



Evolutionary Relationship of Genomic Insulin Sequence in Different Mammalian Species: A Computational Approach

M. A Hashem^{1,*}, Neena Islam², Md. Moinul Abedin Shuvo¹, Md. Arifuzzaman¹

¹Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong (USTC), Foy's Lake, Chittagong, Bangladesh

²National Centre for Control of Rheumatic Fever and Heart Diseases, Sher-E-Bangla Nagar, Dhaka, Bangladesh

Email address:

hashem_bmb@yahoo.com (M. A Hashem), neenaislam@yahoo.com (N. Islam), moinulshuvo@yahoo.com (Md. M. A. Shuvo), larif67@yahoo.com (Md. Arifuzzaman)

*Corresponding author

To cite this article:

M. A Hashem, Neena Islam, Md. Moinul Abedin Shuvo, Md. Arifuzzaman. Evolutionary Relationship of Genomic Insulin Sequence in Different Mammalian Species: A Computational Approach. *Ecology and Evolutionary Biology*. Vol. 1, No. 2, 2016, pp. 23-28.

doi: 10.11648/j.eeb.20160102.13

Received: August 17, 2016; **Accepted:** September 5, 2016; **Published:** September 22, 2016

Abstract: Genomic insulin is located on the short arm of chromosome 11 in human genome. It is a well studied polypeptide hormone, consists of 110 amino acids which start with signaling peptide of 1-24 amino acids, B-chain of 25-54 amino acids, C-peptide of 55-89 amino acids and end with A-chain of 90-110 amino acids. Insulin, produced by the beta cell of the pancreas in response to glucose stimuli, binds to its receptor rapidly due to receptor autophosphorylation and primordially regulates nutritional metabolic pathways. In this study we have depicted and explored evolutionary conservation rate, insight into structure and phylogenetic connection of insulin molecule among eight mammalian species; *Homo sapiens* (Human), *Bos taurus* (Cattle), *Cavia porcellus* (Guinea pig), *Canis lupus familiaris* (Dog), *Gorilla gorilla* (Western gorilla), *Ovis aries* (Sheep), *Pan troglodytes* (Chimpanzee), *Pongo pygmaeus* (Orangutan) using Computational Biology. The analysis of physico-chemical characteristics, secondary and 3-D structure prediction of insulin in different species identified phylogenetically most related species. The major findings are that genomic insulin from Human and Dog has a lowest genetic distance of 0.13 of the mammalian species studied. Human and Guinea pig has the next lowest genetic distance of 0.39 and 69.1% identical at the amino acid level. Whereas Human and Western gorilla has genetic distances of 0.00 and 100% identical at the amino acid level and share a common node on the phylogenetic tree. Physico-chemical study also shows that these sequences show high leucine content (18.2%) with high instability index (>40) except Sheep and Cattle has low leucine and instability index (<40). The sequence analysis among species has allowed us to know the manner in which the insulin has evolved over million-year period. This study result provides rapid comprehensive information to calculate the amino acid sequences in relations to evolutionary conservation rates as well as molecular phylogenetics.

Keywords: Molecular Phylogenetics, Genomic Insulin, Multiple Sequence Alignment, Pairwise Distances, Physico-chemical Characteristics, Secondary Structure, 3D Structure Prediction

1. Introduction

Since the first appearance of life on Earth, development has been from simple life forms to more complex ones. There has been a progressive change over time, and over many generations, to produce different species from a common ancestor. In Evolutionary Biology, the phylogenetic tree represents the evolutionary histories of all organisms on the earth using molecular evidence. Evolutionary evidence

obtained from comparative studies of specific features as well as homologous characteristics of existing organisms to determine the evolutionary relationship among different species. Analyses of evolutionary relationship of central molecular networks uncover underlying ancestral variation that can be targeted by biomedical study to develop insights and interventions into disease [1]. The study of evolution changed dramatically with the discovery of genomic insulin and represents as an evolutionary evidence. Genomic

sequences of insulin and insulin-like signaling molecules are rapidly evolving within amniotes [1]. Evolutionary relationships in different mammalian species can be studied by comparing their genomic insulin sequences.

