Journal of Applied Sciences and Nanotechnology

Journal homepage: jasn.uotechnology.edu.iq

Bioinformatics Analysis of BMP15 Gene and Bone Morphological Protein-15 Sequence

[1]Maryam D. Kamel[*], [1]Ali A. Issa, [1]Rand J. Sattar, [1]Rusul H. Sami, [1]Yasser A. Hussein, [2]Doaa S. El-Sayed, [1]Adawiya J. Haider, [1]Bilal H. Jasim

[1]Department of Applied Sciences, University of Technology - Iraq
[2]Department of Chemistry, Faculty of Science, Alexandria University - Egypt

A R T I C L E I N F O

Article history:
Received: March, 16, 2024
Accepted: May, 27, 2024
Available online: June, 10, 2024

Keywords:
BMP15,
SOPMA,
ProtParam,
Swiss model

Corresponding Author:
Maryam D. Kamel
maryam.d.kamel@uotechnology.edu.iq

A B S T R A C T

This study characterized the physicochemical properties and predicted the structure, function, and protein-protein interactions of bone morphogenetic protein 15 (BMP15) using bioinformatics tools. BMP15 was found to have a molecular weight of 45055.01 Daltons, sequence length of 392 amino acids, and an extinction coefficient of 55390 at 280 nm with a basic isoelectric point. Secondary structure analysis revealed BMP15 consists mostly of random coil (63.78%), followed by alpha helix (20.66%) and extended strand (15.56%) as well as beta turns. Amino acids with high coil structure like glycine and alanine, which are hydrophobic and flexible, represented the highest concentrations. Transmembrane helix prediction identified four helices located from inside to outside and three from outside to inside. SWISS-MODEL generated four protein structure models corresponding to sequences (Q6PX77.1. A, 5vqf.2. A, 5ntu.1. A, and 5hly.1. A) with sequence identities of (75.38%, 20.83%, 20.77% and 20.33%) respectively. Results correlate BMP15 with oocyte maturation and granulosa cell activation in follicular development. This comprehensive bioinformatics analysis of BMP15 properties, structure, and interactions provides a framework for further study of genetically inherited infertility, drug design, and new protein analysis.

https://doi.org/10.53293/jasn.2024.7315.1280, Department of Applied Sciences, University of Technology - Iraq.

1. Introduction

Proteins are components belonging to the Transforming Growth Factor (TGF-β) superfamily that includes a set of signalling molecules known as bone morphogenetic proteins (BMPs) such as Fibroblast Growth Factors (FGFs), Growth Differentiation Factor 9 (GDF9), Bone Morphogenetic Protein 6 (BMP6), and Bone Morphogenetic Protein 15 (BMP15). These paracrine factors are produced from oocytes (ODPFs) of the mammalian. They are crucial for various cellular processes such as growth, differentiation and development [1, 2]. Bone Morphogenetic Protein 15 (BMP-15), which is encoded by the BMP15 gene in humans, is considered an essential protein for ovarian folliculogenesis regulation, plays a role in maintaining adult tissue homeostasis, inhibits the expression of follicle-stimulating hormone receptors, and increases the production of kit ligands as well as crucial for

embryogenesis and evolvement. Moreover, BMP15 affects female fertility and reproductive health because it influences granulosa cell mitosis and thus controls granulosa cell proliferation and differentiation [3]. The cytogenetic location of the BMP15 gene is at Xp11.22 [4], Which is located on the short arm (p) of chromosome X (sexual chromosome) at position 11.22 (band eleventh sub-band two, sub-sub-band two). BMP15 consists of 392 amino acids divided between two exons; the first exon encodes for 109 amino acids, while the second exon encodes for 283 amino acids. Both of the exons from the homo-dimer structure of the BMP15 protein [5]. Sequence analysis can be used to specify the properties of various regions, such as active sites and functional areas of the DNA, and to interpret the evolutionary relation between the categorised groups (phylogenetic analysis) [6], and contrasting an unknown sequence with millions of other sequences found in the DNA information gene banks [7]. However, no previous study has investigated using bioinformatics tools to analyse the BMP15 protein. Therefore, this research aims to propose a methodology for protein-protein interaction prediction and protein-ligand identification from the sequences of amino acids, and it classifies them by the protein affinities to ligands.

2. Computational Strategy

2.1. Study Design

The procedure of the research project is shown in the flowchart diagram, Fig. 1:

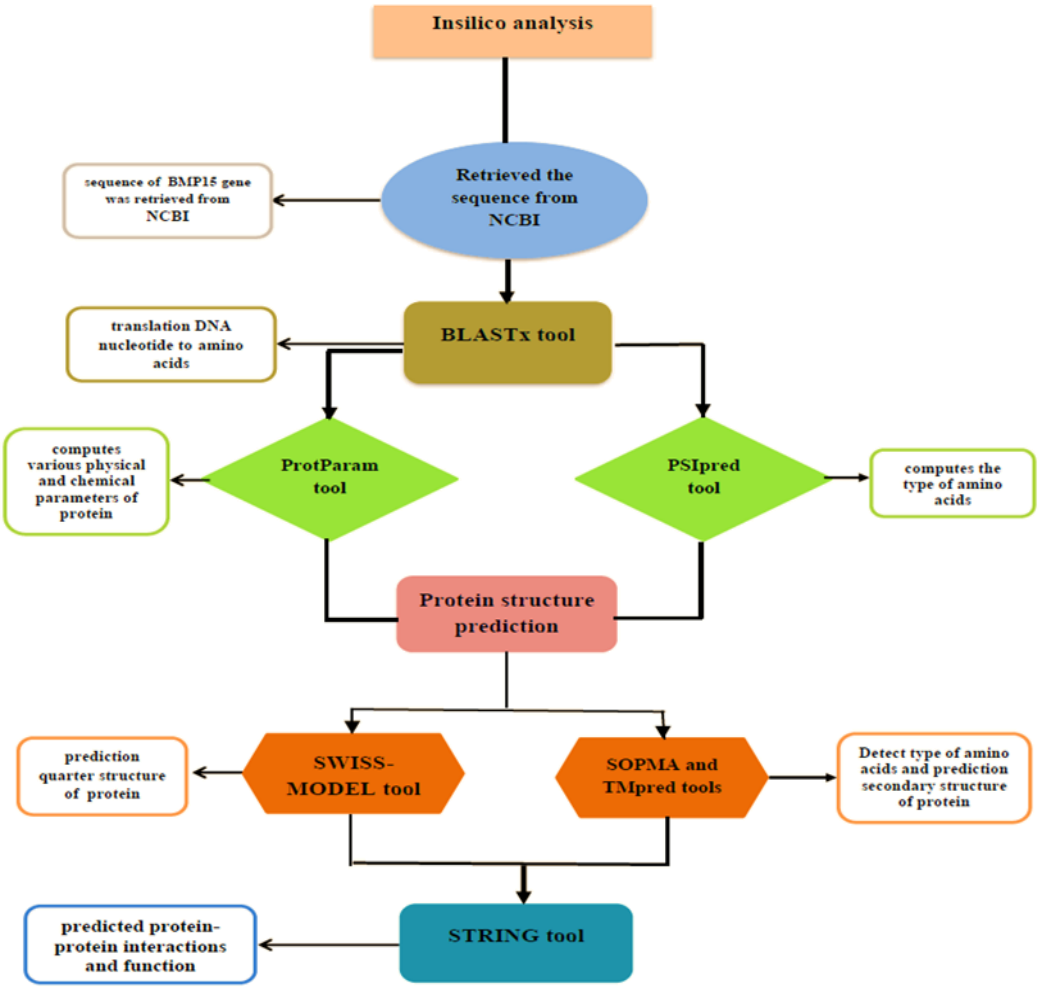


Figure 1: The flowchart diagram of the study design.

2.2. Insilco Study

Many web servers were utilised in this study for sequence retrieval, translation of DNA nucleotide, protein sequence analysis such as computing physicochemical properties, and Protein Structures Prediction [8]. In silico inspection, a program is used on the National Centre for Biotechnology Information (NCBI) website.

2.3. Gene Sequence Retrieval

The Sequences of the BMP15 gene retrieved in FASTA format from NCBI.

2.4. Converted the DNA Sequence of The BMP15 Gene to Amino Acids through a Translation Tool

using Augustus server version 4.0 (http://augustus.gobics.de/submission) translation of DNA nucleotide to amino acids [9].

2.5. Compute Physicochemical Properties Prediction

Protein Structure Prediction Server (PSIPRED) tool to detect the type of amino acids of the protein structure [10]. ProtParam (protein parameters) online tool to identify the BMP15 protein's physicochemical characteristics for the human sequence retrieved from NCBI. The parameters calculated by Prot Param are the molecular weight, theoretical pI, atomic composition, amino acid composition, instability index, aliphatic index, and grand average of hydropathicity (GRAVY) [11].

