

Esercizio: Traduzione della sequenza del trascritto MYOZ 1 (vista in una diapositiva delle lezioni)

>gi|21359948|ref|NM_021245.2| Homo sapiens myozenin 1 (MYOZ1), mRNA
GTTTCTCCCTAAGTGTCTTTGGATCTCAGGCTCTAGGTGCAATGTGAAGGGGAGTCCCTGGGCAGACTGATCCCTGGC
TCAGACAGTTCAGTGGGAGAATCCCAAAGGCCTTTTCCCTCCTTCTGAGCCTCCGGGCAAGGAGGGAGGGATCTTGGTT
CCAGGTCTCAGTACCCCTGTGCCATTTGAGCTGCTTGCCTCATCATCTCTATTAATAACCAACTTCCCTCCCCACT
GCCAGTGTGCCCCACGCCTGCCAGCTCGTGTCTCCGGTCACAGCAGCTCAGTCTCCAAAGCTGCTGGACCCAGG
GAGAGCTGACCCTGCCCGAGCAGCCGGCTGAATCCACCTCCACAATGCCGCTCTCAGGAACCCCGGCCCTAATAAGAA
GAGGAAATCCAGCAAGCTGATCATGGAACCTCACTGGAGGTGGACAGGAGAGCTCAGGCTTGAACCTGGGCAAAAAAGATCA
GTGTCCCAAGGGATGTGATGTTGGAGGAACCTGTCGCTGCTTACCAACCCGGGGCTCCAAGATGTTCAAACTGCGGCAGATG
AGGGTGGAGAAGTTTATTTATGAGAACCACCCTGATGTTTTCTCTGACAGCTCAATGGATCACTTCCAGAAGTTCTTCC
AACAGTGGGGGGACAGCTGGGCACAGCTGGTCAGGGATTCTCATAACAGCAAGAGCAACGGCAGAGGGCGCAGCCAGGCAG
GGGCAGTGGCTCTGCCGGACAGTATGGCTCTGATCAGCAGCACCATCTGGGCTCTGGGTCTGGAGCTGGGGGTACAGGT
GGTCCC CGGGCCAGGCTGGCAGAGGAGGAGCTGCTGGCACAGCAGGGGTTGGTGAGACAGGATCAGGAGACCAGGCAGG
CGGAAGGAAAAACATACACTGTGTTCAAGACCTATATTTCCCATGGGAGCGAGCCATGGGGGTTGACCCCCAGCAAAA
AAATGGAACCTGGCATTGACCTGGCCTATGGGCCAAAGCTGAACCTCCCAAATATAAGTCTTCAACAGGACGGCA
ATGCCCTATGGTGGATATGAGAAGGCCTCCAAACGCATGACCTTCCAGATGCCCAAGTTTGACCTGGGGCCCTTGCTGAG
TGAACCCCTGGTCTCTACAACCAAACCTCTCCAACAGGCCTTCTTTCAATCGAACCCCTATTCCCTGGCTGAGCTCTG
GGGAGCCTGTAGACTACAACGTGGATATTGGCATCCCTTGGATGGAGAAAACAGAGGAGCTGTGAGGTGTTTTCTCCTCT
GATTTGCATCATTTCCCTCTCTGGCTCCAATTTGGAGAGGGAATGCTGAGCAGATAGCCCCATTGTTAATCCAGTATC
CTTATGGGAATGGAGGGAAAAAGGAGAGATCTACCTTCCATCCTTTACTCCAAGTCCCCACTCCACGCATCCTTCTCA
CCAACCTCAGAGCTCCCTTCTACTTGCTCCATATGGAACCTGCTCGTTTATGGAATTTGCTCTGCCACCAGTAACAGTCA
ATAAACTTCAAGGAAAATGAAAAAAA

Traduzione con **Expsy Translate tool**

Strand '+' (5'3') (si ricavano 3 frame)

