

Structural analysis and functional prediction of suspect *Mus musculus* integrin

Group 11

Gao peng

Han guangwei

Wang haiming

Lang yifei



Structural analysis and functional prediction of suspect *Mus musculus integrin*

- Background
- Primary Structure Analysis
- Structure Prediction
- Function Analysis



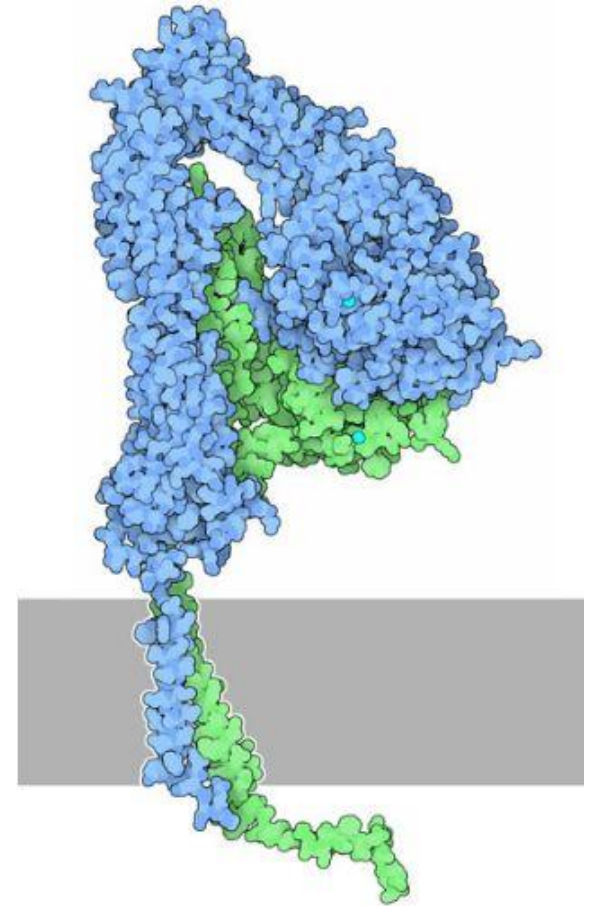
Background

Integrins are receptors that mediate the attachment between a cell and the tissues that surround it, such as other cells or the extracellular matrix (ECM). In signal transduction, integrins pass information about the chemical composition of the ECM into the cell. Therefore, they are involved in cell signaling and the regulation of cell cycle, shape, and motility.



Background

Integrins are ubiquitous membrane receptor family which includes 18 different α subunits and 8 β subunits forming more than 20 $\alpha\beta$ heterodimers. Integrin has a lot of ligands include such as several bacterial and viral proteins. Integrins play a key role in vascular endothelial cell and tumour cell adhesion, lymphocyte transport, tumor growth and infection.



Primary Structure Analysis

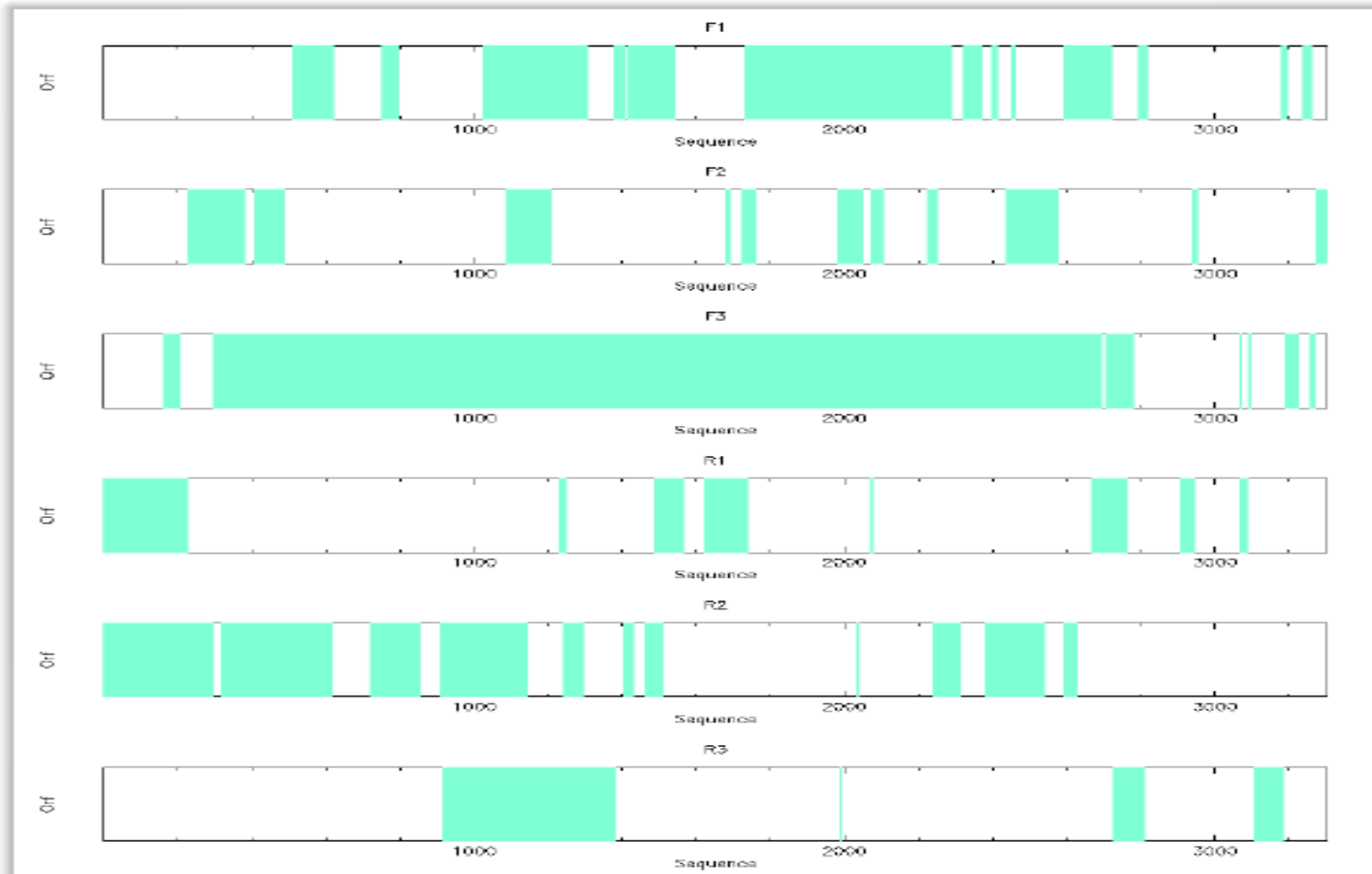
The original cDNA sequence(3304bp):

