

ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ



ΤΜΗΜΑ ΧΗΜΕΙΑΣ

ΔΟΜΙΚΗ ΒΙΟΧΗΜΕΙΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΡΓΑΣΙΑ ΚΥΡΟ-ΟΔΥΣΣΕΥΣ ΚΟΣΜΑΤΟΥ (8957)

Θεωρητικό Μέρος

Στόχος της εργασίας αυτής είναι η υπολογιστική μελέτη της πρωτεΐνης Bone Morphogenetic Protein 3 (Human), δηλαδή της οστεϊκής μορφογενετικής πρωτεΐνης του ανθρώπου.

Η μελέτη αφορά στον γενετικό κώδικα της πρωτεΐνης και την αμινοξική ακολουθία της, στο οποίο κομμάτι εκτελείται και ομοιοπαράθεση μεταξύ των διάφορων αμινοξικών ακολουθιών οι οποίες λαμβάνονται από διαφορετικές βάσεις δεδομένων πρωτεϊνών.

Στη συνέχεια πραγματοποιείται έλεγχος των περιοριστικών ενδονουκλεασών οι οποίες μπορούν να δράσουν στη πρωτεΐνη.

Τέλος, χρησιμοποιούνται υπολογιστικά προγράμματα για τη μοντελοποίηση της πρωτεΐνης και την άντληση περαιτέρω δεδομένων.

Πειραματικό Μέρος

Αρχικά επιλέγεται η οστεϊκή μορφογενετική πρωτεΐνη 3 του ανθρώπου από τη βάση δεδομένων UniProt (<https://www.uniprot.org>). Από εκεί λαμβάνονται τα εξής γενικά δεδομένα:

Protein | **Bone morphogenetic protein 3**

Gene | **BMP3**

Organism | *Homo sapiens (Human)*

Η ίδια ιστοσελίδα παρέχει επιπλέον πληροφορίες για τη λειτουργία της πρωτεΐνης.

Η BMP3 ρυθμίζει αρνητικά την οστεϊκή πυκνότητα. Δρά ανταγωνιστικά ως προς άλλες οστεϊκές μορφογενετικές πρωτεΐνες, καταστέλλοντας την ικανότητα τους να επιφέρουν διαφοροποίηση στα προγονικά οστεϊκά κύτταρα, και στην ικανότητα τους για οστεοποίηση.

Από την ίδια βάση δεδομένων λαμβάνεται επίσης η αμινοξική αλληλουχία της πρωτεΐνης σε FASTA format. (single letter code)

10	20	30	40	50
MAGASRLFL	WLGCFVSLA	QGERPKPPFP	ELRKAVPGDR	TAGGGPDSEL
60	70	80	90	100
QPQDKVSEHM	LRLYDRYSTV	QAARTPGSLE	GGSQPWRPRL	LREGNTVRSF
110	120	130	140	150
RAAAAETLER	KGLYIFNLTS	LTKSENILSA	TLYFCIGELG	NISLSCPVS
160	170	180	190	200
GCSHHAQRKH	IQIDLSAWTL	KFSRNQSQLL	GHLSVDMAKS	HRDIMSWLSK
210	220	230	240	250
DITQLLRKAK	ENEEFLIGFN	ITSKGRQLPK	RRLPFPEPYI	LVIYANDAAIS
260	270	280	290	300
EPESVVSLLQ	GHRNFPTGTV	PKWDSHIRAA	LSIERRKKRS	TGVLLPLQNN
310	320	330	340	350
ELPGAHEYQYK	KDEVWEERKP	YKTLQAQAP	KSKNKKKQRK	GPHRKSQTLQ
360	370	380	390	400
FDEQTLKKAR	RKQWIEPRNC	ARRYLKVDFA	DIGWSEWIIIS	PKSFDAYYCS
410	420	430	440	450
GACQFPMPKS	LKPSNHATIQ	SIVRAVGVVP	GIPEPCCVPE	KMSSLSILFF
460	470			
DENKNVVLKV	YPNMTVESCA	CR		

