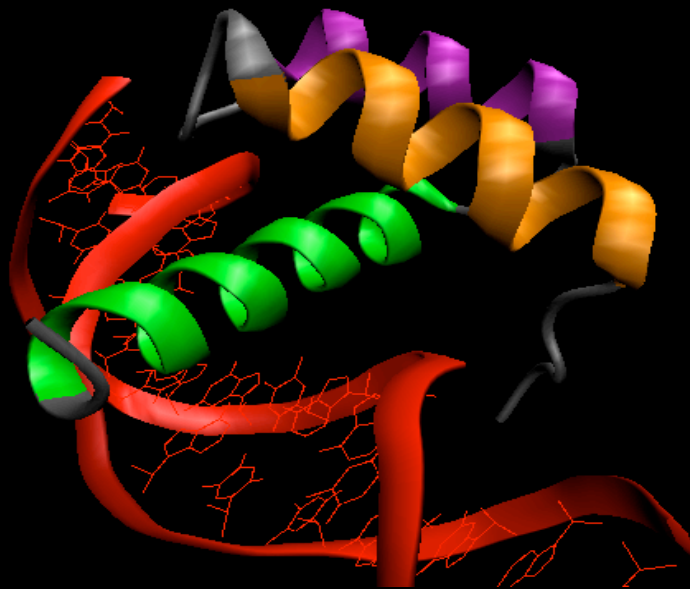


TRANSCRIPTION FACTORS:

AN INSIGHT INTO DNA BINDING DOMAINS AND MOLECULAR DOMESTICATION



Rosa Martínez, Lúdia Mateo, Aina Ollé, Clara Pérez

Outline

☐ Part1: Transcription factors

☐ Introduction

☐ Methodology

☐ Transcription factor families:

- Zinc fingers (Zn fingers)
- Basic Leucine Zipper (bZip)
- Basic-helix-loop-Helix (bHLH)
- Helix Turn Helix (HTH)

☐ Part2: Molecular domestication of DNA binding domains: CAG

☐ Introduction

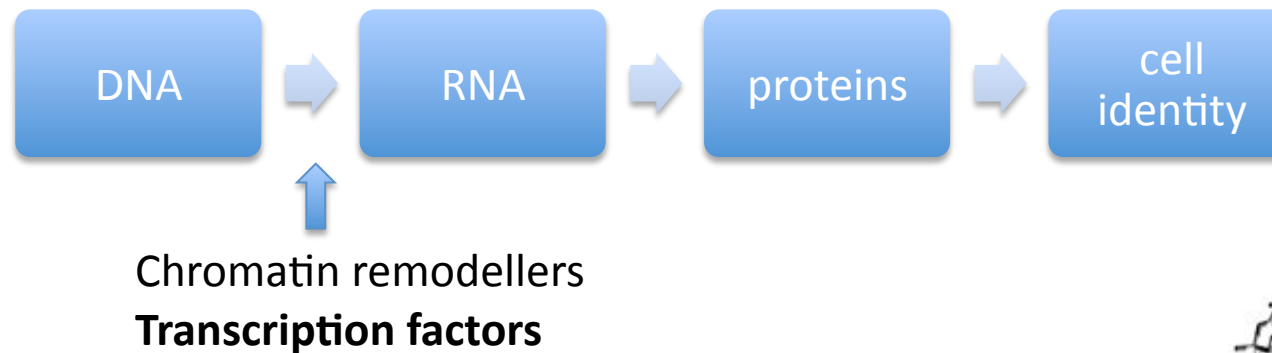
☐ Modelling strategies

☐ Final models

☐ Sequence analysis

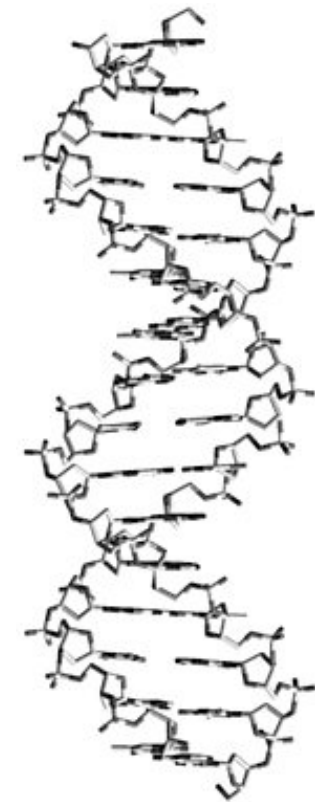
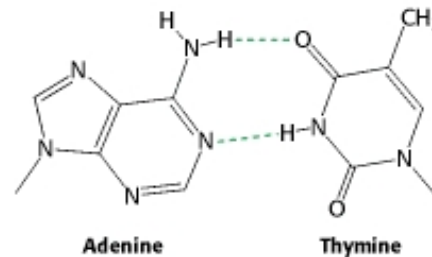
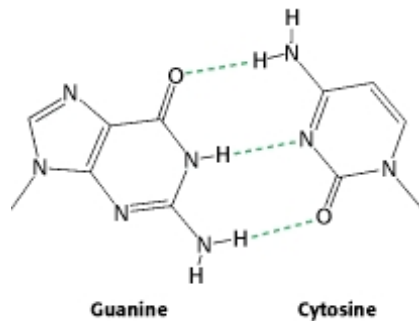
☐ Conclusions

Introduction



Sequence-specific DNA recognition:

- Sequence-dependent DNA structure and flexibility
- Physicochemical signatures of base pairs



DNA in B form

Introduction

Recognition based on loops



TBP (1YTB)



NF-kB (1SVC)

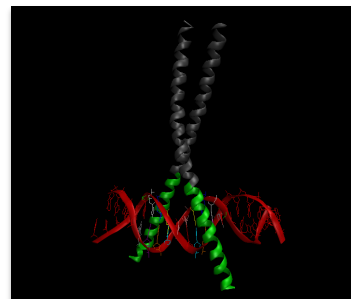


Met repressor (1CMA)

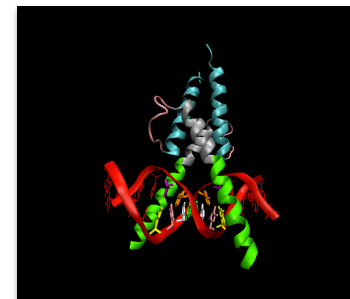
Recognition based on α -helices



Zif268(1ZAA)



C/EBPa (1NWQ)



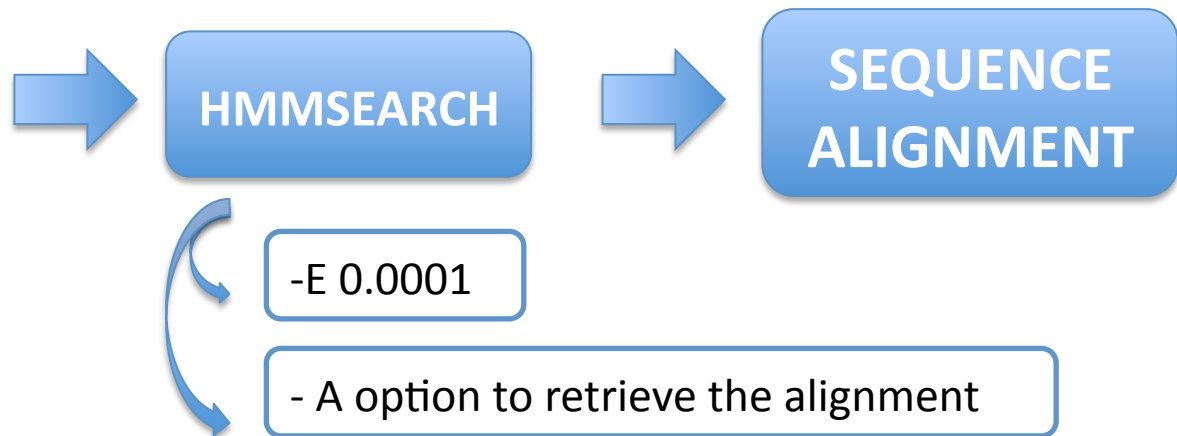
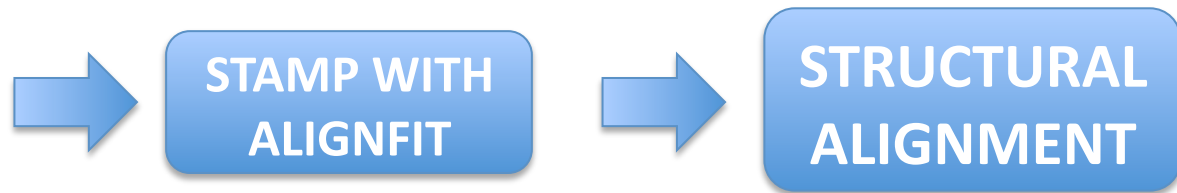
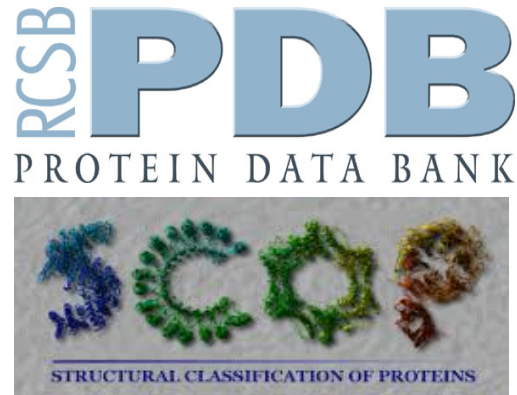
MyoD(1MDY)



Antp (9ANT)

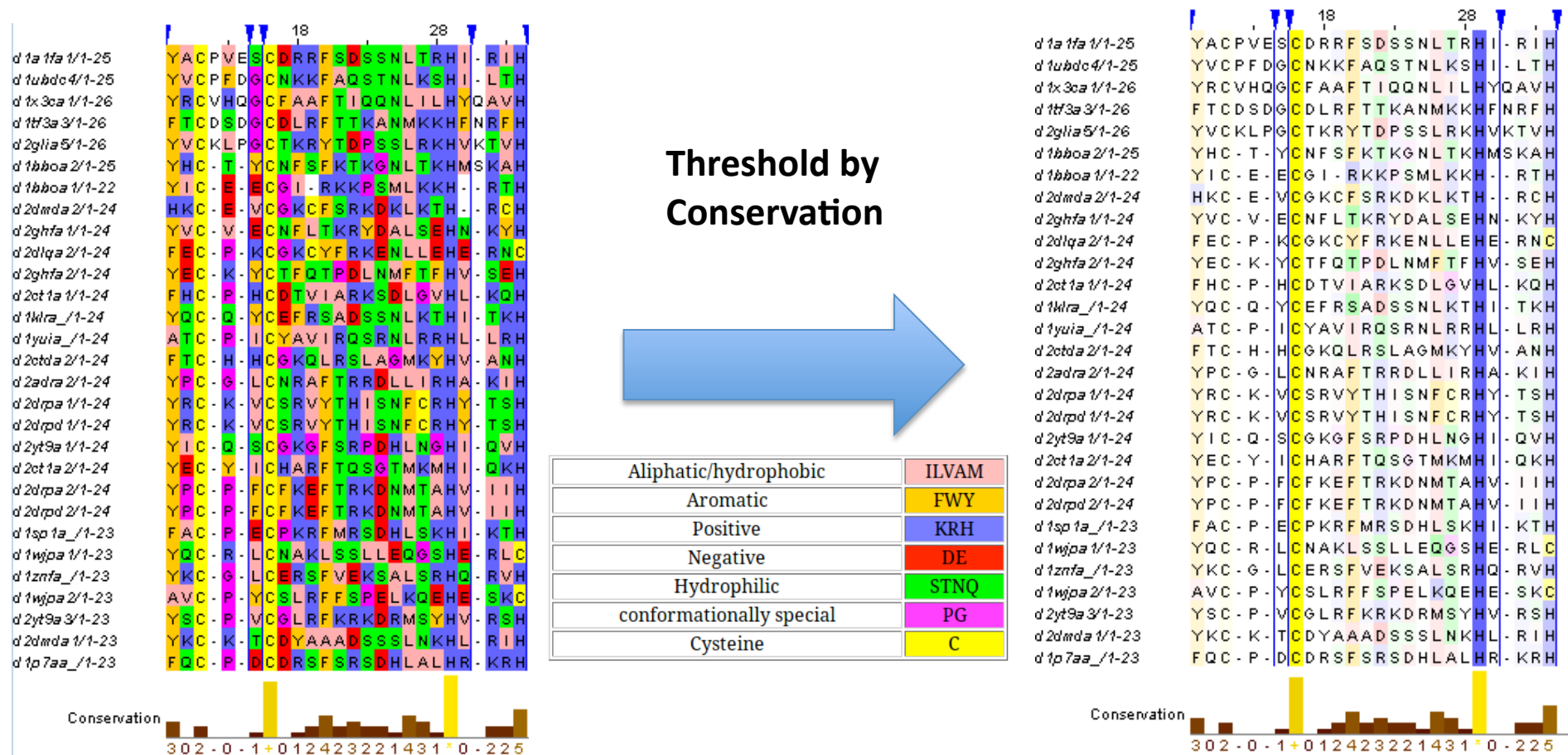
Methodology

- Alignments :



Methodology

- JalView representation of the alignments:



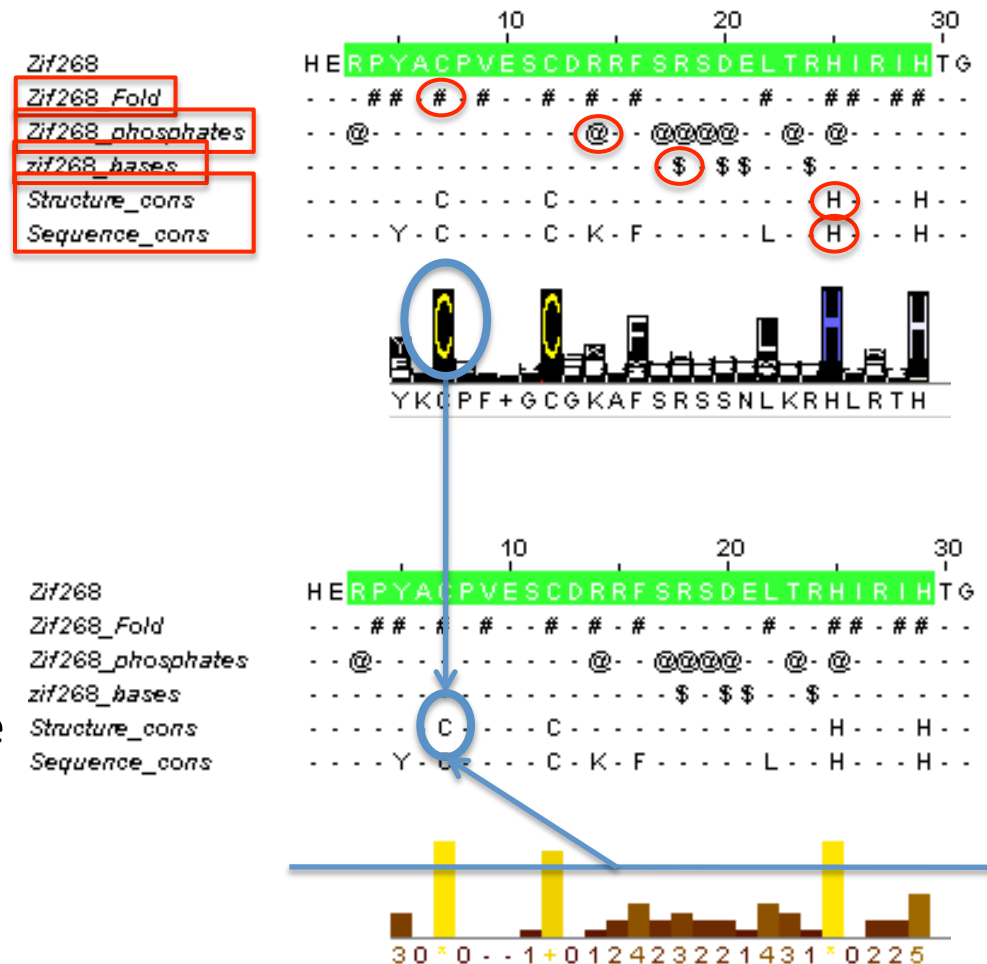
Methodology

- Main interactions of the example:

- # → Fold
- @ → Non-specific interactions
- \$ → Specific interactions
- aa. letter → Structure conservation
- aa. letter → Sequence conservation

- Consensus logo

- Manual annotation of the most conserved residues: conservation score



Zinc Fingers TFs

Zn finger domain:

28-30 amino acids

SCOP classification:

- Class: g: Small proteins
- Fold: g.37: beta-beta-alpha zinc fingers
- Superfamily: g.37.1 beta-beta-alpha zinc fingers
- Family: g.37.1.1: Classic zinc finger, C2H2

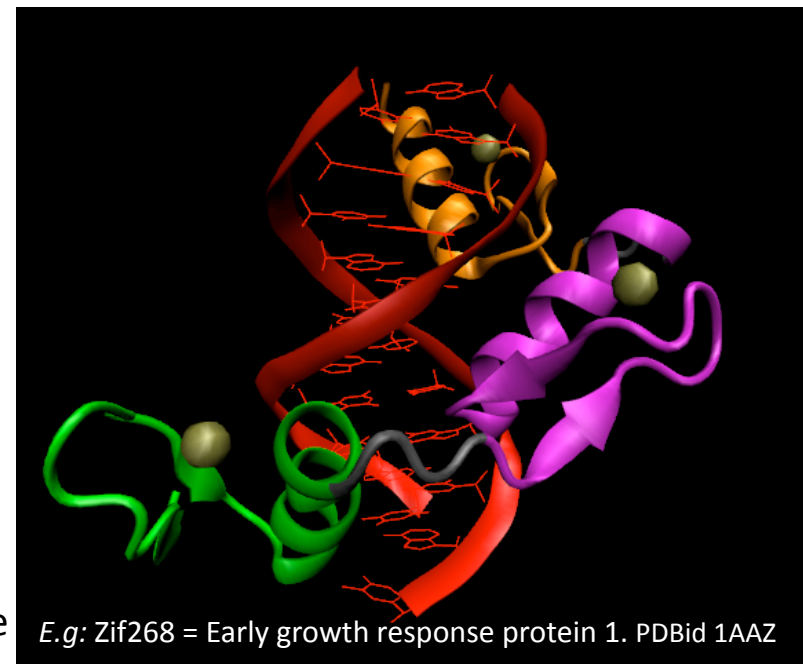
•Functions:

- Replication, reparation, transcription
- Metabolism, signaling
- Cell proliferation, apoptosis

•**Tandem** → major groove DNA

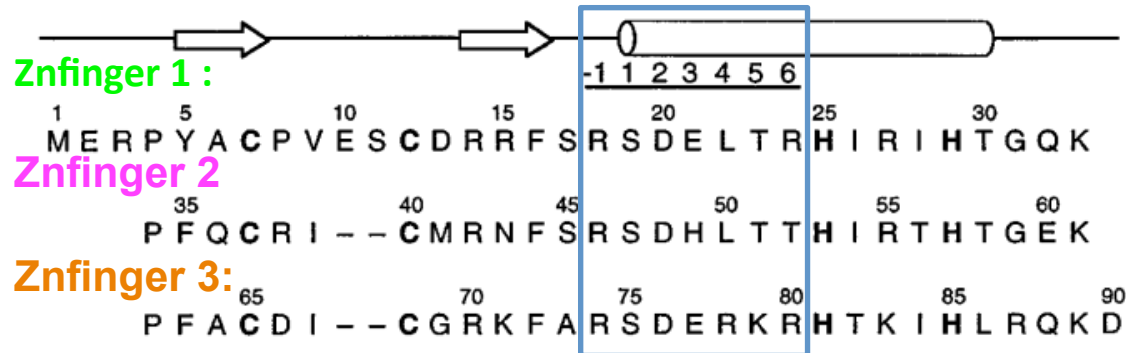
•**Interaction** with biomolecules: DNA, RNA, proteins.

•**Example:** Zif268 = **EGR-1** (Early growth response protein 1) → critical role in neuronal plasticity



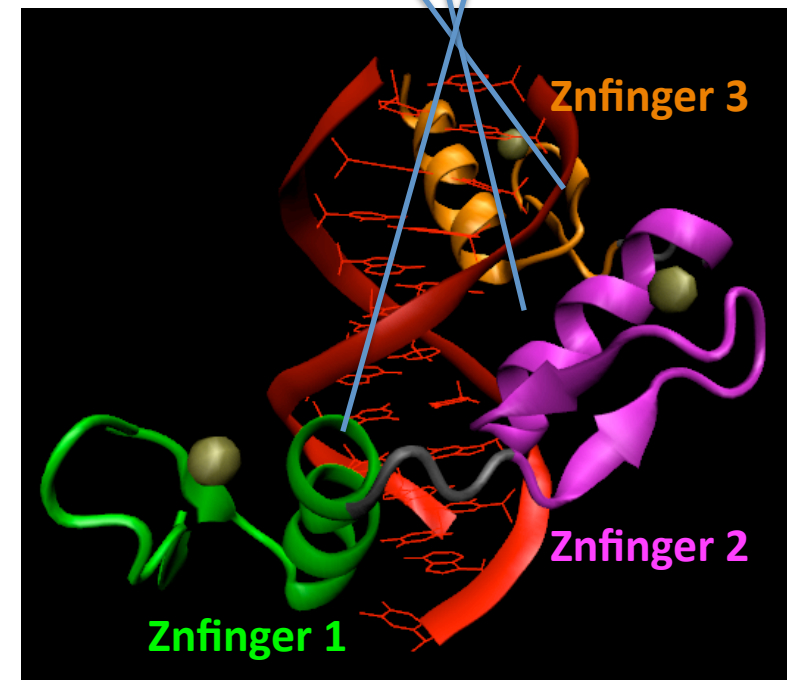
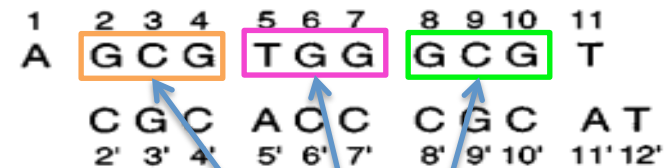
Cys2His2 zn finger (Zif268): binding to DNA

Zif268 Zn fingers sequence:



- N-terminal portion of alfa-helix
- 3 bp subsite per finger → "antiparallel"

DNA-binding site:



Zinc Fingers TFs



- Entries bound to DNA: 31
- ↳ Structures superimposed: 59



- PFAM: Zn fingers (PF00096)
- Hmmer hits (<0.0001):10663

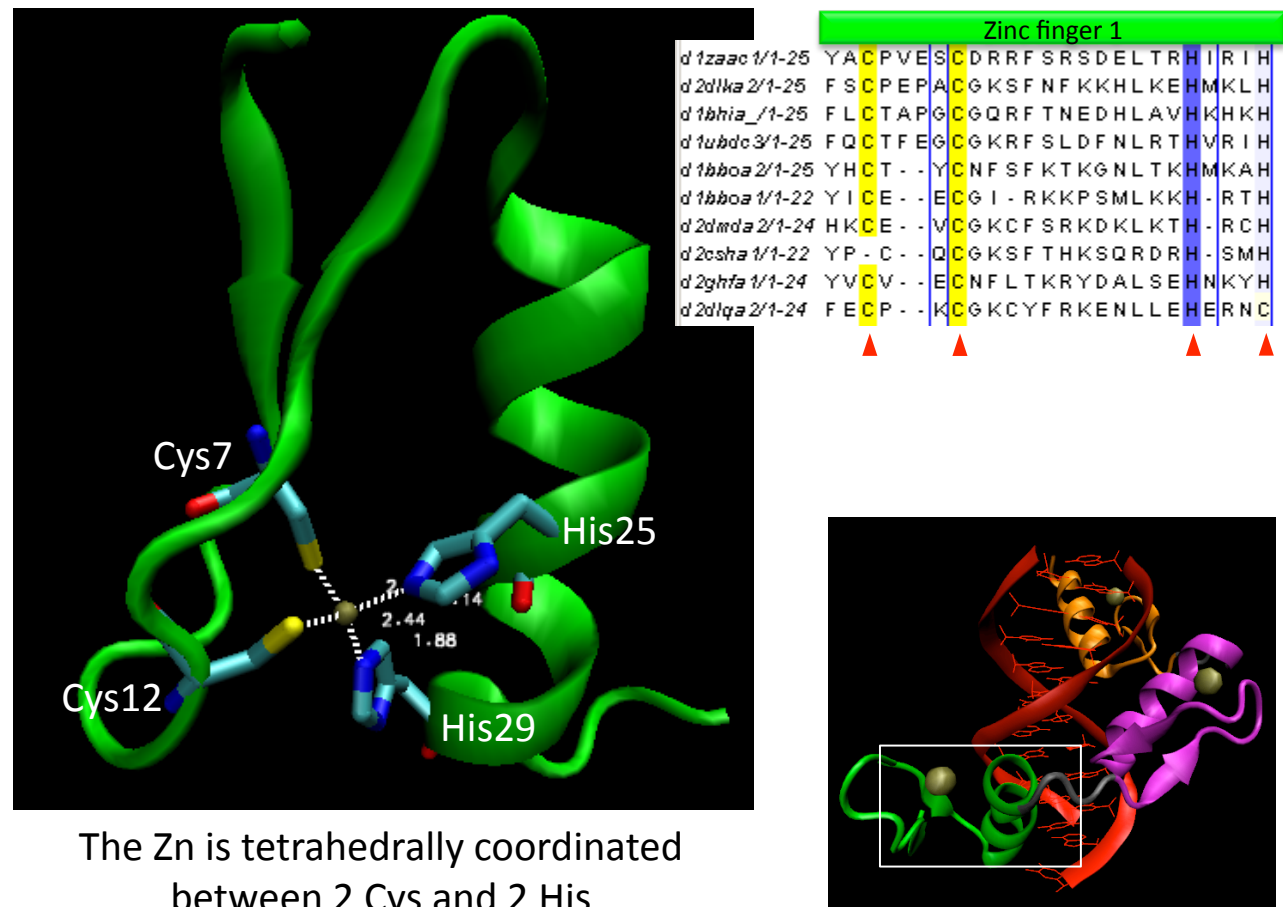
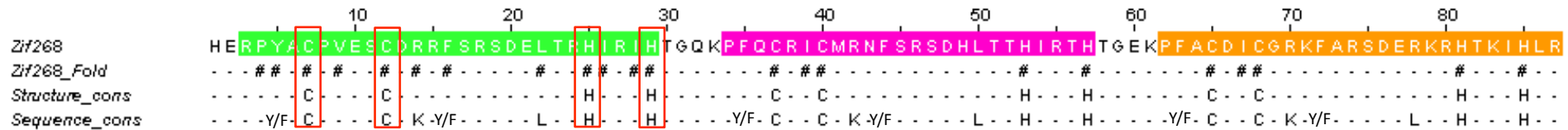
↳ Sequences aligned: 10663



Superimposition of the Zn finger domains

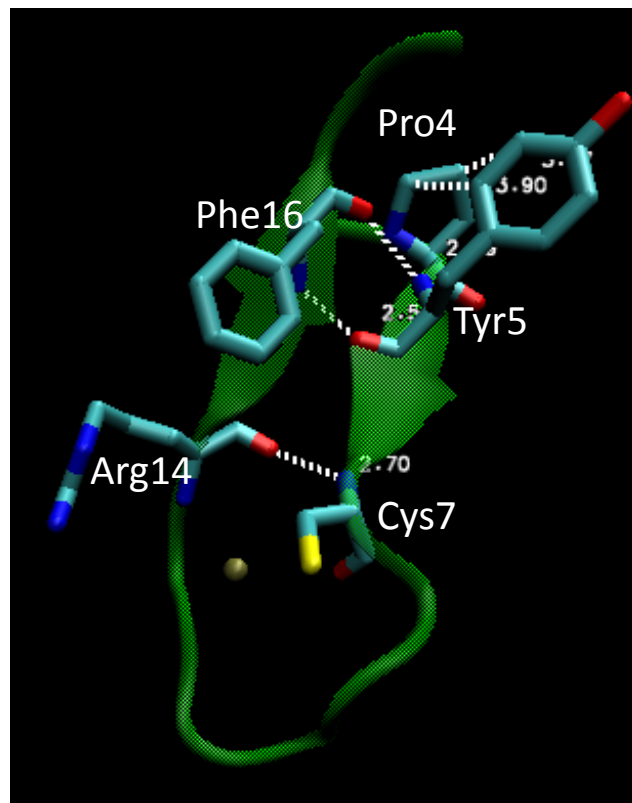
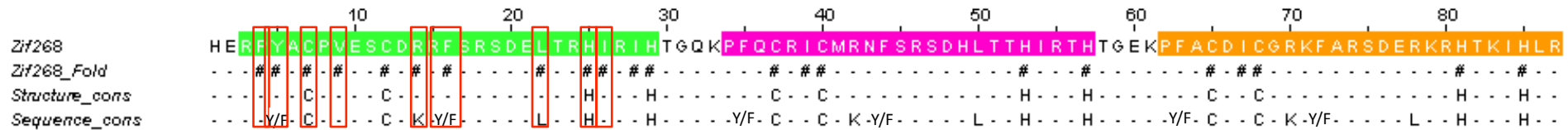
Cys2His2 zn finger (Zif268): structure

Each Zn finger domain Cys2His2 consist in a beta- hairpin and alfa-helix. The structure is maintained by the hydrophobic core, the zinc ion and the interaction between residues.

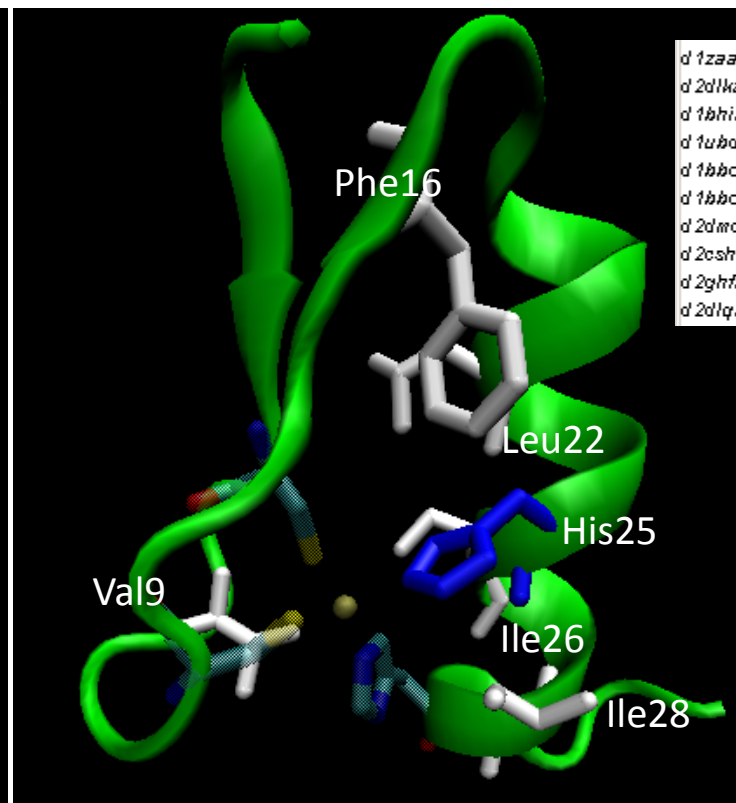


Cys2His2 zn finger (Zif268): structure

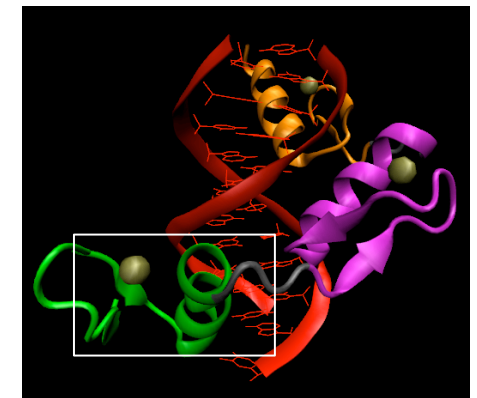
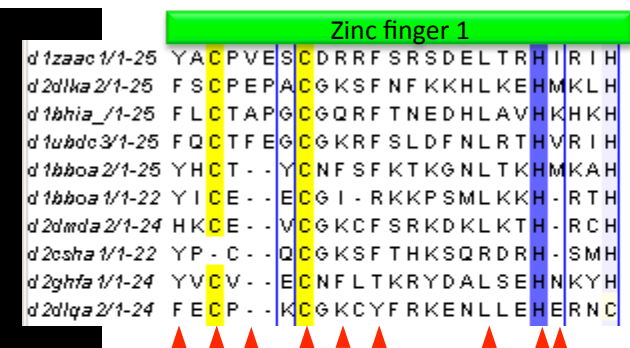
Each Zn finger domain Cys2His2 consist on a beta- hairpin and alfa-helix. The structure is maintained by the hydrophobic core, the zinc ion and the interaction between residues.



Interactions between residues



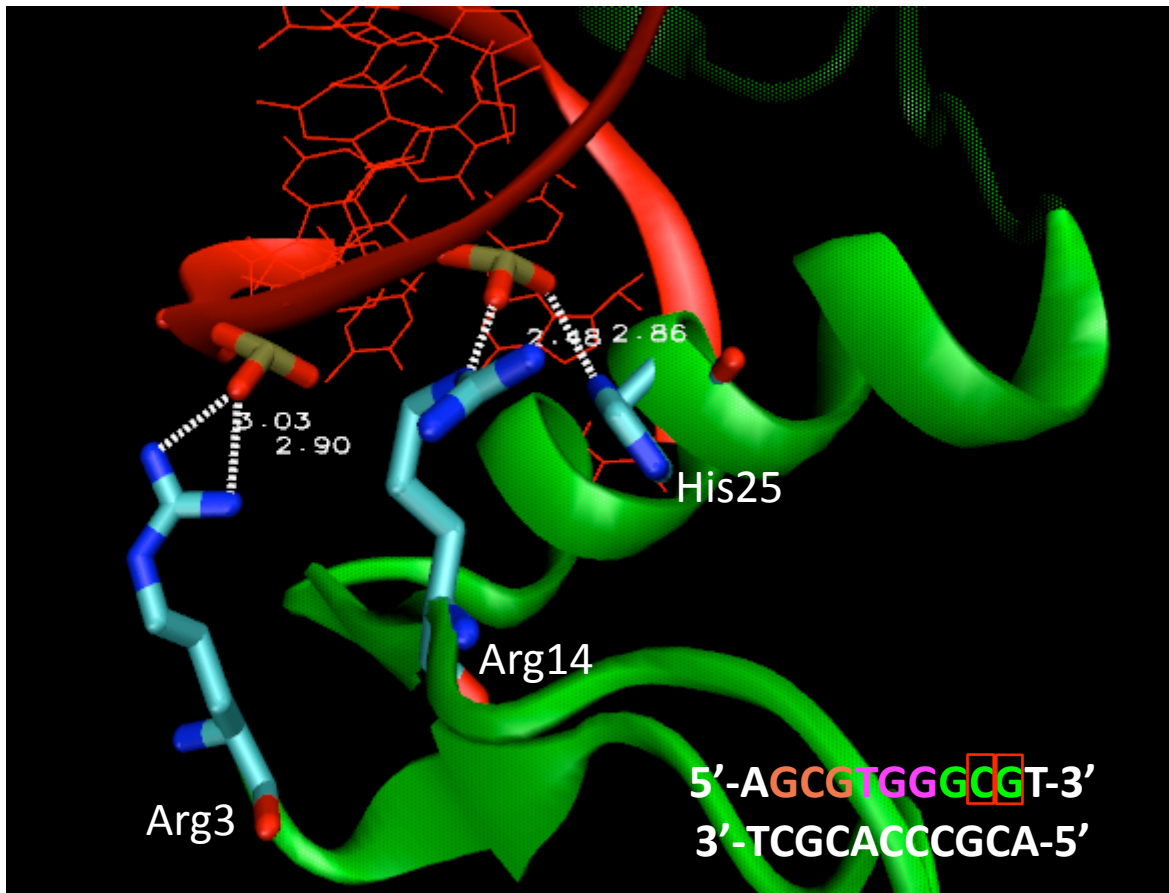
Hydrophobic core: highly conserved hydrophobic residues



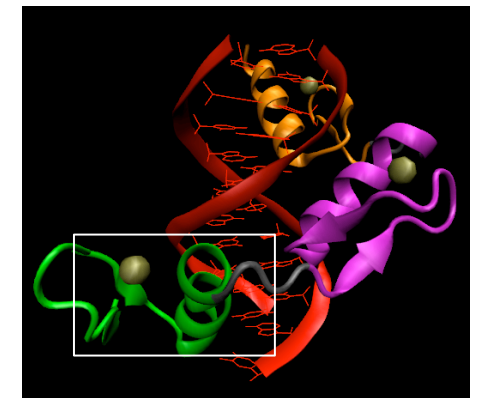
Cys2His2 zn finger (Zif268): phosphate contacts

H-bonds to the DNA backbone are the main inespecific contacts

	10	20	30	40	50	60	70	80										
Zif268	HERPYACPVESCDRRFSRDELTRHIRHTGQKPFQCRICMRNFSRSDHLTTHIRHTTGEKPFACDICGRKFARSDEKRRHTKIHRL																	
Zif268_phosphates	@-@																	

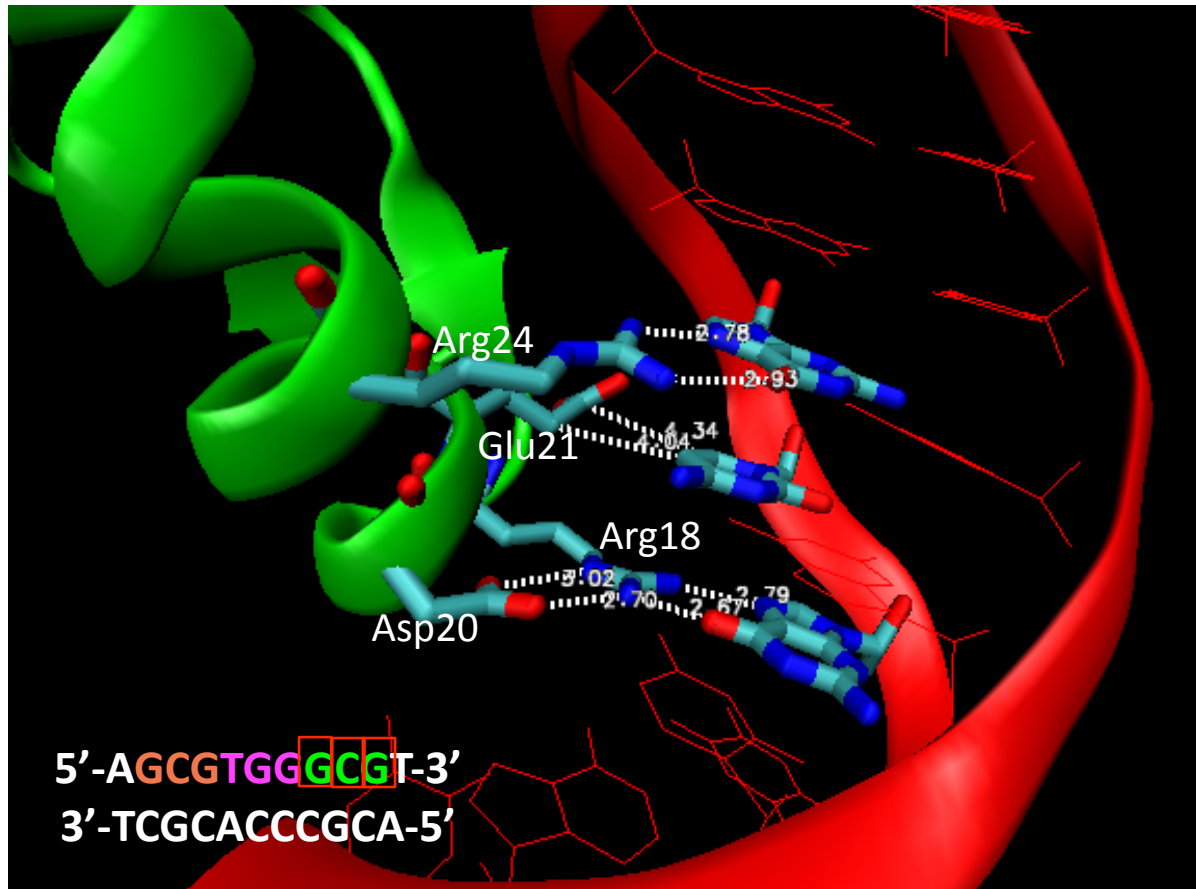
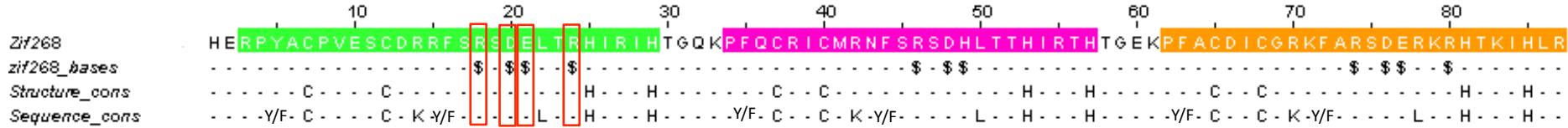


Zinc finger 1																									
d1zaac 1/1-25	Y	A	C	P	V	E	S	C	D	R	R	F	S	R	S	D	E	L	T	R	H	I	R	I	H
d2dlka 2/1-25	F	S	C	P	E	P	A	C	G	K	S	F	N	F	K	K	H	L	K	E	H	M	K	L	H
d1bhia 1/1-25	F	L	C	T	A	P	G	C	G	Q	R	F	T	N	E	D	H	L	A	V	H	K	K	H	H
d1ubdc 3/1-25	F	Q	C	T	F	E	G	C	G	K	R	F	S	L	D	F	N	L	R	T	H	V	R	I	H
d1bhba 2/1-25	Y	H	C	T	-	-	Y	C	N	F	S	F	K	T	K	G	N	L	T	K	H	M	K	A	H
d1bhba 1/1-22	Y	I	C	E	-	-	E	C	G	I	-	R	K	K	P	S	M	L	K	K	H	-	R	T	H
d2mda 2/1-24	H	K	C	E	-	-	V	C	G	K	C	F	S	R	K	D	K	L	K	T	H	-	R	C	H
d2csa 1/1-22	Y	P	-	C	-	-	Q	C	G	K	S	F	T	H	K	S	Q	R	D	R	H	-	S	M	H
d2ghfa 1/1-24	Y	V	C	V	-	-	E	C	N	F	L	T	K	R	Y	D	A	L	S	E	H	N	K	Y	H
d2dlqa 2/1-24	F	E	C	P	-	-	K	C	G	K	C	Y	F	R	K	E	N	L	L	E	H	E	R	N	C



Cys2His2 zn finger (Zif268): bases contacts

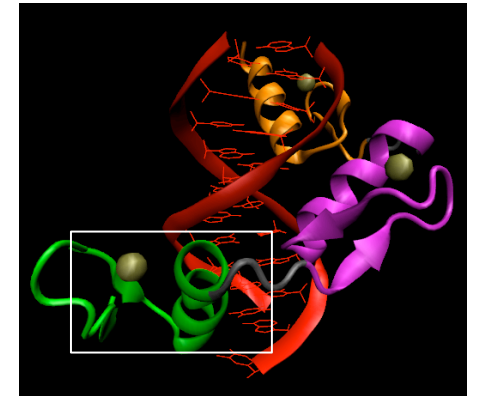
H-bonds between Zn fingers 1-3 and residues are the same but not for Zn finger 2 → different 3 bp binding site



Zinc finger 1

d1zaac 1/1-25	YAC	PVES	SCDRRFSRSD	ELTRH	IRIRH
d2dlka 2/1-25	FSC	PEPA	CGKSFNFK	KKHLKE	HMKLH
d1bhia_1/1-25	F	LCTAF	CGQRF	TNEDHLAV	HKHKK
d1ubdc3/1-25	FQ	CTFEG	CGKRFSLDF	NLNRTH	TVRIRH
d1hbca 2/1-25	YHCT	- - Y	CNFSFKT	KGNLTK	HMKAH
d1hbca 1/1-22	YICE	- - E	CGI	RKKPSMLKKH	- RTH
d2mda 2/1-24	HKCE	- - V	CGKCF	SRKDKLKT	H - RCH
d2sha 1/1-22	YP	- C - -	QCGKSFT	THKSQRDRH	- SMH
d2ghfa 1/1-24	YVCV	- - E	CNFLT	KRYDALSE	HNKYH
d2lqa 2/1-24	FEC	P - K	CGKCYFR	KENLLE	HERNC

