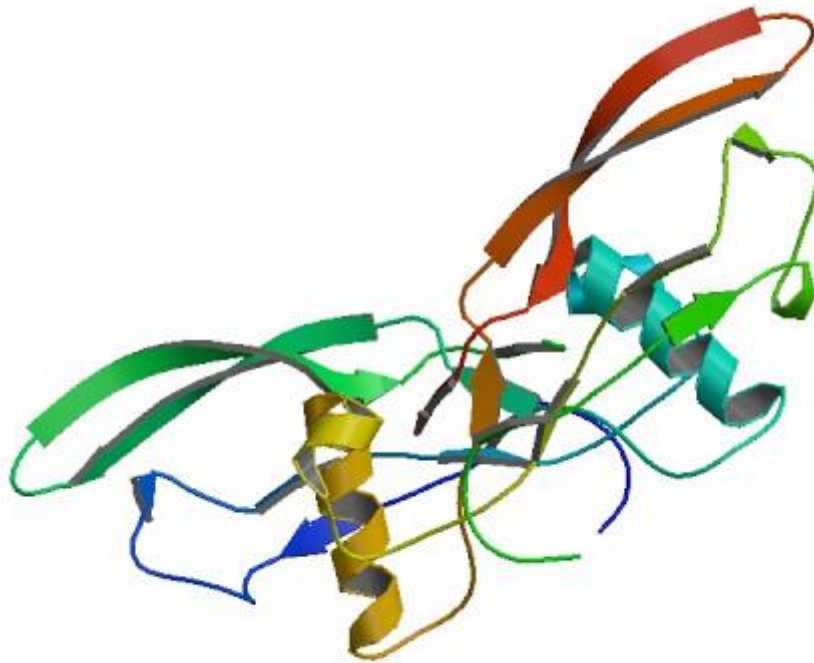


# Άσκηση στο μάθημα Δομική Βιοχημεία και Στοιχεία Βιοπληροφορικής

Αναφορά Αποτελεσμάτων



Καθηγ.: Χολή-Παπαδοπούλου Θεοδώρα

Επώνυμο / Όνομα: Heben Volha

Αριθμός Μητρώου: 9060

Ημερομηνία: 17/05/2019

Τίτλος: Οστική Μορφογενετική Πρωτεΐνη 3

Κωδικός **P12645**

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

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.

Πρωτεΐνη: οστική μορφογενετική πρωτεΐνη 3

Γονίδιο: BMP3

Οργανισμός: Homo sapiens (Human)

Λειτουργία: ρυθμίζει αρνητικά την οστική πυκνότητα. Ανταγωνίζεται την ικανότητα ορισμένων οστεογονικών BMP να προάγουν την οστεοπρογονική διαφοροποίηση (δηλαδή διαφοροποίηση των μυοκυττάρων σε οστεοκύτταρα) και οστεογένεση.

Η αμινοξική ακολουθία της πρωτεΐνης σε FASTA format:

```
>sp|P12645|BMP3_HUMAN Bone morphogenetic protein 3 OS=Homo sapiens OX=9606 GN=BMP3 PE=1 SV=2
MAGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGGGPDSELQPQDKVSEHM
LRLYDRYSTVQAARTPGSLEGGSQWRPRLREGNTVRSFRAAAETLERKGLYIFNLTS
LTKSENILSATLYFCIGELGNISLSCPVS GGCSHHAQRKHIQIDLSAWTLKFSRNQSQLL
GHLSVDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYI
LVYANDAAISEPESVSSLQGHRNFPTGTVPKWDSHIRAALSIERRKKRSTGVLLPLQNN
ELPGA EYQYKKDEVWEERKPYKTLQAQAEKSKNKKKQKRGPHRKSQTLQFDEQTLKKAR
RKQWIEPRNCARRYLKVDFA DIGWSEWIIISPKSFDAYYCSGACQFPMPKSLKPSNHATIQ
SIVRAVGVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESACAR
```

Βάσεις δεδομένων της νουκλεοτιδικής αλληλουχίας:

- EMBL
- GenBank
- DDBJ

**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.

```
TCAGCGTTGGAGTGGAGACGGCGCCCGCAGCGCCCTGCGCGGGTGAGGTCCGCGCAGCTG
CTGGGGAAGAGCCACCTGTGAGGCTGCGCTGGGTGAGCGCAGCAAGTGGGGCTGGCCGC
```

TATCTCGCTGCACCCGGCCGCTCCCGGGCTCCGTGCGCCCTCGCCCCAGCTGGTTTGA  
GTTCAACCCTCGGCTCCGCCGCCGGCTCCTTGCGCCTTCGGAGTGTCCCGCAGCGACGCC  
GGGAGCCGACGCGCCGCGCGGTACCTAGCCATGGCTGGGGCGAGCAGGCTGCTCTTTCT  
GTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGCAGGGAGAGAGACCGAAGCCACCTTTCCC  
GGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTGGTGGCCCGGACTCCGAGCT  
GCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAGCACGGT  
CCAGGCGGCCCCGGACACCGGGTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCCTCGGCT  
CCTGCGCGAAGGCAACACGGTTCGAGCTTTCCGGCGGCAGCAGCAGAACTCTTGAAAG  
AAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTTGTCTGC  
CACACTGTATTTCTGTATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGG  
AGGATGCTCCCATCATGCTCAGAGGAAACACATTCAGATTGATCTTTCTGCATGGACCT  
CAAATTCAGCAGAAACCAAAGTCAACTCCTTGCCATCTGTCAAGTGGATATGGCCAAATC  
TCATCGAGATATTATGTCCTGGCTGTCTAAAGATATCACTCAATTCTTGAGGAAGGCCAA  
AGAAAATGAAGAGTTCCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAA  
GAGGAGGTTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTT  
TGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAATTTTCCCACTGGAAGTGT  
TCCCAAATGGGATAGCCACATCAGAGCTGCCCTTTCCATTGAGCGGAGGAAGAAGCGCTC  
TACTGGGGTCTTGCTGCCTCTGCAGAACAACGAGCTTCTGGGGCAGAATACCAGTATAA  
AAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCAGGCTCAGGCCCTGA  
AAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGACGCTCCA  
ATTTGATGAGCAGACCCTGAAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTG  
CGCCAGGAGATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTC  
CCCCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCTCATGCCAAAGTC  
TTTGAAGCCATCAAATCATGCTACCATCCAGAGTATAGTGAGAGCTGTGGGGTCTGTTCC  
TGGGATTCTGAGCCTTGCTGTGTACCAGAAAAGATGTCCTCACTCAGTATTTTATTCTT  
TGATGAAAATAAGAATGTAGTGCTTAAAGTATACCCTAACATGACAGTAGAGTCTTGCGC  
TTGCAGATAACCTGGCAAAGAACTCATTTGAATGC

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

>ENA|AAH96269|AAH96269.1 Homo sapiens (human) bone morphogenetic protein 3  
ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTGAGCCTGGCG  
CAGGGAGAGAGACCGAAGCCACCTTTCCCGGAGCTCCGCAAAGCTGTGCCAGGTGACCGC  
ACGGCAGGTGGTGGCCCGGACTCCGAGCTGCAGCCGCAAGACAAGGTCTCTGAACACATG  
CTGCGGCTCTATGACAGGTACAGCACGGTCCAGGCGGCCCGGACACCGGGCTCCCTGGAG  
GGAGGCTCGCAGCCCTGGCGCCCTCGGCTCCTGCGCGAAGGCAACACGGTTTCGAGCTTT  
CGGGCGGCAGCAGCAGAACTCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATCG  
CTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGAGAGCTAGGA  
AACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACAC  
ATTGAGATTGATCTTTCTGCATGGACCTCAAATTCAGCAGAAACCAAAGTCAACTCCTT  
GGCCATCTGTCAAGTGGATATGGCCAAATCTCATCGAGATATTATGTCCTGGCTGTCTAAA  
GATATCACTCAATTCTTGAGGAAGGCCAAAGAAAATGAAGAGTTCCTCATAGGATTTAAC

ATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGTTACCTTTTCCAGAGCCTTATATC  
TTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATCAAGCTTACAG  
GGACACCGGAATTTTCCCACTGGAAGTGTCCCAAATGGGATAGCCACATCAGAGCTGCC  
CTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGCTGCCTCTGCAGAACAAAC  
GAGCTTCCTGGGGCAGAATACCAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCT  
TACAAGACCCTTCAGGCTCAGGCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAAG  
GGGCCTCATCGGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGAAAAAGGCAAGG  
AGAAAGCAGTGGAATTGAACCTCGGAATTGCGCCAGGAGATACCTCAAGGTAGACTTTGCA  
GATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTTTGATGCCTATTATTGCTCT  
GGAGCATGCCAGTTCCCCATGCCAAAGTCTTTGAAGCCATCAAATCATGCTACCATCCAG  
AGTATAGTGAGAGCTGTGGGGGTCTTCCTGGGATTCTGAGCCTTGCTGTGTACCAGAA  
AAGATGTCTCACTCAGTATTTTATTCTTTGATGAAAATAAGAATGTAGTGCTTAAAGTA  
TACCCTAACATGACAGTAGAGTCTTGCCTTGAGATAA

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

#### ORIGIN

```
1  tcagcgttgg agtggagacg gcgcccgcag cgccctgcgc gggtagagtc cgcgagctg
61  ctggggaaga gcccacctgt caggctgcgc tgggtcagcg cagcaagtgg ggctggccgc
121 tatctcgctg caccggcgcc cgtcccgggc tccgtgcgcc ctgccccag ctggtttggg
181 gttcaaccct cggtccgccc gccggctcct tgcgccttcg gagtgtcccg cagcgacgcc
241 gggagccgac gcgcccgcgc ggtacctagc catggctggg gcgagcaggc tgctctttct
301 gtggctgggc tgcttctgcg tagcctggc gcaggagag agaccgaagc cacctttccc
361 ggagctccgc aaagctgtgc caggtgaccg cacggcaggt ggtggcccg actccgagct
421 gcagccgcaa gacaaggtct ctgaacacat gctgcggctc tatgacaggt acagcacggt
481 ccaggcggcc cggacaccgg gctccctgga gggaggctcg cagccctggc gccctcggtc
541 cctgcgcgaa ggcaacacgg ttcgcagctt tcgggcggca gcagcagaaa ctcttgaag
601 aaaaggactg tatactttca atctgacatc gctaaccaag tctgaaaaca ttttgtctgc
661 cacactgtat ttctgtattg gagagctagg aaacatcagc ctgagttgtc cagtgtctgg
721 aggatgctcc catcatgctc agaggaaaca cattcagatt gatctttctg catggaccct
781 caaattcagc agaaaccaa gtcaactcct tggccatctg tcagtggata tggccaaatc
841 tcatcgagat attatgtcct ggctgtctaa agatatcact caattcttga ggaaggccaa
901 agaaaatgaa gaggttcctca taggatttaa cattacgtcc aagggacgcc agctgcaaaa
961 gaggaggtta ctttttcag agccttatat cttggtatat gccaatgatg ccgccatttc
1021 tgagccagaa agtgtggtat caagcttaca gggacaccgg aattttccca ctggaactgt
1081 tcccaaatgg gatagccaca tcagagctgc cttttccatt gagcggagga agaagcgctc
1141 tactggggtc ttgctgcctc tgcagaacaa cgagcttcct ggggcagaat accagtataa
1201 aaaggatgag gtgtgggagg agagaaagcc ttacaagacc cttcaggctc aggcccctga
1261 aaagagtaag aataaaaaga aacagagaaa ggggcctcat cggaagagcc agacgtcca
1321 atttgatgag cagaccctga aaaaggcaag gagaaagcag tggattgaac ctcggaattg
1381 cgccaggaga tacctcaagg tagactttgc agatattggc tggagtgaat ggattatctc
1441 cccaagtcc tttgatgcct attattgctc tggagcatgc cagttcccca tgccaaagtc
1501 tttgaagcca tcaaatcatg ctaccatcca gagtatagtg agagctgtgg gggctgttcc
```

```

1561 tgggattcct gagccttgct gtgtaccaga aaagatgtcc tcactcagta ttttattcctt
1621 tgatgaaaat aagaatgtag tgcttaaagt ataccctaac atgacagtag agtcttgcg
1681 ttgcagataa cctggcaaag aactcatttg aatgc

```

//

**GenBank – mRNA Translation (Accession Number AAH96269, Sequence Version AAH96269.1,**

**Keywords MGC.):**

ORIGIN

```

1  magasrllfl wlgcfcvsla qgerpkppfp elrkavpgdr tagggpdsel qpqdkvsehm
61  lrlydrystv qaartpgsle ggsqpwrprl lregntvrsf raanaetler kglyifnlts
121 ltksenilsa tlyfcigelg nislscpvsg gcshhaqrkh iqidlsawtl kfsrnqsqll
181 ghlsvdmaks hrdimswlsk ditqflrkak eneeffligfn itskgrqlpk rrlpfpepyi
241 lvyandaaais epesvssslq ghrnfptgtv pkwdshiraa lsierkkrs tgvllplqnn
301 elpgaeyqyk kdevweerkp yktlqaqape ksknkkkqrk gphrksqtlq fdeqtlkkar
361 rkqwieprnc arrylkvdfa digwsewiis pksfdayycs 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  tcagcgttgg agtggagacg gcgcccgcag cgccctgcgc gggtgaggtc cgcgcagctg
61  ctgggggaaga gcccacctgt caggctgcgc tgggtcagcg cagcaagtgg ggctggccgc
121 tatctcgctg caccggcgcc cgtcccgggc tccgtgcgcc ctgccccag ctggtttgga
181 gttcaaccct cggtccgcc gccggctcct tgcgccttcg gagtgtcccg cagcgacgcc
241 gggagccgac gcgccgcgcg ggtacctagc catggctggg gcgagcaggc tgctctttct
301 gtggctgggc tgcttctgcg tgagcctggc gcaggagag agaccgaagc cacctttccc
361 ggagctccgc aaagctgtgc cagggtgacc cagggcaggt ggtggcccg actccgagct
421 gcagccgcaa gacaaggtct ctgaacacat gctgcggctc tatgacaggt acagcacggt
481 ccaggcggcc cggacaccgg gctcccctga gggaggctcg cagccctggc gccctcggct
541 cctgcgcgaa ggcaacacgg ttcgcagctt tcgggcggca gcagcagaaa ctcttgaag
601 aaaaggactg tatatcttca atctgacatc gctaaccaag tctgaaaaca ttttgtctgc
661 cacactgtat ttctgtattg gagagctagg aaacatcagc ctgagttgtc cagtgtctgg
721 aggatgctcc catcatgtc agaggaaaca cattcagatt gatctttctg catggaccct
781 caaattcagc agaaacaaaa gtcaactcct tggccatctg tcagtggata tggccaaatc
841 tcatcgagat attatgtcct ggctgtctaa agatatcact caattcttga ggaaggccaa
901 agaaaatgaa gaggttcctca taggatttaa cattacgtcc aagggacgcc agctgccaaa
961 gaggaggtta cttttccag agccttata cttggtatat gccaatgatg ccgccatttc
1021 tgagccagaa agtgtggtat caagcttaca gggacaccgg aattttccca ctggaactgt
1081 tcccaaatgg gatagccaca tcagagctgc cttttccatt gagcggagga agaagcgctc
1141 tactgggggtc ttgttgccct tgcagaacaa cgagcttcct ggggcagaat accagtataa
1201 aaaggatgag gtgtgggagg agagaaagcc ttacaagacc cttcaggctc aggccctga
1261 aaagagtaag aataaaaaga aacagagaaa ggggcctcat cggaaagacc agacgtcca
1321 atttgatgag cagaccctga aaaaggcaag gagaaagcag tggattgaac ctcggaattg
1381 cgccaggaga tacctcaagg tagactttgc agatattggc tggagtgaat ggattatctc
1441 cccaagtcc tttgatgcct attattgctc tggagcatgc cagttcccca tgccaaagtc
1501 tttgaagcca tcaaatcatg ctaccatcca gagtatagtg agagctgtgg gggtcgttcc
1561 tgggattcct gagccttgct gtgtaccaga aaagatgtcc tcactcagta ttttattcctt
1621 tgatgaaaat aagaatgtag tgcttaaagt ataccctaac atgacagtag agtcttgcg
1681 ttgcagataa cctggcaaag aactcatttg aatgc

