"MOLECULAR DOCKING AND VIRTUAL SCREENING TO FIND NOVEL LIGANDS FOR PTP1B, A DRUG TARGET FOR DIABETES TYPE 2"

DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY IN INFORMATION TECHNOLOGY

(**BIO-INFORMATICS**)

Under the Supervision of

Dr. B. S. Sanjeev

Asst. Prof.

IIIT-Allahabad

Submitted By

Surjeet Singh

M. Tech. IT (Bio-Informatics)

IIIT-Allahabad

MB200407

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY
ALLAHABAD, U.P., 211012

1.1.1 INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

1.1.1.1Allahabad

(Deemed University)

of India)

Date: _____

WE DO HEREBY RECOMMEND THAT THE THESIS WORK PREPARED UNDER OUR SUPERVISION BY **SURJEET SINGH** ENTITLED "**MOLECULAR DOCKING AND VIRTUAL SCREENING TO FIND NOVEL LIGANDS FOR PTP1B, A DRUG TARGET FOR DIABETES TYPE 2**" BE ACCEPTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF TECHNOLOGY IN INFORMATION TECHNOLOGY (BIOINFORMATICS) FOR EXAMINATION.

COUNTERSIGNED

THESIS ADVISOR

DEAN (ACADEMIC)

NDIAN INSTITUTE OF INFORMATION TECHNOLOGY

Allahabad

(Deemed University)

(A Centre of Excellence in Information Technology Established by

3

4

4.1.1.1.1

4.1.1.1.2 CERTIFICATE OF APPROVAL*

The foregoing thesis is hereby approved as a creditable study in the area of

Information Technology carried out and presented in a manner satisfactory to warrant its acceptance as a pre-requisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve the thesis only for the purpose for which it is submitted.

COMMITTEE ON _____

FINAL EXAMINATION _____

FOR EVALUATION

OF THE THESIS _____

*Only in case the recommendation is concurred in

Acknowledgement

I am highly grateful to the honorable Director, IIIT-Allahabad, **Prof. M. D. Tiwari**, for his ever helping attitude and encouraging us to excel in studies. Besides, he has been a source of inspiration during my entire period of M.Tech. at IIITA.

I am thankful to **Prof. G. C. Nandi**, Dean Academics, IIIT Allahabad for providing all the necessary requirements and for his moral support for this dissertation work as well during the whole course of M. Tech.

The most notable source of guidance was my advisor, **Dr. B.S. Sanjeev,** Assistant Professor. I owe his a great deal of thanks for taking me under his wing and allowing me to soak up some of his knowledge and insight.

It gives me great pleasure to express my deep sense of gratitude and my humble gratefulness to **Prof. Krishna Misra, Coordinator of Indo Russian Center for Biotechnology** for her simple way of teaching. I am grateful to her for initiating and encouraging me to work in the field of Bioinformatics.

I am also grateful to **Dr T. Lahiri**, **Dr. C.V.S. Siva Prasad** Assistant Professors, **Mr. Vikram Katju** and **Mr. Pritish Varadwaj** Faculty Associates, M. Tech Information Technology (Bioinformatics) programme, IIIT- Allahabad for their support and motivation; through out my research project work.

I am also thankful to my friend ***Professor. Jasjeet Bagla*** for extending valuable support to use super computing facility at *Harish Chandra Research Institute* , Allahabad.

I am also thankful to my friends ***Ms. Vandana Kumari, Ms Monika***, ***Mr. Rohita Sinha, Mr. Sailendra Nautiyal, Mr. Senthil Natesan*** *and* ***Mr. Yogesh Joshi*** for helping me to understand some programming problem.

I am also thankful to rest of the classmates for their friendliness during my work. I am also thankful to them for helping me in my project work and also some kind of discussion regarding my work which helps me to understand the concept regarding my work.

This acknowledgement will not complete until I pay my respectful homage to my family especially my parents, whose enthusiasm to see this work complete was as infectious as their inspiration.

Finally, I also acknowledge Ministry of Human Resource and Development (MHRD), Govt. of India for their financial support at IIIT- Allahabad.

Surjeet Singh

Abstract

Type 2 diabetes is characterized by an impaired insulin action, early insulin resistance, and a decline and ultimate failure of beta cell function. In addition, both fasting and post-meal glucose levels in the blood are high. Compared to controls, type II diabetics have reduced insulin-stimulated glucose disposal in all insulin sensitive tissues, particularly skeletal muscle, liver, and fat. Improving tissue sensitivity to insulin is a major clinical goal to help ameliorate not only abnormal glucose metabolism, but also some of the cardiovascular risk factors that accompany this syndrome. Protein-tyrosine phosphatases (PTPases) that function as negative regulators of the insulin signaling cascade have been identified as novel targets for the therapeutic enhancement of insulin action in insulin-resistant disease states. Recent studies have provided compelling evidence that one of the main functions of the intracellular enzyme PTPase 1B (PTP1B) is to suppress insulin action. Reducing PTP1B abundance not only enhances insulin sensitivity and improves glucose metabolism but also protects against obesity induced by high-fat feeding. Inhibition of PTP1B in insulin target tissues using pharmaceutical agents has shown enhanced insulin signaling and glucose tolerance in preclinical models. PTP1B inhibitors may eventually find an important clinical role as novel insulin sensitizers in the management of type 2 diabetes and the metabolic syndrome. This project aimed to find out the potent inhibitors of PTP1B by molecular docking and virtual screening of large database of more than 4.1 million ligands.

List of Abbreviations

ASO	Antisense Oligonucleotides
CAP	C-Cbl-Associated Protein
CD45	Protein Tyrosine Kinase (a type of)

DIO	Diet-Induced Obesity
DMS	Distributed Molecular Surface
ER	Endoplasmic Reticulum
GLUT4	Glucose Transporter-4
GS	Glycogen Synthase
GSK3	GlycogenSsynthase Kinase 3
IGF-IR	Insulin-like growth factor I Receptor
IRS	Insulin Receptor Substrate
JAK2	Janus kinase 2 (a protein tyrosine kinase)
KO	Knock Out
LAR	Receptor-like Transmembrane PTP
PDB	Protein Data Bank
PDGFR	Platelet-derived growth factor receptor
PDK1	Phosphoinositide-Dependent Kinase-1
PI3K	Phosphatidylinositol 3-Kinase
PKC	Protein kinase C (lambda/Zeta)
PTP	Protein Tyrosine Phosphatase
PTP1B	Protein Tyrosine Phosphatase 1-B
RMSD	Root Mean Square Deviation
RTK	Receptor Tyrosine Kinase
SHP2,	Protein phosphatase 2
TC-PTP	T cell protein tyrosine phosphatase
UCSF	University of California, San-Francisco
WHO	World Health Organization

TABLE OF CONTENTS

5

LIST OF FIGURES

6

LIST OF TABLES

7 Introduction

It is predicted that 300 million people worldwide will suffer from type 2 diabetes by the year 2025 [WHO-1999]. This sobering statistic makes the search for agents to intervene in type 2 diabetes ever more pressing. Insulin is a key regulator of intermediary metabolism, and defective cellular action of insulin results in insulin resistance, an

increasingly common condition afflicting the industrialized world [Kahn-1998]. Insulin resistance is present when insulin produces less than the expected biological response in various tissues that are the physiologic targets for the metabolic actions of insulin. Insulin resistance is a fundamental pathogenic flaw leading to the "metabolic syndrome," a condition which underlies the increased cardiovascular risk profile in the majority of overweight and obese patients and is often associated with impaired glucose tolerance or frank diabetes [Law, R. E.-1998]. The increased cardiovascular risk in this syndrome is due to its association with a panoply of abnormalities including endothelial dysfunction, an atherogenic lipid profile, hypertension and abnormal thrombolysis In the setting of insulin resistance, elevations in fasting and post-prandial glucose levels are secondarily determined by the secretory function of the pancreatic β-cells [Polonsky, K-1996]. Frank type 2 diabetes with overt hyperglycemia occurs when there has been a progressive failure of the pancreas to supply the increased insulin needs of the body. Maneuvers that enhance tissue insulin sensitivity, including increasing activity and promoting weight loss, can facilitate insulin stimulated glucose disposal and ameliorate the impaired glucose tolerance as well as many of the risk factors that accompany the metabolic syndrome. Since lifestyle modification has been difficult for patients to adopt for prolonged periods of time, there is a definite need for new pharmaceutical agents that can ameliorate insulin resistance.

Previous experience with these so-called "insulin sensitizers" has shown that by improving insulin sensitivity, they may be able to ameliorate several cardiovascular risk factors that accompany the metabolic syndrome, and also slow the progression of glucose intolerance to frank diabetes [Goldstein, B. J.-1999]. In efforts to develop pharmaceutical agents that enhance insulin sensitivity, specific components of the insulin action pathway have been targeted. In particular, interest has grown in regulatory events in insulin signaling that affect the tyrosine phosphorylation of components involved in the transmission of the insulin signal [Ostman, A-2001].

PTP1B, a protein-tyrosine phosphatase (PTPase) enzyme that has a key role in the negative regulation of protein tyrosine phosphorylation in the insulin action cascade [Byon, J. H.-1998]. PTP-1B has two proposed specific functions:

1.) It associates with the insulin receptor and dephosphorylates it, thereby disrupting the ability of the receptor to respond to insulin binding;
2.) It dephosphorylates JAK-2, an important mediator of the leptin signaling cascade.

For these reasons, PTP-1B has been an important target for synthetic inhibitors as potential treatment for diabetes. Inhibition of PTP1B might be expected to enhance insulin action in diabetes, as well as improving insulin resistance in patients with the metabolic syndrome.

The dominant technique for the identification of new lead compounds in drug discovery is the physical screening of large libraries of chemicals against a biological target (high throughput screening). An alternative approach, known as virtual screening, is to computationally screen large libraries of chemicals for compounds that complement targets of known structure and experimentally test those that are predicted to bind well.

In this study DOCK 5.4 and other accessory programs were used to screen and rank database of over 4.1 million ligands.

8 Literature Review

8.1 Overview of Insulin Signaling Pathway

Insulin is secreted from pancreatic β-cells in response to increasing glucose concentrations in the blood. The hormone binds to its receptor, a tetrameric complex composed of two α- and two β-subunits [Saltiel, AR,-2001]. Binding of the hormone to the extracellular α-subunits triggers a conformational change that activates the intrinsic tyrosine kinase activity of the intracellular β-subunit via autophosphorylation of specific tyrosine residues in the activation loop. Some phosphorylated residues (outside the activation loop) act as docking sites for IR substrates (IRSs), which in turn become phosphorylated by the receptor tyrosine kinase (RTK). The phosphorylated IRSs serve as adaptor proteins and recruit phosphatidylinositol 3-kinase (PI3K) via the regulatory subunit. PI3K then catalyzes the conversion of phosphatidylinositol to the 3,4-bis- and

3,4,5-trisphosphates that stimulate the activity of phosphoinositide-dependent kinase-1 (PDK1). The phosphoinositide-dependent kinases activate Akt, or PKB, via phosphorylation of a critical serine and threonine residue. The series of protein phosphorylations on the signaling molecules downstream of the insulin receptor culminates in the uptake of glucose into cells by the glucose transporter GLUT4. The mechanism by which GLUT4-containing vesicles become activated (downstream of PI3K). Akt and the atypical PKCζ/λ (also activated via PI3K phosphorylation) have been implicated in the process [Standaert, ML-2001]. Irrespective of its involvement in the activation of GLUT4 vesicles, Akt appears to participate in the pathway by phosphorylating glycogen synthase kinase 3 (GSK3) to promote glycogen synthesis via glycogen synthase (GS) [Moule, SK-1997]. GSK3 is constitutively active and phosphorylates GS to inactivate this enzyme, which is required for the incorporation of glucose (in the form of UDP-glucose) into glycogen. Phosphorylation of GSK3 by Akt inactivates the kinase and relieves its block on GS. In addition to the above pathway, a PI3K-independent pathway appears to be required for insulin-dependent glucose uptake into cells. This c-Cbl-associated protein (CAP)/Cbl-dependent pathway apparently provides a second signal that influences GLUT4 vesicle translocation via lipid rafts to effect glucose uptake [Bickel, PE-2002].

