

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

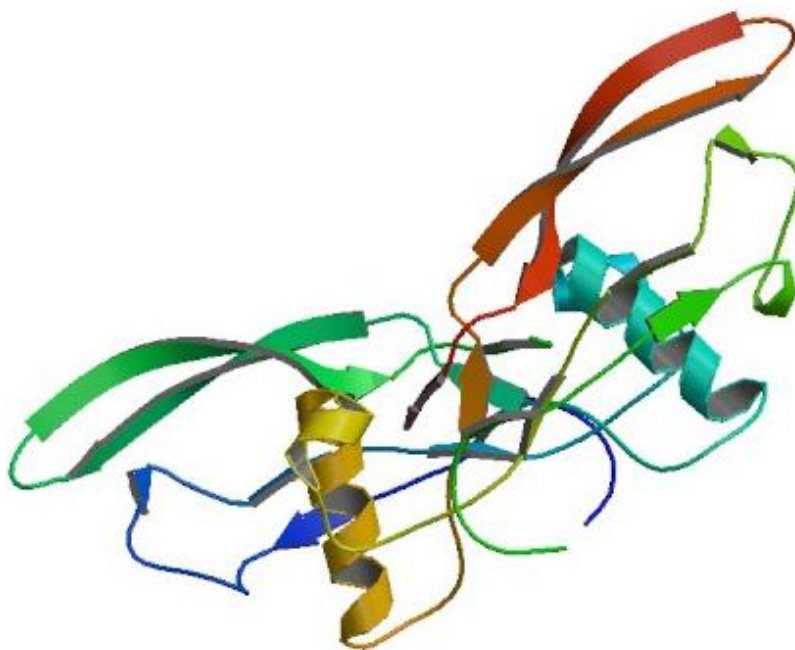
ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΕΡΓΑΣΤΗΡΙΟ ΒΙΟΧΗΜΕΙΑΣ



**Άσκηση στο μάθημα Δομική Βιοχημεία και Στοιχεία
Βιοπληροφορικής**

Αναφορά Αποτελεσμάτων



Υπεύθυνη Καθηγήτρια: Χολή- Παπαδοπούλου Θ.

Υπεύθυνος Εργασίας: Παυλίδης Αλέξανδρος

ΑΕΜ: 9012

Ημερομηνία: 15/05/2019

Τίτλος: Οστική Μορφογενετική Πρωτεΐνη 3 (Bone Morphogenetic Protein 3)

Κωδικός: **P12645**

Στην παρούσα εργασία μελετάται η οστική μορφογενετική πρωτεΐνη 3 (BMP 3/ Bone Morphogenetic Protein 3).

Λειτουργία: Οι Οστικές Μορφογενετικές Πρωτεΐνες (BMPs) είναι αυξητικοί παράγοντες, που ανήκουν στην υπερ-οικογένεια των αυξητικών παραγόντων TGF beta. Διακατέχουν σημαντικό ρόλο σε μία πληθώρα λειτουργιών στον οργανισμό, κάποιες από τις οποίες είναι τόσο ο σχηματισμός, η διατήρηση και η επιδιόρθωση οστών όσο και η εμβρυϊκή ανάπτυξη. Δρουν μέσω παρακρινούς ή αυτοκρινούς συστήματος, και μέσω πρόσδεσης στην κυτταρική επιφάνεια υποδοχέων κινασών Ser/ Thr πυροδοτούν ένα καταρράκτη αντιδράσεων-σημάτων μέσω άλλων σημάτων Smad (πρωτεΐνες μεταφορείς μηνυμάτων σημαντικές για τη ρύθμιση της κυτταρικής ανάπτυξης).

Ειδικότερα, η BMP3 προκαλεί σχηματισμό χόνδρων και οστών. Λειτουργεί επίσης στην επαγωγή του μεσοδερμικού, την ανάπτυξη των δοντιών, το σχηματισμό των άκρων και την αποκατάσταση των θραυσμάτων. ρυθμίζει αρνητικά την οστική πυκνότητα. Ανταγωνίζεται την ικανότητα ορισμένων οστεογονικών BMP να προάγουν την οστεοπρογονική διαφοροποίηση (δηλαδή διαφοροποίηση των μυοκυττάρων σε οστεοκύτταρα) και οστεογένεση.

Αρχικά μπαίνουμε στη διεύθυνση της βάσης δεδομένων πρωτεϊνικών ακολουθιών SPROT: <http://www.expasy.ch/sprot/>. Στη συνέχεια διαλέγουμε την τράπεζα δεδομένων επιλέγοντας το link List of UniProtKB/Swiss-Prot (reviewed) entries. Στο πεδίο πάνω δεξιά πληκτρολογούμε το **BMP 3** και ακολούθως επιλέγουμε την πρώτη πρωτεΐνη με τον κωδικό καταχώρησης **P12645 (Homo Sapiens/ Human)**. Μπαίνουμε στη σελίδα της πρωτεΐνης και αντλούμε από εκεί τις παρακάτω πληροφορίες:

Η αμινοξική ακολουθία της πρωτεΐνης που προκύπτει από το FASTA format είναι:

```
>sp|P12645|BMP3_HUMAN Bone morphogenetic protein 3 OS=Homo sapiens  
OX=9606 GN=BMP3 PE=1 SV=2  
MAGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGGGPDSELQP  
QDKVSEHM
```

LRLYDRYSTVQAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAAETLERKG
LYIFNLTS
LTKSENILSATLYFCIGELGNISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRN
QSQLL
GHLSVDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRL
PFPEPYI
LVYANDAAISEPESVSSSLQGHRNFPTGTVPKWDSHIRAALSIERRKKRSTGV
LLPLQNN
ELPGA EYQYKKDEVWEERKPYKTLQAQAPEKSKNKKKQRKGPHRKSQTLQF
DEQTLKKAR
RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFPMPKSLKP
SNHATIQ
SIVRAVGVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESACR

Οι βάσεις δεδομένων που μπορούν να επιλεγούν για τη μελέτη είναι:

- EMBL
- GenBank
- DDBJ

Στη συγκεκριμένη εργασία αναζητήθηκαν οι πληροφορίες της πρωτεΐνης στη βάση δεδομένων EMBL, σύμφωνα με την οποία προκύπτουν τα παρακάτω δεδομένα:

EMBL (Accession Number BC096269, Sequence Version BC096269.2, Keywords MGC.):

>ENA|BC096269|BC096269.2 Homo sapiens bone morphogenetic protein 3, mRNA (cDNA clone MGC:

116934 IMAGE:40006302), complete cds.

TCAGCGTTGGAGTGGAGACGGCGCCCGCAGCGCCCTGCGCGGGTGAGGTCCGCG
CAGCTG

CTGGGGAAGAGCCACCTGTCAGGCTGCGCTGGGTCAGCGCAGCAAGTGGGGCT
GGCCGC

3

TATCTCGCTGCACCCGGCCGCGTCCCGGGCTCCGTGCGCCCTCGCCCCAGCTGGT
TTGGA

GTTCAACCCTCGGCTCCGCCGCCGGCTCCTTGCGCCTTCGGAGTGTCCCGCAGCG
ACGCC

GGGAGCCGACGCGCCGCGCGGGTACCTAGCCATGGCTGGGGCGAGCAGGCTGCT
CTTTCT

GTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGCAGGGAGAGAGACCGAAGCCACC
 TTTCCC
 GGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTGGTGGCCCCGGA
 CTGAGCT
 GCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAG
 CACGGT
 CCAGGCGGCCCCGGACACCGGGCTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCC
 TCGGCT
 CCTGCGCGAAGGCAACACGGTTTCGCAGCTTTTCGGGCGGCAGCAGCAGAACTCT
 TGAAAG
 AAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTTG
 TCTGC
 CACACTGTATTTCTGTATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTG
 TCTGG
 AGGATGCTCCCATCATGCTCAGAGGAAACACATTCAGATTGATCTTTCTGCATGG
 ACCCT
 CAAATTCAGCAGAAACCAAAGTCAACTCCTTGGCCATCTGTCAGTGGATATGGCC
 AAATC
 TCATCGAGATATTATGTCCTGGCTGTCTAAAGATATCACTCAATTCTTGAGGAAG
 GCCAA
 AGAAAATGAAGAGTTCCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCT
 GCCAAA
 GAGGAGGTTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCC
 ATTTT
 TGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAATTTTCCCACTGGA
 ACTGT
 TCCCAAATGGGATAGCCACATCAGAGCTGCCCTTTCCATTGAGCGGAGGAAGAA
 GCGCTC
 TACTGGGGTCTTGCTGCCTCTGCAGAACAAACGAGCTTCCTGGGGCAGAATACCAG
 TATAA
 AAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCAGGCTCAGGC
 CCCTGA
 AAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGA
 CGCTCCA
 ATTTGATGAGCAGACCCTGAAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCG
 GAATTG
 CGCCAGGAGATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATT
 ATCTC
 CCCCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCCCATGCCAA
 AGTC
 TTTGAAGCCATCAAATCATGCTACCATCCAGAGTATAGTGAGAGCTGTGGGGGTC
 GTTCC
 TGGGATTCTTGAGCCTTGCTGTGTACCAGAAAAGATGTCCTCACTCAGTATTTTAT
 TCTT
 TGATGAAAATAAGAATGTAGTGCTTAAAGTATACCCTAACATGACAGTAGAGTCT
 TGCGC
 TTGCAGATAACCTGGCAAAGAAGTCAATTTGAATGC

**EMBL– mRNA Translation (Accession Number AAH96269, Sequence Version
 AAH96269.1, Keywords MGC.):**

>ENA|AAH96269|AAH96269.1 Homo sapiens (human) bone morphogenetic protein

ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTG
AGCCTGGCG

CAGGGAGAGAGACCGAAGCCACCTTTCCCGGAGCTCCGCAAAGCTGTGCC
AGGTGACCGC

ACGGCAGGTGGTGGCCCGGACTCCGAGCTGCAGCCGCAAGACAAGGTCTC
TGAACACATG

CTGCGGCTCTATGACAGGTACAGCACGGTCCAGGCGGCCCGGACACCGGG
CTCCCTGGAG

GGAGGCTCGCAGCCCTGGCGCCCTCGGCTCCTGCGCGAAGGCAACACGGT
TCGCAGCTTT

CGGGCGGCAGCAGCAGAACTCTTGAAAGAAAAGGACTGTATATCTTCAA
TCTGACATCG

CTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGA
GAGCTAGGA

AACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCA
GAGGAAACAC

ATTCAGATTGATCTTTCTGCATGGACCCTCAAATTCAGCAGAAACCAAAG
TCAACTCCTT

GGCCATCTGTGAGTGGATATGGCCAAATCTCATCGAGATATTATGTCCTGG
CTGTCTAAA

GATATCACTCAATTCTTGAGGAAGGCCAAAGAAAATGAAGAGTTCCTCAT
AGGATTTAAC

4

ATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGGTACCTTTTCCAGA
GCCTTATATC

TTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATC
AAGCTTACAG

GGACACCGGAATTTTCCCACTGGAAGTGTCCCAAATGGGATAGCCACAT
CAGAGCTGCC

CTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGCTGCCTCTG
CAGAACAAC

GAGCTTCCTGGGGCAGAATACCAGTATAAAAAGGATGAGGTGTGGGAGG
AGAGAAAGCCT

TACAAGACCCTTCAGGCTCAGGCCCTGAAAAGAGTAAGAATAAAAAGA
AACAGAGAAAG

GGGCCTCATCGGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGAA
AAAGGCAAGG

AGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGATACCTCAAGGT
AGACTTTGCA

GATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTTTGATGCCTAT
TATTGCTCT

GGAGCATGCCAGTTCCCCATGCCAAAGTCTTTGAAGCCATCAAATCATGC
TACCATCCAG

AGTATAGTGAGAGCTGTGGGGGTCGTTCTGGGATTCCTGAGCCTTGCTGT
GTACCAGAA

AAGATGTCCTCACTCAGTATTTTATTCTTTGATGAAAATAAGAATGTAGTG
CTTAAAGTA

TACCCTAACATGACAGTAGAGTCTTGCGCTTGCAGATAA

GenBank (Accession Number BC096269, Sequence Version BC096269.2,
Keywords MGC.):

ORIGIN

1 tcagcgttgg agtggagacg gcgcccgcag cgccctgcgc gggtaggtc cgcgcagctg
61 ctggggaaga gccacactgt caggctgcgc tgggtcagcg cagcaagtg ggctggccgc
121 tatctcgtg caccggccg cgtcccggc tccgtgcgc ctcgcccag ctggtttgga
181 gttcaacct cggtccgcc gccggctct tgcgcctcg gagtgtccg cagcgacgcc
241 gggagccgac gcgccgcgc ggtacctagc catggctggg gcgagcagcg tgctcttct
301 gtggctgggc tgcttctcg tgagcctggc gcagggagag agaccgaagc caccttccc
361 ggagctccgc aaagctgtgc caggtgaccg cacggcaggt ggtggcccg actccgagct
421 gcagccgcaa gacaaggtct ctgaacacat gctgcggctc tatgacaggt acagcacggt
481 ccaggcggcc cggacaccgg gtccttga gggaggtcg cagccctggc gccctcggt
541 cctgcgcgaa ggcaacacgg ttcgcagctt tcggcgga gcagcagaaa ctcttgaag
601 aaaaggactg tatatctca atctgacatc gctaaccaag tctgaaaaca tttgtctgc
661 cacactgtat ttctgtattg gagagctagg aaacatcagc ctgagttgtc cagtgtctgg
721 aggatgctcc catcatgctc agaggaaaca cattcagatt gatcttctg catggaccct