2.6. Protein Structure Prediction

A self-optimized prediction method with alignment, SOPMA (version 3.0), was used to predict the secondary structure [12], and using the TMpred program to determine the Hydropathy plot for BMP15 [13]. When the sequence is provided in FASTA format, the structural details of the protein sequence are offered, such as B-sheets, helices, and coils. Four fundamental processes are involved in the modelling process: finding a structure that exhibits homology with the target, choosing the optimal template with the highest degree of identity with the target sequence once it has been aligned with the target, and modelling the structure, using.

SWISS-MODEL is a service that models protein structural homology completely automatically. Using homology modelling techniques, the Swiss-model automated system models the quaternary structure of a protein based on its amino acid sequence [14].

2.7. Protein-Protein Interaction

Using the STRING (search tool for the retrieval of interacting genes/proteins) program Version: 12.0 to predict protein-protein interactions and function has been predicted [15]. When the input was provided in the FASTA format of the amino acid sequence [16].

3. Results and Discussion

3.1. The BMP15 Gene Sequence Retrieved from NCBI

The sequence of the BMP15 gene consists of 2 exons and one intron, which is composed of 1679 bp of DNA; the sequence of the exons consists of 1227 Pb. This codon region of the BMP15 gene encodes to Bone Morphogenetic Protein 15, which consists of 392 amino acids. As shown below in Fig. 2. The 17 amino acid signal peptide and the first segment of the propertied region are encoded by the first exon of the BMP15 protein. In contrast, the second exon encodes the remaining section of the propertied region and the full anticipated 125 amino acids mature domain. The codon region consists of two exons; the first exon encodes for 109 amino acids while the second exon encodes for 283 amino acids together to form the BMP15 protein, which consists of 392 amino acids that form a homo-dimer structure of the protein.

bone morphogenetic protein 15 preproprotein [Homo sapiens]

Sequence ID: NP_005439.2 Length: 392 Number of Matches: 1

See 6 more title(s) See all Identical Proteins(IPG)

Range 1: 1 to 392 GenPept Graphics ▼ Next Match ▲ Previ

Score	Expect	Method	Identities	Positives	Gaps
815 bits(2105)	0.0	Compositional matrix adjust.	392/392(100%)	392/392(100%)	0/392(0%)

```
Query  1    MVLLSILRILFLCELVLFMEHRAQMAEGGQSSIALLAEAPTLPLIEELLEESPGEQPRKP  60
            MVLLSILRILFLCELVLFMEHRAQMAEGGQSSIALLAEAPTLPLIEELLEESPGEQPRKP
Sbjct  1    MVLLSILRILFLCELVLFMEHRAQMAEGGQSSIALLAEAPTLPLIEELLEESPGEQPRKP  60

Query  61   RLLGHSLRYMLELYRRSADSHGHPRENRTIGATMVRLVKPLTNVARPHRGTWHIQILGFP  120
            RLLGHSLRYMLELYRRSADSHGHPRENRTIGATMVRLVKPLTNVARPHRGTWHIQILGFP
Sbjct  61   RLLGHSLRYMLELYRRSADSHGHPRENRTIGATMVRLVKPLTNVARPHRGTWHIQILGFP  120

Query  121  LRPNRGLYQLVRATVVYRHHLQLTRFNLSCHVEPWVQKNPTNHFPSSEGDSSKPSLMSNA  180
            LRPNRGLYQLVRATVVYRHHLQLTRFNLSCHVEPWVQKNPTNHFPSSEGDSSKPSLMSNA
Sbjct  121  LRPNRGLYQLVRATVVYRHHLQLTRFNLSCHVEPWVQKNPTNHFPSSEGDSSKPSLMSNA  180

Query  181  WKEMDITQLVQQRFWNNKGHRILRLRFMCQQQKDSGGLELWHGTSSLDIAFLLLYFNDTH  240
            WKEMDITQLVQQRFWNNKGHRILRLRFMCQQQKDSGGLELWHGTSSLDIAFLLLYFNDTH
Sbjct  181  WKEMDITQLVQQRFWNNKGHRILRLRFMCQQQKDSGGLELWHGTSSLDIAFLLLYFNDTH  240

Query  241  KSIRKAKFLPRGMEEFMERESLLRRTRQADGISAEVTASSSKHSGPENNQCSLHPFQISF  300
            KSIRKAKFLPRGMEEFMERESLLRRTRQADGISAEVTASSSKHSGPENNQCSLHPFQISF
Sbjct  241  KSIRKAKFLPRGMEEFMERESLLRRTRQADGISAEVTASSSKHSGPENNQCSLHPFQISF  300

Query  301  RQLGWDHWIIAPPFYTPNYCKGTCLRVLRDGLNSPNHAIIQNLINQLVDQSVPRPSCVPY  360
            RQLGWDHWIIAPPFYTPNYCKGTCLRVLRDGLNSPNHAIIQNLINQLVDQSVPRPSCVPY
Sbjct  301  RQLGWDHWIIAPPFYTPNYCKGTCLRVLRDGLNSPNHAIIQNLINQLVDQSVPRPSCVPY  360

Query  361  KYVPISVLMIEANGSILYKEYEGMIAESCTCR  392
            KYVPISVLMIEANGSILYKEYEGMIAESCTCR
Sbjct  361  KYVPISVLMIEANGSILYKEYEGMIAESCTCR  392
```

Figure 2: Shown the result of the translation of exons of the BMP15 gene to the amino acid sequences of Bone morphogenetic protein 15 via the BLASTx program.

3.2. Primary Sequence Analysis

As shown in Fig. 3, the primary structure analysis results indicate that the BMP15 consists of non-polar and hydrophobic residues, which include Tryptophan, Alanine, Phenylalanine, Glycine, Isoleucine, Leucine, Methionine, Proline, and Valine. The hydrophilic residues include Cysteine, Asparagine, Glutamine, Serine, Threonine, and Tyrosine. The BMP15 showed a hydrophobic nature because hydrophobic residues represent the highlight contents that count 188 compared to the lower contents of the hydrophilic residues count of 106 [17, 18].

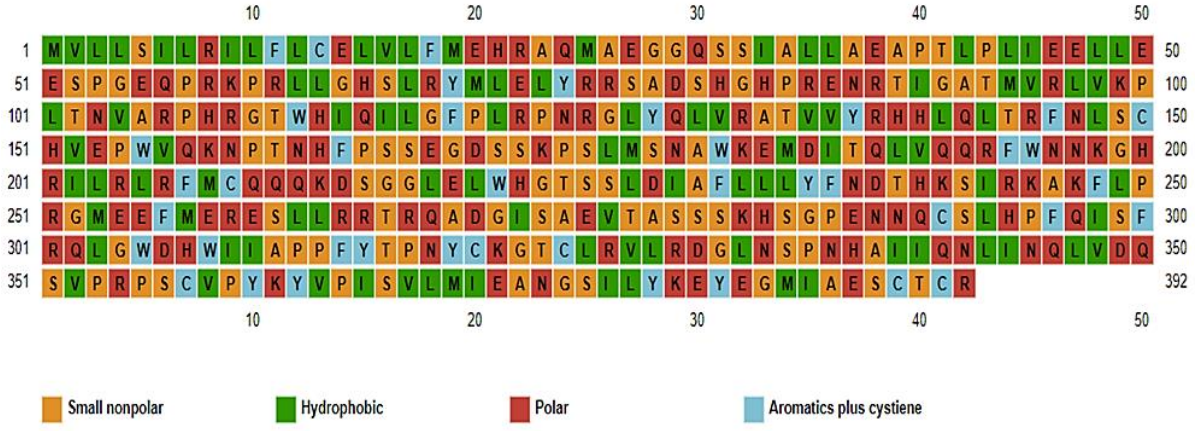


Figure 3: The types of BMP15 protein amino acids predicted via the PSIpred tool.

3.3. Physio-Chemical Properties of BMP15

Table 1 illustrates the findings of predicting physiochemical characteristics with ProtParam using the amino acid sequence.

Table 1: The physio-chemical properties of BMP15 protein by ProtParam tool.