Insulin is a very old unique protein that may have derived more than a billion years ago [2], and share the same modular organization of their precursor, including N-terminal signal peptide followed by three domains (A, B and C domains). Insulin is synthesized within the beta cells of the islets of Langerhans as a precursor preproinsulin, and after removal of the signal peptide, proinsulin folds to form the exact tertiary structure, and the removal of C-peptide by cleaved by endoproteolytic enzymes [3]. The yielding of A and B domain are covalently linked to form mature insulin [4].

The mature structure of insulin consists of two polypeptide chains (B-chain of 30 amino acids, and A-chain of 21 amino acids), cross-linked by disulfide bonds. Genomic insulin of different mammalian species is structural building blocks that define as the function of a protein through their various combinations and also be the units of the evolutionary history of proteins as well as genomes that contain them.

Illustrating the entire evolutionary history of each insulin domain, from the genesis of a new domain or domain combination to the loss and transfer of a domain from a specific genome, can further our understanding of some of the fundamental unsolved problems in an evolutionary relationship, such as the trial in the early evolution of life.

2. Methods and Materials

2.1. Obtaining Genomic Insulin Sequences

Genomic Insulin sequences of eight different species retrieved from NCBI protein database in FASTA format. [<http://www.ncbi.nlm.nih.gov/>]. The accession numbers of these sequences are *Homo sapiens* (AAA59172.1), *Bos taurus* (ACD35246.1), *Cavia porcellus* (AAA37041.1), *Canis lupus familiaris* (P01321.1), *Gorilla gorilla* (Q8HXV2.1), *Ovisaries* (AAB60625.1), *Pan troglodytes* (P30410.1), *Pongo pygmaeus* (Q8HXV2.1);

2.2. Multiple Sequence Alignment

These genomic insulin sequences have analyzed on ClustalW [<http://www.ebi.ac.uk/clustalw/>] for the multiple sequence alignment. Sequences also analyzed using Geneious 7.1.2 [5] and a ClustalW algorithm are use to align multiple sequences in parallel.

2.3. Construction of Phylogenetic Tree

Sequences have aligned with ClustalW by the MEGA5 and output file of this program has used for the construct phylogenetic tree by using the Maximum Likelihood method and 10000 replicates use for bootstrap statistical test.

2.4. Pairwise Distances

Genetic distances have measured among these sequences by using MEGA 5 [6].

2.5. Physico-chemical Characteristics

The ExPASy ProtParam tool has used to compute different properties of genomic insulin including theoretical [pI], a number of positively [Arg + Lys] and negatively charged [Asp + Glu] amino acids, extinction co-efficient, instability index, aliphatic index and Grand Average of Hydropathicity [GRAVY].

The crystallization tendency of these sequences have determined by the CRYSTAL2 web server [7] [<http://biomine-ws.ece.ulberta.ca/CRYSTALP2.html>].

2.6. Characterization of Secondary Structure

Secondary structure properties of these sequences, including Alpha helix, 3_{10} Helix, Pi helix, Beta bridge, Extended strand, Beta turns, Beta region, Random coil, Ambiguous states and other states have analyzed by the SOPMA[8][http://npsa-bil.ibcp.fr/cgi-bin/npsa_automat.pl?pa=ge/NPSA/npsa_sopma.html].

2.7. 3D Structure Prediction

The 3D structures of these mammalian species are predicted by using the 3D-JIGSAW [version 3] comparative modeling server [9] [http://bmm.cancerresearchuk.org/~populus/populus_submit.html].

3. Results & Discussion

3.1. Multiple Sequence Alignment

Multiple sequence alignment of different color boxes are green, green-brown, and red color represent that 100%, 30-99%, and 0-29% identical respectively between species (Figure 1).

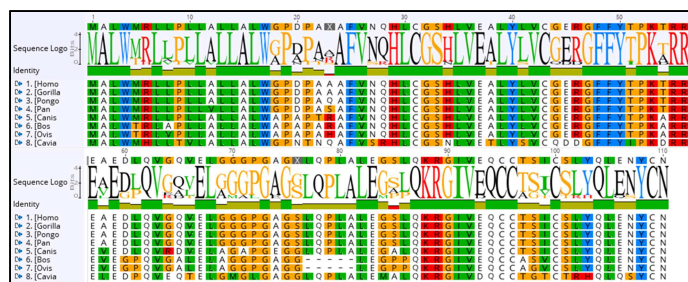


Figure 1. Multiple sequence alignment shows sequences similarities among mammalian species by Geneious 7.1.2 [5].

