

**Augmentation of Hematopoietic Stem cell  
proliferation to increase the efficiency of Stem cell  
transplantation**

*A Major Project dissertation submitted*

*in partial fulfilment of the requirement for the degree of*

**Master of Technology**

**In**

**Bioinformatics**

*Submitted by*

**Ankita Nagpal**

**(2K11/BIO/01)**

**Delhi Technological University, Delhi, India**

*Under the supervision of*

**Dr Jaigopal**



Department of Biotechnology  
Delhi Technological University  
(Formerly Delhi College of Engineering)  
Shahbad Daultpur, Main Bawana Road,  
Delhi-110042, INDIA



## CERTIFICATE

This is to certify that the M. Tech. dissertation entitled “Augmentation of Hematopoietic Stem cell proliferation to increase the efficiency of Stem cell transplantation” submitted by **ANKITA NAGPAL (2K11/BIO/01)** in partial fulfilment of the requirement for the award of the degree of Master of Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate’s own work carried out by her 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 Jaigopal**

(Project Mentor)

Department of Bio-Technology

Delhi Technological University

(Formerly Delhi College of Engineering, University of Delhi)

# **DECLARATION**

I **Ms. Ankita Nagpal (2K11/BIO/01)**, student of M. Tech. (Bioinformatics), Delhi Technological University have completed the project titled “Augmentation of Hematopoietic Stem cell proliferation to increase the efficiency of Stem cell transplantation” for the award of Degree of M. Tech. (Bioinformatics), for academic session 2011-13. The information given in this project is true to the best of my knowledge.

**Ankita Nagpal**

**(2K11/BIO/01)**

# **ACKNOWLEDGEMENT**

**It gives me immense pleasure to thank all those people who have been instrumental in the completion of my project.**

**In the first place I sincerely thank Dr. R.P. Tripathi, Director of INMAS, for allowing me to do this research work. I would like to extend my vote of thanks to Dr. Rajiv Vij (Scientist 'F') without whom this project would not have been possible in INMAS.**

**I would like to express my heartfelt gratitude to my supervisor Dr. G.U. Gurudatta (Scientist 'F') for allowing me to work under his guidance and providing me with all the facilities during my project work.**

**I am deeply indebted to Mr. Pawan Kumar Raghav for his keen interest, constructive criticism, unceasing encouragement and guidance. I express my gratitude to him for sincerely helping me and boosting my morale to work hard.**

**I express my deep sense of thankfulness to Dr Jai Gopal whose enthusiastic zeal boosted me for the successful completion of my work.**

**I gratefully acknowledge Dr. Yogesh Kumar Verma (Scientist 'C'), Mr. Neeraj Satija, Mr. Siddharth Pandey and Mr Vikas for their constant support.**

**Finally, yet importantly, I would like to express my heartfelt thanks to my beloved parents, for their blessings and my friends, for their help and wishes for the successful completion of this project.**

**ANKITA NAGPAL  
2K11/BIO/01**

# CONTENTS

| TOPIC  | PAGE NO |
|--|---------|
| <i>LIST OF FIGURES</i>   | 1       |
| <i>LIST OF TABLES</i>  | 2       |
| <i>LIST OF ABBREVIATIONS</i>                                       | 3       |
| ABSTRACT   | 5       |
| 1. INTRODUCTION  | 6       |
| 2. REVIEW OF LITERATURE  | 7       |
| 2.1 Hematopoietic Stem Cells (HSC) and Transplantation             | 7       |
| 2.2 c-KIT and its role in hematopoiesis                            | 8       |
| 2.2.1 Signal transduction through c-Kit                            | 9       |
| 2.2.1.1 Negative regulation of Kit signaling by SHP-1 and SHP-2    | 11      |
| 2.3 SHP-1 and its Structure  | 12      |
| 2.3.1 Structure of SHP-1   | 12      |
| 2.3.2 Auto-inhibited Structure of SHP-1                            | 13      |
| 2.4 Protein Tyrosine Phosphatases                                  |         |
| 2.4.1 Classification of PTPs                                       | 14      |
| 2.4.2 Mechanism of Dephosphorlation of protein tyrosine phosphates | 16      |
| 3. METHODOLOGY   |         |
| 3.1 Sequence and Structure analysis                                | 18      |
| 3.2 Search for SHP1/2 inhibitors from PUBCHEM                      | 18      |
| 3.3 Dataset preparation  | 18      |
| 3.4 Pharmacophore modeling and building of 3D-QSAR model           | 18      |
| 3.4.1 Ligand preparation   | 19      |

|   |    |
|---|----|
| 3.4.2 Creating Pharmacophore Sites and Common Pharmacophore Hypothesis Generation | 20 |
| 3.4.3 Scoring Pharmacophore Hypothesis  | 20 |
| 3.4.4 Building of QSAR model  | 20 |
| 3.5 Building of new substituent's   | 21 |
| 3.6 Validation of Physiochemical properteis                                       | 21 |
| 3.7 Molecular Docking   | 21 |
| 3.8 3D QSAR of substituted ligands  | 22 |
| <b>4. RESULTS</b>   |    |
| 4.1 Sequence and structural analysis  | 23 |
| 4.2 Dataset and biological activity   | 23 |
| 4.3 Pharmacophore modelling   | 31 |
| 4.4 Building of QSAR model  | 37 |
| 4.5 Building of new substituent's   | 42 |
| 4.6 Validation of Physiochemical properteis                                       | 46 |
| 4.7 Molecular docking   | 46 |
| 4.8 Testing of substituent's using 3D QSAR model development                      | 49 |
| <b>5. DISCUSSION</b>  | 53 |
| <b>6. CONCLUSION AND FUTURE PRESPECTIVE</b>                                       | 54 |
| <b>7. REFERENCES</b>  | 56 |
| <b>8. APPENDIX</b>  | 59 |

# LIST OF FIGURES

| <b>Figures</b>   | <b>Page No</b> |
|--|----------------|
| Figure1 c-Kit expression in hematopoietic cells  | 9              |
| Figure2 Signal transduction molecules binding to the activated c-Kit receptor  | 10             |
| Figure3 Structure of SHP-1   | 12             |
| Figure 4 Classification and Substrate Specificity of PTPs  | 15             |
| Figure5 Superimposition of SHP-1 and HePTP   | 24             |
| Figure6 Common pharmacophore for active ligands  | 32             |
| Figure7 Compounds in dataset and common pharmacophore obtained   | 36             |
| Figure8 QSAR visualization of various substituent's' affect  | 38             |
| Figure9 Fitness graph between observed activity v/s PHASE predicted activity   | 39             |
| Figure10 R <sub>1</sub> ,R <sub>2</sub> ,R <sub>3</sub> R <sub>4</sub> ,R <sub>5</sub> substituents attached to common pharmacophore | 43             |
| Figure11 Fitness graph between observed activity v/s PHASE predicted activity for substituent's                                      | 50             |
| Figure12 Shows substituent's effect in compound 47   | 52             |

# LIST OF TABLES

| <b>Tables</b>   | <b>Page No</b> |
|---|----------------|
| Table1 Protein interacting with Kit phosphotyrosine residue                         | 11             |
| Table2 RMSD Values of members of non receptor tyrosine phosphatases                 | 24             |
| Table3 Binding energy of compounds with SHP-1(PDB_ID 2B30)                          | 25             |
| Table4 Shows the structure, IC <sub>50</sub> values and binding energy of compounds | 27             |
| Table5 Phase hypothesis   | 32             |
| Table6 3D-QSAR statistical parameters   | 39             |
| Table7 Fitness and PHASE predicted activity data for all compounds                  | 41             |
| Table8 Shows various substituents built.  | 44             |
| Table 9 Shows compounds with Lipinski filter application.                           | 47             |
| Table10 Shows the binding energy and interacting residues of substituent's          | 49             |
| Table11 QSAR statistical results for the model                                      | 50             |
| Table12 Phase predicted activity  | 51             |



# LIST OF ABBREVIATIONS

|                  |  |
|------------------|--|
| HSC              | Hematopoietic Stem Cells                       |
| SCF              | Stem Cell Factor                               |
| SHP-1            | Src Homology 2 Domain Containing Phosphatase 1 |
| SHP-2            | Src homology 2 Domain Containing Phosphatase 2 |
| HePTP            | Hematopoietic Tyrosine Phosphatase             |
| QSAR             | Quantitative Structure Activity Relationship   |
| PTPs             | Protein Tyrosine Phosphatases                  |
| PTKs             | Protein Tyrosine Kinases                       |
| DSPs             | Dual Specific Phosphatases                     |
| RMSD             | Root Mean Square Deviation                     |
| PTENs            | Phosphatase and Tensin homologs                |
| LMW              | Low-Molecular-Weight Phosphatases              |
| NRPTPs           | Non Receptor Tyrosine Phosphatase              |
| Cid              | Compound ID                                    |
| BE               | Binding Energy                                 |
| K <sub>i</sub>   | Inhibition Constant                            |
| IC <sub>50</sub> | Inhibitory Concentration at 50%                |
| STEP             | Striatal-Enriched Protein Tyrosine Phosphatase |
| PTP1B            | Protein Tyrosine Phosphatase 1B                |
| TCPTP            | T-cell Protein Tyrosine Phosphatase            |
| AID              | Assay ID                                       |

|                |  |
|----------------|--|
| NSC 87877      | SHP-1inhibitor   |
| ESC            | Embryonic Stem Cell  |
| LOO            | Leave One Out Cross Validation                               |
| F              | Fischer coefficient  |
| S              | Standard Deviation   |
| R <sup>2</sup> | Correlation Coefficient                                      |
| Q <sup>2</sup> | Leave one out cross validation                               |
| COMFA          | Comparative Molecular Field Analysis                         |
| COMSIA         | Comparative Molecular Similarity Indices Analysis            |
| ADMET          | Absorption, Distribution, Metabolism, Excretion and Toxicity |
| APS            | Antigen Presenting Cells                                     |
| LCMS           | Liquid Chromatography-MassSpectrometry                       |

## **ABSTRACT**

Hematopoietic Stem Cells (HSC) have capability to self renew and give rise to all blood cell types. Hematopoiesis is a process of generation of all blood cells (red, white and platelets) in bone marrow. These blood cells develop from HSC's released into blood stream along with small number of stem cells. Nowadays bone marrow transplantation (BMT) or hematopoietic stem cell transplantation is an effective way to treat cancer because chemotherapy and radiation, are toxic to the bone marrow. Therefore, it is necessary to increase the number of HSC's in blood prior reaching the bone marrow. The objective of the work is to increase the proliferation of HSC's needed during transplantation. HSC's have two main characteristics of proliferation and differentiation which is regulated by c-Kit, a protein tyrosine kinase family receptor. Binding of Stem Cell Factor to c-KIT results in receptor dimerization and activates c-KIT activity. In contrary Src homology region2 (SH-2) domain-containing phosphatase-1 (SHP-1) binds to phosphotyrosine residue 570 of c-KIT and negatively regulates proliferation of HSC's. In order to increase proliferation we have designed inhibitors against SHP-1 to inactivate enzyme activity. Till date only NSC87877 was reported to inhibit SHP-1/2, but we have found some more inhibitors which have higher binding energy with SHP-1 than NSC87877. On the basis of our screening results we have developed a pharmacophore and QSAR model and generated analog ligands. Out of which ligands which followed Lipinski rule of 5 and binding energy better than NSC87877 with SHP-1 were further tested using 3D-QSAR model development.

## 1. INTRODUCTION

Stem cells are unspecialized cells in the human body that are capable of becoming specialized cells, each with new specialized cell functions.. Stem cells can be classified into four broad types based on their origin, viz. stem cells from embryos; stem cells from the foetus; stem cells from the umbilical cord; and stem cells from the adult (Bongso *et al.*, 2005).Among the adult stem cells comes the hematopoietic stem cells. Hematopoietic stem cells have ability of haematopoiesis. Haematopoiesis is the the production and maintenance of blood stem cells and their proliferation and differentiation into the cells of peripheral blood. The hematopoietic stem cell is derived early in embryogenesis from mesoderm and becomes deposited in very specific hematopoietic sites within the embryo (Bongso *et al.*, 2005).Transplantation of stem cells from the original transplant recipient into secondary and tertiary irradiated recipients reconstitutes hematopoiesis with resultant normal life spans. Transplantation requires two essential properties proliferation to replenish the stem cell compartment (self-renewal) and lifelong production of blood (Pearce W *et al.*, 2008). During transplantation high number of HSC is needed as the cells reaching target eventually decreases. The work here aims at increasing the proliferation of the (HSC) cells needed during transplantation Signaling by stem cell factor and Kit, its receptor, plays important roles in hematopoiesis Stem cell factor exists as both a soluble and a membrane-bound glycoprotein while Kit is a receptor protein-tyrosine kinase. C-Kit consists of an extracellular domain, a transmembrane segment, a juxtamembrane segment, and a protein kinase domain that contains an insert of about 80 amino acid residues. Binding of stem cell factor to Kit results in receptor dimerization and activation of protein kinase activity. The activated receptor becomes autophosphorylated at tyrosine residues that serve as docking sites for signal transduction molecules containing SH2 domains Shp-2 phosphatase and Shp-1 phosphatase bind to the phosphotyrosine at 568 and 570 residue in juxtamembrane region of C-kit. SHP-1 binds to the phosphotyrosine residue 570 of C-kit and negatively regulate proliferation of HSC (Roskoski, Robert. 2005; Ronnstrand, L. 2004). The Ligand based drug designing approach has been used to design inhibitors against (SHP-1) to increase proliferation. HePTP hematopoietic tyrosine phosphatase (HePTP) and SHP-1 are the members of nonreceptor tyrosine phosphatase family and have similar structure. The rmsd between HePTP and SHP-1 was calculated using PYMOL and came out to be .656 Two immunoassay AID\_1077 and AID\_1059 were found where NSC87877 (only inhibitor known of SHP-1) was tested against hematopoietic tyrosine phosphatase (HePTP) .So the inhibitors taken from these assays were docked against (SHP-1) PDB\_ID (2B30).Inhibitors which showed better binding energy ,good activity and similar structure were used as a dataset for pharmacophore and QSAR studies using PHASE module of Schrodinger-9.The 3D –QSAR obtained from AAADRRR-190 hypothesis was found to be statistically good  $r^2=.89$  and  $q^2=.81$ ,Fishcer ratio=60.9,Pearson R=.94.This model was used to design 83 substituent's which were docked against (SHP-1) using autodock 4.2. The substituent's which showed best binding energy were used as test set to validate the QSAR model developed.

## 2. REVIEW OF LITERATURE

### 2.1 Hematopoietic Stem Cells (HSC) and Transplantation

Stem cells are undeveloped cells capable of proliferation, self renewal, conversion to differentiated cells, and regenerating tissues. Stem cells are unspecialized cells in the human body that are capable of becoming specialized cells, each with new specialized cell functions. The best example of a stem cell is the bone marrow stemcell that is unspecialized and able to specialize into blood cells, such as white blood cells and red blood cells, and these new cell types have special functions, such as being able to produce antibodies, act as scavengers to combat infection and transport gases (Bongso *et al.*, 2005). They serve as a repair system by being able to divide without limit to replenish other cells. When a stem cells divides, each new cell has the potential to either remain as a stem cell or become another cell type with new special functions, such as blood cells, brain cells, etc There are two main types of stem cells, embryonic and nonembryonic. Embryonic stem cells (ESC) are pluripotent because they can differentiate into all cell types; nonembryonic stem cells (non-ESC) are multipotent because their potential to differentiate into cell types is more limited (Tuch, Bernard.2006). Embryonic stem cells generate every specialized cell in the human body. Adult stem cells are located in tissues throughout the body and function as a reservoir to replace damaged or aging cells. Under physiologic conditions, adult stem cells are traditionally thought to be restricted in their differentiation to cell lineages of the organ system in which they are located.. Embryonic stem cells have great promise and versatility but, compared with adult stem cells, are currently difficult to control due to their tendency to form tumors containing all types of tissue, ie, teratomas (Pearce W *et al.*, 2008)

Stem cells can be classified into four broad types based on their origin, viz. stem cells from embryos; stem cells from the fetus; stem cells from the umbilical cord; and stem cells from the adult (Forbes S *et al.*, 2002).Among the adult stem cells comes the hematopoietic stem cells. Bone marrow possesses stem cells that are hematopoietic and mesenchymal in origin. The hematopoietic stem cell is derived early in embryogenesis from mesodermal and becomes deposited in very specific hematopoietic sites within the embryo. These sites include the bone marrow, liver, and yolk sac. Hematopoietic stem cells have ability of haematopoiesis. Haematopoiesis is the the production and maintenance of blood stem cells and their proliferation and differentiation into the cells of peripheral blood. All of the mature blood cells in the body are generated from a relatively small number of hematopoietic stem cells (HSCs) and progenitors (Smith,Clayton.2002). HSCs are able to generate every lineage found in the hematopoietic system including red blood cells, platelets, and a variety of lymphoid and myeloid cells. Some of the most important lymphoid cells include natural killer (NK) cells,T cells, and B cells, while important myeloid cells include granulocytes, monocytes, macrophages, microglial cells, and dendritic cells. Each of these cell types can be generated from a single HSC, and each HSC has an enormous capacity to generate large numbers of these cells over many years and perhaps even decades (Smith, C. 2002).

Transplantation of stem cells from the original transplant recipient into secondary and tertiary irradiated recipients reconstitutes hematopoiesis with resultant normal life spans. Transplantation requires two essential properties proliferation to replenish the stem cell compartment (self-renewal) and lifelong production of blood (Pearce W *et al.*, 2008). Hematopoietic stem cells (HSCs) play a key role in hematopoietic system that functions mainly in homeostasis and immune response. HSCs transplantation has been applied for the treatment of several diseases. However, HSCs persist in the small quantity within the body, mostly in the quiescent state. HSC maintenance, balance between self-renewal and proliferation are essential requirement for advancement of HSC expansion and transplantation in the future. Hematopoiesis and HSC development are the key role to improve efficient HSC expansion for the transplantations (Chotinantakul *et al.*, 2012).

## **2.2 c-KIT and its role in hematopoiesis.**

Stem Cell Factor (also known as SCF, kit-ligand, KL, or steel factor) is a cytokine that binds to the c-Kit receptor (CD117). SCF can exist both as a transmembrane protein and a soluble protein. This cytokine plays an important role in hematopoiesis. SCF plays an important role in the hematopoiesis during embryonic development. Sites where hematopoiesis takes place, such as the fetal liver and bone marrow, all express SCF (Gali *et al.*, 1994). SCF has been shown to increase the survival of HSCs in vitro and contributes to the self renewal and maintenance of HSCs in-vivo. HSCs at all stages of development express the same levels of the receptor for SCF (c-Kit) (Kent *et al.*, 2008).

