

Δομική Βιοχημεία

Όνομα: Μαρία
Επώνυμο: Κουργιαντάκη
ΑΕΜ: 8959

Τίτλος: BMP (Bone Morphogenetic Protein, Οστική μορφογενετική πρωτεΐνη)

Αναφορά Αποτελεσμάτων

Κωδικός: P12643

Πρωτεΐνη: Bone morphogenetic protein 2 του οργανισμού homo sapiens

Γονίδιο: BMP2

Λειτουργία:

Επάγει τον σχηματισμό των οστών και των χόνδρων. Επιπλέον, δίνει το έναυσμα για την διαφοροποίηση των μυοβλαστών (myoblasts) σε οστεοβλαστούς (osteoblasts) μέσω του EIF2AK3-EIF2A-ATF4 μονοπατιού. Η BMP2 ενεργοποιεί τον EIF2AK3 (μεταφραστικός παράγοντας) και διεγείρεται η φωσφορυλίωση του EIF2A (μεταφραστικός παράγοντας) με αποτέλεσμα την αυξημένη έκφραση του γονιδίου του ATF4 (μεταγραφικός παράγοντας) που παίζει σημαντικό ρόλο στην διαφοροποίηση των οστεοβλαστών. Επίσης, διεγείρει την διαμεμβρανική πρωτεΐνη TMEM119 που ρυθμίζει την έκφραση του ATF4.

FASTA

>sp|P12643|BMP2_HUMAN Bone morphogenetic protein 2 OS=Homo sapiens OX=9606 GN=BMP2 PE=1 SV=1

MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSM
FGLKQRPTPSRDAVVPPYMLDLYRRHSGQPGSPAPDHRLEERAASRANTVRSFHHHEESLEE
LPETSGKTTTRRFFFNLSIPTEEFITSAELQVFREQMQLGNNSSFHHRINIYEI IKPA
TANSKFVTRLLDTRLVNQNASRWESFDVTPAVMRWTAQGHANHG FVVEVAHLEEKQGV
SRHVRISRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFPFLADHLNSTNHAIVQTLVNSVNSKI PKAC
CVPTELSAISMLYLDENEKVVLKNYQDMVVEGCGCR

ΑΠΟ CCDS

Blue: alternating exons.

Red : amino acids encoded across a splice junction.

Nucleotide Sequence (1191 nt):

ATGGTGGCCGGGACCCGCTGTCTTCTAGCGTTGCTGCTTCCCCAGGTCCTCCTGGGCGGCGGGCTGGCC
TCGTTCCGGAGCTGGGCGCAGGAAGTTCGCGGCGGGCTCGTCGGGCGGCCCTCATCCAGCCCTCTGA
CGAGGTCCTGAGCGAGTTCGAGTTGCGGCTGCTCAGCATGTTTCGGCCTGAAACAGAGACCCACCCAGC
AGGGACGCCGTGGTGGCCCCCTACATGCTAGACCTGTATCGCAGGCACTCAGGTCAGCCGGGCTCACC
CCCGAGACCACCGTTGGAGAGGGCAGCCAGCCAGCCAACTGTGCGCAGCTTCCACCATGAAGAATC
TTTGGAAGAACTACCAGAAACGAGTGGGAAAACAACCCGGAGATTCTCTTTAATTTAAGTTCTATCCCC
ACGGAGGAGTTTATCACCTCAGCAGAGCTTCAGGTTTTCCGAGAACAGATGCAAGATGCTTTAGGAAACA
ATAGCAGTTTCCATCACC GAATTAATATTTATGAAATCATAAAACCTGCAACAGCCAACTCGAAATTC
CGTGACCAGACTTTTGGACACCAGGTTGGTGAATCAGAAATGCAAGCAGGTGGGAAAGTTTGTATGTCACC
CCCGCTGTGATGCGGTGGACTGCACAGGGACACGCCAACCATGGATTTCGTGGTGGAGTGGCCACTTGG
AGGAGAAACAAGGTGTCTCCAAGAGACATGTTAGGATAAGCAGGTCTTTGCACCAAGATGAACACAGCTG
GTCACAGATAAGGCCATTGCTAGTAACTTTGGCCATGATGGAAAAGGGCATCCTCTCCACAAAAGAGAA