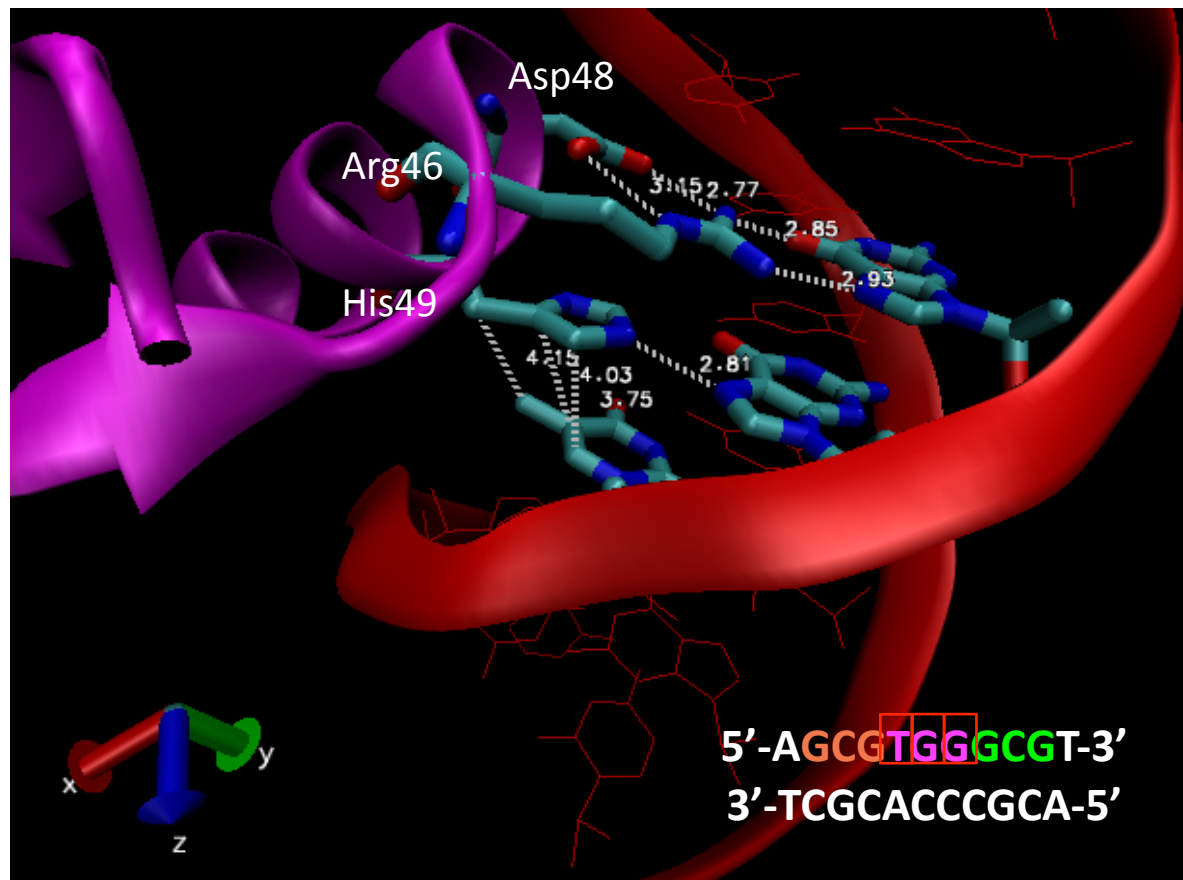
▲ ▲ ▲



Cys2His2 zn finger (Zif268): bases contacts

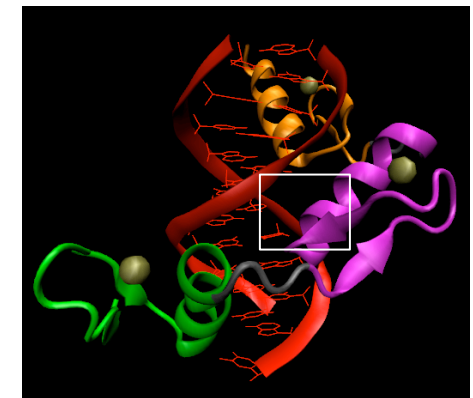
H-bonds between Zn fingers 1-3 and residues are the same but not for Zn finger 2 → different 3 bp binding site

	10	20	30	40	50	60	70	80		
Zif268	HERPYACP	VESCDRRFS	RSDLTRHIR	HTGQK	PFQCRICMRNFS	RSDH	LTTHIRTH	TGEK	PFACDICGRKFARS	DERKRHTKIHLR
zif268_bases		-\$-\$-\$	-\$		-\$-\$-\$				-\$-\$-\$	-\$
Structure_cons		C.....C	H.....H	C.....C		H.....H	C.....C			H.....H
Sequence_cons		-Y/F- C.....C	-K-Y/F-.....L.....H.....H	-Y/F- C.....C	-K-Y/F-.....L.....H.....H	-Y/F- C.....C	-K-Y/F-.....L.....H.....H			L.....H.....H

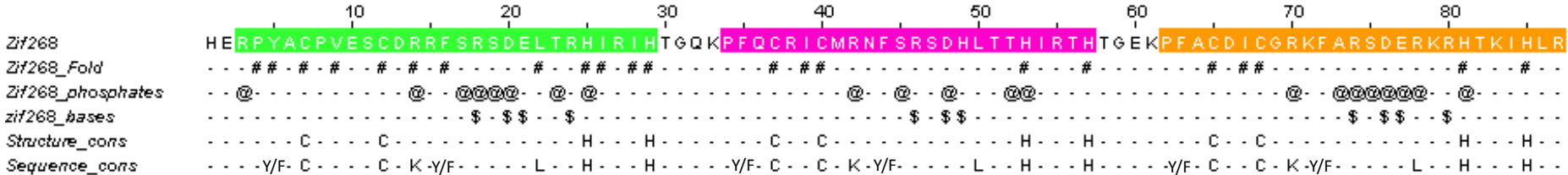


	Zinc finger 2																							
d1zaac 2/1-23	F	Q	C	R	-	I	C	M	R	N	F	S	R	S	D	H	L	T	T	H	I	R	T	H
d1zaac 3/1-23	F	A	C	D	-	I	C	G	R	K	F	A	R	S	D	E	R	K	R	H	T	K	I	H
d2mda 3/1-23	F	K	C	Q	-	I	C	P	Y	A	S	R	N	S	S	Q	L	T	V	H	L	R	S	H
d2cota 1/1-23	Y	K	C	D	-	E	C	G	K	A	F	I	Q	R	S	H	L	I	G	H	H	R	V	H
d2cota 2/1-23	Y	K	C	D	-	E	C	G	K	S	F	S	H	S	S	D	L	S	K	H	R	R	T	H
d1ubdc 2/1-25	H	V	C	A	-	E	C	G	K	A	F	V	E	S	S	K	L	K	R	H	-	L	V	H
d2dtga 1/1-26	Y	K	C	S	-	S	C	S	Q	Q	F	M	Q	K	K	D	L	Q	S	H	M	K	L	H

▲ ▲ ▲



Final summary



Structural align.
Consensus logo
(59 str)



Sequence align.
Consensus logo
(10663 seq)



The basic leucine zipper (bZIP) TFs

bZIP domain:

60-80 amino acids

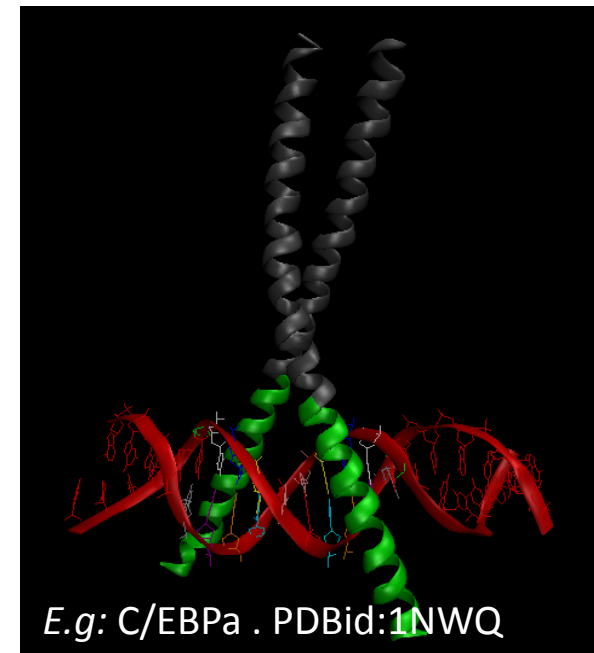
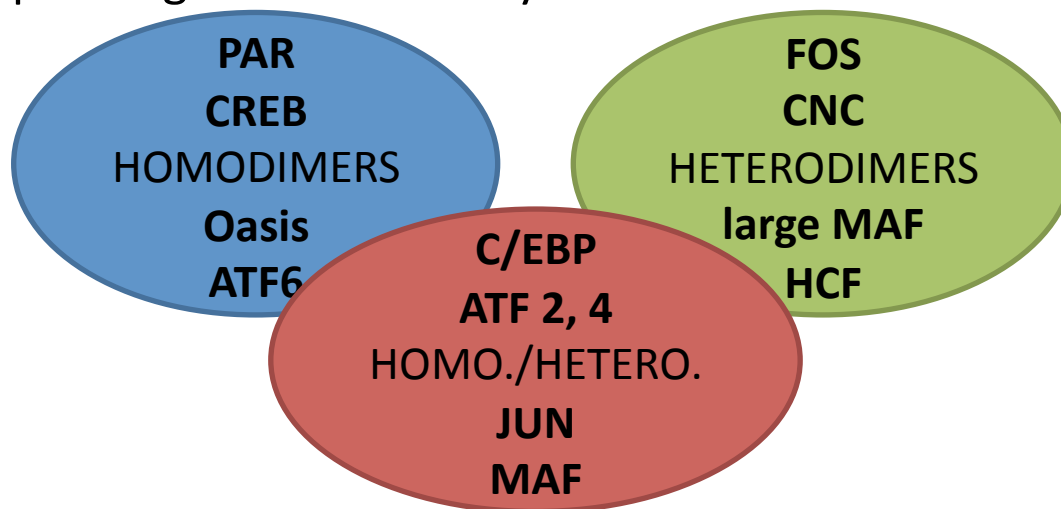
SCOP classification:

- Class: h: Coiled coil proteins
- Fold: h. 1: Parallel coiled-coil
- Superfamily: h.1.3: Leucine zipper domain
- Family: h.1.3.1: Leucine zipper domain

FUNCTIONS:

- Mitogenic stimulation
- Stress response
- Cytokine stimulation
- Development

TWO REGIONS: basic (N-ter) and leucine zipper (C-ter)
Different DNA binding sites and dimerization properties depending on its subfamily



5'-TATTGCGCAATA-3'
3'-ATAACGCGTTAT-5'

The basic leucine zipper (bZIP) TFs



- 13 entries bound to DNA



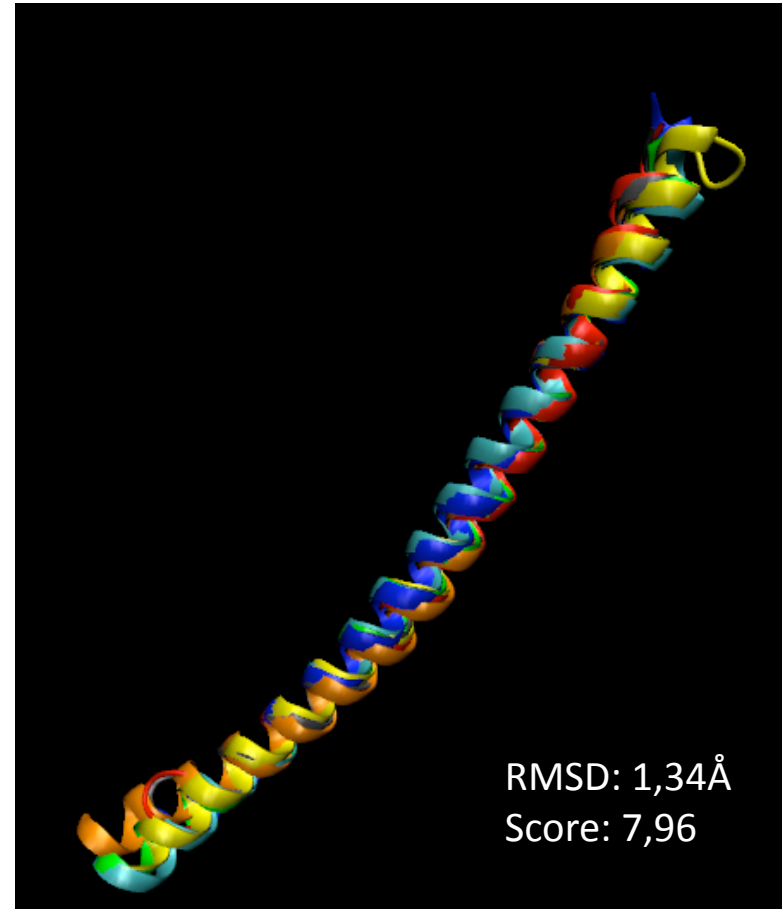
Structures superimposed: 7
Discarded → DNA binding domain:
bHLH or *Zn-finger*



- 277 hmmer hits (<0.0001)



Sequences aligned: 273
Discarded: those presenting
uniquely homology with the
leucine zipper region)



Superimposition of the crystalized bZIP domains:
C/EBPa, C/EBPb, C-Fos, C-Jun, CREB, cAMP ATF2
and GCN4.

C/EBPa: dimerisation contacts/fold

Amphipathic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two π - π' interactions and an interhelical hydrogen bond (N-H)

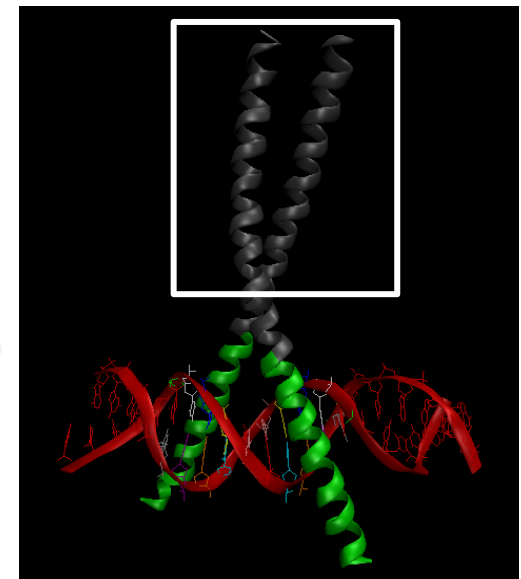
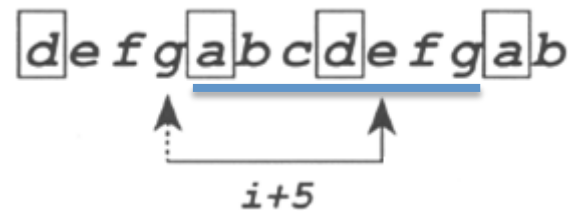
[illegible]

LEUCINE ZIPPER

312 322 332

RNVETQOKVLELTSDNDRLRKRVQQLSRELDTL - RG ---
RMKQLEDKVEELLSKNYHLENEVARLKKLVGER - - - - -
YVKSLENRVAVLENQNKTLIEELKALKDLYSHK - - - - -
RIARLEEKVKTLLKAQNSELASTANMLREQVAQLK - QKVM
LNLETOHKVLELTAENERLQKKVEQLSRELSTL - RNLFK
LTDTLQAEQTDQLEDEKSALQTEIANLLKEKEKLE - FILA
WVQSLLEKKAEDLSSLNGQLQSEVTLRLNEVAQLKQLLLA

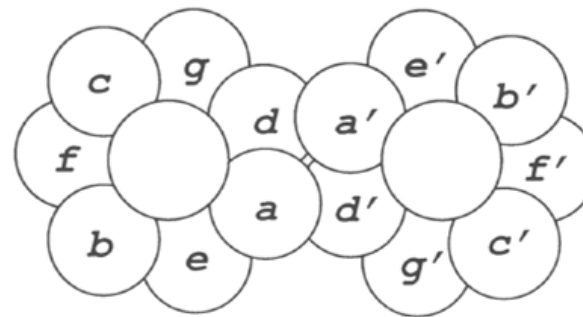
▲ ▲ ▲ ▲ ▲ ▲ ▲ ▲



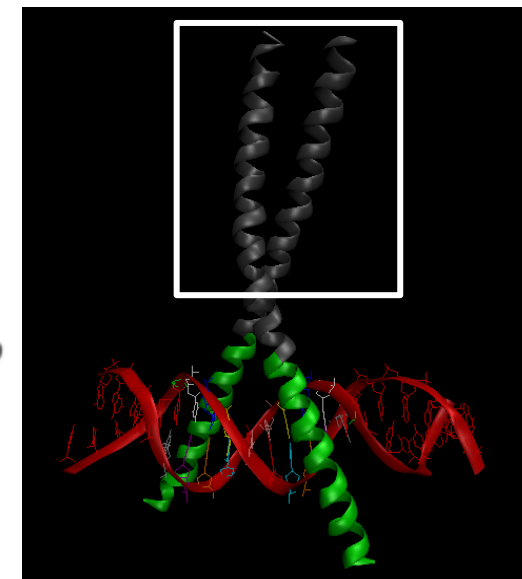
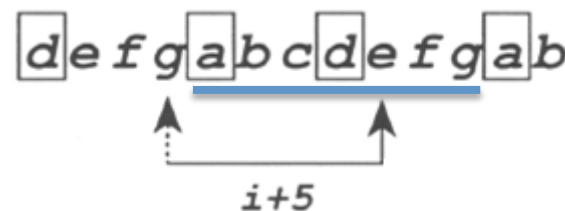
C/EBPa: dimerisation contacts/fold

Amphypatic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two g-e' interactions and a interhelical hydrogen bond (N-N)

	283	288	293	298	303	308	313	318	323	328	333	338
<i>C/EBPa_rat</i>	GSNSNEYRVRERNNIAVRKSRDKAKQRNVETOQKVLELTSDNDRRLRKRVEQLSRELDTLRG											
<i>C/EBPa_fold</i>	-----#-----#-----##-----#-----#-----#-----##-----#-----##-----											
<i>Structure_cons</i>	-----RN--AA-KSR-K-----E-K--L--N--L--V--L-----											
<i>Sequence_cons</i>	-----NR-A-----R-R-----LE-----L-----											



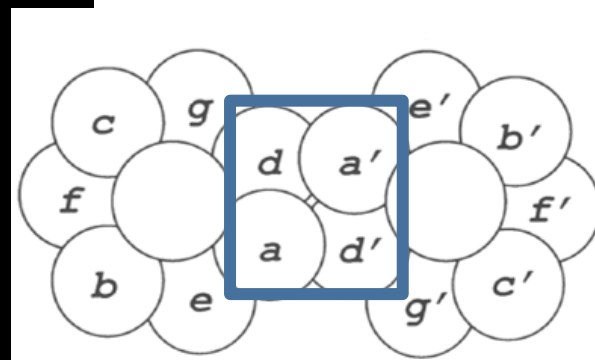
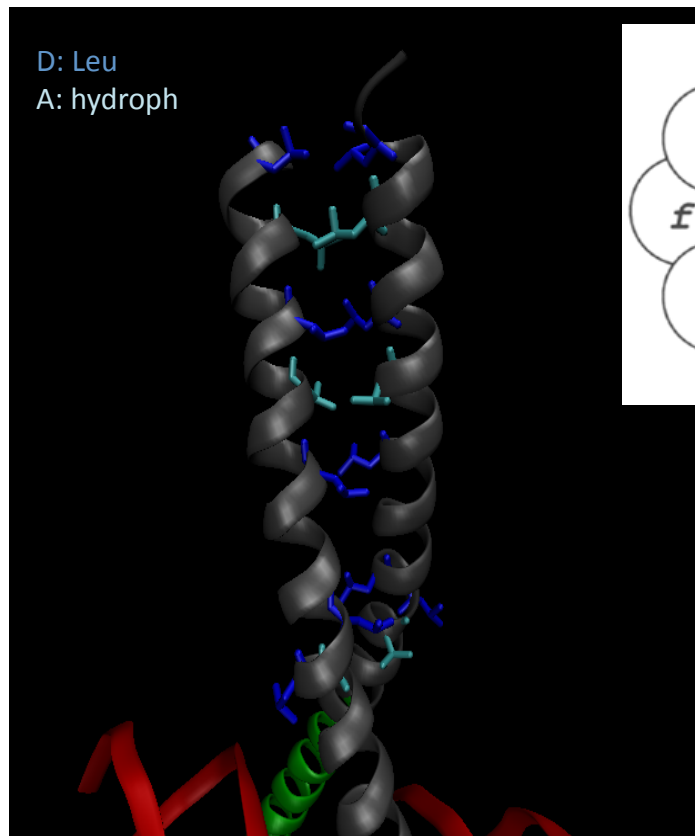
LEUCINE ZIPPER											
	312		322		332						
NVETOQKVLELTSDNDRRLRKRVEQLSRELDTL-RG---											
MKQLEDKVEELLSKNYHLENEVARLKKLVGER-----											
VKSLENRVAVLENQNKTLIEELKALKDLYSHK-----											
IARLEEKVKTLKAQNSELASTANMLREQVAQLK-QKVM											
NLETQHKVLELTAENERLQKKVEQLSRELSTL-RNLFK											
TDTLQAEITDQLEDEKSALQTEIANLLKEKEKLE-FILA											
VQSLEKKAEDLSSLNGQLQSEVTLLRNEVAQLKQLLLA											



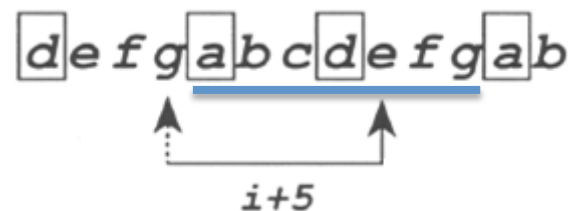
C/EBPa: dimerisation contacts/fold

Amphypatic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two g-e' interactions and a interhelical hydrogen bond (N-N)

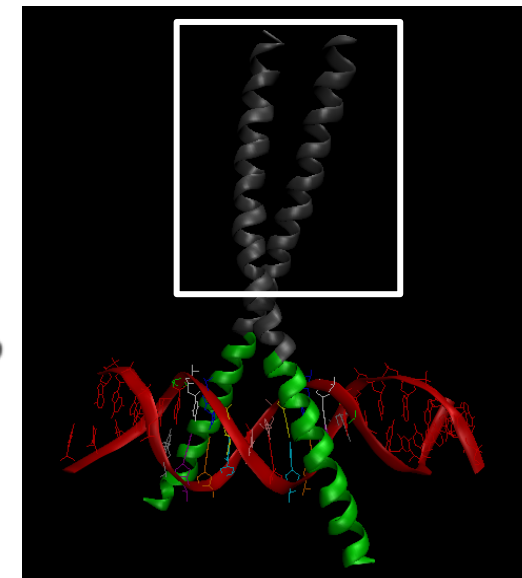
	283	288	293	298	303	308	313	318	323	328	333	338
<i>C/EBPa_rat</i>	GSNSNEYRVRRERNNAVRKSRDKAKQRNVETOQKVLELTSDNDRLRKRVEQLSRELDTLRG											
<i>C/EBPa_fold</i>	-----#-----#-----##-----#-----#-----#-----#-----#-----##-----											
<i>Structure_cons</i>	-----RN--AA-KSR-K-----E-K--L--N--L--V--L-----											
<i>Sequence_cons</i>	-----NR-A-----R-R-----LE-----L-----											



hydrophobic core



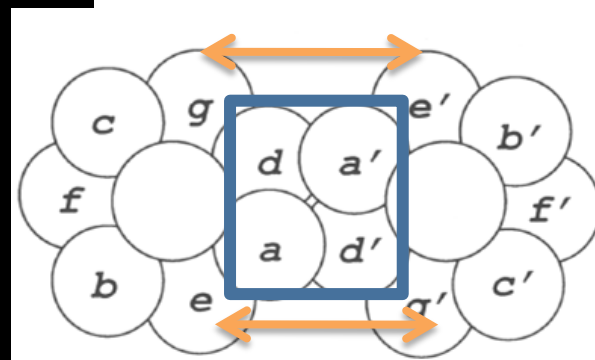
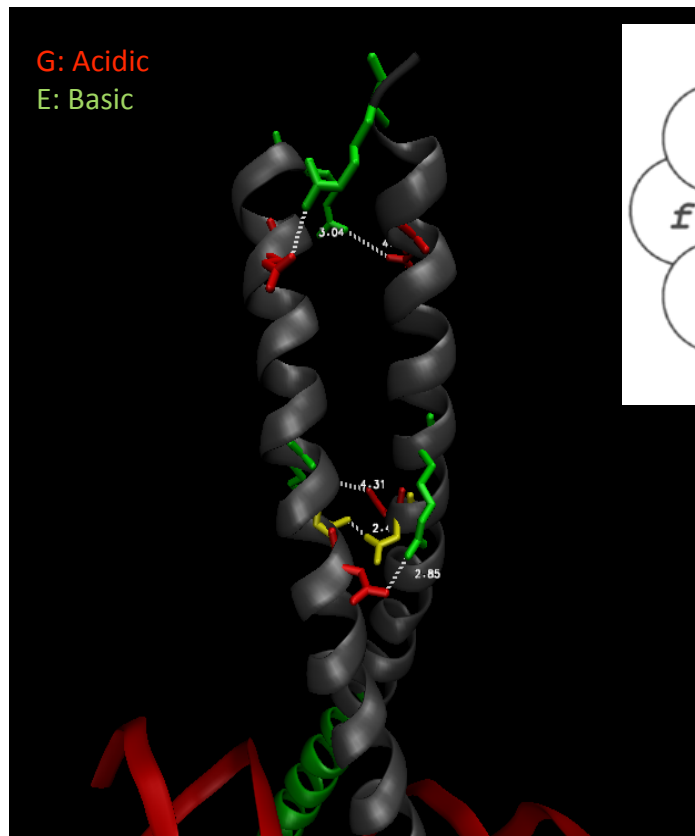
LEUCINE ZIPPER											
	312		322		332						
NVETQOKVLELTSDNDRLRKRVEQLSRELDTL-RG---											
MKQLEDKVEELLSKNYHLENEVARLKKLVGER-----											
VKSLENRVAVLENQNKTLIEELKALKDLYSHK-----											
IARLEEKVKTLKAQNSELASTANMLREQVAQLK-QKVM											
NLETQHKVLELTAENERLQKKVEQLSRELSTL-RNLFK											
TDTLQAEETDQLEDEKSALQTEIANLLKEKEKLE-FILA											
VQSLEKKAEDLSSLNGQLQSEVTLLRNEVAQLKQLLLA											



C/EBPa: dimerisation contacts/fold

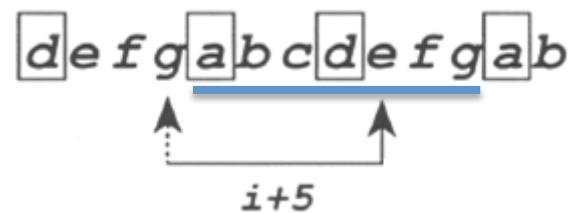
Amphypatic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two g-e' interactions and a interhelical hydrogen bond (N-N)

	283	288	293	298	303	308	313	318	323	328	333	338
<i>C/EBPa_rat</i>	GSNSNEYRVRRERNNAVRKSRDKAKQRNVETQOKVLELTSDNDRLRKRVVEQLSRELDTLRG											
<i>C/EBPa_fold</i>	-----#-----#-----##-----#-----#-----#-----##-----##-----											
<i>Structure_cons</i>	-----RN--AA-KSR-K-----E-K--L--N--L--V--L-----											
<i>Sequence_cons</i>	-----NR-A-----R-R-----LE-----L-----											

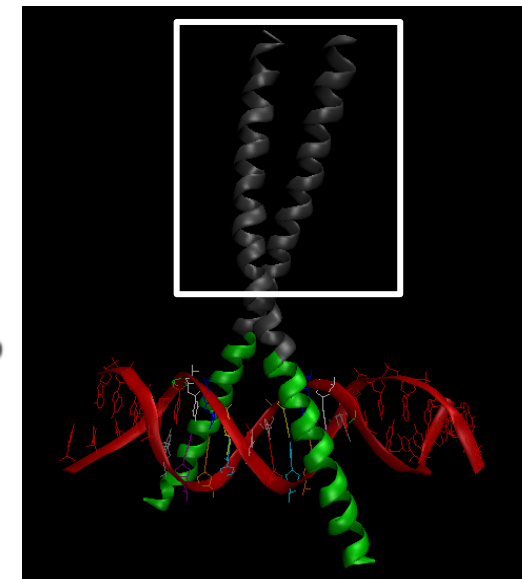


hydrophobic core

electrostatic interactions



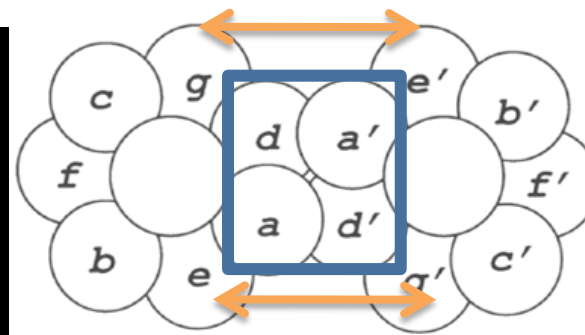
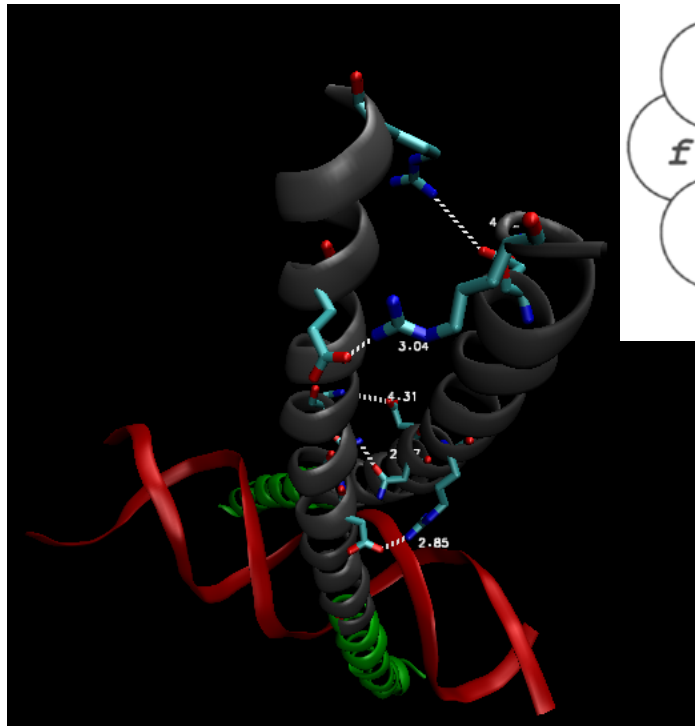
LEUCINE ZIPPER											
	312		322		332						
	NVETQOKVLELTSDNDRLRKRVVEQLSRELDTL-RG---										
	MKQLEDKVEELLSKNYHLENEVARLKKLVGER-----										
	VKSLENRVAVLENQNKTLIEELKALKDLYSHK-----										
	IARLEEKVKTLKAQNSELASTANMLREQVAQLK-QKVM										
	NLETQHKVLELTAENERLQKKVEQLSRELSTL-RNFLFK										
	TDTLQAEITDQLEDEKSALQTEIANLLKEKEKLE-FILA										
	VQSLEKKAEIDLSSLNGQLQSEVTLLRNEVAQLKQLLLA										



C/EBPa: dimerisation contacts/fold

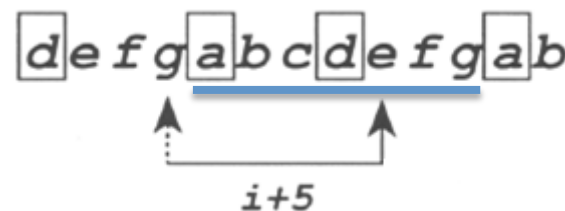
Amphypatic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two g-e' interactions and a interhelical hydrogen bond (N-N)

	283	288	293	298	303	308	313	318	323	328	333	338
<i>C/EBPa_rat</i>	GSNSNEYRVRRERNNAVRKSRDKAKQNVETQOKVLELTSDNDRLRKRVQEQLSRELDTLRG											
<i>C/EBPa_fold</i>	-	-	-	-	-	-	-	-	-	-	-	-
<i>Structure_cons</i>	-	-	-	-	-	-	-	-	-	-	-	-
<i>Sequence_cons</i>	-	-	-	-	-	-	-	-	-	-	-	-

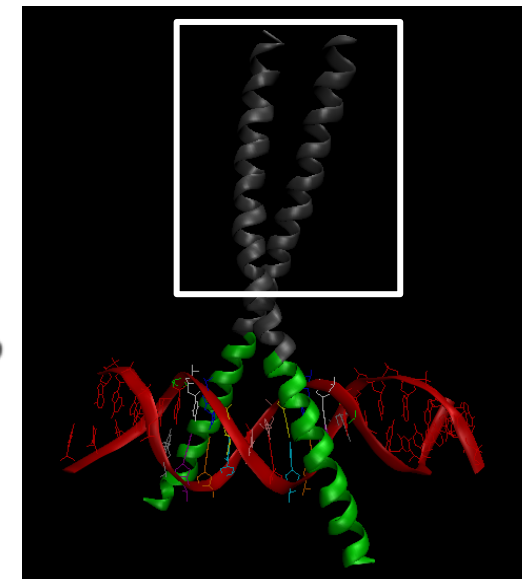


hydrophobic core

electrostatic interactions

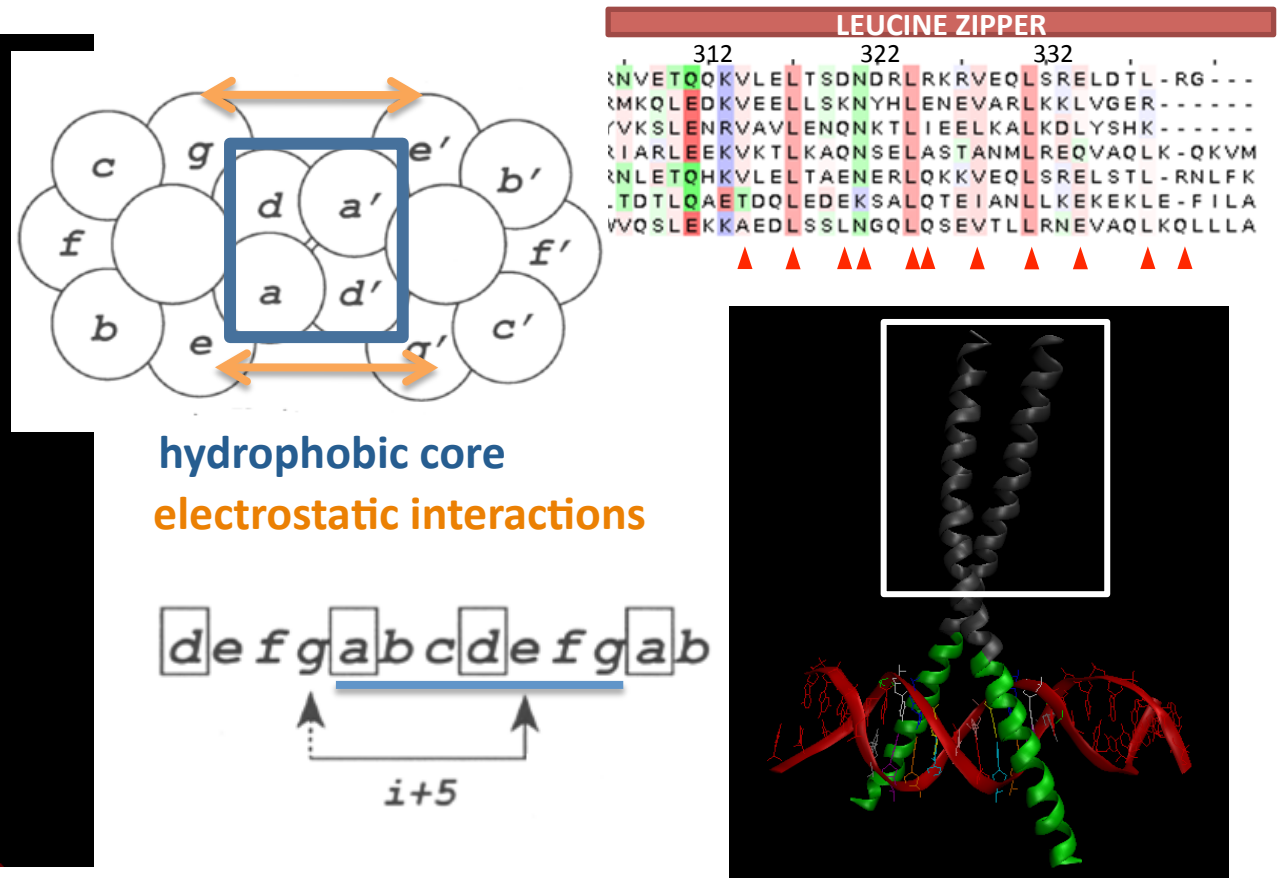
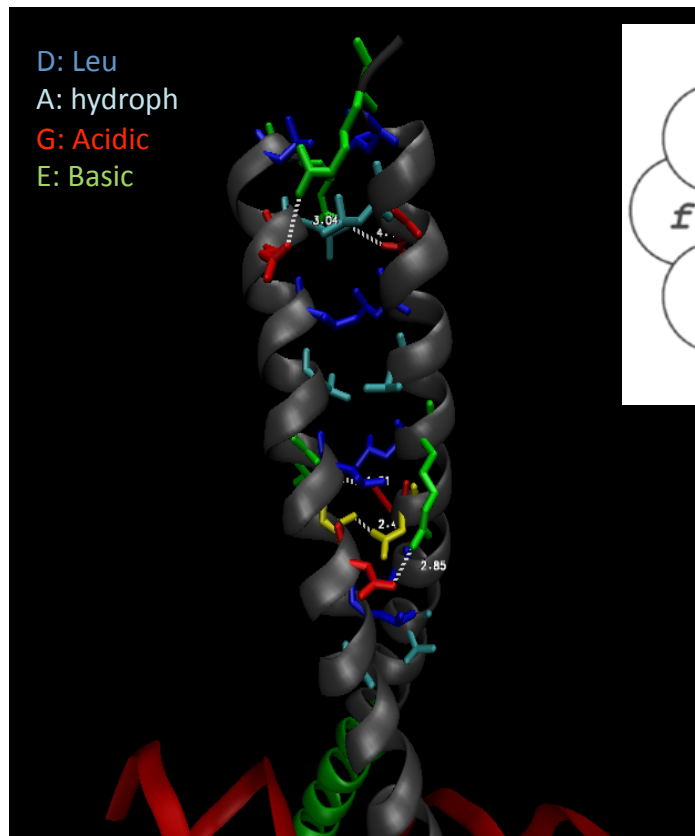
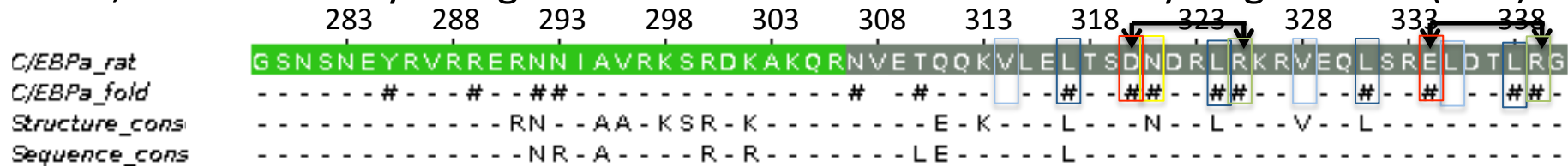


LEUCINE ZIPPER											
	312		322		332						
NVETQOKVLELTSDNDRLRKRVQEQLSRELDTL-RG---											
MKQLEDKVEELL SKNYHLENEVARLKKLVGER-----											
VKSLENRVAVLENQNKTLIEELKALKDLYSHK-----											
IARLEEKVKTLKAQNSELASTANMLREQVAQLK-QKVM											
NLETQHKVLELTAENERLQKKVEQLSRELSTL-RNLFK											
TDTLQAE T DQLEDEKSALQTEIANLLKEKEKLE-FILA											
VQSLEKKAEDLSSLNGQLQSEVTLRNEVAQLKQLLLA											



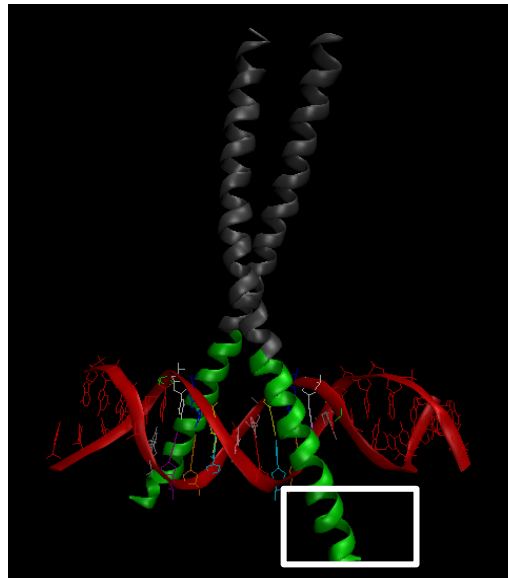
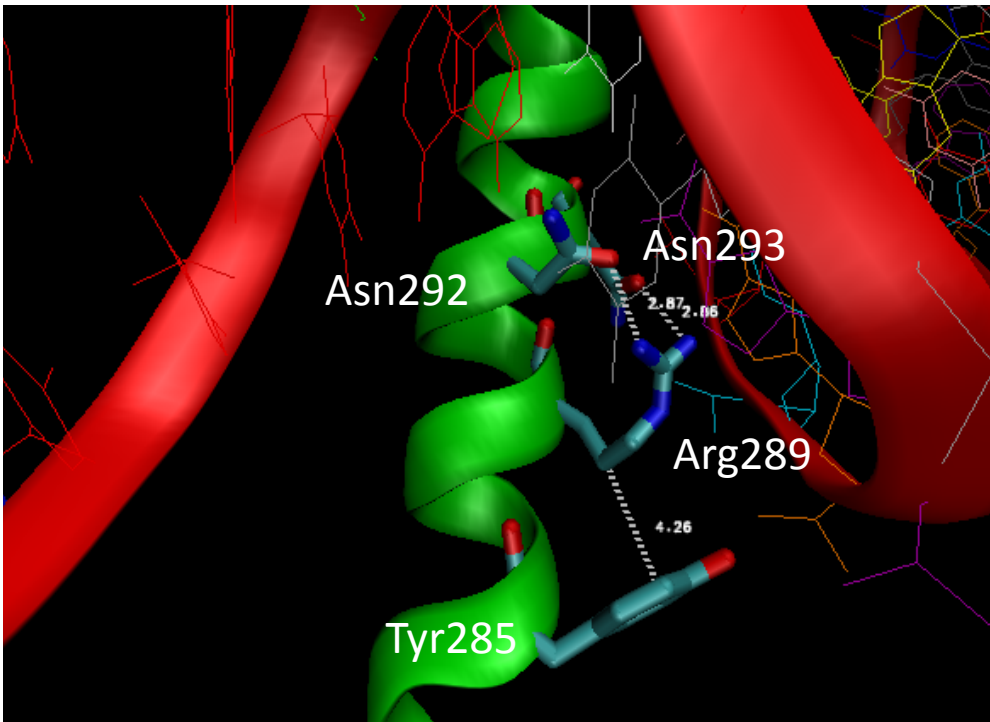
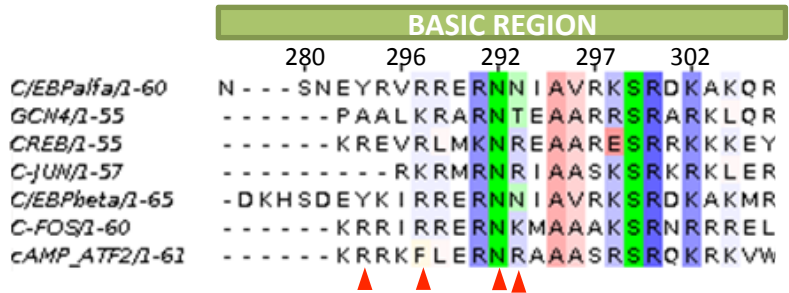
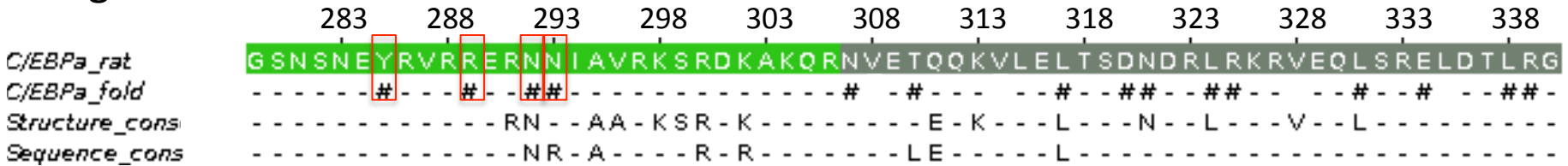
C/EBPa: dimerisation contacts/fold

Amphypatic helices forming the coiled coil. Leucines and hydrophobic residues in the core, surrounded by two g-e' interactions and a interhelical hydrogen bond (N-N)



C/EBPa: dimerisation contacts/fold

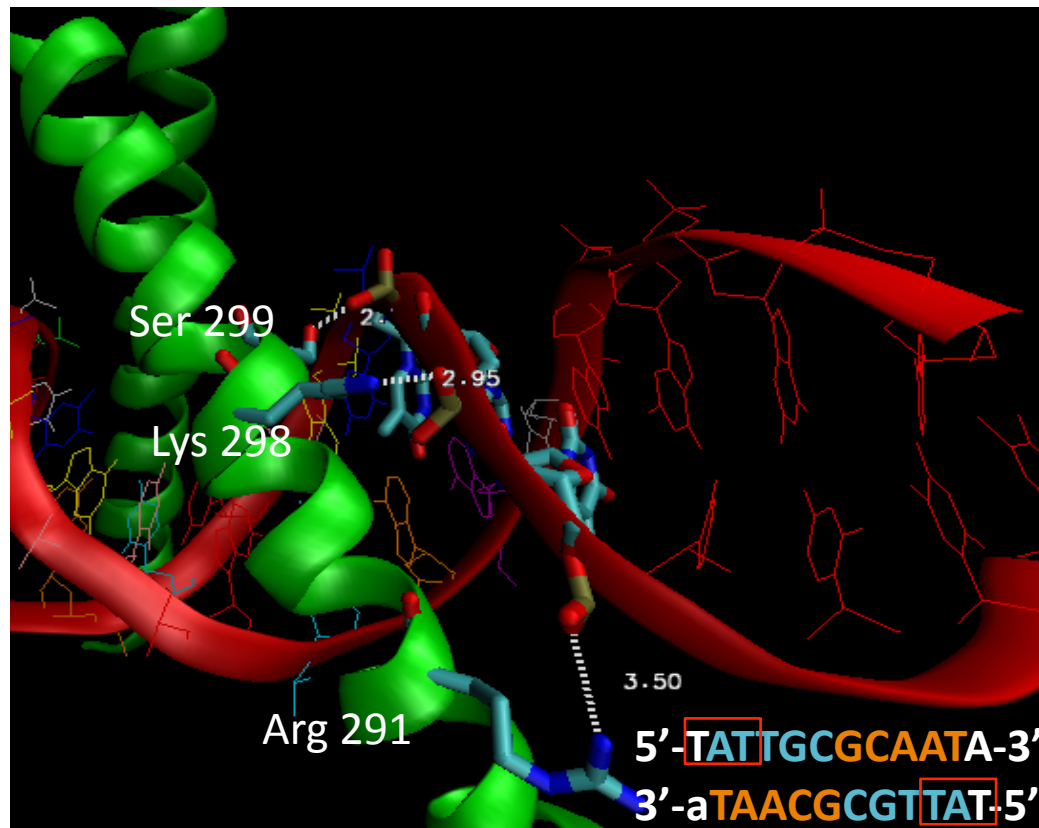
The protein-DNA interface is stabilized by interactions between side chains of basic region aminoacids



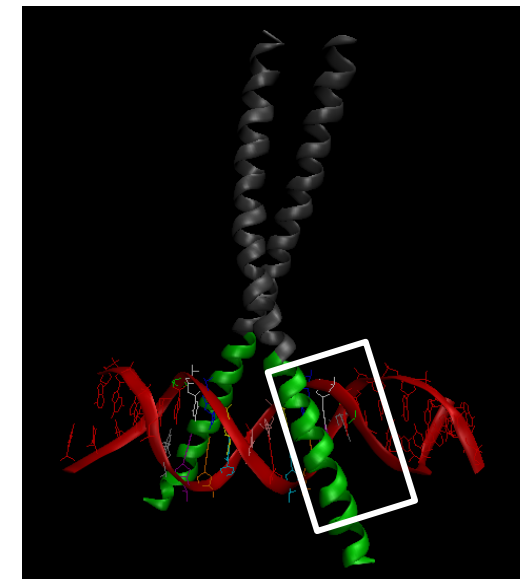
C/EBPa: phosphate contacts

Most non-specific contacts are hydrogen bonds between phosphates and either basic or polar aminoacids.

