
Master Biomedizin 2019

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

3) MSA

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? Use UniProt.

Extra region in P02226: “IGNESN”.

b. How could that have happened in evolution?

Deletion in P11582, or insertion in P02226.

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

Pablo Mier
munoz@uni-mainz.de

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)

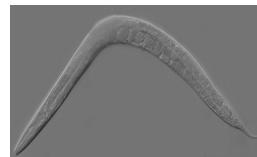
69-132 (zebrafish)



Human
(*Homo sapiens*)



Cattle
(*Bos taurus*)



Worm
(*Caenorhabditis elegans*)



Zebra fish
(*Danio rerio*)

9

```

sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVLLLSGQPAASAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK 60
sp|Q3SZ22|XBP1_BOVIN     MVVVAPAQSPAAGAPKVLLLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK 60
*****
*****

sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
sp|Q3SZ22|XBP1_BOVIN     RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
*****
*****

sp|P17861|XBP1_HUMAN      EKTHTGLVVENQELRQRLGMDALVAEEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL 180
sp|Q3SZ22|XBP1_BOVIN     EKTHTGLVVENQELRQRLGMDALVTEEEAETKGNAGLVAGSAESAALRLRAPLQQVQAQL 180
*****
*****

sp|P17861|XBP1_HUMAN      SPLQNISPWILAVLTQIQSLISCFWTTWTQSCSSNALPQSLPAWRSSQSTQKDPVP 240
sp|Q3SZ22|XBP1_BOVIN     SPLQNISPWTLALTTLQTLSTSCWAFCTWTQSCSSDVLPQSLPAWRSSQKWTQKDPVP 240
*****
*****

sp|P17861|XBP1_HUMAN      YQPPFLCQWGRHQPSWKPLMN 261
sp|Q3SZ22|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-LISGQKASTGAAQGGYSRSISVMIPNQASSDSDSTTS-G 55
sp|P17861|XBP1_HUMAN     MVVVAAAPNPADGTPKVLL-LISGQPAASAAGAPAG---QALPLMVPAQRGASPEAASGGLP 56
sp|Q3SZ22|XBP1_BOVIN     MVVVAPAQSPAAGAPKVLL-LISGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP 56
*****

tr|G5EE07|G5EE07_CAEL    GPRKRERLNHLSQEEKMDRRKLKNRVAAQNARDKKKERSAKIEDVMRDLVEENRRLRAEN 107
tr|Q8UVQ5|Q8UVQ5_DANRE   PLRKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKAKMGELEQQVLELEENQKLHVEN 115
sp|P17861|XBP1_HUMAN     QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEENQKLLLEN 116
sp|Q3SZ22|XBP1_BOVIN     QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEENQKLLLEN 116
*****

tr|G5EE07|G5EE07_CAEL    ERLRRQKNLNMNQNESMYMEENNENLMNSNDACIQNVVYEEEVVGEVAPVVVGGED 167
tr|Q8UVQ5|Q8UVQ5_DANRE   LLLRDKTSDLLSENEELRLGL--DTLETKEQVQVLE-----SAVSDLG--LVTGSSE 165
sp|P17861|XBP1_HUMAN     QLLREKTHGLVVENQELRQLGM--DALVAEEEEAE-----AKGNEVR--PVAGSAE 163
sp|Q3SZ22|XBP1_BOVIN     QLLREKTHGLVVENQELRQLGM--DALVTEEEAE-----TKGNAG--LVAGSAE 163
*****

tr|G5EE07|G5EE07_CAEL    RRAFESAAFINEPQQWEQARSTSINNINISNQLRRMDSKKNNTISVDMYLTIIISILCNHMD 227
tr|Q8UVQ5|Q8UVQ5_DANRE   SAAL----RLRVPPQVQVQAQSPNLKTPWILTALALQTLISLISCLVFWTSLTPSSSSRQ 221