AAACGTCAAGCCAAACACAAACAGCGGAAACGCCTTAAGTCCAGCTGTAAGAGACACCCTTTGTACGTGG
ACTTCAGTGACGTGGGGTGGGAATGACTGGATTGTGGCTCCCCGGGGTATCACGCCTTTTACTGCCACGG
AGAATGCCCTTTTCCTCTGGCTGATCATCTGAAGTCCACTAATCATGCCATTGTTTCAGACGTTGGTCAAC
TCTGTTAACTCTAAGATTCTTAAGGCATGCTGTGTCCCGACAGAACTCAGTGCTATCTCGATGCTGTACC
TTGACGAGAATGAAAAGTTGTATTAAAGAACTATCAGGACATGGTTGTGGAGGGTTGTGGGTGTCGCTA
G

Translation (396 aa):

MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSMFGLKQRPTPS
RDAVVPPYMLDLYRRHSGQPGSPAPDHRLEERAASRANTVRSFHHESLEELPETSGKTTRRFFFNLS
SIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPATANSKFPVTRLLDTRLVNQNASRWESFDVT
PAVMRWTAQGHANHGFFVEVAHLEEKQGVSKRHVRISRLHQDEHSWSQIRPLLVTFGHDGKGHPLHKRE
KRQAKHKQRKRLKSSCKRHPLYVDFSDVGWNDWIVAPPGYHAFYCHGECPPFLADHLNSTNHAIVQTLVN
SVNSKIPKACCVPTELSAISMLYLDENEKVVLKNYQDMVVEGCGCR

Δεν υπήρχαν δεδομένα για την BMP2 από GenBank και DDBJ.

EXPASY-TRANSLATION

Μετάφραση ολόκληρης της νουκλεοτιδικής αλυσίδης (1191 νουκλεοτίδια):

5'3' Frame 1

MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASS
GRPSSQPSDEVLSEFELRLLSMFGGLKQRPTPSRDAVVPP
YMLDLYRRHSGQPGSPAPDHRLEERAASRANTVRSFHHES
SLEELPETSGKTTRRFFFNLSIPTEEFITS AELQVFRE
QMQDALGNNSSFHHRINIYEIIKPATANSKFPVTRLLDT
RLVNQNASRWESFDVTPAVMRWTAQGHANHGFFVEV
AHLEEKQGVSKRHVRISRLHQDEHSWSQIRPLLVTFG
HDGKGHPLHKREKRQAKHKQRKRLKSSCKRHPLYVDF
SDVGWNDWIVAPPGYHAFYCHGECPPFLADHLNSTNH
AIVQTLVNSVNSKIPKACCVPTELSAISMLYLDENEKV
VLKNYQDMVVEGCGCR Stop

Μετάφραση μπλε αλληλουχίας + 4 νουκλεοτίδια μπροστά:

5'3' Frame 1

ESLEELPETSGKTTRRFFFNLSIPTEEFITS AELQVFR
EQMQDALGNNSSFHHRINIYEIIKPATANSKFPVTRLLD
TRLVNQNASRWESFDVTPAVMRWTAQGHANHGFFVE
VAHLEEKQGVSKRHVRISRLHQDEHSWSQIRPLLVTF
GHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHPLYVD
FSDVGWNDWIVAPPGYHAFYCHGECPPFLADHLNSTN
HAIVQTLVNSVNSKIPKACCVPTELSAISMLYLDENEKV
VLKNYQDMVVEGCGCR Stop

Μετάφραση μόνο της μπλε νουκλεοτιδικής αλληλουχίας:

5'3' Frame 3

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

Red: open reading freams

MULTALIN

Έγινε σύγκριση τεσσάρων αλληλουχιών: 1) ολόκληρη η FASTA , 2) η αλληλουχία αμινοξέων από τα δεδομένης της CCDS , 3) η αλληλουχία αμινοξέων όπως μεταφράστηκε το γονίδιο μέσω της διαδικασίας expasy → proteomics→tools→translate και 4) η αλληλουχία αμινοξέων με μετάφραση μόνο του μπλε τμήματος της νουκλεοτιδικής αλληλουχίας μέσω της διαδικασίας expasy → proteomics→tools→translate

Τα αποτελέσματα έδειξαν πως οι 3 πρώτες αλληλουχίες ταυτίζονται πλήρως ενώ 4^η αλληλουχία ουσιαστικά ξεκινά 116 αμινοξέα μετά τις τρεις πρώτες και από το 117^ο αμινοξύ έως το 396^ο οι τέσσερις αλληλουχίες ταυτίζονται πλήρως.

