



Using Bioinformatics and NCBI Tools to Sequence and Structure Analysis of Transcription Factor 7 Like 2 Gene (TCF7L2) in Iraqi Diabetic Mellitus Type II Patients

Batool K. Queen*, Abbas A. Mohammed

Department of Applied Sciences, University of Technology – Iraq

Article information

Article history:

Received: August, 26, 2021

Accepted: October, 17, 2021

Available online: March, 05, 2022

Keywords:

TCF7L2 gene,

Diabetic Mellitus type II,

Bioinformatics

*Corresponding Author:

Batool K. Queen

batoolkareem97@gmail.com

Abstract

In this paper, the study aimed to analyze the sequence and structure of Transcription Factor 7 Like 2 gene (TCF7L2) in Iraqi Diabetic Mellitus Type II (T2DM) Patients and was compared with standard sequence form National Center of Bioinformatics Tools (NCBI) using BLAST. Ten blood samples of (T2DM) Iraqi patients, was collected from Al-Mustansiriya University National Diabetes Centre in Baghdad Province/Iraq (17-65 year). DNA was extracted from whole blood of patients' samples using the Quick-DNA™ Blood MiniPrep kit, then it was sent to Korea at Macrogen Corporation Company, where they used automated DNA sequencing for sequence analysis. The results of the sequence analysis of (T2DM) patients' samples, we found: six missense mutations, one deletion mutation, and three silent mutations detected using BLAST in (NCBI). All mutations appeared at the same sites of the gene which controls the rate of genetic information transcription that indicates to have a relationship with (T2DM). These mutations were recorded on the (NCBI). The physicochemical properties of (TCF7L2) determined in the present study included; alpha-helical structure and 3-Dimension structure appeared contrast when compared with the gene template. Briefly, mutations effected (TCF7L2) which influences the structure, physicochemical properties of the protein, and the secretion of insulin hormone which maintains glucose level in blood.

DOI: [10.53293/jasn.2021.4198.1079](https://doi.org/10.53293/jasn.2021.4198.1079), Department of Applied Sciences, University of Technology

This is an open access article under the CC BY 4.0 License.

1. Introduction

The Transcription Factor 7 Like 2 gene (TCF7L2) polymorphisms and development of Diabetes Mellitus Type II (T2DM) remains a big area of study especially in Iraqi diabetic patients. Bioinformatics Tools will provide a standardized method to compare the analysis, structure, and sequence of (TCF7L2) in Iraqi patients with (TD2M), with a standard sequence from the National Centre of Biotechnology Information (NCBI). Bioinformatics is defined as understanding and organizing the information associated with these molecules on a large scale by conceptualizing biology in terms of molecules (in the sense of Physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science, and statistics). In a nutshell, bioinformatics is a molecular biology information management system with several applications [1].

The amount of data collected in the form of databases and computer processing capability has been increasing at an exponential rate, which explains why bioinformatics has progressed quickly in the scientific area. The data of COVID19 which is new and dangers to collected, now are available in bioinformatics tools and can be collected for COVID19 monitoring through an intelligent network of medical biosensors, thereby improving healthcare quality [2]. Bioinformatics consists of statistics control, analysis, and modelling over all domain names [3]. Type II diabetes (T2DM) is a long-term condition in which the body's ability to produce or respond to the hormone insulin is impaired. Or when the body is unable to use the insulin it produces, resulting in high blood glucose levels (hyperglycemia) [4], Resistance of Insulin and relative (rather than absolute) insulin shortage characterizes diabetes II that affects (90 – 95%) of diabetics' patients [5]. Several factors are believed to have a contact responsible for this kind of diabetes like lifestyle, stress, and environmental conditions, in addition, that in most cases of type II diabetes, several genes played a little role in the overall disease. TCF7L2 is one of the well-studied genes for diabetes susceptibility in most populations, as of 2011, more than 36 genes have been discovered that contribute to the risk of type 2 diabetes [6, 7]. TCF7L2 gene, which is found on the chromosome (10q25.2-q25.3), there are 19 exons in this gene, 5 exons are optional. TCF7L2 is considered a member of the high mobility group box (HMGB) family with 619 amino acids and have a molecular mass of 67919 Da. Have plays a vital function in the Wnt signalling pathway [8]. TCF7L2 works as a transcription factor that influences the transcription function of numerous genes, allowing it to perform a wide range of actions within the cell[9]. However, TCF7L2 does not directly regulate glucose metabolism in β -cells, but regulates glucose metabolism in pancreatic and liver tissues[10]. Also reported to activate Wnt target genes, which consequently increases the reproduction of β -cells in the pancreas, TCF7L2 variations have since become essential genetic materials in determining the dangers of type II of diabetic mellitus all around the world by increasing susceptibility to type II diabetes through decreasing the production of glucagon-like peptide-1 (GLP-1)[9]. The TCF7L2 allele is the riskiest of the frequent genetic variations, increasing the risk of diabetes by 1.5 times [11]. Our study aimed to analyze the sequence and structure of the TCF7L2 gene in Iraqi T2DM patients and how the mutations or variation effects the gene and increasing the potential risk of diabetes. However, it's important to mention that, this study is the first in Iraq.