	283	288	293	298	303	308	313	318	323	328	333	338
<i>C/EBPa_rat</i>	GSNSNEYRVRRE RNN IAVRKS RDKAKQRNVETQQKVLELTSDNDR LRKRVEQLSRELD TLRG											
<i>C/EBPa_phosphates</i>	-----@-@-----@-@-----@-@-----											
<i>Structure_cons</i>	-----RN--AA-KSR-K-----E-K--L--N--L--V--L-----											
<i>Sequence_cons</i>	-----NR-A---R-R-----LE-----L-----											

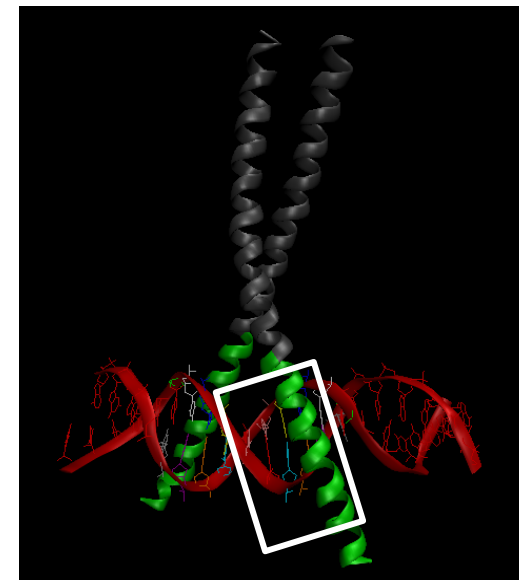
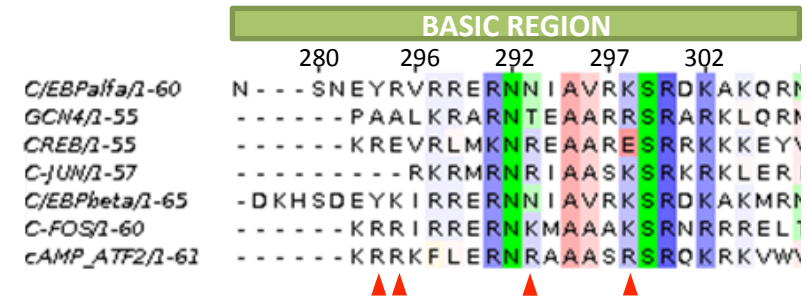
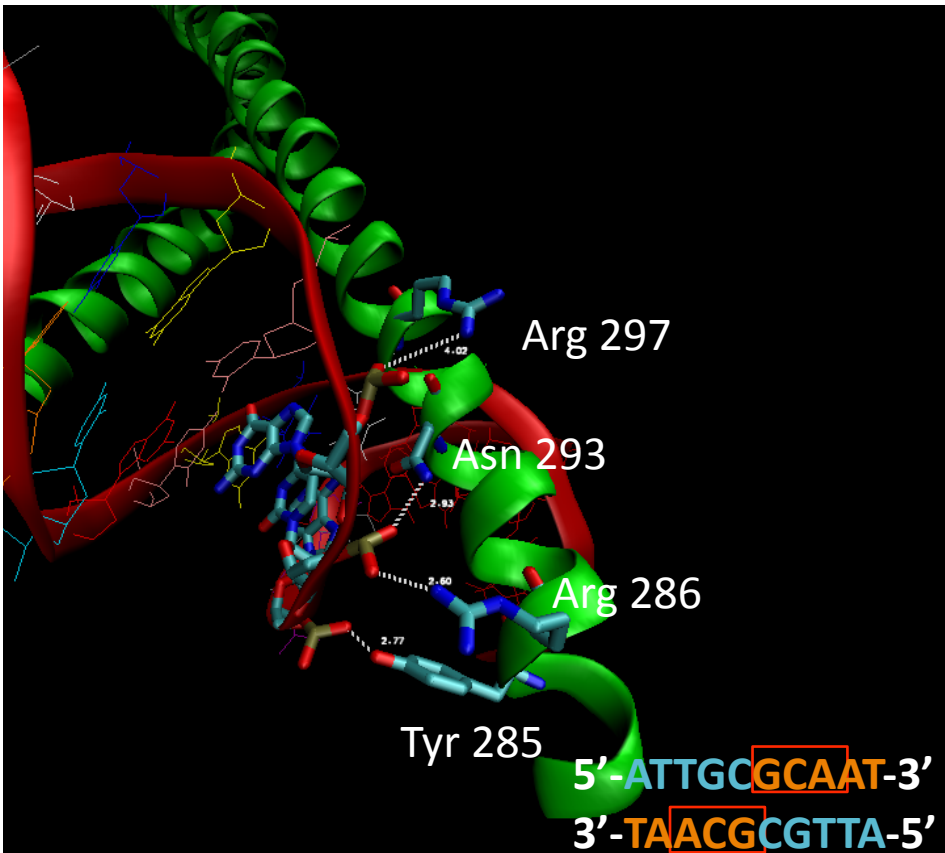
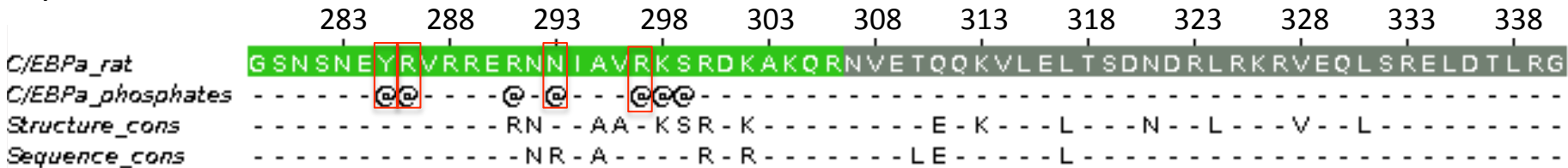


	BASIC REGION				
	280	296	292	297	302
<i>C/EBPa/1-60</i>	N---	SNEYRVRRE	RNN IAVRKS	RDKAKQR	
<i>GCN4/1-55</i>	-----	PAALKRARNT	EAARRS	RARKLQR	
<i>CREB/1-55</i>	-----	KREVRLMKN	REAARE	SRRKKKEY	
<i>C-JUN/1-57</i>	-----	RKRMRNR	IAASK	SRKKLER	
<i>C/EBPbeta/1-65</i>	-DKHSDEYK	IRRE RNN IAVRKS	RDKAKMR		
<i>C-FOS/1-60</i>	-----	KRRIRRE	RNMMAAK	SRNRREL	
<i>cAMP_ATF2/1-61</i>	-----	KRRKFLER	NRAAAS	SRQKRKVW	



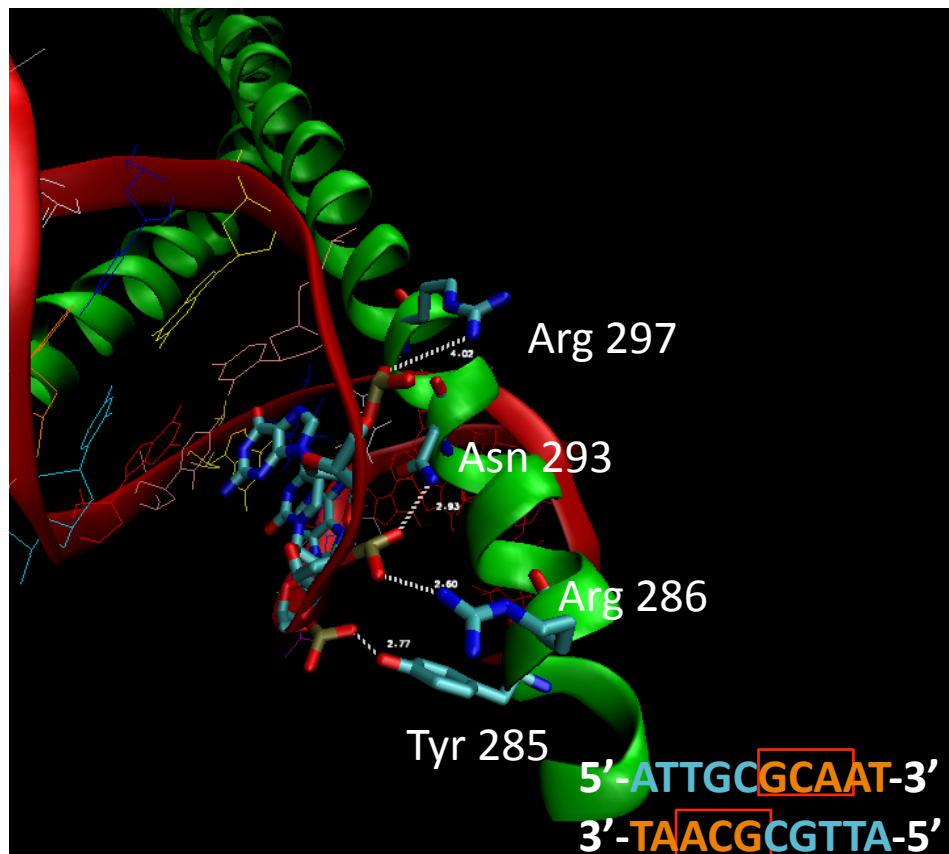
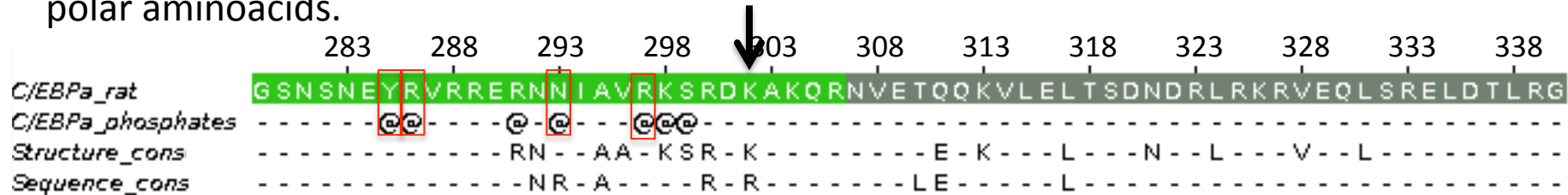
C/EBPa: phosphate contacts

Most non-specific contacts are hydrogen bonds between phosphates and either basic or polar aminoacids.

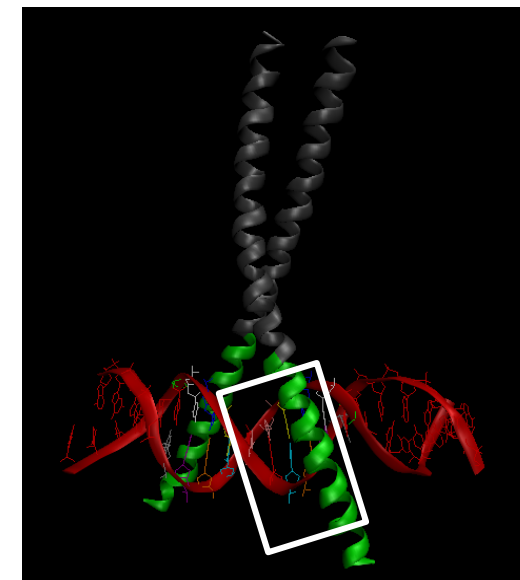


C/EBPa: phosphate contacts

Most non-specific contacts are hydrogen bonds between phosphates and either basic or polar aminoacids.

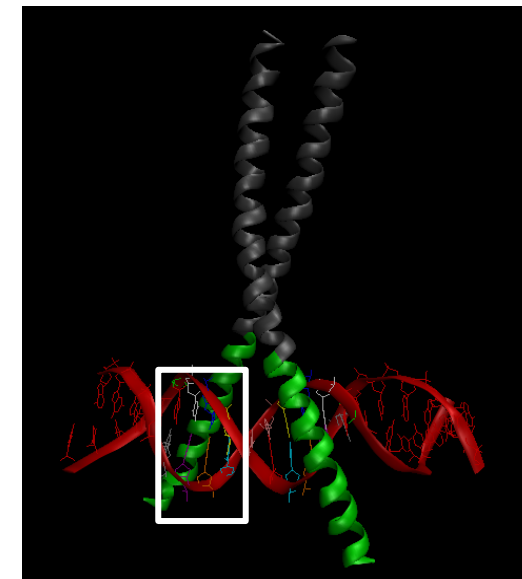
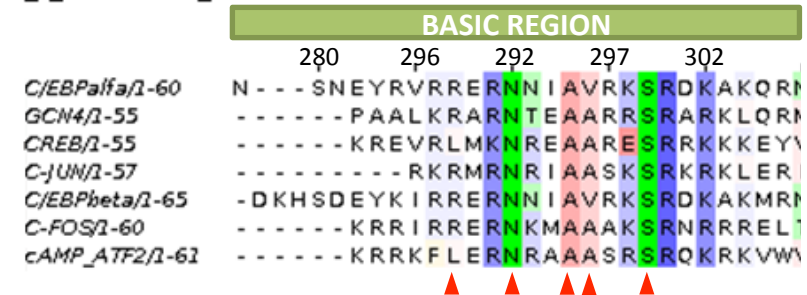
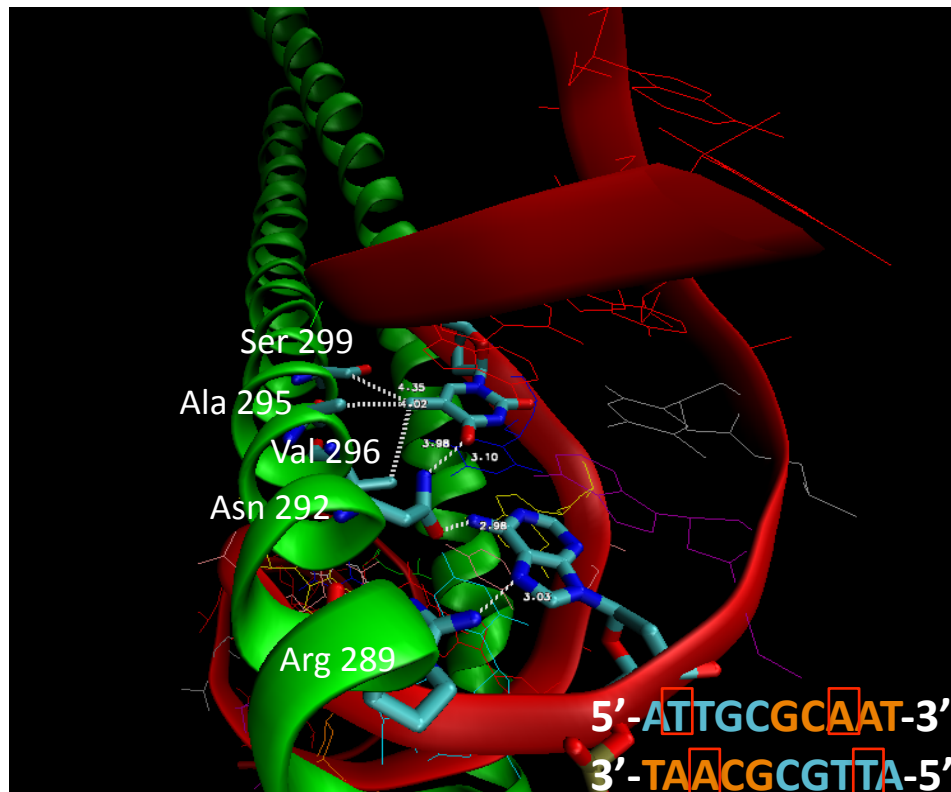
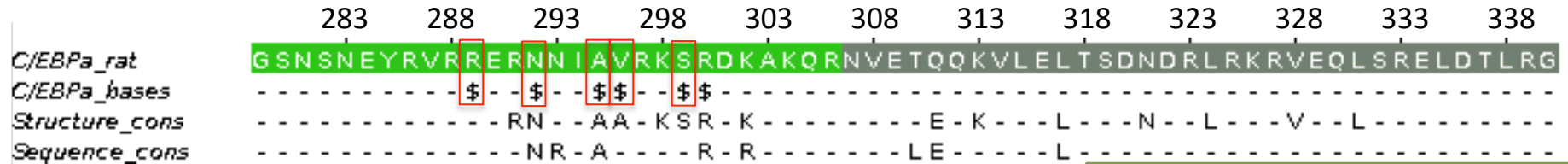


	BASIC REGION					
	280	296	292	297	302	
<i>C/EBPa</i> 1-60	N---	SNEYRVRERNNIAVRKSRDKAKQR				
<i>GCN4</i> 1-55	-----	PAALKRARNTAAARRSRARKLQR				
<i>CREB</i> 1-55	-----	KREVRMLKNREAAARESRKKKEY				
<i>C-JUN</i> 1-57	-----	RKRMRNRRIAASKSRKKLER				
<i>C/EBP</i> eta1-65	-DKHSDEYKIRREERNIAVRKSRDKAKMR					
<i>C-FOS</i> 1-60	-----	KRRIRREERNKMAAAKSRNRREL				
<i>cAMP_ATF2</i> 1-61	-----	KRRKFLERNRAAASRSRQKRKVW				



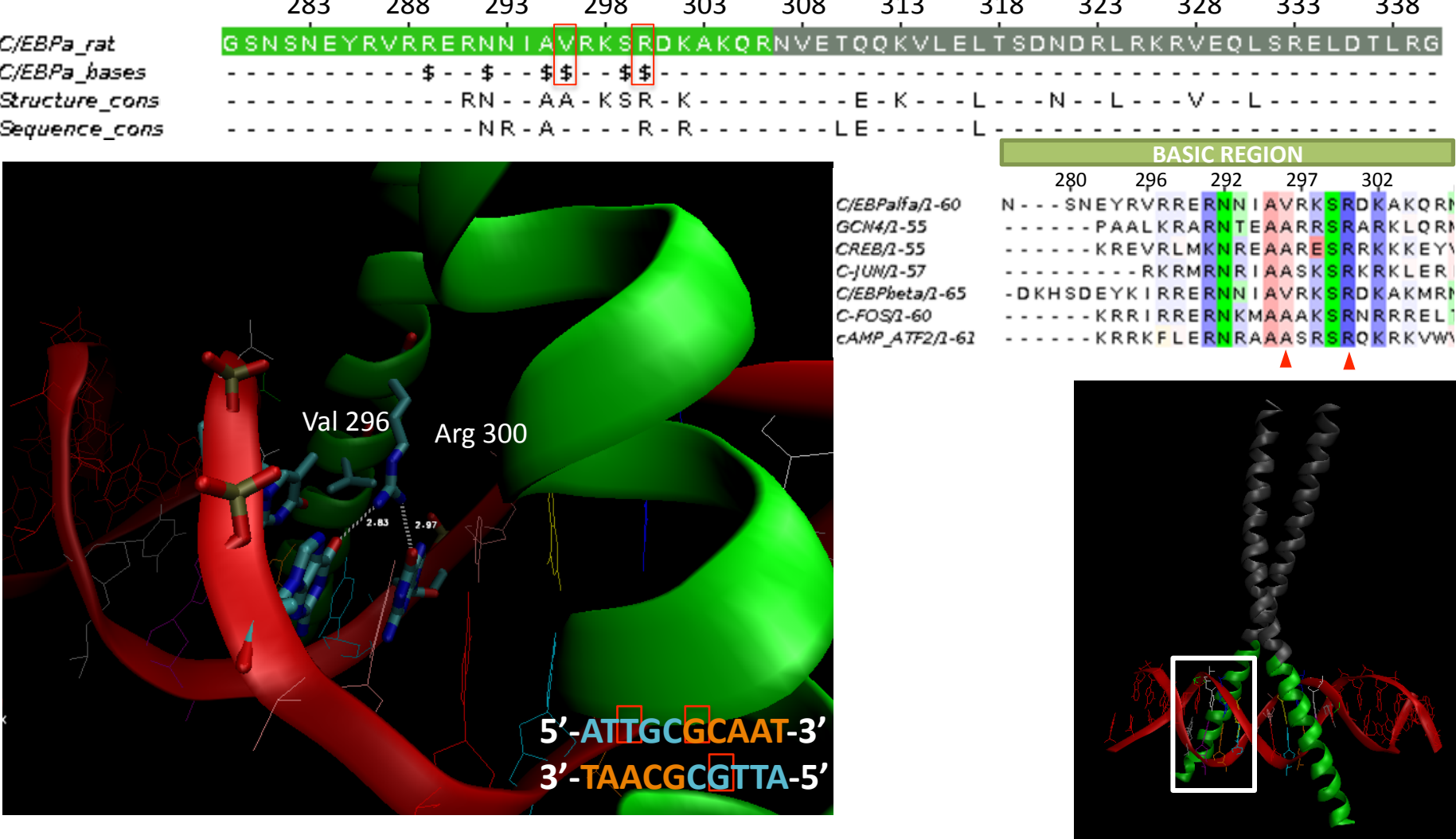
C/EBPa: base contacts

The invariant Asn292 is involved in two specific base contacts, while Ser299, Ala295 and Val296 make VDW contacts with T-4.



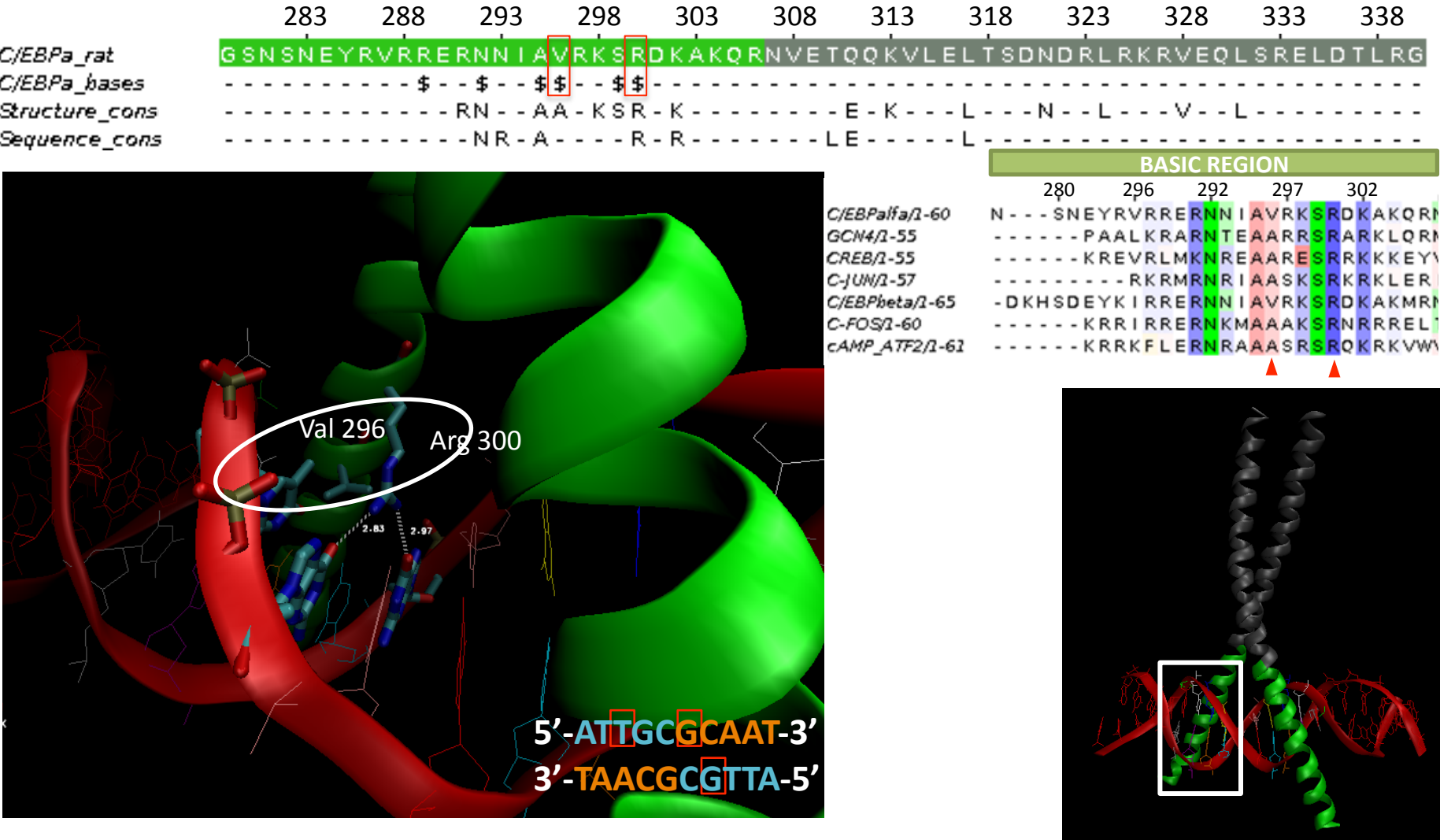
C/EBPa: base contacts

Val 296 gives DNA binding specificity for the C/EBP subfamily.

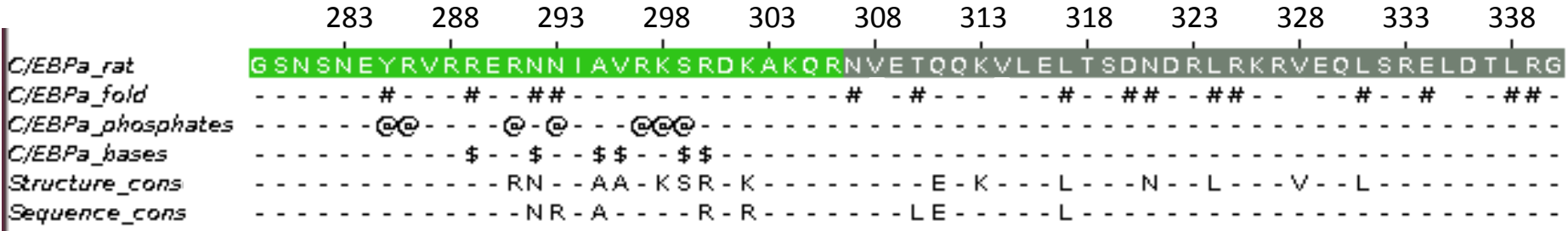


C/EBPa: base contacts

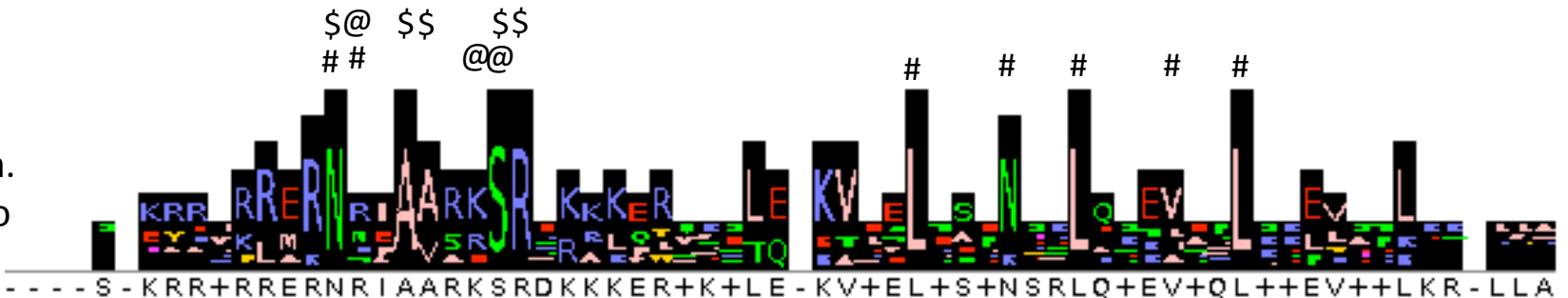
Val 296 gives DNA binding specificity for the C/EBP subfamily.



bZIP: final summary



Structural align.
Consensus logo
(7 str)



Sequence align.
Consensus logo
(273 seq)



The basic-helix–loop-helix (bHLH) TFs

bHLH domain:

60 amino acids

SCOP classification:

- Class a: all alpha
- Fold: a.38: HLH-like
- Superfamily: a.38.1: HLH, helix-loop-helix DNA-binding domain
- Family: a.38.1.1: HLH, helix-loop-helix DNA-binding domain

Dimerisation: Homo/heterodimers

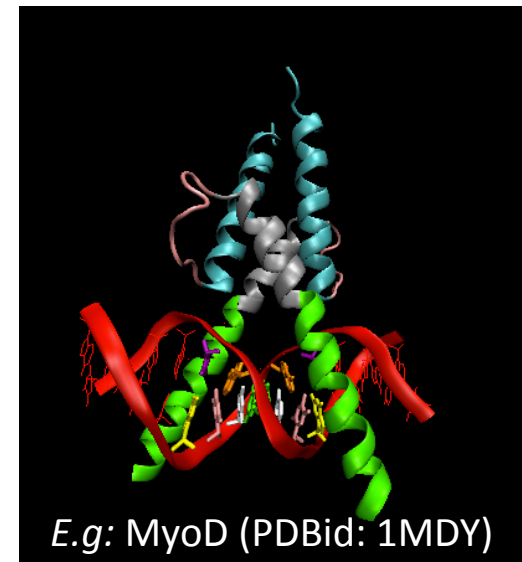
“Classical” binding site: E-box: 5’-CANNTG-3’

TFs with bHLH domain may have **other domains**:

e.g. Zip: dimerization

Functions (mammals):

- Development: MyoD, NeuroD
- Cell cycle (cancer): Myc
- Other: circadian clock control: Clock, BMAL1



The basic-helix–loop-helix (bHLH) TFs



- 11 entries bound to DNA



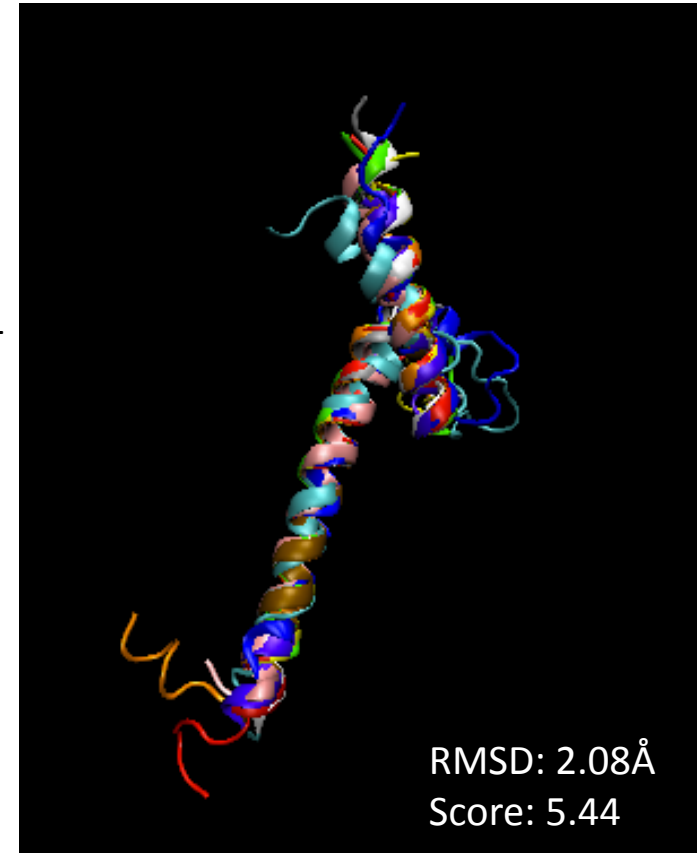
Structures superimposed (non-redundant): 11



- PFAM: HLH (PF00010)
- Hmsearch hits: 645 (E-value <0.0001)

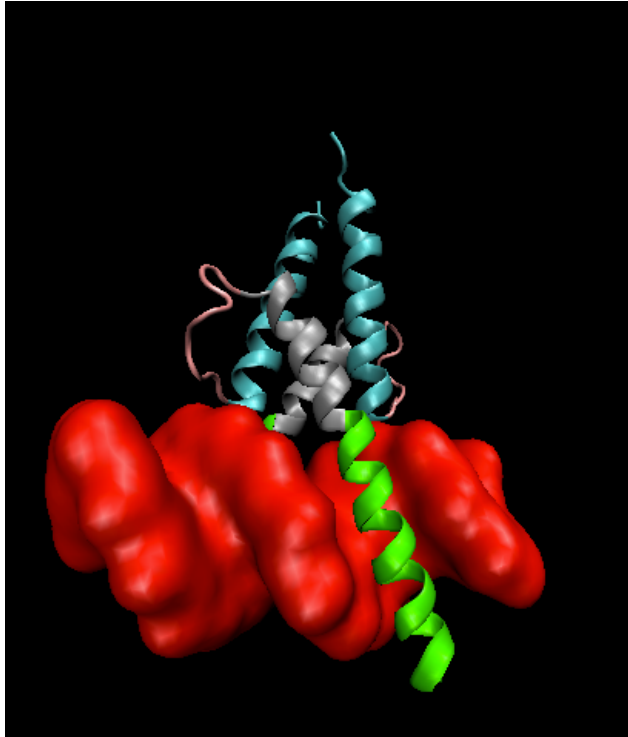


Sequences aligned: 619
(ID family, which lacks basic region
and doesn't bind DNA,
was discarded).

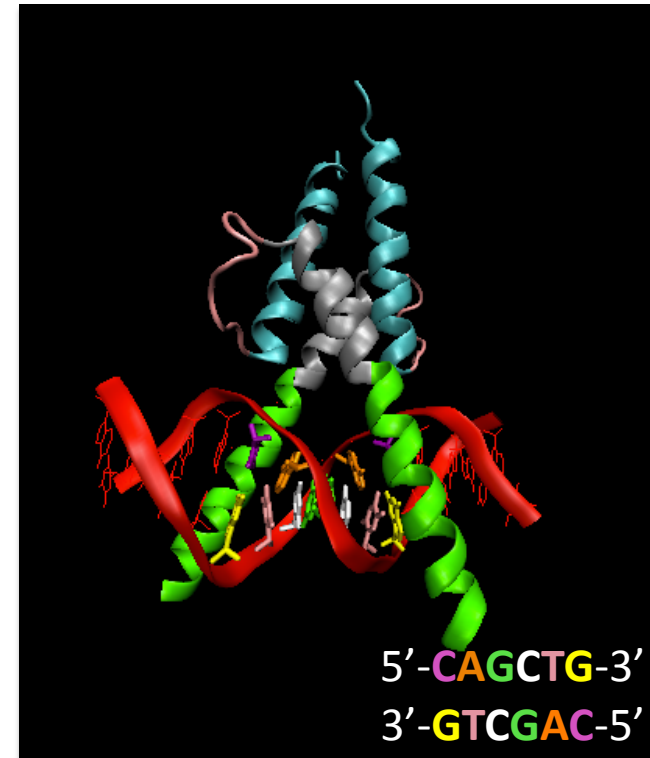


Superimposition of the crystalized bHLH domains

MyoD (bHLH domain): structure



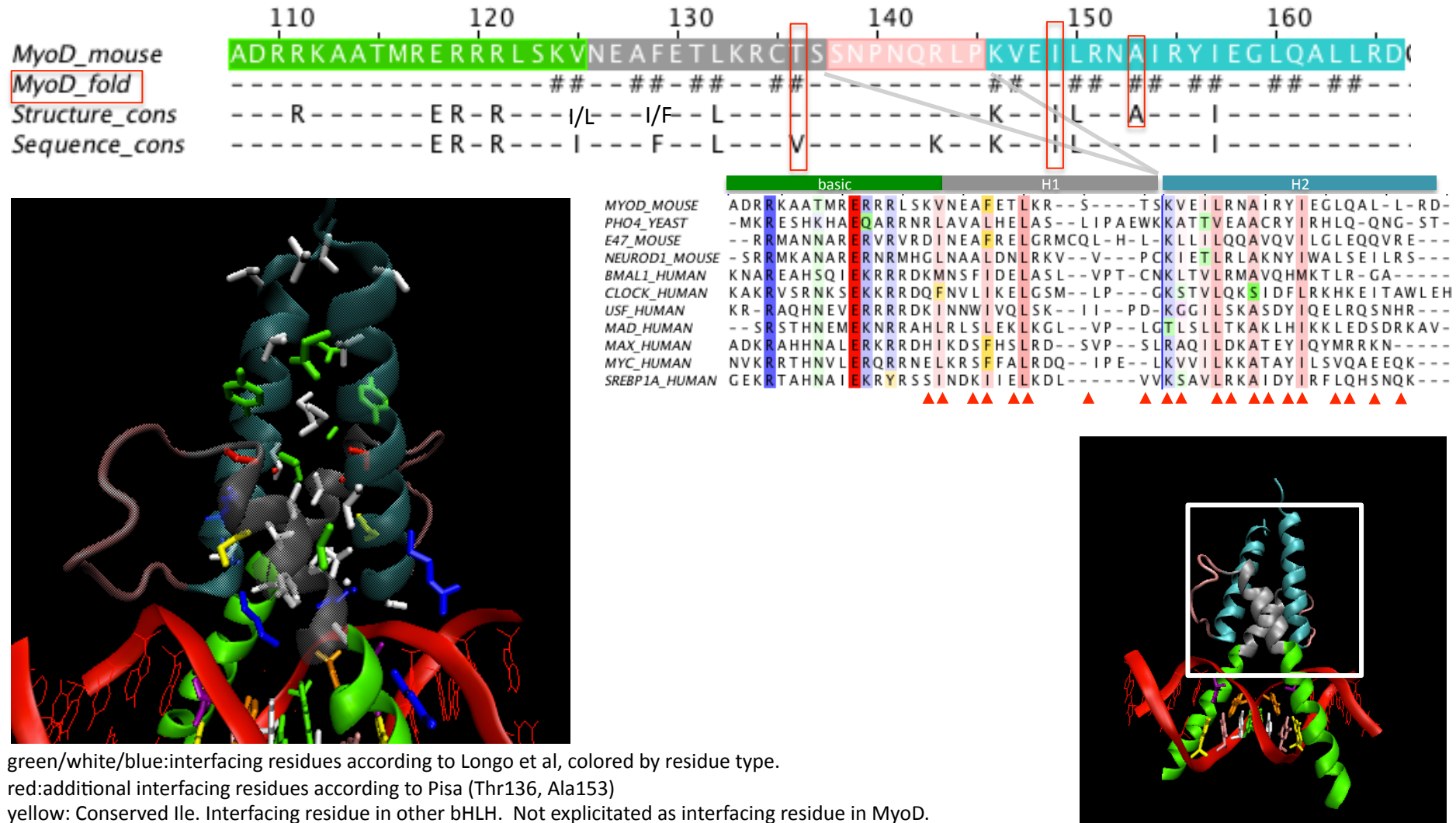
bHLH (residues 108-166):
-**basic region**: DNA binding (major groove)
-**H1**, **LOOP**, **H2**: dimerisation



MyoD preferred binding site:
5'-CAGCTG-3'
Each monomer does the same interactions (symmetry).

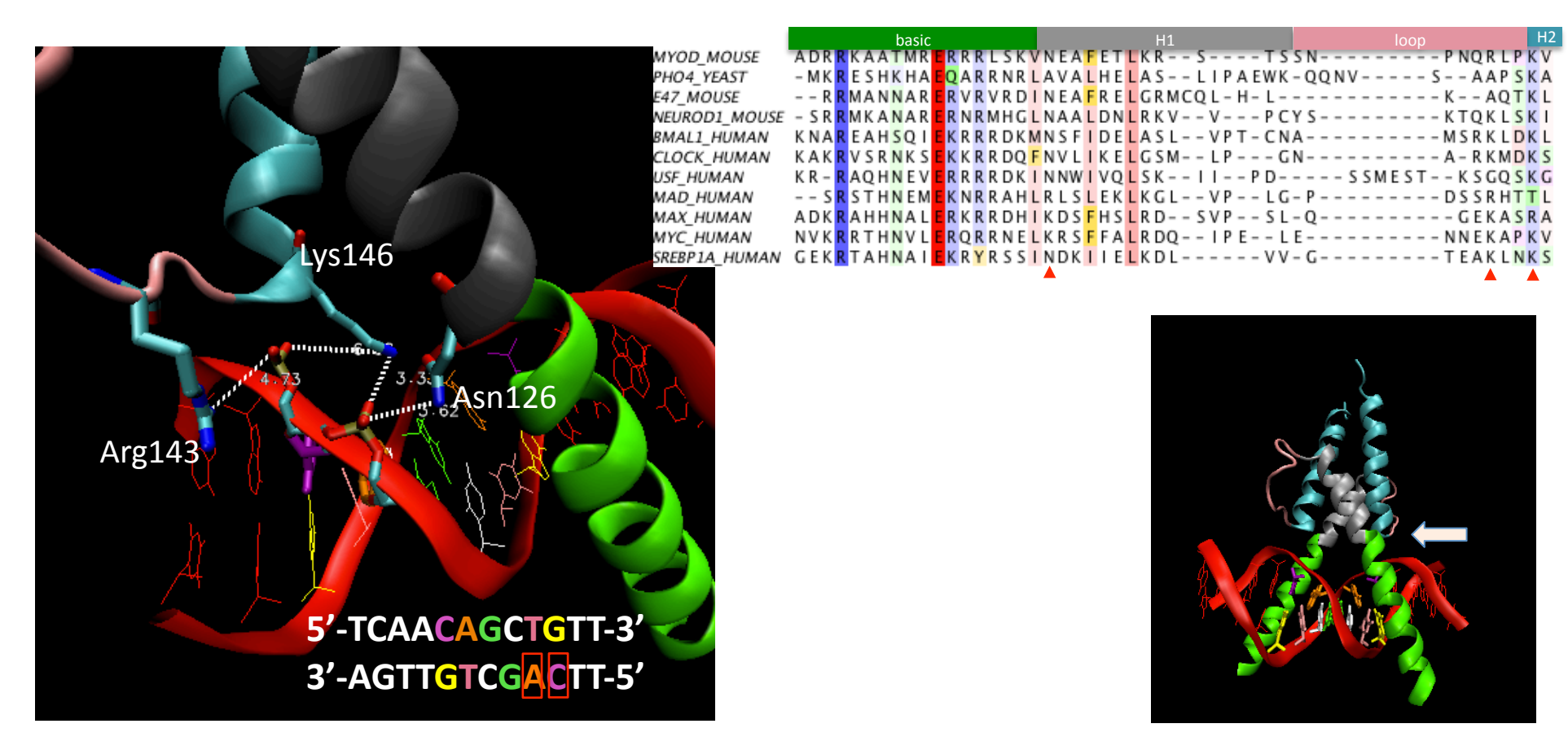
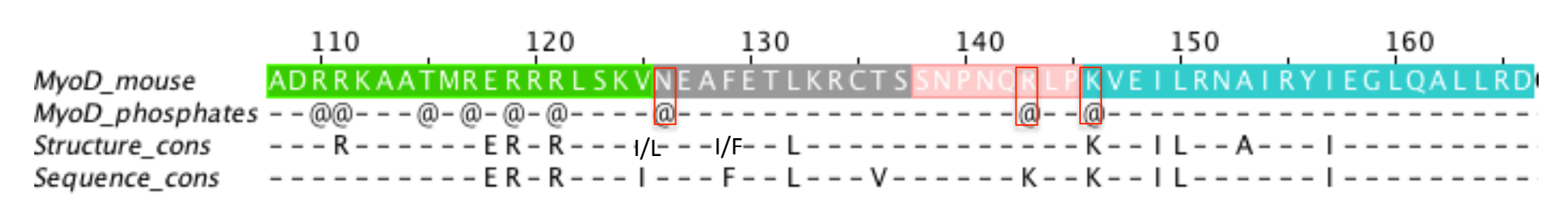
MyoD: dimerisation contacts/fold

The dimer is stabilised by packing interactions (VDW) and the hydrophobic effect. Hydrophobic side-chains of H1 and H2 are importantly involved in these contacts.