Λαμβάνεται επίσης η αλληλουχία αμινοξέων απο άλλες βάσεις δεδομένων, όπως η GENBANK:

```
>ENA|BC096270|BC096270.2 Homo sapiens bone morphogenetic protein 3, mRNA (cDNA clone MGC:116935 IMAGE:40006303), complete cds.
TCAGCGTTGGAGTGGAGACGGCGCCCGCAGCGCCCTGCGCGGGTGAGGTCCGCGCAGCTG
CTGGGGAAGAGCCACCTGTGAGGCTGCGCTGGGTGAGCGCAGCAAGTGGGGCTGGCCGC
TATCTCGCTGCACCCGGCGCGTCCCGGGCTCCGTGCGCCCTCGCCCCAGCTGGTTTGGGA
GTTCAACCTCGGCTCGCCGCCGGCTCCTTGCCTTCGAGTGTCCCGCAGCGACGCC
GGGAGCCGACGCGCGCGCGGTACCTAGCCATGGCTGGGGCGAGCAGGCTGCTTTTCT
GTGGCTGGGCTGCTTCTGCGTGAGCTGGCGCAGGGAGAGAGACCAGGACCACTTTCCC
GGAGCTCCGCAAGCTGTGCGAGGTGACCGCACGGCAGGTGGTGGCCCGACTCCGAGCT
GCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAGCAGGT
CCAGGCGGGCCGGACACCGGGCTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCCTCGGCT
CCTGCGCGAAGGCAACACGGTTCGCGAGCTTTCGGGCGGCGAGCAGAGAACTCTTGAAAG
AAAAGGACTGTATATCTTCAATCTGACATCGCTAACCAAGTCTGAAAACATTTGTCTGC
CACACTGTATTTCTGTATTGGAGAGCTAGGAAACATCAGCCTGAGTTGTCCAGTGTCTGG
AGGATGCTCCCATCATGCTCAGAGGAAACACATTGAGTTGATCTTTCTGCATGGACCT
CAAAATTCAGCAGAAACCAAGTCAACTCCTTGCCATCTGTGAGTGGATATGGCCAAATC
TCATCGAGATATTATGCTCTGGCTGTCTAAGATATCACTCAATTCTTGAGGAAGGCCAA
AGAAAAATGAAGAGTTCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAA
GAGGAGGTTACCTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTT
TGAGCCAGAAAGTGTGGTATCAAGCTTACAGGGACACCGGAATTTCCCACTGGAACTGT
TCCCAAATGGGATAGCCACATCAGAGCTGCCCTTTCCATTGAGCGGAGGAAGAAGCGCTC
TACTGGGGTCTTGTGCTCTGCAGAACACGAGCTTCTGGGGCAGAAATACCAATATAA
AAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCTTCAGGCTCAGGCCCTGA
AAAGAGTAAGAATAAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGACGCTCA
ATTTGATGAGCAGACCTGAAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTG
CGCCAGGAGATACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTC
CCCCAAGTCTTTGATGCCTATTATTGCTCTGGAGCATGCCAGTTCCTCCATGCCAAAGTC
TTTGAAGCCATCAAAATCATGCTACCATCCAGAGTATAGTGAGAGCTGTGGGGTCTTCC
TGGGATTCCTGAGCCTTGTGTACCAAGAAAGATGTCCTCACTCAGTATTTTATCTT
TGATGAAAATAAGAAATGATGCTTAAAGTATACCTAACATGACAGTAGAGTCTTGC
TTGCAGATAACCTGGCAAGAACTCATTGAAATGC
```

Και η βάση δεδομένων CCDS, από όπου λαμβάνεται και η αλληλουχία νουκλεοτιδίων για την συγκεκριμένη πρωτεΐνη:

Nucleotide Sequence (1419 nt):

```
ATGGCTGGGGCGAGCAGGCTGCTCTTTCTGTGGCTGGGCTGCTTCTGCGTGAGCCTGGCGCAGGGAGAGA
GACCGAAGCCACCTTTCCCGGAGCTCCGCAAAGCTGTGCCAGGTGACCGCACGGCAGGTGGTGGCCCGGA
CTCCGAGCTGCAGCCGCAAGACAAGGTCTCTGAACACATGCTGCGGCTCTATGACAGGTACAGCACGGTC
CAGGCGGGCCGGACACCGGGCTCCCTGGAGGGAGGCTCGCAGCCCTGGCGCCCTCGGCTCCTGCGCGAAG
GCAACACGGTTTCGAGCTTTGCGGCGGCAGCAGCAGAACTCTTGAAAGAAAAGGACTGTATATCTTCAA
TCTGACATCGCTAACCAGTCTGAAAACATTTTGTCTGCCACACTGTATTTCTGTATTGGAGAGCTAGGA
AACATCAGCCTGAGTTGTCCAGTGTCTGGAGGATGCTCCCATCATGCTCAGAGGAAACACATTGAGATTG
ATCTTTCTGCATGGACCCTCAAATTGAGCAGAAACCAAAGTCAACTCCTTGGCCATCTGTGAGTGGATAT
GGCCAAATCTCATCGAGATATTATGCTCTGGCTGTCTAAAGATATCACTCAACTCTTGAGGAAGGCCAAA
GAAAATGAAGAGTTCTCATAGGATTTAACATTACGTCCAAGGGACGCCAGCTGCCAAAGAGGAGGTTAC
CTTTCCAGAGCCTTATATCTTGGTATATGCCAATGATGCCGCCATTTCTGAGCCAGAAAGTGTGGTATC
AAGCTTACAGGGACACCGGAATTTCCCACTGGAAGTGTCCCAATGGGATAGCCACATCAGAGCTGCC
CTTTCCATTGAGCGGAGGAAGAAGCGCTCTACTGGGGTCTTGTGCTCTGCAGAACACGAGCTTCCTG
GGGCAGAATACCAAGTATAAAAAGGATGAGGTGTGGGAGGAGAGAAAGCCTTACAAGACCCTTCAGGCTCA
GGCCCTGAAAAGAGTAAGAATAAAAAGAAACAGAGAAAGGGGCCTCATCGGAAGAGCCAGACGCTCCAA
TTTGATGAGCAGACCCTGAAAAGGCAAGGAGAAAGCAGTGGATTGAACCTCGGAATTGCGCCAGGAGAT
ACCTCAAGGTAGACTTTGCAGATATTGGCTGGAGTGAATGGATTATCTCCCCAAGTCCTTTGATGCCCTA
TTATTGCTCTGGAGCATGCCAGTTCCTCATGCCAAAGTCTTTGAAGCCATCAAATCATGCTACCATCCAG
AGTATAGTGAGAGCTGTGGGGGTGCTTCTGGGATTCTGAGCCTTGCTGTGTACCAGAAAAGATGTCCT
CACTCAGTATTTTATCTTTGATGAAAATAAGAATGTAGTGCTTAAAGTATACCCTAACATGACAGTAGA
GTCTTGCGCTTGACAGATAA
```

