TRANSCRIPTION FACTORS

Structural Biology, 2024
Mar Castaño, Mariona Colomer and Francina Barceló

Table of contents

Introduction

Transcription factors (TFs) are proteins that bind to specific DNA-regulatory sequences in order to **control gene expression**.

This could lead to changes in gene transcription and protein synthesis, which would then impact cellular function. Their activity is essential to determine function and respond to cellular environments

Introduction

Transcription factors are classified based on the characteristics of their DNA-binding domains.

The three most abundant superclasses have in common that an α-helix is exposed in such a way that it binds into the major groove of the DNA:

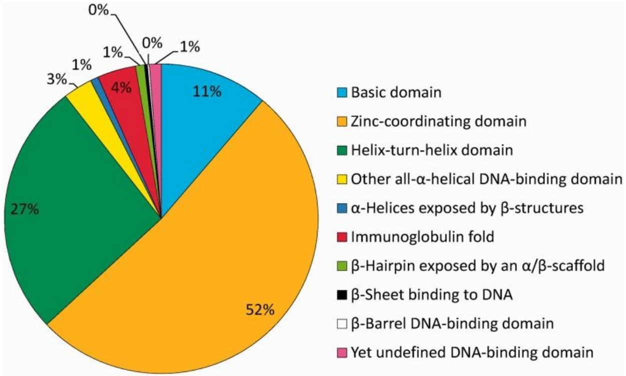


Fig1. Superclass distribution of human TF genes

Wingender E, Schoeps T, Haubrock M, Dönitz J. TFClass: a classification of human transcription factors and their rodent orthologs. Nucleic Acids Res. 2015 Jan;43(Database issue):D97-102. doi: 10.1093/nar/gku1064. Epub 2014 Oct 31. PMID: 25361979; PMCID: PMC4383905.

Basic domains

Zinc-coordinating DNA-binding domains

Helix-turn-helix domains

c-JUN

Basic domains

Basic region:

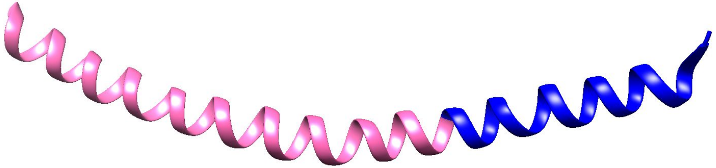
- Free in solution has a random structure
- Alpha-helically folded upon binding to DNA.

Class:

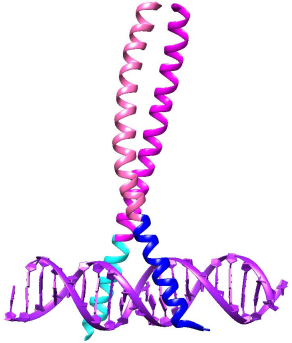
Basic leucine zipper factors (bZIP)
Fos, Jun, CREB, C/EBP, ATF-4

Basic helix-loop-helix factors (bHLH)

Basic helix-span-helix factors (bHSH)



A specific alpha-helical dimerization domain provides the linkage between two DNA-contacting basic regions which adopt helical conformation →

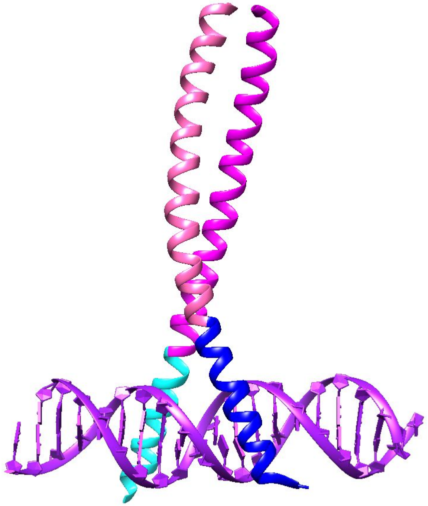


bZIP

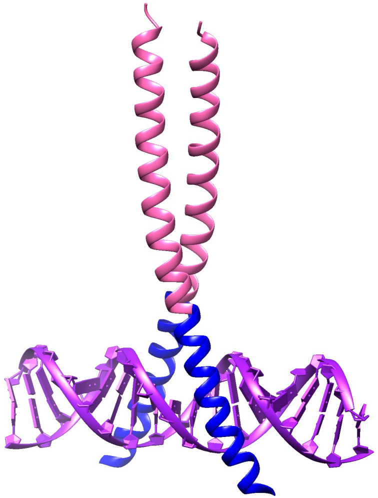
The bZIP TFs plays an essential role in several processes in eukaryotic cells, from early development to tumorigenesis.

40-80 amino acid long bZIP domain with two motifs:

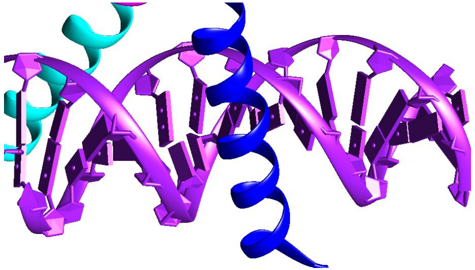
- **Basic region**: bind TF to its target DNA.
- **Leucine zipper**: needed for TF dimerization.



Heterodimer

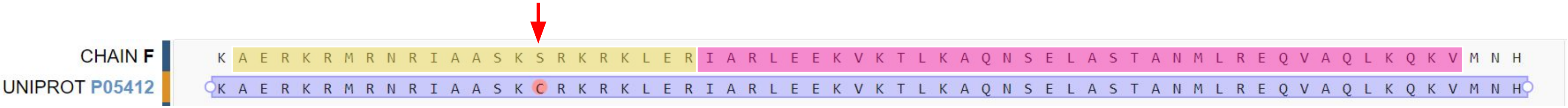


Homodimer



c-JUN

- Transcriptional activator.
- Homodimer or heterodimer (c-Fos).
- Regulation of cell growth and differentiation.
- Deregulated expression is implicated in the oncogenic transformation of cells.



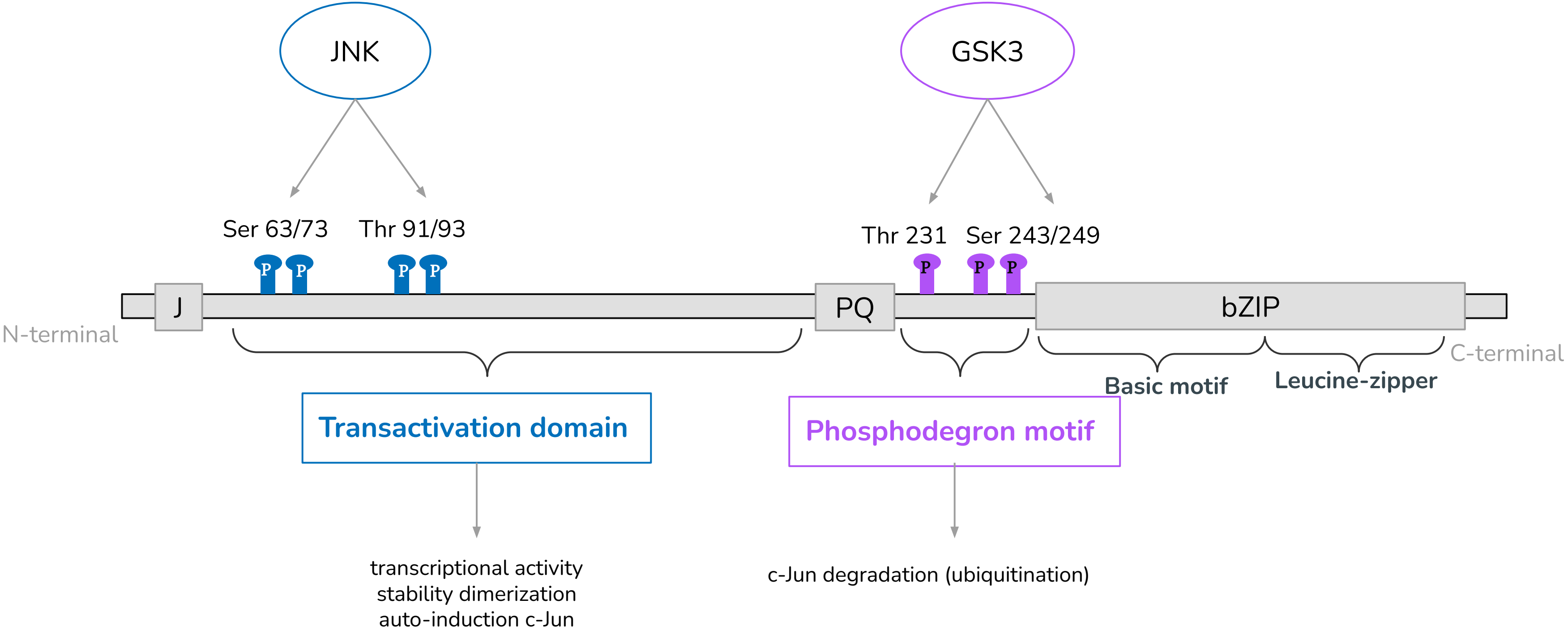
CHAIN F K A E R K R M R N R I A A S K S R K R K L E R I A R L E E K V K T L K A Q N S E L A S T A N M L R E Q V A Q L K Q K V M N H

UNIPROT P05412 K A E R K R M R N R I A A S K C R K R K L E R I A R L E E K V K T L K A Q N S E L A S T A N M L R E Q V A Q L K Q K V M N H

Basic region

Leucine zipper region

C-JUN



Transactivation domain

transcriptional activity
stability dimerization
auto-induction c-Jun

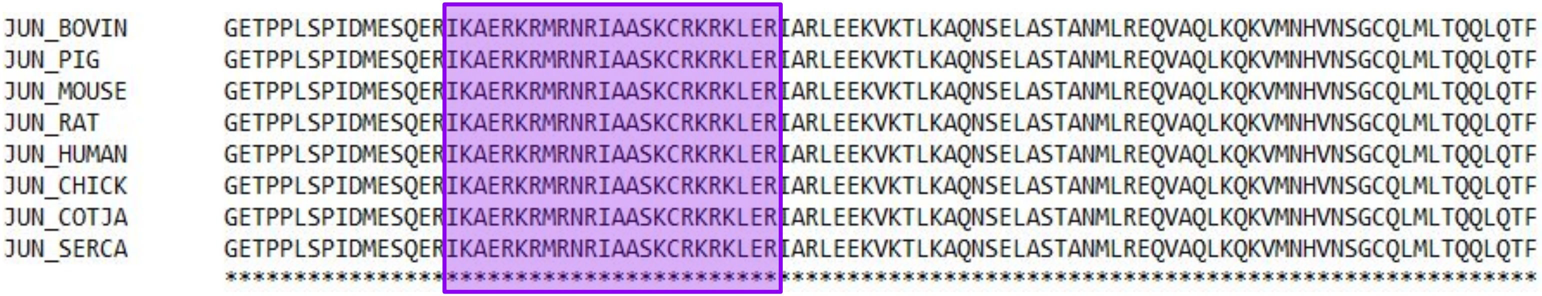
Phosphodegron motif

c-Jun degradation (ubiquitination)

SCOP CLASSIFICATION: c-Jun

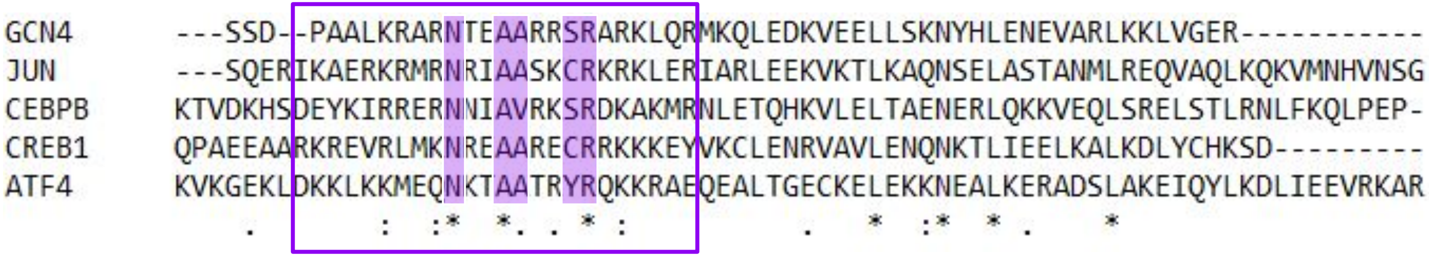
CLASS	All alpha proteins
FOLD	Left-handed parallel coiled-coil
SUPERFAMILY	Leucine zipper-like
FAMILY	Leucine zipper domain
PROTEIN	c-Jun
SPECIES	Human (Homo Sapiens)

Homology: Basic region



Basic region

```
JUN_BOVIN     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_PIG       GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_MOUSE     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_RAT       GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_HUMAN     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_CHICK     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_COTJA     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_SERCA     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
              ***********************************************************************************************
```

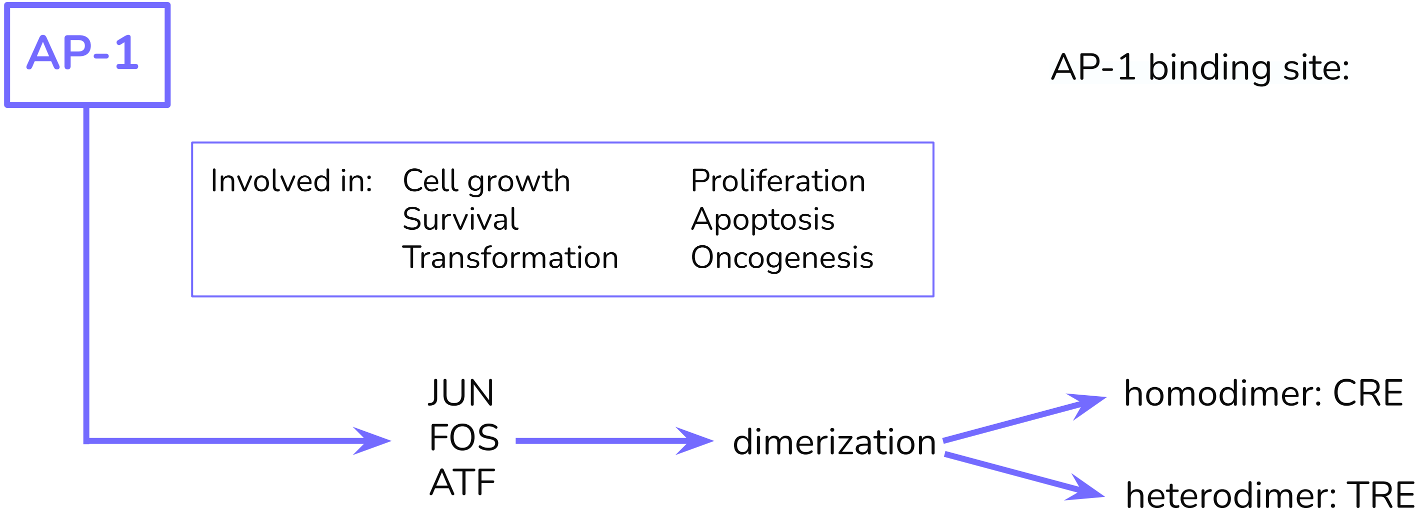


```
GCN4       ---SSD--PAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER-----------
JUN        ---SQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSG
CEBPB      KTVDKHSDEYKIRRERNNIAVRKSRDKAKMRNLETQHKVLELTAENERLQKKVEQLSRELSTLRNLFKQLPEP-
CREB1      QPAEEAARKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD---------
ATF4       KVKGEKLDKKLKKMEQNKTAATRYRQKKRAEQEALTGECKELEKKNEALKERADSLAKEIQYLKDLIEEVRKAR
               .       :  :*  *. . * :          .   *  :*  * .    *
```

Hydrogen bonds → Arg (R), Asn (N) and Lys (K) → hydrogen atom is covalently bonded to an electronegative atom

Van der Waals → Ala (A) and Ser (S) → temporary fluctuation in the distribution of electrons

Basic region



AP-1

Involved in:
- Cell growth
- Survival
- Transformation
- Proliferation
- Apoptosis
- Oncogenesis

JUN
FOS
ATF

dimerization

homodimer: CRE

heterodimer: TRE

AP-1 binding site:

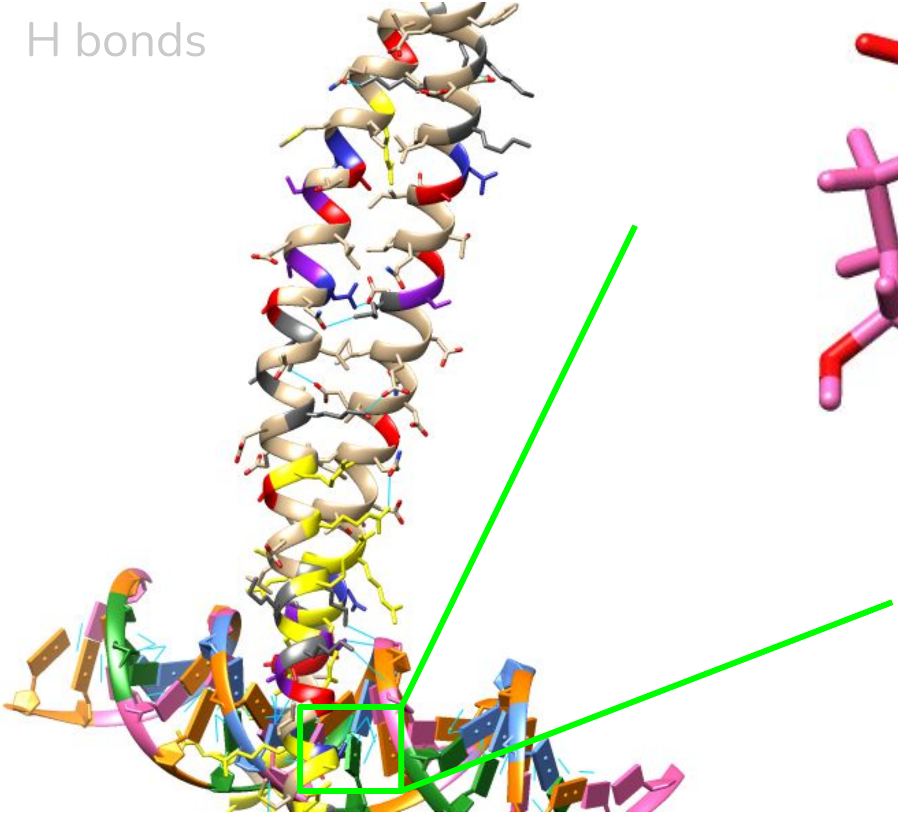
5'-TGACTCA-3'
3’-ACTGAGT-5’