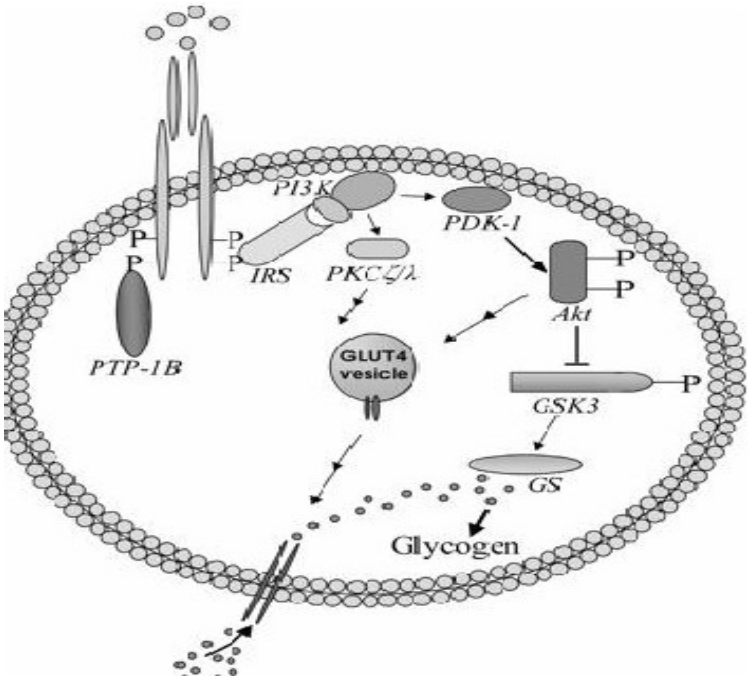

Figure 1 Insulin-signaling pathways [Ernest Asante-2003]

As substantial amount of knowledge has accrued regarding processes that initiate and propagate insulin signaling to influence glucose uptake; events that lead to signal termination, however, are not that well understood. An emerging hypothesis that is gaining acceptance involves the dephosphorylation of key tyrosine residues in the activation loop of the receptor. It has been postulated that the level of receptor activation is determined by the opposing actions of receptor phosphorylation vis-à-vis dephosphorylation. The role of PTPs in the deactivation of IR is therefore taking on much significance in insulin signaling. Thus inhibition of the IR phosphatase should provide an attractive approach for intervention in type 2 diabetes. The search for the enzyme that dephosphorylates the IR has implicated a number of PTPs**[**Ernest Asante-2003**].**

8.2 Insulin Signaling and Protein Tyrosine Phosphatases

Nonspecific inhibitor vanadate, which acts as an insulin-mimetic agent, implicated PTPs in insulin signaling [Fantus, IG-1994]. Because many of the signaling proteins downstream of the IR tend to be phosphorylated on serine and threonine residues, research efforts focused on the identification of the PTP(s) that dephosphorylated the IR. This is because PTPs are enzymes that remove phosphate groups from key tyrosine residues on signaling proteins. The enzymes are classified as PTPs on the basis of an invariant and catalytically essential cysteine residue that is part of a unique signature motif: (I/V)HCX$_5$R(S/T)[Barford, D-1998]. Enzymes in this family fall into two main groups, receptor or nonreceptor, depending on whether they possess or do not possess a transmembrane domain. Using gene knockout approaches, transgenic mice in which specific PTP genes have been overexpressed, and other biochemical approaches, several PTPs, including PTP1B, PTPα, PTPε, SHP2, and LAR, have been implicated as negative regulators of insulin signaling.

8.3 Insulin Signaling and Protein Tyrosine Phosphatases-1B (PTP1B)

The most convincing evidence that PTP1B is involved in the insulin-signaling pathway originates from the phenotype of the PTP1B KO [Klaman, LD-2000] mouse and, more recently, from results of PTP1B ASO [Zinker, BA-2002] treatments in diabetic rodents. The PTP1B KO mouse has generated a number of surprising results and has provided insights into a number of presumptive roles for the phosphatase in vivo. It was assumed that the disruption of the PTP1B gene in mice would result in either lethality or a significant susceptibility to tumor formation, since this phosphatase has been shown, at least in cell culture, to be involved in the attenuation of many growth factor receptor kinase-signaling pathways, including IGF-IR, PDGFR, EDGFR, and IR, to name a few [Roome, J-1988][lint, AJ-1999]. Neither of these possibilities was observed; the mice were viable and long-lived without a significant increase in tumor formation. The reason for this may be that results derived from cell culture studies may not accurately reflect the

function of PTP1B in vivo. It also seems possible that there may be a compensation for the PTP1B deficiency by other PTPs. Although we cannot completely rule out compensatory effects, the phenotype we have observed in the PTP1B KO mice.

8.4 Negative Regulator of Insulin Signaling: PTP1B

PTP1B deficiency in mice results in enhanced insulin sensitivity, as demonstrated by a significant reduction in fed glucose levels that is maintained with one-half the circulating insulin levels [Elchebly, M-1999]. Additionally, there is increased insulin-stimulated phosphorylation of the IR in muscle and liver and an improved glucose clearance in glucose and insulin tolerance tests. The loss of PTP1B potentiates insulin's activity, which suggest that PTP1B is a negative regulator of insulin signaling. This place PTP1B downstream of the IR, and presumably it functions to dephosphorylate and inactivate the IR. Alternatively, or in addition to its activity on the IR, PTP1B potentially can attenuate insulin signaling by dephosphorylating IRSs or possibly other phosphotyrosyl insulin-dependent signaling molecules yet to be identified. Although proof that PTP1B directly interacts with the IR in a cellular or in vivo context is not unequivocal, there is a significant amount of evidence to suggest that this is probably the case.

Catalytically inactive mutants of PTP1B can "trap" the activated IR in immunoprecipitation protocols. It has been suggested that an NH_2-terminal domain of PTP1B that includes tyrosines 152 and 153 is required for IR binding [Dadke, S-2000]. Additional evidence supporting a direct interaction between PTP1B and IR comes from kinetic and structural studies with the IR activation segment [Salmeen, A-2000]. A very interesting observation was on the order of dephosphorylation of a triphosphorylated peptide derived from the activation segment of the receptor by various PTPs. Ramachandran et al. [Ramachandran, C-1992] found that the receptor-type PTPs CD45 and LAR preferentially dephosphorylated the single phosphotyrosine residue, whereas TC-PTP, the most closely related PTP to PTP1B, displayed no phosphotyrosyl preference. In contrast, PTP1B showed a very strong preference for the tandem pTyr motif, suggesting that this feature

was a strong determinant for PTP1B substrate binding and specificity. The recent crystallization of the IR activation segment in complex with PTP1B has established a structural basis for this selectivity [Salmeen, A-2000]. It was observed that there are extensive interactions between the tandem pTyr residues and PTP1B such that pTyr 1162 is located within the catalytic site and pTyr-1163 is positioned in the adjacent secondary pTyr-binding site. From these data, it seems likely that PTP1B may directly dephosphorylate the activated IR.

An outstanding issue that remains to be resolved involves how the endoplasmic reticulum (ER)-localized PTP1B can actually interact with the IR that becomes activated on the plasma membrane. As a possible explanation, it has been reported that the ER can come into close proximity with the plasma membrane and that, under some conditions (i.e., phagocytosis), both membranes may even fuse together [Gagnon, E-2002]. Thus it seems possible that PTP1B and the IR can come within close proximity for dephosphorylation to take place. It is also possible that, once the activated IR is internalized into endosomes, it may be directed to specific locations on the ER, where it becomes dephosphorylated, as has been suggested for the dephosphorylation of the platelet-derived and epidermal growth factor receptors by PTP1B[Haj, FG-2002].

8.5 PTP1B & Insulin Resistance: Functional Relationship

Both the PTP1B KO mouse and the PTP1B ASO-treated diabetic animals display increased insulin sensitivity [Klaman, LD-2000]. Treatment of *ob/ob* and *db/db* mice with PTP1B-specific ASO reduced protein levels of the phosphatase in liver and fat by 50% and resulted in normalization of glucose levels in these preclinical insulin-resistant mouse models. Hence, a $\geq$50% reduction in PTP1B protein level by genetics or ASO is sufficient to cause insulin sensitization and alleviate insulin resistance. Currently, it is not clear whether the improvement in the insulin resistance observed in these models is a consequence of a correction in the dysregulation in the insulin receptor/PTP1B equilibrium or

is due to an overall enhancement in insulin action that results from a reduction in the levels of a negative regulator.

Recently, it was reported that reversible oxidation of PTP-1B may control the amount of functionally active PTP1B available in the cell. Goldstein and colleagues have shown that activation of IR results in the production of H_2O_2 and a concomitant transient oxidation and inactivation of PTP-1B. They have also demonstrated that, depending on the type of fat depot, there was a significant difference in the level of oxidized-inactive PTP1B, and they suggested that increased levels of active PTP1B could contribute to insulin resistance [Wu, X-2001]. Several groups have also reported that phosphorylation of PTP1B by both the IR and other protein kinases also affects PTP1B enzyme activity[Tao, J-2001]. Understanding what regulates the in vivo activity of PTP-1B and how the phosphatase interacts with the IR would enhance efforts to develop potent inhibitors for the enzyme.

