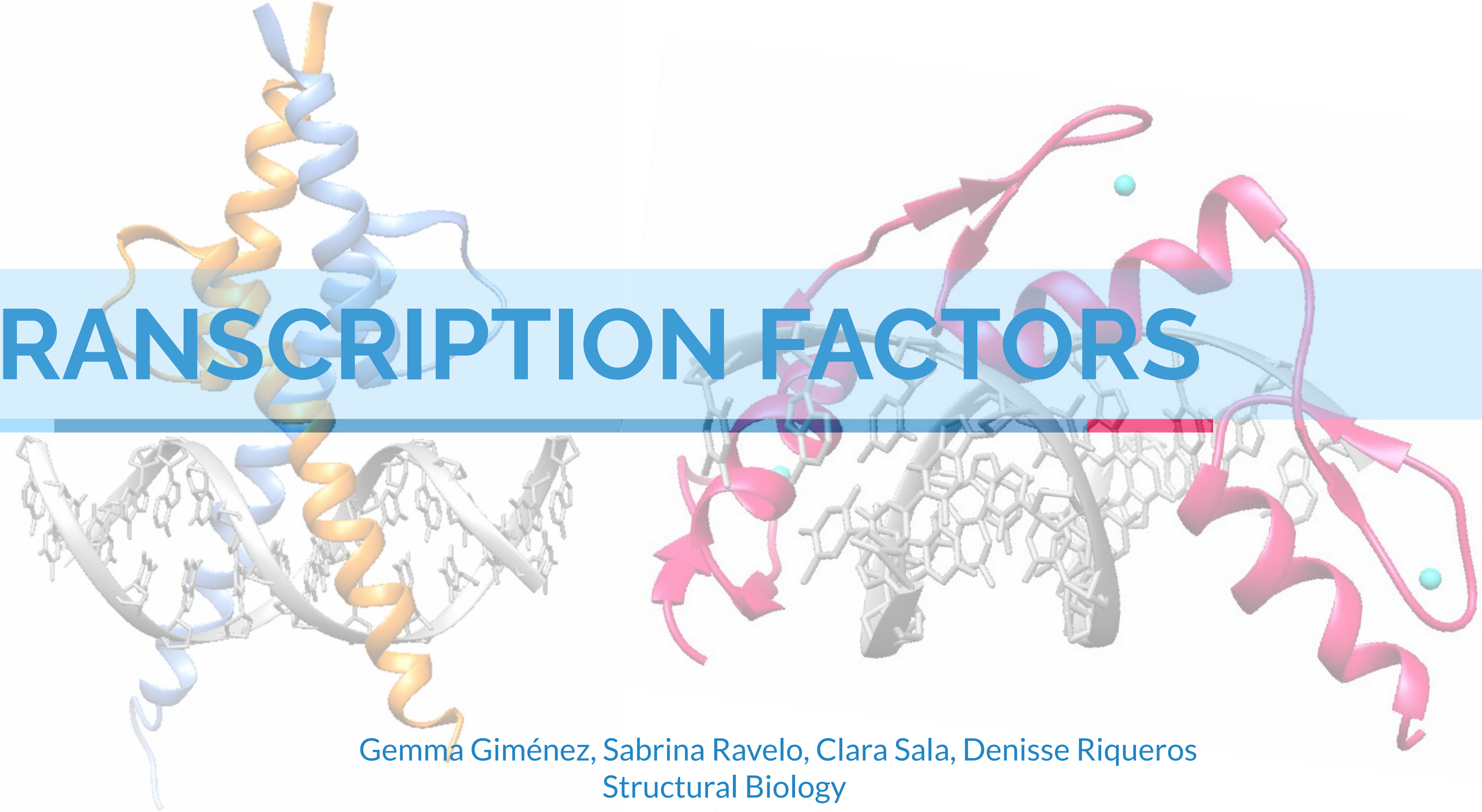


TRANSCRIPTION FACTORS

Gemma Giménez, Sabrina Ravelo, Clara Sala, Denisse Riqueros
Structural Biology
March 2022

Introduction

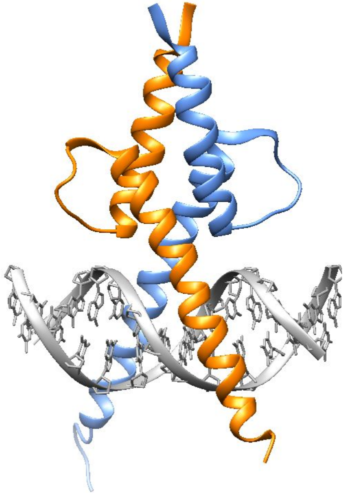
Transcription factors are proteins that **bind to DNA-regulatory sequences**, localized in the 5-upstream region of target genes in order to promote, or block the recruitment of RNA polymerase.

This may result in **increased or decreased gene transcription**, protein synthesis, and subsequent altered cellular function.

Proteins

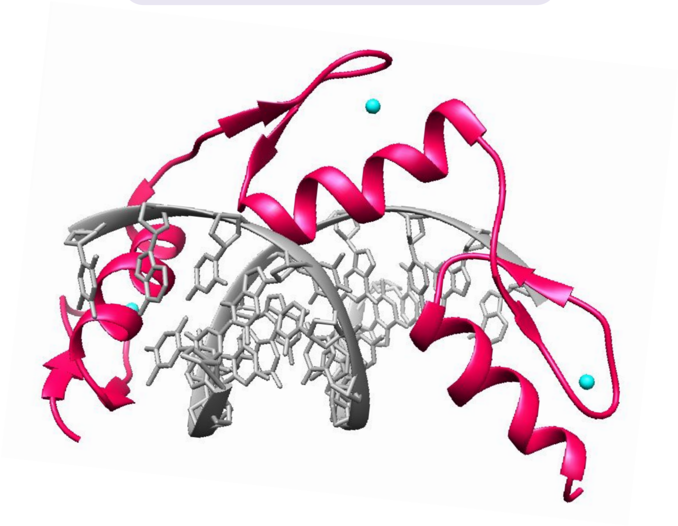
BASIC HELIX-LOOP-HELIX DOMAIN (b-HLH)

MyoD



ZINC-COORDINATING DNA BINDING DOMAIN

Zif268



1.

Helix-loop-helix

Helix-loop-helix

The bHLH is a protein structural motif that characterizes one of the largest families of **dimerizing transcription factors**. They are able to form **homodimers** or **heterodimers**.

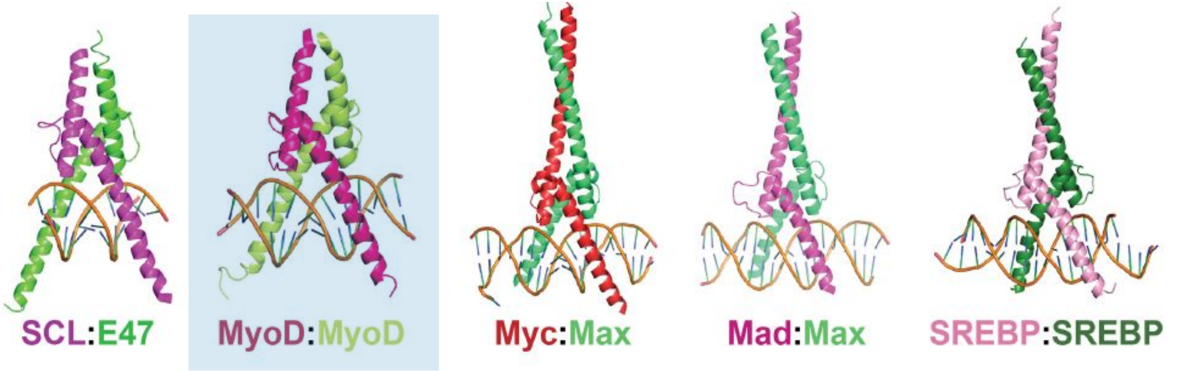
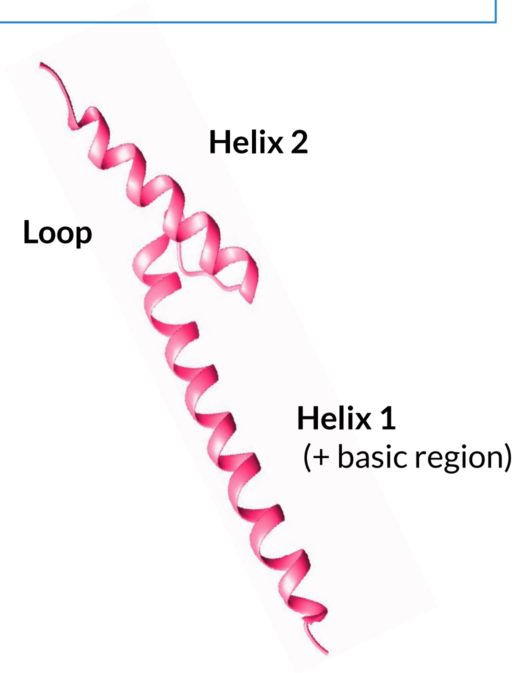


Figure 1. Murre C. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* 2019 Jan 1;33(1-2):6-25.



The bHLH domains fold into a **parallel**, **left-handed, four-helix bundle** with a well-defined **hydrophobic core,** that lies just above the DNA.

E-box: **CANNTG**

MyoD

Biological function

Myoblast determination protein 1 (MyoD) is part of the myogenic regulatory factor (MRF) family of TFs that control the determination and differentiation of skeletal muscle cells

FUNCTION

- **Master regulatory factor → enables myoblast proliferation during development**
- **Reactivation during muscle regeneration (exercise or tissue damage)**

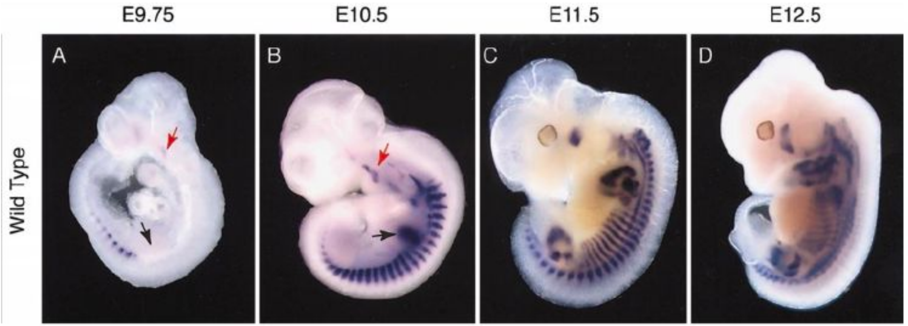


Figure 2. Whole-mount in situ hybridization for MyoD mRNA in wild-type in mice embryos. Chen JC, Ramachandran R, Goldhamer DJ. Essential and redundant functions of the MyoD distal regulatory region revealed by targeted mutagenesis. Dev Biol. 2002 May 1;245(1):213-23.

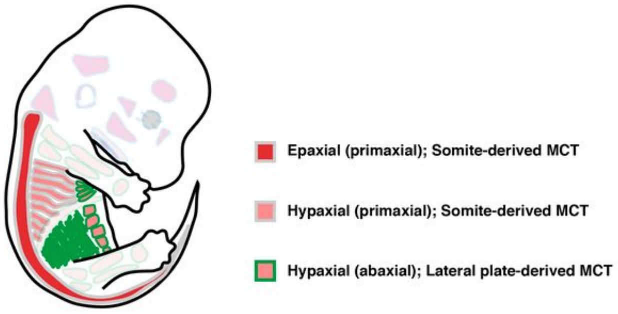


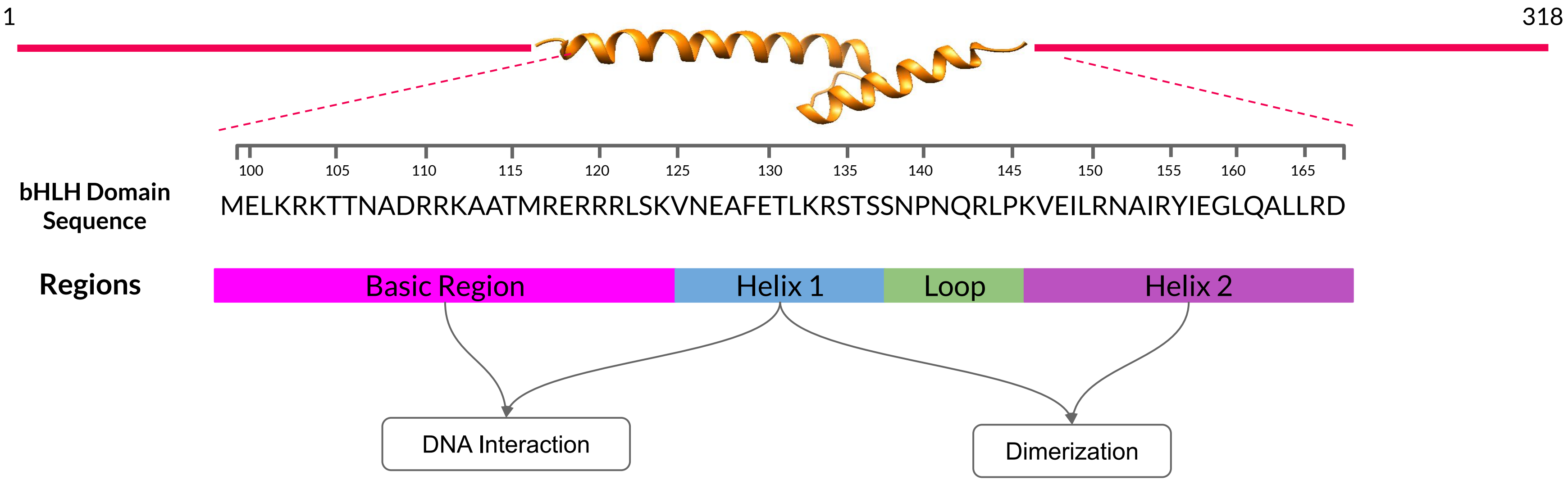
Figure 3. Development of axial muscles in tetrapods. Sefton EM, Kardon G. Connecting muscle development, birth defects, and evolution: An essential role for muscle connective tissue. Curr Top Dev Biol. 2019;132:137-176.

SCOP classification

Class	All alpha proteins
Fold	HLH-like *4-helices; bundle, closed, left-handed twist; 2 crossover connections*
Superfamily	HLH, helix-loop-helix DNA-binding domain
Family	HLH, helix-loop-helix DNA-binding domain
Protein	Myod B/HLH domain
Species	Mouse (Mus musculus)

General structure

Sequence and regions



bHLH Domain Sequence

MELKRKTTNADRRKAATMRERRRLSKVNEAFETLKRSTSSNPNQRLPKVEILRNAIRYIEGLQALLRD

Regions

Basic Region | Helix 1 | Loop | Helix 2

Basic Region → DNA Interaction

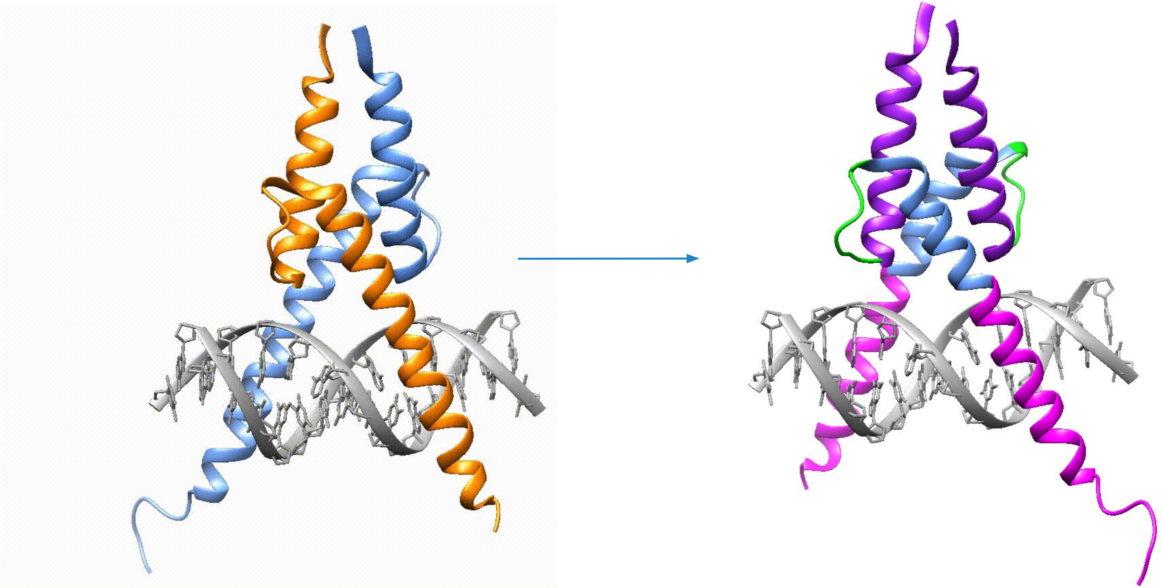
Helix 1 → DNA Interaction

Helix 1 → Dimerization

Helix 2 → Dimerization

General structure

3D structure



MyoD

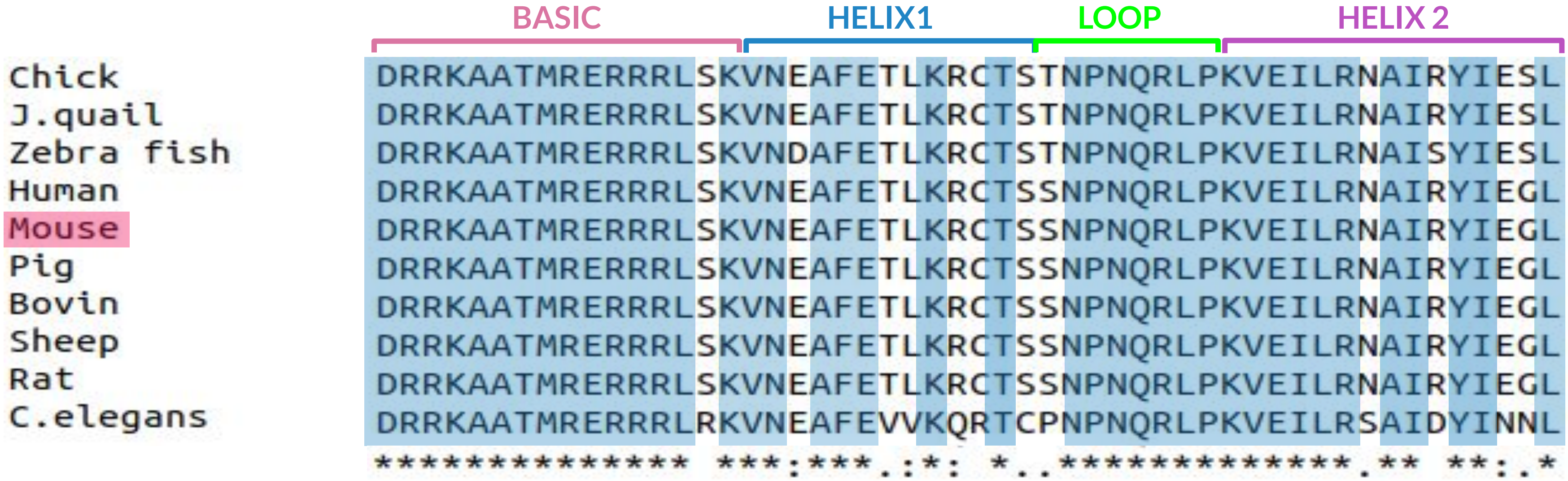
- Basic region
- Helix 1
- Helix 2
- Loop

Sequence and structure analysis

MSA: Evolution among different species

Methodology and multiple sequence alignment results

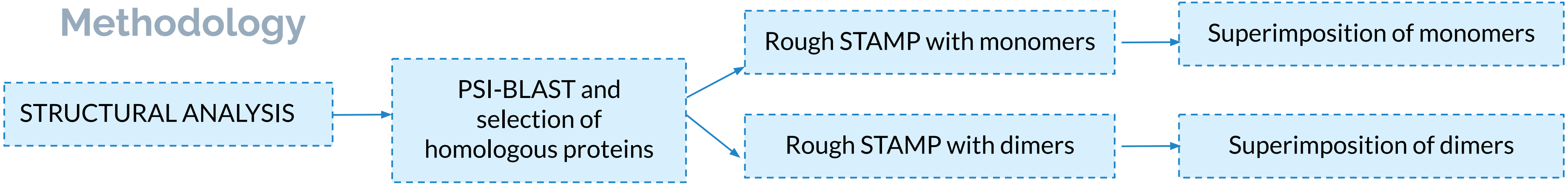
SEQUENCE ANALYSIS → PSI-BLAST and selection of homologous proteins → MSA alignment using ClustalW



```
Chick          DRRKAATMRERRRLSKVNEAFETLKRCTSTNPNQRLPKVEILRNAIRYIESL
J.quail        DRRKAATMRERRRLSKVNEAFETLKRCTSTNPNQRLPKVEILRNAIRYIESL
Zebra fish     DRRKAATMRERRRLSKVNDAFETLKRCTSTNPNQRLPKVEILRNAISYIESL
Human          DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
Mouse          DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
Pig            DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
Bovin          DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
Sheep          DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
Rat            DRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGL
C.elegans      DRRKAATMRERRRLRKVNEAFEVVKQRTCPNPNQRLPKVEILRSAIDYINNL
               ************** ***:***.:*: *..*************.** **:.*
```

Structural alignment: bHLH family

Methodology



STRUCTURAL ANALYSIS

PSI-BLAST and selection of homologous proteins

Rough STAMP with monomers

Superimposition of monomers

Rough STAMP with dimers

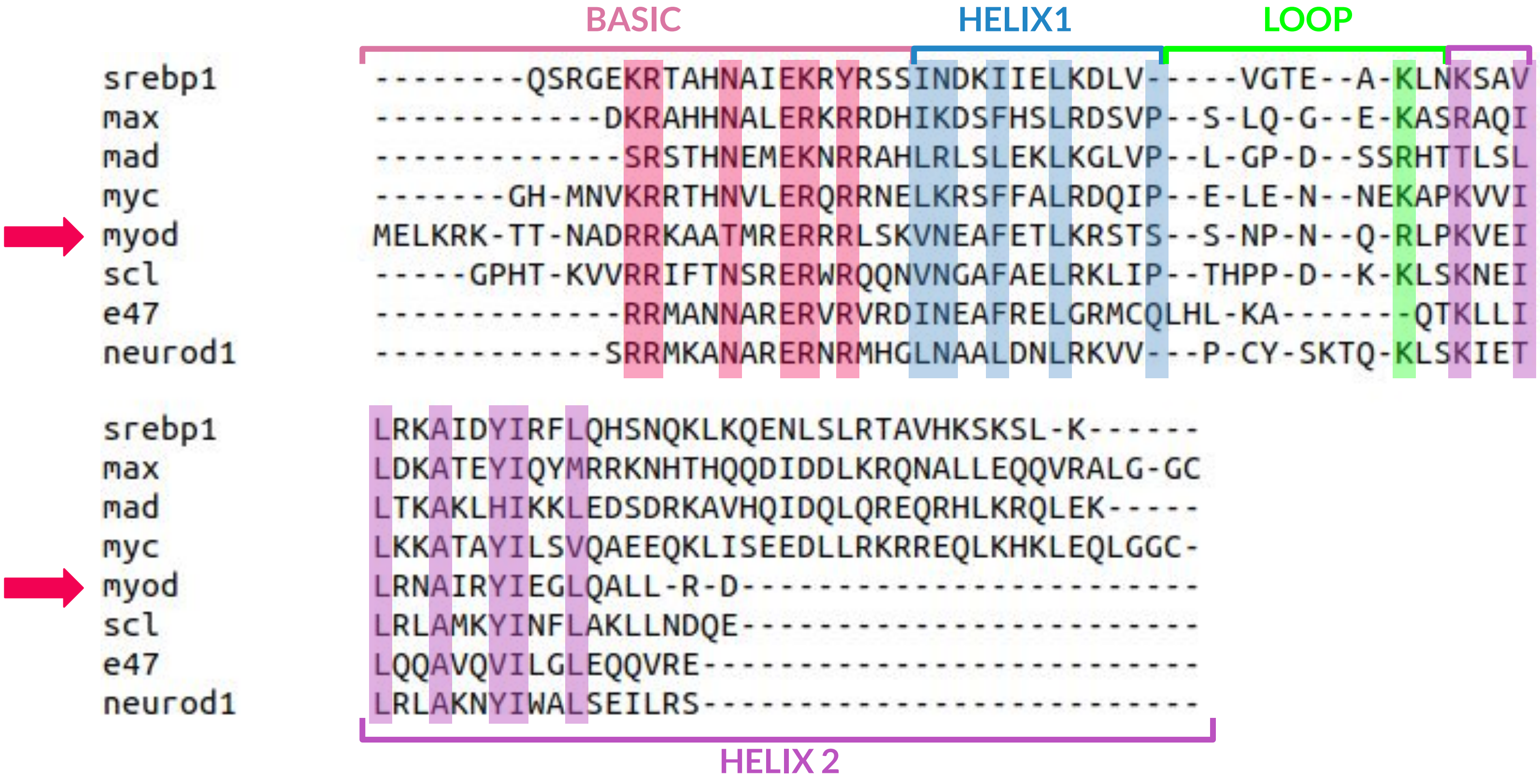
Superimposition of dimers

PDB ID	Resolution	Protein	Length
1MDY	2.80 Å	MyoD	68
		MyoD	62
1AM9	2.30 Å	SREBP1	82
		SREBP1	82
1NKP	1.80 Å	Myc	88
		Max	83

PDB ID	Resolution	Protein	Length
1NLW	2.00 Å	Mad	80
		Max	76
2YPB	2.87 Å	SCL	91
		E47	82
2QL2	2.50 Å	NeuroD1	60
		E47	60

Same family structures

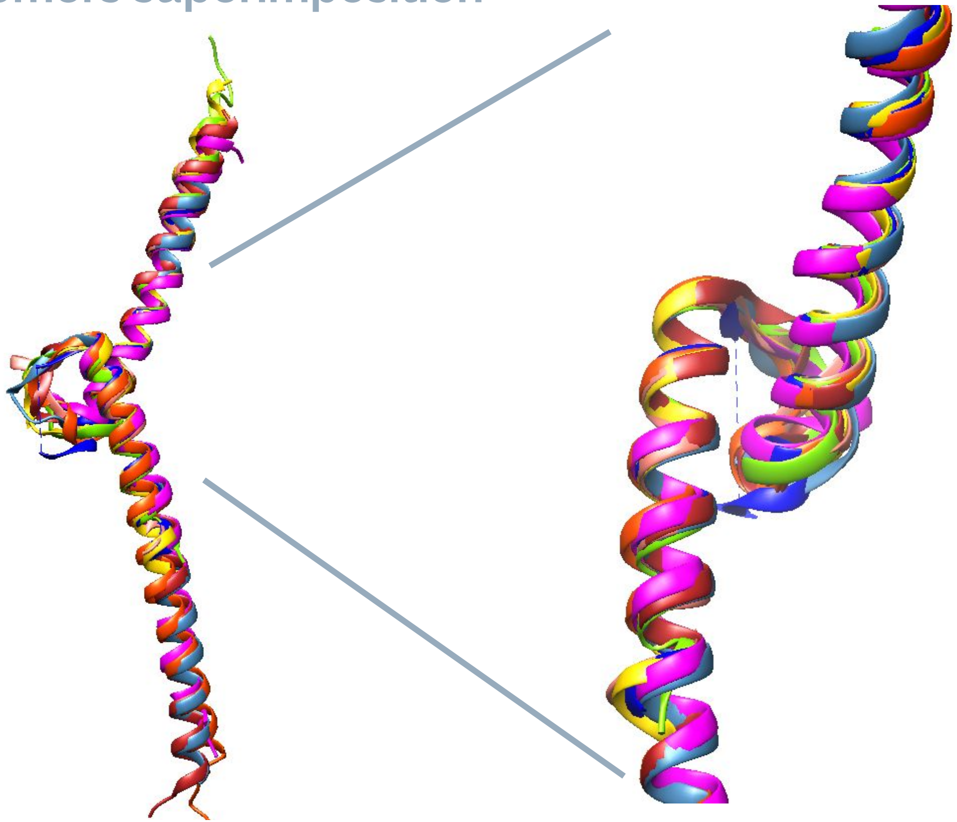
Structural alignment (monomers)



```
srebp1     --------QSRGEKRTAHNAIEKRYRSSINDKIIELKDLV-----VGTE--A-KLNKSAV
max        ------------DKRAHHNALERKRRDHIKDSFHSLRDSVP--S-LQ-G--E-KASRAQI
mad        -------------SRSTHNEMEKNRRAHLRLSLEKLKGLVP--L-GP-D--SSRHTTLSL
myc        -------GH-MNVKRRTHNVLERQRRNELKRSFFALRDQIP--E-LE-N--NEKAPKVVI
myod       MELKRK-TT-NADRRKAATMRERRRLSKVNEAFETLKRSTS--S-NP-N--Q-RLPKVEI
scl        -----GPHT-KVVRRIFTNSRERWRQQNVNGAFAELRKLIP--THPP-D--K-KLSKNEI
e47        -------------RRMANNARERVRVRDINEAFRELGRMCQLHL-KA-------QTKLLI
neurod1    ------------SRRMKANARERNRMHGLNAALDNLRKVV---P-CY-SKTQ-KLSKIET

srebp1     LRKAIDYIRFLQHSNQKLKQENLSLRTAVHKSKSL-K------
max        LDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALG-GC
mad        LTKAKLHIKKLEDSDRKAVHQIDQLQREQRHLKRQLEK-----
myc        LKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC-
myod       LRNAIRYIEGLQALL-R-D------------------------
scl        LRLAMKYINFLAKLLNDQE------------------------
e47        LQQAVQVILGLEQQVRE--------------------------
neurod1    LRLAKNYIWALSEILRS--------------------------
```

