BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 11(2)

Comparative sequence-structure analysis of Aves insulin

Md Mirazul Islam*, M Aktaruzzaman & Zahurin Mohamed

Pharmacogenomics Lab, Department of Pharmacology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia; Md Mirazul Islam – E mail: mirazul.du.bc@gmail.com; *Corresponding author

Received December 15, 2014; Accepted January 26, 2015; Published February 28, 2015

Abstract:
Normal blood glucose level depends on the availability of insulin and its ability to bind insulin receptor (IR) that regulates the downstream signaling pathway. Insulin sequence and blood glucose level usually vary among animals due to species specificity. The study of genetic variation of insulin, blood glucose level and diabetics symptoms development in Aves is interesting because of its optimal high blood glucose level than mammals. Therefore, it is of interest to study its evolutionary relationship with other mammals using sequence data. Hence, we compiled 32 Aves insulin from GenBank to compare its sequence-structure features with phylogeny for evolutionary inference. The analysis shows long conserved motifs (about 14 residues) for functional inference. These sequences show high leucine content (20%) with high instability index (>40). Amino acid position 11, 14, 16 and 20 are variable that may have contribution to binding to IR. We identified functionally critical variable residues in the dataset for possible genetic implication. Structural models of these sequences were developed for surface analysis towards functional representation. These data find application in the understanding of insulin function across species.

Key words: Aves Insulin, blood glucose level, sequence-structure analysis, physic-chemical properties.

Background:
Diabetes mellitus (DM) is one of the predominant diseases that affect presently ~ 382 million people all over the world and its incidence is expected to increase to 592 million by 2035(according to international diabetes federation). Insulin level or binding ability to IR is the major determinant factor of DM. Insulin is a globular protein central to the regulation of vertebrate carbohydrate metabolism. It is one of the most important hormones, carrying messages that describe the amount of available sugar from moment to moment in the blood. Insulin is the primary regulator of carbohydrate homeostasis and has effect on lipid and protein metabolism **[1, 2].** The mechanism of action of these hormones is mediated by their specific binding to the Insulin Receptor (IR) **[3].** The binding of insulin to IR leads to activation of the tyrosine kinase function of the intracellular part of the receptor and subsequent transporter activation as well as increase cellular uptake of glucose **[4].** A confirmatory repeat blood sugar level ≥140 mg/100 ml proved valuable in defining a high risk group for diabetes in human **[5].** But there is an unusually high blood glucose level found in birds without diabetes or any associated consequences **(Figure 1).** Normal plasma glucose levels in some birds is three to four times higher than human **[6].** May be birds have some intrinsic mechanism to control blood glucose levels without showing diabetic symptoms. Comparative analysis of Aves insulin may gives some ideas about the mechanism.

Insulin is made in the pancreas and added to the blood after meals when sugar levels are high. This signal then spreads throughout the body, to the liver, muscles and fat cells. Insulin tells these organs to uptake glucose from the blood and stores in the form of glycogen or fat. The mechanism of insulin binding to insulin receptor and signal transduction through the transmembrane domain has vital role to maintain blood glucose levels **[7].** Structure of insulin is the key to protein

function and interaction to IR. Evolution of insulin gene and its promoter has started over a 450 million-year period **[8].** Protein structure analysis can provide lots of complex information about protein functions related disorders. Wet lab based research requires the trial and error method and cannot make a prediction before the original result. This problem can be overcome by the use of computational biology. Alteration in protein structure leads to altered protein function which in turn leads to development of diseases **[9].** The target of this research is to give an intrinsic view of Aves insulin that may suggest an important idea about control mechanism of blood sugar level as well as recombinant human insulin development.