SCF binds to the c-Kit receptor (CD 117), a receptor tyrosine kinase. c-Kit is expressed in HSCs, mast cells, melanocytes, and germ cells. It is also expressed in hematopoietic progenitor cells including erythroblasts, myeloblasts, and megakaryocytes. However, with the exception of mast cells, expression decreases as these hematopoietic cells mature and c-Kit is not present when these cells are fully differentiated. Fig 1 shows the expression of C-kit in hematopoietic cells. Signaling by stem cell factor and Kit, its receptor, plays important roles in gametogenesis, hematopoiesis, mast cell development and function, and melanogenesis (Roskoski, R. 2005; Ronnstrand, L. 2004) Stem cell factor exists as both a soluble and a membrane-bound glycoprotein while Kit is a receptor protein-tyrosine kinase. Kit consists of an extracellular domain, a transmembrane segment, a juxtamembrane segment, and a protein kinase domain that contains an insert of about 80 amino acid residues. Binding of stem cell factor to Kit results in receptor dimerization and activation of protein kinase activity. The activated receptor becomes autophosphorylated at tyrosine residues that serve as docking sites for signal transduction molecules containing SH2 domains. The adaptor protein APS, Src family kinases, and **Shp2** tyrosyl phosphatase bind to phosphotyrosine 568. **Shp1** tyrosyl phosphatase and the adaptor protein Shc bind to phosphotyrosine 570. These residues occur in the juxtamembrane segment of Kit (Ronnstrand, L. 2000)

## Kit Expression in Hematopoietic Cells

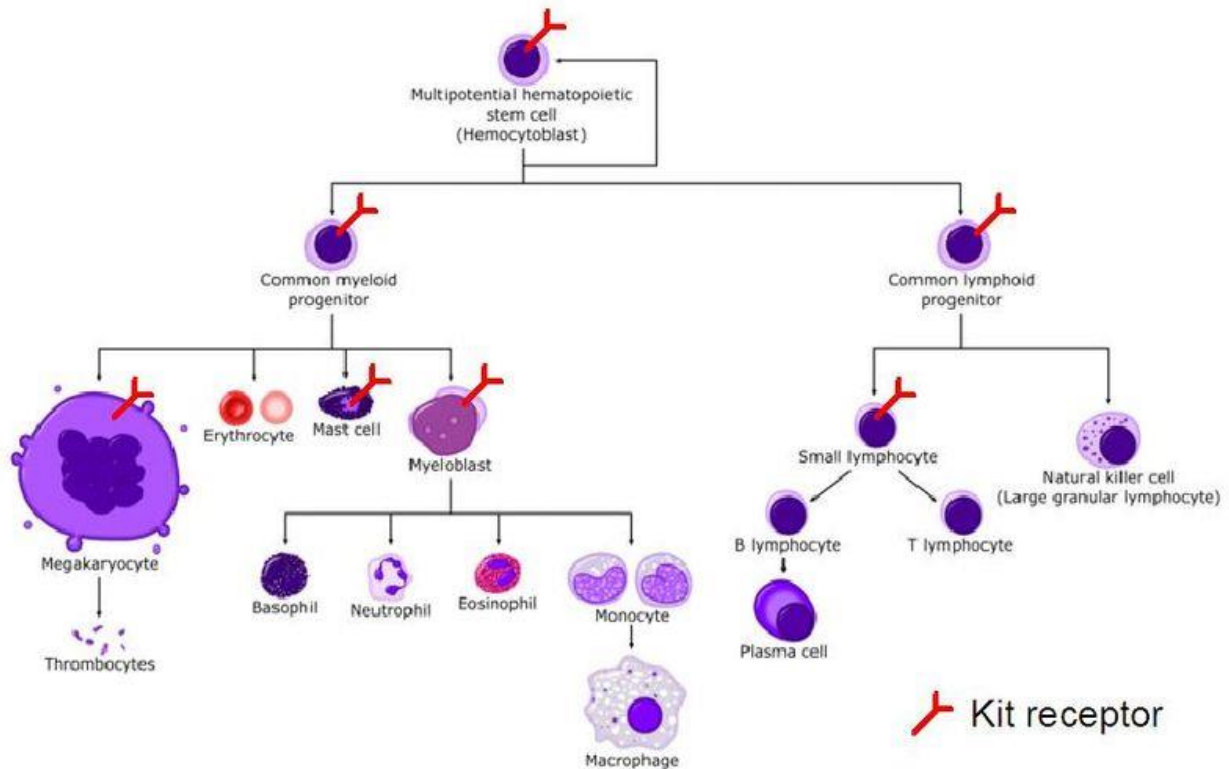


Figure 1 c-Kit expression in hematopoietic cells (Ronnstrand, L. 2004)

### 2.2.1 Signal transduction through c-Kit

Stem cell factor exists as both a soluble and a membrane-bound glycoprotein while Kit is a receptor protein-tyrosine kinase. The complete absence of stem cell factor or Kit is lethal. Deficiencies of either produce defects in red and white blood cell production, hypopigmentation, and sterility. Gain-of-function mutations of Kit are associated with several human neoplasms including acute myelogenous leukemia, gastrointestinal stromal tumors, and mastocytomas.

Kit consists of an extracellular domain, a transmembrane segment, a juxtamembrane segment, and a protein kinase domain that contains an insert of about 80 amino acid residues. Binding of stem cell factor to Kit results in receptor dimerization and activation of protein kinase activity. The activated receptor becomes autophosphorylated at tyrosine residues that serve as docking sites for signal transduction molecules containing SH2 domains. The adaptor protein APS, Src family kinases, and **Shp2** tyrosyl phosphatase bind to phosphotyrosine **568**. **Shp1** tyrosyl phosphatase and the adaptor protein Shc bind to phosphotyrosine **570**. C-terminal Src kinase homologous kinase and the adaptor Shc bind to both phosphotyrosines 568 and 570. These residues occur in the juxtamembrane segment of Kit. Three residues in the kinase insert domain are phosphorylated and attract the adaptor protein Grb2 (Tyr703), phosphatidylinositol 3-kinase (Tyr721), and phospholipase C (Tyr730). Phosphotyrosine 900 in the distal kinase domain binds phosphatidylinositol 3-kinase which in turn binds the adaptor protein Crk. Phosphotyrosine 936,

also in the distal kinase domain, binds the adaptor proteins APS, Grb2, and Grb7. Kit has the potential to participate in multiple signal transduction pathways as a result of interaction with several enzymes and adaptor proteins (Roskoski, Robert. 2005).

Various proteins interact with c-Kit phosphotyrosine residues. Table 1 (Roskoski, Robert. 2005) shows the binding site of these proteins on c-Kit, their function and their biological effect.

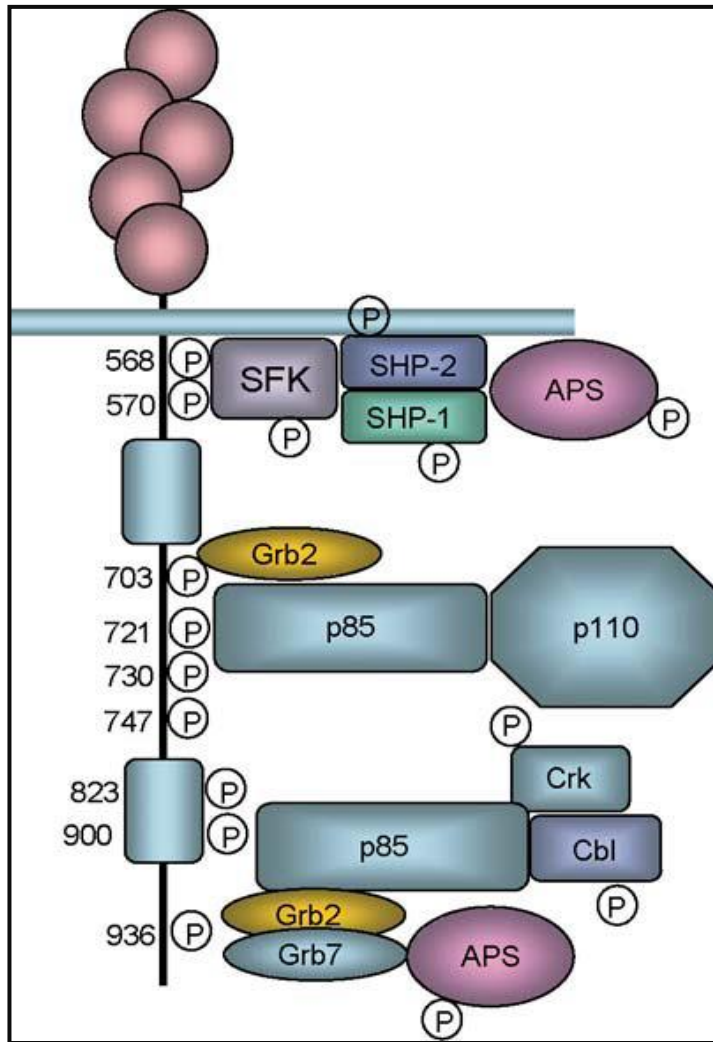


Figure 2 Signal transduction molecules binding to the activated c-Kit receptor. Upon ligand binding, c-Kit dimerizes, and its intrinsic tyrosine kinase activity is activated, leading to phosphorylation of key residues. These residues constitute high-affinity binding sites for signal transduction molecules. The numbers refer to tyrosine residues phosphorylated in c-Kit, and the corresponding signal transduction molecule is depicted. (Roskoski, Robert. 2005)

As seen from Figure 2 various proteins interact with c-Kit phosphotyrosine residues. Table 1 (Roskoski, Robert. 2005) shows the binding site of these proteins on c-Kit, their function and their biological effect.



Table1 Protein interacting with Kit phosphotyrosine residue

| Human residue | Protein    | Function  | Biological effect      |
|---------------|------------|---|------------------------|
| 568           | APS        | Adaptor containing PH andSH <sub>2</sub> domain                               | Kit degradation        |
| 568           | SHP-2      | SH <sub>2</sub> domain containing phosphatase2                                | Inhibits Proliferation |
| 568           | Src Kinase | Non receptor protein tyrosine kinase family                                   | Proliferation          |
| 568,570       | Chk        | Csk homology kinase   | Not reported           |
| 568,570       | Shc        | SH <sub>2</sub> containing transformingproteinC1;bindsGrb2                    | Not reported           |
| 570           | SHP-1`     | SH <sub>2</sub> domain containing phosphatase1                                | Inhibits Proliferation |
| 703           | Grb2       | Growth factor receptor-bound protein2   | Not reported           |
| 721           | P85        | Regulatory subunit of PI-3 kinase   | Survival,proliferation |
| 730           | Plcγ       | Phospholipase Cγ  | Survival,proliferation |
| 823           |            | Kit activation loop tyrosine phosphorylation site                             | Proliferation          |
| 900           | P85        | Regulatory subunit of PI-3 Kinase   | Proliferation          |
| 900           | Crk        | Adaptor protein that contains one SH <sub>2</sub> ,two SH <sub>3</sub> domain | Proliferation          |
| 936           | APS        | Adaptor containing PH and SH <sub>2</sub> domains                             | Kit degradation        |
| 936           | Grb2       | Growth factor receptor bound protein2   | Not reported           |
| 936           | Grb7       | Growth factor receptor bound protein7   | Not reported           |

### 2.2.1.1 Negative regulation of c-Kit signaling by SHP-1 phosphatase and SHP-2 phosphatase

Shp1 is a cytosolic phosphotyrosyl phosphatase containing two tandem SH2 domains, a phosphatase domain and a C-terminal tail. Shp1 occurs primarily in hematopoietic and epithelial cells, and it is a negative regulator of growth factor signaling. Besides inhibiting Kit signaling, Shp1 diminishes the growth-promoting properties of the colony-stimulating factor 1, erythropoietin, and interleukin 3 receptors, an effect mediated either directly by receptor dephosphorylation or indirectly dephosphorylation of receptor-associated protein-tyrosine kinases. Shp1 plays a role in the control of signaling cascades that couple growth factor receptors to hemopoietic cell differentiation. The N-terminal-SH2 domain of Shp1 blocks its catalytic domain and keeps the enzyme in an inactive conformation (Kozłowski *et al.*, 1998). One plausible notion for enzyme regulation involves the recruitment and binding of this Shp1 SH2 domain to target phosphotyrosine residues with concomitant phosphatase activation

Shp2, like Shp1, is a cytosolic phosphotyrosyl phosphatase containing two tandem SH2 domains, a phosphatase domain, and a C-terminal tail. Shp2, in contrast to Shp1, occurs in many types of cells. The SH2 domains of Shp2 target this enzyme to phosphotyrosines in a variety of growth factor receptors and other signaling molecules. Thus, Shp1 and Shp2 can negatively modulate Kit signaling by interacting with these specific phosphotyrosine residues(Kozłowski *et al.*, 1998).

## 2.3 SHP-1 and its Structure

SHP-1 belongs to the family of non-receptor protein tyrosine phosphatases (PTPs) and generally acts as a negative regulator in a variety of cellular signaling pathways. SHP-1 is predominantly expressed in hematopoietic and epithelial cells and behaves mainly as a negative regulator of signaling pathways in lymphocytes (Yang *et al.*, 2002).

The crystal structure of SHP-1 shows that it consists of three domains. The residues 1–108 and 116–208 fold as two Src homology 2 domains, the N-SH2 and C-SH2 domains, respectively. Residues 270–532 fold as the typical PTP domain, a highly twisted ten-stranded  $\beta$  sheet flanked by four helices on the convex side and two helices and a  $\beta$  hairpin from the concave side. The architecture of the three domains is compact. The two SH2 domains look like two antennas of the global PTP domain in the overall view. Crystal structure revealed that both SH2 domains of SHP-1 have the typical SH2 domain fold, which consists of a central four-stranded  $\beta$ -sheet with an  $\alpha$  helix on either side. The phosphopeptide-binding sites of both SH2 domains face away from the PTP domain and are fully exposed on the surface of the molecule. Nevertheless, the spatial arrangement of the two SH2 domains on the PTP domain significantly differs. In contrast to the N-SH2 domain that strongly interacts with the PTP domain, the C-SH2 domain is tethered around and extends to the surface of the catalytic domain and has no significant direct interactions with the PTP domain (Yang *et al.*, 2003; Wang *et al.*, 1999).

### 2.3.1 Structure of SHP-1

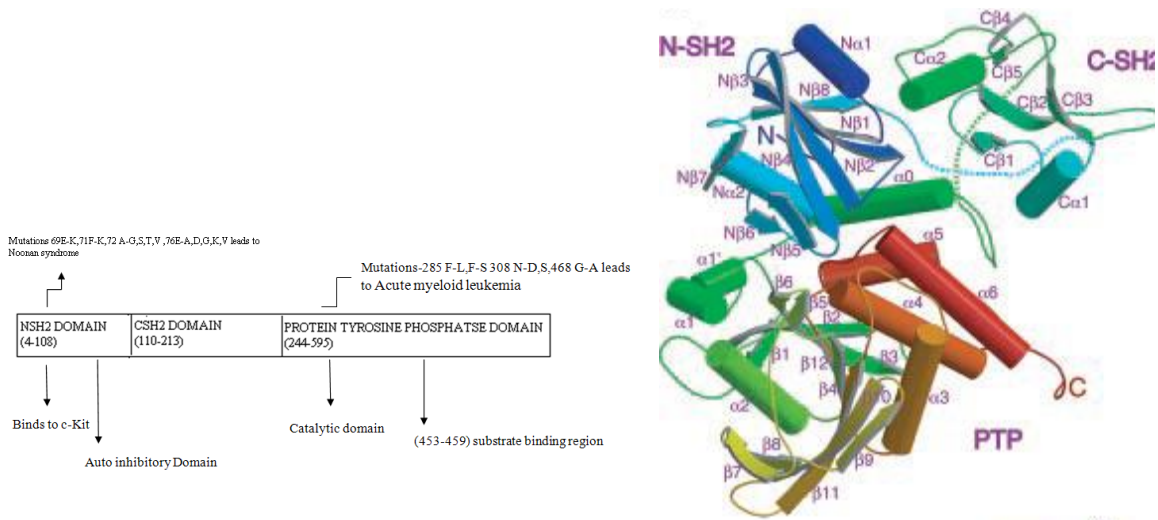


Figure 3 Structure of SHP-1. A schematic drawing of structure with color ramping from blue (N terminal) to red (C terminal)

### 2.3.2 Auto-inhibited Structure of SHP-1

In the absence of ligands for the tandem SH2 domain of these PTPs the more N-terminal SH2 domain folds onto the catalytic domain to block substrate access by the insertion of a loop on the backside of the SH2 domain into the catalytic pocket of the PTP. When the SH2 domains of this inhibited form of SHP1 encounter a tyrosine phosphorylated ligand, the closed conformation opens and the enzyme is activated some 100-fold. Under physiological conditions, SH2 domain ligand binding also juxtaposes the PTP domain to its substrates, which contain, or associate with, the phosphorylated SH2 domain ligand(Yang *et al.*, 2003).

Previous studies have revealed that both SH2 domains of SHP-1 could bind to tyrosine phosphorylated immunoreceptor tyrosine-based inhibitory motif peptides. However, the crystal structure of the ligand-free SHP-1 supports that the N-SH2 domain, instead of C-SH2 or both domains, acts the auto-inhibition role. The interaction between N-SH2 domain and PTP domain is extensive, whereas the C-SH2 domain does not show significant interface with either of the other two domains. Cys455, the catalytic nucleophile, is located at the base of the active-site cleft. In the auto-inhibited conformation of SHP-1, it appears that the N $\beta$ 4-N $\beta$ 5 loop of the N-SH2 domain is protruding to the catalytic PTP domain to directly block the entrance to the active site, which prevents the cysteine residue from exposing to the substrate. This inactive conformation is stabilized by various interactions including the salt bridge between Asp61 and Lys362, and the hydrogen bonds in the residue pairs of Ser59/His422, Gly60/ Gln506, and Asn58/Gln502. In addition, the extensive interactions around the protruding N $\beta$ 4-N $\beta$ 5 loop are present at the interface between the N-SH2 domain and the PTP domain(Yang *et al.*,1998;2013).

### 2.3.3 Activation Mechanism of SHP-1

The highly mobile C-SH2 domain functions as an antenna to search for phosphopeptides. Binding of phosphopeptide to the C-SH2 domain results in large conformational changes that restore the distorted conformation of the neighboring N-SH2 domain and subsequently opens up its phosphopeptide-binding pocket to harbor a second phosphopeptide molecule. These events can weaken the auto-inhibiting interaction on the interface between the N-SH2 and PTP domains and permit the subsequent synergistic opening up of the active site of the PTP domain. This mechanism is consistent with the notion that a truncated SHP-1 lacking the C-SH2 domain would be activated to a much lesser extent than full-length SHP-1(Yang *et al.*, 2003). . It also is envisioned that a much larger change in the relative positions of the two SH2 domains will occur due to the mobility of the C-SH2 domain when they are simultaneously bound by biphosphorylated peptides, which can lead to greater movement of the N-SH2 domain. This movement should optimize the opening up of the active site of the PTP domain and can qualitatively explain why biphosphorylated peptides can activate SHPs at a 10-fold higher level than monophosphorylated peptides(Yang *et al.*, 2003;Wang *et al.*,1999).

## 2.4 Protein Tyrosine Phosphatases

A cornerstone of many cell-signalling events rests on reversible phosphorylation of tyrosine residues on proteins. The reversibility relies on the co-ordinated actions of protein tyrosine kinases and protein tyrosine phosphatases (PTPs), both of which exist as large protein families (Stoker *et al.*, 2005). PTPs regulate a wide range of signalling pathways. PTPs work antagonistically with Protein Tyrosine Kinases (PTKs) and inhibit cell proliferation. PTPs are a group of enzymes that remove phosphate groups from phosphorylated tyrosine residues on proteins. Phosphorylation of proteins is one of the posttranslational modifications, which is reversible and plays a critical role in the regulation of many cellular functions. As a consequence, maintaining an appropriate level of protein tyrosine phosphorylation is essential for many cellular functions (Anderson *et al.*, 2001). Tyrosine-specific protein phosphatases catalyse the removal of a phosphate group attached to a tyrosine residue. These enzymes are key regulatory components in signal transduction pathways cell cycle control, and are important in the control of cell growth, proliferation, differentiation and transformation. PTPs have been implicated in regulation of many cellular processes, such as cell growth, cellular differentiation, Mitotic cycles, Oncogenic transformation (Anderson *et al.*, 2001).

All PTPases carry the highly conserved active site motif C(X)5R (PTP signature motif), employ a common catalytic mechanism, and possess a similar core structure made of a central parallel beta-sheet with flanking alpha-helices containing a beta-loop-alpha-loop that encompasses the PTP signature motif (Dixon *et al.*, 1998)

Individual PTPs may be expressed by all cell types, or their expression may be strictly tissue-specific. Most cells express 30% to 60% of all the PTPs, however hematopoietic and neuronal cells express a higher number of PTPs in comparison to other cell types. T cells and B cells of hematopoietic origin express around 60 to 70 different PTPs. The expression of several PTPs is restricted to hematopoietic cells, for example, LYP, SHP1, CD45, and HePTP (Alonso *et al.*, 2004).

Of the 107 PTP genes, 11 are catalytically inactive 2 dephosphorylate mRNA and 13 dephosphorylate inositol phospholipids. Thus, 81 PTPs are active protein phosphatases with the ability to dephosphorylate phosphotyrosine (Alonso *et al.*, 2004).

