CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Identification of Biomarkers Associated with Prediabetic Insulin Resistance

by

Anam Ayub

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2021

I dedicate this thesis to my parents and my teachers.

CERTIFICATE OF APPROVAL

Identification of Biomarkers Associated with Prediabetic Insulin Resistance

by

Anam Ayub

(MBS191004)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Rehan Zafar Paracha	NUST, Islamabad
(b)	Internal Examiner	Dr. Arshia Amin Butt	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
April, 2021

Dr. Sahar Fazal
Head
Dept. of Biosciences & Bioinformatics
April, 2021

Dr. Muhammad Abdul Qadir
Dean
Faculty of Health & Life Sciences
April, 2021

Author's Declaration

I, **Anam Ayub** hereby state that my MS thesis titled "**Identification of Biomarkers Associated with Prediabetic Insulin Resistance**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Anam Ayub)

Registration No: MBS191004

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Identification of Biomarkers Associated with Prediabetic Insulin Resistance**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been dully acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Anam Ayub)

Registration No: MBS191004

Acknowledgement

To begin with the best name of **Almighty Allah**, most gracious and merciful, whose bounties are unbounded, whose benevolence is everlasting, whose advantages are uncountable, whose being is eternal, whose mercy is unlimited, whose provisions are un-finishing and whose love is our existence, whose worship is our faith. Words appear too little to express my gratitude to Allah Almighty who has bestowed me with more than I deserve.

I would like to specifically express my sincere gratitude to Capital University of Science and Technology (CUST) Islamabad for providing me an opportunity to do MS Biosciences and achieve my aim to do better research. I would like to extend special thankfulness to my supervisor, **Dr. Syeda Marriam Bakhtiar**, for her regular aid, encouragement, and guidance throughout this thesis. Special gratitude to my friends and colleagues for assisting me and encouraging me.

In the end, I specific my gratitude to my dad and mom, and siblings for imparting me unfailing support and non-stop encouragement during my years of studies and writing thesis.

(Anam Ayub)

Abstract

prediabetes is defined a person with week fasting glucose or week glucose tolerance who are with high risk for type 2 diabetes mellitus. The patients with week fasting glucose or week glucose tolerance are at increased risk for developing type 2 diabetes mellitus, cardiovascular disease and many other distinct metabolic abnormalities. there are no of metabolic disease that is associated with Diabetes mellitus due to chronic high blood glucose levels Insulin resistance is a syndrome that rises no of other abnormalities and different physical results that present mostly in insulin resistant persons. In this study, we have used in-silico approach for identification of biomarkers in prediabetic insulin resistance. Firstly, we have collected some biomarkers from literature review and name them as set no 1 and the generate interaction network of set no 1 gene, in interaction network there are many genes that interact with set no 1 gene and noted the genes and name them set no 2, in interaction network many genes from set no 1 are not appear so we have total 40 different genes then enrichment analysis were done on all the genes to check their biological, molecular and cellular annotation. Through uniport the ids of all the genes and protein sequences were obtained. The genes that are collected by intraction network and literature mining were checked by PDB through BLASTP to check the same three-dimensional proteins structures and then ProSA server were used to check the energy profiles of all the modeled proteins. Through proSA server we have examined all proteins structure have negative Z score. Then protein protein doking were done through GRAMM-X web server and Cluspro, abd all the docked proteins were checked in discovery studio. Through docking we have examined that ARG13 was most repeated amino acid in all docked proteins. The molecular docking of set 1 biomarker and set 2 novel biomarkers was done to identify the most common interacting amino acids residues. It is observed that majority of the set 2 novel biomarkers which showed interactions with the set 1 biomarkers in gene interaction network. Interact with the same amino acid of their corresponding receptor, Therefore, during interaction of two sets of biomarkers, the novel biomarkers cause mutations in these amino acids which result in the

development of disease. And mutations in these amino acids due to interacting novel biomarkers result in the decrease stability of receptor gene.

Keywords: insulin resistance, prediabetes, biomarkers

Contents

List of Figures

List of Tables

Abbreviations

FFA	Free-Fatty Acid
FPG	Fasting Plasma Glucose
GDM	Gestational Diabetes Mellitus
IGF	Impaired Fasting Glycemia
IGT	Impaired Glucose tTolerance
IR	Insulin Receptor
IR	Insulin Resistance
LADA	Late-onset Autoimmune Diabetes of the Adult
MODY	Maturity-Onset Diabetes of the Young
NIDDM	Non-Insulin-Dependent Diabetes Mellitus
T1DM	Type 1 Diabetes Mellitus
T2DM	Type 2 Diabetes Mellitus
VLVD	Very Low-Density Lipoprotein

Chapter 1

Introduction

Diabetes mellitus is a complex metabolic disease that is characterized by consistently elevated glucose levels in the blood [1]. Due to rapid increase of this disease it is leading cause of global morbidity and mortality. Due to genetic and environmental factors there are several types of diabetes, and these are categorized according to their etiology. Most abundantly types of diabetes are type 1 diabetes Mellitus (T1DM), type 2 diabetes Mellitus (T2DM) [2] and Gestational Diabetes Mellitus (GDM)[3].maturity-onset diabetes of the young (MODY) and late-onset autoimmune diabetes of the adult (LADA) are the less common form of diabetes [4]. The most common form of diabetes that affecting worldwide is T2DM, about 95 % of the world's population have type 2 diabetes)[3]. A different condition marked by weekly activity and/or insulin secretion.

Type 2 diabetes mellitus (T2DM) is a varied metabolic disorder, that is due to three metabolic disorders: (i) week or irregular insulin secretion from the pancreas, (ii)due to cells that are resistant to insulin [5], and (iii) irregular glucose uptake in the visceral area [6]. Due to these defects in body different metabolic disorders produced in body includes; hyperglycemia, hyperinsulinemia – If insulin is least responsive to the cells, blood insulin and hyperlipidemia accumulation [7], and hyperlipidemia.This disease is usually produced by impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT) – also calles as prediabetes. When the blood glucose level is higher than normal but not enough higher that produced

type 2 diabetes referred as Pre-diabetes. Excessive urination, increased appetite, increased thirst, exhaustion, and weight reduction are the signs that lead to hyperglyceremia. Over time, the signs can include blurred vision, recurrent infection and slow wound cure. This is why not only Type 2 diabetes has all these signs [8]. There are also many individuals with asymptomatic disorders, i.e. where they have no of the above symptoms, therefore, early diagnosis of diabetes does not occur [9] And such persons are typically recognized by T2DM when T2DM presents itself or itself as permanent complications [10].

When normal and higher insulin level produced in the result of very week biological response then the condition is defined as insulin resistance [11]. naturally this refers to decreased sensitivity to insulin refereed glucose disposal [12]. Insulin resistance is a syndrome that rises no of other abnormalities and different physical results that present mostly in insulin resistant persons. All tissues of body give different response to insulin dependence and sensitivity. Given tissue differences in insulin dependence and sensitivity, the symptoms of insulin resistance syndrome are likely to indicate the composite effects of excess insulin and variable resistance to its behavior due to tissue variations in insulin dependency and sensitivity. Metabolic syndrome is the clinical diagnostic agency that distinguishes all individuals at high risk for insulin resistance-related (cardiovascular) morbidity [11] [12].

At the whole-body level, the insulin actions are influenced by the interaction of other hormones Insulin, though the dominant hormone driving metabolic processes in the fed state, acts in concert with growth hormone and IGF1; in response to insulin growth hormone is secreted, among other stimuli, preventing insulin-induced hypoglycemia. Other counter-regulatory hormones involved catecholamines, glucagon and glucocorticoids. These hormones drive metabolic processes in the fasting state. Glucagon promotes glycogenolysis, gluconeogenesis and ketogenesis. The ratio of insulin to glucagon describes the degree of dephosphorylation or phosphorylation of the related enzymes [13]. Glucocorticoids promote muscle catabolism, gluconeogenesis and lipolysis while Catecholamines promote lipolysis and glycogenolysis. When these hormones secrete in excess amount, they

may contribute to insulin resistance in particular locations, but does not account for the huge majority of insulin resistant conditions. In most cases Insulin resistance is believed to be apparent at the cellular level via post-receptor defects in insulin signaling. Despite promising findings for a number of laboratory animal insulin-signaling defects, they are currently unclear in their relevance for human insulin resistance. Possible mechanisms include down-regulation of insulin receptor tyrosine phosphorylation, delicate genetic polymorphisms, IRS protein, or GLUT 4 function abnormalities or PIP-3 kinase [14].

Glucose uptake is largely insulin dependent via GLUT 4 and is around 60 to 70% of all-body insulin mediated uptake [15]. Muscle glycogen synthesis is decreased In insulin resistance; this seems largely mediated by decreased intracellular glucose translocation[16]. In relation to protein turnover, a study did not show any difference between type 2 diabetic insulin resistants and controls, even if this was at the cost of hyperinsulinemia in this euglycemic clamp hyperinsulinemia study. The effects on adipose tissue are similar In insulin resistance, but the increased free fatty acid flux in liver tends to promote hepatic very low-density lipoprotein (VLDL) production[17]. Whereas the compensatory hyperinsulinemia normally proceeds to remove ketogenesis. In addition, because the activity of lipoprotein lipase depends on insulin and is hindered by insulin resistance, peripheral intake of VLDL triglycerides is also decreased. These pathways lead to insulin resistance hypertriglyceridemia [18].

Compensational hyperinsulinemia happens if pancreatic β-cell secretion rises to maintain normal blood glucose levels in muscle and adipose muscle resistance. Insulin resistance syndrome refers to the category of abnormalities and associated physical effects that more often occur in people with insulin resistance. Given the insulin dependence and sensitivity tissue differences, the manifestations of insulin resistance syndrome may indicate the composite effects of excess insulin and its variable effects [12][13]. Insulin resistance (IR) was identified as a type 2 diabetes (T2D) precursor [19][20] and cardiovascular disease [21][22][23]. In a number of conditions, including obesity, the IR and compensative hyperinsulinemia are normal. IR is a major pathophysiological cause of dysglycemia (decreased fasting

glycemia, IFG, and decreased glucose tolerance, IGT) and T2D When coupled with b-cell dysfunction [24]. IR disorders were also associated with high risk cardiovascular conditions (HCVs) such as hypertension, dyslipidemia and atherosclerosis [25][26]. However, we do not fully appreciate these comparisons.

In the majority of patients with reduced tolerance to glucose (IGT) or non-insulin-dependent diabetes mellitus (NIDDM), the resistance to insulin induced glucose uptake is present in -25% of people with normal oral tolerance to glucose. In such situations, degradation in glucose tolerance can be avoided only if the p-cell can increase its response to insulin confinement and preserve its chronic hyperinsulinemia status.

If this purpose is not accomplished, the glucose homeostasis is seriously decompensated. The relation between the resistance to insulin and glucose intolerance is greatly strengthened by changes in the concentration of plasma-free fatty acid (FFA) in the system. NIDDM patients are immune to the suppression of plasma FFA but plasma FFA levels can be lowered by relatively low increases in insulin. Therefore, increased levels of FFA plasma circulating can be prevented when large quantities of insulin are released.

Plasma FFA concentration will not be blocked normally, If hyperinsulinemia cannot be sustained, and the resulting increase in plasma FFA concentration will lead to increased hepatic glucose production. Since these events are occurring in people who are very Immune to insulin induced glucose absorption, the potential for severe rapid hyperglycemia under these conditions is evident even small increases in hepatic glucose output.

Although hyperinsulinemia may prevent the glucose homeostasis in insulin-resistant individuals from decompensating frankly, the endocrine pancreas responds in a compensatory manner not without expense. Insulin resistant hyperglycemia and hyperinsulinemia are treated or untreated in patients with hypertension [27]. Insulin resistance create many defects in body, and obesity being one of the major factors involved in the progress of insulin resistance [2]. The occurrence of Hyperglycemia in obese people is usually followed by this. Prediabetes is an acronym

for subjects with a higher risk of diabetes with glucose deficiency and/or glucose impairment. In both patients, there is an increased risk of developing diabetes mellitus or cardiovascular disease of type 2 [24]. Pradiabetes is a disorder in which people have elevated blood glucose or A1c levels than average, but not as high as normal. Studies show that prediabetic people who lose weight and increase physical activity can avoid or delay type 2 diabetes and restore normal blood glucose levels in some cases [28].

Prediabetes is a high-risk condition for diabetes and has a higher-than-average glycaemic component than the diabetes threshold. 5–10% Every year of people with prediabetes will progress to diabetes, Conversion to normoglycemia with the same proportion. The prevalence of prediabetes worldwide is growing, and experts expect above 470 million people to have prediabetes by 2030. Prediabetes is related to insulin resistance and β-cell dysfunctio at the same time — anomalies that occur before changes in glucose can be detected.

Observational evidence suggests that pre-diabetes and early nephropathy are associated with chronic kidney diseases, small fibrenopathy, diabetic retinopathy and rising risk of macrovascular diseases. The risk of diabetes prediction can be optimized by multifactorial risk assessments using non-invasive behavior and metabolic features based on the blood, as well as glycaemia values. Lifestyle improvement is a crucial factor in prevention of diabetes, with 40-70% relative risk reduction for prediabetic individuals. The collection of data also shows possible pharmacotherapy benefits [29].

Prediabetes is normally a high-risk condition for diabetes development, as defined as blood glucose levels that are higher than average, but lower than diabetes thresholds. The criteria that are used to Diagnose for prediabetes have changed with the time and vary dependent on the institution of origin. With reference to WHO, People who have one of two distinct states are at high risk of developing diabetes: IFG (impaired fasting glucose), defined as a FPG (fasting plasma glucose) concentration of $\geq$6·1 and $<$ 7·0 mmol/L, without IGT (impaired glucose tolerance); and IGT, defined as an FPG concentration of $<$ 7·0 mmol/L and a

2h post load plasma glucose concentration of ≥7·8 and <11·1 mmol/L, measured during a 75 g OGTT (oral glucose tolerance test)[30]. The American Diabetes Association (ADA) has a lower IFG cutoff value (FPG 5·6–6·9 mmol/L) that has been added to IGT with 5·7–6·4 percent glycated hemoglobin A1c (HbA1c) as a new category for high-risk individuals [31].

Any identified pre-diabetes that is restricted to IGT or IFG does not include any other risk factors for diabetes such as type 2 diabetes family history or metabolic syndrome. Another critical comment on the word prediabetes is that the many subjects with either IFG or IGT will not development to type 2 diabetes. That is why, another name might be better.

Up to now the WHO taskforce has not implemented "intermediate hyperglycemia" [30]. Another probable name is "border line diabetes, but currently this term has no formal meaning and is not recommended. This analysis uses the word "prediabetes," but acknowledges that it is not generally acceptable or predicts the conversion to diabetes at all times [32].

Both insulin resistance and b-cell dysfunction function have been observed during prediabetes even during early stages and are needed for the majority of prediabetes hyperglycemia. The earliest abnormality appears to be insulin resistance (IR), although in studies which investigate the course of pre-diabetes pathogenicity a substantial heterogeneity between individuals and populations exists.

This is no surprise; studies of genome-wide sequencing studies (GWAS) have shown that at least 60 genes confer a risk of type 2 diabetes growth, most of them linked to b-cell biology. Researchers recorded IR and early sporadic hyperglycemia about 13 years prior to diagnosis. Blood glucose was close to normal possibly due to compensatory mechanisms to improve pancreatic beta cell insulin development, before they were defeated 2-6 years prior to diagnosis at a time when there was more sustained hyperglycemia that also had to do with declining IR. few years later, with now clear hyperglycemia, a steeper decrease in the pancreatic b-cell function appeared to lead to a clinical analysis of diabetes. But what causes the plasma glucose levels to increase in precisely both fasting and fed condition?

The pathophysiology of IFG and IGT was investigated through euglycemic hyperinsulinemia clamping studies. Their data indicate that the isolated IFG (normal post meal blood glucose [BG]), but not any evidence of hepatic IR, was associated with increased glucose (i.e., glycogenolysis was normally suppressed). Those with a combined IFG/IGT, however, had combined increased gluconeogenesis, insulin-reduced glycogenolysis failure, and extrahepatic IR decreased glucose disposal [33].

Biomarkers are biological molecules in biological fluid (serum, urine, salive), tissues, for example genes, proteins and metabolites) to assist in demonstrating normal or irregular processes in the body or the conditions of the disease [34]. Furthermore, they are very significant for the early discovery of diseases [34].

A significant proportion have struggled to resolve the clinical trials for lack of sensitivity, specificity and reproducibility despite numerous studies to classify a great number of biomarkers. That is why, before sending it for clinical trials it is important that biomarkers to be appropriately validated first. In a clinical setting there are three types of biomarkers: (i) diagnostic (used for identification of identification), (ii) prognostic (for predicted result or progress of a disease), and (iii) theragnostic (for the identification of appropriate treatment) [35].

1.1 Aims and Objectives

Insulin resistance result is onset of diabetes. Therefore, if we can identify or diagnose insulin resistance at its earlier stage we can stop or reduce the progression of diabetes. This study is designed with following objectives:

1. Identification of biomarkers linked to prediabetic insulin resistance.
2. To Identify novel biomarkers associated with prediabetic insulin resistance through protein proteins interaction network
3. Identification of common interacting amino acid residue between biomarkers

1.2 Scope of study

The biomarker detection in this study may provide valuable information for the prognosis of diseases, in predicting of response to therapies, antagonistic events and drug interactions, and in establishing baseline risk. the development of new predictive, diagnostic and prognostic treatments. These biomarkers will also play a growing role in the development and evolution of new medicines.

Chapter 2

Literature Review

2.1 Diabetes

A group of metabolic diseases characterized by high glucose levels i.e. hyperglycemia is called diabetes mellitus (DM). Reasons for this condition are defects in insulin secretion or insulin action or may be both. Long term damage of various organ, non-functioning and total failure of vital organs i.e. kidnies, eyes, nerves and heart are associated conditions of hyperglycemia or DM [36].

Diabtese development consists of various pathogenic pathways. These pathways can range from autoimmune obiletration of the cells of insulin producing organ i.e. pancreas which give rise to insulin deficiency in the body. Due to this deficiency various abnormalities becomes resistant to insulin action. Abnormalities arises in many biochemicals such as carbohydrates, proteins and fatty acids' metabolism due to the deficient action of insulin on target tissues. insufficient action of insulin is dure to inadequate secretion of insulin or no response from tissues to produced insulin. This can happen at one or more different points in complex hormonal action pathways. Primary cause of hyperglycemia is still unknown as abnormality in either insulin secretion or action may coesxist in a patient or only one type also can be a cause of diabetes mellitus [37]. Dibetes mellitus can be divided into different types based on interaction between genetic or environmental factors

and is describes as per its etiology. Type 1 Diabetes Mellitus (T1DM), Type 2 Diabetes Mellitus (T2DM) [1] and Gestational Diabetes Mellitus (GDM) are the most common types of diabetes [38]. There are some other types of diabtes as well such as maturity-onset diabetes of the young (MODY) and late-onset autoimmune diabetes of the adult (LADA), but these are very less common [4]. One of the most common type of DM is T2DM which affects almost 95% of the whole world's population [38].

2.2 Outcomes of Diabetes

Diabetes mellitus (DM) is a very multifaceted metabolic disorder which is characterized by persistent very high glucose levels in the blood[1]. as per the International Diabetes Federation it is one of the biggest global health emergency in twenty first century [39]. In 2015, worldwide its occurrence was of one in eleven adults (1/11) and the projected occurrence of the impaired glucose tolerance was one in fifteen (1/15) adults. These ratios are expected to increase further, specifically in urban population.

These increasing indices lead to high medical and economic burden. Almost 12% of global health budget is being expended on diabtese worldwide [39]. Many studies are being conducted on various populations in this regard. A similar study directed recently in the Romanian population and it showed that diabetes mellitus is one of the main health care glitches of their medical system. its occurrence is 11.6% and the prediabetes' is 16.5%[40][41].

It is a foremost international health problem, which is currently affecting almost 246 million people around the world. In coming 30 years this number is expected to be doubled [42], Majority of the cases are of type 2 diabetes [43][44]. In diabetes some structural changes develops in the brain: cerebral atrophy and lacunar infarcts, blood flow changes of both hypo and hyper perfusion [41]. In diabetic patients, it is found that brain volumes are reduced which restricts to the hippocampus. while it is also observed that there is an inverse association between

glycemic control and hippocampal volumes. In some studies and experiments HbA1C is said to be only and major predictor of hippocampal volumes [41][45]. Type 1 diabetes is insulin dependent and people having DM1 need unlike and more complex management of their hyperglycemia as compared to DM2. Frequent monitoring and adjustment of insulin doses, maintenance of diet and exercise are the prerequisites of Type 1 diabetes management and control. This type of diabetes starts at a very early age [46].

Vascular lesions are also a complication of this disease. These lesions may develop into small and large blood vessels. This mechanism of forming vascular lesions is multifactorial. Activation of various chemical pathways such as protein kinase C, accumulation of sorbitol depletion of myoinositol is due to derangements in vessels because of high glucose influx. This give rise to vascular lessions [47][48]. Microvasular complications also includes coronary heart diseases, peripheral vascular diseases, retinopathy, nephropathy, stroke and erectile dysfunction etc. There is a continuous relationship between glycemic controls and the incidence and progression of microvascular complications [49]. There are some unmodifiable risk factors for type II diabetes mellitus which include age, ethnicity, family history (genetic predisposition), history of gestational diabetes, and low birth weight. Its incidence and frequency increases with age. In 2005, it was reported by the CDCP (Centers for Disease Control and Prevention) that the incidence of diabetes between people of aged 20 years or elder was 20.6 million (9.6% of the people of respective age group). It is also estimated that prevalence of diabetes increases with age (10.3 million people aged 60 years, elder or 20.9% of respective age group, had diabetes) [28].

There are many associated risk factors of diabetes and it can disturb many systems in the body with time which is also a cause of serious complications. These c complications can be divided into micro and macro vascular. neuropathy (Nervous system damage), nephropathy (renal system damage) and retinopathy (eye damage)grouped into microvascular damage or complications [32]. While cardiovascular diseases, stroke, and peripheral vascular disease lie in macrovascular group. Peripheral vascular disease may cause bruises and injuries that do not heal

and lead to gangrene and amputation. NHANES data from 1999 to 2004 shows that occurrence of microvascular complications is much more higher than that of macrovascular complications) [50]. These complications may be episodic (eg foot ulcers or infections) which can be treated and re occur many times. These usually begin slightly, but with time results in prolonged damage to the organ leading to complete non-functioning of organs. Other complications comprise dental disease, deceased resistance to various infections i.e influenza and pneumonia, and other birth related difficulties in pregnant women having diabetes. Type I and type II diabetes have similar complications to some extent but frequency and age of occurrence may vary accordingly.

