Regular Article

PHYSICAL CHEMISTRY RESEARCH
Published by the
Iranian Chemical Society
www.physchemres.org
info@physchemres.org

Phys. Chem. Res., Vol. 11, No. 3, 653-673, September 2023
DOI: 10.22036/pcr.2022.349495.2129

In Silico Analysis Predicting the Structural and Functional Effects of High-risk nsSNPs in the Human *GCK* Gene Associated with Gestational Diabetes

J. Khamlich[a,b,*], I. Douiyeh[a,b] A. Saih[b], A. Baggar[a], A. Kettani[b] and A. Safi[a]

[a]*Laboratory Biochemistry Environment and Agri-food, Department of Biology, Faculty of Science and Technics Mohammedia, Hassan II University Casablanca, Morocco*

[b]*Laboratory of Biology and Health, URAC 34, Faculty of Sciences Ben M'Sik Hassan II University of Casablanca, Morocco*

(Received 29 June 2022, Accepted 30 September 2022)

To assess the effect of missense variants and the degree of pathogenicity of each nsSNP on the function, structure, and stability of the glucokinase protein (GCK), several algorithms were utilized including PHD-SNP, PROVEAN, SNPs&GO, PolyPhen 2.0, SIFT, MutPred, I-Mutant, MUpro, Consurf, and STRING. We have evaluated the flexibility levels of the residues in GCK protein using the PredFlexy server and then we used the DynOmics server to study the molecular dynamics and understand the correlated communications between the residues. Towards the end, TM-align and the PyMol program were used to analyze the topology and structural similarity between the native model and the generated mutants. In total, seven out of eight nsSNPs were found to be the most damaging variants and to exhibit a deleterious effect on the structure of the GCK protein, and probably on its function. This in silico study gives information on functional polymorphisms that impact the structure, stability, and function of the GCK protein, and consequently susceptibility to Gestational Diabetes.

Keywords: GCK, Gestational diabetes, Missense variants, Prediction, Bioinformatics tools, Genetic susceptibility

INTRODUCTION

Gestational diabetes (GD) is one of the most common complications of pregnancy and is a widespread condition with a prevalence that has increased by more than 30% in a number of nations, which has prompted researchers to take a major interest in it.

GD can occur as a result of increased levels of blood hormones whose function is antagonistic to insulin, resulting in insulin resistance. The other case is the occurrence of GD due to increased insulin resistance during pregnancy stimulated by a genetic predisposition to impaired islet β-cell function [1].

Several genes are involved essentially *IRS-1*, *TCF7L2*, *KCNJ11*, *PPARgamma,* and *GCK*, the latter having been the target of several research teams either by bioinformatics approaches or by experimental studies. See that glucokinase is the essential enzyme in the regulation of glucose homeostasis [2,3].

The GLUT family of glucose transporters is functionally dependent on hexokinase (HK) phosphorylation of intracellular glucose to sustain the glucose gradient throughout the plasma membrane. The first enzyme of the glycolytic pathway is hexokinase present in mammalian cells. This enzyme engages glucose towards catabolism by catalyzing its phosphorylation into glucose 6-phosphate with ATP as the phosphate donor [4].

Therefore, human glucose metabolism is closely regulated by the activity of glucokinase (GCK), a 12-exon gene located on 7p15.3-p15.1 that is responsible for the expression of glucokinase (a monomeric protein of 465 amino acids and weighing approximately 50 kD).

The initial and limiting step of glycolysis in the liver and pancreas is catalyzed by Glucokinase protein which catalyzes the ATP-dependent conversion of glucose to glucose

*Corresponding author. E-mail: jihane.khamlich-etu@etu.univh2c.ma

6-phosphate [5,6]. Glucokinase is a protein that is mainly related to disorders of carbohydrate metabolism. This explains its involvement in the occurrence of several pathologies in this sense. A mutation in the *GCK* gene responsible for the synthesis of the glucokinase protein is associated with the development of several pathologies, including neonatal hypoglycemia in infants and children, maturity-onset diabetes 2 in young people, also called Mody, as well as gestational diabetes, which is the subject of this study and which has been demonstrated by several teams [7,8]. M. Capuano *et al.* (2012) validated a study on the association between the *GCK* gene and Type 2 Maturity Onset Diabetes of the Young (MODY2) the study which was composed of 3 parts, experimental on 66 diabetic children from southern Italy suspected of MODY2, statistical and a small bioinformatics part to which they revealed six new mutations of *GCK* and they elucidated the structure function link between human GCK alteration and MODY2 [9]. The subject that was treated also by S. Costantini *et al*. (2014) who showed the involvement of certain heterozygous mutations with loss of function in the gene of *GCK* in the provocation of diabetes to maturity of the subtype GCK of young people [10]. In the same context, a study was done to show the association between gestational diabetes and glucokinase by Samreen Siddiqui *et al.* (2018) who worked on a sample of 154 women from northern India which they associated with 3 predictive tools of pathogenicity to which they did not find an association between the 2 parameters, The result was inconsistent with the case-control study by Shaat and his team that reported the involvement of *GCK* gene variants in increasing the risk of gestational diabetes in Scandinavian women [11,12]. An nsSNP is a mutation in which one amino acid is replaced by another, which can lead to the generation of a mutant protein with functional and structural alterations that can cause disease, making the identification of the most deleterious nsSNPs resulting in particular effects in humans a real challenge for researchers.

This is feasible today, through in silico approaches that are fast, inexpensive, and above all reliable and with which deleterious nsSNPs in a target gene can be predicted and identified.

For this purpose, our workflow was based on the prediction of the most deleterious non-synonymous nucleotide polymorphisms (nsSNPs) in the *GCK* gene that is associated with gestational diabetes. Using different tools, we examined the functional and structural effects of the nsSNPs to investigate their potentially deleterious effects on the gene. Alternatively, of the confirmation by biological experiment, the study attempts to give an effective method for the fast and cost-effective targeting of deleterious nsSNPs that will be candidates for subsequent functional verification experiments.

This type of approach will help us both deepen our understanding of how these deleterious mutations may alter protein function and identify important gaps that need to be filled and experimentally proven in future studies.

MATERIALS AND METHODS

Datasets

The datasets of the corresponding human *GCK* gene were retrieved after their identification (in November 2020) from the *Ensembl* database (http://www.ensembl.org/) and then focused on the nsSNPs which were subjected to an in-depth study to detect their effects on the protein.

The amino acid sequence was extracted from the Uniprot database and the *GCK* gene and protein information were obtained from the OMIM database (https://www.omim.org/entry/138079); (https://www.uniprot.org/uniprot/Q9BYF1.fasta).

The strategy followed to select and study nsSNPs with structural and functional impact is presented in the figure below (Fig. 1).

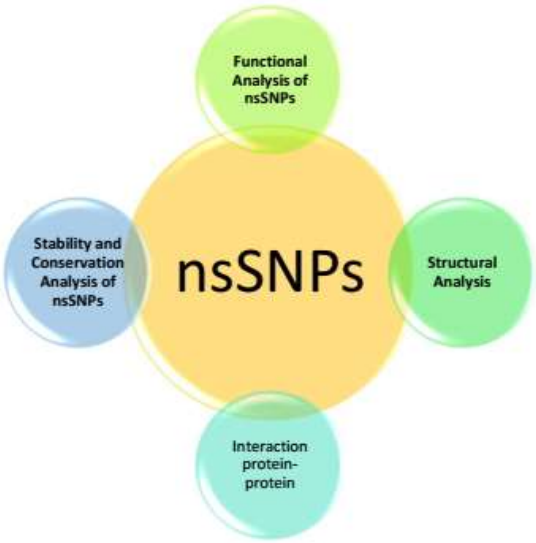


Fig. 1. Graphical illustration of the steps taken to explore the most plausible nsSNPs in the *GCK* gene.