Translation (472 aa):

```
MAGASRLFLWLGCFCVSLAQGERPKPPFPELRKAVPGDRTAGGGPDSELQPQDKVSEHMLRLYDRYSTV
QAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAAEFLERKGLYIFNLTSLTSENILSATLYFCIGELG
NISLSCPVSGGCSHHAQRKHIQIDLSAWTLKFSRNQSQLLGHLSVDMAKSHRDIMSWLSKDITQLLRKAK
ENEEFLIGFNITSKGRQLPKRRLPFPEPYILVYANDAAISEPESVSSSLQGHRNFPTGTVPKWDSHIRAA
LSIERRKKRSTGVLLPLQNNELPGAQYQYKKDEVWEERKPYKTLQAQAPKSKNKKKQKRGPHRKSQTLQ
FDEQTLKKARRKQWIEPRNCARRYLKVDFADIGWSEWIIISPKSFDAYYCSGACQFMPKSLKPSNHATIQ
SIVRAVGVPVGIPEPCCVPEKMSSLSILFFDENKNVVLKVYPNMTVESACR
```

Στη συνέχεια, το εργαλείο το οποίο χρησιμοποιείται είναι το TRANSLATE TOOL της ιστοσελίδας expasy.org. Το εργαλείο αυτό προσφέρει τη δυνατότητα μετάφρασης νουκλεοτιδικής αλληλουχίας, σε διάφορα πλαίσια ανάγνωσης. Χρησιμοποιώντας τη νουκλεοτιδική αλληλουχία την οποία παρέχει η βάση δεδομένων CCDS, και τοποθετώντας τη στο εργαλείο TRANSLATE TOOL λαμβάνονται τα παρακάτω:

<p>5'3' Frame 1</p> <p>MAGASRLFLWLCFCVSLAQGERPKPPPELRAKAVPGDRTAGGQPDSELPQDQVSEHMLRLDYRSTVQAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAAEATLERKGLYIFNLTSLKSENILSATLYFCIGELGNISLSCPVSGGC SHHAQRKHIOIDLAWTLKFSRNSQQLGHLVDMAKSHRDIWSLSDITQLLRKAKENEEFLGNITSKGRQLPKRRLPFPEYVILVYANDAAISEPVSVLSQGHNFPTGTVPKWDHSHRAALSIERKKRSTGVLLPLQNNELPG AEYQYKKDEVWEERKPYKTLQAQPEKSKNKKQRKQPHRSQTLQFDEQTLKKARRKQWIEPRNCARRYLKVPDADIGWSEWISPKSFDAYYCSGACQFPMKSLKPSNHATIOQSVRAVGVVPGIPEPCCVPEKMSSLSILFFDENKNV VLKVIPNMTVESACACR-</p>
<p>5'3' Frame 2</p> <p>WLGRAGSCFCWAASA-AWRERDRSHLSRSSAKLCQVTAQVARTPSCSRKTRSLNTOCGSMTGTARSRRPGHRAFWREARSPGALGSCAKATRAFAAFGRQOQKLLKEKDCISSI-HR-PSLKTFCPLPHCISVLES-ETSA-VVQCLEDA PIMLRGNTFRLLIFLHGSPNSAETHVNSLAICQWIPNLTIELCPGLKRISLNS-GRPKMKSSS-DLTLRPDASCQCGVYLQSLISWYMPMPPFLSKQWYQAYRDTGIFPLEIFPNGIATSELPFLSGGRSALLGSCCLCRTTSFLG QNTSIKMRMGRRSLTRPFRPLRPLKRVKRNREGLIGARRSNLMSRP-KRQGESSLNGLIAPGDTSR-TLQILAGVNLSPSPMLPIIALEHASSPCQSL-SHQIMLPSRV--ELWGSFLGLSLAVYQRCPHSVFYSLMKIRM- CLKYTLT-Q-SLALAD</p>
<p>5'3' Frame 3</p> <p>GWGEQAALSVAGLLRLREPAGRETEATFPGAPQSCAR-PHGRWWPGLRAAAARQGL-THAAAL-QVQHGPGGPDGTGLPGGRLAALAPSAPARRQHGSQLSGGSSRNS-KKRTVYLSQDIANOV-KHFVCHTVFLYWRARKHQPESLSSVNRML PSCSEETHSD-SFCMDPQIQKPKSTPFWPSVSGYQISSRYIVLAV-RYHSTLEEGQRK-RVPHRI-HYVQGTAAKEEVTFSRALYLGICQ-CRHF-ARKCGIKLTGTPEFSHWNCSCMG-PHSCPPH-AEEALYGLAASAEQASWG RIPV-KG-GVGGEKALQDPGSGP-KE-E-KETEKASSEEDAPI--ADPEKGEKAVID-TSELQEIQQRLCRYWLE-MDYLPQVL-CLLLNWSMEVPHAKVFEATKSCYHPEYSESOGGRSWS-ALLCTRKDVLTYQFYL--K-ECS A-SIP-HDSRLRLQI</p>
<p>3'5' Frame 1</p> <p>LSASARLYCHVRVYFKHYILFIKE-NTE-CHLFWYTARLNRPNRDPHSSHYTLDSMI-WLQRLNHGELACSRATIGIKLGGDNPTPANICKVYLEVSPGAIPRNPPLSPCLFQGLLIKLERLALPMRPLSLFLFILTLFRGLSLKGL VRLSLPHLILFILVFCPRKLVVLQROQDPSRALLPPLNGKSSDVAIPFGNSSSGKIPVSD-A-YHTFWLRNGGIIGYQDIRLWKR-PPLWQLASLGRNVKSYELFIFFGLPQELSDIFRQPGHNISMRFQHIH-QMAKELTVLSAEFE GPCRKINLNVFPLSMGASSRHWITQADVS-LSNTEIQCGRQNVRLG-RQIEDIQSFSSKFCOCPKAAANRVAFQEPRAPGLRASLGQARCPGRLDRAVVFIEPQHVFRDLVRLQLGVRRATTCAVWHSFAELRERWLRLSLRQA HAAEAQPKQEPARPSH</p>
<p>3'5' Frame 2</p> <p>YLQADSTVMLGYTLSTTFLFSSKNKILSEIDFSGTQCGSGIPGTTPTALTILNMVA-PDGFKDFCMGNWHAPEQ--ASKDLGEIHSLOPISAKSTLYLLAQFRGSIHCFLLAFFRVCSNNSVWLFR-GPFLCFPLFLLFSGA-A-RVL -GLSSHTSSFLYISAPGSSLCFRGSKTVERFFLRSMERAAALMWLSHLGTVPVGKFRCPKLDITLSSGEMAAASLAYTKI-GSGKGNLLFGSWRPLDVMLNFMNNSFSFLAFLKS-VISLDSQDIISR-DLAISTDRWPRS-LWFLNLN VHAERSI-MCFL-A-WEHPDGTGLRLMFFSSPIQKYSVADKMFSDLVSDVRLKIYSPFLSRVSAARAKLRTVLPSSRSRGQCGEPPSPREPGRVAAWTVLYLS-SRSMCSETLSOGCSSESSEGGPPAVRSPGTALRSSGKGGLSPCARL TQKQPSHRKSSLLAPA</p>
<p>3'5' Frame 3</p> <p>ICKRRTLLSC-GIL-ALHSYFHQRIKY-VRTSFLVHSAQESQERPPQLSLYSGW-HDLMASKTLAWGTGMLQSNRNRHQTWGR-SIHSSQYLQSLP-GISWRNSEVQSTAFSLPFGSAGHQIGASGSSDEAPFSVSFYSGQPEPEGSC KAFSPPTPHFFYTGLPQEARCSAAEARQP-SASSAQWKGQL-CGYPIWEQFQWENSQVPSLIPHLAQKRWHRHWPRIKALEKVTSSLAGVFWT-C-IL-GTLHFLWPSRVE-YL-TART-YLDEIWPLYLTDGGVDFGFC-I-G SMQKQDQSECVSSEHDSILQTLDNSS-CFLAQYRNTVWQTKCFQTLWMSD-RYTLVFPQFLLLPESCEPCCLRAGAEGARAASLPFGSPVSGPPGCTCHRAAACVQRCLAAAAARSPGHHLPCGHLAQLOGAPKVASVSLPAPGS RRSSPATERAACSPQP</p>

