IDENTIFICATION OF TPX INHIBITORS AGAINST PATHOGEN-HOST INTERACTIONS

by

Utku Deniz

B.S., Chemical Engineering, Boğaziçi University, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2014

ACKNOWLEDGEMENTS

It is my pleasure to thank all the people who made this thesis study possible. Foremost, I would like to express my deep gratitude to my supervisor, Assoc. Prof. Elif Özkırımlı Ölmez and my co-supervisor, Prof. Kutlu Özergin Ülgen, for the guidance and encouragement they provided at all levels of this research through these two years.

During the period of two years, many friends are helpful to color my life. I have to acknowledge to my colleague, Begüm Alaybeyoğlu, for her assistances in many aspects whenever I needed. I would also like to thank Sinan Koç for his friendship, entertainment and support. I also thank the people whom I met at the university, and shared experiences and memories in life, Merve, Barış, Elif, Deniz, Selcen, and Emre.

Last but the most important, I owe more than thanks to my family members, which includes my parents and an elder sister, for their financial support and encouragement throughout my life. Without their support, it is impossible for me to finish my college and graduate education seamlessly. I dedicate this thesis study to my marvelous family.

ABSTRACT

IDENTIFICATION OF TPX INHIBITORS AGAINST PATHOGEN-HOST INTERACTIONS

Antibiotic resistance is a major health issue as due to the emergence of multidrug resistant bacteria. Thus, there is an increasing need for the identification of new drug targets and discovery of novel scaffolds in the combat with infectious diseases. One mechanism that can be targeted is the pathogen-host interaction. Effector proteins of pathogenic bacteria invade the host cell through type III secretion system (T3SS) and disrupt the cellular signaling mechanism. Previous studies demonstrated that thiol peroxidase, Tpx, is functional in the assembly of T3SS and its inhibition by salicylidene acylhydrazides prevents the secretion of pathogenic effectors. The aim of this thesis study is to carry out virtual screening and molecular docking to identify potential Tpx inhibitors. Both ligand-based and structure-based pharmacophore models were developed to screen the ZINC database of 500,000 compounds based on 3D similarity to the chosen pharmacophore hypotheses. Molecular docking of filtered 10,000 compounds was performed with Glide module of Maestro molecular modeling package to predict binding modes of the ligands. Top scoring hits were further analyzed by efficiency indices, strain energy corrections, and their absorption, distribution, metabolism and excretion (ADME) and druglikeness properties. Common scaffolds of the selected clusters were used for substructure search and 31 hits with high docking scores, fitness values to the hypothesis and binding efficiency index (BEI), low molecular weight and high percentage of human oral absorption (HOA) values were obtained. As a final outcome, eight ligands with different chemotypes including n-benzyl formamide, 1,2-dimethoxybenzene and phenoxyethanol were proposed as potential inhibitors of Tpx after induced fit docking and selectivity analysis.

ÖZET

PATOJEN-KONAK HÜCRE ETKİLEŞİMLERİNE KARŞI TPX İNHİBİTÖRLERİNİN BELİRLENMESİ

Antibiyotik direnci ilaca karşı dirençli bakterilerin çoğalmasıyla ciddi bir sağlık sorunu haline gelmiştir. Bu yüzden, bulaşıcı hastalıkla savaşta, yeni ilaç hedeflerinin ve yeni kimyasal yapıların belirlenmesine artan bir ihtiyaç bulunmaktadır. İlaç hedefi olarak kullanılan mekanizmalardan biri patojen-konak hücre etkileşimleridir. Patojenik bakterilerin efektör proteinleri, konak hücreyi tip III salgı sistemi vasıtasıyla istila ederler ve hücresel sinyal mekanizmasını bozarlar. Tiol peroksidaz proteininin tip III salgı sisteminin kurulmasında işlevi olduğunun ve salisilidin açilhydrazit türevlerinin bu proteinin fonksiyonlarını bloke ettiği geçmişteki çalışmalarla gösterilmiştir. Bu çalışmanın amacı sanal tarama ve moleküler kenetleme yöntemleriyle potansiyel Tpx inhibitörleri belirlemektir. Hem mevcut ligand yapılarını hem de protein yapısını temel alan farmakoför modeller oluşturulmuş ve seçilen hipoteze üç boyutlu benzerliğe göre ZINC veri tabanı taranmıştır. Tarama sonucunda elde edilen on bin tane molekülün bağlanma bölgelerini tahmin etmek için Maestro moleküler modelleme paketinin Glide modülü ile moleküler kenetleme gerçekleştirilmiştir. Doklamadan en yüksek skoru elde eden bileşikler, etkinlik katsayıları, gerginlik enerji düzeltimi, emme, dağıtım, metabolizma ve boşaltım (ADME) ve ilaç benzerliği özelliklerine göre detaylı olarak incelenmiştir. Seçilen kümelerin ortak yapıları kullanılarak tekrar tarama yapılmıştır ve 31 tane bileşik yüksek doklama skoru, hipoteze olan uygunluk değerleri ve bağlanma etkinlik indisleri, düşük molekül ağırlığı ve yüksek ağızdan emilim yüzdesi baz alınarak seçilmiştir. İndüklenmiş doklama ve seçicilik analizlerin de yapılarak, n-benzil formamit, 1,2-dimetoksibenzen and fenoksietanol içeren değişik kimyasal yapılardaki sekiz ligand potansiyel Tpx inhibitörleri olarak önerilmektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two dimensional
3D	Three dimensional
AccptHB	Hydrogen bond acceptor
ADME	Adsorption, distribution, metabolism and excretion
AhpE	Alkyl hydroperoxide reductase E
BCP	Bacterioferritin comigratory protein
BEI	Binding efficiency index
DonorHB	Hydrogen bond donor
FISA	Hydrophilic component of total solvent accessible surface area
FOSA	Hydrophobic component of total solvent accessible surface area
Grx	Glutaredoxin
HOA	Human oral absorption
IC_{50}	Half maximal inhibitory concentration
Mol_MW	Molecular weight in Da
MOE	Molecular operating environment
NMR	Nuclear magnetic resonance
PDI	Protein disulfide isomerase
PISA	Phi component of total solvent accessible surface area
PLS	Partial least squares
Prx	Peroxiredoxin
PSA	Polar surface area of polar nitrogen and oxygen atoms
QPlogPo/w	Predicted octanol/water partition coefficient

QSAR	Quantitative structure activity relationship
RMSE	Root mean square error
SAR	Statistical molecular design
SASA	Total accessible solvent surface area in square angströms
SD	Standard deviation
SE	Sedimentation equilibrium
SEI	Surface efficiency index
SP	Standard precision
T3SS	Type III secretion system
Tpx	Thiol peroxidase
Trx	Thioredoxin
TrxR	Thioredoxin reductase
TSA	Thiol specific antioxidant
vdW	Van der Waals
WPSA	Weakly polar component of total solvent accessible surface area
XP	Extra precision

1. INTRODUCTION

Yersinia species (*Yersinia pestis*, *Yersinia pseudotuberculosis*, and *Yersinia enterocolitica*), known to be facultative parasites, induce infectious diseases such as Yersiniosis [1] and plague, as well as autoimmune diseases like reactive arthritis and Crohn`s disease [2, 3]. These pathogenic organisms utilize the type III secretion system (T3SS) to transmit their powerful effectors into the host cytoplasm. These effectors inhibit the cellular signaling mechanism and prevent the defensive response in the host cell [4].

A class of salicylidene acylhydrazide compounds were shown to inhibit T3SS [5, 27] by binding to thiol peroxidase, Tpx, which is functional in the assembly of T3SS, in *Yersinia pseudotuberculosis*. Inhibition of Tpx by salicylidene acylhydrazides prevents the expression of type III secretion system [6]. Tpx enhances the bacterial defense system against reactive oxygen species by utilizing reducing forms of thioredoxin (Trx1) and thioredoxin reductase (TrxR) to reduce alkyl hydroperoxides. It also plays a key role in activation of transcription factors and signaling [7]. The structure of Tpx has been determined by crystallography [8] and the active site residues that contribute to ligand binding have been determined by NMR [6]. In this thesis, the aim was to discover novel Tpx inhibitors using computational techniques.

Tpx inhibitors were identified using a workflow that included ligand-based or structure-based pharmacophore modeling followed by molecular docking and substructure search. The pharmacophore hypotheses were used to screen 500,000 compounds from the ZINC database. Multistep docking of filtered molecules with the pharmacophore model was performed by taking ligand flexibility into account. Then, the selected compounds with high docking scores were filtered using strain energy corrections and ADME properties. The compounds were divided into clusters based on the pairwise similarity value, which is higher than 0.4. Common scaffolds of selected clusters were used for substructure search. The druglike compounds from different chemotypes with high selectivity toward Tpx were further analyzed by induced fit docking and interaction analysis.

This thesis is organized as follows: Theoretical Background section includes information about type III secretion system, the structure and active binding site of Tpx, previous work on the inhibition of T3SS and Tpx, and virtual screening methods. The computational tools utilized for all virtual screening applications throughout the study including hypothesis selection, docking and post-docking applications are explained in the Material and Methods section. Computational results of two different pharmacophore models along with structure-based and ligand-based multistep docking with GLIDE are presented in detail in the Results and Discussion section. Conclusions and Recommendations part covers the summary of general workflow of the study, the significant results about proposed compounds and recommendations for future work.

2. BACKGROUND

2.1. Type III Secretion System Mechanism

Pathogenic *Yersinia* organisms (*Yersinia pestis*, *Yersinia pseudotuberculosis*, and *Yersinia enterocolitica*) utilize an efficient method for transmitting their powerful effectors into cytoplasm of the host immune system. Then, these effectors modulate host cellular functions and prevent the cells` response to infection. This secretion system consists of three basic components: the Ysc injectisome, which spreads into the bacterial membranes, the Yop effectors, and the Yop translocators, which are used to deliver the effectors across the eukaryotic plasma membrane [9].

The ability of the type III secretion systems to deliver virulence factors as effector proteins directly into host cells was first discovered for Yersinia effector proteins. The proteins are secreted directly from the bacterial cell into the eukaryotic cell, also known as “the host” cell [10]. In the widely known Gram-negative bacteria, about 20 proteins play a role in type III secretion and among these proteins, nine are the common ones and they display a high sequence similarity [11].

Type III secretion mechanism, which is shown in Figure 2.1 consists of a base, needle, tip and translocon [12]. Between inner and outer membranes of gram-negative bacteria, the base is placed and it includes several membrane rings. These porous rings serve as a base for the needle structure. The needle structure, or injectisome, exists between the cytoplasm of gram-negative bacteria and host cell, and facilitates the delivery of bacterial effector proteins. The needle passes through both inner and outer membranes of bacteria. In order to transmit bacterial effector proteins to the target cell, the tip protein, which is placed in the end of the needle, is employed and it constitutes a small hole on the cell. First effectors that are secreted into the host cell cytoplasm form a channel called translocon within the host membrane and other effectors are inserted through this translocon. The bacterial effectors manipulate the actin polymerization system of the host cell. Actin is a component of the cytoskeleton and it also plays a significant role in mobility and in changes in cell shape. By means of these effectors, the bacterium can

utilize the host cell`s machinery. After the bacterium enters the host cell it is able to secrete other effectors more easily and it can penetrate neighboring cells and quickly infect the whole issue [12, 13].

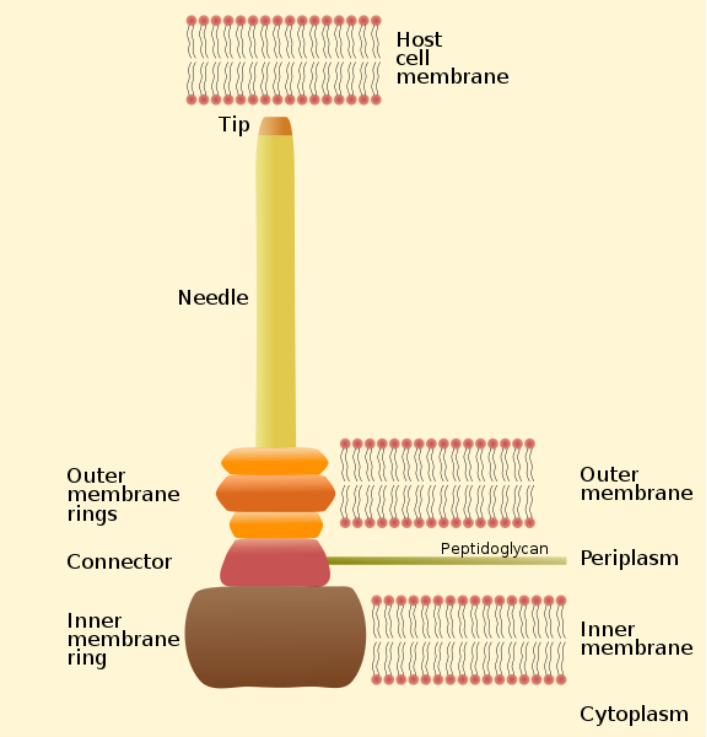


Figure 2.1. Type III secretion system apparatus [12].

2.2. Thiol Peroxidase (Tpx)

Thiol peroxidase (TPX) is functional in the assembly of T3SS. Tpx is a member of the Thioredoxin (TRX) superfamily, Peroxiredoxin (PRX) family and Atypical 2-cys PRX subfamily [14], and it is part of an oxidative stress defense system that utilizes thioredoxin and thioredoxin reductase to reduce alkyl hydroperoxides. Furthermore, it is an antioxidant which inhibits hydrogen peroxide response. Tpx is also involved in activation of transcription factors and signaling [15].

Thioredoxin superfamily includes a large and diverse group of proteins which have thioredoxin-fold domain. Many members of TRX superfamily have a redox active CXXC motif and interact with two substrates in two distinct steps. Some families in this superfamily including thioredoxin, glutaredoxin (GRX) and protein disulfide isomerase

(PDI) function as protein disulfide oxidoreductases (PDO) by altering the redox state of target proteins via the reversible oxidation of their active site dithiol. Thiol peroxidase as the member of PRX family does not function as PDO [16, 17].

Peroxiredoxin family is composed of different classes of peroxiredoxins as PRX5-like, Typical 2-cys PRX, 1-cys PRX, Bacterioferritin Comigratory Protein (BCP), Alkyl Hydroperoxide Reductase E (AhpE) and Atypical 2-cys PRX [18]. They have no cofactors such as metals or prosthetic groups as the distinction criterium from other peroxidases. Peroxiredoxins generally use thioredoxin to recharge after reduction reactions. The oxidized form of Prx is inactive and it requires donation of electron from reduced Trx in order to restore its catalytic activity. The first step of catalytic reaction, which is common to all peroxiredoxins, is the nucleophilic attack by the peroxidatic cysteine on the peroxide. The oxygen-oxygen bond is cleaved and cysteine sulfenic acid intermediate is formed. The second step of the reaction is the resolution of the intermediate and it distinguishes the Prx subfamilies. The presence or absence of the resolving cysteine constitutes the 2-cys or 1-cys peroxiredoxin types. The resolving cysteine of 2-cys Prxs is placed either on the same chain (atypical) or on the second chain (typical) of a functional homodimer [19, 20].

Atypical 2-cys Prx subfamily is known as Trx-dependent thiol peroxidase (Tpx). The genealogical tree of thiol peroxidase is displayed in Figure 2.2. This family is composed of

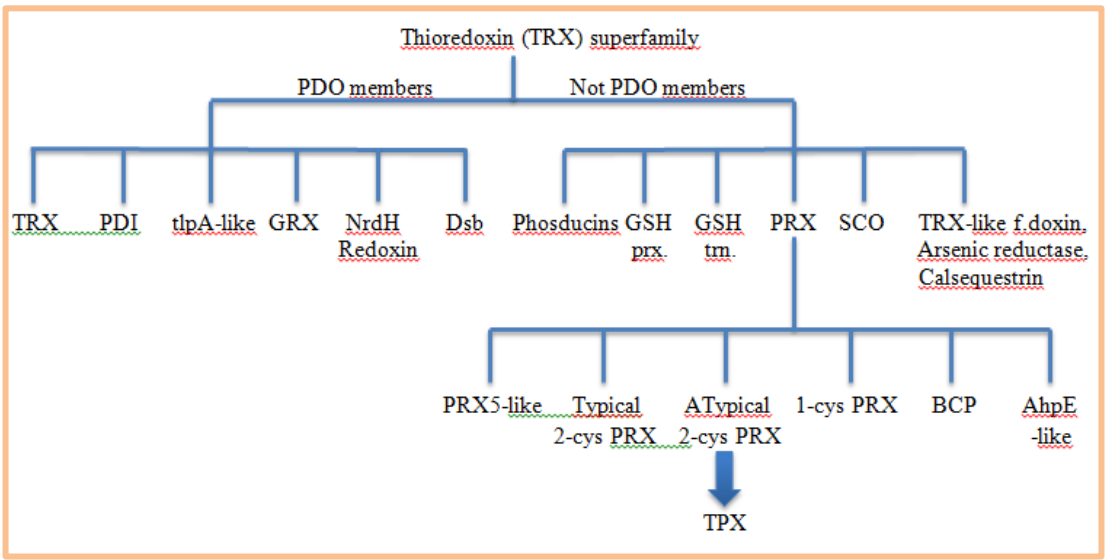


Figure 2.2. The genealogical tree of thiol peroxidase.

peroxiredoxins which include peroxidatic and resolving cysteines and are similar to the homodimeric thiol specific antioxidant (TSA). The oxidized or peroxidatic cysteine (C_p) is resolved via formation of intramolecular disulfide bond with conserved resolving cysteine (C_r) on the same chain. Tpx is a bacterial periplasmic peroxidase which differs from other peroxiredoxins by displaying substrate specificity toward alkyl hydroperoxides over hydrogen peroxide. It enhances the bacterial defense system against reactive oxygen species [21]. Its functionality depends on the reducing forms of Thioredoxin (Trx1) and Thioredoxin reductase (TrxR). Tpx consists of three cysteine residues as C61, C95 and C82. The resolving cysteine (C61) and the peroxidasic cysteine (C95) form the redox active disulphide bond [22]. The third cysteine (C82) does not play an important role in the redox activities of Tpx and does not form any covalent interactions [23]. In the oxidized form of Tpx, the formation of disulphide bond allows substrate access, whereas in the reduced form, C61 blocks the active site region [23].

Tpx regulates the expression of type III secretion system in order to invade the host cell cytoplasm. Tpx was determined to be present solely as a dimer in solution via sedimentation equilibrium (SE) data [23]. Salicylidene acylhydrazide class of compounds has been shown to target Tpx [6]. The three dimensional structure of Tpx was determined by crystallography [8] and the active site residues that contribute to ligand binding were determined by NMR [6]. In this thesis study, Tpx in its oxidized dimeric form (PDB code: 3ZRD) in *Yersinia pseudotuberculosis* is used in docking calculations. The crystal structure of Tpx (Figure 2.3) suggests that the binding pocket is mostly hydrophobic.

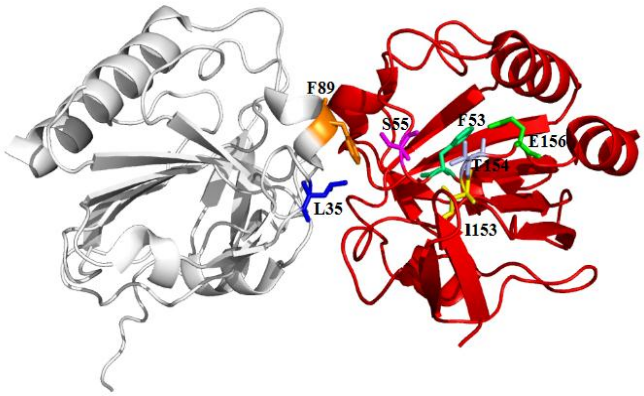


Figure 2.3. The Tpx dimer and the active site residues at the dimer interface. This figure was generated by Pymol [24].

2.3. Previous Work on Inhibition of Type III Secretion System and Tpx in *Yersinia pseudotuberculosis*

There are several studies in literature regarding inhibition of type III secretion mechanism and Tpx. O-acyl salicylanilides were first identified as potential inhibitors against type III secretion system in *Yersinia pseudotuberculosis* [25]. These molecules were served as a starting point for the identification of novel ligands that target virulence factors. The structure of O-acyl salicylanilide via ACD/ChemSketch software is displayed in Figure 2.4, in which Z is a substituted phenyl ring of 6 to 30 carbon atoms including substituents, R is a substituted or unsubstituted alkyl of phenyl group of 2 to 30 carbon atoms including substituents and M is a radical group consisting of –C=N, –F·, –NO2, –H lower alkyl or haloalkyl.

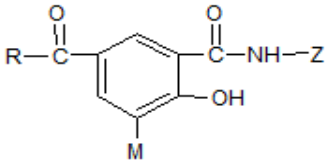


Figure 2.4. The structure of O-acyl salicynilide.

23 small molecules of a class of acylated hydrazones of different salicylaldehydes, (Figure 2.5), that target type III secretion system of *Yersinia pseudotuberculosis* in vitro have been identified by screening commercially available molecules with gene reporter assays. Inhibitory concentrations of these molecules were also reported [5].

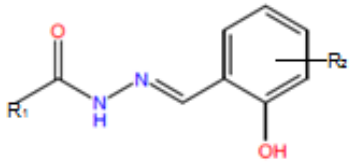


Figure 2.5. The structure of salicylidene acylhydrazide.

In the study performed by Anna M. Kauppi and her colleagues, 27 molecules of a class 2-arylsulfonylamino-benzanilides were examined for inhibition of T3SS in *Yersinia* organisms. This set of compounds was found to be selective for T3SS in *Yersinia pseudotuberculosis* but had no effect on bacterial growth [26].

Small molecule inhibitors of T3SS were also been identified by statistical molecular design (SAR) and multivariate quantitative structure-activity relationship (QSAR) analysis. 50 salicylidene acylhydrazides were synthesized in high purity. Based on biological evaluation, 18 compounds were classified as active because they showed at least 40% inhibition at 50 μM concentration [27].

In 2012, three classes of compounds were retrieved to prevent T3SS at low concentrations without affecting growth of the bacteria. These are sulfonylamino benzanilides, salicylanilides and salicylidene acylhydrazides. It was determined that the salicylidene acylhydrazide class of compounds unlike sulfonylamino bezanilides and salicylanilides influence both secretion and motility without affecting transcription [28].

In the same year, a model of Tpx with a salicylidene acylhydrazide derivative, ligand ME0052 (Figure 2.6), has been built by docking with Molecular Operating Environment (MOE) [23]. One hydrogen bond between p-hydroxyl of ME0052 and Ile 153 carbonyl of Tpx, which induced to the chemical shift in the neighboring Thr 154 amide, was observed.

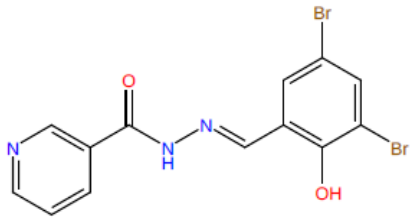


Figure 2.6. The salicylidene acylhydrazide class of ligand ME0052 [23].

In this thesis study, 27 active salicylidene acylhydrazides which display at least 40% inhibition at 50 μM ligand concentration (9 compounds from [5] and 18 molecules from [27]) were utilized for building the ligand-based pharmacophore model. In literature, inhibition percentages at different ligand concentrations were available [5, 27] and half maximal inhibitory concentration (IC_{50}) values of inhibitors were obtained using statistical correlation. 2D structures of these inhibitors are displayed in Table A.1.

2.4. Virtual Screening

Virtual screening (VS) is a computational technique that is related to quick search of libraries of chemical structures in order to identify those structures which are most likely to bind to a target receptor, typically a protein [29]. With the aid of VS, one can reduce the enormous chemical space of millions of compounds to a number that can be studied in detail by experiments or further computational methods.

Virtual screening aims to attain novel chemical structures that bind to the macromolecular target. Then, success of virtual screening is evaluated in terms of finding interesting new scaffolds. Low hit rates of interesting scaffolds are much preferred over high hit rates of already known scaffolds. There are two main virtual screening techniques as ligand-based and structure-based. Structure-based methods utilize structural information of the receptor protein. Ligand-based methods tend to find results that are closely related to known active molecules.

Structure-based virtual screening executes docking of candidate ligands into a target receptor followed by applying a scoring function to estimate the likelihood that the ligand will generally bind to a receptor with high affinity. However, in some cases such as side effects or hERG channel blocking, the low affinity can be required. Except blind docking and site map calculations, binding site of the target protein should be known because the docking process includes sampling of the coordinate space of the binding site and scoring each possible ligand pose that is then taken as the predicted binding mode for that compound. Docking process is generally carried out by using ligand flexibility and rigid receptor [30, 31].

Once a pose has been generated for a compound in the binding site it is scored to rank the quality of the pose with respect to other poses for the compound, and with respect to other molecules in the database. There is a wide choice of scoring functions available, and they can be categorized as being physical-based (force-field), empirical or knowledge-based. The physical scoring functions are based on atomic force fields. Empirical scoring functions are calculated using physicochemical properties such as hydrogen-bond counts to estimate the binding free energy. The overall binding free energy consists of several free

energy terms corresponding to hydrogen bonding, hydrophobic interactions, entropic changes and interactions with metal ions [32].

Ligand-based virtual screening carries out different approaches in attaining ligands that can interact with the target molecule. First approach is to utilize 2D chemical similarity analysis methods to search a database of chemical structures against one or more active ligand structures. It is faster than pharmacophore building approach but using three-dimensional forms of the ligands in addition to structures is generally more successful compared to 2D similarity alone [33].

The second approach is related with pharmacophore building if enough information about the ligands which bind to the target receptor is known. The proposed ligand is compared to the pharmacophore model for its consistency [34]. As stated in the IUPAC definition; "a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure optimal supramolecular interactions with a specific biological target structure and to block its biological response. The pharmacophore can be considered as the largest common denominator shared by a set of active molecules." Typical pharmacophore features include hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations and anions. The features need to match different chemical groups with similar properties in order to attain novel ligands.

Moreover, if the biological activities of known inhibitors are available, quantitative structure-activity relationship (QSAR) model can be utilized to validate the best hypothesis [35]. In pharmacophore-based QSAR model, ligands are aligned to the pharmacophore feature points and QSAR descriptors are calculated. Then, QSAR equation is obtained using partial least squares method [36]. Active ligands are divided into training set and test set. The training set is used to generate QSAR model and test set is utilized for prediction of the activity in order to validate the proposed model. With an increasing number of PLS factors, models are constructed and regression is carried out. The accuracy of the model increases with increasing number of PLS factors and maximum possible PLS factor can be no larger than twenty percent of the training set molecules [35].

The quality of QSAR model is evaluated by statistical parameters such as SD, R^2, F, RMSE, Q^2 and Pearson-R values. SD is the standard deviation of the regression, F is the variance ratio, RMSE is root-mean-square error of predicted actives, Q^2 value is similar to R^2 in terms of the predicted activities and Pearson-R represents the value for the correlation between the predicted and observed activity for the test set. SD, R^2 and F display the predictive ability of QSAR model on training set compounds, whereas RMSE, Q^2 and Pearson-R measure the predictive ability of QSAR model on test set molecules.

3. METHODS

All virtual screening applications for the discovery of potential Tpx inhibitors were carried out via Schrödinger`s Small Molecule Drug Discovery Suite 2013 on Linux operating system using HP xw6600 Workstation. The programs utilized in Schrödinger`s Suite were: Protein Preparation Wizard [37, 38] for the preparation of the receptor structure, LigPrep [39] for ligand preparation, ConfGen [40, 41] for multiple conformation generations of ligands, Phase [35, 42-43] for ligand-based pharmacophore modeling and QSAR model, Glide [44-46] for ligand docking and formation of active binding site of the receptor protein, Induced Fit [47, 48] for accurate prediction of binding conformation of ligands using conformational changes in receptor active binding sites, QikProp [49] for prediction of the pharmacokinetic properties for absorption, distribution, excretion and metabolism (ADME) [50] of ligands and Canvas [51, 52] for hierarchical clustering and scaffold decomposition among the selected ligands. Maestro interface [53] was used to access all required modules. The workflow of the proposed study is displayed in Figure 3.1.

3.1. Receptor Structure Preparation

The crystal structure of the receptor protein was retrieved from the Protein Data Bank (PDB) (PDB id: 3ZRD). Protein structure was prepared using Protein Preparation Wizard (PPW) [38] of Schrödinger Suite 2013. Missing hydrogens were added to heavy atoms and coordinates of missing loops and side chains (chain A residues: 4, 17, 141, 167, chain B residues: 17, 24, 26, 33, 67, 124, 167) were filled via Prime module [54, 55]. After the structural defects were fixed, the undesired parts of the structure, such as water molecules, were deleted. No water molecules were included because there was no mention of important water mediated contacts around active binding site in the literature. Only two chains (A and B) of Tpx receptor were kept for further operations.

