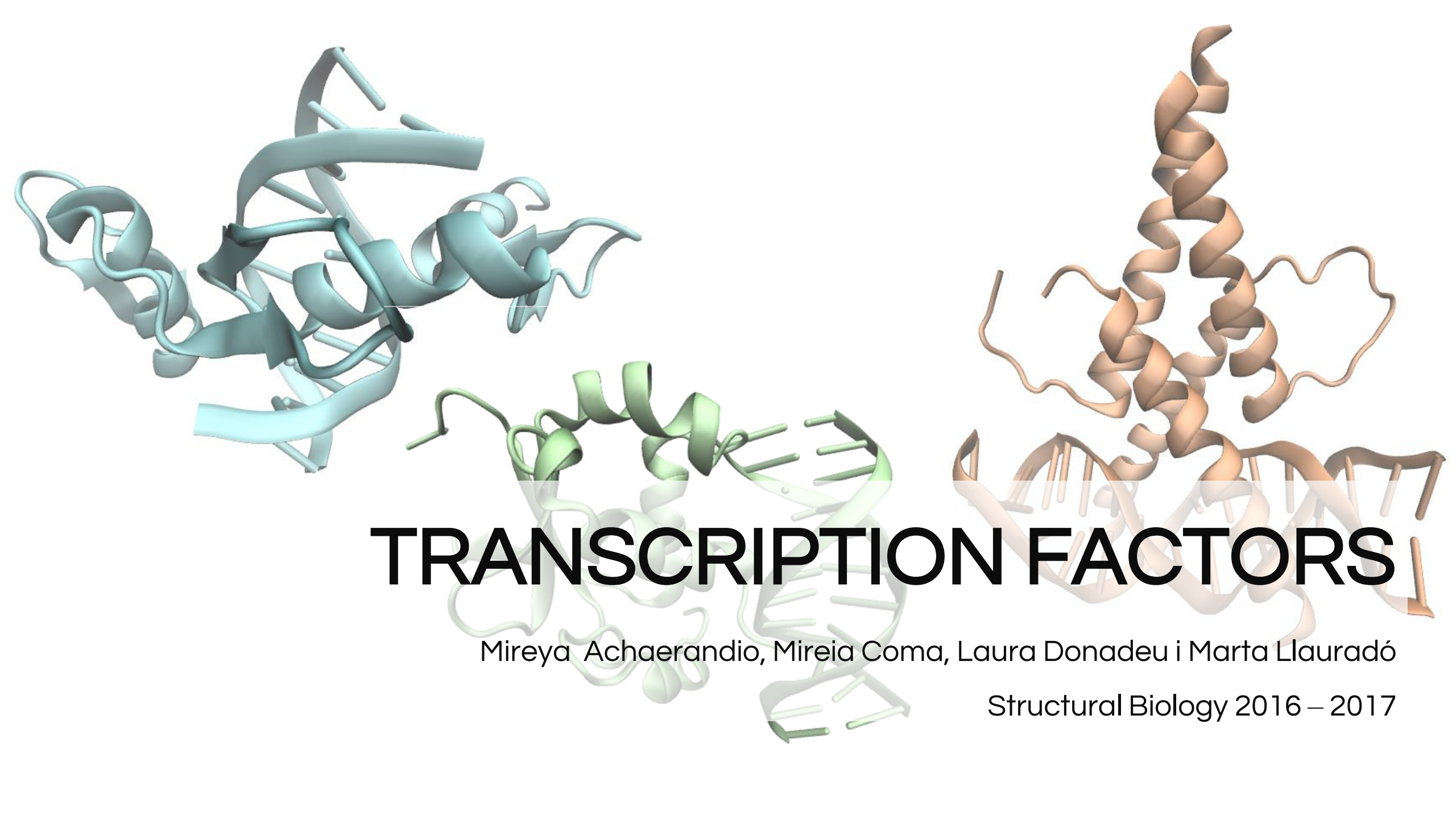


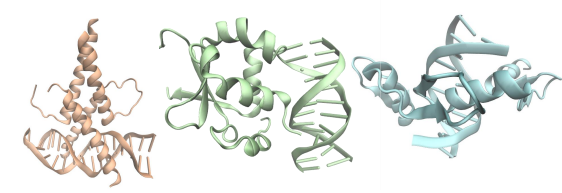


TRANSCRIPTION FACTORS

Mireya Achaerandio, Mireia Coma, Laura Donadeu i Marta Llauradó

Structural Biology 2016 – 2017

Index



General Introduction



Methodology



Zn Finger: KLF4



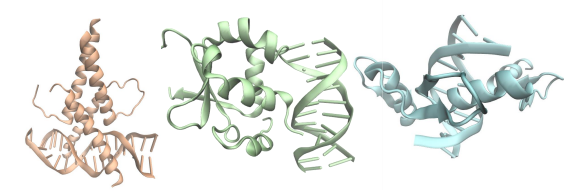
Helix-Turn-Helix: HSF1



Helix-Loop-Helix: MITF



Conclusions



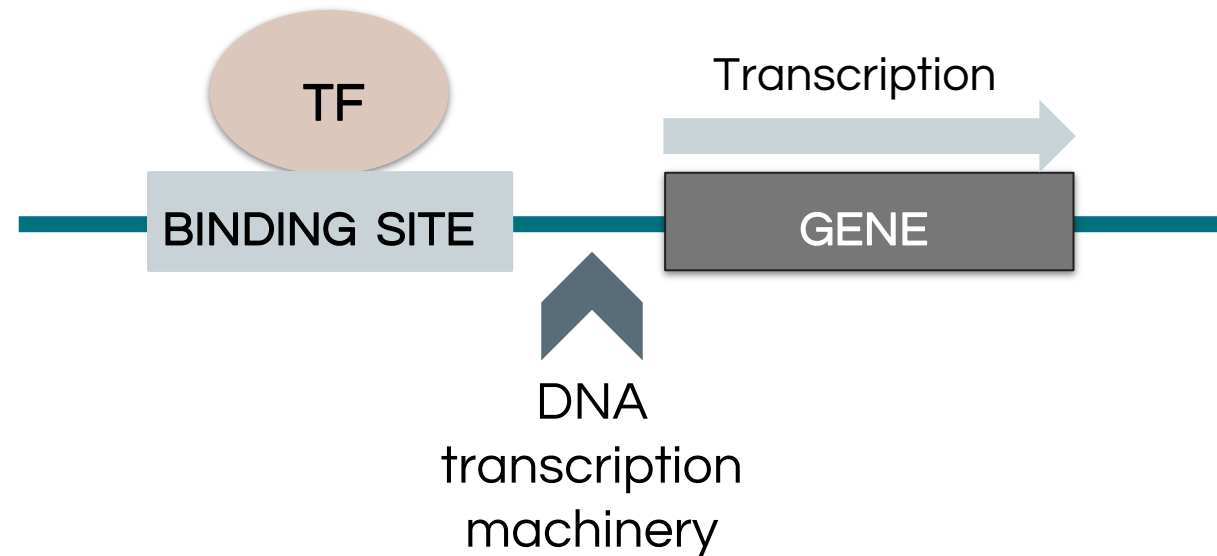
Introduction

What is a transcription factor?

Regulation of DNA transcription

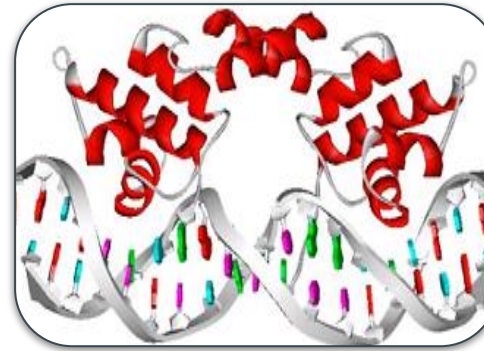
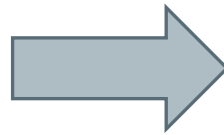
Recognition of specific DNA sequences

Reorganization of local nucleosome structures

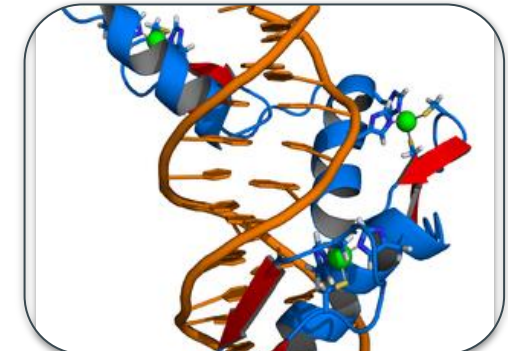


Introduction

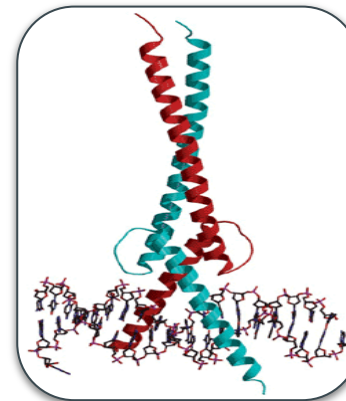
Four different DNA-binding motifs:



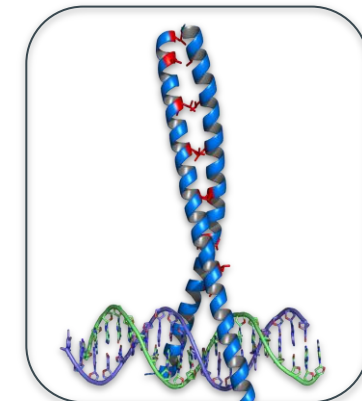
Helix-turn-helix
(HTH)



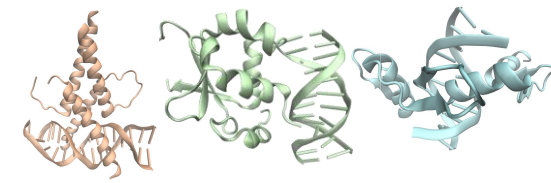
Zinc fingers
(Zn fingers)



Helix-loop-helix
(bHLH)



Leucine zippers
(bZIP)

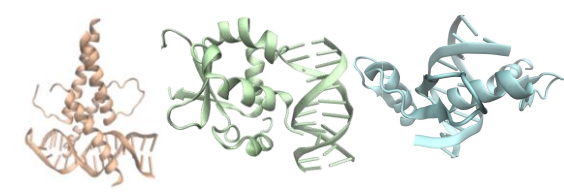


HLH: Taking Myc to the Max. Nature Structural Biology. 2003;10(3):157-157.

bZIP: Cremallera de leucina [Internet]. Es.wikipedia.org. 2017 [cited 3 March 2017]. Available from: <https://es.wikipedia.org/wiki/Cremallera-de-leucina>

HTH: From Molecule to Man - Protein [Internet]. Swift.cmbi.ru.nl. 2017 [cited 3 March 2017]. Available from: <http://swift.cmbi.ru.nl/gv/students/mtom/SUP-1.html>

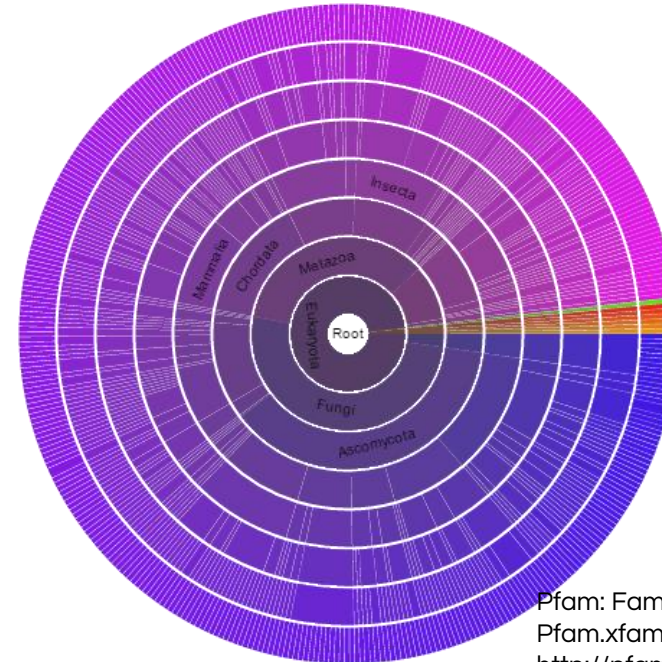
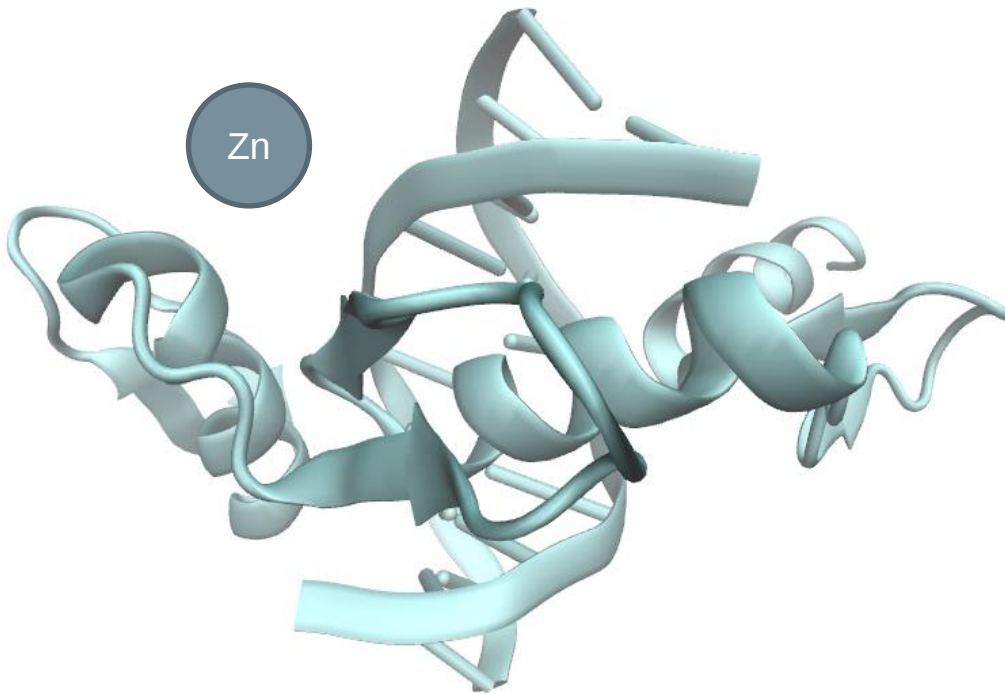
ZnF: Zinc finger [Internet]. En.wikipedia.org. 2017 [cited 3 March 2017]. Available from: <https://en.wikipedia.org/wiki/Zinc-finger>



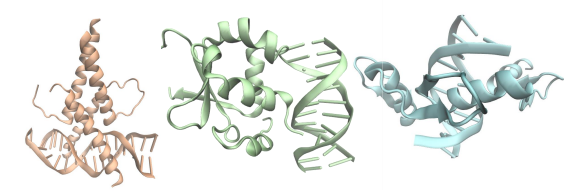
Zinc fingers

Introduction

- Critical binding interactions with DNA
- Particularly important in gene regulation
 - ↗ Activation
 - ↘ Inhibition

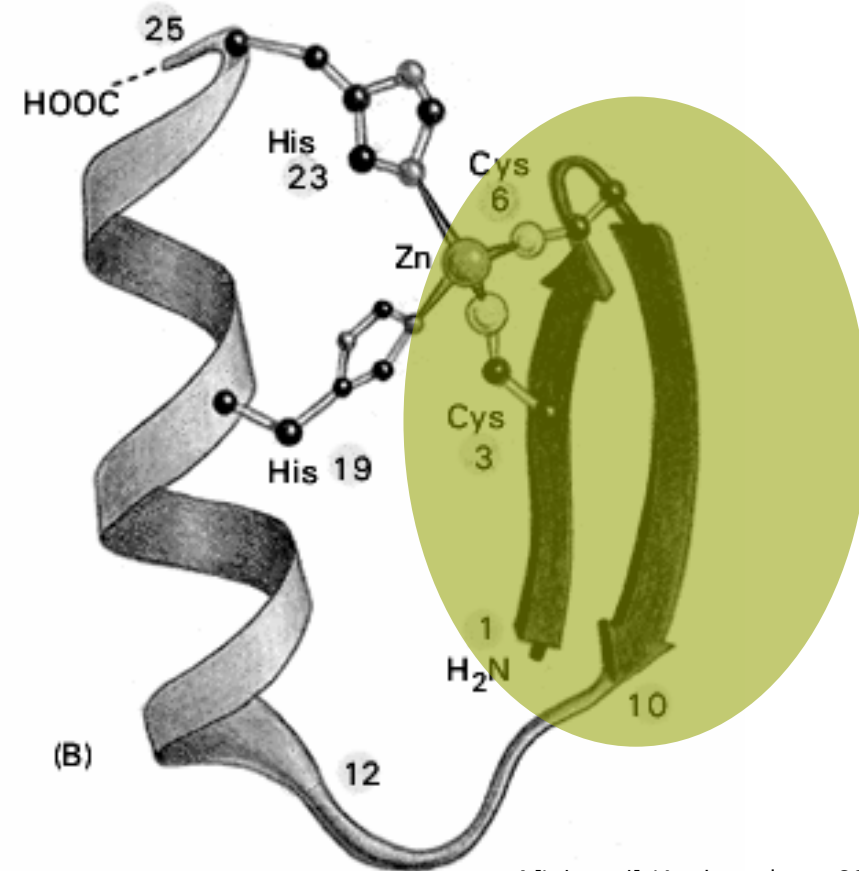
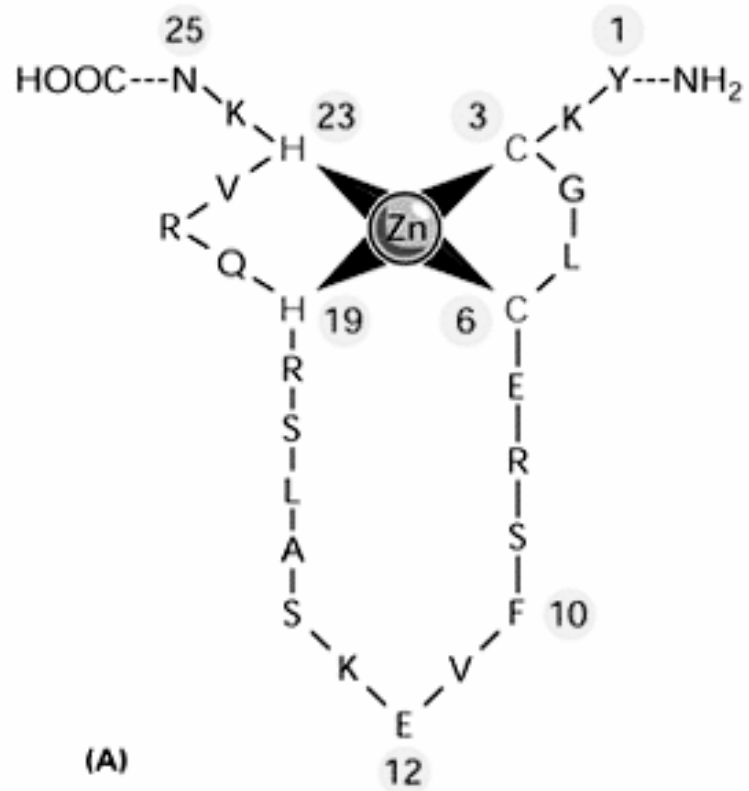


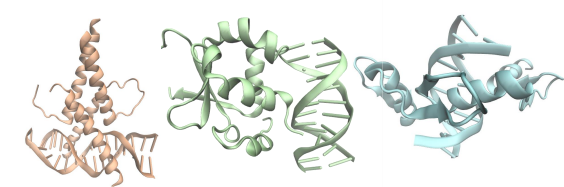
Pfam: Family: zf-C2H2 (PF00096) [Internet].
Pfam.xfam.org. 2017 [cited 3 March 2017]. Available from:
<http://pfam.xfam.org/family/PF00096#tabview=tab7>



Zinc Fingers

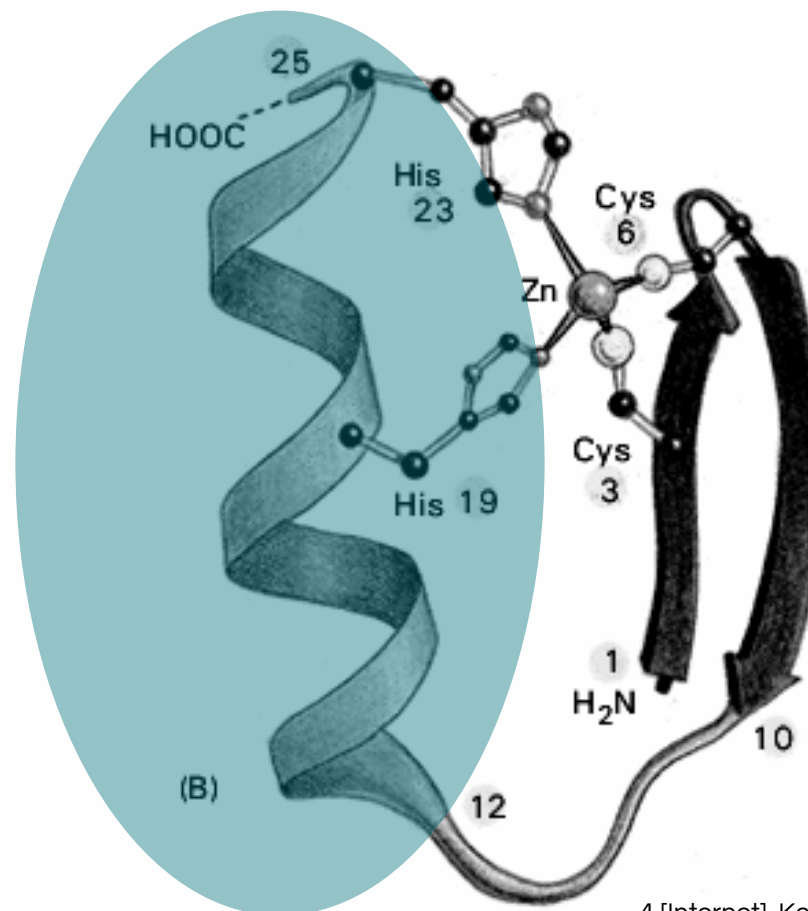
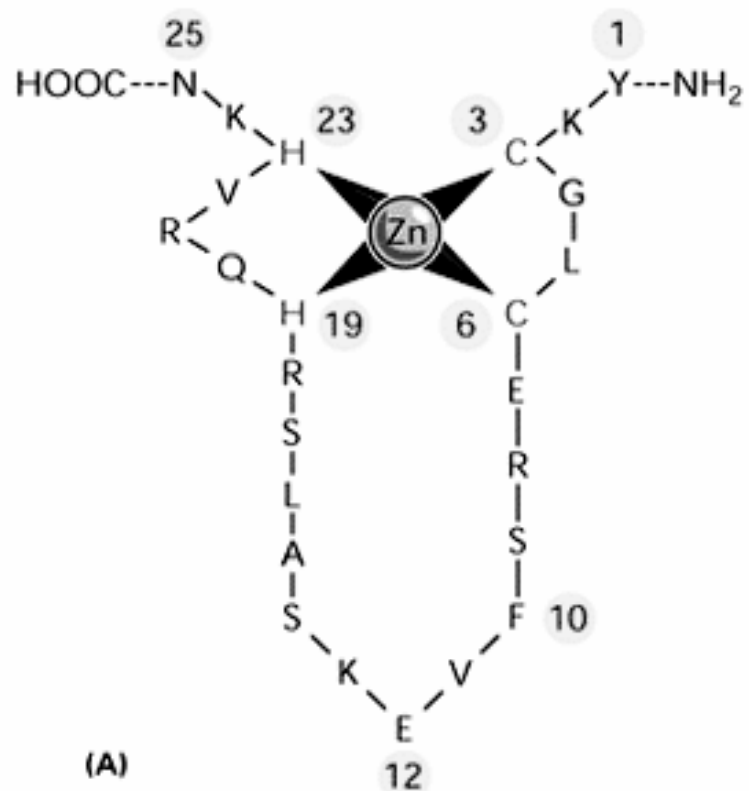
Introduction



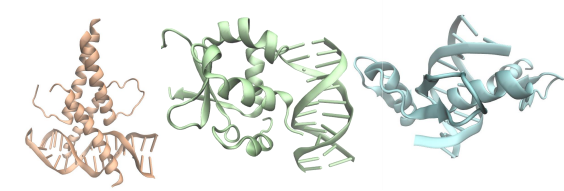


Zinc Fingers

Introduction

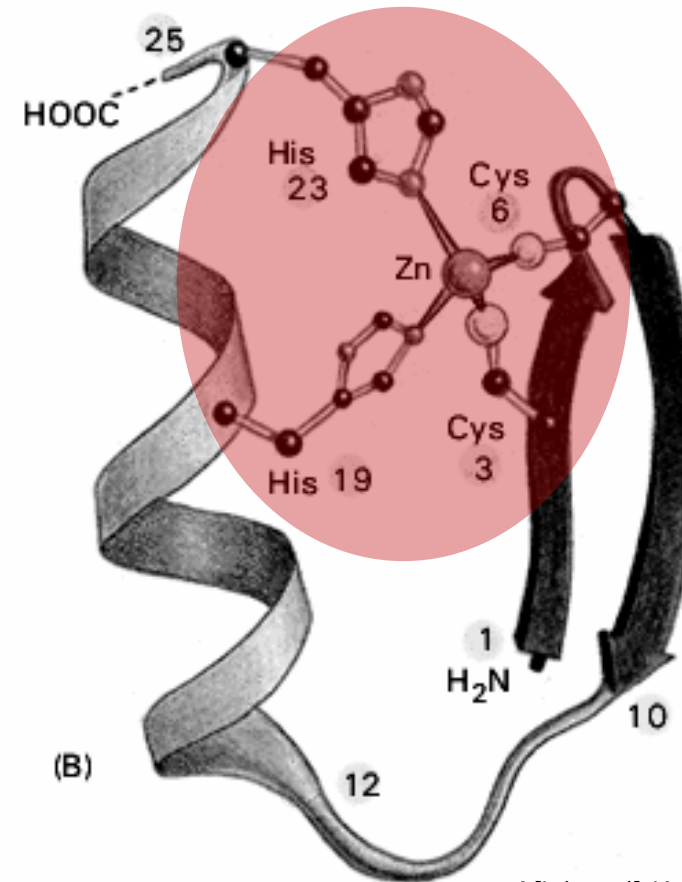
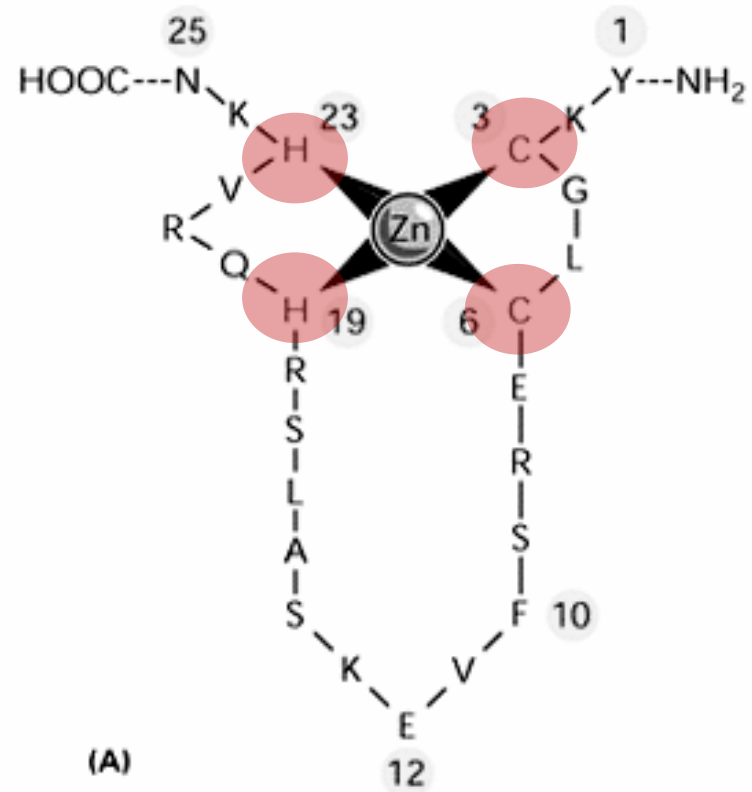


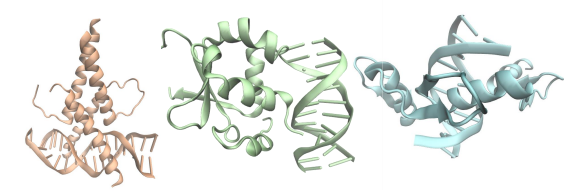
4 [Internet]. Kc.njnu.edu.cn. 2017 [cited 3 March 2017].
Available from:
<http://kc.njnu.edu.cn/swxbx/shuangyu/8.htm>



Zinc Fingers

Introduction





Helix-turn-helix

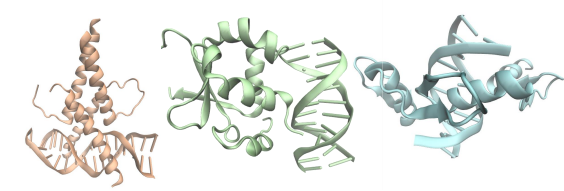
Introduction

HTH motifs are found in all known DNA binding proteins that regulate gene expression

Wide structural diversity
Di-helical
Tri-helical
Tetra-helical
Winged helix-turn-helix

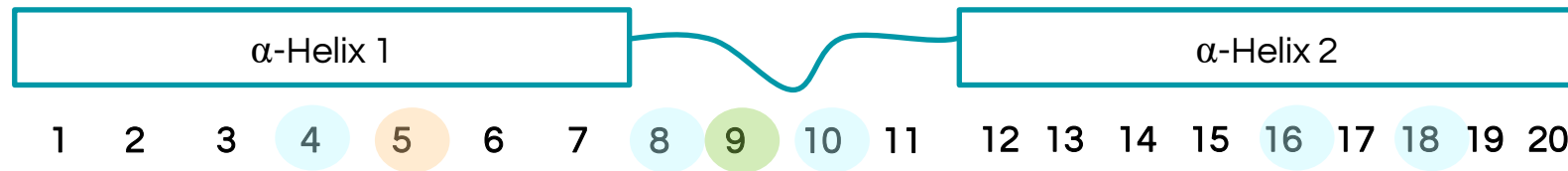


Pfam: Family: HTH_3 (PF01381) [Internet]. Pfam.xfam.org. 2017 [cited 3 March 2017]. Available from: <http://pfam.xfam.org/family/PF01381#tabview=tab7>

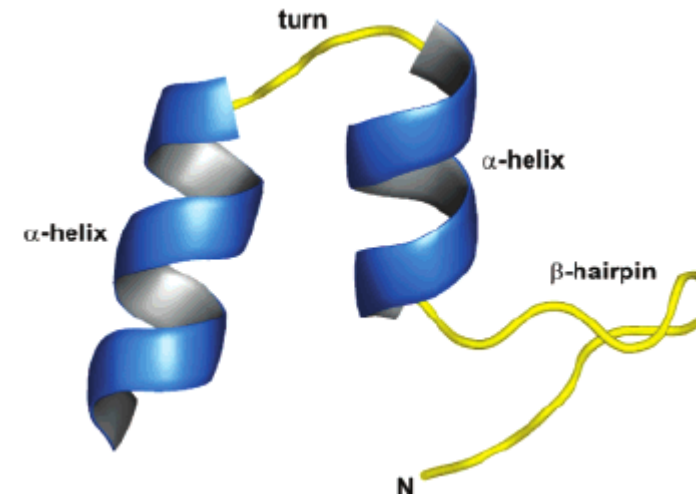


Helix-turn-helix

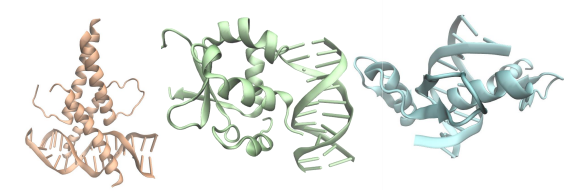
Introduction



- Hydrophobic residues
- Glycine or Alanine
- Glycine



[Internet]. 2017 [cited 3 March 2017]. Available from: <https://www.researchgate.net/figure/7267631-fig1-Figure-1-Depiction-of-the-ankyrin-repeat-motif-which-contains-33-amino-acid-residues>



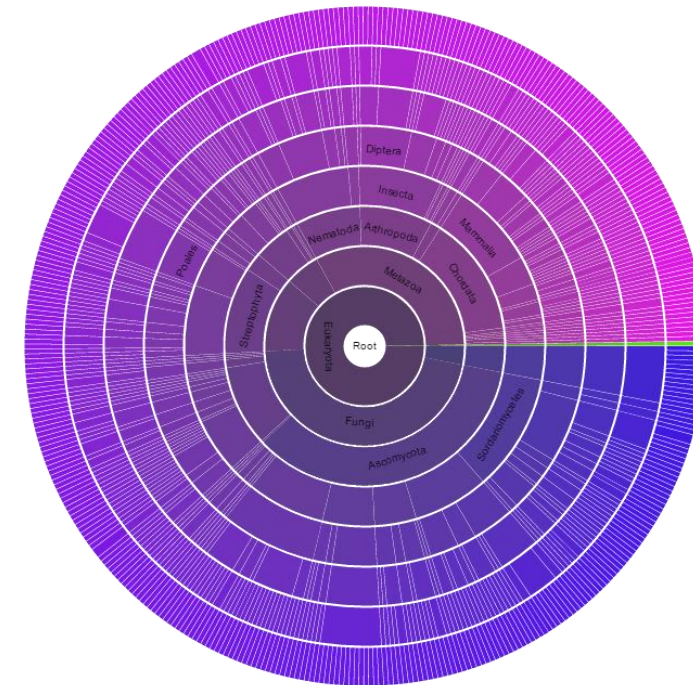
Helix-loop-helix

Introduction

Important for many **developmental processes**:

- Neurogenesis
- Myogenesis
- Hematopoiesis
- Pancreatic development

HLH domains
b/HLH
b/HLH/Z



Pfam: Family: HLH (PF00010) [Internet]. Pfam.xfam.org. 2017 [cited 3 March 2017]. Available from: <http://pfam.xfam.org/family/PF00010#tabview=tab7>

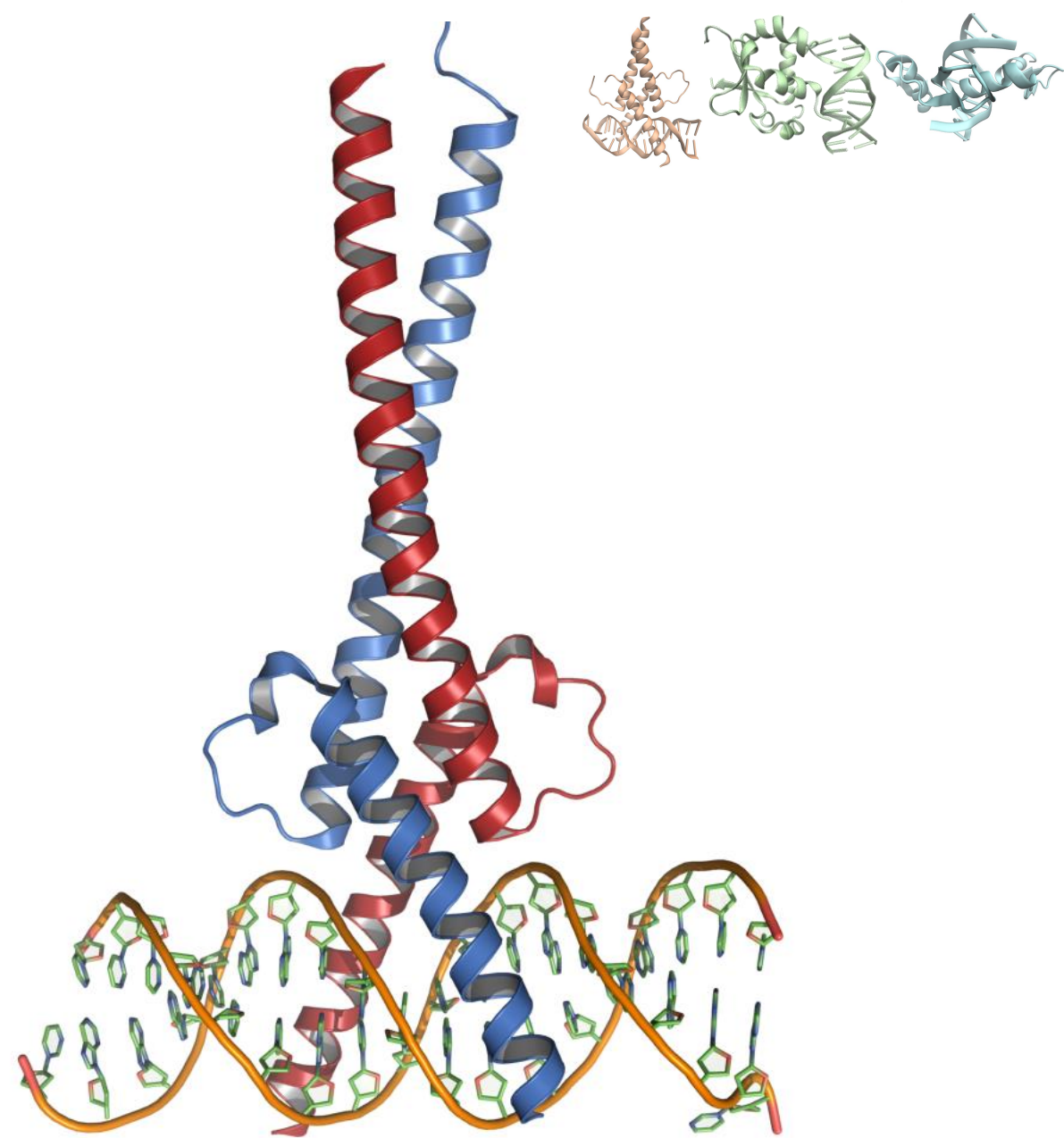
Helix-loop-helix

Introduction

2 α -helix connected by a loop



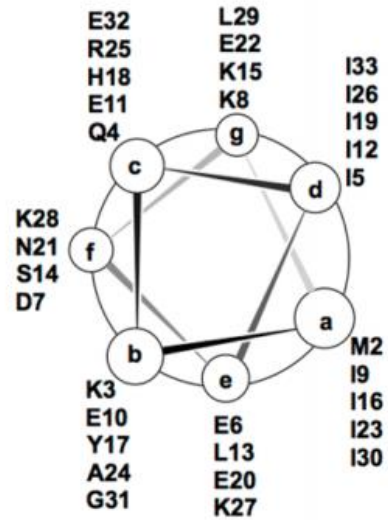
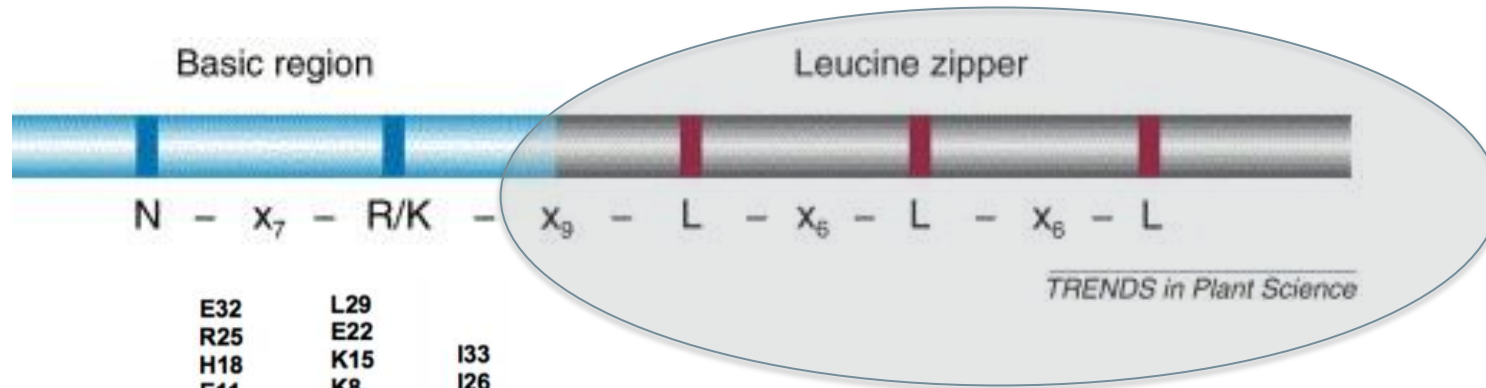
DNA E-box



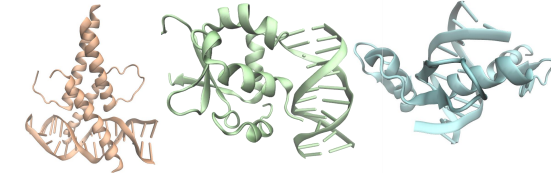
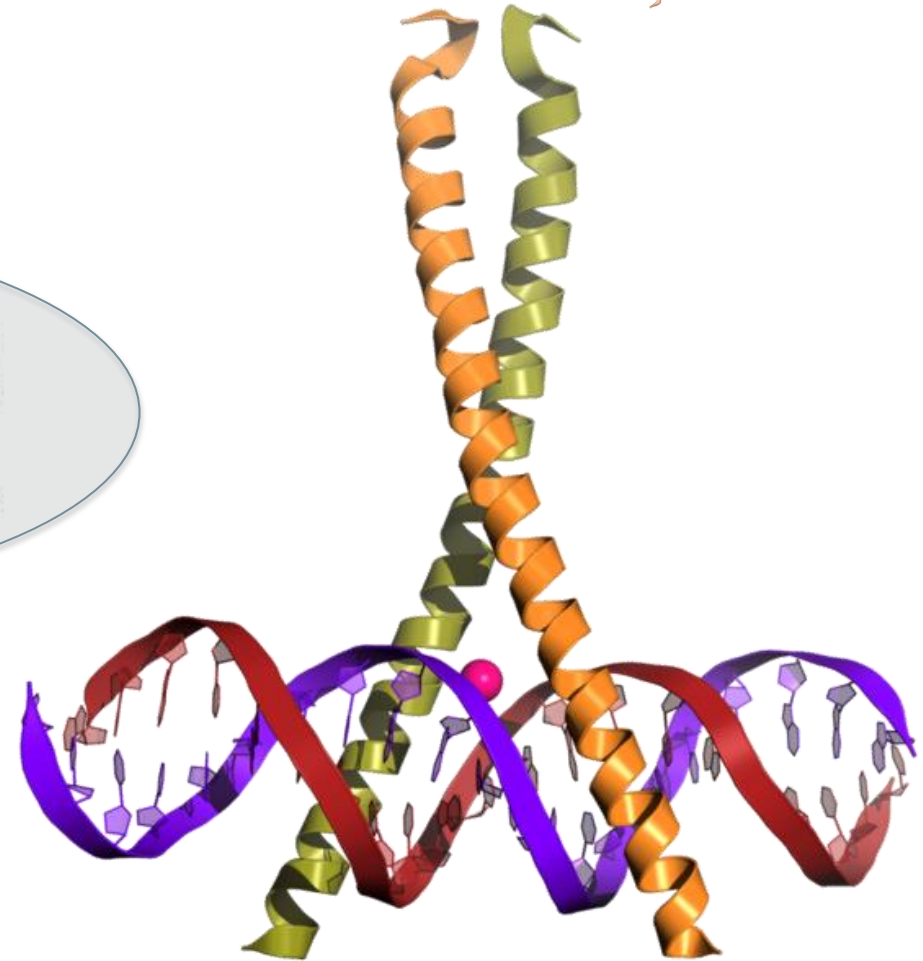
Myc [Internet]. Es.wikipedia.org. 2017 [cited 3 March 2017]. Available from: <https://es.wikipedia.org/wiki/Myc>

Leucine Zippers

Introduction



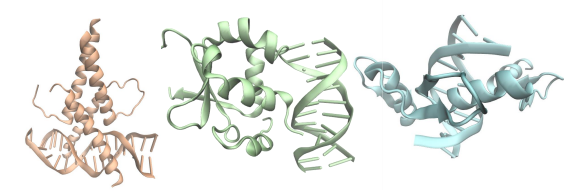
AMPHIPATHIC HELIX



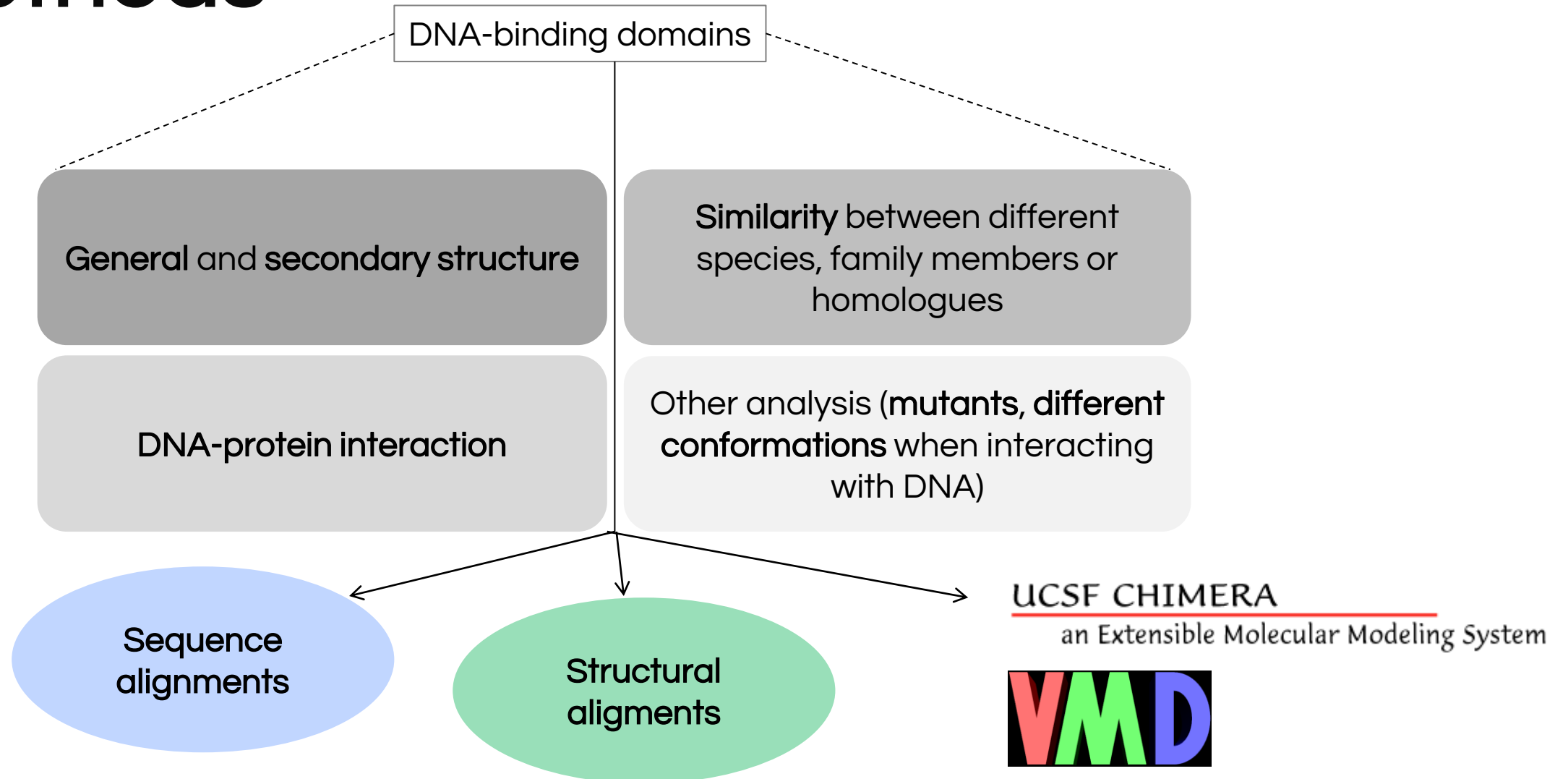
Nepal M, Sheedlo M, Das C, Chmielewski J. Accessing Three-Dimensional Crystals with Incorporated Guests through Metal-Directed Coiled-Coil Peptide Assembly. *Journal of the American Chemical Society*. 2016;138(34):11051-11057.