Το πρώτο πλαίσιο ανάγνωσης είναι αυτό στο οποίο το γονίδιο κωδικοποιείται χωρίς να διακόπτεται.

Πέρα απο τα διάφορα πλαίσια ανάγνωσης, δοκιμάστηκε και η μετάφραση του μπλέ τμήματος της νουκλεοτιδικής αλληλουχίας.

<p>5'3' Frame 1</p> <p>KLLKEKDCISSI-HR-PSLKTFCPLPHCISVLES-ETSA-VVQCLEDAPIMLRGNTFRLLIFLHGSPNSAETHVNSLAICQWIPNLTIELCPGLKRISLNS-GRPKMKSSS-DLTLRPDASCQCGVYLQSLISWYMPMPPFLSKQWYQ AYRDTGIFPLEIFPNGIATSELPFLSGGRSALLGSCCLCRTTSFLGQNTSIKMRMGRRSLTRPFRPLRPLKRVKRNREGLIGARRSNLMSRP-KRQGESSLNGLIAPGDTSR-TLQILAGVNLSPSPMLPIIALEHASSPCQ SL-SHQIMLPSRV--ELWGSFLGLSLAVYQRCPHSVFYSLMKIRM-CLKYTLT-Q-SLALAD</p>
<p>5'3' Frame 2</p> <p>NS-KKRTVYLSQDIANOV-KHFVCHTVFLYWRARKHQPESLSSVNRMLPSCSEETHSD-SFCMDPQIQKPKSTPFWPSVSGYQISSRYIVLAV-RYHSTLEEGQRK-RVPHRI-HYVQGTAAKEEVTFSRALYLGICQ-CRHF-ARKCGIK LTGTPEFSHWNCSCMG-PHSCPPH-AEEALYGLAASAEQASWGRIPIV-KG-GVGGEKALQDPGSGP-KE-E-KETEKASSEEDAPI--ADPEKGEKAVID-TSELQEIQQRLCRYWLE-MDYLPQVL-CLLLNWSMEVPHAK VFEATKSCYHPEYSESOGGRSWS-ALLCTRKDVLTYQFYL--K-ECSA-SIP-HDSRLRLQI</p>
<p>5'3' Frame 3</p> <p>TLERKGLYIFNLTSLKSENILSATLYFCIGELGNISLSCPVSGGCASHHAQRKHIOIDLAWTLKFSRNSQQLGHLVDMAKSHRDIWSLSDITQLLRKAKENEEFLGNITSKGRQLPKRRLPFPEYVILVYANDAAISEPVSVLSQGHNFPTGTVPKWDHSHRAALSIERKKRSTGVLLPLQNNELPGAEYQYKKDEVWEERKPYKTLQAQPEKSKNKKQRKQPHRSQTLQFDEQTLKKARRKQWIEPRNCARRYLKVPDADIGWSEWISPKSFDAYYCSGACQFPMKSLKPSNHATIOQSVRAVGVVPGIPEPCCVPEKMSSLSILFFDENKNVVLKVIPNMTVESACACR-</p>
<p>3'5' Frame 1</p> <p>LWHGELACSRATIGIKLGGDNPTPANICKVYLEVSPGAIPRNPPLSPCLFQGLLIKLERLALPMRPLSLFLFILTLFRGLSLKGLVRLSLPHLILFILVFCPRKLVVLQROQDPSRALLPPLNGKSSDVAIPFGNSSSGKIPVSD-A-YHTFWLRNGGIIGYQDIRLWKR-PPLWQLASLGRNVKSYELFIFFGLPQELSDIFRQPGHNISMRFQHIH-QMAKELTVLSAEFEGPCRKINLNVFPLSMGASSRHWITQADVS-LSNTEIQCGRQNVRLG-RQIEDIQSFSSKFCOCPKAAANRVAFQEPRAPGLRASLGQARCPGRLDRAVVFIEPQHVFRDLVRLQLGVRRATTCAVWHSFAELRERWLRLSLRQAHAEEAQPKEQPARPSH</p>
<p>3'5' Frame 2</p> <p>PMGNWHAPEQ--ASKDLGEIHSLOPISAKSTLYLLAQFRGSIHCFLLAFFRVCSNNSVWLFR-GPFLCFPLFLLFSGA-A-RVL-GLSSHTSSFLYISAPGSSLCFRGSKTVERFFLRSMERAAALMWLSHLGTVPVGKFRCPKLDITLSSGEMAAASLAYTKI-GSGKGNLLFGSWRPLDVMLNFMNNSFSFLAFLKS-VISLDSQDIISR-DLAISTDRWPRS-LWFLNLNVAERSI-MCFL-A-WEHPDGTGLRLMFFSSPIQKYSVADKMFSDLVSDVRLKIYSPFLSRVSAARAKLRTVLPSSRSRGQCGEPPSPREPGRVAAWTVLYLS-SRSMCSETLSOGCSSESSEGGPPAVRSPGTALRSSGKGGLSPCARLTQKQPSHRKSSLLAPATQKQPSHRKSSLLAPA</p>
<p>3'5' Frame 3</p> <p>LAWGTGMLQSNRNRHQTWGR-SIHSSQYLQSLP-GISWRNSEVQSTAFSLPFGSAGHQIGASGSSDEAPFSVSFYSGQPEPEGSCKAFSPPTPHFFYTGLPQEARCSAAEARQP-SASSAQWKGQL-CGYPIWEQFQWENSQVPSLIPHLAQKRWHRHWPRIKALEKVTSSLAGVFWT-C-IL-GTLHFLWPSRVE-YL-TART-YLDEIWPLYLTDGGVDFGFC-I-GSMQKQDQSECVSSEHDSILQTLDNSS-CFLAQYRNTVWQTKCFQTLWMSD-RYTLVFPQFLLLPESCEPCCLRAGAEGARAASLPFGSPVSGPPGCTCHRAAACVQRCLAAAAARSPGHHLPCGHLAQLOGAPKVASVSLPAPGSRRSSPATERAACSPQP</p>

