

---

# Master Biomedizin 2016

MSA

## 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*)

# Homology + MSA

\*Images from: UniProt, TCoffee

9

■	Your list...3Z0VB	Entry	Entry name	Protein names	Gene names	Organism
<input type="checkbox"/>	Q90WY9	Q90WY9	Q90WY9_XENLA	Survival motor neuron		Xenopus laevis
<input type="checkbox"/>	Q16637	Q16637	SMN_HUMAN	Survival motor neuron protein	SMN1 SMN, SMNT SMN2 SMNC	Homo sapiens
<input type="checkbox"/>	P97801	P97801	SMN_MOUSE	Survival motor neuron protein	Smn1 Smn	Mus musculus
<input type="checkbox"/>	Q98SU9	Q98SU9	Q98SU9_CHICK	Survival motor neuron protein	SMN	Gallus gallus
<input type="checkbox"/>	Q9W6S8	Q9W6S8	SMN1_DANRE	Survival motor neuron protein 1	smn1 smn	Danio rerio (Zebrafish)
<input type="checkbox"/>	W4XFQ6	W4XFQ6	W4XFQ6_STRPU	Uncharacterized protein	Sp-Smn1	Strongylocentrotus purpuratus
<input type="checkbox"/>	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 <sup>i</sup>	91 – 151	61	Tudor  PROSITE-ProRule annotation	