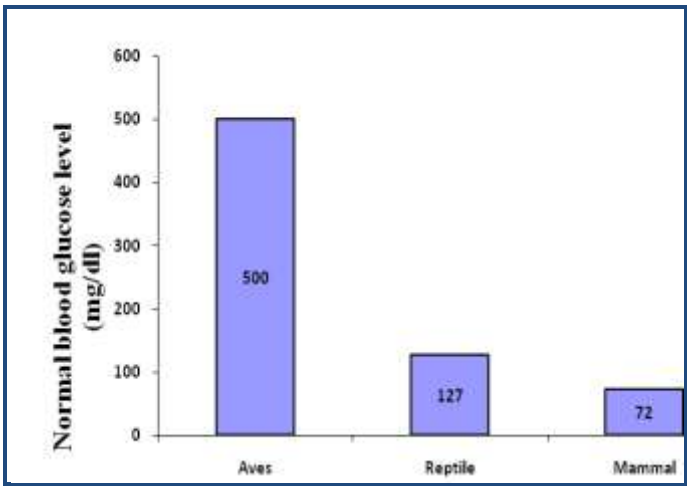

Figure 1: Variation of normal blood glucose level [15]. There is a drastic fluctuation of normal blood glucose level among Aves, Reptile and Mammal. Aves glucose level is four times higher than Reptile and seven times higher than Mammal.

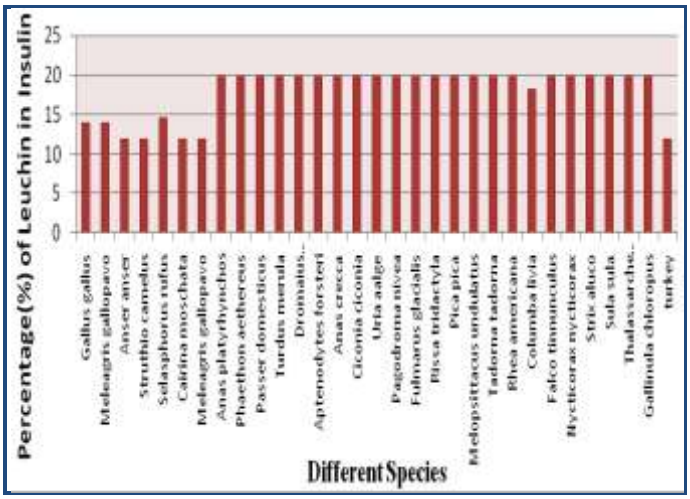

Figure 2: Percentage of Leuchin in Aves insulin. Around 20 % of total amino acid in Aves insulin is Leuchin. It is hydrophobic amino acid and the reason behind this high percentage of Leu is not fully known yet.

Methodology:

Protein sequence retrieval

Thirty two Aves insulin and ten mammalian insulin sequences were collected from UniProt (http://www.uniprot.org/). We preferred most commonly available mammal and all Aves sequences found UniProte database until mid June, 2013. Those sequences were used for further analysis by online or freely available computational tools.

Analysis of Physico-chemical properties

The ProtParam tool (http://web.expasy.org/protparam/) of ExPASy was used to compute amino acid composition (%), molecular weight, theoretical isoelectric point (pI), number of positively and negatively charged residues, extinction coefficient, instability and aliphatic index, Grand Average of Hydropathy (GRAVY).