Καθώς και η μετάφραση την αλληλουχίας, χωρίς το μπλέ κομμάτι.

<p>5'3' Frame 1</p> <p>MAGASRLFLWLCFCVSLAQGERPKPPPELRAKAVPGDRTAGGQPDSELPQDQVSEHMLRLDYRSTVQAARTPGSLEGGSQPWRPRLREGNTVRSFRAAAASLKPSNHATIOQSVRAVGVVPGIPEPCCVPEKMSSLSILFFDENKNV VLKVIPNMTVESACACR-</p>
<p>5'3' Frame 2</p> <p>WLGRAGSCFCWAASA-AWRERDRSHLSRSSAKLCQVTAQVARTPSCSRKTRSLNTOCGSMTGTARSRRPGHRAFWREARSPGALGSCAKATRAFAAFGRQOHL-SHQIMLPSRV--ELWGSFLGLSLAVYQRCPHSVFYSLMKIRM- CLKYTLT-Q-SLALAD</p>
<p>5'3' Frame 3</p> <p>GWGEQAALSVAGLLRLREPAGRETEATFPGAPQSCAR-PHGRWWPGLRAAAARQGL-THAAAL-QVQHGPGGPDGTGLPGGRLAALAPSAPARRQHGSQLSGGSSIFEAIRKSCYHPEYSESOGGRSWS-ALLCTRKDVLTYQFYL--K-ECS A-SIP-HDSRLRLQI</p>
<p>3'5' Frame 1</p> <p>LSASARLYCHVRVYFKHYILFIKE-NTE-CHLFWYTARLNRPNRDPHSSHYTLDSMI-WLQRCOCPKAAANRVAFQEPRAPGLRASLGQARCPGRLDRAVVFIEPQHVFRDLVRLQLGVRRATTCAVWHSFAELRERWLRLSLRQA HAAEAQPKQEPARPSH</p>
<p>3'5' Frame 2</p> <p>YLQADSTVMLGYTLSTTFLFSSKNKILSEIDFSGTQCGSGIPGTTPTALTILNMVA-PDGFKDAARAKLRTVLPSSRSRGQCGEPPSPREPGRVAAWTVLYLS-SRSMCSETLSOGCSSESSEGGPPAVRSPGTALRSSGKGGLSPCARL TQKQPSHRKSSLLAPA</p>
<p>3'5' Frame 3</p> <p>ICKRRTLLSC-GIL-ALHSYFHQRIKY-VRTSFLVHSAQESQERPPQLSLYSGW-HDLMASKMLLPESCEPCCLRAGAEGARAASLPFGSPVSGPPGCTCHRAAACVQRCLAAAAARSPGHHLPCGHLAQLOGAPKVASVSLPAPGS RRSSPATERAACSPQP</p>

Στη συνέχεια, για να συγκρίνουμε την ορθότητα των αποτελεσμάτων μας, καθώς και των βιβλιογραφικών δεδομένων από τις διάφορες βάσεις πρωτεϊνών, έλαβε χώρα ομοιοπαράθεση των διαφόρων πρωτεϊνικών αλυσιδών, χρησιμοποιώντας το εργαλείο multalin.

