Study of protein interaction in Apoptotic pathway in Diabetes

A Major Project Dissertation submitted
In partial fulfilment of the requirement for the degree of

Master of Technology
In
Bioinformatics

Submitted by

Binod Koirala

(2K11/BIO/03)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. PRAVIR KUMAR

Associate Professor

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, New Bawana Road, Delhi-110042

DELHI TECHNOLOGICAL UNIVERSITY, New Delhi

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"Study of Protein interaction in Apoptotic pathway in Diabetes"**, submitted by **Binod Koirala (2K11/BIO/03)** in partial fulfillment of the requirement for the award of the degree of Master of Technology, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by him under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date:

Dr. Pravir Kumar
Associate Professor
Department of Bio-Technology
Delhi Technological University
(Formerly Delhi College of Engineering, University of Delhi)

ACKNOWLEDGEMENT

I express my sincere thanks and deepest gratitude to my guide **Dr. Pravir Kumar** (Associate Professor) Department of Biotechnology, Delhi Technological University for giving his sagacious guidance, advice and supervision in compiling the work presented here

I am greatly thankful to all respected faculty of the Department of Biotechnology, DTU for their efficient Teaching, generous support and providing me clear concept in Biotechnology.

Mere words are not enough to express my feelings towards my Parents**.** It was their dream and guidance that made me strong enough to pursue further studies.

Binod Koirala
2K11/BIO/03

DECLARATION

I hereby declare that the work, which is being presented in project entitled **"Study of protein interaction in Apoptotic pathway in Diabetes"** in partial fulfilment of the requirement for the award of MASTER OF TECHNOLOGY in BIOINFORMATICS degree, is an authentic record of my own work carried out under the supervision of Dr. Pravir Kumar, Associate Professor, Department of Biotechnology, Delhi Technological University.

The project was undertaken as a part of academic curriculum according to the University rules and norms and it has not commercial interest and motive, it is my original work. It is not submitted to any other organization for any purpose.

BINOD KOIRALA

Roll No. 2K11/BIO/03

CONTENT

TOPIC	PAGE NO.
List of Figure	*... VII*
List of Tables	*... VII*
List of Abbreviations	*... VIII*
1. ABSTRACT	… 9
2. INTRODUCTION	… 10
2.1Background	
2.2 Aim and Objective of the Study	
3. REVIEW OF LITERATURE	… 12
3.1Apoptosis	
3.2 Apoptotic pathway regulation	
3.3 Heat shock proteins and apoptosis	
3.4 Drug Designing	
3.5 Structure based drug designing	
3.6 Ligand based drug designing	
3.7 Molecular docking	
3.8 Docking algorithims	
3.9 Virtual screening	
4. METHODOLOGY	… 22
4.1Protein-protein interaction	
4.2 Shortest pathway analysis	
4.3 Physiochemical parameter computation	
4.4 Findsite	
4.5 Argus lab	
4.6 Jmol	

5. RESULTS … 30

5.1Bcl-2 interaction network
5.2Caspase interaction network
5.3Shortest pathway analysis
5.4Physiochemical parameter computation
5.5Study of Bcl-2
5.6 Virtual Screening of Bcl-2
5.7 Docking of Zinc and KEGG compounds

6. DISCUSSION … 49

7. CONCLUSION … 50

8. REFERENCES … 51

9. APPENDIX … 55

9.1 ZINC Database

9.2 KEGG database

9.3 Lipinski rule

LIST OF FIGURES

Figure 1: Protocol of Structure based Drug designing … 18

Figure 2: Target based virtual screening … 21

Figure 3: Flowchart of the virtual database screening … 28

Figure 4: Protein-protein interaction network of Bcl-2 … 30

Figure 5: Protein-protein interaction network of Caspase … 32

Figure 6: Virtual Screened Zinc compounds … 41

Figure 7: Virtual Screened KEGG compounds … 43

LIST OF TABLES

Table 1: Physio-Chemical parameters of interacted proteins … 37

Table 2: 3D-View of Zinc compounds … 45

Table 3: 3D-View of KEGG compounds … 47

Table 4: Docking score … 48

LIST OF ABBREVIATION

3D:	3 Dimensions
Apaf-1:	Apoptotic protease activating factor 1
BAX:	Bcl-2–associated X protein
Bcl2:	B cell lymphoma 2
DED:	Death effector domain
FADD:	Fas associated death domain
GSK3β:	Glycogen Synthase kinase 3 Beta
Hif1α:	Heat Inducible factor 1 alpha
HSPs:	Heat Shock proteins
MSA:	Multiple Sequence Alignment
PDB:	Protein Databank
PDK1:	Phosphoinositide-dependent protein kinase-1
PDK2:	Phosphoinositide-dependent protein kinase-2
PI3K:	Phosphoinositide-3-Kinase
PIP2:	Phosphatidylinositol biphosphate 2
PIP3:	Phosphatidylinositol biphosphate 3
PKB:	Protein kinase B
PPI:	Protein-protein Interaction
SBDD:	Structure-Based Drug Design
SGK1:	Serum and Glucose Induced Kinase 1
T1D:	Type1 Diabetes
T2D:	Type2 Diabetes

Study of Protein Interaction in Apoptotic Pathway in Diabetes

Binod Koirala
Delhi Technological University

ABSTRACT

Diabetes is the common non-communicable disease. At the molecular level, pancreatic β cell decreased due to insulin resistance, disturbance in Glut-receptors and excessive accumulation of sugar in the blood. These molecular events trigger the apoptotic pathway that plays a major role in the development of insulin deficiency and the progression of the disease. Proteins that are included in the Bcl-2 family and caspase family are the major regulators of programmed cell death. The proteins interacted with Bcl-2 and caspase also has the important role in this process. The study of different kinds of physiochemical parameters of these proteins is helpful to understand the apoptotic pathway. As Bcl-2 is a major regulator of this process, the Bcl-2 inhibitors are promising drugs for diabetes. *In-silico* screening of compounds from free databases (ZINC and KEGG) of more than 2 million structures is carried to find out the potential inhibitors.

Keywords: Diabetes, Bcl-2,Caspase, Apoptotic pathway, Protein-protein Interaction Network, *In-silico* Screening, ZINC,KEGG

INTRODUCTION

2.1 Background

Diabetes mellitus is a metabolic disorder that is characterized by the inability to maintain blood glucose levels under the physiological conditions. This chronic disease is affecting nearly 300 million individuals worldwide. Both types of diabetes (type 1 and type 2) are characterized by loss of functional pancreatic beta cell mass causing different degrees of insulin deficiency. Since insulin is the primary anabolic hormone that regulates blood glucose level, type 1 diabetics require a continuous supply of insulin for survival.

Glucose is one of the body's main sources of energy. In normal physiology, the body maintains blood glucose levels within a narrow range (80-120mg/dl). Blood glucose is balanced between endogenous appearance from the liver (through glycogenolysis and gluconeogenesis) and kidneys, exogenous appearance from the intestines (following a meal), and utilization of glucose by all tissues. Two gross metabolic conditions exist. When fasting, the body relies primarily on glucose stored in the form of glycogen and fatty acids stored in the form of triglycerides to fuel its metabolic needs. After a meal, glucose absorbed from the gut is used to replenish glycogen and fat stores diminished while fasting.

The body regulates the processes that control the production and storage of glucose by secreting the endocrine hormone, insulin, from the pancreatic B-cells. Insulin facilitates anabolic metabolism throughout the body. An increase in insulin above basal concentrations (2-12 mU/l) will decrease the release of glucose from the liver and increase glucose uptake into insulin-receptive tissues. This has the net effect of decreasing endogenous blood glucose appearance. There are many substances in the body that promote and inhibit insulin secretion, refining the detail to which the B-cells react to changes in the body's metabolic state. Glucose is by far the dominant stimulus for insulin secretion, establishing a direct relationship between insulin secretion and the blood glucose level in the body. When glucose concentrations increase, insulin concentrations will increase as well – a classical negative feedback system that keeps glycemia within very narrow range.

In diabetes, there is an uncoupling of blood glucose levels and the concentration of insulin that prevents the proper regulation of glycemia. Instead of a narrow glycemic range, blood glucose deviations can extend from hypoglycemia (less than 60 mg/dl) into hyperglycemia (fasting blood glucose greater than 126 mg/dl, postprandial blood glucose greater than 200 mg/dl).

This can be the result of a complete insulin deficiency, which is classified as insulin-dependent diabetes mellitus (type 1diabetes). However, the predominant form of diabetes is non-insulin-dependent diabetes mellitus (type 2 diabetes). Those afflicted with type 2 diabetes are commonly overweight with a

sedentary lifestyle. An abnormally high resistance to insulin causes sustained hyperglycemia, especially following meals. A third class of diabetes, gestational diabetes, presents itself during pregnancy and is a health concern for the mother and the developing fetus. Diabetes continues to be the most significant risk factor for the development of blindness, kidney failure, and neuropathy

2.2 Objective of the study

In this work, we have studied protein-protein interaction network of two major regulators Bcl-2 family and Caspase family proteins in apoptotic pathway.
The main objectives of this study are:

- To identify the major interacted proteins with Bcl-2 and caspase family proteins.
- To study the physio-chemical parameters of these interacted proteins.
- To identify drug like compounds for the treatment of diabetes using bioinformatics tools.

REVIEW OF LITERATURE

3.1 Apoptosis

Apoptosis is a coordinated series of events for the programmed execution of cell death, and plays an important role in the maintenance of tissue homeostasis. Defects in apoptosis regulatory machinery are implicated in a variety of pathological states; inadequate apoptosis may contribute to oncogenesis, while excess apoptosis is the underlying cause for cell loss during HIV/AIDS, neurodegeneration, and diabetes mellitus.

3.2 Apoptotic pathway regulation

A host of physical, chemical and biological factors can trigger apoptotic death by activating complex yet tightly controlled intracellular signal transduction pathways The extrinsic pathway is activated upon ligation of the cell surface death receptor(s), which in turn activates downstream effector mechanisms orchestrated by the caspase family of cysteine proteases (Green, 2005). The prototype example of death signaling via the extrinsic pathway is the Fas(CD95) death receptor, which brings about the assembly of the death inducing signaling complex(DISC), a multi-protein complex comprising of the cytoplasmic aspect of the Fas(CD95) receptor, the adaptor protein FADD (Fas-associated death domain containing protein), and pro-caspase 8. Assembly of this complex is the initiating signal for the processing of pro-caspase 8 to its active form. Depending upon the level of caspase 8 activity, the ensuing signal can either directly activate downstream caspase cascade or involve the intermediacy of the mitochondrial death pathway for efficient execution (Scaffidi *et al.,* 1998). The intrinsic or mitochondria dependent pathway characteristically involves cross-talk between caspases, pro-apoptotic members of the Bcl-2 family (e.g. Bax and Bad) and death amplification factors, such as cytochrome *c* and apoptosis-inducing factor, released from the mitochondria (Green, 2005). In contrast, anti-apoptotic members of the Bcl-2 family, such as Bcl-2 and Bcl-xL, blunt intrinsic death signaling by blocking the recruitment of pro-apoptotic members to the mitochondria (Green, 2005). Thus, there exists a tight intracellular balance between these two function ally contrasting Bcl-2 family members, and the fate of the cell is determined by the tilt in the ratio towards one or the other. Additionally, the phosphorylation status of some of the Bcl-2 members could also determine their distinct role(s) during cell death or survival signaling; phosphorylation status of Bad is critical in maintaining cell survival, as in this conformation Bad is anchored to 14-3-3 and unable to translocate to the mitochondria. Upon initiation of an apoptotic signal or inactivation of the kinase, unphosphorylated Bad is no longer able to bind 14-3-3, is free to translocate to the mitochondria and trigger mitochondrial outer membrane permeablization (Xiao & Singh, 2006; Zhou *et al.,* 2002).