2.3 Diagnosis and Treatment

Management of diabetes is to keep levels of blood sugar close to normal as possible. This control should be achieved without causing level of blood glucose to be low. This can be achieved typically by life style changes such as dietary modifications, weight control, exercise and walk and regular use of appropriate medication.

To actively participate in management and treatment of disease. Knowledge of the disease is most important thing. As associated complications and risk factors are very uncommon and less severe in patients who have ample knowledge of the disease and manage their blood glucose levels very well [51][20].

As per the American College of Physicians, 7-8% level of HbA1c is the goal of treatment of diabetes [52]. There are some other risk factors which can accelerate the negative effects of the disease. These are smoking, hypertension, obesity, metabolic disorders, stress and lack of proper exercise. Special footwear is widely used to reduce the risk of ulcers on diabetic feet although the evidence for its effectiveness remains the same [53]. There are many tests to diagnose diabetes. The current diagnostic methods depend only on blood glucose levels monitoring, and are invasive as they all require blood samples. ADA (The American Diabetes Association)prescribe that people having no symptoms should be tested for T2DM

by one of the following tests: OGTT (oral glucose tolerance) test, FPG (plasma glucose)test, and HbA1c (Hemoglobin A1c) test[54][10] these methods have been recently approved by the ADA (American Diabetes Association) in 2010 and by WHO (World Health Organization) in 2011. Random blood plasma levels are also used as tests but results are not as much reliable as other mentioned tests [55].

2.4 Insulin

A very important polypeptide hormone which regulates metabolism of carbohydrates is insulin. It is derived from a latin word "insula" which means "island". It is because this hormone is produced in the islets of langerhenns of pancreas [56]. Metabolism of fatty acids, protiens and carbohydrates is regulated by insulin and it promotes the absorption of glucose into muscle especially skeletal muscles, liver and blood [57].

The end product of this absorbed glucose is either glycogen which is formed by glycolysis or triglycerides which is accomplished by lipolysis. Glucose can be converted into both i.e. glycerides or glycogen in liver as well [57]. High levels of insulin control the level of glucose production and secretion from liver by usually inhibiting it [58].

In many tissues protein synthesis also affected by insulin which is circulating in blood. So, it is an anabolic hormone which promotes conversion of small molecules in the cells into large molecules in the blood. On contrary decreased levels of insulin in blood pose an opposite effect and promotes catabolism. Major catabolic process in reversal of body fats. Specialized cells are sensitive to glucose levels in blood. These cells are called Beta cells. These cells respond to high glucose levels in the blood and secrets insulin in response and this effect is reversed when level of glucose is low in blood. Insulin secretion is inhibited by low levels of glucose [59]. Insulin helps in transportation of blood glucose into the various cells of body where it is metabolized into required products to produce energy. Glucose concentration in blood is regulated in this way. When there is high levels of glucose in the blood,

insulin acts to increase reuptake of glucose by muscles cells and fat cells [56][60]. Chromosome 11p15 is the locus of INS gene and it is precusros of insulin [61].

In some mammals, such as mice, there are two insulin genes. one of these is a homologous of almost all mammals (Ins2), and the other is a retrospective copy that includes a sequence of stimulants but lacks an intron (Ins1). Both rodent insulin genes are active [62].

2.5 Structure and Function of insulin

The molecule of insulin consists of 51 aminoacids arranged in two polypeptide chains which are linked by disulphide bonds. Structure is shown in fig 1.1. these chains are called alpha and beta chain. The alpha or A chain has 21 residues of amino acids. It also has additional loop of disulphide bond between A 6 and A11 (fig 1.1). whereas the beta chain or B chain consists of 30 residues of amino acids.

Primary molecular structure of insulin is known and studies from than 50 species of animals [63][64]. Although during its working it acts as a single molecule i.e. monomer, while during its storage and biosynthesis it assembles itself into a dimer. It also sometimes assembles itself in hexamers in the presence of zinc. X-ray analysis of insulin insulin molecule has revealed all the three structures i.e. monomeric, dimeric and hexameric [65].

Insulin's crystal structure has been studied well and well documented. It describes the activity anf binding affinity of insulin to receptors. It has been studies extensively and further elaborations and researches are welcomed for insulin structure and structural activity relationship [66].

According to a study and review, downstream signaling of insulin receptors interacts with other signaling pathways of growth factors i.e. IGFI and IGF2 [67]. This approach is very helpful in identifying and demonstration of the importance of insulin receptor ligand agonists. Because this can reveal potential mimetic of insulins as therapeutic agents for the treatment of diabtese [66].

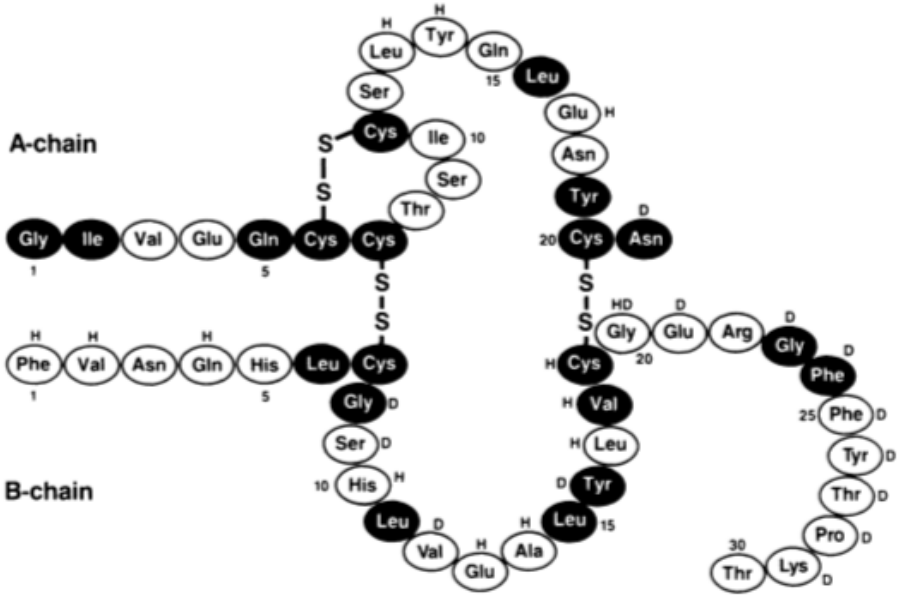

FIGURE 2.1: Human insulin's primary structure. The black residues are those amino acids which vary among different [63][64].

The monomeric 3D structure of insulin was first discovered through x-ray crystallography technique in 1926 [68]. As discussed in earlier section Insulin in humans consists of 51 amino acids and its molecular weight is 5808Da. The biologically active form of insulin which is circulating one has two monomeric chains. These chains are alpha and beta and consists of twenty-one and thirty amino acids respectively. These are linked by strong disulphide bonds at position A7-B7 and A20-B19.

At very low concentrations i.e. micromolecular concentrations it gets arranged in a very beautiful symmetric structure with the help of zinc. This structure is hexameric [66]. In insulin synthesis and activity, the alpha cells get cues from neighboring bets cells and secretes glucagon hormone in the blood in exactly opposite fashion [59]. This secretion is regulated as per glucose concentration in the blood i.e if it is low, the secretion will be high and vice versa. Glucagon acts to increase the concentration or level of glucose by stimulating biochemical processed in the liver. These processed may be gluconeogenesis or glycogenolysis [57][59]. Primary mechanism of glucose homeostasis is the secretion of glucagon and insulin in the blood as a result of blood glucose concentrations [59].

2.6 Biosynthesis of Insulin

The insulin contains 51 amino acids and have a molecular weight of 5.8kDa. However, the gene responsible for insulin encodes a precursor molecule known as preproinsulin which is mainly a 110 amino acid precursor. Similar to other secretary proteins, this insulin precursor has a hydrophic terminal which is signal peptide of N-terminal. This hydrophobic peptide interacts with signal recognition particles known as cytosolic ribonucleoproteins (SRP))[69]. Insulin is produced by beta cells within pancreas in mammals. Pancreas is primarily and exocrine gland and almost one to three million pancreatic cells (islets) form this endocrine part. This endocrine part is only 2% of the pancreatic mass. While within pancreas these beta cells accounts 65 to 80% of all cells [69].

Insulin is produced from beta cells in pancreas and has a very major role in regulating metabolism of various biochemicals i.e., carbohydrates and fats. It is synthesized as a polypeptide which is known as preproinsulin. This insulin precursor harbors a signal peptide which consists of 24 residues. This directs the budding of polypeptide from endoplasmic reticulum (ER) in the cell. Further this signal peptide is broken down into polypeptides which is translocated to the ER. There it forms proinsulin precursor. The proinsulin is further packed and folded in the proper structure and confirmation by disulphide bonds. This insulin is then transported across golgi apparatus network. In golgi complex this molecule is converted into active insulin molecule by action of endopeptidases prohormone convertor which is PC1, PC2 and exprotease carboxy peptidases. for the formation of C peptide, this molecule is broken down at two different positions. So, a result mature insulin is formed which contains to chains i.e. A chain and B chain. These chains are connected to each other by disuplphide linkages and A chain also has an interchain bond as well [70][71].

As discussed in earlier section, initially insulin is formed as a precursor in beta cells of pancreas islets. Short time after its synthesis and assembly in endoplasmic reticulums, this precursor of insulin in transported to complex glogi networsk within the cell. Granules are formed in trans glogi network (TGN) and these

are immature in nature. This process of transportation may take upto thirty minutes. The two chain insulin molecules further undergo maturation to active insulin molecule by the action of endopeptidases in the cell. These endopeptidases cut the C peptide from insulin molecule at a position of 64 and 65 having lysine and arginine respectively.

The second cite of cleavage is between 31 and 32 amino acid [72]. This mature insulin molecule is packed within mature granules which are waiting for further metabolic signals and nerve stimulation of vegal nerves for exocytosis from the cell in the blood circulation. Many signaling mechanisms involved in the regulation of insulin secretion from the beta cells of pancreas. Phosphorylation of glucose through glucokinases is a mechanism involved after transportation through transporters (Glut 2). To yield ATP (adenosine triphosphate) Glucose-6 phosphate undergoes a rate limiting step and stepwise metabolism followed by oxidative phosphorylation. Increased intracellular ATP concentration lowers the output of K1 and ultimately leads to beta cells depolarization. Voltage sensitive Ca2 1 channels are opened and this Ca2 is transported inside the cell. When levels of Ca2 are increased inside cell, active kinases gets stimulated and regulates insulin exocytosis and secretion into blood circulation [73].

2.7 Action of Insulin

Different types of tissues respond in a different way to insulin. Though tissue sensitivity to insulin associates with the levels of insulin receptors which are articulated on the plasma membrane. Now it has been clear that different components' assembly of insulin chemical signaling pathway is responsible for conferring specificity of insulin on target cells.

So, transport of insulin dependent glucose takes place only in adipose tissues and skeletal muscles. This is because these cells have insulin dependent transporter named as Glut4. Same as inhibition of glucose formation from glycogen with the help of insulin is specific to kidneys and liver. While on the other hand, effect on

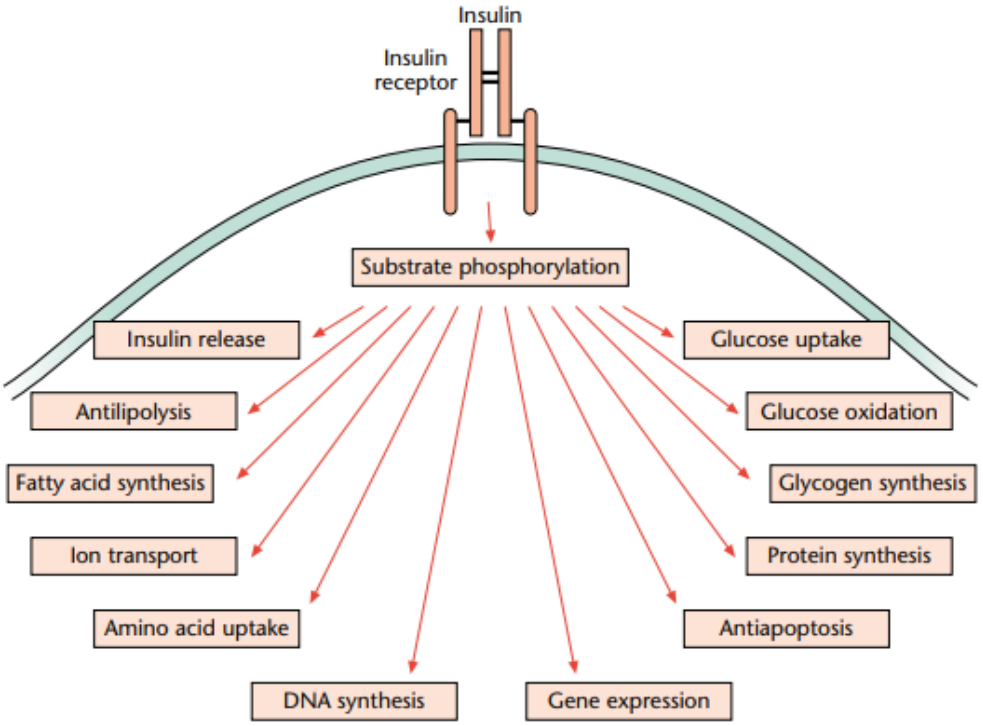

FIGURE 2.2: Action of insulin in cells. Multiple effects of insulin and mediation by receptor binding and tyrosine kinase [73].

ion transport. Synthesis of DNA and synthesis of proteins appears to be universal [73].

The Insulin employs multiple effects in the cell. Action of insulin is mediated by binding of molecule to its receptors. As a result phosphorylation of the receptors and other substrates by tyrosine kinase as shown in fig 2.2

2.8 Insulin Resistance

Insulin resistance a very well-known phenomenon and defined where normal or increased levels of insulin produces a weak biological response [11]. characteristically this refers to decreased sensitivity of insulin to insulin mediated glucose dumping[59]. Muscles usually account upto 60 to 70% of the entire body insulin dependent uptake of glucose. This uptake is done by GLUT4 transporters [15]. If muscular glycogen synthesis is reduced it is called insulin resistance. The resistance to insulin tends to be reduced with the translocation of intracellular glucose

[59]. As per previous studies, no difference has been reported between insulin resistant type II diabetics and control. Although this was due to hyperinsulinemia in hyperinsulinemic clamp study [17]. In resistance of insulin, the effects are similar on adipose tissues, whereas flux of free fatty acids promotes hepatic VLDL production (very low density lipoproteins [17] while ketogenesis usually remains repressed by the compensatory hyperinsulinemia. As lipoprotein lipases activity depends on insulin and dysfunctional by insulin resistance, the peripheral uptake of triglycerides from very low-density lipoproteins are also decreased. These biochemical mechanisms account for observed hyper triglyceridemic insulin [18].

Insulin resistance (IR) is said to be a precursor of type II diabetes mellitus (T2DM) and some other cardiovascular diseases[19][74] [21][22][11]. Insulin resistance and compensatory hyperinsulinemia commonly found in many different conditions which includes obesity as well. A syndrome is a group of various interlinked abnormalities os as insulin resistance which is related to physical consequences and very common in insulin resistant patients or individuals. Given the tissue difference between insulin dependence and sensitivity, insulin resistance syndrome is more likely to reflect a combination of insulin-dependent effects and variable action resistance. Metabolic syndrome represents the clinical analyst who identifies individuals at high risk for insulin-resistant (cardiovascular) disease [11] [12].

The relationship between obesity, insulin resistance, and type II diabetes has long been valued. a study has established that IMCL (intramyocellular lipid) content is a better forecaster of muscle insulin resistance than fat mass in hale and hearty, young, inactive and lean individuals [121][75] and nonobese, nondiabetic but insulin- resistant adults [76] and children[77], signifying a potential contributing role for IMCL mediator(s) in this process. Almost half century ago, Randle and colleagues in their study suggested that insulin resistance of muscles linked with obesity could be accredited to high fatty acid oxidation that bounds insulin-stimulated glucose utilization [77]. Higher delivery and oxidation of fat molecules leads to accumulation of phosphofructokinase inhibitors which are citrates is basic concepts of this model. Phosphofructokinase is very important enzyme which lead

to increase in intra myocellular glucose 6 phosphate and glucose. These ultimately impairs glucose uptake and utilization [77]. this has been illustrated in figure 2.3.

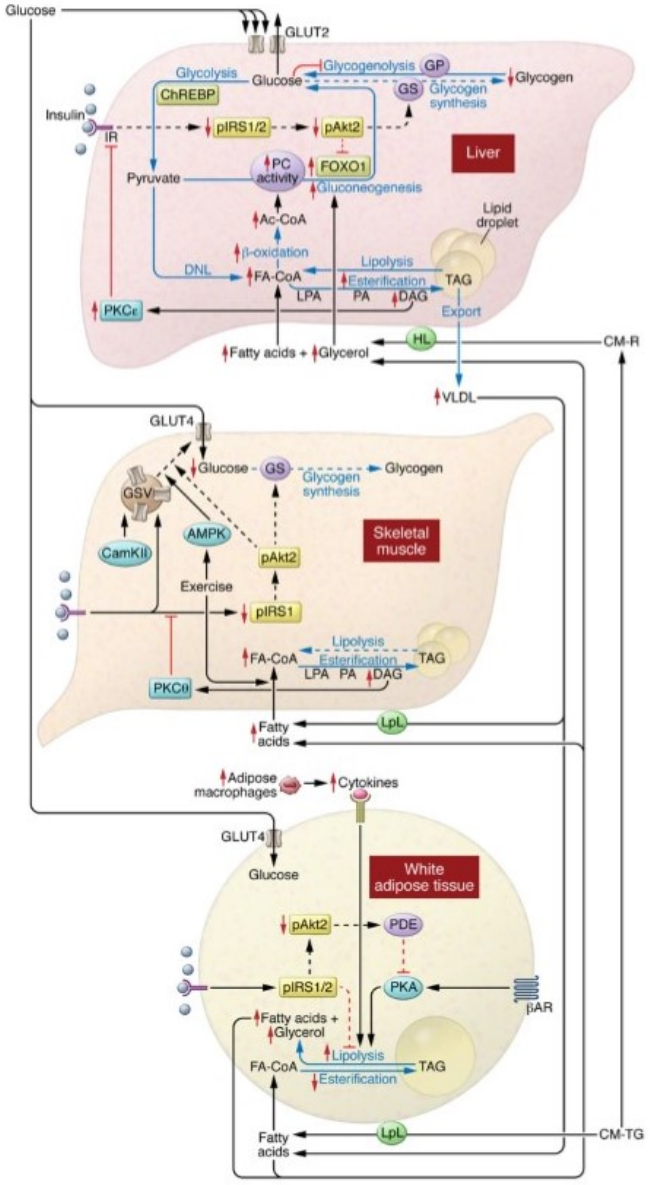

FIGURE 2.3: Insulin resistance mechanism.

DAG-mediated activation of PKCε in liver impairs hepatic insulin signaling, confining insulin-stimulated hepatic glycogen formation. Hepatic synthesis of lipids continues undiminished. In the skeletal muscles, DAG-mediated activation of PKCθ blights muscle insulin signaling, obstructing insulin-stimulated muscle glucose uptake and leads to higher levels of glucose delivery to the liver. Exercise

can still help to promote glucose uptake. In the adipose tissues, cytokine release from ATMs and promotes adipose lipolysis which leads to higher release of fatty acids (FAs). This further drive hepatic lipid synthesis and activate hepatic gluconeogenesis via acetyl-CoA–arbitrated (Ac-CoA–mediated) activation of PC and glycerol, increasing glucose production via substrate push. (IR :insulin receptor; GP:glycogen phosphorylase; GS:glycogen synthase; LPA(lysophosphatidic acid); HL (hepatic lipase); CM-R: chylomicron remnants; VLDL(very low–density lipoprotein); CAM-KII :calmodulin kinase II; IRS1/2: insulin receptor substrate 1/2; βAR: β-adrenergic receptor; CM-TG: chylomicron- triglycerides)[78].

2.9 Prediabetes

The term prediabetes denotes to lessened fasting glucose and/or impaired glucose tolerance in subjects who are at higher risk for T2DM (type II diabetes mellitus). Though both types of patients have an increased risk for evolving type 2 diabetes mellitus and cardiovascular diseases, they evident distinguished metabolic irregularities[2]. Glycemic variables that are higher than that of normal values but lower than threshold value is increased risk for onset of diabtese and this is termed as prediabetes (intermediate hyperglycemia) The ratio of prediabetic to convert to diabetics is 5–10% per year. This is same as changing back to normoglycemia. Existence of prediabetes is increasing globally and as per specialists it is projected to be more than 470 million people by end of 2030 [29].

Table 2.1: A comparison of normal, pre-diabetes and diabetes grounded on three diagnostic methods [79].

	FPG (Fasting plasma glucose)	PG in OGTT (Plasma glucose in oral glucose tolerance test)	A1C
Normal values	<100mg/dL or 5.5mmol/L	<140mg/dL or 7.8mmol/L	<5.7% or 39mmol/mol

Table 2.1 continued from previous page

	FPG (Fasting plasma glucose)	PG in OGTT (Plasma glucose in oral glucose tolerance test)	A1C
Pre-diabetes values	$\geq$ 100mg/dL or 5.5mmol/L	$\geq$ 140mg/dL or 7.8 mmol/L	$\geq$ 5.7% OR 39mmol/mol
Diabetes values	$\geq$126mg/dL or 7.8 mmol/L	$\geq$200mg/dL or 11.1 mmol/L	$\geq$6.5% or 48mmol/mol

All three diagnostic tests have their pros and cons. Typically fasting plasma glucose level greater than or equal to 140mg/dL is a very specific but insensitive test for prediction of diabetes mellitus [79]. Almost all individuals with FPG $\geq$ 140 mg/dL will have plasma glucose in oral tolerance test equal to or greater than 200mg/dL . But a substantial portion of individuals (depending on the population) with PG in OGTT $\geq$ 200 mg/dL will not have an FPG $\geq$ 140 mg/dL.

Glycation of proteins is measured by HbA1C test and it indicates chronic hyperglycemia, but this test does not directly prove increased blood glucose levels [80]. As per WHO, if people have impaired fasting glucose or impaired glucose tolerance, they are increased risk of having diabetes.