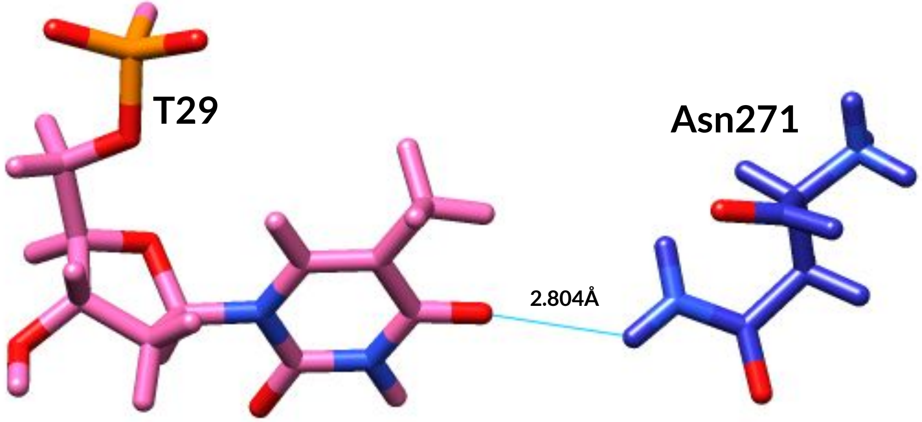
- Specific bases contacts between: N-terminal region c-Jun with DNA in the major groove, over a whole turn
- Only asymmetric major groove contact Fos/Jun + DNA → Arginine 279 of Jun

