

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, Ανθρώπινη οστική μορφογενετική πρωτεΐνη)



Κούρτη Ελένη

A.E.M. : 8861

Επιβλέπων Καθηγητής: Χολή-Παπαδοπούλου Θεοδώρα

Θεσσαλονίκη, Ιούνιος 2019

1 ΜΕΛΕΤΗ ΤΗΣ BMP

Στην παρούσα αναφορά αποτελεσμάτων παρουσιάζονται λεπτομερώς τα βήματα που έγιναν στον υπολογιστή για την μελέτη της BMP ανθρώπινης πρωτεΐνης, δηλαδή της οστικής μορφογενετικής πρωτεΐνης. Τα «εργαλεία» που χρησιμοποιήθηκαν για τον σκοπό αυτό είναι τα εξής:

- i. Protein Data Basis: Βάσεις Δεδομένων Πρωτεϊνών, πρόκειται για το σύνολο των αρχείων πληροφοριών καθώς και τα απαραίτητα εργαλεία που εξασφαλίζουν την πρόσβαση των ερευνητών σε πληροφορίες σχετικές με τις πρωτεΐνες.
- ii. Translate Tool: Εργαλείο Μετάφρασης, πρόκειται για εργαλείο που επιτρέπει την μετάφραση νουκλεοτιδικής αλληλουχίας (είτε DNA είτε RNA) σε πρωτεΐνη^[1].
- iii. Multalin Tool: Χρησιμοποιήθηκε για την ομοιοπαράθεση αλληλουχιών, δηλαδή την σύγκρισή τους και την εύρεση κοινών τμημάτων.

Στο δεύτερο μέρος της εργασίας μελετήθηκαν τα προβλεπόμενα σημεία στην αλληλουχία της πρωτεΐνης, στα οποία είναι δυνατόν να δράσουν περιοριστικά ένζυμα. Για αυτή την διαδικασία χρησιμοποιήθηκαν:

- iv. NEB Cutter V2.0 και
- v. Sequence extractor

Οι πρακτικές λειτουργίες των εργαλείων αυτών αναλύονται στην αντίστοιχη ενότητα.

Η παρούσα εργασία, επίσης, περιέχει περιγραφή των χαρακτηριστικών ιδιοτήτων της αλληλουχίας της πρωτεΐνης. Οι πληροφορίες αυτές αντλούνται με τη βοήθεια προγραμμάτων όπως α. protein workshop και β. RasMol

2 ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ

2.1 Προσδιορισμός της πρωτεΐνης ενδιαφέροντος από την βάση δεδομένων

Πρώτο βήμα ήταν η αναζήτηση της πρωτεΐνης ενδιαφέροντος στη βάση δεδομένων UniProt . Επιλέχθηκε τυχαία μια από τις προσφερόμενες επιλογές που προκύπτουν από την αναζήτηση για την ανθρώπινη οστική μορφογενετική πρωτεΐνη. Αυτή η αναφορά αποτελεσμάτων, λοιπόν, αφορά την BMP7 του ανθρώπου με κωδικό P18075. Από την ίδια ιστοσελίδα λαμβάνονται οι παρακάτω πληροφορίες:

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

Πρόκειται για πρωτεΐνη που σχετίζεται με τον σχηματισμό χόνδρων και οστών και παίζει σημαντικό ρόλο στη ρύθμιση του ασβεστίου στον οργανισμό και την ομοιοστάση των οστών, ενώ παράλληλα θεωρείται επαγωγικός παράγοντας υπεύθυνος για το φαινόμενο της επιθηλιακής οστεογένεσης^[2].

2.2 Λήψη των αλληλουχιών

Συνεχίζοντας την πλοήγηση στην ιστοσελίδα της UniProt, λήφθηκε η αμινοξική ακολουθία της πρωτεΐνης σε FASTA FORMAT ^[2], η οποία παρουσιάζεται παρακάτω:

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

```
MHVRSLRAAPHSFVALWAPLFLRLSALADFSLDNEVHSSFIHRRRLRSQERREMQREILS  
ILGLPHRPRPHLQGKHNSAPMFMLDLYNAMAVEEGGGPGGQGFSPYPKAVFSTQGPPLAS  
LQDSHFLTDADMVMSFVNLVEHDKEFFHPRYHHREFRFDLSKIPEGEAVTAAEFRIYKDY  
IRERFDNETFRISVYQVLQEHLGRESDLFLDLSRTLWASEEGWLVFDITATSNHWVNP  
HNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPFMVAFKATEVHFIRSIRSTGSKQRS  
QNRSKTPKNQEALRMANVAENSSDQRQACKKHELYVSFRDLGWQDWIIAPEGYAAYYCE  
GECAPFLNSYMNATNHAIVQTLVHFINPETVVPKPCCAPTQLNAISVLYFDDSSNVILKKY  
RNMVVRACGCH
```

Το FASTA FORMAT είναι single-letter code, δηλαδή κάθε αμινοξύ αντιπροσωπεύεται από ένα γράμμα.