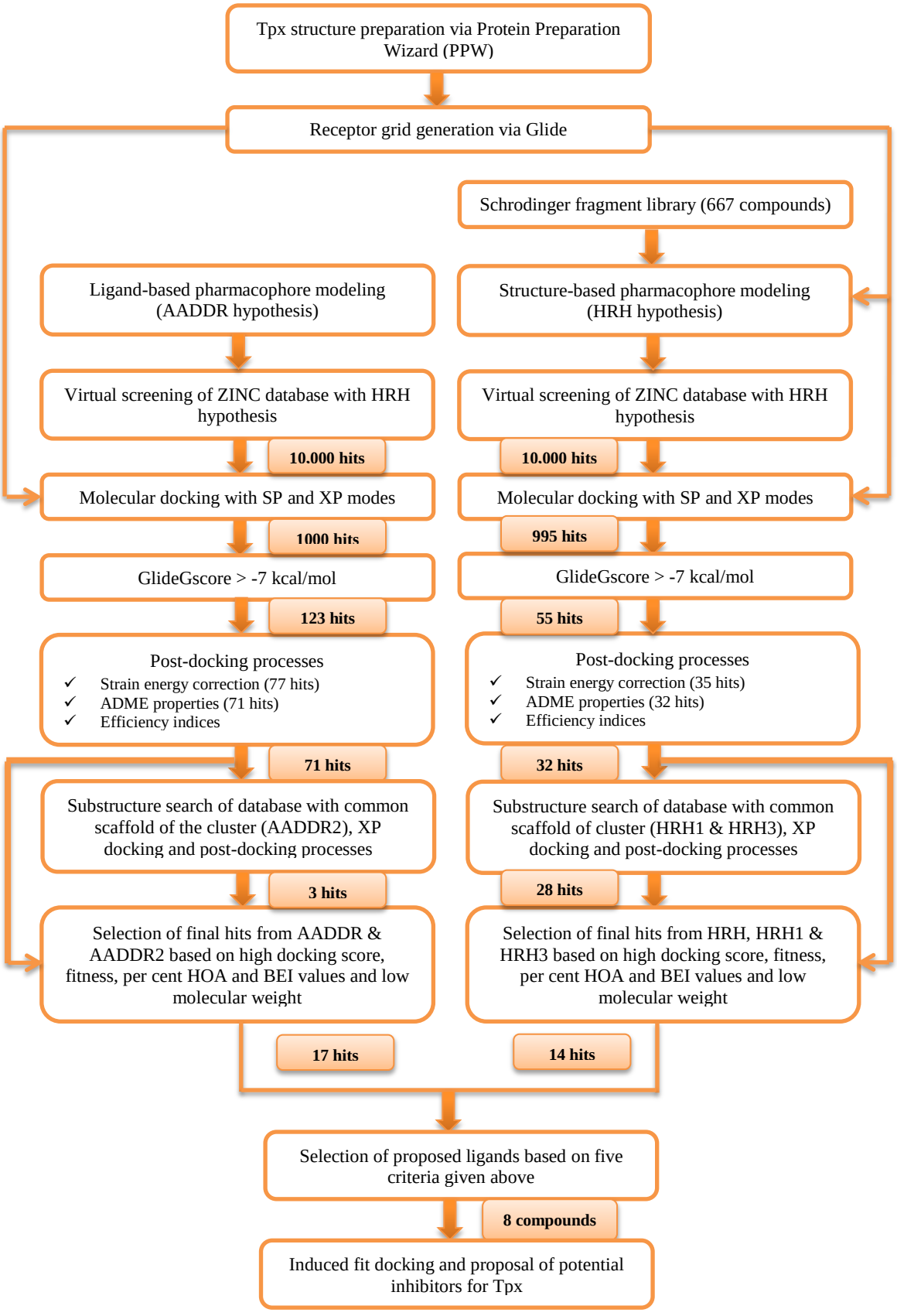


Figure 3.1. Workflow of the identification of Tpx inhibitors.

Hydrogen-bonding network was optimized upon choosing the protonation state by reorienting hydroxyl and thiol groups, amide groups of asparagines and glutamines and imizadole ring of histidines since orientations of these hydroxyl or thiol groups and amide groups cannot be determined from X-ray structure. Protonation states of aspartic acid, glutamic acid and histidine and tautomeric states of histidine imidazole group were predicted at neutral pH. At the last stage of protein preparation, all protein atoms were energy minimized by setting the maximum heavy atom RMSD to 0.30 Å using the OPLS_2005 (Optimized Potentials for Liquid Simulations) force field. The steps of the protein preparation process are displayed in Figure 3.2.

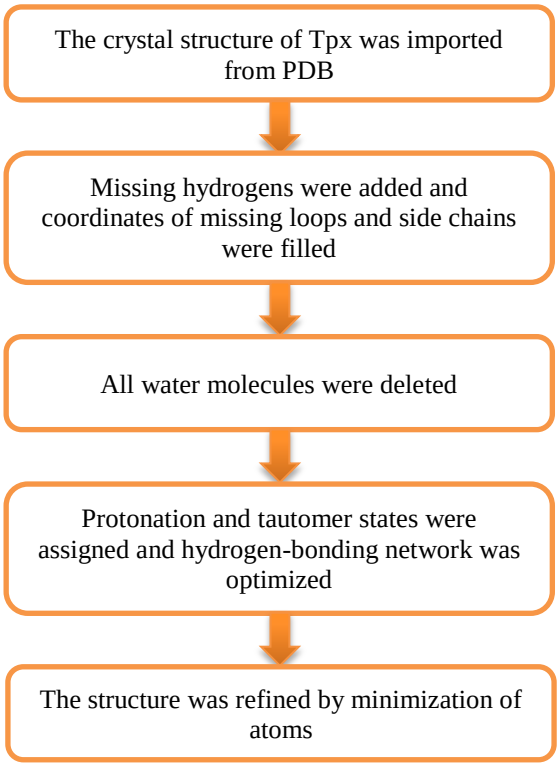


Figure 3.2. Workflow of the protein preparation process.

3.2. Receptor Grid Generation

Small identical grids are required for Glide docking that represent shape and features of the receptor binding site. Electrostatic energy of the protein structure is calculated using these grids. These energy values of the grids are utilized in molecular docking to use the scoring and screening time effectively. The files retrieved from the Protein Preparation Wizard were used to generate the grid via the Receptor Grid Generation panel [46]. This

panel includes five tabs, namely receptor, site, constraints, rotatable groups and excluded volumes.

In the Receptor tab, the prepared receptor protein was taken to the Maestro workspace. As receptor flexibility is not allowed by Glide, van der Waals radii of nonpolar atoms with partial charges is reduced to gain little flexibility. The scaling value of 1.0 (no scaling) with 0.25 partial charge cutoff was used as default value. Since there was no mention in literature that Tpx has tight and buried binding site, the van der Waals radius of atoms were not scaled.

In the Site tab, Glide utilizes two 3D cubic boxes to perform the energy calculations. The center of the two boxes is the same. In the enclosing or outer box, all ligand atoms to be docked should be placed and the center of the grid box was at the center of the Tpx active binding site including Leu 35 and Phe 89 residues of chain A and Phe 7, Phe 53, Pro 54, Thr 58, Cys 61, Val 65, Phe 94, Leu 127, Arg 133, Ile 153 and Thr 154 residues of chain B [23]. The center of the docked ligands must reside the inner box. The dimensions of the inner box were left as default (10 Å).

Four types of constraints including positional, hydrogen bond, metal or hydrophobic can be applied in Glide docking and these constraints are defined in the Constraints tab. A positional constraint with Isoleucine 153 residue was added to the specified grid because Ile 153 was shown to make a hydrogen bond interaction with the salicylidene acylhydrazide class of ligand [23] and induced to the conformational shift of Thr 154.

In the Rotatable Groups tab, hydroxyl groups of Ser, Thr, Tyr can be allowed some flexibility in ligand binding to form a more favorable protein-ligand complex. Ser 92, Thr 78, Tyr 118 residues were not close to the active binding site and therefore, rotations were not permitted.

In the Excluded Volumes tab, some regions in the active site can be prevented from being occupied by ligands. Throughout the operations, excluded volumes were not used.

3.3. Generation of Pharmacophore Models

In computer-aided drug design, combination of the results of multiple virtual screening protocols gives rise to high-quality lead compounds against a target protein [56-58]. In this thesis study, two different pharmacophore models were applied to find different chemotypes of inhibitors for Tpx.

3.3.1. Structure Based Pharmacophore Modeling

Structure-based pharmacophore modeling involves an analysis of the chemical features of the active site and their relationships and a subsequent model assembly with selected features. E-pharmacophores script [32, 59] with Glide and Phase [43] softwares was utilized for hypothesis generation for Tpx receptor.

The Schrödinger fragment library [46], which includes 667 conformers of 441 unique fragments, was docked to Tpx without any constraints using extra precision (XP) mode of Glide. Docking calculations were recorded in Glide XP descriptor file for post-docking. E-pharmacophores script was used to obtain information about energetic terms of docking obtained from Glide XP descriptor file.

E-pharmacophores script maps the energetic contributions of all conformers onto the atom centers and energy terms are summed on each atom center. The summed energy values of each atom lead to detection of possible pharmacophore sites. These sites were ranked based on their energy terms and the best ranking pharmacophore sites were used to generate the pharmacophore hypothesis. A set of six pharmacophore features is available in Phase hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobic group (H), negatively charged group (N), positively charged group (P) and aromatic group (R). The hypothesis can be made up of three to seven pharmacophore sites. HRH hypothesis was constructed to filter the database of 500,000 compounds generated from ZINC and the number of structures was reduced to 10,000 based on the fitness scores to the hypothesis.

3.3.2. Ligand Based Pharmacophore modeling

27 salicylidene acylhydrazide derivatives which were previously shown to inhibit Tpx [5, 27] were used to generate hypotheses via Develop Common Pharmacophore Hypotheses application of Phase module that has five steps: Prepare Ligands, Create Sites, Find Common Pharmacophores, Score Hypothesis and Build QSAR Model.

In the Prepare Ligands step, 2D structures of the known inhibitors which were drawn manually on the Maestro interface, and LigPrep [39] was used to generate low-energy 3D structures of these ligands. All possible ionization states of ligands were generated at neutral pH and maximum 32 stereoisomers were produced for each ligand. The energy of the structures was minimized with ConfGen [41] and maximum 100 conformers per ligand were selected.

In the Create Sites step, the possible pharmacophore sites for each ligand were created. Hydrogen bond acceptor site was located on an atom carrying at least one lone pair such as -O, -N. Hydrogen bond donor site was placed on a hydrogen atom which is attached to a heavy atom such as -OH, -NH, $-NH_2$. Hydrophobic sites were composed of chains of at most 4 carbons, rings and isopropyl groups. Positively or negatively charged group sites existed on the charged atom of the group. At the center of each ring, aromatic ring sites were placed.

Common pharmacophores were produced using the created sites. Common pharmacophore features which were observed in each inhibitor with the identical spatial arrangement were grouped. A variant list for common pharmacophores, which includes a number of feature types, was generated by Phase. In this thesis study, maximum five points in a variant were selected to generate a ligand-based hypothesis because use of more points decreased the likelihood of finding common pharmacophores amongst the set of known inhibitors.

In the Scoring Hypothesis step, overall quality of each hypothesis was measured based on its survival score, which is a combination of alignment (site) score, vector score and volume score. Alignment (site) score is the root-mean squared deviation (RMSD) in

the site-point positions of the pharmacophore. Vector score is the average cosine of the angles formed by corresponding pairs of vector features as acceptors, donors and aromatic rings in the structure. Volume score is the overlap of van der Waals spheres of the non-hydrogen atoms in each pair. The fitness score of ligands to pharmacophore hypothesis ranges from -1.0 to 3.0, a fitness score of -1.0 can be considered as 0% proximity to selected pharmacophore hypothesis, whereas 3.0 corresponds to exact match (100% proximity) of all site points to ligand features.

The QSAR model was created for the hypotheses, which have highest survival scores, using Build QSAR Model tab. Since the activities of known inhibitors for Tpx are available in terms of their IC50 concentrations, QSAR model was built to obtain the best hypothesis. In this tab, the ligands were aligned to the set of pharmacophore sites in the selected hypothesis, and fitness scores were obtained. The space occupied by the aligned ligands was defined by the collection of uniform small cubes of 1 Å on each side. Each ligand can be represented by a set of bit values (zero for no occupancy or one for occupancy). Once the occupancies were determined, partial least squares (PLS) method were applied and a regression coefficient was found for every bit. Using these regression coefficients, structural features within a molecule that increases or decreases the activity were determined via QSAR model. 27 active ligands were manually divided into training set (17 compounds) and test set (10 compounds) with the selection of PLS factor as three. Regression was performed with the training set molecules and in order to validate the model, test set compounds were used.

Five-point hypothesis AADDR gave the best three-dimensional alignment to the compounds in terms of site, volume, vector and total survival scores and it was validated by QSAR model. Then, this hypothesis was employed in screening of the ZINC database to reduce the molecules to 10,000 which were based on the fitness scores to the hypothesis.

3.4. Molecular Docking Protocol

After database filtering with the selected hypothesis, each molecule in the database was docked to the binding site of the target receptor. The docking process included sampling the coordinate space of the binding site and scoring each possible ligand

interaction that was the predicted binding mode of the compound. As the conformation of compound in bound form can be different from the unbound conformation, the conformational flexibility of ligands was taken into account by generating conformations for each ligand. There are many docking programs available such as Glide, AutoDock or GOLD and they differ in scoring functions they utilize, ligand and protein flexibility options, and the CPU time for docking of one molecule. In this thesis study, Glide was employed for docking with flexible ligand and rigid target receptor in its crystal structure conformation.

Ligand poses were scored using Schrödinger`s GlideGscore scoring function which is based on ChemScore [45]. ChemScore is a function derived from a set of 82 protein-ligand complexes, for which measured binding affinities are available. Glide utilizes random search algorithm and uses an empirical scoring function which is displayed in Equation 3.1.

$$GlideScore = (0.05 \times vdW) + (0.15 \times Coul) + Lipo + H_{bond} + Metal + Rewards + Rot\ B + Site \quad (3.1)$$

where vdW term is the Van der Waals energy, Coul term represents the Coulomb energy, Lipo stands for lipophilic term derived from hydrophobic grid potential, H_{bond} is the hydrogen-binding term, Metal stands for metal-binding term and only the interactions with anionic or highly polar acceptor atoms are included, Rewards term represents rewards and penalties for various features, such as hydrophobic enclosure and correlated hydrogen bonds, RotB is the penalty for freezing rotatable bonds, Site is the polar interactions in the active site.

Three different docking modes are available in Glide including high-throughput virtual screening, standard precision (SP) and extra precision (XP). They differ in the scoring functions and sampling [45]. Scoring functions of HTVS and SP are fundamentally same; however, XP scoring function contains several penalties including desolvation or charge penalty. XP also includes improvements to the hydrophobic reward which is in the form of lipophilic-lipophilic pair terms of the SP scoring function, detection of π-π stacking and pi-cation interactions. HTVS is recommended to scan very large number of

compounds using a highly restricted sampling. SP mode performs screening with more through sampling and moderate rapidity. XP, on the other hand, requires more CPU time as a result of more extensive sampling with greater conformational requirements in order to eliminate false positives.

3.4.1. Virtual Screening Workflow

Virtual Screening Workflow (VSW) unit was used in the molecular docking step. The compounds filtered from ZINC database via selected pharmacophore hypothesis were docked to Tpx dimer structure with pre-generated grid file using SP and XP modes consecutively. The advantage of VSW is that the compounds from SP mode pass to the XP stage without any intervention. If SP and XP modes are performed separately, the strain, which emerges during the first stage, should be eliminated by energy minimization. The positional constraint with Isoleucine 153 and ligand flexibility were applied in both docking modes. Maximum ten poses per compound were generated in both docking modes. Ligand poses were ranked based on GlideGscore and only best scoring hit was kept in final list. Top 10 per cent of the docked ligands according to SP Gscore were docked using XP mode (1000 compounds). The threshold GlideGscore value for the ligands, which were then used in post-docking, was selected based on the distribution of the GlideGscores. Output file as a pose viewer file with the extension of pv.mae was generated including the Tpx receptor and the final hits.

3.5. Post Docking Processes

Docking results of the compounds were analyzed to find potential inhibitors of Tpx by applying some post-docking steps. In this study, the high-ranking poses were analyzed based on ligand strain, efficiency indices, ADME and druglikeness properties, and protein-ligand interactions.

Ligands are docked to a rigid receptor via Glide and the lack of receptor flexibility is compensated by ligand strain values. Ligands with high strain values due to rigid receptor docking were identified via ligand strain calculations using the Strain-Rescore script. The ligands with higher than 4 kcal/mol energy difference between bound and free

conformations received penalties and quarter of this energy difference was added to the Gscore by MacroModel module [60], resulting in corrected Gscore values.

Ligand efficiency is an indicator of ligand affinity normalized by size and it provides guidance in the lead selection of the drug discovery process [61]. After strain energy corrections, the ligands were further evaluated by calculations of the ligand efficiency index (LEI), which is the Gibbs free energy per the number of non-hydrogen atoms in the compound, binding efficiency index (BEI), which relates potency values to molecular weight on a per kilodalton scale and surface efficiency index (SEI), which monitors the potency gains as related to the increase in polar surface area (PSA) normalized by 100 $Å^2$ [61-64].

The pharmacokinetic properties for absorption, distribution, excretion and metabolism (ADME) [50] were evaluated using the QikProp module [49]. Additionally, druglikeness of ligands according to Lipinski`s rule of five [65] were also predicted by QikProp. The residues that interacted with the ligands and played a significant role in binding were identified using the Structure Interaction Fingerprint Tool (SIFT) [66].

3.6. Similarity Analysis and Substructure Search

After about 1000 ligands for both ligand-based and structure-based pharmacophore models were docked to Tpx in XP mode, a threshold docking score was selected based on the distribution of GlideGscores. Investigation of similarity between ligands with docking scores higher than the threshold value was carried out to determine a diverse set of chemical scaffolds as potential inhibitors of Tpx using Canvas module [52]. Fingerprint (2D) similarity was determined using atom pairs based on the Tanimoto Coefficient (TC) [67]. ‘Atom pair’ is defined by the shortest path among the non-hydrogen atoms, and a similarity coefficient such as Tanimoto Coefficient is computed using common atom pairs.

Substructure search is the process of finding similar structures to a query in order to detect other compounds with inhibition potential against a target molecule. After hierarchical clustering and scaffold decomposition were carried out in the Canvas module, the common scaffolds of the ligands with favorable BEI values were selected and the

ZINC database was screened with that scaffold again in order to obtain additional high scoring ligands which were missed in the pharmacophore filtering stage [68].

3.7. Induced Fit Docking

Induced fit docking (IFD) [47, 48] predicts ligand binding modes and conformational changes in the target receptor by allowing side chain motions in the receptor. After identification of a set of compounds, their binding mode was further examined by IFD. Initially, Glide XP docking of selected hits to the rigid receptor was performed. Prime was employed to generate maximum 20 poses for each ligand and residues that are located within 5 Å of 20 poses. Receptor side chain flexibility for those residues was introduced while the coordinates of other residues and the backbone remained fixed. Finally, the ligands were redocked in Glide XP mode into the lowest energy induced fit receptor structure with no scaling and their final scores were obtained.

3.8. Selectivity Analysis

Basic Local Alignment Search Tool (BLAST) [69] was utilized to identify sequentially similar proteins to Tpx. After the similar proteins were found, the proposed ligands were docked to these proteins in XP mode without any constraint to verify that the selected ligands are selective for Tpx. The grid files were generated such that they included residues corresponding to the residues of Tpx which mostly interacted with the proposed ligands.

4. RESULTS AND DISCUSSION

Pathogen-host interaction is one of the mechanisms that can be targeted in the design of novel inhibitors against multidrug resistant bacteria. The key protein in pathogenic bacteria to overcome the defense mechanism of host cell is Tpx. Within the scope of this study, potential Tpx inhibitors were identified via pharmacophore modeling, docking and substructure search approach. The pharmacophore models, which were generated either based on the binding site structure or a set of previously identified Tpx inhibitors, were used to filter the ZINC database using 3D similarity to the selected pharmacophore hypothesis. Molecular docking by Glide was applied to predict the binding modes of the compounds in the filtered database. Strain energy, ADME and druglikeness properties, similarity and efficiency indices were calculated to filter the compounds obtained from docking operations. Common scaffolds in the top scoring hits were identified via Canvas, and substructure search with the selected scaffold was carried out to filter the database and obtain other compounds. The surviving molecules were analyzed via induced fit docking and selectivity to other similar proteins and were proposed as potential inhibitors of Tpx.

4.1. Receptor Structure Preparation and Grid Generation

The coordinates of the receptor protein was taken from PDB (id: 3ZRD) and it was prepared using Protein Preparation Wizard module. Bond orders were automatically assigned and missing hydrogens were added. The coordinates of missing loops and side chains (chain A residues: 4, 17, 141, 167, chain B residues: 17, 24, 26, 33, 67, 124, 167) were predicted by Prime. Out of 226 water molecules present in the coordinate file, all were deleted because there was no mention of important water mediated contacts in the literature.

Hydrogen-bonding network was optimized at neutral pH by reorienting hydroxyl and thiol groups, the imidazole ring in histidine and amide groups of asparagine and glutamine. Finally, energy minimization of all atoms using OPLS_2005 force field was carried out via Impref utility with the maximum heavy atom RMSD of 0.30 Å. The structure of Tpx before and after preparation is displayed in Figure 4.1.

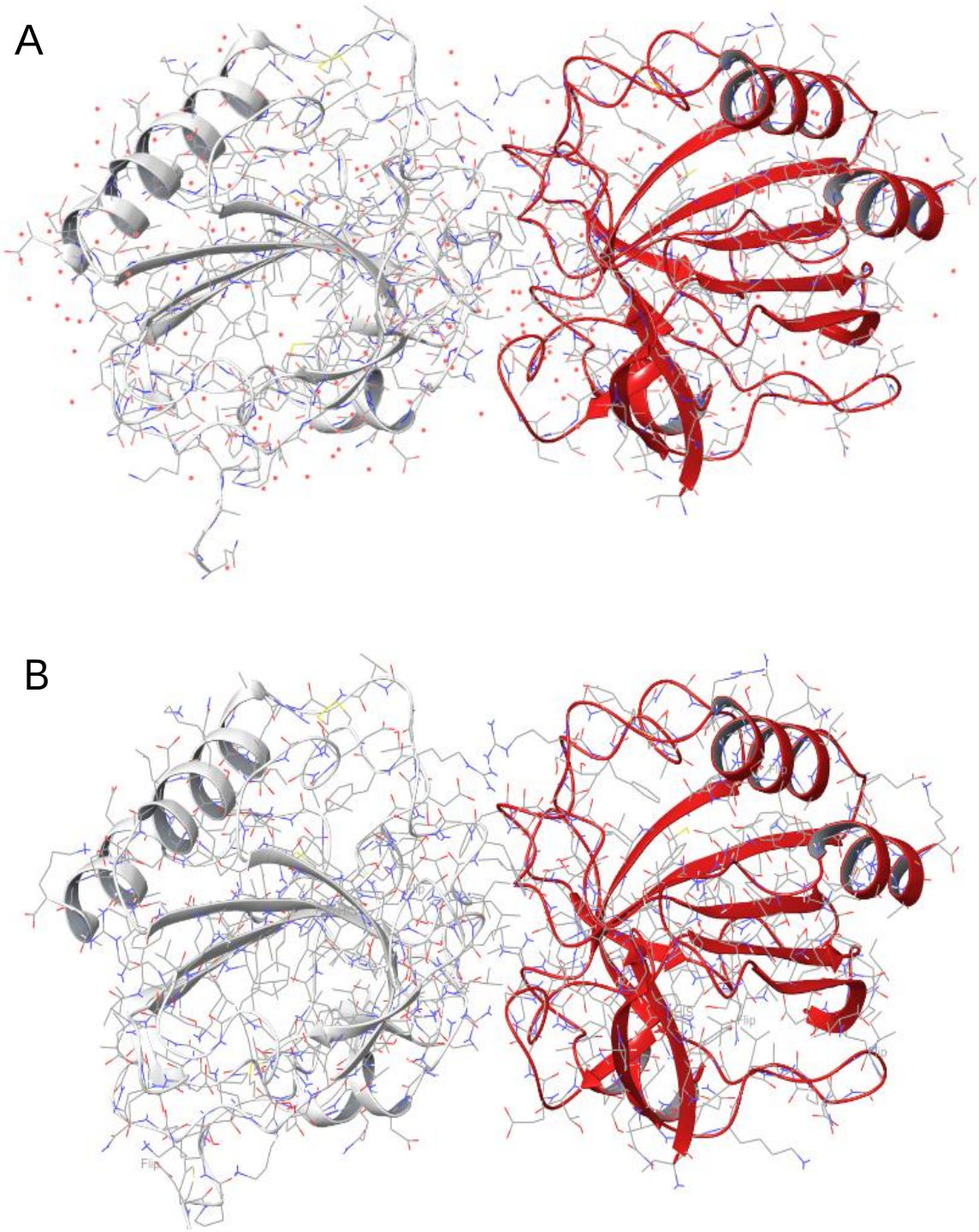


Figure 4.1. Structure of Tpx - before (A) and after (B) preparation. Backbone is displayed in cartoon and side chains are in licorice representation. A and B chains are shown in grey and red, respectively. Water molecules are in red circles.

The binding site of Tpx was represented by small identical grids. The center of the grid box was chosen as the centroid of active binding site residues including Leu 35 and Phe 89 of chain A and Phe 7, Phe 53, Pro 54, Thr 58, Cys 61, Val 65, Phe 94, Leu 127, Arg

133, Ile 153 and Thr 154 of chain B. The dimensions of inner box in which the center of a ligand to be docked must reside were selected as 10 Å. Rotation of hydroxyl groups on threonines, tyrosines and serines were not allowed because the residues as Thr 78, Tyr 112 and Ser 92 were not close to the active site. A positional constraint with Ile 153 was used because of its importance in binding specified in literature [23]. The van der Waals scaling value of nonpolar atoms were left as default value of 1.0 (no scaling) with 0.25 partial charge cutoff. The boxes which define the receptor binding site are shown in Figure 4.2.

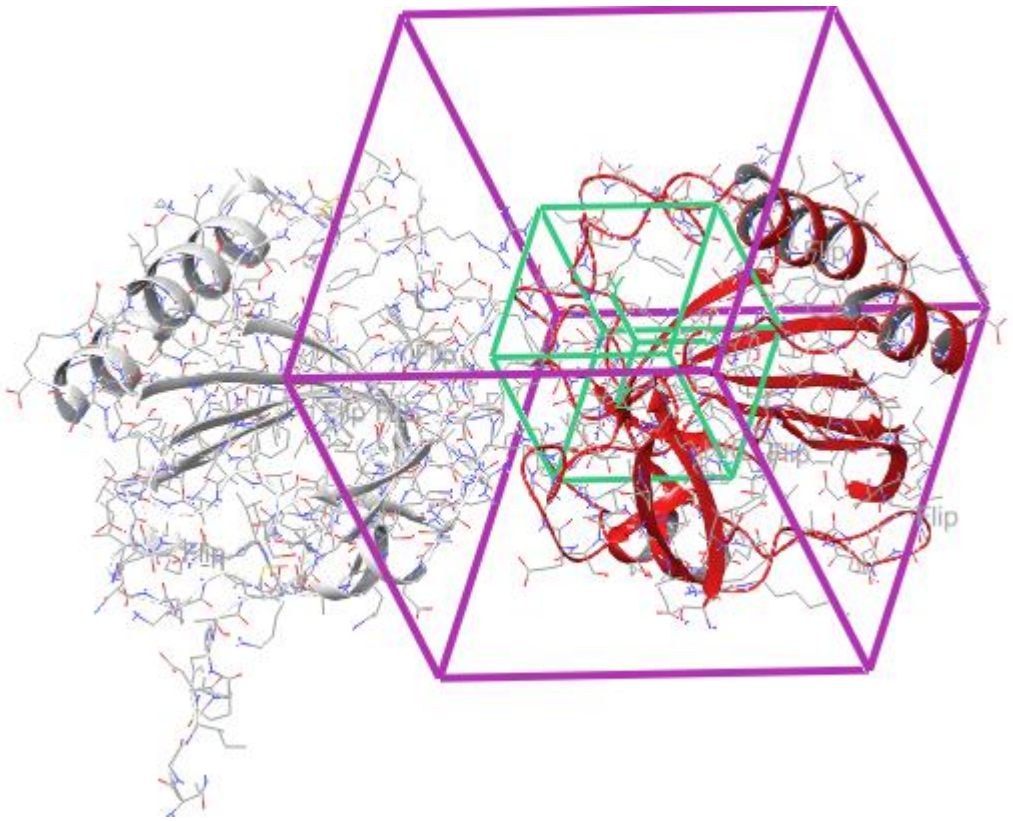


Figure 4.2. Outer (purple) and inner (green) boxes for Tpx which were obtained in Receptor Grid Generation Panel.

