

Supplementary Information

DNA–TCP complex structures reveal a unique recognition mechanism for TCP transcription factor families

Yi Zhang¹, Yong-ping Xu¹, Ju-kui Nie¹, Hong Chen¹, Genji Qin¹, Bo Wang¹, and
Xiao-Dong Su^{1, *}

¹ State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences,
and Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing,
100871, China.

* Lead contact

* Correspondence and requests for materials should be addressed to X-D. S (email:
xdsu@pku.edu.cn).

Contents

Supplementary Table 1-2

Supplementary Figure 1-9

Supplementary Table 1. Data collection and refinement statistics.

	TCP10-49M& 12bp class II DNA	TCP10-DBD& three-site class II DNA	TCP10-DBD& one-site class II DNA	TCP10-DBD& two-site class II DNA	TCP10-DBD& 1M class II DNA
Data Collection Statistics					
Space Group	P6 ₁ 22	C2	P3 ₂ 21	P2 ₁	P2 ₁
Wavelength (Å)	1.1000	0.9792	1.0800	1.0800	1.0800
Number of Reflections	258574 (25559)	119561(9705)	171649 (15391)	123838(11594)	119391(12397)
Cell dimensions					
a, b, c (Å)	68.1 68.1 210.8	74.6 76.8 46.0	78.8 78.8 82.8	78.8 76.9 83.1	78.6 76.9 83.2
α , β , γ (°)	90.0 90.0 120.0	90.0 103.6 90.0	90.0 90.0 120.0	90.0 113.9 90.0	90.0 113.7 90.0
Resolution (Å)	29.21 - 2.90 (3.01 - 2.90)	44.72 - 1.92 (1.99 - 1.92)	35.56-2.65 (2.75-2.65)	29.10-2.99 (3.10-2.99)	29.16-3.04 (3.15-3.04)
Rmerge (%)	11.7 (61.5)	10.3 (43.4)	9.0 (81.4)	16.5 (76.6)	10.0 (46.2)
CC _{1/2}	0.999 (0.975)	0.996 (0.866)	0.999 (0.978)	0.986 (0.754)	0.997 (0.867)
I/ σ (I)	38.8 (8.9)	13.7 (2.2)	27.6 (2.82)	8.8 (2.8)	14.5 (4.7)
Completeness (%)	99.5 (97.0)	99.8 (98.6)	96.4 (68.0)	96.8 (77.9)	97.1 (86.1)
Multiplicity	37.0 (39.1)	6.2 (5.1)	19.2(17.9)	6.7 (6.9)	6.8 (7.2)
Wilson B-factor	52.17	29.39	39.81	63.86	69.73
Refinement Statistics					
Reflections used in refinement	6982 (653)	19272 (1888)	8616 (586)	17928 (1428)	17178 (1515)
Reflections used for R _{free}	699 (65)	1928 (189)	862 (57)	1783 (145)	1732 (154)
R _{work} /R _{free} %	0.18/0.24 (0.19/0.26)	0.18/0.21 (0.27/0.35)	0.21/0.25 (0.32/0.34)	0.21/0.27 (0.30/0.39)	0.21/0.26 (0.27/0.34)
No. non-hydrogen atoms	1442	1826	1323	4487	4779
Macromolecules	1442	1712	1323	4487	4779
Solvent	-	114	-	-	-
Protein residues	119	141	102	370	391
B factor (Å ²)					
Protein	49.73	34.56	41.06	59.05	67.37
Solvent	-	37.25	-	-	-
R.m.s deviations					
Bond Lengths (Å)	0.008	0.008	0.011	0.011	0.012
Bond Angles (°)	0.94	1.02	1.07	1.12	1.19
Ranachandran (%)					
Preferred region	96.46	100.00	97.96	95.25	93.83
Allowed region	3.54	0.00	2.04	4.75	6.17
Outliers	0.00	0.00	0.00	0.00	0.00

	TCP15-DBD	TCP15-DBD& 12bp class I DNA
Data Collection Statistics		
Space Group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	1.1000	0.9792
Number of Reflections	89783 (8355)	667441 (64406)
Cell dimensions		
a, b, c (Å)	26.6 51.6 71.5	71.2 112.8 156.8
α, β, γ (°)	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	19.72 – 2.00 (2.07 – 2.00)	18.05-3.00 (3.11-3.00)
Rmerge (%)	15.2 (115.6)	26.0 (83.4)
CC _{1/2}	0.998 (0.861)	0.998 (0.994)
I/σ (I)	14.2 (2.1)	19.25 (3.62)
Completeness (%)	99.8 (100)	93.3 (71.9)
Multiplicity	12.8 (12.8)	25.9 (25.3)
Wilson B-factor	26.47	68.33
Refinement Statistics		
Reflections used in refinement	7075 (693)	24198 (1832)
Reflections used for R _{free}	708 (69)	1210 (111)
R _{work} /R _{free} %	0.18/0.22 (0.22/0.31)	0.20/0.25 (0.30/0.39)
No. non-hydrogen atoms	728	5399
Macromolecules	708	5399
Solvent	20	-
Protein residues	92	475
B factor (Å ²)		
Protein	34.69	60.84
Solvent	36.48	-
R.m.s deviations		
Bond Lengths (Å)	0.009	0.012
Bond Angles (°)	0.96	1.16
Ranachandran (%)		
Preferred region	96.59	94.55
Allowed region	3.41	5.45
Outliers	0.00	0.00