Impaired fasting glucose (IFG) is defined as a fasting plasma glucose levels of $\geq$6·1 and <7·0 mmol/L, but without reduced glucose tolerance; and impaired glucose tolerance is defined as an fasting plasma glucose concentration of <7·0 mmol/L and a 2 hour post load plasma glucose concentration of $\geq$7·8 and <11·1 mmol/L. this definition is subject to measuring during a 75 g oral glucose tolerance test (OGTT).

Same threshold for impaired glucose tolerance but a lower cutoff for impaired fasting glucose is applied by the American Diabetes Association (ADA). ADA has presented glycated hemoglobin A1c (HbA1c) 5·7–6·4% as a new group for higher diabetes risk[31]. Any pre-diabetes definition limited to IGT and / or IFG does not

include other risk factors for diabetes, such as family history of type 2 diabetes or metabolism. Another reproach of the term pre-diabetes is that many individuals with IFG or IGT will not develop type 2 diabetes.

That is why, another name is preferred. So far the "intermediate hyperglycemia" of the WHO interim group has not been widely accepted.[30]. Another name for this is "border line diabetes", but this term is not presently suggested and has no prescribed definition [32]. In 2012 CDC (Centers for Disease Control) had estimated that 86 million people or 1/3 adults had diabetes in the US. However almost 90% of individuals did not know that they are diabetic.

In 2015, the International Diabetes Federation projected that the worldwide occurrence of impaired glucose tolerance (IGT) in adults was 318 million and anticipated to reach 482 million by 2040[81]. The annual development rate to diabetes is 5–10% [82]. Older individuals, having severe insulin resistance (IR), low insulin secretion, and other associated diabetes risk factors are more likely to progress [83].

2.10 Prediabetic Insulin Resistance

Both insulin resistance and b-cell dysfunction function have been observed during prediabetes even during early stages and are needed for the majority of prediabetes hyperglycemia. The earliest abnormality appears to be IR (insulin resistance), although in studies which investigate the course of pre-diabetes pathogenicity a substantial heterogeneity between individuals and populations exists.

This is no surprise; studies of genome-wide sequencing studies (GWAS) have shown that at least 60 genes confer a risk of type 2 diabetes growth, most of them linked to b-cell biology. Researchers have reported signs of IR and early intermittent hyperglycemia approximately 13 years earlier the diagnosis of diabetes.

Blood glucose was close to normal possibly due to compensatory mechanisms to improve pancreatic beta cell insulin development, before they were defeated 2-6

years prior to diagnosis at a time when there was more sustained hyperglycemia that also had to do with declining IR. A few years later, with now clear hyperglycemia, a steeper decrease in the pancreatic b-cell function appeared to lead to a clinical diagnosis of diabetes. But what causes the plasma glucose levels to increase in precisely both fasting and fed condition?

The pathophysiology of IFG and IGT was investigated through euglycemic hyperinsulinemia clamping studies. Their data indicate that the isolated IFG (normal post meal blood glucose [BG]), but not any evidence of hepatic IR, was associated with increased glucose (i.e. glycogenolysis was normally suppressed). Those with a combined IFG/IGT, however, had combined increased gluconeogenesis, insulin-reduced glycogenolysis failure, and extrahepatic IR decreased glucose disposal [33].

2.11 Biomarkers

Biological molecules which could help in the indication of some normal or abnormal body processes or some specific disease are called biomarkers. These are present in body fluids or more specifically biological fluids i.e., saliva, serum and urine etc and mainly are genes, proteins and some metabolites of biochemicals [34]. Moreover, these biological molecules are also very important for diagnosis and early detection of abnormalities in the body which lead to disease [34].

Extensive research work has been done on identification of many biomarkers but there are some limitations due to which many researches have not extended beyond clinical trials. This is because the specificity, sensitivity and reproductivity are inadequate. So, before clinical trials it is of great importance that biomarkers should be properly validated.

Biomarkers used in clinical settings are of three types which are diagnostic biomarkers, prognostics and threputic biomarkers. Diagnostic are those which are used in disease identification while prognostic biomarkers are those which are used to predict outcomes of disease [35]. Some common clinically approved biomarkers are

Hcg (human chorionic gonadotropin) and PSA (prostate specific antigen). hCG is important biomarker of pregnancy and found in urine while PSA helps in the diagnosis of prostate cancer [84].

Despite the numerous biomarker advances, theragnostic biomarkers are still very limited in number [85]. Researched has suggested that these biomarkers can be helpful for accurate dosage, prediction of responses for a particular treatment, can maximize efficacy of drugs and minimize toxicity of drugs for relevant individuals [42].

They are a very long and difficult discovery. A good biomarker should be reproducible, sensitive, specific and must be standardized [34]. In recent years several approaches to identifying new biomarkers have been employed [85]. For the discovery and validation of new biomarkers, computational biology has played very important instrumental role in the detection [42]. .

2.12 Use of Biomarkers

as application of biomarkers can be as under;

- disease screening
- disease characterization (e.g. trinucleotide repeats)[86].
- To rule out, diagnose, staging, and monitor illnesses
- prognosis information
- individualize the therapeutic interferences by response monitoring responses to therapies and predicting consequences in response to them
- prediction of adverse drug reactions
- To predict and guide drug toxicity treatment (e.g., of serum concentrations measurements following medication overdose)

- cell types identification (e.g. histological markers). Use of biomarkers for optimization and individualization of doses of many drugs is not very welcoming approach. For example effects of warfarin on coagulation process is a complex model and includes two genetic biomarkers which produce very little therapeutic effect [87] . These are also used at numerous stages of drug discovery and development.

- it is also used in screening compounds as a target in drug discovery i.e. cyclooxygenase activity measurement for the identification of potential anti-inflammatory agents.

- Biomarkers also used as endpoints in pharmacodynamic studies, e.g. for prevention of heart diseases , serum cholesterol used as a marker for the action of a drugs and in pharmaco kinetic and pharmacodynamics studies[88].

2.13 Identification of Biomarkers

Good understanding with the pathophysiology of the disease is a basic and first step in identification of biomarkers. Factors associated with it also of basic importance. Therefore, a better understanding of any disease makes it workable and easy for identification of the various factors which may be associated with it.

Biomarkers associated with mechanism and etiology are best suited for prediction and proper diagnosis of the heart diseases in heart failure. This is also helpful for studying development and progression of the disease. The second step in biomarker identification is mechanism understanding, whereby intervention can affect the pathophysiology of disease [88]. Use of biomarkers is very useful for drug development process as well. It is used in studying different aspects of illness and monitoring the fruitful effects of applied interventions. However, as the events linking pathogenesis to outcomes are usually complex, the easier it is to recognize biomarkers that can diagnose a disease, track the reaction to a medication, and test disease progression are the more aware of the underlying abnormalities of

drug conditions and mechanics. For undeveloped and clinical pharmacologists as well as others involved in the identification of biomarkers accumulation of this information is a challenge [88].

2.14 Insilco Identification of Biomarkers

These tools are reported in literature for the following functions and after that we are selected the best tools for the specific function.

TABLE 2.2: Available tools for different analysis

Tool name	Function	Reference
Go term finder, David, Go term mapper	for gene annotation	[130][131][132]
Blast, FASTA, MUSCLE	For sequence alignment	[133][134][135]
Swiss model, Fold X, I TASSER, Hpred	For homology modeling	[136][137][123][138]
Saves, Verify 3D	For structure validation	[139][140]
CATH, Pock drug, p2rank	For active pocket	[141][142][142]
Autodock vina, Arguslab, ClusPro	For modeling and simulation	[143][144][141]
PLIP, Struct2net, Prise	For protein interactions	[145][146][147]
Webgivi, David, TopFun	For enrichment analysis	[134][131][148]

2.15 Biomarker Associated with Diabetes

Good understanding with the pathophysiology of the disease is a basic and first step in identification of biomarkers. Factors associated with it also of basic importance. Therefore, a better understanding of any disease makes it workable and easy for identification of the various factors which may be associated with it.

Table 2.3: Available list of biomarkers used in diagnosis and prognosis of diabetes [89].

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
Traditional biomarkers	Hba1c	There is β3 subunit portion of hemoglobin and attachment of glucose to its amino- terminal part produces Hba1c	Levels of Hba1c is a reflection of continuing glycemia	[90] [91]
	FA(Fructosamine)	It is a keto amine which is produced by glycosylation of all proteins of serum (primarily albumin)	Increased level of glucose concentrations directly affects FA level	[92]
	(GA) glycated albumin	It is produced by Albumin glycosylation is calculated by the GA fraction to total albumin	Serum concentration of GA is linked with diabetes and prediabetes condition	[93] [94]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	1,5 AG	It is a alimentary monosaccharide, and plasma absorptions of it is oppositely linked with plasma glucose level.	There are low plasma levels of 1,5 AG in persons with compromised glycemic controls in comparison with normoglycemic.	[95]
	OGTT	It is oral glucose tolerance and a direct measure of fasting and 2-hour post plasma glucose levels	Elevated levels are linked with prediabetes and diabetes.	[96]
Inflammatory biomarkers	CRP	It is derived from a primary marker of biosynthesis which is response of acute phase (il-6-dependent hepatic marker)	Directly linked to type II diabetes.	[97] [98]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	IL-6	It is an immunoregulator and involved in sustaining glucose balance and metabolism by its action on β3 cells of pancreas, adipocytes, hepatocytes, and skeletal muscles.	It is related with ype II diabetes and IR(insulin resistance).	[99] [100] [98]
	Wbcs	WBCs number is a immunity and inflammatory marker and it's an indicator of inflammation and vascular (micro and macro) complications in diabetes.	WBCs count is a predictive of impaired insulin action, T2DM, vascular heart diseases.	[101] [102] [103]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	Fibrinogen	Fibrinogen is responsible for blood thickness, platelet accumulation, and fibrin for clotting	It is associated with: prediabetes and feebly with diabetes. It is also related to increased 1-hour post levels of glucose and atherosclerosis	[104] [105]
	PAI-I	It is indicator of lowered fibrinolysis and its lowered activity leads to substandard coagulation147,148	PAI-I is an Independent/ predictor of diabetirs	[105] [106]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	IL-I8	IL-18 levels rise during hyperglycemia.	Directly associated with high risk of type II DM. it is also an indicator of progression from prediabetic state to diabetes.	[107]
	IL-IRA	It is an anti-inflammatory biomarker which rises when pathway of IL-1 is induced by glucose and free fatts via over eating.	Related with: •Its level rises in diabetes and prediabetic state and also associated with decreased sensitivity of insulin.	[108] [100]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
Novel biomarkers	Adiponectin	High insulin resistance (IR) and obesity are associated with lesser levels of adiponectin, while increased levels are observed in various prevention and intervention groups.	It comes from adipose tissues and it has anti-inflammatory, anti-therogenic and insulin sensitizing effects.	[109]
	FetA	Feta is associated with higher risk of developing type II DM with some other complications.	It is a glycoprotein of liver and it'sbeen projected to encourage lipid-induced IR through the signaling pathwayof TLR4, which results in production of various inflammatory cytokines	[110] [111]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	A-HB	α-HB is found to be remarkably linked with IR which does not depend on sex, age, body mass index, and collection site. It is related with decrease in glycine and serine. These are upstream of α-KB	It is an organic derivative which is produced through the combination of α-KB. α-KB is a product of amino\ acid catabolism.	[112] [113] [112]
	L-GPC	L-GPC is a negative indicator of type II DM progression.	it is a metabolite which is formed in the liver by the enzyme phospholipase A2 and by acyltransferase of lecithin cholesterol in the circulation.	[112] [114] [115]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	Lp(a)	There is an reverse relationship of Lp(a) with occurrence of prediabetic state and T2DM.	It is a lipoprotein that lead to atherogenesis.	[116] [117]
	HDL	HDL-C encourages insulin secretion and its low concentration may lead to development from prediabetes to diabetes.	It is a major lipoprotein.	[118]
	Ceramide	It is positively linked with prediabetes and type II DM.	Lipid molecule	[119]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	Ferritin and transferrin	It is strongly linked with higher risk of prediabetes and diabetes. Through radical formation, β3-cell oxidation stress and damage to DNA iron contributes to insulin resistance. Furthermore, catalytic iron excites the production of reactive oxidative species, hepatic dysfunction, and β3-cell apoptosis, and all these contribute to IR.	Ferritin is an intracellular protein which stores and control release of iron.	[120] [121]
	MBL -associated serine proteases	MASP1 is found to completely related with prediabetes, diabetes, and cardiovascular diseases.	These are important enzymes for activation of complement system (lectin pathway) and innate immune responses.	[98] [122]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	THBS1	THBS1 is positively linked with increased prediabetes prevalence, insulin resistance, metabolic dysregulation and adipose inflammation in T2DM and obesity.	THBS1 initiates formation of multiprotein multiplexes that modulate cellular phenotypes.	[122]
	GPLD1	GPLD1 is positively linked with Type II diabetes and prediabetic state and other novel biomarkers.	It has a assumed role in the pathway of glycosylphosphatid ylinositol compound for insulin-mimetic singling although exact mechanism is still unknown.	[122]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	MiRNA	Various miRNAs have been found elevated in individuals who are prediabetics. Mir-192 and 193b found to linked with IFG, IGT, triglycerides and fatty liver index.it also have role in regulation of insulin.	Various types of MiRNAs are involved in many cells signaling pathways and functions i.e. differentiation, growth, regulation of tumor proteins and gene expression which lead to higher oxygen and lower ATP formation, thus promotes insulin synthesis.	[127] [128] [129]

Table 2.3 continued from previous page

Category	Name of biomarkers	Mechanism of action	Association with deglycation	Ref
	Acylcarnitine	Prediabetic state associated inflammation and insulin resistance (IR) has increased levels of acyl-carnitine.	Acyl-carnitines interact with important inflammatory marker which is NF-Kβ3, and promote IR.	[123] [124] [125]

Chapter 3

Material and Methods

3.1 Extraction of Pre Diabetic Insulin Resistance Genes Through Text Mining

By literature review through manual process, 25 different genes were selected that were related to prediabetic insulin resistance. They play role in insulin resistance in human body. these genes were obtained through text mining were named as set 1

3.2 Identification of Novel Markers Linked to Already Known Genes Through Gene Interactions Networks

Fun coup 2.0: Genes network interactions were developed for the genes selected through literature review by fun coup. It was concluded that there are so many other genes they were interacted with the gene they were entered in fun coup server. fun coup database (http://FunCoup.sbc.su.se) is a bioinformatics tool that was used to generate the network interactions within genes and gene products it is also

used to create global network interactions[149]. Genes were noted and were used as novel genes (ENO2, PRDX1, ALDH2, PYGB, MDH2, ACTB, TUBB, EEF2, PGK1, RACK1, PGD, PGM1, VCP, HSPA8, EEF1A1, HSP90AB1, TUBB4B, ENO1, TALDO1, PSMD2, MDH2, ACSS2, VDAC2, HSPAL, SLC25A5, HSPD1, ATTP51B, GPI) as listed above. So now two sets are obtained. First was selected through literature review by manual method, and the other was selected by genes list that is interacted with the genes of first list in network interaction. biomarkers identified through gene interactions network were saved as set 2.

3.3 Enrichment Analysis

Enrichment analysis were done on the genes that were placed in set 1 and set 2. The enrichment analysis gives different outcomes like molecular, cellular and biological annotation of provided gene set. Tissue specificity of selected genes were also checked by enrichment analysis.

Measurement in the reciprocal scrutiny of large-scale genomic data functional relations between an experimentally results-induced gene or subject protein and a database of gene-protein sets.that is why a constant used method was to apply an over representation base enrichment analysis.

However, there are four benefits to this approach (1) It tests the functional overlapping gene/protein combination (2) it concerns the missing annotations gene (3) It considered the physical interaction network structure set between the gene/protein subjected set (4) The tissue-specific gene/protein interaction can be established [150].

3.4 Retrieval of Protein Sequences

the genes that were selected through manual process (set no 1) and the genes that were interacted with the set no 1 gene (set no 2) both protein sequences were

downloaded through uniport knowledgebase in FASTA (canonical) format. Then pasted all these sequences in word document.

uniport stands for universal protein resource that it gives us a stable, easily accessible and brief resource on protein sequences and functional annotation[151].Uniprot database was made or produced by three great bioinformatic institute and it was comprised of four different groups, UniProt Archive, the UniProt Knowledgebase, the UniProt Referenced Clusters and the UniProt Metagenomic and Environmental Sequence Database [152].

3.5 Gene Annotation

By using Software Tool for Researching Annotation of Proteins (STRAP) tool Both the set 1 and 2 genes were articulated, and by using uniport identifier [153] to check the enrichment analysis and the results were shown by this tool in almost 2-3 hours with respect to no of genes. The Uniprot IDs of all the proteins of set no 1 and set no 2 were saved. Then the uniport IDS were entered in STRAP tool to get annotation of proteins in tables form and number of GO charts of proteomic data.

3.6 Identification of Similar Protein Structure

The BLASTp was usually used because the protein sequence had collected by uniport. The protein sequences that had already collected by uniport past in query sequence box in fasta format and blast the query sequence this procedure shows the similarities with already present proteins sequences of different species in data base.

BLAST (blast.ncbi.nlm.nih.gov) is a program that compared primary biological sequences information like proteins, amino acids and nucleotides of DNA and ran. blast data base contains a vast number of proteins and nucleotide sequences of

different sequences. Through blast we can compare subject sequence with already present sequences and identified about the subject sequence [154]. NCBI was a big data base founded in 1988, this data base majorly works on biotechnology and biomedicine and it provided a lot of many other bioinformatic tools. Major data bases that work in ncbi is gene bank and pubmed (referenced data base for biological literature), blast (basic local alignment search tool), genome, snp, genes and proteins [155].

3.7 Domains Sequence and Structure Prediction

The different collected proteins were examined in the Protein data bank by BLASTP for the presence of comparable three-dimensional structures of protein. Then by using Robetta server the possible domains of each selected protein sequence were predicted and the domains structure predictions were made using the Robetta server [156]. The Robetta server gives ab initio along with comparative models of proteins domains.

3.8 Validation of Predicted Domains

To check the energy profiles of the modeled proteins were evaluated by using a ProSA server [54]. For understanding of biological process at molecular level the structural model of proteins availability was very important. proSA server was used to check the error at theoretical and experimental level in structural biological. ProSA server is basically used for validation of structure prediction and modeling of experimental protein structures. The selected genes proteins sequences that were present in PDb format are enter in this server and check and validate the structure of proteins one by one and note down the results.

Protein structure analysis is widely used to check 3D models of protein structures for potential errors. This server is widely used because it gives results within

seconds even for a large molecule [157]. It gives the Z score and energy values of all given protein one by one. The selected genes proteins sequences that were present in PDb format are enter in this server and check and validate the structure of proteins one by one and note down the results.

3.9 Cross Checking of Model's Validation

MetaMQAPII and phy 2 servers was also used to check the structures of all the proteins and they were also used for validation for 3d structure of given protein. The pdb files of proteins sequences are enter with the name of given protein. For the results the email id was also written in required box. These servers take time of few minutes or may be hours for results.

3.10 Docking and Identification of Common Interacting Amino Acids

Discovery studio is a software that were established for researchers for feigning small molecule and macromolecule systems. It was established and disseminated by Dassault Systemes BIOVIA. Discovery studio gives us no of different application e.g. Simulation, ligand design, pharmacophore modeling and structure-base design it were helpful in drug design and protein modeling research [161]. Docking results were analyzed in discovery studio, the interacting amino acids were identified between the docked complexes and surface analysis was performed to determine the accuracy of docking.

3.11 Protein-Protein Docking

In computational and experimental biology due to growing needs it required consistent computational procedure for checking of modeling and interactions between

proteins. A GRAMM-X web server [158] was used to check the interactions between one pair of proteins structures. In this server the ligand and receptor docking done and check the interactions between the protein sets. The set no 1 gene were used as receptors and set no 2 genes were used as the ligend, through this procedure the protein- protein docking were done. The results were collected through email. The docking results were also cross checked using Cluspro server to determine whether accurate docking has been done or not. Cluspro (an automated docking and discrimination method for the prediction of protein complexes).

In theoretical and computational biology, the predication of proteins structure was very important and challenging. Docking of proteins are very important to check the cellular function and organization. through cluspro server the protein-protein docking were done with good surface complementarity[159]. The cluspro server was widely use for protein-protein docking, it provides a simple page where two files were required in pdb format one was receptor and other one is ligand. This method gives no of different docked protein structures, model score of docked proteins[160].

3.12 Research Method

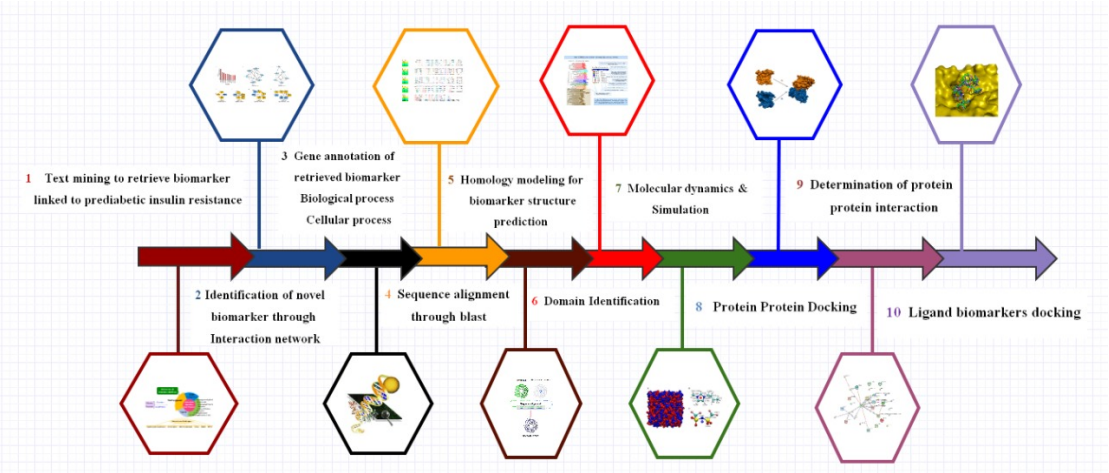

FIGURE 3.1: Research Methodology

Chapter 4

Result and Discussion

The biomarkers that is associated with prediabetes and insulin resistance are also collected through literature review by manual process and name them as set no 1 and the generate interaction network of set no 1 gene, in interaction network there are many genes that interact with set no 1 gene and noted the genes and name them set no 2, in interaction network many genes from set no 1 are not appear so we have total 40 different genes then enrichment analysis were done on all the genes to check their biological, molecular and cellular annotation. Through uniport the ids of all the genes and protein sequences were obtained. The genes that are collected by interaction network and literature mining were checked by PDB through BLASTP to check the same three-dimensional proteins structures and then ProSA server were used to check the energy profiles of all the modeled proteins.