2. Material and Methods

2.1. Blood Sample Collection

Ten Blood samples were collected from Iraqi diabetic mellitus type II patients from Al-Mustansiriya University National Diabetes Centre in Baghdad province/Iraq (aged between 17 - 65 years). For each patient drawn five millilitres of whole blood and this process was done under aseptic conditions. Whole blood was collected in a sterile EDTA tube for a genetic test [12]. The blood samples were subjected to freezing at -20C.

2.2. DNA Extraction from Blood Samples

The extraction of DNA has been used to extract the human DNA from the whole blood of patients by using a protocol in Quick-DNA™ Blood MiniPrep (Catalog Nos. D3024 & D3025) was optimized to extract the human DNA from fresh or frozen blood.

2.3. Estimation of DNA Concentration and Purity

The DNA extracted from whole blood of diabetic mellitus type II patient's concentration and purity was estimated by using The Quantus™ Fluorometer Promega/USA. The purity of DNA was determined at wavelength 260nm-280nm, which represents the ratio of the absorbance and the purity (1.6 - 1.8) means DNA high purity [13].

2.4. Agarose Gel Electrophoresis

After DNA extraction, the presence and integrity of extracted DNA were confirmed by using agarose gel electrophoresis [14]. The samples were carefully loaded into the individual wells of the gel. Then electrical power was turned for 1-2 h of 7 v\c2. After that, the DNA migrates from the cathode pole (-) to the anode pole (+) (shown in Figure 1). The red safe nucleic acid stained bands in the gel were visualized using UV light with a wavelength of 336 nm.

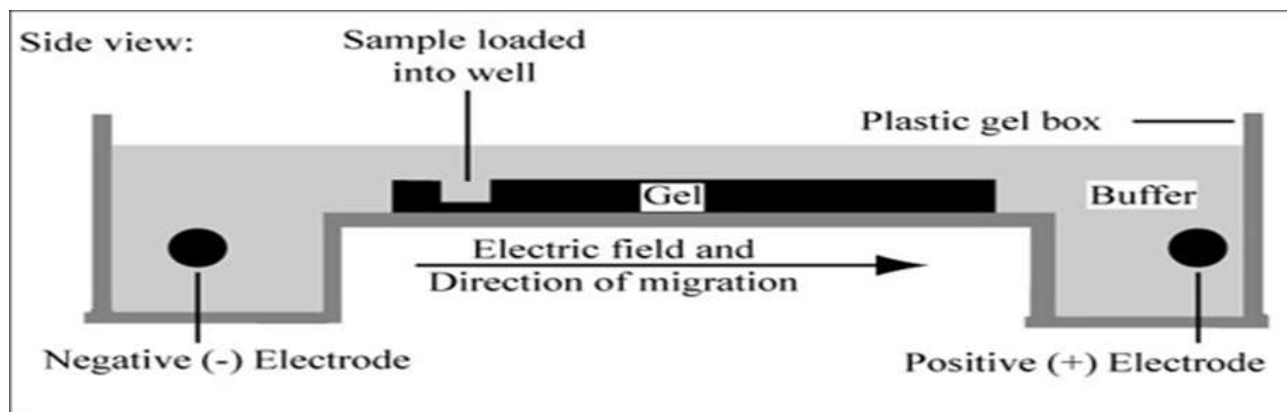


Figure 1: The agarose gel electrophoresis system. (Figure by MIT Open Courseware.).

2.5. Primer for TCF7L2 Gene

The full (TCF7L2) gene sequences were taken from the Genomic Database of the (NCBI), the reference sequence ID: NG_012631.1. Integrated DNA Technologies investigated dissolved primers (Table 1) in free DdH₂O after they were lyophilized; to get it as a stock solution in an optimal concentration of 100 pmol/l with kept it at -20, to prepare a primer suspended of 10 pmol/l concentration as work. To reach a final volume of 100: 1, 90: 1 of free DdH₂O water with 10: 1 of the stock solution (IDT/ company, Canada).

