Master Biomedizin 2016

MSA

Pablo Mier
munoz@uni-mainz.de

9

a. Using the protein "Q90WY9", find its orthologs in the following organisms: human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), zebrafish (*Danio rerio*), purple sea urchin (*Strongylocentrotus purpuratus*) and honey bee (*Apis mellifera*). Choose reviewed entries whenever possible. Save the sequences in a fasta file.

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

Human
(*Homo sapiens*)

Mouse
(*Mus musculus*)

Chicken (rooster)
(*Gallus gallus*)

Zebrafish
(*Danio rerio*)

Sea urchin
(*Strongylocentrotus purpuratus*)

Honey bee
(*Apis mellifera*)

Images from: UniProt, TCoffee

9

	Your list:...3Z0VB	Entry	Entry name		Protein names	Gene names	Organism
☐	Q90WY9	Q90WY9	Q90WY9_XENLA		Survival motor neuron		Xenopus laev
☐	Q16637	Q16637	SMN_HUMAN		Survival motor neuron protein	**SMN1** SMN,SMNT **SMN2** SMNC	Homo sapiens
☐	P97801	P97801	SMN_MOUSE		Survival motor neuron protein	Smn1 Smn	Mus musculu
☐	Q98SU9	Q98SU9	Q98SU9_CHICK		Survival motor neuron protein	**SMN**	Gallus gallus
☐	Q9W6S8	Q9W6S8	SMN1_DANRE		Survival motor neuron protein 1	smn1 smn	Danio rerio (Z
☐	W4XFQ6	W4XFQ6	W4XFQ6_STRPU		Uncharacterized protein	**Sp-Smn1**	Strongylocent
☐	A0A088A467	A0A088A467	A0A088A467_APIME		Uncharacterized protein	LOC100577491	Apis mellifera

a. 7 sequences.

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

(Q16637, SMN_HUMAN)

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

Region

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

Pablo Mier
munoz@uni-mainz.de

JGU | imb Institute of Molecular Biology

10

a. Using the human protein "Q02078", find its orthologs in the following organisms: orangutan (*Pongo abelii*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), cattle (*Bos taurus*), pig (*Sus scrofa*) and chicken (*Gallus gallus*). Save the sequences in a fasta file.

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?

Orangutan
(*Pongo abelii*)

Mouse
(*Mus musculus*)

Rat
(*Rattus norvegicus*)

Cattle
(*Bos taurus*)

Pig
(*Sus scrofa*)

Chicken (rooster)
(*Gallus gallus*)

Homology + MSA

10

	Your list:...F86JW	Entry	Entry name		Protein names	Gene names	Organism	Length
☐	Q02078	Q02078	MEF2A_HUMAN		Myocyte-specific enhancer factor 2A	MEF2A MEF2	Homo sapiens (Human)	507
☐	Q5REW7	Q5REW7	MEF2A_PONAB		Myocyte-specific enhancer factor 2A	MEF2A	Pongo abelii (Sumatran orangutan) (Pongo pygmaeus abelii)	494
☐	Q2MJT0	Q2MJT0	MEF2A_RAT		Myocyte-specific enhancer factor 2A	Mef2a	Rattus norvegicus (Rat)	495
☐	Q60929	Q60929	MEF2A_MOUSE		Myocyte-specific enhancer factor 2A	Mef2a	Mus musculus (Mouse)	498
☐	A2VDZ3	A2VDZ3	MEF2A_BOVIN		Myocyte-specific enhancer factor 2A	MEF2A	Bos taurus (Bovine)	492
☐	A2ICN5	A2ICN5	MEF2A_PIG		Myocyte-specific enhancer factor 2A	MEF2A	Sus scrofa (Pig)	507
☐	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 = 4 Q
Pig = 4 Q
Cattle = 5 Q
Mouse = 6 Q
Rat = 7 Q
Orangutan = 9 Q
Human = 11 Q

