Research Article

Structural Insights of *Megalobrama amblycephala* Major Histocompatibility Complex Class I Alpha and II Alpha Antigen Proteins Revealed by *in Silico* Analysis

Ngoc Tuan Tran* and Gui-Tang Wang
Institute of Hydrobiology, Chinese Academy of Sciences, and the Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Wuhan, Hubei Province, 430072, P. R. China
University of Chinese Academy of Sciences, Beijing 100039, P. R. China

Wei-Min Wang
College of Fisheries, Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education/Key Lab of Freshwater Animal Breeding, Ministry of Agriculture, Huazhong Agricultural University, Wuhan, Hubei 430070, China
Collaborative Innovation Center for Efficient and Health Production of Fisheries in Hunan Province, Changde 41500, China

Abstract

Major histocompatibility complex (MHC) encodes for major histocompatibility antigens which are important in the adaptive immunity of organisms. In this study, the two blunt snout bream (*Megalobrama amblycephala*) (*Ma*-)MHC Iα and MHC IIα antigen proteins were retrieved from the NCBI database and characterised using *in silico* approaches. Physicochemical characterisations of *Ma*-MHC Iα and *Ma*-MHC IIα, including theoretical isoelectric point (pI=5.33 and 4.62, respectively), extinction coefficient (EC=66,600/66,350 and 35,410/35,785 $M^{-1}.cm^{-1}$, assuming all pairs of cysteine residues form cysteines/reduced), instability index (II=37.0 and 35.6), aliphatic index (AI=76.3 and 79.5), grand average hydropathy (GRAVY=-0.475 and -0.197), were analysed. *Ma*-MHC Iα is a membrane protein, whereas *Ma*-MHC IIα is soluble. Disulphide linkages: Cys114-Cys178 and Cys214-Cys272 were found in *Ma*-MHC Iα, and Cys30-Cys179 in *Ma*-MHC IIα. Secondary structure prediction showed that random coils were predominant and followed by alpha helices, extended strands and beta turn. Secondary structures were assigned using 3-D model-based secondary structure approach. The 3-D structures were modelled and validated. The models of proteins were similar to theirs counterparts in human (PDB ID: 2rfxA for *Ma*-MHC Iα) and rat (1fngA for *Ma*-MHC IIα). Further, the ligand-binding sites, enzyme commission (EC) and gene ontology (GO), were predicted based on structure-based functional annotation. This study conducted the analyses of characterisations, structures and functional annotation of *Ma*-MHC Iα and *Ma*-MHC IIα proteins applying *in silico* methods for the first time. The findings provide important

***Correspondence** : tranntts@gmail.com DOI 10.14456/tijsat.2016.18

information for further studies on the extraction, purification, separation and specific functions, e.g. in resistance to infections, of these proteins in blunt snout bream.

Keywords: *Megalobrama amblycephala*; *in silico* analysis; MHC Iα; MHC IIα

1. Introduction

Major histocompatibility complex (MHC) is a large genomic region encoding for major histocompatibility antigens and plays important functions in adaptive immunity of vertebrates [1, 2]. The MHC genes are composed of two main subgroups of immunologically active molecules: class I (MHC I) and class II (MHC II) [3]. Basically, MHC I proteins are heterotrimers, including two alpha and beta-2-microglobulin polypeptide chains which function in displaying intracellular proteins to cytotoxic CD8+ T cells [3-6]. MHC II proteins are heterodimers, consisting of two alpha and beta homogenous peptide chains, encoded by two separate genes [7]. In aquatic animals, both MHC I and II genes have been isolated and characterised in numerous fish species such as zebrafish (*Brachydanio rerio*) [8], Japanese flounder (*Paralichthys olivaceus*) [6], turbot (*Scophthalmus maximus*) [9], half-smooth tongue sole (*Cynoglossus semilaevis*) [10], large yellow croaker (*Pseudosciaena crocea*) [11], miiuy croaker (*Miichthys miiuy*) [12], Nile tilapia (*Oreochromis niloticus*) [13]. Both MHC I and MHC II have also been regarded to be studied as crucial immune-related components in blunt snout bream (*Megalobrama amblycephala*), one of major aquaculture species in China [14, 15]. The basic information of these two genes is available in the NCBI (National Center for Biotechnology Information) database. However, the studies on structure and function of these genes isolated from blunt snout bream using *in silico* analysis have not yet been reported in detail so far. In this current study, the efforts on characterising the physicochemical properties and homology modelling of these two proteins (*Ma*-MHC Iα and *Ma*-MHC IIα) were undertaken. This study aimed to investigate the structural, physiochemical and functional properties of *Ma*-MHC Iα and *Ma*-MHC IIα proteins in blunt snout bream with the usage of *in silico* approaches. Our results are useful for further studies on the extraction, purification, separation and specific functions, e.g. immune-related functions, of the proteins.