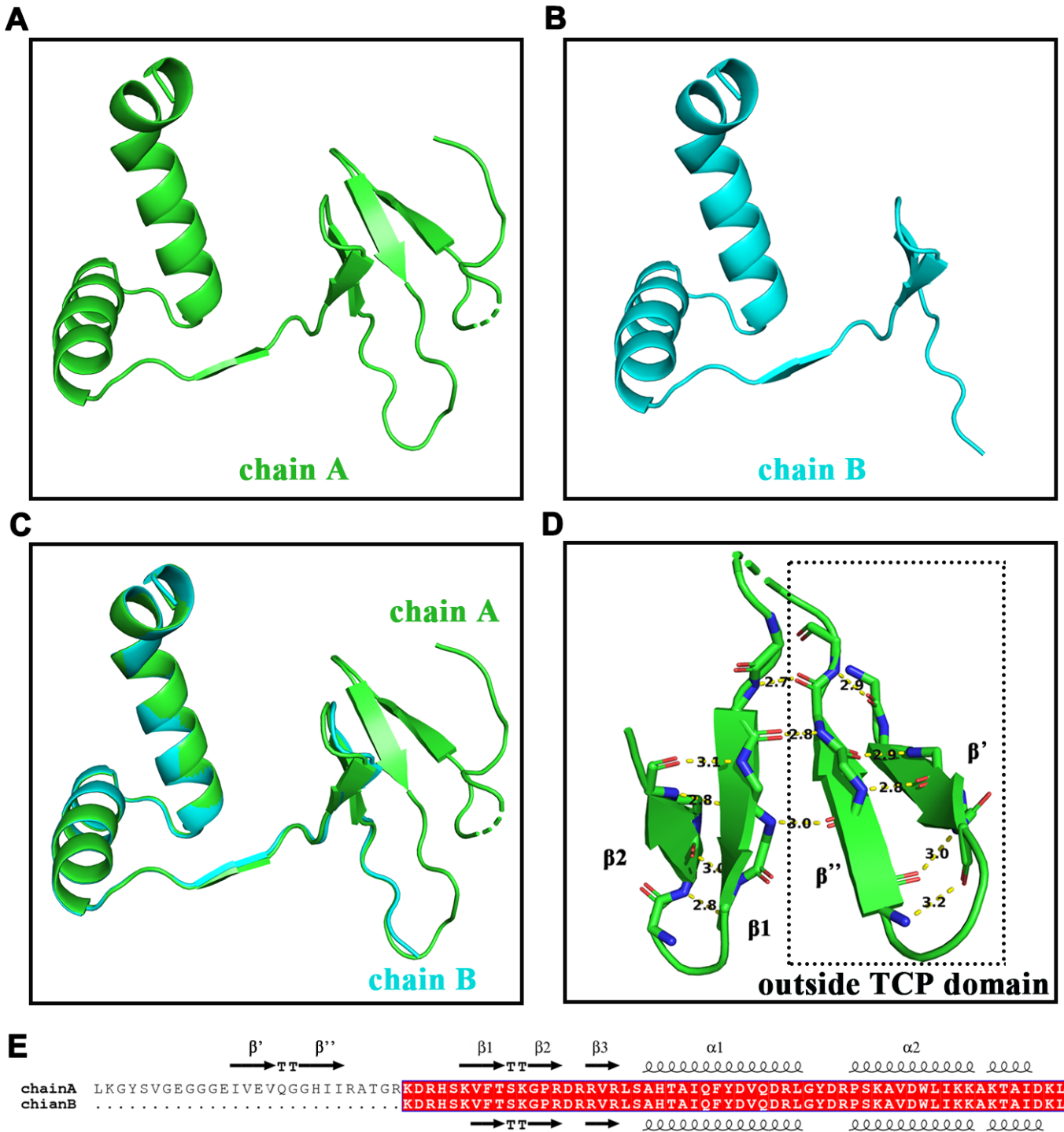
Values in bracket indicate those of the highest resolution shell.

Supplementary Table 2. DNA sequences used in each assay.

class II DNA	Sequence	
three-site class II DNA	5'-AT <u>GTGGTCCCCAC</u> T-3'	3'-A CACCAGGGGTG AT-5'
two-site class II DNA	5'-AT <u>GTGGTCCCGTG</u> T-3'	3'-A CACCAGGGCAC AT-5'
one-site class II DNA	5'-AT AGAG <u>TCCCGTG</u> T-3'	3'-A TCTCAGGGCAC AT-5'
one-site class II DNA (for crystallization)	5'-GAG <u>GCCCCCCC</u> CATAATA-3'	3'-TCTCCGGGGGGGTATTA-5'
1M class II DNA	5'-AT <u>GTGGTCCCCAG</u> T-3'	3'-A CACCAGGGCAC AT-5'
12 bp class II DNA	5'-AT <u>GTGGTCCCC</u> -3'	3'-A CACCAGGGG T-5'

class I DNA	Sequence	
three-site class I DNA	5'-AT <u>GTGGGTCCCAC</u> T-3'	3'-A CACCCAGGGTG AT-5'
two-site class I DNA	5'-AT <u>GTGGGTCCCGTG</u> T-3'	3'-A CACCCAGGCAC AT-5'
one-site class I DNA	5'-AT AGAG <u>GTCCCGTG</u> T-3'	3'-A TCTCCAGGCAC AT-5'
12 bp class I DNA	5'-AT <u>GTGGGTCCCC</u> -3'	3'-A CACCCAGGGG T-5'

