**Introduction**

Protein with its amino acid are important for maintaining structure of cells, making antibodies to work properly, regulate the growth of hormones with enzymes and contributes to the repairing mechanisms. Organism living in marine and fresh water consists of protein with high amino acid proportion. Fish is a diverse group of organisms that habituates in different aquatic environment and holds prime importance in food industry. Biologically, fish muscle proteins contain all essential nutrients like milk, meat and egg protein. This protein varies in amount from species to species. Globally the consumption of fish production by human is about 77 percent. Fish was chosen as a sample source because there are many different verities of fish and source of protein for many fish species are readily available. Furthermore, fish is very nutritious part of man’s diet since it is rich in vitamins, minerals and all essential amino acids in right proportions. Study of muscle genes and proteins will be beneficial for human since both vertebrates and invertebrates have muscle proteins in common.

**Keywords:** Sequence analyses; Homology modeling; Structural analyses; Vertebrates; Invertebrates

**Materials and Methods**

**Protein retrieval and sequence analysis**

Protein sequences of fish muscle were retrieved from UniProt Knowledgebase database and NCBI using accession no. G1ERR8, Q9PV76, E6ZGD0, Q9PRF1, F8K8N3, Q1L5K3, E6ZHF3, gi|5726351, Q8AW95, gi|59858543, Q58H26, Q9NAS5, E6ZHF3, gi|5726351, Q8AW95, gi|59858543, Q58H26, Q9NAS5, E6ZHF3, gi|5726351, Q8AW95, gi|59858543, Q58H26, Q9NAS5, E6ZHF3, gi|5726351, Q8AW95, gi|59858543, Q58H26, Q9NAS5.
Homology Modeling, Phylogeny and Different Computational Approaches. MOJ Proteomics Bioinform 2(3): 00047. DOI: 10.15406/mojpb.2015.02.00047

After 3D model was constructed evaluation was performed using PSVS and WHAT IF. PSVS was used for assessment of 3D template of known structure with the help of protein structure validation web server (PSVS) [10]. Secondary structure features such as helices, strands, coils, acidic and basic residues, domains, transmembrane topology were predicted using Swiss PDB viewer and PSIPRED. NetTurnP and NetSurfP was used for beta turns and protein surface accessibility prediction. Beta turns formation are important in folding, stability and origin of protein. These features computed by ProtParam were molecular weight, theoretical pl, amino acid composition, atomic composition, extinction coefficient, estimated half life, aliphatic index and grand average of hydrophobicity (GRAVY).

Prediction of secondary structure

Secondary structure of muscle proteins were computed using SWISS PDB Viewer [8], PSIPRED [9], NetTurnP [10] and NetSurfP [11]. Secondary structure features such as helices, strands, coils, acidic and basic residues, domains, transmembrane topology were predicted using Swiss PDB viewer and PSIPRED. NetTurnP and NetSurfP was used for beta turns and protein surface accessibility prediction. Beta turns formation are important in folding, stability and origin of protein.

Validation of 3D structure

After 3D model was constructed evaluation was performed using PSVS and WHAT IF. PSVS was used for assessment of 3D model which integrates information from various structure evaluation software including RPF, PROCHECK, MolProbity, Verify 3D, Prosisa II, and other structure validation software. Stereochemistry analyses were performed using WHAT IF. Deep View was used for visualizing 3D structure [11].

Functional analyses of fish muscle proteins

To study the function of muscle proteins ProtFunc [13] was used. This server utilizes information from other prediction server of DAS annotation viewer related to post translational modification then finally categorize the information in form of cellular role, enzyme class and gene ontology features. NCBI’s Conserved Domain Database (CDD) [14] was used for finding conserved domain in protein sequence.

Submission of the model in protein model database (PMDB)

The models generated for actin, actinin, dystrophin, gelsolin, M2 protein, plastin 3, thymosin, troponin was successfully submitted in Protein model database (PMDB) [15] having PMID: PM0078304, PM0078303, PM0078298, PM0078299, PM0078300, PM0078301, PM0078302 and PM0078305.

Phylogenetic analysis of fish muscle proteins

This section includes multiple sequence alignment of proteins, phylogenetic tree construction and its evaluation, performed using following computational approach. Phylogenetic trees of 10 fish muscle proteins including actin, actinin, dystrophin, fimbrin, gelsolin, myosin heavy chain, spectrin, titin, troponin and troponin were made. BLAST analysis of selected proteins was performed against non redundant databases by setting parameters on default. Then sequences with highest identity greater than 70% were collected for multiple sequence alignment. The same strategy was repeated for each selected protein and step by step sequences were collected for multiple sequence alignment. Computational tools including Clustal X [16], MEGA [17] and DIVEIN [18] were used for understanding the evolutionary significance of fish muscle proteins.