Analysis of the Most Deleterious nsSNPs to the *GCK* Gene by *In Silico* Tools

In order to achieve our objective and detect the deleterious effect of GCK protein nsSNPs, we used several *in silico* tools to predict the pathogenicity of the nsSNPs tested such as PROVEAN, SNPs&GO, SIFT, PHD-SNP, and PolyPhen 2.0 as shown in Table 1, or to reveal their functional effect on the protein such as MutPred, I-Mutant,

Table 1. Tools Used to Identify the most Pathogenic and Damaging nsSNPs

Tool	Description	Availability	Ref.
SNPs&GO	SNPs&GO is a tool to determine the presence or absence of a mutation/disease relationship by exploiting the functional annotation of the protein, seeing that it brings together protein sequence-derived information, evolutionary information and function as encoded in Gene Ontology terms. Probability values > 0.5 for each variant is predicted as an nsSNP disease.	https://snps-and-go.biocomp.unibo.it/snps-and-go/	[13]
SIFT (sorting intolerant from tolerant)	SIFT is an algorithm that uses sequence homology to calculate the probability that substitution will result in an adverse effect on protein function. It uses a query protein that is searched in a protein database to obtain homologous sequences with which it performs an alignment and based on the amino acid composition a score that it calculates. The SIFT score varies between 0 and 1, and if it is between 0 and 0.05 it should affect the protein function.	https://sift.bii.a-star.edu.sg	[14]
Polymorphism Phenotyping v2 (PolyPhen-2)	An automated tool that predicts the effect of substitution on a given protein by generating output elements that classify nsSNPs into benign, possibly damaging, and probably damaging based on phylogenetic and structural features of the sequence as input elements. The server also outputs a count score that is the basis for these predictions, where score 1 is considered the most damaging.	http://genetics.bwh.harvard.edu/pph2/	[15]
PROVEAN	The server gives a prediction for a protein sequence, accepts as input a protein sequence and amino acid combinations, performs a BLAST search to find support sequences that have already been calculated and produce PROVEAN scores. When the final score is less than -2.5, the server predicts that the variant is harmful, and when it is greater than that value, it predicts it as neutral.	http://provean.jcvi.org/index.php	[16]
PhD-SNP (Predictor of human Deleterious Single Nucleotide Poly morphisms)	It is an online server, used to inform about the deleterious or neutral character of the SNP requires as input element the protein sequence, the position, and the new residue.	https://snps.biofold.org/phd-snp/pages/PhD-SNP_HelpOld.html	[17]

MUpro, Consurf, STRING in Table 2, as well as others aiming to demonstrate their structural effect on the GCK protein which are Swiss-Model, QMEAN, Verify 3D, Errat, Ramachandran plot, TM-align, PyMol, DynOmics ENM, andPredyflexy in Table 3.

To do this, First, we used several approaches to predict pathogenicity and deletion of *GCK* gene nsSNPs through different web servers, namely PROVEAN, SNPs&GO, SIFT, PHD-SNP, and PolyPhen 2.0, which generated results classified into two categories, "neutral" or "deleterious".

Table 2. Tools Used to Study the Conservation and Stability of the Protein

Tool	Description	Availability	Ref.
MutPred	a tool that generates molecular mechanisms potentially responsible for disease and returns interpretable pathogenicity score distributions, which is used to demonstrate the impact of variants on protein structure and function that can be used to guide experimental studies of phenotype-altering variants. The output emits a p-value (where the p-value is less than 0.05 and the p-value is greater than 0.01) were considered as reliable and very reliable hypotheses respectively.	http://mutpred.mutdb.org	[18]
I-Mutant	In order to verify the stability of the target protein, I-Mutant 2.0 was used, which is a web server based on a support vector machine that helps to detect any change in stability during a mutation of the single-point protein from the sequence information. The tool predicts the Reliability Index (RI) of results ranging from 0 to 10, with 10 being the highest reliability, as well as a value for predicting energy change (DDG).	https://folding.biofold.org/i-mutant/i-mutant2.	[19]
MUpro	To ensure the result provided by I-Mutant, we used MUpro, a server that can forecast the modifications in protein stability for single amino acid mutations, it is based on both; neural networks and support vector machines, that each will provide us with a different result that are the value of energy change (DDG) and the sign of energy change with a score to measure the confidence of prediction that varies between -1 and 1	http://mupro.proteomics.ics.uci.edu	[20]
Consurf	The ConSurf server is a bioinformatics tool used to estimate the rate of evolution of positions of an amino acid in a protein based on phylogenetic relationships between homologous sequences. This technique allows to determine the functional sections of a protein by analyzing the degree of this conservation which varies from 1 to 9. Thus, degree 9 corresponds to the most conserved residue, the lower the score, the weaker the conservation.	https://consurf.tau.ac.il	[21]
STRING	Then, the STRING web server was used to build a network including physical and functional interactions of a target protein. The basic unit of this server is the functional association, so it highlights the proteins that participate in a given biological function.	https://string-db.org/	[22]

Table 3. The Tools Used to Verify the Quality of the 3D Structures of the Models that were Used to Identify Structural Modifications

Tool	Description	Availability	Ref.
Swiss-Model	SWISS-MODEL workspace is a web server dedicated to the modeling of protein structure homology; it uses a database deduced from the PBB. This will allow a model generation based on sequences or templates. The models are annotated with information on the experimental method used, resolution, mean strength potential to allow rapid retrieval of relevant structural information when selecting the model.	http://swissmodel.expasy.org/workspace/	[23]
QMEAN	which stands for Qualitative Model Energy Analysis with its clustering method, provides a z-score to evaluate the quality of the model by comparing it to reference structures of the same size present in the PDB and solved by experimental techniques.	https://swissmodel.expasy.org/qmean/help	[24]
Verify 3D	As we used Verify 3D tool to determine the compatibility of the amino acid sequence with the 3D structure of the protein, 65% of the amino acids should have a score higher than 0.2 to validate the model according to this software.	http://services.mbi.ucla.edu/Verify_3D/	[25]
ERRAT	Errat software was also used to examine the statistics of non-bonded interactions between different types of atoms so that it could then plot the values of the error function versus the position of residues.	http://services.mbi.ucla.edu/ERRAT/	[26]
Ramachandran plot	For the Ramachandran plot, it gives an idea of the conformation of a protein. The ideal result for this software is to have more than 90% of the residues in the central regions see that the proportion of residues in the central regions is one of the best guides of stereochemical quality.	http://services.mbi.ucla.edu/SAVES/Ramachandran/	[27]
TM-align	As we used the TM-align server to compare the 2 structures of mutated and native proteins, see that this tool allows the superposition of 2 structures based on a residual-residual alignment and on the calculation of a modeling score (TM-score) and value (RMSD). The output elements are a TM-score that must be between 0 and 1, of which 1 designates a perfect match between the structures as well as an RMSD value that when it is high it implies a greater variation between the native and mutated structures.	https://zhanglab.ccmb.med.umich.edu/TM-align/	[28]
PyMol	is a tool that provides a high-quality interface and requires a PDB file for the purpose of visualizing, analyzing, and manipulating the predictions of residues and evolutionary sites in protein structures	https://pymol.org	[29]
DynOmics ENM	is a server that enables for the efficient generation of information on collective dynamics. This tool gathers three components: an evaluation of the collective movements of biomolecules as well as that of key sites involved in chemical processes, although resolution exchanges between the complete atomic and CG representations.	http://enm.pitt.edu/	[30]
Predyflexy	The server is used to provide a prediction of both the flexibility of the proteins and the local structure of the proteins, which is determined using the normalized B-factors and RMSF, the latter calculated by the GROMACS tool after superimposing each snapshot structure on the initial conformation. It gives a result in two forms: a graph that illustrates the sequence according to the prediction of flexibility (which varies from 0 to 2 for highly flexible residue) and the confidence index (which varies from 1 to 19 and its increase reflects the accuracy of the prediction) and a table that includes the same information on each amino acid of the sequence of a target protein.	https://www.dsimb.inserm.fr/dsimb_tools/predyflexy/index.html	[31]

Prediction of Functional and Structural Changes

To predict functional changes, the Mutpred tool was used to identify molecular mechanisms potentially responsible for gestational diabetes. I-Mutant 2.0 and MUpro were utilized to discover the effects of high-risk nsSNPs on the stability of the protein upon amino acid substitution by examining the free energy. The ConSurf web server was used to evaluate the rate of amino acid position evolution in the GCK protein based on phylogenetic relationships between homologous sequences and the STRING tool was used to construct a protein-protein network including physical and functional interactions of a target protein.

Validation of the Generated Models

To predict structural changes we used the Swiss-Model algorithm, to create 3D structures for the native protein and the mutant which were validated using the following verification tools (Ramachandran plot, Verify 3D, Errat, and QMEAN) and then we used the TM-align Server to compare the structures of the native protein with the mutants by superimposing the structures based on the residual-residual alignment, on the calculation of the modeling score (TM score) and the root mean square deviation (RMSD).

The root mean square fluctuation (RMSF) was calculated using the PredyFlexy server to predict the dynamic class of an amino acid residue (flexible, intermediate, rigid), and also to estimate whether each residue is likely to generate conformational changes within the protein.

And towards the end, we used the DynOmics ENM server to provide a prediction of both protein flexibility and local protein structure, which is determined using normalized B-factors and RMSF, the latter being calculated by the GROMACS software after superimposing each instantaneous structure on the initial conformation.

Calculation of the Cumulative Score of nsSNPs

In order to give a sufficiently reliable prediction, we calculated the cumulative score using the Sum Excel function, which was based on the aggregation of the prediction results of all tools used for each nsSNP (PHD-SNP, PROVEAN, SIFT, SNPs&GO, PolyPhen 2.0, I-Mutant, Mutpred, MUpro, ConSurf, and TM-align) [32]. Subsequently, we set a restricted cumulative score value (which is 8/10 software that showed a positive prediction 1) when the results of the ten software tools were combined, each substitution predicted by at least 8 tools as deleterious was classified to be high-risk pathogenic for GCK. This makes this step a crucial part of predicting the most deleterious nsSNPs based on the results of all the servers we used to increase the reliability of our results.

