
Master Biomedizin 2018

- 1) UCSC & UniProt
- 2) Homology
- 3) MSA
- 4) Phylogeny

8

a. The proteins “P11582” and “P02226” are paralogs, but they differ in length (152 vs 161 amino acids). Is there an extra region in P02226, or the extra amino acids are dispersed along the protein?

Extra region in P02226: “IGNESN”.

b. How could that have happened in evolution?

Deletion in P11582, or insertion in P02226.

8

P11582	GLBE_CHITH	1	-MKFIILALCVA--AASALSGDQIGLVQSTYGVKVGDSVGILYAVFKADPTIQAAFPQFV	57
P02226	GLB7A_CHITH	1	MKFFAVLALCIVGAIASPLSADQAALVKSTWAQVRNSEVEILAAVFTAYPDIQARFPQFA	60
			* :****: . ** **.*. **.*: :*: . . . * ** **.* * * ** *	
P11582	GLBE_CHITH	58	GKDLDAIKGGAEFSTHAGRIVGFLGGVIDD-----LPIIGKHVDALVATHKPRGVTHAQ	111
P02226	GLB7A_CHITH	61	GKDVASIKDTGAFATHAGRIVGFVSEIIALIGNESNAPAVQTLVGQLAASHKARGISQAQ	120
			: :*. . *:**: . :* * : . . *.*: ** *: : **	
P11582	GLBE_CHITH	112	FNNFRAAFIAYLKGHVDTAAVEAAWGATFDAFFGAVFAKM	152
P02226	GLB7A_CHITH	121	FNEFRAGLVSYVSSNVAWNAAAESAWTAGLDNIFGLLFAAL	161
			**:* ** . :*: : . :* : **.*: ** * : * : ** : ** :	

- a.** Extra region in P02226: “IGNESN”.
- b.** Deletion in P11582, or insertion in P02226.

9

a. Both “P17861” (XBP1_HUMAN) and “Q3SZZ2” (XBP1_BOVIN) are “X-box binding protein 1” proteins. Can you detect which region/s of these proteins is/are important for their function? Why? Use Clustal Omega.

What should you do to detect them? No. They are too similar. We would need a protein from a more distant organism.

b. Add the proteins “G5EE07” (G5EE07_CAEEL) and “Q8UVQ5” (Q8UVQ5_DANRE) to the study in 8a. Are you able to identify that region/s now? Why? Use Clustal Omega.

Yes. They are not as similar. bZIP (basic-leucine zipper) domain in positions:

70-133 (human)

70-133 (cattle)

61-117 (worm)

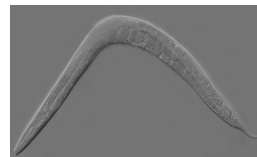
69-132 (zebrafish)



Human
(*Homo sapiens*)



Cattle
(*Bos taurus*)



Worm
(*Caenorhabditis elegans*)



Zebra fish
(*Danio rerio*)

9

```

sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVL LLSGQPA SAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK 60
sp|Q3SZZ2|XBP1_BOVIN     MVVVAPAQSPAAGAPKVL LLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK 60
*****
*****

sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
sp|Q3SZZ2|XBP1_BOVIN     RQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
*****
*****

sp|P17861|XBP1_HUMAN      EKTHGLVVENQELRQRLGMDALVAEEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL 180
sp|Q3SZZ2|XBP1_BOVIN     EKTHGLVVENQELRQRLGMDALVTEEEAEATKGNAGLVAGSAESAALRLRAPLQQVQAQL 180
*****
*****

sp|P17861|XBP1_HUMAN      SPLQNI SPWILAVLT LQIQSLISCFWTTWTQSCSSNALPQSLPAWRSSQSTQKDPVP 240
sp|Q3SZZ2|XBP1_BOVIN     SPLQNI SPWTLALT LQTLSTSCWAFCTWTQSCSSDVLPLQSLPAWRSSQKWTQKDPVP 240
*****
*****

sp|P17861|XBP1_HUMAN      YQPPFLCQWGRHQPSWKPLMN 261
sp|Q3SZZ2|XBP1_BOVIN     YRPPLLHPWGRHQPSWKPLMN 261
*****
*****

```

a. No. They are too similar. We would need a protein from a more distant organism.