Multiple Sequence Alignment through Clustal X

Clustal X [16] is a widely used multiple sequence alignment tool which is completely coded in C++. Clustal X, which is desktop version of Clustal W was used for multiple sequence alignment in order to get knowledge about structure, function, location, stability and origin of protein. FASTA formatted file containing amino acid sequences was loaded to Clustal X as given by opening file menu. These amino acid sequences were selected by performing BLAST analysis of fish muscle proteins against non redundant protein sequence databases. The sequences with lower E-value and identity greater than 70% were chosen for multiple sequence alignment. The alignment was performed in Clustal X by setting parameters as gap opening 20, gap extension 20, delay divergent sequences 30, negative matrix off and protein weight matrix used was Gonnet series. Nexus, Clustal and FASTA was marked for an output.

Construction of Phylogenetic tree by using MEGA

MEGA [17] stands for Molecular Evolutionary Genetics Analysis used for evolutionary study of DNA and protein sequences. It is a desktop application which was used for comparative study of homologous sequences belonging to different species and
Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches

Copyright: ©2015 Khalid et al.

Results and Discussion

The present study was to perform sequence and structure analysis of fish muscle proteins. The protein sequences were retrieved from Uniprot database and NCBI with accession number as G1ERR8, Q9PV76, E6ZGD0, Q9PRFi, F8KB8N3, Q1LSK3, E6ZHFE3, gi|5726351, Q8AW95, gi|9858543, Q58H26, Q9NASS5, gi|185132813, Q8UVF6 and gi|49901349.

Protein sequence analysis

BindN was used for predicting DNA and RNA binding residues for fish muscle proteins which is useful for understanding protein-nucleic acid interaction. The degree of conservation of amino acid depicts the structural and functional importance. The positions which evolve rapidly are considered as variable while positions which evolve slowly are known to be conserved. This tool was used for identification of functional region in fish muscle proteins. Consurf was explored for estimation of evolutionary conserved amino acids in protein which was based on phylogenetic relationship inferred from homologous sequences (Table 1).

**Table 1:** Binding residues with conserved amino acids predicted by BindN and ConSurf.

| Protein | Total No. of Residues | No. of Exposed Residues According to Neural Network Algorithm | No. of Buried Residues According to Neural Network Algorithm | No. of Functional Residues (Highly Conserved and Exposed) | No. of Structural Residues (Highly Conserved and Buried) | Predicted DNA Binding Residues | Predicted RNA Binding Residues |
|---------|-----------------------|-------------------------------------------------------------|-------------------------------------------------------------|-----------------------------------------------------------|----------------------------------------------------------|-----------------------------|-------------------------------|
| Actin   | 103                   | 52                                                          | 51                                                          | 21                                                        | 4                                                        | 22                          | 22                           |
| Actinin | 110                   | 64                                                          | 46                                                          | 20                                                        | 12                                                       | 21                          | 24                           |
| Dystrophin | 40                | 29                                                          | 11                                                          | 6                                                         | 4                                                        | 15                          | 16                           |
| Filamin | 1343                  | 0                                                           | 0                                                           | 0                                                         | 0                                                        | 262                         | 303                          |
| Gelsolin | 730                  | 458                                                         | 235                                                         | 112                                                       | 47                                                       | 147                         | 186                          |
| M1      | 196                   | 0                                                           | 0                                                           | 0                                                         | 0                                                        | 37                          | 31                           |
| M2      | 190                   | 115                                                         | 75                                                          | 25                                                        | 15                                                       | 40                          | 33                           |
| Myosin  | 43                    | 28                                                          | 15                                                          | 17                                                        | 8                                                        | 7                           | 9                            |
| Nebulin | 57                    | 43                                                          | 14                                                          | 14                                                        | 7                                                        | 29                          | 33                           |
| Plastin | 627                   | 405                                                         | 221                                                         | 83                                                        | 46                                                       | 103                         | 122                          |
| Spectrin | 220          | 154                                                         | 66                                                          | 32                                                        | 8                                                        | 40                          | 49                           |
| Thymosin | 42                    | 38                                                          | 4                                                           | 7                                                         | 0                                                        | 12                          | 13                           |
| Titin   | 129                   | 80                                                          | 49                                                          | 33                                                        | 17                                                       | 21                          | 39                           |
| Tropomyosin | 284              | 213                                                         | 73                                                          | 56                                                        | 7                                                        | 45                          | 69                           |
| Troponin | 223                  | 186                                                         | 37                                                          | 39                                                        | 5                                                        | 81                          | 121                          |

PROFEAT is a bioinformatics server used for calculating structural and chemical features of protein from primary sequence data. These features provides knowledge about biological properties of proteins and peptides. Thus in order to compute the structural and physicochemical features of proteins and peptides PROFEAT was used. All fish muscle proteins were found as non-allergen (Table 2).

