

Insights from Modeling the 3D Structure of DNA–CBF3b Complex

Kuo-Chen Chou*

Gordon Life Science Institute, 13784 Torrey Del Mar, San Diego, California 92130

Received May 9, 2005

A 3-dimensional model for a DNA–protein interaction has been developed. The protein component is the 61-residue fragment (res. 11–71) of subunit b of the yeast centromeric DNA binding factor 3, CBF3b. The CBF3b fragments bind to the 17 base pairs (5'-CGGAGGACTGTCTCCG-3') as a symmetric homodimer, with each folded into three distinct conformations: a compact, zinc-binding domain (res. 11–44); an extended linker (res. 45–57); and an α -helical dimerization element (res. 58–71). The DNA fragment in the complex is featured by a relatively straight conformation with only slight deviation from a standard B-structure, and a large part of the major groove not blocked by the protein. The large DNA open area provides the necessary space for the other subunits of CBF3 to bind coordinately with CBF3b, fully consistent with the observation that the cooperation of all four CBF3 components is absolutely required to constitute an activity that specifically interacts with centromere DNA. The model also provides a footing for further considering the possible binding arrangements of the other three subunits, namely CBF3a, CBF3c, and CBF3d.

Keywords: DNA–protein interaction • CBF3 • palindromic 17-bp DNA • Zn-linker-helix motif • centromere • kinetochore • segment matching

I. Introduction

The length of DNA is far greater than the size of the nucleus compartment in which it dwells. To make it fit into this compartment, the DNA has to be highly condensed thru various packing interactions. For example, the shortest human chromosome contains 4.6×10^7 base pairs of DNA, which is equivalent to 14 000 μm of extended DNA. In its most condensed state during mitosis, the chromosome is about 2 μm long, implying that the DNA is condensed with a packing ratio of $14\,000/2 = 7000$. To achieve the desired packing ratio, DNA is not packaged directly into the final structure of chromatin. Instead, it contains several hierarchies of organization thru interacting with different proteins in different manners.

Chromosome segregation in mitosis and meiosis depends on the interactions among the centromere, kinetochore, and spindle apparatus (see, e.g., ref 1). Centromeres are highly complex chromosomal substructures involved in essential aspects of chromosome transmission during cell division. Kinetochores are specialized protein complex thru which the replicated chromosomes will bind to a distinct subset of the microtubules of the mitotic spindle. The functional centromere DNA is organized into three domains: CDEI, CDEII, and CDEIII, of which CDEIII is essential for centromere function.¹

CBF3 (centromere DNA binding factor 3) is a key component of the *S. cerevisiae* kinetochore that binds selectively to CDEIII, a highly conserved DNA sequence found in centromeres and in several promoters, and that is essential for chromosome segregation.¹ The importance of CBF3 has been further strength-

ened by the fact that it is necessary for the movement of the centromeres along microtubules, that it plays a role in the attachment of chromosomes to the spindle, and that its mutations can result in a cell cycle arrest.¹

CBF3 consists of four subunits: CBF3a, CBF3b, CBF3c, and CBF3d. All the four subunits are indispensable for cell growth, and are considered to be core components of the yeast kinetochore. Although it is not quite clear how the CBF3 components interact with the centromere DNA to form the CBF3–DNA complex, the zinc finger protein CBF3b is most likely involved in the specific interaction of CBF3 with the centromere CDEIII according to the existing data. This is because CBF3b is the only one of the four subunits that contains a known DNA binding domain, a Zn_2Cys_6 type zinc finger domain that is known to recognize base triplets,² and is essential for chromosome segregation and the CBF3–DNA complex formation.³

Therefore, to understand the mechanism of chromosome segregation at a deeper level, and gain some insights into the roles and whereabouts of the other subunits during the interaction of CBF3 with the centromere DNA, it is instructive to first develop a 3D (dimensional) model of the binding interaction between the centromere DNA and CBF3b. The present study was initiated in an attempt to address this problem.