Συγκεκριμένα ομοιοπαράτέθηκαν μεταξύ τους: η αλληλουχία αμινοξέων σε FASTA format από το site UniProt, η αλληλουχία από τη βάση δεδομένων CCDS, και η μετεφρασμένη αλληλουχία που λήφθηκε παραπάνω από το εργαλείο translate tool.

	10	20	30	40	50	60
1xxxxx0	M	A	G	A	S	R
3xxxxx2	M	A	G	A	S	R
2xxxxx1	M	A	G	A	S	R
Consensus	M	A	G	A	S	R
Prim.cons.	M	A	G	A	S	R
	70	80	90	100	110	120
1xxxxx0	L	R	L	Y	D	R
3xxxxx2	L	R	L	Y	D	R
2xxxxx1	L	R	L	Y	D	R
Consensus	L	R	L	Y	D	R
Prim.cons.	L	R	L	Y	D	R
	130	140	150	160	170	180
1xxxxx0	L	T	K	S	E	N
3xxxxx2	L	T	K	S	E	N
2xxxxx1	L	T	K	S	E	N
Consensus	L	T	K	S	E	N
Prim.cons.	L	T	K	S	E	N
	190	200	210	220	230	240
1xxxxx0	G	H	L	S	V	M
3xxxxx2	G	H	L	S	V	M
2xxxxx1	G	H	L	S	V	M
Consensus	G	H	L	S	V	M
Prim.cons.	G	H	L	S	V	M
	250	260	270	280	290	300
1xxxxx0	L	V	Y	A	N	A
3xxxxx2	L	V	Y	A	N	A
2xxxxx1	L	V	Y	A	N	A
Consensus	L	V	Y	A	N	A
Prim.cons.	L	V	Y	A	N	A
	310	320	330	340	350	360
1xxxxx0	E	L	P	G	A	E
3xxxxx2	E	L	P	G	A	E
2xxxxx1	E	L	P	G	A	E
Consensus	E	L	P	G	A	E
Prim.cons.	E	L	P	G	A	E

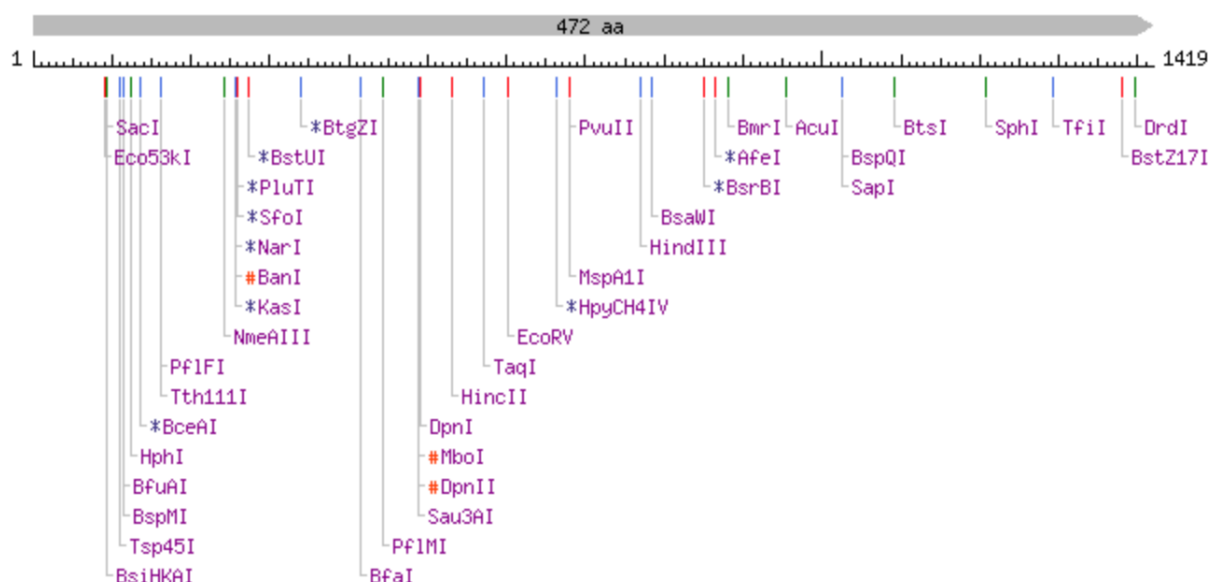
	370	380	390	400	410	420
1xxxxx0	RKQWIEPRNCARRYLKVD	FADIGWSEWII	SPKSF	DAYYCSGACQ	FMPKSLKPSNH	ATI
3xxxxx2	RKQWIEPRNCARRYLKVD	FADIGWSEWII	SPKSF	DAYYCSGACQ	FMPKSLKPSNH	ATI
2xxxxx1	RKQWIEPRNCARRYLKVD	FADIGWSEWII	SPKSF	DAYYCSGACQ	FMPKSLKPSNH	ATI
Consensus	RKQWIEPRNCARRYLKVD	FADIGWSEWII	SPKSF	DAYYCSGACQ	FMPKSLKPSNH	ATI
Prim.cons.	RKQWIEPRNCARRYLKVD	FADIGWSEWII	SPKSF	DAYYCSGACQ	FMPKSLKPSNH	ATI

	430	440	450	460	470
1xxxxx0	SIVRAVGVP	GIPECCVPEKMSSLSIL	FFDENKNVVLKVYP	NMTVESCACR	
3xxxxx2	SIVRAVGVP	GIPECCVPEKMSSLSIL	FFDENKNVVLKVYP	NMTVESCACR	
2xxxxx1	SIVRAVGVP	GIPECCVPEKMSSLSIL	FFDENKNVVLKVYP	NMTVESCACR	
Consensus	SIVRAVGVP	GIPECCVPEKMSSLSIL	FFDENKNVVLKVYP	NMTVESCACR	
Prim.cons.	SIVRAVGVP	GIPECCVPEKMSSLSIL	FFDENKNVVLKVYP	NMTVESCACR	