Non-specific DNA	Sequence	
1#	5'- AT ACTAAAAATGT T-3'	3'-A TGATTTTTTACAA T-5'
2#	5'- AT ACAAAAAAAGT T-3'	3'-A TGTTTTTTTCAA T-5'

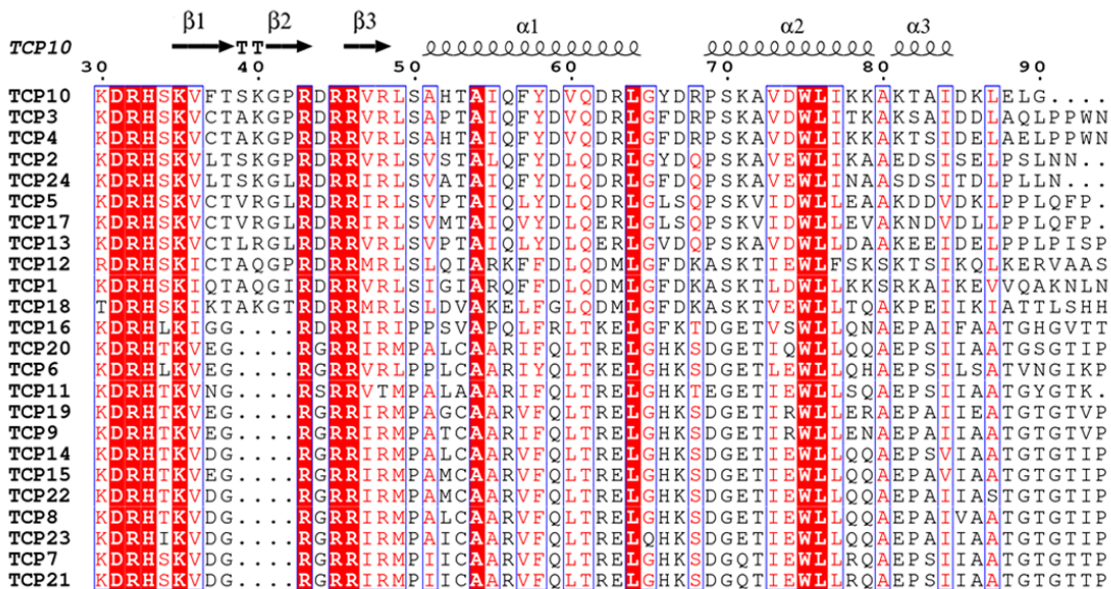
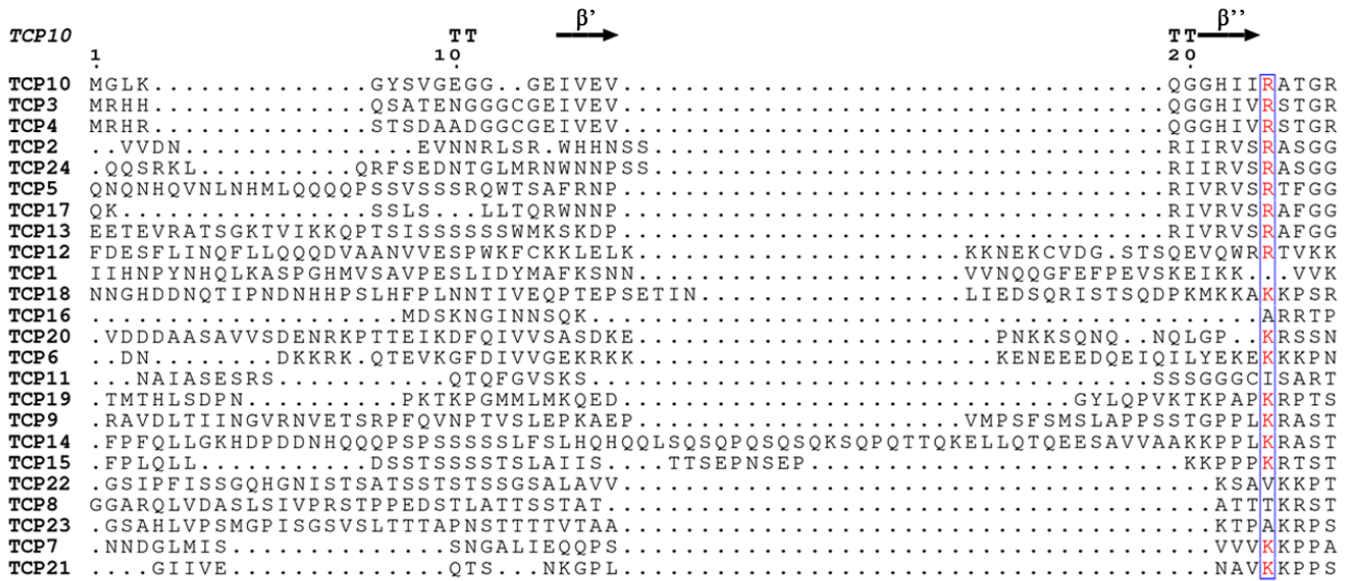


Supplementary Figure 1. Structure alignment of chain A and chain B of AtTCP10-DBD.

A–C, Structure alignment of chain A (green) & chain B (cyan) of AtTCP10-DBD. Chain A has 27 amino acids more than chain B in the N-terminal, forming two β strands (β' and β'').