II. Method

Like CBF3b, the yeast protein GAL4, which binds to DNA as a transcriptional activator of several genes that encode galactose-metabolizing enzymes, also contains a Zn_2Cys_6 type zinc finger domain^{4,5} that is known to recognize base triplets. In other words, both CBF3b and GAL4 are among a set of homologous

* To whom correspondence should be addressed. E-mail: kchou@san.rr.com.

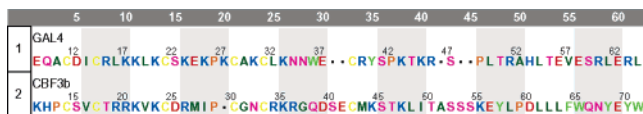


Figure 1. Sequence alignment between GAL4 and CBF3b. The alignment was performed by using the PILEUP program in the GCG package⁹ with the focus on the sequences that contains the putative binding sites with DNA. The amino acids are colored according to their function: acidic – red; basic – blue; neutral hydrophilic – pink; aliphatic – dark green; aromatic – light green; thiol containing – yellow; and imino – orange.

proteins with the conserved Zn₂Cys₆ fungal-type,⁶ and hence they should have the same DNA-binding motif. Accordingly, to develop the 3D structure of CBF3b–DNA complex, it is rational to use the crystal structure of GAL4–DNA complex as a template.⁷

The sequence of CBF3b (organism source *Saccharomyces cerevisiae*) was taken from refs 3 and 8. The entire sequence consists of 608 amino acid residues. The putative binding site of CBF3b with DNA begins at res. 14 and ends at res. 42, with a length of 29 residues (see Swiss-Prot code CBF3B_YEAST or accession number P40969). The sequence of GAL4 (organism source *Saccharomyces cerevisiae*) was taken from refs 4–5. The binding site of GAL4 with DNA begins at res.11 and ends at res.38, with a length of 28 residues (see Swiss-Prot code GAL4_YEAST or accession number P04386).

The sequence alignment between GAL4 and CBF3b was performed by the PILEUP program in the GCG package,⁹ with the focus on the fragments that contains the aforementioned putative binding regions with DNA. The aligned results are given in Figure 1, where the amino acids are colored according to their function: acidic – red; basic – blue; neutral hydrophilic – pink; aliphatic – dark green; aromatic – light green; thiol containing – yellow; and imino – orange.

On the basis of the sequence alignment (Figure 1) and the atomic coordinates of the crystal structure, 1d66.pdb, of GAL4–DNA complex,⁷ the 3D structure of CBF3b–DNA complex was derived by means of the “segment matching” or “coordinate reconstruction” approach.^{10–13} The rationale of the approach is based on the finding that most hexapeptide segments of protein structure can be clustered into only 100 structurally different classes.¹⁴ Thus, comparative models can be built by using a subset of atomic positions from a selected template structure as “guiding” positions, as well as by identifying and assembling short, all-atom segments that fit these guiding positions. The template structure is usually homologous to the targeted one and is preferably with a high structural resolution. The entire operation involves the following steps: (1) breaking the targeted chain into many short sequence segments; (2) searching the database, which contains more than 5200 high-resolution crystal protein structures, for matching the segments according to the sequence alignment of Figure 1 and the “guiding” positions of the template protein chain (1d66.pdb); (3) fitting the coordinates of the matched segments into the growing target structure under the monitor to avoid any van der Waal overlap until all atomic coordinates of the targeted structure were obtained; (4) repeating the process 10 times and generating an average model, followed by a global energy minimization to create the final 3D structure. The segment matching approach was previously used to model the structure of the protease domain of caspase-8¹⁵ before the crystal coordinates of caspase-3 were released.¹⁶ To deal with the