781 caaattcagc agaaaccaa gtcaactcct tggccatctg tcagtggata tggccaaatc
 841 tcatcgagat attatgtcct ggctgtctaa agatatact caattctga ggaaggccaa
 901 agaaaatgaa gagttcctca taggatttaa cattacgtcc aaggagcggc agctgccaaa
 961 gaggagggtta cctttccag agccttatat cttggtatat gccaatgatg ccgccatttc
 1021 tgagccagaa agtgtggtat caagcttaca gggacaccgg aattttcca ctggaactgt
 1081 tcccaaatgg gatagccaca tcagagctgc ctttccatt gagcggagga agaagcgctc
 1141 tactggggtc ttgctgcctc tgcagaacaa cgagcttctt ggggcagaat accagtataa
 1201 aaaggatgag gtgtgggagg agagaaagcc ttacaagacc ctcaggctc aggccctga
 1261 aaagagtaag aataaaaaga aacagagaaa ggggcctcat cggaagagcc agacgtcca
 1321 atttgatgag cagaccctga aaaaggcaag gagaaagcag tggattgaac ctcggaattg
 1381 cgccaggaga tacctcaagg tagactttgc agatattggc tggagtgaat ggattatctc
 1441 cccaagtcc ttgatgcct attattgctc tggagcatgc cagttccca tgccaaagtc
 1501 ttgaagcca tcaaatcatg ctaccatcca gagtatagtg agagctgtgg gggtcgttcc
 5
 1561 tgggattcct gagccttgc gtgtaccaga aaagatgtcc tcaactagta tttattctt
 1621 tgatgaaat aagaatgtag tgctaaagt atacctaac atgacagtag agtcttgcgc
 1681 ttgcagataa cctggcaaag aactcattg aatgc //

GenBank – mRNA Translation (Accession Number AAH96269, Sequence Version AAH96269.1, Keywords MGC.):

ORIGIN

1 magasrllfl wlgcfcvsla qgerpkppfp elrkavpgdr tagggpdsel qpqdkvsehm
 61 lrydrystv qaartpgsle ggsqpwrprl lregntvrsf raaaetler kglyifnlts
 121 ltksenilsa tlyfcigelg nislscpvsg gcshhaqrkh iqidlsawtl kfsrnqsll
 181 ghlsvdmaks hrdimswlsk ditqflrkak eneeffligfn itskgrqlpk rrlpfpepyi
 241 lvyandaais epesvssllq ghrnfptgtv pkwdshiraa lsierkkrs tgvllplqnn
 301 elpgaeyqyk kdevweerkp yktlqaqape ksknkkkqrk gphrksqtlq fdeqtlkkr
 361 rkqwieprnc arrylkvdfa digwsewiis pksfdaiyys gacqfmpks lkpsnhatiq
 421 sivravgvvp gipepccvpe kmsslsilff denknvvlkv ypnmtvesca cr
 //

DDBJ (Accession Number BC096269, Sequence Version BC096269.2, Keywords MGC.):

BASE COUNT 417 a 452 c 478 g 368 t ORIGIN 1 tcagcgttg
agtggagacg gcgcccgcag cgcctgcgc gggtagagtc cgcgcagctg 61 ctggggaaga gccacactgt caggetgcgc
tgggtcagcg cagcaagtgg ggctggccgc 121 tatctcgtg caccggccg cgtccgggc tccgtgcgc ctcgcccag
ctggtttgga 181 gttcaacct cggctccgc gccggctct tgcgccttc gagtgccg cagcgacgc 241
gggagccgac gcgcccgcg ggtacctagc catggctggg gcgagcaggc tgctttct 301 gtggctggc tcttctgc
tgagcctggc gcagggagag agaccgaagc caccttccc 361 ggagctcgc aaagctgtgc caggtgaccg cacggcaggt
ggtggcccg actccgagct 421 gcagccgcaa gacaaggtct ctgaacacat gtcgcgctc tatgacaggt acagcacggt
481 ccaggcgcc cggacaccgg gctccctgga gggaggctc cagccctggc gcctcggct 541 cctgcgcgaa
ggcaacacgg ttcgcagct tcggcgga gcagcagaaa ctctgaaag 601 aaaaggactg tatacttca atctgacatc
gtaaccaag tctgaaaaca tttgtctgc 661 cacactgtat ttctgtatt gagagctagg aaacatcagc ctgagttgc
cagtgtctgg 721 aggatgctc catcatgctc agaggaaaca cattagatt gatcttctc catggaccct 781 caaattcagc
agaaaccaa gtaactcct tggccatctg tcagtggata tggccaaatc 841 tcatcgatg attatgctt ggctgtctaa
agatacact caattctga ggaaggccaa 901 agaaaatgaa gagttctca taggatttaa cattacgtcc aaggagcgc
agctgcaaa 961 gaggaggta ctttccag agcctatat ctggtatat gccaatgatg ccgccatttc 1021 tgagccagaa
agtgtggtat caagcttaca gggacaccgg aatttcca ctggaactgt 1081 tccaaatgg gatagccaca tcagagctgc
ctttccatt gagcggagga agaagcgtc 1141 tactgggctc ttgctgctc tgcagaacaa cgagcttct ggggcagaat
accagtataa 1201 aaaggatgag gtgtgggagg agagaaagcc ttacaagacc ctcaggctc agccctga 1261
aaagagtaag aataaaaaga aacagagaaa ggggcctcat cggaagagcc agacgtcca 1321 attgatgag cagaccctga
aaaaggcaag gagaagcag tggattgaac ctggaattg 1381 cgccaggaga tacctcaagg tagacttgc agatattggc
tggagtgaat ggattatctc 1441 cccaagtc ttgatgctc attattgctc tggagcatgc cagttccca tgcctaaatc 1501
ttgaagcca tcaatcatg ctaccatca gatatagtg agagctgtg ggtcgttc 1561 tgggattcct gagcctgtc
gtgtaccaga aaagatgctc tcaatcagta tttattct 1621 tgatgaaat aagaatgtag tgctaaagt atacctaac
atgacagtag agtctgcgc 1681 ttgcagataa cctggcaaag aactcattg aatgc //

6

DDBJ – mRNA Translation (Accession Number AAH96269, Sequence Version AAH96269.1, Keywords MGC.):

BEGIN

1 MAGASRLLFL WLGCFCVSLA QGERPKPPFP ELRKAVPGDR TAGGGPDSEL
QPQDKVSEHM

61 LRLYDRYSTV QAARTPGSLE GGSQPWRPRL LREGNTVRSF RAAAAETLER
KGLYIFNLTS

121 LTKSENILSA TLYFCIGELG NISLSCPVSF GCSHHAQRKH IQIDLSAWTL
KFSRNQSLL

181 GHLSVDMAS HRDIMSWLSK DITQFLRKAK ENEEFLIGFN ITSKGRQLPK
RRLPFPEPYI

241 LUYANDAAIS EPESVSSLQ GHRNFPTGTV PKWDSHIRAA LSIERRKRS
TGVLLPLQNN

301 ELPGAQYQYK KDEVWEERKP YKTLQAQAPK KSKNKKKQK GPHRKSQTLQ
FDEQTLKKAR

361 RKQWIEPRNC ARRYLKVDFA DIGWSEWIS PKSFDAYYCS GACQFPMPKS
LKPSNHATIQ

421 SIVRAVGVP GIPEPCCVPE KMSSLSILFF DENKNVVLKV YPNMTVESCA CR

//

The Consensus CDS (CCDS) project (Sequence IDs included in CCDS 3588.1)

Blue highlighting indicates alternating exons. Red highlighting indicates amino acids encoded across a splice junction.