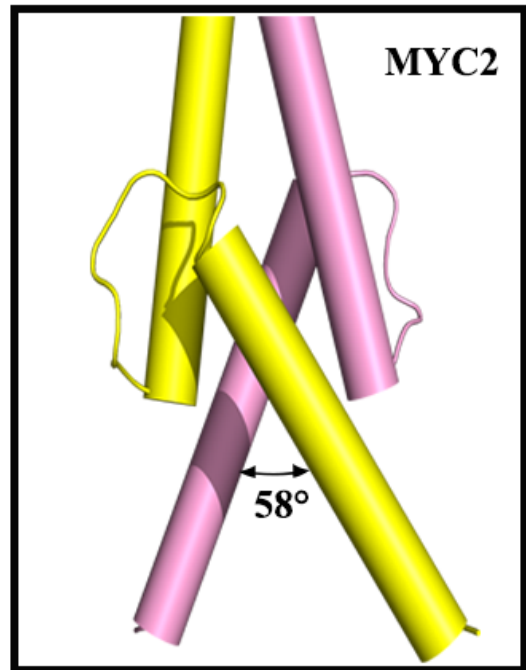
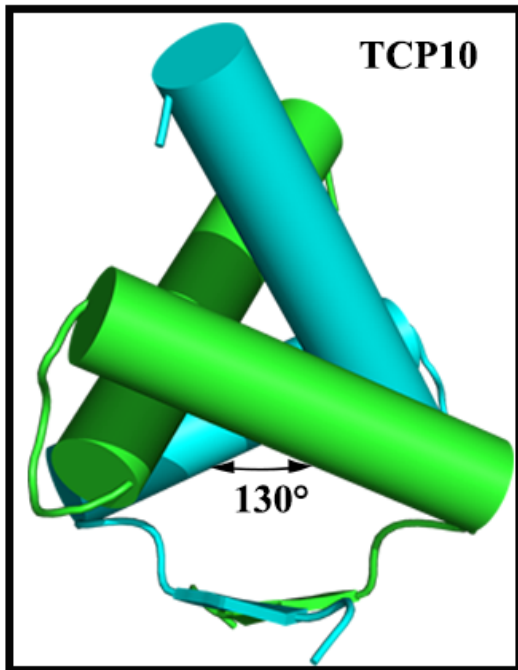
D, The β sheet formed by β' , β'' , $\beta 1$ and $\beta 2$ in chain A.

E, Sequence alignment of chain A & chain B of AtTCP10-DBD.



Supplementary Figure 2. Sequence alignments of AtTCP10 (residue 1–90) with other 23 members.

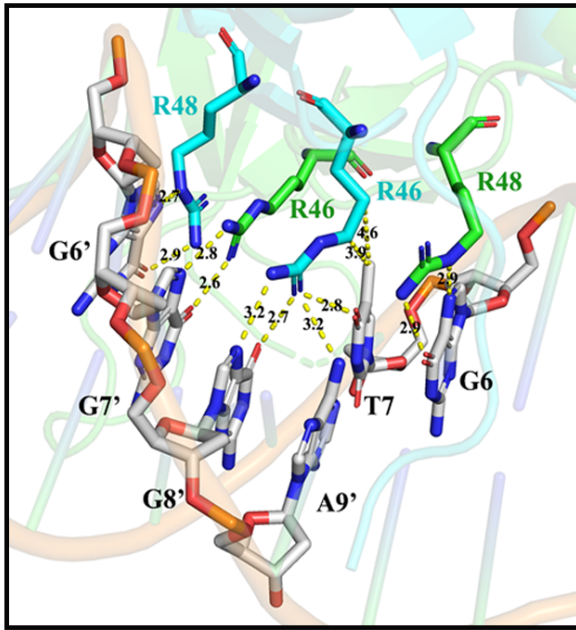
Besides β1–β3 and α1–α2, there are two more β strands (β' and β'') in the structure. From sequence conservation, the β' and β'' may only exist in the small branch of TCP10/TCP3/TCP4, other structural motif/topology may occur in other subfamilies.



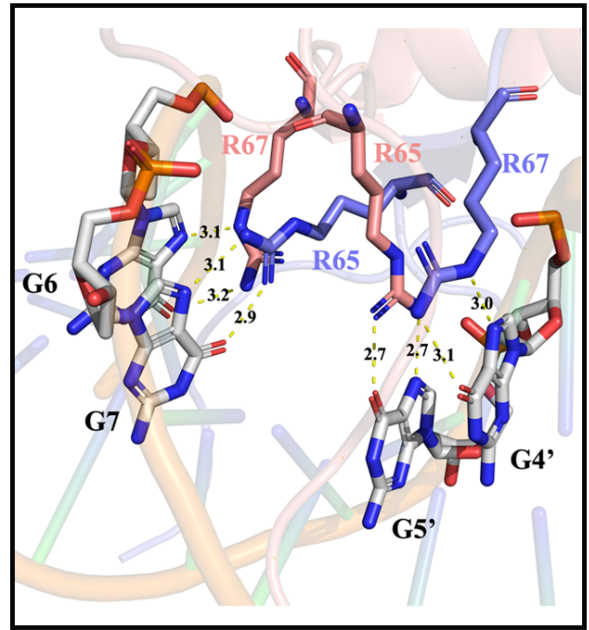
Supplementary Figure 3. Comparison of TCP domain with bHLH domain.

Comparison of the tilt angles of two $\alpha 1$ helices between TCP10 and canonical bHLH protein (MYC2).