### 2.4.1 Classification of PTPs

The **class I** PTPs, are the largest group of PTPs with 99 members, which can be further subdivided into 38 classical PTPs and 61 VH-1-like or dual-specific phosphatases (DSPs) Classical PTPs can be further divided into 21 receptor tyrosine phosphatase and 17 nonreceptor-type PTPs. Dual-specific phosphatases (DSPs) can be further divided into 11 MAPK phosphatases (MPKs), 3 Slingshots, 3 PRLs, 4 CDC14s, 19 atypical DSPs, 5 Phosphatase and tensin homologs (PTENs), 16 Myotubularins

Dual-specificity phosphatases (dTyr and dSer/dThr) dual-specificity protein-tyrosine phosphatases. Ser/Thr and Tyr dual-specificity phosphatases are a group of enzymes with both Ser/Thr and tyrosine-specific protein phosphatase activity able to remove the serine/threonine or the tyrosine-bound phosphate group from a wide range of phosphoproteins, including a number of enzymes that have been phosphorylated under the action of a kinase. Dual-specificity protein phosphatases (DSPs) regulate mitogenic signal transduction and control the cell cycle, **Class II** LMW (low-molecular-weight) phosphatases, or acidphosphatases, act on tyrosine phosphorylated proteins, low-MW aryl phosphates and natural and synthetic acyl phosphates. The class II PTPs contain only one member, low-molecular-weight phosphotyrosine phosphatase (LMPTP). The **Class III** Cdc25 phosphatases (dTyr and/or dThr) PTPs contains three members, CDC25 A, B, and C. The **Class IV** pTyr-specific phosphatases contains four members, Eya1-4

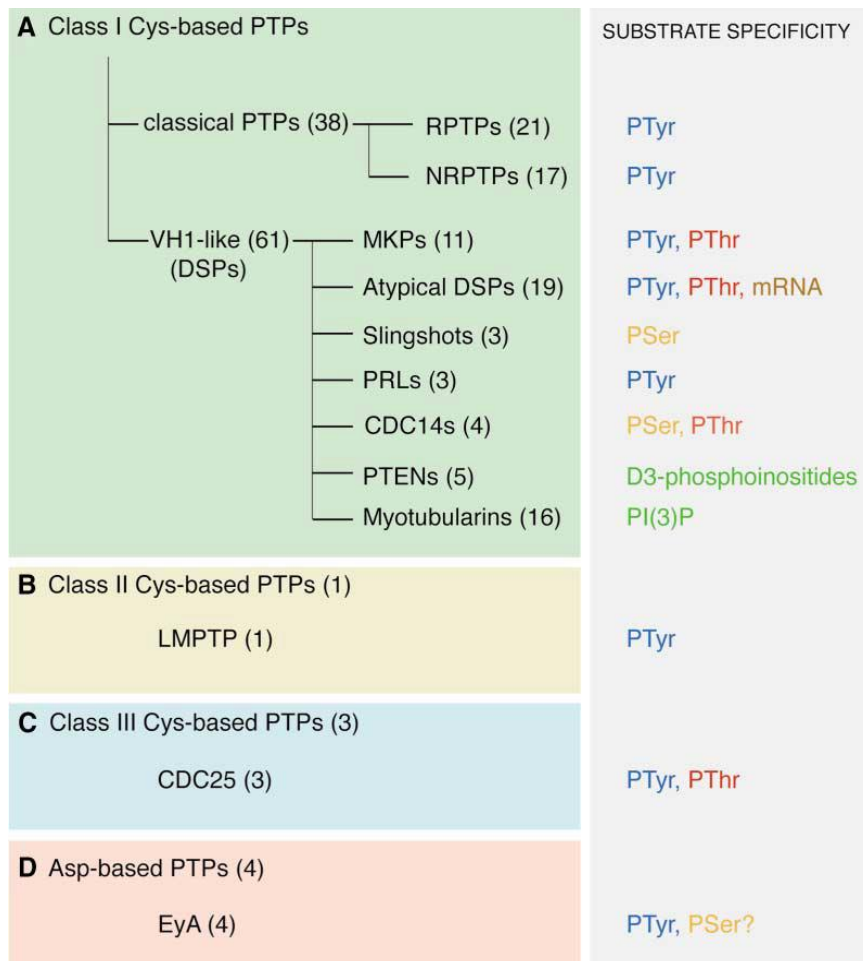


Figure 4 Classification and Substrate Specificity of PTPs .The PTP families are color coded: class I Cysbased PTPs (green), class II Cys-based PTPs(pale yellow), class III Cys-based PTPs (pale blue), and Asp-based PTPs (pink). The substrate specificity of each group or class of PTPs is listed (Alonso *et al.*,2004).

## 2.4.2 Mechanism of Dephosphorylation of protein tyrosine phosphatases

The capacity of PTPs to dephosphorylate phosphotyrosine residues selectively on their substrates plays a pivotal role in initiating, sustaining and terminating cellular signaling. As seen from figure 3 PTPs can be grouped into two types. The first group is the tyrosine-specific PTPs that dephosphorylate protein substrates on tyrosine. Tyrosine-specific PTPs comprise receptor-like PTPs and non-transmembrane PTPs. The second group is the DSPs (dual-specificity phosphatases) that dephosphorylate protein substrates on tyrosine, serine and threonine residues, as well as lipid substrates (Anderson *et al.*, 2001).

PTPs consist of non catalytic N- and C-terminal segments and the catalytic domain. The catalytic domains of classical PTPs contain approx. 280 residues and comprise 22 invariant and 42 highly conserved residues that fall within ten consensus motifs. It has been shown that both the catalytic domain and noncatalytic segments of the PTPs contribute to the substrate specificity *in vivo* (Tigani *et al.*,2007). Whereas noncatalytic domains may target the PTPs to specific intracellular compartments in which the effective local concentration of substrate is high the PTP catalytic domains themselves confer site-selective protein dephosphorylation by recognizing both the phosphotyrosine residue to be dephosphorylated and its flanking amino acids in the substrate (Tigani *et al.*,2007).

A structural feature that is highly conserved among PTPs is the catalytic, or PTP loop (also known as the signature motif). This PTP loop comprises of the sequence HC(X5)R or(I/V)HCXAGXXR(S/T)G which defines the PTP family and this sequence is referred to as the 'PTP signature motif'. Another conserved loop, the recognition loop, plays an important role in substrate recognition. The residues Val and Tyr assist the substrate's insertion into catalytic site Ser of the PTP loop forms a hydrogen bond with the the recognition loop, stabilizing the active site cleft. A third conserved loop is the WPD loop. The cysteine and arginine residues in the PTP signature motif are essential for catalytic activity. The cysteine residue acts in the first step of catalysis wherein the sulfur atom of the thiolate group serves as a nucleophile and attacks the substrate phosphate (Jia *et a.*,1995).The arginine residue contributes to substrate binding and stabilizes the cysteine-phosphate intermediate. Another important motif integral to PTP catalysis is the WPD (Trp-Pro-Asp) loop, which becomes displaced by several angstroms [8–12 Å (1 Å=0.1 nm)] upon substrate binding and closes around the side chain of the phosphotyrosine residue. This conformational change positions the invariant aspartate residue (Asp181 in PTP1B) in the WPD loop in a position that allows it to act as a general acid for the first step of catalysis. This step involves protonating the phenolic oxygen of the tyrosyl leaving group, thus **cleaving the phosphate off tyrosine**, to form the cysteine-phosphate intermediate. This same aspartate residue also acts as a general base in the second step of catalysis, which, together with a highly conserved glutamine residue (Gln262 in PTP1B), co-ordinates an essential water molecule to promote the hydrolysis of the cysteine-phosphate intermediate (Tigani *et al.*,2007; Pannifer *et al.*,1998).

The nonreceptor transmembrane protein tyrosine phosphatases (NRPTPs) includes various members like PTPN6/**SHP-1**, PTPN7/**HePTP**, PTPN11/**SHP-2**(Alonso *et al.*, 2004)

Hematopoietic tyrosine phosphatase (HePTP) a 38 kDa phosphatase another member of non receptor transmembrane protein tyrosine phosphatases (NRPTPs) as is SHP-1 . it is strongly expressed in T cells. It is composed of a C-terminal classical PTP domain (residues 44–339) and a short N-terminal extension (residues 1–43) that functions to direct HePTP to its physiological substrates (Mustelin *et al.*, 2005). HePTP shares similar structure with SHP-1. The rmsd value calculated between HePTP and SHP-1 using pymol came out to be 0 .656. It can be seen that the residues and the structures which is responsible for the inhibitor binding is conserved in both Heptp and shp-1.so inhibitors (including NSC87877 only inhibitor in PUBCHEM for SHP-1) against HePTP were used for pharmacophore and QSAR studies.

### **3. MATERIALS AND METHODS**

#### **3.1 Sequence and Structure analysis**

Various members of nonreceptor tyrosine phosphatases were superimposed using Pymol. PTPN1, PTPN2, PTPN9, PTPN5, PTPN7/HePTP were superimposed to SHP1/PTPN6 and SHP-2/PTPN11. This superimposition was done to calculate structure deviation between various members of non receptor tyrosine phosphatase on the basis of RMSD calculated using Pymol.

#### **3.2 Search for SHP1/2 inhibitors from PUBCHEM**

SHP-1 inhibitors were searched from PUBCHEM. Only NSC87877 (CID 16654632) have been reported as potential inhibitor of SHP-1 in PUBCHEM (<http://pubchem.ncbi.nlm.nih.gov>). Two assays from PUBCHEM i.e. AID\_1077 and AID\_1059, have been identified wherein NSC87877 was tested against HePTP. However the IC<sub>50</sub> values of NSC87877 was repeated for HePTP in two different assays ids (AID 1077 and AID 1059). In AID\_1077, 147 compounds and in AID\_1059, 57 compounds have been reported. These 57 were also common in AID\_1077. So (147 compounds) found were used for the preparation of Dataset.

#### **3.3 Dataset preparation**

The Protein SHP-1 (PDB\_ID: 2B30) was downloaded from the RCSB PDB site. The inhibitors were collected from PUBCHEM in SDF format. Pyrx ) was used to screen the ligands which showed comparable and better binding energy than NSC87877 (cid 16654632) against SHP-1 (PDB\_ID: 2B30). The inhibitors which showed better binding energy were further selected on the basis of similar structure and activity to build the dataset. So the inhibitors which showed better binding energy than NSC8787, belonged to congeneric series (similar in structure) and had good activity were selected to be the members of the dataset that was used for pharmacophore and QSAR studies. The clustering feature of pubchem was used for building this congeneric series. Finally a dataset of 24 compounds was built which was further used for QSAR studies.

#### **3.4 Pharmacophore modeling and building of 3D-QSAR model**

A successful 3D-QSAR study was performed to establish relationship between the spatial three-dimensional pharmacophoric features and inhibition activity of a class of inhibitors in the dataset. Present 3D-QSAR study was performed with the dataset of 24 compounds (including NSC87877) with well defined inhibitory activity given as IC<sub>50</sub> values in  $\mu$ M concentration. For the correlation purpose IC<sub>50</sub> values then converted to their molar values and subsequently calculated to free energy-related terms, i.e.,  $-\log (1/IC_{50})$ . This dataset was then chosen for generating common pharmacophore hypotheses and then performing QSAR analysis.

PHASE-3.1 module of Maestro-9 (Phase 3.1, Schrödinger, LLC, 2009) molecular modeling software was used to generate 3D pharmacophore models for selected series of inhibitors. A pharmacophore conveys characteristics of the three-dimensional arrangement of the



pharmacophoric elements which are important to be critical for binding. A given hypothesis may be combined with known activity data to create a 3D-QSAR model that identifies overall aspects of molecular structure that govern activity.

#### **3.4.1 Ligand preparation**

The structures were sketched using maestro builder toolbar and were imported to develop pharmacophore model panel of the PHASE with their respective activity values. The ligands were assigned as actives and inactives by giving an appropriate activity threshold value 5.6. The activity threshold value was selected in the basis of dataset activity distribution and active ligands are chosen to derive a set of suitable pharmacophores. Sketched structures were energy minimized/cleaned up by Ligprep module using OPLS\_2005 force field (LigPrep, Schrödinger, LLC, 2009) and proper protonation states were assigned with the ionizer subprogram at pH  $7.2 \pm 0.2$ . Conformation generation is an important step in PHASE. The conformations were generated with the help of ConfGen method taking GB/SA solvent model using OPLS\_2005 (MacroModel, Schrödinger, LLC, 2009) force field. About 1000 conformers were generated per structure ensuring 50 step minimization. The minimized conformers were filtered using a relative energy parameter limitation of 10kcal/mol and a minimum atom deviation of 1Å. Thus lowest energy non-redundant conformers of a ligand were used for pharmacophore model development. A couple of conformer was defined as identical if the relative distance between them is below 1Å.

#### **3.4.2 Creating Pharmacophore Sites and Common Pharmacophore Hypothesis Generation**

According to the bioactivity, the molecules were divided into actives and inactives, setting the maximum and minimum values in the activity threshold window of PHASE. Pharmacophore sites of a ligand are represented in the 3D space by a set of points. These points coincide with various chemical characteristics with type, location, and directionality, which facilitate non-covalent bonding with the receptor sites. The pharmacophore features like hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic/Non-polar group (H), negatively ionizable (N), positively ionizable (P), and aromatic ring (R) were used to create the pharmacophore sites for the energy-calculated ligands. Tree-based partition algorithm is used by PHASE for detection of common pharmacophore from a set of variants taking maximum tree depth 3. To find common pharmacophore, PHASE algorithm use an exhaustive analysis of k-point pharmacophore match picked from the conformations of a set of active ligands on the basis of inter site distances, and then find all spatial arrangements of pharmacophore features those are common to at least 8 of 10 active ligands. The generated pharmacophores match different set of actives eliminating the chance of its exclusiveness toward a small subset of ligands. The different pharmacophore hypotheses were further examined by using a scoring function, so that it produced the best alignment of the ligands.

### 3.4.3 Scoring Pharmacophore Hypothesis

The scoring of the pharmacophore hypotheses was done in relation to the information from the active ligands considering various geometric and heuristic factors. The alignment to a reference pharmacophore is considered according to RMSD of the site points and the average cosine of the vectors keeping their tolerance 1.2 Å and 0.5, respectively. To get the reference ligand from the most active set, upper 10% was considered for score calculation. For further refinement, volume scoring was performed in order to measure quantitatively of how each non-reference ligand is superimposing with the reference ligand. Here, the cutoff for volume scoring was kept at 1.00 for the non-reference pharmacophores.

The resulting pharmacophore was then scored and ranked. The scoring was done to identify the best candidate hypothesis, and which provided an overall ranking of all the hypotheses. The scoring algorithm included the contributions from the alignment of site points and vectors, volume overlap, selectivity, number of ligands matched, relative conformational energy, and activity. Among which best hypothesis AAADRRR.190 was selected on the basis of score and discrimination of active and nonactive molecules i.e if active molecules score well, the hypothesis could be invalid as it does not discriminate between active and inactive.

### 3.4.4 Building of QSAR model

To produce a statistically significant 3D-QSAR model, the first and the foremost requirement is the alignment of ligands; therefore, to execute the QSAR study, a pharmacophore-based alignment was considered. The PHASE algorithm uses a very flexible approach for the development of 3D QSAR model. It considers a rectangular grid of 1 Å grid distance in a 3D space. Thus, it creates cubes of said dimension in the 3D space. The atoms of the molecules which are considered as overlapping Vander Waal spheres fall inside these cubes depending on the volume of the atomic spheres. These occupied cube spaces are termed as volume bits. A volume bit is allocated for each different class of atom that occupies a cube. There are six atom classes, viz. two hydrogen bond acceptor (A), one positively ionizable (P), and two aromatic ring (R) used for classifying the atom characteristics. The total number of volume bits consigned to a specified cube is based on how many training set molecules occupy that cube. A single cube may represent the occupation by one or various atoms or sites, and even those from the same molecule or may be from unlike molecules of the training set. Thus, a molecule may be represented by a binary string concurrent to the occupied cubes, and also the various types of atomic sites that exist in those cubes. To create an atom-based QSAR model, these volume bits which encode the geometrics and chemical characteristics of the molecule are regarded as independent variables in PLS (Partial Least square) regression analysis. The maximum PLS factor that can be taken is  $N/5$ , where  $N$  is the number of ligands present in the training set. In this study, a significant 3D-QSAR model was generated using AAADRRR-190 hypothesis. For QSAR model generation, training and test partition was done by random selection method. Atom-based model selection criterion was chosen for model building. PLS factor was set as 03, the maximum number of PLS factors in each model can be  $1/5$  the total number of training set

molecules. More the PLS factor value, more will be the reliability of models. Various models have been generated and the best model was selected on the basis of the statistical significance.

### **3.5 Building of new substituent's**

The QSAR model generated (which is statistically significant) was used to design novel compounds with better biological activity. The 3D –QSAR model built explains how and at what extent electron withdrawing, hydrophobic and H- donor properties should be modified to achieve better biological activity of inhibitors. The QSAR model generated shows blue regions where electron withdrawing, hydrophobic, and H –donor groups can be added to common pharmacophore and could lead to better biological activity. Also the red regions shows the sites where addition of these groups is prohibited for the inhibitors biological activity. On the basis of the QSAR model developed about 83 substituent's were build using permutation and combination with the help of Chem draw(ChemDraw Std 13.0 Suite).

### **3.6 Validation of Physiochemical Properties**

Lipinski rule of 5 was used for drug designing was used. it states that that most "drug-like" molecules have  $\log P \leq 5$ , molecular weight  $\leq 500$ , number of hydrogen bond acceptors  $\leq 10$ , and number of hydrogen bond donors  $\leq 5$ . Molecules violating more than one of these rules may have problems with bioavailability. The rule is called "Rule of 5", because the border values are 5, 500,  $2*5$ , and 5. So Lipinski filter was applied to all the new substituent's formed using molinformatics server.

### **3.7 Molecular docking**

The ligands which passed the Lipinski filter were converted into pdb format using Chemdraw(ChemDraw Std 13.0 Suite). Autodock 4.2(Sotriffer *et al.*,2000) was used for docking of these ligands with SHP-1 (PDB\_ID: 2B30). All polar hydrogens were added to the receptor and Kollman United Atomic Charges were computed. For all ligands we added polar hydrogens and computed the Gasteiger charges. The structure of ligands and receptor were than saved in Pdbqt format used for docking calculations. The grid definition, adjusted to the receptor active site which included residues (Lys-277,Asn-278) was set to grid box with the dimension of (x=48 y=58, z=48),grid center (x=15.463, y=28.47, z=42.85) and spacing of 0.375.The ligands which showed binding energy greater than NSC87877were further used as test set to validate the QSAR model developed.

### **3.8 3D QSAR of substituted ligands**

To validate the QSAR model developed the substituent's which passed Lipinski filter and showed better binding energy than NSC87877with SHP-1 were used as test set for QSAR model

development and for testing these compounds. The inhibition constant calculated from autodock were converted into  $-\text{Log } k_i$  and it was used as activity.

## 4. RESULTS

### 4.1 Sequence and Structure analysis

Sequence and structural analysis was done to find the sequence and structure conservation among various members of non receptor tyrosine phosphatase. The Root Mean Square deviation values calculated for various members of nonreceptor tyrosine phosphatases with SHP-1, SHP-2 are shown in Table2. From the Table2 it can be seen that HePTP and SHP-1 bear close resemblance to each other as Root Mean Square deviation value between the two comes out to be.656A..Also Figure5 shows that the inhibitor binding site and the structure is conserved in both HePTP and SHP-1. In Figure 5 SHP-1 structure is shown in red color and HePTP is shown in green color .The inhibitor binding site ie. LYS 277, ASP 278 and ARG 459 in SHP-1(blue color) and ASP 105, TYR 106 and ARG276 (Pink color) in HePTP is highly conserved

### 4.2 Dataset and biological activity

All the inhibitors of the AID\_1077 and AID\_1059 have been listed in Table3.AID \_1007 (Fluorescent Secondary assay for dose –response confirmation of chemical inhibitors of HePTP) and AID \_1059(In vitro HePTP Dose response Colorimetric assay for SAR study) are two immunoassays .AID\_ 1077 had 147 compounds and AID \_1059 had 57 compounds which were common in AID\_1077 and AID\_1059.Table3 also shows the binding energy of these inhibitors with SHP-1(PDB\_ID 2B30) which have been calculated using Pyrx. NS87877 (Cid 16654632) the only inhibitor of SHP-1 reported in PUBCHEM has been highlighted in Table3. The inhibitors showing binding energy better than NSC87877(Cid 16654632) have been used for creating dataset as shown in Table4.Also from Table3 it can be seen that not all the inhibitors which have binding energy greater than NSC87877 have been used for creating the dataset. This is because the inhibitors needed for QSAR studies belong to congeneric series (similar in structure).So some of the inhibitors although had binding energy greater than NSC87877 but did not have similar structures so were not taken in the dataset. For building congeneric series clustering feature of PUBCHEM was used. All the compound were taken from the same node indicating a congeneric series.Table4 shows the inhibitors included in dataset (on the basis of binding energy, activity and common structures).Table4 also shows structure, binding energy and IC<sub>50</sub> values of compounds in dataset .All the compounds which had binding energy greater than NSC87877,good activity and similar structures are included in the dataset.