Nucleotide Sequence (1419 nt):

ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGC
AGGGAGAGA
GACCGAAGCCACCTTTCCCGGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTG
GTGGCCCGGA
CTCCGAGCTGCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTA
CAGCACGGTC
CAGGCGGCCCGGACACCGGGCTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCCTCGGCTC
CTGCGCGAAG
GCAACACGGTTTCGCAGCTTTCTGGGCGGCAGCAGCAGAACTCTTGAAAGAAAAGGACTGT
ATATCTTCAA
TCTGACATCGCTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGA
GAGCTAGGA
AACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACAC
ATTCAGATTG
ATCTTTCTGCATGGACCCTCAAATTCAGCAGAAACCAAAGTCAACTCCTTGCCATCTGTC
AGTGGATAT
GGCCAAATCTCATCGAGATATTATGTCCTGGCTGTCTAAAGATATCACTCAACTCTTGAGG
AAGGCCAAA
GAAAATGAAGAGTTCCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAAG
AGGAGGTTAC
CTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAG
TGTGGTATC
AAGCTTACAGGGACACCGGAATTTTCCCACTGGAAGTGTCCCAAATGGGATAGCCACAT
CAGAGCTGCC
CTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGCTGCCTCTGCAGAACAAAC
GAGCTTCCTG
GGGCAGAATACCAGTATAAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACC
CTTCAGGCTCA
GGCCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCC
AGACGCTCCAA
TTTGATGAGCAGACCCTGAAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTG
CGCCAGGAGAT
ACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTT
TGATGCCTA
TTATTGCTCTGGAGCATGCCAGTTCCCCATGCCAAAGTCTTTGAAGCCATCAAATCATGCT
ACCATCCAG
AGTATAGTGAGAGCTGTGGGGGTCGTTCTTGGGATTCTTGAGCCTTGCTGTGTACCAGAA
AAGATGTCCT
CACTCAGTATTTTATTCTTTGATGAAAATAAGAATGTAGTGCTTAAAGTATACCCTAACAT
GACAGTAGA GTCTTGCGCTTGACAGATAA

Translation (472 aa):

MAGASRLFLWLGCFCVSLAQGERPKPPFELRKAVPGDRTAGGGPDSELQPQDKVSEHMLR
LYDRYSTV

QAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAAETLERKGLYIFNLSTLTKSENILSATLYF
CIGELG
NISLSCPVS GGCSHHAQRKHIQIDLSAWTLKFSRNQS QLLGHLSVDMAKSHR DIMSWLSKDIT
QLLRKAK
ENEEFLIGFNITSKGRQLPKRRLPFPEPYILVYANDAAISEPESV VSSLQGHRNFPTGTVPKWDS
HIRAA
LSIERRKKRSTGVLLPLQNNELPGA EYQYKKDEVWEERKPYKTLQAQAPEKSKNKKKQRKGP
HRKSQTLQ
FDEQTLKKARRKQWIEPRNCARRY LKVD FADIGWSEWISPKSFDAYYCSGACQFPMPKSLKP
SNHATIQ SIVRAVG VVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESACR

Στη συνέχεια μεταβαίνουμε στη σελίδα <https://www.expasy.org/proteomics> και επιλέγουμε το εργαλείο **Translate** • nucleotide sequence translation • **[more]**. Έπειτα, εισαγάγουμε μια ακολουθία DNA 7 στο πλαίσιο που εμφανίζεται (χωρίς αριθμούς και κενά) και πατάμε το κουμπί TRANSLATE SEQUENCE. Τα αποτελέσματα περιέχουν τρία πλαίσια για την κατεύθυνση 5'3' (μετάφραση του νοηματικού κλώνου) και τρία πλαίσια για την κατεύθυνση 3'5' (μετάφραση του αντινοηματικού κλώνου) με τα Ανοιχτά Πλαίσια Ανάγνωσης (Open Reading Frames) επισημασμένα με «κόκκινο». Επιλέγουμε το πλαίσιο (frame) που περιέχει τα λιγότερα Stop-κωδικόνια:

5'3' Frame 1

Met AGASRLLFLWLGCFCVSLA QGERPKPPFPELRKAVP
GDRTAGGGPDSELQPQDKVSEH Met LRLYDRYSTVQAA
RTPGSLEGGSQPWRPRLREGNTVRSFRAAAAETLERK
GLYIFNLSTLTKSENILSATLYFCIGELGNISLSCPVS GG
CSHHAQRKHIQIDLSAWTLKFSRNQS QLLGHLSVD Met A
KSHRDI Met SWLSKDITQLLRKAKENEEFLIGFNITSKGR
QLPKRRLPFPEPYILVYANDAAISEPESV VSSLQGHRNF
PTGTVPKWDSHIRA ALSIERRKKRSTGVLLPLQNNELP
GA EYQYKKDEVWEERKPYKTLQAQAPEKSKNKKKQRK
GPHRKSQTLQFDEQTLKKARRKQWIEPRNCARRY LKV
DFADIGWSEWISPKSFDAYYCSGACQFP Met PKSLKPSN
HATIQSIVRAVG VVPGIPEPCCVPEK Met SSLSILFFDENK
NVVLKVYPN Met TVESCACR Stop

3'5' Frame 1

LSASARLYCHVRVYFKHYILIFIKE Stop NTE Stop GH LFW
YTARLRNPRNDPHSSHYTLDGS Met I Stop WLQRLWHGEL
ACSR AIIGIKGLGGDN PFTPANICKVYLEVSPGAIPRFNP
LLSPCLFQGLLIKLERLALP Met RPLSLFLFILTLFRGLSL
KGLVRLSLLPHLILFILVFCPRKLVVLQRQQDPSRALLP

PLNGKGSSDVAIPFGNSSSSGKIPVSL **Stop** A **Stop** YHTFWLR
NGGIIGIYQDIRLWKR **Stop** PPLWQLASLGRNVKSYEELF
IFFGLPQELSDIFRQPGHNISM **Met** RFGHIH **Stop** Q **Met** AKELT
LVSAEFEGPCRKINLNVFPLS **Met** **Met** GASSRHWTTQADV
S **Stop** LSNTETIQCGRQNVFRLG **Stop** RCQIEDIQSFSFKSFCC
CRPKAANRVAFAPGLRASLQGARCPRGLDRAVP
VIEPQHVFRLDLVRLQLGVRATTCTRAVTWHSFAELRER
WLRSLSLRQAHAEEAQPQKEQPARPSH