```

//

## **DDBJ – mRNA Translation (Accession Number AAH96269, Sequence Version AAH96269.1,**

**Keywords MGC.):**

BEGIN

```
1  MAGASRLFL  WLGCFCVSLA  QGERPKPPFP  ELRKAVPGDR  TAGGGPDSEL  QPQDKVSEHM
61  LRLYDRYSTV  QAARTPGSLE  GGSQPWRPRL  LREGNTRVSF  RAAAAETLER  KGLYIFNLTS
121 LTKSENILSA  TLYFCIGELG  NISLSCPVSQ  GCSHHAQRKH  IQIDLSAWTL  KFSRNQSQLL
181 GHLSVDMAKS  HRDIMSWLSK  DITQFLRKAK  ENEEFLLGFN  ITSKGRQLPK  RRLPFPEPYI
241 LUYANDAAIS  EPESVSSLQ  GHRNFPTGT  PKWDSHIRAA  LSIERRKKRS  TGVLLPLQNN
301 ELPGAQYQYK  KDEVWEERKP  YKTLQAQAP  KSKNKKKQK  GPHRKSQTLQ  FDEQTLKKAR
361 RKQWIEPRNC  ARRYLKVDFA  DIGWSEWII  PKSFDAYYCS  GACQFMPK  LKPSNHATIQ
421 SIVRAVGVP  GIPEPCVPE  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):**

```
ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGCAGGGAGAGA
GACCGAAGCCACCTTTCCCGGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTGGTGGCCCGGA
CTCCGAGCTGCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAGCACGGTC
CAGGCGGCGCCGACACCGGGCTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCCTCGGCTCCTGCGCGAAG
GCAACACGGTTCGAGCTTTGCGGCGGCAGCAGCAGAACTCTTGAAAGAAAAGGACTGTATATCTTCAA
TCTGACATCGCTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGAGAGCTAGGA
AACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACACATTGAGATTG
ATCTTTCTGATGGACCTCAAATTCAGCAGAAACCAAGTCAACTCCTTGGCCATCTGTGAGTGGATAT
GGCCAAATCTCATCGAGATATTATGCTCTGGCTGTCTAAAGATATCACTCAACTCTTGAGGAAGGCCAAA
GAAAATGAAGAGTTCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGTTAC
CTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATC
AAGCTTACAGGGACACCGGAATTTTCCACTGGAAGTGTCCCAAATGGGATAGCCACATCAGAGCTGCC
CTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGTGCTGCTGCTGAGAACACGAGCTTCCTG
GGGCAGAATACAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCTTCAGGCTCA
GGCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCTCATCGGAAGAGCCAGACGCTCCAA
TTTGATGAGCAGACCTGAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGAT
ACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTTTGATGCCTA
TTATTGCTCTGGAGCATGCCAGTTCCTCATGCCAAAGTCTTTGAAGCCATCAAATCATGCTACCATCCAG
AGTATAGTGAGAGCTGTGGGGGTGTTCTGTTGGGATTCTGAGCCTTGCTGTGTACCAGAAAAGATGTCCT
CACTCAGTATTTTATTCTTTGATGAAAATAAGAATGTAGTGCTTAAAGTATACCCTAACATGACAGTAGA
GTCTTGCCTTGCAGATAA
```

### **Translation (472 aa):**

