Trends in Bioinformatics

ISSN 1994-7941
Short Communication

Conservation Pattern, Homology Modeling and Molecular Phylogenetic Study of BMP Ligands

1Sudarshan Chatterjee, 1Madhuri Haque, 2Md. Shaifur Rahman, 2Hossen M. Jamil, 2Naznin Akhtar, 3S.M. Abdul-Awal, 1Md. Shahedur Rahman and 2Sikder M. Asaduzzaman

1Department of Biotechnology and Genetic Engineering, Jahangirnagar University, 1342 Dhaka, Bangladesh
2Tissue Banking and Biomaterial Research Unit, Atomic Energy Research Establishment, 1349 Dhaka, Bangladesh
3Biotechnology and Genetic Engineering Discipline, Khulna University, 9208 Khulna, Bangladesh

Abstract

Background: Bone Morphogenetic Proteins (BMPs) play many important roles in embryogenesis and metamorphosis throughout the life of vertebrates and other animals. The BMPs are multifunctional and some BMPs do similar functions which lead us to believe that it may have existed before the evolution and they may evolve from a single BMP. Additionally, less is known regarding phylogeny and conservation based conserved molecular mechanism of BMPs. Hence, evolutionary relationship of 16 BMP ligands (phylogenetic as well as protein sequence conserved patterns) were done. Materials and Methods: For this study, protein sequences were retrieved from UniProtKB, homology modeling was executed by Swiss-model using 3jr.1A and 2qcn.1A as template, followed by MSA and phylogeny. Later conserved regions of BMP ligand were compared. Protein subfamily determination was done in Zebra and supported by the phylogenetic data. Results: Remarkably, similar region of conserved area were observed and different disulfide linkage pattern had been identified. It was found notable patterns in C1-C4, C2-C5 and C3-C6 that all but BMP3 and BMP15 do not contain 7th Cys. Phylogeny study indicate, according to evolutionary clock, GDF11 and BMP15 were more distantly diverged taxa, BMP3A and BMP3B had same point origin and GDFS, 6 and 7 are homologous. These seven proteins, as per this study indicates that, those are evolved at simultaneously during evolution, whereas other nine forms monophyletic taxa. Conclusion: However, more studies needs to explore on this. As per homology modelling studied, BMP ligands shares common evolutionary origin, some of the members are highly diverged indicating ancient evolutionary history of these protein. This study will be useful for wet lab once the evolutionary relationship gets established.

Key words: Bone morphogenetic protein, bioinformatic analysis, conservation pattern, homology modeling, molecular phylogenetic

Received: April 25, 2016  Accepted: May 17, 2016  Published: September 15, 2016


Corresponding Author: Md. Shaifur Rahman, Tissue Banking and Biomaterial Research Unit, Atomic Energy Research Establishment, 1349 Dhaka, Bangladesh

Copyright: © 2016 Sudarshan Chatterjee et al. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.
INTRODUCTION

The Bone Morphogenetic Protein (BMP), Growth and Differentiation Factor (GDF) belongs to large TGF-β superfamily. Based on sequence similarities new TGF-β family members is added to increase the number and revealed the newer functions. It is beyond the scope of the article to present the biology of TG factors in any detail. However, BMP family comprises the largest members and most of them are multifunctional and few individual BMP shares many functions. For instances, BMP2, 4, 6, 7 and 9 are capable of inducing bone formation in ectopic sites while BMP4 is also a critical determinant of the non-osteogenic embryological patterning of mammals. Moreover, some of BMP family members shares common pathways and followed conserved molecular mechanism. These multiple functions and sharing feature of BMPs family members are suggesting by the fact that all BMP-like molecules may have an evolutionary relation and same origin point. For example, chromosomal assignments of TGF family members indicate that these genes have become widely dispersed during their evolution. Many investigators have isolated, sequenced and showed crystallographic structure of BMPs. It is widely accepted that the degree and extent of sequence homology can suggest a phylogenetic relationships within the any protein superfamily, for instance BMPs family. For distantly related genes the definition of homologous (orthologous) or duplicated (paralogous) genes is often difficult, a phylogenetic study can often solve this problem. Evolutionary information from sequence data may therefore allow a prediction of the possible biochemical properties of these otherwise uncharacterized protein.

