

## Supplementary Information

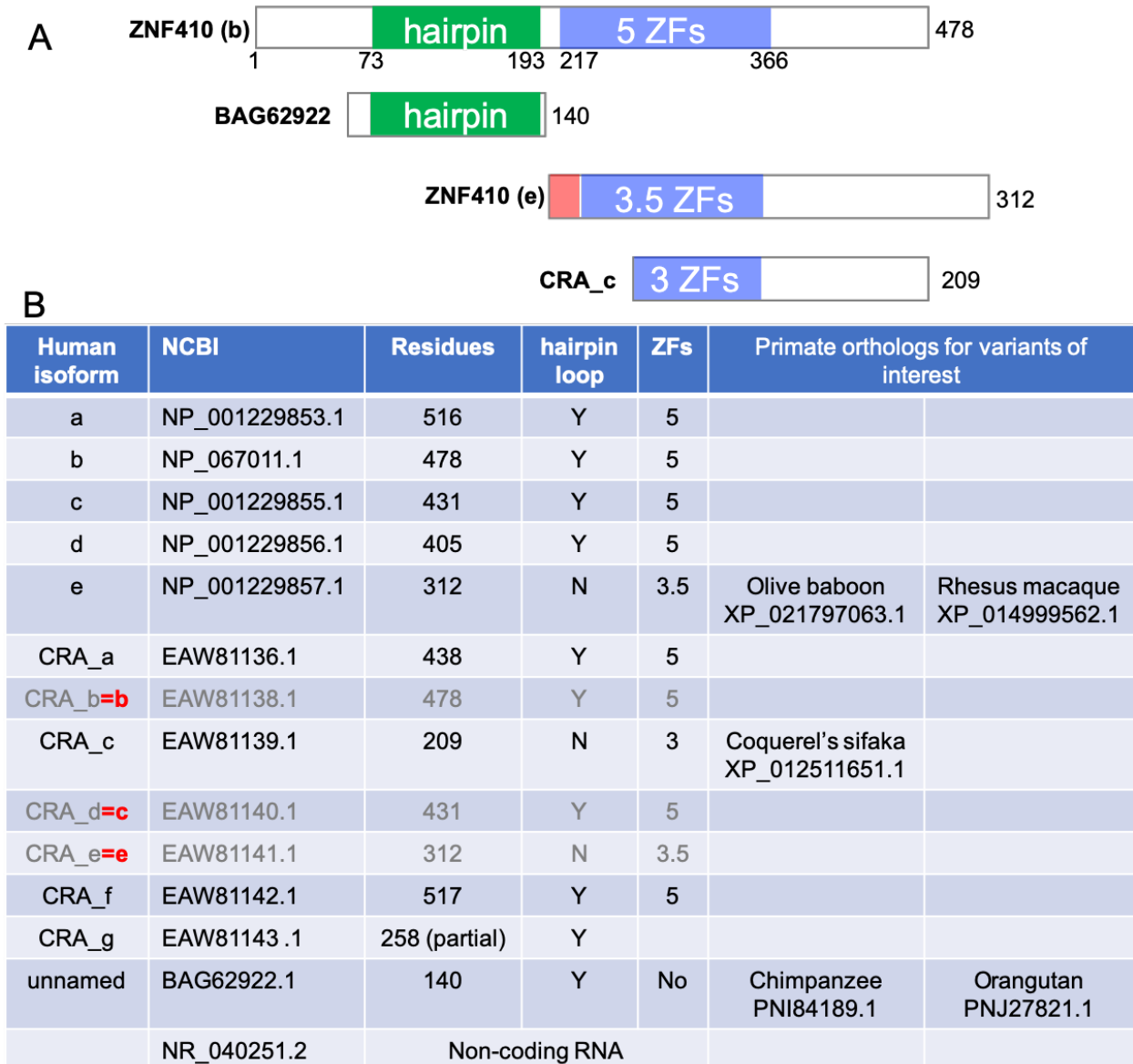
### **Allosteric autoregulation of DNA binding via a DNA-mimicking protein domain: a biophysical study of ZNF410-DNA interaction using small angle X-ray scattering**