```
MAGASRLFLWLGCFCVSLAQGERPKPPFELRKAVPGDRTAGGGPDSELQPQDKVSEHMLRLYDRYSTV
QAARTPGSLEGGSQWRPRLREGNTRVSFRAAAAETLERKGLYIFNLSTLTKSENILSATLYFCIGELG
NISLSCPVSQGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLLGHLSVDMAKSHRDISWLSKDITQLLRKAK
ENEFLIFNITSKGRQLPKRRLPFPEPYILUYANDAAISEPESVSSLQGHRNFPTGTPKWDSHIRAA
LSIERRKKRSTGVLLPLQNNELPGAQYQYKDEVWEERKPYKTLQAQAPEKSKNKKKQKQPHRKSQTLQ
FDEQTLKKARRKQWIEPRNCARRYLKVDFA DIGWSEWII PKSFDAYYCSGACQFMPKSLKPSNHATIQ
SIVRAVGVPGIPEPCVPEKMSSLSILFFDENKNVVLKVYPNMTVESCA CR
```

Στη συνέχεια μεταβαίνουμε στη σελίδα <https://www.expasy.org/proteomics> και επιλέγουμε το εργαλείο [Translate](#) • nucleotide sequence translation • [\[more\]](#). Έπειτα, εισαγάγουμε μια ακολουθία DNA

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 AGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGG  
GPDSELQPQDKVSEH Met LRLYDRYSTVQAARTPGSLEGGSQPWRP  
RLLREGNTVRSFRAAAAETLERKGLYIFNLTSLTKSENILSATLYFCI  
GELGNISLSCPVS GGCSHHAQRKHIQIDLSAWTLKFSRNQSQLLGH  
LSVD Met AKSHRDI Met SWLSKDITQLLRKAKENEEFLIGFNITSKGRQ  
LPKRRLPFPEPYILVYANDAAISEPESSVSSLQGHRNFPTGTVPKW  
DSHIRAALSIERRKKRSTGVLLPLQNNELPGA EYQYKKDEVWEERK  
PYKTLQAQAPEKSKNKKKQRKGPHRKSQTLQFDEQTLKKARRKQW  
IEPRNCARRYLKVDFADIGWSEWIISP KSFDAYYCSGACQFP Met PK  
SLKPSNHATIQSIVRAVG VVPGIPEPCCVPEK Met SSLSILFFDENKN  
VVLKVYPN Met TVESCACR Stop

#### 3'5' Frame 1

LSASARLYCHVRVYFKHYILIFIKE Stop NTE Stop GHLFWYTARLRNPR  
NDPHSSHYTLDGS Met I Stop WLQRLWHGELACSR AIIGIKGLGGDNP  
FTPANICKVYLEVSPGAIPRFNPLLSPCLFQGLLIKLERLALP Met RPL  
SLFLFILTLFRGLSLKGLVRLSLLPHLILFILVFCPRKLVVLQRQQDP  
SRALLPPLNGKGSSDVAIPFGNSSSGKIPVSL Stop A Stop YHTFWLRN  
GGIIGIYQDIRLWKR Stop PPLWQLASLGRNVKSYEELFIFFGLPQELS  
DIFRQPGHNIS Met RFGHIH Stop Q Met AKELTLVSAEFEGPCRKINLNV  
FPLS Met Met GASSRHWTTQADVS Stop LSNT EIQCGRQNVFRLG Stop R  
CQIEDIQSFSFKS FCCCRPKAANRVAFAQEPRAPGLRASLQGARC  
PGRLDRAVPVIEPQHVFRLDLVRLQLGVRATTCRAVTWHSFAELRER  
WLRSLSLRQAHA EAAQPQKEQPARPSH

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:

GCAGCAGAAACTCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTGTCTGCCACACTGTATTTCTG  
TATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACACATTAGATTGATCTTTCTG  
CATGGACCCTCAAATTCAGCAGAAACCAAAGTCAACTCCTTGGCCATCTGTCTAGTGGATATGGCCAAATCTCATCGAGATATTATGTCCTGGCTG  
TCTAAAGATATCACTCAACTCTTGAGGAAGGCCAAAGAAAATGAAGAGTTCCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAA  
GAGGAGGTTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGAC  
ACCGGAATTTTCCACTGGAACGTTCCTCAATGGGATAGCCACATCAGAGCTGCCCTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTC

TTGCTGCCTCTGCAGAACACGAGCTTCCTGGGGCAGAATACCAAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCA  
GGCTCAGGCCCTGAAAAGAGTAAGAATAAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGA  
AAAAGGCAAGGAGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATT  
ATCTCCCCCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCCATGCCAAAG

Translation:

#### 5'3' Frame 1

AAETLERKGLYIFNLTSLTKESENILSATLYFCIGELGNISLSCPVS G G  
CSHHAQRKHIQIDLSAWTLKFSRNQS QLLGHL SVD **Met** AKSHRDI **Met**  
SWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYILVY  
ANDAAISEPESV VSSLQGHRNFPTGTVPKWDSHIRAALS IERRKKR  
STGVLLPLQNNELPGA EYQYKKDEVWEERKPYKTLQAQ APEKSKN  
KKKQRKGPHRKSQTLQFDEQTLKKARRKQWIEPRNCARRY LKVDF  
ADIGWSEWIISP K SFDAYYCSGACQFP **Met** PK

#### 3'5' Frame 1

LWHGELACSR A IIGIKGLGGDNPFT PANICKVYLEVSPGAIPRFNPL  
LSPCLFQGLLIKLERLALP **Met** RPLSLFLFILTLFRGLSLKGLVRLSLL  
PHLILFILVFCPRKLVVLQRQQDPSRALLPPLNGKGSSDVAIPFGNS  
SSGKIPVSL **Stop** A **Stop** YHTFWLRNGGIIGIYQDIRLWKR **Stop** PPLWQL  
ASLGRNVKSYEELFIFFGLPQELSDIFRQPGHNIS **Met** RFGHIH **Stop** Q  
**Met** AKELTLVSAEFEGPCRKINLNVFPLS **Met** **Met** GASSRHWTTQADVS  
**Stop** LSNTEIQCGRQNVFRLG **Stop** RCQIEDIQSFSFKSFCC

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:

AAACTCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGA  
GAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACACATTGAGATTGATCTTTCTGCATGGAC  
CCTCAAATTCAGCAGAAACCAAGTCAACTCCTTGCCATCTGTAGTGGATATGGCCAAATCTCATCGAGATATTATGCTCTGGCTGTCTAAAG  
ATATCACTCAACTCTTGAGGAAGGCCAAAGAAAATGAAGAGTTCCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGG  
TTACCTTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAA  
TTTTCCCACTGGAAGTGTCCCAAATGGGATAGCCACATCAGAGCTGCCCTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGCTGC  
CTCTGCAGAACACGAGCTTCTGGGGCAGAATACCAAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCAGGCTCAG  
GCCCCTGAAAAGAGTAAGAATAAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGACGCTCCAATTTGATGAGCAGACCCTGAAAAGGC  
AAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCC  
CCAAGTCCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCCATGCCAAAG

Translation:

#### 5'3' Frame 3

TLERKGLYIFNLTSLTKESENILSATLYFCIGELGNISLSCPVS G G CSH  
HAQRKHIQIDLSAWTLKFSRNQS QLLGHL SVD **Met** AKSHRDI **Met** SWL  
SKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYILVYAND



AAISEPESVSSSLQGHRNFPTGTVPKWDSHIRAALSIERRKKRSTG  
VLLPLQNNELPGA EYQYKKDEVWEERKPYKTLQAQAPEKSKNKKK  
QRKGPHRKSQTLQFDEQTLKKARRKQWIEPRNCARRYLKVDFADI  
GWSEWIISPKSFDAYYC SGACQFP **Met** PK

### 3'5' Frame 1

LWHGELACSR A IIGIKGLGGDNPFTPANICKVYLEVSPGAIPRFNPL  
LSPCLFQG L LIKLERLALP **Met** RPLSLFLFILTLFRGLSLKGLVRLSL  
PHLILFILVFCPRKLVVLQRQQDPSRALLPPLNGKGS SDVAIPFGNS  
SSGKIPVSL **Stop** A **Stop** YHTFWLRNGGIIGIYQDIRLWKR **Stop** PPLWQL  
ASLGRNVKSYEELFIFFGLPQELSDIFRQPGHNIS **Met** RFGHIH **Stop** Q  
**Met** AKELTLVSAEFEGPCRKINLNVFPLS **Met** **Met** GASSRHWTTQADV S  
**Stop** LSNTEIQCGRQNVFRLG **Stop** RCQIEDIQSFSFKS

Έπειτα, κάνουμε την πολλαπλή ομοπαράθεση (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 ομοπαράθεση 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
BMP3_HUMAN						
sequence1	MAGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGGGPDSELQPQDKVSEHM					
sequence2	-----					
Consensus						
Prim.cons.	MAGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGGGPDSELQPQDKVSEHM					
	70	80	90	100	110	120
BMP3_HUMAN						
sequence1	LRLYDRYSTVQAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAATLERKGLYIFNLTS					
sequence2	-----FRAAAATLERKGLYIFNLTS					
Consensus						
Prim.cons.	LRLYDRYSTVQAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAATLERKGLYIFNLTS					
	130	140	150	160	170	180
BMP3_HUMAN						
sequence1	LTKSENILSATLYFCIGELGNISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLL					
sequence2	LTKSENILSATLYFCIGELGNISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLL					
Consensus	LTKSENILSATLYFCIGELGNISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLL					
Prim.cons.	LTKSENILSATLYFCIGELGNISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLL					
	190	200	210	220	230	240
BMP3_HUMAN						
sequence1	GHLSDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYI					
sequence2	GHLSDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYI					
Consensus	GHLSDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYI					
Prim.cons.	GHLSDMAKSHRDIMSWLSKDITQLLRKAKENEEFLIGFNITSKGRQLPKRRLPFPEPYI					
	250	260	270	280	290	300
BMP3_HUMAN						
sequence1	LVYANDAAISEPESVSSLQGHRNFTGTVPKWDSHIRAALSIERRKKRSTGVLLPLQNN					
sequence2	LVYANDAAISEPESVSSLQGHRNFTGTVPKWDSHIRAALSIERRKKRSTGVLLPLQNN					
Consensus	LVYANDAAISEPESVSSLQGHRNFTGTVPKWDSHIRAALSIERRKKRSTGVLLPLQNN					
Prim.cons.	LVYANDAAISEPESVSSLQGHRNFTGTVPKWDSHIRAALSIERRKKRSTGVLLPLQNN					
	310	320	330	340	350	360
BMP3_HUMAN						
sequence1	ELPGAQYQKKDEVWEERKPYKTLQAQAPKSKNKKKQKRGPHRKSQTLQFDEQTLKKAR					
sequence2	ELPGAQYQKKDEVWEERKPYKTLQAQAPKSKNKKKQKRGPHRKSQTLQFDEQTLKKAR					
Consensus	ELPGAQYQKKDEVWEERKPYKTLQAQAPKSKNKKKQKRGPHRKSQTLQFDEQTLKKAR					
Prim.cons.	ELPGAQYQKKDEVWEERKPYKTLQAQAPKSKNKKKQKRGPHRKSQTLQFDEQTLKKAR					

	370	380	390	400	410	420
BMP3_HUMAN	RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFMPK	SLKPSNHATIQ				
sequence1	RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFMPK	-----				
sequence2	RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFMPK	-----				
Consensus	RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFMPK					
Prim.cons.	RKQWIEPRNCARRYLKVDFADIGWSEWIISPKSFDAYYCSGACQFMPKSLKPSNHATIQ					

	430	440	450	460	470
BMP3_HUMAN	SIVRAVGVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESCACR				
sequence1	-----				
sequence2	-----				
