

In silico analysis of myostatin protein of *Labeo calbasu*

Abstract

Myostatin is a protein released in autocrine fashion which exerts negative effect for skeletal muscle growth during myoblast proliferation and differentiation in teleost fishes. Myostatin is a member of the TGF- β family encoded by MSTN gene, also known as growth differentiation factor 8 (GDF-8). MSTN gene knock-out exhibited hypertrophy and hyperplasia of skeletal muscle, but over expression or systemic administration of myostatin causes muscle atrophy. Blocking of myostatin activity helps in generating higher muscle growth rate and mutation in MSTN gene creates double-muscle phenotype. Therefore, in silico analysis of Myostatin in an Indian major carp species *Labeo calbasu* has carried out in the present study for understanding structural organization of the concerned protein. Amino acid sequence having the accession number AEN75195.1 was retrieved from NCBI. The sequence was subjected to ExPasy ProtParam, PSIPRED and GORIV, PDB sum and SWISSMODEL/Workspace for finding out the physicochemical properties, secondary structure prediction and template selection procedure respectively. The predicted 3D structure of Myostatin by Swiss model server was validated through RAMPAGE server and then submitted to Protein Model Database (PMDb) having the ID No: PM0081037. The extracellular location was also analysed by Hum-mPLoc 2.0. The structural examination of Myostatin will help in development of antagonistic protein or inhibitor, which can shed a new direction for the growth promotion of piscine species during various periods of life cycle.

Keywords: myostatin, *in silico*, homology modeling, *labeo calbasu*, GORIV, expasy protparam, MSTN, uniprotKB

Volume 3 Issue 5 - 2017

Samik Acharjee, Anil Datt Upadhyay, Ajit Kumar Roy, Rumpi Ghosh

Department of Bioinformatics, Central Agricultural University, India

Correspondence: Ajit Kumar Roy, Bioinformatics Center, College of Fisheries, Lembucherra, Tripura, Central Agricultural University, India, 799210, Email akroy1946@yahoo.co.in

Received: June 12, 2017 | **Published:** August 29, 2017

Abbreviations: GDF, growth differentiation factor; PMDB, protein model database; NCBI, national center for biotechnology information; GRAVY, grand average of hydropathicity

Introduction

Myostatin is a protein released from myocytes acting in autocrine fashion for inhibiting the myogenesis. It is a member of the TGF- β protein family encoded by MSTN gene which acts as a negative regulator for skeletal muscle growth during myoblast proliferation and differentiation in teleost fishes.¹ Due to its involvement in growth processes it is also known as growth differentiation factor 8 (GDF-8). MSTN gene Knock out showed both the hypertrophy and hyperplasia of skeletal muscle,² whereas, over expression or systemic administration of it resulted muscle atrophy.^{3,4} In agricultural and veterinary fields, treatment for blocking of myostatin activity is in practice which helps in generating higher muscle growth rate. Generally, it is expressed in skeletal muscle and the mutation in MSTN gene creates double-muscle phenotype^{2,5,6} in several domesticated species like Belgian blue and Piedmontese cattle⁷ or Texel sheep.⁸ However from fishes such as rainbow trout,^{9,10} atlantic salmon (*Salmo salar*),¹¹ zebra fish (*Danio rerio*),^{12,13} sea bream (*Sparus aurata*),¹⁴ Orange spotted grouper (*Epinephelus coioides*),¹⁵ medaka (*Oryzias latipes*),¹⁶ baramundi (*Lates calcarifex*),¹⁷ sea perch (*Lateolabrax japonicus*)¹⁸ and catfish (*Ictalurus punctatus*)¹⁹ etc the presence of Myostatin genes were well reported. Studies on myotomal musculature of fishes are interesting area of research since few decades. In vivo blocking of Myostatin gene expression in skeletal muscle of fishes was also carried out by Terova et al.²⁰ The structural exploration of Myostatin in an Indian major carp species like *Labeo calbasu* has a potential significance, which can ultimately results to a conception of development of antagonistic protein or inhibitor protein

for generating higher growth in fishes during different stages of life cycle.