bHLH family structures

Monomers superimposition

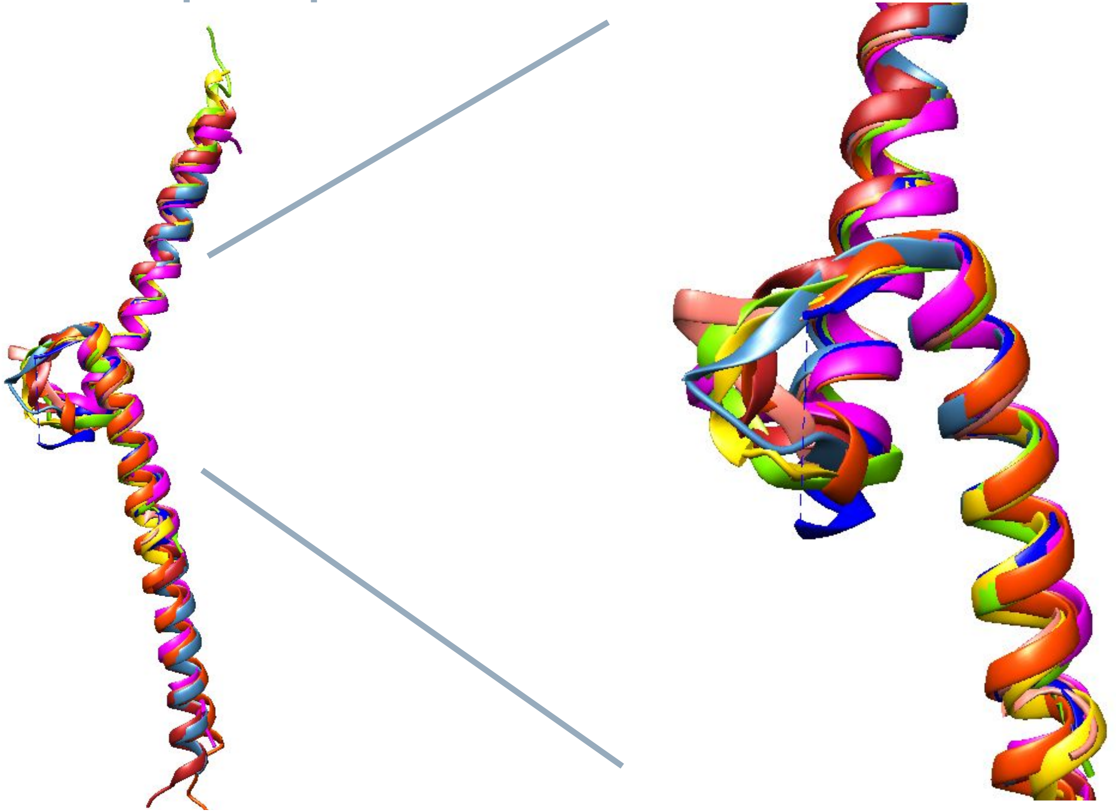


Sc = **6.74** RMS = **0.47** Len = **72** nfit = **49**

- ● SREBP1
- ● Myc
- ● NeuroD1
- ● Mad
- ● SCL
- ● MyoD
- ● Max
- ● E47

bHLH family structures

Monomers superimposition

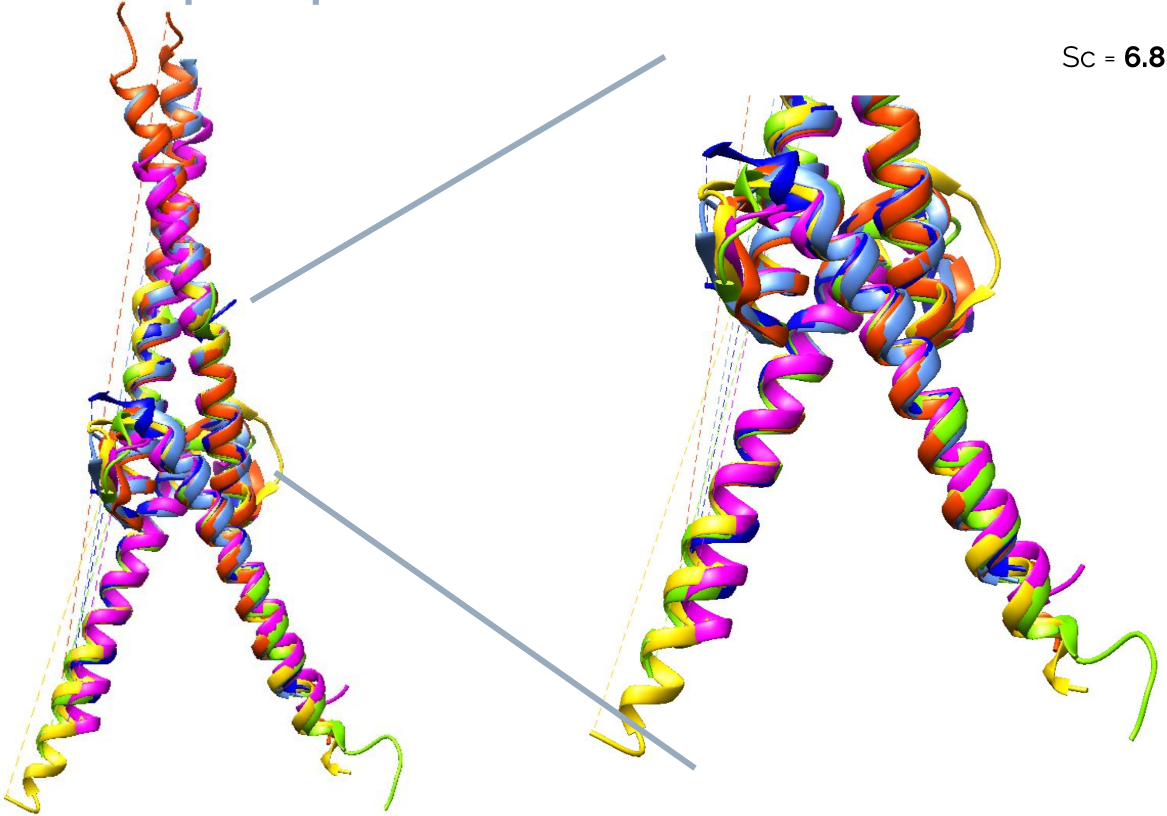


Sc = **6.74** RMS = **0.47** Len = **72** nfit = **49**

- ● SREBP1
- ● Myc
- ● NeuroD1
- ● Mad
- ● SCL
- ● MyoD
- ● Max
- ● E47

bHLH family structures

Dimers superimposition



Sc = **6.82** RMS = **0.84** Len = **173** nfit = **103**

- ● SREBP1-SREBP1
- ● Myc
- ● NeuroD1
- ● Mad-Max
- ● SCL-E47
- ● MyoD-MyoD
- ● Myc-max
- ● E47-NeuroD1

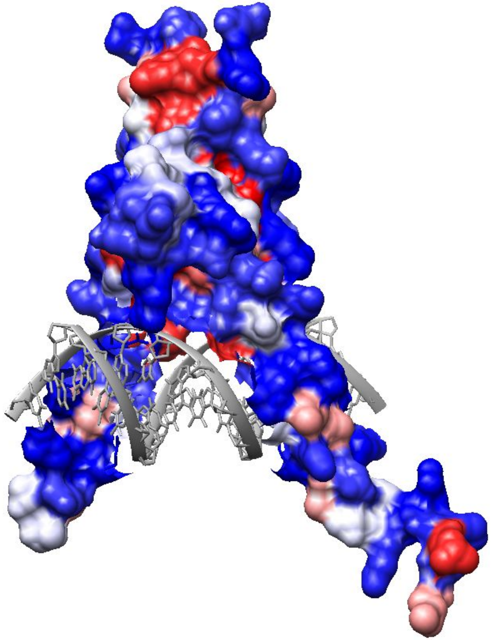
Important residues and interactions

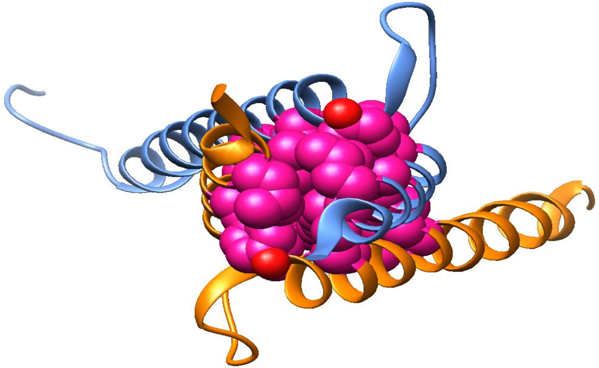
Dimer interactions

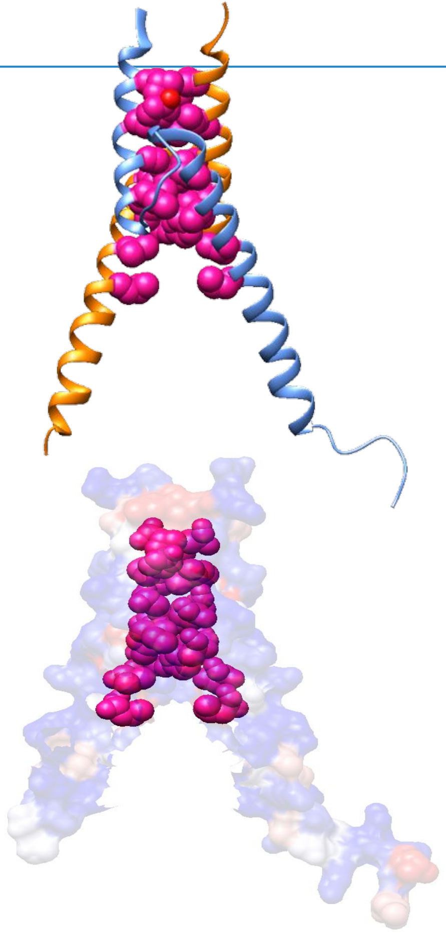
DNA-protein interactions

Important residues and interactions

Dimer interactions - Hydrophobic core





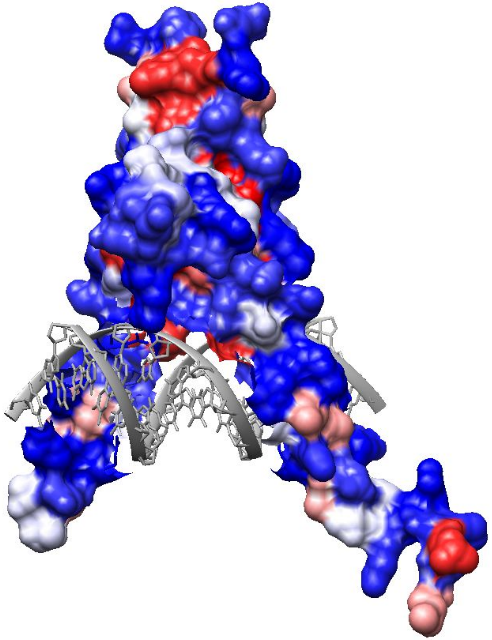


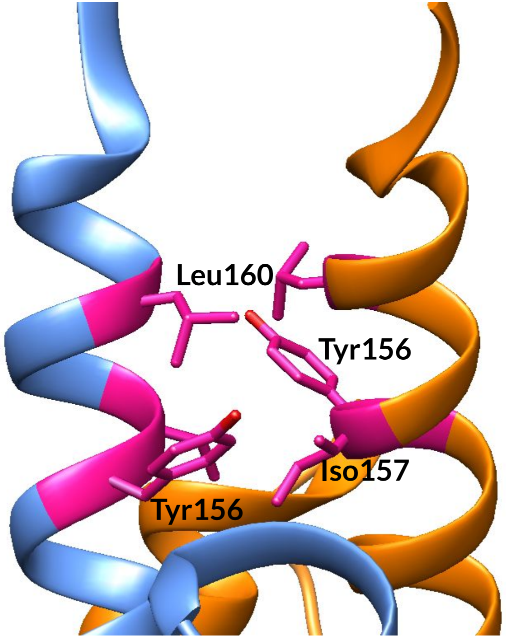
Hydrophilic Hydrophobic

🟦🟦⬜🟥🟥

Important residues and interactions

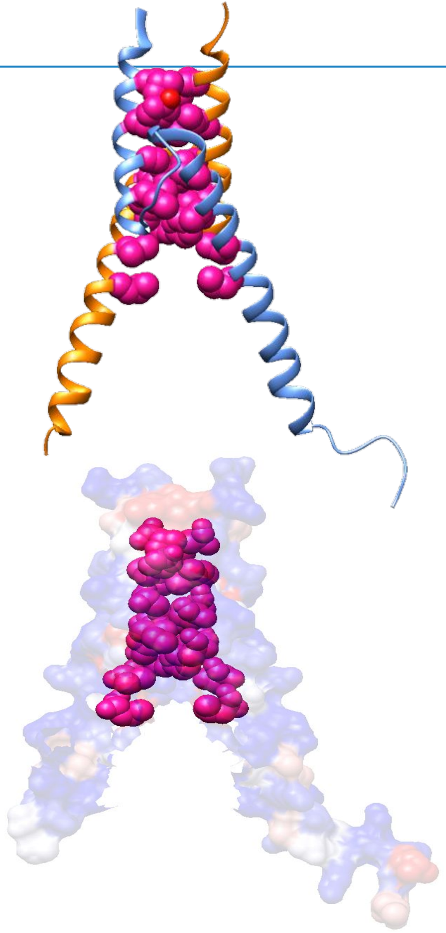
Dimer interactions - Hydrophobic core





Hydrophilic Hydrophobic

🟦🟦⬜🟥🟥



Important residues and interactions

Dimer interactions - Hydrophobic core

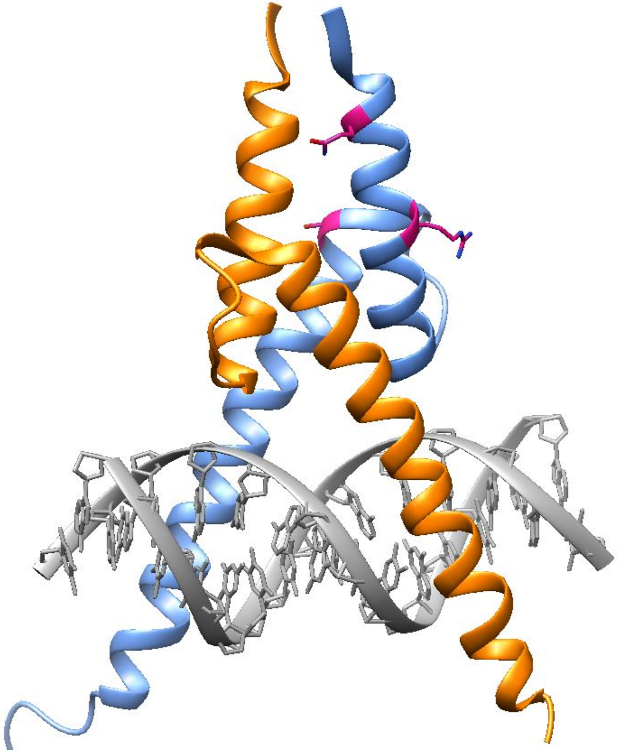
```
srebp1   INDKIIELKDLV KSAVLRKAIDYIRFLQHSNQKLKQENLSLRTAVHKSKSL-K------
max      IKDSFHSLRDSV RAQILDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALG-GC
mad      LRLSLEKLKGLV TLSLLTKAKLHIKKLEDSDRKAVHQIDQLQREQRHLKRQLEK-----
myc      LKRSFFALRDQI KVVILKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC-
myod     VNEAFETLKRST KVEILRNAIRYIEGLQALL-R-D------------------------
scl      VNGAFAELRKLI KNEILRLAMKYINFLAKLLNDQE------------------------
e47      INEAFRELGRMC KLLILQQAVQVILGLEQQVRE--------------------------
neurod1  LNAALDNLRKVV KIETLRLAKNYIWALSEILRS--------------------------
```

Val125 Phe129 Leu132

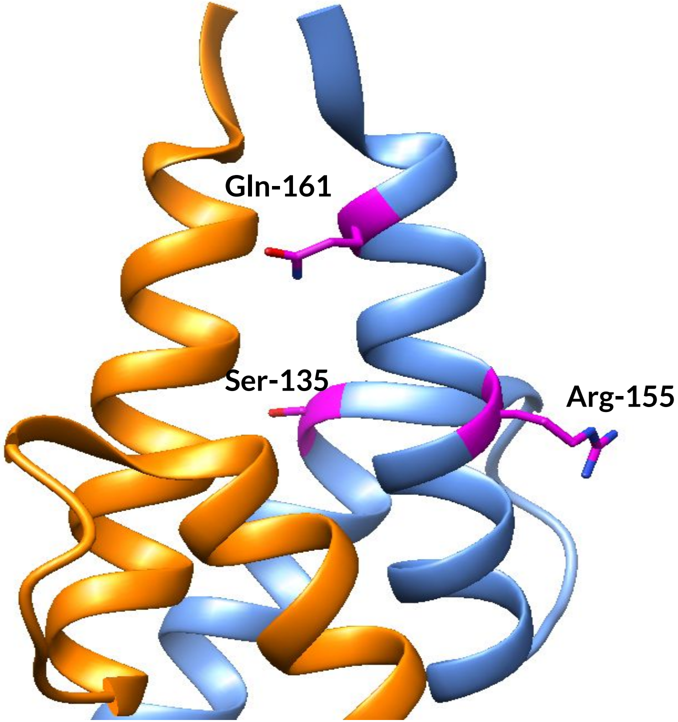
Ile149 Leu150 Ala153 Tyr156 Ile157 Leu160

Important residues and interactions

Dimer interactions - Dimerization specificity



Ser135
Arg155
Gln161



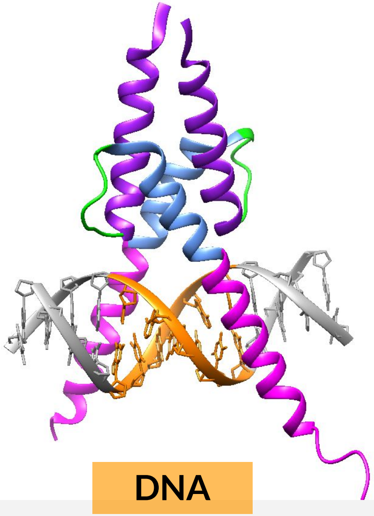
Important residues and interactions:

Dimer interactions - Dimerization specificity

```
srebp1    INDKIIELKDLV KSAV LRKAIDYIRFLQHSNQKLKQENLSLRTAVHKSKSL-K------
max       IKDSFHSLRDSV RAQI LDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALG-GC
mad       LRLSLEKLKGLV TLSL LTKAKLHIKKLEDSDRKAVHQIDQLQREQRHLKRQLEK-----
myc       LKRSFFALRDQI KVVI LKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC-
myod      VNEAFETLKRST KVEI LRNAIRYIEGLQALL-R-D----------------------
scl       VNGAFAELRKLI KNEI LRLAMKYINFLAKLLNDQE----------------------
e47       INEAFRELGRMC KLLI LQQAVQVILGLEQQVRE------------------------
neurod1   LNAALDNLRKVV KIET LRLAKNYIWALSEILRS------------------------
```

Dna interactions

Specific and non-specific interactions



DNA

5’ TCAACAGCTGTTGA 3’
3’ AGTTGTCGACAACT 5’

SPECIFIC INTERACTIONS

-Direct contact between the basic region of MyoD and the **DNA base pairs**.

-Hydrogen bonds or **van der Waals** interactions

Primary basis for **sequence selectivity**

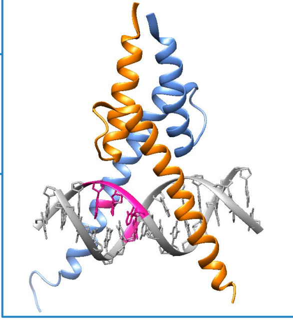
NON SPECIFIC INTERACTIONS

-Independent of nucleotide sequence.

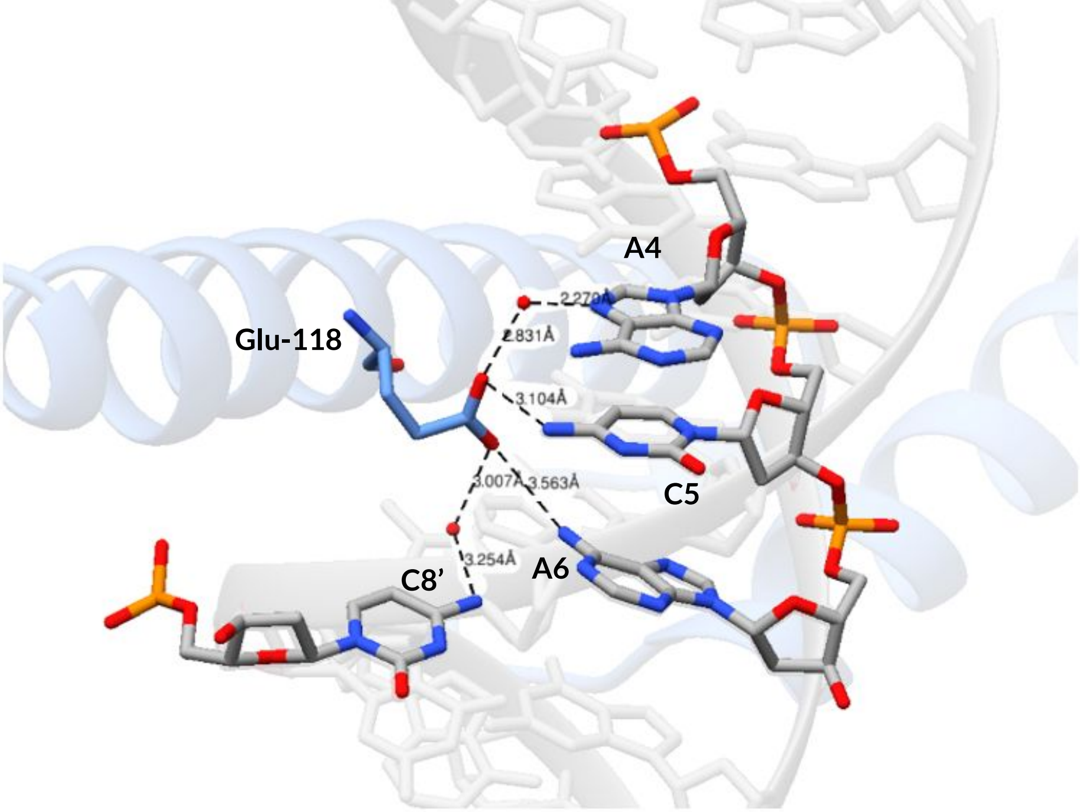
- **Electrostatic interactions** between basic residues and the sugar-phosphate backbone

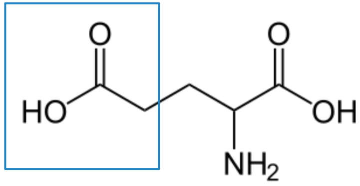
Important for **stability**

Important residues and interactions



DNA-protein interactions: SPECIFICITY

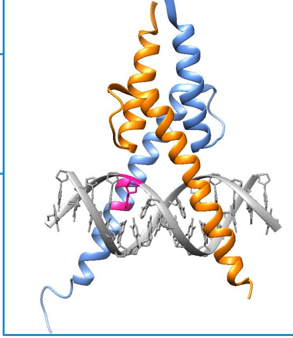




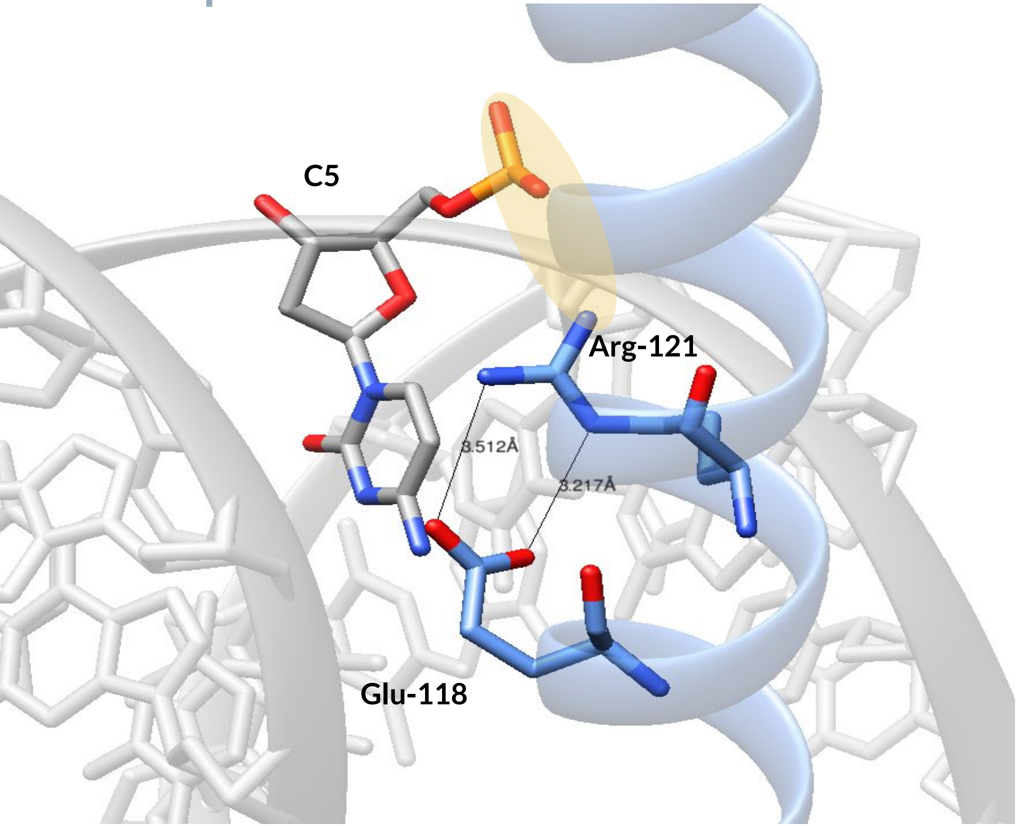
Glu-118

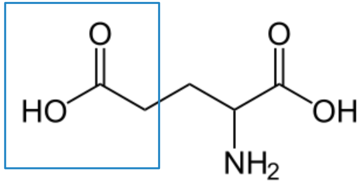
5’ TCA**ACA**GCTGTTGA 3’
3’ AGTTGT**C**GACAACT 5’

Important residues and interactions

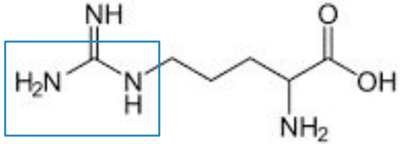


DNA-protein interactions: SPECIFICITY





Glu-118



Arg-121

5’ TCAACAGCTGTTGA 3’
3’ AGTTGTCGACAACT 5’

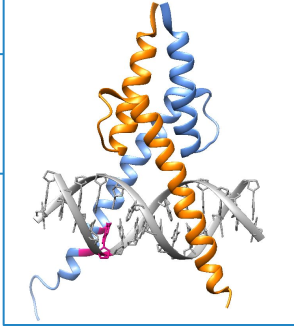
Important residues and interactions

DNA-protein interactions

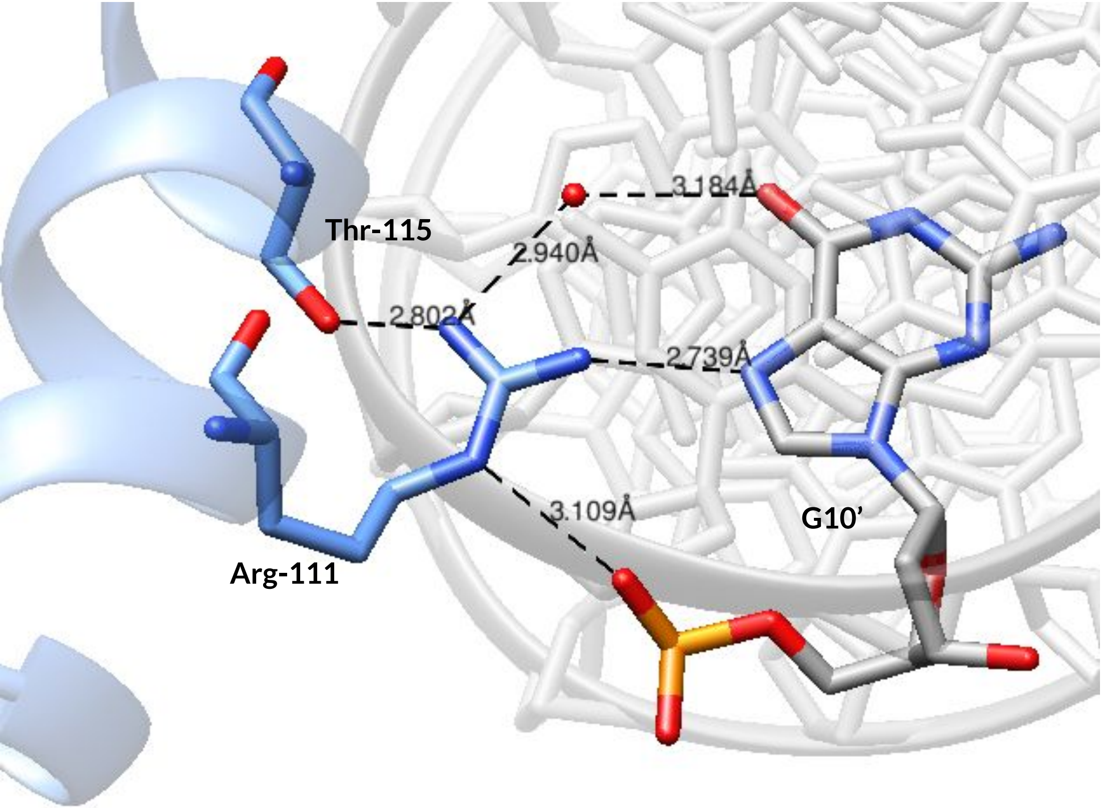
```
srebp1   --------QSRGEKRTAHNAIEKRYRSSI
max      ------------DKRAHHNALERKRRDHI
mad      -------------SRSTHNEMEKNRRAHL
myc      -------GH-MNVKRRTHNVLERQRRNEL
myod     MELKRK-TT-NADRRKAATMRERRRLSKV
scl      -----GPHT-KVVRRIFTNSRERWRQQNV
e47      -------------RRMANNARERVRVRDI
neurod1  ------------SRRMKANARERNRMHGL
```

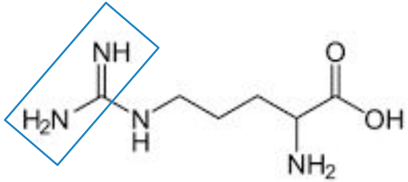
Glu118 Arg121

Important residues and interactions

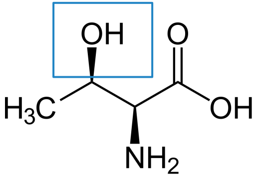


DNA-protein interactions: SPECIFICITY





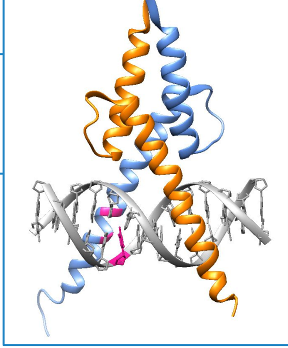
Arg-111



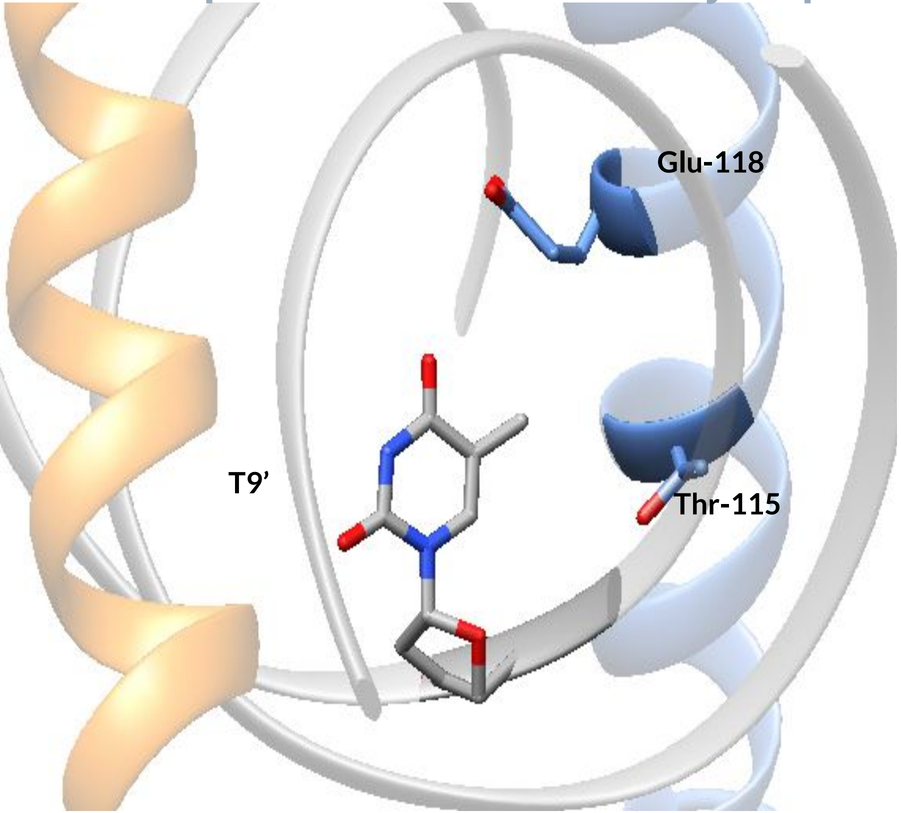
Thr-115

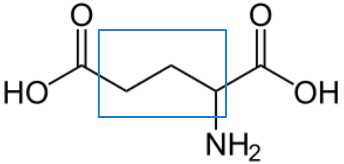
5’ TCAACAGCTGTTGA 3’
3’ AGTTGTCGACAACT 5’