RESULTS

Dataset

We started the work with preliminary research on the GCK gene, in which we were interested in the Homo sapiens data, in this research 80413 SNPs were detected in totality among which we find 1572 in the 3'UTR region, 618 in the 5'UTR region, 52295 in the intron region, 1459 synonymous variants and 2783 nsSNPs. Only the last category of nsSNPs have been investigated during this study and examined for further analysis, as shown in the graphical representation in Fig. 2.

In total, 12 different web servers and bioinformatics programs were used to evaluate the effects of the selected nsSNPs. In this study, the aim is to rely on a set of algorithms that perform complementary tasks to increase the reliability of the results generated.

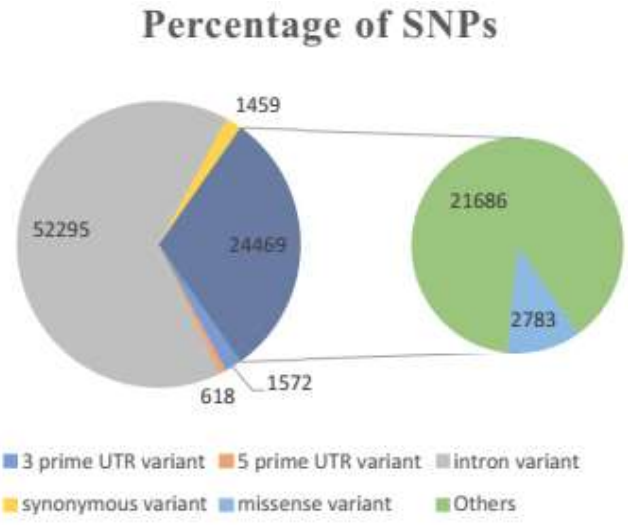


Fig. 2 The diagram represents the percentage distribution of SNPs; nsSNPs, 3'UTR region, 5'UTR region, intron region, and other SNP types in the GCK gene.

In order to distinguish the functional effects of the mutation on the protein we used the following algorithms: PROVEAN, SNPs&GO, SIFT, PHD-SNP, and PolyPhen 2.0, then the tools I. Mutant and MUpro tools were used to determine the effects of mutations on the stability of the GCK protein and then Consurf, MutPred, Swiss-Model, and STRING tools were used to determine the impact of mutations on the functional and structural level and also to determine the interactions between proteins

Identification of Damaging nsSNPs

To extrapolate deleterious nsSNPs that can significantly alter the structure or function of the GCK protein, five servers namely PROVEAN, SNPs&GO, SIFT, PHD SNP, and PolyPhen were used.

And since each of these algorithms uses different parameters to evaluate nsSNPs as deleterious or neutral, we set a score of 4 tools out of 5 to determine the pathogenicity of an nsSNP, out of 8 nsSNPs, 7 were predicted to be deleterious nsSNPs in all computational algorithms, the result of our analysis in has been summarized in Table 4 below.

Characterization of Protein Stability Changed by Mutations

A stability analysis was performed on the 7 nsSNPs that were predicted to be deleted at high risk to detect changes in *GCK* stability in terms of Reliability Index (RI) and a DDG value, this was tested by MUpro and I-Mutant. The result revealed that the 7 deleterious nsSNPs caused a decrease in protein stability as Table 5 indicates.

Analysis of the Evolutionary Conservation of the GCK Protein

An analysis by the Consurf web server was carried out to predict the conservation regions located in the GCK protein to identify high-risk structural and functional residues of the protein using evolutionary conservation and accessibility to solvents, see that mutations located in highly conserved regions can directly affect the protein function of this gene. The analysis revealed that the residues namely *I130T, C371F, V181M, V182M* are moderately conserved and buried, while the amino acid substitution namely *E237k* is moderately conserved and exposed, and for *E236K* and *T206P* they are exposed and highly conserved. The results of this analysis are presented in Fig. 3 and the details on the conservation score and prediction are in Table 6.

Prediction of Molecular Change of Substitutions by MutPred

MutPred server was developed to classify an amino acid substitution as either deleterious or neutral, the prediction is based on changes at the atomic and molecular level of the substitution related to the disease, which means stability, accessibility to solvents, transmembrane helix, catalytic residues, and intrinsic disorder. MutPred calculates the G-score to predict the probability of substitution to be deleterious which is supposed to be 0.5 or more, an analysis

Table 4. Screening for Possible Deleterious nsSNPs *GCK* Gene Using Different *In Silico* Tools

RS	Substitution	SIFT	Score	PolyPhen 2.0	Score	SNP& Go	Score	PHD-SNP	Score	PROVEAN	Score
rs1036483919	*I 130T*	Del	0	Pb.D	0.571	Dis	10	N	5	Del	-3.79
rs587780345	*V181M*	Del	0	Pro.D	0.998	Dis	10	N	3	Del	-2.856
rs587780345	*V182M*	Del	0	Pro.D	0.999	Dis	10	N	4	Del	-2.748
rs587780346	*T206P*	Del	0	Pro.D	0.999	Dis	10	Dis	8	Del	-5.864
rs587780347	*E236K*	Del	0.03	Pb.D	0.451	Dis	10	Dis	9	Del	-3.865
rs587780347	*E237K*	Del	0.03	Pro.D	0.991	Dis	9	Dis	2	Del	-3.217
rs193922331	*S263P*	Del	0	B	0.044	Dis	10	Dis	4	N	-1.728
rs587780343	*C371F*	Del	0	Pro.D	0.999	Dis	10	Dis	7	Del	-10.088

Del = Deleterious; Pb.D = Possibly damaging; Pro.D = Probably damaging; B = Benign; Dis = Disease; N = Neutral.

Fig. 3. Analysis of amino acid residues conserved during the evolution of *GCK* human gene by ConSurf whose color-coding bar indicates the conservation score.

Table 5. Validation Result of Protein Stability Change by Using I-Mutant 2.0 and MUpro

		I-Mutant			MUpro	
RS	Substitution	Stability	RI	DDG	Stability	DeltaG
rs1036483919	*I130T*	D.S	4	-0.43	D.S	-1.911
rs587780345	*V181M*	D.S	1	0.17	D.S	-0.62
rs587780345	*V182M*	D.S	6	-0.16	D.S	-0.361
rs587780346	*T206P*	D.S	5	-2.72	D.S	-1.54
rs587780347	*E236K*	D.S	9	-2.73	D.S	-0.84
rs587780347	*E237K*	D.S	6	-1.49	D.S	-1.028
rs193922331	*S263P*	D.S	3	-0.14	D.S	-0.46
rs587780343	*C371F*	D.S	9	-1.96	D.S	-0.860

DDG = Free energy change value; RI = Reliability index; Decrease stability = D.S.

Table 6. Evolutionary Conservancy of Amino Acids in GCK Analyzed by Consurf

nsSNP Id	Residue	Conservation score	Prediction
rs1036483919	*I130T*	6	Buried residue
rs587780345	*V181M*	7	Buried residue
rs587780345	*V182M*	8	Buried residue
rs587780346	*T206P*	9	Highly conserved and exposed
rs587780347	*E236K*	9	Highly conserved and exposed
rs587780347	*E237K*	5	Buried residue
rs193922331	*S263P*	5	Average and exposed
rs587780343	*C371F*	8	Buried residue

Table 7. Analysis of the Structural and Functional Effect of nsSNPs of the Human GCK Gene by the MutPred Server

Substitution	MutPred 2 score	Molecular mechanisms with P-values <= 0.05	P-value
I130T	0.740	Altered metal binding	8.8e-03
		Loss of relative solvent accessibility	0.02
		Gain of allosteric site at F133	0.02
		Altered stability	0.04
		Gain of sulfation at Y125	0.01
V181M	0.694	Loss of relative solvent accessibility	0.02
		Altered metal binding	0.05
V182M	0.834	Loss of relative solvent accessibility	0.03
E236K	0.934	Altered metal binding	3.6e-03
		Gain of allosteric site at E236	2.9e-03
		Gain of relative solvent accessibility	0.03
		Altered DNA binding	7.8e-03
		Loss of catalytic site at E236	0.05
		Altered ordered interface	0.02
E237K	0.769	Altered metal binding	9.9e-04
		Gain of allosteric site at Y234	5.7e-03
		Loss of catalytic site at E236	0.05
T206P	0.939	Altered metal binding	2.7e-03
		Gain of allosteric site at N204	1.1e-03
		Altered ordered interface	0.03
		Loss of helix	0.03
		Gain of catalytic site at N204	2.5e-03
		Altered DNA binding	0.03
S263P	0.662	Loss of strand	0.02
		Altered metal binding	0.03
C371F	0.930		

was made through the latter and the results are represented in Table 7.

Validation of the Generated Models

The crystal structure of human glucokinase (GCK) is already present in Protein Data Bank (PDB), crystallization is done by DIFFRACTION X-RAY 10.2210/pdb1V4S/PDB (ID 1V4S, with resolution equal to 2.4 Å), but homology modeling on the native model has been performed to model and target the mutants. For that, we used the homology modeling tool Swiss Model to produce the 3D structure of the native protein and the 7 mutants of the GCK protein, to discover whether the 7 high-risk missense variants modify the native structure of the GCK protein.

