In-Silico Analysis and Structural Prediction of *Catalase* Protein in Emu (*Dromaius novaehollandiae*) Through Homology Modeling

Thashi Bharadwaj[1], Rexon G[2], Tony Grace[3*]

[#1]*PhD student, Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Padanakkad, Kasaragod, Kerala, India*

[2]*Assistant Professor, Department of Genomic Science, School of Biological Sciences, Central University of Kerala, Padanakkad, Kasaragod, Kerala, India*

***Abstract*—** *Applicability of immune defense proteins from non-mammalian species can be beneficial in devising novel disease management strategies to overcome the concerns of resistance to current antibiotic therapeutics. Understanding the proteins involved in antimicrobial defense against infections can also aid in developing new generation of antibiotics. Evolution of new emerging diseases and atypical host diseases, it is imperative that we have in depth understanding of the presence and evolution of different proteins involved in defense mechanisms. Here we report the identification and structural dynamics of Catalase enzyme in Emu, a member of the oldest class of order of living birds.*

Keywords— *Structure prediction, Homology modeling, Catalase, Computational analysis, Immune defense.*

I. INTRODUCTION

In the past few decades, emergence of many drug resistant strains of microorganisms has given rise to increasing outbreak of unmanageable diseases[1] thus resulting in progressively difficult disease management practices.[2]Better understanding of the host immune defence mechanisms could potentially help us in finding novel therapeutic methods which would aid the prevalent disease management practices. Immune responses are inducible phenomena ensuing from a close association between environment, pathogen signal recognition systems and gene expression mechanism.[3-5] After pathogen recognition, activation of several transduction pathways occurs which in turn leads to the activation of transcription factors that induce gene expression.[6,7]Considerable interest has been developed in elucidating avian immune response mechanisms, understanding the proteins involved in immune regulation and the resulting transcriptional regulation.[8-11]Apart from that it could also endow us with the rational foundation for developing strategies to control various diseases. Immune system genes are believed to be under keen selective pressures to cope with pathogenic attacks and therefore are subject to protein-level sequence changes. This theory is reinforced by studies suggesting that in chicken (*Gallus gallus*) immune defence genes are evolving under greater positive selection than other genes.[12]

Emu (*Dromaius novaehollandiae*), a member of ratite family of flightless birds, is among the oldest Order of living birds, *Palaeognathiformes*.[13] Their phylogenetic relationship like other members of this family (emu, kiwi, rhea, cassowary), can be linked to the flightless ancestors that are believed to have inhabited the Southern Hemisphere 80 to 90 million years ago.[14] They are native to Australia, but found widely dispersed across varying habitats types of the continent. Successful wild and captive breeding capacity of emus is attributed to its ability to tolerate wide range of temperature owing to effective thermoregulation, varied diet patterns and heightened immunity.[15,16] Though morphological variation among emu is very less, very little data is available on their genetic variability. Phylogenetic relationships among ratite birds have been examined through mitochondrial sequence analysis in several studies.[17] However emu genomic information is yet not publicly available as of now.

Emergence and blossoming of emu industry in the 1980s, for the purpose of commercial production of meat, oil, and leather, has heightened the interest and importance in characterizing emu pedigrees.[13] Today, emus are farmed worldwide including in India for their fat (emu oil) and meat.[18-19] Emu oils popularly claimed by marketers to possess anti-inflammatory and anti-oxidative effects.[20] When we endeavor to introduce this species in a new habitat, care and precaution is required. Population size of Emus bred at each farm appears variable over time. Further the lack of proper information of the genetic variation in wild populations has limited the evaluation of expected variations in farmed stocks. More studies and data are a requisite, which is crucial in pedigree analysis, inbreeding assessment, and identifying markers for desirable heritable traits. This would hopefully aid in improving their genetic management.

Catalase gene (CAT1) codes for the antioxidant enzyme *catalase* that has a critical role in body's oxidative stress defense mechanism.[21]It is mostly restricted in peroxisomes and contains iron porphyrin enzyme having a key functional role in scavenging of reactive oxygen species (ROS). It removes unwarranted H_2O_2(hydrogen peroxide) during developmental process or biotic/abiotic stress thus avoiding oxidative damage and thus protecting the cell from toxic effects of hydrogen peroxide[22].It is found in most of the aerobically respiring organisms and serves to protect cells from the toxic effects of hydrogen peroxide. It also plays a major role in promoting the growth of various types of cells.

