"*In silico* screening of novel drug candidates for diabetes mellitus"

By

Shabana Bibi
Reg. No.1456505

A PhD thesis submitted to the
Division of Environmental and Life Engineering,
Graduate school of Engineering
Maebashi Institute of Technology
Japan
June 2017

TABLE OF CONTENTS

ABSTRACT

Diabetes mellitus is known as blood-sugar disease. The pancreas fails to perform its appropriate function to stimulate insulin production in diabetic patients. The prevalence of type 2 diabetes mellitus (T2DM) has increased dramatically during recent decades and now it is a serious global health burden. According to the International Diabetes Federation 2015 report, the ratio of diabetic patients in the world is one out of eleven adults. Diabetes mellitus and its related complications are major causes of death in various countries.

Most diabetes medicines nowadays available and have approval from FDA (United States Food & Drug Administration), but unfortunately, they could not approach satisfactory levels of blood sugar (glucose) in patients suffering diabetes mellitus and possess numerous adverse effects. Thus novel classes of anti-diabetic drugs are required. The efforts established by computer-aided drug design (CADD) are desirable because the CADD techniques can screen numerous available databases to produce novel and effective virtual candidates and decrease the time and costs to develop new drugs. The computer-aided drug design, especially virtual screening, is a widely-used technique for lead identification and lead optimization. The contribution of CADD techniques in the identification of antidiabetic agents has been discussed in this dissertation.

Most of the diabetes patients cannot afford diabetic medicine in low-income countries and prefer to eat a healthy diet or some alternative low-priced plant-based products. The use of alternative medicine is increased in the world for lowering blood glucose in diabetic patients. While some highly developed countries people prefer plant-based treatments because they are safe and effective with few side effects.

Protein tyrosine phosphatase non-receptor type 1 (PTPN1) inhibitory drugs for T2DM are a hot research target because to inhibit PTPN1 could efficiently ameliorate insulin resistance with normal plasma glucose level in patients of T2DM.

I identified novel antidiabetic agents along with knowledge of plant extracts which possess antidiabetic activity by computer-aided drug design methods. I concluded that the antidiabetic agents show the appropriate mode of interactions with *Canavalia ensiformis* protein; hence it proved their mechanism of action as controller of diabetes by stimulating insulin secretion. The identified lead and designed analogs based on it can be recommended for laboratory tests to confirm their antidiabetic activity. While the plant extract isosilybin has the possibility to become a PTPN1 inhibitor with antidiabetic activity. The isosilybin can be recommended for laboratory tests and further analyses to confirm its activity.

In chapter 1, I introduced the background and current status of CADD for diabetes mellitus, my research goals and the strategies used in this dissertation.

In chapter 2, by computational analysis of *Canavalia ensiformis* protein, I demonstrated that it conserved amino acid sequence homologous to human insulin protein, and it is also evident from the literature review that leguminous plants contain the insulin-like sequence homologous to animal insulin. The plant insulin (UniProt ID: Q7M217) used as alternative source of human insulin showed its mechanism of action in terms of optimal binding mode with available antidiabetic drugs. A biphenyl derivative was screened as a lead compound (WO2007067614) and designed its analogs. Molecular docking analyses showed that four analogs are recommended as antidiabetic agents with suitable drug-like properties as compared with a standard antidiabetic drug (aleglitazar).

In chapter 3, plant-derived PTPN1 inhibitors possessing antidiabetic activity were used for pharmacophore model generation. The pharmacophore-based screening of plant-derived compounds of the ZINC database was conducted using ZINCpharmer; screened hits were assessed to evaluate their drug-likeness, pharmacokinetics, detailed binding behavior and aggregator possibility. The crystal structure of PTPN1 (PDB ID: 3EAX) was used as a molecular target for docking analyses of screened dataset. Through the virtual screening and in silico pharmacology protocols ZINC30731533 (isosilybin) was identified as a lead compound with optimal properties.

In chapter 4, I sum-ups the achievement and originality of this research work and reviewed the integration of computational methods used to produce fruitful results in the discovery of antidiabetic drugs and clarified my research outcomes warrant new protocols in the design/discovery of potential drug-like virtual hits based on the available biological data. It concluded with the significant aspects of the current research scheme in the area of drug discovery of plant-derived proteins and compounds for future functional food and medicinal research.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ADMET	Absorption, Distribution, Metabolism, Excretion and Toxicity
AR	Aromatic ring
CADD	Computer-aided drug design
CYP1A2	Cytochrome p4501A2
CYP2C19	Cytochrome p4502C19
CYP2C9	Cytochrome p4502C9
DM	Diabetes mellitus
EMA	European medicines agency
FDA	Food and drug administration
FGI	Functional group inter conversion
HBA	Hydrogen bond acceptor
HBD	Hydrogen bond donor
HR	Hydrophobic region
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory constant
LBDD	Ligand based drug design
logP/clogP	Partition coefficient between n-octanol and water log (Coctanol/Cwater)
MW	Molecular weight
NMR	Nuclear magnetic resonance
PAINS	Pan assay interference compounds
PDB	Protein data bank
PSA	Polar surface area
PTP1B	Protein tyrosine phosphatase 1B
PTPN1	Protein tyrosine phosphatase non receptor type 1
QSAR	Quantitative structure-activity relationships
QSPR	Quantitative structure-properties relationships
RB	Rotatable bonds
RMSD	Root-mean-square deviation
SAS	Synthetic accessibility score
SBDD	Structure based drug design
SMILES	Simplified molecular-input line-entry system
T1DM	Type 1 diabetes mellitus
T2DM	Type 2 diabetes mellitus
TC	Tanimoto coefficient

CHAPTER 1

INTRODUCTION

1.1. Diabetes mellitus

Diabetes mellitus (DM) is a group of diseases that result from high levels of blood glucose and that depend on insulin production and action. It involves multiple disorders of abnormal carbohydrates, lipids and protein metabolism [1]. People with diabetes may develop serious complications such as heart disease, stroke, kidney failure, blindness and premature death. DM is a diverse and complicated disorder that is characterized by persistent hyperglycemia. It has been called a “third killer” of human health [2]. Hypoglycemic medication is used to lower the blood sugar level in the body or to treat other severe symptoms of DM. These medications can be categorized into insulin and insulin preparations, which are used only parenterally and hypoglycemic medicine that can be administered orally [3].The 2014 National Diabetes Statistics Report revealed that from 2010 to 2012, the number of American diabetic patients increased from 25.8 million to 29.1 million, and that the DM prevalence rate for adults aged 20 years and older increased from 11.3% to 12.3% [4]. The International Diabetes Federation recently reported that the number of people with diabetes is expected to rise from 382 million to 592 million by 2035. Most people with diabetes live in low and middle-income countries [5, 6].

1.1.1. Types and treatments for diabetes mellitus

There are two most important categories of diabetes mellitus (DM); type 1 known as T1DM and type 2 known as T2DM but there is another third type is diabetes known as gestational diabetes belongs to pregnant women's.

T1DM is an autoimmune disorder in which the immune system is activated to terminate the pancreatic cells function to produce insulin [7, 8]. T1DM is usually ten to fifteen percent of all type of diabetic cases [8]. Its indications are frequent and also life-threatening. Its diagnosis is quite rapid and managed with insulin injections only depends upon the condition of patients. T1DM does not depend on the lifestyle, but if someone has T1DM, regular diet and exercise can reduce the chance of development of other complications e.g. damage to kidney, limbs, and eyes [7].

T2DM is a progressive disorder in which body develops resistance to regular insulin functions and losses its capacity to regulate sufficient insulin in the pancreas [7-9]. T2DM is related with risk factors; unstable lifestyle, genetic and family history. Usually, eighty-five to ninety percent of all diabetic cases belong to T2DM. There are no specific indications; normally situation can go undetected and being realized in old age. There is currently no treatment for T2DM; which can manage the condition properly but healthy food, lifestyle adaptations, and proper medicine can improve the situation to decrease the risk of development into progressive complications especially cardiovascular disorders [8, 9].

Gestational diabetes is in between five percent to ten percent cases found in the pregnant women. Usually, in the initial situation, it can manage with the regime of healthy food and physical exercises. But sometimes it is managed with insulin injections

in the period of pregnancy and ended this situation after delivery of baby but the risk is still there for baby and mother both to develop T2DM in rest of their life [7].

Complications of DM are a frequent heart attack, stroke, and the collapse of blood vessels, kidney diseases, nervous disorders, eye infections, and pregnancy complications [9, 10].

Although there are antidiabetic medications currently approved by the U.S. FDA to treat patients with type 2 diabetes, most do not achieve appropriate glycemic control, and some have severe side effects. Successful treatment of type 2 diabetes, therefore, requires new drugs with improved mechanisms of action. In our review, I describe the use of computational tools for the discovery and design of new anti-diabetic drugs that are not currently approved, but that may lower glucose levels and decrease the risk of hypoglycemia, which is a major difficulty to control level of glucose and important for treatments which increase levels of insulin [6].

Table 1.1: Approved drugs for type 2 diabetes.

Therapeutic Class of Compound	Mechanism of Action	Approved Drugs	Date of First Compound Approved	Adverse Effects and/or Comments
Biguanide	Increases insulin sensitivity, suppresses glucose production in the liver	Phenformin, Metformin	1957 (EMA), 1995 (FDA)	Nausea, vomiting, diarrhea and flatulence; If taken with meals, avoid use in patients with renal or hepatic impairment or with CHF, because of increased risk for lactic acidosis.
Second generation sulfonylureas	Stimulate insulin secretion from the pancreas	Glimepiride, Glipizide, Gliclazide, Glibenclamide(Glyburide), Gliquidone	Glibenclamide (Glyburide):1969 (EMA),1984 (FDA)	Hypoglycemia and weight gain
Insulin: regular human insulin, NPH insulin,	Helpful in lowering blood glucose	Regular insulin, Bovine insulin	Regular insulin:1982 (FDA),1984 (EMA), Bovine insulin:1922	Severe hypoglycemia and weight gain. A new administration

insulin aspart, insulin lispro, insulin glargine, insulin detemir, insulin levemir				form of inhaled insulin has been recently approved (2014) (Afrezza) for type 1 and type 2 diabetes.
Alpha-glucosidase inhibitor	Delay complex carbohydrate absorption	Acarbose, Miglitol, Voglibose	Acarbose:1991 (EMA),1995 (FDA)	Flatulence, diarrhea, abdominal pain. Less effective than other agents, it is considered in all elderly patients with mild diabetes.
Glinides	Stimulate insulin secretion from the pancreas	Repaglinide, Nateglinide	Repaglinide:1998 (EMA), 1997 (FDA)	Hypoglycemia and weight gain; the precaution is to take with meals to control rapid onset. Some partial agonists are in clinical trials. An example is INT131 (previously known as AMG-131), which progressed through the phase 2 clinical trials. C333H is a novel partial agonist in preclinical development.
Thiazolidinediones	Increase peripheral tissue insulin sensitivity	Pioglitazone, Rosiglitazone	Rosiglitazone:1999 (FDA), 2000 (EMA)	Edema, it should be avoided in patients with heart failure. These agents can cause or exacerbate CHF contra indicated in patients with NYHA class III or IV heart failure.
Amylin analogue	Slowing of gastric emptying, suppression of elevated glucagon, stimulation of satiety.	Pramlintide	Pramlintide:2005 (FDA)	Approved for type 1 and 2 diabetes, nausea, hypoglycemia when combined with other anti-diabetic drugs (*e.g.* insulin).
GLP-1 agonists	Stimulation of glucose dependent insulin release, suppression of elevated glucagon levels,	Exenatide, Liraglutide, Exenatide extended-release, Lixisenatide, Albiglutide, Dulaglutide	Exenatide: 2005 (FDA), 2006 (EMA), Liraglutide:2010 (FDA), 2009 (EMA), Exenatide ER: 2012 (FDA), Lixisenatide: 2013 (EMA),	Only injectable drug, weight loss, nausea, vomiting, diarrhea and acute pancreatitis. Risk for medullary thyroid cancer, pancreatitis

	reduction of gastrointestinal motility		Albiglutide: 2014 (FDA, EMA), Dulaglutide: 2014 (FDA)	or pancreatic cancer. Not confirmed in clinical trials by FDA and EMA. Many oral GLP-1 agents are under trail for TD. ORMD-0901 NN9924, NN9926, NN9927, NN9928, TTP054, ZYOG1, NN9924, ORMD-0901, TTP054 have reached Phase 2
DPP4 inhibitor	Slow inactivation of incretin hormones	Sitagliptin, Vildagliptin, Saxagliptin, Linagliptin, Alogliptin	Sitagliptin:2006 (FDA), 2007 (EMA), Vildagliptin:2008 (EMA), Saxagliptin: 2009 (FDA, EMA), Linagliptin:2011 (FDA), Alogliptin: 2013 (FDA)	Risk for medullary thyroid cancer, pancreatitis or pancreatic cancer. Not confirmed in clinical trials by FDA and EMA. Few agents are under clinical: ARI-2243 (Phase 1), Teneligliptin (Phase 1), Omarigliptin (Phase 3), Trelagliptin (Phase 3)
Bile acid sequestrant	Possibly activation of the farnesoid X receptor / bile acid receptor	Colesevelam	Colesevelam:2008 (FDA)	Constipation, nausea and dyspepsia. Primary a lipid lowering drug with additional glucose lowering effects. Mechanism of action for diabetes control is unknown.
Dopamine agonist	Central modification of insulin resistance	Bromocriptine	Bromocriptine:2009 (FDA)	Orthostatic hypotension, nausea. Mechanism of action for diabetes control is unknown.
SGLT2 inhibitor	Reduction of the renal threshold for glucose excretion	Dapagliflozin, Canagliflozin, Empagliflozin	Dapagliflozin:2012 (EMA), 2014 (FDA), Canagliflozin:2013 (FDA), Empagliflozin:2014 (FDA, EMA)	Genital infections and possible diuretic effects. Other favorable effects of SGLT2 inhibitors include a reduction in both body weight

			and blood pressure. Still some agents are under trails to improve the effects, *e.g.* Ertugliflozin (Phase 3), EGT0001442 (Phase 2), luseogliflozin (TS-071) (Phase 1).

The most important function of anti-DM drugs is to stimulate the insulin via pancreatic cells and improve sensitivity of cells toward insulin hormone and it's normally utilized along insulin. Various therapeutic classes of DM medications are present in the market and the reason to choose a medicine based on the type of DM (age factor, situations of the diabetic person and other critical issues). Twelve classes of anti-DM drugs are currently available and approved (table 1.1).

There are ten more classes that have new mechanisms of action, which are in various phases of clinical trials shown in table 1.2. These therapeutic classes provide novel compounds that show improved safety and tolerability profiles for known adverse effects related to marketed agents such as gastrointestinal side effects, hypoglycemia risk and weight gain. Further optimization and clinical studies will help to generate a useful drug in a short period of time from these compounds. These agents may potentially control glucose levels and improve outcomes in patients with T2DM. I expect computer-aided drug design techniques to contribute in improvement of the compounds and acceleration of novel diabetes drug development [6].

Table 1.2: Drugs under development for type 2 diabetes

Therapeutic Class of Compound	Mechanism of Action	Adverse Effects and/or Comments
11 beta-hydroxysteroid	Improves lipid profiles,	Risk of glucose intolerance, insulin resistance, dyslipidemia,

dehydrogenase type 1 inhibitor	fasting glucose levels and hepatic insulin sensitivity	and hypertension. Some agents are under trails. There are no long-term studies available beyond 3 months: PF-00915275 (Phase 1), INCB13739 (Phase 2), MK-0916 (Phase 2).
Glycogen phosphorylase inhibitor	Potential target of hepatic glucose production	In early development: oral agents have shown promising results in animals and humans.
Glucokinase activator	Activate key enzyme to increase hepatic glucose metabolism	Hyperlipidemia, hyperglycemia and Cardiovascular risk. Several drugs are currently in phase 2 clinical trials: PF-04937319 (Phase 2), AZD1656 (Phase 2).
G protein–coupled receptor 119 agonist	Activation induces insulin release and increases secretion of glucagon-like peptide 1 and gastric inhibitory peptide	Low potential for hypoglycemia. Several agents are in clinical trials: DS-8500 (Phase 2), MBX2982 (Phase 2), GSK1292263 (Phase 2).
PTP1B/PTPN1 inhibitor	Negatively regulates insulin in a signal pathway that helps to increase leptin and insulin release.	Reduces adipose tissue storage of triglyceride under conditions of over-nutrition and was not associated with any obvious toxicity. No weight gain, indicating another substantial advantage for diabetic patients, who are frequently obese and at high cardiovascular risk. Some agents are currently in clinical trials: TTP814 (Phase 1/2), ISIS-PTP1BRx (Phase 2).
Glucagon-receptor antagonist	Block glucagon from binding to hepatic receptors, thereby decreasing gluconeogenesis.	Low potential for hypoglycemia. Several agents are under trails: BAY 27-9955 (Phase 1), LGD-6972 (Phase 1), MK-0893 (Phase 2), MK-3577 (Phase 2), LY-2409021 (Phase 2).
Hepatic carnitine palmitoyltransferase 1 (CPT1) inhibitors	CPT1 is a mitochondrial enzyme involved in fatty acid metabolism makes CPT1 important in many metabolic disorders such as diabetes. Inhibition decreases gluconeogenesis	Since its crystal structure is not known, its exact mechanism of action remains to be determined. Only limited data available. One agent is in clinical trials: Teglicar (Phase 2).
Diacylglycerol acyltransferase (DGAT)-1 inhibitors	Inhibition of DGAT-1 enzyme responsible for final step in triglyceride synthesis – weight loss, improved insulin sensitivity, decreased cholesterol and triglycerides	Gastrointestinal side effects (nausea, diarrhea, vomiting). Several agents are in clinical trials: DS-7250 (Phase 2), P7435 (Phase 1)
Sirtuin1 (SIRT1) activators	Enhance glucose production and lipid metabolism, insulin signaling and pancreatic insulin secretion.	SIRT1activation improves glucose homeostasis and insulin resistance. Very early development. One agent is in clinical trials: SRT3025 (Phase 1)
Glucocorticoid receptor antagonist	Liver specific glucocorticoid receptor antagonist; reduction of hepatic glucose production.	Early in development, Only limited data is available. One agent is in clinical trials: ISIS-GCGRRx (Phase 1).

1.1.2. Protein tyrosine phosphatase non-receptor type 1 (PTPN1)

Some of important drugs are currently under development for T2DM. PTPN1 could be one upcoming possible oral therapeutic option for glycemic control and weight

management. PTPN1 knockout mice shown anti-DM activity by subsequently normalizing blood glucose levels and improves insulin sensitivity [11]. PTPN1 inhibition is a novel approach for the treatment of DM and PTPN1 inhibitors represent attractive medicinal activity in experimental studies for DM, obesity and cancer treatment [12, 13]. Recent studies demonstrate that biochemical and pharmacological confirmation for PTPN1 as important negative regulator of insulin along with leptin hormone. PTPN1 mechanism of action for T2DM is shown in figure 1.1 [13].

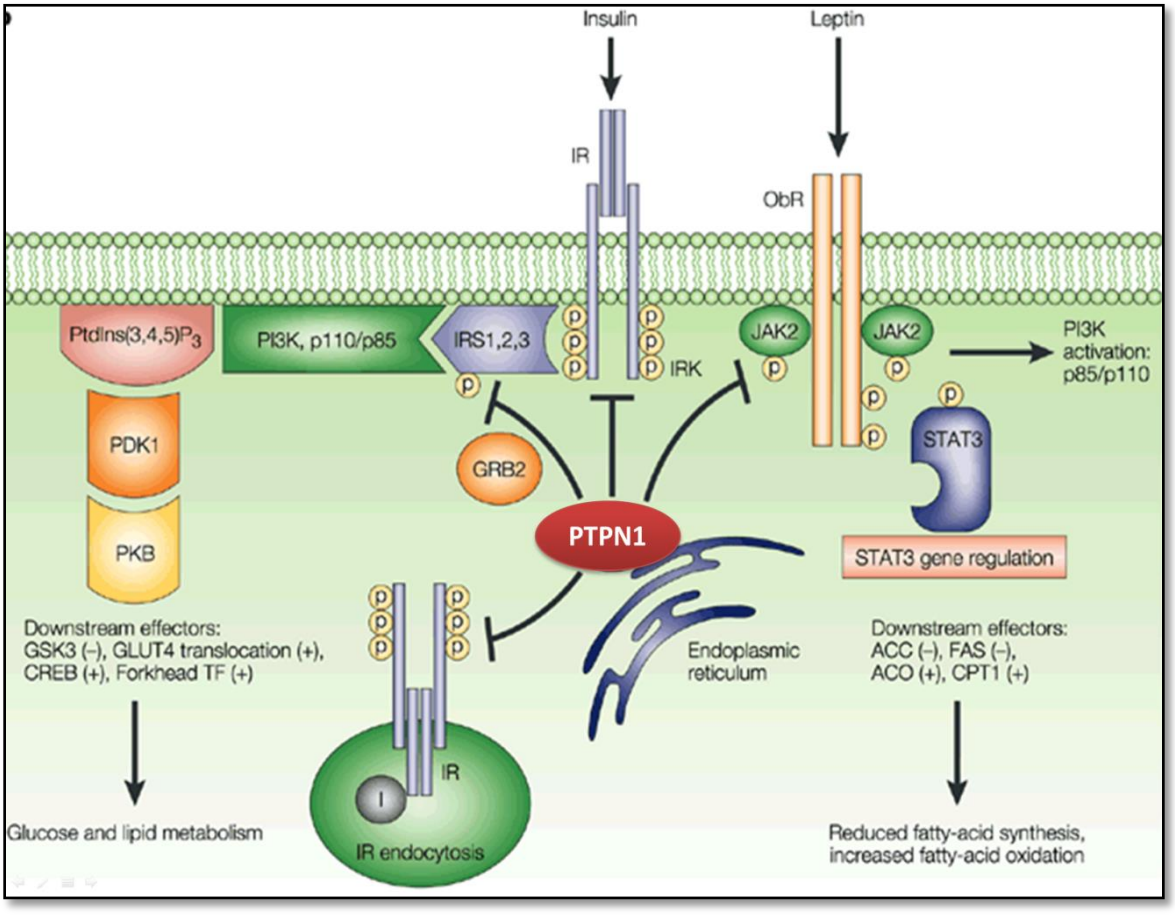


Figure 1.1: Protein tyrosine phosphatase non-receptor type 1 (PTPN1) in insulin and leptin signaling pathway [13].