Table2 RMSD Values of various members of non receptor tyrosine phosphatases

| Members     | SHP1/PTPN6 | SHP2/PTPN11 |
|-------------|------------|-------------|
| PTPN1/PTP1B | 0.729      | 0.665       |
| PTPN2/TCPTP | 0.734      | 0.667       |
| PTPN9       | 0.962      | 0.799       |
| PTPN5/STEP  | 0.788      | 0.747       |
| PTPN7/HePTP | 0.656      | 0.673       |

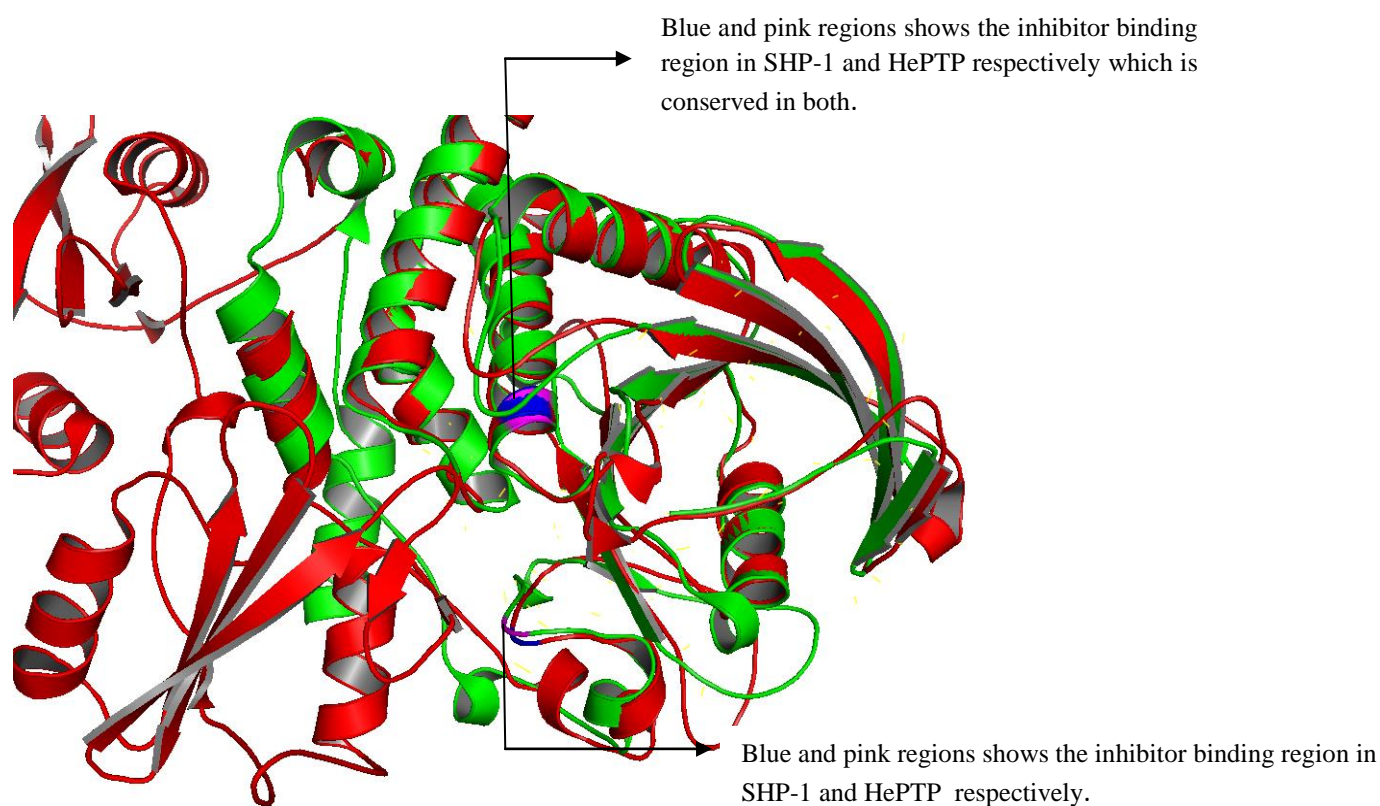


Figure5 Superimposition of SHP-1(shown in red colour) PDB\_ID:2B30 and HePTP (shown in green colour) PDB\_ID 1ZCO.The rmsd value between both structure was found to be 0 .656.The regions shown in blue (residues lysine 277and asparagine 278and arginine 459) in SHP-1 responsible for inhibitor binding site are conserved with regions shown in pink (asparagine 105,Tyrosine106 and arginine276) in HePTP.

Table3 shows all compounds of the assays and Binding energy of compounds with SHP-1(PDB\_ID 2B30) calculated using Pyrx

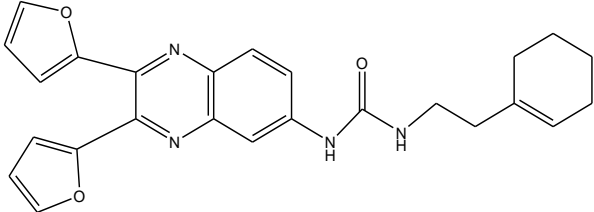
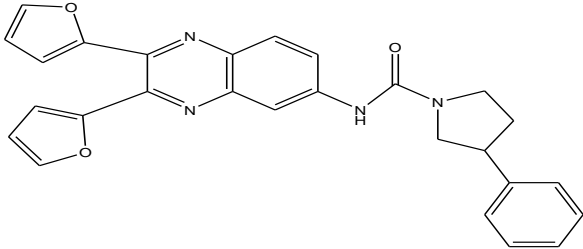
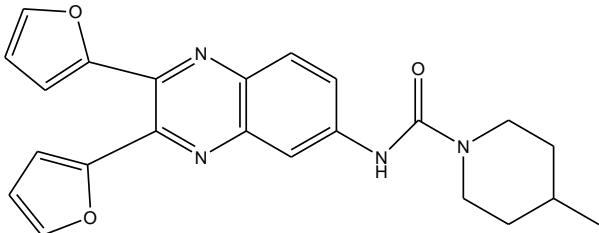
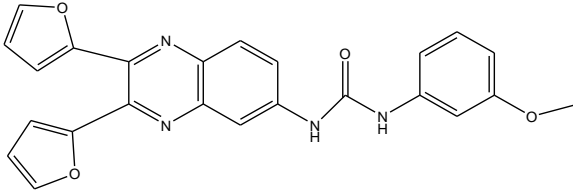
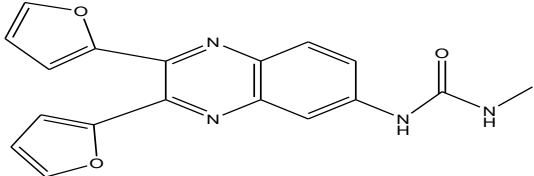
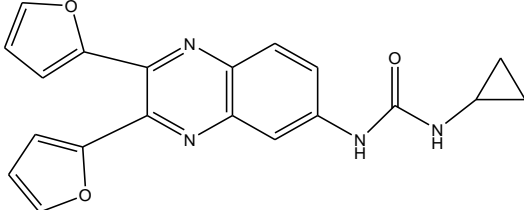
| Compound ID  | Binding Energy | Compound ID  | Binding Energy |
|--------------|----------------|--------------|----------------|
| cid 24178230 | -9.5           | cid 3157647  | -7.6           |
| cid 1209211  | -9.5           | cid 1228861  | -7.4           |
| cid 1299058  | -9.4           | cid 5346285  | -7.4           |
| cid 24178225 | -9.3           | cid 2266660  | -7.4           |
| cid 2214811  | -9.3           | cid 44182133 | -7.3           |
| cid 654089   | -9.3           | cid 889983   | -7.3           |
| cid 1331726  | -9.3           | cid 1516220  | -7.2           |
| cid 4715351  | -9.3           | cid 3157646  | -7.2           |
| cid 1789     | -9.3           | cid 16654632 | -7.1           |
| cid 4715351  | -9.3           | cid 2229326  | -6.9           |
| cid 128895   | -9.2           | cid 3000187  | -6.9           |
| cid 372955   | -9.0           | cid 3266419  | -6.9           |
| cid 24178231 | -8.9           | cid 2878586  | -6.9           |
| cid 2925154  | -8.9           | cid 1213466  | -6.9           |
| cid 892446   | -8.9           | cid 891589   | -6.9           |
| cid 2924978  | -8.9           | cid 3157647  | -6.9           |
| cid 3239711  | -8.9           | cid 1072898  | -6.9           |
| cid 24178215 | -8.9           | cid 1357397  | -6.9           |
| cid 24178232 | -8.8           | cid 646406   | -6.8           |
| cid 901652   | -8.8           | cid 2928673  | -6.8           |
| cid 1329592  | -8.7           | cid 3136927  | -6.5           |
| cid 1331766  | -8.7           | cid 44229061 | -6.5           |
| cid 24178226 | -8.5           | cid 2012947  | -6.5           |
| cid 176598   | -8.4           | cid 8853383  | -6.5           |
| cid 24178233 | -8.4           | cid 2230267  | -6.3           |
| cid 2301472  | -8.3           | cid 1282000  | -6.3           |
| cid 1092683  | -8.3           | cid 2230291  | -6.3           |
| cid 762708   | -8.3           | cid 2226406  | -6.3           |
| cid 1072900  | -8.3           | cid 4302116  | -6.3           |
| cid 5076888  | -8.2           | cid 1329592  | -6.3           |
| cid 3453217  | -8.2           | cid 16654893 | -6.2           |
| cid 2258411  | -8.2           | cid 16654891 | -6.1           |
| cid 901652   | -8.2           | cid 16654890 | -6.1           |
| cid 20110352 | -8.2           | cid 16654691 | -6.1           |
| cid 2240797  | -8.2           | cid 646096   | -6.1           |
| cid 403950   | -7.8           | cid 2940938  | -6.1           |
| cid 6492412  | -7.7           | cid 1589738  | -6.1           |
| cid 5766720  | -7.7           | cid 16217011 | -6.0           |
| cid 6000533  | -7.7           | cid 2869196  | -6.0           |
| cid 24178237 | -7.6           | cid 2975102  | -6.0           |
| cid 24178227 | -7.6           | cid 2930528  | -6.0           |
| cid 3239711  | -7.6           | cid 2300608  | -6.0           |

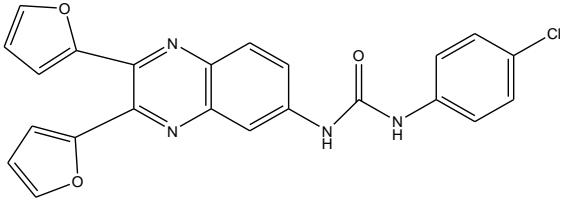
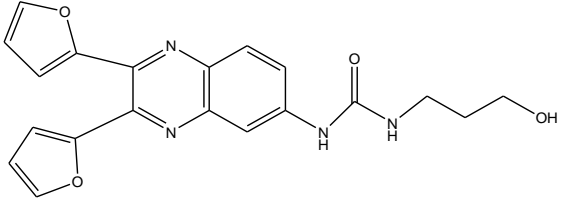
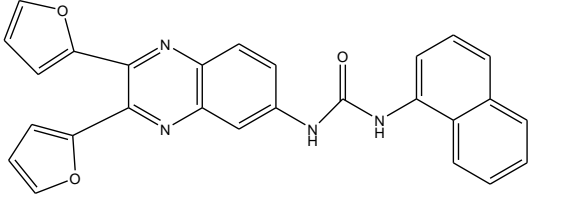
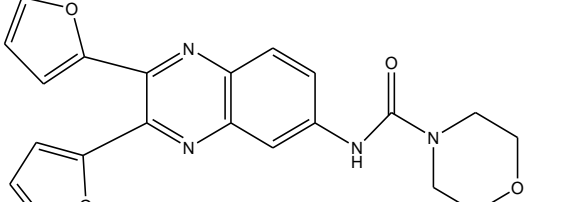
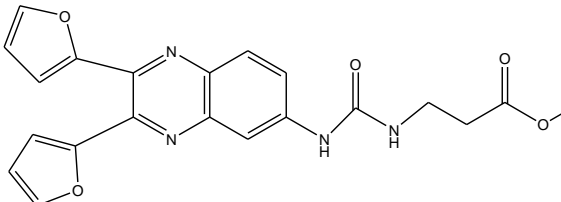
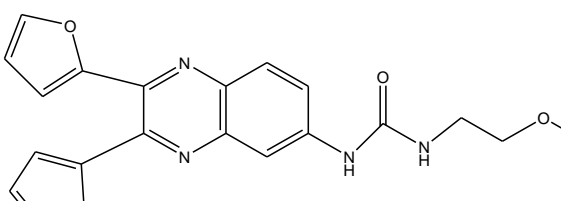
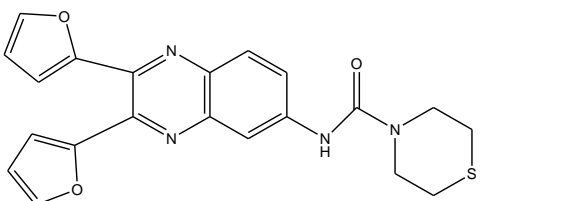
Table3 (continued) shows all compounds of the assays and Binding energy of compounds with SHP-1(PDB\_ID 2B30) calculated using Pyrx

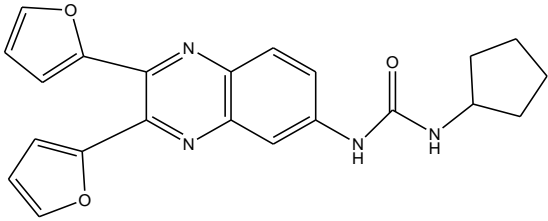
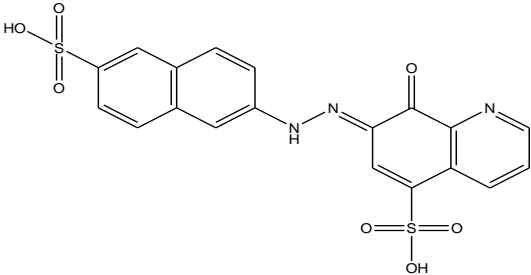
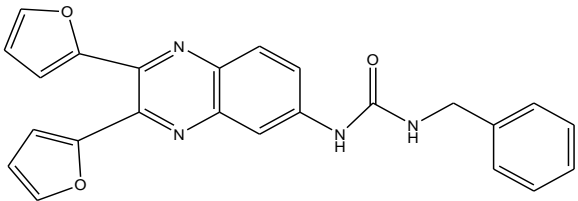
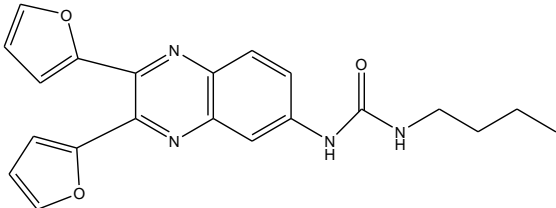
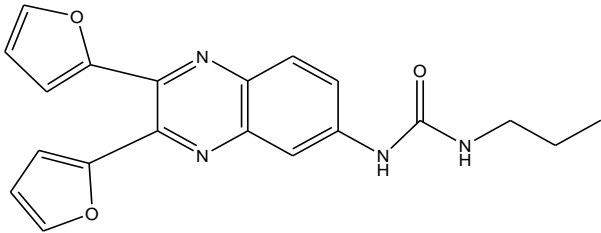
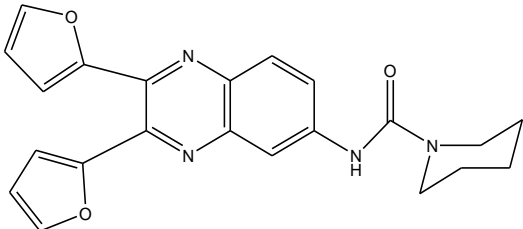
| Compound ID  | Binding Energy | Compound ID  | Binding Energy |
|--------------|----------------|--------------|----------------|
| cid 2921964  | -6.0           | cid 2975144  | -5.5           |
| cid 2914536  | -5.9           | cid 254573   | -5.5           |
| cid 2173774  | -5.9           | cid 4969416  | -5.4           |
| cid 5341943  | -5.9           | cid 2545467  | -5.4           |
| cid 65456566 | -5.9           | cid 5765581  | -5.3           |
| cid 24761488 | -5.9           | cid 9512029  | -5.2           |
| cid 818221   | -5.9           | cid 1328767  | -5.1           |
| cid 889170   | -5.9           | cid 24178232 | -5.0           |
| cid 1209230  | -5.9           | cid 71599    | -5.0           |
| cid 3883207  | -5.8           | cid 2207086  | -5.0           |
| cid 5336454  | -5.8           | cid 16654969 | -5.0           |
| cid 2826665  | -5.8           | cid 2269367  | -5.0           |
| cid 2266660  | -5.7           |              |                |
| cid 9595032  | -5.6           |              |                |
| cid 5765582  | -5.6           |              |                |
| cid 1756795  | -5.6           |              |                |
| cid 2260301  | -5.6           |              |                |
| cid 1435211  | -5.6           |              |                |
| cid 3182456  | -5.6           |              |                |
| cid 44182134 | -5.5           |              |                |
| cid 5504142  | -5.5           |              |                |
| cid 2062730  | -5.5           |              |                |

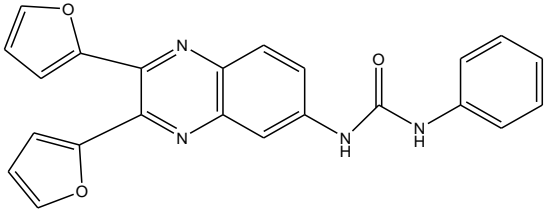
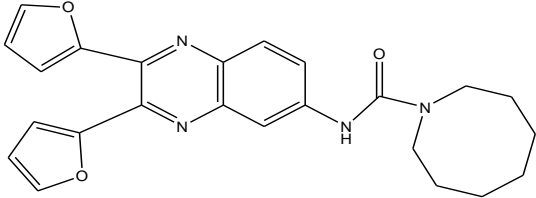
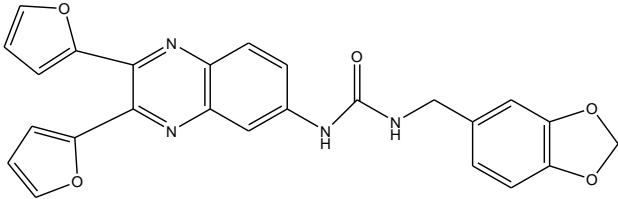
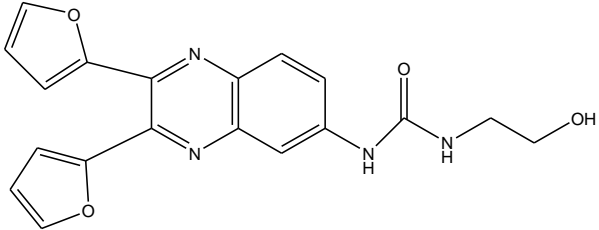
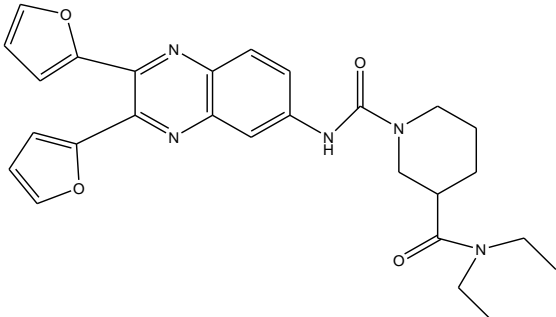


Table4 shows the structure, IC<sub>50</sub> values and binding energy of compounds

| Compound ID | Structure   | -Log Ic50 | B.E  |
|-------------|---|-----------|------|
| cid 1209211 |   | 5.560     | -9.5 |
| cid 2925154 |    | 6.250     | -8.9 |
| cid 2924978 |  | 6.070     | -8.9 |
| cid 762708  |  | 5.550     | -8.3 |
| cid 1092683 |  | 6.080     | -8.3 |
| cid 1329592 |  | 5.400     | -8.7 |

|              |   |       |      |
|--------------|---|-------|------|
| cid 24178225 |    | 6.020 | -9.3 |
| cid 2301472  |    | 6.070 | -8.3 |
| cid 24178230 |    | 5.890 | -9.5 |
| cid 1072900  |   | 6.150 | -8.3 |
| cid 24178237 |  | 6.160 | -7.6 |
| cid 3157647  |  | 6.700 | -7.6 |
| cid 24178215 |  | 5.470 | -8.9 |

|              |  |       |      |
|--------------|--|-------|------|
| cid 24178233 |     | 5.950 | -8.4 |
| cid 16654632 |     | 5.530 | -7.1 |
| cid 24178232 |     | 5.810 | -8.8 |
| cid 24178227 |    | 6.400 | -7.6 |
| cid 3239711  |  | 5.400 | -7.6 |
| cid 372955   |   | 5.530 | -9.0 |

|              |   |       |      |
|--------------|---|-------|------|
| cid 24178231 |    | 6.240 | -8.9 |
| cid 2214811  |    | 5.730 | -9.3 |
| cid 1299058  |   | 5.670 | -9.4 |
| cid 1331766  |  | 6.000 | -9.3 |
| cid 24178226 |  | 6.320 | -8.5 |

---

### 4.3 Pharmacophore modeling

Several seven-point common pharmacophore hypotheses with various combination of site were generated using active molecules. Minimum sites 4 and Maximum site 7 were considered to have optimum combination of sites or features common to active compounds. The molecules were classed into active and inactive based on activity threshold for identifying the pharmacophore features considering highest active molecule. The pharmacophore models were ranked on the basis of alignment to the active compounds. The “survival” scoring (s) function identifies the best candidate from the generated models and assigns an overall ranking of all the hypotheses. The scoring algorithm includes contributions from the alignment of site points and vectors, volume overlap, selectivity, number of ligands matched, relative conformational energy, and activity. However, the model should also discriminate between active and inactive molecules. If inactive molecules scores well, the hypotheses could be invalid as it does not discriminate between actives and inactives. The various hypothesis generated and their scores are listed in table 5. Among which hypothesis AAADRRR.190 was selected on the basis of score and discrimination of active and nonactive molecules i.e if active molecules score well, the hypothesis could be invalid as it does not discriminate between active and inactive.