Understanding the gene and protein structural and sequence physiological parameters can help us in better understanding and comparing the oxidative stress mechanisms in organisms and human beings. The variations if present in Emu can also indicate and help us gain insights on the stress defense mechanisms a non-native organism goes through when introduced.

2. Materials and Methods

2.1 Tissue Materials

Tissue samples from adult male emu birds were collected from slaughterhouses. As per slaughterhouse records, the birds were sourced from emu farms of Nileshwar, Kerala, India. Tissues from multiple organs (liver, heart and lungs) were collected in tubes with RNAlater.

2.2 RNA Extraction

Total RNA of tissue samples were extracted using Trizol reagent according to the manufacturer's protocol (Ambion®). The absorbance ratio at 260 nm, 280 nm were used to assess the quality of the RNA and the 28 S and 18 S were visualized through electrophoresis. DNase I was used to remove DNA contamination. The first-strand cDNA synthesis was completed using Transcriptor first strand cDNA synthesis kit (Roche, Inc).

2.3 Primer designing and Sequencing

The primers used in the study were designed using the freely available online primer-designing tool Primer3 (http://bioinfo.ut.ee/primer3/).[23] The template sequence used for designing the primer was based on the available *Catalase* sequence of other avian species retrieved from NCBI database.. The designed primers were used to obtain PCR products using cDNA as the template DNA.

2.4 Primary and Secondary Sequence Analysis of Catalase protein

All *in-silico* analyses were carried out using tools available under the ExPASy server.[24]The DNA sequence was translated to protein sequence using the translate tool in ExPASy (http://web.expasy.org/translate/). The physico-chemical parameters of the obtained sequence were analyzed using ProtParam tool(http://web.expasy.org/protparam/). The parameters considered for analysis include molecular weight, theoretical PI, amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY).
TMHMM(http://www.cbs.dtu.dk/services/TMHMM/) was used to predict the presence of transmembrane region of the protein.[25] Secondary structure prediction was done via SOPMA server protein (https://npsa-prabi.ibcp.fr/).[26]

2.5 Prediction and validation of 3D structure

The three dimensional structure was predicted and validated using Phyre[2] (Protein Homology/analogy Recognition Engine V2.0).[27] It is considered as a very reliable and user friendly interface used for protein modelling. The structure generated predicted by the Phyre[2] server was retrieved in the PDB format and viewed using RASMOL visualization tool.[28]The secondary structural indicators like alpha helix, beta sheets, random coils and extended strands were analyzed using RasMol. Structure was validated through Ramachandran plot and stereo chemical analysis of the predicted structure was carried out via the PROCHECK software downloaded from the SAVES server (The Structure Analysis and Verification Server version 4).[29]

3. Results and Analysis

3.1 Primary and secondary sequence analysis

The BLAST result (Fig 1) showed maximum homology with the avian *Catalase* protein sequences available in NCBI database.

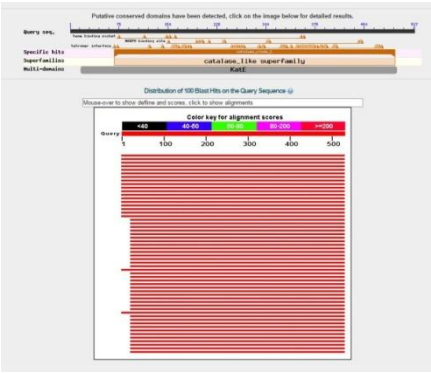

Fig: 1.* BLAST result for *Catalase* protein from *Dromaius novaehollandiae* against *NCBI database

FASTA sequence of the *Catalase* protein sequence was subjected to ProtParam analysis (Fig 2). The protein was found to have 527 amino acid residues with a molecular weight of 59648.0 (Table 1). The

theoretical PI was found to be 6.74. The number of positively and negatively charged residues in the sequence obtained was 62 and 64 respectively. Aliphatic index, defined as the relative volume occupied by aliphatic side chains that represents the thermal stability of globular protein was 66.96. Grand Average of Hydropathy (GRAVY) of the protein sequence was -0.591.

