

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΧΗΜΕΙΑΣ

Γ' ΚΑΤΕΥΘΥΝΣΗ: ΧΗΜΙΚΗ ΣΥΝΘΕΣΗ – ΒΙΟΧΗΜΕΙΑ &

ΒΙΟΕΦΑΡΜΟΓΕΣ



ΔΟΜΙΚΗ ΒΙΟΧΗΜΕΙΑ & ΣΤΟΙΧΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΦΟΡΑ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

BMP (Bone Morphogenetic Protein) Human

Εργαστήριο Βιοχημείας, Τμήμα Χημείας, Α.Π.Θ.

Βάσεις Δεδομένων Πρωτεϊνικών Οξέων και Πρωτεϊνών

Genbank, EMBL, Swiss-Prot, Protein Data Bank, PIR, PROSITE, TrEMBL, UniProt

SWISS-PROT
Protein Knowledgebase
TrEMBL
Uniprot Knowledgebase

Η SWISS-PROT έχει την έδρα της στο Ευρωπαϊκό Ινστιτούτο Βιοπληροφορικής (European Bioinformatics Institute), ιδρύθηκε το 1986 και είναι συνδεδεμένη με 90 άλλες τράπεζες.

PIR
Protein Information Resource

Η Protein Information Resource (PIR) βρίσκεται στο πανεπιστήμιο του Σακραμέντο και είναι τμήμα του Σπένσο Ιδρύματος Βιοτροπικής Έρευνας (NBRF) του Η.Π.Α. και παρέχονται μία σειρά από βάσεις δεδομένων που σχετίζονται με πρωτεϊνικές ακολουθίες και συνάδεται από συμπληρωματικές αρχές.

PROSITE
Database of protein families and domain

Η PROSITE είναι βάση ταξινόμησης σε οικογένειες πρωτεϊνικών ακολουθιών και αποτελείται παρόμοιες ακολουθίες (sequence domains).

Βάσεις δεδομένων δομικής βιοχημείας

Οι βάσεις δεδομένων που σχετίζονται με την τρισδιάστατη δομή είναι οι: Protein Data Bank (PDB), η CATH και η SCOP.

PDB
Protein Data Bank

Η Protein Data Bank (PDB) αποτελεί τη μοναδική βάση πληροφοριών όπου είναι καταγεγραμμένες οι τρισδιάστατες δομές βιολογικών μακρομορίων και ιδρύθηκε το 1971 στο Brookhaven National Laboratories (BNL). Οι πληροφορίες στην PDB εκτός από τις συντεταγμένες των ατόμων που αποτελούν τη δομή περιλαμβάνουν και πληροφορίες σχετικά με τη δομή όπως βιολογικές πληροφορίες, πληροφορίες για

ΒΑΣΙΛΙΚΗ Δ. ΝΙΚΟΛΟΠΟΥΛΟΥ

A.E.M. 8992

ΕΠΙΒΛΕΠΩΝ: ΚΑΘΗΓΗΤΡΙΑ ΧΟΛΗ-ΠΑΠΑΔΟΠΟΥΛΟΥ

ΘΕΟΔΩΡΑ

ΘΕΣΣΑΛΟΝΙΚΗ 2019

1. ΕΙΣΑΓΩΓΗ ΣΤΗ ΜΕΛΕΤΗ ΤΗΣ BMP

Σε αυτή την εργασία μελετήθηκε η BMP (Bone Morphogenetic Protein) Human, δηλαδή η οστική μορφογενετική πρωτεΐνη του ανθρώπου και οι διάφορες ακολουθίες νουκλεοτιδίων/αμινοξέων της. Ταυτόχρονα παρουσιάζονται με λεπτομέρεια τα βήματα που ακολουθήθηκαν.

Για την μελέτη αυτή χρησιμοποιήθηκαν τα παρακάτω:

- 1) Protein Data Bases: είναι οι βάσεις δεδομένων για πρωτεΐνες, οι οποίες συνίστανται από ένα σύνολο αρχείων με πληροφορίες για της πρωτεΐνες καθώς και από διάφορα εργαλεία που εξασφαλίζουν την πρόσβαση των ερευνητών σε πληροφορίες σχετικές με τις πρωτεΐνες.
- 2) Translate Tool: είναι το εργαλείο μετάφρασης, δηλαδή μεταφράζει την νουκλεοτιδική ακολουθία του DNA ή του RNA σε ακολουθία αμινοξέων (πρωτεΐνη).
- 3) Multalin Tool: είναι ένα εργαλείο στο οποίο μπορούμε να συγκρίνουμε τις αλληλουχίες και να βρούμε τα κοινά τμήματά τους.

Στη συνέχεια, πραγματοποιήθηκε η μελέτη της ακολουθίας των νουκλεοτιδίων της πρωτεΐνης και βρέθηκαν τα σημεία στα οποία μπορούν να δράσουν οι περιοριστικές ενδονουκλεάσες. Οι περιοριστικές ενδονουκλεάσες είναι ένζυμα, τα οποία απομονώνονται από τα βακτήρια και καταλύουν την υδρόλυση φωσφοδιεστερικών δεσμών σ' ένα νουκλεϊκό οξύ, σε συγκεκριμένες θέσεις.

Για τον σκοπό αυτό χρησιμοποιήθηκαν τα παρακάτω:

- 1) NEB Cutter V2.0: εργαλείο που εκτελεί την εύρεση των περιοριστικών ενδονουκλεασών.
- 2) Sequence extractor: εργαλείο που μας δίνει ένα «χάρτη» των περιοριστικών ενδονουκλεασών της ακολουθίας που μελετάμε.

Τέλος, μελετήθηκαν διάφορα χαρακτηριστικά γνωρίσματα που αφορούν μια πρωτεΐνη, όπως η κρυσταλλική απεικόνισή της, οι α-έλικες, β-ελάσματα που

περιέχει, καθώς και απεικονίσεις των σημείων όπου βρίσκονται κάποια συγκεκριμένα αμινοξέα.