5'3' Frame1

VSP **Stop** VLLWISGSRNCVKGSPWAD **Stop** SLAQTQVQWENPKGLFPPS **Stop** ASQGGRDLGSRVSVPPVPFELLALIIS
INNQLPSPASAPTQAQLVFSGHSSSVLQSCWTPGRADHCPSSRLNPPPQCRSQEPRPLIRGNPAS **Stop** SWNSLEVD
RRAQA **Stop** TWAKRSVSQG **Met** **Stop** CWRNCRCLPTGAPRCSNCGR **Stop** GWRSLF **Met** RTTL **Met** FSLTAQWITSRSSF
QQWGDSDWAQLVRDSHTARATAEAAAARQGAVALPDS **Met** ALISSTIWALGLELGVQVVPARLAEELLQAQGLVRQDQET
RQAEKENISLCSRPIFPHGSEPWGLTPSKKWNLALTCWP **Met** GKPLNFPNISPSTGRQCP **Met** VD **Met** RRPPNA **Stop** PS
RCPSLTWGPC **Stop** VNPWSSTTKTSPTGLLSIEPLFP **Stop** ALGSL **Stop** TTTWILASPW **Met** EKQRSCEVFPPLICII
SPLWLQFEG **Met** LSR **Stop** PPLLIQYPYNGGKKERSTFPSFTPSPHSTHPSPTQSSPSTCSIWNLVYGICATSNS
Q **Stop** TSRK **Met** KKKKKKKKKK

5'3' Frame2

FLPKCFFGSALGA **Met** **Stop** RGVPGQTDPWLRQFSGRI PKAFSLLPEPPGKEGGILVPGSQYPLCHLSCLRSSLLITN
FPPPLPVLPPRLPSSCSPVTAAQSSKAAGPQGELTTARAAG **Stop** IHLHNAALRNPGP **Stop** **Stop** EEEIQQADHGTHWR
WTGELRLEPGQKDQCPKGCVDVGGTVAAYQPLQDVQTAADGGEVYL **Stop** EPP **Stop** CFL **Stop** QLNGSLPEVPSNSGG
TAGHSWSGILIQEQRRRQPRGRQWL CRTVWL **Stop** SAAPSGLWVWSWGYRWSRGPWQRRSCWHSRGW **Stop** DRIRRP
GRRRRKTYHCVDLYFP **Met** GASHGG **Stop** PPAKNGTWH **Stop** PAGLWQS **Stop** TSQI **Stop** VLQODGNALWWI **Stop** E
GLQTHDLPDAQV **Stop** PGALAE **Stop** TPGLQPKPQQAFFQSNPYS LAELWGACRLQRGYWHPLGWRNRGAVRCFL **St**
op FASFPLSGSNLEREC **Stop** ADSPHC **Stop** SSIL **Met** **Met** EGKRRDLFPFHLLQVPTPRILPHQLRAPLLLPYGTCS
F **Met** EFALPPVTVNKLQK **Stop** KKKKKKKKKK

5'3' Frame3

FSLSASLDLRL **Stop** VQCEGESLGRILPGSDSSVGESQRPFPSFLSLRARREGSWFQGLSTPCA **Stop** AACAHHL **Sto**
p **Stop** PTLPHCQCCPHACPARVLRSQLSPPKLLDPRES **Stop** PLPEQPAESTST **Met** PLSGTPAPNKKRKS **Met**
ELTGGGQESSGLNLGKKSIVPRDV **Met** LEELSLLTNRGSK **Met** FKLRL **Met** RVEKFIYENHPDVFS **Met** DHFQKFLP
TVGGQLGTAGQGFYSKSNRGGSQAGGSGSAGQYSDQHHLGS GSGAGGTGGPAGQAGRGAAGTAGVGETGSGDQAG
GEGKHITVFKTYISPWER **Met** GVDPOQK **Met** ELGIDLLAYAKAELPKYKSFNRTA **Met** PYGGYEKASKR **Met** TFQ **Met**
PKFDLGPLLSEPLVLYNQLNSNRPSFNRTPIPLWSSGEPVDYNDIGIPLDGETEEL **Stop** GVSSSDLHFFPSLAPIWRG
NAEQIAPIVNPVSLWEWREKGEIYLSILYSKSPHASFLTNSLPFYLLH **Met** EPARLWNLCHQ **Stop** QSINFKENEKK
KKKKKKKK