Genomic insulin has found identical among eight mammalian species. It is also found that *Homo sapiens* (Human) and *Gorilla gorilla* (Western gorilla) are highly conserved in their sequences.

3.2. Phylogenetic Connection

Today, molecular phylogenetics helps to construct and evaluates hypothesis about historical patterns of ancestry, divergence and descent in the form of the phylogenetic tree. It shows evolutionary relationships predicted from the multiple sequence alignment.

The phylogenetic tree based on genomic insulin sequences of the eight mammalian species such as *Homo sapiens* (Human), *Bos taurus* (Cattle), *Cavia porcellus* (Guinea pig), *Canis lupus familiaris* (Dog), *Gorilla gorilla* (Western gorilla), *Ovis aries* (Sheep), *Pan troglodytes* (Chimpanzee),

and *Pongo pygmaeus* (Orangutan) form a phylogenetically related clusters or sub-groups (Figure 2). This phylogenetic tree revealed that two monophyletic of six major phylogenetic groups produced. These monophyletic groups are *Homo sapiens* (Human), *Gorilla gorilla* (Western gorilla) and *Bos taurus* (Cattle), *Ovis aries* (Sheep) and relatively recent common ancestor and therefore seen to be phylogenetically closer to each other. The result also exhibits a fundamental diversity among all the genomic insulin as *Homo sapiens* (Human) is closely related to the *Gorilla gorilla* (Western gorilla) rather than to the *Pan troglodytes* (Chimpanzee) and *Pongo pygmaeus* (Orangutan). *Bos taurus* (Cattle) is closely related to the *Ovis aries* (Sheep) rather than to the *Canis lupus familiaris* (Dog). The remaining *Cavia porcellus* (Guinea pig) as the out-group that is the most distantly related of the eight mammalian species.

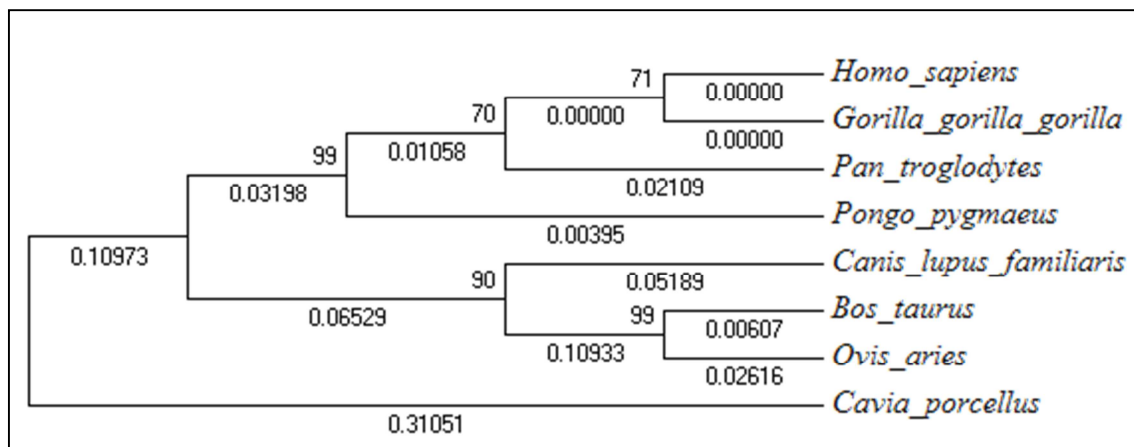


Figure 2. Molecular Phylogenetic analysis of eight mammalian species by using Maximum Likelihood method based on the JTT matrix-based model [10]. The tree with the highest log likelihood (-615.3939) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. All positions containing gaps and missing data have eliminated. There have a total of 105 positions in the final data set. Evolutionary analyses have conducted in MEGA5 [6].