Parameter	BMP15
Number of amino acids	392
Molecular weight	45055.01
Theoretical pI	9.28
Total number of negatively charged residues (Asp + Glu)	35
Total number of positively charged residues (Arg + Lys)	46
Total number of atoms	6333
Extinction coefficients	55390
half-life	30 hours
instability index (II)	52.69
Aliphatic index	88.80
Grand average of hydropathicity (GRAVY):	-0.345

Protein structure stability can be predicted by analysing physico-chemical parameters with the ProtParam program. An instability index value was used to measure the strength of the protein structure. If the index value is lower than 40, it refers to a stable protein, while a higher index value of 40 refers to an unstable protein. Results show that Bone morphogenetic protein 15 has a molecular weight of 45055.01 daltons with a sequence length of 392. The BMP15 protein recorded an extinction coefficient of 55390 at 280nm, indicating that the protein can absorb light at 280 nm. The extinction coefficient can be used to calculate the concentration of a protein in the solution. The pH level where a protein surface is charged, however, its net charge remains zero, is known as the isoelectric point (PI). The BMP15 protein displayed a basic PI. The isoelectric point of the basic buffers is 9.28, which is soluble and thus can be utilised in buffer systems for protein purification [19]. A protein aliphatic index is a relative volume occupied by its aliphatic side chains, including leucine, isoleucine, valine, and alanine. It was determined at 88.80 and might be evidence for thermos-stability. The Grand Average hydropathy (GRAVY) values demonstrated the protein's hydrophobic character. The BMP15 showed insoluble property due to its GRAVY value of -0.345 [20]. The negative value of protein sequences may suggest that the protein is stable. Hydrophobic amino acids may be implicated in ligand binding and recognition.

3.4. Secondary Structure Prediction

The SOPMA tool was used to predict the protein's secondary structure through computational methods. The program was used to receive the protein main sequence for the SOPMA examination. The previous program provided details on the helices, coils, and strands of bone morphological protein 15. SOPMA analyses predicted the secondary structure of proteins shown in Fig. 4. The secondary structure analysis revealed that the helix structure extended 81 residues, representing 20.66% of the protein. On the other hand, the strand structure extended 61 residues, representing 15.56% of the protein, while the random coil structure extended 250 residues, representing 63.78% of the protein, as shown in Fig. 5. This study concluded that random coils have the highest structure, followed by helix structure, whereas strand structure has the lowest structure in bone morphological protein 15.

```
        10        20        30        40        50        60        70
         |         |         |         |         |         |         |
MVLLSILRILFLCELVLFMEHRAQMAEGGQSSIALLAEAPTLPLIEELLEESPGEQPRKPRLLGHSLRYM
hhhhhhhheeecccchhhhhhhccccccccchhhhhcccchhhhhhhhhcccccccccccccchhhh
LELYRRSADSHGHPRENRTIGATMVRLVKPLTNVARPHRGTWHIQILGFPLRPNRGLYQLVRATVVYRHH
hhhhhhccccccccccccccchheeehhccccccccccccceeeeeeecccccchhhhhhhhhhccc
LQLTRFNLSCHVEPWVQKNPTNHFPSSEGDSSKPSLMSNAWKEMDITQLVQQRFWNNKGHRILRLRFMCQ
ccccccccceeeeecccccccccchhhhhccccccccceeeeehhhhhhhcccccccceeeeeeccc
QQKDSGGLELWHGTSSLDIAFLLLYFNDTHKSIRKAKFLPRGMEEFMERESLLRRTRQADGISAEVTASS
ccccccccccccccccccceeeeeccccccccchhccccccccccccccccccccccccccc
SKHSGPENNQCSLHPFQISFRQLGWDHWIIAPPFYTPNYCKGTCLRVLRDGLNSPNHAIIQNLINQLVDQ
ccccccccccccceeeeeccccccceeccccceeeecccccccccccccchhhhhhhhhhccc
SVPRPSCVPYKYVPISVLMIEANGSILYKEYEGMIAESCTCR
ccccccccccccceeeeeccccceeeehccchheeeccccc

Sequence length :   392
```

Figure 4: Prediction of the secondary structure features (Alpha helix (Hh), Extended strand (Ee) and Random coil (Cc), of bone morphogenetic protein 15 (BMP15) via SOPMA tool.

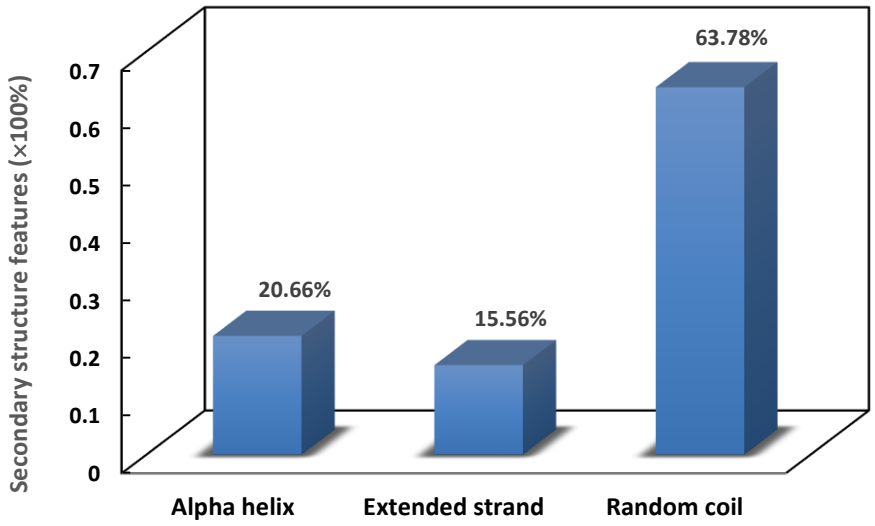


Figure 5: The percentage of the secondary structure features (helix, B-strand, and coil) of BMP15 protein.

Leucine is hydrophobic and thus tends to be buried within hydrophobic protein cores. Moreover, it tends to be within alpha-helices, opposite to the beta strands. Bone morphological protein-15 has a high coil structure due to its composition, which has a high concentration of both hydrophobic glycine and flexible lanine. Additionally, proline is characterised by its ability to break the arrangement of the secondary structures and form links within the polypeptide chains. Most non-polar amino acid side chains, such as Ala, Val, Leu, Ile, and Phe, are found in transmembrane segments. The highly polar CO-NH groups (peptide bonds) are found in the polypeptide backbone of transmembrane segments participating in the hydrogen bonding (H-bonds) to reduce the energetic cost of transferring the peptide bonds into a hydrocarbon interior. This H-bonding is accomplished easily within the alpha-helices because all the peptide bonds are H-bonded internally [21].

3.5. Transmembrane Prediction

One crucial area of bioinformatics is the prediction of transmembrane helices in integral membrane proteins. The most effective techniques to date are to predict individual transmembrane helices and the entire protein topology or the total number of transmembrane helices and their orientation about the membrane [22]. Reliable techniques to predict topology and distinguish between soluble and membrane proteins have significant uses in genome

research because they can be employed to identify general patterns in the evolution of the membrane proteins [21]. TM base is a database of transmembrane proteins and their helical membrane-spanning domains. The transmembrane helices analysis indicated that four helices are located from the inside to the outside among the locations of the accession number sequence. In comparison, three helices are located from the outside to the inside. The findings suggested that transmembrane topology is completely theoretical. However, there is a possibility that all transmembrane helices have been identified, but it might be applied cautiously.

Inside → outside | outside → inside
1- 17 (17) 1365| 1- 17 (17) 1381
(25- 45 (21) 98 ++) |
(221- 239 (19) 343) | 220- 240 (21) 523 +
(357- 378 (22) 89) | (362- 378 (17) 233 +)

The prediction plot should be used for the topology assignment of unknown proteins shown in Fig. 6.