Interestingly, the kinase involved in Bad phosphorylation is a critical cell survival serine threonine kinase downstream of PI3K and regulated by the protein phosphatase PTEN (Sun *et al.,* 1999). In

addition to its role in promoting tumorigenesis, Akt/PKB also is a critical regulator of pancreatic β cell mass and function and carbohydrate metabolism via insulin dependent uptake of glucose. Upregulation of Akt activity strongly reinforces survival pathways and blunts death signaling in a variety of cell types, and its downregulation renders the intracellular milieu permissive for death execution (Datta *et al.,* 1997). The latter has been documented in the dysfunction and loss of insulin secreting β cells associated with diabetes mellitus.

Indeed, increased rate of apoptosis has been linked to the pathophysiology of diabetes mellitus, and the PI3K/Akt pathway is critically involved in regulating cell fate decision.

Insulin resistance and deficiency are characteristic hallmarks of Type 2 diabetes mellitus. However, in addition, deposition of pancreatic amyloid consisting of islet amyloid polypeptide (IAPP) is also a common finding. The relative contribution of insulin insensitivity and deficiency in the pathogenesis of Type 2 diabetes has long been a subject of debate. Insulin resistance precedes insulin deficiency and is thus suspected to play a causative role (Reaven, 1988). However, defects in insulin sensitivity alone are inadequate for diabetic hyperglycemia; in separate cases (i.e., obesity and genetic insulin receptor defects), compensatory insulin production by pancreatic B-cells is capable of maintaining normaglycemia. It is believed that the disease can arise due to failure of β cell insulin preproduction to counteract the progressive insensitivity to insulin

.

Insulin secretion in response to a nutrient load presents a biphasic feature. The first phase of secretion, due to the release of preformed insulin, is rapid and relatively intensive and is followed by a longer and less intense second phase resulting from the synthesis of new insulin. Insulin deficiency can be a consequence of compromised insulin secretion per β cell and/or reduction in total β -cell mass. Although defects in both phases of insulin secretion are noted in glucose intolerance/diabetes (definitions available onwww.diabetes.org), there is no solid evidence that they play a causative role in diabetes (Masiello, 2006). On the other hand, cross-sectional studies of sample size with appropriate control subjects have presented compelling data supporting a role for B-cell loss in the pathogenesis of Type 2 diabetes (Butler *et al.,* 2003, 2004). The important findings include: (i) β-cell mass can increase during insulin resistance, which appears to represent a compensatory state; (ii) β-cell loss is present in diabetes; (iii) β-cell deficit correlates with the level of impairment in glucose tolerance, from around 40% in glucose intolerance to approximately 60% in Type 2 diabetes patients; (iv) β-cell death can directly lead to insulin deficiency. It appears that loss of 60% or more B-cells elicits manifestations of diabetes, particularly in the presence of insulin resistance (Butler *et al., 2003, 2004*).

There is experimental evidence that β-cell volume is reduced in patients with Type 2 diabetes, and recent reports comparing pancreatic tissues from Type 2 diabetic patients and non-diabetic subjects showed significantly higher rate of apoptosis in diabetic islets as opposed to the non-diabetic counterparts (Butler *et al.,* 2003, 2004). In transgenic diabetic animal models that express IAPP, typical features of β-cell apoptosis were observed concomitantly with a reduction in B-cell mass. Similarly, elevated activities of caspases-3 and caspase -8 in B-cells from islets of Type 2 diabetic patients, which

could be inhibited by the anti-diabetic agent metformin, were reported in a separate study, thus reinforcing the link between apoptosis and β-cell loss during Type 2 diabetes (Marchetti *et al.,* 2004).

Apoptotic death of B-cells has been observed in response to diverse stimuli, such as glucose, cytokines, IAPP, and free fatty acids (FFA). In primary cultures of murine β-cells, glucose induces apoptosis in a dose dependent manner, starting at 11mM (Efanova *et al.,* 1998). A decrease in anti-apoptotic Bcl-xL and reciprocal increases in pro-apoptotic proteins, Bad, Bid and Bik, have been noted (Federici *et al.,* 2001). Cytokine evoked β-cell apoptosis follows a drop in mitochondrial transmembrane potential and is inhibited by the over-expression of Bcl-2 (Saldeen, 2000). In addition to engaging the intrinsic death pathway in β-cell loss, glucose- and cytokine-mediated apoptosis have also been linked to the activation of the extrinsic death pathway, such as glucose-induced increase in cell surface Fas(CD95) expression and glucose-mediated as well as cytokine-mediated activation of caspases-8 (Maedler *et al.,* 2001).

Direct contact of IAPP fibrils also induces β-cell apoptosis, which is dependent on protein biosynthesis. Treatment of B-cells with IAPP fibrils increases mRNA levels of p53 and arrests cells before the S phase. The mRNA level of p21WAF/CIP1 (a p53- and cell cycle associated protein) also fluctuates, increasing initially and then falling. By the time the latter occurs, cells progress through the cell cycle with a defect in DNA repair related to p21, eventually leading to apoptosis (Zhang *et al.,* 1999). IAPP fibrils can also elicit apoptosis through activating JNK and p38MAPK as well as caspase-3, which is virtually completely blocked by the respective pharmacological inhibitors (Rumora *et al.,* 2002).

Similarly, FFA triggers β-cell apoptosis in a caspase dependent manner and by down-regulating Bcl-2mRNA (Lupi *et al.,* 2002). In addition, FFA-induced apoptosis is linked to down regulation of Akt phosphorylation, which is reversed by expression of constitutively active Akt. Down regulation of Akt has also been observed as a central mechanism in β-cell apoptosis induced by dominant negative form of the hepatocyte nuclear factor-1α. In contrast, activation of Akt (i.e., by simvastatin in islet transplantation) inhibits the activation of apoptosis machinery such as Bad, cytochrome *c* release and caspase-9, which protects β-cell from apoptotic death and helps maintain β-cell mass (Contreras *et al.,* 2002).The protective role of active Akt is further corroborated by data demonstrating that, troglitozone, the thiazolidinedione compound in clinical use for the treatment of diabetes, blocks β-cell apoptosis in human islets by activating Akt signaling *(*Lin et *al.,* 2005). However, it should be pointed out that functional loss of Akt alone might not be adequate to trigger β-cell apoptosis, as Akt knockout increases apoptosis sensitivity but does not specifically lead to β-cell death. Interestingly, functional defect in Akt predisposes to β-cell apoptosis and development of diabetes in the presence of defective insulin signaling (Ueki *et al.,* 2006). These findings underscore the importance of functional Akt as somewhat of a rhesostat in controlling β-cell death and the consequential development of diabetes. Indeed, the involvement of Akt in β-cell physiology might go beyond apoptosis and include the regulation of insulin secretion, β-cell volume, neogenesis and replication (reviewed in Elghazi *et al., 2006*).

Type 1 diabetes remains a common disease, although it accounts for only a relative small proportion of the overall diabetes population. Its onset occurs early in life due to rapid loss of pancreatic β-cells by cell-mediated autoimmunity; pathological infiltration of islets with lymphocytes occurs prior to clinical manifestations. Both helper (CD4+) and cytotoxic (CD8+) T cells are actively involved, and are under the influence of MHC loci as well as non-MHC determinants. Indeed, the link between hyperactive immune response and Type 1 diabetes is so striking that T cell depleted NOD (non-obese diabetic) mice fail to develop diabetes. In contrast, adoptive transfer of T cells from a diseased donor to a disease-free recipient concomitantly transfers the disease.

Infection-associated molecular mimicry has been a popular hypothesis for the development of autoimmunity in Type 1 diabetes. However, there is evidence that early developmental remodeling and/or homeostasis of β-cell mass involves β-cell apoptosis (Trudeau et al., 2000), which might trigger autoimmunity. In NOD mice, CD8+ cells are primed in the pancreatic lymph nodes, remain undetectable for up to 3–4 weeks after birth and progressively increase to maximal levels by 8 weeks of age. T cell priming is achieved by antigen-presenting cells that transport β-cell antigens specifically to pancreatic lymph nodes (Hoglund *et al.,* 1999). Although the relationship between β-cell apoptosis and auto-immunity remains to be fully established, there is emerging evidence that T cell-induced apoptosis is a dominant effector mechanism in Type 1 diabetes. In this regard, pancreatic β-cells derived from newly diagnosed patients with Type 1 diabetes were found to have increased cell surface expression of Fas(CD95) as compared to β-cells from healthy subjects that did not constitutively express detectable Fas(CD95). The apoptotic signal is then delivered via the Fas(CD95)L (Fas-ligand) expressed on infiltrating T lymphocytes (Stassi *et al.,* 1997). Furthermore, expression of dominant-negative Fas(CD95) or neutralizing antibodies to Fas(CD95)L significantly blocks apoptosis, maintains adequate β-cell function, blocks adoptive transfer of diabetes by primed T cells, and retards the course of diabetes development (Allison *et al.,* 2005). In contrast, transgenic expression of Fas(CD95) in β-cells accelerates the course of the disease (*Petrovsky et al., 2002*). A similar mechanism has been demonstrated for cytokine-induced β-cell death (Riachy *et al.,* 2006), thus highlighting the role of the extrinsic apoptotic pathway. Other studies provide evidence to implicate mitochondria as well in β-cell loss leading to Type 1 diabetes (Chang et al., 2004).

Similar to Type 2 diabetes, inhibition of Akt is implicated in Type 1 diabetes (Storling *et al.,* 2005) and activation of Akt (i.e., by insulin-like growth factor) delays the onset of the diabetes (Chen *et al.,* 2004). Recent findings have also implicated the transcription factor NF-κB in B-cell apoptosis more so in Type 1 than Type 2 diabetes (Cnop *et al.,* 2005). However, there still is a fair amount of controversy on whether NF-κB is upregulated or downregulated in β-cell apoptosis in diabetes.