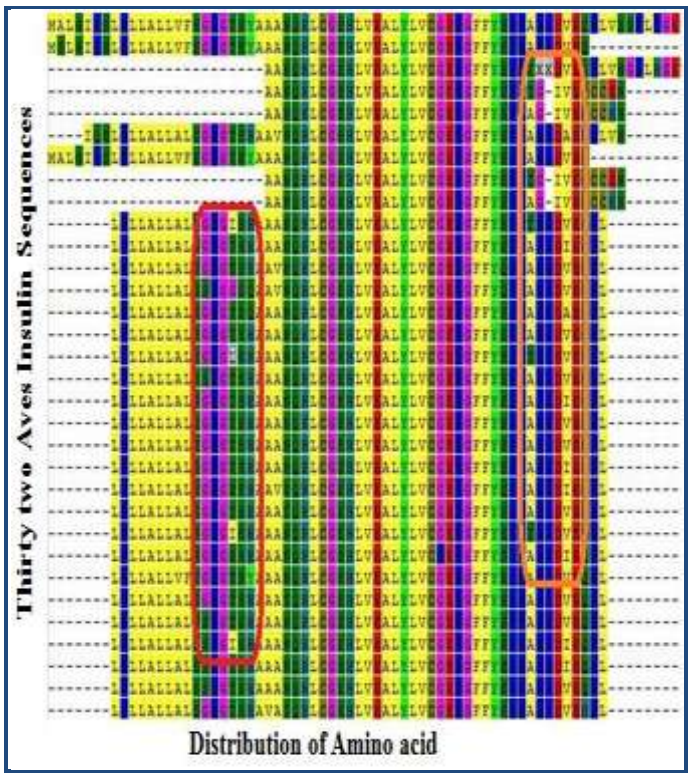

Figure 3: Multiple (thirty two) Aves insulin sequences alignment using MEGA5.1 software. It is clearly found that two particular regions are usually show variability. Left region show Ile and Tyr are variable and right region show Thr, Gly and Ala are variable.

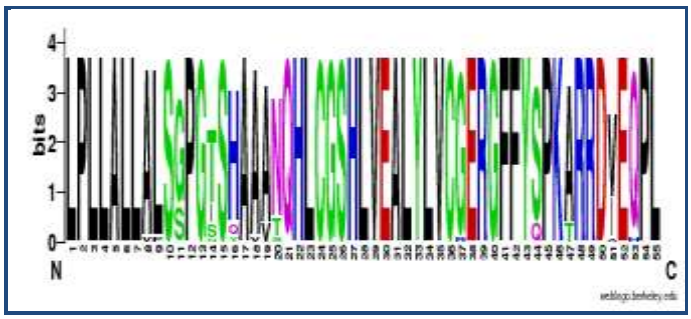

Figure 4: WebLogo representation of Aves Insulin. The amino acid types and position are shown on the x axis. The overall height of the amino acid stacks, plotted on the y axis, indicates the sequence conservation at a given position, while the height of individual symbols within a stack indicates the relative frequency of an amino acid at that position. Amino acids are color coded according to their type as basic (blue), hydrophobic (black), polar/nonpolar (green), and acidic (red).

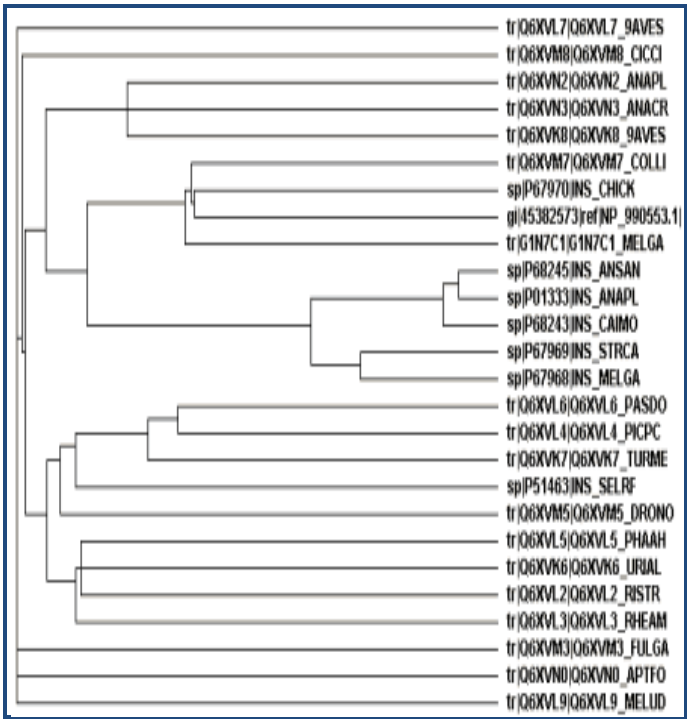

Figure 5: Phylogenetic tree of Aves insulin sequences by using Neighbor-Joining Method. Closely related species like Anser anser (>sp|P68245|)and Cairian moschata (>sp|P68243|) are found nearby but the location of Melopsittacus undulates (>tr|Q6XVL9|) is different from them, as it is not there neighbor species.