Ελέγχθηκε η διαθεσιμότητα αλληλουχιών από άλλες Data Basis:

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

Παρακάτω παρουσιάζονται οι αλληλουχίες νουκλεοτιδίων από την DDBJ ^[3] :

1.

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

```
GCCCCGAGCCCCGAGCCCCGGGTAGCGCGTAGAGCCGGCGCGATGCACGTGCGCTCACTGC
GAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCTTGCTGCGCTCCG
CCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACCGGCGCCTCC
GCAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCCATTTTGGGCTTGCCCCACC
GCCCCGCGCCCGCACCTCCAGGGCAAGCACAACTCGGCACCCATGTTTCATGCTGGACCTGT
ACAACGCCATGGCGGTGGAGGAGGGCGGGCGGGCCGCGGCCAGGGCTTCTCCTACCCCT
ACAAGGCCGTCTTCAGTACCCAGGGCCCCCCTCTGGCCAGCCTGCAAGATAGCCATTTCC
TCACCGACGCCGACATGGTCATGAGCTTCGTCAACCTCGTGGAACATGACAAGGAATTC
TCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTTTCCAAGATCCCAGAAGGGG
AAGCTGTACGGCAGCCGAATTCCGGATCTACAAGGACTACATCCGGGAACGCTTCGACA
ATGAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAGCACTTGGGCAGGGAATCGG
ATCTCTTCTGCTCGACAGCCGTACCCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTGTTG
ACATCAGCCACCAGCAACCACTGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGC
TCTCGGTGGAGACGCTGGATGGGCAGAGCATCAACCCCAAGTTGGCGGGCCTGATTGGGC
GGCACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTCTTCAAGGCCACGGAGGTCC
ACTTCCGCAGCATCCGGTCCACGGGGAGCAAACAGCGCAGCCAGAACCGCTCCAAGACGC
CCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAGAACAGCAGCAGCGACCAGA
GGCAGGCCTGTAAGAAGCACGAGCTGTATGTGAGCTTCCGAGACCTGGGCTGGCAGGACT
GGATCATCGCGCCTGAAGGCTACGCCGCCTACTACTGTGAGGGGGAGTGTGCCTTCCCTC
TGAACTCCTACATGAACGCCACCAACCACGCCATCGTGACAGCGCTGGTCCACTTCATCA
ACCCGGAACGGTGCCCAAGCCCTGCTGTGCGCCACGCAGCTCAATGCCATCTCCGTCC
TCTACTTCGATGACAGCTCCAACGTATCCTGAAGAAATACAGAAACATGGTGGTCCGGG
CCTGTGGCTGCCACTAGCTCCTCCGAGAATTACAGACCCTTTGGGGCCAAGTTTTCTGGA
TCCTCCATTGCTTCGCTTGGCCAGGAACAGCAGACCAACTGCCTTTTGTGAGACCTTCC
CCTCCCTATCCCCAACTTTAAAGGTGTGAGAGTATTAGGAAACATGAGCAGCATATGGCT
TTTGATCAGTTTTTTCAGTGGCAGCATCCAATGAACAAGATCCTACAAGCTGTGCAGGCAA
AACCTAGCAGAAAAAAAACAACGCATAAAGAAAAAATGGCCGGGGCCAGGTCATTGGCTG
GGAAGTCTCAGCCATGCACGGACTCGTTTCCAGAGGTAATTATGAGCGCCTACCAGCCAG
GCCACCCAGCCGTGGGAGGAAGGGGGCGTGGCAAGGGGTGGGCACATTGGTGTCTGTGCG
AAAGGAAAATTGACCCGGAAGTTCCTGTAATAAATGTCACAATAAAACGAATGAATGAAA
AAAAAAAAAAAAA
```

2.

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

```
ATGCACGTGCGCTCACTGCGAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCC
CTGTTCTTGCTGCGCTCCGCCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGC
TTCATCCACCGGCGCCTCCGCAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCCTCTCC
ATTTTGGGCTTGCCCCACGCGCGCCGCACCTCCAGGGCAAGCACAACTCGGCACCC
```