Insulin binds to its receptor (IR) and induces conformational changes to activate insulin receptor kinase domain (IRK) in cytoplasmic part of IR. Activated receptor undergoes autophosphorylation of tyrosine residues and phosphorylate insulin receptor substrate (IRS) activates phosphatidylinositol-3-Kinase (PI3K) via interacting with p85 subunit and activates the catalytic subunit p110. Activation of P13K encourages downstream effectors which monitor the translocation of glucose transporter 4 (GLUT4) and cellular glucose endorsement in muscle and deactivates glycogen-synthase kinase 3 (GSK3). Leptin hormone is cooperative in metabolic homeostasis along with PTPN1. Leptin binds to its receptor (obR) and proceeds phosphorylation of Janus kinase 2 domain's (JAK2), and it stimulates the JAK signals to STAT pathway and perhaps the P13K pathway (mechanism not clear). STAT3 pathway start by JAK2 phosphorylation encourages translocation of STAT3 towards the nucleus. STAT3 encourages gene reactions which reduce transcription of acetyl coenzyme-A carboxylase (ACC), decreasing malonyl CoA in addition to fatty acid synthesis, while accumulative fatty acid oxidation. Cytosolic PTPN1 dephosphorylates insulin receptors and leptin receptors to terminate the process [12, 13]. Hence, slight variations in the expression or action of PTPN1 enzyme with respect to insulin receptor could disturb insulin signaling and contribute to insulin resistance in T2DM patients.

1.1.3. Natural source of anti-diabetic medication

Bioactive natural products with therapeutic potential for DM are abundantly available and some are beyond exploration by conventional methods. Natural medicines are usually safe, inexpensive, and easily accessible while sometimes it's more efficacious

than a synthetic medicine [14]. Several databases of natural drug-like compounds are useful to find important lead compounds for many disease treatments. Small molecules and secondary metabolites have been economically designed and synthesized by nature for the benefit of evolution; in other words, they have been evolutionarily selected [15]. Natural products contain various types of biologically relevant privileged structures that have saved millions of lives, which render them a continuous source of inspiration for the discovery of new drugs [16]. These plant-based compounds assist as excellent initial points for exploring biologically applicable chemical space [17]. Therefore, identification of natural products that are capable of modulating protein functions in pathogenesis-related pathways is the heart of drug discovery and development [18]. Until now, distinct natural products have been chemically modified and driven to become Food and Drug Administration (FDA) approved drugs [19]. Natural products and their derivatives in 1981 to 2010, accounted for 74.8% of all drugs approved by the FDA [20].

Merits of plant-based medicine have been proved in development of numerous drugs. Metformin FDA approved drug used from long time drug for management of T2DM, is derived from the guanidine which were obtained from *Galegine officinalis* [20]. Various studies in investigation of plant-based antidiabetic agents are discussed in details [21]. Some of these plant-based medicines are better extracted and use in crude form as is the common practice in traditional anti-DM medicine. In addition, the combined effect of the constituent anti-DM agents could be better than a single agent acting alone.

It is necessary to get that food which gives you maximum vitamins and minerals required for good health. Various research displays that person affected by DM are more expected to use supplements as medicine than the person without DM. Summary of

National Health Survey demonstrated that 22 percent people affected by DM use herbal therapies. While additional research confirmed that 31 percent DM patients use dietary supplement. Various ethnic individuals in the world; Hispanics or Latino, African-Americans population and Native Americans society also has routine of eating additional dietary supplements [22].

Insulin-like material glucokinin was present in plant sources and microbes that exhibited similar functions to those of insulin in vertebrates [23]. The presence of insulin-type peptides confirms in bacteria and fungi also [24, 25]. Ample research has demonstrated insulin-type molecule is present in *Momordica charantia* [26]. They showed the related features of a protein of animal insulin in plants. Xavier-Filho et al. retrieved information that suggested insulin was present in plants. Their results suggested that the insulin-type protein with the conserved sequence as of bovine insulin was expressed in plants family *Leguminosae.* These old-style treatments are hopeful as anti-DM medicine. So it is an urgent need to shift the focus of research on the way to the plant-based origin of insulin and it should elicit less adverse outcomes as compared to commercially available drugs for hyperglycemia and DM [27].

1.2. Computer-aided drug design for diabetes mellitus

1.2.1. Status of computer-aided drug design for fatal diseases

The average cost of launching a new drug onto the market is estimated to 1.8 billion dollars [28], and few drugs make it to the market. From 1999 to 2008, only 50 compounds were approved by the FDA in the U.S., out of which 17 were identified as arising from target-based drug design methods [29]. This suggests that experimental

libraries made by conventional high-throughput screening take more time, and that the results are not always efficient for developing novel drugs [6].

Computer-aided drug design provides advantages for experimental findings, mechanisms of action and new suggestions for molecular structures for new synthesis, and it can help in making cost-effective decisions before the costly process of drug synthesis begins. Numerous compounds were discovered and/or optimized using computational methods and they have reached the clinical stage of drug development or have even gained U.S. Food and Drug Administration (FDA) approval [30, 31]. Computer-aided drug design can increase the hit rate of novel anti-diabetic drug-like compounds because it better uses a large chemical search space to find a suitable target compared with traditional high-throughput screening and combinatorial chemistry. Several studies have compared conventional high-throughput screening and virtual screening, and virtual screens had hit rates of tenfold to 1700-fold those of conventional screening [6, 32-36]. Computational methods are required because the amount of biological data has increased and manual screening against such data requires much time and human resources. Computer-aided drug design methods have been used in the development of therapeutic molecules for over three decades. The increasing use of this method is reflected in the number of publications about computer-aided drug design in fatal diseases. Publications on computer-aided drug design for the top 3 most fatal diseases [6, 37, 38] are shown in Figure 1.2.

Diabetes has the third most papers published on computer aided drug design, but the number of published papers for diabetes was half of what it was for cancer or HIV. Thus, there is still room for improvement in antidiabetic drug design with the help of computational techniques [6].

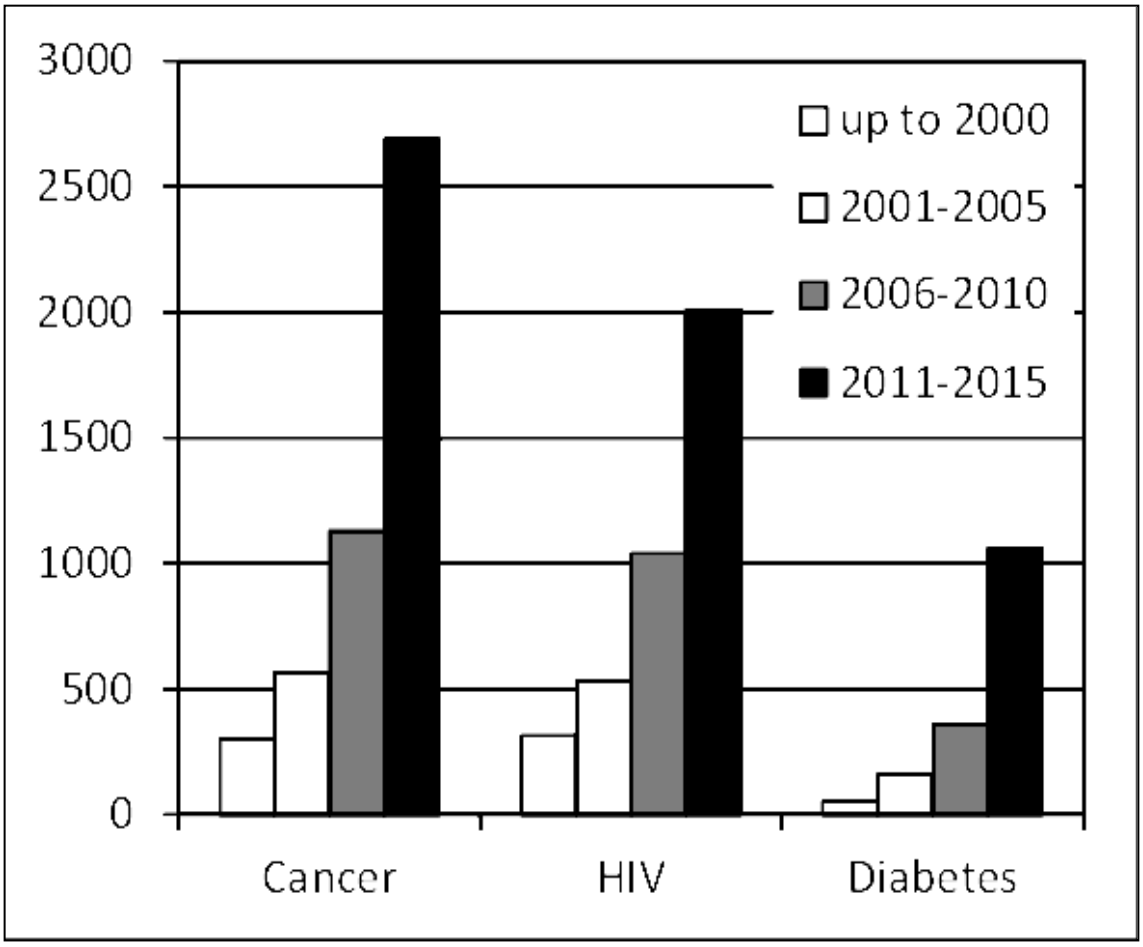


Figure 1.2: The number of publications related to computer-aided drug design and diseases. Key words used in the Google Scholar search (scholar.google.com) were as follows: computer-aided drug design and disease; e.g. diabetes.

1.2.2. Concepts of drug design, discovery and development

The basics regarding drug design programs are identification, design of compounds or dataset of compounds that can generate the preferred medicinal properties. The success rate of any drug design scheme depends on the creativity and interplay of different techniques at the same ground includes biotechnology, bioinformatics, genomics, genetics, proteomics, structural biology, pharmacology, medicinal chemistry, and pharmacokinetics [39]. The anti-DM drug design is a complex process that requires expertise from multidisciplinary fields.

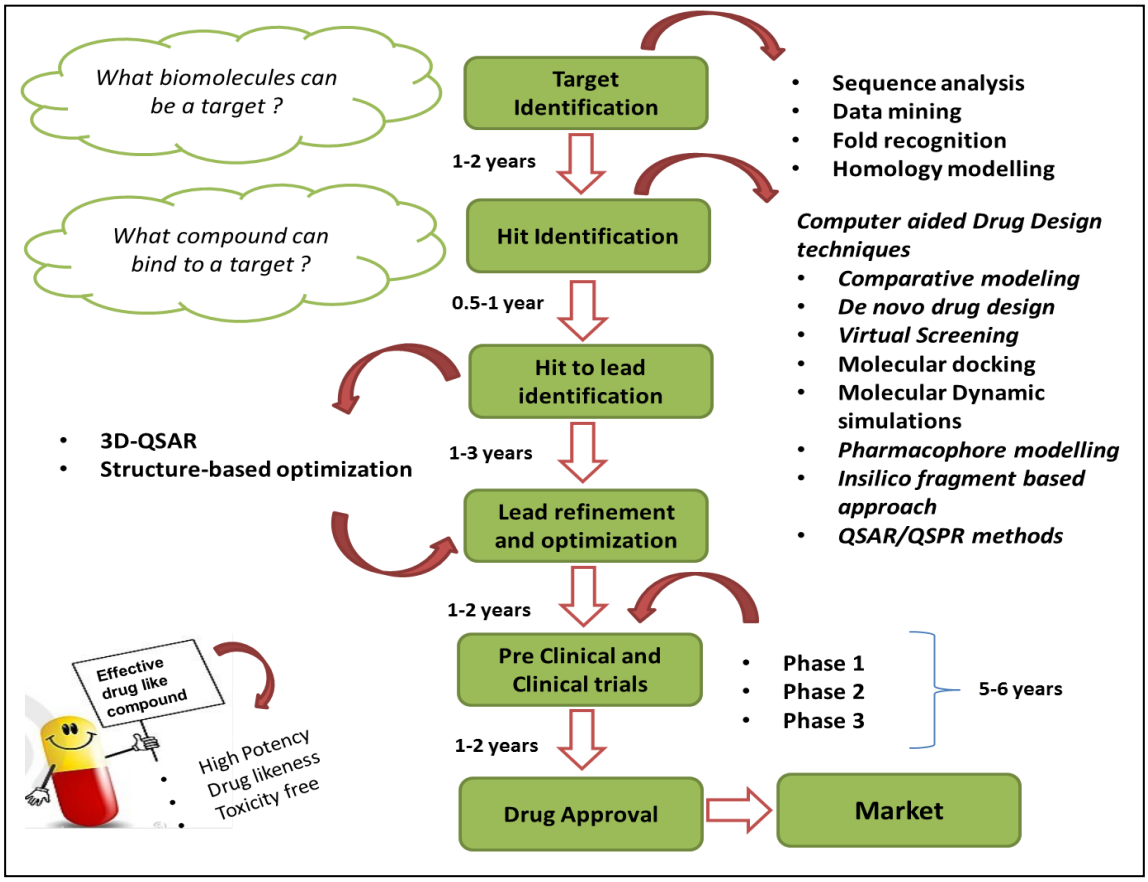


Figure 1.3: Concepts of drug design, discovery and development and impact of computational methodologies.

***Hit= Virtual candidates that can fit to the target binding site.**

*** Lead= a most active virtual candidate with preferred biological activity.**

*** QSAR and QSPR= Quantitative structure activity/properties relationship of chemical compounds.**

The primary phase in the pipeline of drug discovery includes; Selection of a validated drug target. Following various phases of lead identification and optimization, next step is pre-clinical or animal tests, and ultimate phase of clinical trials using human beings [40] shown in Figure 1.3. The identification of a potent drug for diabetes which reaches the appropriate glycemic control is a costly procedure. Usually, a new drug with FDA approval needs approximate 10 years before introducing to market [41]. Possibly, most of the anti-DM drugs are not accepted in the late clinical phase because it exhibits some toxic effects or due to less efficacious. It has been stated that total cost of each drug discovery and development process is almost US $2.6 billion [42]. A comparatively cheap explanation is to use computational methods which can be used to rank target proteins and drug candidates that have the anticipated properties to ultimately develop an efficacious drug. Actually, the early twenty percent of the procedure of drug development is contributed by computer-aided drug design. Drug design to develop effective anti-DM drugs is extremely complex and expensive practice with unpredictable outcomes. To reduce these problems, CADD becomes gradually popular owing to low cost and least investment in manpower by using database resources of chemical compounds (Figure 1.3). CADD methods are essential to aid identification of conventional drug targets involved in insulin signaling pathways, design of new lead compound and structural modification of lead compound to improve aspects of its binding affinity, pharmacokinetic and pharmacodynamics parameters.

The design typically features small molecules that can interact with target protein/enzymes and inhibits their function. The distinction stems from whether a 3D structure of a protein is available and used in the design process. Structure-based methods of drug design can proceed with the only existence of target protein structure and modeling software for building ligands in the projected binding pocket. However, further insights delivered by the assessment of molecular energies for the bonding process are the center of current structure-based methods of drug designing [43]. Ligand-based methods do not require 3D structure of protein but analyze the structure-activity relationship of chemical compounds that have been tested in the biological assay for its target function. One seeks patterns in the assay results to suggest potential modifications of the compounds yield enhanced activity. The upside is that a target structure is not required; the downside is that substantial activity data are needed [44].

1.2.3. Current computational techniques

Drug development requires extensive clinical testing and is a costly process. There are two main phases involved in creating a new drug: the discovery phase and the clinical testing phase. *In silico* approaches, including virtual high throughput screening, and de novo structure-based rational drug design, has been established as tools in the discovery phase [6]. Virtual screening emerged for finding novel drug-like compounds. *In silico* virtual screening has become a reliable, cost effective and time-saving technique that is complementary to in vitro screening for the discovery and optimization of potent lead and hit compounds. There are two broad categories of screening techniques; the ligand-based virtual screening and receptor-based virtual screening, to select candidate compounds that are likely to interact favorably with the target binding sites from a

chemical database. The three-dimensional structure of protein or protein-ligand complex is helpful in lead identification using molecular modeling. Quantitative structure-activity relationship (QSAR), pharmacophore and biological assays can be helpful to optimize and design new leads. Structure-based drug design helps to provide potent and significant compounds more productively in the drug discovery process. Structure-based virtual screening is used more frequently than the ligand-based virtual screening (322 to 107 studies) [6, 45].

Virtual screening uses high-performance computing to screen large chemical databases and prioritize compounds for synthesis. Current databases allow rapid virtual screening of up to 100,000 molecules per day using parallel computing techniques [46]. The databases of three-dimensional structures directly available for virtual screening are [6]:

- Advanced Chemistry Development [47]
- InfoChem GmgH database [48]
- MDPI database [49]
- National Cancer institute open database compound [50]
- Thomson index chemicus database [51]
- Tripos discovery research screening libraries [52]
- ZINC database [53]

They contain libraries that have been experimentally determined. Several computer programs have been developed and used in research leading to drug discoveries for various diseases. They are based on computational techniques of drug design, using

different algorithms and scoring functions. Some of the programs for virtual screening and docking studies are [6]:

- AutoDock[54]
- CLC drug discovery work bench [55]
- Dock [56]
- FlexX [57]
- FRED [58]
- Glide [59]
- GOLD [60]
- MOE [61]

Several remarkable drug design applications using docking tools have been mentioned in our review. Pharmacophore modeling, or ligand-based virtual screening, is an efficient method to increase hit rates in drug discovery research [6].

- Catalyst [62]
- LigandScout 4.0 [63,64]
- MOE (pharmacophore module) [61]
- Phase [65]

These are widely used computer programs for pharmacophore elucidation and virtual screening. The effective pharmacophore models depend on two factors: the definite understanding and placement of pharmacophoric features, and the alignment method used for overlaying the three-dimensional pharmacophore model with a set of ligand compounds of screened data [6, 66].

QSAR methods can be used to optimize lead compounds. Modern three-dimensional QSAR methods involve the interaction fields around a molecule by calculating the interaction energy in a grid. The well-known three dimensional QSAR techniques are; comparative molecular field analyses [67] and comparative molecular similarity index analyses [68] to predict activity and correlates the biological dataset of chemical compounds. These approaches calculate molecular properties including steric, electronic, hydrogen bonding, and hydrophobic fields. Some of the programs used in research and that are available for two-dimensional and three dimensional QSAR analyses are [6]:

- CODESSA [69]
- Dragon [70]
- QSARpro-Vlife science [71]
- SYBYL-Xsuit [72]

Another type of program is the versatile and advanced software for molecular modeling and simulation, which has broad applications to many-particle systems, includes [6]:

- AMBER [73]
- CHARMM [74]
- GROMAC [75]

1.3. Motivation

By IDF (International Diabetes Federation) 2015 report; ratio of diabetic patients in the world is one out of 11 adults. Diabetes mellitus (DM) and its related complications are major causes of death in various countries. Despite continuous efforts of the international communities to reduce the impact of DM on poor and developed countries, there is steadily rise in the number of diabetic patient because of high cost and low availability of medications (specially in poor countries).

Available anti-DM drugs approved by FDA could not approach sufficient blood sugar (glucose) levels in patients suffering from DM, and there were many side effects affiliated with these medicines as I mentioned in chapter 1. Therefore, a new class of potential candidates is urgently needed. Efforts established on CADD techniques can mine numerous databases, generate novel and powerful virtual hits, and decrease the time period and cost need for discovery of novel anti-DM drug.

Drug design efforts in this way are most expected for development of potential drugs if the target is novel mechanism of action. Such methods could lead to anti-DM prescriptions with functional and structural difference with respect to available drugs and shows novel approach to reach appropriate glycemic control. As DM is a disease of all poor and developed countries, cost effective technologies have to be used to find the novel and potential entities. I have identified small drug-like compounds that have potential and may helpful in development of new anti-DM drugs.

1.4. Research goals and strategies

By using CADD methods I want to contribute in the successful discovery of novel antidiabetic drug candidates. Number of anti-DM drugs and recombinant insulin are accessible to DM patients, but with severe side effects. My goal is to discover novel candidate compounds which should be safe and harmonious to human body.

Numerous new medicines and their active ingredients are derived from plants because it's cheap and safe source of drugs. Merits of plant-based medicine have been justified through development of some drugs. An example is the metformin, a FDA approved drug, used for a long time for management of Type 2 diabetes mellitus, had been derived from the plant source. The presence of plant proteins whose genomic sequences are similar to those of animal insulin encourages confirming its activity as insulin and evaluating its action with respect to diabetic medicine. It could produce therapeutically significant effects for diabetic patients.

Computer-based screening of large databases has shown compatibility with various *in silico* procedures such as molecular docking and pharmacophore generation. *In silico* drug-likeness and pharmacokinetic estimations adds knowledge to reduce the adverse outcomes of chemical compounds. Contributions of computer-aided drug design approaches in the identification of plant-based anti-DM virtual candidates have been explained in this thesis.

1.5. Thesis outline

Chapter 1 introduces target disease (diabetes mellitus) with its types and complications. Importance of natural products used in diabetes treatment is briefly explained. Status of CADD for top 3 most fatal diseases (Cancer, HIV and diabetes) has been demonstrated. CADD approaches have contributed to successful identification of new anti-DM drug candidates and highlighting currently FDA-approved medicines for DM with the newly discovered diabetes drugs also that appeared in the development phase and could attain the appropriate glucose control and reduce the threat of hyperglycemia, which is the main cause of glucose imbalance and an important concern for anti-DM therapies which enhance insulin production.

Chapter 2 focus on plant insulin protein present in *Canavalia ensiformis* used as the target protein for identification of potential anti-DM agents. Identification of most active compound from a set of eight compounds with desired biological activities for a validated molecular target has explained in detail. Analogs design of lead compound using functional group inter-conversion approach has demonstrated. Molecular docking analyses showed that the four analogs could be used as anti-DM agents. Binding energies and binding interactions of the analogs have been explained in detail.

Chapter 3 focuses on pharmacophore modeling based on the information of known biological activities of plant-based PTPN1 compounds. Shared feature pharmacophore model has been established. Molecular superimposition algorithm works in order to organize the 3D structure of the input dataset in a way that chemical features of compounds located in similar positions in each pharmacophore model. Pharmacophore-based screening of natural compounds of ZINC database has been

conducted. Molecular docking analysis explained binding features of selected drug-like hits with the target protein. Identified hits were assessed for their aggregator potential to compare with previously reported aggregators. By virtual screening and *in silico* pharmacology protocols; identified a lead compound with best results is explained in detail.

Chapter 4 sum-ups the achievement and originality of this research work. This chapter reviews the integration of computational methods used to produce fruitful results in the discovery of anti-DM drugs and explains my research outcomes warrant new protocols in the field of CADD. It concluded with significant aspects of the current research scheme in the area of drug discovery of plant-derived proteins and compounds for future functional food and medicinal research.