2. Materials and Methods

2.1 Protein sequence retrieval, and physicochemical and functional characterisations

The sequences of blunt snout bream (*Ma*-) major histocompatibility complex class I alpha (*Ma*-MHC Iα) (Accession no. AFI42189.1) and class II alpha (*Ma*-MHC IIα) (AGV52142.1) were obtained from the NCBI's database (http://www.ncbi.nlm.nih.gov/) under the FASTA format and used for further analyses. Physicochemical characterisation were analysed using ProtParam tool (http://web.expasy.org/protparam/), each of which characters based on different validation parameters [16]. Theoretical isoelectric point (pI), the pH value at which the net charge of a protein is zero, was calculated using the compute pI/MW tool. Extinction coefficient (EC) indicates how much light proteins absorb in water at 280 nm [17]. Instability index (II) is a measure to evaluate the stability of proteins in a test tube, the II value is <40, indicating the proteins are probably stable and *vice versa* [18]. Aliphatic index (AI) is a positive factor for the increase of thermal stability of proteins, directly involving in the mole fraction of aliphatic side chains (alanine, isoleucine, leucine and valine) [19]. Grand average hydropathy (GRAVY) for proteins are calculated as the sum of hydropathic values of all amino acids, divided by the number of residues in the sequences [20].

SOSUI (http://harrier.nagahama-i-bio.ac.jp/sosui/) was used to determine the types of protein and CYS_REC (http://linux1.softberry.com/) to detect the positions of cysteines and the presence of disulphide bonds and their bonding patterns containing in the proteins.

2.2 Protein structure analysis

The secondary structure of protein was predicted using Self-Optimized Prediction Method by Alignment (SOPMA) server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) according to default parameters (Window width: 17, similarity threshold: 8 and number of states: 4). The secondary structure of the modeled protein was assigned employing the STRIDE program (http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py), which used hydrogen-bond energies and main chain dihedral angles to identify helices, coils and strands [21]. Three-dimensional (3-D) structures were generated employing SWISS-MODEL server (http://swissmodel.expasy.org/). The modelled structures were selected on the basis of sequence identity with the Protein Data Bank (PDB) templates [22].

The stereochemical quality and accuracy of predicted models were checked using Procheck (http://services.mbi.ucla.edu/PROCHECK/), ProQ (http://www.sbc.su.se/~bjornw/ProQ/ProQ.html) and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php). The best models were selected following criteria: the total number of residues (>90%) were in the most favoured regions and additional allowed region and an overall G-factor value (>-0.5) (analysed using Procheck server) [23], LGscore and MaxSub (ProQ) was over 1.5 and 0.1, respectively [24], and the Z-Scores (ProSA) were within the range of the typical scores for native proteins of the similar size and the plots of residue energy values were negative [25, 26].

COFACTOR web server (http://zhanglab.ccmb.med.umich.edu/COFACTOR/) was used to perform the global structure match using TM-align structural alignment. TM-score of the structural alignment between the query structure and known structures in the protein database was calculated to assess the global structural similarity, being scored from 0 to 1, where TM-score=1 indicates the perfect match between two structures. Scores below 0.17 correspond to randomly chosen unrelated proteins, whereas a score higher than 0.5 implies generally the same fold [27].