Οι πληροφορίες αυτές αντλήθηκαν από τα παρακάτω προγράμματα:

- 1) Protein Workshop: είναι ένα πρόγραμμα που απεικονίζει σε 3D τις πρωτεΐνες, σε ένα υψηλής ποιότητας στυλ «κορδέλας». Ακόμη το Protein Workshop προσφέρει διάφορες επιλογές στις οποίες γίνεται να αλλάξει το στυλ απεικόνισης, τα χρώματα και η ορατότητα της πρωτεΐνης. Τέλος, μπορούν να προστεθούν ετικέτες σε άτομα και κατάλοιπα.
- 2) RasMol: είναι ένα πρόγραμμα ηλεκτρονικών υπολογιστών που χρησιμοποιείται κυρίως για την απεικόνιση και διερεύνηση βιολογικών μακρομοριακών δομών, όπως αυτές που υπάρχουν στην Protein Data Bank.

2. ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

2.1 Προσδιορισμός της πρωτεΐνης από την βάση δεδομένων

Αρχικά πραγματοποιήθηκε η αναζήτηση της πρωτεΐνης BMP στη βάση δεδομένων UniProt. Έπειτα, από την λίστα των οστικών μορφογενετικών πρωτεϊνών του ανθρώπου που εμφανίστηκε, επιλέχθηκε η BMP7 με κωδικό P18075. Από την ίδια ιστοσελίδα λήφθηκαν και οι παρακάτω πληροφορίες:

Protein | Bone morphogenetic protein 7
Gene | BMP7
Organism | *Homo sapiens (Human)*

Η λειτουργία της: προκαλεί τον σχηματισμό χόνδρων και οστών. Ακόμη, μπορεί να αποτελέσει οστεοεπαγωγικό παράγοντα υπεύθυνος για το φαινόμενο της επιθηλιακής οστεογένεσης. Τέλος, παίζει ρόλο στη ρύθμιση του ασβεστίου και στην ομοιόσταση των οστών.

2.2 Λήψη αλληλουχιών

Προχωρώντας προς τα κάτω στην ιστοσελίδα της UniProt, λήφθηκε η ακολουθία των αμινοξέων της πρωτεΐνης σε FASTA FORMAT (single-letter code, δηλαδή κάθε αμινοξύ αντιπροσωπεύεται από ένα γράμμα), η οποία είναι η εξής:

```
>sp|P18075|BMP7_HUMAN Bone morphogenetic protein 7 OS=Homo sapiens OX=9606 GN=BMP7 PE=1 SV=1
```

```
MHVRSLRAAAPHSFVALWAPLFLLRSALADFSLDNEVHSSFIHRRLRSQERREMQREILS  
ILGLPHRPRPHLQGKHNSAPMFMLDLNAMAVEEGGGPGGQGFSPYPKAVFSTQGPPLAS  
LQDSHFLTADVMVSFVNLVEHDKEFFHPRYHHREFRFDLSKIPEGEAVTAAEFRIYKDY  
IRERFDNETFRISVYQVLQEH LGRESDLFLLDSRTLWASEEGWL VFDITATSNHWV VNP  
HNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPFMVAF F KATEVHF RSIRSTGSKQRS  
QNRSKTPKNQEALRMANVAENSSSDQRQACKKHELYVSFRDLGWQDWIIAPEGYAAYYCE  
GECAPFLNSYMNATNHAIVQTLVHFINPETVPKPCCAPTQLNAISVLYFDDSSNVILKKY  
RNMVVVRACGCH
```

Ακόμη, βρέθηκαν και οι ακολουθίες από άλλες Data Bases:

GenBank ⁱ	AL157414 Genomic DNA No translation available.
DDBJ ⁱ	BC008584 mRNA Translation: AAH08584.1
CCDS ⁱ	CCDS13455.1

Ακολουθία νουκλεοτιδίων από την DDBJ:

1.

>ENA|BC008584|BC008584.1 Homo sapiens bone morphogenetic protein 7, mRNA (cDNA clone MGC:17201 IMAGE:4183402), complete cds.

```
GCCCCGAGCCCCGAGCCCCGGGTAGCGCGTAGAGCCGGCGCGATGCACGTGCGCTCACTGC
GAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCTGCTGCGCTCCG
CCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACCGGCGCCTCC
GCAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCCATTTTGGGCTTGCCCCACC
GCCCCGCGCCCGCACCTCCAGGGCAAGCACAACTCGGCACCCATGTTTCATGCTGGACCTGT
ACAACGCCATGGCGGTGGAGGAGGGCGGCGGGCCCCGGCGGCCAGGGCTTCTCCTACCCCT
ACAAGGCCGTCTTCAGTACCCAGGGCCCCCTCTGGCCAGCCTGCAAGATAGCCATTTCC
TCACCGACGCGGACATGGTCATGAGCTTCGTCAACCTCGTGGAACATGACAAGGAATTCT
TCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTTTCCAAGATCCCAGAAGGGG
AAGCTGTACGGCAGCCGAATTCCGGATCTACAAGGACTACATCCGGGAACGCTTCGACA
ATGAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAGCACTTGGGCAGGGAATCGG
ATCTCTTCTGCTCGACAGCCGTACCCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTGTTG
ACATCACAGCCACCAGCAACCACTGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGC
TCTCGGTGGAGACGCTGGATGGGCAGAGCATCAACCCCAAGTTGGCGGGCCTGATTGGGC
GGCACGGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTCTTCAAGGCCACGGAGGTCC
ACTTCCGCAGCATCCGGTCCACGGGGAGCAAAACAGCGCAGCCAGAACCGCTCCAAGACGC
CCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAGAACAGCAGCAGCGACCAGA
GGCAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCGAGACCTGGGCTGGCAGGACT
GGATCATCGCGCCTGAAGGCTACGCCGCTACTACTGTGAGGGGGAGTGTGCCTTCCCTC
TGAACCTCCTACATGAACGCCACCAACCACGCCATCGTGCAGACGCTGGTCCACTTCATCA
ACCCGGAACGGTGCCCAAGCCCTGCTGTGCGCCACGCAGCTCAATGCCATCTCCGTCC
TCTACTTCGATGACAGCTCCAACGTCATCCTGAAGAAATACAGAAACATGGTGGTCCGGG
CCTGTGGCTGCCACTAGCTCCTCCGAGAATTGAGACCCTTTGGGGCCAAGTTTTTCTGGA
TCCTCCATTGCTCGCCTTGGCCAGGAACCAGCAGACCAACTGCCTTTTGTGAGACCTTCC
CCTCCCTATCCCCAACTTTAAAGGTGTGAGAGTATTAGGAAACATGAGCAGCATATGGCT
TTTGATCAGTTTTTTCAGTGGCAGCATCCAATGAACAAGATCCTACAAGCTGTGCAGGCAA
AACCTAGCAGAAAAAAAAAACAACGCATAAAGAAAAAATGGCCGGGCCAGGTCATTGGCTG
GGAAGTCTCAGCCATGCACGGACTCGTTTCCAGAGGTAATTATGAGCGCCTACCAGCCAG
GCCACCCAGCCGTGGGAGGAAGGGGGCGTGGCAAGGGGTGGGCACATTGGTGTCTGTGCG
AAAGGAAAATTGACCCGGAAGTTCTGTAAATAAATGTCACAATAAAACGAATGAATGAAA
AAAAAAAAAAAAAA
```

