IJBPAS, March, 2021, 10(3): 948-960

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

DOCKING STUDIES AND PROTEIN – PROTEIN INTERACTION ANALYSIS TO UNVEIL THE PROTEINS RESPONSIBLE FOR HYPERTHYROIDISM IN HUMAN

JABALIA N AND BANSAL H*

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Sector 125, Noida-201303, U.P., India

***Corresponding Author: Hina Bansal: E Mail: hinabansal@gmail.com; Phone (M): +91-9811099082; ORCID: 0000-0003-1683-1581**

Received 17th May 2020; Revised 19th June 2020; Accepted 4th July 2020; Available online 1st March 2021

https://doi.org/10.31032/IJBPAS/2021/10.3.5405

ABSTRACT

Hyperthyroidism is a common endocrine disorder, the complications of untreated disease include hypertension, osteoporosis and cardiovascular system. There are many anti - thyroid drugs but the main shortcoming is that the hyperthyroidism often comes back after they are discontinued. That's why it still needs attention for more effective treatments. The aim of the present work was to present an outline of the structural and functional aspects of protein responsible for hyperthyroidism from *Homo sapiens*. First, the genes responsible for hyperthyroidism were identified and investigated through different computational tools and servers to exploit its functional analysis, phylogenetic assessment, 3D structural modelling and quality analysis of the generated 3D model. The 3D structure of all the selected proteins those structures were not present in PDB database, were generated. Further, protein-protein interaction analysis was also performed to observe the functional involvement of these proteins in metabolic pathway. The docking study was performed against all proteins which were having maximum involvement in metabolic pathways. Thus, the present work may facilitate *in-vitro* and *in-vivo* studies to design novel and potential inhibitors against hyperthyroidism.

Keywords: Endocrine disorder, metabolic pathway, phylogenetic analysis, 3D structural modelling

1. INTRODUCTION

Hyperthyroidism in human is a state in which thyroid gland produces thyroid hormones in excess due to its over activity and circulates them in blood. The thyroid gland is a butterfly-shaped gland places in the anterior portion of the neck just below the adams apple **[1-3]**. The thyroid gland wraps around the trachea. It consists of two lobes connected by an isthmus **[4]**. The thyroid gland secretes thyroid hormones, thyroxine (T4), triiodothyronine (T3) and calcitonin, which help to regulate the human body metabolism and effects many processes like growth **[2]**. The thyroid hormones that majorly involved in metabolic processes are T3 and T4 **[3]**. Usually the common cause of hyperthyroidism is Graves' disease with female preponderance. Problems linked with hyperthyroidism comprise hypertension, osteoporosis, atrial fibrillation, effect on heart and cardiovascular system **[5, 6]**. It is observed that it affects more females in comparison to males **[7]**. Thyroid disease follows the autoimmune mechanism which is the main cause of this gender-specificity also **[8]**. There are three treatment options – anti thyroid medications, radioablation and surgery. Each of these modalities has its own merits and demerits. Heavy dose of antithyroid drugs give the immediate treatment to the patient suffering from thyroid diseases **[3]**. Ever since the introduction of anti-thyroid drugs, till date these have been the first line treatment options. The main shortcoming of these drugs is that the underlying hyperthyroidism often comes back after they are discontinued. That's why it still needs attention for more effective treatments.

In the current study, we performed a computational and protein – protein interaction network analysis approach for unveiling the proteins involved in hyperthyroidism. In our study, the first step was to identify the proteins responsible for hyperthyroidism. Next, structural analysis was performed to reveal the 3D structure of all the selected proteins. Further, phylogenetic analysis was performed for all the genes related to hyperthyroidism. Significant protein-coding genes responsible for hyperthyroidism were found out through microarray data analysis. STRING database was used to identify protein interaction network of the selected genes. KEGG Pathway studies was performed to predict the associations of the protein sequences. The docking of selected protein related to hyperthyroidism and involved in metabolic pathways was performed against anti thyroid

drugs selected from literature **[9]**. Thus, this study will provide an insight into the various dimensions of hyperthyroidism which will help the readers to effectively use the information for their research endeavors.