Interactions: Basic region

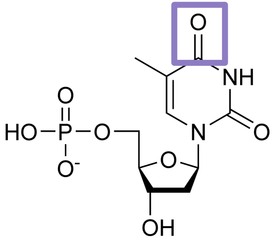
IKAERKRMR**N**RIAASKCRKRKLER

H bonds

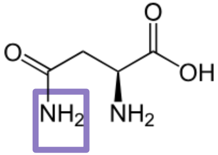




Thymidine



Asparagine

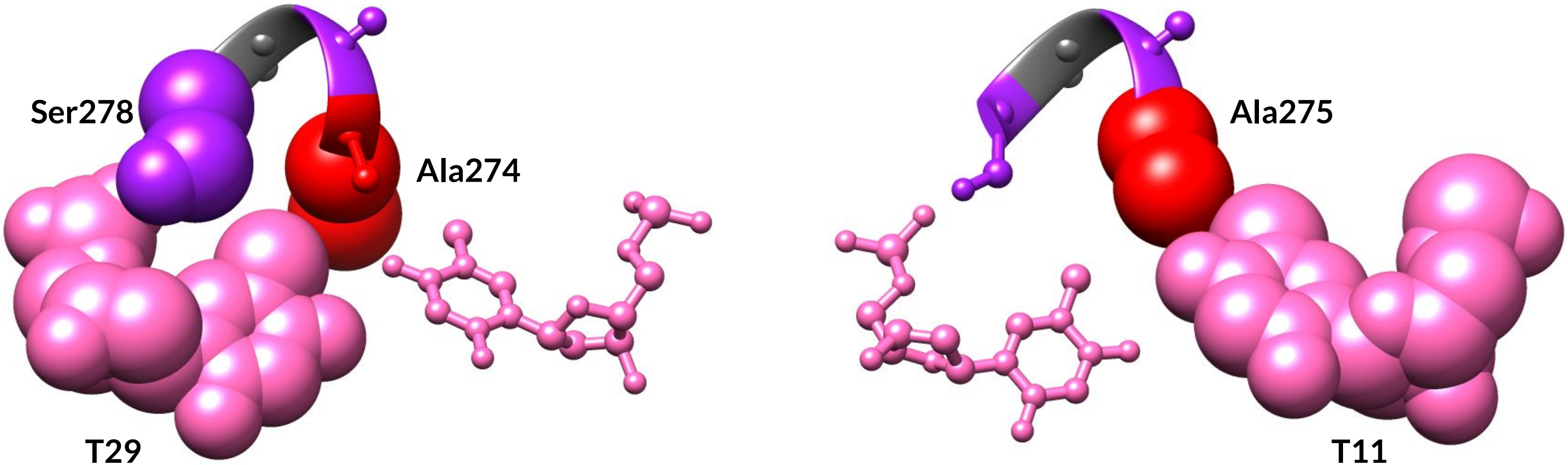


Interactions: Basic region

IKAERKRMRNRIAASKCRKRKLER

S

Van der Waals → hydrophobic interaction



cohesion and some properties

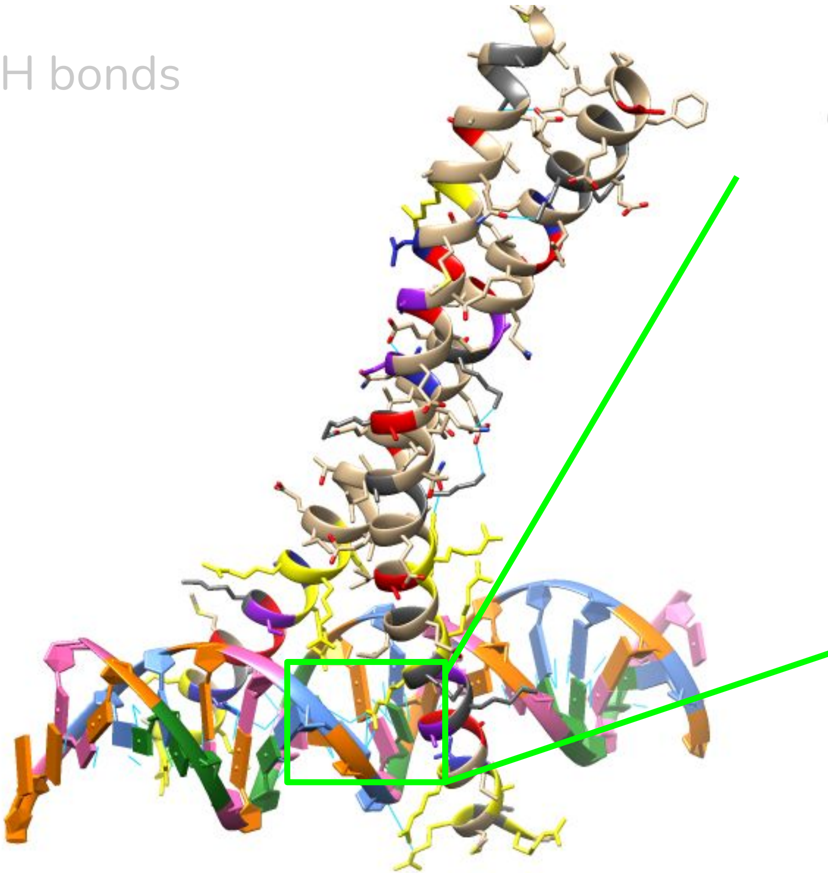
boiling

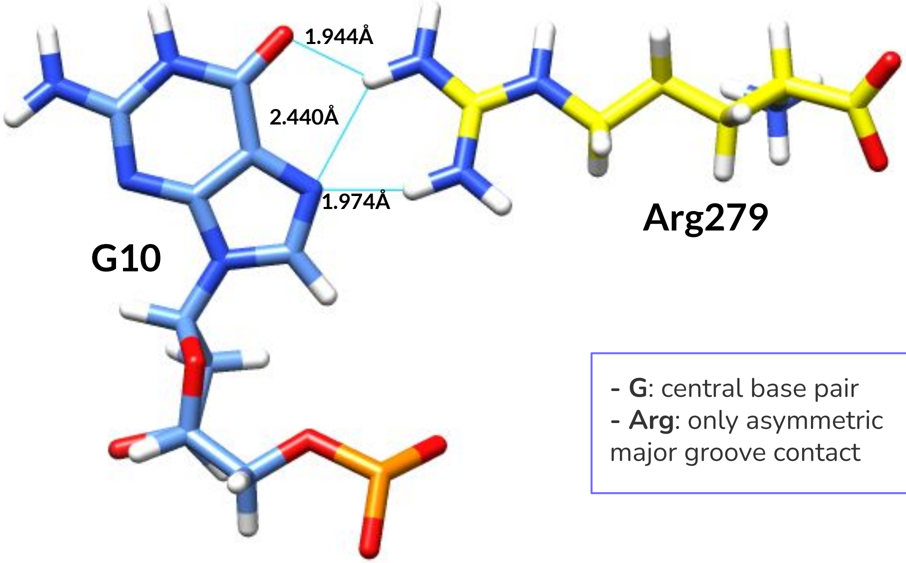
solubility

Interactions: Basic region

IKAERKRMRNRIAASKCRKRKLER

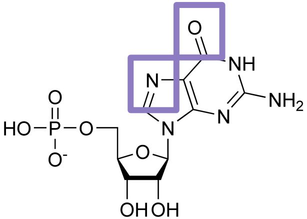
H bonds



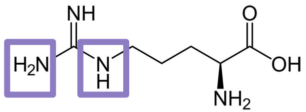


- **G**: central base pair
- **Arg**: only asymmetric major groove contact

Guanosine



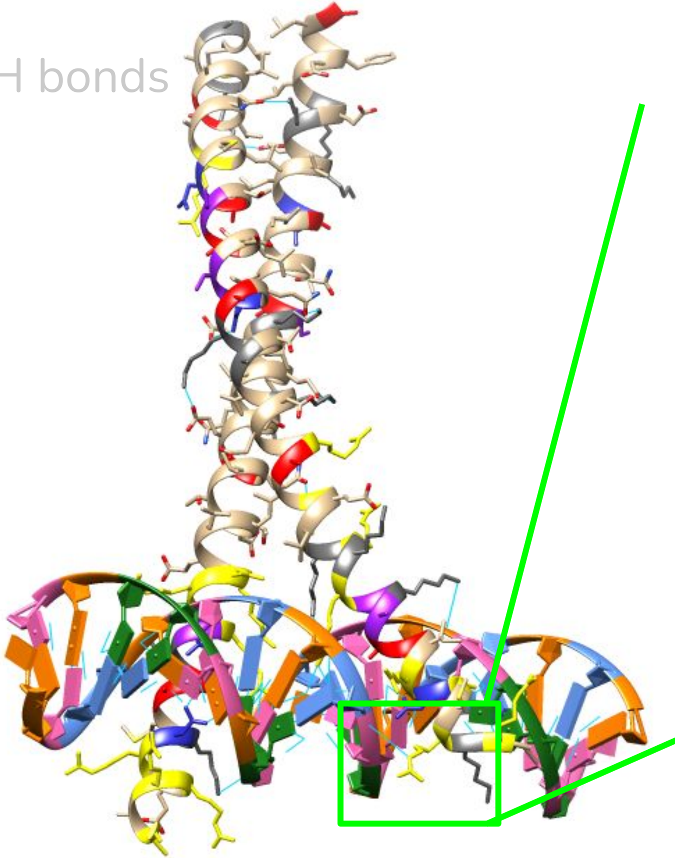
Arginine

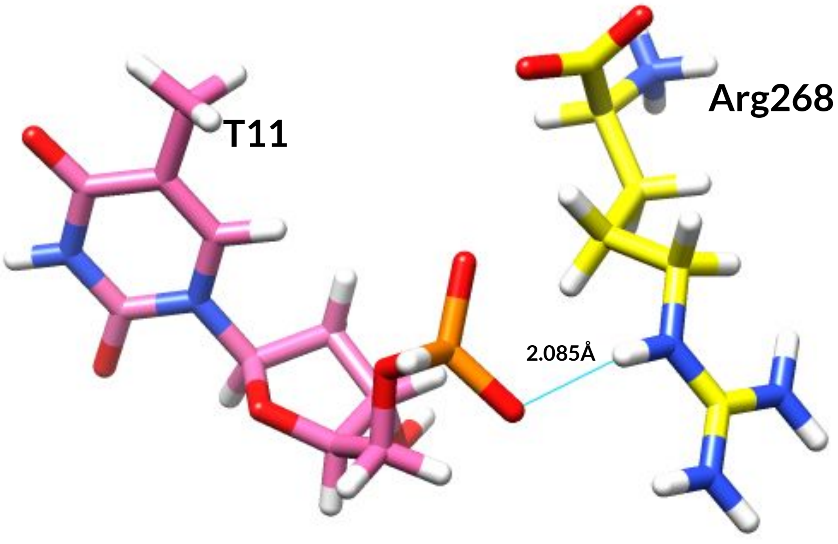


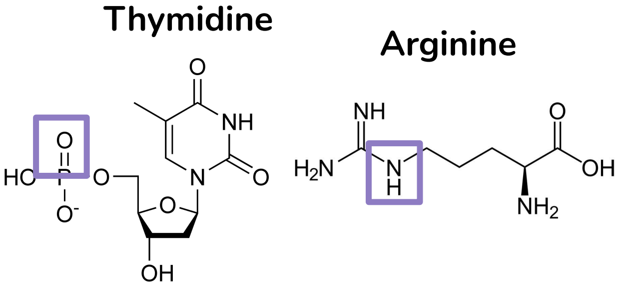
Interactions: Basic region

IKAERKRMRNRIAASKCRKRKLER

H bonds



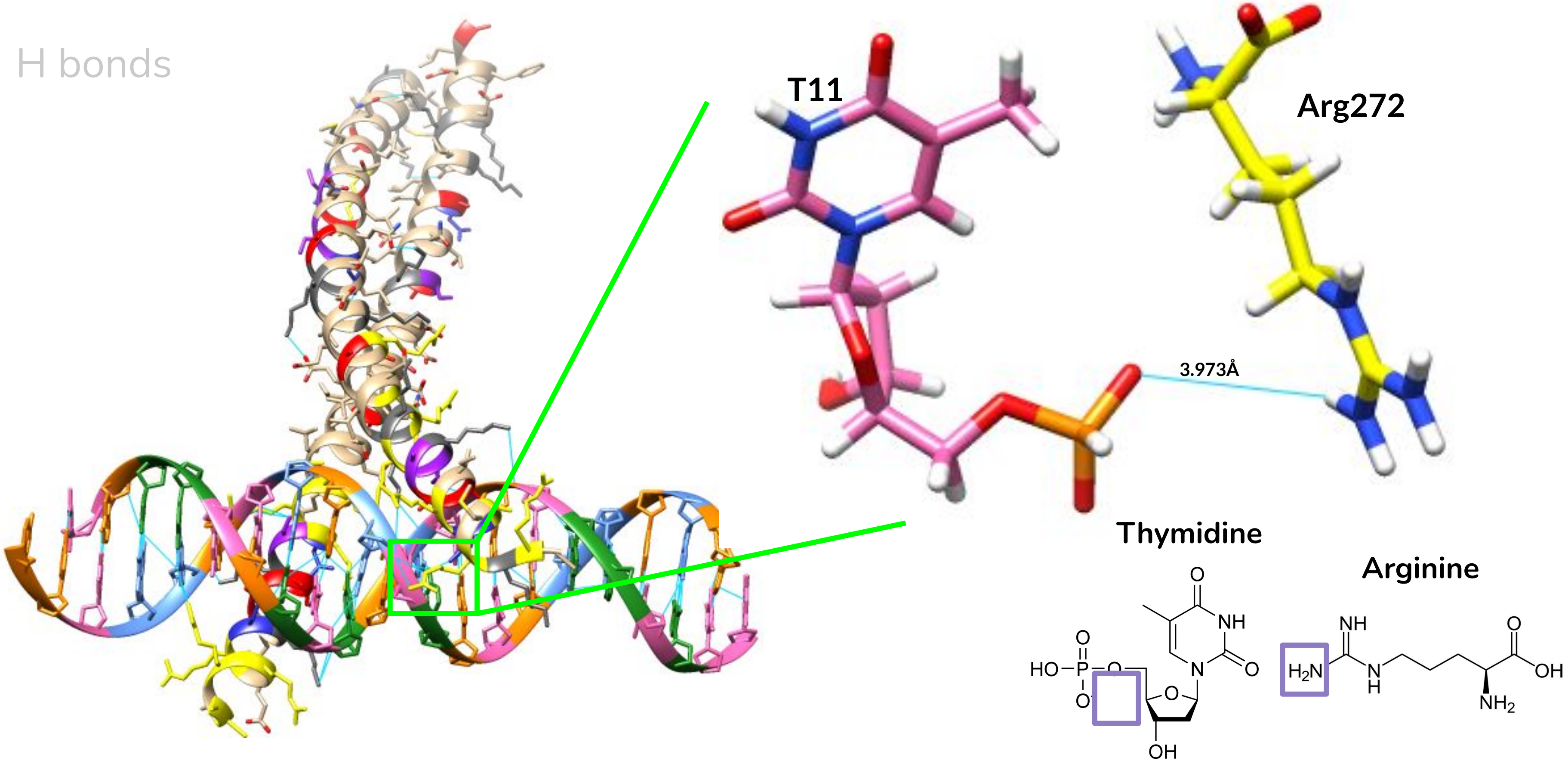




Interactions: Basic region

IKAERKRMRN**R**IAASKCRKRKLER

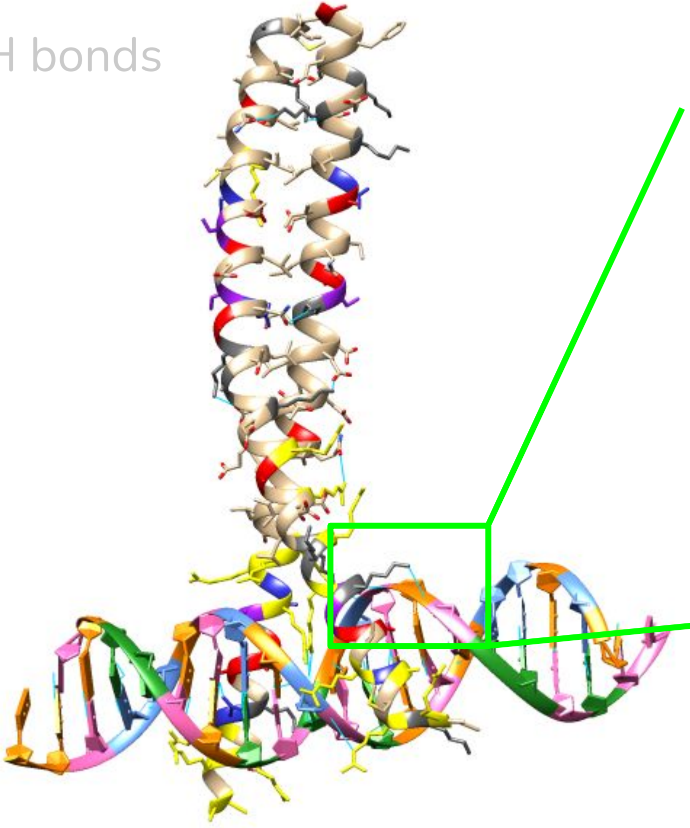
H bonds

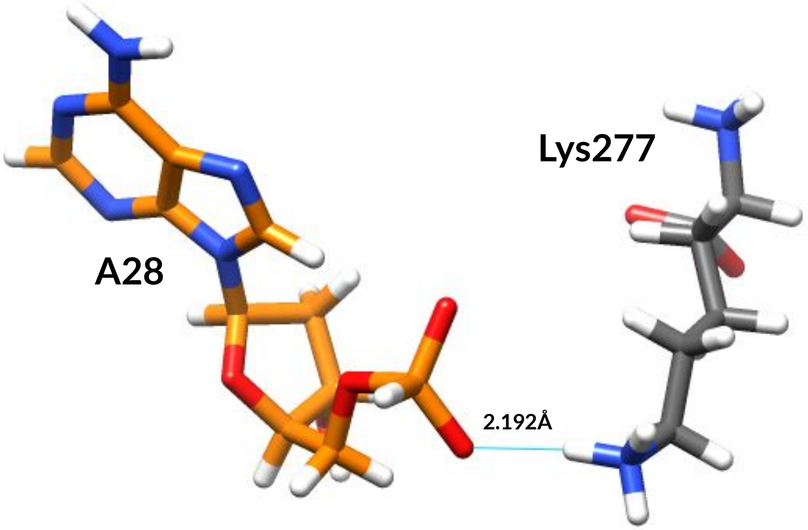


Interactions: Basic region

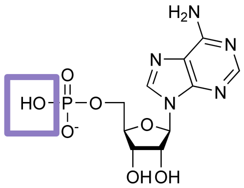
IKAERKRMRNRIAASKCRKRKLER

H bonds

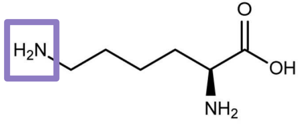




Adenosine



Lysine



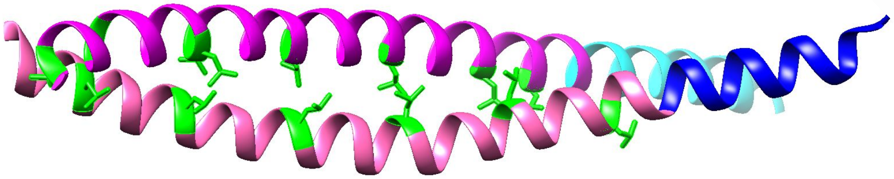
Leucine zipper

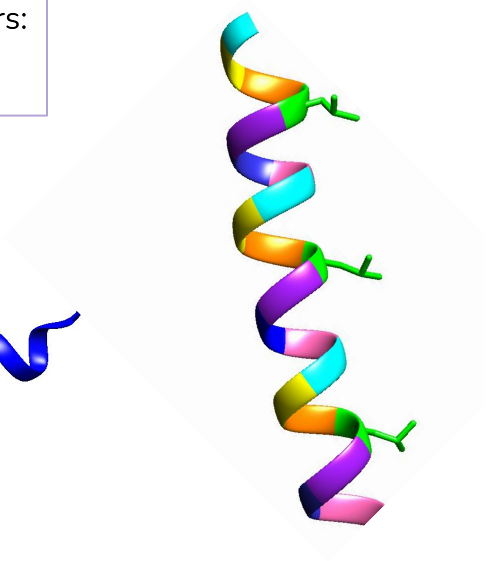
- Dictates dimerization specificity: prerequisite for DNA-binding
- Heptad coil: 7 residue modules
- Leucine at position d: repetition every 7 residues
- Can form either homodimers or heterodimers: Knobs-into-holes model

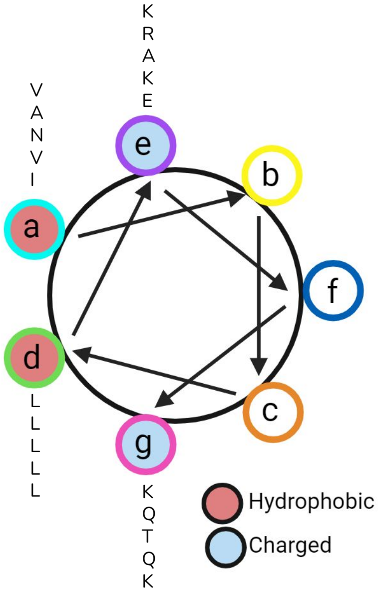
Heptad coil

a b c d e f g

IARLEEK VKTLKAQ NSELAST ANMLREQ VAQLKQK

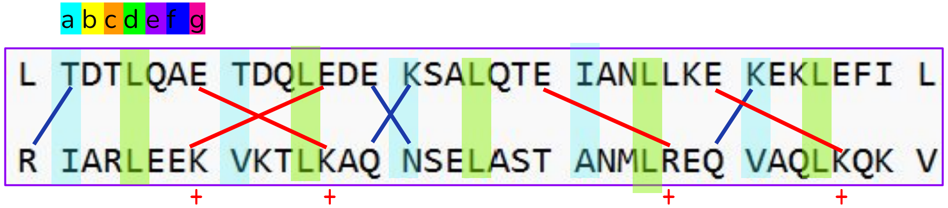






Knobs-into-holes model

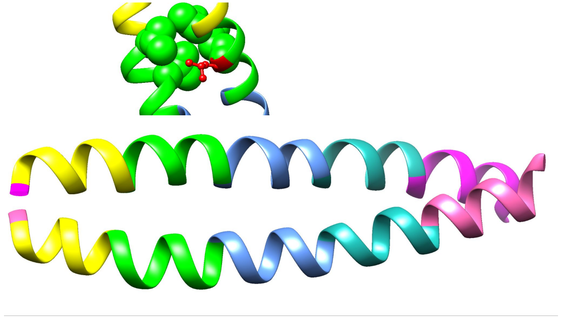
Interaction c-Jun / c-Fos:


a b c d e f g

c-Fos: L TDTLQAE TDQLEDE KSALQTE IANLLKE KEKLEFI L

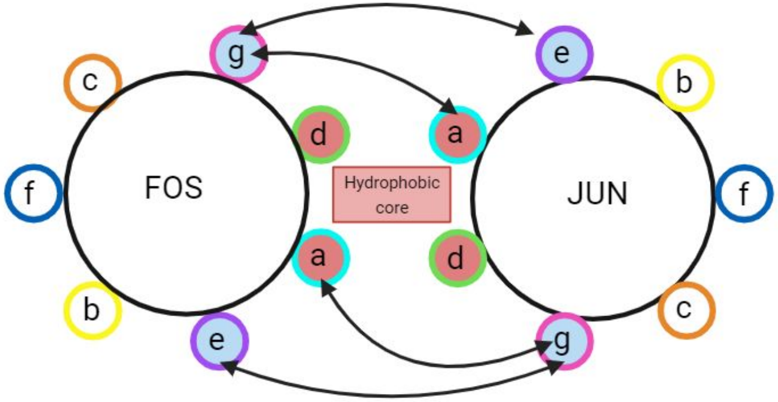
c-Jun: R IARLEEK VKTLKAQ NSELAST ANMLREQ VAQLKQK V

+ + + +



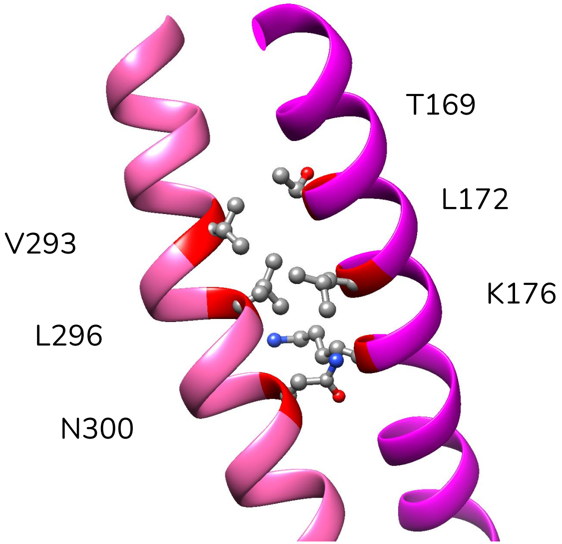
- Hydrophobic a-a and d-d interactions
- Electrostatic e-g and a-g interactions

— Hydrogen Bonds
— Salt bridges

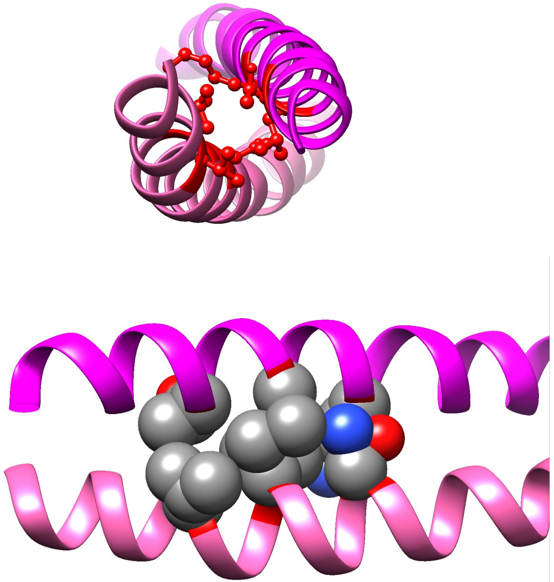


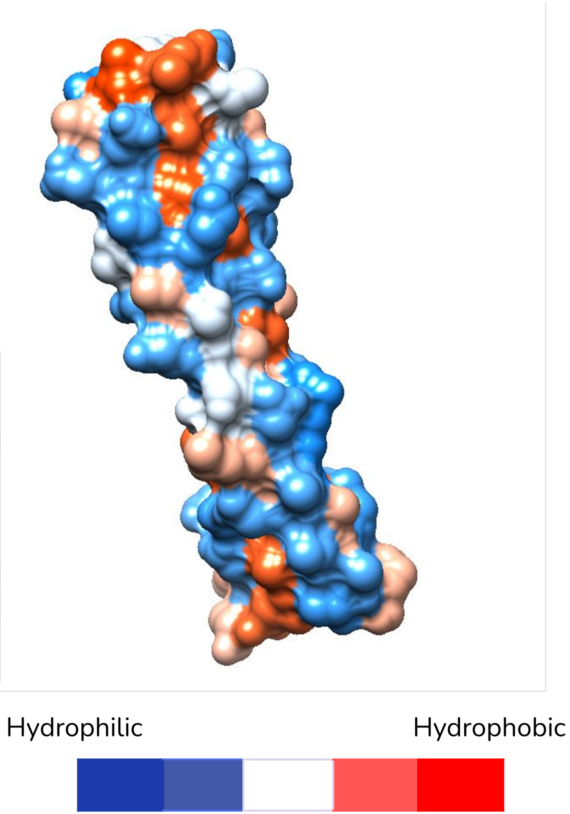
Knobs-into-holes model

Hydrophobic a-a and d-d interactions:



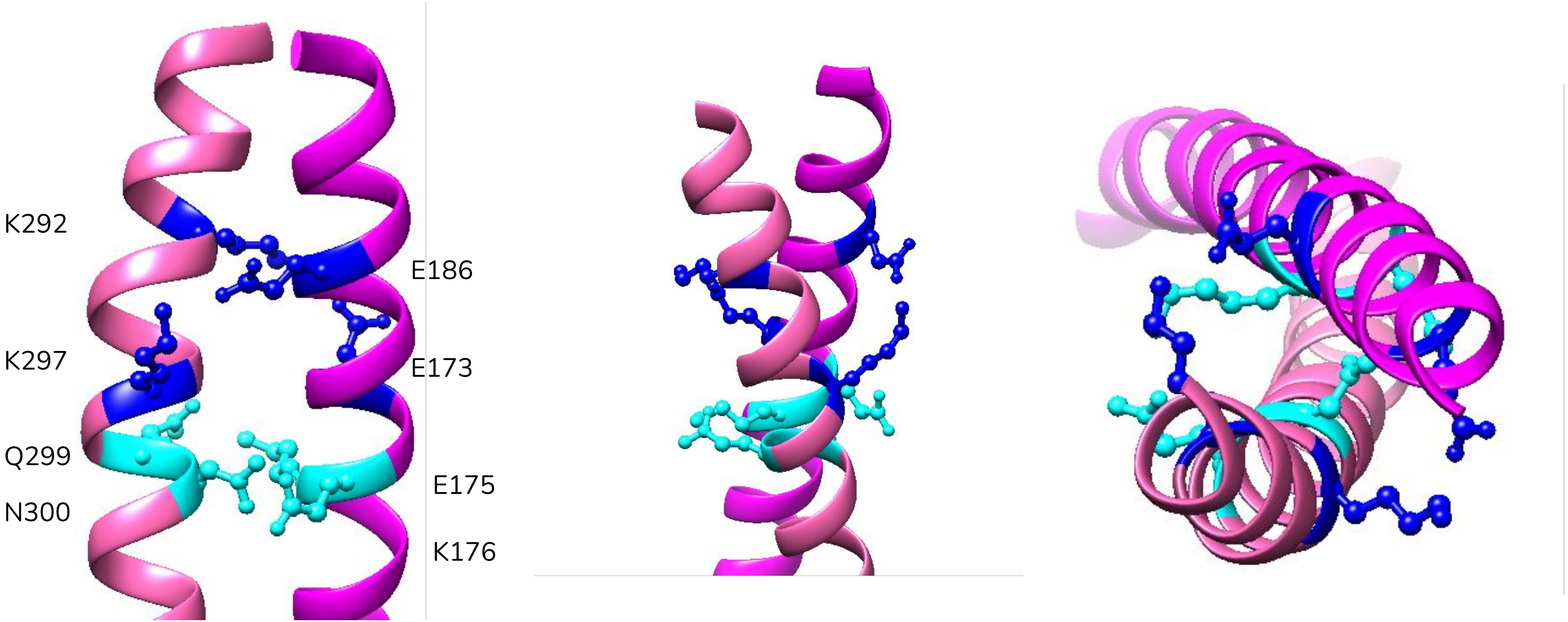
a-a: V293 + T169
d-d: L296 + L172
a-a: N300 + K176





Knobs-into-holes model

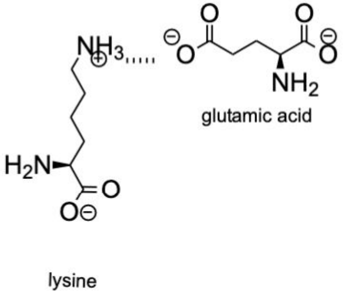
Electrostatic e-g and a-g interactions:



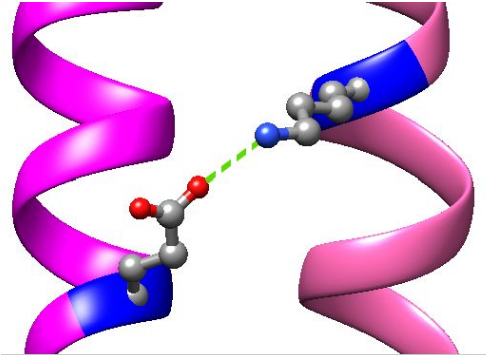
Knobs-into-holes model

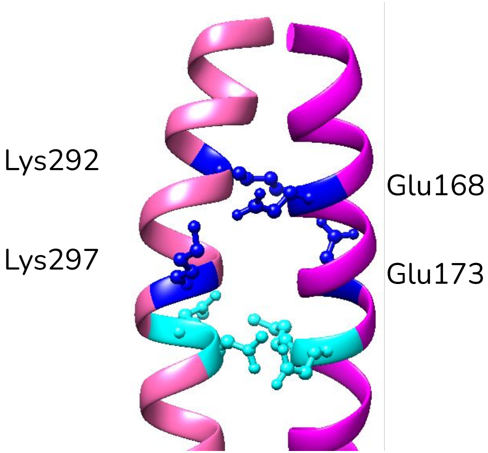
Salt bridges

g - e	Lys292 - Glu173	2.938 Å
e - g	Lys297 - Glu168	2.801 Å

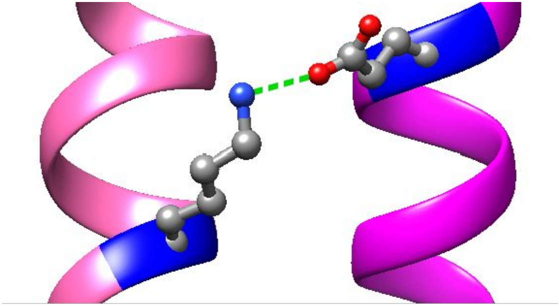


Glu173 - Lys292





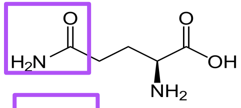
Lys297 - Glu168



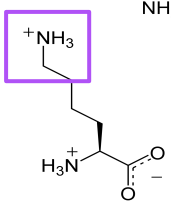
Knobs-into-holes model

Hydrogen Bond

Glutamine

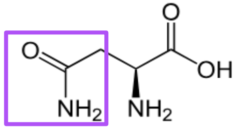


Lysine

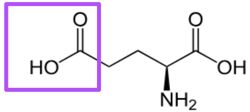


g -a	Gln299- Lys176	2.472 Å
a - g	Asn300 - Glu175	2.227Å

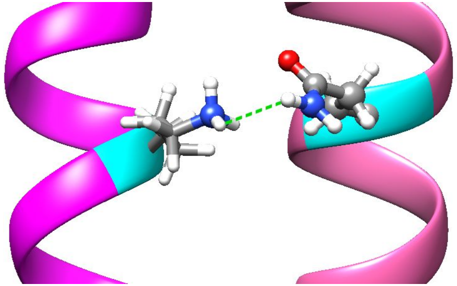
Asparagine

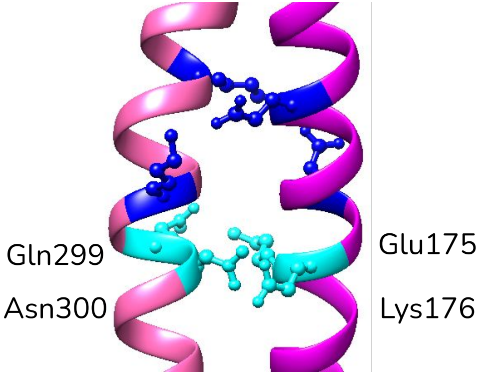


Glutamic acid

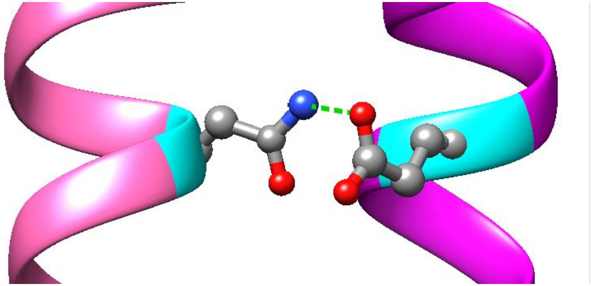


Lys176- Gln299

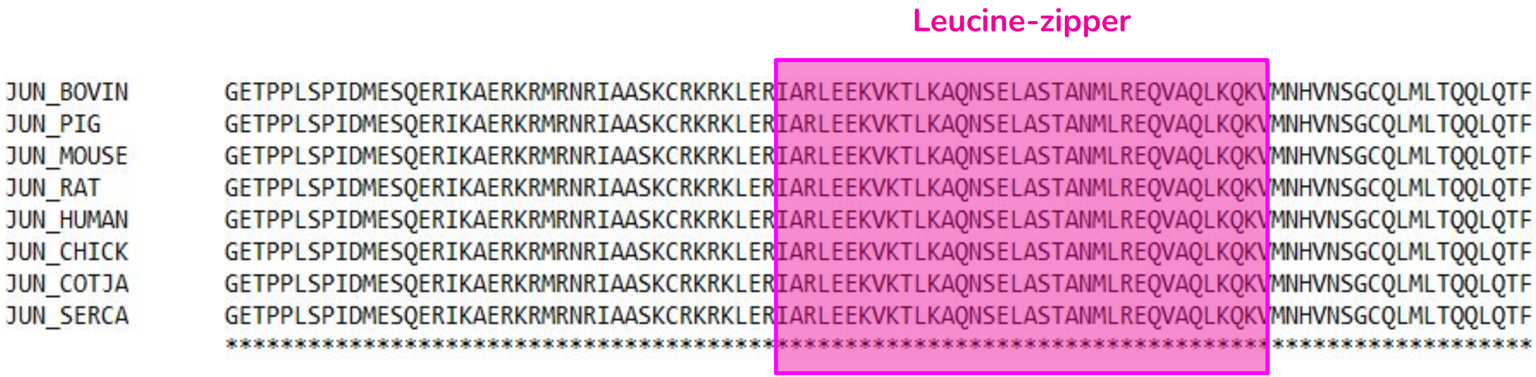




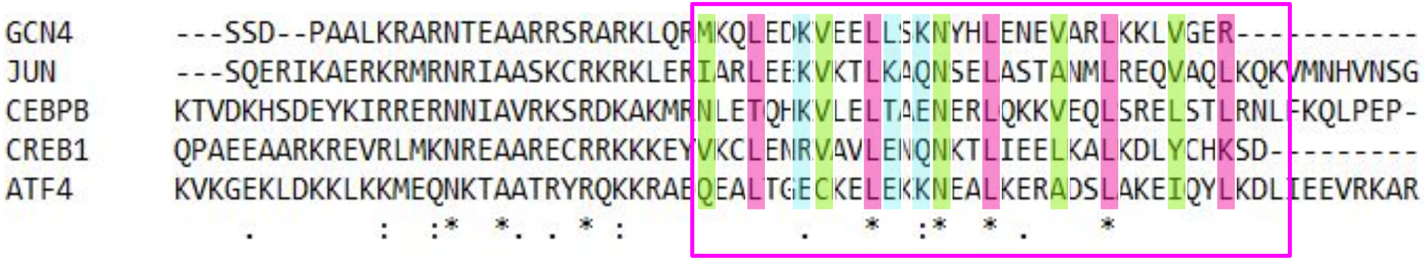
Asn300 - Glu175



Homology: Leucine zipper

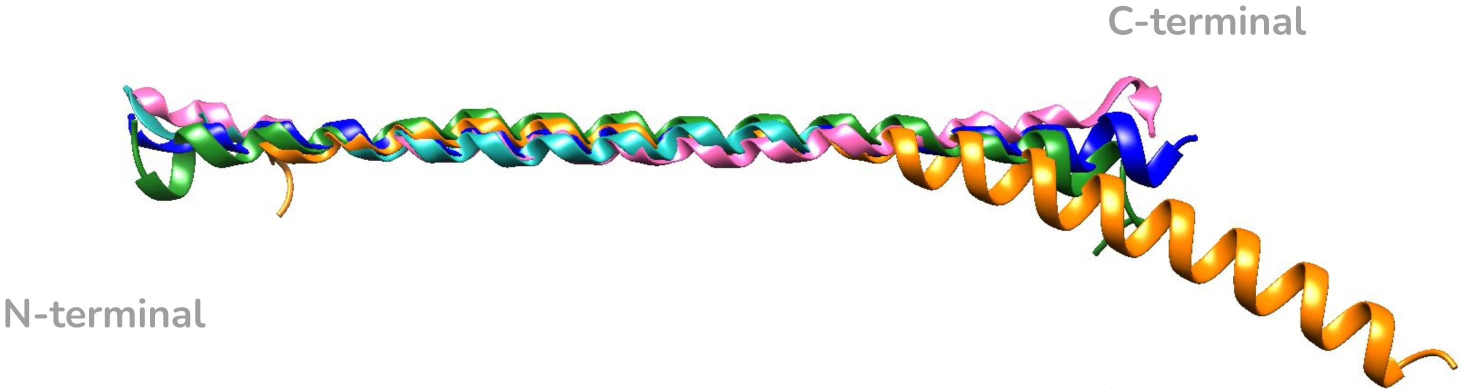


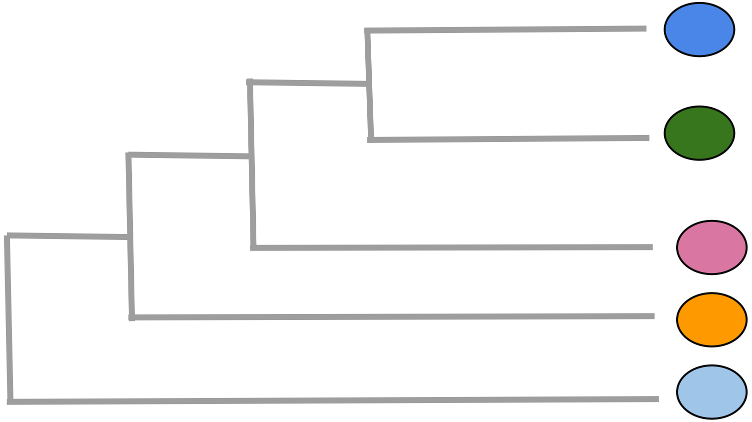
```
JUN_BOVIN   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_PIG     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_MOUSE   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_RAT     GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_HUMAN   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_CHICK   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_COTJA   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
JUN_SERCA   GETPPLSPIDMESQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSGCQLMLTQQLQTF
            ***********************************************************************************************
```



```
GCN4     ---SSD--PAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVGER-----------
JUN      ---SQERIKAERKRMRNRIAASKCRKRKLERIARLEEKVKTLKAQNSELASTANMLREQVAQLKQKVMNHVNSG
CEBPB    KTVDKHSDEYKIRRERNNIAVRKSRDKAKMRNLETQHKVLELTAENERLQKKVEQLSRELSTLRNLFKQLPEP-
CREB1    QPAEEAARKREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYCHKSD---------
ATF4     KVKGEKLDKKLKKMEQNKTAATRYRQKKRAEQEALTGECKELEKKNEALKERADSLAKEIQYLKDLIEEVRKAR
             .       :  :*  *. . * :          .   *  :*  *.    *
```

Superimposition





Score	5.56
RMSD	1.06

CREB1
JUN
CEBPB
ATF4
GCN4

KLF4

Zinc Fingers

Small, functional, independently folded domain that requires **coordination of one or more zinc ions** to stabilize its structure by a different combination of cysteines and histidines.

Nuclear receptors with C4 zinc fingers

Other C4 zinc finger-type factors

C2H2 zinc finger factors

DM-type intertwined zinc finger factors

CXXC zinc finger factors

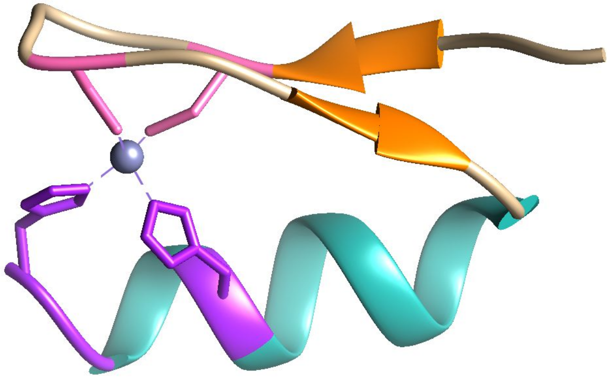
C2HC

C3H zinc finger factors

C2CH THAP-type zinc finger factors

Zinc Fingers - C2H2 family

- Two Cys and two His residues bonded tetrahedrally to a Zn ion.
- Consists of a short antiparallel β-sheet formed by two strands and hairpin turn, followed by an α-helix.
- They are transcription factors, meaning that they function by identifying particular DNA sequences.



KLF4: Biological function

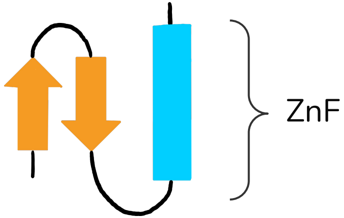
- Transcription factor
- SP/KLF superfamily: three zinc-finger motifs
- Binds to the DNA sequence: **5'-GAGGCGTGGC-3’**
- Indispensable for terminal maturation of epithelial tissues
- They are also potentials targets for cancer therapy

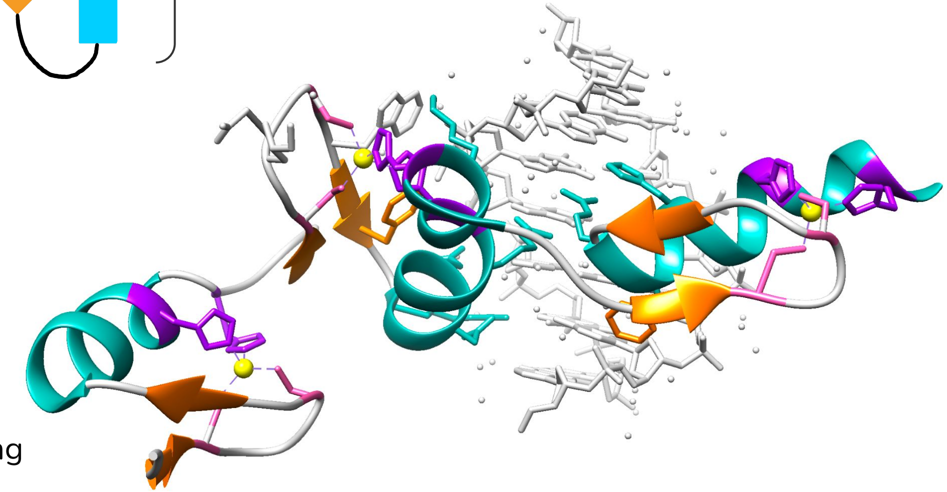
SCOP Classification

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	Krüppel-like factor 4, GKLF
SPECIES	Mouse (*Mus musculus*)

General structure

- 3 ZnF
- **ββα fold:** 2 antiparallel β-sheets + α-helix
- Hydrophobic core + Zn : stable structure





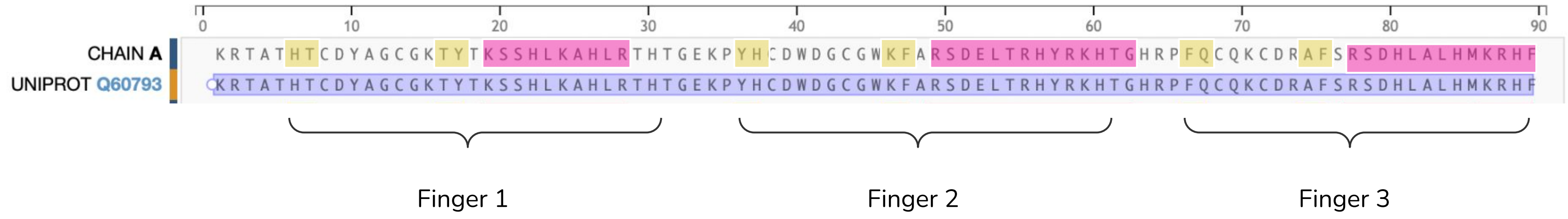
α-helixes

Specific DNA binding

β-sheets

Non-specific DNA-phosphate binding

2WBS - PDB information



CHAIN **A** KRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQKCDRAFSRSDHLALHMKRHF

UNIPROT Q60793 KRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHCDWDGCGWKFARSDELTRHYRKHTGHRPFQCQKCDRAFSRSDHLALHMKRHF

Antiparallel β-sheet

α-helix

Consensus sequence of C2H2 family

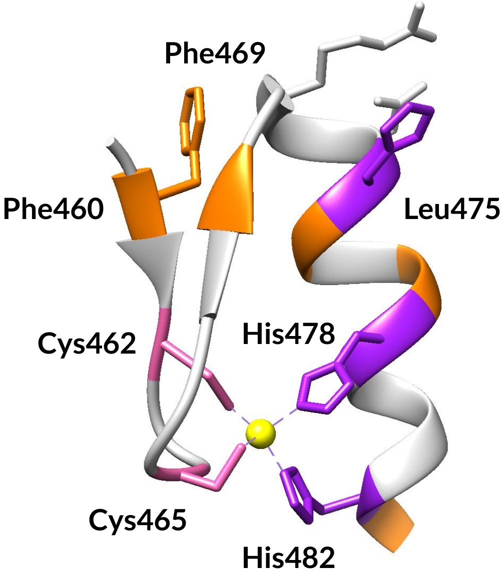
ψ-X-C-X2-5-C-X3-ψ-X5-ψ-X2-H-X3-5-H

Hydrophobic residue

Cysteine residue

Histidine residue

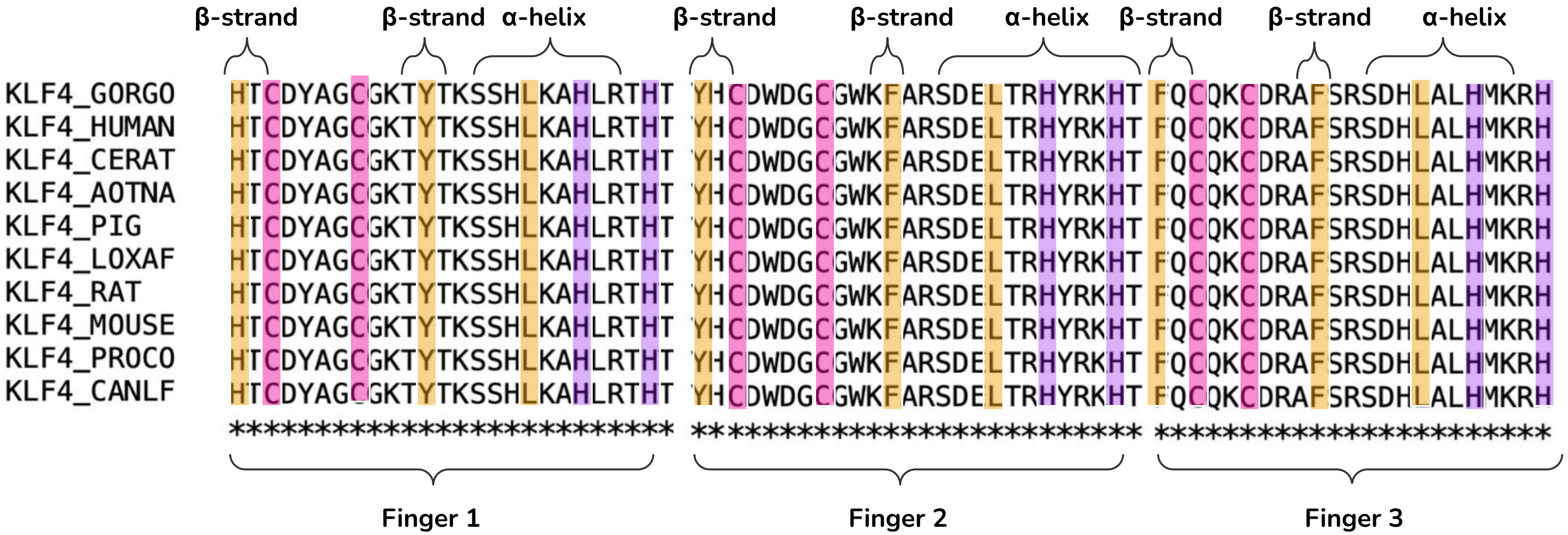
X = any residue



Finger 3

Consensus sequence - different species

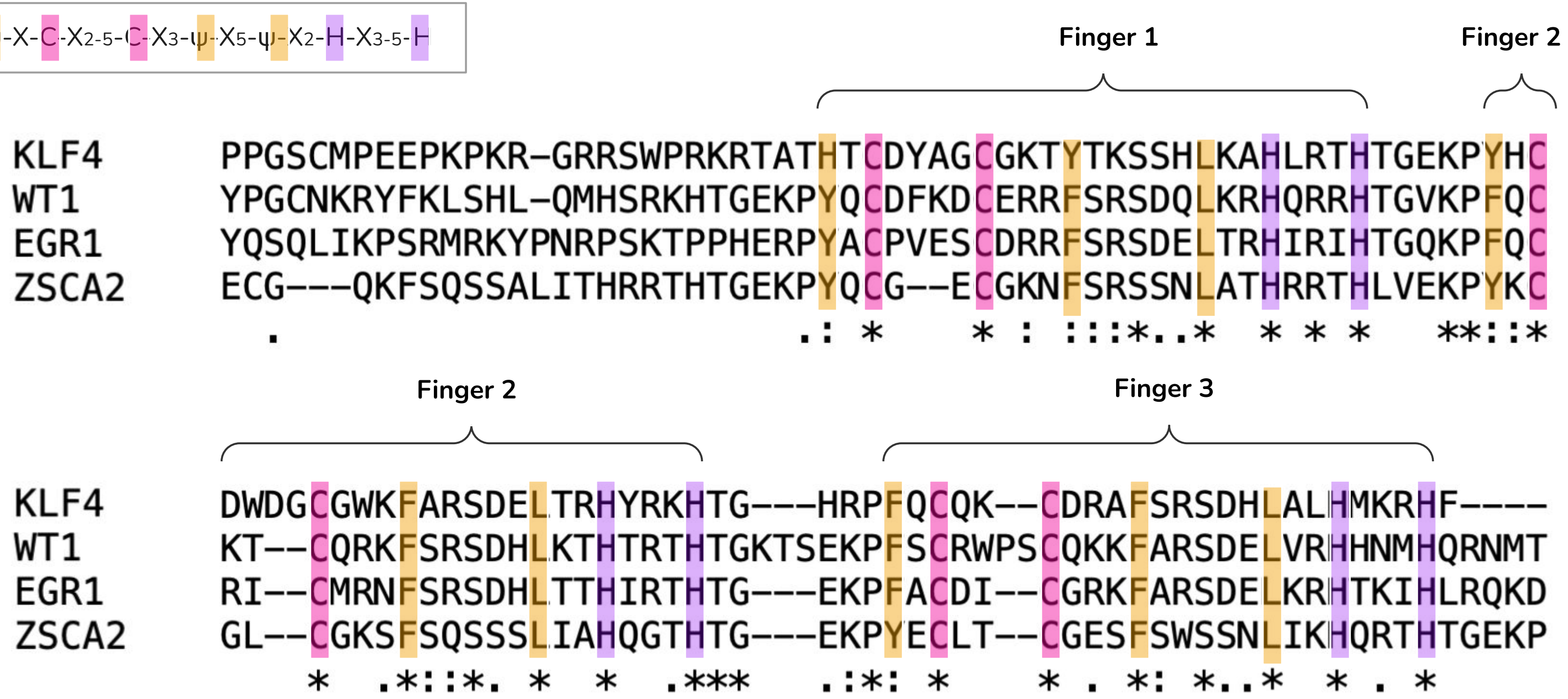
ψ-X-C-X_{2-5}-C-X_3-ψ-X_5-ψ-X_2-H-X_{3-5}-H



```
KLF4_GORGO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
             ************************** ************************** ***********************
```