ATGTTTCATGCTGGACCTGTACAACGCCATGGCGGTGGAGGAGGGCGGCGGGCCCCGGCGGC
CAGGGCTTCTCCTACCCCTACAAGGCCGTCTTCAGTACCCAGGGCCCCCTCTGGCCAGC
CTGCAAGATAGCCATTTCCCTCACCGACGCCGACATGGTCATGAGCTTCGTCAACCTCGTG
GAACATGACAAGGAATTCTTCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTT
TCCAAGATCCCAGAAGGGGAAGCTGTACGGCAGCCGAATTCCGGATCTACAAGGACTAC
ATCCGGGAACGCTTCGACAATGAGACGTTCGGATCAGCGTTTATCAGGTGCTCCAGGAG
CACTTGGGCAGGGAATCGGATCTCTTCTGCTCGACAGCCGTACCCTCTGGGCCTCGGAG
GAGGGCTGGCTGGTGTGTTTGACATCACAGCCACCAGCAACCACTGGGTGGTCAATCCGCGG
CACAACTTGGGCCTGCAGCTCTCGGTGGAGACGCTGGATGGGCAGAGCATCAACCCCAAG
TTGGCGGGCCTGATTGGGCGGCACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTC
TTCAAGGCCACGGAGGTCCACTTCCGACGATCCGGTCCACGGGGAGCAAACAGCGCAGC
CAGAACCCTGCCAAGACGCCCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAG
AACAGCAGCAGCAGACCAGAGGCAGGCCTGTAAGAAGCAGAGCTGTATGTCAGCTTCCGA
GACCTGGGCTGGCAGGACTGGATCATCGCGCCTGAAGGCTACGCCGCCTACTACTGTGAG
GGGAGTGTCCTTCCCTCTGAACTCCTACATGAACGCCACCAACCAGCCATCGTGCAG
ACGCTGGTCCACTTCATCAACCCGGAACGGTGCCCAAGCCCTGCTGTGCGCCACGCAG
CTCAATGCCATCTCCGTCCTCTACTTCGATGACAGCTCCAACGTCATCCTGAAGAAATAC
AGAAACATGGTGGTCCGGGCCTGTGGCTGCCACTAG

Στη συνέχεια παρουσιάζονται και οι διαθέσιμες νουκλεοτιδικές αλληλουχίες από την CCDS ^[4] :

Πλήρης Νουκλεοτιδική Αλληλουχία:

Nucleotide Sequence (1296 nt):

ATGCACGTGCGCTCACTGCGAGCTGCGGCGCCGCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCTCTGC
TGCGCTCCGCCCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACCGGCGCCTCCG
CAGCCAGGAGCGGCGGGAGATGCAGCGCGAGATCTCTCCATTTTGGGCTTGCCCCACCGCCCGCGCCCCG
CACCTCCAGGGCAAGCACAACTCGGCACCCATGTTTCATGCTGGACCTGTACAACGCCATGGCGGTGGAGG
AGGGCGGCGGGCCCGGCGGCGCAGGGCTTCTCTACCCCTACAAGGCCGTCTTCAGTACCCAGGGCCCCC
TCTGGCCAGCCTGCAAGATAGCCATTTCTCACCAGCGCCGACATGGTCATGAGCTTCGTCAACCTCGTG
GAACATGACAAGGAATTCTTCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTTTCCAAGATCC
CAGAAGGGGAAGCTGTACGGCAGCCGAATTCCGGATCTACAAGGACTACATCCGGGAACGCTTCGACAA
TGAGACGTTCCGGATCAGCGTTTATCAGGTGCTCCAGGAGCACTTGGGCAGGGAATCGGATCTCTTCTTG
CTCGACAGCCGTACCTCTGGGCCTCGGAGGAGGGCTGGCTGGTGTGTTGACATCACAGCCACCGCAACC
ACTGGGTGGTCAATCCGCGGCACAACCTGGGCCTGCAGCTCTCGGTGGAGACGCTGGATGGCAGAGCAT
CAACCCCAAGTTGGCGGGCCTGATTGGGCGGCACGGGCCCCAGAACAAGCAGCCCTTCATGGTGGCTTTC
TTCAAGGCCACGGAGGTCCACTTCCGACGATCCGGTCCACGGGGAGCAAACAGCGCAGCCAGAACCGCT
CCAAGACGCCCAAGAACCAGGAAGCCCTGCGGATGGCCAACGTGGCAGAGAACAGCAGCAGCGACAGAG
GCAGGCCTGTAAGAAGCAGCAGCTGTATGTCAGCTTCCGAGACCTGGGCTGGCAGGACTGGATCATCGCG
CCTGAAGGCTACGCCGCCTACTACTGTGAGGGGGAGTGTGCTTCCCTCTGAACTCCTACATGAACGCCA
CCAACCACGCCATCGTGACAGCGCTGGTCCACTTCATCAACCCGGAACGGTGCCCAAGCCCTGCTGTGC
GCCACGCAGCTCAATGCCATCTCCGTCTCTACTTCGATGACAGCTCCAACGTCATCCTGAAGAAATAC
AGAAACATGGTGGTCCGGGCCTGTGGCTGCCACTAG

Μεταφρασμένη Αλληλουχία :

Translation (431 aa):