Materials and methods

Sequence retrieval

The amino acid sequence of Myostatin of *Labeo calbasu* was retrieved from National Center for Biotechnology Information (NCBI) having the accession number AEN75195.1. The FASTA format of the sequences were downloaded and used for further analysis. The retrieved sequence was verified by peptide search in the UniProtKB having the entry no. G3EIK5.

Physicochemical properties of myostatin

ExPasy ProtParam server was used for determination of physicochemical properties of the Myostatin protein.²¹ The molecular weight, theoretical pI, the % of amino acid composition, total number of negative and positive residue, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index, grand average of hydropathicity (GRAVY) were analyzed by this web tool.

Secondary structure prediction

Secondary structure of the Myostatin was analysed by PSIPRED²² and GORIV²³ methodology.

Template selection

Template search for the query protein, Myostatin was performed through PDBsum database which presented 114 hits.²⁴ The template 3hh2 (A) was selected showing 89% sequence identity. The template sequence was also cross verified by SWISSMODEL/Workspace which displayed 88.99% sequence identity with the query sequence.²⁵

Homology modeling

By homology modeling, three dimensional structure of Myostatin was predicted using Swiss model server.^{26,27}

Validation of the prediction

The predicted structure of the Myostatin was validated through Ramachandran plot by utilizing RAMPAGE server.²⁸

Protein model database

The predicted model was also submitted to the Protein Model Database (PMDb). PMDB is a three dimensional protein models database.²⁹

Subcellular localization

Hum-mPLoc 2.0 was used to predict the subcellular localization of the concerned protein Myostatin.³⁰

Results and discussion

Analysis of retrieved sequences

The retrieved sequence of myostatin from NCBI and UniProtKB heaving the accession number AEN75195.1 and entry no. G3EIK5 was subjected to multiple sequence alignment through Clustal Omega for finding out the sequence similarity. The multiple sequence alignment showed the almost close sequence similarity or identity. But, a little number of differences exists among the retrieved sequences of the different databases (Figure 1), which denotes that clustal Omega or clustal X is a useful web tool for finding out such differences. Therefore, it is required to consider any one database sequence before proceeding for further analysis during such studies. Moreover, the retrieved sequence of Myostatin from NCBI heaving the accession number AEN75195.1 was considered as the query sequence for further analysis.