2.3 Structure-based functional annotation

Annotations on ligand-binding sites, enzyme commission (EC) and gene ontology (GO) were performed using I-TASSER server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/), which structurally matches the 3-D model of *Ma*-MHC Iα and *Ma*-MHC IIα to the known templates in protein function databases.

3. Results and Discussion

3.1 Physicochemical and functional characterisations

The physicochemical characterisation of *Ma*-MHC Iα and *Ma*-MHC IIα was calculated using ProtParam tool. The theoretical isoelectric point (pI) value of *Ma*-MHC Iα and *Ma*-MHC IIα proteins was computed as 5.33 and 4.62, respectively, indicating the acidic character of these proteins. The pI value may be principally used for purifying proteins through performing isoelectric focusing method [28]. The extinction coefficient (EC) measured at 280 nm was 66,600/66,350 (for *Ma*-MHC Iα) and 35,410/35,785 $M^{-1}.cm^{-1}$ (for *Ma*-MHC IIα) when assumed that all pairs of cysteine residues form cysteines or reduced, respectively. Herein, the high EC values imply the high concentration of cysteine, tryptophan and tyrosine presenting in the protein sequences [17]. The instability index (II) value is a measure to assess the stability

of proteins in a test tube [18]. The II value of *Ma*-MHC Iα and *Ma*-MHC IIα was correspondingly 37.0 and 35.6, indicating both proteins are stable (II <40, [18]). The AI of *Ma*-MHC Iα and *Ma*-MHC IIα was 76.3 and 79.5, respectively. The high AI values indicate the stability in a wide temperature range of these proteins. The grand average hydropathy (GRAVY) was computed as -0.475 (*Ma*-MHC Iα) and -0.197 (*Ma*-MHC IIα) [20]. Herein, the negative values of GRAVY infer these proteins are hydrophilic and better solubility in water.

Based on SOSUI analysis, *Ma*-MHC Iα was classified as a membrane protein with two transmembrane regions of ^{3}SVMLLLLGAHLAYAGTHSLKYFY^{25} (secondary type) and ^{297}IGIIVGALAAVVLVILIGVAGYV^{319} (primary type), while *Ma*-MHC IIα was soluble. The results obtained by CYS_REC showed that both proteins contain cysteine and disulphide bonds in pair. The most probable pattern of pairs of cysteine residues located in Cys114-Cys178 and Cys214-Cys272 was predicted in *Ma*-MHC Iα and Cys30-Cys179 in *Ma*- MHC IIα.

3.2 Protein structure analysis

The secondary structure of *Ma*-MHC Iα and *Ma*-MHC IIα were predicted using SOPMA (Figure 1). The results revealed that four main secondary elements, including alpha helices, extended strand, beta turn and random coils, were found in protein sequences. Whereas, the remaining elements, comprising 3_{10} helix, Pi helix, beta bridge, bend region, and ambiguous states, were absent. The *Ma*-MHC Iα consists of a dominating number of random coils (38.3%), followed by alpha helices (25.4%), extended strands (21.9%) and beta turn (14.4%). The *Ma*-MHC IIα comprising random coils (38.0%) was prevailing and was followed by extended strands (30.3%), alpha helices (24.4%), and beta turn (7.3%). Albeit the results showed the predominance of random coils in the protein secondary structure, random coils are basically considered as the patterns of lacking of regular secondary structures in a protein sequence [29]. This indicates the highly regular structures of the protein based on their actual polypeptide backbone chain which are mainly generated by alpha helix and beta sheet [30]. These secondary structures are a link between the linear information and the 3-D structures of proteins.

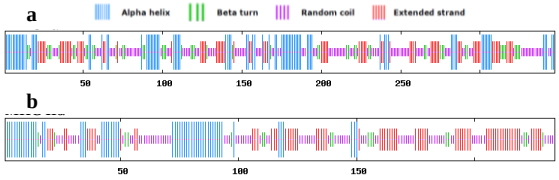

Figure 1. Prediction of secondary structure of (a) *Ma*-MHC Iα and (b) *Ma*-MHC IIα using SOPMA.