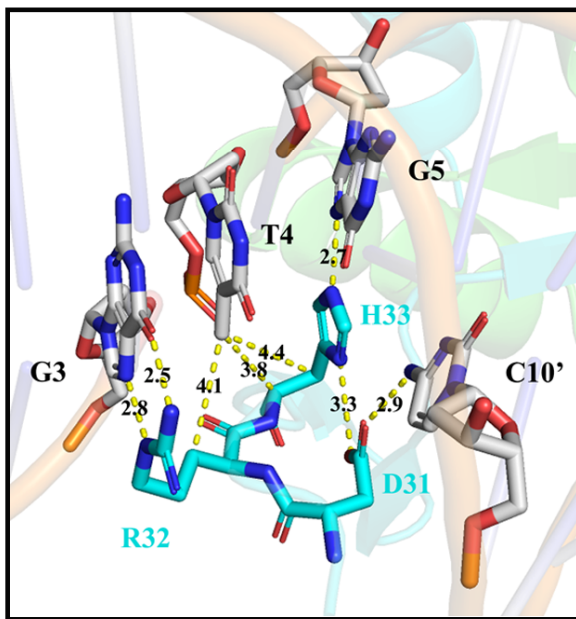
A RVR saddle



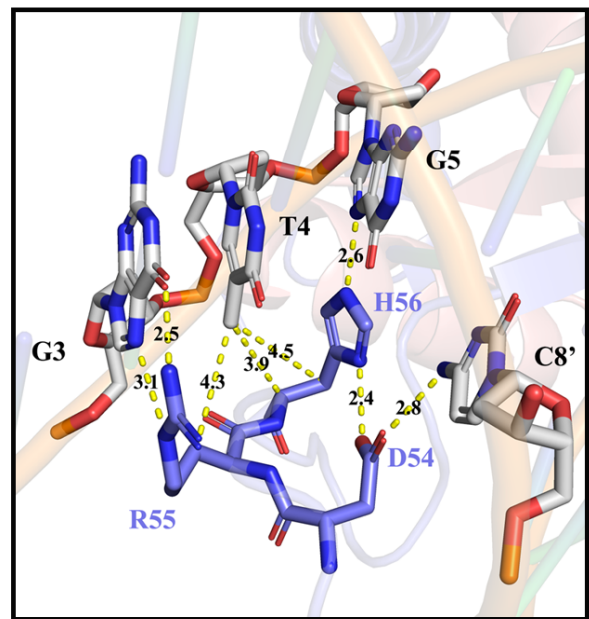
C RIR saddle



B DRH loop



D DRH loop



Supplementary Figure 4. The details of interactions in AtTCP10-DBD complex and AtTCP15-DBD complex.

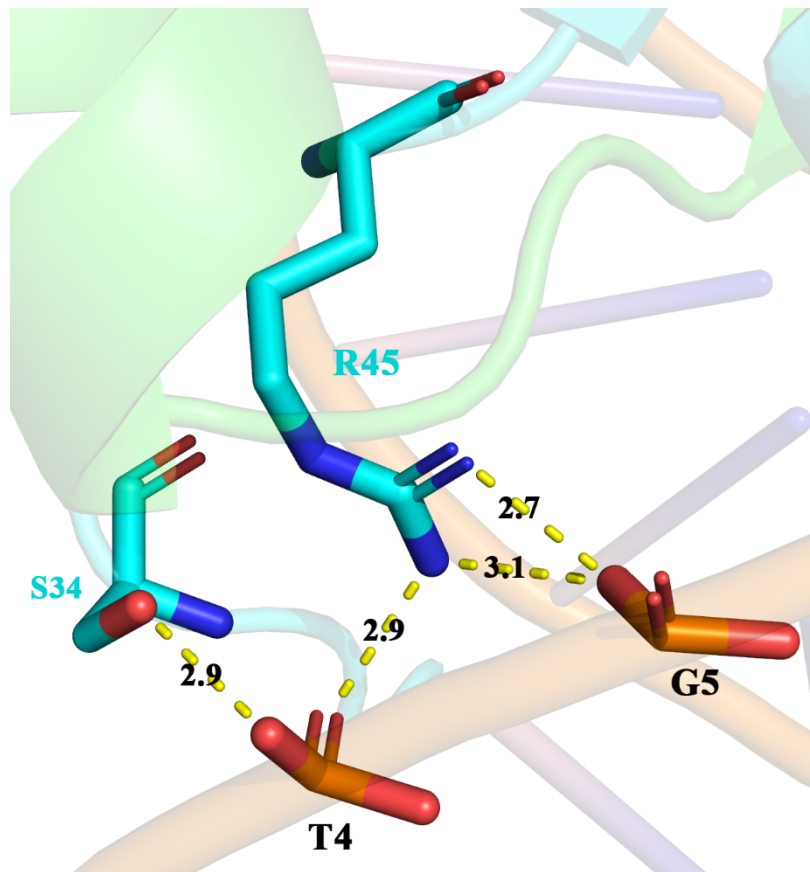
A, Base-specific interactions in the RVR saddle of AtTCP10-DBD complex are shown as follows: Arg^{46A} (green), Arg^{48A} (green) and Arg^{48B} (cyan) recognize G7', G6 and G6' respectively. Arg^{46B} (cyan) recognizes T7, G8' and A9'. Among these base-specific interactions, the following hydrogen bonds play important roles: the side chain nitrogen of Arg⁴⁸ with O6 and N7 of G6 and G6'; the guanidyl of Arg^{46A} with O6 and N7 of G7'; the guanidyl of Arg^{46B} with O6 and N7 of G8'; the NH₂⁺ group of Arg^{46B} with the hydroxyl of A9' and O4 of T7. Furthermore, Arg^{46B} also forms hydrophobic interactions with the T7 methyl group.