SWISS-MODEL was used (PDB ID 6e0i.1.A) as a model to predict the 3D models and the structures were then visualized. The 7 deleterious nsSNPs were individually substituted in the native sequence, the sequences were submitted to the homology modeling tool SWISS-MODEL and then 3D structures models were generated for all mutants. After the generation of the native model of the target protein as well as the different mutants, we proceeded to the validation and verification of these models using five structural descriptors by inserting the structure as PDB form such as Verify 3D, Errat as well as the output data are taken from Swiss-Model Ramachandran Plot, QMEAN, and GMQE score, to evaluate the overall quality of the generated models.

The open-source program PyMol was used to superimpose two protein structures of the native model with each mutant (Table 8), this superimposition which was evaluated on the basis of the RMSD calculation the results are mentioned below in (Table 9). All the models we have generated have been visualized by (DassaultSystèmes BIOVIA, Discovery Studio Modeling Environment, Release 4.5, San Diego: DassaultSystèmes, 2015).

Comparative Modelling

To analyze the topology and the structural similarity between the native model and the generated mutants the TM-align tool was used. The tool calculated a TM and RMSD score to measure the average distance between the skeletons of the superimposed proteins. The tool provides 2 scores as a result: an RMSD value that must be greater than 0.15 to be considered significant and a TM-align score that is normally between 0 and 1, a TM-align score < 0.2 indicates no similarity between two structures, and a TM-align score > 0.5 means that the structures share the same fold [28, 33]. As we used the RMSD result that was generated during the overlay see that it is the basic element for this program the results were also mentioned in Table 9 as we calculated the average generated by the two tools.

Protein Flexibility

In order to categorize amino acid residues into flexible, intermediate, or rigid sites and evaluate the levels of residue dynamics in a target protein we used the predicted root mean square fluctuations (RMSF) and B-factors obtained from the PredyFlexy prediction server, seeing that this will direct us towards amino acid residues that are highly flexible and that are essential for proteins during conformational changes and folding because they grant them great movement.

RMSF is the value of the displacement of a particular atom with respect to the reference structure, its calculation is usually based on the alignment of the rigid bodies of the structures in each frame of the simulation at the reference coordinates, and a high RMSF value reflects the presence of significant fluctuation [34].

So, our analysis performed by PredyFlexy as represented in Table 10 showed that residues *T206P, S263P, V181M, E236K,* and *E237K* shared strong to moderate flexibility scores respectively with confidence indexes ranging from 6 to 15, while residues *I130T, V182M, C371F* were identified as rigid with low index scores.

Dynamic Cross-correlation Matrix Analysis by the DynOmics Server

Dynamic Cross-Correlation Matrix analysis was performed to understand the correlated communications between residues by taking into account the environment, prediction of potential functional sites, and reconstruction of conformers of all atoms from structures [35].

The result showed that compared to the wild type, the variants *i.e.*, *V182M, T206P, and C371F* slightly increased the degree of positive (red color) and negative (blue color) correlations observed in the native GCK inversely to the variant *S263P* which was decreased, resulting in a significant

Table 8. 3D Superimposed Structures of Native (Blue) and Mutants (I130T: Cyan, V181M: Red, V182M: Green, C371F: Orange, E236K: Purple, E237K: Magenta, T206P: Yellow, and S263P: White) GCK Investigated by Swiss Model and PyMol

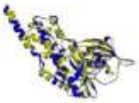
Residue	Superimposed Structures	Properties/ Structure Validation Server
Native Structure		Ramachandran Favoured: 96.93% QMEAN: -0.22 GMQE :0.92 ERRAT :(Overall Quality Factor) : 98.8584 VERIFY 3D: (PASS) 96.94% of the residues have averaged 3D-1D score >= 0.2
I130T		Ramachandran Favoured:97.59% QMEAN :0.17 GMQE :0.92 ERRAT :(Overall Quality Factor) :99.0805 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2
V181M		Ramachandran Favoured:97.37% QMEAN :0.15 GMQE :0.92 ERRAT :(Overall Quality Factor) :99.3213 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2
V182M		Ramachandran Favoured: 96.71% QMEAN: -0.27 GMQE :0.92 ERRAT :(Overall Quality Factor) :98.6301 VERIFY: (PASS) 96.94% of the residues have averaged 3D-1D score >= 0.2
T206P		Ramachandran Favoured:96.71% QMEAN: -0.23 GMQE :0.92 ERRAT :(Overall Quality Factor) :98.8584 VERIFY: (PASS)96.94% of the residues have averaged 3D-1D score >= 0.2
E236K		Ramachandran Favoured:96.93% QMEAN: -0.23 GMQE :0.92 ERRAT :(Overall Quality Factor) :99.3135 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2
E237K		Ramachandran Favoured :97.37% QMEAN :0.34 GMQE :0.92 ERRAT :(Overall Quality Factor) :99.3213 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2
S263P		Ramachandran Favoured: 96.93% QMEAN: -0.32 GMQE:0.91ERRAT: (Overall Quality Factor): 98.8584 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2
C371F		Ramachandran Favoured: 97.59% QMEAN :0.46 GMQE :0.92 ERRAT :(Overall Quality Factor) :98.8532 VERIFY: (PASS) 96.07% of the residues have averaged 3D-1D score >= 0.2

Table 9. Structural Alignment Comparing Mutant and Wild-type GCK Models by TM-align

Model	TM-score	RMSD TM-align	RMSD Pymol	RMSD MOY
I130T	0.99887	0.27	0.054	0.16
V181M	0.99994	0.06	0.061	0.06
V182M	1.00000	0.02	0.014	0.017
T206P	0.99999	0.02	0.013	0.016
E236K	1.00000	0.01	0.002	0.006
E237K	0.99994	0.06	0.060	0.06
S263P	0.99999	0.02	0.008	0.01
C371F	0.99886	0.27	0.061	0.16

Table 10. Results of Flexibility Prediction by Using PredyFlexy Server

Model	I130T	V181M	V182M	T206P	E236K	E237K	S263P	C371F
Predicted Flexibility Class	0	1	0	2	1	1	2	0
Confidence index	17	14	13	14	14	15	6	16

Table 11. Correlation between Predicted and Observed Fluctuations of GCK Native and Mutants

GCK structures	Native	*I130T*	*V181M*	*V182M*	*T206P*	*E236K*	*E237K*	*S263P*	*C371F*
Correlation between observed and predicted fluctuations	0.53	0.53	0.53	0.55	0.58	0.55	0.53	0.51	0.55

correlation in the movement of the residues was noticed in the Dynamic Cross-Correlation Matrix maps analysis as shown in Fig. 4 and Table 11.

Calculation of the Cumulative Score of nsSNPs

We represented on the file (Supplementary Table 1) the nsSNPs studied according to all the tools used in the study to be able to judge their degrees of pathogenicity through the calculation of a cumulative score, the latter where we mentioned the neutral prediction by the value 0 and deleterious by 1.

After combining the result of the ten algorithms, the following amino acid substitutions: *I130T, C371F, V182M, E236K,* and *T206P* were ranked as the most deleterious nsSNPs with a cumulative score of 9 by the ten tools, while two variants *V181M, E237K* obtained a cumulative score of 8 by the ten tools while only one nsSNP *S263P* with a score lower than eight. So, after this step, seven nsSNPs (*I130T, C371F, V182M, E236K, T206P, V181M, E237K*) were evaluated as high-risk pathogenic nsSNPs of the GCK gene and were selected for the rest of the study. This step allowed us to predict the most deleterious nsSNPs with more accuracy and reliability (Show Supplementary Table 1).