3.3. Pairwise Distances

The pairwise distances analysis among eight mammalian species revealed percent identity and divergences of each sequences pair. Result from the Table 1 exhibits that genomic insulin from *Homo sapiens* (Human) and *Canis lupus familiaris* (Dog) has a lowest genetic distance of 0.13 of the

mammalian species studies. *Homo sapiens* (Human) and *Cavia porcellus* (Guinea pig) have the next lowest genetic distances of 0.39 and 69.1% identical at the amino acid levels. Whereas *Homo sapiens* (Human) and *Gorilla gorilla* (Western gorilla) have genetic distances of 0.00 and 100% identical at the amino acid levels (Table 1).

Table 1. Estimation of Evolutionary Divergence between Sequences.

Mammalian species	1	2	3	4	5	6	7	8
1. <i>Homo sapiens</i>								
2. <i>Gorilla gorilla</i>	0.00							
3. <i>Pongo pygmaeus</i>	0.01	0.01						
4. <i>Pan troglodytes</i>	0.02	0.02	0.02					
5. <i>Canis lupus familiaris</i>	0.13	0.13	0.13	0.14				
6. <i>Bos taurus</i>	0.18	0.18	0.18	0.19	0.14			
7. <i>Ovis aries</i>	0.19	0.19	0.19	0.20	0.17	0.03		
8. <i>Cavia porcellus</i>	0.39	0.39	0.38	0.41	0.43	0.48	0.46	

The numbers of amino acid substitutions per site from between sequences are shown. Standard error estimates are shown above the diagonal and were obtained by a bootstrap procedure (10000 replicates). Analyses were conducted using

the Poisson correction model [12]. The analysis involved 8 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 105 positions in the final dataset. Evolutionary analyses were

conducted in MEGA5 [6].

3.4. Physico-chemical Characteristics

In the physico-chemical characteristics (Table 2.) shows that all of the genomic insulin have found that less acidic molecules which represented as theoretical (pI) that is

calculated by positively (Asp+Glu) and negatively (Arg+Lys) charged amino acids. From the study of instability index, surprising that all of the genomic insulin are unstable with high leucine content (18.2%) (Table 3.) except *Bos taurus* (Cattle) and *Ovis aries* (Sheep) [11].

Table 2. Physico-chemical characteristics of eight mammalian genomic insulins.

Genomic insulin	Theoretical (pI)	'-' charged residues (Asp+Glu)	'+' charged residues (Arg+Lys)	Extinction Coefficient	Instability Index	Aliphatic Index	(GRAVY)	Crystallization Coefficient
<i>Homo sapiens</i>	5.22	10	7	17335	40.33	102.91	0.193	0.708
<i>Gorilla gorilla</i>	5.22	10	7	17335	40.33	102.91	0.193	0.708
<i>Pongo pygmaeus</i>	5.22	10	7	17335	40.33	102.00	0.145	0.709
<i>Pan troglodytes</i>	5.22	10	7	17335	40.33	103.73	0.191	0.709
<i>Canis lupus familiaris</i>	5.61	11	9	17335	44.73	107.36	0.207	0.728
<i>Bos taurus</i>	6.70	8	8	17335	33.91	95.81	0.187	0.547
<i>Ovis aries</i>	6.26	8	7	17335	33.10	97.62	0.226	0.576
<i>Cavia porcellus</i>	5.15	11	7	15845	57.88	93.09	-0.017	0.751

Table 3. Amino Acid composition of eight mammalian genomic insulins (in %).