B, Interactions in one of the DRH loops of AtTCP10-DBD complex are shown as follows: Asp³¹ recognizes C10'; Arg³² recognizes G3 and T4; and His³³ recognizes G5 and T4. Asp³¹ also interacts with His³³. The following bonds play important roles: the carboxyl group of Asp³¹ with the amino groups of C10 and C10'; Arg³² with G3 and G3'; the imidazole group of His³³ with G5 and G5'. The side chains of Arg³² and His³³ also form hydrophobic interactions with the methyl groups of T4 and T4'.

C, Base-specific interactions in the RIR saddle of AtTCP15-DBD complex are shown as follows: Arg^{67A} (pink), Arg^{67B} (slate), Arg^{65A} (pink) and Arg^{65B} (slate) recognize G6, G4', G7 and G5' respectively.

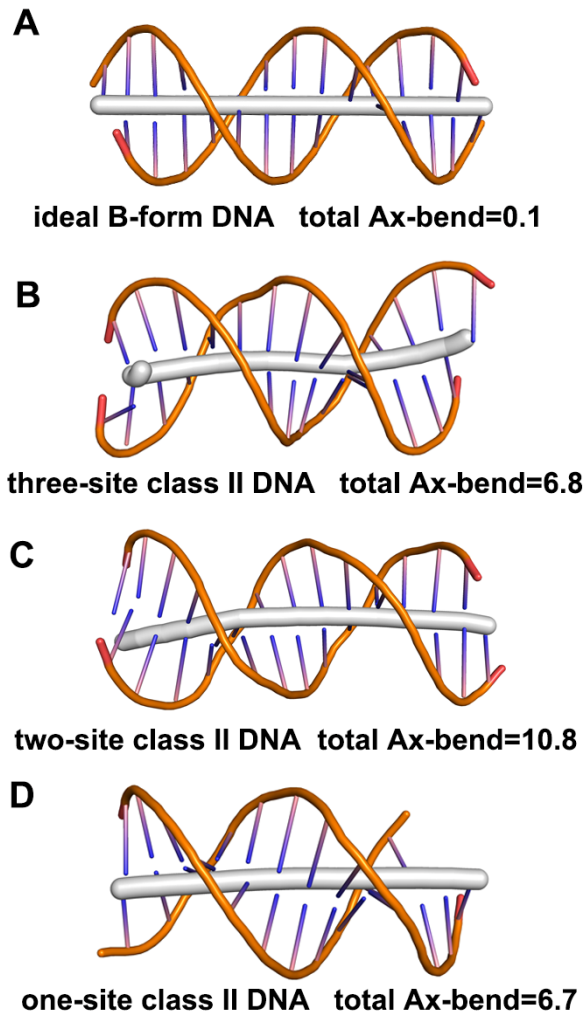
D, Interactions in one of the DRH loops of AtTCP15-DBD complex are shown as follows: Asp⁵⁴ recognizes C8'; Arg⁵⁵ recognizes G3 and T4; and His⁵⁶ recognizes G5 and T4. Asp⁵⁴ also interacts with His⁵⁶.

The distances (Å) of all the hydrogen bonds and hydrophobic interactions with bases are shown.



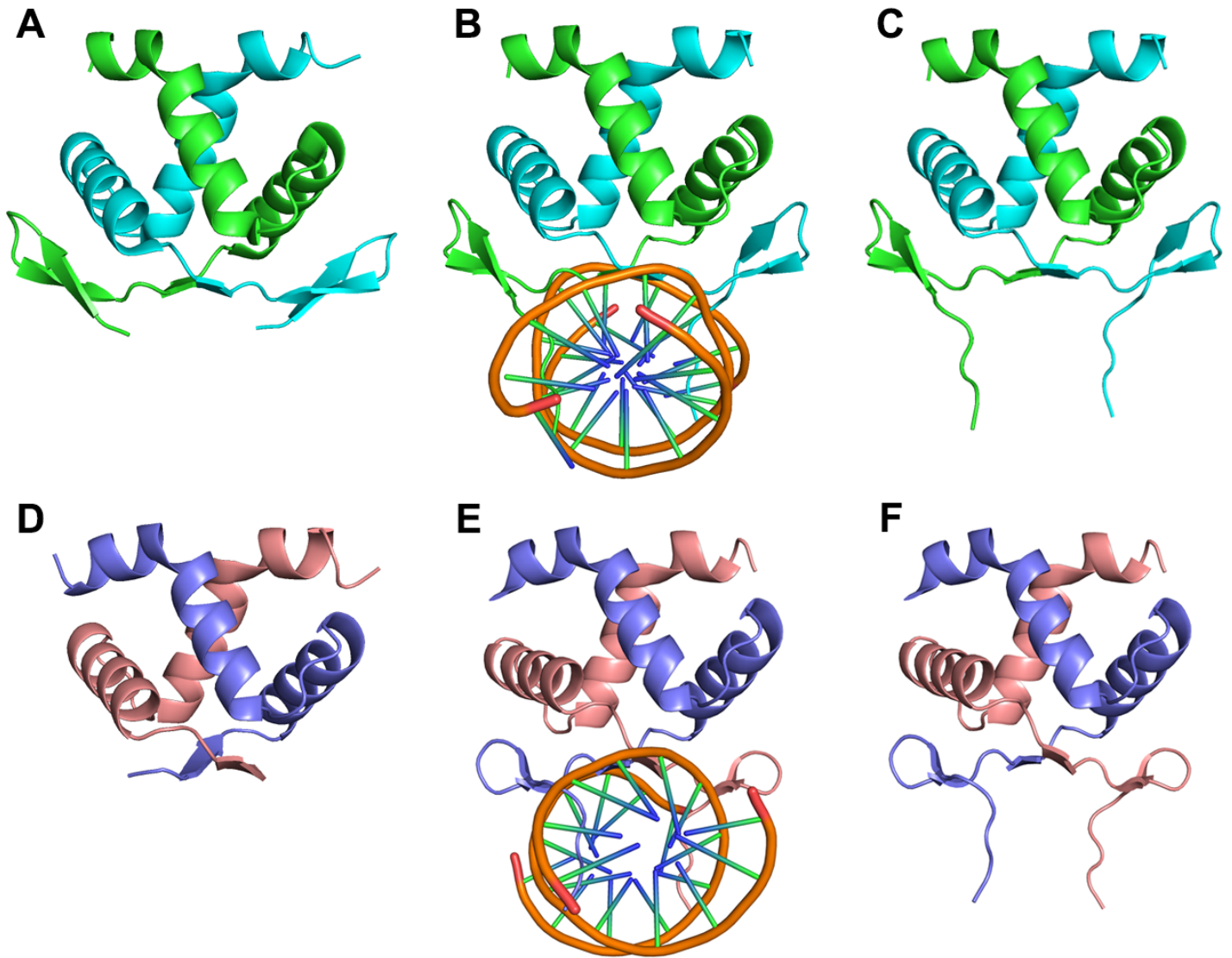
Supplementary Figure 5. The interactions between protein residues and DNA backbone.