Analysis of Protein-protein Interaction Network Characterization by STRING

Mutation can affect the function and structure of a protein, so a mutated protein can interact with other proteins

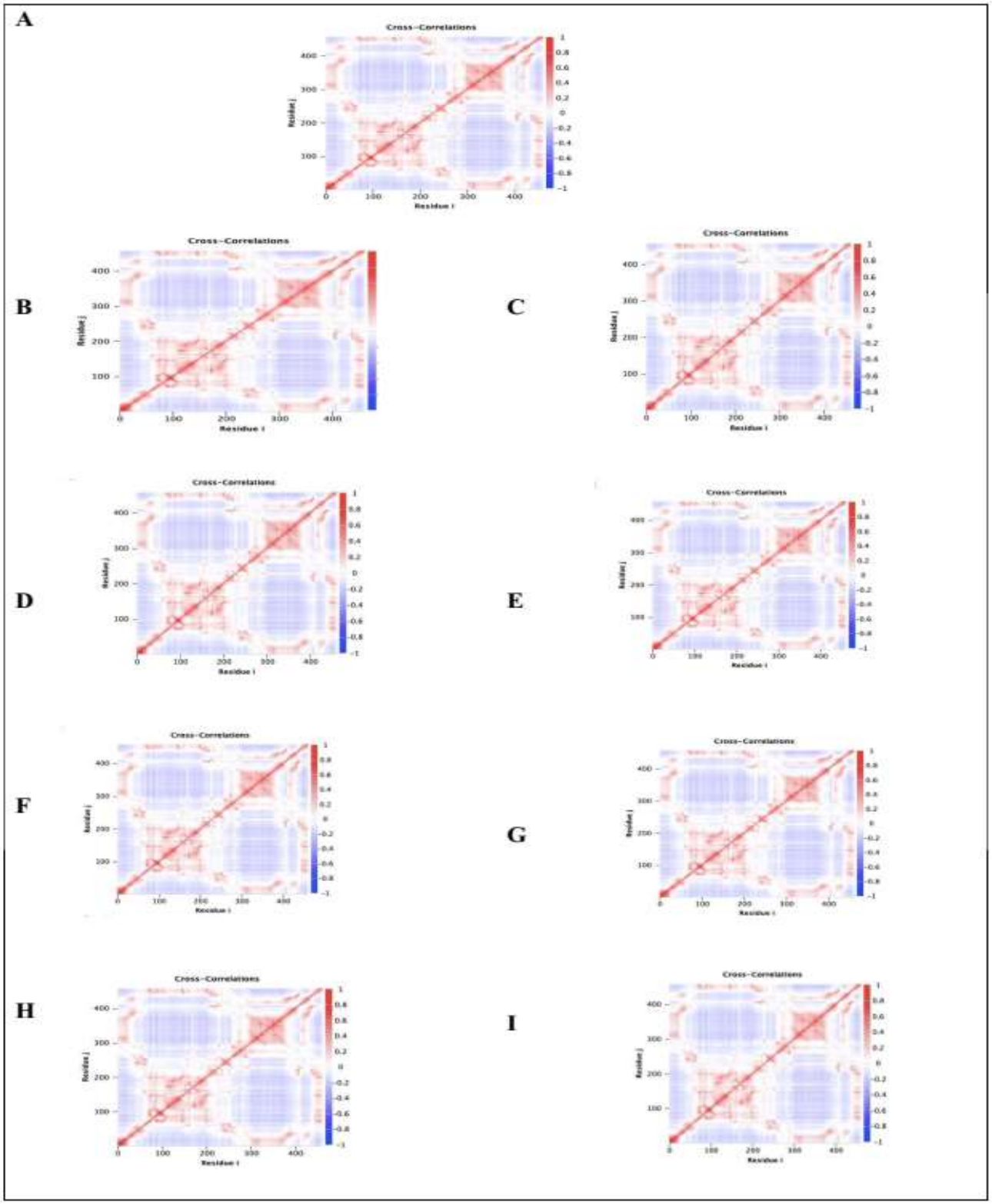


Fig. 4. Results of Dynamic Cross-Correlation Map Analysis (DCCM) of native GCK and mutants: A (wild type), B(*I130T*), C (*C371F*), D (*V181M*), E (*V182M*), F (*E237K*), G (*E236K*), H (*T206P*), I (*S263P*).

and cause phenotypic effects. The interactions of the GCK protein with other proteins were predicted with the online software STRING for associated gene screening. Then, we noted a total of 11 nodes (as shown in Fig. 5) namely GPI (Glucose-6-phosphate isomerase); GCKR (glucokinase regulatory protein); G6PD (glucose-6-phosphate1-dehydrogenase); PGM1 (phosphoglucomutase-1); PDX1 (pancreas/duodenum homeobox protein 1); H6PD (bifunctional endoplasmic protein); TREH (trehalase); NeuroD1 (neurogenic differentiation factor 1); INS (insulin) and G6PC.

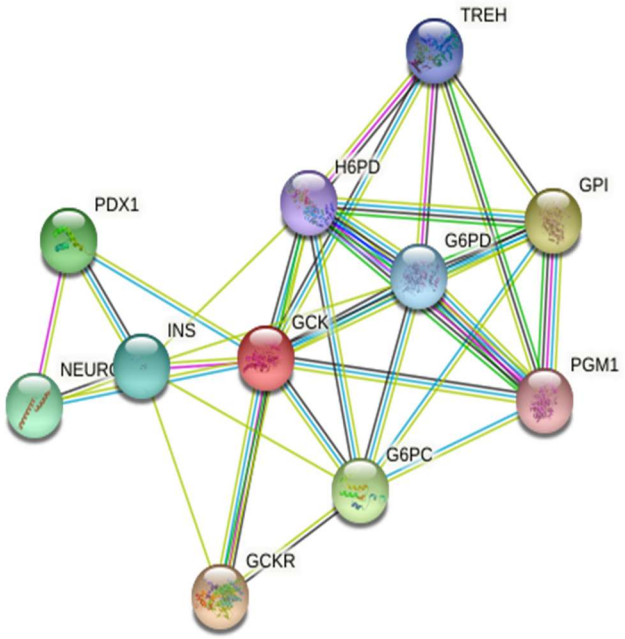


Fig. 5. Protein-protein interaction network of Glucokinase illustrated by STRING server.

DISCUSSION

Glucokinase is primarily synthesized in pancreatic β-cells, where it catalyzes the final phase of insulin secretion, and in hepato-parenchymal cells of the liver, where it contributes in glycogen production. And given its crucial role in preserving glucose homeostasis, its function is largely regulated at several levels. Because it is vulnerable to numerous protein-protein interactions, post-translational modification events, and other processes that have a variety of physiological effects, glucose kinase has the unusual capacity to self-regulate its own activity via sluggish conformational dynamics.

Several diseases have been associated with mutations in this gene, which are either activating mutations resulting in congenital hyperinsulinemia clinically, although several diabetes diseases are caused by loss-of-function mutations [36].

These include chronic kidney disease, hypertension, and especially diseases directly related to glycemic metabolism. In fact, several studies have correlated high fasting blood glucose levels 1 and 2 h after OGTT with the *rs1799884* GCK variant in different populations European, Thai and Scandinavian [37,38], including hyperglycemia, GCK-MODY, and gestational diabetes, for which several studies have examined its association with GCK as a candidate gene, and for which a number of genetic variants have been identified [39-41].

Besides the clinical studies, bioinformatics studies were carried out with the aim of studying the catalytic process of human glucokinase, where they proceeded by homology modeling in order to construct a glucokinase complex with ATP, glucose, and Mg^{2+} (GMAG complex), with which they carried out a simulation of molecular dynamics and calculated the free energy of binding. It was reported as a result that *Lys169Asn* which is one of the natural mutations of the GCK gene has an extremely important functional role in glucose metabolism as it enhances the binding of glucokinase to both ATP and glucose as it binds ATP and glucose together and participates directly in glucose phosphorylation, the result which is key to understand the catalytic mechanism of GK and the cause of MODY pathogenic mechanism due to glucokinase mutation[42]. On the experimental side Ramasammy. R and his team showed in 2021 that the GCK gene polymorphism is associated with gestational diabetes through an experimental analysis performed on a set of 419 samples collected from 210 pregnant women with gestational diabetes and 209 controls, in which they showed the association between the AA genotype (ORs = 2.9) and the A allele (ORs = 2.4) with the increased risk of gestational diabetes [3].

Also, It has been shown by Shaat *et al*. (2006) that common polymorphisms in the GCK gene increase the risk of GD in Scandinavian women [38]. Anette P. Gjesing and her

team also conducted a study in Denmark (2017) in which they reported a prevalence of 5.8% of *GCK*, *HNF1A*, *HNF4A*, *HNF1B*, or *INS* variants in women diagnosed with GD, with the involvement of six variants (*Gly73Arg, Gly73Glu, Ala209Val, Ser442Pro*) of the *GCK* gene that was considered to be possibly pathogenic [43].

In the same context, several studies have experimentally shown the association of some deleterious variants of the gene with the disease and have been reported to influence susceptibility to MG which was in agreement with our result this concerns (*rs193922331* and *rs587780346*)indicating that increased attention to the screening of this gene in women with MG was warranted [44-47]. Maria Negahdar *et al.* (2012) showed that the serine residue at position 263 is associated with GCK-MODY diabetes through protein misfolding leading to destabilization and increased rate of degradation as well as reduced GCK catalytic activity in pancreatic β cells [44]. Another line of research Yael Gozlan *et al.* (2012) have shown experimentally by sequencing and confirmed by DGGE that a mutation at the threonine residue at position 206 is strongly linked with GCK-MODY and they have also supported their result by a bioinformatics approach using the NEST program to which they suggested the effect of the T206P mutation on the amino acids M210 and C233 which are essential for the enzymatic activity of GCK [46].

Among SNPs, nsSNPs are associated with a single substitution of an amino acid that will subsequently encode an erroneous codon which may affect the structure and functionality of the protein.