Genomic insulin	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
<i>Homo sapiens</i>	9.1	4.5	2.7	1.8	5.5	6.4	7.3	10.9	1.8	1.8	18.2	1.8	1.8	2.7	5.5	4.5	2.7	1.8	3.6	5.5
<i>Gorilla gorilla</i>	9.1	4.5	2.7	1.8	5.5	6.4	7.3	10.9	1.8	1.8	18.2	1.8	1.8	2.7	5.5	4.5	2.7	1.8	3.6	5.5
<i>Pongo pygmaeus</i>	8.2	4.5	2.7	1.8	5.5	7.3	7.3	10.9	1.8	1.8	18.2	1.8	1.8	2.7	5.5	4.5	2.7	1.8	3.6	5.5
<i>Pan troglodytes</i>	7.3	4.5	2.7	1.8	5.5	6.4	7.3	10.9	1.8	1.8	18.2	1.8	1.8	2.7	5.5	5.5	2.7	1.8	3.6	6.4
<i>Canis lupus familiaris</i>	6.4	4.5	3.6	4.5	5.5	8.2	5.5	9.1	2.7	1.8	16.4	1.8	3.6	2.7	3.6	3.6	6.4	1.8	2.7	5.5
<i>Bos taurus</i>	13.3	5.7	2.9	0.0	5.7	4.8	7.6	11.4	1.9	1.0	15.2	1.9	1.0	2.9	7.6	2.9	1.9	1.9	3.8	6.7
<i>Ovis aries</i>	12.4	4.8	2.9	0.0	5.7	4.8	7.6	12.4	2.9	1.0	15.2	1.9	1.0	2.9	7.6	1.9	1.9	1.9	3.8	7.6
<i>Cavia porcellus</i>	10.9	6.4	2.7	1.8	5.5	5.5	8.2	8.2	1.8	1.8	18.2	1.8	1.8	2.7	5.5	2.7	2.7	1.8	3.6	6.4

Additionally, high aliphatic index indicates that more thermally stable. Aliphatic index of *Canis lupus familiaris* (107.36), *Pan troglodytes* (103.73), *Homo sapiens* (102.91), *Gorilla gorilla* (102.91) and *Pongo pygmaeus* (102.00) classifies them as most thermostable, closely followed by other mammalian species, *Ovis aries* (97.62), *Bos taurus* (95.81), and *Cavia porcellus* (93.03).

Furthermore, Grand average of hydropathicity index (GRAVY) indicates hydrophobic or hydrophilic character of protein. GRAVY of all genomic insulin have found hydrophobic except *Cavia porcellus* (-0.017). Extinction coefficient for all sequences are observed high rather than *Cavia porcellus* (Guinea pig). This prediction is useful for protein-protein interaction.

Crystallization coefficient values of genomic insulin are observed within the range of 0.406 to 0.751 (Table 2).

3.5. Characterization of Secondary Structure

In the analysis of secondary structure of eight mammalian genomic insulin sequences result in the predominance of random coil which is followed by alpha helix, extended strand, and beta sheet, respectively (Table 4). The high value of random coil bears crucial significance in protein tertiary structure and related functions. For *Canis lupus familiaris* (Dog), it is found that alpha helix (52.73%) exceed the random coil.

Table 4. Secondary structure prediction of eight mammalian genomic insulins (in %).

Genomic insulin's	α helix	3_{10} Helix	Pi Helix	β bridge	Extended strand	β turn	Bend region	Random coil	Ambiguous states
<i>Homo sapiens</i>	47.27	0.00	0.00	0.00	14.55	9.09	0.00	29.09	0.00
<i>Gorilla gorilla</i>	47.27	0.00	0.00	0.00	14.55	9.09	0.00	29.09	0.00
<i>Pongo pygmaeus</i>	48.19	0.00	0.00	0.00	13.64	10.91	0.00	27.27	0.00
<i>Pan troglodytes</i>	42.73	0.00	0.00	0.00	20.00	10.00	0.00	27.27	0.00
<i>Canis lupus familiaris</i>	52.73	0.00	0.00	0.00	11.82	10.00	0.00	25.45	0.00
<i>Bos taurus</i>	39.05	0.00	0.00	0.00	14.29	10.48	0.00	36.19	0.00
<i>Ovis aries</i>	39.09	0.00	0.00	0.00	15.24	9.52	0.00	36.19	0.00
<i>Cavia porcellus</i>	43.64	0.00	0.00	0.00	11.82	15.45	0.00	29.09	0.00

3.6. 3D Structure of Genomic Insulin

3D structure prediction results show that five model for each of the given sequences and ranked them according to the scores of ramachandran plot. High ranked 3D model has selected for these sequences. (Figure 3.–Figure 10.)

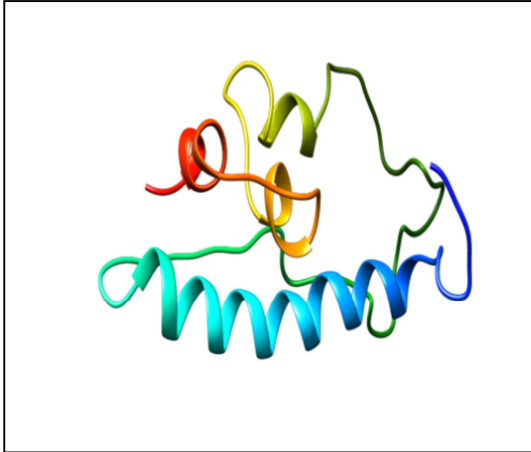


Figure 3. 3D Structure of Insulin (*Homo sapiens*).