4.2. Determination of Potential Inhibitors of Tpx by Structure-Based Pharmacophore Modeling

Structure-based pharmacophore modeling was carried out via E-pharmacophores script of Phase. Initially, docking of small fragments obtained from Schrodinger Fragment Library [46] to the binding site of Tpx was performed. Then, using energetic terms and binding modes of fragments to the receptor, descriptive information was collected and

possible pharmacophore sites were determined. Finally, hypothesis was constructed using selected pharmacophore sites.

Structure-based pharmacophore sites were created by unconstrained docking of small fragments to the binding site of Tpx receptor. Via the information from binding modes and energetic terms of 441 unique fragments (667 conformers), seven pharmacohore sites were determined. The coordinates, ranks, scores and types of seven sites are given in Table 4.1.

Table 4.1. Coordinates, ranks and scores of seven pharmacophore features by E-pharm.

Rank	Feature	Type	Score (kcal/mol)	X	Y	Z
1	H1489	H	-7.61	18.780	14.224	12.427
2	R2265	R	-5.10	18.941	12.567	14.860
3	H1450	H	-2.52	17.325	10.284	11.799
4	R1807	R	-2.30	19.133	12.162	12.954
5	H1401	H	-1.72	19.981	12.722	14.860
6	A577	A	-1.11	15.473	11.221	10.982
7	R2170	R	-1.07	16.562	10.491	11.721

The pharmacophore sites included three hydrophobic groups consistent with the information that the binding pocket is mostly hydrophobic [23]. Since as the number of sites increases the pharmacophore model becomes more restrictive, hypothesis with three sites was selected in order to eliminate as few compounds as possible from the database. Distances between two pharmacophore site points are listed in Table 4.2. Consequently, the hypothesis based on the rank and inner-site distances was created using H1489, R2265 and H1401 (HRH hypothesis). For example, H1450 was not used in the hypothesis instead of H1401 despite its high rank because the distance of H1450 site with the other sites used in the hypothesis was quite high (> 6Å). HRH hypothesis in the binding site of Tpx structure and the inter distances of the hypothesis sites are shown in Figure 4.3. ZINC database was filtered with HRH hypothesis and 10,000 ligands with the minimum fitness of 1.90 passed to docking step.

Table 4.2. Inner-site distances of seven pharmacophore sites.

Site1	Site2	Distance (Å)	Site1	Site2	Distance (Å)
A577	H1401	6.051	H1450	H1489	7.942
A577	H1450	3.138	H1450	R1807	4.703
A577	H1489	5.646	H1450	R2170	2.478
A577	R1807	2.808	H1450	R2265	8.236
A577	R2170	4.704	H1489	R1807	3.259
A577	R2265	6.862	H1489	R2170	9.391
H1401	H1450	6.324	H1489	R2265	3.032
H1401	H1489	5.076	R1807	R2170	6.189
H1401	R1807	3.681	R1807	R2265	4.079
H1401	R2170	7.379	R2170	R2265	9.301
H1401	R2265	4.096			

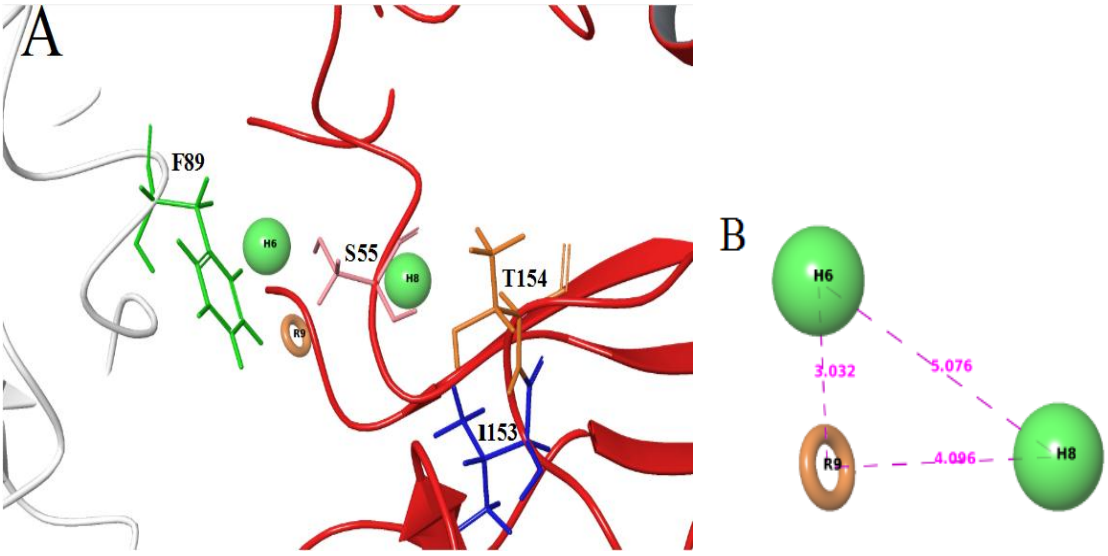


Figure 4.3. (A) Location of pharmacophore sites in the Tpx binding site with Phe 89 (green) of A chain and Ile 153 (blue), Ser 55 (pink) and Thr 154 (orange) of B chain. (B) Distances between HRH hypothesis sites (right). Aromatic ring (R) site is in orange and hydrophobic (H) sites are in green.

4.2.1. Molecular Docking with HRH Hypothesis

The docking studies were performed via Glide application of Schrödinger [46] with two docking modes which were standard precision (SP) and extra precision (XP). Virtual Screening Workflow (VSW) was utilized to pass top scored 10% ligands from SP docking to XP docking. The filtered database and grid file were used as input to VSW application for molecular docking of ligands that were flexible while the receptor was kept rigid. Docking and scoring were not allowed for compounds with more than 300 atoms and rotatable bonds. For both docking modes, maximum ten poses per compound were generated and only best scoring pose was taken to the final hit list.

Initially, 10,000 compounds obtained from the virtual screening step were docked to the Tpx active site in SP mode using Ile 153 interaction as a positional constraint. 995 molecules with highest SP score were then used in XP docking. The GlideGscore values of the 995 compounds (Figure 4.4) were between -9.1 and -2.4 kcal/mol and displayed that a relatively sharp decrease to -7.0 kcal/mol followed by a significantly smaller decrease. Therefore, GlideGscore of -7.0 kcal/mol was selected as the cutoff value and 55 molecules survived.

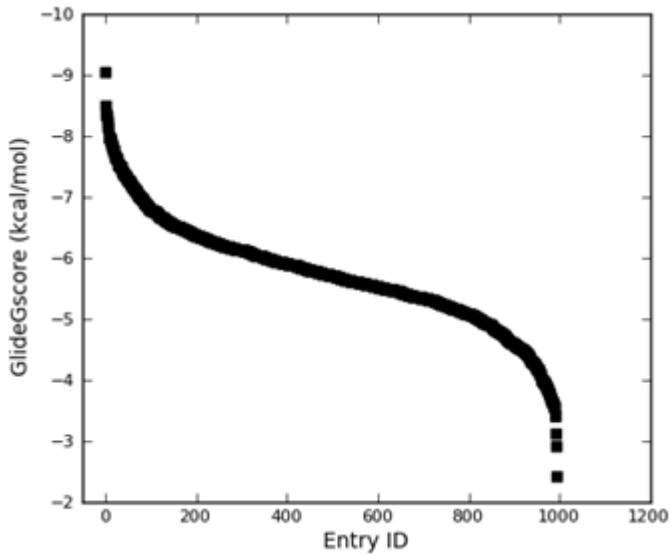


Figure 4.4. GlideGscore values for top scoring 995 compounds filtered with HRH hypothesis and docked using XP docking.

4.2.2. Post Docking Analyses of Selected Compounds

The binding efficiency index (BEI) and surface efficiency index (SEI) values of the 55 compounds were calculated based on the descriptive equations [62, 64].

$$LEI = \frac{\Delta G}{\#\ of\ non-hydrogen\ atoms} \tag{4.1}$$

$$\Delta G\ (kcal/mol) = -RTlnK_i \tag{4.2}$$

$$p(K_i) = -\log(K_i \times 10^{-9}) \tag{4.3}$$

$$BEI = \frac{p(K_i)}{Molecular\ weight\ (kDa)} \tag{4.4}$$

$$SEI = \frac{p(K_i)}{Polar\ surface\ area/100\ Å^2} \tag{4.5}$$

The efficiency indices of 55 ligands are displayed in Table 4.3 and Figure 4.5. The compounds on the BEI versus SEI plot were placed above the diagonal line suggesting that the inhibitors were mostly polar despite the existence of apolar groups in the compound which satisfy the hydrophobic sites of the hypothesis.

Table 4.3. Efficiency indices of 55 compounds of HRH hypothesis.

ZINC ID	LEI	# of non-H atoms	ΔG ($\frac{kcal}{mol}$)	p(K_i)	MW (kDa)	BEI	PSA (Å^2)	SEI
ZINC12378903	-0.301	26	-7.826	3.260	432.44	7.539	118.73	2.746
ZINC16891564	-0.314	27	-8.478	2.782	410.54	6.776	101.24	2.748
ZINC00186392	-0.439	19	-8.341	2.882	318.18	9.059	101.42	2.842
ZINC01275072	-0.347	24	-8.328	2.892	380.29	7.604	63.74	4.537
ZINC40984430	-0.244	31	-7.564	3.452	467.54	7.384	71.31	4.841
ZINC05640330	-0.329	23	-7.567	3.450	381.92	9.033	50.21	6.871
ZINC08627941	-0.302	25	-7.550	3.462	383.92	9.019	71.68	4.830
ZINC30769045	-0.301	23	-6.923	3.922	374.48	10.474	83.68	4.687
ZINC16383080	-0.399	20	-7.980	3.147	300.74	10.465	40.64	7.744
ZINC16383076	-0.398	20	-7.960	3.162	300.74	10.513	40.64	7.780
ZINC19868300	-0.241	29	-6.989	3.874	455.60	8.503	40.63	9.534
ZINC00185281	-0.378	19	-7.182	3.732	309.34	12.066	115.47	3.232
ZINC06793908	-0.294	25	-7.350	3.609	375.47	9.612	92.07	3.920

Table 4.3. Efficiency indices of 55 compounds of HRH hypothesis (cont.).

ZINC ID	LEI	# of non-H atoms	ΔG ($\frac{kcal}{mol}$)	p(K_i)	MW (kDa)	BEI	PSA (Å^2)	SEI
ZINC57992551	-0.343	20	-6.860	3.969	336.42	11.796	47.18	8.412
ZINC40984420	-0.217	32	-6.944	3.907	493.60	7.915	77.17	5.063
ZINC01071498	-0.268	29	-7.772	3.300	397.43	8.302	59.56	5.540
ZINC40984473	-0.234	30	-7.020	3.851	450.54	8.548	84.42	4.562
ZINC02168326	-0.349	16	-5.584	4.904	305.50	16.054	30.45	16.107
ZINC26781246	-0.319	19	-6.061	4.555	343.42	13.262	78.47	5.804
ZINC40959674	-0.218	33	-7.194	3.724	472.54	7.880	89.80	4.146
ZINC16736487	-0.263	26	-6.838	3.985	399.42	9.976	99.02	4.024
ZINC04013595	-0.543	13	-7.059	3.823	208.32	18.349	59.40	6.435
ZINC16638691	-0.313	21	-6.573	4.179	326.44	12.802	50.56	8.266
ZINC19798670	-0.312	22	-6.864	3.966	325.37	12.188	86.75	4.572
ZINC40959638	-0.214	33	-7.062	3.820	473.53	8.068	103.77	3.682
ZINC12297152	-0.277	23	-6.371	4.327	377.44	11.465	70.97	6.097
ZINC16736490	-0.256	26	-6.656	4.118	401.44	10.259	100.28	4.107
ZINC01649089	-0.433	15	-6.495	4.236	253.40	16.717	26.65	15.894
ZINC03043149	-0.368	20	-7.360	3.602	302.19	11.919	18.26	19.725
ZINC00099607	-0.351	18	-6.318	4.366	285.34	15.301	57.87	7.545
ZINC17020109	-0.408	18	-7.344	3.614	262.74	13.753	34.15	10.581
ZINC12079331	-0.229	29	-6.641	4.129	443.55	9.309	87.24	4.733
ZINC40959641	-0.210	33	-6.930	3.917	473.53	8.272	101.82	3.847
ZINC27675619	-0.349	17	-5.933	4.648	322.48	14.415	41.40	11.228
ZINC12279762	-0.244	28	-6.832	3.989	413.52	9.647	81.42	4.900
ZINC12371164	-0.347	19	-6.593	4.164	308.53	13.497	100.90	4.127
ZINC40959592	-0.202	34	-6.868	3.963	489.57	8.094	107.08	3.701
ZINC12371364	-0.346	19	-6.574	4.178	305.39	13.682	40.90	10.215
ZINC40959659	-0.207	33	-6.831	3.990	472.54	8.443	91.98	4.338
ZINC19839291	-0.382	18	-6.876	3.957	270.32	14.637	33.40	11.847
ZINC12218805	-0.315	22	-6.930	3.917	318.39	12.303	72.77	5.383
ZINC03851018	-0.258	26	-6.708	4.080	398.48	10.239	113.55	3.593
ZINC11935292	-0.327	21	-6.867	3.963	301.41	13.150	35.39	11.199
ZINC19169013	-0.359	18	-6.462	4.260	279.34	15.252	76.99	5.534
ZINC08456196	-0.239	28	-6.692	4.092	433.91	9.430	64.98	6.297
ZINC12031154	-0.231	25	-5.775	4.764	430.54	11.066	51.40	9.270
ZINC40984503	-0.198	31	-6.138	4.498	493.60	9.113	78.94	5.698
ZINC40969736	-0.246	25	-6.150	4.489	399.49	11.238	65.98	6.804
ZINC19075501	-0.274	26	-7.124	3.775	384.21	9.825	77.24	4.887
ZINC06239267	-0.245	26	-6.370	4.328	413.86	10.457	91.88	4.710
ZINC09196338	-0.355	17	-6.035	4.574	278.40	16.428	57.65	7.933
ZINC21231292	-0.262	24	-6.288	4.388	391.44	11.210	133.69	3.282
ZINC21142911	-0.262	26	-6.812	4.004	373.36	10.724	113.28	3.535
ZINC12373752	-0.243	27	-6.561	4.188	395.48	10.589	44.18	9.480
ZINC02253070	-0.201	28	-5.628	4.872	482.58	10.096	106.87	4.559

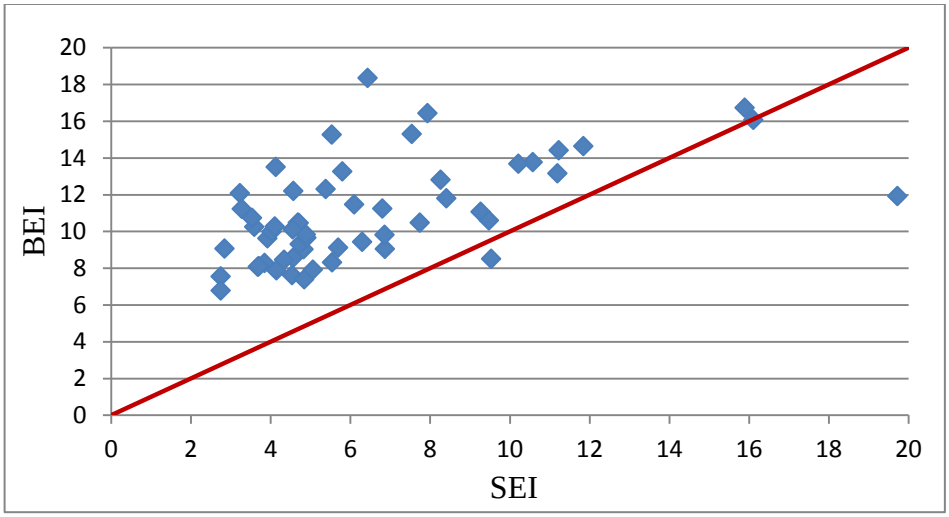


Figure 4.5. Binding efficiency versus surface efficiency indices of 55 hits of HRH hypothesis.

Ligands with too much strain were identified by applying strain energy correction via Strain-Rescore script [46]. The ligands with more than 4 kcal/mol energy difference between bound and free conformations received ligand strain penalties and quarter of the strain penalty was added to the GlideGscore. As a result, 20 compounds were eliminated (Table 4.4).

Table 4.4. GlideGscores (> -7 kcal/mol), fitness values to HRH hypothesis and strain corrected Gscores of ligands. Eliminated molecules are indicated by strikethrough.

ZINC ID	Glide GScore	Fitness value	Bound energy (kcal/mol)	Free energy (kcal/mol)	Strain energy (kcal/mol)	Strain penalty	Strain Gscore
ZINC12378903	-9.04	2.05	60.65	57.57	3.08	0.00	-9.04
ZINC16891564	-8.49	1.91	34.36	31.28	3.08	0.00	-8.49
ZINC00186392	-8.35	1.91	24.42	22.01	2.41	0.00	-8.35
ZINC01275072	-8.33	1.97	12.92	10.10	2.82	0.00	-8.33
ZINC40984430	-8.31	1.97	36.89	33.00	3.89	0.00	-8.31
~~ZINC05640330~~	~~-8.23~~	~~1.92~~	~~32.32~~	~~25.13~~	~~7.19~~	~~0.80~~	~~-7.43~~
ZINC08627941	-8.15	2.00	25.96	23.30	2.66	0.00	-8.15
~~ZINC30769045~~	~~-8.12~~	~~1.94~~	~~43.49~~	~~34.80~~	~~8.69~~	~~1.17~~	~~-6.95~~
ZINC16383080	-7.97	1.92	52.75	52.57	0.18	0.00	-7.97
ZINC19868300	-7.96	1.98	85.57	82.76	2.81	0.00	-7.96
~~ZINC00185281~~	~~-7.94~~	~~2.00~~	~~38.29~~	~~34.00~~	~~4.29~~	~~0.07~~	~~-7.87~~
~~ZINC06793908~~	~~-7.94~~	~~1.97~~	~~74.49~~	~~62.63~~	~~11.86~~	~~1.97~~	~~-5.97~~
~~ZINC57992551~~	~~-7.88~~	~~2.03~~	~~17.41~~	~~11.03~~	~~6.38~~	~~0.60~~	~~-7.28~~

Table 4.4. GlideGscores (> -7 kcal/mol), fitness values to HRH hypothesis and strain corrected Gscores of ligands. Eliminated molecules are indicated by strikethrough (cont.).

~~ZINC40984420~~	~~-7.81~~	~~1.96~~	~~40.36~~	~~31.91~~	~~8.45~~	~~1.11~~	~~-6.71~~
ZINC01071498	-7.76	1.94	80.59	77.25	3.34	0.00	-7.76
ZINC40984473	-7.72	1.97	32.27	30.44	1.83	0.00	-7.72
ZINC02168326	-7.67	1.99	32.37	31.58	0.79	0.00	-7.67
ZINC26781246	-7.66	1.91	37.01	35.11	1.90	0.00	-7.66
ZINC40959674	-7.64	1.92	47.89	44.51	3.38	0.00	-7.64
~~ZINC16736487~~	~~-7.63~~	~~1.99~~	~~39.32~~	~~34.57~~	~~4.75~~	~~0.19~~	~~-7.44~~
ZINC04013595	-7.61	1.94	11.91	10.49	1.42	0.00	-7.61
ZINC16638691	-7.51	2.06	39.80	36.81	2.99	0.00	-7.51
ZINC19798670	-7.49	1.94	33.62	31.94	1.68	0.00	-7.49
~~ZINC40959638~~	~~-7.49~~	~~1.92~~	~~50.66~~	~~44.84~~	~~5.82~~	~~0.46~~	~~-7.03~~
ZINC12297152	-7.47	2.04	48.98	47.36	1.62	0.00	-7.47
ZINC16736490	-7.42	2.00	38.18	36.30	1.88	0.00	-7.42
ZINC01649089	-7.37	1.91	19.92	18.71	1.21	0.00	-7.37
ZINC03043149	-7.37	1.98	33.36	33.35	0.01	0.00	-7.37
ZINC00099607	-7.36	2.12	22.22	20.78	1.44	0.00	-7.36
ZINC17020109	-7.35	1.98	19.02	17.94	1.08	0.00	-7.35
~~ZINC12079331~~	~~-7.34~~	~~1.95~~	~~72.87~~	~~64.70~~	~~8.17~~	~~1.04~~	~~-6.30~~
ZINC40959641	-7.34	1.92	45.20	42.19	3.01	0.00	-7.34
ZINC27675619	-7.32	1.93	10.22	9.43	0.79	0.00	-7.32
ZINC12279762	-7.31	1.95	62.11	61.75	0.36	0.00	-7.31
ZINC12371164	-7.29	1.96	26.49	23.96	2.53	0.00	-7.29
~~ZINC40959592~~	~~-7.27~~	~~1.92~~	~~39.32~~	~~32.74~~	~~6.58~~	~~0.65~~	~~-6.62~~
ZINC12371364	-7.26	2.00	34.40	34.24	0.16	0.00	-7.26
~~ZINC40959659~~	~~-7.25~~	~~1.92~~	~~52.81~~	~~43.35~~	~~9.46~~	~~1.37~~	~~-5.88~~
ZINC19839291	-7.25	1.91	22.31	21.65	0.66	0.00	-7.25
ZINC12218805	-7.25	2.09	47.01	46.36	0.65	0.00	-7.25
ZINC03851018	-7.22	1.91	26.65	24.87	1.78	0.00	-7.22
~~ZINC11935292~~	~~-7.19~~	~~1.97~~	~~25.84~~	~~18.97~~	~~6.87~~	~~0.72~~	~~-6.47~~
ZINC19169013	-7.19	2.01	42.43	39.64	2.79	0.00	-7.19
ZINC08456196	-7.16	1.97	41.43	40.27	1.16	0.00	-7.16
~~ZINC12031154~~	~~-7.16~~	~~2.01~~	~~61.84~~	~~53.84~~	~~8.00~~	~~1.00~~	~~-6.16~~
~~ZINC40984503~~	~~-7.14~~	~~1.96~~	~~43.78~~	~~38.50~~	~~5.28~~	~~0.32~~	~~-6.82~~
~~ZINC40969736~~	~~-7.14~~	~~1.99~~	~~79.53~~	~~60.47~~	~~19.06~~	~~3.77~~	~~-3.37~~
ZINC19075501	-7.13	1.95	17.67	17.47	0.20	0.00	-7.13
~~ZINC06239267~~	~~-7.11~~	~~1.95~~	~~56.40~~	~~52.06~~	~~4.34~~	~~0.09~~	~~-7.02~~
ZINC09196338	-7.10	1.91	33.84	32.91	0.93	0.00	-7.10
ZINC21231292	-7.07	2.03	14.60	13.08	1.52	0.00	-7.07
~~ZINC21142911~~	~~-7.06~~	~~1.95~~	~~26.45~~	~~21.43~~	~~5.02~~	~~0.26~~	~~-6.80~~
~~ZINC12373752~~	~~-7.04~~	~~1.99~~	~~67.80~~	~~49.05~~	~~18.75~~	~~3.69~~	~~-3.35~~
~~ZINC02253070~~	~~-7.03~~	~~2.00~~	~~91.40~~	~~63.11~~	~~28.29~~	~~6.07~~	~~-0.96~~
~~ZINC19799833~~	~~-7.00~~	~~1.91~~	~~47.77~~	~~40.47~~	~~7.30~~	~~0.83~~	~~-6.17~~

The remaining 35 compounds were further analyzed with their pharmacokinetic properties as absorption, distribution, metabolism and excretion (ADME) via QikProp program of Maestro interface. QikProp compares the predicted properties of molecules with those of 95% of known drugs. The key descriptors for ADME criteria as SASA and its contributors FOSA, FISA, PISA and WPSA, PSA and percentage of Human Oral Absorption were evaluated based on the acceptable ranges and the compounds that were out of these ranges were eliminated from further analyses.

Number of stars (#stars) displays the number of descriptor values that fall outside the 95% range of similar values for known drugs. SASA is described as the total solvent accessible surface area ($Å^2$). FOSA is the hydrophobic component of the SASA, which is related with a saturated carbon and attached hydrogen. FISA is the hydrophilic component of the SASA, which is associated with the SASA on N, O, and H on heteroatoms. PISA is the π component of the SASA that is referred to a carbon and attached hydrogen. WPSA is the weakly polar component of the SASA, which is related with the polar components as electronegative atoms such as halogens, phosphorus and sulphur atoms. PSA is the van der Waals surface area of polar nitrogen and oxygen atoms. Percentage of human oral absorption (HOA%) is accepted as good when higher than 80% and poor when lower than 25%. For each compound, the properties described above with their acceptable ranges are displayed in Table 4.5.

Table 4.5. ADME properties for selected 35 hits of HRH hypothesis. Out-of-range property is displayed in bold and eliminated molecules are tabulated by strikethrough.

ZINC ID	# stars	SASA	FOSA	FISA	PISA	WPSA	PSA	HOA %
		300.0 1000.0	0.0 750.0	7.0 330.0	0.0 450.0	0.0 175.0	7.0 200.0	<25% low >80% high
ZINC12378903	1	751.73	227.22	155.89	266.28	102.34	118.73	100.00
ZINC16891564	1	730.95	124.29	175.86	393.08	37.71	101.24	81.94
ZINC00186392	0	520.10	47.98	179.09	133.75	159.28	101.42	77.22
ZINC01275072	0	639.72	78.53	93.47	307.32	160.41	63.74	100.00
ZINC40984430	0	811.46	441.84	47.34	284.39	37.89	71.31	100.00
ZINC08627941	0	709.14	238.19	138.81	276.87	55.27	71.68	100.00
ZINC16383080	1	511.01	229.29	50.09	98.97	132.56	40.64	94.91
~~ZINC19868300~~	~~1~~	~~815.41~~	~~623.98~~	~~**4.41**~~	~~187.03~~	~~0.00~~	~~40.63~~	~~100.00~~
ZINC01071498	0	668.89	267.69	66.95	253.11	81.15	59.56	95.93
ZINC40984473	1	816.06	459.14	65.88	291.04	0.00	84.42	93.50
ZINC02168326	1	654.69	542.74	64.51	47.44	0.00	30.45	100.00
ZINC26781246	0	674.53	523.18	84.44	66.91	0.00	78.47	84.60
ZINC40959674	0	839.70	361.73	89.25	388.72	0.00	89.80	100.00
ZINC04013595	1	467.99	131.75	123.89	176.15	36.21	59.40	91.69
ZINC16638691	0	657.26	452.46	34.63	170.17	0.00	50.56	3.85
ZINC19798670	0	578.96	188.42	125.28	265.27	0.00	86.75	92.07
ZINC12297152	0	668.92	532.47	56.80	79.39	0.00	70.97	89.81
ZINC16736490	1	698.45	349.26	132.66	172.80	43.73	100.28	78.59
ZINC01649089	1	541.77	303.33	50.39	162.51	25.55	26.65	96.69
~~ZINC03043149~~	~~3~~	~~459.75~~	~~170.33~~	~~29.93~~	~~36.55~~	~~**222.94**~~	~~18.26~~	~~100.00~~
ZINC00099607	0	598.64	266.85	57.52	274.28	0.00	57.87	100.00
ZINC17020109	0	533.05	74.34	58.82	349.25	50.27	34.15	100.00
ZINC40959641	0	841.31	367.58	115.38	358.34	0.00	101.82	88.62
ZINC27675619	0	604.56	360.69	37.98	142.15	63.74	41.40	100.00
ZINC12279762	0	768.95	561.07	95.55	112.32	0.00	81.15	88.04
ZINC12371164	0	559.94	274.99	153.55	98.57	32.83	100.90	70.88
ZINC12371364	0	572.53	347.17	35.58	149.20	40.59	40.90	100.00
ZINC19839291	0	530.87	266.44	59.68	115.90	88.85	33.40	76.46
ZINC12218805	0	622.83	306.26	118.12	161.04	37.41	72.77	79.50
ZINC03851018	0	697.06	253.48	150.39	255.01	38.18	113.55	93.99
ZINC19169013	0	558.29	362.22	84.47	111.60	0.00	76.99	71.49
ZINC08456196	0	747.37	336.67	47.85	261.15	101.70	64.98	100.00
~~ZINC19075501~~	~~3~~	~~577.23~~	~~30.01~~	~~141.99~~	~~164.59~~	~~**240.65**~~	~~77.24~~	~~100.00~~
ZINC09196338	0	603.32	521.45	71.13	10.74	0.00	57.65	88.89
ZINC21231292	0	631.54	232.41	190.71	187.04	21.39	133.69	73.61

ADME properties of three molecules (ZINC19868300, ZINC03043149 and ZINC19075501) were outside the acceptable range of FISA and WPSA and 32 compounds remained after this step.