Through proSA server we have examined all proteins structure have negative Z score. Then protein protein doking were done through GRAMM-X web server and Cluspro, abd all the docked proteins were checked in discovery studio. Through docking we have examined that ARG13 was most repeated amino acid in all docked proteins. The molecular docking of set 1 biomarker and set 2 novel biomarkers was done to identify the most common interacting amino acids residues. Total 40 different genes then enrichment analysis were done on all the genes to check their biological.

4.1 Extraction of Pre Diabetic Insulin Resistance Genes Through Text Mining

By literature review through manual process, 25 different genes were selected that are placed in table 1.1. there were many articles on pub med and google scholar that were reviewed and then selected these 25 genes that play role in prediabetes and insulin resistance.

Table 4.1: genes list from text mining

Sr.no	Gene Name	Reference
1	HNF4A(MODY1)	[162]
2	GCK(MODY2)	[163]
3	HNF1A(MODY3)	[164]
4	PKM	[165]
5	FASN	[165]
6	ACACA	[165]
7	PEPCK	[165]
8	INSR	[166]
9	IRS1	[167]
10	IRS2	[167]
11	GUT4	[166]
12	IGF1	[166]
13	GCKR	[166]
14	SIRT1	[168]
15	GAPDH	[168]
16	PKLR	[167]
17	SSTR2	[168]
18	FOS	[165]
19	GCK	[165]
20	APOA1	[165]
21	PCK2	[165]

Table 4.1 continued from previous page

Sr.no	Gene Name	Reference
22	G6PC	[165]
23	APOC3	[166]
24	GLP1R	[166]

In table 4.1 there are 24 different genes that are collected by literature review through manual process all of these genes are related to prediabetes ,diabetes and insulin resistance.

4.2 Identification of Novel Markers Linked to Already Known Genes Through Gene Interactions Networks

fun coup database (http://FunCoup.sbc.su.se) is a bioinformatics tool that is used to generate the network interactions within genes and gene products it is also used to create global network interactions[149]. genes network interactions were selected from literature review by fun coup. It was concluded that there are so many other genes that are interacted with the gene that were entered in fun coup server. These genes were placed in table 1.2 and used as novel genes. The interaction network of literature mining gene was shown in fig 4.1 there were no of different genes that interact with set no 1 gene and made an interaction. The highlighted genes were those which are present in set no 1 gene and made an interaction with other genes.

TABLE 4.2: genes list through interaction network

Sr.no	Gene Name	Sr.no	Gene Name
1	ENO2	16	YWHAE
2	PRDX1	17	HSP90AB1

Table 4.2 continued from previous page

Sr.no	Gene Name	Sr.no	Gene Name
3	ALDH2	18	TUBB4B
4	PYGB	19	ENO1
5	MDH2	20	NPEEPS
6	ACTB	21	TALDO1
7	TUBB	22	PSMD2
8	EEF2	23	MDH2
9	PGK1	24	ACSS2
10	RACK1	25	VDAC2
11	PGD	26	HSPAL
12	PGM1	27	SLC25A5
13	VCP	28	HSPD1
14	HSPA8	29	ATTP51B
15	EEF1A1	30	GPI

The gene that were made intraction network with set no 1 genes are placed in table 4.2. There were few genes from set 1 which were not present in interaction network so those genes were excluded.

4.3 Enrichment Analysis

Measuring functional relation between from an experimentally resultant gene or protein of subject and a database of recognized gene/protein sets in mutual tasks in the scrutiny of large scale functional genomic data.so for this a constant used approach is to apply an over representation base enrichment analysis [150]. Through enrichment analysis the gene annotation at three different levels like cellular level, molecular level and biological level were occur on the set 1 and set 2 genes. From the gene annotation different functions of genes at different level were checked.

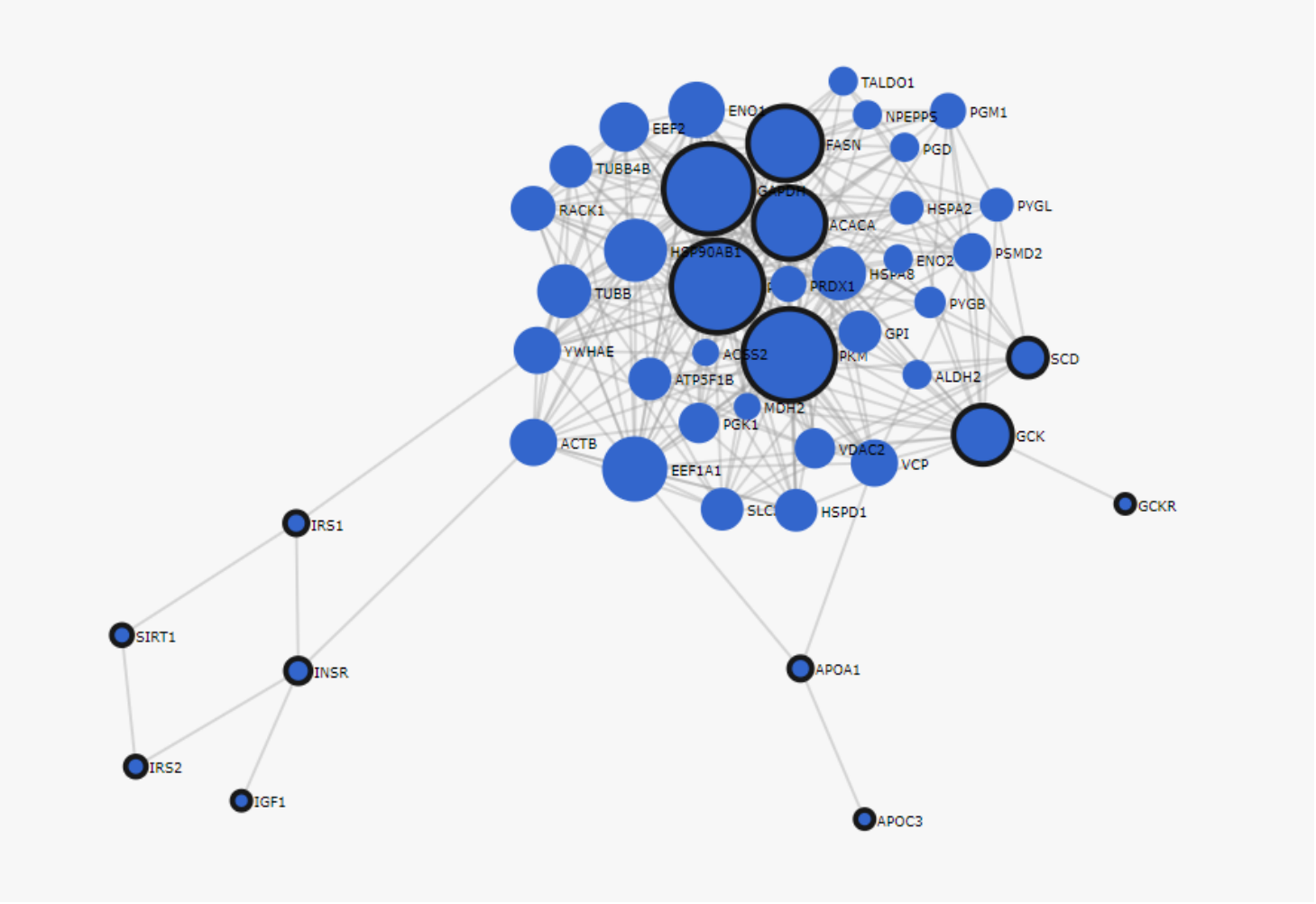

FIGURE 4.1: interaction network of genes by funcoup (http://FunCoup.sbc.su.se) database. The bold and highlighted circles show the interactions with other genes. Highlighted genes are those that are in set no 1 gene.

In table 4.3 molecular annotation through enrichment analysis were done in which XD score(these values in bold digits because they are significant) and overlapped genes were identified mostly overlapped genes are IRS1, IRS2 and INSR. Tissue specificity of all the genes from set no 1 and 2 were also checked by enrich net server that's XD- score was varies from 8.97 to 1.09 at different tissue types at different level. The other results like biological and cellular annotation were placed in supplementary data.

TABLE 4.3: Molecular Annotation through enrichment analysis

Annotation (pathway/ process)	**Significance of network distance distribution**	**Significance of overlap (Fisher -test, q-value)**	**Dataset size (uploaded gene set)**	**Dataset size (pathway gene set)**	**Dataset size (overlap)**
insulin -like growth factor receptor binding	2.0538	0.00041	20	13	IRS1 INSR IGF1 IRS2
phosphatidylinositol 3-kinase binding	1.3269	0.00081	20	20	IRS1 INSR
lipopolysaccharide	0.8020	1.00	19	11	HSPD1
ATPase activity, coupled	0.8020	1.00	19	11	HSPA8

AMP binding	0.7339	1.00	19	12	ACSS2
enzyme regulator activity	0.5839	1.00	19	15	PSMD2
cell surface binding	0.5133	1.00	19	17	HSPD1
oxidore ductase activity, acting on the alde hyde or oxo group of donors, NAD or NADP as acceptor	0.5133	1.00	19	17	ALDH2
kinesin binding	0.4575	1.00	19	19	ACTB

4.4 Retrieval of Protein Sequences for Both Set 1 and Set 2

Uniport server were used to checked the selected genes id and their protein sequences. uniport stands for universal protein resource that it gives us a stable,

easily accessible and brief resource on protein sequences and functional annotation[151].So the genes that were selected through manual process (set no 1) and the genes that were interacted with the set no 1 genes (set no 2) both protein sequences were downloaded through uniport knowledgebase in FASTA (canonical) format.

These gene ids were following(P35568,Q9Y4H2,P05019,Q14397,P35557,P14618, P49327,Q13085,P06213,P04406,P30613, P14618, P02647, Q16822, P02656, P09104, Q06830, P05091, P11216, P40926, P60709, Q7JJU6, P13639, P00558, P52209, P36871, P55072, P11142, P68104, P62258, P08238, P68371, P06733, P37837, Q13200, P40926, Q9NR19, P45880, P05141, P10809, Q92643)by clicking on gene ids the protein sequences were obtained Then pasted all these sequences in word document. The FASTA sequences of all the proteins were mention in supplementary data. These protein sequences were got through gene ids that is obtained by uniport Kb.

4.5 Gene Annotation

All the genes that were obtained from manual method and through interaction network were articulated by using Software Tool for Researching Annotation of Proteins (STRAP) tool.by uniport identifier (The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK)[4]. All differential proteins have been obtained by Uniprot ID. Then the uniport ides were inserted in Software Tool for Researching Annotation of Proteins tool to produce protein annotation tables and diversity of GO charts of proteomic data.

In tale 4.4 gene annotation were done in which gene ids, primary name, primary length and functions of gene were checked through STRAO Tool. Through strap tool there are no of different functions of a single cell were identified. GCK play role in regulating insulin; IRS1 mediated the role in cellular processers of insulin; IGF1 that is called as insulin like growth factors that are obtained from plasma and functionally play role in insulin; GCKR that play role in glucokinases;

Table 4.4: Gene Annotation through Strap Tool

Gene Name	**Uniprot Id**	**Name**	**Primary gene name**	**Length**	**Function**
L0HBA3	L0HBA3	Phospho transferase	GCK	51	Glucokinase (GCK) guides the release of insulin and the metabolism of the liver glucose.
P20823	P35568	Insulin receptor substrate 1	IRS1	1242	Can mediate insulin regulation of different cell processes.
Q9Y4H2	Q9Y4H2	Insulin receptor substrate 2	IRS2	1338	Regulation of different insulin cellular processes can be mediated.

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
P05019	P05019	Insulin-like growth factor I	IGF1		Insulin-like growth factors are structurally and functionally linked to insulin and are plasma isolated, but have dramatically increased activity to promote growth. Stimulates glucose delivery to osteoblastic bone-derived cells and operates at much lower levels than insulin.
Q14397	Q14397	Glucokinase regulatory protein	GCKR		Glucokinase (GCK) is regulated by forming an inactive enzyme complex.

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
	Q86Y46	Keratin, type II cytoskeletal 73	KRT73		Specific portion of intermediate keratin filaments in the hair follicle's internal root sheath (IRS).
P35557	P35557	Hexokinase-4	GCK		Phosphorylation to hexosis 6-phosphate catalyzed, including hexosis of D-glucose, D-fructose and D-mannose.

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
A7LFL1	A7LFL1	Phosphotransferase	GCK	465	Transfers sugars into the cell, including glucose, mannose and mannitol. Has a role in hair growth.
P14618	P14618	Pyruvate kinase PKM	PKM		Specific components of intermediary keratin filaments in the hair follicle's inner root sheath (IRS).

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
P49327	P49327	Fatty acid synthase	FASN		Fatty acid synthetase is a multifunctional enzyme that catalyze long-chain saturated fatty acid de novo biosynthesis in the presence of NADPH starting from the acetile-CoA and malonyl-CoA.
A0A0U1RQF0	A0A0U1RQF0	3-hydroxyacyl-[acyl-carrier-protein] dehydratase	FASN	2509	The multifunctional protein contains 7 catalysts.

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
Q13085	Q13085	Acetyl-CoA carboxylase 1	ACACA		Cytosolic enzyme which catalyzes acetyl-CoA carboxylation into malonyl-CoA.
Q86SK9	Q86SK9	Stearoyl-CoA desaturase 5	SCD5		Stearyl-CoA desaturase that 4uses O(2) and reduced cytochrome b5 electrons to introduce saturated fatty acyl-CoA substrates with the first double connections.

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
O00767	000767	Acyl-CoA desaturase	SCD	359	The role of lipid biosynthesis plays a significant role. The role of genes involved in lipogenesis is significant in regulating the expression of mitochondrial oxidation of fatty acid.
P35558	P35558	Phosphoenolpyruvate carboxykinase, c ytosolic [GTP]	PCK1		Cytosolic phosphoenolpyruvate carboxykinase which catalyzes the reversible decarboxylation and oxaloacetate (OAA) phosphorylation and functions as the gluconeogensis-restrictive enzyme

Table 4.4 continued from previous page

Gene Name	Uniprot Id	Name	Primary gene name	Length	Function
Q16822	Q16822	Phosphoenolpyruvate carboxykinase [GTP], mitochondrial	PCK2		A metabolism task that generates lactate glucose.
P06213	P06213	Insulin receptor	INSR		Tyrosine kinase receptor that mediates insulin pleiotropic activities.
P14672	P14672	Solute carrier family 2, facilitated glucose transporter member 4	SLC2A4		A glucose supplementary transporter regulated by insulin that plays a key role in blood circulation removal. Insulin response is 4 controlled by its intracellular role.

KRT73 play role in roots of hairs; hexokinases-4 play role in the formation of D-fructose and D mannose; phosphotransferase play role in transport sugar in cell. PKM pay role in hair growth; FASN important in the formation of long chains fatty acids from acetyl-CoA; 3-hydroxyacyl-dehydratase as acyl carrier this protein done 7 catalytic functions.

ACACA this is a cytosolic enzyme and performed function in rate limiting step; SCD5 it gives mixture of unsaturated fatty acids; PCK1 it perfume different function in low and high glucose level at low level it catalyze cataplerotic and at high level it catalyze anaplerotic;SLC2A4 play role in transport of glucose from extracellular milieu into cell.

4.6 Sequence Alignment to Find Similar Proteins in PDB

From NCBI the protein blast were done where the compression of primary biological information like proteins, amino acids and nucleotides of DNA and RNA. The blast gives us alot of proteins and nucleotide sequences of different sequences.

BAST (blast.ncbi.nlm.nih.gov) is a program that compared primary biological sequences information like proteins, amino acids and nucleotides of DNA and RNA. blast data base contains a vast number of proteins and nucleotide sequences of different sequences.

Through blast we can compare subject sequence with already present sequences and identified about the subject sequence [154].The different collected proteins were examined through the Protein data bank by BLASTP for the presence of similar three-dimensional structures of proteins. The pblast was usually used because the protein sequence of required genes were collected by uniprot. These protein sequences had already collected by uniport past in query sequence box in fasta format and blast the query sequence this procedure shows the similarities with already present proteins sequences of different species in data base.

TABLE 4.5: Multiple Sequence Alignment of all the Genes of set no 1 and 2

Gene name	**Description**	**Scientific name**	**Max score**	**Total score**	**Query cover**	**E value**	**Per ident**
GCK	Chain A, Glucokinase Isoform 3	Homo sapiens	102	102	100%	3e-28	98.04%
GCKR	Chain A, Glucokinase Regulatory Protein		1288	1288	100%	0.0	100.00%
IGF1	Chain A, INSULIN-LIKE GROWTH FACTOR-I	Homo sapiens	149	149	35%	1e-47	
IRS1	Chain A, INSULIN RECEPTOR SUBSTRATE 1	Homo sapiens	560	560	21%	0.0	100.00%
IRS2	Chain A, INSULIN RECEPTOR SUBSTRATE 1	Homo sapiens	320	320	20%	3e-100	60.44%
INSR	Chain A, Insulin receptor	Homo sapiens	2783	2783	98%	0.0	98.60%
PKM	Chain A, Pyruvate kinase M2	Homo sapiens	1097	1097	100%	0.0	100.00%

Table 4.5 continued from previous page

Gene name	**Description**	**Scientific name**	**Max score**	**Total score**	**Query cover**	**E value**	**Per ident**
PKLR	Chain A, Pyruvate kinase PKLR	Homo sapiens	1097	1097	94%	0.0	100.00%
ACACA	Chain A, Acetyl-CoA carboxylase 1	Homo sapiens	4888	4888	100%	0.0	99.96%
SITR2	Sirt2 in complex with a 13-mer trifluoro acetylated Ran peptide	Homo sapiens	736	736	91%	0.0	100.00%
PKM	Chain A, Pyruvate kinase M2	Homo sapiens	1097	1097	100%	0.0	100.00%

The sequence was run in blast to check the similarity of sequences with other genes. Lower the E(expected value) value means closer the result to query sequence, The E value is the number of hits that are supposed to occur when your query sequence is checked for random sequences in the database. expected value in table 4.5 all the query sequences have value 0 or less the o so the query sequence of proteins were present in pdb and identification of query sequences were almost 90-100%

in pdb. The remaining data from multiple sequence alignment were placed in supplementary data.

4.7 Domains Identification and their Structure Prediction

By using Robetta server the structure predictions were made using the de novo Rosetta method[156]. The Robetta server provide proteins domain both ab initio and comparative models.

To check the energy profiles of the displayed proteins were estimated by using a Protein structure analysis server [54]. For understanding of biological process at molecular level the structural model of proteins availability was very important. proSA server was used to check the error at theoretical and experimental level in structural biological. ProSA server were basically used for structure prediction and modeling of experimental protein structures.

ProSA was widely used to check 3D models of protein structures for potential errors. This server was widely used because it gives results within seconds even for a large molecule[157] It gives the Z score and energy values of all given protein one by one.the selected genes proteins sequences that were present in PDb formate are enter in this server and check and validate the structure of proteins one by one and note down the results.

The z score of all the genes of set no 1 and 2 were noted (ACACA -13.35, ACOD -1.21, APOA -5.07, APOC3 -0.37, FASN -11.78, G3P -9-9, GCK -8.84, GCKR -4.6, IGF1 -4.6, INSR -9.83, IRS1 -5.18, IRS2 -5.38, KPYM -10.89, PKLR -10.89, SIRT1 , ENO2 -5.34, PRDX1 -7.55, ALDH2 -9.85, PYGB -11.52, MDH2 -9.45, ACTB -10.09, TUBB -8.78, EEF2 -12.41, PGK1 -11.52, RACK1 -3.3, PGD -11.78, PGM1 -12.1, VCP -12.96, HSPA8 -8.72, EEF1A1 -9.95, YWHAE -6.8, HSP90AB1 -9.05, TUBB4B -8.78, ENO1 -10.34, TALDO1 -8.8, MDH2 -9.45, ACSS2 -9.4, VDAC2 -4.74, SLC25A5 -4.9, HSPD1 -10.25) As we know that the lower the Z

score higher the validity of structure of proteins, as all the structures Z score values were negative so the predicted structures of proteins were accurate.