Pfam: Family: bZIP1 (PF00170) [Internet]. Pfam.xfam.org. 2017 [cited 3 March 2017]. Available from: <http://pfam.xfam.org/family/PF00170>

Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T et al. bZIP transcription factors in Arabidopsis. 2017.



Methods



Methods

Sequence alignments

1. *Jackhmmer* in Uniprot

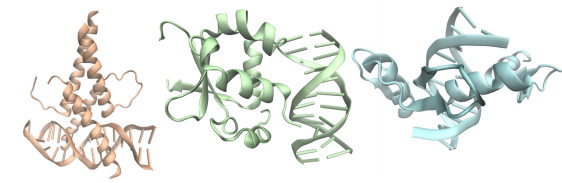


2. Different sequences of different species:
- E-value
 - Score

Pfam

1. Our domain of interest → .sto
2. *hmmbuild* → .hmm

hmmalign



Structural alignments

RCSB **PDB**
PROTEIN DATA BANK

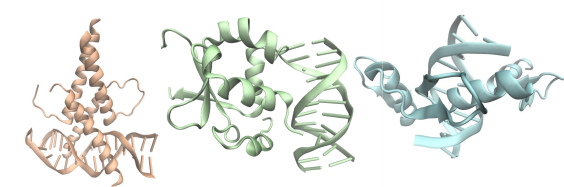
1. Rough STAMP

Sc ↑

RMS ↓

nfit ↑

Sc	
5,5 - 9,8	High structural similarity → Functional or evolutionary relationship
2'5 – 5'5	More distant structures → Not always a relationship
<2	Very low structural similarity



Methods

Sequence alignments

```
HMMER3/ [3.0 | March 2010]
NAME HSF_DNA-bind
ACC PF00447.12
DESC HSF-type DNA-binding
LENG 104
ALPH amino
RF no
CS no
MAP yes
DATE Tue Sep 27 00:48:17 2011
NSEQ 116
EFFN 1.823120
CKSUM 2614862979
GA 21.00 21.00
TC 21.00 21.20
NC 20.90 20.90
STATS LOCAL MSV -9.5960 0.71661
STATS LOCAL VITERBI -10.7616 0.71661
STATS LOCAL FORWARD -3.8688 0.71661
HMM
R S T V W Y G H I K L M N P Q
m->m m->i m->d i->m i->i d->m d->d
COMPO 2.74351 4.70511 3.03673 2.63232 2.82295 3.18039 3.46593 2.92735 2.45879 2.54435 3.71388 2.87060 3.59630 2.88227
2.86908 2.57350 2.91016 2.85669 4.41099 3.34413
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146
2.89801 2.37887 2.77519 2.98518 4.58477 3.61503
0.01566 4.56041 5.28276 0.61958 0.77255 0.00000 *
1 2.15702 4.93368 3.04594 2.33364 4.17667 3.35352 3.64963 3.15665 2.31217 3.06750 4.00295 3.06941 3.01381 2.87491
3.00744 2.49206 2.48679 3.02196 5.43289 3.52241 1 - -
2.68618 4.42225 2.77519 2.73123 3.46354 2.40513 3.72494 3.29354 2.67741 2.69355 4.24690 2.90347 2.73739 3.18146
```

```
# jackhmmmer :: iteratively search a protein sequence against a protein database
# HMMER 3.0 (March 2010); http://hmmerr.org/
# Copyright (C) 2010 Howard Hughes Medical Institute.
# Freely distributed under the GNU General Public License (GPLv3).
# - - - - -
# query sequence file: HSF1human.fa
# target sequence database: /cursos/20431/BE/databases/blastdat/uniprot_sprot.fasta
# - - - - -
```

Query: sp|Q00613|HSF1_HUMAN [L=529]
Description: Heat shock factor protein 1 OS=Homo sapiens GN=HSF1 PE=1 SV=1

Scores for complete sequences (score includes all domains):

--- full sequence ---				--- best 1 domain ---			-#dom-		Sequence	Description
E-value	score	bias	E-value	score	bias	exp	N			
+	0	1046.3	2.5	0	1046.2	1.8	1.0	1	sp P38532 HSF1_MOUSE	Heat shock factor protein 1 OS=Mus mus
+	0	1190.1	2.5	0	1189.9	1.8	1.0	1	sp Q00613 HSF1_HUMAN	Heat shock factor protein 1 OS=Homo sa
+	0	1051.4	2.9	0	1051.2	2.0	1.0	1	sp Q08DJ8 HSF1_BOVIN	Heat shock factor protein 1 OS=Bos tau
+	2.5e-249	833.2	0.4	4.2e-170	571.7	0.0	2.0	2	sp P38529 HSF1_CHICK	Heat shock factor protein 1 OS=Gallus
+	1.8e-167	563.0	5.4	7.3e-149	501.6	0.0	2.1	2	sp P41154 HSF_XENLA	Heat shock factor protein OS=Xenopus l
+	2e-86	295.6	0.1	1e-70	243.7	0.0	2.1	2	sp P38530 HSF2_CHICK	Heat shock factor protein 2 OS=Gallus
+	1.1e-85	293.1	0.6	1.1e-74	256.7	0.0	2.1	2	sp Q03933 HSF2_HUMAN	Heat shock factor protein 2 OS=Homo sa
+	2.6e-85	291.8	0.1	6.4e-75	257.6	0.0	2.1	2	sp P38533 HSF2_MOUSE	Heat shock factor protein 2 OS=Mus mus

Methods

Sequence alignments

1. *Jackhmmer* in Uniprot

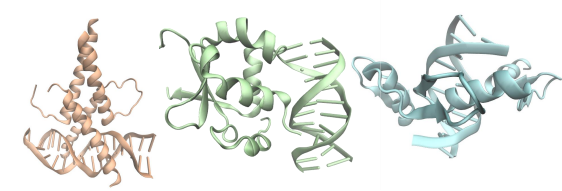


2. Different sequences of different species:
- E-value
 - Score

Pfam

1. Our domain of interest → .sto
2. *hmmbuild* → .hmm

hmmalign



Structural alignments



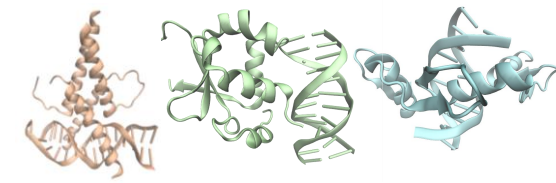
1. Rough STAMP

Sc ↑

RMS ↓

nfit ↑

Sc	
5,5 - 9,8	High structural similarity → Functional or evolutionary relationship
2'5 – 5'5	More distant structures → Not always a relationship
<2	Very low structural similarity



Methods

Structural alignments

STAMP Structural Alignment of Multiple Proteins

Version 4.4 (May 2010)

by Robert B. Russell & Geoffrey J. Barton

Please cite PROTEINS, v14, 309-323, 1992

Running roughfit.

Sc = STAMP score, RMS = RMS deviation, Align = alignment length

Len1, Len2 = length of domain, Nfit = residues fitted

Secs = no. equivalent sec. strucs. Eq = no. equivalent residues

%I = seq. identity, %S = sec. str. identity

P(m) = P value (p=1/10) calculated after Murzin (1993), JMB, 230, 689-694

(NC = P value not calculated - potential FP overflow)

	No.	Domain1	Domain2	Sc	RMS	Len1	Len2	Align	NFit	Eq.	Secs.	%I	%S	P(m)
Pair 1	5hdg	1fbu	5.63	1.15	94	90	91	68	68	0	57.35	100.00	7.58e-21	
Pair 2	5hdg	1hks	6.87	1.34	94	106	105	84	84	0	65.48	100.00	2.29e-33	
Pair 3	1fbu	1hks	5.01	1.58	90	106	95	68	67	0	50.75	100.00	4.40e-17	

Reading in matrix file HSF.mat...

Doing cluster analysis...

Cluster: 1 (5hdg & 1hks) Sc 6.87 RMS 1.34 Len 105 nfit 84

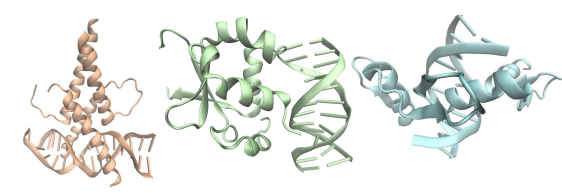
See file HSF.1 for the alignment and transformations

Cluster: 2 (1fbu & 5hdg 1hks) Sc 6.98 RMS 1.33 Len 99 nfit 71

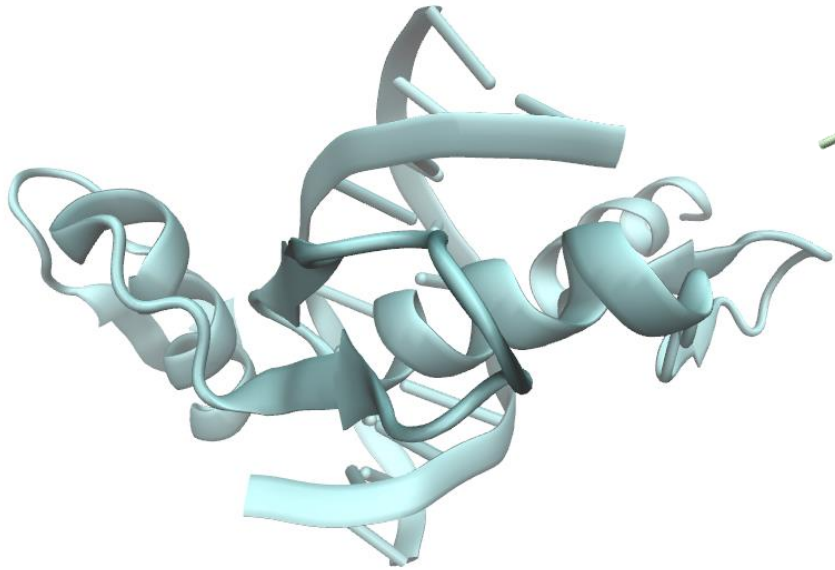
See file HSF.2 for the alignment and transformations

If the Sc is not
correct → Alignfit

Introduction

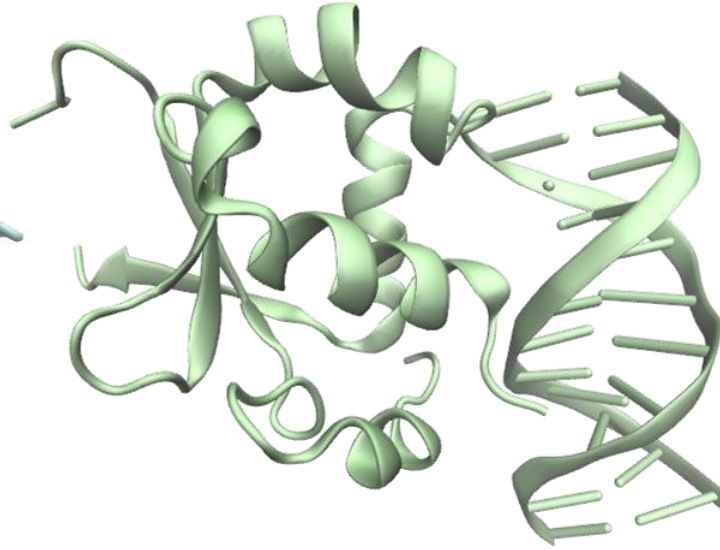


Zinc finger



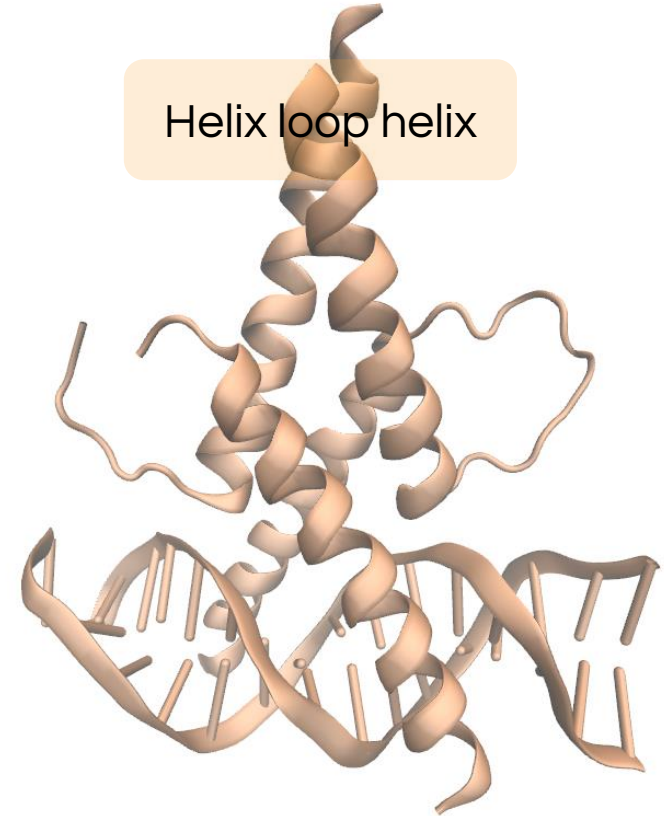
Krüppel-like factor 4 (*Mus musculus*)

Helix turn helix

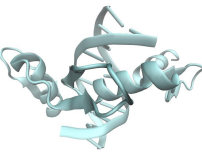


Heat-shock factor 1
(*Homo sapiens*)

Helix loop helix



Microphthalmia-associated
transcription factor (*Mus musculus*)



KLF4

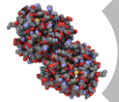
Index



Classification



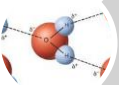
Biological function



General structure

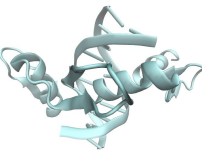


Secondary structure



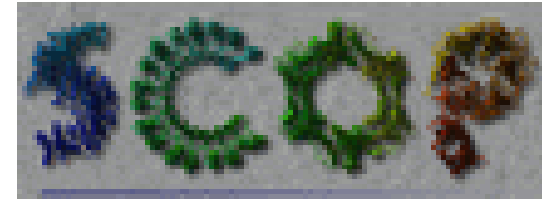
Important residues and interactions





KLF4

Classification



Family: Classic zinc finger, C2H2

Lineage

Root

Scop

Class

Small proteins

Fold

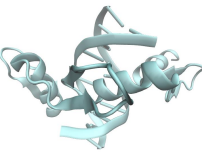
Beta-Beta-Apha Zinc Fingers

Superfamily

Beta-Beta-Apha Zinc Fingers

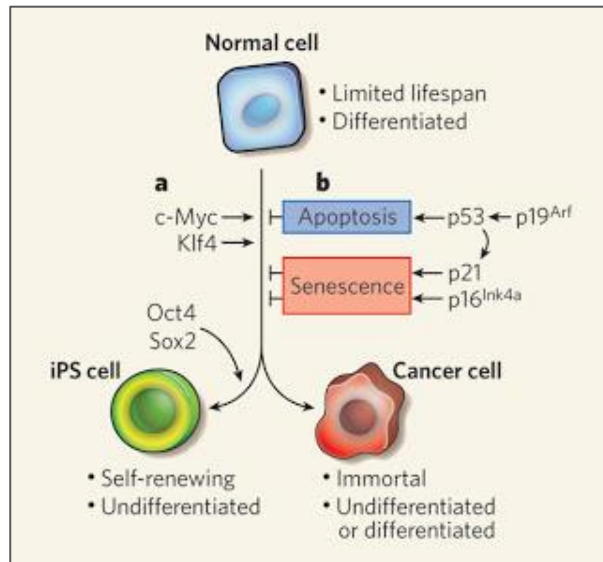
Family

Classic Zinc Finger, C2H2



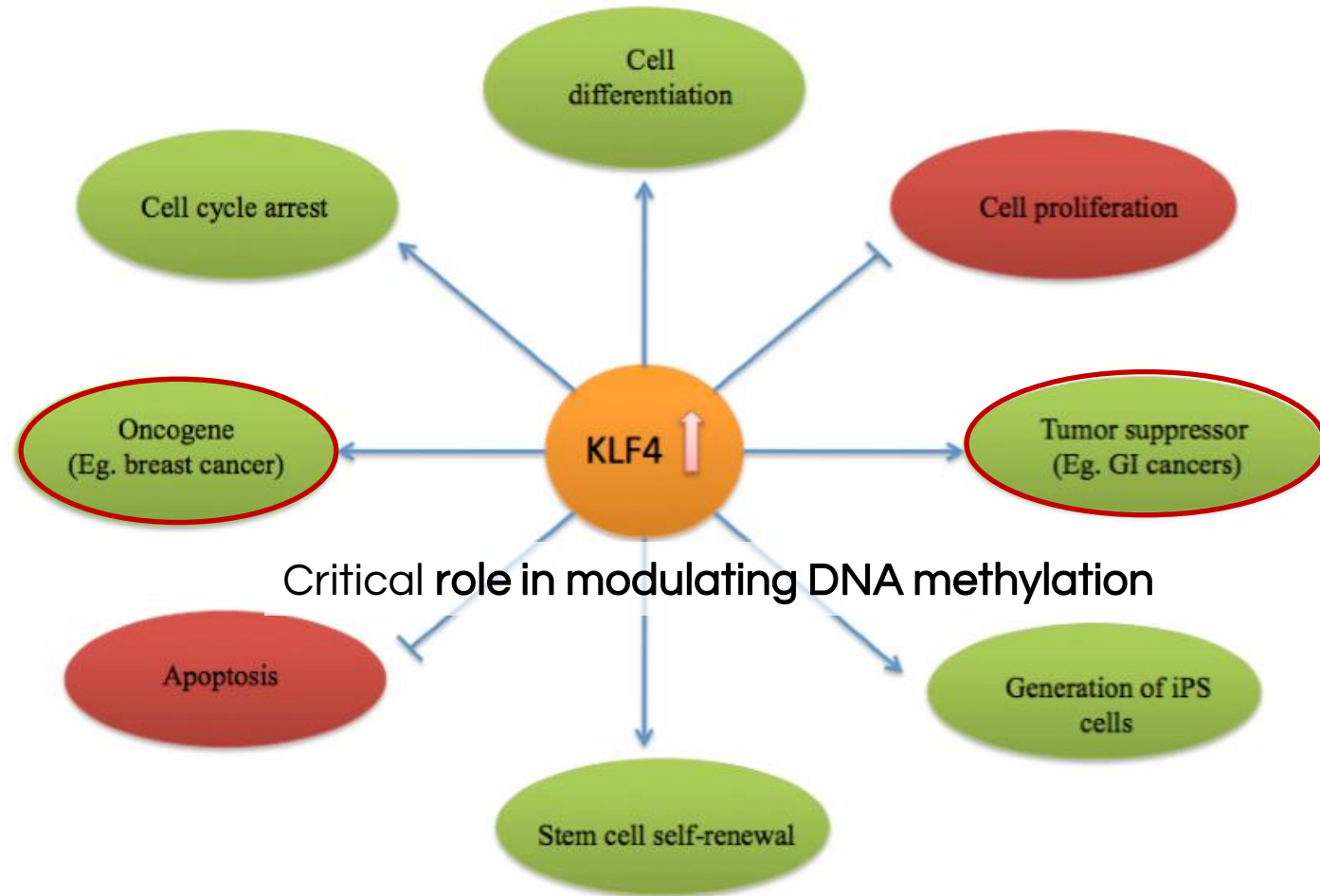
KLF4

Biological function

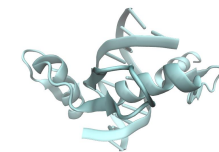


Yamanaka Factors

Role of Krüppel-like factors in leukocyte development, function, and disease.
Am Soc Hematology 2010.

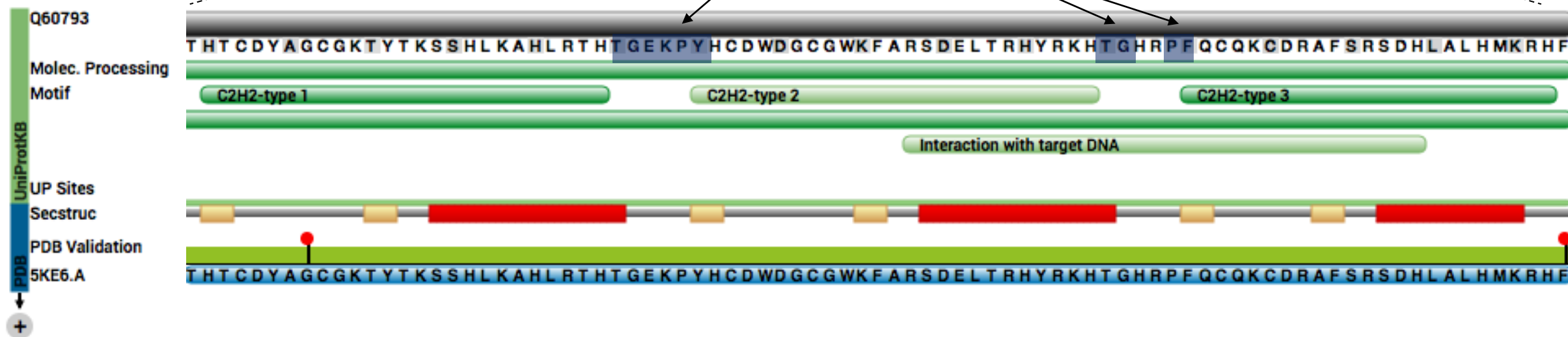
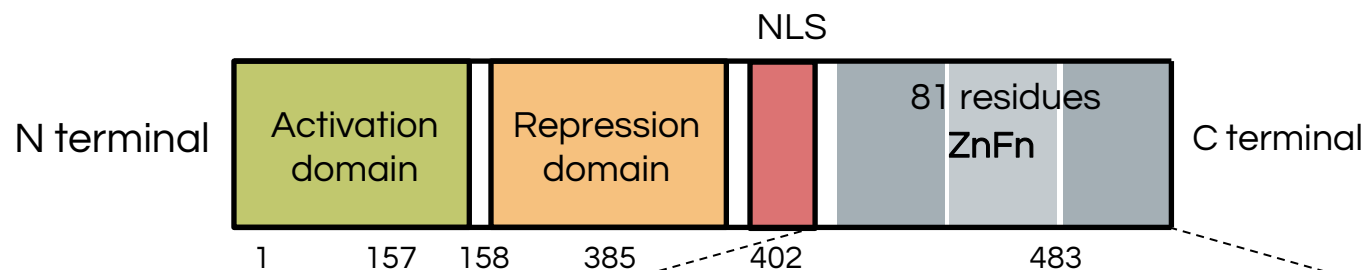


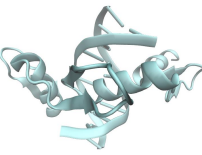
Ai et al. Kruppel like factor 4 (KLF4): a transcription factor with diverse context-dependent functions. *Gene Ther Mol Biol* Vol 13, 194-203, 2009



KLF4

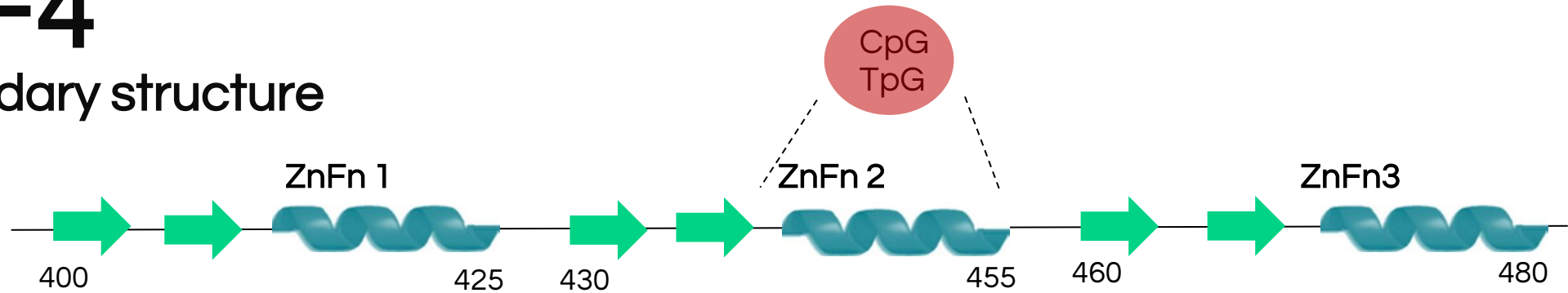
General structure





KLF4

Secondary structure

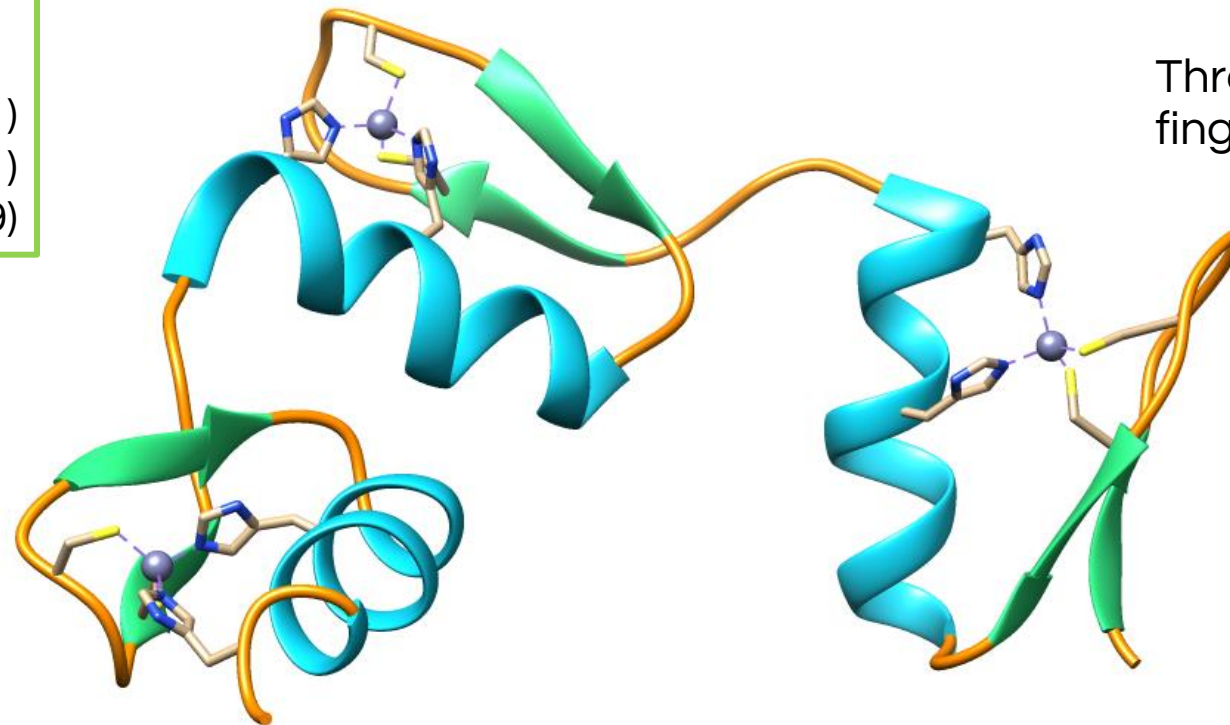


Beta-strands:

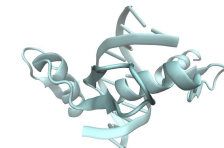
$\beta 1a$: (400-401) $\beta 1b$: (410-411)
 $\beta 2a$: (430-431) $\beta 2b$: (440-441)
 $\beta 3a$: (460-461) $\beta 3b$: (468-469)

Helix-alpha:

$\alpha 1$: (414-425)
 $\alpha 2$: (444-455)
 $\alpha 3$: (472-480)



Three tandem C2H2
fingers



KLF4

Secondary structure

Beta-strands:

β 1a: (400-401) β 1b: (410-411)
 β 2a: (430-431) β 2b: (440-441)
 β 3a: (460-461) β 3b: (468-469)

Helix-alpha:

α 1: (414-425)
 α 2: (444-455)
 α 3: (472-480)

ZnFn 1

ZnFn 2

ZnFn 3

sp|Q60793|KLF4_MOUSE
 sp|043474|KLF4_HUMAN
 tr|F6VVB3|KLF4_MACMU
 tr|A7YWE2|KLF4_BOVIN
 tr|H2QXN2|KLF4_PANTR
 tr|D3U661|KLF4_FELCA
 tr|Q90XE8|KLF4_DANRE
 tr|Q28D29|KLF4_XENTR
 tr|Q52JJ4|KLF4_PIG
 tr|G9JZK4|KLF4_BUBBU
 tr|A0A1A8AVD8|KLF4_NOTFU
 tr|A0A0B4UDA9|KLF4_RABIT
 tr|G1ERW8|KLF4_XENLA
 tr|C0KKX2|KLF4_NOTVI
 tr|M4T5C8|KLF4_MICLV
 tr|A0A077JN77|KLF4_CYNPY
 tr|A0A1A8HS30|KLF4_NOTKU
 tr|A0A142FHL6|KLF4_9CICH
 tr|A0A1A8PAT3|KLF4_9TELE
 tr|A0A166NG74|KLF4_CAPHI

atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHSCefpGCGKTYTKSSHLKAHLRTHt
 atHTCEyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyvgCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCEyaGCGKTYTKSSHLKAHLRTHt
 atHNCeypGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt
 atHTCDyvgCGKTYTKSSHLKAHLRTHt
 atHTCDfvGCGKTYTKSSHLKAHLRTHt
 atHTCDyvgCGKTYTKSSHLKAHLRTHt
 atHTCDyaGCGKTYTKSSHLKAHLRTHt

kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCSweGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDweGCGWKFARSDDELTRHYRKH
 kpYHCDwdGCGWKFARSDDELTRHYRKH

hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCCHLCERAFSRSDHLALHMKRHm-
 hrpFQCQRCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-
 hrpFQCQKCDRAFSRSDHLALHMKRHf-

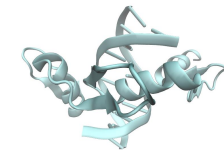
Zebra fish

Eastern newt

zebra mbuna

403: Negatively charged 421: Polar residues

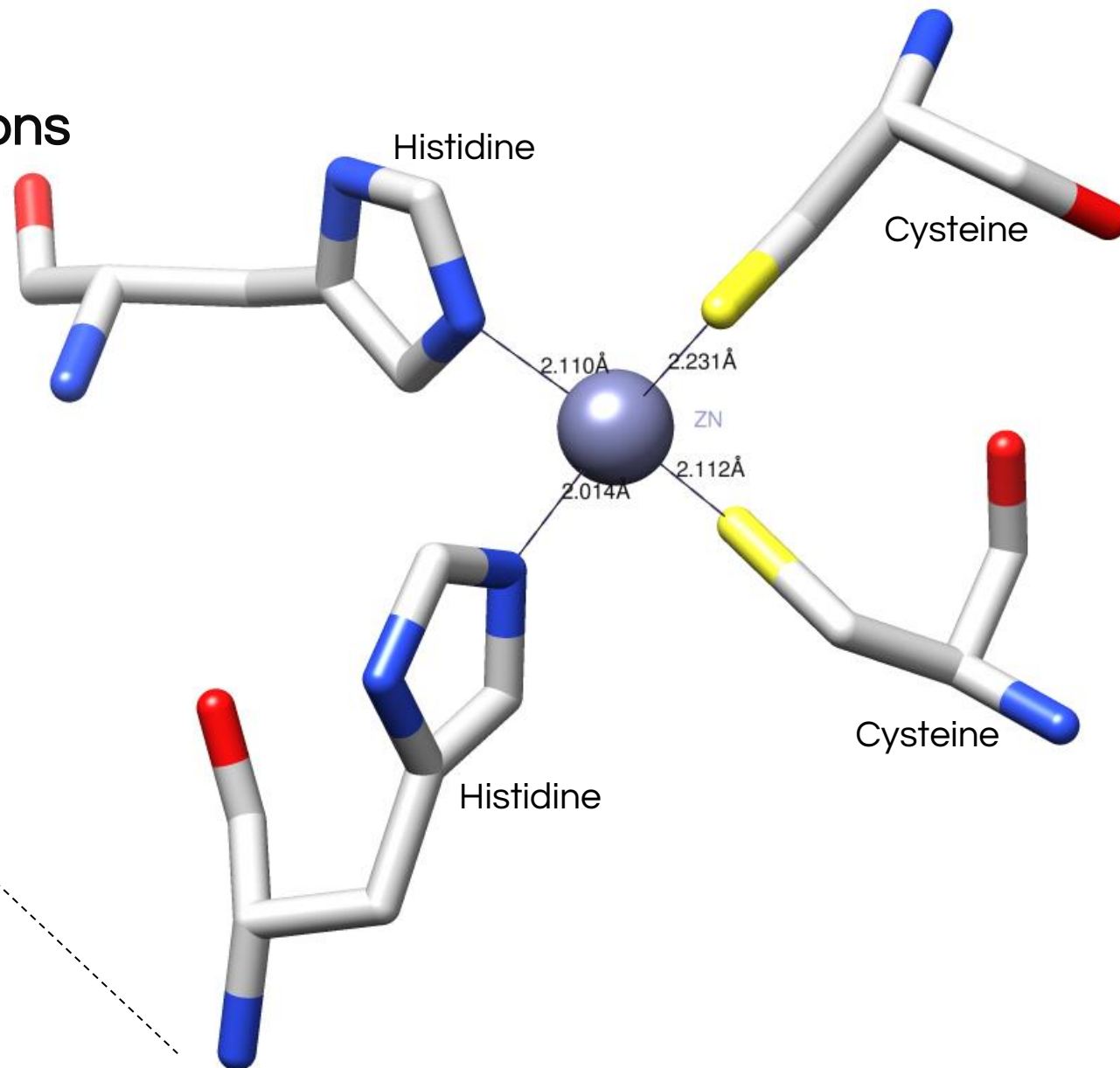
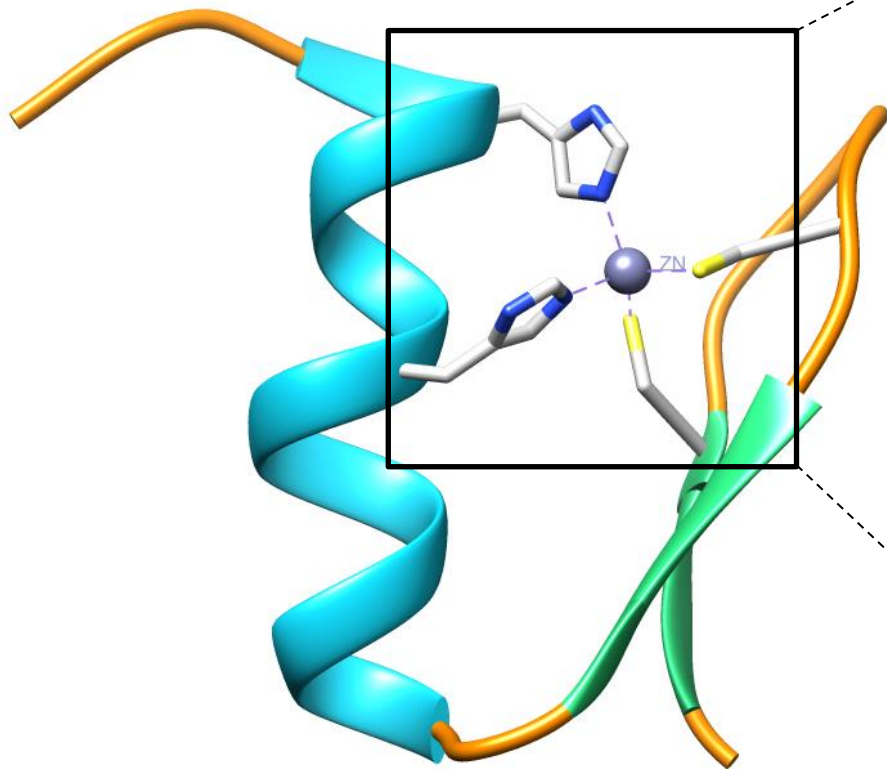
466: Negatively charged



KLF4

Important residues and interactions

Residues that interact with the zinc

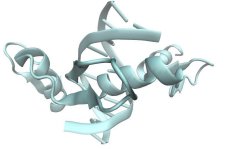


KLF4

Important residues and interactions

Residues that interact with the zinc

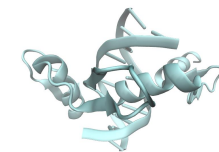
Zn finger domain	Residues numbers	
ZnFn1	Cys 402	His 420
	Cys 407	His 424
ZnFn2	Cys 432	His 478
	Cys 437	His 482
ZnFn3	Cys 463	His 478
	Cys 465	His 482



ZnFn 1

sp Q60793 KLF4_MOUSE	atHTCDyaGCGKTYTKSSHLKAHLRTHt
sp 043474 KLF4_HUMAN	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr F6VVB3 KLF4_MACMU	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A7YWE2 KLF4_BOVIN	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr H2QXN2 KLF4_PANTR	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr D3U661 KLF4_FELCA	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr Q90XE8 KLF4_DANRE	atHSCEfpGCGKTYTKSSHLKAHMRTHt
tr Q28D29 KLF4_XENTR	atHTCEyaGCGKTYTKSSHLKAHLRTHt
tr Q52JJ4 KLF4_PIG	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr G9JZK4 KLF4_BUBBU	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A1A8AVD8 KLF4_NOTFU	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A0B4UDA9 KLF4_RABIT	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr G1ERW8 KLF4_XENLA	atHTCEyaGCGKTYTKSSHLKAHLRTHt
tr C0KKX2 KLF4_NOTVI	atHNCEypGCGKTYTKSSHLKAHMRTHt
tr M4T5C8 KLF4_MICLV	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A077JN77 KLF4_CYNPY	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A1A8HS30 KLF4_NOTKU	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A142FHL6 KLF4_9CICH	atHTCDfvGCGKTYTKSSHLKAHHRTHt
tr A0A1A8PAT3 KLF4_9TELE	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A166NG74 KLF4_CAPHI	atHTCDyaGCGKTYTKSSHLKAHLRTHt

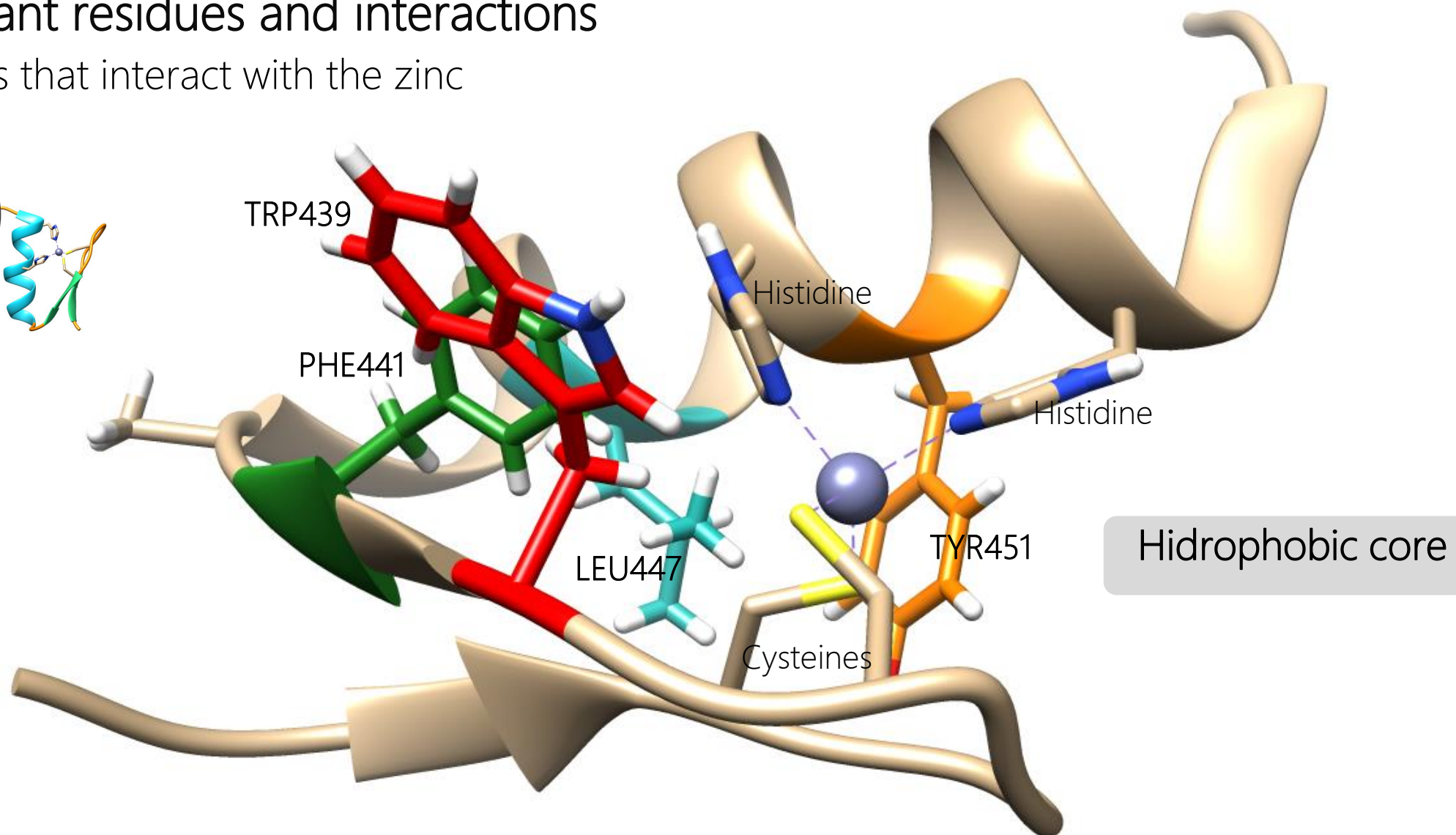
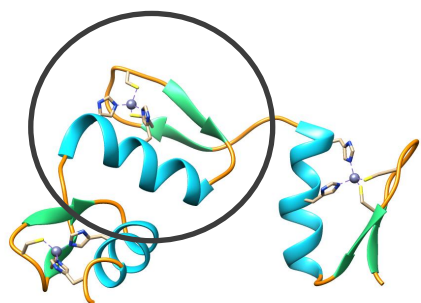
402 407 420 424