Prediction of 3D structure by using homology-modeling approach

An important term used in structure prediction is homology modeling which refers to prediction of three-dimensional structure of protein by using template of known 3D structure. The 3D structure of protein provides knowledge about function of protein and activity of an enzyme. Structure prediction also plays key role in bioinformatics in terms of medicine and biotechnology. First BLAST database was searched to find the best template of known structure with highest identity. BLAST search with default parameters were performed against PDB to find best template. The template having maximum identity was selected for homology modeling to study the protein of interest. Then 3D model was generated by using template of known structure with the help of protein structure prediction web server (PS*). Template used for predicting 3D model was 1DXA for actin, ITJT_A for actinin, 1DXF_A for dystrophin, 2FGH_A for gelsolin, 2JDF_A for M2 protein, 1A0A_A for plastin 3, 1HJO_A for thymosin.
and 1JID_E for tropinin (Figure 1-8).

**Table 2:** Protein family name predicted by PROFEAT.

| Protein | Protein Functional Family Prediction |
|---------|-------------------------------------|
| Titin   | All lipid binding protein, ion binding, chlorophyll biosynthesis, calcium binding, TC 3A 1 ATP binding cassette (ABC) family, motor protein, actin binding, magnesium binding. |
| Filamin | Cell adhesion, zinc binding, all lipid binding proteins, virulence, metal binding, antigen, actin binding, and DNA repair. |
| Spectrin| All lipid binding proteins, metal binding, actin binding, calcium binding. |
| M1      | Iron binding, transferases, alky or aryl groups, all lipid binding proteins, zinc binding, structural protein (matrix protein, core protein, viral occlusion body, keratin), oxidoreductases acting on CH-CH group of donors, lipid metabolism, transferases including acyl transferases, all DNA binding, metal binding, lyses including carbon oxygen lyses, DNA repair. |
| M2      | Transmembrane, transferases are including glycotransferases, iron binding, copper binding, oxidoreductases acting on heme group of donors, magnesium binding. |
| Actinin | rRNA binding protein, zinc binding, DNA repair; calcium binding, magnesium binding, TC 3A 1 ATP binding cassette (ABC) family. |
| Gelsolin| Zinc binding, actin capping, transferases including glycotransferases, all lipid binding protein, metal binding, actin binding, photosystem 1, calcium binding. |
| Actin   | Zinc binding, all DNA binding, actin binding. |
| Tropomyosin | All lipid binding protein, actin binding, copper binding. |
| Troponin| Copper binding. |
| Plastin 3| Zinc binding, transferases transferring phosphorous containing groups, glycotransferases, metal binding, all lipid binding protein, actin binding, calcium binding, pore forming toxins (proteins and peptides), transferases transferring one carbon groups, photosystem 1, carbon binding. |

**Figure 1:** Actinin 3D structure  
**Figure 2:** Dystrophin 3D structure  
**Figure 3:** M2 protein 3D structure  
**Figure 4:** Plastin 3D structure  
**Figure 5:** Actin 3D structure  
**Figure 6:** Gelsolin 3D structure  
**Figure 7:** Thymosin 3D structure  
**Figure 8:** Troponin 3D structure

After construction of 3D model evaluation was performed using PSVS and WHAT IF. PSVS was used to determine the Ramachandran plot to assure the quality of the model. The result of the Ramachandran plot of all predicted models showed greater than 90% residues in favorable region representing that it is a reliable and good quality model (Table 3). A model having more than 90% residues in favorable region is considered as good quality model. 3D model was further evaluated by WHAT IF, which after performing stereochemical analysis indicated that predicted models are correct.