In addition to ADME properties, druglikeness properties were also predicted by QikProp and were assessed according to Lipinski`s Rule of Five [65]. It considers molecular weight (< 500 Da), number of hydrogen bond donors (≤ 5) and acceptors (≤ 10) and octanol-water partition coefficient (≤ 5). Octanol-water partition coefficient (QPlogPo/w) gives an estimate about compound`s lipophilicity; up to a certain limit, higher lipophilicity results in permeation through biological membranes easily but can lead to a poor aqueous solubility. For each compound, the druglikeness properties are represented in Table 4.6. For all hits in the list, no violations to these properties were observed.

Table 4.6. Druglikeness properties for selected 35 hits of HRH hypothesis.

ZINC ID	Mol_MW (Da)	DonorHB	AccptHB	QPlogPo/w	Rule of Five
	0.0 500.0	0.0 5.0	0.0 10.0	-2.0 6.5	# violations
ZINC12378903	432.44	2.00	6.25	4.29	0
ZINC16891564	410.54	5.00	9.50	2.28	0
ZINC00186392	318.18	3.00	6.00	1.56	0
ZINC01275072	380.29	2.00	3.50	4.39	0
ZINC40984430	467.54	1.00	8.00	4.24	0
ZINC08627941	383.92	4.00	3.50	4.58	0
ZINC16383080	300.74	1.00	4.50	2.69	0
ZINC01071498	397.43	0.00	7.00	3.35	0
ZINC40984473	450.54	1.00	9.00	3.47	0
ZINC02168326	305.50	2.00	1.75	4.83	0
ZINC26781246	343.42	3.00	8.35	1.92	0
ZINC40959674	472.54	1.00	9.00	4.18	0
ZINC04013595	208.32	3.00	1.50	2.43	0
ZINC16638691	326.44	1.00	6.00	3.47	0
ZINC19798670	325.37	4.00	5.00	2.54	0
ZINC12297152	377.44	0.00	8.95	2.01	0
ZINC16736490	401.44	1.00	9.00	2.30	0
ZINC01649089	253.40	1.00	4.20	3.00	0
ZINC00099607	285.34	1.00	4.00	3.84	0
ZINC17020109	262.74	1.00	2.25	4.06	0
ZINC40959641	473.53	1.00	10.00	3.51	0
ZINC27675619	322.48	1.00	4.50	3.74	0
ZINC12279762	413.52	0.00	9.00	2.83	0
ZINC12371164	308.35	2.00	5.50	1.58	0
ZINC12371364	305.39	1.00	5.00	2.97	0
ZINC19839291	270.32	1.00	5.70	1.66	0
ZINC12218805	318.39	1.00	5.50	2.00	0
ZINC03851018	398.48	4.00	5.50	3.60	0
ZINC19169013	279.34	2.00	5.00	1.53	0
ZINC08456196	433.91	1.00	6.00	4.98	0
ZINC09196338	278.40	1.00	6.00	2.27	0
ZINC21231292	391.44	2.00	6.00	2.22	0

The interaction diagrams were generated for the remaining 32 compounds using Structure Interaction Fingerprint Tool (SIFT) [66] of Schrödinger to identify the ligand-protein interactions. Side chain or backbone hydrogen bonding and $\pi - \pi$ stacking interactions of 32 ligands with the residues of Tpx are shown in Table 4.7.

Table 4.7. Interactions of the survived 32 hits with the residues of Tpx.

	RESIDUE							
ZINC ID	Ile 153	Glu 156	Ser 55	Phe 7	Phe 89	Ala 62	Thr 58	Phe 53
ZINC12378903	H (b)	H (s), H (s)	H (b)		π- π			
ZINC16891564	H (b)	H (s)	H (b)					
ZINC00186392	H (b)	H (s), H (s)	H (b)					
ZINC01275072	H (b)	H (s)		π- π				
ZINC40984430		H (s)						
ZINC08627941	H (b), H (b)	H (s), H (s)						
ZINC16383080		H (s)	H (b)					
ZINC01071498		H (s)						
ZINC40984473		H (s)						
ZINC02168326		H (s)						
ZINC26781246	H (b)	H (s)						
ZINC40959674		H (s)	H (b)		π- π	H (b)		
ZINC04013595	H (b)	H (s), H (s)			π- π			
ZINC16638691		H (s)						
ZINC19798670	H (b), H (b)	H (s), H (s)					H (s)	
ZINC12297152		H (s)						
ZINC16736490		H (s)	H (b)		π- π			
ZINC01649089		H (s)	H (b)					
ZINC00099607		H (s)		π- π	π- π			
ZINC17020109		H (s)						
ZINC40959641		H (s)	H (b)		π- π	H (b)		
ZINC27675619		H (s)						
ZINC12279762		H (s)					H (s)	
ZINC12371164		H (s)						
ZINC12371364		H (s)						
ZINC19839291		H (s)	H (b)					
ZINC12218805		H (s)		π- π		H (b)		
ZINC03851018	H (b)	H (s)						pi
ZINC19169013		H (s)	H (b)					
ZINC08456196		H (s)	H (b)					
ZINC09196338		H (s)	H (b)					
ZINC21231292		H (s), H (s)			π- π	H (b)		

All ligands made side chain hydrogen bonding interactions with Glu 156 of chain B. Furthermore, twelve ligands made hydrogen bonds with the Ser 55 backbone and nine ligands made hydrogen bonds with Ile 153 backbone. Also, seven ligands made $\pi - \pi$ stacking interactions with Phe 89 of chain A.

These 32 compounds were examined by pairwise similarity based on the Tanimoto Coefficient (TC) which has a range from 0 to 1 with higher values indicating higher similarity. Compounds with higher than 0.4 pairwise similarity values were divided into three clusters (Cluster 1: ZINC40984430, ZINC40984473 and ZINC08456196, Cluster 2: ZINC40959674 and ZINC40959641, Cluster 3: ZINC12371164, ZINC12371364 and ZINC19169013) with single-linkage hierarchical clustering via Canvas module [52]. BEI and SEI values of the compounds in the clusters are displayed in Figure 4.6.

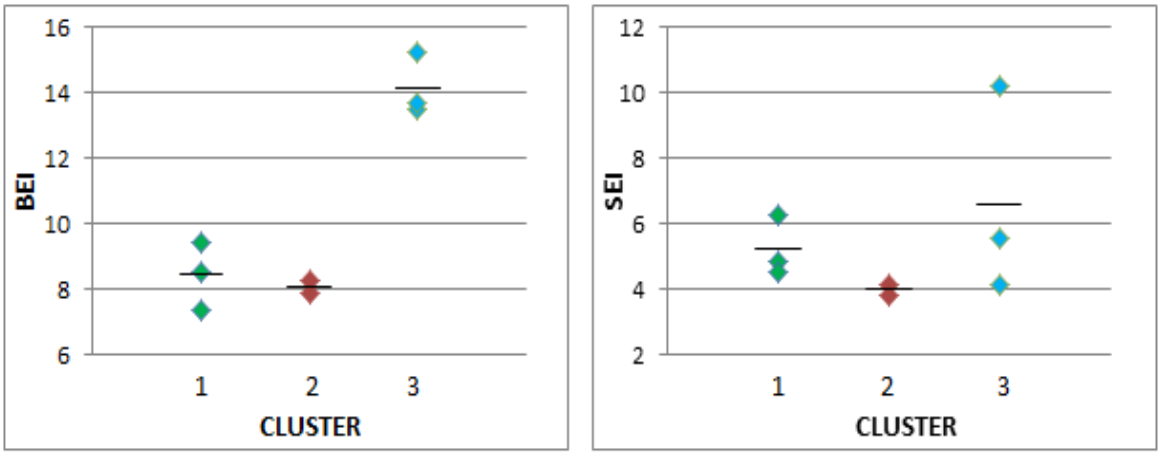


Figure 4.6. Binding efficiency (left) and surface efficiency (right) indices for the compounds in three clusters. The average BEI and SEI values of the compounds in each cluster are displayed by strikethrough.

The compounds of cluster 3 had high BEI values implying more druglikeness [62, 64]. In addition, BEI and SEI values of the compounds in cluster 1 were similar and higher than the values of the compounds in cluster 2. Therefore, cluster 1 and cluster 3 were used in substructure search. The common scaffold of each cluster (1 and 3) and seven compounds with high BEI values which did not fall into any cluster were used as query in substructure search of the ZINC database (Table 4.8).

Table 4.8. 2D structures of the ligands in clusters 1 and 7 and of the ligands having BEI values higher than 14 and common scaffolds of the ligands in each cluster.

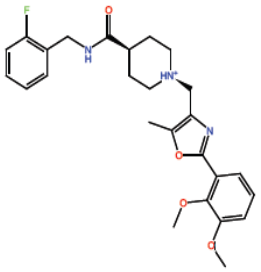
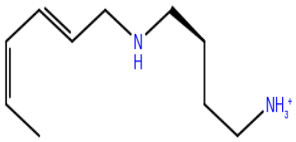
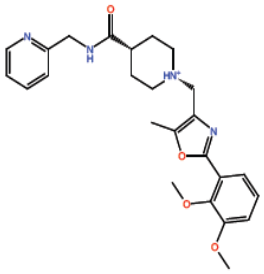
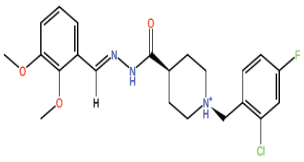
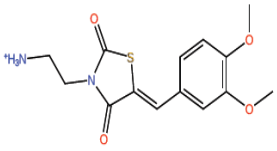
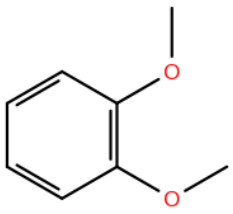
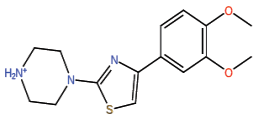
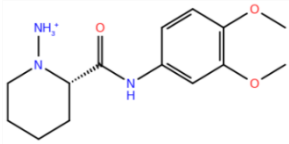
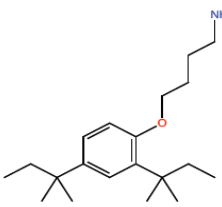
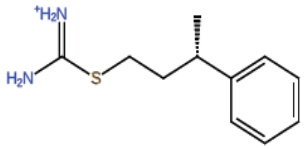
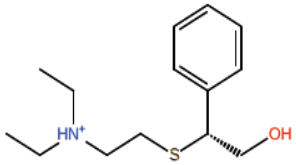
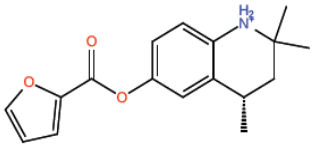
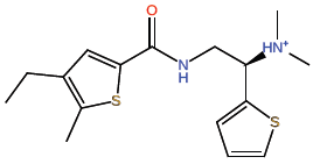
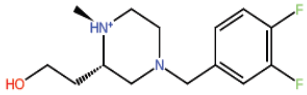
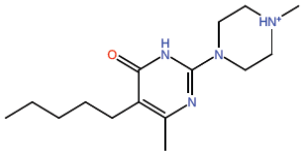
ZINC ID	Cluster	BEI	2D Structure	Common scaffold
ZINC409844430	1	7.38		
ZINC40984473		8.55		
ZINC08456196		9.43		
ZINC12371164	3	13.50		
ZINC12371364		13.68		
ZINC19169013		15.25		
ZINC02168326	-	16.05		-

Table 4.11. 2D structures of the ligands in clusters 1 and 7 and of the ligands having BEI values higher than 14 and common scaffolds of the ligands in each cluster (cont.).

ZINC ID	Cluster	BEI	2D Structure	Common scaffold
ZINC04013595	-	18.35		-
ZINC01649089	-	16.72		-
ZINC00099607	-	15.30		-
ZINC27675619	-	14.42		-
ZINC19839291	-	14.64		-
ZINC09196338	-	16.43		-

Substructure search with the high ranking compounds did not yield any new compounds but using the chemical scaffolds of cluster 1 and cluster 3 yielded a novel set of compounds that were missed in the structure-based pharmacophore modeling and docking of filtered compounds with HRH hypothesis. Figure 4.7 displays the binding site of Tpx and the compatibility of HRH hypothesis with representative ligands of clusters 1 and 3.

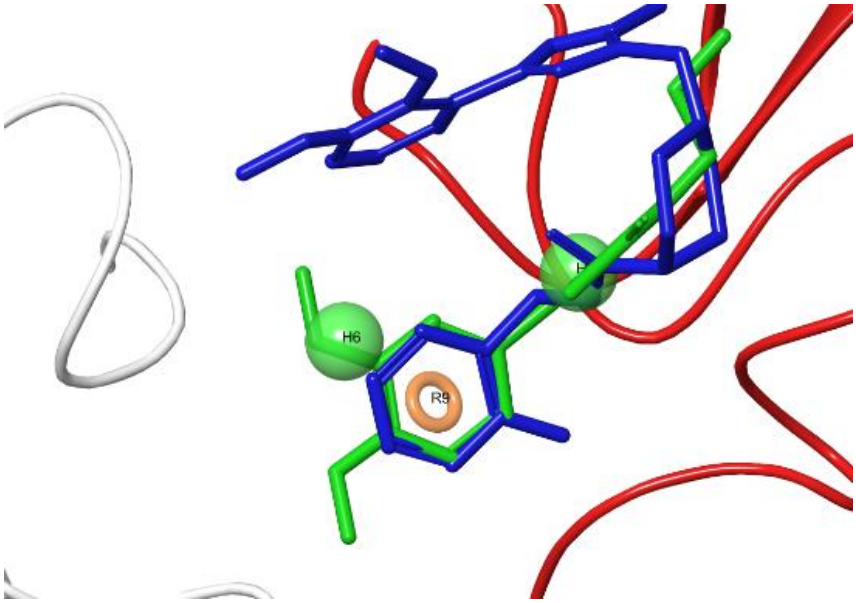


Figure 4.7. Tpx binding site with HRH hypothesis and representative ligands of cluster 1 (blue) and cluster 3 (green). The fitness values of these ligands to the HRH hypothesis are 1.97 and 1.96, respectively.

4.2.3. Substructure search with the scaffold of cluster 1

The substructure search with the scaffold of cluster 1 (Table 4.8), which includes n-benzylformamide that is used for hair dye preparations, generated 1350 hits with GlideGscores between -9.5 and -2.4 kcal/mol. 1222 of these ligands were not obtained in the first hit set found from the molecular docking of HRH hypothesis. Since the three ligands in cluster 1 had GlideGscores between -8.3 and -7.1 kcal/mol and the average GlideGscore value of the 32 hits retrieved from HRH hypothesis was -7.6 kcal/mol, the cutoff value in substructure search was set to -8.0 kcal/mol and 17 hits were retrieved. 15 of these hits were not obtained in the hit set retrieved from the docking of HRH hypothesis. The fitness values of 17 ligands to the HRH hypothesis were also obtained.

Strain-energy values of these 17 hits were calculated and 14 ligands were eliminated. GlideGscores, fitness values to the HRH hypothesis and strain-corrected scores are given in Table 4.9.

Table 4.9. GlideGscores, fitness values to HRH hypothesis and strain corrected Gscores of 17 ligands obtained from substructure search of cluster 1. Eliminated molecules are indicated by strikethrough.

ZINC ID	Glide GScore	Fitness value	Bound energy (kcal/mol)	Free energy (kcal/mol)	Strain energy (kcal/mol)	Strain penalty	Strain Gscore
ZINC55099370	-9.53	1.35	55.06	51.11	3.95	0.00	-9.53
~~ZINC12305346~~	~~-9.17~~	~~0.99~~	~~78.62~~	~~57.28~~	~~21.35~~	~~4.34~~	~~-4.83~~
~~ZINC12298763~~	~~-8.94~~	~~-0.32~~	~~79.07~~	~~38.97~~	~~40.10~~	~~9.03~~	~~0.09~~
ZINC12306731	-8.85	0.56	47.87	43.96	3.91	0.00	-8.85
~~ZINC57991179~~	~~-8.62~~	~~0.73~~	~~53.62~~	~~46.26~~	~~7.37~~	~~0.84~~	~~-7.78~~
~~ZINC12315897~~	~~-8.46~~	~~-0.22~~	~~57.65~~	~~42.13~~	~~15.51~~	~~2.88~~	~~-5.58~~
~~ZINC12303764~~	~~-8.45~~	~~1.11~~	~~73.24~~	~~62.84~~	~~10.40~~	~~1.60~~	~~-6.85~~
~~ZINC12373325~~	~~-8.45~~	~~0.66~~	~~65.79~~	~~57.32~~	~~8.47~~	~~1.12~~	~~-7.33~~
~~ZINC01650373~~	~~-8.43~~	~~0.48~~	~~30.33~~	~~21.19~~	~~9.15~~	~~1.29~~	~~-7.14~~
~~ZINC04435023~~	~~-8.34~~	~~1.01~~	~~51.36~~	~~33.97~~	~~17.39~~	~~3.35~~	~~-4.99~~
~~ZINC19792649~~	~~-8.24~~	~~0.36~~	~~84.99~~	~~39.99~~	~~45.00~~	~~10.25~~	~~2.52~~
~~ZINC04435086~~	~~-8.23~~	~~1.16~~	~~38.82~~	~~32.09~~	~~6.73~~	~~0.68~~	~~-7.55~~
ZINC12373624	-8.19	1.53	47.21	44.64	2.57	0.00	-8.19
~~ZINC20102922~~	~~-8.11~~	~~-0.67~~	~~40.19~~	~~29.65~~	~~10.54~~	~~1.64~~	~~-6.47~~
~~ZINC19799304~~	~~-8.09~~	~~0.25~~	~~84.55~~	~~41.59~~	~~42.95~~	~~9.74~~	~~1.65~~
~~ZINC20450657~~	~~-8.06~~	~~0.77~~	~~70.73~~	~~63.83~~	~~6.90~~	~~0.73~~	~~-7.33~~
~~ZINC04550415~~	~~-8.05~~	~~-0.31~~	~~50.46~~	~~30.31~~	~~20.15~~	~~4.04~~	~~-4.01~~

The remaining 3 ligands were further analyzed for ADME and druglikeness properties. No violations to these properties existed (Table 4.10, Table 4.11). BEI and SEI values of three hits were also calculated (Table 4.12).

Table 4.10. ADME properties for three hits found from substructure search of cluster 1.

ZINC ID	# stars	SASA	FOSA	FISA	PISA	WPSA	PSA	HOA %
		300.0 1000.0	0.0 750.0	7.0 330.0	0.0 450.0	0.0 175.0	7.0 200.0	<25% low >80% high
ZINC55099370	0	712.53	414.71	56.02	214.98	26.82	50.34	72.98
ZINC12306731	1	749.77	462.98	29.06	257.73	0.00	35.42	93.01
ZINC12373624	0	739.56	522.19	15.17	169.70	32.50	32.93	100.00

Table 4.11. Druglikeness properties for three hits obtained from substructure search of cluster 1.

ZINC ID	Mol_MW (Da)	DonorHB	AccptHB	QPlogPo/w	Rule of Five
	0.0 500.0	0.0 5.0	0.0 10.0	-2.0 6.5	# violations
ZINC55099370	432.44	2.00	6.25	4.29	0
ZINC12306731	410.54	5.00	9.50	2.28	0
ZINC12373624	318.18	3.00	6.00	1.56	0

Table 4.12. Efficiency indices of three ligands.

ZINC ID	BEI	SEI
ZINC55099370	5.38	4.44
ZINC12306731	6.13	7.08
ZINC12373624	6.76	9.08

The ligand-protein interaction diagrams were generated for three ligands. Side chain or backbone hydrogen bonding and $\pi - \pi$ stacking interactions of these ligands with the residues of Tpx are displayed in Table 4.13.

Table 4.13. Interactions of three ligands with the residues of Tpx.

	RESIDUE					
ZINC ID	Ile 153	Glu 156	Phe 7	Phe 89	Asp 57	Thr 154
ZINC55099370	H (b)	H (s)		π- π	H (b)	H (b)
ZINC12306731		H (s)		π- π	H (b)	
ZINC12373624	H (b)	H (s)	π- π			

These three ligands made side chain hydrogen bonding interactions with Glu 156. Two out of three ligands also made backbone hydrogen bonding interactions with Ile 153 and Asp 57, and π- π stacking interaction with Phe 89.

4.2.4. Substructure search with the scaffold of cluster 3

When the common scaffold of cluster 3 (Table 4.8), 1,2-dimethoxybenzene that exerts an excellent sedative effect, was used in substructure search of ZINC database, 3024 hits were found with GlideGscores between -9.7 and -2.0 kcal/mol. Out of 3024 hits, 2792 were different from the 995 hits obtained after docking. In order to be consistent with the case of cluster 1, the cutoff Gscore value was set to -8.0 kcal/mol and 82 ligands were retrieved and fitness values of these ligands to the HRH hypothesis were obtained. Out of these 82, 57 ligands with high strain were eliminated. The strain corrected Gscores and fitness values of 25 ligands to the HRH hypothesis are represented in Table 4.14.

Table 4.14. Strain Gscores and fitness values of the survived 25 ligands to the HRH hypothesis.

ZINC ID	Strain GScore	Fitness value
ZINC12298678	-9.03	1.30
ZINC12243105	-8.91	1.07
ZINC08497999	-8.88	1.42
ZINC32616896	-8.87	1.13
ZINC12315856	-8.68	-0.24
ZINC19792239	-8.62	-0.47
ZINC12298618	-8.38	0.23
ZINC05955710	-8.36	0.34
ZINC00359882	-8.29	-0.18
ZINC08497972	-8.28	0.74
ZINC12867573	-8.28	0.13
ZINC19691207	-8.24	0.98
ZINC09357466	-8.19	-0.53
ZINC05241975	-8.17	1.30
ZINC20309176	-8.17	0.36
ZINC08497993	-8.17	-0.39
ZINC06703639	-8.15	0.43
ZINC40959760	-8.11	0.17
ZINC05910022	-8.10	1.37
ZINC40959677	-8.06	-0.67
ZINC32616892	-8.03	0.72
ZINC40959801	-8.03	-0.19
ZINC08428543	-8.03	0.36
ZINC48711249	-8.01	-0.52
ZINC12373528	-8.00	-0.48

The remaining 25 compounds were further analyzed for ADME and druglikeness properties to find any violations (Table 4.15, Table 4.16).

Table 4.15. ADME properties for 25 hits found from substructure search of cluster 3.

ZINC ID	# stars	SASA	FOSA	FISA	PISA	WPSA	PSA	HOA %
		300.0 1000.0	0.0 750.0	7.0 330.0	0.0 450.0	0.0 175.0	7.0 200.0	<25% low >80% high
ZINC12298678	0	780.48	419.89	54.63	228.98	76.97	54.60	91.34
ZINC12243105	0	758.54	380.82	55.97	321.75	0.00	57.10	100.00
ZINC08497999	0	629.64	273.10	56.14	261.99	38.41	46.86	100.00
ZINC32616896	0	755.40	464.57	102.54	188.29	0.00	79.31	92.00
ZINC12315856	0	841.20	589.34	60.27	191.59	0.00	66.00	90.42
ZINC19792239	1	825.94	474.02	106.29	245.63	0.00	82.98	76.59
ZINC12298618	0	738.68	603.09	33.15	102.44	0.00	58.18	89.32
ZINC05955710	0	620.13	349.56	70.84	199.73	0.00	72.26	91.60
ZINC00359882	0	650.17	289.39	45.09	244.04	71.64	43.21	100.00
ZINC08497972	0	637.22	276.61	56.09	257.49	47.02	47.11	100.00
ZINC12867573	0	700.38	466.87	95.82	137.69	0.00	82.30	79.01
ZINC19691207	0	660.84	372.06	164.97	59.28	64.53	115.49	43.38
ZINC09357466	0	754.00	203.84	183.82	366.33	0.00	129.04	86.38
ZINC05241975	0	661.32	345.60	31.67	210.09	73.96	46.87	100.00
ZINC20309176	0	684.95	306.99	156.35	162.19	59.41	113.46	77.77
ZINC08497993	0	640.77	281.62	54.79	304.36	0.00	47.03	100.00
ZINC06703639	0	660.02	287.33	56.24	259.43	57.02	45.98	100.00
ZINC40959760	1	892.92	536.58	96.25	260.10	0.00	101.56	88.29
ZINC05910022	0	551.26	230.28	20.71	261.86	38.41	26.36	100.00
ZINC40959677	1	900.34	532.26	96.12	271.96	0.00	100.20	88.54
ZINC32616892	0	772.90	511.11	62.98	198.81	0.00	58.48	100.00
ZINC40959801	1	604.56	360.69	37.98	142.15	63.74	41.40	100.00
ZINC08428543	0	768.95	561.07	95.55	112.32	0.00	81.15	88.04
ZINC48711249	0	559.94	274.99	153.55	98.57	32.83	100.90	70.88
ZINC12373528	1	572.53	347.17	35.58	149.20	40.59	40.90	100.00

Table 4.16. Druglikeness properties for 25 hits obtained by substructure search of cluster 3.

ZINC ID	Mol_MW (Da)	DonorHB	AccptHB	QPlogPo/w	Rule of Five
	0.0 500.0	0.0 5.0	0.0 10.0	-2.0 6.5	# violations
ZINC12298678	408.49	2.00	6.70	4.05	0
ZINC12243105	485.63	1.00	8.00	4.77	0
ZINC08497999	319.38	2.00	4.70	3.48	0
ZINC32616896	399.49	2.00	6.70	3.71	0
ZINC12315856	444.62	2.00	8.20	4.06	0
ZINC19792239	444.57	3.00	9.15	3.03	0
ZINC12298618	366.50	1.00	7.70	3.09	0
ZINC05955710	367.40	1.00	6.75	2.72	0
ZINC00359882	349.86	2.00	4.20	4.27	0
ZINC08497972	319.38	2.00	4.70	3.55	0
ZINC12867573	347.41	1.00	7.50	2.06	0
ZINC19691207	390.28	5.00	8.70	-0.21	0
ZINC09357466	461.47	1.00	9.00	3.27	0
ZINC05241975	362.86	1.00	5.70	3.99	0
ZINC20309176	460.91	0.00	8.00	2.84	0
ZINC08497993	301.39	2.00	4.70	3.44	0
ZINC06703639	335.83	2.00	4.70	3.83	0
ZINC40959760	480.56	1.00	9.75	3.51	0
ZINC05910022	275.32	1.00	3.00	3.50	0
ZINC40959677	480.56	1.00	9.75	3.57	0
ZINC32616892	395.50	1.00	5.00	4.99	0
ZINC40959801	482.60	1.00	9.00	4.15	0
ZINC08428543	355.43	2.00	8.90	1.71	0
ZINC48711249	355.44	1.00	6.20	3.56	0
ZINC12373528	391.53	2.00	8.20	2.92	0
ZINC12298678	408.49	2.00	6.70	4.05	0
ZINC12243105	485.63	1.00	8.00	4.77	0
ZINC08497999	319.38	2.00	4.70	3.48	0
ZINC32616896	399.49	2.00	6.70	3.71	0
ZINC12315856	444.62	2.00	8.20	4.06	0
ZINC19792239	444.57	3.00	9.15	3.03	0
ZINC12298618	366.50	1.00	7.70	3.09	0

For all hits in the list, no violations to these properties were observed. The hit list of 995 compounds obtained from the docking operations of HRH hypothesis did not include these final hits.

Binding efficiency and surface efficiency index values of these 25 ligands were also calculated (Table 4.17).

Table 4.17. Efficiency index values of the survived 25 ligands.

ZINC ID	BEI	SEI
ZINC12298678	5.79	4.33
ZINC12243105	5.05	4.29
ZINC08497999	6.02	4.10
ZINC32616896	6.24	3.14
ZINC12315856	5.94	4.00
ZINC19792239	6.04	3.24
ZINC12298618	8.45	5.32
ZINC05955710	7.22	3.67
ZINC00359882	8.37	6.77
ZINC08497972	4.20	2.83
ZINC12867573	8.44	3.56
ZINC19691207	7.59	2.56
ZINC09357466	6.10	2.18
ZINC05241975	8.28	6.41
ZINC20309176	6.54	2.66
ZINC08497993	10.00	6.41
ZINC06703639	9.02	6.59
ZINC40959760	6.34	3.00
ZINC05910022	11.11	11.60
ZINC40959677	6.44	3.09
ZINC32616892	7.86	5.31
ZINC40959801	6.45	3.36
ZINC08428543	9.23	3.27
ZINC48711249	8.80	6.14
ZINC12373528	8.02	4.76

The survived 25 compounds were further assessed by ligand-protein interactions (Table 4.18).

Table 4.18. Interactions of the survived 25 ligands with the residues of Tpx.