2. MATERIAL AND METHODS

A schematic representation of the methodology is showing in **Figure 1**. All the genes related to hyperthyroidism were identified from NCBI **[10]** and UniProt database **[11]** and their protein sequences, Uniprot ID and the accession number was also retrieved and recorded. Chromosomal location of these genes was observed from the Ensembl database **[12]**. 3D structure of all the selected genes related to hyperthyroidism were searched in Protein Data Bank **[13]**. Proteins those structures were not presented in PDB were modelled by swiss model workspace **[14]** and evaluated by Ramachandran Plot Analysis (Rampage) **[15]**. The phylogenetic tree of all the selected genes was constructed by MEGA X **[16]** based on neighbor-joining method. Next, Protein interaction network analysis of the selected genes responsible for hyperthyroidism and physically interact with each other to accomplish cellular functions, such as metabolism, cell cycle control, and signal transduction was performed by using STRING database **[17, 18]**. The analysis has been performed for identifying key proteins having utmost metabolic functional associations. KEGG pathway analysis database **[19, 20]** was used to predict the pathway associations of the protein sequences. Microarray analysis of Hyperthyroidism dataset was performed by Gene Expression Omnibus (GEO) **[21]**. The dataset obtained from GEO for hyperthyroidism was GSE71956 (dataset for Grave's disease).R programming has been used to analyze and graphically visualization of the datasets. It is publicly available software environment which runs on operating systems like windows, UNIX and Mac **[22]**. The docking of selected proteins related to hyperthyroidism and involved in metabolic pathways were done by Swissdock **[23]**. The drugs available for the treatment of hyperthyroidism were identified from drugbank **[24]**. The chemical structures of receptor and ligand were automatically retrieved by Swissdock from Zinc database **[25]**. The docking results were obtained in terms of energy Kcal/mol.

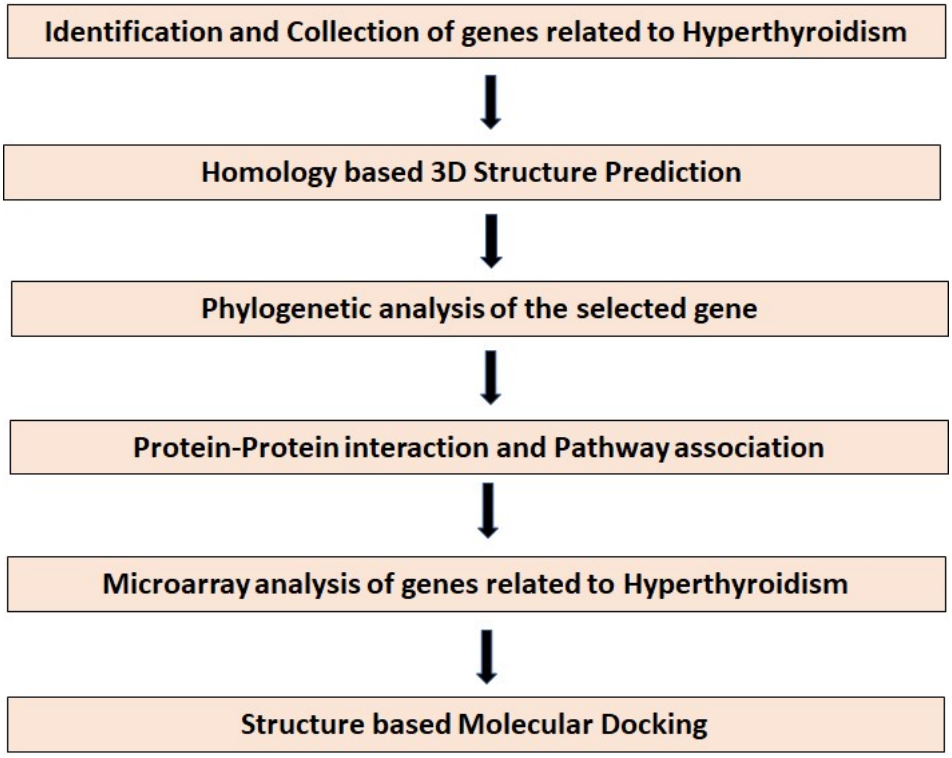


Figure 1: A schematic representation of the methodology

3. RESULTS AND DISCUSSIONS

Identification and collection of genes related to hyperthyroidism

In total 20 genes were selected from protein sequence databases. Their accession number and protein sequences in fasta format **(Supplementary file 1)** were retrieved from UniProt and NCBI database. The chromosomal location of the selected genes was retrieved from Ensembl database **(Table 1)**.