Evolutionary histories of proteins study have been intensified by the growing genomic data and large uncharacterized protein sequence records. More so, proteins that are a part of a large family are being conveniently identified, characterized and classified with the help of their comparative or homologous features and structural figure since all protein families exhibit some level of similarity and relationship. Functional characterization of an unknown protein can be analysis by its predicted 3D structure through computational methods and later proved by wet lab experimental search. But, the prediction of protein model have some short comings such as the determination techniques are time consuming and complicated nature. Fortunately, different computational methods have been employed to bridge the gap between the number of known sequences and 3D models. Template-based protein structure modeling techniques is one of very interesting methods that rely on principles of natural proteins from the theory of evolution viewpoint. In this study, 16 major BMP ligands (Table S1) were analyzed to understand their protein sequence conversed patterns and evaluate phylogenetic relationship among these protein family members. Overall phylogeny unfolds the evolutionary relationship; however more studies needs to explore on this. Model evaluation were also performed to check the reliability of the models. This study will be useful for wet lab once the evolutionary relationship gets established.

Table S1: Human BMP ligands with their alternative name, function their Swissprot accession number, gene name, amino acid residue and corresponding RCSB PDB ID were analyzed during this study

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Alternative name</th>
<th>Functions</th>
<th>Accession No.</th>
<th>Gene name</th>
<th>Amino acid residues</th>
<th>PDB Id</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP4</td>
<td>BMP2B</td>
<td>Induces cartilage and bone formation</td>
<td>P12644</td>
<td>BMP4</td>
<td>408</td>
<td>N/A*</td>
</tr>
<tr>
<td>BMP2</td>
<td>BMP2A</td>
<td>Induces cartilage and bone formation</td>
<td>P12643</td>
<td>BMP2</td>
<td>396</td>
<td>3bmp</td>
</tr>
<tr>
<td>BMP10</td>
<td>N/A</td>
<td>Cardiac development</td>
<td>O95393</td>
<td>BMP10</td>
<td>424</td>
<td>N/A*</td>
</tr>
<tr>
<td>BMP15</td>
<td>GDF9b</td>
<td>Follicle-stimulating hormone activity</td>
<td>O95972</td>
<td>BMP15</td>
<td>392</td>
<td>N/A*</td>
</tr>
<tr>
<td>GDF2</td>
<td>BMP9</td>
<td>Potent circulating inhibitor of angiogenesis</td>
<td>Q9UK05</td>
<td>GDF2</td>
<td>429</td>
<td>4mpl</td>
</tr>
<tr>
<td>BMP5</td>
<td></td>
<td>Chondrogenesis</td>
<td>P22003</td>
<td>BMP5</td>
<td>454</td>
<td>N/A*</td>
</tr>
<tr>
<td>BMP6</td>
<td>Vrg1</td>
<td>Osteoblast differentiation, chondrogenesis</td>
<td>P22004</td>
<td>BMP6</td>
<td>513</td>
<td>2s52</td>
</tr>
<tr>
<td>BMP7</td>
<td>OP1</td>
<td>Osteoinductive, development of kidney and eye</td>
<td>P18075</td>
<td>BMP7</td>
<td>431</td>
<td>1bmp</td>
</tr>
<tr>
<td>BMP8A</td>
<td>OP2</td>
<td>Bone morphogenesis, maintenance of spermatogenesis</td>
<td>Q72526</td>
<td>BMP8A</td>
<td>402</td>
<td>N/A*</td>
</tr>
<tr>
<td>BMP8B</td>
<td>OP3</td>
<td>Bone induction, calcium regulation and bone homeostasis</td>
<td>P34820</td>
<td>BMP8B</td>
<td>402</td>
<td>N/A*</td>
</tr>
<tr>
<td>BMP3A</td>
<td>Osteogenin</td>
<td>Negative regulator of bone morphogenesis</td>
<td>P12645</td>
<td>BMP3A</td>
<td>472</td>
<td>2eqq</td>
</tr>
<tr>
<td>BMP3B</td>
<td>GDF10</td>
<td>Endochondral bone formation and bone morphogenesis inhibitor</td>
<td>P55107</td>
<td>GDF10</td>
<td>478</td>
<td>N/A*</td>
</tr>
<tr>
<td>GDF5</td>
<td>BMP12</td>
<td>Enhances tendon healing and bone formation</td>
<td>P43026</td>
<td>GDF5</td>
<td>501</td>
<td>1wag</td>
</tr>
<tr>
<td>GDF6</td>
<td>BMP13</td>
<td>Chondrogenesis and hypertrophy</td>
<td>Q6KF10</td>
<td>GDF6</td>
<td>455</td>
<td>N/A*</td>
</tr>
<tr>
<td>GDF7</td>
<td>BMP14</td>
<td>Tendon and ligament formation, repair and development of sensory neurons</td>
<td>Q724PS</td>
<td>GDF7</td>
<td>450</td>
<td>N/A*</td>
</tr>
<tr>
<td>GDF11</td>
<td>BMP11</td>
<td>Development of spinal cord during embryogenesis, eye, pancreas development, kidney formation and skeleton patterning</td>
<td>O95390</td>
<td>GDF11</td>
<td>407</td>
<td>N/A*</td>
</tr>
</tbody>
</table>