Strand '-' 3'5' (si ricavano 3 frame)

3'5'Frame1

FFFFFFFFFFFFIFLEVY **Stop** LLLVAEQIP **Stop** TSRFH **Met** EQVEGEL **Stop** VGEEGCWEVGLGVKDGKVDLSFFPPFP
Stop GYWINNGGYLLSIPSPNWSQRGE **MetMet** QIRGGNTSOLLCSFIQGDANIHVVVYRLPRAQPGNRGSIERRPVGEV
 LVVEDQGF^TQQGPQVKLGHLEGHAFGGLLISTIGHCRPVEGLIFGKFSFGPIGQOVNAKFHFLLVGNPHGSLPWGNIGLE
 HSD **Met** FSFSACLVS **Stop** SCLTNPCASSSSSASLARGTTCTPSSRPRAQ **Met** VLLIRAILSGRATAPCLAAASAVALA
 V **Stop** ESLTSCAQLSPHCWKELEVIH **Stop** AVRENIRVVLINKLLHPLHPQFEHLGAPVGKQRQFLQHHPWDTDLFAQ
 VQA **Stop** ALLSTSSEFHDQLAGFPLLIRGRGS **Stop** ERHCGGGFSRLLGQWSALPGVQQLWRTELL **Stop** PENTSWAGVG
 AALAVGEGSWLLIE **MetMet** SASSSNGTGGTETLEPRSLPPCPEAQEGGKRPLGFSSH **Stop** TV **Stop** ARDQSAQGLPFTL
 HLEPEIQRST **Stop** GE

3'5'Frame2

FFFFFFFFFFFFSFLKFIDCYWWQSKFHKRAGSIWSK **Stop** KGSELVLRKDAWSGDLE **Stop** **Met** ER **Stop** ISPFSLSHS
 KDTGLT **Met** GAICSAFPLQIGAREGK **Stop** CKSEETPHSSSVSPSKG **Met** PISTL **Stop** STGSPELSOGIGVRLKEGLL
 ERFWL **Stop** RTRGSLSKGPRSNLGIWKV **Met** RLEAFSYPP **Stop** GIAVLLKDLYLGSSALAP **Stop** ASRS **Met** PSSIFCW
 GSTP **Met** ARSHGEI **Stop** VLNTVICFPSPPAWSPDPVSPTPAVPAAPPLPAWPAGPPVPPAPDPEPRWCC **Stop** SEPYCP
 AEPLPPAWLPPLPLLLLYENP **Stop** PAVPSCPPTVGRNFWK **Stop** SIELSEKTSGWFS **Stop** INFSTLICRSLNILEPRL
 VSSDSSSNITSLGTLIFLPRFKPELSCPPPVS **Met** ISLLDFLFLGAGVPESGIVEVDSAGCSGSQLSLGSSSFGGLS
 CCDRRTRAGQAWGQHWQWGREVGY **StopStop** **StopStop** AQAQ **Met** AQGVLRPWNQDPSLLARRLRKEGKGLWDSPT
 ELSEPGISLPRDPSHCT **Stop** SLRSKEALREK