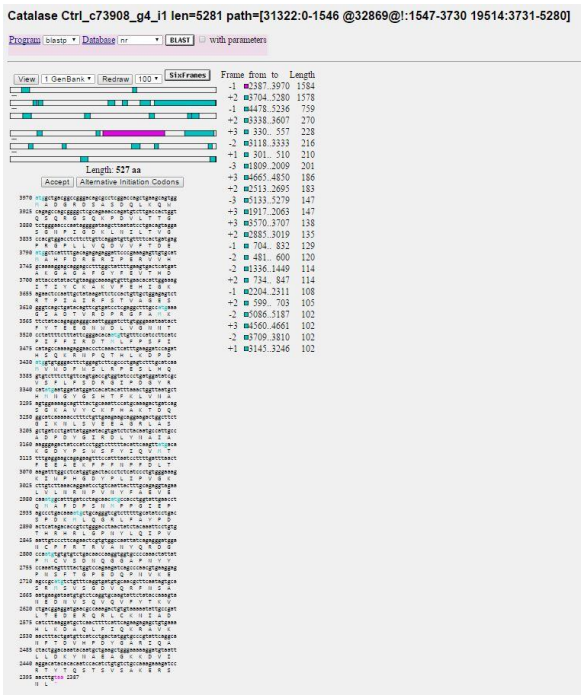

Fig: 2. Open reading frame obtained through ORF finder

Element	Symbol	Number of atoms
Carbon	C	2661
Hydrogen	H	4063
Nitrogen	N	743
Oxygen	O	790
Sulfur	S	17
Total		8274

Table 1: Elemental composition of catalase protein in Emu

According to SOPMA secondary structure prediction tool there were 24.86 % alpha helix, 24.29% extended strands, 11.39 % beta turns and 39.47% random coils in *Catalase* (Fig 3). According to TMHMM server prediction, no transmembrane helices were found in this protein. SOPMA predicted secondary structure was further validated using another tool PSPIRED (Fig 4).

Similar prediction results were obtained in both SOPMA and PSPIRED analysis,

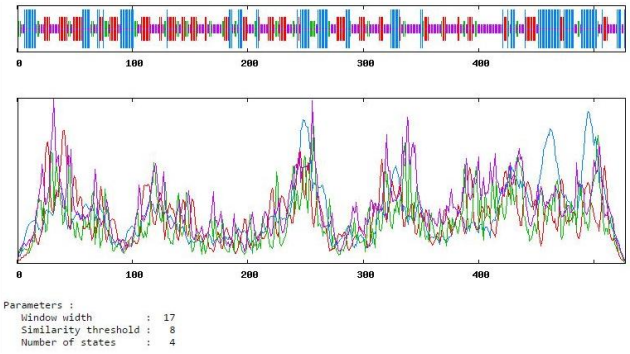

Fig: 3. Secondary structure of Catalase protein predicted by SOPMA

Fig:4. The secondary structure of Catalase protein predicted by PSIPRED

3.2 3D Structure generation and validation

The predicted three dimensional structure of *Catalase* generated through Phyre[2]-webportal and visualized using RASMOL revealed that the predicted protein had 499 groups, 3996 atoms, 4109 bonds and 301hydrogen bonds (Fig. 5). These are three dimensionally packed via19 helices, 57 turns and 20 strands.

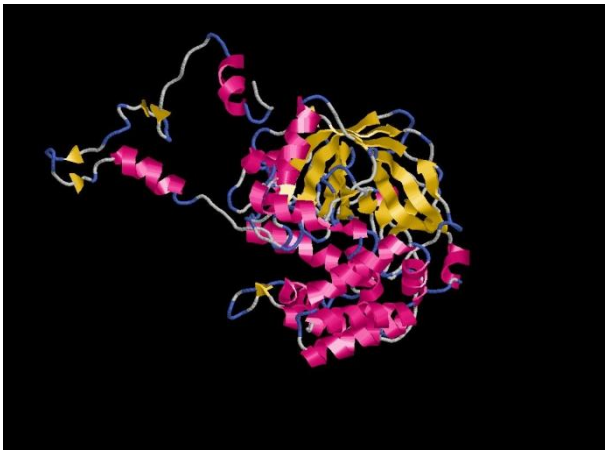

Fig:5. Catalase predicted structure in Rasmol viewer

Structural analysis of the predicted protein using Ramachandran plot validated that 91.1% residues fell in the favoured region enabling protein functionality. The results of RM plot analysis and statistical probabilities are shown in Figure 6 and Table 2 respectively.