MetaMQAPII and phy 2 servers was also used to check the structures of all the proteins and they were also used for validation for 3d structure of given protein. The pdb files of proteins sequences are enter with the name of given protein. For the results the email id was also written in required box.These servers take time of few minutes or may be hours for results. The Z score is also called as standard

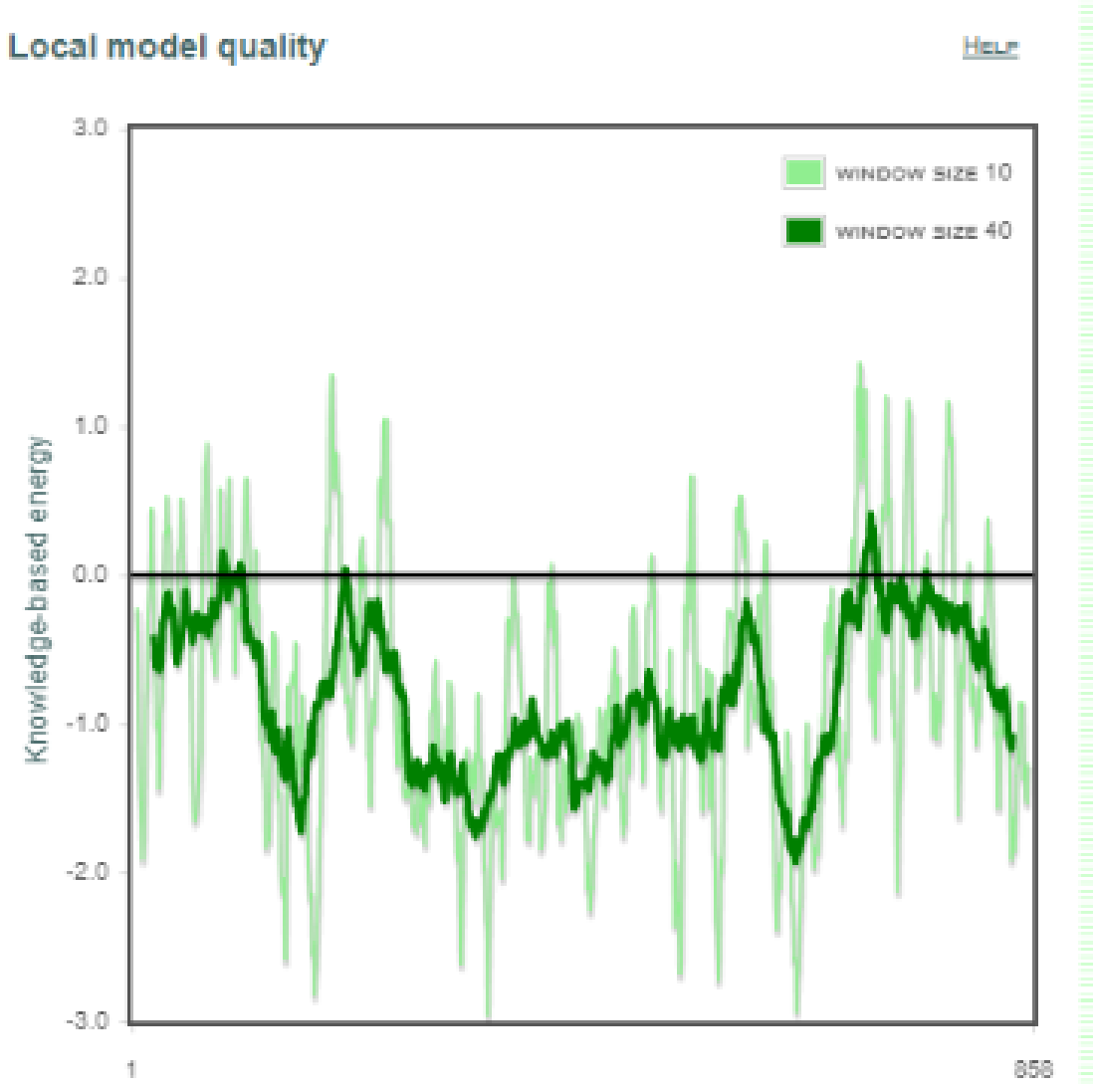

FIGURE 4.2: Fig: 4.1. The protein structure validation of EEF2 were represent in this graph and Z score value is -12.41. Z-score is a numerical measurement that gives a value's relationship to the mean of a group of values.

score is a very useful statistic because it (a) enables us to estimate the possibility of a score that occurs within our normal distribution and (b) Allows one to compare

two scores from the various usual distributions. In this graph the prediction level of protein structure were checked.

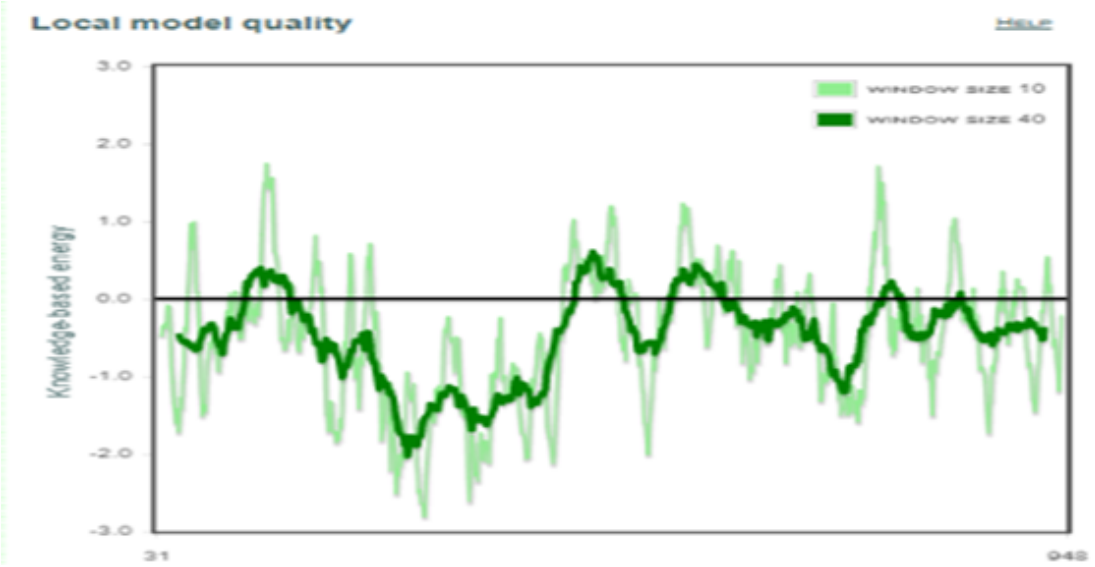

FIGURE 4.3: The protein structure validation of INSR were represent in this graph and Z score value is -9.83.

Two values were obtained one is negative and other is positive, if the z score is positive its means the value of a protein structure is above the mean and if the value is negative its mean the value of protein structure is below the mean. Through this also checked the minimum threshold level of query proteins. The remaining proteins graphs were present in supplementary data.

4.8 Protein-Protein Docking

In computational and experimental biology due to growing needs it required consistent computational procedure for checking of modeling and interactions between proteins. A GRAMM-X web server[158] was used to check the interactions between one pair of proteins structures. In this server the ligand and receptor docking done and check the interactions between the protein sets. The set no 1 gene were used as receptors and set no 2 genes were used as the ligend, through this procedure the protein- protein docking were done. The results were collected through email. Cluspro (an automated docking and discrimination method for the prediction of

protein complexes) was also used to cross validate the results of docking, both the tools provided similar docking reults. In theoretical and computational biology, the predication of proteins structure was very important and challenging. Docking of proteins are very important to check the cellular function and organization. through cluspro server the protein- protein docking were done with good surface complementarity [159]. The cluspro server was widely use for protein-protein docking, it provides a simple page where two files were required in pdb format one was receptor and other one is ligand. This method gives no of different docked protein structures, model score of docked proteins[160].

4.9 Analysis of Docking

Through the docking the interaction between ligand and protein were checked because docking results were giving the amino acid through which two proteins interact because amino acid is the functional unit of protein.

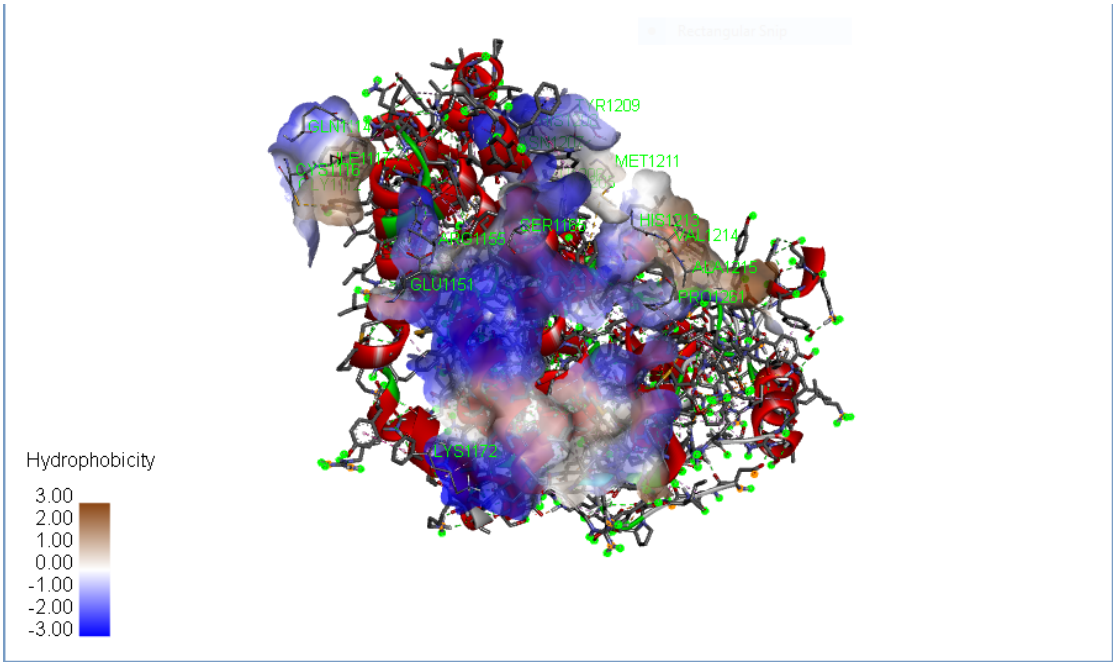

FIGURE 4.4: Docking result of ACACA and ENO1.The figure shows that there are many amino acids that help in interaction of two proteins. In the docking between ACACA-ENO1, ACACA is the receptor and ENO1 is ligand.

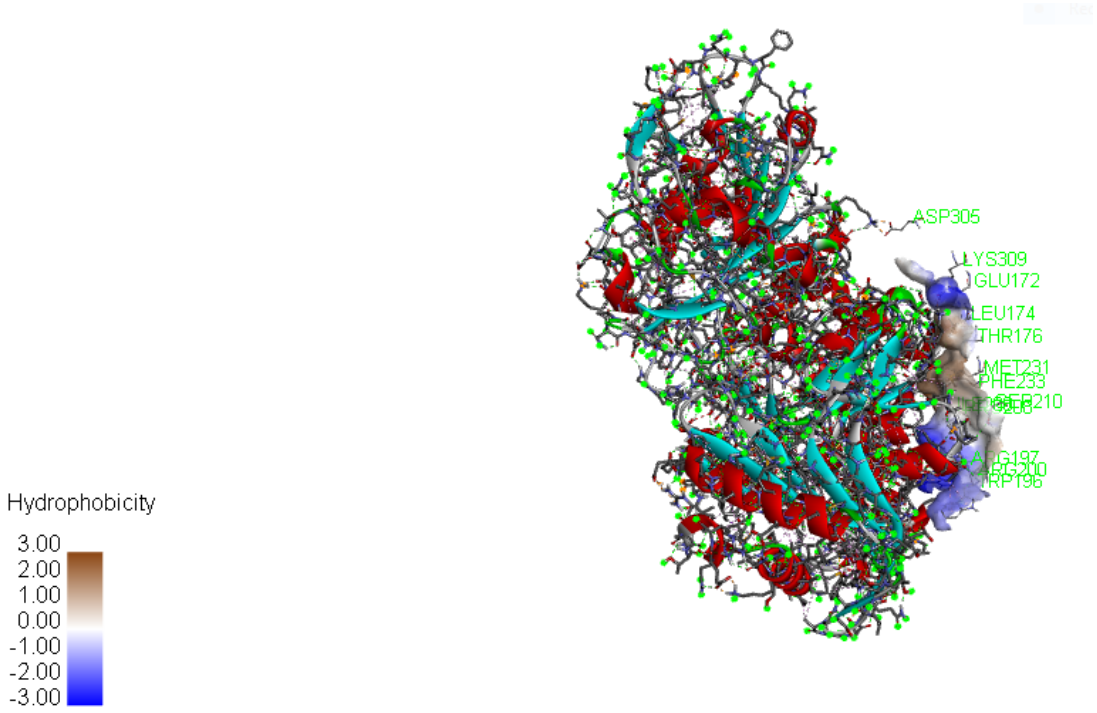

FIGURE 4.5: The docking result G3P AND PGM1The picture shows that there are many amino acids that help in interaction of two proteins. In the docking between G3P-PGM1, G3P is the receptor and PGM1 is ligand.

Discovery studio was a software that were established for researchers for feigning small molecule and macromolecule systems. It was established and disseminated by Dassault Systemes BIOVIA.

Discovery studio gives us no of different application e.g. Simulation, ligand design, pharmacophore modeling and structure-base design it were helpful in drug design and protein modeling research [161].

4.10 Identification of Common Amino Acids in Interacting Complexes to Confirm Set 2 Genes as Novel Biomarkers

Docking results were analyzed in discovery studio, the interacting amino acids were identified between the docked complexes and surface analysis was performed, and

interacted amino acids were noted. The molecular docking of set 1 biomarker and set 2 novel biomarkers was done to identify the most common interacting amino acids residues, shown in table 4.6.

Table 4.6: Most Frequent Amino Acids as a Result of Docking

Sr.no	**Amino acids**	**Repetition**
1	TYR215	9
2	ARG400	10
3	ASP35	10
4	LEU10	10
5	LEU39	10
6	LYS194	10
7	MET231	10
8	THR187	10
9	ARG200	11
10	HIS218	11
11	LYS18	11
12	LYS186	11
131	PHE233	11
14	TYR45	11
15	ARG40	12
16	TRP15	12
17	ARG16	13
18	ARG21	13
19	TYR42	13
20	TRP196	14
21	ARG197	15
22	ARG13	23

From table 4.6 It is observed that majority of the set 2 novel biomarkers which showed interactions with the set 1 biomarkers in gene interaction network. Interact

with the same amino acid of their corresponding receptor.

Therefore, during interaction of two sets of biomarkers, the novel biomarkers cause mutations in these amino acids which result in the development of disease. And mutations in these amino acids due to interacting novel biomarkers result in the decrease stability of receptor gene.

4.10.1 Identification of Biomarkers Linked to Prediabetic Insulin Resistance.

The biomarkers that is associated to prediabeties and insulin resistance were retrieved through literature mining and we were obtained 25 different biomarkers through this process, and named them as set no 1. To Identify novel biomarkers associated with prediabetic insulin resistance through protein proteins interaction network.

The interaction network was generated through funcoup server of the biomarkers that were obtained through literature mining, in interaction network there are many genes that interact with set no 1 gene and noted the genes and name them set no 2. The set no 2 genes were used as noval biomarkers.

4.10.2 Identification of Common Interacting Amino Acid Residue between Biomarkers

To accomplished this aim protein protein docking were done set no 1 gene were placed as receptors and set no 2 genes were placed as ligand, then there are few amino acids that were repeated again and again.

So ARG13 is the mostly repeated amino acid in the result of docking so we can say mutations in these amino acids due to interacting novel biomarkers result in the decrease stability of receptor gene, this mutation in the biomarker may cause the diseases.

4.11 Discussion

The Identification of biomarkers linked to prediabetic insulin resistance. The biomarkers that is associated to prediabetes and insulin resistance were retrieved through literature mining and we were obtained 24 different biomarkers through this process, and named them as set no 1. All of these genes were obtained in manual process.

To Identify novel biomarkers associated with prediabetic insulin resistance through protein proteins interaction network, the interaction network was generated through funcoup server of the biomarkers that were obtained through literature mining, in interaction network there are many genes that interact with set no 1 gene and noted the genes and name them set no 2. The set no 2 genes were used as novel biomarkers.

Identification of common interacting amino acid residue between biomarkers, To accomplished this aim protein proteins docking were done set no 1 gene were placed as receptors and set no 2 genes were placed as ligand, then there are few amino acids that were repeated again and again so ARG13 is the mostly repeated amino acid in the result of docking so we can say mutations in these amino acids due to interacting novel biomarkers result in the decrease stability of receptor gene, this mutation in the biomarker may cause the diseases.

Many people work on prediabetes and insulin resistance that leads to type 2 diabetes, DR. Polonsky Work on MODY1 and MODY2 genes that play role in diabetes and check the mutation in these genes if the mutation in MODY1 and MODY2 then it leads to type 2 diabetes. If the mutation in MODY1 gene occur then it create defect in bete cell and more and more insulin produce in body and insulin resistance created[164][169]. So, we were checked the biomarkers that create prediabetes and insulin resistance in body.

A .T.HATTERSLEY check the function of GCK gene that play role in hyperglycaemia if the mutation in this gene then the body leads to prediabetes and then

diabetes[170]. GCK play an important role in glucose phosphorylation if the mutataion in this genes then the body no work properly, we were also select this gene as a biomarker and observe that it play an important role in body and maintain the glucose level of body[163].

In shanghi china a Raquel Villegas and there collages work on the gene IRS1,IRS2, INSR and IRS3 and check their work in young age people and observed that these gene play important role in glucose pathway if any mutation occur in these gene then the body create insulin resistance and prediabetes and finaly diabetes[166].

So, we were all these genes and no of other genes that play role in glucose management and observed their function. the aims and objectives of my study in which identification of biomarkers and then novel biomarkers that is retrieved by interaction network and then protein proteins docking were done and check the effect of these biomarkers in progression of insuin resistance and prediabetes and we were find some repeated amino acids in which if mutation occur then the body leads to type 2 diabetes.

Chapter 5

Conclusions and Recommendations

The biomarkers that is associated with prediabetes and insulin resistance are also collected through literature review by manual processw and name them as set no 1 and the generate interaction network of set no 1 gene, in interaction network there are many genes that interact with set no 1 gene and noted the genes and name them set no 2, in interaction network many genes from set no 1 are not appear so we have total 40 different genes then enrichment analysis were done on all the genes to check their biological, molecular and cellular annotation. Through uniport the ids of all the genes and protein sequences were obtained. The genes that are collected by interaction network and literature mining were checked by PDB through BLASTP to check the same three-dimensional proteins structures and then ProSA server were used to check the energy profiles of all the modeled proteins. Through proSA server we have examined all proteins structure have negative Z score. Then protein protein doking were done through GRAMM-X web server and Cluspro, abd all the docked proteins were checked in discovery studio. Through docking we have examined that ARG13 was most repeated amino acid in all docked proteins. The molecular docking of set 1 biomarker and set 2 novel biomarkers was done to identify the most common interacting amino acids residues. It is observed that majority of the set 2 novel biomarkers which showed interactions with the set 1

biomarkers in gene interaction network. Interact with the same amino acid of their corresponding receptor, Therefore, Common interacting amino acids were checked using molecular docking of both sets of biomarkers. It was observed that the most interacting common amino acids were ARG13, In majority of the docked complexes. It was inferred from docking that set 2 biomarkers interacts with these amino acids of set 1 biomarkers, so they may be potential biomarkers for pre diabetic insulin resistance. Identified biomarkers associated with prediabetes could further be confirmed or validated in our population using different wet lab technologies. The identified biomarker was in different form it may be a protein, gene, micro-RNA and any metabolite; so, there were different methods in wet lab to validate all type of biomarkers. However, if the biomarker was a gene then sequence alignment and polymorphism PCR done; if the biomarker was a metabolite then electrophoresis and spectrophotometry was done, if the biomarker was a protein then ELISA (enzyme linked immunosorbent assay) were done and if the biomarkers was a micro-RNA then PCR (polymerase chain reaction) was done. This will further lead to identification of real biomarkers in our population that could be used for designing diagnosis techniques specific for pre-diabetic insulin resistance. Preventing prediabetic insulin resistance will inhibit the progression of deadly diseases like diabetes, cardiovascular diseases and complex syndromes such as metabolic syndrome.

Bibliography

[1] 2013 Sheean et al., "NIH Public Access," Bone, vol. 23, no. 1, pp. 1–7, 2008.

[2] E. Ferrannini, A. Gastaldelli, and P. Iozzo, "Pathophysiology of Prediabetes," Med. Clin. North Am., vol. 95, no. 2, pp. 327–339, 2011.

[3] S. N. Seclen, M. E. Rosas, A. J. Arias, E. Huayta, and C. A. Medina, "Prevalence of diabetes and impaired fasting glucose in Peru: report from PERU-DIAB , a national urban population-based longitudinal study," pp. 1–7, 2015.

[4] M. Stumvoll, B. J. Goldstein, and T. W. Van Haeften, "Type 2 diabetes: principles of pathogenesis and therapy," vol. 365, 2010.

[5] L. A. Ahmad and J. P. Crandall, "Type 2 Diabetes Prevention: A Review," vol. 28, no. 2, pp. 53–59, 2010.

[6] J. L. Leahy, "Pathogenesis of Type 2 Diabetes Mellitus," vol. 36, pp. 197–209, 2005.

[7] E. Nicolaides and C. J. H. Jones, "Type 2 diabetes – implications for macrovascular mechanics and disease," vol. 2, no. 1, pp. 9–12, 2002.

[8] "Symptoms of Diabetes and Their Association With the Risk and Presence of," vol. 30, no. 11, 2007.

[9] R. Valdez, "Detecting Undiagnosed Type 2 Diabetes: Family History as a Risk Factor and Screening Tool," vol. 3, no. 4, pp. 722–726, 2009.

[10] I. Classification, "Standards of Medical Care in Diabetes d 2014," vol. 37, no. October 2013, pp. 14–80, 2014.

[11] W. T. Cefalu, "Insulin resistance: Cellular and clinical concepts," Proc. Soc. Exp. Biol. Med., vol. 226, no. 1, pp. 13–26, 2001.

[12] G. Reaven, "The metabolic syndrome or the insulin resistance syndrome? Different names, different concepts, and different goals," Endocrinol. Metab. Clin. North Am., vol. 33, no. 2, pp. 283–303, 2004.

[13] P. Anni, S. Guzzo, A.Fargion, Dongiov, S. Colombo, L. Valenti, and A. L. Fracanzani, "Iron and insulin resistance," Aliment. Pharmacol. Ther. Suppl., vol. 22, no. 2, pp. 61–63, 2005.

[14] S. B. Wheatcroft, I. L. Williams, A. M. Shah, and M. T. Kearney, "Pathophysiological implications of insulin resistance on vascular endothelial function," Diabet. Med., vol. 20, no. 4, pp. 255–268, 2003.

[15] U. Smith, "Impaired ('diabetic') insulin signaling and action occur in fat cells long before glucose intolerance - Is insulin resistance initiated in the adipose tissue?," Int. J. Obes., vol. 26, no. 7, pp. 897–904, 2002.

[16] S. J. Hunter and W. T. Garvey, "Insulin action and insulin resistance: Diseases involving defects in insulin receptors, signal transduction, and the glucose transport effector system," Am. J. Med., vol. 105, no. 4, pp. 331–345, 1998.

[17] S. M. Grundy, "What is the contribution of obesity to the metabolic syndrome?," Endocrinol. Metab. Clin. North Am., vol. 33, no. 2, pp. 267–282, 2004.

[18] R. M. Krauss and P. W. Siri, "Metabolic abnormalities: Triglyceride and low-density lipoprotein," Endocrinol. Metab. Clin. North Am., vol. 33, no. 2, pp. 405–415, 2004.

[19] H. Ginsberg, J. M. Olefsky, and G. M. Reaven, "Further evidence that insulin resistance exists in patients with chemical diabetes," Diabetes, vol. 23, no. 8, pp. 674–678, 1974.

[20] K. M. V. Narayan and M. B. Weber, "Clinical risk factors, DNA variants, and the development of type 2 diabetes," N. Engl. J. Med., vol. 360, no. 13, pp. 1360–1361, 2009.

[21] R. A. DeFronzo and E. Ferrannini, "Insulin resistance: A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atheros clerotic cardiovascular disease," Diabetes Care, vol. 14, no. 3, pp. 173–194, 1991.