Gundeep Kaur<sup>1,8</sup>, Ren Ren<sup>1,8</sup>, Michal Hammel<sup>2</sup>, John R. Horton<sup>1</sup>, Jie Yang<sup>1</sup>, Yu Cao<sup>1</sup>, Chenxi He<sup>3</sup>, Fei Lan<sup>3</sup>, Xianjiang Lan<sup>4,5</sup>, Gerd A. Blobel<sup>5,6</sup>, Robert M. Blumenthal<sup>7</sup>, Xing Zhang<sup>1,\*</sup>, and Xiaodong Cheng<sup>1,\*</sup>

<sup>1</sup>Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; <sup>2</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California, USA; <sup>3</sup>Shanghai Key Laboratory of Medical Epigenetics, International Laboratory of Medical Epigenetics and Metabolism, Ministry of Science and Technology, Institutes of Biomedical Sciences, Fudan University and Key Laboratory of Carcinogenesis and Cancer Invasion, Ministry of Education, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai, 200032, China; <sup>4</sup>Department of Systems Biology for Medicine, School of Basic Medical Sciences; Department of Liver Surgery and Transplantation, Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China; <sup>5</sup>Division of Hematology, the Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>6</sup>Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>7</sup>Department of Medical Microbiology and Immunology, and Program in Bioinformatics, The University of Toledo College of Medicine and Life Sciences, Toledo, OH 43614, USA

<sup>8</sup>These authors contributed equally to this work.

\*Correspondence: XZhan21@mdanderson.org; XCheng5@mdanderson.org



**Figure S1.** There are at least nine distinct isoforms for human ZNF410. **(A)** Cartoons of selected isoforms. **(B)** Three are particularly noteworthy, each of them is present in at least one other primate species. The unnamed isoform (BAG62922.1) begins at an internal methionine, but ends before any of the ZNFs. The other two are just the reverse – isoform e (NP\_001229857 or EAW81141) is substituted for ZF1 and the first cysteine of ZF2, and CRA\_c (EAW81139) begins just ahead of ZNF3.

Common name	Species	Clade	ZNF410 Ortholog	Cover	Identity (%)	Length (aa)
Human (isoform a)	<i>Homo sapiens</i>	Primate	NP_001229853.1	(100)	(100)	516
Rhesus macaque	<i>Macaca mulatta</i>	Primate	XP_014999559.2	93	95.67	478
Angolan colobus	<i>Colobus angolensis</i>	Primate	XP_011804978.1	98	90.20	506
Common marmoset	<i>Callithrix jacchus</i>	Primate	XP_002754185.2	93	95.46	478
Malayan flying lemur	<i>Galeopterus variegatus</i>	Dermoptera	XP_008565230.1	93	92.99	478
Mouse	<i>Mus musculus</i>	Glire	NP_659082.1	93	93.40	478
American beaver	<i>Castor canadensis</i>	Glire	XP_020022334.1	93	93.20	478
Guinea pig (cavy)	<i>Cavia porcellus</i>	Glire	XP_003472457.1	93	93.20	478
European rabbit	<i>Oryctolagus cuniculus</i>	Glire	XP_002719703.1	93	95.05	478
Arabian camel (dromedary)	<i>Camelus dromedarius</i>	Laurasiatheria	XP_010974834.2	93	92.99	478
Sperm whale	<i>Physeter catodon</i>	Laurasiatheria	XP_007109074.1	93	92.99	472
Cow	<i>Bos taurus</i>	Laurasiatheria	NP_001338875.1	93	92.37	478
Goat	<i>Capra hircus</i>	Laurasiatheria	XP_005686117.1	93	92.37	478
Horse	<i>Equus caballus</i>	Laurasiatheria	XP_001490078.2	93	92.37	478
Cat	<i>Felis catus</i>	Laurasiatheria	XP_003987863.1	93	91.55	478
Dog	<i>Canis lupus familiaris</i>	Laurasiatheria	XP_022277914.1	93	91.96	478
Polar Bear	<i>Ursus maritimus</i>	Laurasiatheria	XP_008698131.1	93	91.96	479
Chinese pangolin	<i>Manis pentadactyla</i>	Laurasiatheria	XP_036769129.1	93	91.96	477
Large flying fox	<i>Pteropus vampyrus</i>	Laurasiatheria	XP_023392808.1	93	91.98	479
Common vampire bat	<i>Desmodus rotundus</i>	Laurasiatheria	XP_024423383.1	93	90.52	478
Star-nosed mole	<i>Condylura cristata</i>	Laurasiatheria	XP_004682125.1	93	88.87	459
Elephant	<i>Loxodonta africana</i>	Afrotheria	XP_003408805.1	93	85.83	490
Armadillo	<i>Dasypus novemcinctus</i>	Xenarthra	XP_012379386.1	92	83.0	468
Koala	<i>Phascolarctos cinereus</i>	Metatheria	XP_020841968.1	93	78.17	499
Platypus	<i>Ornithorhynchus anatinus</i>	Monotremata	XP_028931293.1	92	72.88	467