Important residues and interactions

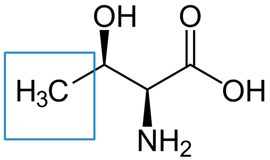


DNA-protein interactions - Hydrophobic Pocket





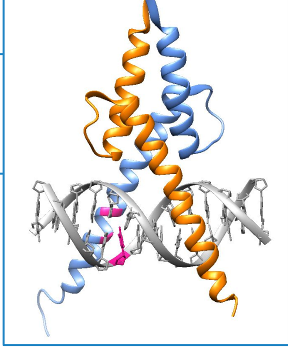
Glu-118



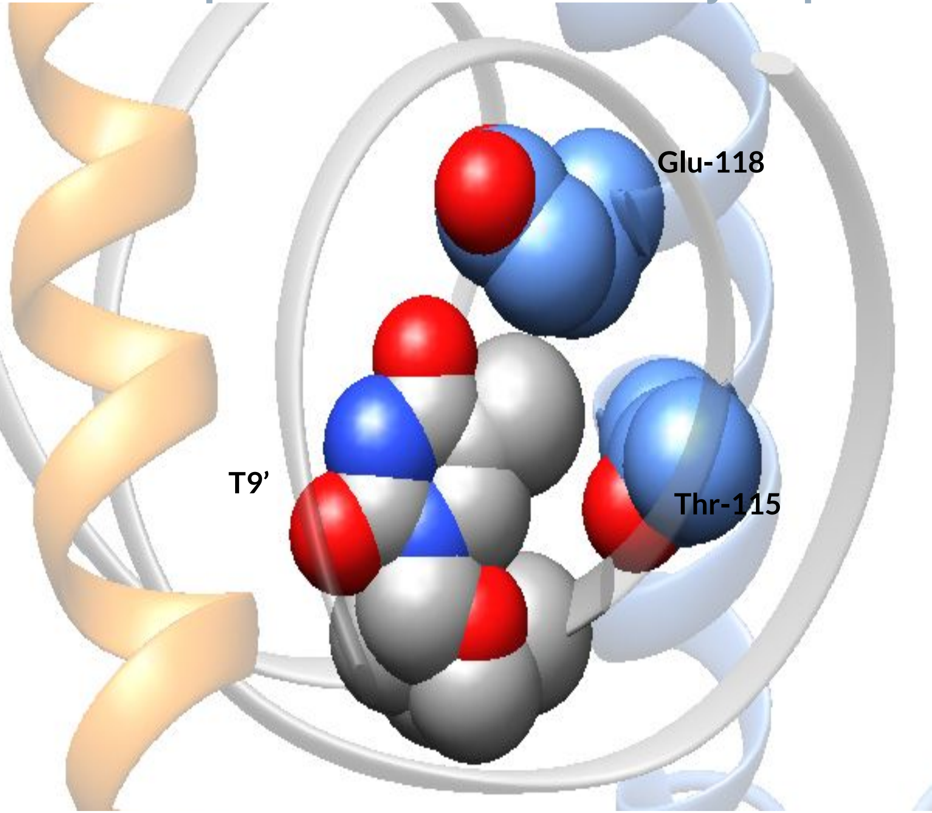
Thr-115

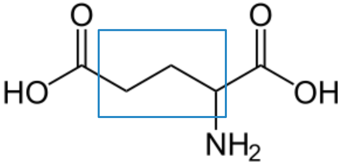
5’ TCAACAGCTGTTGA 3’
3’ AGTTG**T**CGACAACT 5’

Important residues and interactions

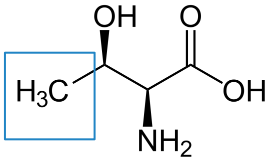


DNA-protein interactions - Hydrophobic Pocket





Glu-118

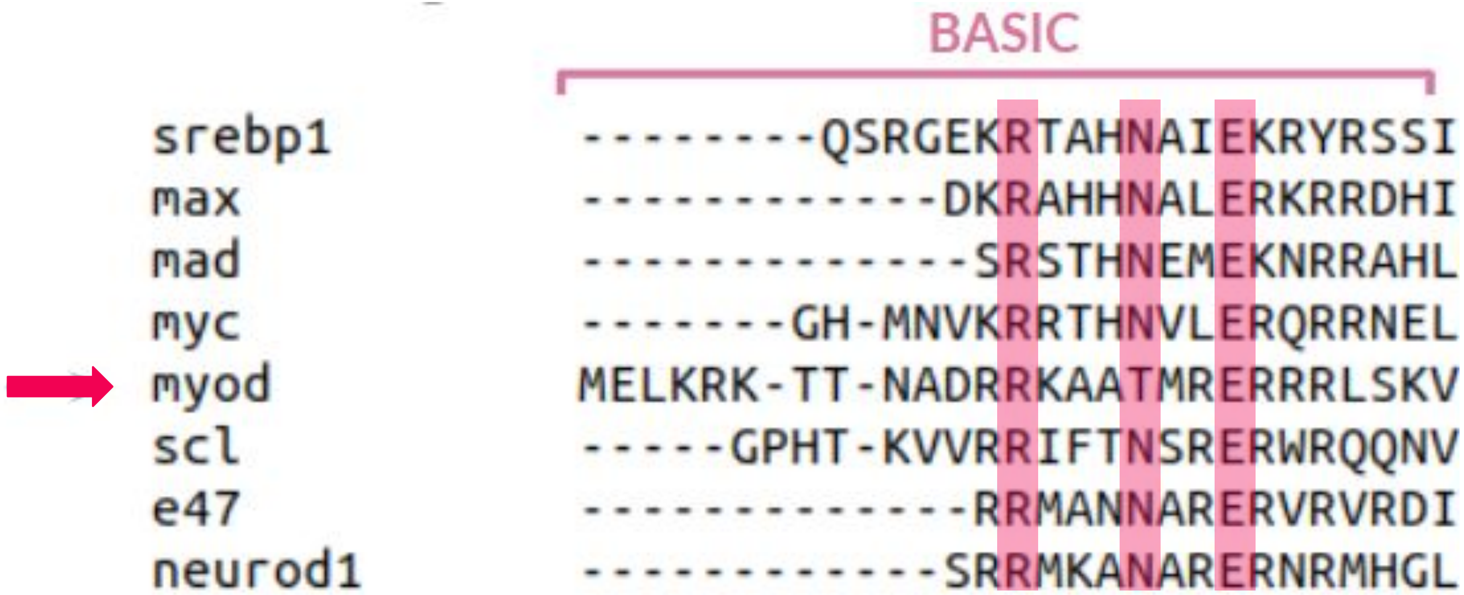


Thr-115

5' TCAACAGCTGTTGA 3'
3' AGTTG**T**CGACAACT 5'

Important residues and interactions

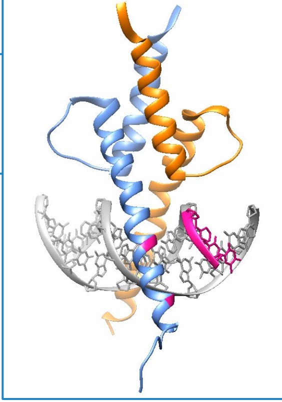
DNA-protein interactions



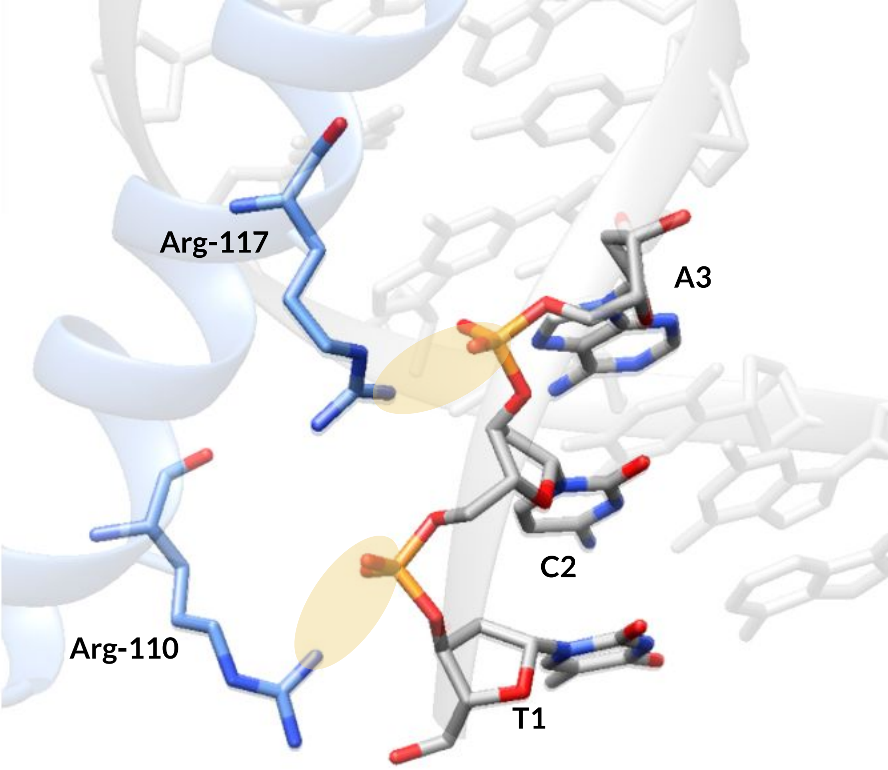
```
srebp1   --------QSRGEKRTAHNAIEKRYRSSI
max      ------------DKRAHHNALERKRRDHI
mad      -------------SRSTHNEMEKNRRAHL
myc      -------GH-MNVKRRTHNVLERQRRNEL
myod     MELKRK-TT-NADRRKAATMRERRRLSKV
scl      -----GPHT-KVVRRIFTNSRERWRQQNV
e47      -------------RRMANNARERVRVRDI
neurod1  ------------SRRMKANARERNRMHGL
```

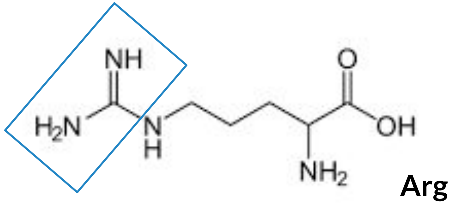
Arg111 Thr115 Glu118

Important residues and interactions



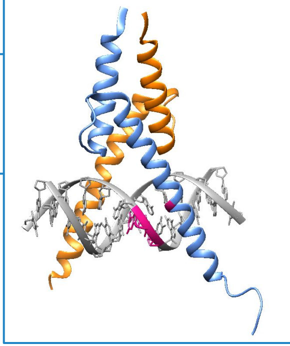
DNA-protein interactions: STABILIZATION



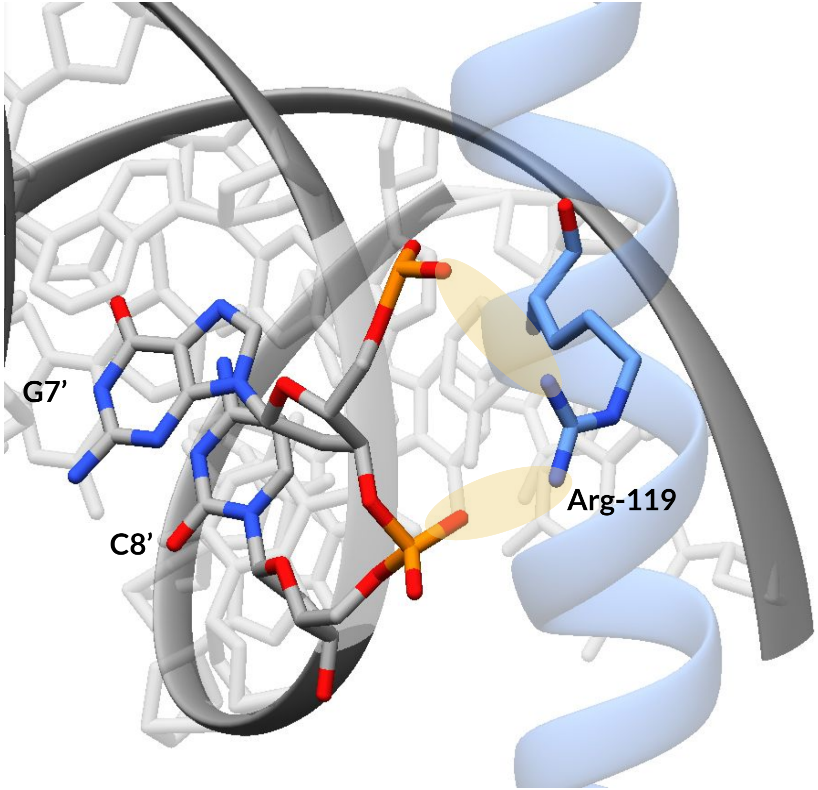


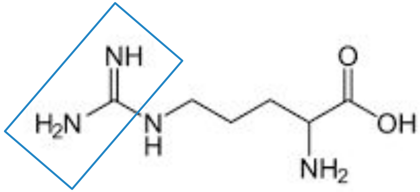
5’ **TCA**ACAGCTGTTGA 3’
3’ AGTTGTCGACAACT 5’

Important residues and interactions



DNA-protein interactions: STABILIZATION



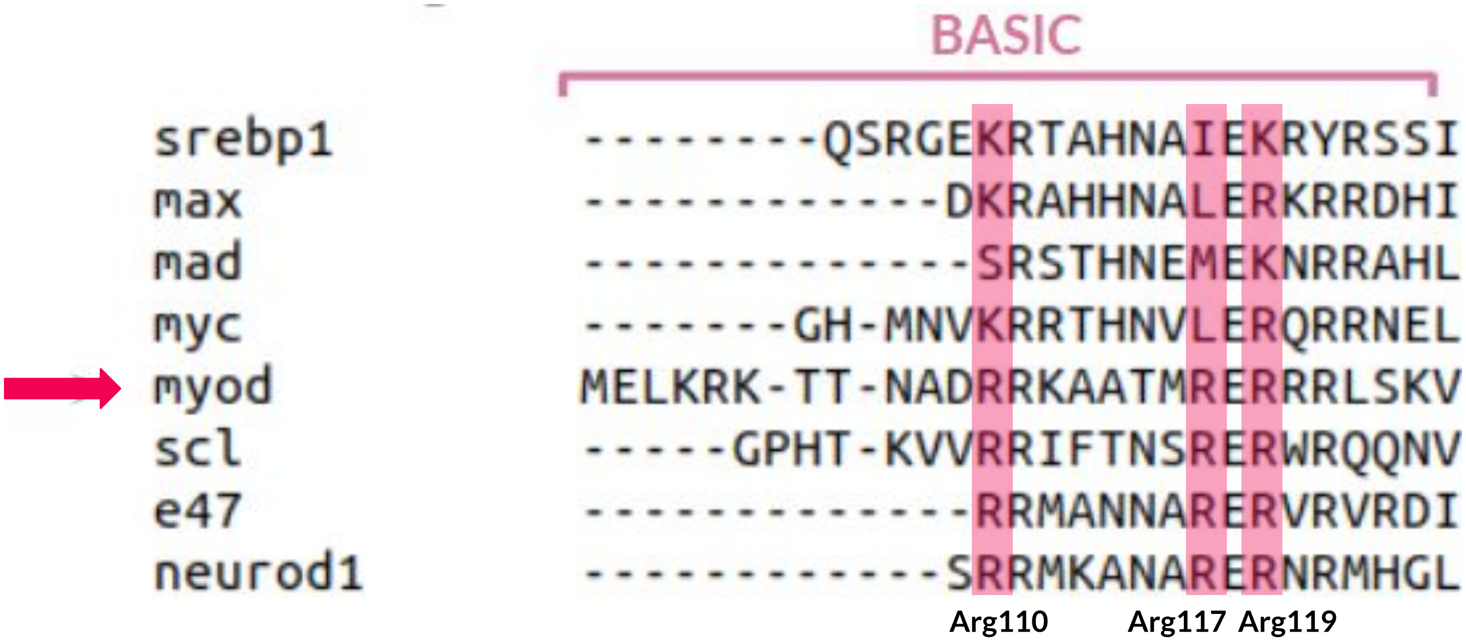


Arg-119

5' TCAACAGCTGTTGA 3'
3' AGTTGT<u>CG</u>ACAACT 5'

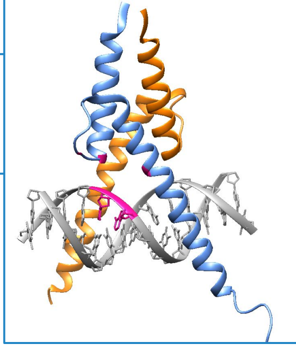
Important residues and interactions

DNA-protein interactions

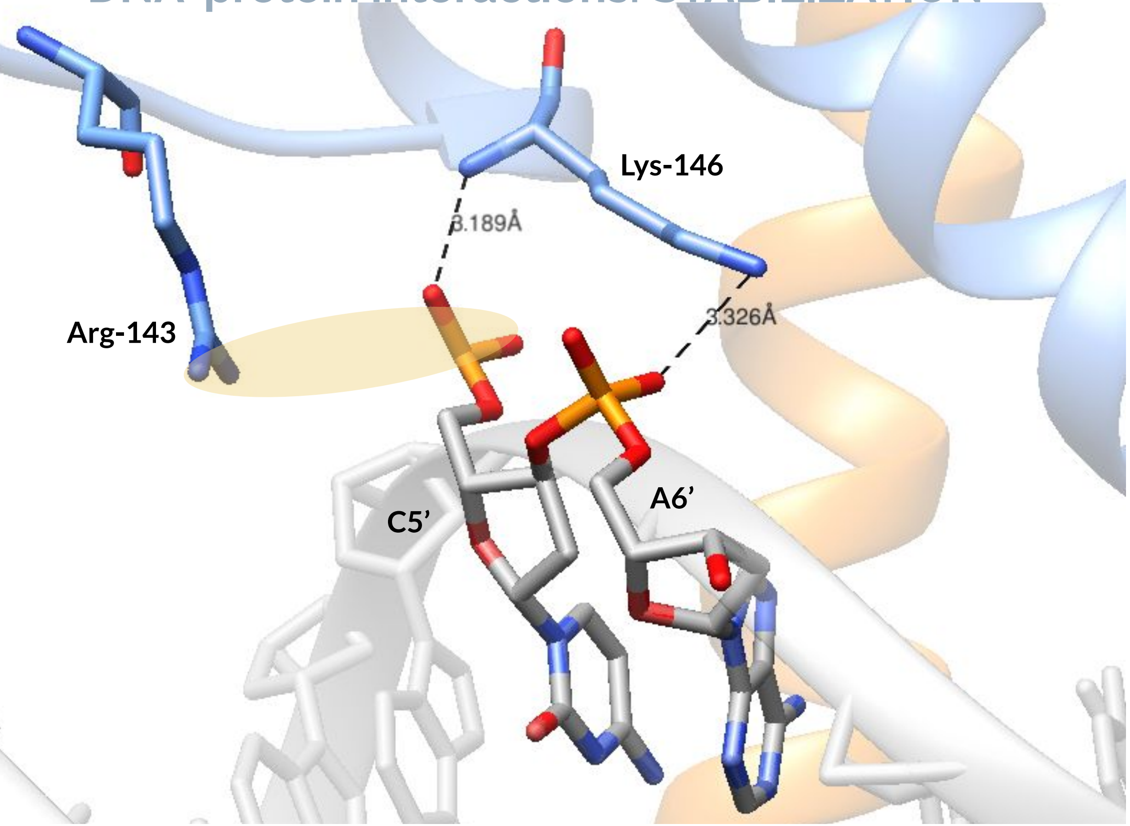


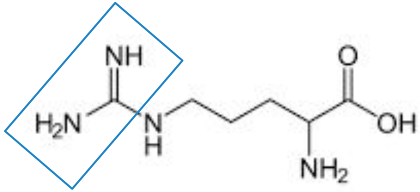
```
srebp1      --------QSRGEKRTAHNAIEKRYRSSI
max         ------------DKRAHHNALERKRRDHI
mad         -------------SRSTHNEMEKNRRAHL
myc         -------GH-MNVKRRTHNVLERQRRNEL
myod        MELKRK-TT-NADRRKAATMRERRRLSKV
scl         -----GPHT-KVVRRIFTNSRERWRQQNV
e47         -------------RRMANNARERVRVRDI
neurod1     ------------SRRMKANARERNRMHGL
```

Important residues and interactions

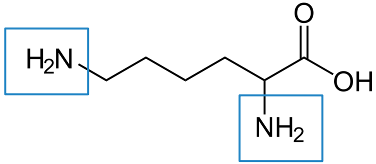


DNA-protein interactions: STABILIZATION





Arg-143

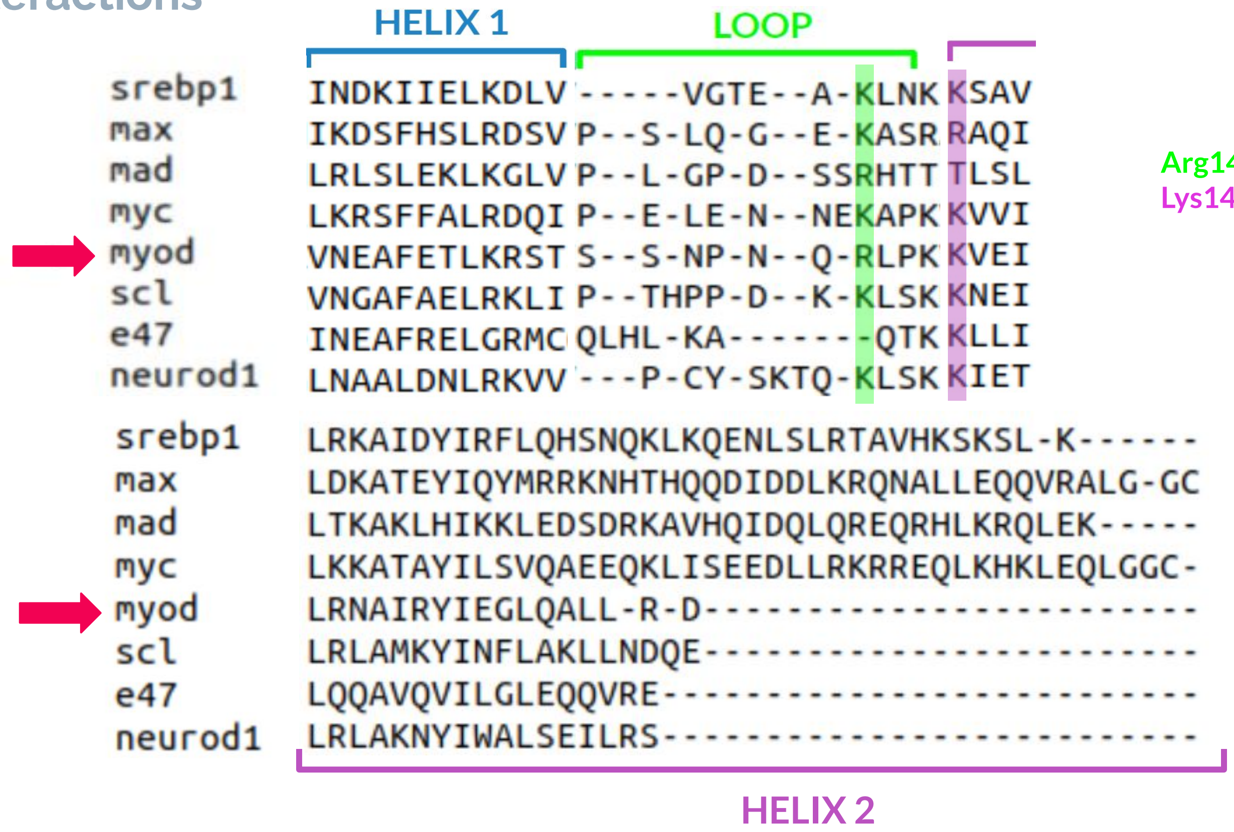


Lys-146

5' TCAACAGCTGTTGA 3'
3' AGTTGTCGACAACT 5'

Important residues and interactions

DNA-protein interactions

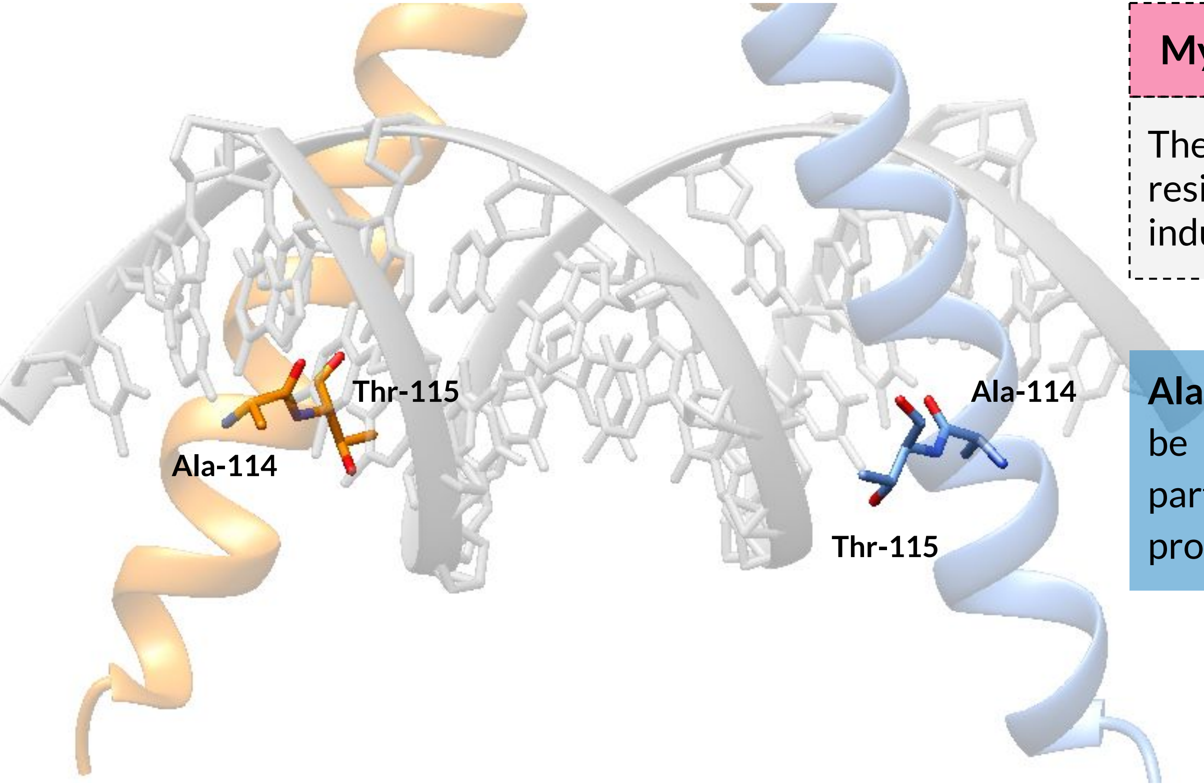


```
srebp1     INDKIIELKDLV -----VGTE--A-KLNK KSAV
max        IKDSFHSLRDSV P--S-LQ-G--E-KASR RAQI
mad        LRLSLEKLKGLV P--L-GP-D--SSRHTT TLSL
myc        LKRSFFALRDQI P--E-LE-N--NEKAPK KVVI
myod       VNEAFETLKRST S--S-NP-N--Q-RLPK KVEI
scl        VNGAFAELRKLI P--THPP-D--K-KLSK KNEI
e47        INEAFRELGRMC QLHL-KA-------QTK KLLI
neurod1    LNAALDNLRKVV ---P-CY-SKTQ-KLSK KIET

srebp1     LRKAIDYIRFLQHSNQKLKQENLSLRTAVHKSKSL-K------
max        LDKATEYIQYMRRKNHTHQQDIDDLKRQNALLEQQVRALG-GC
mad        LTKAKLHIKKLEDSDRKAVHQIDQLQREQRHLKRQLEK-----
myc        LKKATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC-
myod       LRNAIRYIEGLQALL-R-D------------------------
scl        LRLAMKYINFLAKLLNDQE------------------------
e47        LQQAVQVILGLEQQVRE--------------------------
neurod1    LRLAKNYIWALSEILRS--------------------------
```

Arg143
Lys146

Important residues and interactions

DNA-protein interactions: MYOGENIC CODE



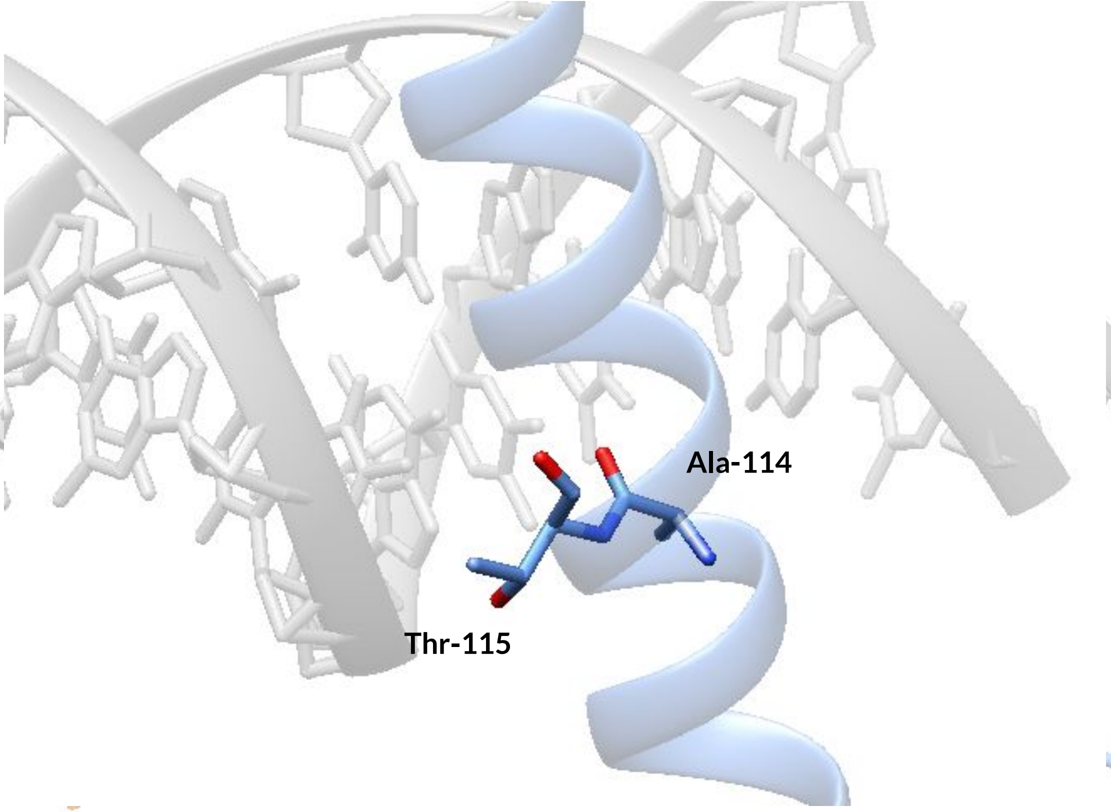
Myogenic code

The myogenic code is a sequence of residues absolutely required for dominant induction of myogenesis by the MRFs.