Ακολούθως, επιστρέφουμε στη διεύθυνση της βάσης δεδομένων CCDS και αυτήν τη φορά επιλέγουμε από τη νουκλεοτιδική ακολουθία τη μεταφράσιμη αλληλουχία (νουκλεοτίδια επισημασμένα με «μπλε») μαζί με 7 νουκλεοτίδια πριν το μεταφράσιμο τμήμα (5' GCAGCAG 3'). Επαναλαμβάνουμε τη διαδικασία μετάφρασης με τη βοήθεια του εργαλείου Translate • nucleotide sequence translation • [more] από την κατηγορία Proteomics, τα αποτελέσματα παρουσιάζονται παρακάτω:

Nucleotide sequence:

GCAGCAGAACTCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATC
GCTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTG

TATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGGAGGAT
GCTCCCATCATGCTCAGAGGAAACACATTCAGATTGATCTTTCTG

CATGGACCCTCAAATTCAGCAGAAACCAAAGTCAACTCCTTGGCCATCTG
TCAGTGGATATGGCCAAATCTCATCGAGATATTATGTCCTGGCTG

TCTAAAGATATCACTCAACTCTTGAGGAAGGCCAAAGAAAATGAAGAGTT
CCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAA

GAGGAGGTTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGC
CGCCATTTCTGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGAC

ACCGGAATTTTCCCACTGGAAGTGTCCCAAATGGGATAGCCACATCAGA
GCTGCCCTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTC

8

TTGCTGCCTCTGCAGAACACGAGCTTCCTGGGGCAGAATACCAGTATAA
AAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCA

GGCTCAGGCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGG
CCTCATCGGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGA

AAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAG
ATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATT

ATCTCCCCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTC
CCCATGCCAAAG

Translation: 5'3' Frame 1 A A E T L E R K G L Y I F N L T S L T K S E N I L S A T
L Y F C I G E L G N I S L S C P V S G G C S H H A Q R K H I Q I D L S A W T L
K F S R N Q S Q L L G H L S V D Met A K S H R D I Met S W L S K D I T Q L L
R K A K E N E E F L I G F N I T S K G R Q L P K R R L P F P E P Y I L V Y A N
D A A I S E P E S V V S S L Q G H R N F P T G T V P K W D S H I R A A L S I E
R R K K R S T G V L L P L Q N N E L P G A E Y Q Y K K D E V W E E R K P Y
K T L Q A Q A P E K S K N K K K Q R K G P H R K S Q T L Q F D E Q T L K K
A R R K Q W I E P R N C A R R Y L K V D F A D I G W S E W I I S P K S F D A
Y Y C S G A C Q F P Met P K

3'5' Frame 1 L W H G E L A C S R A I I G I K G L G G D N P F T P A N I C K V Y
L E V S P G A I P R F N P L L S P C L F Q G L L I K L E R L A L P Met R P L S L
F L F I L T L F R G L S L K G L V R L S L L P H L I L F I L V F C P R K L V V L
Q R Q Q D P S R A L L P P L N G K G S S D V A I P F G N S S S G K I P V S L
Stop A Stop Y H T F W L R N G G I I G I Y Q D I R L W K R Stop P P L W Q L A
S L G R N V K S Y E E L F I F F G L P Q E L S D I F R Q P G H N I S Met R F G
H I H Stop Q Met A K E L T L V S A E F E G P C R K I N L N V F P L S Met Met
G A S S R H W T T Q A D V S Stop L S N T E I Q C G R Q N V F R L G Stop R C
Q I E D I Q S F S F K S F C C Τέλος, επιστρέφουμε στη διεύθυνση της βάσης
δεδομένων CCDS και αυτήν τη φορά επιλέγουμε από τη νουκλεοτιδική ακολουθία
μόνο τη μεταφράσιμη αλληλουχία. Επαναλαμβάνουμε τη διαδικασία μετάφρασης με
τη βοήθεια του εργαλείου Translate • nucleotide sequence translation • [more] από
την κατηγορία Proteomics, τα αποτελέσματα παρουσιάζονται παρακάτω:

Nucleotide sequence:

AAACTCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATCGCTAACC
AAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGA

GAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCA
TCATGCTCAGAGGAAACACATTCAGATTGATCTTTCTGCATGGAC

CCTCAAATTCAGCAGAAACCAAAGTCAACTCCTTGGCCATCTGTCAAGTGG
ATATGGCCAAATCTCATCGAGATATTATGTCCTGGCTGTCTAAAG

ATATCACTCAACTCTTGAGGAAGGCCAAAGAAAATGAAGAGTTCCTCATA
GGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGG

TTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATT
TCTGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAA

TTTTCCCACTGGAAGTGTTCCTCCAAATGGGATAGCCACATCAGAGCTGCCCT
TTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGCTGC

CTCTGCAGAACAAACGAGCTTCCTGGGGCAGAATACCAGTATAAAAAGGAT
GAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCAGGCTCAG

GCCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATC
GGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGAAAAAGGC

AAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGATACCTC
AAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCC

CCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCCCATGC
CAAAG

Translation:

5'3' Frame 3 T L E R K G L Y I F N L T S L T K S E N I L S A T L Y F C I G E L G
N I S L S C P V S G G C S H H A Q R K H I Q I D L S A W T L K F S R N Q S Q L
L G H L S V D Met A K S H R D I Met S W L S K D I T Q L L R K A K E N E E F
L I G F N I T S K G R Q L P K R R L P F P E P Y I L V Y A N D

9

A A I S E P E S V V S S L Q G H R N F P T G T V P K W D S H I R A A L S I E R
R K K R S T G V L L P L Q N N E L P G A E Y Q Y K K D E V W E E R K P Y K
T L Q A Q A P E K S K N K K K Q R K G P H R K S Q T L Q F D E Q T L K K A
R R K Q W I E P R N C A R R Y L K V D F A D I G W S E W I I S P K S F D A Y
Y C S G A C Q F P Met P K

3'5' Frame 1 L W H G E L A C S R A I I G I K G L G G D N P F T P A N I C K V Y
L E V S P G A I P R F N P L L S P C L F Q G L L I K L E R L A L P Met R P L S L
F L F I L T L F R G L S L K G L V R L S L L P H L I L F I L V F C P R K L V V L
Q R Q Q D P S R A L L P P L N G K G S S D V A I P F G N S S S G K I P V S L
Stop A Stop Y H T F W L R N G G I I G I Y Q D I R L W K R Stop P P L W Q L A
S L G R N V K S Y E E L F I F F G L P Q E L S D I F R Q P G H N I S Met R F G
H I H Stop Q Met A K E L T L V S A E F E G P C R K I N L N V F P L S Met Met
G A S S R H W T T Q A D V S Stop L S N T E I Q C G R Q N V F R L G Stop R C
Q I E D I Q S F S F K S