8.6 PTP1B AND OBESITY RESISTANCE

An unexpected phenotype of the PTP1B KO mouse was its resistance to diet-induced obesity (DIO). Because insulin promotes the storage of glucose and fat, it was expected that PTP1B KO mice would be rather more susceptible to obesity. Not only are the homozygous mice resistant to DIO, but the heterozygotes also display this phenotype, suggesting that an ~50% reduction in PTP1B levels would be sufficient for insulin sensitization and obesity resistance. In fact, this was recently validated by studies with the PTP1B ASO [Rondinone, CM-2002]. Several factors appear to contribute to the obesity resistance phenotype. For instance, PTP1B KO mice have been reported to exhibit enhanced leptin sensitivity [Zabolotny, JM-2002]. It has been suggested that this may be due to PTP-1B acting as a negative regulator of leptin signaling by dephosphorylating the leptin receptor-associated kinase Jak2 [Cheng, A-2002]. Although a role for PTP1B in leptin signaling seems possible, the studies reported with the PTP1B KO mouse do not conclusively implicate a role for the phosphatase in leptin signaling. This is because in both the

ob/ob/ PTP1B double knockout and the *PTP1B–/–* mice treated with gold thioglucose to ablate leptin-responsive hypothalamic neurons, leptin signaling is absent. The resulting mice, however, were more resistant to obesity than their respective controls [Zabolotny, JM-2002]. Because leptin signaling is absent and therefore cannot be influenced by increased signaling (via an absence of PTP1B), the results would indicate that other mechanisms besides enhanced leptin signaling contribute to the obesity resistance. Indeed, in both models of the PTP1B KO mice that lacked leptin signaling, insulin sensitivity was maintained at the level of the control lean mice. Because insulin and leptin sensitivity are very tightly coupled [Ceddia, RB-2002], the enhanced leptin sensitivity observed in the *PTP1B–/–* mice could be an indirect effect of the insulin-sensitive phenotype and not necessarily a direct effect of PTP1B on leptin signaling. More work is required to show definitively that PTP1B has a direct role in leptin signaling.

8.7 Structure of Protein Tyrosine Phosphatases (PTP 1B)

PTP 1B , a 37-kD protein comprised of a single domain, is topologically organized into 8 alpha helices and 12 beta sheets. Spanning the length of the molecule is a highly twisted, 10-stranded, mixed beta sheet. Beta-12, the final beta strand of this structure's primary sequence, is located proximally to a central region containing parallel strands with the order beta-3, beta-12, beta-4, and beta-11. This motif, in turn, is flanked by a series of anti-parallel beta strands.

The center of the sheet is bounded on either side by alpha helix 2 on one side and helices alpha-3 and alpha-4 on the adjacent side. Following alpha-4, the chain is observed to form a four-helix bundle consisting of helices alpha-5, alpha-6, alpha-3, and alpha-4.

Figure 2 Secondary structure sequence of PTP1B

A short alpha helix, alpha-1, is found at the top of a beta-sheet neighboring beta-1. This segment denotes the beginning of the conserved PTP domain. Above the central beta-sheet is an additional anti-parallel beta-sheet containing beta-5 and beta-6. The carboxyl terminus of the chain is situated at the terminus of alpha-6. The non-conserved amino-terminal of the chain folds into two alpha-helices labeled alpha-1' and alpha-2'. These helices are found to wrap about the NH2-terminus of alpha-6.

Situated on loop 15, a segment which connects beta-12 and alpha-4, is the cysteine residue (Cys215) critical to the catalytic function of this enzyme. Intervening loops connecting secondary structural elements are observed to converge about this re- gion. Most of the 27 invariant residues in the sequences of PTP domains for which activ- ity toward phosphotyrosine-containing proteins or peptides has been reported are found on these loops. Invariant residues more distally located from the catalytic site are buried and likely stabilize the protein's tertiary structure.

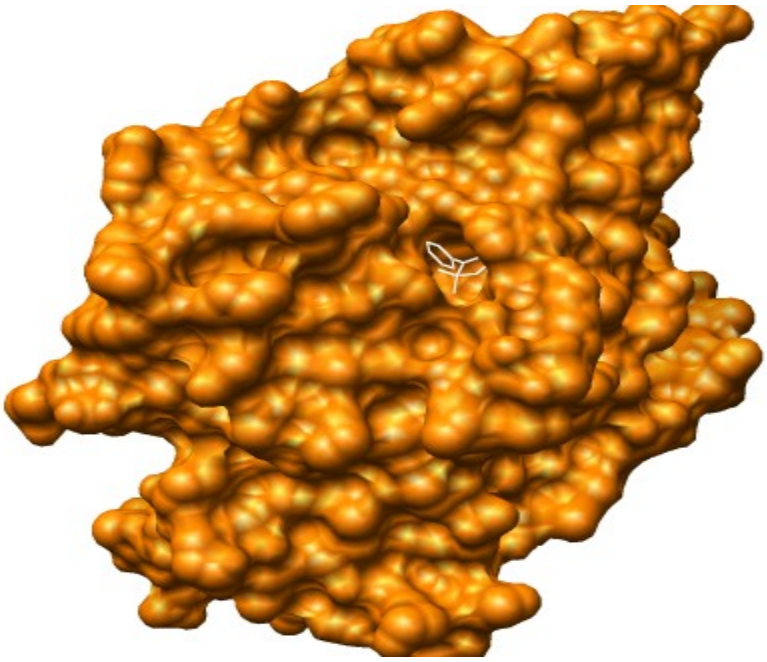

Figure 3 PTP1B complexed with (Inhibitor) 5-phenyl-1,2,5-thiadiazolidin-3-ONE 1,1-dioxide

8.8 Virtual Screening

The dominant technique for the identification of new lead compounds in drug discovery is the physical screening of large libraries of chemicals against a biological target (high-throughput screening). An alternative approach, known as virtual screening, is to computationally screen large libraries of chemicals for compounds that complement targets of known structure, and experimentally test those that are predicted to bind well. Such receptor-based virtual screening faces several fundamental challenges, including sampling the various conformations of flexible molecules and calculating absolute binding energies in an aqueous environment. Nevertheless, the field has recently had important successes: new ligands have been predicted along with their receptor-bound structures — in several cases with hit rates (ligands discovered per molecules tested) significantly greater than with high-throughput screening. Even with its current limitations, virtual

screening accesses a large number of possible new ligands, most of which may then be simply purchased and tested. For those who can tolerate its false-positive and false-negative predictions, virtual screening offers a practical route to discovering new reagents and leads for pharmaceutical research.

PTP1B protein complexed with 5-phenyl-1,2,5-thiadiazolidin-3-ONE 1,1-dioxide (PDB Code : 2BGE, Resolution: 1.80 Angstrom) has been used for the virtual screening and molecular docking of PTP1B. The primary objective of this study was to find more potent and selective PTP1B inhibitors with better pharmacological profiles than 5-phenyl-1,2,5-thiadiazolidin-3-one 1,1-dioxide. As a first step, virtual screening [BRIAN-04] was carried out using this protein structure of PTP1B and removing 5-phenyl-1,2,5-thiadiazolidin-3-one 1,1-dioxide. In this study DOCK 5.4 and other accessory programs were used to screen and rank database of over 4.1 million ligands.

9 Materials and Methods

The following software programs and database were used for molecular docking and virtual screening of ligand database.

1. **DMS**
2. **SPHGEN**
3. **SHOWBOX**
4. **SHOWSPHERE**
5. **GRID**
6. **Chimera**

7. **DOCK 5.4 version**
8. **ZINC -** A Free Database of Commercially Available Compounds for Virtual Screening

9.1 Dock working principle

Step 1: Start with crystal coordinates of target receptor

Step 2: Generate molecular surface for receptor. This is performed by DMS program. Only surface for the active site needs to be generated.

Step 3: Generate spheres to fill the active site. The shape of cavities in the receptor is used to define spheres; the centers of the spheres become potential locations for ligand atoms.

Step 4: Matching. Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand. Typically on the order of tens of thousands of orientations are generated for each ligand molecule.

Step 5: Scoring. Each oriented molecule is then scored for fit. There are currently 3 scoring schemes:

- Shape scoring, which uses a loose approximation to the Lennard-Jones potential
- Electrostatic scoring, which uses the program DELPHI to calculate electrostatic potential
- Force-field scoring, which uses the AMBER potential.

The top scoring orientation for each molecule is then saved, and used to compare to scores of other molecules. The final result is ordered by score.

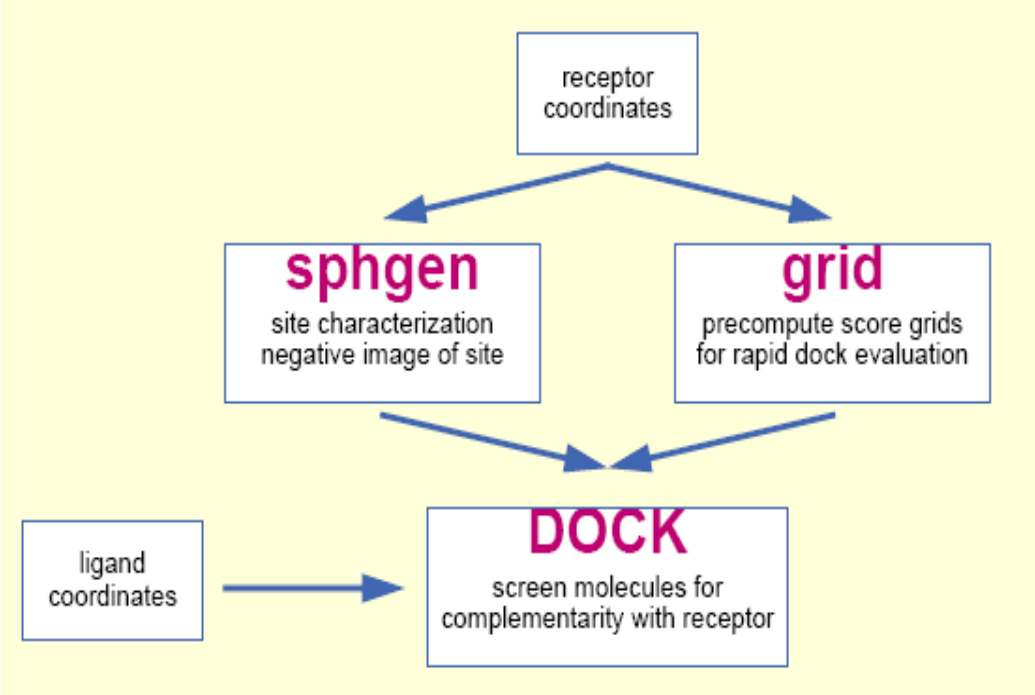

Figure 4 Main programs in DOCK suite

9.2 Preparing Molecules for DOCKing:

Description of the steps required to prepare molecules as input for a DOCK run that attempts to predict the orientation of a ligand in an active site.