Analysis of Secondary structural properties

Secondary structural properties of the protein including alpha helix, 3_{10} helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous states and other states were computed by the use of SOPMA (Self Optimized Prediction Method with Alignment, http://npsapbil.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) tool of NPS (Network Protein Sequence Analysis) **[10]**.

Prediction of functional properties

The motif prediction analysis was carried out with the help of Expasy's prosite tool. For functional analysis, the motifs of the insulin protein sequences were identified by using Prosite (http://prosite.expasy.org/). Input data type was in FASTA format and motifs were scanned against prosite patterns.

Identification of Signature Logo using Web tool

Logo of Aves insulin was generated using Web Logo tool (http://weblogo.berkeley.edu/). In this overall height of the stack indicates the sequence conservation at that position, while the heights of the symbols within the stack indicate the relative frequency of each amino acid at that position.

Sequence alignment

Insulin sequences were align by using MEGA5.1 and identify the variable region that may be responsible for functional activity of high plasma glucose level. Direct comparison between human and turkey insulin sequence is given below and box shows the changes of amino acids. It indicate completely different types of (in term of hydrophobic and hydrophilic) amino acid changes in between this two species.

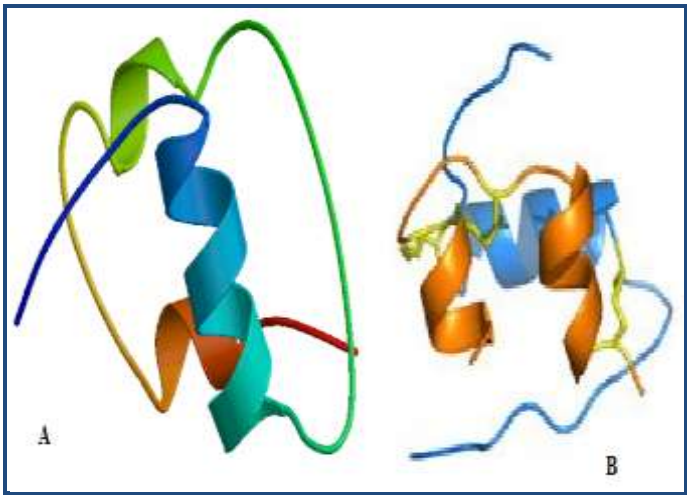

Figure 6: 3D structure of Turkey (left) and Human (right) Insulin was determined using SWISS-Model. A chain (light red) and B chain (light blue) are linked via two inter-chain disulfide bond among Cys residues and one intra-chain bond in A chain.

Insulin Sequence for Human

Phe-Val-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala-Gly-Ile-Val-Glu-Gln-Cys-Cys-Ala-Ser-Val-Cys-Ser-Leu-Thr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

Insulin Sequence for Turkey:

Ala-Ala-Asn-Gln-His-Leu-Cys-Gly-Ser-His-Leu-Val-Glu-Ala-Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Ser-Pro-Glu-Ala-Gly-Ile-Val-Glu-Gln-Cys-Cys-His-Asn-Thr-Cys-Ser-Leu-Thr-Gln-Leu-Glu-Asn-Tyr-Cys-Asn

Phylogenetic analysis

Thirty two sequences of Aves insulin were aligned by ClustalW tool and output file of this program was used for generation of phylogenetic tree (http:// www.ebi.ac.uk/Tools/msa/clustalw2/) by using Neighbor-Joining method.