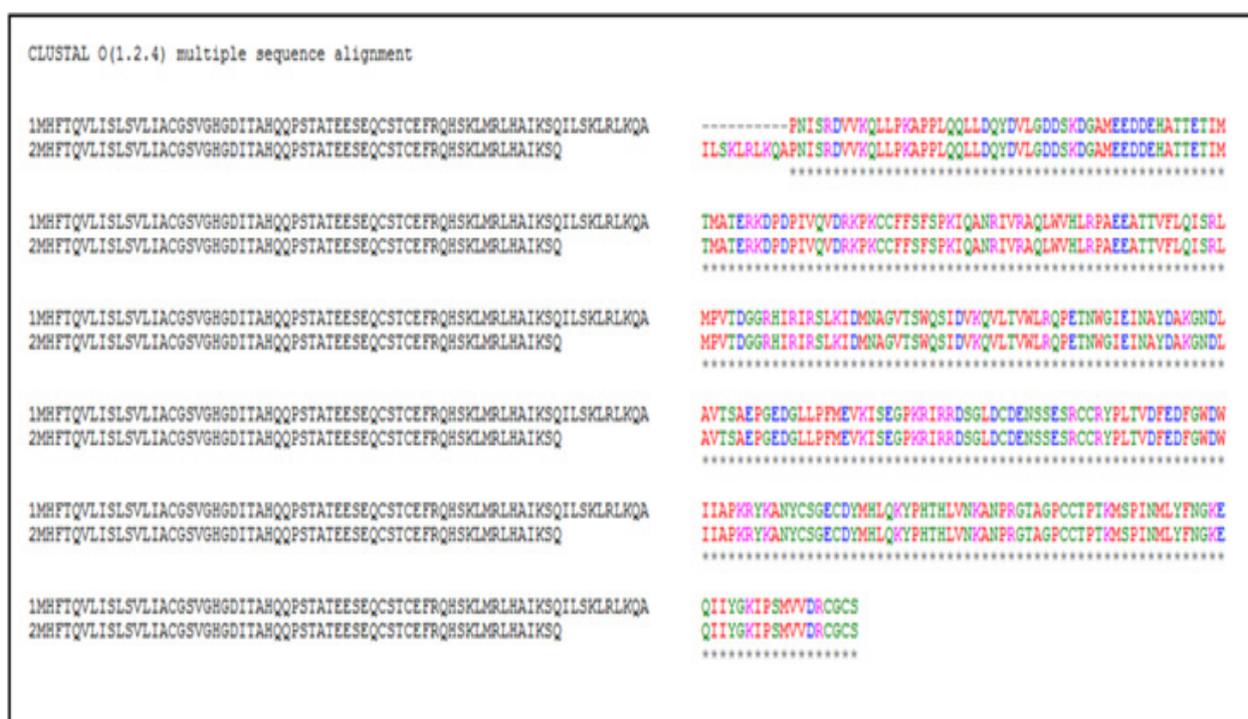


Figure 1 Multiple sequence alignment between the retrieved sequences of Myostatin from NCBI and UniProtKB.

Physicochemical properties of myostatin

The physicochemical properties of Myostatin were analyzed by ExPASy protParam server. Myostatin is a 378 amino acid long protein with the estimated molecular weight 42613.74kDa. The theoretical isoelectric point (pI) is 6.42 and the amino acid composition showed the maximum presence of Leucine (7.7%) and minimum presence of Tryptophan (1.6%) (Table 1). The total number of negatively charged and positive charged residues of Myostatin are (Asp+Glu)-48 and (Arg+Lys)-45, respectively. The formula $C_{1863}H_{2959}N_{527}O_{565}S_{26}$ of Myostatin contains 5940 atoms in total. The extinction coefficients of the concerned protein were calculated at 280 nm in water ($M^{-1}cm^{-1}$) which presented the values of ext. coefficient 47285 and Abs 0.1% (=1g/l) 1.110, assuming all pairs of Cys residues form cystines; but the ext. coefficient becomes 46410 and Abs 0.1% (=1g/l) is 1.089 when assumed all Cys residues are reduced. The N-terminal of the sequence of the protein is Methionine (Met). The estimated half-life of Myostatin presented 30hours in case of in-vitro culture condition of mammalian reticulocytes; whereas more than 20hours for in-vivo

culture condition of yeast and more than 10hours in case of *E. coli*. The estimated instability index (II) of the Myostatin is 54.27 which classify the protein as unstable. Aliphatic index 78.41 of the protein measures its considerable thermostability along with the relative volume occupied by aliphatic side chains. The value -0.430 of the GRAVY indicates that hydrophilic nature of the protein.

Secondary structure prediction

The secondary structure of Myostatin predicted by PSIPRED server revealed the presence of helix region, strand region and coil regions (Figure 2). PSIPRED works on PSI-blast homology search algorithm.²² GOR IV analysis suggested the presence of alpha helix 23.54%, extended strands 21.16% and coiled region of 55.29%.

Template identification

For homology modeling, template 3hh2 (A) was identified showing 89% sequence similarity along with z-score of 987.78 from the PDBsum database (Figure 3). The selected template contains

the crystal structure of myostatin: follistatin 288 complexes by X ray diffraction technique (2.15Å). The template sequence was also

verified by SWISSMODEL/Workspace possessing the no. 3hh.2.1.A with 88.99% sequence identity with the query sequence (Figure 4).