2.

>ENA|AAH08584|AAH08584.1 Homo sapiens (human) bone morphogenetic protein 7

```
ATGCACGTGCGCTCACTGCGAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCC
CTGTTCTGCTGCGCTCCGCCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGC
```

TTCATCCACCGGCGCCTCCGCAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCC
ATTTTGGGCTTGCCCCACCGCCCGCGCCCGCACCTCCAGGGCAAGCACAACTCGGCACCC
ATGTTTCATGCTGGACCTGTACAACGCCATGGCGGTGGAGGAGGGCGGCGGGCCCGGCGGC
CAGGGCTTCTCCTACCCCTACAAGGCCGTCTTCAGTACCCAGGGCCCCCTCTGGCCAGC
CTGCAAGATAGCCATTTCTCTACCGACGCCGACATGGTCATGAGCTTCGTCAACCTCGTG
GAACATGACAAGGAATTCTTCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTT
TCCAAGATCCCAGAAGGGGAAGCTGTCACGGCAGCCGAATTCCGGATCTACAAGGACTAC
ATCCGGGAACGCTTCGACAATGAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAG
CACTTGGGCAGGGAATCGGATCTCTTCTGCTCGACAGCCGTACCCCTCTGGGCCTCGGAG
GAGGGCTGGCTGGTGTGTTTGACATCACAGCCACCAGCAACCACTGGGTGGTCAATCCGCGG
CACAACCTGGGCCTGCAGCTCTCGGTGGAGACGCTGGATGGGCAGAGCATCAACCCCAAG
TTGGCGGGCCTGATTGGGCGGCACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTC
TTCAAGGCCACGGAGGTCCACTTCCGCAGCATCCGGTCCACGGGGAGCAAACAGCGCAGC
CAGAACCGCTCCAAGACGCCCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAG
AACAGCAGCAGCGACCAGAGGCAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCGA
GACCTGGGCTGGCAGGACTGGATCATCGCGCCTGAAGGCTACGCCGCCTACTACTGTGAG
GGGGAGTGTGCCTTCCCTCTGAACTCCTACATGAACGCCACCAACCACGCCATCGTGCAG
ACGCTGGTCCACTTCATCAACCCGAAACGGTGCCCAAGCCCTGCTGTGCGCCACGCAG
CTCAATGCCATCTCCGTCTCTACTTCGATGACAGCTCCAACGTATCCTGAAGAAATAC
AGAAACATGGTGGTCCGGGCCTGTGGCTGCCACTAG

Ακολουθία νουκλεοτιδίων από την CCDS:

Η πλήρης Νουκλεοτιδική Αλληλουχία:

Nucleotide Sequence (1296 nt):

ATGCACGTGCGCTCACTGCGAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCTCTGC
TGCGCTCCGCCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACCGGCCTCCG
CAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCCATTTTGGGCTTGCCCCACCGCCCGCGCCCG
CACCTCCAGGGCAAGCACAACTCGGCACCCATGTTTCATGCTGGACCTGTACAACGCCATGGCGGTGGAGG
AGGGCGGCGGGCCCGGCGGCCAGGGCTTCTCCTACCCCTACAAGGCCGTCTTCAGTACCCAGGGCCCCC
TCTGGCCAGCCTGCAAGATAGCCATTTCTCTACCGACGCCGACATGGTCATGAGCTTCGTCAACCTCGTG
GAACATGACAAGGAATTCTTCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTTCCAAGATCC
CAGAAGGGGAAGCTGTCACGGCAGCCGAATTCCGGATCTACAAGGACTACATCCGGGAACGCTTCGACAA
TGAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAGCACTTG66CAGGGAATCGGATCTCTTCTG
CTCGACAGCCGTACCCCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTGTTGACATCACAGCCACCAGCAACC
ACTGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGCTCTCGGTGGAGACGCTGGATGGGCAGAGCAT
CAACCCCAAGTTGGCGGGCCTGATTGGGCGGCACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTC
TTCAAGGCCACGGAGGTCCACTTCCGCAGCATCCGGTCCACGGGGAGCAAACAGCGCAGCCAGAACCGCT
CCAAGACGCCCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAGAACAGCAGCAGCGACCAGAG
GCAGGCCTGTAAGAAGCACGAGCTGTATGTCAGCTTCCGAGACCTGGGCTGGCAGGACTGGATCATCGCG
CCTGAAGGCTACGCCGCCTACTACTGTGAGGGGGAGTGTGCCTTCCCTCTGAACTCCTACATGAACGCCA
CCAACCACGCCATCGTGCAGACGCTG6TCCACTTCATCAACCCGAAACGGTGCCCAAGCCCTGCTGTGC
GCCACGCAGCTCAATGCCATCTCCGTCTCTACTTCGATGACAGCTCCAACGTATCCTGAAGAAATAC
AGAAACATGGTGGTCCGGGCCTGTGGCTGCCACTAG