Alanine 114 and Threonine 115 (AT) may be required to set the basic domain in a particular conformation required for the proper binding to the E-box.

Important residues and interactions

DNA-protein interactions: MYOGENIC CODE



Myogenic code

The myogenic code is a sequence of residues absolutely required for dominant induction of myogenesis by the MRFs.

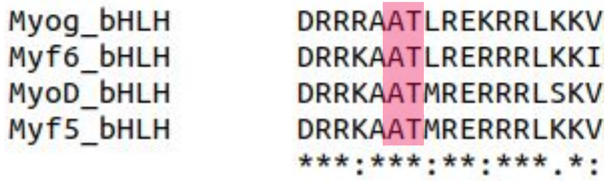
Alanine 114 and Threonine 115 (AT) may be required to set the basic domain in a particular conformation required for the proper binding to the E-box.

ANNEX: Same family structures

Multiple sequence alignment (MyoD in different species)

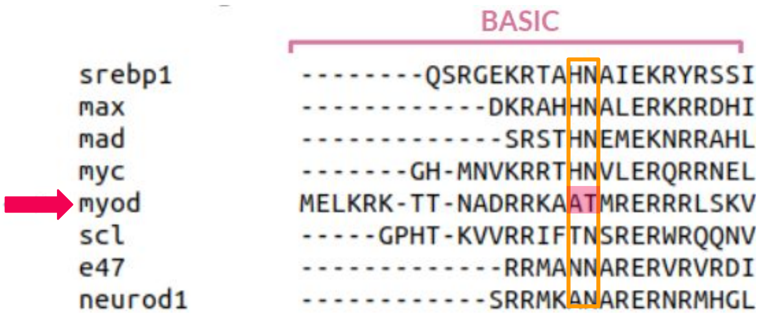
```
Chick     DRRKAATMRERRRLSKV
Cotja     DRRKAATMRERRRLSKV
Danre     DRRKAATMRERRRLSKV
Human     DRRKAATMRERRRLSKV
Mouse     DRRKAATMRERRRLSKV
Pig       DRRKAATMRERRRLSKV
Bovin     DRRKAATMRERRRLSKV
Sheep     DRRKAATMRERRRLSKV
Rat       DRRKAATMRERRRLSKV
Caeel     DRRKAATMRERRRLRKV
          ************** **
```

Multiple sequence alignment (MRF family)



```
Myog_bHLH     DRRRAATLREKRRLKKV
Myf6_bHLH     DRRKAATLRERRRLKKI
MyoD_bHLH     DRRKAATMRERRRLSKV
Myf5_bHLH     DRRKAATMRERRRLKKV
              ***:***:**:***.*:
```

Structural alignment (monomers)



```
srebp1     --------QSRGEKRTAHNAIEKRYRSSI
max        ------------DKRAHHNALERKRRDHI
mad        -------------SRSTHNEMEKNRRAHL
myc        -------GH-MNVKRRTHNVLERQRRNEL
myod       MELKRK-TT-NADRRKAATMRERRRLSKV
scl        -----GPHT-KVVRRIFTNSRERWRQQNV
e47        -------------RRMANNARERVRVRDI
neurod1    ------------SRRMKANARERNRMHGL
```

Interactions summary

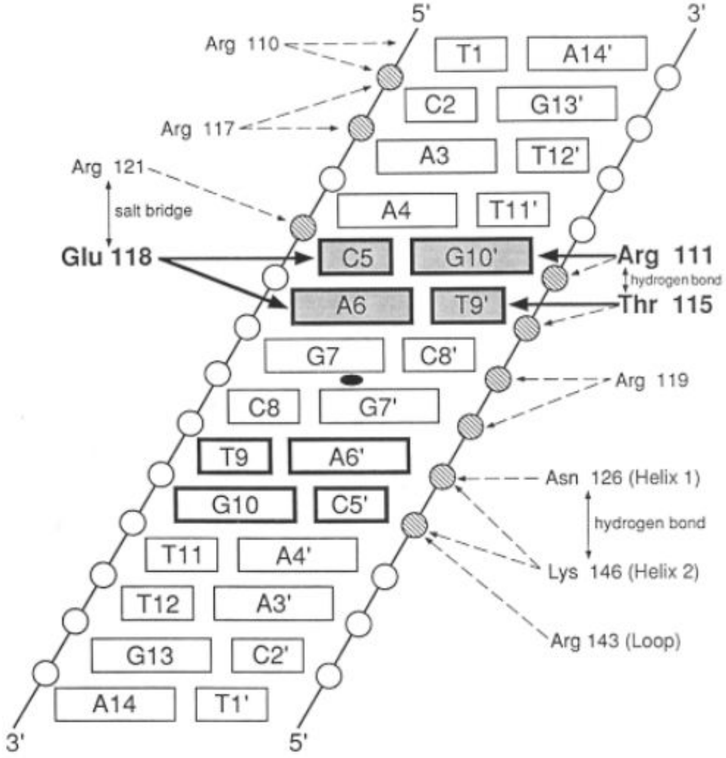


Figure 4. Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*. 1994 May 6;77(3):451-9.

2.

Zinc Fingers

Zinc fingers

Small, functional, independently folded domain that requires coordination of one or more **zinc ions** to stabilize its structure.

1. C2H2 family
2. Zinc Ribon
3. Zinc binding loops
4. Gag Knuckle
5. Zn2/Cys6
6. Metallothionein
7. Treble clef
8. TAZ2 domain

→ **C2H2 family**

Zn is coordinated by two Cys and two His residues

They are transcription factors that function by recognition of specific DNA sequences

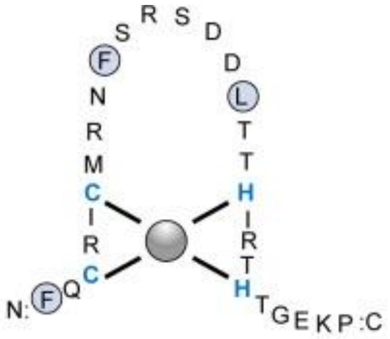


Figure 5. William J, Lennarz, M. Daniel Lane. (2013) Encyclopedia of biological Chemistry: Zinc Fingers. Academic Press

Zif268

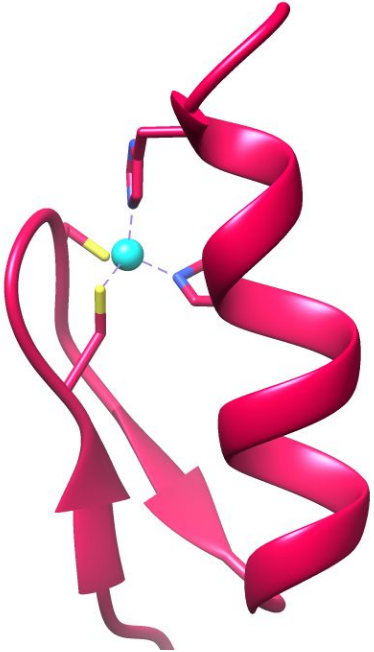
Zif268/EGR1

Basic features

- Transcription factor
- Binds to the DNA sequence 5'-GCG(T/G)GGGCG-3’
- Key role in the regulation of cell survival, proliferation invasion and cell death.
- Oncogenic role → gastric and pancreatic cancer
- Tumor suppressor → gliomas and melanocytomas

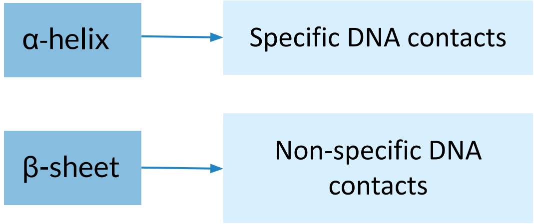
SCOP classification

Class	*Small proteins* *Usually dominated by metal ligand, heme, and/or disulfide bridges*
Fold	*Beta-beta-alpha zinc fingers* *(N-terminal beta-hairpin and C-terminal alpha-helical region; each part provides two zinc-coordinating residues)*
Superfamily	*Beta-beta-alpha zinc fingers*
Family	*Classic Zinc Fingers, C2H2*
Protein	*ZIF268*
Organisme	*Mus musculus (Mouse)*



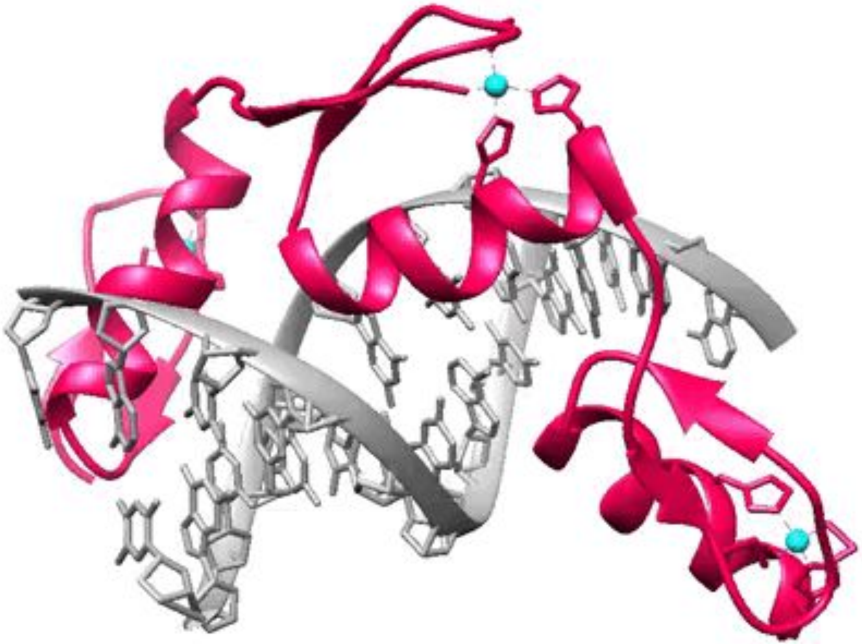
Zif268: general structure

- 3 Zinc fingers
- Antiparallel β-sheet + α-helix
- Zn + hydrophobic core

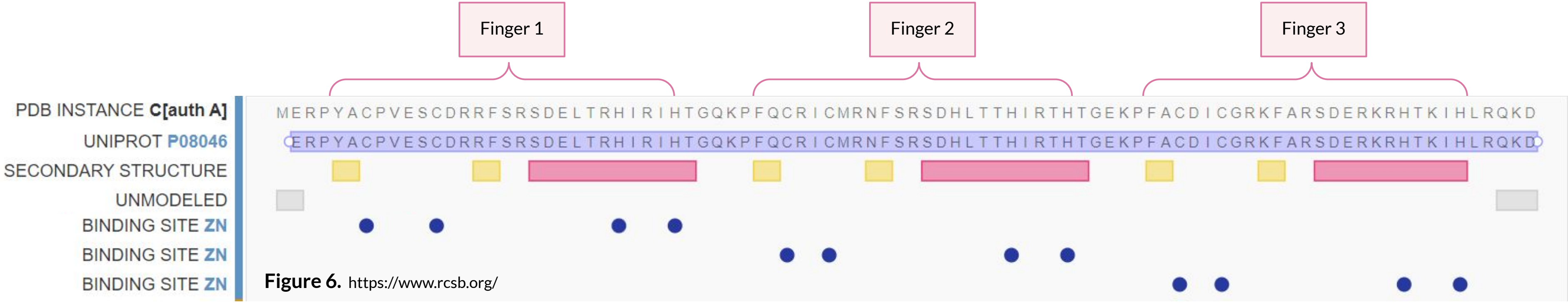


α-helix → Specific DNA contacts

β-sheet → Non-specific DNA contacts



1AAY Second structure correlation



PDB INSTANCE **C[auth A]** MERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRTHTGEKPFACDICGRKFARSDERKRHTKIHLRQKD

UNIPROT **P08046** ERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRTHTGEKPFACDICGRKFARSDERKRHTKIHLRQKD

SECONDARY STRUCTURE

UNMODELED

BINDING SITE **ZN**

BINDING SITE **ZN**

BINDING SITE **ZN**

Figure 6. https://www.rcsb.org/

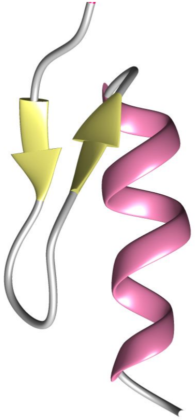
MERPYACPVESCDRRFSRSDELTRHIRIHTGQK

PFQCRI--CMRNFSRSDHLTTHIRTHTGEK

PFACDI--CGRKFARSDERKRHTKIHLRQKD

Antiparallel β-sheet

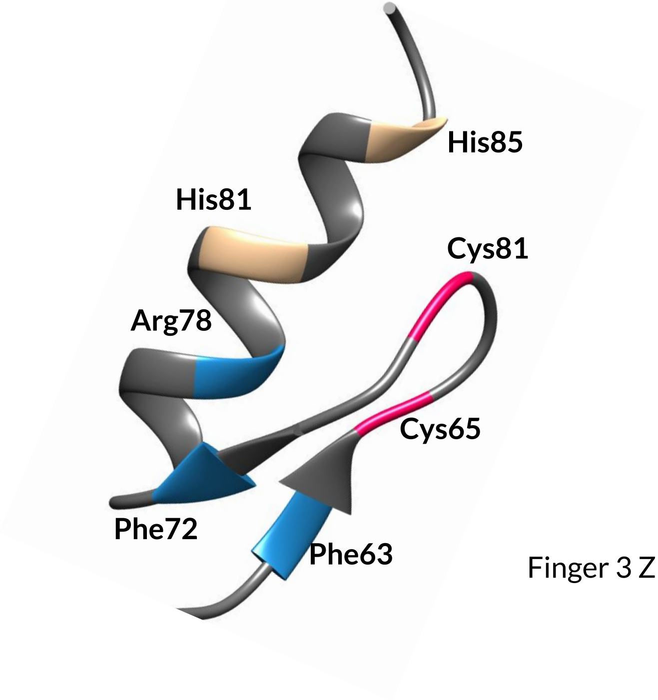
α - helix



C2H2 family: Consensus sequence

$-Z-X_{1-2}-C-X_{2-4}-C-X_{3}-Z-X_{5}-Z-X_{2}-H-X_{3-5}-H$

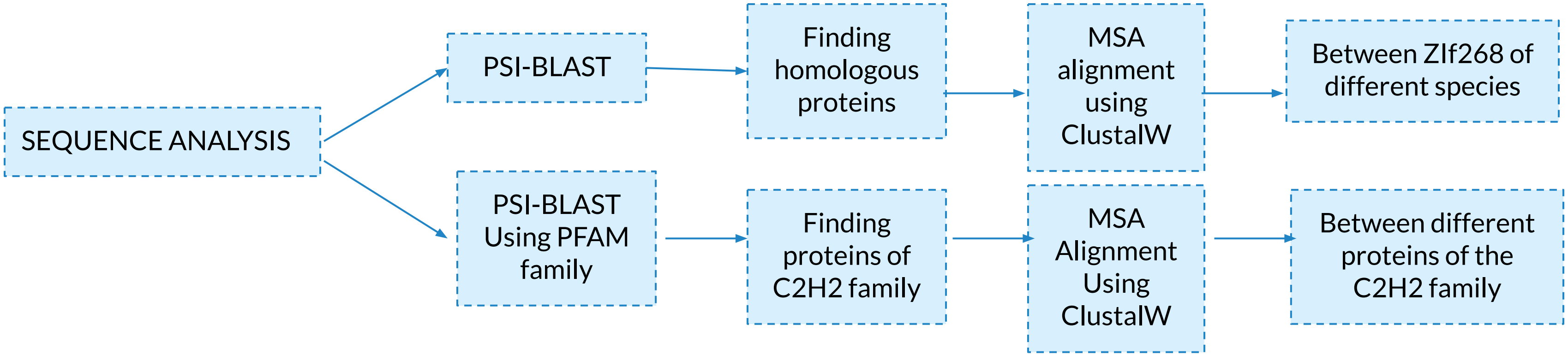
C = Cysteine residue
Z = Hydrophobic residue
H = Histidine residue
X = Any residue



Finger 3 Zif268

Sequence and structure analysis

Methodology



SEQUENCE ANALYSIS

PSI-BLAST

Finding homologous proteins

MSA alignment using ClustalW

Between Zlf268 of different species

PSI-BLAST Using PFAM family

Finding proteins of C2H2 family

MSA Alignment Using ClustalW

Between different proteins of the C2H2 family

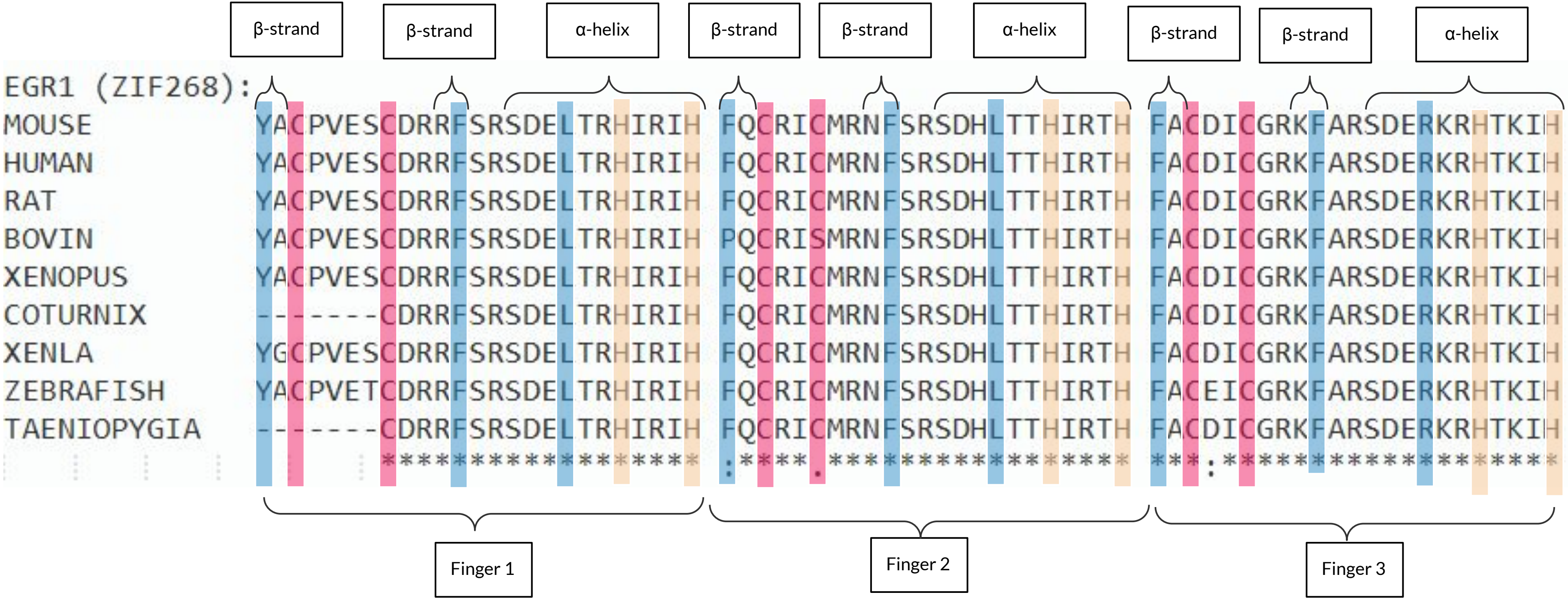
Methodology

STRUCTURAL ANALYSIS → Selection of C2H2 family proteins → STAMP → Alignfit → Superimposition

PDB ID	Resolution	Protein	Length	Organism
4r2e	1.80 Å	Wilms Tumor Protein	93	*Homo sapiens*
1aaY	1.60 Å	Zif268	90	*Mus musculus*
1ubd	2.50 Å	YY1	124	*Homo sapiens*
2gli	2,60 Å	Gli	155	*Homo sapiens*

Analyzing consensus sequence residues

MSA: Residues conservation among different species



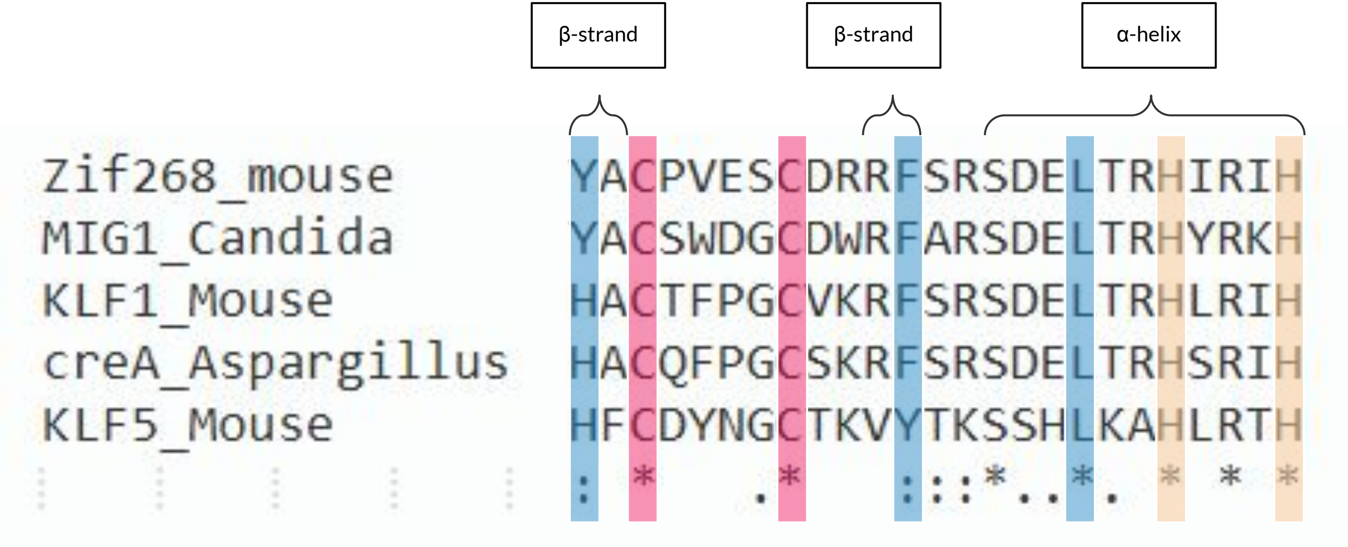
```
EGR1 (ZIF268):
MOUSE         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT           YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN         YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS       YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX      -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA         YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH     YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA   -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                     ****************** :****.***************** ***:*******************
```

🟥 Cys 🟦 Hydrophobic 🟧 His

-Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H

MSA: Residues conservation among different C2H2 ZnF

Clustal 2.1 multiple sequence alignment

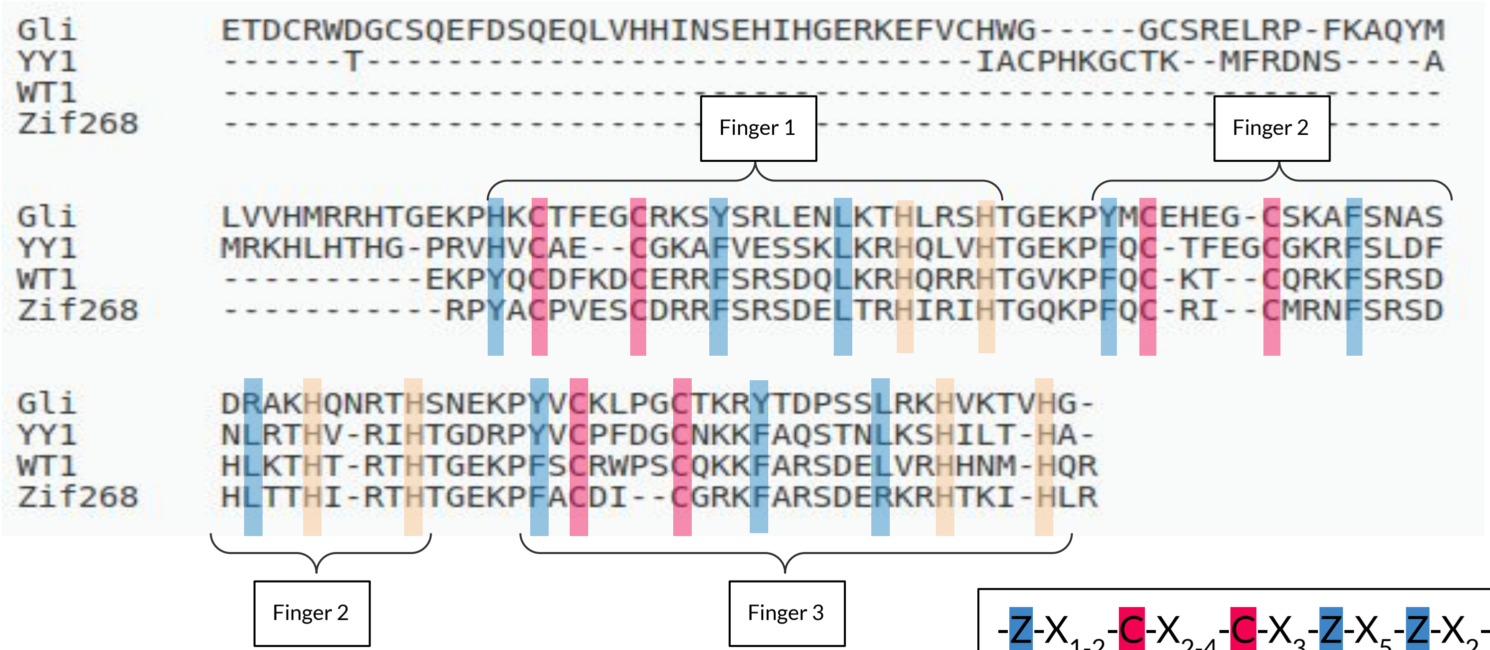


```
Zif268_mouse         YACPVESCDRRFSRSDELTRHIRIH
MIG1_Candida         YACSWDGCDWRFARSDELTRHYRKH
KLF1_Mouse           HACTFPGCVKRFSRSDELTRHLRIH
creA_Aspargillus     HACQFPGCSKRFSRSDELTRHSRIH
KLF5_Mouse           HFCDYNGCTKVYTKSSHLKAHLRTH
                     : *   .*   :::*..*. * * *
```