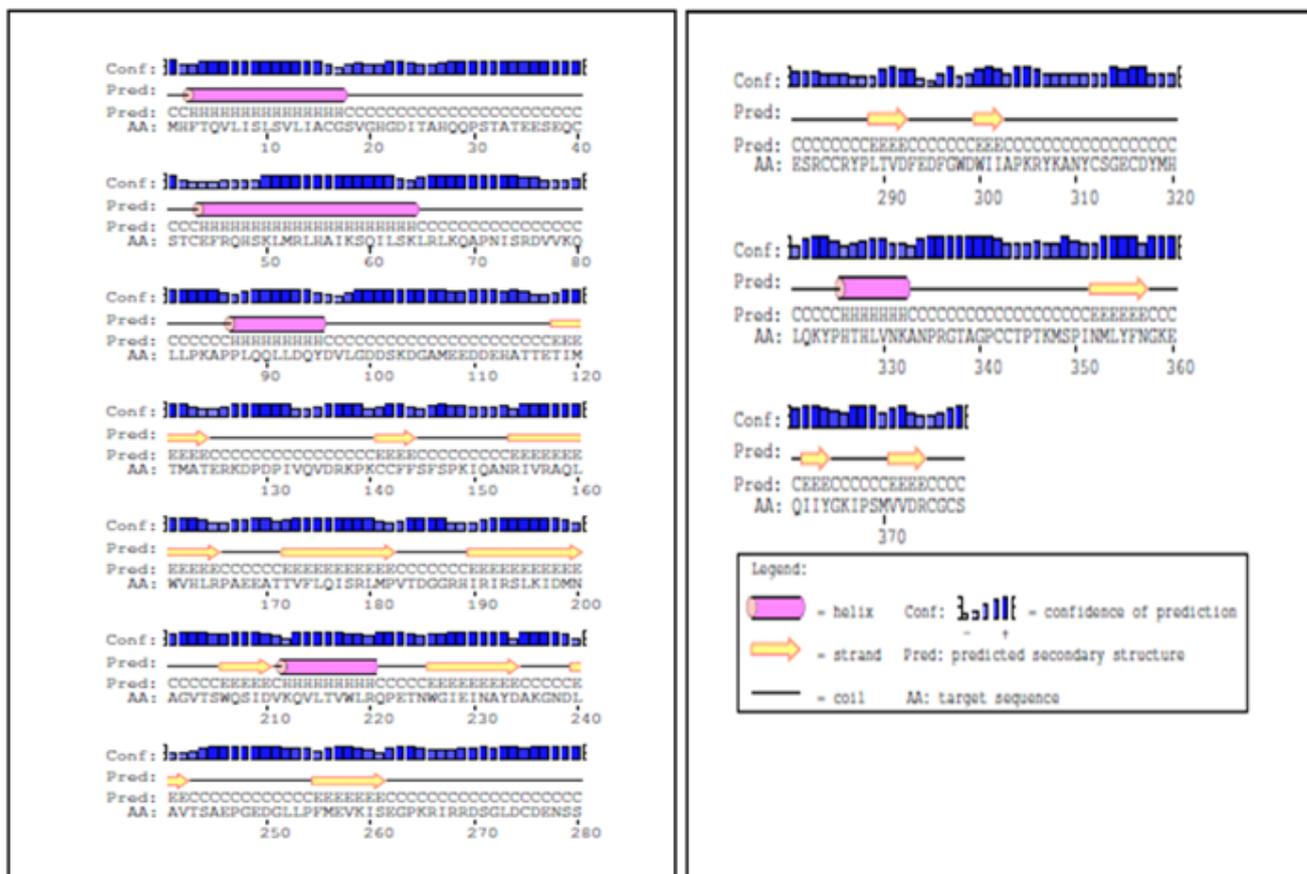


Figure 2 Secondary structure of Myostatin by PSIPRED.