[22] E. Ferrannini et al., "Insulin resistance, insulin response, and obesity as indicators of metabolic risk," J. Clin. Endocrinol. Metab., vol. 92, no. 8, pp. 2885–2892, 2007.

[23] H. N. Ginsberg, "Insulin resistance and cardiovascular disease," J. Clin. Invest., vol. 106, no. 4, pp. 453–458, 2000.

[24] M. A. Abdul-Ghani and R. A. DeFronzo, "Pathophysiology of prediabetes," Curr. Diab. Rep., vol. 9, no. 3, pp. 193–199, 2009.

[25] G. M. Reaven, "Insulin resistance: A chicken that has come to roost," Ann. N. Y. Acad. Sci., vol. 892, no. 650, pp. 45–57, 1999.

[26] G. M. Reaven et al., "Insulin secretion and action in noninsulin-dependent diabetes mellitus. Is insulin resistance secondary to hypoinsulinemia?," Am. J. Med., vol. 75, no. 5 PART 2, pp. 85–93, 1983.

[27] B. Singh, "Surrogate markers of insulin resistance: A review," World J. Diabetes, vol. 1, no. 2, p. 36, 2010.

[28] F. Facts and O. N. Diabetes, "centers for disease control and prevention. National Diabetes Fact Sheet," Nj, vol. 07470, pp. 973–696, 2011.

[29] A. G. Tabák, C. Herder, W. Rathmann, E. J. Brunner, and M. Kivimäki, "Diabetes 1 Prediabetes: a high-risk state for diabetes development," Lancet, vol. 379, no. 9833, pp. 2279–2290, 2012.

[30] I. D. F. Consultation, "A European Evidence-Based Guideline for the Prevention of Type 2 Diabetes," vol. 42, pp. 3–36, 2010.

[31] D. Care and S. S. Suppl, "2 . Classi fi cation and Diagnosis of Diabetes: Standards of Medical Care in Diabetes d 2018," vol. 41, no. January, pp. 13–27, 2018.

[32] D. O. F. Diabetes, "Definition and description of diabetes other categories of glucose," vol. 33, 2010.

[33] C. M. Edwards, "P rediabe t es A Worldwide Epidemic," 2016.

[34] N. Mangalakumar, "Scholarship at UWindsor An Adaptive Clustering Algorithm for Gene Expression Time- Series Data Analysis An Adaptive Clustering Algorithm for Gene Expression Time-Series Data Analysis," 2018.

[35] J. C. C. Immunol, C. M. Mcgeough, and A. J. Bjourson, "Clinical & Cellular Immunology Diagnostic , Prognostic and Theranostic Genetic Biomarkers for Rheumatoid Arthritis," 2012.

[36] K. G. M. M. Alberti and P. Z. Zimmet, "Definition , Diagnosis and Classification of Diabetes Mellitus and its Complications Part 1: Diagnosis and Classification of Diabetes Mellitus Provisional Report of a WHO Consultation," pp. 539–553, 1998.

[37] C. Of, D. Mellitus, O. C. Of, G. Regulation, and D. O. F. Diabetes, "Diabetes mellitus and other categories of description of diabetes," vol. 28, 2005.

[38] 6th Ed. [online] Available at: http://www.idf.org/sites/default/ files/EN _6E _Atlas _Full _0.pdf [Accessed 17 March 2015]

[39] B. Paulweber et al., "A European Evidence-Based Guideline for the Prevention of Type 2 Diabetes," vol. 42, pp. 3–36, 2010.

[40] M. Mota et al., "Prevalence of diabetes mellitus and prediabetes in the adult Romanian population: Predator study †," pp. 1–9, 2015.

[41] B. Sv et al., "The association between Diabetes mellitus and Depression Pathophysiological mechanisms," vol. 9, no. 2, pp. 120–125, 2016.

[42] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030," Diabetes Res. Clin. Pract., vol. 94, no. 3, pp. 311–321, 2011.

[43] L. M. Fa, "Outcomes of diabetes care: a population-based study," vol. 15, no. 4, 2003.

[44] T. Ninomiya et al., "Albuminuria and Kidney Function Independently Predict Cardiovascular and Renal Outcomes in Diabetes," pp. 1813–1821, 2009.

[45] S. M. Gold, I. Dziobek, V. Sweat, and A. Tirsi, "Hippocampal damage and memory impairments as possible early brain complications of type 2 diabetes," pp. 711–719, 2007.

[46] C. D. Moulton, J. C. Pickup, and K. Ismail, "Depression and diabetes 2 The link between depression and diabetes: the search for," LANCET Diabetes Endocrinol., vol. 3, no. 6, pp. 461–471, 2015.

[47] P. Pathway, "Pathogenesis of Diabetic Nephropathy Ryuichi Kikkawa * and Masakazu Haneda," 1997.

[48] R. Kikkawa, "Chronic complications in diabetes mellitus," no. 2000, pp. 1996–1998, 2017.

[49] R. Donnelly, A. M. Emslie-smith, I. D. Gardner, and A. D. Morris, "Vascular complications of diabetes Microvascular complications Macrovascular complications," vol. 320, no. April, 2000.

[50] A. D. A. Statements, "American Association of Clinical Endocrinologists and American Diabetes ," vol. 32, no. 6, 2009.

[51] H. M. Belmont, "No Title," vol. 123, no. 12, pp. 961–974, 1995.

[52] A. Qaseem, T. J. Wilt, D. Kansagara, and C. Horwitch, "Hemoglobin A 1c Targets for Glycemic Control With Pharmacologic Therapy for Nonpregnant

Adults With Type 2 Diabetes Mellitus: A Guidance Statement Update From the American College of Physicians," no. March 2017, 2018.

[53] P. R. Cavanagh, "Therapeutic footwear for people with diabetes," vol. 20, no. December 2003, pp. 51–55, 2004.

[54] R. E. Chance and B. H. Frank, "Research, development, production, and safety of biosynthetic human insulin," Diabetes Care, vol. 16, no. 3 SUPPL. 3, pp. 133–142, 1993.

[55] "Society for endocrinology , metabolism and diabetes of southern Africa: abstracts of papers The following are abstracts of papers presented at the 10th Meeting of the Society for Endocrinology , Metabolism and Diabetes of Southern Africa held at the Athe," no. September 1970, pp. 25–29.

[56] I. S. Johnson, "Human Insulin from Recombinant DNA Technology Author (s): Irving S . Johnson," Science (80-.)., vol. 219, no. 4585, pp. 632–637, 1983.

[57] V. B. Khyade, "Herbals and Their Compounds Targeting Pancreatic Beta Cells for the Treatment of Diabetes," Int. J. Sci. Stud., vol. 6, no. 3, pp. 1–44, 2018.

[58] P. Sonksen, J. Sonkserr, C. London, S. T. Hospital, and L. Sei, "Insulin: understanding its action in health and disease," Br. J. Anaesth., vol. 85, no. 1, pp. 69–79, 2000.

[59] J. H. Koeslag, P. T. Saunders, and E. Terblanche, "A reappraisal of the blood glucose homeostat which comprehensively explains the type 2 diabetes mellitus-syndrome X complex," J. Physiol., vol. 549, no. 2, pp. 333–346, 2003.

[60] K. Docherty and D. F. Steiner, "Post-translational proteolysis in polypeptide hormone biosynthesis.," Annu. Rev. Physiol., vol. 44, no. 4, pp. 625–638, 1982.

[61] N. P. Group, "Insulin and its receptor: structure , function and evolution," no. 7, pp. 1351–1362, 2004.

[62] M. Shiao, B. Liao, M. Long, and H. Yu, "Adaptive Evolution of the Insulin Two-Gene System in Mouse," 2008.

[63] A. Wollmer, B. Rannefeld, B. R. Johansen, K. R. Hejnaes, P. Balschmidt, and F. B. Hansen, "Phenol-Promoted Structural Transformation of Insulin in Solution *," vol. 368, no. August, pp. 903–911, 1987.

[64] L. Langkjar, "Insulin Structure and Stability," no. 1922, 1988.

[65] U. Derewenda, Z. Derewenda, G. G. Dodson, R. E. Hubbard, and F. Korber, "Molecular structure of insulin: The insulin monomer and its assembly," vol. 45, 1989.

[66] P. De Meyts, "Insulin and its receptor: structure , function and evolution," no. 7, pp. 1351–1362, 2004.

[67] P. De Meyts and J. Whittaker, "Structural biology of insulin and igf1 receptors: implications for drug design," 2002.

[68] C. M. Yip and M. D. Ward, "Atomic Force Microscopy of Insulin Single Crystals: Direct Visualization of Molecules and Crystal Growth," Biophys. J., vol. 71, no. 2, pp. 1071–1078, 1996.

[69] P. F. Egea, R. M. Stroud, and P. Walter, "Targeting proteins to membranes: structure of the signal recognition particle," pp. 213–220, 2005.

[70] Z. Vajo, J. Fawcett, and W. C. Duckworth, "Recombinant DNA Technology in the Treatment of Diabetes: Insulin Analogs," vol. 22, no. 5, pp. 706–717, 2001.

[71] A. P-, B. Ahmad, P. R. D. I. Ri, and O. A. B. Gh, "Pharmacology of insulin RE IO) L ED N IM IN," vol. 44, no. 0, pp. 10–14, 2003.

[72] W. C. Duckworth, R. G. Bennett, and F. G. Hamel, "Insulin Degradation: Progress and Potential *," no. November 1998, 2014.

[73] S. Najjar and M. College, "Insulin Action: Molecular Basis of Diabetes Insulin Action: Basic Mechanism," pp. 1–10.

[74] N. Pulizzi et al., "Clinical Risk Factors, DNA Variants, and the Development of Type 2 Diabetes," 2008.

[75] M. Krssak et al., "Rapid communication Intramyocellular lipid concentrations are correlated with insulin sensitivity in humans: a 1 H NMR spectroscopy study," 1999.

[76] G. Perseghin et al., "Intramyocellular Triglyceride Content Is a Determinant of in Vivo Insulin Resistance in Humans A 1 H - 1 3 C Nuclear Magnetic Resonance Spectroscopy Assessment in Offspring of Type 2 Diabetic Parents," vol. 48, no. August, 1999.

[77] I. Sensitivity et al., "H Nuclear Magnetic Resonance Spectroscopy in Lean," vol. 51, April, 2002.

[78] V. T. Samuel and G. I. Shulman, "The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux," vol. 126, no. 1, 2016.

[79] R. Marium et al., "From Pre-Diabetes to Diabetes: Diagnosis ," no. 55, pp. 1–30.

[80] T. Pros and D. D. With, "The Pros and Cons of Diagnosing Diabetes With A1C," vol. 34, 2011.

[81] A. D. Pradhan, J. E. Manson, J. E. Buring, and P. M. Ridker, "C-Reactive Protein, Interleukin 6, and Risk of Developing Type 2 Diabetes Mellitus," vol. 286, no. 3, pp. 327–334, 2001.

[82] N. Bansal, "Prediabetes diagnosis and treatment: A review," vol. 6, no. 2, pp. 296–303, 2015.

[83] S. E. Kahn, "The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes," pp. 3–19, 2003.

[84] J. R. Prensner, J. R. Prensner, M. A. Rubin, J. T. Wei, and A. M. Chinnaiyan, "Beyond PSA: The Next Generation of Prostate Cancer Biomarkers," vol. 3, 2012.

[85] D. Kaufman, "Journal of Clinical & Cellular Immunology," no. November, 2015.

[86] C. E. Pearson, "Repeat Associated Non-ATG Translation Initiation: One DNA , Two Transcripts , Seven Reading Frames , Potentially Nine Toxic Entities," vol. 7, no. 3, pp. 1–5, 2011.

[87] P. Kesteven et al., "A Randomized Trial of Genotype-Guided Dosing of Warfarin," 2013.

[88] J. K. Aronson and R. E. Ferner, "Biomarkers — A General Review," no. March, pp. 1–17, 2017.

[89] B. Dorcely, K. Katz, R. Jagannathan, S. S. Chiang, I. J. Goldberg, and M. Bergman, "Novel biomarkers for prediabetes , diabetes , and associated complications," pp. 345–361, 2017.

[90] O. Articles, "Use of HbA1c for Diagnoses of Diabetes and Prediabetes: Comparison with Diagnoses Based on Fasting and 2-Hr Glucose Values and Effects of Gender , Race , and Age," vol. 12, no. 5, pp. 258–268, 2014.

[91] R. M. Bookchin and P. M. Gallop, "Structural biology of insulin and igf1 receptors: implications for drug design," vol. 32, no. 1, pp. 86–93, 1968.

[92] R. T. Ribeiro, M. P. Macedo, and J. F. Raposo, "HbA1c , Fructosamine , and Glycated Albumin in the Detection of Dysgly- caemic Conditions," pp. 14–19, 2016.

[93] E. Danese, M. Montagnana, A. Nouvenne, and G. Lippi, "Advantages and Pitfalls of Fructosamine and Glycated Albumin in the Diagnosis and Treatment of Diabetes," 2015.

[94] A. E. Sumner et al., "A1C Combined With Glycated Albumin Improves Detection of Prediabetes in Africans: The Africans in America Study," vol. 39, no. February, pp. 271–277, 2016.

[95] C. Cl et al., "Screening for type 2 diabetes and prediabetes in obese youth: evaluating alternate markers fructosamine , and glycated albumin," pp. 1–6, 2015.

[96] N. Furusyo, T. Koga, M. Ai, and S. Otokozawa, "Utility of glycated albumin for the diagnosis of diabetes mellitus in a Japanese population study: results from the Kyushu and Okinawa Population Study (KOPS) Atherosclerosis Risk in Communities Japanese Society for Clinical Chemistry," pp. 3028–3036, 2011.

[97] A. D. Pradhan, J. E. Manson, J. E. Buring, and P. M. Ridker, "C-Reactive Protein, Interleukin 6, and Risk of Developing Type 2 Diabetes Mellitus," vol. 286, no. 3, pp. 327–334, 2001.

[98] P. Závodszky and P. Gál, "Multiple roles of complement MASP-1 at the interface of innate immune response and coagulation," Mol. Immunol., pp. 1–10, 2014.

[99] T. Suzuki et al., "Interleukin-6 Enhances Glucose-Stimulated Insulin Secretion From Pancreatic b -Cells Potential Involvement of the PLC-IP 3 – Dependent Pathway," vol. 60, no. February, pp. 537–547, 2011.

[100] J. Kim, R. A. Bachmann, and J. Chen, "Interleukin-6 and Insulin Resistance," vol. 80, no. 08, pp. 613–633.

[101] B. Vozarova, C. Weyer, R. S. Lindsay, R. E. Pratley, C. Bogardus, and P. A. Tataranni, "High White Blood Cell Count Is Associated With a Worsening of Insulin Sensitivity and Predicts the Development of Type 2 Diabetes," vol. 51, no. October 2001, 2002.

[102] C. Mertoğlu and M. Günay, "Neutrophil-Lymphocyte Ratio and Platelet-Lymphocyte Ratio as Useful Predictive Markers of Prediabetes and Diabetes Mellitus," Diabetes Metab. Syndr. Clin. Res. Rev., 2016.

[103] M. Verdoia, A. Schaffer, L. Barbieri, G. Aimaretti, P. Marino, and F. Sinigaglia, "Impact of diabetes on neutrophil-to-lymphocyte ratio and its relationship to coronary artery disease," Diabetes Metab., pp. 1–8, 2015.

[104] V. Grossmann et al., "Profile of the Immune and Inflammatory Response in Individuals With Prediabetes and Type 2 Diabetes," pp. 1–9, 2015.

[105] A. Festa, R. D. Agostino, R. P. Tracy, and S. M. Haffner, "Plasminogen Activator Inhibitor-1 Predict the The Insulin Resistance Atherosclerosis Study," vol. 51, no. APRIL, 2002.

[106] A. Festa et al., "States of Glucose Tolerance," 1999.

[107] R. From et al., "Elevated Levels of Interleukin-18 Predict the Development of Type 2 Diabetes," vol. 54, no. October, 2005.

[108] M. Y. Donath, "Inflammation as a Sensor of Metabolic Stress in Obesity and Type 2 Diabetes," vol. 152, no. November, pp. 4005–4006, 2011.

[109] M. A. J. S. Hipley, "Adiponectin Trajectories Before Type 2 Diabetes Diagnosis," vol. 35, no. phase 1, 2012.

[110] F. Schick et al., "Impact of the Adipokine Adiponectin and the Hepatokine Fetuin-A on the Development of Type 2 Diabetes: Prospective Cohort- and Cross-Sectional Phenotyping Studies," vol. 9, no. 3, pp. 1–8, 2014.

[111] D. Pal et al., "Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance," Nat. Med., no. July, pp. 1–10, 2012.

[112] W. E. Gall et al., "a -Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population," vol. 5, no. 5, 2010.

[113] S. V. E. R. R. E. L. A. N. D. Aas, "The formation of 2-hydroxybutyric acid in experimental animals," vol. 58.

[114] J. Cobb et al., "Journal of Diabetes Science and Technology," 2013.

[115] T. J. Cunningham, L. Yao, and A. Lucena, "Product inhibition of secreted phospholipase A2 may explain lysophosphatidylcholines ' unexpected therapeutic properties," vol. 10, pp. 1–10, 2008.

[116] A. Dotevall, S. Johansson, and A. Rosengren, "Increased levels of triglycerides , BMI and blood pressure and low physical activity increase the risk of diabetes in Swedish women . A prospective 18-year follow-up of the BEDA * study," pp. 615–622, 2004.

[117] L. Ding et al., "Serum lipoprotein (a) concentrations are inversely associated with T2D , prediabetes , and insulin resistance in a middle-aged and elderly Chinese population," vol. 56, pp. 920–926, 2015.

[118] T. D. Filippatos and E. C. Rizos, "Small High-Density Lipoprotein (HDL) Subclasses are Increased with Decreased Activity of HDL-Associated Phospholipase A 2 in Subjects with Prediabetes," pp. 547–555, 2013.

[119] M. Guasch-ferr and M. A. Mart, "Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis," vol. 39, no. May, pp. 833–846, 2016.

[120] A. Aregbesola, "Serum ferritin and glucose homeostasis: change in the association by glycaemic state," no. October 2014, pp. 507–514, 2015.

[121] J. Huang et al., "Iron Overload and Diabetes Risk: A Shift From Glucose," vol. 60, no. January, 2011.

[122] C. Von Toerne et al., "biomarkers associated with prediabetes: the KORA F4 study," Diabetologia, 2016.

[123] X. Zhang, C. Zhang, L. Chen, X. Han, and L. Ji, "Human serum acylcarnitine profiles in different glucose tolerance states," Diabetes Res. Clin. Pract., no. 11, pp. 1–7, 2014.

[124] P. Kovacs, M. Stumvoll, G. M. Fiedler, M. Mai, A. To, and A. B. Leichtle, "Serum Levels of Acylcarnitines Are Altered in Prediabetic Conditions," vol. 8, no. 12, 2013.

[125] S. H. Adams et al., "Plasma Acylcarnitine Profiles Suggest Incomplete Long-Chain Fatty Acid b -Oxidation and Altered Tricarboxylic Acid Cycle Activity in Type 2 Diabetic African-American Women 1 – 3," pp. 1073–1081, 2009.

[126] S. A. Summers, "Progress in Lipid Research Ceramides in insulin resistance and lipotoxicity," vol. 45, pp. 42–72, 2006.

[127] M. Párrizas et al., "Circulating miR-192 and miR-193b Are Markers of Intervention," vol. 100, no. February, pp. 407–415, 2015.

[128] L. Kong, J. Zhu, W. Han, X. Jiang, M. Xu, and Y. Zhao, "Significance of serum microRNAs in pre-diabetes and newly diagnosed type 2 diabetes: a clinical study," pp. 61–69, 2011.

[129] L. Sun, B. Jiang, W. Li, and J. Zou, "MicroRNA-15a positively regulates insulin synthesis by inhibiting uncoupling protein-2 expression," Diabetes Res. Clin. Pract., vol. 91, no. 1, pp. 94–100, 2011.

[130] R. M. Flight and P. D. Wentzell, "Potential Bias in GO:: T ermFinder," vol. 10, no. 3, 2009.

[131] D. W. Huang et al., "DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists," vol. 35, pp. 169–175, 2007.

[132] M. Smid and L. C. J. Dorssers, "GO-Mapper: functional analysis of gene expression data using the expression level as a score to evaluate Gene Ontology terms," vol. 20, no. 16, pp. 2618–2625, 2004.

[133] S. F. Altschul et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," vol. 25, no. 17, pp. 3389–3402, 1997.

[134] W. R. Pearson, "Finding Protein and Nucleotide Similarities with fasta," no. March, pp. 1–25, 2016.

[135] R. C. Edgar, "Musele: a multiple sequence alignment method with reduced time and space complexity," vol. 19, pp. 1–19, 2004.

[136] A. Waterhouse et al., "Swiss Model: homology modelling of protein structures and complexes," vol. 46, no. May, pp. 296–303, 2018.

[137] O. Buβ, J. Rudat, and K. Ochsenreither, "FoldX as Protein Engineering Tool: Better Than Random Based Approaches," Comput. Struct. Biotechnol. J., vol. 16, pp. 25–33, 2018.

[138] K. Kumar and A. Prakash, "Structure-based functional annotation of hypothetical proteins from Candida dubliniensis: a quest for potential drug targets," 3 Biotech, pp. 561–576, 2015.

[139] H. C. Lee, K. Lai, M. T. Lorenc, M. Imelfort, C. Duran, and D. Edwards, "Bioinformatics tools and databases for analysis of next-generation sequence data," vol. 11, no. 1, pp. 12–24, 2011.

[140] M. Ullah, J. Hira, T. Ghosh, N. Ishaque, and N. Absar, "A Bioinformatics Approach for Homology Modeling and Plasmodium falciparum 3D7," J. Young Pharm., vol. 4, no. 4, pp. 261–266.

[141] S. R. Comeau, D. W. Gatchell, and S. Vajda, "ClusPro: an automated docking and discrimination method for the prediction of protein complexes," vol. 20, no. 1, pp. 45–50, 2004.

[142] A. T. R. Laurie and R. M. Jackson, "Methods for the Prediction of Protein-Ligand Binding Sites for Structure- Based Drug Design and Virtual Ligand Screening," pp. 395–406, 2006.

[143] D. Seeliger, B. L. De Groot, and V. Pymol, "Ligand docking and binding site analysis with PyMOL and Autodock / Vina," pp. 417–422, 2010.