Μεταφρασμένη Αλληλουχία:

Translation (431 aa):

MHVRSLRAAAPHSFVALWAPLFLRLRSALADFSLDNEVHSSFIHRRRLRSQERREMQRREILSILGLPHRPRP
HLQKGKHSAPMFMLDLYNAMAVEEGGGPGGGFSYPYKAVFSTQGPPLASLQDSHFLTDADMVMSFVNLV
EHDKEFFHPRYHHREFRFDLSKIPEGEAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFL
LDSRTLWASEEGWLVDITATSNHWWNPRHNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPFMVA
FKATEVHFRSIRSTGSKQRSQNRSKTPKNQEALRMANVAENSSSDQRQACKKHELYVSFRDLGWQDWIIA
PEGYAAYYCEGECAPFLNSYMNATNHAIVQTLVHFINPETVVPKCCAPTQLNAISVLYFDDSSNVILKKY
RNMVVRACGCH

2.3 Μετάφραση των νουκλεοτιδίων

Η μετάφραση της πλήρους νουκλεοτιδικής ακολουθίας από την CCDS, της BMP7, έγινε με το TRANSLATE TOOL το οποίο βρίσκουμε στην ιστοσελίδα www.expasy.org. Αρχικά πραγματοποιήθηκε η διαδικασία της μετάφρασης για όλη την νουκλεοτιδική αλληλουχία. Μόλις γίνει η μετάφραση της νουκλεοτιδικής ακολουθίας το σύστημα μας εμφανίζει τα διάφορα πιθανά αναγνωστικά πλαίσια.

Το πρώτο αναγνωστικό πλαίσιο όπου το γονίδιο κωδικοποιείται από την αρχή έως το τέλος του χωρίς να διακόπτεται η ακολουθία του είναι το εξής:

5' Frame 1

Met HVRSLRAAAPHSFVALWAPLFLRLRSALADFSLDNEVHSSFIHRRRLRSQERRE Met QREILSILGLPHRPRPHLQKGKHSAP Met F Met
LDLYNA Met AVEEGGGPGGGFSYPYKAVFSTQGPPLASLQDSHFLTDADMVMSFVNLVEHDKEFFHPRYHHREFRFDLSKIPEGE
EAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFLDSRTLWASEEGWLVDITATSNHWWNPRHNLGLQLSVETLD
GQSINPKLAGLIGRHGPQNKQPF Met VAFKATEVHFRSIRSTGSKQRSQNRSKTPKNQEALR Met ANVAENSSSDQRQACKKHELYV
SFRDLGWQDWIIAPEGYAAYYCEGECAPFLNSY Met NATNHAIVQTLVHFINPETVVPKCCAPTQLNAISVLYFDDSSNVILKKYRN Met
VVRACGCH Stop

Έπειτα, μεταφράστηκε η ακολουθία από τα πρώτα χρωματισμένα με μπλε νουκλεοτίδια (εξώνια) μέχρι το τέλος και προέκυψε μια εντελώς διαφορετική ακολουθία αμινοξέων:

5' Frame 1

WN Met TRNSSTHATTIESSGLIFRSQKGLSRQPNSSGSTRITSGNAST Met RRSQSAFIRCSRSTWAGNRSSCSTAVPSGPRRRAG
WCLTSQPPATTGWSIRGTTWACSSRWRRW Met GRASTPSWRA Stop LGGTGPRSTSSPSWWLSSRPRRSTSAASGPRGANSAAARTAP
RRPRTKPCGWPTWQRTAAATGRPVIRSTSC Met SASETWAGRTGSSRLKATPPTTVRGSVPSL Stop TPT Stop TPPTTPSCRRWSTS
STRKRCPSPAVRPRSS Met PSPSSTS Met TAPTSS Stop RNTETWWSGPVAAT

Στη συνέχεια, μεταφράστηκαν οι μπλε ακολουθίες συραμμένες η μία δίπλα στην άλλη:

```
5'3' Frame 1
WN Met TRNSSTHATTIESSGLIFPRSQKGLSRQPNSSGSTRITTSGNAST Met RRRSGSAFIRCSRSTWAGQSINPKLAGLIGRHGPQNK
QPF Met VAFFKATEVHFERSIRSTGSKQRSQNRSKTPKNQEALR Met ANVAGLDHRA Stop RLRRLLL Stop GGVCLPSELLHERHQPRHRA
DA
```

Τέλος, μεταφράστηκε η ακολουθία με μερικά νουκλεοτίδια πριν το πρώτο μπλε νουκλεοτίδιο μέχρι το τέλος. Και σε αυτή την περίπτωση το γονίδιο δεν διακόπτεται:

```
5'3' Frame 1
Met SFVNLVEHDKFEFFHPRYHHREFRFDLSKIPGEAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFLLDSRTLWASE
EGWLVEFDITATSNHWVYNPRHNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPF Met VAFFKATEVHFERSIRSTGSKQRSQNRSK
TPKNQEALR Met ANVAENSSSDQRQACKKHELYVSFRDLGWQDWIIAPEGYAAYYCEGECAPLNSY Met NATNHAIVQTLVHFINPE
TVPKPCCAPTQLNAISVLVFDSSNVILKKYRN Met VVRAGGCH Stop
```

2.4 Σύγκριση ομοιότητας ακολουθίας αμινοξέων

Η διαδικασία της ομοιοπαράθεσης πραγματοποιήθηκε με την χρήση του εργαλείου Multalin. Πιο συγκεκριμένα, συγκρίθηκαν οι παρακάτω ακολουθίες των αμινοξέων:

- 1) Η ακολουθία FASTA FORMAT, από την ιστοσελίδα UniProt (ώριμη μορφή γονιδίου της BMP7 Human).
- 2) Η μεταφρασμένη ακολουθία από την CCDS
- 3) Η ακολουθία από τη μετάφραση της πλήρους νουκλεοτιδικής αλληλουχίας (από το 2.3).