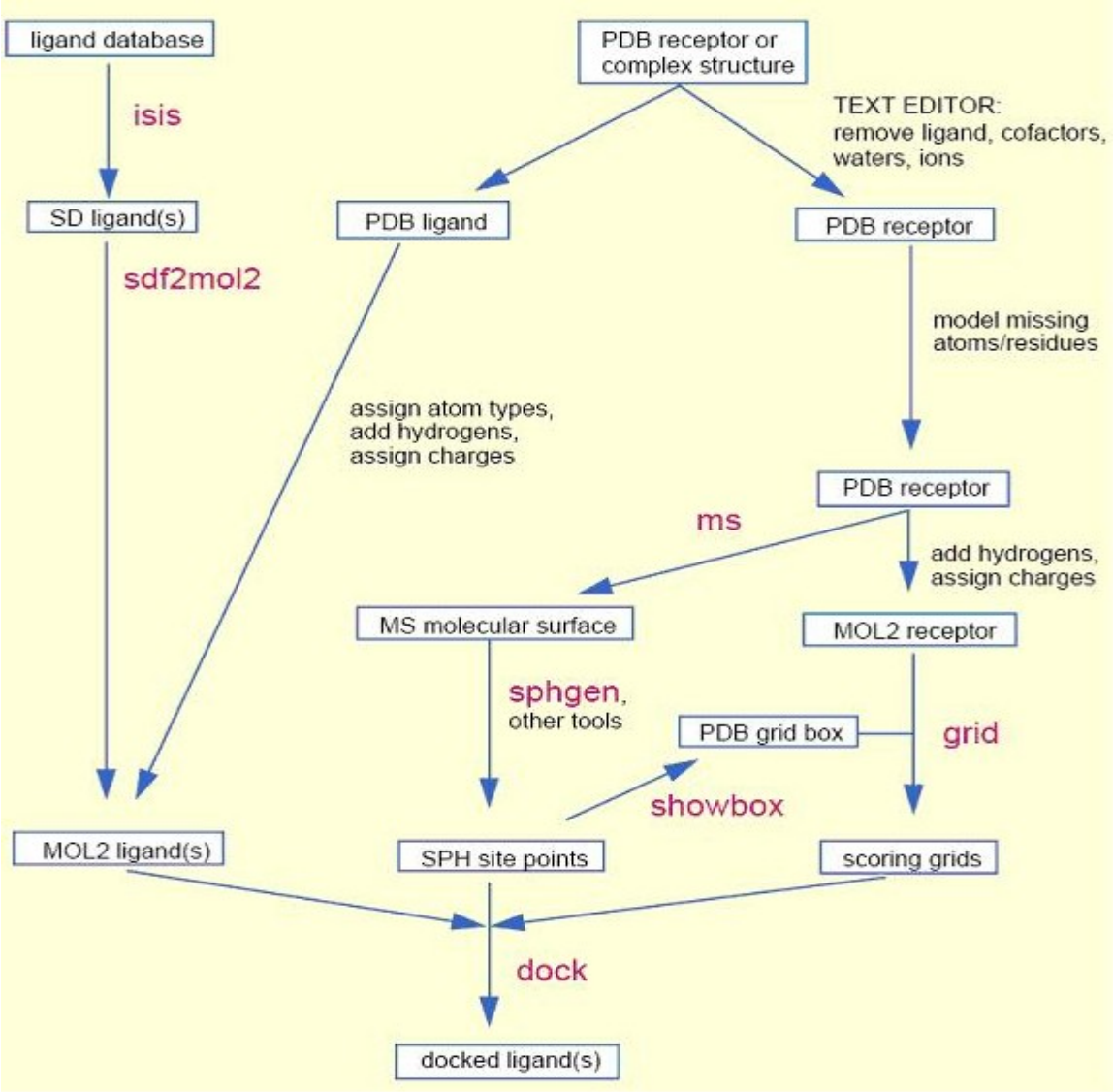

Figure 5 Docking Task Flowchart

STEP 1: Examine the pdb file

The first step in any docking project is selecting the file that will be used for the structure of the target.

STEP 2: Prepare the ligand file

a) Create file containing only ligand atoms and connectivity (lig.pdb).

b) Add hydrogens to the ligand (lig_H.pdb). It is critical that the atom type, bond order, and protonation state of the ligand are checked at this point for accuracy. Mistakes in any of these values will effect later charge calculations and docking results.

c) Calculate charges for the ligand and save in mol2 format (lig_charged.mol2).

STEP 3: Prepare the receptor file

a) Create new file containing receptor and all other molecules thought to be required for binding (rec.pdb). Importantly, though, if other molecules are included at this point, they should be added to the rec.pdb file and be considered part of the receptor for the rest of the steps. All other components of the pdb file should be removed.

b) Modify residues as necessary (rec_fixed.pdb). Almost every protein target needs some level of modeling before it can be used in docking. Once again, choices of what to model and what to ignore are highly system dependent and should be considered on a case by case basis. **Below is a list of common issues**:

1. Residues listed in protein sequences that are incomplete or missing

2. Protonation states of amino acid (i.e. histidine) that vary based on chemical environment

3. The biological unit of the protein is a dimer or trimer, even though only the monomer has been included in the structure file

4. Unnatural amino acids have been incorporated into the receptor structure

c) Add hydrogens, calculate charges, and save in mol2 format (rec_charged.mol2).

The above tasks i.e. removing solvent molecules, adding H atoms, adding charges and saving in mol2 format can be accomplished by the software, called **CHIMERA** from UCSF [PETTERSON-04].

9.3 Generating Spheres

STEP 1: Generate the molecular surface of target. This was done by the program **DMS** [CONNOLLY-83,CONNOLLY-83a]

9.3.1 DMS

DMS is "Distributed MS" a version of the solvent-accessible surface computing program. The molecular surface resembles the van der Waals surface of a molecule, except that crevices between atoms are smoothed over and interstices too small to accommodate the probe are eliminated. The surface includes cavities in the interior of the molecule, even if they are not accessible to a solvent molecule coming from the outside.

The molecular surface calculated is that traced out by the surface of the probe sphere rather that the probe spheres's center. According to Richards' definition the molecular surface consists of two parts: contact surface and reentrant surface. The contact surface is made up of those parts of the molecular van der Waals surface that can actually be in contact with the surface of the probe. The reentrant surface is defined by the interior-facing part of the probe when it is simultaneously in contact with more than

one atom. DMS reports the amounts of contact and reentrant surface area, and the combined total surface area on the standard error output.

The input file must be in the Protein Data Bank (PDB) format. The first letter or first two letters of the atom name is used to determine the element type. By default, implicit hydrogens are included for carbon, nitrogen and oxygen atoms, thus aromatic carbons and nitrogens will have van der Waals radii that are somewhat too big. Note that only amino acid and nucleic acid residues will be included unless -a is also specified.

OPTIONS:

The flags may be in any order. The meanings of the flags are described below:

-a Include all atoms, not just those in amino acid and nucleic acid residues.

-d Change the density of points on the surface. Density is a factor affecting the density of points on the surface; the default of 1.0 produces about 5 points per square angstrom. Only values between 0.1 and 10.0 are permitted. For large proteins, a density of 0.5 is recommended.

-g Write all the informative messages to file, instead of the standard error output. Genuine errors still go to the standard error output. This file is not rewound at any time, so messages from several runs may be accumulated.

-i Calculate the molecular surface only for those residues and atoms specified in file, but keeping the rest of the molecule for collision checks. The file consists of a series of lines such as the following:

ASP 205 CA

TYR 13 *

GLY 116 FRM

HIS 178 TO

The asterisk means all atoms of the residue and the ‘‘FRM’’ and ‘‘TO’’ mean all residues from 116 to 178 inclusive. The sequence number may contain letters, and if the PDB input file contains chain identifiers, then those should be appended on the right of the sequence

number. Residue insertion codes should be placed between the sequence number and any chain identifier. Residues contained in HETATM records should have an asterisk appended to the end of the residue identifier. The surface generated using the -i flag is not always the same as the surface generated by running the entire molecule and afterwards selecting out the desired atoms. The first surface will not include reentrant surface lying between an atom in the –i file and atoms not in the file.

-n Include the unit normals to the surface with each surface point record.

-v Produce more verbose output. DMS will announce each computation phase as it is entered as well as a count of the atom types in the molecule and the number of computation requests handled by each host that participated in the DMS calculation.

-o the output is written to file. This flag is not optional.

-w Change the water probe radius from the default radius of 1.4 angstroms. This parameter must be between 1.0 and 201.0.

The output consists of a series of atom and surface point records, with the same format for the first six fields. Each atom is followed by the surface points (if any) which belong to it. These first six fields are in the following format: residue name, sequence number, atom name, x coordinate, y coordinate, z coordinate. For an atom record, the seventh field is ‘‘A.’’ For a surface point record, the seventh field begins with an ‘‘S,’’ followed by a ‘‘C,’’ ‘‘R,’’ or ‘‘S’’ according to whether the point is part of contact, reentrant, or ‘‘saddle’’ surface (‘‘saddle’’ is a type of reentrant surface where the probe is in contact with exactly two atoms). This is followed a digit used for depicting different density levels. The eighth field is the molecular surface area associated with the point in square angstroms. If the -n flag is specified, the next three fields are the unit normal vector pointing outward from the surface. Informative messages and errors are written to the standard error output unless a -g file is specified.

The chemical elements and radii that the program handles are detailed in the table below. The program gets these values from the file /usr/local/midas/resource/dms/radii. If there is a file in the current directory called radii, then DMS will use that file instead. So in

order to add uncommon elements or use different radii, one should copy the default file and modify it. The file format is documented in the file itself.

Element	Radius
H	1.20
C	1.90
N	1.50
O	1.40
F	1.35
P	1.90
S	1.85
Cl	1.80
Fe	0.64
Cu	1.28
Zn	1.38
Br	1.95
I	2.15
Other	1.90

Table 1 DMS Radius for various elements

Usage: dms input_file -a –n –w 1.4 –v –o output_file

Example: dms receptor.pdb –a –n –w 1.4 –v –o rec.ms

STEP 2: Generate spheres

To generate spheres from the molecular surface and normal vectors, the program **sphgen** that is distributed as an accessory to DOCK (add reference) was used.

SPHGEN [KUNTZ-82] generates sets of overlapping spheres to describe the shape of a molecule or molecular surface. For receptors, a negative image of the surface invaginations is created; for a ligand, the program creates a positive image of the entire molecule. Spheres are constructed using the molecular surface described by Richards (1977) calculated with the program dms. Each sphere touches the molecular surface at two points and has its radius along the surface normal of one of the points. For the receptor, each sphere center is "outside" the surface, and lies in the direction of a surface normal vector. For a ligand, each sphere center is "inside" the surface, and lies in the direction of a reversed surface normal vector. Spheres are calculated over the entire surface, producing approximately one sphere per surface point.

This very dense representation is then filtered to keep only the largest sphere associated with each receptor surface atom. The filtered set is then clustered on the basis of radial overlap between the spheres using a single linkage algorithm. This creates a negative image of the receptor surface, where each invagination is characterized by a set of overlapping spheres. These sets, or "clusters," are sorted according to numbers of constituent spheres, and written out in order of descending size. The largest cluster is typically the ligand binding site of the receptor molecule. The program showsphere writes out sphere center coordinates in PDB format and may be helpful for visualizing the clusters.

For generating spheres using **sphgen**, you must generate a file named INSPH, the format of which is below:

rec.ms #molecular surface file

R	#sphere outside of surface (R) or inside surface (L)
X	#specifies subset of surface points to be used (X=all points
0.0	#prevents generation of large spheres with close surface contacts (default=0.0)
4.0	#maximum sphere radius in Angstroms (default=4.0)
1.4	#minimum sphere radius in Angstroms (default=radius of probe)
rec.sph	#clustered spheres file

To generate the spheres, simply use the command "sphgen" in the same folder that contains the INSPH file. The output will be two files: rec.sph, which contains the spheres in clusters, and OUTSPH, which contains general information about the calculation.
NOTE: For sphgen to work, the comments above - labeled by # - must be removed from the INSPH file. Also, if the calculation has been run before, all output files (including OUTSPH and rec.sph) must be removed from the working directory. Finally, for technical reasons, Sphgen cannot handle more than 9999 spheres. If the target is large target, spheres can be generated by selecting a subsection of the protein using a visualization program and using it to generate the molecular surface and spheres.