The 3-D structures of *Ma*-MHC Iα and *Ma*-MHC IIα was rendered based on the sequence similarity against the PDB's database performing SWISS-MODEL server. The best templates used for modelling *Ma*-MHC Iα and *Ma*-MHC IIα were 2yf5.1.A (36.7% sequence identity) and 1fng.1.A (28.9%), respectively. The 3-D structures of those proteins were shown in Figure 2 a/b-1.

The stereochemical quality of both predicted models was tested performing online servers (Table 1). Based on Procheck's analysis, most of residue number (>90%) were found in favoured (88.5 and 87.7% for *Ma*-MHC Iα and *Ma*-MHC IIα, respectively) and allowed regions (9.9 and 11.6%). The overall G-factor values of two models were 0.01 and -0.15, respectively, being higher than the acceptable values (a cut-off value of -0.5). These results validated the good quality of predicted models [23]. The Lgscore (1.844 and 1.483 for *Ma*-MHC Iα and *Ma*-MHC IIα, respectively), and MaxSub (0.217 and 0.177) employing ProQ revealed the correct models, with the Lgscore was ≥1.5 and MaxSub was >0.1 [24]. The Z-score (*Ma*-MHC Iα: -8.12 and *Ma*-MHC IIα: -5.38) obtained by ProSA indicates the close relationship to typical native structures [25] (Figure 2 a/b-2). The

plot of residue energies of these models was mostly negative values (Figure 2 a/b-3).

Table 1. Assessments of the 3-D structures for *Ma*-MHC Iα and *Ma*-MHC IIα.

Validation index	Models	
	***Ma*-MHC Iα**	***Ma*-MHC IIα**
Procheck analysis:		
Most favoured regions (%)	88.5	87.7
Additional allowed regions (%)	9.1	10.4
Generously allowed regions (%)	0.8	1.2
Disallowed regions (%)	1.6	0.6
Non- proline and non-glycine residues	243	163
Overall average G-factor	0.01	-0.15
ProQ:		
Lgscore	1.844	1.483
MaxSub	0.217	0.177
ProSA		
Z-*score*	-8.12	-5.38

In addition, the secondary structure assignment of the two proteins was carried out using STRIDE (Figure 3). The results showed that the helical positions of both proteins were aligned with those of their templates (data not shown). In *Ma*-MHC Iα, the proportion of helical residues, the conserved domains, obtained for both the sequence-based and 3-D model-based secondary structure approaches was approximately similar (26.0 vs. 25.4%), while in *Ma*-MHC IIα the percentage was little less (12.5 vs. 24.4%). This suggests the good quality to both of the secondary structures.

Altogether, these results confirmed the reliable and good quality of predicted models for both *Ma*-MHC Iα and *Ma*-MHC IIα. To our knowledge, data about tertiary structure of these proteins in blunt snout bream is still absent, this study provides the basic structural insights, facilitating in studying of functional properties of these proteins in the fish.

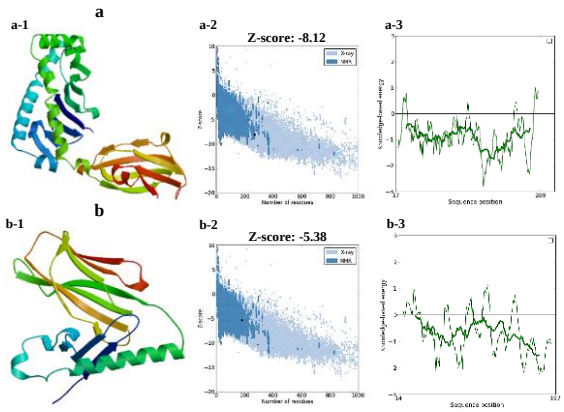

Figure 2. Structural models of (a) *Ma*-MHC Iα and (b): *Ma*-MHC IIα and the *Z*-score and energy plot of predicted models obtained from ProSA analysis results.