b. Yes. They are not as similar. bZIP (basic-leucine zipper) domain in positions:

70-133 (human)
70-133 (cattle)
61-117 (worm)
69-132 (zebrafish)

```

tr|G5EE07|G5EE07_CAEL    -----MSNYPKRIYVLPARHVAAPQPMAPKRALP---TEQVVAQLLGDDMGPS 47
tr|Q8UVQ5|Q8UVQ5_DANRE   MVVVT---AGTGGAHKVL-LISGQKSAAGAAQGGYRSISVMIPNQASSDSDSTTS-G 55
sp|P17861|XBP1_HUMAN     MVVVAAAPNPADGTPKVL-LLSGQPA SAAGAPAG---QALPLMVPAQRGASPEAASGGLP 56
sp|Q3SZZ2|XBP1_BOVIN     MVVVAPAQSPAAGAPKVL-LLSGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP 56
*****

tr|G5EE07|G5EE07_CAEL    GPRKRERLNHLSPEEKMDRRKLN RVAAQNARDKKERSAKIEDVMRDLVEENRRLRAEN 107
tr|Q8UVQ5|Q8UVQ5_DANRE   PLRKRQRLTHLSPEEKALRRKLN RVAAQTARDRKKAKMGELEQQVLELEENQKLHVEN 115
sp|P17861|XBP1_HUMAN     QARKRQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLEN 116
sp|Q3SZZ2|XBP1_BOVIN     QARKRQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLEN 116
*****

tr|G5EE07|G5EE07_CAEL    ERLRRQNKNLMMNQNESMYMEENNENLMNSNDACIQNVVYEEEVVGEVAPVVVGGED 167
tr|Q8UVQ5|Q8UVQ5_DANRE   LLLRDKTSDLLSENEELRLGL--DTLETKEQVQVLE-----SAVSDLG--LVTGSSE 165
sp|P17861|XBP1_HUMAN     QLLREKTHGLVVENQELRQRLGM--DALVAEEEEAE-----AKGNEVR--PVAGSAE 163
sp|Q3SZZ2|XBP1_BOVIN     QLLREKTHGLVVENQELRQRLGM--DALVTEEEAE-----TKGNAG--LVAGSAE 163
*****

tr|G5EE07|G5EE07_CAEL    RRAFESAAFINEPQQWEQARSTSIINNINISNQLRRMDSKKNNTISVDMYLTIIISILCNHMD 227
tr|Q8UVQ5|Q8UVQ5_DANRE   SAAL----RLRVPPQVQVQAQSPNLKTPWILTALALQTLISLISCLVFWTSLTPSSSSRQ 221
sp|P17861|XBP1_HUMAN     SAAL----RLRAPLQQVQAQLSPLQNI SPWILAVLT LQIQSLISCFWTTWTQSCSSNA 219
sp|Q3SZZ2|XBP1_BOVIN     SAAL----RLRAPLQQVQAQLSPLQNI SPWTLALT LQTLSTSCWAFCTWTQSCSSDV 219
*****

tr|G5EE07|G5EE07_CAEL    RNKKMDTSNKKSSNISRAQAESSIDSLLATLRKEQTMQRLVQADPCTHLQKRVKHFRRIP 287
tr|Q8UVQ5|Q8UVQ5_DANRE   TFLKHSLSRSSCWGVQESKYLPPHLQLWGPQLSWKPLMN----- 263
sp|P17861|XBP1_HUMAN     LPQSLPAWRSSQSTQKDPVPYQPPFLCQWGRHQPSWKPLMN----- 261
sp|Q3SZZ2|XBP1_BOVIN     LPQSLPAWRSSQKWTQKDPVPYRPPLLHPWGRHQPSWKPLMN----- 261
*****

```

Domain ⁱ	70 – 133
Domain ⁱ	70 – 133
Domain ⁱ	61 – 117
Domain ⁱ	69 – 132

64	bZIP	PROSITE-ProRule annotation
64	bZIP	PROSITE-ProRule annotation
57	BZIP (basic-leucine zipper)	InterPro annotation
64	BZIP (basic-leucine zipper)	InterPro annotation



P17861 (human)
Q3SZZ2 (cattle)
G5EE07 (worm)
Q8UVQ5 (zebrafish)

10

a. Using the protein “Q90WY9”, find its orthologs in the following organisms: human (*Homo sapiens*) [Q16637], mouse (*Mus musculus*) [P97801], chicken (*Gallus gallus*) [Q98SU9], zebra fish (*Danio rerio*) [Q9W6S8], purple sea urchin (*Strongylocentrotus purpuratus*) [W4XFQ6] and honey bee (*Apis mellifera*) [A0A088A467]. Choose reviewed entries whenever possible.

b. Which regions of the previous sequences are important for their function? Why? Use T-Coffee.

