

---

# Master Biomedizin 2017

- 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?
- b.** How could that have happened in evolution?

## 8

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

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



## 9

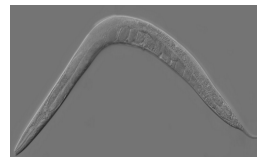
- a.** Both “P17861” (XBP1\_HUMAN) and “Q3SZZ2” (XBP1\_BOVIN) are “X-box binding protein 1” proteins. Can you detect which region/s of these proteins is/are important for their function? Why? Use Clustal Omega.  
What should you do to detect them?
- b.** Add the proteins “G5EE07” (G5EE07\_CAEEL) and “Q8UVQ5” (Q8UVQ5\_DANRE) to the study in 8a. Are you able to identify that region/s now? Why? Use Clustal Omega.



Human  
(*Homo sapiens*)



Cattle  
(*Bos taurus*)



Worm  
(*Caenorhabditis elegans*)



Zebra fish  
(*Danio rerio*)

9

```

sp|P17861|XBP1_HUMAN      MVVVAAAPNPADGTPKVL LLSGQPA SAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK 60
sp|Q3SZ22|XBP1_BOVIN     MVVVAPAQSPAAGAPKVL LLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK 60
*****
sp|P17861|XBP1_HUMAN      RQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
sp|Q3SZ22|XBP1_BOVIN     RQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLENQLLR 120
*****
sp|P17861|XBP1_HUMAN      EKTHGLVVENQELRQRLGMDALVAEEAEAKGNEVRPVAGSAESAALRLRAPLQQVQAQL 180
sp|Q3SZ22|XBP1_BOVIN     EKTHGLVVENQELRQRLGMDALVTEEEAETKGNAGLVAGSAESAALRLRAPLQQVQAQL 180
*****
sp|P17861|XBP1_HUMAN      SPLQNI SPWILAVLT LQIQSLISCFWTTWTQSCSSNALPQSLPAWRSSQSTQKDPVP 240
sp|Q3SZ22|XBP1_BOVIN     SPLQNI SPWTLALT LQTLSTSCWAFCTWTQSCSSDVLPLQSLPAWRSSQKWTQKDPVP 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-LISGQKSAAGGYSRISVMIPNQASSDSDSTTS-G-P 55
sp|P17861|XBP1_HUMAN     MVVVAAAPNPADGTPKVL-LLSGQPA SAAGAPAG---QALPLMVPAQRGASPEAASGGLP 56
sp|Q3SZ22|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|Q3SZ22|XBP1_BOVIN     QARKRQRLTHLSPEEKALRRKLN RVAAQTARDRKKARMSELEQQVVDLEENQKLLLEN 116
*****
tr|G5EE07|G5EE07_CAEL    ERLRRQNKNLMMNQNESMYMEENNENLMNSNDACIQNVVYEEEVVGEVAPVVVGGED 167
tr|Q8UVQ5|Q8UVQ5_DANRE   LLLRDKTSDLLSENEELRLGL--DTLETKEQVQVLE-----SAVSDLG--LVTGSSE 165
sp|P17861|XBP1_HUMAN     QLLREKTHGLVVENQELRQLGM--DALVAEEAE-----AKGNEVR--PVAGSAE 163
sp|Q3SZ22|XBP1_BOVIN     QLLREKTHGLVVENQELRQLGM--DALVTEEEAE-----TKGNAG--LVAGSAE 163
*****
tr|G5EE07|G5EE07_CAEL    RRAFESAAFINEPQQWEQARSTSIINNINISNQLRRMDSKKNNTISVDMYLTIIISILCNHMD 227
tr|Q8UVQ5|Q8UVQ5_DANRE   SAAL----RLRVPPQVQVQASPNLKTSPWILTALALQTLISLISCLVFWTSLTPSSSSRQ 221
sp|P17861|XBP1_HUMAN     SAAL----RLRAPLQQVQAQLSPLQNI SPWILAVLT LQIQSLISCFWTTWTQSCSSNA 219
sp|Q3SZ22|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|Q3SZ22|XBP1_BOVIN     LPQSLPAWRSSQKWTQKDPVPYRPPLLHPWGRHQPSWKPLMN----- 261
*****

```

Domain <sup>i</sup>	70 – 133
Domain <sup>i</sup>	70 – 133
Domain <sup>i</sup>	61 – 117
Domain <sup>i</sup>	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*), mouse (*Mus musculus*), chicken (*Gallus gallus*), zebra fish (*Danio rerio*), purple sea urchin (*Strongylocentrotus purpuratus*) and honey bee (*Apis mellifera*). Choose reviewed entries whenever possible.
- 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*)



Zebra fish  
(*Danio rerio*)



Sea urchin  
(*Strongylocentrotus purpuratus*)



Honey bee  
(*Apis mellifera*)





## 11

- 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*).
- 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

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

```

```

420 QQQQQQQQQQQPPPP-PQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
410 QQQQQQQQQQP-PPPPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 465
412 QQQQQQQQPQPQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 466
418 QQQQQQPQQQ-PP-PP-PPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 469
412 QQQQP-----QP-PPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 463
420 QQQQPPPPSQAPQPQPQPQP-QPQPQARQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
419 PQQQPPQPQPQPQP-----PQPPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 470
*** ** * : * * * * * : * *

```