Genome-wide Analysis of BMP Family Proteins in Zebrafish *(Danio Rerio)* Genome

Farhana Mostofa[1], Nur Fatihah Binti Mohd Yusoff[2], Siti Aqlima Ahmad[1], Andsyahida Ahmad[1,*]
[1]Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
[2]Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

ABSTRACT

Bone morphogenetic proteins (BMPs) are a group of signalling molecules which belongs to TGF beta super family and known to be a potent inducer of bone and cartilage formation. They also have an important role in cartilage & skeletal disorders such as osteoarthritis and osteoporosis. Since zebrafish is a well validated model for genetic studies and developmental biology, genome information of zebrafish was used to analyse the protein of interests. First, thirty-eight BMP family members of zebrafish were identified using HMM profiles in HMMER and NCBI databases. Next, protein sequences were analysed using the MEGA-X, MEME, Gene Structure Display Server 2.0 & ProtParam programs to characterise the sequence relationship of zebrafish BMP proteins and to determine their physicochemical properties. From the phylogenetic analysis, five major clades were inferred. Among these clades, a group of TGF, Inhibin and BMP proteins were clustered distinctly. Most of the BMP proteins contain five motifs which were annotated as TGF beta and TGF beta pro peptide domains. These protein lengths were ranged from 347-501 residues with the molecular weight ranging from 40201.62 - 54217.72 Da and isoelectronic point (pI) values between 6.13 to 9.85. These pI values showed 83% of the proteins are basic in nature. These data that can be used for subsequent analyses on the regulation of identified BMP proteins in zebrafish in response to inflammatory states. Additionally, manipulation of their sequences can be useful for a genetic study through a CRISPR technology on cartilage and skeletal disease model in zebrafish.

Keywords: bioinformatic analysis, bmp family proteins, bone and skeletal disease, zebrafish

INTRODUCTION

BMPs are a family of signalling molecules that belong to the TGF beta superfamily and are known to be a crucial inducer of bone and cartilage development.(MacFarlane et al. 2017).To date, over 20 BMPs family members have been recognized. BMP-1, which is procollagen C proteinase containing 730 amino acid residues with rich cysteine residue, is a regulatory factor for bone growth and belongs to the family of metalloproteinases. The remaining BMPs belong to the multifunctional growth factors, transforming growth factor (TGF-β) superfamily. TGF-β superfamily mainly includes four sub-families , *i.e.* TGF-β subfamily, BMPs subfamily, growth and differentiation factors (GDFs) subfamily, and activins/inhibins subfamily(Ducy and Karsenty, 2000).Members of the BMP family affect all aspects of bone, cartilage and joint biology (Salazar et al. 2016). They play an important role incartilage & skeletal disorders such as osteoarthritis and osteoporosis.Members of BMP family for example, BMP2 and BMP4 has role in chondrocyte hypertrophy and cartilage degradation.TGF-βs and BMPs regulate postnatal joint cartilage homeostasis, thus dysregulated TGF-β and BMP signalling are often associated with osteoarthritis in both human disease and mouse models(Wu et al. 2016). Since, zebrafishis a well validated model for genetic studies and developmental

biology, and it also shares 80% genome similarity to human genome(Carnovali et al. 2019). Therefore, in our workgenome information of zebrafish was usedto analyseBMP family proteins in zebrafish to study about bone and skeletal disease using *in silico* approaches.

METHODOLOGY

First, BMP 7 gene from human genome was selected based on literature review and HMM profile of that gene was identified from Pfam (http://pfam.xfam.org/) database.Then,thirty-eight BMP family members of zebrafish were identified using HMM profiles of BMP7 gene, HMMER (http://hmmer.org/)tools and NCBI (https://www.ncbi.nlm.nih.gov/) databases. Next, protein sequences were analysed using the MEGA-X (https://www.megasoftware.net/) and maximum likelihood method to determine phylogenetic relationship among identified BMP family members.The reliability of the estimated trees was evaluated by Bootstrap method with 1000 replications. MEME (https://meme-suite.org/meme/), in-silico tools was used to determine motif and sequence logos of protein of interests. Gene Structure Display Server 2.0 (http://gsds.gao-lab.org/) used to identify gene structure of the thirty-eight BMP proteins & ProtParam (https://web.expasy.org/protparam/) programs used to determine their physicochemical properties such as molecular weights, positively charged residues, negatively charged residues, protein length, Instability index, Aliphatic index, GRAVY.

RESULTS

From the phylogenetic analysis, five major clades were inferred. In clade I GDF proteins are included, in clade II BMP proteins, in clade III nodal proteins, in clade IV TGF proteins and in Clade V inhibin proteins.Among these clades, a group of TGF, Inhibin and BMP proteins were clustered distinctly (Figure **1**). Most of the BMP proteins contain five motifs which were annotated as TGF beta and TGF beta pro peptide domains except Lefty 1, Lefty 2, GdF8, Myostasin contains four motifs, BMP2b contains 3 and GDF6-A like contains only two motifs. Signature of Motifs are as follows Motif I: C-X-G-X-C, Motif II: C, Motif III: C-X-F, Motif Iv: M-X-X-X-C-X-C, Motif V: C-X-C. From sequence logo analysis, it is found that cystine and Glycine amino acids are most prominent in BMP family(data not shown).Among 38 genes, few genes for example GDF2, BMP2, BMP7a TGF beta 2 & Inhibin beta 2 don't contain introns.These protein lengths were ranged from 347-501 AA residues (data not shown). From the physicochemical result analysis, it was found that molecular weight of the 38 proteins ranges from 40201.62 - 54217.72. Based on computed PI values, most of the proteins are basic in nature and contains more positively charged residue (1883) than negatively charged residues (1605)(Table 1).According to instability index (II) data, most of the proteins has value more than 40, meaning they are unstable (Saleem and Rajput, 2020). In line with II value, Aliphatic index of proteins are high (71.07-90.28) indicating these BMP proteins are thermally stable. GRAVY score of proteins represents hydrophilic nature of proteins with good solubility and it is found that all proteins have negative values {-0.225-(-0.786)} for GRAVY so they are hydrophilic in natureand isoelectronic point (pI) values between 6.13 to 9.85. These pI values showed 83% of the proteins are basic in nature (Table **1**).