MHVRSLRAAAPHSFVALWAPLFLRLSALADFSLDNEVHSSFHRRRLRSQERREMQRREILSILGLPHRPRP
HLQGHNSAPMFMLDLYNAMAVEEGGGPGGQGSYPYKAVFSTQGPPLASLQDSHF LT DADMVMSFVNLV
EHDKEFFHPRYHHREFRFDLSKIPEGEAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFL
LDSRTLWASEEGWLVDITATSNHVVNPRHNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPFMVAF
FKATEVHFRSIRSTGSKQRSQNRSKTPKNQEALRMANVAENSSSDQRQACKKHEL VVSFRDLGWQDWIIA
PEGYAAYYCEGECAFP LNSYMNATNHAIVQTLVHF INPETVPKPCCAPTQLNAISVLYFDDSSNVILKKY
RNMVVRACGCH

2.3 Μετάφραση των νουκλεοτιδίων

Σε αυτήν την υποενότητα παρουσιάζονται τα αποτελέσματα από την μετάφραση της πλήρους νουκλεοτιδικής αλληλουχίας από την CCDS, της οστικής μορφογενετικής πρωτεΐνης του ανθρώπου. Η μετάφραση έγινε με το TRANSLATE TOOL που προσφέρει η ιστοσελίδα www.expasy.org. Η πρώτη δοκιμή μετάφρασης έγινε για όλη την νουκλεοτιδική αλληλουχία. Μετά το πέρας της διαδικασίας μετάφρασης το σύστημα προσφέρει διάφορα πιθανά αναγνωστικά πλαίσια. Παρακάτω παρουσιάζεται το πρώτο αναγνωστικό πλαίσιο στο οποίο το γονίδιο κωδικοποιείται από την αρχή έως το τέλος του και δεν διακόπτεται^[5].

Στην συνέχεια επιχειρήθηκε διαφορετικό τμήμα από την πλήρη αλληλουχία της CCDS.

A.

5'3' Frame 1

Met HVRSLRAAAPHSFVALWAPLFLRLRSALADFSLDNEVHSSFIHRRRLRSQERRE Met QREILSILGLPHRPRPHLQGGKHSAP Met F Met LDLYNA Met AVEEGGGPGGGGFSYPYKAVFSTQGPPLASLQDSHFLTAD Met V Met SFVNLVEHDKEFFHPRYHHREFRFDLSKIEG EAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFLLDSRTLWASEEGWLVFDITATSNHWVNPVRHNLGLQLSVETLD GQSINPKLAGLIGRHGPNKQPF Met VAFFKATEVHFRSIRSTGSKQRSQNRSKTPKNQEALR Met ANVAENSSSDQRQACKKHELYV SFRDLGWQDWIIAPEGYAAAYYCEGECAPFLNSY Met NATNHAIVQTLVHFINPETVPKPCCAPTQLNAISVLVYFDDSSNVILKKYRN Met VVRACGCH Stop

Συγκεκριμένα χρησιμοποιήθηκε η αλληλουχία από τα πρώτα επισημασμένα με μπλε χρώμα έως το τέλος. Όπως αναμενόταν το αποτέλεσμα είναι τελείως διαφορετικό:

B.

5'3' Frame 1

WN Met TRNSSTHATTIESSGLIFPRSQKGKLSRQPNNSGSTRITTSNGAST Met RRRSGSAFIRCSRSTWAGNRISSCSTAVPSGPRRRAG WCLTSQPPATTGWSIRGTTWACSSRWRRW Met GRASTPSWRA Stop LGGTGPRITSSPSWWLSSRPRTSASGPRGANSAAARTAP RRPRTKPCGWPTWQRTAAATRGPRVRSSTSC Met SASETWAGRTGSSRLKATPPTTVRGSVPSL Stop TPT Stop TPPTTSCRRWSTS STRKRCPSPAVRPRSS Met PPSSTSS Met TAPTSS Stop RNTETWWSGPVAAT

Παρακάτω παρουσιάζεται η μετάφραση από την συρραφή των μπλέ τμημάτων της πλήρους αλληλουχίας.

Γ.

5'3' Frame 1

WN Met TRNSSTHATTIESSGLIFPRSQKGKLSRQPNNSGSTRITTSNGAST Met RRRSGSAFIRCSRSTWAGQSINPKLAGLIGRHGPQNK QPF Met VAFFKATEVHFRSIRSTGSKQRSQNRSKTPKNQEALR Met ANVAGLDHRA Stop RLRLRLLL Stop GGVCLPSELLHERHQPRHRA DA

Τέλος μεταφράστηκε η περιοχή λίγο πριν το πρώτο μπλε τμήμα έως το τέλος. Και σε αυτή την περίπτωση το γονίδιο δεν διακόπτεται:

Δ.