STEP 3: Select subset of spheres to be used in DOCK

OPTION 1: Use the largest cluster generated by sphgen. The clusters contained in the rec.sph file are ranked according to size (number of spheres in the cluster). In most cases, the largest cluster is typically the ligand binding site of the molecule. To visualize the spheres, the program **showsphere** can be used, distributed as an accessory to DOCK.
This program also has an input file, shown below:

rec.sph	#sphere cluster file
1	#cluster number to process (<0 = all)

N #generate surface as well as pdb file

selected_cluster.pdb #name for output file

To convert the selected sphere file to pdb format, use the command:

$ showsphere < sphgen_cluster.in

If we are satisfied with the spheres in this cluster, create a new file containing only the selected cluster (**sphgen_cluster.sph**).

OPTION 2: Select spheres within some radius of a desired location. If the active site is known, we can select spheres within a radius of a set of atoms that describes the site. To do this, we need to use the program **sphere_selector**, which is distributed as an accessory with DOCK. **sphere_selector** will take the output from SPHGEN and select all spheres with a user-defined radius of a target molecule. The target molecule can be anything (i.e. known ligand, receptor residue, etc) as long as it is in proper MOL2 format. The required input for sphere_selector is:
USAGE:

sphere_selector<sphere_cluster_file.sph> <set_of_atoms.mol2> <radius>

Here, we select all spheres within 10.0 Angstroms root mean square deviation (RMSD) from every atom of the crystal structure of the ligand, using the command

$ sphere_selector rec.sph lig_charged.mol2 10.0

Step 4: Construction of box to enclose spheres

SHOWBOX is an interactive program that allows visualization of the location and size of the grids that will be calculated by the program grid, using any graphics program that can display PDB format. The user is asked whether the box should be automatically constructed to enclose all of the spheres in a cluster. If so, the user must also enter a value for how closely the box faces may approach a sphere center (how large a “cushion” of space is desired and the sphere cluster filename and number. If not, the user is asked whether the box will be centered on manually entered coordinates or a sphere cluster center of mass. Depending on the response, the coordinates of the center or the sphere cluster filename and number are requested. Finally, the user must enter the desired box dimensions (if not automatic) and a name for the output PDB-format box file.

Usage:

$ showbox < box.in

The input file box.in has the following details

Y

5

Selected_sphere.sph

1

rec_box.pdb

9.4 Scoring GRID construction

GRID creates the grid files necessary for rapid score evaluation in DOCK. Two types of scoring are available: contact and energy scoring. The scoring grids are stored in files ending in *.cnt and *.nrg respectively. When docking, each scoring function is applied independent of the others and the results are written to separate output files. GRID also computes a bump grid which identifies whether a ligand atom is in severe

steric overlap with a receptor atom. The bump grid is identified with a *.bmp file extension. The file containing the bump grid also stores the size, position and grid spacing of all the grids.

GRID must be run command line from a standard UNIX shell. It reads a parameter file containing field/value pairs using the following command:

USAGE:

```
$   grid -i grid.in
```

grid.in

```
compute_grids             yes
grid_spacing              0.3
output_molecule           no
contact_score             yes
contact_cutoff_distance   4.5
chemical_score            no
energy_score              yes
energy_cutoff_distance    12.0
atom_model                a
attractive_exponent       6
repulsive_exponent        12
distance_dielectric       yes
dielectric_factor         4
bump_filter               yes
bump_overlap              0.75
receptor_file             rec_fixed.mol2
box_file                  rec_box.pdb
vdw_definition_file       parameters/vdw_AMBER_parm99.defn
score_grid_prefix         grid
```

9.5 Running DOCK

DOCK [MENG-92, MENG-93, SHOICHET-92] must be run command line from a standard UNIX shell. It read a parameter file containing field/value pairs using the following command:

$ dock/bin/dock5 –i parameter.in

9.5.1 Example parameter input file:

```
ligand_atom_file                /ligand/ligand.mol2
ligand_outfile_prefix           ligand
limit_max_ligands               no
write_orientations              no
write_conformations                     no
initial_skip                    0
calculate_rmsd                  yes
rank_ligands                    yes
num_scored_poses_written        1
cluster_ranked_poses            no
orient_ligand                   yes
automated_matching              yes
receptor_site_file              selected_spheres.sph
max_orientations                1000
critical_points                 no
chemical_matching               no
use_ligand_spheres              no
```

```
flexible_ligand                 no
bump_filter                     yes
bump_grid_prefix                ../grid/grid
max_bumps                       2
score_molecules                 yes
grid_score_primary              yes
grid_score_secondary            yes
grid_score_vdw_scale            1
grid_score_es_scale             1
grid_score_grid_prefix          ../grid/grid
minimize_ligand                 yes
minimize_anchor                 no
minimize_flexible_growth        no
minimize_final_pose             yes
use_advanced_simplex_parameters no
simplex_max_cycles              1
simplex_score_converge          0.1
simplex_cycle_converge          1.0
simplex_trans_step              1.0
simplex_rot_step                0.1
simplex_tors_step               10.0
simplex_final_max_iterations    1000
simplex_random_seed             0
atom_model                      all
vdw_defn_file                   parameters/vdw_AMBER_parm99.defn
flex_defn_file                  parameters/flex.defn
flex_drive_file                 parameters/flex_drive.tbl
chem_defn_file                  parameters/chem.defn
```

9.6 ZINC - Database of Commercially Available Compounds for Virtual Screening

ZINC [JOHN-05] contains molecules, each with 3D structure using catalogs of compounds from vendors. The molecules have been assigned biologically relevant protonation states and are annotated with properties such as molecular weight, calculated log P, and number of rotatable bonds. Each molecule in the library contains vendor and purchasing information and is ready for docking.

9.6.1 Salient criteria for ZINC

1. Compounds should be purchasable for rapid testing of docking hypotheses.
2. Subsets of molecules with variable properties such as functional groups, molecular weight, and calculated log P should be easy to create and manipulate.
3. The database must support multiple protonation models, tautomeric forms, stereo chemistries (e.g. racemic mixtures as well as stereo chemically pure compounds), regioisomeric forms (E/Z isomerism), suppliers, and 3D conformational sampling.
4. It should be possible to annotate molecules using both numeric and alphanumeric data.
5. It should be easy to add new molecules, tag or remove those that are no longer available and fix those that have errors.
6. The database should be quick to search and download, and it should be straightforward to obtain regular updates.

9.6.2 Methods followed in making ZINC

Molecules are filtered out with formula weight greater than 700, calculated log P greater than 6 and less than 4, number of hydrogen-bond donors greater than 6, number of hydrogen-bond acceptors greater than 11, and number of rotatable bonds greater than 15. Molecules containing an atom other than H, C, O, N, F, S, P, Cl, Br,or I are removed. Exceptions are made to include a number of actual drugs that violate these constraints;

these rules are guidelines toward making the database loosely conform to current opinion in the field.

9.6.3 Various programs used in preparing ZINC database:

1. Conversion of 2D SDF files to isomeric SMILES by OpenEye's convert.py tool (OpenEye Scientific Software, http://www.eyesopen.com)

2. To desalt the molecules and filter out undesirable molecules – OpenEye's filter.1.0.2 program
3. To ensure uniqueness in the database, canonical representations are calculated with OpenEye's OEchem library.
4. OpenEye's Omega program(calculates accessible conformations relatively accurately and efficiently) is used to generate initial 3D models from unambiguous isomeric SMILES.
5. Schrodinger's ligprep program (www.schrodinger.com) is employed to create relevant, correctly protonated forms of the molecule between pH 5 and 9.5. For example this includes deprotonating carboxylic acids and tetrazoles and protonating most aliphatic amines.
6. The semiempirical quantum mechanical program AMSOL calculates the partial atomic charges and atomic desolvation penalties for a single 3 D conformation of each protonation state, stereoisomer, and tautomer.
7. 3D conformatons are distilled into a flexibase format using mol2db program.

Molecules in ZINC are annotated by molecular property. These include molecular weight, number of rotatable bonds, calculated log P (OpenEye's wang's algorithm) number of hydrogen-bond acceptors, number of hydrogen-bond donors, number of chiral

double bonds(E/Z isomerism), polar and apolar desolvation energy (in kcal/mol), net charge and number of rigid fragments. Each molecule is also annotated with the vendor and original catalog number of each commercial source of that compound. Molecules are available for download in SMILES, mol2, sdf and DOCK-flexibase formats.

Subset name	**log P**	**Molecular weight**	**hydrogen-bond donor**	**hydrogen-bond acceptor**
lead-like	< 4 > -2	< 350 > 150	≤ 3	≤ 6
drug-like	≤ 5	≤ 500	≤ 5	≤ 10
Fragment-like	< 3 > -2	≤ 250	< 3	< 6
Greasy-leads	< 6 > 2	< 350	-	-

Table 2 ZINC Ligand Properties

10 Results and Discussion

The following classes of ligands are downloaded from ZINC database for virtual screening:

Class of chemicals	Number of Molecules
Drug like	25,29,908
Lead like	7,51,695
Greasy	8,29,114
Fragment like	56,526
Total	41,67,243

Table 3 Chemical classes for screening

Virtual screening was carried out for finding novel ligands of PTP1B with over 4.1 millions ligands of various chemical properties. Top 10 molecules from all chemical class except are presented here along with their energy score of both van der Waals and electrostatic components.