Consensus					
Prim.cons.	SIVRAVGVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESCACR				

**Alignment data :**  
Alignment length : 472  
Residues conserved for 90 % or more (upper-case letters) : 303 is 64.19 %  
Residues conserved for 50 % and less than 90 % (lower-case letters) : 7 is 1.48 %  
Residues conserved less than 50 % (white space) : 162 is 34.32 %  
V conserved positions (!) : 0 is 0.00 %  
\_M conserved positions (\$) : 0 is 0.00 %

Στην επόμενη άσκηση μελετούνται τα ένζυμα περιορισμού αλληλουχίας DNA της πρωτεΐνης BMP 3. Πληκτρολογούμε τη διεύθυνση <http://nc2.neb.com/NEBcutter2/> του εργαλείου **NEBcutter V2.0**, εμφανίζεται ένα παράθυρο στο οποίο εισαγάγουμε τη νουκλεοτιδική ακολουθία της πρωτεΐνης από τη βάση δεδομένων CCDS, κάτω από το παράθυρο επιλέγουμε το NEB enzymes και πατάμε Submit:

← → ↺ ⌂ ⓘ Μη ασφαλής | nc2.neb.com/NEBcutter2/cutshow.php?name=28674403-FASTA

Εφαρμογές

**NEBcutter**

Linear Sequence: FASTA

Display: - NEB single cutter restriction enzymes  
- Main non-overlapping, min. 100 aa ORFs

GC=51%, AT=49%

Cleavage code  
✗ | blunt end cut  
⬇ | 5' extension  
⬆ | 3' extension  
⬇ | cuts 1 strand

Enzyme name code  
Available from NEB  
Has other supplier  
Not commercially available  
\*: cleavage affected by CpG meth.  
.: cleavage affected by other meth.  
(enz.name): ambiguous site

1 472 aa 1419

SacI Eco53kI BstXI BstII PvuII BsrI AclI BtsI SphI TfiI DndI BstZ171  
EcoRI BstUI PfuI SfoI NarI BstI KasI NheIII PstI HincII DpnI HpaI BfuAI BspMI Tsp4SI BstH9AI  
BstEII HhaI BspRI HincII DpnI HpaI BfuAI BspMI Tsp4SI BstH9AI  
BstEII HhaI BspRI HincII DpnI HpaI BfuAI BspMI Tsp4SI BstH9AI

Main options  
New DNA  
Custom digest  
View sequence  
ORF summary  
Save project  
Print

Availability  
All commercial  
All

Display  
2 cutters  
3 cutters

Zoom  
Zoom in  
More...

Minimum ORF length to display: 100 aa OK

List  
0 cutters  
1 cutters  
All sites  
Save all sites  
Flanking enzymes

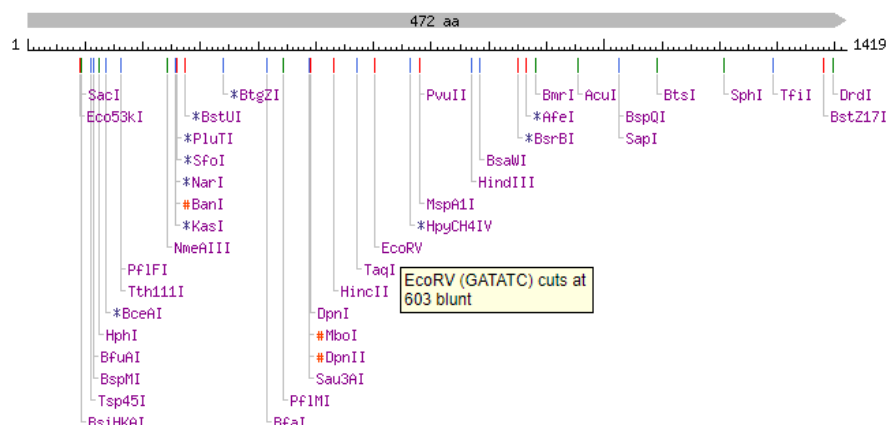
Στη συνέχεια, επιλέγουμε μερικά ένζυμα από το εργαλείο **NEBcutter V2.0** με βάση τον αριθμό θέσεων κοπής της DNA ακολουθίας: BamHI, EcoRI, KpnI, SmaI – δεν κόβουν καμία φορά την DNA αλληλουχία, EcoRV, HindIII, KasI – κόβουν την αλληλουχία μόνο 1 φορά, MseI, RsaI – κόβουν την αλληλουχία 2 φορές, HinfI,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

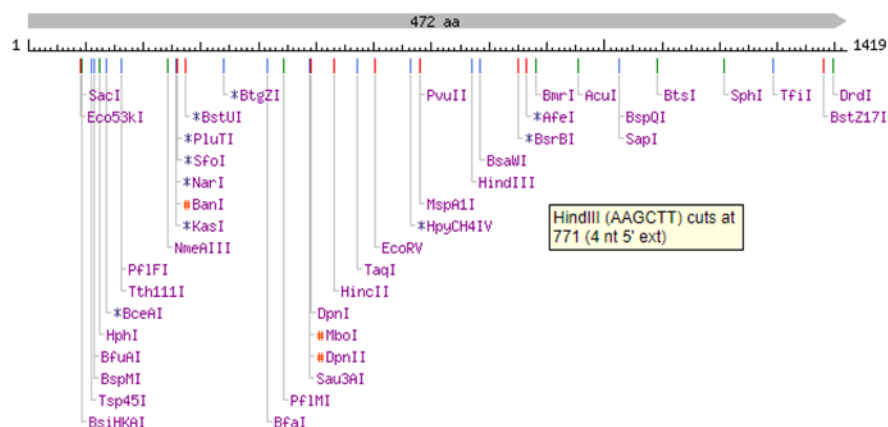
3' CTA|TAG 5'



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

5' A|AGCTT 3'

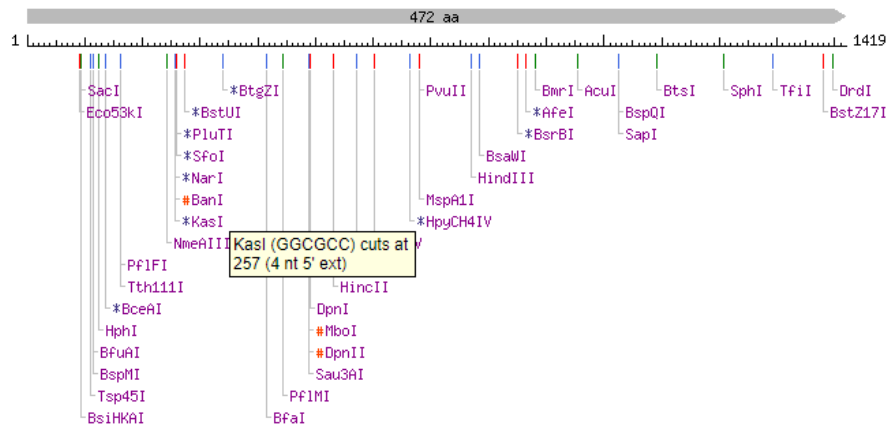
3' TTCGA|A 5'



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

5' G|GCGCC 3'

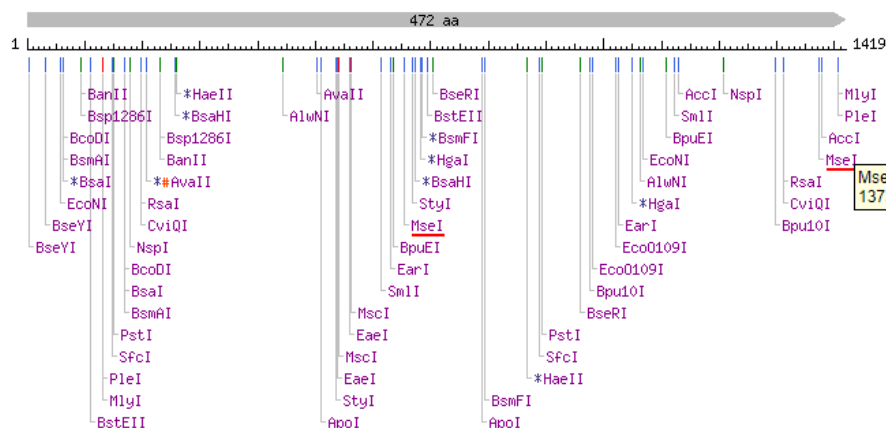
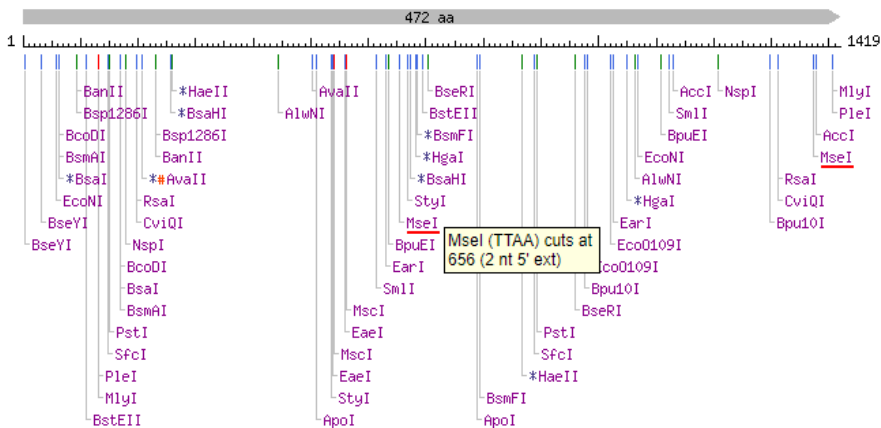
3' CCGCG|G 5'



MseI: κόβει στις θέσεις 656 και 1373 της νουκλεοτιδικής αλληλουχίας και δίνει 5' προεξέρχοντα άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

5' T|TAA 3'

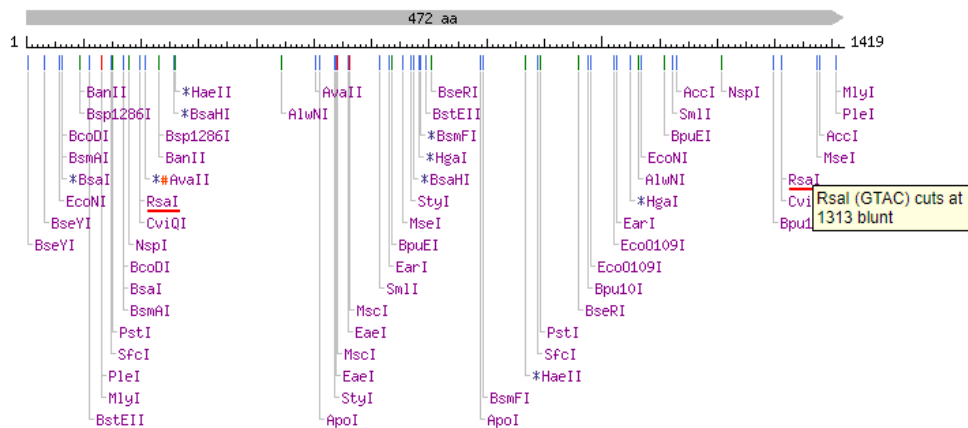
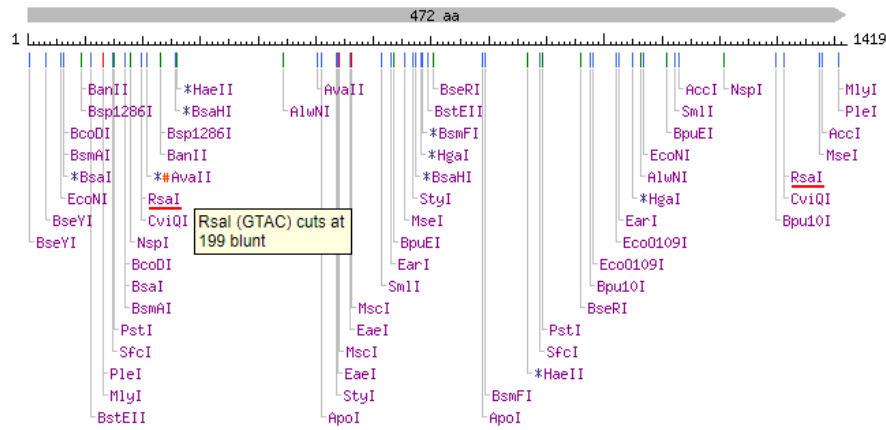
3' AAT|T 5'



RsaI: κόβει στις θέσεις 199 και 1313 της νουκλεοτιδικής αλληλουχίας και δίνει τυφλά άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

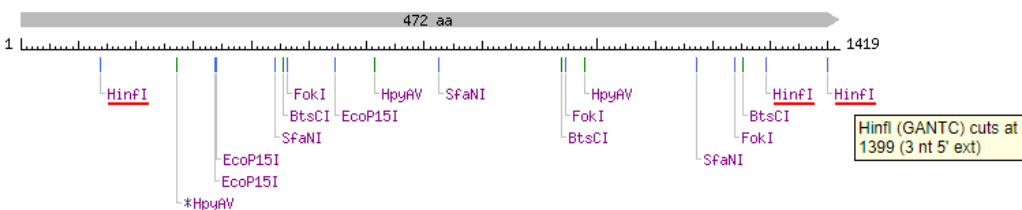
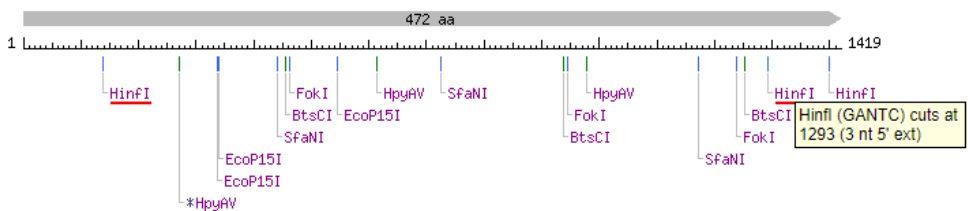
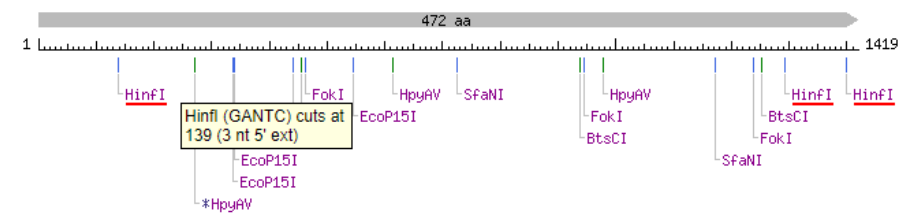
5' GT|AC 3'

3' CA|TG 5'



HinfI: κόβει στις θέσεις 139, 1293 και 1399 της νουκλεοτιδικής αλληλουχίας και δίνει 5' προεξέρχοντα άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

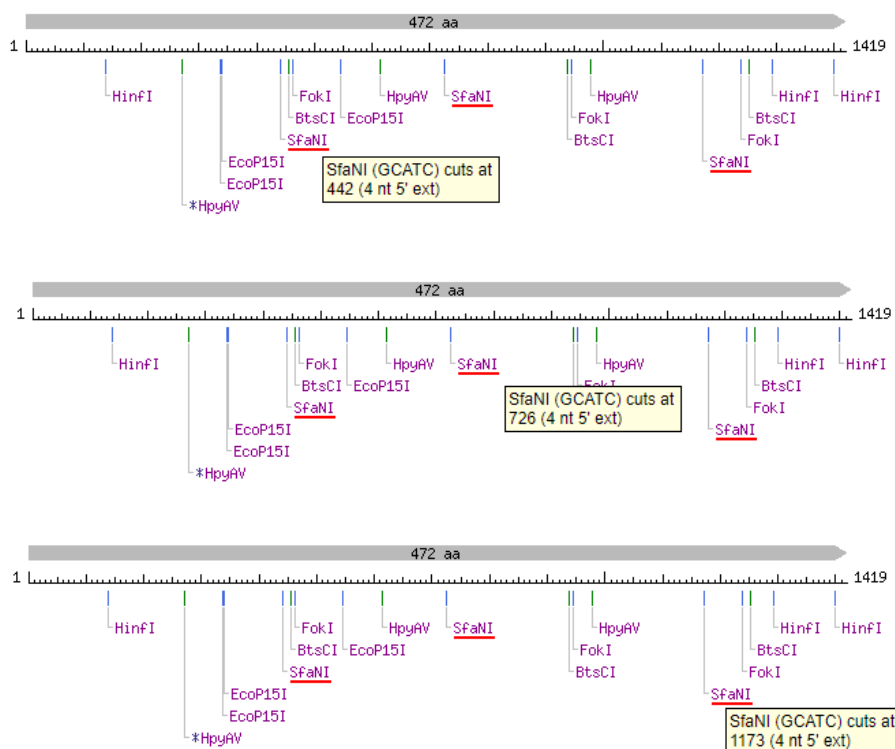
5' G|ANTC 3'  
3' CTNA|G 5'



SfaNI: κόβει στις θέσεις 442, 726 και 1173 της νουκλεοτιδικής αλληλουχίας και δίνει 5' προεξέρχοντα άκρα στο DNA. Οι θέσεις περιορισμού του ενζύμου είναι:

5' GCATC(N)<sub>5</sub>| 3'

3' CGTAG(N)<sub>9</sub>| 5'



Ακολουθως, 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, ακολουθούμενες 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.

# Sequence Extractor - Results

Length: 1419 bp.

- Mouse over items on the map to view additional information. Click on two restriction sites or two compatible PCR primers to generate a product (requires JavaScript).

RestStart

PCRStart