[144] K. Bano, F. Bano, and N. Akhtar, "Conformational analysis (geometry optimization) of nucleosidic antitumor antibiotic showdomycin by ArgusLab 4 software Conformational analysis (geometry optimization) of nucleosidic antitumor antibiotic showdomycin by arguslab 4 software," no. February, 2009.

[145] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, and M. Schroeder, "PLIP: fully automated protein – ligand interaction profiler," vol. 43, no. April, pp. 443–447, 2015.

[146] R. Singh, D. Park, J. Xu, R. Hosur, and B. Berger, "Struct2Net: a web service to predict protein – protein interactions using a structure-based approach," vol. 38, no. May, pp. 508–515, 2010.

[147] D. Rauh, G. Klebe, and M. T. Stubbs, "Understanding Protein – Ligand Interactions: The Price of Protein Flexibility," pp. 1325–1341, 2004.

[148] Y. Zhou et al., "for the analysis of systems-level datasets," Nat. Commun., no. 2019.

[149] T. Schmitt, C. Ogris, and E. L. L. Sonnhammer, "FunCoup 3.0: Database of genome-wide functional coupling networks," Nucleic Acids Res., vol. 42, no. D1, pp. 380–388, 2014.

[150] E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, and A. Valencia, "EnrichNet: Network-based gene set enrichment analysis," Bioinformatics, vol. 28, no. 18, pp. 451–457, 2012.

[151] T. U. Consortium, "The Universal Protein Resource (UniProt)," vol. 36, no. November 2007, pp. 190–195, 2008.

[152] T. U. Consortium, "The Universal Protein Resource (UniProt) in 2010," vol. 38, no. October 2009, pp. 142–148, 2010.

[153] V. N. Bhatia, D. H. Perlman, C. E. Costello, and M. E. Mccomb, "Software Tool for Researching Annotations of Proteins: Open-Source Protein Annotation Software with Data Visualization," vol. 81, no. 23, pp. 9819–9823, 2009.

[154] J. Ye, S. Mcginnis, and T. L. Madden, "Blast: improvements for better sequence analysis," vol. 34, pp. 6–9, 2006.

[155] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes , transcripts and proteins," vol. 33, pp. 501–504, 2005.

[156] D. E. Kim, D. Chivian, and D. Baker, "Protein structure prediction and analysis using the Robetta server," Nucleic Acids Res., vol. 32, no., pp. 526–531, 2004.

[157] M. Wiederstein and M. J. Sippl, "ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins," Nucleic Acids Res., vol. 35, no. SUPPL.2, pp. 407–410, 2007.

[158] A. Tovchigrechko and I. A. Vakser, "GRAMM-X public web server for protein-protein docking," Nucleic Acids Res., vol. 34, no. WEB. SERV. ISS., pp. 310–314, 2006.

[159] S. R. Comeau, D. W. Gatchell, S. Vajda, and C. J. Camacho, "ClusPro: An automated docking and discrimination method for the prediction of protein complexes," Bioinformatics, vol. 20, no. 1, pp. 45–50, 2004.

[160] D. Kozakov et al., "The ClusPro web server for protein-protein docking," Nat. Protoc., vol. 12, no. 2, pp. 255–278, 2017.

[161] D Studio and Discovery Studio Life Science Modeling and Simulations Researchgaye.Net, pp. 1-8, 2008.

[162] G. I. Bellt et al., "n OEl," vol. 88, no. February, pp. 1484–1488, 1991.

[163] A. T. Hattersley et al., "Linkage of type 2 diabetes to the glucokinase gene," Lancet, vol. 339, no. 8805, pp. 1307–1310, 1992.

[164] B. I. Hospital, "of the the Programmed Failure of the for Insulin Resistance," vol. 334, no. 12, pp. 777–783, 1996.

[165] P. R. Langlais, L. J. Mandarino, and W. T. Garvey, "Normal physiology of insulin action Mechanisms of insulin signal transduction," 2015.

[166] R. Villegas et al., "Genetic Variation and Insulin Resistance in Middle-Aged Chinese Men," Ann. Hum. Genet., vol. 79, no. 5, pp. 357–365, 2015.

[167] M. Benito, "Tissue specificity on insulin action and resistance: Past to recent mechanisms," Acta Physiol., vol. 201, no. 3, pp. 297–312, 2011.

[168] M. Benito, "Tissue-specificity of insulin action and resistance*," Arch. Physiol. Biochem., vol. 117, no. 3, pp. 96–104, 2011.

Annexure A

Enrichment Analysis

Table A1: Biological Annotation

Study feature	RR (95% CI)	Study feature	RR (95% CI
Prospective Cohort	1.43 (1.16-1.77)	RCT	1.16 (0.77-1.73)
DVT only	1.51 (1.13-2.01)	DVT and/or PE	1.29 (1.01-1.64)
Idiopathic VTE	1.05 (0.75-1.47)	Idiopathic and secondary VTE	1.53 (1.24-1.89)
Started before 1995	1.26 (0.99-1.62)	Started in 1995 or after	1.49 (1.13-1.98)
Follow-up ≤2 years	1.48 (1.00-2.20)	Follow-up> 2 years	1.33 (1.08-1.65)
Follow-up ≤4 years	1.26 (0.98-1.62)	Follow-up >4 years	1.53 (1.16-202)
OAT 3 months	1.59 (1.10-2.30)	OAT ≥6 months	1.07 (0.71-1.59)
APS and NA deficiencies included	1.49 (1.14-1.96)	APS and NA deficiencies excluded	1.26 (0.97-1.63)
High quality	1.30 (1.00-1.70)	Non-high quality	1.43 (1.10-1.85)

Table A2: Cellular Annotation

Annotation (pathway/ process)	Signif icance of network distance distri bution (XD-Score)	Signif icance of overlap (Fisher test, q-value)	Dataset size (uploaded gene set)	Dataset size (pathway gene set)	Dataset size (overlap)
axon part	0.7312*	5.1e-01	19	12	YWHAE
M band	0.6735*	5.1e-01	19	13	ENO1
NuA4 histone acetyl transferase complex	0.5812*	5.4e-01	19	15	ACTB
kinesin complex	0.4549*	5.7e-01	19	19	YWHAE
photo receptor inner segment	0.4549*	5.7e-01	19	19	ENO2
melan osome	0.3683	1.0e-03	19	93	YWHAE PRDX1 HSP90AB1 HSPA8

Table A3: Molecular Annotation

Annotation (pathway/ process)	Significance of network distance distribution (XD-Score)	Significance of overlap (Fisher-test, q-value)	Dataset size (uploaded gene set)	Dataset size (pathway gene set)	Dataset size (overlap)
insulin-like growth factor receptor binding	2.0538	0.00041	20	13	IRS1 INSR IGF1 IRS2
phosphatidy linositol 3-kinase binding	1.3269	0.00081	20	20	IRS1 INSR

Table A4: Domain Identification

Job id	Target	Length	Position of Conserved region
44482	HNF4A	474	1-8, 34-56, 130-148, 163-178, 378-394, 413-467
44491	HNF1A	631	1-6, 34-80, 91-105, 147-154, 174-203, 217-234, 248-249, 283-376, 386-406, 465-491, 499-519, 535-616, 628-632

Table A4 continued from previous page

Job id	Target	Length	Position of Conserved region
44577	FAS	1260	30-33, 408-423,474-492, 845-858,963-983,1067, 1099-1115,1162-1171
44578	FAS	1251	295,602-609,850-868, 928-957,1248-1251
44579	A0A0U1RQF0	1680	33,408-422, 474-492, 541-544, 701-705, 963-984, 1099-1114, 1171-1175, 1519-1530,
44580	A0A0U1RQF0	829	178-190, 429-444, 505-533, 825-830
44492	IRS1	1242	1-13, 133-144, 262-283, 302-423, 440-460, 471-485, 494-551, 554-585, 588-612, 621-694, 708-762, 772-861, 870-894, 904-1011, 1015-1085, 1095-1176, 1187-1228, 1240-1242

Table A4 continued from previous page

Job id	Target	Length	Position of Conserved region
44493	IRS2	1338	1-13, 63-65, 80-86, 89, 148-159, 164-176, 297-320, 332-496, 517-534, 550-570, 572-574, 576-596, 603-626, 634-654, 661-675, 682-764, 786-801, 808-884, 901-917, 932-1035, 1047-1070, 1085-1107, 1116-1187, 1189, 1197-1244, 1263-1282, 1302-1305, 1334, 1336-1338
44494	IGF1	195	49, 74-90, 110-125, 140-195
44495	GCKR	625	1-9, 445-446, 604-625
44497	K2C73	540	1-47, 52-53, 55-70, 76, 172-176, 456-486, 490-492, 494-540
44498	HXK4	465	1-13, 457-458, 460-465
44499	A7LFL1	465	1-13, 457-458, 460-465
44507	KPYM	531	1-27, 401,
44508	SCD5	330	1-14, 29-38, 320-330
44509	ACOD	359	1-14, 17-49, 345-360
44510	PCKGC	622	1-3, 5-7, 12,13,14-18, 621,622
44511	PCKGM	640	1-6, 8,34,36,37,639-640

Table A4 continued from previous page

Job id	Target	Length	Position of Conserved region
44512	INSR	1382	1-9,662-666, 680-728, 746-764,774-790, 989,1117-1132,1307-1352, 1365-1375, 1378-1382
44513	GLUT4	509	1-22,260-276, 482-502, 504-509
44514	PTN1	435	31,33-39, 204, 336-350, 433-435
44516	CCL2	99	24-27,89-99
44697	G3P	335	1, 334,335
44698	KPYR	574	1-10, 44-68, 442-448,
44700	KPYM	531	1-27, 401
44701	SSR2	369	1-12, 23-35, 235-255, 328-370
44702	FOS	380	1-3, 5-54, 82-145, 207-260, 273-291, 316, 352-373,
44703	HXK4	465	1-13, 457-458, 460-465
44704	APOA1	267	23-25, 29-31, 48-61, 150-181, 221-241, 260, 264-267
44705	PCKGM	640	1-6,8, 34,36-37, 639-640
44706	G6PC	357	349-350, 353-357
44707	APOC3	99	21-25,28-29, 43-60,87-100
44708	GLP1R	463	1-6, 8-9, 24-29, 49-60, 339-343, 423-463,