sp|P17861|XBP1_HUMAN     SAAL----RLRAPLQQVQAQLSPLQNISPWILAVLTQIQSLISCFWTTWTQSCSSNA 219
sp|Q3SZ22|XBP1_BOVIN     SAAL----RLRAPLQQVQAQLSPLQNISPWTLALTTLQTLSTSCWAFCTWTQSCSSDV 219
*****

tr|G5EE07|G5EE07_CAEL    RNKKMDTSNKSSNISRAQAESSIDSLLATLRKEQTMQRLVQADPCTHLQKRVKHFRRIP 287
tr|Q8UVQ5|Q8UVQ5_DANRE   TFLKHSLSRSSCWGVQESKYLPPHLQLWGPQLSWKPLMN----- 263
sp|P17861|XBP1_HUMAN     LPQSLPAWRSSQSTQKDPVPYQPPFLCQWGRHQPSWKPLMN----- 261
sp|Q3SZ22|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)
Q3SZ22 (cattle)
G5EE07 (worm)
Q8UVQ5 (zebrafish)

***) Homology + MSA**

10

a. Using the protein “Q90WY9”, find its orthologs in the following organisms: human (*Homo sapiens*) [Q16637; 55.2%], mouse (*Mus musculus*) [P97801; 54.8%], chicken (*Gallus gallus*) [Q98SU9; 60.7%], zebra fish (*Danio rerio*) [Q9W6S8; 56.9%], purple sea urchin (*Strongylocentrotus purpuratus*) [W4XFQ6; 37.5%] and honey bee (*Apis mellifera*) [A0A088A467; 37.4%]. 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 frog)
<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	DSDIWDDTALIKAYDKAVSSFKRA	KNEDCTI
sp Q16637 SMN_H	MAMSSGGSGGGVPEQEDSVLFRRTGTGSD	DSDIWDDTALIKAYDKAVSFKHA	KNGDICE
sp P97801 SMN_M	MAMSGGGAG---SEQEDTVLFRRTGTGSD	DSDIWDDTALIKAYDKAVSFKHA	KNGDICE
tr Q98SU9 Q98SU9	MA-----GRVLFRRGAGQSD	DSDMWDDTALIKAYDKAVSFKNA	KNGDCE
sp Q9W6S8 SMN1	MAN-----GAEDVFCRGRTGSD	DSDIWDDTALIKAYDKAVSFKNA	LKGDGAT
tr W4XFQ6 W4XFQ6	MAD-----RSGDVFRVDR-SDN	DSDIWDDTALIKAYDKAISYVKG	TKDGSEKE
tr A0A088A467 A	MAD-----DNVLFIRNGNST	HSSEVWDDTALIKAYDKAINLKE	VGKRMGI-
cons	**
tr Q90WY9 Q90WY9	GAETEEKNPTRK-KNNKKNRSRKKCNAPLKKWRIGDTCNAVSESDGNIYSATISSIDAKRR		
sp Q16637 SMN_H	TSGK--PKTTPKR-KPAKKNKSQKKNTASLQWKVGDGKSAIWSDEGCIYPATIASIDFKRE		
sp P97801 SMN_M	TPDK--PKGTARR-KPAKKNKSQKKNTAPLKQWKVGDGKSAIWSDEGCIYPATITSIDFKRE		
tr Q98SU9 Q98SU9	PSDKQEQRAGVKR-KNSKKNRNRKSNAPLKQWKVGDGKSAIWSDEGNIYPATIASINLRKG		
sp Q9W6S8 SMN1	PDEN--DNPQKKR-KNNKKNKSQKKCNAPDKEWQVGDGKSAIWSDEGNIYPATITSDQEKQ		
tr W4XFQ6 W4XFQ6	AR---SKP--KR-KRGGKKK-NKKNLVESQTKWKVGDGKSAIWSDEGNIYPATITSDQEKQ		
tr A0A088A467 A	G-----IKNSECKQKLQKQSKLTSPPYKWKIVGAPCRAYVSEDEGIYEAITTIYENNG		
cons
tr Q90WY9 Q90WY9	TCIVVYSGYGNSEEHSLADLRFPDTEASESDQDRDQEQEINGD	EH-STDESD	
sp Q16637 SMN_H	TCVVVYTGYNREEQNLSDDLSPTECANNIEQNAQENENES	QV-STDESE	
sp P97801 SMN_M	TCVVVYTGYNREEQNLSDDLSPTECANSTEQNTQ--ENES	QV-STDDSE	
tr Q98SU9 Q98SU9	TCVVYTYGYNKEEQNLADLLPDSDET--NENET	PY-STDESE	
sp Q9W6S8 SMN1	TCVVFYTYGYNKEEQNLSDLLTEPPDMDDEDALKT--ANVKET	ES-STDESD	
tr W4XFQ6 W4XFQ6	SCIVRYTYGYNKEEKRLSDFSEEAETSVAASN--KALEN	GYDSMEWTD	
tr A0A088A467 A	SCIVKFVGYGNTKEVELHSLLESLQSQIAQQK--NAMEHFKDEENEEISESNF-STNMNS		
cons
tr Q90WY9 Q90WY9	RSSRSQSKDPQNPRTPKS---SQWNGQFPVP	PPPFMPGF--GRH-GEKLDQA-	
sp Q16637 SMN_H	NSRSPGNKS---DNIPKPS---APWNSFLPPP	PP--MPGPRLGPGKP-GLKFNGPP	