TGGCGGCCGCAGGAGAGGGGGGAGGAGAAAAGCAGAGAAGAGTGGGGGCTGGCGGCCGAGGAAAAAGGAAGGGTCTCCGAGAGTAGACGCGCGGGG
GACTCGGCGAGGGTGCCTCCGAGCAGCGACAACCTCCGAGCGTCCCAGCGGGCCAGCGAGGAGGATGGTGGCGGCCGGGCGCGGACCAGCCCCGGCCGCG
GGCGCCGTGAGCCGGAGCGCAGCGCCCCGCATGCGGGCTGCGGTCCCCGGCCTCGGCCCCGCTCCGCCCCCGCGAGCGCCCCAGCCGAGCGGCGCGCATC
ATGCCGCGGGTGCCCGCGACCCCTCTACGCCTGTCTGCTCGGGCTCTGCGCGCTCGTTCCGCGCCTCGCAGGGCTCAACATATGCACTAGTGGAAGTGCCAC
CTCGTGTGAAGAATGCCTGTTGATCCACCCAAAATGTGCCTGGTGTCTCAAAGAGTACTTTGGCAATCCACGGTCCATCACCTCTCGGTGTGACCTGAAGG
CAAACCTCATCCGGAATGGCTGTGAAGGTGAGATTGAGAGTCCAGCCAGCAGCACCCACGTCTCCGGAACCTACCTCTCAGCAGCAAGGGTTCAGTGC
CACGGGCTCTGACGTCATCCAGATGACGCCCGAGGAGATTGCACTGAGCCTCCGGCCAGGCGAGCAGACTACGTTCCAGCTGCAGGTGCGCCAGGTGGAG
GACTACCCTGTAGACCTGTACTACCTGATGGACCTCTCCCTCTCCATGAAGGATGACTTGGAGAACATCCGGAGCCTGGGCACCAAGCTTGGCGAGGAAA
TGAGGAAGCTCACTAGTAACTTCCGCTTAGGTTTCGGGTCTTTTGTGACAAGGACATCTCTCCTTCTCCTACACGGCACCGAGATAACCAGACCAATCCGT
GTATTGGTTACAAGTTATTCCCAACTGCGTCCCCTCCTTCGGGTTCCGGCATCTGCTGCCTCTCACAGACAGAGTCGACAGCTTCAACGAGGAAGTGAGG
AAGCAGAGGGTGTCCCGAACCAGATGCCCCGAGGGGGGTTTGTGCGGTCTCCAGGCTGCTGTCTGCAAGGAGAAGATCGGATGGCGAAAAGAT
GCTCTGCACTTGCTGGTGTTCACAACAGACGATGTGCCCCACATCGCACTGGATGGAAAAGTGGGTGGCCTGGTCCAGCCCCACGATGGCCAGTGTACCT
GAATGAAGCCAATGAGTACACAGCCTTAACCAGATGGACTATCCATCGCTTGCCTTGCCTTGGGGAGAAGCTGGCAGAGAACAATATCAACCTCATTTTG
CTGTGACGAAGAACCCTATATGCTCTACAAGAATTTTACAGCCCTGATACCTGGAACCACTGTGGAGATTTTGCATGGAGATTTCAAAAATATTATTCAA
CTGATTATCAATGCGTACAGTAGCATCCGGGCTAAAGTGGAGCTGTCACTGTGGGATCAGCCAGAAGACCTTAAATCTCTTCTTCACTGCCACCTGCCAAGA
TGGCATATCTTACCCTGGTCAGAGGAAGTGTGAGGGTCTGAAGATTGGGGACACGGCATCCTTTGAAGTGTCCGTGGAGGCTCGGAGCTGCCCGGCAG
CAAGCAGCACAGTCTTTCACCTTGAGGCCCGTGGGCTCCCGGACAGTCTGCAAGTGGAAGTCGCCTACAATTGCACATGCGGCTGTAGCACGGGGCTGG
AGCCCAACAGTGCCAGATGCAGTGGGAATGGAACATACACCTGTGGGCTGTGCGAGTGTGACCCCGCTACCTGGGCACTAGGTGCGAGTGCCAGGAGG
GGGAGAACCAGAGCGGGTACCAGAACCCTGTGCCGGGAGGCAGAGGGCAAGCCTCTGTGCAGCGGGCGTGGAGAGTGTAGCTGCAACCAGTGCTCCTGCT
TCGAGAGTGAGTTCGGGAGGATCTACGGACCTTTCTGCGAGTGTGACAGCTTTTCTGTGCCAGAAACAAGGGCGTCTATGCTCAGGCCATGGAGAGTGT
CACTGTGGAGAATGCAAATGCCACGCAGGTTACATTGGGGACAATTGTAAGTGTCTCAACAGACGTCAGCACATGCAGGGCCAAGGATGGGCAGATCTGCA
GTGACCGAGGCCGTTGTGTCTGTGGCAGTGCCAGTGACAGAGCCTGGAGCCTTGGGGAGACGTGTGAGAAGTGCCCAACCTGCCCGGATGCTTGCAG
CTCTAAGAGAGACTGTGTGAATGCTTGTACTTACCAGGGGAAAACCTGACAACCAGACCTGCCACCACAGTGCAAAAGATGAGGTGATCACGTGGGTA
GACACCATCGTCAAAGATGACCAGGAGGCTGTGCTTTGCTTCTACAAAACCTGCTAAGGACTGCGTTATGATGTTTCACTACACAGAAGTGCCTAATGGGA
GGTCCAACCTTGACGGTCTCCGGGAGCCAGAATGTGGAAGTGCCTTCAAGTGCCTTCAACAAGTGCCTTCAACAAGTGCCTTCAACAAGTGCCTTCAACA
AAACCCCTGTACAGAAAGCCACTCTCCACACACTGTGATTTGCGCTTCAACAAGTGCCTTCAACAAGTGCCTTCAACAAGTGCCTTCAACAAGTGCCT
TGGCTGGAGGGGACTAAGGATGAAGACTCTGGCGTGCCTTGGACTTCTGGACATTTGCTCAGCTAGCTAGGCACGCACGGATAATGGAGATGCCCT
CCATTGAGCCCTAAGGGACCTGGTAGCCACACAGCGGGCCACAGGCACTTGGGGCACTTCCCTCCAAGCCAGGGAAGCAAGGAGACTCTGGTGTCTC
AGCTTCCCCTCTGCCGCTCCAGCTTGTGTCTC



Primary Structure Analysis

The CDS analysis by tool PlotORF



The right
CDS
prediction



Primary Structure Analysis

The CDS sequence (2399bp):