Caspase activation and regulation is carried out in three steps:

1. **Initiator Caspase**

- Initiator caspases include caspase-2, caspase-8 and caspase-9 and caspase-10 which generally present in inert form in cytosol and require homodimerization for the activation.
- Activation of caspase-9, 8 and 2 depends upon apoptotic protease-activating factor 1 (Apaf-1) - apoptosome, death-inducing signaling complex (DISC), and PIDDosome, respectively.
- Caspase-8 activation is mediated by binding of Fas-L ligand to the transmembrane death receptors, Fas which belongs to the family of Tumor necrosis factor. Fas bind to the Fas associated death domain (FADD). This binding allows the aggregation of FADD and emergence of death effector domain (DEDs). These emerged DEDs induce oligomerization of procaspase-8 by interacting to their prodmain DEDs and allow formation of complex known as death inducing signal complex. In DISC, two linear subunits of procaspase-8 interact which leads to the auto-activation of caspase-8.
- Procaspase-8 can be activated vigorously or mildly which may lead to the activation of downstream procaspase-3 or mitochondrion-mediated pathway by truncated Bid into active form, tbid respectively. This tbid triggers the release of cytochrome c, apoptosis-inducing factor (AIF) and other molecules from mitochondria, and apoptosis induces.
- Caspase-9 activation is mediated by the intrinsic factor Bcl-2 which remains in association with Apaf-1 in normal condition. But during cell damage, it gets deattached from Apaf-1 and allows the complex formation called Apoptosome which includes oligomerized apoptotic protease factor-1 (Apaf-1), cytochrome c, the cofactor dATP/ATP and procaspase-9. Interaction between CARDs of N-terminal of Apaf-1 and the prodomain of procaspase-9 activates caspase-9 which in turn, activates procaspase-3 and 7. This activated caspase-3 then activates procaspase-9 and form a positive feedback loop (Fan *et al.,* 2005).

2. **Executioner Caspase**
- Executioner caspase include caspase-3, caspase-6 and caspase-7.
- These caspase present in the cytosol as inactive dimers and activated by limited proteolysis within their interdomain linker segment by the initiator caspase.

3. **Terminator or Cleaning Caspase**
- Terminator Caspase includes caspase-1, 4, 5, 10, 11 and 12. Activation of these caspsases leads to the processing of pro-inflammatory cytokines, thus, has specialized role in inflammation (Cullen *et al.,* 2009).

3.3 Heat Shock Proteins and Apoptosis

Heat shock proteins help in maintaining cellular homeostasis and cell survival. Hsps act as cytoprotective protein thus inhibit the apoptosis response. Major members of Hsps group prevent apoptosis by inhibiting caspase activation. Oligomerization of Apaf complex is inhibited by Hsp90 whereas Hsp27 and Hsp70 participates by inhibiting signaling pathway from surface inhibitors such as TNF-α and Fas receptors. There are various exciting examples which supports Hsps role in apoptosis inhibition and maintaining cellular homeostasis. For e.g. Inhibition of JNK activation by Hsp70 which is required for apoptosis response thus Hsp70 blocks JNK signaling. Another example is delimitation of bleb formation by Hsp70 associated with F-actin ring which helps in bleb formation by actin myosin formation (Sreedhar *et al.,* 2004; Beere, L. 2004).

3.4 Drug Designing

Drug designing is an important aspect in medical science. It is inventive process of finding medication for a disease. A drug can be designed using traditional drug designing method which is a random screening process and is very complex. It takes around 10-12 years and a lot of money to bring a drug in the market. An alternative approach to traditional drug designing is computer aided drug designing that uses computational resources, biological information present in the online databases and software packages to find the drug (Ooms, F. 2000).

Computer aided drug designing can be done by two approaches:

1. Structure based drug designing
2. Ligand based drug designing

3.5 Structure based drug designing

SBDD method is protein structure centric. It means based on the protein structure i.e. its amino acid residues, active site pocket. This method relies on finding of ligand molecule best fitted in the binding pocket of target structure (Verlinde *et al.,* 1994; Mandal *et al.,* 2009).

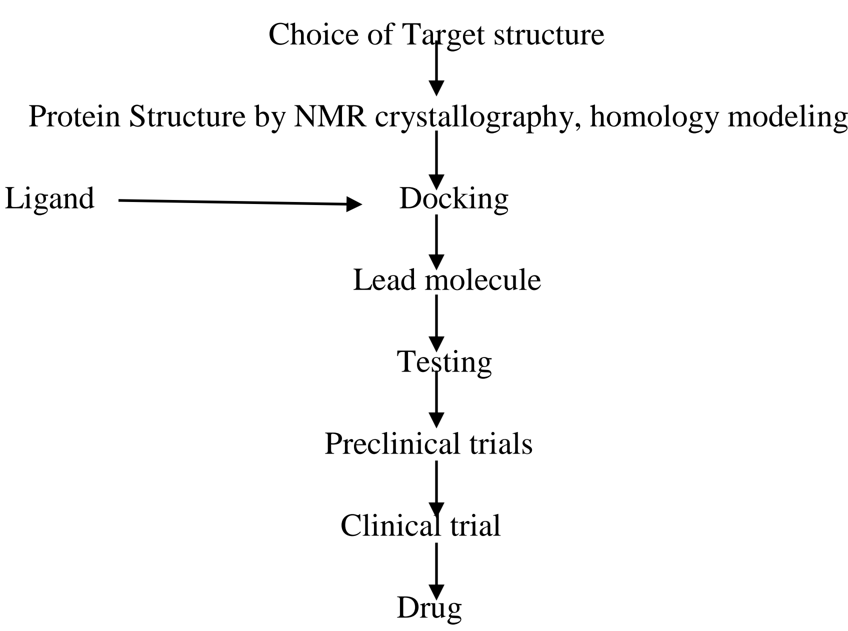


Figure 1: Protocol of Structure Based Drug Designing

3.6 Ligand based drug designing

Ligand based drug designing is based upon the already available drug or ligand molecule. This ligand molecule is required to develop a pharmachore model which describes the minimum structural feature required to fit in the binding pocket of target protein (Acharya *et al.,* 2011).

3.7 Molecular Docking

Molecular docking is an important tool in drug discovery. It is an approach in which a ligand molecule is allowed to fit in active site of the target protein with preferred orientation and optimum pose using computational methods. For this, active site knowledge is required. It increases the efficiency of docking. Sometimes, active site of protein is already known or it can be predicted by comparing the target structure with the family of protein sharing similar function. If it does not help, then there are various online server for predicting the binding pockets such as Q –site Finder, POOL server, POCKET and GRID (Meng *et al.,* 2011).

A simple docking process involves two steps:

1. Identifying the different conformation of ligand in the binding pocket of protein.
2. Then aligning these conformations on the basis of molecular docking score.

3.8 Docking Algorithms

There are various algorithms used by different software in molecular docking. These are Matching algorithm, incremental construction method, Monte carlo method, Stochastic methods, Genetic algorithm and molecular dynamics.

Stochastic method is a widely used docking algorithm and it modifies the ligand conformation to find out the conformational space. Genetic algorithm and Monte carlo method belong to the class of stochastic method (Reddy *et al.,* 2007).

Genetic Algorithm

This algorithm is based upon Darwin theory of evolution. Degree of freedom of ligand is considered as gene. These genes constitute chromosomes which defines the pose of ligand. Mutation and crossover are the two genetic operators in genetic algorithm and their effect results in new ligand structure. Software relies on GA are Autodock 4.2, Argus lab 4.0 and GOLD.

Monte Carlo Method

In this method, number of conformation obtained is pre-detemined. This method relies upon bond rotation, rigid-body translation or rotation for identifying the pose of ligand. Energy based selection criterion is used to test the obtained conformation. If the criterion limit is passed then conformation is saved and other conformation is predicted.

Scoring Function

Scoring function mains objective is to differentiate among different poses of a ligand in the binding pocket of receptor protein and then arrange them according to their binding affinities.

Type of Scoring Function

Force field based

This function depends upon the force field. It is based on energy landscape theory according to which native structure resides in deep and narrow well on potential energy surface. This function is a summation of all force filed energy such as H-bond, vander waal forces, electrostatic forces and protein atoms.

Empirical

This scoring function decomposes binding energy in component such as ionic interaction, H-bond interaction, binding entropy and hydrophobic interactions.

Knowledge based scoring

It is aimed to reproduce experimental structures. Major advantage of this scoring function is its computational simplicity that helps in efficient screening of large libraries.

Consensus scoring

As the name suggest, scores on the basis of a consensus obtained after combining different scores from different scoring schemes and if it follows same pattern in all schemes then the score is accepted.

3.9 Virtual screening

Virtual screening is a process of screening of billions of compounds against a receptor protein. It is a high throughput screening. HTS is based upon on massive trial and error to identify the pharmacological active compound. It is a lucrative technology to find out the hit molecules (Singh *et al.,* 2006; Reddy *et al.,* 2007).

Virtual Screening can be done by ways;

1. Structure based virtual screening
2. Ligand based virtual screening

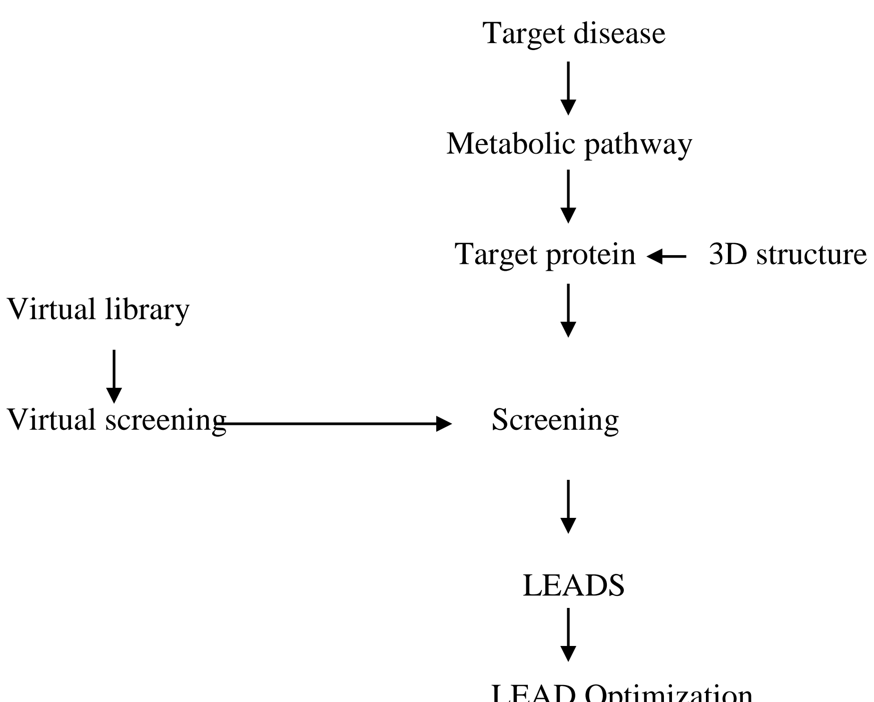


Figure 2: Target based virtual screening

METHODOLOGY

4.1 Protein-Protein interaction-

Protein- protein interaction is characterized by a number of parameters. Protein- protein interaction itself is defined by stoicheometry, its affinity, the kinetics of its information and dissociation and its particular features. Here, to study protein- protein interaction, STRING database is used. STRING (search tool for the retrieval of interacting genes/proteins) is a biological database and web resource of known and predicted protein – protein interactions.