Έπειτα, κάνουμε την πολλαπλή ομοιοπαράθεση (multiple sequence alignment) με τον εξής τρόπο: ξαναμπαίνουμε στη διεύθυνση <https://www.expasy.org/>, επιλέγουμε την κατηγορία **Proteomics** και το εργαλείο **Multalin • Multiple sequence alignment • [more]**. Στο παράθυρο που εμφανίζεται μεταφέρουμε διαφορετικές μεταφρασμένες ακολουθίες νουκλεοτιδίων για τη σύγκριση ομοιότητάς τους. Πρώτα εισαγάγουμε ολόκληρη νουκλεοτιδική ακολουθία της BMP 3 σε μορφή FASTA, μετά – ακολουθία που αποτελείται από τη μεταφράσιμη αλληλουχία μαζί με 8 μεταφρασμένα νουκλεοτίδια πριν το μεταφράσιμο τμήμα (sequence1). Τέλος, εισαγάγουμε μόνο τη μεταφράσιμη ακολουθία (sequence 2) και πατάμε το SUBMIT.

Παρακάτω παρουσιάζονται τα αποτελέσματα που παίρνουμε. Τα αμινοξέα που εμφανίζονται «κόκκινα» στην ομοιοπαράθεση είναι ταυτόσημα, τα «πράσινα» αμινοξέα μπορεί να είναι ταυτόσημα ή παρόμοιου φορτίου αλλά δεν βρίσκονται σε όλες τις ακολουθίες που αναλύθηκαν ενώ τα «μαύρα» δεν παρουσιάζουν καμία ομοιότητα μεταξύ τους.

	10	20	30	40	50	60
S1xxxx0	MIPGNRMLMVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQ\$HELLRDFEAT					
S2xxxx1	MIPGNRMLMVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQ\$HELLRDFEAT					
Consensus	MIPGNRMLMVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQ\$HELLRDFEAT					
Prim.cons.	MIPGNRMLMVLLCQVLLGGASHASLIPETGKKKVAEIQGHAGGRRSGQ\$HELLRDFEAT					
	70	80	90	100	110	120
S1xxxx0	LLQMFGRLRRRPQPSKSAVIPDYMRLDLYRLQSGEEEEQIHSTGLEYPERPASRANTVRSF					
S2xxxx1	LLQMFGRLRRRPQPSKSAVIPDYMRLDLYRLQSGEEEEQIHSTGLEYPERPASRANTVRSF					
Consensus	LLQMFGRLRRRPQPSKSAVIPDYMRLDLYRLQSGEEEEQIHSTGLEYPERPASRANTVRSF					
Prim.cons.	LLQMFGRLRRRPQPSKSAVIPDYMRLDLYRLQSGEEEEQIHSTGLEYPERPASRANTVRSF					
	130	140	150	160	170	180
S1xxxx0	HHEEHLENIPGTSENSAFRFLNLSSIPENEVISSAELRLFREQVDQGPDWERG\$FHRINI					
S2xxxx1	HHEEHLENIPGTSENSAFRFLNLSSIPENEVISSAELRLFREQVDQGPDWERG\$FHRINI					
Consensus	HHEEHLENIPGTSENSAFRFLNLSSIPENEVISSAELRLFREQVDQGPDWERG\$FHRINI					
Prim.cons.	HHEEHLENIPGTSENSAFRFLNLSSIPENEVISSAELRLFREQVDQGPDWERG\$FHRINI					
	190	200	210	220	230	240
S1xxxx0	YEVMKPPAEVVP\$GHLITRLLDTRLVHHNVTRWETFDVSPAVLRW\$TREKQPNYGLAIEVTH					
S2xxxx1	YEVMKPPAEVVP\$GHLITRLLDTRLVHHNVTRWETFDVSPAVLRW\$TREKQPNYGLAIEVTH					
Consensus	YEVMKPPAEVVP\$GHLITRLLDTRLVHHNVTRWETFDVSPAVLRW\$TREKQPNYGLAIEVTH					
Prim.cons.	YEVMKPPAEVVP\$GHLITRLLDTRLVHHNVTRWETFDVSPAVLRW\$TREKQPNYGLAIEVTH					
	250	260	270	280	290	300
S1xxxx0	LHQTRTHQGQ\$HVRISRSLPQSGSNWAQLRPLLVTFGHDGRGH\$ALTRRRRAKRSPKH\$HSQR					
S2xxxx1	LHQTRTHQGQ\$HVRISRSLPQSGSNWAQLRPLLVTFGHDGRGH\$ALTRRRRAKRSPKH\$HSQR					
Consensus	LHQTRTHQGQ\$HVRISRSLPQSGSNWAQLRPLLVTFGHDGRGH\$ALTRRRRAKRSPKH\$HSQR					
Prim.cons.	LHQTRTHQGQ\$HVRISRSLPQSGSNWAQLRPLLVTFGHDGRGH\$ALTRRRRAKRSPKH\$HSQR					

	310	320	330	340	350	360
S1xxxx0	ARKKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLNSTNHAIVQTL					
S2xxxx1	ARKKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLNSTNHAIVQTL					
Consensus	ARKKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLNSTNHAIVQTL					
Prim.cons.	ARKKKNKNCRRHSLYVDFSDVGWNDWIVAPPGYQAFYCHGDCPFPLADHLNSTNHAIVQTL					
	370	380	390	400		
S1xxxx0	VNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKKNYQEMVVEGCGCR					
S2xxxx1	VNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKKNYQEMVVEGCGCR					
Consensus	VNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKKNYQEMVVEGCGCR					
Prim.cons.	VNSVNSSIPKACCVPTELSAISMLYLDEYDKVVLKKNYQEMVVEGCGCR					

Συμπέρασμα: Οι S1(FASTA) και S2(Translate) είναι ίδιες, άρα οι αλληλουχίες είναι ίδιες και αρχίζουν να μοιάζουν από το 10^ο αμινοξύ.