**Table 3:** Tabulated form of predicted structure of fish muscle proteins illustrating template and target used with some physicochemical properties predicted by ProtParam.

| PMDB ID | Protein ID | Target Protein | PDB Template | Ramachandron Plot % score | Lengh of a.a | Molecular Weight | Theoretic Pl |
|---------|------------|----------------|--------------|---------------------------|--------------|-----------------|-------------|
| PM0078304 | Q5H126 | Actin | 1D4X_A | 96.7% | 103 | 11630 | 5.71 |
| PM0078303 | Q8AW95 | Actin | 1TJT_A | 98% | 110 | 12470 | 9.47 |
| PM0078298 | Q9PV76 | Dystrophin | 1DX_A | 91.7% | 40 | 4532 | 8.36 |
| PM0078299 | gi|59858543 | Gelsolin | 2FGH_A | 91.7% | 730 | 81360.5 | 5.54 |
| PM0078300 | E6ZHF3 | M2 protein | 2JD_F | 93.2% | 190 | 23107.3 | 7.56 |
| PM0078301 | gi|49901349 | Plastin 3 | 1A0_A | 93.2% | 190 | 76149.5 | 5.95 |
| PM0078302 | Q8UVF6 | Thymosin | 1HI0_A | 97.3% | 42 | 4851.5 | 5.31 |
| PM0078305 | gi|185132813 | Troponin | 1JJD_E | 100% | 75 | 9256 | 9.86 |

**Citation:** Khalid S, Idrees S, Khalid H, Hussain B, Tiwari S, et al. (2015) *Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches*. MOJ Proteomics Bioinform 2(3): 00047. DOI: 10.15406/mojpb.2015.02.00047
Visualization of 3D structures was performed using DEEP VIEW. Secondary structure of muscle proteins were computed using SWISS PDB Viewer and PSIPRED. NCBI’s Conserved Domain Database (CDD) was used for finding conserved domain in protein sequence. Secondary structure features (Table 4) such as helices, strands, coils, acidic and basic residues, domains, transmembrane topology were predicted using Swiss PDB viewer, CDD and PSIPRED.

Secondary structure of protein plays important role in protein classification, predicting structural changes and function of protein.

NetTurnP and NetSurfP was used for beta turns (Table 5) and protein surface accessibility prediction. Beta turns are non repetitive structures. Beta turns formation are important in folding, stability of proteins and molecular recognition processes. DIANNA [5] was used for cysteine classification and prediction of disulfide connectivity, which provides useful information related to secondary structure since disulphide bonds, helps in stabilizing the folding of protein.

**Functional analyses of fish muscle proteins**

To study the function of muscle proteins ProtFunc (Table 6) was used. This study predicted that all muscle proteins have functional importance and were found to be involved in different body functions. Titin and Dystrophin was found to play role in translation, were classified as an enzyme, helps in immune response and acts as lyases. Filamin was functionally categorized as purines and pyrimidines, was classified as an enzyme, acts as lyases and important structural protein. Spectrin was known to be involved in regulatory functions, was classified as nonenzyme and acts as an important growth factor. M1 was found to play role in amino acid biosynthesis, was classified as an enzyme and acts as ligase. M2 was found to play role in energy metabolism, acts as an enzyme and helps in transcription regulation. Nebulin was known to be involved in regulatory functions, was classified as non enzyme and plays role in transcription. Actinin was found to play role in translation, was classified as nonenzyme and acts as an important growth factor. Gelsolin was found essential in central intermediary metabolism, was classified as an enzyme and acts as hydrolases. Actin was found to play role in amino acid biosynthesis, was classified as an enzyme and acts as ligase.

### Table 4: Prediction of secondary structure features of fish muscle proteins.

| PMDB ID   | Helices | Strands | Coils | Acidic Residues | Basic Residues | Domains | Motif |
|-----------|---------|---------|-------|-----------------|----------------|---------|-------|
| PM0078304 | 37      | 28      | 39    | 12              | 7              | 1       | 16    |
| PM0078303 | 71      | 0       | 40    | 12              | 17             | 1       | 148   |
| PM0078298 | 15      | 0       | 26    | 2               | 3              | 1       | 17    |
| PM0078299 | 158     | 252     | 321   | 100             | 83             | 6       | 102   |
| PM0078300 | 12      | 86      | 93    | 18              | 19             | 2       | 147   |
| PM0078301 | 139     | 0       | 117   | 86              | 78             | 6       | 108   |
| PM0078302 | 33      | 0       | 10    | 10              | 9              | 1       | 26    |
| PM0078305 | 67      | 0       | 10    | 10              | 9              | 0       | 125   |

### Table 5: Summarized table of total number of Beta turns, cysteines, disulphide bond predicted by Net turn P and DIANNA.