CHAPTER 2

LEAD IDENTIFICATION AND OPTIMIZATION OF PLANT INSULIN-BASED ANTIDIABETES DRUGS

2.1. Abstract

Objective: Diabetes mellitus (DM) depends on multiple factors involved in pancreatic disorders and becomes the third leading cause of deaths in humans. The presence of plant proteins whose genomic sequences are similar to those of animal insulin has been demonstrated. I wished to discover anti-DM drugs having high inhibitory activity based on plant protein.

Methods: Computer-aided molecular docking methods were applied using Auto Dock Vina software.

Results: I have selected a plant protein with UniProt identification Q7M217 insulin in *Canavalia ensiformis* as the target protein for DM. I have identified an active lead compound among eight candidate compounds based on significant interactions with protein molecule and half-maximal inhibitory concentration (IC_{50}) values. I have designed four analogs of the lead compound. Molecular docking analyses showed that the four analogs could be used as anti-DM agents with suitable drug-like properties as compared with a standard compound for the treatment of DM (aleglitazar). These analogs can also be used for future studies.

Conclusion: The present study has identified an anti-DM compound, a biphenyl derivative, based on plant insulin. I have designed its analogs using a functional group

inter-conversion approach. Our computer-aided study provided information on binding energies and binding interactions of the analogs to predict their anti-DM activity.

Keywords: Diabetes mellitus, Plant insulin, Lead identification and optimization, Computer-aided drug design

2.2. Introduction

Insulin hormone regulates blood sugar levels. If insulin is not present in the body, cell could not utilize the energy from blood sugar factory to uphold the metabolic events within a body. Frederick Grant Banting and Charles Best (1921) took out insulin from dogs. It was introduced into a 14-year-old male with DM in 1922 as medicine for this disease [76].

Insulin reached to approval by the US Food and administration (FDA) in 1939 [77]. It is used in the homeostasis of blood sugar and lipids, growth and progress of tissue, and responds to elevated levels of glucose and amino acids in blood. It regulates metabolism by tissue-specific mechanisms such as protein phosphorylation and altered functions and shows different gene expression. The physiological discorded of T2DM; by insulin resistance includes pancreatic beta cells, skeletal muscles, liver and fat storing tissues. Similar to insulin secretion and glucagon suppression, the combination of stimulation (high post-meal level of blood glucose) is owned by humans to maintain plasma glucose levels at about ≈5 mmol [77-80]. Glucose-based insulin regulatory mechanism of beta cells has been described [81]. Pancreatic beta cells respond to increased levels of sugar/glucose in plasma by secretion of insulin. Protein-facilitated

glucose transporter 2 on the membrane of beta cells has high Michaelis constant. The maximum rate can be achieved by a system that allows fast equilibration of glucose across the membrane. Glucokinase promotes the phosphorylation of glucose and encourages conversion to a glycolytic cycle, an important step in determining glucose-stimulated insulin secretion [81]. The released insulin connects with insulin receptor (IR), a transmembrane heterodimer of twice alpha and beta subunits retained by a disulfide bonding. Isoforms IR type 1 and IR type 2 have various affinities to bind with insulin with the extracellular domains. Fluctuating affinity for insulin has advantage over insulin resistance, but it is a controversial matter and still incompletely understood. Insulin binding facilitates the interaction and autophosphorylation of three tyrosine residues in the control domain and raises the activity of the enzyme tyrosine kinases. Then, intracellular phosphorylation of insulin receptor substrate, (IRS type 1) takes place. IRS type 1 is a typical adapter protein expresses four isoforms, of which IRS type 1 and IRS type 2 get involved in the balance of glucose and glucagon to control levels of blood glucose and T2DM [82, 83].

Insulin hormone in plants is not accepted by plant science researchers [84]. Insulin is the main glucose controlling hormone and it was originally isolated from pancreatic tissue of an animal [85]. Plant life does not carry pancreas and so glucose does not precede major metabolites. Several studies suggested that chemicals similar to animal insulin exist in plants and extracts from these chemical substances alter the metabolism of the seedlings, so it was proved that insulin protein is present in plants [86, 87]. Khanna and his colleagues stated that the similar to insulin, glucokinin is present in plants and microbes that exhibited similar functions to those of insulin in vertebrates [23]. Therefore,

in some studies, insulin-like peptides have been reported in the living organism such as bacteria and fungi [24, 25].

Further studies which proved the possibility of the existence of insulin-type molecules in *Momordica charantia* was conducted by Ng and his colleagues [26]. It proves that similar features of a protein of animal insulin found in plants. *Momordica charantia* when co-administered with the conventional drugs and tested in clinical studies for its combined effects proves that it produced positive interactions with the drugs and significantly reduce the serum glucose at half of the regular dose with metformin [88], while with glibenclamide also shows remarkable reduction in serum glucose at half of regular dose of glibenclamide [88]. In other experimental studies, it is proved that combined therapy of metformin with *Momordica charantia* presented improved hypoglycemic activity in normal, streptozotocin induced- and alloxan-diabetic rats [89, 90]. Positive interactions of plant extracts with antidiabetic drugs could improve the situation of diabetes worldwide in terms of enhanced drug bioactivity and side effects.

The “Human Genome Project” brought revolution which permitted comparison analyses of sequence of nucleotides and proteins through bioinformatics approaches to identify common proteins that could exist across different living organisms [91, 92].

Xavier-Filho and his fellows retrieved the information that suggested insulin was present in plants. Their results suggested that insulin protein express common amino-acid sequence as bovine insulin in plants of family *Leguminosae* [93]. Koona and his fellows tested this assumption that plant species contain sequences similar to that of animal insulin by phylogenetic analyses of different categories of insulin. They predicted protein domains and demonstrated that molecules similar to insulin are present in plant species.

In addition, domains common to the sequence of insulin are present in *Bauhinia purpurea, Canavalia ensiformis, and Vigna unguiculata*. Proteins similar to insulin may have role in the development of plants and show metabolic activities [94].

Bauhinia purpurea (orchid tree) is a member of family *Leguminosae*. It is an average-size deciduous tree, the components of which are used as medicine for body pain, restlessness, fever, dropsy, rheumatism, seizures, and septicemia [95]. Plant bark functions as an astringent in the management of diarrhea and its isolated chemicals are useful in the treatment of stomach ulcers. The plant has pharmacologic actions on the central nervous system and has cardiotonic, hypoglycemic, lower blood cholesterol, oxidation inhibition and anti-hepatotoxicity activities [96]. Leaves of orchid tree are widely used to cure abrasions and muscular damages [97]

Canavalia ensiformis (horse bean/ Jack beans) is a member of family *Leguminosae.* It is found in the Central America and West Indian islands. Jack beans cultivation widely found in the humid tropical region of Asia and Africa. *Canavalia ensiformis* seeds have been reported to possess anti-hypercholesterolemic and hypoglycemic properties [98]. Its extracts have been tested on alloxan-induced DM rats, showed good activity against hyperlipidemia and hyperketonemia, and it has been shown to be potential anti-DM agents. Oral administration of an aqueous extract of the seeds of *Canavalia ensiformis* has been shown to reduce urinary and blood levels of glucose and to elevate levels of triacylglycerol, ketonic group, and level of cholesterol related with DM [99].

Vigna unguiculata (cowpea) is a member of family *Leguminosae* found in various regions of Asia and Africa [100], owns a three-lobed leaf and long slender pods. It

reaches maturity in sixty days when sowed. The protein sequence of *Vigna unguiculata* shows similar sequence to bovine/animal insulin to sequence of the plant-insulin extracted from the cowpea seed-coat [27, 101]. These old-style treatments have encouraging future in DM management. Adverse effects were reported as compared in commercial drugs available for hypoglycemia [27]. In the present study, I carried out bioinformatics studies of molecular docking to identify new drugs for DM treatment using plant extracts with similar sequences to those of animal insulin.

I employed a ligand-based drug design and revealed diverse classes of small drug-like compounds to be potential candidates for DM treatment. Moreover, I computed molecular modeling and docking studies for the lead compound, which was identified from the test dataset of anti-DM compounds [102] and modified for optimization of its activity. These results will provide a deeper understanding of the inhibitory behavior of the compound and be valuable in the development of anti-DM drugs.

2.3. Materials and methods

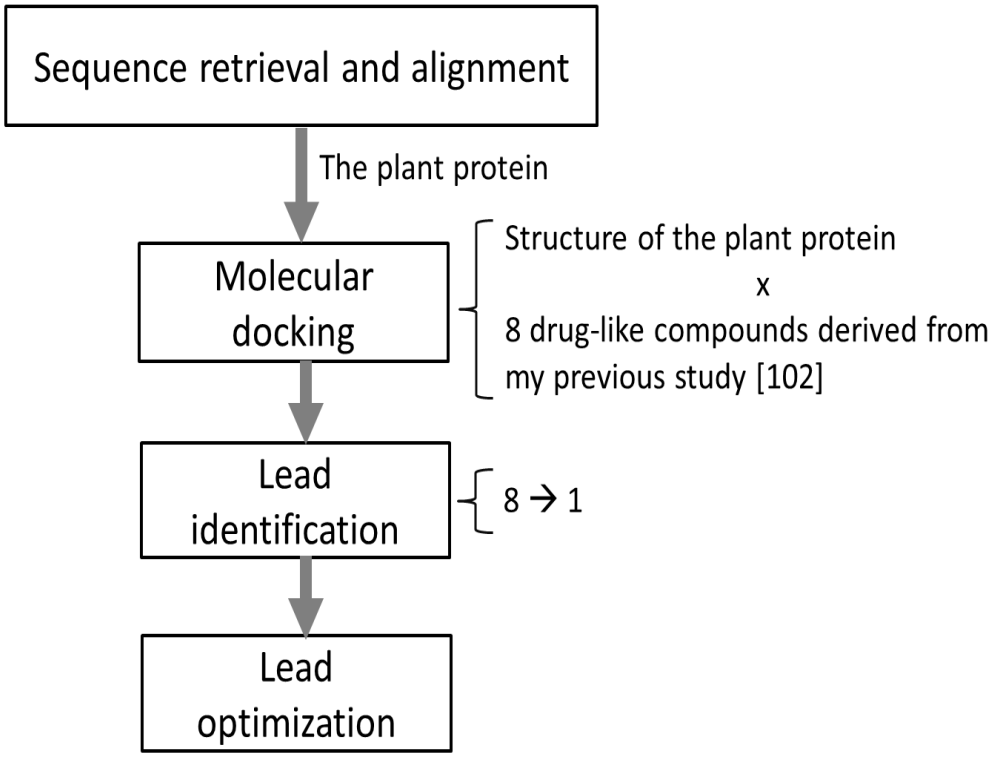


Figure 2.1: Schematic workflow summarizing the methods used to identify plant insulin-based antidiabetic drugs.

Diabetic mellitus was selected as target disease to start this study. As results of human insulin sequence similarity search by BlastP [103], *Canavalia ensiformis* found with highest sequence similarity of 56% with 88.2 maximum bits score. Anti-DM compounds were retrieved from my previous study [102]. Insulin protein isolated by plant source was subjected to identify the most active anti-DM compound by molecular docking and detail interaction analysis. Lead compound was identified and optimized by analogs design using functional group inter-conversion approach. Schematic workflow

summarizing computer-aided drug design methods used in this study is shown in Figure 2.1.

BlastP [103] is used to for identification of homologs of human insulin by using the insulin sequence with 110 amino acid length as input query sequence. Query sequence was submitted in FASTA format and results retrieved in the HTML format. Blast search for the identical sequences present in the database with respect to query sequence. While performing BlastP sequence similarity search using NCBI portal. The results were given in graphical format showing the most identical hits, domain knowledge and family of protein and also provide the table of sequence identical to the input query sequence with certain Blast scores and similarity percentages, with the alignment for each sequence with respect to query sequence shown in Table 2.1 and 2.2.

Molecular docking analyses were undertaken to evaluate the most preferred geometry of protein–ligand complexes. Anti-DM compounds were analyzed for a target protein binding using Auto Dock v 4.0 and Auto Dock Vina [54, 104]. The docking phase is, in general, meaningful with its two components: target protein and ligand. Molecular docking simulations identify native or similar to native configurations of docked complexes.

Docking steps were conducted in a specific sequence. Briefly, water molecules were excluded from target protein structure, and then the input was provided to analytical docking tool. Marsili-Gasteiger partial charges were calculated for the target protein by Auto Dock v 4.0 [54, 104]. And then the protein structure was examined for the missing atoms. So when missing atoms confirmed, hydrogen atoms were added by selecting the default parameters. After these modifications, the protein structure was obtained, and the

ligand was prepared for docking experiment. Marsili-Gasteiger partial charges were calculated for the ligand [105]. Then some moiety exhibiting torsions in ligands was defined. To choose torsion for flexible docking, rotatable connections were altered into non-rotatable connections and vice versa. The number assigned for active torsions was marked as the most atoms. After preparation of protein and a ligand structure, an inflexible residue was set by utilizing the GRID modules of Auto Dock v4.0. A flexible macromolecule was then obtained. Auto Dock Vina was used for molecular docking. This software outputs various energy conformations. Among these, the lowest energy conformation against each docked ligand was selected and docking results for the selected dataset are generated.

For understanding the results of molecular docking, the therapeutic target protein was docked with the test set, and the interactions between of binding pocket of the protein molecule and ligands must be found. There are three types of interactions found in the docked complex: hydrophobic interactions, ionic interactions, and hydrogen bonding. Interaction analyses were conducted by using Visual Molecular Dynamic (VMD) program [106]. The interaction results were considered within a distance of 4Å. The detail binding behavior of each docked complex was analyzed (Table 2.2).

2.4. Results and discussion

Target identification and selection is an important step in initiating drug design. The plant insulin protein was used as the target protein in this study. I extracted plant-insulin 3D structure from the public source of MODBASE database [107] to examine it as a substitute source of human insulin protein. The three-dimensional (3D) structure of

protein isolated from *Canavalia ensiformis* with identification number Q7M217 in figure 2.2 shows two representations.

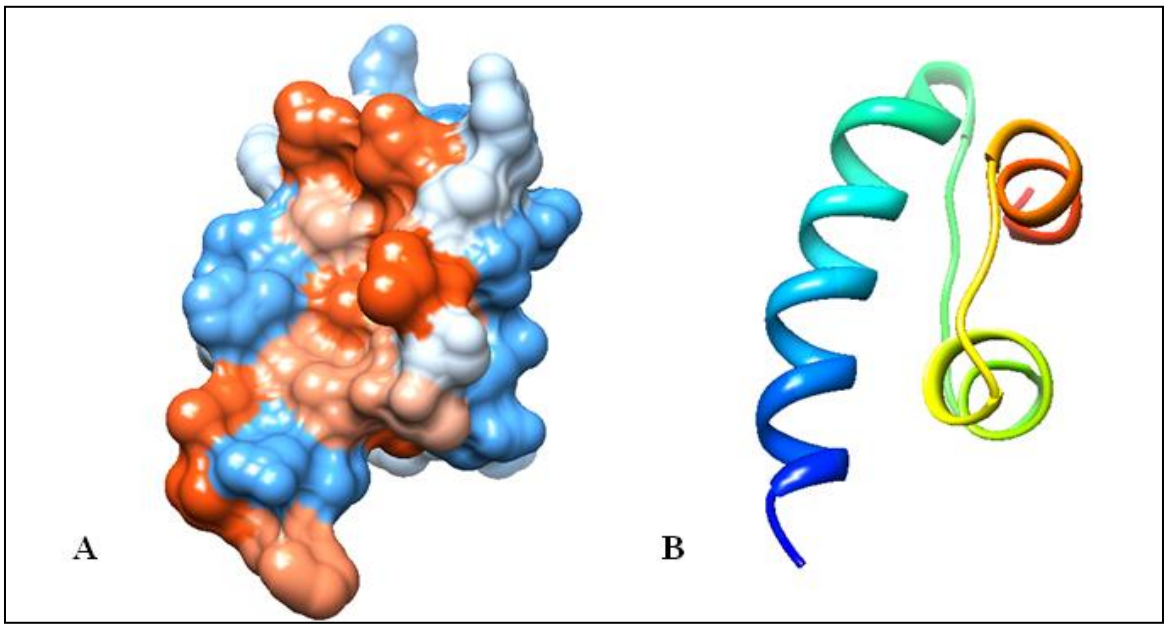


Figure 2.2: Structure of plant insulin extracted from *Canavalia ensiformis*, identification number Q7M217 [A] protein hydrophobic surface and [B] ribbon representations generated by chimera software.

The insulin-like growth factor segments of human insulin are conserved to the insulin sequence in *Bauhinia purpurea, Canavalia ensiformis and Vigna unguiculata* [94]. These plants are members of the class: *Leguminosae*. I selected *Canavalia ensiformis* for testing as an insulin source because it has been tested in wet laboratory experiments and because it has a highly identical homolog to human insulin protein (table 2.1 & table 2.2).

In a wet laboratory experiment, a protein extracted from *Canavalia ensiformis* was acknowledged by anti-human insulin antibodies that lower the level of blood glucose

in alloxanized mice (suggesting that the plant insulin has biologic potential against DM), and found to have evolutionary characteristics similar to those of human insulin [108].

The reason to select the most identical insulin-like protein *Canavalia ensiformis* in this study is depicted by sequence similarity search. Results summary of human insulin sequence similarity search by BlastP [103] is shown in table 2.1 and the align sequences are shown in table 2.2. *Canavalia ensiformis* shows the highest sequence similarity of 56% with 88.2 maximum bits score. While *Vigna unguiculata* with 72.4 bits score shows 49% sequence similarity and *Bauhinia purpurea* with 65.5 bits score shows 67% sequence similarity with human insulin protein.

BlastP a freely available web tool searches for the identical and specific hits as homologs. They represent a reliable association between the protein query sequence (human insulin sequence) and a domain model. Figure 2.3 displays putative conserved domain and information of the superfamily retrieved against the query sequence used as input to BlastP. Conserved IIGF-insulin-like domains shown in dark green bar and IIGF-like superfamily shown in light green bar concluded the function of the model protein. IIGF-like superfamily is a large class of evolutionary proteins which own diverse hormonal activities and its subfamily is insulin and insulin-like growth factors.

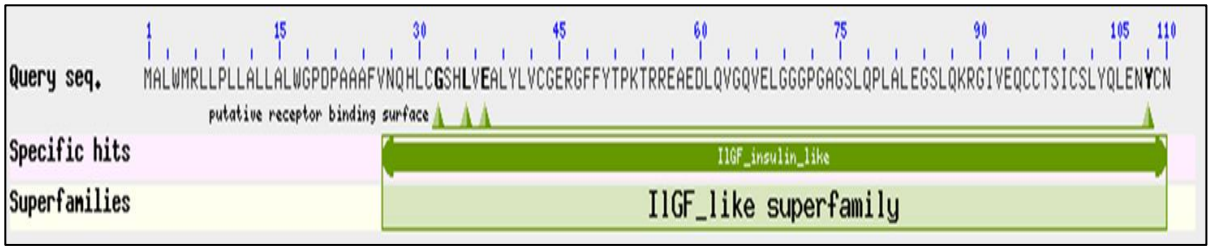


Figure 2.3: Graphical summary of the database sequence aligned to the query sequence.

Table 2.1: Summary of the alignment results of three top scored plant insulin hits against human insulin by BlastP.

Top scored hits	Accession ID	Source	Max score	Total score	Query cover	E value	Identity	Positives	Gaps
1	A59151	Canavalia ensiformis (jack bean)	88.2	88.2	78%	1e-22	56%	56%	40%
2	P83770.1	Vigna unguiculata (cowpea)	72.4	72.4	78%	3e-16	49%	50%	40%
3	721138A	Bauhinia purpurea (camel's foot tree)	65.5	110	58%	1e-13	67%	79%	43%

Table 2.2: Sequence alignment for human insulin and three top scored plant insulin hits.

Top scored hits	Protein description and sequence alignments against Query (human insulin)
1	Insulin precursor - jack bean (fragments) / Canavalia ensiformis (jack bean) (Sequence length: 51) Query 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84 FVNQHLCGSHLVEALYLVCGERGFFYTPK Sbjct 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA------------------------------ 30 Query 85 SLQKRGIVEQCCTSICSLYQLENYCN 110 GIVEQCC S+CSLYQLENYCN Sbjct 31 -----GIVEQCCASVCSLYQLENYCN 51
2	RecName: Full=Insulin-like protein; Contains: Rec Name: Full=Insulin-like protein B chain; Contains: Rec Name: Full=Insulin-like protein A chain / Vigna unguiculata (cowpea) (Sequence length: 51) Query 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84 FVNQHL GSHLVEALYLV GERGFFYTPK Sbjct 1 FVNQHLXGSHLVEALYLVXGERGFFYTPKA------------------------------ 30 Query 85 SLQKRGIVEQCCTSICSLYQLENYCN 110 GIVEQ S+ SLYQLENY N Sbjct 31 -----GIVEQXXASVXSLYQLENYXN 51
3	Insulin / Bauhinia purpurea (camel's foot tree) (sequence length: 51) Query 12 ALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKT 54 ++ +L+ + F NQHLCGSHLVEALYLVCGERGFFYTPK Sbjct 9 SVCSLYQLENYCNFANQHLCGSHLVEALYLVCGERGFFYTPKA 51

I used aleglitazar (Roche, Basel, Switzerland) [109] with a half-maximal inhibitory concentration (IC_{50}) value of 0.019 μM as a standard drug for DM. I collected data for aleglitazar from PubChem [110], which provides authenticated chemical structure and all related information of drugs and which is organized by the US National Institutes of Health. Aleglitazar is a type of sensitizer used for T2DM treatment to reduce the complications of cardiovascular morbidity and mortality. In T2DM patients, aleglitazar can control levels of lipids and glucose in a synergistic manner while eliciting limited side effects and toxicity [110]. I designed and evaluated novel candidate compounds based on a comparison with aleglitazar.

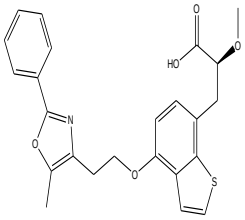
I generated a test dataset of eight compounds (table 2.3) by perusing studies of anti-DM drugs [102]. The dataset was considered highly active owing to their low IC_{50} values (μM). The rule of five [111] used to evaluate drug-likeness of chemical compounds and the results integrate the pharmacokinetics of these compounds from a previous study [112]. Compound structures in the test dataset were made by Chem Draw Ultra 8.0 [113]. The compounds and their bioavailability in the form of IC_{50} values are listed in table 2.3.