**Figure S2.** Phylogeny of ZNF410 among 25 diverse mammals, chosen based (1).

1. Hecker, N. and Hiller, M. (2020) A genome alignment of 120 mammals highlights ultraconserved element variability and placenta-associated enhancers. *Gigascience*, **9**, giz159.

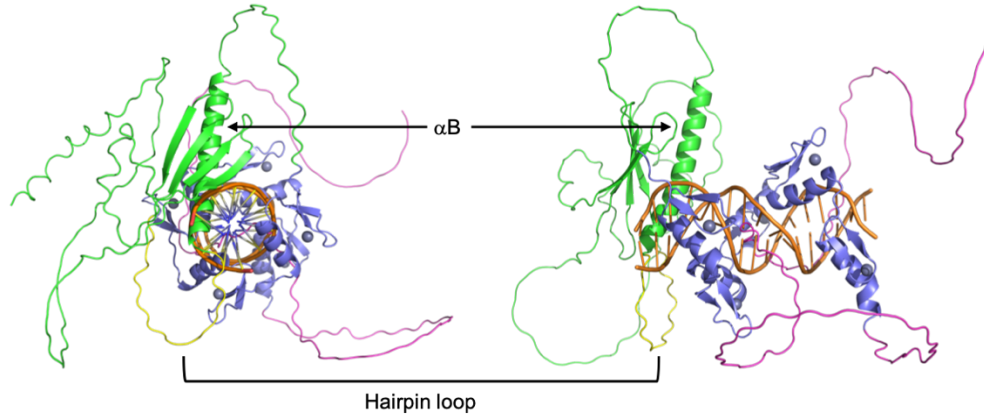
(A) Vertebrate orthologs	Mammalia	Aves	Reptilia	Amphibia	Osteichthyes
	<i>Homo sapiens</i>	<i>Gallus gallus</i>	<i>Anolis carolinensis</i>	<i>Xenopus tropicalis</i>	<i>Lepisosteus oculatus</i>
	NP_001229855.1	NP_001006480.2	XP_003214434.1	NP_001072669.1	XP_015205560.1
	human	chicken	green anole lizard	western clawed frog	shortnose gar
Cover (%)	(100)	87	98	98	98
E value	(0.0)	0.0	0.0	2e-151	2e-142
Identity (%)	(100)	86	69	63	58
Length (aa)	431 (isoform c)	445	467	403	448
<b>(B) Substitutions relative to human ortholog</b>					
Hairpin loop	(60 aa)	6	8	11	18
ZFs	(115 aa)	4	7	21	20
<b>(C) Number of ZNF410 binding sites within 1 KB of CHD4 promoter (all in the same orientation)</b>					
	11	9	10	11	4

(D)