| Protein Name | No. of Beta Turns | No. of Predicted Cysteines | No. of Predicted Disulfide Bonds |
|--------------|-------------------|----------------------------|----------------------------------|
| Actin        | 21                | 4                          | 0                                |
| Actinin      | 0                 | 2                          | 1                                |
| Filamin      | 766               | 21                         | 10                               |
| Gelsolin     | 0                 | 9                          | 4                                |
| M1           | 30                | 5                          | 2                                |
| M2           | 0                 | 11                         | 5                                |
| Plastin      | 170               | 8                          | 4                                |
| Spectrin     | 32                | 2                          | 1                                |
| Titin        | 47                | 3                          | 1                                |
| Dystrophin   | 10                | 0                          | 0                                |
| Thymosin     | 1                 | 0                          | 0                                |

Citation: Khalid S, Idrees S, Khalid H, Hussain B, Tiwari S, et al. (2015) *Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches*. MOJ Proteomics Bioinform 2(3): 00047. DOI: 10.15406/mojpb.2015.02.00047
Table 6: Protein function predicted by ProtFunc.

| Protein | Protein Function Predicted by ProtFunc |
|---------|----------------------------------------|
| Titin   | Play role in translation, classified as an enzyme, help in immune response, and acts as lyases. |
| Dystrophin | Play role in translation, classified as an enzyme, help in immune response, and acts as lyases. |
| Filamin | Functionally categorized as purines and pyrimidines, classified as an enzyme, acts as lyases and important structural protein. |
| Spectrin | Known to be involved in regulatory functions, classified as nonenzyme, acts as an important growth factor. |
| M1      | Play role in amino acid biosynthesis, classified as an enzyme, act as a ligase. |
| M2      | Play role in energy metabolism, acts as an enzyme, known to be involved in transcription regulation. |
| Nebulin | Known to be involved in regulatory functions, classified as non enzyme, play role in transcription. |
| Actinin | Play role in translation, classified as nonenzyme, acts as a growth factor. |
| Gelsolin | Play role in central intermediary metabolism, classified as an enzyme, acts as hydrolases. |
| Actin   | Play role in energy metabolism, classified as an enzyme and acts as an important growth factor. |
| Troponin | Play role in translation, classified as nonenzyme. |
| Thymosin | Play role in translation, classified as nonenzyme, acts as an important hormone. |

Submission of the model in protein model database (PMDB)

The models generated for actin, actinin, dystrophin, gelsolin, M2 protein, plastin 3, thymosin, troponin was successfully submitted in Protein model database (PMDB) and can be find using PM0078304, PM0078303, PM0078298, PM0078299, PM0078300, PM0078301, PM0078302 and PM0078305.

Phylogenetic analysis of fish muscle proteins

By inferring phylogeny novel type of relationship was predicted among species including Amphiichthys koelzi, Oryzias latipes, Dicentrarchus labrax, Plecoglossus altivelis, Daniorerio, Salmosalar, Macrobrachium rosenbergii and Anisakis simplex. Comparative study of actin, actinin, plastin 3 or fimbrin, gelsolin, myosin, spectrin, tropomyosin and troponin fish protein revealed the genetic divergence in to two major lineages. Phylogenetic topology of titin and dystrophin muscle protein revealed the genetic divergence into four lineages (Figure 9-18).

Figure 9: Phylogenetic tree of Actin
Figure 10: Phylogenetic tree of Actinin
Figure 11: Phylogenetic tree of Dystrophin; Figure 12: Phylogenetic tree of Fimbrin; Figure 13: Phylogenetic tree of Gelsolin; Figure 14: Phylogenetic tree of Myosin
Figure 15: Phylogenetic tree of Spectrin; Figure 16: Phylogenetic tree of Titin; Figure 17: Phylogenetic tree of Troponymosin; Figure 18: Phylogenetic tree of Troponin

Citation: Khalid S, Idrees S, Khalid H, Hussain B, Tiwari S, et al. (2015) Ab-Initio Prediction of Sequence and Structural Biology of Fish Muscle Proteins Using Homology Modeling, Phylogeny and Different Computational Approaches. MOJ Proteomics Bioinform 2(3): 00047. DOI: 10.15406/mojpb.2015.02.00047
Statistical evaluation of phylogenetic tree

To computes the statistical measurements related to diversity and divergence from pairwise distance DIVEIN (Table 7) was used.