*N/A: No experimental structure available
MATERIALS AND METHODS

Sequence and structure retrieval: Sequence data of BMP2, 3, 3B, 4, 5, 6, 7, 8, 8B, 10, 15 and GDF2, 5, 6, 7, 11 were retrieved from UniProtKB (http://www.uniprot.org/) as FASTA format. Crystallographic structures for BMP2, 3, 6, 7 and GDF2, 5 were collected from RCSB-PDB (http://www.rcsb.org/pdb/)46. Rest of the protein tertiary structures those are not available in RCSB-PDB, protein structural modeling by Swiss-model (http://swissmodel.expasy.org/) were executed to derive predictive structure25,26.

Sequence alignment and analysis of conserved position: The retrieved sequences were analyzed using ClustalW-MSA for obtaining pairwise distance27. Calculation was done in MEGA6.06. Web based aligner T-Coffee (http://www.tcoffee.org/Projects/tcoffee/) was used to generate a reliable MSA28 and this data were saved in Phylip format those aim to be incorporated in Phylogenetic Web Repeater (POWER: http://power.nhri.org.tw/power/home.htm)29,30. Multiple Sequence Comparison by Log-Expectation (MUSCLE): http://www.drive5.com/) was used to analyze the conserved and consensus positions of the C-terminal and results were observed in JAVA runtime environment viewer JALVIEW31,32.

Protein subfamily predictions and phylogenetic tree construction: In Zebra web-server (http://biokinet.belozerksy.msu.ru/zebra), MSA files were given in FASTA format with “.fa” extension was used to predict the subfamilies within a BMPs family. This web interface identified amino acid residues responsible for functional discrimination and selected hotspots for directed evolution33. The POWER web interface (http://power.nhri.org.tw/) was applied for phylogenetic tree construction and analysis which based on the Dayhoff-PAM and PROTDIST method34. The KITSCH (http://caps.ncbs.res.in/ iws/ phylip_fileskitsch.html) program (Fitch-Margoliash least squares method) was used for distance analysis of phylogenetic tree35. The reliability of the estimated trees was evaluated by Bootstrap method with 1000 and 500 replications.

Homology modeling of BMP ligands and conservation pattern determination: To infer structural characteristics of BMP ligands, homology modeling was performed and predicted 3D structures using the Swiss-model (http://swissmodel.expasy.org)36. Ten proteins (Table 1) were subjected to model. Figure 1-6 were generated with RASWIN program36,37. Sequence identity above 30% is a relatively good predictor of the expected accuracy of a model38. Model quality is assessed...
with the local composite scoring function QMEAN. The weights of QMEAN have been specifically retrained for Swiss-model, leading to more accurate local quality predictions. In addition, global QMEAN scores are calculated as indicators for the overall model quality. Further, a combined quality estimate is provided, which combines the QMEAN estimate with the GMQE obtained from the target-template alignment. The resulting GMQE is expressed as a number between zero and one, where higher numbers indicate higher reliability. Later, PDB files of BMPs structures from both RCSB-PDB and homology modeled with corresponding chain identifier were uploaded in the Consurf server to perform MSA using Multiple Alignment using Fast Fourier Transform (MAFFT). Consurf server was run in “Conseq mode” which retrieved us conserved amino acid residues with their specific position. The homologues were collected from UNIPROT with search algorithm PSI-BLAST. The E-value of PSI-BLAST E-value was settled at 0.0001 and the numbers of iterations were 3. The chosen sequences by the server were manually set between the ranges of maximal identity of 95% to minimal identity of 35%. Neighbor joining tree were constructed with maximum likelihood distance. Conservation scores were calculated by Bayesian method and model of substitution for proteins. For functional protein interaction network, STRING database was used to identify the known and predicted protein interactions partners of BMPs both direct (physical) and indirect (functional) associations.  