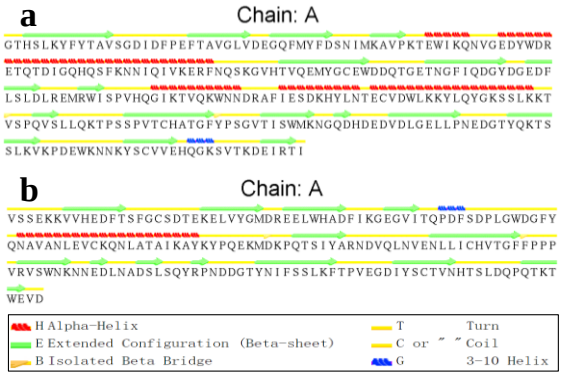

Figure 3. Secondary structure assignment of the models of *Ma*-MHC Iα (a) and *Ma*-MHC IIα (b), generated by STRIDE.

Structure similarity analysis was performed using COFACTOR server (Table 2). The TM-scores >0.5 indicates the two investigated proteins generally have the same fold [27]. Our results revealed that a very high level of structural conservation between *Ma*-MHC Iα and *Ma*-MHC IIα with theirs counterparts in organisms, including human, rat and mouse, where TM-scores are more than 0.8. Among these structural analogues, both *Ma*-MHC Iα and *Ma*-MHC IIα were similar to *Homo sapiens* HLA I histocompatibility (PDB ID: 2rfxA) and *Mus musculus* MHC II I-EK (1fngA), showing the TM-score value of 0.878 and 0.972, respectively, the structurally identical between these models are shown in Figure 4.

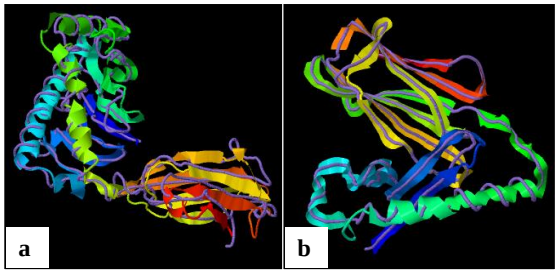

Figure 4. Structural alignment between *Ma*-MHC IIα and *Ma*-MHC Iα (shown in cartoon) and its counterpart structures of *Homo sapiens* HLA class I histocompatibility antigen (PDB ID: 2rfxA) and *Mus musculus* MHC class II I-EK (1fngA) (backbone trace), respectively, generated by COFACTOR server.

3.3 Structure-based functional annotation

To functionally annotate, the modelled structures were queried in the I-TASSER. The ligand-binding sites were predicted, with peptide was a highly significant confidence, on the basis of structure similarity with the rat minor histocompatibility antigen complex RT1-AA/MTF-E (PDB: 1ed3A, C-score=0.81) for *Ma*-MHC I and the human CF34 TCR in complex with HLA-B8/FLR (3ffcA, C-score=0.32) for *Ma*-MHC IIα (Table 3).

Enzyme Commission (EC) analysis indicated that receptor protein-tyrosine kinase (EC number: 3.1.26.4, $Cscore^{EC}$=0.304) and triose-phosphate isomerase (5.3.1.1, $Cscore^{EC}$=0.303) was found hitting with the *Ma*-MHC Iα and *Ma*-MHC IIα, respectively (Table 4).

Consensus prediction of GO terms predicted that both proteins attending to several biological functions molecular function, biological process, and cellular component in organism (Table 5).

Table 2. Top five identified structural analogs in Protein Data Bank (PDB) based on COFACTOR prediction.

Protein	PDB ID	Protein name	Species	TM-Score	RMSD[a]	IDEN[a]	Cov.
Ma-MHC Iα	2rfxA	HLA I histocompatibility	*Homo sapiens*	0.878	2.43	0.351	0.978
	3bzfC	HLA I histocompatibility	*Homo sapiens*	0.861	2.51	0.356	0.971
	2icnA	Zinc-α-2-glycoprotein	*Homo sapiens*	0.855	2.43	0.269	0.967
	1fruA	Neonatal FC receptor	*Rattus norvegicus*	0.853	2.48	0.251	0.963
	4gupA	MHC I-related protein	*Homo sapiens*	0.852	2.26	0.368	0.956
Ma-MHC IIα	1fngA	MHC II I-EK	*Mus musculus*	0.972	0.48	0.289	0.978
	1bx2D	HLA-DR2	*Homo sapiens*	0.947	1.19	0.307	0.973
	3c6lG	TCR 2W20 alpha chain	*Mus musculus*	0.943	1.05	0.261	0.978
	4d8pA	HLA-DQA1	*Homo sapiens*	0.929	1.22	0.261	0.967
	1es0A	H-2 II histocompatibility	*Mus musculus*	0.927	1.13	0.246	0.962