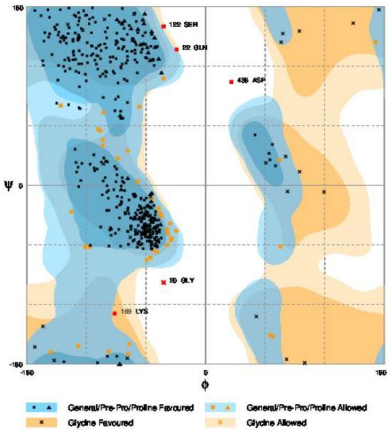

Fig:6. Ramachandran plot of predicted structure of Catalase protein

Residue Position Feature	Probability
Residue in most favoured region	453 (91.1%)
Residues in allowed regions	39 (7.8%)
Residues in outlier regions	5 (1.0%)

Table: 2. Ramachandran plot statistics for predicted Catalase protein in Emu

4. CONCLUSION

Catalase enzyme plays a major role in the oxidative stress defense mechanism. The oxidative stress mechanism of non-native organisms in a foreign organism can be different from its inherent mechanism. Understanding the major differences could potentially help us to regulate and negotiate the deleterious effects of oxidative stress induced cellular damage. In our study, the structure of catalase from Emu as predicted through the various *in-silico* tools was found to be highly stable and showed closed similarity with those of the predicted structures of *Catalase* enzyme found in other organisms. The primary and secondary structures predictions gave clear indications on the parameters including molecular weight, theoretical PI, amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY). As expected and already known, the obtained sequence was not a transmembrane protein and contained no signal peptides. Thus we conclude from our study that by utilizing robust computational tools, we could further predict various other protein structures in Emu which can be utilized to help us understand the organism's bio-physiology and thus could also throw light into the evolutionary pattern of the proteins, and in general the ecological and evolutionary significance of this bird as this species belongs to the oldest order of living birds.

ACKNOWLEDGEMENTS

Central University of Kerala.

REFERENCES

[1] DaSilva, E. and Iaccarino, M., Emerging diseases: a global threat. Biotechnology Advances, 17(4), pp.363-384, 1999.

[2] Bloom, David E., and David Canning. Epidemics and economics.na, 2006.

[3] Akira, S., Uematsu, S. and Takeuchi, O., Pathogen recognition and innate immunity.Cell, 124(4), pp.783-801, 2006.

[4] Mogensen, T.H., Pathogen recognition and inflammatory signaling in innate immune defenses. Clinical microbiology reviews, 22(2), pp.240-273, 2009.

[5] Thompson, M.R., Kaminski, J.J., Kurt-Jones, E.A. and Fitzgerald, K.A., Pattern recognition receptors and the innate immune response to viral infection.Viruses, 3(6), pp.920-940, 2011

[6] Aderem, A. and Ulevitch, R.J., Toll-like receptors in the induction of the innate immune response. Nature, 406(6797), pp.782-787 2000.

[7] Bonizzi, G. and Karin, M., The two NF-κB activation pathways and their role in innate and adaptive immunity. Trends in immunology, 25(6), pp.280-288, 2004.

[8] de Geus, E.D., Tefsen, B., van Haarlem, D.A., van Eden, W., van Die, I. and Vervelde, L., Glycans from avian influenza virus are recognized by chicken dendritic cells and are targets for the humoral immune response in chicken. Molecular immunology, 56(4), pp.452-462, 2013.

[9] . Staines, K., Young, J.R. and Butter, C., Expression of chicken DEC205 reflects the unique structure and function of the avian immune system. PloS one, 8(1), p.e51799, 2013.