	RESIDUE							
ZINC ID	Ile 153	Glu 156	Ser 55	Phe 7	Phe 89	Ala 62	Asp 57	Thr 154
ZINC12298678	H (b)	H (s), H (s)			π- π			
ZINC12243105		H (s)	H (b)					
ZINC08497999	H (b)	H (s)						
ZINC32616896		H (s)				H (b)		
ZINC12315856	H (b)	H (s)	H (b)		π- π			
ZINC19792239		H (s)				H (b)		
ZINC12298618	H (b)	H (s)	H (b)					
ZINC05955710	H (b)	H (s)	H (b)					
ZINC00359882		H (s)		π- π		H (b)		
ZINC08497972		H (s)			π- π			
ZINC12867573		H (s)	H (b)					
ZINC19691207	H (b)	H (s)					H (b)	H (b)
ZINC09357466		H (s)					H (b)	
ZINC05241975		H (s)						
ZINC20309176	H (b)						H (b)	
ZINC08497993		H (s)				H (b)		
ZINC06703639		H (s)				H (b)		
ZINC40959760		H (s)	H (b)					
ZINC05910022		H (s)						
ZINC40959677		H (s)	H (b)	π- π				
ZINC32616892		H (s)						
ZINC40959801		H (s)	H (b)					
ZINC08428543		H (s), H (s)				H (b)		
ZINC48711249		H (s)						
ZINC12373528	H (b)	H (s)						

The survived 25 ligands interacted mostly with Ile 153, Glu 156, Ser 55 and Ala 62 residues of Tpx. 24 out of 25 ligands made side chain hydrogen bonding interactions with Glu 156. 8 hits made backbone hydrogen bonding interactions with Ser 55 and Ile 153. Six ligands made backbone hydrogen bonding interactions with Ala 62.

The locations of final hits obtained by substructure searches of clusters 1 and 3 on Tpx binding site with the representative ligands of cluster 1 (ZINC40984430) and cluster 3 (ZINC12371164) of HRH hypothesis are given in Figure 4.8.

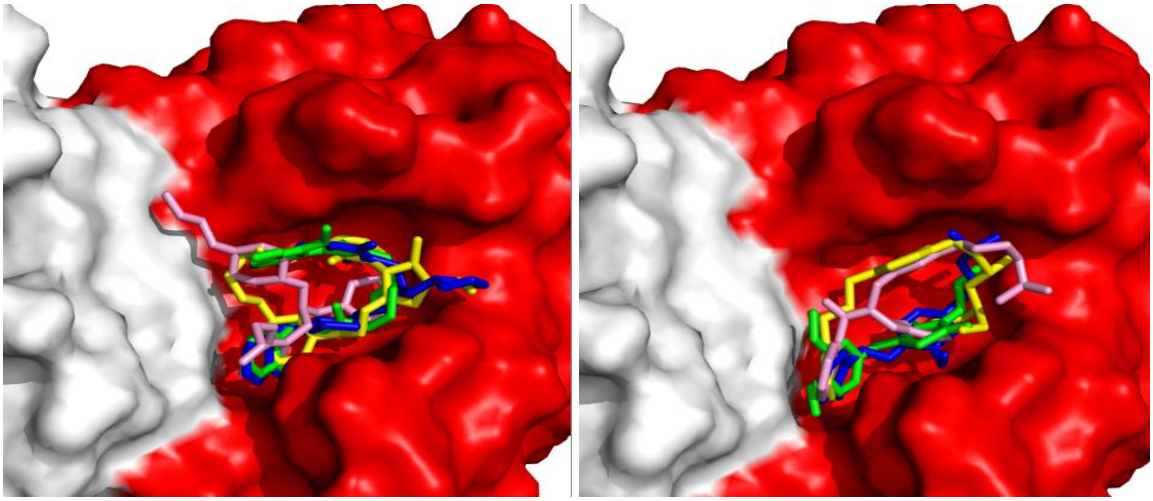


Figure 4.8. Three highest docking scored ligands obtained by substructure search of cluster 1 (left) and cluster 3 (right) in yellow, pink and blue. The representative ligands of cluster 1 (ZINC40984430) and cluster 3 (ZINC12371164) are also shown in green.

4.3. Ligand-Based Pharmacophore Modeling Based on Previously Identified Tpx Inhibitors

The salicylidene acylhydrazide type inhibitors of Tpx were used to build ligand-based pharmacophore models. 2D structures of 27 experimentally identified inhibitors in previous studies [5, 27] were converted into 3D via LigPrep [39] with different tautomerization and ionization states at neutral pH. For twenty-seven inhibitors, 570 energetically minimized conformers were generated by ConfGen [41] and were docked to Tpx using Glide XP without any constraints. For each inhibitor, the conformer which had the highest GlideGscore value was kept for protein-ligand interaction analysis. For the 23 inhibitors, GlideGscores of their highest ranked conformer and the key interactions that they made with Tpx are listed in Table 4.19.

Table 4.19. Docking scores and Tpx interactions of experimentally [5, 27] identified Tpx inhibitors.

Compound	GlideGscore (kcal/mol)	Interactions with Tpx
1	-5.91	Asp 57, Ser 55
2	-6.18	Ile 153, Ser 55
3	-4.88	Arg 93, Asp 57, Phe 89, Ser 55
4	-7.90	Glu 156, Ile 153, Phe 53, Ser 55
5	-7.10	Glu 156, Ser 55
6	-5.16	Ile 153, Phe 7, Phe 53, Ser 55
7	-6.04	Ile 153, Ser 55
8	-5.29	Ile 153
9	-5.44	Asp 57, Phe 89
10	-5.47	Phe 53, Ser 55
11	-5.24	Thr 58
12	-5.00	Asp 57, Phe 53, Phe 89, Ser 55
13	-6.91	Ala 62, Ser 55
14	-5.93	Asp 57, Glu 156, Ile 153, Phe 53, Ser 55
15	-5.41	Ile 153, Phe 53, Ser 55
16	-6.39	Ser 55
17	-5.51	Ile 153, Phe 7, Phe 53
18	-4.25	Ser 55
19	-6.17	Asp 57, Phe 53
20	-5.34	Glu 156, Ile 153, Phe 89, Ser 55
21	-5.40	Ile 153, Phe 53, Ser 55
22	-6.68	Asp 57
23	-8.43	Glu 156, Ser 55
24	-8.72	Glu 156, Ile 153, Ser 55
25	-7.34	Glu 156, Ile 153, Ser 55
26	-7.07	Glu 156, Ser 55
27	-5.70	Ala 62, Ser 55

Average GlideGscore of the 27 inhibitors was -6.11 kcal/mol and the key interactions were observed between Glu 156, Ile 153, Ser 55 and Phe 53 residues of the Tpx receptor and the ligands. The location of high ranking ligands in the Tpx dimer interface is shown in Figure 4.9.

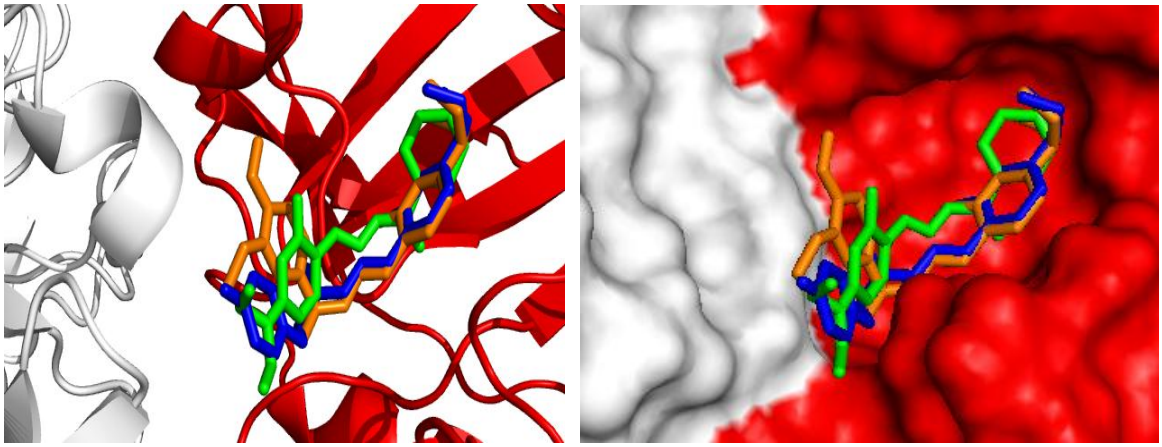


Figure 4.9. Ribbon and surface representations of Tpx (A chain in grey and B chain in red) with 3 top ranking ligands. Compounds, 4, 23 and 24, which have the highest docking scores, are in green, orange and blue.

Based on the detection of possible pharmacophore sites of 27 experimentally identified ligands and survival scores of hypotheses between three and seven sites, two pharmacophore models including five pharmacophore sites (AADDR and AAADR) were generated. The scores of 18 hypotheses produced by Phase are displayed in Table 4.20.

Table 4.20. Scores of 5-point hypotheses constructed from known Tpx inhibitors.

ID	Align Score	Vector Score	Volume Score	Survival Score
AADDR.20	0.879	0.984	0.758	3.621
AAADR.45	0.877	0.981	0.757	3.614
AAADR.8	0.697	0.962	0.697	3.356
AADDR.15	0.652	0.957	0.625	3.235
AAADR.41	0.649	0.912	0.485	3.045
AADDR.18	0.630	0.914	0.488	3.033
AAADR.42	0.649	0.912	0.453	3.014
AAADR.44	0.657	0.897	0.449	3.003
AADDR.19	0.655	0.904	0.444	3.003
AADDR.24	0.479	0.904	0.616	2.999
AADDR.22	0.655	0.881	0.448	2.984
AAADR.10	0.649	0.876	0.445	2.969
AAADR.43	0.599	0.949	0.418	2.966
AAADR.6	0.610	0.861	0.445	2.917
AADDR.21	0.520	0.948	0.447	2.915
AAADR.47	0.515	0.947	0.449	2.911
AADDR.17	0.468	0.923	0.416	2.807
AAADR.9	0.461	0.919	0.412	2.792

The quality of each hypothesis was measured by its survival score, which is a weighted combination of site, vector, and volume scores. Survival scores of top two hypotheses were fairly close to each other and subscores that constitute survival score did not point to a single hypothesis that represents all parameters best. Consequently, these two hypotheses were further investigated with QSAR method.

4.3.1. Quantitative Structure-Activity Relationship (QSAR) Model

The hypothesis that yielded the most accurate ligand activity prediction via QSAR model was chosen for database filtering rather than the highest survival score. Fitness of each ligand to the selected hypotheses is represented in Table 4.21.

Table 4.21. Fitness of all Tpx inhibitors to top two hypotheses.

Compound	AADDR.20	AAADR.45
1	2.12	2.09
2	2.91	2.86
3	1.92	1.87
4	2.72	2.69
5	2.00	1.94
6	2.86	2.89
7	2.84	2.85
8	2.23	2.20
9	2.80	2.72
10	2.35	2.34
11	2.38	2.37
12	2.76	2.76
13	2.61	2.61
14	2.71	2.72
15	2.62	2.63
16	2.89	2.90
17	3.00	2.87
18	2.61	2.62
19	2.67	2.69
20	2.84	2.84
21	2.79	2.79
22	2.84	2.78
23	2.73	2.70
24	2.60	2.58
25	2.78	2.77
26	2.78	2.78
27	2.85	2.88

The QSAR model was generated from 27 inhibitors that have a range of known activities using partial least squares (PLS) regression method, as described in methods section. IC_{50} values of these 27 inhibitors were obtained using statistical correlation of inhibition percentages at different ligand concentrations in Excel (Table 4.22, Table B.1).

Table 4.22. IC_{50} values of 27 active Tpx inhibitors.

Compound	IC_{50} (μM)	Compound	IC_{50} (μM)
1	25.56	15	39.31
2	20.44	16	71.83
3	29.02	17	32.17
4	27.04	18	23.56
5	16.89	19	45.89
6	20.11	20	46.21
7	25.76	21	42.92
8	14.49	22	31.13
9	13.31	23	26.73
10	38.01	24	39.93
11	44.18	25	43.16
12	63.38	26	50.00
13	37.18	27	66.37
14	49.15		

27 active ligands were manually divided into a training set (17 compounds) and a test set (10 compounds). Test set includes compounds 2, 5, 7, 8, 11, 12, 18, 19, 22 and 24. Regression was carried out with three PLS factor because maximum PLS factor can be no higher than twenty percent of the number of training set molecules. Robustness of the models was internally validated by statistical parameters as root mean square error, standard deviation and variance ratio (Table 4.23).

Table 4.23. QSAR results for top two hypotheses.

Set	Parameters	AADDR.20	AAADR.45
Training	SD	0.086	0.123
	R^2	0.961	0.951
	F	125.5	122.7
Test	RMSE	0.102	0.143
	Q^2	0.894	0.801
	Pearson-R	0.946	0.912

For one PLS factor, regression resulted in 0.774 R^2 and 0.737 Q^2 values and for two PLS factor, these parameters were 0.864 and 0.837, respectively.

SD value measures the strength of the fit to the training set compounds. From Table 4.23, it is seen that lowest standard deviation was observed in the hypothesis AADDR.20. R^2 value displays how model successfully interprets structure-activity relationship for training set compounds. The R^2 value for the hypothesis AADDR.20 was observed higher, although AAADR.45 had also good coverage on the training set. Large values of variance ratio (F) indicate a more statistically significant regression and AADDR.20 fits better to the training set. Therefore using three PLS factors, satisfactory model fit to the training set compounds was observed with AADDR.20. The predicted activities and known activities of training set compounds by taking the negative logarithm of the activities in micromolar (pIC_{50}) are represented in Table 4.24 and Figure 4.10.

Table 4.24. Actual and predicted activities of training set compounds for AADDR.20.

Compound	Actual Activity (pIC_{50})	Predicted Activity (pIC_{50})	Error (%)
1	4.592	4.63	0.83
3	4.537	4.56	0.51
4	4.568	4.59	0.48
6	4.697	4.73	0.70
9	4.876	4.90	0.49
10	4.420	4.45	0.68
13	4.430	4.42	0.23
14	4.308	4.34	0.74
15	4.405	4.35	1.25
16	4.144	4.11	0.82
17	4.493	4.48	0.29
20	4.335	4.31	0.58
21	4.367	4.34	0.62
23	4.573	4.55	0.50
25	4.365	4.38	0.34
26	4.301	4.29	0.26
27	4.367	4.34	0.62

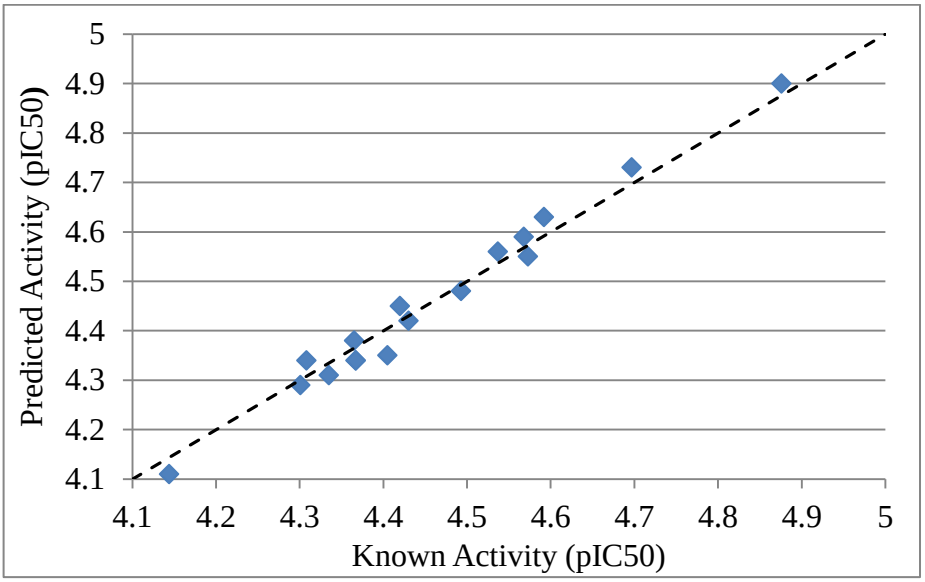


Figure 4.10. Scatter plot of activities of training set compounds for AADDR.20.

Predictive ability of the model is also important for a good QSAR model. Quality of the model fit on test set compounds was measured by RMSE, Q^2 and Pearson-R values. Q^2 is analogous to R-squared statistic and it comes from applying QSAR model to the test set. The correlation between predicted and actual activities is determined by Pearson-R value for which increased numbers are more favorable. From Table 4.23, hypothesis AADDR.20 displayed the best prediction on the test set compounds with 0.894 Q^2 and 0.946 Pearson-R values, and lower root-mean-square error as 0.102. The actual and predicted activities of test set compounds for the hypothesis AADDR.20 are displayed in Table 4.25 and Figure 4.11.

Table 4.25. Actual and predicted activities of test set compounds for AADDR.20.

Compound	Actual Activity (pIC_{50})	Predicted Activity (pIC_{50})	Error (%)
2	4.690	4.67	0.43
5	4.772	4.81	0.80
7	4.589	4.60	0.24
8	4.839	4.79	1.01
11	4.355	4.33	0.57
12	4.198	4.23	0.76
18	4.628	4.66	0.69
19	4.338	4.30	0.88
22	4.507	4.46	1.05
24	4.399	4.45	1.15

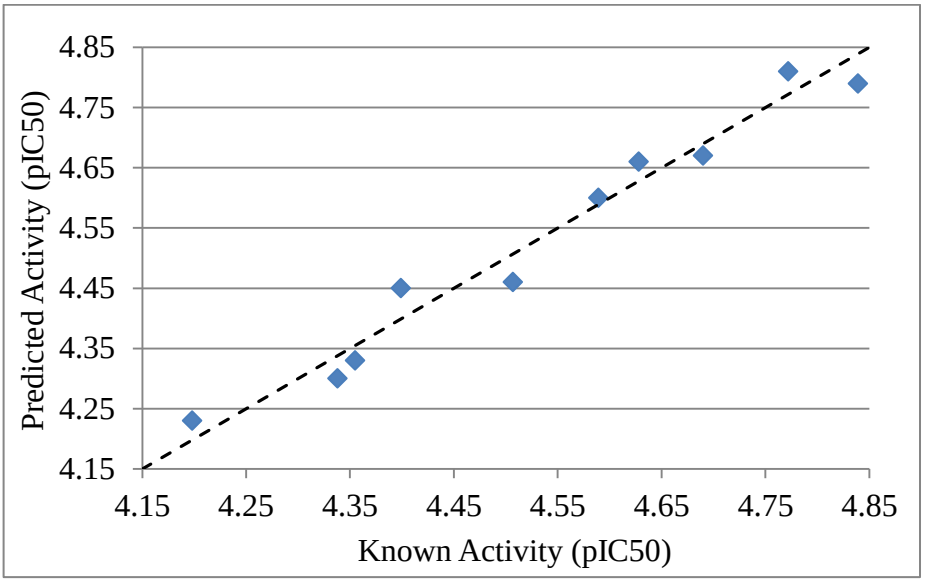


Figure 4.11. Scatter plot of activities of test set compounds for AADDR.20.

The accuracy of this QSAR model was validated by constituting other QSAR models with different training and test sets. 10 different models, which most accurately predict the activities, were selected. The compounds in test sets and the percent average error between actual and predicted activities in QSAR models are displayed in Table 4.26.

Table 4.26. The statistical parameters of different QSAR models with percentage error values between actual and predicted activities.

QSAR Model	Test Set Compounds	R^2	Q^2	Average IC_{50} Error (%)
1	2, 5, 7, 8, 11, 12, 18, 19, 22, 24	0.961	0.894	0.65
2	1, 3, 4, 6, 10, 13, 15, 20, 21, 26	0.957	0.891	0.95
3	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0.951	0.885	1.24
4	1, 2, 3, 4, 5, 6, 7, 8, 9, 27	0.947	0.881	1.29
5	1, 2, 3, 4, 5, 6, 7, 8, 9, 17	0.945	0.878	1.31
6	1, 10, 11, 12, 13, 14, 15, 16, 17, 18	0.931	0.873	2.17
7	9, 10, 11, 12, 13, 14, 15, 16, 17, 18	0.929	0.871	2.32
8	1, 19, 20, 21, 22, 23, 24, 25, 26, 27	0.921	0.863	2.75
9	9, 19, 20, 21, 22, 23, 24, 25, 26, 27	0.917	0.859	2.96
10	4, 19, 20, 21, 22, 23, 24, 25, 26, 27	0.914	0.855	3.43

Average error values between actual and predicted activity values in 10 different QSAR models are also represented in error bars (Figure 4.12). Maximum and minimum error values in each QSAR model are displayed in strikethrough.

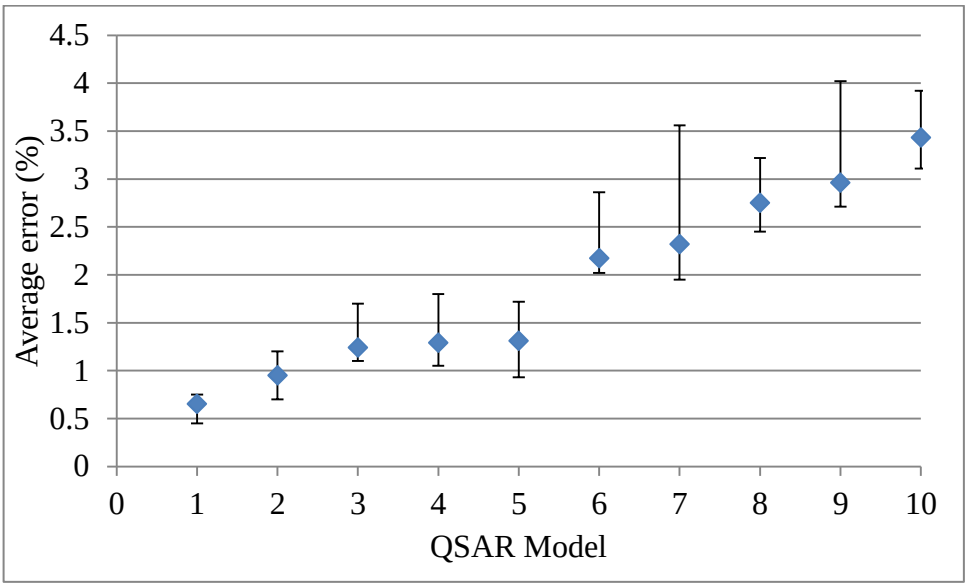


Figure 4.12. The quality of 10 QSAR models in error bars.

Overall, five-point hypothesis AADDR with two hydrogen acceptors (A), two hydrogen donors (D) and one ring group (R) gave the best three-dimensional alignment to the compounds in terms of site, volume, vector and total survival scores, and satisfactory statistical significance and predictive ability on all ligands when it was combined with a QSAR model. AADDR site points matched with compound 17 [27] as 3.00 fitness value (Figure 4.13). The pharmacophore model was used to search the ZINC database and to reduce the database to 10,000 hits with fitness values between 2.90 and 1.25.

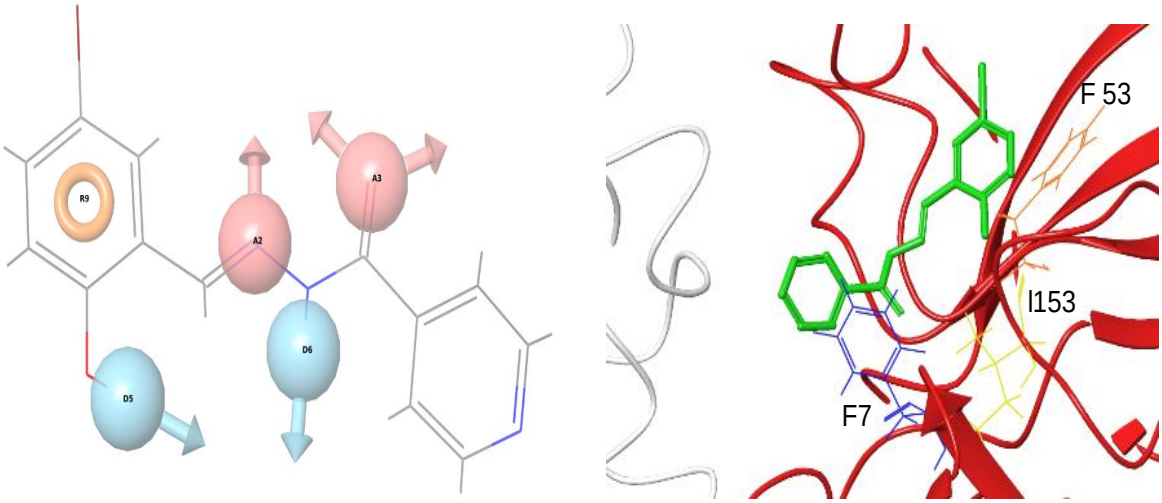


Figure 4.13. Ligand-based pharmacophore hypothesis AADDR with its sites (left). The compound 17 [27] which had fitness of 3.00 to AADDR hypothesis in green tube with Ile 153 (yellow), Phe 7 (blue) and Phe 53 (orange) (right).

4.3.2. Molecular Docking of the Selected Ligands with AADDR Hypothesis

10,000 ligands that matched the AADDR hypothesis were docked to Tpx in SP and XP modes with a positional constraint on Isoleucine 153 and 1000 compounds with GlideGscore values between -9.7 and -1.2 kcal/mol (Figure 4.14) were obtained. As the docking scores showed a relatively sharp decrease to -7.0 kcal/mol, the GlideGscore cutoff value of -7.0 kcal/mol (consistent with structure-based docking) was used for the hits of AADDR hypothesis resulting in 123 ligands.

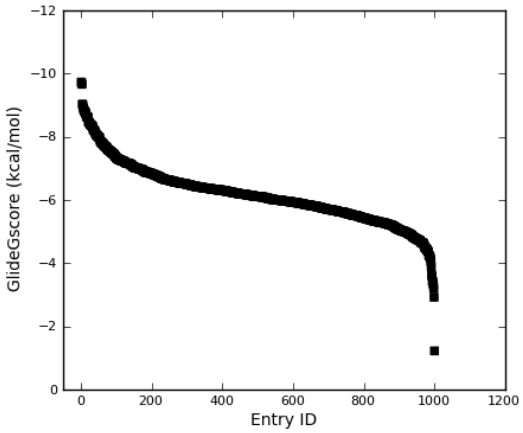


Figure 4.14. GlideGscore values for top scoring 1000 compounds filtered with AADDR hypothesis and docked using XP docking.

4.3.3. Post Docking Analyses of AADDR Hits

46 of the 123 hits were eliminated due to high ligand strain. Of the remaining 77 hits, one violated the Lipinski`s Rule with a octanol-water partition coefficient higher than five and it was eliminated. Five other hits were eliminated due to high PISA (π component of SASA) values. Finally 71 hits with high GlideGscores and good ADME properties remained. Evaluation of the BEI and SEI values showed that the inhibitors were polar (Figure 4.15). The properties and docking scores of the final 71 ligands are given in Table 4.27.

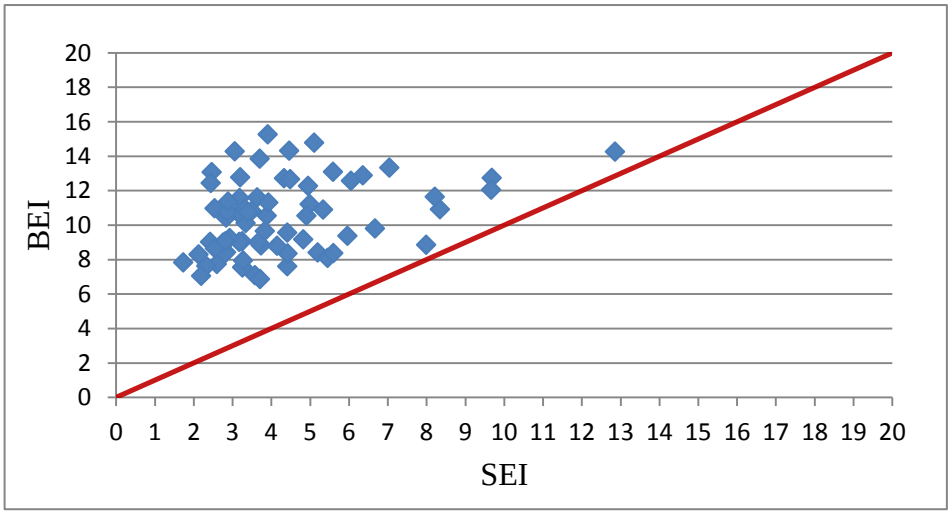


Figure 4.15. BEI versus SEI values of survived hits of AADDR hypothesis.

Table 4.27. The properties of final 71 compounds retrieved by ligand-based pharmacophore modeling based on previously identified Tpx inhibitors.