$$-Z-X_{1-2}-C-X_{2-4}-C-X_{3}-Z-X_{5}-Z-X_{2}-H-X_{3-5}-H$$

🟥 Cys 🟦 Hydrophobic 🟧 His

Structural analysis among different C2H2 ZnF

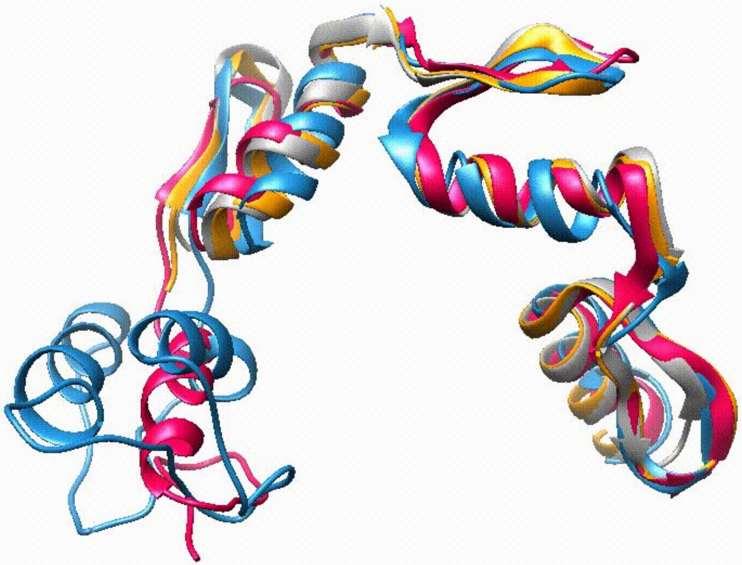


```
Gli       ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG----GCSRELRP-FKAQYM
YY1       ------T-----------------------------IACPHKGCTK--MFRDNS----A
WT1       -----------------------------------------------------------
Zif268    -----------------------------------------------------------

Gli       LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1       MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1       ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268    -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli       DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1       NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1       HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268    HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

$-Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H$

Cys Hydrophobic His

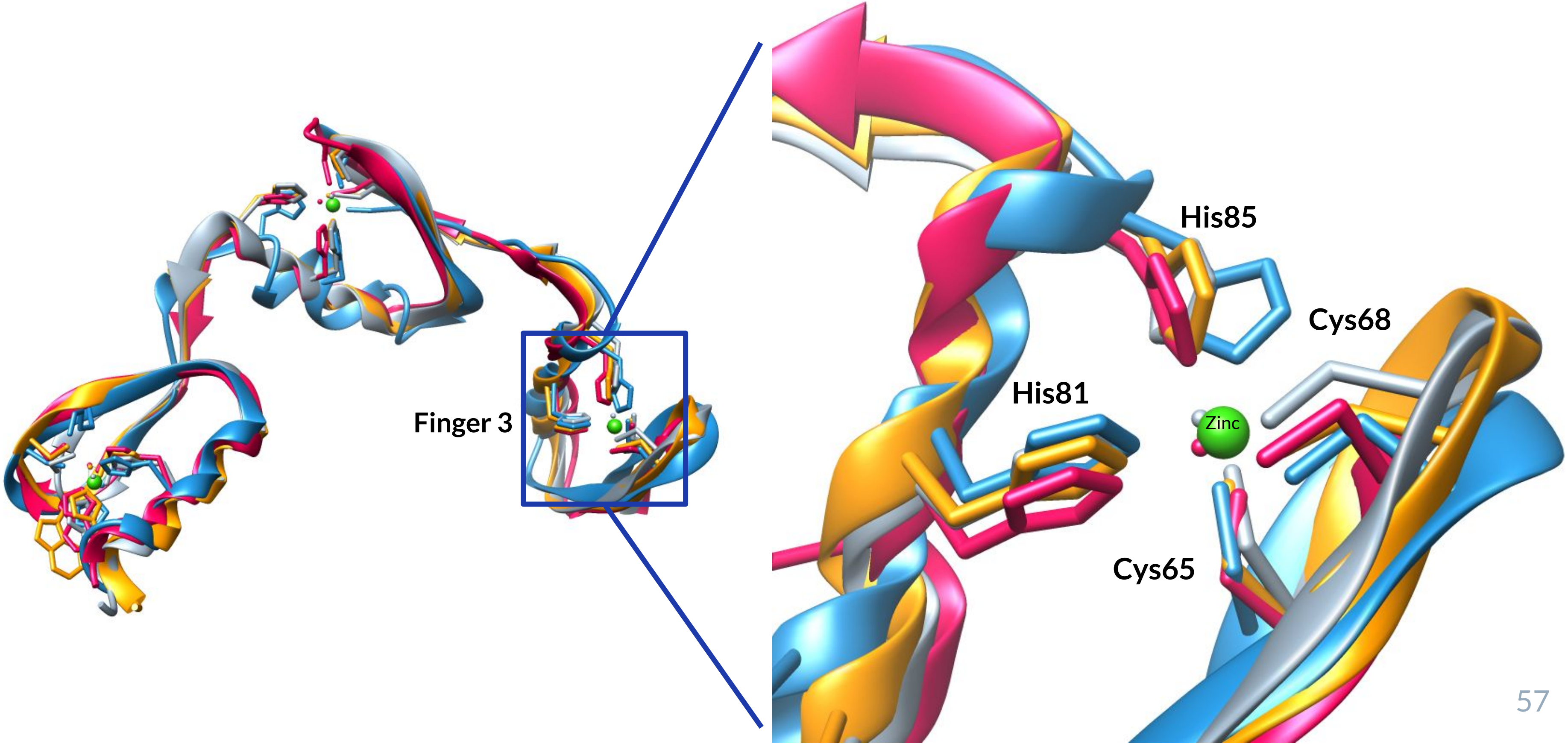
Superimposition



● YY1C
● GliA
● WT1A
● Zif268A

Sc = **6.18** RMS = **1.74** Len = **156** nfit = **82**

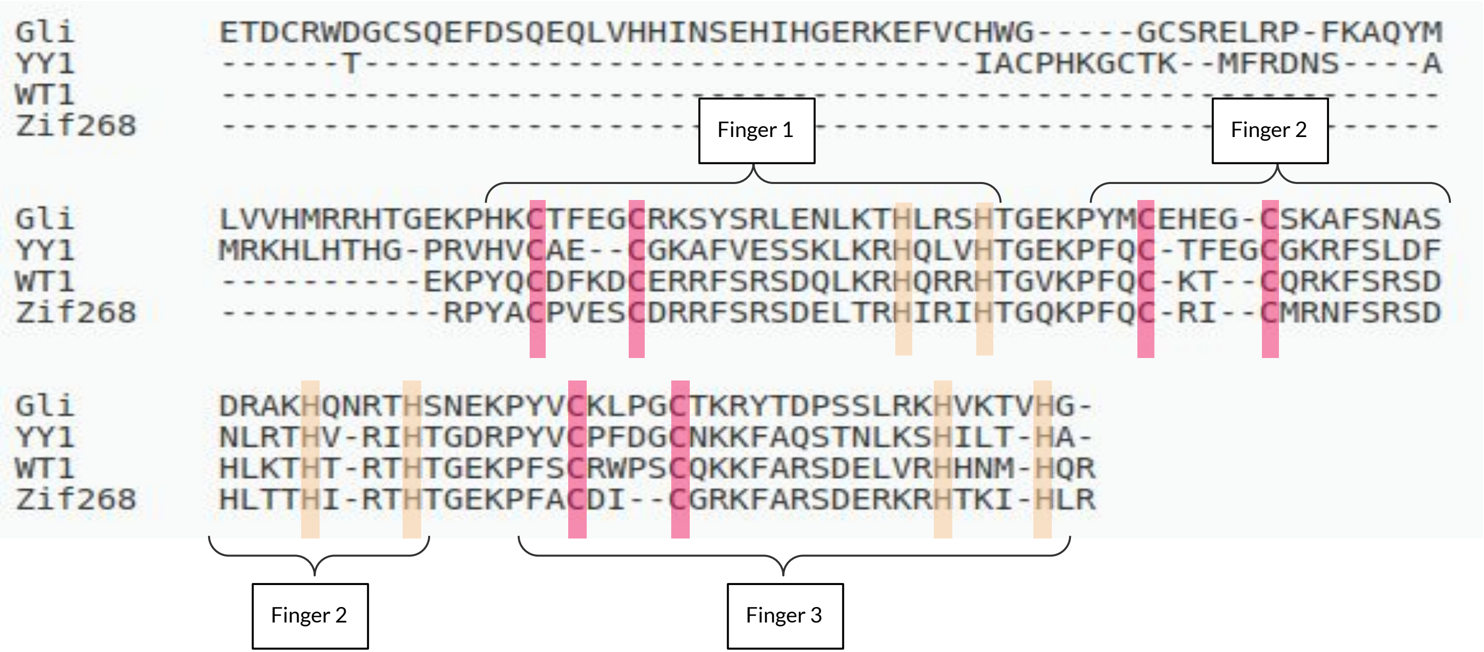
Superimposition with Zn atoms



Key residues

(Consensus sequence)

Conservation of residues



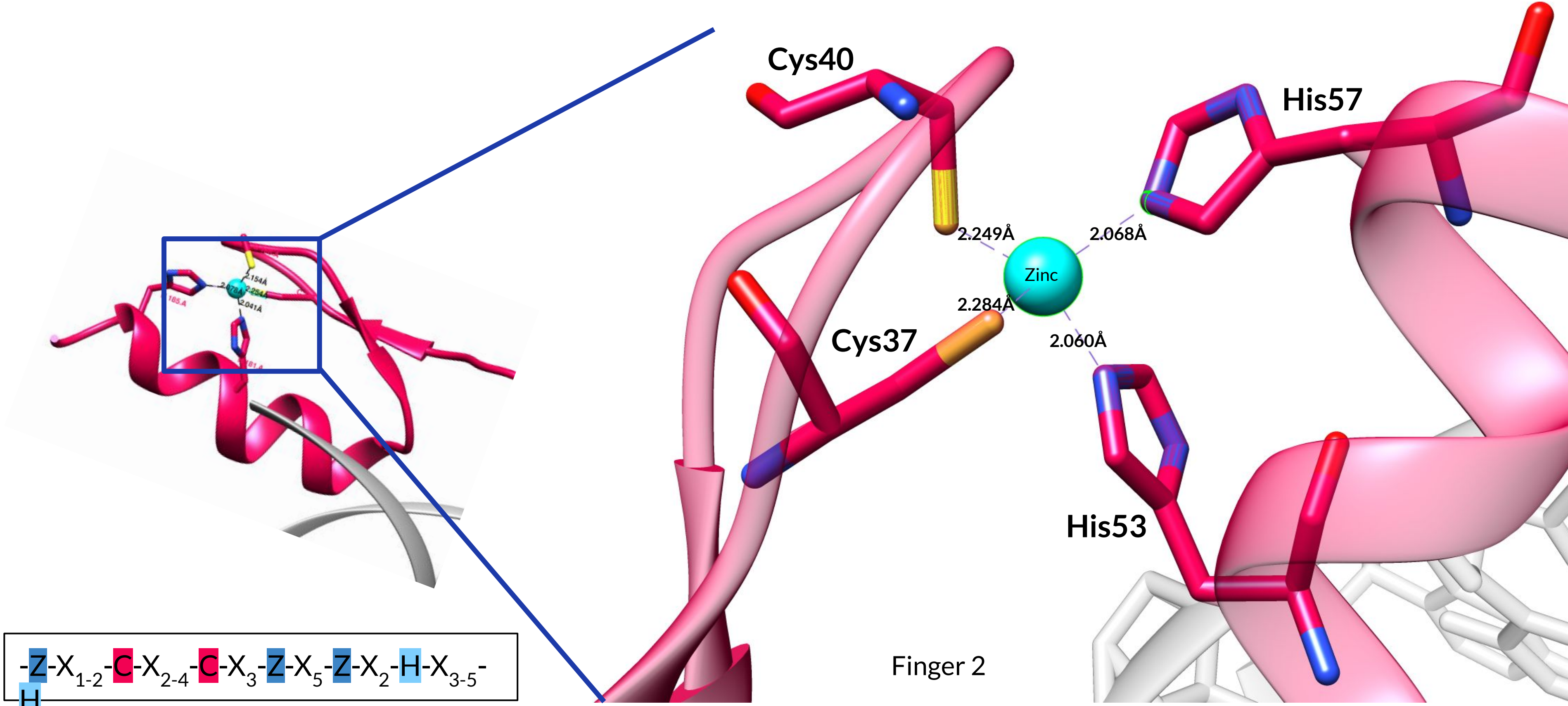
```
Gli      ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG----GCSRELRP-FKAQYM
YY1      ------T-----------------------------IACPHKGCTK--MFRDNS----A
WT1      -----------------------------------------------------------
Zif268   -----------------------------------------------------------

Gli      LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1      MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1      ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268   -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli      DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1      NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1      HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268   HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

🟥 Cys

🟧 His

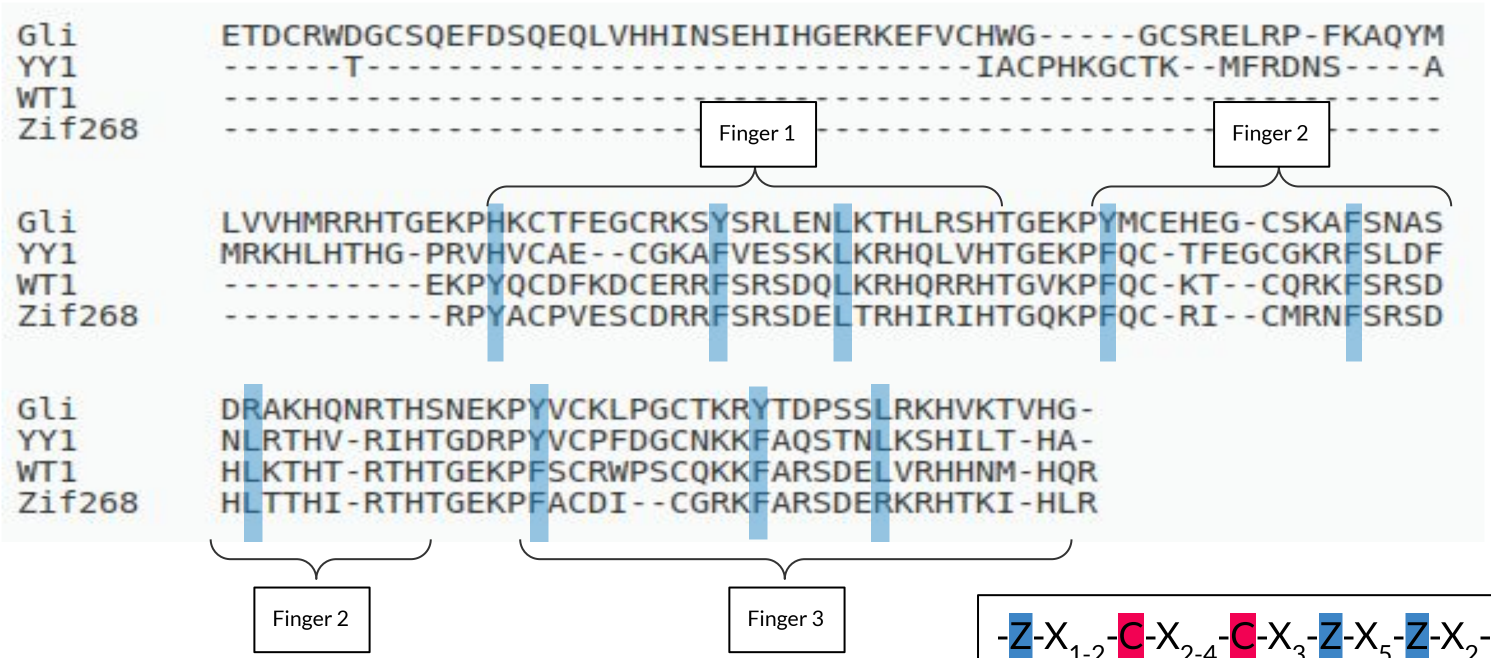
$-Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H$

Residues involved in Zn binding



$-Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H$

Conservation of residues

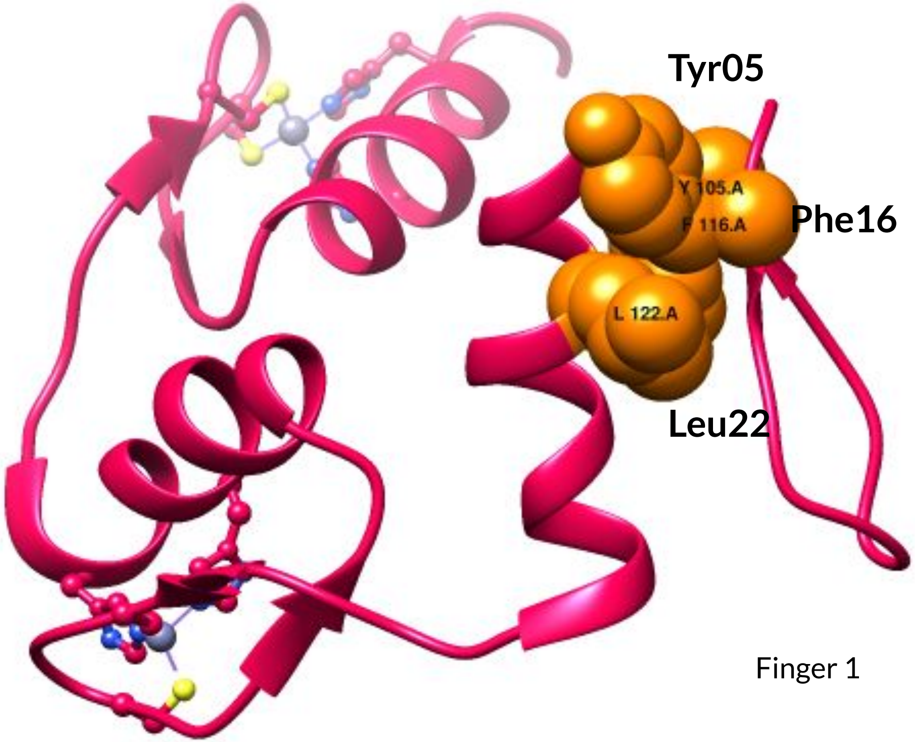


```
Gli      ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG-----GCSRELRP-FKAQYM
YY1      ------T------------------------------IACPHKGCTK--MFRDNS----A
WT1      ------------------------------------------------------------
Zif268   ------------------------------------------------------------

Gli      LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1      MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1      ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268   -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli      DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1      NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1      HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268   HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

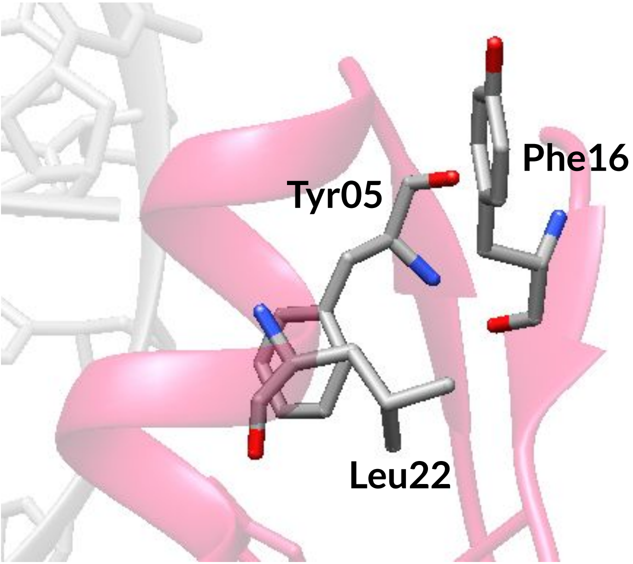
$-Z-X_{1-2}-C-X_{2-4}-C-X_{3}-Z-X_{5}-Z-X_{2}-H-X_{3-5}-H$

Hydrophobic

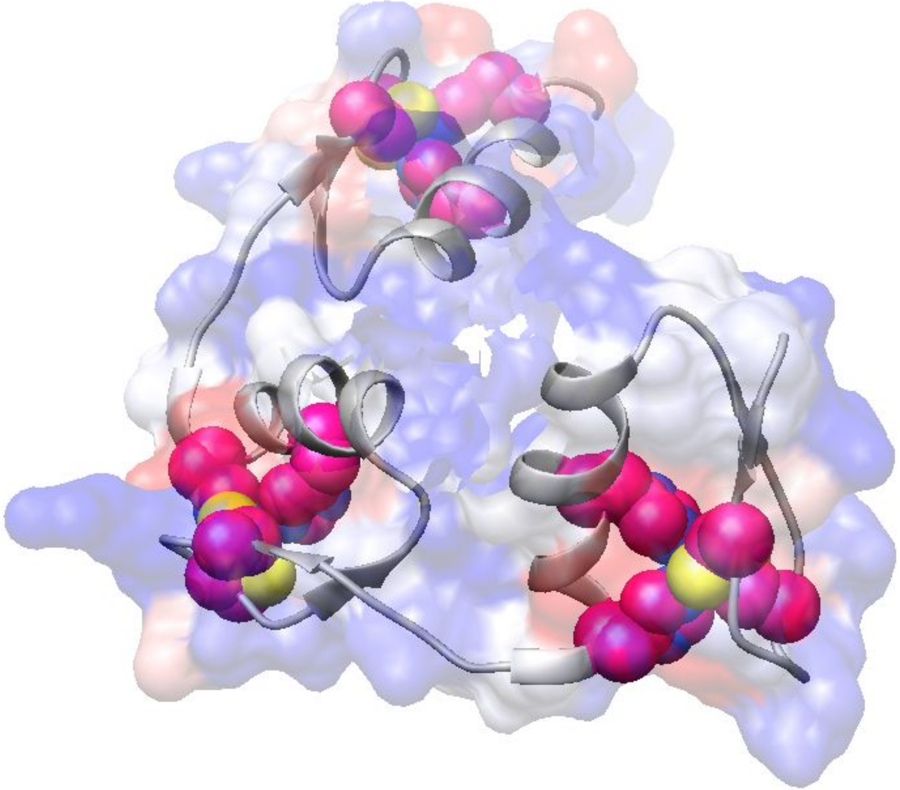
Hydrophobic core

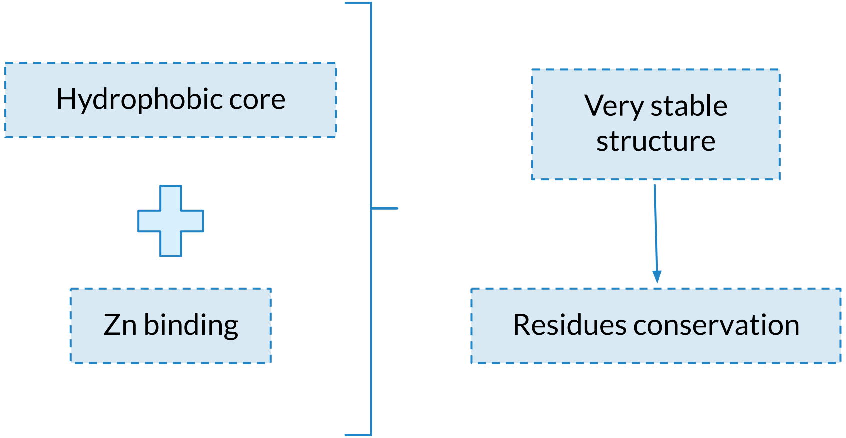


Finger 1



Hydrophobic core in the different fingers





Hydrophobic core

+

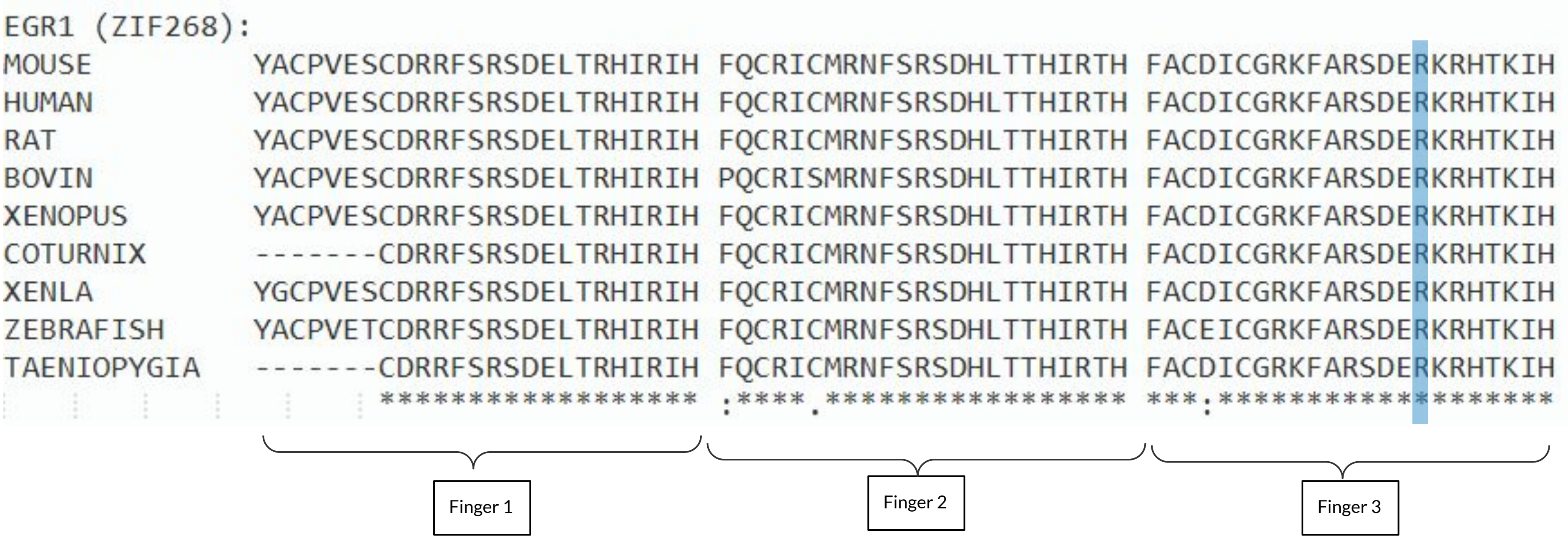
Zn binding

Very stable structure

Residues conservation

Hydrophilic 🟦🟦⬜🟥🟥 Hydrophobic

Residues conservation among species



```
EGR1 (ZIF268):
MOUSE          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT            YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN          YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS        YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX       -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA          YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH      YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA    -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                      ****************** :****.***************** ***:*******************
```

Finger 1

Finger 2

Finger 3

Cys Hydrophobic His

Zn Finger 3: Arg78 residue

1) Aliphatic carbon side chain of this Arg residue contributes to the proper formation of the hydrophobic core
2) Leu mutants had no ability to bind DNA
3) Arg78 water-mediated interaction with phosphate is essential for DNA binding

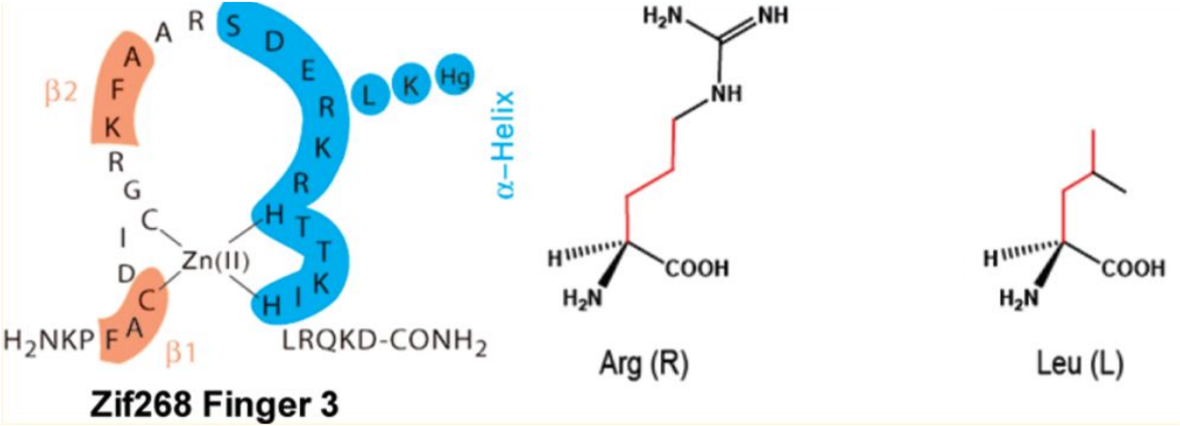
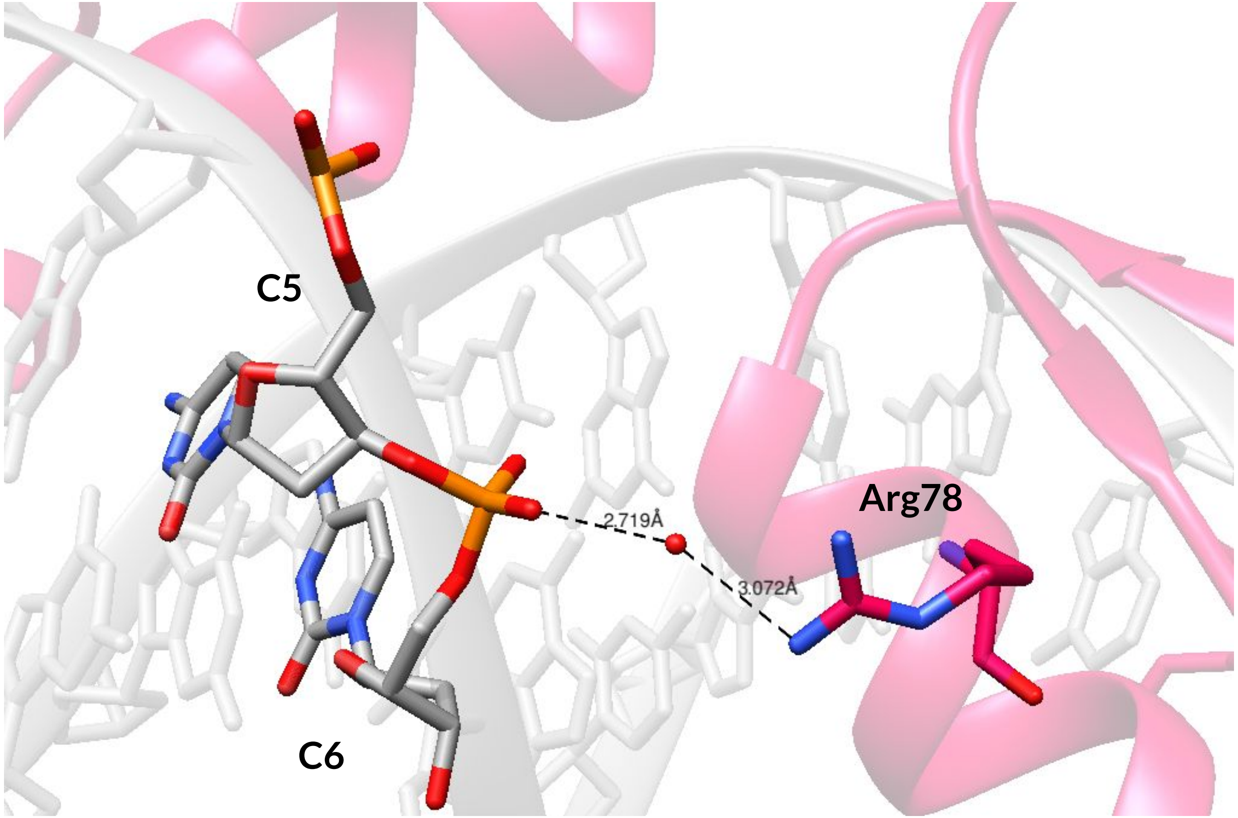


Figure 7. Negi, S., Imanishi, M., Sasaki, M., Tatsutani, K., Futaki, S., & Sugiura, Y. (2011). An arginine residue instead of a conserved leucine residue in the recognition helix of the finger 3 of Zif268 stabilizes the domain structure and mediates DNA binding. Biochemistry, 50(28), 6266-6272.

Zn Finger 3: Arg78 residue

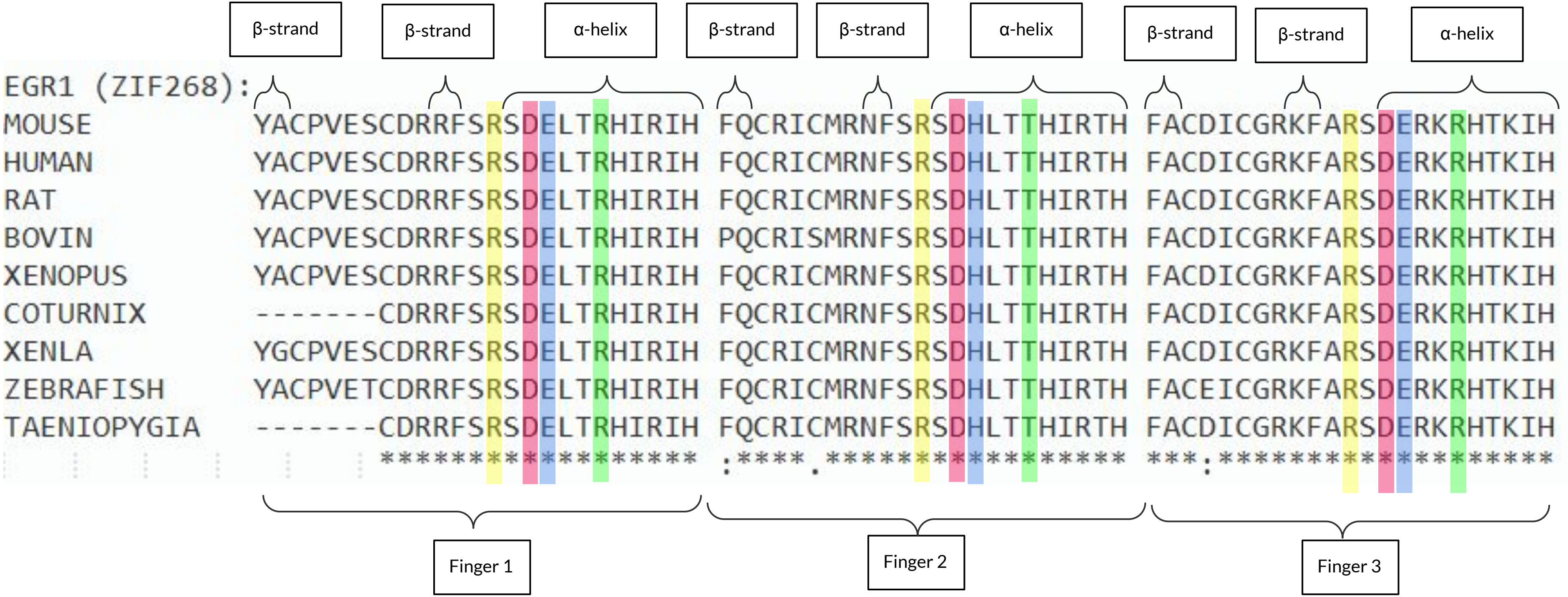


Analyzing residues involved in DNA binding

Key residues

(DNA binding specific)

Residues conservation among species



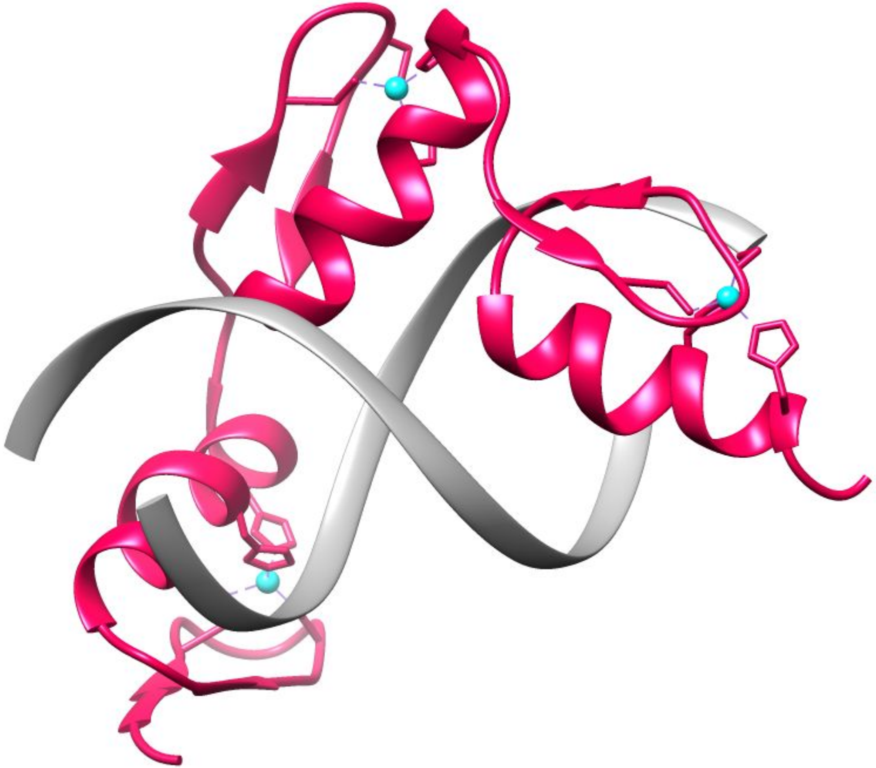
β-strand β-strand α-helix β-strand β-strand α-helix β-strand β-strand α-helix