5'3' Frame 1

Met SFVNLVEHDKEFFHPRYHHREFRFDLSKIEGEAVTAAEFRIYKDYIRERFDNETFRISVYQVLQEHLGRESDFLLDSRTLWASE EGWLVFDITATSNHWVNPVRHNLGLQLSVETLDGQSINPKLAGLIGRHGPQNKQPF Met VAFFKATEVHFRSIRSTGSKQRSQNRSK TPKNQEALR Met ANVAENSSSDQRQACKKHELYVSFRDLGWQDWIIAPEGYAAAYYCEGECAPFLNSY Met NATNHAIVQTLVHFINPETVPKPCCAPTQLNAISVLVYFDDSSNVILKKYRN Met VVRACGCH Stop

2.4 Ομοιοπαράθεση

Σ αυτή την υποενότητα παρατίθενται τα αποτελέσματα ομοιοπαράθεσης, δηλαδή σύγκρισης ομοιότητας των αμινοξικών αλληλουχιών:

- της αλληλουχίας FASTA FORMAT, που βρέθηκε από την UniProt και αποτελεί την ώριμη μορφή γονιδίου που αποδίδει την BMP 7 Human ^[2]
- της μεταφρασμένης αλληλουχίας από την CCDS
- καθώς και της αλληλουχίας από το πρώτο αποτέλεσμα **A.** μετάφρασης της πλήρους νουκλεοτιδικής αλληλουχίας.

Η διαδικασία πραγματοποιήθηκε με τη βοήθεια του εργαλείου Multalin ^[6].

	10	20	30	40	50	60
1xxxxx0	MHVRS	LRAAAPH	SFVAL	WAPL	FLLRS	ALADFS
2xxxxx1	MHVRS	LRAAAPH	SFVAL	WAPL	FLLRS	ALADFS
3xxxxx2	MHVRS	LRAAAPH	SFVAL	WAPL	FLLRS	ALADFS
Consensus	MHVRS	LRAAAPH	SFVAL	WAPL	FLLRS	ALADFS
Prim. cons.	MHVRS	LRAAAPH	SFVAL	WAPL	FLLRS	ALADFS
	70	80	90	100	110	120
1xxxxx0	ILGLPH	RPRPH	LQGH	NISAP	FMFLD	LYNAM
2xxxxx1	ILGLPH	RPRPH	LQGH	NISAP	FMFLD	LYNAM
3xxxxx2	ILGLPH	RPRPH	LQGH	NISAP	FMFLD	LYNAM
Consensus	ILGLPH	RPRPH	LQGH	NISAP	FMFLD	LYNAM
Prim. cons.	ILGLPH	RPRPH	LQGH	NISAP	FMFLD	LYNAM
	130	140	150	160	170	180
1xxxxx0	LQDSHF	LTDAD	MWMSF	VNLVE	HDK	EFFHPR
2xxxxx1	LQDSHF	LTDAD	MWMSF	VNLVE	HDK	EFFHPR
3xxxxx2	LQDSHF	LTDAD	MWMSF	VNLVE	HDK	EFFHPR
Consensus	LQDSHF	LTDAD	MWMSF	VNLVE	HDK	EFFHPR
Prim. cons.	LQDSHF	LTDAD	MWMSF	VNLVE	HDK	EFFHPR
	190	200	210	220	230	240
1xxxxx0	IRERFD	NETFR	ISVYQ	VLQEH	LGRES	DLFL
2xxxxx1	IRERFD	NETFR	ISVYQ	VLQEH	LGRES	DLFL
3xxxxx2	IRERFD	NETFR	ISVYQ	VLQEH	LGRES	DLFL
Consensus	IRERFD	NETFR	ISVYQ	VLQEH	LGRES	DLFL
Prim. cons.	IRERFD	NETFR	ISVYQ	VLQEH	LGRES	DLFL
	250	260	270	280	290	300
1xxxxx0	HNILGL	QSVET	LDGQS	INPKL	AGLIG	RHGPQ
2xxxxx1	HNILGL	QSVET	LDGQS	INPKL	AGLIG	RHGPQ
3xxxxx2	HNILGL	QSVET	LDGQS	INPKL	AGLIG	RHGPQ
Consensus	HNILGL	QSVET	LDGQS	INPKL	AGLIG	RHGPQ
Prim. cons.	HNILGL	QSVET	LDGQS	INPKL	AGLIG	RHGPQ