In this study, we selected highly deleterious mutations of the autosomal GCK gene located on chromosome 7p13 and possessing 12 exons and responsible for the expression of glucokinase (a monomeric protein of 465 amino acids and weighing approximately 50 kD) [48,49]. The genomic data of the variations were extracted from the Ensembl database, from which we selected 8 non-synonymous mutations that were predicted to be the most deleterious and located in conserved positions, the latter of which underwent a thorough analysis through several software programs based on different algorithms to allow annotation of nsSNPs and to detect structural changes upon the selected mutations.

rs1036483919 (I130T). Isoleucine is replaced by threonine at position 130. This mutation has been predicted by 4 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein as it is placed in a moderately conserved and buried region which has been demonstrated by the Consurf server, the mutation will also generate several changes at the molecular level such as Altered Metal-binding, Loss of Relative solvent accessibility, Gain of Allosteric site at F133, Altered Stability and Gain of Sulfation at Y125 the result which was demonstrated by the Mutpred server and validated with a score of 0.740 which is considered significant, as we used the TM-align and Pymol tools to see if the mutant model and superimpose with the native model the result showed a difference and a non-superimposition between the 2 structures with an RMSD of 0.16 which is considered significant.

rs587780343 (C371F). Cysteine is replaced by Phenylalanine at position 371. This mutation has been predicted by 5 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein. The residue is placed in a fairly conserved area and buried with a score of 8 as declared by the Consurf tool, the mutation has also generated a change at the molecular level which manifests itself as an Altered Metal Binding with a very significant p-value (3.6e-03), the average RMSD raised from the two tools used TM-align and Pymol showed a significant difference between the 2 native and mutant structures with an RMSD of 0.16.

rs587780345 (V181M). Valine is replaced by Methionine at position 181. This mutation has been predicted by 4 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein, as it is placed in a moderately conserved and buried region as demonstrated by Consurf with a score of 7, the mutation will also generate several changes at the molecular level such as Loss of Relative solvent accessibility and an Altered Metal-binding with a score of 0.69 which is considered significant since it is higher than 0.5, the average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2 native and mutant structures with an RMSD of 0.06.

rs587780345 (V182M). Valine is replaced by Methionine at position 182. This mutation has been predicted by 4 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein, as it is placed in a fairly conserved and buried area, as demonstrated by Consurf with a score of 8, the mutation has also generated a

change at the molecular level which manifests itself as Loss of Relative solvent accessibility with a significant p-value (0.03), the average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2 native and mutant structures with an RMSD of 0.017.

rs587780347 (E237K). Glutamic acid is replaced by Lysine at position 237. This mutation has been predicted by 5 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein, the residue is placed in a buried area, the mutation will also generate several changes at the molecular level such as Altered Metal-binding, Altered Metal-binding, and Gain of Allosteric site at Y234, with a significant score of 0.769, the average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2native and mutant structures with an RMSD of 0.06.

rs587780347 (E236K). Glutamic acid is replaced by Lysine at position 236. This mutation has been predicted by 5 tools as damaging, and its occurrence will cause a decrease in the stability of the GCK protein. The residue is located on the surface of the protein and is a highly conserved area with a conservation score of 9 which is very significant, the mutation will also generate several changes at the molecular level such as Gain of Allosteric site at E236, Gain of Relative solvent accessibility, Loss of Catalytic site at E236 and Altered Ordered interface, the average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2 native and mutant structures with an RMSD of 0.006.

rs587780346 (T206P). Threonine is replaced by Proline at position 206. This mutation has been predicted by 5 tools as damaging, and that its occurrence will cause a decrease in the stability of the GCK protein. The residue is located on the surface of the protein and is a highly conserved area with a conservation score of 9 which is very significant, the mutation will also generate many changes at the molecular level which occurs in Altered Metal-binding, Gain of Allosteric site at N204, Altered Ordered interface, Loss of Helix, Gain of Catalytic site at N204, Altered DNA binding with a MutPred score of 0.939, the average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2 native and mutant structures with an RMSD of 0.016.

rs193922331 (S263P). Serine is replaced by Proline at position 263. This mutation has been predicted by 3 tools as damaging and that its occurrence will cause a decrease in the stability of the protein, it is placed in an exposed and average area with a score of 5 as it generated molecular changes according to the Mutpred tool; Loss of Strand and an Altered Metal-binding; with a score of 0.662. The average RMSD raised from the two tools used TM-align and Pymol showed a slight not overlap between the 2native and mutant structures with an RMSD of 0.01.

After the generation of the mutant models by homology with the SWISS-MODEL tool, we had to ensure the quality of our models by using several quality verifications tools (Ramachandran Plot, QMEAN, and GMQE score, Errat and Verify 3d). These tools confirmed the validity of our models and that they are within the required standards.

The interactions of the GCK protein with other proteins were predicted with the online software STRING for associated gene screening, a total of 11 nodes were involved in the protein-protein interaction network (PPI), as shown in the figure below with an enriched PPI p-value of 1.2e-06 and an average local clustering coefficient of 0.84. Thus the result showed 11 main genes evaluated by the degree of connectivity in the PPI network which are GPI (glycolytic enzyme); GCKR (inhibits glucokinase by forming an inactive complex with this enzyme); G6PD (the main function of this enzyme is to provide reducing power (NADPH)); G6PC (forms with the glucose-6-phosphate transporter (SLC37A4/ G6PT) the complex responsible for the production of glucose by glycogenolysis and gluconeogenesis); PGM1 (this enzyme participates in both the degradation and synthesis of glucose); PDX1 (particularly involved in glucose-dependent regulation of insulin gene transcription); INS (which decreases the concentration of glucose in the blood); H6PD (oxidizes glucose-6-phosphate and glucose); TREH (intestinal trehalase is probably involved in the hydrolysis of ingested trehalose, belongs to the glycosyl hydrolase 37 families); NeuroD1 (acts as a transcriptional activator). Thus, the results show that the GCK protein has protein interactions that are particularly involved in carbohydrate metabolism. Any change in the structure and function of the protein can affect its ability to interact with other molecules. The STRING map showed the interaction of glucokinase with 11 different proteins, which was experimentally confirmed [50-54].

To analyze the topology and the structural similarity between the native model and the generated mutants the TM-align tool was used. The tool calculated a TM and RMSD score to measure the average distance between the skeletons of the superimposed proteins. The RMSD results for modelled mutants 1 and 4 were significant for pathogenicity considering that RMSD values above 0.15 are considered significant structural perturbations that could have functional implications for the protein. Eight nsSNPs out of 2783 nsSNPs on the GCK protein were considered the most pathogenic nsSNPs. All eight nsSNPs were included for further analysis (Table 10). The results showed that 7 of the 8 substitutions resulted in a highly significant degree of pathogenicity. Amino acid substitutions in GCK*: I130T, C371F, V182M, E236K, and T206P* had the highest cumulative score (CS = 9), predicted to be damaged by 9 out of 10 tools.

The nsSNPs on which we have worked, some of them have already undergone experimental studies that have treated the link between the mutation of the nsSNP and the appearance of the pathology, the thing that we have confirmed in silico, whereas the others have still not been studied experimentally for this, this study is going to be very interesting for the teams that want to work on this disease because it's going to allow them to narrow down the number of nsSNPs from 2783 to 7 so they're going to target the nsSNPs that have proven a high level of pathogenicity in *in silico* to prove that experimentally, and focus on those that are most likely to be damaged and involved.

CONCLUSION

The present *in silico* analysis of nsSNPs in the human GCK gene concluded that the mutations *I130T, C371F, V181M, V182M, E237K, E236K, and T206P* are the most deleterious nsSNPs among the gene variants. The seven nsSNPs were predicted to be damaging, located in fairly conserved regions, and also affecting the stability of the protein. Therefore, these mutations would probably affect the function of the GCK protein. Therefore, the results of this study confirm the previous findings and may be useful as a target for teams seeking to work further on this topic and to experimentally prove the link between these variants and gestational diabetes. Therefore, additional confirmation of the outcomes generated from the study is suggested through clinical and/or laboratory examinations.

ACKNOWLEDGMENTS

The authors are thankful to everyone who contributed to this work for providing encouragement and facilities.

Abbreviations

GCK: Glucokinase protein
PDB: Protein Data Bank
GD: Gestational Diabetes
HK: Hexokinase
nsSNPs: SNPs of non-synonymous coding
RMSD: Root mean square deviation
SNP: Single nucleotide polymorphisms

REFERENCES

[1] Rosik, J.; Szostak, B.; Machaj, F.; Pawlik, A., The Role of Genetics and Epigenetics in the Pathogenesis of Gestational Diabetes Mellitus. *Ann. Hum. Genet*. **2019**, *84*. https://doi.org/10.1111/ahg.12356.

[2] Ao, D.; Zhao, Q.; Song, J. -Y.; Liu, Z.; Wang, Y.; Wang, H. -J.; Yang, H. -X., The Association of the Glucokinase Rs4607517 Polymorphism with Gestational Diabetes Mellitus and Its Interaction with Sweets Consumption in Chinese Women. *Public Health Nutr.* **2021**, *24* (9), 2563-2569. https://doi.org/10.1017/S1368980020000609.