**Required for interaction
with MAPKs**   **Beta domain**

```
Q02078 MEF2A_HUMAN   241 ATG-ANSLGKVMPTKSPPPPGGGNLGMNSRKPDLRVVIPPSSKGMMPPLSEEEELELNTQ 299
Q5REW7 MEF2A_PONAB   239 ATG-ANSLGKVMPTKSPPPPGGGNLGMNSRKPDLRVVIPPSSKGMMPPL-------NTQ 289
Q2MJT0 MEF2A_RAT     241 NTG-ANSVGKVMPTKSPPPPGGGSVGMNSRKPDLRVVIPPSSKGMMPPL-------NAQ 291
Q60929 MEF2A_MOUSE   239 NTG-ANSLGKVMPTKSPPPPGGGSLGMNSRKPDLRVVIPPSSKGMMPPLSEEEELELNAQ 297
A2VDZ3 MEF2A_BOVIN   241 TTG-ANSLGKVMPTKSPPPPGGGSLGMNSRKPDLRVVIPPSSKGMMPPL-------NTQ 291
A2ICN5 MEF2A_PIG     241 TTG-ANSLGKVMPTESPPPPGGGNLGMNSRKPDLRVVIPPSSKGMMPPLSEEEELELNTQ 299
Q9W6U8 MEF2A_CHICK   239 TAGGGNGLGKVMPTKSPPPPGGGGLGMNNRKPDLRVVIPPSSKGMMPPLTEEDELELNTQ 298
                         :*  .*.:*****:*******  :***.*************************  *:*

Q02078 MEF2A_HUMAN   420 QQQQQQQQQQQPPPP-PQPQPQPPQPQPRQEMGRSPVDSLSSSSSSYDGSDREDPRGDFH 478
Q5REW7 MEF2A_PONAB   410 QQQQQQQQQP----PPPPPQPQPPQPQPRQEMGRSPVDSLSSSSSSYDGSDREDPRGDFH 465
Q2MJT0 MEF2A_RAT     412 QQQQQQQPQQQPPPQPPPQ-----PQPQPQPRQEMGRSPVDSLSSSSSSYDGSDREDPRGDFH 466
Q60929 MEF2A_MOUSE   418 QQQQQQPQQQ----PP----PQPQPQPRQEMGRSPVDSLSSSSSSYDGSDREDPRGDFH 469
A2VDZ3 MEF2A_BOVIN   412 QQQQQP-------QP-PPPPPQAPQPQPRQEVGRSPVDSLSSSSSSYDGSDREDPRGDFH 463
A2ICN5 MEF2A_PIG     420 QQQQPPPPSQAPQPQPPQPQP-QPQPQARQEMGRSPVDSLSSSSSSYDGSDREDPRGDFH 478
Q9W6U8 MEF2A_CHICK   419 PQQQPPQQPQPPQP--------PQQPPQRQEMGRSPVDSLSSSSSSYDGSDREDPRSDFH 470
                         ***             **  ***:***********************:.***
```

Pablo Mier
munoz@uni-mainz.de

JG|U   imb
Institute of
Molecular Biology

12

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.
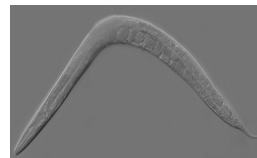What should you do to detect them?

b. Add the proteins "G5EE07" (G5EE07_CAEEL) and "Q8UVQ5" (Q8UVQ5_DANRE) to the study in 12a. Are you able to identify that region/s now? Why? Use Clustal Omega.

Human
(*Homo sapiens*)

Cattle
(*Bos taurus*)

Worm
(*Caenorhabditis elegans*)

Zebrafish
(*Danio rerio*)