Results & Discussion:

The Physiochemical characterization, secondary structure properties, motif and phylogenetic analysis was done by using different computational tools for 32 Aves insulin sequences. Insulin sequences contain leucine around 20% of their amino acids which is significantly higher than other **(Figure 2).** The total number of positively (Arg + Lys) and negatively (Asp + Glu) charged residues were quite same, that's why pI was ~7. Extinction coefficient for all Insulin was observed higher. High extinction coefficient means higher concentration of lysine, tryptophan and tyrosine. This prediction is useful to study protein-protein interaction studies. The higher aliphatic index indicates higher thermostability and higher concentration of alanine, valine, isoleucine and leucine occupying the relative volume of a protein. A protein is stable or not can be described by its instability index. Instability index for Insulin in most case is higher than 40 and thus describing these proteins as unstable **[11].**

Average of Hydropathy (GRAVY) was computed for all the members. A broad range of GRAVY value was observed from 0.304 to -0.006 for Insulin **Table 1 (see supplementary**

material). SOPMA analysis was done for all insulin members and it showed a high value for random coil in all the members **Table 2 (see supplementary material).** High value for random coil bears important significance in the study of protein tertiary structure and related functions. Functional analysis of these proteins includes identification of important motifs **Table 3 (see supplementary material).** Only eight proteins show functional motif within 32 sequences**.** These motifs were 14 amino acids in length arise because specific residues and regions proved to be important for the biological function of a group of proteins, which are conserved in both structure and sequence during evolution **[12].** For observing variability of Aves insulin sequences, MEGA5.1 software was used **(Figure 3).** WebLogo was designed (from weblogo.berkeley.edu) to show variable and constant amino acid position **(Figure 4).** Amino acid position 11, 14, 16 and 20 are variable on species to species. Phylogenetic tree was constructed with distance based Neighbor-Joining method. A number of clusters were found **(Figure 5)**. 3D structure of turkey and human insulin are determined (**Figure 6**) Proteins in close evolutionary relationship may be analyzed together for their involvement in similar biological processes. Another important finding is the structural differences between IR. There are sequence deletion found between 743 to 755 and 1007 to 1012 in Aves IR (comparison between human and turkey). This IR and insulin sequences differences may play a major role for binding affinity as well as intracellular signaling pathway that control blood glucose level.

Conclusion:

In this study, detail information of Aves insulin was sequentially identified using various computational tools. Insulin is related to diabetic, a group of metabolic diseases in which a person has high blood sugar, either because the pancreas does not produce enough insulin, or because cells do not respond to the insulin that is produced **[13].** Some information influence to carry on the study like South Asian people have higher blood glucose levels than white European people **[14]** and SNP alleles in the IR gene are associated with typical migraine **[15].** Present investigation and information may provide a possible explanation for high blood glucose in Aves as well as species specificity of insulin. This information will help to design effective recombinant insulin for therapeutic application. However, this finding is not enough to establish the hypothesis and need further study and validation by experimental approaches.

Acknowledgment:

We would like to thank Dr. A.H.M.Nurun Nabi (Biochemistry and Molecular Biology, University of Dhaka) for his inspiration to do this work.

Conflict of Interest:

Authors declared no conflict of interest.

References:

[1] Saltiel AR & Kahn CR, *Nature* 2001 **414**: 799 [PMID: 11742412]
[2] Kitamura T *et al. Annu Rev Physiol.* 2003 **65**: 313 [PMID: 12471165]
[3] Nakae J *et al. Endocr Rev.* 2001 **22**: 818 [PMID: 11739335]
[4] Kasuga M *et al. Nature* 1982 **298**: 667 [PMID: 6178977]
[5] O'Sullivan JB *et al. JAMA*. 1965 **194**: 587 [PMID: 5897230]
[6] O'Donnell JA *et al. J Wildl Dis*. 1978 **14**: 479 [PMID: 739587]
[7] Schaffer L, *Eur J Biochem*. 1994 **221**: 1127 [PMID: 8181471]
[8] Hay CW & Docherty K, *Diabetes* 2006 **55**: 3201 [PMID: 17130462]
[9] Saha A *et al. Bioinformation* 2013 **9**: 42 [PMID: 23390343]
[10] Geourjon C & Deleage G, *Comput Appl Biosci*. 1995 **11:** 681 [PMID: 8808585]
[11] http://web.expasy.org/docs/expasy_tools05.pdf
[12] Rani S *et al. Bioinformation* 2013 **9**: 572 [PMID: 23888098]
[13] Shoback, edited by David G. Gardner, Dolores (2011). Greenspan's basic & clinical endocrinology (9th ed.). New York: *McGraw-Hill Medical*. pp. Chapter 17.
[14] Mostafa SA *et al. Diabetes Care.* 2012 **35**: 1746 [PMID: 22699291]
[15] Delgado C *et al. J Wildl Dis.* 2011 **47**: 523 [PMID: 21719817]