ZINC ID	Strain Gscore (kcal/mol)	Fitness	Mol_MW (Da)	SASA	QPlogPo/w	HOA %	BEI
			0.0 500.0	300.0 1000.0	-2.0 6.5	<25% low >80% high	
ZINC12314954	-8.96	1.59	343.40	652.32	3.07	89.42	7.09
ZINC26948418	-8.82	1.34	357.74	593.19	0.94	48.74	7.84
ZINC00803235	-8.82	1.47	363.46	696.44	3.67	100.00	7.61
ZINC02692864	-8.73	1.44	313.44	683.23	4.28	100.00	10.93
ZINC05073475	-8.68	1.75	343.42	650.35	3.26	100.00	8.42
ZINC02001245	-8.63	1.32	230.27	480.39	1.23	69.72	11.58
ZINC12928831	-8.48	1.28	359.34	707.22	2.76	71.80	7.74
ZINC01569737	-8.47	1.33	261.32	523.84	-0.26	56.44	10.66
ZINC04279523	-8.44	1.36	350.25	604.34	3.74	100.00	8.86
ZINC20309142	-8.36	1.51	435.51	686.38	3.75	80.84	7.57
ZINC01292370	-8.34	1.29	419.33	682.42	4.57	93.78	6.87
ZINC00625890	-8.33	1.36	402.45	709.81	1.44	67.32	8.76
ZINC57992380	-8.27	1.25	398.46	735.35	1.57	72.28	8.42
ZINC13810826	-8.22	1.26	416.48	686.09	1.85	75.81	9.06
ZINC26780521	-8.19	1.48	299.41	639.47	1.85	100.00	12.75
ZINC20309181	-8.16	1.39	445.86	675.26	2.19	60.46	7.64
ZINC04102500	-8.10	1.36	436.90	734.66	1.91	70.03	8.36
ZINC20309890	-8.04	1.40	452.55	746.34	3.75	85.02	8.82
ZINC01252080	-8.00	1.27	284.74	491.57	-1.64	37.48	13.08
ZINC02339978	-7.98	1.47	388.85	639.00	4.93	100.00	8.10
ZINC20309903	-7.87	1.53	439.47	755.14	2.60	62.17	8.98
ZINC02227339	-7.85	1.47	368.43	644.55	4.74	100.00	9.38
ZINC01718125	-7.78	1.25	283.37	599.57	3.40	100.00	11.65
ZINC16545157	-7.75	1.68	266.25	500.91	-0.79	53.64	12.45
ZINC01575542	-7.72	1.41	369.42	649.68	-0.37	38.82	9.03
ZINC01283334	-7.71	1.30	350.44	652.87	3.74	89.92	9.57
ZINC36500739	-7.69	1.44	496.52	731.48	2.14	57.42	7.06
ZINC00633243	-7.68	1.26	414.46	679.85	1.71	73.40	9.03
ZINC02337617	-7.67	1.47	382.46	645.37	4.94	100.00	9.81
ZINC18249750	-7.65	1.64	356.43	696.16	1.94	73.40	10.13
ZINC16525119	-7.64	1.52	238.24	470.82	-0.87	59.99	14.28
ZINC20309809	-7.60	1.57	455.47	749.58	2.16	56.79	8.65
ZINC11592555	-7.60	1.62	308.38	614.67	-0.23	41.52	12.77
ZINC02443796	-7.57	1.48	388.85	622.41	4.71	100.00	8.38
ZINC00536176	-7.57	1.34	289.35	515.77	2.31	88.17	12.90
ZINC01700848	-7.55	1.31	256.24	493.17	-0.14	70.65	14.79
ZINC60382125	-7.51	1.31	330.43	678.66	2.84	91.22	13.34
ZINC32836064	-7.51	1.37	352.31	617.35	2.32	80.42	10.56

Table 4.27. The properties of final 71 compounds retrieved by ligand-based pharmacophore modeling based on previously identified Tpx inhibitors (cont.).

ZINC ID	Strain Gscore (kcal/mol)	Fitness	Mol_MW (Da)	SASA	QPlogPo/w	HOA %	ZINC ID
ZINC12385542	-7.49	1.31	360.43	686.93	3.14	89.77	10.91
ZINC01247044	-7.47	1.30	300.38	548.46	2.12	82.69	12.58
ZINC19877995	-7.46	1.34	298.29	509.18	0.32	82.79	12.71
ZINC26949215	-7.45	1.40	372.83	626.49	0.78	68.86	10.72
ZINC00127686	-7.40	1.79	307.36	609.22	1.60	69.85	11.60
ZINC00806563	-7.39	1.67	356.47	712.14	2.86	87.14	11.22
ZINC06193011	-7.37	1.39	470.59	711.02	3.42	94.78	8.35
ZINC11971663	-7.31	1.29	381.47	633.67	2.05	96.99	10.55
ZINC06746173	-7.30	1.66	314.38	553.51	1.82	85.30	13.10
ZINC04649681	-7.28	1.37	322.41	601.95	4.25	100.00	12.06
ZINC16925214	-7.26	1.28	307.35	534.85	0.43	73.30	14.32
ZINC07977647	-7.26	1.29	324.26	566.41	3.75	100.00	11.32
ZINC39612159	-7.26	1.43	442.86	747.73	2.77	65.85	8.31
ZINC46587617	-7.23	1.28	424.46	721.96	4.77	100.00	9.17
ZINC02522626	-7.23	1.42	277.37	579.44	0.08	52.32	15.27
ZINC01312082	-7.19	1.26	303.32	561.12	1.00	100.00	13.86
ZINC36485022	-7.16	1.34	468.94	711.60	3.65	91.96	9.01
ZINC04982797	-7.15	1.42	344.16	546.64	0.42	68.40	10.93
ZINC09310251	-7.14	1.76	462.55	832.57	4.09	100.00	8.80
ZINC12561670	-7.14	1.32	499.57	762.06	4.60	100.00	7.53
ZINC02373214	-7.14	1.38	426.87	690.45	2.10	77.76	9.25
ZINC08393981	-7.13	1.61	389.34	602.99	3.01	90.41	9.67
ZINC16490913	-7.13	1.27	348.35	596.30	0.90	74.23	10.83
ZINC04385391	-7.11	1.46	326.35	641.39	3.14	95.09	12.28
ZINC05865540	-7.09	1.29	364.37	655.95	0.81	57.13	10.98
ZINC04181422	-7.09	1.81	365.39	685.23	1.17	70.81	10.43
ZINC05073884	-7.08	1.36	327.47	674.65	4.53	100.00	14.27
ZINC13810748	-7.06	1.25	389.41	685.01	1.31	71.90	10.77
ZINC57993543	-7.03	1.26	393.43	726.70	2.59	82.33	10.76
ZINC17055169	-7.02	1.38	320.35	609.94	3.95	95.56	12.67
ZINC09962923	-7.00	1.58	459.51	684.73	2.30	76.46	9.06
ZINC06473864	-7.00	1.82	486.33	736.49	2.58	86.74	7.94
ZINC17071195	-7.00	1.27	340.34	662.98	2.59	69.42	11.36

The final hits were also analyzed for Tpx interactions (Table 4.28). The survived 71 ligands interacted mostly with Ile 153, Glu 156, Ser 55, Phe 89 and Asp 57. 53 ligands out of 71 ligands made side chain hydrogen bonds with Glu 156. 35 hits made backbone hydrogen bonds with Ser 55. Furthermore, 21 ligands made $\pi - \pi$ stacking interaction with Phe 89.

Table 4.28. Interaction of 71 hits with the residues of Tpx.

ZINC ID	Ile 153	Glu 156	Ser 55	Phe 7	Phe 53	Phe 89	Thr 154	Asp 57	Ala 62
ZINC12314954	H (b)	H (s)							
ZINC26948418	H (b)	H (s)	H (b)						
ZINC00803235	H (b)	H (s)		π- π	π- π				
ZINC02692864	H (b)	H (s)							
ZINC05073475	H (b)	H (s)							
ZINC02001245		H (s)	H (b)			π- π	H (b)		
ZINC12928831		H (s)					H (b)		
ZINC01569737		H (s)	H (b) H (b)			π- π			
ZINC04279523	H (b)								
ZINC20309142	H (b)							H (b)	
ZINC01292370		H (s)	H (b)		π- π				
ZINC00625890	H (b)	H (s)						H (b)	H (b)
ZINC57992380		H (s)	H (b)						
ZINC13810826		H (s)	H (b)						
ZINC26780521	H (b)	H (s)							
ZINC20309181	H (b)							H (b)	
ZINC04102500	H (b)	H (s)						H (b)	H (b)
ZINC20309890	H (b)							H (b)	
ZINC01252080	H (b)	H (s)					H (b)		H (b)
ZINC02339978	H (b)	H (s)	H (b)	π- π					
ZINC20309903	H (b)							H (b)	
ZINC02227339	H (b)	H (s)	H (b)	π- π	π- π				
ZINC01718125	H (b)	H (s)							
ZINC16545157	H (b)	H (s)	H (b)			π- π			
ZINC01575542		H (s)	H (b)		π- π	π- π	H (b) H (b)		
ZINC01283334		H (s)							
ZINC36500739		H (s)	H (b)						
ZINC00633243			H (b)			π- π	H (b)	H (b)	
ZINC02337617	H (b)	H (s)	H (s)	π- π	π- π				
ZINC18249750		H (s)				π- π			H (b)
ZINC16525119		H (s)	H (b)						
ZINC20309809	H (b)							H (b)	
ZINC11592555	H (b) H (b)		H (b)						
ZINC02443796	H (b)	H (s)	H (b)	π- π	π- π				
ZINC00536176		H (s)	H (b)						
ZINC01700848		H (s)	H (b)						
ZINC60382125		H (s)	H (b)		π- π	π- π			
ZINC32836064		H (s)	H (b)				H (b)		
ZINC12385542	H (b)	H (s)				π- π	H (b)		
ZINC01247044		H (s)	H (b)			π- π			
ZINC19877995			H (b)						

Table 4.28. Interaction of 71 hits with the residues of Tpx (cont.).

ZINC ID	Ile 153	Glu 156	Ser 55	Phe 7	Phe 53	Phe 89	Thr 154	Asp 57	Ala 62
ZINC26949215		H (s)	H (b)			π- π			
ZINC00127686	H (b)	H (s)				π- π	H (b)		
ZINC00806563		H (s)				π- π			
ZINC06193011		H (s)						H (b)	
ZINC11971663		H (s)			π- π				
ZINC06746173		H (s)					H (b)		
ZINC04649681	H (b)	H (s)							
ZINC16925214	H (b)	H (s)	H (b)			π- π			
ZINC07977647		H (s)	H (b)			π- π			
ZINC39612159		H (s)			π- π, π- π	π- π		H (b) H (b)	
ZINC46587617			H (b)						H (b)
ZINC02522626	H (b)	H (s), H (s)					H (b)		H (b)
ZINC01312082	H (b)	H (s)							
ZINC36485022	H (b) H (b)		H (s)	π- π		π- π		H (b)	
ZINC04982797		H (s)	H (b)						H (b)
ZINC09310251						π- π		H (b)	
ZINC12561670			H (b)			π- π		H (b)	
ZINC02373214	H (b)	H (s)		π- π				H (b)	H (b)
ZINC08393981			H (b)						
ZINC16490913		H (s)	H (b)						
ZINC04385391						π- π		H (b)	H (b)
ZINC05865540		H (s)	H (b)						H (b)
ZINC04181422		H (s)				π- π		H (b)	
ZINC05073884		H (s)							
ZINC13810748			H (b)			π- π			H (b)
ZINC57993543			H (b)			π- π			H (b)
ZINC17055169		H (s)	H (b)		π- π				
ZINC09962923		H (s)	H (b)			π- π		H (b)	
ZINC06473864						π- π		H (b)	
ZINC17071195			H (b)			π- π		H (b)	H (b)

The 71 compounds were clustered based on Tanimoto Coefficient (TC) and eight clusters comprising compounds with more than 0.4 similarity were identified using single-linkage hierarchical clustering via Canvas module (Cluster 1: ZINC12314954, ZINC12385542, Cluster 2: ZINC00625890, ZINC13810826, ZINC04102500, ZINC00633243, ZINC02373214, ZINC13810748, Cluster 3: ZINC02692864, ZINC04279523, ZINC26780521, ZINC01718125, ZINC05073884, Cluster 4: ZINC36500739, ZINC36485022, Cluster 5: ZINC20309142, ZINC20309181,

ZINC20309890, ZINC20309903, ZINC20309809, Cluster 6: ZINC02339978, ZINC02227339, ZINC02337617, ZINC02443796, Cluster 7: ZINC18249750, ZINC00806563, and Cluster 8: ZINC01292370, ZINC01283334). BEI and SEI values of the compounds in all clusters are shown in Figure 4.16.

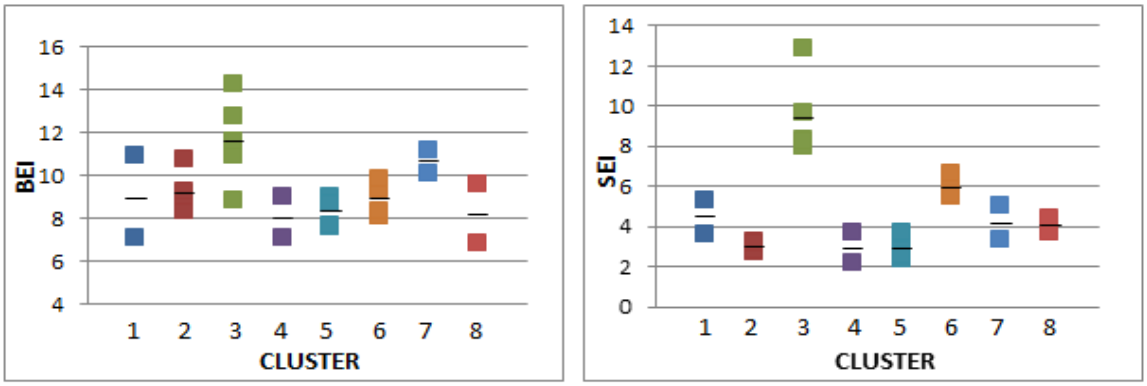


Figure 4.16. Binding efficiency (left) and surface efficiency (right) indices for the compounds in eight clusters. The average BEI and SEI values of the compounds in each cluster are displayed by strikethrough.

BEI and SEI values of the compounds in clusters 2, 3, 6 and 7 were similar and higher than the values of the compounds in other clusters. Thus, clusters 2, 3, 6 and 7 were picked out for substructure search. The consistency of representative ligands of each cluster with the AADDR hypothesis is displayed in Figure 4.17.

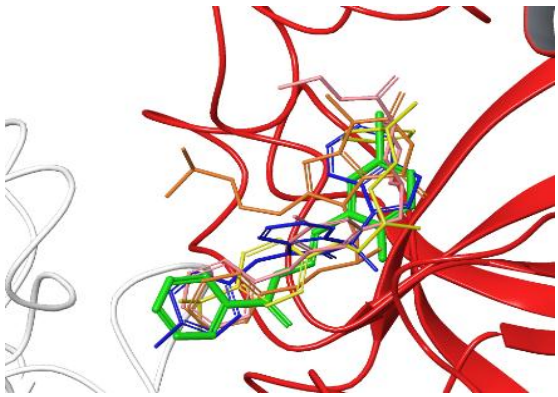


Figure 4.17. Tpx binding site with the ligand of AADDR hypothesis in green tube and the representative ligands of cluster 2 (orange), 3 (yellow), 6 (blue) and 7 (pink) in thin tubes. The fitness values of these ligands to the AADDR hypothesis 1.36, 1.44, 1.47 and 1.64, respectively.

Finally, the common scaffold of each cluster (2, 3, 6 and 7) and four compounds with high BEI values which did not fall into any cluster were used as query in substructure search of the ZINC database (Table 4.29).

Table 4.29. 2D structures of the ligands in clusters 2, 3, 6 and 7 and of the ligands having BEI values more than 14 and common scaffolds of the ligands in each cluster.

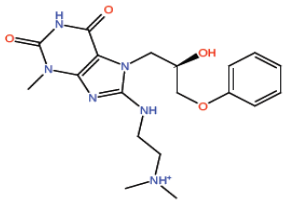
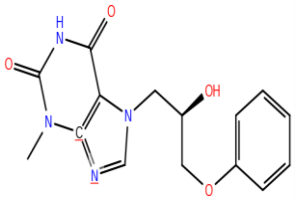
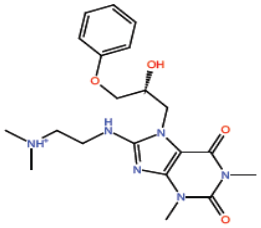
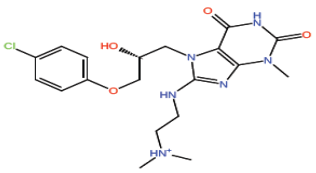
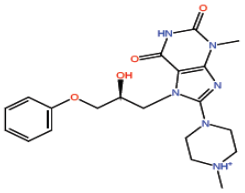
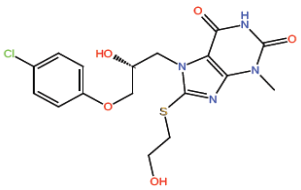
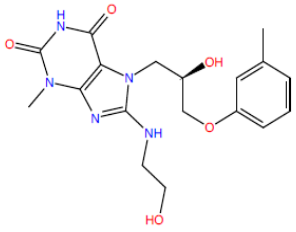
ZINC ID	Cluster	BEI	2D Structure	Common scaffold
ZINC00625890	2	8.76		
ZINC13810826		9.06		
ZINC04102500		8.36		
ZINC00633243		9.03		
ZINC02373214		9.25		
ZINC13810748		10.77		

Table 4.29. 2D structures of the ligands in clusters 2, 3, 6 and 7 and of the ligands having BEI values more than 14 and common scaffolds of the ligands in each cluster (cont.).

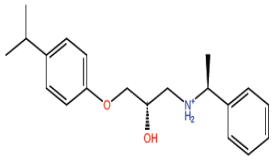
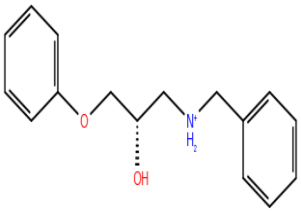
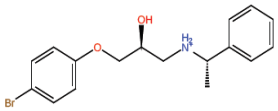
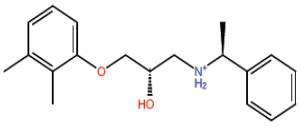
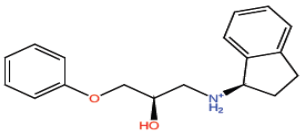
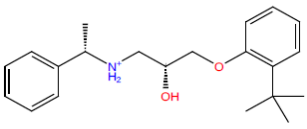
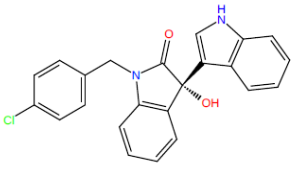
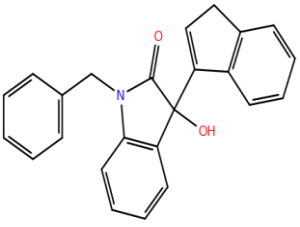
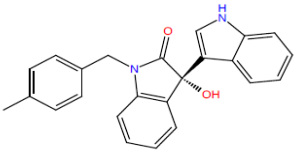
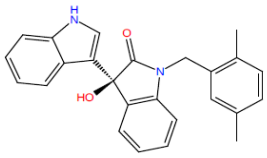
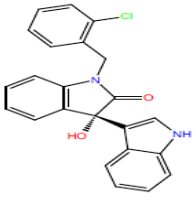
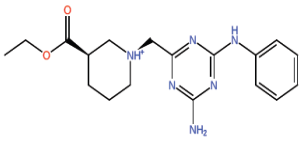
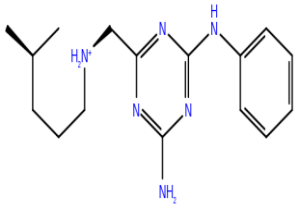
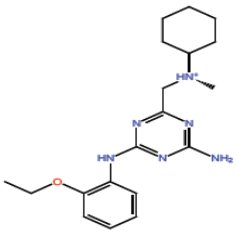
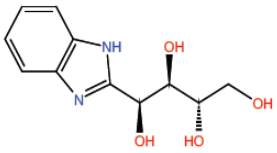
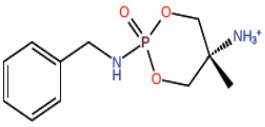
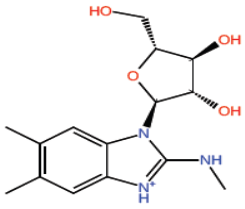
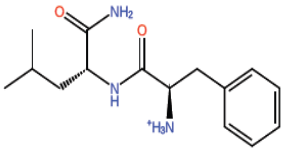
ZINC ID	Cluster	BEI	2D Structure	Common scaffold
ZINC02692864	3	10.93		
ZINC04279523		8.86		
ZINC26780521		12.75		
ZINC01718125		11.65		
ZINC05073884		14.27		
ZINC02339978	6	8.10		
ZINC02227339		9.38		
ZINC02337617		9.81		
ZINC02443796		8.38		

Table 4.29. 2D structures of the ligands in clusters 2, 3, 6 and 7 and of the ligands having BEI values more than 14 and common scaffolds of the ligands in each cluster (cont.).

ZINC ID	Cluster	BEI	2D Structure	Common scaffold
ZINC18249750	7	10.13		
ZINC00806563		11.22		
ZINC16525119	-	14.28		-
ZINC01700848		14.79		-
ZINC16925214		14.32		-
ZINC02522626		15.27		-

The compounds identified by substructure search with the scaffolds of cluster 3, 6 and 7 had high strain energy violations and poor docking scores (lower than -8.0 kcal/mol). The substructure search based on the high BEI compounds yielded no new compounds due to their large size. Only, using the chemical scaffold of cluster 2 yielded a novel set of

compounds that were missed in the ligand-based pharmacophore modeling and docking approach.

4.3.4. Substructure Search with the Scaffold of Cluster 2

After ZINC database was screened with the scaffold of cluster 2 (Table 4.29), which includes phenoxyethanol that is used in cosmetics, vaccines and pharmaceuticals as a preservative, 110 hits were found with GlideGscores between -8.7 and -3.1 kcal/mol and an average fitness value of 1.16. Out of 110 hits, 96 were different from the 1000 hits retrieved by molecular docking of ligands based on AADDR hypothesis. Since the average GlideGscore value of 71 hits retrieved from AADDR hypothesis was -7.7 kcal/mol, the GlideGscore cutoff value was selected as -8.0 kcal/mol and 5 ligands remained. Three of these five hits were not obtained in the hit set retrieved from the docking of AADDR hypothesis. Two of these ligands were eliminated due to high strain energy. The three ligands showed good ADME properties. The properties of the final three hits, which were different from the 1000 hits based on AADDR hypothesis, are given in Table 4.30.

Table 4.30. The properties of final 3 hits with their StrainGscores and efficiency indices.

ZINC ID	Strain Gscore (kcal/mol)	Mol_MW (Da)	SASA	QPlogPo/w	HOA %	BEI	BEI
		0.0 500.0	300.0 1000.0	-2.0 6.5	<25% low >80% high		
ZINC13810798	-8.68	388.43	684.70	0.88	57.78	6.78	1.75
ZINC04038300	-8.39	427.50	769.84	3.84	100.00	6.69	2.58
ZINC02408672	-8.22	387.44	725.36	3.18	95.69	7.70	2.67

The consistency of binding conformations of final hits with the representative ligand of cluster 2 is displayed in Figure 4.18.

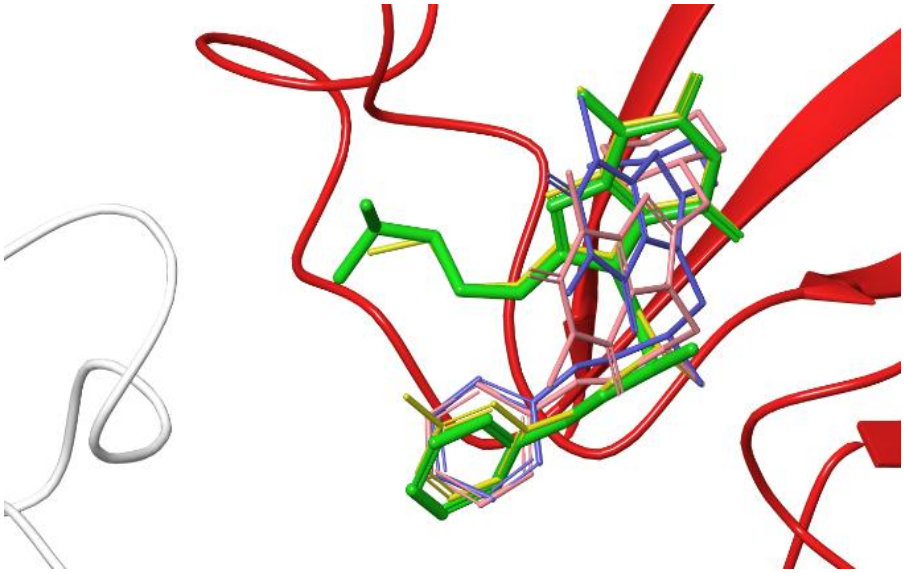


Figure 4.18. Three ligands attained from the common scaffold of cluster 2 of AADDR hypothesis in yellow, pink and light blue thin tubes, respectively with the representative ligand of cluster 2 (ZINC00625890) in green tube on Tpx structure.

These three ligands made backbone hydrogen bonds with Ile 153 and side chain hydrogen bonds with Glu 156.

The ligands from the substructure search of HRH hypothesis, ZINC55099370 (Table 4.9), and ZINC12298678 (Table 4.14), which had 1.35 and 1.30 fitness to HRH hypothesis and the ligand from the substructure search of AADDR hypothesis, ZINC13810798 (Table 4.30), which had 1.23 fitness to AADDR hypothesis, were obtained as novel compounds. Thus, the substructure search based on the scaffolds identified after docking revealed new compounds which did not fall into the first sets of 10,000 compounds retrieved from database searches.

4.4. Proposed Compounds and Their Interaction Analyses

60 hits with average molecular weight (MW) of 372 Da were obtained via the structure-based pharmacophore modeling route while 74 hits with average MW of 366 Da were obtained via the ligand-based pharmacophore modeling path. 60 hits had an average FOSA (hydrophobic component of SASA) of 359.22 and octanol-water partition coefficient (QPlogPo/w**)** of 3.23 whereas 74 hits had an average FOSA of 216.48 and

QPlogPo/w of 2.34. Despite having similar average sizes, the molecules from the structure-based route are more hydrophobic as indicated by their average FOSA and octanol-water partition coefficient values. This is an expected result because the structure-based pharmacophore model includes two hydrophobic sites and one ring whereas the ligand-based pharmacophore model includes two acceptors, two donors and one ring.

The pool of ligands retrieved from ligand-based (74 hits) and structure-based (60 hits) pharmacophore models were further analyzed to find most drug-like compounds based on docking scores, fitness values to the hypothesis, molecular weight, percentage of human oral absorption (HOA) and BEI values. Molecular weight term did not appear explicitly in final selection score equation because BEI also includes molecular weight term in the denominator. The selection of final hits was optimized using the following simple scoring scheme based on the contribution of each factor equally:

$$\text{Final selection score} = 0.25\text{GlideGscore} + 0.25\text{Fitness} + 0.25\text{HOA} + 0.25\text{BEI} \quad (4.6)$$

Via these optimization criteria, 14 hits from structure-based and 17 hits from ligand-based pharmacophore models were picked out (Table 4.31).

Table 4.31. Properties of final hits obtained from two pharmacophore models and substructure searches of clusters of related pharmacophore models.