Boxed regions, because they are conserved in evolution.



Human
(*Homo sapiens*)



Mouse
(*Mus musculus*)



Chicken (rooster)
(*Gallus gallus*)



Zebra fish
(*Danio rerio*)



Sea urchin
(*Strongylocentrotus purpuratus*)



Honey bee
(*Apis mellifera*)

Homology + MSA

*Images from: UniProt, TCoffee

■ Your list: 3Z0VB	Entry	Entry name	Protein names	Gene names	Organism
<input type="checkbox"/>	Q90WY9	Q90WY9	Q90WY9_XENLA	Survival motor neuron	Xenopus laevis (African clawed toad)
<input type="checkbox"/>	Q16637	Q16637	SMN_HUMAN	Survival motor neuron protein	Homo sapiens (Human)
<input type="checkbox"/>	P97801	P97801	SMN_MOUSE	Survival motor neuron protein	Mus musculus (Mouse)
<input type="checkbox"/>	Q98SU9	Q98SU9	Q98SU9_CHICK	Survival motor neuron protein	Gallus gallus (Chicken)
<input type="checkbox"/>	Q9W6S8	Q9W6S8	SMN1_DANRE	Survival motor neuron protein 1	Danio rerio (Zebrafish)
<input type="checkbox"/>	W4XFQ6	W4XFQ6	W4XFQ6_STRPU	Uncharacterized protein	Strongylocentrotus purpuratus
<input type="checkbox"/>	A0A088A467	A0A088A467	A0A088A467_APIME	Uncharacterized protein	Apis mellifera (Honeybee)

a. 7 sequences.

b. Boxed regions, because they are conserved in evolution.

(Q16637, SMN_HUMAN)

Feature key	Position(s)	Length	Description	Graphical view
Domain ⁱ	91 – 151	61	Tudor PROSITE-ProRule annotation	

Region

Feature key	Position(s)	Length	Description	Graphical view
Region ⁱ	13 – 44	32	P1 (binding site for GEMIN2)	
Region ⁱ	97 – 209	113	Required for interaction with RPP20/POP7	
Region ⁱ	240 – 267	28	P2 (binding site for SNRNPB)	
Region ⁱ	279 – 294	16	Required for interaction with SYNCRIP	