TGGCGGCCCGCAGGAGAGGGGGGAGGAGAAAAGCAGAGAAGAGTGGGGGCTGGCGGCCGAGGAAAAAGGAAGGGTCTCCGAGAGTAGACGCGCGGGG
GACTCGGCGAGGGTGCCTCCGGAGCAGCGACAACCTCCGAGCGTCCCAGCGGGCCAGCGAGGAGGATGGTGGCGGCCGGGCGCGGACCAGCCCCGGCCGCG
GGCGCCGTGAGCCGGAGCGCAGCGCCCCGGCATGCGGGCTGCGGTCCCCGGCCTCGGCCCCGCTCCGCCCCCGCGAGCGCCCCAGCCGAGCGGCGCGCATC
ATGCCGCGGGTGCCCGCGACCCCTCTACGCCTGTCTGCTCGGGCTCTGCGCGCTCGTTCCGCGCCTCGCAGGGCTCAACATATGCACTAGTGGAAGTGCCAC
CTCGTGTGAAGAATGCCTGTTGATCCACCCAAAATGTGCCTGGTGTCTCAAAGAGTACTTTGGCAATCCACGGTCCATCACCTCTCGGTGTGACCTGAAGG
CAAACCTCATCCGGAATGGCTGTGAAGGTGAGATTGAGAGTCCAGCCAGCAGCACCCACGTCTCCGGAACCTACCTCTCAGCAGCAAGGGTTCCAGTGC
CACGGGCTCTGACGTCATCCAGATGACGCCGAGGAGATTGCAAGTGCAGCTCCGGCCAGGCGAGCAGACTACGTTCCAGCTGCAGGTGCGCCAGGTGGAG
GACTACCCTGTAGACCTGACTACCTGATGGACCTCTCCCTCTCCATGAAGGATGACTTGGAGAACATCCGGAGCCTGGGCACCAAGCTTGGCGAGGAAA
TGAGGAAGCTCACTAGTAACTTCCGCTTAGGTTTCGGGTCTTTTGTGACAAGGACATCTCTCCTTTCTCTACACGGCACCGAGATAACCAGACCAATCCGT
GTATTGGTTACAAGTTATCCCCAACTGCGTCCCCTCTTCGGGTTCCGGCATCTGCTGCCTCTCACAGACAGAGTCGACAGCTTCAACGAGGAAGTGAGG
AAGCAGAGGGTGTCCCGAACCAGAGATGCCCCGAGGGGGGGTTTGTGCGGTCTCCAGGCTGCTGTCTGCAAGGAGAAGATCGGATGGCGAAAAGAT
GCTCTGCACTTGTGTGTTACAACAGACGATGTGCCCCACATCGCACTGGATGGAAAACCTGGGTGGCCTGGTCCAGCCCCACGATGGCCAGTGTACCT
GAATGAAGCCAATGAGTACACAGCCTTAACCAGATGGACTATCCATCGCTTGCCTTGCTTGGGGAGAAGCTGGCAGAGAACAATATCAACCTCATTTTTG
CTGTGACGAAGAACCCTATATGCTCTACAAGAATTTACAGCCCTGATACCTGGAACCACTGTGGAGATTTTGCATGGAGATTCCAAAAATATTATTCAA
CTGATTATCAATGCGTACAGTAGCATCCGGGCTAAAGTGGAGCTGTCAAGTGTGGGATCAGCCAGAAGACCTTAATCTCTTCTTACTGCCACCTGCCAAGA
TGGCATATCTTACCCTGGTCAGAGGAAGTGTGAGGGTCTGAAGATTGGGGACACGGCATCCTTTGAAGTGTCCGTGGAGGCTCGGAGCTGCCCCGGCAGA
CAAGCAGCACAGTCTTACCTTGAGGCCCGTGGGCTTCCGGGACAGTCTGCAGGTGGAAGTCGCCTACAATTGCATGCGGCTGTAGCAGGGGCTGG
AGCCCAACAGTGCCAGATGCAGTGGGAATGGAACATACACCTGTGGGCTGTGCGAGTGTGACCCCGGCTACCTGGGCACTAGGTGCGAGTGCCAGGAGG
GGGAGAACCAGAGCGGGTACCAGAACCCTGTGCCGGGAGGCAGAGGGCAAGCCTCTGTGCAGCGGGCGTGGAGAGTGTAGCTGCAACCAGTGCTCCTGCT
TCGAGAGTGAGTTCGGGAGGATCTACGGACCTTTCTGCGAGTGTGACAGCTTTTCTGTGCCAGAAACAAGGGCGTCTATGCTCAGGCCATGGAGAGTGT
CACTGTGGAGAATGCAAATGCCACGCAGGTTACATTGGGGACAATTGTAAGTGTCAACAGACGTCAGCACATGCAGGGCCAAGGATGGGCAGATCTGCA
GTGACCGAGGCCGTTGTGTCTGTGGGCAAGTGCAGTGCACAGAGCCTGGAGCCTTTGGGGAGACGTGTGAGAAGTGCCCAACCTGCCCGGATGCTTGCAG
CTCTAAGAGAGACTGTGTGAATGCTTGTACTTACCAGGGGAAACCTGACAACCAGACCTGCCACCACCAGTGCAAAGATGAGGTGATCACGTGGGTA
GACACCATCGTCAAAGATGACCAGGAGGCTGTGCTTTGCTTCTACAAAACCTGCTAAGGACTGCGTTATGATGTTACAGCTACACAGAAGTCCCAATGGGA
GGTCCAACCTGACGCTCCCGGAGCCAGTGTGGAAGTGCACCCCAATGCCATGCCATCTGCTGGTGTGGTTGGCAGCATCCTCTGATTGGGATG
GCACTCTTGCCATCTGGAAGTGTCTGCTACCATCCACAGACCCGAGAGTTTGGCAAGTTCCTAAGTGTGAGCGCTCCAGGGCCCGTTATGAAATGGCCTC
AAACCCCTGTACAGAAAGCCATCTCCACACACACTGTCGATTTGCTTCAACAAGTTCACAAAATCCTACAATGGCTCAGTGGACTGAGGCTCCTGGA
TGGCTGGAGGGGACTAAGGATGAAGACTCTGGCGTGCCTTGGACTTCTGGACCATTTGCTCACGCTAGCTAGGCACGCACGGATAATGGAGATGCCCT
CCATTGAGCCCTAAGGGACCTGGTAGCCACACAGCGGGCCACAGGCACTTGGGGCCACTTCCCTCCAAGCCAGGGAAGCAAGGAGACTCTGGTGTCTC
AGCTTCCCCTCTGCCGCTCCAGCTTGTGTCTC



Primary Structure Analysis

The corresponding AA sequence:

MPRVPATLYACLLGLCALVPRLAGLNICTSGSATSCEECLLIHPKCAWCSKEYFGNPRSI
TSRCDLKANLIRNGCEGEIESPASSTHVLRNLPLSSKGSSATGSDVIQMTQPQEIIVSLRPG
EQTTFQLQVRQVEDYPVDLYYLMDSL SMKDDLENIRSLGTKLAEEMRKLTSNFR LGF
GSFVDKDISPFSYTAPRYQTNPCIGYKLF PNCVPSFGFRHLLPLTDRVDSFNEEVRKQRV
SRNRDAPEGGFDAVLQAAVCKEKIGWRKDALHLLVFTTDDVPHIALDGKLGGLVQPH
DGQCHLNEANEYTASNQMDYPSLALLGEKLAENNINLIFAVTKNHYMLYKNFTALIPG
TTVEILHGDSKNIIQLIINAYSSIRAKVELSVWDQPEDLNLFFTATCQDGISYPGQRKCEG
LKIGDTASFEVSVEARSCPGRQAAQSFTLRPVGFRDSLQVEVAYNCTCGCSTGLEPNSA
RCSGNGTYTCGLCECDPGYLGTRCECQEGENQSGYQNLCREAEGKPLCSGRGECSCN
QCSCFESEFGRIYGPFCEDSFSCARNKGVLC SGHGEGCHCGECKCHAGYIGDNCNCST
DVSTCRAKDGQICSDRGRVCVGCQCCTEPGAFGETCEKCPTCPDACSSKRDCVECLL
HQGKPDNQTCHHQCKDEVITWVDTIVKDDQEAVLCFYKTAKDCVMMFSYTELPNGR
SNLTVLREPECGSAPNAMTILLAVVGSILLIGMALLAIWKLLVTIHDRREFAKFQSERSR
ARYEMASNPLYRKPISTHTVDFAFNKFNKSYNGSVD



Primary Structure Analysis

BLAST result (BLAST X)

The result shows the target sequence has a high homology with integrin beta-5.

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NP_001139356.1	integrin beta-5 isoform 1 precursor [Mus musculus]	1467	1467	99%	0.0	100%	UGM
NP_671480.2	integrin, beta 5 precursor [Rattus norvegicus]	1439	1439	99%	0.0	97%	UGM
NP_034710.2	integrin beta-5 isoform 2 precursor [Mus musculus]	1395	1460	99%	0.0	100%	UGM
XP_003495486.1	PREDICTED: integrin beta-5-like [Cricetulus griseus]	1386	1386	97%	0.0	97%	GM
NP_002204.2	integrin beta-5 precursor [Homo sapiens]	1373	1373	99%	0.0	92%	UGM
NP_001253135.1	integrin beta-5 precursor [Macaca mulatta]	1372	1372	99%	0.0	92%	G
XP_002716465.1	PREDICTED: integrin, beta 5 [Oryctolagus cuniculus]	1364	1364	99%	0.0	91%	UGM
XP_002758770.1	PREDICTED: integrin beta-5 [Callithrix jacchus]	1351	1351	99%	0.0	90%	GM
NP_777104.1	integrin beta-5 precursor [Bos taurus]	1346	1346	99%	0.0	91%	UGM
NP_001233598.1	integrin beta-5 precursor [Sus scrofa]	1343	1343	99%	0.0	91%	UG



Primary Structure Analysis

BLAST result (BLAST P)

We basically got a same result with the BLAST X.