Compound	ZINC ID	GlideGscore (kcal/mol)	Fitness	Mol_MW (Da)	% HOA	BEI
HRH[a]						
1	ZINC12378903	-9.04	2.05	432.44	100.00	7.54
2	ZINC00186392	-8.35	1.91	318.18	77.22	9.06
3	ZINC16383080	-7.97	1.92	300.74	94.91	10.47
4	ZINC02168326	-7.67	1.99	305.50	100.00	16.05
5	ZINC04013595	-7.61	1.94	208.32	91.69	18.35
6	ZINC16638191	-7.51	2.06	326.44	100.00	12.80
7	ZINC01649089	-7.37	1.91	253.40	96.69	16.72
8	ZINC00099607	-7.36	2.12	285.34	100.00	15.30
9	ZINC12371364	-7.26	2.00	305.39	100.00	13.68
10	ZINC19169013	-7.19	2.01	279.34	71.49	15.25
11	ZINC09196338	-7.10	1.91	278.40	88.89	16.43
HRH1[b]						
12	ZINC12373624	-8.19	1.53	442.62	100.00	6.76
HRH3[c]						
13	ZINC12298678	-9.04	1.30	408.49	91.34	5.79
14	ZINC08497999	-8.88	1.42	319.38	100.00	6.02
AADDR[d]						
15	ZINC12314954	-8.96	1.59	343.40	89.42	7.09
16	ZINC00803235	-8.82	1.47	363.46	100.00	7.61
17	ZINC02692864	-8.73	1.44	313.44	100.00	10.93
18	ZINC05073475	-8.68	1.75	343.42	100.00	8.42
19	ZINC02001245	-8.63	1.32	230.27	69.72	11.58
20	ZINC01569737	-8.47	1.33	261.32	56.44	10.66
21	ZINC26780521	-8.19	1.48	299.41	100.00	12.75
22	ZINC01718125	-7.78	1.25	283.37	100.00	11.65
23	ZINC16545157	-7.75	1.68	266.25	53.64	12.45
24	ZINC16525119	-7.64	1.52	238.24	59.99	14.28
25	ZINC01700848	-7.55	1.31	256.24	70.65	14.79
26	ZINC00127686	-7.40	1.79	307.36	69.85	11.60
27	ZINC06746173	-7.30	1.66	314.38	85.30	13.10
28	ZINC16925214	-7.26	1.28	307.35	73.30	14.32
29	ZINC02522626	-7.23	1.42	277.37	52.32	15.27
30	ZINC04385391	-7.11	1.46	326.35	95.09	12.28
AADDR2[e]						
31	ZINC02408672	-8.22	1.23	387.44	95.69	7.70

[a] The hits retrieved by structure-based pharmacophore modeling (HRH).
[b] by substructure search of cluster 1 of HRH.
[c] by substructure search of cluster 3 of HRH.
[d] by ligand-based pharmacophore modeling (AADDR)
[e] by substructure search of cluster 2 of AADDR.

These proposed compounds were analyzed for their interactions with the Tpx binding site residues. The summary of the interactions of these final 31 hits is displayed in Table 4.32.

Table 4.32. Interactions of final 31 ligands with Tpx.

HRH						
	Ile 153	Glu 156	Ser 55	Phe 89	Thr 154	Asp 57
1	H (b)	H (s), H (s)	H (b)	π- π		
2	H (b)	H (s), H (s)	H (b)			
3		H (s)	H (b)			H (b)
4		H (s)				
5	H (b)	H (s), H (s)				
6		H (s)				
7		H (s)	H (b)			
8		H (s)		π- π		
9		H (s)				H (b)
10		H (s)	H (b)			
11		H (s)	H (b)			
HRH1						
12	H (b)	H (s)			H (b)	
HRH3						
13		H (s), H (s)				
14		H (s)				
AADDR						
	Ile 153	Glu 156	Ser 55	Phe 89	Thr 154	Ala 62
15	H (b)	H (s)				
16	H (b)	H (s)		π- π		
17	H (b)	H (s)				
18	H (b)	H (s)				
19		H (s)	H (b)			
20		H (s)	H (b), H (b)			
21	H (b)	H (s)				
22	H (b)	H (s)				H (b)
23	H (b)	H (s)	H (b)	π- π		
24		H (s)	H (b)			
25		H (s)	H (b)			
26	H (b)	H (s)			H (b)	
27		H (s)			H (b)	
28	H (b)	H (s)	H (b)			
29	H (b)	H (s), H (s)			H (b)	
30				π- π	H (b)	
AADDR2						
31	H (b)	H (s)		π- π		

Except compound 30, all compounds make hydrogen bonding interaction with side chain of Glu 156. Because the AADDR hypothesis and Ile 153 were placed nearby (Figure 6), more compounds retrieved via ligand-based pharmacophore modeling made hydrogen bonds with Ile 153 backbone than those obtained via structure-based pharmacophore route.

Eventually, eight ligands with high docking scores, good fitness to hypotheses and high BEI values, 1, 4, 12, 14, 15, 16, 17 and 26 were picked out. Binding modes of the proposed ligands and the binding site residues that interact with the ligand are shown in Figure 4.19.

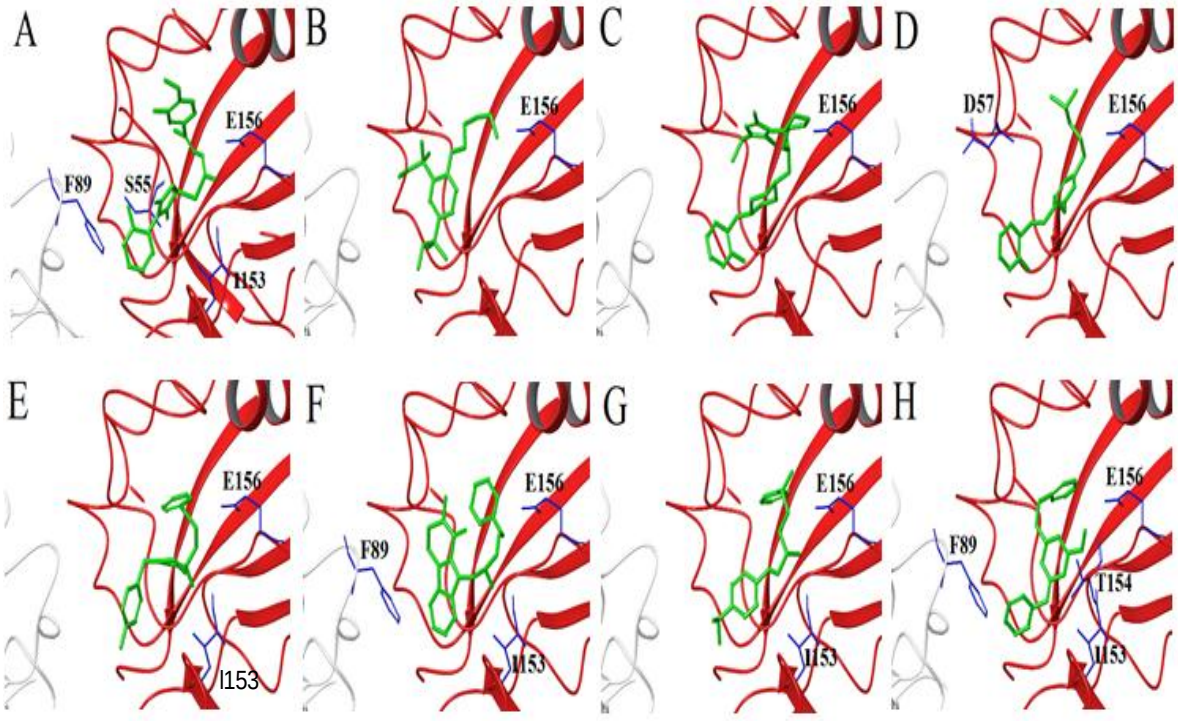


Figure 4.19. Binding modes of the ligands (A) 1, (B) 4, (C) 12, (D) 14, (E) 15, (F) 16, (G) 17 and (H) 26 in the Tpx binding site and the residues that interact with the ligand.

The consistency of these ligands with the hypothesis sites was checked by superimposition of HRH and AADDR hypotheses to the ligands (Table 4.33). The distances between the hypothesis sites and the ligand atoms which represent those hypothesis sites are displayed. These eight ligands were further assessed by induced fit docking and selectivity analysis.

Table 4.33. Superimposition of HRH and AADDR hypotheses to ligands with distances of respective ligand atoms to the site points which are displayed by an arrow.

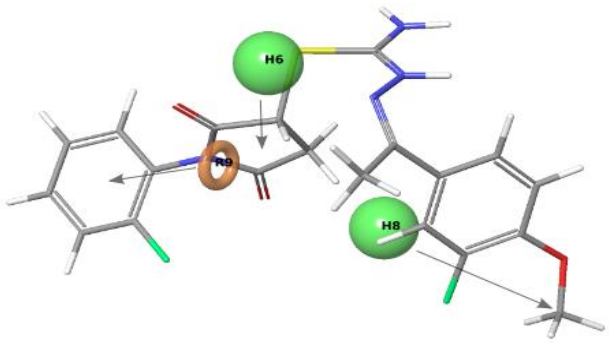
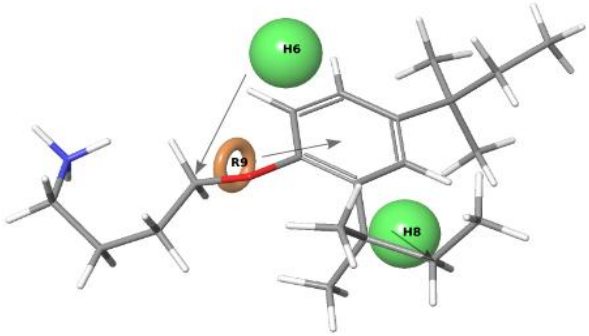
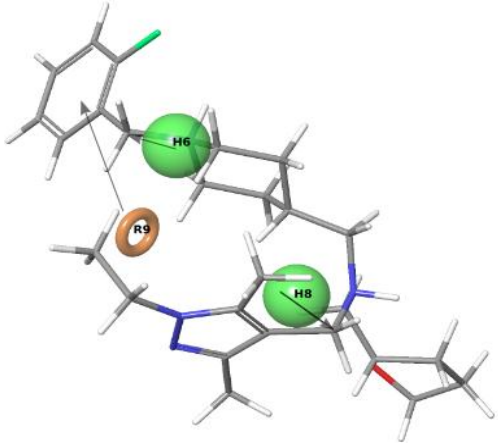
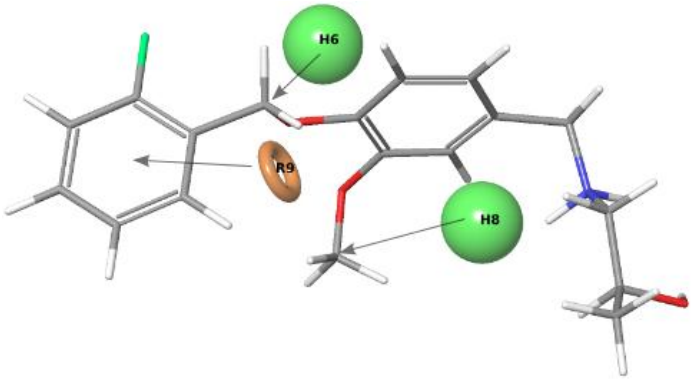
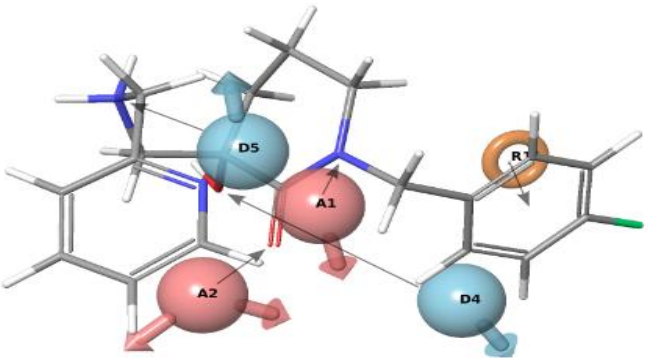
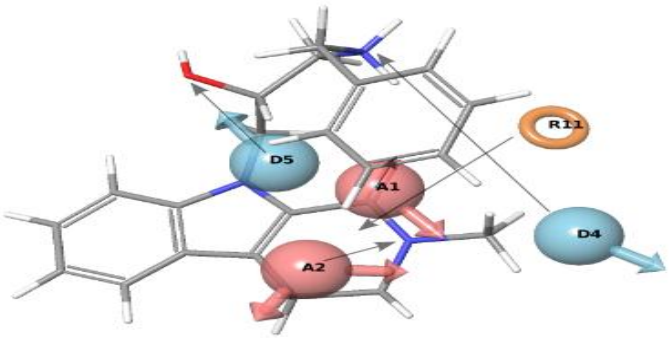
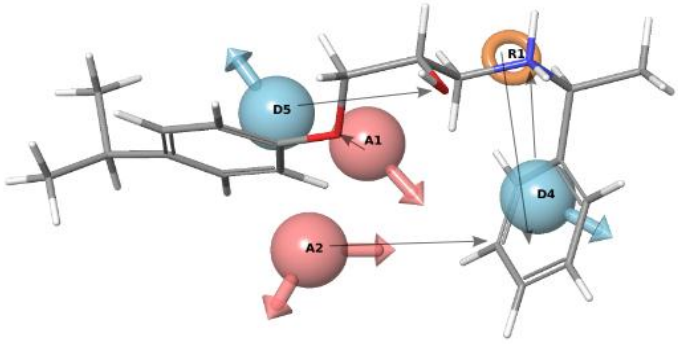
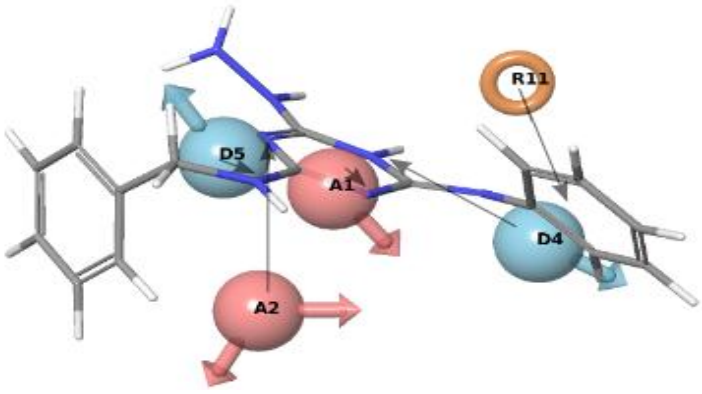
Ligand		Site	Distance (Å)
1		H6 R9 H8	1.81 2.63 3.06
4		H6 R9 H8	2.87 2.18 2.07
12		H6 R9 H8	5.88 1.92 2.08
14		H6 R9 H8	3.23 1.58 4.77

Table 4.33. Superimposition of HRH and AADDR hypotheses to ligands with distances of respective ligand atoms to the site points which are displayed by an arrow (cont.).

Ligand		Site	Distance (Å)
15		A1 A2 D4 D5 R11	1.16 1.69 4.41 3.14 1.44
16		A1 A2 D4 D5 R11	1.00 2.60 6.33 2.93 2.66
17		A1 A2 D4 D5 R11	1.64 4.35 2.92 4.52 3.16
26		A1 A2 D4 D5 R11	0.75 3.47 2.86 1.71 2.51

4.5. Induced Fit Docking

The binding conformations of the proposed eight ligands were assessed by allowing side chain flexibility in addition to the ligand flexibility during docking. The highest IFD scored poses for each ligand as well as rigid docking scores and their interactions with Tpx are displayed in Table 4.34.

Table 4.34. Rigid and IFD docking results and interaction analysis of final eight ligands.

Compound		Score	Ile 153	Glu 156	Ser 55	Phe 7	Phe 53	Phe 89	Pro 54	Thr 154
1	Rigid docking	-9.04	H (b)[a]	H (s)[b]	H (b)			π-π[c]		
	IFD	-7.58	H (b)	H (s)	H (b)			π-π		
4	Rigid docking	-7.67		H (s)						
	IFD	-8.52		H (s)		π-π				H (b)
12	Rigid docking	-8.19		H (s)						
	IFD	-10.42		H (s)				π-π		
14	Rigid docking	-8.88	H (b)	H (s)						H (b)
	IFD	-10.07		H (s)		ππ			π-π	H (b)
15	Rigid docking	-8.96	H (b)	H (s)						
	IFD	-10.70	H (b)	H (s)	H (b)	ππ		π-π	π-π	
16	Rigid docking	-8.82	H (b)	H (s)				π-π		
	IFD	-10.39	H (b)	H (s)		π-π			π-π	
17	Rigid docking	-8.73	H (b)	H (s)						
	IFD	-10.83	H (b)	H (s)		π-π				
26	Rigid docking	-7.40	H (b)	H (s)				π-π		H (b)
	IFD	-10.56	H (b), H (b)	H (s)	H (b)	π-π	π-π			

[a] backbone hydrogen bonding interaction.
[b] side chain hydrogen bonding interaction.
[c] pi-pi stacking interaction.

For the compounds 4, 12 and 26, IFD score was high despite having relatively low rigid docking scores. Thus, the selection of proposed compounds based on not only good docking and fitness scores but also low molecular weight, high BEI and % HOA values allowed the identification of additional compounds. For example, the ligand 26 binding score increased considerably from -7.40 kcal/mol to -10.56 kcal/mol upon IFD. Its binding modes in the presence and absence of protein flexibility and its interaction maps with the binding site residues are displayed in Figure 4.20.

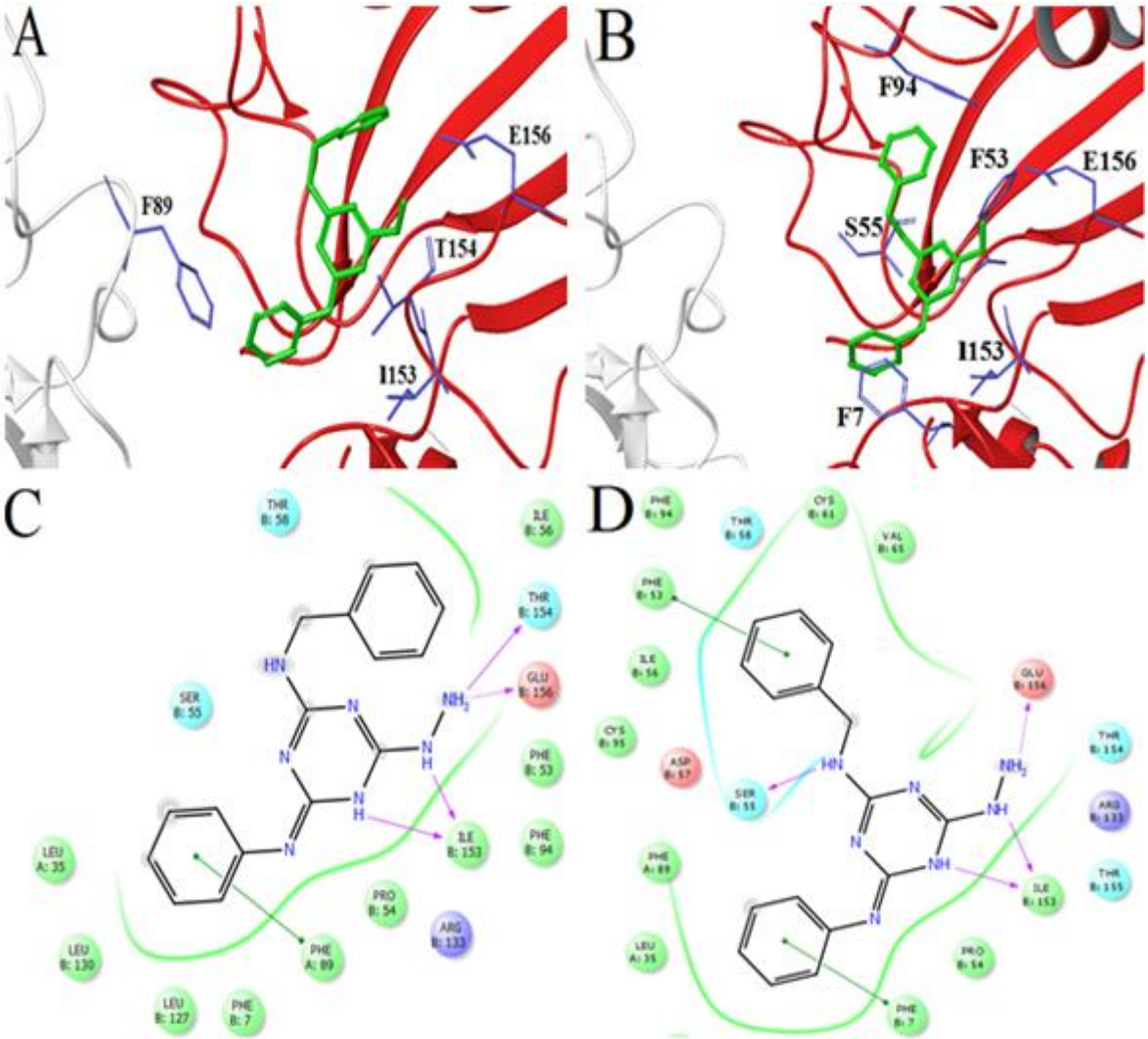


Figure 4.20. The binding mode of ligand 26 (green) in the rigid docked form (A) and in the IFD mode (B) and the residues which interacts with the ligand. (C) and (D) are the interaction maps of ligand 26 in the binding site. Purple and dashed lines represent backbone and side-chain hydrogen bonding interaction, respectively.

Superimposition results of the conformations of rigid docked form and highest docking scored IFD poses for each ligand are displayed in Figure 4.21.

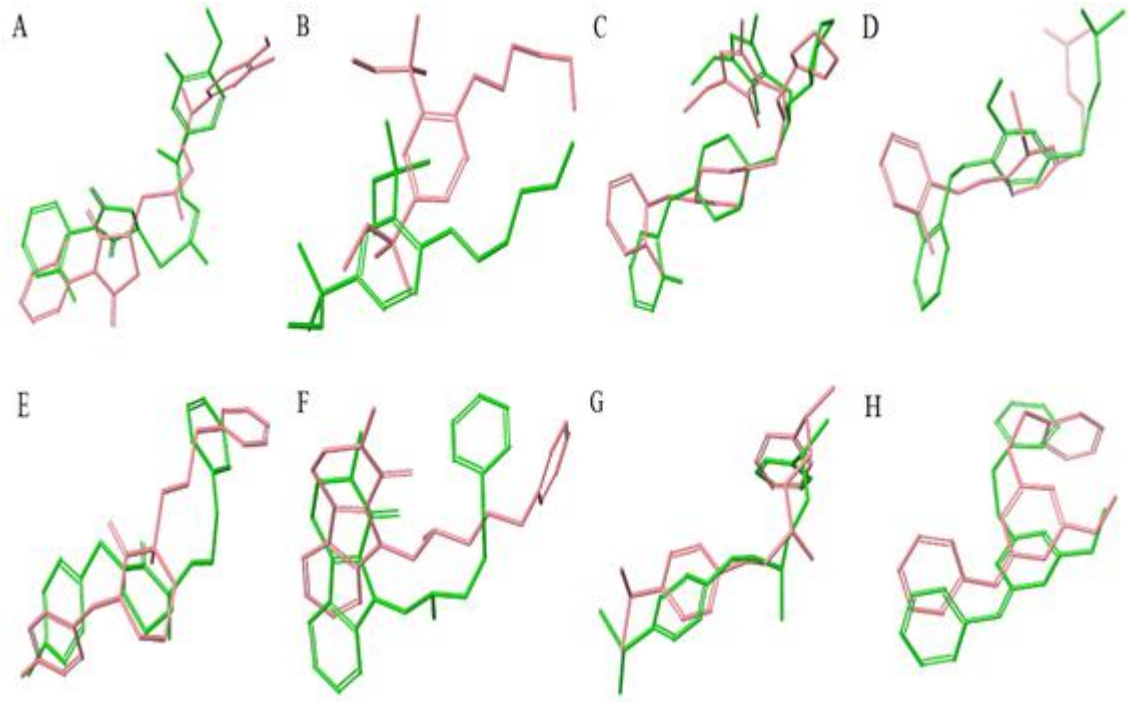


Figure 4.21. Rigid (pink) and IFD (green) binding conformations of ligands (A) 1, (B) 4, (C) 12, (D) 14, (E) 15, (F) 16, (G) 17 and (H) 26 in the binding site of Tpx.

The proposed compounds made interactions mostly with Ile 153, Ser 55, Glu 156 and Phe 89 residues similar to the 27 experimentally identified ligands but the docking scores were higher than those of the experimental ligands suggesting the proposed ligands bound to Tpx more favorably.

4.6. Selectivity Analysis

Basic Local Alignment Search Tool (BLAST) [22] was utilized to identify sequentially similar proteins to Tpx. An atypical 2-cys peroxiredoxin from *Staphylococcus aureus* (PDB code: 3P7X) was found out to be most similar to Tpx with a query cover percentage of 84% and maximum identity percentage of 49%. The other similar proteins were peroxiredoxin-like protein from *Aquifex aeolicus* (PDB code: 2YZH), bacterioferritin comigratory protein homolog (Bcp4) from *Sulfolobus solfataricus* (PDB code: 3HJP) and a

dimeric Peroxiredoxin Q from *Sulfolobus tokodaii* (4G2E). Sequence alignment results of these proteins are shown in Table 4.35.

Table 4.35. Sequnce alignment results attained by NCBI-BLAST [70].

Pdb Id	Name	Query Cover	E-value	Maximum Identity
3P7X	An atypical 2-cys peroxiredoxin from *Staphylococcus aureus*	84%	3e-43	49%
2YZH	Peroxiredoxin-like protein from *Aquifex aeolicus*	83%	3e-43	41%
3HJP	Bacterioferritin comigratory protein homolog (Bcp4) from *Sulfolobus solfataricus*	73%	6e-17	36%
4G2E	A dimeric PrxQ from *Sulfolobus tokodaii*	76%	2e-16	34%

To verify that the proposed ligands are selective for Tpx, the eight ligands were docked to these proteins in XP mode without any constraint. The grid was generated such that it included residues corresponding to Phe 7, Phe 53, Ser 55, Asp 57, Phe 89, Ile 153, Thr 154 and Glu 156. Docking results and interactions of eight compounds with each protein are displayed in Table 4.36.

Table 4.36. Similarity analysis of four proteins to Tpx.

Ligand		Protein	Glide Gscore	Ile 153	Glu 156	Ser 55	Phe 7	Phe 89	Asp 57	Thr 154
1	ZINC12378903 -9.04 kcal/mol Fitness: 2.05	3P7X	-4.58							
		2YZH	-6.38							
		3HJP	-3.76							
		4G2E	-2.27		Glu 141[a],s					Thr 139,b
4	ZINC02168326 -7.67 kcal/mol Fitness: 1.99	3P7X	-3.51							
		2YZH	-5.31							
		3HJP	-5.10							
		4G2E	-3.10							
12	ZINC12373624 -8.19 kcal/mol Fitness: 1.53	3P7X	-5.58							
		2YZH	-5.06							
		3HJP	-3.70							
		4G2E	-1.57							
14	ZINC08497999 -8.88 kcal/mol Fitness: 1.42	3P7X	-4.58							
		2YZH	-5.85							
		3HJP	-4.39							
		4G2E	-4.40							
15	ZINC12314954 -8.96 kcal/mol Fitness: 1.59	3P7X	-5.22							
		2YZH	-5.41							
		3HJP	-3.84							
		4G2E	-2.83		Glu 141,s					Thr 139,b
16	ZINC00803235 -8.82 kcal/mol Fitness: 1.47	3P7X	-4.80							
		2YZH	-6.44							
		3HJP	-4.94					Phe 75, π-π		
		4G2E	-4.90							
17	ZINC02692864 -8.73 kcal/mol Fitness: 1.44	3P7X	-5.23							
		2YZH	-6.69							
		3HJP	-3.30							
		4G2E	-5.33							
26	ZINC00127686 -7.40 kcal/mol Fitness: 1.79	3P7X	-4.85				Phe 6, π-π	Phe 75, π-π		
		2YZH	-5.29							
		3HJP	-3.92							
		4G2E	-3.77							

[a] Correspondence of residues of Tpx in the sequence of the related protein.

GlideGscore and interaction results indicate that the proposed ligands are selective for Tpx since they bound poorly to the proteins (Glide Gscores below than -6.7 kcal/mol) which are similar to Tpx. Furthermore, the binding site residues of the human ortholog peroxiredoxin 3 (29% sequence identity in 71% query cover) and the Tpx paralog, the uridine phosphorylase (PDB code: 4I2V), (29% sequence identity in 43% query cover), are different from those of Tpx suggesting that the identified ligands will not bind to these proteins, either.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

In this thesis, an integrated study of pharmacophore modeling, molecular docking followed by substructure search of ZINC database with the common scaffolds obtained after docking was successfully applied to identify potential inhibitors of Tpx. Two kinds of pharmacophore models, ligand-based, AADDR, based on previously identified Tpx inhibitors and structure-based, HRH, based on the information of the active binding site were built to screen the ZINC database. In order to select and validate the best hypothesis in ligand-based pharmacophore modeling, quantitative structure-activity relationship (QSAR) model was developed by generating training and test sets out of 27 known Tpx inhibitors, and AADDR hypothesis was statistically significant (0.961 R^2, 0.102 RMSE and 0.946 Pearson-R values) and predictive ability on all ligands when it was associated with the QSAR model. After database filtering with HRH and AADDR hypotheses was carried out, molecular docking was applied and docked compounds which had better GlideGscore than -7.0 kcal/mol (55 ligands for HRH, and 123 compounds for AADDR) were subjected to post docking evaluations using calculations of efficiency indices, strain energy corrections, and ADME and druglikeness properties and a set of compounds were identified as putative inhibitors. In addition, substructure search of common scaffolds of clusters including n-benzylformamide (a key component in hair dyes [71]), 1,2-dimethoxybenzene (a sedative and commonly known as veratrol [72]) and phenoxyethanol (a preservative in cosmetics, vaccines and pharmaceuticals [73]) was used to select additional 31 compounds out based on high docking score, favorable fitness value to pharmacophore model, low molecular weight, high BEI and % HOA values. Induced fit docking was performed to find favorable binding conformations for selected eight ligands with high docking score, fitness and BEI value. In induced fit docking, the side chains of the active site residues are allowed to move and the most significant movements were observed for Ile 153, Glu 156 and Phe 53 residues. The eight ligands from different chemotypes were proposed as potential inhibitors for Tpx. Out of these eight ligands, compound 4 (ZINC02168326) is the trypanothione (parasite analog of glutathione) reductase [74], compound 14 (ZINC08497999) is the inhibitor of human tyrosyl-DNA

phosphodiesterase (TDP1) [75], related with cancer treatment, compound 15 (ZINC12314954) is the HIV-1 integrase inhibitor [76], compound 16 (ZINC00803235) is the inhibitor of Hepatisis C Virus (HCV) [77] and Transforming Growth Factor β (TGF-b), which plays a role in immunity and cancer [78], and compound 17 (ZINC02692864) is inhibitor of the malarial parasite plastid [79]. These promising hits can be a starting point for the design of novel ligands against attacks of pathogenic organisms on cells.