tr Q90WY9 Q90WY9	MAGL-----EDGGEVLFRRGAGQSD	DSDIWDDTALIKAYDKAVSFKRA	KNEDCTI
sp Q16637 SMN_H	MAMSSGGSGGGVPEQEDSVLFRRTGTGSD	DSDIWDDTALIKAYDKAVSFKHA	KNGDICE
sp P97801 SMN_M	MAMSGGGAG---SEQEDTVLFRRTGTGSD	DSDIWDDTALIKAYDKAVSFKHA	KNGDICE
tr Q98SU9 Q98SU9	MA-----GRVLFRRGAGQSD	DSDMWDDTALIKAYDKAVSFKNA	KNGDICE
sp Q9W6S8 SMN1	MAN-----GAEDVFCRGRTGSD	DSDIWDDTALIKAYDKAVSFKNA	KNGDICE
tr W4XFQ6 W4XFQ6	MAD-----RSGDVFRVDR-SDN	DSDIWDDTALIKAYDKAVSFKNA	KNGDICE
tr A0A088A467 A	MAD-----DNVLFIRNGNST	DSDIWDDTALIKAYDKAVSFKNA	KNGDICE
cons	**
tr Q90WY9 Q90WY9	GAETEEKNPTRK-KNNKKNRSRKKCN	PLKKWRIGDTCNAVSEEDGNIYSATISSIDAKRR	
sp Q16637 SMN_H	TSGK--PKTTPKR-KPAKKNKSQKKNTA	PLQKWKVGDGKSAIWEEDGCIYPTATISIDFKRE	
sp P97801 SMN_M	TPDK--PKGTARR-KPAKKNKSQKKNTA	PLQKWKVGDGKSAIWEEDGCIYPTATISIDFKRE	
tr Q98SU9 Q98SU9	PSDKQEQRAGVKR-KNSKKNRNRKSNAT	PLQKWKVGDGKSAIWEEDGNIYSATISSIDAKRR	
sp Q9W6S8 SMN1	PDEN--DNPQKKR-KNNKKNRSRKKCN	PLQKWKVGDGKSAIWEEDGNIYSATISSIDAKRR	
tr W4XFQ6 W4XFQ6	AR---SKP--KR-KRGGKKK-NKKNLV	PLQKWKVGDGKSAIWEEDGNIYSATISSIDAKRR	
tr A0A088A467 A	G-----IKNSECKQKLQKQSKLTS	PLQKWKVGDGKSAIWEEDGNIYSATISSIDAKRR	
cons
tr Q90WY9 Q90WY9	TCIVVYSGYGNSEEHSLADLRFP	TEAESESDQDRDQEQEINGD	EH-STDESD
sp Q16637 SMN_H	TCVVVYTGYNREEQNLSLLSP	TEAESESDQDRDQEQEINGD	EH-STDESD
sp P97801 SMN_M	TCVVVYTGYNREEQNLSLLSP	TEAESESDQDRDQEQEINGD	EH-STDESD
tr Q98SU9 Q98SU9	TCVVVYTGYNKEEQNLADLLP	TEAESESDQDRDQEQEINGD	EH-STDESD
sp Q9W6S8 SMN1	TCVVVYTGYNKEEQNLADLLP	TEAESESDQDRDQEQEINGD	EH-STDESD
tr W4XFQ6 W4XFQ6	TCIVVYTGYNKEEQNLADLLP	TEAESESDQDRDQEQEINGD	EH-STDESD
tr A0A088A467 A	TCIVVYTGYNKEEQNLADLLP	TEAESESDQDRDQEQEINGD	EH-STDESD
cons
tr Q90WY9 Q90WY9	RSSRSQSKDPQNPRTPKS	SQWNGQFPVP	PPPFMPGF
sp Q16637 SMN_H	NSRSPGNKS	DNIPKPS	APWNSFLPPP
sp P97801 SMN_M	HSRSLRSK	AHSKSKA	APWNSFLPPP
tr Q98SU9 Q98SU9	KSSQSHHNNEN-NCTKARF	SPKNLRFPIP	PT-PPGL
sp Q9W6S8 SMN1	RSFTPKSGHAKHKSNNFPMGPPSWFSPFP	PPPPPHF	KKMDGRRGEGP
tr W4XFQ6 W4XFQ6	HSQSPMHPSPG-SGAGPRK	RSHHPPPP	QHPPPP
tr A0A088A467 A	RKYS-GEKM-DCSEET	KTYKHQFMP	GTSTFN
cons
tr Q90WY9 Q90WY9	-----HPFLSGWPP	-----PFLGPPMI	PPPPPMSP
sp Q16637 SMN_H	PPPPPPPPHLLSCWLP	PPPSGPII	PPPPPIPC
sp P97801 SMN_M	PPPLPPLPPFLPCWMP	PPPSGPII	PPPPPIPC
tr Q98SU9 Q98SU9	-----PPFLSCWPP	PPFAGPPI	PPPPPMGP
sp Q9W6S8 SMN1	-----GPSFGWPP	MIPLGPMI	PPPPPMSP
tr W4XFQ6 W4XFQ6	-----SPYPSGWP	PPPPPMSP	PPPPPMSP
tr A0A088A467 A	-----VMADT	PPPPPMSP	PPPPPMSP
cons
tr Q90WY9 Q90WY9	D-AG-EDDEALGSMILSIWMSGYHTGYLGLKQGRME	SI	
sp Q16637 SMN_H	D-SL-DDADALGSMILSIWMSGYHTGYLGLKQGRME	SI	
sp P97801 SMN_M	D-CL-DDTDLGSMILSIWMSGYHTGYLGLKQGRME	SI	
tr Q98SU9 Q98SU9	D-SL-EDDEALGSMILSIWMSGYHTGYLGLKQGRME	SI	
sp Q9W6S8 SMN1	D-FQ-EDDEALGSMILSIWMSGYHTGYLGLKQGRME	SI	
tr W4XFQ6 W4XFQ6	D-L-EDMDKEALHSLMSWMSGYHTGYLGLKQGRME	SI	
tr A0A088A467 A	KLDP-NDAAELSSMLMSWYISGFHTGYLGLKQGRME	SI	
cons
tr Q90WY9 Q90WY9	-----GKPP-----HQ-K		
sp Q16637 SMN_H	-----SHSL-----N		
sp P97801 SMN_M	-----SHT-----N		
tr Q98SU9 Q98SU9	-----EREA-----YL-K		
sp Q9W6S8 SMN1	-----SKKS-----HR-K		
tr W4XFQ6 W4XFQ6	DKVRETKEPAPSQPAEESQDSNOET		
tr A0A088A467 A	-----KNC-----		
cons