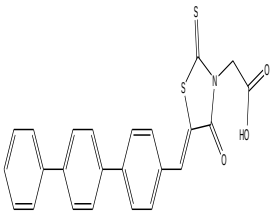
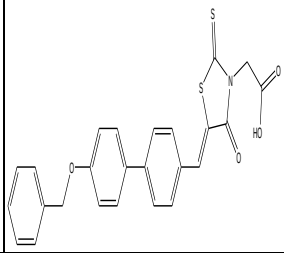
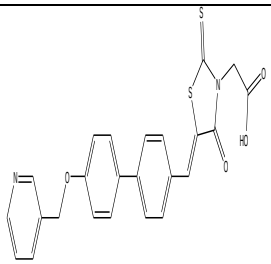
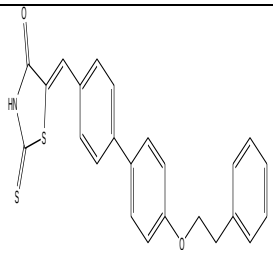
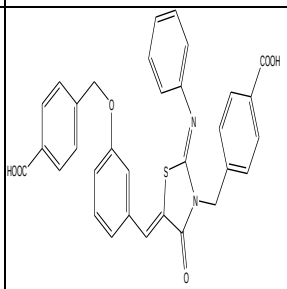
I evaluated interactions of compounds with protein molecule using Auto Dock and Auto Dock Vina [104]. By employing docking analyses, different confirmations of compounds were provided as docked complex with the target protein molecule. I generated ten most active conformations for each ligand ranked based on binding affinities of the ligand with the protein molecule. I selected the optimal confirmation from these ten confirmations (having a minimum value of the root-mean-square

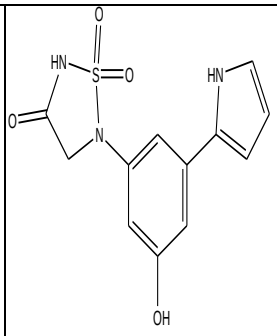
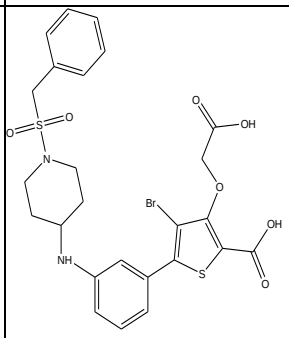
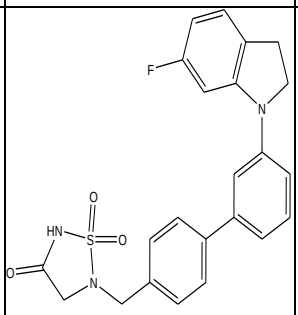
deviation) based on the computed energies of compounds docked with the protein molecule for analyses of their binding behavior.

Furthermore, I analyzed the two-dimensional (2D) and three-dimensional (3D) structures of the ligand and plant target protein. Amino acids involved in the interactions in the relevant binding pocket were studied. The test dataset was docked with the target protein. Amino acids in the protein pocket were identified within a distance of 10Å. The residues beneficial for interactions and comprise protein pocket were: HIS4, HIS5, HIS10, ALA14, ALA30, PHE24, PHE25, VAL12, TYR16, TYR26, TYR49, THR27, CYS7, CYS36, CYS37, CYS41, CYS50, LYS29, LEU3, LEU11, LEU17, LEU43, LEU46, VAL40, GLN8, GLN35, GLN45, GLY8, GLY31, GLU13, ASN3, ASN48, and SER39 (table 2.3).

Table 2.3: Structures and binding interactions of the standard drug, aleglitazar, and eight test compounds (T1-T8) including amino acid data in the target protein pocket and binding energies

Name	Structure	IC_{50} (µM)	Hydrogen bonding		Ionic interaction		Hydrophobic interaction		Binding Energy (Kcal/mol)
			Amino acids	Distance (Å)	Amino acids	Distance (Å)	Amino acids	Distance (Å)	
Aleglitazar		0.019	O-HIS10:NE2	3.21	None		C-HIS10:CD2 C-ALA14:CA C-ALA14:CB C-LEU11:CD2 C-LEU11:CD2 C-CYS7:CA C-SER39:C C-VAL40:CA C-VAL40:CA	3.82 3.93 3.71 4.03 3.83 3.82 3.90 3.94 3.73	-7.7

T1		0.53	S-GLN8:N	4.00	None	C-TYR26:CD2 C-TYR26:CB C-PHE24:CE2 C-PHE24:CZ C-PHE24:CZ C-TYR16:CB C-TYR16:CB C-TYR16:CD2 C-TYR16:CD2 C-TYR16:CE2 C-VAL12:CG1 C-VAL12:CG1 C-VAL12:CG2 C-VAL:C	3.95 3.85 3.95 3.75 3.89 3.45 4.00 3.90 3.75 3.69 4.00 3.94 3.71 3.75	-8.5
T2		0.48	O-SER39:N N-CYS37:O	3.95 3.55	None	C-GLU13:C C-ALA14:CA C-ALA14:CB C-LEU43:CD2 C-LEU46:CD2 C-LEU11:CD2 C-VAL40:CG1 C-VAL40:CB C-VAL40:CB C-VAL40CG2	3.71 3.40 3.40 3.40 3.76 4.00 3.96 3.99 3.46 3.77	-7.8
T3		1.10	O-GLN45:NE2 S-ASN48:OD1 O-GLN45:N	3.22 3.47 3.89	None	C-TYR49:CE1 C-TYR49:CE1 C-ASN48:CB C-ALA30:CA C-PHE25:CD2 C-GLN35:CD	3.75 3.75 3.29 3.80 3.84 3.96	-7.7
T4		1.24	HN-TYR26:O	3.91	None	C-TYR16:CD2 C-TYR16:CB C-GLU13:CG C-GLU13:CD C-VAL12:CB C-VAL12:CG1 C-TYR26:CB	3.58 3.78 3.51 3.97 3.66 3.85 3.75	-7.5
T5		0.22	O-CYS7:SG O-ASN3:ND2 H-ASN3:ND2	4.02 3.08 2.81	NH-GLU13:O	3.99	C-GLU13:CB C-ALA14:CB C-HIS10:C C-HIS10:CB C-LEU11:CD2 C-LEU46:CD2 C-CYS41:CB C-LEU11:CD2 C-LEU6:CD2	3.77 3.91 3.84 3.79 3.75 3.45 4.04 3.75 3.99	-8.1

T6		0.08	NH-TYR26:O O-SER9:NH S-CYS41:NH	4.04 3.24 3.81	NH-HIS5:0 NH-HIS5:ND1	1.97 3.06	C-HIS5:CA C-HIS5:ND1 C-TYR26:CB C-TYR26:CE C-VAL12:CB C-VAL12:CG1 C-VAL12:CG1 C-VAL12:CG2 C-TYR26:CB C-TYR26:CE C-PHE24:CE2 C-PHE24:CZ	3.88 4.00 3.06 3.81 3.71 3.67 3.86 4.00 3.94 3.67 3.67 3.90	-5.9
T7		0.005	O-LEU11:N H-ASN3:ND2 O-ASN3:ND2 H-ASN3OD1 H-CYS36:O H-SER39:O H-VAL40:N H-CYS41:N	3.73 2.62 3.15 3.74 3.47 3.55 3.71 3.95	None		C-LEU43:CD2 C-ALA14:CA C-ALA14:CB C-ALA14:CB C-VAL40:CG2	3.76 3.71 3.73 3.66 3.82	-7.6
T8		0.13	O-HIS5:N N-TYR26:OH	3.67 3.17	None		C-GLY8:C C-VAL12:CG1 C-VAL12:CG2 C-PHE24:CE2 C-PHE24:CZ C-TYR26:CB C-TYR26:CB C-TYR26:CB C-TYR26:CZ C-TYR26:CZ	3.84 3.87 3.79 3.78 3.56 3.65 3.90 3.64 3.92 3.92	-8.0

I considered most of the essential amino acids present in active site of plant protein that was similar to human insulin protein. One study reported insulin in the testa of *Canavalia ensiformis* [108]. Our docking results revealed that the residues present in the active site of target protein involved in the interaction with the selected ligands for DM.

I selected the best conformation of the docked complex out of ten poses based on the criterion of minimum binding affinity and identified and generated the interactions by VMD [106] (table 2.3). VMD software enables labeling and provides the calculation of

the distance between residues of the particular ligand in a protein active site. Important interactions identified in the test dataset included ionic (COOH-NH3 or NH2-COOH), hydrogen (N-O, O-N, O-O) and hydrophobic interactions (C-C). All interactions were calculated < 4Å of the distance between the active residues of the ligand and protein.

I selected a lead compound, which is an anti-diabetic synthetic compound with publication number: WO2007067614 shown as T6 in table 2.3. from the dataset of eight compounds that had desired biologic activities on a validated molecular target. In general, a lead compound can be modified to produce another compound with a better profile by removing unwanted properties to avoid unwanted side effects.

Compounds used as potential leads can be synthetic and semi-synthetic compounds, as well as proteins in marine organisms, plants and animals [114]. The lead compound I selected was from the synthetic source in the test dataset (table 2.3). I conducted lead identifications by a computer-aided approach involving virtual screening, pharmacophore mapping, and molecular docking analyses [112,115-116].

In general, an appropriate potential drug candidate is a compound with fewer side effects or is more efficacious [117]. The lead compound may not necessarily become a drug candidate. To avoid such a situation, lead optimization can be advantageous in lead identification. One pharmaceutical company reported on the methods of the identification and optimization of lead compounds [118]. I identified a lead compound based on the binding interactions, lowest inhibitory values in terms of IC_{50}, and docking score. Figure 2.4 demonstrate the binding behavior of lead compound with target protein. I made analogs of the compound to obtain the most active anti-DM drugs. Table 2.4 demonstrates the analogs designed by modifying the functional groups to make the

compound more efficacious. The designed analog compounds from this study need to be tested for ADMET properties. Four analogs were recommended after analyses of the lead compound. Table 2.4 demonstrates analogs structures designed by lead compound with their International Union of Pure and Applied Chemistry nomenclature generated by ChemDraw Ultra 8.0 [113]. These analogs were created by addition or removal of the structural moiety or by replacement of each moiety with another present in the structure of the most active compound. First analog had a functional group comprising a sulfur atom and a hydrogen atom (-SH) at the position of (-OH). The second analog was made by a nucleophilic substitution (though its activity was dependent upon the electronic nature of the substituent). The third analog was made by the reduction of a ketone group. The fourth analog was made by removal of a steric blocker to improve the binding character of the compound. This method of analog design improved binding interactions with the target protein. Table 2.4 also listed the possible interactions and binding energies of the analog set within the distance of 10Å of a pocket of the protein molecule. The target protein (figure 2.2) showed a better binding interaction with our test dataset. Thus, I proposed it as a candidate to confirm its activity in future studies.

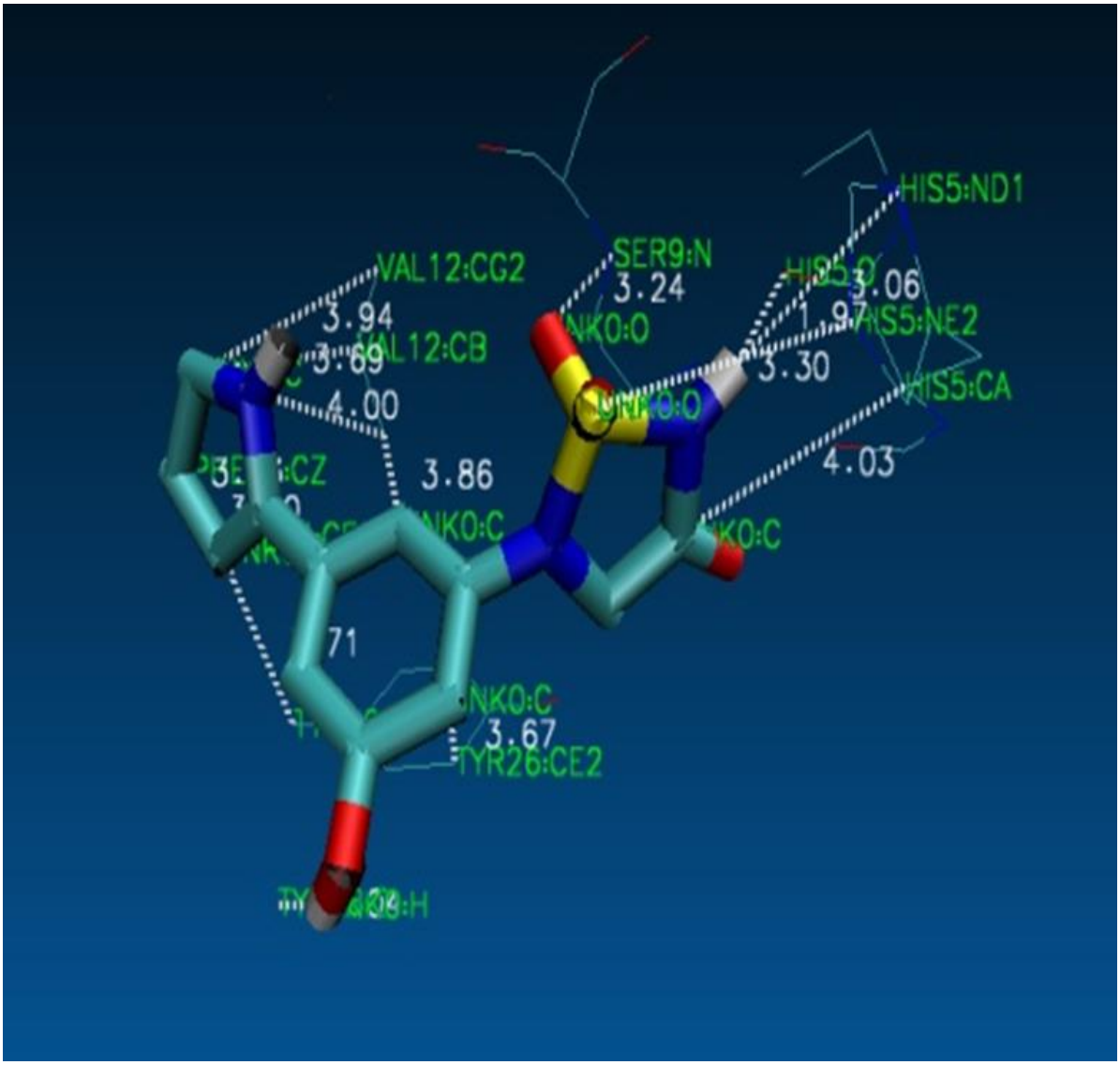
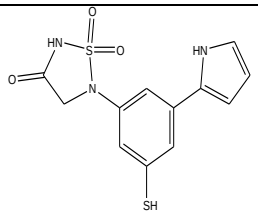
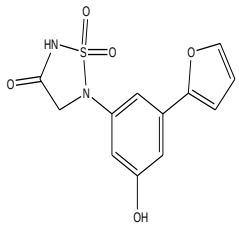
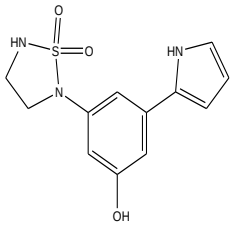


Figure 2.4: Binding interaction of docked lead compound T6 with active-site residues of target protein characterized in bond formation. Red highlights hydrogen bond acceptors and blue highlights the hydrogen bond donors, white highlights hydrogen bonds and yellow highlights halogens atom.

Table 2.4: Analogs of the lead compound (T6) along with interactions and binding affinities of the analogs with those of the target protein pocket

No.	FGI	Structure and IUPAC name	Hydrogen bonding		Ionic interaction		Hydrophobic interaction		Binding Energy (Kcal/mol)
			Amino acids	Distance (Å)	Amino acids	Distance (Å)	Amino acids	Distance (Å)	
1	Functional group conversion	 5-[3-Mercapto-5-(1*H*-pyrrol-2-yl)-phenyl]-1,1-dioxo-1λ^6-[1,2,5]thiadiazolidin-3-one	NH-CYS36: O O-CYS41:NH S-CYS41:NH	1.76 2.80 3.81	None		C-HIS10:CG C-HIS10:CG C-CYS7:CA C-CYS7:CA C-LEU11:CD2 C-ALA14:CB	3.90 3.81 3.66 3.78 3.88 3.63	-5.8
2	Nucleophic substituent	 5-(3-Furan-2-yl-5-hydroxy-phenyl)-1,1-dioxo-1λ^6-[1,2,5]thiadiazolidin-3-one	NH-SER39: O NH-CYS41: O NH-CYS41:SO	3.41 3.52 3.95	None		C-LEU3:CD2 C-ALA14:CB C-LEU11:CD2 C-CYS7:C C-CYS7:CA C-CYS7:CA	3.66 3.67 3.96 3.98 3.60 3.93	-5.9
3	Reduction of ketonic group	 3-(1,1-Dioxo-1λ^6-[1,2,5]thiadiazolidin-2-yl)-5-(1H-pyrrol-2-yl)-phenol	NH-TYR25: O O-SER9: N S-CYS41: S	3.99 3.24 3.89	NH-HIS5:0	3.66	C-CYS7:CA C-CYS7:CA C-HIS5:CA C-HIS5:ND1 C-VAL12:CG1 C-VAL12:CG1 C-VAL12:CG2 C-TYR26:CB C-TYR26:CE	3.76 3.80 4.00 3.06 3.86 3.00 3.34 3.77 3.37	-5.9

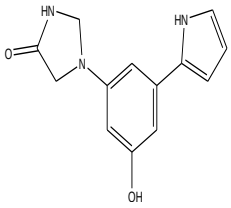
4	Removal of steric blocker	 1-[3-Hydroxy-5-(1H-pyrrol-2-yl)-phenyl]-imidazolidin-4-one	NH-TYR49: O 3.21 NH-CYS50: O 3.51	None	C-LEU3:CA 3.22 C-ALA14:CB 3.51 C-CYS7:CA 3.76 C-CYS7:CB 3.52 C-LEU11:CD2 3.99	-5.9

Table 2.5 enlists the drug-like properties of designed analogs with respect to standard anti-DM drug. These four analogs are small drug-like molecules following Lipinski's [111] and veber's rules [119] of drug-likeness. Lipinski's rules for druglikeness limits molecular weight (MW) to less than 500 Dalton, logP values to less than 5, hydrogen bond acceptors (HBA) to less than 10, and hydrogen bond donors (HBD) to less than 5 and veber's rules limits rotatable bonds (RB) to less than 10 while value of polar surface area (PSA) to less than 120 Å. By the universal idea an oral biologically efficacious drug candidate should not violate except one property as described [111].

Table 2.5: Summary of drug-like properties of analogs and a standard antidiabetic drug

Chemical compounds	MW	LogP	HBA	HBD	PSA	RB
Analog - 1	309.02	0.244	6	2	66.1	2
Analog - 2	294.03	-0.778	7	2	44.73	3
Analog - 3	279.07	-0.439	6	3	61.53	3

Analog - 4	243.1	-0.313	5	3	47.53	3
Aleglitazar	437.51	5.1	7	1	110	9

2.5. Conclusion

Natural products have been suggested to be the best sources of medicines for the treatment of DM [120]. Nearly 80% of the world population use traditional medicines: they prefer plant-based drugs for primary health care [121]. Safe and effective use of natural products can ensure that plant-based medicines are more harmonious with biologic systems. I identified an anti-DM compound based on plant insulin: a synthetic compound with publication number: WO2007067614 shown as T6 in table 2.3. I designed its analogs using a functional group inter-conversion approach. Our computer-aided approach provided information on binding energies and binding interactions of the analogs to predict their anti-DM activities. Several studies highlights the combined treatments of plant extracts and conventional drugs significantly enhanced the effect with respect to improved plasma glucose and insulin levels as compared to individual drug treatment. So this plant-based study could be helpful in future to understand the plant insulin and antidiabetic drug interactions.

CHAPTER 3

AN INTEGRATED COMPUTATIONAL APPROACH FOR PLANT-BASED PROTEIN TYROSINE PHOSPHATASE NON-RECEPTOR TYPE 1 INHIBITORS

3.1. Abstract

Background: The protein tyrosine phosphatase non-receptor type 1 (PTPN1) is a novel target for the type 2 diabetes mellitus. According to the International Diabetes Federation 2015 report, one out of 11 adults suffers from diabetes mellitus globally.

Objective: Current anti-diabetic drugs can cause life-threatening side-effects. The present study proposes a pipeline for the development of effective and plant-derived anti-diabetic drugs that may be safer and better tolerated.

Methods: Plant-derived protein tyrosine phosphatase non-receptor type 1 inhibitory enzymes possessing antidiabetic activity less than 10μM were used as a training set. A common feature pharmacophore model was generated. Pharmacophore-based screening of plant-derived compounds of the ZINC database was conducted using ZINCpharmer. Screened hits were assessed to evaluate their drug-likeness, pharmacokinetics, detailed binding behavior, and aggregator possibility based on their physiochemical properties and chemical similarity with reported aggregators.

Results: Through virtual screening and in silico pharmacology protocols isosilybin (ZINC30731533) was identified as a lead compound with optimal properties. This compound can be recommended for laboratory tests and further analyses to confirm its activity as PTPN1 inhibitor.

Conclusion: The present study identified a plant-derived anti-diabetic virtual lead compound with the potential to inhibit PTPN1, which may be helpful to enhance insulin production. This computer-aided study could facilitate the development of novel pharmacological inhibitors for diabetes treatment.

Keywords: Computer-aided drug design, diabetes mellitus, flavonoids, isosilybin, protein tyrosine phosphatase non-receptor type 1, common feature pharmacophore modeling, molecular docking, pharmacokinetics.

3.2. Introduction

Plant-based medicine is a way to treat diabetes mellitus (DM). Traditional medicine has employed a huge collection of plant-derived treatments effective in the treatment of blood glucose imbalance and diabetes mellitus [122,123]. Experimental studies have shown that the sequence of the insulin-like growth factor domain of animal insulin protein is similar to the plant-insulin sequence found in *Canavalia ensiformis*, *Vigna unguiculata* and *Bauhinia purpurea* [94]. Computer-aided molecular docking methods were applied to human insulin protein [116] and plant insulin present in *Canavalia ensiformis* to identify anti-diabetic compounds [76].

In our previous study, I discussed FDA approved DM medicines; insulin, biguanides, second generation sulfonylureas, alpha- glucosidase inhibitors, glinides, glucagon-type peptide 1 receptor as agonist activity, thiazolidinediones, bile-acid sequestrants, newly developed drugs based on dipeptidyl peptidase-4 (DPP-4) inhibitors, dopamine activators, amylin analogs, and sodium-dependent glucose cotransporter-2 inhibitors in detail [6]. However, currently available anti-DM drugs possess side effects

such as headache, stomach upset, peripheral edema, increase in weight, and hypotension [124]. Therefore, compounds with ideal properties to stimulate insulin signaling pathway are required [125].