```

                                     CfoI
                                     HhaI
M  A  G  A  S  R  L  L  F  L  W  L  G  C  F  C  V  S  L  A  Q  G  E  R  P  K  P
ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGCAGGGAGAGACCGAAGCC
      ^10      ^20      ^30      ^40      ^50      ^60      ^70
TACCGACCCCGCTCGTCCGACGAGAAAGACACCGACCCGACGAAGACGCACTCGGACCGCGTCCCTCTCTTGGCTTCGG
-----
                                     5' GATC 3' HinfI
                                     3' CTNAG 5' HinfI
                                     HinfI
      AluI
      HpaII
      MspI
P  F  P  E  L  R  K  A  V  P  G  D  R  T  A  G  G  G  P  D  S  E  L  Q  P  Q  D
ACCTTTCCCGGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTGGTGGCCCGGACTCCGAGCTGCAGCCGCAAG
      ^90      ^100      ^110      ^120      ^130      ^140      ^150
TGGAAAGGGCCTCGAGGCGTTTCGACACGGTCCACTGGCGTGCCGTCCACCACCGGGCCTGAGGCTCGACGTCGGCGTTC
-----
                                     5' GTAC 3' RsaI
                                     3' CATG 5' RsaI
                                     HpaII
                                     MspI
      RsaI
      HaeIII
      AluI
      PstI
K  V  S  E  H  M  L  R  L  Y  D  R  Y  S  T  V  Q  A  A  R  T  P  G  S  L  E
ACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAGCACGGTCCAGGCGGGCCCGGACACCGGGCTCCCTGGAG
      ^170      ^180      ^190      ^200      ^210      ^220      ^230
TGTTCAGAGACTTGTGTACGACGCCGAGATACTGTCCATGTCGTGCCAGGTCCGCCGGGCCTGTGGCCCGAGGGACCTC
-----
      5' GGCGCC 3' KsiI
      3' CCGCGG 5' KsiI
      CfoI
      HhaI
      NarI
      BstUI
      CfoI
      HhaI
      AluI
G  G  S  Q  P  W  R  P  R  L  L  R  E  G  N  T  V  R  S  F  R  A  A  A  A  E  T
GGAGGCTCGCAGCCCTGGCGCCCTCGGCTCCTGCGCGAAGGCAACACGGTTCGAGCTTTTCGGGCGGCAGCAGCAGAAAC
      ^250      ^260      ^270      ^280      ^290      ^300      ^310
CCTCCGAGCGTCGGGACCGCGGGAGCCGAGGACGCGCTTCCGTTGTGCCAAGCGTCGAAAGCCCGCCGTCGTCGTCTTTG
-----
      MboII
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
TCTTGAAAGAAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTTGTCTGCCACACTGTATT
      ^330      ^340      ^350      ^360      ^370      ^380      ^390
AGAACTTTCTTTTCTGACATATAGAAGTTAGACTGTAGCGATTGGTTCAGACTTTTGTAAACAGACGGTGTGACATAA
-----
```



3' CTACG 5' N 5NNNN SfaNI

AluI DdeI DdeI

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

TCTGTATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACAC

^410 ^420 ^430 ^440 ^450 ^460 ^470

AGACATAACCTCTCGATCCTTTGTAGTCGGACTCAACAGGTCACAGACCTCCTACGAGGGTAGTACGAGTCTCCTTTGTG

-----

MboI NdeII HincII HaeIII

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 M

ATTCAGATTGATCTTTCTGCATGGACCCTCAAATTCAGCAGAAACCAAAGTCAACTCCTTGGCCATCTGTCTCAGTGGATAT

^490 ^500 ^510 ^520 ^530 ^540 ^550

TAAGTCTAACTAGAAAGACGTACCTGGGAGTTTAAGTCGTCTTTGGTTTCAGTTGAGGAACCGGTAGACAGTCACCTATA

-----

5' GATATC 3' EcoRV

3' CTATAG 5' EcoRV

HaeIII TaqI EcoRV HaeIII

A K S H R D I M S W L S K D I T Q L L R K A K E N E E

GGCCAAATCTCATCGAGATATTATGTCTGGCTGTCTAAAGATATCACTCAACTCTTGAGGAAGGCCAAAGAAAATGAAG

^570 ^580 ^590 ^600 ^610 ^620 ^630

CCGTTTGTAGTAGCTCTATAATACAGGACCGACAGATTTCTATAGTGAGTTGAGAACTCCTTCCGGTTTCTTTTACTTC

-----

5' TTAA 3' MseI

3' AATT 5' MseI

MseI

MboII AluI PvuII

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

AGTTCCTCATAGGATTTAATACATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGGTTACCTTTTCCAGAGCCTTATATC

^650 ^660 ^670 ^680 ^690 ^700 ^710

TCAAGGAGTATCCTAAATTGTAATGCAGGTTCCCTGCGGTCGACGGTTTCTCCTCCAATGGAAGGTCTCGGAATATAG

-----

5' AAGCTT 3' HindIII

3' TCGAA 5' HindIII

3' CTACG 5' N 5NNNN SfaNI

AluI HpaII

HindIII MspI

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

TTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAATTTTCCAC

^730 ^740 ^750 ^760 ^770 ^780 ^790

AACCATATACGGTTACTACGGCGGTAAAGACTCGGTCTTTTACACCATAGTTCGAATGTCCCTGTGGCCTTAAAGGGTG

-----

MboII

CfoI HhaI

AluI

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

TGGAACGTGTTCCCAAATGGGATAGCCACATCAGAGCTGCCCTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCT

^810 ^820 ^830 ^840 ^850 ^860 ^870

ACCTTGACAAGGGTTTACCCTATCGGTGTAGTCTCGACGGGAAAGGTAAGTTCGCTCCTTCTTCGCGAGATGACCCCA

-----

PstI AluI

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

TGCTGCCTCTGCAGAACACGAGCTTCCTGGGGCAGAATACCAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCT

^890 ^900 ^910 ^920 ^930 ^940 ^950

ACGACGGAGACGTCTTGTGCTCGAAGGACCCCGTCTTATGGTCATATTTTCTTACTCCACACCCTCCTCTCTTTTCGGA

-----

HaeIII

DdeI HaeIII

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

TACAAGACCCTTCAGGCTCAGGCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCA

^970 ^980 ^990 ^1000 ^1010 ^1020 ^1030

ATGTTCTGGGAAGTCCGAGTCCGGGGACTTTTCTCATTCTTATTTTCTTTGTCTCTTTCCCGGAGTAGCCTTCTCGGT

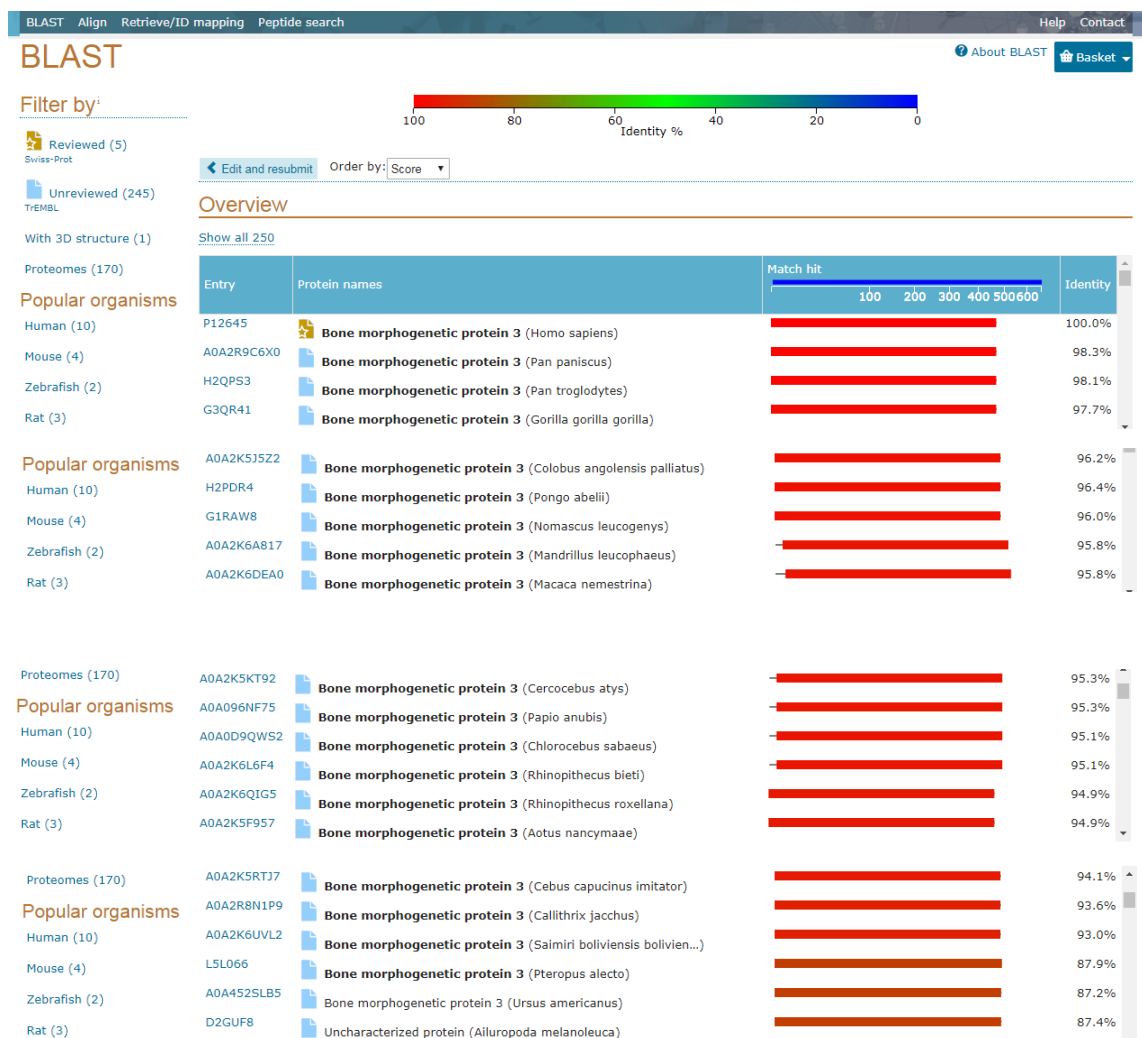
-----