```
EGR1 (ZIF268):
MOUSE          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT            YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN          YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS        YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX       -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA          YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH      YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA    -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                      ****************** :****.***************** ***:*******************
```

Finger 1 Finger 2 Finger 3

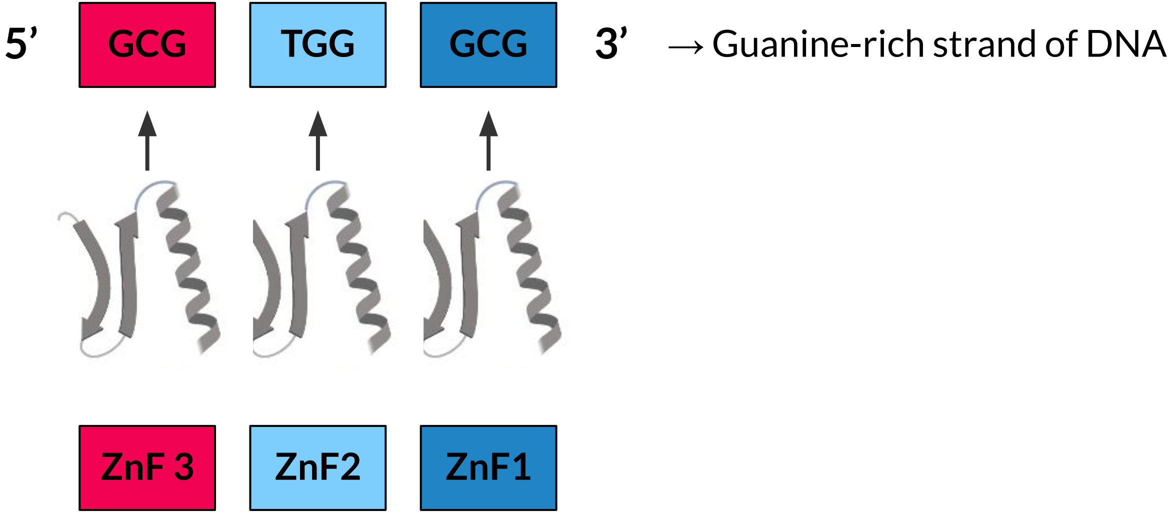
Position -1 **Position 2** **Position 3** **Position 6**

DNA Specific Interactions



DNA-binding consensus sequence

Zif268 consensus binding site



5’ GCG TGG GCG 3’ → Guanine-rich strand of DNA

ZnF 3 ZnF2 ZnF1

Residues involved in specific DNA binding

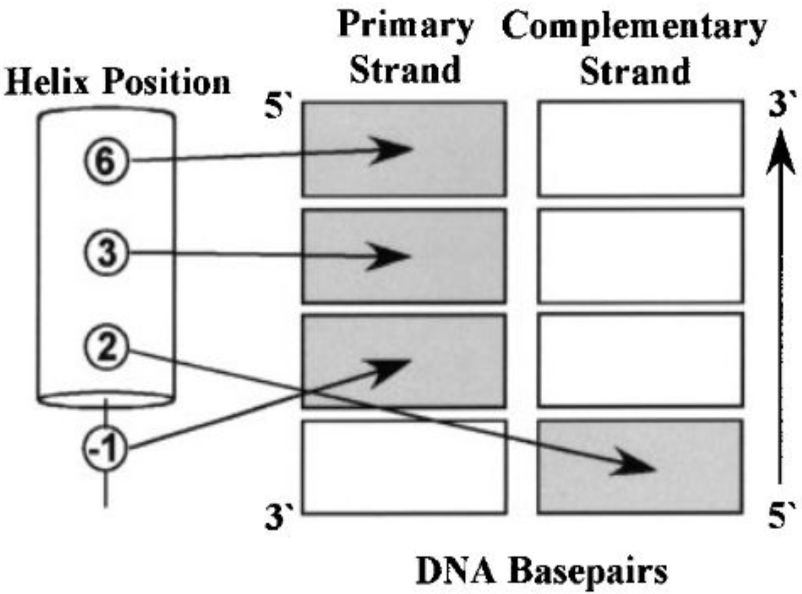


Figure 8. Miller, J. C., & Pabo, C. O. (2001). Rearrangement of side-chains in a Zif268 mutant highlights the complexities of zinc finger-DNA recognition. *Journal of molecular biology*, *313*(2), 309-315.

Important residues

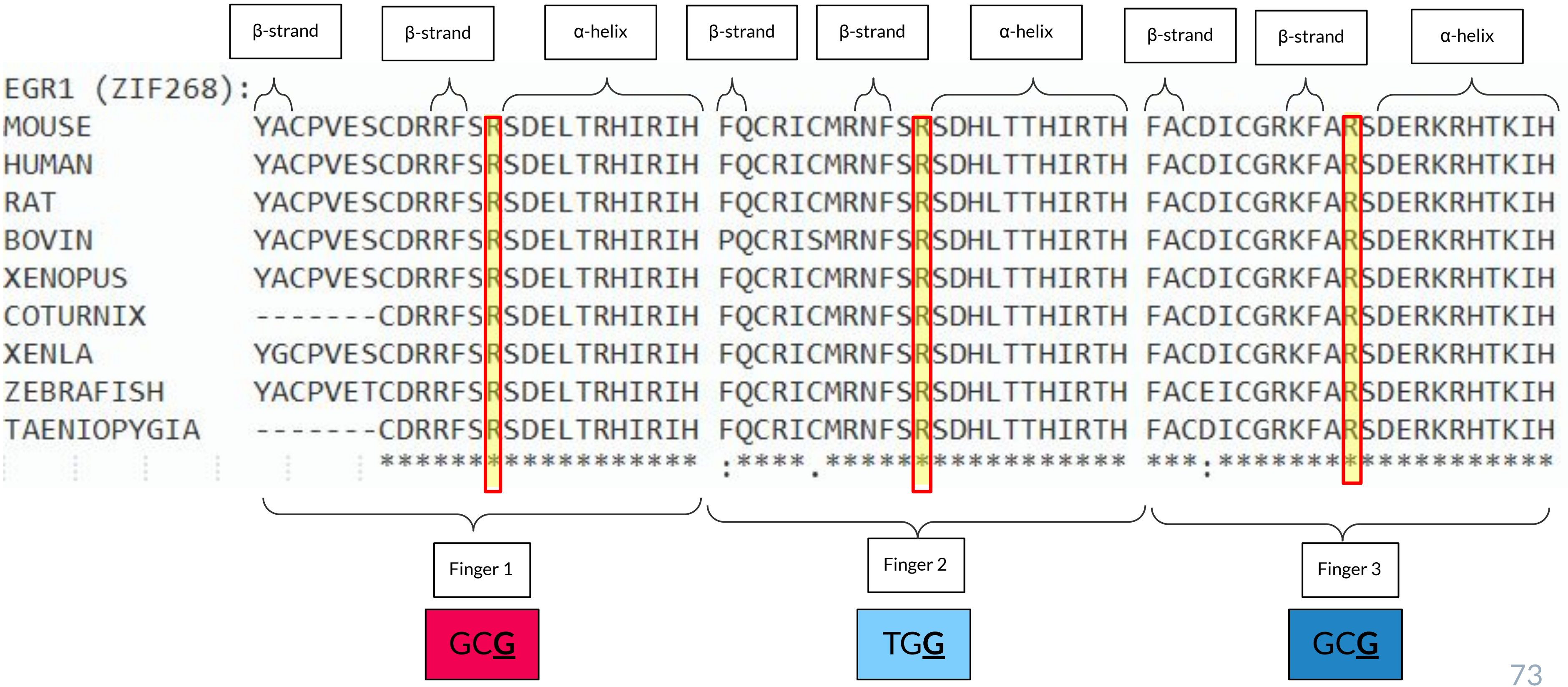
Finger 1 → Arg118, Asp120,Arg124 (positions -1,2,6)

Finger 2 → Arg146, Asp148, His149 (positions -1,2,3)

Finger 3 → Arg174, Asp176, Arg180 (positions -1,2,6)

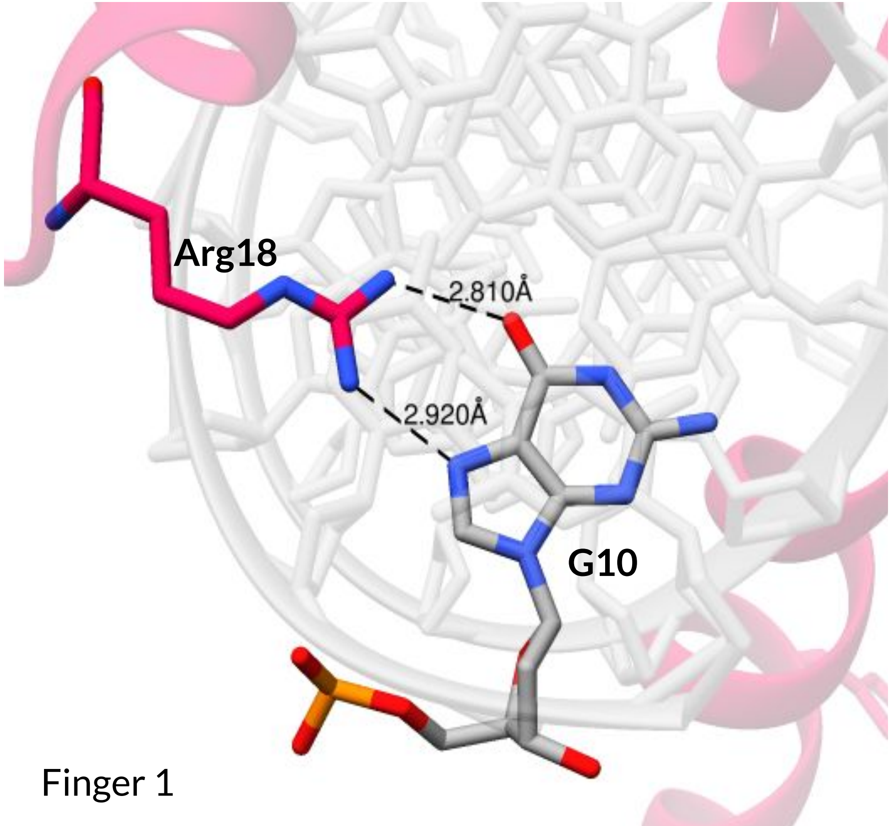
Position -1

Position -1 Position 2 Position 3 Position 6

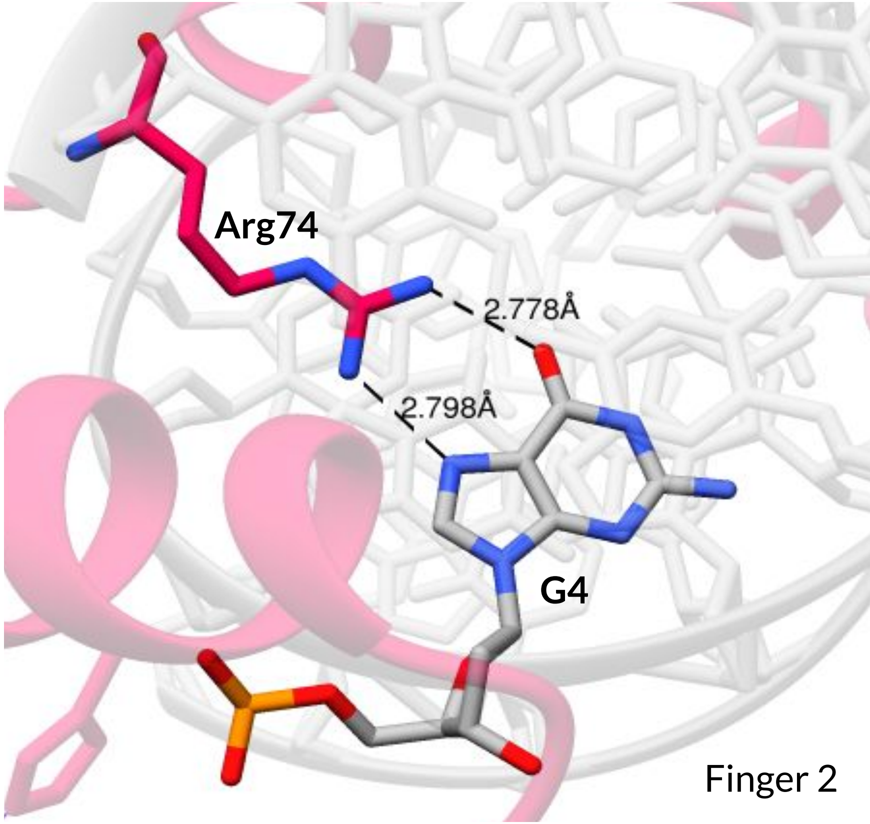


```
EGR1 (ZIF268):
MOUSE          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT            YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN          YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS        YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX       -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA          YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH      YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA    -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                      ****************** :****.***************** ***:*******************
```

Position -1

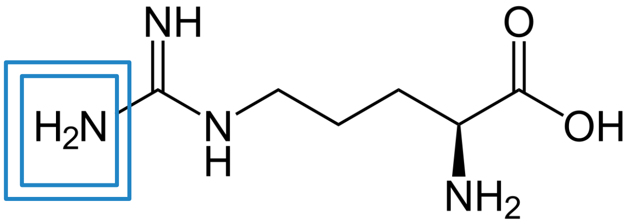


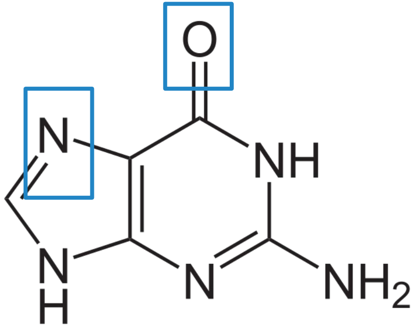
Finger 1



Finger 2

Position -1

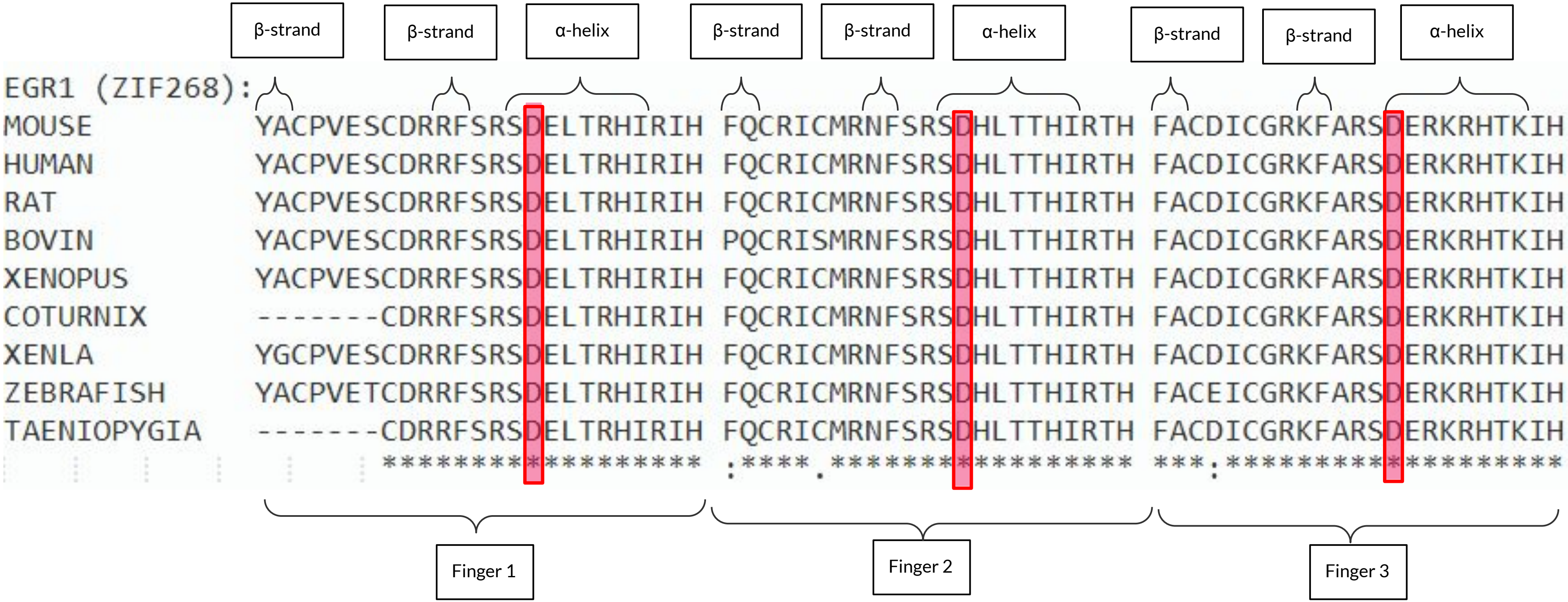

Arginine


Guanine

R18A mutation results in 100-fold loss affinity

Position 2

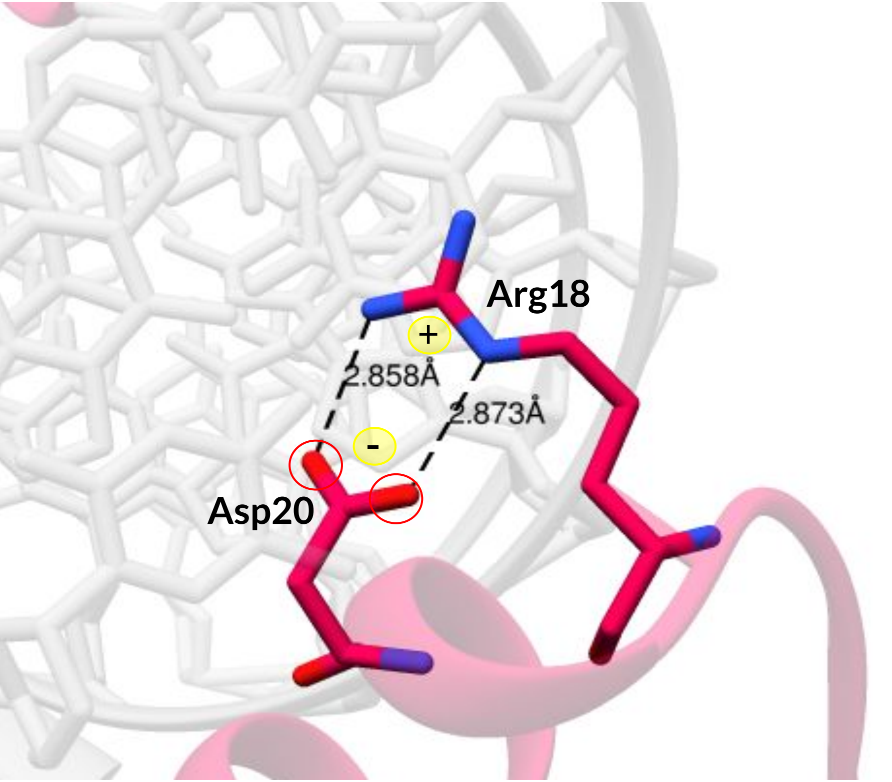
Position -1 **Position 2** Position 3 Position 6



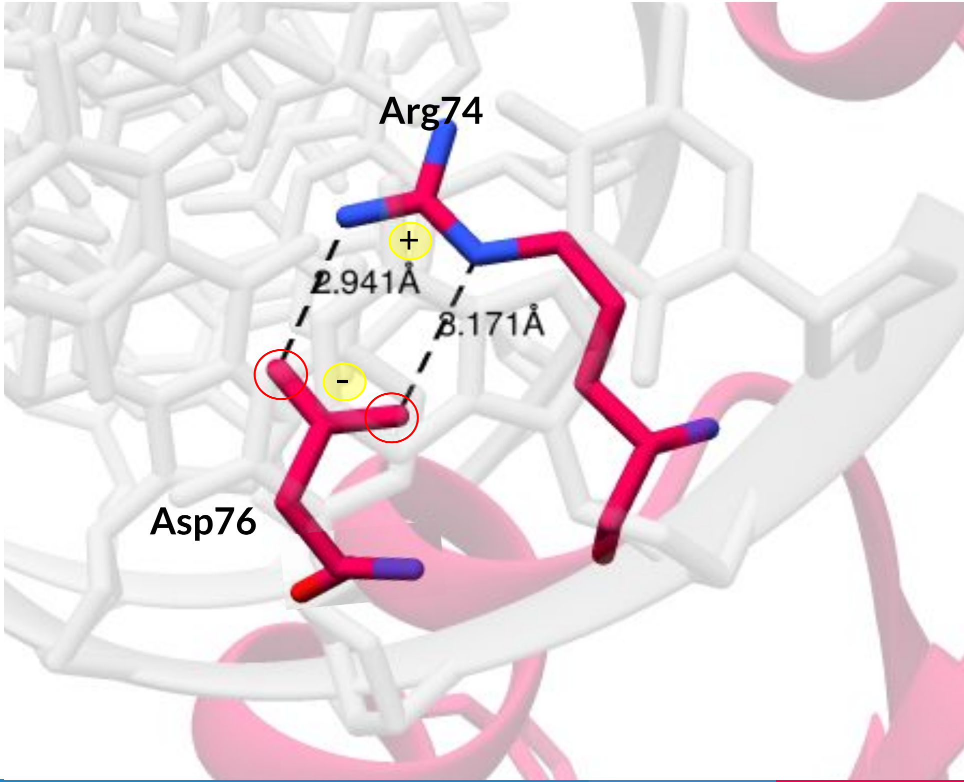
```
EGR1 (ZIF268):
MOUSE         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT           YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN         YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS       YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX      -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA         YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH     YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA   -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                     ****************** :****.***************** ***:*******************
```

Position 2: Interaction with position -1 residue

Finger 1

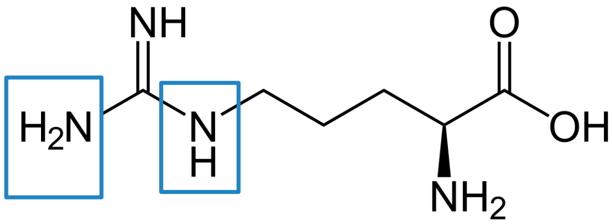


Finger 2

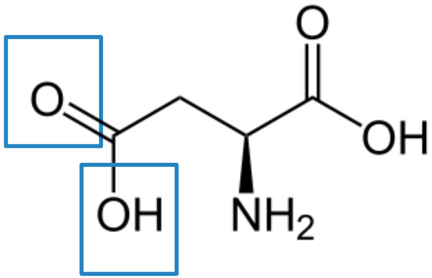


Position 2: Interaction with position -1 residue

Arginine

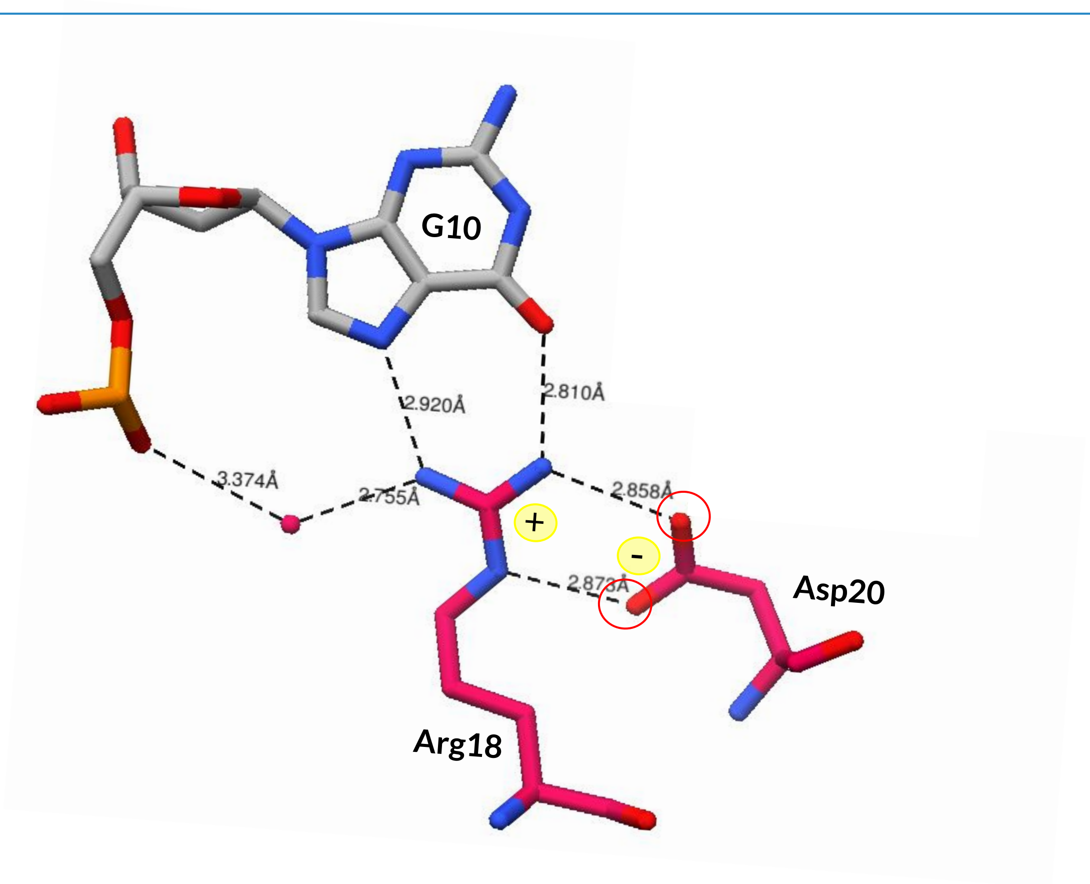


Aspartic



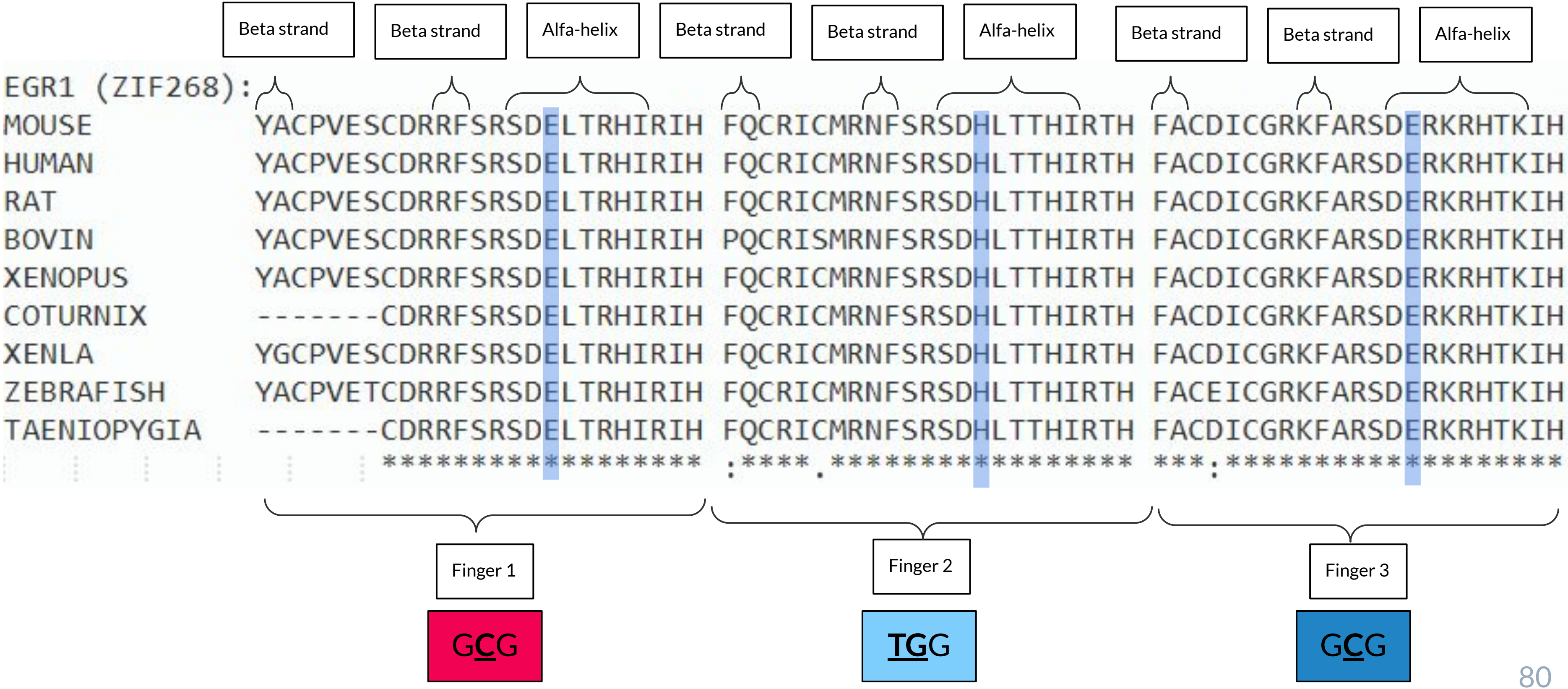
D20A mutation results in a reduced specificity for guanine

Interaction between position -1 and position 2



Position 3

Position -1 Position 2 **Position 3** Position 6