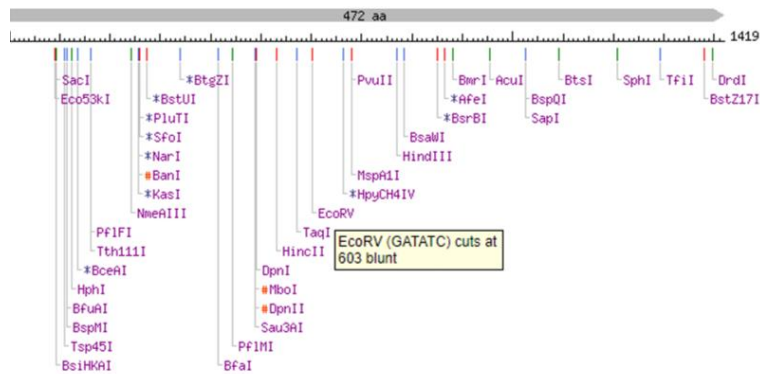
Στη συνέχεια, επιλέγουμε μερικά ένζυμα από το εργαλείο **NEBcutter V2.0** με βάση τον αριθμό θέσεων κοπής της DNA ακολουθίας: BamHI, EcoRI, KpnI, SmaI – δεν κόβουν καμία φορά την DNA αλληλουχία, EcoRV, HindIII, KasI – κόβουν την αλληλουχία μόνο 1 φορά, MseI, RsaI – κόβουν την αλληλουχία 2 φορές, HinfI, SfaNI – κόβουν την αλληλουχία 3 φορές. Το εργαλείο αυτό μας δίνει τα εξής αποτελέσματα:

EcoRV: κόβει στη θέση 603 της νουκλεοτιδικής αλληλουχίας και δίνει τυφλά άκρα στο DNA.

Οι θέσεις περιορισμού του ενζύμου είναι:

5' GAT|ATC 3'

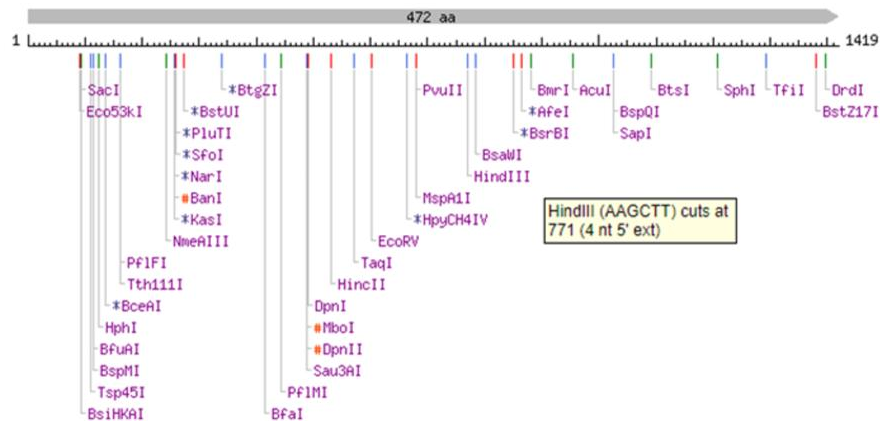
3' CTA|TAG 5'



HindIII: κόβει στη θέση 771 της νουκλεοτιδικής αλληλουχίας και δίνει 5' προεξέρχοντα άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

5' A|AGCTT 3'

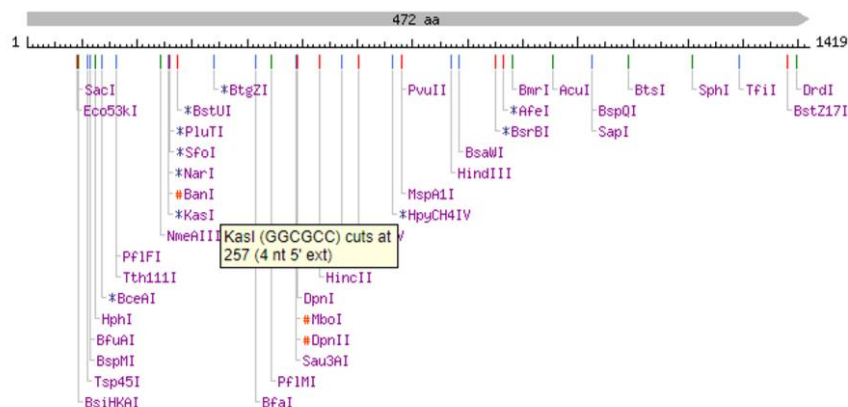
3' TTCGA|A 5'



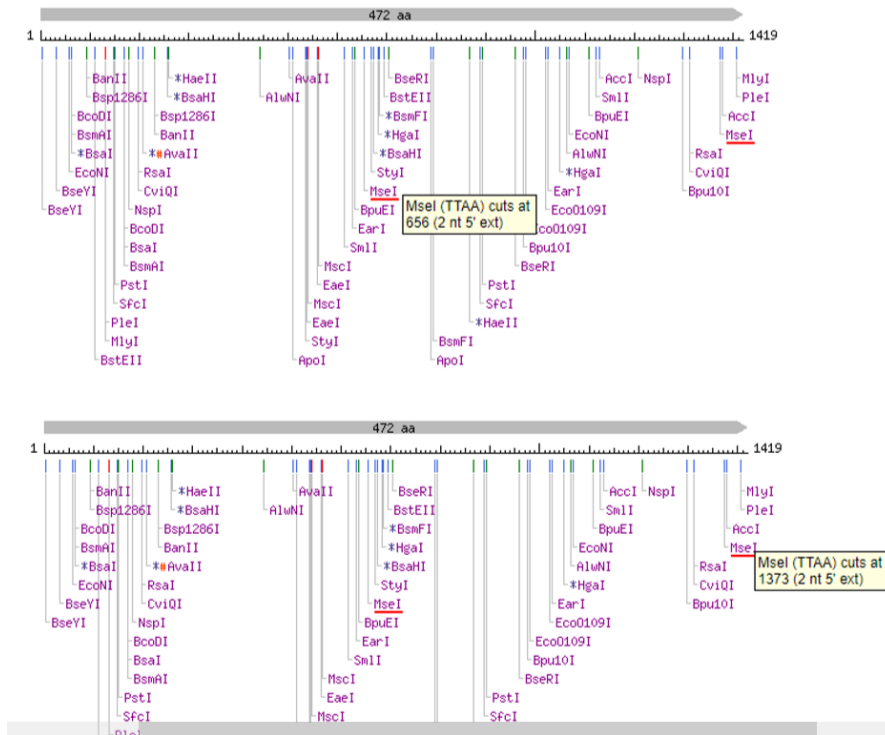
KasI: κόβει στη θέση 257 της νουκλεοτιδικής αλληλουχίας και δίνει 5' προεξέρχοντα άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

5' G|GCGCC 3'

3' CCGCG|G 5'