The STRING database contains information from numerous sources, including experimental data, computational prediction methods and public text collectios. It is freely accessible and it is regularly updated. The latest version 9.0 contains information on about 5.2 millions proteins from 1133 species. STRING has been developed by a consortium of academic institutions including CPR, EMBL, KU, SIB, TUD and UZH.

Protein-protein interaction networks are an important ingredient for the system-level understanding of cellular processes. Such networks can be used for filtering and assessing functional genomics data and for providing an intuitive platform for annotating structural, functional and evolutionary properties of proteins. Exploring the predicted interaction networks can suggest new directions for future experimental research and provide cross-species predictions for efficient interaction mapping

The data is weighted and integrated and a confidence score is calculated for all protein interactions. Results of the various computational predictions can be inspected from different designated views. There are two modes of STRING: Protein-mode and COG-mode. Predicted interactions are propagated to proteins in other organisms for which interaction has been described by inference of orthology. A web interface is available to access the data and to give a fast overview of the proteins and their interactions. A plug-in for cytoscape to use STRING data is available. Another possibility to access data STRING is to use the application programming interface (API) by constructing a URL that contains the request.

Like many other database that store protein association knowledge STRING imports data from experimentally derived protein-protein interactions through literature curation. Furthermore, STRING also store computationally predicted interactions from: (i) text mining of scientific texts, (ii) interactions computed from genomic features, and (iii) interactions transferred from model organisms based on orthology.

All predicted or imported interactions are benchmarked against a common reference of functional partnership as annotated done by KEGG (Kyoto Encyclopedia of Genes and Genomes)

STRING imports protein association knowledge from databases of physical interaction and databases of curated biological pathway knowledge (MINT, HPRD, BIND, DIP, BioGRID, KEGG, Reactome, IntAct, EcoCyc, NCI-Nature Pathway Interaction Database, GO). Links are supplied to the originating data of the respective experimental repositories and database resources.

A large body of scientific texts (SGD, OMIM, FlyBase, PubMed) are parsed to search for statistically relevant co-occurrences of gene names.

Predicted data

- Neighborhood: Similar genomic context in different species suggest a similar function of the proteins.
- Fusion-fission events: Proteins that are fused in some genomes are very likely to be functionally linked (as in other genomes where the genes are not fused).
- Occurrence: Proteins that have a similar function or an occurrence in the same metabolic pathway must be expressed together and have similar phylogenetic profile.
- Co expression: Predicted association between genes based on observed patterns of simultaneous expression of genes.

4.2 SHORTEST PATHWAY ANALYSIS

VISANT-

It is an integrative visual analysis tool biological network and pathways. In this work it is used for shortest pathway analysis. Computationally predicted functional associations are produced in CAGT. Most of experimentally determined interaction data are collected from MIPS, BIND and HPRD etc. and updated periodically. Some large experimental data, such as chip-chip data for transcriptional regulatory networks ,is collected independently. In addition to the computational predicted interactions produced in lab, predictome database also collects experimentally determined ones by either directly importing from literature or by synchronizing with existing interaction databases such as Biogrid, MIPS, BIND, HPRD etc.

- Independently created interaction databases often use different naming systems for proteins/genes. For example, MINT uses a protein's UniProt id in protein-protein interaction for Homo sapiens which, however, is represented using gene's HGNC (HUGO Gene Nomenclature Committee) id in Biogrid.
- Genetic associations very often need to be compared to protein-protein interactions.
- High throughput gene expression results must often be mapped onto a protein-protein interaction network or pathway.

Because genes often encode multiple proteins, using gene id to represent both gene and protein is a natural choice; otherwise, uncertainty may arise when integrating interaction data. For example, for a gene with two splice variants, the variant to use when integrating interactions between the MINT (protein-based) and Biogrid databases (gene-based) will be ambiguous. The shortcoming of this solution is that certain splice variants may only share a subset of interactions. This dilemma can be overcome if alternative splicing knowledge is represented using a metanode, an undergoing process of this project.

4.3 Physio-chemical parameter computation

ProtParam

ProtParam computes various physico-chemical properties that can be deduced from a protein sequence. No additional information is required about the protein under consideration. The protein can either be specified as a Swiss-Prot/TrEMBL accession number or ID, or in form of a raw sequence. White space and numbers are ignored. If anyone provides the accession number of a Swiss-Prot/TrEMBL entry, he will be prompted with an intermediary page that allows him to select the portion of the sequence on which he would like to perform the analysis. The choice includes a selection of mature chains or peptides and domains from the Swiss-Prot feature table (which can be chosen by clicking on the positions), as well as the possibility to enter start and end position in two boxes. By default (i.e. if anyone leave the two boxes empty) the complete sequence will be analyzed.

The calculated parameters

The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Molecular weight and theoretical pI are calculated as in Compute pI/Mw. The amino acid and atomic compositions are self-explanatory. All the other parameters will be explained below.

Extinction coefficients

The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it.

It has been shown that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and Cystine (cysteine does not absorb appreciably at wavelengths >260 nm, while cystine does) at a

given wavelength, the extinction coefficient of the native protein in water can be computed using the following equation:

E(Prot) = Numb(Tyr)*Ext(Tyr) + Numb(Trp)*Ext(Trp) + Numb(Cystine)*Ext(Cystine)

Where (for proteins in water measured at 280 nm): Ext (Tyr) = 1490, Ext (Trp) = 5500, Ext(Cystine) = 125;

The absorbance (optical density) can be calculated using the following formula:

Absorb(Prot) = E(Prot) / Molecular weight

Two values are produced by ProtParam based on the above equations, both for proteins measured in water at 280 nm. The first one shows the computed value based on the assumption that all cysteine residues appear as half cystines (i.e. all pairs of Cys residues form cystines), and the second one assuming that no cysteine appears as half cystine (i.e. assuming all Cys residues are reduced). Experience shows that the computation is quite reliable for proteins containing Trp residues; however there may be more than 10% error for proteins without Trp residues. Cystine is the amino acid formed when of a pair of cysteine molecules are joined by a disulfide bond.

In vivo half-life

The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. ProtParam relies on the "N-end rule", which relates the half-life of a protein to the identity of its N-terminal residue; the prediction is given for 3 model organisms (human, yeast and E.coli). The N-end rule originated from the observations that the identity of the N-terminal residue of a protein plays an important role in determining its stability in vivo. The rule was established from experiments that explored the metabolic fate of artificial beta-galactosidase proteins with different N-terminal amino acids engineered by site-directed mutagenesis. The beta-gal proteins thus designed have strikingly different half-lives in vivo, from more than 100 hours to less than 2 minutes, depending on the nature of the amino acid at the amino terminus and on the experimental model (yeast in vivo; mammalian reticulocytes in vitro, Escherichia coli in vivo). In addition, it has been shown that in eukaryotes, the association of a destabilizing N-terminal residue and of an internal lysine targets the protein to ubiquitin-mediated proteolytic degradation. Note that the program gives an estimation of the protein half-life and is not applicable for N-terminally modified proteins.

The instability index provides an estimate of the stability of your protein in a test tube. Statistical analysis of 12 unstable and 32 stable proteins has revealed that there are certain dipeptides, the occurrence of which is significantly different in the unstable proteins compared with those in the stable ones. The authors of this method have assigned a weight value of instability to each of the 400 different dipeptides (DIWV). Using these weight values it is possible to compute an instability index (II) which is defined as:

$$II = (10/L) * \sum_{i=1}^{i=L-1} DIWV(x[i]x[i+1])$$

Where: L is the length of sequence

DIWV(x[i]x [i+1]) is the instability weight value for the dipeptide starting in position i.

A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.

GRAVY (Grand Average of Hydropathy)

The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence.

4.4 Find Site

This software is freely available to all academic users and not-for-profit institutions. Compound libraries for virtual screening libraries can be downloaded from http://cssb.biology.gatech.edu/findsite/.Currently, two libraries are available: KEGG compound and KEGG drug. If anyone would like to run findsite screen against a different library, can either prepare it by himself or send the compounds in SDF, MOL2 or SMI. Library will be formatted for use with FINDSITE and will be made available to the academic community from http://cssb.biology.gatech.edu/findsite/.

Scoring functions for ligand-based virtual screening

Molecular fingerprints are bit strings that represent the structural and chemical features of organic compounds (Daylight Theory Manual). Tanimoto coefficient is the most popular measure to quantify the similarity of two sets of bits (e.g. molecular fingerprints). Classical Tanimoto coefficient is defined as:

$$T = c \div (a+b+c)$$

where *a* is the count of bits on in the 1st string but not in the 2nd string, *b* is the count of bits on in the 2nd string but not in the 1st string, and *c* is the count of the bits on in both strings. In addition to the classical Tanimoto coefficient, the overlap between two molecular fingerprints can be measured by the average Tanimoto coefficient:

$$\mathbf{A= (T+T') \div 2}$$

Where T' is the Tanimoto coefficient calculated for bit positions set off rather than set on. In ligand-based virtual screening using findsite screen, if T or A scoring function is selected, the Tanimoto coefficient calculated between a library compound and the centroid of each ligand cluster reported by findsite is weighted by the cluster fraction:

$$\mathbf{S = \Sigma wi\ Xi}$$

Where S is the score assigned to the library compound, n is the number of chemically dissimilar ligand clusters within a given binding pocket, w_i is the fraction of ligand.