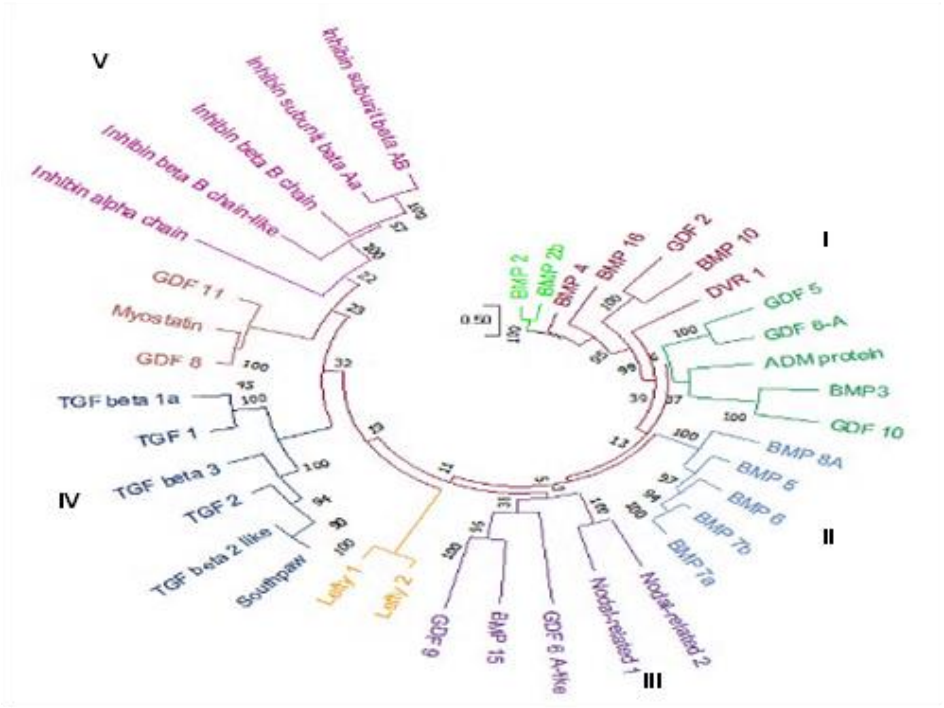

Figure 1. Five major clades were inferred from Phylogenetic tree in which BMP, TGF beta & Inhibin protein clustered distinctly.

Table1. Physicochemical Properties of 38 identified BMP Family Proteins.

Physicochemical properties	Values
Molecular weights	40201.62 - 54217.72
pI	6.13-9.85
Total number of positively charged values Asp+ G/Glu	1883
Total number of negatively charged residue	1605
Protein Length	347-501
Instability Index	41.07-81.74
AliphaticIndex	71.07- 90.28
GRAVY	-0.225-(-0.786)

CONCLUSION

These findings can be used for subsequent analyses on the regulation of identified BMP proteins in zebrafish in response to inflammatory states. Additionally, manipulation of their sequences can be useful for a genetic study through a CRISPR technology on cartilage and skeletal disease model in zebrafish.

ACKNOWLEDGEMENT

This research was financially supported by Ministry of Higher Education Malaysia under Fundamental Research Grant Scheme (FRGS) (Project No: 02-01-04-SF1211).

REFERENCES

1. Carnovali, Marta, et al. “Zebrafish Models of Human Skeletal Disorders: Embryo and Adult Swimming Together.” BioMed Research International, edited by Antoni Camins, vol. 2019, Hindawi, 2019, p. 1253710.
2. Ducy, Patricia, and Gerard Karsenty. “The Family of Bone Morphogenetic Proteins.” Kidney International, vol. 57, no. 6, 2000, pp. 2207–14.
3. MacFarlane, Elena Gallo, et al. “TGF-β Family Signaling in Connective Tissue and Skeletal Diseases.” Cold Spring Harbor Perspectives in Biology, vol. 9, no. 11, Cold Spring Harbor Laboratory Press, Nov. 2017, p. a022269.
4. Salazar, Valerie S., et al. “BMP Signalling in Skeletal Development, Disease and Repair.” Nature Reviews Endocrinology, vol. 12, no. 4, 2016, pp. 203–21.
5. Saleem, Afnan, and Shiveeli Rajput. “Insights from the in Silico Structural, Functional and Phylogenetic Characterization of Canine Lysyl Oxidase Protein.” Journal of Genetic Engineering and Biotechnology, vol. 18, no. 1, 2020, p. 20.
6. Wu, Mengrui, et al. “TGF-β and BMP Signaling in Osteoblast, Skeletal Development, and Bone Formation, Homeostasis and Disease.” Bone Research, vol. 4, no. 1, 2016, p. 16009,