[3] Ramasammy, R.; Munisammy, L.; Sweta, K.; Selvakumar, S.; Velu, K.; Rani, J.; Kajalakshmy, S., Association between GCK Gene Polymorphism and Gestational Diabetes Mellitus and Its Pregnancy Outcomes. *Meta Gene*. **2021**, *28*, 100856. https://doi.org/10.1016/j.mgene.2021.100856.

[4] Markwardt, M. L.; Seckinger, K. M.; Rizzo, M. A., Regulation of Glucokinase by Intracellular Calcium Levels in Pancreatic β Cells. *J. Biol. Chem*. **2016**, *291* (6), 3000-3009. https://doi.org/10.1074/jbc.M115.692160.

[5] Su, J.; Zheng, N.; Li, Z.; Huangfu, N.; Mei, L.; Xu, X.; Zhang, L.; Chen, X., Association of GCK Gene DNA Methylation with the Risk of Clopidogrel Resistance in

Acute Coronary Syndrome Patients. *J. Clin. Lab. Anal.* **2019**, *34*. https://doi.org/10.1002/jcla.23040.

[6] Sternisha, S. M.; Miller, B. G., Molecular and Cellular Regulation of Human Glucokinase. *Arch. Biochem. Biophys.* **2019**, *663*, 199-213. https://doi.org/10.1016/j.abb.2019.01.011.

[7] Santoro, N.; Zhang, C. K.; Zhao, H.; Pakstis, A. J.; Kim, G.; Kursawe, R.; Dykas, D. J.; Bale, A. E.; Giannini, C.; Pierpont, B.; *et al.* Variant in the Glucokinase Regulatory Protein (GCKR) Gene Is Associated with Fatty Liver in Obese Children and Adolescents. *Hepatology*. **2012**, *55* (3), 781-789. https://doi.org/10.1002/hep.24806.

[8] Kim, C., Gestational Diabetes: Risks, Management, and Treatment Options. *Int. J. Womens Health*. **2010**, *2*, 339-351. https://doi.org/10.2147/IJWH.S13333.

[9] Capuano, M.; Garcia-Herrero, C. M.; Tinto, N.; Carluccio, C.; Capobianco, V.; Coto, I.; Cola, A.; Iafusco, D.; Franzese, A.; Zagari, A.; *et al*. Glucokinase (GCK) Mutations and Their Characterization in MODY2 Children of Southern Italy. *PLOS ONE*. **2012**, *7* (6), e38906. https://doi.org/10.1371/journal.pone.0038906.

[10] Costantini, S.; Malerba, G.; Contreas, G.; Corradi, M.; Marin Vargas, S. P.; Giorgetti, A.; Maffeis, C., Genetic and Bioinformatics Analysis of Four Novel GCK Missense Variants Detected in Caucasian Families with GCK-MODY Phenotype. *Clin. Genet.* **2015**, *87* (5), 440-447. https://doi.org/10.1111/cge.12406.

[11] Siddiqui, S.; Waghdhare, S.; Gopi, S.; Bhargava, A.; Panda, M.; Radha, V.; Mohan, V.; Dubey, S.; Jha, S., GCK Gene Screening and Association of GCK Variants With Gestational Diabetes in North Indian Population. *Clin. Med. Insights Endocrinol Diabetes*. **2018**, *11*, 1179551418806896. https://doi.org/10.1177/1179551418806896.

[12] Shaat, N.; Ekholm, E.; Lernmark, A.; Ivarsson, S.; Lynch, K.; Parikh, H.; Almgren, P.; Berntorp, K.; Groop, L., Erratum: Common Variants in MODY Genes Increase the Risk of Gestational Diabetes Mellitus. *Diabetologia*. **2006**, *49*, 2226-2227. https://doi.org/10.1007/s00125-006-0350-0.

[13] Calabrese, R.; Capriotti, E.; Fariselli, P.; Martelli, P. L.; Casadio, R., Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins. *Hum. Mutat*. **2009**, *30* (8), 1237-1244. https://doi.org/10.1002/humu.21047.

[14] Sim, N. -L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P. C., SIFT Web Server: Predicting Effects of Amino Acid Substitutions on Proteins. *Nucleic. Acids. Res.* **2012**, *40* (W1), W452-W457. https://doi.org/10.1093/nar/gks539.

[15] Adzhubei, I.; Jordan, D. M.; Sunyaev, S. R., Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet*. **2013**, *Chapter 7*, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76.

[16] Choi, Y.; Chan, A. P., PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels. *Bioinformatics*. **2015**, *31* (16), 2745-2747. https://doi.org/10.1093/bioinformatics/btv195.

[17] Capriotti, E.; Calabrese, R.; Casadio, R., Predicting the Insurgence of Human Genetic Diseases Associated to Single Point Protein Mutations with Support Vector Machines and Evolutionary Information. *Bioinformatics*. **2006**, *22* (22), 2729-2734. https://doi.org/10.1093/bioinformatics/btl423.

[18] Pejaver, V.; Urresti, R.; Lugo-Martinez, J.; Kymberleigh, A. P.; Ning Lin, G.; Nam, H.; Mort, M.; N. Cooper, D.; Sebat, J.; M. Iakoucheva, L.; *et al*. Inferring the Molecular and Phenotypic Impact of Amino Acid Variants with MutPred2|Nat. Commun. **2020**.

[19] Capriotti, E.; Fariselli, P.; Casadio, R., I-Mutant2.0: Predicting Stability Changes upon Mutation from the Protein Sequence or Structure. *Nucleic. Acids. Res.* **2005**, *33* (suppl_2), W306-W310. https://doi.org/10.1093/nar/gki375.

[20] Cheng, J.; Randall, A.; Baldi, P., Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins*. **2006**, *62* (4), 1125-1132. https://doi.org/10.1002/prot.20810.

[21] Ben Chorin, A.; Masrati, G.; Kessel, A.; Narunsky, A.; Sprinzak, J.; Lahav, S.; Ashkenazy, H.; Ben-Tal, N., ConSurf-DB: An Accessible Repository for the Evolutionary Conservation Patterns of the Majority of

PDB Proteins. *Protein. Sci*. **2020**, *29* (1), 258-267. https://doi.org/10.1002/pro.3779.

[22] Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; *et al*. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic. Acids. Res*. **2019**, *47* (D1), D607-D613. https://doi.org/10.1093/nar/gky1131.

[23] Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics*. **2006**, *22* (2), 195-201. https://doi.org/10.1093/bioinformatics/bti770.

[24] Benkert, P.; Tosatto, S. C. E.; Schomburg, D., QMEAN: A Comprehensive Scoring Function for Model Quality Assessment. *Proteins*. **2008**, *71* (1), 261-277. https://doi.org/10.1002/prot.21715.

[25] Lüthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of Protein Models with Three-Dimensional Profiles. *Nature* **1992**, *356* (6364), 83-85. https://doi.org/10.1038/356083a0.

[26] Colovos, C.; Yeates, T. O., Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein. Sci*. **1993**, *2* (9), 1511-1519. https://doi.org/10.1002/pro.5560020916.

[27] Hollingsworth, S. A.; Karplus, P. A., A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins. *Biomol. Concepts*. **2010**, 271-283. https://doi.org/10.1515/bmc.2010.022.

[28] Zhang, Y.; Skolnick, J., TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic. Acids. Res*. **2005**, *33* (7), 2302-2309. https://doi.org/10.1093/nar/gki524.

[29] Lua, R. C.; Lichtarge, O., PyETV: A PyMOL Evolutionary Trace Viewer to Analyze Functional Site Predictions in Protein Complexes. *Bioinformatics*. **2010**, *26* (23), 2981-2982. https://doi.org/10.1093/bioinformatics/btq566.

[30] Li, H.; Chang, Y. -Y.; Lee, J. Y.; Bahar, I.; Yang, L. -W., DynOmics: Dynamics of Structural Proteome and Beyond. *Nucleic. Acids. Res.* **2017**, *45* (W1), W374-W380. https://doi.org/10.1093/nar/gkx385.

[31] de Brevern, A. G.; Bornot, A.; Craveur, P.; Etchebest, C.; Gelly, J. -C., PredyFlexy: Flexibility and Local Structure Prediction from Sequence. *Nucleic. Acids. Res.* **2012**, *40* (W1), W317-W322. https://doi.org/10.1093/nar/gks482.

[32] AlGhamdi, N. A.; Alsuwat, H. S.; Borgio, J. F.; AbdulAzeez, S., Emerging of Composition Variations of SARS-CoV-2 Spike Protein and Human ACE2 Contribute to the Level of Infection: *In Silico* Approaches. *J. Biomol. Struct. Dyn*. **2020**, 1-12. https://doi.org/10.1080/07391102.2020.1841032.

[33] Mistri, M.; Tamhankar, P. M.; Sheth, F.; Sanghavi, D.; Kondurkar, P.; Patil, S.; Idicula-Thomas, S.; Gupta, S.; Sheth, J., Identification of Novel Mutations in HEXA Gene in Children Affected with Tay Sachs Disease from India. *PLoS One*. **2012**, *7* (6). https://doi.org/10.1371/journal.pone.0039122.