**RESULTS AND DISCUSSION**

**Pairwise distances from Multiple Sequence Alignment (MSA) and conservation pattern of BMP:** For understanding a protein family it is good to know the protein sequences, structures and conservation patterns. The BMPs were classified into subfamily according to structural similarity, previously reported. However, amino acid sequence and position specific information were used for classification. Sequence similarity was found to vary among the studied 16 BMP proteins. The MEGA pairwise distances score between the BMP sequences was estimated from 1.744 (maximum) to 0.011 (minimum) with an average of 1.208. Pairwise distances of all members are given in four different charts in Fig. 1.  

The results indicated that BMPs sequences shared different level of distances between themselves and this may be a hint of their evolutionary disperses. Six cysteine residues were identified as highly conserved at 501, 620, 624, 659, 692 and 694 positions of multiply aligned consensus sequence. The C-terminal region of BMP ligands were found very conserved (Fig. 2).  

**Phylogenetic tree reconstruction and analysis:** The evolutionary history inferred from phylogenetic analysis is usually depicted as branching, tree like diagrams that represent an estimated pedigree of the inherited relationships among molecules. The phylogenetic tree analysis made possible to explore the evolutionary history and divergence of BMP protein. In this study, two different studies are used to generate unrooted trees, one relying on protein distance matrix (Protdist) and another is maximum parsimonious tree relying on Fitch-Margoliash least squares method. In the unrooted tree resulted from protdist method, GDF11 was most diverged and BMP15 was second most diverged taxon (Fig. 3a). The BMP3A and 3B formed a monophyletic group and GDF5, 6 and 7 formed another monophyletic group. Ligands in subfamily one formed a single monophyletic taxon which later divided and formed into four different monophyletic taxon. In the maximum parsimonious tree topology were established with 500 bootstrap replicates (Fig. 3b). All leaf nodes and internal nodes were same for both trees except for BMP15 and GDF11.  

**Protein subfamily determination:** The BMP shares common evolutionary origin but some of the members are highly diverged which indicated an ancient evolutionary history of these proteins. Their clustering pattern in tree demanded categorizing into three subfamily which was done in Zebra and supported by the phylogenetic data. Previously, BMP family members were classified into four categories: BMP-2/4, BMP-5/6/7/8a/8b, BMP-9/10 and BMP-12/13/14. Selected 16 BMP proteins from this study were found to be clustered into three subfamilies. Nine ligands BMP2, 4, 5, 6, 7, 8a, 8b, 10 and GDF2 were categorized into subfamily-1. The rest of the members were classified into subfamily-2 comprising BMP15, GDF5, 6, 7 and subfamily-3 comprising BMP3A, 3B, GDF11.  