Region

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

tr	Q90WY9	Q90WY	MAGL	-----	EDGGEVLFRRGAGQSD	-----	DSDIWDDTALIKAYDKAVSFKRAL	KNE	CTI
sp	Q16637	SMN_H	MAMSSGGSGGGVPEQ	EDSVLFRRTGTGQSD	-----	DSDIWDDTALIKAYDKAVASFKHAL	KNGD	ICE	
sp	P97801	SMN_M	MAMSGSGGAG	-----	SEQEDTVLFRRTGTGQSD	-----	DSDIWDDTALIKAYDKAVASFKHAL	KNGD	ICE
tr	Q98SU9	Q98SU	MA	-----	GRVLFRRTGTGQSD	-----	DSDMWDDTALIKAYDKAVASFKNAL	KNGD	CSE
sp	Q9W6S8	SMN1	MAN	-----	GAEDVVFRCRTGTGQSD	-----	DSDIWDDTALIKAYDKAVASFKNAL	KNGD	GAT
tr	W4XFQ6	W4XFQ	MAE	-----	RSQDVVFRVDR	SDN	DSDIWDDTALIKAYDKAISYVKGMI	KNGD	GE
tr	A0A088A467	A	MAD	-----	DNVLFIRNGNST	HS	EDVWDDTALIKAYDKAINKAKEE	VKG	MG
cons			**				*****		*
tr	Q90WY9	Q90WY	GAETEEKNPRTKR	KNNKKNRSRKKCNAA	KKWRIGDTCNAVWSE	DGNIYSATISSIDAKRR			
sp	Q16637	SMN_H	TSGK	--PKTTPKR	KPAKKNKSQKKNTAAS	LQWVKVGDKCAIWS	DGCIYPATIASIDFKRE		
sp	P97801	SMN_M	TPDK	--PKGTARR	KPAKKNKSQKKNTAT	LQWVKVGDKCAIWS	DGCIYPATIASIDFKRE		
tr	Q98SU9	Q98SU	PSDKQEQRAGVKR	KNSKKNRNKNSNAVPL	KQWVKVGDSCNAVWSE	DGNIYPATIASINLKR			
sp	Q9W6S8	SMN1	PQEN	--DNPQKKR	KNNKKNKSQKKCNAA	DKWVQVGDSCYAFWS	DGNIYPATIASIDFKRE		
tr	W4XFQ6	W4XFQ	AR	--SKP	--KR	KRGKKKK	NKKNLVPS	DKWVKVGDSCVFTED	EQVYSAVVKAINHKKT
tr	A0A088A467	A	G	-----	IKNSECKQLQKQSKLT	SKPYK	KWIVGAPCRAVYSE	DGEIYEAITIKIYENNG	
cons							*****		*
tr	Q90WY9	Q90WY	TCIVVYSGYGNSEEHSLADLR	FDTSEAESDQRDQEQEINGD	-----	EH	STDESD		
sp	Q16637	SMN_H	TCVVYTYGNGNREEQNLSDLLS	PTCEVANNIEQNAQENENES	-----	QV	STDESE		
sp	P97801	SMN_M	TCVVYTYGNGNREEQNLSDLLS	PTCEVANSTEQNTQ	-----	QV	STDDSE		
tr	Q98SU9	Q98SU	TCVVYTYGNGNKEEQNLADLLP	ASDET	-----	PE	STDESE		
sp	Q9W6S8	SMN1	TCVVFYTYGNEEQNLSDLLP	PDMDDEDALKT	-----	ANVKET	ES	STDESD	
tr	W4XFQ6	W4XFQ	SCIVRYTYGNEEQNLSDLLS	SEAEATSVASVNS	-----	KAELEN	GYDSMEWTD		
tr	A0A088A467	A	SCIVKFVGYGNTKELHSLLE	SLGLQSQIAQQK	-----	NAMEH	KFDEENEEISESFN	STMMNS	
cons			*****				*****		*
tr	Q90WY9	Q90WY	RSRSRSHQSKDPQNRTPKS	-----	SQWNGQFPPV	-----	PPPFMPGF	-----	GRH-GEKLDQA
sp	Q16637	SMN_H	NSRSPGNKS	-----	DNIPKPS	-----	APWNSFLPPP	-----	PP--MPGRLPGKPK-GLKFNQGP
sp	P97801	SMN_M	HSSRSLRKS	-----	AHKSKA	-----	APWTSFLPPP	-----	PP--MPGSGLPGKPK-GLKFNQGP
tr	Q98SU9	Q98SU	KSSQSHHNN	-----	NCTKARF	-----	SPKNLRFP	-----	PT--PPGL--GRH-GSKFRTL
sp	Q9W6S8	SMN1	RSFTPKQSGHAKHKS	SNFPMGPPSWFSPFP	-----	PPPPPPHF	-----	KKMDGRRGEGP	
tr	W4XFQ6	W4XFQ	HSOSPMHPSG	-----	SGAGPRK	-----	RSHHPPPPHQP	PPPPHPS	MTN-P--LGYT
tr	A0A088A467	A	RKYS-GEKM	-----	DCSEET	-----	KTYKHQFMP	-----	GTSFN
cons							*****		*
tr	Q90WY9	Q90WY	-----	HPFLSGWPP	-----	PFLPGPMIP	PPPPPMSP		
sp	Q16637	SMN_H	PPPPPPPHLLSCWLP	-----	PFPSGPMIP	PPPPPMSP			
sp	P97801	SMN_M	PPPPPLPPPPFLPCWMP	-----	PFPSGPMIP	PPPPPMSP			
tr	Q98SU9	Q98SU	-----	PPFLSCWPP	-----	PFAPGPMIP	PPPPPMSP		
sp	Q9W6S8	SMN1	-----	GPSFPGWPP	-----	MPLPGPMIP	PPPPPMSP		
tr	W4XFQ6	W4XFQ	-----	SPYPGSWYPPHQA	PPPPPMSP	PLPFA	PPWGSAAQRI	SMPTT	PPPHPLPE
tr	A0A088A467	A	-----	VMADT	-----	MPAP	-----	PLPPQLMA	
cons							*****		*
tr	Q90WY9	Q90WY	D-AC	EDDEALGSMILISWYMSGYHTGYLGLKQGRM	ESSI	-----			
sp	Q16637	SMN_H	D-S	DDADALGSMILISWYMSGYHTGYMGRONK	EGRC	-----			
sp	P97801	SMN_M	D-S	DDTALGSMILISWYMSGYHTGYMGRONK	EGKC	-----			
tr	Q98SU9	Q98SU	D-S	EDDEALGSMILISWYMSGYHTGYLGLKQSR	MAAL	-----			
sp	Q9W6S8	SMN1	D-FG	EDDEALGSMILISWYMSGYHTGYMGLRQGR	EA	-----			
tr	W4XFQ6	W4XFQ	D-L	LEMDKEALHSMILISWYMSGYHTGYEGMKSK	SSHSATS	SKPKEDGSTPRREQGRQTRTR			
tr	A0A088A467	A	KLPA	NDAEALSSMLMSWYISGFHTGYHLKQAE	NOTKR	-----			
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	DKVRETKEPAP	SOPAEESQDSNQET	-----				
tr	A0A088A467	A	-----	KNC	-----				
cons							*****		*