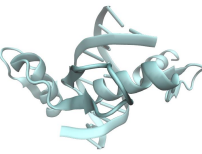


KLF4

Important residues and interactions

Residues that interact with the zinc

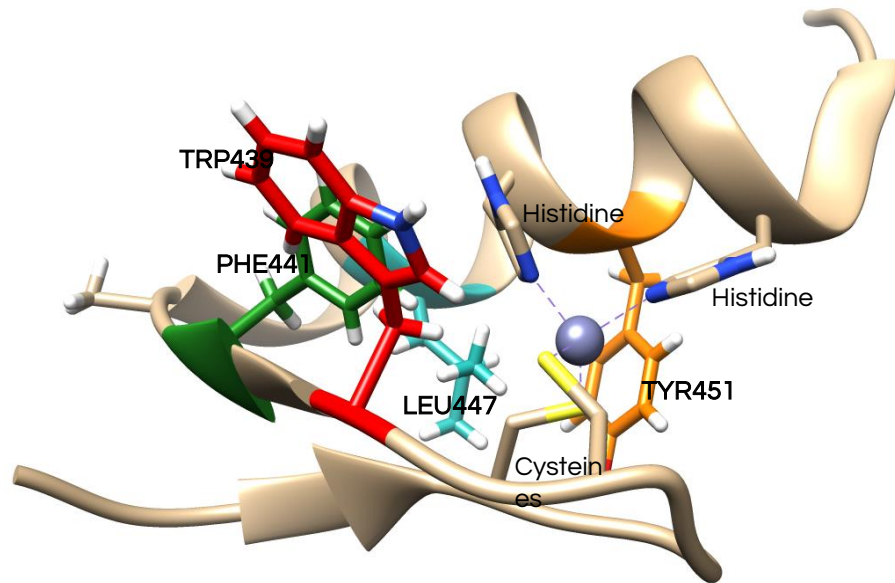




KLF4

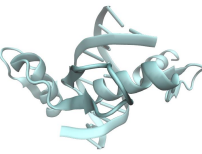
Important residues and interactions

Residues that interact with the zinc



ZnFn 2

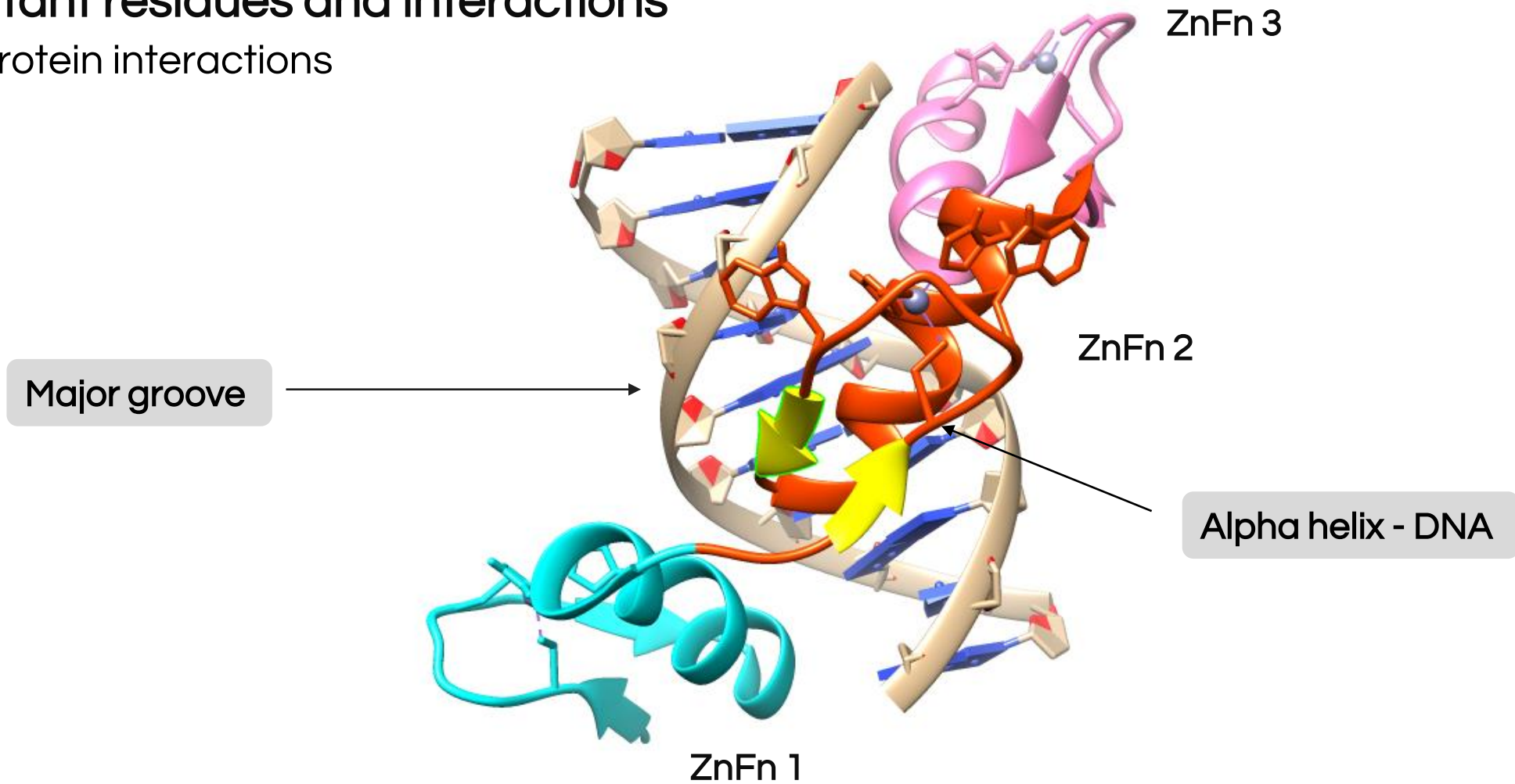
sp	Q60793	KLF4_MOUSE	kpYHCDwdGCGWKFARSDELTRHYRKH
sp	043474	KLF4_HUMAN	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	F6VVB3	KLF4_MACMU	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	A7YWE2	KLF4_BOVIN	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	H2QXN2	KLF4_PANTR	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	D3U661	KLF4_FELCA	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	Q90XE8	KLF4_DANRE	kpYHCSweGCGWKFARSDELTRHYRKH
tr	Q28D29	KLF4_XENTR	kpYHCDweGCGWKFARSDELTRHYRKH
tr	Q52JJ4	KLF4_PIG	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	G9JZK4	KLF4_BUBBU	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	A0A1A8AVD8	KLF4_NOTFU	kpYHCDweGCGWKFARSDELTRHYRKH
tr	A0A0B4UDA9	KLF4_RABIT	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	G1ERW8	KLF4_XENLA	kpYHCDweGCGWKFARSDELTRHYRKH
tr	C0KKX2	KLF4_NOTVI	kpYHCNweGCGWKFARSDELTRHYRKH
tr	M4T5C8	KLF4_MICLV	kpYHCDwdGCGWKFARSDELTRHYRKH
tr	A0A077JN77	KLF4_CYNPY	kpYHCDweGCGWKFARSDELTRHYRKH
tr	A0A1A8HS30	KLF4_NOTKU	kpYHCDweGCGWKFARSDELTRHYRKH
tr	A0A142FHL6	KLF4_9CICH	kpYHCDweGCGWKFARSDELTRHYRKH
tr	A0A1A8PAT3	KLF4_9TELE	kpYHCDweGCGWKFARSDELTRHYRKH
tr	A0A166NG74	KLF4_CAPHI	kpYHCDwdGCGWKFARSDELTRHYRKH

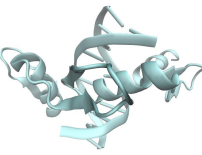


KLF4

Important residues and interactions

DNA-protein interactions

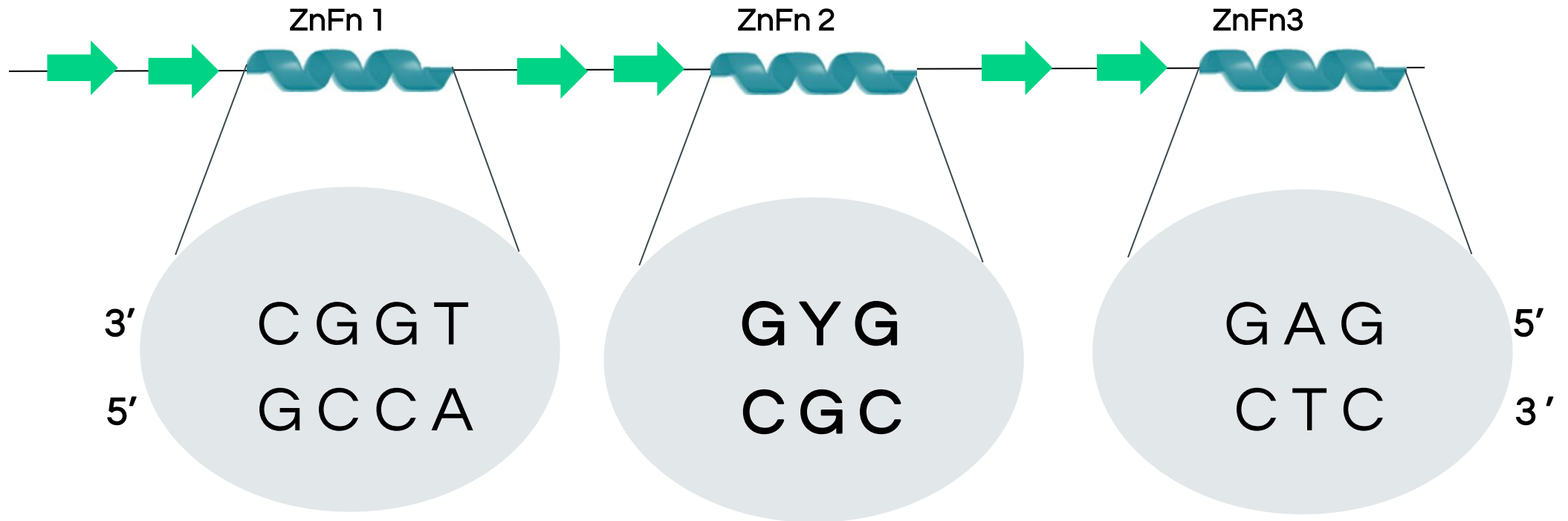


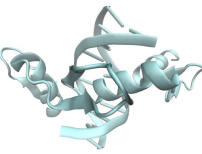


KLF4

Important residues and interactions

DNA-protein interactions



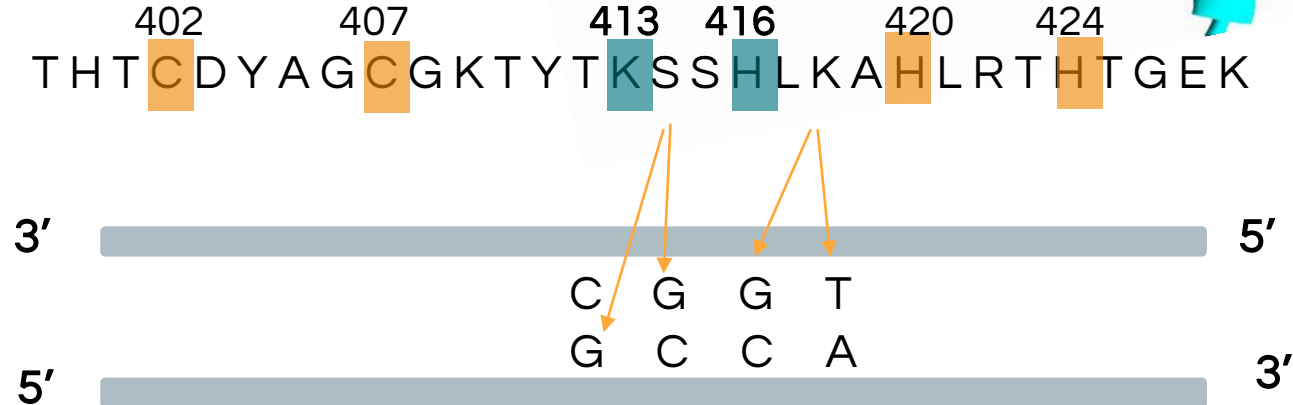


KLF4

Important residues and interactions

DNA-protein interactions

ZnFn 1



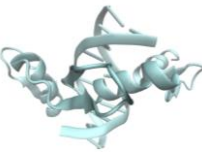
SPECIFIC INTERACTIONS:

NH₃ group of Lys 413 - G₉ (supper strand) and G₁ (lower strand) Hydrogen bonds

NH group of His 416 - TpG dinucleotide

→ Hydrogenbond with G₈ → Super base pair

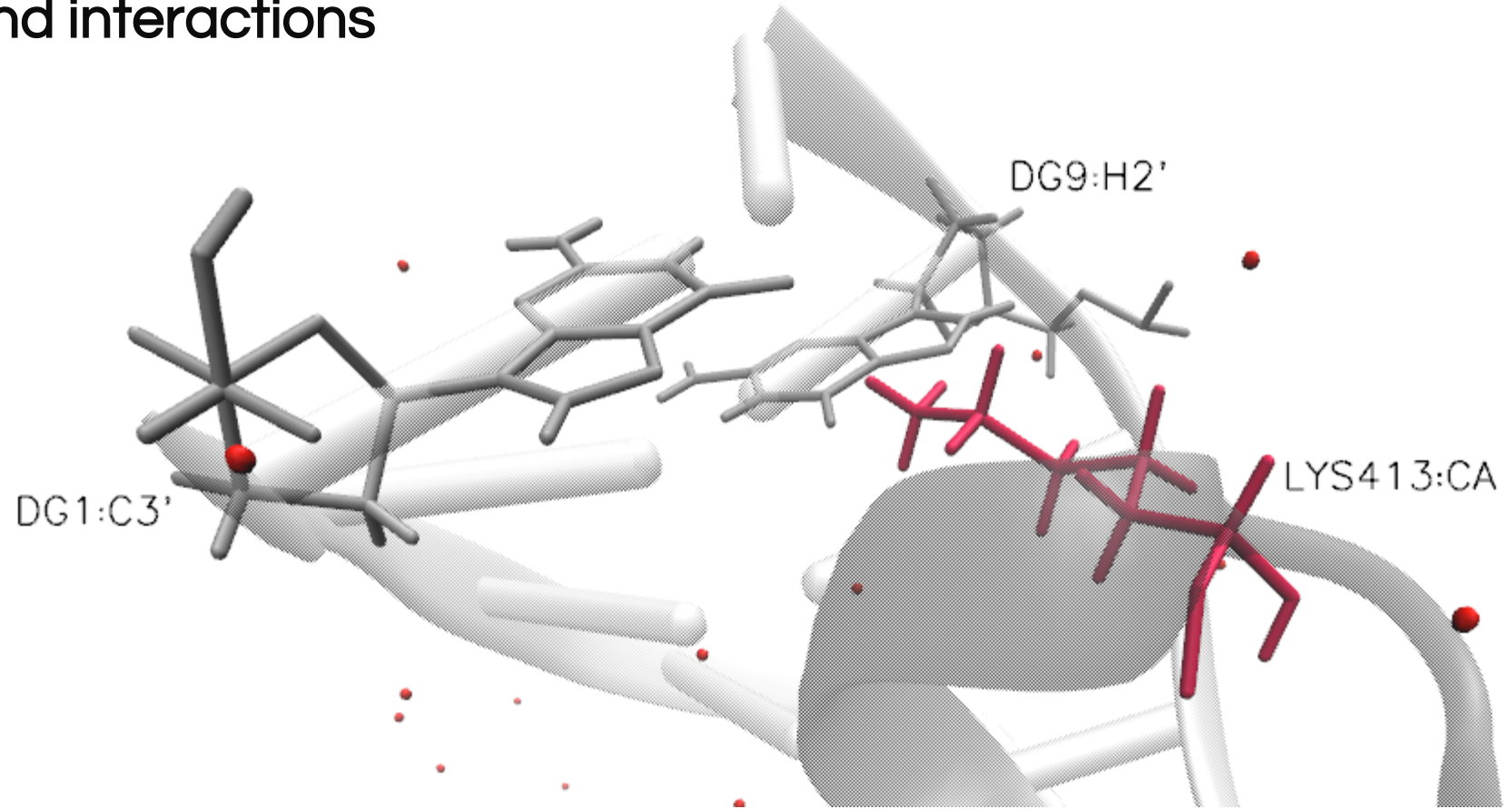
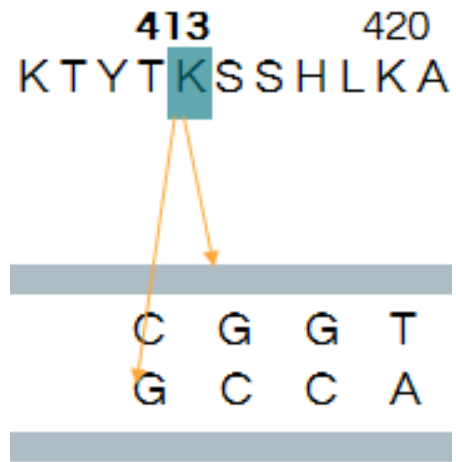
→ Stacking interaction with T₇

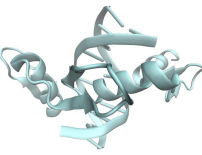


KLF4

Important residues and interactions

DNA-protein interactions



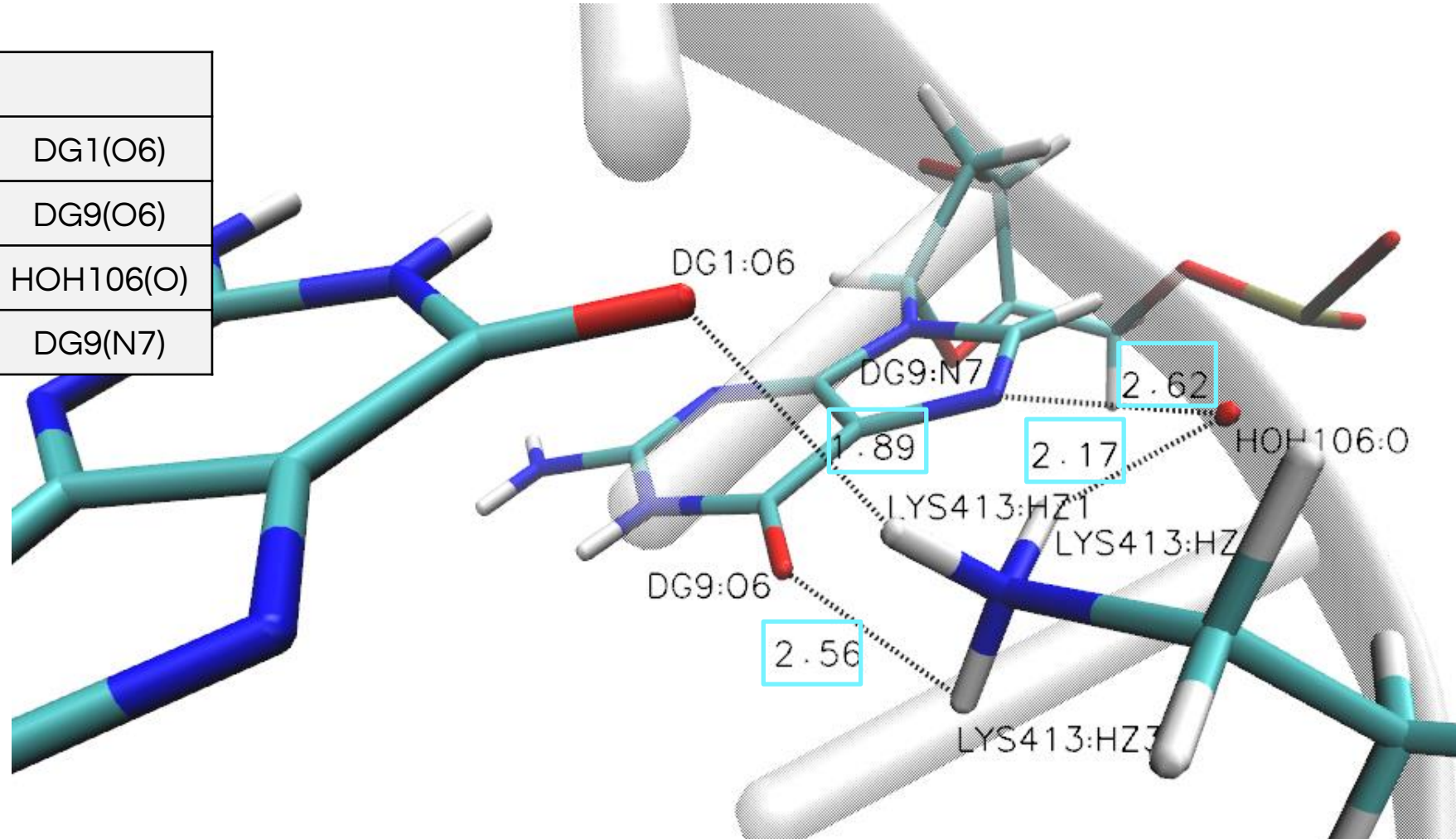
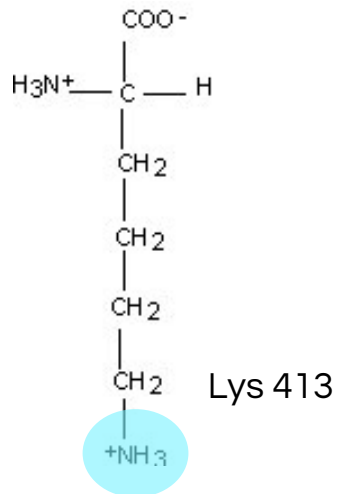


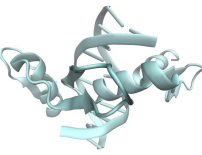
KLF4

Important residues and interactions

DNA-protein interactions

Hydrogen bonds		
Lys 413 (-NH ₃ ⁺)	1,89 Å	DG1(O6)
Lys 413 (-NH ₃ ⁺)	2,56 Å	DG9(O6)
Lys 413 (-NH ₃ ⁺)	2,17 Å	HOH106(O)
HOH106(O)	2,62 Å	DG9(N7)

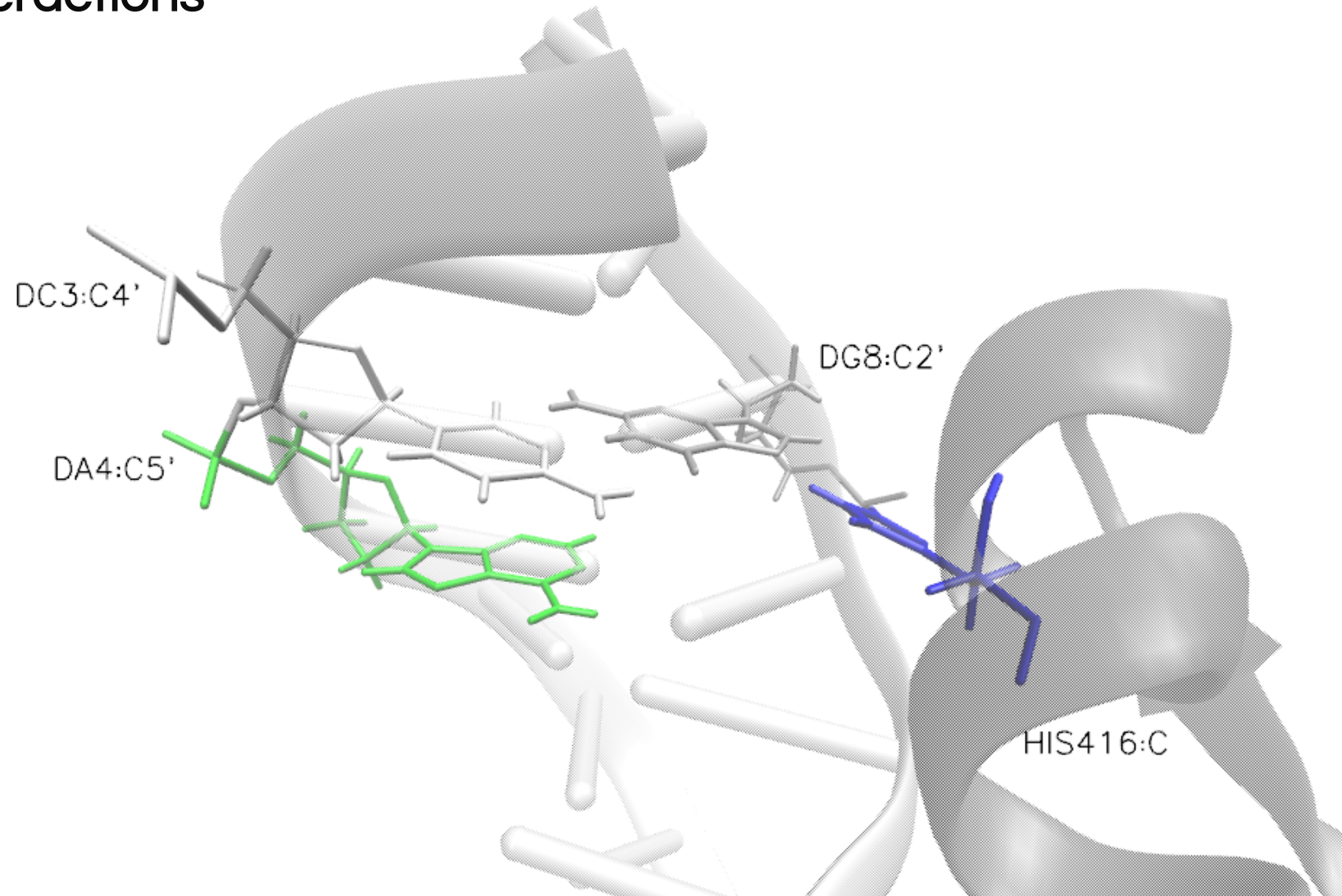
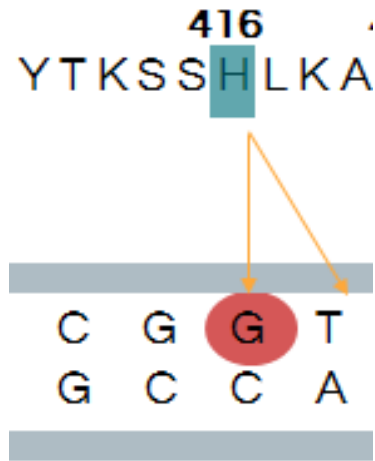


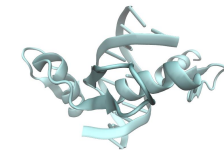


KLF4

Important residues and interactions

DNA-protein interactions



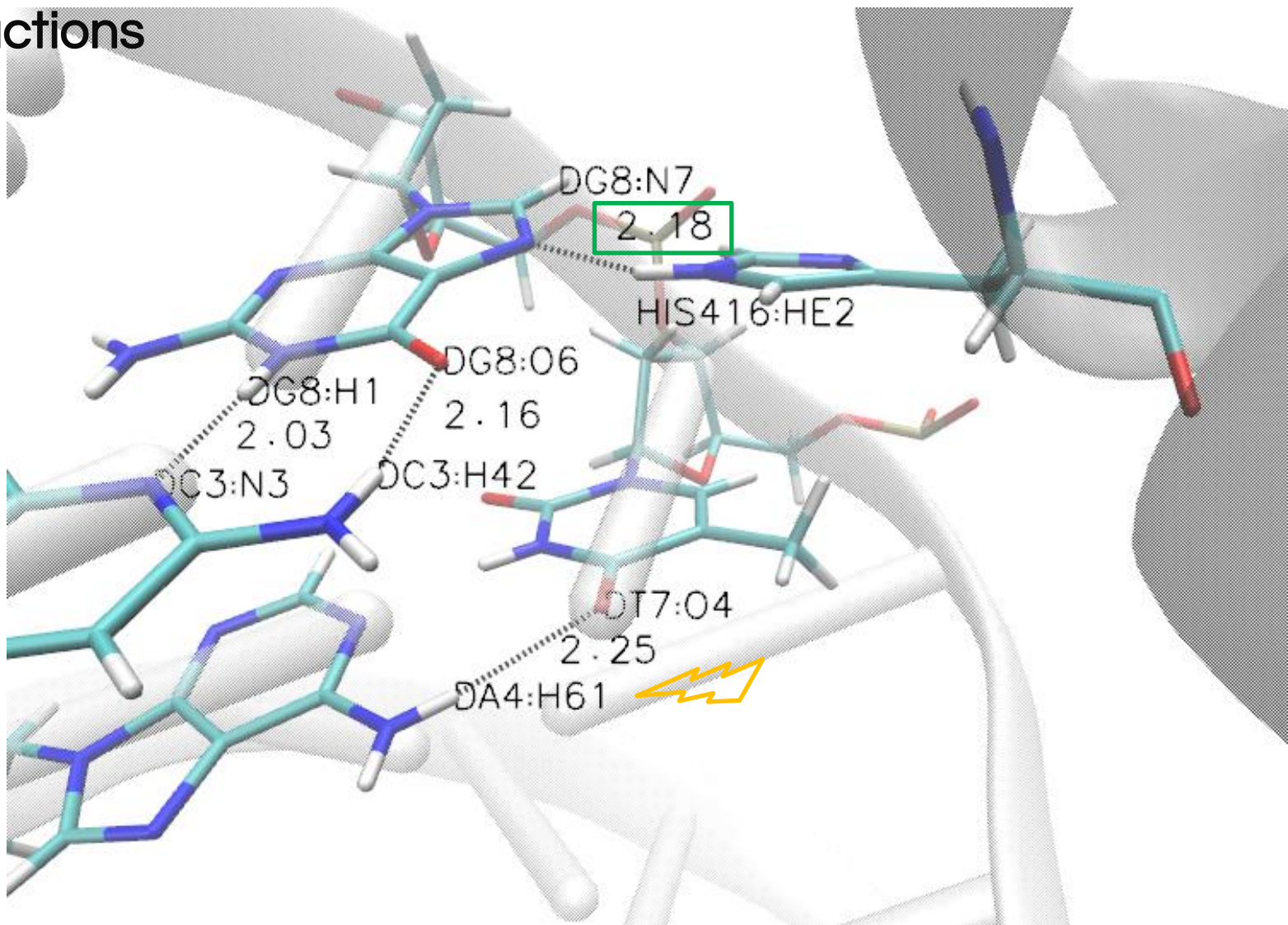
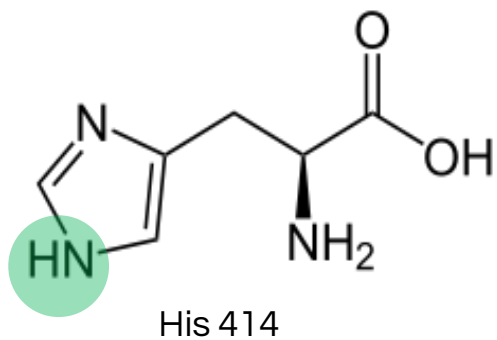


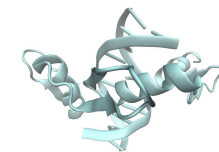
KLF4

Important residues and interactions

DNA-protein interactions

Hydrogen bonds		
His 414 (-NH)	2,18 Å	DG8(N7)
DG8 (H1)	2,03 Å	DC3(N3)
DG8 (O6)	2,16 Å	DC3(H42)
DT7 (O4)	2,25 Å	DA4 (H61)

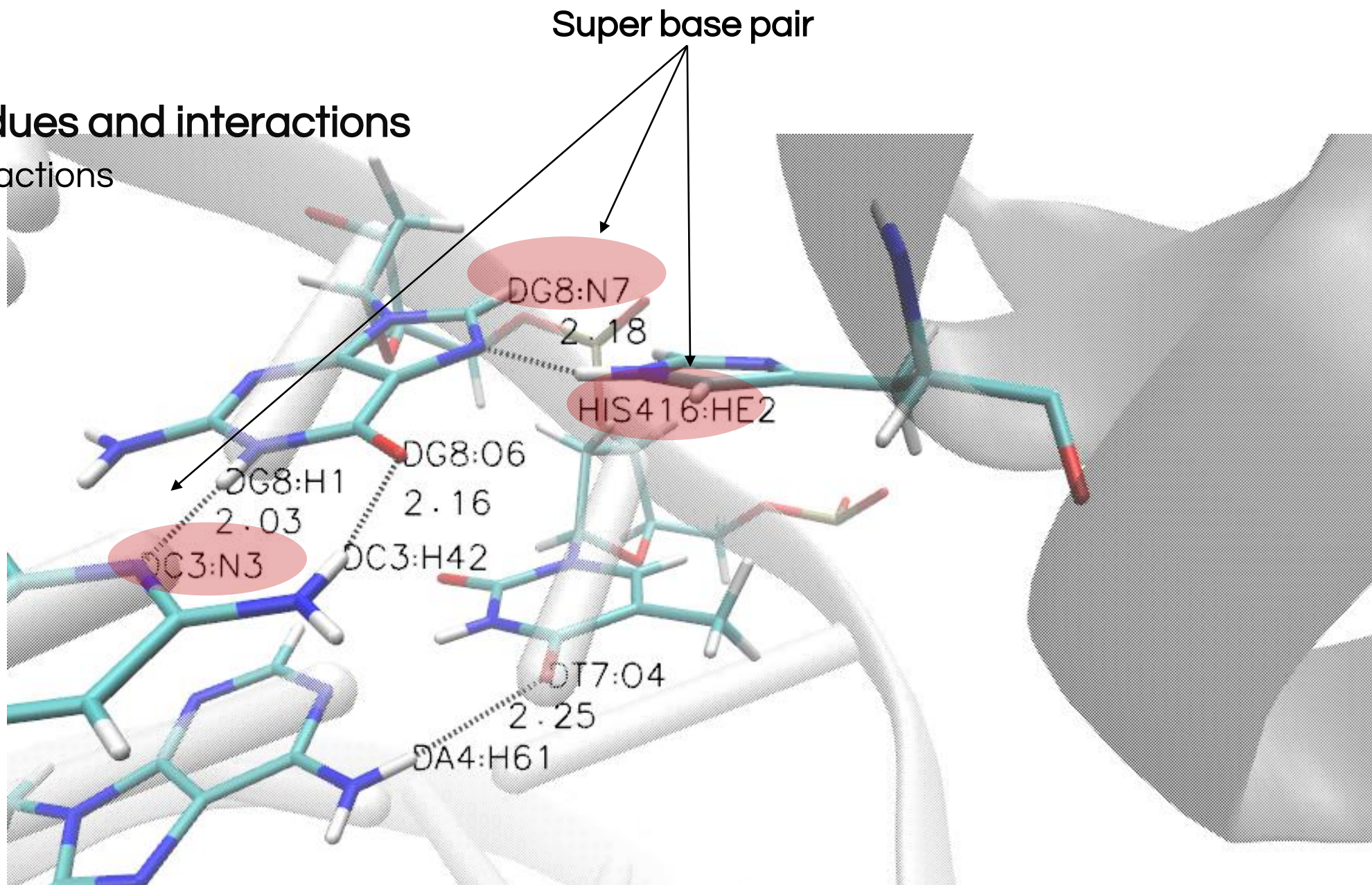


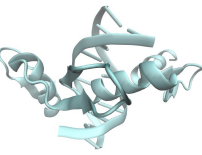


KLF4

Important residues and interactions

DNA-protein interactions

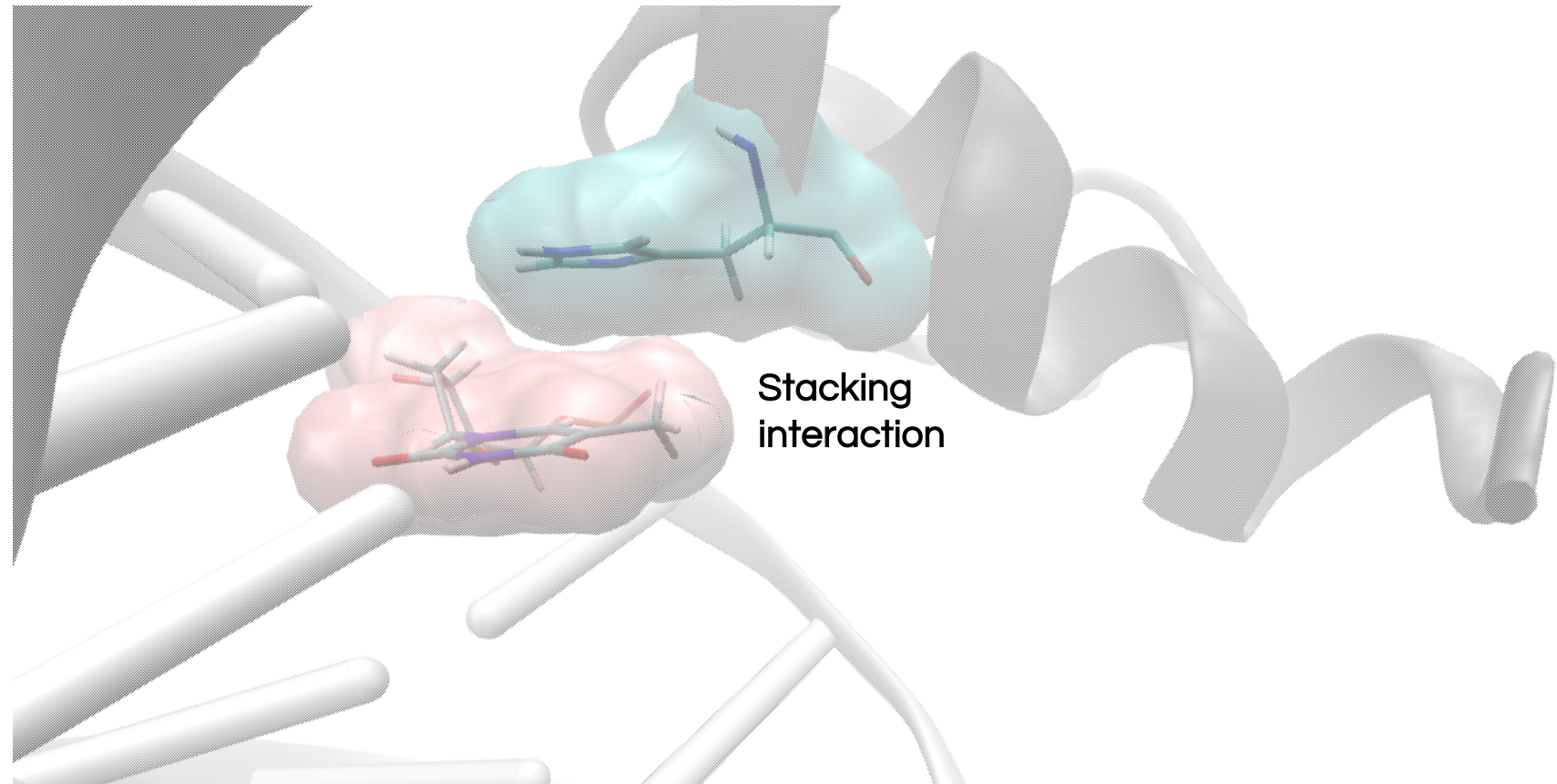
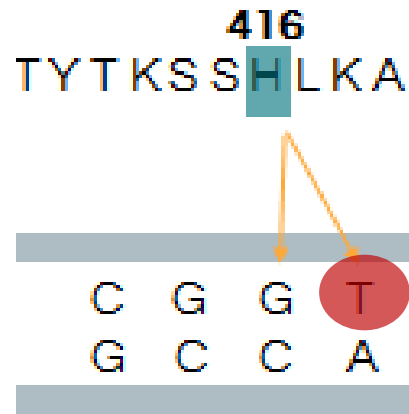




KLF4

Important residues and interactions

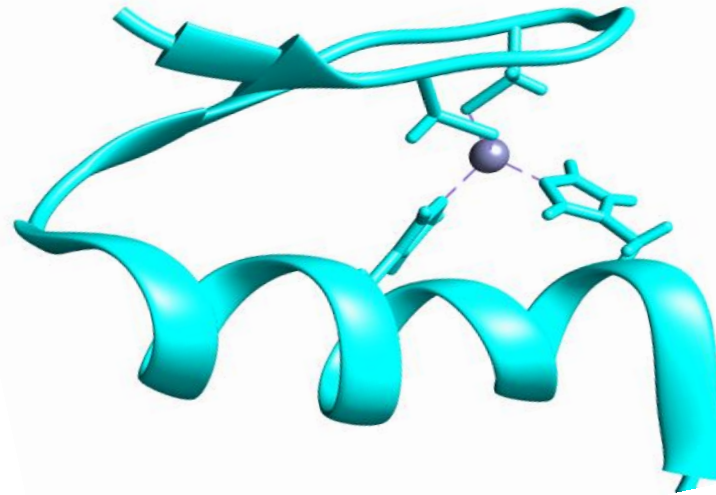
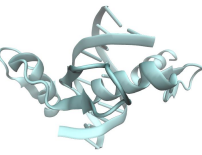
DNA-protein interactions



KLF4

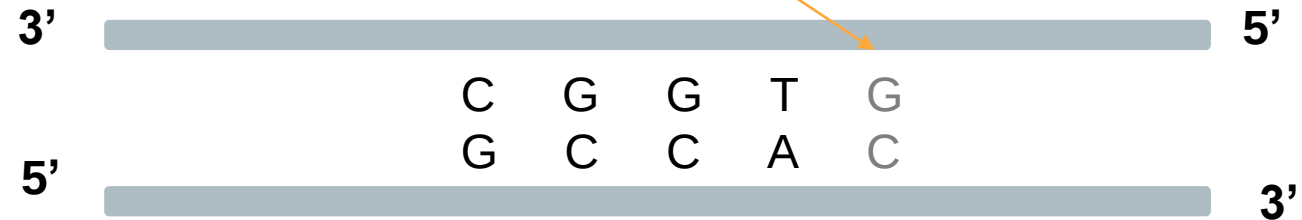
Important residues and interactions

DNA-protein interactions



ZnFn 1

402 407 411 420 424
THT **C** DYAG **C** GKT **Y** TKSSHLKA **H** LRT **H** TGEK



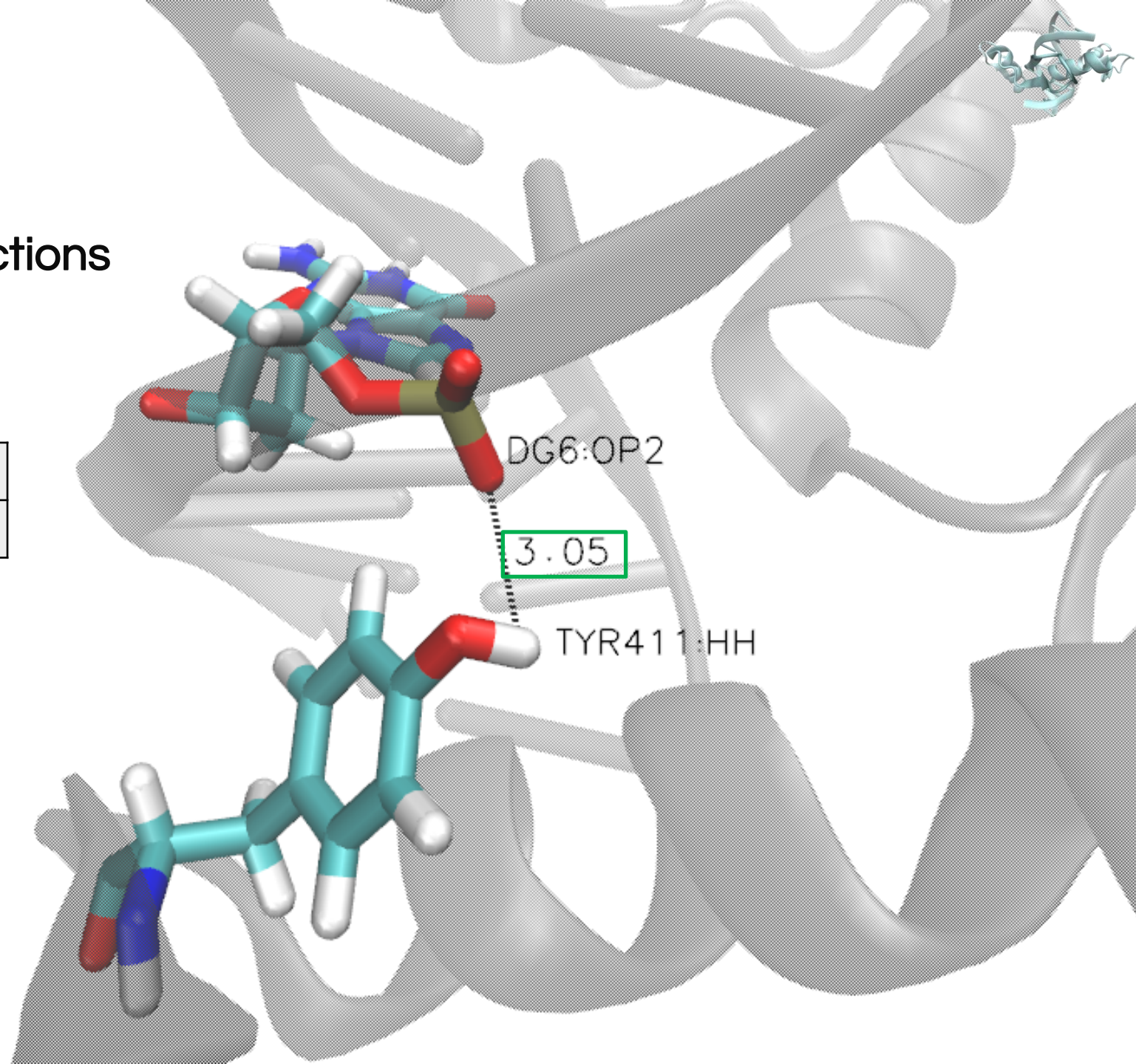
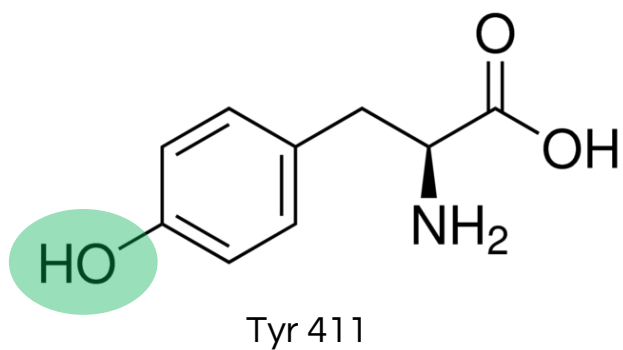
UNESPECIFIC INTERACTIONS: -OH group of **tyrosine** - Phosphate from **Guanine**

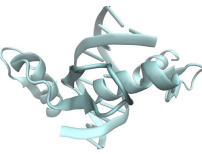
KLF4

Important residues and interactions

DNA-protein interactions

Hydrogen bond		
Tyr 411 (-OH)	3,05 Å	DG6(OP)





KLF4

Important residues and interactions

DNA-protein interactions

All conserved

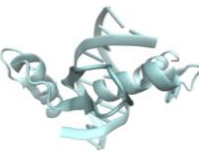
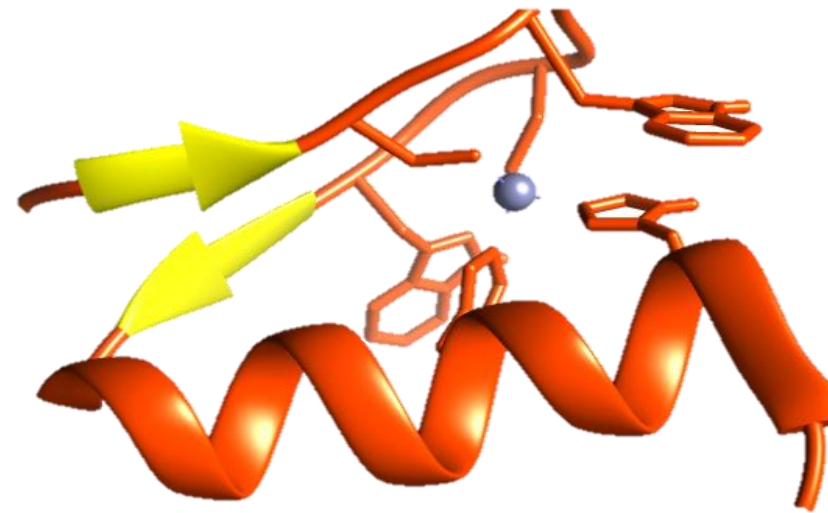
ZnFn 1

sp Q60793 KLF4_MOUSE	atHTCDyaGCGKTYTKSSHLKAHLRTHt
sp 043474 KLF4_HUMAN	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr F6VVB3 KLF4_MACMU	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A7YWE2 KLF4_BOVIN	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr H2QXN2 KLF4_PANTR	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr D3U661 KLF4_FELCA	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr Q90XE8 KLF4_DANRE	atHSCEfpGCGKTYTKSSHLKAHMRTHt
tr Q28D29 KLF4_XENTR	atHTCEyaGCGKTYTKSSHLKAHLRTHt
tr Q52JJ4 KLF4_PIG	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr G9JZK4 KLF4_BUBBU	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A1A8AVD8 KLF4_NOTFU	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A0B4UDA9 KLF4_RABIT	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr G1ERW8 KLF4_XENLA	atHTCEyaGCGKTYTKSSHLKAHLRTHt
tr C0KKX2 KLF4_NOTVI	atHNCEypGCGKTYTKSSHLKAHMRTHt
tr M4T5C8 KLF4_MICLV	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A077JN77 KLF4_CYNPY	atHTCDyaGCGKTYTKSSHLKAHLRTHt
tr A0A1A8HS30 KLF4_NOTKU	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A142FHL6 KLF4_9CICH	atHTCDFvGCGKTYTKSSHLKAHHRTHt
tr A0A1A8PAT3 KLF4_9TELE	atHTCDyvGCGKTYTKSSHLKAHHRTHt
tr A0A166NG74 KLF4_CAPHI	atHTCDyaGCGKTYTKSSHLKAHLRTHt