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
NP_001139356.1	integrin beta-5 isoform 1 precursor [Mus musculus]	1731	1731	100%	0.0	100%	UGM
NP_671480.2	integrin, beta 5 precursor [Rattus norvegicus]	1698	1698	100%	0.0	97%	UGM
NP_034710.2	integrin beta-5 isoform 2 precursor [Mus musculus]	1644	1644	94%	0.0	100%	UGM
XP_003495486.1	PREDICTED: integrin beta-5-like [Cricetulus griseus]	1639	1639	97%	0.0	97%	GM
XP_002716465.1	PREDICTED: integrin, beta 5 [Oryctolagus cuniculus]	1601	1601	100%	0.0	91%	UGM
NP_002204.2	integrin beta-5 precursor [Homo sapiens]	1588	1588	100%	0.0	92%	UGM
NP_777104.1	integrin beta-5 precursor [Bos taurus]	1586	1586	100%	0.0	91%	UGM
NP_001253135.1	integrin beta-5 precursor [Macaca mulatta]	1583	1583	100%	0.0	92%	G
NP_001233598.1	integrin beta-5 precursor [Sus scrofa]	1580	1580	100%	0.0	91%	UG
XP_003275621.1	PREDICTED: integrin beta-5-like [Nomascus leucogenys]	1577	1577	97%	0.0	92%	GM



Primary Structure Analysis

Multiple alignment results

The result shows there are obviously homology between these sequences and our target sequence.

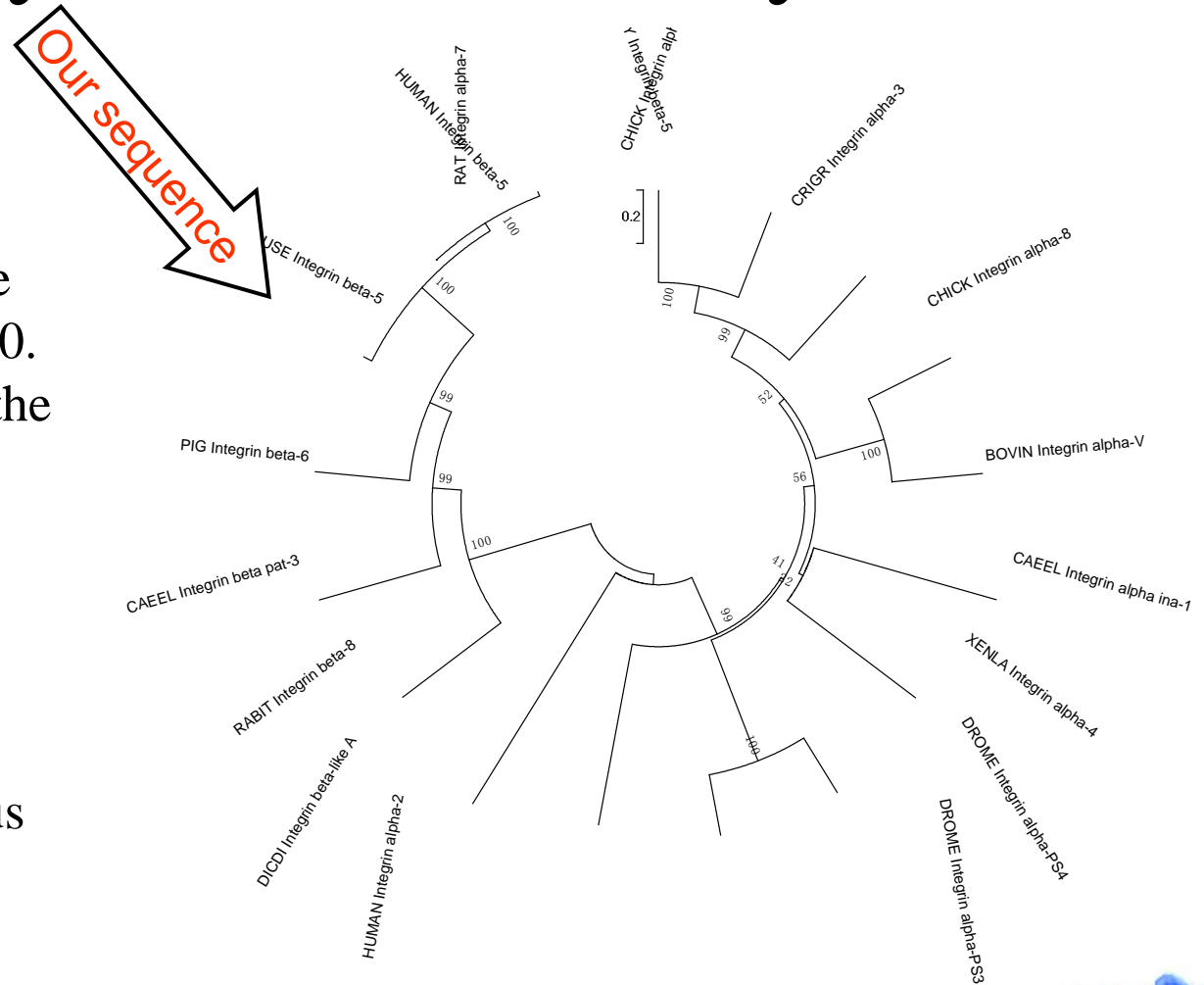
Date of job execution	Jun 7, 2012
Running time	37.9 seconds
Identical positions	621
Identity	62.664%
Similar positions	99
Program	clustalo



Primary Structure Analysis

The neighbor-joining tree constructed by MEGA 5.0. Our target sequence has the highest homology with homo sapiens integrin beta5.

Now we are pretty much sure that our target sequence is mus musculus integrin beta-5.



Primary Structure Analysis

The basic information of this peptide

Molecular weight = 88051.79

Residues = 799

Average Residue Weight = 110.202

Isoelectric Point = 5.9701

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	277	34.668
Small	(A+B+C+D+G+N+P+S+T+V)	438	54.819
Aliphatic	(A+I+L+V)	200	25.031
Aromatic	(F+H+W+Y)	78	9.762
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	430	53.817
Polar	(D+E+H+K+N+Q+R+S+T+Z)	369	46.183
Charged	(B+D+E+H+K+R+Z)	191	23.905
Basic	(H+K+R)	97	12.140
Acidic	(B+D+E+Z)	94	11.765