PDBsum
Sequence:

MHFTQVLISLSVLIACGSVGHGDIHQQPSTATEESEQCSTCEFRQHSKLMRLHAIKSQILSKLRLKQAPNISRDVVVKQL
LLPKAPLQQLLDQYDVLGDDSKDGAMEEDEHATTETIMTWATERKDPPIVQVDRKPKCCFFSFKIQANRIVRAQLN
HLRPAEEATTVFLQISRLMPVTDGGRHIRIRSLKIDNAGVTSWQSIDVQVLTWLRQPETNNGIEINAYDAKNDLAVT
SAEPGEDOLLPFMEVKISEGPKRIRRDGLDCDENSESRCCRYPLTVDFEDFGNDWI IAPKRYKANYCSGECDCYMLQKY
FHTLNVKANPRGTAGPCTPTKMSFINMLYFNGKEQIIYKIPSMVDRCCGS

Sequence length: 378 residues.

Your sequence search returned the following 114 hits:

PDB code	Model	%-tsg	a.a.	z-	Ligands	Protein name
		Length	identity	overlap		
				score		
1. 3hh2(A)	X- ray 2.15Å	109	89.0%	109	987.8	PO4, CIT. Crystal structure of the myostatin:follistatin 288 complex
2. 3sek(B)	X- ray 2.40Å	109	89.0%	109	987.8	NAG. Crystal structure of the myostatin:follistatin-like 3 complex
3. 5e4q(A)	X- ray 1.50Å	108	88.8%	107	970.7	PG4. Crystal structure of human growth differentiation factor 11 (gdf-11)
4. 5ihw(A)	X- ray 2.35Å	109	88.8%	107	970.6	PO4, FLC. Crystal structure of the gdf11:follistatin 288 complex
5. 5f3b(C)	X- ray 1.76Å	106	88.1%	109	956.3	GOL. Structure of myostatin in complex with chimeric rk35 antibody
6. 5f3h(I)	X- ray 2.70Å	104	87.8%	107	939.2	Structure of myostatin in complex with humanized rk35 antibody
7. 3rijr(A)	X- ray 3.05Å	340	27.8%	349	490.8	NAG, NAG-NAG. Crystal structure of pro-tgf beta 1
8. 5il1(A)	X- ray 2.25Å	95	78.9%	109	459.2	MPD. Crystal structure of gdf8

Figure 3 Template 3hh2 (A) from the PDBsum database.

Homology modeling

Homology modeling predicted the 3D structure of the Myostatin protein of *Labeo calbasu*. The structural conformation was generated from Swiss model server aligning the query sequence to the template sequence (Figure 5A & 5B). Model quality was estimated by assessing the QMEAN score, which stands for qualitative model energy analysis is composite scoring function describing the major geometrical aspects of protein structures. QMEAN was tested on several standard decoy sets including a molecular dynamics simulation decoy set as well as on a comprehensive data set. QMEAN shows a statistically significant improvement over nearly all quality measures describing the ability of the scoring function to identify the native structure and discriminate good from bad models.³¹

Validation of the prediction

The validation of the predicted structure (Figure 6A & 6B) of the proteins was done through Ramachandran plot (phi/psi). The stereochemical analysis of RAMPAGE server showed number of residues in favoured region is 96.3% and in allowed region is 3.7% (Figure 6), which indicates that predicted models is likely to be correct.

Protein model database

The generated model for myostatin was successfully submitted in the Protein Model Database (PMD) heaving the PMID: PM0081037 (Figure 7).

<input type="checkbox"/>	5fo.1.H	Transforming growth factor beta-1		34.30	X-ray, 3.5Å	hetero-oligomer	8 x CA ^{CS} , 3 x BMA ^{CS} , 26 x NAG ^{CS} , 6 x MN ^{CS} , 11 x MAN ^{CS}	▼
<input type="checkbox"/>	3rjr.1.A	Transforming growth factor beta-1		36.05	X-ray, 3.0Å	homo-dimer	1 x NAG-NAG ^{CS} , 1 x NAG ^{CS}	▼
<input type="checkbox"/>	3rjr.1.B	Transforming growth factor beta-1		36.05	X-ray, 3.0Å	homo-dimer	1 x NAG-NAG ^{CS} , 1 x NAG ^{CS}	▼
<input checked="" type="checkbox"/>	3hh2.1.A	Growth/differentiation factor 8		88.99	X-ray, 2.1Å	hetero-oligomer	None	▼
<input type="checkbox"/>	3sek.1.C	Growth/differentiation factor 8		88.99	X-ray, 2.4Å	hetero-oligomer	2 x NAG ^{CS}	▼
<input type="checkbox"/>	5f3b.1.F	Growth/differentiation factor 8		88.99	X-ray, 1.8Å	hetero-oligomer	None	▼
<input type="checkbox"/>	5ji.1.B	Growth/differentiation factor 8		88.99	X-ray, 2.2Å	homo-dimer	None	▼
<input type="checkbox"/>	5ji.1.A	Growth/differentiation factor 8		88.99	X-ray, 2.2Å	homo-dimer	None	▼
<input type="checkbox"/>	3hh2.1.A	Growth/differentiation factor 8		88.89	X-ray, 2.1Å	hetero-oligomer	None	▼