Annexure B

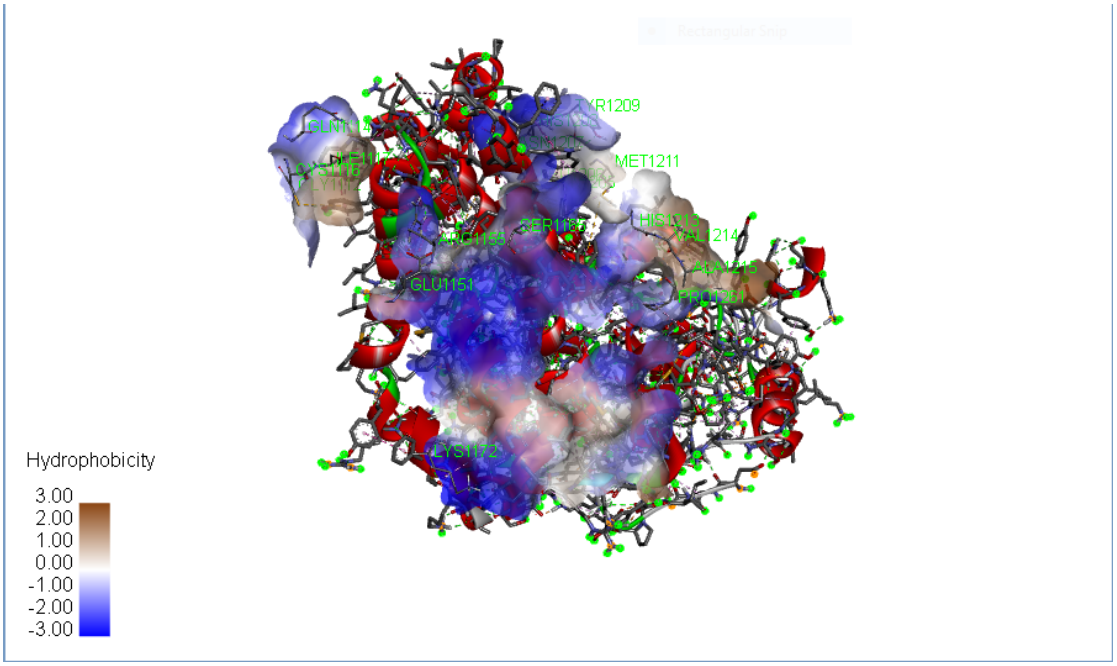

FIGURE A1: ACACA-ACOD1

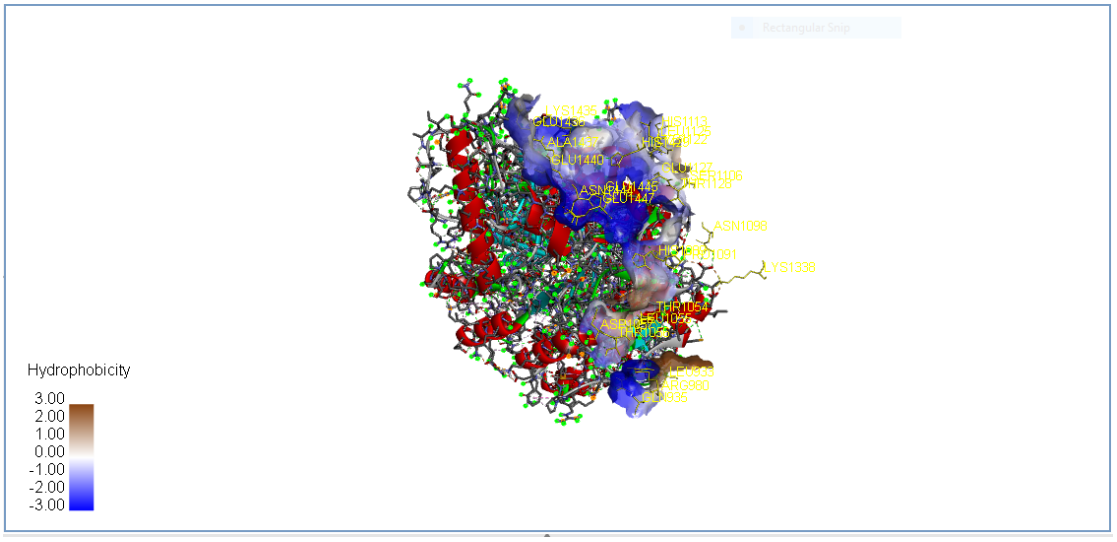

FIGURE A2: ACACA-PKM

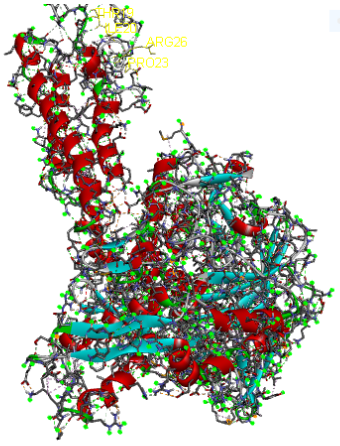

FIGURE A3: ACOD1-HSPA8

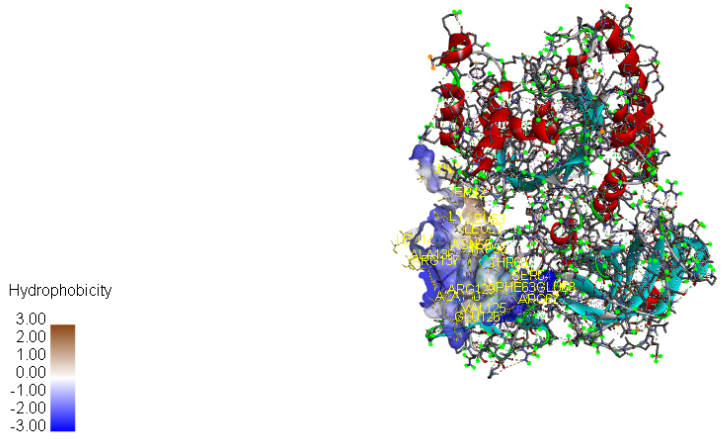

FIGURE A4: APOA1-EEF1A1

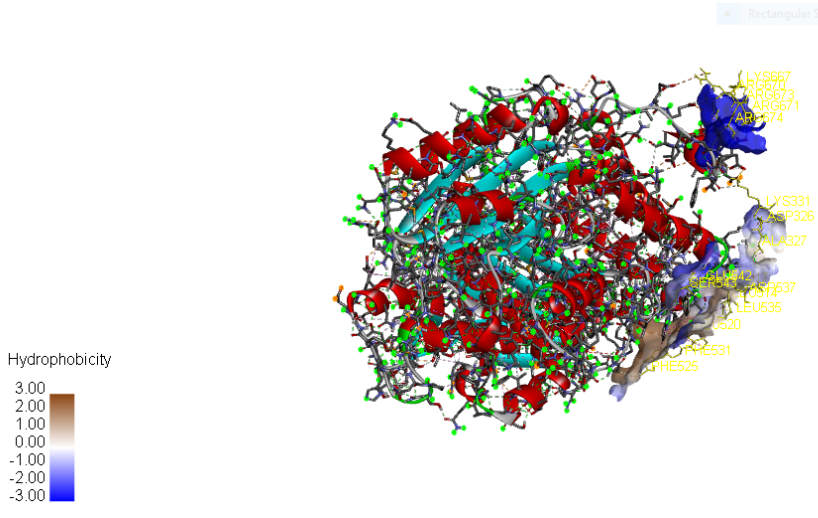

FIGURE A5: FASND1-TUBB4B

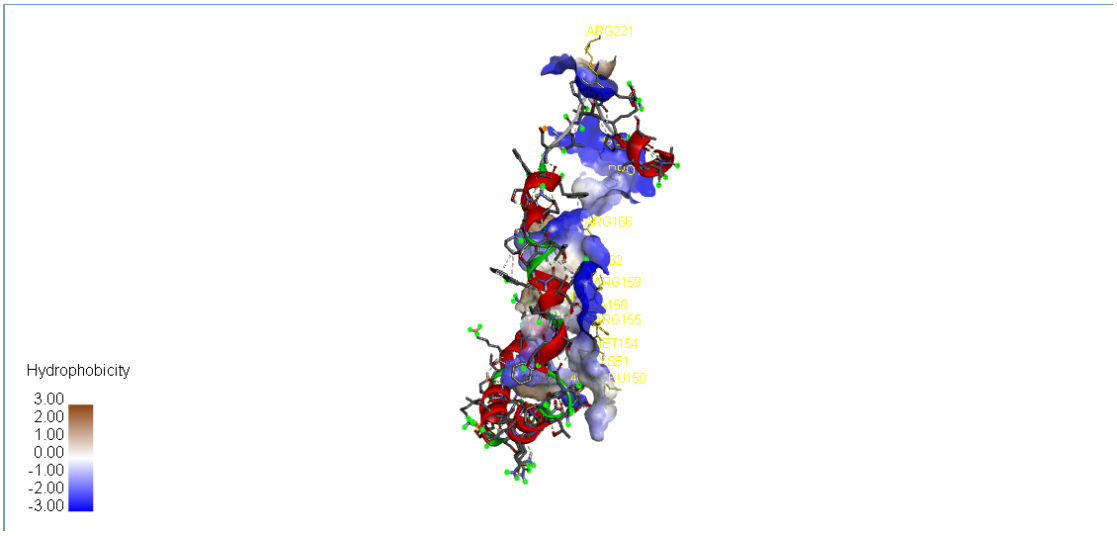

FIGURE A6: APOC3-APOA1

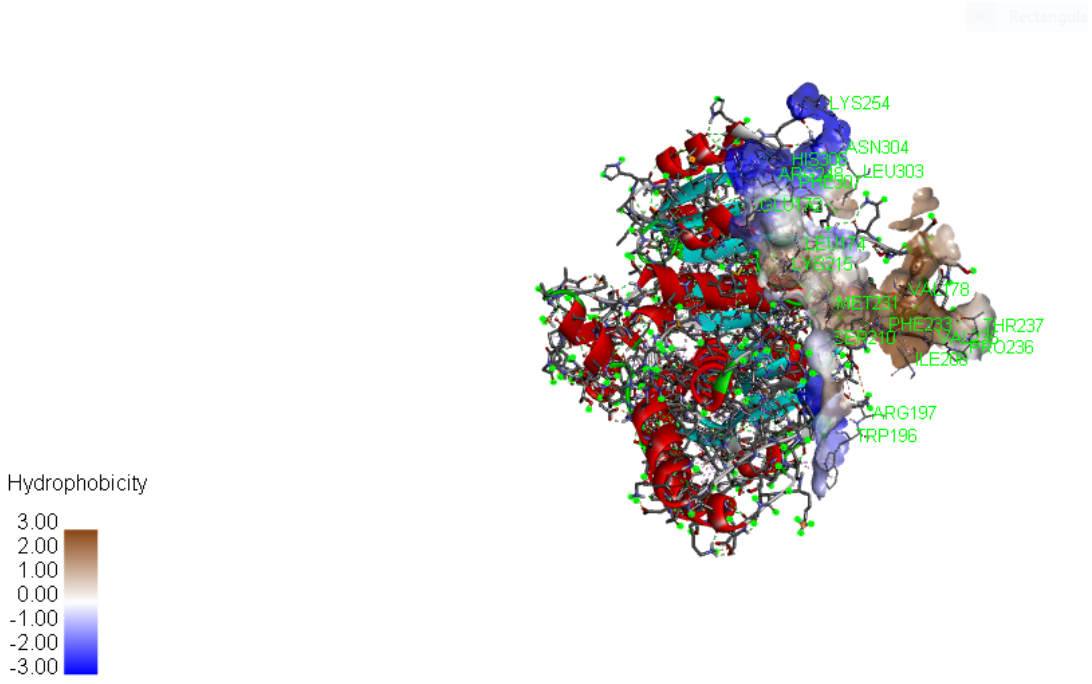

FIGURE A7: G3P-MDH2

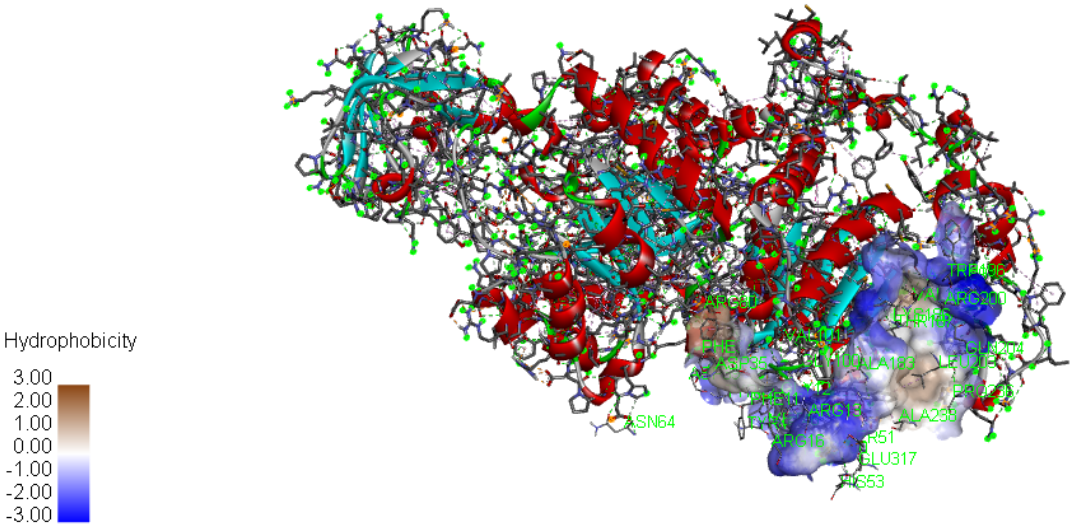

FIGURE A8: G3P-PKLR

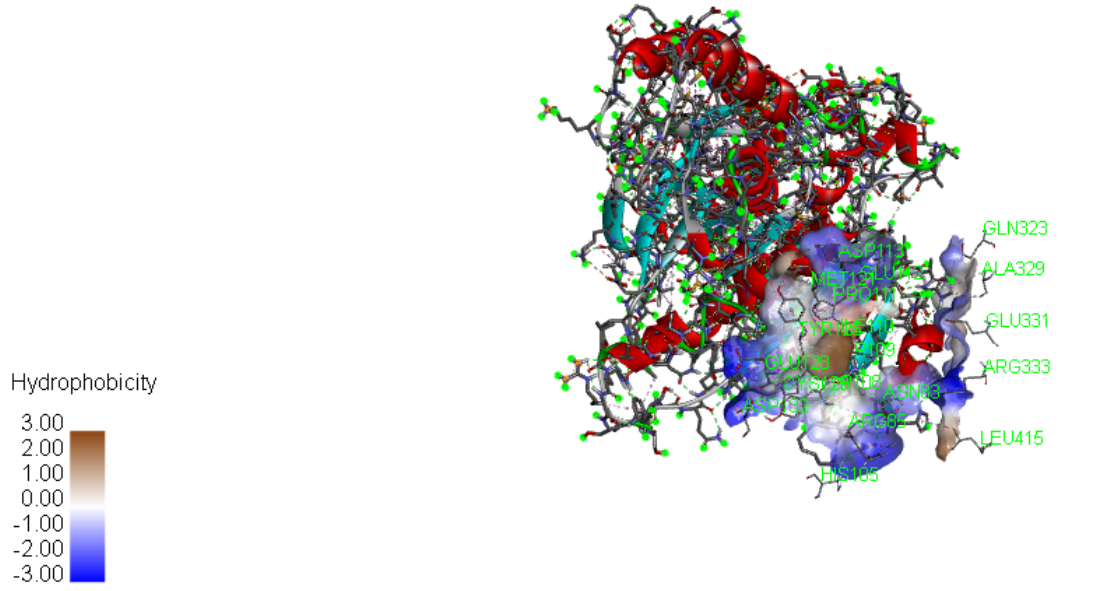

FIGURE A9: GCK-MDH2

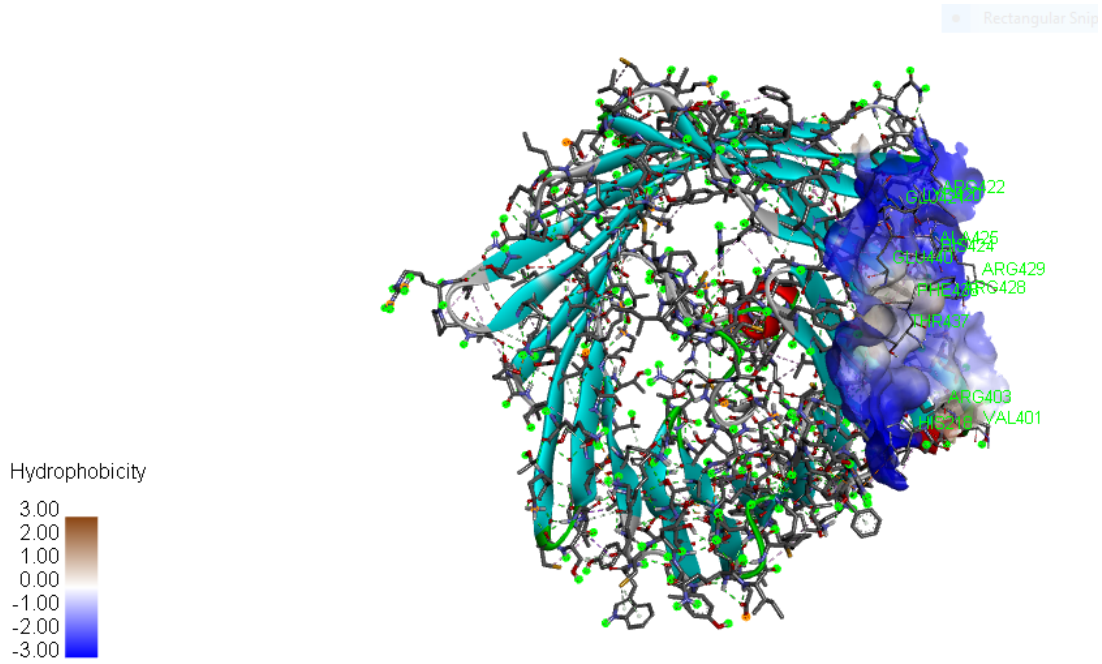

FIGURE A10: GCK-VDAC2

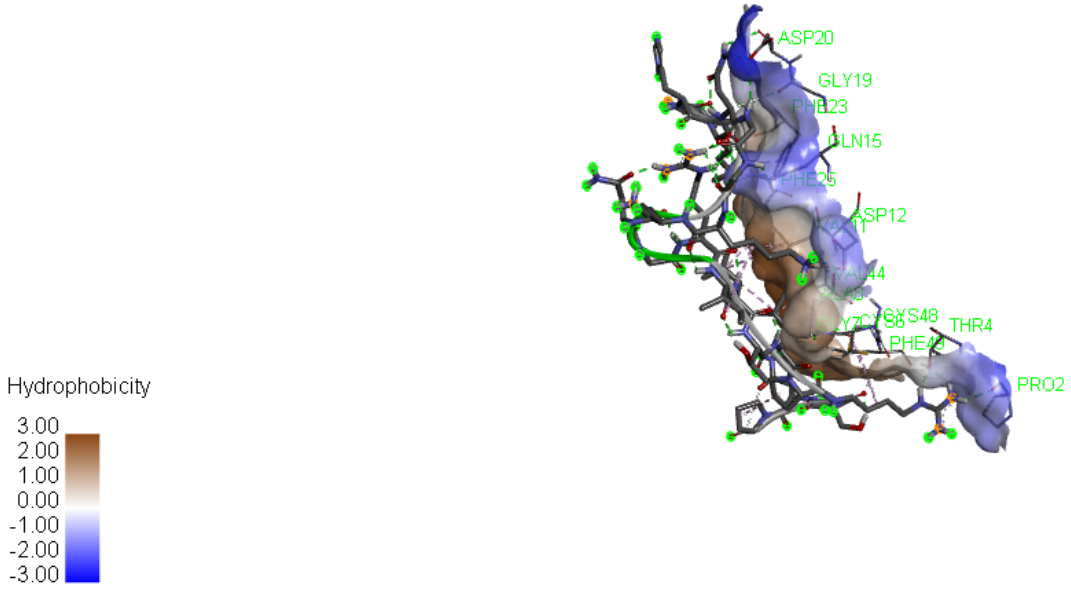

FIGURE A11: IGF1D1-INSR5

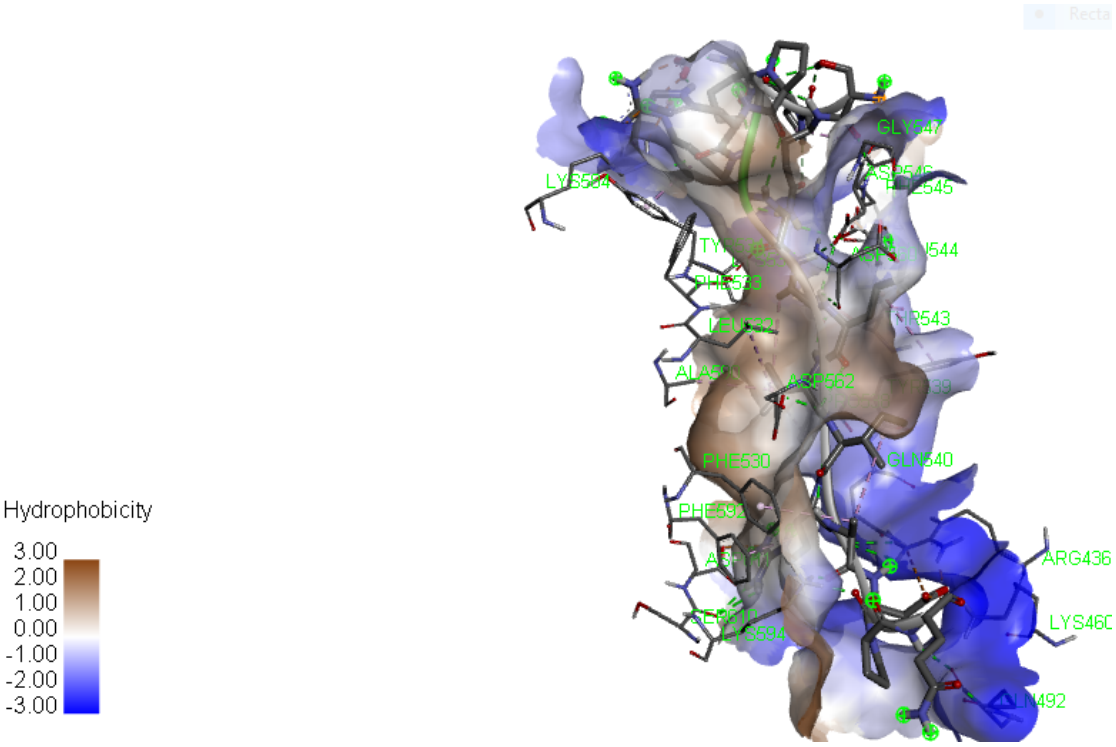

FIGURE A12: INSRD1-IRF2D11

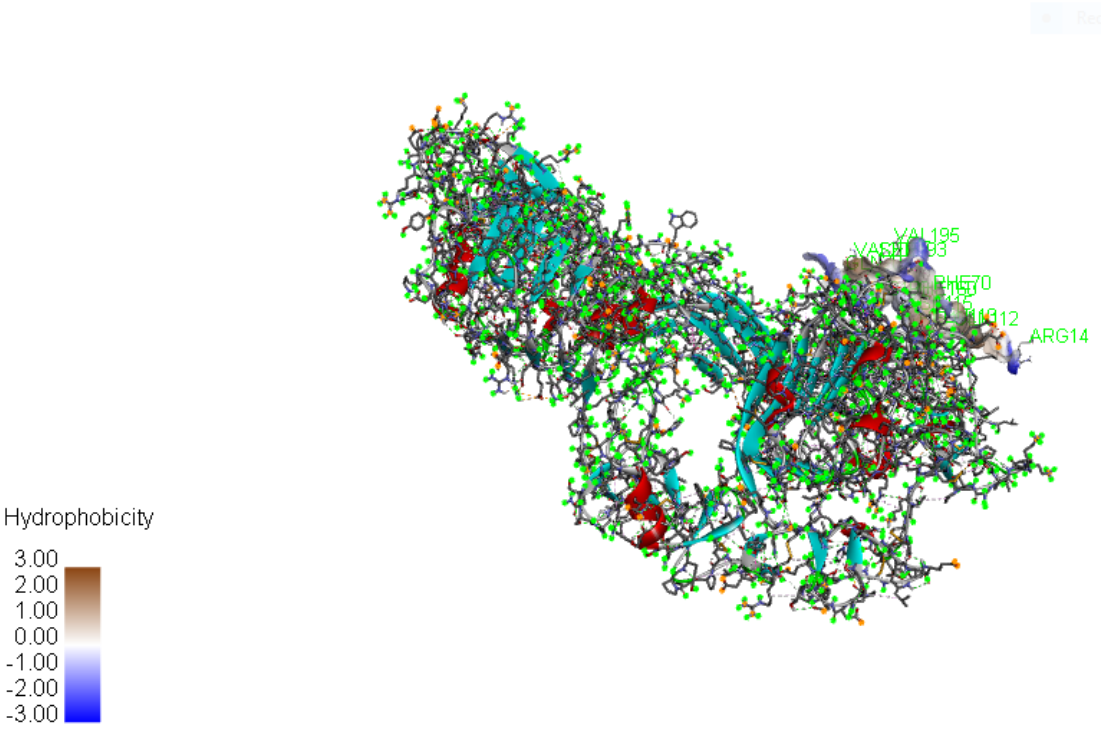

FIGURE A13: IRS1D1-INSRD1

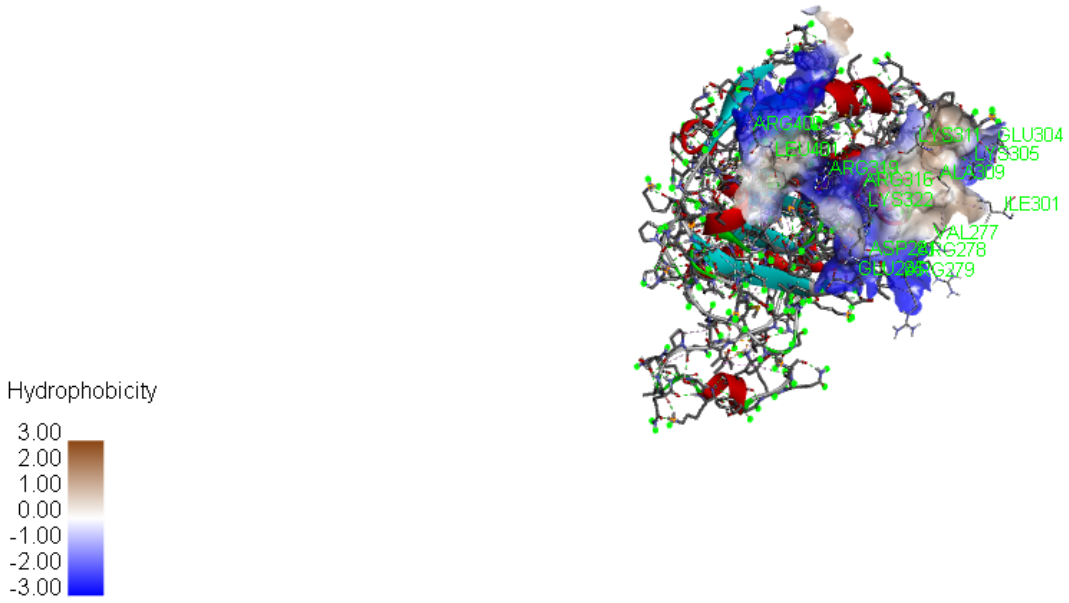

FIGURE A14: KPYM-PRDX1

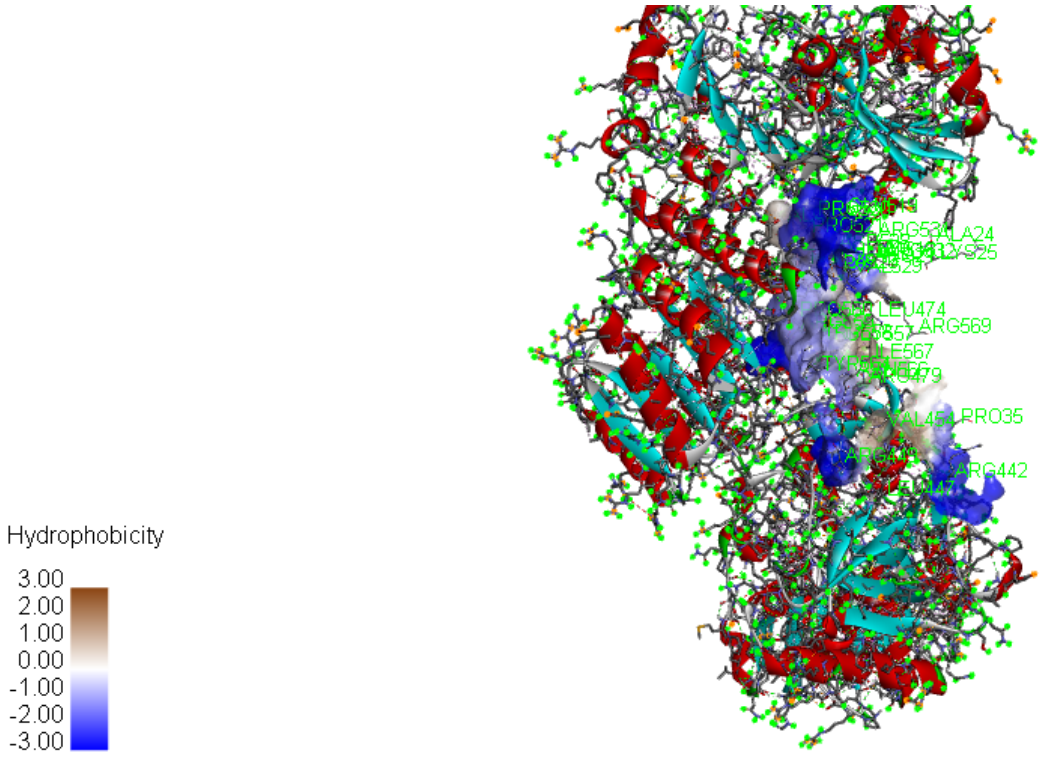

FIGURE A15: PKLR-FSND1

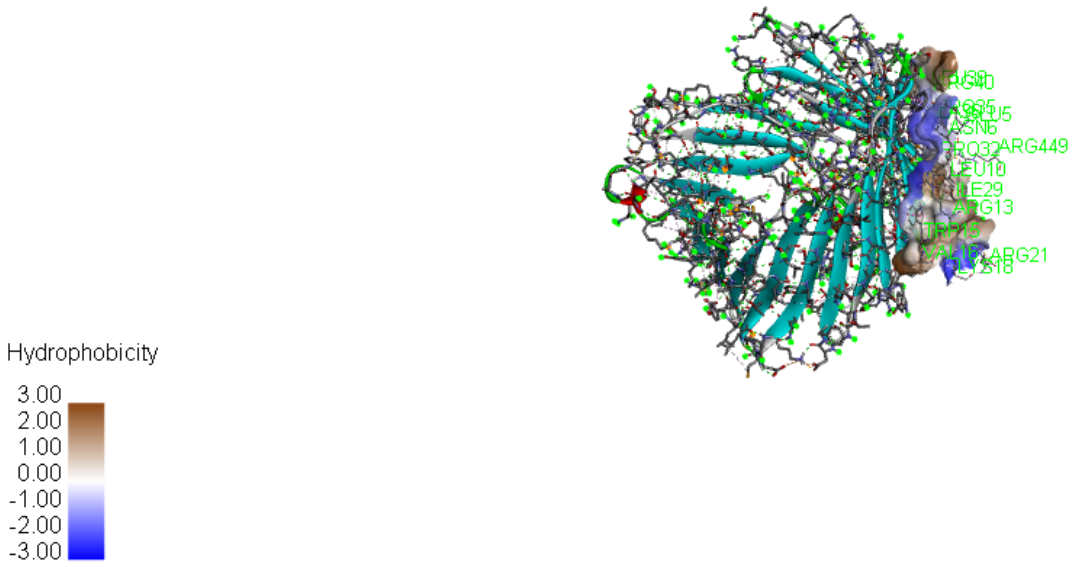

FIGURE A16: PKLR-VDAC2

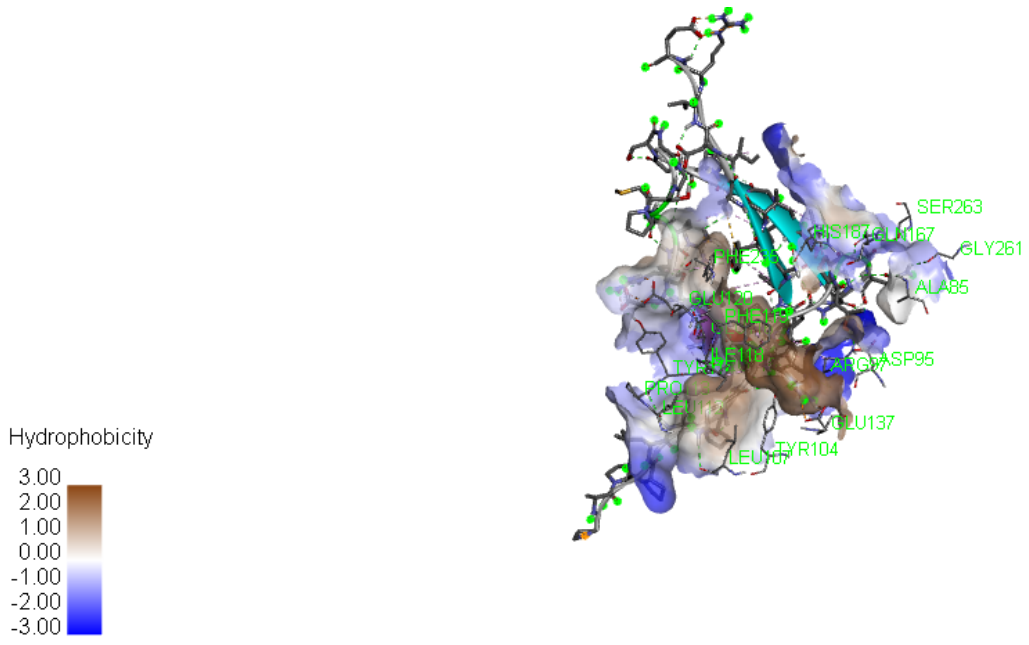

FIGURE A17: SIRT1-IRS2D1