situation of lacking a proper template, the 3D structure of the catalytic domain of caspase-3 was first derived by using the crystal structure of caspase-1 as a template. Subsequently, the caspase-3 structure thus obtained was used as a template to further model the protease domain of caspase-8. After the crystal coordinates of caspase-3 protease domain were finally released and the crystal structure of the caspase-8 protease domain was determined,¹⁷ it was found that the RMSD (root-mean-square-deviation) for all the backbone atoms of the caspase-3 protease domain between the crystal and computed structures was 2.7 Å, while the corresponding RMSD was 3.1 Å for caspase-8, and only 1.2 Å for its core structure, indicating that the predicted structures of caspase-3 and -8 were quite close to their crystal structures. Shortly afterward, this method was successively applied to model the CARDs (caspase recruitment domains) of Apaf-1, Ced-4, and Ced-3 by using the NMR structure of the RAIDD CARD¹⁸ as a template, and to model the Cdk5–Nck5a* complex¹⁹ as well as the protease domain of caspase-9.²⁰ Two years after the computed Cdk5–Nck5a* complex structure was published,¹⁹ the crystal structure of the complex was determined.²¹ It was found that the predicted Cdk5 and the crystal Cdk5 are almost the same. Also, it was observed by these authors²¹ that, upon the binding of Cdk5 and Nck5a* (or p25), the buried surface area was 3400 Å², which was very close to 3461 Å² derived from the computed structure.^{19,22} Meanwhile, stimulated by the computed Cdk5–Nck5a*–ATP structure, the molecular truncation experiments were conducted²³ with the conclusion that the experimental results “confirm and extend specific aspects of the original predicted computer model”. The segment match approach was also used to predict the tertiary structure of β-secretase zymogen,²⁴ leading to a compelling elucidation of why the prodomain of β-secretase did not suppress activity like in a strict zymogen, as observed by Shi et al.²⁵ and Benjannet et al.²⁶ Recently, the segment match approach was further used to develop the 3D structures of extracellular domains for the subtypes 1, 2, 3, and 5 of GABA-A receptors,²⁷ and the structures thus obtained can be successfully used to clarify the long-standing ambiguity regarding the subunit arrangement (clockwise or counterclockwise) in the heteropentamers, providing useful insights for understanding the molecular action mechanism of the receptors.

III. Results and Discussion

The predicted 3D structure for the CBF3b–DNA complex is illustrated in Figure 2a and b. In Figure 2a the DNA fragment is colored in yellow, the A-chain of CBF3b in red, and the B-chain in purple. The DNA fragment contains 19 base pairs (bp) with sequence given below:

$$\left\{ \begin{array}{l} \text{CCGGAGGACAGTCCTCCGG} \\ \text{GGCCTCTGACAGGAGGCC} \end{array} \right\} \quad (1)$$

in which the following 17-bp palindrome

$$\left\{ \begin{array}{l} \text{CGGAGGACAGTCCTCCG} \\ \text{GCCTCTGACAGGAGGC} \end{array} \right\} \quad (2)$$

is the binding site for CBF3b. The CBF3b fragments bind to the 17-bp DNA site as a symmetric homodimer, with each folded into three distinct conformations (Figure 2): (1) a compact, zinc-binding domain (res. 11–44); (2) an extended linker (res. 45–57); and (3) an α-helical dimerization element (res. 58–71). The paired helices of the dimerization element