10.1 Greasy top 10 Ligands

1)

Name: ZINC00969372
RMSD: 0
Energy Score: -56.777107
vdw: -35.822250
es: -20.954855

2-[4-oxo-2-(4-tert-butylcyclohexylidene)aminoimino-thiazolidin-5-yl]acetic

2)

Name: ZINC02136559
RMSD: 0
Energy Score: -50.638535
vdw: -33.773861
es: -16.864672

3-[4-(4-methoxyphenyl)phenyl]-2-oxo-propanoic

3)

Name: ZINC02433415
RMSD: 0
Energy Score: -50.438122
vdw: -34.422985
es: -16.015137

[No name assigned by Zinc]

4)

Name: ZINC02066448
RMSD: 0
Energy Score: -50.314754
vdw: -30.562346
es: -19.752409

2-[4-[(2-methyl-4-oxo-3H-quinazolin-3-yl)iminomethyl]phenoxy]acetic

5)

Name: ZINC01521756
RMSD: 0
Energy Score: -49.210175
vdw: -35.411560
es: -13.798615

3-hydroxy-2-[4-(4-isobutylphenyl)thiazol-2-yl]amino-propanoic

6)

Name: ZINC00068435
RMSD: 0
Energy Score: -48.916840
vdw: -26.795036
es: -22.121805

4-[3-(4-chlorophenyl)-3-oxo-prop-1-enyl]aminobenzoic

7)

Name: ZINC00160629
RMSD: 0
Energy Score: -48.898354
vdw: -27.933342
es: -20.965012

3-[4-(4-chlorophenyl)sulfanyl-3-nitro-phenyl]acrylic

8)

Name: ZINC01259447
RMSD: 0
Energy Score: -48.785091
vdw: -30.304573
es: -18.480518

3-[4-(3-ethylphenyl)phenyl]propanoic

9)

Name: ZINC01523271
RMSD: 0
Energy Score: -48.370918
vdw: -29.605032

es: -18.765884

3-[4-(4-ethoxy-3-sec-butyl-phenyl)thiazol-2-yl]aminopropanoic

10)

Name: ZINC00453384
RMSD: 0
Energy Score: -48.139896
vdw: -34.194965
es: -13.944932

2-[4-(2-phenoxypropanoylamino)phenoxy]acetic

10.2 Drug like top 10 ligands

1)

Name: ZINC01001250
RMSD: 0
Energy Score: -59.343987
vdw: -30.486544
es: -28.857441

4-[[4-[(2-carboxymethyloxyphenyl)methylidene]-3-methyl-5-oxo-1H-pyrazolyl]]benzoic

2)

Name: ZINC00969372
RMSD: 0
Energy Score: -56.777107
vdw: -35.822250
es: -20.954855

2-[4-oxo-2-(4-tert-butylcyclohexylidene)aminoimino-thiazolidin-5-yl]acetic

3)

Name: ZINC01257923
RMSD: 0
Energy Score: -56.659447
vdw: -31.582424
es: -25.077021

2-[4-(3-hydroxyaminocarbonylphenyl)phenyl]aminoacetic

4)

Name: ZINC02912978
RMSD: 0
Energy Score: -55.980484
vdw: -23.972197
es: -32.008289

2-[3-carboxy-4-(3- carboxypropylamino)phenyl]sulfonylaminobenzoic

5)

Name: ZINC02788645
RMSD: 0
Energy Score: -54.718060
vdw: -28.874233
es: -25.843826

2-[4-(carboxymethoxy)phenyl]thiazolidine-4-carboxylic

6)

Name: ZINC01226912
RMSD: 0
Energy Score: -53.940453
vdw: -30.952578
es: -22.987873

4-[3-(4-bromophenyl)aminosulfonyl-4-methyl-phenyl]carbonylaminobenzoic

7)

Name: ZINC02093264
RMSD: 0
Energy Score: -52.947525
vdw: -29.159521
es: -23.788004

4-[[3-(2-carboxyethyl)-4,8-dimethyl-2-oxo-chromen-7-yl]oxymethyl]benzoic

8)

Name: ZINC02809902
RMSD: 0
Energy Score: -52.394634
vdw: -8.363735
es: -44.030899

(3-phosphonooxyphenoxy)phosphonic

9)

Name: ZINC00977508
RMSD: 0
Energy Score: -52.388203
vdw: -32.380440
es: -20.007763

4-[[5-[(amino-dicyano-dimethyl-BLAHylidene)methyl]-2-furyl]]benzoic

10)

Name: ZINC00537457
RMSD: 0
Energy Score: -52.217987
vdw: -34.037426
es: -18.180559

2-(4-styrylpiperazinyl)carbonylaminopropanoic

10.3 Lead like top 10 ligands

1)

Name: ZINC00969372
RMSD: 0
Energy Score: -56.777107
vdw: -35.822250
es: -20.954855

2-[4-oxo-2-(4-tert-butylcyclohexylidene)aminoimino-thiazolidin-5-yl]acetic

2)

Name: ZINC00537457
RMSD: 0
Energy Score: -52.217987
vdw: -34.037426
es: -18.180559

2-(4-styrylpiperazinyl)carbonylaminopropanoic

3)

Name: ZINC01521178
RMSD: 0
Energy Score: -50.931358
vdw: -31.497229
es: -19.434132

3-hydroxy-2-BLAHylamino-propanoic

4)

Name: ZINC00215106
RMSD: 0
Energy Score: -50.759781
vdw: -26.026064
es: -24.733717

5-[(3-carbamoyl-4-ethyl-5-methyl-2-thienyl)amino]-5-oxo-pentanoic

5)

Name: ZINC01068949
RMSD: 0
Energy Score: -50.670303
vdw: -30.232044
es: -20.438257

2-[4-(3-amino-2-cyano-3-thioxo-prop-1-enyl)phenoxy]acetic

6)

Name: ZINC00203271
RMSD: 0
Energy Score: -50.572426
vdw: -31.478683
es: -19.093742

2-(6-ethoxy-4-methyl-quinazolin-2-yl)sulfanylacetic

7)

Name: ZINC00296700
RMSD: 0

Energy Score: -50.320896

vdw: -23.358137

es: -26.962759

3-[3-(3-hydroxy-3-oxo-prop-1-enyl)phenyl]acrylic

8)

Name: ZINC01257923

RMSD: 0

Energy Score: -50.260750

vdw: -25.411350

es: -24.849401

2-[4-(3-hydroxyaminocarbonylphenyl)phenyl]aminoacetic

9)

Name: ZINC00296251

RMSD: 0

Energy Score: -49.083397

vdw: -30.282198

es: -18.801199

3-(6-hydroxy-2-isopropyl-benzofuran-5-yl)acrylic

10)

Name: ZINC00391981
RMSD: 0
Energy Score: -48.658440
vdw: -24.429632
es: -24.228809

benzene-1,3,5-tricarboxylic

10.4 Fragment like top 10 ligands

1)

Name: ZINC02458587
RMSD: 0
Energy Score: -47.771179
vdw: -26.791307
es: -20.979874

5-[(5-amino-3-pyridyl)]thiophene-2-carboxylic

2)

Name: ZINC00874633
RMSD: 0
Energy Score: -47.692165
vdw: -27.364004
es: -20.328161

5-(4-chlorophenyl)isoxazole-3-carboxylic

3)

Name: ZINC00173374
RMSD: 0
Energy Score: -46.571960
vdw: -26.905027
es: -19.666931

5-(4-chlorophenyl)-2H-pyrazole-3-carboxylic

4)

Name: ZINC01034281
RMSD: 62.36
Energy Score: -46.357246
vdw: -27.482958
es: -18.874289

4-BLAHylbenzoic

5)

Name: ZINC02237034
RMSD: 0
Energy Score: -46.113373
vdw: -20.564249
es: -25.549124

2-(sulfanylmethyl)-1H-benzoimidazole-5-carboxylic

6)

Name: ZINC00105246

RMSD: 0

Energy Score: -46.009293

vdw: -22.753155

es: -23.256140

pyridine-2,6-dicarboxylic

7)

Name: ZINC00295225

RMSD: 0

Energy Score: -45.945549

vdw: -25.749449

es: -20.196100

2-(3-oxoisoindolylidene)acetic

8)

Name: ZINC03260103

RMSD: 0

Energy Score: -45.741486

vdw: -28.218630

es: -17.522858

oxoBLAHcarboxylic

9)

Name: ZINC00155997

RMSD: 0

Energy Score: -45.617790

vdw: -25.352243

es: -20.265547

3-(3-hydroxyphenyl)acrylic

10)

Name: ZINC03748658
RMSD: 0
Energy Score: -45.444221
vdw: -26.490753
es: -18.953466

5-isopropyl-1H-indole-2-carboxylic

10.5 DOCK result of inhibitor (5-phenyl-1,2,5-thiadiazolidin-3-ONE 1,1-dioxide) with PTP1B

Name: Inhibitor.pdb
RMSD: 4.87387
Energy Score: -31.339895
vdw: -31.339895
es: 0.000000

10.6 Top 5 ligands from all classes of ligands

1)

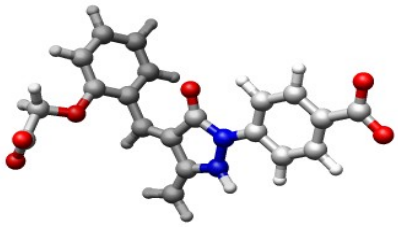

Figure 6 ZINC01001250

Name: ZINC01001250
RMSD: 0
Energy Score: -59.343987
vdw: -30.486544
es: -28.857441

4-[[4-[(2-carboxymethyloxyphenyl)methylidene]-3-methyl-5-oxo-1H-pyrazolyl]]benzoic

2)

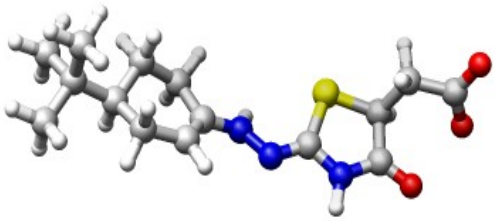

Figure 7 ZINC00969372

Name: ZINC00969372
RMSD: 0
Energy Score: -56.777107
vdw: -35.822250
es: -20.954855

2-[4-oxo-2-(4-tert-butylcyclohexylidene)aminoimino-thiazolidin-5-yl]acetic

3)

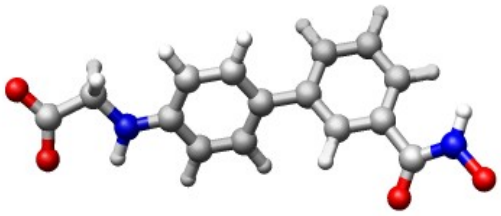

Figure 8 ZINC01257923

Name: ZINC01257923
RMSD: 0
Energy Score: -56.659447
vdw: -31.582424
es: -25.077021

2-[4-(3-hydroxyaminocarbonylphenyl)phenyl]aminoacetic

4)

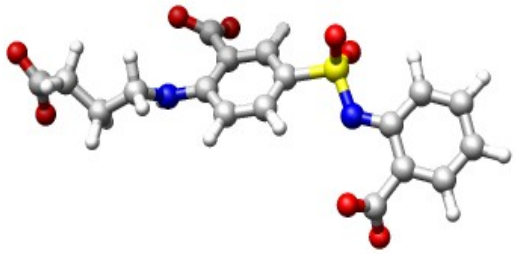

Figure 9 ZINC02912978

Name: ZINC02912978
RMSD: 0
Energy Score: -55.980484
vdw: -23.972197
es: -32.008289

2-[3-carboxy-4-(3- carboxypropylamino)phenyl]sulfonylaminobenzoic

5)

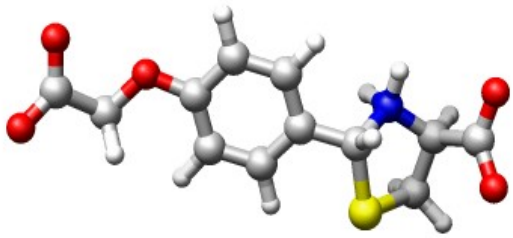

Figure 10 ZINC02788645

Name: ZINC02788645
RMSD: 0
Energy Score: -54.718060
vdw: -28.874233
es: -25.843826

2-[4-(carboxymethoxy)phenyl]thiazolidine-4-carboxylic

10.7 PTP1B with top 5 ligands

1)