Figure 4 Template 3hh.2.1.A from the SWISSMODEL/Workspace.

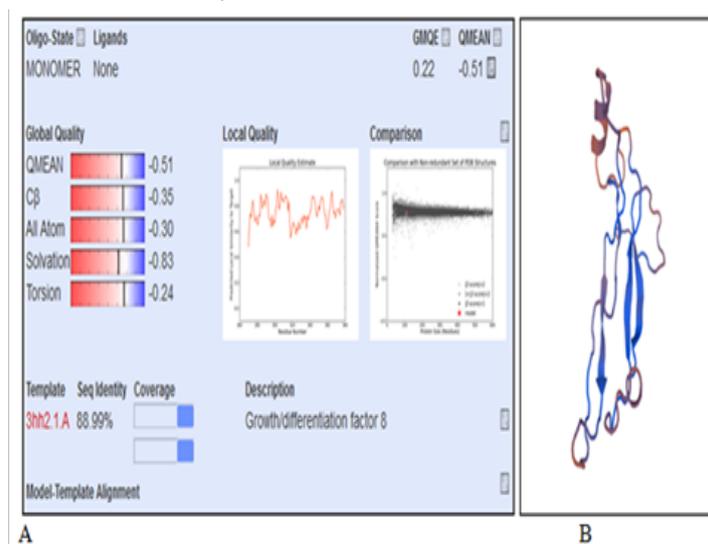


Figure 5 Homology modeling myostatin
 A. Model report and
 B. Modeled structure.

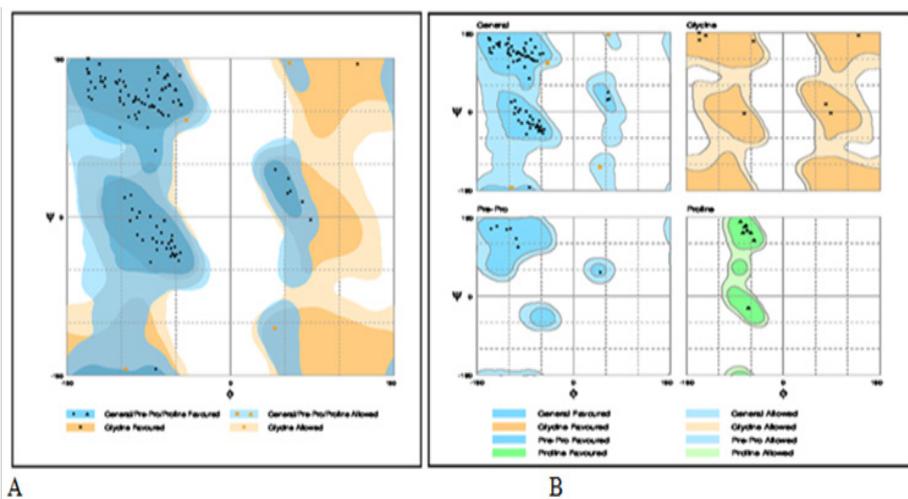


Figure 6 Ramachandran plot of the predicted structure of myostatin.



Figure 7 Protein Model Database (PMDb) with submission ID: PM0081037.