	310	320	330	340	350	360
1xxxxx0	QNRSK	TPKIQ	EALRM	ANVAEN	SSDQR	QACKK
2xxxxx1	QNRSK	TPKIQ	EALRM	ANVAEN	SSDQR	QACKK
3xxxxx2	QNRSK	TPKIQ	EALRM	ANVAEN	SSDQR	QACKK
Consensus	QNRSK	TPKIQ	EALRM	ANVAEN	SSDQR	QACKK
Prim. cons.	QNRSK	TPKIQ	EALRM	ANVAEN	SSDQR	QACKK
	370	380	390	400	410	420
1xxxxx0	GECAF	PLNSV	YHATN	HAIVQ	TLVHF	INP
2xxxxx1	GECAF	PLNSV	YHATN	HAIVQ	TLVHF	INP
3xxxxx2	GECAF	PLNSV	YHATN	HAIVQ	TLVHF	INP
Consensus	GECAF	PLNSV	YHATN	HAIVQ	TLVHF	INP
Prim. cons.	GECAF	PLNSV	YHATN	HAIVQ	TLVHF	INP
	430					
1xxxxx0	RIWV	VRACG				
2xxxxx1	RIWV	VRACG				
3xxxxx2	RIWV	VRACG				
Consensus	RIWV	VRACG				
Prim. cons.	RIWV	VRACG				

Το συμπέρασμα είναι ότι οι αλληλουχίες ταυτίζονται. Όπως ήταν αναμενόμενο, το FASTA FORMAT και η μεταφρασμένη αλληλουχία που προσφέρονται βιβλιογραφικά, είναι ίδιες. Όμως και η αλληλουχία A. που επιλέχθηκε σαν γονίδιο από την διαδικασία της μετάφρασης του πρώτου αναγνωστικού πλαισίου, φαίνεται ότι επαληθεύεται και αντιπροσωπεύει όντως το γονίδιο για την BMP 7.

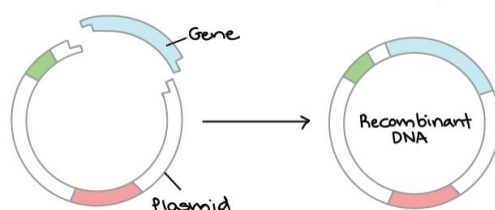
3 DNA CLONING ΚΑΙ ΕΡΓΑΛΕΙΑ ΠΡΟΒΛΕΨΗΣ CUTTING

3.1 Λίγα Στοιχεία Γενετικής Μηχανικής

Η τεχνολογική πρόοδος και η πρόοδος της επιστήμης, οδήγησαν στη δυνατότητα δημιουργίας μορίων DNA in vitro και την εισαγωγή τους σε ζωντανούς οργανισμούς. Η εξέλιξη αυτή αναφέρεται ως «ανασυνδασμένο DNA» ή recombinant, και υπάγεται στην γενετική μηχανική.

Για να παραχθεί το ανασυνδασμένο DNA, χρησιμοποιούνται περιοριστικά ένζυμα που απομονώνονται από βακτήρια και ονομάζονται νουκλεάσες περιορισμού (restriction nucleases) και καταλύουν την υδρόλυση φωσφοδιεστερικών δεσμών σ' ένα νουκλεϊκό οξύ, σε ειδικές θέσεις. Συνεπώς μπορεί να χρησιμοποιηθούν για την παραγωγή μιας ειδικής ομάδας κλασμάτων DNA από ένα γονιδίωμα ^[7].

Κατά την κλωνοποίηση του DNA, που ορίζεται ως η διεργασία παραγωγής πολλών ταυτόσημων αντιγράφων ενός μορίου DNA ^[7], το γονίδιο στόχος που επιλέγεται για κλωνοποίηση εισάγεται σε πλασμίδιο. Γι αυτό, στα ίδια σημεία είναι απαραίτητο να «κοπεί» από τα ίδια περιοριστικά ένζυμα και το πλασμίδιο στο οποίο θα εισαχθεί στη συνέχεια το τμήμα DNA.



Εικόνα 3.1: Εισαγωγή νουκλεϊκής αλληλουχίας σε πλασμίδιο με τη βοήθεια περιοριστικών νουκλεασών ^[8]

Στην διάθεση του ερευνητή υπάρχουν εργαλεία που προσφέρονται για την εύρεση των κατάλληλων περιοριστικών νουκλεασών και την εξακρίβωση των σημείων που αυτές μπορούν να «κόψουν» την αλληλουχία υπό μελέτη.

Στην παρακάτω ενότητα, λοιπόν, παρουσιάζεται η πρόβλεψη κατακερματισμού της αλληλουχίας της BMP από ένζυμα ικανά να «κόψουν» 2 και 3 φορές την αλληλουχία με την βοήθεια των εργαλείων :

- i. NEB Cutter V2.0 : Εκτελεί την εύρεση των νουκλεασών περιορισμού χρησιμοποιώντας γενετικό κώδικα E.Coli ^[9]
- ii. Sequence extractor: προσφέρει επίσης έναν «χάρτη» περιοριστικών ενζύμων της υπο μελέτη αλληλουχίας ^[10]

[illegible]

Ενδεικτικά θα παρουσιαστούν αναλυτικά τα σημεία «κοπής» 3 εκ των νουκλεασών περιορισμού που κόβουν την αλληλουχία 2 φορές , με τη βοήθεια του Sequence Extractor tool ^[10]:

PluTI :

M H V R S L R A A A P H S F V A L W A P L F L L R S A
ATGCACGTGCGCTCACTGCGAGCTGC**GGCGC**GCACAGCTTCGTGGCGCTCTGGGCACCCCTGTTCTCTGCTGCGCTCCGC
 ^10 ^20 ^30 ^40 ^50 ^60 ^70
TACGTGCACGCGAGTGACGCTCGACGCG**CGCGG**CGTGTCTGAAGCACCGCGAGACCCGTGGGGACAAGGACGACGCGAGGCG

Και

L A D F S L D N E V H S S F I H R R L R S Q E R R E M
CCTGGCCGACTTCAGCCTGGACAACGAGGTGCACTCGAGCTTCATCCACC**GGCGC**CTCCGCAGCCAGGAGCGGCGGGAGA
 ^90 ^100 ^110 ^120 ^130 ^140 ^150
GGACCGGCTGAAGTCGGACCTGTTGCTCCACGTGAGCTCGAAGTAGGTGGC**CGCGG**AGGCGTCGGTCTCGCCGCCCTCT

EcoRI:

S F V N L V E H D K E F F H P R Y H H R E F R F D L
TGAGCTTCGTCAACCTCGTGGAACTGACAAGG**AATTC**TTCCACCCACGCTACCACCATCGAGAGTTCCGGTTTGATCTT
 ^410 ^420 ^430 ^440 ^450 ^460 ^470
ACTCGAAGCAGTTGGAGCACCTTGTAAGTCT**CCTTA**GAAGGTGGGTGCGATGGTGGTAGCTCTCAAGGCCAACTAGAA

Και

S K I P E G E A V T A A E F R I Y K D Y I R E R F D N
TCCAAGATCCCAGAAGGGGAAGCTGTACGGCAGCCG**AATTC**GGATCTACAAGGACTACATCCGGGAACGCTTCGACAA
 ^490 ^500 ^510 ^520 ^530 ^540 ^550
AGGTTCTAGGTCTTCCCTTCGACAGTGCCGTCG**GCTTA**AGGCCTAGATGTTCTGATGTAGGCCCTTGCGAAGCTGTT

BsrI:

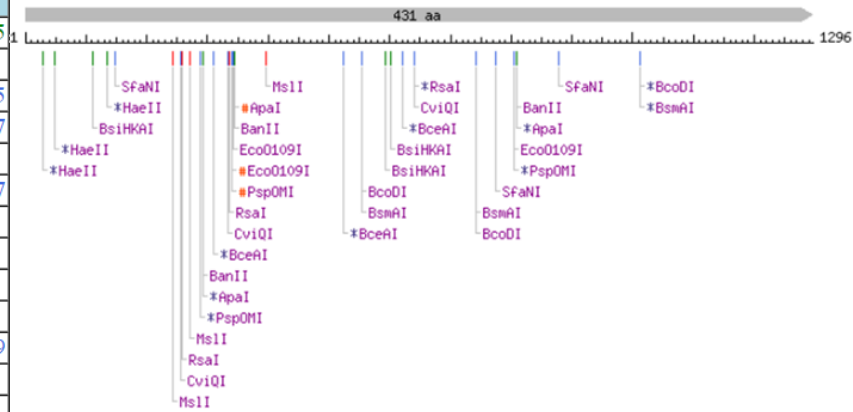
T L W A S E E G W L V F D I T A T S N H W V V N P R
GTACCCTCTGGGCCTCGGAGGAGGCTGGCTGGTGTGACATCACAGCCACCAGCAACCACTG**GGT**GGTCAATCCGCGG
 ^650 ^660 ^670 ^680 ^690 ^700 ^710
CATGGGAGACCCGGAGCCTCCTCCGACCGACCACAACTGTAGTGTCGGTGGTCTGTTGGTGAC**CCACC**AGTTAGGCGCC

Και

I 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
GATCATCG**GGC**CTGAAGGCTACGCCGCTACTACTGTGAGGGGGAGTGTGCCTTCCCTCTGAACTCCTACATGAACGCCA
 ^1050 ^1060 ^1070 ^1080 ^1090 ^1100 ^1110
CTAGTAG**GGG**GACTTCCGATGCGGCGGATGATGACACTCCCCCTCACACGAAGGGAGACTTGAGGATGTACTTGCGGT

Αντίστοιχα επιλέγοντας την ένδειξη 3 cutters, βρίσκουμε ότι 13 περιοριστικά ένζυμα ανταποκρίνονται στην κοπή της αλληλουχίας 3 φορές στα παρακάτω σημεία:

#	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 _↓ RGCY ⁺ 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



4 ΔΥΝΑΤΟΤΗΤΕΣ ΕΥΡΕΣΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΓΝΩΡΙΣΜΑΤΩΝ ΑΛΛΗΛΟΥΧΙΑΣ ΠΡΩΤΕΪΝΗΣ

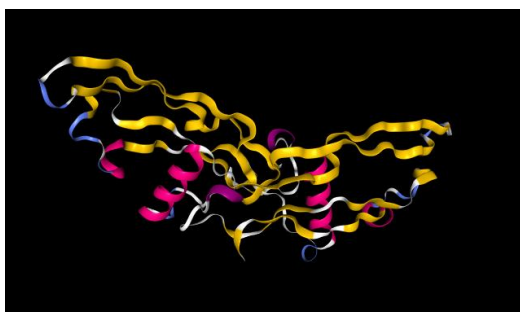
Με τη βοήθεια της επιστήμης της βιοπληροφορικής, είναι πλέον δυνατή η εύρεση πλήθους πληροφοριών που αφορούν μια πρωτεΐνη με το πάτημα μερικών κουμπιών. Μέσω προγραμμάτων όπως το RasMol αλλά και το Java Protein Workshop, μπορεί κανείς να βρει πληροφορίες για μια πρωτεΐνη όπως:

- Την κρυσταλλική απεικόνισή της
- Οπτικοποιημένες πληροφορίες για το που βρίσκονται συγκεκριμένα αμινοξέα
- Πόσες α -έλικες ή πόσα β -ελάσματα διαθέτει

Τα στοιχεία αυτά εμφανίζονται στον χρήστη, μετά από «άνοιγμα» του .pdb file με τα προαναφερθέντα προγράμματα. Το αρχείο .pdb διατίθεται από τις βάσεις δεδομένων πρωτεϊνών, και αποθηκεύεται στον υπολογιστή.

Στη συγκεκριμένη περίπτωση παρατίθενται τα χαρακτηριστικά που αφορούν την πρωτεΐνη BMP7 Human με κωδικό P18075, η οποία επιλέχθηκε τυχαία για την εκπόνηση αυτής της εργασίας. Από την σελίδα της RCSB PDB ^[12] για την πρωτεΐνη, αποθηκεύτηκε στον υπολογιστή το αρχείο τύπου .pdb . Στη συνέχεια το αρχείο αυτό «ανοίγεται» στον υπολογιστή και βρίσκουμε τα παρακάτω στοιχεία:

- ✚ Η ανάλυση της, από την συγκεκριμένη πηγή που χρησιμοποιείται είναι 2.00 Angstroms ^[11].
- ✚ Η πρωτεΐνη είναι μια **μοναδική αλυσίδα** που αποτελείται από **139 κατάλοιπα (residues)** .
- ✚ Η πρωτεΐνη διαθέτει **3 α -έλικες** , 13%, και **9 β -ελάσματα**, 33%
- ✚ Η κρυσταλλική απεικόνιση της πρωτεΐνης είναι :



- ✚ Η πρωτεϊνική αλυσίδα μπορεί να περιέχει ετεροάτομο N-ακετυλο-D-γλυκοζαμίνης. (να προσθέσω φωτο από protein workshop)
- ✚ Διαθέτει γύρω του 71 μόρια νερού.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. <https://web.expasy.org/translate/> , [5 Μαρτίου 2019]
2. <https://www.uniprot.org/uniprot/P18075> , [5 Μαρτίου 2019]
3. <https://www.ebi.ac.uk/ena/data/view/AAH08584> [6 Μαρτίου 2019]
4. <https://www.ncbi.nlm.nih.gov/CCDS/CcbsBrowse.cgi?REQUEST=CCDS&GO=MainBrowse&DATA=CCDS13455.1> [6 Μαρτίου 2019]
5. <https://web.expasy.org/translate/> [6 Μαρτίου 2019]
6. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_multalin.html [6 Μαρτίου 2019]
7. Ν. Ανάγνου, Π. Παπαζαφείρη, Ι. Παπαματθαϊάκης, Κ. Σταματόπουλος, 2016, *Βασικές Αρχές Κυτταρικής Βιολογίας – 3^η Έκδοση*, Εκδόσεις Πασχαλίδης Broken Hill, Νικόσια
8. <https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/overview-dna-cloning> [21 Μαρ. 19]
9. <http://nc2.neb.com/NEBcutter2/> [21 March 2019]
10. <https://www.bioinformatics.org/seqext/> [21 Μαρ. 19]
11. Greenwald, J., Groppe, J., Gray, P., Wiater, E., Kwiatkowski, W., Vale, W., Choe, S., The BMP7/ActRII Extracellular Domain Complex Provides New Insights into the Cooperative Nature of Receptor Assembly, (2003) *Mol.Cell* **11**: 605-617
12. <https://www.rcsb.org/> [8 Μαΐου 2019]