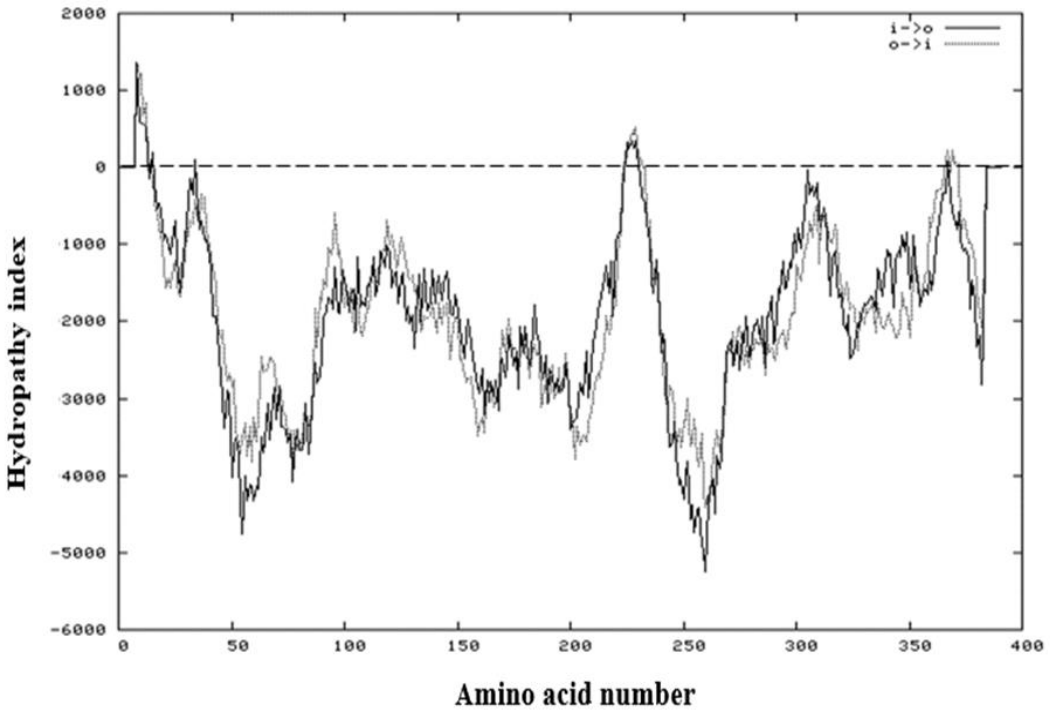
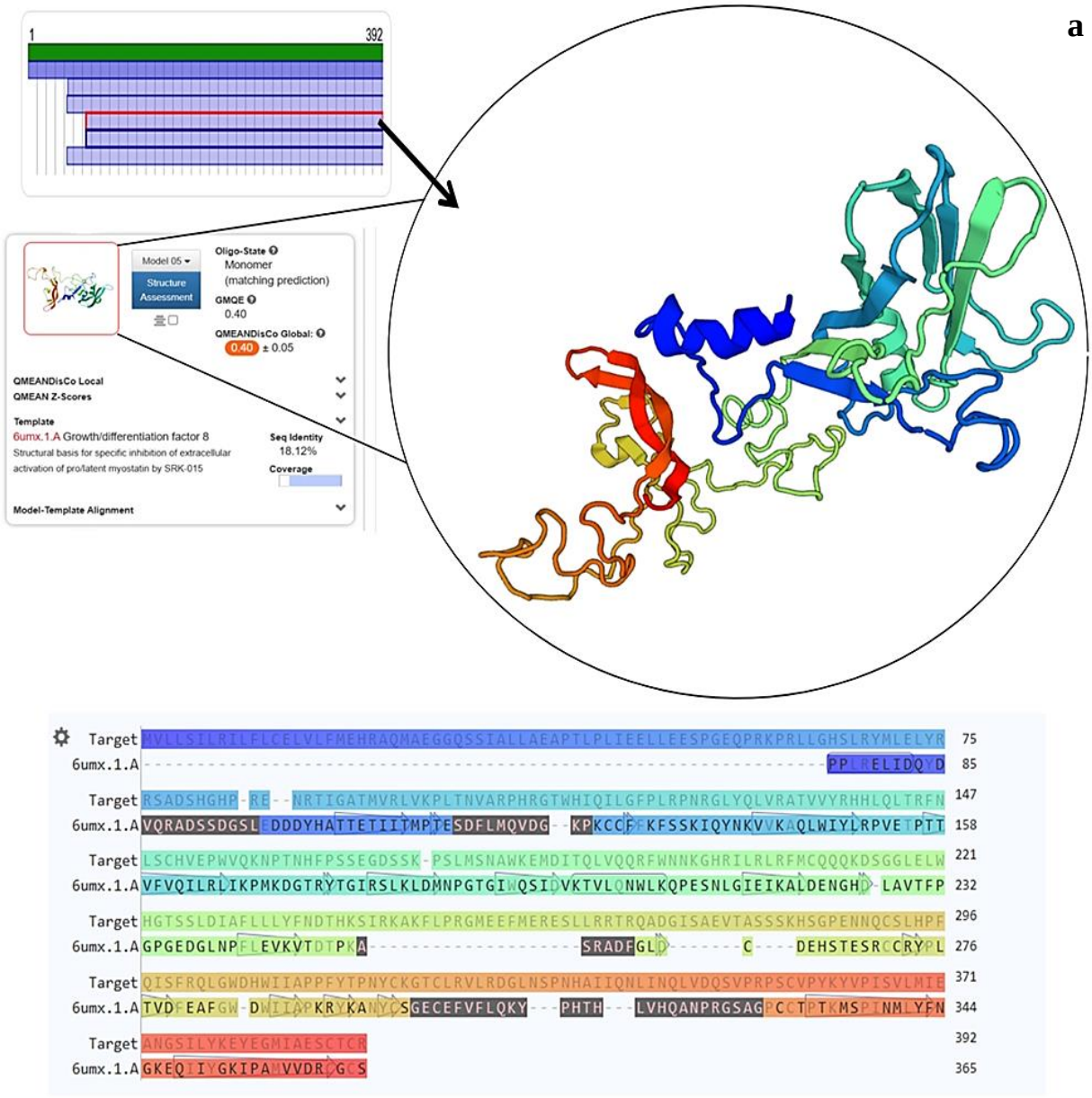


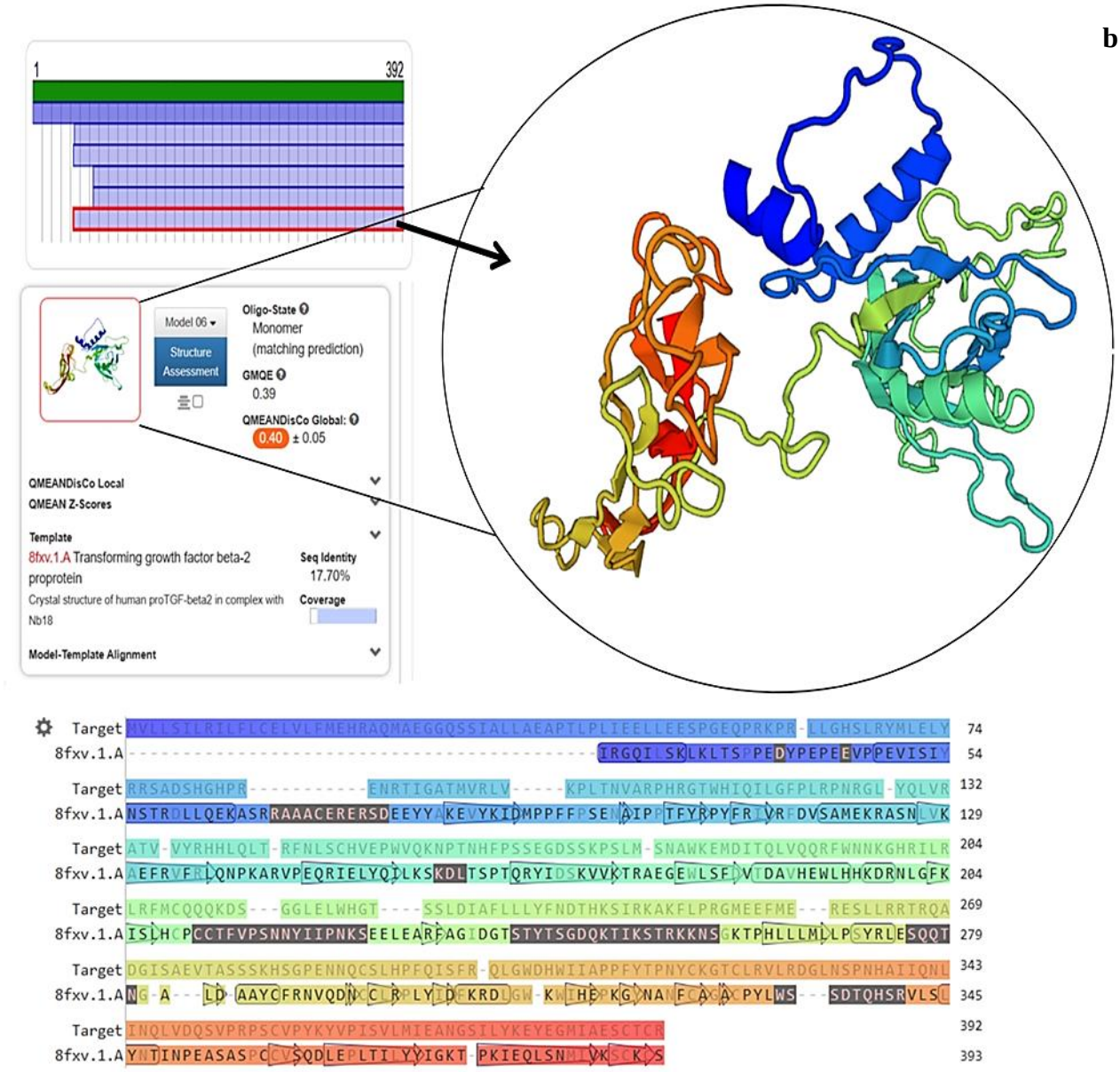
Figure 6: Transmembrane region prediction, TMpred graph of BMP15 using TMPRED tool.