Molecular targets for pharmacological treatments of DM has been studied to develop unique anti-DM agents, including protein tyrosine phosphatase non-receptor type 1 (PTPN1) previously also known as protein tyrosine phosphatase 1B (PTP1B), peroxisome proliferator-activated receptor gamma, pyruvate dehydrogenase kinase, beta 3 adrenoceptors, glycogen synthase kinase 3, DPP-4, cannabinoid receptors, and fructose bisphosphatases enzymes [126, 127]. The protein tyrosine phosphatases accelerate the protein tyrosine dephosphorylation in the regulatory mechanism of insulin via dephosphorylation of activated auto phosphorylated insulin receptor and downstream substrate proteins [128]. The PTPN1 has been a target for management of the diabetes disease and obesity [129], and PTPN1 knockout mice had insulin sensitivity and tolerance to diet-induced obesity [11, 130]. Recent technical advances in biochemical synthesis proved the discovery of potential synthetic PTPN1 candidates, but complications for example polarity besides less enzyme selectivity remained to be controlled [131]. The uses of plant-derived products have appreciated as an alternative source for discovery of PTPN1 inhibitory candidates [132]. In vitro and in vivo methods confirmed that natural products are beneficial for the discovery of new and potential PTPN1 inhibitors [117].

In the present study, I discussed structural, biological and molecular activities of diverse plant-derived PTPN1 compounds reported in last decades. I use computer-aided drug design (CADD) strategies for identification of novel compounds having PTPN1

inhibitory activity from the ZINC dataset of plant-derived compounds, which will be beneficial for medicinal chemist and pharmacologists to develop new PTPN1 inhibitors with anti-DM activity [133].

3.3. Material and methods

3.3.1. Pharmacophore modeling and computer-based screening of ZINC database

In recent years, various experimental approaches have been developed to investigate flavonoids with PTPN1 inhibitory activity by incorporating novel approaches to previously tested models to improve their anti-DM activity. Botanical information, chemical structure and physicochemical properties of natural flavonoids with PTPN1 inhibitory activity were selected from reported data (Table 3.1) [134-139]. Eleven compounds were used as a training set based on their physiochemical properties, Lipinski's filter, and IC50 values less than 10μM. These 11 compounds were used for pharmacophore modeling using LigandScout 4.1 [63]. ChemDraw Ultra 8.0 software [113] is used for sketching chemical structure of training dataset and saved in Protein Data Bank (PDB) format. Consequently, these files were used as input to LigandScout 4.1. A pharmacophore fit model was generated using the 11 compounds of the training set and used for screening of the plant-derived set of ZINC database. Table 3.2 shows pharmacophore features of the training set and common feature of a selected pharmacophore model. Pharmacophore features of the most appropriate model were also generated for each compound displayed in Table 3.3.

Screening procedures were performed using a shared feature pharmacophore modeling approach for best flexible conformation exploration using ZINCpharmer

[140].The identified hits as an outcome of the database search were subjected to drug-like filtration. Data Warrior [141] was used to calculate the physiochemical properties and toxicity estimation. Lipinski's filter was applied to the screened dataset [111]. Partition coefficient logP values limit is less than 5 or clogP values limit is less than 6, molecular weight limit is less than 500 Dalton, hydrogen bond acceptors limit is less 10, and hydrogen bond donors limit is less than 5. Veber's rule (rotatable bonds must be less than 10 while value of polar surface area must be less than 120 Å) was also considered because molecular flexibility of selected compounds is dependent on the number of rotatable bonds, an important property which influences bioavailability of drugs [119,142]. Compounds were short-listed based on drug-likeness and were subsequently checked for toxicity using four criteria (mutagenicity, tumorigenicity, irritant and reproductive effects).

3.3.2. Molecular docking and interaction analyses

The prerequisite of docking for the PTPN1 data set is the data of the target protein structure. The data of the protein (ID: 3EAX) was downloaded from the PDB; its X-ray crystallographic structure demonstrated a high resolution of 1.9 Å [143]. The existing active site for ligand binding used in the nuclear magnetic resonance (NMR) study of 3EAX structure was used for the molecular docking study of screened hits. Plant-derived compounds from the ZINC database, after passing through drug-like filters, were docked into the optimal binding site of the PTPN1 protein using the CLC drug discovery workbench tool [55]. A detail investigation of active site of the target protein was performed to check if the significant residues which are responsible for activity were included in the binding site or not. The compound which holds the selected binding site

was specified. After compound selection, docking was performed. The docking results summary was displayed in the project with dock scores. A 3D view was selected for manual inspection of the structural features of the docked complex involved in binding. The docked complex in a mol format file was imported to chimera [144] and was saved in PDB format for the binding interaction by Ligplot. The ligand-protein interactions were predicted using Ligplot [145]. It generates 2D schematic diagrams of a docked complex to explain interactions with hydrophobic moieties and hydrogen bonds having a distance within 4 Å.

3.3.3. ADME calculation and aggregator advisor prediction

In silico ADME (absorption, distribution, metabolism and excretion) calculations are steadily gaining interest in computer aided drug discovery [146]. These methods are used here to shortlist data with suitable pharmacokinetic (ADME) and toxicity profiles in early phase of PTPN1 drug discovery. SwissADME [147] is used to calculate the pharmacological profile (drug action/effects within an organism) of selected 15 hits from the perspective of drug discovery. This online tool is also used to determine toxic structural moieties and synthetic accessibility of selected hits.

Aggregator possibility evaluation for selected 15 hits was conducted by Aggregator advisor [148]. This online tool compares the chemical similarity of known aggregator compounds based on Tanimoto coefficient calculation and physiochemical properties. Input was provided using a SMILES file extension to obtain a reliable ADME prediction and information on already reported aggregates for each compound.

3.3.4. Lead identification

After the systematic analyses of all compounds, a lead compound was identified as a chemical compound that has the best pharmacological or biological activity against the PTPN1 therapeutic target. In detail, the lead compound was identified based on the best drug-likeness, pharmacokinetic properties, molecular docking and best binding interactions with the significant residues involved in binding the NMR structure of the target protein. The aggregator properties were used along with the other parameters to shortlist hits as a lead compound.

3.4. Results and discussion

To find potential plant-derived PTPN1 inhibitors and to deliver an idea for drug design, I have used both ligand and structure based methods. In our study an integrated computational approach has been applied for ligand-based pharmacophore modeling of reported PTPN1 inhibitors and database screening to retrieve diverse plant-derived chemical scaffolds. Molecular docking has been applied in order to clarify the behavior of PTPN1 enzyme by the binding of plant-derived compounds with different binding affinities. It helps to identify significant residues involve in PTPN1 inhibition. I used integrated strategies for structural insights along with medicinal chemistry prospective, which will robust bio-assay method to enable the design of potential and selective PTPN1 inhibitors. The significance of physiochemical properties and ADME/toxicity filters for inhibitor design is also emphasized in the search of active plant-derived PTPN1 inhibitors. Completely similar reported aggregator compound is a useful reference point in designing inhibitor with better physiochemical properties and possibility of inhibition of PTPN1 enzyme for in vivo testing with a plant-derived lead compound [133].

I have discussed in our previous research that most of FDA approved anti-DM medicines cannot achieve a satisfactory blood sugar levels in DM person, and have numerous adverse effects; therefore, novel anti-DM medicines are desirable. The CADD techniques can be exploited to screen efficient hits from various databases and to reduce the time and cost to discover novel anti-DM medications [6].

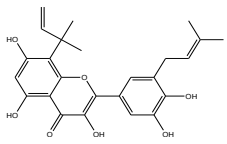
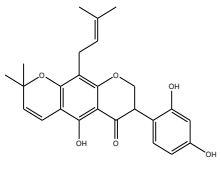
The merits of plant-based medicine have been demonstrated for various diseases. Metformin is most used medication for T2DM, derivative of guanidine which is obtained from Galegine Officinalis [20]. The potential advantages of plant-based medicines over synthetic medicines include less side-effects, increased efficacy, increased availability, and lower cost. Some of these plant-based medicines are better extracted and used in a crude form as is the common practice in traditional anti-diabetic medicine [14]. It is mostly used an adjunct to oral antidiabetic therapy to reach appropriate glycemic control.

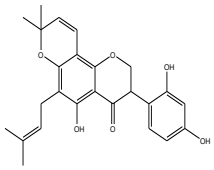
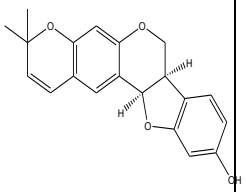
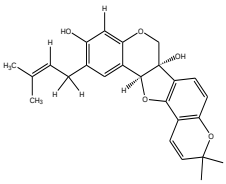
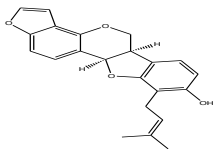
PTPN1 is an acceptable therapeutic target that can be effectively targeted for the management of T2DM. Regardless of the accessibility of numerous synthetic PTPN1 inhibitors, their use often entails side-effects, some of which are life-threatening. However, to date, PTPN1 inhibitors are far from achieving regulatory approval from the FDA. Therefore, it is necessary to identify novel hits that have potential to evolve into effective inhibitors of PTPN1. The present study focuses on a pipeline for the development of effective plant-derived anti-DM drugs that are safer and better tolerated when compared with synthetic alternatives [133].

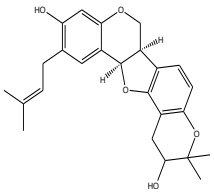
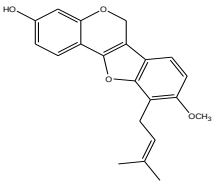
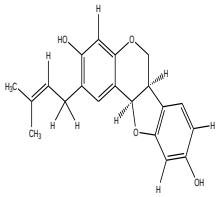
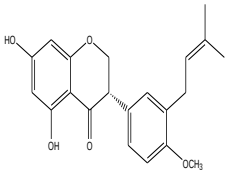
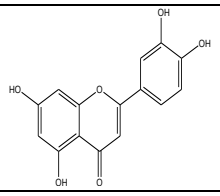
Flavonoids have extensive variety of biological activities, including anti-DM, anti-oxidative, anti-inflammatory, anti-allergic, anti-proliferative, anti-viral, anti-cancer activities, stomach and liver protection [149]. Initially, I selected 30 compounds reported

in the literature in the last decade whose PTPN1 inhibitory activity is less than 10 μM. Drug-likeness filters were applied and 11 compounds fulfilled the criteria of drug-likeness. These are flavonoid compounds with PTPN1 inhibitory activity. Inhibitory activities against PTPN1 evaluated using experimental analyses, 2D structures, and physiochemical properties confirming drug-like properties, origin and botanical information concerning source plant species are listed in Table 3.1. This information is useful to understand the importance of these plants for anti-DM medicines.

Table 3.1: Selected compounds that possess protein tyrosine phosphatase non-receptor type 1 inhibitory activity used as a training set.

Compounds	2D structure	Potency (IC_{50})	Physiochemical properties	Source	Place of origin	Ref.
F1		4.3 μM	MW: 424.491 cLogP: 5.5969 HBA: 6 HBD: 3 PSA: 96.22 RB: 3	*Broussonetia Papyrifera* (Extract of roots)	China	[134]
F2		2.6 μM	MW: 424.491 cLogP: 5.5969 HBA: 6 HBD: 3 PSA: 96.22 RB: 3	*Erythrina addisoniae* (EtOAc extract of the stem bark)	West tropical Africa	[135]

F3		4.1 μM	MW: 422.475 cLogP: 5.4331 HBA: 6 HBD: 3 PSA: 96.22 RB: 3	*Erythrina addisoniae* (EtOAc extract of the stem bark)	West tropical Africa	[135]
F4		7.6 μM	MW: 324.375 cLogP: 4.4262 HBA: 4 HBD: 1 PSA: 47.92 RB: 0	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[136]
F5		8.8 μM	MW: 406.476 cLogP: 5.2725 HBA: 5 HBD: 2 PSA: 68.15 RB: 2	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[136]
F6		6.0 μM	MW: 350.413 cLogP: 5.6154 HBA: 4 HBD: 2 PSA: 58.92 RB: 3	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[136]

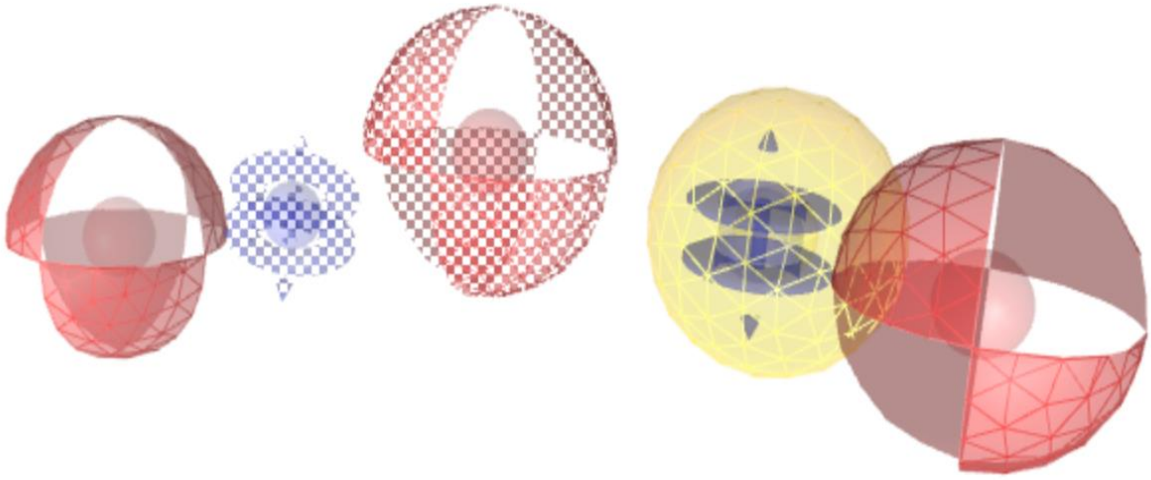
F7		9.7 μM	MW: 408.492 cLogP: 5.5966 HBA: 5 HBD: 2 PSA: 68.15 RB: 2	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[137] [138]
F8		4.1 μM	MW: 336.386 cLogP: 4.9685 HBA: 4 HBD: 1 PSA: 51.83 RB: 3	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[137]
F9		7.6 μM	MW: 324.375 cLogP: 4.9125 HBA: 4 HBD: 2 PSA: 58.92 RB: 2	*Erythrina abyssinica* (Extract of stem bark)	Africa (Nigeria)	[137]
F10		4.6 μM	MW: 354.401 cLogP: 4.6821 HBA: 5 HBD: 2 PSA: 75.99 RB: 4	*Erythrina addisoniae* (EtOAc extract of roots)	West tropical Africa, Nigeria, Congo.	[138]
F11		6.70 μM	MW: 288.254 cLogP: 1.81 HBA: 6	Abundantly present in various fruits and vegetables,	Various regions in the world	[139]

			HBD: 4 PSA: 107.22 RB: 1	e.g. *Salvia tomentosa*, *Aiphanesaculeata*.		

To achieve the goal of the ligand-based pharmacophore modeling using LigandScout 4.1, eleven compounds were used as input. Possible lowest energy conformations for each compound were generated (Table 3.2) and all conformations of least flexible compounds were then aligned. For a configurable number of best alignment solutions; common pharmacophoric features were interpolated and ten hypothetical pharmacophore models were created. The score generated for these models is shown in Figure 3.1. The pharmacophore models were ranked by means of several adjustable scoring parameters taking into account chemical feature overlap, steric overlap, or both. Pharmacophore models are set of common feature pharmacophores created by processing all compounds of the dataset. If minimum three common functional features can be identified by alignment and interpolation process, common feature pharmacophore generation is considered to be successful [63].

The model 1 with the highest score is selected for a database search to retrieve similar hits from the plant-derived set of the ZINC database. Common features are important for the activity of compounds. The pharmacophore generated using 11 compounds contained four types of pharmacophore features: hydrophobic regions (HRs), aromatic rings (ARs), hydrogen bond acceptors (HBAs), hydrogen bond donors (HBDs). Table 3.2 shows the total features present in each compound and the number of features which are common and the best fit to make the best model. Six features are the best fit to

generate the best pharmacophore model. The best pharmacophore model (model 1) in this study contains three HBAs, two ARs and one HR, as shown in Figure 3.1. Red spheres represent HBAs, purple spheres represent ARs and yellow spheres represent HRs for overlay of each compound of dataset upon the pharmacophore generated as shown in Table 3.3.



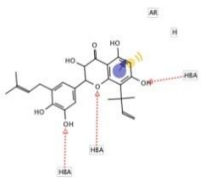
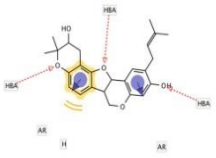
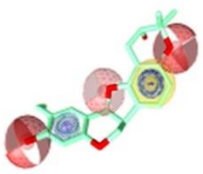
Name	Score	Name	Score	Name	Score	Name	Score	Name	Score
Model-1	0.8565	Model-3	0.8489	Model-5	0.8484	Model-7	0.8057	Model-9	0.7792
Model-2	0.8492	Model-4	0.8486	Model-6	0.8371	Model-8	0.7845	Model-10	0.7639

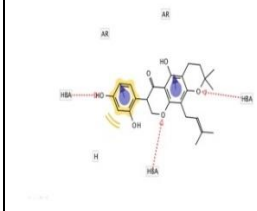
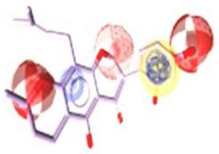
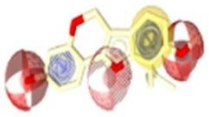
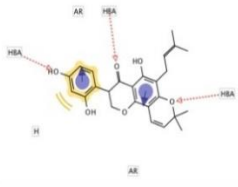
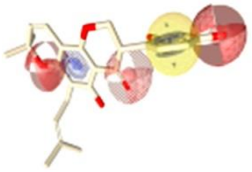
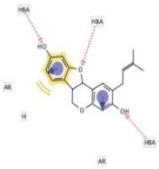
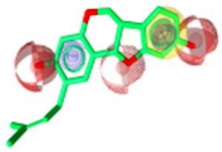
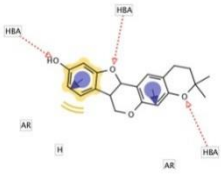
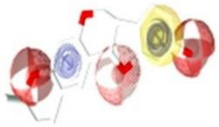
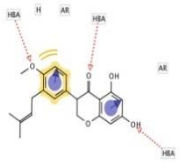
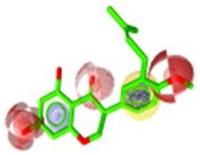
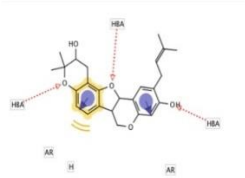
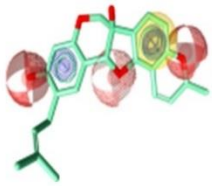
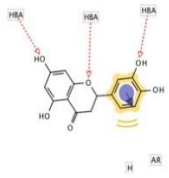
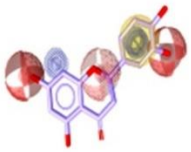
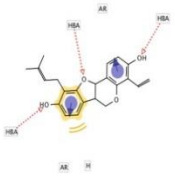
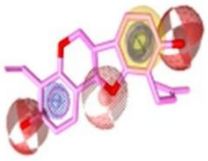
Figure 3.1: Ten pharmacophore hypothetical models (lower panel) were generated for eleven compounds using LigandScout 4.1. Six features are the best fit to generate the best pharmacophore model. The proposed pharmacophore model (model 1 shown in upper panel) used in this study contains three HBAs (red spheres), two ARs (purple spheres) and one HR (yellow spheres).

Table 3.2: Pharmacophore features of the training set and common pharmacophore feature of a selected pharmacophore model.

Compounds	HR	AR	HBA	HBD	Number of confirmations	Common pharmacophoric feature	Pharmacophoric fit
F1	5	2	7	5	54	5	56.30
F2	5	2	6	3	65	6	64.01
F3	6	2	6	2	28	6	65.02
F4	3	2	4	1	4	6	65.37
F5	6	2	5	2	21	6	65.34
F6	5	2	4	2	23	6	64.62
F7	5	2	5	2	49	6	65.34
F8	4	3	2	1	5	5	48.49
F9	3	2	4	2	19	6	65.34
F10	3	2	4	2	106	6	65.07
F11	1	2	6	4	6	5	56.62

Table 3.3: Overlay of training set compounds upon the pharmacophore generated using LigandScout 4.1.

Sr #	2D Pharmacophore	3D Pharmacophore	Sr #	2D Pharmacophore	3D Pharmacophore
F1			F7		

F2			F8		
F3			F9		
F4			F10		
F5			F11		
F6					

*Red spheres represent hydrogen bond acceptors, yellow spheres represent hydrophobic regions, and purple spheres represent aromatic rings.

Significant research has been conducted on achieving PTPN1 inhibition and many developed compounds have reached stage I or II clinical trials only to be discarded because of bioavailability and toxicity issues. The development of PTPN1 inhibitors is challenging because of potency/affinity, selectivity, and cell permeability issues. I

applied approximate filters and safety protocols to find a potent lead compound with sufficient oral bioavailability, highest docking score with favorable binding interactions, acceptable toxicity estimations, favorable pharmacokinetics, aggregation information and medicinal chemistry safety parameters [133].

.