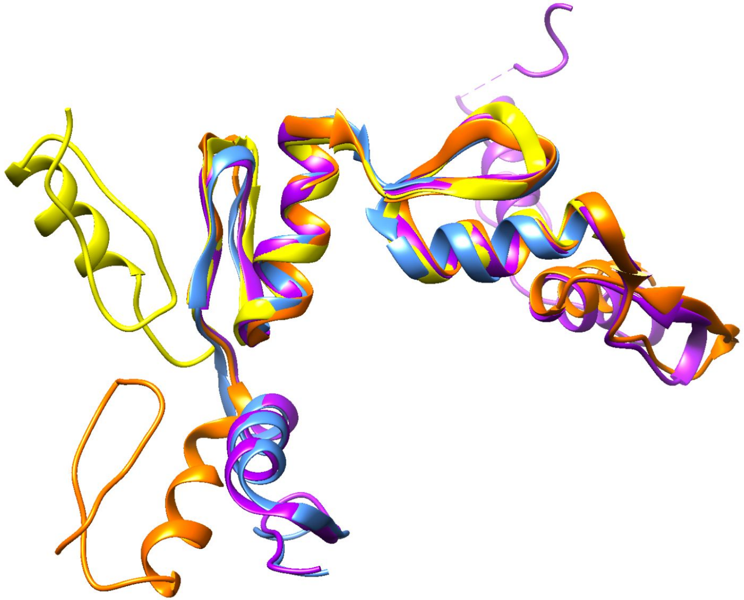
Consensus sequence - same family structures

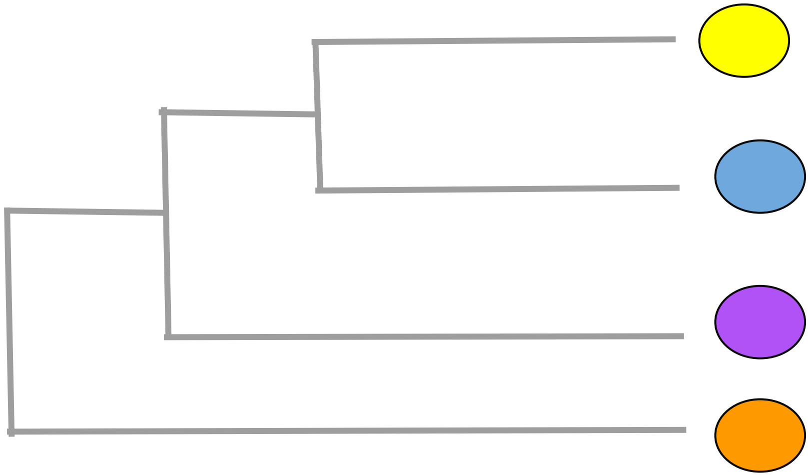
ψ-X-C-X$_{2\text{-}5}$-C-X$_3$-ψ-X$_5$-ψ-X$_2$-H-X$_{3\text{-}5}$-H



```
KLF4     PPGSCMPEEPKPKR-GRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHC
WT1      YPGCNKRYFKLSHL-QMHSRKHTGEKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC
EGR1     YQSQLIKPSRMRKYPNRPSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC
ZSCA2    ECG---QKFSQSSALITHRRTHTGEKPYQCG--ECGKNFSRSSNLATHRRTHLVEKPYKC
           .                       .: *    * : :::*..*  * * *   **::*

KLF4     DWDGCGWKFARSDELTRHYRKHTG---HRPFQCQK--CDRAFSRSDHLALHMKRHF----
WT1      KT--CQRKFSRSDHLKTHTRTHTGKTSEKPFSCRWPSCQKKFARSDELVRHHNMHQRNMT
EGR1     RI--CMRNFSRSDHLTTHIRTHTG---EKPFACDI--CGRKFARSDELKRHTKIHLRQKD
ZSCA2    GL--CGKSFSQSSSLIAHQGTHTG---EKPYECLT--CGESFSWSSNLIKHQRTHTGEKP
             *  .*::*. *  *  .***   .:*: *    * . *: *..*  * . *
```

Superimposition

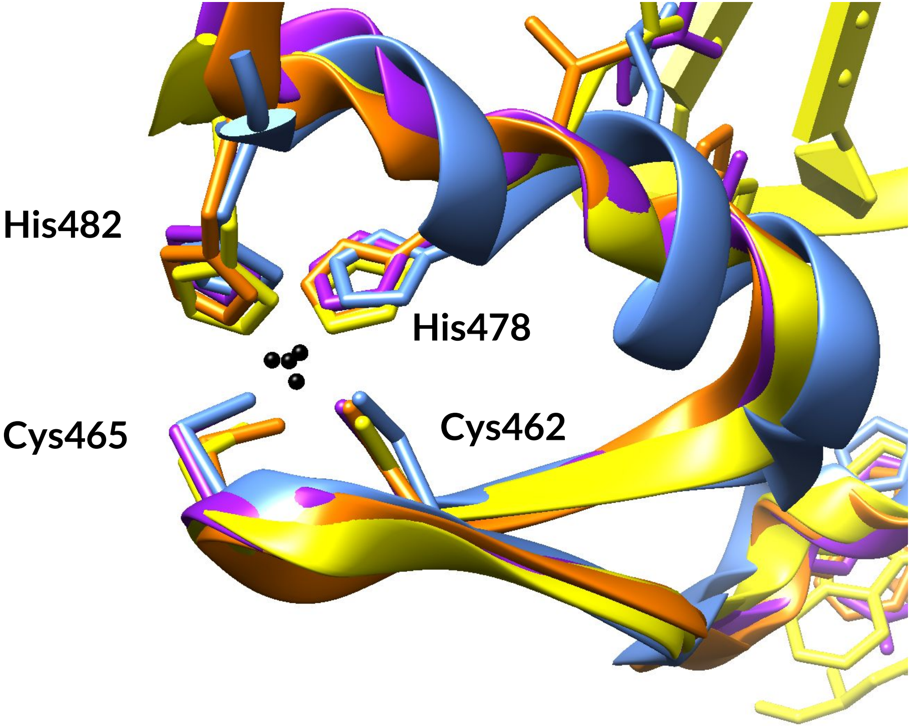




Score	5.58
RMSD	1.18

KLF4 WT1 ZSCA2 EGR1

Superimposition with Zn

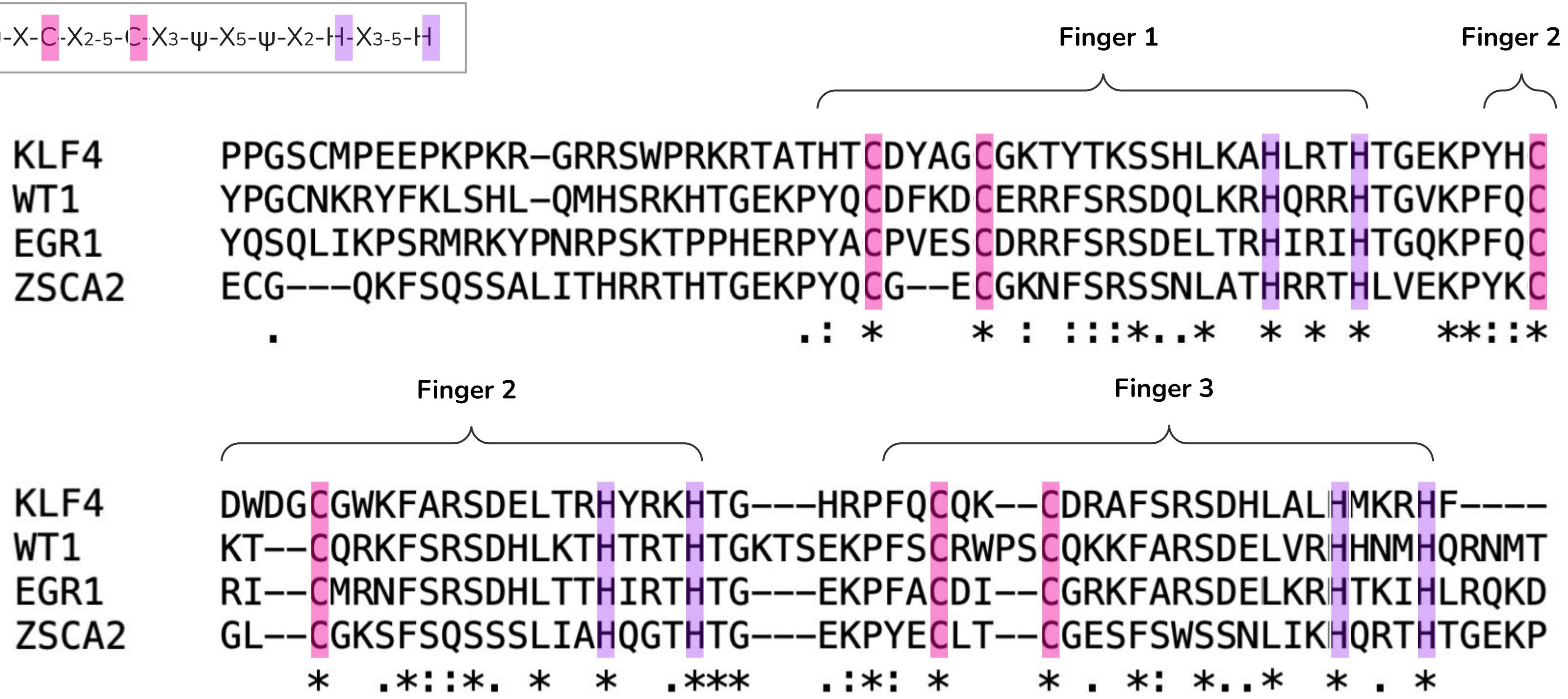


KLF4
WT1
ZSCA2
EGR1

Score	5.58
RMSD	1.18

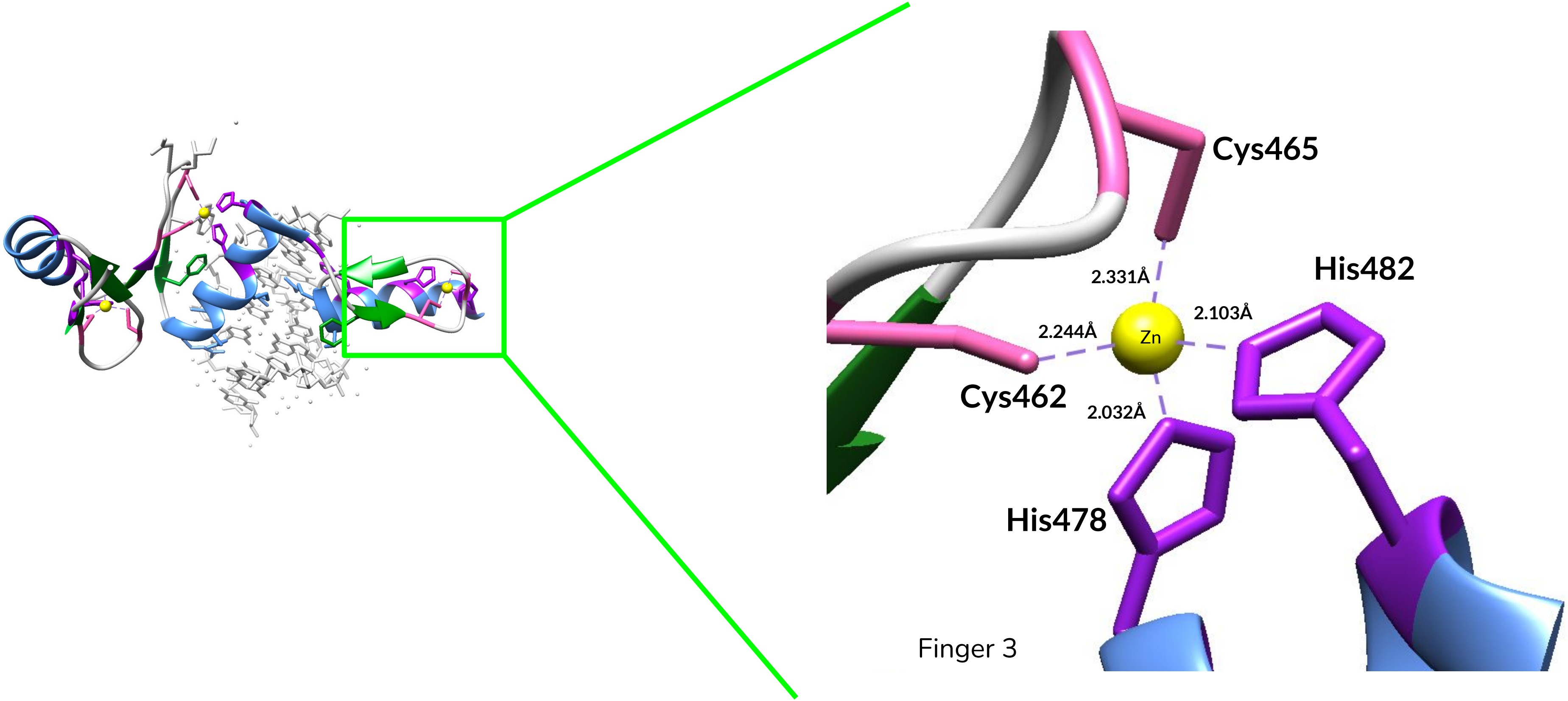
Conservation of residues

$\psi\text{-X-C-}X_{2\text{-}5}\text{-C-}X_{3}\text{-}\psi\text{-}X_{5}\text{-}\psi\text{-}X_{2}\text{-H-}X_{3\text{-}5}\text{-H}$



```
KLF4      PPGSCMPEEPKPKR-GRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHC
WT1       YPGCNKRYFKLSHL-QMHSRKHTGEKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC
EGR1      YQSQLIKPSRMRKYPNRPSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC
ZSCA2     ECG---QKFSQSSALITHRRTHTGEKPYQCG--ECGKNFSRSSNLATHRRTHLVEKPYKC
            .                       .: *    * : :::*..*  * * *   **::*

KLF4      DWDGCGWKFARSDELTRHYRKHTG---HRPFQCQK--CDRAFSRSDHLALHMKRHF----
WT1       KT--CQRKFSRSDHLKTHTRTHTGKTSEKPFSCRWPSCQKKFARSDELVRHHNMHQRNMT
EGR1      RI--CMRNFSRSDHLTTHIRTHTG---EKPFACDI--CGRKFARSDELKRHTKIHLRQKD
ZSCA2     GL--CGKSFSQSSSLIAHQGTHTG---EKPYECLT--CGESFSWSSNLIKHQRTHTGEKP
              *  .*::*. *  *  .***   .:*: *    * . *: *..*  * . *
```

Residues involved in Zn binding

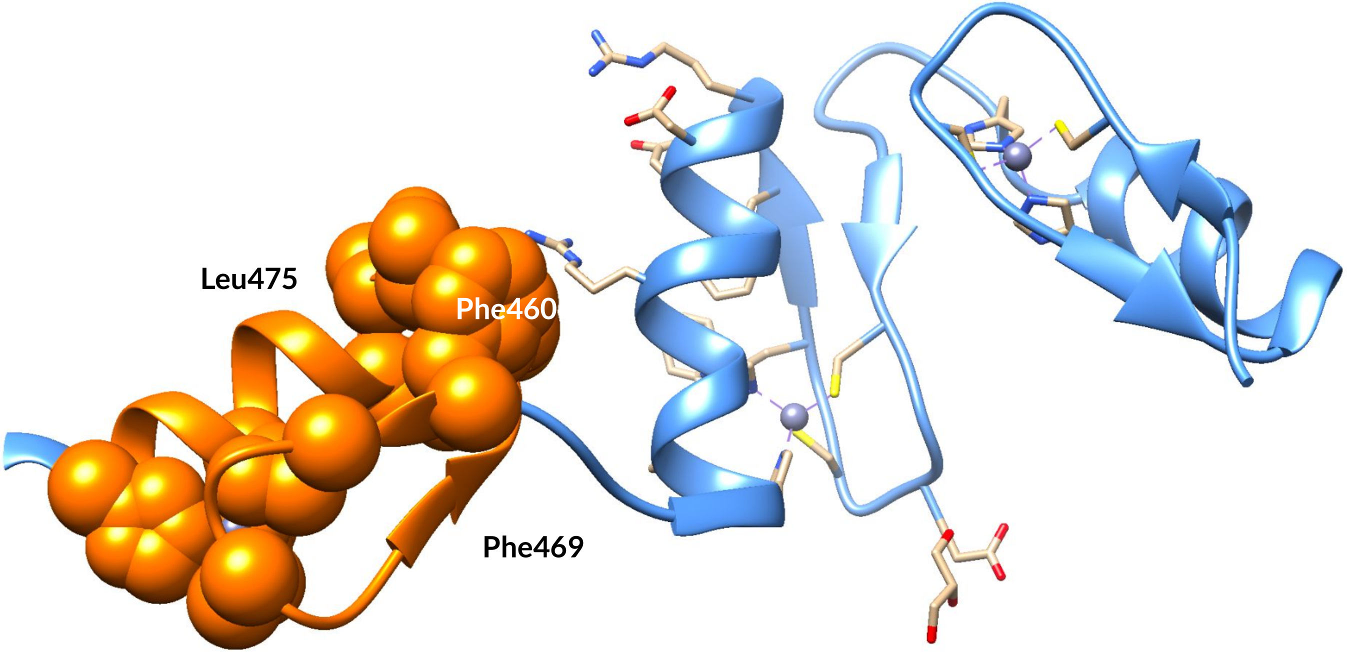


Conservation of residues

$\psi\text{-X-C-X}_{2\text{-}5}\text{-C-X}_{3}\text{-}\psi\text{-X}_{5}\text{-}\psi\text{-X}_{2}\text{-H-X}_{3\text{-}5}\text{-H}$