Table 1 Amino acid composition of myostatin

Amino Acids	No's.	Percentage (%)
Ala (A)	22	5.80%
Arg (R)	22	5.80%
Asn (N)	12	3.20%
Asp (D)	26	6.90%
Cys (C)	14	3.70%
Gln (Q)	20	5.30%
Glu (E)	22	5.80%
Gly (G)	21	5.60%
His (H)	11	2.90%
Ile (I)	25	6.60%
Leu (L)	29	7.70%

Table Continued..

Amino Acids	No's.	Percentage (%)
Lys (K)	23	6.10%
Met (M)	12	3.20%
Phe (F)	10	2.60%
Pro (P)	23	6.10%
Ser (S)	27	7.10%
Thr (T)	22	5.80%
Trp (W)	6	1.60%
Tyr (Y)	9	2.40%
Val (V)	22	5.80%

Subcellular localization

By the application of Hum-mPLoc 2.0, extra cellular localization was obtained for the Myostatin protein.

Conclusion

The in-silico investigation of Myostatin revealed the physicochemical and structural parameters of the protein involved in prohibition or inhibition of skeletal muscle growth. Based on the findings, it could be concluded that Myostatin is an unstable, hydrophilic protein with secondary structure of alpha helix 23.54%, extended strands 21.16% and coiled region of 55.29%. The generated 3D structure of the Myostatin protein of *Labeo calbasu* by following homology modeling method was submitted to Protein Model Database (PMDB) which can be retrieved by the PMID no: PM0081037. Moreover, it might be resolved that the predicted 3D structure exhibited favoured region of 96.3% and allowed region of 3.7% which indicates that model is expected to be correct in prediction. The modeled structure of Myostatin can further utilized for predicting or designing blocker or inhibitor of the concerned protein. It can shed a new direction for the growth promotion of piscine species in a rapid manner in different developmental stages of life cycle. Further, extra cellular localization of Myostatin protein is also predicted by web tool which reduces the cost and effort during the experimental examination processes such as: immunohistochemistry, immunofluorescence etc. X-ray crystallography and NMR spectroscopy are most convenient method of 3D structure prediction, but requires enormous time and financial support, however also a tedious method in application. Moreover, the utility of computational and web tools minimizes these efforts. It is also useful for filling up the gaps generated within the large amount of data with available sequences and solved structures.

Acknowledgements

Authors are thankful to the Dean, College of Fisheries, Central Agricultural University (I), Lembucherra, Tripura (W) for encouragement and moral support. This work has been carried out under DBT funded project "Establishment of Bioinformatics Infrastructure Facility for Biology Teaching through Bioinformatics" and hence financial support by the Department of Biotechnology, Ministry of Science and Technology, Govt. of India, New Delhi is duly acknowledged.

Conflict of interest

The author declares no conflict of interest.