Alignment data :

Alignment length : 396

Residues conserved for 90 % or more (upper-case letters) : 280 is 70.71 %

Residues conserved for 50 % and less than 90 % (lower-case letters) : 116 is 29.29 %

Residues conserved less than 50 % (white space) : 0 is 0.00 %

IV conserved positions (!) : 0 is 0.00 %

LM conserved positions (\$) : 0 is 0.00 %

FY conserved positions (%) : 0 is 0.00 %

NDQEBZ conserved positions (#): 0 is 0.00 %

Sequence 0001 : FASTAx0 (396 residues).

Sequence 0002 : CCDSxx1 (396 residues).

Sequence 0003 : expasy (396 residues).

Sequence 0004 : bluexx3 (280 residues).

10	20	30	40	50	60	
FASTAx0						
MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSM						
CCDSxx1						
MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSM						
expasy						
MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSM						
bluexx3						

Consensus						
mvagtrcllalllpqvllggaaglvpelgrrkfaaassgrpssqpsdevlsefelrllsm						
Prim.cons.						
MVAGTRCLLALLLPQVLLGGAAGLVPELGRRKFAAASSGRPSSQPSDEVLSEFELRLLSM						

	70	80	90	100	110	
120						
FASTAx0						
FGLKQRPTPSRDAVVPPYMLDLYRRHSGQPGSPAPDHRLERAASRANTVRSFHHEESLEE						
CCDSxx1						
FGLKQRPTPSRDAVVPPYMLDLYRRHSGQPGSPAPDHRLERAASRANTVRSFHHEESLEE						
expasy						
FGLKQRPTPSRDAVVPPYMLDLYRRHSGQPGSPAPDHRLERAASRANTVRSFHHEESLEE						
bluexx3						
-----SLEE						
Consensus						
fglkqrptpsrdavvppymldlyrrhsgqpgspapdhrleraasrantvrsfhheeSLEE						
Prim.cons.						
FGLKQRPTPSRDAVVPPYMLDLYRRHSGQPGSPAPDHRLERAASRANTVRSFHHEESLEE						

	130	140	150	160	170	
180						
FASTAx0						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						
CCDSxx1						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						
expasy						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						
bluexx3						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						
Consensus						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						
Prim.cons.						
LPETSGKTTTRFFFNLSIPTEEFITS AELQVFREQMDALGNNSSFHHRINIYEIIKPA						

	190	200	210	220	230	
240						
FASTAx0						
TANSKFPVTRLLDTRLVNQNASRWESFDVTPAVMRWTAQGHANHG FVVEVAHLEEKQGV S						
CCDSxx1						
TANSKFPVTRLLDTRLVNQNASRWESFDVTPAVMRWTAQGHANHG FVVEVAHLEEKQGV S						
expasy						
TANSKFPVTRLLDTRLVNQNASRWESFDVTPAVMRWTAQGHANHG FVVEVAHLEEKQGV S						

```

bluexx3
TANSKFPVTRLDDTRLVNQNASRWESFDVTPAVMRWTAQGHANHGFFVEVAHLEEKQGV
Consensus
TANSKFPVTRLDDTRLVNQNASRWESFDVTPAVMRWTAQGHANHGFFVEVAHLEEKQGV
Prim.cons.
TANSKFPVTRLDDTRLVNQNASRWESFDVTPAVMRWTAQGHANHGFFVEVAHLEEKQGV

                250      260      270      280      290
300
                |          |          |          |          |
|
FASTAx0
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
CCDSxx1
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
expasy
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
bluexx3
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
Consensus
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP
Prim.cons.
KRHVRI SRSLHQDEHSWSQIRPLLVTFGHDGKGHPLHKREKRQAKHKQRKRLKSSCKRHP

                310      320      330      340      350
360
                |          |          |          |          |
|
FASTAx0
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC
CCDSxx1
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC
expasy
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC
bluexx3
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC
Consensus
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC
Prim.cons.
LYVDFSDVGWNDWIVAPPGYHAFYCHGECFFPLADHLNSTNHAIVQTLVNSVNSKIPKAC

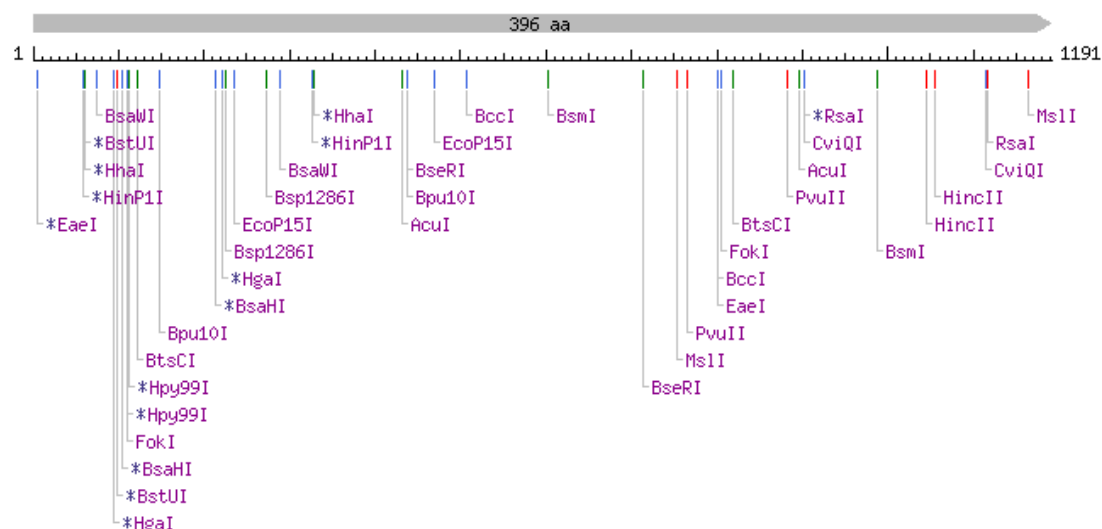
                370      380      390
                |          |          |
FASTAx0      CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR
CCDSxx1      CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR
expasy      CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR
bluexx3      CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR
Consensus    CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR
Prim.cons.    CVPTELSAISMLYLDENEKVVVLKKNYQDMVVEGCGCR