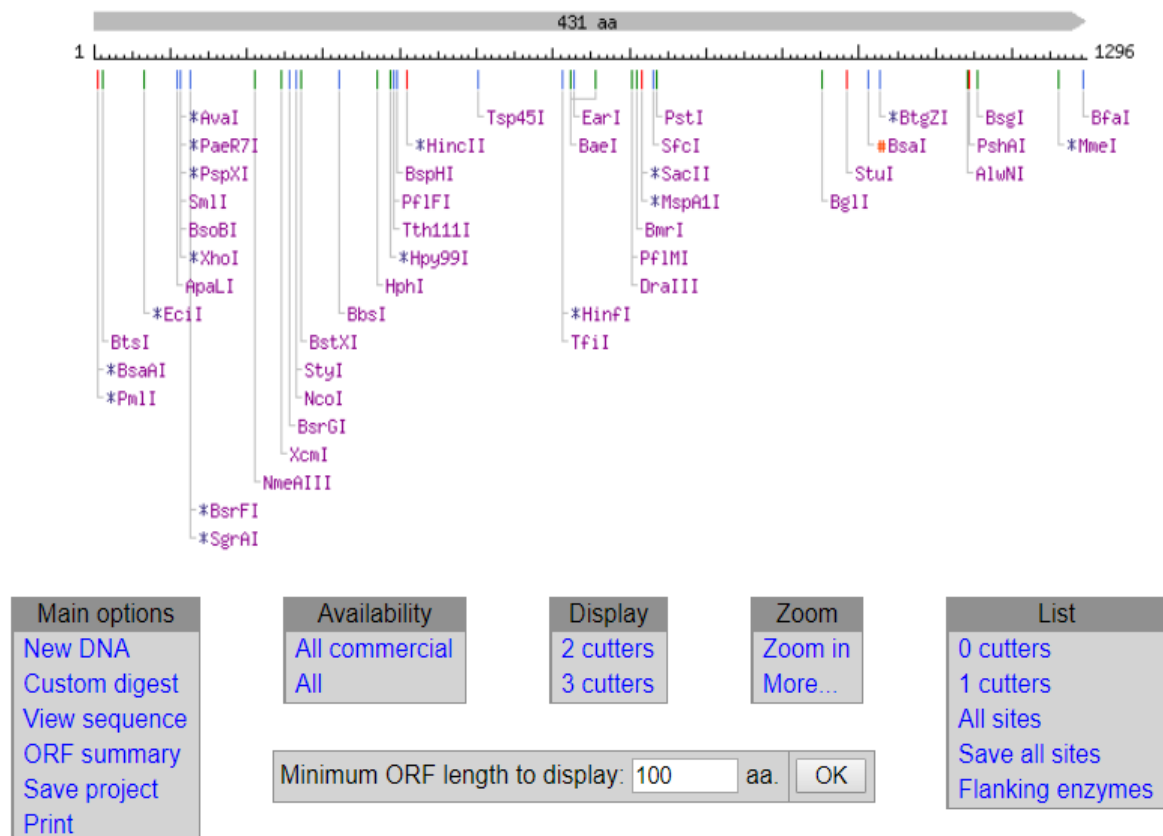
Η σύγκριση τους φαίνεται στην επόμενη σελίδα:

	10	20	30	40	50	60
1xxxxx0	MHVRS	LRAAAPHSFVALWAPLFL	LLRSALADFS	LDNEVHSSFIHRLRSQERREMQREILS		
2xxxxx1	MHVRS	LRAAAPHSFVALWAPLFL	LLRSALADFS	LDNEVHSSFIHRLRSQERREMQREILS		
3xxxxx2	MHVRS	LRAAAPHSFVALWAPLFL	LLRSALADFS	LDNEVHSSFIHRLRSQERREMQREILS		
Consensus	MHVRS	LRAAAPHSFVALWAPLFL	LLRSALADFS	LDNEVHSSFIHRLRSQERREMQREILS		
Prim.cons.	MHVRS	LRAAAPHSFVALWAPLFL	LLRSALADFS	LDNEVHSSFIHRLRSQERREMQREILS		
	70	80	90	100	110	120
1xxxxx0	ILGLPHRPRPHLQGKHNSAPMF	MLDLYNAMAVEEGGPGGQGF	SYPYKAVFSTQGPPLAS			
2xxxxx1	ILGLPHRPRPHLQGKHNSAPMF	MLDLYNAMAVEEGGPGGQGF	SYPYKAVFSTQGPPLAS			
3xxxxx2	ILGLPHRPRPHLQGKHNSAPMF	MLDLYNAMAVEEGGPGGQGF	SYPYKAVFSTQGPPLAS			
Consensus	ILGLPHRPRPHLQGKHNSAPMF	MLDLYNAMAVEEGGPGGQGF	SYPYKAVFSTQGPPLAS			
Prim.cons.	ILGLPHRPRPHLQGKHNSAPMF	MLDLYNAMAVEEGGPGGQGF	SYPYKAVFSTQGPPLAS			
	130	140	150	160	170	180
1xxxxx0	LQDSHFLTADMMVMSFVNL	VEHDKEFFHPRYHHREFR	FDLSKIPEGEAVTAAEFRIYKDY			
2xxxxx1	LQDSHFLTADMMVMSFVNL	VEHDKEFFHPRYHHREFR	FDLSKIPEGEAVTAAEFRIYKDY			
3xxxxx2	LQDSHFLTADMMVMSFVNL	VEHDKEFFHPRYHHREFR	FDLSKIPEGEAVTAAEFRIYKDY			
Consensus	LQDSHFLTADMMVMSFVNL	VEHDKEFFHPRYHHREFR	FDLSKIPEGEAVTAAEFRIYKDY			
Prim.cons.	LQDSHFLTADMMVMSFVNL	VEHDKEFFHPRYHHREFR	FDLSKIPEGEAVTAAEFRIYKDY			
	190	200	210	220	230	240
1xxxxx0	IRERFDNETFRISVYQVLQ	EHLGRESDFLLDSRTLWASEEGWLV	FDITATSNHMMVNPR			
2xxxxx1	IRERFDNETFRISVYQVLQ	EHLGRESDFLLDSRTLWASEEGWLV	FDITATSNHMMVNPR			
3xxxxx2	IRERFDNETFRISVYQVLQ	EHLGRESDFLLDSRTLWASEEGWLV	FDITATSNHMMVNPR			
Consensus	IRERFDNETFRISVYQVLQ	EHLGRESDFLLDSRTLWASEEGWLV	FDITATSNHMMVNPR			
Prim.cons.	IRERFDNETFRISVYQVLQ	EHLGRESDFLLDSRTLWASEEGWLV	FDITATSNHMMVNPR			
	250	260	270	280	290	300
1xxxxx0	HNLGLQLSVETLDGQSNPK	LAGLIGRHGPQNKQPFMVAFF	KATEVHFRSIRSTGSKQRS			
2xxxxx1	HNLGLQLSVETLDGQSNPK	LAGLIGRHGPQNKQPFMVAFF	KATEVHFRSIRSTGSKQRS			
3xxxxx2	HNLGLQLSVETLDGQSNPK	LAGLIGRHGPQNKQPFMVAFF	KATEVHFRSIRSTGSKQRS			
Consensus	HNLGLQLSVETLDGQSNPK	LAGLIGRHGPQNKQPFMVAFF	KATEVHFRSIRSTGSKQRS			
Prim.cons.	HNLGLQLSVETLDGQSNPK	LAGLIGRHGPQNKQPFMVAFF	KATEVHFRSIRSTGSKQRS			
	310	320	330	340	350	360
1xxxxx0	QNRSKTPKNQEALRMANVA	ENSSDQRQACKKHLYVS	FRDLGWQDIIAPEGYAAYYCE			
2xxxxx1	QNRSKTPKNQEALRMANVA	ENSSDQRQACKKHLYVS	FRDLGWQDIIAPEGYAAYYCE			
3xxxxx2	QNRSKTPKNQEALRMANVA	ENSSDQRQACKKHLYVS	FRDLGWQDIIAPEGYAAYYCE			
Consensus	QNRSKTPKNQEALRMANVA	ENSSDQRQACKKHLYVS	FRDLGWQDIIAPEGYAAYYCE			
Prim.cons.	QNRSKTPKNQEALRMANVA	ENSSDQRQACKKHLYVS	FRDLGWQDIIAPEGYAAYYCE			
	370	380	390	400	410	420
1xxxxx0	GECAPLNSYNNATNHAIV	QTLVHFINP	ETVPKPCCAPTQLNAISVLY	FDSSNVILKKY		
2xxxxx1	GECAPLNSYNNATNHAIV	QTLVHFINP	ETVPKPCCAPTQLNAISVLY	FDSSNVILKKY		
3xxxxx2	GECAPLNSYNNATNHAIV	QTLVHFINP	ETVPKPCCAPTQLNAISVLY	FDSSNVILKKY		
Consensus	GECAPLNSYNNATNHAIV	QTLVHFINP	ETVPKPCCAPTQLNAISVLY	FDSSNVILKKY		
Prim.cons.	GECAPLNSYNNATNHAIV	QTLVHFINP	ETVPKPCCAPTQLNAISVLY	FDSSNVILKKY		
	430					
1xxxxx0	RNMVVRACGCH					
2xxxxx1	RNMVVRACGCH					
3xxxxx2	RNMVVRACGCH					
Consensus	RNMVVRACGCH					
Prim.cons.	RNMVVRACGCH					

Καθώς οι ακολουθίες που είναι χρωματισμένες με κόκκινο είναι όμοιες, οδηγούμαστε στο συμπέρασμα ότι και οι τρεις ακολουθίες ταυτίζονται μεταξύ τους. Αν και ήταν αναμενόμενο για την 1 και 2 ακολουθίες να ταυτίζονται καθώς προέρχονται από την βιβλιογραφία, ωστόσο και η 3 φαίνεται πως αντιπροσωπεύει το γονίδιο της BMP7.