Table 7: Summarized table with statistical measurements of phylogenetic tree including protein, number of taxa, likelihood log, parsimony, tree size, gamma Shape parameter, mean, standard deviation and median analyzed by DIVEIN server.

| Sr. # | Proteins | No. of taxa | Log Likelihood | Parsimony | Tree Size | Gamma Shape Parameter | Mean | S.D | Median |
|-------|----------|-------------|----------------|------------|-----------|------------------------|------|-----|--------|
| 1     | Actin    | 11          | -1038.71898    | 39         | 0.76396   | 0.529                 | 0.1401248 | 0.2126536 | 0.0354923 |
| 2     | Actinin  | 12          | -49.07.2736    | 354        | 1.26628   | 0.585                 | 0.325363  | 0.1710674 | 0.3750267 |
| 3     | Dystrophin| 14          | -748.1.6288    | 3435       | 13.20001  | 2.06809               | 0.4517532 | 0.4039974 | 0.2799059 |
| 4     | Filbrin  | 10          | -423.3.815     | 448        | 0.92222   | 0.799                 | 0.2598686 | 0.147638  | 0.3098932 |
| 5     | Gelsolin | 11          | -597.6.7554    | 691        | 1.37231   | 0.699                 | 0.3301184 | 0.2391237 | 0.1970655 |
| 6     | Spectrin | 11          | -958.7.4472    | 363        | 0.36001   | 0.153                 | 0.0774055 | 0.0494607 | 0.0822327 |
| 7     | Myosin   | 18          | -1111.1.806    | 984        | 1.16158   | 0.784                 | 0.2049471 | 0.1153865 | 0.1627676 |
| 8     | Titin    | 26          | -875.18009     | 49         | 0.30294   | 0.897                 | 0.0445636 | 0.0248786 | 0.0452123 |
| 9     | Troponymosin| 25        | -2981.71234    | 400        | 1.75808   | 0.471                 | 0.1964246 | 0.1039309 | 0.1429878 |
| 10    | Troponin | 17          | -3877.24952    | 490        | 5.14144   | 0.352                 | 0.5318019 | 0.2927219 | 0.3902863 |

BindN [3] was used for prediction of DNA and RNA binding residues in order to understand the function of DNA and RNA binding proteins. Filamin protein was found to have greater number of DNA and RNA binding residues. In filamin 262 DNA residues with 303 RNA residues were predicted. In plastin 3 protein 103 DNA and 122 RNA residues were found. In troponin predicted DNA residues were 81 and RNA residues were 121 in number. Thus BindN showed that selected fish muscle proteins are good binding proteins. ConSurf [4] was explored for estimation of evolutionary conserved amino acids in protein which was based on phylogenetic relationship inferred from homologous sequences. In actin number of functional residue predicted was 21 whereas in myosin 17 residues, in dystrophin 6, in titin 33, in spectrin 32, in M2 protein 26 amino acids were highly conserved and exposed. Filamin protein was found to have high number of functionally conserved amino acids with 225 residues. Study of conserved position of these amino acids contributes to structural and functional knowledge. Thus from ConSurf study it was found these muscle proteins have structural and functional importance.

DIAANNA [5] was used for cysteine classification and prediction of disulfide connectivity. In gelsolin, plastin 3 and M2 protein four disulfide bonds were predicted. In M1 protein 2 disulfide bonds whereas in spectrin 2 and titin 1 disulfide bond was predicted. Filamin protein was found to have greater number of disulfide bond. Two cysteines were predicted in spectrin 2, and actin. In titin 3, in plastin 38, in M2 protein 11, in M1 protein 5, in gelsolin 9 and in filamin 21 cysteines were predicted. This knowledge helps us to understand secondary structure of protein since disulfide bonds play important role for stabilizing the folding process in protein. In addition knowledge of disulfide bond with cysteine also provides information for genome annotation. PROFEAT [6] is a bioinformatics server used for calculating structural and chemical features of protein from primary sequence data. These features provide knowledge about biological properties of proteins and peptides. Thus in order to compute the structural and physicochemical features of proteins and peptides PROFEAT was used. All fish muscle proteins were found as non allergen.

