Master Biomedizin 2019

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



3) MSA



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 P02226 GLB7A_CHITH	1 1	-MKFIILALCVAAASALSGDQIGLVQSTYGKVKGDSVGILYAVFKADPTIQAAFPQFV MKFFAVLALCIVGAIASPLSADQAALVKSTWAQVRNSEVEILAAVFTAYPDIQARFPQFA * :****:. ** **.** .**:**:* ** ***.* * ***	57 60
P11582 GLBE CHITH P02226 GLB7A_CHITH	58 61	GKDLDAIKGGAEFSTHAGRIVGFLGGVIDDLPNIGKHVDALVATHKPRGVTHAQ GKDVASIKDTGAFATHAGRIVGFVSEIIALIGNESNAPAVQTLVGQLAASHKARGISQAQ ***: :** *:********:. :*	111 120
P11582 GLBE CHITH P02226 GLB7A_CHITH	112 121	FNNFRAAFIAYLKGHVDYTAAVEAAWGATFDAFFGAVFAKM FNEFRAGLVSYVSSNVAWNAAAESAWTAGLDNIFGLLFAAL **:***::*:.::* :.**.*:* : :* :** :	152 161

a. Extra region in P02226: "IGNESN".



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	<pre>sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN</pre>	MVVVAAAPNPADGTPKVLLLSGOPASAA MVVVAPAQSPAAGAPKVLLLSGOPASAA **********************************	GGAPAGRALPVMVPGQQGASPEGASGVP .****:***:****************************	PQARK 60 ***** 60 NQLLR 120 NQLLR 120 VQAQL 180 VQAQL 180 ***** 180 KDPVP 240	a. No. They are too similar. We would protein from a more distant organism.	need a
	sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN	YQPPFLCQWGRHQPSWKPLMN 261 YRPPLLHPWGRHQPSWKPLMN 261				
		* *** * ******	tr G5EE07 G5EE07_CAEEL tr Q8UVQ5 Q8UVQ5_DANRE sp P17861 XBP1_HŪMAN sp Q3SZZ2 XBP1_BOVIN	MVVVT AGTGGA MVVVAAAPNPADGT	PKRIYVLPARHVAAPQPQRMAPKRALPTEQVVAQLLGDDMGPS HKVL-LISGKQSASTGAAQGGYSRSISVMIPNQASSDSDSTTSG-P PKVL-LLSGQPASAAGAPAGQALPLMVPAQRGASPEAASGGLP PKVL-LLSGQPAATGGAPAGRALPVMVPGQQGASPEGASGVPP *:::::::::::::::::::::::::::::::::::	47 55 56 56
 b. Yes. They are not as simila zipper) domain in positions: 70-133 (human) 70-133 (cattle) 61-117 (worm) 		r. bZIP (basic-leucine	tr G5EE07 G5EE07_CAEEL tr Q8UVQ5 Q8UVQ5_DANRE sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN	GPRKRERLNHLSGE PLRKRQRLTHLSPE QARKRQRLTHLSPE QARKRQRLTHLSPE	EKMDRRKLKNRVAAQNARDKKKERSAKIEDVMRDLVEENRRLRAEN EKALRKLKNRVAAQTARDRKKAKMGELEQQVLELELENQKLHVEN EKALRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN EKALRRKLKNRVAAQTARDRKKARMSELEQQVVDLEEENQKLLLEN ** **************	107 115 116 116
			tr G5EE07 G5EE07_CAEEL tr Q8UVQ5 Q8UVQ5_DANRE sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN	ERLRRQNKNLMNQQI RLLRDKTSDLLSEN QLLREKTHGLVVEN	NESVMYMEENNENLMNSNDACIYQNVVYEEEVVGEVAPVVVVGGED EELFORLGLDTLETKEQVQVLESAVSDLGLVTGSSE QELFORLGMDALVAEEAEAKGNEVRPVAGSAE QELFORLGMDALVTEEAETKGNGAGLVAGSAE	167 165 163 163
69	9-132 (zebrafish)		tr G5EE07 G5EE07_CAEEL tr Q8UVQ5 Q8UVQ5_DANRE sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN	SAAL RLRVPP(SAAL RLRAPL(SAAL RLRAPL(QWEQARSTSINNNISNQLRRMDSKKNNTISVDMYLTIISILCNHMD QQVQAQQSPNLKTSPWILTALALQTLSLISCLVFWTSLTPSSSSRQ QQVQAQLSPLQNISPWILAVLTLQIQSLISCWAFWTTWTQSCSSNA QQVQAQLSPLQNISPWTLMALTLQTLSLTSCWAFCSTWTQSCSSDV * **::::::::::::::::::::::::::::::::::	227 221 219 219
			tr G5EE07 G5EE07_CAEEL tr Q8UVQ5 Q8UVQ5_DANRE sp P17861 XBP1_HUMAN sp Q3SZZ2 XBP1_BOVIN	TFLKHRSLSRSSCW LPQSLPAWRSSQRS	SRAQAESSIDSLLATLRKEQTVMQRLVQADPCTHLQKRVKHFRRIP WGVQESKYLPPHLQLWGPHQLSWKPLMN TQKDPVPYQPPFLCQWGRHQPSWKPLMN TQKDPVPYRPPLLHPWGRHQPSWKPLMN : * .* : *::	287 263 261 261
	Domain ¹	70 – 133 64 bZIF	PROSITE-ProRule annotation 🔻		P17861 (human)	
	Domain ⁱ	70 – 133 64 bZI	P 💞 PROSITE-ProRule annotation 👻		Q3SZZ2 (cattle)	
	Domain ¹	61 – 117 57 BZIP	(basic-leucine zipper) 🖋 InterPro annotation 👻		G5EE07 (worm)	
	Domain ¹	69 – 132 64 BZI	P (basic-leucine zipper) & InterPro annotation 🔻		Q8UVQ5 (zebrafish)	