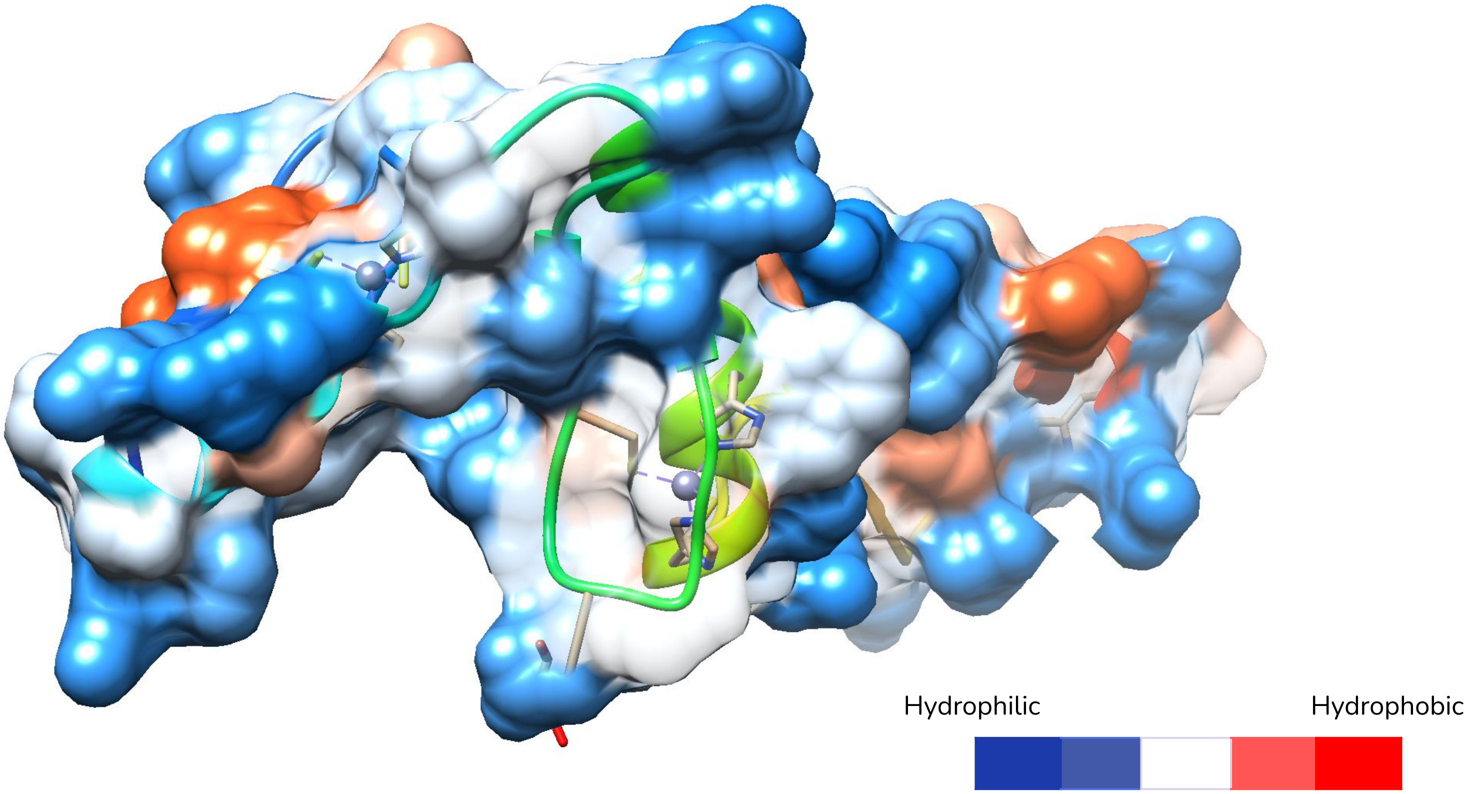
```
KLF4      PPGSCMPEEPKPKR-GRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHC
WT1       YPGCNKRYFKLSHL-QMHSRKHTGEKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC
EGR1      YQSQLIKPSRMRKYPNRPSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC
ZSCA2     ECG---QKFSQSSALITHRRTHTGEKPYQCG--ECGKNFSRSSNLATHRRTHLVEKPYKC
            .                       .: *    * : :::*..*  * * *   **::*

KLF4      DWDGCGWKFARSDELTRHYRKHTG---HRPFQCQK--CDRAFSRSDHLALHMKRHF----
WT1       KT--CQRKFSRSDHLKTHTRTHTGKTSEKPFSCRWPSCQKKFARSDELVRHHNMHQRNMT
EGR1      RI--CMRNFSRSDHLTTHIRTHTG---EKPFACDI--CGRKFARSDELKRHTKIHLRQKD
ZSCA2     GL--CGKSFSQSSSLIAHQGTHTG---EKPYECLT--CGESFSWSSNLIKHQRTHTGEKP
              *  .*::*. *  *  .***   .:*: *    * . *: *..*  * . *
```

Hydrophobic core

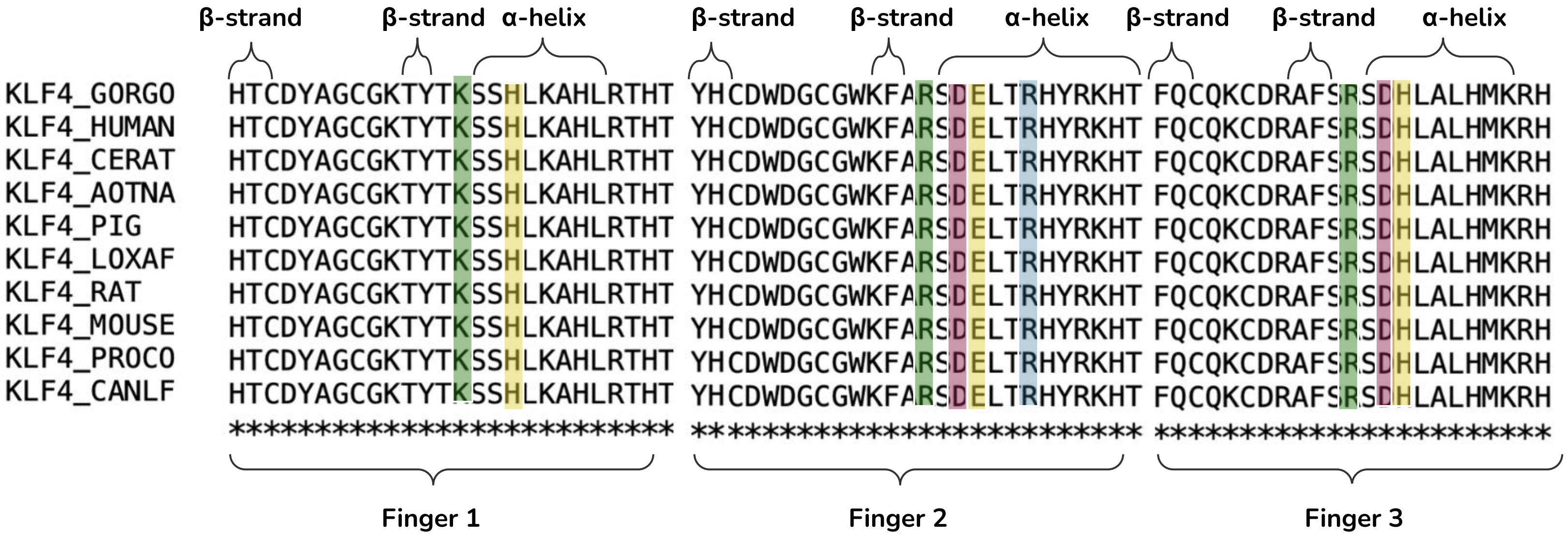


Finger 3

Hydrophobic core



Specific DNA-binding - residues conservation among species

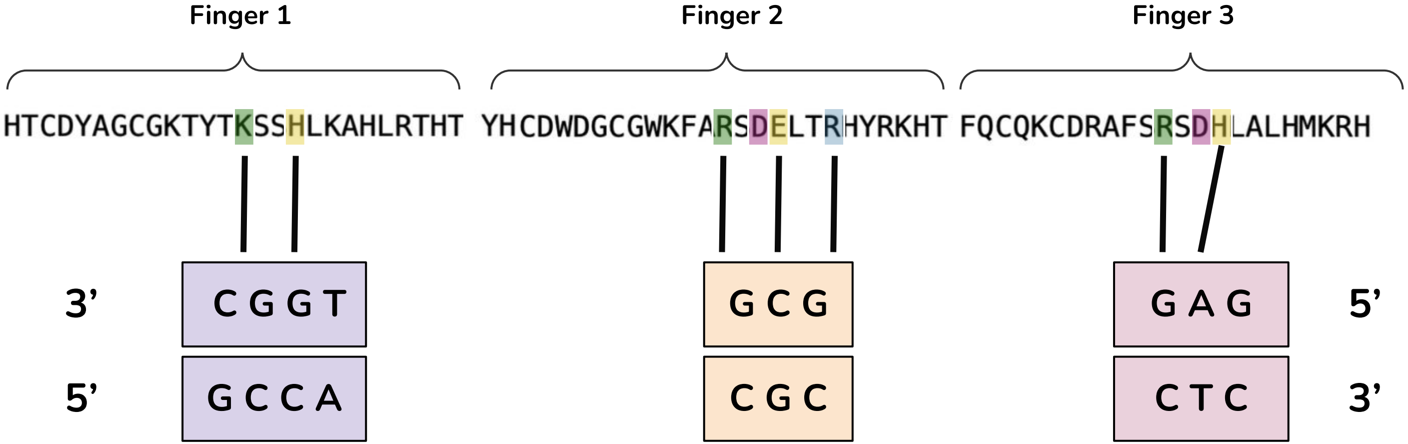


```
KLF4_GORGO    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG      HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT      HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
              ************************** ************************** ***********************
```

🟩 Position -1
🟥 Position 2
🟨 Position 3
🟦 Position 6

DNA-binding consensus sequence

KLF4 consensus binding site



Finger 1

Finger 2

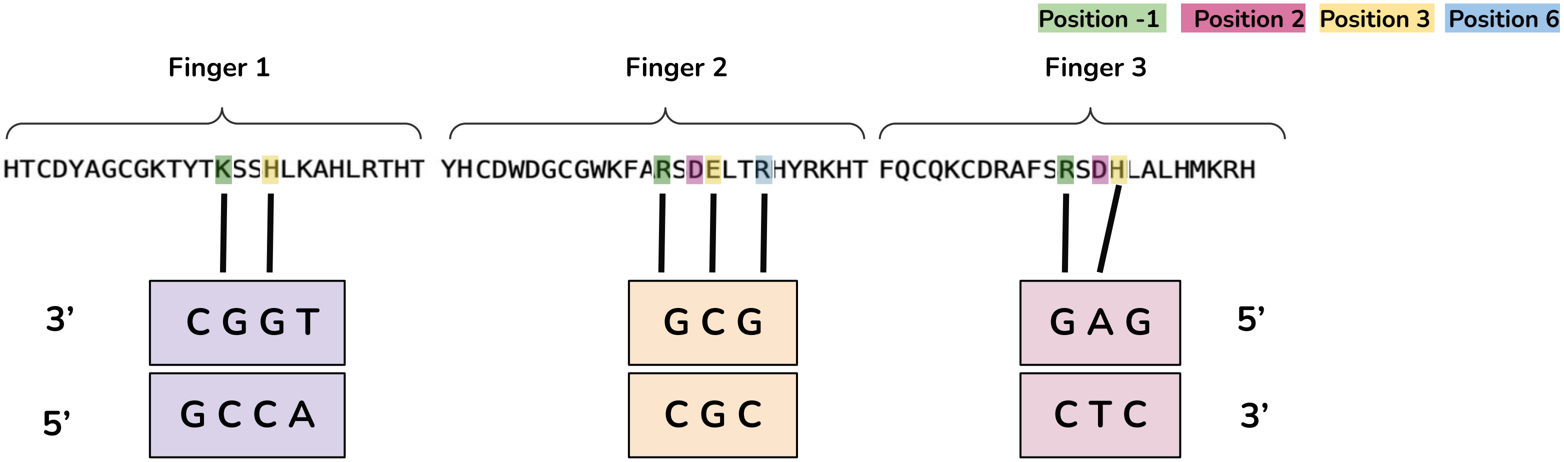
Finger 3

HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH

3’ C G G T G C G G A G 5’

5’ G C C A C G C C T C 3’

Residues involved in specific DNA binding



Position -1 Position 2 Position 3 Position 6

Finger 1 Finger 2 Finger 3

HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH

3’ C G G T G C G G A G 5’

5’ G C C A C G C C T C 3’

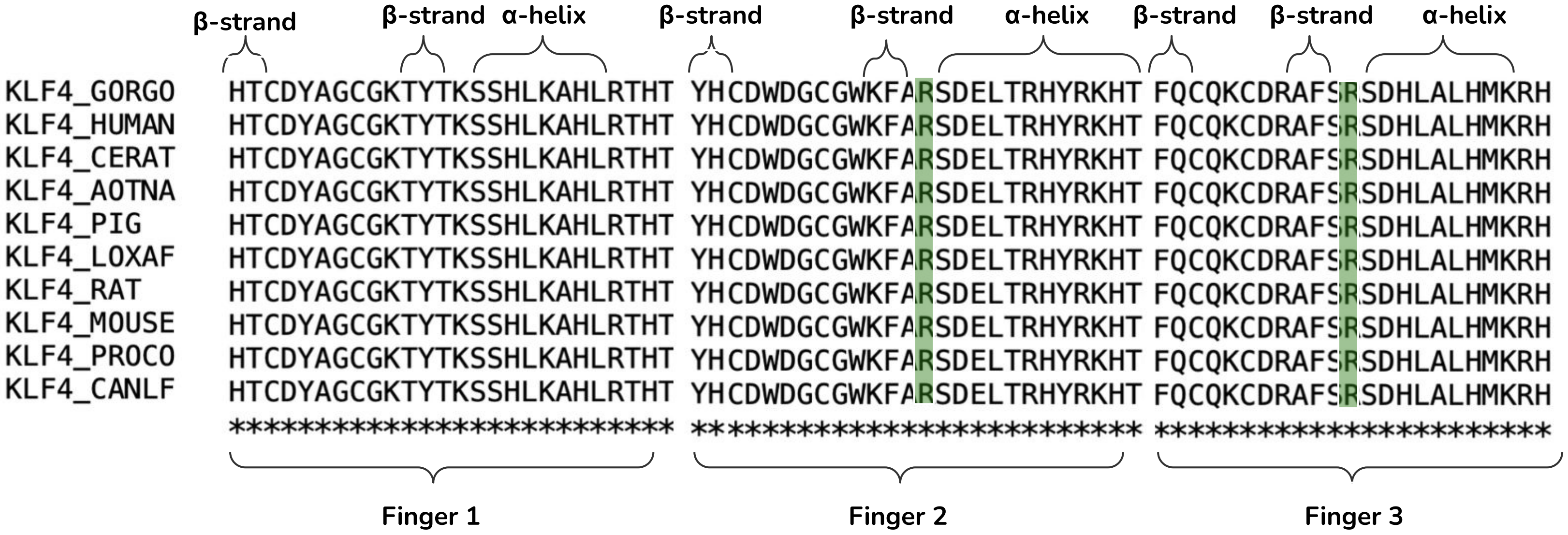
Important residues

Finger 1 → Lys413, His416 (positions -1,3)

Finger 2 → Arg443, Asp445, Glu446, Arg449 (positions -1,2,3,6)

Finger 3 → Arg471, Asp473, His474 (positions -1,2,3)

Position -1



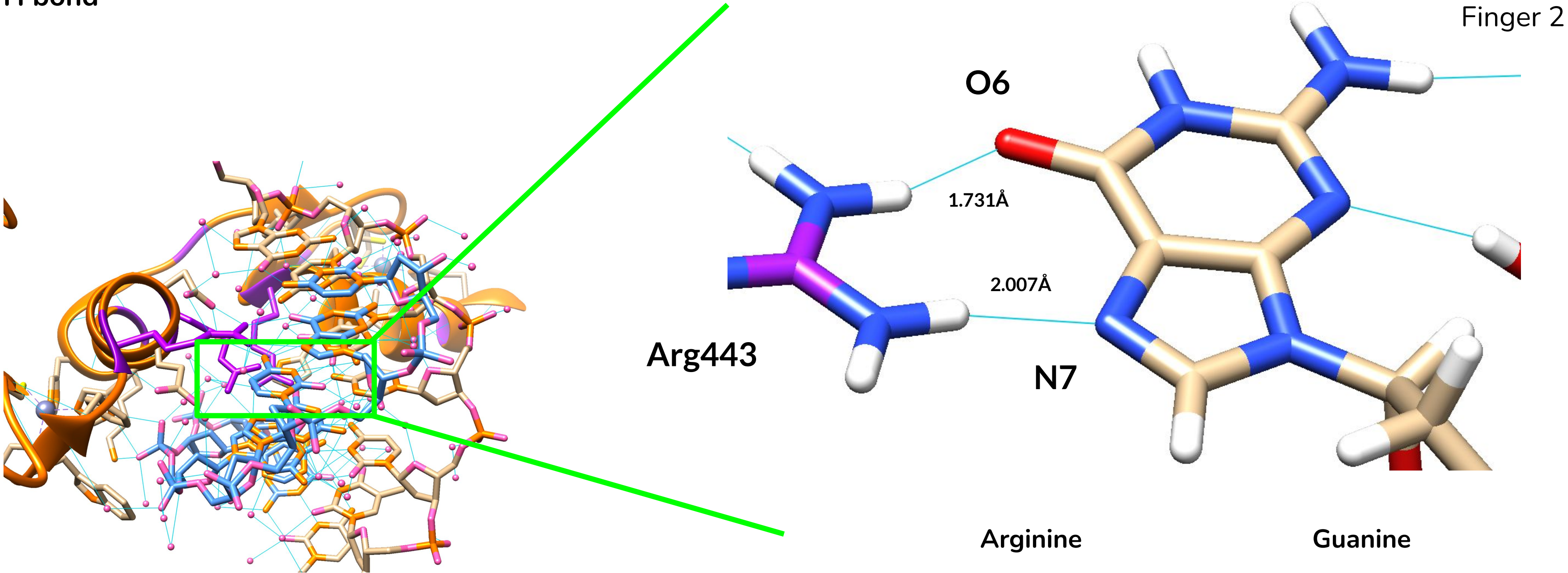
```
KLF4_GORGO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
             ************************** ************************** ***********************
```