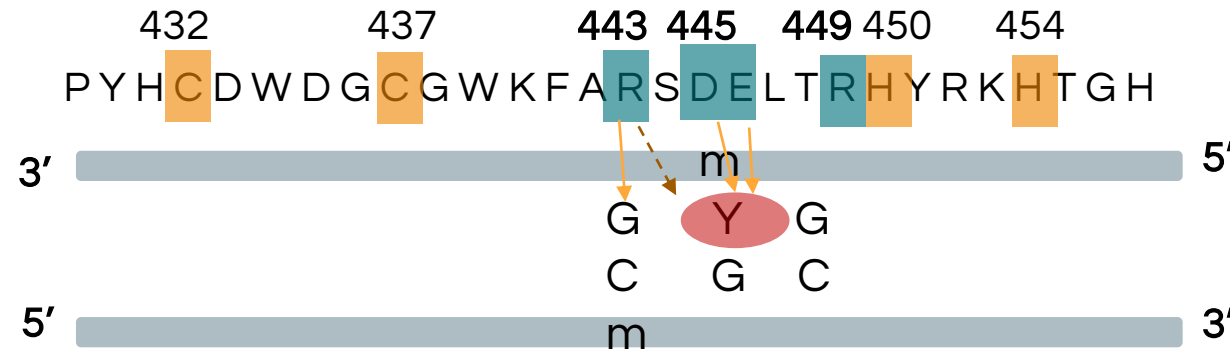
KLF4

Important residues and interactions

DNA-protein interactions



ZnFn 2



SPECIFIC INTERACTIONS:

Arg 443 (NH₃)

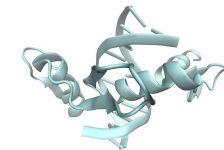
- Guanine (N, O) Hydrogen bond
- 5m-Thymine (CH₃) Hydrogen bond CH-O type
- Asp445 (O) Salt bridge

Asp 445 (N) - 5m-Thymine (CH₃)

Van der Waals

Glu 446 (-COOH) - 5m-Thymine (CH₃)

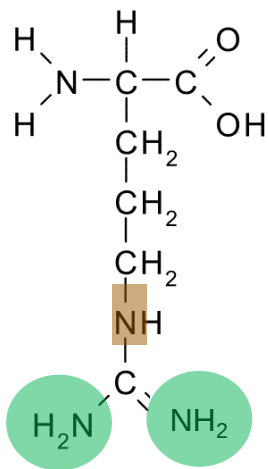
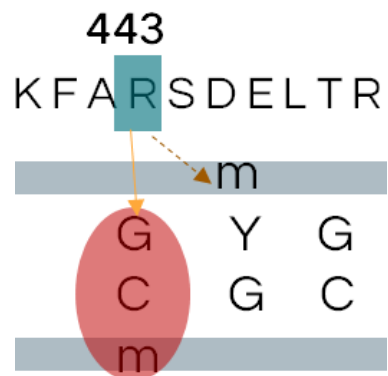
Hydrogen bond CH..O type



KLF4

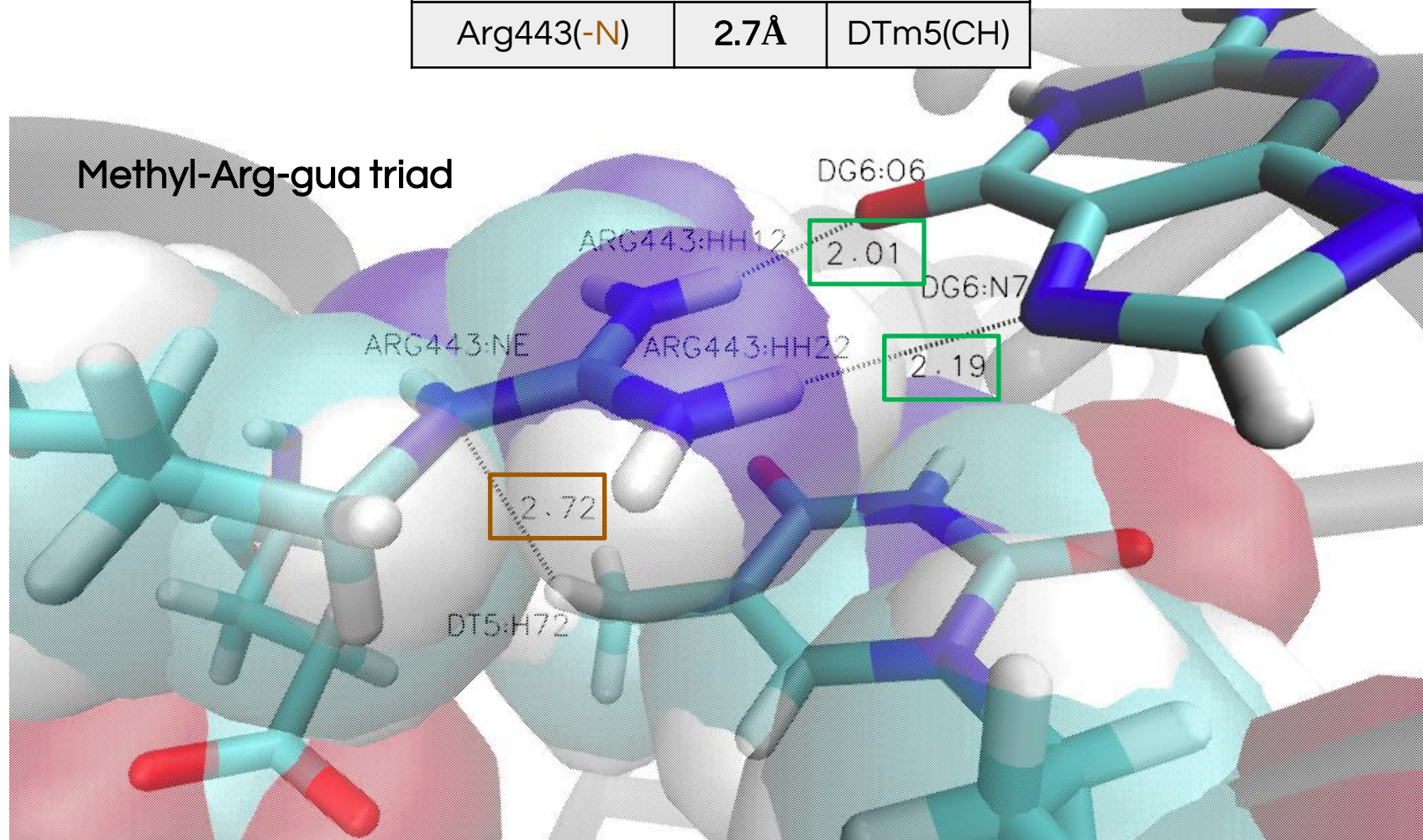
Important residues and interactions

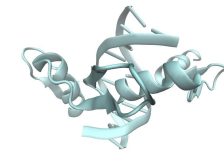
DNA-protein interactions



Arg443

Hydrogen bonds		
Arg443 (-NH)	2,01 Å	DG6(O)
Arg443 (-NH)	2,19 Å	DG6(N)
Van der Waals		
Arg443 (-N)	2.7Å	DTm5(CH)





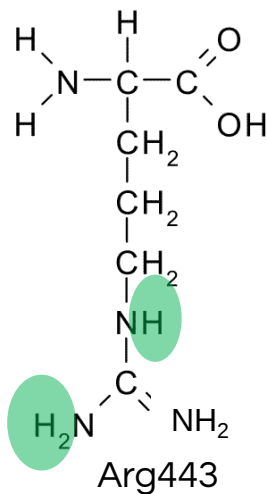
KLF4

Important residues and interactions

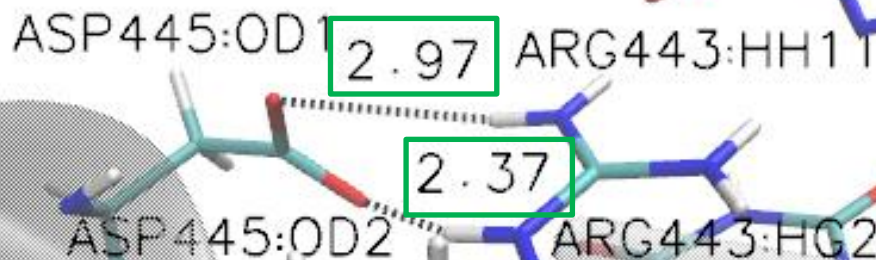
DNA-protein interactions

443 445
A R S D E L T R

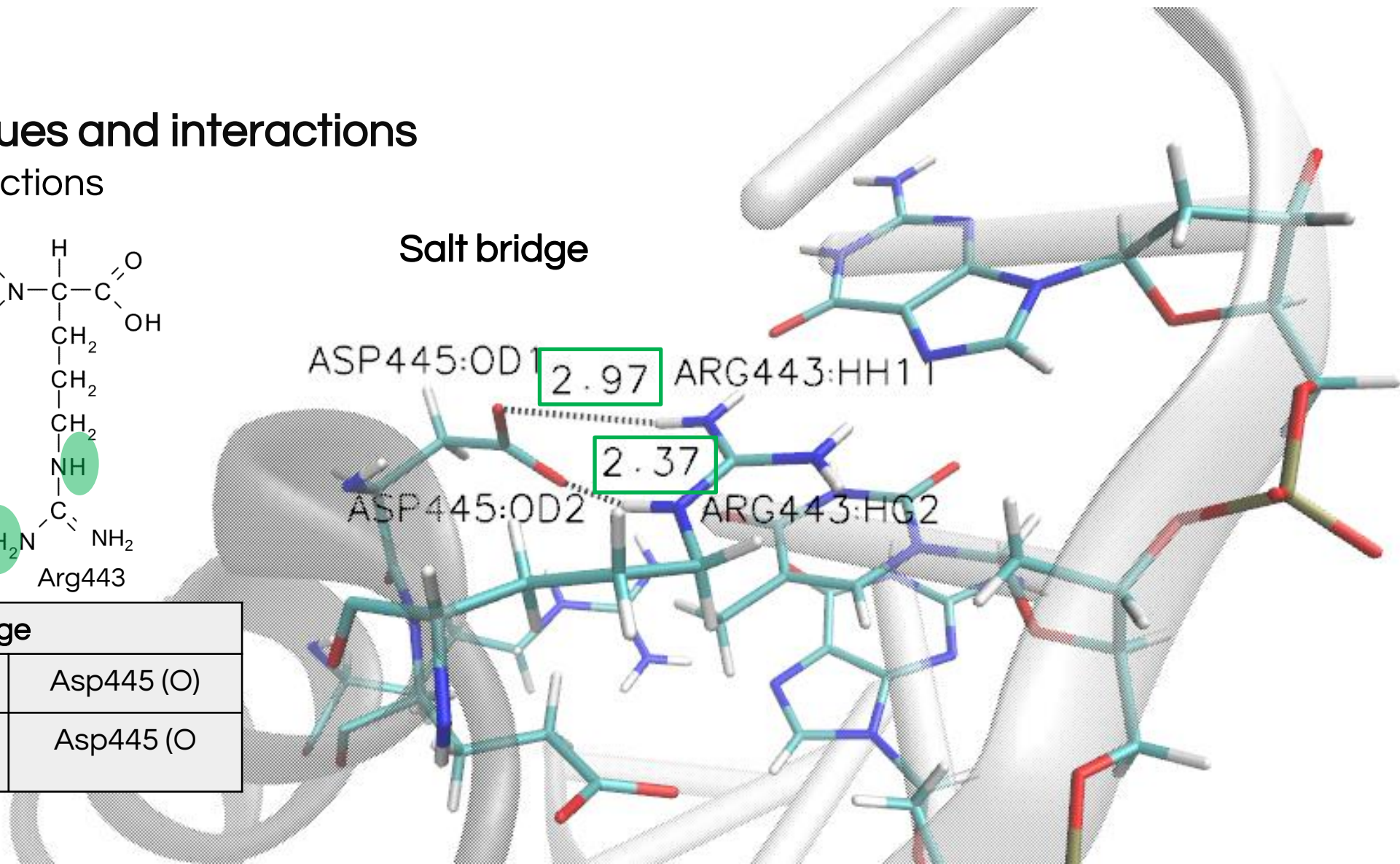
↻
m
G T G
C A C
m

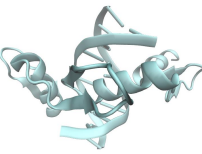


Salt bridge



Salt bridge		
Arg443 (-NH)	2,97 Å	Asp445 (O)
Arg443 (-NH)	2,37 Å	Asp445 (O)

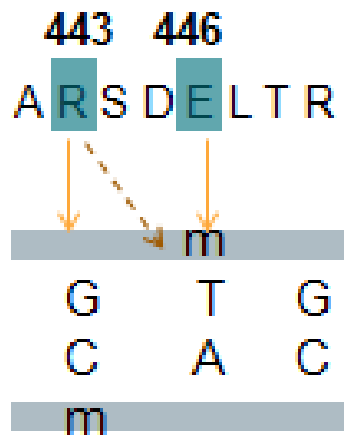




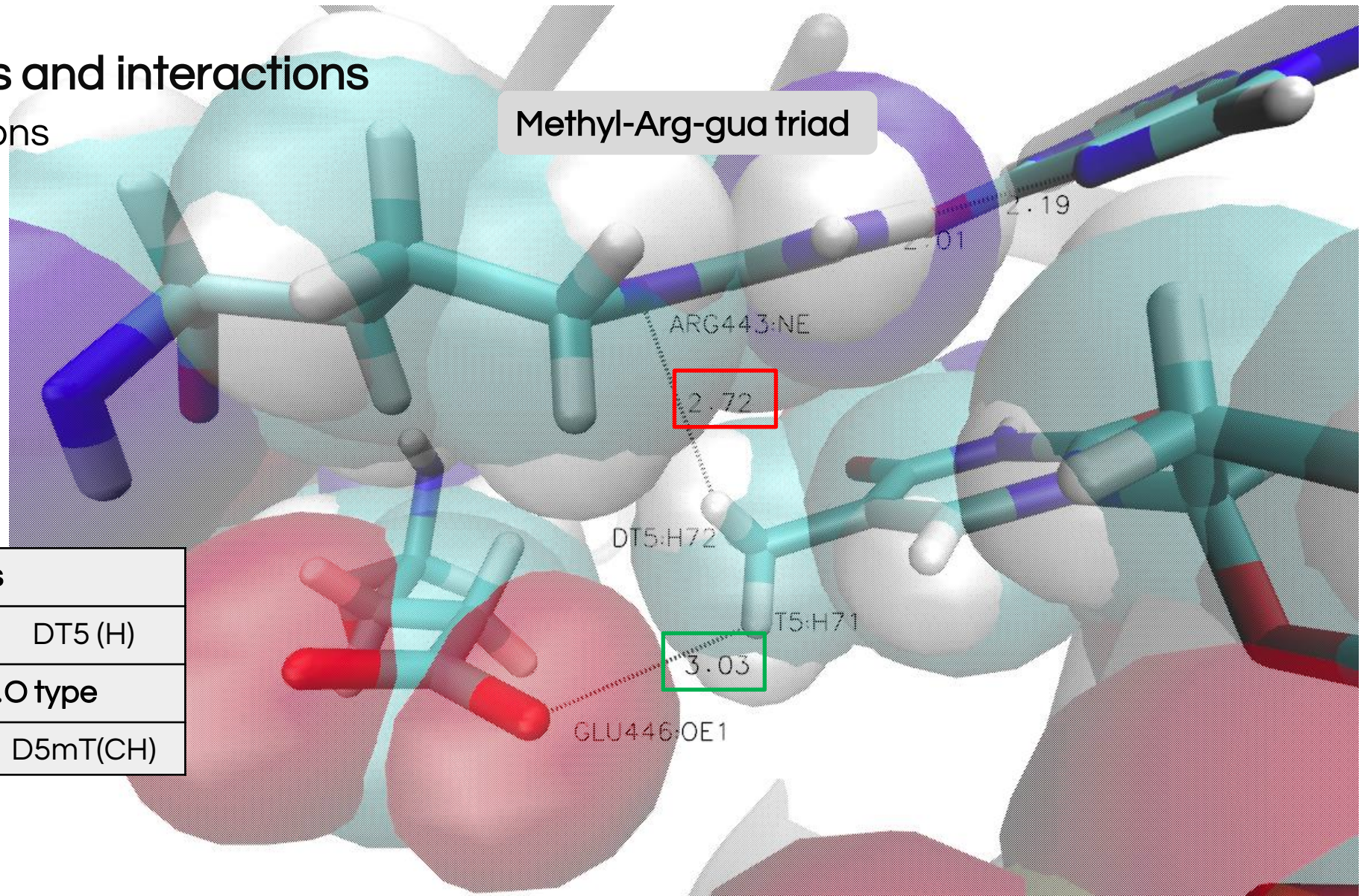
KLF4

Important residues and interactions

DNA-protein interactions



Van der Waals		
Arg443 (-N)	2,72 Å	DT5 (H)
Hydrogen bone CH...O type		
Glu446(O)	3,03 Å	D5mT(CH)



KLF4

Important residues and interactions

DNA-protein interactions

Cluster: 1 (2wbu & 4m9e)

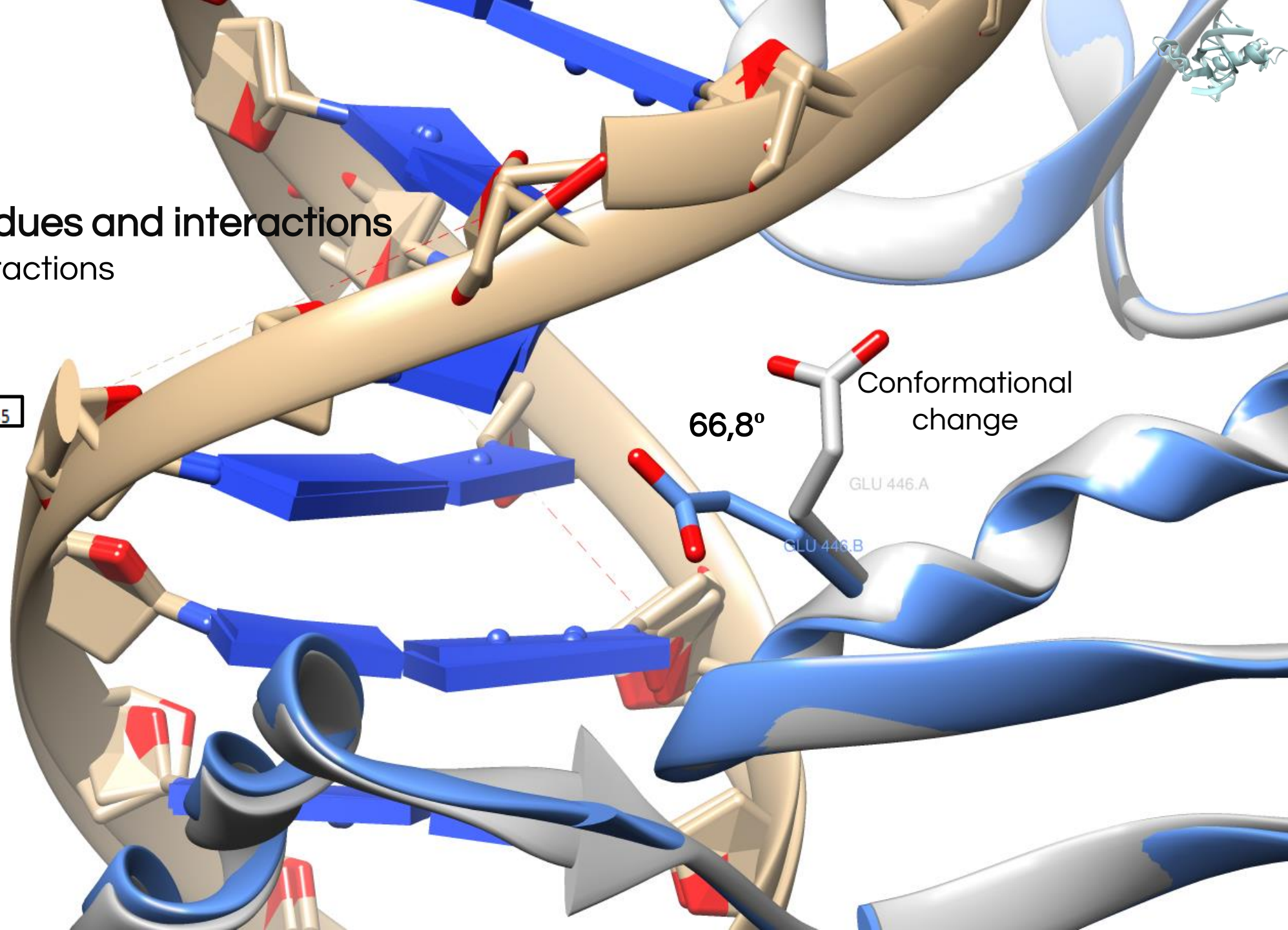
Sc	9.62	RMS	0.45	Len	85	nfit	85
----	------	-----	------	-----	----	------	----



Methylated DNA



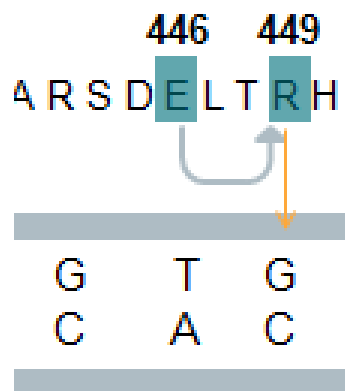
Unmethylated DNA



KLF4

Important residues and interactions

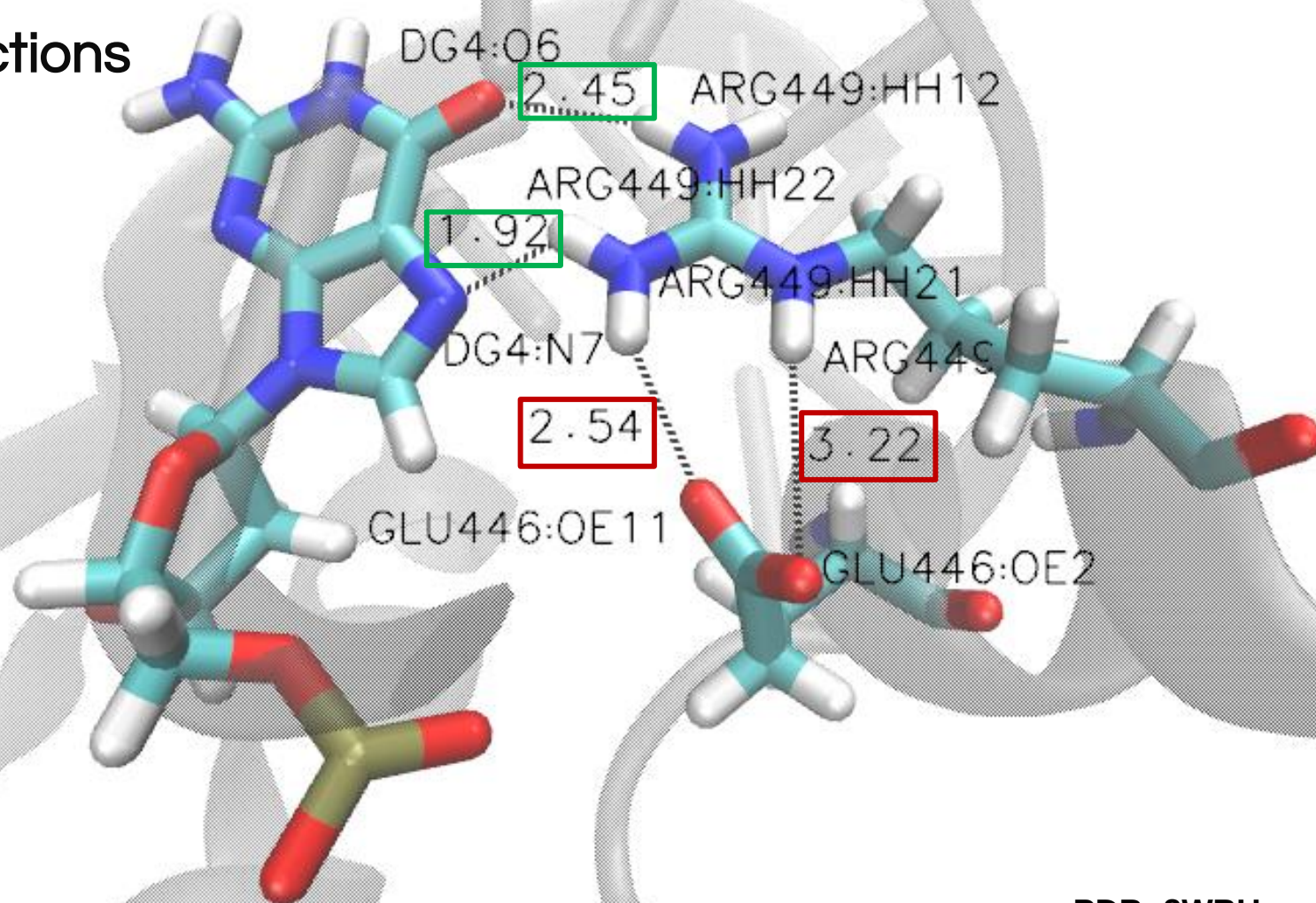
DNA-protein interactions



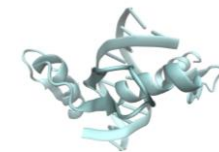
Unmethylated DNA

Hydrogen bonds

Glu446 (O)	2,56 Å	Arg449 (H)
Glu446 (O)	3,22 Å	Arg4439(H)
Arg449 (H)	1,92 Å	DG4(N)
Arg449(H)	2,45 Å	DG4(O)



PDB: 2WBU

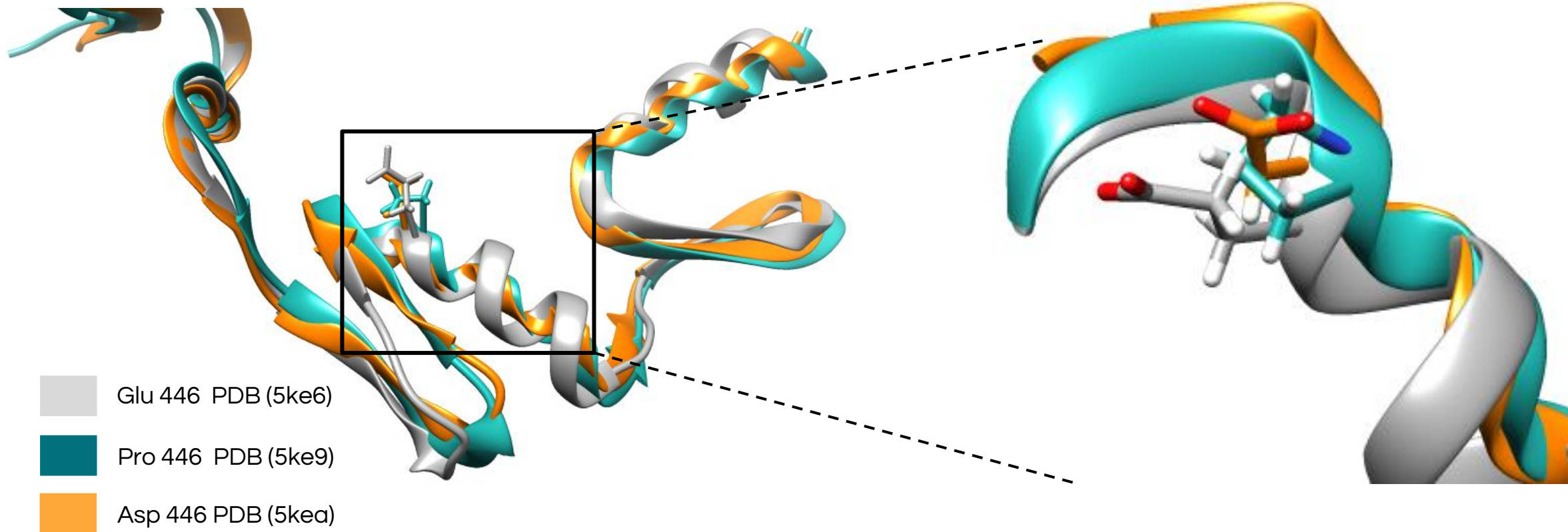


KLF4

Important residues and interactions

DNA-protein interactions

Glu446 (E446) 5ke6
Pro446 (E446P) 5ke9
Asp446 (E446D) 5kea



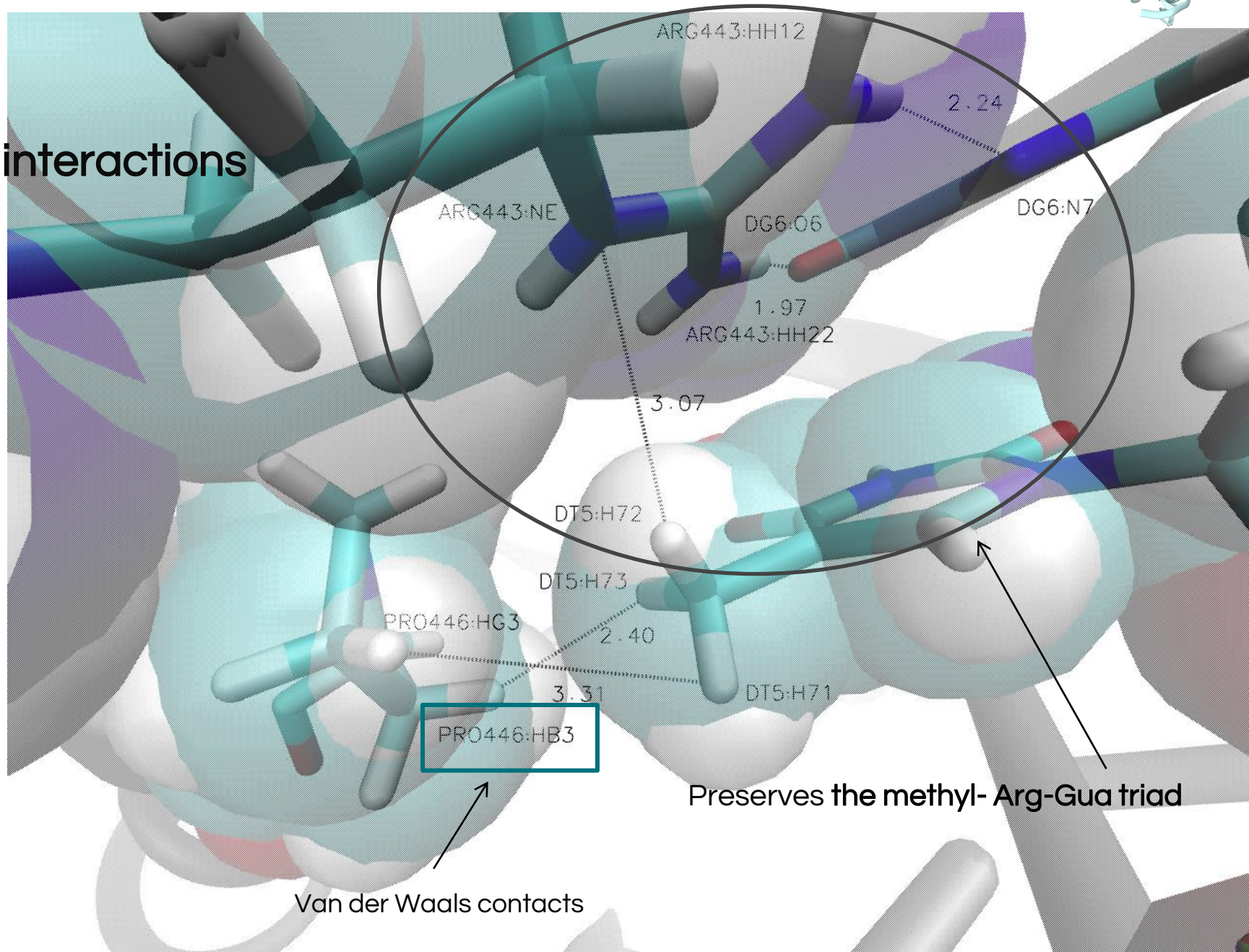
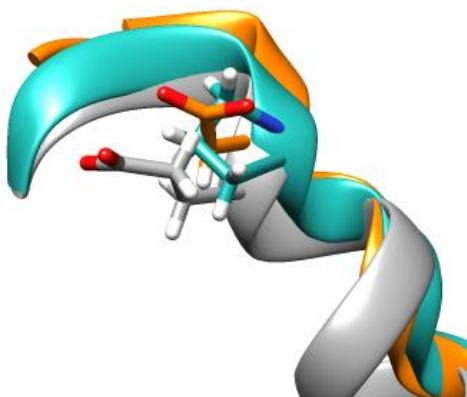
Cluster: 2 (5ke6 & 5ke9 5kea) Sc 9.41 RMS 1.49 Len 85 nfit 85

KLF4

Important residues and interactions

DNA-protein interactions

Pro 446 PDB (5ke9)

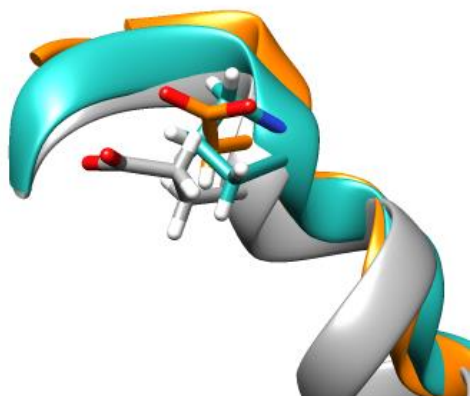


KLF4

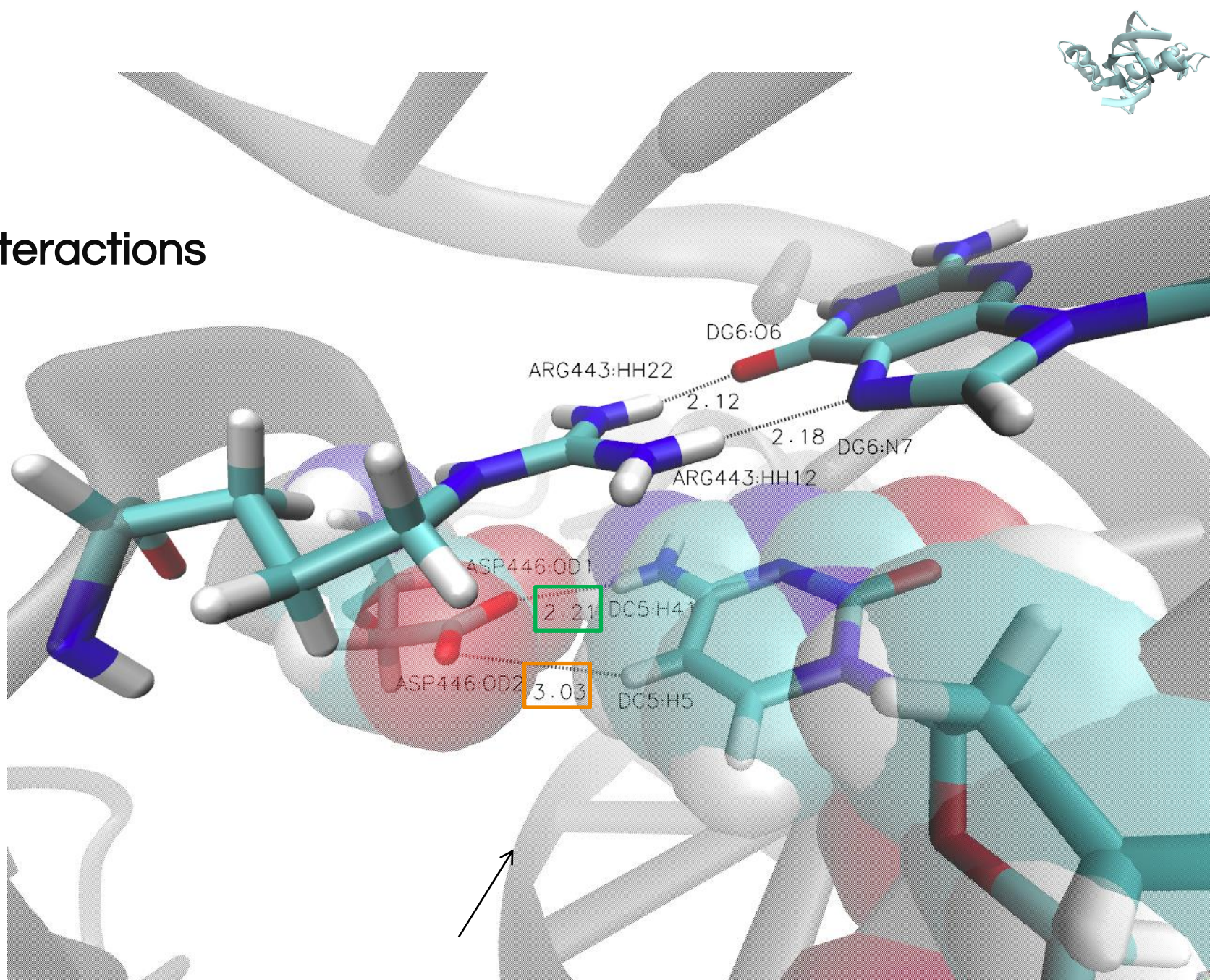
Important residues and interactions

DNA-protein interactions

 Asp 446 PDB (5kea)



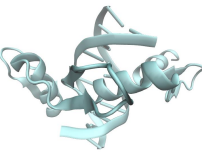
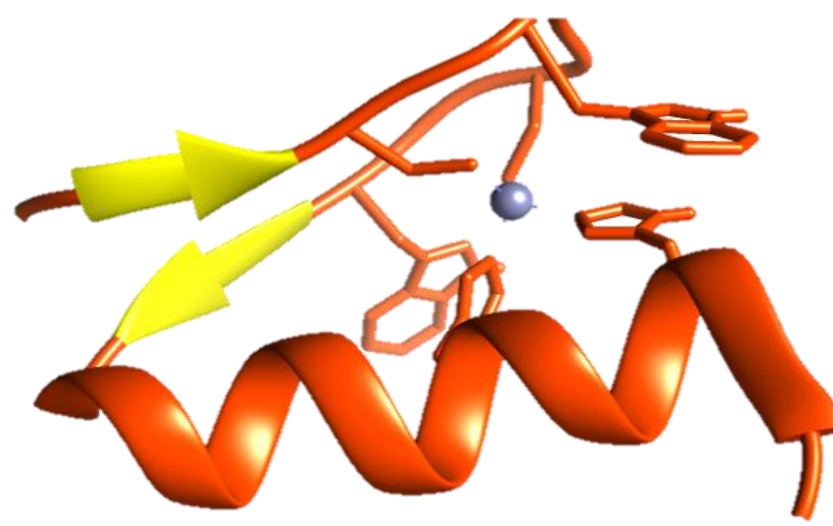
Hydrogen bonds (CH..O type)		
Asp446 (-O)	2,21 Å	DC5(-NH)
Asp446(-O)	3,03 Å	DC5(-CH)



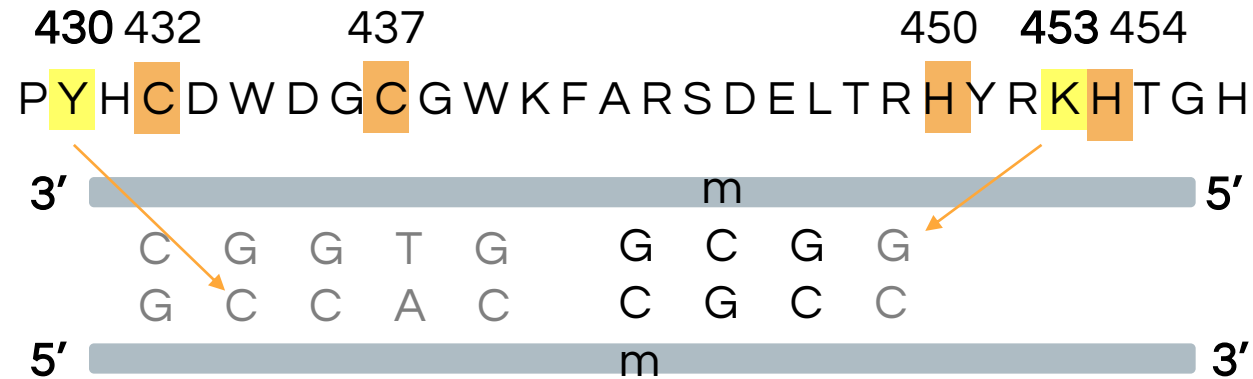
KLF4

Important residues and interactions

DNA-protein interactions



ZnFn 2



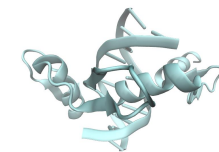
UNSPECIFIC INTERACTIONS:

Tyr 430 (COOH) – C2 (Phosphate backbone)

Hydrogen bond

Lys 453 (NH₃) – G3 (Phosphate backbone)

Hydrogen bond

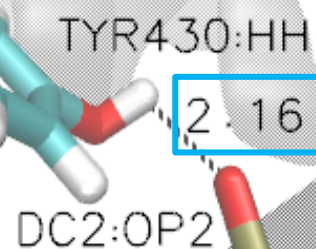
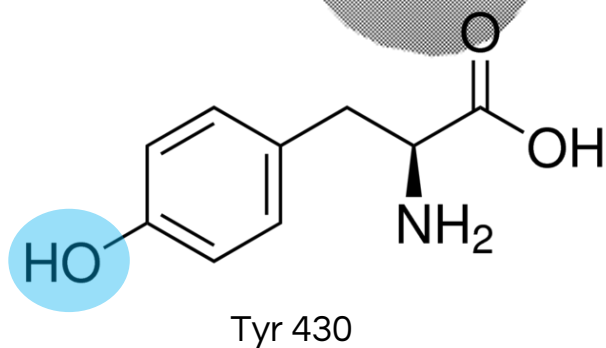


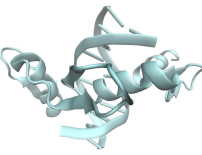
KLF4

Important residues and interactions

DNA-protein interactions

Hydrogen bonds		
Tyr430 (-OH)	2,16 Å	DC2(-OP)



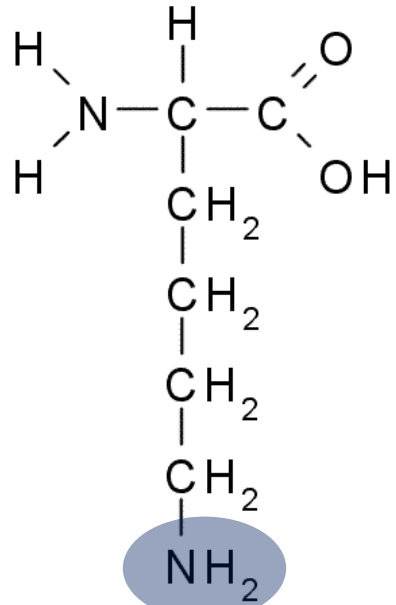


KLF4

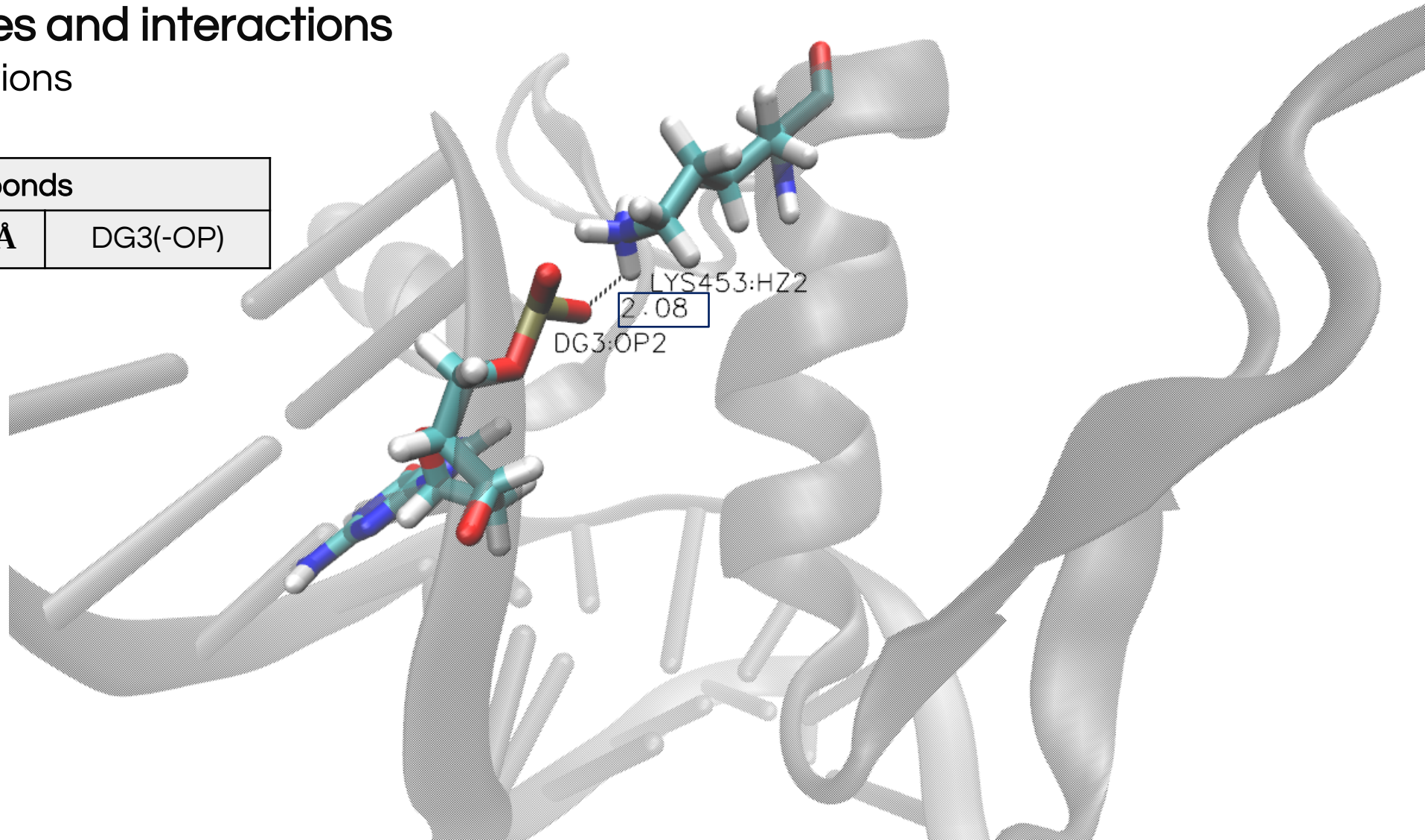
Important residues and interactions

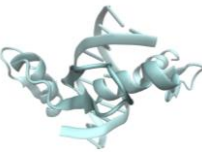
DNA-protein interactions

Hydrogen bonds		
Lys 453 (-NH ₃)	2,08 Å	DG3(-OP)