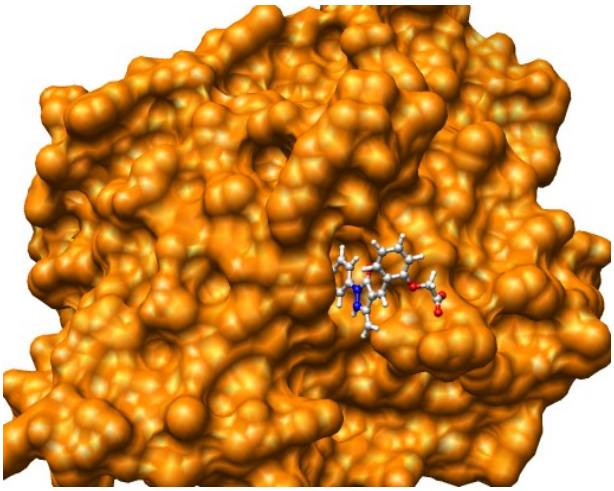

Figure 11 PTP1B and ZINC01001250

2)

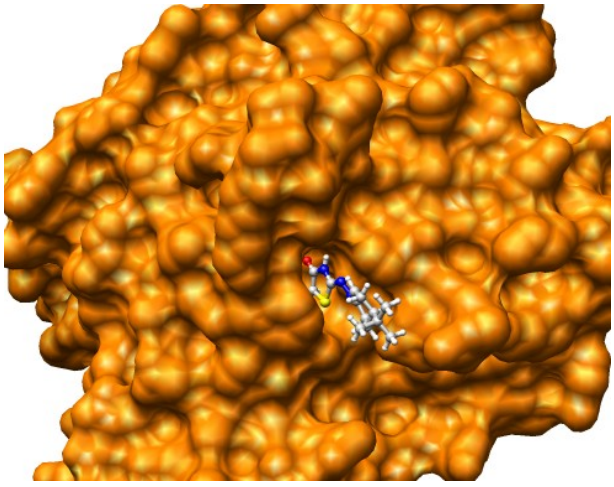

Figure 12 PTP1B and ZINC00969372

3)

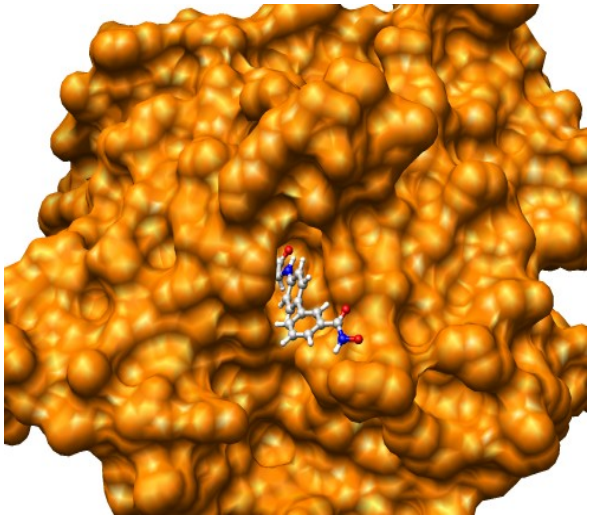

Figure 13 PTP1B and ZINC01257923

4)

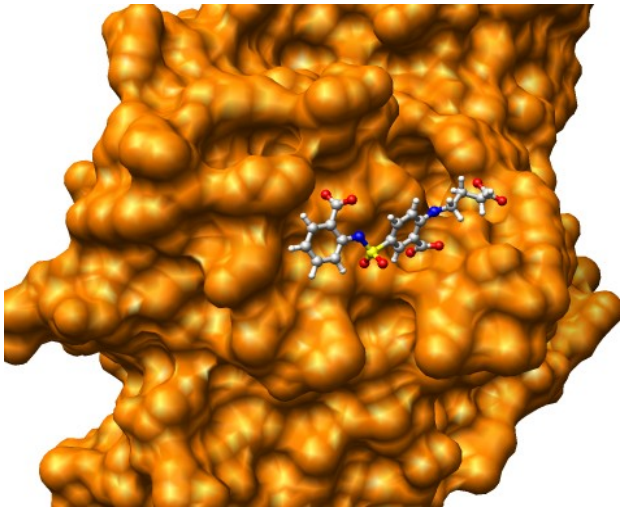

Figure 14 PTP1B and ZINC02912978

5)

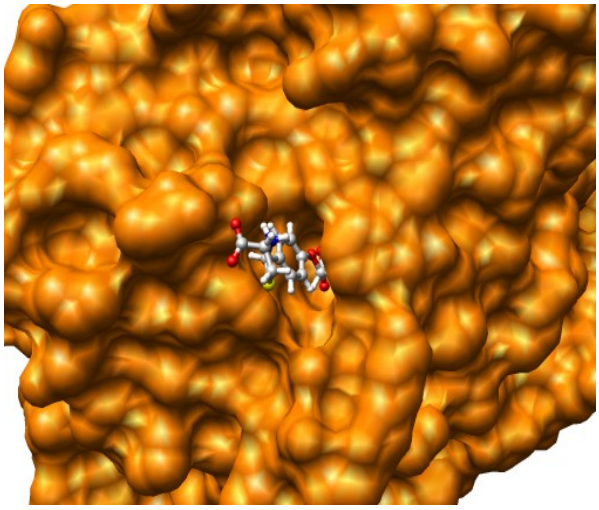

Figure 15 PTP1B and ZINC02788645

About 41,67,243 ligand molecules of various chemical properties were screened by DOCK program and there are quite good number of ligand molecules were surpassing the specific competitive inhibitor 5-PHENYL-1,2,5-THIADIAZOLIDIN-3-ONE 1,1-

DIOXIDE in terms of energy score consisting of van der Waals and electrostatic components. These ligands, especially belonging to drug like, chemical class of Zinc database are putatively better ligands for PTP1B inhibition, though biochemical experimental confirmation of them is must.

11 Conclusion

This project was aimed at finding novel lead molecules for the selective competitive inhibition of PTP1B protein, an important protein that plays crucial role in the insulin resistance and obesity. The X ray crystal structure (resolution 1.80 Angstrom) of PTP1B along with specific inhibitor (5-phenyl-1,2,5-thiadiazolidin-3-ONE 1,1-dioxide) (PDB code 2BGE) was considered as receptor molecule after removing all water molecules and 5-phenyl-1,2,5-thiadiazolidin-3-ONE 1,1-dioxide.

This protein was used for virtual screening for finding ligands having better energy scores using the program DOCK and database of commercially available chemical moieties over 4.1 million in numbers. There were good number of ligands with better energy scores compare to 5-phenyl-1,2,5-thiadiazolidin-3-one 1,1-dioxide being documented, though their efficacy, toxicity and pharmacokinetic properties must be

studied experimentally.

12 Future Proposal

Virtual Screening is the first and the foremost important step in the process of drug designing process. In our project we have done the virtual screening of the PTP1B protein against the database of 4.1 million ligands. Top 40 leads have been selected out of this process with the help of DOCK5.4.

These 40 ligands are showing good binding affinity towards the drug target(PTP1B) and it is proposed that further large scale optimization of these probable leads will definitely provide a potent inhibitor for the target. Optimization process will comprise of better scoring of the various conformations of the given leads which will require sophisticated computational environment and many days. Combination of Docking programs and scoring function can be useful for the better estimation of the Binding

affinity and search of more conformational space. Further optimization will be the QSAR(Quantitative Structural Activity Relationship) study of the ligands and study of their physiochemical properties. Since the binding affinity of the present leads is comparable to the already existing inhibitor of the target protein, we are sure of getting a potent inhibitor out of the proposed leads.

13 Bibliography

1. [HOTAMISLIGIL-05a]Inflammation, stress and diabetes, Kathryn E, Wellen and G S Hotamisligil, The journal of Clinical Investigation, Review, Vol. 115 No. 5 May 2005.
2. [JOHN-05] John J. Irwin and Brian K. Shoichet : ZINC - A Free Database of Commercially Available Compounds for Virtual Screening, , J. Chem. Inf. Model. 2005, 45, 177-182 (http://zinc.docking.org)
3. [LAZAR-05]Mitchell A. Lazar: How obesity causes diabetes: Not a tall tale, , Type 2 Diabetes viewpoint, Science, Vol. 307, 21 January 2005.

4. [LOWELL-05]Bradford B. Lowell and Gerald I. Shulman: Mitochondrial dysfunction and type 2 diabetes, , Type 2 Diabetes viewpoint, Science, Vol. 307, 21 January 2005.
5. [NAKATANI-05] Nakatani Y, et.al: Involvement of ER stress in insulin resistance and diabetes. J. Biol. Chem. Vol. 280, No. 1, January 7 2005, pp. 847-851.
6. [SCHWARTZ-05] Michael W. Schwartz and Daniel Porte Jr.: Diabetes, obesity and the brain, Type 2 Diabetes viewpoint, Science, Vol. 307, 21 January 2005.
7. [BRIAN-04] Brian K. Shoichet: Virtual Screening of Chemical Libraries, *Nature*, Vol. 432, 16 December 2004. (Insight commentary).
8. [JORGENSEN-04]The many roles of computation in Drug Discovery, William L. Jorgensen, Science, Vol. 303, 19 March 2004 (Drug Discover, special section).
9. [PETTERSON-04] Petterson. E. F.et.al: UCSF-Chimera – A visualization system for exploratory research and analysis, , J. Comput. Chem. 25(13): 1605-1612 (2004).Chimera home page: (http://www.cgl.ucsf.edu/chimera)
10. [ERNEST-03]Ernest Asante-Appiah and Brian P. Kennedy Protein tyrosine phosphatases: the quest for negative regulators of insulin action -2003
11. [BICKEL-02]Bickel, PE. Lipid rafts and insulin signaling. Am J Physiol Endocrinol Metab 282: E1-E10, 2002
12. [CEDDIA-02]Ceddia, RB, Koistinen HA, Zierath JR, and Sweeney G. Analysis of paradoxical observations on the association between leptin and insulin resistance. *FASEB J* 16: 1163-1176, 2002
13. [CHENG-02]Cheng, A, Uetani N, Simoncic PD, Chaubey VP, Lee-Loy A, McGlade CJ, Kennedy BP, and Tremblay ML. Attenuation of leptin action and regulation of obesity by protein tyrosine phosphatase 1B. *Dev Cell* 2: 497-503, 2002
14. [GAGNON-02]Gagnon, E, Duclos S, Rondeau C, Chevet E, Cameron PH, Steele-Mortimer O, Paiement J, Bergeron JJ, and Desjardins M. Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* 110: 119-131, 2002