Arg⁴⁵ has electrostatic interactions with T4 & T4' and G5 & G5'. H-bonds include Arg⁴⁵ to T4 & T4' and G5 & G5', Ser34 to T4 & T4'



Supplementary Figure 6. The helical axis and the total value of Ax-bend of ideal B-form DNA and bound class II three-site, two-site, one-site DNA.

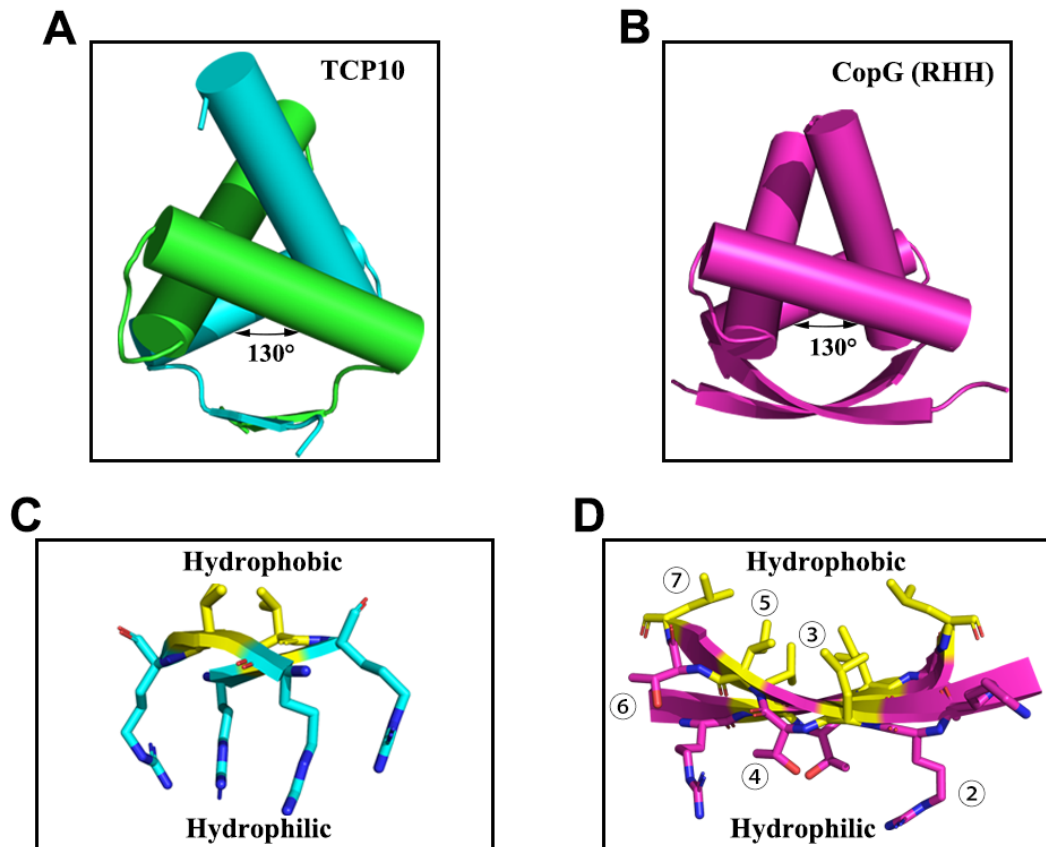
The helical axis and the total value of Ax-bend (the angle between axes of the DNA helix in two adjacent base pairs) of (A) ideal B-form DNA and (B) three-site class II DNA, (C) two-site class II DNA, (D) one-site class II DNA from complex structures.



Supplementary Figure 7. Comparison of the class II and class I TCP domain in the apo form and in complex with DNA.