Tertiary structure prediction

All selected genes were searched for their existing 3-dimensional structure in PDB database and recorded with their PDB ID in **Table 1**. The genes, those structures were not reported in PDB database were modelled using swiss model workspace. The 3D structure of total 9 genes out of 20 selected genes, were modelled and represented in **Table 2**. The generated structures were evaluated by Rampage **(Table 3)**. It has been observed that 3-D model structure of protein sequence B7U540 has shown best sequence identity w.r.t its template and favored region was 97.6% after validation.

Phylogenetic analysis

Phylogenetic analysis was performed for all the selected proteins referring genes related to hyperthyroidism. A phylogenetic tree was

constructed for all the selected 20 proteins through neighbor-joining method using MEGA X. The resulting tree represents the ancestral relationship among all the genes and infer their evolutionary history **(Figure 2)**. The tree divided the sequences into two major clusters, these clusters were again divided into sub-clusters. The analysis revealed that all the genes responsible for hyperthyroidism had originated from a common ancestor and during evolution diverged further into subgroups.

Protein-protein interaction and pathway association

Functional associations between selected genes responsible for hyperthyroidism have been predicted by STRING database **(Figure 3)**. The network forms interactions between nodes representing genes. Out of all the selected proteins 6 proteins named THRB, TSHR, TG, FCRL3, CTLA4 and SLCO1C1 are showing interactions. The color difference in the interaction lines of the string network represent different indications. The yellow line indicates text-mining, the red line represents the occurrence of fusion and light blue line indicates database. Proteins AZGP1, LMOD1, GPR174 and LHPP do not have any interaction with the other proteins. Proteins KCNJ18 and CACNA1S also have interactions with each other.

KEGG Automatic Annotation Server (KAAS) was used to identify the pathway association for all the 20 selected genes responsible for hyperthyroidism. The results revealed that out of 20 selected sequences 7 protein sequences associated with 88 KEGG pathways. The complete list of pathways is given in **Supplementary file 2**. Proteins involved in metabolic pathways were further used for docking studies.

Microarray analysis of genes

Microarray analysis of hyperthyroidism dataset was performed by GEO. The dataset obtained from GEO for hyperthyroidism was GSE71956 which include both the test and the control samples. There were total 49 samples out of which 31 samples were from patients having Grave's disease and rest were from healthy human sample. R programming has been used to analyze and graphically visualization of the datasets. **Figure 4** shows the BoxPlot analysis of gene dataset GSE71956. The statistical analysis of hyperthyroidism dataset retrieved from GEO provides an in-sight into genes responsible for grave's disease.

Docking Analysis

Proteins related to hyperthyroidism and involved in metabolic pathways were used for docking studies (Supplementary file 2). Hence, docking of six protein sequences

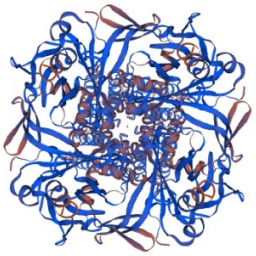
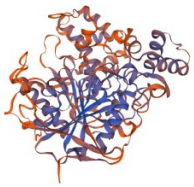
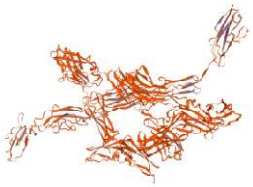
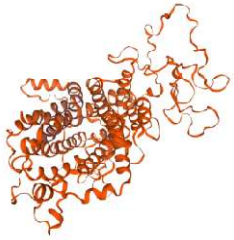
(P16473, Q13698, P63092, Q9H008.2, P16410.3, P10828.2) involved in metabolic pathways of human against the drugs carbimazole, methimazole and propylthiouracil **(Table 4)**. As the structure of protein P01266 was not available in protein databank and the modeled structure through Swiss model workspace was also not showing good level of confidence, so did not consider for docking. The docking results showed minimum binding energy -7.70 Kcal/mol between the protein THRB against the drug propylthiouracil **(Figure 5)**.