```

MboII
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
GACGCTCCAATTTGATGAGCAGACCCTGAAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGAT
^1050 ^1060 ^1070 ^1080 ^1090 ^1100 ^1110
CTGCGAGGTTAAACTACTCGTCTGGGACTTTTTCCGTTCTCTTCGTCACCTAACTTGGAGCCTTAACGCGGTCTCTCTA
-----
3' CTACG 5' N 5NNNN S
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
ACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTTTGATGCCTATTATTGCTCT
^1130 ^1140 ^1150 ^1160 ^1170 ^1180 ^1190
TGGAGTTCATCTGAAACGTCTATAACCGACCTCACTTACCTAATAGAGGGGGTTTACGAACTACGGATAATAACGAGA
-----
faNI
SphI
G A C Q F P M P K S L K P S N H A T I Q S I V R A V G
GGAGCATGCCAGTTCCCATGCCAAAGTCTTTGAAGCCATCAAATCATGCTACCATCCAGAGTATAGTGAGAGCTGTGGG
^1210 ^1220 ^1230 ^1240 ^1250 ^1260 ^1270
CCTCGTACGGTCAAGGGGTACGGTTTCAGAACTTCGGTAGTTTAGTACGATGGTAGGTCTCATATCACTCTCGACACCC
-----
5' GANTC 3' HinfI 5' GTAC 3' RsaI
3' CTNAG 5' HinfI 3' CATG 5' RsaI
DdeI
HinfI RsaI DdeI
V V P G I P E P C C V P E K M S S L S I L F F D E N K
GGTCGTTCTGGGATTCTGAGCCTTGCTGTGTACCAGAAAAGATGTCCTCACTCAGTATTTTATTCTTTGATGAAAATA
^1290 ^1300 ^1310 ^1320 ^1330 ^1340 ^1350
CCAGCAAGGACCCTAAGGACTCGGAACGACACATGGTCTTTTCTACAGGAGTGAGTCATAAAATAAGAACTACTTTTAT
-----
RestEnd
PCREnd
5' TTAA 3' MseI 5' GANTC 3' HinfI
3' AATT 5' MseI 3' CTNAG 5' HinfI
CfoI
MseI HinfI HhaI
N V V L K V Y P N M T V E S C A C R Z
AGAATGTAGTGCTTAAAGTATAACCCTAACATGACAGTAGAGTCTTGCGCTTGCGAGATAA
^1370 ^1380 ^1390 ^1400 ^1410
TCTTACATCACGAATTTTCATATGGGATTGTACTGTCTCATCTCAGAACGCGAACGTCTATT
-----

```

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%).



Τέλος, ανοίγουμε στο πρόγραμμα **RasMol** το αρχείο PDB της πρωτεΐνης, το οποίο κατεβάζουμε από την ιστοσελίδα της βάσης δεδομένων Protein Data Bank in Europe (<https://www.ebi.ac.uk/pdbe/entry/pdb/2QCQ>).

Στο παράθυρο με τις εντολές σε μορφή κειμένου, πληκτρολογούμε τις παρακάτω εντολές, ώστε να αλλάξουμε το background σε «λευκό», το αμινοξύ 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> 
```