DEEP VIEW [11] was used for analyzing secondary structure features such as coils, ribbons, acidic and basic residues. In gelsolin 158 helices, 252 strands and 321 coils were predicted. In actin 37 helices, 28 strands and 39 coils were predicted. In actinin 71 helices and 40 coils were predicted. Dystrophin was found to contain 15 helices and 26 coils. In M2 protein 12 helices, 86 strands and 93 coils were predicted. 139 helices and 117 coils were predicted from plastin 3D model. In case of thymosin 33 helices where as in troponin 67 helices were predicted with 10 coils in both proteins. PSIPRED integrates several protein structure prediction methods on one platform. PSIPRED [9] was used for prediction of protein structure, transmembrane topology prediction and for recognition of folds and domains. Homology modeling approach was used to predict three dimensional structures. Homology modeling refers to prediction of tertiary structure of protein of interest using template of known 3D structure with homologous sequence. WHAT IF and PSVS [11] was used for structure validation and evaluating stereochemistry of 3D model. The identification of a conserved domain footprint may be the only clue towards cellular or molecular function of a protein, as it indicates local or partial similarity to other proteins, some of which may have been characterized experimentally [15]. Template used for predicting 3D model was 1D4X_A for actin, 1T7T_A for actinin, 1DXX_A for dystrophin, 2FGH_A for gelsolin, 2JDF_A for M2 protein, 1AOA_A for plastin 3, 1HJO_A for thymosin and 1JD_E for tropnin. After validation 3D models were successfully submitted to PMDB [15] as PM0078304, PM0078303, PM0078298, PM0078299, PM0078300, PM0078301, PM0078302 and PM0078305. Protein 3D structure is important in understanding protein interactions, function and their localization [19]. Structure prediction refers to the prediction of 3D structure from its amino acid sequence. Number of motifs found in actin was 16, in actinin 148, in dystrophin 17, in gelsolin 102, in M2 protein 147, in plastin 108, in thymosin 26 and in troponin 125. CDD [14] is a large resource which contains manually curates domain models and provides information about sequence, structural and functional
relationship. Six domains were predicted in gelsolin and plastin 3. In actin, actinin, dystrophin and thymosin one domain was found. The main objective of this study was to explore the structural and functional importance of novel fish muscle proteins.

Fish muscle [1] was found as an excellent model for performing sequence and structural analysis. Sequence analysis was carried out using different bioinformatics tools to understand structure, function and evolution of fish muscle proteins with significant features. Homology modeling technique was applied for predicting 3D structure. This 3D structure is important in understanding protein interaction, function and its localization. Structural knowledge has allowed us to identify functionally important residues and disulfide linkages. Furthermore 3D knowledge of proteins will contribute to design efficient drugs. Phylogenetic analysis of ten fish muscle proteins including actin, actinin, filamin, dystrophin, myosin, gelsolin, titin, spectrin, tropomyosin, and troponin were performed. In order to construct phylogenetic trees sequences were aligned by Clustal X using gap opening penalty 30, gap extension penalty 20 and GONNET protein weight matrix [16]. The phylogenetic tree was generated in MEGA 4 using maximum Likelihood approach [17]. The bootstrap was performed using 1000 replications [20]. Thus, novel type of relationship was predicted among species including Amphichthys koelzi, Oryzias latipes, Dicentrarchus labrax, Plecoglossus altivelis, Danio rerio, Salmo salar, Macrobrachium rosenbergii and Anisakis simplex.

Comparative study of actin, actinin, plastin3 or filamin, gelsolin, myosin, spectrin, tropomyosin and troponin fish protein revealed the genetic divergence into two major lineages. Phylogenetic topology of titin and dystrophin muscle protein revealed the genetic divergence into four lineages. The phylogenetic study have application in various fields of biology including systematic, bioinformatics and comparative genomics. Statistically phylogenetic trees were analyzed by DIVEIN predicting number of taxa, values of log likelihood, gamma shape parameter, mean, standard deviation and median. Titin was found to include highest number of taxa, 26 species a and smaller number of taxa was observed in Fimbrin protein with 10 species. This comparative study will be beneficial for predicting the function of individual genes and mechanism of inherited diseases by comparing the genetic material of different species.

Conclusion

Overall evidence from in silico approaches revealed that fish muscle proteins have structural and functional significance. Future functional research can be conducted via exploring the proteins of model organisms for using it as a diagnostic tool for designing effective vaccines utilizing structure based drug designing approach.

Acknowledgement

We acknowledge the entire team member specially Prof. Vasco Azevedo and Prof. Bilal Hussain for their support.

References

1. Fernandez DA, Calvo J (2009) Fish muscle: the exceptional case of Notothenioids. Fish Physiol Biochem 35(1): 43-52.

2. Campbell AN (2006) Comparative Proteomics Kit 1: Protein Profile Module. Bio Rad Laboratories pp. 1049-1051.

3. Wang L, Brown SJ (2006) BindN: a web based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. Nucleic Acids Res 34(2): 243-248.