*) Homology + MSA



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*)



-	Marca 11 - 4 - 0 - 0 - 0 - 0 - 0								
	Your list:3Z0VB	Entry 🔻	Entry name 븆		Protein names 🗣 🛛 🗎	Gene names 🗣	Organism 🗣	tr Q90WY9 Q90WY sp Q16637 SMN_H sp P97801 SMN_M	MAGLEDGGEVLFRRGAGQSD -DSDIWDDTALIKAYDKAVSSFKRALKNE MAMSSGGSGGGVPEQEDSVLFRRGTGQSD -DSDIWDDTALIKAYDKAVASFKHALKNG
	Q90WY9	Q90WY9	Q90WY9_XENLA		Survival motor neuron		Xenopus laevis (Afric	sp P97801 SMN_M tr Q98SU9 Q98SU sp Q9W6S8 SMN1 tr W4XFQ6 W4XFQ	MAMGSGGAGSEQEDTVLFRAGTGQSD -DSDIWDDTALIKAYDKAVASFKHALKNU MAMGSGGAGSEQEDTVLFRAGTGQSD -DSDIWDDTALIKAYDKAVASFKHALKNU MANGRULFRAGAGQSD -DSDIWDDTALIKAYDKAVASFKNALKNU MANGRUVFCRGTGQSD -DSDIWDDTALIKAYDKAVASFKNALKNU MADDNVLFIRGNGNSTEHSEDVWDDSALIKAYDKAINLAKEWVGKF
	Q16637	Q16637	SMN_HUMAN	☆	Survival motor neuron protein	SMN1 SMN, SMNT SMN2 SMNC	Homo sapiens (Huma	n tr A0A088A467 A	
								cons	** ************************************
	P97801	P97801	SMN_MOUSE	☆	Survival motor neuron protein	Smn1 Smn	Mus musculus (Mous	e tr Q90WY9 Q90WY sp Q16637 SMN H	GAETEEKNPRTKR-KNNKKNRSRKKCNAPLKKWRIGDTCNAVWSEDGNIYSATISSI TSGKPKTTPKR-KPAKKNKSQKKNTARSLQQWKVGDKCSAIWSEDGCIYPATIASI
	Q98SU9	Q98SU9	Q98SU9_CHICK		Survival motor neuron protein	SMN	Gallus gallus (Chicke	sp 010637 SMN_M n) sp P97801 SMN_M tr 098SU9 098SU sp 09W6S8 SMN1 tr W4XFQ6 W4XFQ	TPDKPKGTARR-KPAKKNKSQKKNATPLKQWKVQDKSAIMSEDGLTPAIIASI PDK-VKGTARR-KPAKKNKSQKKNATPLKQWKVGDSCAVWSEDGNVYPAIIASI PSDKQEQRAGVKR-KNSKKNNKSRKRCNAPDLKQWKVGDSCAVWSEDGNVYPAIIASI PQENDNPGKKR-KNNKKNKSRKRCNAPDKEWQVGDSCYAFWSEDGNLYTATISI ARSKP-KR-KRGKKK-NKKNLV <u>S</u> CTKWKVGDRCKSVFTEDEQVYSAVVKAI GIKNSECKQKLQKQSKLTSKPYKKWIVGAPCRAVYSEDGEIYEAIITKI
	Q9W6S8	Q9W6S8	SMN1_DANRE	☆	Survival motor neuron protein 1	smn1 smn	Danio rerio (Zebrafish) tr A0A088A467 A	
	W4XFQ6	W4XFQ6	W4XFQ6 STRPU		Uncharacterized protein	Sp-Smn1	Strongylocentrotus pu	cons	
			····· (···					tr Q90WY9 Q90WY	TCIVVYSGYGNSEEHSLADLRFPDTSEAESDORDOEOEINGDEH-S
	A0A088A467	A0A088A467	A0A088A467_APIMI	=	Uncharacterized protein	LOC100577491	Apis mellifera (Honey	tr Q985U9 Q985U	TCIVVYSGYGNSEEHSLADLRFPDTSEAESDQRDQEQEINGDEH-S TCVVVYTGYGNREEQNLSDLLSPTCEVANNIEQNAQENENESQV-S TCVVVYTGYGNREEQNLSDLLSPTCEVANSTEQNTQENESQV-S TCVVYTGYGNKEEQNLADLLPPASDET
		↑						tr W4XFQ6 W4XFQ tr A0A088A467 A cons	SCIVRYTGYONEEEKRLSDLFSEEFAETSVASVNS-KAELEN
									·····
	a. 7 sequ	luences.						tr Q90WY9 Q90WY sp Q16637 SMN_H sp P97801 SMN_M tr Q98SU9 Q98SU sp Q9W6S8 SMN1 tr W4XFQ6 W4XFQ	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGPRLGPGKP -GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP -GLK KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGL GRH -GSK RSFTPQKSGHAKHKSKSNFPMGPSWFPSFPPG PPPPPHF KKMDGRR HSOSPMHPSG SGAGPRK RSHHPPPPHOPHPPPPHPSS MTH -P TH -P TH -P TH -P RSHHPPPPHOPHPPPHPSS MTH -P RSHHPPPPHOPHPPPHPSS MTH -P