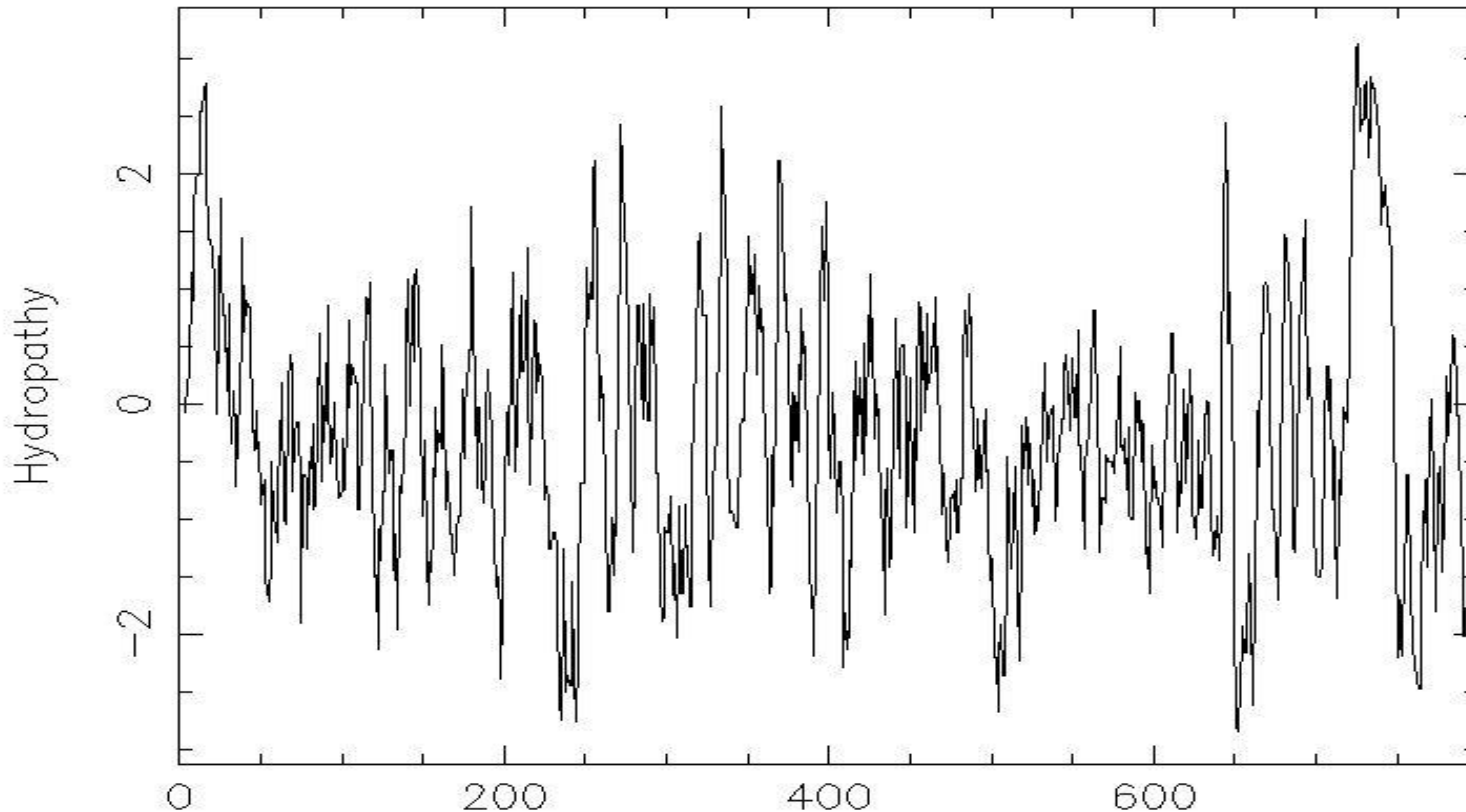


Primary Structure Analysis

Hydropathy analysis by pepwindow

The result shows this peptide has a variable hydropathy.

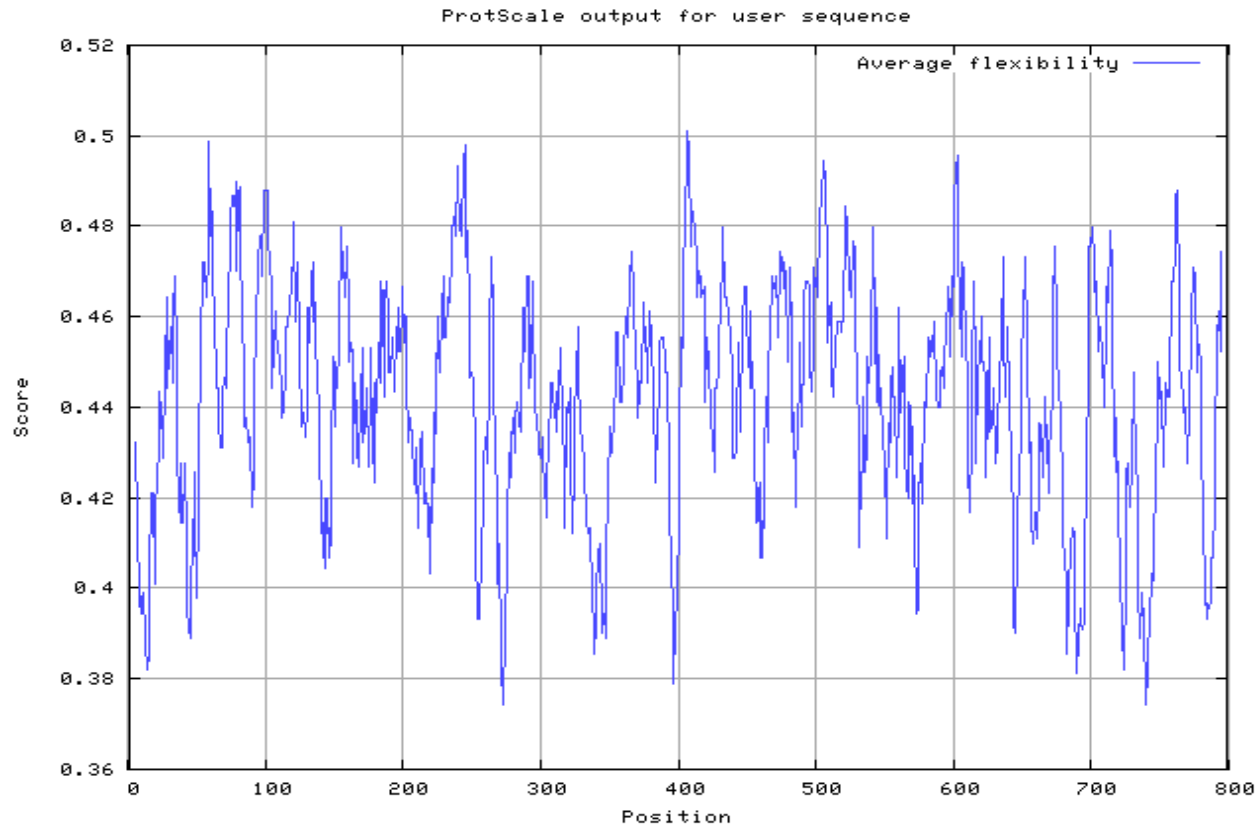
Kyte-Doolittle Plot



Primary Structure Analysis

The flexibility analysis of our target peptide

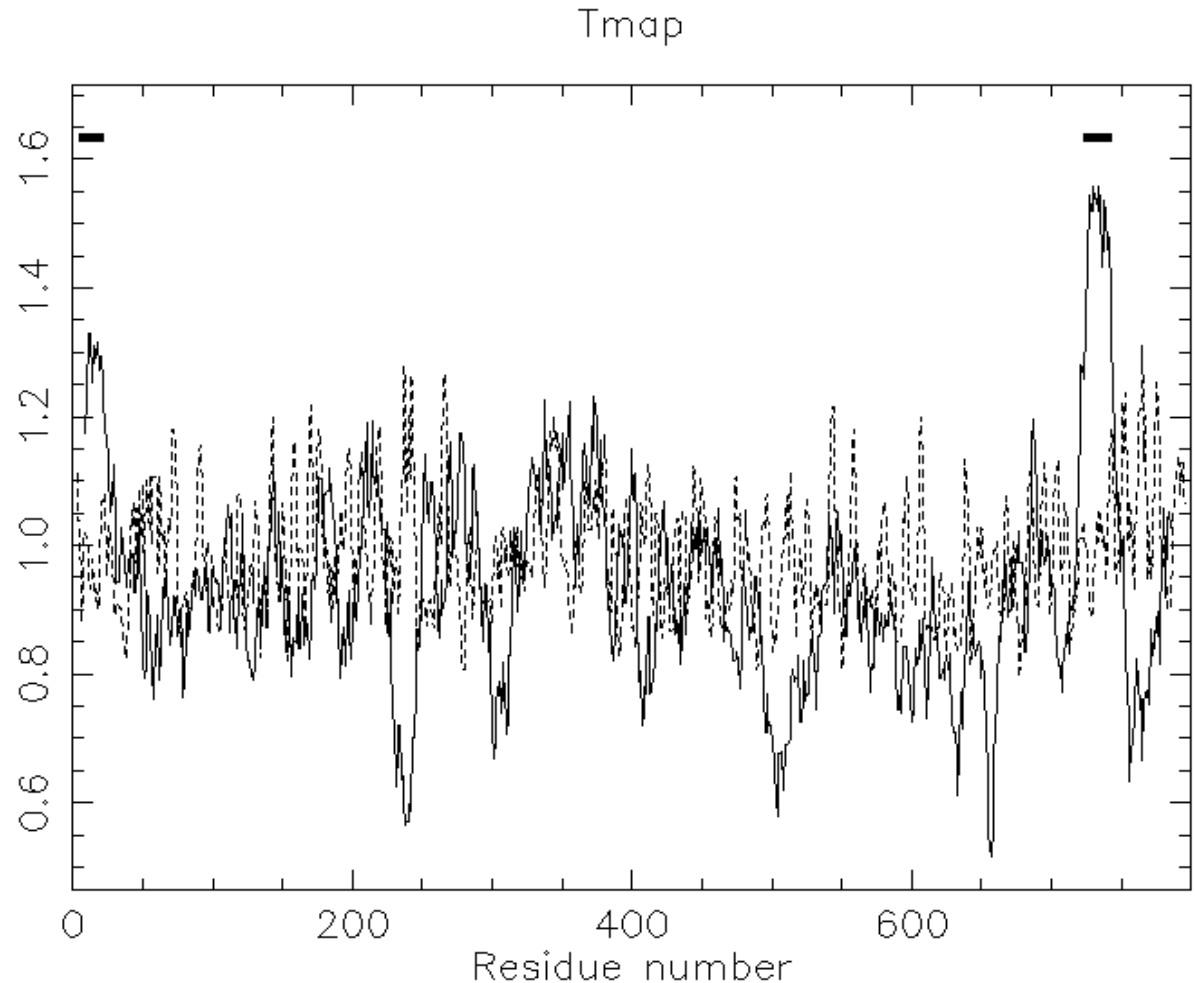
The region with higher flexibility may combine the secondary structure.