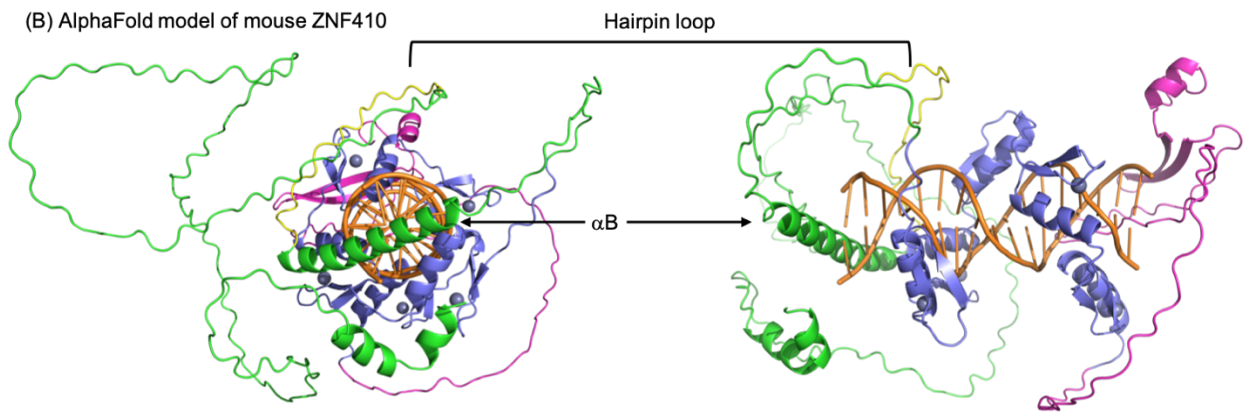
		*** ** * * *
H. sapiens	GLGSSAEHLVVFQDEAEDSGNDFLSSESTDSSTPWFLRVQELAHDSLIAATRAQLAKNAKTSSNGE	
G. gallus	GLGSPAEHLVVFQDEVEDSGNDFLSHDSTDSSTPWFLRVQELAHDSLIAATRAQLAKNAKASNNGE	
A. carolinensis	GLGSLSEHLVVFQDEAEDSGNDFLSNNSSTDSSTPWFLRVQELAHDSLIAATRAQLAKNAKASNNGE	
X. tropicalis	GLTSPSEHLVVFQDEAEDSGNDFLSNDCTDSNTPWYLRVQELAHDSLIAATRAQLAKSVKTSSVGE	
L. oculatus	GLAASAEPLVVFVEDEVD---NDISTGSESADGSTPWYLRVQELAHDSLIAATRAQLAKDAKASNSGD	*** * * * *

**Figure S3.** (A) Summary of five vertebrate ZNF10 orthologs. (B) Number of substitutions within the hairpin loop and ZF region are listed. (C) Number of ZNF410 binding sites within 1 KB 5' to CHD4 promoters are listed. (D) Sequence alignment of the corresponding hairpin loop. Acidic residues in proposed DNA-mimicking segments of human (top) and gar (bottom) orthologs are indicated with red asterisks.

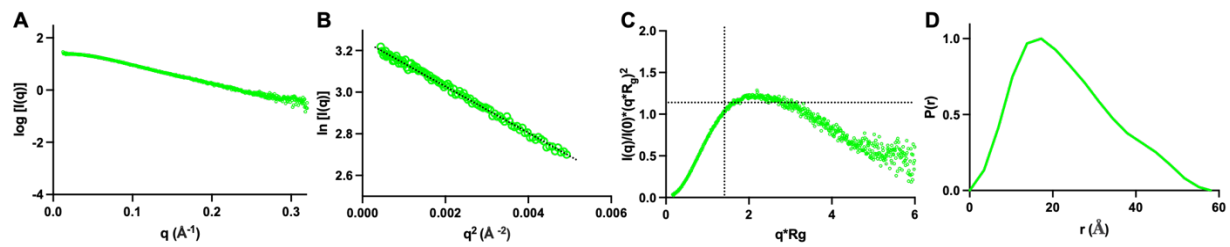
(A) AlphaFold model of Bos Taurus ZNF410