Lys 453





KLF4

Important residues and interactions

DNA-protein interactions

All conserved

ZnFn 2

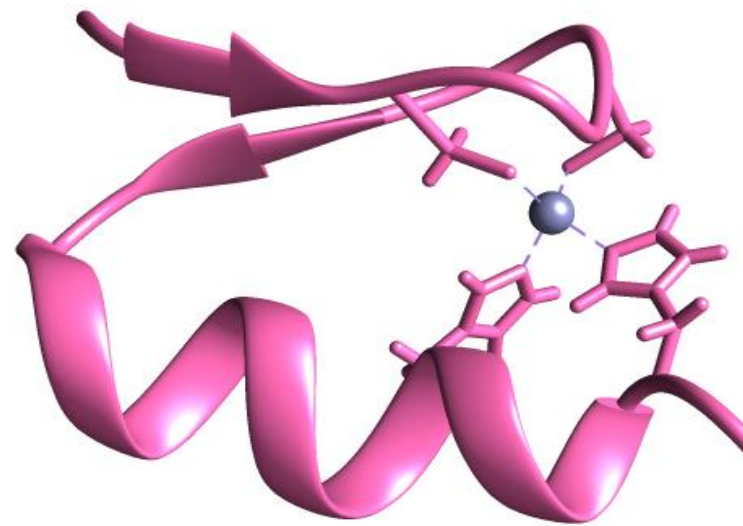
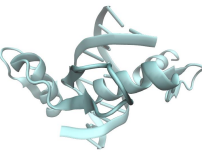
sp | Q60793 | KLF4_MOUSE
sp | 043474 | KLF4_HUMAN
tr | F6VVB3 | KLF4_MACMU
tr | A7YWE2 | KLF4_BOVIN
tr | H2QXN2 | KLF4_PANTR
tr | D3U661 | KLF4_FELCA
tr | Q90XE8 | KLF4_DANRE
tr | Q28D29 | KLF4_XENTR
tr | Q52JJ4 | KLF4_PIG
tr | G9JZK4 | KLF4_BUBBU
tr | A0A1A8AVD8 | KLF4_NOTFU
tr | A0A0B4UDA9 | KLF4_RABIT
tr | G1ERW8 | KLF4_XENLA
tr | C0KKX2 | KLF4_NOTVI
tr | M4T5C8 | KLF4_MICLV
tr | A0A077JN77 | KLF4_CYNPY
tr | A0A1A8HS30 | KLF4_NOTKU
tr | A0A142FHL6 | KLF4_9CICH
tr | A0A1A8PAT3 | KLF4_9TELE
tr | A0A166NG74 | KLF4_CAPHI

kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCSweGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCNweGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDweGCGWKFARSDDELTRHYRKH
kpYHCDwdGCGWKFARSDDELTRHYRKH

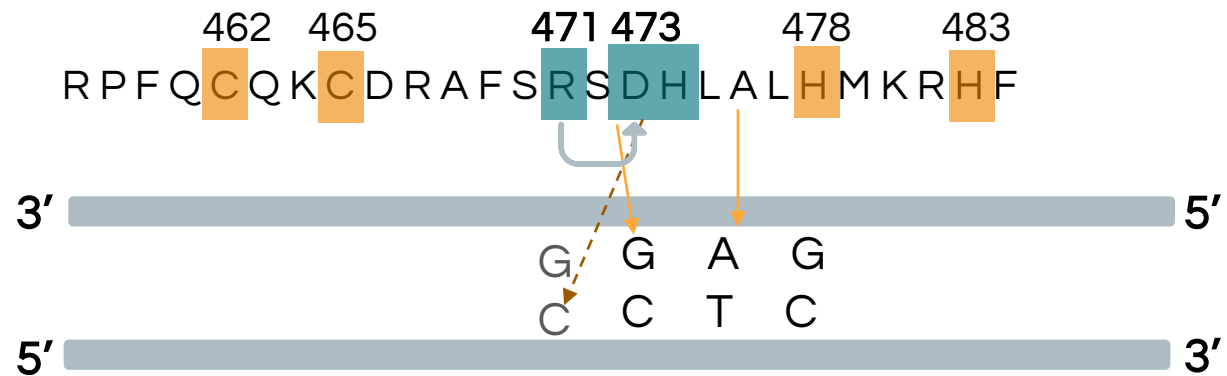
KLF4

Important residues and interactions

DNA-protein interactions



ZnFn 3



SPECIFIC INTERACTIONS:

Arg 471 (NH₂) – G3 (-CO and nitrogen)

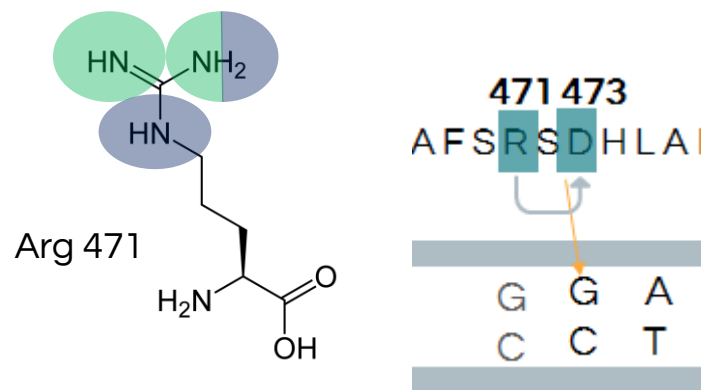
Arg 471 (NH₂) – Asp473 (-CO)

Asp 473 (-CO) – C7 (N)

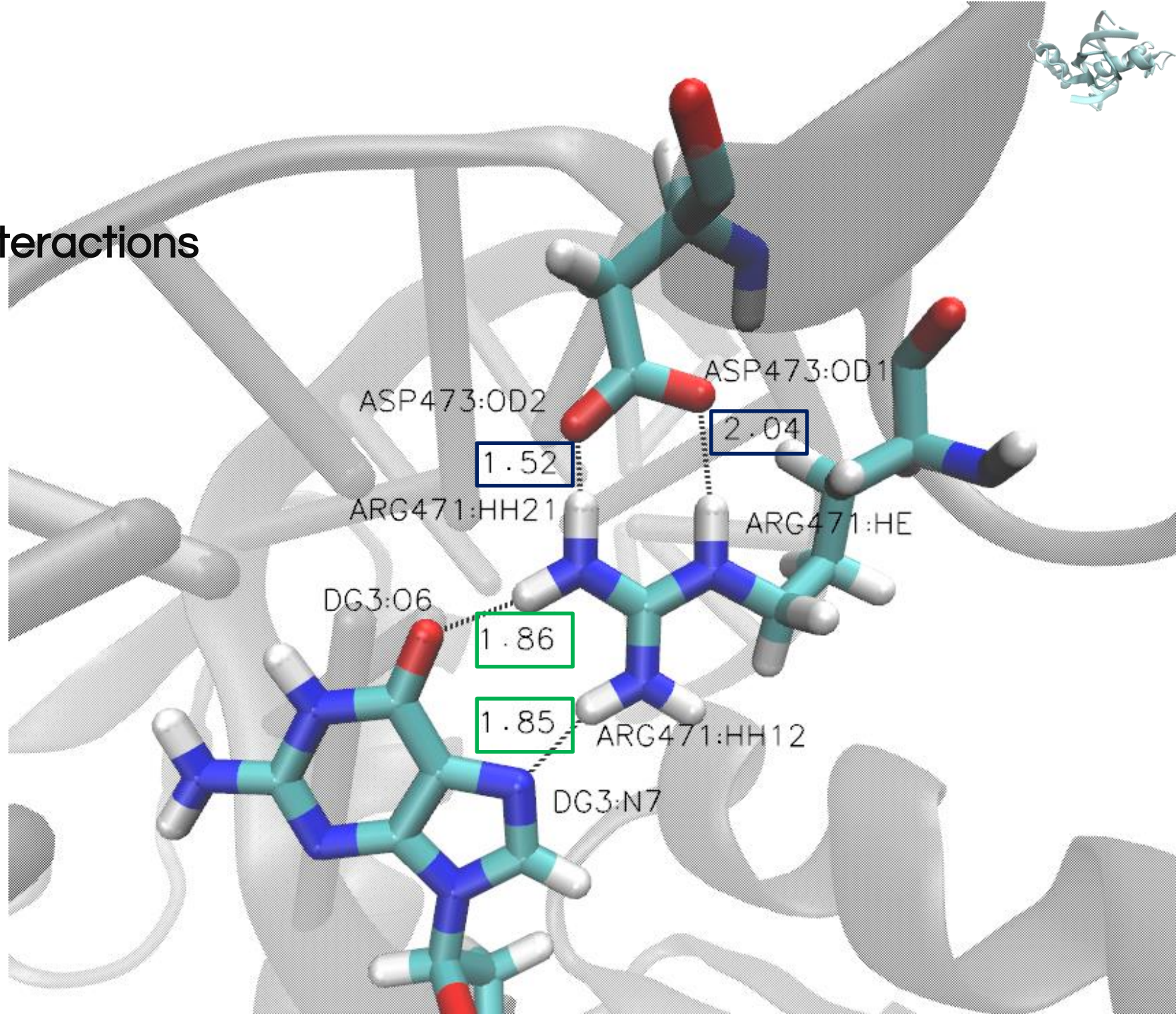
KLF4

Important residues and interactions

DNA-protein interactions



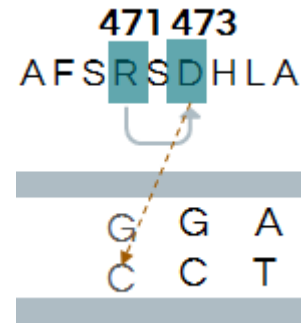
Hydrogen bonds		
Arg 471(-NH ₂)	1,86 Å	DG3(O)
Arg 471(-NH ₂)	1,85 Å	DG3(N)
Arg 471(-NH ₂)	1,52 Å	Asp473(O)
Arg 471(-NH ₂)	2,04 Å	Asp473(O)



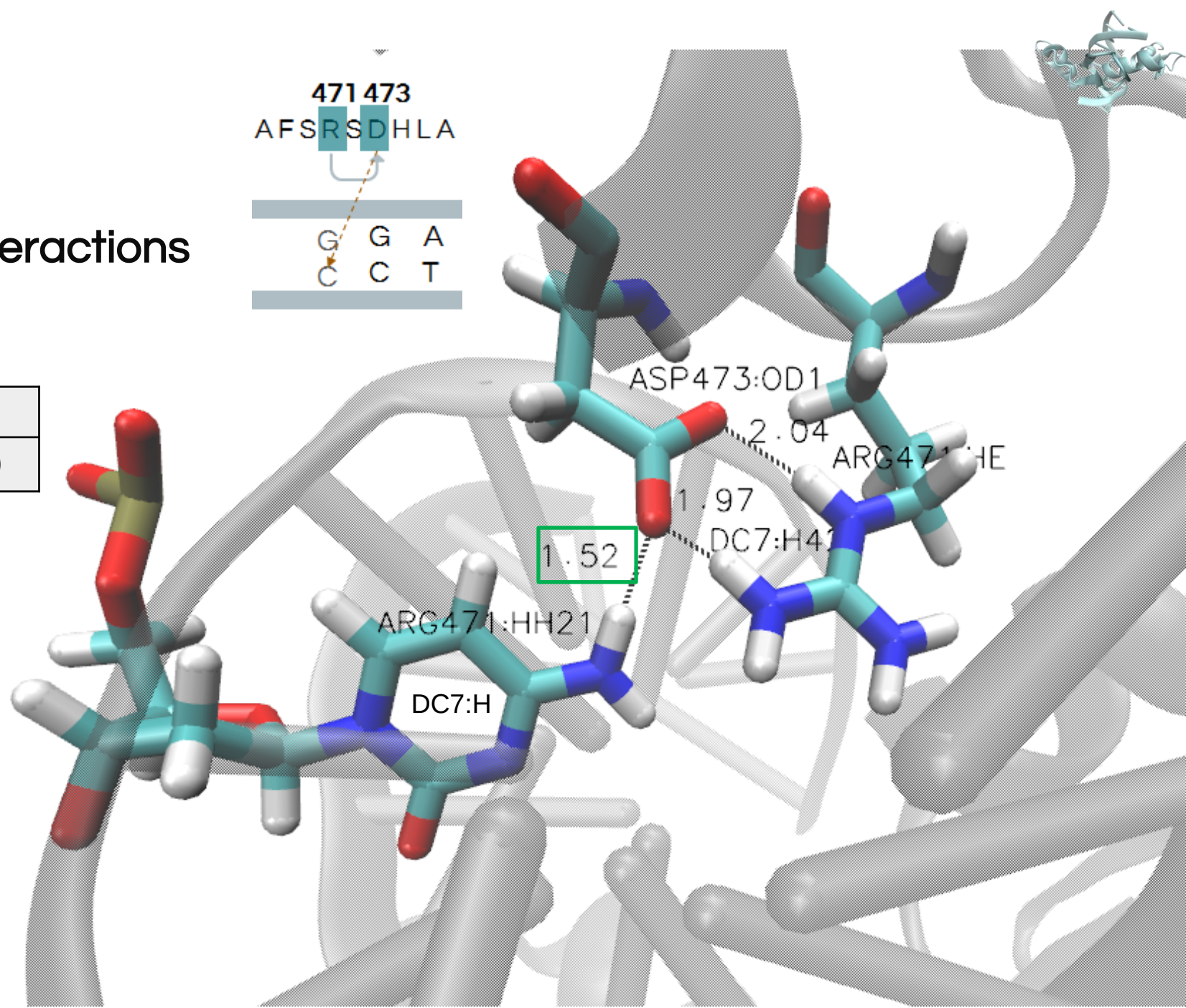
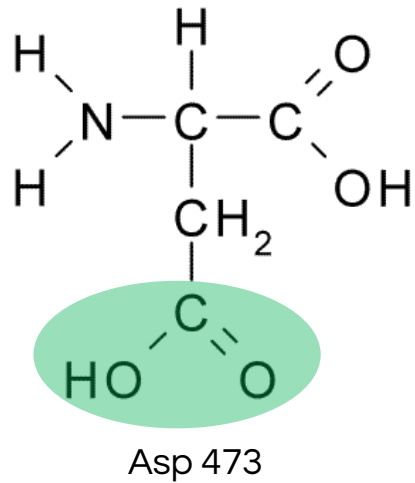
KLF4

Important residues and interactions

DNA-protein interactions



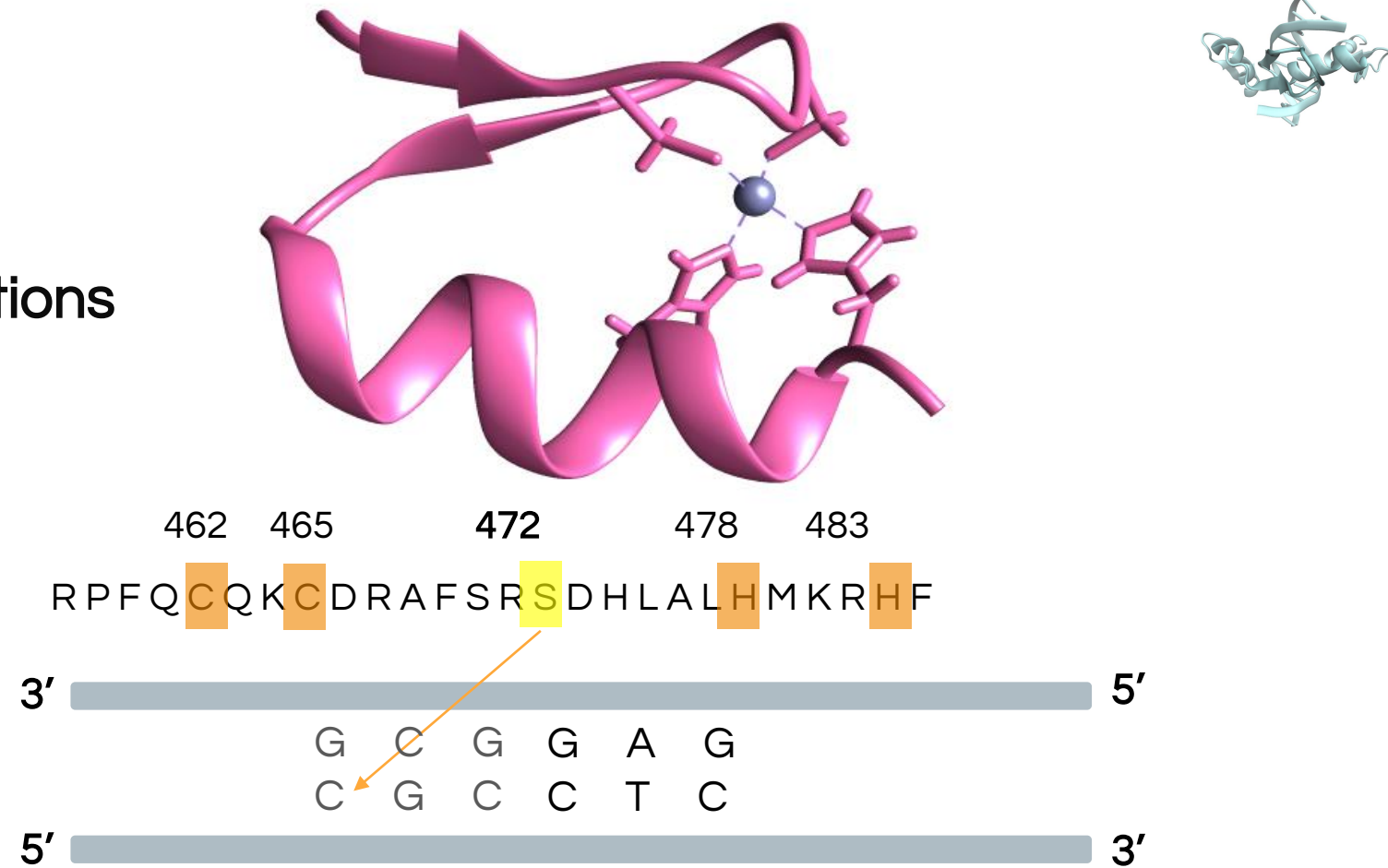
Hydrogen bond		
Asp 473(-COOH)	1,52 Å	DC7'(NH)



KLF4

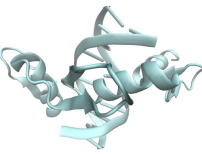
Important residues and interactions

DNA-protein interactions



UNSPECIFIC INTERACTIONS:

Ser472 (NH₂) – Phosphate backbone of the C5'

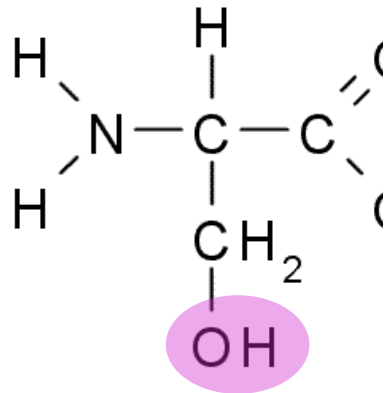


KLF4

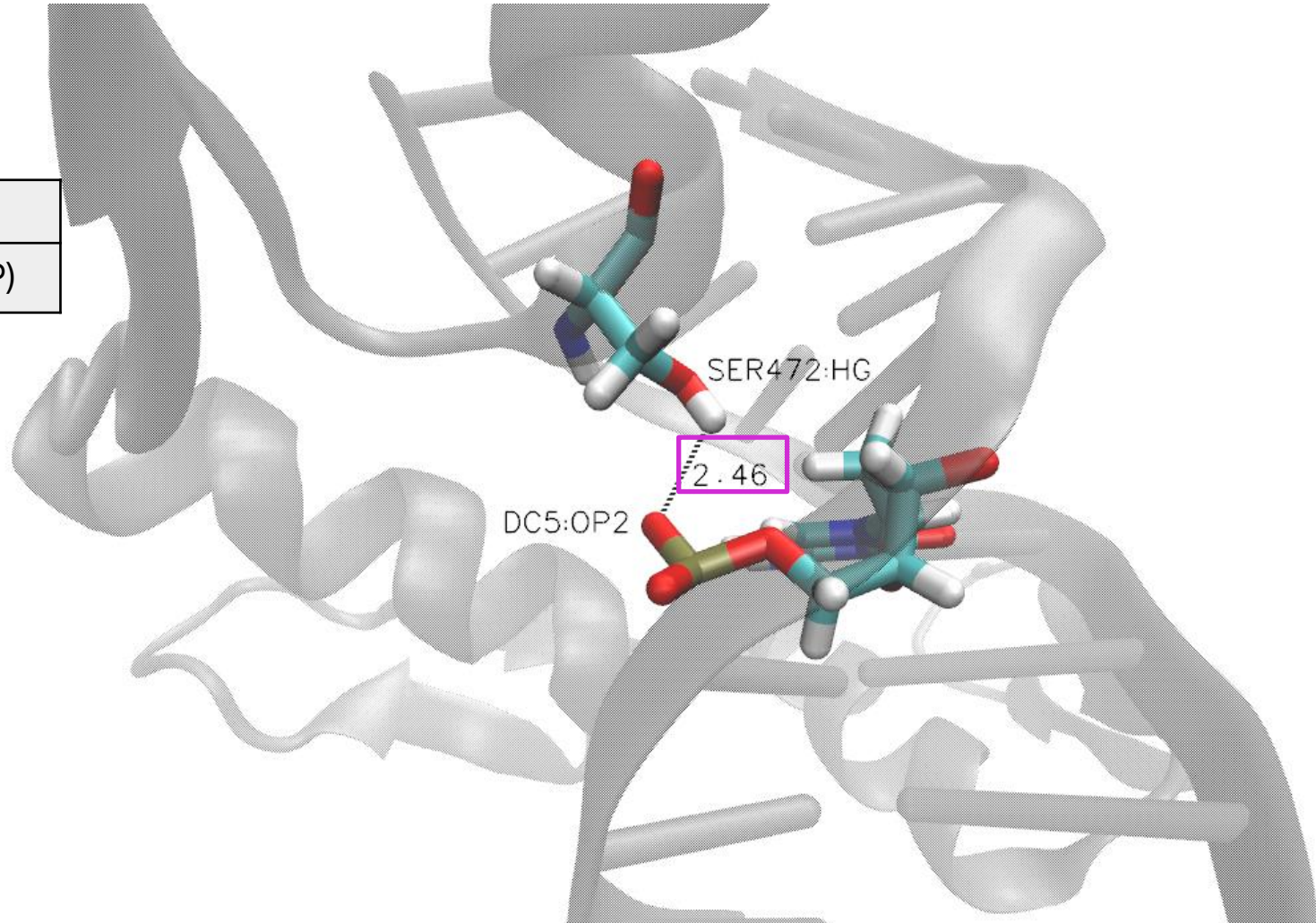
Important residues and interactions

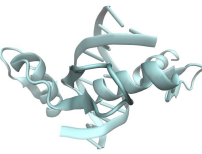
DNA-protein interactions

Hydrogen bond		
Ser 472 (-OH)	2,46 Å	DC5'(OP)



Ser 472





KLF4

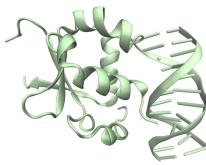
Important residues and interactions

DNA-protein interactions

All conserved

ZnFn 3

sp Q60793 KLF4_MOUSE	hrpFQCQKCDRAFSRSDHLALHMKRHf-
sp 043474 KLF4_HUMAN	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr F6VVB3 KLF4_MACMU	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A7YWE2 KLF4_BOVIN	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr H2QXN2 KLF4_PANTR	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr D3U661 KLF4_FELCA	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr Q90XE8 KLF4_DANRE	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr Q28D29 KLF4_XENTR	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr Q52JJ4 KLF4_PIG	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr G9JZK4 KLF4_BUBBU	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A1A8AVD8 KLF4_NOTFU	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A0B4UDA9 KLF4_RABIT	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr G1ERW8 KLF4_XENLA	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr C0KKX2 KLF4_NOTVI	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr M4T5C8 KLF4_MICLV	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A077JN77 KLF4_CYNPY	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A1A8HS30 KLF4_NOTKU	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A142FHL6 KLF4_9CICH	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A1A8PAT3 KLF4_9TELE	hrpFQCQKCDRAFSRSDHLALHMKRHf-
tr A0A166NG74 KLF4_CAPHI	hrpFQCQKCDRAFSRSDHLALHMKRHf-



HSF1

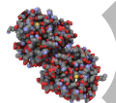
Index



Classification



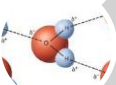
Biological function



General structure

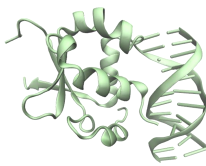


Secondary structure



Important residues and interactions





HSF1

Classification



Family: Heat-shock transcription factor

Lineage

Root

Scop

Class

All alpha proteins

Fold

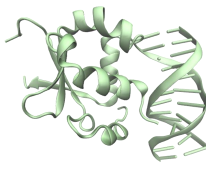
DNA/RNA-binding –helical bundle

Superfamily

Winged helix DNA-binding domain

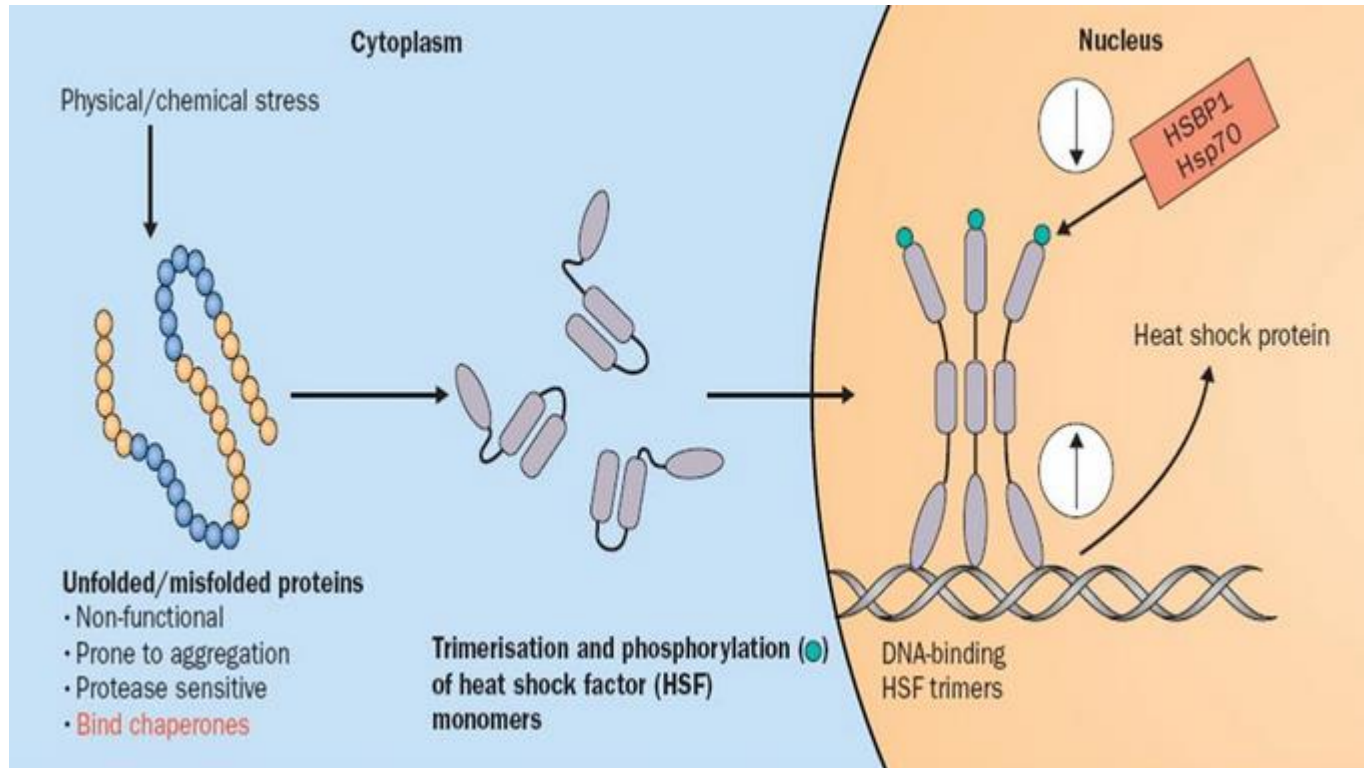
Family

Heat-shock transcription factor



HSF1

Biological function



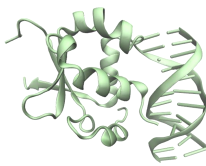
Central role in **protein homeostasis**



Heat Shock Response (HSR)

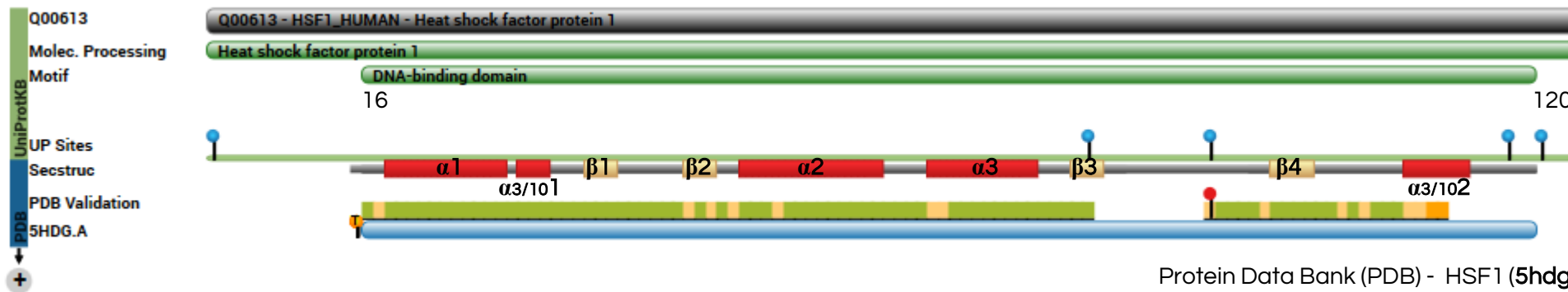
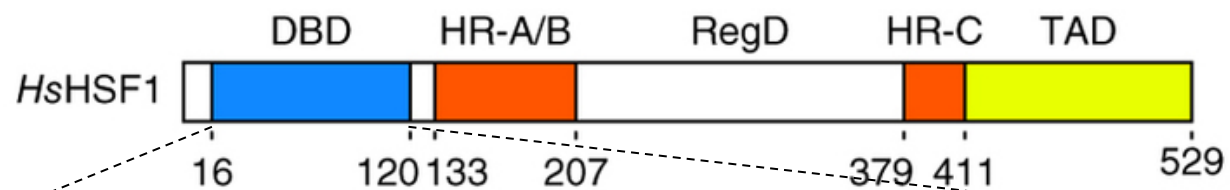
Heat shock factors (HSF)
Heat shock element (HSE)
Heat shock proteins (HSPs)

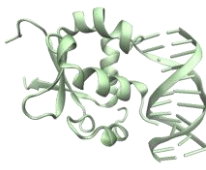
- 50 – 90 kDa
- Conserved N-terminal winged HTH DBD.



HSF1

General structure





HSF1

Secondary structure



Helix-alpha:

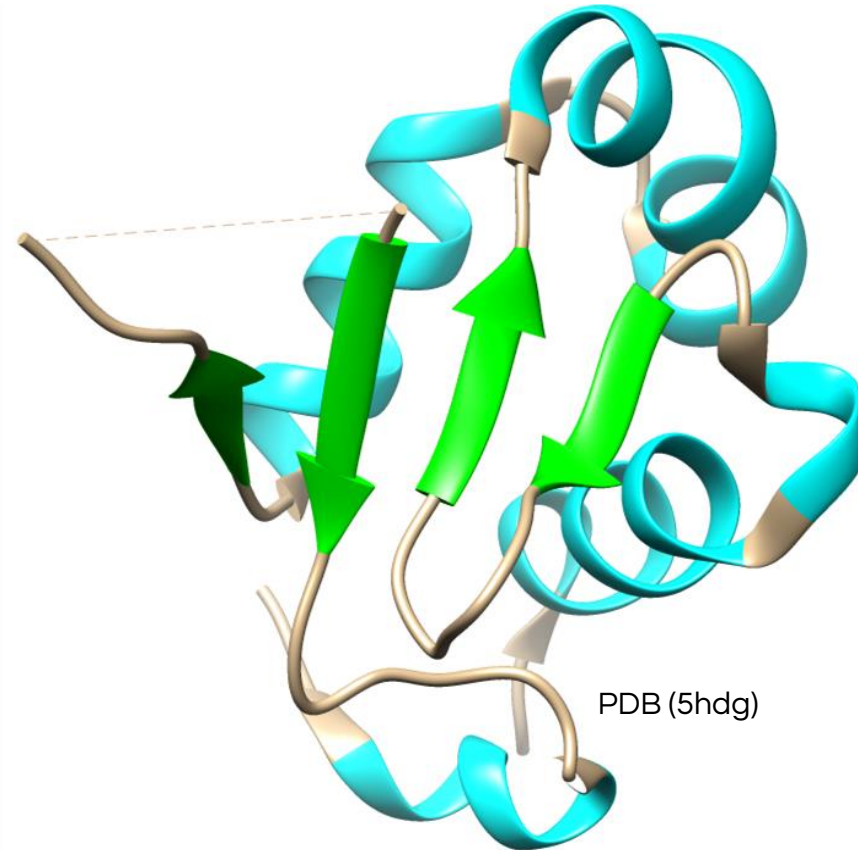
$\alpha 1$: (17-27)
 $\alpha 2$: (49-61)
 $\alpha 3$: (66-75)

Helix 3/10:

$\alpha 3/101$: (29-31)
 $\alpha 3/102$: (109-114)

Beta-strands:

$\beta 1$: (35-37)
 $\beta 2$: (44-46)
 $\beta 3$: (79-81)
 $\beta 4$: (97-100)



PDB (5hdg)

HSF1

Secondary structure

Helix-alpha:

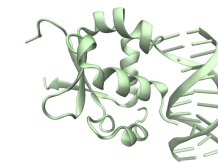
$\alpha 1$: (17-27)
 $\alpha 2$: (49-61)
 $\alpha 3$: (66-75)

Helix 3/10:

$\alpha_{3/10} 1$: (29-31)
 $\alpha_{3/10} 2$: (109-114)

Beta-strands:

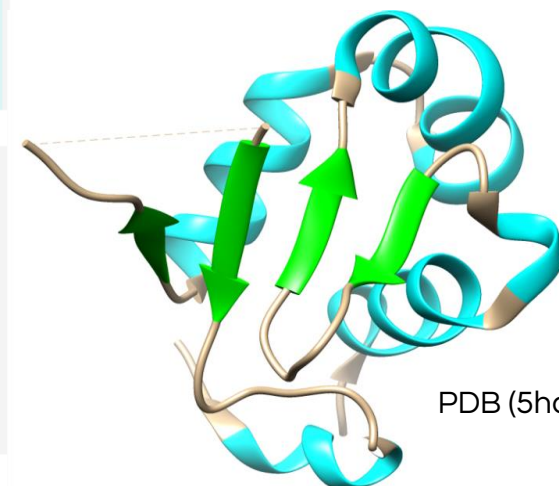
$\beta 1$: (35-37)
 $\beta 2$: (44-46)
 $\beta 3$: (79-81)
 $\beta 4$: (97-100)



sp Q00613 HSF1_HSAPIENS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGQFAKEVLPKYF
sp P38532 HSF1_MMUSCULUS	dlavgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGQFAKEVLPKYF
sp Q08DJ8 HSF1_BTaurus	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVLDQGQFAKEVLPKYF
sp P38529 HSF1_GGALLUS	aaaavgagpggsnvsAFLTKLWTLVEDPETDPLICWSPSGNSFHVFDQGQFAKEVLPKYF
sp P41154 HSF_XLAEVIS	--mdphgtcggsnvpAFLAKLWTLVEDPDTDPLICWSPEGNSFHVFDQGQFAKEVLPKYF
sp P22813 HSF_DMELANOASTER	mhsygdaaaigsgvpAFLAKLWRLVDDADTNRLICWTKDGQSFVIQNQAQFAKELLPLNY
sp P22121 HSF_KLACTIS	khhskklsttrarpAFVNKLWSMVNDKSNEKFIHWSTSGESIVVPRERFVQEVLPKYF
sp Q02953 HSF_SPOMBE	vgmsgsgssqrkitQFSNKLNMVNDSSSTDLSIRWSDRGDSFLVIGHEDFAKLVLPRYF
sp P10961 HSF_SCEREVISIAE	prqttrryqshksrpAFVNKLWSMLNDDSNTKLIQWAEDGKSFIVTNREEFVHQILPKYF

sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp Q08DJ8 HSF1_BTaurus	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLENIKRK
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK
sp P22813 HSF_DMELANOASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLDQIKRK
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ

sp Q00613 HSF1_HSAPIENS	VT-svstlksedikirqdsvtklldtdvqlmkgkqecmdskllamkhenealwrevaslrq
sp P38532 HSF1_MMUSCULUS	VT-svstlksedikirqdsvtklldtdvqlmkgkqecmdskllamkhenealwrevaslrq
sp Q08DJ8 HSF1_BTaurus	VT-svstlrsedikirqdsvtklldtdvqlmkgkqesmdskllamkhenealwrevaslrq
sp P38529 HSF1_GGALLUS	VT-svssiknedikvrqdnvtklldtdiqvmkgkqesmdskliamkhenealwrevaslrq
sp P41154 HSF_XLAEVIS	VN-tmsatksdevkvrdsvgklisdvqsmkgkqesidgrllsmkhenealwrevaslrq
sp P22813 HSF_DMELANOASTER	ISNnkgddkgvlkpeamskildvkvmrgrqdnldsrfsamqenevlwreiaslrqkh
sp P22121 HSF_KLACTIS	KSNtnilggttnaevdihillneletvkynqlaiaedlkritkdneumlwenmmarerhq
sp Q02953 HSF_SPOMBE	KA-gsqpveesntslmstisselqniriqqmnlslsriqvdaalwqenmenrerqr
sp P10961 HSF_SCEREVISIAE	KG-ssnnhnspsgngnpangsnipldnaagsnnsnnnissnsffnngllqgktrlrln



PDB (5hdg)



HSF1

Important residues and interactions

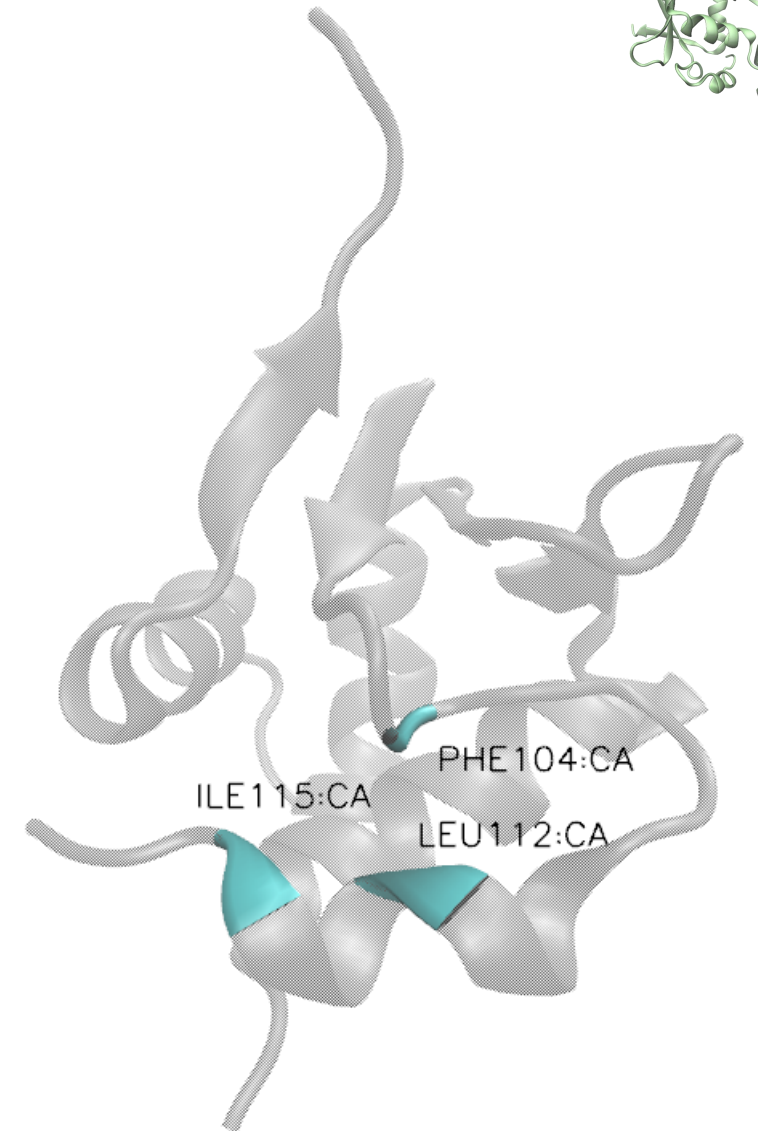
C terminal stabilization

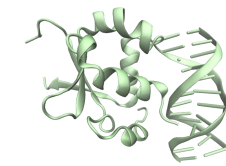
		104	112	115
sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp Q08DJ8 HSF1_BTAURUS	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLENIKRK			
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK			
sp P22813 HSF_DMELANOGASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLDQIKRK			
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ			
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK			
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ			

Phe 104
 Lys 112
 Ile 115

} Hydrophobic interactions with the **core** of the domain

Val 115 in *S. pombe* → Both aliphatic





HSF1

Important residues and interactions

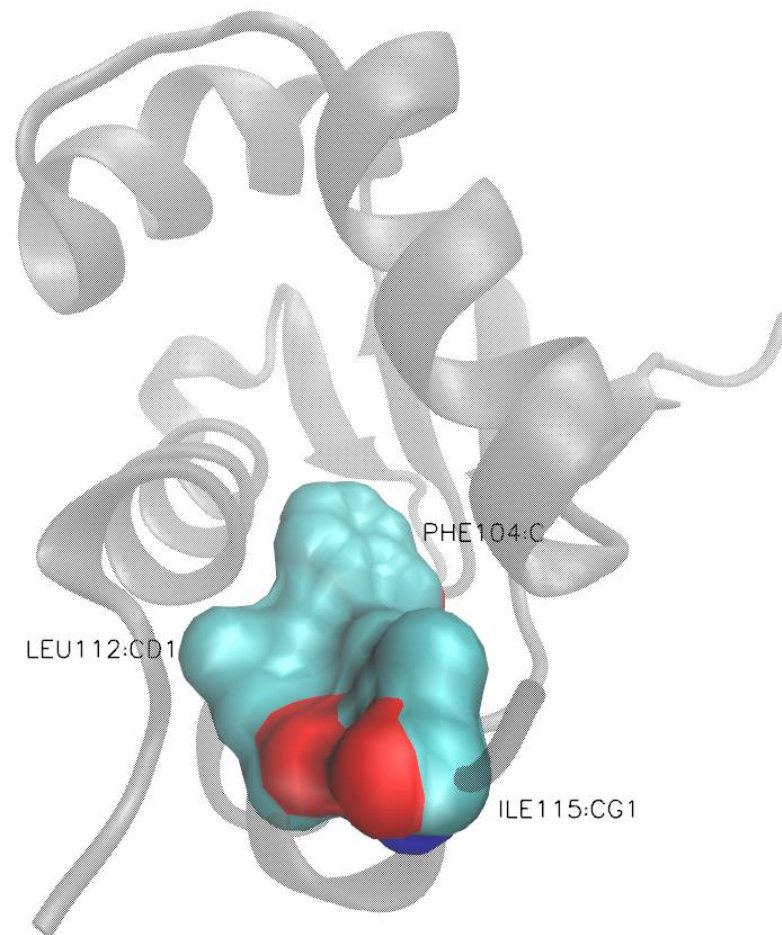
C terminal stabilization

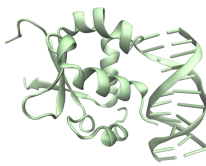
		104	112	115
sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp Q08DJ8 HSF1_BTAURUS	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK			
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLENIKRK			
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK			
sp P22813 HSF_DMELANOASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLDQIKRK			
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ			
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK			
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ			

Phe 104
 Lys 112
 Ile 115

} Hydrophobic interactions with the core of the domain

Val 115 in *S. pombe* → Both aliphatic





HSF1

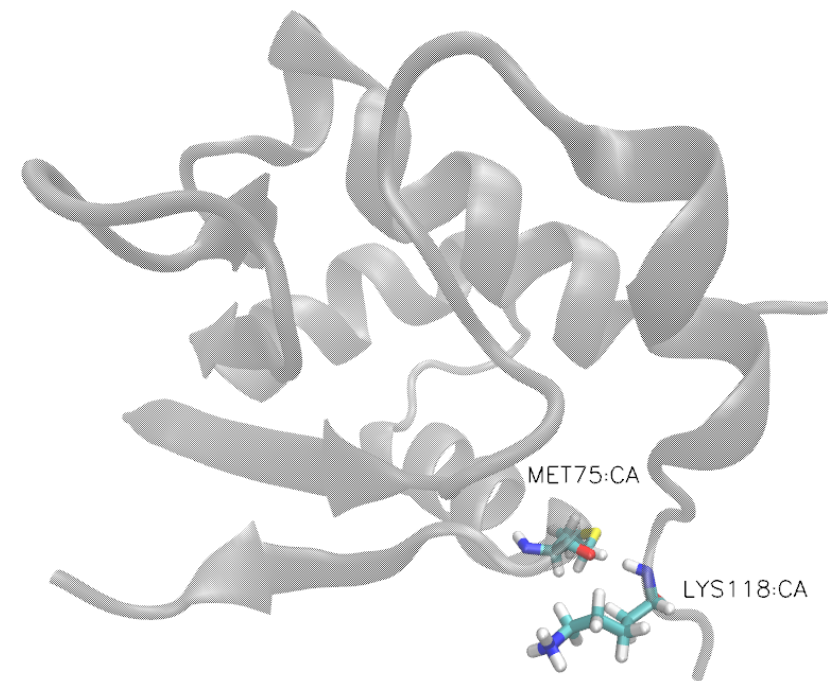
Important residues and interactions

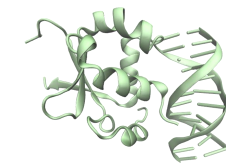
C terminal stabilization

		75	118
sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK		
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK		
sp Q08DJ8 HSF1_BTAURUS	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK		
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLENIKRK		
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK		
sp P22813 HSF_DMELANOASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLDQIKRK		
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ		
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK		
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ		