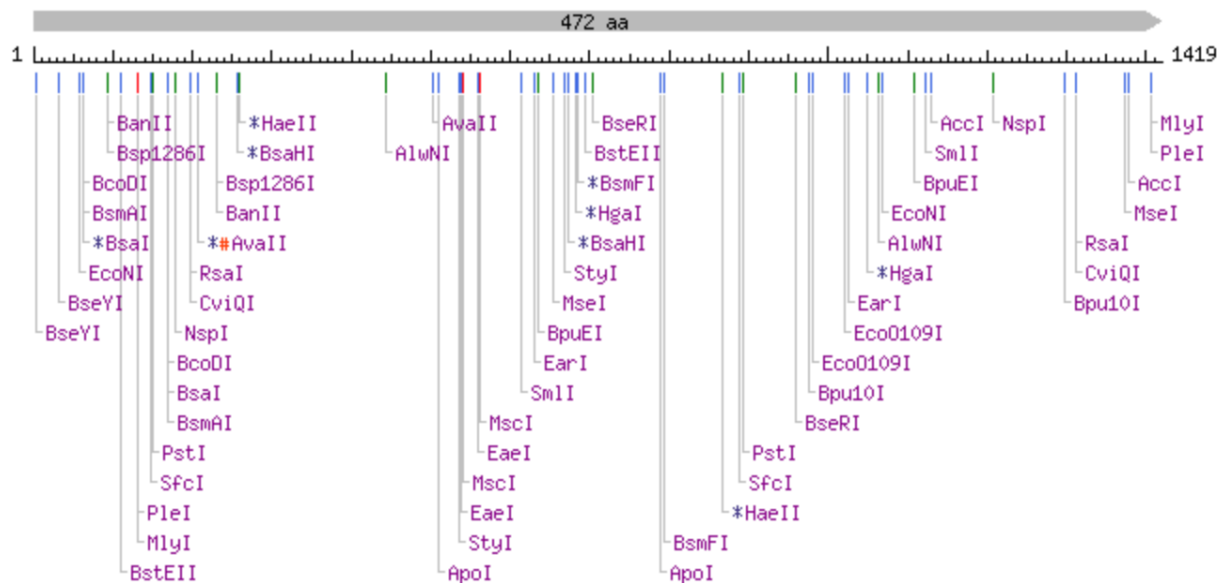
Δεν αποτελεί έκπληξη πως οι αλληλουχίες της βάσης δεδομένων CCDS ταυτίζονται μεταξύ τους, καθώς χρησιμοποιήθηκε η ίδια αλληλουχία νουκλεοτιδίων για τη μετάφραση, αλλά ταυτίζονται και με την αλληλουχία της βάσης UniProt, το οποίο επιβεβαιώνει την αρτιότητα των βιβλιογραφικών δεδομένων.

Στη συνέχεια πραγματοποιήθηκε έλεγχος της δράσης περιοριστικών ενδονουκλεασών επι της πρωτεΐνης, χρησιμοποιώντας το πρόγραμμα Neb cutter V2.0. Εισάγοντας την νουκλεοτιδική αλληλουχία, το πρόγραμμα “προβλέπει” ποιες εμπορικά διαθέσιμες περιοριστικές ενδονουκλεάσες μπορούν να δράσουν επι του γονιδιώματος της πρωτεΐνης, και σε ποια σημεία.

Το πρόγραμμα παράγει μια τεράστια λίστα περιοριστικών ενδονουκλεασών με μορφή γραφήματος ή λίστας. Παρακάτω δίνονται ενδεικτικά γραφήματα και λίστες περιοριστικών ενδονουκλεασών οι οποίες δρουν μια και δύο φορές, αντίστοιχα επι του γονιδιώματος της πρωτεΐνης.