15. [HAJ-02]Haj, FG, Verveer PJ, Squire A, Neel BG, and Bastiaens PI. Imaging sites of receptor dephosphorylation by PTP1B on the surface of the endoplasmic reticulum. *Science* 295: 1708-1711, 2002
16. [RONDINONE-02]Rondinone, CM, Trevillyan JM, Clampit J, Gum RJ, Berg C, Kroeger P, Frost L, Zinker BA, Reilly R, Ulrich R, Butler M, Monia BP, Jirousek MR, and Waring JF. Protein tyrosine phosphatase 1B reduction regulates adiposity and expression of genes involved in lipogenesis. *Diabetes* 51: 2405-2411, 2002
17. [ZABOLOTNY-02]Zabolotny, JM, Bence-Hanulec KK, Stricker-Krongrad A, Haj F, Wang Y, Minokoshi Y, Kim YB, Elmquist JK, Tartaglia LA, Kahn BB, and Neel BG. PTP1B regulates leptin signal transduction in vivo. *Dev Cell* 2: 489-495, 2002
18. [ZINKER-02]Zinker, BA, Rondinone CM, Trevillyan JM, Gum RJ, Clampit JE, Waring JF, Xie N, Wilcox D, Jacobson P, Frost L, Kroeger PE, Reilly RM, Koterski S, Opgenorth TJ, Ulrich RG, Crosby S, Butler M, Murray SF, McKay RA, Bhanot S, Monia BP, and Jirousek MR. PTP1B antisense oligonucleotide lowers PTP1B protein, normalizes blood glucose, and improves insulin sensitivity in diabetic mice. *Proc Natl Acad Sci USA* 99: 11357-11362, 2002
19. [ALAN-01]Alan R. Saltiel and C.Ronald Kahn: Insulin signaling and the regulation of glucose and lipid metabolism, , Nature, Vol. 414, 13 December 2001.
20. OSTMAN-01[]Ostman, A.; Bohmer, F. D. *Trends in Cell Biology.* 2001, *11* (6), 258-266.
21. [SALTIEL-01]Saltiel, AR, and Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* 414: 799-806, 2001
22. [TAO-01]Tao, J, Malbon CC, and Wang HY. Insulin stimulates tyrosine phosphorylation and inactivation of protein-tyrosine phosphatase 1B in vivo. *J Biol Chem* 276: 29520-29525, 2001
23. [WU-01] Wu, X, Hoffstedt J, Deeb W, Singh R, Sedkova N, Zilbering A, Zhu L, Park PK, Arner P, and Goldstein BJ. Depot-specific variation in protein-tyrosine phosphatase activities in human omental and subcutaneous adipose tissue: a potential contribution to differential insulin sensitivity. *J Clin Endocrinol Metab* 86: 5973-5980, 2001

24. [DADKE-00] Dadke, S, Kusari J, and Chernoff J. Down-regulation of insulin signaling by protein-tyrosine phosphatase 1B is mediated by an N-terminal binding region. *J Biol Chem* 275: 23642-23647, 2000
25. [KLAMAN-00]Klaman, LD, Boss O, Peroni OD, Kim JK, Martino JL, Zabolotny JM, Moghal N, Lubkin M, Kim YB, Sharpe AH, Stricker-Krongrad A, Shulman GI, Neel BG, and Kahn BB. Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol Cell Biol* 20: 5479-5489, 2000
26. [SALMEEN-00]Salmeen, A, Andersen JN, Myers MP, Tonks NK, and Barford D. Molecular basis for the dephosphorylation of the activation segment of the insulin receptor by protein tyrosine phosphatase 1B. *Mol Cell* 6: 1401-1412, 2000
27. [VINCENT-00] Vincent Aguirre, et.al: The c-Jun NH_2-terminal Kinase promotes insulin resistance during association with Insulin Receptor Substrate-1 and phosphorylation of Ser^{307}, J. Biol. Chem. Vol. 275, No. 12, 24 March 2000, pp. 9047-9054
28. [ELCHEBLY-99]Elchebly, M, Payette P, Michaliszyn E, Cromlish W, Collins S, Loy AL, Normandin D, Cheng A, Himms-Hagen J, Chan CC, Ramachandran C, Gresser MJ, Tremblay ML, and Kennedy BP. Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* 283: 1544-1548, 1999
29. [GOLDSTEIN-99]Goldstein, B. J. Diabetes Technology & Therapeutics 1999, *1* 267-275.
30. [STANDAERT-99]Standaert, ML, Bandyopadhyay G, Insulin activates protein kinases C-zeta and C-lambda by an autophosphorylation-dependent mechanism and stimulates their translocation to GLUT4 vesicles and other membrane fractions in rat adipocytes. *J Biol Chem* 274: 25308-25316, 1999
31. [WHO-99]*WHO* Fact Sheet 138: 1999
32. [BRADFORD-98]Barford, D, Das AK, and Egloff MP. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. Annu Rev Biophys Biomol Struct 27: 133-164, 1998

33. [BYON-98]Byon, J. H.; Kusari, A. B.; Kusari, J. Molecular & Cellular Biochemistry 1998, 182(1-2), 101-108.
34. [HSUEH-98]Hsueh, W. A.; Law, R. E. *Am. J. Med.* 1998, *105*(1A), 4S-14S.
35. [KAHN-98]Kahn, B. B.; Rossetti, L. *Nature Genetics* 1998, *20*(3), 223-225.
36. [LINT-97]lint, AJ, Tiganis T, Barford D, and Tonks NK. Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc Natl Acad Sci USA* 94: 1680-1685, 1997
37. [MOULE-97]Moule, SK, Welsh GI, Edgell NJ, Foulstone EJ, Proud CG, and Denton RM. Regulation of protein kinase B and glycogen synthase kinase-3 by insulin and beta-adrenergic agonists in rat epididymal fat cells. Activation of protein kinase B by wortmannin-sensitive and -insensitive mechanisms. *J Biol Chem* 272: 7713-7719, 1997
38. [POLONSKY-96]Polonsky, K. S.; Sturis, J.; Bell, G. I. New England Journal of Medicine 1996, 334(12), 777-783.
39. [FANTUS-94]Fantus, IG, Ahmad F, and Deragon G. Vanadate augments insulin-stimulated insulin receptor kinase activity and prolongs insulin action in rat adipocytes. Evidence for transduction of amplitude of signaling into duration of response. *Diabetes* 43: 375-383, 1994
40. [HURLEY-93]Hurley, TR, Hyman R, and Sefton BM. Differential effects of expression of the CD45 tyrosine protein phosphatase on the tyrosine phosphorylation of the lck, fyn, and c-src tyrosine protein kinases. *Mol Cell Biol* 13: 1651-1656, 1993
41. [MENG-93] Meng, E.C., Gschwend, D.A., Blaney, J.M. and Kuntz, I.D, Orientational sampling and rigid-body minimization in molecular docking, Proteins, 17(3): 266-278, 1993
42. [WALTON-93]K.M. Walton, J.E. Dixon. protein tyrosine phosphatases. Annu. Rev. Biochem. 62, 101 (1993).
43. [MENG-92] Meng, E.C., Shoichet, B.K. and Kuntz, I.D. Automated docking with grid-based energy evaluation. J. Comp. Chem. 13: 505-524, 1992.

44. [RAMACHANDRAN-92]Ramachandran, C, Aebersold R, Tonks NK, and Pot DA. Sequential dephosphorylation of a multiply phosphorylated insulin receptor peptide by protein tyrosine phosphatases. *Biochemistry* 31: 4232-4238, 1992
45. [SHOICHET-92] Shoichet, B.K., Bodian, D.L. and Kuntz, I.D. Molecular docking using shape descriptors. J. Comp. Chem. 13(3): 380-397, 1992.
46. [ROOME-88]Roome, J, O'Hare T, Pilch PF, and Brautigan DL. Protein phosphotyrosine phosphatase purified from the particulate fraction of human placenta dephosphorylates insulin and growth-factor receptors. *Biochem J* 256: 493-500, 1988
47. [CONNOLLY-83] Connolly, M.L: Analytical molecular surface calculation. J. Appl. Cryst. 16: 548-558, 1983.
48. [CONNOLLY-83a] Connolly, M.L.: Solvent-accessible surfaces of proteins and nucleic acids. Science, 221: 709-713, 1983.
49. [KUNTZ-82] Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E, A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161: 269-288, 1982

Appendix

SYBYL MOL2 format

This format is used for general molecule input and output of DOCK. Although previous versions of DOCK supported an extended PDB format to store molecule inform−ation, the current version now uses MOL2 as the primary molecule format. This format

has the advantage of storing all the necessary information for atom features, position, and connectivity. It is also a standardized format that other modeling programs can read.

Of the many record types in a MOL2 file, DOCK recognizes the following: MOLECULE, ATOM, BOND, SUBSTRUCTURE and SET. In the MOLECULE record, DOCK utilizes information about the molecule name and number of atoms, bonds, substructures and sets. In the ATOM record DOCK utilizes information about the atom names, types, coordinates, and partial charges. In the BOND record, DOCK utilizes the atom identifiers for the bond. In the SUBSTRUCTURE record, DOCK records the fields, but does not utilize them. The SET records are entirely optional. They are used only in special circumstances, like when ligand flexibility is considered.

Example

This example file illustrates all the elements of the MOL2 file read and written by dock. It includes optional SET records which are used by the ligand flexibility routines.

```
@<TRIPOS>MOLECULE
Histidine
20 20 1 0 2
SMALL
GASTEIGER
****
Histidine with Main Chain as Rigid Anchor


@<TRIPOS>ATOM
1    N1   -1.0947   0.5371    1.7186 N.4     1 <1> 0.2252
2    C2   -0.9885   0.9170    0.2765 C.3     1 <1> 0.0213
3    C3   -0.2043  -0.1565   -0.4766 C.3     1 <1> 0.0354
4    C4   -2.3725   1.0376   -0.3154 C.2     1 <1> 0.0897
5    O5   -2.7546   2.1336   -0.8057 O.co2   1 <1> -0.5442
6    C6    1.1797  -0.2771    0.1153 C.2     1 <1> 0.0328
7    N7    2.2791   0.4215   -0.2757 N.pl3   1 <1> -0.3074
8    C8    1.5387  -1.0911    1.1173 C.2     1 <1> 0.0462
9    C9    3.3256   0.0285    0.4990 C.2     1 <1> 0.0853
10   N10   2.9039  -0.8872    1.3511 N.2     1 <1> -0.2465
11   H11  -1.6259   1.2643    2.2287 H       1 <1> 0.2001
```

```
12   O12  -3.1452   0.0423    -0.3188 O.co2  1 <1> -0.5442
13   H13  -0.1461   0.4545     2.1242 H      1 <1> 0.2001
14   H14  -0.4726   1.8710     0.1898 H      1 <1> 0.0918
15   H15  -0.7202  -1.1105    -0.3899 H      1 <1> 0.0385
16   H16  -0.1270   0.1200    -1.5261 H      1 <1> 0.0385
17   H17   2.3126   1.1114    -1.0125 H      1 <1> 0.1528
18   H18   0.8943  -1.7774     1.6466 H      1 <1> 0.0845
19   H19   4.3357   0.4040     0.4286 H      1 <1> 0.1000
20   H20  -1.5855  -0.3703     1.8010 H      1 <1> 0.2001
@<TRIPOS>BOND
1  1  2  1
2  2  3  1
3  2  4  1
4  3  6  1
5  4  5  ar
6  6  7  1
7  6  8  2
8  7  9  1
9  8 10 1
10 9 10 2
11 1 11 1
12 4 12 ar
13 1 13 1
14 2 14 1
15 3 15 1
16 3 16 1
17 7 17 1
18 8 18 1
19 9 19 1
20 1 20 1
@<TRIPOS>SUBSTRUCTURE
1 **** 1 TEMP 0 **** **** 0 ROOT
@<TRIPOS>SET
ANCHOR STATIC ATOMS <user> **** Anchor Atom Set
1 2
RIGID STATIC BONDS <user> **** Rigid Bond Set
2 1 3
```