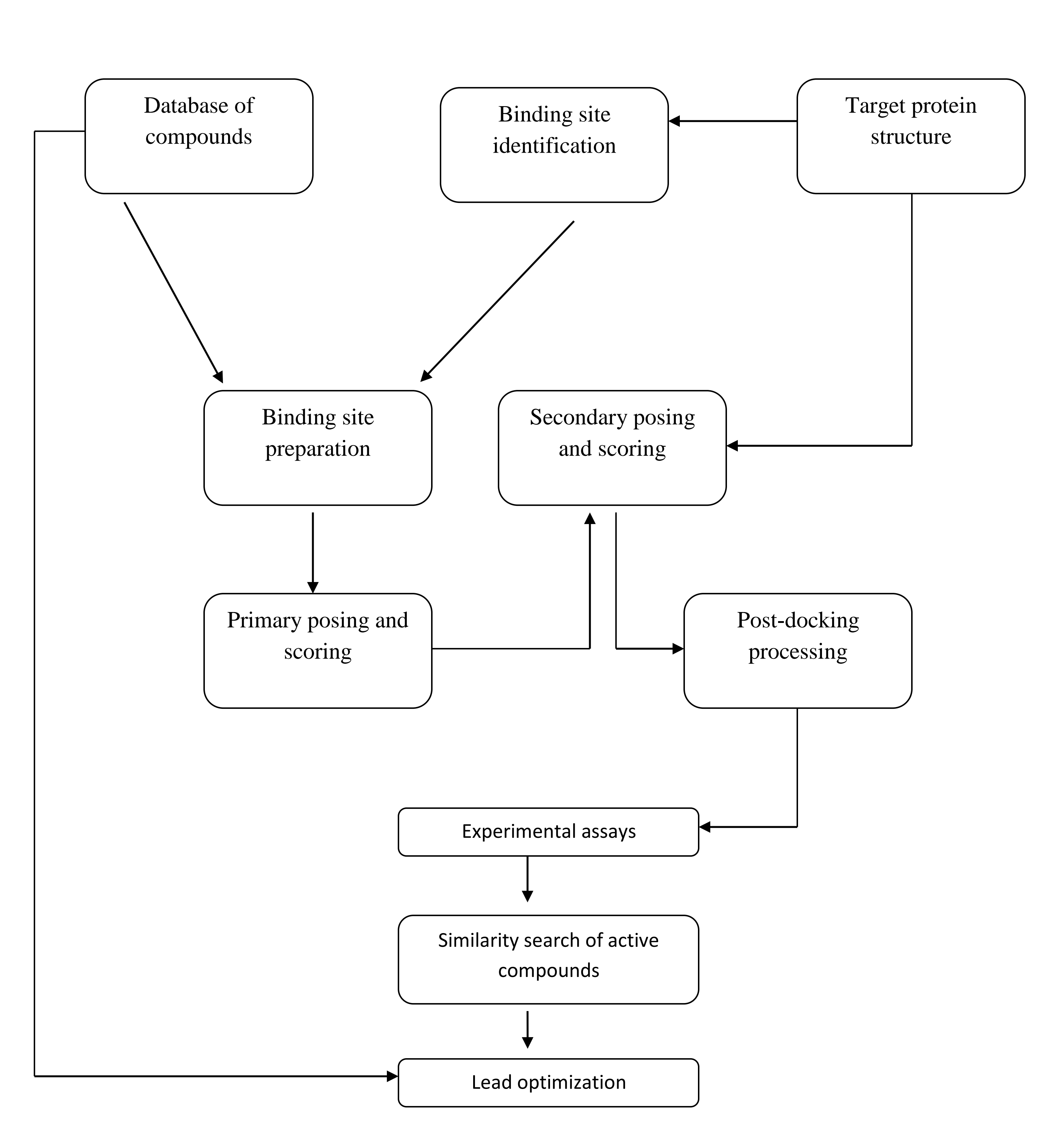


Fig.3: Flowchart of the virtual database screening

4.5 Argus lab

ArgusLab is a molecular modeling, graphics, and drug design program for Windows operating systems. It's getting a little dated by now, but remains surprisingly popular.
ArgusLab is freely licensed. No need to sign anything. It is is not allowed to redistribute ArgusLab from other websites or sources. However, may link to this website from own websites.
Before dock a molecule, it is need to define the atoms that make up the Ligand (drug, inhibitor, etc.) and the Binding Site on the protein where the drug binds. The crystal structure of a protein used can be freely accessed and downloaded from the Brookhaven Protein Databank (PDB) and represents an important result of a European collaborative antibacterial research project where several Slovenian researchers were involved.

- Download and open the PDB structure 2XA0.
- Create "Ligand and Binding Site groups".
- Dock the ligand into the defined biding site.
- Analyze and save the results.

4.6 Jmol

Mol is an open-source Java viewer for three-dimensional chemical structures, with features for chemicals, crystals, materials and biomolecules. Features include reading a variety of file types and output from quantum chemistry programs and animation of multi-frame files and computed normal modes from quantum programs.

RESULTS

5.1 Bcl-2 interaction network

In the study of protein-Protein interaction network of Bcl-2, the following result is obtained. It shows closely interacted proteins with Bcl-2. All the interacted proteins functional analysis is collected from String database itself along with the interacted proteins.

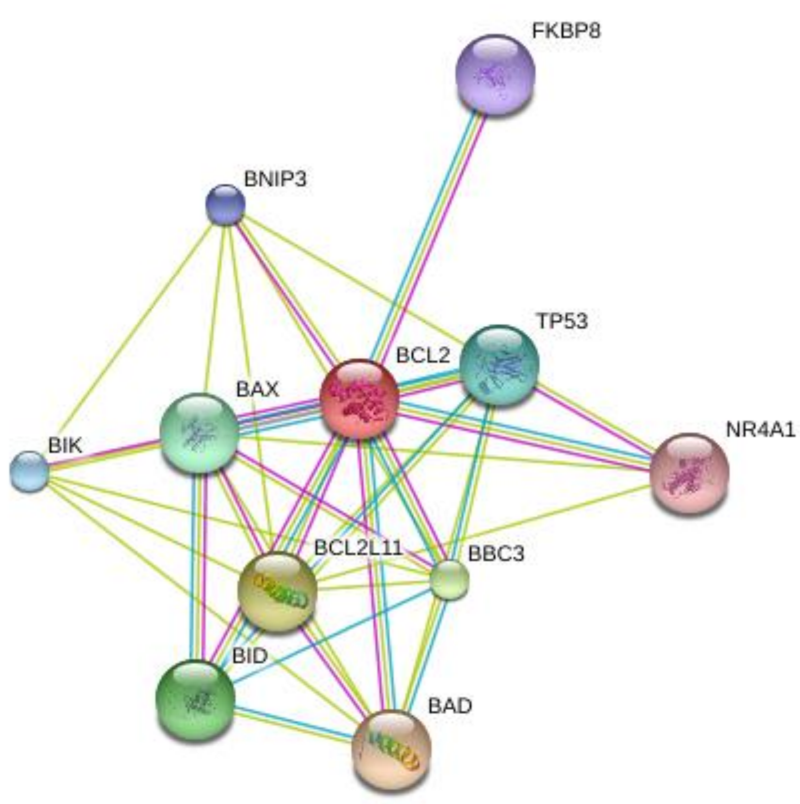


Fig. 4: Protein- Protein interaction network of Bcl-2

Bcl-2 suppresses apoptosis in a variety of cell systems including factor-dependent lymphohematopoietic and neural cells.It regulates cell death by controlling the mitochondrial membrane permeability. It appears to function in a feedback loop system with caspases and inhibits caspase activity either by preventing the release of cytochrome c from the mitochondria and/or by binding to the apoptosis-activating factor (APAF-1) .

BAD promotes cell death and successfully competes for the binding to Bcl-X (L), Bcl-2 and Bcl-W, thereby affecting the level of heterodimerization of these proteins with BAX. It can reverse the death repressor activity of Bcl-X(L), but not that of Bcl-2 (By similarity) and appears to act as a link between growth factor receptor signaling and the apoptotic pathways .

BCL2L11 induces apoptosis.The isoform BimL is more potent than isoform BimEL. Isoform Bim-alpha1, isoform Bim-alpha2 and isoform Bim-alpha3 induce apoptosis, although less potent than the isoforms BimEL, BimL and BimS. Isoform Bim-gamma induces apoptosis.

BBC 3 is essential mediator of p53-dependent and p53-independent apoptosis.

BID is the major proteolytic product p15 BID allows the release of cytochrome c (By similarity).The Isoform 1, isoform 2 and isoform 4 induce ICE-like proteases and apoptosis. Isoform 3 does not induce apoptosis. Counters the protective effect of Bcl-2 .

BAX accelerates programmed cell death by binding to, and antagonizing the apoptosis repressor BCL2 or its adenovirus homolog E1B 19k protein and Induces the release of cytochrome c, activation of CASP3, and thereby apoptosis .

Tp53 acts as a tumor suppressor in many tumor types; induces growth arrest or apoptosis depending on the physiological circumstances and cell type. It is involved in cell cycle regulation as a trans-activator that acts to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinases. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. Implicated in Notch signaling cross-over.

BIK accelerates programmed cell death and binds to the apoptosis repressors Bcl-X(L), BHRF1, Bcl-2 or its adenovirus homolog E1B 19k protein suppresses this death-promoting activity. Does not interact with BAX .

BCL2/adenovirus E1B 19kDa interacting protein 3 is the apoptosis-inducing protein that, which can overcome BCL2 suppression. It may play a role in repartitioning calcium between the two major intracellular calcium stores in association with BCL2 .

FKBP8 constitutively inactive PPiase, which becomes active when bound to calmodulin and calcium. It seems to act as a chaperone for BCL2, targets it to the mitochondria and modulates its phosphorylation state. The BCL2/FKBP8/calmodulin/calcium complex probably interferes with the binding of BCL2 to its targets. The active form of FKBP8 may therefore play a role in the regulation of apoptosis .

5.2 Caspase interaction network

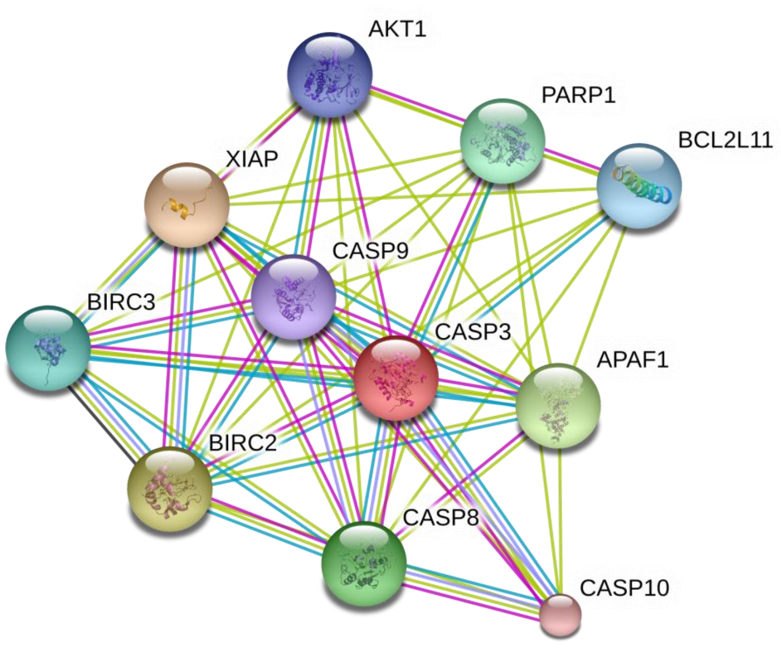


Fig. 5: Caspase protein-protein interaction network

Casp3 is involved in the activation cascade of caspases responsible for apoptosis execution. At the onset of apoptosis it proteolytically cleaves poly(ADP-ribose) polymerase (PARP) at a '216-Asp-|-Gly-217' bond. It cleaves and activates sterol regulatory element binding proteins (SREBPs) between the basic helix-loop- helix leucine zipper domain and the membrane attachment domain and cleaves and activates caspase-6, -7 and -9. Involved in the cleavage of huntingtin .