TM-score is an assessment of the structural alignment between the query structure and known structures in the protein database. RMSD[a] is the average root mean square deviation between residues that are structurally aligned by TM-align; IDEN[a] is the percentage sequence identity in the structurally aligned region; Cov. is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein

Table 3. Top two identified ligand binding sites in Protein Data Bank (PDB) based on I-TASSER prediction.

Protein	PDB Hit	C-score	Cluster size	Lig Name	Ligand Binding Site Residues
Ma-MHC Iα	1ed3A	0.81	1548	PEPTIDE	23, 25, 58, 73, 77, 79, 80, 81, 83, 84, 87, 88, 90, 91, 94, 98, 108, 112, 129, 136, 156, 159, 160, 166, 169, 170, 173, 177, 181, 185
	3fqnA	0.02	25	PEPTIDE	21, 23, 25, 58, 77, 80, 81, 84, 112, 170, 173, 177, 181, 185
Ma-MHC IIα	3ffcA	0.32	383	PEPTIDE	2, 3, 5, 6, 9, 10, 13, 24, 28, 42, 44, 51, 65, 68, 69, 71, 74, 75, 78, 82, 86, 90
	3s5lD	0.2	189	PEPTIDE	24, 26, 37, 46, 47, 58, 66, 67, 68, 69, 72, 75, 79, 82, 83, 85, 86, 90, 93

Table 4. Top five identified Enzyme Commission (EC) numbers and active sites in Protein Data Bank (PDB) based on I-TASSER prediction.

Protein	PDB ID	Cscore^{EC}	TM-score	RMSD^{a}	IDEN^{a}	Cov	EC Number	EC name
Ma-MHC Iα	1e27A	0.304	0.72	2.07	0.361	0.775	3.1.26.4	Receptor protein-tyrosine kinase
	1es0B	0.156	0.429	3.84	0.139	0.536	4.1.1.15	Ribonuclease H
	2hnhA	0.149	0.374	6.75	0.056	0.634	2.7.7.7	Glutamate decarboxylase
	1ti6A	0.147	0.37	6.61	0.043	0.625	1.97.1.2	Pyrogallol hydroxytransferase
Ma-MHC IIα	2iamA	0.303	0.727	1.34	0.303	0.761	5.3.1.1	Triose-phosphate isomerase
	3dtfA	0.156	0.43	5.2	0.02	0.671	2.6.1.42	Branched-chain-amino-acid transaminase
	3dtgA	0.156	0.43	5.2	0.025	0.671	2.6.1.42	Branched-chain-amino-acid transaminase
	2zxcA	0.147	0.41	5.15	0.021	0.624	3.5.1.23	Ceramidase
	1e27A	0.147	0.426	3.97	0.156	0.551	3.1.26.4	Ribonuclease H

Cscore^{EC} is the confidence score for the EC number prediction (ranging in between 0-1). TM-score is an assessment of the structural alignment between the query structure and known structures in the protein database. RMSD^{a} is the average root mean square deviation between residues that are structurally aligned by TM-align; IDEN^{a} is the percentage sequence identity in the structurally aligned region; Cov. is the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein

Table 5. The consensus prediction of GO terms of the two models based on I-TASSER prediction.