				they	are conserved in	evolution		sp 016637 SMN_H sp P97801 SMN_M tr 098SU9 098SU	NSRSPGNKS ONIKPKS APWNSFLPPP PP MPGPRLGPGKP -GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP -GLK KSSQSHHNEN NCTKARF SPKNLRFPI PT PPGL GRH -GSK RSFTPQKSGHAKHKSKSNFPMGPPSWFPSFPPG PPPPPHB KKMDGRR HSOSPMHPSG SGAGPRK RSHHPPPPHOPHPPPHPSS WTH -P MTH -P WTH -P
			s, because	they	are conserved in o	evolution.		sp Q16637 SMN_H sp P97801 SMN_M tr Q98SU9 Q98SU sp Q9W6S8 SMN1 tr W4XFQ6 W4XFQ	NSRSPGNKS ONIKPKS APWNSFLPPP PP MPGPRLGPGKP -GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP -GLK KSSQSHHNEN NCTKARF SPKNLRFPI PT PPGL GRH -GSK RSFTPQKSGHAKHKSKSNFPMGPPSWFPSFPPG PPPPPHB KKMDGRR HSOSPMHPSG SGAGPRK RSHHPPPPHOPHPPPHPSS WTH -P MTH -P WTH -P
			s, because	they				sp 016637 SMN H sp P97801 SMN M tr 0985U9 0985U sp 09W658 SMN1 tr W4XF06 W4XFQ tr A0A088A467 A cons	NSRSPONKS ONIKPKS APWNSFLPPP PP MPGPRLGPGKP -GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP -GLK KSSQSHHNEN NCTKARF SPKNLRFPIP PPT PPGL CRH - GSK RSFTPOKSGHAKKSKSNFPMGPSWFPSFPFG PPPPPPHF KKMDGRR HSQSPMHPSG SGAGPRK RSHHPPPPPHQPHPPPPHPSS MTH - P 1 RKYS - GEKM DCESEET KTYKHQFMP PPFL SGWPP PFL PGPPMTPPP
Featu	b. Boxed			they	(Q16637, SMN_HUN			sp (016637 [SMN_H sp P97801 [SMN_M tr (0985U9 [0985U sp (098588 SMN] tr (W4XFQ tr (A0A088A467 [A cons tr (090WY9 [090WY sp (016637 [SMN_M sp [P97801 [SMN_M	NSRSPONKS ONIKPKS APWNSFLPPP PP MPGPRLGPGKP GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP GLK KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGL GRH GSK RSFTPOKSGHAKHSKSNFPMGPPSWFPSFPFG PPPPPHFF KKMDGRR HSQSPMHPSG - SGAGPRK RSHHPPPPPHQPHPPPHPSS MTH - P RKYS - GEKM DCESEET KTYKHQFMP HPFLSGWPP PFLPGPPMIPPPHPPHQPHPPPHPSS MTH - P PPPPPPPHLLSCWLP PFLPGPPMIPPP PPPPPPPPLPPPFLCWMP PFLSGPPIIPPP PPPLSPPP
	b. Boxed	d regions	Length D	escription	(Q16637, SMN_HUN	IAN)		sp 016637 SMN H sp P97801 SMN M tr 0985U9 0985U sp 09W658 SMN1 tr W4XF06 W4XFQ tr A0A088A467 A cons	NSRSPONKS DNIKPKS APWNSFLPPP PP MPGPRLGPGKP GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP GLK KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGL GRH -GSK RSFTPQKSGHAKHKSKSNFPMGPSWEPSFPGG PPPPPHPHF KKMDGRR HSOSPMHPSG SGACPK RSHHPPPPPHPPPHPPPHPSS MTH GTSI RKYS - GEKM DCESEET KTYKHQFMP GTSI HPFLSGWPP PFLPGPPHILLSCWL P PPPPPPPHLLSCWL P PPPPPPPPLPUPPFLPCWPP PPPLPPPFLPCWPP PPFLSCWPP PFLSCWPP MTH SPENDER SFPCSWPPHOAPPPPPPPPPPPPPMPPPPPADAPPPPPP PPFLSCWPP PFPAGPPLIPPP SPPCSWPPHOAPPPPPPPPPPPPPPMPPPPPMSSPLFFAPWGSPAAORIFSMPPTPPPH