2018 (T-Coffee)

tr | Q90WY9 | Q90WY | MAGL-----EDGGEVFLRRRGAGQSD--DSDIWDDDLTIKAYDKAVASFFKRALKNKDECT
sp | Q16637 | SMN H | MAMSGGSGGGVPPEQDSVLFRRGTGQSD--DSDIWDDELTIKAYDKAVAFKFHALKNKDICE
sp | P97801 | SMN M | MAMSGGGAG---SEQEDTVLFRRGTGQSD--DSDIWDDELTIKAYDKAVASFKNALKNKDICE
tr | Q98SU9 | Q98SU | MA-----GRVLFRRGAGQSD--DSMDMDDTLTIKAYDKAVASFKNALKNDCOSE
sp | Q9W6S8 | SMN1 | GAEDVFVCRCGTQSD--DSDIWDDELTIKAYDKAVASFKNALKNKGEGAT
tr | W4XFQ6 | W4XFQ | MAE-----RSGDVFRVRDR-SDN--DSDIWDDELTIKAYDKAISYYVKMTKDGSEKE
tr | A0A088A467 | A | MAD-----DNVLFIRNGNSNTEHSEDVVDDSLAIKAYDKAINLAKEVEVGKRMGI-

cons

tr | Q90WY9 | Q90WY | GAETEEKNPRTKR-KNNKKNSRKCKNAAPLKWKWRIGDTCNNAVWSEGDNIYSATISSIDAKRR
sp | Q16637 | SMN H | TSGK--PKTTPKR-KPAKKNKSKQKNATAASLQQWKVGDKCSAIWSEDDGCYIPATIASIDFKR
sp | P97801 | SMN M | TPDK--PKGTAR-KPAKKNKSKQKNATTPLKLQWKVGDKCSAVWSEDDGCYIPATIASIDFKR
tr | Q98SU9 | Q98SU | PSDKQEQRAVGRK-KNSKKNSNRNKSNAVPLKWVGDSNAVWSEDDGVNYPATIASINIKRG
sp | Q9W6S8 | SMN1 | PQEN--DNPGKKR-KNNKKNSRKCKNAAPQEKWVGDSYAFVWSEDDGNLTATITTSVDQEGE
tr | W4XFQ6 | W4XFQ | AR--SKP--KR-KRGGKKK-NKKNLVPSQTWKVGDRCKSVFTEDEQVSVAVKAINHKKT
tr | A0A088A467 | A | G-----TKNSECKQLKQKSKLTSKPYYKWIWGAPCRADVISEDGEIYAIIITKIYENG

cons

tr | Q90WY9 | Q90WY | TCIVVYSGYGNEEHSLADLRFPDTSEAESDQRDQEQTINGD-----EH-STDESD
sp | Q16637 | SMN H | TCVVVYTGYGNEEQLNDLLSPICEVANNIEQNAQENES-----QV-STDESE
sp | P97801 | SMN M | TCVVVYTGYGNEEQLNDLLSPTCEVANSTEQNTO--ENES-----QV-STDESE
tr | Q98SU9 | Q98SU | TCVVTYTYGNGKEEQLNDLLPPASDET-----NENET-----PY-STDESE
sp | Q9W6S8 | SMN1 | TCVFVYTYGNEEQLNDLLTPEPDMDADALT--ANVKET-----E5-STDESD
tr | W4XFQ6 | W4XFQ | SCIVRYTYGNGEEKRLSDTFESEAEATSVASVMS-KAELEN-----GYDSMEWT
tr | A0A088A467 | A | SCIVFYGYGNETEKLHLSLESEGLQSQAQQK--NAMEHKFDENEIESNF-STMMS