3'5'Frame3

FFFFFFFFFFFFHFP **Stop** SLLTVTGGRANSINEQVPYGASRRGALSW **Stop** GR **Met** RGVGTWSKGWKGRSLLFPSIPIRIL
 D **Stop** QWGLSAQHLSLKLPERGNDANQRRKHLTAPLFLHPRGCQYPRCSLQAPQSSARE **Stop** GFD **Stop** KKACWRGFG
 CRGPGVHSARAPGQTWASGRSCVWRPSHIHHRALPSC **Stop** RTYIWEVQLWPHRPAGQCQVPPFAGGQPPWLAP **Met** GKY
 RS **Stop** TQ **Stop** YVFLRLPGLLILSHQPLLCQQLLLCQPGPRDHLYPQLQTQSPDGAADQSHTVQRSHCPLPGCRLCRC
 SCC **Met** R^IPDQLCPAVPPLLEGTSGSDPLSCQRKHQGGSHK **Stop** TSPSSAAV **Stop** TSWSPGW **Stop** AATVPPTSHPL
 GH **Stop** SFCPGSSLSPPVHLQ **Stop** VP **Stop** SACWISSY **Stop** GPGFLRAALWRWIQPAARAVVSSPWGPAALD **Stop**
 AAVTGEHELGRGGSTGSGGGKLVINRDDRKQLKWHRGY **Stop** DPGTKIPPSLPGSGRREKAFGILPLNCLSQGSVCP
 GTPLHIAPRA **Stop** DPKKHLGR

ORF Finder (Open Reading Frame Finder)

PubMed
Entrez
BLAST
OMIM
Taxonomy
Structure

Anonymous

Frame	from	to	Length
+3	366	1265	900
+1	580	1080	501
-1	864	1292	429
-2	215	424	210
+2	1364	1540	177
-2	35	187	153
-1	90	209	120
-2	1460	1570	111
-3	574	684	111

filamento '+' forward: GTTTCTCCCTAAGTGCTTCTTTGGATCTCAGGCTCT

Frame +1 [GTT][TCT][CCC][TAA] ecc GTGCTTCTTTGGATCTCA

Frame +2 G[TTT][CTC][CCT][AAG] ecc TGCTTCTTTGGATCTCA

Frame +3 GT[TTC][TCC][CTA][AGT] ecc GCTTCTTTGGATCTCA

filamento '-' Reverse: AGAGCCTGAGATCCAAAGAAGCACTTAGGGAGAAAC

Frame -1 [AGA][GCC][TGA][GAT] ecc CCAAAGAAGCACTTAGGGAGAAAC

Frame -2 A[GAG][CCT][GAG][ATC] ecc CAAAGAAGCACTTAGGGAGAAAC

Frame -3 AG[AGC][CTG][AGA][TCC] ecc AAAGAAGCACTTAGGGAGAAAC

Se si conosce a priori la direzionalità della regione codificante, allora si può tradurre la sequenza nucleotidica in tre differenti 'frame'.

Se non si conosce la direzionalità, si può invece tradurre in $3 + 3 = 6$ 'frame' differenti (tre sul filamento '+' e tre sul filamento '-').

Con la traduzione in silico, come si può capire quale è il 'frame giusto', cioè quello che probabilmente codifica la proteina ?

Metodo pratico:

ricordando che:

- la traduzione parte da un codone che codifica per la metionina (ATG o AUG per mRNA)
- una 'open reading frame' (ORF) è la sequenza nucleotidica compresa tra il primo codone (metionina) e lo stop;
- considerando che ci sono 64 differenti codoni (4^3) e che tre rappresentano lo stop alla traduzione, se la sequenza fosse casuale, ci dovremmo aspettare circa uno stop ogni 20 codoni (con 61 codoni codificanti $\rightarrow 61/3 \approx 20$).

Se nella traduzione in silico vediamo un frame con una lunga ORF, possiamo, con una elevata probabilità, affermare che questo rappresenta il frame giusto.

Analizzate i frame relativi a MYOZ1 soprascritti ed individuate il presunto 'giusto frame' di lettura.

Controllate poi se corrisponde effettivamente con l'ORF sottolineata nella sequenza gi|21359948 scritta inizialmente.