	b. Boxed	d regions	Length D	escription	(Q16637, SMN_HUN	IAN)		sp (016637 SMN H sp P97801 SMN M tr (0985U9 0985U sp (098658 SMN1 tr W4XF06 W4XF0 tr A0A088A467 A cons tr (090WY9 090WY sp (016637 SMN H sp P97801 SMN M tr (0985U9 0985U sp (09858 SMN1 tr W4XF06 W4XF0	NSRSPGNKS DNIKPKS APWNSFLPPP PP PMGSGLGPGKP GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP GLKI KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGGL GRH GSKI RSFTPQKSGHAKHKSKSNFPMGPPSWFPSFPPG PPPPPPHF KKMOGRR HSOSPMHPSG SGGRPK GKM GSKI HSOSPMHPSG SGGRPK RSHTPPPPPHQPHPPPPHPSS GTSI GTSI GTSI KYS-GEKM DCESEET KTYKHQFMP PFLSGPPIPPI PPPPPPPHLLSCWLP PFPSGPPIIPPI PPPPPPPPHLLSCWLP PFSGPPIIPPI PPPSGPPIIPPI PPPSGPPIIPPI PPPLSCWPP PFLSCWP PFPAGPPLIPPI GPSFPGWPP PFPAGPPLIPPI GPSFPGMPP STYPGSWYPPH0APPPPMMSPLPFAPWGSPAAORISTSMPTPPMI MTPLGPPMIPPI
Doma	b. Boxed	d regions	Length D	escription	(Q16637, SMN_HUN	IAN)		sp (16637 SMN H sp P97801 SMN M tr (985U9 Q985U sp (99658 SMN1 tr (48476) W4XFQ tr (48476) W4XFQ tr (484884467 A cons tr (990WY9 Q90WY sp (16637 SMN M tr (985U9 Q985U sp (99658 SMN1 tr (48476) W4XFQ tr (484884467 A cons	NSRSPCNKS DNIKPKS APWNSFLPPP PP MPGPRLGPGKP-GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP-GLK KSSQSHNEN NCTKARF SPKNLRFPIP PT PPGL GRH-GSK RSFTPQKSGHAKKSKSNFPMGPPSWEPSFPFG PPPPPPHF KKMDGRR HSOSPMHPSG SGACPRK RSHHPPPPHQPHPPPHQPHPPPHPSS MTH-Q RKYS - GEKM DCESEET KTYKHQFMP GTSI HPFLSGWPP PFLPGPPIPP PPPPPPPPHLLSSCWLP PFPSGPPIIPPP PPPPPPPPHLPPPFLPCWMP PFPSGPPIIPPP PPPPLPPPFLPCWMP PFPSGPPIIPPP GPSFPGWPP PFPLSCWPP SPYPGSWYPPHQAPPPPMPPPPMMSPLPFAPWGSPAAQRIFSMPTPPPH WMADT MPAP PLPI
Doma	b. Boxed	d regions	Length D	escription	(Q16637, SMN_HUN	IAN)		sp (016637 [SMN H sp (97801 [SMN H tr (098509 [Q9850] sp (09858 [SMN1 tr (404884467 [A cons tr (090WY9 [090WY sp (016637 [SMN H sp (016637 [SMN H sp (09850] [Q9850] tr (09858 [SMN1 tr (09858 [SMN1 tr (404884467 [A cons tr (090WY9 [090WY	NSRSPCNKS DNIKPKS APWNSFLPPP PP MPGPRLGPGKP-GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP-GLK KSSQSHNEN NCTKARF SPKNLRFPIP PT PPGL GRH-GSK RSFTPQKSGHAKKSKSNFPMGPPSWEPSFPFG PPPPPPHF KKMDGRR HSOSPMHPSG SGACPRK RSHHPPPPHQPHPPPHQPHPPPHPSS MTH-Q RKYS - GEKM DCESEET KTYKHQFMP GTSI HPFLSGWPP PFLPGPPIPP PPPPPPPPHLLSSCWLP PFPSGPPIIPPP PPPPPPPPHLPPPFLPCWMP PFPSGPPIIPPP PPPPLPPPFLPCWMP PFPSGPPIIPPP GPSFPGWPP PFPLSCWPP SPYPGSWYPPHQAPPPPMPPPPMMSPLPFAPWGSPAAQRIFSMPTPPPH WMADT MPAP PLPI