**Protein homology model data:** Protein homology modeling executed in this study was that effort to generate the structural information that has not been experimentally obtained yet. The structures of homologous proteins are generally better conserved than their sequences. This phenomenon is demonstrated by the prevalence of Structurally Conserved Regions (SCRs) even in highly divergent protein families. Template based protein modeling were performed for 10 BMP ligands that have no experimental 3D structural data available (Table 1). All the templates that have been used are with more than 20% sequence identity and some are above 30%. All the modeled protein in this study contains a good GMQE (around 50%) which supports the reliability of these modeled proteins.
Fig. 1(a-d): MSA scores of protein sequences of 16 human BMP ligands
Conserved pattern of BMP ligands: However, BMPs are highly conservative in structures. But, residue swapping selectively alters BMP orienting ability and structure lead to distinct function. The structural data comprising the conserved amino acid residues in the BMP is represented in Fig. 4. Cysteine plays very important role in stabilization of protein structure at higher level because disulfide bridges formed by cysteine residues are permanent component of protein primary structure, it can also change secondary structure by steric constraints. Cysteine related disulfide bridges are permanent element for stabilization of the tertiary structure and in most cases; interchain S-S bridges are absolute condition for quaternary structure to exist. In this study, the entire cysteine residue with their position and found a remarkable pattern of C1-C4, C2-C5 and C3-C6. All but BMP3 and BMP15 do not contain the 7th cysteine residue.

N-terminal region is rich in variable amino acid residues which indicate these regions are flexible to mutations.
First 200 amino acids are exposed to outer environment and are rarely functionally or structurally involved. On the contrary, C-terminal region is highly rich in conserved amino acid residues and these conserved amino acids are of functional and structural significance. Cysteine residues were specially found to be conserved and their disulfide linking pattern is C1-C4, C2-C5 and C3-C6. The conserved cysteine residues are given in Table 2 and found to be following three distinct patterns which are shown in Fig. 5.

One pattern consist three disulphide bond and variable number of free cysteine residues as in BMP2 and most other BMPs. Another pattern consist four disulphide bonds and five free cysteine residues as in GDF11. The last pattern is for BMP proteins forming homodimers like BMP3A, BMP3B and BMP6. They contain an A-chain and B-chain which are identical molecules forming dimer and they are bridged by the same amino acids from each chains.

Protein-protein interaction among BMP ligands: From the interaction string image, BMP receptors are seems to be the central functional partner to be interacting with BMP ligands. Effectors like SMADs, receptors like ACVR and other proteins like noggin and chordin are seen to be interacting with BMP (Fig. 6). From this network that scores of all the nodes ranges from 0.998-0.999, therefore it is deducible that all the proteins are strongly interconnected finally, protein interaction image has given their signaling and interaction information briefly.
**5.** (a-c): Disulphide linking pattern, (a) Pattern 1: There are three disulphide linkage as they are assumed to be monomer, (b) Pattern 2: Consists four linkage, it is a monomer too and (c) Pattern 3: Consists of seven disulfide linkage three from each monomer, a common disulphide bridge between two monomer.

**Fig. 6:** Protein-protein network between the proteins of BMP ligands and STRING image shows the proteins and their functional partners in colored circles and the interaction with colored straight lines.
CONCLUSION

The BMPs are a complex family of extracellular proteins. The results reported in this study, along with the accumulating data and models from other experimental study are brought together to present the evolutionary history of BMPs. Phylogenetic analyses of this study suggest that human BMPs are a differentiated group of proteins rendering similar function and belongs to a monophyly taxon although few of them are highly diverged. The C-terminal fold of BMPs and conserved nature of amino acid residues seems to have the potential for many different uses, depending on the protein expression and interactions. Hypotheses of BMPs function can now be placed in a phylogenetic context and it can be anticipated that the phylogenetic framework presented here will provide important insights into the history of changing functions of these apparently simple but versatile proteins.

SIGNIFICANCE STATEMENTS

Bone Morphogenetic Proteins (BMPs) family are being increasingly studied due to their contribution in embryonic development and physiological functions in human body throughout life. Practically, recombinant BMP2 and 7 are able to induce bone formation on scaffold and ectopic site. It has already been used in different clinical applications such as open fracture of long bones, non-unions and vertebral arthrodesis and lumbar fusion. This study reported about the template-based protein 3D structure modeling; molecular phylogenetics, conserved pattern analysis of 16 BMP ligands and interaction partner identification which could provide a broader understanding of this protein family. From functional standpoint, this study will be useful for wet lab to study molecular mechanisms of action and significance in dental, cartilage and bone-related diseases once the structure and evolutionary relationship gets established.

ACKNOWLEDGMENT

The authors would like to thank TBRRU, Atomic Energy Research Establishment, 1349 Dhaka for the support of software during the bioinformatic study.

REFERENCES