[10] Gong, Q., Qu, N., Niu, M., Qin, C., Cheng, M., Sun, X. and Zhang, A., Immune responses and protective efficacy of a novel DNA vaccine encoding outer membrane protein of avian Pasteurellamultocida. Veterinary immunology and immunopathology, 152(3), pp.317-324, 2013.

[11] Min, W., Kim, W.H., Lillehoj, E.P. and Lillehoj, H.S., Recent progress in host immunity to avian coccidiosis: IL-17 family cytokines as sentinels of the intestinal mucosa. Developmental & Comparative Immunology, 41(3), pp.418-428, 2013.

[12] Tim Downing, Paul Cormican, Cliona O'Farrelly, Daniel G Bradley and Andrew T Lloyd, Evidence of the adaptive evolution of immune genes in chicken BMC Research Notes, 2:254 doi:10.1186/1756-0500-2-254, 2009.

[13] Hammond, E.L., Lymbery, A.J., Martin, G.B., Groth, D. and Wetherall, J.D., Microsatellite analysis of genetic diversity in wild and farmed emus (Dromaius novaehollandiae). Journal of Heredity, 93(5), pp.376-380, 2002.

[14] Roff, D.A., The evolution of flightlessness: is history important?.Evolutionary Ecology, 8(6), pp.639-657, 1994.

[15] Maina, J.N. and King, A.S., The lung of the emu, Dromaiusnovaehollandiae: a microscopic and morphometric study. Journal of anatomy, 163, p.67, 1989.

[16] Herd, R.M. and Dawson, T.J., Fiber digestion in the emu, Dromaiusnovaehollandiae, a large bird with a simple gut and high rates of passage. Physiological Zoology, pp.70-84, 1984.

[17] Baker, A.J., Daugherty, C.H., Colbourne, R. and McLennan, J.L., Flightless brown kiwis of New Zealand possess extremely subdivided population structure and cryptic species like small mammals. Proceedings of the National Academy of Sciences, 92(18), pp.8254-8258, 1995.

[18] Jeengar, M.K., Kumar, P.S., Thummuri, D., Shrivastava, S., Guntuku, L., Sistla, R. and Naidu, V.G.M., Review on emu products for use as complementary and alternative medicine. Nutrition, 31(1), pp.21-27, 2015.

[19] Warale, R.H., Chauhan, H.D., Parmar, D., Kulkarni, R.C., Srivastava, A.K., Makwana, R.B., Pawar, M.M. and Bhagwat, S.R., Emu Farming: An Alternative to Indian Poultry, 2014.

[20] Bennett, D.C., Code, W.E., Godin, D.V. and Cheng, K.M., Comparison of the antioxidant properties of emu oil with other avian oils. Animal Production Science, 48(10), pp.1345-1350, 2008.

[21] Guan, L.M., Zhao, J. and Scandalios, J.G., Cis-elements and trans-factors that regulate expression of the maize Cat1 antioxidant gene in response to ABA and osmotic stress: H2O2 is the likely intermediary signaling molecule for the response. The Plant Journal, 22(2), pp.87-95, 2000.

[22] Su, Y., Guo, J., Ling, H., Chen, S., Wang, S., Xu, L., Allan, A.C. and Que, Y., Isolation of a novel peroxisomal catalase gene from sugarcane, which is responsive to biotic and abiotic stresses. PloS one, 9(1), p.e84426, 2014.

[23] Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G.,. Primer3—new capabilities and interfaces. Nucleic acids research, 40(15), pp.e115-e115, 2012.

[24] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D. and Bairoch, A., ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research, 31(13), pp.3784-3788, 2003.

[25] Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology, 305(3), pp.567-580, 2001.

[26] Geourjon, C. and Deleage, G., SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer applications in the biosciences: CABIOS, 11(6), pp.681-684, 1995.

[27] Kelley, L.A. and Sternberg, M.J., Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols, 4(3), pp.363-371, 2009.

[28] Sayle, R. and Bissell, A., RasMol: A program for fast, realistic rendering of molecular structures with shadows. In Proceedings of the 10th Eurographics UK (Vol. 92, pp. 7-9), 1992.

[29] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M., PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography, 26(2), pp.283-291, 1993.