Domai Regio	b. Boxed	d regions Position(s) 91 – 15:	Length D	escription Idor 🦑 PR	(Q16637, SMN_HUN) OSITE-ProRule annotation v	fian) Graphical view		sp (016637 SMN H sp (97801 SMN M tr (098509 Q9850 sp (09858 SMN1 tr (484506 W4XFQ tr A0A088A467 A cons tr (090WY9 Q90WY sp (016637 SMN M tr (098509 (09850 sp (09858 SMN1 tr (40488A467 A cons tr (090WY9 Q90WY sp (016637 SMN M	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGPGLGPGKP - GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP - GLKI KSSQSHNKEN NCTKARF SPKNLRFPIP PT PPGL GRH - GSKI RSFTPQKSGHAKHKSKSNFPMGPDSWFPSFPPG PPPPPPHF KKMDGRR HSOSPMHPSG SGAGPRK RSHHPPPPHQPHPPPPHPPSS MTH -P RKYS - GEKM DCESEET
Doma Regio	b. Boxed	d regions	Length D	escription	(Q16637, SMN_HUN) OSITE-ProRule annotation v	IAN)		sp (016637 SMN H sp P97801 SMN M tr (0985U9 Q985U sp (09858 SMN1 tr (4040884467 A cons tr (090WY9 Q90WY sp (016637 SMN H tr (09858 SMN1 tr (4040884467 A cons tr (090WY9 Q985U sp (09658 SMN1 tr (4040884467 A cons tr (090WY9 Q90WY sp (016637 SMN H sp P97801 SMN M tr (09850 Q9850 tr (09858 SMN1 tr (09850 Q9850 tr (09858 SMN1 tr (09858 SMN1	NSRSPCNKS DNIKPKS APWNSFLPPP PP MPGPRLGPGKP-GLK HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP-GLK KSSQSHNEN NCTKARF SPKNLRFPIP PT PPGL GRH-GSK RSFTPOKSGHAKKSKSNFPMGPSWEPSFPFG PPPPPHPHFSS MTH-P- RSSSCHMEN DCESEET KTYKHOFMP GTS RKYS-GEKM DCESEET KTYKHOFMP PFLPGPPHPPPPPPPPPPPPPPPPPPPLSS MTH-P- HPFLSGWPP PFLPGPPMPPPPPPPPPPPPPPPPPPPPPPPPFLPC PPPPPPPPPFLSCWLP PFPSGPPIIPPP PPFLSCWPP PFPSGPPIIPPP PPFLSCWPP PFPSGPPIIPPP SPYPGSWYPPHQAPPPPPPPPMSPLFFAPWGSPAAQRIPSMPPPPPH SWYPGSWYPPHQAPPPPPPPPMSPLFFAPWGSPAAQRIPSMPTPPPH SWADALGSMLISWYMSGYHTGYYLGLKQGRMESSI PLP SLP DDFALGSMLISWYMSGYHTGYYMGFRONKKECC D - SLP ODDALGSMLISWYMSGYHTGYYMGFRONKKECC D - SLP ODDALGSMLISWYMSGYHTGYYMGGRONKKECC D - SLP ODDALGSMLISWYMSGYHTGYYMGGRONKKECC D - FGE EDDEALGSMLISWYMSGYHTGYYMGGRONKKECC D - FDEDEALGSMLISWYMSGYHTGYYMGGRONKKEAAL
Doma Regio Featu	b. Boxed	d regions Position(s) 91 – 15:	Length D 61 T Length D	escription Idor 🏈 PR	(Q16637, SMN_HUN) OSITE-ProRule annotation v	fian) Graphical view		sp (016637 SMN H sp (026637 SMN H sp (028509) 02850 sp (02858 SMN1 tr (028509) 02850 tr 4040884467 A cons tr (028509) 02850 sp (02853 SMN H tr (028509) 02850 sp (02858 SMN1 tr (02850 02850 sp (02854 SMN H tr (02850 029650 sp (02854 SMN H tr (02850 029650 SMN H tr (02850 029650 SMN H tr (02850 029650 SMN H tr (02850 02850 SMN H tr (02850 02850 SMN1 tr (02850 02850 SMN1 tr (02850 SMN1 H tr (02850	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGSGLGPGKP - GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP - GLKI KSSQSHHNEN - NCTKARF SPKNLRFPIP PT PPGL GRH - GSKI RSFTPQKSGHAKHKSKSNEPMGPDSWEPSFPPG PPPPPPHF KKMDGRRI HSQSPMHPSG SGAGPRK