sp P97801 SMN_M	HSRSLRSK---AHSKSKA---APWTSFLPPP	PP--MPGSLGPGKP-GLKFNGPP	
tr Q98SU9 Q98SU9	KSSQSHHNNEN--NCTKARF--SPKNLRFPIP	PT--PPGL--GRH-GSKFRTL-	
sp Q9W6S8 SMN1	RSFTPKSGHAKHKSNSNFPMPGPPSWFSPFPPG	PPPPPHF--KKMDGRRGEGP-	
tr W4XFQ6 W4XFQ6	HSQSPMHPSPG--SGAGPRK---RSHHPPPPQHPPPPSPSS	MTH-P--LGYT-	
tr A0A088A467 A	RKYS-GEKM--DCESEET---KTYKHQFMP	GTSTFN--	
cons
tr Q90WY9 Q90WY9	-----HPFLSGWPP-----PFLGPPMIPPPMPSP		
sp Q16637 SMN_H	PPPPPPPPHLLSCWLP-----PFPSPGPIIPPPPIPC		
sp P97801 SMN_M	PPPLPPLPPFLPCWMP-----PFPSPGPIIPPPPIISP		
tr Q98SU9 Q98SU9	-----PPFLSCWPP-----PFPAGPPIIPPPMPGP		
sp Q9W6S8 SMN1	-----GPSFGWPP-----MIPLGPPMIPPPMPSP		
tr W4XFQ6 W4XFQ6	-----SPYPSWYPPHQAPPPMPPPPMMSPLPFAWGPAAQRIIPSPPTPPHPLPE		
tr A0A088A467 A	-----VMADT-----MPPAP--PLPPQLMA		
cons
tr Q90WY9 Q90WY9	D-AG-EDDEALGSMILSIWMSGYHTGYLGLKQGRME	SSI	
sp Q16637 SMN_H	D-SL-DDADALGSMILSIWMSGYHTGYMGFRONQKE	ORC	
sp P97801 SMN_M	D-CL-DDTDALGSMILSIWMSGYHTGYMGFRONKKE	CKC	
tr Q98SU9 Q98SU9	D-SL-EDDEALGSMILSIWMSGYHTGYLGLKQSRME	IAL	
sp Q9W6S8 SMN1	D-FQ-EDDEALGSMILSIWMSGYHTGYMLGRGRKE	IAA	
tr W4XFQ6 W4XFQ6	D-L-EMDKALHSLMSWMSGYHTGYEGMKKSKTS	HSATSKPKEDGSTPRREQGROTRTR	
tr A0A088A467 A	KLDP-NDAAELSSMLMSWYISGFHTGYHGLKQAEKN	TKR	
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	DKVRETKEPAPSQPAEESQDSNQET		
tr A0A088A467 A	-----KNC-----		
cons

2019 & 2018 (T-Coffee)

[illegible]

2017 (T-Coffee)

[illegible]

11

a. Using the human protein “Q02078”, find its orthologs in the following organisms: orangutan (*Pongo abelii*) [Q5REW7; 91.1%], rat (*Rattus norvegicus*) [Q2MJT0; 93.3%], mouse (*Mus musculus*) [Q60929; 89.5%], cattle (*Bos taurus*) [A2VDZ3; 93.9%], pig (*Sus scrofa*) [A2ICN5; 95.7%] and chicken (*Gallus gallus*) [Q9W6U8; 82.9%]. Choose reviewed entries whenever possible.

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

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

Required for interaction
with MAPKs Beta domain

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

```

241 ATG-ANSLGKVMPTKSPPPPGGGNLG MNSRKPDLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 ATG-ANSLGKVMPTKSPPPPGGGNLG MNSRKPDLRVVIPPSSKGMMPPL ----- N T Q 289
241 NTG-ANSVGKVMPTKSPPPPGGGSVG MNSRKPDLRVVIPPSSKGMMPPL ----- N A Q 291
239 NTG-ANSLGKVMPTKSPPPPGGGSLG MNSRKPDLRVVIPPSSKGMMPPL S E E E E L E L N A Q 297
241 TTG-ANSLGKVMPTKSPPPPGGGSLG MNSRKPDLRVVIPPSSKGMMPPL ----- N T Q 291
241 TTG-ANSLGKVMPTESPPPPGGGNLG MNSRKPDLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 TAGGGNGLGKVMPTKSPPPPGGGGLG MNNRKPDLRVVIPPSSKGMMPPL T E E D E L E L N T Q 298
: * . * . : * * * * * : * * * * * : * * . * * * * * * * * * * * * * * * * *

```

```

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

```