MyoD: phosphate contacts

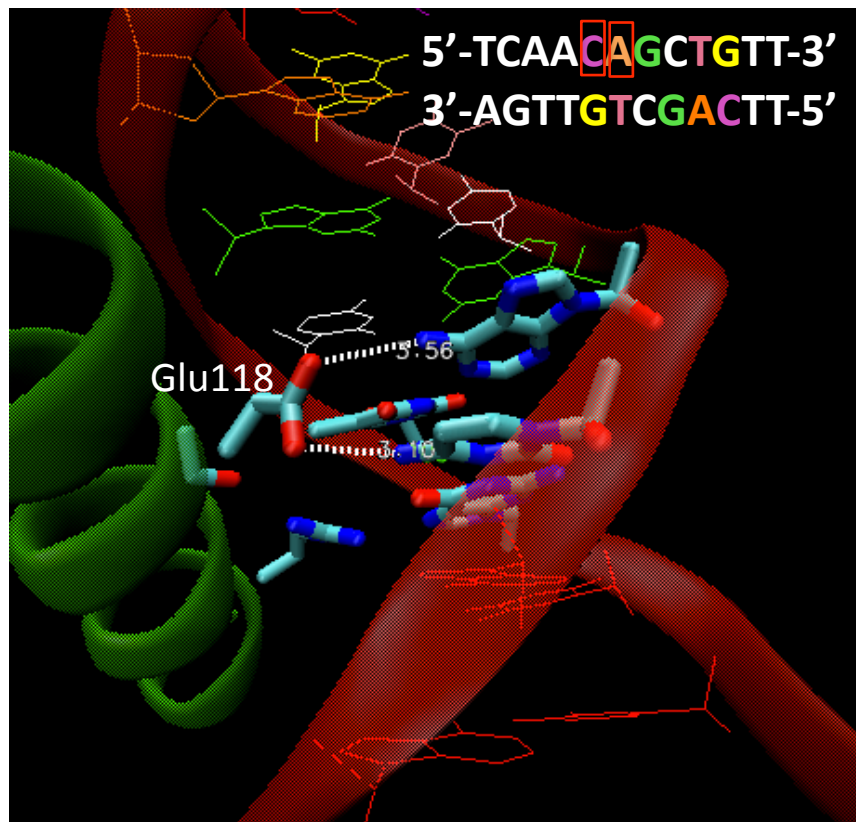
Residues outside the basic region also contact phosphates.



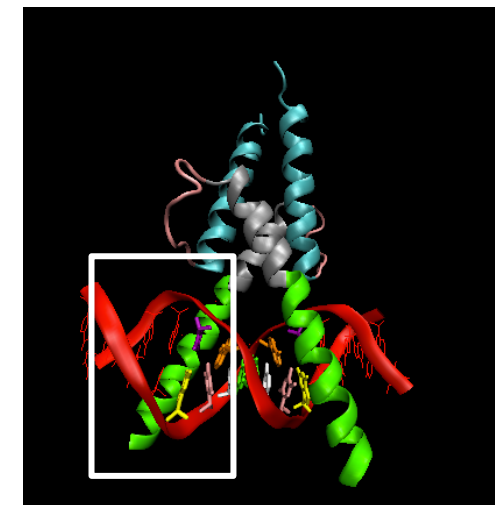
MyoD: base contacts

Arg111, Thr115 and Glu118 are responsible for sequence-specific contacts with the bases.
No direct contacts are observed with the two central bases.

	110	120	130	140	150	160				
<i>MyoD_mouse</i>	ADRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGLQALLRD									
<i>MyoD_bases</i>	---\$---\$---\$-----									
<i>Structure_cons</i>	---R-----ER-R---I/L---I/F---L-----K---IL--A---I-----									
<i>Sequence_cons</i>	-----ER-R---I---F---L---V-----K---K---IL-----I-----									



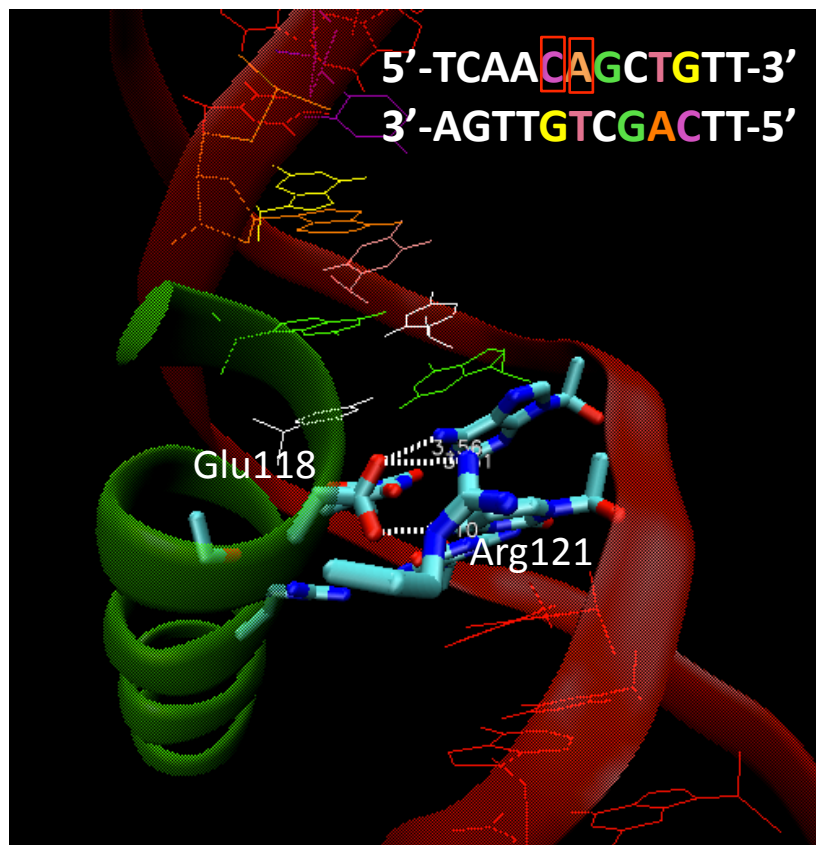
	basic									
MYOD_MOUSE	ADRR	KAATMR	ERRRLSKV							
PHO4_YEAST	-MKR	ESHKHA	QARRNRL							
E47_MOUSE	--RR	MANNAR	ERVVRDI							
NEUROD1_MOUSE	-SRR	MKANAR	ERNRMHGL							
BMAL1_HUMAN	KNAR	EAHSQI	EKRRRDQM							
CLOCK_HUMAN	KAKR	VS RNKS	EKKRRDQF							
USF_HUMAN	KR-	RAQHNEV	ERRRRDKI							
MAD_HUMAN	--SR	STHNEM	EKNRRRAHL							
MAX_HUMAN	ADKR	AHHNAL	ERKRRDHI							
MYC_HUMAN	NVKR	RTHNVL	ERQRRNEL							
SREBP1A_HUMAN	GEKR	TAHNAI	EKRYSSTI							



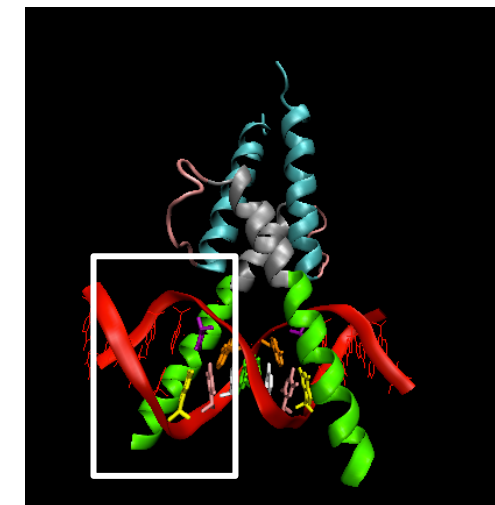
MyoD: base contacts

Arg111, Thr115 and Glu118 are responsible for sequence-specific contacts with the bases. No direct contacts are observed with the two central bases.

	110	120	130	140	150	160				
<i>MyoD_mouse</i>	ADRRKAATMRERRRLSKVNEAFETLKRCTSSNPNQRLPKVEILRNAIRYIEGLQALLRD									
<i>MyoD_bases</i>	---\$---\$---\$-----									
<i>Structure_cons</i>	---R-----ER-R---I/L---I/F---L-----K---IL--A---I-----									
<i>Sequence_cons</i>	-----ER-R---I---F---L---V-----K---K---IL-----I-----									



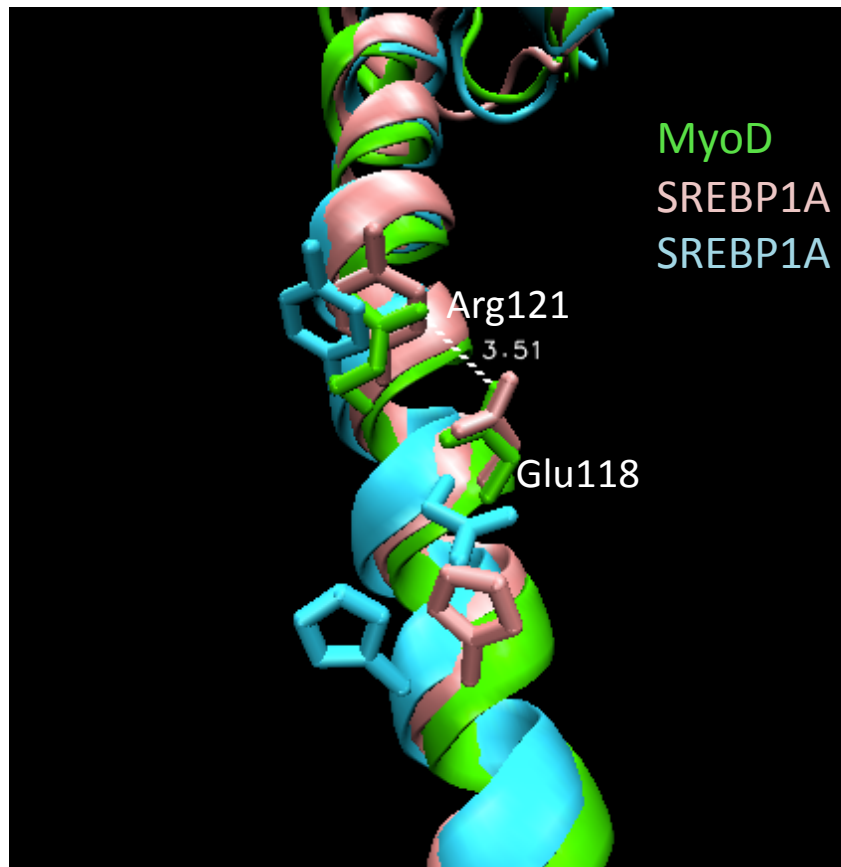
	basic									
MYOD_MOUSE	ADRR	KAATMR	ERRRLSKV							
PHO4_YEAST	-MKR	ESHKHA	QARRNRL							
E47_MOUSE	--RR	MANNAR	ERVVRDI							
NEUROD1_MOUSE	-SRR	MKANAR	ERNRMHGL							
BMAL1_HUMAN	KNAR	EAHSQI	EKRRRDKM							
CLOCK_HUMAN	KAKR	VS RNKS	EKKRRDQF							
USF_HUMAN	KR-	RAQHNEV	ERRRRDKI							
MAD_HUMAN	--SR	STHNEM	EKNRRRAHL							
MAX_HUMAN	ADKR	AHHNAL	ERKRRDHI							
MYC_HUMAN	NVKR	RTHNVL	ERQRRNEL							
SREBP1A_HUMAN	GEKR	TAHNAI	EKRYSII							



The importance of sequence

A specific amino acid sequence determines a specific DNA binding site.

Arg→Tyr: change in the conformation of Glu due to loss of a salt-bridge.



5'-CAGCTG

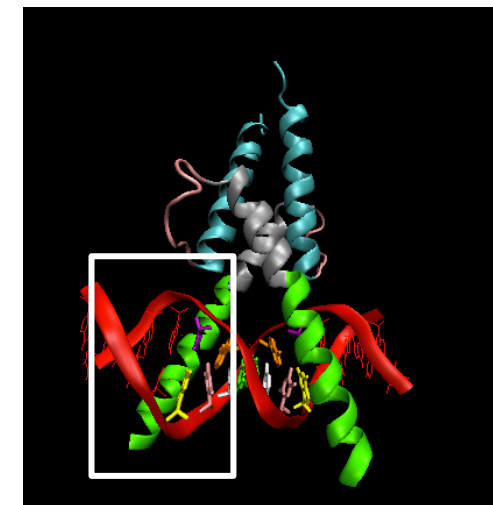
Canonical E-box: CACGTG

StRE—5' -BPyCACNCCAPy-3'

(B: anything except T, Py:pyrimidine)

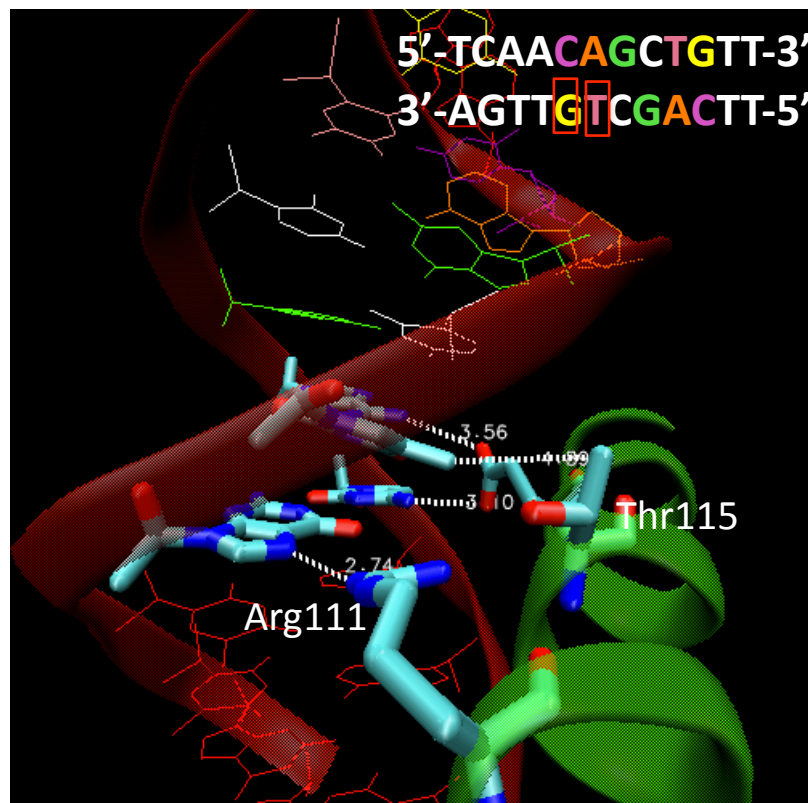
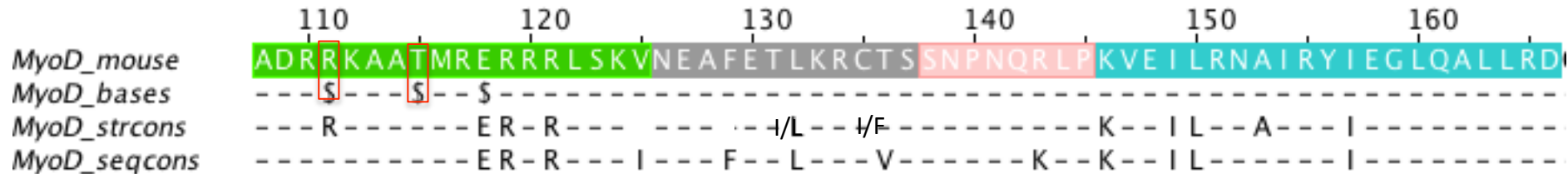
e.g.: 5' -ATCACCCAC-3

	basic									
MYOD_MOUSE	A	D	R	R	K	A	A	T	M	R
PHO4_YEAST	-	M	K	R	E	S	H	K	H	A
E47_MOUSE	-	-	R	R	M	A	N	N	A	R
NEUROD1_MOUSE	-	S	R	R	M	K	A	N	A	R
BMAL1_HUMAN	K	N	A	R	E	A	H	S	Q	I
CLOCK_HUMAN	K	A	K	R	V	S	R	N	K	S
USF_HUMAN	K	R	-	R	A	Q	H	N	E	V
MAD_HUMAN	-	-	S	R	S	T	H	N	E	M
MAX_HUMAN	A	D	K	R	A	H	H	N	A	L
MYC_HUMAN	N	V	K	R	R	T	H	N	V	L
SREBP1A_HUMAN	G	E	K	R	T	A	H	N	A	I

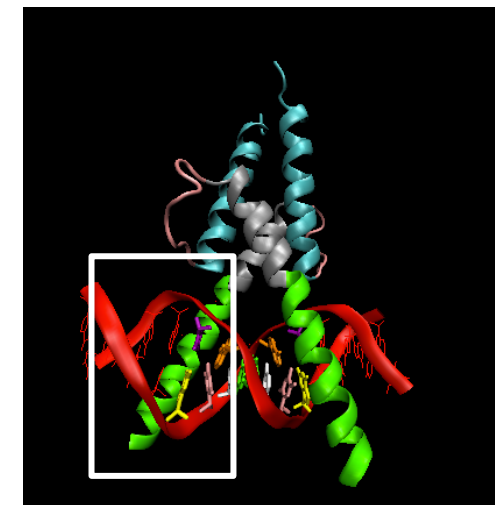


MyoD: base contacts

Arg111, Thr115 and Glu118 are responsible for sequence-specific contacts with the bases. No direct contacts are observed with the two central bases.

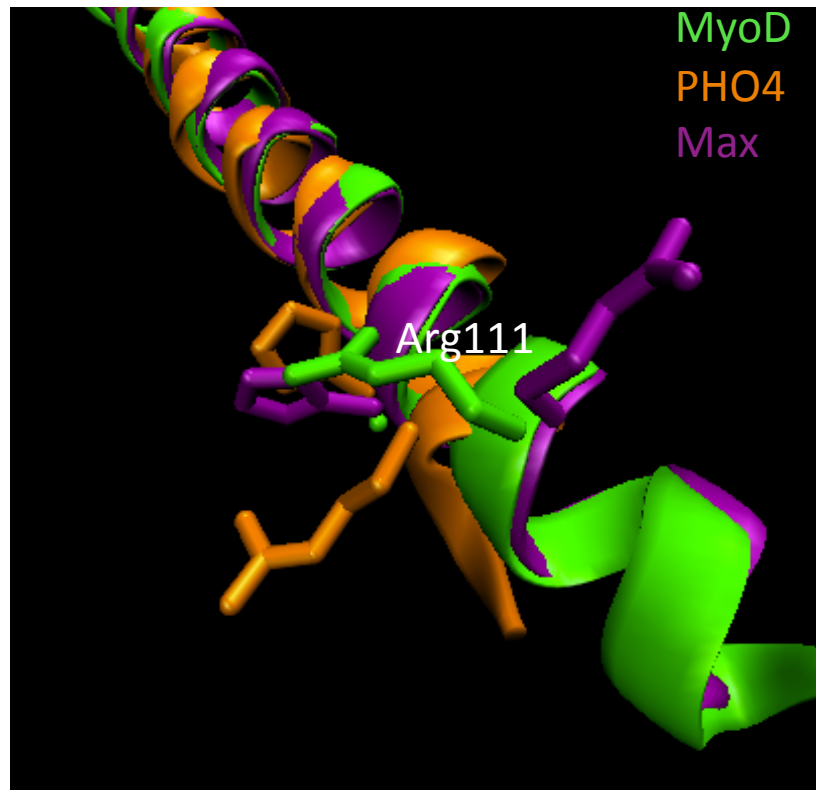
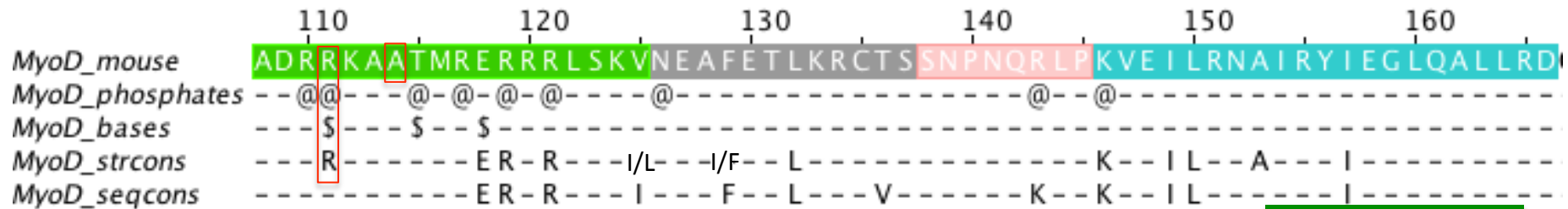


	basic									
MYOD_MOUSE	ADR	R	K	A	A	T	M	R	E	R
PHO4_YEAST	-	M	K	R	E	S	H	K	H	A
E47_MOUSE	-	-	R	R	M	A	N	N	A	R
NEUROD1_MOUSE	-	S	R	R	M	K	A	N	A	R
BMAL1_HUMAN	K	N	A	R	E	A	H	S	Q	I
CLOCK_HUMAN	K	A	K	R	V	S	R	N	K	S
USF_HUMAN	K	R	-	R	A	Q	H	N	E	V
MAD_HUMAN	-	-	S	R	S	T	H	N	E	M
MAX_HUMAN	A	D	K	R	A	H	H	N	A	L
MYC_HUMAN	N	V	K	R	R	T	H	N	V	L
SREBP1A_HUMAN	G	E	K	R	T	A	H	N	A	I

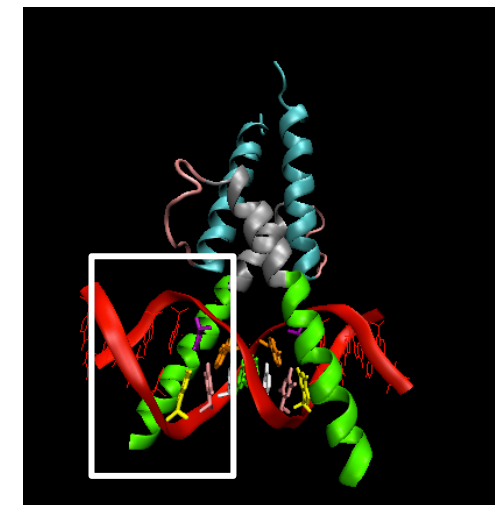


The importance of sequence

Conservation at one position does not always imply same structure/function. Surrounding residues matter!



	basic														
MYOD_MOUSE	ADR	R	K	A	A	T	M	R	E	R	R	L	S	K	V
PHO4_YEAST	-MK	R	E	S	H	K	H	A	E	Q	A	R	N	R	L
E47_MOUSE	--	R	R	M	A	N	N	A	R	E	R	V	R	D	I
NEUROD1_MOUSE	-SR	R	M	K	A	N	A	R	E	R	N	M	H	G	L
BMAL1_HUMAN	KNA	R	E	A	H	S	Q	I	E	K	R	R	R	D	K
CLOCK_HUMAN	KAK	R	V	S	R	N	K	S	E	K	K	R	R	D	Q
USF_HUMAN	KR-	R	A	Q	H	N	E	V	E	E	R	R	R	D	K
MAD_HUMAN	--S	R	S	T	H	N	E	M	E	K	N	R	A	H	L
MAX_HUMAN	ADK	R	A	H	H	N	A	L	E	E	K	R	R	D	H
MYC_HUMAN	NVK	R	T	H	N	V	L	E	R	Q	R	R	N	E	L
SREBP1A_HUMAN	GK	R	T	A	H	N	A	I	E	K	R	Y	R	S	S



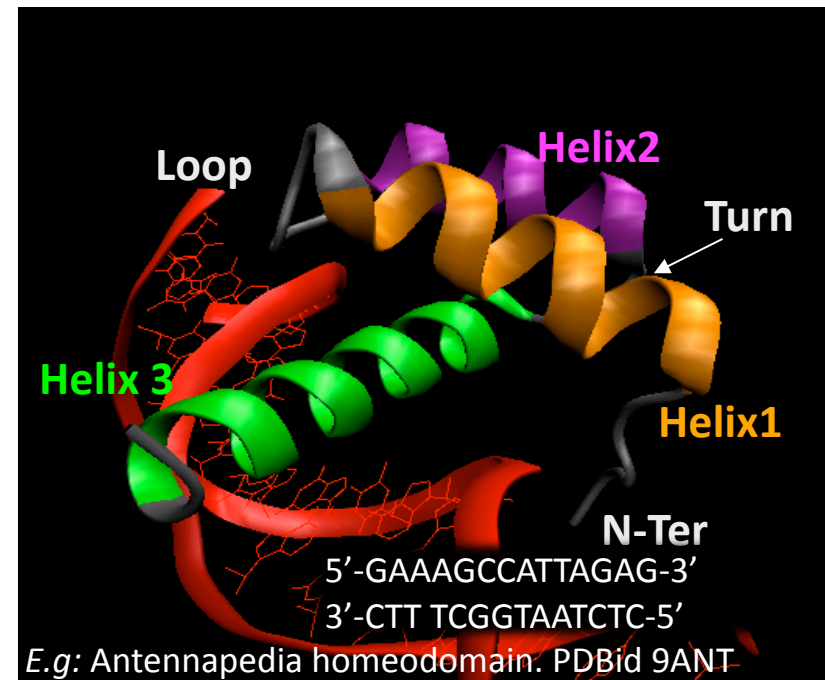
The helix turn helix (HTH) TFs

HTH domain:

60 amino acids

SCOP classification:

- Class a: all alpha
- Fold: a.4: DNA/RNA-binding 3-helical bundle
- Superfamily: a.4.1: Homeodomain-like
- Family: a.4.1.1: Homeodomain



DNA binding site: preference for 5'-ATTA-3'.

In vivo cooperative binding gives them the specificity they need

Tandemly arranged DNA binding domains: POU, ZnF, LeuZ

Functions:

- Embryonic axial patterning (Antennapedia, Engrailed, Bicoid)
- Pluripotency maintenance (NANOG)

The helix turn helix (HTH) TFs



- 41 protein domains (65 structures)



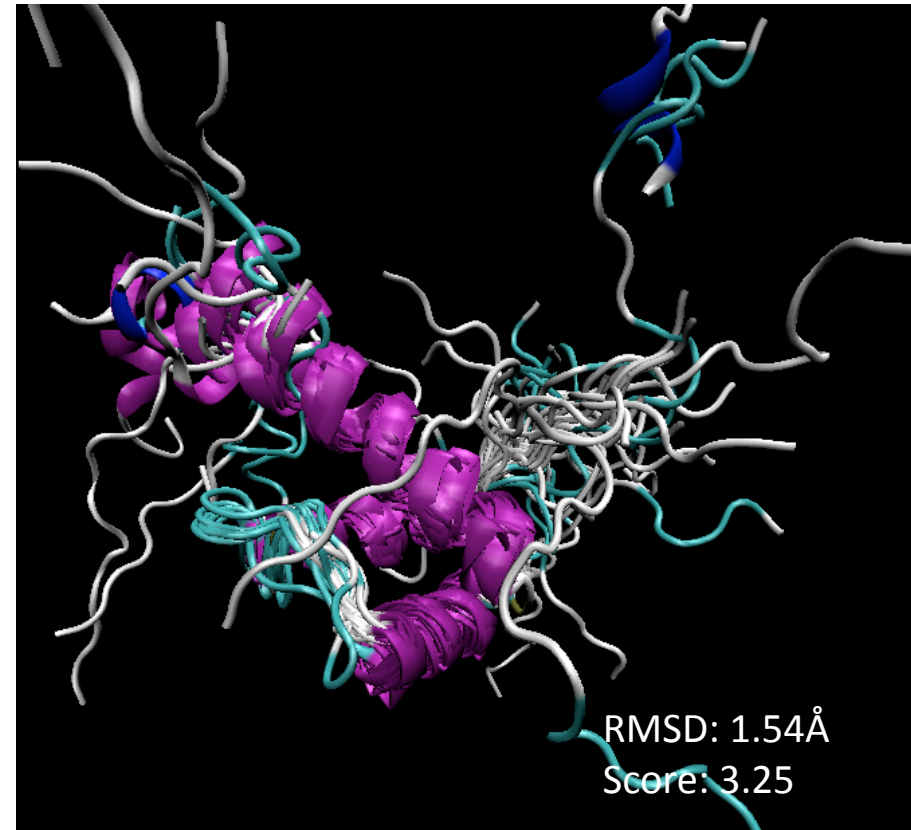
Superimposition of **39 domain coordinates files** from SCOP
(two sequences with a large insertion were removed)



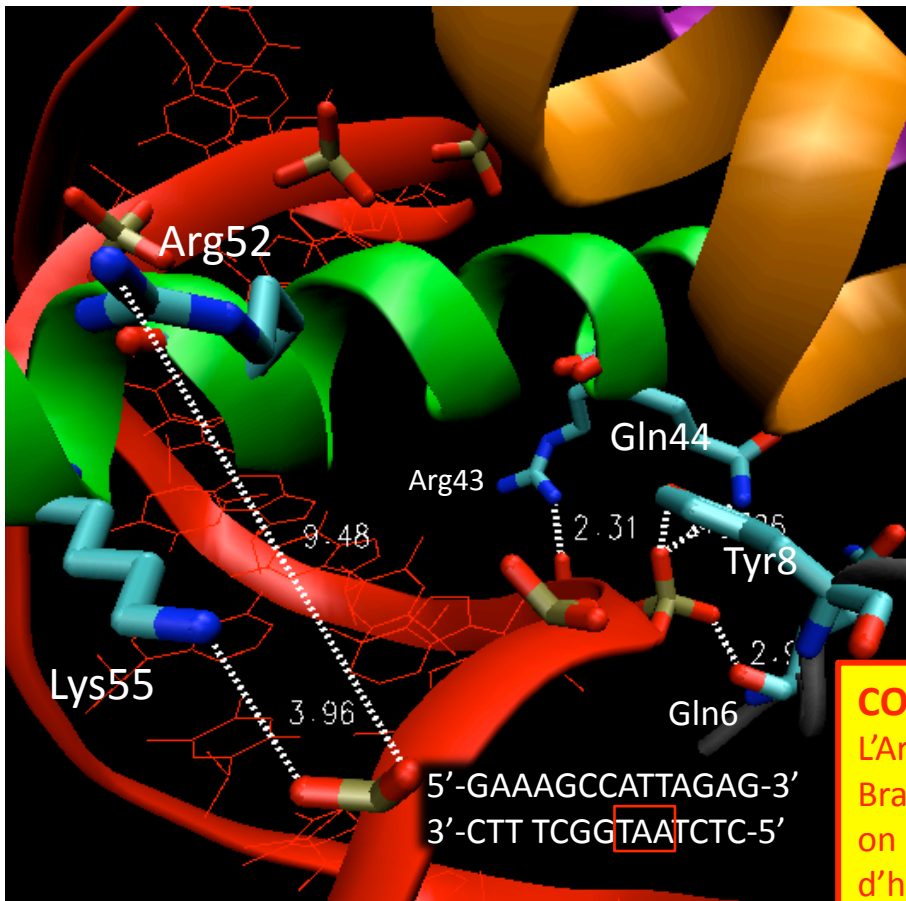
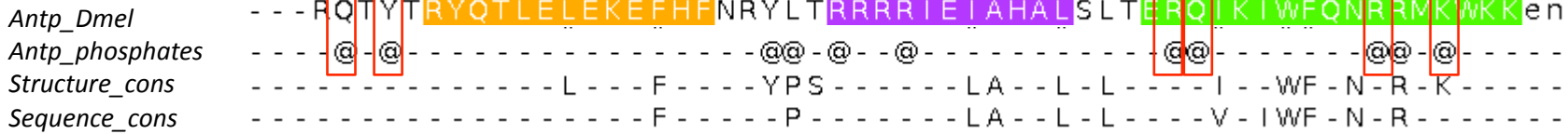
- Pfam Homeobox (PF00046)
- hmmsearch hits (E-value<0.0001)



Sequences aligned: **1613** fragments



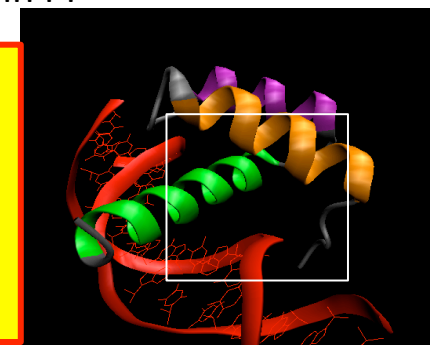
Antp: Phosphate contacts



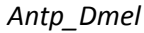
9anth/1-56	RO	TY	TR	YQ	TL	EL	KE	FH	NR	YL	TR	RR	RI	EA	HA	LS	LT	ER	QI	KI	WF	QN	RR	MK	WK	-K	
2cra/1-58	R	I	P	YS	K	G	Q	L	R	E	L	E	R	E	Y	A	N	K	F	I	T	K	D	K	R	R	K
1pufa/1-77	RC	P	Y	T	K	H	Q	L	E	L	E	K	E	F	L	N	M	Y	L	T	R	D	R	Y	E	V	A
2e1o/1-57	Q	V	R	F	S	N	D	Q	T	I	E	L	E	K	K	F	E	Q	K	Y	L	S	P	P	E	R	K
1x2n/1-62	R	G	V	L	P	K	H	A	T	N	V	M	R	S	W	L	F	H	P	P	Y	T	E	D	E	K	K
1b72a/1-68	R	T	N	F	T	R	Q	L	T	R	O	L	E	K	E	F	H	N	K	Y	L	S	R	A	R	R	V
1akha/1-49	-	-	-	I	S	P	Q	A	R	A	F	L	E	E	V	F	R	K	O	S	L	N	S	K	E	E	V
1b8b/1-58	R	R	N	F	S	K	O	A	E	I	L	N	E	Y	F	N	P	Y	L	T	S	E	A	E	E	L	A
21lx/1-60	R	T	F	S	Q	Q	L	E	N	E	A	T	F	N	R	Y	P	D	M	S	T	R	E	E	I	A	V
1b8ia/1-62	R	O	T	Y	T	R	Y	Q	T	L	E	L	E	K	E	F	H	N	H	Y	L	T	R	R	R	I	E
1f1ja/1-65	R	T	T	S	A	S	O	L	D	E	L	A	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E	E
1oet/1-60	R	T	S	I	E	T	N	I	R	V	A	L	E	K	S	F	L	N	Q	K	P	T	S	E	E	I	T
3hdda/1-56	R	T	A	F	S	S	E	Q	L	A	R	L	K	R	E	F	H	N	R	Y	L	T	E	R	R	Q	L
3hdda/1-55	R	T	A	F	S	S	E	Q	L	A	R	L	K	R	E	F	H	N	R	Y	L	T	E	R	R	Q	L
1ig7/1-58	R	T	P	F	T	A	Q	L	L	A	L	E	R	K	F	R	K	O	Y	L	S	I	A	E	R	A	E
1igga/1-57	R	T	A	F	T	R	D	Q	L	G	R	L	E	K	E	F	Y	E	N	Y	V	S	R	P	R	C	E
1iggb/1-57	R	T	A	F	T	R	D	Q	L	G	R	L	E	K	E	F	Y	E	N	Y	V	S	R	P	R	C	E
1f1jb/1-58	R	T	T	F	S	A	S	O	L	D	E	L	A	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E
2f1jc/1-59	R	T	T	F	S	A	S	O	L	D	E	L	A	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E
1au7a/1-58	R	T	T	S	I	A	A	K	D	A	L	E	R	H	F	G	H	S	K	P	S	S	Q	E	I	M	R
1au7b/1-58	R	T	T	S	I	A	A	K	D	A	L	E	R	H	F	G	H	S	K	P	S	S	Q	E	I	M	R
1aplc/1-59	G	H	R	F	T	K	E	N	V	R	I	L	S	W	F	A	N	P	L	Y	D	T	K	G	L	E	N
1apld/1-58	G	H	R	F	T	K	E	N	V	R	I	L	S	W	F	A	N	P	L	Y	D	T	K	G	L	E	N

Electrostatic: Arg43, Arg52, Lys55
HB: (Gln6), Tyr8, Gln44

CORRECCIÓ BALDO & NURIA:
L'Arg52 surt al llibre de text de
Branden & Tooze però no al paper
on descriuen el cristall!! No l'hauria
d'haver tingut en compte, està
massa lluny... sorry!



Antp: Phosphate contacts



Antp_phosphates

Structure_cons

Sequence_cons

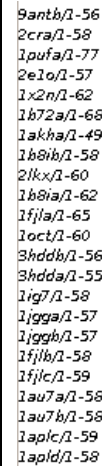


Arg28

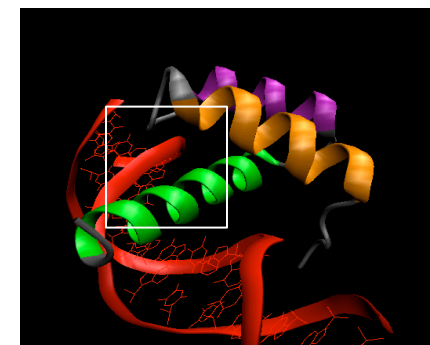
Arg31

Arg53

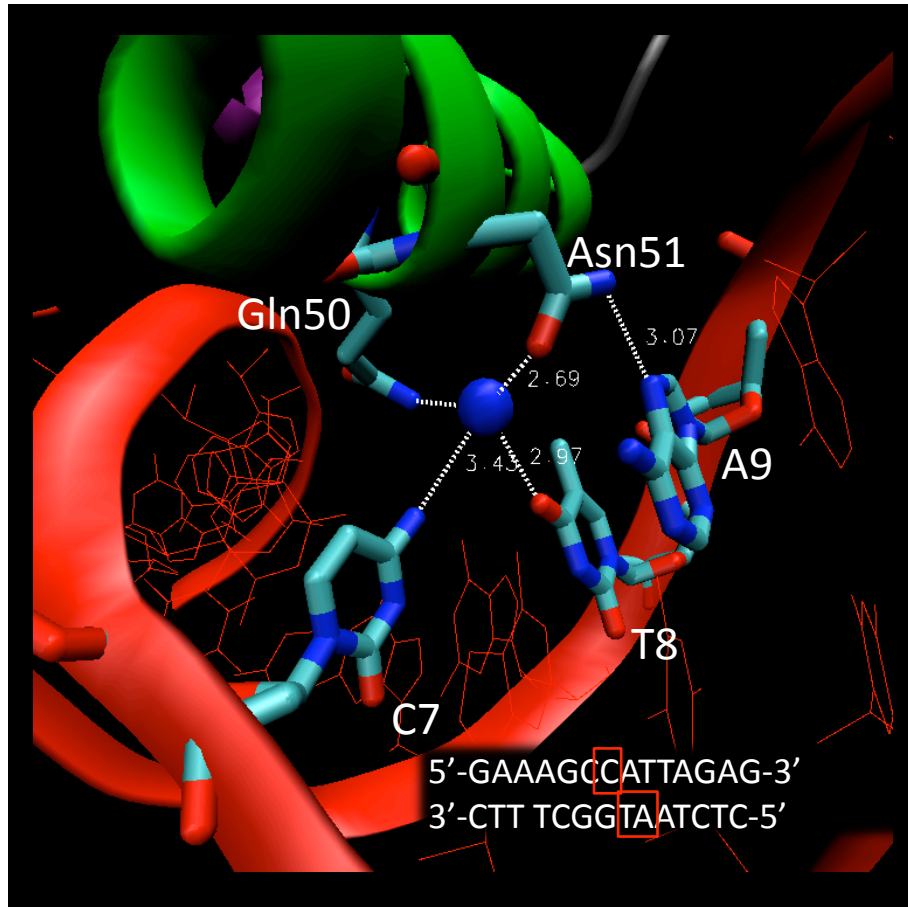
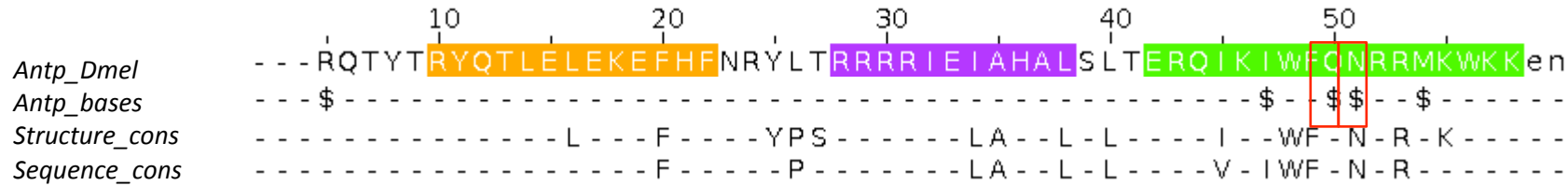
5'-GAAAGCCATTAGAG-3'
3'-CTTTCGGTAATCTC-5'