Table 1: The TCF7L2 gene-specific primers sequence.

Primer	Sequence	T _m (°C)	GC	Product size
Forward	5'- TCATAGGGGTCTGGCTTGGA- 3'	57.8	55	888bp
Revers	5'-TGATGAACGGGGACTTGCTC- 3'	57.3	55	

2.5. Polymerase Chain Reaction

The reaction mix was made by mixing the primer (working solution) and template DNA with Taq PCR Master Mix, then vortexing to avoid salt concentration discrepancies. Then prepared a mixed reaction as in Table 2. In thermal cycler placed PCR tubes and run with 5 v/cm².on (2%) agarose gel than started the programed system of cycles in Table 3.

Table 2: The Maxime PCR kit (i-Taq) Components with their Concentration.

Components	Concentration
Taq PCR Pre Mix	5µl
Forward primer	5 picomols/µl (1µl)
Reverse primer	5 picomols/µl(1µl)
DNA	1.5 µl
Distill water	16.5 µl
Final volume	25µl

Table 3: The thermal cycle program-specific interaction phases.

No.	Phase	Tm (°C)	Time	No. of cycle
1-	Initial Denaturation	95°C	3 min.	1 cycle
2-	Denaturation -2	95°C	45sec	3 cycle
3-	Annealing	64°C	45sec	
4-	Extension-1	72°C	45sec	
5-	Extension -2	72°C	7min.	1 cycle

By Agarose gel electrophoresis, PCR genetic (DNA) products, and ladder marker resolved, five μ l of PCR amplified products were poured into separate wells with attention in the 1.5% agarose gel (1.5% agarose/100ml 1X TBE buffer) and set to run at 75V for 60 minutes. Red safe Nucleic acid staining solution used to stain the gel with the DNA bands can be seen obviously [15].

2.6. Sequencing of DNA

The outcome of PCR product for (T2DM) Iraqi patients who gave clear band after PCR analysis for TCF7L2 gene, sent to Korea/ Macrogen Company, for sequencing.

2.7. Compared Between TCF7L2 DNA Sequence Gene of T2DM Samples and TCF7L2 Sequence Gene Retrieved from NCBI

For detection of the location of the mutation in TCF7L2 gene in patients samples, using BLAST online server (<https://blast.ncbi.nlm.nih.gov/Blast.cgi/>) [16].

2.8. DNA Sequence of TCF7L2 Gene Translation to Amino Acid Sequence

For translation of nucleotide sequence of (TCF7L2) gene to amino acid sequence, we are using (BLASTX) Basic Local Alignments Search Tool BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi/>) [17].

2.9. Physicochemical Properties Prediction

The PortParam online server tool (<https://web.expasy.org/protparam/>) used to investigate the physicochemical properties of TCF7L2 protein that effect amino acids sequence and the way that interact across a sequence in which the feature of the protein like: theoretical PI, amino acid compositions, molecular weight, instability index, and the other parameters can by calculate using this tool [18].

2.10. Protein Structure Prediction.

Using the PSIPred online program that gave a result with the highly accurate secondary structure of the protein that includes alpha helix, beta-strands and coils will obtain from the sequences of amino acid in the home page of the web address (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html/>)[19]. PHYRE2 online tool for tertiary structure (<https://string-db.org/cgi/>) [20].

3. Results and Discussion

The TCF7L2 DNA was extracted from the whole blood samples following the standard procedure that was used in many genetic kinds of research and studies according to this cite [21]. The quantification of DNA measured by Quantus™ Fluorometer was (1-50) ng/ μ l and the purity ranged was (1.6-1.8). The result demonstrated that the purity of extracted DNA in the total samples was sufficiently high for PCR analysis as shown in Figure (2) in light fluorescents color.



Figure 2: Under UV light and after staining with Red Safe nucleic acid stain, human genomic DNA extraction from whole blood were shown the bands on 1.5% agarose gel at 30 min in gel electrophoresis.

3.1 PCR

After the extraction of pure DNA from the all blood samples, PCR technique used to amplification the segments of extracted DNA obtained from diabetic patients. By using two primers that utilised to amplify the (TCF7L2) gene exon as specific primers, which were designed by (NCBI) software and gave identity I.d. (100%) with Homo sapiens (TCF7L2) gene by BLAST and obtained by IDT (Integrated DNA Technologies Company, Canada). As shows in Figure (3), the result of amplification segments of patient's DNA by primers that gave pure PCR bands with size (888 bp) after electrophoresed at 5 v/cm².on (2%) agarose gel.