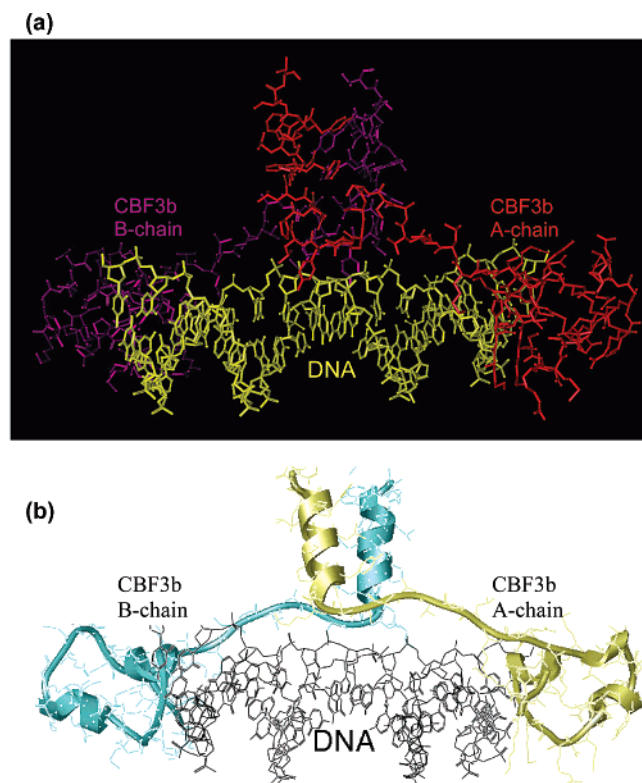


Figure 2. Computed 3D structure of the CBF3b-DNA complex: the CBF3b fragments bind to DNA as a symmetric homodimer. (a) The dot-and-stick drawing where the DNA fragment is colored in yellow, the A-chain of CBF3b in red, and the B-chain in purple. (b) The ribbon and dot-and-stick drawing where the DNA fragment is colored in black, the A-chain of CBF3b in yellow, and the B-chain in light blue. See the text for further explanation.

are almost parallel and projected away from the DNA along the 2-fold axis of the complex (Figure 2b). The compact domain is held together by two zinc ions, tetrahedrally coordinated by six cysteine residues thru the following mode: Cys-14, Cys-17, Cys-24, and Cys-30 are the ligands for one zinc ion, and Cys-14, Cys-30, Cys-33, and Cys-42, the ligands for the other; Cys-14 and Cys-30 are shared by both ions. The zinc-binding domain, lying in the major groove near each end of the DNA fragment, contacts its three base pairs, and hence plays an important role in recognizing and selectively binding the DNA site. As shown from Figure 2, the DNA binding site for CBF3b is a roughly symmetrical, 17-bp element with a highly conserved CCG sequence at either end, reading outward 5' to 3' from the dyad (see eq 2). Accordingly, the binding mode between DNA and CBF3b can be characterized as Zn-linker-helix motif.

Furthermore, the palindromic 17-bp DNA in the complex is featured by (1) a relatively straight conformation with only slight deviation from a standard B-structure, and (2) a considerably large part of the major groove not blocked by the protein. The large open area provides the necessary space for CBF3a, CBF3c, and CBF3d to bind coordinately with CBF3b, fully consistent with the observation that the cooperation of all four CBF3 components is absolutely required to constitute an activity that specifically interacts with centromere DNA.¹ These findings provide useful insights for further considering the possible binding arrangements for the other three subunits of CBF3.

IV. Conclusion

The DNA–protein interaction is one of the most important processes in the cycle of life because the accessibility of genetic information depends on the ability of proteins to recognize and interact with DNA. The binding interaction of CBF3b with DNA is featured by the Zn-linker-helix motif. There are many other different DNA–protein interaction motifs, such as helix-turn-helix motif,²⁸ helix-loop-helix motif,²⁹ β sheet motif,³⁰ and leucine zipper motif.³¹ Even for the zinc finger motif there are some other types,³² such as helix–Zn–sheet type,³³ and helix–Zn–helix type.³⁴ All these motifs and binding types share an important feature: they use either an α helix or a β sheet to recognize the major groove of the DNA, fitting into or making important contacts with it. As a consequence, the major groove would be mostly blocked. In contrast to that, the current Zn-linker–helix motif, or the Zn_2Cys_6 fungal type, uses a small Zn-containing domain to recognize a conserved CCG triplet at each end of the site thru direct contacts with the major groove and use an extended linker to bring the helix away from the DNA. Such a binding motif actually leaves the large part of the DNA major groove open, thereby making the necessary accessible space available for the other three subunits of CBF3. The 3D model developed here can serve as a footing for further investigation into the interactions between DNA and CBF3, particularly for the in-depth understanding how the other subunits are involved during the interaction process.