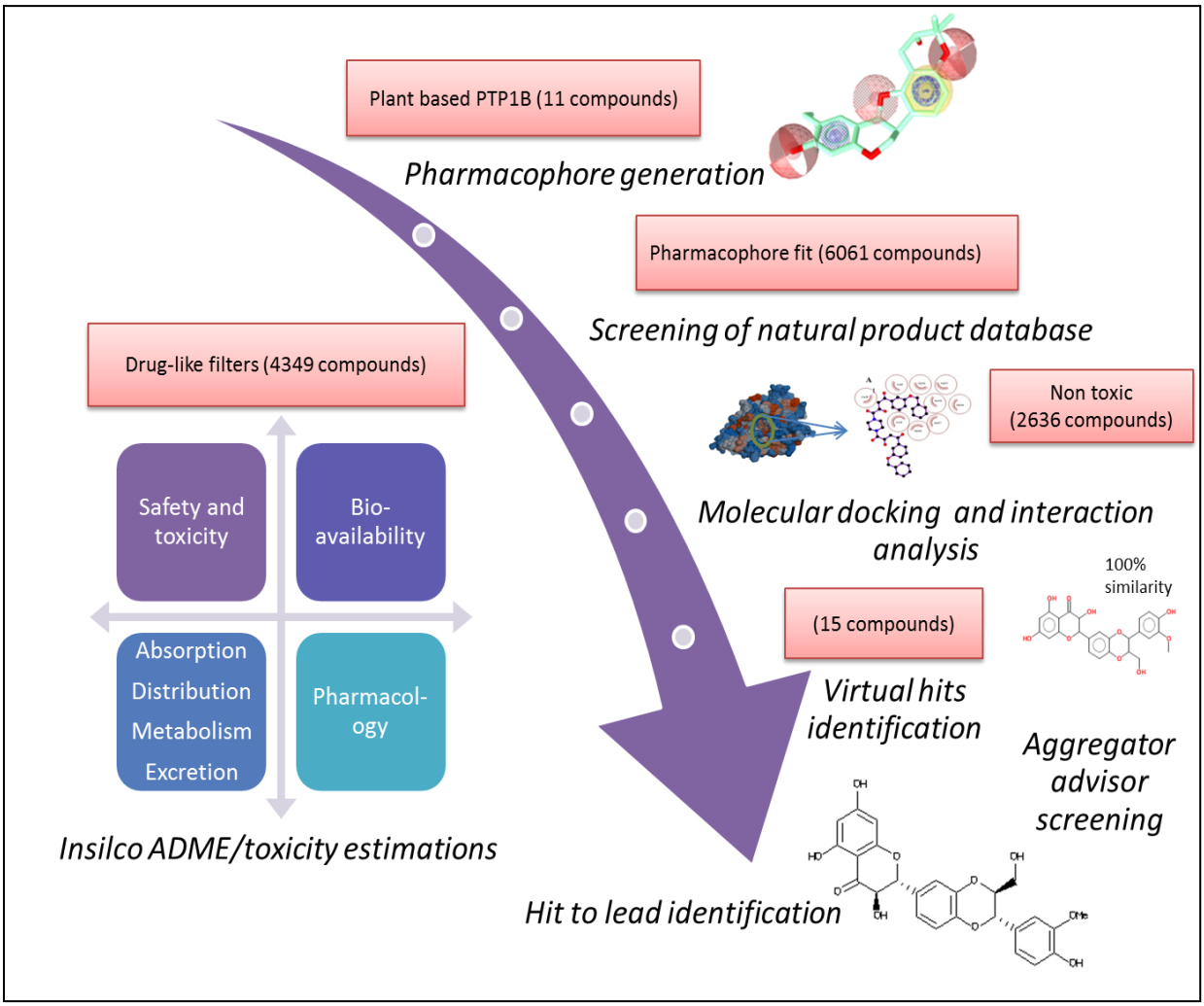


Figure 3.2: Schematic workflow summarizing the screening of protein tyrosine phosphatase non-receptor type 1 inhibitors using computer aided drug design.

To identify new plant-derived PTPN1 hits, pharmacophore based screening was conducted against 22,723,923 plant-derived compounds from the ZINC database and 6061 compounds fit with the pharmacophore query. Several filters were selected before screening by ZINCpharmer. These criteria dictated that the molecular mass of the compound should be less than 500 Daltons, the number of rotatable bonds should not be more than 10 [150], the maximum hits per configuration should be one, the maximum hits per molecule should be one and the maximum root-mean-square deviation (RMSD) for screened hits should be 1.5. The identified pharmacophore fit compounds from the databases were required to be drug-like. Therefore, Data Warrior was used for screening to calculate the physiochemical properties and toxicity estimation. Lipinski's and Veber's rules were applied to the screened dataset and reduced it to 4349 compounds. The dataset was subjected to a toxicity estimation using four criteria (mutagenicity, tumorigenicity, irritant and reproductive effects) and 2636 compounds with no risk of toxicity were retrieved. These compounds were used for molecular docking using the CLC drug discovery workbench and the 15 top scoring compounds were considered for interaction analyses using Ligplot. Then ADME calculations were performed for the selected 15 top scored hits to confirm their pharmacokinetic profile and medicinal parameters. Summary of drug-likeness and pharmacokinetic properties of the selected 15 selected hits is shown in Table 3.4. Many of the selected hits were shown to interact with the cytochrome P450 isoforms. Tyr46, Asp48, Ser216, Ala217 and Arg221 residues of target proteins are mostly involved in the binding of 15 hits as shown in Table 3.5. Summary of ligand-protein binding analyses of selected 15 hits is shown in Table 3.6. Synthetic accessibility score (SAS) based on the fragmental analyses of the structures of virtual hits is

acceptable for easy synthesis in laboratories. The 11 hits demonstrated no drug safety alerts. The aggregator potential through comparison with previously reported aggregators is shown only for the four hits demonstrated in Table 3.7. A lead compound was identified after a series of filters based on a CADD scheme (Figure 3.2).

Table 3.4: Summary of drug-likeness and pharmacokinetic properties of the 15 selected virtual hits.

Virtual hits	Oral bioavailability							Pharmacokinetic properties	Log Kp (skin permeation)	Water solubility	Toxicity estimation
	MW	cLogP	HBA	HBD	RB	PSA	B-Score				
ZINC06137783	450.581	3.2659	7	2	5	73.91	0.55	GI absorption BBB permeant P-gp substrate CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor	−6.19	Moderately soluble	No risk
ZINC04259062	464.545	2.8826	8	2	4	119.22	0.55	GI absorption P-gp substrate CYP1A2 inhibitor CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor	−7.34	Moderately soluble	No risk
ZINC03841413	460.553	2.4036	9	2	5	116.43	0.55	GI absorption P-gp substrate CYP3A4 inhibitor	−8.02	Soluble	No risk
ZINC04277683	458.516	3.016	8	2	4	90.98	0.55	GI absorption P-gp substrate CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor	−7.31	Moderately soluble	No risk
ZINC04259056	476.507	3.1168	8	2	4	90.98	0.55	GI absorption P-gp substrate CYP1A2 inhibitor CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor	−7.35	Moderately soluble	No risk
ZINC04259064	458.497	2.0691	9	2	4	103.87	0.55	GI absorption P-gp substrate CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor	−7.84	Soluble	No risk
ZINC05535232	415.448	3.7749	7	2	1	89.95	0.55	GI absorption CYP2C19 inhibitor CYP2C9 inhibitor CYP3A4 inhibitor	−6.90	Moderately soluble	No risk
ZINC04237088	445.521	4.4229	5	0	2	64.41	0.55	GI absorption BBB permeant CYP2C19 inhibitor CYP2C9 inhibitor CYP3A4 inhibitor	−5.91	Moderately soluble	No risk
ZINC13733603	421.448	3.5857	7	0	5	74.3	0.55	GI absorption CYP1A2 inhibitor CYP2C19 inhibitor	−5.97	Moderately soluble	No risk

								CYP2C9 inhibitor CYP2D6 inhibitor CYP3A4 inhibitor			
ZINC41585804	472.539	3.3751	7	1	6	71.47	0.55	GI absorption BBB permeant P-gp substrate CYP2C19 inhibitor CYP2C9 inhibitor CYP2D6 inhibitor	−6.20	Moderately soluble	No risk
ZINC00004749	316.308	2.4978	6	4	1	107.22	0.55	GI absorption BBB permeant P-gp substrate CYP2C19 inhibitor CYP2C9 inhibitor CYP2D6 inhibitor	−6.20	Moderately soluble	No risk
ZINC02093367	424.451	4.2868	5	0	4	61.83	0.55	GI absorption CYP2C19 inhibitor CYP2C9 inhibitor CYP3A4 inhibitor	−5.40	Moderately soluble	No risk
ZINC30731533	482.44	2.1266	10	5	4	155.14	0.55	GI absorption CYP3A4 inhibitor	−7.89	Moderately soluble	No risk
ZINC00968072	274.271	2.3608	5	4	1	90.15	0.55	GI absorption P-gp substrate	−7.02	Soluble	No risk
ZINC13722309	485.512	2.4048	9	0	5	116.82	0.55	GI absorption CYP2C19 inhibitor CYP2C9 inhibitor CYP3A4 inhibitor	−7.30	Moderately soluble	No risk

* The toxicity estimation used four major criteria (mutagenicity, tumorigenicity, reproductive effects and irritant effects).

The selected hits showed acceptable bioavailability and pharmacokinetic profile. These virtual hits follow the rule of five and Veber's rule of drug-likeness. A toxicity estimation based on four criteria was also acceptable for the 15 hits. There was no predicted mutagenic, tumorigenic, irritant and reproductive risk to a patient. The Log Kp (skin permeation) values were in an acceptable range, demonstrating that the cell permeability power of the selected candidates is sufficient. Water solubility was assessed because a weakly soluble chemical candidate restricts the in-vivo bioactivity because of less dissolution in the intestinal solutions subsequently involves in its oral administration [151]. The predicted solubility values for the selected hits were within a suitable range. In silico pharmacokinetic principles (rules that explain how a body affects/deals a drug with absorption, distribution, metabolism, and elimination) are used to assess if a drug-like compound is likely to be safe and effective for therapeutic management in patients. In

addition, an estimation of cytochrome P450 (e.g. cytochrome p4501A2, cytochrome p4502C19 and cytochrome p4502C9) was performed for its most critical isoforms by SwissADME predictor shown in Table 3.4. The cytochrome P450 is a superfamily which performs important functions such as absorption, metabolism, and elimination of a drug from liver [152]. If a drug does not metabolize and accumulate for long time in the body will cause toxicity. Many hits were shown to interact with the cytochrome P450 isoforms. In silico estimation for gastrointestinal (GI) drug absorption is acceptable, all selected hits shown positive response towards oral administration.

The molecular docking was conducted to estimate the most favorable geometry of the ligand bound to a PTPN1 macromolecule. I considered 2636 compounds for the molecular docking study. PTPN1 was used as a molecular target (PDB ID: 3EAX) [137].

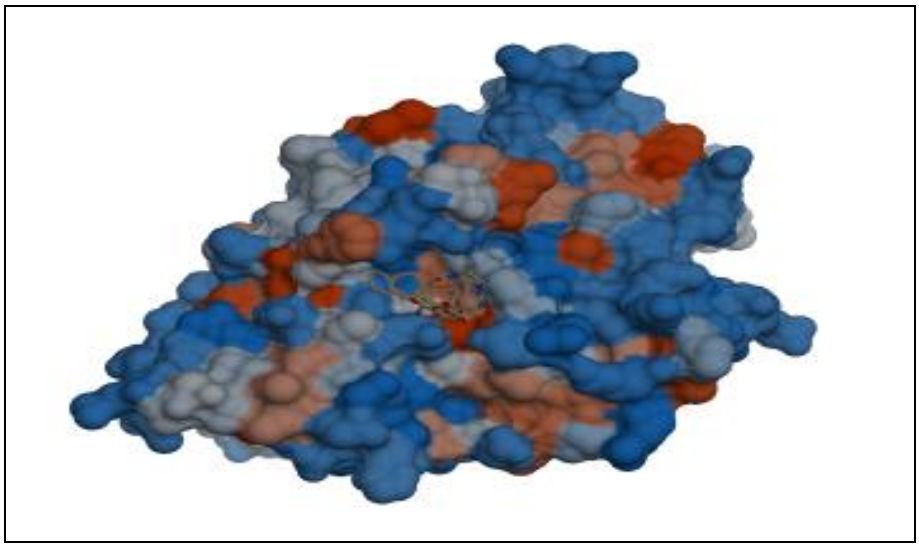


Figure 3.3: Hydrophobic surface and the active binding site of the 3EAX protein showing LZP ligand that is co-crystallized and overlaid at the active site, as generated using chimera.

I used the CLC drug discovery workbench for analyses of molecular binding of our screened dataset within the protein binding site. A comprehensive study of active site of PTPN1 X-rays crystal structure was conducted. I found that most of the protein binding site is hydrophobic. Figure 3.3 shows the molecular surface recognition of 3EAX generated by using chimera software. Figure 3.4 shows results of a detailed 2D interactions analysis of the target protein complex and three potential hits. Figure 3.4 (A) shows that Tyr46, Asp48, Val49, Ser216, Ala217, Gly220, Arg221, Gln262, and Gln266 are interacting residues within 4 Å of the protein binding site of 3EAX. Molecular docking simulations identified hydrogen and hydrophobic bindings with significant residues of PTPN1 target protein as shown in Table 3.5.

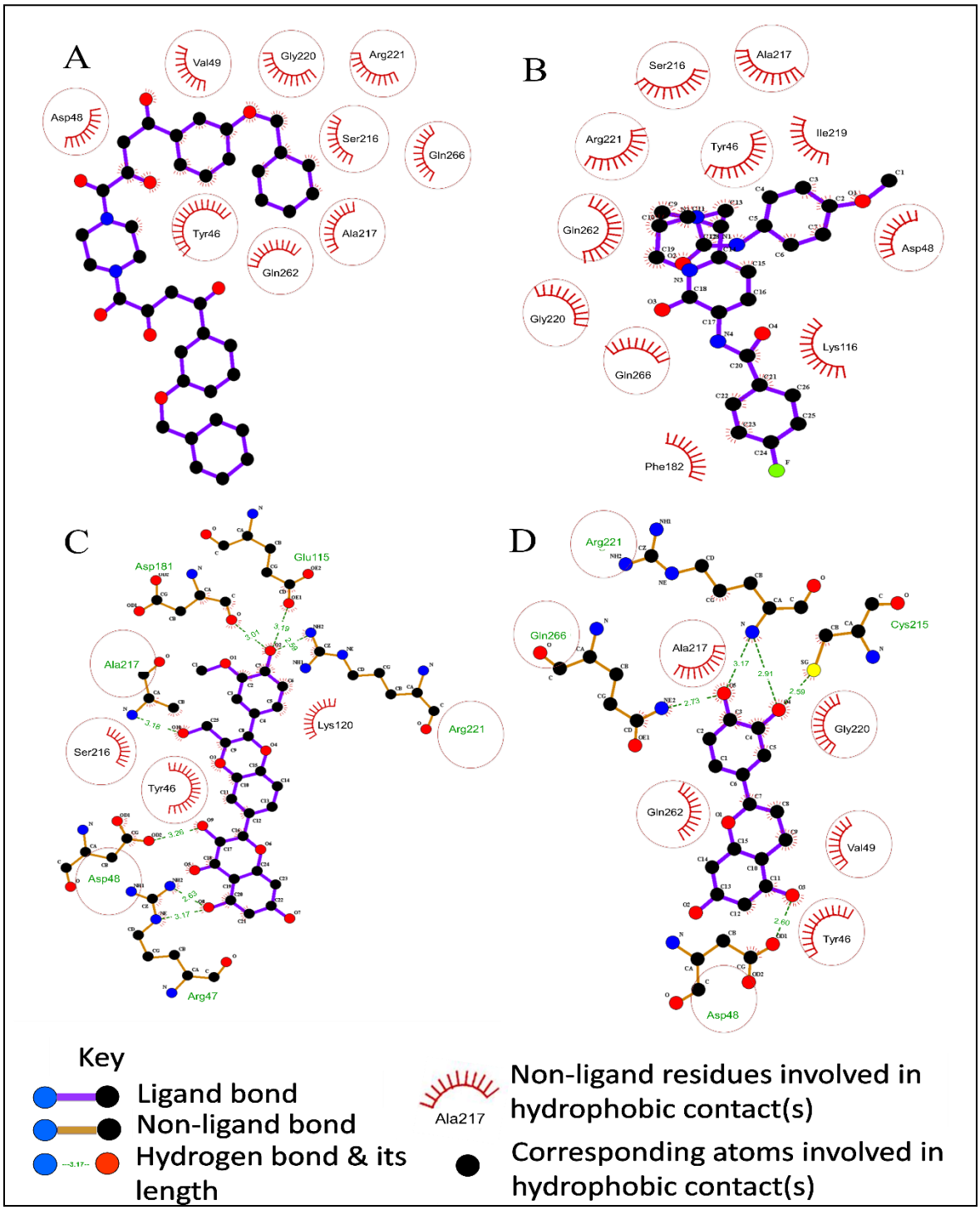


Figure 3.4: Schematic representation of the binding mode of ligands with protein tyrosine phosphatase non-receptor type 1 inhibitors (PDB ID: 3EAX). The protein site is hydrophobic and the NMR structure of the 3EAX protein complex bonded with LZP is shown in (A). Conserved interacting residues of the binding site of the

target protein bonded with the virtual hits (B). ZINC04259056 shows only hydrophobic bonding (C). ZINC30731533 shows large network of hydrophobic and hydrogen bonding (D). ZINC00968072 also shows large network of hydrophobic and hydrogen bonding. Conserved interacting residues are displayed in red circles.

Table 3.5: Conserved interacting residues within the binding site of the target protein of the top scored 15 virtual hits.

Virtual hits	Tyr46	Asp48	Val49	Ser216	Ala217	Gly220	Arg221	Gln262	Gln266
ZINC06137783	+	+	−	+	+	+	+	−	+
ZINC04259062	+	+	−	+	+	+	+	+	−
ZINC03841413	+	−	−	+	+	+	+	+	+
ZINC04277683	+	+	−	+	+	+	+	−	+
ZINC04259056	+	+	−	+	+	+	+	+	+
ZINC04259064	+	+	−	+	+	+	+	+	−
ZINC05535232	+	+	+	+	+	+	+	−	−
ZINC04237088	−	−	−	+	+	−	+	+	+
ZINC13733603	+	+	−	+	+	+	+	−	−
ZINC41585804	−	+	−	+	+	+	+	+	+
ZINC00004749	+	+	−	+	−	−	+	+	+
ZINC02093367	+	+	−	−	+	+	+	+	−
ZINC30731533	+	+	−	+	+	−	+	−	−
ZINC00968072	+	+	+	−	+	+	+	+	+
ZINC13722309	+	+	−	+	+	−	+	+	−

* If key residues are present in the binding interaction within 4Å of the binding site of the target protein, then this is represented by "+". If the residues are not present this is represented by "–".

Table 3.6: Summary of molecular docking analyses of selected 15 virtual hits

Virtual hits	RMSD	Score	Ligand conformation penalty	Interacting residues in the active binding site	No. of hydrogen bonds	No. of hydrophobic bonds	Total number of bonds
ZINC06137783	0.95	−55.125	1.76	Tyr46, Asp48, Phe182, Gly183,Cys215, Ser216, Ala217, Ile219,Gly220, Arg221, Gln266.	0	29	29
ZINC04259062	0.95	−55.133	4.07	Asp48, Lys116, Ala217, Ser216,Ile219, Gly220, Arg221, Gln262.	0	28	28
ZINC03841413	0.95	−53.418	2.25	Tyr48, Phe182, Gly183, Asp184,Cys215, Ser216, Ala217, Ile219, Gly220,Arg221,Gln262, Gln266.	0	29	29
ZINC04277683	0.95	−52.773	4.26	Tyr46, Asp48, Lys116, Phe182,Gly183, Cys215, Ser216,Ile219,Ala217,Gly220,Arg221, Gln266.	0	29	29
ZINC04259056	0.95	−52.545	4.26	Tyr46, Asp48, Lys116, Phe182,Ser216, Ala217, Ile219, Gly220, Arg221, Gln262, Gln266.	0	28	28
ZINC04259064	0.95	−51.956	4.70	Tyr46, Asp48, Lys116, Phe182,Ser216, Ala217, Ile219, Gly220, Arg221, Gln262.	1 [Lys116: (2.88Å)]	29	30
ZINC05535232	1.03	−51.241	0.35	Tyr46, Asp48, Val49, Lys116,Lys120, Cys215,Ser216, Ala217, Gly220, Arg221.	2 [Lys116: (3.00Å)] [Ser216: (3.24Å)]	33	35
ZINC04237088	1.11	−51.215	6.53	Phe182, Gly183,Cys215,	0	29	29

				Ser216, Ala217, Ile219, Gly220, Arg221, Gln262, Thr263, Gln266.			
ZINC13733603	0.68	−51.169	3.00	Phe182, Gly183, Cys215, Ser216,Ala217, Ile219, Gly220, Arg221,Gln262, Thr263, Glu266.	0	29	29
ZINC41585804	0.57	−50.542	5.69	Tyr46, Asp48, Phe182, Gly183,Ser216, Ala217,Gly220, Arg221, Gln262, Thr263, Gln266.	3 [Gln266: (3.01Å)] [Arg221: (3.12Å)] [Tyr46: (2.82Å)]	28	31
ZINC00004749	0.84	−50.387	0.36	Asp48, Tyr46, Trp179, Ser216,Arg221,Gln262,Thr263, Gln266.	4 [Asp48: (2.63Å)] [Ser216: (2.99Å)] [Arg221: (2.73Å)] [Arg221: (2.64Å)]	21	25
ZINC02093367	0.94	−50.372	3.44	Tyr46, Asp48, Cys215, Ala217,Ile219, Gln262, Arg221, Gly220.	3 [Arg47: (2.80Å)] [Arg47: (3.01Å)] [Arg47: (2.89Å)]	20	23
ZINC30731533	0.79	−50.331	2.61	Tyr46, Arg47, Asp48, Glu115,Lys120, Asp181,Ser216, Ala217, Arg221.	7 [Asp48: (3.26Å)] [Arg47: (2.63Å)] [Arg47: (3.17Å)] [Asp181: (3.01Å)] [Ala217: (3.18Å)] [Glu115: (3.29Å)] [Arg221: (2.59Å)]	30	37
ZINC00968072	0.82	−50.315	0.36	Tyr46,Val49,Asp48,Cys215, Ala217,Arg221,Gly220, Gln266.	5 [Gln266: (2.73Å)] [Arg221: (3.17Å)] [Arg221: (2.91Å)] [Cys215: (2.59Å)] [Asp48: (2.60Å)]	20	25

ZINC13722309	0.54	−50.239	3.01	Tyr46, Arg47, Asp48, Gln262, Ala217, Ile219, Cys215.	3 [Arg221: (3.12Å)] [Ser216: (3.23Å)] [Ser216: (3.10Å)]	37	40

On the basis of the best docking score, 15 hits were selected to identify a potent lead compound. The interactions of the active conformation of the best scoring 15 hits with the target protein were identified using Ligplot. The saved conformation for the docked complex was subjected to detailed interactions analyses. The docked files were uploaded to Ligplot to obtain its schematic representation of the hydrogen bonding and hydrophobic interactions. Detailed interactions of all docked complexes are shown in Table 3.6. These hits resulted from the hydrogen bonds and the hydrophobic moiety interacting with the significant residues in the range of 4 Å, and the calculated hydrogen bonds distances have been demonstrated.

The optimal binding mode of the three hits ZINC04259056, ZINC30731533 and ZINC00968072 having dock scores of −52.545, −50.331 and −50.315, is shown in Figure 3.4 (B, C, D), respectively. The 2D analyses of these docked complexes revealed the significant residues involved in the binding interactions of the selected hits. Common binding residues are marked with red circles to highlight them (Figure 3.4). The diagram provides a schematic representation of the docked complex. There are many hydrophobic interactions and so only residues are shown for clarity. Nine residues for the binding of the NMR protein (3EAX) to its ligand (LZP) are shown in Figure 3.4 (A). While (B) ZINC04259056 and (D) ZINC00968072 show that eight interacting residues are common

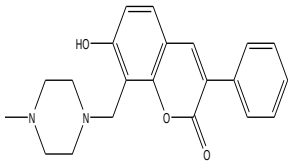
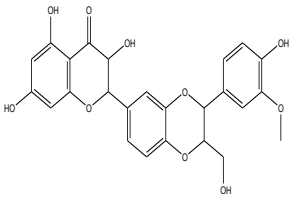
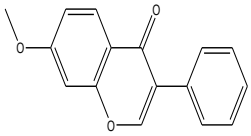
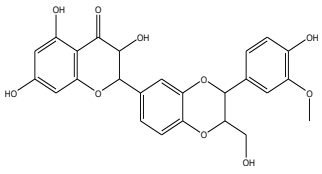
and (C) ZINC30731533 shows that five interacting residues are common to the 3EAX docked complex. Red circles are demonstrating similar binding residues (Figure 3.4).