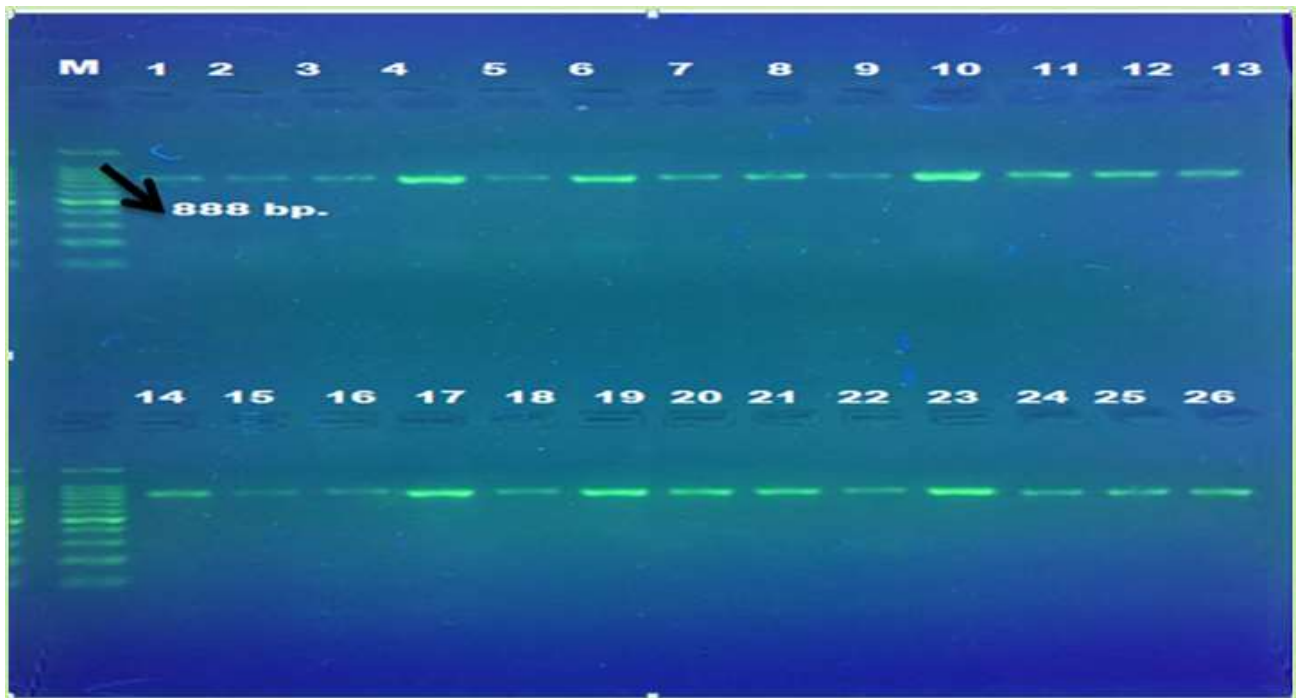


Figure 3: PCR product the band (888 bp) for (TCF7L2) gene. The product was electrophoresis in 1x TBE buffer for 1:30 hours on 2% agarose at 5 v/cm².

3.2. Sequence Analysis of Samples

Samples sequencing results that obtained from Korea/Macrogen Corporation Company using an automated DNA sequencing after compared with the sequence of (TCF7L2) gene. Showed six missense mutations, one deletion mutation and three silent mutations in Table 4with details. These are available in the (NCBI) Reference Sequence by using (Bioeditpro.version:7.0.0).

Table 4: The simple results of sequencing and type of mutation in DNA level and RNA levels with the effect that occurs after translation.

NO.	Wild genetic code of DNA	Mutation in genetic code of DNA	Type of substitution	Wild genetic code of RNA	Mutations in genetic code of RNA	Effect of translation	Location
1	GTC	A T C	Transition	GUC Valine	A U C Lucine	Missenses mutation	104323
2	GGC	A G C	Transition	GGC Glycine	A G C Serine	=	104533
3	TGT	A G T	Transversion	UGU Cysteine	AGU Serine	=	104074
4	GAT	G G T	Transition	GAU Aspartic acid	G G U Glycine	=	104621
5	CAA	C A T	Transversion	CAA Glutamine	CAU Histidine	=	103889
6	GAA	G G A	Transition	GAA Glycine	G G A Glycine	Silent mutation	10387
7	TGT	T GT	Deletion	UGU Cysteine	T GU	Frameshift mutation	104074

The primary structure of the nucleotide sequence of TCF7L2 gene for diabetic patients was translated to amino acid sequence by using the BLASTX tool from NCBI. The structure analysis provided the physiochemical properties of TCF7L2 for all samples by using the PortParam server as shown in Table 5.