<u>G</u> C G

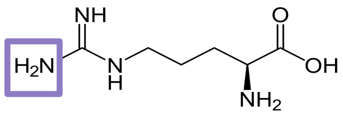
<u>G</u> A G

Position -1

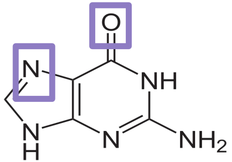
H bond



Arginine

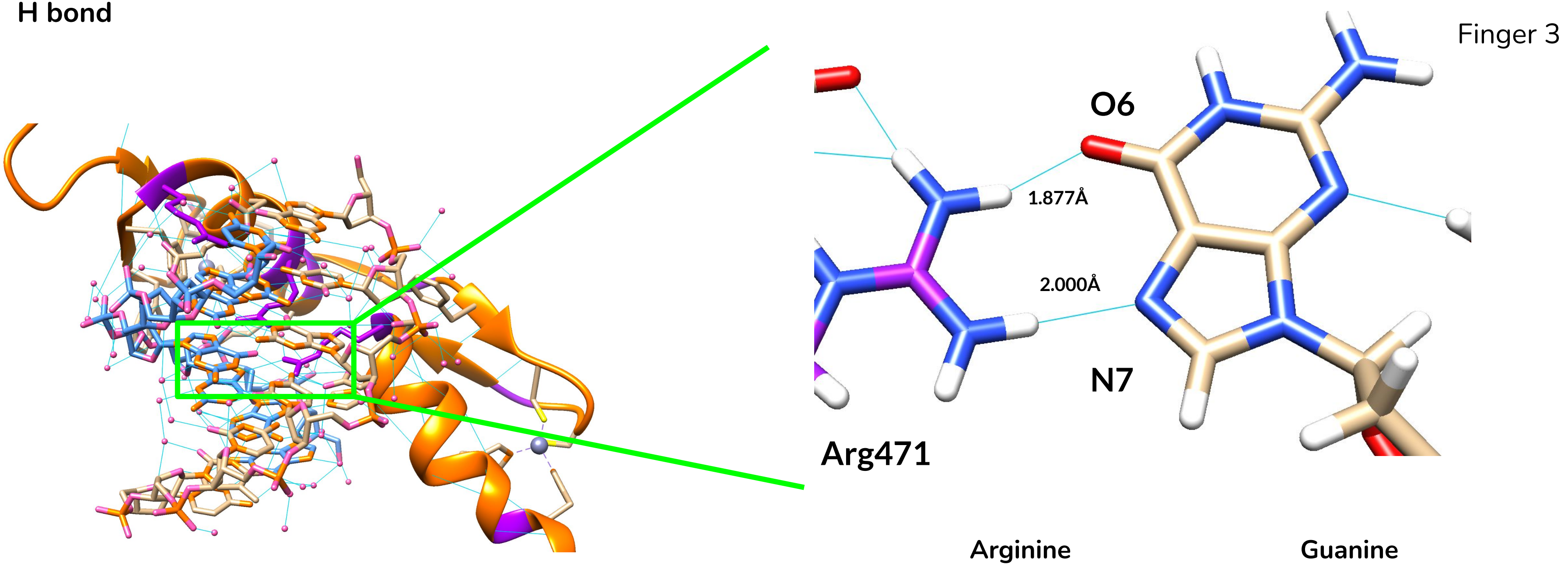


Guanine

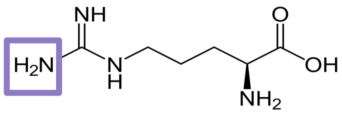


Position -1

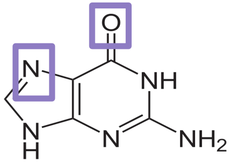
H bond



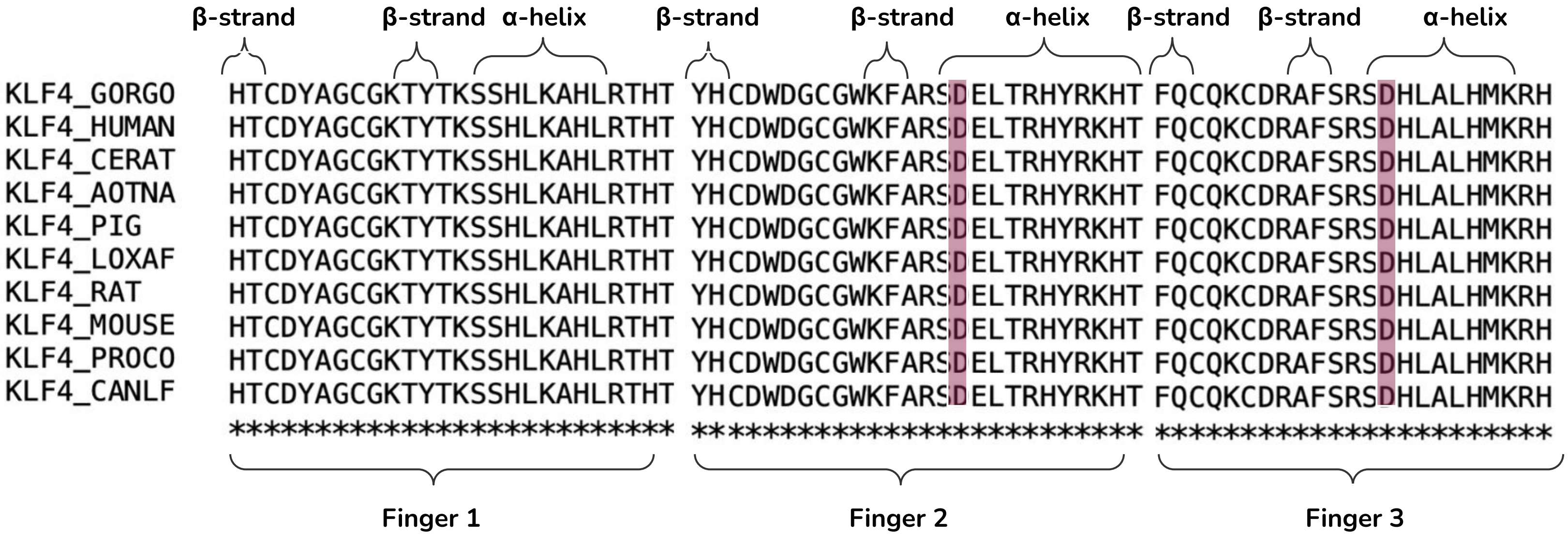
Arginine



Guanine



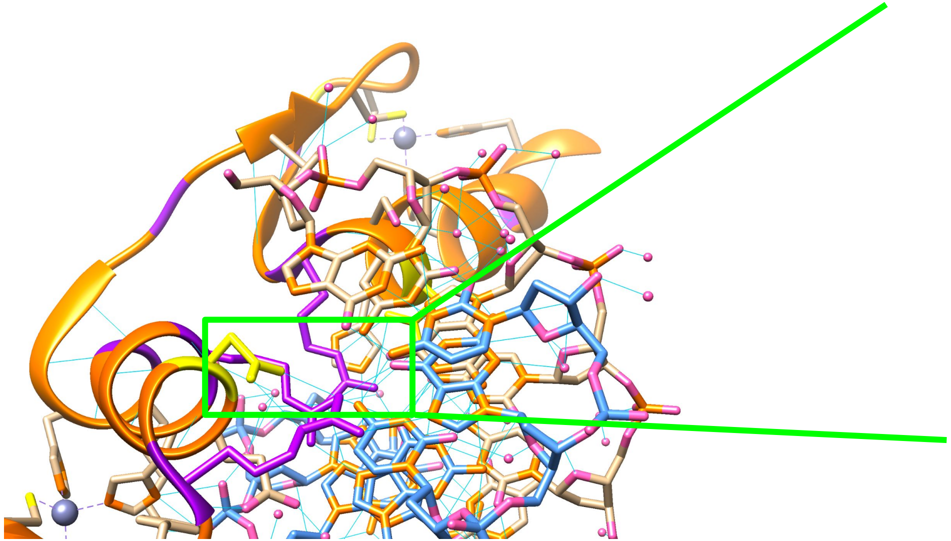
Position 2



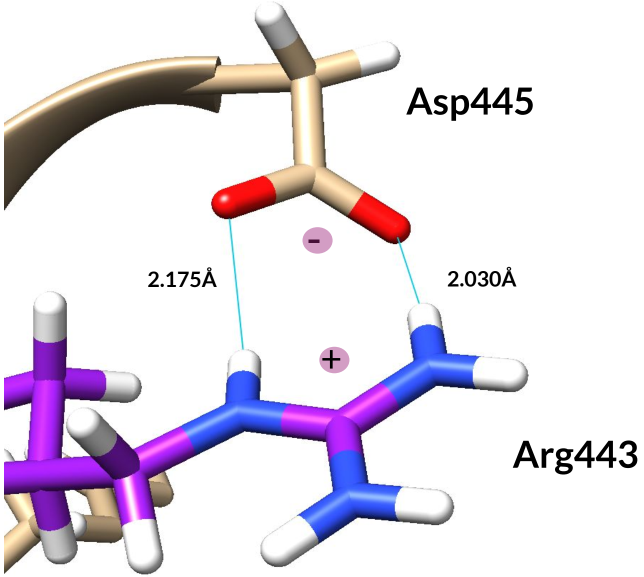
```
KLF4_GORGO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT     HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF   HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
             ************************** ************************** ***********************
```

Position 2

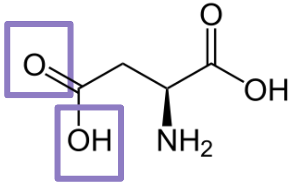
Salt bridge



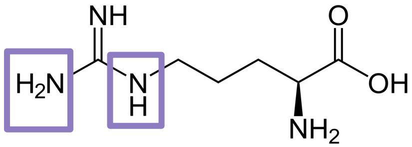
Finger 2



Aspartic

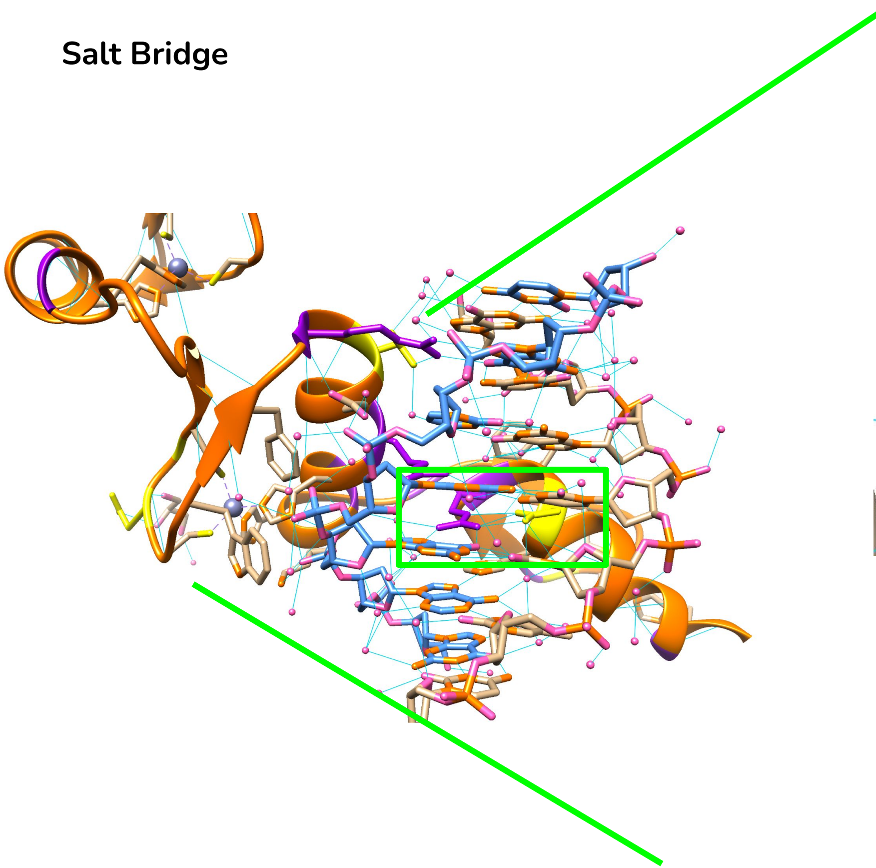


Arginine

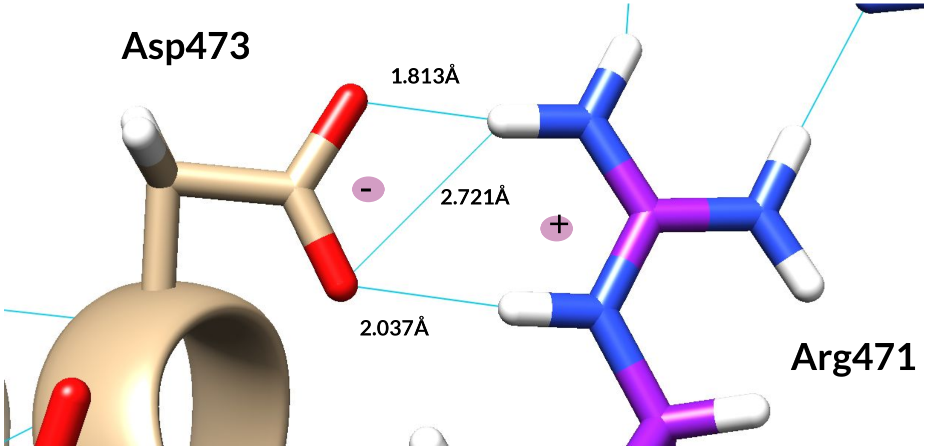


Position 2

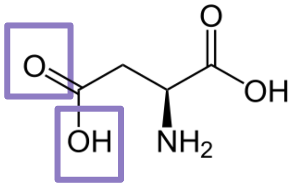
Salt Bridge



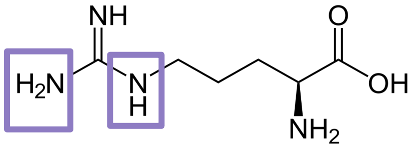
Finger 3



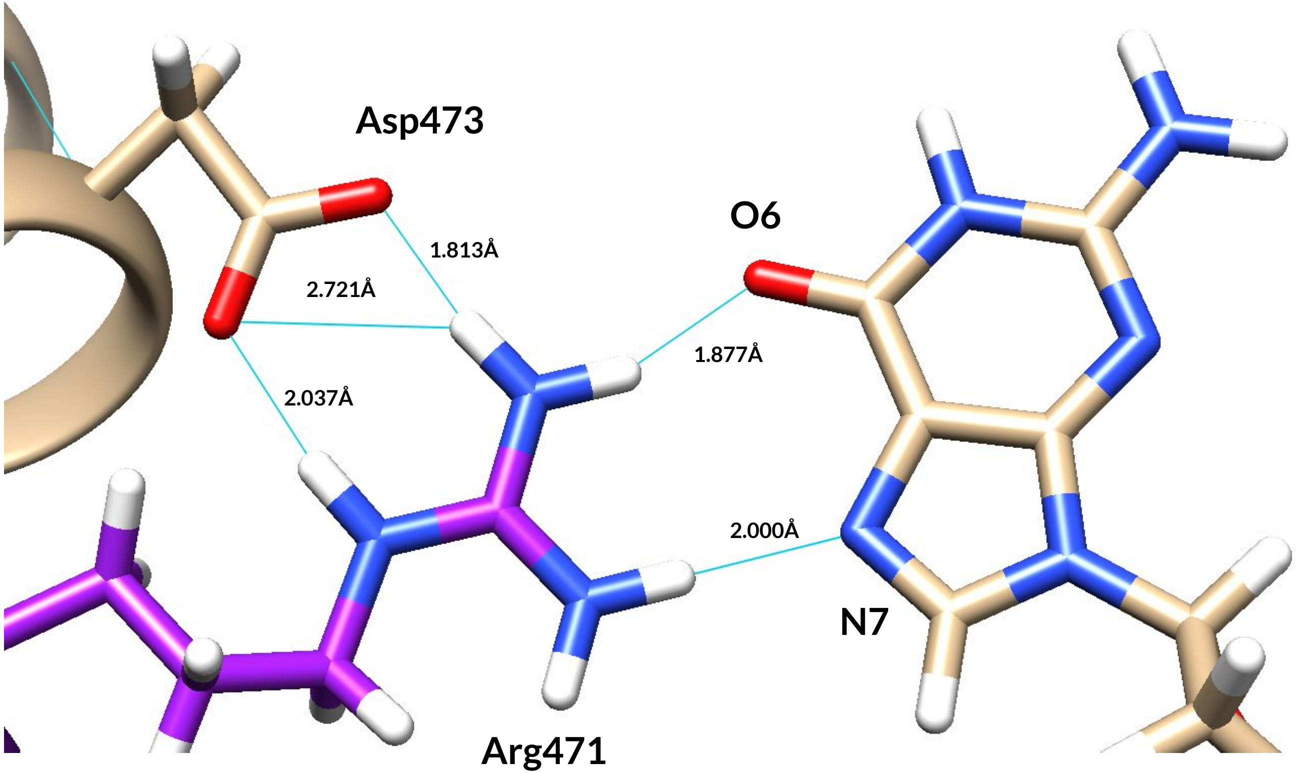
Aspartic



Arginine

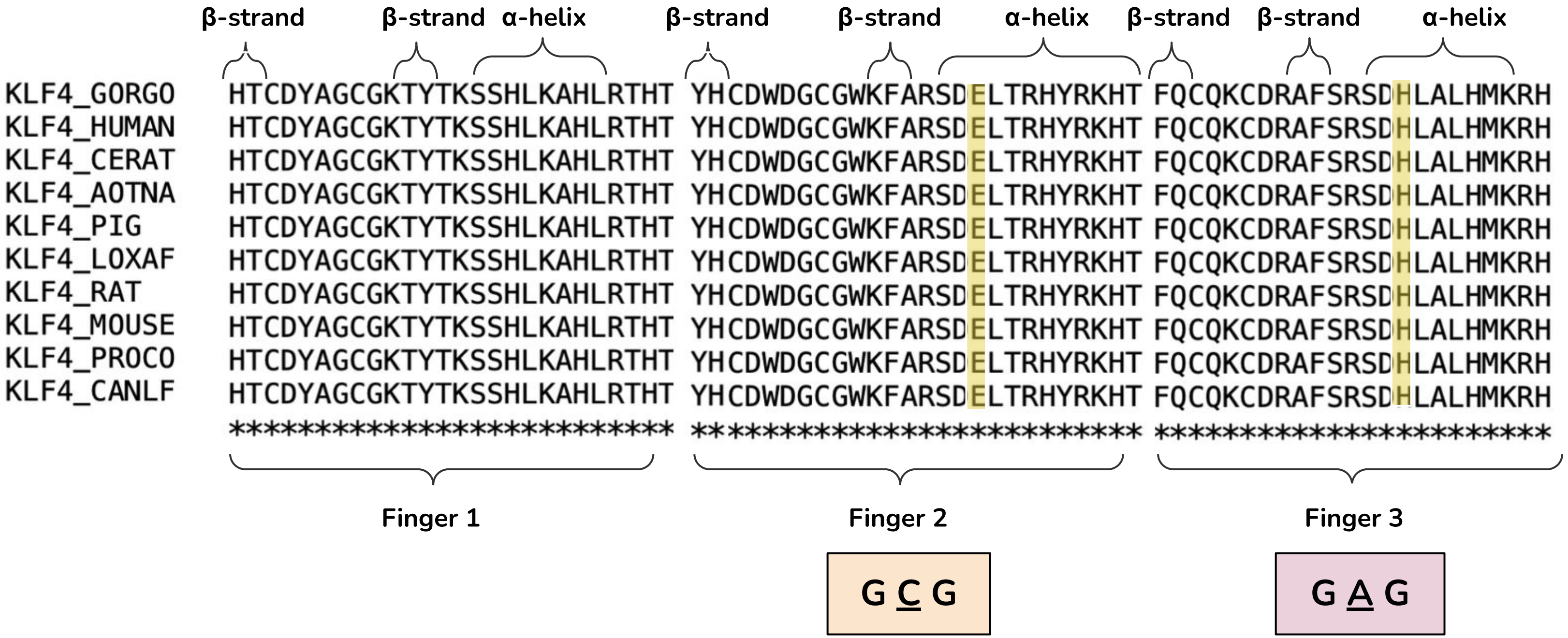


Interaction between position -1 and position 2



Finger 3

Position 3



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

```
KLF4_GORGO  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF  HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
            ************************** ************************** ***********************
```