Edited by P Kangueane

Citation: **Islam *et al.*** Bioinformation 11(2): 067-072 (2015)

BIOINFORMATION

open access

Supplementary material:

Table 1: Various parameters computed using Expasy's ProtParam tool of Aves Insulin

Protein	Accession No.	Length	Molecular Weight	pI	-R	+R	EC	II	AI	Gravy
Gallus gallus	>sp|P67970	107	11980.7	6.27	10	9	14815	55.42	91.12	-0.051
Meleagris gallopavo	>sp|P67968	107	11996.7	6.27	10	9	14815	58.37	90.19	-0.076
Turkey	>sp|P67968	51	5698.4	5.78	4	2	6335	28.53	78.43	-0.006
Anser anser anser	>sp|P68245	51	5716.4	4.98	5	2	6335	15.45	76.47	-0.078
Struthio camelus	>sp|P67969	51	5698.4	5.78	4	2	6335	28.53	78.43	-0.006
Selasphorus rufus	>sp|P51463	103	11377.9	5.80	11	7	6335	36.90	90.87	-0.158
Cairina moschata	>sp|P68243	51	5716.4	4.98	5	2	6335	15.45	76.47	-0.078
Meleagris gallopavo	>sp|P67968	51	5698.4	5.78	4	2	6335	28.53	78.43	-0.006
Anas platyrhynchos	>tr|Q6XVN2	55	5930.8	7.01	4	4	3105	38.83	111.82	0.253
Phaethon aethereus	>tr|Q6XVL5	55	5902.8	7.01	4	4	3105	46.81	108.36	0.209
Passer domesticus	>tr|Q6XVL6	55	5943.8	7.01	4	4	3105	39.38	110.00	0.196
Turdus merula	>tr|Q6XVK7	55	5934.8	6.91	4	4	3105	53.12	110.00	0.191
Dromaius novaehollandiae	>tr|Q6XVM5	55	5851.7	6.91	4	4	3105	42.33	103.09	0.155
Aptenodytes forsteri	>tr|Q6XVN0	55	5888.8	7.01	4	4	3105	38.83	106.55	0.204
Anas crecca	>tr|Q6XVN3	55	5930.8	7.01	4	4	3105	38.83	111.82	0.253
Ciconia ciconia	>tr|Q6XVM8	55	5918.8	7.01	4	4	3105	50.32	106.55	0.196
Uria aalge	>tr|Q6XVK6	55	5902.8	7.01	4	4	3105	46.81	108.36	0.209
Pagodroma nivea	>tr|Q6XVL7	55	5888.8	7.01	4	4	3105	38.83	106.55	0.204
Fulmarus glacialis	>tr|Q6XVM3	55	5888.8	7.01	4	4	3105	38.83	106.55	0.204
Rissa tridactyla	>tr|Q6XVL2	55	5902.8	7.01	4	4	3105	46.81	108.36	0.209
Pica pica	>tr|Q6XVL4	55	5930.8	7.01	4	4	3105	52.25	111.82	0.251
Melopsittacus undulatus	>tr|Q6XVL9	55	5875.8	7.01	4	4	3105	38.83	106.55	0.255
Tadorna tadorna	>tr|Q6XVK8	55	5930.8	7.01	4	4	3105	38.83	111.82	0.253
Rhea americana	>tr|Q6XVL3	55	5987.9	8.06	4	5	3105	53.23	108.36	0.133
Columba livia	>tr|Q6XVM7	55	5976.9	6.91	4	4	4595	43.13	102.91	0.264
Falco tinnunculus	>tr|Q6XVM4	55	5875.8	7.01	4	4	3105	38.83	106.55	0.255
Nycticoraxhoactli	>tr|Q6XVL8	55	5918.8	7.01	4	4	3105	50.32	106.55	0.196
Strix aluco	>tr|Q6XVL1	55	5944.9	7.01	4	4	3105	58.31	115.45	0.296
Sula sula	>tr|Q6XVK9	55	5932.8	7.01	4	4	3105	58.31	108.36	0.202
Thalassarche melanophrys	>tr|Q6XVM6	55	5918.8	7.01	4	4	3105	50.32	106.55	0.196
Gallinula chloropus	>tr|Q6XVM2	55	5912.8	7.08	4	4	3105	29.93	110.00	0.304