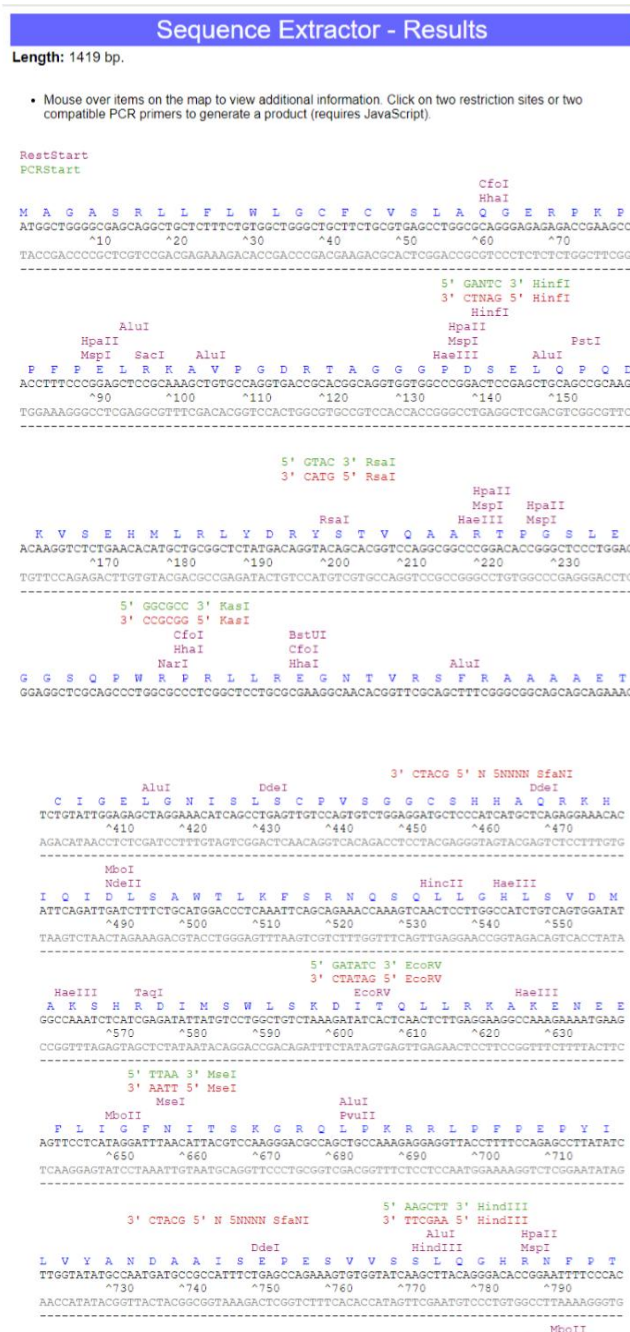
3' AAT|T 5'



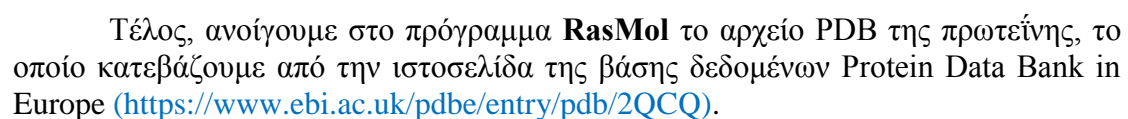
3' CA|TG 5'



Ακολουθώς, μπαίνουμε στην ιστοσελίδα <http://molbiol-tools.ca/PCR.htm>, επιλέγουμε το εργαλείο **Sequence Extractor** και στο πάνω παράθυρο εισαγάγουμε τη νουκλεοτιδική ακολουθία της πρωτεΐνης από τη βάση δεδομένων CCDS (ακολουθία DNA), ενώ στο κάτω παράθυρο – τις αλληλουχίες που αναγνωρίζει το συγκεκριμένο ένζυμο περιορισμού, ακολουθούνμενες από κενό και μετά το όνομα του ενζύμου (επιλέξαμε τα ένζυμα που παρουσιάστηκαν πριν). Παρακάτω παρουσιάζονται τα αποτελέσματα ανάλυσης ακολουθίας του DNA, όπου φαίνονται οι θέσεις κοπής από τα ένζυμα περιορισμού που επιλέξαμε.



Εάν θέλουμε να γίνει σύγκριση της πρωτοταγούς ακολουθίας της P12645 με άλλες όμοιες ή ομόλογες σε διαφορετικά ποσοστά κάθε φορά πρωτεΐνες, επιλέγουμε την ένδειξη BLAST στο πάνω αριστερό μέρος της σελίδας της πρωτεΐνης <https://www.uniprot.org/uniprot/P12645>. Παρακάτω παρουσιάζονται τα αποτελέσματα σύγκρισης με μία συγκεκριμένη σειρά αρχίζοντας από την πρωτεΐνη με την υψηλότερη ομολογία / ομοιότητα (μέχρι και την πρωτεΐνη με ποσοστό ομοιότητας 87,4%).



19

κυστεΐνη της πρωτεΐνης να χρωματίσουμε σε «μωβ» και να φαίνεται το συγκεκριμένο αμινοξύ μόνο σε μορφή ribbons, ενώ ολόκληρη η πρωτεΐνη να είναι σε μορφή wireframe:

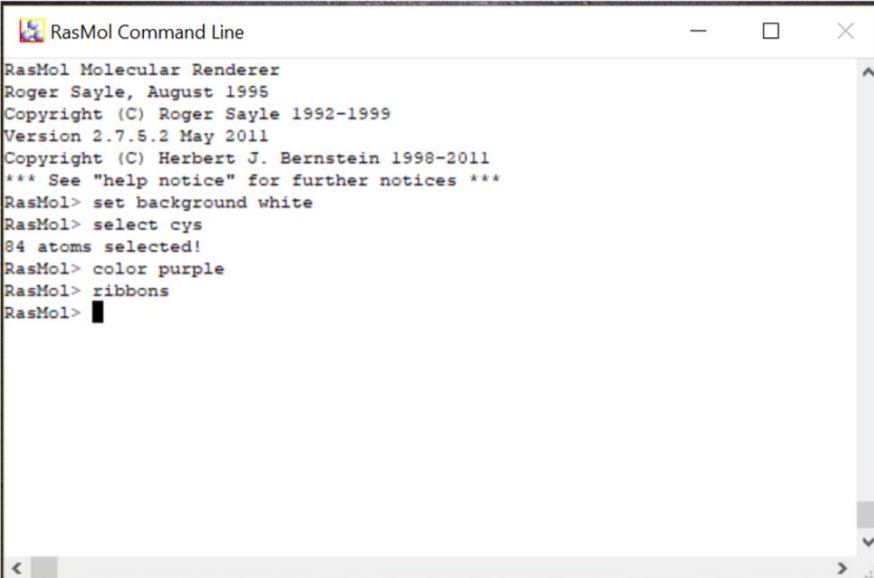
set background white

select cys

color purple

ribbons

Στην επόμενη σελίδα παρουσιάζονται οι εντολές στο **RasMol** και εικόνα της πρωτεΐνης που παίρνουμε:



```
RasMol Command Line
RasMol Molecular Renderer
Roger Sayle, August 1995
Copyright (C) Roger Sayle 1992-1999
Version 2.7.5.2 May 2011
Copyright (C) Herbert J. Bernstein 1998-2011
*** See "help notice" for further notices ***
RasMol> set background white
RasMol> select cys
84 atoms selected!
RasMol> color purple
RasMol> ribbons
RasMol> █
```