```
EGR1 (ZIF268):
MOUSE          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN          YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT            YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN          YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS        YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX       -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA          YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH      YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA    -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                      ****************** :****.***************** ***:*******************
```

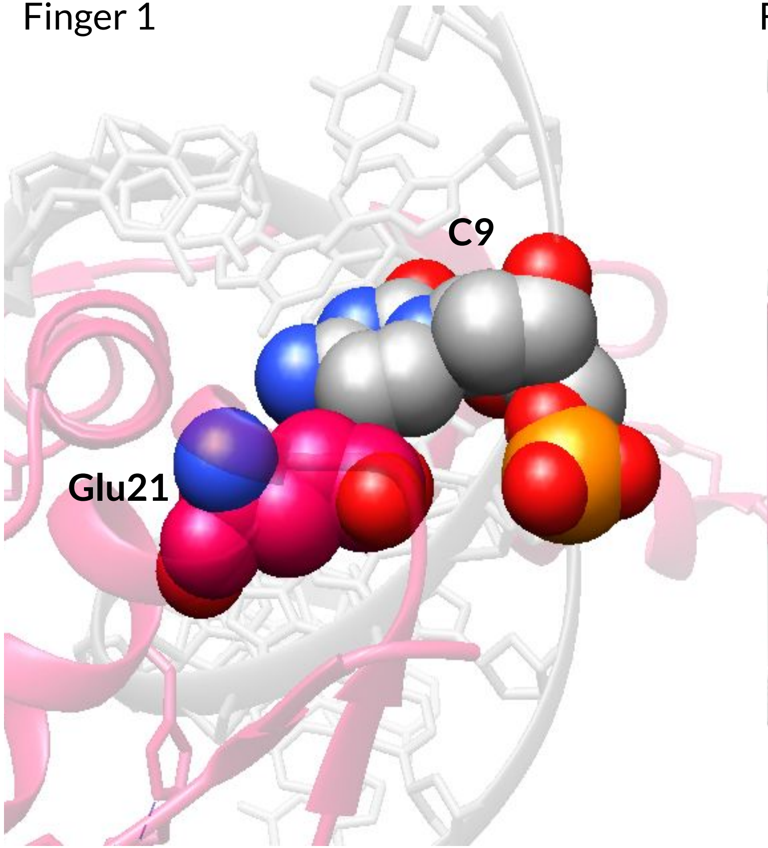
GCG

TGG

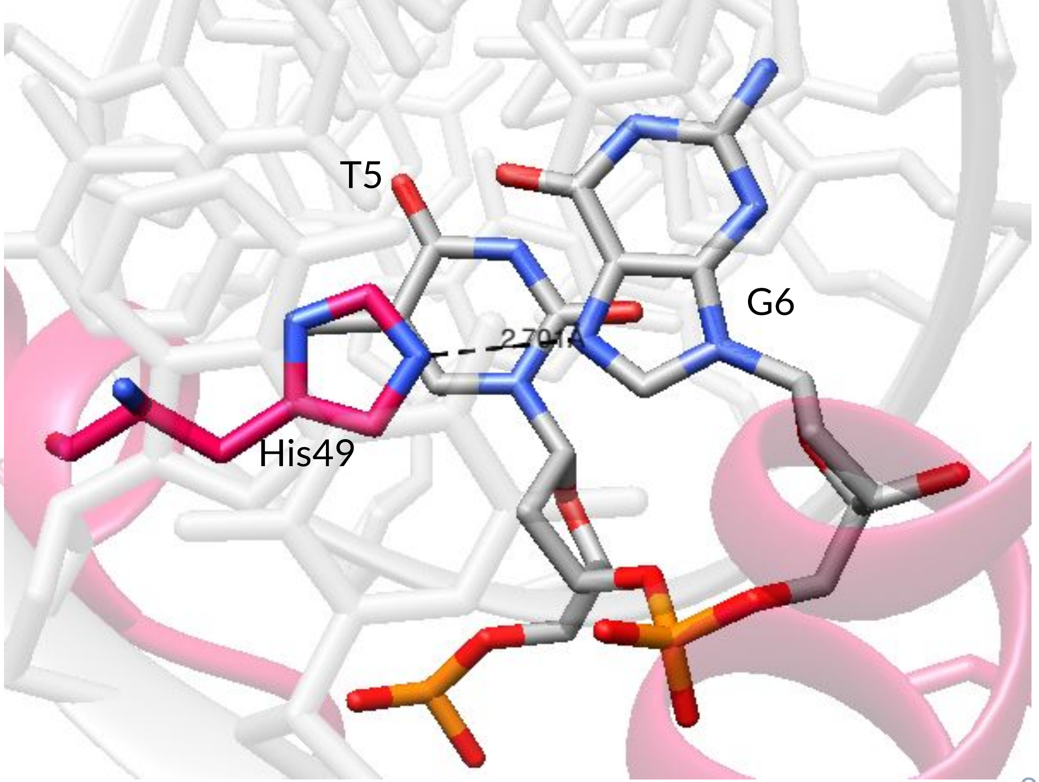
GCG

Position 3

Finger 1

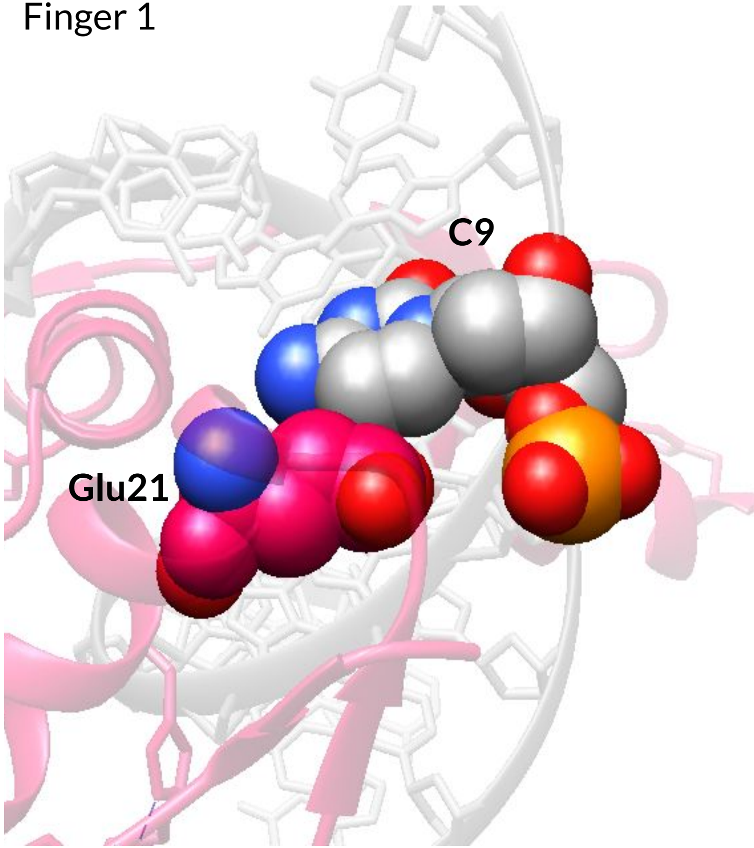


Finger 2

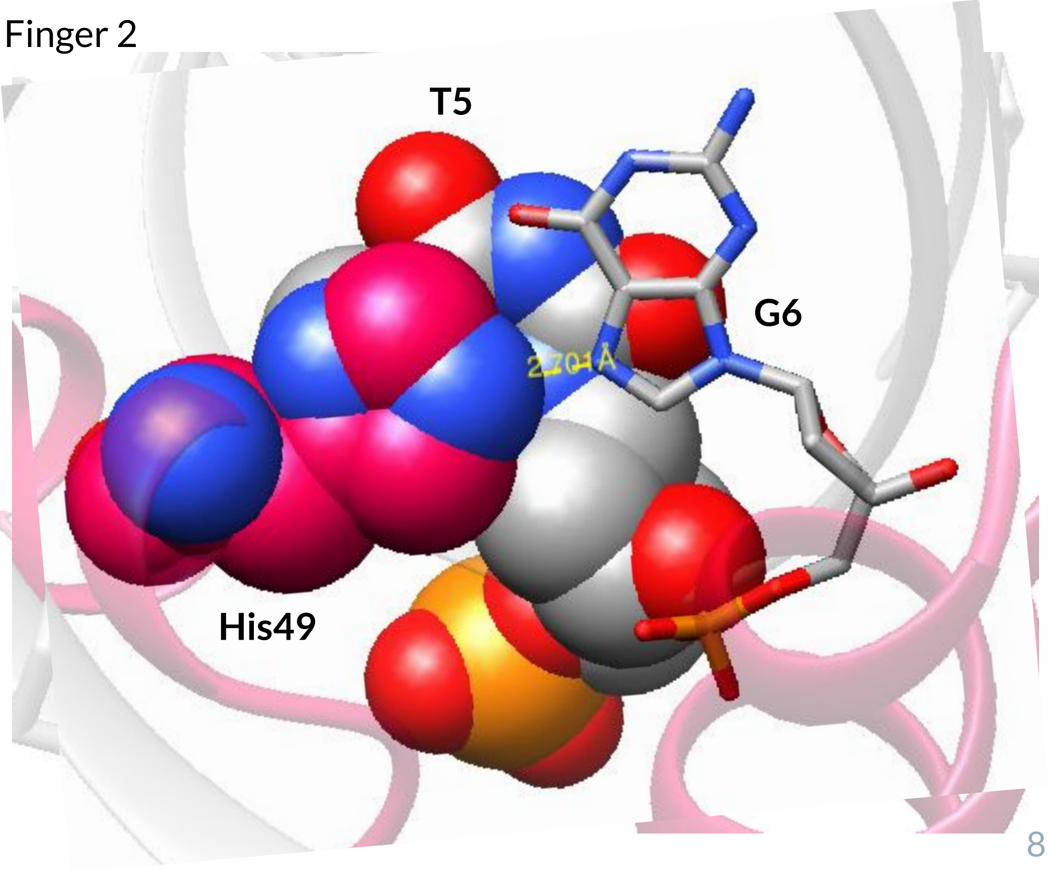


Position 3

Finger 1



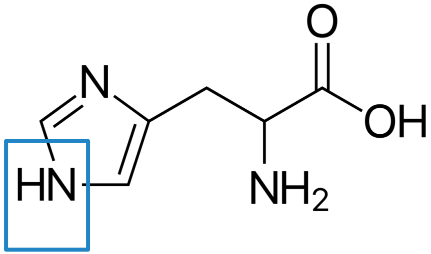
Finger 2



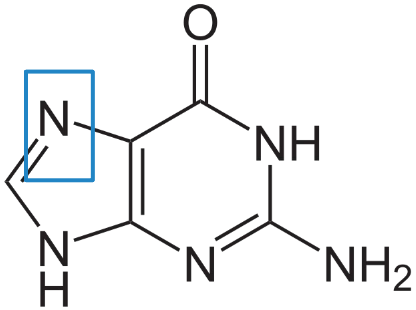
Position 3

H bond

Histidine

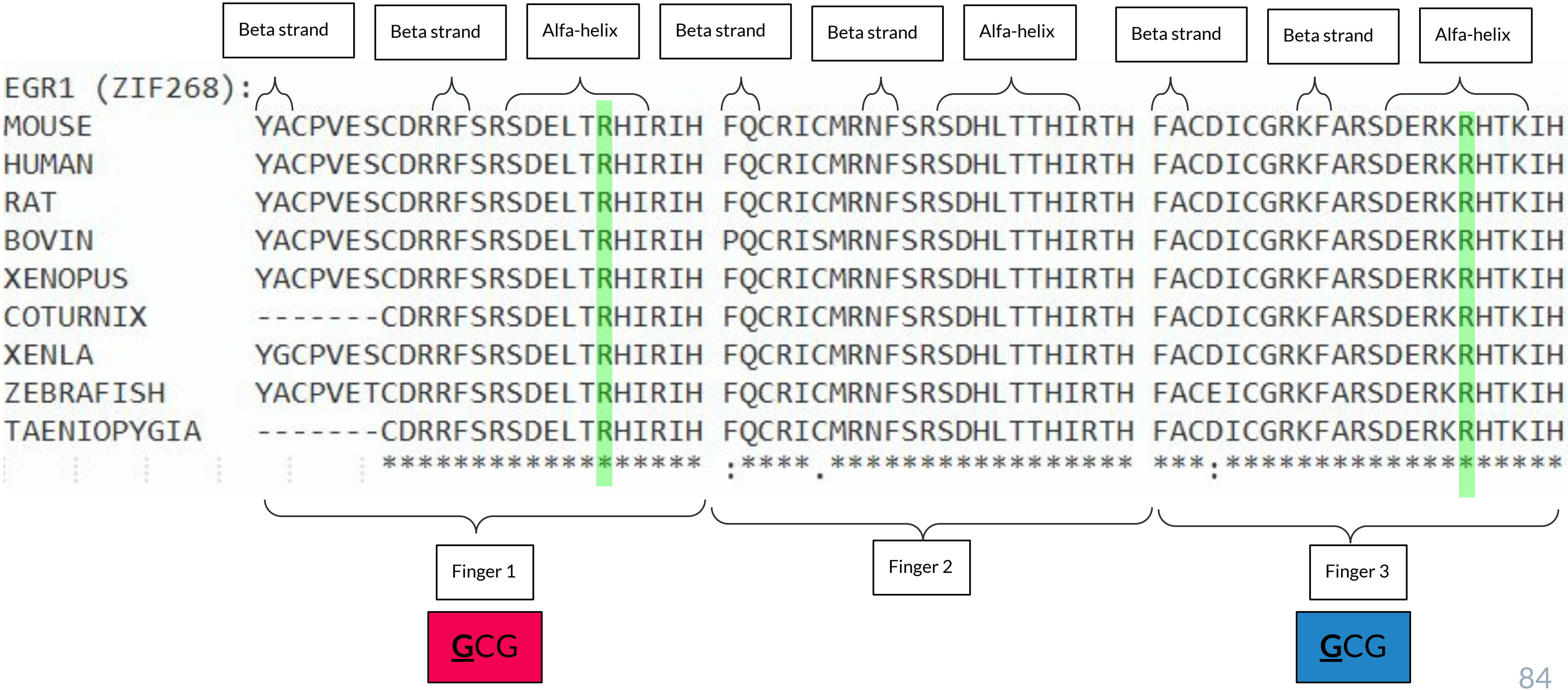


Guanine



Position 6

Position -1 Position 2 Position 3 **Position 6**



Beta strand

Beta strand

Alfa-helix

Beta strand

Beta strand

Alfa-helix

Beta strand

Beta strand

Alfa-helix

```
EGR1 (ZIF268):
MOUSE         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
HUMAN         YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
RAT           YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
BOVIN         YACPVESCDRRFSRSDELTRHIRIH PQCRISMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENOPUS       YACPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
COTURNIX      -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
XENLA         YGCPVESCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
ZEBRAFISH     YACPVETCDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACEICGRKFARSDERKRHTKIH
TAENIOPYGIA   -------CDRRFSRSDELTRHIRIH FQCRICMRNFSRSDHLTTHIRTH FACDICGRKFARSDERKRHTKIH
                     ****************** :****.***************** ***:*******************
```

Finger 1

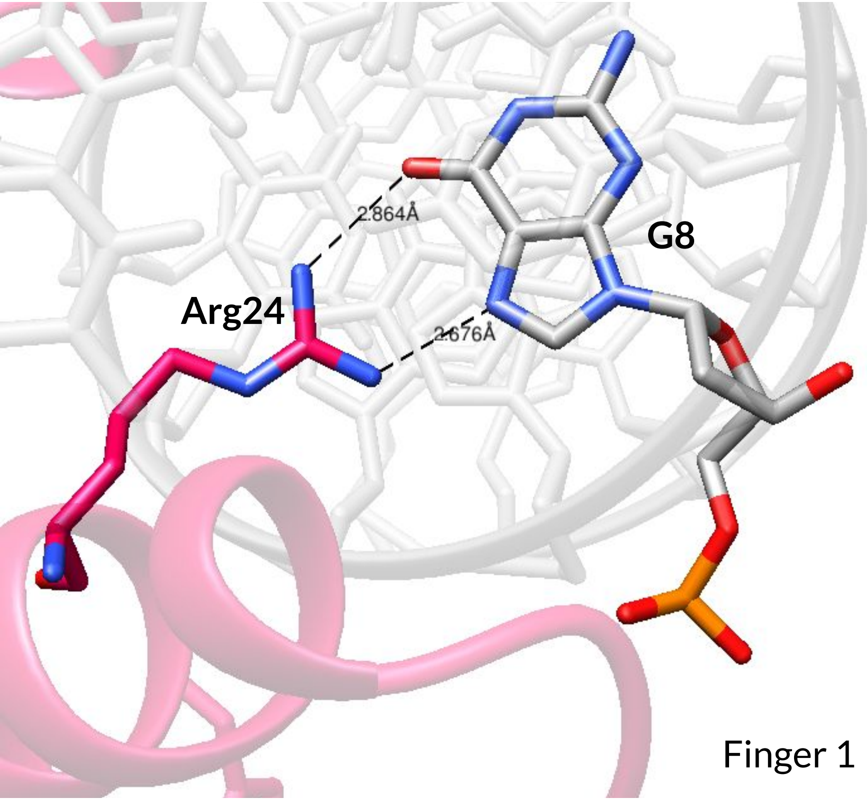
Finger 2

Finger 3

GCG

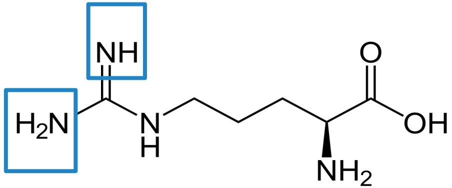
GCG

Position 6

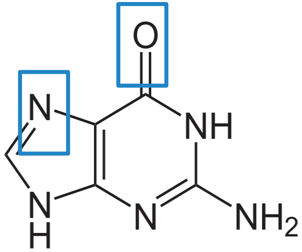


Finger 1

Arginine



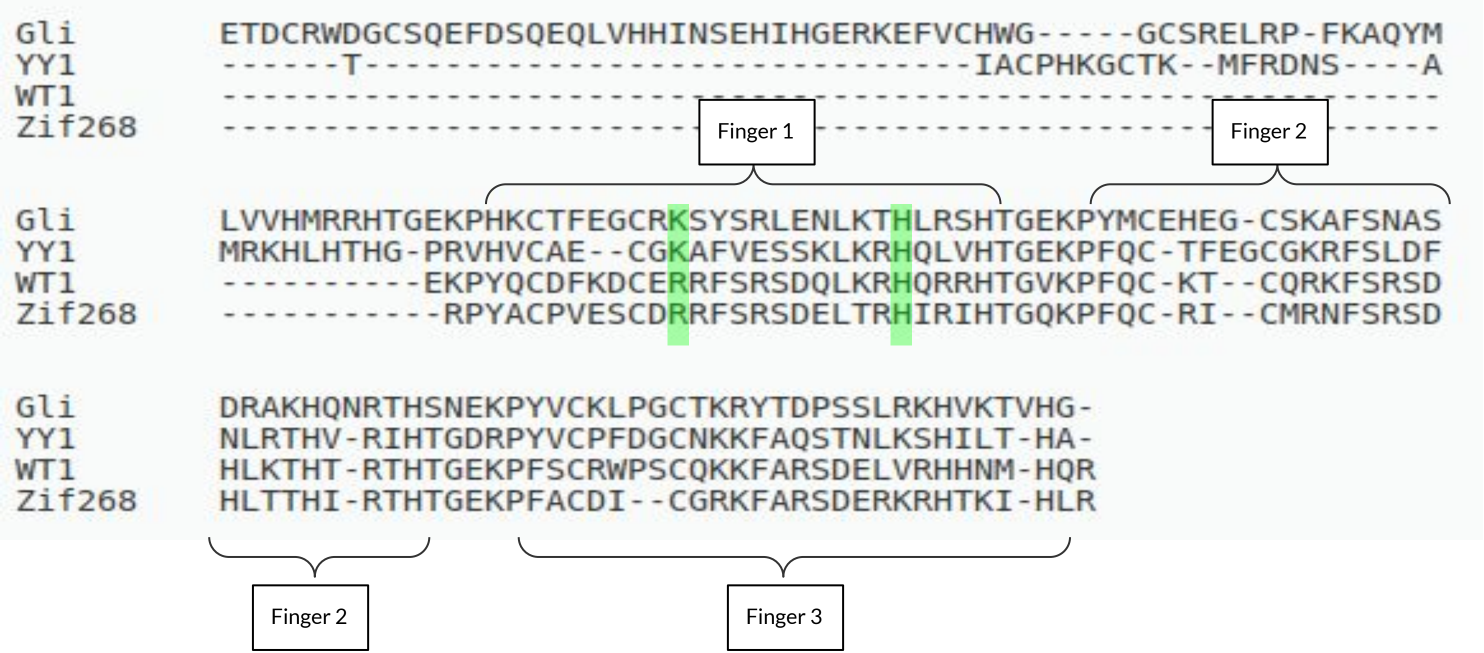
Guanine



Key residues

(non-specific DNA binding)

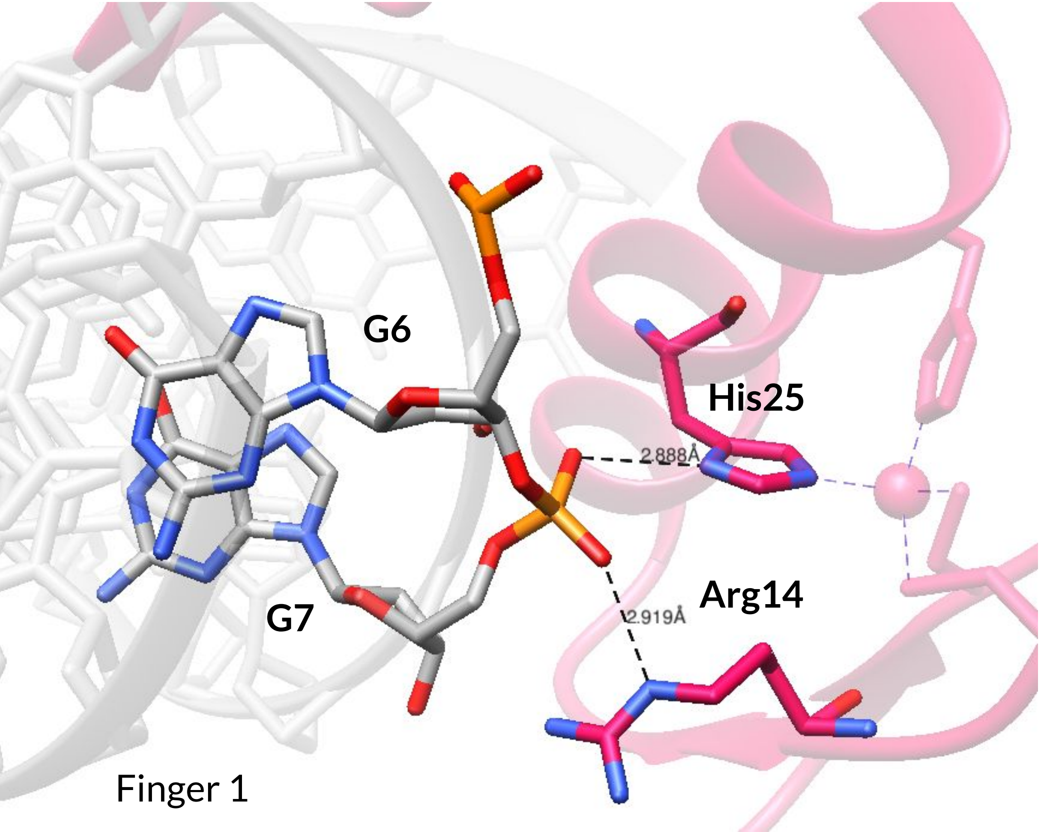
DNA non-specific interactions: Arg 14 and His25



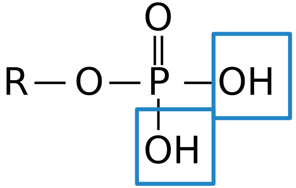
```
Gli      ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG-----GCSRELRP-FKAQYM
YY1      ------T------------------------------IACPHKGCTK--MFRDNS----A
WT1      ------------------------------------------------------------
Zif268   ------------------------------------------------------------

Gli      LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1      MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1      ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268   -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli      DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1      NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1      HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268   HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

Residue involved in non-specific interaction

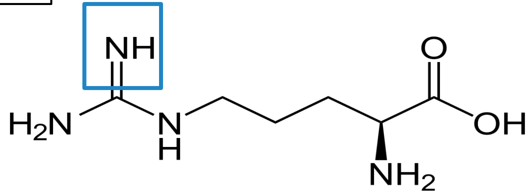
DNA non-specific interactions: Arg 14 and His25



