, J. Fernandez-Recio, J.Y. Joung, J.Y. Kim, K. Joo, J. Lee, D. Kozakov, S. Vajda, S. Mottarella, D.R. Hall, D. Beglov, A. Mamonov, B. Xia, T. Bohnuud, C.A. Del Carpio, E. Ichiishi, N. Marze, D. Kuroda, S.S. Roy Burman, J.J. Gray, E. Chermak, L. Cavallo, R. Oliva, A. Tovchigrechko, S.J. Wodak, Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment, Proteins 84 (Suppl 1) (2016) 323–348.
[20] P.J. Kundrotas, Z. Zhu, J. Janin, I.A. Vakser, Templates are available to model nearly all complexes of structurally characterized proteins, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 9438–9441.
[21] I. Russo Krauss, V. Spiridonova, A. Pica, V. Napolitano, F. Sica, Different duplex/quadruplex junctions determine the properties of anti-thrombin aptamers with mixed folding, Nucleic Acids Res. 44 (2016) 983–991.
[22] I. Russo Krauss, A. Pica, A. Merlino, L. Mazzarella, F. Sica, Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer, Acta Crystallogr D Biol Crystallogr 69 (2013) 2403–2411.
[23] M. Gao, J. Skolnick, DBD-Hunter: a knowledge-based method for the prediction of DNA-protein interactions, Nucleic Acids Res. 36 (2008) 3978–3992.
[24] S. Emamzadah, L. Tropia, I. Vincenti, B. Falquet, T.D. Halazonetis, Reversal of the DNA-binding-induced loop L1 conformational switch in an engineered human p53 protein, J. Mol. Biol. 426 (2014) 936–944.
[25] N.K. Vidangos, J. Heideker, A. Lyubimov, M. Lamers, Y. Huo, J.G. Pelton, J. Ton, J. Gralla, J. Berger, D.E. Wemmer, DNA recognition by a sigma(54) transcriptional activator from Aquifex aeolicus, J. Mol. Biol. 426 (2014) 3553–3568.
[26] S. Poddar, D. Chakravarty, P. Chakrabarti, Structural changes in DNA-binding proteins on complexation, Nucleic Acids Res. 46 (2018) 3298–3308.
[27] K.J. Newberry, R.G. Brennan, The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus, J. Biol. Chem. 279 (2004) 20356–20362.
[28] G. Liu, X. Liu, H. Xu, X. Liu, H. Zhou, Z. Huang, J. Gan, H. Chen, L. Lan, C.G. Yang, Structural insights into the redox-sensing mechanism of MarR-type regulator AbfR, J. Am. Chem. Soc. 139 (2017) 1598–1608.
[29] M. Bonnet, M. Kurz, S. Mesa, C. Briand, H. Hennecke, M.G. Grutter, The structure of Bradyrhizobium japonicum transcription factor FixK(2) unveils sites of DNA binding and oxidation, J. Biol. Chem. 288 (2013) 14238–14246.
[30] D.W. Rodgers, S.C. Harrison, The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites, Structure 1 (1993) 227–240.
[31] M.B. Shevtsov, S.D. Streeter, S.J. Thresh, A. Swiderska, J.E. McGeehan, G.G. Kneale, Structural analysis of DNA binding by C.Csp231I, a member of a novel class of R-M controller proteins regulating gene expression, Acta Crystallogr D Biol Crystallogr 71 (2015) 398–407.
[32] A.W. Ghoorah, M.D. Devignes, M. Smail-Tabbone, D.W. Ritchie, Spatial clustering of protein binding sites for template based protein docking, Bioinformatics 27 (2011) 2820–2827.
[33] I. Johansson-Åkhe, C. Mirabello, B. Wallner, Predicting Protein-Peptide Interaction Sites Using Distant Protein Complexes as Structural Templates, bioRxiv, 2018, p. 398768.
[34] S. Muratcioglu, E. Guven-Maiorov, O. Keskin, A. Gursoy, Advances in template-based protein docking by utilizing interfaces towards completing structural interactome, Curr. Opin. Struct. Biol. 35 (2015) 87–92.
[35] L.C. Xue, J. Rodrigues, D. Dobbs, V. Honavar, A. Bonvin, Template-based protein-protein docking exploiting pairwise interfacial residue restraints, Briefings Bioinf. 18 (2017) 458–466.