```

ΕΝΔΟΝΟΥΚΛΕΑΣΕΣ ΠΕΡΙΟΡΙΣΜΟΥ

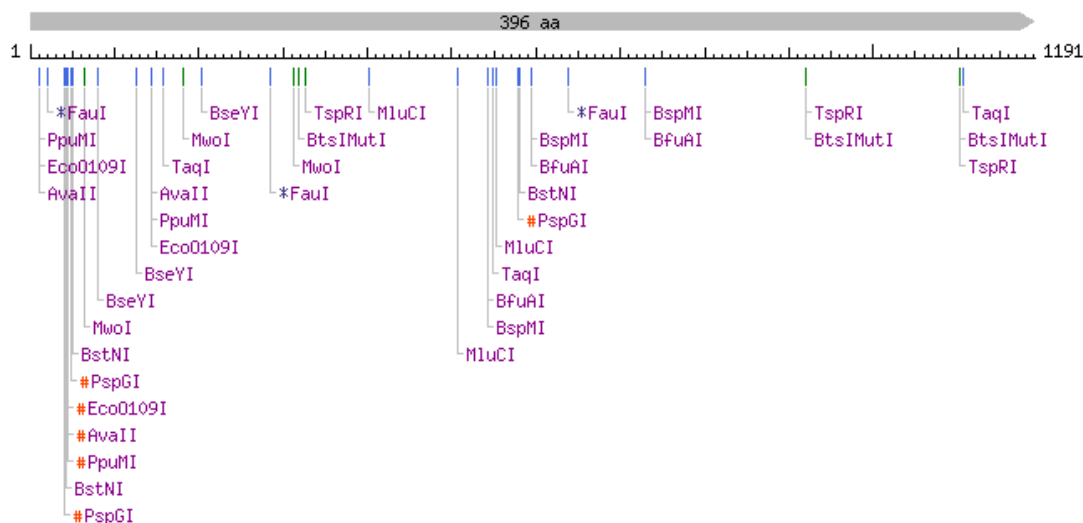
Από το NEBcutter βρίσκουμε ποιες ενδονουκλεάσες και σε ποία/πόσα σημεία κόβουν τη αλληλουχία DNA του γονιδίου BMP2.