RSHHPPPPPH0PHPPPPHSS MTH - P I RKYS - GEKM DCESEET KTYKH0FMP PPPPPPPHLLSCWLP PFESGPPIIPPP PPPPPPPHLLSCWLP PFESGPPIIPPP PPPPPPPHLLSCWLP PFESGPPIIPPP SPFSGWPPHQAPPPPMPPPPMMSPLPFAPWGSPAAQRIPSMPTPPPHI SPFGSWYPPHQAPPPPMPPPPMMSPLPFAPWGSPAAQRIPSMPTPPPHI VMADT MPPAP PLPI VMADT PLPI PSF EDDEALGSMLISWYMSGYHTGYYLGLKQGRMESSI D- SIP EDDEALGSMLISWYMSGYHTGYYLGLKQGRMEKSKTS D - AG - EDDEALGSMLISWYMSGYHTGYYLGLKQGRMKE GKC D - SPF EDDEALGSMLISWYMSGYHTGYYLGLKQGRMKKSKTS D - FG - EDDEALGSMLISWYMSGYHTGYYMGFRONKKE GKC D - SPF EDDEALGSMLISWYMSGYHTGYYMGFRONKKE GKC D - LUEMOKEALHAMLSWYMSGYHTGYYMGFRONKKKSKYTSSATSKPKEDGSTPRREOGRK KLPU-NDAEALSSMLMSWYISGFHTGYYMGLRQGRKEAAA D - LUEMCKEALHAMLSWYMSGYHTGYYMGLRQGRKEAAL
Domai Regio	b. Boxed	d regions Position(s) 91 – 15: Position(s)	Length D 61 Tu Length D 32 P	escription Idor 🖋 PR escription L (binding	(Q16637, SMN_HUN) OSITE-ProRule annotation v	fian) Graphical view		sp (016637 [SMN H sp (97801 [SMN M tr (098509 (09850 sp (090658 [SMN1 tr (4040884467 [A cons tr (090WY9 (090WY sp (016637 [SMN H sp (096537 [SMN H sp (09658 [SMN1 tr (098509 (09850 tr (4040884467 [A cons tr (090WY9 (090WY sp (016637 [SMN M tr (098509 (090WY sp (090658 [SMN1 tr (098509 (09850 sp (090658 [SMN1 tr (4040884467 [A cons	D-I-CEDDEALGSMLISWYMSGYHIGYYMGLRUGRKEAAA D-LEEMKEALHSMLMSWYMSGYHIGYYGKKSKTSSHSATSKPKEDGSTPRREOGR KLPU-NDAEALSSMLMSWYISGFHTGYYHGLKQAEKNGTKR-
Domai Regio Featu Region	b. Boxed	d regions Position(s) 91 – 15 Position(s) 13 – 44	Length D 61 Tu Length D 32 P	escription Idor 🖋 PR escription L (binding	(Q16637, SMN_HUN) OSITE-ProRule annotation v site for GEMIN2)	fian) Graphical view		sp (016637 [SMN H sp (97801 [SMN M tr (985U9 (985U) sp (99658 [SMN1 tr (4040884467 [A cons tr (090WY9 (090WY sp (016637 [SMN H sp (97801 [SMN M tr (0985U9 (985U) sp (99658 [SMN1 tr (4040884467 [A cons tr (090WY9 (090WY sp (016637 [SMN M tr (0985U9 (0985U) sp (090658 [SMN1 tr (4040884467 [A cons tr (090WY9 (090WY sp (090658 [SMN1 tr (4040884467 [A cons tr (090WY9 (090WY sp (016637 [SMN H	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGSGLGPGKP - GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PT PPGGL GRH-GSKI KSSQSHHNEN - NCTKARF SPKNLRFPIP PT PPGL GRH-GSKI RSFTPQKSGHAKHKSKSNEPMGPDSWEPSFPPG PPPPPPHF KKMOGRR HSQSPMHPSG SGAGPRK RSHHPPPPPHOPPPPPSS MTH -P I RKYS-GEKM DCESEET KTYKHQFMP PFESGPPIPP PPPPPPPHLLSCWLP PFESGPPIIPPP PPPPPPPHLLSCWLP PFESGPPIIPPP PPPPPPPHLLSCWLP PFESGPPIIPPP SPFSGWPPHQAPPPPMPPPPMMSPLPFAPWGSPAAQRIPSMPTPPPH SPFGGSWYPPHQAPPPPMPPPPPMMSPLPFAPWGSPAAQRIPSMPTPPPH VMADT MPPAP PLP D-AG-EDDEALGSMLISWYMSGYHTGYYLGLKQGRMESSI MPPAP PLP D-GL DDDALGSMLISWYMSGYHTGYYLGLKQGRMESSI