Primary Structure Analysis

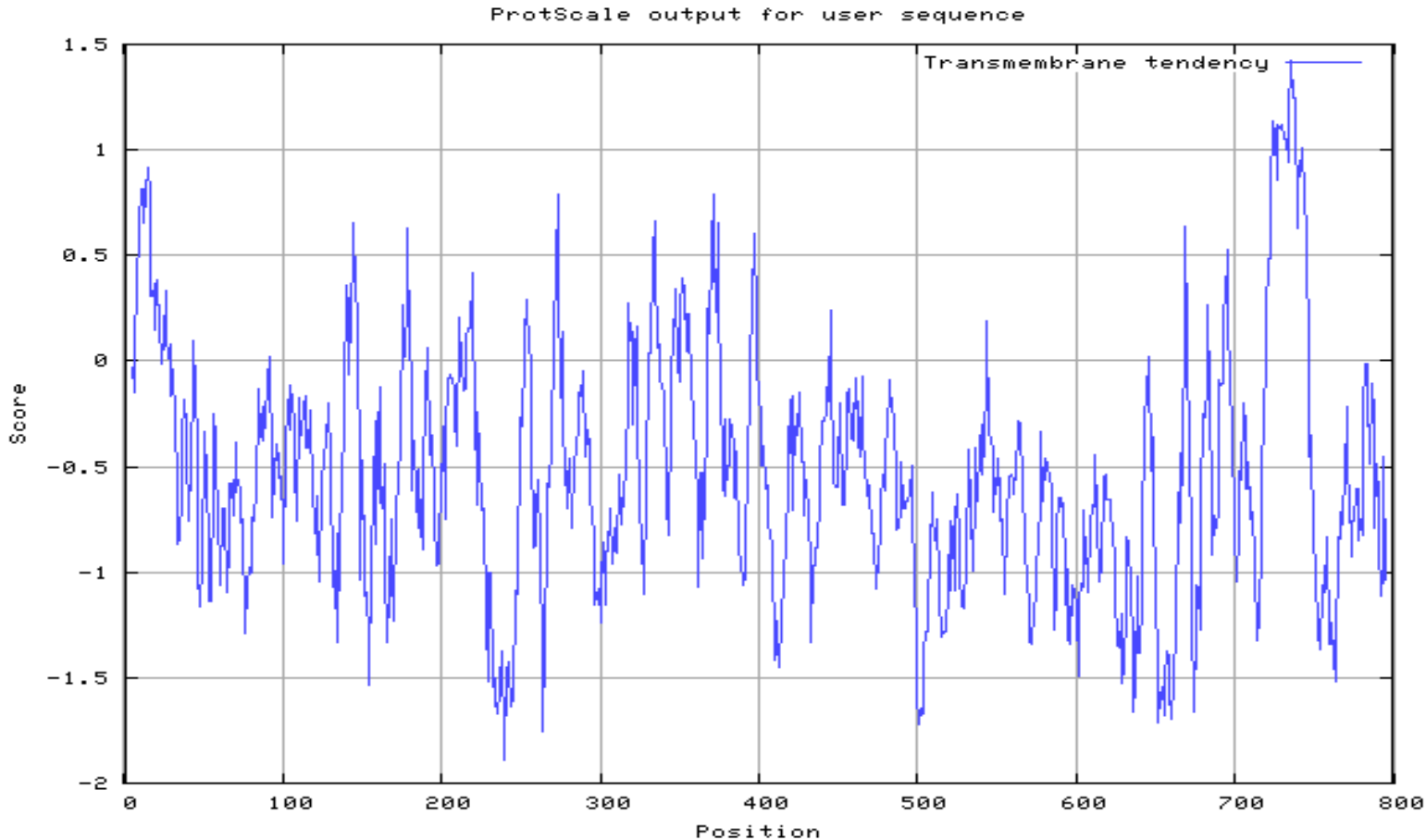
The Transmembrane analysis of our target sequence (Tmap)

The result shows there may exist one major transmembrane regions.