Table 5: The physiochemical properties of TCF7L2 protein in diabetic patients and the standard from NCBI.

Group	NO. of sample	M.Wt(Molecular weight)	PI	Instability Index(II)
Diabetic Patient	1	5889.72	5.16	48.72
	2	6123.13	5.42	59.78
	3	5194.02	5.44	74.92
	4	5278.18	5.16	81.90
NCBI Ref		7482.73	5.39	40.46

After using PortParam serves, the results show that the substitution of amino acid has many effects on the TCF7L2 protein and the function as a transcription factor for β -cell in the pancreas. It observed that the physicochemical properties of the protein in Table 5, which the stability of protein determined by Instability Index (II) and should be less than 40. So, the substitution in the gene effects on (II) which probably consider the protein with above of 40 (II) as unstable protein, than the primary structure-dependent statistical (II) that was created for predicting protein stability in vivo [22]. The protein's Isoelectric Point (PI) is the pH at which it carries no net charge, so it will precipitate in an acidic buffer if the PI is less than 7, and soluble in a basic buffer

if the PI is more than 7. The (PI) of TCF7L2 in T2DM patients were (5.16) while the (PI) of TCF7L2 that was retrieved from (NCBI) was 5.39 so they are likely to precipitate in acidic buffers [23]. These results are approved with other studies that conformed to the mutation effects on primary structure and physicochemical characteristics especially protein stability [24].

3. 3. Protein Structure Analysis

The TCF7L2 protein secondary structure was determined by using PSIPred online software program that gave alpha helix predicated and beta-turn and random coil for samples groups of T2DM patients and healthy control, Figures 4 & 5 show the secondary structure of T2DM patient and (NCB) Rf seq. alpha helix predicted values in patient (27.6%), while in TCF7L2 (34.8%) and the coil predicated values (patient = 58.8%, TCF7L2=55.6%), and the β -turn predicates values shown (patient=13.6%, TCF7L2=9.6%).