4. Ashkenazy H, Erez E, Martz E, Pepko T, Ben-Tal N (2010) Consurf2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 38(Web Server Issue): W529-W533.

5. Ferré F, Cote P (2005) DIANNA: a web server for disulfide connectivity prediction, Nucleic Acids Res 33(Web Server Issue): W250-W252.

6. Rao HB, Zhu E, Yang GB, Li ZR, Chen YZ (2011) Update of PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequences, Nucleic Acids Res 39(Web Server issue): W385-W390.

7. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy: the proteomics server for in depth protein knowledge and analysis, Nucleic Acids Res 31(13): 378-4378.

8. Arnold K, Kiefer F, Kopp J, Battey JN, Podvinec M, et al. (2009) The Protein Model Portal. J Struct Funct Genomics 10(1): 1-8.

9. Kousounadis A, Redfern OC, Jones DT (2009) Improving classification in protein structure databases using text mining, BMC Bioinformatics 129.

10. Petersen B, Lundegaard C, Petersen TN (2010) NetTurnP: Neural Network Prediction of Beta turns by use of Evolutionary information and Predicted Protein Structure Features, PLoS One 5(11): e11.15079.

11. Idrees S, Nadeem S, Kanwal S, Elsan B, Yousaf A, et al. (2012) In silico sequence analysis, homology modeling and function annotation of Ocmimbusbinalis hypothetical protein GCT28_OOCA, Journal of Bioautomation 16(2): 111-118.

12. Chen CC, Hwang JK, Yang JM (2006) P52: protein structure prediction server, Nucleic Acids Res 34(Web Server Issue): 152-157.

13. Jensen LJ, Gupta R, Blom N, Deves D, Tamames J, et al. (2002) Prediction of human protein functions from post translational modifications and localization features, J Mol Biol 319(5): 1257-1265.

14. Marchler-Bauer A, Lu S, Anderson JR, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a conserved domain database for the functional annotation of proteins, Nucleic Acids Res 39(Database Issue): D225-D229.

15. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia, Proteins 66(4): 778-795.

16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA (2007) Clustal W and Clustal X version 2.0, Bioinformatics 23(21): 2947-2948.

17. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist centric analysis, Nucleic Acids Res 31(13): 3784-3790.

18. Arnold K, Kiefer F, Kopp J, Battey JN, Podvinec M, et al. (2009) The Protein Model Portal. J Struct Funct Genomics 10(1): 1-8.

19. Koussounadis A, Redfern OC, Jones DT (2009) Improving classification in protein structure databases using text mining, BMC Bioinformatics 129.

20. Petersen B, Lundegaard C, Petersen TN (2010) NetTurnP: Neural Network Prediction of Beta turns by use of Evolutionary information and Predicted Protein Structure Features, PLoS One 5(11): e11.15079.

21. Idrees S, Nadeem S, Kanwal S, Elsan B, Yousaf A, et al. (2012) In silico sequence analysis, homology modeling and function annotation of Ocmimbusbinalis hypothetical protein GCT28_OOCA, Journal of Bioautomation 16(2): 111-118.

22. Chen CC, Hwang JK, Yang JM (2006) P52: protein structure prediction server, Nucleic Acids Res 34(Web Server Issue): 152-157.

23. Jensen LJ, Gupta R, Blom N, Deves D, Tamames J, et al. (2002) Prediction of human protein functions from post translational modifications and localization features, J Mol Biol 319(5): 1257-1265.

24. Marchler-Bauer A, Lu S, Anderson JR, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a conserved domain database for the functional annotation of proteins, Nucleic Acids Res 39(Database Issue): D225-D229.

25. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia, Proteins 66(4): 778-795.

26. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA (2007) Clustal W and Clustal X version 2.0, Bioinformatics 23(21): 2947-2948.

27. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist centric analysis, Nucleic Acids Res 31(13): 3784-3790.

28. Deng W, Maust RS, Nickle DC, Lernach GH, Liu Y, et al. (2010) DIVEIN: A web server to analyze Phylogenies, Sequence divergence, Diversity and Informative sites, Biotechniques 48(S): 405-408.

29. Jones DT (2001) Protein structure prediction in genomics. Brief Bioinform 2(2): 111-125.

30. Pavlopoulos GA, Soldatos TG, Barbosa-Silva A, Schneider R (2010) A web server for computing structural and physicochemical features of proteins and peptides from amino acid sequences, Nucleic Acids Res 38(Web Server Issue): W385-W390.