## 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

\*Images from: UniProt

10

■	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 = 4 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 MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 ATG-ANSLGKVMPTKSPPPPGGGNLG MNSRKPDRLRVVIPPSSKGMMPPL ----- N T Q 289
241 NTG-ANSVGKVMPTKSPPPPGGGSVG MNSRKPDRLRVVIPPSSKGMMPPL ----- N A Q 291
239 NTG-ANSLGKVMPTKSPPPPGGSLG MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N A Q 297
241 TTG-ANSLGKVMPTKSPPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL ----- N T Q 291
241 TTG-ANSLGKVMPTESPPPPGGGSLG MNSRKPDRLRVVIPPSSKGMMPPL S E E E E L E L N T Q 299
239 TAGGGNGLGKVMPTKSPPPPGGGSLG MNNRKPDRLRVVIPPSSKGMMPPL T E E D E L E L N T Q 298
: * . * . : * * * * * : * * * * * : * * . * * * * * : * *

```

```

420 QQQQQQQQQQQPPPP-PQPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
410 QQQQQQQQQP-----PPPPQPQPQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 465
412 QQQQQQPQQPPPPPPQ-----PQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 466
418 QQQQQQPQQ-----PP-----PQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 469
412 QQQQP-----QP-PPPPQPQPQPQEVGRSPVDSLSSSSSSSYDGSREDPRGDFH 463
420 QQQPPPPSQAPQPQPQPQP-QPQPQARQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 478
419 PQQPQPQPQPQP-----PQPQPQEMGRSPVDSLSSSSSSSYDGSREDPRGDFH 470
*** ** * * * * * * * * * * * * * * * * * *

```

## 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?

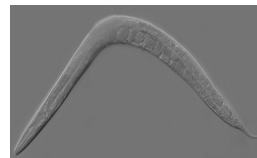
**b.** Add the proteins “G5EE07” (G5EE07\_CAEEL) and “Q8UVQ5” (Q8UVQ5\_DANRE) to the study in 12a. Are you able to identify that region/s now? Why? Use Clustal Omega.



Human  
(*Homo sapiens*)



Cattle  
(*Bos taurus*)



Worm  
(*Caenorhabditis elegans*)



Zebrafish  
(*Danio rerio*)

12

```

sp|P17861|XBP1_HUMAN      MVVVAAPNPADGTPKVL LLSGQPA SAAGAPAGQALPLMVPAQRGASPEAASGGLPQARK 60
sp|Q3SZZ2|XBP1_BOVIN     MVVVAQAQSPAAGAPKVL LLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPPQARK 60
*****
*****

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

sp|P17861|XBP1_HUMAN      EKT HGLVVENQELRQRLGMDALVAEEEEAEKGNVVRPVAGSAESAALRLRAPLQQVQAQL 180
sp|Q3SZZ2|XBP1_BOVIN     EKT HGLVVENQELRQRLGMDALVTEEEAETKGNAGLVAGSAESAALRLRAPLQQVQAQL 180
*****
*****