Finger 1

Finger 2

G C G

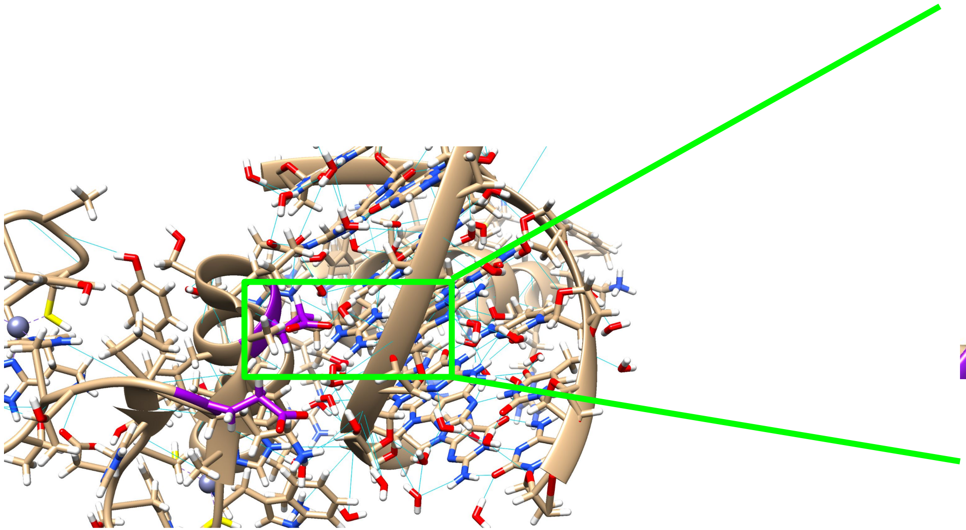
Finger 3

G A G

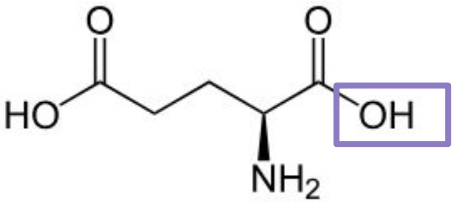
Position 3

Finger 2

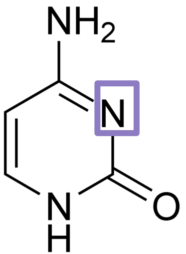
H bond



Glutamic acid



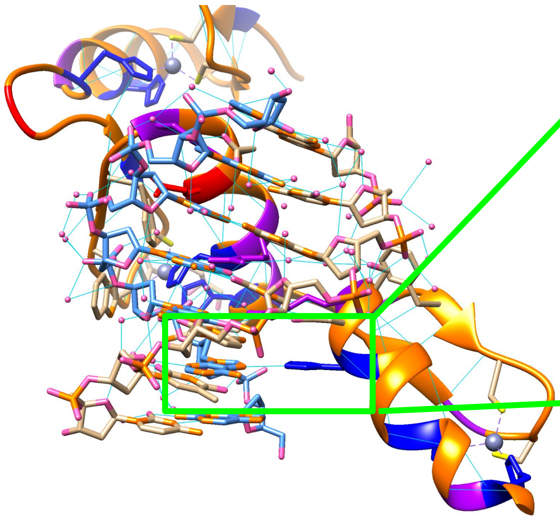
Cytosine

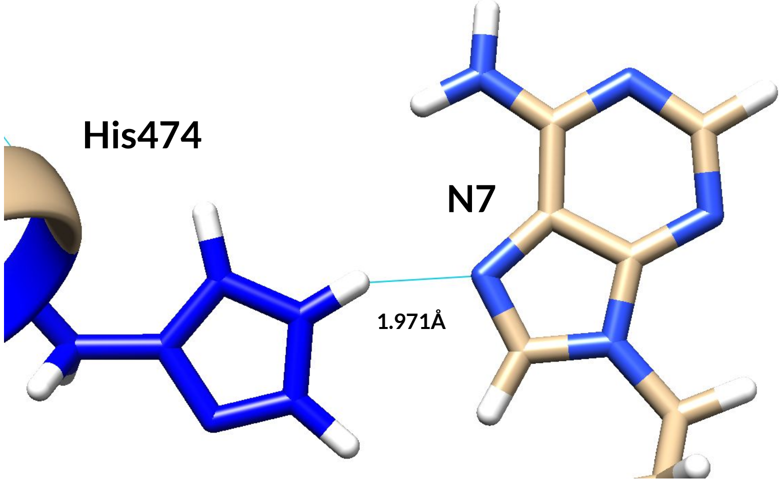


Position 3

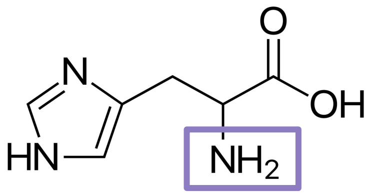
H bond

Finger 3

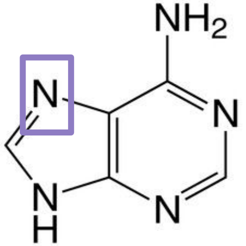




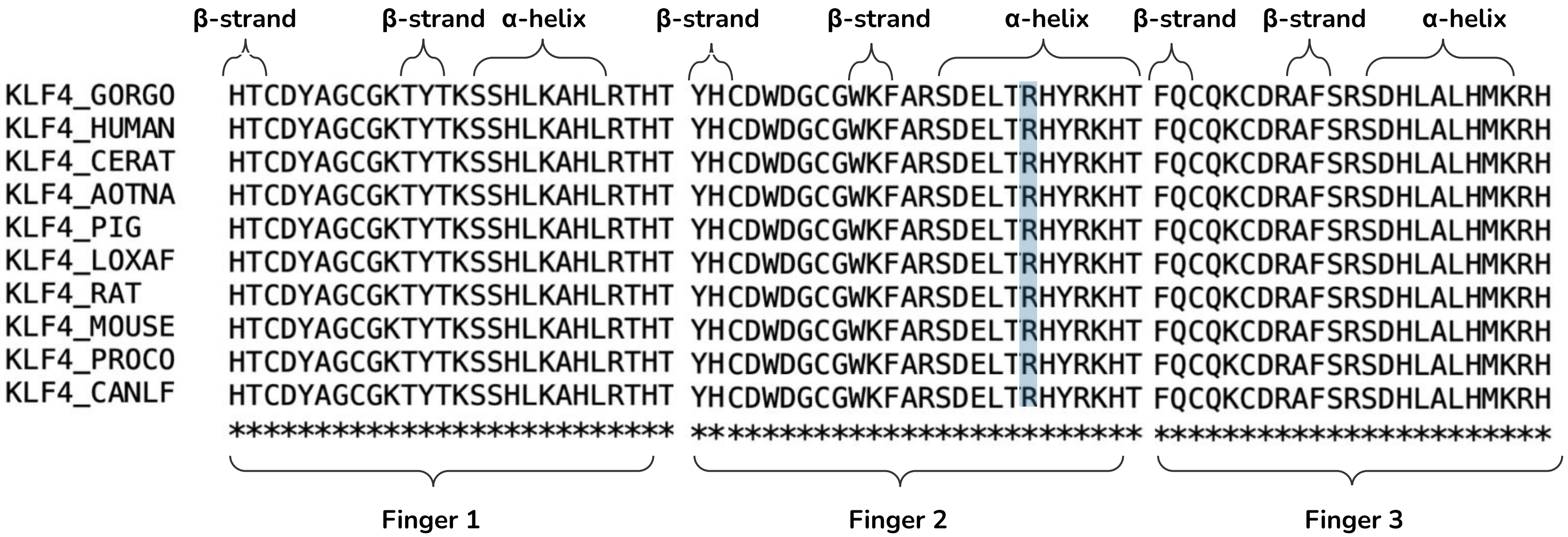
Histidine



Adenine



Position 6



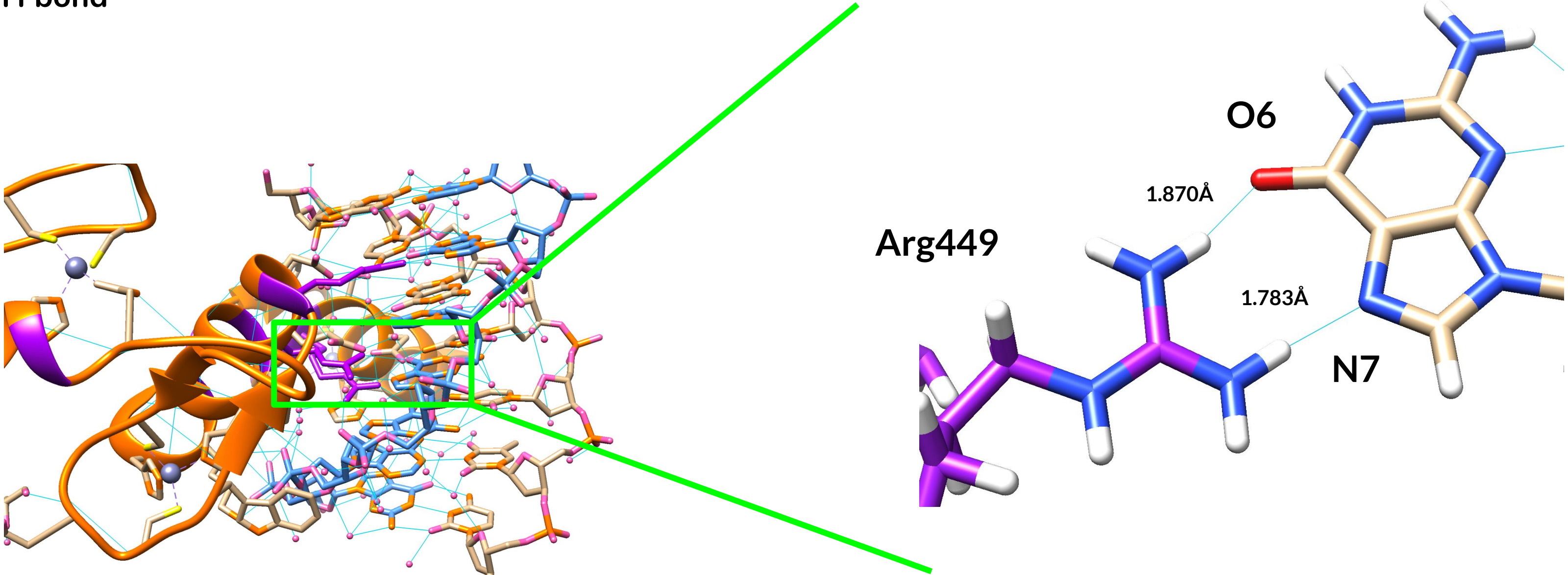
```
KLF4_GORGO    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_HUMAN    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CERAT    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_AOTNA    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PIG      HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_LOXAF    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_RAT      HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_MOUSE    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_PROCO    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
KLF4_CANLF    HTCDYAGCGKTYTKSSHLKAHLRTHT YHCDWDGCGWKFARSDELTRHYRKHT FQCQKCDRAFSRSDHLALHMKRH
              ************************** ************************** ***********************
```

G C G

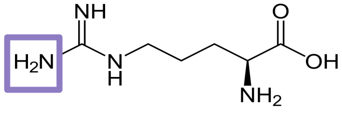
Position 6

H bond

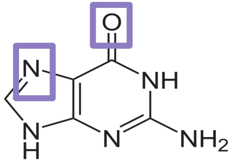
Finger 2



Arginine



Guanine



Non-specific DNA-phosphate binding residues

```
KLF4      PPGSCMPEEPKPKR-GRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHC
WT1       YPGCNKRYFKLSHL-QMHSRKHTGEKPYQCDFKDCERRFSRSDQLKRHQRRHTGVKPFQC
EGR1      YQSQLIKPSRMRKYPNRPSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQC
ZSCA2     ECG---QKFSQSSALITHRRTHTGEKPYQCG--ECGKNFSRSSNLATHRRTHLVEKPYKC
            .                       .: *    * : :::*..*  * * *   **::*

KLF4      DWDGCGWKFARSDELTRHYRKHTG---HRPFQCQK--CDRAFSRSDHLALHMKRHF----
WT1       KT--CQRKFSRSDHLKTHTRTHTGKTSEKPFSCRWPSCQKKFARSDELVRHHNMHQRNMT
EGR1      RI--CMRNFSRSDHLTTHIRTHTG---EKPFACDI--CGRKFARSDERKRHTKIHLRQKD
ZSCA2     GL--CGKSFSQSSSLIAHQGTHTG---EKPYECLT--CGESFSWSSNLIKHQRTHTGEKP
              *  .*::*. *  *  .***   .:*: *    * . *: *..:  * . *
```

Finger 1 → Lys409

Finger 2 → Trp439, Ala442

Finger 3 → Arg467, Ser470

Conclusions

JUN	KLF4
- There is a highly conservation between species and also between homologous in the two main domains. - Basic region binds a specific binding site of the DNA, the AP-1 site. - Leucine Zipper region is a highly conserved and functionally important structure that plays a central role in gene expression regulation in eukaryotic cells by mediating the dimerization of transcription factors and their subsequent DNA binding.	- During evolution, important residues have been conserved - In the C2H2 family, the hydrophobic core holds together the antiparallel beta-sheets and the alpha-helices. - Hydrophobic residues are essential for protein stability. - Sequence conservation is not observed in the residues involved in nonspecific DNA-phosphate binding but it is in those involved in specific DNA binding.

PEM Questions

1. bZIP TFs:

a) Have two motifs: DNA-binding basic region and leucine zipper.
b) Both of them are conserved within evolution.
c) Both are correct.
d) Free in solution becomes alpha-helically folded.
e) All of them are correct.

2. Transcription factors from AP1:

a) Are involved in cell growth, proliferation, survival, apoptosis, transformation, oncogenesis.
b) Are Jun and KLF4.
c) Both are correct.
d) Only binds TRE sites.
e) All of them are correct.

3. The most important interactions between Jun and DNA are:

a) Van der Waals.
b) Hydrogen bonds.
c) Both are correct.
d) Salt bridges.
e) All of them are correct.

4. Leucine zipper region is characterized by:

a) Alanine residue repeats every fourth position
b) Leucine residue repeats every seventh position
c) Electrostatic interactions between d and g positions
d) Hydrophobic interactions between residues in e position
e) All are correct

5. Which of the following residues are implicated in both hydrophobic interactions and electrostatic interaction?

a) Val293 and Thr196
b) Asn300 and Lys176
c) Lys176 and Val294
d) Leu296 and Leu173
e) Glu186 and Thr196

PEM Questions

6. Which is the consensus sequence of C2H2 family?

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

7. Which are the two Zn fingers that most contribute to the DNA binding specificity?

a) Finger 1 and finger 3
b) Finger 1 and finger 2
c) Finger 2 and finger 3
d) None of the above
e) All of them are exactly the same

8. Regarding the hydrophobic core, select the true statement:

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

9. Select the correct answers about KLF4:

1. According to SCOP it belongs to the helix loop helix family.
2. It has 4 zinc fingers
3. Each of its fingers contains a β sheet and two α helix
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

10. Choose the correct answer about KLF4 and its specific binding to DNA:

a) The residues involved in this interaction with the DNA are extremely conserved among different species.
b) The arginine residues from helix position −1 are stabilized by a salt bridge with aspartate residues of helix position 2.
c) The contribution of the first zinc finger is smaller than that of the two C-terminal zinc fingers
d) The arginine residues of helix positions -1 and 6 confer the most of the DNA-binding specificity.
e) All of them are true

Bibliography

- Branden, C.-I., & Tooze, J. (1999). Introduction to Protein Structure (2nd ed.). CRC Press.
- Cruzalegui, F. H., Hardingham, G. E., & Bading, H. (1999). c-Jun functions as a calcium-regulated transcriptional activator in the absence of JNK/SAPK1 activation. The EMBO journal, 18(5), 1335–1344. https://doi.org/10.1093/emboj/18.5.1335
- Glover, J. N., & Harrison, S. C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature, 373(6511), 257–261. https://doi.org/10.1038/373257a0
- Junius, F. K., O'Donoghue, S. I., Nilges, M., Weiss, A. S., & King, G. F. (1996). High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. The Journal of biological chemistry, 271(23), 13663–13667. https://doi.org/10.1074/jbc.271.23.13663
- Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Krüppel-like transcription factors. *Genome Biol* 4, 206 (2003). https://doi.org/10.1186/gb-2003-4-2-206
- Liu, Y., Olanrewaju, Y. O., Zheng, Y., Hashimoto, H., Blumenthal, R. M., Zhang, X., & Cheng, X. (2014). Structural basis for Klf4 recognition of methylated DNA. *Nucleic acids research*, *42*(8), 4859–4867. https://doi.org/10.1093/nar/gku134
- Lively, T. N., Nguyen, T. N., Galasinski, S. K., & Goodrich, J. A. (2004). The basic leucine zipper domain of c-Jun functions in transcriptional activation through interaction with the N terminus of human TATA-binding protein-associated factor-1 (human TAF(II)250). The Journal of biological chemistry, 279(25), 26257–26265. https://doi.org/10.1074/jbc.M400892200
- New human TF classification. (n.d.). Edgar-wingender.de. Retrieved February 27, 2024, from http://www.edgar-wingender.de/huTF_classification.html
- Pace, N. J., & Weerapana, E. (2014). Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules*, *4*(2), 419–434. https://doi.org/10.3390/biom4020419
- Presnell, J. S., Schnitzler, C. E., & Browne, W. E. (2015). KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes. *Genome biology and evolution*, *7*(8), 2289–2309. https://doi.org/10.1093/gbe/evv141
- Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nikolakaki, E., & Woodgett, J. R. (1991). Phosphorylation of c-jun mediated by MAP kinases. Nature, 353(6345), 670–674. https://doi.org/10.1038/353670a0

Bibliography

- Sharma, R., Choi, K. J., Quan, M. D., Sharma, S., Sankaran, B., Park, H., LaGrone, A., Kim, J. J., MacKenzie, K. R., Ferreon, A. C. M., Kim, C., & Ferreon, J. C. (2021). Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. Nature communications, 12(1), 5579. https://doi.org/10.1038/s41467-021-25761-7
- Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., Heinemann, U., & Carstanjen, D. (2011). The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cellular and molecular life sciences : CMLS*, *68*(18), 3121–3131. https://doi.org/10.1007/s00018-010-0618-x
- Suske, G., Bruford, E., & Philipsen, S. (2005). Mammalian SP/KLF transcription factors: bring in the family. *Genomics*, *85*(5), 551–556. https://doi.org/10.1016/j.ygeno.2005.01.005
- Suske G. (1999). The Sp-family of transcription factors. *Gene*, *238*(2), 291–300. https://doi.org/10.1016/s0378-1119(99)00357-1
- Swamynathan S. K. (2010). Krüppel-like factors: three fingers in control. *Human genomics*, *4*(4), 263–270. https://doi.org/10.1186/1479-7364-4-4-263