Η παρακάτω εικόνα δείχνει τις ενδονουκλεάσες που κόβουν σε 2 σημεία:



#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AcuI	CTGAAG(N) ₁₄ △NN▽	list	433/431, 897/895
2	BccI	CCATCNNNN▽N△	list	509/510, 802/803
3	Bpu10I	CC▽TNA△GC	list	148/151, 438/441
4	BsaHI	GR▽CG△YC	list	*106/108, *215/217
5	BsaWI	W▽CCGG△W	list	75/79, 290/294
6	BseRI	GAGGAG(N) ₈ △NN▽	list	439/437, 715/713
7	BsmI	GAATG△CN▽	list	603/601, 988/986
8	Bsp1286I	G△DGCH▽C	list	227/223, 274/270
9	BstUI	CG▽CG	list	*61, *100
10	BtsCI	GGATG△NN▽	list	124/122, 819/817
11	CviQI	G▽TA△C	list	903/905, 1116/1118
12	EaeI	Y▽GGCC△R	list	*5/9, 801/805
13	EcoP15I	CAGCAG(N) ₂₅ ▽NN△	list	237/239, 470/472
14	FokI	GGATG(N) ₉ ▽NNNN△	list	111/115, 806/810
15	HgaI	GACGC(N) ₅ ▽(N) ₅ △	list	*95/100, *223/228
16	HhaI	G△CG▽C	list	*61/59, *329/327
17	HinP1I	G▽CG△C	list	*59/61, *327/329
18	HincII	GTY▽RAC	list	1047, 1056
19	Hpy99I	△CGWCG▽	list	*111/106, *114/109
20	MslI	CAYNN△NNRTG	list	755, 1165
21	PvuII	CAG▽CTG	list	767, 884
22	RsaI	GT▽AC	list	*904, 1117

Η παρακάτω εικόνα δείχνει τις ενδονουκλεάσες που κόβουν την αλληλουχία σε 3 σημεία:

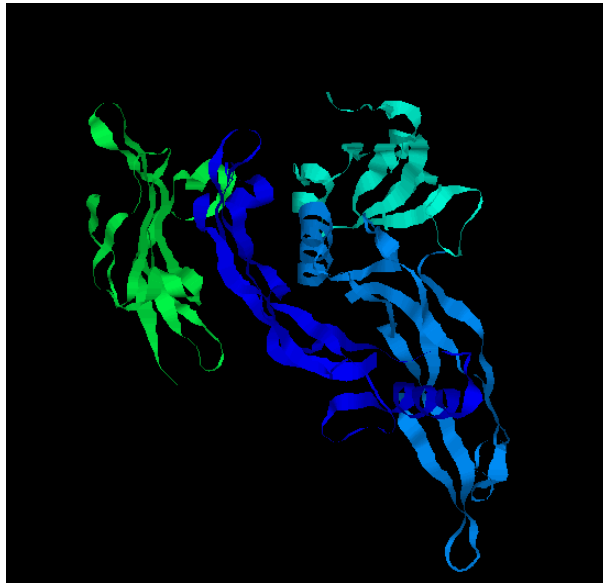


#	Enzyme	Specificity	Sites & flanks	Cut positions (blunt - 5' ext. - 3' ext.)
1	AvaII	G [▼] WGC _▲ C	list	11/14, #45/48, 144/147
2	BfuAI	ACCTGCNNNN [▼] NNNN _▲	list	543/547, 596/600, 731/735
3	BseYI	C [▼] CCAG _▲ C	list	81/85, 128/132, 205/209
4	BspMI	ACCTGCNNNN [▼] NNNN _▲	list	543/547, 596/600, 731/735
5	BstNI	CC [▼] W _▲ GG	list	43/44, 52/53, 582/583
6	BtsIMutI	CAGTG _▲ NN [▼]	list	320/318, 921/919, 1104/1102
7	EcoO109I	RG [▼] GNC _▲ CY	list	11/14, #45/48, 144/147
8	FauI	CCCGCNNNN [▼] NN _▲	list	*22/24, *285/287, *639/641
9	MluCI	[▼] AATT _▲	list	402/406, 509/513, 553/557
10	MwoI	GCNN _▲ NNN [▼] NNGC	list	65/62, 182/179, 313/310
11	PpuMI	RG [▼] WGC _▲ CY	list	11/14, #45/48, 144/147
12	PspGI	[▼] CCWGG _▲	list	#41/46, #50/55, #580/585
13	TaqI	T [▼] CG _▲ A	list	158/160, 550/552, 1108/1110
14	TspRI	_▲ NNCASTGNN [▼]	list	327/318, 921/912, 1104/1095

Από την uniprotKB βρίσκουμε τα PDB entries που είναι διαθέσιμα και επιλέγουμε αυτό με το μικρότερο resolution, για την BMP2 αυτό είναι 1.85Å και ο κωδικός για την PDB είναι 2H62. Στη συνέχεια πηγαίνουμε στη σελίδα RCSB-PDB και κατεβάζουμε το αρχείο .pdb της 2H62 προκειμένου να το χρησιμοποιήσουμε για να πάρουμε απεικόνιση της πρωτεΐνης στο RasMol.