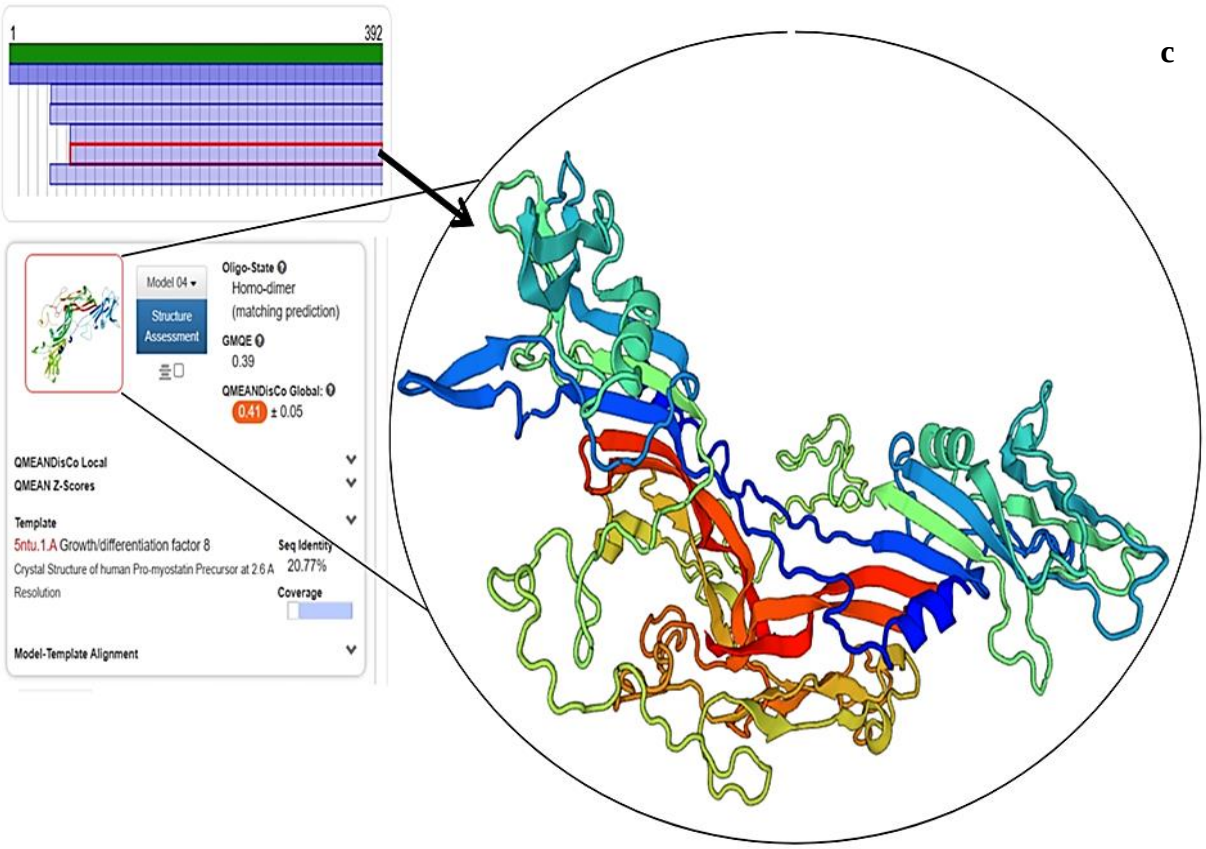
The TMpred program was used to determine the hydropathy plot for BMP15. The hydropathy index is shown on the y-axis, while the amino acid residues in the sequence are shown on the x-axis. Negative values show hydrophilic qualities, and positive values indicate hydrophobic attributes. The present study suggested two potential transmembrane hydrophobic sections separated by a sizable hydrophilic domain. The core region is dominated by the sequence locations shown in brackets. Interestingly, a significant score was higher than 500 (the BMP15 total score was 1888). Therefore, based on the results above, it is possible to predict that the transmembrane is significant.

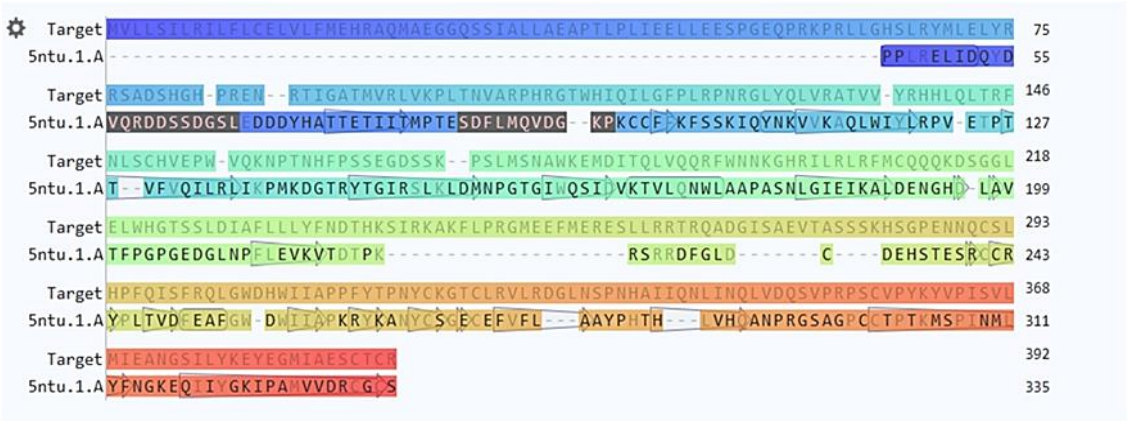
3.6. SWISS Model

The four models in this study have been obtained via a fully automated process that model's construction, alignment, and template selection by the server. This method follows the four phases of the homology modelling approach and is used to predict the three-dimensional structures of the proteins. The arrangement of the top four templates is displayed in Fig. 7.


a
1
392
Model 05
Structure Assessment
Oligo-State
Monomer
(matching prediction)
GMQE
0.40
QMEANDisCo Global:
0.40 ± 0.05
QMEANDisCo Local
QMEAN Z-Scores
Template
6umx.1.A Growth/differentiation factor 8
Structural basis for specific inhibition of extracellular activation of pro/latent myostatin by SRK-015
Seq Identity
18.12%
Coverage
Model-Template Alignment
Target
6umx.1.A
75
85
147
158
221
232
296
276
371
344
392
365


b
1
392
Model 06
Structure Assessment
Oligo-State
Monomer
(matching prediction)
GMQE
0.39
QMEANDisCo Global:
0.40 ± 0.05
QMEANDisCo Local
QMEAN Z-Scores
Template
8fxv.1.A Transforming growth factor beta-2 proprotein
Crystal structure of human proTGF-beta2 in complex with Nb18
Seq Identity
17.70%
Coverage
Model-Template Alignment
Target
8fxv.1.A


1
392
Model 04
Structure Assessment
Oligo-State
Homo-dimer
(matching prediction)
GMQE
0.39
QMEANDisCo Global:
0.41 ± 0.05
QMEANDisCo Local
QMEAN Z-Scores
Template
5ntu.1.A Growth/differentiation factor 8
Crystal Structure of human Pro-myostatin Precursor at 2.6 A Resolution
Seq Identity
20.77%
Coverage
Model-Template Alignment
c