HB: Tyr25

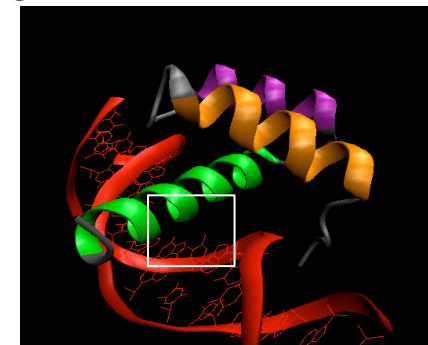


Antp: Base contacts



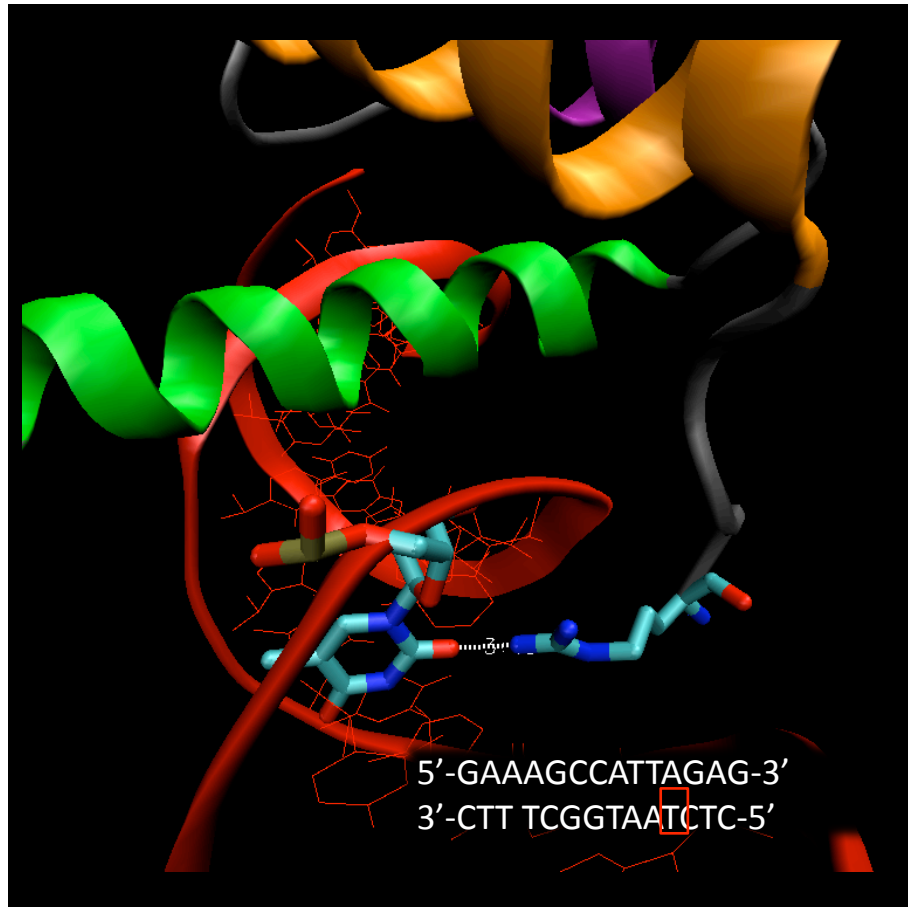
9antb/1-56	R	Q	T	Y	T	R	Y	Q	L	E	L	E	K	E	F	H	N	R	Y	L	T	R	R	R	R	I	E	A	H	A	L	S	L	T	E	R	Q	I	K	I	W	F	C	N	R	R	M	K	W	K	-	K	
2cra/1-58	R	I	P	Y	S	K	G	Q	L	R	E	L	E	R	E	Y	A	N	K	F	I	T	K	D	K	R	R	I	S	A	A	T	S	L	S	E	R	Q	I	T	I	W	F	Q	N	R	R	V	K	E	-	K	
1pufa/1-77	R	C	P	Y	T	K	H	O	L	E	L	E	K	E	F	L	N	M	Y	L	T	R	D	R	R	Y	E	V	A	R	L	L	N	L	T	E	R	Q	V	K	I	W	F	Q	N	R	R	M	K	M	-	K	
2e1o/1-57	Q	V	R	F	S	N	D	Q	T	I	E	L	E	K	K	F	E	Q	K	Y	L	S	P	P	E	R	K	R	L	A	K	M	L	Q	L	S	E	R	Q	V	K	T	W	F	Q	N	R	A	K	W	-	R	
1x2n/1-62	R	G	V	L	P	K	H	A	T	N	V	M	R	S	W	L	F	H	P	Y	P	T	E	D	E	K	K	Q	I	A	A	Q	T	N	L	T	L	L	Q	V	N	N	W	F	I	N	A	R	R	I	-	L	
1b72a/1-68	R	T	N	F	T	T	R	Q	L	E	L	E	K	E	F	H	N	K	Y	L	S	R	A	R	R	V	E	I	A	A	T	L	E	N	E	T	Q	V	K	I	W	F	Q	N	R	R	M	K	Q	-	K		
1akha/1-49	-	-	-	I	S	P	O	A	R	A	F	L	E	E	V	F	R	K	Q	S	L	N	S	K	E	E	E	V	A	K	K	G	I	T	P	L	Q	V	R	W	F	I	N	K	R	M	R	S	-	-			
1b8ib/1-58	R	R	N	F	S	K	Q	A	S	E	I	L	N	E	Y	F	Y	N	P	Y	P	S	E	E	A	K	E	L	A	R	K	C	G	I	T	S	Q	V	S	N	W	F	G	N	K	R	I	Y	-	K			
2lxx/1-60	R	T	H	F	T	S	Q	Q	L	O	E	L	E	A	T	F	O	N	R	Y	P	D	M	S	T	R	E	E	I	A	V	W	T	N	L	T	E	A	R	V	R	W	F	K	N	R	A	K	W	-	R		
1b8ia/1-62	R	Q	T	Y	T	R	Y	Q	T	L	E	L	E	K	E	F	H	N	H	Y	L	T	R	R	R	I	E	M	A	H	A	L	S	L	T	E	R	Q	I	K	I	W	F	Q	N	R	R	M	K	L	-	K	
1fjla/1-65	R	T	T	F	S	A	S	Q	L	D	E	L	E	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E	E	L	A	Q	R	T	N	L	T	E	A	R	I	Q	V	W	F	Q	N	R	A	R	L	-	R	
1oct/1-60	R	T	S	I	E	T	N	I	R	V	A	L	E	K	S	F	L	N	Q	K	P	T	S	E	E	I	T	M	I	A	D	Q	L	N	M	E	K	E	V	I	R	W	F	C	N	R	R	O	K	E	-	R	
3hddh/1-56	R	T	A	F	S	S	E	Q	L	A	R	L	K	R	E	F	N	N	R	Y	L	T	E	R	R	R	Q	L	S	S	E	L	G	L	N	E	A	Q	I	K	I	W	F	Q	N	K	R	A	K	I	-	K	
3hdda/1-55	R	T	A	F	S	S	E	Q	L	A	R	L	K	R	E	F	N	N	R	Y	L	T	E	R	R	R	Q	L	S	S	E	L	G	L	N	E	A	Q	I	K	I	W	F	Q	N	K	R	A	K	I	-	K	
1ig7/1-58	R	T	P	F	T	T	A	Q	L	L	A	L	E	R	K	F	R	K	Q	Y	L	S	I	A	E	R	A	E	F	S	S	S	L	S	L	T	E	T	Q	V	K	I	W	F	Q	N	R	A	K	A	-	R	
1igga/1-57	R	T	A	F	T	R	D	O	L	G	R	L	E	K	E	F	Y	E	N	Y	V	S	R	P	R	R	C	E	L	A	A	Q	L	N	L	P	E	S	T	I	K	V	W	F	Q	N	R	R	M	K	D	-	K
1iggb/1-57	R	T	A	F	T	R	D	O	L	G	R	L	E	K	E	F	Y	E	N	Y	V	S	R	P	R	R	C	E	L	A	A	Q	L	N	L	P	E	S	T	I	K	V	W	F	Q	N	R	R	M	K	D	-	K
1fjlb/1-58	R	T	T	F	S	A	S	Q	L	D	E	L	E	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E	E	L	A	Q	R	T	N	L	T	E	A	R	I	Q	V	W	F	Q	N	R	A	R	L	-	K	
1fjlc/1-59	R	T	T	F	S	A	S	Q	L	D	E	L	E	R	A	F	E	T	Q	Y	P	D	I	Y	T	R	E	E	L	A	Q	R	T	N	L	T	E	A	R	I	Q	V	W	F	Q	N	R	A	R	L	-	K	
1au7a/1-58	R	T	T	I	S	I	A	A	K	D	A	L	E	R	H	F	G	H	S	K	P	S	S	Q	E	I	M	R	M	A	E	E	L	N	L	E	K	E	V	R	W	F	C	N	R	R	O	K	E	R	V		
1au7b/1-58	R	T	T	I	S	I	A	A	K	D	A	L	E	R	H	F	G	H	S	K	P	S	S	Q	E	I	M	R	M	A	E	E	L	N	L	E	K	E	V	R	W	F	C	N	R	R	O	K	E	R	V		
1aplc/1-59	G	H	R	F	T	K	E	N	V	R	I	L	E	S	W	F	A	N	P	Y	L	D	T	K	G	L	E	N	L	M	K	N	T	S	L	S	R	I	Q	I	K	N	W	V	S	N	R	R	K	E	-	K	
1apld/1-58	G	H	R	F	T	K	E	N	V	R	I	L	E	S	W	F	A	N	P	Y	L	D	T	K	G	L	E	N	L	M	K	N	T	S	L	S	R	I	Q	I	K	N	W	V	S	N	R	R	K	E	-	K	

Water mediated HB: Asn51, Gln50, C7, T8
 Direct HB: Asn51-A9



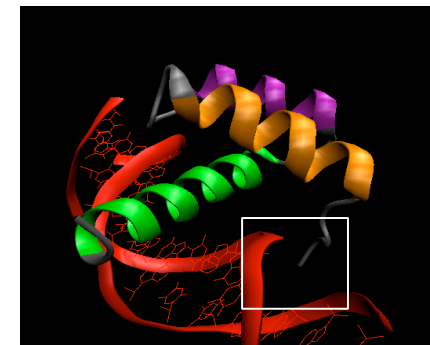
Antp: Base contacts

		10		20		30		40		50			
<i>Antp_Dmel</i>	--	RQTYT	RYQ	TELEKEF	HF	NRYLT	RRRR	IEIAHAL	SLT	ERQIK	IWFQNR	RMKWK	Ken
<i>Antp_bases</i>	--	\$									\$	\$	\$
<i>Structure_cons</i>	--			L		F		YPS		LA		L	L
<i>Sequence_cons</i>	--					F		P		LA		L	L

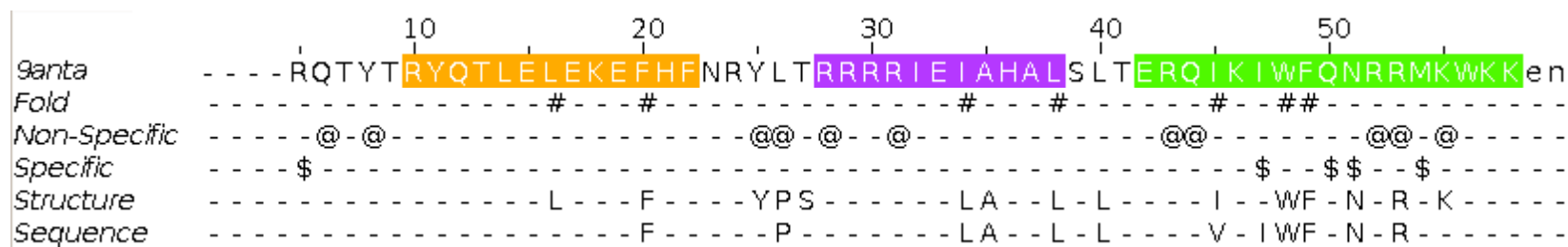


9antb/1-56	RQTYTRYQTLELEKEFHNNRYLT	RRRRRIEIAHALSLTERQIKI	IWFQNR	RMKWK-K
2cra/1-58	RIPYSKGQLRELEREYANKFI	TDKDKRKISAAATLSERQIT	IWFQNR	RVKE-K-K
1pufa/1-77	RCPTKHTQLELEKEFLNNMYL	TRDRRYEVARLLNLTERQVKI	IWFQNR	RMKM-K-K
2e1o/1-57	QVRFSNDQITIELEKKFEQKYL	SPPERKRLAKMLQLSERQVKT	WFO	NRAKW-RR
1x2n/1-62	RGVLPKHATNVMRSWLFHPYPT	EDEKKQIAAQTNLTLLQVNNWF	INARRI	-LQ
1b72a/1-68	RTNFTTRLTLELEKEFHNNYL	SRARRVEIAATLELNETQVKI	IWFQNR	RMKQ-K-K
1akha/1-49	--ISPOARAFLEEVRFKQSL	NSKEKEEVAKKCGITPLQVRVWF	INKMRS	--
1b8ib/1-58	RRNFSKQASEILNEYFYNFY	PSEAAKEELARCGITVSQVSN	WFG	NKIRY-K-K
2lxx/1-60	RTHTSSQQLQELEATFONRYP	DMSTREEIAVWNTLEARVRWF	KNNRAK	-RR
1b8ia/1-62	RQTYTRYQTLELEKEFHNNYL	TRRRRIEMAHALSLTERQIKI	IWFQNR	RMKL-K-K
1fjla/1-65	RTTFSASQDLELERAFETQYP	DIYTREELAQRTNLTEARIQV	WFO	NRRARL-R-K
1oct/1-60	RTSIENTNIRVALEKSLFNQKP	TSEETMIADQLNMEKEVIRV	WFC	NNRQKEK-R
3hddh/1-56	RTAFSSSEQLARLKREFNNRYL	TERRRQQLSSELGLNEAQIKI	IWFQNR	KRAKIK-K
3hdda/1-55	RTAFSSSEQLARLKREFNNRYL	TERRRQQLSSELGLNEAQIKI	IWFQNR	KRAKIK-K
1ig7/1-58	RTPFTTAQLLALERKFRKQYL	SIAERAEFSSSLSLTETQVKI	IWFQNR	RAKAK-R
1igga/1-57	RTAFTRDOLGRLEKEFYENYV	SRPRCELAQAQNLPESTIKV	WFO	NRRMKD-R
1iggb/1-57	RTAFTRDOLGRLEKEFYENYV	SRPRCELAQAQNLPESTIKV	WFO	NRRMKD-R
1fjlb/1-58	RTTFSASQDLELERAFETQYP	DIYTREELAQRTNLTEARIQV	WFO	NRRARL-R-K
1fjlc/1-59	RTTFSASQDLELERAFETQYP	DIYTREELAQRTNLTEARIQV	WFO	NRRARL-R-K
1au7a/1-58	RTTISIAAKDALERHFGHSKP	SSQEIIMRMAEELNLEKEVVRV	WFC	NNRQREKRV
1au7b/1-58	RTTISIAAKDALERHFGHSKP	SSQEIIMRMAEELNLEKEVVRV	WFC	NNRQREKRV
1aplc/1-59	GHRFTKENVRILESWFANPYL	DTKGLENLMKNTLSRIQIKNW	VS	NNRRKE--K
1apld/1-58	GHRFTKENVRILESWFANPYL	DTKGLENLMKNTLSRIQIKNW	VS	NNRRKE--K

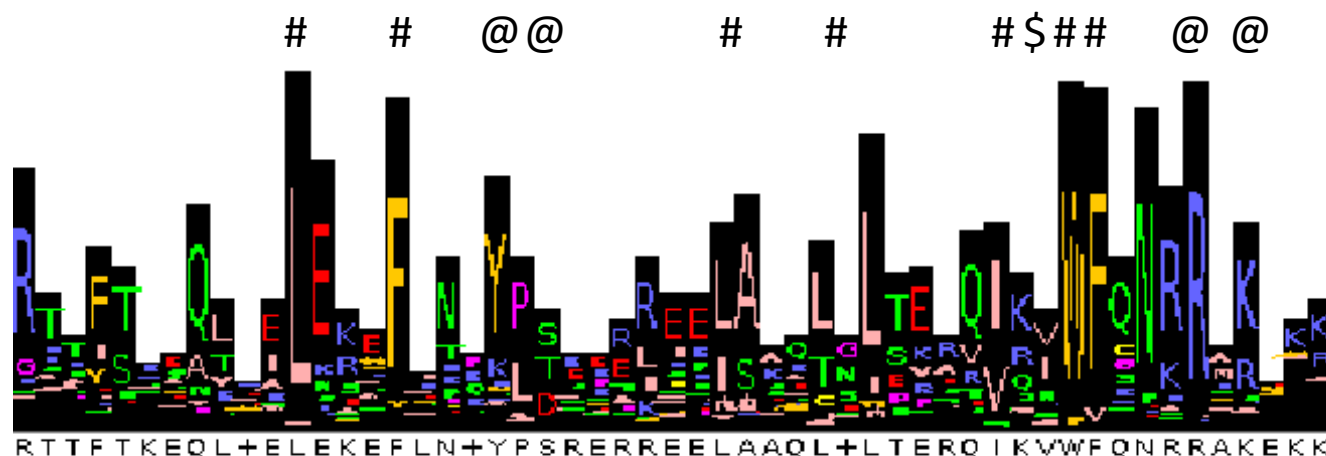
HB: Arg5 T11



Antp: Summary



Structural align.
Consensus logo
(39 str)



Sequence align.
Consensus logo
(1613 seq)



Outline

☐ Part1: Transcription factors

☐ Introduction

☐ Methodology

☐ Transcription factor families:

- Zinc fingers (Zn fingers)
- Basic Leucine Zipper (bZip)
- Basic-helix-loop-Helix (bHLH)
- Helix Turn Helix (HTH)

☒ Part2: Molecular domestication of DNA binding domains: CAG

☒ Introduction

☒ Modelling strategies

☒ Final models

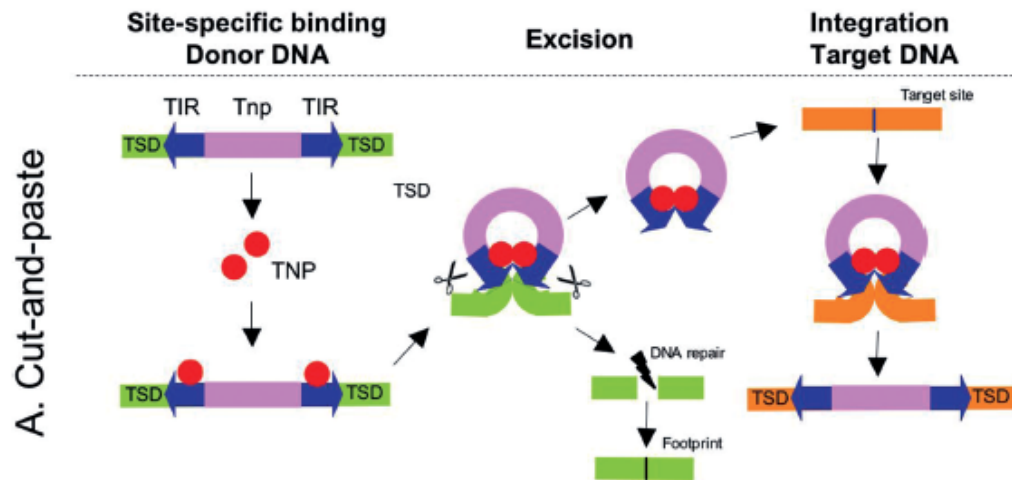
☒ Sequence analysis

☐ Conclusions

Introduction

Transposable Elements: Jumping Genes

Transposons encode the enzymes necessary for their own mobilization.



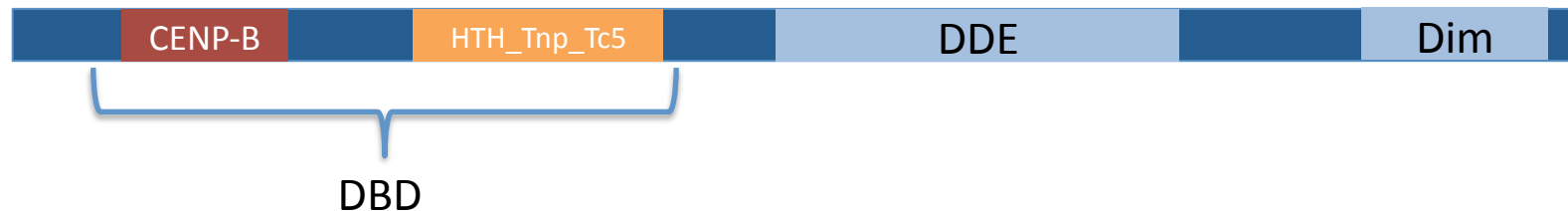
Tn5 transposase. PDBid: 1MUH

Introduction

Molecular Domestication: Turning junk into gold

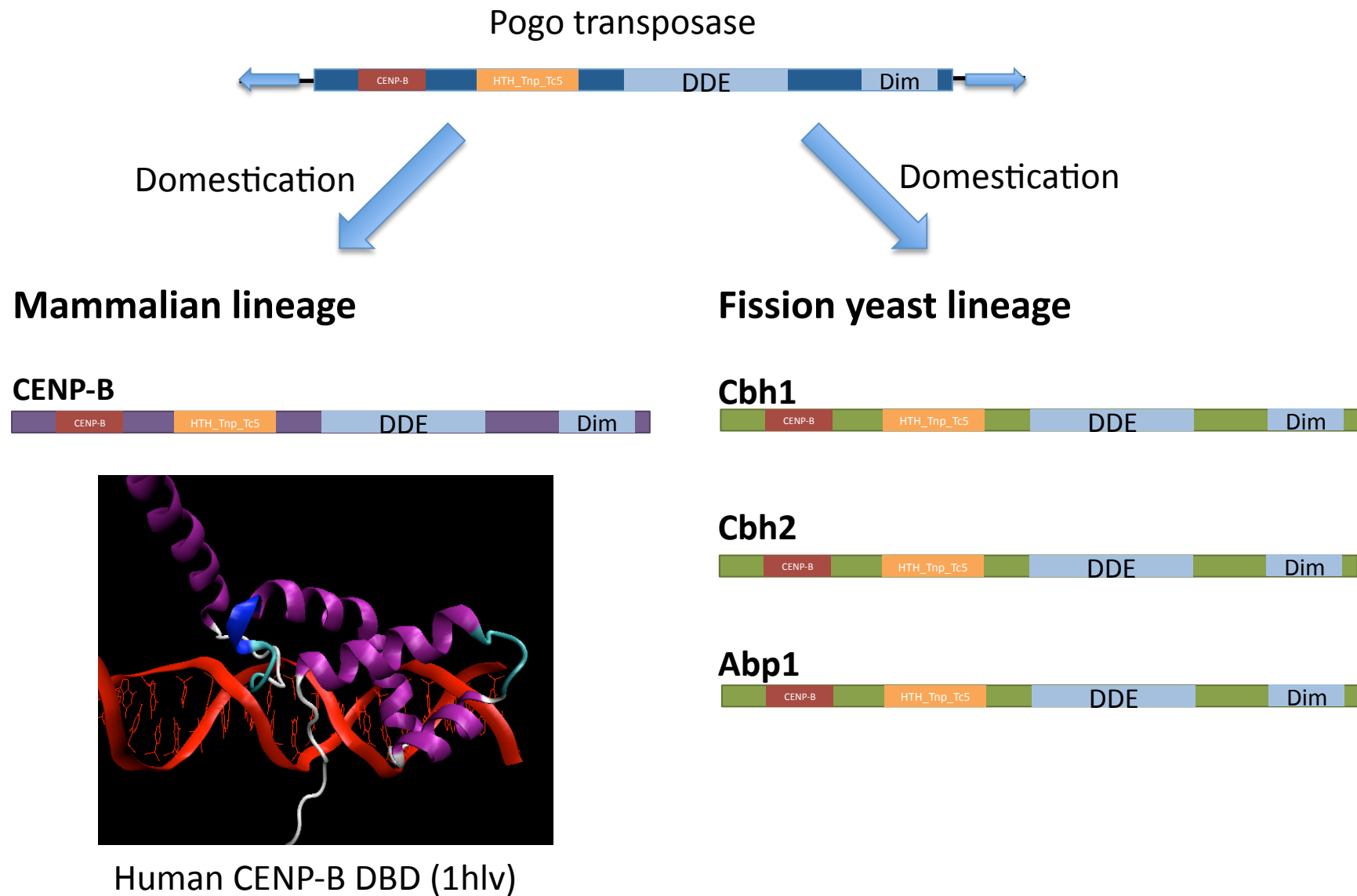
Transposases have been repeatedly recruited by the host genome during evolution.

Pogo transposase (class II) has a DNA binding domain, an endonuclease domain and a dimerization domain.

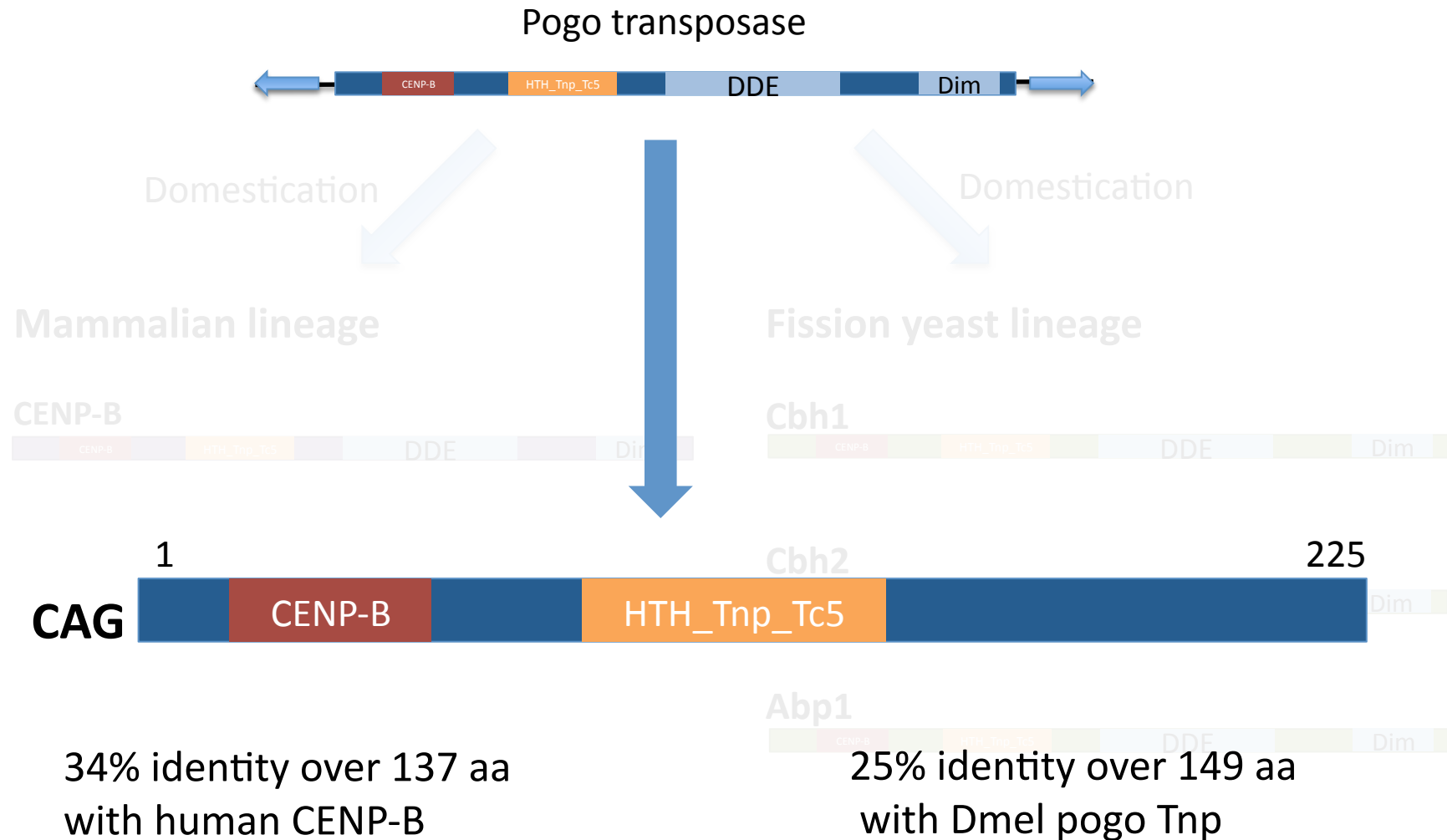


The DBD is responsible for the specific binding to transposon ends.

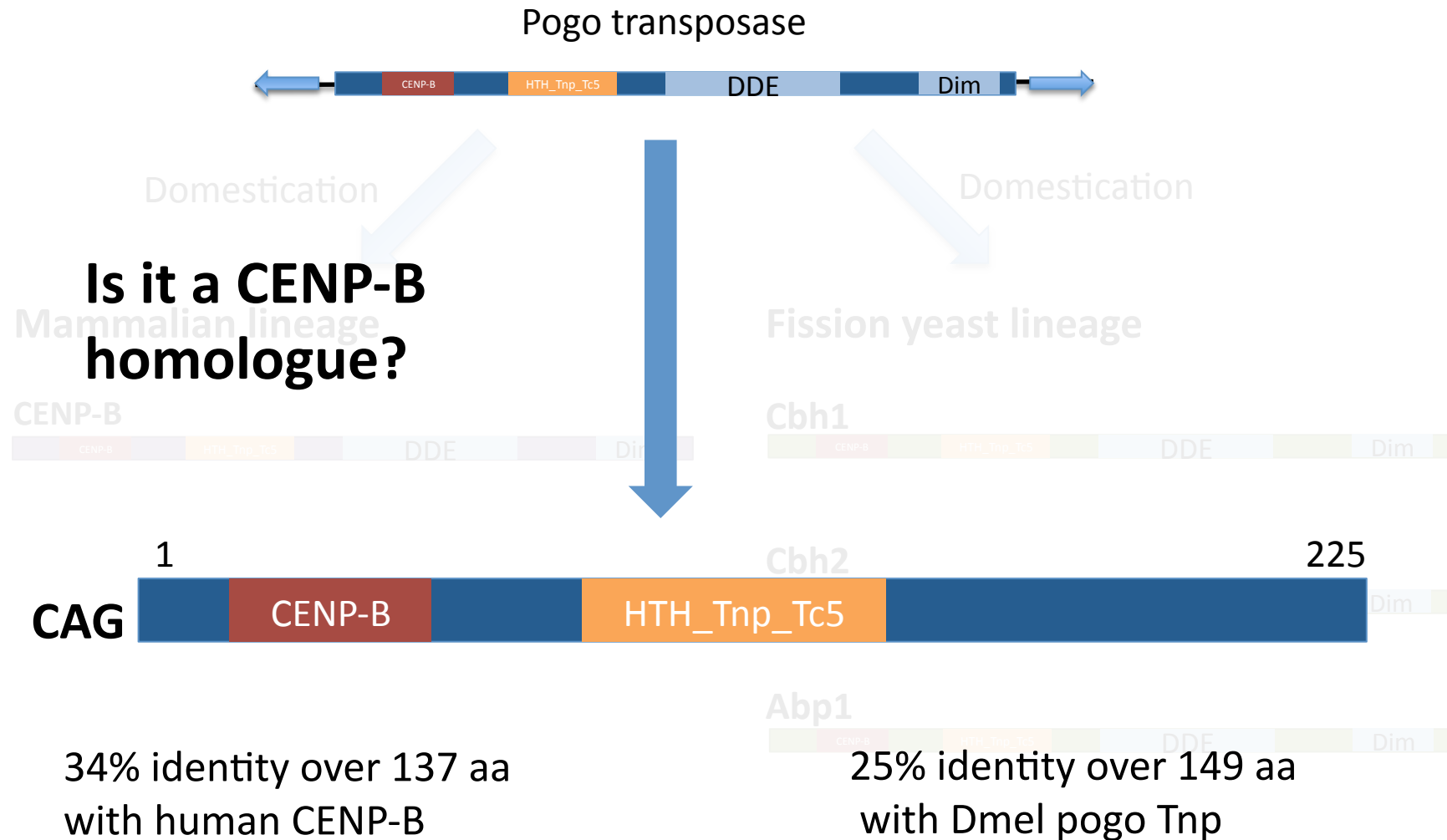
Introduction



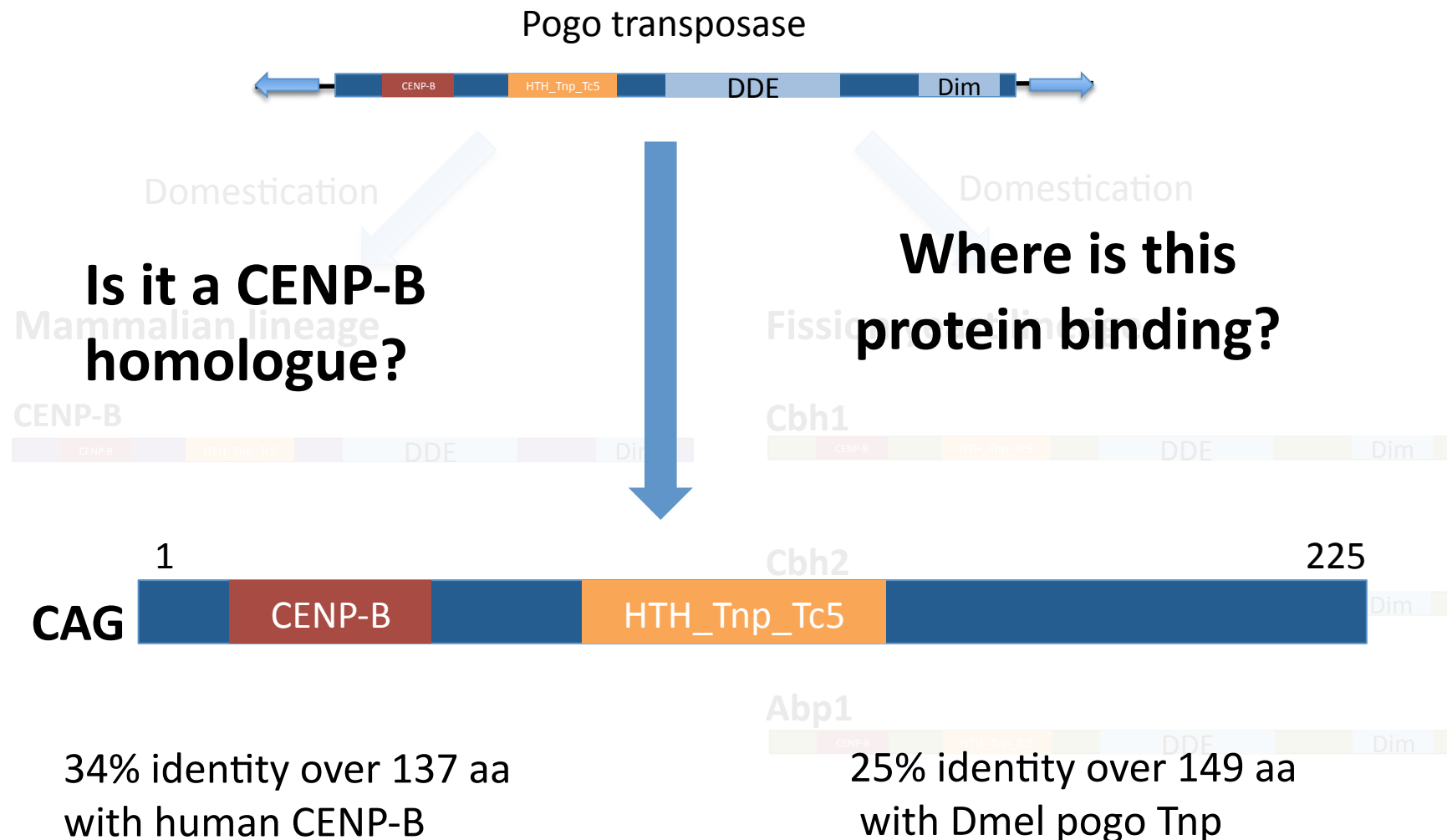
Introduction



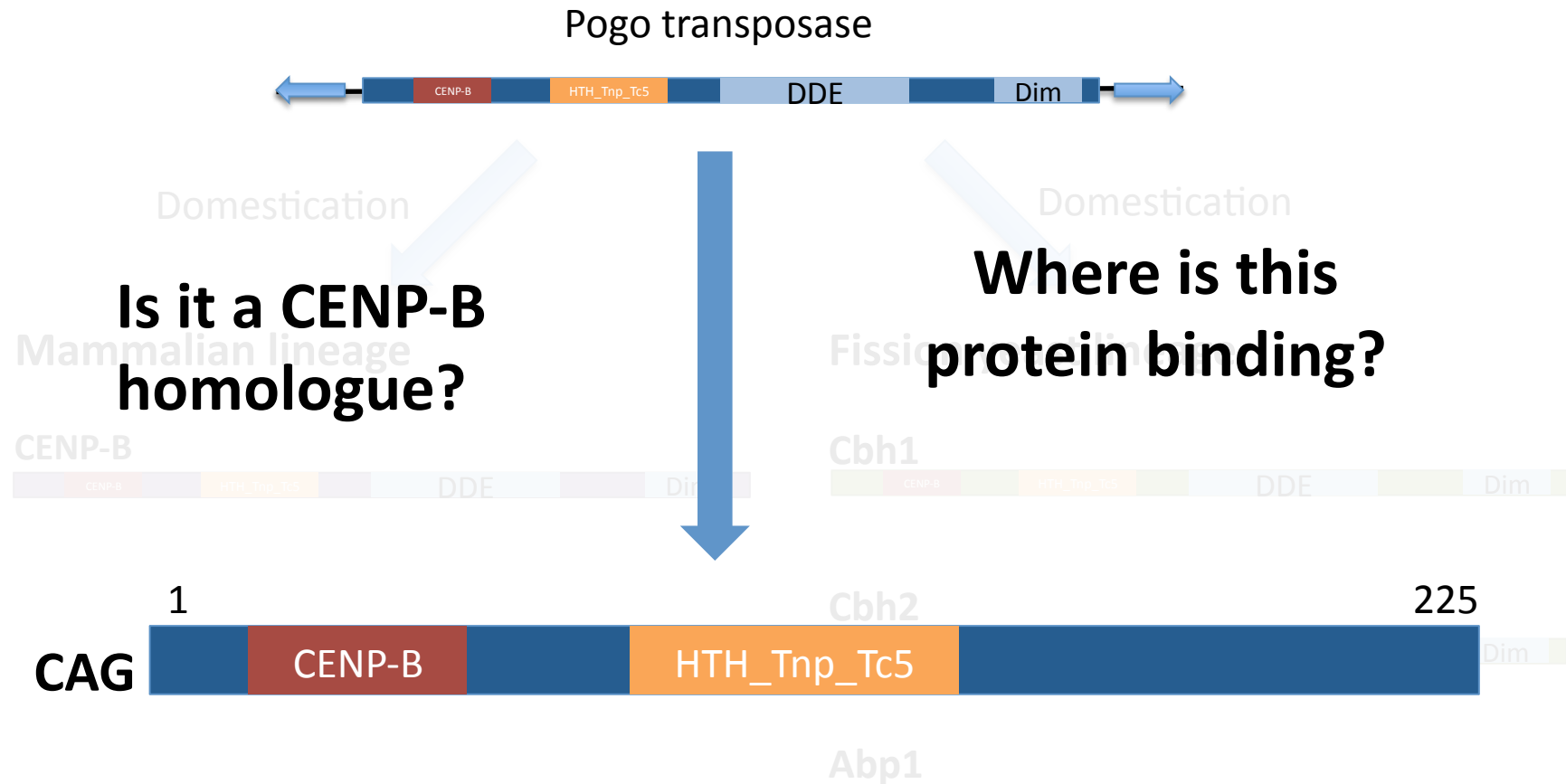
Introduction



Introduction



Introduction



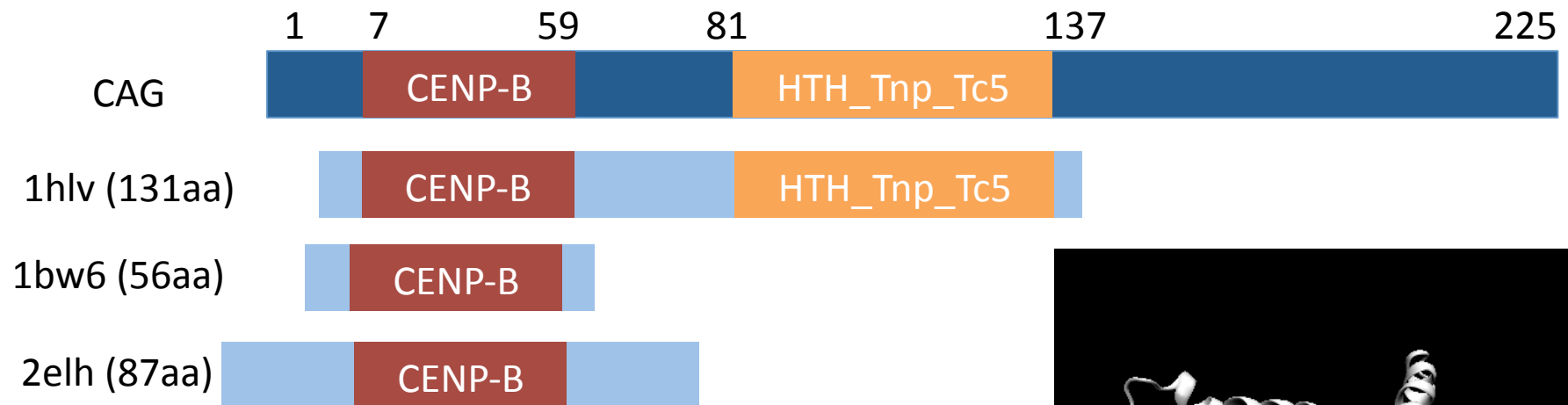
Modelling and MSA analysis

Methods: overview

1. Look for matching HMM (hmmscan)



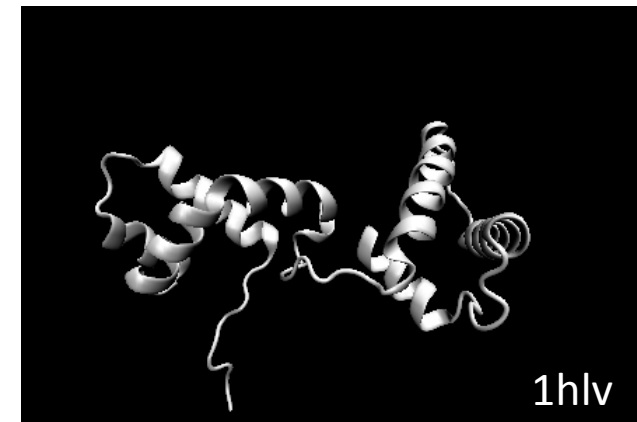
2. Search for templates: hmmsearch with best HMM (CENP-B) in pdb database.



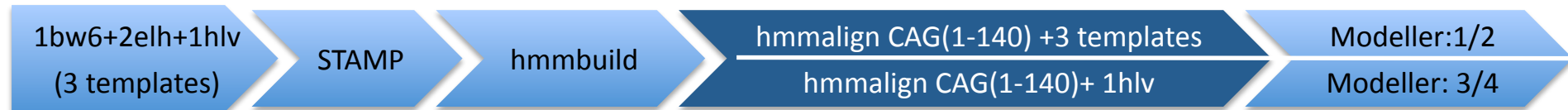
3. Sequence alignments

4. Modelling (Modeller and servers)

5. Evaluation of the models



Modelling strategies



Mod1/2

```

CAG      -----mpservrk---PRTSLTLEEKMEVIOSEQE--RNKLSVRDLAK-RFNIGK
1bw6     -----MGPKRRQLTFREKSRIIQEVEENPDLRKGEIAR-RFNIPP
2elh     gssgssgmnmirmatkakrpl-----RSLTPRDKIHAIQRIH--DGESKASVA-RDIGVPE
1hlv     -----MGPKRRQLTFREKSRIIQEVEENPDLRKGEIAR-RFNIPP

CAG      TQAADILKHKOSIKEGLL-SGELKLNmRRNPLSQRGAQIDEMCFDWFSRVRTENIPISG
1bw6     STLSTILKNKRAILASE-----
2elh     STLRGWCKNEDKLRFMSR-----QsATDNLCAD-ALGDKMD-----
1hlv     STLSTILKNKRAILASERKYGVASTCR-KTNKLSPY-DKLEGLLLIAWFQQIRAAGLPVKG

CAG      EMVRKKAKQLAVELGHSNFSTSSGWLEKWRKT--hmvryndtqds
1bw6     -----
2elh     -----
1hlv     IILKEKALRIAEEELGMDDFTASNGWLDRFRRRRS-----
  
```

Mod3/4

```

CAG      MPSEVRVK---PRTSLTLEEKMEVIOSEQE--RNKLSVRDLAKRFNIGKTQAADILKHKOSI
1hlv     -----MGPKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTILKNKRAI

CAG      KEGLL-SGELKLNmRRNPLSQRGAQIDEMCFDWFSRVRTENIPISGEMVRKKAKQLAVE
1hlv     LASERKYGVASTCR-KTNKLSPY-DKLEGLLLIAWFQQIRAAGLPVKGIIILKEKALRIAEE

CAG      LGHSNFSTSSGWLEKWRKT--HNVRYNDTGDS
1hlv     LGMDDFTASNGWLDRFRRRRS-----
  
```

Modelling strategies

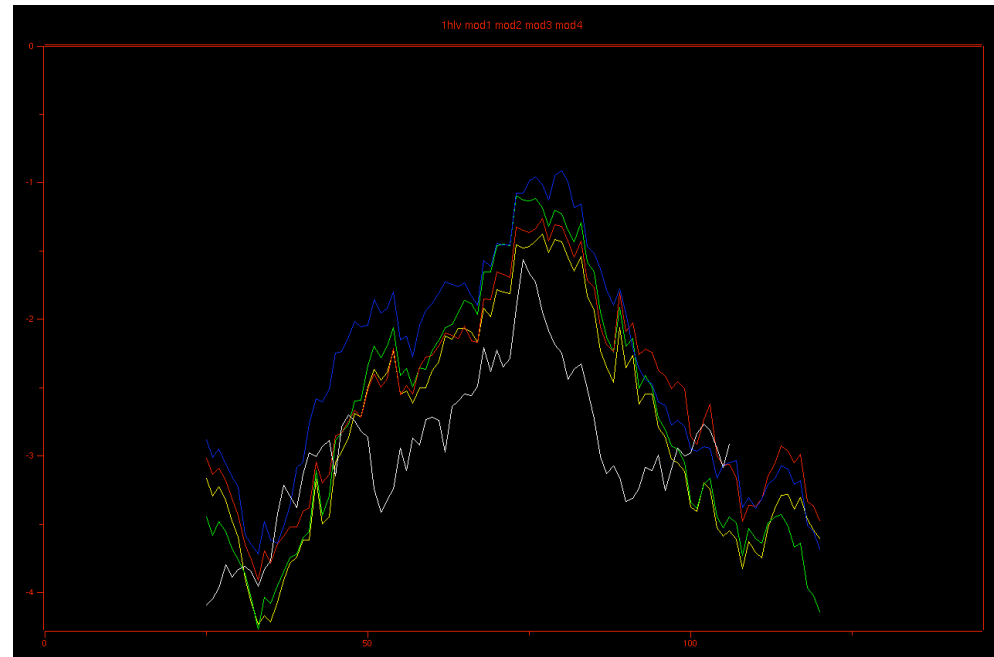
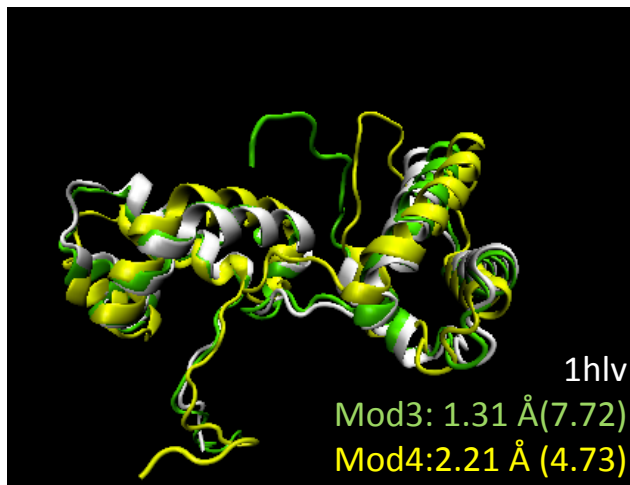
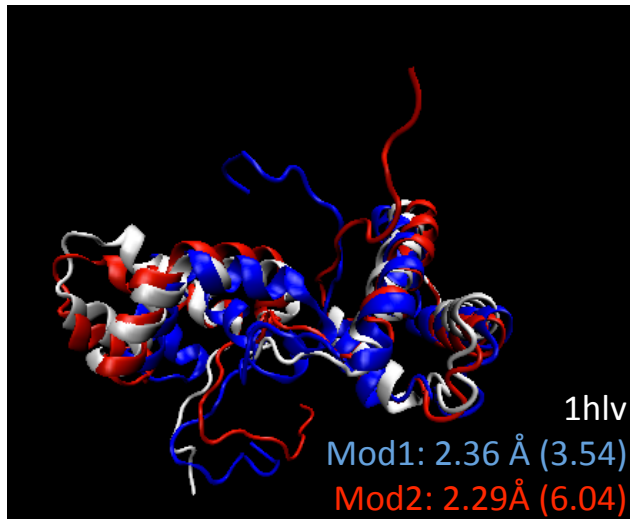
1bw6+2elh+1hlv
(3 templates)

STAMP

hmmbuild

hmmalign CAG(1-140) +3 templates
hmmalign CAG(1-140)+ 1hlv

Modeller:1/2
Modeller: 3/4



	Z-comb	Z-pair	Z-surf
1hlv	-6.66	-6.34	-3.47
Mod1	-5.99	-5	-3.69
Mod2	-6.31	-4.86	-4.30
Mod3	-6.70	-5.40	-4.33
Mod4	-6.44	-5.13	-4.22