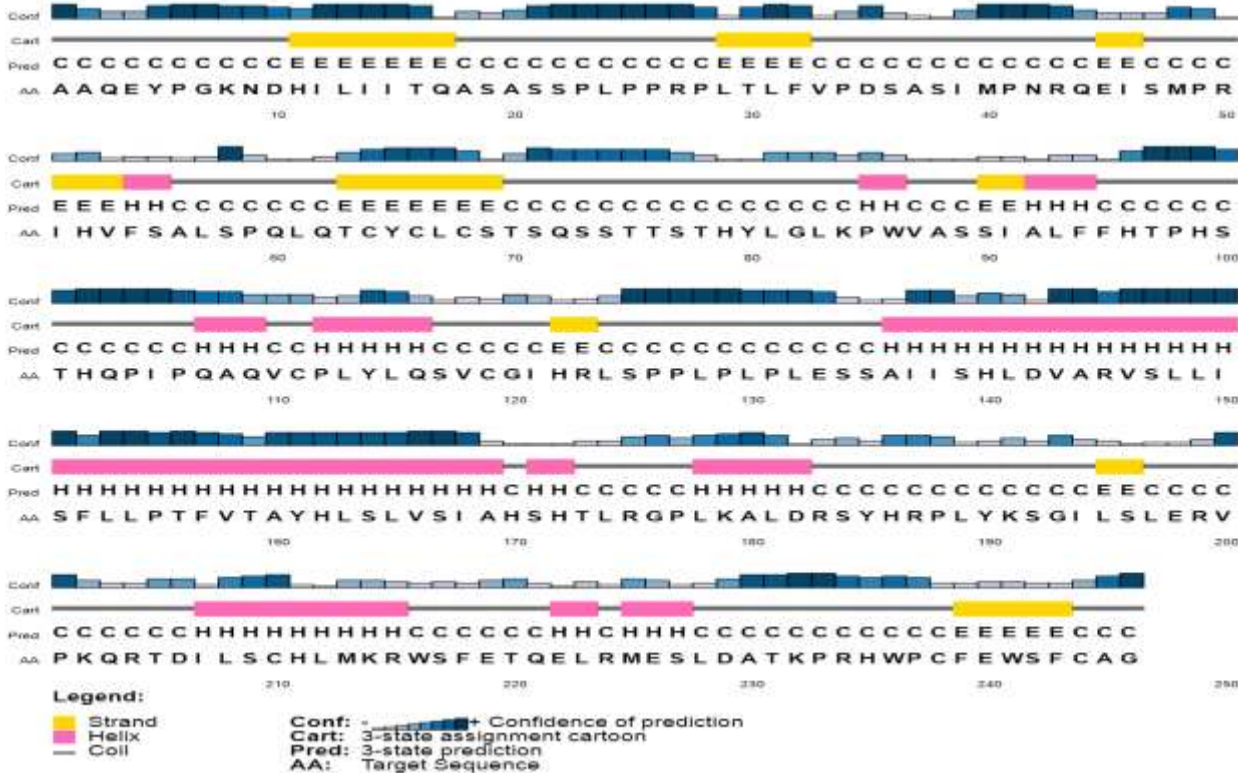


Figure 4: The secondary structure of TCF7L2 in T2DM Iraqi patients.

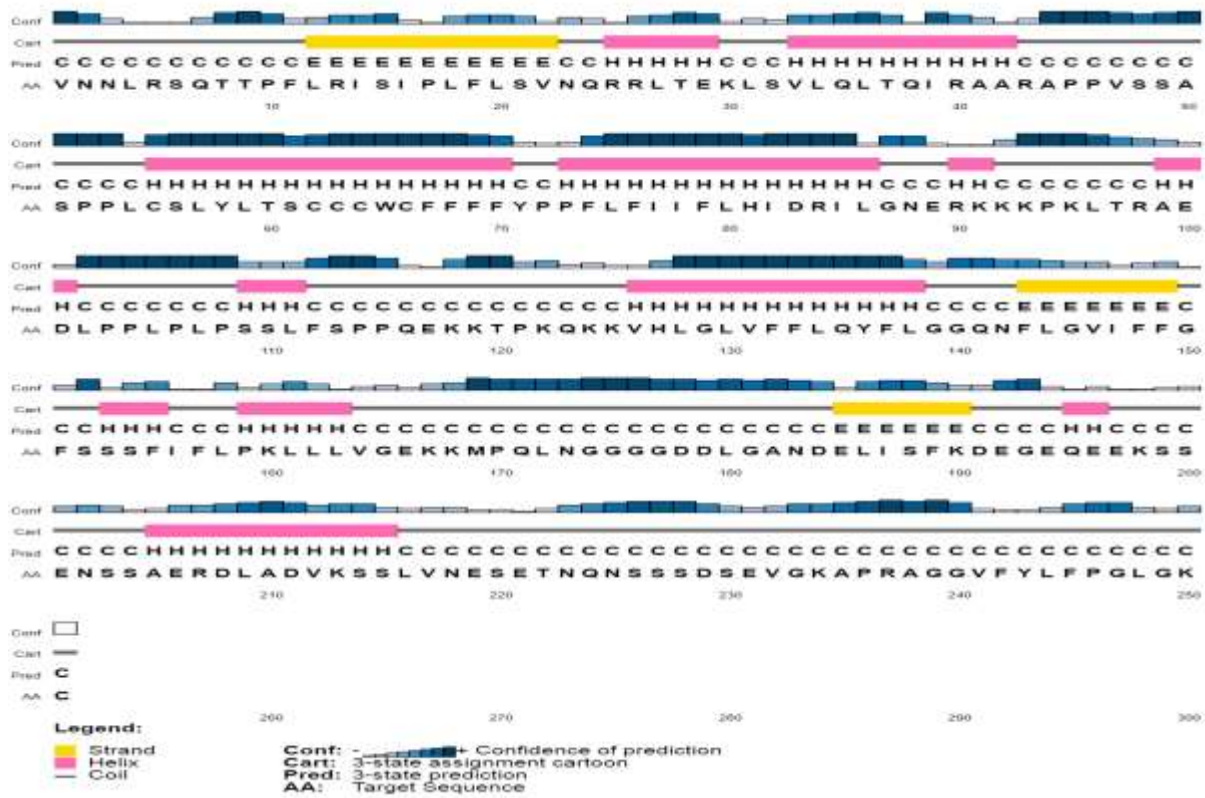


Figure 5: The secondary structure of TCF7L2 that retrieved from (NCBI).

The results of TCF7L2 protein tertiary structure for patients with T2DM that compared with the 3D structure of the TCF7L2 protein from NCBI by using the Phyre2, those the tertiary structure of TCF7L2 gene in patients with diabetes appeared out of the templates the provided by Phyre2 tool (Figure 6). The highest % I.d (94%) was used to predict the 3D structure.