X-linked inhibitor of apoptosis (Apoptotic suppressor) has E3 ubiquitin-protein ligase activity. It mediates the proteasomal degradation of target proteins, such as caspase-3, SMAC or AIFM1. It is inhibitor of caspase-3, -7 and -9 and mediates activation of MAP3K7/TAK1, leading to the activation of NF-kappa-B

APAF1 mediates the cytochrome c-dependent autocatalytic activation of pro-caspase-9 (Apaf-3), leading to the activation of caspase-3 and apoptosis. This activation requires ATP. Isoform 6 is less effective in inducing apoptosis.

Caspase 8 is the most upstream protease of the activation cascade of caspases responsible for the TNFRSF6/FAS mediated and TNFRSF1A induced cell death. Binding to the adapter molecule FADD recruits it to either receptor. The resulting aggregate called death- inducing signaling complex (DISC) performs CASP8 proteolytic activation. The active dimeric enzyme is then liberated from the DISC and frees to activate downstream apoptotic proteases. Proteolytic fragments of the N-terminal propeptide (termed CAP3, CAP5 and CAP6) are likely retained in the DISC.

DNA fragmentation factor (45kDa) alpha polypeptide is an Inhibitor of the caspase-activated DNase (DFF40). Poly (ADP-ribose) polymerase 1 is involved in the base excision repair (BER) pathway, by catalyzing the poly(ADP-ribosyl)ation of a limited number of acceptor proteins involved in chromatin architecture and in DNA metabolism. This modification follows DNA damages and appears as an obligatory step in a detection/signaling pathway leading to the reparation of DNA strand breaks and mediates the poly(ADP- ribosyl)ation of APLF and CHFR. It Positively regulates the transcription of MTUS1 and negatively regulates the transcription of MTUS2/TIP150 .

Baculoviral IAP repeat-containing 2 is an apoptotic suppressor. The BIR motifs region interacts with TNF receptor associated factors 1 and 2 (TRAF1 and TRAF2) to form an heteromeric complex, which is then recruited to the tumor necrosis factor receptor 2 (TNFR2) .

Fas (TNFRSF6)-associated via death domain is also an apoptotic adaptor molecule that recruits caspase-8 or caspase-10 to the activated Fas (CD95) or TNFR-1 receptors. The resulting aggregate called the death-inducing signaling complex (DISC) performs caspase-8 proteolytic activation and active caspase-8 initiates the subsequent cascade of caspases mediating apoptosis.

Granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1) enzyme is necessary for target cell lysis in cell- mediated immune responses. It cleaves after Asp and seems to be linked to an activation cascade of caspases (aspartate-specific cysteine proteases) responsible for apoptosis execution. It cleaves caspase-3, -7, -9 and 10 to give rise to active enzymes mediating apoptosis.

Akt-v-akt murine thymoma viral oncogene homolog 1 is a general protein kinase capable of phosphorylating several known proteins and phosphorylates TBC1D4. Signals downstream of phosphatidylinositol 3-kinase (PI(3)K) to mediate the effects of various growth factors such as platelet-derived growth factor (PDGF), epidermal growth factor (EGF), insulin and insulin-like growth factor I (IGF-I). It plays a role in glucose transport by mediating insulin-induced translocation of the GLUT4 glucose transporter to the cell surface. It mediates the anti apoptotic effects of IGF-I and insulin-stimulated protein.

Casp9 (apoptosis-related cysteine peptidase) is involved in the activation cascade of caspases responsible for apoptosis execution. Its binding of caspase-9 to Apaf- 1 leads to activation of the protease which then cleaves and activates caspase-3 and proteolytically cleaves poly(ADP-ribose) polymerase (PARP).

Parp1-poly (ADP-ribose) polymerase 1 is also involved in the base excision repair (BER) pathway, by catalyzing the poly(ADP-ribosyl)ation of a limited number of acceptor proteins involved in chromatin architecture and in DNA metabolism. This modification follows DNA damages and appears as an obligatory step in a detection/signaling pathway leading to the reparation of DNA strand breaks. It mediates the poly (ADP- ribosyl)ation of APLF and CHFR and Positively regulates the transcription of MTUS1 and negatively regulates the transcription of MTUS2/TIP150.

Casp10 (apoptosis-related cysteine peptidase) is involved in the activation cascade of caspases responsible for apoptosis execution. It is recruited to both Fas- and TNFR-1 receptors in a FADD dependent manner. It may participate in the granzyme B apoptotic pathways and cleaves and activates caspase- 3, -4, -6, -7, -8, and -9. It also hydrolyzes the small- molecule substrates, Tyr-Val-Ala-Asp-|-AMC and Asp-Glu-Val-Asp.

Birc3-baculoviral IAP repeat-containing 3 is an apoptotic suppressor. The BIR motifs region interacts with TNF receptor associated factors 1 and 2 (TRAF1 and TRAF2) to form an heteromeric complex, which is then recruited to the tumor necrosis factor receptor 2 (TNFR2).

5.3 Shortest pathway analysis:

Protein –protein interactions (PPIs) play critical roles in the biological processes. Here I have identified some such shotest pathways where Bcl-2 family and other proteins are interacted in apoptotic pathway by protein-protein interaction network analysis.

Shortest path:

HSPD1-HSPA1A-BCL-2

HSPD1-RAF1-BCL-2

HSPD1-PRKACA-BCL-2

HSPD1-MYC-BCL-2

HSPD1-TP53-BCL-2

HSPD1-BAX-BCL-2

HSPD1-HSP90AA1-BCL-2

BCL-2-CASP3-HSPD1

CASP3-HSPD1

CASP3-UBC-BCL-10

CASP3-CASP9-BCL-10

CASP3-CASP8-BCL-10

CASB-PRKCQ-BCL-10

CASP3-ELAVL1-BCL-10

CASP3-BTRC-BCL-10

CASP3-BIRC4-BCL-10

CASP3-BIRC3-BCL-10

CASP3-BIRC2-BCL-10

CASP3-AKT1-BCL10

HSPD1-CASP3

HSPD1-CDK9-BCL-10

HSPD1-COPS5-BCL-10

HSPD1-UBC-BCL-10

HSPD1-1KBKG-BCL-10

BCL-10-BIRC3-CASP3

5.4 Physio-chemical parameter computation:

To, understand more about the interacted proteins, the different kinds of physiochemical parameters are calculated using protporam.

Protein	No. of amino acids	Molecular weight	Ex. Coefficient	Theoretical pI	Instability index	GRAVY	Half life (hrs)	Stability
BCL-2	239	26265.8	45045	6.75	51.63	-0.136	30	unstable
BAD	168	18391.8	30480	6.60	80.60	-1.152	30	unstable
BCL2L11	198	22170.9	20190	8.43	95.13	-0.713	30	unstable
BBC3	261	26498.2	11500	11.98	68.75	-0.680	30	unstable
BID	241	26836.3	21345	6.60	62.19	-0.401	30	unstable
BAX	218	24219.9	41605	7.74	48.10	-0.156	30	unstable
TP53	393	43653.1	36035	6.33	73.59	-0.756	30	unstable
BIK	160	18015.9	14105	4.21	62.38	-0.643	30	unstable
BNIP3	194	21540.9	24980	6.31	58.26	-0.696	30	unstable
FKBP8	413	44648.8	34420	4.78	58.18	-0.274	30	unstable
NR4N1	598	64463.3	45600	6.82	59.60	-0.242	30	unstable
CASP3	277	31607.8	26400	6.09	40.39	-0.457	30	unstable
XAIP	497	56684.8	73560	6.22	42.63	-0.541	30	unstable
BIRC2	618	69899.6	75175	6.27	52.83	-0.387	30	unstable
APAF1	1248	141839.9	191690	5.95	40.36	-0.268	30	unstable
CASP8	538	61835.5	34560	5.15	46.62	-0.496	30	unstable
PARP1	**1014**	**113083**	**120055**	**8.99**	**37.22**	**-0.536**	**30**	**stable**
BIRC3	604	68371.5	68185	5.71	49.74	-0.346	30	unstable
AKT1	**480**	**55686.4**	**65695**	**5.75**	**35.47**	**-0.575**	**30**	**stable**
CASP9	416	46280	30200	5.73	49.35	-0.237	30	unstable
CASP10	522	58993.5	36160	6.18	45.77	-0.330	30	unstable

Table 1-Physio-chemical parameters of interacted proteins

5.5 Study of Bcl-2:

In this work the following parameters of Bcl-2 are identified.

POCKET 1 3 1.0000

- pocket number (1)
- number of structure templates used to identify this pocket (3)
- Fraction of templates that share this pocket (1.0000).

TEMPLATE 1 1BXLA 181 148 0.556 0.581 0.421 2.233

TEMPLATE 2 2W3LA 141 136 0.971 0.956 0.410 0.967

TEMPLATE 3 2YXJA 137 134 0.637 0.627 0.403 1.064

- Template number(1),(2),(3)
- Template PDB-ID (1BXLA),(2W3LA),(2YXJA)
- Number of residues (181),(141),(137)
- Alignment length – number of residues aligned to the target by fr-TM-align (148),(136),(134)
- Global sequence identity to the target (0.556),(0.971),(0.637)
- TM-score to the target structure (0.581),(0.956),(0.627)
- Cα-RMSD [Å] of the aligned region (0.421),(0.410),(0.403),sequence identity calculated over the residues aligned by fr-TM-align (2.233),(0.967),(1.064)

LIGAND 1 1BXLA00 121 1723.930 1.3537 827.49

LIGAND 2 2W3LA00 42 575.209 6.9766 84.46

LIGAND 3 2YXJA00 56 813.266 9.8428 168.33

- ligand PDB-ID (1BXLA),(2W3LA)
- number of heavy atoms (121),(42),(56)
- molecular weight (1723.930),(575.209),(813.266)
- logP (1.3537),(6.9766),(9.8428)
- PSA (polar surface area) (827.49),(84.46),(168.33)

CENTER 58.736 -25.561 -30.705

Coordinates of the predicted pocket center (x, y, z).