A–C, The apo structure of class II OsPCF6 (5zkt), the complex structure of class II AtTCP10-DBD with DNA, and the complex structure without DNA

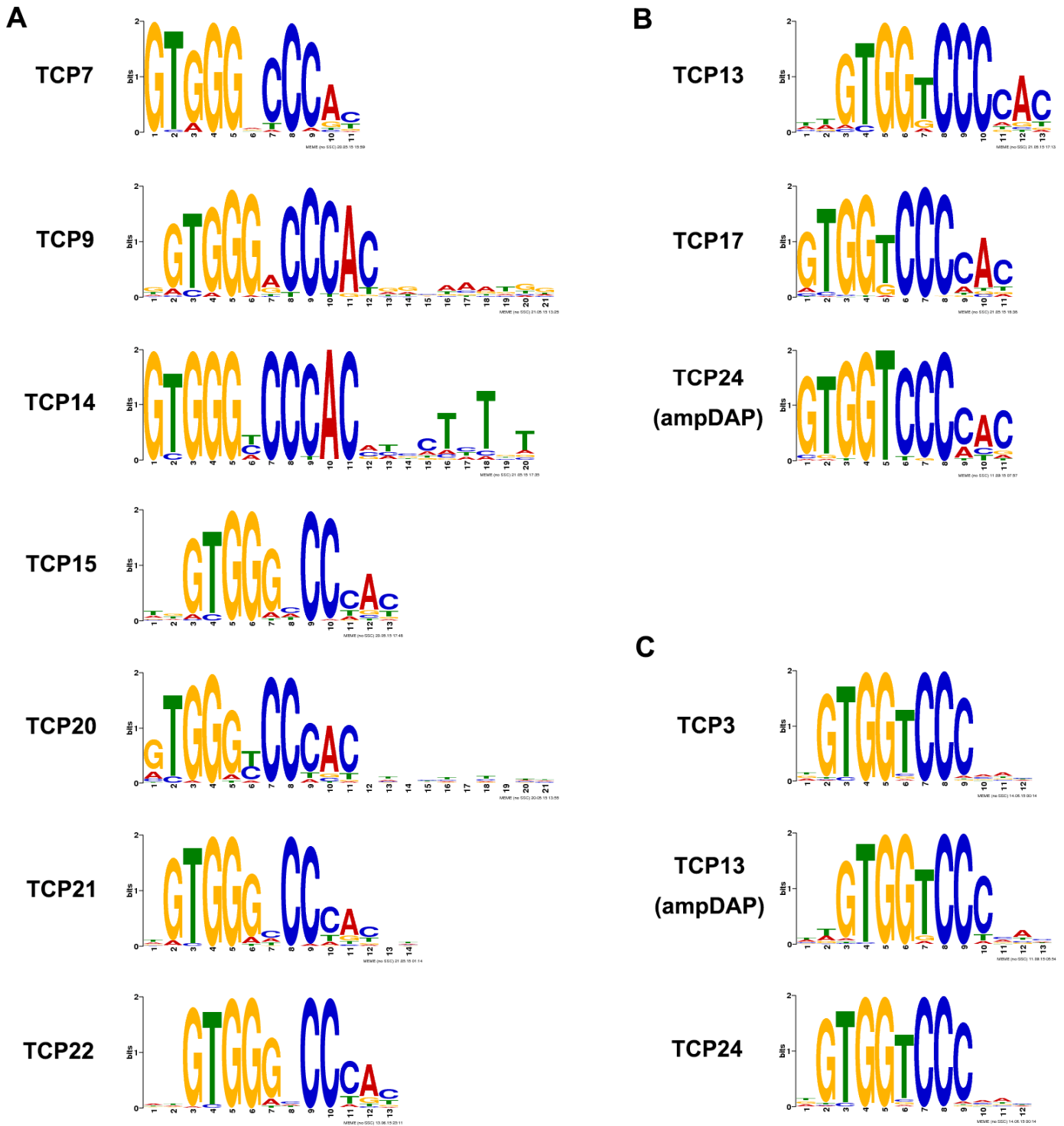
D–F, The apo structure of class I AtTCP15-DBD, the complex structure of class I AtTCP15-DBD with DNA, and the complex structure without DNA



Supplementary Figure 8. Comparison of the TCP domains with RHH domains

A–B, Comparison of the tilt angles of the two $\alpha 1$ helices in the AtTCP10-DBD and bacterial RHH domains (GopG).

C–D, The β -sheet interacting with DNA in the TCP domain of AtTCP10-DBD and RHH domain. (C) In the TCP domain of TCP10, the $\beta 3$ strand is shorter, and the side chain of valine (yellow) is oriented toward the hydrophobic core while the side chains of arginine (positions 46 and 48) interact with the DNA; (D) In the bacterial RHH domain, residues ③, ⑤, ⑦ in the β -sheet with hydrophobic side chains (yellow) are oriented toward the hydrophobic core. Residues ②, ④, ⑥ are oriented toward away from the core and interact with the DNA nucleotides.



Supplementary Figure 9. DAP-seq and ampDAP-seq results of TCP proteins.

A, DAP-seq results of class I TCP proteins AtTCP7, AtTCP9, AtTCP14, AtTCP15, AtTCP20, AtTCP21 and AtTCP22 show the binding motif for class I TCPs is GTGGGNCCCAC.

B, DAP-seq results of class II TCP proteins AtTCP13 and AtTCP17 and ampDAP-seq result of AtTCP24 show the binding motif for class II TCPs is GTGGTCCCCAC.

C, DAP-seq results of class II TCP proteins AtTCP3 and AtTCP24 and ampDAP-seq result of AtTCP13 show the binding motif for class II TCPs is GTGGTCCC.