Target
5ntu.1.A
75
55
146
127
218
199
293
243
368
311
392
335

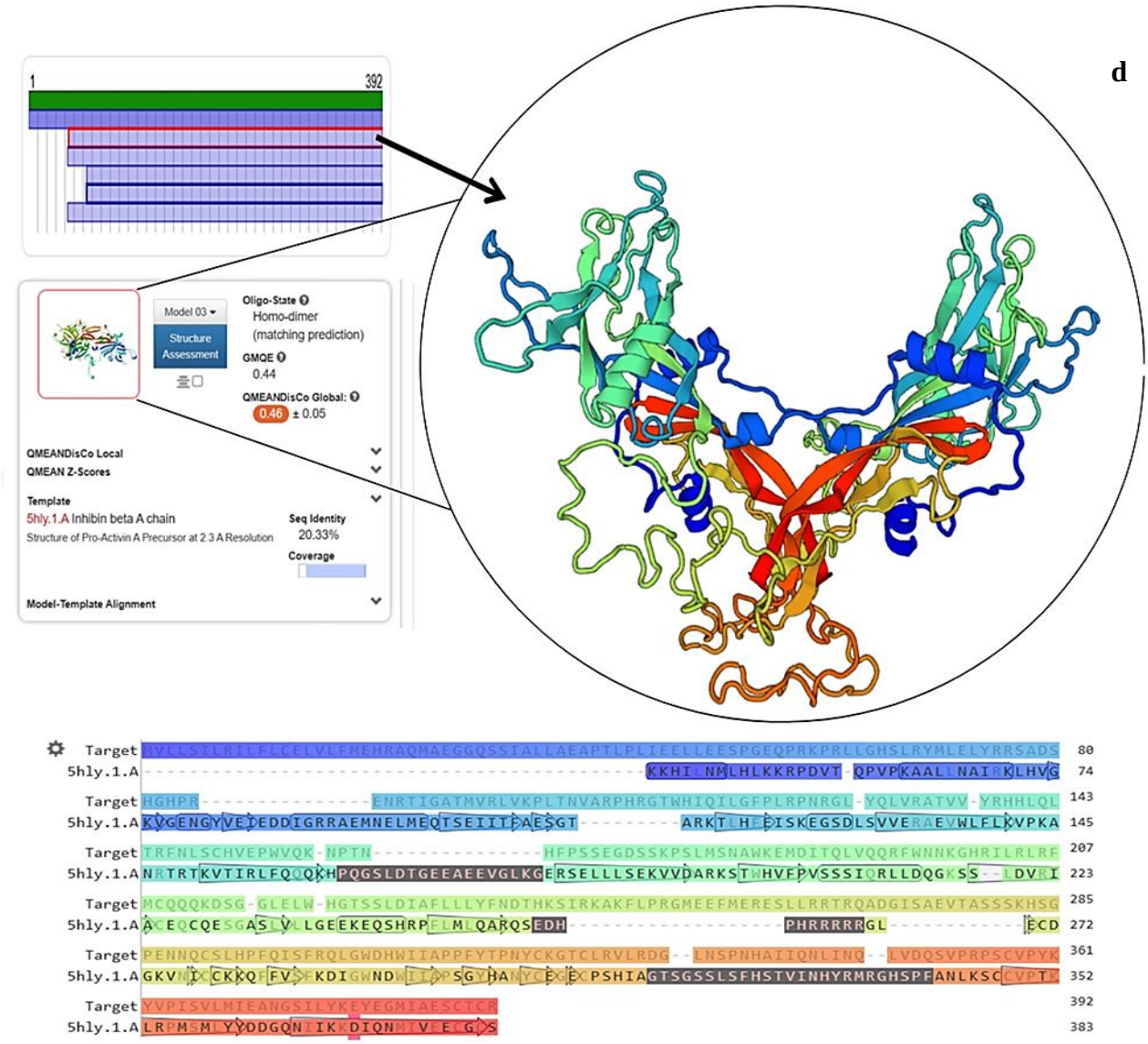


Figure 7: The quaternary structures of four templates of Protein Homology/analogy, (**a**) model one (Q6PX77.1.A) with sequence identity 75.38%, (**b**) model two (5vqf.2.A) with sequence identity 20.83%, (**c**) model three (5ntu.1.A) with sequence identity 20.77% and (**d**) model four (5hly.1.A) with sequence identity 20.33%, which recognition by Swiss model tool.

a. Using a template search to find structural homology, The SWISS-MODEL template library searched for the templates using Blast and BMPlits. BLAST [22] was used to scan the target sequence against the primary amino acid sequence. It was discovered that 345 templates matched the target sequence. This list was filtered by a heuristic down to 50; the top templates are listed in Table 2.

Table 2: The list of four top templates against the primary amino acid sequence of BMP15

No.	Template	Sequence Identity	BiounitOligo State	Description
1.	Q6PX77.1. A	75.38	Monomer	Bone morphogenetic protein 15
2.	5vqf.2. A	20.83%	homo-dimer	Transforming growth factor beta-1 Crystal Structure of pro-TGF-beta 1
3.	5ntu.1. A	20.77%	homo-dimer	Growth/differentiation factor 8 Crystal Structure of Human Pro-myostatin
4.	5hly.1. A	20.33%	homo-dimer	Inhibin beta A chain Structure of Pro-Activin a Precursor

The template (5hly.1.A) Inhibin beta A chain Structure of Pro-Activin A Precursor matches the target sequence (20.33%). This percentage was less than the other three templates in the top when the (5ntu.1.A) Growth/differentiation factor 8 Crystal Structure of human Pro-myostatin matches the target sequence (20.77%) this percentage was less than the other two templates in the top, while the second template in the top was (5vqf.2.A) Transforming Growth Factor beta-1 Crystal Structure of pro-TGF-beta one that matches 20.83% of the target sequence, all these templates have homo-dimer structure. The best template (Q6PX77.1. A) Bone morphogenetic protein 15 that has a Monomer structure matches (75.38%) of the target sequence and is used to predict the protein sequence structure. This process, known as protein structure prediction, involves inferring the three-dimensional structure of a protein depending on its amino acid sequence. The protein structure prediction includes secondary, tertiary, and quaternary structures based on the primary structure, which is essential to determine their functions. There are several methods for protein structure prediction. The template-based technique is called Homology modelling for 3-dimensional structure creation. This model can be applied to the primary sequence when matching the sequence of a protein with a known structure. These techniques provide precise extraction of complex characteristics from protein sequence data. Predicted protein structures with high accuracy are used in drug development, antibody design, understanding protein-protein interactions, and molecular interactions. Therefore, studying and predicting the quaternary structure of a protein depending on its amino acid sequence is valuable in both computational biology and chemistry and holds great promise for future studies.

b. Template Selection: Based on the target template's alignment features, each identified template's quality has been forecasted. Model 2 was chosen because it was of a higher caliber than the other models, and the templates with the best quality were then selected for model building.

c. Model Building: Models are built based on the target-template alignment. Coordinates conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field in the case of loop modelling with Promod-II [23]. If it does not give satisfactory results, an alternative model is built with MODELLER [24].

d. Estimating Model Quality: The QMEAN scoring function has been used to evaluate the global and per-residue model quality [25]. For improved performance, weights of the individual QMEAN terms have been trained specifically for SWISS-MODEL, as shown in Fig. 8.

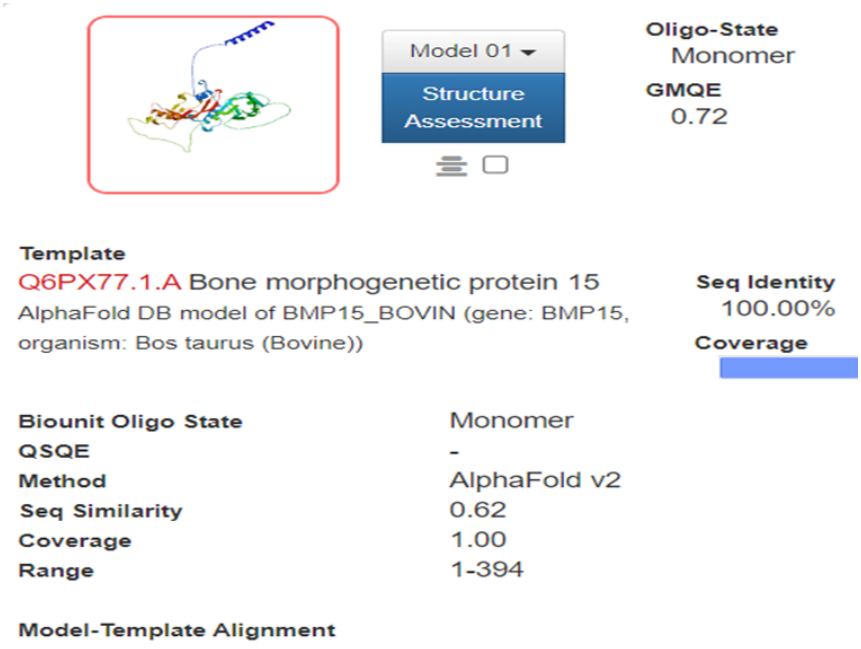


Figure 8: The Bone morphogenetic protein 15 model with a monomer structure built using the SWISS model tool.

3.7. Models Ligands

The ligands from the template structure are homologously transferred to the model when the key requirements are provided [26]. In the template database, the ligands have annotations indicating their biological relevance, and the model is in contact with the ligand without any conflicts between the ligand and the protein or between the target and the template. In addition, the interaction of the residues with the ligand should be conserved (Fig. 9). The unavailability of these four requirements will prevent a particular ligand from being included in the model. The model summary provides information about the deleted ligands and the reasons for their deletion. Using computational methods to predict the proteins -' interaction and detect the ligands follow four main techniques: analysis of the protein's sequence, the study of the structure of the protein, prediction of the ligands of protein, and prediction of the protein's function. Identifying the proteins that bind to physiologically active ligands is an important step in drug development. Computationally, methods are considered practical for experimental research due to time-saving and lower costs for protein-protein interaction prediction. The description of small-molecule ligands is useful for creating research methods for further drug development studies and the analysis of novel amino acid sequences. Many techniques are used for protein description, which lists the characteristics of the whole amino acid sequence [27]. The typical characteristics are the Smiths-Waterman pairwise alignment normalized scores. Physio-chemical properties correlate with residue pairings, residue distribution, and other factors, and they are used in several methods to construct sequence descriptors of whole amino acid sequences.