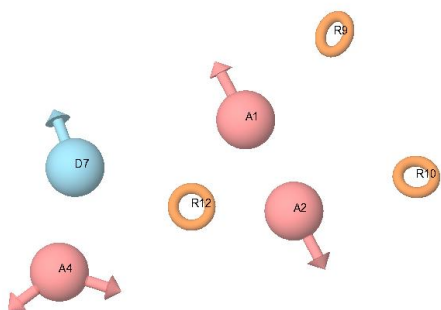
The best pharmacophore hypothesis AAADRRR-190 Figure 6 was selected for further QSAR study. The above mentioned 3D pharmacophore hypothesis Figure 6 encompass the following features: three hydrogen bond acceptor (A) in pink color, one hydrogen bond donors (D) in sky blue color and three Aromatic ring (R) in yellow color. The hypothesis was selected on the basis of survival score. Survival weighted is combination of the vector, site, volume, and survival scores, and a term for the number of matches, a large value of survival score indicates the better fitness of the active ligands on the common pharmacophore and validates the model. So from Table 5 it can be seen that hypothesis AAADRRR.190 has the best survival score 3.581 which is highest among all the hypothesis generated. The Inactive score, Vector, site, Volume and selectivity of hypothesis came out to be 1.787, 0.948, .88, .757 and 2.579 respectively. Survival inactive survival score for actives with a multiple of the survival score for inactive subtracted. Site score, this score measures how closely the site points are superimposed in an alignment to the pharmacophore of the structures that contribute to the hypothesis, based on the RMS deviation of the site points of a ligand from those of the reference ligand. Selectivity estimate of the rarity of the hypothesis, High selectivity means that the hypothesis is more likely to be unique to actives. All these values are highest among the selected hypothesis among all the hypothesis (Table 4). All the molecules showed good alignment with good fitness score ranging 3.00 (for highest) to 1.95 (for lowest active).

The 2D representation of the AAADRRR- 190 hypothesis is given in Figure 6c. Figure 7 shows structure of all the compounds and common pharmacophore obtained.

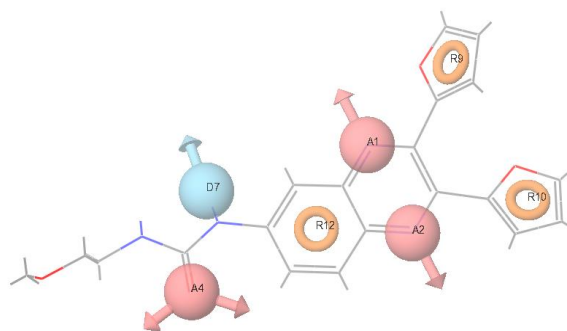
Table5 Phase hypothesis

| Phase hypothesis | Survival | Survival inactive | Vector | Site | Volume | Selectivity |
|------------------|----------|-------------------|--------|------|--------|-------------|
| AAADRRR.190      | 3.581    | 1.787             | .948   | .88  | .757   | 2.579       |
| AAADRRR.263      | 3.543    | 1.769             | .947   | .87  | .730   | 2.476       |
| AAADRRR.235      | 3.540    | 1.755             | .947   | .87  | .720   | 2.475       |

6a



6b



6c

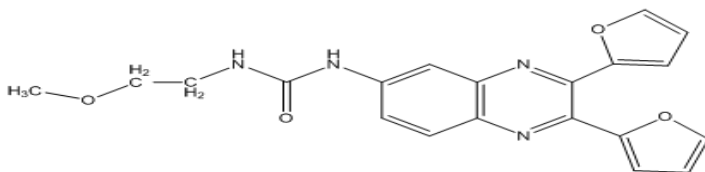
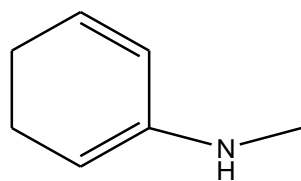
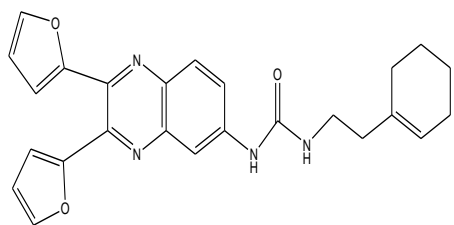


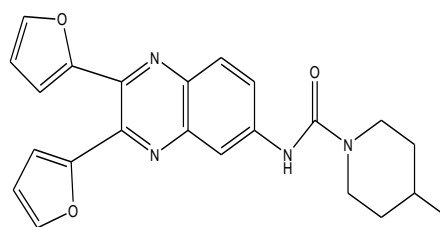
Figure. 6a and 6b shows common pharmacophore for active ligands [three hydrogenbond acceptor (A) in pink color, one hydrogen bond donors (D) in sky blue color, and three aromatic ring (R) in yellow color. 6c shows 2D representation of common pharmacophore.



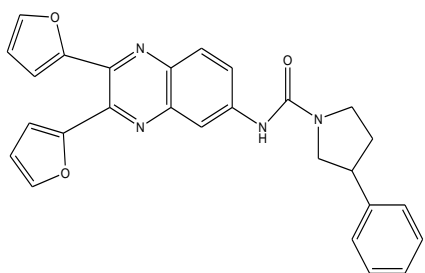
Common backbone



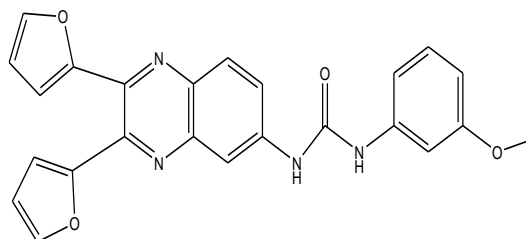
cid 1209211



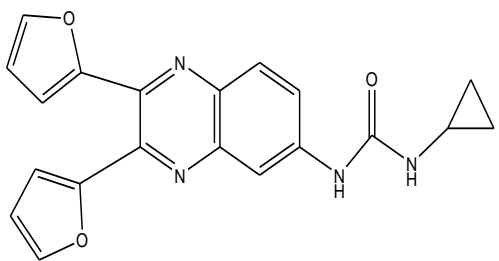
cid 2924978



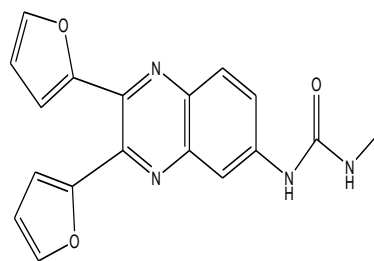
cid 2925154



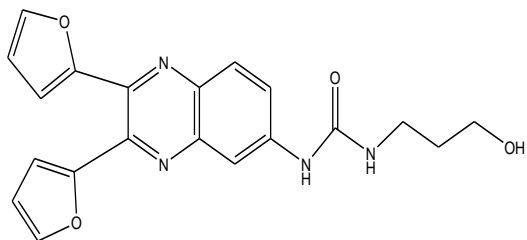
cid 762708



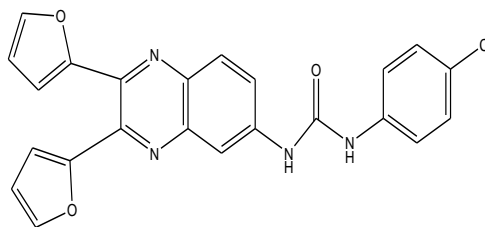
cid 1329592



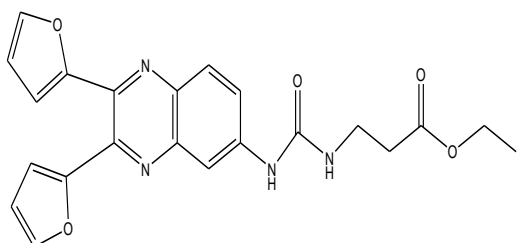
cid 1092683



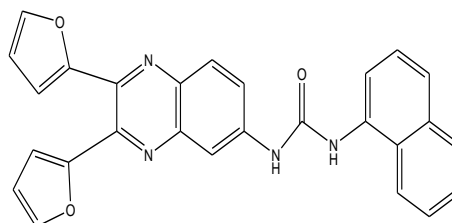
cid 2301472



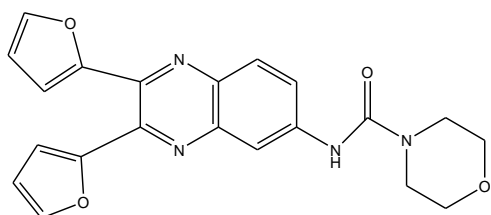
cid 24178225



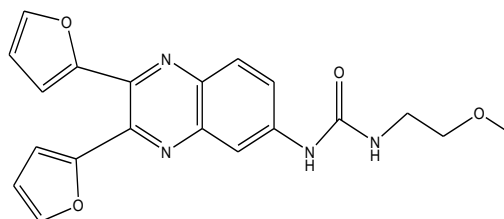
cid 24178237



cid 24178230

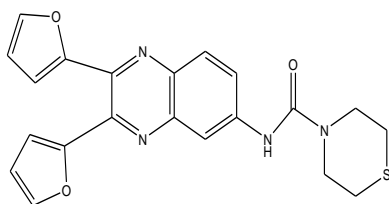


cid 1072900

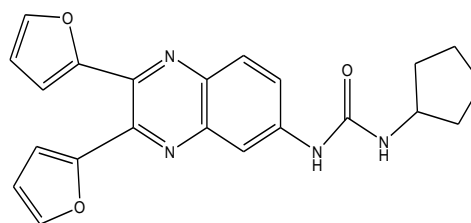


cid 3157647

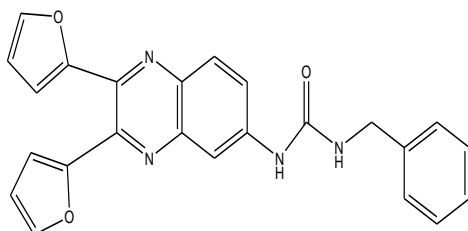




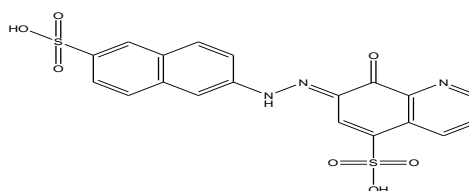
cid 24178215



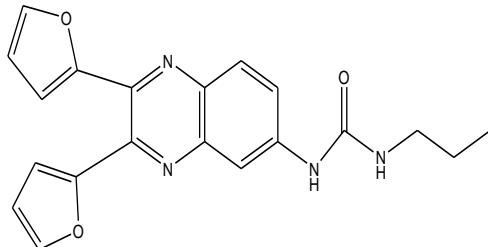
cid 24178233



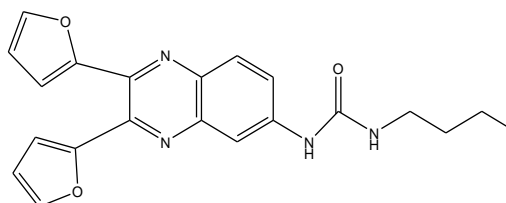
cid 24178232



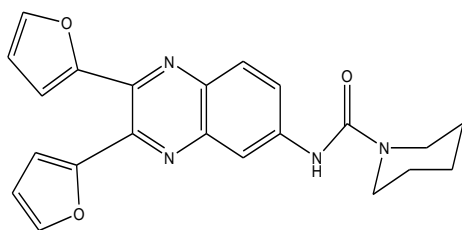
cid 16654632



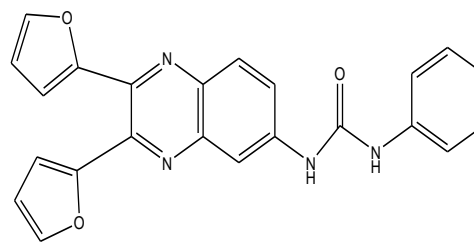
cid 3239711



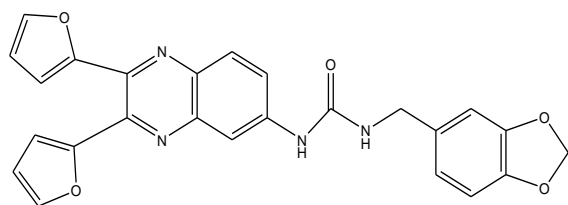
cid 24178227



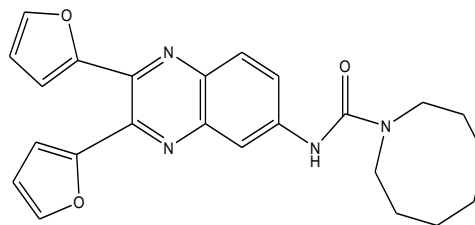
cid 372955



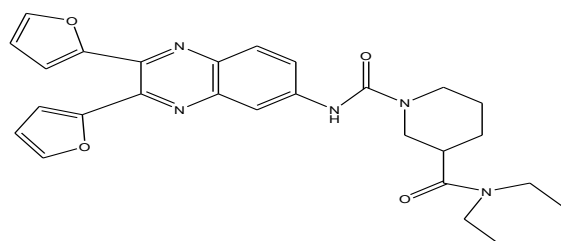
cid 24178231



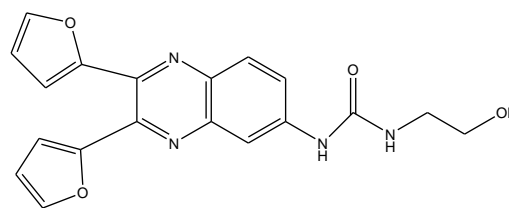
cid 1299058



cid 2214811



cid 24178226



cid 1331766

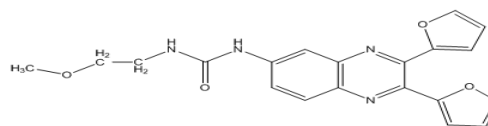
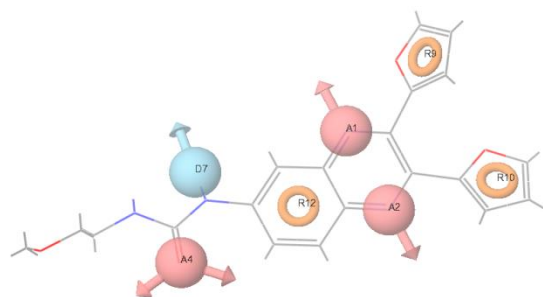


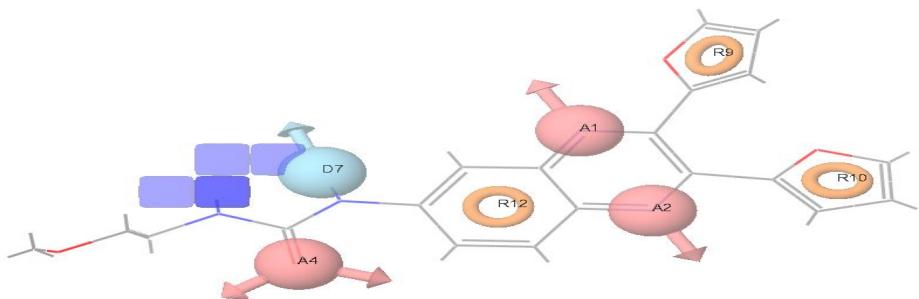
Figure7 Shows the structure of compounds in dataset and common pharmacophore obtained

#### 4.4 Building of 3D QSAR model

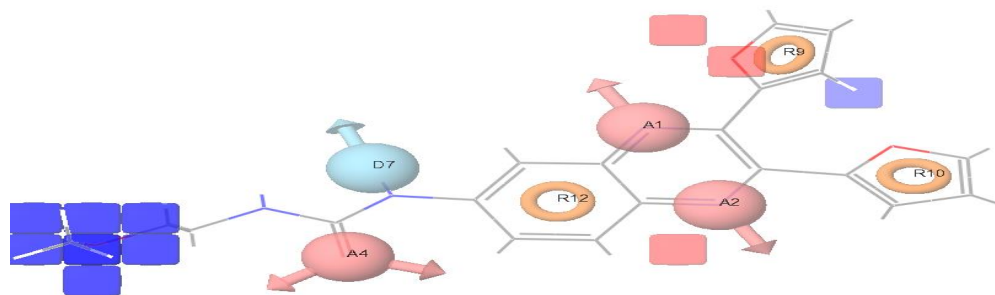
A 3D-QSAR study has been performed successfully on the series of inhibitors to understand the effect spatial arrangement of structural features on biological activity of these inhibitors. Result of the 3D-QSAR can be visualized from Figure 8. The blue cubes in 3D plots of the 3D pharmacophore regions refer to ligand regions in which the specific feature is important for better activity, whereas the red cubes demonstrate that particular structural feature or functional group is not essential for the activity or likely to reason for decreased in biological activity. From Figure 8a it can be seen that H donor can be added to carbon just adjacent to acceptor shown in common pharmacophore. Figure 8b shows that electron accepting group can be added near one of the rings and to carbon chain. Figure 8c shows hydrophobic groups can be added near one of the rings. All these changes can lead to increase in biological activity of inhibitors. The statistical results of 3D-QSAR study are summarized in Table 6. The reliability of the present 3D-QSAR analysis can be justified by the fact that all statistical measures are significant. The model generated showed statistically good results with  $r^2$  (Correlation Coefficient) = 0.897, correlation coefficient  $r$  measures how closely the observed data tracks the fitted regression line.

$q^2 = 0.8173$ .  $q^2$  is leave one out cross validation. It involves using a single observation from original sample as the validation determinant and remaining observation as training data. This is repeated such that each observation in the sample is used once as validation data. The statistical significance of model was also confirmed by a high value of Fischer ratio = 60.9 Fischer ratio indicates while the fit of the data to the regression line is excellent, how can one decide if this correlation is based purely on chance. A very low value of RMSE = 0.11 again indicates the model is significant. One of the other parameter which signify model predictivity is Pearson R. Its value of 0.94 shows that correlation between predicted and observed activity for test compound is excellent. The fitness graph can be visualized from Figure 9. The performance of QSAR model on the training and test set molecules is shown in Figure 9. The solid line in the plot indicates the hypothetical "best fit" line between the predicted and experimental activity. The scatter plot indicates a good relation between the predicted and experimental activity with  $r^2 = 0.81$ . Table 7 shows that 3PLS factors were taken for QSAR model development by randomly selecting 17 compounds as training set and 7 compounds as test set. It also shows the pharm set and also fitness of compounds. PLS factors was set as 03, the maximum number of factors in each model can be 1/5 total no of training molecules. So for our model number of PLS factors was taken to be 3 in our case. The QSAR model showed good statistical results and was then considered for building of new substituents.