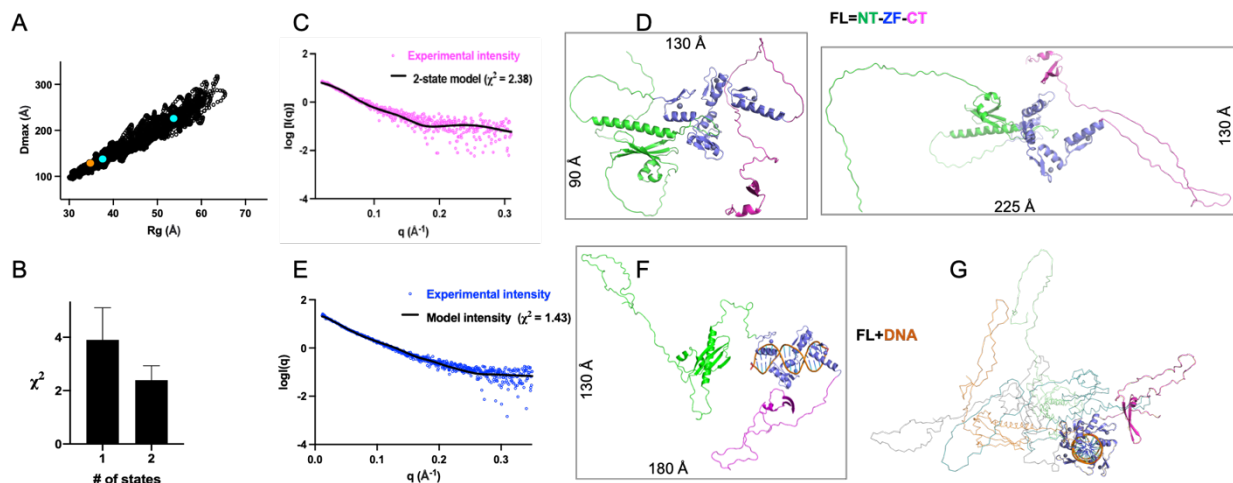
(B) AlphaFold model of mouse ZNF410



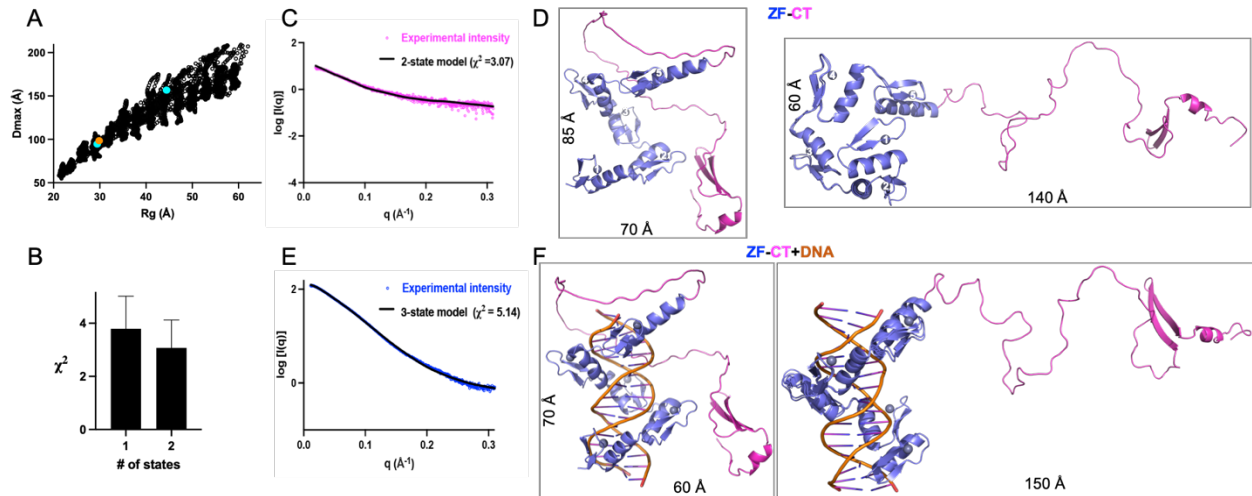
**Figure S4.** Two orthogonal views of AlphaFold predictions of (A) *Bos taurus* and (B) mouse ZNF410 orthologs. The predicted corresponding long  $\alpha B$  helix blocks the DNA binding.