Abbreviations: 3D, 3-dimensional; CBF3, centromere DNA binding factor 3; bp, base pair.

Acknowledgment. The author would like to thank the two anonymous referees whose constructive comments have greatly strengthened the presentation of this paper.

References

- Lechner, J.; Ortiz, J. Minireview: The *Saccharomyces cerevisiae* kinetochore. *FEBS Lett.* **1996**, *389*, 70–74.
- Wu, H.; Yang, W.; Barbas, C. F. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 344–348.
- Lechner, J. A zinc finger protein, essential for chromosome segregation, constitutes a putative DNA binding subunit of the *Saccharomyces cerevisiae* kinetochore complex, Cbf3. *EMBO J.* **1994**, *13*, 5203–5211.
- Laughon, A.; Gesteland, R. F. Primary structure of the *Saccharomyces cerevisiae* GAL4 gene. *Mol. Cell. Biol.* **1984**, *4*, 260–267.
- Gadhavi, P. L.; Raine, A. R. C.; Alefounder, P. R.; Laue, E. D. Complete assignment of the 1H NMR spectrum and secondary structure of the DNA binding domain of GAL4. *FEBS Lett.* **1990**, *276*, 49–53.
- Pan, T.; Coleman, J. E. GAL4 transcription factor is not a “zinc finger” but forms a $\text{Zn}(2)\text{-Cys}(6)$ binuclear cluster. *Proc. Natl. Acad. Sci., U.S.A.* **1990**, *87*, 2077–2081.
- Marmorstein, R.; Carey, M.; Ptashne, M.; Harrison, S. C. DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* **1992**, *356*, 408–414.
- Strunnikov, A. V.; Kingsbury, J.; Koshland, D. CEP3 encodes a centromere protein of *Saccharomyces cerevisiae*. *J. Cell Biol.* **1995**, *128*, 749–760.
- Devereux, J.; Genetic Computer Group (GCG): Madison, Wisconsin, 1994.
- Claessens, M.; Cutsem, E. V.; Lasters, I.; Wodak, S. Modelling the polypeptide backbone with ‘spare parts’ from known protein structures. *Protein Eng.* **1989**, *4*, 335–345.
- Jones, T. A.; Thirup, S. Using known substructures in protein model building and crystallography. *EMBO J.* **1986**, *5*, 819–822.
- Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. Knowledge-based prediction of protein structures and design of novel molecules. *Nature (London)* **1987**, *326*, 347–352.
- Chou, K. C.; Nemethy, G.; Pottle, M.; Scheraga, H. A. Energy of stabilization of the right-handed $\beta\beta$ crossover in proteins. *J. Mol. Biol.* **1989**, *205*, 241–249.