5.6 VIRTUAL SCREENING OF BCL-2 FROM ZINC AND KEGG LIBRARY

ZINC00037358

N-(2, 6-dimethylphenyl)[pyrazole-3-carbonyl amino)cyclohexyl]carboxamide

ZINC20774134

Name-3(S)-3-benzyl-4-(2,6-dimethyl pyrimidine-4-carbonyl)piperazin-2-one

ZINC20600904

1-[(1, 5-dimethyl-1H pyrazole-3-yl]-3-(2-methyl phenyl)-3-pyrrolidinol]

ZINC04260722

3-benzyl-3,5,9-triazabicyclo[4.3.0]nonane-2,4-dione

ZINC12437966

(3S*, 4S*)-4-{ethyl [2-(1H-pyrazol-1-yl)ethyl]amino}-1-(4-methoxybenzoyl)-3-pyrrolidinol

ZINC20535531

N-3-[3, 5-dimethyl-1H-pyrazol-1-yl]-N-(1-methyl-5-oxo-3-pyrrolidinyl)benzamide

Fig.6-Virtually screened zinc compounds

C01952

Polyarginine

C04303

N-Benzoyl-D-arginine-4-nitroanilide; Bz-Arg-p-nitroanilide

C00306

Bradykinin

C12267

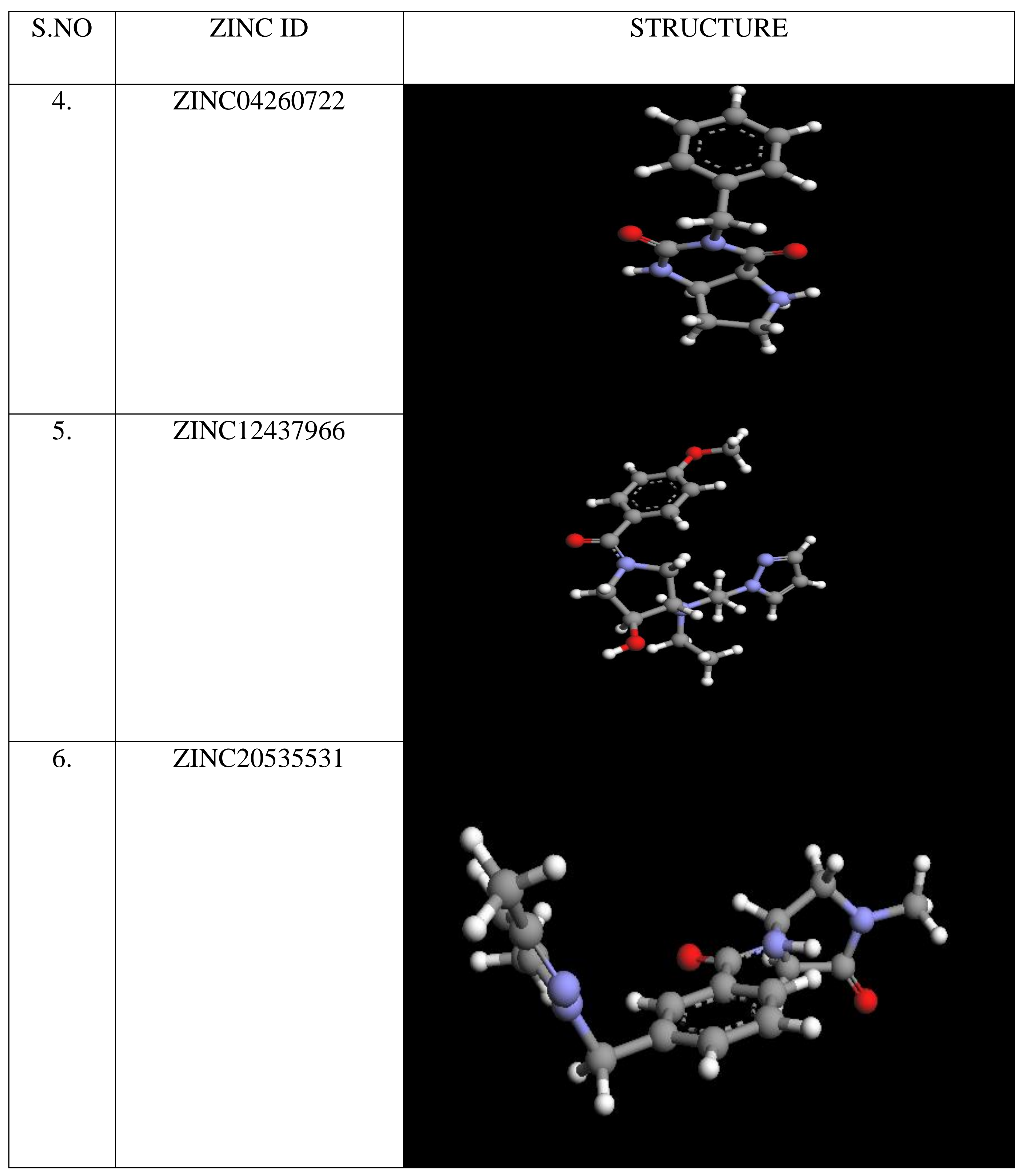
Bacillomycin D

C12270

NN-Acetyl-aspartyl-glutamate: NAAG; Isospaglumic acid

Fig.7-Virtually screened KEGG compounds

S.NO	ZINC ID	STRUCTURE
1.	ZINC00037358	
2.	ZINC20774134	
3.	ZINC20600904	



S.NO	ZINC ID	STRUCTURE
4.	ZINC04260722	
5.	ZINC12437966	
6.	ZINC20535531	

Table2a-3D view of ZINC compounds

S.NO	KEGG ID	STRUCTURE
1.	C01952	
2.	C04303	
3.	C00306	

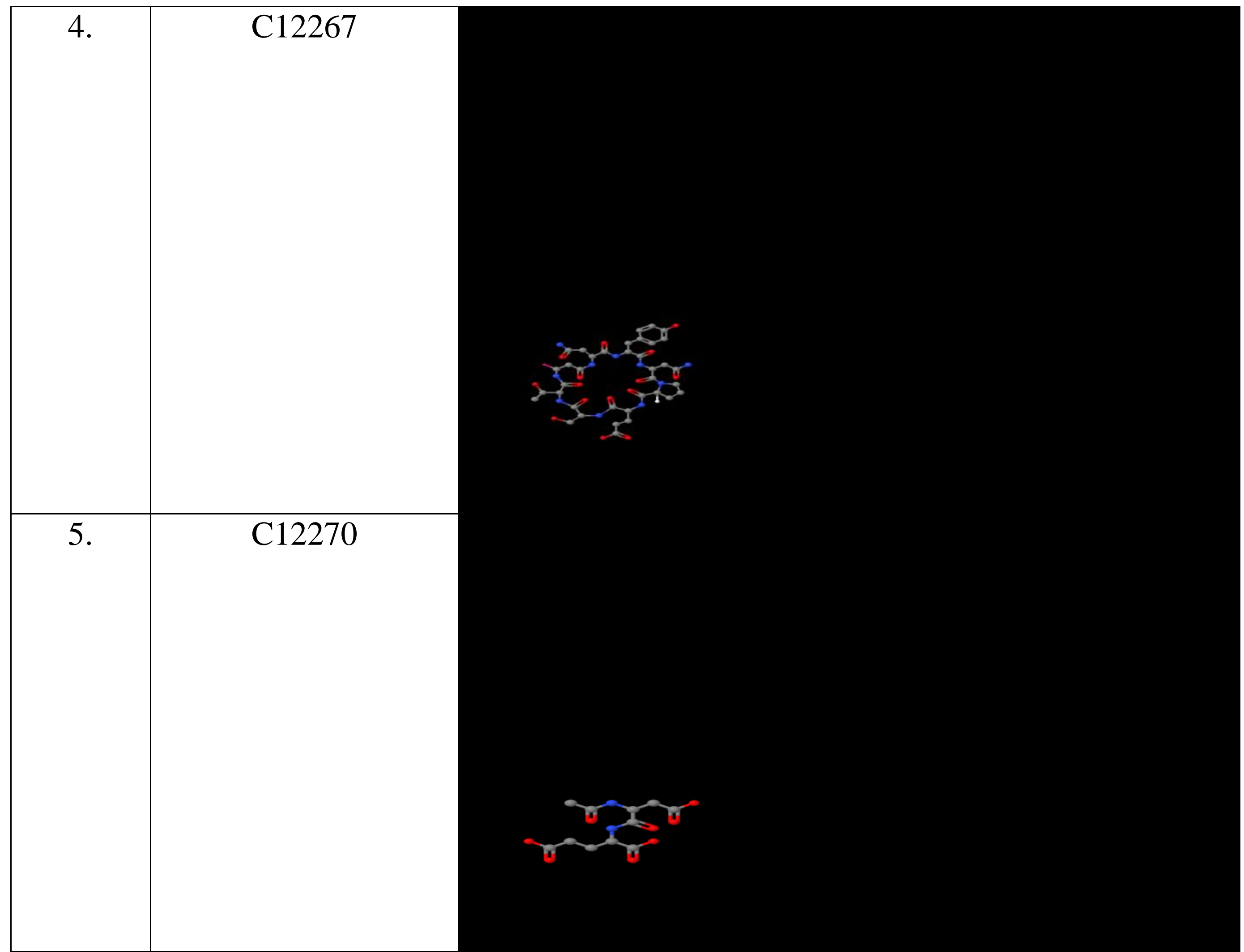
4.	C12267	
5.	C12270	

Table2b-3D view of KEGG compounds

5.7 DOCKING OF ZINC AND KEGG COMPOUNDS-

The docking score of each screened compound from both ZINC and KEGG library are tabulated.

ZINC ID	DOCK SCORE	KEGG ID	DOCK SCORE
ZINC00037358	-12.60	C01952	-11.80
ZINC20774134	-12.52	C04303	-9.42
ZINC20600904	-10.54	C00306	-9.23
ZINC04260722	-10.25	C12267	-8.75
ZINC12437966	-8.40	C12270	-8.63
ZINC20535531	-7.90		

Table3: Docking score

DISCUSSION

The above results show that the virtual screening can be used to find out the drug like compounds. Among the virtual screening compounds poly- arginine plays important role in diabetes as an inhibitor. Stabilization of complexed insulin against enzymatic degradation by trypsin and α-chymotrypsin is observed especially for the high molecular weight PEGylated arginine-based derivative. Insulin release rates in simulated intestinal fluid are controlled by the length of PEG chains and the presence of arginine end-groups (Tsiourvas *et al.,* 2012).

From protein-protein interaction analysis different interacted proteins are obtained. After analyzing the physio-chemical properties of these proteins, two proteins AKT-1 and PARP-1 have instability index less than 40(table1). This classifies that these two proteins are stable. This may be a potential target to further research of apoptotic pathway.

In this work we carried out a virtual screening of new potential Bcl-2 inhibitors starting from an analysis of the structural features. Application of Lipinski's rule of five on virtually screened compounds describes molecular features important for a drug's pharmacokinetics in the human body, including their absorption, distribution, metabolism, and excretion (ADME). However, the rule cannot predict if a compound is pharmacologically active. In recent years, this rule was used in virtual high-throughput screenings as a filter to broadly discriminate compounds.

CONCLUSION

Computational biology offers a fruitful approach to study in biological systems owing to its high-throughput screening ability. In this study, protein- protein interaction study explores closely interacted proteins and their shortest pathway with mediators. These proteins bio-chemical analysis will help to further analysis of the apoptotic pathway in diabetes. In virtual screening, ZINC and KEGG library are used to find out inhibitors. This methodology has the advantage that it deals with commercially available compounds, which thus do not require to be synthesized. It reduces the time and cost of the drug development process.

In future, molecular dynamics and simulation study of both Bcl-2 and virtual screened compounds will be helpful to more understanding the thermodynamic properties .But, this whole process is totally *In silico* study of the protein and other compounds. So, the result will be depending on used software accuracy.