Prosa energy
profiles and
Z-scores

Modelling strategies

hmmalign CAG(1-140) +3 templates+CENP-B.hmm

Modeller:5/6

hmmalign CAG(1-140)+ 1hlv+CENP-B.hmm

Modeller: 7/8

Mod5/6

```
CAG  -----MPSEVRKPRSTLTLEEKMEVIOEQERNK-LSVRDLAKRFNIGKTQAADI
1bw6 -----MG-PKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTI
2elh  GSSGSSGMNIRMGTGKGRPLRSLTPRDKIHAIQRIHDGE--SKASVARDIGVPESTLRGW
1hlv  -----MG-PKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTI
```

```
CAG  LKHKQSIKEG-LLSGELKLNQMRNPLSQRGAQIDEMCFDWFSSRVRTENIPISGEMVRKK
1bw6 LKNKRAILAS-E-----
2elh CKNEDKLRFMRSQSATDNLCADALGDKM-----
1hlv LKNKRAILAS-ERKYGVASTCRKTNKLSPYDKLEGLLIAWFQQIRAAGLPVKGIIKEKA
```

```
CAG  AKQLAVELGHSNFTSSGWLEKWRKTHNVRYNDTGDS
1bw6 -----
2elh -----
1hlv LRIAEEELGMDDFTASNGWLDREFRRRS-----
```

Mod7/8

```
CAG  MPSEVRKPRSTLTLEEKMEVIOEQERNK-LSVRDLAKRFNIGKTQAADILKHKQSIKEG
1hlv ----MG-PKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTILKNKRAILAS
```

```
CAG  LLSGELKLNQMRNPLSQRGAQIDEMCFDWFSSRVRTENIPISGEMVRKKAKQLAVELGHS
1hlv ERKYGVASTCRKTNKLSPYDKLEGLLIAWFQQIRAAGLPVKGIIKEKALRIAEEELGMDD
```

```
CAG  NFSTSSGWLEKWRKTHNVRYNDTGDS
1hlv FTASNGWLDREFRRRS-----
```

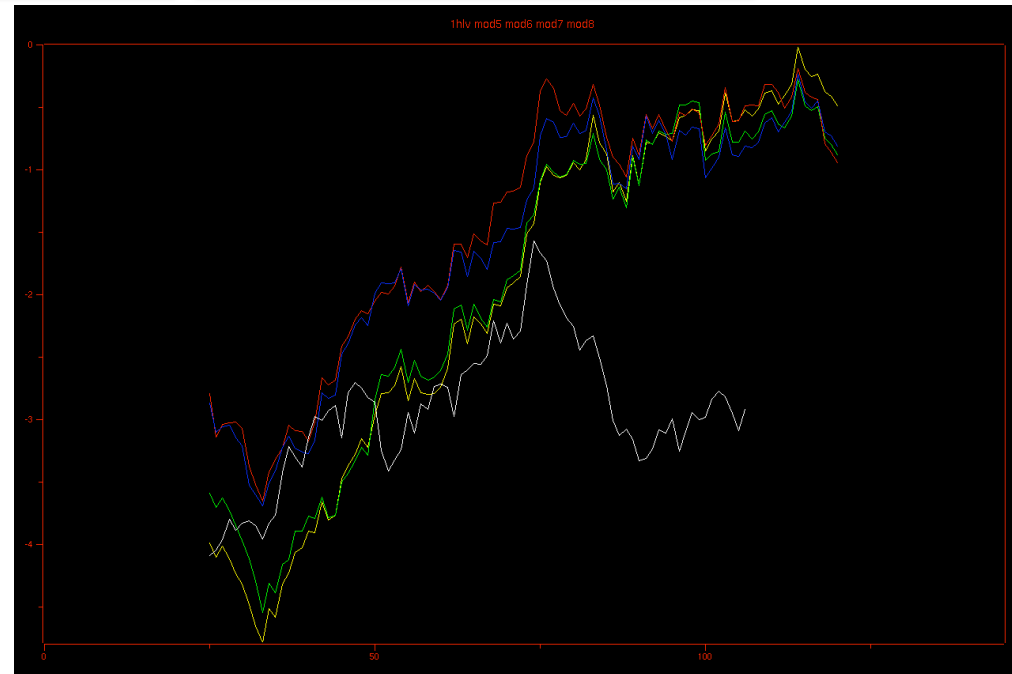
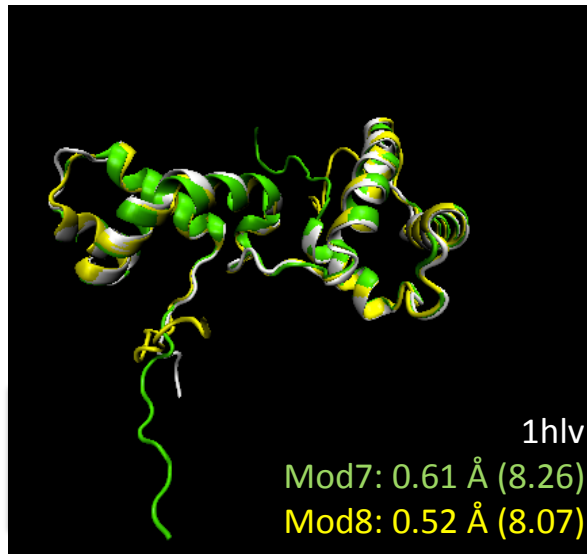
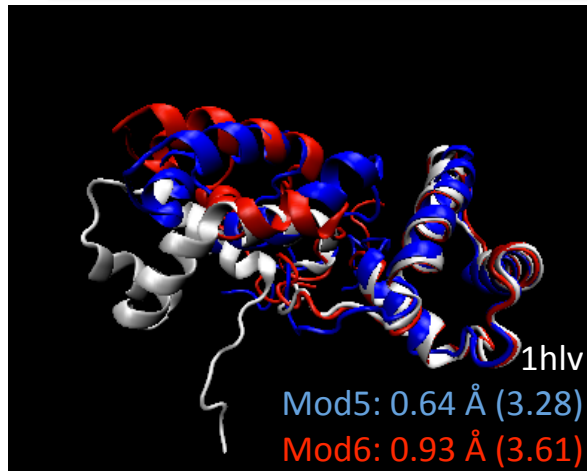
Modelling strategies

hmmalign CAG(1-140) +3 templates+CENP-B.hmm

Modeller:5/6

hmmalign CAG(1-140)+ 1hlv+CENP-B.hmm

Modeller: 7/8



	Z-comb	Z-pair	Z-surf
1hlv	-6.66	-6.34	-3.47
Mod5	-4.17	-4.17	-1.87
Mod6	-3.86	-4.25	-1.34
Mod7	-4.76	-4.81	-2.10
Mod8	-4.86	-4.75	-2.30

Prosa energy
profiles and
Z-scores

Modelling strategies

hmmalign CAG(1-225) + 1hlv + CENP-B.hmm
hmmalign CAG(1-225) + 1hlv + Tc5.hmm
clustalw CAG(1-225) + 1hlv

secondary structure calculation of 1hlv

Manual
modification
of the clustal
alignment

Modeller:
Frank1/2

```
1hlv      ----mg-PKRRQLTFREKSRIIQEVEENpdlRKGEIARRFNIPPSTLSTILKNKRAILAS
CAG_D1_1_64 mpsevrRKPRTSLTLEEKMEVIQSQERNK-LSVRDLAKRFNIGKTQAADILKHKQSIKEG

1hlv      erkygvastcrktnklspydkglliawfqqiraaglpvkgiilkekalriaeelgmdd
CAG_D1_1_64 llsge-----

1hlv      ftasngwldrfrrrrs
CAG_D1_1_64 -----
```

```
1hlv      mgpkrrqlt 1hlv      ----MG-PKRRQLTFREKSRIIQEVEENPDLRKGEIARRFNIPPSTLSTILKNKRAILAS
CAG_D2_81_137 ----- CAG      MPSEVRVKPRTSLTLEEKMEVIQSQERNK-LSVRDLAKRFNIGKTQAADILKHKQSIKEG

1hlv      vastcrktn 1hlv      ERKYGVASTCRKTNKLSP-YDKLEGLLLIAWFQQIRAAGLPVKGIIILKEKALRIA-EELGM
CAG_D2_81_137 ----- CAG      LLSGELKLNQMRNPLSQRGAQIDEMCFDWFSSRVRTENIPISGEMVRKKAKOLAVE-LGH

1hlv      NGWLDLRFRR 1hlv      DDFTASNGWLDLRFRRRS
CAG_D2_81_137 SGWLEKWRK CAG      SNFSTSSGWLEKWRKTHN
```

```
1hlv      ----MG-PKR
CAG      MPSEVRVKPRTSLTLEEKMEVIQSQERNK-LSVRDLAKRFNIGKTQAADILKHKQSIKEG
          . * ** * . * * * . . . . . * . . * . . * . .

1hlv      ERKYGVASTCRKTNKLSP-YDKLEGLLLIAWFQQIRAAGLPVKGIIILKEKALRIA-EELGM
CAG      LLSGELKLNQMRNPLSQRGAQIDEMCFDWFSSRVRTENIPISGEMVRKKAKOLAVE-LGH
          . . . * ** . . . . * * . . . * . . . * * * . .

1hlv      DFTASNGWLDLRFRRRS-----
CAG      NFSTSSGWLEKWRKTHNVRYNDTGDSLQLQEFAILVKSEPISENKDDCEPYPVTLIEPI
          * . . * * . . * . .

1hlv      -----
CAG      YSTEEAMQLARLKEFAKDDYASYQQLISLENQSWKWNIFKKELP
```

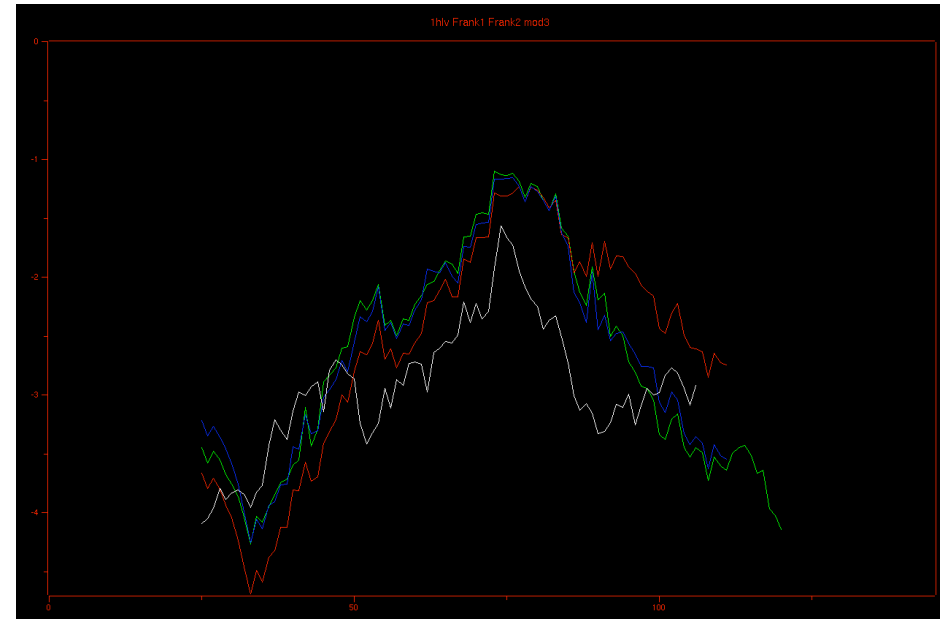
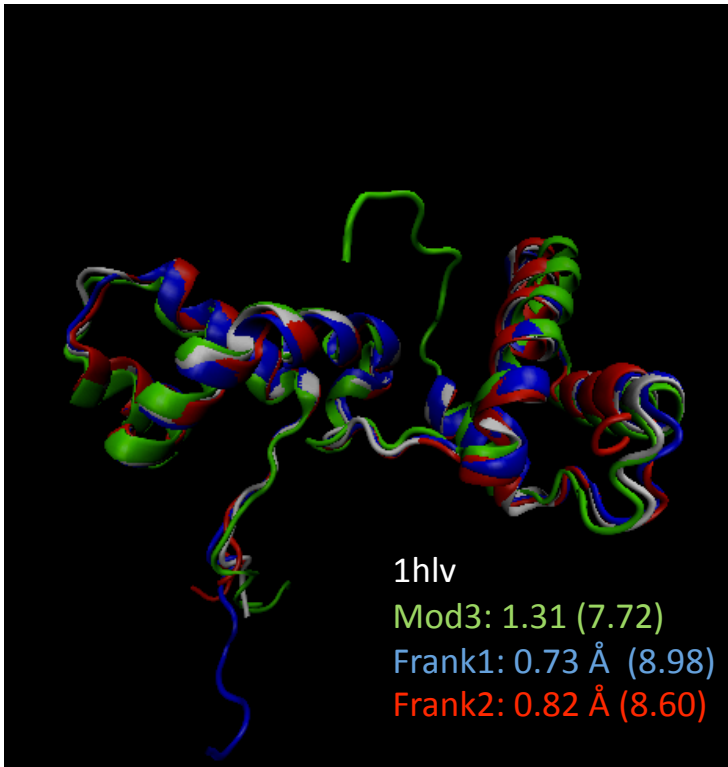
Modelling strategies

hmmalign CAG(1-225) + 1hlv + CENP-B.hmm
hmmalign CAG(1-225) + 1hlv + Tc5.hmm
clustalw CAG(1-225) + 1hlv

secondary structure calculation of 1hlv

Manual
modification
of the clustal
alignment

Modeller:
Frank1/2



Prosa energy
profiles and
Z-scores

	Z-comb	Z-pair	Z-surf
1hlv	-6.66	-6.34	-3.47
Frank1	-6.49	-5.10	-4.23
Frank2	-6.34	-4.99	-4.13
Mod3	-6.70	-5.40	-4.33

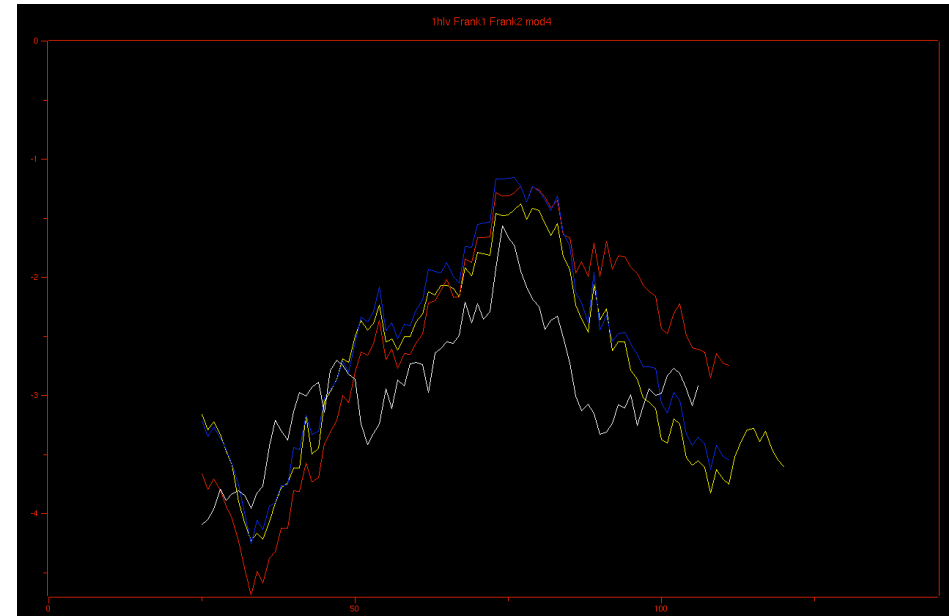
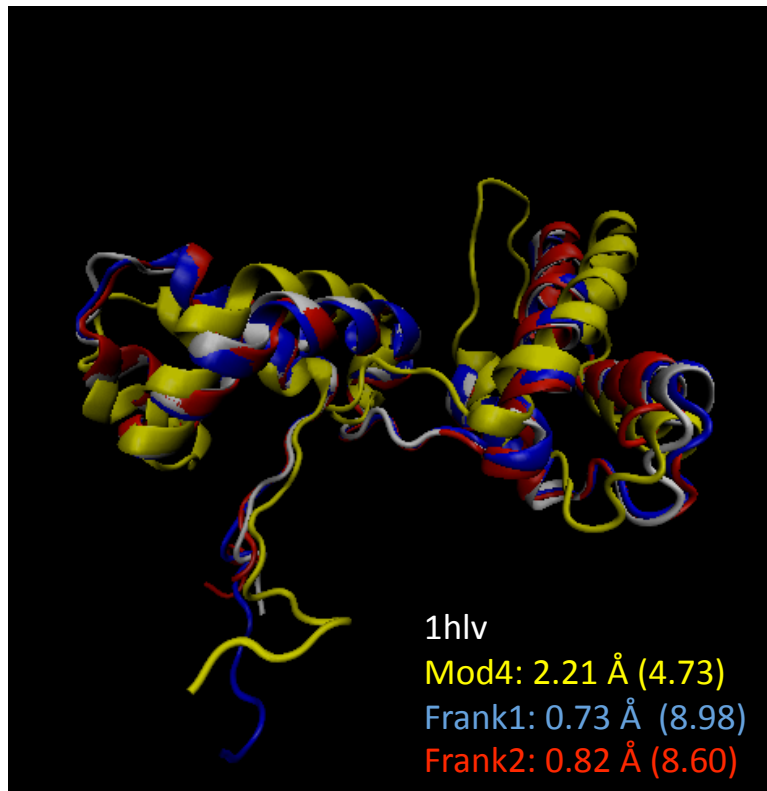
Modelling strategies

hmmalign CAG(1-225) + 1hlv + CENP-B.hmm
hmmalign CAG(1-225) + 1hlv + Tc5.hmm
clustalw CAG(1-225) + 1hlv

secondary structure calculation of 1hlv

Manual
modification
of the clustal
alignment

Modeller:
Frank1/2

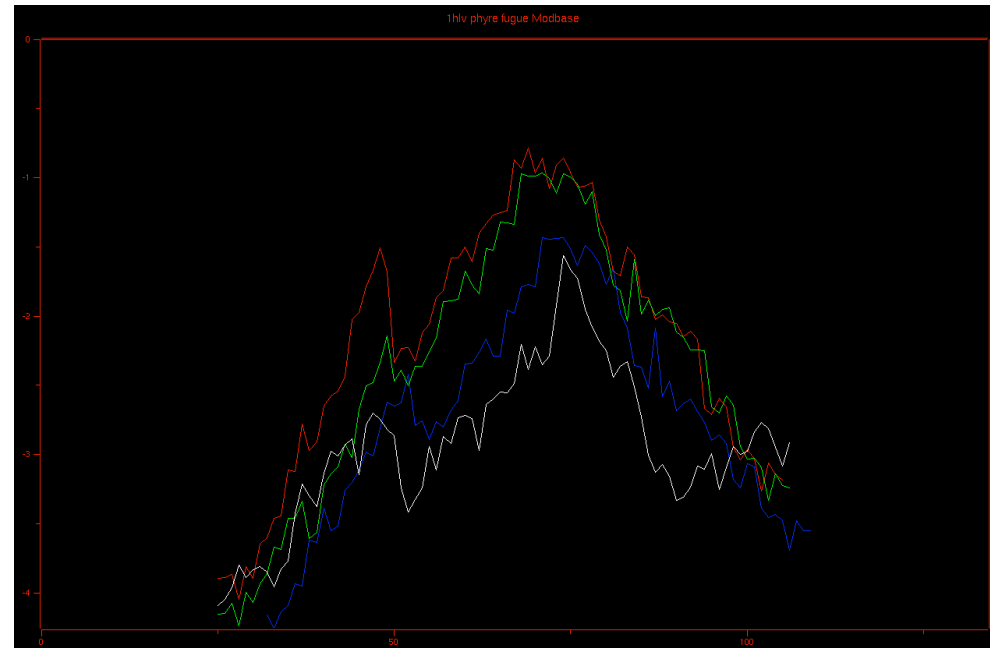
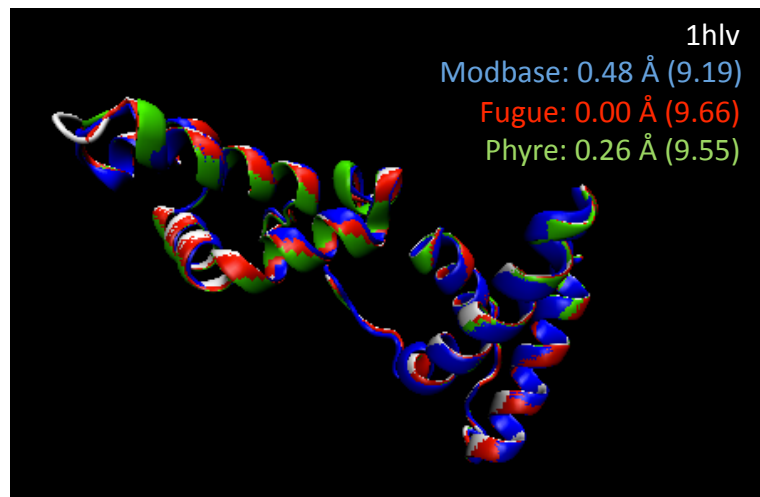


Prosa energy
profiles and
Z-scores

	Z-comb	Z-pair	Z-surf
1hlv	-6.66	-6.34	-3.47
Frank1	-6.49	-5.10	-4.23
Frank2	-6.34	-4.99	-4.13
Mod4	-6.44	-5.13	-4.22

Modelling strategies

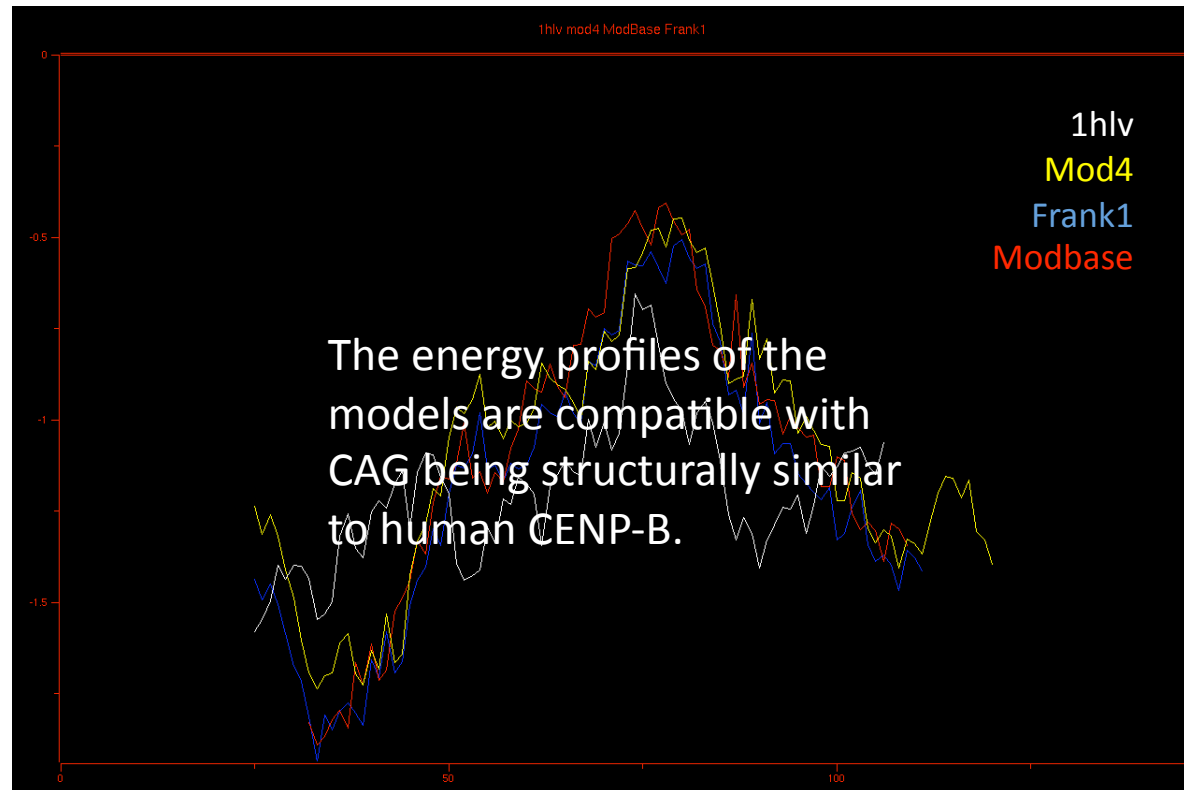
Servers (Fugue, Phyre) and Modbase models:



Prosa energy
profiles and
Z-scores

	Z-comb	Z-pair	Z-surf
1hlv	-6.66	-6.34	-3.47
Modbase	-6.89	-5.26	-4.47
Phyre	-6.45	-5.20	-4.03
Fuge	-6.20	-4.98	-3.90

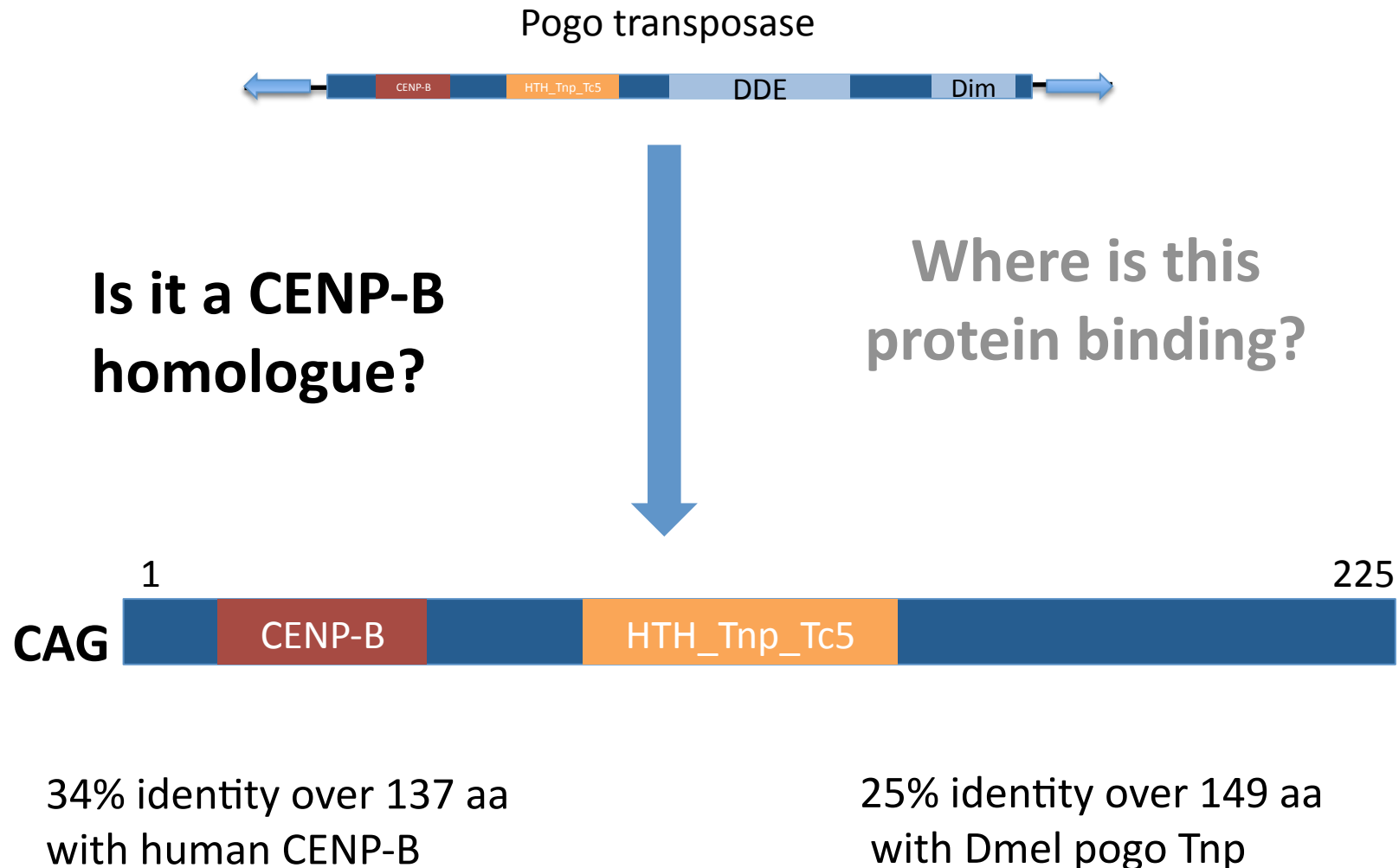
Final models



Procheck:

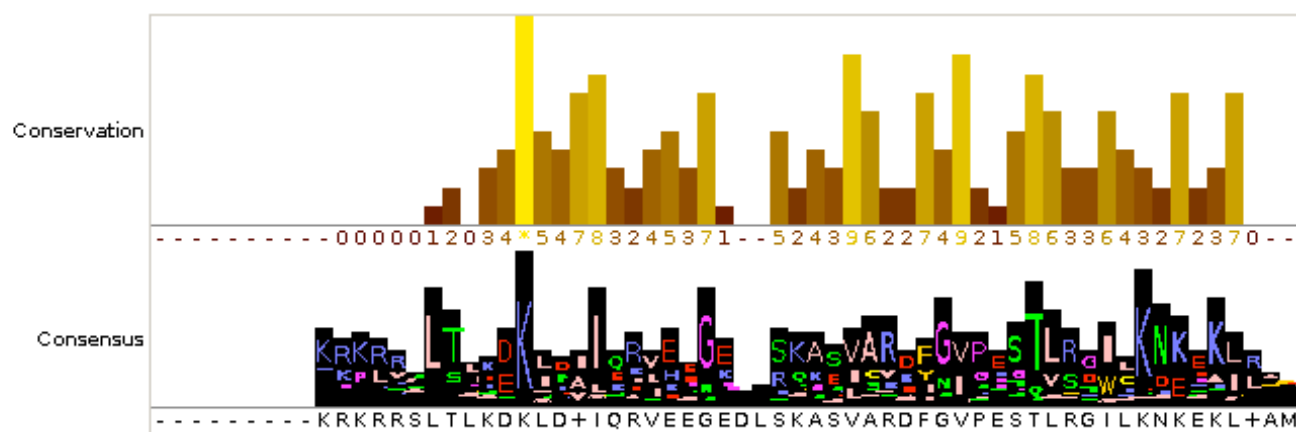
	CORE	ALLOW	GENER	DISALL
1hlv	91,3%	8,7%	0,0%	0,0%
Mod4	89.3%	9.9%	0.0%	0.8%
Frank1	91,9%	7,3%	0,0%	0,8%
ModBase	92,2%	6,0%	0,9%	0,9%

Introduction

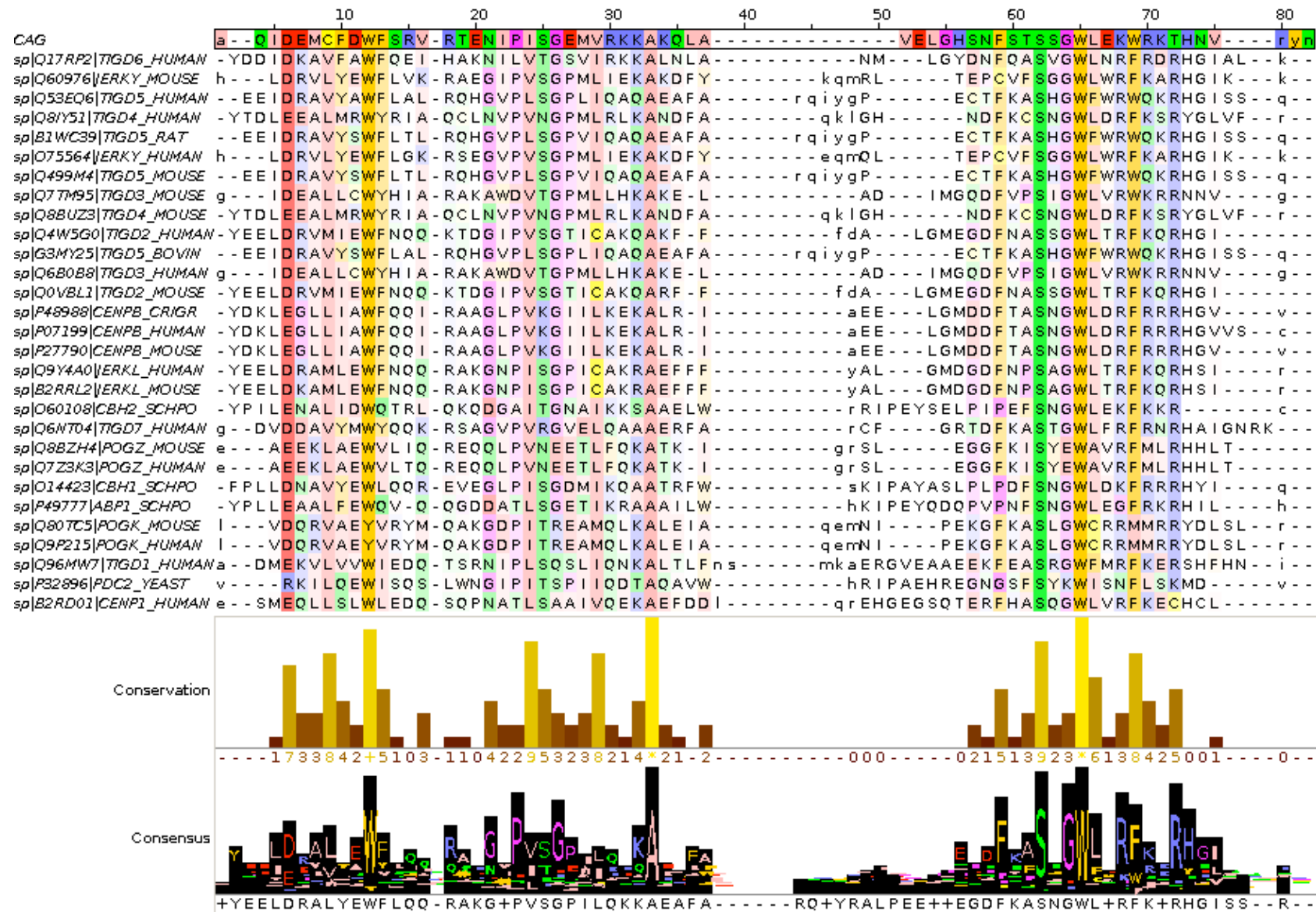


Multiple Sequence Alignment analysis

	10	20	30	40	50	60
CAG	---	mpserVRKPR	SLTLEEKMEV	IOSQERNK	-ISVRDLAKRFN	IGKTOAADILKHKQSIKEG
sp Q680B8 TIGD3_HUMAN	---	KKKLHALSLAEK	IQVLELLDES	K-mSQSEVARRR	FQVSQPOISRI	CKNKEKLLA-
sp P07199 CENPB_HUMAN	---	PKRRQLTFREK	SRIIQEVEEN	PdIRKGEIARR	FNIPPSTLST	ILKNKRAILA-
sp B2RD01 CENP1_HUMAN	---	KRERKAITLDL	KLEVLRREFE	AGE--KLSQIA	KALDLAISTV	ATIRDSKEKIK
sp Q88UZ3 TIGD4_MOUSE	---	KKKKSLSIEEK	IDINAVESG	K--KKAIEAAE	YGIKKNSLSS	IMKNKDKVLE-
sp Q53EQ6 TIGD5_HUMAN	---	---	KDKLQAIERV	KGGE--RQASV	CRDFGVPGG	TLRGWLKDEPKLR--
sp B2ARL2 VERKL_MOUSE	---	KRKRVLTLIKD	KLDIIKKLED	GG--SSKQLAVI	YGIGETT	VRDIRKNKEKI---
sp Q8IY51 TIGD4_HUMAN	---	KKKKSLSIEEK	IDINAVESG	K--KKAIEAAE	YGIKKNSLSS	IMKNKDKVLE-
sp Q9VBW9 DANR_DROME	---	KRPLRNLTPND	KVRAIQRIHNG	E--TKASVSRD	IGVPEST	LRGWCKNEQKLRFM
sp Q6NT04 TIGD7_HUMAN	---	ky---TTLNLEE	KMKVLSRIEAGR	--SLKSVMD	EFGISKSTFY	DIKKNKLLILD-
sp Q0VBL1 TIGD2_MOUSE	---	KRKRVLTLIKD	KLDIIKKLEEG	N--SFKKLSVLY	GIGESTVRD	IKKNKERI---
sp P27790 CENPB_MOUSE	---	PKRRQLTFREK	SRIIQEVEEN	PdIRKGEIARR	FNIPPSTLST	ILKNKRAILA-
sp Q60976 VERKY_MOUSE	---	KRKRVLTLKEK	IDICTRLERGE	--SRKALMQE	YNVGMSTLYD	IKAHKAQLLRF
sp Q29CW2 DAN_DROPS	---	KRPLRSLTPRD	KIHAIQRIHDGE	--SKASVARD	IGVPEST	LRGWCKNEQKLRFM
sp B0X560 DAN_CULQU	---	KRPLRHLTATD	KIDAIQRIHDGE	--SKASVARD	IGVPEST	LRGWCKNEEKLRY-
sp Q17RP2 TIGD6_HUMAN	---	KKRRQFSLEE	KMKVVGAVDSG	K--RKGDVAKE	FGITPSTLST	FLKDRTKF---
sp Q96MW7 TIGD1_HUMAN	---	RKSRTSLTLN	KLEMIKLSEEG	M--SKAEIGRR	LGLLRQT	VSQVVNAKEKFLKE
sp Q499M4 TIGD5_MOUSE	---	---	KDKLQAIERV	KGGE--RQASV	CRDFGVPGG	TLRGWLKDEPKLR--
sp P48988 CENPB_CRIGR	---	PKRRQLTFREK	SRIIQEVEEN	PdIRKGEIARR	FNIPPSTLST	ILKNKRAILA-
sp Q9VBW6 DAN_DROME	---	KRPLRSLTPRD	KIHAIQRIHDGE	--SKASVARD	IGVPEST	LRGWCKNEQKLRFM
sp Q4W5G0 TIGD2_HUMAN	---	KRKRVLTLIKD	KLDIIKKLEEG	I--SFKKLSV	VYGIGESTVRD	IKKNKERI---
sp B1WC39 TIGD5_RAT	---	---	KDKLQAIERV	KGGE--RQASV	CRDFGVPGG	TLRGWLKDEPKLR--
sp Q16IB4 DAN_AEDAE	---	KRPLRHLTATD	KIDAIQRIHDGE	--SKASVARD	IGVPEST	LRGWCKNEEKLRY-
sp Q7TM95 TIGD3_MOUSE	---	KKKLHALSLAEK	IQVLELLDES	K-mSQSEVARRR	FQVSQPOISRI	CKNKEKLLA-
sp G3MY25 TIGD5_BOVIN	---	---	KDKLQAIERV	KGGE--RQASV	CRDFGVPGG	TLRGWLKDEPKLR--
sp Q9Y4A0 VERKL_HUMAN	---	KRKRVLTLIKD	KLDIIKKLED	GG--SSKQLAVI	YGIGETT	VRDIRKNKEKI---
sp Q29CW0 DANR_DROPS	---	KRPLRNLTPND	KVRAIQRIHNG	E--TKASVSRD	IGVPEST	LRGWCKNEQKLRFM



Multiple Sequence Alignment analysis



Multiple Sequence Alignment analysis

Pogo transposase



Is it a **CENP-B** homologue?

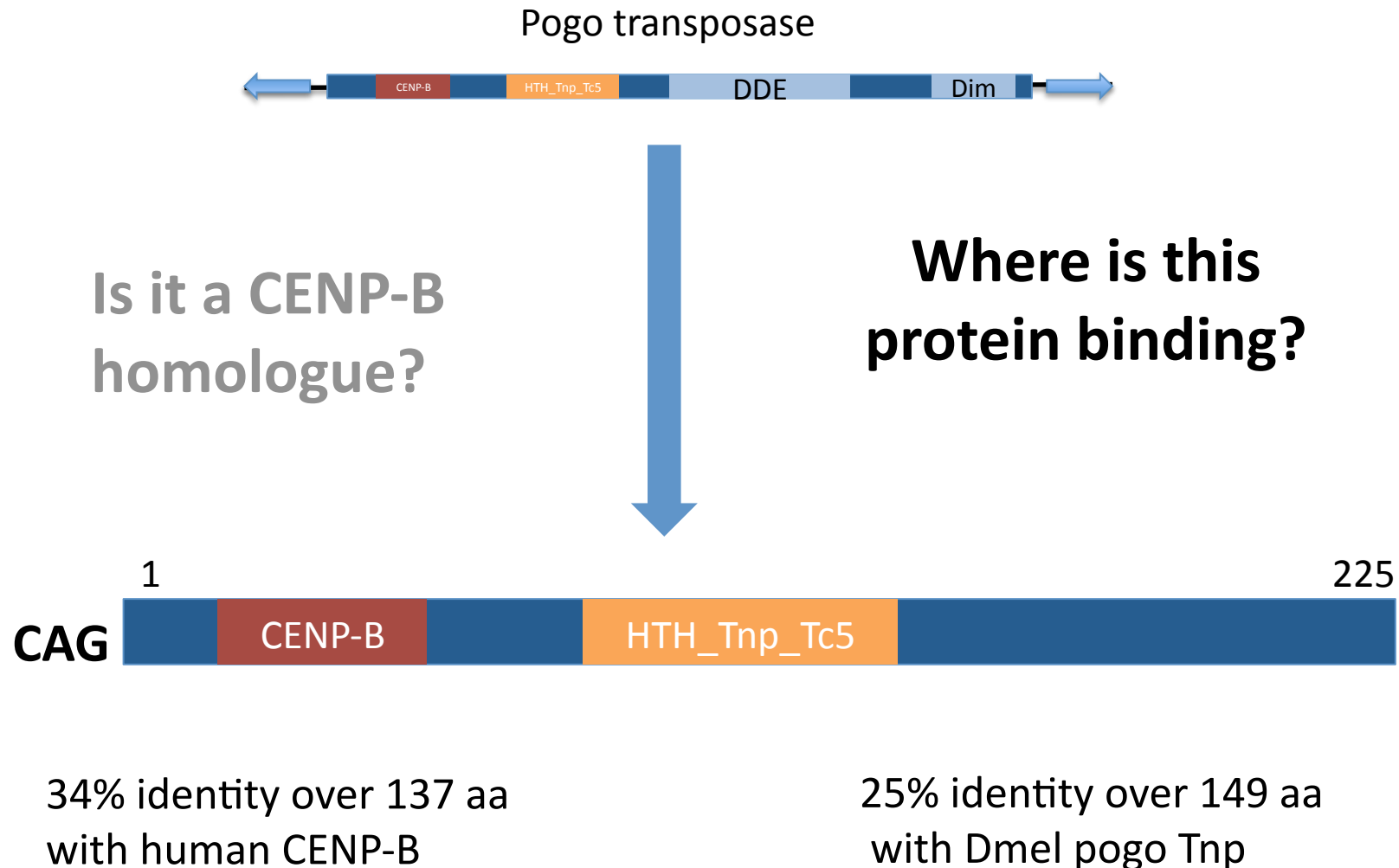
Where is this protein binding?



34% identity over 137 aa
with human CENP-B

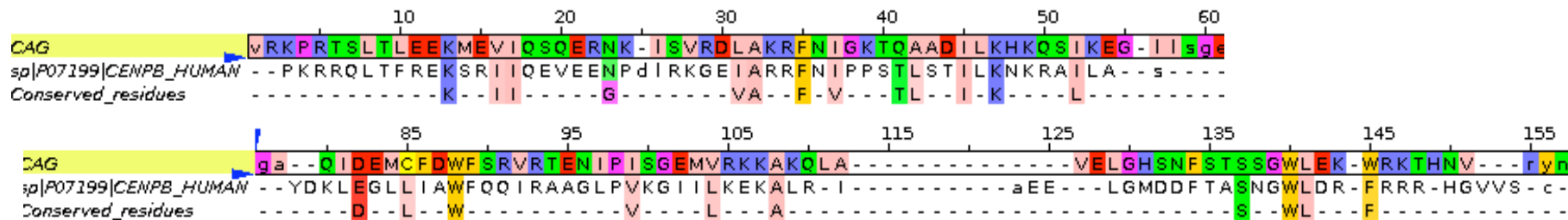
25% identity over 149 aa
with Dmel pogo Tnp

Multiple Sequence Alignment analysis

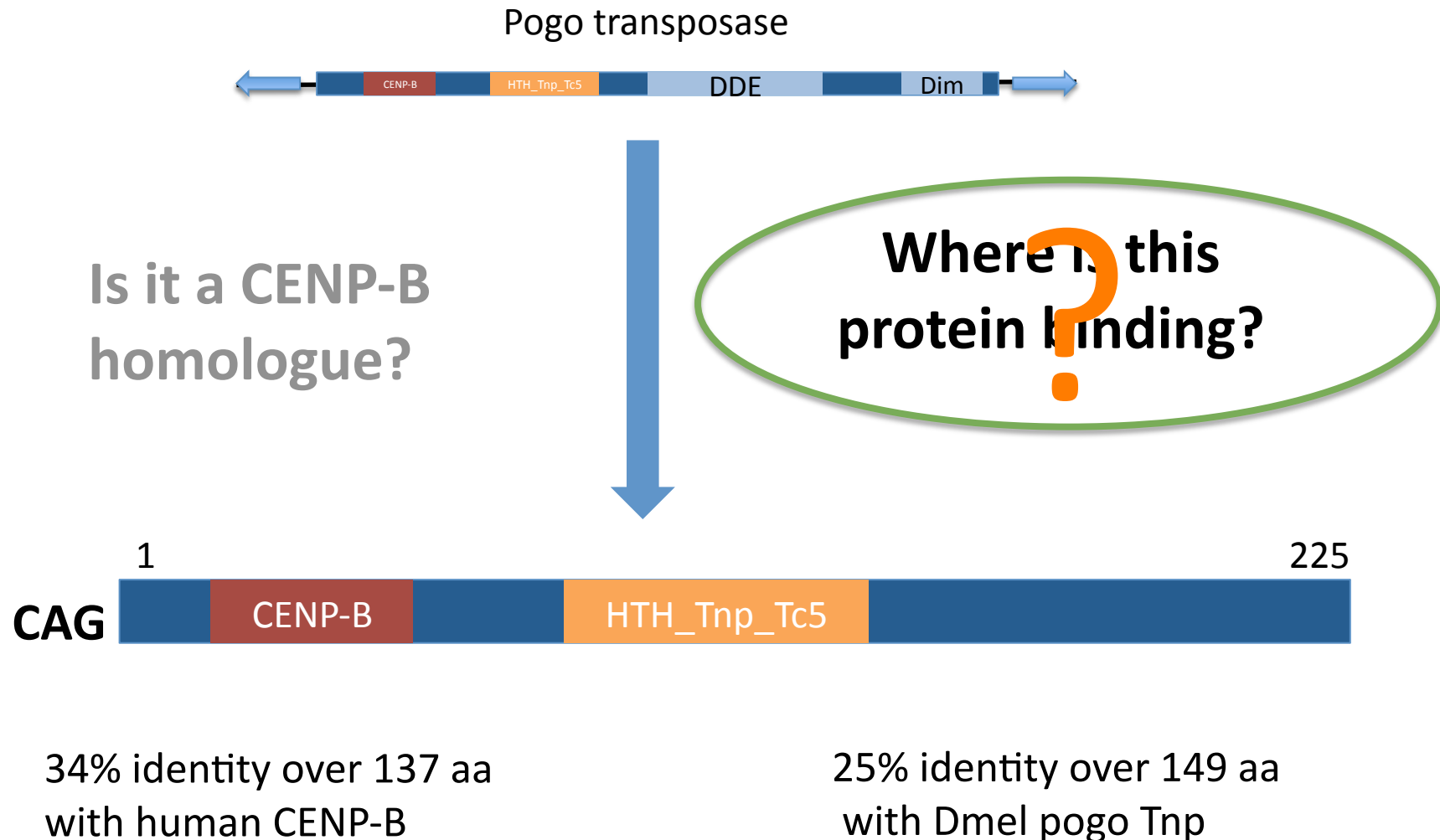


Multiple Sequence Alignment analysis

CAG shares many conserved residues with human CENP-B and other members of the family.