Lys 118 - Met 75 Hydrogen bond

→ Gln 118 in *K. lactis* → Both polar



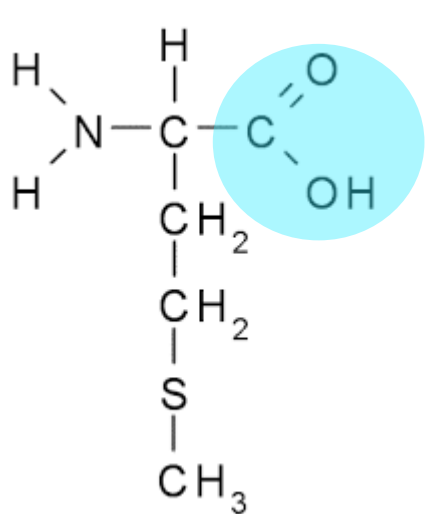


HSF1

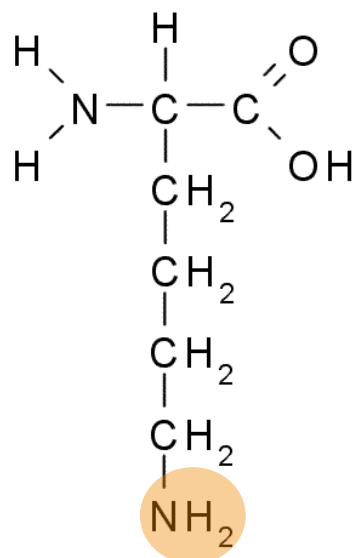
Important residues and interactions

C terminal stabilization

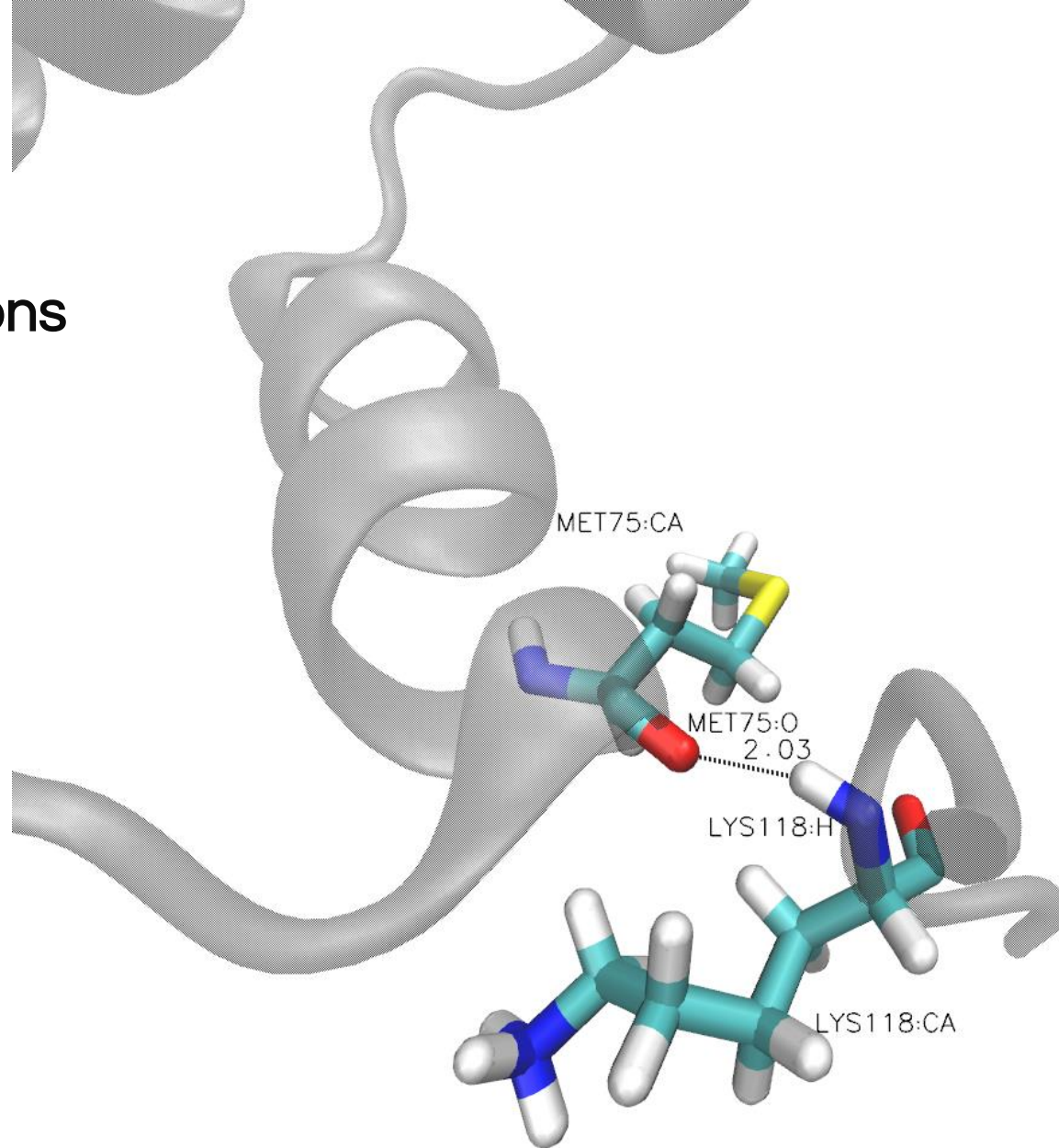
Hydrogen bonds		
Met 75 (-COOH)	2,03 Å	Lys 118 (-NH ₂)

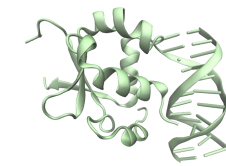


Met 75



Lys 118



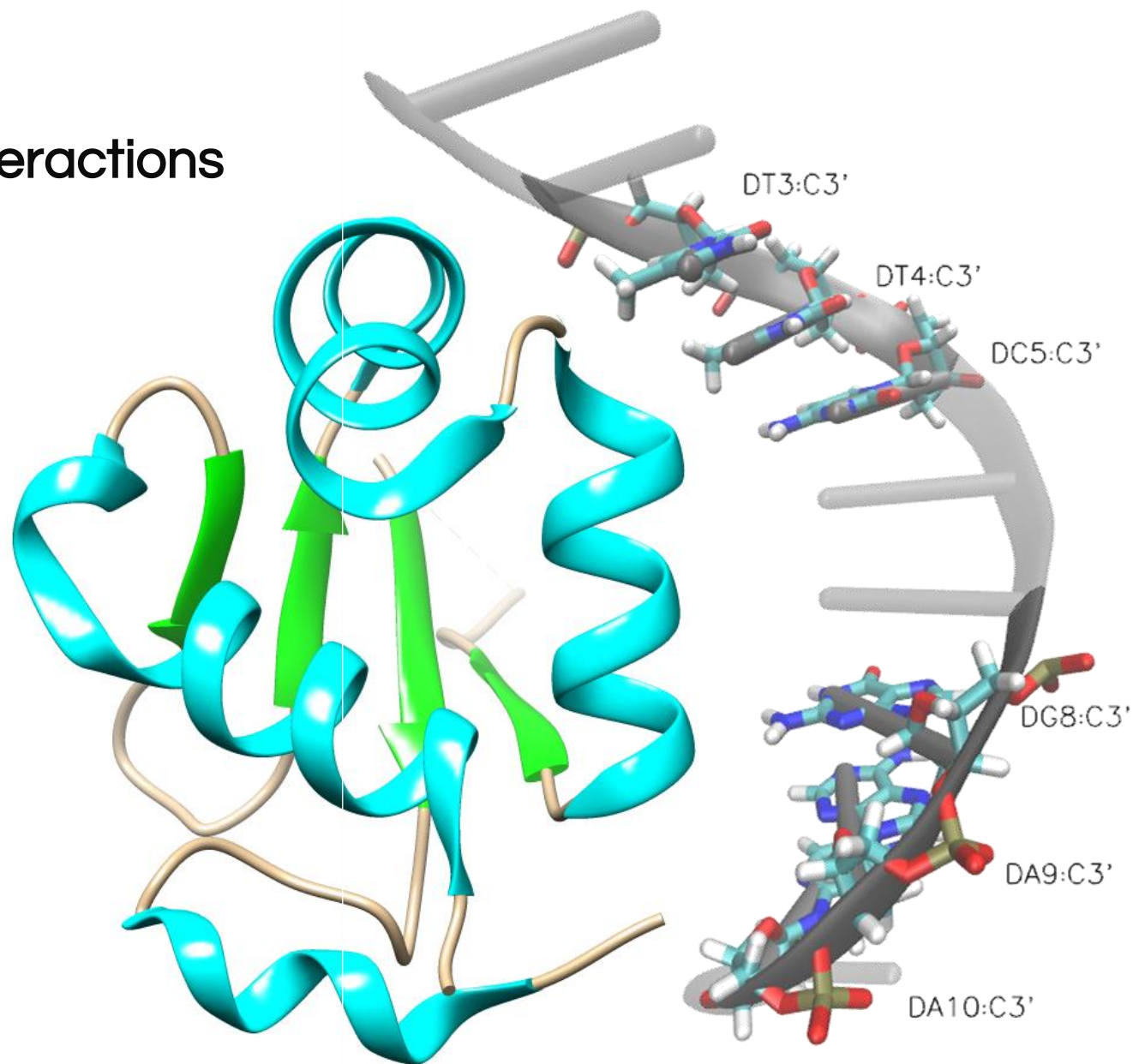


HSF1

Important residues and interactions

DNA-protein interactions

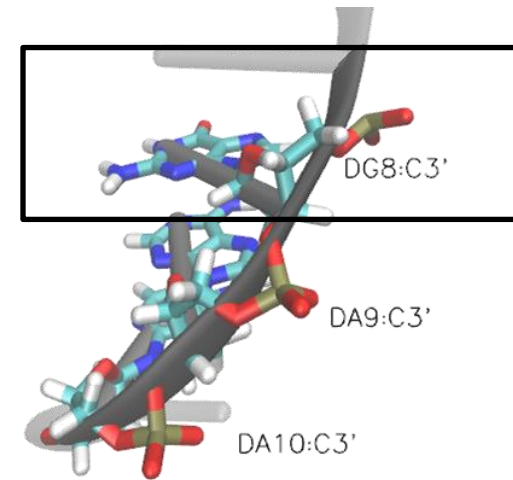
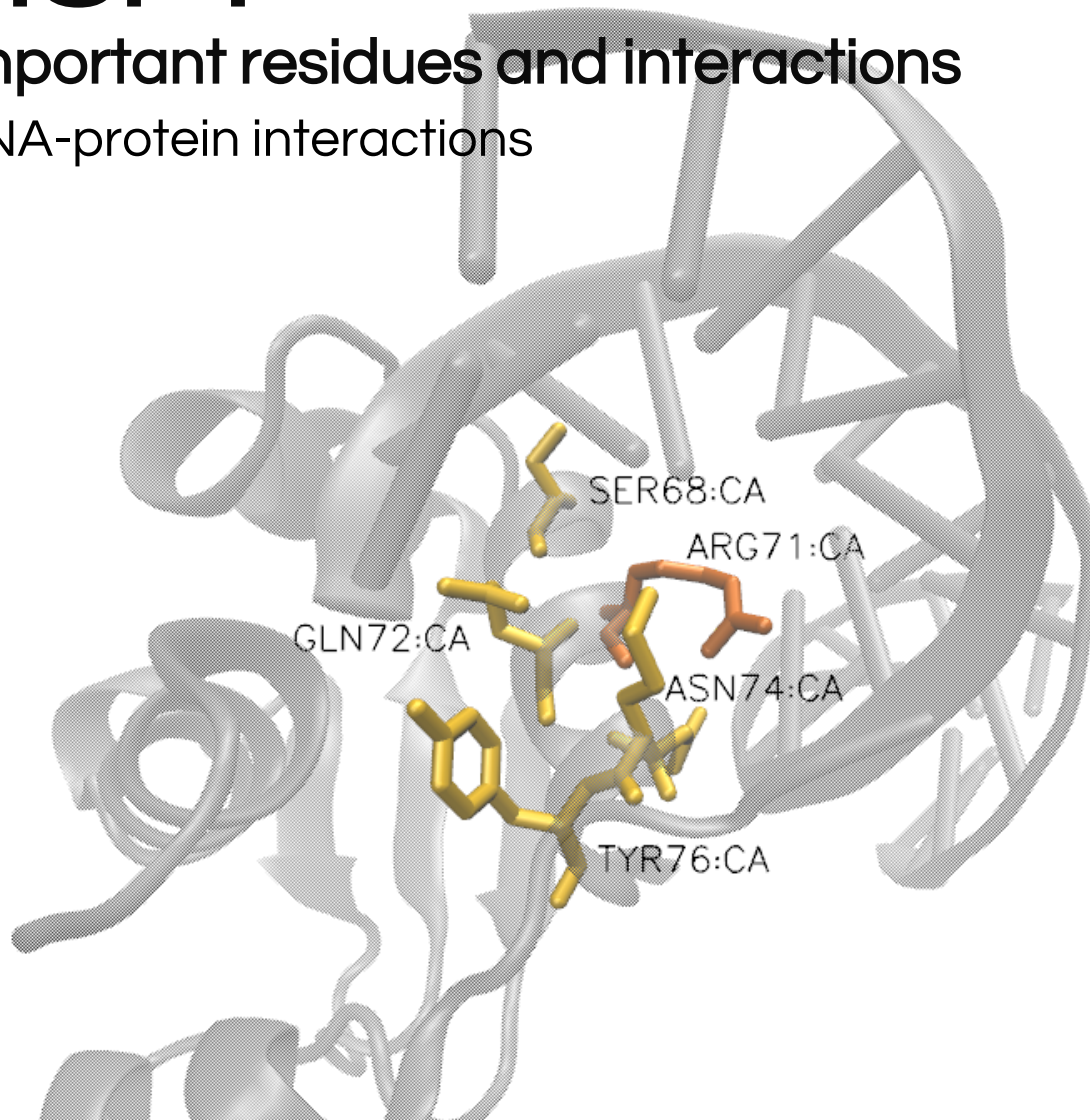
- Mainly in helix $\alpha 3$
- Heat shock element:
nTTCnGAAn



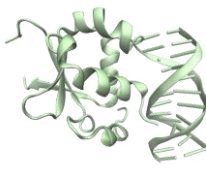
HSF1

Important residues and interactions

DNA-protein interactions



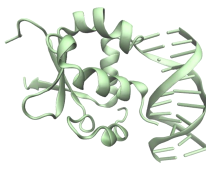
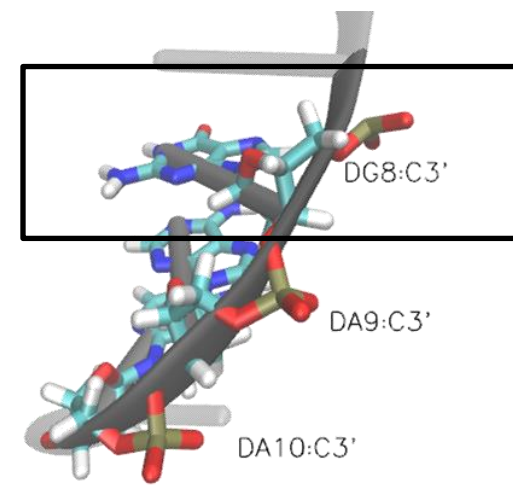
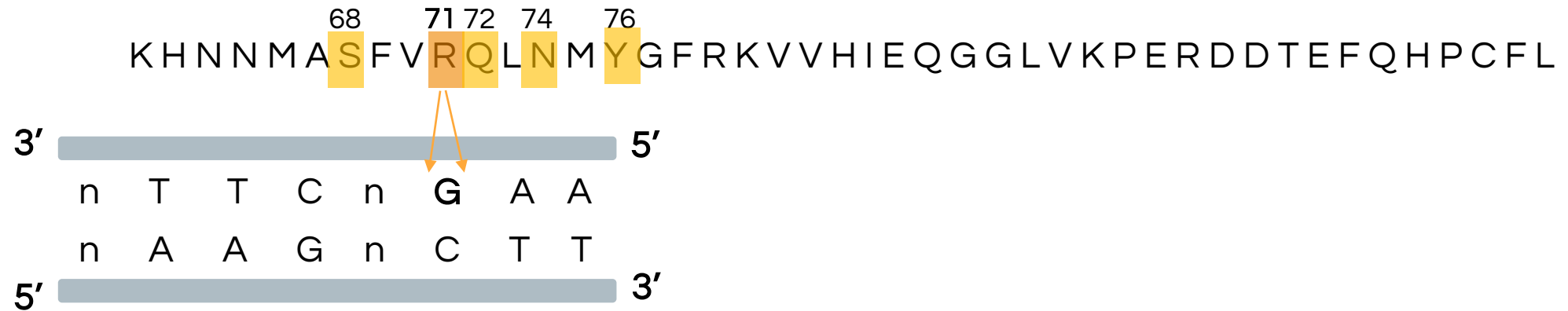
62 68 71 72 74 76
KHNNMASFVRQLNMYGFRKVV
HIEQGGLVKPERDDTEFQHPCF
104



HSF1

Important residues and interactions

DNA-protein specific interactions



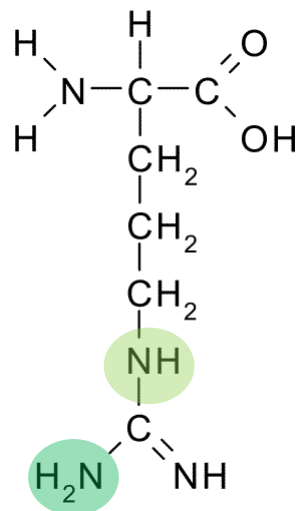
SPECIFIC INTERACTIONS:	Guanidinium group of Arg 71	- nGAAn	Parallel hydrogen bonds
	Guanidinium group of Arg 71	- Arg 71 (-COOH)	Hydrogen bond
	Aliphatic group of Arg 71	- nGAAn	Hydrophobic interaction

HSF1

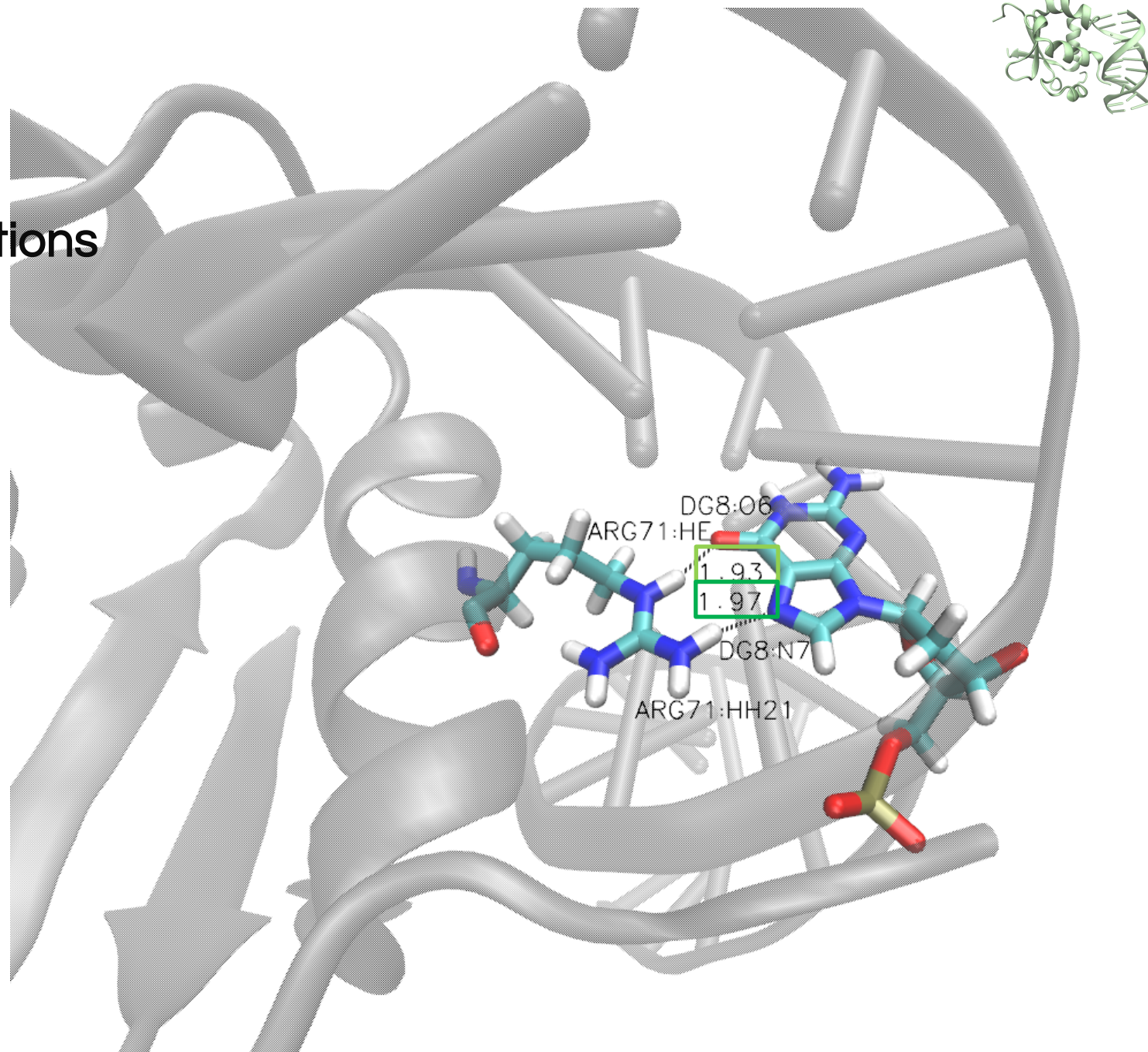
Important residues and interactions

DNA-protein specific interactions

Hydrogen bonds		
Arg 71 (-NH ₂)	1,97 Å	DG8 (N7)
Arg 71 (-NH)	1,93 Å	DG8 (N6)



Arg 71

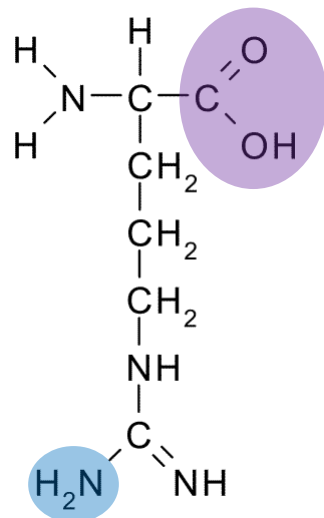


HSF1

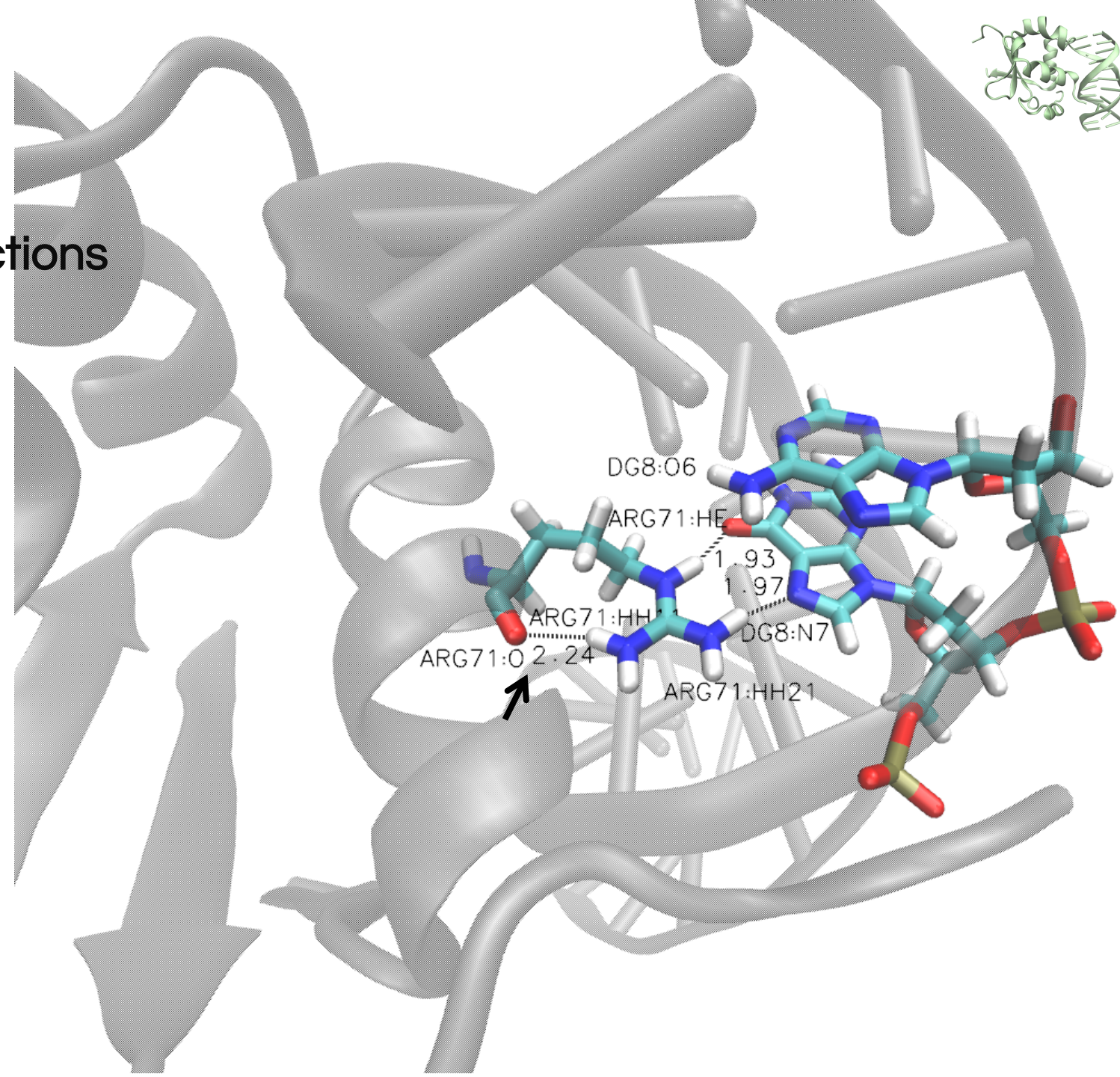
Important residues and interactions

DNA-protein specific interactions

Hydrogen bonds		
Arg 71 (-NH ₂)	2,24 Å	Arg71 (-COOH)



Arg 71

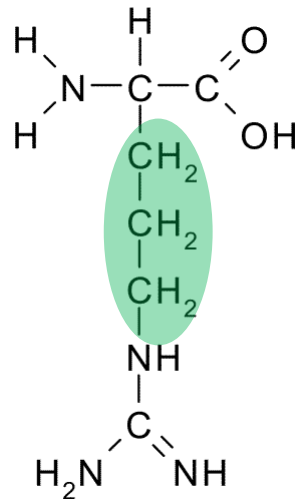


HSF1

Important residues and interactions

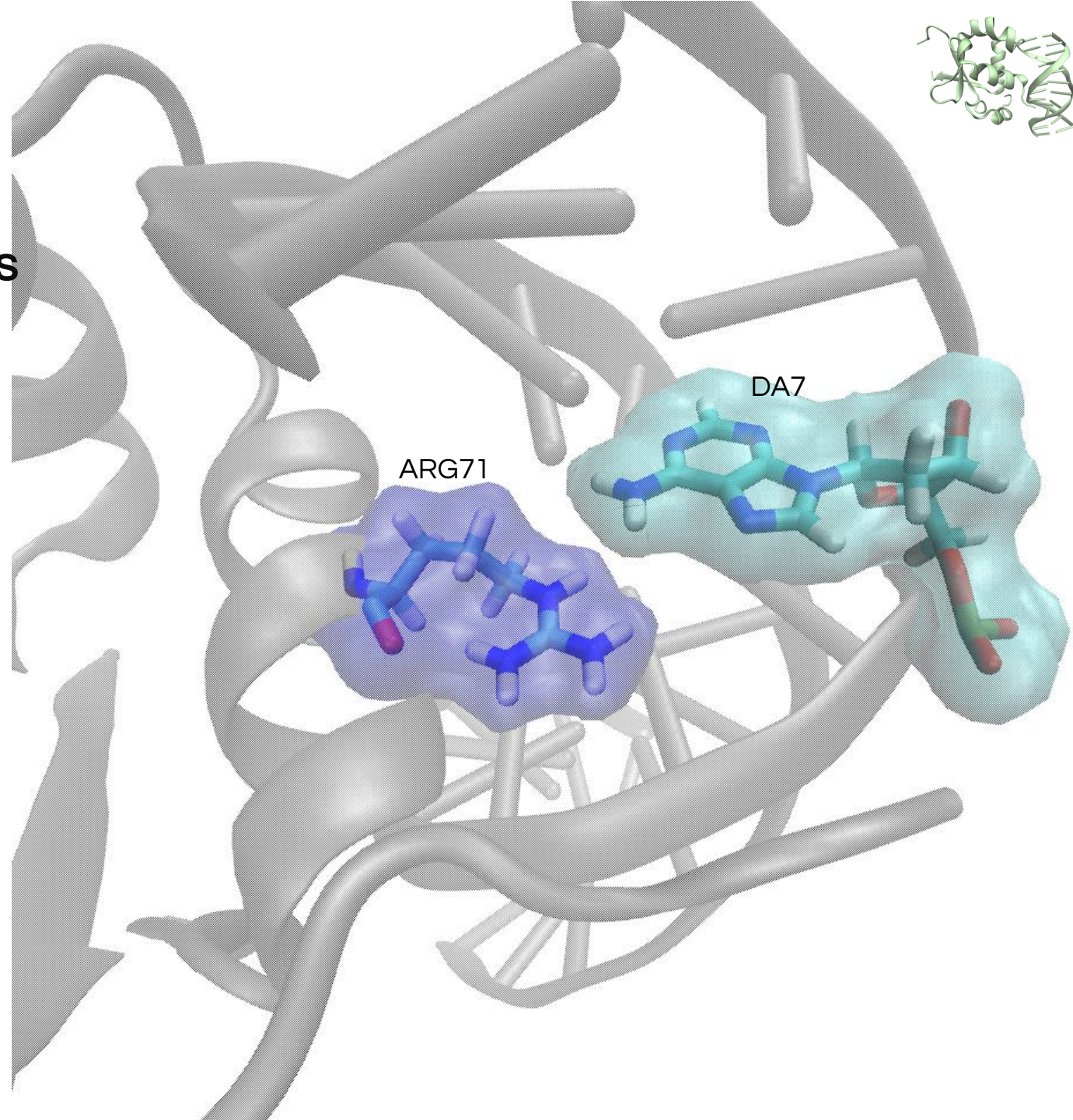
DNA-protein specific interactions

Hydrophobic interactions		
Arg 71 (aliphatic)	2,24 Å	DA7



Arg 71

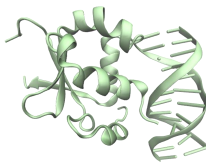
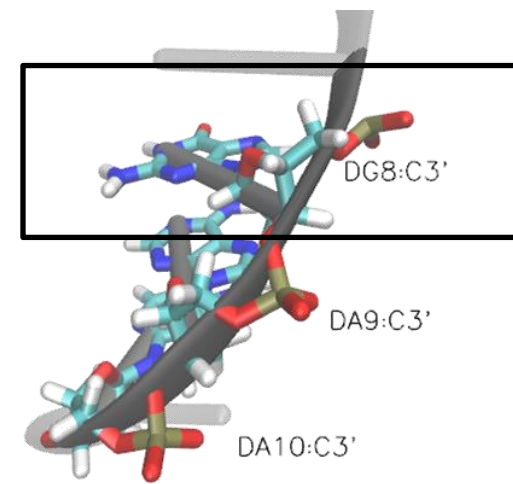
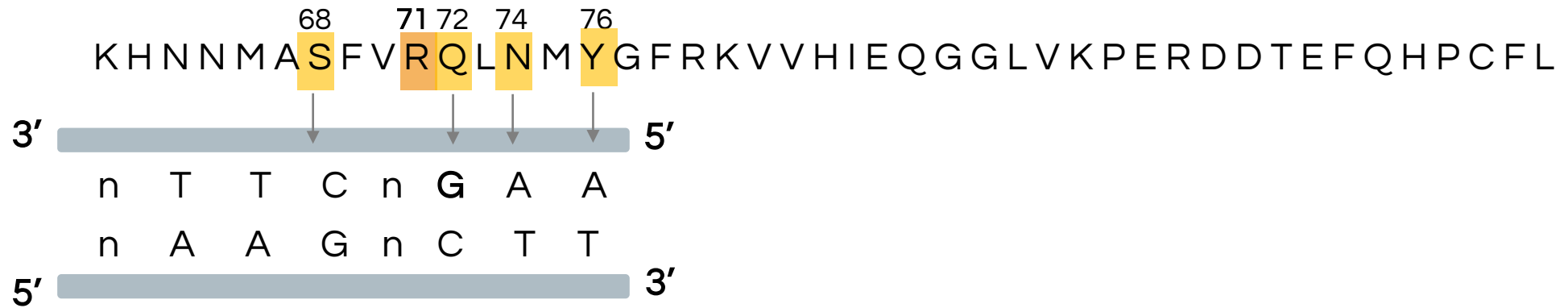
↑
Base
preceding G



HSF1

Important residues and interactions

DNA-protein unspecific interactions



UNSPECIFIC INTERACTIONS:

Stabilization of specific interactions

Ser 68

Gln 72

Asn 74

Tyr 76

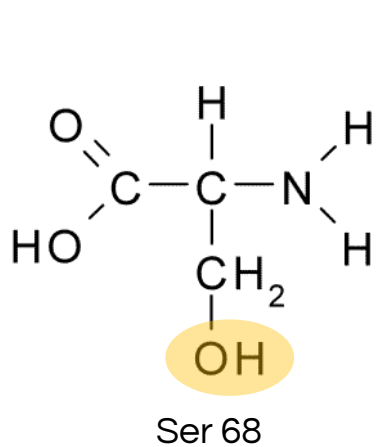
} Hydrogen bonds with the phosphate backbone

HSF1

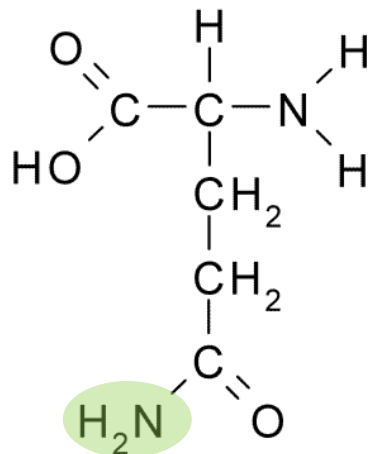
Important residues and interactions

DNA-protein unspecific interactions

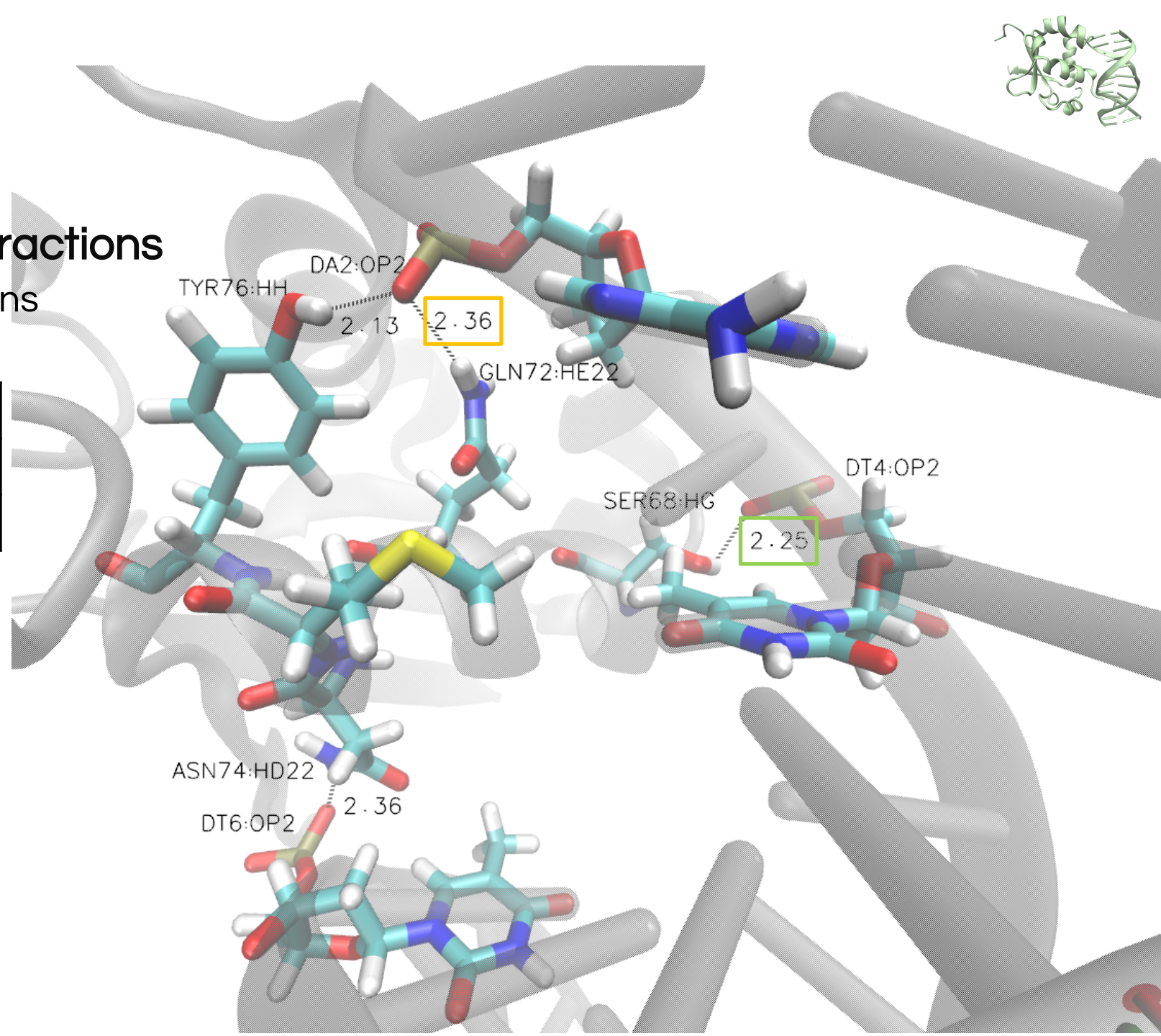
Hydrogen bonds		
Ser 68 (-OH)	2,25 Å	DT4 (OP2)
Gln 72 (-NH ₂)	2,36 Å	DA2 (OP2)



Ser 68



Gln 72

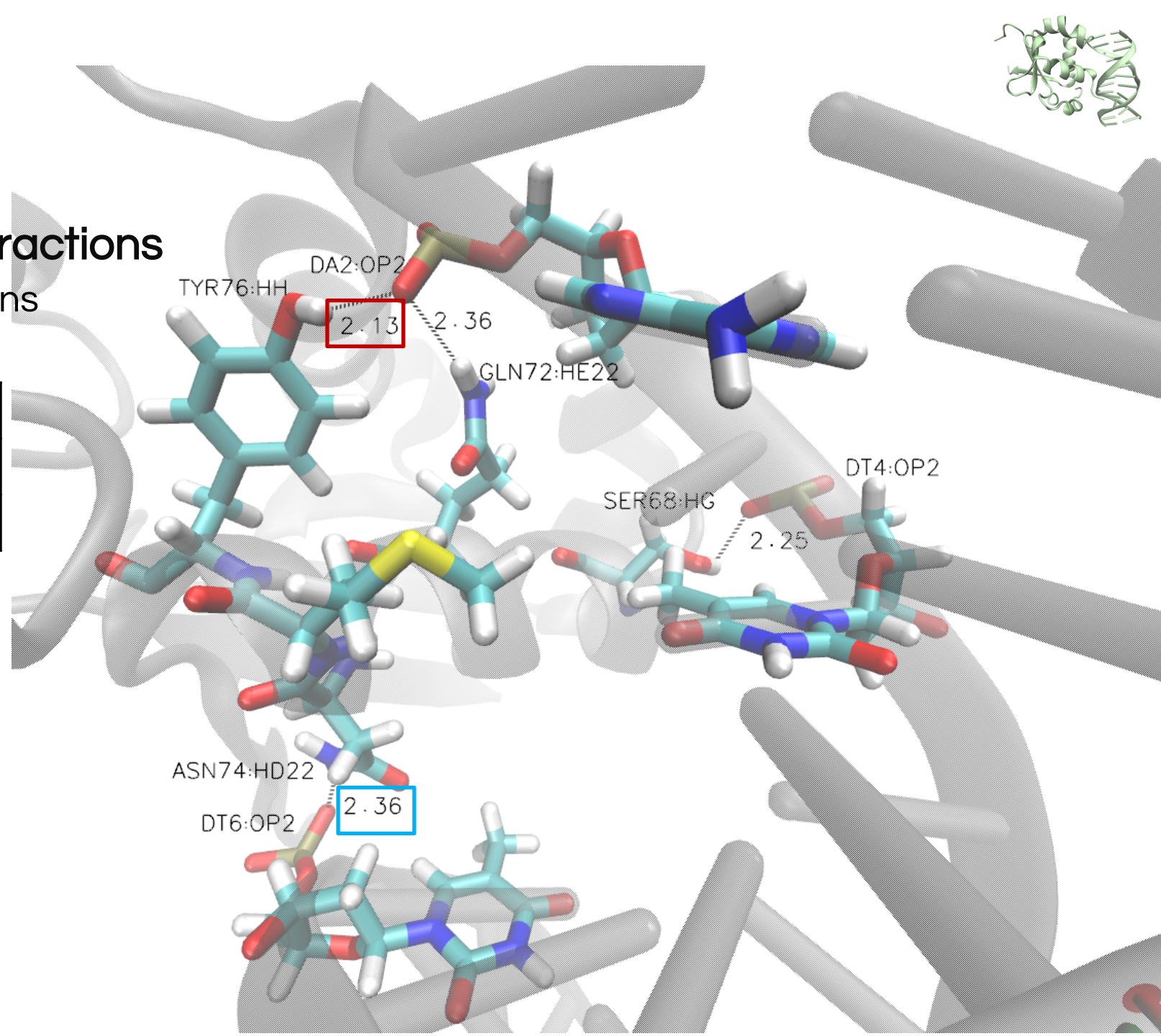
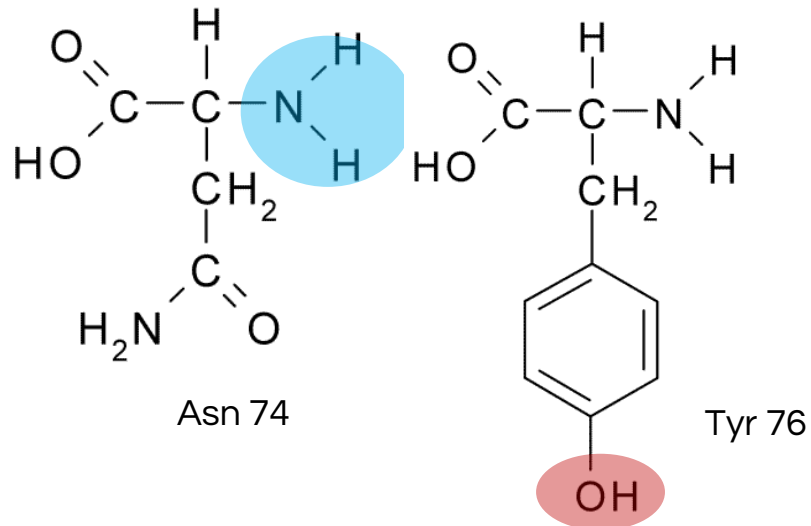


HSF1

Important residues and interactions

DNA-protein unspecific interactions

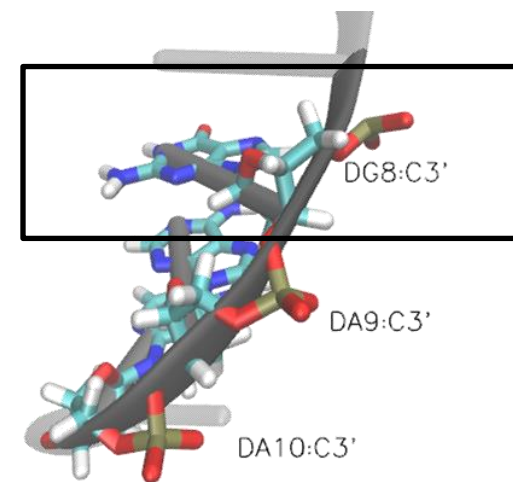
Hydrogen bonds		
Asn 74 (-NH ₂)	2,36 Å	DT6 (OP2)
Tyr 76 (-OH)	2,13 Å	DA2 (OP2)



HSF1

Important residues and interactions

DNA-protein interactions. Sequence alignment



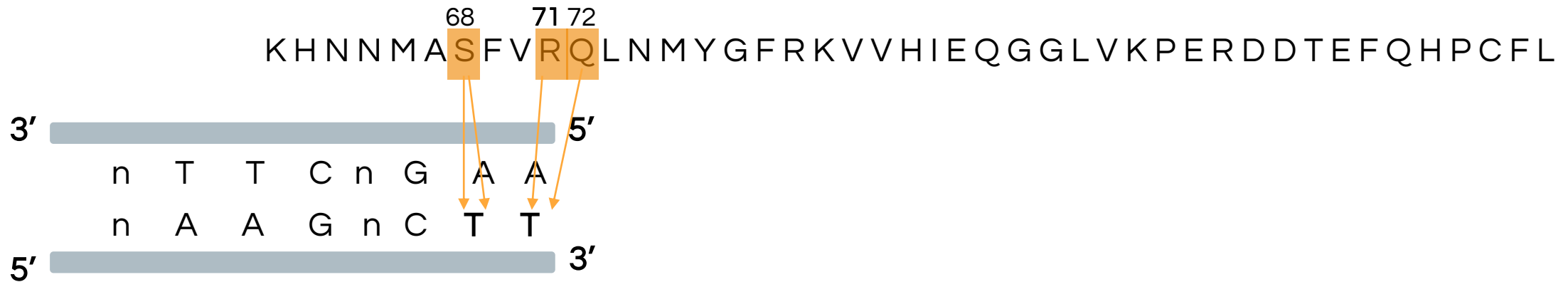
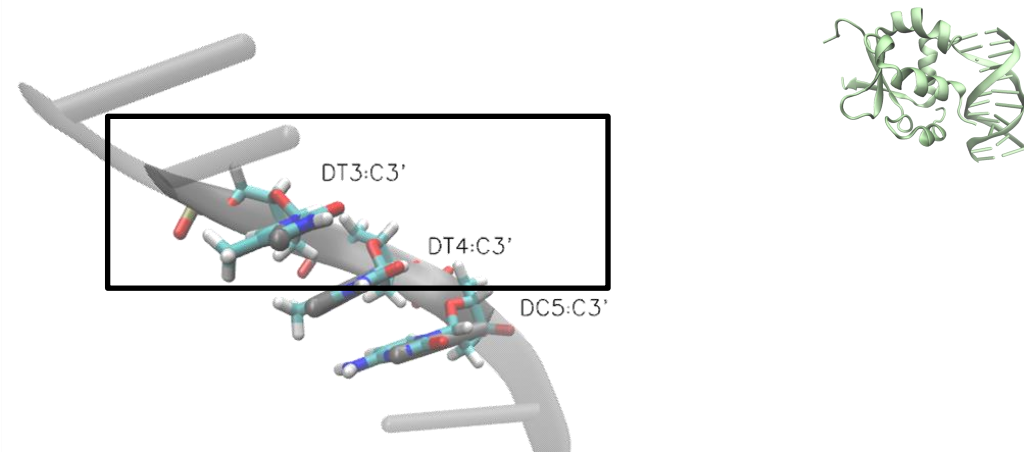
sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp Q08DJ8 HSF1_BTAURUS	KHSNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLLENIKRK
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK
sp P22813 HSF_DMELANOGASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLLDQIKRK
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIIRQ

All conserved

HSF1

Important residues and interactions

DNA-protein specific interactions



SPECIFIC INTERACTIONS:

Ser 68

Arg 71

Gln 72

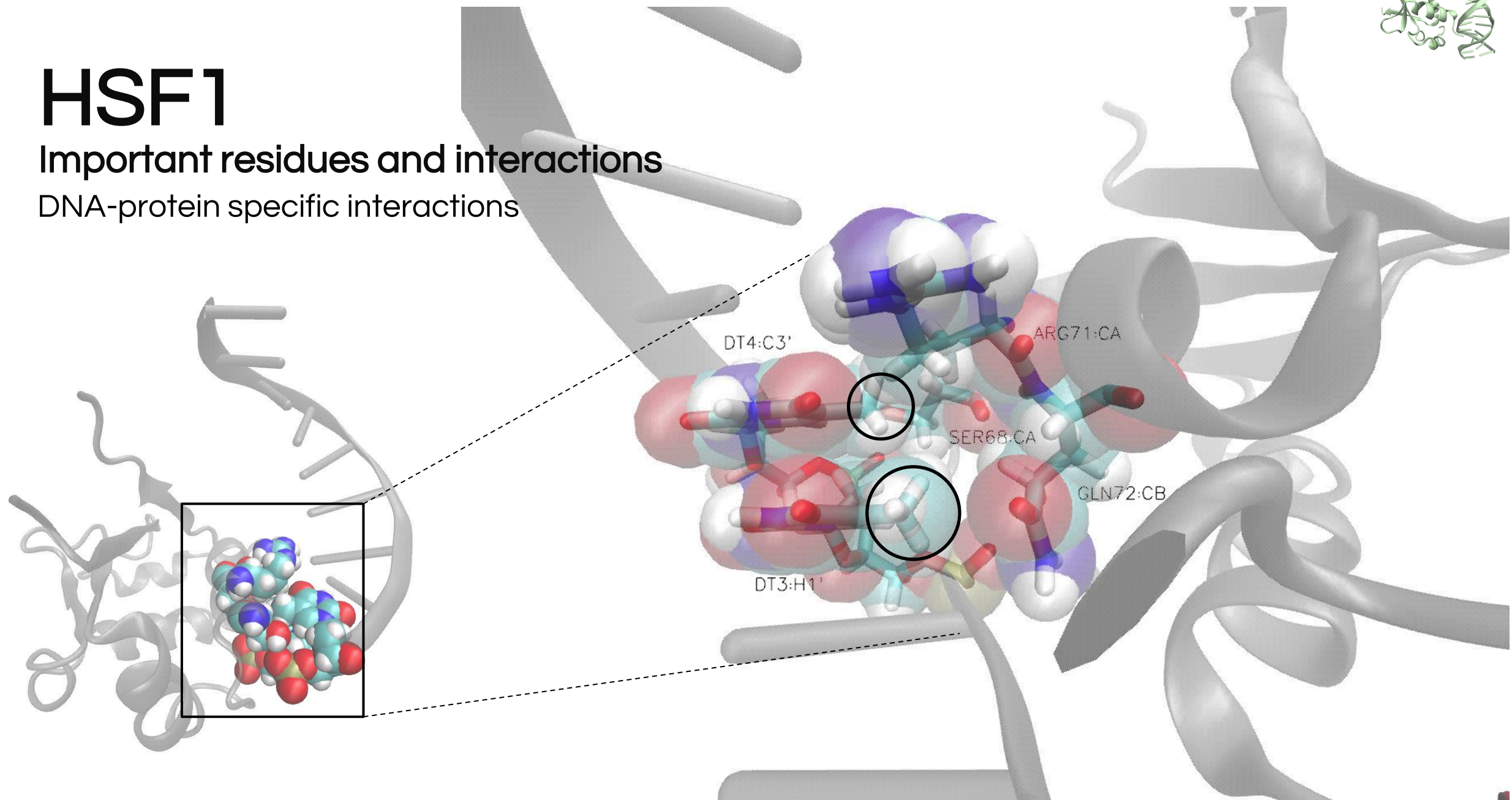
- nAAGnCTTn

Van der Waals

HSF1

Important residues and interactions

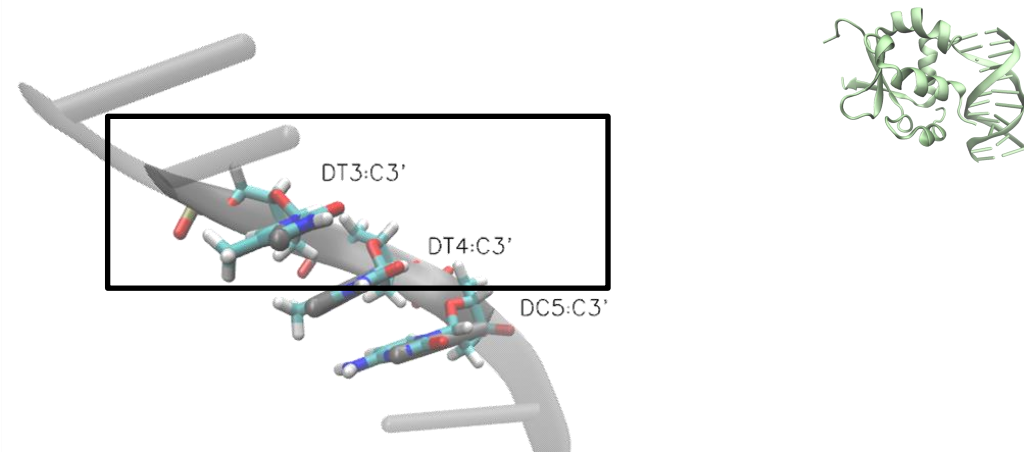
DNA-protein specific interactions



HSF1

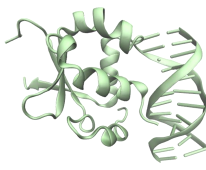
Important residues and interactions

DNA-protein specific interactions



sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp Q08DJ8 HSF1_BTAURUS	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHHPYFIRGQEHLLLENIKRK
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHHPYFIRGQEQLLENIKRK
sp P22813 HSF_DMELANOGASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLLDQIKRK
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ

All conserved



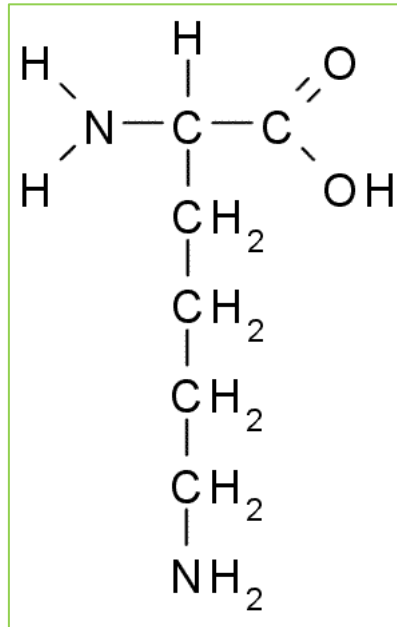
HSF1

Important residues and interactions

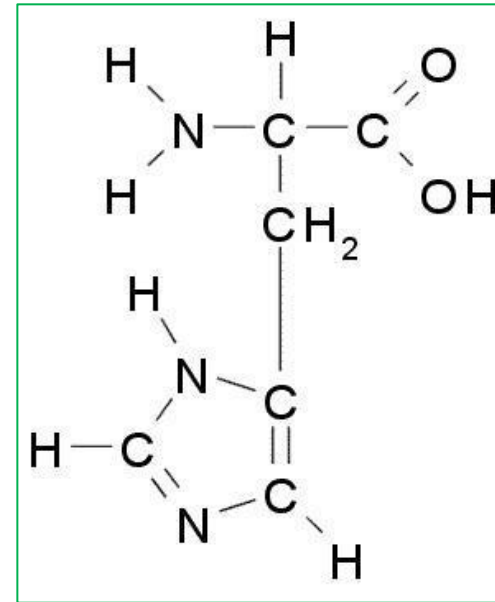
DNA-protein stabilization interactions

1. **Lys** 62 – **His** 63
2. **Arg** 79 – **Lys** 80
3. **Arg** 117 – **Lys** 118

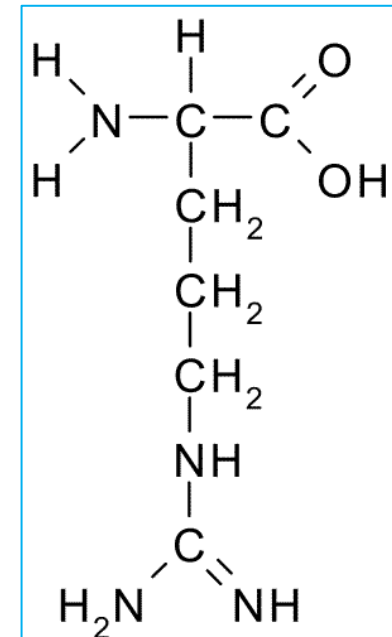
Three pairs of
basic residues



Lysine



Histidine



Arginine

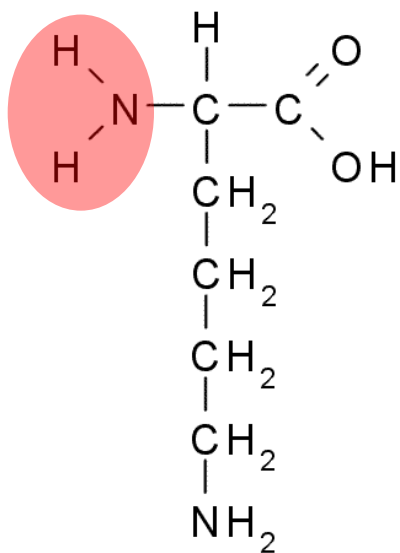


HSF1

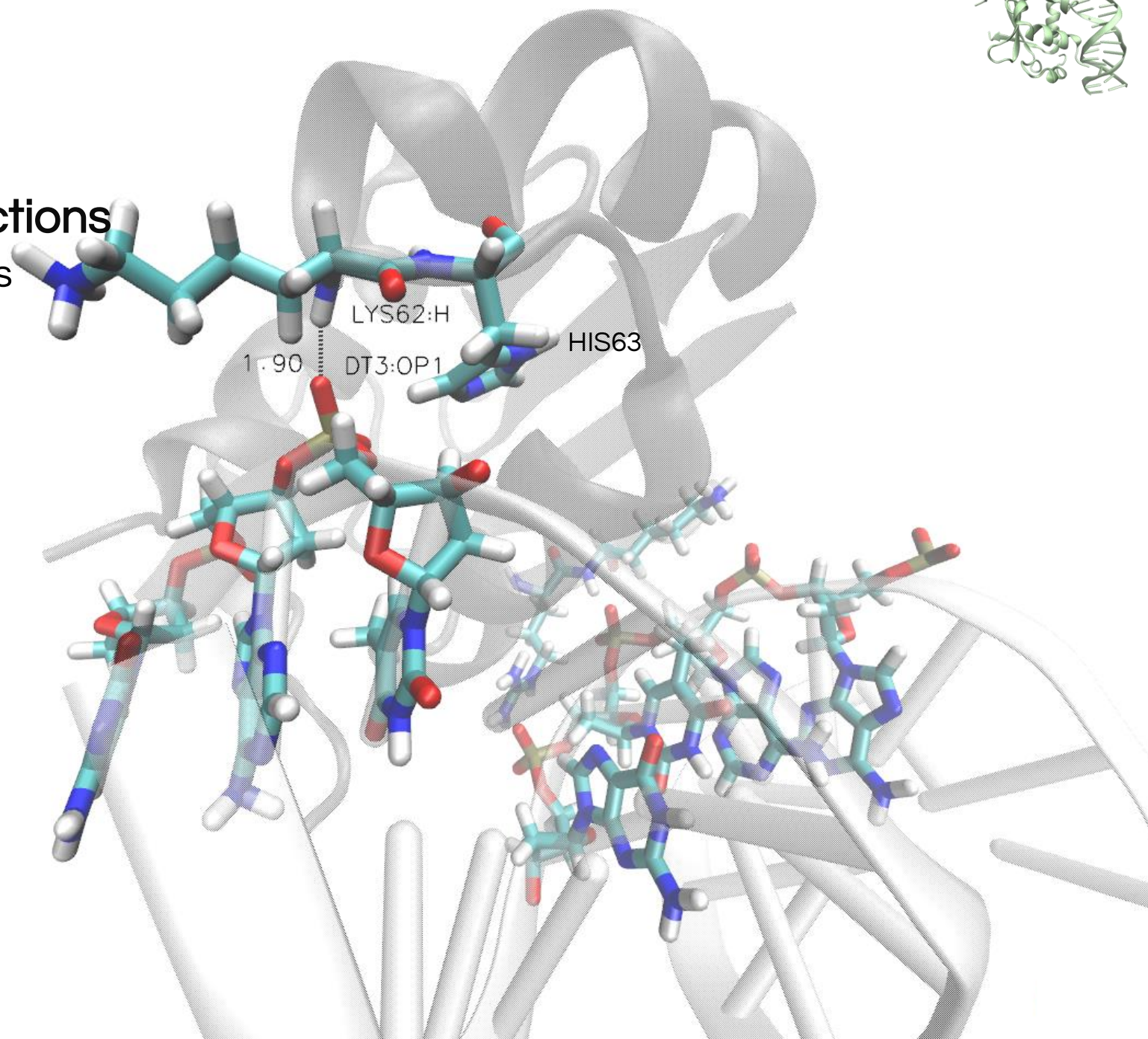
Important residues and interactions

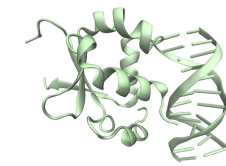
DNA-protein stabilization interactions

1	Hydrogen bonds		
	Lys 62 (-NH ₂)	1,90 Å	DT3 (OP1)



Lys 62



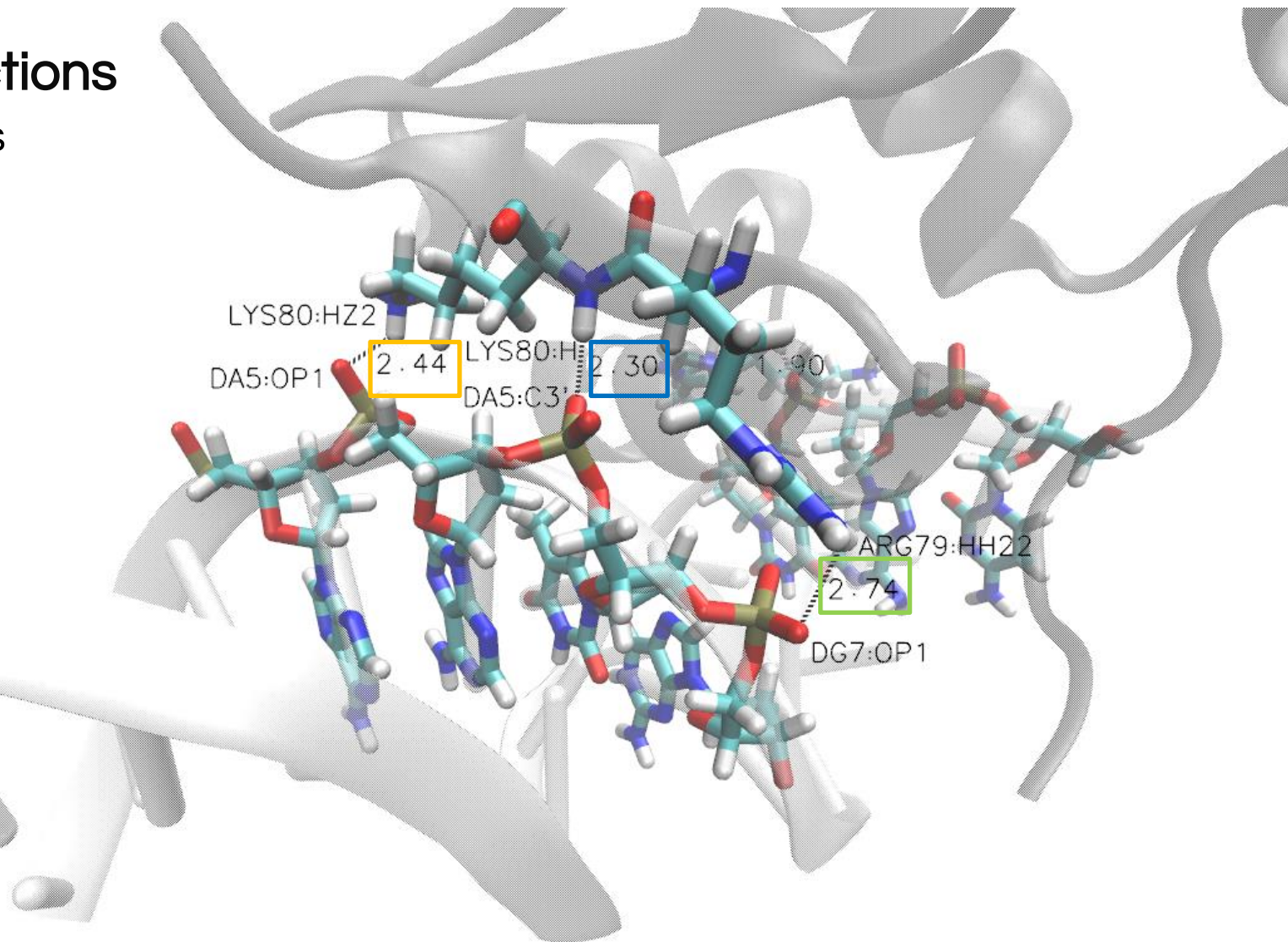
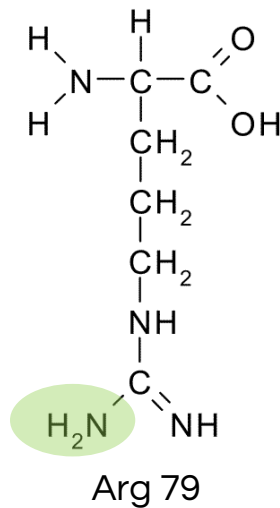
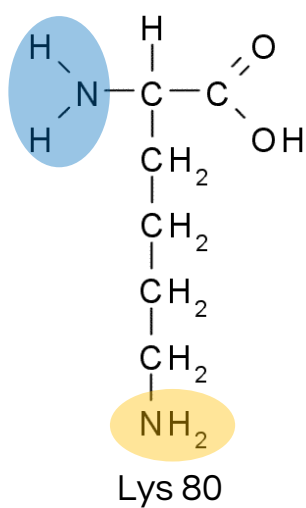


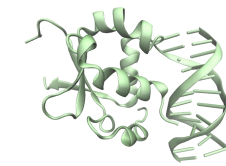
HSF1

Important residues and interactions

DNA-protein stabilization interactions

2	Hydrogen bonds	
Arg 79 (-NH ₂)	2,74 Å	DG7 (OP1)
Lys 80 (-NH ₂)	2,44 Å	DA5 (OP1)
Lys 80 (-NH ₂)	2,30 Å	DA5 (C3)



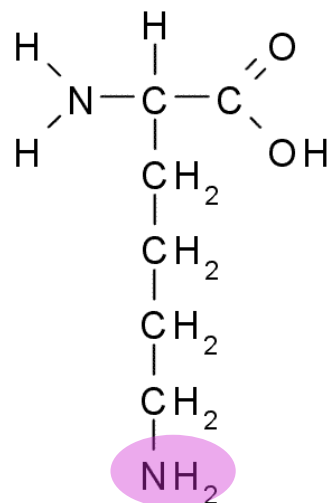


HSF1

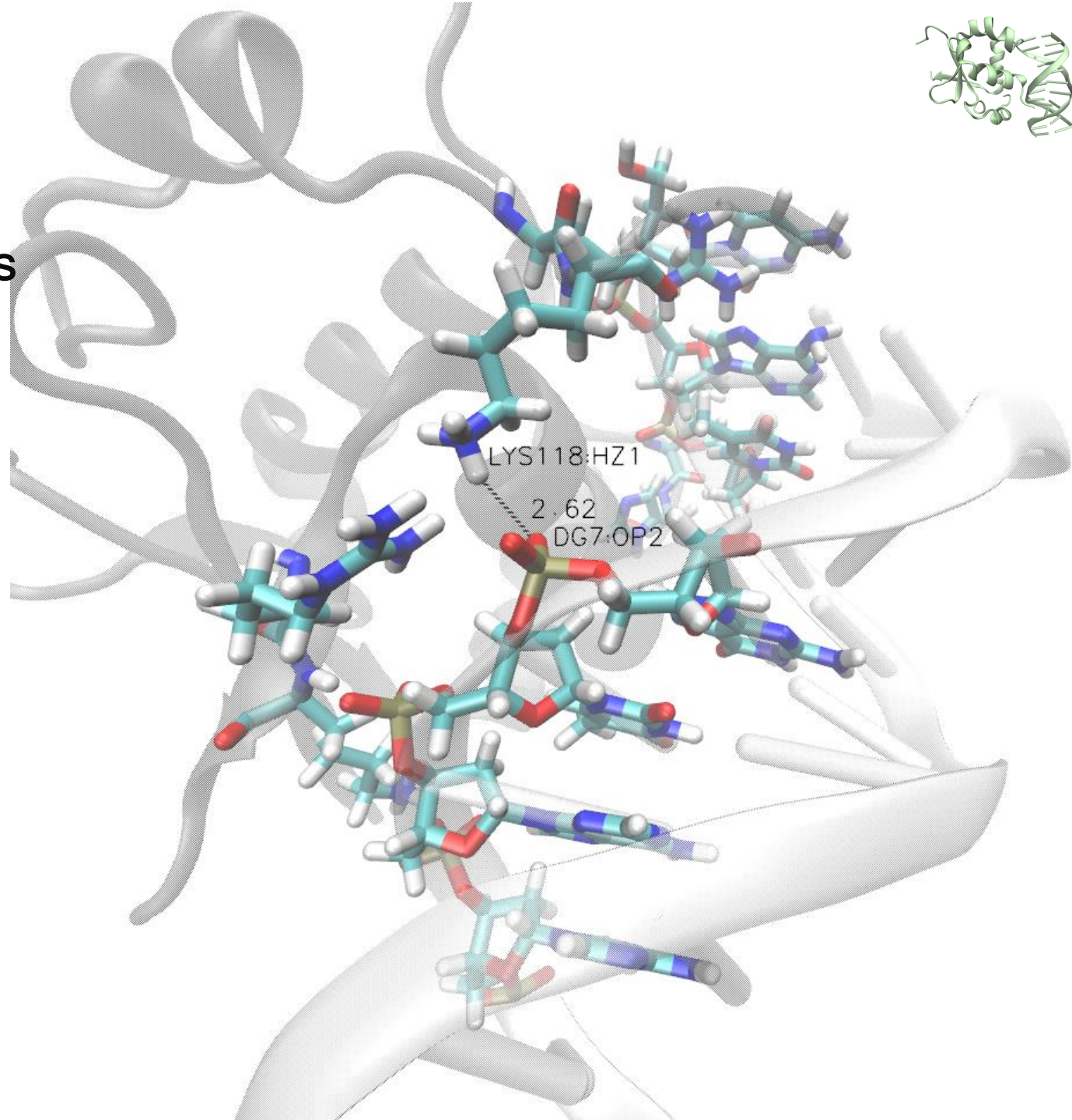
Important residues and interactions

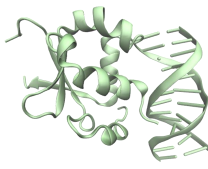
DNA-protein stabilization interactions

3	Hydrogen bonds	
Lys 118 (-NH ₂)	2,62 Å	DG7 (OP2)



Lys 118





HSF1

Important residues and interactions

DNA-protein stabilization interactions

Some of the **Lys** residues can be

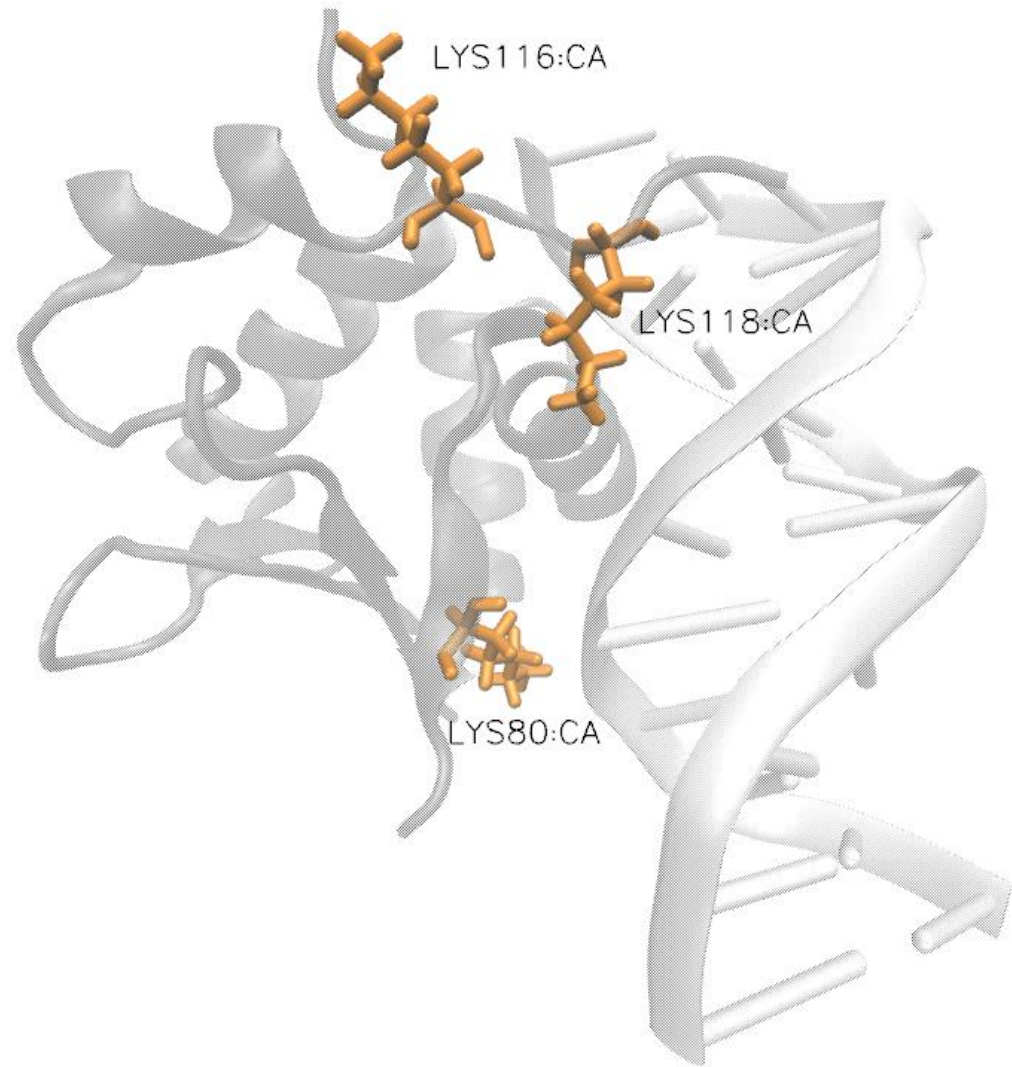
acetylated

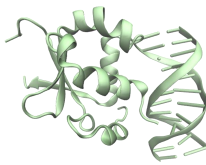


Lose the positive charge



Lowering overall affinity to DNA.





HSF1

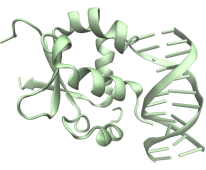
Important residues and interactions

DNA-protein stabilization interactions

sp Q00613 HSF1_HSAPIENS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38532 HSF1_MMUSCULUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp Q08DJ8 HSF1_BTAURUS	KHSNMAFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPCFLRGQEQLLENIKRK
sp P38529 HSF1_GGALLUS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eKDDTEFQHPYFIRGQEHLLENIKRK
sp P41154 HSF_XLAEVIS	KHNNMASFVRQLNMYGFRKVVHIEQGGLV--Kp-eRDDTEFQHPYFIRGQEQLLENIKRK
sp P22813 HSF_DMELANOASTER	KHNNMASFIRQLNMYGFHKITSIDNGGLR-fD---RDEIEFSHPFFKRNSPFLLDQIKRK
sp P22121 HSF_KLACTIS	KHSNFASFVRQLNMYGWHKVQDVKSGSMLsnN---DSRWEFENENFKRGKEYLLENIVRQ
sp Q02953 HSF_SPOMBE	KHNNFSSFVRQLNMYGFHKVPHIQQGVQLQ--SdspNELLEFANPNFQRDQPELLCLVTRK
sp P10961 HSF_SCEREVISIAE	KHSNFASFVRQLNMYGWHKVQDVKSGSIQ--Ss-sDDKWQFENENFIRGREDLLEKIIRQ



Arg 79 → His 79 in *D. melanogaster*, *K. lactis*, *S. pombe* and *S. cerevisiae* → Both positively charged and polar

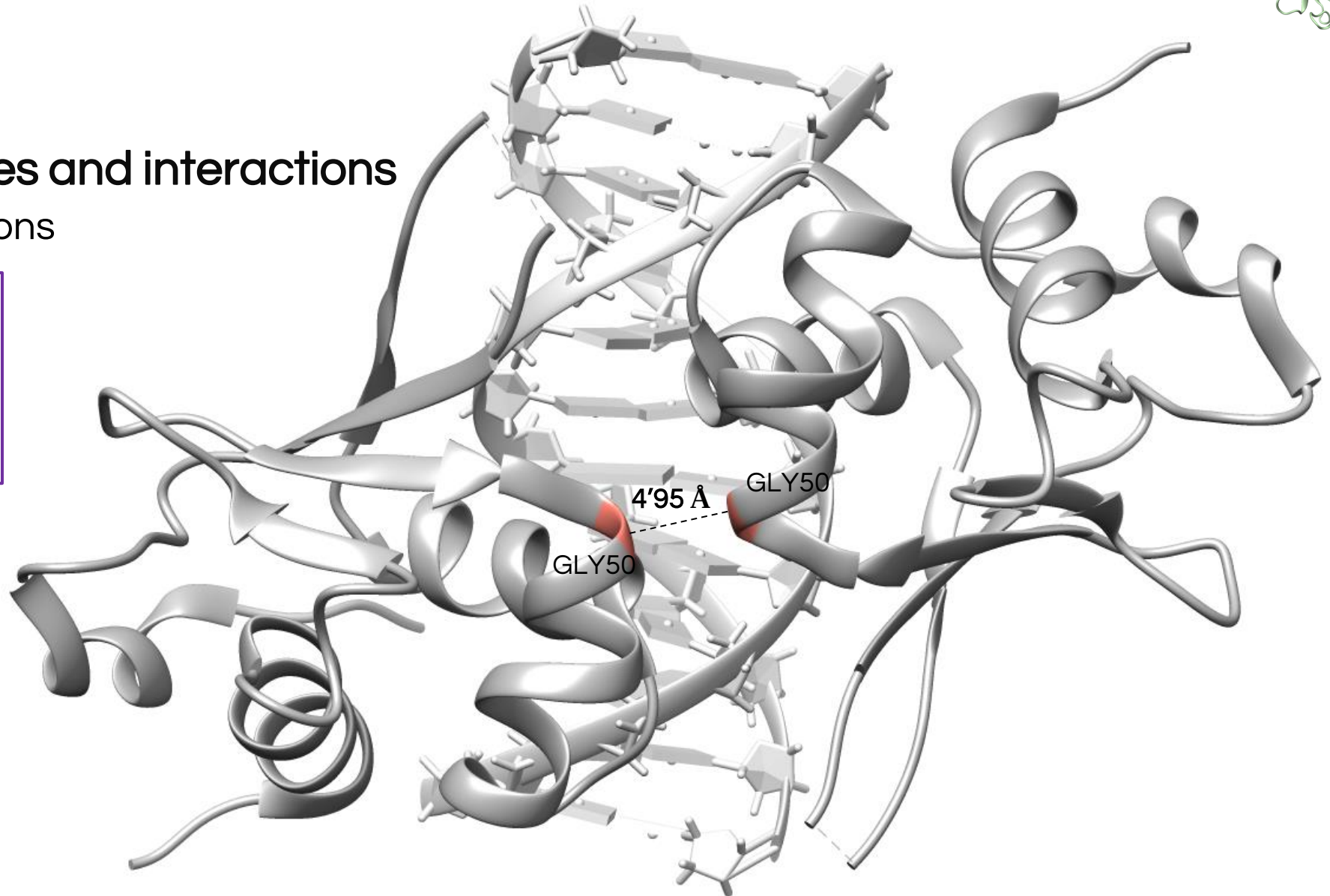


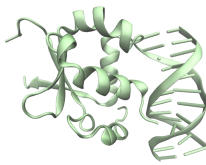
HSF1

Important residues and interactions

DBD – DBD interactions

- Gln 50 - Gly 50
- Other hydrophobic interactions





HSF1

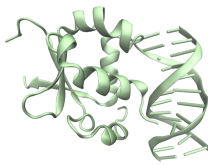
Important residues and interactions

DBD – DBD interactions

sp Q00613 HSF1_HSAPIENS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P38532 HSF1_MMUSCULUS	dlavgpagaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp Q08DJ8 HSF1_BTAURUS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVLDQGGQFAKEVLPKYF
sp P38529 HSF1_GGALLUS	aaaavgagpggsnvsAFLTKLWTLVEDPETDPLICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P41154 HSF_XLAEVIS	--mdphgtcggsnvpAFLAKLWTLVEDPDTDPLICWSPEGNSFHVFDQGGQFAKEVLPKYF
sp P22813 HSF_DMELANOGASTER	mhsygdaaaigsgvpAFLAKLWRLVDDADTNRLICWTKDGQS FVIQNQAQFAKELLPLNY
sp P22121 HSF_KLACTIS	khhskkklsttrarpAFVNKLWSMVNDKSNEKFHWSTSGESIVVPNRERFVQEVLPKYF
sp Q02953 HSF_SPOMBE	vgmsgsgssqnkitQFSNKLYNMVNDSSSTD SLIRWSDRGDSFLVIGHEDFAKLVLPRYF
sp P10961 HSF_SCEREVISIAE	prqttrryqshksrpAFVNKLWSMLNDDSN TKLIQWAEDGKS FIVTNREEFVHQILPKYF

Gly 50 → Ala 50 in *D. melanogaster* → Both **little** and apolar

Gly 50 → Glu 50 in *K. lactis*, *S. pombe*, *S. cerevisiae* → **Glu**
is **polar** and **negatively charged**



HSF1

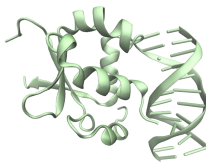
Important residues and interactions

DBD – DBD interactions

sp Q00613 HSF1_HSAPIENS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P38532 HSF1_MMUSCULUS	dlavgpagaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp Q08DJ8 HSF1_BTAURUS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVLDQGGQFAKEVLPKYF
sp P38529 HSF1_GGALLUS	aaaavgagpggsnvsAFLTKLWTLVEDPETDPLICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P41154 HSF_XLAEVIS	--mdphgtcggsnvpAFLAKLWTLVEDPDTDPLICWSPEGNSFHVFDQGGQFAKEVLPKYF
sp P22813 HSF_DMELANOASTER	mhsygdaaaigsgvpAFLAKLWRLVDDADTNRLICWTKDGQS FVIQNQAQFAKELLPLNY
sp P22121 HSF_KLACTIS	khhskkklsttrarpAFVNKLWSMVNDKSNEKFHWSTSGESIVVPNRERFVQEVLPKYF
sp Q02953 HSF_SPOMBE	vgmsgsgssqnkitQFSNKLYNMVNDSSSTD SLIRWSDRGDSFLVIGHEDFAKLVLPRYF
sp P10961 HSF_SCEREVISIAE	prqttrryqshksrpAFVNKLWSMLNDDSN TKLIQWAEDGKSFIVTNREEFVHQILPKYF

Gln 49 → Arg 49 in *K. lactis* and *S. cerevisiae* or His 49 in *S. pombe* → Both changes are also polar but positively charged

Gln 51 → Arg 51 in *K. lactis*, Asp 51 in *S. pombe* and Glu 51 in *S. cerevisiae* → All changes are also polar but charged amino acids: Arg +, Asp - and Glu -.



HSF1

Important residues and interactions

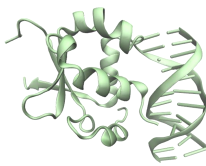
DBD – DBD interactions

sp Q00613 HSF1_HSAPIENS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P38532 HSF1_MMUSCULUS	dlavgpagaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp Q08DJ8 HSF1_BTAURUS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVLDQGGQFAKEVLPKYF
sp P38529 HSF1_GGALLUS	aaaavgagpggsnvsAFLTKLWTLVEDPETDPLICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P41154 HSF_XLAEVIS	--mdphgtcggsgnvpAFLAKLWTLVEDPDTDPLICWSPEGNSFHVFDQGGQFAKEVLPKYF
sp P22813 HSF_DMELANOGASTER	mhsygdaaaigsgvpAFLAKLWRLVDDADTNRLICWTKDGQSFVIQNQAQFAKELLPLNY
sp P22121 HSF_KLACTIS	khhskkklsttrarpAFVNKLWSMVNDKSNEKFHWSTSGESIVVPNRERFVQEVLPKYF
sp Q02953 HSF_SPOMBE	vgmsgsgssqnkitQFSNKLYNMVNDSSSTDLSIRWSDRGDSFLVIGHEDFAKLVLPRYF
sp P10961 HSF_SCEREVISIAE	prqttrryqshksrpAFVNKLWSMLNDDSN TKLIQWAEDGKSFIVTNREEFVHQILPKYF

Asp 48 → Asn 48 in *D. melanogaster*, *K. lactis*
and *S. cerevisiae* and Gly 48 → Both changes
are little but not charged

Lys 54 → Gln 54 in *K. lactis* → Both polar but
Gln not charged

Lys 54 → His 54 in *S. cerevisiae* → Both
positively charged



HSF1

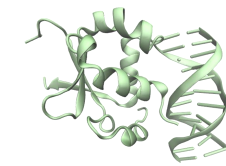
Important residues and interactions

DBD – DBD interactions

sp Q00613 HSF1_HSAPIENS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P38532 HSF1_MMUSCULUS	dlavgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp Q08DJ8 HSF1_BTAURUS	dlpvgpgaagpsnvpAFLTKLWTLVSDPDTDALICWSPSGNSFHVLDQGGQFAKEVLPKYF
sp P38529 HSF1_GGALLUS	aaaavgagpggsnvsAFLTKLWTLVEDPETDPLICWSPSGNSFHVFDQGGQFAKEVLPKYF
sp P41154 HSF_XLAEVIS	--mdphgtcggpsnvpAFLAKLWTLVEDPDTDPLICWSPEGNSFHVFDQGGQFAKEVLPKYF
sp P22813 HSF_DMELANOASTER	mhsygdaaaigsgvpAFLAKLWRLVDDADTNRLICWTKDGQS FVIQNQAQFAKELLPLNY
sp P22121 HSF_KLACTIS	khhskkklsttrarpAFVNKLWSMVNDKSNEKFIHWSTSGESIVVPNRERFVQEVLPKYF
sp Q02953 HSF_SPOMBE	vgmsgsgssqnkitQFSNKLYNMVNDSSSTD SLIRWSDRGDSFLVIGHEDFAKLVLPRYF
sp P10961 HSF_SCEREVISIAE	prqttrryqshksrpAFVNKLWSMLNDDSN TKLIQWAEDGKS FIVTNREEFVHQILPKYF

Low conservation is showed, specially in the more distant species, like yeasts

→ These residues are not as important as the ones that interact with DNA.