REFERENCES

Bastaki, S (2005). Review: Diabetes mellitus and its treatment. Int J Diabetes & Metabolism. *13,* 111-134

Baumeister, W; Walz, J; Zuhl, F; Seemuller, E (1998). The proteasome: Paradigm of a self-compartmentalizing protease. *Cell. 92,* 367–380.

Beere, LM (2004). The stress of dying: the role of heat shock proteins in the regulation of apoptosis. Journal of Cell Science. *117*, 2641-2651.

Belova, L; Sharma, S; Brickley, DR; Nicolarsen, JR; Patterson, C; Conzen, SD (2006). Ubiquitin–proteasome degradation of serum- and glucocorticoid-regulated kinase-1 (SGK-1) is mediated by the chaperone-dependent E3 ligase CHIP. Biochem. J. 400, 235–244.

Bento, CF; Fernandes, R; Ramalho, J; Marques, C; Shang, F; Taylor, A; Pereira, P (2010). The Chaperone-Dependent Ubiquitin Ligase CHIP Targets HIF-1a for Degradation in the Presence of Methylglyoxal. PLOS ONE. *11*, e15062.

Bernasconi, A; Segre, AM (2000). Ab Initio Methods for Protein Structure Prediction: A New Technique based on Ramachandran Plots. ERCIM News No. *43,* 13.

Biedasek, K; Andres, J; Mai, K; Adams, S; Spuler, S; Fielitz, J., Spranger, J (2011). Skeletal Muscle 11beta-HSD1 Controls Glucocorticoid- Induced Proteolysis and Expression of Ubiquitin E3 Ligases Atrogin-1 and MuRF-1. PLOS ONE. *6*, e16674.

Breitschopf, K; Bengal, E; Ziv, T; Admon, A; Ciechanover, A (1998). A novel site for ubiquitination: the N-terminal residue and not internal lysines of MyoD, is essential for conjugation and degradation of the protein. *EMBO J. 17,* 5964–5973.

Broome, CS; Kayani, AC; Palomero, J; Dillmann, WH; Jackson, MJ; Mestril, R; McArdle, A (2006). Effect of lifelong overexpression of HSP70 in skeletal muscle on age-related oxidative stress and adaptation after nondamaging contractile activity. *FASEB. 20,* E855–E860.

Dodd, SL; Hain, B; Senf, SM; Judge, AR (2009). Hsp27 inhibits IKKβ-induced NF-ƙB activity and skeletal muscle atrophy. FASEB. *23,* 3415-3423.

Fan, TJ; Han, LJ; Cong, RS; Liang, J (2005). Caspase Family Proteases and Apoptosis. Acta Biochimica et Biophysica Sinica. *37,* 719–727.

Fox, SP; Gasparini, G; Harris, AL (2001). Angiogenesis: pathological, prognostic, and growth-factor pathways and their link to trial design and anticancer drugs. *Lancet Oncol. 2*, 278–89.

Frith, M.C., Saunders, N.F.W., Kobe, B; Bailey, TL (2008). Discovering sequence motifs with arbitrary insertions and deletions. *PLOS Computational Biology*. 4, e1000071.

Gepts, W (1965). Pathologic anatomy of the pancreas in juvenile diabetes mellitus. Diabetes. *14,* 619–633.

Gysemans, CA; Pavlovic, D; Bouillon, R; Eizirik, DL; Mathieu, C (2001). Dual role of interferon-gamma signalling pathway in sensitivity of pancreatic beta cells to immune destruction. Diabetologia. *44,* 567-574.

Hu, FB (2011). Globalization of Diabetes: The role of diet, lifestyle and genes. Diabetes Care. *34,* 1249–1257.

Irwin, JJ; Shoichet, BK (2005). ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model. *45,* 177-182.

Jiang, W; Wang, S; Xiao, M; Lin, Y; Zhou, L; Lie, Q; Xiong, Y; Guan, KL; Zao, S (2011). Acetylation regulates gluconeogenesis by promoting PEPCK1 degradation via recruiting the UBR5 ubiquitin Mol Cell. *43,* 33-44.

Kim, SJ; DeStefano, M; Oh, WJ; Wu, C; Finlan, M; Liu, D; Su, B; Jacinto, E (2012). mTOR Complex 2 Regulates Proper Turnover of Insulin Receptor Substrate-1 via the Ubiquitin Ligase Subunit Fbw8. Mol Cell. *48,* 875-887.

King, H; Aubert, RE; Herman, WH (1998). Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections, *Diabetes Care*. 1414–1431.

Kota, SK; Meher, LK; Jammula, S; Kota, SK; Krishna, SVS; Modi, KD (2012). Aberrant angiogenesis: The gateway to diabetic complications. Indian journal of Endocrinology and Metabolism. *16*, 6.

Laskowski , RA; MacArthur, MW; Moss, DS; Thornton, JM (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.* 26, 283-291.

Laskowski, RA; Rullmannn, JA; MacArthur, MW; Kaptein, R; Thornton, JM (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*. 8, 477-486.

Letunic, I; Doerks, T; Bork, P (2012). SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research. *40,* D302–D305.

Li, F; Xie, P; Fan, Y; Zhang, H; Zheng, L; Gu, D; Patterson, C; Li, H (2009). C Terminus of Hsc70-interacting Protein Promotes Smooth Muscle Cell Proliferation and Survival through Ubiquitin-mediated Degradation of FoxO1. J Biol Chem. *30*, 20090–20098.

Lokireddy, S; Wijesoma, WI; Teng, S; Bonala, S; Gluckman, PD; McFarlane, C; Sharma, M; Kambadur R (2012). The Ubiquitin Ligase Mul1 Induces Mitophagy in Skeletal Muscle in Response to Muscle-Wasting Stimuli. Cell Metabolism. *16,* 613-624.

Mandrup-Poulsen , T (2001). Beta-cell apoptosis: stimuli and signaling. *Diabetes. 50,* S58–S63

Martin, A; Ochagavia, ME; Rabasa, LC; Miranda, J; Fernandez-de-Cossio, J; Bringas, R (2010). SOFTWARE Open Access BisoGenet: a new tool for gene network building, visualization and analysis. BMC Bioinformatics. *11,* 1-9.

Meng, X; Zang, H; Mezei, M; Cui, M (2011). Molecular Docking: A powerful approach for structure-based drug discovery. Current Computational Aided Drug Design. *2*, 146–157.

Morcus, F; Lamanna, C; Sikora, M; Izaguirre, J (2008). Cytoprophet: a *Cytoscape* plug-in for protein and domain interaction networks inference. *Systems biology. 24, 2265–2266.*

Oever, IAM; Raterman, HG; Nurmohamed, MT; Simsek, S (2010). Endothelial Dysfunction, Inflammation and Apoptosis in Diabetes Mellitus. Mediators of Inflammation. *2010*, 792393.

Ooms, F (2000). Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry. Current Medicinal Chemistry. *7,* 141-158.

Peters, JM; Harris, JR; Finley, D. (1998). Ubiquitin and the biology of the cell. Plenum Press, New York and London, pp. 1-472.

Pugh, CW; Ratcliffe, PJ (2003). Regulation of angiogenesis by hypoxia: Role of the HIF System. Nature Medicine. *9*, 6.

Reddy, AS; Pati, SP; Kumar, PP; Sastry, GN (2007). Virtual Screening in Drug Discovery – A Computational Perspective. Current Protein and Peptide Science. *8,* 331-353.

Saltiel, AR; Kahn, CR (2001). Insulin signalling and the regulation of glucose and lipid metabolism. Nature. *414*, 799–806.

Scheffner, M; Huibregtse ,JM; Vierstra, RD; Howley, PM (1993). The HPV-16 E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell*. 75, 495–505.

Schultz, J; Milpetz, F; Bork, P; Ponting CP (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. 95*, 5857–5864.

Shannon, P; Markiel, A; Ideker, T (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. *13,* 1498-2504.

Singh, N; Cheve, G; Ferguson, DM; McCurdy, CR (2006). A combined ligand-based and target-based drug design approach for G-protein coupled receptors: application to salvinorin A, a selective kappa opioid receptor agonist. J Comput Aided Mol Des. *20,* 471–493.

Sreedhar, AS; Csermely, P (2004). Heat shock proteins in the regulation of apoptosis: new strategies in tumor therapy A comprehensive review. Pharmacology & Therapeutics. 101, 227– 257.

Thompson, JD; Higgins, DG; Gibson, TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. *22,* 4673-80.

Thomsen, R; Christensen, MH (2006). MolDock: A New Technique for High-Accuracy Molecular Docking. *J. Med. Chem. 49,* 3315–3321.

Tong, W; Wei, Y; Murga, LF; Ondrechen, MJ; Williams, RJ (2009). Partial Order Optimum Likelihood (POOL): Maximum Likelihood Prediction of Protein Active Site Residues Using 3D Structure and Sequence Properties. PLoS Comp Biol, 2009. 5, e1000266.

Tytell, M; Hooper, PL 2001). Heat shock proteins: new keys to the development of cytoprotective therapies. *Expert Opin Ther Targets*. *5,* 267–287.

Vanhaesebroeck, B; Leevers, SJ; Ahmadi, K; Timms, J; Katso, R; Driscoll, PC; Woscholski, R; Parker, PJ; Waterfield, MD (2001). Synthesis and Function of 3-Phosphorylated Inositol Lipids. Annu. Rev. Biochem. *70*, 532-602.

Verlinde, CLMJ; Hol, WGJ (1994). Structure-based drug design: progress, results and challenges. Structure. 2, 577-587.

Xu, X; Sarikas, A; Dolios, S; Lefontant, PJ; Tsai, SC; Zhu, W; Nakajima, H; Nakajima, HO; Field, HJ; Wang, R; Pan, ZQ (2012). Identification of insulin receptor substrate 1's degradation determinants for signaling cullin-RING Ubiquitin E3 Ligase 7-mediated ubiquitination. The American Society for Biochemistry and Molecular Biology. *M112*, 405209.

APPENDIX

ZINC DATABASE**:** ZINC is a free database of commercially available compounds for virtual screening. ZINC contains over 21 million purchasable compounds in ready to dock, 3D formats. ZINC is provided by the Shoichet Laboratory in the department of the pharmaceutical chemistry at the University of California, San Francisco. ZINC is used by investigators (generally people with training as biologists or chemists) in pharmaceutical companies, biotech companies, and research universities.

KEGG DATABASE: KEGG (Kyoto Encyclopedia of Genes and Genomes) is a database resource that integrates genomic, chemical and systemic functional information. In particular, gene catalogs from completely sequenced genomes are linked to higher-level systemic functions of the cell, the organism and the ecosystem. In KEGG database KEGG compounds and KEGG drugs are available for virtual screening.

LIPINSKI’S RULE: This rule indicates the drug like behavior of a compound. Lipinski's rule states that, in general, an orally active drug has no more than one violation of the following criteria:

- The molecular weight is less than 500.
- Not more than 5 hydrogen bond donors.
- Not more than 10 hydrogen bond acceptors.
- An octanol-water partition coefficient (logP) is less than 5.