2.5 Σημεία στα οποία δρουν οι περιοριστικές ενδονουκλεάσες

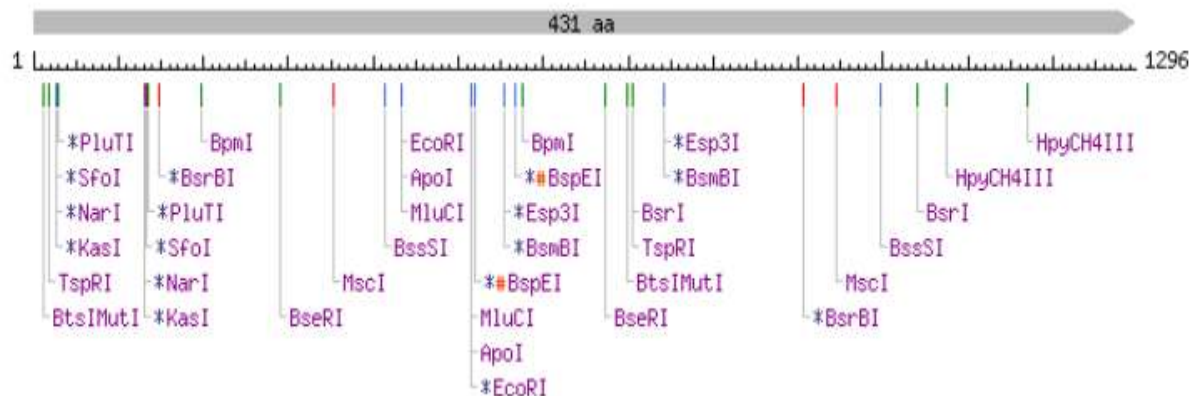
Σε αυτό το σημείο πραγματοποιήθηκε η διαδικασία πρόβλεψης των σημείων της ακολουθίας των νουκλεοτιδίων, τα οποία μπορούν να κοπούν από τις περιοριστικές ενδονουκλεάσες δύο και τρεις φορές. Αρχικά η ακολουθία της BMP7 από την CCDS βιβλιοθήκη, τοποθετήθηκε στην ιστοσελίδα του NEB Cutter Tool, επιλέχθηκαν τα πεδία Circular και NEB Enzymes και πατήθηκε το κουμπί Submit. Έτσι εμφανίστηκε η παρακάτω εικόνα:



Σε αυτή την εικόνα εμφανίζονται όλες οι περιοριστικές ενδονουκλεάσες. Στο γκρι κουτάκι «Display» επιλέχθηκε πρώτα το «2 cutters» και μας βγήκε μια λίστα με 19 ένζυμα (περιοριστικές ενδονουκλεάσες), τα οποία κόβουν την ακολουθία των νουκλεοτιδίων της BMP7 δύο φορές. Έπειτα, επιλέχθηκε το «3 cutters» και μας βγήκε μια λίστα με 13 ένζυμα (περιοριστικές ενδονουκλεάσες), τα οποία κόβουν την ακολουθία τρεις φορές. Τέλος, ως παράδειγμα, με την χρήση του Sequence Extractor tool βρέθηκαν αναλυτικά τα σημεία κοπής για τρία από τα ένζυμα που κόβουν δύο φορές την ακολουθία των νουκλεοτιδίων.

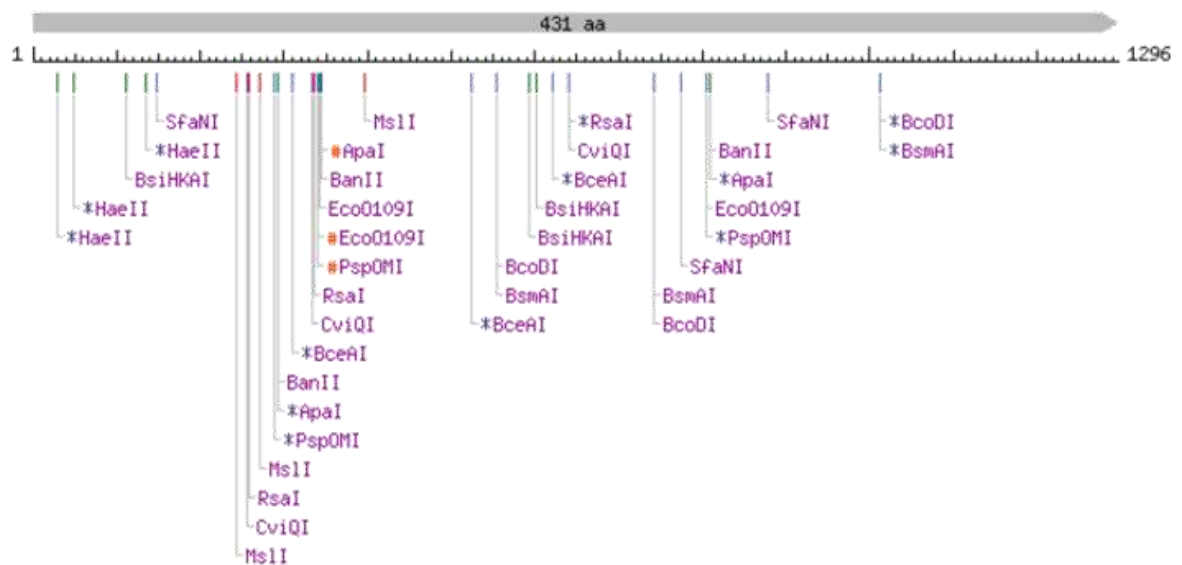
Περιοριστικές ενδονουκλεάσες που κόβουν 2 φορές την ακολουθία και τα σημεία κοπής τους:

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	ApoI	R [▼] AATT [▲] Y	list	433/437, 517/521
2	BpmI	CTGGAG(N) ₁₄ [▲] NN [▼]	list	199/197, 577/575
3	BseRI	GAGGAG(N) ₈ [▲] NN [▼]	list	292/290, 673/671
4	BsmBI	CGTCTCN [▼] NNNN [▲]	list	*556/560, *742/746
5	BspEI	T [▼] CCGG [▲] A	list	*#521/525, *#569/573
6	BsrBI	CCG [▼] CTC	list	*150, *908
7	BsrI	ACTG [▲] GN [▼]	list	706/704, 1042/1040
8	BssSI	C [▼] ACGA [▲] G	list	415/419, 997/1001
9	BtsIMutI	CAGTG [▲] NN [▼]	list	13/11, 699/697
10	EcoRI	G [▼] AATT [▲] C	list	433/437, *517/521
11	Esp3I	CGTCTCN [▼] NNNN [▲]	list	*556/560, *742/746
12	HpyCH4III	AC [▲] N [▼] GT	list	1075/1074, 1170/1169
13	KasI	G [▼] GCGC [▲] C	list	*27/31, *131/135
14	MluCI	[▼] AATT [▲]	list	433/437, 517/521
15	MscI	TGG [▼] CCA	list	355, 946
16	NarI	GG [▼] CG [▲] CC	list	*28/30, *132/134
17	PluTI	G [▼] GCGC [▼] C	list	*31/27, *135/131
18	SfoI	GGC [▼] GCC	list	*29, *133
19	TspRI	[▲] NNCASTGNN [▼]	list	20/11, 706/697