Protein	Molecular Function	Biological Process	CellularComponent
Ma-MHC Iα	GO:0005515 (0.85)	GO:0001916 (0.53)	GO:0042612 (0.85)
	GO:0042605 (0.32)	GO:0002485 (0.53)	GO:0009897 (0.53)
		GO:0042590 (0.32)	GO:0016021 (0.53)
		GO:0042742 (0.32)	
Ma-MHC IIα	GO:0005515 (0.83)	GO:0019882 (0.83)	GO:0042613 (0.83)
		GO:0006955 (0.83)	

4. Conclusion

In this study, the physicochemical and functional characteristics of the two proteins *Ma*-MHC Iα and *Ma*-MHC IIα were firstly characterised using *in silico* analysis. Physicochemical characters of the proteins are acidic, stable, hydrophilic, and high concentration of cysteine, tryptophan and tyrosine. Disulphide linkages were predicted through the protein sequences. Secondary structure revealed the random coils were predominant and followed by other elements (alpha helices, extended strands and beta turn). Further, 3-D structures of the proteins were predicted and validated using computational tools, which are useful for functional analysis of experimentally derived crystal structures. This current study provides a better understanding about the physiological and functional characterisations, structures and functional annotation of both *Ma*-MHC Iα and *Ma*-MHC IIα of blunt snout bream. The information is important for further studies on the extraction, purification, separation and specific functions, e.g. in resistance to infections, facilitating improvement activity of the proteins.

5. References

[1] Edwards, S.V. and Hedrick, P.W., Evolution and Ecology of MHC Molecules: From Genomics to Sexual Selection, Trends in Ecology & Evolution, Vol.13, No.8, pp.305-311, 1998.

[2] Barribeau, S.M., Villinger, J. and Waldman, B., Major Histocompatibility Complex Based Resistance to a Common Bacterial Pathogen of Amphibians, PLoS One, Vol.3, No.7, pp.e2692, 2008.

[3] Dixon, B., van Erp, S.H., Rodrigues, P.N., Egberts, E. and Stet, R.M., Fish Major Histocompatibility Complex Genes: An Expansion, Developmental & Comparative Immunology, Vol.19, No.2, pp.109-133, 1995.

[4] Bjorkman, P.J., Saper, M., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D., Structure of the Human Class I Histocompatibility Antigen, HLA-A 2, Nature, Vol.329, No.6139, pp.506-512, 1987.

[5] Rothbard, J. and Gefter, M.L., Interactions Between Immunogenic Peptides and MHC Proteins, Annual Review of Immunology, Vol.9, No.1, pp.527-565, 1991.

[6] Srisapoome, P., Ohira, T., Hirono, I. and Aoki, T., Cloning, Characterization and Expression of cDNA Containing Major Histocompatibility Complex Class I, IIα and IIβ Genes of Japanese Flounder *Paralichthys olivaceus*, Fisheries Science, Vol.70, No.2, pp.264-276, 2004.

[7] Klein, J. and Figueroa, F., Evolution of the Major Histocompatibility Complex, Critical Reviews in Immunology, Vol.6, No.4, pp.295-386, 1985.

[8] Takeuchi, H., Figueroa, F., O'hUigin, C. and Klein, J., Cloning and Characterization of Class I MHC Genes of the Zebrafish, *Brachydanio rerio*, Immunogenetics, Vol.42, No.2, pp.77-84, 1995.

[9] Zhang, Y.-X. and Chen, S.-L., Molecular Identification, Polymorphism, and Expression Analysis of Major Histocompatibility Complex Class IIA and B Genes of Turbot (*Scophthalmus maximus*), Marine Biotechnology, Vol.8, No.6, pp.611-623, 2006.

[10] Xu, T.-J., Chen, S.-L., Ji, X.-S. and Sha, Z.-X., Molecular Cloning, Genomic Structure, Polymorphism and Expression Analysis of Major Histocompatibility Complex Class IIA and IIB Genes of Half-Smooth Tongue Sole (*Cynoglossus semilaevis*), Fish & Shellfish Immunology, Vol.27, No.2, pp.192-201, 2009.

[11] Yu, S., Ao, J. and Chen, X., Molecular Characterization and Expression

Analysis of MHC Class II α and β Genes in Large Yellow Croaker (*Pseudosciaena crocea*), Molecular Biology Reports, Vol.37, No.3, pp.1295-1307, 2010.

[12] Xu, T., Sun, Y., Shi, G., Cheng, Y. and Wang, R., Characterization of the Major Histocompatibility Complex Class II Genes in Miiuy Croaker, PloS One, Vol.6, No.8, pp.e23823, 2011.