Table 1: List of genes related to hyperthyroidism with UNIPROT/NCBI-ID, PDB-ID and chromosomal locations

S. No.	Uniprot ID/ Accession No.	Gene name	Protein name	Chromosomal location	Organism	PDB ID
1	P16473	TSHR	Thyrotropin receptor	Chromosome 14:80954989-81146302	*Homo sapiens*	2XWT
2	B7U540	KCNJ18	Inward rectifier potassium channel 18	Chromosome 17:21692523-21704612	*Homo sapiens*	-
3	Q13698	CACNA1S	Voltage-dependent L-type calcium channel subunit alpha-1S	Chromosome 1:201039512-201112566	*Homo sapiens*	2VAY
4	P01266	TG	Thyroglobulin	Chromosome 8: 132866958-133134903	*Homo sapiens*	-
5	P16410	CTLA4	Cytotoxic T-lymphocyte protein 4	Chromosome 2: 203867786-203873960	*Homo sapiens*	1AH1
6	Q96P31	FCRL3	Fc receptor-like protein 3	Chromosome 1: 157674321-157700857	*Homo sapiens*	-
7	P63092	GNAS	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	Chromosome 20: 588839712-58911192	*Homo sapiens*	5G53
8	Q9BXC1	GPR174	Probable G-protein coupled receptor 174	Chromosome X: 7914466379175315	*Homo sapiens*	-
9	P25311	AZGP1	Zinc-alpha-2-glycoprotein	Chromosome 7: 99966720-99976042	*Homo sapiens*	3ES6
10	Q9H008	LHPP	Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	Chromosome 10: 124461834124617888	*Homo sapiens*	2XYD
11	P16410.3	CTLA4	Cytotoxic T-lymphocyte protein 4	Chromosome 2: 203,867,77203,873965	*Homo sapiens*	1H6E
12	P10828.2	THRB	Thyroid hormone receptor beta	Chromosome 3: 24,117,153-24,495,756	*Homo sapiens*	4ZO1
13	NP_001018046	TSHR isoform 2	Thyrotropin receptor isoform 2	Chromosome 14: 80,954,98981,146,302	*Homo sapiens*	2XWT
14	Q9NYB5.1	SLCO1C1	Solute carrier organic anion transporter family member 1C1	Chromosome 12: 20,695,35520,753,386	*Homo sapiens*	-
15	NP_001136098.1	TSHR isoform 3	Thyrotropin receptor isoform 3 precursor	Chromosome 14: 80,954,98981,146,302	*Homo sapiens*	2XWT
16	NP_001176	AZGP1	Zinc-alpha-2-glycoprotein precursor	Chromosome 7: 99,966,720-99,976,042	*Homo sapiens*	6R2U
17	NP_001139416.1	SLCO1C1 Isoform 4	Solute carrier organic anion transporter family member 1C1 isoform 4	Chromosome 12: 20,695,355-20,753,386	*Homo sapiens*	-

18	**NP_001139417**	**SLCO1C1 Isoform 3**	**Solute carrier organic anion transporter family member 1C1 isoform 3**	**Chromosome 12: 20,695,355-20,753,386**	***Homo sapiens***	**-**
19	**NP_001139418**	**SLCO1C1 Isoform 1**	**Solute carrier organic anion transporter family member 1C1 isoform 1**	**Chromosome 12: 20,695,355-20,753,386**	***Homo sapiens***	**-**
20	**NP_036266.2**	**LMOD1**	**Leiomodin-1**	**Chromosome 1: 201,896,456-201,946,588**	***Homo sapiens***	**-**

Table 2: Tertiary structure prediction of proteins responsible for hyperthyroidism

Accession no.	Template	Sequence Identity (%)	Sequence Coverage	Structure
B7U540	**3SPH.1.A**	**88.92**	**42-372**	
P01266	**5YDJ.1.A**	**32.69**	**2197-2731**	
Q96P31	**6IAA.1.A**	**22.04**	**107-564**	
Q9BXC1	**5UNG.1.A**	**28.52**	**16-307**	
Q9NYB5.1	**6E9C.1.A**	**16.63**	**44-649**	