cons

tr | Q90WY9 | Q90WY | RSSRSHQSKDPQNRTEPKS-----SQWNQGFPPV-----PPPFMPGF--GRH-GEKLDQA-
sp | Q16637 | SMN H | NSRSPGNKS--DNIKPKS--APWSNFLPPP-----PP-MPGPLRPGKP-GLKFNGPP
sp | P97801 | SMN M | HSSRSLSK--AHSKSA--APWTSFLPPP--PP-MPGSLGPGKP-GLKFNGPP
tr | Q98SU9 | Q98SU | KSSQSHHEN--NCKTARF--SPKNLRFPIP--PT--PPGL--GRH-GSKFRTL-
sp | Q9W6S8 | SMN1 | RSTFTOKSGHAHKHSKSNFMPPGPSFWFSPFGP-----PPPPPHF--KKMDGRARGEPP-
tr | W4XFQ6 | W4XFQ | HSQSPMHPSG--SGAGRPK--RSHHPPPPHQHPPPPHPSS--MTH-P--LGVT-
tr | A0A088A467 | A | RKYS-GEKM--DCESEET--KTYKHQFMP-----GTSFN--

cons

tr | Q90WY9 | Q90WY | -----HPFLSGWPP-----PFLPGPMPIPPPPMSP
sp | Q16637 | SMN H | PPPPPPPPHLLSCWLPP-----PFPSPGPIIPPPPPICP
sp | P97801 | SMN M | PPPPLPPPPFLPCWMP-----PFPSPGPIIPPPPPISP
tr | Q98SU9 | Q98SU | -----PPFLSGWPP-----PFPAGPLIIPPPPPGP
sp | Q9W6S8 | SMN1 | GPSFPGWPP-----MIPLGPMIIPPPMSP
tr | W4XFQ6 | W4XFQ | SPYPGSWYPHQAAPPMPMPMPMMSPLPFAWPWGSPAQRISPMPTTPPPHPLPE
tr | A0A088A467 | A | -----VMADT-----MPPAP--PLLPOLMA

cons

tr | Q90WY9 | Q90WY | D-AC-EDDEALGSM LISWYMSGYHTGYLGLKQGRMESSI-----
sp | Q16637 | SMN H | D-SL-DDADALGSM LISWYMSGYHTGYMGFRONQKEGRC-----
sp | P97801 | SMN M | D-CL-DDTDALGSM LISWYMSGYHTGYMGFRONKKEGKC-----
tr | Q98SU9 | Q98SU | D-SP-EDDEALGSM LAWYMSGYHTGYLGLKQSRMEAAL-----
sp | Q9W6S8 | SMN1 | D-FG-EDDEALGSM LISWYMSGYHTGYMGLROGRKEAAA-----
tr | W4XFQ6 | W4XFQ | D-LEEMDK EALHSLMSWYSGYHTGYEGMKSKTSSH SATSKPKEDGSTRPREQGRQTRTR
tr | A0A088A467 | A | KLPD-NDAEALSMLMSWYISGFHTGYHGLKQAEKNQTKR-----

cons

tr | Q90WY9 | Q90WY | -----GKPP-----HQ-K
sp | Q16637 | SMN H | -----SHSL-----N
sp | P97801 | SMN M | -----SHT-----N
tr | Q98SU9 | Q98SU | -----EREA-----YL-K
sp | Q9W6S8 | SMN1 | -----SKKS-----HR-K
tr | W4XFQ6 | W4XFQ | DKVRETKEPAPSQAEESSQDSNQET
tr | A0A088A467 | A | -----KNC-----

cons

2017 (T-Coffee)

[illegible]

11

a. Using the human protein “Q02078”, find its orthologs in the following organisms: orangutan (*Pongo abelii*) [Q5REW7], rat (*Rattus norvegicus*) [Q2MJT0], mouse (*Mus musculus*) [Q60929], cattle (*Bos taurus*) [A2VDZ3], pig (*Sus scrofa*) [A21CN5] and chicken (*Gallus gallus*) [Q9W6U8].

b. Do you think the evolution is pressuring these sequences?