References

- Lee SJ, McPherron AC. Regulation of myostatin activity and muscle growth. *Proc Natl Acad Sci USA*. 2001;98(16):9306–9311.
- McPherron AC, Lee SJ. Double muscling in cattle due to mutations in the myostatin gene. *Proc Natl Acad Sci USA*. 1997;94(23):12457–12461.
- McPherron AC, Lawler AM, Lee SJ. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*. 1997;387(6628):83–90.
- Fanzani A, Conraads VM, Penna F, et al. Molecular and cellular mechanisms of skeletal muscle atrophy: an update. *J Cachexia Sarcopenia Muscle*. 2012;3(3):163–179.
- Grobet L, Martin LJ, Poncelet D, et al. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet*. 1997;17(1):71–74.
- Kambadur R, Sharma M, Smith TPL, et al. Mutations in myostatin (GDF8) in Double-Muscling Belgian Blue and Piedmontese Cattle. *Genome Research*. 1997;7(9):910–915.
- Culley G. *Observations on livestock*. 4th ed. London, UK: G Woodfall; 1807.
- Laville E, Bouix J, Sayd T, et al. Effects of a quantitative trait locus for muscle hypertrophy from Belgian Texel sheep on carcass conformation and muscularity. *J Anim Sci*. 2004;82(11):3128–3137.
- Garikipati DK, Gahr SA, Rodgers BD. Identification, characterization, and quantitative expression analysis of rainbow trout myostatin-1a and myostatin-1b genes. *J Endocrinol*. 2006;190(3):879–888.
- Garikipati DK, Gahr SA, Roalson EH, et al. Characterization of rainbow trout myostatin-2 genes (rtMSTN-2a and -2b): genomic organization, differential expression, and pseudogenization. *Endocrinology*. 2007;148:2106–2115.
- Ostbye TK, Bardal T, Vegusdal A, et al. Molecular cloning of the Atlantic salmon activin receptor IIB cDNA-Localization of the receptor and myostatin in vivo and in vitro in muscle cells. *Comp Biochem Physiol*. 2007;2(2):101–111.
- Xu C, Wu G, Zohar Y, et al. Analysis of myostatin gene structure, expression and function in zebrafish. *J Exp Biol*. 2003;206(Pt 22):4067–4079.
- Kerr T, Roalson EH, Rodgers BD. Phylogenetic analysis of the myostatin gene sub-family and the differential expression of a novel member in zebrafish. *Evolution & development*. 2005;7(5):390–400.
- Maccatrozzo L, Bargelloni L, Radaelli G, et al. Characterization of the myostatin gene in the gilthead seabream (*Sparus aurata*): sequence, genomic structure, and expression pattern. *Mar Biotechnol*. 2001;3(3):224–230.

15. Ko CF, Chiou TT, Chen TT, et al. Molecular cloning of myostatin gene and characterization of tissue-specific and developmental stage-specific expression of the gene in orange spotted grouper, *Epinephelus coioides*. *Mar Biotechnol*. 2007;9(1):20–32.
16. Chisada S, Okamoto H, Taniguchi Y, et al. Myostatin-deficient medaka exhibit a double-muscling phenotype with hyperplasia and hypertrophy, which occur sequentially during post-hatch development. *Dev Biol*. 2011;359(1):82–94.
17. Santis DC, Jerry DR. Differential tissue-regulation of myostatin genes in the teleost fish *Lates calcarifer* in response to fasting. Evidence for functional differentiation. *Mol Cell Endocrinol*. 2011;335(2):158–165.
18. Ye HQ, Chen SL, Sha ZX, et al. Molecular cloning and expression analysis of the myostatin gene in sea perch (*Lateolabrax japonicus*). *Biotechnol*. 2007;9(2):262–272.
19. Kocabas AM, Kucuktas H, Dunham RA, et al. Molecular characterization and differential expression of the myostatin gene in channel catfish (*Ictalurus punctatus*). *Biochim Biophys Acta*. 2002;1575(1-3):99–107.
20. Terova G, Rimoldi S, Bernardini G, et al. Inhibition of myostatin gene expression in skeletal muscle of fish by in vivo electrically mediated dsRNA and shRNAi delivery. *Mol Biotechnol*. 2013;54(2):673–684.
21. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis on the ExPASy server. In: John M Walker, editor. *The Proteomics Protocols Handbook*. USA: Humana Press; 2005. p. 571–607.
22. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402.
23. Sen TZ, Jerigen R, Kloczkowski A. GOR V server for protein secondary structure prediction. *Bioinformatics*. 2005;21(11):2787–2788.
24. Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic*. 2001;29(1):221–222.
25. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2):195–201.
26. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009;1:S162–173.
27. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*. 2014;42:W252–258.
28. Lovell c, Davis IW, Arendall WB, et al. Structure validation by C α geometry: phi, psi and C β deviation. *Proteins*. 2003;50(3):437–450.
29. Castrignano T, Meo PDO, Cozzetto D, et al. The PMDB protein model database. *Nucleic Acids Res*. 2006;34:D306–309.
30. Shen HB, Chou KC. A top-down approach to enhance the power of predicting human protein subcellular localization: Hum-mPLoc 2.0. *Anal Biochem*. 2009;394(2):269–274.
31. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. 2011;27(3):343–350.