8a



8b



8c

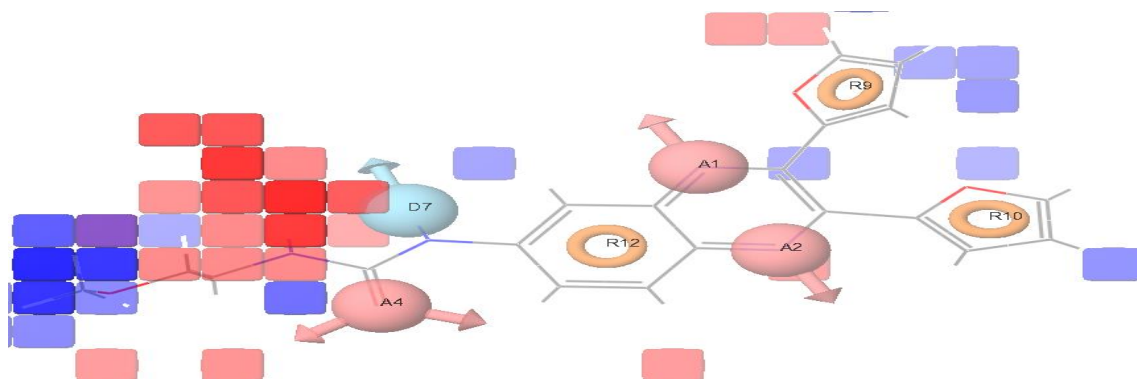


Figure8 QSAR visualization of various substituent's affect: 8a: hydrogen-bond donor effect, 8b: electron withdrawing feature,8c:hydrophobic effect. Blue regions shows where various substituent's like hydrogen bond donor, electron withdrawing and hydrophobic substituent's can

be substituted. Red regions shows the region in which substitution can decrease biological activity and is not allowed Figure8 shows 3D-pharmacophore regions around compounds. For the selected pharmacophore blue and red cubes represent favorable and unfavorable regions, respectively. Molecular substitutions which increase the number of blue cubes will definitely lead to increases biological activity, while molecular substitutions which increase the number of red cubes will lead to decreased activity. Figure8a shows the H donor characteristic for the selected hypothesis .The blue cubes near donor7 and its neighbor carbon indicates that addition of H donor here will increase biological activity of inhibitors. Figure8b shows Electron withdrawing characteristic for selected hypothesis. Visual analysis of Figure shows that addition of electron withdrawing group near ring9 will increase biological activity of inhibitors. The potency of compound can be increased by addition of small electron withdrawing groups like fluoro, chloro, bromo etc on ring9 and the carbon chain. Figure 8c shows the effect of positive and negative hydrophobic potential. it can be deduced from fig that hydrophobic group are well tolerated near ring9 and ring10 while the substitution of hydrophobic group near donor attached to ring 12 is unacceptable or may decrease the biological activity of the inhibitors.

Table6 3D-QSAR statistical parameters

| PLS factors | SD     | r <sup>2</sup> | F     | P          | RMSE   | q <sup>2</sup> | Pearson R |
|-------------|--------|----------------|-------|------------|--------|----------------|-----------|
| 1           | 0.2271 | 0.6496         | 27.8  | 9.378e-005 | 0.1719 | 0.6242         | 0.9546    |
| 2           | 0.1275 | 0.897          | 60.9  | 1.233e-007 | 0.1198 | 0.8173         | 0.9467    |
| 3           | 0.0817 | 0.9607         | 105.9 | 2.184e-009 | 0.1749 | 0.6111         | 0.711     |

SD standard deviation of the regression, r<sup>2</sup> for the regression, F variance ratio. Large values of F indicate a more statistically significant regression, P significance level of variance ratio. Smaller values indicate a greater degree of confidence, RMSE root-mean-square error, q<sup>2</sup> for the predicted activities, Pearson R value for the correlation between the predicted and observed activity for the test set.

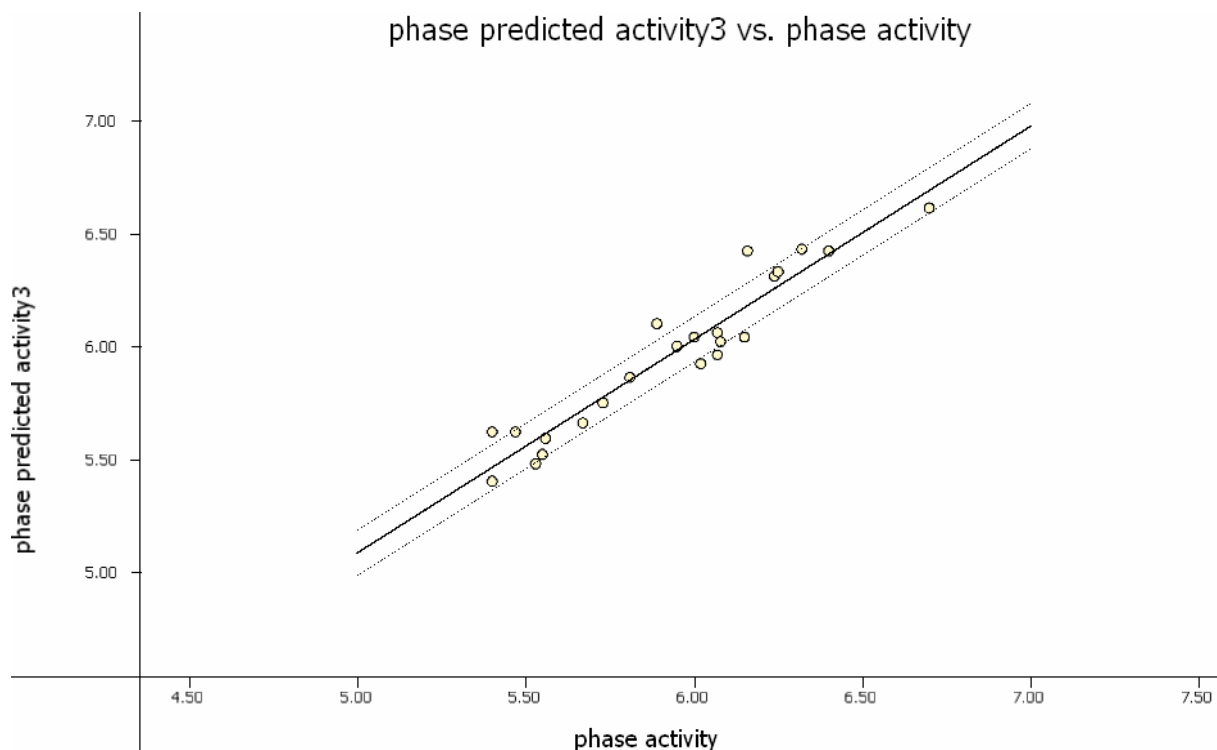


Figure9 Fitness graph between observed activity versus PHASE predicted activity for training and test set compounds

Validity of the model can be expressed by internal predictivity ( $q^2 = 0.81$ ) which is obtained by leave-one-out (LOO) or leave n out method. The  $q^2$  by leave-one-out method is more reliable and robust statistical parameter than  $r^2$  because it is obtained by external validation method of dividing the dataset into training and test set. The large value of F (60.9) indicates a statistically significant regression model, which is supported by the small value of the variance ratio (P), an indication of a high degree of confidence. Further small values of standard deviation of the regression (0.12) and RMSE (0.119) make an obvious implication that the data used for model generation are best for the QSAR analysis. Apart from the above-mentioned features, PLS factor also confirms the reliability of the model. In this study, number of PLS factor was taken as 3 and for each increment it gives one equation and there should be stepwise improvement eachtime the model generated. In addition to the above parameters it is interesting to note that active ligands are closely fitted to the regression line and inactive ligands are scattered. The solid line in the plot indicates the hypothetical “best fit” line between the predicted and experimental activity. The scatter plot indicates a good relation between the predicted and experimental activity with  $r^2=0.81$ .

Table7 Fitness and PHASE predicted activity data for all compounds

| Cid      | QSAR set | Activity | PLS factors | Phase predicted activity | Pharm set | Fitness |
|----------|----------|----------|-------------|--------------------------|-----------|---------|
| 1209211  | Training | 5.560    | 1 2 3       | 5.89,5.56,5.59           | Inactive  | 2.75    |
| 2925154  | Test     | 6.250    | 1 2 3       | 6.16,6.36,6.33           | Active    | 2.57    |
| 2924978  | Test     | 6.070    | 1 2 3       | 5.87,5.99,5.96           | Active    | 2.16    |
| 762708   | Training | 5.550    | 1 2 3       | 5.62,5.52,5.60           | Inactive  | 1.95    |
| 1092683  | Training | 6.080    | 1 2 3       | 6.06,6.12,6.02           | Active    | 2.34    |
| 1329592  | Training | 5.400    | 1 2 3       | 5.47,5.34,5.40           | Inactive  | 2.34    |
| 24178225 | Training | 6.020    | 1 2 3       | 6.22,6.09,5.92           | Active    | 2.33    |
| 2301472  | Training | 6.070    | 1 2 3       | 6.21,6.14,6.06           | Active    | 2.51    |
| 24178230 | Test     | 5.890    | 1 2 3       | 6.12,6.04,6.10           | Active    | 2.77    |
| 1072900  | Training | 6.150    | 1 2 3       | 5.72,5.99,6.04           | Active    | 2.12    |
| 24178237 | Test     | 6.160    | 1 2 3       | 6.24,6.27,6.42           | Active    | 2.89    |
| 3157647  | Training | 6.700    | 1 2 3       | 6.26,6.39,6.61           | Active    | 3.00    |
| 24178215 | Training | 5.470    | 1 2 3       | 5.60,5.69,5.62           | Inactive  | 2.13    |
| 24178233 | Training | 5.950    | 1 2 3       | 5.95,5.95,6.00           | Active    | 2.71    |
| 16654632 | Training | 5.530    | 1 2 3       | 5.44,5.44,5.48           | Inactive  | 2.34    |
| 24178232 | Test     | 5.810    | 1 2 3       | 6.02,5.81,5.86           | Active    | 2.79    |
| 24178227 | Training | 6.400    | 1 2 3       | 6.28,6.43,6.42           | Active    | 2.68    |
| 3239711  | Test     | 5.400    | 1 2 3       | 5.56,5.59,5.62           | Inactive  | 2.35    |
| 372955   | Test     | 5.530    | 1 2 3       | 5.54,5.46,5.52           | Inactive  | 2.66    |
| 24178231 | Training | 6.240    | 1 2 3       | 6.30,6.37,6.31           | Active    | 2.71    |
| 2214811  | Training | 5.730    | 1 2 3       | 5.84,5.69,5.75           | Active    | 2.77    |
| 1299058  | Training | 5.670    | 1 2 3       | 5.96,5.62,5.66           | Inactive  | 2.73    |
| 1331726  | Training | 6.000    | 1 2 3       | 5.72,6.09,6.04           | Active    | 2.22    |
| 24178226 | Training | 6.320    | 1 2 3       | 6.30,6.32,6.43           | Active    | 2.84    |

## 4.5 Building of new substituent's

The results shown in Figure8 were used to build substituents. Figure10 shows how the QSAR results have been used to add substituent's  $R_1$ ,  $R_2$ ,  $R_3$ ,  $R_4$ , and  $R_5$ . Figure10a, 10b, 10c shows how  $R_1$ (hydrogen donor),  $R_2, R_3, R_4$ (electron withdrawing group),  $R_5$ (hydrophobic groups) have been substituted according to QSAR results. So 83 substituents were built as can be seen from Table8. Different substituents were attached at  $R_1, R_2, R_3, R_4, R_5$  (Table8).

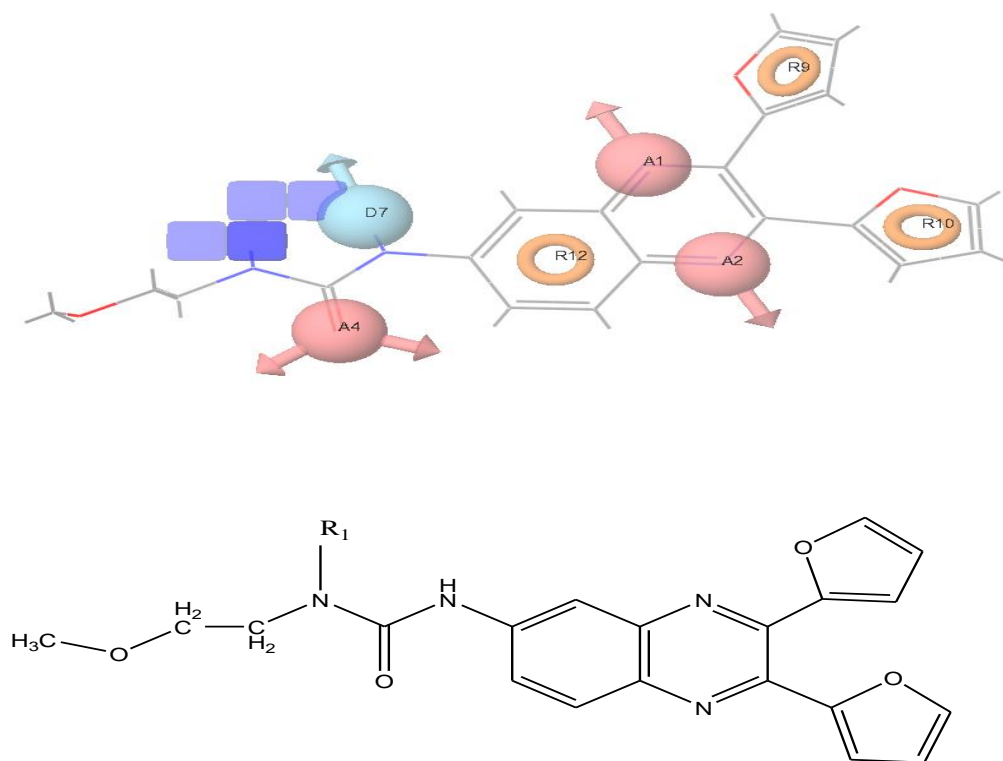
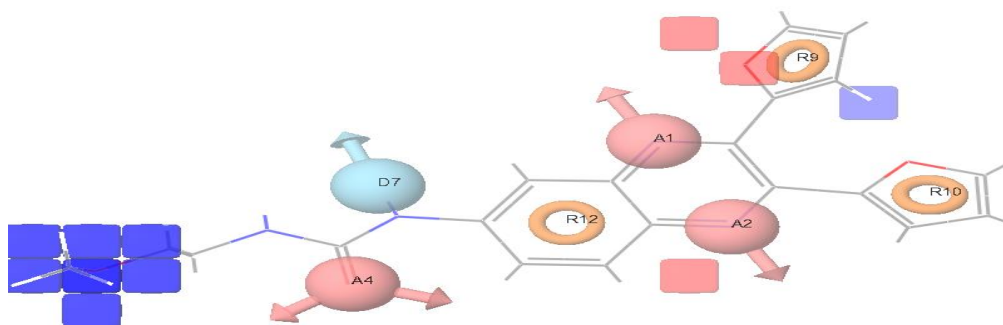


Figure10a H-donor group  $R_1$  attached according to QSAR result.





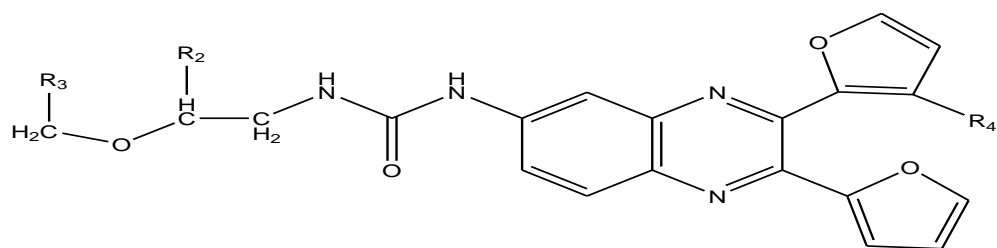


Figure10b Electron withdrawing group R<sub>2</sub>, R<sub>3</sub>, R<sub>4</sub> attached according to QSAR result.

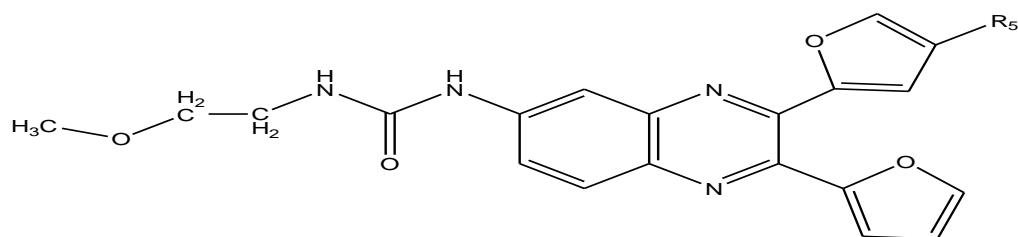
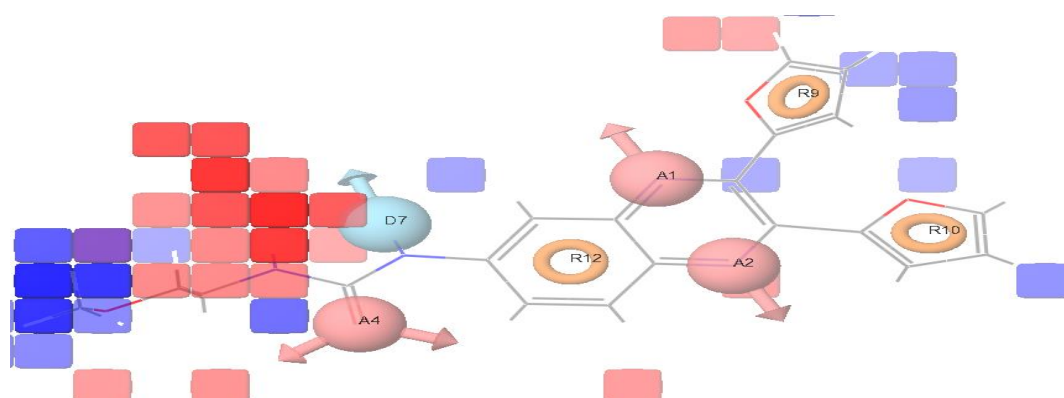


Figure10c Hydrophobic group R<sub>5</sub> attached according to QSAR result

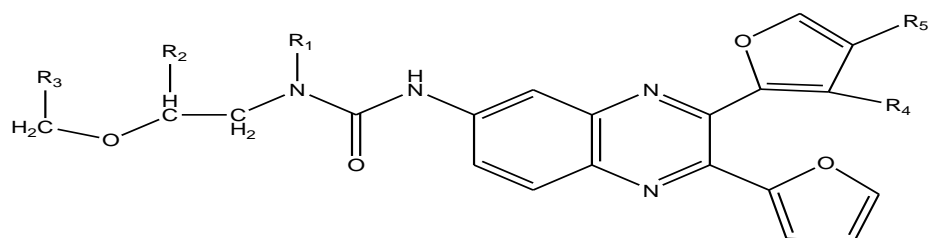


Figure10 R<sub>1</sub>, R<sub>2</sub>, R<sub>3</sub>, R<sub>4</sub>, R<sub>5</sub> substituents attached to common pharmacophore.