Στη πραγματικότητα η BMP2 είναι μια μεγάλη διμερής πρωτεΐνη ενώ τα αρχεία της PDB αναφέρονται μόνο σε τμήμα της πρωτεΐνης που επιτεύχθηκε η X-ray ανάλυση του.

Στο RasMol παίρνουμε διάφορες εικόνες για τη δομή της 2H62. Παρακάτω δίνονται απεικονίσεις με χρωματισμό ανά αλυσίδα και έπειτα με χρωματισμό με βάση τη δευτεροταγή δομή.

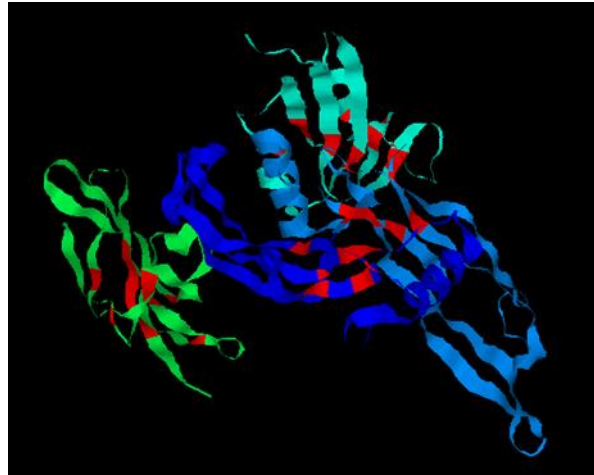


Από τη δεύτερη απεικόνιση βρίσκουμε ότι έχει 8 έλικες και 26 β-φύλλα.

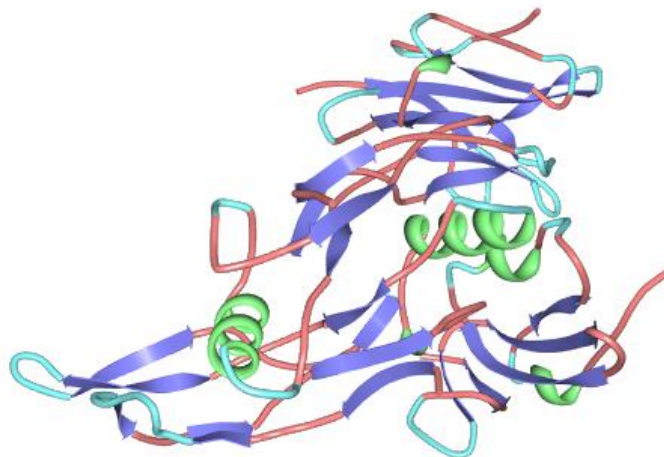
Τέλος, από το RasMol command line δίνουμε εντολές για συγκεκριμένες απεικονίσεις/επισημάνσεις. Για παράδειγμα στη παρακάτω εικόνα με τις εντολές:

- Select cys
- Color red

Χρωματίζονται και οι 304 κυστεΐνες της πρωτεΐνης σε κόκκινο χρώμα.



Ακόμα, 3D απεικονίσεις για την 2H62 μπορούμε να πάρουμε και μέσα από την σελίδα της PDB από την επιλογή Protein Workshop και κάνοντας Launch RCSB-protein workshop. Η παρακάτω εικόνα δίνει την απεικόνιση της 2H62 με βάση τη 2^ο ταγή δομή όπως δίνεται από το protein workshop.



Η απεικόνιση αυτή φαίνεται να διαφέρει λίγο από αυτήν του RasMol, για παράδειγμα εδώ φαίνονται 6 έλικες. Οι διαφορές οφείλονται στο ότι η κάθε εφαρμογή μπορεί να βασιστεί σε διαφορετικές πηγές. Πιο συγκεκριμένα, στη σελίδα RCSB-PDB και στην ενότητα Sequence δίνονται πληροφορίες για την 2^ο ταγή δομή τόσο με βάσει τους συγγραφείς όσο και με βάση την DSSP. Ενώ οι συγγραφείς απαριθμούν 6 έλικες στις 4 αλυσίδες, η DSSP δίνει σύνολο 9 ελίκων για τις 4 αλυσίδες. Μια 3^η πηγή, η STRIDE δίνει σύνολο 8 ελίκων. Παρόμοιες αποκλίσεις παρατηρούνται για τον αριθμό των β-φύλλων. Βλέπουμε δηλαδή μια ποικιλία προβλέψεων για τη δευτεροταγή δομή.