[13] Pang, J.-C., Gao, F.-Y., Lu, M.-X., Ye, X., Zhu, H.-P. and Ke, X.-L., Major Histocompatibility Complex Class IIA and IIB Genes of Nile Tilapia *Oreochromis niloticus*: Genomic Structure, Molecular Polymorphism and Expression Patterns, Fish & Shellfish Immunology, Vol.34, No.2, pp.486-496, 2013.

[14] Ma, X.-Q., Liu, Z.-Z., Li, S.-F., Tang, W.-Q. and Yang, J.-Q., Full Length cDNA Cloning and Tissue Expression of Major Histocompatibility Complex (MHC) Class I From Blunt Snout Bream (*Megalobrama amblycephala*), Journal of Shanghai Ocean University, Vol.20, No.1, pp.34-43, 2011.

[15] Luo, W., Zhang, J., Wen, J.-F., Liu, H., Wang, W.-M. and Gao, Z.-X., Molecular Cloning and Expression Analysis of Major Histocompatibility Complex Class I, IIA and IIB Genes of Blunt Snout Bream (*Megalobrama amblycephala*), Developmental & Comparative Immunology, Vol.42, No.2, pp.169-173, 2014.

[16] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., Protein Identification and Analysis Tools on the ExPASy Server, Springer, 2005.

[17] Gill, S.C. and Von Hippel, P.H., Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data, Analytical Biochemistry, Vol.182, No.2, pp.319-326, 1989.

[18] Guruprasad, K., Reddy, B.B. and Pandit, M.W., Correlation Between Stability of a Protein and Its Dipeptide Composition: A Novel Approach for Predicting *in vivo* Stability of a Protein from Its Primary Sequence, Protein Eng., Vol.4, No.2, pp.155-161, 1990.

[19] Ikai, A., Thermostability and Aliphatic Index of Globular Proteins, Journal of Biochemistry, Vol.88, No.6, pp.1895-1898, 1980.

[20] Kyte, J. and Doolittle, R.F., A Simple Method for Displaying the Hydropathic Character of a Protein, Journal of Molecular Biology, Vol.157, No.1, pp.105-132, 1982.

[21] Heinig, M. and Frishman, D., STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins, Nucleic Acids Research, Vol.32, No.suppl. 2, pp.W500-W502, 2004.

[22] Fiser, A., Template-Based Protein Structure Modeling, In: Computational Biology (Fenyö, D. editor), Springer, 2010.

[23] Ramachandran, G., Ramakrishnan, C. and Sasisekharan, V., Stereochemistry of Polypeptide Chain Configurations, Journal of Molecular Biology, Vol.7, No.1, pp.95-99, 1963.

[24] Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L. and Elofsson, A., A Study of Quality Measures for Protein Threading Models, BMC Bioinformatics, Vol.2, No.1, pp.5, 2001.

[25] Wiederstein, M. and Sippl, M.J., ProSA-web: Interactive Web Service for the Recognition of Errors in Three-Dimensional Structures of Proteins, Nucleic Acids Research, Vol.35, No.suppl. 2, pp.W407-W410, 2007.

[26] Sippl, M.J., Recognition of Errors in Three-Dimensional Structures of Proteins, Proteins: Structure, Function, and Genetics, Vol.17, No.4, pp.355-362, 1993.

[27] Zhang, Y. and Skolnick, J., TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score, Nucleic Acids Research, Vol.33, No.7, pp.2302-2309, 2005.

[28] Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E., Reference Points for Comparisons of Two-Dimensional Maps of Proteins from Different Human Cell Types Defined in a pH Scale Where Isoelectric Points Correlate with Polypeptide Compositions, Electrophoresis, Vol.15, No.1, pp.529-539, 1994.

[29] Cai, S. and Singh, B.R., Identification of β-turn and Random Coil Amide III Infrared Bands for Secondary Structure Estimation of Proteins, Biophysical Chemistry, Vol.80, No.1, pp.7-20, 1999.

[30] Pauling, L., Corey, R.B. and Branson, H.R., The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain, Proceedings of the National Academy of Sciences, Vol.37, No.4, pp.205-211, 1951.