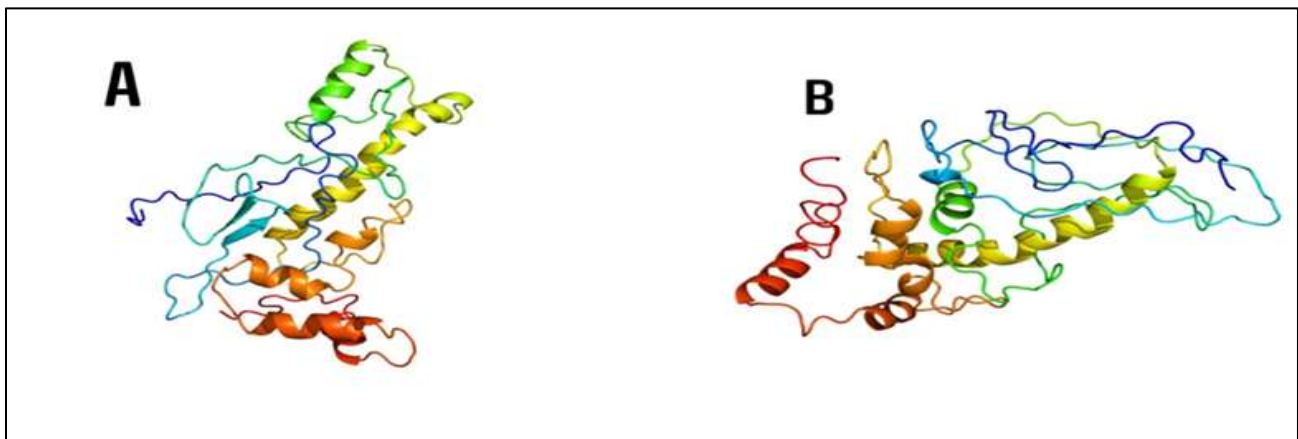


Figure 6: The 3-dimension structure of TCF7L2 protein in A) T2DM Iraqi Patient B) TCF7L2 protein from NCBI.

When compared to TCF7L2 retrieved from NCBI, the mutations on the TCF7L2 gene for T2DM patients caused different effects on the structure of the protein, as shown in the structural analysis result. This result proved the agreement with the studies [25] that demonstrated the mutations cause large changes in the sequences that effected on the structure of the protein.

4. Conclusions

In the present study, the mutations in the TCF7L2 gene in T2DM Iraqi patients affect the physicochemical properties of TCF7L2 protein such as molecular weight, PI, and the stability of protein when compared with TCF7L2 retrieved from NCBI. Also, the mutations effected the structure of TCF7L2 by changing the number and position of alpha-helix, beta-turn, and coil, which lead to the loss of the function of TCF7L2 protein as a transcription factor.

Acknowledgment

The authors express their sincere appreciation to Wahj Al-DNA Laboratory, University of Technology, Baghdad, Iraq for its help in some matters of the current study.

Conflict of Interest

There are no conflicts of interest regarding the publication of this manuscript.