HSF1

Superimposition between homologues

- PDB 5hdg (*Homo sapiens*)
- PDB 1fbu (*Kluyveromyces lactis*)
- PDB 1hks (*Drosophila melanogaster*)

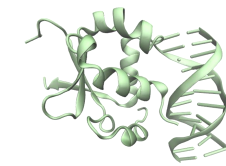
Cluster: 2 (1fbu & 5hdg 1hks)

Sc	6.98	RMS	1.33	Len	99	nfit	71
----	------	-----	------	-----	----	------	----

5hdg	HSAPIENS	--HVPAFLTKLWTLVSDPDTDALICWSP-SGNSFHVFDQGQFAKEVLPKYFKHNNMASFV
1hks	DMELANOGASTER	GSGVPAFLAKLWRLVDDADTNRLICWTK-DGQSFVIQNQAQFAKELLPLNYKHNNMASFI
1fbu	KLACTIS	----PAFVNKLWSMVNDKSNEKFIHWSTSG-ESIVVPNRERFVQEVLPKYFKHSNFASFV

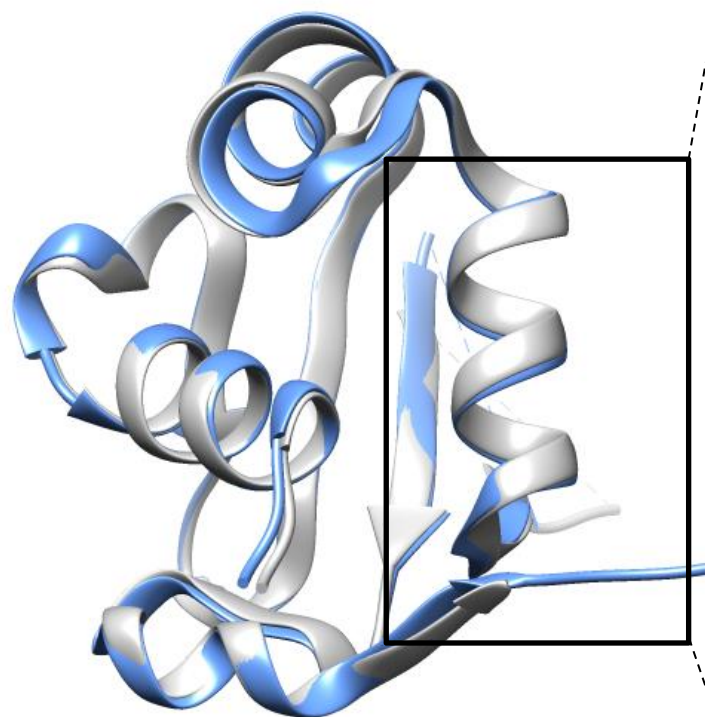
5hdg	HSAPIENS	RQLNMYGFRKVV-----HIE-DTEFQH-PCFLRGQEQLLENIKR-
1hks	DMELANOGASTER	RQLNMYGFHKITS-IDN-----GG-LRFDR--DEIEFSH-PFFKRNSPFLLDQIKRK
1fbu	KLACTIS	RQLNMYGWHKVQDV--KSGSMLSNNDS-----RWEFENER-HA-----



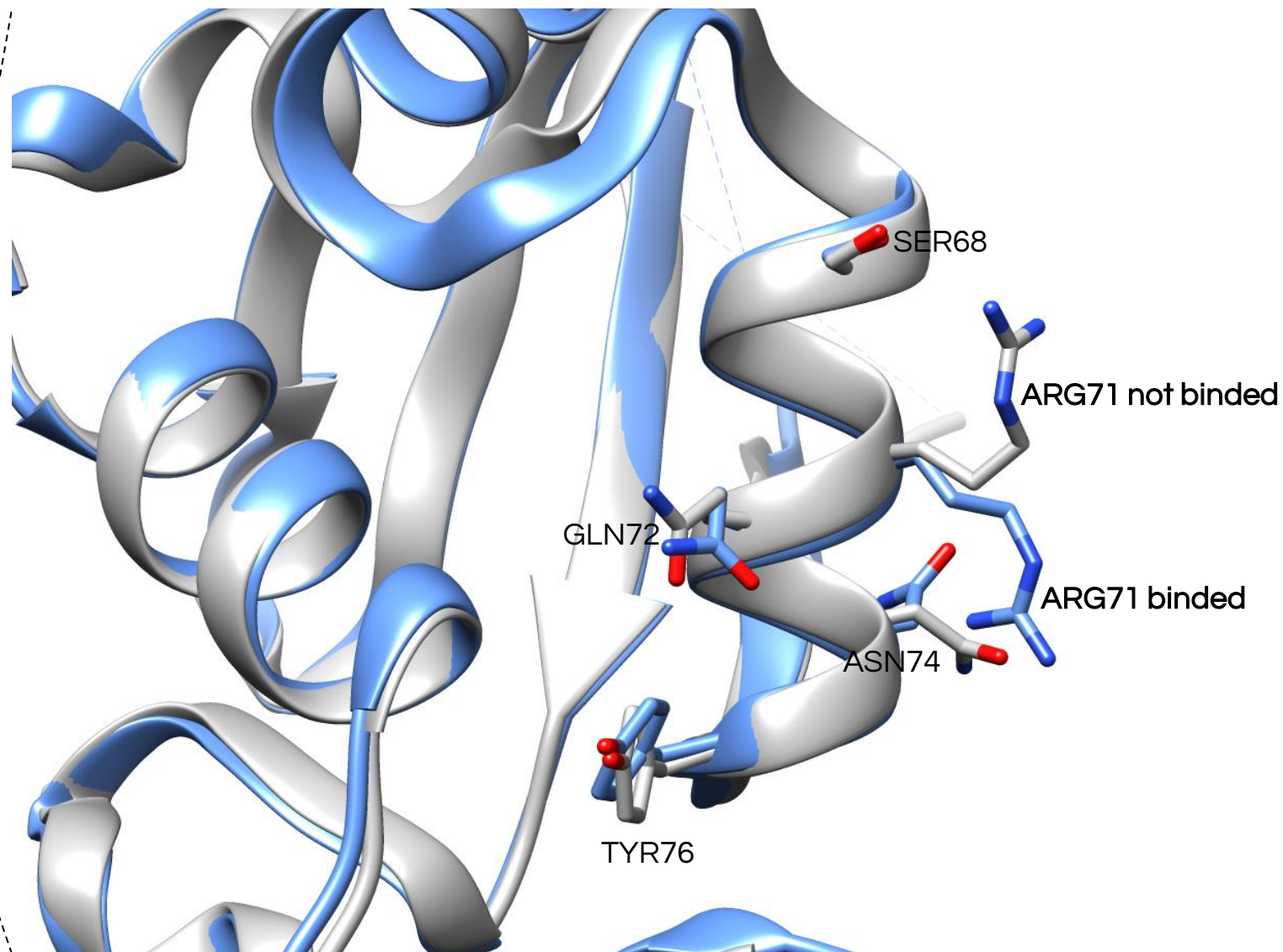


HSF1

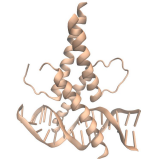
DNA binding/not binding



■ DNA binded ■ DNA not binded



Cluster: 1 (5d5v & 5hdg) Sc 8.72 RMS 0.51 Len 96 nfit 90



MITF

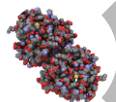
Index



Classification



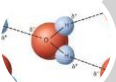
Biological function



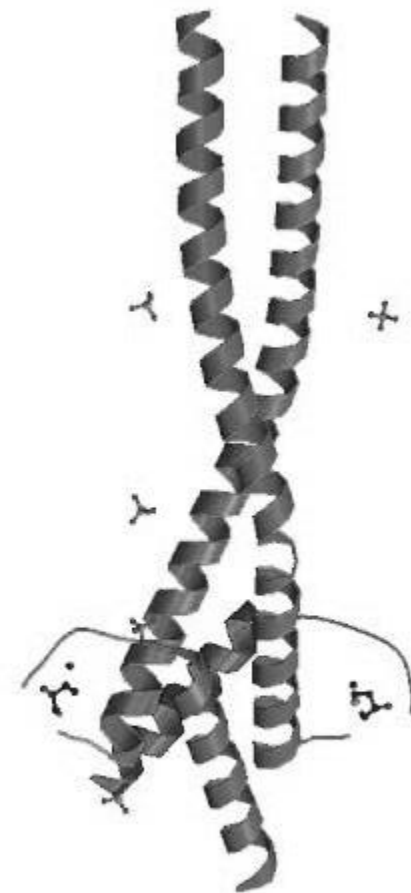
General structure

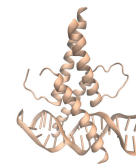


Secondary structure



Important residues and interactions





MITF

SCOP classification



Family: Helix-loop-helix DNA-binding domain

Lineage

Root

Scop

Class

All alpha proteins

Fold

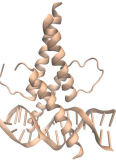
HLH-like

Superfamily

HLH, helix-loop-helix DNA-binding domain

Family

HLH, helix-loop-helix DNA-binding domain



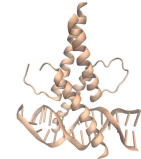
MITF

Biological function

- Regulation of genes with essential roles in **cell differentiation**, **proliferation** and **survival**:
 - e.g. Melanocyte development: **tyrosinase** (TYR) and **tyrosinase-related protein 1** (TYRP1).
- Mutations of the MITF protein are related to various **syndromes** and **diseases**:
 - Waardenburg syndrome 2A (WS2A)
 - Tietz Albinism-Deafness Syndrome (TADS)
 - Cutaneous malignant melanoma 8 (CMM8)
 - Microphthalmia

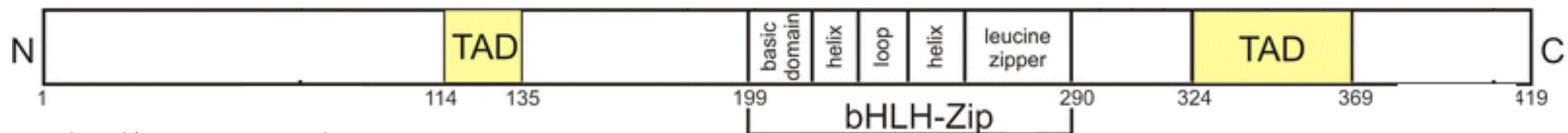


Author: Etan J. Tal

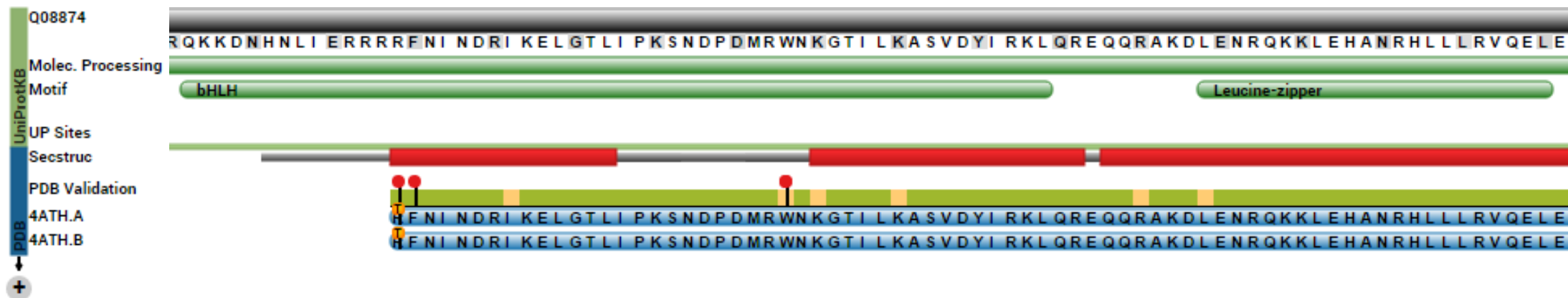


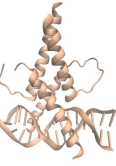
MITF

General structure



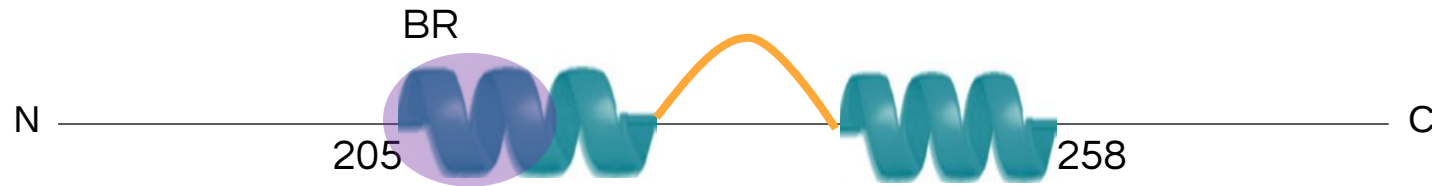
Adapted from Hartman ML and Czyz M, 2014.





MITF

Secondary structure



Helix- α

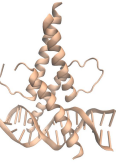
$\alpha 1$: (205-230)

Basic region: 205-221

$\alpha 2$: (243-258)

Loop:

(231-242)



MITF

Important residues and interactions

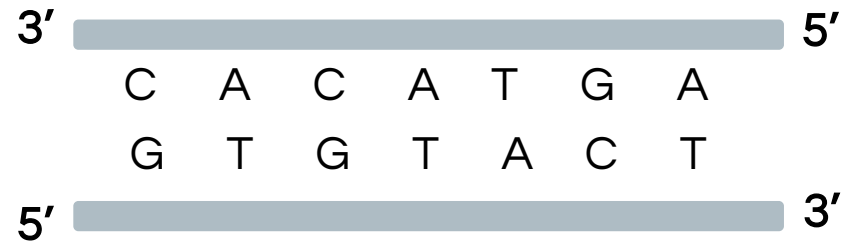
DNA-protein specific interactions

- MITF binds to 2 DNA motifs:

- the E-box

- the **M-box**:

K K D N **H** N L **I** **E** R R R R F N I N D R I K E L G T L





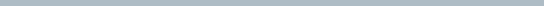
DNA-protein specific interactions

3' 5'

C A C A T G A

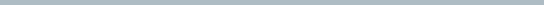
G T G T A C T

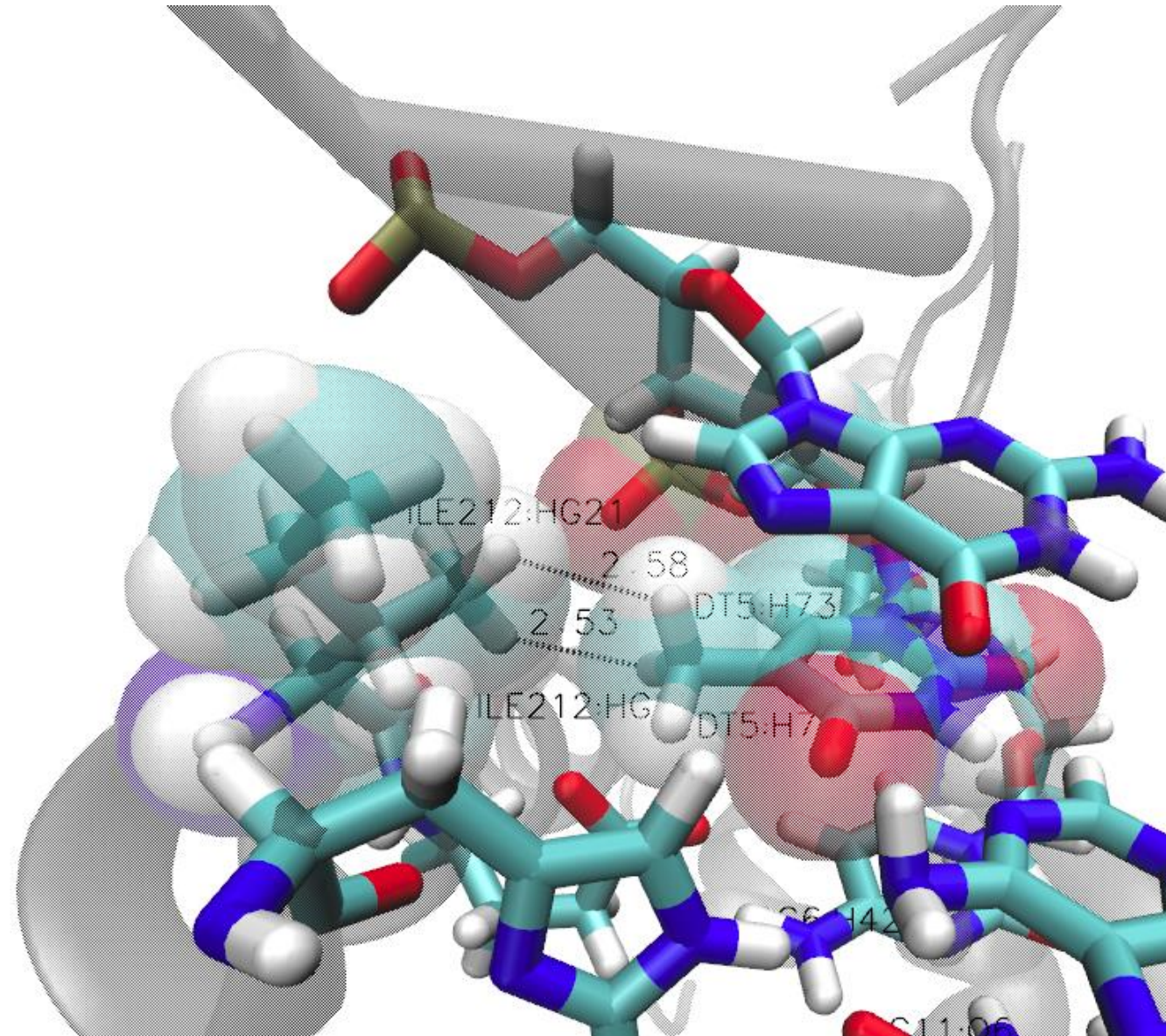
5' 3'

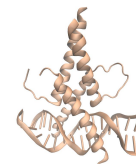
3'  5'

C A C G T G

G T G C A C

5'  3'



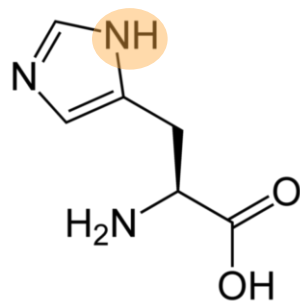


MITF

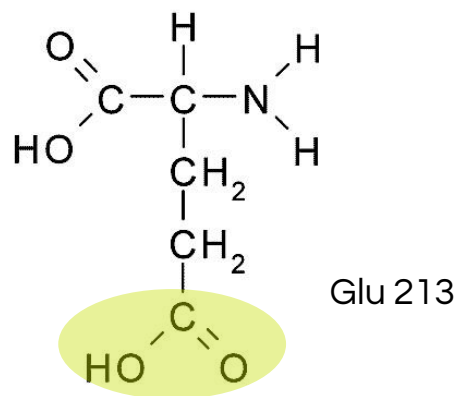
Important residues and interactions

DNA-protein specific interactions

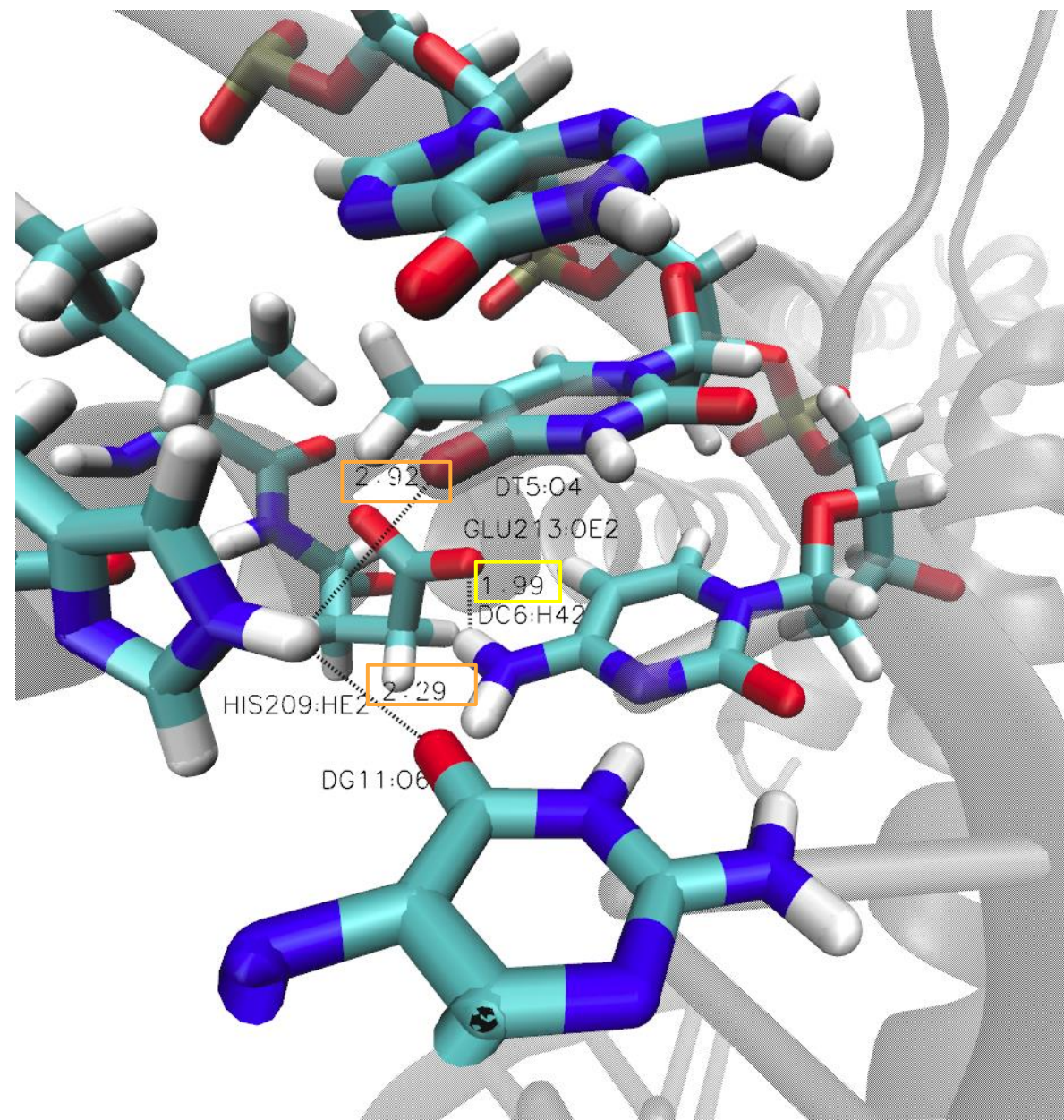
Hydrogen bonds		
His 209 (NH)	2,92 Å	DT5(O4)
	2,29 Å	DG11(O6)
Glu 213 (COO ⁻)	1,99 Å	DC6 (H42)

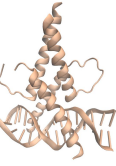


His 209



Glu 213





MITF

Important residues and interactions

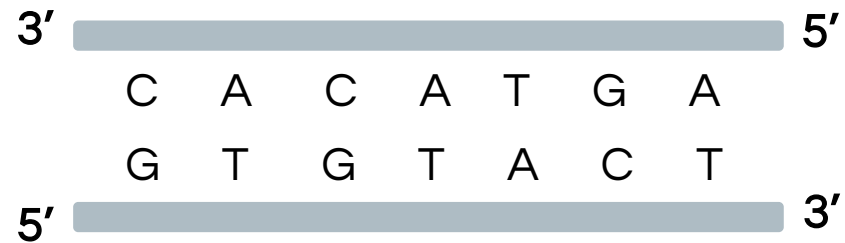
DNA-protein unspecific interactions

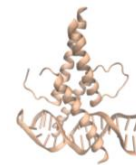
- MITF binds to 2 DNA motifs:

- the E-box

- the M-box:

(...) R R **R** R F N I **N** D R I K E (...) I **L** K A S V D Y I R K



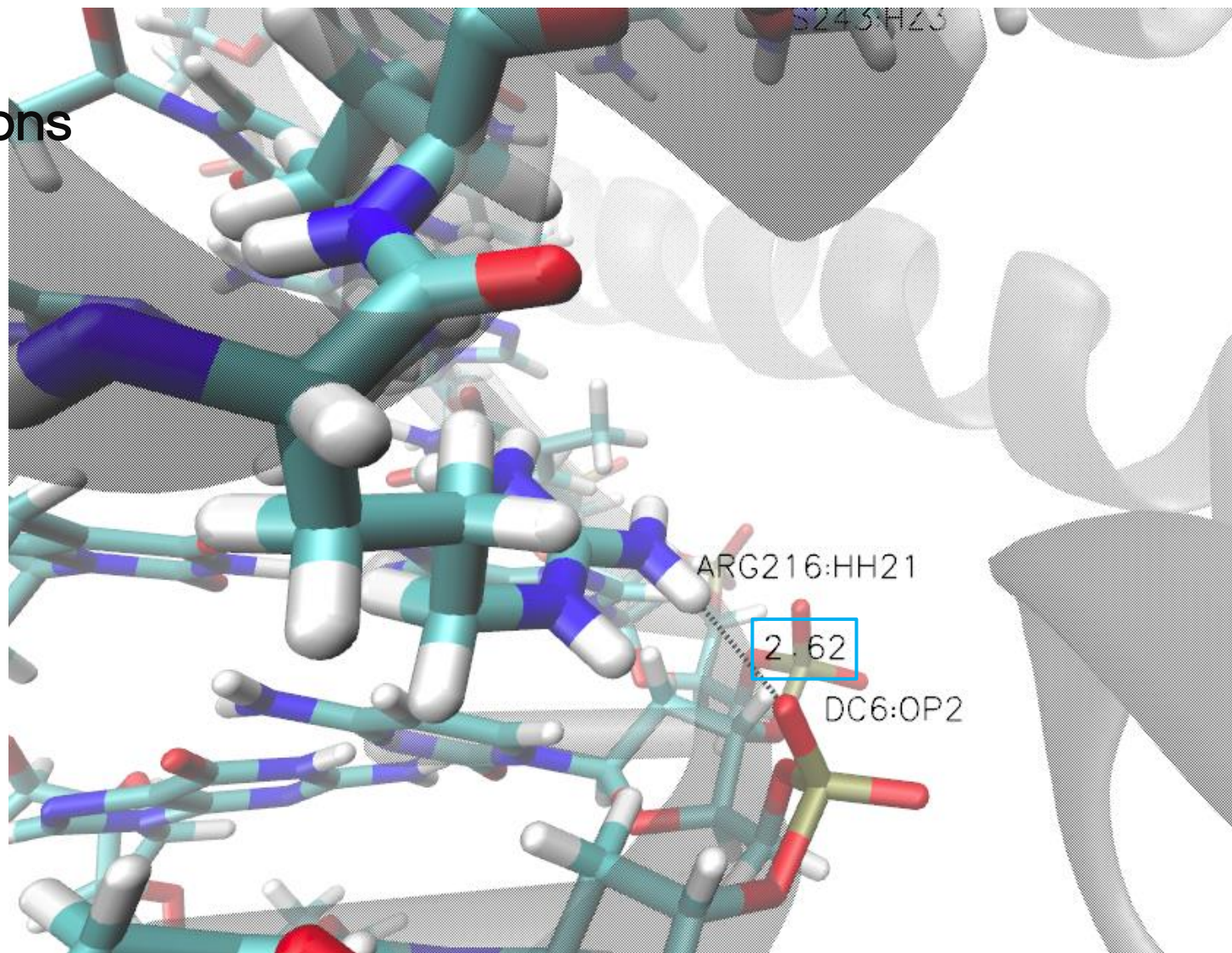
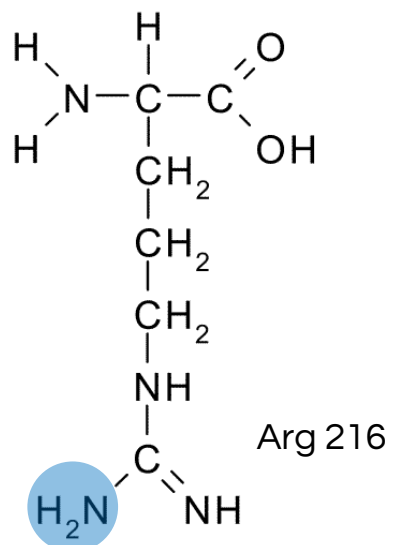


MITF

Important residues and interactions

DNA-protein unspecific interactions

Hydrogen bonds		
Arg 216 (NH ₂)	2,62 Å	DC6 (OP2)

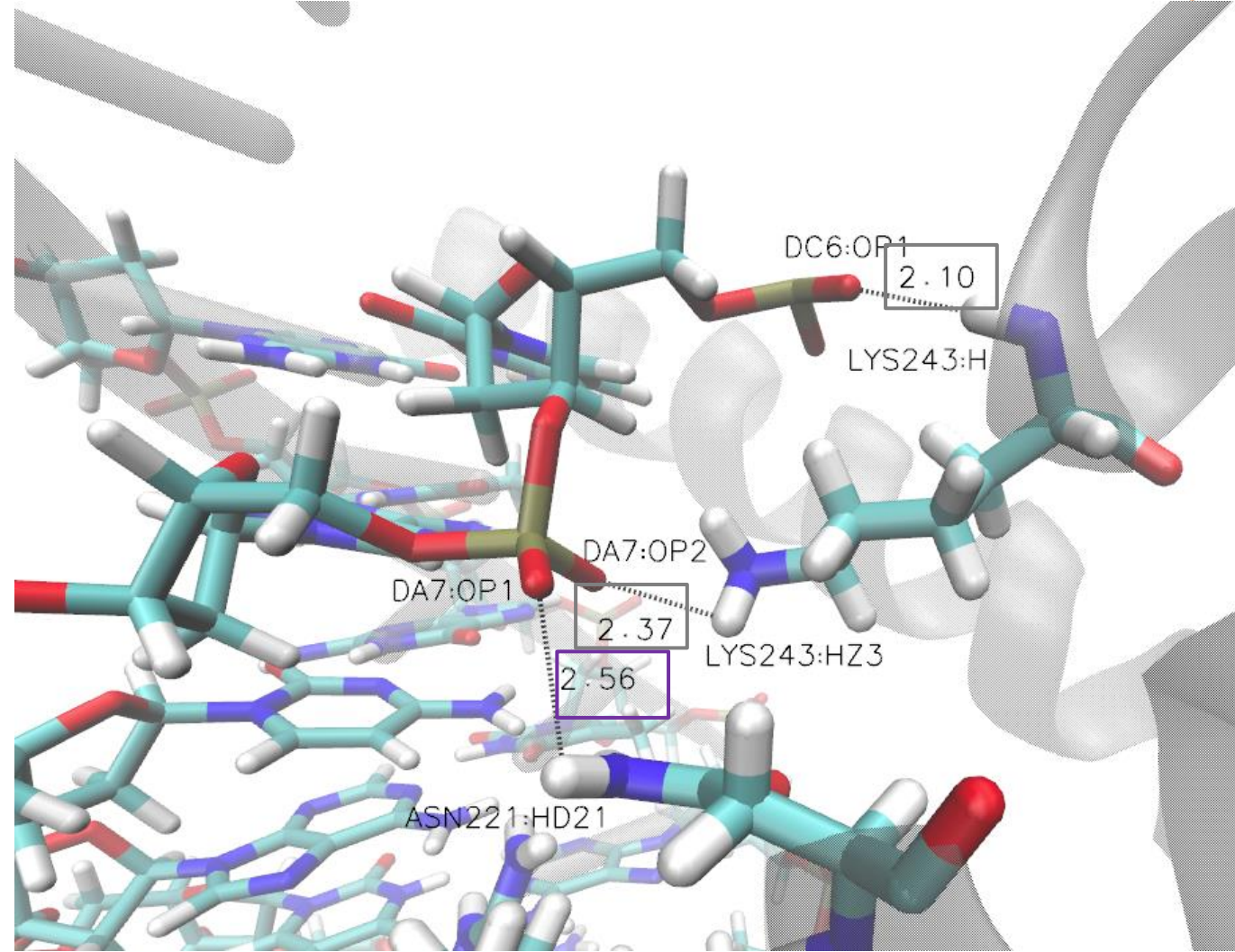
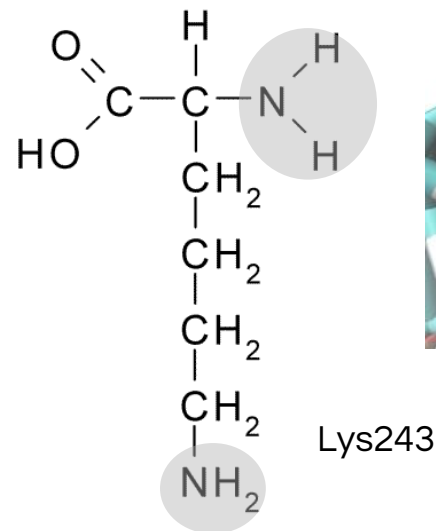
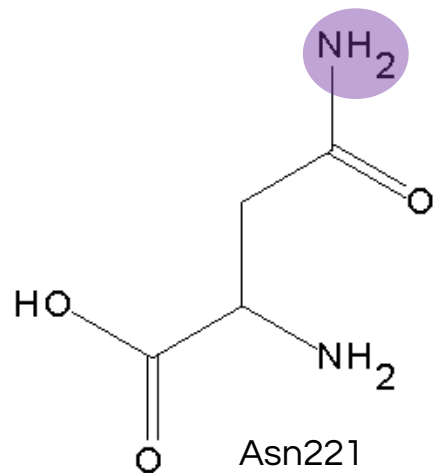


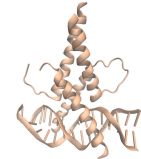
MITF

Important residues and interactions

DNA-protein unspecific interactions

Hydrogen bonds		
Asn 221 (NH ₂)	2,56 Å	DA7(OP1)
Lys 243 (NH ₂)	2,37 Å	DA7(OP2)
	2,10 Å	DC6(OP1)



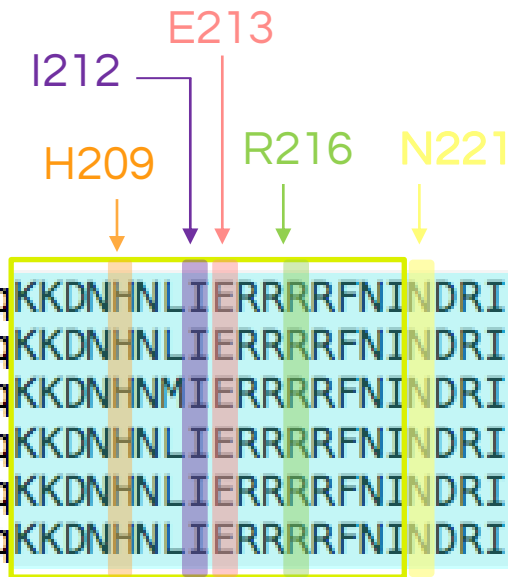


MITF

Sequence alignments

sp|Q08874|MITF_MMUSCULUS
sp|075030|MITF_HSAPIENS
tr|Q6WSQ9|MITF_DMELANOASTER
tr|073871|MITF_GGALLUS
tr|F6ZKQ4|MITF_ECABALLUS
tr|F1LQV3|MITF_RNORVEGICUS

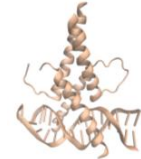
ppgltisnscpanlpnikreltacifptesealarakero
ppgltisnscpanlpnikreltacifptesealarakero
dilqndsfnfdknfnelsikqepqnltdaemnalakdro
gnqsmpppglnisnscpanlpnikreltesealarakero
ppgltisnscpanlpnikreltacifptesealarakero
ppgltisnscpanlpnikreltacifptesealarakero



sp|Q08874|MITF_MMUSCULUS
sp|075030|MITF_HSAPIENS
tr|Q6WSQ9|MITF_DMELANOASTER
tr|073871|MITF_GGALLUS
tr|F6ZKQ4|MITF_ECABALLUS
tr|F1LQV3|MITF_RNORVEGICUS

KELGTLIPKSN-----DPDMRWKGTILKASVDYIRKLQreqqrakdlenrqkklehan
KELGTLIPKSN-----DPDMRWKGTILKASVDYIRKLQreqqrakelenrqkklehan
KELGTLIPKGSdafyevVRDIRPNKGTILKSSVDYIKCL-khevtrlrqnelrqrqvelq
KELGTLIPKSN-----DPDMRWKGTILKASVDYIRKLQreqqrakelenrqkklehan
KELGTLIPKSN-----DPDMRWKGTILKASVDYIRKLQreqqrakelenrqkklehan
KELGTLIPKSN-----DPDMRWKGTILKASVDYIRKLQreqqrakdlenrqkklehan

↑
K243



MITF

Mutations and health implications

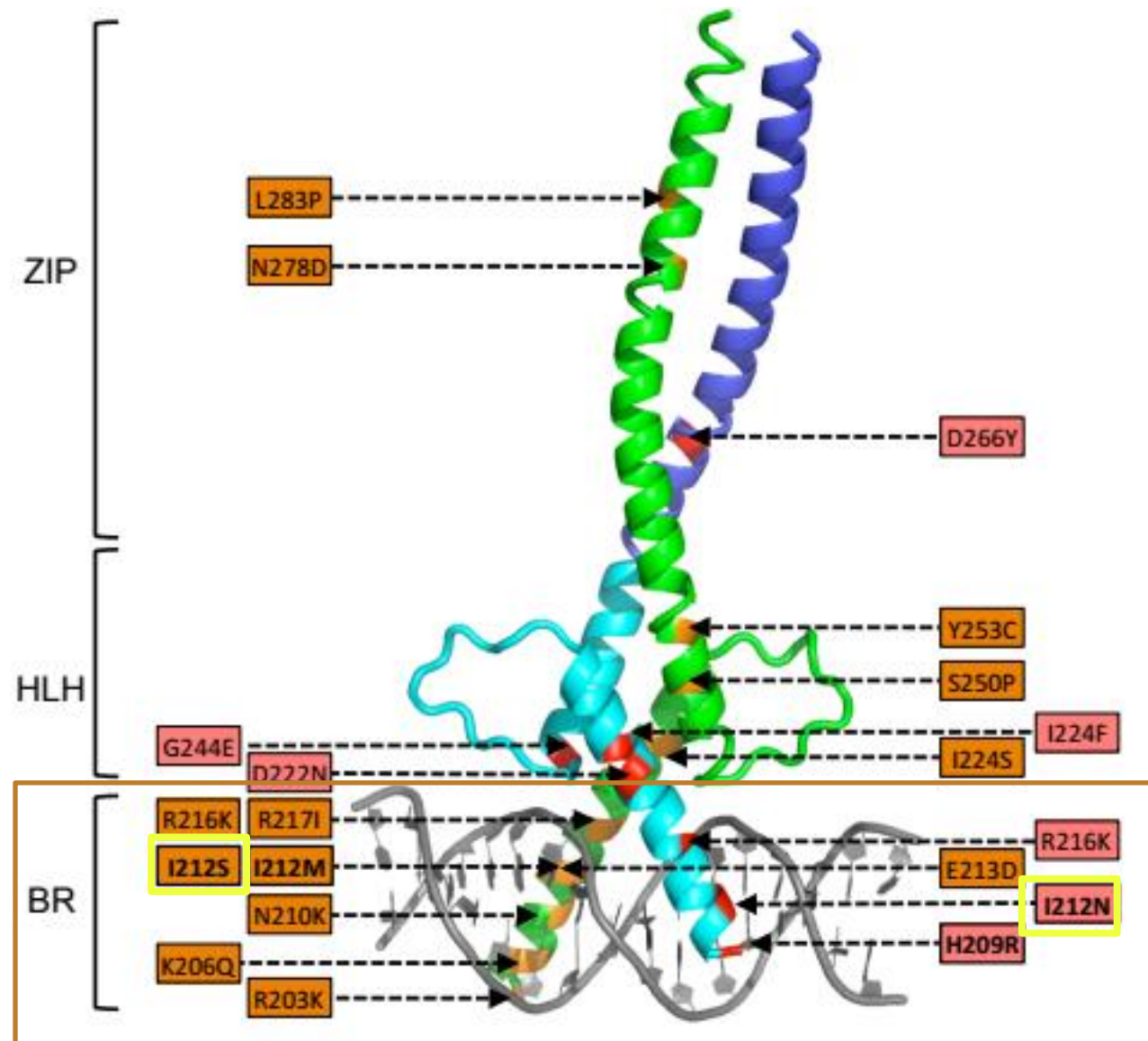
WSA2

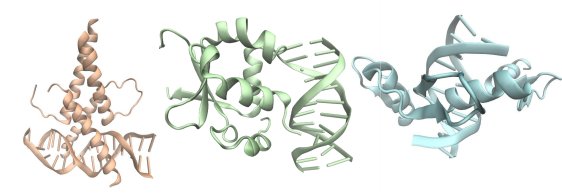
TADS

Mutations on the BR

I212S
I212N

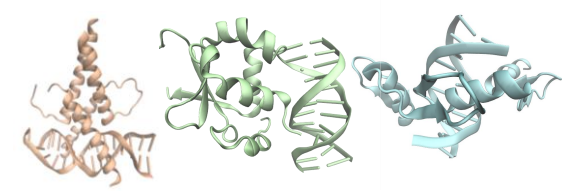
(non polar → polar)





Conclusions

- The most important **residues** for the **DNA interaction** are the **most conserved** in both, structure and sequence alignments.
- A **mutation** in one of these residues can **change** the **overall conformation** of the interactions.
- The **structure of the proteins change** when they are **bond** to the **DNA** and in some cases it also changes when they are bound to **methyalted DNA**.
- **Mutations on the DNA binding domain** can lead to **serious health issues**.



PEM questions

1. Which of these DNA binding motifs is found in humans?

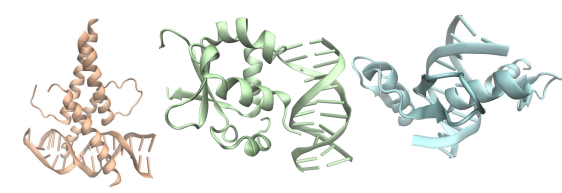
- a) Helix-loop-helix.
- b) Leucine zippers.
- c) Helix-turn-helix.
- d) Zinc fingers.
- e) All of them.

2. Characteristics of the zinc fingers:

- a) They are bind to a zinc ion for more stability.
- b) They are mainly formed by two β sheets and one α helix.
- c) Both are correct.
- d) The β sheets are antiparallel.
- e) All are correct.

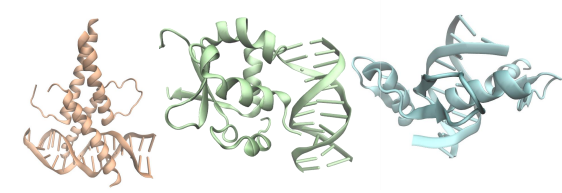
3. About inespecific interactions, between which structures and residues occur?

- a) DNA side chain – Protein side chain.
- b) DNA backbone – DNA side chain.
- c) DNA backbone – Protein backbone.
- d) DNA backbone – Protein side chain.
- e) None of them.



PEM questions

4. The specific interaction between HSF1 and the DNA is formed by the union of some aminoacids and:
- a) Cytosines and adenines.
 - b) Only adenines.
 - c) Cytosines and guanines.
 - d) Two thymines and one guanosine.**
 - e) None of them.
5. The transcription factor KLF4 interacts with the:
- a) Minor groove of the DNA.
 - b) Major groove of the DNA.**
 - c) Both are correct.
 - d) An specific sequence of the minor groove.
 - e) All of them are correct.
6. Which feature makes the His416 in the first Zinc Finger of the KLF4 important for the DNA interaction?
- a) It forms an stacking interaction with its aromatic ring and the one from the Thymine 7 base pair.
 - b) The stacking interaction perturbs the union between the T and the A.
 - c) Both are correct.
 - d) Histidine also interacts with the G, which interacts with the C of the complementary strand, forming a super base pair.
 - e) All are correct.**



PEM questions

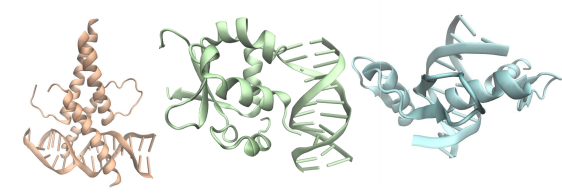
7. Helix-loop-Helix transcription factors are important in:

- 1) Neurogenesis.
 - 2) Myogenesis.
 - 3) Hematopoiesis.
 - 4) Pancreatic development.
- a) 1, 2 and 3.
 - b) 2 and 4.
 - c) 1 and 3.
 - d) 4.

e) 1, 2, 3 and 4.

8. The Heat shock factor 1 DNA binding domain is formed for:

- a) Three helix-alpha.
 - b) Two helix 3/10.
 - c) Four beta-strands.
- d) All of them.**
 - e) None of them.



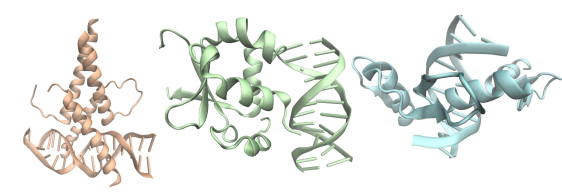
PEM questions

9. Choose the correct answer:

- a) MITF is an bHLHzip transcription factor.
- b) MITF can specifically bind to the E-box or the M-box.
- c) The two statements above are correct.**
- d) MITF interacts with DNA mostly through residues on the loop.
- e) All the statements are correct.

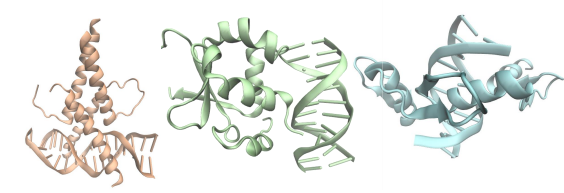
10. Regarding the M-box of the MITF factor, choose the correct statements:

- 1) Unlike the canonical E-box, M-box isn't a palindromic motif.
 - 2) All bHLH transcription factors have preference for binding to the M-box.
 - 3) In MITF, Isoleucine 212 is a key residue for specific binding to the M-box.
 - 4) Isoleucine 212 interacts with an adenine residue through a hydrogen bond.
- a) 1, 2 and 3.
 - b) 1 and 3.**
 - c) 3 and 4.
 - d) 4.
 - e) 1, 2, 3 and 4.



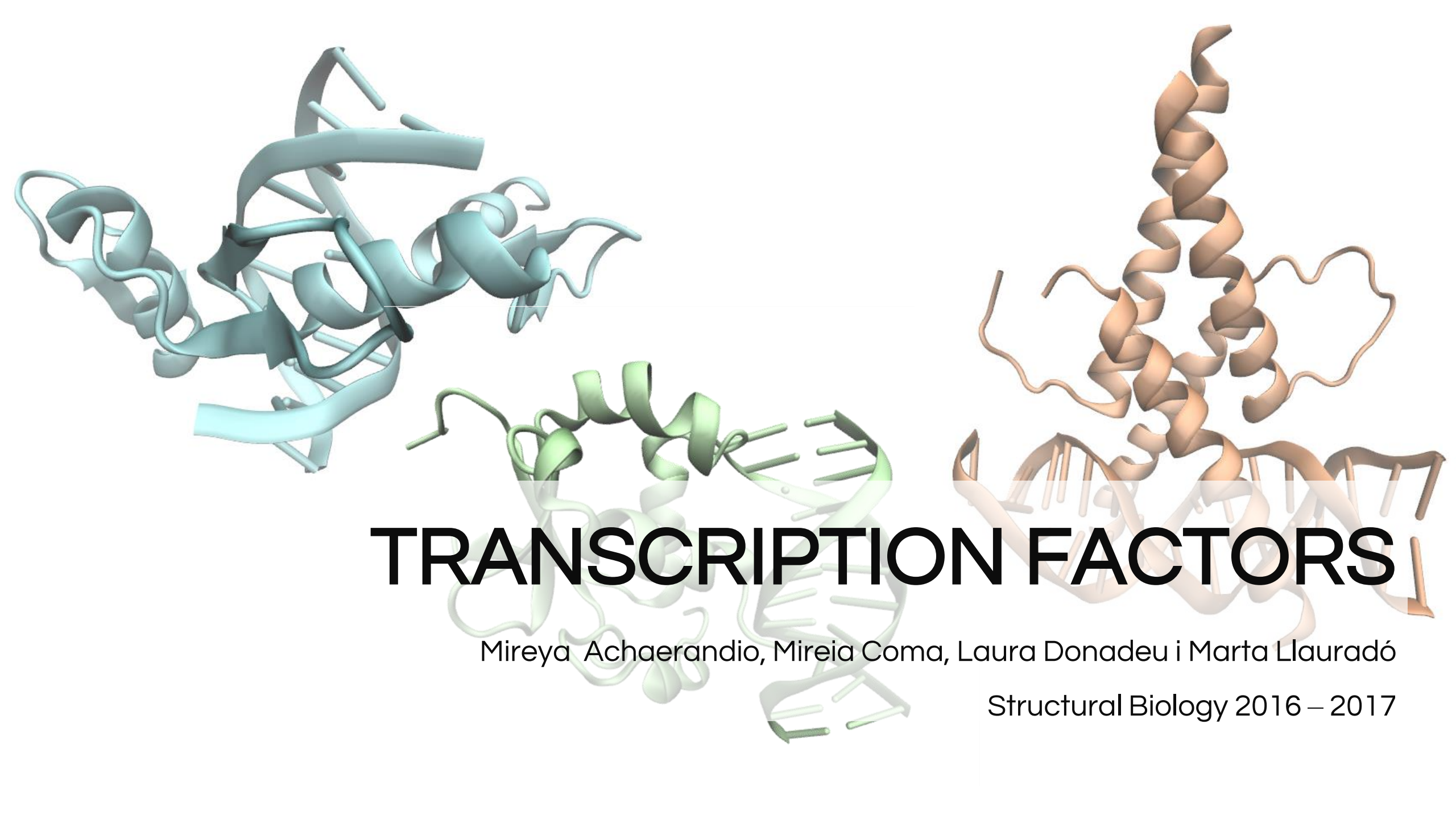
Bibliography

- Harrison C, Bohm A, Nelson H. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science*. 1994;263(5144):224-227.
- Harrison S, Aggarwal A. DNA Recognition by Proteins with the Helix-Turn-Helix Motif. *Annual Review of Biochemistry*. 1990;59(1):933-969.
- Hashimoto H, Wang D, Steves A, Jin P, Blumenthal R, Zhang X et al. Distinctive Klf4 mutants determine preference for DNA methylation status. *Nucleic Acids Research*. 2016;774.
- Hartman M, Czyz M. MITF in melanoma: mechanisms behind its expression and activity. *Cell Mol Life Sci*. 2014; 72(7):1249-1260.
- Jones S. An overview of the basic helix-loop-helix proteins. *Genome Biol*. 2004; 5(6): 226.
- Laity J, Dyson HJ and Wright P. DNA-induced α -Helix Capping in Conserved Linker Sequences is a Determinant of Binding Affinity in Cys2-His2 Zinc Fingers. *Mol. Biol*. 2000 (295):719-727.
- Liu Y, Olanrewaju Y, Zheng Y, Hashimoto H, Blumenthal R, Zhang X et al. Structural basis for Klf4 recognition of methylated DNA. *Nucleic Acids Research*. 2014;42(8):4859-4867.7
- Massari ME, Murre C. Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms. *Mol Cell Biol*. 2000; 20(2): 429-440.



Bibliography

- Neudegger T, Verghese J, Hayer-Hartl M, Hartl F, Bracher A. Structure of human heat-shock transcription factor 1 in complex with DNA. *Nature Structural & Molecular Biology*. 2016;23(2):140-146.
- Pearson R, Fleetwood J, Eaton S, Crossley M, Bao S. Krüppel-like transcription factors: A functional family. *International Journal of Biochemistry*. 2008(40):1996-2001.
- Pogenberg V, Ogmundsdottir M, Bergsteinsdottir K, Schepsky A, Phung B, Deineko V et al. Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. *Genes & Development*. 2012;26(23):2647-2658.
- Schuetz A, Nana D, Rose C, Zocher G, Milanovic M, Koenigsmann J et al. The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cellular and Molecular Life Sciences*. 2011;68(18):3121-3131.
- Tassabehji M, Newton V, Read A. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat Genet*. 1994; 8(3): 251-255.
- Voss THager G. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics*. 2013;15(2):69-81.
- Vangapandu H, Ai W. Kruppel like factor 4 (KLF4): a transcription factor with diverse context-dependent functions. *Gene Ther Mol Biol*/2009(13):194-203. Schuetz A, Nana D, Rose C, Zocher G, Milanovic M, Koenigsmann J, et al. The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cell. Mol. Life Sci*. 2011(68):3121–3131.



TRANSCRIPTION FACTORS

Mireya Achaerandio, Mireia Coma, Laura Donadeu i Marta Llauradó

Structural Biology 2016 – 2017