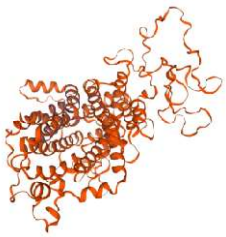
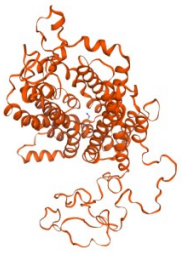
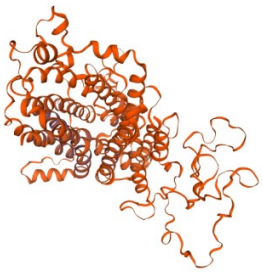
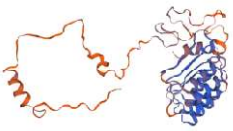
NP_001139416.1	6E9C.1.A	16.82	66-530	
NP_001139417	6E9C.1.A	15.07	44-599	
NP_001139418	6E9C.1.A	16.19	44-648	
NP_036266.2	4RW1.1.b	50.23	317-600	

Table 3: Ramachandran plot analysis of hyperthyroidism protein sequences

Accession No.	Rampage		
	Favoured region (%)	Allowed region (%)	Outlier region (%)
B7U540	97.6	1.8	0.6
P01266	91.0	6.6	2.4
Q96P31	90.1	6.8	3.1
Q9BXC1	95.9	2.4	1.7
Q9NYB5.1	84.1	10.3	5.6
NP_001139416.1	83.6	11.7	4.8
NP_001139417	86.8	7.9	5.2
NP_001139418	86.7	9.8	3.5
NP_036266.2	86.9	9.6	3.5

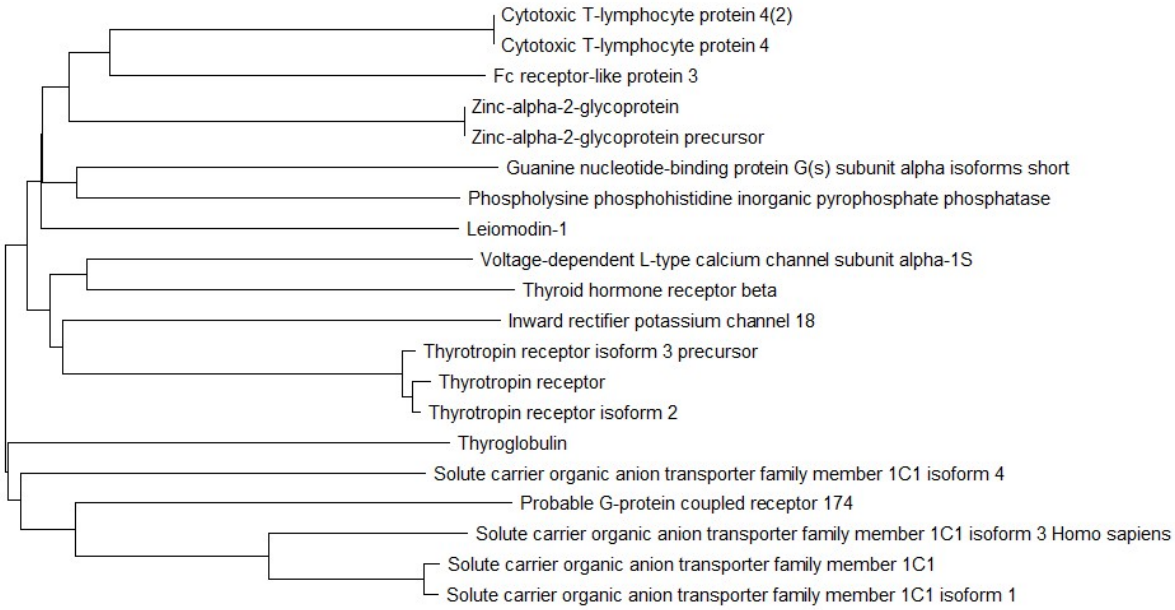


Figure 2: Phylogenetic analysis of selected genes responsible for hyperthyroidism