Περιοριστικές ενδονουκλεάσες που κόβουν 3 φορές την ακολουθία και τα σημεία κοπής τους:

#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	ApaI	G _↓ GGCC ⁺ C	list	*293/289, #346/342, *809/805
2	BanII	G _↓ RGY ⁺ C	list	293/289, 346/342, 809/805
3	BceAI	ACGGC(N) ₁₂ ⁺ NN _↓	list	*310/312, *524/526, *623/625
4	BcoDI	GTCTCN ⁺ NNNN _↓	list	556/560, 742/746, *1013/1017
5	BsiHKAI	G _↓ WGCW ⁺ C	list	113/109, 593/589, 602/598
6	BsmAI	GTCTCN ⁺ NNNN _↓	list	556/560, 742/746, *1013/1017
7	CviQI	G ⁺ TA _↓ C	list	258/260, 335/337, 641/643
8	EcoO109I	RG ⁺ GNC _↓ CY	list	#342/345, 343/346, 806/809
9	HaeII	R _↓ GCGC ⁺ Y	list	*31/27, *49/45, *135/131
10	MslI	CAYNN ⁺ NNRTG	list	244, 271, 397
11	PspOMI	G ⁺ GGCC _↓ C	list	*289/293, #342/346, *805/809
12	RsaI	GT ⁺ _↓ AC	list	259, 336, *642
13	SfaNI	GCATC(N) ₅ ⁺ NNNN _↓	list	149/153, 776/780, 878/882



T I A P E G Y A A Y Y C E G E C A F P L N S Y M N A T
GATCATCGCGCTGAAGGCTACGCCCTACTACTGTGAGGGGAGTGTGCCTTCCCTCTGAACCTACATGAACGCCA
 ^1050 ^1060 ^1070 ^1080 ^1090 ^1100 ^1110
CTAGTAGCGCGGACTTCCGATGCGGCGGATGATGACACTCCCCCTCACACGGAAGGGAGACTTGAGGATGTACTTGCGGT

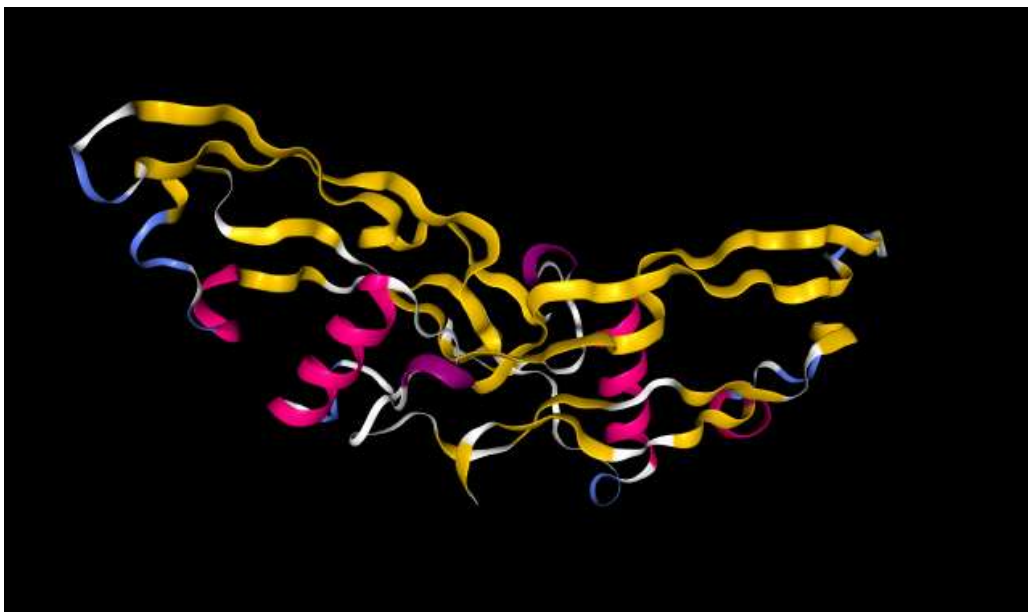
2.6 Χαρακτηριστικά γνωρίσματα της πρωτεΐνης

Για την εύρεση των διαφόρων χαρακτηριστικών γνωρισμάτων της BMP7 πραγματοποιήθηκε πλοήγηση στην ιστοσελίδα RCSB Protein Data Bank. Από εκεί το .pdb file αποθηκεύτηκε στον υπολογιστή και ανοίχθηκε με το Notepad και αντλήθηκαν οι παρακάτω πληροφορίες:

1. Η ανάλυση της, από την συγκεκριμένη πηγή που χρησιμοποιείται είναι 2.00 Angstroms.
2. Η πρωτεΐνη έχει μια και μοναδική αλυσίδα που αποτελείται από 139 κατάλοιπα (residues) .
3. Η πρωτεΐνη διαθέτει 3 α-έλικες (ποσοστό 13%) και 9 β-ελάσματα (ποσοστό 33%).

Τέλος ανοίγοντας το αρχείο με το Java Protein Workshop βρέθηκαν οι εξής πληροφορίες:

1. Η κρυσταλλική απεικόνιση της πρωτεΐνης:



2. Η πρωτεϊνική αλυσίδα μπορεί να περιέχει ετεροάτομο N-ακετυλο-D-γλυκοζαμίνης.
3. Η πρωτεΐνη έχει 71 μόρια νερού.