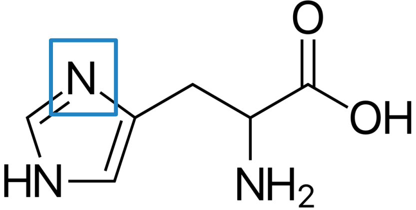
Phosphate group



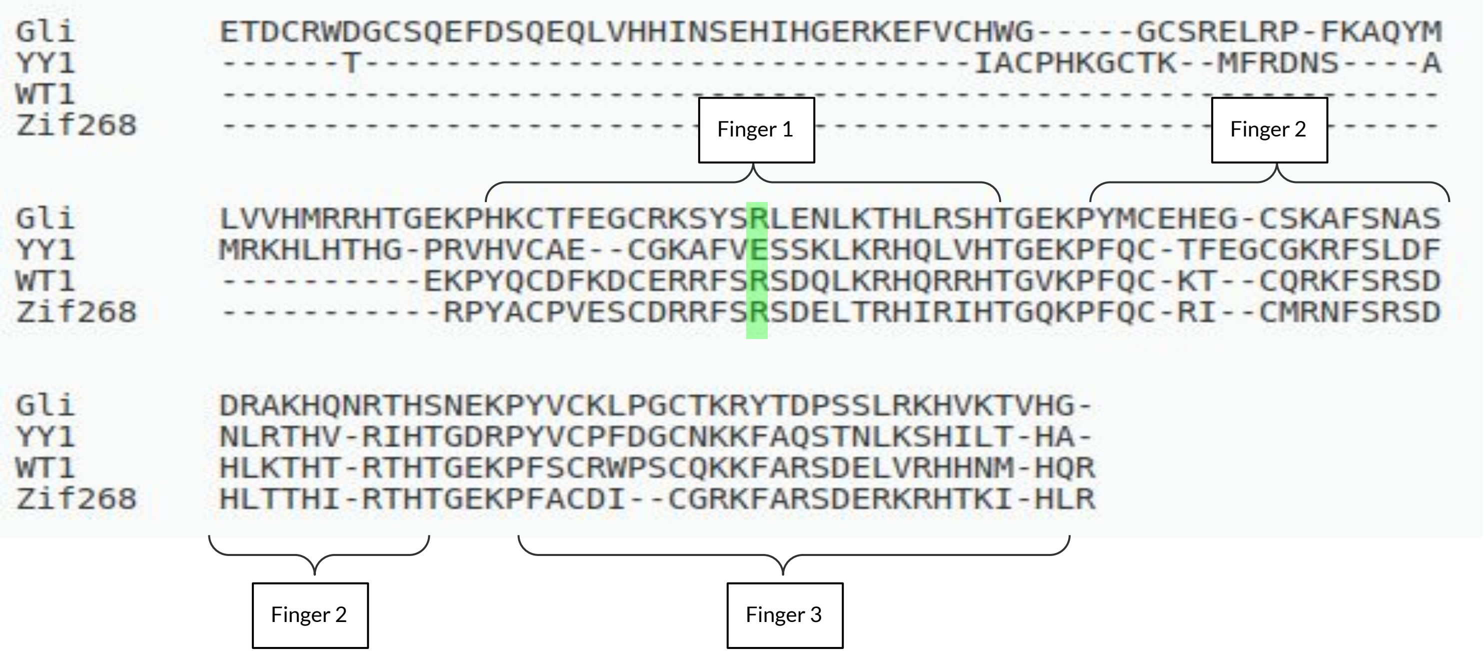
Arginine



Histidine



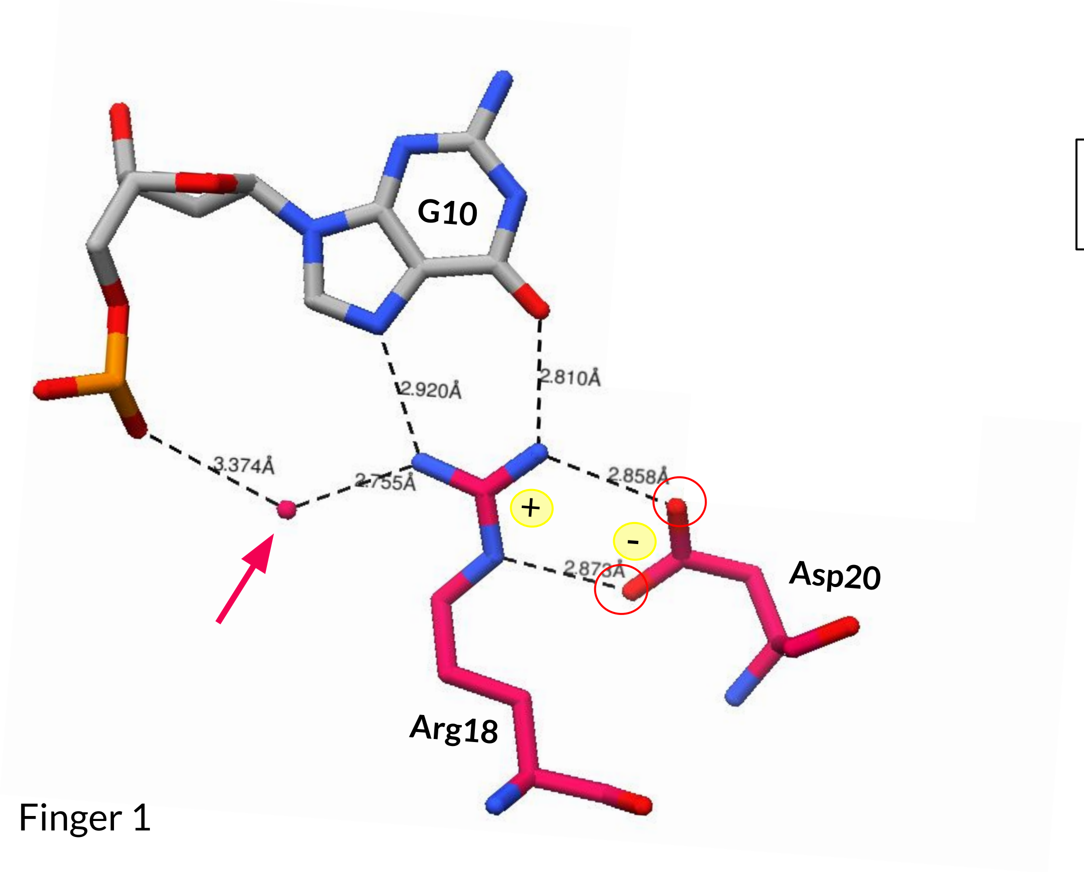
DNA non-specific Interactions: Arg 18



```
Gli       ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG-----GCSRELRP-FKAQYM
YY1       ------T------------------------------IACPHKGCTK--MFRDNS----A
WT1       ------------------------------------------------------------
Zif268    ------------------------------------------------------------

Gli       LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1       MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1       ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268    -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli       DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1       NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1       HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268    HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

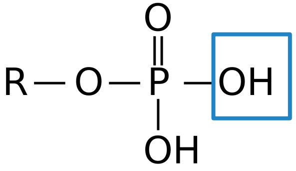
Residue involved in non-specific interaction

DNA nos-specific interactions: Arg 18

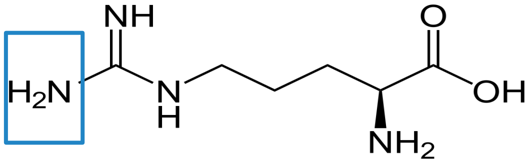


Finger 1

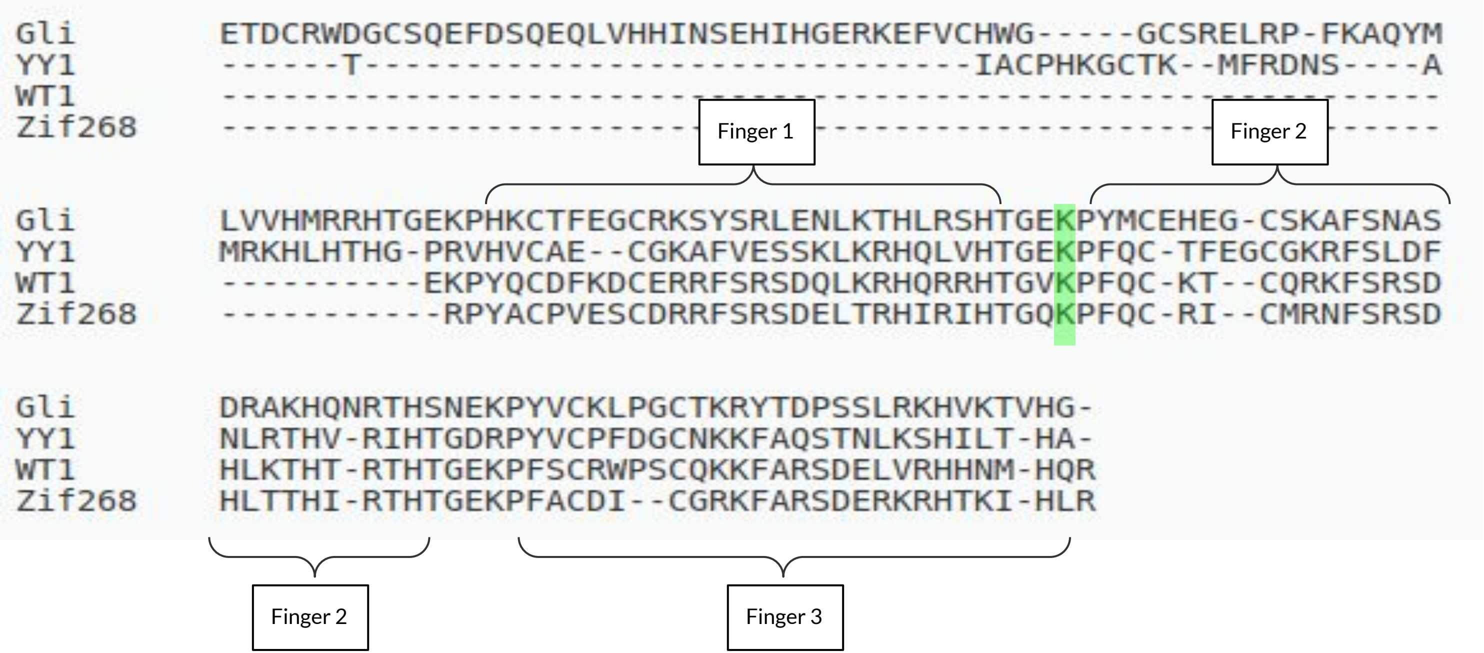
Phosphate group



Arginine



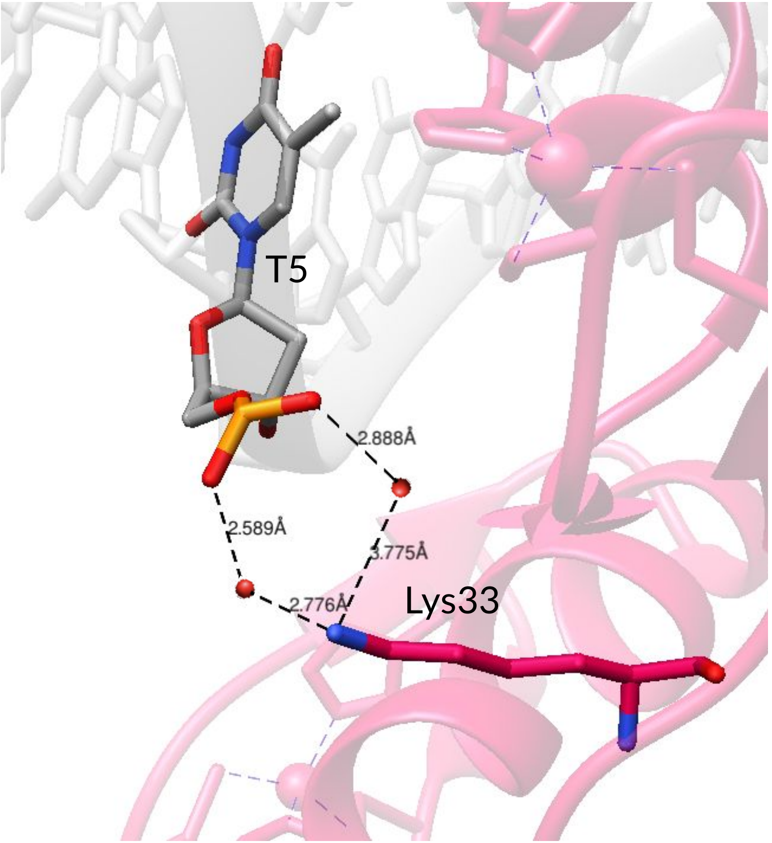
DNA non-specific interactions: Lys33



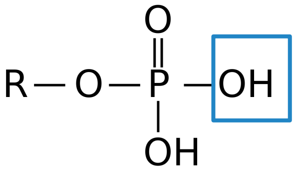
```
Gli       ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG-----GCSRELRP-FKAQYM
YY1       ------T------------------------------IACPHKGCTK--MFRDNS----A
WT1       ------------------------------------------------------------
Zif268    ------------------------------------------------------------

Gli       LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1       MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1       ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268    -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli       DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1       NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1       HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268    HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

Residue involved in non-specific interaction

DNA non-specific Interactions: Lys33

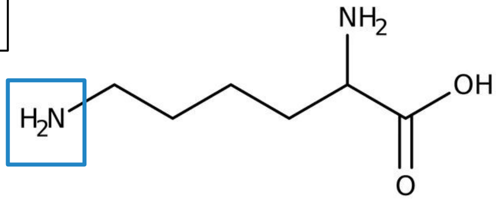


Phosphate group



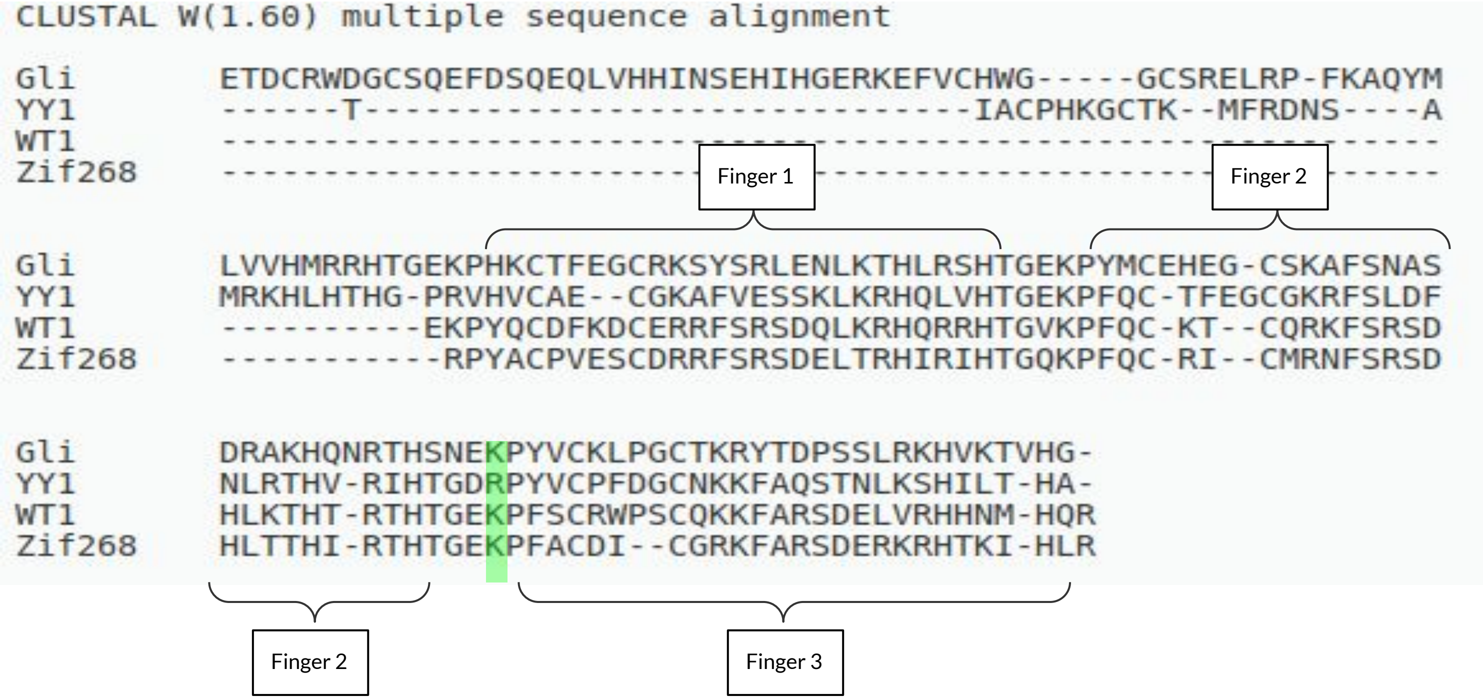
$R-O-P(=O)(OH)-OH$

Lysine



Linker region between finger 1 and finger 2

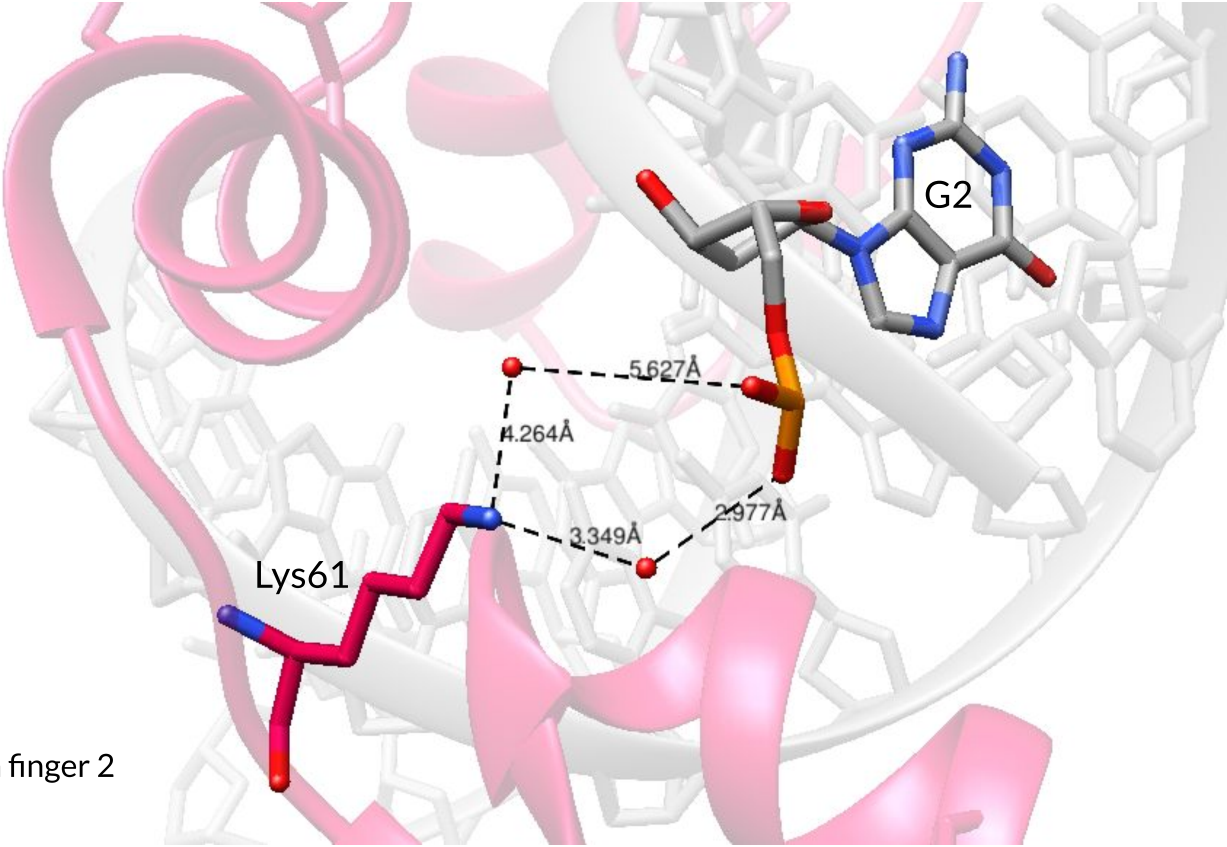
DNA non-specific interactions: Lys61



```
CLUSTAL W(1.60) multiple sequence alignment

Gli       ETDCRWDGCSQEFDSQEQLVHHINSEHIHGERKEFVCHWG----GCSRELRP-FKAQYM
YY1       ------T-----------------------------IACPHKGCTK--MFRDNS----A
WT1       -----------------------------------------------------------
Zif268    -----------------------------------------------------------

Gli       LVVHMRRHTGEKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCEHEG-CSKAFSNAS
YY1       MRKHLHTHG-PRVHVCAE--CGKAFVESSKLKRHQLVHTGEKPFQC-TFEGCGKRFSLDF
WT1       ----------EKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC-KT--CQRKFSRSD
Zif268    -----------RPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC-RI--CMRNFSRSD

Gli       DRAKHQNRTHSNEKPYVCKLPGCTKRYTDPSSLRKHVKTVHG-
YY1       NLRTHV-RIHTGDRPYVCPFDGCNKKFAQSTNLKSHILT-HA-
WT1       HLKTHT-RTHTGEKPFSCRWPSCQKKFARSDELVRHHNM-HQR
Zif268    HLTTHI-RTHTGEKPFACDI--CGRKFARSDERKRHTKI-HLR
```

Residue involved in non-specific interaction

DNA non-specific Interactions: Lys61



Linker region between finger 2 and finger 3

Conclusions

Conclusions

Structure	MyoD and Zif268, which are structurally different can perform a similar function as transcription factors and bind to DNA.
Conservation	Conservation of HLH and ZnF domains is high, demonstrating the huge importance of them for their function as transcription Factors.
Hydrophobic and basic residues	Hydrophobic residues are essential for protein stability and basic residues are highly important for DNA-binding.
DNA-binding specificity	These transcription factors have different DNA binding specificity since the sequences interacting are not the same.
Stability and function	Stabilization of its structure is required in order to perform their function of DNA binding.

Thank you for your attention!

Pem questions

1. Select the correct answers about MyoD:

1. According to SCOP it belongs to the helix loop helix family.
2. It binds specifically to the E-box (CAGCTG)
3. Some hydrophobic residues help to stabilize the protein.
4. Its helix 2 is the one that binds to the DNA.
    - a) 2 and 4
    - b) 1, 2 and 3
    - c) 1, 2, 3 and 4
    - d) 4
    - e) None the above

2. b/HLH proteins like MyoD bind to DNA through the region:

- a. Helix 1 (H1)
- b. Basic region
- c. Helix 2 (H2)
- d. Loop
- e. All of the above

3. In transcription factors, the contacts made with the phosphates of the DNA are:

- a. Specific contacts
- b. π stacking
- c. Unspecific contacts
- d. Water mediated contacts
- e. Hydrophobic contacts

4. The binding domain of MyoD recognizes the:

- a. Major groove of the DNA.
- b. Minor groove of the DNA
- c. Both of the above are correct
- d. An specific region of minor groove of the DNA
- e. All of them are correct.

5. Select the <u>WRONG</u> answer/s about the conservation of MyoD:

1. Basic residues needed for DNA-binding in helix 1 are highly conserved among species and other bHLH proteins
2. Hydrophobic residues are not conserved among species because they are not needed for the proper binding to the DNA
3. The myogenic code is conserved among the Myogenic Regulatory Factor family (MRFs)
4. Residues located in the loop are highly conserved among other members of the bHLH family of transcription factors.
    - a) 2 and 4
    - b) 1, 2 and 3
    - c) 1, 2, 3 and 4
    - d) 4
    - e) None the above

Pem questions

6.Select the correct answers about Zif268:

1. According to SCOP it belongs to the helix loop helix family.
2. It has 4 zinc fingers
3. It is involved in the transcriptional activation of numerous muscle-specific genes
4. It is a transcription factor

    a) 1,2,3,4
    b) 1,2,3
    c) 1 y 3
    d) 2 y 4
    e) 4

7.Which of the following are specific interactions of Zif268 with DNA? Indicate the true statement.

a) Arg 18 of finger 1 with guanine
b) Lys33 with the 5′ phosphate of base 5
c) Cys17 with Zn ion
d) Lys61 with the 5’ phosphate of base 2
e) Arg98 with 5’ phosphate of base 3

8.Which is the consensus sequence of C2H2 family?

a) .-Z-X_{1-2}-F-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-C
b) -Z-X_{1-2}-F-X_{2-4}-C-X_3-Z-X_5-Z-X_2-C-X_{3-5}-H
c) -Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-Y-X_{3-5}-H
d) -Z-X_{1-2}-C-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H
e) -Z-X_{1-2}-G-X_{2-4}-C-X_3-Z-X_5-Z-X_2-H-X_{3-5}-H

9.Which are the most different zinc fingers regarding the DNA binding specificity?

a. Finger 1 and 3
b. Finger 1 and finger 5
c. Finger 1 and finger 2
d. Finger 4 and finger 3
e. All of them are exactly the same

10.Regarding the hydrophobic core of Zif268, select the false statement:

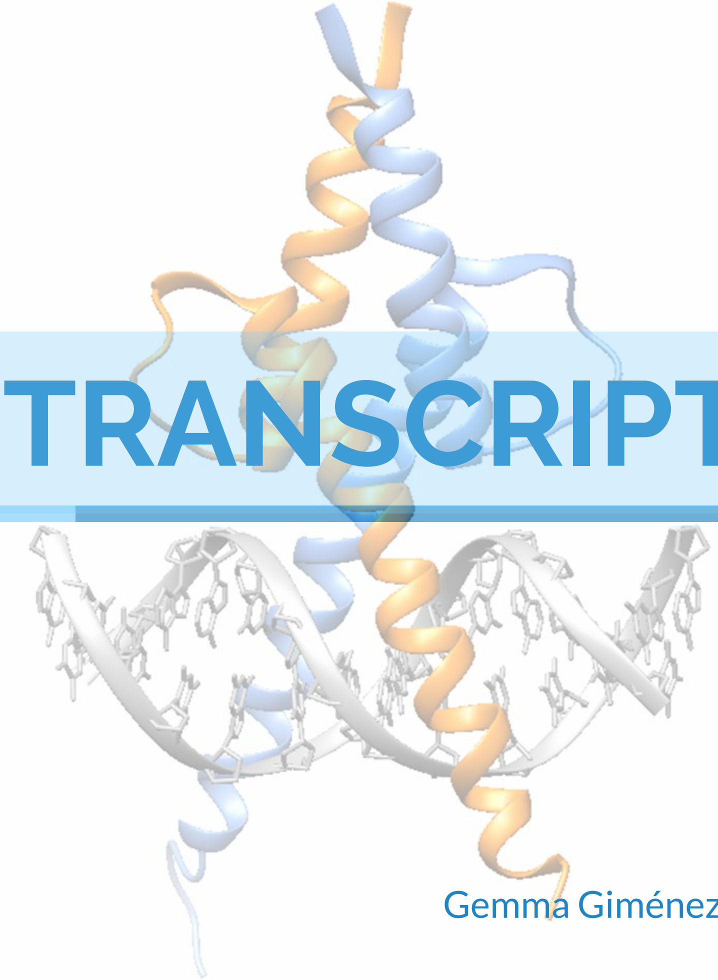
a. It allows the stabilization of the zinc finger structure
b. The residues involved are highly conserved
c. In the three zinc fingers the residues forming the hydrophobic core are exactly the same.
d. Tyr05, Phe16 and Leu22 are some residues forming the hydrophobic core.
e. All of the above are true

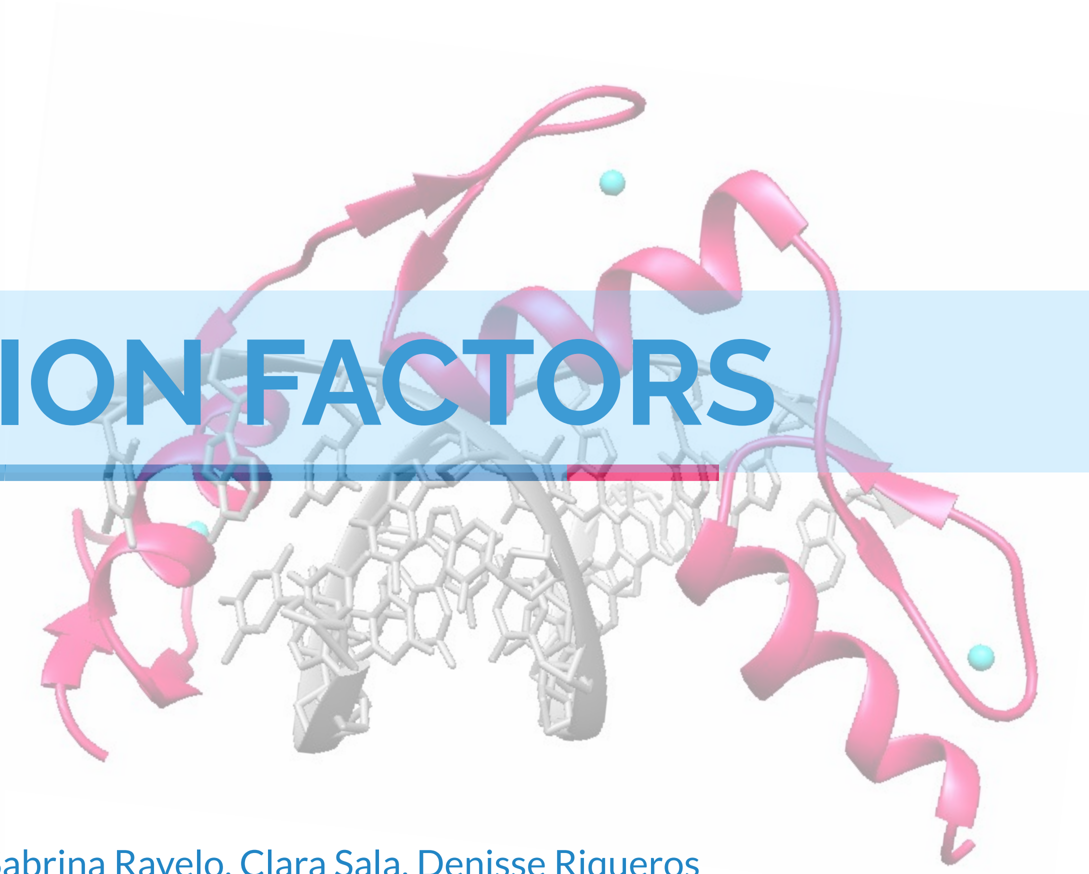
Bibliography

1. Branden C, Tooze J. Introduction to protein structure. 2a ed. New York: Garland; 1999.
2. Finkelstein AV, Ptitsyn OB. Protein Physics: A Course of Lectures. San Diego: Elsevier Science & Technology; 2002.
3. Creighton TE. The Biophysical chemistry of nucleic acids & proteins. S.l: Helvetian Press; 2010.
4. Murre C. Helix-loop-helix proteins and the advent of cellular diversity: 30 years of discovery. *Genes Dev.* 2019;33(1-2):6-25.
5. Phillips SE. Built by association: structure and function of helix-loop-helix DNA-binding proteins. *Structure*. 1994;2(1):1-4.
6. Weirauch MT, Hughes TR. A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell Biochem*. 2011;52:25-73.
7. Pérez Sirkin DI, Lafont AG, Kamech N, Somoza GM, Vissio PG, Dufour S. Conservation of Three-Dimensional Helix-Loop-Helix Structure through the Vertebrate Lineage Reopens the Cold Case of Gonadotropin-Releasing Hormone-Associated Peptide. *Front Endocrinol (Lausanne)*. 2017;8:207.
8. Ma PC, Rould MA, Weintraub H, Pabo CO. Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*. 1994;77(3):451-9.
9. Grove CA, De Masi F, Barrasa MI, Newburger DE, Alkema MJ, Bulyk ML, Walhout AJ. A multiparameter network reveals extensive divergence between C. elegans bHLH transcription factors. *Cell*. 2009;138(2):314-27.
10. Shirakata M, Friedman FK, Wei Q, Paterson BM. Dimerization specificity of myogenic helix-loop-helix DNA-binding factors directed by nonconserved hydrophilic residues. *Genes Dev*. 1993;7(12A):2456-70.
11. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009;10(4):252-63.
12. Jones S. An overview of the basic helix-loop-helix proteins. *Genome Biol*. 2004;5(6):226.
13. Murre C, Bain G, van Dijk MA, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR, Stuiver MH. Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta*. 1994;1218(2):129-35.

Bibliography

14. Davis RL, Cheng PF, Lassar AB, Weintraub H. The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell*. 1990;60(5):733-46.
15. Wendt H, Thomas RM, Ellenberger T. DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem*. 1998;273(10):5735-43.
16. Hernández-Hernández JM, García-González EG, Brun CE, Rudnicki MA. The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin Cell Dev Biol*. 2017;72:10-18.
17. Rebar, E. J., & Pabo, C. O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*. 1994:263(5147), 671-673.
18. Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., ... & Raschellà, G. Zinc-finger proteins in health and disease. *Cell death discovery*. 2017;3(1):1-12.
19. Laity, J. H., Lee, B. M., & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*. 2001;11(1):39-46.
20. Wang, B., Guo, H., Yu, H., Chen, Y., Xu, H., & Zhao, G. The role of the transcription factor EGR1 in cancer. *Frontiers in Oncology*. 2021;*11*:775.
21. Rebar, E. J., & Pabo, C. O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*. 1994;*263*(5147):671-673.
22. Pace, N. J., & Weerapana, E. Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules*. 2014;*4*(2):419-434.
    Negi, S., Imanishi, M., Sasaki, M., Tatsutani, K., Futaki, S., & Sugiura, Y. An arginine residue instead of a conserved leucine residue in the recognition helix of the finger 3 of Zif268 stabilizes the domain structure and mediates DNA binding. *Biochemistry*. 2011;*50*(28):6266-6272.
23. Elrod-Erickson, M., Rould, M. A., Nekludova, L., & Pabo, C. O. Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger–DNA interactions. *Structure*. 1996;*4*(10):1171-1180.
24. Pavletich, N. P., & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*. 1991;*252*(5007):809-817.



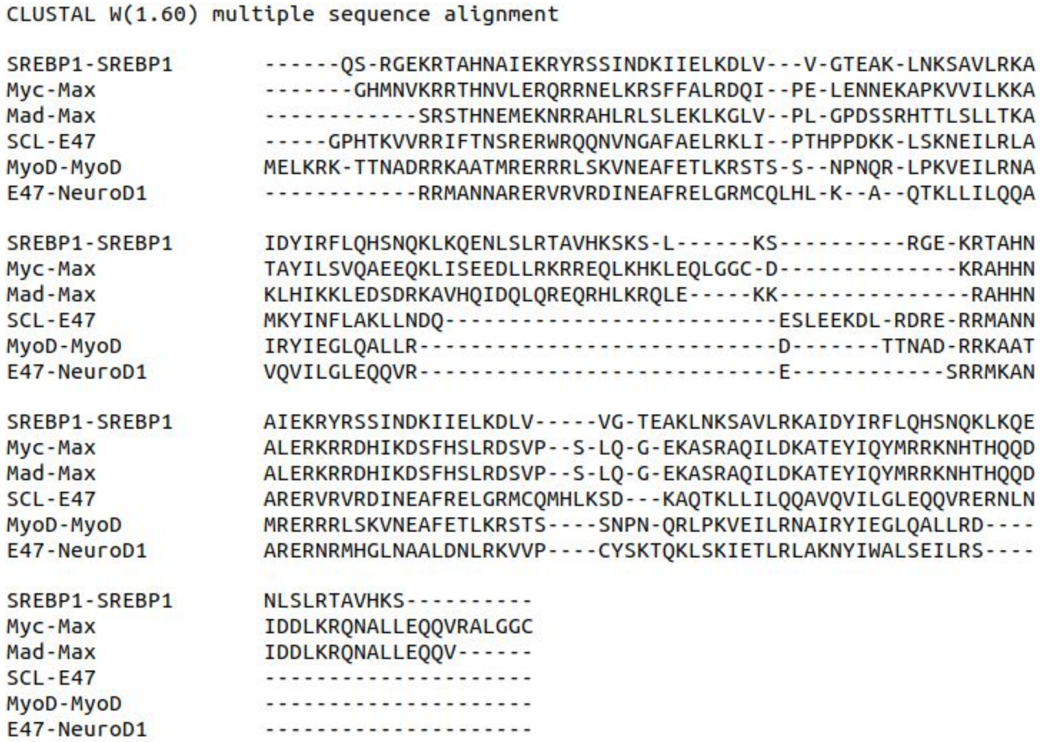


TRANSCRIPTION FACTORS

Gemma Giménez, Sabrina Ravelo, Clara Sala, Denisse Riqueros
Structural Biology
March 2022

ANNEX: Same family structures

Structural alignment (dimers)



```
CLUSTAL W(1.60) multiple sequence alignment

SREBP1-SREBP1       ------QS-RGEKRTAHNAIEKRYRSSINDKIIELKDLV---V-GTEAK-LNKSAVLRKA
Myc-Max             -------GHMNVKRRTHNVLERQRRNELKRSFFALRDQI--PE-LENNEKAPKVVILKKA
Mad-Max             ------------SRSTHNEMEKNRRAHLRLSLEKLKGLV--PL-GPDSSRHTTLSLLTKA
SCL-E47             -----GPHTKVVRRIFTNSRERWRQQNVNGAFAELRKLI--PTHPPDKK-LSKNEILRLA
MyoD-MyoD           MELKRK-TTNADRRKAATMRERRRLSKVNEAFETLKRSTS-S--NPNQR-LPKVEILRNA
E47-NeuroD1         ------------RRMANNARERVRVRDINEAFRELGRMCQLHL-K--A--QTKLLILQQA

SREBP1-SREBP1       IDYIRFLQHSNQKLKQENLSLRTAVHKSKS-L------KS----------RGE-KRTAHN
Myc-Max             TAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC-D--------------KRAHHN
Mad-Max             KLHIKKLEDSDRKAVHQIDQLQREQRHLKRQLE-----KK---------------RAHHN
SCL-E47             MKYINFLAKLLNDQ--------------------------ESLEEKDL-RDRE-RRMANN
MyoD-MyoD           IRYIEGLQALLR----------------------------D-------TTNAD-RRKAAT
E47-NeuroD1         VQVILGLEQQVR----------------------------E------------SRRMKAN

SREBP1-SREBP1       AIEKRYRSSINDKIIELKDLV-----VG-TEAKLNKSAVLRKAIDYIRFLQHSNQKLKQE
Myc-Max             ALERKRRDHIKDSFHSLRDSVP--S-LQ-G-EKASRAQILDKATEYIQYMRRKNHTHQQD
Mad-Max             ALERKRRDHIKDSFHSLRDSVP--S-LQ-G-EKASRAQILDKATEYIQYMRRKNHTHQQD
SCL-E47             ARERVRVRDINEAFRELGRMCQMHLKSD---KAQTKLLILQQAVQVILGLEQQVRERNLN
MyoD-MyoD           MRERRRLSKVNEAFETLKRSTS----SNPN-QRLPKVEILRNAIRYIEGLQALLRD----
E47-NeuroD1         ARERNRMHGLNAALDNLRKVVP----CYSKTQKLSKIETLRLAKNYIWALSEILRS----

SREBP1-SREBP1       NLSLRTAVHKS----------
Myc-Max             IDDLKRQNALLEQQVRALGGC
Mad-Max             IDDLKRQNALLEQQV------
SCL-E47             ---------------------
MyoD-MyoD           ---------------------
E47-NeuroD1         ---------------------
```