Domai Regio Featu Region	b. Boxed	d regions Position(s) 91 – 15 Position(s) 13 – 44	Length D 61 T Length D 32 P 1113 R	escription ador & PR escription L (binding equired for	(Q16637, SMN_HUN) OSITE-ProRule annotation v site for GEMIN2)	fian) Graphical view		sp (016637 SMN H sp (03603 SMN H sp (038509 03850 sp (03850 03850 tr (038509 03850 tr (0380467 A cons tr (036637 SMN H sp (036637 SMN H sp (03850 03850 tr (03850 03850 tr (03850 03850 tr (0380467 A cons tr (03809 03850 tr (0380467 A cons tr (03850 03850 sp (03850 03850 tr (0380467 A cons tr (03850 03850 tr (0380467 A cons tr (03850 03850 tr (03850 03850 tr (03850 03850 tr (03850 03850 sp (03850 03850 03850 sp (03850 03850 03850 sp (03850 03850 03850 03850 03850 sp (03850 03850	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGSGLGPGKP - GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP - GLKI KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGL GRH - GSKI RSFTPQKSGHAKHKSKSNFPMGPPSWFPSFPPG PPPPPPHF KKMOGRR HSQSPMHPSG SGAGPRK RSHHPPPPHOPPPPPPHSS MTH - P 1 RKYS - GEKM DCESEET KTYKHQFMP GTSI HPFLSGWPP PFPSGPPIIPPP PPPPPPPHLLSCWLP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP GFSFPGWP PFPSGPPIIPPP GFSFPGWP PFPSGPPIIPPP MPPA
Domai Regio Featu Region Region	b. Boxed	d regions Position(s) 91 – 152 Position(s) 13 – 44 97 – 203	Length D 1 61 T Length D 1 Length D 1 1 32 P 1 113 R 2 28 P	escription idor & PR escription L (binding equired for 2 (binding	(Q16637, SMN_HUN DOSITE.ProRule annotation 👻	fian) Graphical view		sp (016637 SMN H sp (016637 SMN M tr (09850) 90850 sp (09858 SMM1 tr (4040884467 A cons tr (090WY9) (090WY sp (016637 SMN H sp (016637 SMN M tr (098509 (09850 sp (09858 SMN1 tr (4040884467 A cons tr (090WY9) (090WY sp (016637 SMN M tr (098509 (09850 tr (A040884467 A cons tr (090WY9) (090WY sp (016637 SMN M tr (4040884467 A cons tr (090WY9) (090WY sp (016637 SMN M tr (09858 SMN1 tr (4040884467 A cons	NSRSPGNKS DNIKPKS APWNSFLPPP PP MPGSGLGPGKP - GLKI HSSRSLRSK AHSKSKA APWTSFLPPP PP MPGSGLGPGKP - GLKI KSSQSHHNEN NCTKARF SPKNLRFPIP PT PPGL GRH - GSKI RSFTPQKSGHAKHKSKSNFPMGPPSWFPSFPPG PPPPPPHF KKMOGRR HSQSPMMPSG SGAGPRK RSHHPPPPPPHOPPPPSS MTH - P I RKYS - GEKM DCESEET KTYKHOFMP PFLPGPPIIPPP PPFLSCWPP PFPSGPPIIPPP PPPPPPPPHLLSCWLP PFPSGPPIIPPP PFPSGPPIIPPP PFPSGPPIIPPP PFLSCWPP PFPSGPPIIPPP PFPSGPPIIPPP SPYFGSWYPPHQAPPPPMPSPLPFAPWGSPAAQRIPSMPTPPH VMADT MPAAP WADT MPAAP CL- DDFALGSMLISWYMSGYHTGYYLGLKOGRMESSI CL- DDFALGSMLISWYMSGYHTGYYLGLKOGRMEAL FG EDDEALGSMLISWYMSGYHTGYYLGLKOGRMEAL FG EDDEALGSMLISWYMSGYHTGYYLGLKOGRMEAL FG EDDEALGSMLISWYMSGYHTGYYLGLKOGRKKSKTSS SHSL N SHSL N SHSL N SHT N