Have they evolved? If yes, how?

Did they gain or lose any domain/motif/region?

Yes. Differences in:

Beta domain (LCR Glu, E)

Glutamines. Maximum Q stretch:

Chicken = 3 Q

Rat = 7 Q

Pig = 4 Q

Orangutan = 9 Q

Cattle = 5 Q

Human = 11 Q

Mouse = 6 Q



Orangutan
(*Pongo abelii*)



Mouse
(*Mus musculus*)



Rat
(*Rattus norvegicus*)



Cattle
(*Bos taurus*)



Pig
(*Sus scrofa*)



Chicken (rooster)
(*Gallus gallus*)

Homology + MSA

*Images from: UniProt

11

■	Your list:...F86JW	Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/>	Q02078	Q02078	MEF2A_HUMAN	Myocyte-specific enhancer factor 2A	MEF2A MEF2	Homo sapiens (Human)	507
<input type="checkbox"/>	Q5REW7	Q5REW7	MEF2A_PONAB	Myocyte-specific enhancer factor 2A	MEF2A	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	494
<input type="checkbox"/>	Q2MJT0	Q2MJT0	MEF2A_RAT	Myocyte-specific enhancer factor 2A	Mef2a	Rattus norvegicus (Rat)	495
<input type="checkbox"/>	Q60929	Q60929	MEF2A_MOUSE	Myocyte-specific enhancer factor 2A	Mef2a	Mus musculus (Mouse)	498
<input type="checkbox"/>	A2VDZ3	A2VDZ3	MEF2A_BOVIN	Myocyte-specific enhancer factor 2A	MEF2A	Bos taurus (Bovine)	492
<input type="checkbox"/>	A2ICN5	A2ICN5	MEF2A_PIG	Myocyte-specific enhancer factor 2A	MEF2A	Sus scrofa (Pig)	507
<input type="checkbox"/>	Q9W6U8	Q9W6U8	MEF2A_CHICK	Myocyte-specific enhancer factor 2A	MEF2A	Gallus gallus (Chicken)	499

a. 7 sequences.

b. Yes. Differences in:

Beta domain (LCR Glu, E)
Glutamines. Maximum Q stretch:

Chicken = 3 Q
Pig = 4 Q
Cattle = 5 Q
Mouse = 6 Q
Rat = 7 Q
Orangutan = 9 Q
Human = 11 Q

Q02078 MEF2A_HUMAN
Q5REW7 MEF2A_PONAB
Q2MJT0 MEF2A_RAT
Q60929 MEF2A_MOUSE
A2VDZ3 MEF2A_BOVIN
A2ICN5 MEF2A_PIG
Q9W6U8 MEF2A_CHICK

Q02078 MEF2A_HUMAN
Q5REW7 MEF2A_PONAB
Q2MJT0 MEF2A_RAT
Q60929 MEF2A_MOUSE
A2VDZ3 MEF2A_BOVIN
A2ICN5 MEF2A_PIG
Q9W6U8 MEF2A_CHICK

Required for interaction
with MAPKS

Beta domain

```

241 ATG-ANSLGKVMPTKSPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 ATG-ANSLGKVMPTKSPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL ----- N T Q 289
241 NTG-ANSVGKVMPTKSPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL ----- N A Q 291
239 NTG-ANSLGKVMPTKSPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N A Q 297
241 TTG-ANSLGKVMPTKSPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL ----- N T Q 291
241 TTG-ANSLGKVMPTESPPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 TAGGGNGLGKVMPTKSPPPGGGSLG MNNRKPDRLRVVIPPSSKGMMPPL T E E D E L E L N T Q 298
: * . * . : * * * * * : * * * * * : * * . * * * * * : * *

```

```

420 QQQQQQQQQQQPPPP-PQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
410 QQQQQQQQQQP-PPPPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 465
412 QQQQQQQQQQPQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 466
418 QQQQQQQQQQP-PP-PP-PPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 469
412 QQQQQP-----QP-PPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 463
420 QQQQPPPPSQAPQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
419 PQQQPPQPQPQPQP-----PQPQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 470
*** ** * : * * * * * : * * * * * : * *

```