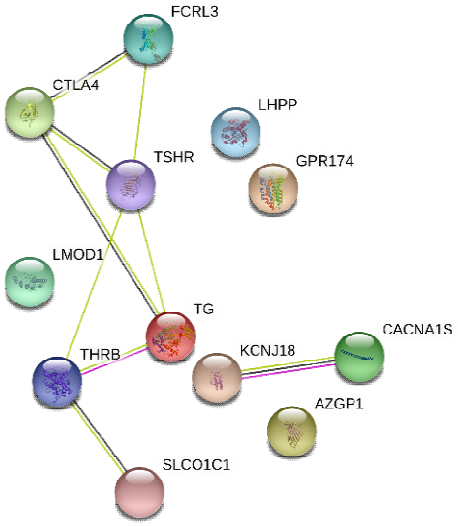


Figure 3: String analysis of selected genes responsible for hyperthyroidism

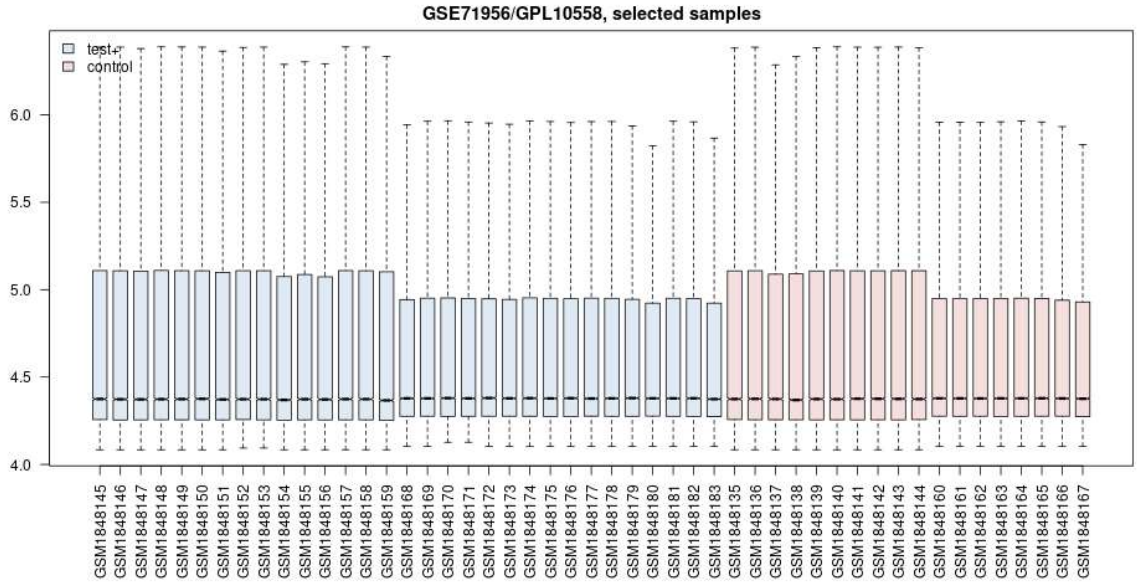


Figure 4: Boxplot of dataset GSE71956
: Control; : Test

Table 4: Docking results of Swissdock in terms of full fitness/ΔG Kcal/mol

Protein sequence	Template	Ligand		
		Carbimazole	Methimazole	Propylthiouracil
P16473	2XWT	-934.69/-6.19	-919.54/-5.86	-981.69/-6.54
Q13698	2VAY	-1251.40/-6.31	-1237.48/-6.05	-1297.46/-6.49
P63092	5G53	1042.88/-7.01	-1023.20/-6.14	-1085.34/-6.81
Q9H008.2	2XYD	-1787.00/-6.29	-1773.05/-6.20	-1832.40/-6.42
P16410.3	1H6E	-1787.00/-6.29	-1773.05/-6.20	-1832.40/-6.42
P10828.2	4ZO1	-1505.60/-7.22	-1487.02/-6.70	-1552.56/-7.70

Figure 5: Docking of hyperthyroidism protein (P10828) with propylthiouracil

4. CONCLUSION

Although various conventional and experimental therapies have been directed for the treatment of hypothyroidism, but it often comes back after they are discontinued. Hence, an *in-silico* approach has been applied to give an insight towards the hormonal disease i.e. hyperthyroidism. In the current research sequential, structural and interactional aspects were included which will facilitate the understanding of genes responsible for hyperthyroidism. The phylogenetic analysis revealed that all genes related to the disease were originated from the same ancestor. Further, docking results showed the closest association between the protein THRB and the drug propylthiouracil. In conclusion, the present approach expedites the computational study of endocrine disorder which may be beneficial for academic and industrial proposes after experimental and validation studies.

Acknowledgements

The authors are highly grateful to founder President Dr. Ashok K Chauhan and Chancellor Mr. Atul Chauhan Amity University Uttar Pradesh, Noida, India for providing necessary support and facilities.