Seven compounds show prominent hydrophobic binding interactions and eight compounds also show hydrogen bonds. Hydrogens bonds are also of great importance in PTPN1 inhibitor design. In general, for most effective inhibition; inhibitory compounds should interact with the most possible surface residues of protein binding pocket. Therefore, the inhibitory compounds must have polar amino acid and be charged anionically at functional pH [153]. In addition, the inhibitory compound must be firmly anchored by the establishment of hydrogen bonds with particular amino acid residues and inhibitory functional groups in the protein binding pocket [153,154]. However, interactions with polar amino acids will reduce the ability of PTPN1 inhibitors to cross cell membrane and to access the cytosolic PTPN1 [153]. Hydrogen and hydrophobic interaction together contribute in PTPN1 inhibition. It contributes in strong bonding of selected hits with the active binding site of respective protein. ZINC13722309 shows 40 binding interactions with the target protein, while ZINC30731533 shows 37 binding interactions and ZINC05535232 shows 35 binding interactions. ZINC30731533 shows the best binding mode with the hydrophobic moiety and the polar surface residues of the protein pocket. The dock scores of the top 15 hits are ≥ -50. The RMSD values are in the range of 0.5 to 1.1, which are considered to be acceptable values. The ligand conformation penalty is the conformational restriction energies that involve binding of flexible ligands in the protein pocket.

Aggregator advisor was used for 15 selected hits to check their aggregator possibility (Table 3.7). While using this online tool default affinity range (0.1 to 10 μM) was selected. On the basis of lipophilicity, LogP and chemical similarity thresholds (Tanimoto coefficient), three hits (ZINC03841413, ZINC04259064 and ZINC00968072) were not previously reported as aggregator and have not shown any similarity with known aggregator in the database. However, some hits were not similar to any known aggregators in the database and would require appropriate controls for possible aggregation if analyzed in vitro. Four hits (ZINC41585804, ZINC00004749, ZINC02093367, and ZINC30731533) with similarly threshold, LogP, and chemical structure of known aggregators are shown in Table 3.7 [155].

Table 3.7: Summary of aggregator advisor results and medicinal alerts for selected 15 virtual hits.

Virtual hits	Aggregator likelihood				Medicinal chemistry		
	LopP	TC	Structure of similar compound	Comments	SAS	PAINS	Brenk
ZINC06137783	4.4	-	-	Not similar to any known aggregator in in-house database.	5.13	No alerts	No alerts
ZINC04259062	3.0	-	-	Not similar to any known aggregator in in-house database.	4.82	No alerts	No alerts
ZINC03841413	2.3	-	-	Has not been previously reported as an aggregator, or to be similar to an aggregator.	5.00	No alerts	No alerts
ZINC04277683	3.1	-	-	Not similar to any known aggregator in in-house database.	4.80	No alerts	No alerts

ZINC04259056	3.3	-	-	Not similar to any known aggregator in in-house database.	4.82	No alerts	No alerts
ZINC04259064	1.9	-	-	Has not been previously reported as an aggregator, or to be similar to an aggregator.	4.83	No alerts	No alerts
ZINC05535232	3.5	-	-	Not similar to any known aggregator in in-house database.	4.51	No alerts	No alerts
ZINC04237088	4.1	-	-	Not similar to any known aggregator in in-house database.	4.68	No alerts	No alerts
ZINC13733603	4.7	-	-	Not similar to any known aggregator in in-house database.	4.28	No alerts	No alerts
ZINC41585804	4.9	72%		Reported as a colloidal aggregator.	3.82	Undesirable alerts	No alerts
ZINC00004749	2.4	78%		Reported as a colloidal aggregator.	3.82	Undesirable alerts	No alerts
ZINC02093367	6.0	71%		Reported as a colloidal aggregator.	3.94	No alerts	Undesirable alerts
ZINC30731533	1.5	100%		Reported as a colloidal aggregator.	4.92	No alerts	No alerts

ZINC00968072	2.5	-	-	Has not been previously reported as an aggregator, or to be similar to an aggregator.	3.07	Undesirable alerts	Undesirable alerts
ZINC13722309	3.0	-	-	Not similar to any known aggregator in in-house database.	4.87	No alerts	No alerts

Selected 15 hits were also checked from a medicinal chemistry perspective, in terms of the presence of any toxic moieties, PAINS also known as pan assay interference compounds and Brenk filters, to determine an oral bioavailability and drug safety profile [147]. ZINC41585804, ZINC00004749, ZINC02093367, and ZINC00968072 showed high-risk structural alerts. Consequently, these four compounds should be discarded during the initial phase of drug development to avoid possible toxic effects. The remaining 11 compounds demonstrated no drug safety alerts and potential starting points for further studies. While ZINC30731533 shows the complete similarity (100%) to compound that has been reported as aggregator (Table 3.7). The aggregator likelihood assessment suggested a reference point for identified lead compound suitable for future study.

Eleven compounds with acceptable physiochemical properties and without any expected toxicity or medicinal chemistry alerts were identified. The ZINC30731533 hit showed the best results in all the in silico protocols applied in the current study. It showed binding interactions, with a large network of hydrophobic interactions along with seven hydrogen bonds with the most important polar residues of the 3EAX NMR target protein structure. This compound did not show any toxicity risks like mutagenicity,

tumorigenicity, reproductive effects and irritant effects. The pharmacokinetic calculations were also favorable. The aggregator likelihood with previous reported active compound was 100%, and it suggested ZINC30731533 is a potent lead compound through computational methods.

ZINC30731533 is known as isosilybin (major active constituent of silymarin); an abundant flavonolignans identified in milk thistle. Silymarin is famous as Chinese traditional medicine for overindulgence of food or indigestion treatments. When an adjunct to oral diabetes therapy is used; it showed that level of fasting blood glucose reduces and maintain HbA1c in animal models and in diabetic patients. It seems to increase insulin sensitivity. But more research is required to confirm its efficacy in management of diabetes mellitus [156]. Hence ZINC30731533 is suitable for in vivo studies to validate its PTPN1 inhibitory activity, with the potential for development of an antidiabetic drug.

In this study I focused on identifying new candidates as anti-DM agents. By using structural information of already modeled PTPN1 structure (PDB id: 3EAX), detail binding behavior of PTPN1 is studied. Computational techniques were systematically used in this study to produce the results. It is worthy to note that I identified a lead compound (isosilybin; an active constituent of silymarin) as a PTPN1 inhibitor. Our recommendation is to test isosilybin in laboratories to confirm its activity as PTPN1 inhibitor [133]. I highlighted the importance of PTPN1 enzyme in this study which is involved in many biological processes, however, design and development of PTPN1 inhibitor is a hot research topic for treatment of obesity and cancer, as well as for T2DM [143]. The proposed virtual lead, isosilybin, is a flavonoid compound which shows

extensive variety of biological activities because flavonoids are beneficial for treatment of various diseases [149]. Furthermore, DM is a complicated disease, and DM patients usually have other complications along with the disease. In this regard, the proposed lead compound is expected to show multiple treatments for diabetic patients who have other complications.

3.5. Conclusion

By using computer-aided drug design methodologies successfully identified plant-derived therapeutic hits with the potential to inhibit the PTPN1 target, which may be helpful to enhance insulin production. The newly identified lead isosilybin (ZINC30731533) in this systematic study was without any predicted toxic effects and showed the best binding mode with the PTPN1 therapeutic target. Therefore, the lead compound is expected to function as an anti-diabetic drug after subsequent testing and validation. The present study could aid in the development of PTPN1 inhibitors for diabetes treatment.

CHAPTER 4

CONCLUSION AND FUTURE PROSPECTS

This dissertation aims to identify novel candidates as anti-DM agents. Drug discovery of antidiabetic drugs was delayed in previous 30 years, but now, as a result of detail understanding of molecular targets, scientific knowledge, advance technology, progress in drug design methods shows remarkable improvements.

Computer-aided molecular docking methods were applied to human insulin protein [116] and plant insulin present in *Canavalia ensiformis* to identify anti-diabetic compounds (chapter 2). Safe and effective use of natural products can ensure that plant-based medicines are more harmonious with biological systems. The use of some plants extracts, alone or in combinations with conventional diabetic medications proves as beneficial [88-90] but due to complication in the structure of phytochemicals and its bioactivities, mechanism is not clearly understood; plant extracts can show numerous favorable activities in several metabolic pathways. The present study confirms plant proteins genomic sequences are similar to those of animal insulin (table 2.1 and table 2.2) and evaluate its action with diabetic medicine. It could be therapeutically significant for diabetic patients. I selected a lead compound, which is an anti-diabetic synthetic compound with publication number: WO2007067614 shown as T6 in table 2.3 from the dataset of eight compounds that had desired biologic activities on a validated molecular target. Lead compound was screened as a novel plant insulin-based compound (through molecular docking analysis) and its four analogs were confirmed as antidiabetic agents with appropriate drug-like properties compared to the standard compound (aleglitazar).

Computer-aided approach provided information on binding energies and binding interactions of the analogs to predict their anti-DM activities. These analogs need to synthesize in lab and test for their pharmacokinetic and pharmacodynamics effects.

PTPN1 inhibition can deceases adipose tissue storage of triglyceride lower than the conditions of over-nutrition and was not related with any severe toxicity. No weight increase, indicating additional substantial benefit for anti-DM patients, who are often obese along with cardiovascular risk. I have well established computer-aided pipeline to highlight new PTPN1 inhibitors (chapters 3). Computer-based screening is applied to identify the most promising anti-DM agents from the plant-derived set of ZINC databases [133]. Screening was on the basis of the pharmacophoric features (model 1 shown in figure 3.1) of reported PTPN1 inhibitors and binding mode of PTPN1 protein structure (3EAX). Through the screening pipeline isosilybin (ZINC30731533) was screened as a PTPN1 inhibitor. Isosilybin was confirmed by oral bioavailability (Lipinski rule and veber's rule) and pharmacological activity (ADME and toxicity estimations). Isosilybin is a major active constituent of Silymarin. Silymarin demonstrated increase in insulin sensitivity in diabetic patients, but the exact mechanism of action was not clearly understood. My computer-aided approach confirmed that the isosilybin (ZINC30731533) acts as a potential PTPN1 inhibitor and mechanism of PTPN1 inhibition is clearly understood for diabetes mellitus [157]. It might lead in future development of potential PTPN1 inhibitor.

Computational methods are used to identify mode of interaction of therapeutic target and previously unknown bio-activities for known plant-derived data. Subsequently, it will add information to progress companies made functional food ingredients and

dietary supplements. At this stage, it is significant to highlight that while finding bioactive insulin-like proteins or anti-DM compounds; identification of the plant-derived proteins/compounds is likely important it will add information to chemical synthesis of novel and unique natural products which could be valuable functional food, dietary supplements or an anti-DM medicine.

In future, the results could be useful as substructures for molecular dynamic simulations and wet lab experimental studies which will not only proceed to the new vision of drug design and discovery and may offer an effective therapy for diabetes mellitus.

I have succeeded in screening novel drug candidates as anti-DM agents along with knowledge of plant extracts which possess anti-DM activity by computer-aided drug design methods. My perspective of the methods is to prevent huge cost and hectic work of wet lab experiment-based drug discovery. I have developed computer-aided screening pipelines based on open source software, off-the-shell software, and desktop personal computer. I generated and confirmed an inexpensive scheme available to the academic institutes and developing countries.

ACKNOWLEDGMENT

All praise and exaltation is due to Allah (S.W.T) the creator and sustainer of all worlds. First and foremost I would like to express my gratitude and thanks giving to Him for providing me the boundaries and blessings to complete this work.

First and foremost, I would like to express my sincere appreciation to my supervisor, Prof. Katsumi Sakata, in the Graduate School of Engineering/Environmental and Life Engineering at Maebashi institute of technology for his guidance. He has taught me, both consciously and unconsciously to prove me a good researcher for future.

I owe my sincere gratitude and appreciation to Prof. Nakamura Kensuke, in the Graduate School of Engineering/Environmental and Life Engineering at Maebashi institute of technology for his thought provoking guidance and support as a chief member of the supervisory committee. I would like to convey my cordial thanks to Prof. Hitoshi Miyazaki and Prof. Fukuchi Satoshi at Maebashi institute of technology and Prof. Toshiyuki Saito at National Institute of Radiological Science for valuable comments and advices to produce a useful piece of research work.

I am always thankful to Otsuka Toshimi Scholarship Foundation which supports me financially to fulfil my dream and embroils me in many cultural activities in Japan.

I am thankful to all my family, my teachers and my friends for their moral support. I am highly indebted to my parents, for their expectations, support, and encouragement throughout the completion of this doctorate degree.

I gratefully acknowledge Maebashi institute of Technology for providing me all the essential facilities to complete this research.

I pray to Allah (S.W.T) that may He bestow me with success in future also and shower His blessed knowledge upon me for the betterment of mankind.
Ameen

REFERENCE

[1] Gandhi, G.R., Sasikumar, P. (2012) Antidiabetic effect of Merremia emarginata Burm. F in streptozotocin induced diabetic rats. *Asian Pacific Journal of Tropical Biomedicine*, 2(4), 281-286.

[2] Li, W.L., Zheng, H.C., Bukuru, J., De Kimpe, N. (2004) Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus. *Journal of Ethnopharmacology*, 92(1), 1-21.

[3] Warjeet, S.L. (2011) Traditional medicinal plants of Manipur as antidiabetics. *Journal of Medicinal Plants Research*, 5(5), 677-687.

[4] American Diabetes Association. www.diabetes.org/diabetes-basics/statistics/?loc=db-slabnav.

[5] International Diabetes Federation. IDF Diabetes Atlas, 6th Edition. www.idf.org/sites/default/files/EN_6E_Atlas_Full_0.pdf.

[6] Bibi, S., Sakata, K. (2016) Current Status of Computer-Aided Drug Design for Type 2 Diabetes. *Current Computer Aided Drug Design*, 12(2), 167-177.

[7] Diabetes and the Eyes. http://www.mdfoundation.com.au/page57714536.aspx.

[8] Diabetes Australia.www.diabetesaustralia.com.au

[9] Malik, V.S., Popkin, B.M., Bray, G.A., Despres, J.P., Hu, F.B. (2010). Sugar Sweetened Beverages, Obesity, Type 2 Diabetes and Cardiovascular Disease risk. *Circulation*, 121(11), 1356-1364.

[10] Complication of Diabetes. http://www.idf.org/complications-diabetes.

[11] Elchebly, M., Payette, P., Michaliszyn, E., Cromlish, W., Collins, S., Loy, A.L., Normandin, D., Cheng, A., Himms-Hagen, J., Chan, C. C., Ramachandran, C., Gresser, M.J., Tremblay, M.L., Kennedy, B.P. (1999) Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. Science, 283(5407), 1544-1548.

[12] Yip, S.C., Saha, S., Chernoff, J. (2010) PTP1B: a double agent in metabolism and oncogenesis. *Trends in Biochemical Science*, 35(8), 442-9.

[13] Johnson, T.O., Ermolieff ,J., Jirousek , M.R. (2002) Protein tyrosine phosphatase 1B inhibitors for diabetes, *Nature review, drug discovery*, 1(9), 696-709.

[14] Osadebe, P.O., Odoh, U.E., Uzor, P. (2014) Natural Products as Potential Sources of Antidiabetic Drugs. *British Journal of Pharmaceutical Research*, 4(17), 2075-2095.

[15] Wetzel, S., Bon, R.S., Kumar, K., Waldmann, H. (2011) Biology-Oriented Synthesis. *Angewandte Chemie International Edition in English*, 50(46), 10800-10826.

[16] Newman, D.J., Cragg, G.M. (2007) Natural products as sources of new drugs over

the last 25 years. *Journal of Natural Products*, 70(3), 461-477.

[17] Van-Hattum, H., Waldmann, H. (2014) Biology-oriented synthesis: harnessing the power of evolution. *Journal of the American Chemical Society*, 136(34), 11853-11859.

[18] Koch, M.A., Schuffenhauer, A., Scheck, M., Wetzel, S., Casaulta, M., Odermatt, A., Ertl, P., Waldmann, H. (2005) Charting biologically relevant chemical space: A structural classification of natural products (SCONP). *Proceedings of the National Academy of Sciences*, 102(48), 17272-17277.

[19] Ganesan, A. (2008) The impact of natural products upon modern drug discovery. *Current Opinion in Chemical Biology*, 12(3), 306-317.

[20] Newman, D.J., Cragg, G.M. (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of Natural Products*, 75(3), 311-335.

[21] Patel, D.K., Kumar, R., Laloo, D., Hemalatha, S. (2012) Natural medicines from plant source used for therapy of diabetes mellitus: An overview of its pharmacological aspects. *Asian Pacific Journal of Tropical Disease*, 3, 239-250.

[22] The American Diabetes Association Guide to Herbs & Nutritional Supplements. American Diabetes Association, 2009. http://www.diabetes.org/living-with-diabetes/treatment-and-care/medication/other-treatments/herbs-supplements-and-alternative-medicines/.

[23] Khanna, P., Nag, T.N., Jain, S.C., Mohan, S. (1974) Extraction of insulin from a plant source. Leicester (UK): 3rd. International Congress on Plant Tissue and Cell Cultures.

[24] Le Roith, D., Shiloach, J., Roth, J., Lesnaik, M.A. (1980) Evolutionary origins of vertebrate hormones: Substances similar to mammalian insulins are native to unicellular eukaryotes. *Proceedings of the National Academy of Sciences*, 77(10), 6184-6188.

[25] Le Roith, D., Shiloach, J., Heffron, R., Rubinovitz, C., Tanenbaum, R., Roth, J. (1985) Insulin-related material in microbes: Similarities and differences from mammalian insulins. *Canadian Journal of Biochemistry and Cell Biology*, 63(8), 839-849.

[26] Ng, T.B., Wong, C.M., Li, W.W., Yeung, H.W. (1986) Insulin-like molecules in *Momordica charantia* seeds. *Journal of Ethnopharmacology*, 15(1), 107-117.

[27] Satyanarayana, T., Katyayani, B.M., Latha, H.E. (2006) Hypoglycemic and antihyperglycemic effect of alcoholic extract of Euphorbia leucophylla and its fractions in normal and in alloxan induced diabetic rats. *Pharmacognosy Magazine*, 8(2), 244-253.

[28] Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R., Schacht, A.L. (2010) How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, 9(3), 203-214.

[29] Hurle, M.R., Yang, L., Xie, Q., Rajpal, D.K., Sanseau, P., Agarwal, P. (2013) Computational drug repositioning: from data to therapeutics. *Clinical Pharmacology & Therapeutics*, 93(4), 335-341.

[30] Clark, D.E. (2006) What has computer-aided molecular design ever done for drug discovery? *Expert Opinion on Drug Discovery*, 1(2), 103-110.

[31] Talele, T.T., Khedkar, S.A., Rigby, A.C. (2010) Successful applications of computer aided drug discovery: Moving drugs from concept to the clinic. *Current Topics in Medicinal Chemistry*, 10(1), 127-141.

[32] Kick, E.K., Roe, D.C., Skillman, A.G., Liu, G., Ewing, T.J., Sun, Y., Kuntz, I.D., Ellman, J.A. (1997) Structure-based design and combinatorial chemistry yield low nano molar inhibitors of cathepsin D. *Chemistry & Biology*, 4(4), 297-307.

[33] Paiva, A.M., Vanderwall, D.E., Blanchard, J.S., Kozarich, J.W., Williamson, J.M., Kelly, T.M. (2001) Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. Biochimica et Biophysica Acta, 1545(1-2), 67-77.

[34] Doman, T.N., McGovern, S.L., Witherbee, B.J., Kasten, T.P., Kurumbail, R., Stallings, W.C., Connolly, D.T., Shoichet, B.K. (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1b. *Journal of Medicinal Chemistry*, 45(11), 2213-2221.

[35] Wyss, P.C., Gerber, P., Hartman, P.G., Hubschwerlen, C., Locher, H., Marty, H.P., Stahl, M. (2003) Novel dihydrofolate reductase inhibitors. Structure-based versus diversity-based library design and highthroughput synthesis and screening. *Journal of Medicinal Chemistry*, 46 (12), 2304-2312.

[36] Oshiro, C., Bradley, E.K., Eksterowicz, J.,Evensen, E., Lamb, M.L., Lanctot, J.K., Putta, S., Stanton, R., Grootenhuis, P.D. (2004) Performance of 3d-database molecular docking studies into homology models. *Journal of Medicinal Chemistry*, 47(3), 764-767.

[37] The top 10 causes of death. www.who.int/mediacentre/factsheets/ fs310/en/.

[38] Google scholar. scholar.google.com

[39] Workman, P., Collins, I. (2014) Modern cancer drug discovery: Integrating targets, technologies and treatments. In Cancer drug design and discovery, 2nd edition, Neidle, S., Ed. Academic Press, 3-53.

[40] Carrieri, A., Perez-Nueno, V.I., Lentini, G., Ritchie, D.W. (2013) Recent trends and future prospects in computational gpcr drug discovery: From virtual screening to poly pharmacology. *Current Topics in Medicinal Chemistry*, 13, 1069-1097.

[41] PhRMA 2015 Biopharmaceutical Research Industry Profile. http://www.phrma.org/sites/default/files/pdf/2015_phrma_profile.pdf.

[42] Mullard, A. (2014) New drugs cost us$2.6 billion to develop. *Nature Reviews Drug Discovery*, 13, 877.

[43] Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E.W. Jr. (2014) Computational methods in drug discovery. *Pharmacological Reviews*, 66, 334-395.

[44] Acharya, C., Coop, A., Polli, J.E., Mackerell, A.D.J. (2011) Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach. *Current Computer Aided-Drug Design*, 7(1), 10-22.

[45] Ripphausen, P., Nisius, B., Peltason, L., Bajorath, J. (2010) Quo vadis, virtual screening? A comprehensive survey of prospective applications. *Journal of Medicinal Chemistry*, 53(24), 8461-8467.

[46] Agarwal, A.K., Fishwick, C.W. (2010) Structure-based design of anti-infectives. *Annals of the New York Academy of Sciences*, 1213, 20-45.

[47] Advanced Chemistry Development. www.acdlabs.com.