**Figure S5.** SAXS curves/profile for 17 bp DNA using in our experiments. **(A)** Scattering intensity profile. **(B)** Guinier plot. **(C)** Kratky plot. **(D)** P(r) plot.



**Figure S6. ZNF410 FL exists in compact and extended conformations.** (A) Plot of  $R_g$  vs.  $D_{max}$  values for all 10,000 models obtained in BILBOMD conformational sampling. The initial AlphaFold model is shown as an orange dot. (B) MULTIFOXS analysis suggests ZNF410 FL has the probability to adopt two conformational states (the two blue dots in panel A). (C) Fitting of experimental X-ray scattering profile (magenta dots) to the calculated scattering (black line) of 2-state models. (D) Examples of models representing two states of ZNF410 FL (compact and extended). (E) The fitting of experimental X-ray scattering profile (blue dots) to the calculated scattering (black line) of the fitting model in the presence of DNA. (F) One example of a DNA bound ZNF410 model. (G) A sampling of DNA bound ZNF410 models with varied conformations of NT and CT.



**Figure S7. ZF-CT exists as mixture of compact and extended conformations.** (A) Plot of Rg vs. Dmax values for all 10,000 models obtained in BILBOMD conformational sampling. The starting AlphaFold model is shown as an orange dot. (B) MULTIFOXS analysis suggests ZF-CT has the probability to adopt (at least) two conformational states (two blue dots in panel A). (C) Fitting of experimental scattering profile (magenta dots) to the calculated scattering (black line) of a two-model ensemble. (D) Examples of models represent two states of ZF-CT in the absence of DNA (compact and extended). (E-F) The fitting of experimental X-ray scattering profile (blue dots) to the calculated scattering (black line) from the multi-model states in the presence of DNA (panel F).



**Table S1.** Expression constructs and DNA primers used in this study

Residues	name	pXC#		
1-478	FL	2179		
1-366	NT-ZF	2218		
1-216	NT	2288		
217-366	ZF	2180		
217-478	ZF-CT	2217		
	mutant		Primers used in the mutagenesis	
1-478	S148D	2330	5' -CAGAAGAT <u>GAT</u> GGGAATGATTTCCCTC-3' 3' -TCCTACTCCGTCTTCTACTACCCTTACT-5' (the underlined codon is the intended mutation)	
1-478	S/T-D/E:  (S148D/ S154D/ S155D/ S157D/ T158E/ S160D/ S161D)	2319	(1) (5' -TGTTCAGGGGCCCTGGGATCCATGTTATCAG ATGAGTTAGAAT-3') and the pertinent mutagenic reverse primer (5' -CATGGAATGTC <u>ATCGTCTTCGTCCTCATCGTCGAGGAAATCATTTCCCATC</u> -3') (2) (5' -ATTTCCCTCGACGATGAGGACGAAGACGATGACATT CCATGGTTCCCTCCGG-3') and the 3' reverse primer (5' -TCAGTCACGATGCGGCCGCTCGAGTCATGT CCGTCTTTCAGTTAA-3') The underlined codons are the intended mutations.	
1-478	D/E-A:  (D143A/ E144A/ E146A/ D147A)	2341	(1) 5' -TTGTACAGGCTGCGGCAGCAGCTTCAGGGAATGATT TC-3' and the pertinent mutagenic reverse primer (5' -CATGGAATGTCATCGTCTTCGTCCTCATCGTC <u>GAGGAAATCATTTCCCATC</u> -3') (2) 5' -TTCCCTGAAGCTGCTGCCGAGCCTGTACAAACACT AA-3' and the 3' reverse primer (5' -TCAGTCA CGATGCGGCCGCTCGAGTCATGTCCGTCTTTCAGTTAA-3') The underlined codons are the intended mutations.	
1-366	S/T-D/E	2312	Same as pXC2319	
1-366	D/E-A	2342	Same as pXC2341	