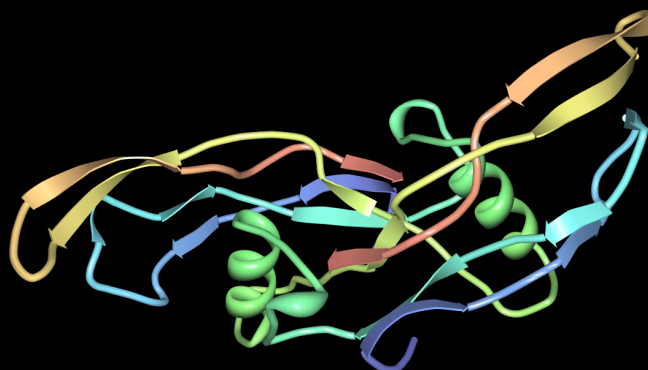


#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AcuI	CTGAAG(N) ₁₄ NN ⁺	list	955/953
2	AfeI	AGC ⁺ GCT	list	*865
3	BanI	G ⁺ GYRC ₂ C	list	#257/261
4	BceAI	ACGGC(N) ₁₂ NN ₂	list	*137/139
5	BfaI	C ⁺ TA ₂ G	list	415/417
6	BfuAI	ACCTGCNNNN ⁺ NNNN ₂	list	115/119
7	BmrI	ACTGGGNNNN ₂ N ⁺	list	881/880
8	BsaWI	W ⁺ CCGG ₂ W	list	785/789
9	BsiHKAII	G ₂ WGCW ⁺ C	list	95/91
10	BspMI	ACCTGCNNNN ⁺ NNNN ₂	list	115/119
11	BspQI	GCTCTTCN ⁺ NNN ₂	list	1027/1030
12	BsrBI	CCG ⁺ CTC	list	*852
13	BstUI	CG ⁺ CG	list	*275
14	BstZ17I	GTA ₂ TAC	list	1380
15	BtgZI	GCGATG(N) ₁₀ NNNN ₂	list	*341/345
16	BtsI	GCAGTG ₂ NN ⁺	list	1093/1091
17	DpnI	GA ₂ TC	list	491
18	DpnII	⁺ GATC ₂	list	#489/493
19	DrdI	GACNN ₂ N ⁺ NGTC	list	1398/1396
20	Eco53ki	GAG ⁺ CTC	list	93
21	EcoRV	GAT ₂ ATC	list	603
22	HincII	GTY ⁺ RAC	list	532
23	HindIII	A ⁺ AGCT ₂ T	list	771/775
24	HphI	GGTGA(N) ₇ N ⁺	list	124/123
25	HpyCH4IV	A ⁺ CG ₂ T	list	*664/666
26	KasI	G ⁺ GCGC ₂ C	list	*257/261
27	MboI	⁺ GATC ₂	list	#489/493
28	MspAII	CMG ⁺ CKG	list	681
29	NarI	GG ⁺ CG ₂ CC	list	*258/260
30	NmeAIII	GCCGAG(N) ₁₉ NN ⁺	list	243/241
31	PfiFI	GACN ⁺ N ₂ NGTC	list	163/164
32	PfiMI	CCAN ₂ NNN ⁺ NTGG	list	445/442
33	PluTI	G ₂ GCGC ⁺ C	list	*261/257
34	PvuII	CAG ⁺ CTG	list	681
35	SacI	G ₂ AGCT ⁺ C	list	95/91
36	SapI	GCTCTTCN ⁺ NNN ₂	list	1027/1030
37	Sau3AI	⁺ GATC ₂	list	489/493
38	SfoI	GGC ⁺ GCC	list	*259
39	SphI	G ₂ CATG ⁺ C	list	1208/1204
40	TaqI	T ⁺ CG ₂ A	list	573/575
41	TfiI	G ⁺ AWT ₂ C	list	1293/1296
42	Tsp45I	⁺ GTSAC ₂	list	112/117
43	Tth111I	GACN ⁺ N ₂ NGTC	list	163/164



#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AccI	GT [↓] MK _↓ AC	list	1130/1132, 1379/1381
2	AlwNI	CAG _↓ NNN [↓] CTG	list	445/442, 1065/1062
3	ApoI	R [↓] AATT _↓ Y	list	511/515, 789/793
4	AvaII	G [↓] GWC _↓ C	list	*#207/210, 503/506
5	BanII	G _↓ RGY [↓] C	list	95/91, 232/228
6	BcoDI	GTCTCN [↓] NNNN _↓	list	63/67, 171/175
7	Bpu10I	CC [↓] TNA _↓ GC	list	977/980, 1298/1301
8	BpuEI	CTTGAG(N) ₁₄ _↓ NN [↓]	list	636/634, 1108/1106
9	BsaHI	GR [↓] CG _↓ YC	list	*258/260, *675/677
10	BsaI	GGTCTCN [↓] NNNN _↓	list	*63/67, 171/175
11	BseRI	GAGGAG(N) ₈ _↓ NN [↓]	list	705/703, 961/959
12	BseYI	C [↓] CCAG _↓ C	list	4/8, 33/37
13	BsmAI	GTCTCN [↓] NNNN _↓	list	63/67, 171/175
14	BsmFI	GGGAC(N) ₁₀ _↓ NNNN _↓	list	*686/690, 794/798
15	Bsp1286I	G _↓ DGCH [↓] C	list	95/91, 232/228
16	BstEII	G [↓] GTNAC _↓ C	list	112/117, 695/700
17	CviQI	G [↓] TA _↓ C	list	198/200, 1312/1314
18	EaeI	Y [↓] GGCC _↓ R	list	540/544, 560/564
19	EarI	CTCTCN [↓] NNN _↓	list	632/635, 1027/1030
20	EcoNI	CCTNN [↓] N _↓ NNAGG	list	58/59, 1069/1070
21	EcoO109I	RG [↓] GNC _↓ CY	list	981/984, 1021/1024
22	HaeII	R _↓ GCGC [↓] Y	list	*261/257, *867/863
23	HgaI	GACGC(N) ₅ [↓] (N) ₅ _↓	list	*683/688, *1050/1055
24	MlyI	GAGTC(N) ₅ _↓	list	133, 1408
25	MscI	TGG [↓] CCA	list	542, 562
26	MseI	T [↓] TA _↓ A	list	656/658, 1373/1375
27	NspI	R _↓ CATG [↓] Y	list	180/176, 1208/1204
28	PleI	GAGTCNNNN [↓] N _↓	list	133/134, 1407/1408
29	PstI	C _↓ TGCA [↓] G	list	152/148, 893/889
30	RsaI	GT [↓] AC	list	199, 1313
31	SfcI	C [↓] TRYA _↓ G	list	148/152, 889/893
32	SmlI	C [↓] TYRA _↓ G	list	615/619, 1123/1127
33	StyI	C [↓] CWWG _↓ G	list	537/541, 668/672

Τέλος, έγινε χρήση του προγράμματος protein workshop της ιστοσελίδας rcsppdb.org, για το προσδιορισμό της κρυσταλλικής δομής και άλλων δομικών πληροφοριών της πρωτεΐνης.



Πέρα απο τη κρυσταλλική δομή, λαμβάνουμε και άλλες πληροφορίες, όπως δηλαδή οτι η πρωτεΐνη αποτελείται απο 2 πανομοιότυπες αλυσίδες, η ανάλυση της είναι 2.21 Angstroms, διαθέτει 98 μόρια νερού γύρω απο το μόριο της πρωτεΐνης και δεν περιέχει ετεροάτομα.