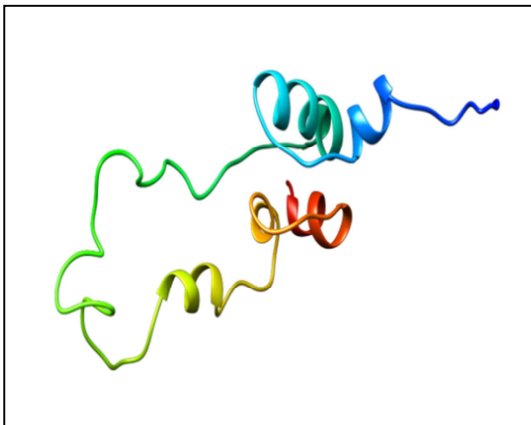


Figure 4. 3D Structure of Insulin (*Gorilla gorilla*).

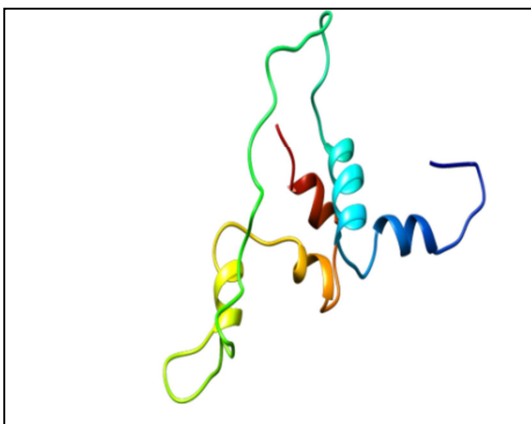


Figure 5. 3D Structure of Insulin (*Pongo pygmaeus*).

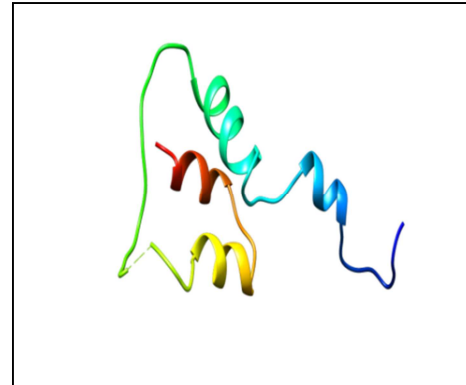


Figure 6. 3D Structure of Insulin (*Pan troglodytes*).

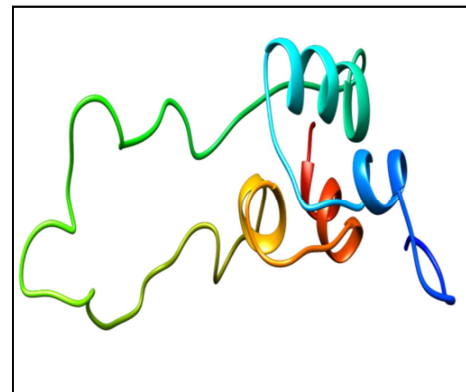


Figure 7. 3D Structure of Insulin (*Canis lupus familiaris*).

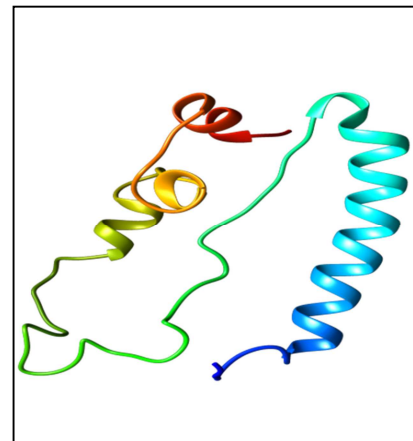


Figure 8. 3D Structure of Insulin (*Bos taurus*).