5.2. Recommendations for Future Studies

In both ligand-based and structure-based screening, additional pharmacophore models can be created by the different features of the ligands and receptor structure to obtain diverse set of inhibitors that may also display good binding affinities. Moreover, the inhibitory performance of the proposed molecules can be tested in vitro by preparing biological assays of the target protein.

Molecular Dynamics (MD) simulations can also be carried out to improve analysis of hits. MD simulations can be performed for both apo state and ligand-bound states of Tpx in order to obtain information about the stability and dynamic properties of the protein-ligand complex.

APPENDIX A: 2D STRUCTURES OF KNOWN TPX INHIBITORS

Table A.1. 2D structures of active inhibitors [5, 27].

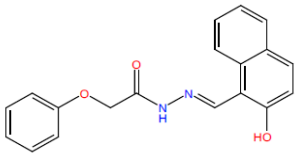
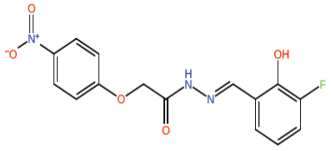
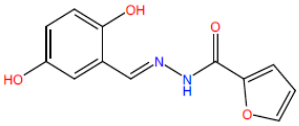
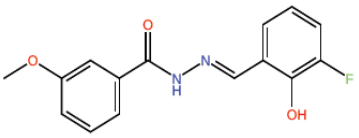
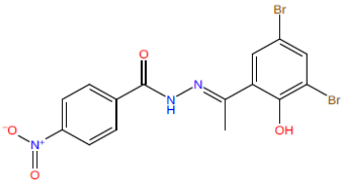
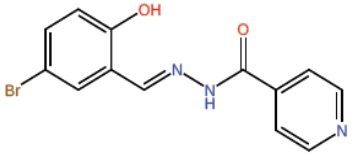
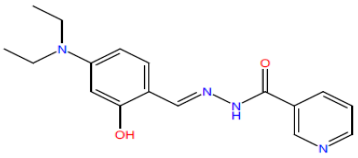
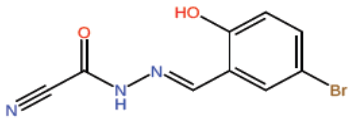
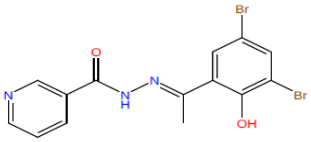
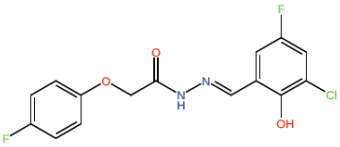
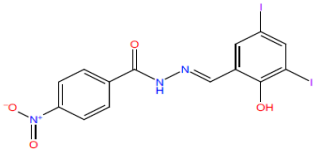
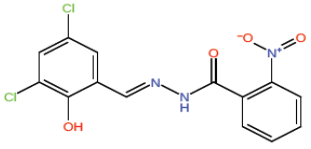
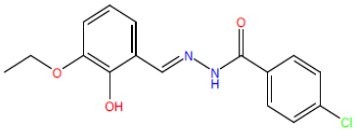
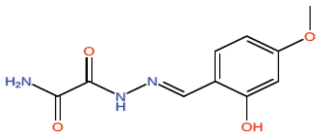
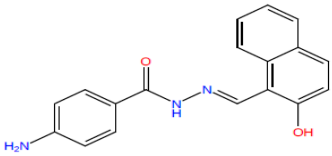
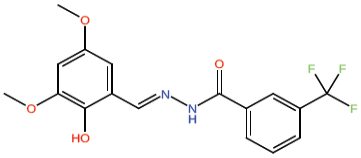
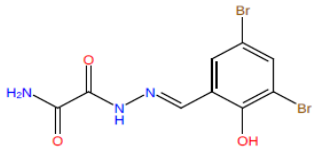
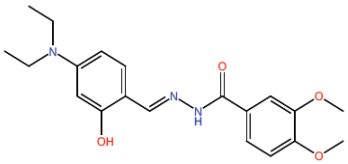
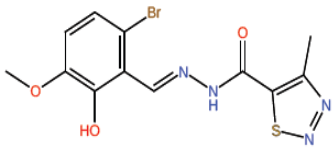
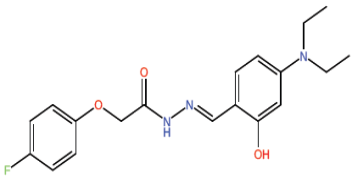
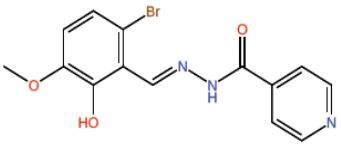
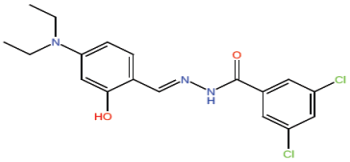
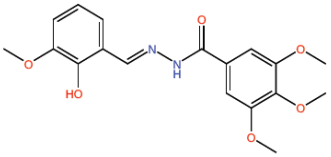
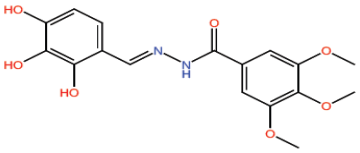
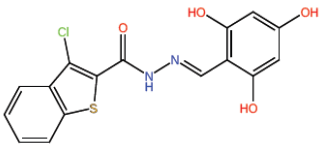
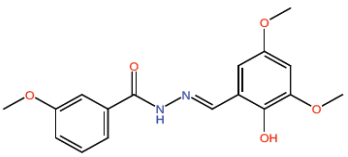
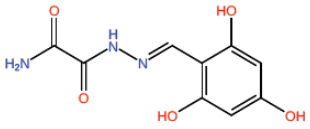
ID	Structure	ID	Structure
1		15	
2		16	
3		17	
4		18	
5		19	
6		20	
7		21	
8		22	

Table A.1. 2D structures of active inhibitors [5, 27] (cont.).

ID	Structure	ID	Structure
9		23	
10		24	
11		25	
12		26	
13		27	
14			

APPENDIX B: PERCENT INHIBITION VERSUS CONCENTRATION DATA

Table B.1. Percent inhibition at different ligand concentrations [5] and related statistical correlation.

	% Inhibition (-y axis) at different ligand concentrations (μM) (-x axis)				
Ligand	10	20	50	100	Statistical correlation
1	30	41	75	89	$y = -0.0093x^2 + 1.6977x + 12.686$
2	22	55	94	102	$y = -0.0184x^2 + 2.8644x - 0.8536$
3	42	47	59	86	$y = 0.0013x^2 + 0.3449x + 38.895$
4	36	45	64	86	$y = -0.0029x^2 + 0.8699x + 28.062$
5	34	61	86	87	$y = -0.0144x^2 + 2.1238x + 18.236$
6	39	49	76	90	$y = -0.0071x^2 + 1.3593x + 25.541$
7	23	46	74	85	$y = -0.0118x^2 + 1.9579x + 7.3959$
8	43	59	73	82	$y = -0.0065x^2 + 1.1099x + 35.281$
9	39	68	77	86	$y = -0.0089x^2 + 1.4119x + 32.785$

Table B.2. Percent inhibition at different ligand concentrations [27] and related statistical correlation.

	% Inhibition (-y axis) at different ligand concentrations (μM) (-x axis)								
Ligand	1.56	3.13	6.25	12.5	25	50	100	200	Statistical correlation
10	-5	14	19	21	34	60	71	-	$y = -0.009x^2 + 1.5817x + 2.8884$
11	-15	9	18	19	27	54	77	-	$y = -0.0069x^2 + 1.4729x - 1.6075$
12	-18	14	17	15	20	43	65	68	$y = -0.0032x^2 + 0.9812x + 0.6681$
13	2	10	21	33	44	70	79	87	$y = -0.0042x^2 + 1.2011x + 11.155$
14	-1	5	17	18	26	51	63	64	$y = 0.00002x^3 - 0.0104x^2 + 1.4178x + 1.236$
15	-24	-5	6	4	25	84	95	97	$y = -0.0068x^2 + 1.9028x - 14.291$
16	-2	7	9	11	17	45	58	74	$y = -0.0024x^2 + 0.8493x + 1.3767$
17	-8	14	13	24	43	87	95	98	$y = -0.0059x^2 + 1.6401x + 3.3405$
18	0	11	6	22	55	86	97	99	$y = 0.00007x^3 - 0.0257x^2 + 2.9092x - 5.19$
19	-6	2	-1	10	15	63	90	94	$y = -0.0049x^2 + 1.4948x - 8.2726$
20	-4	16	8	4	18	61	88	92	$y = -0.0043x^2 + 1.331x - 2.3195$
21	-15	-6	4	8	13	73	94	97	$y = -0.0058x^2 + 1.7157x - 12.952$
22	-10	10	21	24	39	66	74	75	$y = 6E\text{-}05x^3 - 0.0204x^2 + 2.235x - 1.6255$
23	-14	14	14	29	44	66	66	56	$y = 0.0002x^3 - 0.0484x^2 + 3.2787x - 6.8808$
24	-12	11	17	23	35	54	58	50	$y = -0.0113x^2 + 1.7116x - 0.3369$
25	-	19	22	29	34	43	46	46	$y = 2E\text{-}05x^3 - 0.0082x^2 + 0.8767x + 17.233$
26	-10	11	14	19	31	50	58	62	$y = 4E\text{-}05x^3 - 0.0156x^2 + 1.7223x - 1.3533$
27	-7	4	9	13	20	43	63	76	$y = -0.0028x^2 + 0.9502x - 0.7311$

REFERENCES

1. The infectious disease Yersiniosis, "Centers for Disease Control and Prevention (CDC)", http://www.cdc.gov/ncdod/bmd/diseaseinfo/yersinia_g.htm, 2005, [Accessed February 2014].

2. Nikkari, S., R. Merilahti-Palo, R. Saario, K. O. Söderström, K. Granfors, M. Skurnik, and P. Toivanen, "Yersinia-Triggered Reactive Arthritis. Use of Polymerase Chain Reaction and Immunocytochemical Staining in the Detection of Bacterial Components from Synovial Specimens", *Arthritis and Rheumatism*, Vol. 35, No. 6, pp. 682-687, 1992.

3. Malekzadeh, F., C. Alberti, M. Nouraei, H. Vahedi, I. Zaccaria, U. Meinzer, S. Nasseri-Moghaddam, R. Sotoudehmanesh, S. Momenzadeh, R. Khaleghnejad, S. Rashtak, G. Olfati, R. Malekzadeh, and J. P. Hugot, "Crohn`s Disease and Early Exposure to Domestic Refrigeration", *PLoS ONE*, Vol. 4, No. 1, pp. 1-7, 2009.

4. Edgren, T., A. Forsberg, R. Rosqvist, and H. Wolf-Watz, "Type III Secretion in *Yersinia*: Injectisome or Not?", *PLoS Pathogens*, Vol. 8, No. 5, pp. 1-3, 2012.

5. Nordfelth, R., A. M. Kauppi, H. A. Norberg, H. Wolf-Watz, and M. Elofsson, "Small-Molecule Inhibitors Specifically Targeting Type III Secretion", *Infection and Immunity*, Vol. 73, No. 5, pp. 3104-3114, 2005.

6. Wang, D., C. E. Zetterstrom, M. Gabrielsen, K. S. H. Beckham, J. J. Tree, S. E. Macdonald, O. Byron, T. J. Mitchell, D. L. Gally, P. Herzyk, A. Mahajan, H. Uvell, R. Burchmore, B. O. Smith, M. Elofsson, and A. J. Roe, "Identification of Bacterial Target Proteins for the Salicylidene Acylhydrazide Class of Virulence-Blocking Compounds", *Journal of Biological Chemistry*, Vol. 286, No. 34, pp. 29922-29931, 2011.

7. Flohe, L., S. Toppo, G. Cozza, and F. Ursini, "A Comparison of Thiol Peroxidase Mechanisms", *Antioxidants and Redox Signaling*, Vol. 15, No. 3, pp. 763-780, 2011.

8. Gabrielsen, M., C. E. Zetterstrom, D. Wang, K. S. Beckham, and M. Elofsson, "Expression, Purification, Crystallization, and Initial X-Ray Diffraction Analysis of Thiol Peroxidase from Yersinia Pseudotuberculosis", *Acta Crystallographica Section F-Structural Biology and Crystallization Communications*, Vol. 66, pp. 1606-1609, 2010.

9. Cornelis, G.R., "Yersinia Type III Secretion: Send in the Effectors", *Journal of Cell Biology*, Vol. 158, pp. 401-408, 2002.

10. He, S. Y., K. Nomura, and T. S. Whittam, "Type III Protein Secretion Mechanism in Mammalian and Plant Pathogens", *Biochimica et Biophysica Acta*, Vol. 1694, pp. 181-206, 2004.

11. Hueck, C. J., "Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants", *Microbiology and Molecular Biology Reviews*, Vol. 62, No. 2, pp. 379-433, 1998.

12. Wang, Y., L. Zhang, W. L. Picking, W. D. Picking, and R. N. De Guzman, "Structural Dissection of the Extracellular Moieties of the Type III Secretion Apparatus", *Molecular Biosystems*, Vol. 4, No. 12, pp. 1176-1180, 2008.

13. Galan, J. E. and A. Collmer, "Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells", *Science*, Vol. 284, pp. 1322-1328, 1999.

14. Thiol Peroxidase (Tpx), "National Center for Biotechnology Information", http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?hslf=1&uid=cd03014&#seqhr, 2003, [Accessed August 2013].

15. Baker, L. M. S., and L. B. Poole, “Catalytic Mechanism of Thiol Peroxidase from Escherichia Coli: Sulfenic Acid Formation and Overoxidation of Essential Cys61”, *Journal of Biological Chemistry*, Vol. 278, pp. 9203-9211, 2003.

16. Aslund, F., and J. Beckwith, “The Thioredoxin Superfamily: Redundancy, Specificity, and Gray-Area Genomics”, *Journal of Bacteriology*, Vol. 181, No. 5, pp. 1375-1379, 1999.

17. Qi, Y., and N. V. Grishin, “Structural Classification of Thioredoxin-Like Fold Proteins”, *Proteins*, Vol. 58, No. 2, pp. 376-388, 2005.

18. Copley, S. D., W. R. Nowak, and P. C. Babbitt, “Divergence of Function in the Thioredoxin Fold Suprafamily: Evidence for Evaluation of Peroxiredoxins from a Thioredoxin-Like Ancestor”, *Biochemistry*, Vol. 43, No. 44, pp. 13981-13995, 2004.

19. Wood, Z. A., E. Schröder, J. Robin Harris, and L. B. Poole, “Structure, Mechanism and Regulation of Peroxiredoxins”, *Trends in Biochemical Sciences*, Vol. 28, No. 1, pp. 32-40, 2003.

20. Fuuji, J., and Y. Ikeda, “Advances in Our Understanding of Peroxiredoxin, a Multifunctional, Mammalian Redox Protein”, *Redox Report*, Vol. 7, No. 3, pp.123-130, 2002.

21. Dombrecht, B., C. Heusdens, S. Beullens, C. Verrreth, E. Mulkers, P. Proost, J. Vanderleyden, and J. Michiels, “Defence of Rhizobium Etli Bacteroids against Oxidative Stress Involves a Complexly Regulated Atypical 2-Cys Peroxiredoxin“, *Molecular Microbiology*, Vol. 55, No. 4, pp. 1207-1221, 2005.

22. Choi, J., S. Choi, J. Choi, M. K. Cha, I. H. Kim, and W. Shin, “Crystal Structure of Escherichia Coli Thiol Peroxidase in the Oxidized State: Insights into Intramolecular Disulfide Formation and Substrate Binding in Atypical 2-Cys Peroxiredoxins”, *Journal of Biological Chemistry*, Vol. 278, No. 49, pp. 49478-49486, 2003.

23. Gabrielsen, M., K. S. H. Beckham, V. A. Feher, C. Zetterström, D. Wang, S. Müller, M. Elofsson, R. E. Amaro, O. Byron, and A. J. Roe, "Structural Characterization of Tpx from *Yersinia Pseudotuberculosis* Reveals Insights into the Binding of Salicylidene Acylhydrazide Compounds", *PLoS ONE*, Vol. 7, No. 2, pp. 1-9, 2012.

24. PyMOL, "The PyMOL Molecular Graphics System", Version 1.7, Schrödinger, New York, NY, USA. http://www.pymol.org/, accessed at January 2014.

25. Kauppi, A. M., R. Nordfelth, U. Hagglund, H. Wolf-Watz, and M. Elofsson, "Salicylanilides Are Potent Inhibitors of Type III Secretion in Yersinia", *Actin*, Vol. 529, pp. 97-100, 2003.

26. Kauppi, A. M., C. D. Andersson, H. A. Norberg, C. Sundin, A. Linusson, and M. Elofsson, "Inhibitors of Type III Secretion in Yersinia: Design, Synthesis and Multivariate QSAR of 2-Arylsulfonylamino-Benzanilides", *Bioorganic and Medicinal Chemistry*, Vol. 15, No. 22, pp. 6994-7011, 2007.

27. Dahlgren, M. K., C. E. Zetterström, A. Gylfe, A. Linusson, M. Elofsson, "Statistical Molecular Design of a Focused Salicylidene Acylhydrazide Library and Multivariate QSAR of Inhibition of Type III Secretion in the Gram-Negative Bacterium *Yersinia*", *Bioorganic and Medicinal Chemistry*, Vol. 18, pp. 2686-2703, 2010.

28. Tsou, L. K., P. D. Dossa, and H. C. Hang, "Small Molecules Aimed at Type III Secretion Systems to Inhibit Bacterial Virulence", *Medicinal Chemistry and Communications*, Vol. 4, pp. 68-79, 2013.

29. Rester, U., "From Virtuality to Reality – Virtual Screening in Lead Discovery and Lead Optimization: A Medicinal Chemistry Perspective", *Current Opinion in Drug Discovery and Development*, Vol. 11, No. 4, pp. 559-568, 2008.

30. Jones, G., P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and Validation of a Genetic Algorithm for Flexible Docking", Molecular Biology, Vol. 267, No. 3, pp. 727-748, 1997.

31. Rastelli, G., “Emerging Topics in Structure-Based Virtual Screening”, *Pharmaceutical Research,* Vol. 30, No. 5, pp. 1458-1463, 2013.

32. Loving, K., N. K. Salam, and W. Sherman, “Energetic Analysis of Fragment Docking and Application to Structure-Based Pharmacophore Hypothesis Generation”, *Journal of Computer-Aided Molecular Design*, Vol. 23, No. pp. 541-554, 2009.

33. Sun, H., “Pharmacophore-Based Virtual Screening”, *Current Medicinal Chemistry,* Vol. 15, No. 10, pp. 1018-1024, 2008.

34. Horwath, D., “Pharmacophore-Based Virtual Screening”, *Methods in Molecular Biology,* Vol. 672, No. pp. 261-98, 2011.

35. Dixon, S. L., A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw, and R. A. Friesner, “Phase: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening. 1. Methodology and Preliminary Results”, *Journal of Computer-Aided Molecular Design*, Vol. 20, pp. 647-671, 2006.

36. Hasegawa, K. and K. Funatsu, "Partial Least Squares Modeling and Genetic Algorithm Optimization in Quantitative Structure-Activity Relationships", *SAR and QSAR in Environmental Research,* Vol. 11, pp. 189-209, 2000.

37. Sastry, G. M., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments”, *Journal of Computer-Aided Molecular Design,* Vol. 27, No. 3, pp. 221-234, 2013.

38. Protein Preparation, “Schrödinger Suite Protein Preparation Wizard“, Version 2013, Schrödinger, LLC, New York, NY, 2013.

39. Ligand Preparation, “LigPrep”, Version 2.6, Schrödinger, LLC, New York, NY, 2013.

40. Watts, K. S., P. Dalal, R. B. Murphy, W. Sherman, R. A. Friesner, J. C. Shelley, "ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers", *Journal of Chemical Information and Modeling*, Vol. 50, pp. 534-546, 2010.

41. Conformation Generation, "ConfGen", Version 2.4, Schrödinger, LLC, New York, NY, 2013.

42. Dixon, S. L., A. M. Smondyrev, and S. N. Rao, "Phase: A Novel Approach to Pharmacophore Modeling and 3D Database Searching", *Chemical Biology and Drug Design*, Vol. 67, pp. 370-372, 2006.

43. Database Screening, "Phase", Version 3.5, Schrödinger, LLC, New York, NY, 2013.

44. Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, D. E. Shaw, M. Shelley, J. K. Perry, P. Francis, and P. S. Shenkin, "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy", *Journal of Medicinal Chemistry*, Vol. 47, pp. 1739-1749, 2004.

45. Friesner, R. A., R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, and D. T. Mainz, "Extra Precision Glide. Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes", *Journal of Medicinal Chemistry*, Vol. 49, pp. 6177-6196, 2006.

46. Molecular Docking, "Glide", Version 5.9, Schrödinger, LLC, New York, NY, 2013.

47. Sherman, W., T. Day, M. P. Jacobson, R. A. Friesner, and R. Farid, "Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects", *Journal of Medicinal Chemistry*, Vol. 49, pp. 534-553, 2006.

48. Sherman, W., H. S. Beard, and R. Farid, "Use of an Induced Fit Receptor Structure in Virtual Screening", *Chemical Biology and Drug Design*, Vol. 67, pp. 83-84, 2006.

49. Pharmacokinetic Properties, “QikProp”, Version 3.6, Schrödinger, LLC, New York, NY, 2013.

50. Singh, S. S., “Preclinical pharmacokinetics: An Approach towards Safer and Enficacious”, *Current Drug Metabolism*, Vol. 7, No. 2, pp. 165-182, 2006.

51. Sastry, M., S. L. Dixon, J. F. Lowrie, and W. Sherman, ”Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments”, *Journal of Chemical Information and Modeling*, Vol. 50, pp. 771-784, 2010.

52. Substructure Search, “Canvas”, Version 1.6, Schrödinger, LLC, New York, NY, 2013.

53. Maestro Interface, “Maestro”, Version 9.4, Schrödinger, LLC, New York, NY, 2013.

54. Jacobson, M. P., D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, and R. A. Friesner, “A Hierarchical Approach to All-Atom Protein Loop Prediction”, *Proteins*, Vol. 55, pp. 351-357, 2004.

55. Binding Energy Calculations, “Prime”, version 3.2, Schrödinger, LLC, New York, NY, 2013.

56. Tanrikulu, Y., B. Kruger, and E. Proschak, “The Holistic Integration of Virtual Screening in Drug Discovery”, *Drug Discovery Today*, Vol. 18, No. 7-8, pp. 358-364, 2013.

57. Wilson, G. L., and M. A. Lill, “Integrating Structure-Based and Ligand-Based Approaches for Computational Drug Design”, *Future Medicinal Chemistry*, Vol. 3, No. 6, pp. 735-750, 2011.

58. Zander, J., M. Hartenfeller, V. Hahnke, E. Proschak, S. Besier, T. A. Wichelhaus, and G. Schneider, “Multistep Virtual Screening for Rapid and Efficient Identification of

Non-Nucleoside Bacterial Thymidine Kinase Inhibitors". *Chemistry-A European Journal*, Vol. 16, No. 31, pp. 9630-9637, 2010.

59. Salam, N. K., R. Nuti, and W. Sherman, "Novel Method for Generating Structure-Based Pharmacophores Using Energetic Analysis", *Journal of Chemical Information and Modeling*, Vol. 49, pp. 2356-2368, 2009.

60. MacroModel, Version 10.0, Schrödinger, LLC, New York, NY, 2013.

61. Abad-Zapatero, C., "Ligand Efficiency Indices for Effective Drug Discovery", *Expert Opinion on Drug Discovery*, Vol. 2, No. 4, pp. 469-488, 2007.

62. Abad-Zapatero, C., and J. M. Metz, "Ligand Efficiency Indices as Guideposts for Drug Discovery", *Drug Discovery Today*, Vol. 10, No. 7, pp. 464-469, 2005.

63. Abad-Zapatero, C., O. Perisic, J. Wass, A. P. Bento, J. Overington, B. Al-Lazikani, and M. E. Johnson, "Ligand Efficiency Indices for an Effective Mapping of Chemico-Biological Space: The Concept of an Atlas-Like Representation", *Drug Discovery Today*, Vol. 15, pp. 804-811, 2010.

64. Hopkins, A. L., G. M. Keseru, P. D. Leeson, D. C. Rees, and C. H. Reynolds, " The Role of Ligand Efficiency Metrics in Drug Discovery", *Nature Reviews Drug Discovery*, Vol. 13, No. 2, pp. 105-121, 2014.

65. Oprea, T. I., "Property Distribution of Drug-Related Chemical Databases", *Journal of Computer-Aided Molecular Design*, Vol. 14, pp. 251-264, 2000.

66. Deng, Z., C. Chuaqui, and J. Singh, "Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions. *Journal of Medicinal Chemistry*, Vol. 47, No. 2, pp. 337-344, 2004.

67. Ajay, G., W. Bemis, M. A. Murcko, "Can We Learn to Distinguish Between 'Drug-Like' and 'Nondrug-Like' Molecules?", *Journal of Medicinal Chemistry*, Vol. 41, pp.

3314-3324, 1998.

68. Xu, Z., G. Yan, G. Wang, B. Li, J. Zhu, P. Sun, X. Zhang, C. Luo, H. Wang, and W. Zhu, "Combining Pharmacophore, Docking and Substructure Search Approaches to Identify and Optimize Novel B-Raf^{V600E} Inhibitors", *Bioorganic and Medicinal Chemistry Letters*, Vol. 22, No. 17, pp. 5428-5437, 2012.

69. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic Local Alignment Search Tool", *Journal of Molecular Biology*, Vol. 215, No. 3, pp. 403-410, 1990.

70. Basic Local Alignment Search Tool (BLAST), "National Center for Biotechnology Information (NCBI)", http://www. ncbi.nlm.nih.gov, accessed at April 2014.

71. Itou, T., T. Mizooku and E. Nishizawa, US0066141 Patent, 2003.

72. Mori, K., Y. Terajima and K. Yomogida, EP-875902-B1 Patent, 2006.

73. Spaulding, L. A., A. R. Frontauria and W. R. Iacona, WO1092911 Patent, 2013.

74. Horvath, D., "A Virtual Screening Approach Applied to the Search for Trypanothione Reductase Inhibitors", *Journal of Medicinal Chemistry*, Vol. 40, No. 15, pp. 2412-2423, 1997.

75. Weidlich, I. E., T. Dexheimer, C. Marchand, S. Antony, Y. Pommier, and M. C. Nicklaus, " Inhibitors of Human Tyrosyl-DNA Phospodiesterase (hTdp1) Developed by Virtual Screening Using Ligand-Based Pharmacophores", *Bioorganic and Medicinal Chemistry*, Vol. 18, pp. 182-189, 2010.

76. Pace, P., S. A. H. Spieser, and V. Summa, "4-Hydroxy-5-Pyrroline-3-Carboxamide HIV-1 Integrase Inhibitors", *Bioorganic and Medicinal Chemistry Letters*, Vol. 18, No. 14, pp. 3865-3869, 2008.

77. QHTS Assay for Inhibitors of Hepatisis C Virus (HCV), “AID 651820”, https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=651820, accessed at July 2014.

78. QHTS for Inhibitors of TGF-b: Cytotox Counterscreen, “AID 588856”, https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=588856, accessed at July 2014.

79. Primary QHTS for Delayed Death Inhibitors of the Malarial Parasite Plastid, 48 Hour Incubation, “AID 504832”, https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid =504832, accessed at July 2014.