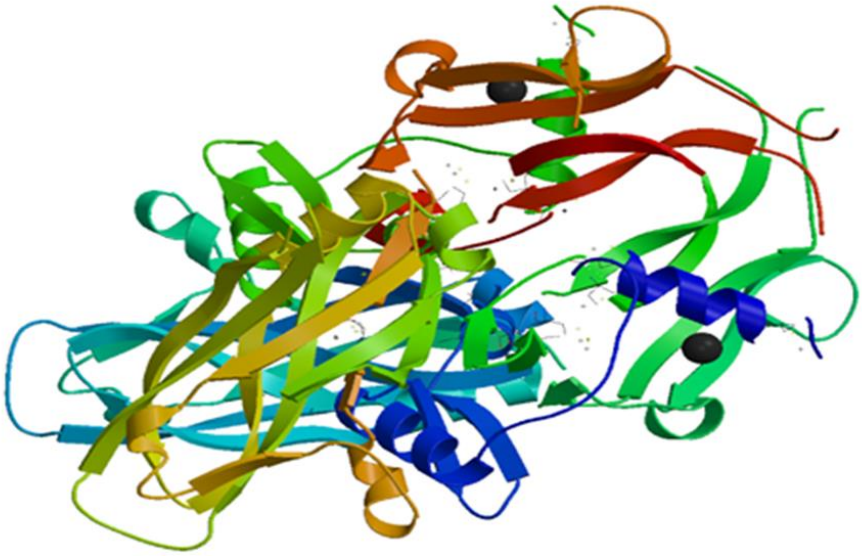


Figure 9: The three-dimensional structure views the structure of the BMP15 protein with ligand binding by the SWISS model tool.

3.8. STRING Analysis

STRING is the software tool and database for known and anticipated protein-protein interactions. It consists of both direct (physical) and indirect (functional) correlations that have been collected from different sources, including high-throughput tests, the literature, genetic context, and (conserved) co-expression [28]. The target sequence is interacting with 20 different proteins. Ten proteins that interact with the function of the target sequence were selected from the list using a heuristic (Fig. 10).

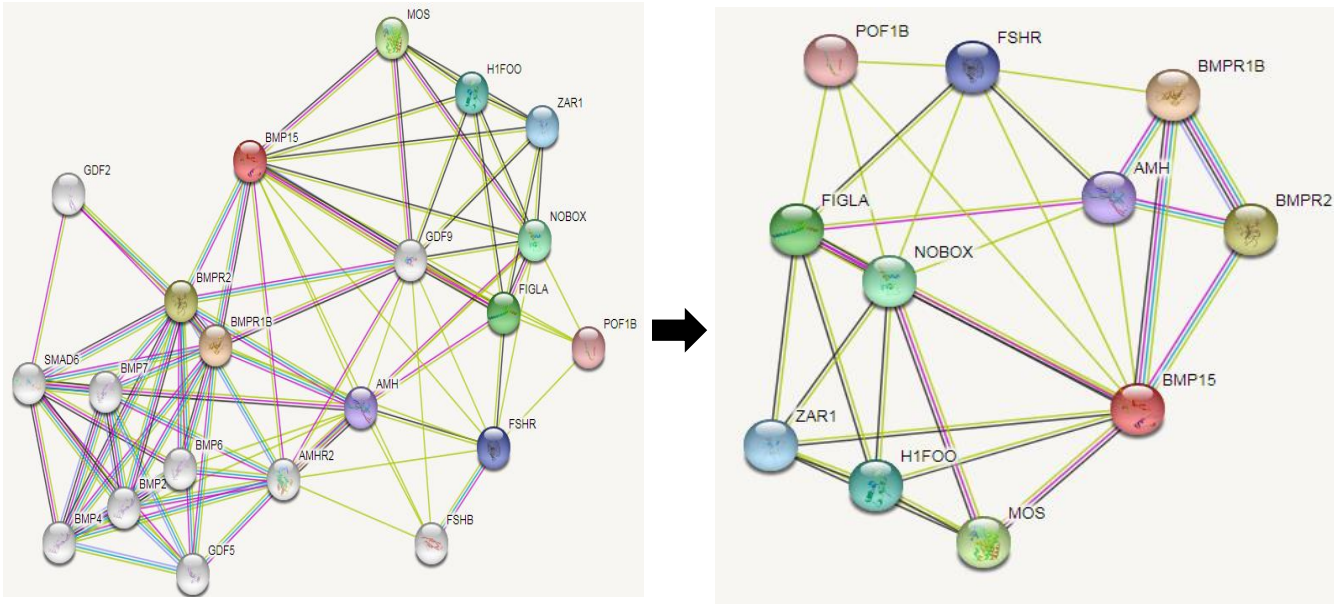


Figure 10: Proteins circles (nodes), and the interaction with the target sequence line colour indicates the type of interaction evidence.

The top protein (BMPR1B) interaction with the target sequence is shown in Fig. 11 due to their interaction in Co-expression, experimentally determined, curated databases, and text mining scored 0.989.

Edges:

Edges represent protein-protein associations	Known Interactions	Predicted Interactions	Others
associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other.	*from curated databases* *experimentally determined*	*gene neighborhood* *gene fusions* *gene co-occurrence*	*textmining* *co-expression* *protein homology*

Your Input:

		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
BMP15	*Bone morphogenetic protein 15; May be involved in follicular development. Oocyte- specific growth/differentiation factor that stimulates folliculogenesis and granulosa cell (GC) growth; Bone morphogenetic proteins (392 aa)*									

Predicted Functional Partners:

BMPR1B	*Bone morphogenetic protein receptor type-1B; On ligand binding, forms a receptor complex consisting of two type II and two t...*					•	•	•	•	0.989
BMPR2	*Bone morphogenetic protein receptor type-2; On ligand binding, forms a receptor complex consisting of two type II and two ty...*					•	•	•	•	0.963
MOS	*Belongs to the protein kinase superfamily. Ser/Thr protein kinase family*				•	•			•	0.847
FIGLA	*Factor in the germline alpha; Germline specific transcription factor implicated in postnatal oocyte-specific gene expression. Pl...*				•	•			•	0.836
NOBOX	*Homeobox protein NOBOX; Transcription factor which may play a role in oogenesis. Binds preferentially to the DNA sequence...*				•				•	0.782
H1FOO	*Histone H1oo; May play a key role in the control of gene expression during oogenesis and early embryogenesis, presumably th...*				•				•	0.773
ZAR1	*Zygote arrest protein 1; Essential for female fertility. May play a role in the oocyte-to-embryo transition (By similarity); Zinc fing...*				•				•	0.772
FSHR	*Follicle-stimulating hormone receptor; Receptor for follicle-stimulating hormone or follitropin. The activity of this receptor is m...*								•	0.769
AMH	*Muellerian-inhibiting factor; This glycoprotein, produced by the Sertoli cells of the testis, causes regression of the Muellerian d...*								•	0.767
POF1B	*Protein POF1B; Plays a key role in the organization of epithelial monolayers by regulating the actin cytoskeleton. May be involv...*								•	0.742

Figure 11: The protein (BMPR1B) interacts is highly with the target sequence.

According to the protein's amino acid sequence, the bone morphological protein 15 (BMP15) may impact follicular growth. Oocyte-specific growth/differentiation factor that promotes the proliferation of granulosa cells (GCs) and folliculo genesis (Fig. 12).

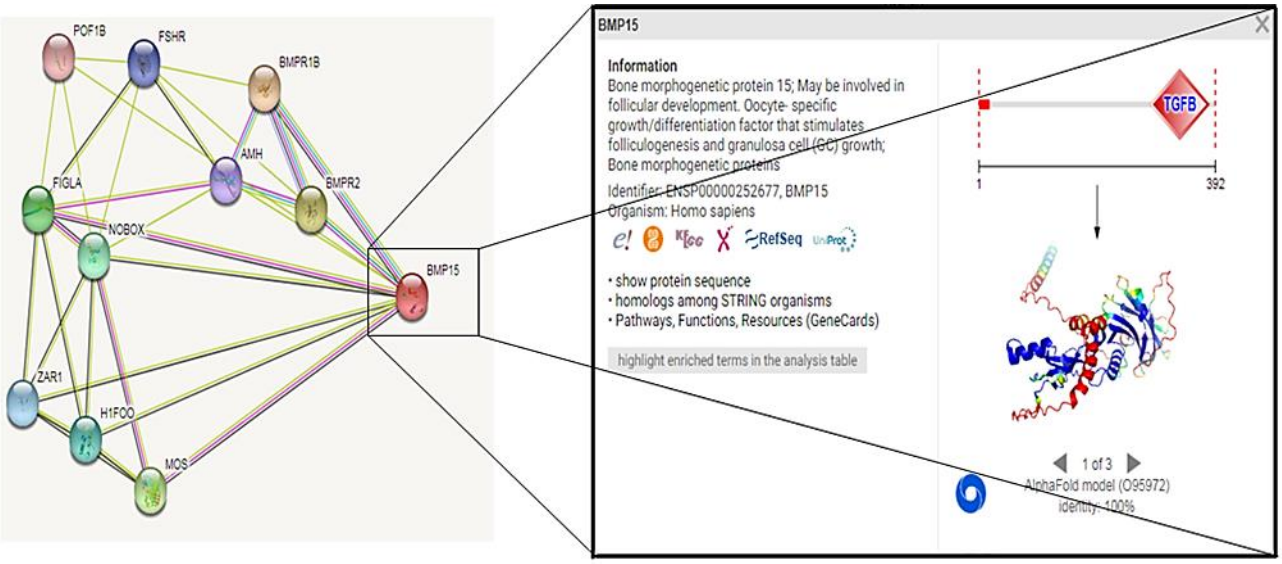


Figure 12: Prediction of the function of target sequence (BMP15 protein) by STRING online software.