PS: Bioinformatics is not an exact science...

2019 & 2018 (T-Coffee)	2017 (T-Coffee)
tr 090WY9 090WY sp 016637 SMN H sp 016637 SMN H mAMSSG6SGGGVPEGDSVLFRRGTGQSD - DSDIWDDTALIKAYDKAVASFKHALKNEDCTI sp 097801 sp 0985U9 tr 0985U9 mAMSSG5GGGVPEGDSVLFRRGTGQSD - DSDIWDDTALIKAYDKAVASFKHALKNGDICE tr 0985U9 sp 0985U9 sp 0985S1 mAnnon -GRVLFRRGTGQSD - DSDIWDDTALIKAYDKAVASFKHALKNGDICE mAnnon -GRVLFRRGTGQSD - DSDIWDDTALIKAYDKAVASFKHALKGEDGAT tr WAXF06 tr AAA08884671A cons *:*	tr(090WY9)(090WYMAGLEDGGEVLFRRGAGOSDDSDIWDDTALIKAYDKAVSSFKRALKNEDCTIsp(016637)SMN HMAMSSGGSGGVPEOEDSVLFRRGTGOSDDSDIWDDTALIKAYDKAVASFKHALKNGDICEsp(0985U9)(0985U)MAMGSGGAGSEQEDTVLFRRGTGOSDDSDIWDDTALIKAYDKAVASFKHALKNGDICEsp(0985U9)(0985U)MA
tr 090WY9 090WY sp 016637 SMN H sp P97801 SMN H tr 098SU9 098SU9 tr 098SU9 098SU1 tr 04XF06 W4XF06 W4XF	tr 090WY9 090WY sp 016637 SMN H sp 016637 SMN H TSGKPK - TTP - • KRKPAKKNKSOKKNTAASLOOWKVGDKCSAIWSEDGCIYPATIASIDFKRE tr 0985U9 0985U sp 09W658 SMN 1 P- · OENDNPGKKRKNNKKNKSOKKNATTPLKOWKVGDSCNAVWSEDGNYPATIASINLKRG sp 09W658 SMN 1 C - OENDNPGKKRKNNKKNKSKKRCNAAPDKEWVGDSCVAFWSEDGNYTATITSVDOEKG tr 4040884467 A RMGIGI - KNSECKQKLQKQSKLTSK PYKKWIVGAPCRAVYSEDGEIYEAIITKIYENNG
tr 090WY9 090WY sp 016637 SMN.H sp 016637 SMN.H tr TCVVVYTGYGNREEQNLSDLLSPTCEVANNTEQNAQENENES 0V-STDESE sp 1978081 SMN.H tr 0985U TCVVVTGYGNREEQNLSDLLSPTCEVANSTEQNTONES 0V-STDDSE tr 0985U TCVVVTGYGNREEQNLSDLLSPTCEVANSTEQNTONENET PY-STDESE sp 090%58 SMN1 TCVVVTTGYGNREEQNLSDLLTEPPOMDEDALKTANVKET ES-STEESD tr VASTFOGNMEECONLSDLLSPTCEVANSTEQUAGALKTANVKET ES-STEESD SCIVKFVGYGNTEKVELHSLLESEGLOSQIAQQKNAMEHKFDEENEEISESNF-STNMNS cons :*:*: : .*** *: * .* .* *:	cons: *:: :* :* * :::** :* :::** :* :::tr 090WY9 090WYTCIVVYSGYGNSEEHSLADLRFPDTSEAESDORDOEOEINGDEHSTDESDRSSRSHOSKDPQNsp 016637 SMN HTCIVVYTGYGNREEONLSDLLSPICEVANNIEONAOENENESOVSTDESENSRSPGNKSDNsp 097801 SMN MTCVVYTGYGNREEONLSDLLSPTCEVANNIEONTOE NESOVSTDDSEHSSRSLRSKAHtr 098SU9 098SUTCVVYTGYGNREEONLADLLPPASDET
tr 090WY9 090WY RSSRSH0SKDPQNRETPKSSQWNG0FPPVPPPMPGFGRH-GEKLD0A- NSRSPGNKSONIKPKSAPWDSFLPPPPPMPGSGLGPGKP-GLKFNGPP sp P97801 SMN. H sp (016637 SMN_H sp 997803 SMN_H tr 098SU9 098SU HSSRSKAAHSKSKAAPWTSFLPPPPPMPGSGLGPGKP-GLKFNGPP KSSQSHHNENNCTKARFSPKNLRFPIPPTPPGLGRH-GSKFRTL- sp (09W6S8 SMN1 tr A0A088A467 A cons tr 090WY9 090WY HFLSGWPP	cons :*:*: *: *** *: *.* : : : :
tr 090WY9 090WY	tr A0A088A467 A GEKMDCESEETKTYKHQFMPGTSFNVMAD cons *
tr 090WY9 090WY D-AC-EDDEALGSMLISWYMSGYHTGYYLGLKQGRMESSI sp 016637 SMN.H D-SL-DDADALGSMLISWYMSGYHTGYYMGFRQN0KEGRC sp 197801 SMN M D-CL-DDTDALGSMLISWYMSGYHTGYYMGFRQN0KEGRC tr 0985U9 0985U D-SP-EDDEALGSMLISWYMSGYHTGYYMGFRQNKEGKC D-SP-EDDEALGSMLISWYMSGYHTGYYLGLKQSRMEAAL sp 098658 SMN1 D-FG-EDDEALGSMLISWYMSGYHTGYYLGLKQSRMEAAL D-FG-EDDEALGSMLISWYMSGYHTGYYEGMKKSKTSSHSATSKPKEDGSTPRREQGRQTRTR tr W4XFQ6 W4XFQ D-LEEMDKEALHSMLMSWYMSGYHTGYYEGMKKSKTSSHSATSKPKEDGSTPRREQGRQTRTR tr A040884A671A KLPD-NDAEALSSMLMSWYSGYHTGYYEGMKKSKTSCHSATSKPKEDGSTPRREQGRQTRTR	tr 090WY9 090WY sp 016637 SMN H sp P97801 SMN M-LSGWPPPFLPGPPMIPPPPMSPDACE-DDEALGSMLISWYMSGYHTGYYLGLKOGRM -LSCWLPPFPSGPPIIPPPPPICPDSLD-DDDALGSMLISWYMSGYHTGYYMGFRONKK tr 098SU9 098SU sp 09W658 SMN1 tr W4XF06 W4XF0 tr A0A088A467 A-LSGWPPPFPGPPIIPPPPMSPDFGE-DDEALGSMLISWYMSGYHTGYYMGFRONKK SPPIPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
cons * :** ***::**:************************	tr Q90WY9 Q90WY ESSIGKPPH-QK sp Q16637 SMN H EG

Pablo Mier munoz@uni-mainz.de



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)



Piq

(Sus scrofa)



Chicken (rooster) (*Gallus gallus*)