References

- [1] N. M. Luscombe, D. Greenbaum, and M. Gerstein, "Mim01-Luscombe-What-Is-Bioinf," p. 346–358, 2001.
- [2] B. A. Taha, "Perspectives of Photonics Technology to Diagnosis COVID–19 Viruses: A Short Review," *Journal of Applied Science and Nanotechnology*, vol. 1, no. 1, p. 1–6, 2021.
- [3] R. Bao, L. Huang, J. Andrade, W. Tan *et al.*, "Review of current methods, Applications, And data management for the bioinformatics analysis of whole exome sequencing," *Cancer Informatics*, vol. 13, no. 3, p. 67–82, 2014.
- [4] N. S. Younus, M. F. Altaee, and Z. A. Sharba, "Correlation of MicroRNAs-122a Gene Expression with Diabetic for Iraqi Patients," *Journal of Applied Science and Nanotechnology*, vol. 1, no. 3, p. 64–72, 2021.
- [5] J. Karalliedde and L. Gnudi, "Diabetes mellitus, a complex and heterogeneous disease, and the role of insulin resistance as a determinant of diabetic kidney disease," *Nephrology Dialysis Transplantation*, vol. 31, no. 2, p. 206–213, 2016.
- [6] K. J. Gaulton, T. Ferreira, Y. Lee, A. Raimondo, *et al.*, "Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci," *Nature Genetics*, vol. 47, no. 12, p. 1415–1425, 2015.
- [7] H. Kolb and S. Martin, "Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes," *BMC Medicine*, vol. 15, no. 1, p. 1–11, 2017.
- [8] L. Prokunina-Olsson, C. Welch, O. Hansson, N. Adhikari, *et al.*, "Tissue-specific alternative splicing of TCF7L2," *Human Molecular Genetics*, vol. 18, no. 20, p. 3795–3804, 2009.
- [9] T. Jin and L. Liu, "The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus," *Molecular Endocrinology*, vol. 22, no. 11, p. 2383–2392, 2008.
- [10] N. Facchinello, E. Tarifeño-Saldivia, E. Grisan, M. Schiavone, *et al.*, "Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration," *Scientific Reports*, vol. 7, no. 1, p. 1–16, 2017.
- [11] C. Herder and M. Roden, "Genetics of type 2 diabetes: Pathophysiologic and clinical relevance," *European Journal of Clinical Investigation*, vol. 41, no. 6, p. 679–692, 2011.
- [12] Z. A. Ghareeb, H. M. A. Alkhafaji, M. K. A. Al-saedi, and I. H. Company, "The Effect of LRP-1 Gene and Mir-205 on Cardiovascular Disease in Iraqi Patients with Familial Hypercholesterolemia," *Journal of Applied Science and Nanotechnology*, vol. 1, no. 3, p. 24–31, 2021.
- [13] P. Desjardins and D. Conklin, "NanoDrop microvolume quantitation of nucleic acids," *Journal of Visualized Experiments*, vol. e2565, no. 45, p. 1–5, 2010.
- [14] M. F. A. Jan, B. Q. H. Al-saadi, H. M. A. Al-khafaji, M. K. Aneed, and I. H. Company, "Primer and Probe Designing to Detect SNP rs 4073 in Interleukin-8 Gene in Iraqi Patients with Bronchial Asthma," *Journal of Applied Science and Nanotechnology*, vol. 1, no. 3, p. 51–57, 2021.
- [15] A. M. Haines, S. S. Tobe, H. J. Kobus, and A. Linacre, "Properties of nucleic acid staining dyes used in gel electrophoresis," *Electrophoresis*, vol. 36, no. 6, p. 941–944, 2015.
- [16] A. Easton, "Brain on Fire – My Month of Madness," *The Yale Journal of Biology and Medicine*, vol. 4, p. 149, 2013.
- [17] M. Dhary Kamel, A. Abdullah Mohammed, and A. Abdulhafidh Ibrahim, "Sequence and Structure Analysis

- of CRP of Lung and Breast Cancer Using Bioinformatics Tools and Techniques,” *Biosciences, Biotechnology Research Asia*, vol. 15, no. 1, p. 163–174, 2018.
- [18] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch, “ExpPASy: The proteomics server for in-depth protein knowledge and analysis,” *Nucleic Acids Research*, vol. 31, no. 13, p. 3784–3788, 2003.
- [19] D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson, and D. T. Jones, “Scalable web services for the PSIPRED Protein Analysis Workbench,” *Nucleic Acids Research*, vol. 41, p. 349–357, 2013.
- [20] V. K. Jatav and S. Gwalior, “Phylogenetic analysis of ras subfamily proteins,” *International Journal of Pharmaceutical Sciences and Research*, vol. 7, no. 3, p.1070-1080, 2016.
- [21] T. Y. Wang, L. Wang, J. H. Zhang, and W. H. Dong, “A simplified universal genomic DNA extraction protocol suitable for PCR,” *Genetics and Molecular Research*, vol. 10, no. 1, p. 519–525, 2011.
- [22] D. G. Gamage, A. Gunaratne, G. R. Periyannan, and T. G. Russell, “Applicability of Instability Index for In vitro Protein Stability Prediction,” *Protein & Peptide Letters*, vol. 26, no. 5, p. 339–347, 2019.
- [23] K. Shimura, W. Zhi, H. Matsumoto, and K. Kasai, “Accuracy in the Determination of Isoelectric Points of Some Proteins and a Peptide by Capillary Isoelectric Focusing: Utility of Synthetic Peptides as Isoelectric Point Markers,” *Analytical chemistry*, vol.72, no.19, p.4747-4757, 2000.
- [24] H. Venselaar, T. A. H. te Beek, R. K. P. Kuipers, M. L. Hekkelman, and G. Vriend, “Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces,” *BMC Bioinformatics*, vol. 11, no. 1, p. 548, 2010.
- [25] N. Sinha and R. Nussinov, “Point mutations and sequence variability in proteins: Redistributions of preexisting populations,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 6, p. 3139–3144, 2001.