12

```
sp|P17861|XBP1_HUMAN   MVVVAAAPNPADGTPKVLLLSGQPASAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK   60
sp|Q3SZZ2|XBP1_BOVIN   MVVVAPAQSPAAGAPKVLLLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK   60
                       ***** *  .**  *:********::: .****:***:***  *:*****.***  ****

sp|P17861|XBP1_HUMAN   RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR   120
sp|Q3SZZ2|XBP1_BOVIN   RQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLENQLLR   120
                       ************************************************************

sp|P17861|XBP1_HUMAN   EKTHGLVVENQELRQRLGMDALVAEEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL   180
sp|Q3SZZ2|XBP1_BOVIN   EKTHGLVVENQELRQRLGMDALVTEEEAETKGNGAGLVAGSAESAALRLRAPLQQVQAQL   180
                       ***********************:****:***    *********************** 

sp|P17861|XBP1_HUMAN   SPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNALPQSLPAWRSSQRSTQKDPVP   240
sp|Q3SZZ2|XBP1_BOVIN   SPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDVLPQSLPAWSSSQKWTQKDPVP   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_CAEEL   ----------MSNYPKRIYVLPARHVAAPQPQRMAPKRALP---TEQVVAQLLGDDMGPS   47
tr|Q8UVQ5|Q8UVQ5_DANRE   MVVVT---AGTGGAHKVL-LISGKQSASTGAAQGGYSRSISVMIPNQASSDSDSTTSG-P   55
sp|P17861|XBP1_HUMAN     MVVVAAAPNPADGTPKVL-LLSGQPASAAGAPAG---QALPLMVPAQRGASPEAASGGLP   56
sp|Q3SZZ2|XBP1_BOVIN     MVVVAPAQSPAAGAPKVL-LLSGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP   56
                          *  :  :: .:   ::         :::    *   *:.  .

tr|G5EE07|G5EE07_CAEEL   GPRKRERLNHLSQEEKMDRRKLKNRVAAQNARDKKKERSAKIEDVMRDLVEENRRLRAEN   107
tr|Q8UVQ5|Q8UVQ5_DANRE   PLRKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKAKMGELEQQVLELELENQKLHVEN   115
sp|P17861|XBP1_HUMAN     QARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN   116
sp|Q3SZZ2|XBP1_BOVIN     TARKRQRLTHLSPEEKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN   116
                          ***:**.***.*** ********** ***:* : . ::*: : .* **::*  **

tr|G5EE07|G5EE07_CAEEL   ERLRRQNKNLMNQQNESVMYMEENNENLMNSNDACIYQNVVYEEEVVGEVAPVVVVGGED   167
tr|Q8UVQ5|Q8UVQ5_DANRE   RLLRDKTSDLLSENEELRQRLGL--DTLETKEQVQVLE------SAVSDLG--LVTGSSE   165
sp|P17861|XBP1_HUMAN     QLLREKTHGLVVENQELRQRLGM--DALVAEEEAE--------AKGNEVR--PVAGSAE   163
sp|Q3SZZ2|XBP1_BOVIN     QLLREKTHGLVVENQELRQRLGM--DALVTEEEAE--------TKGNGAG--LVAGSAE   163
                          . ** .:  *: ::*  :  *     * :  :  *  : : * . .    *.*. :

tr|G5EE07|G5EE07_CAEEL   RRAFESAAFINEPQQWEQARSTSINNNISNQLRRMDSKKNNTISVDMYLTIISILCNHMD   227
tr|Q8UVQ5|Q8UVQ5_DANRE   SAAL----RLRVPPQQVQAQQSPNLKTSPWILTALALQTLSLISCLVFWTSLTPSSSSRQ   221
sp|P17861|XBP1_HUMAN     SAAL----RLRAPLQQVQAQLSPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNA   219
sp|Q3SZZ2|XBP1_BOVIN     SAAL----RLRAPLQQVQAQLSPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDV   219
                          *:         :. * *  **:  :       *  :  :  *      :  *

tr|G5EE07|G5EE07_CAEEL   RNKKMDTSNKSSNISRAQAESSIDSLLATLRKEQTVMQRLVQADPCTHLQKRVKHFRRIP   287
tr|Q8UVQ5|Q8UVQ5_DANRE   TFLKHRSLSRSSCWWGVQESKYLPPHLQLWGPHQLSWKPLMN------------------   263
sp|P17861|XBP1_HUMAN     LPQSLPAWRSSQRSTQKDPVPYQPPFLCQWGRHQPSWKPLMN------------------   261
sp|Q3SZZ2|XBP1_BOVIN     LPQSLPAWSSSQKWTQKDPVPYRPPLLHPWGRHQPSWKPLMN------------------   261
                          .    :   *.      :*    *   .:  * . *:::
```

Domain[i]	70 – 133	64	bZIP PROSITE-ProRule annotation	P17861 (human)
Domain[i]	70 – 133	64	bZIP PROSITE-ProRule annotation	Q3SZZ2 (cattle)
Domain[i]	61 – 117	57	BZIP (basic-leucine zipper) InterPro annotation	G5EE07 (worm)
Domain[i]	69 – 132	64	BZIP (basic-leucine zipper) InterPro annotation	Q8UVQ5 (zebrafish)

Pablo Mier
munoz@uni-mainz.de

JG|U imb Institute of Molecular Biology

13

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?

b. How could that have happened in evolution?

c. How could we check this?

13

```
P11582 GLBE_CHITH      1    -MKFIILALCVA--AASALSGDQIGLVQSTYGKVKGDSVGILYAVFKADPTIQAAFPQFV    57
P02226 GLB7A_CHITH     1    MKFFAVLALCIVGAIASPLSADQAALVKSTWAQVRNSEVEILAAVFTAYPDIQARFPQFA    60
                            * :****:.    ** **.** .**:**:.:*. ..* ** ***.* * *** ****.

P11582 GLBE_CHITH      58   GKDLDAIKGGAEFSTHAGRIVGFLGGVIDD------LPNIGKHVDALVATHKPRGVTHAQ   111
P02226 GLB7A_CHITH     61   GKDVASIKDTGAFATHAGRIVGFVSEIIALIGNESNAPAVQTLVGQLAASHKARGISQAQ   120
                            ***: :**   . *:*********:. :*     *  . * *.*:** **::***

P11582 GLBE_CHITH      112  FNNFRAAFIAYLKGHVDYTAAVEAAWGATFDAFFGAVFAKM                      152
P02226 GLB7A_CHITH     121  FNEFRAGLVSYVSSNVAWNAAAESAWTAGLDNIFGLLFAAL                      161
                            **:***.:::*:...* :.**.*:** * :* :** :** :
```

a. Extra region in P02226: "ALIGNESN".

b. Deletion in P11582, or insertion in P02226.

c. Looking for orthologs for each protein, and checking whether they have the region or not.