Multiple Sequence Alignment analysis



Conclusions

- Selective pressure imposes functional and structural constraints on DNA binding domains of TFs that are reflected at sequence level.
- Structural alignments may be limited due to lack of data. Sequence alignments may introduce noise. The combination of both sequence and structural alignments is very powerful for unravelling functionally important residues.
- In TFs, few residues determine DNA sequence specificities and small changes can greatly modify the binding site. Interaction with other proteins give TFs the high specificity that is observed *in vivo*.
- The DNA binding domain of TFs can be derived from other DNA binding proteins, especially from transposable elements (exaptation).
- The sequence and the predicted structural conservation in CAG suggest that CAG is a putative CENP-B homologue in *Drosophila*.

References

- Brownlie P, Ceska T, Lamers M, Romier C, Stier G, Teo H, et al. The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* 1997;5(4):509-520.
- Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 1993;363(6424):38-45.
- Garvie CW, Wolberger C. Recognition of Specific DNA Sequences. *Mol Cell* 2001;8:937-946.
- Huang J, Weintraub H, Kedes L. Intramolecular regulation of MyoD activation domain conformation and function. *Mol Cell Biol* 1998;18(9):5478-5484.
- Longo A, Guanga GP, Rose RB. Crystal structure of E47-NeuroD1/beta2 bHLH domain-DNA complex: heterodimer selectivity and DNA recognition. *Biochemistry* 2008;47(1):218-229.
- 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):451-459.
- Parraga a, Bellolell L, Ferré-D'Amaré a R, Burley SK. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. *Structure (London, England : 1993)* 1998;6(5):661-672.
- Shimizu T, Toumoto a, Ihara K, Shimizu M, Kyogoku Y, Ogawa N, et al. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. *EMBO J* 1997;16(15):4689-4697.
- Skinner MK, Rawls A, Wilson-rawls J, Roalson EH. Basic Helix-Loop-Helix Transcription Factor Gene Family Phylogenetics and Nomenclature. *Differentiation* 2010;80(1):1-8.
- Nekaludova L, Pabo CO. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc Natl Acad Sci.* 1994; 61: 69948-6952.
- Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO. Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. 1996; 4: 1171-1180.
- Pavletich N, Pabo CO. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science.* 1991; 252: 809-817.
- Matthews JM, Sunde M. Zinc Fingers-Folds for Manu Occasions. *Life.* 2002; 54: 351-355.
- Wolfe SA, Nekludova L, Pabo CO. DNA recognition by Cys2His2 zinc finger protein. *Annu Rev Biophys Biomol Struct.* 1999; 3:183-212.
- Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers. *Nucleic Acids Research.* 2003; 31 (2):532-550.
- Wolfe SA, Grant RA, Elrod-Erickson M, Pabo CO. Beyond the "Recognition code": structures of two Cys2His2 finger/TATA box complexes. *Structure.* 2001; 9:717-723.

References

- Glover JN, Harrison SC. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* 1995;373(6511):257-261.
- Fujii Y, Shimizu T, Toda T, Yanagida M, Hakoshima T. Structural basis for the diversity of DNA recognition by bZIP transcription factors. *Nat Struct Biol.* 2000;7(10):889-893.
- Moitra J, Szilák L, Krylov D, Vinson C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* 1997;36(41):12567-12573.
- O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. *Science* 1989;243(4890):538-542.
- Ramos J, Lazaridis T. Energetic determinants of oligomeric state specificity in coiled coils. *J Am Chem Soc* 2006;128(48):15499-15510.
- Schumacher Ma, Goodman RH, Brennan RG. The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J Biol Chem* 2000;275(45):35242-35247.
- Vinson C, Myakishev M, Acharya A, Mir AA, Moll JR, Bonovich M. Classification of Human B-ZIP Proteins Based on Dimerization Properties. *Mol Cell Biol* 2002;22(18):6321-6335.
- Langosch D, Heringa J. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins* 1998;31(2):150-159.
- Krylov D, Olive M, Vinson C. Extending dimerization interfaces: the bZIP basic region can form a coiled coil. *EMBO J* 1995;14(21):5329-5337.
- Mason JM, Arndt KM. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem* 2004;5(2):170-176.
- Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem* 2003;278(17):15178-15184.
- Krylov D, Mikhailenko I and Vinson C. A thermodynamical scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions. *EMBO J* 1994; 13(12):2849-2861.
- Vinson CR, Hai T, Boyd SM. Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.* 1993; 7: 1047-1058.
- Sinezelle L, Izsvák and Ivics Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* 2009; 66:1073-1093.
- Tanaka Y et al. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *The EMBO Journal.* 2001; 20(23):6612-6618.
- Fraenkel E, Pabo CO. Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex. *Nat Struct Biol.* 1998 Aug;5(8):692-7.

Questions

Which of the following options is describing a characteristic of basic leucine zippers transcription factor?

- a) They form dimers to bind into the DNA major groove
- b) A leucine is repeated through the leucine zipper region every seven aminoacids
- c) **The two above are correct**
- d) The DNA binding region is highly hydrophobic.
- e) All are correct

What is true about the fold of the leucine zippers transcription factors?

- a) Leucine zippers dimerize as coiled-coils.
- b) The hydrophobic core packing is described as “knobs-into-holes”.
- c) **The two above are correct.**
- d) The two helices of the dimer packs in an antiparallel manner.
- e) All are correct.

Questions

Which one is NOT correct about the basic helix loop helix transcription factors:

1. They contact the DNA bases through residues in loops.
 2. They contact the DNA bases through residues in conserved beta strands.
 3. They contact the DNA bases only through water molecules, which is characteristic of this family.
 4. They contact the DNA bases through residues in alpha helices.
- a) 1 and 3
b) 1, 2 and 3
c) 2 and 4
d) 4
e) 1,2,3,4

The basic helix loop helix transcription factors...

1. Have a basic region that makes bonds with the DNA.
 2. Have a basic region that is involved in dimerization.
 3. Have a basic region with highly conserved residues among the members of the family.
 4. Have a loop conserved in structure that is important for sequence-specific DNA recognition.
- a) **1 and 3**
b) 1, 2 and 3
c) 2 and 4
d) 4
e) 1,2,3,4

PEM

Which is/are the conserved residue/es in zinc fingers that are coordinating the zinc atom?

- A. Arg
- B. Arg and Leu
- C. His
- D. Cys
- E. Options C and D**

Relating to zn fingers, ¿which is the correct fold of Cys2His2 family?

- A. Four beta-strands connected by a loop and followed by two helix.
- B. A hairpin followed by an alfa-helix.**
- C. Two helix separated by a short loop.
- D. Cystein rich loops.
- E. Two beta-hairpins connected by a short loop.

PEM

Chose the correct option about homeodomain transcription factors:

- a) The residues of the hydrophobic core of homeodomain establish direct contacts with the DNA backbone.
- b) Homeodomain proteins use a beta-alpha-beta motif for DNA binding.
- c) The first two statements are both true.
- d) In vivo specificity of homeodomain TFs is increased by the interaction with other proteins.**
- e) All above statments are false.

Chose the correct option about homeodomain transcription factors:

- a) A single aminoacid substitution in the central part of the recognition helix can completely modify the target DNA sequence.
- b) Water molecules can participate in short-lived interactions between the recognition helix and DNA.
- c) The first two statments are correct.**
- d) The hydrophobic core of the homeodomain fold is rich in Arginines and Lysines.
- e) All these statments are correct.

PEM

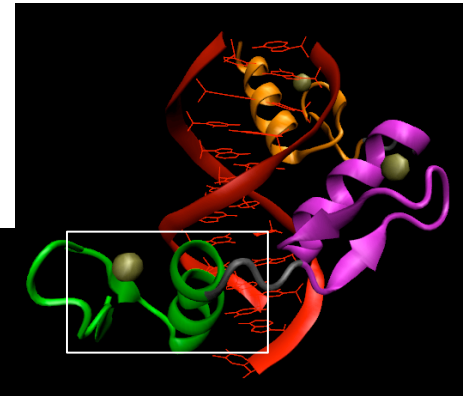
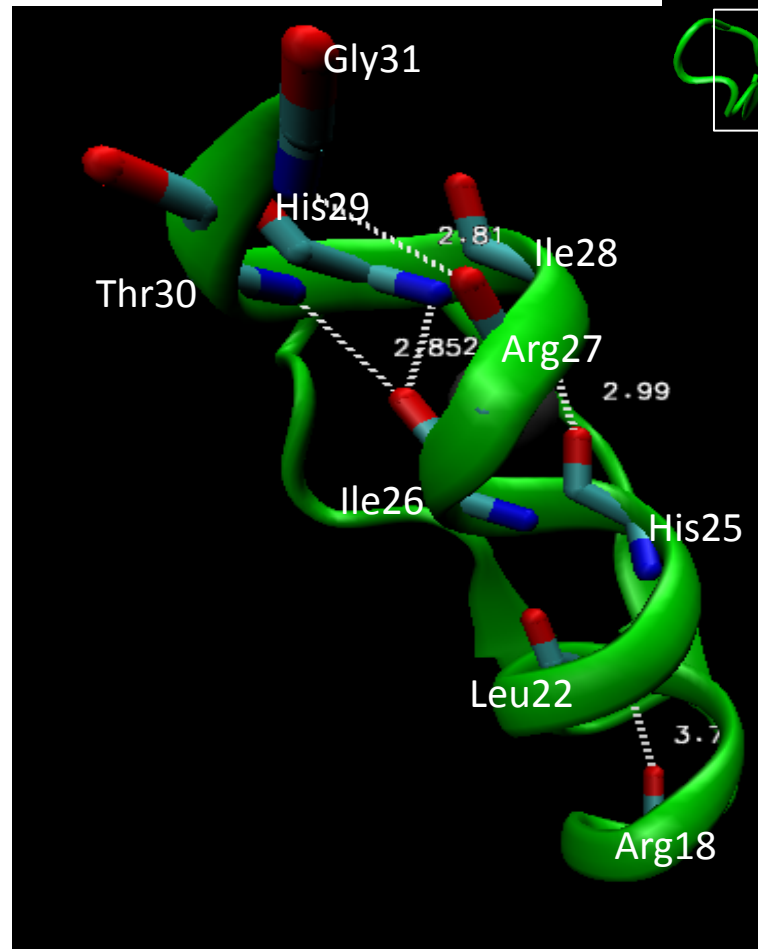
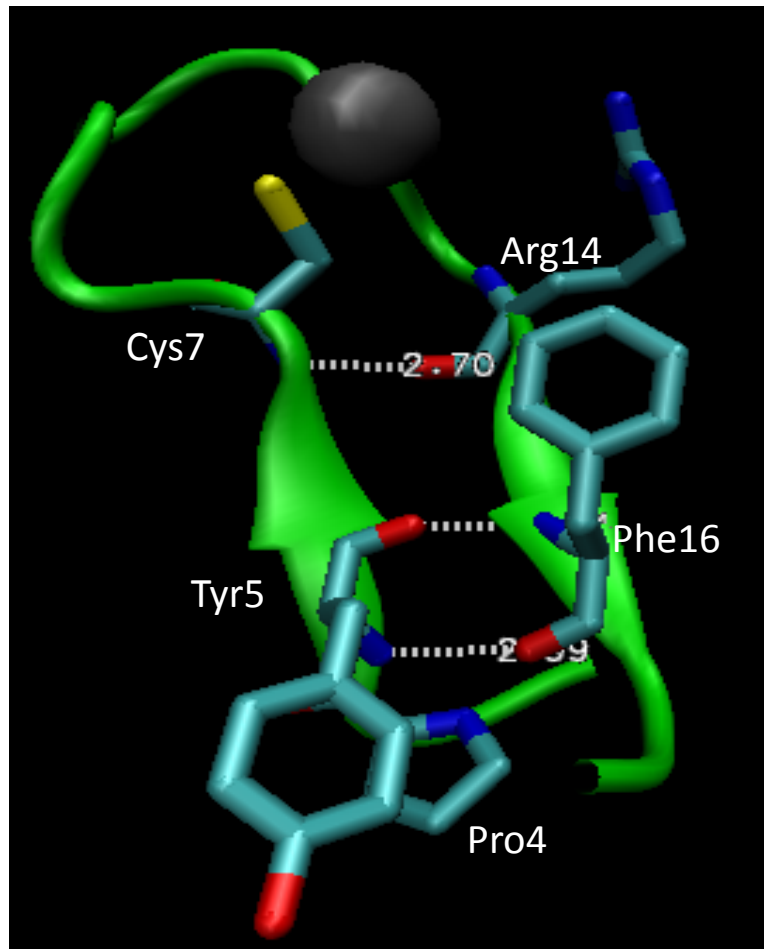
Chose the correct option about molecular domestication:

- a) Molecular domestication consists on the industrial application of biomolecules derived from transposable elements to develop or make useful products.
- b) The DNA binding domain of some human proteins derives from transposable element encoded genes.**
- c) The two first statements are both false.
- d) Transposable elements are selfish molecular parasites that only have detrimental consequences on the host genome.
- e) All statments above are true.

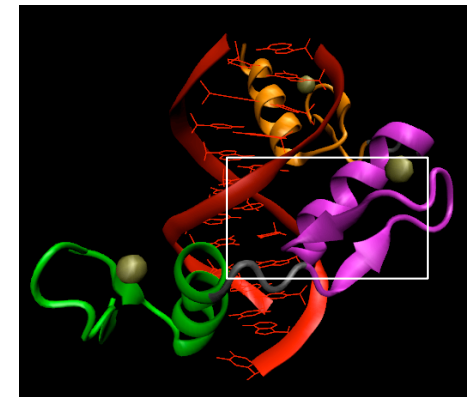
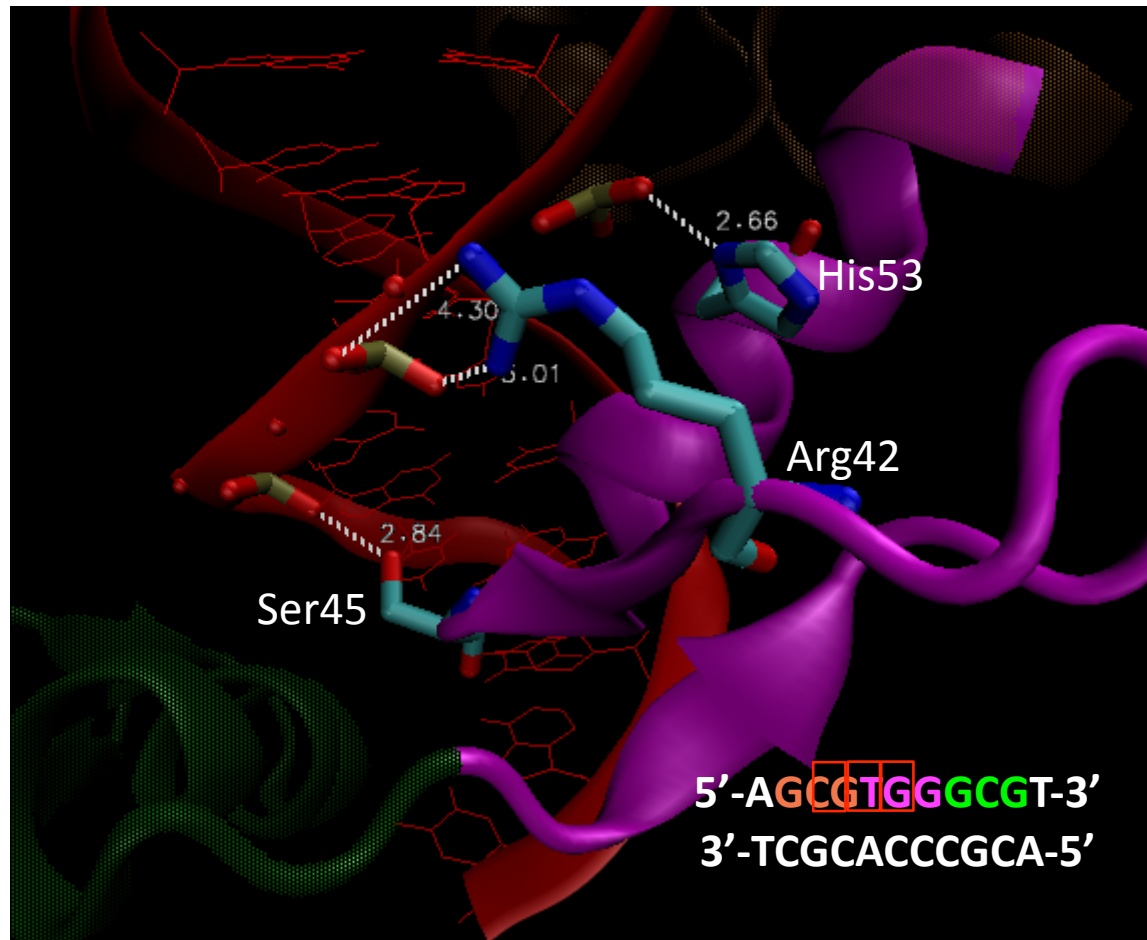
In the evaluation of CAG model...

- a) z-score tells you which model is energetically more stable.
- b) The best models were obtained with a manually curated alignment.**
- c) The two first statments are both false.
- d) It is completely impossible to do homology modelling when the identity is lower than 70% .
- e) All statements above are false.

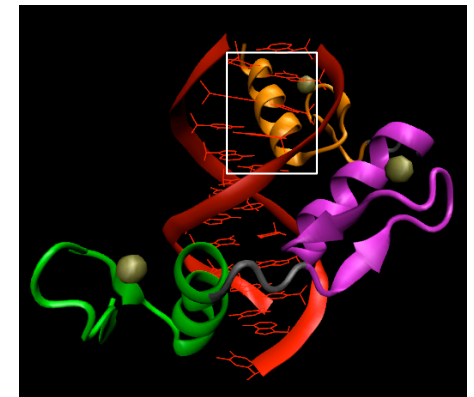
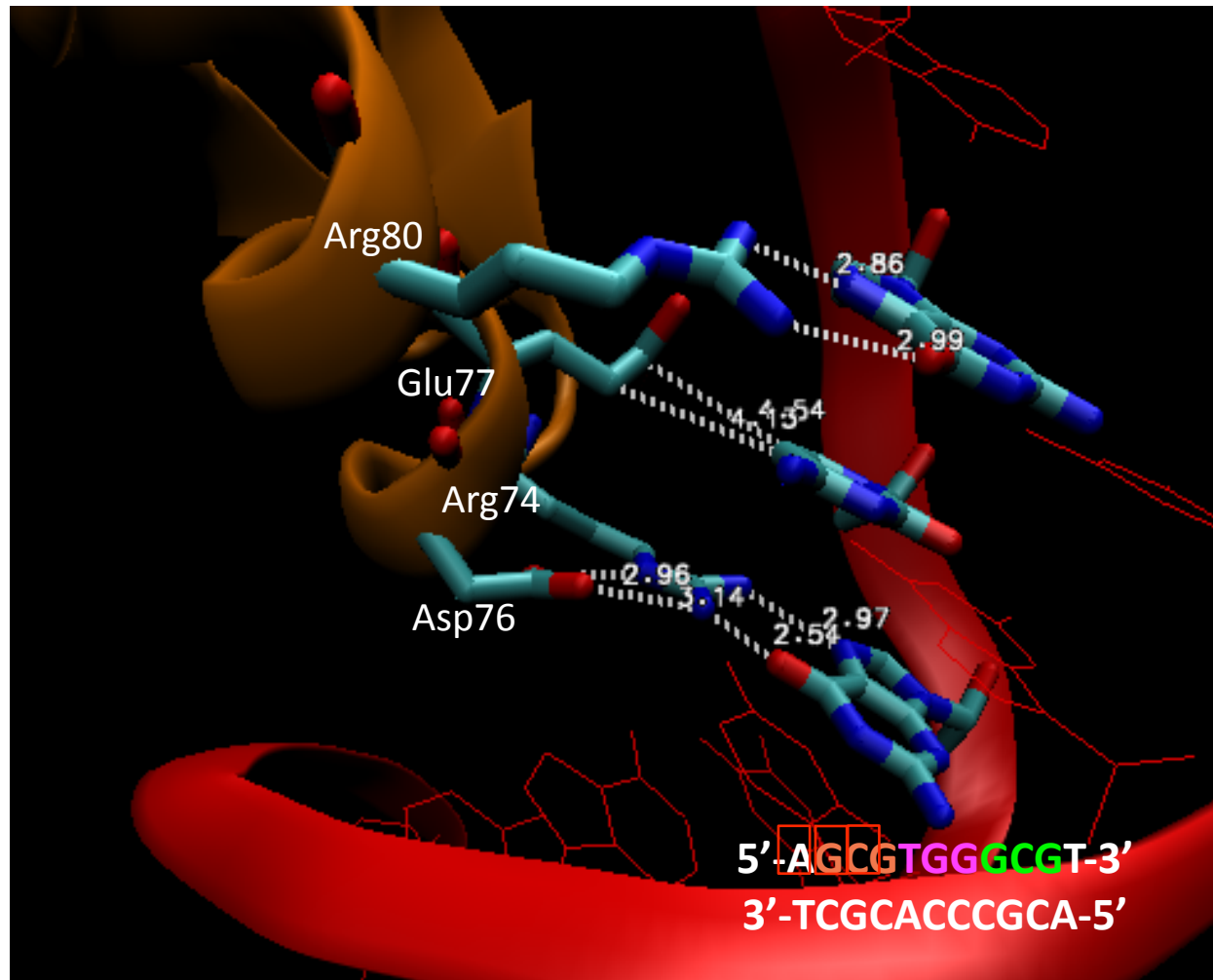
Zif268: additional information



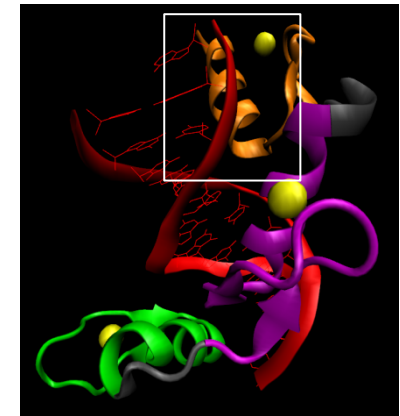
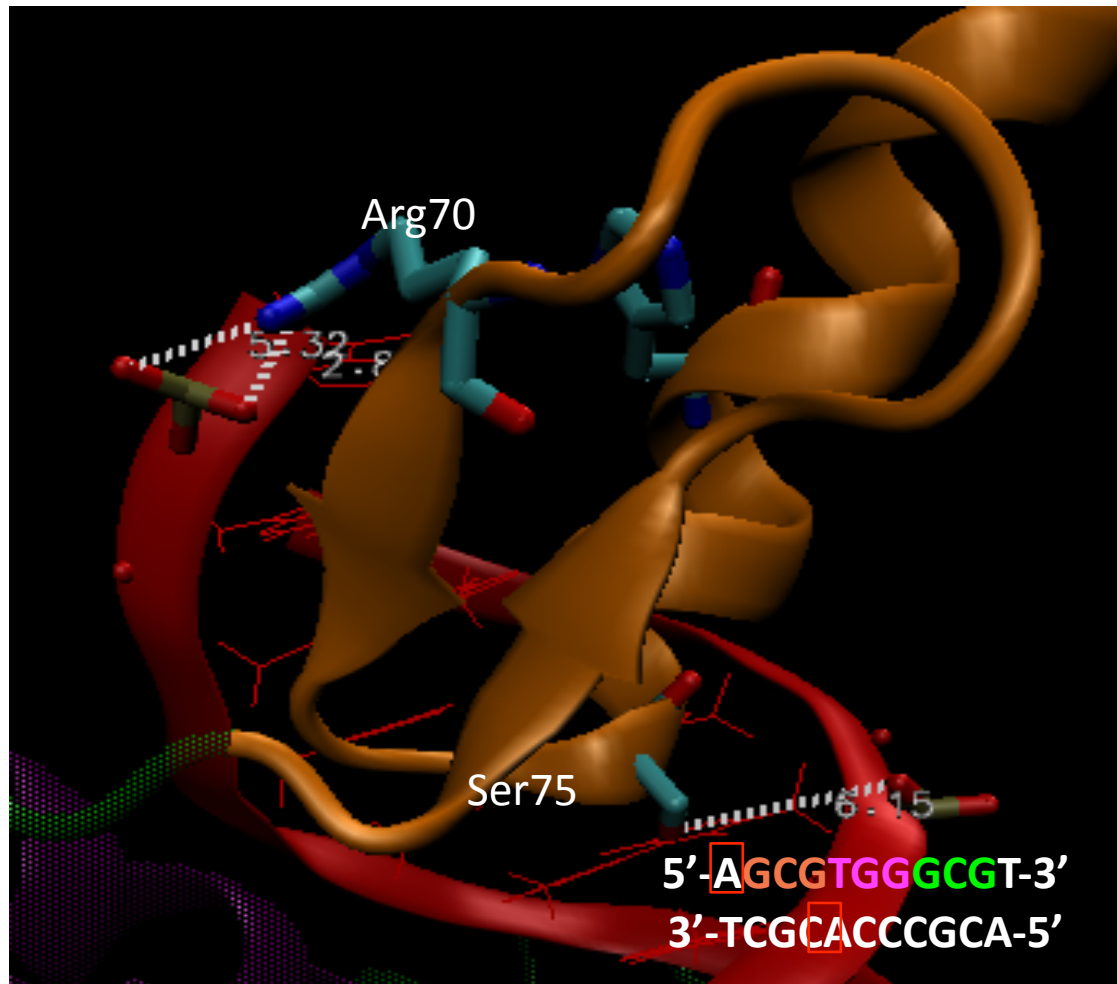
Zif268: additional information



Zif268: additional information

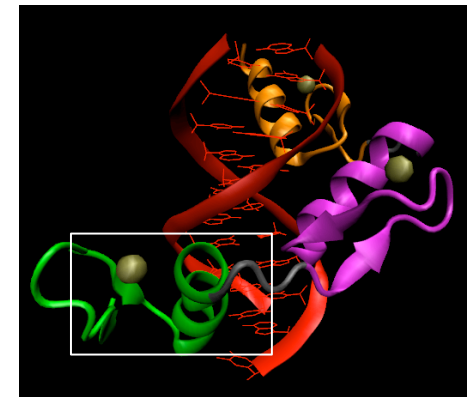
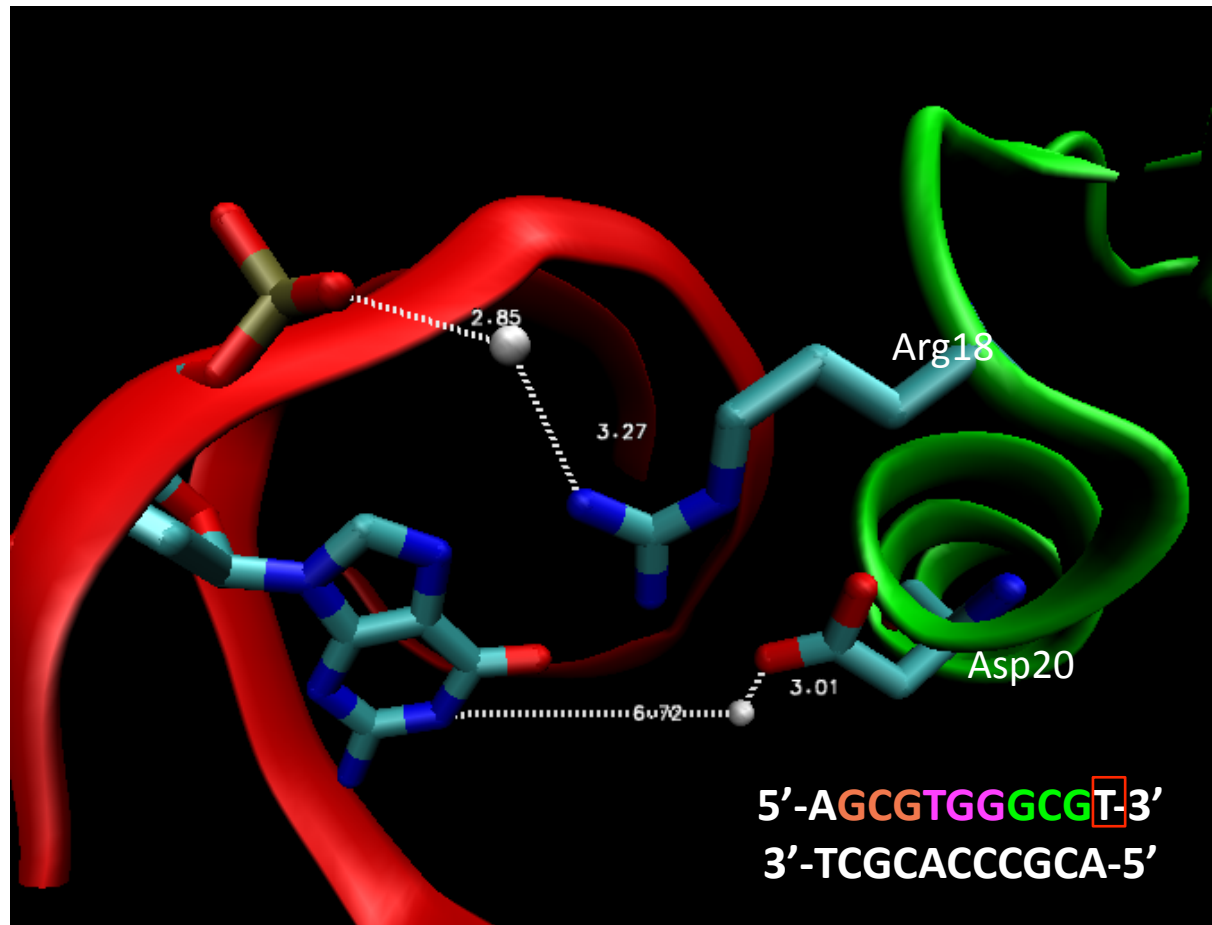


Zif268: additional information



Zif268: additional information

Water-mediated contact help to stabilize the binds to the DNA backbone and to the bases



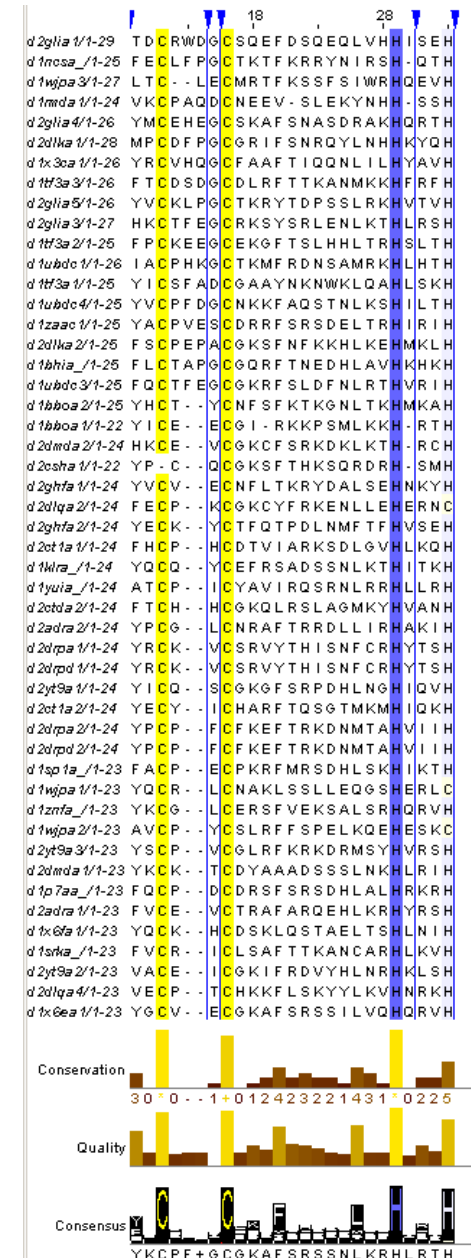
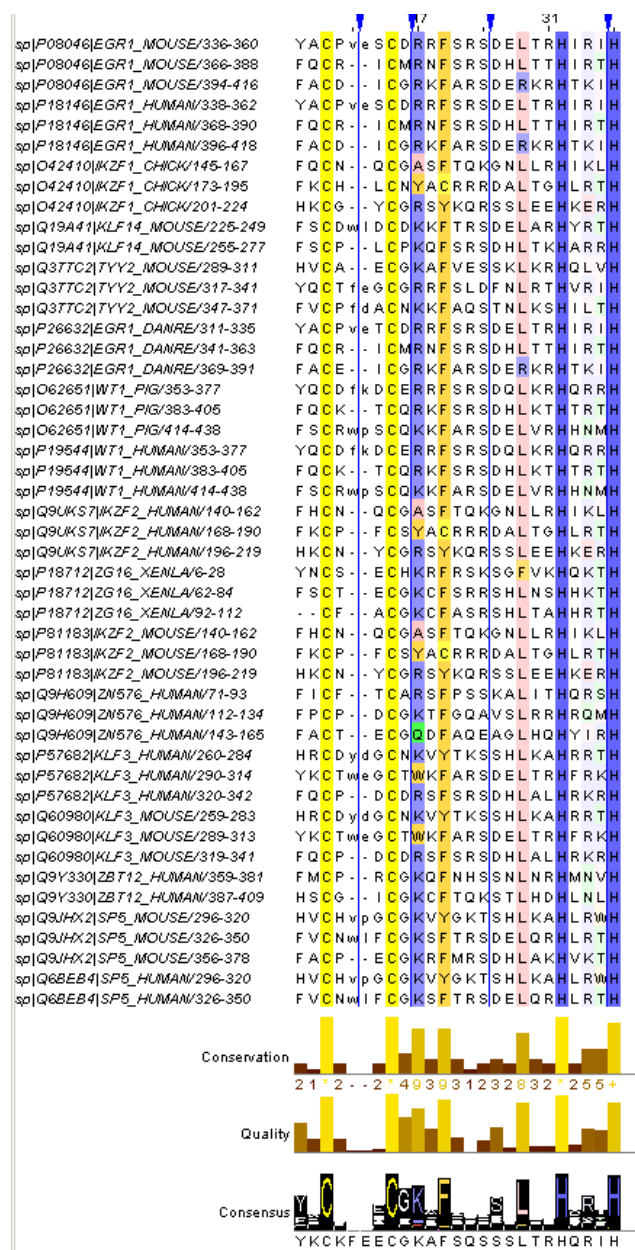
Zif268: additional information

Table 1. Parameters for DNA structures

DNA segment	Pitch, Å	Rise per bp	δ , no.			Groove width, Å		Incl, deg	Disp, Å	Rp, Å	D, Å
			A	B	Int	minor	major				
1. A-DNA (13)	34.6	2.88	15	0	1	10.1	8.2	13.6	-4.8	9.68	-3.76
2. A-DNA (14)	33.5	2.94	15	1	0	10.4	6.9	14.0	-4.1	9.36	-3.82
3. GLI (5)	36.0	3.07	8	8	2	7.4	12.5	9.3	-3.1	10.5	0.90
4. trp repressor (7)	33.1	3.13	2	13	3	7.0	10.6	9.2	-2.3	10.0	1.04
5. Glucocorticoid 1 (8)	35.0	3.24	3	11	4	7.4	11.4	10.5	-1.8	10.0	1.77
6. Zif268 (9)	36.5	3.25	0	15	5	8.2	11.5	9.5	-1.6	9.95	1.37
7. MetJ (10)	35.8	3.32	0	15	3	7.8	11.3	11.9	-1.5	9.8	1.90
8. Engrailed 1 (11)	33.6	3.19	2	15	5	7.5	10.2	10.8	-1.5	9.6	1.47
9. TTK (12)	33.9	3.23	0	12	5	6.8	11.2	8.7	-1.5	9.75	2.00
10. α 2 (15)	34.6	3.33	19	16	5	6.6	11.7	7.3	-1.0	9.7	2.60
11. GCN4 (16)	34.0	3.28	0	34	0	6.3	11.7	4.5	-0.9	9.8	2.52
12. λ repressor 1 (17)	35.5	3.33	1	17	2	6.7	12.1	3.9	-0.9	9.7	2.31
13. 434 repressor 1 (18)	35.5	3.33	0	14	4	5.3	13.4	4.1	-0.6	9.7	3.89
14. Glucocorticoid 2 (8)	34.6	3.35	4	10	4	5.1	13.1	0.2	-0.5	9.6	3.47
15. 434 repressor 2 (18)	32.7	3.17	0	20	0	5.9	11.1	4.8	-0.4	9.7	3.05
16. λ repressor 2 (17)	35.0	3.35	0	18	2	7.1	11.4	6.7	-0.3	9.6	2.77
17. GAL4 (19)	34.1	3.30	0	20	0	5.9	11.8	5.1	0.0	9.3	3.59
18. Engrailed 2 (11)	36.9	3.54	1	15	4	6.0	13.2	1.8	-0.1	9.25	3.44
19. B-DNA (20)	33.2	3.31	3	16	5	4.8	12.3	-0.2	-0.1	9.3	3.58

Structural parameters for selected DNA duplexes. Lengths are given in ångströms; angles are in degrees. Under δ torsion angles, we tabulate the number of angles of B-type ($118^\circ < \delta < 180^\circ$) and the number of angles of A-type ($50^\circ < \delta < 95^\circ$) (6). The number of δ angles between 95° and 118° is tabulated in the Int (intermediate) column. We compute major and minor groove widths by calculating distances between two best-fit helices through the phosphate atoms of the two DNA strands and then subtracting 5.8 Å to account for the van der Waals radii of the phosphate groups. For a DNA fragment making a nearly full helical turn, these groove widths agree well with the NEWHELIX values [for shorter segments, the standard definition of the groove widths (6) used by NEWHELIX is not applicable]. Inclination (Incl), computed for the base-pair centerline of C6–C8 with the NEWHELIX program, is approximately equal to the base-pair-plane inclination adopted by the Cambridge convention. Disp is displacement of the base-pair centerline (C6–C8) from the helical axis. Rp is the radius of the best-fit cylinder through all the phosphates. The parameter D is defined in Fig. 1. In our analysis, we split the DNA into shorter helical segments whenever the rms deviation of P atoms from the best-fit cylindrical surface exceeded 1 Å. When segmentation was necessary, we always attempted to use DNA segments that were contacted by individual domains or subunits of the protein. DNA segments used in this table had the following sequences (reading along one strand). Segments: 1, GGGATCCC; 2, GGGGCCCC; 3, TTGGGTGGT (bp 7–15); 4, GTACTAGTT (mol.2, bp 1–9); 5, GATGTTCTG (bp 10–18); 6, GCGTGGGCGT; 7, TAGACGTCT (bp 1–9); 8, TAATTACCTAA (bp 10–20); 9, TAAGGATA (bp 5–12); 10, CATGTAATTCATT-TACACGC; 11, TCCTATGACTCATCCAG (bp 1–17); 12, GCGGTGATAT (bp 10–19); 13, AGTACAAAC (bp 1–9); 14, CAGAACATC (bp 1–9); 15, TTTCTTGTAT (bp 10–19); 16, ATACCACTGG (bp 1–10); 17, CCGGAGGACA (bp 1–10); 18, TTTGCCATGT (bp 1–10); 19, CGCGAA-TTCGCG. We list parameters for only one half-site of the trp- and GAL4-DNA complexes, because other half-sites are structurally similar. Tramtrack (TTK) parameters were averaged for the two similar copies of the complex present in the crystal (12).