sp|P17861|XBP1_HUMAN      SPLQNI SPWILAVLT LQIQSLISCWAFWTTWTQSCSSNALPQSLPAWRSSQSTQKDPVP 240
sp|Q3SZZ2|XBP1_BOVIN     SPLQNI SPWTLMA LTLQTLSTSCWAFCSWTQSCSSDVL PQLPAWSSSQKWTQKDPVP 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-LISGKQSAAGAAQGGYRSISVMIPNQASSDSDSTTSG-P 55
sp|P17861|XBP1_HUMAN     MVVVAAPNPADGTPKVL-LLSGQPA SAAGAPAG---QALPLMVPAQRGASPEAASGGLP 56
sp|Q3SZZ2|XBP1_BOVIN     MVVVAQAQSPAAGAPKVL-LLSGQPAATGGAPAG---RALPVMVPGQQGASPEGASGVPP 56
*****
*****

tr|G5EE07|G5EE07_CAEL    PRKRERLNHL S QEEKMDRRKLN RVAAQNARDKKERSAKIEDVMRDLVEENRRLRAEN 107
tr|Q8UVQ5|Q8UVQ5_DANRE   PLRKQRRLTHLSPEEKALRRKLN RVAAQTARDRKAKMGELEQQVLELEENQKLHVEN 115
sp|P17861|XBP1_HUMAN     ARKRQRRLTHLSPEEKALRRKLN RVAAQTARDRKAKMSELEQQVVDLEENQKLLLEN 116
sp|Q3SZZ2|XBP1_BOVIN     ARKRQRRLTHLSPEEKALRRKLN RVAAQTARDRKAKMSELEQQVVDLEENQKLLLEN 116
*****
*****

tr|G5EE07|G5EE07_CAEL    ERLRRQKNLNMNQNESVMYMEENNENLMNSNDACIQNVVYEEEVVGEVAPVVVGGED 167
tr|Q8UVQ5|Q8UVQ5_DANRE   RLLRDKTSDLLSENEELRQLGL-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    RRAFESAAFINPEQQWEQARSTSIINNINISNQLRRMDSKKNNTISVDMYLTIIISILCNHMD 227
tr|Q8UVQ5|Q8UVQ5_DANRE   SAAL----RLRVPPQVQAQSPNLKTPWILTALALQTLISLISCLVFWTSLTPSSSSRQ 221
sp|P17861|XBP1_HUMAN     SAAL----RLRAPLQQVQAQLSPLQNI SPWILAVLT LQIQSLISCWAFWTTWTQSCSSNA 219
sp|Q3SZZ2|XBP1_BOVIN     SAAL----RLRAPLQQVQAQLSPLQNI SPWTLMA LTLQTLSTSCWAFCSWTQSCSSDV 219
*****
*****

tr|G5EE07|G5EE07_CAEL    RNKKMDTSNKSSNISRAQAESSIDSLLATLRKEQTMQRLVQADPCTHLOKRVKHFRRIP 287
tr|Q8UVQ5|Q8UVQ5_DANRE   TFLKHSLSRSSCWGVQESKYLPPHLQLWGPHQLSWKPLMN----- 263
sp|P17861|XBP1_HUMAN     LPQSLPAWRSSQSTQKDPVPYQPPFLCQWGRHQPSWKPLMN----- 261
sp|Q3SZZ2|XBP1_BOVIN     LPQSLPAWSSSQKWTQKDPVPYRPPLLHPWGRHQPSWKPLMN----- 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		P17861 (human)
64	bZIP	PROSITE-ProRule annotation		Q3SZZ2 (cattle)
57	BZIP (basic-leucine zipper)	InterPro annotation		G5EE07 (worm)
64	BZIP (basic-leucine zipper)	InterPro annotation		Q8UVQ5 (zebrafish)

## 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  GKDLDAIKGGAEFSTHAGRIVGFLGGVTD-----LPNIGKHVDALVATHKPRGVTHAQ      111
P02226 GLB7A_CHITH     61  GKDVASIKDTGAFATHAGRIVGFVSEITALIGNESNAPAVQTLVGQLAASHKARGISQAQ      120
      ***: :** . *:*****:..* :* :* :* :* :* :* :* :* :* :* :* :* :*
      ***: :** . *:*****:..* :* :* :* :* :* :* :* :* :* :* :* :*

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

```

- Extra region in P02226: "ALIGNESN".
- Deletion in P11582, or insertion in P02226.
- Looking for orthologs for each protein, and checking whether they have the region or not.