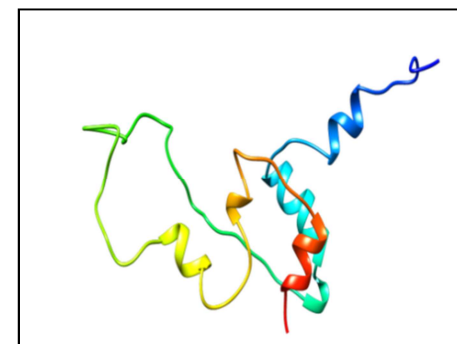


Figure 9. 3D Structure of Insulin (*Ovis aries*).

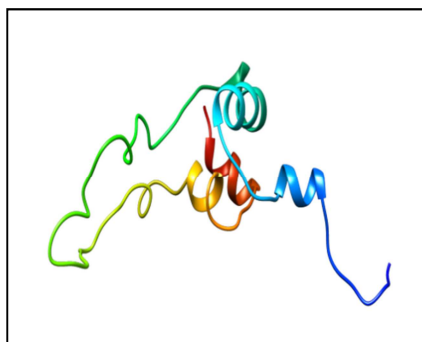


Figure 10. 3D Structure of Insulin (*Cavia porcellus*).

4. Conclusion

Molecular phylogeny is useful in explaining similarities and differences between organisms and for establishing relationships among the population which can be considered as further evidence of evolution. This study manifests that, this mammalian species that have genomic insulin which are similar in the sequence are more closely related than species with fewer similarities in sequence and allowed us to know the manner of evolution over a million-year period. The investigations of these mammalian genomic insulins provide rapid comprehensive information to calculate the amino acid sequence in the relations to evolutionary divergence rate as well as molecular phylogenetics.

Acknowledgments

The authors respect and honor to founder, Vice-Chancellor, Professor Nurul Islam, University of Science and Technology Chittagong, Bangladesh, Professor Nurul Absar, Head and all the teachers of the Department of Biochemistry and Biotechnology, University of Science and Technology Chittagong, Bangladesh for their helpful suggestions and useful comments regarding of this research.

References

- [1] Suzanne E. McGaugh, Anne M. Bronikowski, Chih-Horng Kuo et al., (2015). "Rapid molecular evolution across amniotes of the IIS/TOR network," *Proc Natl Acad Sci U S A*. 112 (22):7055-7060.
- [2] De Souza AM, Lopez JA, (2004). "Insulin or insulin-like studies on unicellular organism: a review," *Barzarch. Boil. Technol.* 47 (60).
- [3] Steiner D, N. Fox, S. Smeekens, S. Ohagi, G. Westermark, and S. Chan, (1993). "New molecular perspectives in islet hormone biosynthesis," *Biochem. Soc. Trans.* 21:139-142.
- [4] Shu Jin Chan and Donald F. Steiner, (2000). "Insulin Through the Ages: Phylogeny of Growth Promoting and Metabolic Regulatory Hormone," *Amer. Zool.* 40: 213-222.
- [5] Kearse M, Moir R, Wilson A et al., (2012). "Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data," *Bioinformatics*, 28 (12): 1647-9.
- [6] Tamura K., Peterson D, Peterson N, Stecher G, Nei M, and Kumar S, (2011). "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods," *Molecular Biology and Evolution* 28: 2731-2739.
- [7] Lukasz k, RazibAa, Aghakhani S, Dick S, Mizianty M, Jahandibeh S, (2009). "CRYSTALP2: sequences-based protein crystallization propensity prediction," *BMC structural Biology* 9:50.
- [8] Geourjon C, Deleage G, (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments," *Comput Appl Biosci*; 11[6]: 681-84.
- [9] Bates PA, Kelly LA, MacCallum RM, Sternberg MJE (2001). "Enhancement of Protein Modelling by Human Intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM." *Proteins: Structure, Function and Genetics*, Suppl5: 39-46.
- [10] Jones D.T, Taylor W.R, and Thornton J.M, (1992). "The rapid generation of mutation data matrices from protein sequences," *Computer Applications in the Biosciences* 8: 275-282.
- [11] http://web.expasy.org/docs/expasy_tools05.pdf.
- [12] Zuckerkandl E. and Pauling L, (1965). "Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.