

A Triplet-plus-One DNA Base recognition Scheme for TFIIIA Zinc fingers 1, 2, 3 and 5.

Much effort has been put into establishing a triplet code for zinc finger protein-DNA recognition. Such a code would in theory match 64 (4^3) short DNA sequences (three base pairs) with 8,000 (20^3) alpha helix variants. It was originally hoped that this triplet code [1] could be the basis for the design of artificial zinc finger proteins for academic, medical and commercial applications [2]. However, the triplet code ignores other known contacts frequently observed between zinc fingers and DNA base pairs [3].

The interactions of TFIIIA zinc fingers 1, 2, 3 and 5 with DNA triplets are shown in the tables below. The canonical alpha helix contacts are compared to those found in the co-crystal structure of TFIIIA and DNA [4]. Amino acids colored in red do not conform to the triplet code and bind instead to the nucleotide base on the complementary DNA (noncoding) strand. Critical contacts (not shown here) are often made to phosphates on the 5'-3' DNA sense (coding) strand by the amino acids in alpha helix positions $\alpha 7$ and $\alpha 10$. The sense strand has the same nucleotide sequence as 5S rRNA.

α -helix positions	ZnF 1	ZnF 2	ZnF 3	Znf 5
-3	tyr	phe		phe
-2				
-1	lys	ser		leu
$\alpha 1$		leu	lys	pro
$\alpha 2$	trp	his	ala	ser
$\alpha 3$	lys	his	asn	arg
$\alpha 4$				
$\alpha 5$			lys	
$\alpha 6$		arg	lys	arg
$\alpha 7$	his	his	his	his

$\alpha 8$				
$\alpha 9$				
$\alpha 10$	lys	thr	arg	val
$\alpha 11$			phe	
$\alpha 12$				

Amino acids in the recognition alpha helices of TFIIIA that bind to DNA.

DNA 5'-3'	dA	dG	dA	dG
TFIIIA alpha helix	$\alpha 6$ ala	$\alpha 3$ lys	-1 lys	$\alpha 2$ trp
ZnF 1	$\alpha 2$ trp	$\alpha 2$ trp	$\alpha 2$ trp	-1 lys

DNA 5'-3'	dG	dG	dG	dT
TFIIIA alpha helix	$\alpha 6$ arg	$\alpha 3$ his	-1 ser	$\alpha 2$ his
ZnF 2	$\alpha 6$ arg	$\alpha 3$ his	$\alpha 2$ his	-1 ser

DNA 5'-3'	dG	dA	dT	dC
TFIIIA alpha helix	α 6 lys	α 3 asn	-1 thr	α 2 ala
ZnF 3	α 6 lys	α 3 asn		α 2 ala

DNA 5'-3'	dG	dG	dT	dA
TFIIIA alpha helix	α 6 arg	α 3 arg	-1 leu	α 2 ser
ZnF 5	α 6 arg	α 3 arg		α 2 ser

The contacts between DNA bases (sense strand) and TFIIIA alpha helices. The predicted interactions are shown above those observed in TFIIIA zinc fingers 1, 2, 3 and 5. A fourth nucleotide that is contacted on the complementary DNA (noncoding) strand is shown colored in red.

Lessons learned

A relationship between positions in the recognition alpha helix and fixed positions in a DNA sequence is not apparent in TFIIIA. This “flexibility” can be explained by the different sizes, shapes, charges and reach of amino acid side chains. Therefore, any code should be based on overall tetramer-tetramer interactions. Four amino acids in the alpha helices (at positions -1, α 2, α 3 and α 6) should be considered as a single unit in binding to a four-base pair sequence. However,

the ability to form an alpha helical structure may require other amino acids to be present. Four DNA quartets with their related amino acid tetramers in TFIIIA are shown in the table below.

Coding strand quartets	dAdGdAdC	dGdGdGdA	dGdAdTdG	dGdGdTdT
-1, α 2, α 3 and α 6	KWKA	SHHR	TANK	LSRR
Zinc finger alpha helix	NkNwKLQAH	TsLhhLTrH	TTKanMKkH	SLPsrLKrH

“Four of the residue positions on the helix N-terminus and exposed face are predicted to provide base specific ligands”. *Toby J. Gibson (1988)* [5].

Bibliography

- [1] DNA-binding Specificities of Cys₂His₂ Zinc Finger Proteins. [[web server](#)]
- [2] The past, present, and future of artificial zinc finger proteins: design strategies and chemical and biological applications. Negi S, Imanishi M, Hamori M, Kawahara-Nakagawa Y, Nomura W, Kishi K, Shibata N, Sugiura Y. (2023) *J Biol Inorg Chem.* 28, 249-261. [[web site](#)]
- [3] Updated understanding of the protein-DNA recognition code used by C2H2 zinc finger proteins. Zhang X, Blumenthal RM, Cheng X. (2024) *Curr Opin Struct Biol.* 87, 102836. [[web site](#)]
- [4] 1TF6 Co-crystal structure of Xenopus TFIIIA zinc finger domain bound to the 5S ribosomal RNA gene Internal control region. Nolte RT, Conlin RM, Harrison SC, Brown RS. (released 1998). [[web site](#)]
- [5] A model for the tertiary structure of the 28 residue DNA-binding motif (‘zinc finger’) common to many eukaryotic transcriptional regulatory proteins. Gibson TJ, Postma JPM, Brown RS, Argos P. (1988) *Protein Engineering* 2, 209-218. [[abstract](#)]