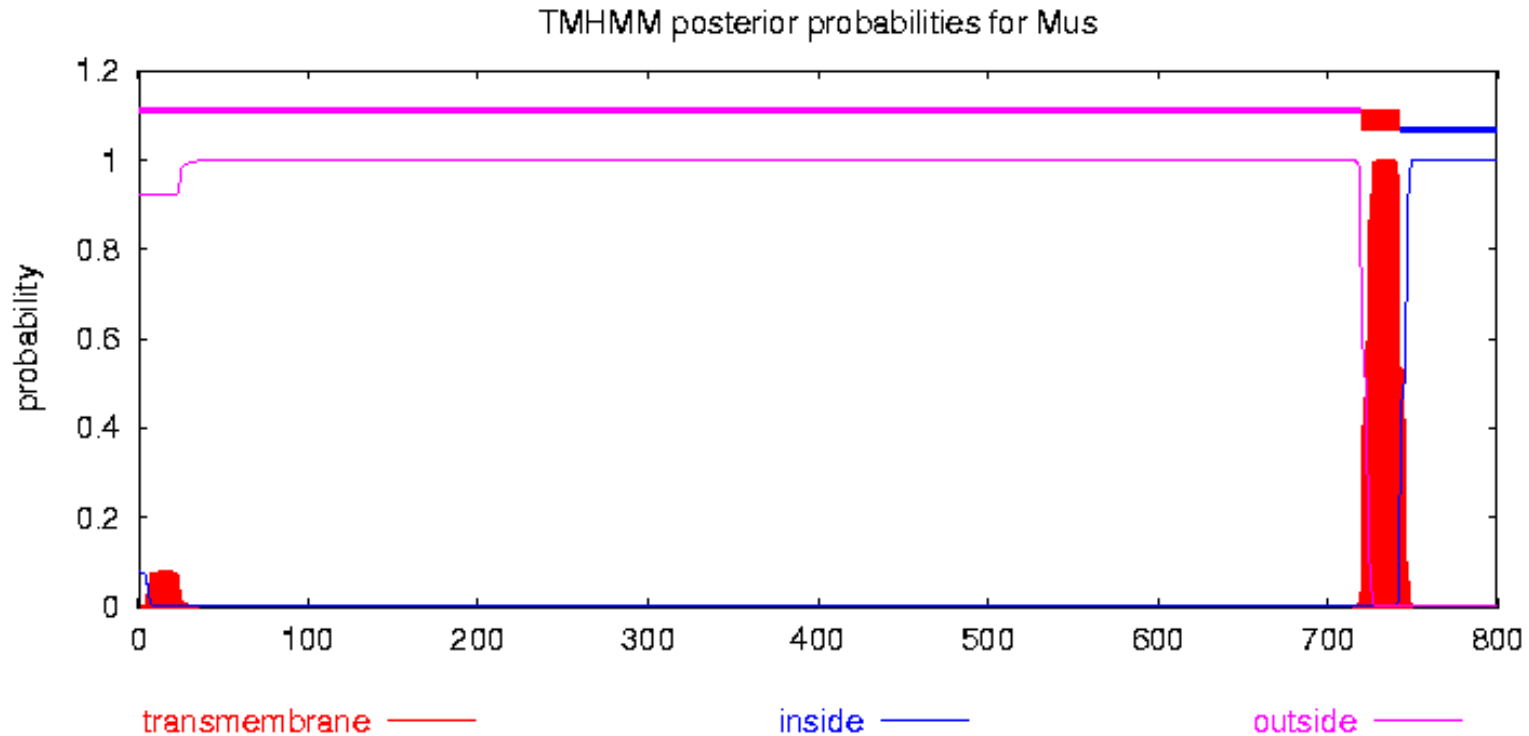
Primary Structure Analysis

The transmembrane analysis from ExPASy and TMHMM shows similar results.



Primary Structure Analysis

The result of TMHMM

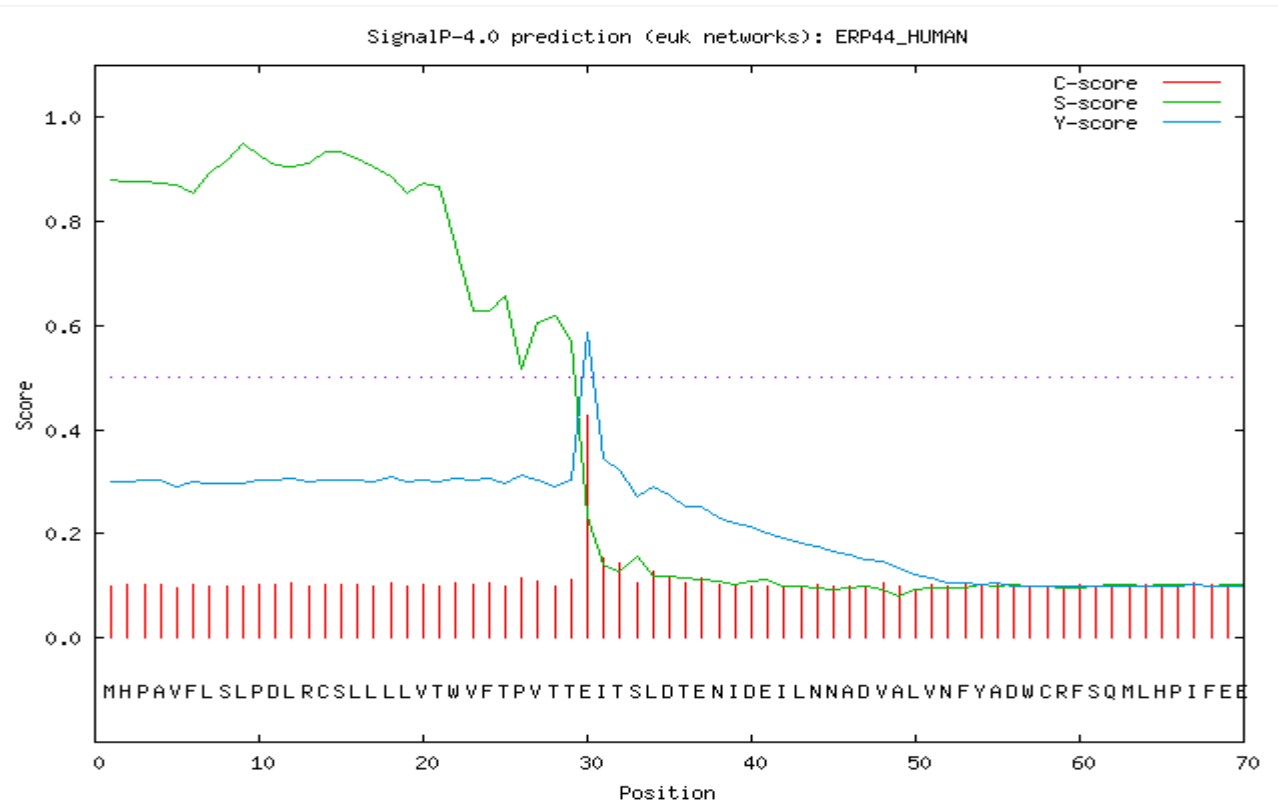


Primary Structure Analysis

The signal-P analysis of our target peptide

The signal-P sequence is :