[48] MDL Drug Data Report. lib.stanford.edu/mdl-drug-data-report.

[49] MDPI Database. www.mdpi.org.

[50] National Cancer Institute Open Database Compounds.cactus.nci.nih.gov.

[51] Thomson Index Chemicus database. thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-anddiscovery/index-chemicus.html.

[52] Tripos Discovery Research Screening Libraries. www.tripos.com.

[53] ZINC database: University of California, San Francisco, USA. www.zinc.docking.org.

[54] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 16, 2785-2791.

[55] CLC drug discovery work bench/Molegro Virtual Docker: CLC Inc A, Denmark, www.clcbio.com/products/clc-drug-discovery-workbench

[56] Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.F., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L., Kuntz, I.D. (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA*, 15(6), 1219-1230.

[57] Kramer, B., Rarey, M., Lengauer, T. (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins*, 37(2), 228-241.

[58] Mcgann, M.R., Almond, H.R., Nicholls, A., Grant, J.A., Brown, F.K. (2003) Gaussian docking functions. *Biopolymers*, 68(1), 76-90.

[59] Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal*

Chemistry, 47(7), 1739-1749.

[60] Verdonk, M.L.; Cole, J.C.; Hartshorn, M.J.; Murray, C.W.; Taylor, R.D. (2003) Improved protein–ligand docking using GOLD. *Proteins*, 52(4), 609-623.

[61] MOE: Chemical Computing Group. www.chemcomp.com.

[62] Catalyst, Accelrys. www.accelrys.com.

[63] Wolber, G., Langer, T. (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *Journal of Chemical Information and Modeling*, 45(1), 160-169.

[64] Wolber, G., Dornhofer, A., Langer, T. (2006) Efficient overlay of small organic molecules using 3D pharmacophores. *Journal of Computer-Aided Molecular Design*, 20(12), 773-788.

[65] Dixon, S.L., Smondyrev, A.M., Knoll, E.H., Rao, S.N., Shaw, D.E., Freisner, R.A. (2006) PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D databases screening: 1. Methodology and preliminary results. *Journal of Computer-Aided Molecular Design*, 20(10-11), 647-671.

[66] Wolber, G., Seidel, T., Bendix, F., Langer, T. (2008) Molecule-pharmacophore super positioning and pattern matching in computational drug design. *Drug Discovery Today*, 13(1–2), 23-29.

[67] Cramer, R.D., Patterson, D.E., Bunce, J.D. (2002) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, 110(18), 5959-5967.

[68] Klebe, G., Abraham, U., Mietzner, T. (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of Medicinal Chemistry*, 37(24), 4130-4146.

[69] CODESSA. www.semichem.com/codessa/.

[70] Mauri, A., Consonni, V., Pavan, M., Todeschini, R. (2006) DRAGON software: An easy approach to molecular descriptor calculations. *MATCH Communications in Mathematical and in Computer Chemistry*, 56(2), 237-248.

[71] VLife. www.vlifesciences.com.

[72] SYBYL-XSuite, Certara USA, Inc. www.certara.com/software/ molecular-modeling-and-simulation/sybyl-x/.

[73] Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz-Jr K.M., Onufriev, C., Simmerling, C., Wang, B., Woods, R.J. (2005) The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16), 1668-1688.

[74] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. (1983) Charmm – a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2), 187-217.

[75] Christen, M., Hunenberger, P.H., Bakowies, D., Baron, R., Burgi, R., Geerke, D.P., Heinz, T.N., Kastenholz, M.A., Krautler, V., Oostenbrink, C., Peter, C., Trzesniak, D., Van Gunsteren, W.F. (2005) The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 26(16), 1719-1751.

[76] Bibi, S., Kulsoom, S., Rashid, H., Sakata, K. (2015) Lead identification and optimization of plant insulin-based antidiabetic drugs through molecular docking analyses. *International Journal of Pharmacy and Pharmaceutical Sciences*, 7(3), 337-343.

[77] Bliss, M. (1982) The discovery of insulin. Chicago: The University of Chicago Press.

[78] Kahn, B.B., Flier, J.S. (2000) Obesity and insulin resistance. *Journal of Clinical Investigation*, 106(4), 473-481.

[79] Lambier, A.M., Scharpe, S., Meester, D. I. (2008) DPP4 inhibitor for diabetes – what next? *Biochemical Pharmacology*, 76, 1637-1643.

[80] Montecucco, F., Steffens, S., Mach F. (2008) Insulin resistance: a pro inflammatory state mediated by lipid-induced signaling dysfunction and involved in atherosclerotic plaque instability. *Mediators of Inflammation*, 20(3), 53-57.

[81] Leibiger, I.B., Leibiger, B., Berggren, P.O. (2008) Insulin signaling in the pancreatic beta-cell. *Annual Review of Nutrition*, 28, 233-251.

[82] Henry, R.R. (2003) Insulin resistance: from predisposing factor to therapeutic target in type2 diabetes. *Clinical Therapeutics*, 25(1), 47-63.

[83] Guo, L., Tabrizchi, R. (2006) Peroxisome proliferator-activated receptor gamma as a drug target in the pathogenesis of insulin resistance. *Pharmacology & Therapeutics*, 111(1), 145-173.

[84] Kuhn, T.S. (1970) The structure of scientific revolutions. Chicago: The University of Chicago Press.

[85] Best, C.H., Smith, R.C., Scott, D.A. (1924) An insulin like material in various tissues of the normal and diabetic animal. *American Journal of Physiology*, 68, 161-182.

[86] Collip, J.P. (1923) Glucokinin. A new hormone in plant tissue. *Journal of Biological Chemistry*, 56(1923), 513-543.

[87] Ellis, M.M., Eyster, W.H. (1923) Some effects of insulin and glucokinin on maize seedlings. *Science*, 58(1513), 541-542.

[88] Tongia, A., Tongia, S.K., Dave. M. (2004) Phytochemical determination and extraction of Momordica charantia fruit and its hypoglycemic potentiation of oral hypoglycemic drugs in diabetes mellitus (NIDDM). *Indian journal of physiology and pharmacology*,48 (2),241-244.

[89] Poonam, T., Prakash, G.P., Kumar, L.V. (2013) Interaction of *Momordica*

charantia with metformin in diabetic rats. *American Journal of Pharmacology and Toxicology*, 8, 102–106.

[90] Vijaya, K., Sunitha, S.P, Husssain, J.A, Sandhya, P., Sujatha, D., Gopireddy, G. (2014) Synergistic antihyperglycemic, antihyperlipidemic and antioxidant effects of *Momordica charantia* and metformin in streptozotocin induced diabetic rats. *World Journal of Pharmaceutical Sciences*. 3, 1901-1980.

[91] Sridhar, G.R., Divakar, C.H., Hanuman, T., Rao, A.A. (2006) Bioinformatics approach to extract information from genes. *International Journal of Diabetes in Developing Countries,* 26(4), 149-151.

[92] Sridhar, G.R., Laksluni, P.V., Apparao, A. (2006) Phylogenetic tree construction of butyrylcholinesterase sequences in life forms. *Journal of the Association of Physicians of India*, 54, 122-123.

[93] Xavier-Filho, J., Oliveira, A.E.A., Silva, L.B., Azevedol, C.V., Venâncio, T.M., Machado, O.L.T., Oliva, M.L., Fernandes, K.V.S., Xavier-Neto, J. (2003) Plant insulin or glucokinin: a conflicting issue. *Brazilian Journal of Plant Physiology,* 15(1), 67-78.

[94] Koona, S.J., Kudipudi, S., Sridhar, G. R., Rao, S.B., Apparao, A. (2010) Plant insulin: an *in silico* approach. *International Journal of Diabetes in Developing Countries,* 30(4), 191-193.

[95] Asolker, L.V., Kakkar, K.K., Chakre, O.J. (2000) Supplement to glossary of Indian medicinal plants. Part-I (A-K). National Institute of Science Communication, New Delhi, 116-117.

[96] Rajanarayana, K., Reddy, M.S., Chaluvadi, M.R., Krishna, D.R. (2001) Bioflavonoids classification, pharmacological, biochemical effects and therapeutic potential. *Indian Journal of Pharmacology,* 33, 2-16.

[97] Chopda, M.Z., Mahajan, R.T. (2009) Wound Healing Plants of Jalgaon District of Maharashtra state, India. *Ethnobotanical Leaflets*, 13, 1-32.

[98] Enyikwola, O., Addy, E.O., Adoga, G.I. (1991) Hypoglycemic effect of *Canavalia ensiformis* (Leguminosae) in albino rats. *Discovery and Innovation*, 3(3), 61-63.

[99] Rachel, N.U. (2003) Effect of aqueous extract of Canavalia ensiformis seeds on hyperlipidaemia and hyperketonaemia in alloxan-induced diabetic rats. *Biokemistri*, 15(1), 7-15.

[100] Herklots, G.A.C. (1972) Vegetables in South-East Asia. London: George Allen & Unwin Ltd.

[101] Venancio, T.M., Oliveira, A.E.A., Silva, L.B., Machado, O.L.T., Fernandes, K.V.S., Xavier-Filho, J. (2003). A protein with sequence homology to bovine insulin is found in the legume Vigna unguiculata (cowpea). *Brazilian Journal of Medical and Biological Research*, 36(9), 1167-1173.

[102] Bibi, S., Kulsoom, S., Rashid, H. (2013) Ligand based approach for pharmacophore generation for identification of novel compounds having

antidiabetic activity. *International Journal of Pharmacy and Pharmaceutical Sciences*, 5(4), 303-314.

[103] Blast P. https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins. National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA

[104] Trott, O., Olson, A.J. (2010) Auto Dock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2), 455-461.

[105] Gasteiger, J., Marsili, M. (1980) Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron*, 36(22), 3219-3228.

[106] Humphrey, W., Dalke, A., Schulten, K. (1996) VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33-38.

[107] Pieper, U., Webb, B.M., Dong, G.Q., Schneidman-Duhovny, D., Fan, H., Kim, S.J., Khuri, N., Spill, Y.G., Weinkam, P., Hammel, M., Tainer, J.A., Nilges, M., Sali, A. (2014) MODBASE, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research*, 42, 336-346.

[108] Oliveira, A.E.A., Machado, O.L.T., Gomes, V.M., Xavier-Neto, J., Pereira, A.C., Vieira, J.G.H., Fernandes, K.V.S. Filho, J.X. (1999) Jack bean seed coat contains a protein with complete sequence homology to bovine insulin. *Protein Peptide Letter*, 6(1), 15-21.

[109] Stirban, A.O., Andjelkovic, M., Heise, T., Nosek, L., Fischer, A., Gastaldelli, A., Herz, M. (2016) Aleglitazar, a dual peroxisome proliferator-activated receptor-α/γ agonist, improves insulin sensitivity, glucose control and lipid levels in people with type 2 diabetes: findings from a randomized, double-blind trial. *Diabetes Obesity and Metabolism*, 18(7), 711-5.

[110] Wang, Y.L., Xiao, J., Suzek, T.O., Zhang, J., Wang, J., Bryant, S.H. (2009) Pub Chem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, 37, 623-633.

[111] Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1-3), 3-26.

[112] Bibi, S., Kulsoom, S., Rashid, H. (2013) Ligand based approach for pharmacophore generation for identification of novel compounds having antidiabetic activity. *International Journal of Pharmacy and Pharmaceutical Sciences*, 5(4), 303-314.

[113] Mills, N. (2006) ChemDraw ultra 8.0. *Journal of the American Chemical Society*, 28(141), 13649-13650.

[114] Tao, J., Wendler, P., Connelly, G., Lim, A., Zhang, J., King, M., Li, T., Silverman,

J.A., Schimmel, P.R., Tally, F.P. (2000) Drug target validation: lethal infection blocked by inducible peptide. *Proceedings of the National Academy of Sciences*, 97(2), 783-786.

[115] Boruah, L., Das, A., Nainwal, L.M., Agarwal, N., Shankar, B. (2013) In-silico drug design: a revolutionary approach to change the concept of current drug discovery process. *International Journal of Biological & Pharmaceutical Research*, 1(1), 60-73.

[116] Bibi, S., Kulsoom, S., Rashid, H. (2013) Insilico approach for lead identification and optimization of antidiabetic compounds. *IOSR Journal of Pharmacy and Biological Sciences*, 7(3), 36-46.

[117] Jones, G., Willett, P. (1995) Docking small-molecule ligands into active sites. *Current Opinion in Chemical Biology*, 6(6), 652-656.

[118] Sawyer, J.S., Anderson, B.D., Beight, D.W. (2003) Synthesis and activity of new aryl-and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-beta type I receptor kinase domain. *Journal of Medicinal Chemistry*, 46(19), 3953-3956.

[119] Veber, D.F., Johnson, S.R., Cheng, H., Smith, B.R., Ward, K.W., Kopple, K.D. (2002) Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 45(12), 2615-2623.

[120] Chauhan, A., Sharma, V.K., Srivastava, P., Kumar, N., Dudhe, R. (2010) Plants having potential antidiabetic activity: a review. *Der Pharmacia Lettre*, 2(3), 369-387.

[121] Sivananthan, M., Elamaran, M. (2013) In vitro evaluation of antibacterial activity of chloroform extract Andrographis paniculata leaves and roots, Durio zibethinus wood bark and Psidium guajava leaves against selected bacterial strains. *International Journal of Biomolecules and Biomedicine*, 3(1), 12-19.

[122] Li, W.L., Zheng, H.C., Bukuru, J., DeKimpe, N. (2004) Natural medicines used in the traditional Chinese medical system for therapy of diabetes mellitus. *Journal of Ethnopharmacology*, 92(1), 1-21.

[123] Potterat, O.; Hamburger, M. (2008) Drug discovery and development with plant-derived compounds. *Progress in Drug Research*, 65, 45, 47-118.

[124] Irons, B.K., Minze, M.G. Drug treatment of type 2 diabetes mellitus in patients for whom metformin is contraindicated. *Diabetes, Metabolic Syndrome and Obesity*, 2014, 18(7), 15-24.

[125] Tiganis, T. (2013) PTP1B and TCPTP—non redundant phosphatases in insulin signaling and glucose homeostasis. *FEBS Journal*, 280(2), 445-458.

[126] Mohler, M.L., He, Y., Wu, Z., Hwang, D.J., Miller, D.D. (2009) Recent and emerging anti-diabetes targets. *Medicinal Research Reviews*, 29(1), 125-195.

[127] Carpino, P.A., Goodwin, B. (2010) Diabetes area participation analysis: a review of companies and targets described in the 2008 - 2010 patent literature. *Expert*

Opinion on Therapeutic Patents, 20(12), 1627-1651.

[128] Wang, L.J., Jiang, B., Wu, N., Wang, S.Y., Shi, D.Y. (2015) Natural and semisynthetic protein tyrosine phosphatase 1B (PTP1B) inhibitors as anti-diabetic agents. *RSC Advances*, 5(60), 48822-48834.

[129] Sun, J., Qu, C., Wang, Y., Huang, H., Zhang, M., Li, H., Zhang, Y., Wang, Y., Zou, W. (2016) PTP1B, A potential target of type 2 diabetes mellitus. *Molecular Biology*, 5(4), 174.

[130] Klaman, L.D., Boss,O., Peroni, O.D., Kim, J.K., Martino, J.L., Zabolotny, J.M., Moghal, N., Lubkin, M., Kim, Y.B., Sharpe, A.H., Stricker-Krongrad, A., Shulman, G.I., Neel, B.G., Kahn, B.B. (2000) Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Molecular and Cellular Biology*, 20(15), 5479-5489.

[131] Combs, A.P. (2010) Recent advances in the discovery of competitive protein tyrosine phosphatase 1B inhibitors for the treatment of diabetes, obesity, and cancer. *Journal of Medicinal Chemistry*, 53(6), 2333-2344.

[132] Jiang, C.S., Liang, L.F., Guo, Y.W. (2012) Natural products possessing protein tyrosine phosphatase 1B (PTP1B) inhibitory activity found in the last decades. *Acta Pharmacologica Sinica*, 33(10), 1217-1245.

[133] Bibi, S., Sakata, K. (2017) An Integrated Computational Approach for Plant-based Protein Tyrosine Phosphatase Non-Receptor Type 1 Inhibitors. *Current Computer-Aided Drug Design*, 13(999).

[134] Chen, R.M., Hu, L.H., An, T.Y., Li, J., Shen, Q. (2002) Natural PTP1B inhibitors from Broussonetia papyrifera. *Bioorganic & Medicinal Chemistry Letters*, 12(23), 3387-3390.

[135] Bae, E.Y., Na, M., Njamen, D., Mbafor, J.T., Fomum, Z.T., Cui, L., Choung, D.H., Kim, B.Y., Oh, W.K., Ahn, J.S. (2006) Inhibition of protein tyrosine phosphatase 1B by prenylated iso flavonoids isolated from the stem bark of Erythrina addisoniae. *Planta Medica*, 72(10), 945-948.

[136] Nguyen, P.H., Le, T.V., Thuong, P.T., Dao, T.T., Ndinteh, D.T., Mbafor, J.T., Kang, K.W., Oh, W.K. (2009) Cytotoxic and PTP1B inhibitory activities from Erythrina abyssinica. *Bioorganic & Medicinal Chemistry Letters*, 19(23), 6745-6749.

[137] Dao, T.T., Nguyen, P.H., Thuong, P.T., Kang, K.W., Na, M., Ndinteh, D.T., Mbafor, J.T., Oh, W.K. (2009) Pterocarpans with inhibitory effects on protein tyrosine phosphatase 1B from Erythrina lysistemon Hutch. *Phytochemistry*, 70(17), 2053-2057.

[138] Nguyen, P.H., Sharma, G., Dao, T.T., Uddin, M.N., Kang, K.W., Ndinteh, D.T., Mbafor, J.T., Oh, W.K. (2012) New prenylated isoflavonoids as protein tyrosine phosphatase 1B (PTP1B) inhibitors from Erythrina addisoniae. *Bioorganic & Medicinal Chemistry*, 20(21), 6459-6464.

[139] Choi, J.S., Islam, M.N., Ali, M.Y., Kim, Y.M., Park, H.J., Sohn, H.S., Jung, H.A. (2014) The effects of C-glycosylation of luteolin on its antioxidant, anti-Alzheimer's disease, anti-diabetic, and anti-inflammatory activities. Achieves of *Pharmacal Research*, 37(10), 1354-1363.

[140] Koes, D.R., Camacho, C.J. (2012) ZINCPharmer: pharmacophore search of the ZINC database. *Nucleic Acids Research*, 40, 409-414.

[141] Sander, T., Freyss, J., Korff, M.V., Rufener, C. (2015) DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling*, 55(2), 460-473.

[142] Pajouhesh, H., Lenz, G.R. (2005) Medicinal chemical properties of successful central nervous system drugs. *NeuroRx*, 2(4), 541-553.

[143] Liu, S., Zeng, L.F., Wu, L., Yu, X., Xue, T., Gunawan, A.M., Long, Y.Q., Zhang, Z.Y. (2008) Targeting inactive enzyme conformation: aryl diketoacid derivatives as a new class of PTP1B inhibitors. *Journal of the American Chemical Society*, 130(50), 17075-17084.

[144] Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612.

[145] Laskowski, R.A., Swindells, M.B. (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778-2786.

[146] Ntie-Kang, F.; Lifongo, L.L.; Mbah, J.A.; Owono-Owono, L.C.; Megnassan, E.; Mbaze, L.M.; Judson, P.N.; Sippl, W.; Efange, S.M. In silico drug metabolism and pharmacokinetic profiles of natural products from medicinal plants in the Congo basin. In Silico Pharmacol., 2013, 1, 12.

[147] Daina, A., Michielin, O., Zoete, V. (2014) iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *Journal of Chemical Information and Modeling*, 54(12), 3284-3301.

[148] Aggregator Advisor. Shoichet Laboratory at UCSF. University of California, San Francisco, USA. http://advisor.bkslab.org/.

[149] Yao, L.H., Jiang, Y.M., Shi, J., Tomas-Barberan, F.A., Datta, N., Singanusong, R., Chen, S.S. (2004) Flavonoids in food and their health benefits. *Plant Foods for Human Nutrition*, 59(3), 113-122.

[150] Nogara, P.A.; Saraiva, R.de.A.; Bueno, D.C.; Lissner L.J.; Corte, C.L.D; Braga M.M.; Rosemberg, D.B.; Rocha, J.B.T. Virtual screening of acetylcholinesterase inhibitors using the lipinski's rule of five and ZINC databank. Biomed. Res. Int., 2015, 2015, 870389.

[151] Khadka, P., Ro, J., Kim, H., Kim. I., Kim, J.T., Kim, H., Cho, J.M., Yun, G., Lee,

J. (2014) Pharmaceutical particle technologies: An approach to improve drug solubility, dissolution and bioavailability. *Asian Journal of Pharmaceutical Sciences*, 9(6), 304-316.

[152] Ogu, C.C., Maxa, J.L. (2000) Drug interactions due to cytochrome P450. *Proceedings (Baylor University. Medical Center)*, 13(4), 421-423.

[153] Wang, Q., Gao, J., Liu, Y., Liu, C. (2012) Molecular dynamics simulation of the interaction between protein tyrosine phosphatase 1B and aryl diketoacid derivatives. *Journal of Molecular Graphics and Modelling*, 38, 186-193

[154] Combs, A.P., Yue, E.W., Bower, M., Ala, P.J., Wayland, B., Douty, B., Takvorian, A., Polam, P., Wasserman, Z., Zhu, W., Crawley, M.L., Pruitt, J., Sparks, R., Glass, B., Modi, D., McLaughlin, E., Bostrom, L., Li, M., Galya, L., Blom, K., Hillman, M., Gonneville, L., Reid, B.G., Wei, M., Becker-Pasha, M., Klabe, R., Huber, R., Li, Y., Hollis, G., Burn, T.C., Wynn, R., Liu, P., Metcalf, B. (2005) Structure-based design and structure-based design and discovery of protein tyrosine phosphatase inhibitors incorporating novel isothiazolidinone heterocyclic phosphotyrosine mimetics. *Journal of Medicinal Chemistry*, 48(21), 6544-6548.

[155] Ferreira, R.S., Simeonov, A., Jadhav, A., Eidam, O., Mott, B.T., Keiser, M.J., McKerrow, J.H., Maloney, D.J., Irwin, J.J., Shoichet, B.K. (2010) Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. *Journal of Medicinal Chemistry*, 53(13), 4891-4905.

[156] Petrie, M., McKay, G., Fisher, M. (2015) Silymarin. Drug notes. *Practical Diabetes*, 32(4), 148-504.

[157] Qian, S., Zhang, M., He, Y., Wang, W., Liu, S. (2016) Recent advances in the development of protein tyrosine phosphatase 1B inhibitors for Type 2 diabetes. *Future Medicinal Chemistry*, 8(11),1239-1258.