[34] Martínez, L., Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis. *PLOS ONE*. **2015**, *10* (3), e0119264. https://doi.org/10.1371/journal.pone.0119264.

[35] Saih, A.; Bouqdayr, M.; Baba, H.; Hamdi, S.; Moussamih, S.; Bennani, H.; Saile, R.; Wakrim, L.; Kettani, A., Computational Analysis of Missense Variants in the Human Transmembrane Protease Serine 2 (TMPRSS2) and SARS-CoV-2. *Biome. Res. Int*. **2021**, *2021*, 9982729. https://doi.org/10.1155/2021/9982729.

[36] George, D. C. P.; Chakraborty, C.; Haneef, S. A. S.; Nagasundaram, N.; Chen, L.; Zhu, H., Evolution- and Structure-Based Computational Strategy Reveals the Impact of Deleterious Missense Mutations on MODY 2 (Maturity-Onset Diabetes of the Young, Type 2). *Theranostics*. **2014**, *4* (4), 366-385. https://doi.org/10.7150/thno.7473.

[37] Freathy, R. M.; Mook-Kanamori, D. O.; Sovio, U.; Prokopenko, I.; Timpson, N. J.; Berry, D. J.; Warrington, N. M.; Widen, E.; Hottenga, J. J.; Kaakinen, M.; *et al*. Variants in ADCY5 and near CCNL1 Are Associated with Fetal Growth and Birth Weight. *Nat. Genet*. **2010**, *42* (5), 430-435. https://doi.org/10.1038/ng.567.

[38] Shaat, N.; Karlsson, E.; Lernmark, A.; Ivarsson, S.;

Lynch, K.; Parikh, H.; Almgren, P.; Berntorp, K.; Groop, L., Common Variants in MODY Genes Increase the Risk of Gestational Diabetes Mellitus. *Diabetologia*. **2006**, *49* (7), 1545-1551. https://doi.org/10.1007/s00125-006-0258-8.

[39] Lima Ferreira, J.; Voss, G.; Sá Couto, A.; Príncipe, R. M., Monogenic Diabetes Caused by GCK Gene Mutation Is Misdiagnosed as Gestational Diabetes - A Multicenter Study in Portugal. *Diabetes Metab. Syndr.: Clin. Res. Tour*. **2021**, *15* (5), 102259. https://doi.org/10.1016/j.dsx.2021.102259.

[40] Fan, R.; Wang, W. -J.; Zhong, Q. -L.; Duan, S. -W.; Xu, X. -T.; Hao, L. -M.; Zhao, J.; Zhang, L. -N., Aberrant Methylation of the GCK Gene Body Is Associated with the Risk of Essential Hypertension. *Mol. Med. Rep.* **2015**, *12* (2), 2390-2394. https://doi.org/10.3892/mmr.2015.3631.

[41] Hishida, A.; Takashima, N.; Turin, T. C.; Kawai, S.; Wakai, K.; Hamajima, N.; Hosono, S.; Nishida, Y.; Suzuki, S.; Nakahata, N.; *et al*. GCK, GCKR Polymorphisms and Risk of Chronic Kidney Disease in Japanese Individuals: Data from the J-MICC Study. *J. Nephrol*. **2014**, *27* (2), 143-149. https://doi.org/10.1007/s40620-013-0025-0.

[42] Zhang, J.; Li, C.; Shi, T.; Chen, K.; Shen, X.; Jiang, H., Lys169 of Human Glucokinase Is a Determinant for Glucose Phosphorylation: Implication for the Atomic Mechanism of Glucokinase Catalysis. *PLoS One*. **2009**, *4* (7), e6304. https://doi.org/10.1371/journal.pone.0006304.

[43] Gjesing, A. P.; Rui, G.; Lauenborg, J.; Have, C. T.; Hollensted, M.; Andersson, E.; Grarup, N.; Sun, J.; Quan, S.; Brandslund, I.; *et al*. High Prevalence of Diabetes-Predisposing Variants in MODY Genes Among Danish Women With Gestational Diabetes Mellitus. *J. Endocr. Soc*. **2017**, *1* (6), 681-690. https://doi.org/10.1210/js.2017-00040.

[44] Negahdar, M.; Aukrust, I.; Johansson, B. B.; Molnes, J.; Molven, A.; Matschinsky, F. M.; Søvik, O.; Kulkarni, R. N.; Flatmark, T.; Njølstad, P. R.; *et al*. GCK-MODY Diabetes Associated with Protein Misfolding, Cellular Self-Association and Degradation. *Biochim. Biophys. Acta*. **2012**, *1822* (11), 1705-1715. https://doi.org/10.1016/j.bbadis.2012.07.005.

[45] Sagen, J. V.; Odili, S.; Bjørkhaug, L.; Zelent, D.; Buettger, C.; Kwagh, J.; Stanley, C.; Dahl-Jørgensen, K.; de Beaufort, C.; Bell, G. I.; *et al*. From Clinicogenetic Studies of Maturity-Onset Diabetes of the Young to Unraveling Complex Mechanisms of Glucokinase Regulation. *Diabetes*. **2006**, *55* (6), 1713-1722. https://doi.org/10.2337/db05-1513.

[46] Gozlan, Y.; Tenenbaum, A.; Shalitin, S.; Lebenthal, Y.; Oron, T.; Cohen, O.; Phillip, M.; Gat-Yablonski, G., The Glucokinase Mutation p.T206P Is Common among MODY Patients of Jewish Ashkenazi Descent. *Pediatr. Diabetes*. **2012**, *13* (6), e14-e21. https://doi.org/10.1111/j.1399-5448.2011.00822.x.

[47] Cao, H.; Shorey, S.; Robinson, J.; Metzger, D. L.; Stewart, L.; Cummings, E.; Hegele, R. A., GCK and HNF1A Mutations in Canadian Families with Maturity Onset Diabetes of the Young (MODY). *Hum. Mutat.* **2002**, *20* (6), 478-479. https://doi.org/10.1002/humu.9090.

[48] Stoffel, M.; Froguel, P.; Takeda, J.; Zouali, H.; Vionnet, N.; Nishi, S.; Weber, I. T.; Harrison, R. W.; Pilkis, S. J.; Lesage, S., Human Glucokinase Gene: Isolation, Characterization, and Identification of Two Missense Mutations Linked to Early-Onset Non-Insulin-Dependent (Type 2) Diabetes Mellitus. *Proc. Natl. Acad. Sci. U S A*. **1992**, *89* (16), 7698-7702.

[49] Matsutani, A.; Janssen, R.; Donis-Keller, H.; Permutt, M. A., A Polymorphic (CA)n Repeat Element Maps the Human Glucokinase Gene (GCK) to Chromosome 7p. *Genomics*. **1992**, *12* (2), 319-325. https://doi.org/10.1016/0888-7543(92)90380-b.

[50] Hu, C.; Zhang, R.; Wang, C.; Yu, W.; Lu, J.; Ma, X.; Wang, J.; Jiang, F.; Tang, S.; Bao, Y.; *et al*. Effects of GCK, GCKR, G6PC2 and MTNR1B Variants on Glucose Metabolism and Insulin Secretion. *PLOS ONE*. **2010**, *5* (7), e11761. https://doi.org/10.1371/journal.pone.0011761.

[51] Vandercammen, A.; Van Schaftingen, E., The Mechanism by Which Rat Liver Glucokinase Is Inhibited by the Regulatory Protein. *Eur. J. Biochem*. **1990**, *191* (2), 483-489. https://doi.org/10.1111/j.1432-1033.1990.tb19147.x.

[52] Yamamoto, Y.; Miyatsuka, T.; Sasaki, S.; Miyashita, K.; Kubo, F.; Shimo, N.; Takebe, S.; Watada, H.;

Kaneto, H.; Matsuoka, T.; *et al.* Preserving Expression of Pdx1 Improves β-Cell Failure in Diabetic Mice. *Biochem. Biophys. Res. Commun*. **2017**, *483* (1), 418-424. https://doi.org/10.1016/j.bbrc.2016.12.128.

[53] Li, H.; Wang, X.; Mao, Y.; Hu, R.; Xu, W.; Lei, Z.; Zhou, N.; Jin, L.; Guo, T.; Li, Z.; *et al*. Long Term Liver Specific Glucokinase Gene Defect Induced Diabetic Cardiomyopathy by up Regulating NADPH Oxidase and down Regulating Insulin Receptor and P-AMPK. *Cardiovasc. Diabetol*. **2014**, *13*, 24. https://doi.org/10.1186/1475-2840-13-24.

[54] Tam, C. H. T.; Ma, R. C. W.; So, W. Y.; Wang, Y.; Lam, V. K. L.; Germer, S.; Martin, M.; Chan, J. C. N.; Ng, M. C. Y., Interaction Effect of Genetic Polymorphisms in Glucokinase (GCK) and Glucokinase Regulatory Protein (GCKR) on Metabolic Traits in Healthy Chinese Adults and Adolescents. *Diabetes*. **2009**, *58* (3), 765-769. https://doi.org/10.2337/db08-1277.