REFERENCES

[1] Little, J.W., Falace, D., Miller, C. and Rhodus, N.L., 2017. Dental Management of the Medically Compromised Patient-E-Book. Elsevier Health Sciences.

[2] Larsen, P.R., Davies, T.F., Schlumberger, M.J. 2003. Thyroid physiology and diagnostic evaluation of patients with thyroid disorders. In: Larsen, P.R., Kronenberg, H.M., Melmed, D., Polonsky, K.S., editors. Williams textbook of endocrinology. 10th ed. Philadelphia: W. B. Saunders; 331-65.

[3] Jameson, J.L., Weetman, A.P. 2005. Disorders of the thyroid gland: anatomy and development. In: Kasper, D.L., Braunwald, E., Fauci, A.S., Hauser, S.L., Longo, D.L., Jameson, J.L., et al., (Eds), Harrison's principles of internal medicine. McGraw- Hill., New York pp. 2105

[4] Mazzaferri, E.L. 1986. The thyroid. In: Mazzaferri EL, editor. Endocrinology. 3rd ed. Medical Examination Publishing, New York

[5] O'Hanlon KM, Baustian GH, Toth DW. Hyperthyroidism. Elsevier; 2004. Available at http://www.firstconsult.com/hyperthyroidism. Accessed October 30, 2004.

[6] Danzi, S., Klein, I. 2004. Thyroid hormone and the cardiovascular

system. *Minerva* Endocrinol.. 29,139 –150.

[7] Canaris, G.J., Manowitz, N.R., Mayor, G., Ridgway, E.C. 2000. The Colorado thyroid disease prevalence study. Arch Intern Med. 160,526 – 530.

[8] Carmona Carmona, C., Bedoya, P., Acevedo, J. and Cardona Arias, J., 2018. Prevalence of Thyroid Disorders in an Institution Providing Health Services in Medellin-Colombia. Transl Biomed, 9, 2. 149.

[9] Gupta, A., Wamankar, S., Gidwani, B., Kaurm C.D. 2016. Herbal drugs for thyroid treatment. Int. J. Pharm. Biol. Sci. 6, 62-70.

[10] The NCBI handbook [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2002 Oct. Chapter 18, The Reference Sequence.

[11] The UniProt Consortium, UniProt: a hub for protein information. Nucleic Acids Res. 43 (2015) D204–D212.

[12] Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G.R., Ruffier, M., Taylor, K., Vullo, A. and Flicek, P. 2015. The Ensembl REST API: Ensembl data for any language. Bioinformatics, 31,143-145.

[13] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.2000. The Protein Data Bank. Nucleic Acids Res. 28,235-242.

[14] Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46,296-303.

[15] Lovell, S.C., Davis, I.W., Arendall, W.B., De Bakker, P.I., Word, J.M. and Prisant, M.G., 2003. & Richardson, DC Structure validation by Cá geometry: Õ, ø and Câ deviation. Proteins: Struct., Funct., Bioinf. 50,437-450.

[16] Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol., 35,1547-1549

[17] Pavlopoulou, A., Karaca, E., Balestrazzi, A. and Georgakilas, A.G., 2019. In Silico Phylogenetic and Structural Analyses of Plant Endogenous Danger Signaling Molecules upon Stress. Oxid med cell longev,

[18] Szklarczyk, D, Franceschini, A, Wyder, S, Forslund, K, Heller, D. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res, 43,447-452.

[19] Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., Kanehisa, M. 2007. KAAS: an automatic genome annotation and pathway recons-truction server. Nucleic acids research, 35,182-185.

[20] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res., 44,457-462.

[21] Barrett, T. and Edgar, R., 2006. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. Methods Enzymol. 411, 352-369.

[22] Kahsar-Miller, M., Boots, L.R., Bartolucci, A., Azziz, R. 2004. Role of a CYP17 polymorphism in the regulation of circulating dehydroepiandrosterone sulfate levels in women with polycystic ovary syndrome. Fertil Steril. 82,973–975.

[23] Grosdidier, A., Zoete, V., and Michielin, O. 2011. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res., 39, 270-277.

[24] Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., Hassanali, M., 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res, 36, D901-D906.

[25] Irwin, J.J. and Shoichet, B.K., 2005. ZINC− a free database of commercially available compounds for virtual screening. J. Chem. Inf. Model., 45, 177-182.