- (14) Unger, R.; Harel, D.; Wherland, S.; Sussman, J. L. A 3-D building blocks approach to analyzing and predicting structure of proteins. *Proteins: Struct. Funct. Bioinform.* **1989**, *5*, 355–373.
- (15) Chou, K. C.; Jones, D.; Heinrichson, R. L. Prediction of the tertiary structure and substrate binding site of caspase-8. *FEBS Lett.* **1997**, *419*, 49–54.
- (16) Rotonda, J.; Nicholson, D. W.; Fazil, K. M.; Gallant, M.; Gareau, Y.; Labelle, M.; Peterson, E. P.; Rasper, D. M.; Ruel, R.; Vaillancourt, J. P.; Thornberry, N. A.; Becker, J. W. The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis. *Nat. Struct. Biol.* **1996**, *3*, 619–625.
- (17) Watt, W.; Koeplinger, K. A.; Mildner, A. M.; Heinrichson, R. L.; Tomasselli, A. G.; Watenpaugh, K. D. The atomic resolution structure of human caspase-8, a key activator of apoptosis. *Structure* **1999**, *7*, 1135–1143.
- (18) Chou, J. J.; Matsuo, H.; Duan, H.; Wagner, G. Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. *Cell* **1998**, *94*, 171–180.
- (19) Chou, K. C.; Watenpaugh, K. D.; Heinrichson, R. L. A Model of the complex between cyclin-dependent kinase 5(Cdk5) and the activation domain of neuronal Cdk5 activator. *Biochem. Biophys. Res. Commun.* **1999**, *259*, 420–428.
- (20) Chou, K. C.; Tomasselli, A. G.; Heinrichson, R. L. Prediction of the Tertiary Structure of a Caspase-9/Inhibitor Complex. *FEBS Lett.* **2000**, *470*, 249–256.
- (21) Tarricone, C.; Dhavan, R.; Peng, J.; Areces, L. B.; Tsai, L. H.; Musacchio, A. Structure and regulation of the Cdk5-p25 (Nck5a) complex. *Mol. Cell* **2001**, *8*, 657–669.
- (22) Chou, K. C. Review: Structural bioinformatics and its impact to biomedical science. *Curr. Med. Chem.* **2004**, *11*, 2105–2134.
- (23) Zhang, J.; Luan, C. H.; Chou, K. C.; Johnson, G. V. W. Identification of the N-terminal functional domains of Cdk5 by molecular truncation and computer modeling. *Proteins: Struct. Funct. Genet.* **2002**, *48*, 447–453.
- (24) Chou, K. C.; Howe, W. J. Prediction of the tertiary structure of the beta-secretase zymogen. *BBRC* **2002**, *292*, 702–708.
- (25) Shi, X. P.; Chen, E.; Yin, K. C.; Na, S.; Garsky, V. M.; Lai, M. T.; Li, Y. M.; Platchek, M.; Register, R. B.; Sardana, M. K.; Tam, M. J.; Thiebeau, J.; Wood, T.; Shafer, J. A.; Gardell, S. J. The pro domain of beta-secretase does not confer strict zymogen-like properties but does assist proper folding of the protease domain. *J. Biol. Chem.* **2001**, *276*, 10366–10373.
- (26) Benjannet, S.; Elagöz, A.; Wickham, L.; Mamarbachi, M.; Munzer, J. S.; Basak, A.; Lazure, C.; Cromlish, J. A.; Sisodia, S.; Checler, F.; Chrétien, M.; Seidah, N. G. Post-translational Processing of beta-Secretase-Amyloid-converting Enzyme) and Its Ectodomain Shedding. *J. Biol. Chem.* **2001**, *276*, 10879–10887.
- (27) Chou, K. C. Modelling extracellular domains of GABA-A receptors: subtypes 1, 2, 3, and 5. *Biochem. Biophys. Res. Commun.* **2004**, *316*, 636–642.
- (28) Harrison, S. C.; Aggarwal, A. K. DNA recognition by protein with the helix-turn-helix motif. *Annu. Rev. Biophys. Biochem.* **1990**, *59*, 933–969.
- (29) Benezra, R.; Davis, R. L.; Lockshon, D.; Turner, D. L.; Weintraub, H. The protein ID: a negative regulator of helix-loop-helix DNA-binding proteins. *Cell* **1990**, *61*, 49–59.
- (30) Knight, K. L.; Sauer, R. T. Biochemical and genetic analysis of operator contacts made by residues within the beta-sheet DNA binding motif of Mnt repressor. *EMBO J.* **1992**, *11*, 215–223.
- (31) Alber, T. Protein-DNA interactions: how GCN4 binds DNA. *Curr. Biol.* **1993**, *3*, 182–184.
- (32) Coleman, J. E. Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins. *Annu. Rev. Biochem.* **1992**, *61*, 897–946.
- (33) Lee, M. S.; Gippert, G. P.; Soman, K. V.; Case, D. A.; Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **1989**, *245*, 635–637.
- (34) Pavletich, N. P.; Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **1991**, *252*, 809–817.

PR050135+