Zif268: additional information



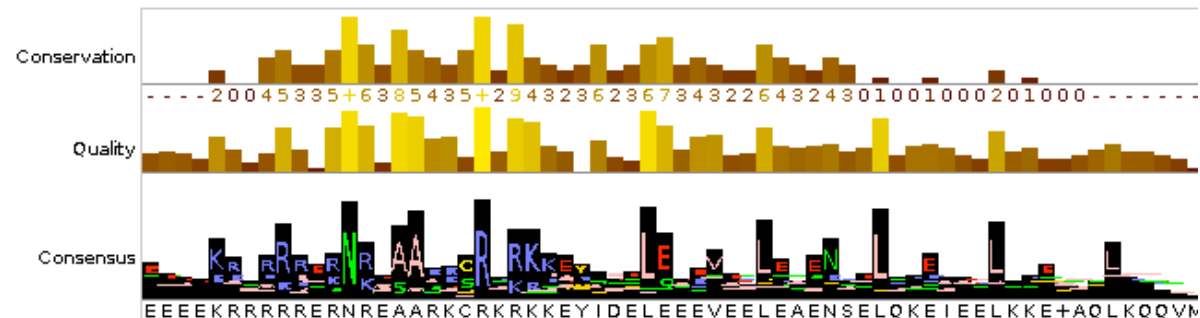
Conservation of the bZIP TFs

SEQUENCE ALIGNMENT

```

sp|Q39162|TGA4_ARATH/78-110  --PDKIQRRLAQNREAAARKSRLRKKAYVQQLETSR-----
sp|P19880|YAP1_YEAST/67-102  -----KQKRTAQNRAAQRAFRRERKERKMKELKKVQSLES I-----
sp|Q39237|TGA1_ARATH/82-118  --PDKIQRRLAQNREAAARKSRLRKKAYVQQLETSRKLKI-----
sp|P14232|TGA1A_TOBAC/74-110 -----KVLRLRLAQNREAAARKSRLRKKAYVQQLENSKLKL IQL-----
sp|P24813|YAP2_YEAST/46-83   -----KSRRTAQNRAAQRAFDRKEAKMKSLQERVELLEQKDA-----
sp|Q00096|CPCA_ASPNG/187-225 --ADPAAMKRARNTAAARKSRARKLERQGERRIIEELERM-----
sp|P23923|HBP1B_WHEAT/46-86  -----KTMRLRLAQNREAAARKSRLRKKAYVQQLENSRLKLTQLEQEL-----
sp|Q9SX27|PAN_ARATH/166-206  -----RTLRLRLAQNREAAARKSRLRKKAYVQQLENRLAQLEEL-----
sp|Q41558|HBP1C_WHEAT/191-231 -----KSLRLRLAQNREAAARKSRLRKKAYIQNLSSRLKLTQLEQEL-----
sp|Q54717|BZPM_DICDI/274-314 -----TSSKNFREKKKEYVKEIESKILALTLENDKLLKKENDSLKT-----
sp|Q39140|TGA6_ARATH/46-86  -----KTLRLRLAQNREAAARKSRLRKKAYVQQLENSRLKLTQLEQEL-----
sp|Q39234|TGA3_ARATH/97-137  ---DKMKRLRLAQNREAAARKSRLRKKAHVQQLLESRLKLSQLEQE-----
sp|P43273|TGA2_ARATH/45-86  ---QKTLRLRLAQNREAAARKSRLRKKAYVQQLENSRLKLTQLEQEL-----
sp|Q64YT2|NFE2_RAT/269-310  -----DIRRRGKNKVAAQNCRRKKLETIVQLERELERLGSEERLLR-----
sp|Q24160|TGA21_TOBAC/168-209 -----KTLRLRLAQNREAAARKSRLRKKAYVQQLENSRLKLSQLEQDLQ-----
sp|Q5EAD3|NFE2_BOVIN/270-311 -----DIRRRGKNKVAAQNCRRKKLETIVQLERELERLGSEERLLR-----
sp|Q16621|NFE2_HUMAN/269-310 -----DIRRRGKNKVAAQNCRRKKLETIVQLERELERLTNERERLLR-----
sp|Q39163|TGA5_ARATH/45-86  ---QKTLRLRLAQNREAAARKSRLRKKAYVQQLENSRLKLTQLEQEL-----
sp|Q08DA8|ATF1_BOVIN/211-267 -PQLKREIRLMKNREAAARECRKKKEYVKCLEENRVAVLENQNKTLIEELKTLKDLYSN-----
sp|Q6PZD9|CEBPE_MOUSE/205-261 ---LEYRLRRERNIAVRKSRDKAKRRIMETQKKVLEYMAENERLRNRVDQLTQELDTLR----
sp|P18846|ATF1_HUMAN/212-268 -PQLKREIRLMKNREAAARECRKKKEYVKCLEENRVAVLENQNKTLIEELKTLKDLYSN-----
sp|Q03935|YAP6_YEAST/220-277 -KTLRNTTAAQNRTAQKAFRRKEKYIKNLEQKSKILLAENNNFKSLNDSLNRDN-----
sp|A1D9Z7|KAPC_NEOFI/95-152  -RPLSTSKRAAQNRAAQRAFRRKEAHIRELEGKVKAYENMGKALQAENYQLREYI-----
sp|Q6NW59|ATF4_DANRE/268-325 -----KKLKKMEQNKTAATRYRQKKRVEQESLNSCESELEKKNRELSEKADSLSREIQYLRDL--
sp|P40917|YAP4_YEAST/236-293  -KPLRNTKRAAQNRSQAQKAFRRREKYIKNLEKSKLFDGLMK-----ENSELKKMIESLKSCL-
sp|Q1LYG4|CR3LB_DANRE/208-265 ERILKKIRRKIRNKQSAQESRKKKEYIDGLESRMAACSAHNHELQKRVFQLEKCNIS-----
sp|Q502F0|CR3LA_DANRE/208-265 EAARKREVRLMKNREAAARECRKKKEYVKCLEENRVAVLENQNKTLIEELKALKDLYC-H-----
sp|P15337|CREB1_RAT/281-338  -----KDKRTAQNRAAQRAFRRERERKMKELDKVQSQLES LNKQSELETKFLRNQVTNLLSEL-
sp|P56095|AP1_KLULA/54-111  EAARKREVRLMKNREAAARECRKKKEYVKCLEENRVAVLENQNKTLIEELKALKDLYC-H-----
sp|P27925|CREB1_BOVIN/265-322 -----RDIRRRGKNKVAAQNCRRKKLENIVELLEQDLHLKDEKEKLLKEKGENDKSLHLLKKQ--
sp|Q16236|NF2L2_HUMAN/499-556

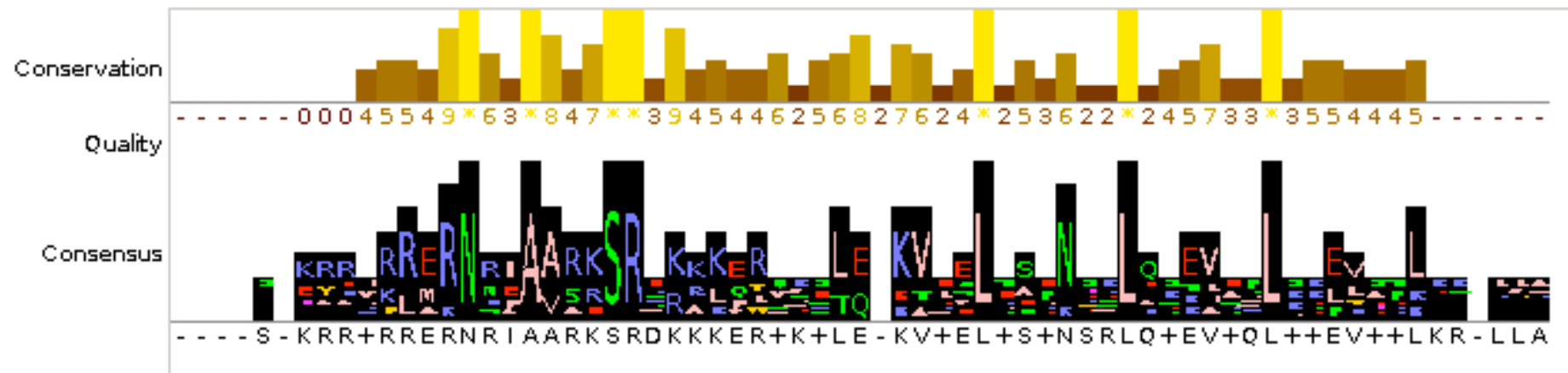
```



bZip: additional information

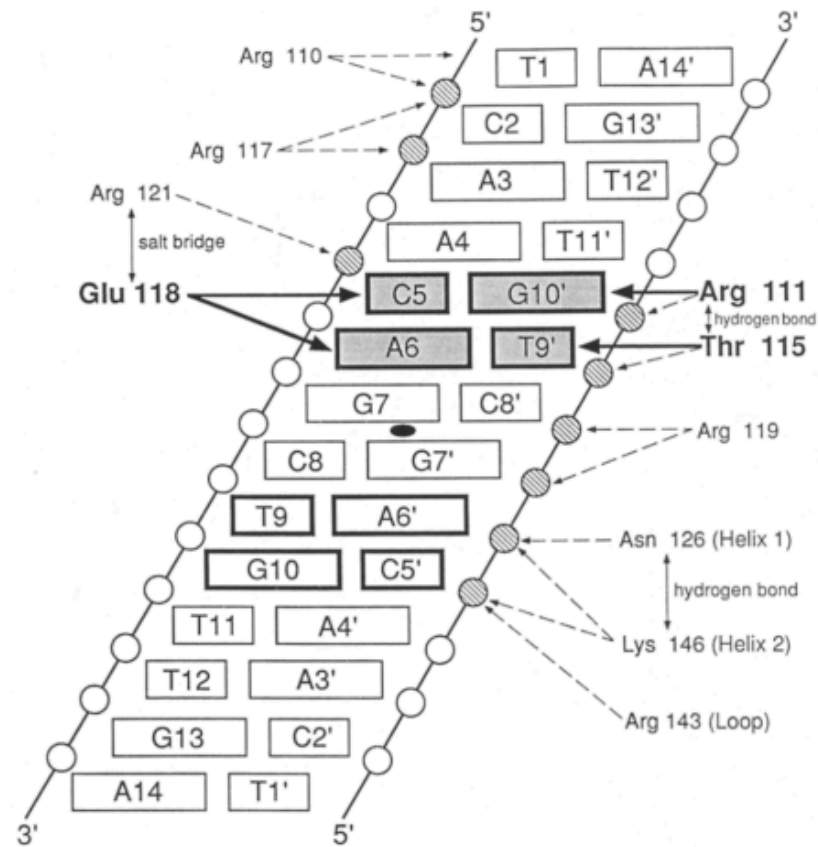
STRUCTURAL ALIGNMENT

	10	20	30	40	50	60
<i>C/EBPalpha</i> /1-60	N---SNEYRVR	RRNNIAVRKSRDKAKQRNVETQ	QKVLELTSDNDR	LRKRVEQLSRELD	TL-RG---	
<i>GCN4</i> /1-55	-----PAALKRARNT	EAARRSRARKLQRMKQLEDKVEEL	LSKNYHLENEVARL	KKLVGER-----		
<i>CREB</i> /1-55	-----KREVR	LMKNREAAARESR	RKKKEYVKSLENRVAVLENQ	NKTLEELKALKDLYSHK-----		
<i>C-JUN</i> /1-57	-----RKRMR	NRRIAASKSR	RKLERIARLEEKVKT	LKAQNSELASTANMLREQVAQLK-QKVM		
<i>C/EBPbeta</i> /1-65	-DKHSDEYKIR	RRNNIAVRKSRDKAKMRNLETQ	HKVLELTAENERLQKKVEQLSRELSTL-RNLFK			
<i>C-FOS</i> /1-60	-----KRRIR	RRERNKMAAAKSR	RNRRELTDTLQAE	TDQLEDEKSALQTEIANLLKEKEKLE-FILA		
<i>cAMP_ATF2</i> /1-61	-----KRRKF	LERNRAAASR	RQKRKVWVQSL	EKKAE	DLSSLNGQLQSEVTLLRNEVAQLKQLLLA	



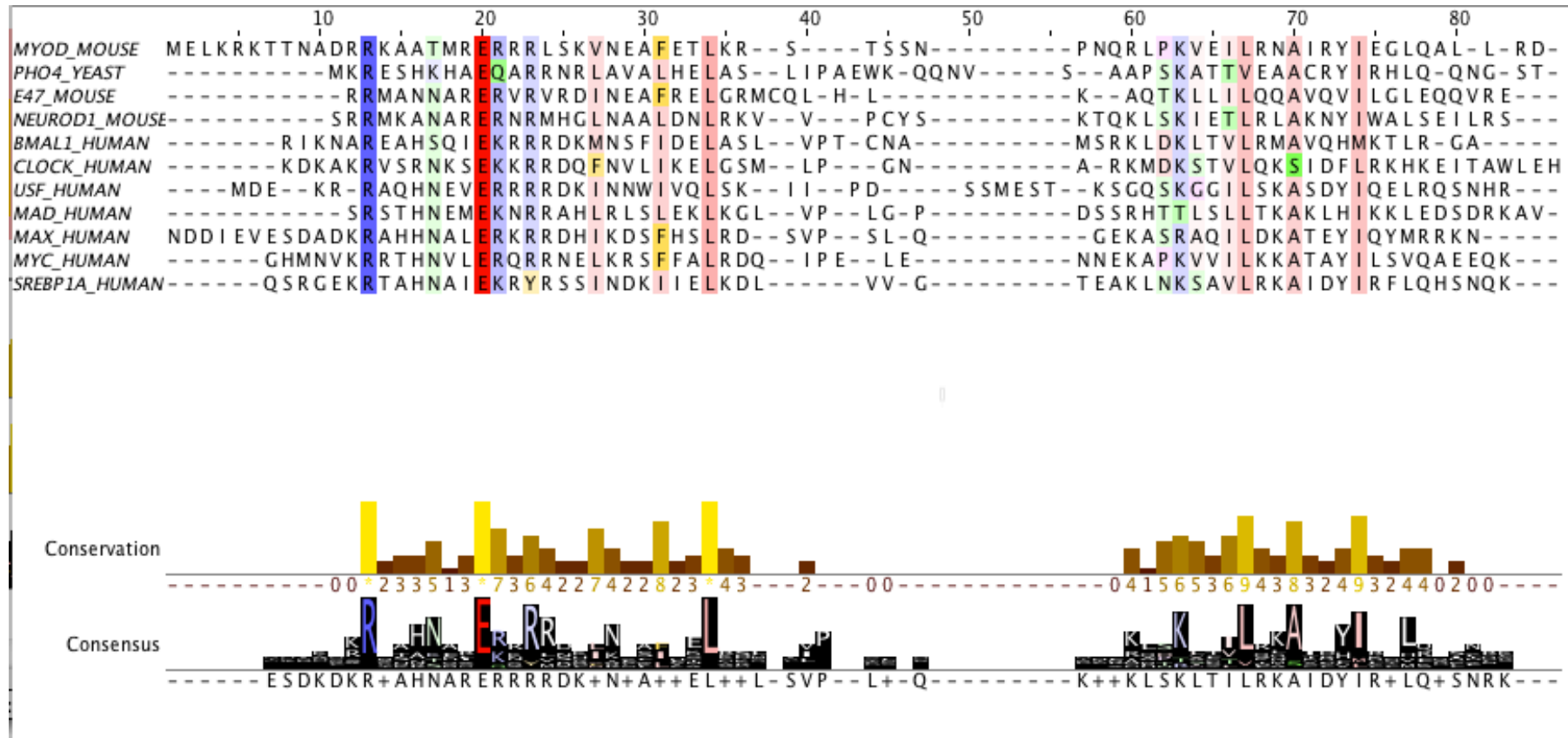
bHLH: additional information

MyoD: contact summary



bHLH: additional information

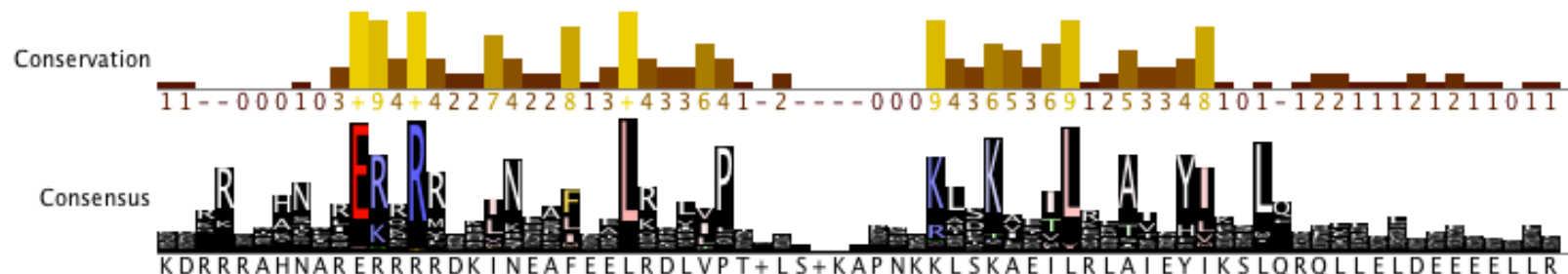
Structural alignment:



bHLH: additional information

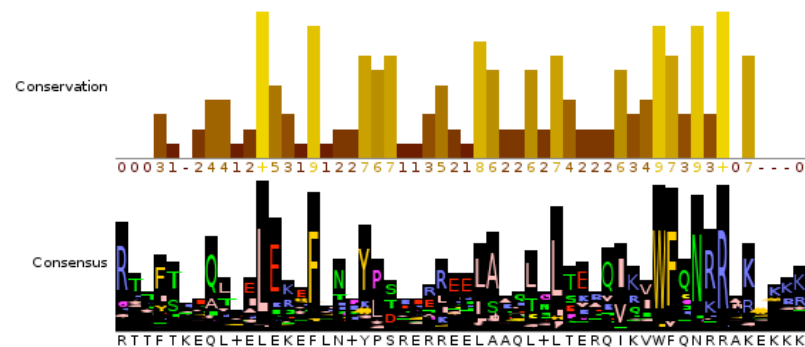
Sequence-based alignment:

	2393	2403	2413	2464	2476	2486	2496
sp P17106 CBF1_YEAST/1-351	kqRKDSHKEV	ERRRRENINTA	INVLSDLLPV-R	---	ESSKAAI	LACAAEY	QKL-ketdeaniekwtlq
sp P16076 MYODB_XENLA/1-287	adRRRAATMR	ERRRLSKVNDA	FETLKRCTST-N	---	PNQRLPKVDI	LRNAISY	DSLQtlldrqeqslypnm
sp P16075 MYOD1_CHICK/1-298	adRRKAATMR	ERRRLSKVNEA	FETLKRCTST-N	---	PNQRLPKVEI	LRNAIRY	ESLQallreqedayypvl
sp P15923 TFE2_HUMAN/1-654	keRRVANNAR	ERLRVRDINEA	FKELGRMCQLhL	---	NSEKQTKLLI	LHQAVSV	LNLEqqvvrernlnpkAAC
sp P15884 ITF2_HUMAN/1-667	keRRMANNAR	ERLRVRDINEA	FKELGRMVQLhL	---	KSDKQTKLLI	LHQAVAV	LSLEqqvvrernlnpkAAC
sp P15881 ITF2_CANFA/1-642	keRRMANNAR	ERLRVRDINEA	FKELGRMVQLhL	---	KSDKQTKLLI	LHQAVAV	LSLEqqvvrernlnpkAAC
sp P15806 TFE2_MOUSE/1-651	keRRVANNAR	ERLRVRDINEA	FKELGRMCQLhL	---	SSEKQTKLLI	LHQAVAV	LSLEqqvvrernlnpkAAC
sp P15375 MYF6_MOUSE/1-242	tdRRKAATLR	ERRRLKKINEA	FEALKRRTVA-N	---	PNQRLPKVEI	LRSAISY	ERLQdlldhrldqqekmqe
sp P15173 MYOG_HUMAN/1-224	vdRRRAATLR	ERRRLKKVNEA	FEALKRSTLL-N	---	PNQRLPKVEI	LRSAIQY	ERLQallsslnqeerdlr
sp P15172 MYOD1_HUMAN/1-320	adRRKAATMR	ERRRLSKVNEA	FETLKRCTSS-N	---	PNQRLPKVEI	LRNAIRY	EGLQallrdqdaappgaa
sp P15171 MYCB_XENLA/1-420	ndKRRTHNVL	ERQRRELKLS	FFALRDQVPRwR	---	NNEKAPKVV	LKKATEYA	----ismqederlriret
sp P14003 HAIR_DROME/1-337	sdRRSNKP	IMEKRRLRARINN	CNLEKTLILD-ATkk	ARHSKLEKAD	LEKTVKHL	QELQRqqaamqqaadpki	
sp P13904 MYODA_XENLA/1-289	adRRKAATMR	ERRRLSKVNEA	FETLKRCTST-N	---	PNQRLPKVEI	LRNAIRY	ESLQallhdqdeafypvl
sp P13903 TWIST_XENLA/1-166	q sQRVMANVR	ERQRTQSLNEA	FSSLRKIPT-L	---	PSDKLSKIQ	T LKLASRY	IDFL-cqvlqsdelksma
sp P13526 ARLC_MAIZE/1-610	gt--KNHVM	SERKRREKLNEM	FLVLKSLLPsiH	---	RVNKAS	LAETIAY	LKELQrrvqelessrepas
sp P13349 MYF5_HUMAN/1-255	mdRRKAATMR	ERRRLKKVNQA	FETLKRCTTT-N	---	PNQRLPKVEI	LRNAIRY	ESLQellreqvenyyslp
sp P13098 ESM8_DROME/1-179	iyQKVKKPML	ERQRARMNKCLDN	LKTLVAE-LR	---	GDDGIMDKAEM	LESAVIFMRQ	--qktppkvaaeeqsl
sp P13097 ESM7_DROME/1-186	qyRKVMKPLL	ERKRRLARINKCL	DELKDLMAE-CVa	TGDAKFEKAD	ILEVTVQHL	RKL-keskhhvpanpeqs	
sp P13096 ESM5_DROME/1-178	lk--VKKPLL	ERQRARMNKCLDT	LKTLVAE-FQ	---	GDDAIMDKAEM	LEAALVF	----mrkqvkkqapvsp
sp P13027 ARR5_MAIZE/1-612	at--KNHVM	SERKRREKLNEM	FLVLKSLLPsiH	---	RVNKAS	LAETIAY	LKELQrrvqelessrepas
sp P12980 LYL1_HUMAN/1-280	vaRRVFTNSR	ERWRQQNVNGA	FAELRKLLPThP	---	PDRKLSKNEV	LRLAMKY	IGFL-vrllrdqaaalaag
sp P12979 MYOG_MOUSE/1-224	vdRRRAATLR	ERRRLKKVNEA	FEALKRSTLL-N	---	PNQRLPKVEI	LRSAIQY	ERLQallsslnqeerdlr
sp P12525 MYCL2_HUMAN/1-357	wtKKKYHSYL	ERKRNRDQSR	FLALRDVEPAIA	---	SCSRVSKVM	LVKATEYLHEL	-aeaeermatekrql
sp P12524 MYCL1_HUMAN/1-364	vtKRKNHNFL	ERKRNRDLRSR	FLALRDQVPTIA	---	SCSKAPKVV	LSKALEYLQAL	-vgaekrmatekrql
sp P12523 MYC_AVIOK/1-416	ndKRRMHNVL	ERQRRELKLS	FFALRDQIPEvA	---	NNEKAPKVV	LKKATEYV	LSIQsdehrliaeeqlr
sp P11420 DA_DROME/1-710	keRRQANNAR	ERIRIRDINEAL	KELGRCMThL	---	KSDKQTKLGI	LNMAVEV	IMTL-eqqvvrernlnpkAA
sp P10627 TWIST_DROME/1-490	snQRVMANVR	ERQRTQSLNDA	FKSLQQLIPT-L	---	PSDKLSKIQ	T LKLATRY	IDFL-crmlsssdslka
sp P10395 MYC_AVIM2/1-416	ndKRRTHNVL	ERQRRELKLS	FFALRDQIPEvA	---	NNEKAPKVV	LKKATEYV	LSIQsdehrliaeeqlr
sp P10166 MYCL1_MOUSE/1-368	vtKRKNHNFL	ERKRNRDLRSR	FLALRDQVPTIA	---	SCSKAPKVV	LSKALEYLQAL	-vgaekkmatekrql
sp P10085 MYOD1_MOUSE/1-318	adRRKAATMR	ERRRLSKVNEA	FETLKRCTSS-N	---	PNQRLPKVEI	LRNAIRY	EGLQallrdqdaappgaa



HTH additional information

	45	73	89	99	112
9anta/1-56	RQTYTRYQTLELEKEFHNNRYL	TRRRRIEIAHALSLTERQIKIWFQ	NNRMKWK-K		
2ecb/1-76	KFEKT-EQLRVLQASFLSSVL	TDEELNRLRAQTKLTRREIDAWFT	EKKKS-KAL		
1s7e/1-50	L-FT--VQRRTLHAIFKKR-P	SKELQITISQQLGLE-LSTVSFFMNR	-----		
2cq/1-59	--VE--PNDTLEKVFVTKYP	DEKRLKGLSKQLDWSVRKIQCFWRHR	Q-DKP		
1ftz/1-70	R-TYT-YQTLLELEKEFHNNRYI	TRRRRIDIANALSLSERQIKIWFQ	NNRMK-KD-		
1wi3/1-71	R-KIS-EALGILQSFHGLYP	DQEAHTLSAQDLKPKHTIIKFFQ	NQRYH-VKH		
2ecc/1-76	G-RKT-EQLAILKSFLLCQWARREDYQKLEQIT	GLRPEIIQWFGDTRYA-LKH			
1ocp/1-67	K-SIE-RVRWSLETMLCPKPS	SLQQITHIANQLGLEKDVVRVWF	CNRR-KGKR-		
1ftt/1-68	R-LFS-AQVYELERRFKQKYL	SAPEREHLSMIHLTPQVKIWFQ	NHRYK-MKR		
1bw5/1-66	R-VLN-KQLHTLRTCYANPRP	DALMKEQLVEMTGLSPRIVRVWF	QNKRC--DK		
1hdp/1-63	RTIE--NVRFALEKSFLLNQKPT	SEEILLIAEQLHMEKEVIRVWF	CNRRQK-EKR		
1akhb/1-78	GHRFT-ENVRILESWFANPYLD	TKGLENLMKNTLSRIQIKNVVSN	NRKR-EKT		
1uhs/1-72	AATMT-DQVEILEYFNKKHP	DPPTTLCLIAAEAGLTEEQTKQWF	KQRLAE-WRR		
2cue/1-68	RTSFT-EQIEALEKEFETHYP	DFARERLAAKIDLPEARIQVWF	SNRRAKW-RR		
1nk2/1-77	RVLFT-AQTYELERRFRQRYL	SAPEREHLSLIRLTPQVKIWFQ	NHRYKT-KR		
1x2m/1-52	-----QPNAILEKVFTTKHP	DEKRLEGLSKQLDWDVRSIQRFWR	QRRNQ---E		
1zq3/1-67	RTTFTSSQIAELEQHFLLGRYL	TAPRLADLSAKLALGTAQVKIWF	KNRRRH-KI		
2cra/1-58	RIPYSGQLRELEREYANKFI	TKDKRRKISAAATSLSERQIT	WFQNRVKE-KK		
1pufa/1-77	RCPYTKHQTLELEKEFLNMYL	TRDRRYEVARLLNLTERQVKIWF	QNRMMKM-KK		
2e1o/1-57	QVRFSDQTILELKKFEQKYL	SPPERKRLAKMLQLSERQVKT	WFQNRRAKW-RR		
1x2n/1-62	RGVLPHATNVMRSWLFHPYP	TEDEKKQIAAQTNLTLLQVNNWF	INARRI-LQ		
1b72a/1-68	RTNFTTRQLTELEKEFHNNRYL	SRARRVEIAATLELNQVVKIWF	QNRMMKQ-KK		
1akha/1-49	--ISPQARAFLEEVFRKQSL	NSKEKEEVAKKCGITPLQVRVWF	INKRMRS---		
1b8ib/1-58	RRNFSKQASEILNEYFYNPYP	SEEAKEELARKCGITVSQVSNWF	GKKIRY-KK		
2lxx/1-60	RTHFTSQQLQELEATFQNRYP	DMSTREEIAVWNTLTERVVRWF	KNRRAKW-RK		
1b8ia/1-62	RQTYTRYQTLELEKEFHNNRYL	TRRRRIEIAHALSLTERQIKIWF	QNRMMKL-KK		
1fjla/1-65	RTTFSASQLDELERAFETQYP	DIYTREELAQRNTLTERIQVWF	QNRRAKL-RK		
1oct/1-60	RTS IETNIRVALEKSFLLNQKPT	SEETMTIADQLNMEKEVIRVWF	CNRRQKEK-R		
3hddb/1-56	RTAFSSQELARLKREFNNRYL	TERRRQQLSSELGLNEAQIKIWF	QNKRAKIK-K		
3hdda/1-55	RTAFSSQELARLKREFNNRYL	TERRRQQLSSELGLNEAQIKIWF	QNKRAKIK-K		
9antb/1-56	RQTYTRYQTLELEKEFHNNRYL	TRRRRIEIAHALSLTERQIKIWF	QNRMMKWK-K		
1ig7/1-58	RTPFTTAQLLALERKFRKQYLS	IAERAEFSSSLSTETQVKIWF	QNRRAKAK-R		
1jgga/1-57	RTAFTRDQLGRLEKEFYENYV	SRPRRCELAAQLNLPESTIKVWF	QNRMMKDK-R		
1jggb/1-57	RTAFTRDQLGRLEKEFYENYV	SRPRRCELAAQLNLPESTIKVWF	QNRMMKDK-R		
1fjlb/1-58	RTTFSASQLDELERAFETQYP	DIYTREELAQRNTLTERIQVWF	QNRRAKL-RK		
1fjlc/1-59	RTTFSASQLDELERAFETQYP	DIYTREELAQRNTLTERIQVWF	QNRRAKL-RK		
1au7a/1-58	RTTISIAAKDALERHFGHSP	SSQIMRMAEELNLEKEVVRVWF	CNRRQREKRV		
1au7b/1-58	RTTISIAAKDALERHFGHSP	SSQIMRMAEELNLEKEVVRVWF	CNRRQREKRV		
1apl/1-59	GHRFTKENVRILESWFANPYLD	TKGLENLMKNTLSRIQIKNVVSN	NRKR-EKT		
1apld/1-58	GHRFTKENVRILESWFANPYLD	TKGLENLMKNTLSRIQIKNVVSN	NRKR-EKT		
1pufb/1-73	RRNFNKQATEILNEYFYNPYP	SEEAKEELAKKCGITVSQVSNWF	GKKIRY-KK		
1b72b/1-73	RRNFNKQATEILNEYFYNPYP	SEEAKEELAKKCGITVSQVSNWF	GKKIRY-KK		

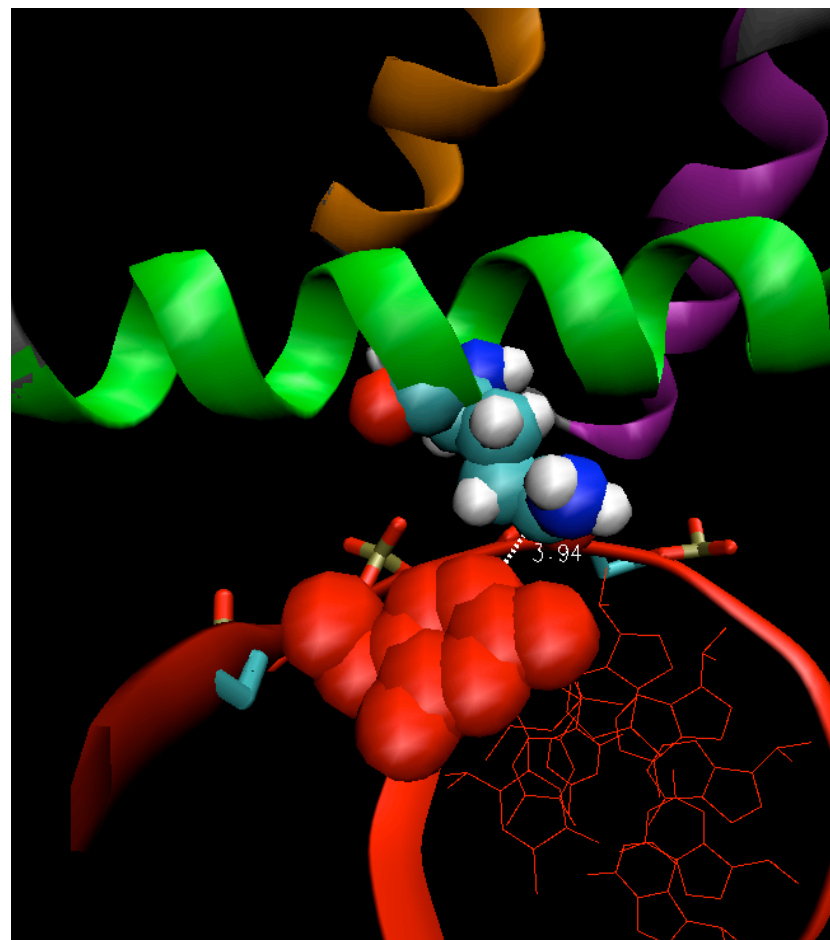


HTH additional information

sp Q90229 DLX3_AMBME/127-183	RKPR	T	I	Y	S	S	Y	Q	L	A	A	L	Q	R	R	F	Q	A	Q	Y	L	A	L	P	E	R	A	E	L	A	A	Q	L	G	L	T	Q	T	Q	V	K	I	W	F	Q	N	R	R	S	K	F	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q6F2E3 IRX1_XENTR/145-186	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	L	N	E	H	R	N	P	Y	P	T	K	G	E	K	I	M	L	A	I	I	T	K	M	T	L	T	Q	V	S	T	W	F	A	N	A	R	R	L	K	K																																																																																																																																																																																																																																																																																																																																																																																			
sp Q3LU38 PROP1_ALOBE/70-126	RRHR	T	T	F	S	A	V	Q	L	E	Q	L	E	S	A	F	G	N	Q	Y	P	D	I	W	A	R	E	S	L	A	R	D	T	G	L	S	E	A	R	I	Q	V	W	F	Q	N	R	R	A	K	Q	R	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q5R7F2 ZHX2_PONAB/446-496	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	K	T	K	E	Q	I	A	H	L	K	A	S	F	L	S	Q	F	P	D	D	A	E	V	Y	R	L	I	E	V	T	G	L	A	R	S	E	I	K	K	W	F	S	D	H	R	Y	R	C	Q	R																																																																																																																																																																																																																																																																																																																																																																									
sp P40764 DLX2_MOUSE/156-212	RKPR	T	I	Y	S	S	F	Q	L	A	A	L	Q	R	R	F	Q	T	Q	Y	L	A	L	P	E	R	A	E	L	A	A	S	L	G	L	T	Q	T	Q	V	K	I	W	F	Q	N	R	R	S	K	F	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q8S897 BLH5_ARATH/248-286	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	F	E	H	P	Y	P	R	D	L	D	K	V	M	L	A	K	Q	T	G	L	T	K	S	Q	V	S	N	W	F	I	N	A	R	V	R	M	-	-	-																																																																																																																																																																																																																																																																																																																																																																																				
sp Q9UMQ3 BARX2_HUMAN/134-190	RRSR	T	I	F	T	E	L	Q	L	M	G	L	E	K	K	F	Q	Q	K	Y	L	S	T	P	D	R	L	D	L	A	Q	S	L	G	L	T	Q	L	Q	V	K	T	W	Y	Q	N	R	R	M	K	W	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q86UP3 ZFH4_HUMAN/2182-2238	RSSR	T	R	F	T	D	Y	Q	L	R	V	L	Q	D	F	F	D	N	A	Y	P	K	D	D	E	I	E	Q	L	S	T	V	L	N	L	P	T	R	V	I	V	W	F	Q	N	A	R	Q	K	A	R	K																																																																																																																																																																																																																																																																																																																																																																																							
sp P24340 HXD9_CHICK/2-58	RKKR	C	P	Y	T	K	Y	Q	T	L	E	L	E	K	E	F	L	N	M	Y	L	T	R	D	R	R	Y	E	V	A	R	I	L	N	L	T	E	R	Q	V	K	I	W	F	Q	N	R	R	M	K	M	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp P15857 HM17_APIME/10-66	RKPR	T	P	F	T	T	Q	L	L	S	L	E	K	K	F	R	K	Q	Y	L	T	I	A	E	R	A	E	F	S	S	S	L	H	L	T	E	T	Q	V	K	I	W	F	Q	N	R	R	A	K	A	K	R																																																																																																																																																																																																																																																																																																																																																																																							
sp B3DM23 P5F13_XENTR/302-357	RKMRT	C	F	D	S	V	L	K	G	R	L	E	G	H	F	M	N	Q	K	P	G	A	R	E	L	A	E	I	A	K	E	L	G	L	E	K	D	V	V	R	V	W	F	C	N	R	R	Q	K	E	K	-	-	-																																																																																																																																																																																																																																																																																																																																																																																					
sp Q15911 ZFH3_HUMAN/2243-2299	RSSR	T	R	F	T	D	Y	Q	L	R	V	L	Q	D	F	F	D	N	A	Y	P	K	D	D	E	F	E	Q	L	S	N	L	L	N	L	P	T	R	V	I	V	W	F	Q	N	A	R	Q	K	A	R	K																																																																																																																																																																																																																																																																																																																																																																																							
sp Q5QMM3 WOX8_ORYS/90-150	-RQR	W	T	P	T	P	M	Q	L	I	L	E	N	I	F	D	N	G	T	P	S	K	Q	K	I	K	D	I	T	A	E	L	Q	I	S	E	T	N	V	Y	N	W	F	Q	N	R	R	A	R	S	K	R																																																																																																																																																																																																																																																																																																																																																																																							
sp Q4PSR7 ATB22_ARATH/70-133	-KKK	K	M	T	S	E	Q	L	K	F	L	E	R	S	F	Q	K	M	K	L	N	P	D	R	K	M	K	L	S	K	E	L	G	L	Q	P	R	Q	I	A	V	W	F	Q	N	R	K	A	R	W	K	N																																																																																																																																																																																																																																																																																																																																																																																							
sp Q6F2E2 VEX1_XENTR/131-186	-RART	K	F	T	A	E	Q	L	E	E	L	E	K	S	F	K	N	R	Y	I	G	S	S	E	K	R	R	L	S	K	V	L	K	L	S	E	N	Q	I	K	T	W	F	Q	N	R	R	M	K	F	K	R																																																																																																																																																																																																																																																																																																																																																																																							
sp Q8SR09 HD2_ENCCU/125-179	--TRAN	F	P	M	D	T	S	Q	L	L	R	S	W	L	K	N	P	Y	P	S	D	A	E	K	A	Y	L	C	Q	K	T	G	L	G	P	A	Q	I	N	N	W	F	I	N	A	R	R	-	-	-																																																																																																																																																																																																																																																																																																																																																																																									
sp Q62066 PHX2A_MOUSE/91-147	RRIR	T	T	F	T	S	A	Q	L	K	E	L	E	R	V	F	A	T	H	Y	P	D	I	Y	T	R	E	E	L	A	L	K	I	D	L	T	E	A	R	V	Q	V	W	F	Q	N	R	R	A	K	F	R	K																																																																																																																																																																																																																																																																																																																																																																																						
sp P14838 HXB5_CHICK/8-64	KRAR	T	A	Y	T	R	Y	Q	T	L	E	L	E	K	E	F	H	N	R	Y	L	T	R	R	R	I	E	I	A	H	A	L	C	L	S	E	R	Q	I	K	I	W	F	Q	N	R	R	M	K	W	K	K																																																																																																																																																																																																																																																																																																																																																																																							
sp Q8C0C0 ZHX2_MOUSE/446-496	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	K	T	K	L	I	A	H	L	K	A	S	F	L	S	Q	F	P	D	D	A	E	V	Y	R	L	I	E	V	T	G	L	A	R	S	E	I	K	K	W	F	S	D	H	R	Y	R	C	Q	R																																																																																																																																																																																																																																																																																																																																																																										
sp Q01822 HXD1_MOUSE/232-285	---RTN	F	S	T	K	Q	L	T	E	L	E	K	E	F	H	N	K	Y	L	T	R	A	R	R	I	E	I	A	N	C	L	Q	L	N	D	T	Q	V	K	I	W	F	Q	N	R	R	M	K	Q	K	K																																																																																																																																																																																																																																																																																																																																																																																								
sp Q7YRX0 RAX2_BOVIN/28-84	RRNR	T	T	F	T	T	Y	Q	L	H	Q	L	E	R	A	F	E	S	H	Y	P	D	V	Y	S	R	E	E	L	A	A	K	V	H	L	P	E	V	R	V	Q	V	W	F	Q	N	R	R	A	K	W	R	R																																																																																																																																																																																																																																																																																																																																																																																						
sp P42586 NKX22_MOUSE/129-185	RKRR	V	L	F	S	K	A	Q	T	Y	E	L	E	R	R	F	R	Q	Y	L	S	A	P	E	R	E	H	L	A	S	L	I	R	L	T	P	T	Q	V	K	I	W	F	Q	N	H	R	Y	K	M	K	R																																																																																																																																																																																																																																																																																																																																																																																							
sp A1YG16 TF2LX_PANPA/74-107	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	K	A	Y	P	S	E	E	E	K	Q	M	L	S	E	K	T	N	L	S	L	L	Q	I	S	N	W	F	I	N	A	R	R	-	-	-																																																																																																																																																																																																																																																																																																																																																																																								
sp Q55144 TLX3_MOUSE/167-223	KKPR	T	S	F	S	R	V	Q	I	C	E	L	E	K	R	F	H	Q	K	Y	L	A	S	A	E	R	A	A	L	A	K	S	L	K	M	T	D	A	Q	V	K	T	W	F	Q	N	R	R	T	K	W	R	R																																																																																																																																																																																																																																																																																																																																																																																						
sp Q9M276 ATB12_ARATH/32-83	----K	R	F	S	E	E	Q	I	K	S	L	E	L	I	F	E	E	T	R	L	E	P	R	K	K	V	Q	V	A	R	E	L	G	L	Q	P	R	Q	V	A	I	W	F	Q	N	K	R	A	R	W	K	-	-	-																																																																																																																																																																																																																																																																																																																																																																																					
sp A1YF22 ZHX1_GORGO/668-717	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	T	P	E	Q	L	H	M	L	K	S	A	F	V	T	Q	W	P	S	P	E	E	Y	D	K	L	A	K	E	S	G	L	A	R	T	D	I	V	S	W	F	G	D	T	R	Y	A	W	K	N																																																																																																																																																																																																																																																																																																																																																																										
sp Q9M2E8 HDG1_ARATH/111-166	KKRY	H	R	H	T	P	K	Q	I	Q	D	L	E	S	V	F	K	C	A	H	P	D	E	K	Q	R	L	D	L	S	R	R	L	N	L	D	P	R	Q	V	K	F	W	F	Q	N	R	R	T	Q	M	K	-	-	-																																																																																																																																																																																																																																																																																																																																																																																				
sp P09631 HXA9_MOUSE/206-262	RKKR	C	P	Y	T	K	H	Q	T	L	E	L	E	K	E	F	L	N	M	Y	L	T	R	D	R	R	Y	E	V	A	R	L	L	N	L	T	E	R	Q	V	K	I	W	F	Q	N	R	R	M	K	M	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q9YGT6 HXA5A_DANRE/191-247	KRPRT	A	Y	T	R	Y	Q	T	L	E	L	E	K	E	F	H	N	R	Y	L	T	R	R	R	I	E	I	A	H	T	L	C	L	S	E	R	Q	I	K	I	W	F	Q	N	R	R	M	K	W	K	K																																																																																																																																																																																																																																																																																																																																																																																								
sp P23463 HXB8_MOUSE/196-252	RRGR	Q	T	Y	S	R	F	Q	T	L	E	L	E	K	E	F	L	N	P	Y	L	T	R	K	R	R	I	E	V	S	H	T	L	A	L	T	E	R	Q	V	K	I	W	F	Q	N	R	R	M	K	W	K	K																																																																																																																																																																																																																																																																																																																																																																																						
sp Q43711 TLX3_HUMAN/167-223	KKPR	T	S	F	S	R	V	Q	I	C	E	L	E	K	R	F	H	Q	K	Y	L	A	S	A	E	R	A	A	L	A	K	S	L	K	M	T	D	A	Q	V	K	T	W	F	Q	N	R	R	T	K	W	R	R																																																																																																																																																																																																																																																																																																																																																																																						
posterior probability	799	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****</

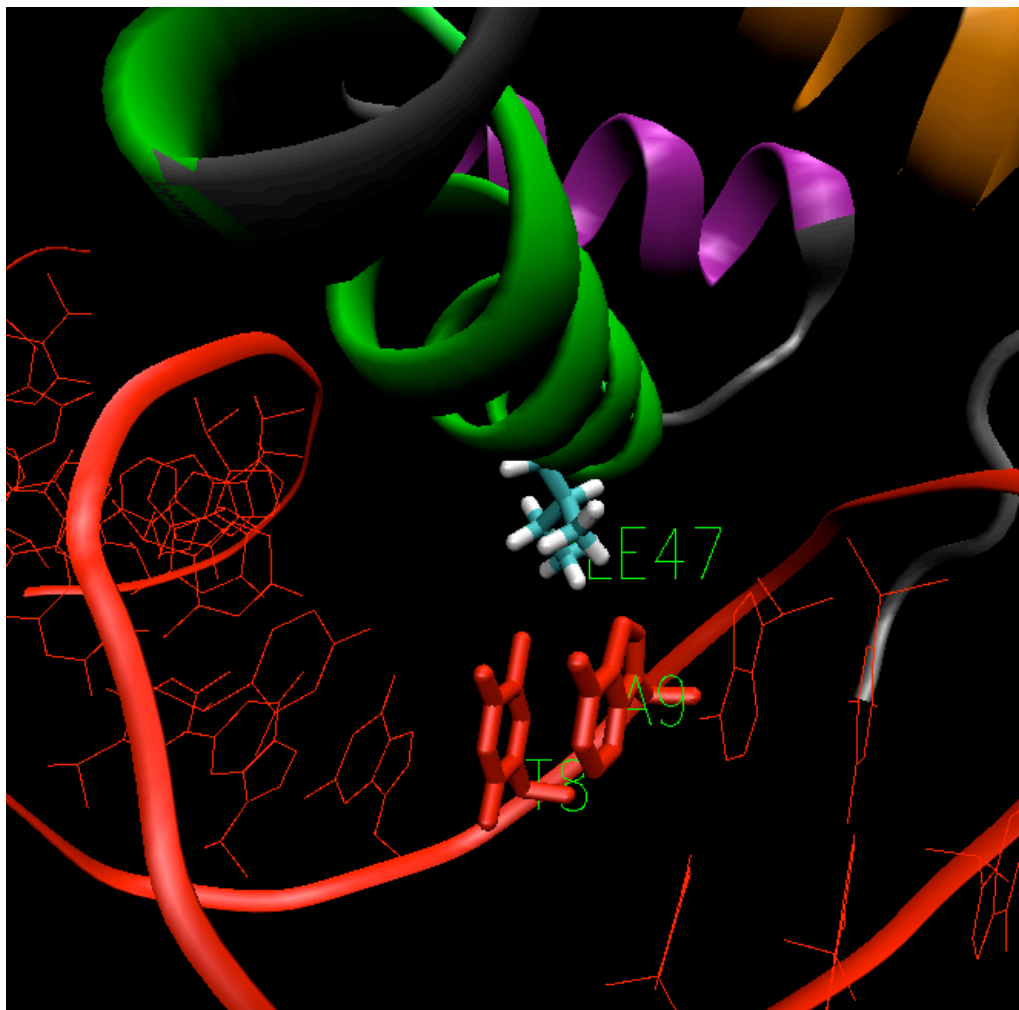
posterior probability 799*****997

HTH additional information



Q50 VdW

HTH additional information



Ile 47

CAG additional information

Initial alignment between CAG and CENP-B:

```

      10      20      30      40      50      60      70      80
cag/1-225  M P S G R V R K P R - - T S L T L E E K M E V I Q S Q E R N - K L S V R D L A K R F N I G K T Q A A D I L K H K Q S I K E G L L S G E L K L N Q M R R N P L S Q R G A Q I D E M
CENP-B[Homo/1-385] M G P - - - - K R - - R Q L T F R E K S R I I Q E V E E N P D L R K G E I A R R F N I P P S T L S T I L K N K R A I L A S E R K Y G V A S T C R K T N K L S P - Y D K L E G L

      90     100     110     120     130     140     150     160     170
cag/1-225  C F D W F S R V R T E N I P I S G E M V R K K A K Q L A V E L G H S N F S A S S C W L E K W R K R H N V R Y N D - - - - - T G D S L D
CENP-B[Homo/1-385] L I A W F Q Q I R A A G L P V K G I I L K E K A L R I A E E L G M D D F T A S N G W L D R F R R R H G V V S C S G V A R A R A R N A A P R T P A A P A S P A A V P S E G S G G S T T

      270     280     290     300     310     320     330     340     350
cag/1-225  - - - - -
CENP-B[Homo/1-385] P C D Y T A N S K G G V T T Q A L A K Y L K A L D T R M A A E S R R V L L L A G R L A A Q S L D T S G L R H V Q L A F F P P G T V H P L E R G V V Q Q V K G H Y R Q A M L L K A I G R

      360     370     380     390
cag/1-225  - - - - - I F K K E L P
CENP-B[Homo/1-385] A R G P G S L R P A A G S H G G P A L C G C R L A G S G A F G H S R L L S

```