So 83 substituents were built as can be seen from Table8. Different substituents were attached at R<sub>1</sub>,R<sub>2</sub>,R<sub>3</sub>,R<sub>4</sub>,R<sub>5</sub> (Table8)

Table8 Shows various substituent's built.

| Compounds | R <sub>1</sub>  | R <sub>2</sub> | R <sub>3</sub> | R <sub>4</sub> | R <sub>5</sub>                |
|-----------|-----------------|----------------|----------------|----------------|-------------------------------|
| 1         | CH <sub>3</sub> | I              | Cl             | F              | C <sub>2</sub> H <sub>5</sub> |
| 2         | CH <sub>3</sub> | Cl             | Cl             | F              | C <sub>2</sub> H <sub>5</sub> |
| 3         | CH <sub>3</sub> | F              | Cl             | F              | C <sub>2</sub> H <sub>5</sub> |
| 4         | NH <sub>2</sub> | I              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 5         | NH <sub>2</sub> | Cl             | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 6         | NH <sub>2</sub> | F              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 7         | OH              | I              | F              | F              | C <sub>3</sub> H <sub>7</sub> |
| 8         | OH              | Cl             | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 9         | OH              | F              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 10        | CH <sub>3</sub> | I              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 11        | CH <sub>3</sub> | Cl             | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 12        | CH <sub>3</sub> | F              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 13        | CH <sub>3</sub> | I              | I              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 14        | CH <sub>3</sub> | Cl             | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 15        | CH <sub>3</sub> | F              | F              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 16        | NH <sub>2</sub> | I              | I              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 17        | NH <sub>2</sub> | Cl             | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 18        | NH <sub>2</sub> | F              | F              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 19        | OH              | Br             | I              | Cl             | C <sub>3</sub> H <sub>7</sub> |
| 20        | OH              | Cl             | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 21        | OH              | F              | F              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 23        | CH <sub>3</sub> | Cl             | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 24        | CH <sub>3</sub> | F              | F              | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 25        | CH <sub>3</sub> | F              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 26        | CH <sub>3</sub> | Cl             | Cl             | F              | C <sub>3</sub> H <sub>7</sub> |
| 27        | CH <sub>3</sub> | I              | I              | F              | C <sub>2</sub> H <sub>5</sub> |
| 28        | OH              | F              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 29        | OH              | Cl             | Cl             | F              | C <sub>2</sub> H <sub>5</sub> |
| 30        | OH              | I              | I              | I              | C <sub>2</sub> H <sub>5</sub> |
| 31        | NH <sub>2</sub> | F              | F              | F              | C <sub>3</sub> H <sub>7</sub> |
| 32        | NH <sub>2</sub> | Cl             | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 33        | NH <sub>2</sub> | I              | F              | F              | C <sub>2</sub> H <sub>5</sub> |
| 34        | CH <sub>3</sub> | F              | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |
| 35        | CH <sub>3</sub> | Cl             | Cl             | Cl             | C <sub>2</sub> H <sub>5</sub> |

|    |                 |    |    |    |                               |
|----|-----------------|----|----|----|-------------------------------|
| 36 | CH <sub>3</sub> | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 37 | OH              | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 38 | OH              | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 39 | OH              | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 40 | OH              | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 41 | OH              | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 42 | OH              | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 43 | OH              | I  | Cl | Cl | C <sub>3</sub> H <sub>7</sub> |
| 44 | OH              | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 45 | OH              | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 46 | NH <sub>2</sub> | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 47 | NH <sub>2</sub> | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 48 | NH <sub>2</sub> | F  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 49 | NH <sub>2</sub> | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 50 | NH <sub>2</sub> | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 51 | NH <sub>2</sub> | Cl | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 52 | NH <sub>2</sub> | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 53 | NH <sub>2</sub> | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 54 | NH <sub>2</sub> | I  | Cl | Cl | C <sub>2</sub> H <sub>5</sub> |
| 55 | CH <sub>3</sub> | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 56 | CH <sub>3</sub> | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 57 | CH <sub>3</sub> | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 58 | CH <sub>3</sub> | Cl | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 59 | CH <sub>3</sub> | Cl | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 60 | CH <sub>3</sub> | Cl | I  | I  | C <sub>3</sub> H <sub>7</sub> |
| 61 | CH <sub>3</sub> | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 62 | CH <sub>3</sub> | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 63 | CH <sub>3</sub> | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 64 | CH <sub>3</sub> | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 65 | CH <sub>3</sub> | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 66 | OH              | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 67 | OH              | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 68 | OH              | F  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 69 | OH              | Cl | I  | I  | C <sub>3</sub> H <sub>7</sub> |
| 70 | OH              | Cl | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 71 | OH              | Cl | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 72 | OH              | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 73 | OH              | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 74 | OH              | I  | I  | I  | C <sub>2</sub> H <sub>5</sub> |
| 75 | NH <sub>2</sub> | F  | I  | I  | C <sub>3</sub> H <sub>7</sub> |

|    |                 |    |   |   |                               |
|----|-----------------|----|---|---|-------------------------------|
| 76 | NH <sub>2</sub> | F  | I | I | C <sub>2</sub> H <sub>5</sub> |
| 77 | NH <sub>2</sub> | F  | I | I | C <sub>2</sub> H <sub>5</sub> |
| 78 | NH <sub>2</sub> | Cl | I | I | C <sub>2</sub> H <sub>5</sub> |
| 79 | NH <sub>2</sub> | Cl | I | I | C <sub>2</sub> H <sub>5</sub> |
| 80 | NH <sub>2</sub> | Cl | I | I | C <sub>2</sub> H <sub>5</sub> |
| 81 | NH <sub>2</sub> | I  | I | I | C <sub>2</sub> H <sub>5</sub> |
| 82 | NH <sub>2</sub> | I  | I | I | C <sub>2</sub> H <sub>5</sub> |
| 83 | NH <sub>2</sub> | I  | I | I | C <sub>2</sub> H <sub>5</sub> |

#### 4.6 Validation of Physicochemical properties

The substituent's build were subjected to Lipinski filter Table 9 shows Log P value, molecular weight, H bond acceptors, H bond donor and TPSA values for all the substituents built. The substituents which passed the lipinski filter have been highlighted. 15 substituents were found to pass all the Lipinski filter. The substituents which passed Lipinski filter were subjected to docking. For likeliness of a compound to become a drug Log P Value should be less than or equal to 5, Molecular weight should be less than 500 daltons, H bond donor should be less than 5, H bond acceptors should be less than 10. All the compounds which followed all the Lipinski filter have been highlighted in green in Table 9. These were further subjected to molecular docking to find out if their binding energy were better than NSC87877.

#### 4.7 Molecular docking

The substituent's which passed Lipinski rule filter were docked with SHP-1 PDB\_ID (2B30). The binding energy and interacting residues are shown in Table 10. All the substituent's showed binding energy greater than NSC87877 (only inhibitor known for SHP-1). Table 10 Shows the binding energy and interacting residues of substituent's built with SHP-1 calculated using autodock. The grid definition, adjusted to the receptor active site which included residues (Lys-277, Asn-278) was set to grid box with the dimension of (x=48 y=58, z=48), grid center (x=15.463, y=28.47, z=42.85) and spacing of 0.375. The parameters were kept same for all the compounds during docking. NSC87877 was taken as a reference molecule during docking.

Table 9 Shows compounds with Lipinski filter application. Validation of physiochemical properties was done by application of Lipinski filter through molinformatics server

| Compound no | Log P | Molecular weight | H acceptors | H donors | TPSA  |
|-------------|-------|------------------|-------------|----------|-------|
| 1           | 7.148 | 596              | 6           | 1        | 67.85 |
| 2           | 4.734 | 495              | 6           | 1        | 67.85 |
| 3           | 4.435 | 488              | 6           | 1        | 67.85 |
| 4           | 6.226 | 581              | 7           | 3        | 93.8  |
| 5           | 4.821 | 489              | 7           | 3        | 93.87 |
| 6           | 4.513 | 473              | 7           | 3        | 93.87 |
| 7           | 6.914 | 596              | 7           | 2        | 88.07 |
| 8           | 4.118 | 490              | 7           | 2        | 88.07 |
| 9           | 4.811 | 474              | 7           | 2        | 88.07 |
| 10          | 6.84  | 580              | 6           | 1        |       |
| 11          | 4.435 | 488              | 6           | 1        | 67.85 |
| 12          | 4.128 | 472              | 6           | 1        | 67.85 |
| 13          | 7.663 | 613              | 6           | 1        | 67.85 |
| 14          | 7.268 | 521              | 6           | 1        | 67.85 |
| 15          | 6.95  | 505              | 6           | 1        | 67.85 |
| 16          | 6.74  | 597              | 7           | 3        | 93.87 |
| 17          | 6.33  | 506              | 7           | 3        | 93.87 |
| 18          | 4.73  | 489              | 7           | 3        | 93.87 |
| 19          | 6.42  | 612              | 7           | 2        | 88.07 |
| 20          | 4.32  | 490              | 7           | 2        | 88.07 |
| 21          | 4.32  | 490              | 7           | 2        | 88.07 |
| 22          | 7.35  | 596              | 6           | 1        | 67.85 |
| 23          | 6.95  | 505              | 6           | 1        | 67.85 |
| 24          | 4.64  | 488              | 6           | 1        | 67.85 |
| 25          | 6.04  | 580              | 6           | 1        | 67.85 |
| 26          | 7.04  | 610              | 6           | 1        | 67.85 |
| 27          | 7.76  | 688              | 6           | 1        | 67.85 |
| 28          | 6.73  | 582              | 7           | 2        | 88.07 |
| 29          | 7.03  | 598              | 7           | 2        | 88.00 |
| 30          | 7.23  | 690              | 7           | 2        | 88.00 |
| 31          | 6.82  | 595              | 7           | 3        | 93.00 |
| 32          | 6.74  | 597              | 7           | 3        | 93.00 |
| 33          | 7.14  | 689              | 7           | 3        | 93.00 |
| 34          | 7.335 | 596              | 6           | 1        | 67.85 |
| 35          | 7.663 | 613              | 6           | 1        | 67.85 |

|    |       |     |   |   |       |
|----|-------|-----|---|---|-------|
| 36 | 7.065 | 704 | 6 | 1 | 67.85 |
| 37 | 4.118 | 490 | 7 | 2 | 88.00 |
| 38 | 6.633 | 507 | 7 | 2 | 88.00 |
| 39 | 6.740 | 597 | 7 | 3 | 93.00 |
| 40 | 5.426 | 495 | 7 | 2 | 88.00 |
| 41 | 6.94  | 523 | 7 | 2 | 88.00 |
| 42 | 6.07  | 568 | 7 | 2 | 88.00 |
| 43 | 6.22  | 612 | 7 | 2 | 88.00 |
| 44 | 7.34  | 615 | 7 | 2 | 88.00 |
| 45 | 7.75  | 706 | 7 | 2 | 88.00 |
| 46 | 4.82  | 489 | 7 | 3 | 93.00 |
| 47 | 4.33  | 490 | 7 | 3 | 93.00 |
| 48 | 7.33  | 596 | 6 | 1 | 67.85 |
| 49 | 4.12  | 505 | 7 | 3 | 93.00 |
| 50 | 6.64  | 522 | 7 | 3 | 93.00 |
| 51 | 7.04  | 614 | 7 | 3 | 93.00 |
| 52 | 6.53  | 597 | 7 | 3 | 93.00 |
| 53 | 7.04  | 614 | 7 | 3 | 93.00 |
| 54 | 7.45  | 705 | 7 | 3 | 93.00 |
| 55 | 7.76  | 688 | 6 | 1 | 67.00 |
| 56 | 7.35  | 596 | 6 | 1 | 67.00 |
| 57 | 7.76  | 688 | 6 | 1 | 67.00 |
| 58 | 7.19  | 596 | 6 | 1 | 67.00 |
| 59 | 7.66  | 613 | 6 | 1 | 67.00 |
| 60 | 7.37  | 718 | 6 | 1 | 67.00 |
| 61 | 7.55  | 688 | 6 | 1 | 67.00 |
| 62 | 6.06  | 704 | 6 | 1 | 67.00 |
| 63 | 7.55  | 688 | 6 | 1 | 67.00 |
| 64 | 7.65  | 704 | 6 | 1 | 67.00 |
| 65 | 7.38  | 796 | 6 | 1 | 67.00 |
| 66 | 6.52  | 582 | 7 | 2 | 88.00 |
| 67 | 7.03  | 598 | 7 | 2 | 88.00 |
| 68 | 7.43  | 690 | 7 | 2 | 88.00 |
| 69 | 6.83  | 598 | 7 | 2 | 88.00 |
| 70 | 5.34  | 615 | 7 | 2 | 88.00 |
| 71 | 6.75  | 706 | 7 | 2 | 88.00 |

|    |       |     |   |   |       |
|----|-------|-----|---|---|-------|
| 72 | 6.233 | 690 | 7 | 2 | 88.00 |
| 73 | 5.750 | 706 | 7 | 2 | 88.00 |
| 74 | 6.145 | 798 | 7 | 2 | 88.00 |
| 75 | 6.610 | 595 | 7 | 3 | 93.87 |
| 76 | 6.226 | 581 | 7 | 3 | 93.87 |
| 77 | 5.450 | 689 | 7 | 3 | 93.87 |
| 78 | 6.533 | 597 | 7 | 3 | 93.87 |
| 79 | 6.040 | 614 | 7 | 3 | 93.87 |
| 80 | 6.400 | 705 | 7 | 3 | 93.87 |
| 81 | 6.930 | 689 | 7 | 3 | 93.87 |
| 82 | 6.453 | 705 | 7 | 3 | 93.87 |
| 83 | 6.858 | 797 | 7 | 3 | 93.87 |

Table10 Shows the binding energy and interacting residues of substituents built with SHP-1calculated using autodock

| Compounds | Binding Energy | Interacting residue of SHP-1 | Interacting residue of ligands |
|-----------|----------------|------------------------------|--------------------------------|
| 2         | -9.10          | Gln83,Lys277                 | NH,O                           |
| 3         | -8.29          | Gln266,Gln83                 | O,NH                           |
| 5         | -8.58          | Gln83,Gln266                 | NH <sub>2</sub> ,O             |
| 6         | -9.23          | Gln83,Gln266                 | NH <sub>2</sub> ,O             |
| 8         | -8.44          | Lys277,Gln83                 | O,OH                           |
| 9         | -8.25          | Glu77,Asn274                 | OH,O                           |
| 11        | -8.74          | Gln83,Gln266                 | O,NH                           |
| 12        | -8.55          | Gln81,Lys277,                | NH,O                           |
| 18        | -8.66          | Asn274,Gln83                 | O,NH <sub>2</sub>              |
| 20        | -8.74          | Gln81,Gln83                  | OH,O                           |
| 21        | -8.41          | Glu77,Asn274                 | OH,O                           |
| 24        | -8.25          | Glu77,Glu83                  | O,NH                           |
| 37        | -8.47          | Gln81,Arg262                 | OH,O                           |
| 46        | -8.94          | Glu77,Lys277                 | NH <sub>2</sub> ,O             |
| 47        | -9.23          | Gln81,Gln83                  | NH <sub>2</sub> ,O             |
| NSC87877  | -7.1           | Lys277,Asp 278               | O                              |

## 4.7 Testing of substituent's using 3D QSAR model development

The substituent's which passed Lipinski filter and binding energy greater than NSC87877 from Table 9 and Table 10 were tested using 3D-QSAR model development. The inhibition constant calculated from autodock were converted into  $-\log k_i$  and  $-\log k_i$  was used as activity. The model generated showed statistically good results with  $r^2=0.8015$ ,  $q^2=0.7215$ . The statistical significance of model was also confirmed by a high value of Fischer ratio =47.8 and a very low value of RMSE =.49. One of the other parameter which signify model predictivity is Pearson R. Its value of .91 shows that correlation between predicted and observed activity for test compound is excellent. Table 10 shows QSAR statistical results and Table 11 shows observed and PHASE predicted activity. Table 11 shows the compounds which were used as test and compounds which were used as training their activity also the activity predicted by PHASE and PLS factor taken. Figure 10 shows the fitness graph between Phase activity and Phase predicted activity. The solid line in the plot indicates the hypothetical "best fit" line between the predicted and experimental activity. The scatter plot indicates a good relation between the predicted and experimental activity with  $r^2=0.80$ . From Figure 11 it can be seen that blue regions for hydrogen-bond donor, electron withdrawing feature and hydrophobic regions are present where we have substituted  $\text{NH}_2$ , F, Cl, Cl,  $\text{C}_2\text{H}_5$  respectively in 47 th compound (the most active compound) hence responsible for its activity.

Table 11 QSAR statistical results for the model

| PLS factors | SD   | $r^2$  | F    | P         | RMSE | $q^2$  | Pearson R |
|-------------|------|--------|------|-----------|------|--------|-----------|
| 1           | .142 | 0.7138 | 38.7 | 1.22e-005 | .39  | 0.563  | 0.871     |
| 2           | .108 | 0.8015 | 47.8 | 1.34e-007 | .49  | 0.7215 | 0.910     |

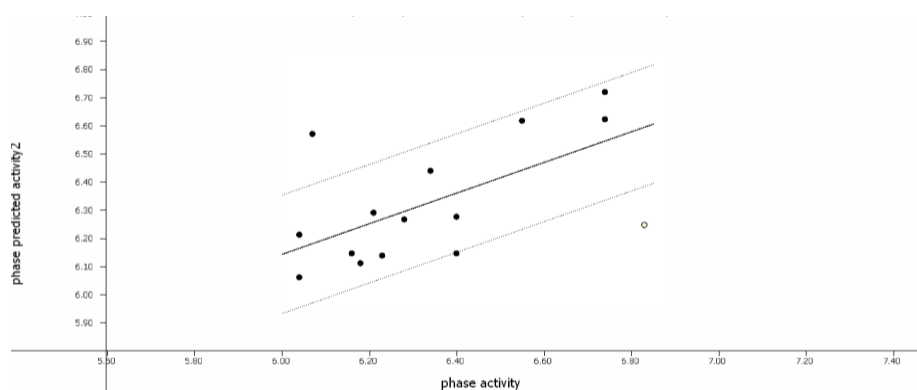


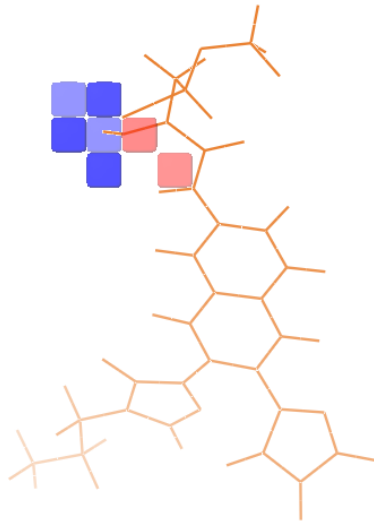
Figure 11 Fitness graph between observed activity versus PHASE predicted activity



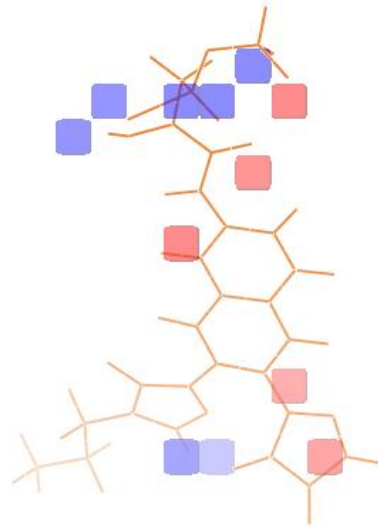
| Compounds | QSAR set | Activity | PLS factors | Phase predicted activity |
|-----------|----------|----------|-------------|--------------------------|
| 2         | Test     | 6.630    | 1,2         | 6.31,6.24                |
| 3         | Test     | 6.070    | 1,2         | 6.57,6.56                |
| 5         | Test     | 6.280    | 1,2         | 6.27,6.26                |
| 6         | Training | 6.740    | 1,2         | 6.66,6.62                |
| 8         | Training | 6.180    | 1,2         | 6.04,6.11                |
| 9         | Training | 6.040    | 1,2         | 6.03,6.06                |
| 11        | Training | 6.400    | 1,2         | 6.31,6.27                |
| 12        | Training | 6.230    | 1,2         | 6.27,6.13                |
| 18        | Training | 6.340    | 1,2         | 6.42,6.43                |
| 20        | Test     | 6.400    | 1,2         | 6.12,6.14                |
| 21        | Training | 6.160    | 1,2         | 6.12,6.14                |
| 24        | Training | 6.040    | 1,2         | 6.38,6.21                |
| 37        | Training | 6.210    | 1,2         | 6.22,6.29                |
| 46        | Training | 6.550    | 1,2         | 6.50,6.61                |
| 47        | Training | 6.740    | 1,2         | 6.61,6.71                |

Table12 Phase predicted activity

12a



12b



12c

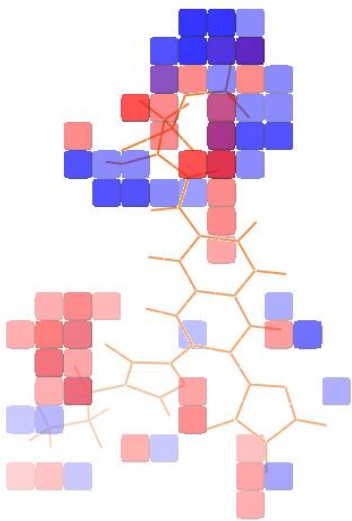


Figure 12 shows a hydrogen-bond donor effect, b: electron withdrawing feature, c: hydrophobic effect in the most active compound in 47. From Figure 11 it can be seen that blue regions for hydrogen-bond donor, electron withdrawing feature and hydrophobic regions are present where we have substituted  $\text{NH}_2$ , F, Cl, Cl,  $\text{C}_2\text{H}_5$  respectively in 47 compound hence responsible for its activity

## 5. DISCUSSION

Transplantation of stem cells from the original transplant recipient into secondary and tertiary irradiated recipients reconstitutes hematopoiesis with resultant normal life spans. Transplantation requires two essential properties proliferation to replenish the stem cell compartment (self-renewal) and lifelong production of blood (Pearce *W et al.*, 2008). During transplantation high number of HSC is needed as the cells reaching target eventually decreases. Hematopoietic stem cells (HSCs) play a key role in hematopoietic system that functions mainly in homeostasis and immune response. HSCs transplantation has been applied for the treatment of several diseases. However, HSCs persist in the small quantity within the body, mostly in the quiescent state. HSC maintenance, balance between self-renewal and proliferation are essential requirement for advancement of HSC expansion and transplantation in the future. Hematopoiesis and HSC development are the key role to improve efficient HSC expansion for the transplantations. So work here has aimed to increase the proliferation of HSC needed during transplantation. SHP-1 which acts as negative regulator of proliferation has been targeted. Both structure based (docking) and ligand based (Pharmacophore modeling and 3D-QSAR) approaches have been used for designing drug against SHP-1. The dataset of inhibitors built was used to carry out Pharmacophore modelling and QSAR studies. Phase module of Schrodinger suite was used. For the generation of pharmacophore model 24 compounds were considered. Minimum sites 4 and Maximum sites 7 were considered to have optimum combination of sites or features common to active compounds. The molecules were classed into active and inactive based on activity threshold for identifying the pharmacophore feature considering highest active molecule. On the basis of Survival score and rescoring of active and inactive molecule AAADRRR.190 with a survival score of 3.581 which is highest among all the hypothesis was selected. Survival score is a combination of Vector, Site, Volume and term for number of matches, a large value of Survival score indicates better fitness of active ligand on the common pharmacophore and validate the model. All the molecules showed good alignment with good fitness score ranging from 3.00 (for highest) to 1.95 (for lowest active). The 3D-QSAR dimensional analysis was carried out based on Pharmacophore alignment. For QSAR model generation dataset was randomly divided into training set of 17 compounds and 07 compounds in test set. As for good model generation 70% of compounds should be in training set and 30% in the test set. PLS factor was set as 03, the maximum number of PLS factors in each model can be 1/5 total number of training molecules. So it was taken to be 3 in our case. More the PLS factor values more will be the reliability of model. Various models have been generated and best model was selected. The reliability of the present 3D-QSAR analysis can be justified by the fact that all statistical measures are significant. The model generated showed statistically good results with  $r^2$  (Correlation Coefficient) = 0.897, correlation coefficient  $r$  measures how closely the observed data tracks the fitted regression line.  $q^2 = 0.8173$ .  $q^2$  is leave one out cross validation. It involves using a single observation from original sample as the validation determinant and remaining observation as training data. This is repeated such that each observation in the sample is used once as validation data. The statistical significance of model was also confirmed by a high value of Fischer ratio = 60.9 Fischer ratio

indicates while the fit of the data to the regression line is excellent, how can one decide if this correlation is based purely on chance. A very low value of RMSE =.11. One of the other parameter which signify model predictivity is Pearson R. Its value of .94 shows that correlation between predicted and observed activity for test compound is excellent. From Figure 8a it can be seen that H donor can be added to carbon just adjacent to acceptor shown in common pharmacophore. Figure 8b shows that electron accepting group can be added near one of the rings and to carbon chain. Figure 8c shows hydrophobic groups can be added near one of the rings. All these changes can lead to increase in biological activity of inhibitors. Based on these QSAR results several substitution were made and various groups were added to form some novel compounds with good biological activity. One of the main goals in drug discovery is identification of innovative small molecular scaffolds exhibiting high binding affinity and selectivity for the target together with a reasonable absorption, distribution, metabolism, excretion and toxicity (ADMET) profile, lead and/or drug likeness. Lipinski rule of 5 is a rule of thumb to evaluate drug likeliness or to determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans. The rule describes molecular properties important for a drugs pharmacokinetics in human body, including its ADMET. So lipinski filter was applied to the substituents for checking their likeliness to become a drug. Substituents which followed Lipinski rule and showed binding energy better than NSC87877 were further tested with QSAR model development. Novel inhibitors showing higher activity can be further tested for ADMET profiling and bioavailability. Then wet lab synthesis can be done.