Table 2: Various parameters computed using Expasy's ProtParam tool of Aves Insulin

Protein	ἀHelix	310Helix	PiHelix	βBridge	Extended Strand	B-Turn	Bend Region	Random Coil	Ambiguous States	Other States
Gallus gallus	44	0	0	0	15	8	0	40	0	0
Meleagris gallopavo	44	0	0	0	15	8	0	40	0	0
Anser anser	20	0	0	0	10	6	0	15	0	0
Struthio camelus	22	0	0	0	8	6	0	15	0	0
Selasphorus rufus	42	0	0	0	15	11	0	35	0	0
Cairina moschata	20	0	0	0	10	6	0	15	0	0
Meleagris gallopavo	22	0	0	0	8	6	0	15	0	0
Anas platyrhynchos	18	0	0	0	10	7	0	20	0	0
Phaethon aethereus	20	0	0	0	11	5	0	19	0	0
Passer domesticus	18	0	0	0	12	7	0	18	0	0
Turdus merula	17	0	0	0	9	6	0	23	0	0
Dromaius novaehollandiae	21	0	0	0	10	5	0	19	0	0
Aptenodytes forster	22	0	0	0	12	4	0	17	0	0
Anas crecca	18	0	0	0	10	7	0	20	0	0
Ciconia ciconia	22	0	0	0	12	6	0	15	0	0

Uria aalge	20	0	0	0	11	5	0	19	0	0
Pagodroma nivea	22	0	0	0	12	4	0	17	0	0
Fulmarus glacialis	22	0	0	0	12	4	0	17	0	0
Rissa tridactyla	20	0	0	0	11	5	0	19	0	0
Pica pica	18	0	0	0	13	7	0	17	0	0
Melopsittacus undulatus	19	0	0	0	12	7	0	17	0	0
Tadorna tadorna	18	0	0	0	10	7	0	20	0	0
Rhea americana	22	0	0	0	13	3	0	17	0	0
Columba livia	15	0	0	0	11	5	0	24	0	0
Falco tinnunculus	19	0	0	0	12	7	0	17	0	0
Nycticorax nycticorax	22	0	0	0	12	6	0	15	0	0
Strix aluco	19	0	0	0	7	6	0	23	0	0
Sula sula	17	0	0	0	10	6	0	22	0	0
Thalassarche melanophrys	22	0	0	0	12	6	0	15	0	0
Gallinula chloropus	15	0	0	0	12	7	0	21	0	0
Turkey	22	0	0	0	8	6	0	15	0	0

Table 3: Motif prediction of Aves Insulin.

Protein	Motif ID	Position	Pattern
Gallus gallus	PS00262	92 - 106	CCHNtCSlyqLenyC
Meleagris gallopavo	PS00262	92 - 106	CCHNtCSlyqLenyC
Anser anser	PS00262	36 - 50	CCENpCSlyqLenyC
Struthio camelus	PS00262	36 - 50	CCHNtCSlyqLenyC
Selasphorus rufus	PS00262	88 - 102	CCHNtCSlyqLenyC
Cairina moschata	PS00262	36 - 50	CCENpCSlyqLenyC
Meleagris gallopavo	PS00262	36 - 50	CCHNtCSlyqLenyC
Turkey	PS00262	36 - 50	CCHNtCSlyqLenyC