Endocrine and paracrine signals regulate ovarian physiology and fertility. They regulate the ovulation quota in many mammals according to their species because, within each species, the GDF9 and BMP15 mRNA expression levels are approximately correlated and impact the ovulation, especially in females of that species [29].

4. Conclusions

The evidence collected from bioinformatics tools was used to identify the structural and functional features of the BMP15 protein. The findings suggested that the BMP15 plays an important role in determining the state of

differentiation and function of granulosa cells and the development of follicles. Additional research on BMP15 and signal interactions is needed to provide a better understanding of follicular development and infertility in women. The functional analysis in this study could be a promising model for further studies in the genetically inherited infertility field due to the correlation between protein, oocyte maturation, and follicular development through the activation of granulosa cells.

Acknowledgement
The authors gratefully acknowledge the technical support provided by the Department of Applied Sciences at the University of Technology in Baghdad, Iraq. We also appreciate the infrastructure assistance given by this department during this study. In addition, we thank the Department of Chemistry at the College of Science of Alexandria University in Alexandria, Egypt for their contributions. This work would not have been possible without the support of these institutions.

Conflict of Interest
The authors declare that they have no conflict of interest.

References
[1] Y. Chen, X. Shan, H. Jiang, L. Sun, and Z. Guo, "Regulation of litter size in sheep (Ovis aries) by the GDF9 and BMP15 genes," *Ann. Agric. Sci.*, vol. 68, no. 2, pp. 148–158, 2023.
[2] K. Sugiura, N. Maruyama, Y. Akimoto, K. Matsushita, and T. Endo, "Paracrine regulation of granulosa cell development in the antral follicles in mammals," *Reprod. Med. Biol.*, vol. 22, no. 1, p. e12538, 2023.
[3] M. D. Kamel and N. J. Kandala, "The Potential Regulatory Role of miR-378 in the Expression of Bone Morphogenetic Protein-15 in Infertile Women with Hyperprolactinemia (HPL)," *Iraqi J. Sci.*, 2023.
[4] S. Kalapahar, S. Sharma, R. Chattopadhyay, and M. Karan, "Overview of Genetics in Infertility," *Bull. Inst. Reprod. Med.*, vol. 39.
[5] M. D. Kamel and N. J. Kandala, "Relationship between BMP15 Gene Expression and FSH in a Sample of Iraqi Infertile Women with Polycystic Ovary Syndrome," *Iraqi J. Biotechnol.*, vol. 21, no. 2, 2022.
[6] V. Nazari, S. Montagud Alario, L. Spilani, V. Dincă, A. Naderi, W. ten Hagen *et al.*, "Co-Evolution of Iolana Blues with Their Host Plants and the Higher Phylogeny of Subtribe Scolitantidina (Lepidoptera, Lycaenidae)," *Diversity*, vol. 16, no. 2, p. 89, 2024.
[7] M. S. Jabir, N. A. Abood, M. H. Jawad, K. Öztürk, H. Kadhim, S. Albukhaty, *et al.*, "Gold nanoparticles loaded TNF-α and CALNN peptide as a drug delivery system and promising therapeutic agent for breast cancer cells," *Mater. Technol.*, vol. 37, no. 14, pp. 3152–3166, 2022.
[8] C. R. Wei, Z. Basharat, and G. C. Lang'at, "In Silico Therapeutic Study: The Next Frontier in the Fight against SARS-CoV-2 and Its Variants," *Drugs Drug Candidates*, vol. 3, no. 1, pp. 54–69, 2024.
[9] K. J. Hoff and M. Stanke, "WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes," *Nucleic Acids Res.*, vol. 41, no. W1, pp. W123–W128, 2013.
[10] R. Azad and S. Mary, "Tentative function and structure prediction of putative genes in the whole genome of gut bacteria ruminococcus champanellensis DSM18848 for the better elucidation of cellular metabolism," in *AIP Conference Proceedings*, AIP Publishing, 2023.
[11] M. D. Kamel, A. A. Mohammed, and A. A. Ibrahim, "Sequence and Structure Analysis of CRP of Lung and Breast Cancer Using Bioinformatics Tools and Techniques," *Biosci. Biotechnol. Res. Asia*, vol. 11, no. 1, pp. 163–174, 2018.
[12] K. V Ashokan, D. S. Mundaganur, and Y. D. Mundaganur, "Catalase: Phylogenetic characterization to explore protein cluster," *J Res Bioinfo*, vol. 1, pp. 1–8, 2010.
[13] J. Singh, K. S. Sandhu, and B. Kaur, "New Insights from the Bioinformatics Tools for Rice Fungal Disease Resistance," in *Fungal Diseases of Rice and Their Management*, Apple Academic Press, 2024, pp. 377–394.
[14] S. Tabassum, F. Munir, A. A. Al Awadh, and Z. Anwar, "Pathological Significance of Cdh1/E-Cadherin Germline Sequence Variants in Breast Cancer Patients," *Exp. Oncol.*, vol. 45, no. 2, pp. 170–179, 2023.
[15] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, *et al.*, "The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible," *Nucleic Acids Res.*, p. gkw937, 2016.

[16] Q. Zheng, L. Guo, R. Yang, Z. Chen, and X. Liu, "Identification of essential genes and drug discovery in bladder cancer and inflammatory bowel disease via text mining and bioinformatics analysis," *Curr. Comput. Aided. Drug Des.*, vol. 20, no. 4, pp. 359–366, 2024.

[17] F. Afkhami, S. Shahbazi, L. Farzadi, and S. Danaei, "Novel bone morphogenetic protein 15 (BMP15) gene variants implicated in premature ovarian insufficiency," *Reprod. Biol. Endocrinol.*, vol. 20, no. 1, p. 42, 2022.

[18] T. K. Sampath and S. Vukicevic, "Biology of bone morphogenetic protein in bone repair and regeneration: A role for autologous blood coagulum as carrier," *Bone*, vol. 141, p. 115602, 2020.

[19] Z. Riaz, M. Hussain, S. Parveen, M. Sultana, S. Saeed, U. Ishaque *et al.*, "In Silico Analysis: Genome-Wide Identification, Characterization and Evolutionary Adaptations of Bone Morphogenetic Protein (BMP) Gene Family in Homo sapiens," *Mol. Biotechnol.*, pp. 1–21, 2023.

[20] J. S. Lolkema and D.-J. Slotboom, "Estimation of structural similarity of membrane proteins by hydropathy profile alignment," *Mol. Membr. Biol.*, vol. 15, no. 1, pp. 33–42, 1998.

[21] E. Wallin and G. Von Heijne, "Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms," *Protein Sci.*, vol. 7, no. 4, pp. 1029–1038, 1998.

[22] A. Sf, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, vol. 25, pp. 3389–3402, 1997.

[23] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling," *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, 1997.

[24] A. Šali and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints," *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, 1993.

[25] P. Benkert, M. Biasini, and T. Schwede, "Toward the estimation of the absolute quality of individual protein structure models," *Bioinformatics*, vol. 27, no. 3, pp. 343–350, 2011.

[26] T. Schmidt, A. Bergner, and T. Schwede, "Modelling three-dimensional protein structures for applications in drug design," *Drug Discov. Today*, vol. 19, no. 7, pp. 890–897, 2014.

[27] D. Karasev, B. Sobolev, A. Lagunin, D. Filimonov, and V. Poroikov, "Prediction of protein–ligand interaction based on sequence similarity and ligand structural features," *Int. J. Mol. Sci.*, vol. 21, no. 21, p. 8152, 2020.

[28] H. Tang, J. Li, J. K. Li, S. H. He, G. Xiang, R. Rong, *et al.*, "BMP6 participates in the pathogenesis of adolescent idiopathic scoliosis by regulating osteopenia," *J. Cell. Physiol.*, vol. 238, no. 11, pp. 2586–2599, 2023.

[29] S. Albukhaty, S. Al-Musawi, S. Abdul Mahdi, G. M. Sulaiman, M. S. Alwahibi, Y. H. Dewir, *et al.*, "Investigation of dextran-coated superparamagnetic nanoparticles for targeted vinblastine controlled release, delivery, apoptosis induction, and gene expression in pancreatic cancer cells," *Molecules*, vol. 25, no. 20, p. 4721, 2020.