## 5. CONCLUSION AND FUTURE PROSPECTIVE

.Signaling by stem cell factor and Kit, its receptor, plays important roles in hematopoiesis . The activated receptor becomes autophosphorylated at tyrosine residues that serve as docking sites for signal transduction molecules containing SH2 domains Shp-2 phosphatase and Shp-1 phosphatase bind to the phosphotyrosine at 568 and 570 residue in juxtamembrane region of C-kit. SHP-1 binds to the phosphotyrosine residue 570 of C-kit and negatively regulate proliferation of HSC. The purpose of this study was to use both ligand based drug designing and structure based drug designing to design inhibitors against SHP-1to increase proliferation of HSC. The dataset of inhibitors built was used to carry Pharmacophore modeling and QSAR so that more inhibitors can be designed and can be tested for their biological activity.The 3D-QSAR obtained from AAADRRR-190 hypothesis was found to be statistically good  $r^2=.89,q^2=.81$ ,Fischer ratio=60.9,PearsonR=.94.So it was easy to draw clear inference to design novel compounds for better biological activity from the QSAR model.The 3D-QSAR model built was used to explain how and to what extent electron withdrawing,hydrophobic and H-donor properties should be modified to achieve better inhibition.The QSAR visualization showed the favorable region for addition of H donor,Electron withdrawing groups and hydrophobic groups.Based on this model some substituents were built.The substituents physiochemical properties were validated using Lipinski rule of 5.The substituents which passed Lipinski filter were docked with SHP-1using autodock..Most of newly designed molecules were found to show interaction with Gln81,Gln83,Gln266,Asn274,Lys277,Asn278,Gln266 of SHP-1.The compounds which showed binding energy greater than NSC87877 and followed Lipinski rule of 5 were tested using 3D QSAR model generation.The QSAR model showed statistical significant results  $r^2=.80,q^2=.72,F=47.8$  thus validating that 15 componds built showed good biological activity.

One of the main goals in drug discovery is identification of innovative small molecular scaffolds exhibiting high binding affinity and selectivity for the target together with a reasonable absorption,distribution,metabolism,excretion and toxicity (ADMET) profile,lead and/or drug likeness.Such chemical entities are likely to be able to enter higher phases of the drug development process..Thus *in vitro* approaches can be used to investigate the ADMET properties of these newly designed chemical entities and to optimize selection of the most suitable candidates for drug development.

Thus the QikProp program can be further used to obtain the ADMET properties of the analogues.it will predict both physically significant descriptors and pharmaceutically relevant properties.

Also virtual screening on the basis of pharmacophore model can be done in order to propose various drug molecules,which have all the essential features.Then wet lab synthesis of these compounds as well characterization via techniques like  $H^1N.M.R$  and L.C.M.S can be done.

## REFERENCES

- Alison,MR; Poulsom,R ;Forbes (2002) . An introduction to stem cells. *J Path* 197: 419–423
- Alonso,A; Sasin,J; Bottini,N; Friedberg, I; Osterman ,A; Godzik,A;Hunter, T; Dixon, J ;Mustelin T (2004 ). Protein tyrosine phosphatases in the human genome . *Cell*.117, 699–711.
- Andersen, J.N; Mortensen, O. H; Peters, G. H., Drake, P. G; Iversen, L. F;Olsen, O. H.,Jansen, P. G; Andersen, H. S; Tonks, N. K; Moller, N. P. (2001).Structural andevolutionary relationships among protein tyrosine phosphatase domains. *Mol. Cell. Biol.* 21, 7117–7136.
- Andrew,W. (2005).Stoker Protein tyrosine phosphatases and signaling *Journal of Endocrinology* .185, 19–33.
- Bongso,A; Eng Hin Lee (2005). Stem cells: their definition, classification and sources.*Stem Cells: From Benchtop to Bedside* .12, 8–12.
- Burt, RK; Loh, Y; Pearce ,W. (2008). Clinical applications of blood-derived and marrow-derived stem cells for nonmalignant diseases. *JAMA.* 8, 925–936.
- Dixon,JE; Denu,JM.(1998). Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr Opin Chem Biol.* 5, 633-41
- Dror. O;(2006). Predicting molecular interactions in silico. An updated guide to pharmacophore identification and its applications to drug design. *Front Med.Chem.*3, 551-584
- Gall, S.; Zseb, K.M; Geissler, E.N.(1994).The kit ligand, stem cell factor, *Adv. Immun.* 55 , 1–96.
- Jia, Z; Barford, D; Flint, A.J; N.K.Tonks (1995). Structural Basis for Phosphotyrosine Peptide Recognition by PTP1B.*Science.* 268,1754-1758.
- John G. Topliss.(1983). *Quantitative Structure-Activity Relationships of Drugs*, Academic Press, New York .337, 1–13
- Kent,D; Copley, M; Benz ,C; Dykstra, B; Bowie, M; Eaves C .(2008). Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway. *Clin. Cancer Res.*7,1926–30.
- Kozlowski, ML; Larose, F; Lee, D.M; Rottapel,R; Siminovitch, KA.(1998) . SHP-1 binds and negatively modulates the c-Kit receptor by interaction with tyrosine 569 in the c-Kit juxtamembrane domain. *Mol. Cell. Biol.* 18, 2089–2099.

Mustelin, T; Tautz, L; Page, R. (2005). Structure of the hematopoietic tyrosine phosphatase (HePTP) catalytic domain: structure of a KIM phosphatase with phosphate bound at the active site. *J. Mol. Biol.* **354**, 150–163

Neel, B. G; Tonks, N. K. (1997) .Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* **9**, 193–204

Pannifer, A; Flint, A; Tonks, N; Barford D. (1998). Visualization of the Cysteinyl-phosphate Intermediate of a Protein-tyrosine Phosphatase by X-ray Crystallography. *The Journal of Biological Chemistry* .**273**, 10454-10462.

Perkins, R.; Fang, H; Tong, W; Welsh, W.J (2003) Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.* **22**, 1666–1679.

Robert F. Gould . Biological Correlations The Hansch Approach, *Advances in Chemistry .Series, No. 114.*

Robert Roskoski Jr (2005). Signaling by Kit protein-tyrosine kinase—The stem cell factor receptor. *Biochemical and Biophysical Research Communications* **337** 1–13

Rönstrand L ( 2004). Signal transduction via the stem cell factor receptor/c-Kit. *Cell. Mol. Life Sci.* **61** ,2535–48.

Simone.B.(2009) Pharmacophore modeling: A continuously evolving tool for computational drug design. *New perspectives in medicinal chemistry. 1*, 13-23

Sun, JP; Zhang ,ZY; Wang ,WQ .(2003). An overview of the protein tyrosine phosphatase superfamily. *Curr Top Med Chem* **3** ,7-12

Tigans, T; Anton M. (2007) Protein tyrosine phosphatase function: the substrate perspective *Biochem. J.* **402**,

Tonks, N. K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **7**, 833–846

Tuch BE (2006). Stem cells—a clinical update *Australian Family Physician* **35** 719–21

Wang, L. L; Blasioli, J; Plas, D. R; Thomas, M. L; Yokoyama, W. M (1999). Crystal Structure of Human Protein Tyrosine Phosphatase SHP-1 in the Open Conformation . *J. Immunol.* **162**, 1318–1323.

Wolber.G ; Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today.* **13** 23-29.

Yang J, Liu L, He D, Song X, Liang X, Zhao ZJ, Zhou GW. ; “Crystal structure of human proteintyrosinephosphatase SHP-1. J Biol Chem. 2003”; **278(8)**:6516–6520

Yang J; Liang X;Niu T;Meng W;Zhao Z; Zhou GW.(1998). Crystal structure of the catalytic domain of protein-tyrosine phosphatase SHP-1. J Biol Chem .**273** 28199–28199.

Yvonne C. ,Martin, A.; “ Practitioner's Perspective of the Role of Quantitative Structure Activity Analysis in Medicinal Chemistry, *J. Med. Chem.*,” **24**, 229 (1981)



## APPENDIX

### 3.5 Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules. Molecular docking is thought of as an optimization problem, which describes the “best-fit” Orientation of a ligand that binds to a particular protein of interest. It is similar to “*lock-and-key*” model, where one is interested in finding the correct relative orientation of the “*key*” which will open up the “*lock*”. Thus the protein can be thought of as the “*lock*” and the ligand can be thought of as a “*key*”. Docking is important as a binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking is the most commonly used in the field of drug design-most drugs are small organic molecules, and docking may be applied to:

**Hit identification**-docking combined with a scoring function can be used to quickly screen large databases of potential drugs in silico to identify molecules that are likely to bind to protein or target of interest.

**Lead optimization**-docking can be used to predict in where and in which relative orientation a ligand binds to a protein (also referred to as the binding mode or pose). This information may be used to design more potent and selective analogs. (Donald *et al.*, vol1)

PyRx is a Virtual Screening software for Computational Drug Discovery that can be used to screen libraries of compounds against potential drug targets. PyRx enables Medicinal Chemists to run Virtual Screening from any platform and helps users in every step of this process - from data preparation to job submission and analysis of the results

AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.

### 3.6 Pharmacophore modelling

The concept of pharmacophore was first introduced in 1990 by Paul Ehrlich, who defined the pharmacophore as “a molecular framework that carries (phoros) the essential features responsible for a drug's (pharmakon) biological activity”. **Ligand-based pharmacophore modeling** has become a key computational strategy for facilitating drug discovery in the absence of a macromolecular target structure. It is usually carried out by extracting common chemical features from 3D structures of a set of known ligands representative of essential interactions between ligands and a specific macromolecular target. In general, pharmacophore generation from multiple ligands (usually called training set compounds) involves two main steps:

Creating the conformational space for each ligand in the training set to represent conformational flexibility of ligands

Aligning the multiple ligands in the training set and determining the essential common features to construct pharmacophore models (Simone, B.2009).

**Molecular alignment** is the major challenging issue in ligand based pharmacophore modeling. The alignment methods can be classified into two categories in terms of their fundamental nature: point –based and property-based approaches (Wolber, G.2008).

The point (**in the point-based method**) can be further differentiated as atoms, fragments or chemical features. In point-based algorithm, pairs of atoms, fragments or chemical feature points are usually superimposed using a least square fitting. The biggest limitation of these approaches is the need for predefined anchor points because the generation of these points can become problematic in the case of dissimilar ligands.

The property-based algorithms make use of molecular field descriptors, usually represented by set of Guassain functions, to generate alignments. The alignment optimization is carried out with some varaint of similarity measure of the intermolecular overlap of the Gaussians as the objective function.

Another challenging problem lies in the practical task of proper selection of training set compounds .This problem, apparently being simple and non technical, often confuses users, even expereinced ones. it has been demonstrated that the type of ligand molecules ,the size of the dataset and its chemical diversity affect the final generated pharmacophore model considerably (Dror, O. 2006).

### 3.7 Quantitative Structure Activity Relationship (QSAR)

**QSAR** stands for “quantitative structure-activity relationships”, is a method that relates chemical structure to biological or chemical activity using mathematical models . If the activity of a set of ligands can be determined, a model can be constructed to describe this relationship. Unlike a pharmacophore model, which encodes only the essential features of an active ligand, the QSAR model allows one to determine the effect of a certain property on the activity of a molecule. For example, the QSAR model may reveal a property to have a highly negative, or alternatively a weak positive effect on ligand activity. Such information is not available using a pharmacophore model (Perkinson *et al.*, 2003).

Quantifying the structure and activity of a ligand is important in the modeling process. Structure quantification is not a trivial problem, since a structure cannot be represented by a mere value. Instead, a set of properties, usually known as the “descriptors”, is computed from the structure and used to quantify it. By using structural descriptors as independent variables and activity as a dependent variable, a model can be built to describe the relationship between the two.

### 3.7.1 Building a QSAR Model:

The process of constructing a QSAR model can be summarized as follows: First, ligands and their activities are collected. Descriptors are calculated and selected before a mathematical modeling method is chosen and the ligand data are then used to construct the QSAR models. After the models are completed, they are tested by internal and external validation procedures. Only then can a QSAR model be used in any practical **applications, such as predicting the activity of a novel compound**. As is the case when building a pharmacophore model, the active ligand set must be gathered from molecular databases or from literature searches before QSAR modeling begins. The process requires not only the collection of ligand structures but also of their activities. Generally, IC<sub>50</sub>s (half maximal inhibitory concentration), EC<sub>50</sub>s (half maximal effective concentration) and  $K_i$  values (inhibition constant) are commonly used to quantify drug activity. However, the quantification of ligand activity as used in QSAR is not limited to pharmacokinetic parameters. Other activity indexes can also be incorporated into model depending on the phenomena one wishes to predict. In addition to structure verification as described in the section on pharmacophore model construction, ligand activity data should also be checked. All activity data should come from the same experimental procedure or assay, and it is preferable if the data comes from the same laboratory, and even the same researcher (Yvonne *et al.*, 1981)

Before a QSAR model can be built, ligand structure descriptors should be ascertained or calculated. Some descriptors obtained directly from data sources or calculated using simple arithmetic operations take into account the specific number of atoms, molecular chain length, molecular mass, *etc.* However, other descriptors may require complex computation, for example pharmacophore-based descriptors molecular field descriptors, which are derived from the interaction of probes and molecules and used in **CoMFA** and **CoMSIA**. It is important that the descriptors are related to the biological or chemical activity which the model will be used to predict. In other words, if a descriptor is not related to activity, one should avoid incorporating the descriptor into the modeling process.

After the activity index (the dependent variable) and descriptors (the independent variables) are prepared for each ligand, a variable selection method and a modeling method can be selected, and a model is built. The selection process If two descriptors represent a similar biological or chemical parameter, one of them should be disregarded. In order to select descriptors, genetic algorithms principle component analysis, artificial neural networks and  $k$ -nearest neighbor approaches can all be used. If a linear model is assumed, some conventional statistical methods, such as the partial least squares method and multiple linear regression can be used. If a nonlinear model is preferred on the other hand, machine learning methods like artificial neural networks or support vector machines can be applied. The main differences among the frequently used QSAR algorithms reside in their means of descriptor generation. For example, most QSAR algorithms, like CoMFA, CoMSIA use similar linear statistical models to explore the relationship

between activity and descriptors, which are calculated by different processes. In CoMFA and CoMSIA, pre-aligned molecules are put onto a grid, or lattice. The descriptors are calculated by the interaction of the molecule and a probe is placed at each intersection of the lattice. The differences between CoMFA and CoMSIA are in the use of different probes and interaction-calculating functions.

In CoMFA, only probes representing steric and electrostatic interactions can be used. In CoMSIA, probes representing hydrophobic and hydrogen bond interactions, in addition to CoMFA probes may be selected. In addition, CoMSIA uses a Gaussian-type function for calculating prober-molecule interaction. By using such a smooth function, the result value is more reasonable than the function used in CoMFA, and defining a cut-off limit to remove invalid values is no longer required.. The fit value describes the goodness of alignment between a ligand and a pharmacophore model and is obtained from a pharmacophore model generated and optimized using known structure and activity data. The model must then be validated before it can be used to predict activity. There are some popular methods used to validate a QSAR model including internal validation approaches (such as the “leave-one-out” or “leave-n-out” cross validation methods , and external validation approaches. In cross validation, one (leave-one-out) or more (leave-n-out) ligand of the training set is excluded. The excluded data is predicted by the model constructed with reduced training set data. These steps are repeated until all data has been excluded and predicted, and the power of a model is determined by the accuracy of prediction. External validation is a widely used method, and is considered important in the QSAR building pipeline. In external validation, the capability of the model is tested using data which is not included in the training set, in contrast to internal validation, which utilizes data taken from the training set to validate the model. In most of the studies, both internal and external validations are performed to ensure the reliability of the model. After the model has passed these strict validation tests, it can be used to predict the activity of novel molecules (John *et al.*, 1983).

### 3.7.2 Statistical concepts

A QSAR generally takes the form of a linear equation

$$\text{Biological Activity} = \text{Const} + (C1 P1) + (C2 P2) + (C3 P3) + \dots$$

where the parameters P1 through Pn are computed for each molecule in the series and the coefficients C1 through Cn are calculated by fitting variations in the parameters and the biological activity

#### a) Standard deviation s:

The standard deviation of the data, s, shows how far the activity values are spread about their average. This value provides an indication of the quality of the guess by showing the amount of variability inherent in the data The standard deviation is calculated as shown below

$s = \sqrt{(\text{compound activity-average activity})^2 + (\text{compound activity-average activity})^2 + \dots / (n-1)}$   
 In the above equation n represents no of compounds. This formula gives how to calculate standard deviation

**b) Correlation coefficient r:**

Variation in the data can be quantified by correlation coefficient r which measures how closely the observed data tracks the fitted regression line

$$r^2 = \frac{\text{Sum-of-Squares of the deviations from the regression line}}{\text{Sum-of-Squares of the deviations from the mean}}$$

$$r^2 = \frac{\text{Regression Variance}}{\text{Original Variance}}$$

Where original variance =  $(\text{compound activity-average activity})^2 + (\text{compound activity-average activity})^2 + \dots$

And regression variance = original variance - variance around the line

$r^2 = 0$  and  $1$ .  $r^2$  of  $0$  means that there is no relationship between activity and parameter  $r^2 = 1$  mean there is perfect correlation.

**c) F statistics:**

While the fit of the data to the regression line is excellent, how can one decide if this correlation is based purely on chance

F statistics is calculated as

$$F_{1,n} = (n-2) \frac{r^2}{1-r^2}$$

This value can be checked in statistical table to determine the significance of regression equation.

**d) Leave one out cross validation  $q^2$ :**

It involves using a single observation from original sample as the validation determinant and remaining observation as training data. This is repeated such that each observation in the sample is used once as validation data.

$$q^2 = (1 - \sum(Y_{\text{predicted}} - Y_{\text{actual}})^2 / \sum(Y_{\text{actual}} - Y_{\text{mea}})^2)$$

where Y is the activity.  $q^2$  should be close to  $r^2$ .