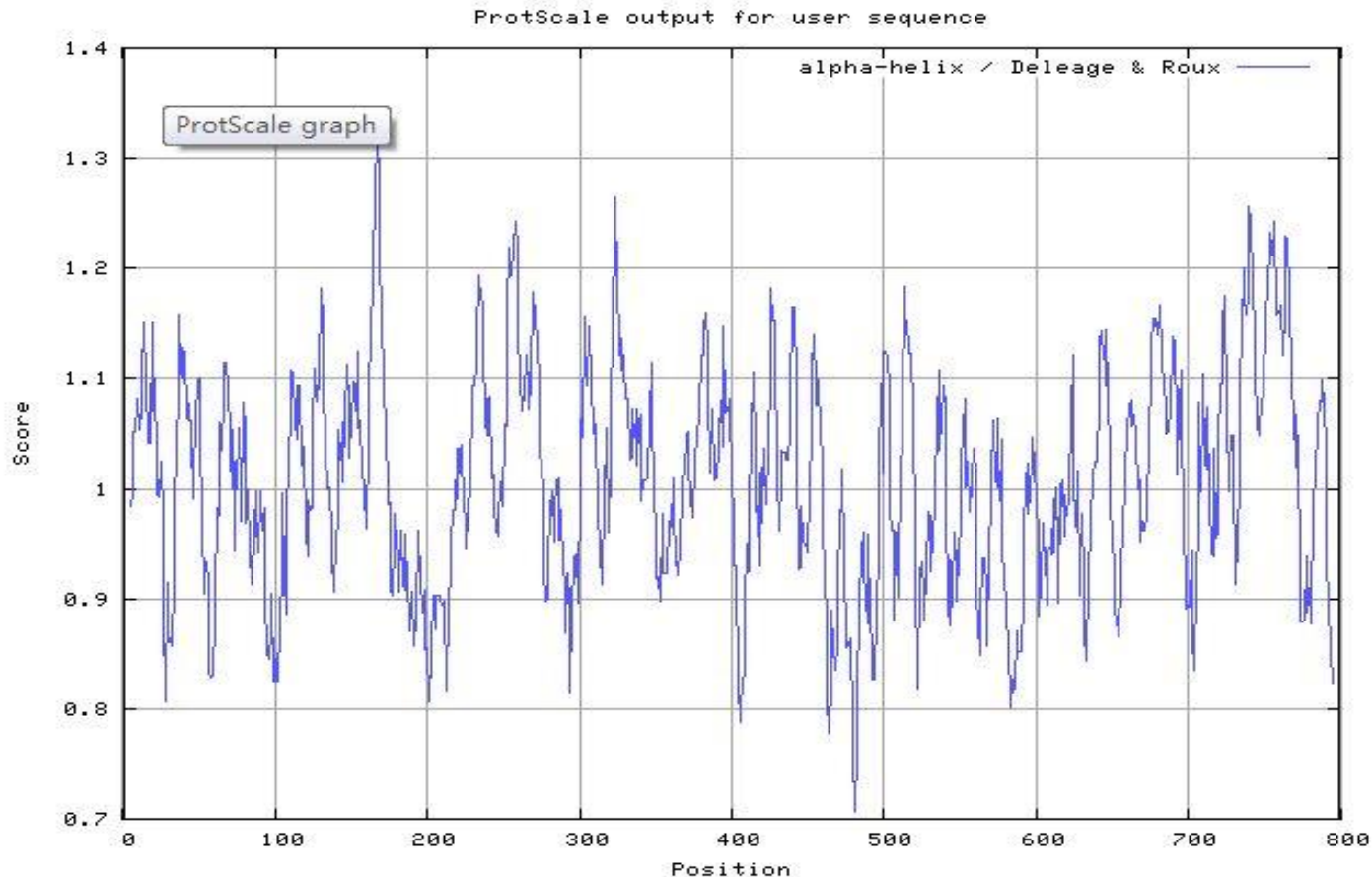
GSAPNAMTILLAVVGSILLIGMALLAIWKLLVTIHD



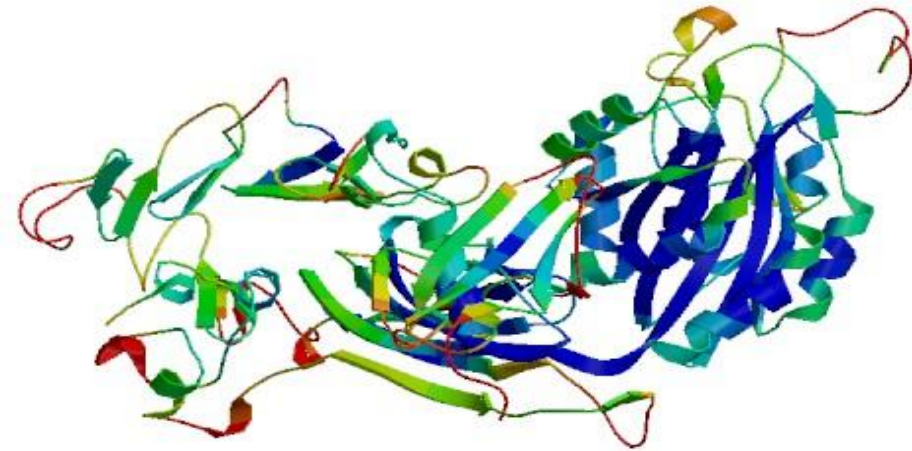
Structure Prediction

The alpha-helix prediction by protscale

The result shows there may exist 11-12 alpha-helices, which match the Swiss-model result.



Structure Prediction

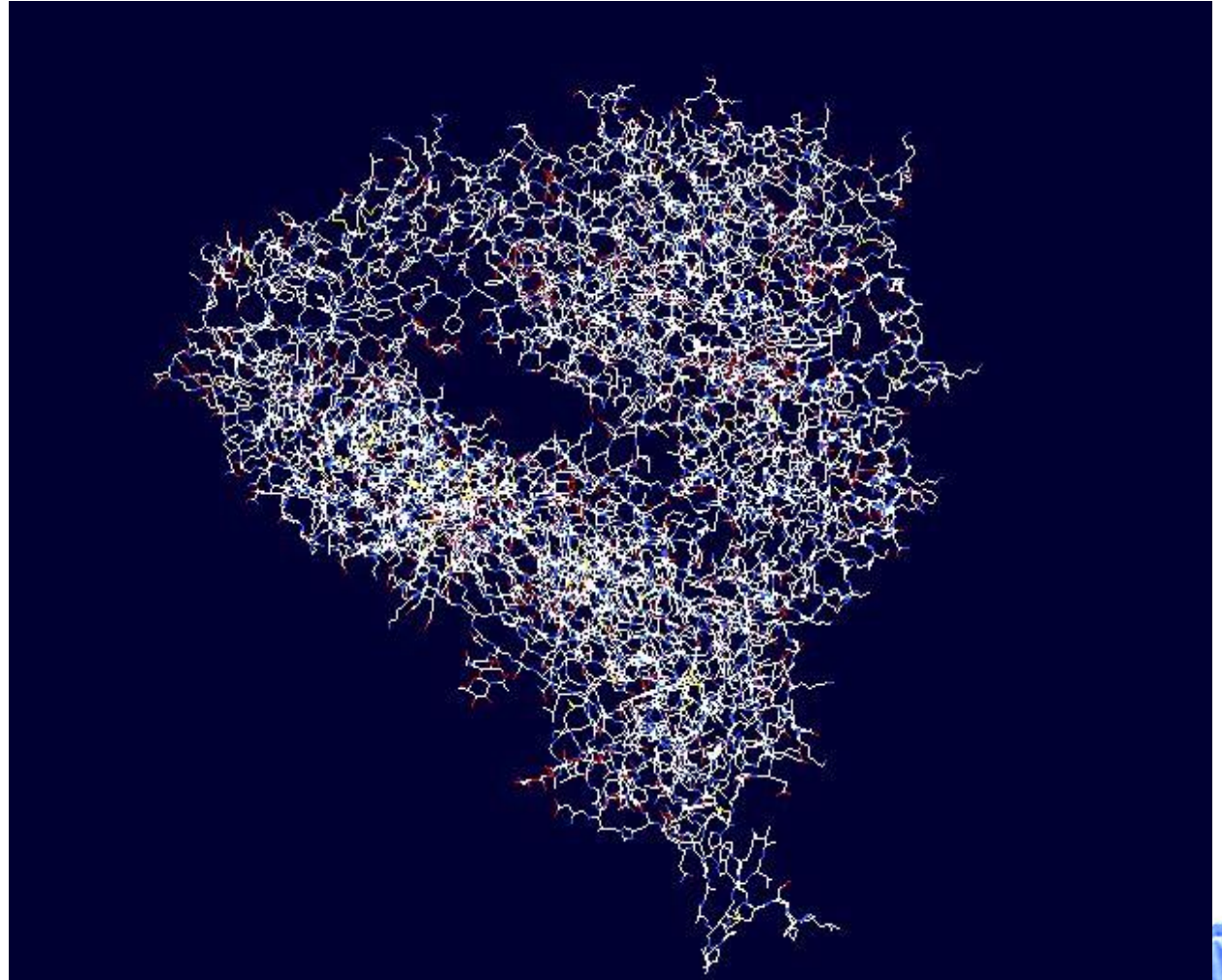


The 3D structure prediction of our target sequence by swiss-model



Structure Prediction

The “magic fit” between our prediction and the complete integrin $\alpha V\beta 3$ in 3D structure. This proved both the basic character of our sequence and its 3D prediction.



Function Analysis

- Above all, this project has proved our sequence to be mus musculus integrin beta-5; the transmembrane analysis and the 3D simulation has made a further certification that the target sequence plays a key role in the signal transduction between ECM and cell.



References

- S. J. Shattil, C. Kim and M. H. Ginsberg (2010) The final steps of integrin activation: the end game. *Nature Reviews Cell and Molecular Biology* 11, 288-300.
- M. A. Arnaout, S. L. Goodman and J.-P. Xiong (2007) Structure and mechanics of integrin-based cell adhesion. *Current Opinion in Cell Biology* 19, 495-507.
- B.-H. Luo, C. V. Carman and T. A. Springer (2007) Structural basis of integrin regulation and signaling. *Annual Review of Immunology* 25, 619-47.
- B.-H. Luo and T. A. Springer (2006) Integrin structures and conformational signaling. *Current Opinion in Cell Biology* 18, 579-586.



Statement

- 由于我们并未进实验室获得第一手的实验数据，我们此次分析的序列为Yu.C.C.、Nandrot.E.F.等作者于2012年提交到genebank数据库中的小鼠整联蛋白序列。



Acknowledgement

- 感谢罗老师一个学期以来的精心指导！
- 感谢我们小组所有成员的合作和努力！
- 感谢同学们的支持！



Thank you!
James!

